SAABM: A framework for combining sentiment analysis and agent-based modelling for dynamic marketplace analysis

A Thesis Submitted for the Degree of Doctor of Philosophy

by

Habiba Daroge

Department of Computer Science

Brunel University

2023

Abstract

The evolution of social media as a digital platform to share consumers experiences of purchasing products has gained a considerable amount of attention from researchers and businesses in recent years. However, the accumulation of large volumes of textual data has presented some challenges in analyzing such unstructured data and capturing complex social phenomena. In light of companies utilizing online platforms to sell their products, the concept of dynamic pricing has been prevalent as a pricing strategy in which the price of a product is continuously adjusted in response to changing market conditions and demand patterns. This element of complexity at which the pace of price fluctuations has evolved over time has resulted in creating advanced models for dynamic marketplace analysis and can be used to understand consumer behaviour and enhance competitiveness for companies. The intersection of dynamic pricing and sentiment analysis presents a unique opportunity to investigate the influence of consumer sentiment on pricing strategies and decision-making processes.

This thesis proposes the Sentiment Analysis and Agent-Based Modelling (SAABM) framework which implements a combination approach of sentiment analysis and Agent-Based Modelling (ABM) to model behavioural complexity. Sentiment analysis and topic modelling is applied in this study based on a case study approach of 100 trainers products including Nike, Puma and Timberland, sourced from Amazon UK consumer reviews. Key insights which were extracted from the data include developing wordclouds for positive, negative and neutral sentiment and applying topic modelling to list the top 10 common topics that were being discussed among consumers about a particular trainers product. Correlation analysis was performed to determine whether there is a correlation between sentiment and price which resulted in a positive correlation for 6 Nike products, 11 Puma products and 11 Timberland products. This exploratory data analysis was used to create an agent-based model to observe interactions between consumers and an Amazon UK seller in a simulation environment.

By incorporating the Bass model (Bass, 1969), coefficient of innovation and social influence were included to investigate what-if scenarios. Visualizations were created to examine how consumers react in the following what-if scenarios: high sentiment and high price, low sentiment and low price, neutral sentiment and neutral price. In addition, to test the robustness of the model, parameter sweeping was implemented which indicated a faster rate of adoption in a smaller market size, increasing the number of innovators increases the social influence diffusion rate and the higher the coefficient of innovation and social influence values result in a higher rate of adoption.

The key contribution of this study is the SAABM framework which is evaluated through the case study approach. One of the findings of the SAABM framework is the integration of sentiment and price variables which influences the adoption threshold of adopting trainers products in a simulated dynamic marketplace environment. Moreover, it provides data-driven insights to be extracted which can aid in data-driven decision making (DDD) to better understand consumer behaviour and to model complex, heterogeneous systems to observe emergent social interactions.

Dedication

Dedicated to my grandfather, Ibrahim Baghdadi, for always instilling the importance of education in us.

Acknowledgements

First and foremost, I am extremely grateful to my supervisors, Dr Anastasia Anagnostou and Dr Crina Grosan for their endless support, advice and patience in the development of my thesis.

I would like to thank the progression review panel, Dr Isabel Sassoon, Dr Stasha Lauria and Dr Diana Suleimenova for their feedback throughout my PhD journey.

I am appreciative of Brunel University's Vice-Chancellor's Conference Prize for Postgraduate Research which enabled me to attend an International conference.

A special shoutout to my cousins Asif Daroge and Farhaan Daroge (may Allah have mercy on him) for acknowledging my position as a Doctoral Researcher on multiple occasions.

Last but by no means least, I would like to thank my good friend Michael for always recognizing my potential, for his consistent encouragement and for his ongoing support.

Declaration & publications

The following papers have been published as a result of the research conducted in this thesis.

- Daroge, H., Anagnostou, A., Grosan, C (2021) ‘An agent-based simulation approach to sentiment analysis in a dynamic marketplace,’ *Computer Science Brunel PhD symposium, Brunel University, London, UK*

- Daroge, H., Anagnostou, A., Grosan, C (2021) ‘An agent-based simulation approach to sentiment analysis in a dynamic marketplace, *Doctoral Research Poster Conference, Brunel University, London, UK*

- Daroge, H., Anagnostou, A., Grosan, C (2021) ‘The application of Sentiment Analysis and Topic Modelling in a Dynamic Marketplace: An Agent-Based Simulation Approach, *Simulation Workshop, UK Operational Research Society, UK.*

Abbreviations

ABM: Agent-Based Modelling

BAG: Bag-Of-Words

BREO: Brunel Research Ethics Online

CRISP-DM: Cross-Industry Standard Process for Data mining

DDD: Data-Driven Decision making

DES: Discrete-Event Simulation

df: DataFrame

LDA: Latent Dirichlet Allocation

LSA: Latent Semantic Analysis

NER: Named Entity Recognition

NLP: Natural Language Processing

ODD: Overview, Design concepts and Details

pLSA: Probabilistic Latent Semantic Analysis

POS: Parts-Of-Speech tagging

SAABM: Sentiment Analysis and Agent-Based Modelling

SD: System Dynamics

SNA: Social Network Analysis

WOM: Word-Of-Mouth

VADER: Valence Aware Dictionary and sEntiment Reasoner

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

This thesis investigates consumer behaviour on sentiment and price using a proposed Sentiment Analysis and Agent-Based Modelling (SAABM) framework which combines sentiment analysis and agent-based modelling (ABM). Sentiment and price are central to this study to investigate the impact of sentiment analysis and dynamic pricing to provide a comprehensive understanding of consumer behaviour leading to data-driven decision-making (DDD) processes.

This introductory Chapter is structured by firstly addressing the motivation and research problem for carrying out this study which is presented in Section 1.2. Section 1.3 highlights the aims and objectives, followed by Section 1.4 which refers to the contributions derived from this research. Section 1.5 addresses the ethical issues mentioned in this research and lastly, Section 1.6 describes the structure of this thesis. Finally, Section 1.7 summarizes this introductory chapter.

1.2 Motivation and Research Problem

With companies such as Amazon increasingly using online platforms to promote their products, online consumer reviews have become a major factor in consumers purchasing decisions (Vimaladevi and Dhanabhakaym, 2012). According to Masters (2019), 89% of consumers are more likely to buy products from Amazon than other e-commerce sites. Amazon is a popular ecommerce website with a large majority of consumers purchasing products via this platform. 93% of customers will read online reviews before making the decision to purchase as opposed to 88% in 2014 (Shrestha, 2016).
In light of the recent Coronavirus pandemic, there has been a surge in online shopping sales globally with potential customers increasingly using online reviews as a factor in their purchasing decisions (Changchit and Klaus 2020).

According to Carter (2022), 89% of consumers view reading online reviews as a necessary tool for purchasing products. Sun, Kang and Zhao (2023) further emphasize this point for companies to take advantage of WOM to promote new products and ultimately for survival in a competitive marketplace. Consumers are utilizing online reviews as a key influence in purchase decisions. These statistics reinforce the importance and interest in analyzing this form of data for both consumers and retailers. Retailers are able to gain insights into consumer experience leading to improved marketing strategies (Ruane and Wallace, 2015; Jain and Reinartz, 2016). MacDonald (2018), states 63% of consumers prefer to shop on websites with product reviews. In addition, by collecting pricing data as well as online review data, retailers can also examine pricing strategies. Moreover, by understanding sentiment derived from online consumer reviews can aid in companies' decision making (Alghamdi, 2013).

Following on from the previous point, 74% of consumers identify word-of-mouth (WOM) from consumer reviews as a key influencer in their purchasing decisions (Kalantaryan, 2022). According to Robertson and Gatignon (1986), potential consumers perceive opinions of individuals who have already purchased the product through WOM as they are unable to evaluate their decision in purchasing the product due to unfamiliarity. Therefore, potential consumers look to trust consumer reviews as a factor of their purchasing decision. Kang *et al.* (2022), mentions individuals even remove products from their online shopping cart based on negative reviews. Hence, with the growth of online shopping, online reviews have become an important platform for potential consumers to aid in their decision-making process.

As these reviews consist of unstructured, noisy text, it is therefore essential to understand and analyze textual data of consumer reviews efficiently. Through Natural Language Processing (NLP), sentiment analysis aims to identify and categorize opinions of a piece of text as either positive, negative, or neutral (Al-Ayyoub *et al*., 2018*).* The study of consumer behaviour has been examined for decades, however, there still seems to be a challenge in understanding consumers buying behaviour and purchase decisions (Alvino, 2019). Traditional approaches to tackle these challenges start showing limitations and there is a need for more advanced, computational methods in understanding consumer behaviour. Søgaard (2022), describes NLP as an emerging field, attracting attention globally in diverse industries. As

well as creating excitement of analyzing textual data, it consequently introduces new challenges in which there is a demand for innovation and overcoming data related obstacles that has an impact on a business (Cao,2016). Therefore, this presents a problem leading to conduct further studies in NLP, particularly sentiment analysis, as demand to understand this concept in increasing. As traditional predictive modelling tools start becoming obsolete due to its lack of features and functions that are needed to tackle existing analytical and methodological problems; there is a need to develop more advanced models that are able to model behavioural complexity.
The rapid pace of price fluctuations has been embraced by online retail markets such as Amazon by amending their prices at the same pace of how the market operates. Dynamic pricing refers to the process of setting and fluctuating product prices over time which can change rapidly and adjusts continuously based on current market factors such as conditions, trends and predictions (Dennis, 2020; Savin, 2019). Zhao (2019) develops a dynamic pricing mechanism model which considers sentiment analysis as the major factor and recognizes that companies which have dynamic pricing mechanisms in place can view the potential of price fluctuations and how this affects customers demand and interest. The advantage of doing this, is that companies remain one step ahead of their competition and contributes to gaining a competitive advantage (Savin, 2019). The application of machine learning on noisy, textual data follows an empirical, data-driven approach for prediction based on a large volume of historical data. Due to the complexity of a dynamic marketplace involving price fluctuations, simulation provides data-driven insights to be extracted using what-if scenarios and essentially provides decision support (Pylianidis *et al.,* 2022). Hence, a multifaceted approach is required in order to address these factors. According to Kelton *et al.,* (2007), simulation has become the preferred tool to investigate complex systems. As Ulbinaite and Moullec (2010), describe a simulation method using ABM as an effective tool to model and analyze consumer behaviour due to its ability to model dynamic and heterogeneous systems in a wide range of domains and is considered as a decision support tool (Shannon, 1975).

As both sentiment analysis and ABM have gained recent attention, there is a challenge in creating a more comprehensive simulation model of consumer behaviour in a dynamic marketplace environment. To address complex, social phenomena such as how sentiment dynamics in agent interactions affect the emergence of collective behaviours and social phenomena in ABM simulations; a combination approach of sentiment analysis and ABM is necessary to overcome these current challenges effectively. Hence, there is a need to investigate understanding how sentiments spread, evolve, and influence the behaviour of agents in a dynamic marketplace context. As the current challenges in analyzing complexity are present, the integration of a combination approach of exploring sentiment analysis within ABM frameworks can evaluate the impact of sentiment on agent behaviours and interactions and the practical applications of sentiment and ABM can advance the understanding of complex social systems and enhance data-driven decision-making. Thus, the identification of a research gap to understand sentiment analysis and dynamic pricing in a marketplace environment needs to be investigated by using a combination approach of machine learning and ABM to tackle these current challenges.

This study addresses this research gap by developing the aim and objectives described in the next Section.

1.3 Aims and objectives

The aim of this thesis is to develop a framework for combining sentiment analysis and ABM for understanding consumer behaviour in a dynamic marketplace.

To achieve this aim, the following objectives have been formulated:

Objective 1: To gain an understanding of the existing research by analyzing and critically evaluating previous studies in the field of sentiment analysis and ABM.

Objective 2: To identify and extract sentiment from textual data collected from online consumer reviews using a NLP approach

Objective 3: To implement ABM for capturing of complex social phenomena for diffusion of innovation.

Objective 4: To develop the SAABM framework for combining sentiment analysis and ABM.

Objective 5: To evaluate the SAABM framework through a case study approach.

1.4 Research contributions

This study attempts to address several research gaps and in doing so makes important contributions which are listed below:

1. Contribution to the body of knowledge
2. Correlation analysis between sentiment and price
3. ABM for diffusion of innovation using the Bass model (Bass, 1969)
4. SAABM framework
5. Evaluate SAABM framework through a case study approach

Contribution 1: Contribution to the body of knowledge.
Although past studies have examined sentiment analysis and ABM of online consumer reviews, the absence of investigating sentiment and price in a dynamic marketplace of trainers products is where this study attempts to address the research gap and contribute to the existing literature.

Contribution 2: Correlation analysis between sentiment and price.
As sentiment is qualitative in nature and dynamic pricing is quantitative, this study employs a mixed-methods research design. This approach seeks to determine whether a correlation exists between sentiment and price for the purposes of exploratory data analysis. This contribution can be fostered on a managerial scale as it allows businesses to be aware of sentiment analysis as a tool to delve deeper into understanding consumer behaviour and allows Amazon sellers to monitor their pricing strategy resulting in DDD (Kim, 2016).

Contribution 3: ABM for diffusion of innovation using the Bass model (Bass, 1969).
The third contribution of this research follows the theoretical lens implemented for ABM through the Bass diffusion of innovation model (Bass, 1969). It has been regarded as one of the most prominent paradigms employed and refined in many fields such as marketing, sociology, and economics (Chang, 2010). This theory supports the prediction of when consumers will adopt an innovation such as trainers products, in addition to simulating and observing complex social phenomena such as social interactions between consumers and sellers.

Contribution 4: SAABM framework

The main contribution of this research is the development of the SAABM framework which reconceptualizes existing frameworks to integrate a combination approach for sentiment analysis and ABM by including sentiment and price variables. Therefore, this study attempts to address this gap in research by collecting data on trainers products from the Amazon UK website. The implications of the results derived from this research contributes to further understanding consumer behaviour through the development of the SAABM framework which will explore what-if scenarios to predict the effects of sentiment and price on the rate of adoption as well as parameter sweeping to test the robustness of the SAABM model.

Contribution 5: Evaluate SAABM framework through a case study approach

By conducting a case study approach of online Amazon UK reviews of trainers products, evaluation of the SAABM framework can be performed.

1.5 Ethical issues

This research study has gained ethical approval from Brunel Research Ethics Online (BREO). For the purposes of this research, online consumer reviews from the Amazon UK website will be collected which consists of the brand of the product, ratings, reviews and date. Individuals who posted the reviews remain anonymous.

1.6 Thesis structure

Chapter 1 presented motivation for carrying out this research as well as the research problem. This Chapter also introduced the aims and objectives which included inspiration for developing the SAABM framework. In this Section, the contributions that this thesis has proposed to the body of knowledge are highlighted, which encompass methodological and practical contributions. Lastly, ethical issues were stated.

Chapter 2 presents a literature review concentrating on the main areas of this study: diffusion of innovation, NLP and simulation. The Chapter is organized into four themes. The first theme introduces the concept of diffusion of innovation which look at the factors that influence the rate of adoption. The second theme explores NLP and how sentiment analysis and topic modelling are applied to unstructured, textual data such as online consumer reviews. Thirdly, simulation techniques are examined which include discrete-event simulation (DES), system dynamics (SD) and ABM followed by simulation software packages. Lastly, a discussion of the literature highlights studies explaining diffusion of innovation, NLP and simulation on consumer behaviour.

Chapter 3 introduces the methodology and investigates research designs. The justification to adopt a case study approach is presented along with a case study description. It proposes the Cross-Industry Standard Process for Data Mining (CRISP-DM) developed by IBM (1996) as a structure for the SAABM framework.

Chapter 4 presents an adaptation of CRISP-DM for exploratory data analysis. Sentiment analysis and topic modelling are applied as a NLP tool to explore sentiment derived from online consumer reviews. After this, correlation analysis is conducted to investigate whether there is a correlation between sentiment and price for the 100 trainers products collected. This analysis is used for the development of the SAABM framework.

Chapter 5 examines ABM on diffusion of innovation. The mechanics of diffusion of innovation are examined through examples of models. An agent-based model is developed without sentiment and price variables to observe the impact on the simulation in the absence of these two variables. The results are presented in the output analysis Section.

The last 2 Chapters build the foundation for the development of the SAABM framework which is presented in Chapter 6. The processes are illustrated diagrammatically, and design concepts explained. In contrast to the model built in Chapter 5, the SAABM framework incorporates sentiment and price, and observations are presented in the output analysis Section. Parameter sweeping is conducted to compare how the changes in parameters affect the simulation on rate of adoption and diffusion rate of social influence. Replication is mentioned as a tool to evaluate the framework. Lastly, contributions are stated along with a summary of the Chapter.

Chapter 7 concludes this research study by summarising the achievement of the aim and objectives. Furthermore, contributions are stated and lastly, limitations of this study pave the way for future work in this area.

1.7 Chapter Summary

This introduction chapter has provided an overview of the structure of this chapter, highlighted the motivation to investigate the research problem of the challenges presented of analyzing textual, online review data and exploring the social interactions in a dynamic marketplace. Thirdly, the aim and objectives have been explained and research contributions illustrated. Ethical issues have been stated and ethical approval has been obtained. Lastly, the thesis structure explains what each of the chapters entail. The next Chapter presents the literature review.

Chapter 2

Literature Review

2.1 Introduction

This Chapter reviews existing research studies on sentiment analysis and simulation on consumer behaviour. This Chapter is structured as follows: Section 2.2 gives an overview of diffusion of innovation with Rogers diffusion of innovation model (2003) being explained in Section 2.2.1. The Bass model (Bass, 1969) is another type of diffusion of innovation model which is explained in Section 2.2.2. Section 2.3 presents NLP techniques followed by simulation methods presented in Section 2.4. Section 2.5 identifies software packages suitable for the application of sentiment analysis and ABM. Section 2.6 presents data mining methodologies. Section 2.7 explores the methodology for the simulation study. Section 2.8 presents an overview of the discussion of literature. Section 2.9 highlights research gaps identified from the literature review and lastly Section 2.10 gives a summary of the Chapter.

2.2 Diffusion of Innovation

The diffusion of innovation is explained as a social process in which individuals respond to an innovation being introduced in a marketplace. The communication of an innovation is important as it is this element that can increase or decrease the rate at which consumers adopt the product (Dearing and Cox, 2018). Models used by businesses need to be able to include the communication channels within the marketplace such as WOM social influences. By analyzing the spread of innovation, diffusion of innovation can be examined. Two common models of diffusion of innovation include Rogers (2003) and Bass model (Bass, 1969) which are explained in further detail below in Subsections 2.2.1 and 2.2.2.

2.2.1 Rogers Diffusion of Innovation model

Rogers (2003) is a common diffusion of innovation model concentrating on the adoption of new innovation based on their stage of adoption and has been implemented in a variety of contexts such as economics, technology, politics and education to name a few (Dooley, 1999). According to Rogers (2003), "An innovation is an idea, practice, or project that is perceived as new by an individual or other unit of adoption." Furthermore, adoption is described as a decision and not adopting leads to the decision of rejecting an innovation. In addition, diffusion is defined as "the process in which an innovation is communicated through certain channels over time among the members of a social system." (Sahin, 2006). An advantage of diffusion of innovation theory is the time aspect the model integrates. According to Rogers (2003), the time feature is overlooked in the majority of behavioural research as the diffusion of innovation process as categorizing adopters and rate of adoptions all include a time aspect.

Anderson *et al.* (1998), analyzed attitudes, skills and behaviour of staff using information technology at University of Alberta. The study implemented Roger's (1995) adopter categories as earlier adopters and mainstream faculty. The results showed the adoption of mainstream faculty to be very low and to increase adoption incentives and training should be examined (Sahin, 2006).

Surendra (2001) studied Rogers (1995) diffusion on predicting web technologies usage by faculty members in a college in Ontario. Access to information was the most significant diffusion factor and administrators were more open to web based educational innovation than professors. The findings of their research concluded that the more positively faculty members perceive diffusion factors of web technologies, the more they are willing to accept the adoption of innovation. Secondly, no relationship was found between years of service and acceptance of innovation, no relationship was found between years of computer usage and acceptance of innovation, no relationship was found between age and acceptance of innovation and lastly, a relationship was observed between the computer and acceptance of innovation.

Rogers' diffusion of innovation theory (2003) suggests there are five types of adopters; innovators, early adopters, early majority, late majority and laggards (Doyle, Garrett and Currie, 2014) which are described below and summarized in Table 1.

- Innovators

Innovators are present at the start of the adoption process and are keen to adopt innovations. They are willing to spend on trying new ideas while being placed on the outside circle of social relationships (Robinson, 2009). They are the first to adopt the product and therefore, willing to take risks. They are open-minded, trend setters and buyers have no indication whether the market will adopt or reject the innovation.

- Early adopters

Early adopters adopt once the benefits of innovation becomes clear. They have sufficient time and money to invest and are a leader in expressing their opinions by providing information about the innovation to potential adopters and may post online reviews about their experience with the product. Although, they do not take as many risks as innovators, they require more information than an innovator in order to make a decision whether to adopt or not to proceed with the adoption of the innovation (Robinson, 2009).

- Early majority

This category of adopters takes longer than innovators and early adopters to adopt the innovation. They dislike complexity with restricted time and money and want proof of the benefits and value for money of the innovation in order to adopt.

- Late majority

The late majority group are skeptical about adopting an innovation and take longer than the previous adopter types to adopt a new product. They typically have less money and interaction with innovators and only invest in solutions that they know have been successfully implemented. Therefore, they require strong peer pressure to persuade late majority to accept innovations.

- Laggards

Laggards are typically the last group of adopters to adopt an innovation and signal that the product is starting to decrease its status as more advanced technologies are being introduced into the market. They are known to be traditional and highly resistant to change. They rarely listen to other opinions and have criticisms of the innovation.

Figure 1 shows these five categories in terms of period adoption and cumulative adoption over time. Adoption is at its peak during the early majority and late majority phases which could be due to social variables such as an increase of product awareness at this stage and an increase in adoptions taking place (Pashaeypoor et al, 2014).

Table 1: Rogers' diffusion of innovation theory (2003) adopter types

Adopter type	Description
Innovators	First to try the innovation and will subsequently influence potential adopters. Risk takers who enjoy being at the cutting edge of the innovation.
Early adopters	Comfortable with adopting the innovation by basing their adoption decision from innovators.
Early majority	Willing to adopt change but slower to adopt than the early adopters. Likely to adopt before the average person.
Late majority	Quite resistant to change so slower at adopting. Bases decision of adoption on the innovation being accepted by the majority of the population.
Laggards	The last to adopt the innovation with a strong resistance to change.

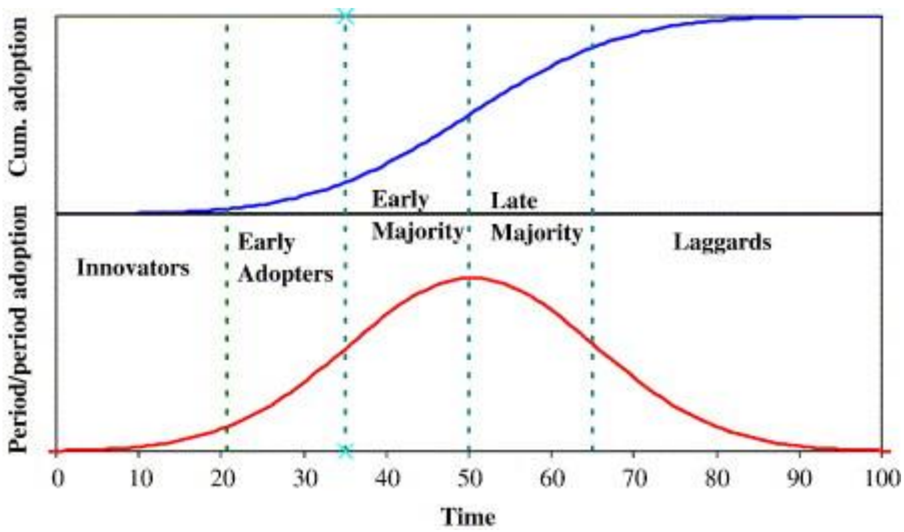


Figure 1. Rogers' diffusion of innovation theory (2003) adapted by Meade and Islam, (2006)

2.2.2 Bass Diffusion Model (Bass, 1969)

Another diffusion of innovation theory is the Bass diffusion model (Bass, 1969) and has been considered as a popular tool that has been widely accepted within academia (Boehner and Gold 2012), to study and forecast diffusion of innovations. It has been applied to a wide range of products including, smartphones, home appliances and non-durable products (Bass, Krishnan, Jain 1994).

According to Norton and Bass (1992), the diffusion of innovation theory consists of 2 adopters. The first adopter type is referred to as innovators or the coefficient of innovation and are described as individuals who have already adopted the product. The second adopter type are described as imitators or the coefficient of imitation and are individuals who are socially influenced to adopt the product as shown in Figure 2.

The Bass model (Bass, 1969) diffusion curve is shown in Figure 3, which describes the number of new adopters over a period of time. According to Bass model (Bass, 1969), the number of innovators who have already adopted the product and pass the message to new potential adopters decrease due to less potential innovators. In contrast, the number of imitators who spread the message through WOM channels increase, enabling a higher influence on imitators to adopt the product. Figure 3 shows imitation influence increases to a peak and then gradually decreases as the market gets saturated (Ismail and Abu, 2013).

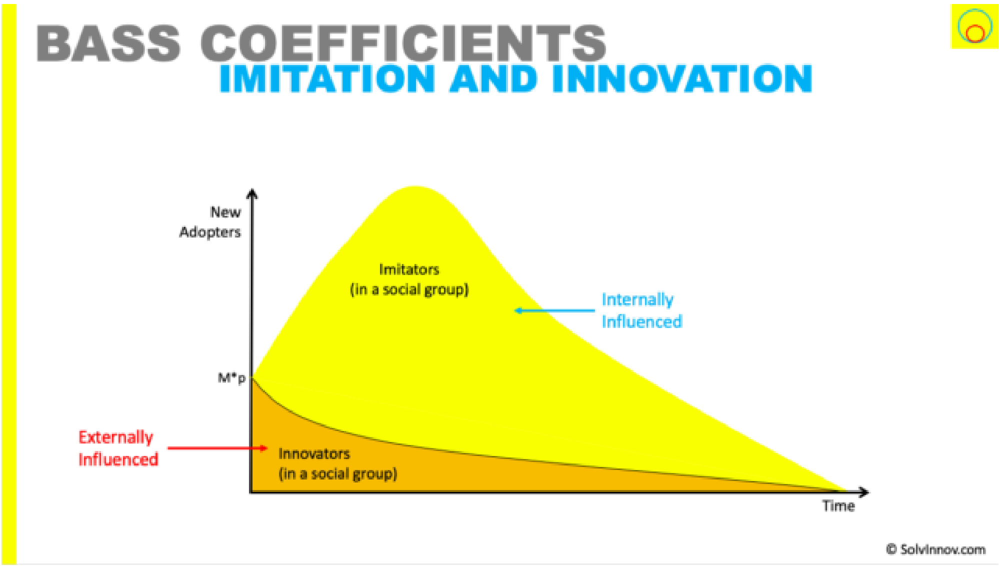


Figure 2. Bass model (Bass, 1969) showing the impact of coefficient of innovation and coefficient of imitation in a market (Tacy, 2021)

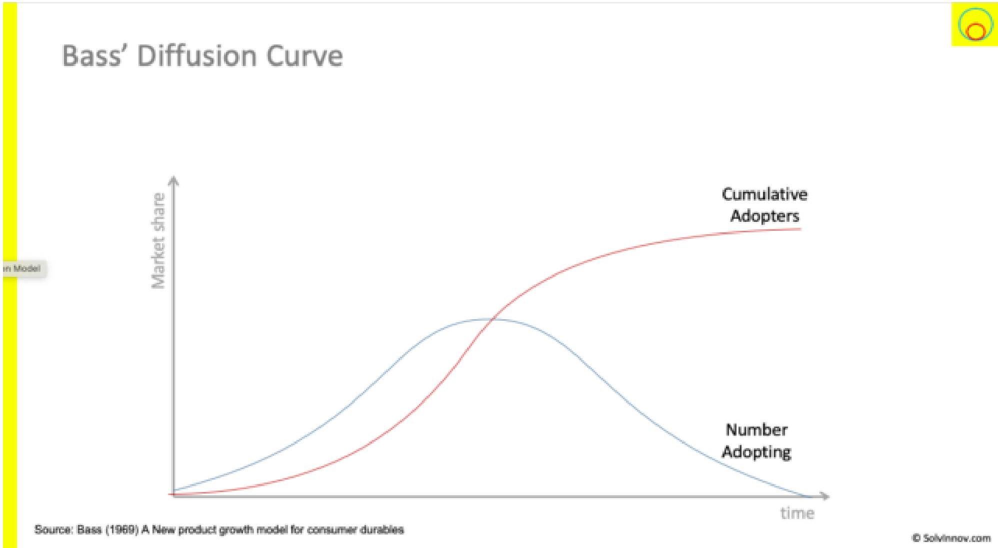


Figure 3: Bass model (Bass, 1969) Diffusion Curve (Tacy, 2021)

The main idea of the model is that the adoption rate of a product comes from two sources: the tendency of consumers to adopt the product independent of social influences and adopt the product because others have adopted which is referred to imitators. One of the assumptions of the Bass model Bass, 1969) is "…the probability of additional first-time adoptions of a new product in a future time is the function of the number of consumers who have already adopted the product" (Wilensky and Rand, 2015). The other assumptions of the Bass model (Bass, 1969) are as follows:

- There are two types of behaviour investigated which are innovation and imitation. Innovation represents individuals who already know about the product and have adopted it. They are also described as a broadcast or external influence as they spread the message (Wilensky and Rand, 2015). Imitation is often described as the social influence because they recommend the product to others through internal influence such as WOM.
- The diffusion process consists of either the consumer adopts, or the consumer waits to adopt.
- All potential adopters eventually make the decision to purchase.
- Repeat purchases are not accounted for.

According to Charan (2015), the basic Bass model (Bass, 1969), is represented by the following equations:

$$n(t) = adopters\ via\ innovation + adopters\ via\ imitation$$

$$adopters\ via\ innovation = p * remaining\ potential = p[M - N(t)]$$

$$adopters\ via\ imitation = q * proportion\ of\ adopters * remaining\ potential$$

(Eq.1)

Where:

n(t) = number of adopters at time t.

N(t) = n (0) + n(1) + n(2)… n (t).

M = total number of potential adopters.

p = coefficient of innovation

q = coefficient of imitation.

Mahajan and Muller (1979), stated in their research on innovation diffusion and new product growth models in marketing that the objective of diffusion model is to represent the level of spread of an innovation among a given set of prospective adopters over time. Bass (1969) assumed that potential adopters of an innovation are influenced by two types of channel communications. Adopters that receive innovation from mass media and other external factors are called innovators while adopters that receive innovation from WOM and observation or other interpersonal factors are called imitators. (Ismail and Abu, 2013).

The spread of innovation can be examined and analyzed using the diffusion of innovation theory (Rogers, 1962) where the rate of technological advancements is measured. By applying this framework, features that could have an influence on the rate of adoption can be investigated (Ratcliff and Doshi, 2013). However, as companies consist of complex systems, it is challenging to predict how innovations diffuse in a dynamic environment. Therefore, numerous factors need to be acknowledged (Frederiks, Stenner, Hobman, 2015), hence, the Bass model (Bass, 1969) created a framework to take into account social influences such as WOM. This is important as Kiesling (2011) describes, "…social influences are a market-dominating factor in the innovation diffusion paradigm." The Bass model (Bass, 1969) has been used extensively over the past 40 years for modelling the adoption of products and services such as consumer durables, computers, technological products, books, movies and more (Charan, 2015).

Teng, Grover and Guttler (2002), conducted a study using the Bass model (Bass, 1969) for comparison of diffusion behaviour involving twenty-five information technologies. Their findings suggest coefficient of imitation was higher and was the main source for adoption rather than coefficient of innovation which had a lower value. The clusters identified showed the result as email in one cluster with low coefficient of imitation and high coefficient of innovation, computers as high coefficient of imitation and low coefficient of innovation and imaging as high coefficient of imitation and coefficient of innovation.

One of the weaknesses of diffusion of innovation theory, is the focus of diffusion of innovation concentrates on the product and innovation and neglects how other factors determine how a product or innovation is adopted (Al-Mamary *et al.,* 2016). In addition, some weaknesses of the Bass model (Bass 1969) are it does not depict individual interactions which is important in understanding diffusion of innovation (Valente, 2005). Another criticism of the Bass model (Bass, 1969) is that it does not consider decision variables which limits a firm's capability of decision-making. Nevertheless, Mahajan *et al.* (2000), and Meade and Islam (2006) propose incorporating the marketing mix variables into the model such as price and advertising.

2.2.3 Justification for Bass model (Bass, 1969)

The Bass model (Bass, 1969) has been applied in numerous empirical studies within different fields enhancing its credibility and reliability as a predictive tool. Unlike other diffusion models, the Bass model (Bass, 1969) is behaviourally grounded in the idea that innovators and imitators both contribute to decision-making for adoption. By adjusting the model's parameters, this research can observe the influence innovators and imitators have on the diffusion rate of trainers products. Furthermore, different what-if scenarios can be undertaken such as changing sentiment and price variables and how this effects adoption dynamics. As the Bass model (Bass,1969) incorporates time this has the advantage of observing how innovation diffuses over time which gives an insight into peak adoption rate and total number of adopters over time. On a managerial scale, the model can provide valuable insights by understanding the potential market size, timing of peak adoption, and the effectiveness of marketing strategies.

As well as the strengths of Rogers (2003) diffusion of innovation theory as mentioned above, there are various limitations. As Rogers (2003) has a more classical approach founded on normal adopter distribution, the Bass model (Bass, 1969) considers the communication process (Mahajan, Muller and Srivastava, 1990). As this study aims to observe the interactions between consumers, seller, and the environment, it is important to be able to investigate imitation factors. As diffusion of innovation theory is based on an organizational level and not the individual level, this study aims to fill this research gap through the SAABM framework by developing an agent-based model to investigate behaviour on a more individual level (Oliveira and Martins, 2011). In addition, as the Bass model (Bass, 1969) is versatile, it will be adopted into this study as it is able to factor in elements such as WOM and social influences. Furthermore, as this study is more interested in exploring the effect of innovators and imitators through social network interaction and not adopter types as explained in Rogers (2003), the Bass model (Bass, 1969) is more suitable to be implemented in this study. As Mahajan *et al.,* (2000), and Meade and Islam (2006) state other variables such as sentiment and price can be added to the Bass model (Bass, 1969) is another justification to use this method for this research. The introduction of sentiment analysis as a NLP tool is explained in the next Section.

2.3 Natural Language Processing (NLP)

According to Bird, Klein and Loper (2009), NLP is experiencing growth in different fields and industry which include scientific, economic and social aspects. Therefore, the human-computer interaction is important for many people to have a working knowledge of NLP.

The following are techniques of NLP: speech recognition, Parts-Of-Speech tagging (POS), Named-Entity Recognition (NER), machine translation, sentiment analysis and topic modelling which are further explained below:

2.3.1 Speech recognition

This NLP tool Involves converting voice commands into text data. This is a particularly challenging task as the way people talk can be unclear such as talking quickly, mumbling words and different accents or incorrect grammar (Kruglyak, 2020). Speech recognition has been applied to a variety of settings. Voice biometrics has been used in online banking and computer security through a voice interpretation algorithm which uses stored voice snippets to monitor the frequency of voice segments through the uttering of various words. Another application is virtual assistants such as Siri which was developed by Apple. Siri allows the voice of the user to be captured to place phone calls, text contacts or converts speech into a text format. Speech recognition is increasingly being applied to homes through voice activated alarm and through smart technology. For example, using the phone to switch the lights on or off. This type of authentication adds another layer of security. In addition, Global Positioning System (GPS) include voice-activated navigation systems, so drivers do not have to manually enter location details, hence improving safety. Another application is in call centres where it is able to transcribe phone calls as speech recognition translates speech from verbal to written text and common queries are able to converse with users through AI chatbots.

2.3.2 Parts-of-Speech tagging (POS)

POS includes grouping words which are according to parts of speech based on a sentence. It is used to identify and extract specific information in a piece of text such as name, location which can be used for developing knowledge bases. The application of POS tagging is in NER and machine translation which are discussed as the next two types of NLP techniques. POS tagging uses machine learning algorithms by training "a large, annotated corpus of text" (Harsha, 2022) and predicting the POS tag for a specific word based on its context. POS tagging is used to understand the grammatical structure of a system by labelling each word and how words relate with each other. It could also be used to characterize homonyms such as 'I am going to *park* the car' and 'we are going to the *park'* depending on the context. Hence, by labelling words using POS tagging, words can be disambiguated, and their intended meaning can be identified more accurately. It can also be used to improve NER and text classification by providing context about the words. Insights can be gained by identifying patterns of language. Some of the steps in POS tagging is firstly data collection. The data collected consists of annotated text which will be used for training and testing purposes for the implementation of the algorithm. POS tags should be annotated for each word. Next, pre-processing splits the text into tokens to ensure text data is clean and ready for modelling. Some tasks to clean the data are, removing punctuation, converting text into lowercase and removing numbers. More information of text pre-processing is mentioned in Chapter 4 where it is regarded as a crucial step performed for sentiment analysis. Next, the processed dataset is split into training and testing datasets. The training dataset will train the POS tagger on the annotated text by creating a statistical model or through a rule-based method (Harsha, 2022). The test set uses the trained model to predict POS tagging on text and performance measured. The next step would be to evaluate and adjust the model if necessary.

2.3.3 Named-Entity Recognition (NER)

This method involves extracting names of locations, people and objects from a piece of text and categorizing them under headings. NER is used in Search Engine Optimization (SEO), customer support and more and is one of the common entity detection methods in NLP (Goyal, 2021). NER automatically scans text and extracts entities such as organization, name, company name, dates, and time. It then classifies the extracted entities into categories and stores them in a database. Some of the applications NER has been applied to are customer support to aid in complaint queries by identifying specific products that consumers are complaining about which helps the customer support team to transfer customers to the appropriate support. As previously mentioned in the application of speech recognition, NER can identify entities in consumer queries through chatbots to understand the context of the conversation hence, improving responses. In a financial context, figures can be extracted from financial reports, thereby being able to quickly identify what is needed and for example can be used to monitor trends in stock prices faster. In an educational setting, researchers can use NER to identify relevant subjects and topics. In addition, NER is increasingly used for recommendation engines to improve and provide personalized recommendations through user searches which leads to using NER for search engines to provide relevant results (Barney, 2023). Some of the benefits of NER are it saves time by automating the extraction process and analyzes important information from unstructured textual data. However, there are some limitations of this approach. A large volume of data may be required for the NER model to be trained on which involves humans manually annotating the text before the model is able to predict. It can also have problems with spelling and spoken word.

2.3.4 Machine translation

This type of NLP technique consists of translating one language into another as well as having the ability to preserve the meaning. It does this by recognizing the grammatical structure of a sentence and the correlation between words. The advantages of machine translation are automated translation which makes it time efficient and cost-effective. Similarly, to NER, machine translation requires a large amount of data to train the model (Jiang and Lu, 2020). Some of the use cases of machine translation are for more effective and efficient communication. Machine translation aims to remove language barriers by translating documents into other languages which can be used in business for stakeholders and customers. Another example is for online reviews for use on the Amazon website which can translate reviews of products from one language to another so potential consumers can read reviews in their own language. In terms of customer service, requests for translation can occur from anywhere in the world.

2.3.5 Sentiment analysis

One of the most common techniques of NLP is sentiment analysis which analyzes and classifies a piece of text as positive, negative or neutral to determine the sentiment of a piece of text. NER as discussed earlier is used in sentiment analysis to extract “product names, brands and other information mentioned in customer reviews, social media posts and other unstructured text” (Barney, 2023). Other use cases of sentiment analysis are it is used to determine sentiment in questionnaires and complaints. There has been more research into feature extraction as product review datasets are being increasingly developed (Liu et al., 2005; Hu and Liu, 2004; Popescu and Etzioni, 2005). Sentiment analysis is mainly conducted for social media analysis. Some of the use cases of sentiment analysis include the opinions of consumers for a product or brand and understanding customer satisfaction which can be very useful information for companies to acquire. Feedback gathered can enable companies to adjust their strategy. For example, if consumers are portraying negative sentiment about a particular product, they can take the necessary steps such as making improvements to the product to be released back into the market. This enables a quick reaction to public sentiment by utilizing DDD. “Businesses can monitor metrics such as brand mentions and sentiments associated with each mention” (Sayedi, 2021). In terms of customer service, sentiment can be extracted to evaluate consumers opinions on satisfaction. Due to the complexity of natural language used by humans, there are still significant improvements that can be made in the field of sentiment analysis contributing to DDD for businesses (Sayedi, 2021).

2.3.6 Topic modelling

Another form of NLP is topic modelling which discovers topics in a piece of text. It is a machine learning technique that uses automation to identify topics based on detecting patterns and recurring words in a piece of text (Murakami *et al.,* 2017). Some of the applications of topic modelling are brand monitoring where key insights can be extracted about a brands image and positioning in a marketplace. For example, according to Swaminathan *et al.,* (2022) "By identifying…themes that emerge from topics, a brand can situate its marketing communications within these topics or emergent themes, thus improving its target market appeal." This is also useful for marketing companies to incorporate stories associated with the brand which has the advantage of creating a relationship with consumers feeling that they can relate more to the product (Fournier and Avery, 2011). Topic modelling has emerged as a Latent Dirichlet Allocation (LDA) method and has been increasingly applied to online social media platforms for opinion analysis, image recognition, marketing and more. Due to topic evolution, specific topics that online users are discussing can be observed as well as identifying trends in topics (Nikolenko, Koltcov, Koltsova, 2017). Previous research on LDA consisted of identifying topics on Twitter (Kim, Jeong, Kim, Kang, and Song 2016), event detection (Qian *et al.,* 2016) and emotion classification (Roberts et al., 2012).

As technology is becoming increasingly utilized globally, emphasizing the importance of human-computer interaction, NLP aspires to design artificially intelligent systems that "are capable of using language as fluently and flexibly as humans do" (Dale, Somers and Moisl, 2000).

2.3.7 Justification for using sentiment analysis and topic modelling for the current study

Referring to the research problem, as this study is interested in analyzing and extracting sentiment from online consumer reviews, sentiment analysis is the most applicable NLP methodology to use. In addition, as mentioned in Chapter 1, this study will examine trainers which looks at three brands, Nike, Puma and Timberland. As topic modelling, identifies themes that emerge from topics it will be interesting to see if brand is frequently mentioned in the online consumer review dataset. By applying sentiment analysis and topic modelling a more comprehensive overview of sentiment and topics being discussed among consumers can be observed. To denote how individuals feel about products or services can create an enhanced contextual understanding which can be effective for DDD. Sentiment analysis is beneficial to be implemented in this study in order to detect whether a correlation exists with dynamic pricing. Deriving sentiment whether positive or negative will be used to observe fluctuations in price and hence, correlation analysis will be performed. Furthermore, as topic modelling identifies the top common topics it will improve understanding of what consumers are saying about the trainers products. This enables a more nuanced perception by creating a holistic view of textual data and extracting insights from such unstructured data. For the purposes of this study sentiment analysis and topic modelling are further discussed below.

2.3.7.1 Sentiment Analysis

Sentiment analysis can be defined as identifying and extracting the subjectiveness and polarity of a piece of text which can take many forms of unstructured data such as reviews, interviews, surveys etc. As social media platforms are gaining popularity there is a multitude of data that can be analyzed to determine whether a piece of text is either positive, negative or neutral. Sentiment analysis can be beneficial to companies to help them in the decision-making process, for example, if public sentiment is being exposed as negative, the company may make some adjustments to the product to create a more positive experience (Rojewska, 2023).

Sentiment analysis is one of the most common tools in NLP as it helps to generalize the opinion of a particular text about a certain topic. As people are expressing their opinions over social media platforms, there is a need to analyze and gain insights into these sentiments which helps a business to understand the social sentiment of their brand, product or service. After a customer has purchased a product, they have the option to leave an online review which influences potential consumers purchasing decision (Jack, L; & Tsai, 2015). This act of leaving a review is fundamental to a business as it can cause a positive or negative impact on the reputation of the company. Therefore, applying sentiment analysis as a tool of NLP can have major advantages to a business. Hence, there is a need to explore sentiment analysis on online reviews where this study looks at examining sentiment analysis based on online reviews written by users on the Amazon UK website. The concept of sentiment analysis has been growing since 2000 and is increasingly being utilized to uncover and analyze attitudes on consumer behaviour in the form of users' opinions (Liu, 2012).

A key part of NLP is discovering how positive or negative the audience feels about a specific topic whether it be tweets, movies or in the case of this research product reviews on trainers. Sentiment analysis is therefore frequently applied in order to gain insight into the opinions of the public (Jason, 2018). Furthermore, Pang and Lee (2008) also illustrate two other factors which lead to a huge burst of research of sentiment analysis such as the growth of machine learning methods in NLP, as there is in upsurge in consumer review websites and an increase in the volume of data with which machine learning algorithms can be trained on.

2.3.7.2 Topic modelling

Topic modelling refers to an unsupervised machine learning tool which detects words and phrases from a set of documents and clustering them into topics. In machine learning and NLP, topic modelling is an unsupervised statistical method for discovering abstract topics that exist within a collection of documents. It scans or mines text to detect frequently used words or phrases and groups them to provide a summary that best represents the information in the document.

The three most common topic modelling methods are Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA) which are described below.

- LSA is used to assess the relationship between documents and the topics contained within it through the application of Singular Value Decomposition (SVD) which uncovers the meaning of words in documents (Qualtrics, no date). However, the advantage of applying this method is due to the large volume of data, more computing resources such as time and storage is necessary (Gandhi, 2022).
- The aim of pLSA is to model co-occurrence data using a probabilistic framework to uncover the underlying semantics. Similarly, to LSA, pLSA reduces its dimensionality by factorizing the sparse co-occurrence matrix (Oneata, 2011). However, pLSA is a more favoured method than LSA providing a probabilistic measure whereas LSA uses factorization using mathematical concepts such as SVD.
- LDA is a popular tool that identifies topics from documents by observing word clusters and word frequencies. LDA estimates words associated with a certain topic and the topics that describe a piece of text (Pascual, 2019). LDA uses a bag-of-words approach (BAG) that

"is a representation of the topic model that highlights key associations for each topic that is linked to a brand" (Swaminathan *et al.,* 2022). An example of an application of LDA is in the customer service field. It can be used to understand and analyze customer complaints through automation (Bastani, Namavari and Shaffer, 2019).

One of the advantages of topic modelling is the ability to automatically scan through various documents to identify trends and topic patterns for example what consumers opinions are about a particular trainers product. Another advantage is that it can create a competitive advantage for companies by gaining a better understanding of what consumers think about a product and hence use this insight to enable DDD. This approach is most suitable to be used in this study as it is "one of the most successful topic models to infer the topics discussed in a collection of documents" (Naskar *et al.,* 2016). In addition, combining topic modelling with sentiment analysis, more topics that are being discussed within a piece of text can be unearthed.

2.4 Introduction to Simulation

Simulation can be used as a powerful tool for modelling and analysis in various fields such as engineering, healthcare, social sciences and economics. Christie et al (2005) highlights the demand of large-scale computer-based simulations being used to predict the behaviour of complex systems. According to Lateef (2010), simulation amplifies real experiences to optimize complex systems. The concept of simulation represents a system or process which reflects its key characteristics and behaviours which offers a versatile and adaptable approach to problem-solving (Negahban and Smith, 2014). Furthermore, simulations can use experimentation or what-if scenarios to test hypotheses and assess the impact of different variables. This is valuable to researchers and businesses to gain insights into the dynamics of systems and to aid in DDD. It can be used as a technological tool to develop, plan and design complex systems to optimize decision-making to evaluate risks, costs and impact on operational performance (Ferreira et al 2020). The most common simulation methodologies used in consumer behaviour is discrete-event simulation (DES), system dynamics (SD) and ABM.

According to Barrett *et al.,* (2008), DES can be applied to real world systems consisting of a queuing theory. For example, each event occurs on a specific process and the result of each event can be an outcome passed onto the next process. DES can be deconstructed into separate processes that autonomously progress through time. Events occur at specific points that change at a specific time and this type of simulation is applied to model these time instances and does not track the state of the system continuously. DES tracks the behaviour of individuals. A DES model typically includes two sets of dynamic objects: entities and resources. Entities are individuals whose behaviour is tracked in detail, and resources are countable items, and their behaviour is not tracked in the model (Pidd, 2004). The changes in state and events where events occur at specific time instances and are tracked continuously is the focus of DES. Following on from this, DES is heavily applied at operational levels such as queuing theory where arrival and service are modelled for example in the context of healthcare where patients arrivals, stay duration of the patient and when they are discharged is observed. In addition, DES is used in queue optimization such as queues in a bank or customer support centre. This theory allows companies to monitor and assess the systems performance and making any adjustments required to eliminate bottlenecks.

Several aspects of human behaviour have been included in DES models, however, the model doesn't take into account interactions. For example, Garnett and Bedford (2004), describe four main factors of human behaviour in DES models in the context of healthcare: human, physiological and environmental, decision-making processes and human psychology factors. The limitations of this are it focusses on the structure of the job rather than individuals who perform the job. By investigating these factors, DES ignores interactions with other healthcare employees working at the company. As DES doesn't allow interactions to be modelled, DES will not be used in this study.

In addition, a study by Günal and Pidd (2006), similar to the previous study and in the context of the healthcare sector take into account human, physiological, environmental and decision-making factors in a DES model. The human factors consisting of doctors multitasking behaviour and workload capacity which is a good example of representing human behaviour and decision-making processes in one model. However, examining interactions are limited and therefore doctors interactions with patients, staff and the environment cannot be examined where these factors can possibly be used to investigate effect on multitasking behaviour. To summarize, DES had been widely applied to modelling healthcare systems such as Accident &

Emergency (A&E) departments, nevertheless, aspects of human behaviour may not be able to be examined using DES models.

SD is another type of simulation modelling where behaviour is examined through the structure of the system consisting of stock and flow networks and information flows rather than an individual level (Sterman, 2000). SD looks beyond separate events and decisions to see the system structure underlying decisions and their combined effects on system performance while, DES focuses on events and the decisions generated from those events. On the other hand, similarly to DES, SD has not been applied to uncover the interactions of individuals behaviour. Therefore, more advanced modelling is needed to model interactions between individuals and the environment which may contribute to the decision-making process. Interactions such as these and interactions with the environment can therefore be modelled using ABM.

2.4.1 ABM

As mentioned in the Chapter 1 data is growing exponentially in numerous fields whether it be marketing, physics, economics through to political science. As this new data is captured, questions arise when analyzing the data and what-if scenario's can be implemented to delve deeper into these questions. As these complex questions are put into practice, ABM can aid in understanding these complex systems (Wilensky and Rand, 2015).

Wilensky and Rand (2015), describe ABM as a new way of doing science by carrying out computer-based experiments. They describe it as one of the primary methodologies that has arisen from complex systems research as simple models are no longer capable of working with large data. Therefore, there is a need to develop more complex models that can aid in visualizing and analyzing this data such as ABM. An agent-based model is described as a powerful tool in helping to understand these complex systems (Xiang *et al.,* 2019). Wilensky and Rand (2015), define ABM as a complex system which can be modelled by producing agents that are generated through simple behavioural rules, identifying the environment which represents the world where agents live, and interactions between agents and the environment. Hence, the main elements of an ABM are stated as agents, environment and interactions where interactions can either be agent-to-agent and agent-to-environment.

According to Wilensky and Rand (2015), *"Agent-based modelling* is a form of computational modelling whereby a phenomenon is modelled in terms of agents and their interactions." Scientists have increasingly used ABM to conduct their research as it enables the user to explore and analyze phenomena and to undertake what-if analysis. Modellers are now able to perform simulations that may not have been capable to perform in the past due to an increase in computational power that is necessary to analyze such large datasets. These features of simulation modelling allow ABM to be developed which has the capability to hold millions of individual agents each with its own properties and behaviours allowing for a greater level of simulating interactions. It should also be noted that many people agree that ABM are easy for people to understand and easier to understand than mathematical representations of the same phenomenon. As Wilensky and Rand (2015), explain, as mathematical representations are constructed from mathematical symbols using equational models, ABMs are constructed from individual objects and simple behavioural rules. As Simudyne (2019) explains, when agents are represented as people, their interactions are classed as social networks which can refer to the flow of information between millions of interacting individuals. The main components of ABM are the agents represented as individuals or autonomous decision-making entities, agents behaviour and the environment (Gilbert & Terna, 2000; Macal & North, 2007).

In the context of consumer behaviour, ABM enables several mechanisms for modelling consumer characteristics such as autonomy and social interaction to gain powerful insights by analyzing patterns as illustrated in the study conducted by Ulbinaite and Moullec (2010). Furthermore, Bell and Mgbemena (2018), states ABM is useful for understanding and predicting consumer behaviour processes as it has the capability to model individual decision-making through heterogeneity and interactions with other agents. In addition, it has the ability to incorporate environmental factors such as dynamic pricing which is useful for this study and allows for the simulation of consumer behaviour.

2.4.2 Basics of ABM

According to Wilensky and Rand (2015), a topology refers to connectivity features which determines the agents neighbourhood including spatial interactions with other agents and interactions with the environment. Some of the common topologies implemented for ABM consist of Cellular Automata, Euclidean space, Geographic Information System (GIS) and network topology (Macal and North, 2009).

Cellular automata use a grid or lattice structure in a two-dimensional space to model the system using transitional rules that allow agents to move from cell to cell in a grid depending on the neighbourhood of the cell (Sudhira *et al.,* 2005). In Euclidean space models, however, agents move in two, three or more dimensional spaces. GIS relates to the movement of agents in a geo-spatial landscape represented by locations in the physical world. By observing its spatial distribution, a large amount of data which looks at large areas of land can be recorded and utilized by environmental scientists, resource management and health organizations (Wilensky and Rand, 2015). According to Najlis and North (2004), GIS includes features in multiple layers which allows modelling of different types of agents within a geographical environment. As the environment changes, so does social behaviour of agents. As O'Sullivan and Torrens (2000), suggest that urban infrastructure changes as interactions between agents and their environment change. In a network topology, links connect agents in order for communication to take place and is based on relationships rather than proximity (Khanzadi and Attar, 2016). Links in static networks are specified at the beginning and do not change. On the other hand, links in dynamic networks change according to the rules in the model and the environment. This can show how the social structure can be impacted. For example, as mentioned earlier in this literature review Chapter, the Bass model (Bass, 1969) consists of an innovator who spreads the message of adoption. This form of communication can be passed on to other agents in a network topology. Chapter 5 details social network analysis as a form of topology suitable to implement in this research.

2.4.2.1 Components of ABM

The components of ABM include agents, interactions and environment which are described below.

- Agents

In an ABM, agents are autonomous and can be an individual, a group of individuals or an organization. Each agent has properties and relationships with other agents (Wilensky and Rand, 2015). Agents have properties which are the internal or external state of the agents which describe an agents current state and as the ABM simulation progresses, the user will be able to inspect each agent as their state changes. The action of each agent affects the action of other agents. Behaviours are the basic way for agents to interact with the world.

- Interactions

As the main idea of ABM is to model phenomena in the world through agents, environment and interactions; the agent-based model has the advantage of modelling quite complex interactions. Interactions can occur between agents and the environment. In addition, as these interactions are updated, agents can update their internal state and take additional actions (Wilensky and Rand, 2015). Hence, the modeller will be able to investigate how the dynamics of the model evolve.

- Environment

The design of the environment consists of factors surrounding the agent as they interact with the model and can affect the decisions of agents.

As illustrated in Figure 4, the ABM cycle (Wilensky and Rand 2015), firstly consists of initializing a world with agents, secondly the agents observe the world which leads to the third step where each agent updates their action based on current observations and depending on their internal model, they go back to observing the world.

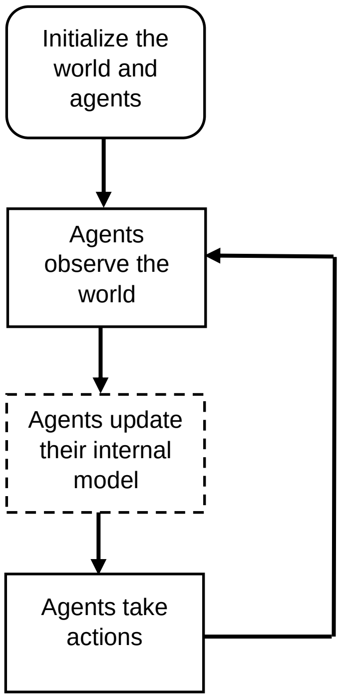


Figure 4: The ABM cycle (Wilensky and Rand, 2015)

Some of the features of ABM include heterogeneity, emergence and complexity which are briefly described below:

- Heterogeneity

ABM allows modelling at each individual or agent level. As behavioural differences exist at an agent level, capturing the heterogeneity with a bottom-up approach allows modelling of richer behaviour.

- Emergence

Another feature of ABMs is its emergent phenomena that appear from agent interactions and where complex systems can be broken down into complex patterns using simple behavioural rules (Simudyne 2019).

- Complexity

Agents can communicate with other agents and whose behaviour is a result of its observations, its knowledge and its interactions with other agents. Agents have internal attributes personal to that agent and its actions are determined through independent decision-making according to the agents personal objectives and depending on its conditions (Kiesling, 2011). Modelling complex social systems is therefore possible through the implementation of ABM (Buchmann, Grossmann and Schwarz, 2016).

2.4.3 Strengths and limitations of ABM

Some of the advantages of ABM are it is a bottom-up approach which allows emergent and dynamic behaviour to be observed whereas a top-modelling approach examines the system as a whole. Another advantage of ABM is attributes and behaviours can be inspected through heterogeneity resulting in the identification of emergent patterns. In addition, agents are able to change their decisions and strategies based on interactions between agents and the environment (Holland, 1996). For example, when modelling social networks, WOM could influence decision-making of its neighbours. As agent-based models describe individuals there is a closer match between ABM and modelling phenomena in the real-world. As such not much training in such a modelling paradigm is required as it is easier to explain and understand which can be beneficial when explaining to stakeholders of a business. Another advantage of ABM is that it considers many assumptions which may be overlooked with other models which is to be explored when designing the model such as knowledge at the foundational level of agents properties and actions as ABM rules are constructed at the individual level as opposed to the aggregate level. As well as investigating the theory of agents behaviours, the model can also be examined to observe if the patterns of behaviour reflect the patterns in the real-world (Wilensky and Rand, 2015). As well as the strengths of ABM, there are some weaknesses. One of the weaknesses of ABM is intensive computational resources such as time, computing power to calculate the behaviour and interactions of all the agents especially if there are a large number of agents and high computational cost. Another weakness is the difficulty in validating and verifying the model (Matthews *et al.* 2007). Moreover, observing every agent where there are a large number of agents in the model can be lost and adds to the difficulty of making decisions.

2.4.4 Output analysis of ABM

To summarize the large amount of data as a result of output analysis from the model, descriptive statistics allows the results to be examined efficiently. By obtaining statistical results such as means and standard deviations and examining these in the form of graphical visualizations, the output from ABM can be further examined and analyzed to better understand the model therefore gaining a deeper understanding of the phenomenon that is being modelled. Graphs can be used to visually observe "…the dynamics and temporal evolution of the model" (Wilensky and Rand, 2015). This uses time-series analysis, as the general trend can be investigated from multiple runs. Furthermore, time-series will be able to show one of the outputs as dynamic patterns of the model.

2.4.5 Justification of ABM

As a result of conducting the literature review, the most common simulation methodologies used in consumer behaviour is DES, SD and ABM. From the literature it is evident that ABM is preferred to simulate consumer behaviour as it allows individuals and their social interactions to be investigated. This methodology uses a bottom-up computational modelling approach looking at system components. By implementing ABM, emphasis is placed on the social interactions between agents and their environment. In contrast, SD uses a top-down approach which looks at the system more holistically rather than the individual components (Ding *et al.* 2018). ABM has the advantage of paralleling real-world systems by engaging in a data-driven way representing social networks. This is further emphasized by Watts (2013), that this approach of combining social network analysis with ABM is essential for modelling complexity as it combines social and computational sciences.ABM is the preferred choice of method to use in this research due to having the capability to simulate interactions between agents (consumers & sellers) and the environment and to also forecast trends. Choosing ABM as opposed to DES allows simulation of movement between consumer agents and seller agents and to analyze its movement pattern. Whereas in DES, “a path must be drawn from point A to point B in order for an entity to move between those two points” (Dubiel and Tsimhoni, 2005). Therefore, it would be difficult to model paths which are not predetermined using DES. One of the advantages of ABM is that agents can change not only their decisions but also their strategies (Holland, 1996). Based on experience, an adaptive agent is able to make different decisions when given the same inputs. ABM creates interest among researchers, scientists and practitioners by investigating a complex systems outlook and “…restructuring the way we understand the world around us” (Wilensky and Rand, 2015). Furthermore, ABM allows experimentation to alter the direction of decision-making which is increasingly important in a complex world where emergent phenomena can be observed. As ABM is a relatively young field, there is a gap for further research in this area that involves further understanding complex systems (Wilensky and Rand, 2015).

2.5 Software packages for the application of sentiment analysis and ABM

2.5.1 Software for sentiment analysis

NLP aims to understand human language through programming languages and software such as Python. According to Turing (2022), Python is favoured as the most popular software for carrying out NLP tasks. Some of the advantages of using Python for NLP are versatile tools and libraries are available for different NLP tasks for example, sentiment analysis, topic modelling and text vectorization. Furthermore, NLP in Python can be integrated with other programming languages to build machine learning models (Turing, 2022). Some of the NLP toolkits are Stanford NLP, RapidMiner, SparkNLP, TextBlob and VADER. Stanford NLP provides pre-trained models for various NLP tasks including sentiment analysis in a range of different languages and domains. It applies tokenization, POS, NER, dependency parsing and conference resolution to name a few. It offers a robust performance but may require more setup compared to TextBlob or

VADER (Manning *et al.,* 2014). RapidMiner uses an integrated data science and machine learning platform for a wide range of data analysis and modelling tasks such as sentiment analysis. It uses pre-built templates for text analysis tasks and a visual interface for creating and deploying machine learning models and data analytics workflows. RapidMiner is more appropriate to use when working with more than textual data as it can handle various data types which are more than just text data. It is applicable for data analytics and modelling, complex machine learning tasks, predictive modelling and advanced data science tasks (Marzukhi *et al.,* 2021). Spark NLP is applicable for NLP tasks in the Apache Spark ecosystem which has been designed for processing large-scale text data using distributed computing. It offers a wide range of NLP tasks, including tokenization, POS, NER, sentiment analysis, and more. It is particularly powerful for big data applications and can efficiently process large datasets in parallel. It provides pre-trained models for multiple languages and domains, making it suitable for a variety of NLP tasks (Kocaman and Talby, 2021). Valence Aware Dictionary and sEntiment Reasoner (VADER) is also a NLP tool used as a lexicon sentiment analysis tool that has been developed specifically for sentiments expressed in social media (Pandey, 2018). It has the capability handle abbreviations, capitalizations, repeated punctuations and emoticons. The output for VADER is a compound score where -1 is very negative and +1 is very positive. To calculate the sentiment score, VADER scans a piece of text for sentiment features, calculates the polarity by summing up the score of features that have been found in the text and normalizing the final score with a range of -1 to +1 (Swarnkar, 2020). The output for TextBlob is a polarity and subjectivity score. Polarity refers to how positive or negative a piece of text is and subjectivity refers to opinions or how subjective a text is. The more subjective a piece of text, the higher the subjectivity score. TextBlob and VADER are two of the most commonly used libraries for sentiment analysis in Python. They both use a lexicon-based approach which "…assumes that sentiment is related to the presence of certain words or phrases in the document," (Hota, Sharma, Verma, 2021). It is regarded as a good primary tool to use for textual analysis. VADER is therefore applicable to be implemented into this research as this study aims to analyze social media data in the form of online consumer reviews. Moreover, as Spark NLP is more suitable for big data and distributed computing environments, TextBlob is more suitable to be used in this research as it requires simple and user-friendly text analysis for basic NLP tasks, such as sentiment analysis and does not need complex, large-scale processing. In addition, VADER is specifically designed to analyze sentiment in social media or informal language and may provide better results for such data compared to the other toolkits discussed. Therefore, this research will use TextBlob (Loria, 2020) and Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto and Gilbert, 2014) for sentiment analysis in Python.

2.5.2 Software for ABM

ABM API's include AnyLogic, GAMA, RePast and NetLogo. AnyLogic is a simulation modelling platform that supports ABM, among other modelling paradigms and has been used in industries such as healthcare, manufacturing and logistics. It uses a multimethod modelling approach which makes it suitable for a broad range of simulation projects that allows the modeller to use a combination of ABM, discrete-event and system dynamics methodologies. It consists of a user-friendly graphical interface where modellers do not need extensive programming experience and supports Java. It is scalable but requires a license as it is a commercial software (Ghali *et al.,* 2022). AnyLogic is better suited for handling complex modelling in various industries and has a smaller user community compared to some other simulation tools, which may result in fewer available and accessible resources such as, tutorials and community support. GAMA is a platform for creating and running agent-based models. It uses a domain-specific language called GAML (GAMA Agent-Based Modelling Language) for defining model entities, behaviours, and interactions. It is designed for users with a more advanced modelling and simulation background who want to create complex simulations for a wide range of applications including spatial 2D and 3D modelling and network modelling (Grignard *et al.,* 2013). Similarly, to AnyLogic, GAMA provides a multimethod modelling approach, allowing users to create models using both agent-based modelling and equation-based modelling such as system dynamics. This flexibility enables the integration of various modelling paradigms. While GAMA is versatile, it may have a steeper learning curve compared to NetLogo. It is more suitable for users who are comfortable with a higher degree of complexity in modelling. Applying rich data into simulation models requires further processing of the data into a form suitable for the simulation approach to be undertaken. NetLogo developed by Wilensky (1999), is composed of an ABM modelling environment or world that simulates a real-world system where two types of agents; stationary agents called patches and mobile agents called turtles are able to move around in the environment and able to observe interactions with other agents that can form networks (Barbosa and Leitão, 2011). This is particularly useful in a social context where interactions can be observed. Furthermore, it is a multi-agent programmable modelling environment that is fully customizable as the user can change the sizes, shapes and colours of the world. In addition, as NetLogo is designed to be easily readable, individuals with little programming skills will benefit from its ease of use. There are also a large community of NetLogo users creating agent-based models and applying it to a wide variety of domains used by academics, professionals and scientists (Wilensky and Rand, 2015). RePast (https://repast.github.io) and NetLogo software (https://ccl.northwestern.edu/netlogo) both have the capability of modelling simulation such as ABM. RePast has adequate support for graphical user interface development, however, NetLogo is easily accessible for the modeller to create graphical output as a ‘point and click’ feature. As RePast uses Java and Python programming languages, NetLogo uses NetLogo modelling language which due to its simplicity, agent-based models can be created using basic programming skills. On the other hand, one of the limitations of NetLogo has been reported as the execution time of running simulations are quite slow (Raab *et al.,* 2022). Despite this, NetLogo has the advantage of having a large library which can be profited by students learning the software.

2.5.3 Justification of NetLogo software

NetLogo has been chosen to be used in this work, due to its good relation between programming effort and simulation speed. It is the most applicable software to be used in this research as it is an ABM platform designed for coding and running agent-based simulations. The analysis of what-if scenarios and forecasting will be able to identify patterns and behaviours that emerge through agent interactions. Another advantage is that the user will be able to learn more about the model through visualizing simulation runs. By visually observing the interactions, allows the researcher to understand and delve deeper into the data. Another benefit of using NetLogo is its BehaviorSpace tool which allows the model to run by altering parameter settings and observing several interactions (Wilensky and Shargel, 2002). Furthermore, “…the syntax is so readable that stakeholders without knowledge of how to build a model can often read the model code and understand what is going on. This helps improve the verifiability of the model.” (Wilensky and Rand, 2015). NetLogo has the advantage of being an open-source platform which is easier to learn and is primarily for educational or research purposes, specifically focusing on ABM.

2.6 Data Mining Methodologies

There are several data mining methodologies such as Sample, Explore, Modify, Model, Assess (SEMMA), Oracle Data Mining (ODM) and Cross-Industry Standard Process for Data Mining (CRISP-DM). SEMMA is developed by Statistical Analysis System (SAS) a software suite developed by SAS Institute. SAS tools and products are aligned with the SEMMA methodology. SEMMA follows a structured approach that emphasizes data preprocessing and data understanding as key steps before modelling. It is especially useful when dealing with data that requires extensive cleaning and transformation. The "Sample" and "Explore" stages involve selecting the data, exploring its characteristics, and identifying data quality issues. Data quality is a central theme in SEMMA, and the "Modify" stage is dedicated to data cleaning and transformation to prepare the data for modelling. It follows an iterative process where data exploration and data modification are carried out in multiple cycles. This allows for continuous improvement in data quality and model development. Furthermore, it is particularly suited for structured data analysis and is commonly used in data warehousing and business intelligence applications (Schock, Dumler and Doepper, (2021). ODM is Oracle's data mining methodology that integrates with Oracle Database allowing data mining tasks to be performed directly within the database environment. This enables efficient data processing and model deployment without data transfer. It offers data mining capabilities within the Oracle ecosystem, making it suitable for projects that heavily rely on Oracle technologies. ODM provides a set of data mining functions and algorithms for tasks like classification, regression, clustering, and anomaly detection. It supports both supervised and unsupervised learning. ODM is well-suited for large-scale data mining tasks, making it particularly suitable for organizations with extensive data stored in Oracle databases (Tamayo *et al.,* 2005). Lastly, CRISP-DM, is tool-agnostic and industry-agnostic and can be implemented with different data mining tools and is applicable to a broad range of industries and domains. CRISP-DM offers a more general approach to data mining, with a focus on understanding business objectives and the methodology can be adapted to different tools and technologies (Schröer, Kruse and Gómez, 2021). It is widely accepted and versatile and as this research is interested in a holistic approach as it covers the entire project lifecycle, this is the most suitable methodology for this study.

2.6.1 CRISP-DM

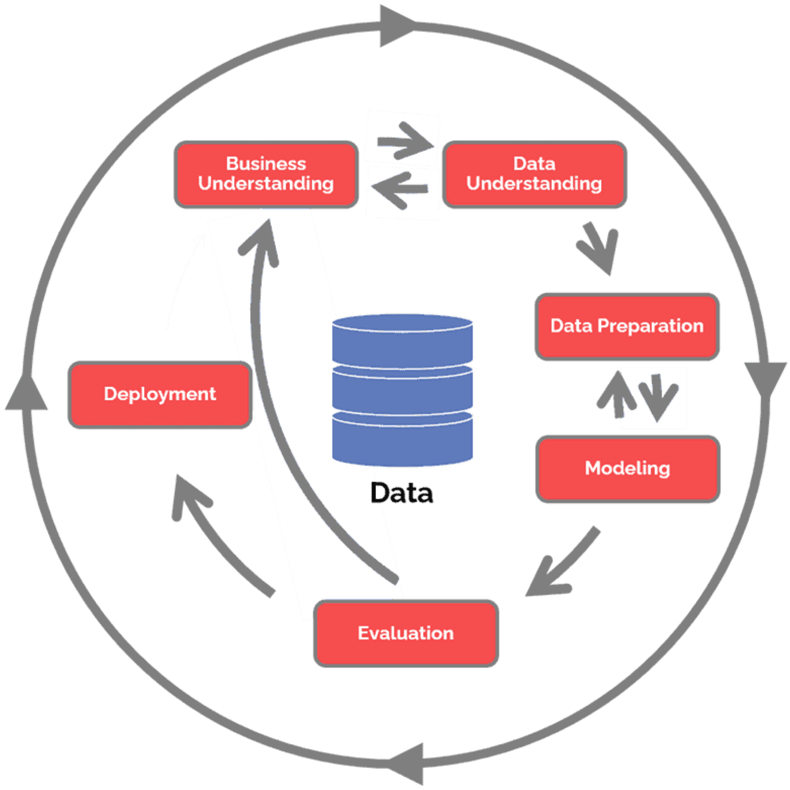


Figure 5: CRISP-DM diagram, (Hotz, 2023).

The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which can be viewed in Figure 5, has increasingly been applied to machine learning projects in recent years. It consists of applying data mining algorithms using advanced instructions and procedures to solve real-world problems. It can be used to aid in planning, organizing and implementing a machine learning project such as sentiment analysis (Luna, 2021). It is a process model for data mining consisting of six iterative phases as follows: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment which are explained next.

- Business Understanding

The first step of the CRISP-DM methodology is to identify and understand the objectives of the project. This involves project objectives, business requirements such as resources, expectations and business success criteria (IBM, 2021).

- Data Understanding

Next is to identify and collect initial data as well as examining and exploring it. Checking the quality of data is important here to verify the quality of the data.

- Data preparation

Data cleaning and formatting should be performed in this step ready to be implemented for modelling.

- Modelling

This stage of the CRISP-DM lifecycle consists of building the model, justification for the choice of model technique used where specific parameters are explored and data mining algorithms are optimized. Moreover, assessing the model is also undertaken at this step by evaluating the model against evaluation criteria such as performance of algorithms.

- Evaluation

In this phase the results are checked against the predefined business objectives and evaluated using performance measures (IBM, 2021). One should ask is the model suitable to be approved for the business and determine whether to proceed to the next step which is deployment or whether iteration needs to continue (Hotz, 2023).

- Deployment

The final stage of the CRISP-DM lifecycle is when the model is ready to be deployed and can take various forms such as a report output or a component for software (Hotz, 2023). This stage can be complex and constant monitoring is required. Improvements within the business can be made if the model is accurate enough to be installed.

2.7 Methodology for simulation study

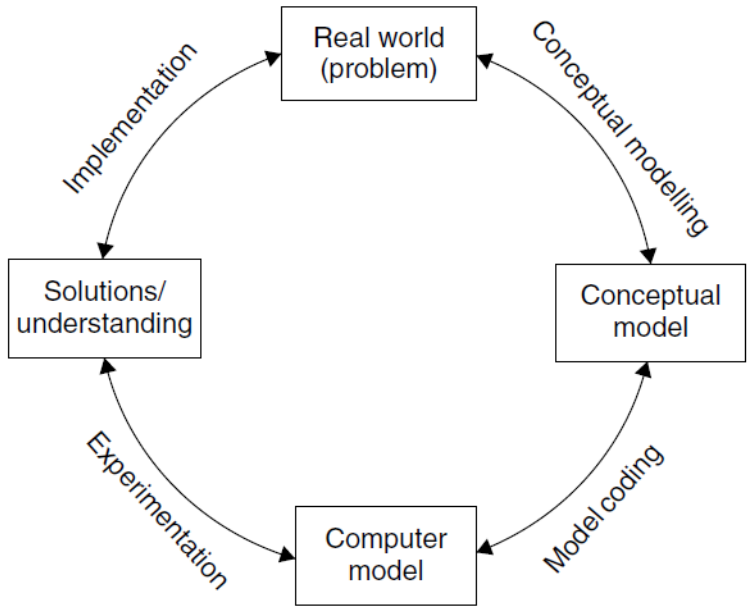


Figure 6: Simulation conceptual diagram (Brooks *et al.,* 2000).

Simulation methodology refers to the systematic approach used to design, develop, and execute computer-based simulations to model and study real-world systems or processes and has been used in various fields, including engineering, science, business, healthcare, and social sciences, to analyze and understand complex systems (Shannon, 1977). A simulation methodology will be implemented for ABM. The simulation will be created using ABM to observe interactions with consumers and seller agent within a dynamic marketplace environment. This enables the capturing of complex social phenomena for diffusion of innovation. The conceptual model can be referred to in Chapter 6 which diagrammatically describes this process. As discussed in Chapter 2, NetLogo will be the platform for developing the agent-based model and visually observing the simulation. In addition, parameter sweeping will be conducted to apply what-if scenarios. As parameter sweeping is performed to understand how the model responds to changes in specific input parameters and to assess the sensitivity of the model to parameter variations, it is an important step for understanding the robustness of the model and for conducting what-if analyses. Parameter sweeping is an important step in the overall model verification and validation process which can be valuable for model refinement and decision-making.

Figure 6 shows the processes involved to conduct a simulation study. The process involves an iteration cycle consisting of the following phases: defining the real-world problem, developing a conceptual model, transforming the model into a computerized version, using the results from the model and implementing them to enable solutions to be provided. The arrows enable movement between the different phases.

- Real world (problem)

  The first stage in a simulation study is to define the problem to be undertaken in order to develop an improved understanding of why the simulation model should be developed. In addition, to investigate the behaviour of the system by using multiple inputs to gain insight.

- Conceptual model

  Describing the simulation model through conceptual modelling, allows concepts to be identified such as determining modelling objectives, inputs, outputs and the analysis of the data. It is a representation of the system and an abstraction of the real world (Robinson, 2013).

- Computer model

  Implementing the model on a computer using model coding to imitate the real world. Using the capabilities of a computer, simulation models can be run, examined and manipulated by transforming the conceptual model into a format that can be understood by a computer (Ferreira *et al*., 2020).

- Solutions/understanding

  This stage is executed by observing the results from the experimentation. For example, what-if scenarios could be conducted in order to obtain a better understanding of the real world. It is also a stage where adjustments can be made to the model if necessary. By implementing simulation in the real world and through simulation-based learning can help businesses in their decision-making (Lateef, 2010).

As the process for a simulation study is circular, the process can be repeated multiple times to improve the simulation and to make adjustments depending on whether the modelling objectives evolve.

2.8 Discussion of literature

This section will review the literature consisting of the following themes, sentiment analysis for consumer behaviour, diffusion of innovation, ABM for consumer behaviour and DDD.

2.8.1 Sentiment analysis for consumer behaviour

A study conducted by Coussement and Poel (2009), investigated the impact of indicators on emotionality towards churn behaviour and compared three classification techniques, Logistic Regression, Support Vector Machines (SVM) and Random Forests (RF), to distinguish churners from non-churners. Their study acknowledged the importance of recognizing customers that are most likely to churn. The support for their argument is reinforced by Rust and Zahorik (1993), that rather than attracting new customers, companies spend on marketing reserves of existing customers. Coussement and Poel (2009), study supports this research for understanding the need to apply new, modern ways of a customer focussed strategy as companies are leaving traditional mass marketing strategies and firms are investing in employees who are able to conduct customer intelligence to improve customer attrition modelling. Their research supports that research into prediction models are growing and there is a requisite to improve the methodology to cater for recent times as acquiring new customers is more expensive than retaining existing customers. As Poel and Larivière (2004), have indicated that even a small amount of change in retention rate can alter significant changes in contribution. Their study reinforced the need to increase prediction power as new information is processed through new sources of media such as emails, blogs and websites. Using call centre emails, they investigated whether several emotions can be expressed by words and if this can enhance the prediction power of a churn system. Their methodology consisted of applying two classifiers, SVM and RF. As well employing these classifiers, they also incorporated a third classification technique; logistic regression. Neslin, Gupta, Kamakura, Lu and Mason (2006) deduce that logistic regression is a robust and recognized technique used in marketing. In addition, Thomas (2000), verifies that is popular among predictive modellers. Therefore, their study uses Logistic Regression as a benchmark for the other two classifiers. The purpose of their study is to support evidence of which classification technique performs best in optimizing the performance of a churn system. This statement is reinforced by expressing the importance of choosing an appropriate classification technique that improves the performance of a churn model. Coussement and Poel (2009), used subscription data from the largest Belgian newspaper company from 2002 to 2005. From call centre emails, they extracted emotionality traits in client/company interactions. Their study was divided into a training and test set to assess predictive abilities of the different classification models. Two different churn models were built for comparison of predictive performance of the different classifiers. Their results show that the model's performance is increased when incorporating emotions viewed by clients in emails. Their study concluded that, "RF improves predictive performance when compared to SVMs and Logistic Regression" (Coussement and Poel 2009).

Kang *et al.* (2022) acknowledges that one of the factors which influence consumers' purchasing decisions are consumer reviews and investigates whether there is a correlation between these two factors. Their study uses LDA for sentiment analysis to identify product quality factors and supporting service factors for new laptop products. They develop a theoretical framework to explore the impact of online reviews on consumer purchasing decisions. Their findings show complexity in the interaction between online review features. In addition, one of the key findings their research suggests there was no significant effect between negative sentiment and variance of online reviews on consumers purchasing decision. Their framework also displays that product features such as volume, polarity, product quality and variance differently affect consumers purchase decision. As their study explores factors and their mechanisms for purchasing decision based on online reviews, there are a few limitations to their research. A limitation of their study is that they only collect data from laptop products and hence other products can be of interest for application of their framework. In addition, their study used JD.com which is an online shopping platform and a popular website in China. Future research could look at other online shopping platforms. Another aspect to their framework could be to include ABM to monitor interactions between consumers.

Sentiment over a period of time has been examined in a political context such as the Brexit referendum. Lansdall-Welfare, Dzogang and Cristianini (2016) study used a lexicon-based approach which measured differing levels of sentiment such as anxiety, sadness, anger, negative and positive effect nationally in the United Kingdom. Their methodology consisted of collecting social media data from Twitter and applying multivariate time series analysis They gathered 10 million tweets over a period of 30 days between 1st June 2016 and 30th June 2016. For each tweet, they collected the date, time, and geographical information. Due to the nature of tweets being restricted to 140 characters at that time, individuals opinions may not be fully representative, and they are more likely to use abbreviations which the analyst should be aware of. Moreover, users may use slang or political jargon and therefore the person who is analyzing the text needs to be aware of cultural expressions. They applied a smoothing function for the daily fluctuations and located points of simultaneous change in the multivariate series, rather than for each component separately. Their results show fluctuations in collective public mood are due to effects such as seasonal and external events. Moreover, they found a significant correlation between the Pound Sterling and Euro exchange rate with change in public sentiment, hours after the Brexit vote. According to Lawani *et al.,* (2019), there is a need to clarify the relationship between reviews and price.

Over recent years, there has been an array of studies concentrating on predicting stock market movements using sentiment analysis. Valencia, Gómez-Espinosa and Valdés-Aguirre (2019), study investigated social media data on the cryptocurrency market. They identified Bitcoin having a lack of research on its behaviour and hence identified Bitcoin, Ethereum, Ripple and Litecoin as a research gap. Cryptocurrencies are becoming an emerging field in the world of financial markets and there is a need to derive insights through the application of sentiment analysis and using machine learning techniques to build a prediction model to predict cryptocurrency stock market movements. They obtained historical market data from the top performing 65 cryptocurrency exchanges which contains the following features such as opening price, highest price, lowest price, closing price and transaction volume for each time-step. In addition to this, social media was gathered from Twitter. As mentioned earlier, the challenge in using tweets are they are restricted in word limit and therefore can lack contextual information. This research aims to overcome this by not having any restrictions of word limit. Their criteria in selecting which tweets to use were based on tweets that were created during the time period of the study, are in the English language and contained the name ‘bitcoin’ or ‘btc’. The advantage of this criteria is the tweet will be specifically related to cryptocurrency. Valence Sentiment Analysis was applied to the tweets which quantifies the degree of pleasure or displeasure of an emotional experience. As financial markets concentrate on valence, their study integrates polarization of opinions as a complimentary dimension to emotional valence. They used a feature vector which included market and social data. Results are compared from neural networks, multi-layer perceptrons (MLP), SVM and RF. Their study concluded performance using MLP was the best for Bitcoin, Ethereum, Ripple and SVM performed best for Litecoin. A limitation of their study was not being able to identify the root cause of behaviour which is one this study aims to uncover. Referring back to the research contribution, an objective is to apply topic modelling to uncover topics and to view which words are being used in association with a brand, identifies key themes and how these evolve over time.

2.8.2 Diffusion of Innovation

Gatti *et al.,* (2014), acknowledges the challenges in measuring and predicting the spread of diffusion across online social networks even though there are multiple adoption models of modelling influence in social networks for products. Particularly, in the area of marketing, companies want to see how users interact to spread content and promotions. Gatti *et al.,* (2014), proposes applying ABM as a modelling paradigm to construct what-if analysis and has identified a research gap in applying this methodology to large online social networks. The main contribution of their study necessitates simulation through user posts using an online social network. They used a case study approach which consisted of investigating Barack Obama's 2012 political campaign using Twitter as the online social network platform. They perform sampling the network, text classification including topic modelling and sentiment analysis, pattern recognition and large-scale simulation. For the agent-based model, Obama is the central user and immediate neighbours, and their associated connections are observed. The proposed method is based on a stochastic multi-agent-based approach where each agent is modelled from the historical data of each user in the network as a Markov Chain process and a Monte Carlo simulation. The main contribution of their work is applying ABM to examine what-if analysis by tuning users' behaviour. However, there are limitations of their study. Firstly, by increasing the volume of data used in the dataset by re-evaluating sampling may improve estimation of the model. Secondly, they have stated in the future work section that improvements could be made on user opinions through topic modelling and sentiment classification techniques. Another limitation of their study is conducting further what-if scenarios to understand the impact change of behaviour would have.

Kotthoff and Hamacher (2022), recognizes the challenge in modelling complex systems particularly diffusion of innovation. Therefore, they suggest ABM as a tool to simulate the complexity of consumer heterogeneity. They focus their study on utilizing a gradient-based calibration method to simulate calibration on an agent-based model with a large dataset consisting of 25 free parameters. Their proposed calibration method was tested on a dataset consisting of photovoltaic systems which uses solar power in homes in Germany during a time period of 2000 to 2016. The agent-based model they created used diffusion of innovation for PV systems diffusion to be able to change the adoption variable for agents to adopt and once of the assumptions are once an agent adopts, they stay in the state of being an adopter. The utility function they employed is directed towards an agents decision process. A limitation of their study is that as decision processes are complex, their model does not adhere to model complexity due to gradients being quite simple. They conclude that their method of incorporating calibration achieves better results that include gradients rather than models that do not include gradients. They document the model using Overview, Design concepts and Details (ODD) which is a document protocol setting out guidelines for the description of ABM. Their study, however, does not cover all aspects of the ODD as their method is more general and not applicable to a specific agent-based model. Their study does not include validation and the concept of overfitting a model as the focus of their study revolves around calibrating ABM using gradients. Their study contributes to scalability which is beneficial for modelling ABM with diffusion of innovation on large scales.

Lengyel *et al.,* (2018), present a complex contagion agent-based model adapted from the Bass diffusion framework (Bass, 1969). Similar to Gatti *et al.,* (2018), Lengyel *et al.,* (2014), also use a case study approach using *International Who is Who (*iWiW) which is a Hungarian online social network. The dataset consists of a period of first adopters from 2002 to 2012. It also comprises of location of users, interpersonal ties, registration dates and last log-in date for each user. The dates are used to identify adoption and churn variables. Adopters and churners are identified for urban scaling laws of innovation over the lifespan of a product. Their findings suggest that there is a difference in comparison to the model where similar ties were followed faster in reality which creates a high assortative mixing in the adoption time. Their research integrates Bass model (Bass, 1969) with ABM using empirical data to fit global diffusion dynamics. Their study used adoption thresholds for early adopters which showed improvement to the network model fit. They conclude that geographical features influence the prediction of adoption.

2.8.3 ABM for consumer behaviour

Yang, Mo and Zhu (2014), propose an agent-based simulation framework, modelling Twitter financial community using Twitter network growth and message propagation mechanism. They use an empirical dataset collected through Twitter API which tracks financial community users using profile information and the tweets and retweets posted by Twitter users. The dataset is split into two parts; the first is for calculating the model statistical and input parameters and the second part is for testing and validation. From the empirical data, they observed broadcasters having the highest follower ratio, followed by acquaintances and lastly, odd users. The ABM they developed uses parameters from the composition of followers imitates the community relation composition with Beta distribution and the mean is representative of the empirical data. They state the results closely match the empirical features of Twitter behaviour's and network dynamics. Scenario analysis conducted, demonstrates that not all users have the same influence for message propagation. They investigated this through the removal of critical nodes which shows no significant difference with retweeting ratio and survival tick or time-step. Through this finding, they have highlighted that the removal of critical nodes can be deployed during extreme events for example, their case study approach consisted of investigating the impact of the associated press hoax incident to the social network community and financial market to suggest policy making through social media. The empirical statistics for the network demographics and aggregated message propagation from Twitter API allowed the pattern to be closely matched to the agent-based model. Their findings suggest that by spreading malicious messages triggers financial market agents to sell a large volume of shares. Scenario analysis was undertaken which concludes that by removing critical nodes such as nodes with the highest betweenness centrality in the network, has the effect of reducing the ratio spread of malicious message to the community. However, they state that incorporating sentiment would allow the distribution to be examined for messages collected in the Twitter financial community which would be able to enhance the accuracy to determine which message would be more likely to be broadcasted.

Schweitzer, Krivachy and Garcia (2020), explore the emergence of emotional interactions between agents and opinion dynamics through ABM. Emotional state of agents such as valence which is described as "pleasure associated with emotions" (Schweitzer, Krivachy and Garcia, 2020) and arousal. ABM is implemented to explore emotions and opinions of individuals. Their "…assumption is that the dynamics of *opinions* is driven by the dynamics of *emotions*" (Schweitzer, Krivachy and Garcia 2020). In their study their driving variable is the emotional state, and the driver variable is individual opinion. They state that their modelling approach differs from others stating that emotions and opinions change on two different time scales as emotions evolve quicker and consequently, the "resulting values for the mean valence and the emotional field…. drive the evolution of opinions" (Schweitzer, Krivachy and Garcia, 2020). A limitation of their study is that they do not have empirical data to validate their model. They also state that their model could potentially be validated with data through the application of sentiment analysis of online user comments including time series data of comments. By applying a lexicon-based technique such as sentiment analysis, they would be able to estimate the value of positive and negative expressions. As, previous literature have used polarization of opinions in social media, (Shin and Lorenz, 2010), (Vicario *et al.,* 2016), Matakos, Terzi and Tsaparas, 2017, Yardi and Boyd, 2010, Zollo *et al.,* 2017) future work could include collecting data from social media platforms such as Reddit (Abisheva *et al.,* 2014) to measure and link the ratio of positive and negative opinions to different emotion scenarios. In addition, by incorporating social network analysis, polarization could be measured to detect different opinions from neighbours in the network for example, interactions between users to further enhance the agent-based model. As their model only considers a one-dimensional opinion space, there is potential to advance this to multi-dimensional to incorporate complexity of mixed opinions to describe how emotions drive opinions in different aspects using a vector. Lastly, topic modelling could be considered as a future avenue to reduce the dimensions of opinions which can then be used to examine polarization and correlations.

Čavoški and Marković (2016), apply ABM to simulate consumer behaviour on business-to-consumer (B2C) e-commerce websites. They employ a utility function which includes a utility threshold to determine if a consumer makes the decision to purchase. The rule is as follows: if the utility function value is below the threshold, the consumer decides not to make the purchase. It is worth noting here that the utility function possesses multiple variables in order to satisfy consumers purchasing decision. Similarly, as other related work that have been discussed in this Chapter, the complexity of decision-making is highlighted by Čavoški and Marković (2016) in which they find ABM to be a suitable methodology to examine the interactions between agents through behavioural rules. Their study implements NetLogo as the platform to create the agent-based model. The agents described in their model consist of consumer, seller and advertisement agents comprised of agent attributes gained from empirical and theoretical data. They conclude that ABM as a methodology is beneficial to implement, contributing to analyzing and monitoring business strategies in e-commerce markets. This tool enables a deeper insight and understanding into consumer behaviour at a managerial level which can aid in improving market segmentation and strategic business decisions. However, they recognize that the aspect of consumers decision-making is complex, and more research is being undertaken in this regard. Therefore, new insights and further study is plausible.

Kerdlap et al (2020) study uses ABM to simulate several scenarios including plastic waste generation, collection routes, sorting, and recycling processes between different entities The simulations were based on real-time evolution using Clementi-Bukit Merah in Singapore to be the location of their case study. Their study used AnyLogic software to create the agent-based model and to observe the interactions. The output of the agent-based model such as material outputs and transportation data were used to determine the environmental impact of the following scenarios: centralized where domestic waste is sorted by a large-scale sorting facility, semi-distributed where domestic waste is sorted at nine small-scale sorting facilities and the third is distributed where the small-scale plastic recycling facilities are distributed and are located within the sorting facilities. Life cycle assessment (LCA) was applied to develop a multi-level model for analyzing the environmental impacts of the different scenarios which resulted in polyethylene terephthalate (PET) bottles and polypropylene (PP) takeaway containers which can increase or decrease the life cycle of greenhouse gas emissions (GHGs) of small-scale distributed sorting and recycling compared to centralized large-scale recycling which is due to GHG from transportation. Their study highlights the importance of planning and designing distributed systems as the environmental impacts are sensitive to the types of trucks used and their GHG. In addition, their study implemented the integration of ABM and LCA to explore different scenarios to support a project plan of distributed systems to lower environmental impacts. Further work in this area could incorporate other what-if scenarios such as different configurations and waste streams. Their study could also apply their research to other locations in Singapore. This review shows how ABM and LCA can be applied in the context of recycling technologies.

2.8.4 DDD

As data is growing exponentially, more companies are investing in analyzing and extracting key insights from data leading to the concept of DDD. McAfee and Brynjolfsson (2012), state businesses who use data, base their decisions on evidence as DDD lead to better and more effective decisions.

A study conducted by Brynjolfsson, Hitt and Kim (2011) investigate whether there is a connection between DDD, and productivity as indicated through case literature and economic theory. Their study uses survey data, in collaboration with McKinsey & Company from 179 public corporations from 2005 to 2009. The survey was aimed at senior management on information systems and business practices. Within the survey, three specific questions were asked in order to build the independent variable as DDD. The main findings they discovered were the coefficient for one of the control variables, Property Plant and Equipment (PPE) was larger than the theoretical value, whereas the coefficient on Information Technology (IT) and other assets were less. However, when combined with DDD, the coefficient on IT is greater as it is one standard deviation higher. The key point that should be gathered from this study is IT aids in enabling DDD as PPE and other assets correlations with market value are not significant and a small positive interaction is identifiable between DDD. However, this point faces criticism as discrepancies of the low direct coefficient of other assets. Conclusively, their study aims to suggest that the correlation of IT capital with DDD results in a higher market value.

Another study conducted by Tambe and Hitt (2011), observed the coefficient estimate on IT measures stays the same while the coefficient estimate on DDD is higher, leading to a 4-6% increase in productivity. To check their assumption, they repeated the study using smaller time periods from 2008-2009. They discovered DDD was practically similar in all time periods of the samples and was followed up by undertaking the Chow test (1960) which is used to examine whether the coefficients of different regression models in one group are different in other groups. The test showed

‘…no significant variation in the DDD coefficient between subperiods.’ (Brynjolfsson, Hitt, Kim, 2011). However, it should be noted that the comparison of the results from the different samples were restricted to 72 firms rather than the 179 firms. The findings show a big difference of an increase in productivity of about 4-6% (Provost and Fawcett, 2013). Their findings conclude that the DDD scale represents higher productivity and evidence for asset utilization. However, further research can be undertaken for more firms. Conclusively, the DDD measurement can be used by senior management to improve decision-making hence, increasing profitability.

Provost and Fawcett (2013), define a relationship between data science, big data and DDD with the opportunity of companies gaining a competitive advantage. Demand is accelerating to hire data scientists with the relevant tools and knowledge to carry out analysis that can derive underlying meaning from the data. Provost and Fawcett (2013) further enhance this statement by mentioning universities are rapidly creating data science programs to keep up with demand of companies hiring employees competent of data science capabilities. Their study also mentions the importance of DDD. They describe that the concept of DDD focusses on making decisions based on data analysis rather than mere instinct. Again, the statistic of a 4-6% productivity increase seems to be the underlying relationship; “And the differences are not small: one standard deviation higher on the DDD scale is associated with a 4–6% increase in productivity. DDD also is correlated with higher return on assets and the relationship seems to be causal” (Provost, Fawcett 2013).

2.9 Research Gap

Table 2 summarizes previous work in the field of sentiment analysis, topic modelling, price as a variable, diffusion of innovation, particularly the Bass model (Bass, 1969) and ABM in comparison to the current research study. It also identifies a research gap where the current paper applies these factors in the development of the SAABM framework to investigate consumer behaviour. To summarize**,** Gatti *et al.* (2014) study incorporates sentiment analysis, topic modelling and ABM in the context of predicting diffusion across social networks. Yang, Mo and Zhu (2014) apply sentiment analysis and a pricing variable in the context of a financial market and applies ABM to model the Twitter financial community. Čavoški and Marković (2016) applies pricing variable and ABM to simulate consumer behaviour on B2C e-commerce websites. Lansdall-Welfare, Dzogang and Cristianini (2017) applies sentiment analysis in a political context using pricing variable. Lengyel *et al.* (2018) focusses on implementing the Bass model (Bass, 1969) on a complex contagion agent-based model. Valencia, Gómez-Espinosa and Valdés-Aguirre (2019) investigate sentiment analysis in the cryptocurrency market. Schweitzer, Krivachy and Garcia (2020) explore sentiment through ABM. Kang *et al.* (2022) investigates sentiment analysis, topic modelling and a pricing variable on consumer purchasing decision on laptop products. Kotthoff and Hamacher (2022) inspect pricing variable, Bass model (Bass, 1969) and ABM utilizing a gradient-based calibration method. This literature review has paved the way for the current research to incorporate sentiment analysis, topic modelling, price variable, Bass model (Bass, 1969) and ABM in the context of trainers products.

Table 2: Comparison of previous work and the current research paper

Authors	Sentiment Analysis	Topic Modelling	Price variable	Diffusion of Innovation (Bass model, Bass 1969)	ABM
Gatti *et al.* (2014)	✓	✓	-	-	✓
Yang, Mo and Zhu (2014)	✓		✓	-	✓
Čavoški and Marković (2016)	-	-	✓	-	✓
Lansdall-Welfare, Dzogang and Cristianini (2017)	✓	-	✓	-	-
Lengyel *et al.* (2018)	-	-	-	✓	✓
Valencia, Gómez-Espinosa and Valdés-Aguirre (2019)	✓	-	✓	-	-
Schweitzer, Krivachy and Garcia (2020)	✓	-	-	-	✓
Kang *et al.* (2022)	✓	✓	✓	-	-
Kotthoff and Hamacher (2022)	-	-	✓	✓	✓
Current paper	✓	✓	✓	✓	✓

As evident from conducting the literature review, there are several research gaps which have been identified as the following:

- Firstly, there is limited literature exploring whether a correlation exists between sentiment and price in the context of trainers products derived from Amazon UK online reviews.
- Secondly, the literature does not adequately explore how sentiment influences consumers purchasing decisions in a dynamic pricing context. Understanding the relationship between consumer sentiment, pricing changes, and buying behaviour is crucial for optimizing pricing strategies and can be a competitive advantage for businesses.
- While both sentiment analysis and ABM have gained significant attention in modelling and simulating complex socioeconomic systems, there is a noticeable gap in the literature when it comes to integrating sentiment analysis within ABM frameworks to create a more comprehensive simulation of human behaviour in dynamic environments. While ABM has been successful in modelling certain aspects of human behaviour, the incorporation of sentiment-driven decision-making processes remains underexplored. Research is needed to investigate how sentiment analysis can improve the realism of agent behaviours, such as purchase decisions and social interactions.
- The literature does not sufficiently address how sentiment dynamics in agent interactions affect the emergence of collective behaviours and social phenomena in ABM. Understanding how sentiments spread, evolve, and influence the behaviour of agents in a dynamic context can therefore be identified as a research gap.
- The literature lacks examples of interdisciplinary collaboration between sentiment analysis and ABM experts in the context of trainers products derived from Amazon UK online reviews. Bridging the gap between these two domains can lead to innovative research that combines sentiment analysis with the modelling capabilities of ABM.
- In comparison to studies reviewed in this literature review chapter, this research will include topics on sentiment Analysis, topic Modelling, price variable, diffusion of innovation specifically (Bass model, Bass 1969) and ABM where previous studies have not yet explored these

concepts together in the context of Amazon online consumer reviews of trainers products.

In conclusion, by conducting research that explores the integration of sentiment analysis and ABM, the impact of sentiment on agent behaviours and interactions, and the practical applications of utilizing a combination approach of sentiment analysis and ABM can advance the understanding of complex social systems and enhance the capabilities of ABM in DDD.

2.10 Chapter summary

This Chapter has reviewed literature in the application of sentiment analysis, topic modelling, diffusion of innovation including the Bass model (Bass, 1969) and ABM.

From the literature review, it is evident that as data in the social media space is growing intensively, there is a research gap in analyzing such unstructured, noisy text data which goes beyond what traditional modelling methods are capable of. By incorporating diffusion of innovation and leveraging the versatility of the Bass model (Bass, 1969), the SAABM framework proposed in this study will be able to factor in elements such as WOM and social influences as well as price and sentiment. In addition, the ABM paradigm offers two advantages for the study of diffusion of innovation; firstly, it enables the modelling of agent heterogeneity and secondly, it allows modelling of interactions in a social network environment. As Railsback and Grimm (2012), describe NetLogo as an ABM tool as being the best platform for beginners along with features being open source and free, NetLogo will be used as the platform to build the agent-based model. It remains to be the most widely used ABM language providing a simplified programming language and graphical interface that allows modellers to "…build, observe, and use ABMs without needing to learn the complexity of a standard programming language" (Wilensky and Rand, 2015). Conclusively, from conducting this literature review, there is little evidence of research about complex human decision making in ABM applied to consumer behaviour in the form of online consumer reviews particularly on trainers products. This thesis identifies two key concepts that will be combined: sentiment analysis and ABM. In review of the literature and the availability of data these two techniques will be evaluated and applied to the research problem.

Chapter 3

Methodology

3.1 Introduction

The previous Chapter has explored several themes of applying sentiment analysis and ABM applied to a variety of domains. This chapter explores research methods and discusses the rationale of adopting a case study approach for this thesis in Section 3.2. Section 3.3 describes applying a case study research design to be implemented in this thesis. Section 3.4 presents a summary of this methodology chapter.

3.2 Research designs

The following types of research design experimental, cross-sectional, longitudinal, comparative and case study will briefly be discussed below.

- Experimental

In an experimental research design, the experimental and control group are identified with regards to the dependent variable. The independent variable undergoes experimental manipulation to observe the effect it has on a dependent variable. Quantitative comparisons are typically employed, for example, in the field of medical drug development, by manipulating dosage variables, researchers can understand how a new drug works and identify potential side effects (Bhat, 2015).

- Cross-sectional

Cross-sectional design involves collecting data on more than one case at a specific point in time; for example, observing a snapshot of the population. Cross-sectional methods can either be quantitative or qualitative. Quantitative can take the form of surveys and qualitative can take the form of interviews or focus groups both from a structured observation at a specific period. They are "then examined to detect patterns of association" (Bryman and Bell, 2011).

- Longitudinal

The difference between cross-sectional and longitudinal study is the latter gathers data over time. This way, researchers are able to observe and identify any changes taking place. An example, of this type of research design is conducting surveys more than once and in the form of a qualitative method, performing interviews on several occasions.

- Comparative

Comparative research involves comparing two or more cases to investigate the differences and commonalities. An example of comparative research design using quantitative methods are surveys “in which there is a direct comparison between two or more cases, including cross-cultural research” (Bryman and Bell, 2011). Qualitatively, ethnographic or interviews could be compared to examine similarities and dissimilarities.

- Case study

The last type of research method discussed in this thesis is the case study approach which involves a “detailed and intensive analysis of a single case” (Bryman and Bell, 2011). It is a common research method through which qualitative and quantitative research can be in a combination format, placing less dependence on using a single approach (Knights and McCabe, 1997). An advantage of this methodological approach is focus is placed on the uniqueness of the case to reveal significant features and to place a deeper understanding of its complexity (Lee, Collier and Cullen, 2007).

3.2.1 Justification for case study approach

Vries (2004) describes case studies as a positivist and interpretivist epistemological methodology and are both an important research approach in Information Systems. To achieve Objective 5 as stated in Chapter 1, the case study approach is the most applicable tool to be used in this research study. Another reason why the use of a case study is justified in this research is the combination of applying qualitative and quantitative data which informs and supplies an empirical foundation to the framework model (Taylor, Bogdan and DeVault, 2016).

Positivism perceives objective information about the world through scientific methods which involves making predictions about the nature of things through systematic methods which can be statistically measured (Bryman and Bell, 2011). This tends to be quantitative research for example, surveys, simulation modelling and statistics. On the other hand, one can argue that interpretivism opposes the positivist approach as scientific methods cannot be used to obtain knowledge about the world but rather to explore an individual's point of view through their behaviour and experiences recognizing this as valuable information. In contrast to positivism, interpretivism is subjective and allows investigation into complex phenomena over time which is useful for sentiment analysis as it allows multiple perspectives to acquired (Purchase, Denize, Olaru, 2014).

As simulation models involve complexity, case study research "provides evidence…between the 'real world' and a simulation of that 'reality.'" (Purchase, Denize and Olaru, 2014). The use of case studies allows a deeper understanding of a particular context by applying a particular conceptual framework such as the SAABM framework proposed in this thesis. Consequently, this study will use the case study approach as it includes qualitative data for sentiment analysis and quantitative data for numerical inputs to be used by the agent-based model as well as taking a positivist and interpretivist approach. Moreover, using this approach is necessary for the synthesis of sentiment analysis and ABM.

3.3 Case study description

As this research aims to develop a framework for combining sentiment analysis and ABM for understanding consumer behaviour in a dynamic marketplace, a case study suitable for this research needs to have the criteria of products that have access to online consumer reviews and have the characteristic of dynamic pricing. Selecting a product from the Amazon.co.uk website enables access to a wide range of products and online reviews that can be collected via a web scraper tool. As there is a market for selling and buying trainers this data uses dynamic pricing. In addition, this data derived from online consumer reviews for the purposes of sentiment analysis and dynamic pricing to determine whether a correlation exists between sentiment and price for trainers products. The choice of this case study is suitable for several reasons; Amazon is a popular platform with a strong online presence for comparing prices and consists of a large volume of customer feedback data, hence, it has a high degree of pricing sensitivity which is applicable for this research. In addition, extracting sentiment in this context; dynamic pricing can be observed more effectively. The integration of sentiment analysis relies on the availability of such data sources, which are often accessible in industries with active customer interactions. In contrast, other case studies which do not consider dynamic pricing and sentiment analysis as significant may not hold the competitive nature of the characteristics and dynamics of the market therefore, affecting the suitability and impact of such a case study. Selecting other case studies which do not have these properties would yield different results depending on the specific context and variables involved.

This research will use trainers sourced from the Amazon.co.uk website using only Amazon retailers and not 3rd party retailers. To collect online review data, a web scraper tool was used. The data extracted consisted of product name, textual review data from the UK, ratings and the date the review was posted. To collect price tracker data, data is collected from camelcamelcamel.com which is a price tracking website monitoring the fluctuations of Amazon prices for the Nike, Puma and Timberland trainers products. This website is applicable to use in this research as it monitors the pricing history of Amazon trainers products.

3.3.1 Amazon marketplace

Amazon is the largest ecommerce retailer in Europe and The United States of America and according to Business Insider (2018), Amazon changes product prices every 10 minutes to remain as competitive as possible at any given period. Due to this, there are constant price fluctuations on Amazon products which adds complexity.

The study of dynamic pricing or adjusting prices due to market conditions is gaining more attention from academia. For example, research in stock market prediction is gaining popularity by emerging from a field of data science and machine learning. Referring back to the research gap, there is a need to tackle and research these challenges using advanced methodological approaches. As mentioned in Chapter 1 with regards to the research problem, there is a growing popularity of consumers posting their experience of products online, representing a new source of collecting information in this format using platforms such as Amazon.co.uk (Feng, Fu and Qin, 2016).

As Amazon stores a wide variety of products, this study will focus on 3 brands of trainers; Nike, Puma and Timberland. The choice of brands was due to availability of data and popularity. Popularity is due to having more reviews for that particular product which will be useful for this research.

The purpose of this research is to not only implement sentiment analysis but to combine this form of NLP with price fluctuations. Hence, trainers have been chosen to be examined in this study due to the global competitive market for trainers products. NLP is a popular tool to analyze textual data with various methods such as finding patterns using word frequencies, topic modelling or information retrieval and extraction. One method that is increasingly being applied to extract useful information from text data is sentiment analysis (Altrabsheh, 2016). Sentiment analysis examines the polarity of the text as either positive, negative or neutral. This method of analyzing such data is explained in Chapter 4.

3.3.2 Trainers

Trainers was seen purely as an athletic footwear but has since become a commercial and fashionable product with a global market value of £69 billion in 2020 and forecasted to value at £105 billion in 2026 (Braithwaite, 2021). As trainers bring in a huge amount of business, footwear, in particular trainers, is now the biggest selling category in the online luxury market. By rising sales online, potential consumers use online buying platforms such as Amazon to buy trainers products. As highlighted in Chapter 1 in the research problem Section, this growing popularity of utilizing an online marketplace to buy products is growing exponentially. These platforms aid in gaining an insight into customer experience for example, comments on the product, delivery experience and posting frequently asked questions. This research is interested in online reviews posted by consumers and ecommerce platforms such as Amazon have a section where consumers can leave reviews for products to help in the decision-making process of other potential customers. This brings in a huge amount of data which can help in extracting and understanding information from these platforms. For example, in this research the focus is on extracting online reviews published by consumers and understanding the sentiment of their posted reviews. The past few years due to lockdowns across the globe influenced by Covid-19, there has been a surge in online shopping with consumers prioritizing comfort, resulting in rising sales of comfortable footwear like trainers (Croudace, 2022).

The sneaker culture as portrayed in America is a global million-pound market where there are high investments and brands such as Nike, Puma and Timberland are creating sneakers to fulfil customers desires. Kawamura (2016) describes sneaker culture as more popularly known in America inspired by the underground culture and hip-hop. Following the launch of

Nike Air Jordans, sneakers have a rise in status as commodities along with celebrity endorsements further promoting investment in sneakers. The significance of using trainers in this study is there is an increase of research in buying and selling trainers globally. Recently founded in 2015, StockX is the world's first, self-proclaimed "stock market of things," (StockX, 2016) buying and selling limited edition sneakers, "with stock market-like visibility" (StockX, 2016). According to Lu (2020), buying and selling encompasses price fluctuations in a dynamic market. Marketplaces such as this are looking to drive customer engagement using social media platforms, hence consumers will be looking for customer reviews to determine their purchasing decision (Chitrakorn, 2020). Recently, the digital age through online platforms has seen a growth in marketing trainers and the resell culture. Following on from the trainers resell culture, was valued at £5 billion in 2019 and predicted to reach £25 billion by 2030 (Kawamura, 2016).

The reasons for choosing the following brands of trainers are due to two factors: availability of data and popularity.

3.3.2.1 Nike

Nike is a leading sportswear brand recognized globally endorsed by many celebrities and sponsorships. According to Statista (2022), Nike is an American multinational corporation and the world's largest supplier and manufacturer of athletic trainers recognized by its distinct trademark, the Nike swoosh. With more than £19 billion in revenue, there is bound to be many consumer reviews posted on the Amazon website which ensures availability of consumer review data to be applied for the process of sentiment analysis.

3.3.2.2 Puma

Puma has been described as one of the most popular brand names in the world carried along with celebrity endorsements such as Usain Bolt (Parsons, 2010). Puma prides itself on affordability, its reputation for creating stylish trainers with an urban feel, stating that luxury doesn't have to be expensive as they provide value at an affordable price. For these reasons, Puma was also selected to derive comments from consumers.

3.3.3.3 Timberland

Timberland has achieved credibility through its footwear range by priding itself on innovation, quality and outdoor wear. With footwear being its leading sales driver, accounting for around 71% of its revenue (Arnson, 2022); Timberland has gained popularity for its high-quality durable fabrics. It's comfort and functional features are represented by its higher price point which will be of interest in this research to examine whether more users comment on this particular brand to influence individual decision-making. One way to measure this is to observe data volume compared to Nike and Puma brands.

3.4 Chapter summary

This methodology chapter explored research designs and the justification of selecting a case study approach to be the most beneficial to be implemented to this study. The case study to be applied in this research uses trainers products gathered from the Amazon UK website due to availability of data and that trainers products have an element of price fluctuations. The next chapter implements CRISP-DM as a methodology to be implemented into this study's research approach for carrying out sentiment analysis.

Chapter 4

NLP for sentiment analysis

4.1 Introduction

This Chapter investigates the application of sentiment analysis to the online review dataset. The structure of this Chapter is as follows; Section 4.2 presents an overview of the application of sentiment analysis to the research problem. Section 4.3 illustrates the methodological approach, Section 4.4 consists of exploratory data analysis, Section 4.5 of correlation analysis and Section 4.6 is a summary of the Chapter.

4.2 Overview of NLP for sentiment analysis

With the accumulation of recent articles and media coverage regarding the significance of analytics and data science across many industries, the use of analytics has been greatly encouraged by numerous successful multi-million-pound companies around the world such as IBM, Microsoft and Amazon. A study documented by IBM showed that applying retail analytics creates a competitive advantage (Waldron, 2013). This statement is further backed up by their study which concluded a supermarket retailer in Venezuela used tools such as analysis and customer insight optimization resulting in a 30% increase in revenue and $7million increase in profit. These methods show positive results and hence why companies are investing more in technologies that uncover consumer insight. More recently, the coronavirus outbreak has attracted several businesses to help analyse big data to minimise spread and to take appropriate precautions (Wall, 2014). By detecting sentiment related to the virus can allow for monitoring of health-related issues through early detection. This, therefore, expresses the need to use advanced technologies where companies can benefit hugely from them. The last few years have shown considerable growth of applying advanced analytics tools and data science techniques to data which have been generated by businesses. According to IDC (2014), the Big Data and Analytics market displays significant growth generating into a ‘multibillion-dollar worldwide opportunity.’

This shows a growing demand for the next couple of years, therefore identifying a gap in literature for further studies and developments into advanced analytics and data science with a focus on customer insight optimisation. With many companies using analytics on their data, there is a sense of competitiveness to perfect insight and knowledge of the data to get ahead. In terms of competitive advantage, data scientists and analysts are being highly sought after to identify and understand consumer behaviour (Foreshew, 2015). The article from The Australian Business review by Foreshew (2015), are now realizing the importance of employees to derive information from textual data with an increase of job vacancies for individuals that possess 'higher level analytical skills that come from…PhD studies' (Foreshew, 2015). Sentiment analysis and ABM is becoming increasingly applied in both academia and across industries which, therefore, implies development of research and practice in this field. Therefore, applying sentiment analysis and topic modelling as a tool of NLP, can have major advantages to a business. Hence, there is a need to explore sentiment analysis on online reviews where this study looks at examining sentiment analysis and topic modelling based on online reviews written by users on the Amazon.co.uk website.

As mentioned previously, the vast amount of textual data poses analytical challenges as traditional approaches start to show limitations due to the unstructured nature of the data. As these reviews consist of unstructured, noisy text; sentiment analysis as a NLP tool aids in better understanding consumer behaviour. By applying sentiment analysis to consumer reviews helps to detect how positive, negative, or neutral a review is which helps to better understand consumer behaviour.

Due to the nature of this research and as mentioned in Chapter 2, this study will implement NLP in the form of sentiment analysis and topic modelling as a computational tool which aims to manipulate and understand human language as humans communicate with each other through natural language (Dale, Somers, Moisl, 2000). According to Bird, Klein and Loper (2009), NLP is experiencing growth in different fields and industry which include scientific, economic and social aspects. In this regard, it is important for individuals working in these sectors and particularly working with textual data to benefit from developing and having a working knowledge of these skills.

Sentiment analysis allows natural language to be characterized according to sentiment polarity for example how positive, negative or neutral a text is. In this study it is particularly important to analyze consumer reviews to detect what users are saying or feeling about trainers products in order to achieve the objectives of this research.

Figure 7 illustrates the methodology diagram based on CRISP-DM, which was explained in Chapter 3, for developing the sentiment analysis model and carrying out correlation analysis. Figure 9 shows the output of the sentiment analysis model which results in wordclouds and topic modelling which can be seen in Section 4.3.

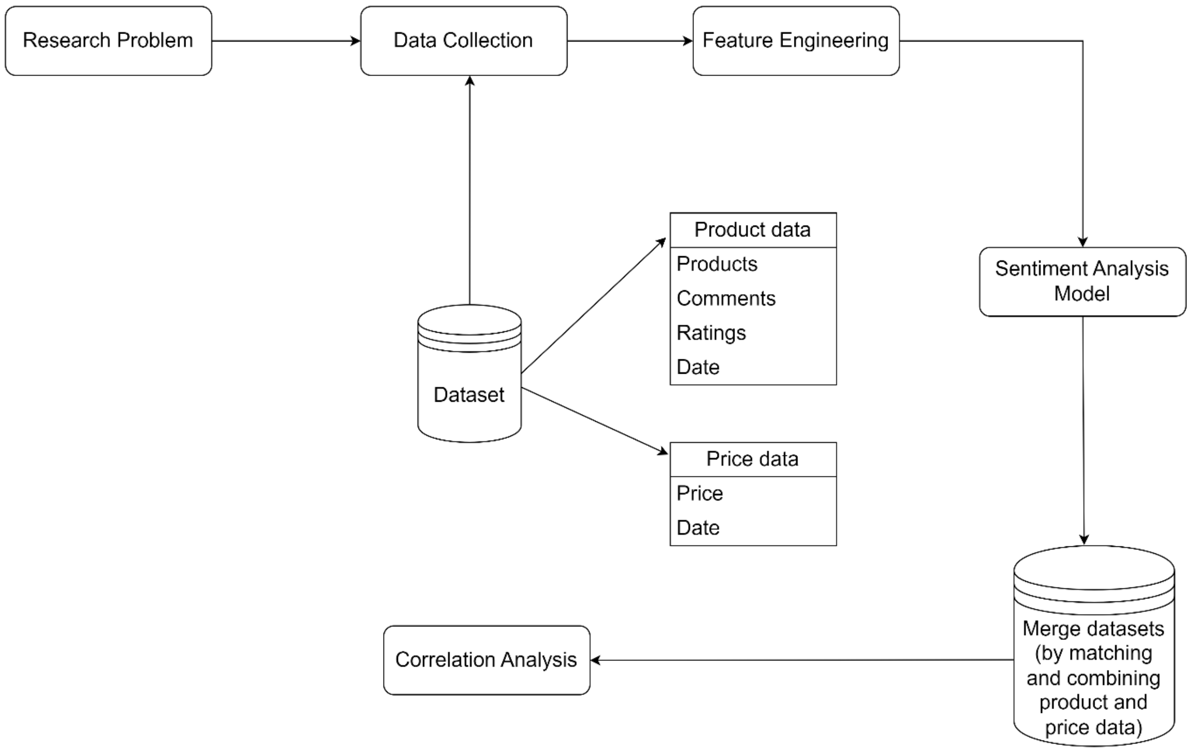


Figure 7: Methodology diagram for developing the sentiment analysis model and correlation analysis (Adapted from CRISP-DM, 1999)

As explained in Chapter 3, the dataset consists of Amazon product data and Amazon price tracker data.

- **Research Problem**
  As stated in the introduction Chapter, this study aims to investigate consumer behaviour on sentiment and price through the application of sentiment analysis and ABM.
- **Data Collection**
  To achieve the aim of this research, two data collection sources are identified:

Product data – Amazon product data is sourced from the Amazon.co.uk website and data extraction is performed using a web scraper tool to collect product details on Nike, Puma and Timberland trainers products. The following data are collected: product data such as the brand, comments which contain textual data for the consumer reviews dataset, ratings which is out of 5 denoting that the higher the rating the more postive the consumers experience and the date for matching purposes with price data.

Price data - The second dataset consists of price tracker data, collected from camelcamelcamel.com which is an Amazon price tracker website which monitors prices and dates for each of the 100 products. Date will also be used for matching purposes with product data.

Information gathered included comments that consumers have posted in the customer reviews section, ratings which are classed as 1 and 2 representing negative comments, 3 as neutral and 4 and 5 representing consumers had something positive to say about the product. Lastly, the date will be collected which is useful for correlation analysis which is explored in Section 4.4.

- **Feature Engineering**

  At this stage, data preprocessing will be applied. As customer reviews entail unstructured, noisy text, the first step in creating the sentiment analysis model is to clean the text so that text from human language can be transferred to a machine-readable format such as features for further processing and to train the model (Chauhan, 2018). The following was carried out for text preprocessing:

  - Tokenization which splits text into smaller pieces or 'tokens' which can be useful for calculating the frequency of words in the text.
  - Removing stop words by removing words which have no significance such as 'a, am, are, an, the' which doesn't necessarily alter the actual meaning of the consumer review.
  - Remove punctuation
  - Remove numbers
  - Remove URLs
  - Lowercase letters

- Removing duplicates
- Remove whitespace
- Remove special characters
- Lemmatization

Text vectorization takes place as a feature engineering technique that transforms tokenized reviews into a numerical vector for NLP model training. Two approaches were undertaken for text vectorization; Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). BoW records the frequency of words in a document and discards information about the order, structure or grammatical details. The second approach, TF-IDF is implemented to transform the piece of text into a structured format, and is commonly used in text mining (Bafna, Pramod, Vaidya, 2016). A vectorizer is formed to assess relevancy of words in a piece of text which classifies the text into positive, negative or neutral by converting the words into an array of numbers (Bhullar *et al.,* 2022).

Another aspect of feature engineering that will be undertaken in this research is topic modelling. Topic modelling aims to uncover topics which is beneficial to understand and extract hidden topics from large volumes of text. Previous studies have used topic modelling for classifying text in a document which identifies topics per document (Li, 2018). As online reviews are large in volume and consist of unstructured data, manually reading through and identifying topics can be very labour intensive, therefore using a topic modelling algorithm such as LDA will enable automation of scanning through text documents and identifying topics. Therefore, this study will use a topic modelling approach for discovery of hidden themes in online consumer reviews.

- **Sentiment Analysis Model**

  The next step in the cycle is to develop a sentiment analysis model. Figure 8 details the process of constructing a sentiment analysis model. Polarity is based on how positive, negative or neutral a piece of text is. The model is built on a rule-based system which uses a lexicon-based approach that counts the number of positive, negative and neutral words in a document and classifies them according to its sentiment.

- **Merge datasets**

  Product and price data are matched and combined in order to perform correlation analysis to observe if there is a correlation between sentiment and price. For example, if there is an online review posted on 27th August 2019 and there is price data available for the exact date; this is considered a match and will be utilized in this study. However, if there is price data available for this date and no online review was posted on the same date and vice versa, this is considered not a match and will be disregarded.

- **Correlation Analysis –** The final step in the methodology diagram is in regards to visually observing whether there is a correlation between sentiment and price in the form of graphical outputs. As part of this step, exploratory data analysis will be carried out which is shown in Section 4.3. The significance of exploratory data analysis is to perform initial investigations of the data which aids in identifying patterns and creating graphical visualizations (Amin, 2020).

Figure 8 shows the sentiment analysis modelling process which includes the following stages: collecting Amazon UK reviews, web scraping, preprocessing, sentiment classification which results in the last stage, sentiment polarity.

- ***Amazon UK reviews***

  The consumer review dataset is formed through the collection of comments that consumers have posted on the Amazon.co.uk website. Consumer reviews are beneficial to analyze as it exposes the opinions of consumers on a product.
- ***Web scraping***

  As mentioned in Section 3.3, a web scraper tool was applied to extract product data such as products, comments, rating and date from the Amazon.co.uk website.
- ***Preprocessing***

  In order for the sentiment analysis model to be implemented, preprocessing is required to clean noisy, textual data to be used for efficient NLP modelling in a format that the sentiment analysis model can read. The steps of preprocessing can be referred to earlier in this Section at the Feature Engineering step. This stage allows raw consumer review data to be transformed and organized into a format so the sentiment analysis model can be implemented by the machine learning model.

- ***Sentiment classification***

  Using a lexicon-based approach, TextBlob and VADER are applied to detect how positive, negative or neutral a review is. In addition, as explained earlier in this Section feature selection uses BOW and TF-IDF to select relevant words from the text for classification purposes.

- ***Sentiment polarity***

  The goal of sentiment analysis is to categorize sentiment according to polarity: positive, negative, or neutral.

Figure 9 shows the output for the sentiment analysis model. After the sentiment analysis model has been created, wordclouds will be generated for exploratory data analysis for visualization purposes. Words that are frequently discussed among consumers are illustrated giving a more in-depth understanding of consumers opinions of the trainers product. These words are associated with either positive, negative, or neutral sentiment. Wordclouds can be used in the initial stages of analysis for further advanced analysis. Moreover, topic modelling will be carried out to deduce the key topics consumers are discussing in the online reviews for the trainers products.

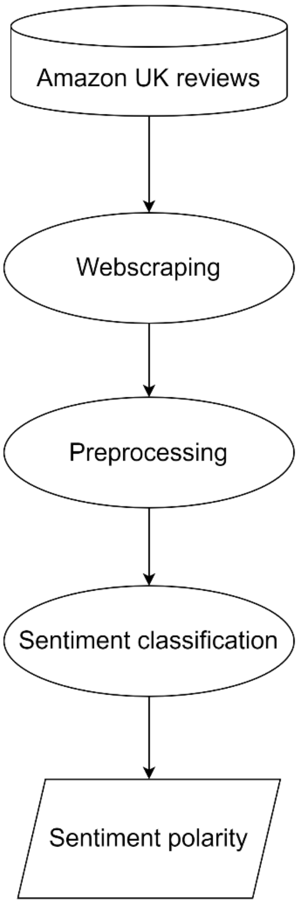


Figure 8: Sentiment Analysis modelling, (Adapted from Medhat, Hassan and Korashy, 2014).

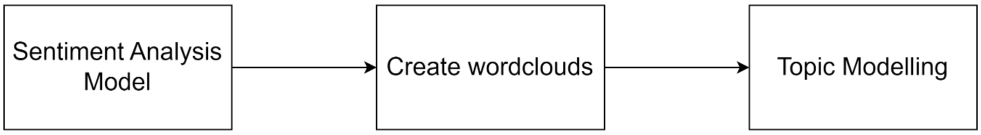


Figure 9: Output for sentiment analysis model

4. 3 Exploratory data analysis

Exploratory data analysis is undertaken to investigate and analyze the dataset to summarize and visualize the data to discover initial patterns and for the purposes of correlation analysis. The advantage of implementing exploratory graphics is that it helps to prepare for a more insightful analysis. Exploration is a critical first step leading to more advanced analysis such as ABM to observe interactions between consumers and sellers in a dynamic marketplace environment. The creation of the sentiment analysis model was performed using Jupyter Notebook which is a web-based interactive platform used in Python.

4.3.1 Preprocessing

Figure 10 shows a dataframe from Jupyter Notebook (Kluyver *et al.,* 2016) which displays a sample of five consumer reviews for the Nike B00CGVQHLM product. A total of 87 reviews were collected from this product. The columns consist of ‘Date’ when the online review was posted by the consumer, ‘ReviewText’ which includes raw textual data extracted from the online reviews and ‘cleanText’ which represents clean textual online review data after preprocessing. From Figure 10, the difference between ‘ReviewText’ and ‘cleanText’ can be observed after text preprocessing has taken place. For example, the first review (which has an index of 0), shows numbers are removed as ‘2’ and ‘0 5’ aren’t in the ‘cleanText’ Section. Stop words are also removed such as ‘go’.

Figure 11 shows a screenshot from Jupyter Notebook after sentiment analysis has taken place. The following columns are presented: ‘Date’ of when the consumers posted the review, ‘Rating’ shows the number of stars out of 5 the consumer gave the product where the higher the rating, the more positive the consumer found the experience, ‘Sentiment’ indicates whether the rating the consumers gave were positive, negative or neutral; a rating of 5 and 4 is categorized as positive sentiment, 3 as neutral sentiment and 2 and 1 as negative sentiment. Next, is the ‘Label’ column where sentiment is labelled as a numerical factor for clearer and a more efficient way to quickly identify which sentiment the review belongs to. Similarly, the category consists of positive sentiment labelled as +1, neutral sentiment as 0 and -1 as negative sentiment. The next column is ‘cleanText’ which was explained earlier in this Section which is the review data after preprocessing has occurred. The next two columns ‘polarity’ and ‘subjectivity,’ uses TextBlob to determine how positive or negative a sentiment is as well as how subjective the comment is. For example, the first comment shows a polarity score of 0.3 indicating slightly positive sentiment. Comparing this to the second comment, which has a polarity score of 1 indicating very high positive sentiment. The subjectivity scores for both comments are high, particularly comment 2 with a score of 1 indicating very subjective reviews. The next four columns, ‘neg,’ ‘neu,’ ‘pos,’ and ‘compound’ show the output scores after VADER has been applied. The compound scores calculates the sum of negative, neutral and positive scores and then normalized between -1 which indicates a very negative sentiment and +1 which indicates very positive sentiment. For example, the first comment shows 4% negative, 51.9% neutral and 44.1% positive. As there is a higher percentage of positive sentiment than negative sentiment this comment is more positive. This review has a compound score 0.9 indicating very high sentiment. Compared to the second review, with a compound score of 0.5 indicating a lower positive sentiment, as the second word is 'dislikes' which has a negative connotation, unlike the first word which is positive, 'awesome.' A summary of each column is provided in Table 3.

	Date	ReviewText	cleanText
0	2018-08-31	2 thumbs nike usually go 0 5 issues fit traine...	thumbs nike usually issues fit trainer look gr...
1	2019-09-04	awesome dislikes bought yesterday already wase...	awesome dislikes bought yesterday already wase...
2	2018-05-03	five stars great fit	five stars great fit
3	2018-03-29	five stars nothing wrong product small	five stars nothing wrong product small
4	2017-01-10	five stars excellent	five stars excellent

Figure 10: DataFrame showing difference between 'ReviewText' and 'cleanText' after preprocessing

	Date	Rating	Sentiment	Label	cleanText	polarity	subjectivity	neg	neu	pos	compound
0	2018-08-31	5	positive	+1	thumbs nike usually issues fit trainer look gr...	0.342857	0.642857	0.040	0.519	0.441	0.9243
1	2019-09-04	5	positive	+1	awesome dislikes bought yesterday already wase...	1.000000	1.000000	0.129	0.574	0.297	0.5423
2	2018-05-03	5	positive	+1	five stars great fit	0.600000	0.575000	0.000	0.233	0.767	0.7650
3	2018-03-29	5	positive	+1	five stars nothing wrong product small	-0.375000	0.650000	0.000	0.662	0.338	0.3724
4	2017-01-10	5	positive	+1	five stars excellent	1.000000	1.000000	0.000	0.351	0.649	0.5719

Figure 11: DataFrame showing application of sentiment analysis

Table 3: Definitions of columns in DataFrame

Date	Date of when consumer left a review on the Amazon.co.uk website.
Rating	Rating entered by a consumer with a range of 1-5 where 1 is considered to be the lowest rating and 5 being the highest.
Sentiment	Based on rating categorization: a rating of 5 and 4 is positive, 3 is neutral and 2 and 1 are negative.
Label	Based on rating categorization: a rating of 5 and 4 is labelled as +1, 3 is 0 and 2 and 1 is -1
cleanText	Text of the review after text preprocessing has been applied.
polarity	The higher the value of polarity, the more positive the piece of text and the lower the value, the more negative the text is.
subjectivity	Float number within the range of 0 and 1 where 0 is very objective and 1 is very subjective.
neg	How negative the review is.
neu	How neutral the review is.
pos	How positive the review is.
compound	Calculates the sum of the valence scores, which is then adjusted according to the algorithm rules and normalized between -1 indicating very negative sentiment and +1 indicating very positive sentiment (Todi, 2019).
length	Length of characters in review.

From conducting the literature review, TextBlob is used to measure polarity and subjectivity scores while Vader is used to measure compound scores. Figure 12 applies TextBlob library in Jupyter Notebook to output the polarity and subjectivity scores which is shown by the top line of code. The next piece of code refers to assigning 'df,' which is the abbreviation for DataFrame, which is the name given to the online review dataset. As the dataset consists of unstructured data, it needs to be converted into a string which stores and allows for manipulation of textual data. The next two lines of code applies TextBlob using polarity and subjectivity metrics. The last three lines of code assign polarity and subjectivity scores to the DataFrame.

As this research is also interested in the compound score which will be used to determine whether there is a correlation between sentiment and price; VADER is also imported as a library which can be seen in Figure 13. Similarly, with TextBlob, the first step is to import VADER. Next, VADER is applied to 'df', and the last piece of code refers to the first five rows of output as seen in Figure 11.

```
from textblob import TextBlob

df['cleanText']= df['cleanText'].astype(str)

pol = lambda x: TextBlob(x).sentiment.polarity
sub = lambda x: TextBlob(x).sentiment.subjectivity


df['polarity'] = df['cleanText'].apply(pol)
df['subjectivity'] = df['cleanText'].apply(sub)
df
```

Figure 12: Coding screenshot for applying TextBlob

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
analyzer = SentimentIntensityAnalyzer()
```

```
sent = df['cleanText'].apply(lambda x: analyzer.polarity_scores(x))
df = pd.concat([df,sent.apply(pd.Series)],1)
```

```
df.head()
```

Figure 13: Coding screenshot for applying VADER

4.3.2 Exploratory data analysis for Nike product

Referring to Figure 11 the third review (index 2), has a polarity of 0.6. As polarity lies within the range of -1 and 1, the first review indicates a positive review while the subjectivity score refers to the review being quite subjective and likely that it signifies personal opinion. In addition, as the consumer gave it a rating of 5; this adheres to positive sentiment. As subjectivity generally refers to opinion, emotion or judgment, the subjectivity of the third review is 0.575 indicating that it is more likely to be the individuals' opinion. The compound score 0.7650 indicates positive sentiment. The compound score for the first review (index 0) is 0.9243 indicating a very high positive sentiment.

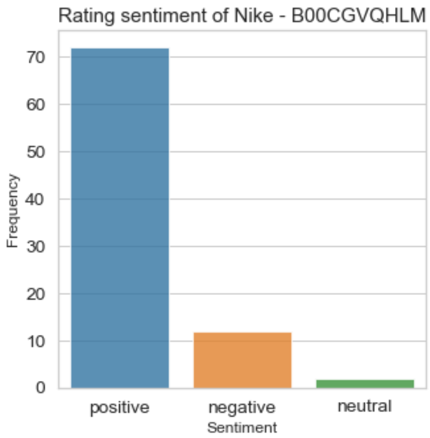
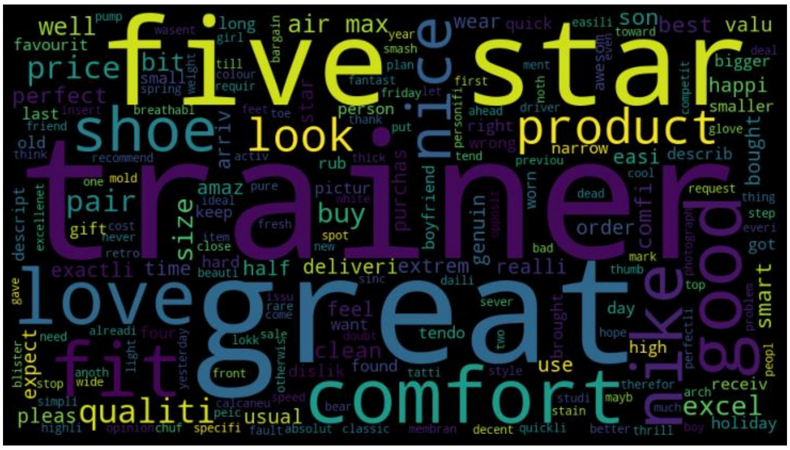


Figure 14: Rating sentiment of Nike product

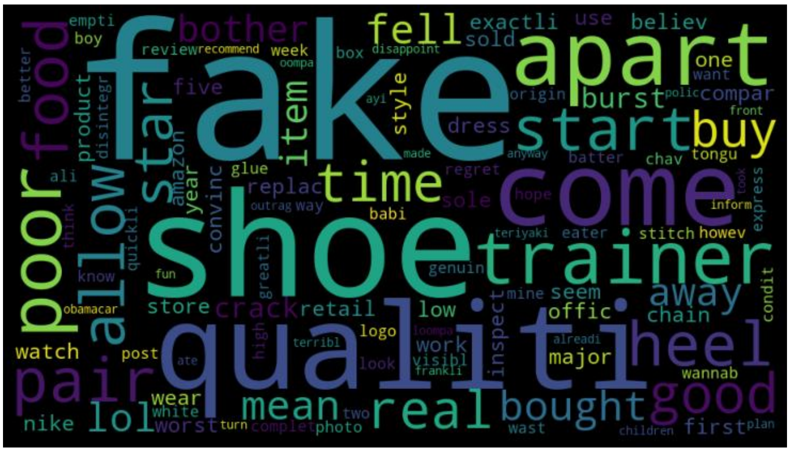
Figure 3 shows the frequency of positive, negative, neutral sentiment. Positive sentiment has the highest frequency indicating the majority of consumers left positive reviews. There are about 11 negative reviews and 1 neutral review for the Nike B00CGVQHLM trainers product.

Figure 15 (a) visualizes sentiment analysis in the form of a wordcloud which represents the most frequent words associated with positive sentiment. Words such as 'great', 'five star' and 'great' are of a larger size and indicates that more consumers have used these terms when describing this Nike trainers product. The word 'trainer' is also highlighted as being one of the main words being expressed as this is describing a footwear product. The next common word is 'great' indicating positive sentiment left by many consumers. Other words associated with this product are, 'size,' 'price,' and 'air max' which are less common. Figure 15 (b) shows words depicting negative sentiment for the Nike B00CGVQHLM product. One of the words which is quite visible from the wordcloud is 'fake.' This is a very negative connotation and the size of the word in the wordcloud denotes that many users have formed the opinion that this product could be fake. On the left side of the wordcloud the word ‘poor’ is also noticeable which may refer to the quality of the product. Figure 15 (c) visualizes neutral words associated with the B00CGVQHLM product. The brand name ‘nike’ is quite common in this wordcloud along with ‘trainer’ and ‘feet’ indicating footwear products. Neutral words denote neither good or bad sentiment and the company reputation would not be affected by this form of emotion (Schaer, 2020).

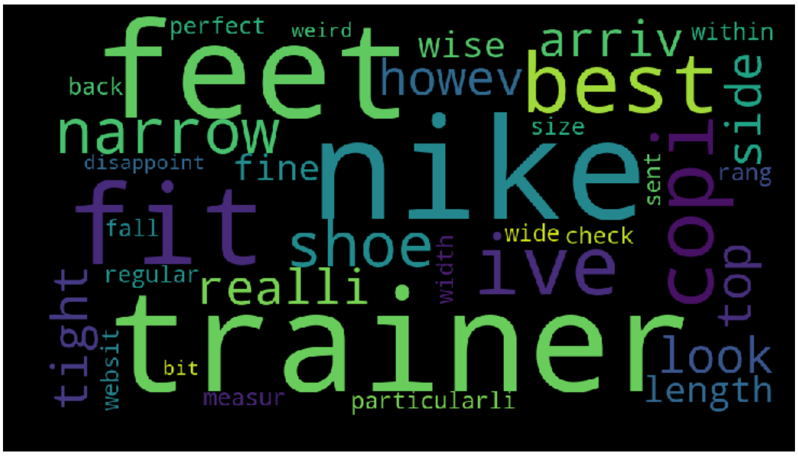
As mentioned in Chapter 2, topic modelling is beneficial for detecting the most common topics from online consumer reviews. Figure 16 illustrates the implementation of topic modelling for the Nike B00CGVQHLM product. ‘Trainers’ represented by the red colour denotes the most common word consumers post for this product which was mentioned more than 25 times. The words ‘great’ and ‘stars’ represented by the orange and gold colour respectively, was mentioned around 22 times. The word ‘good’ represented by the dark green colour was mentioned around 17 times and the brand name ‘nike’ was mentioned 16 times. The brand name ‘nike’ is the fifth most common topic derived from the 34 Nike trainers products that were collected for this study indicating an association with the brand. Brand monitoring through the application of topic modelling is beneficial to companies to improve their decision making. This analysis may also be valuable for branding companies to monitor brand loyalty by observing what consumers are discussing about their brands such as Nike, Puma and Timberland. The words ‘quality,’ ‘comfortable’ and ‘fit’ are the next three common words that appear for this product that refer to the features of the trainers. Lastly, the word ‘nice’ is the tenth most common word. The following words, ‘great,’ ‘good,’ and ‘nice’ describe positive sentiment which is reinforced by the rating sentiment graph for the Nike B00CGVQHLM product which can be observed in Figure 14 showing a high positive sentiment than negative and neutral sentiment.



(a) Positive sentiment



(b) Negative sentiment



(c) Neutral sentiment

Figure 15: Wordclouds for Nike B00CGVQHLM product, visualizing positive, negative and neutral sentiment

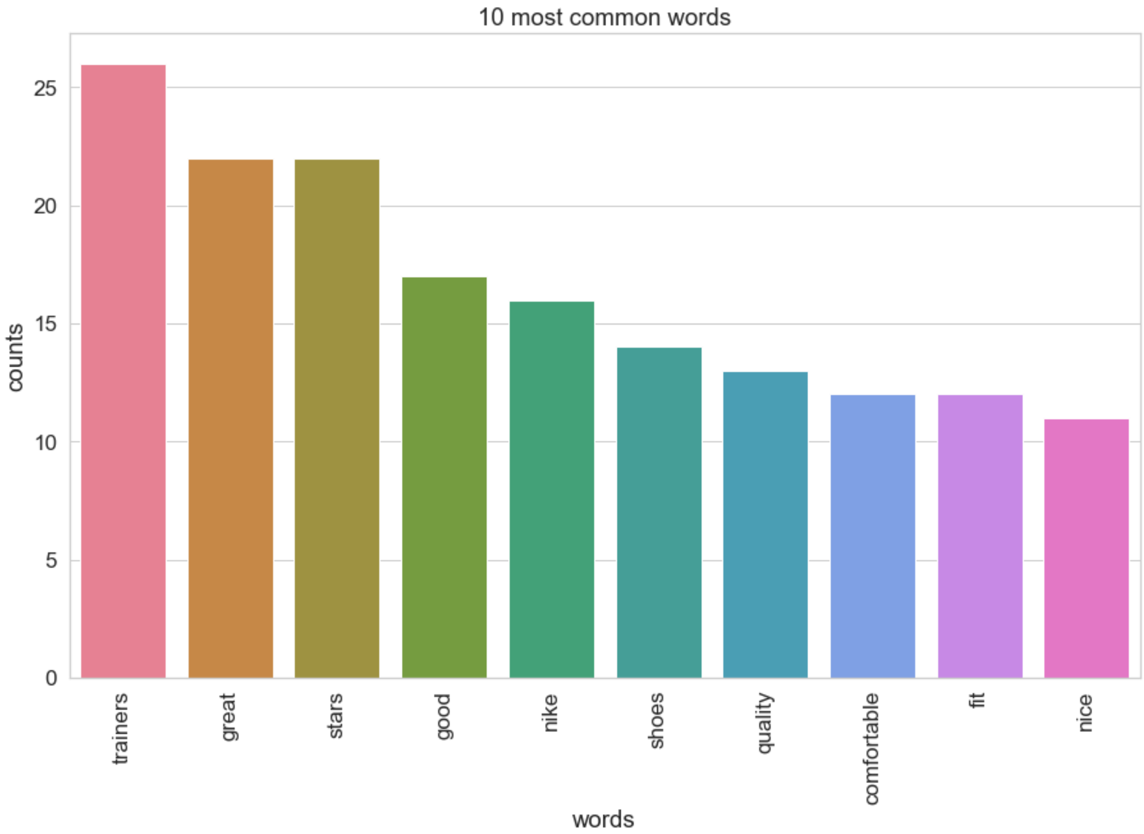


Figure 16: Topic modelling for Nike B00CGVQHLM product

4.3.3 Exploratory data analysis for Puma product

Figure 17 shows that there is a higher positive sentiment rating here compared to negative and neutral sentiment.

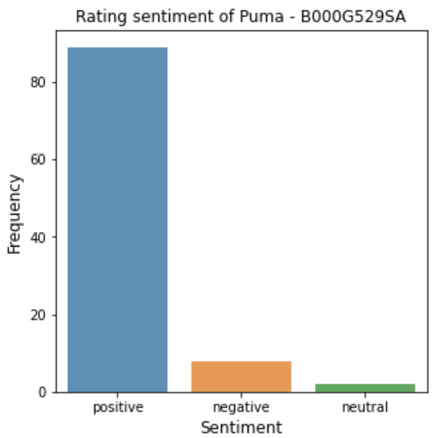
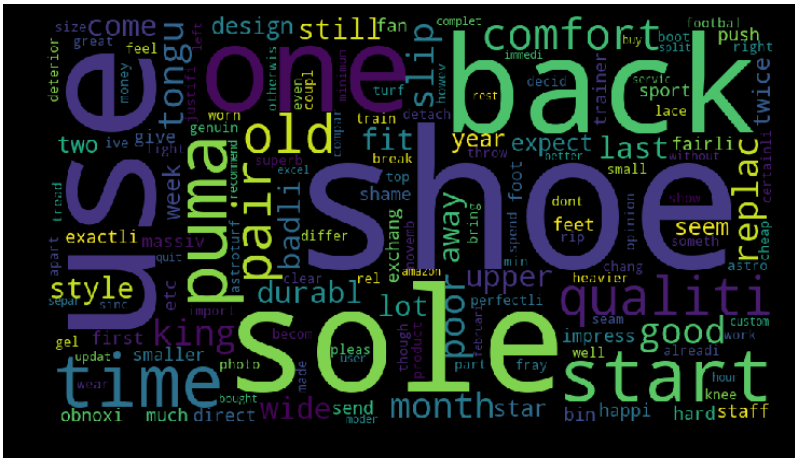


Figure 17: Rating sentiment for Puma B00G529SA product

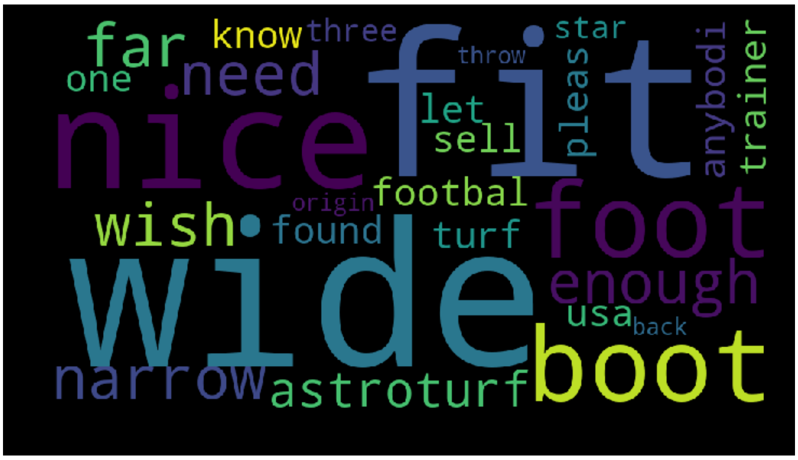
Figure 18 (a) shows the wordcloud generated for the Puma B00G529SA product reviews that were rated positive. Words like ‘great’, ‘fit’ and ‘good’ are larger indicating more recurrent words. Some of the words associated with negative sentiment for the Puma B000G529SA product which can be seen in Figure 18 (b) is ‘old’ and ‘poor’ which may refer to consumers describing the style as quite outdated and insinuating poor quality. ‘Shoe’ is also quite frequent here and can mean the word ‘trainer’ and ‘shoe’ as interchangeable meanings. One insight derived from this wordcloud is the word ‘good’ which should not have been picked up by the algorithm as negative sentiment. Figure 18 (c) shows a wordcloud with neutral sentiment consisting of recurrent words such as ‘fit,’ and ‘wide,’ which both denote how well the product fits the consumer.

Figure 19 shows the topic model for the Puma B00G529SA product. The first most common word which is mentioned more than 35 times is the word, ‘good’ supporting Figure 17 where there is more sentiment derived from this product than negative and neutral. The next most common word is ‘fit’ represented by the orange colour and the seventh most common word, ‘comfortable’ is represented by the turquoise colour. These two words were also mentioned in Figure 16 for the Nike product which again refers to the features of the trainers and how well the trainers fit the consumer. The third most common represented by the gold colour with more than 30 word counts is ‘great’ further reinforcing positive sentiment about this particular Puma product. The fourth most common word is ‘trainers’ mentioned 30 times represented by the dark green colour, followed by ‘pair’ indicated by the green colour. ‘Pair’ could be referred to as a pair of trainers when describing the product. The sixth most common word is ‘puma’ referring to the name of the brand shown by the light green colour. ‘Astro’ is mentioned as the eighth most common word, followed by ‘boots’ and ‘football’ which could relate to a sport such as football as ‘astro’ and ‘boots’ are relatively associated with footwear products worn by football players.

(a) Positive sentiment



(b) Negative sentiment



(c) Neutral sentiment

Figure 18: Wordclouds for B00G529SA product, visualizing positive, negative and neutral sentiment

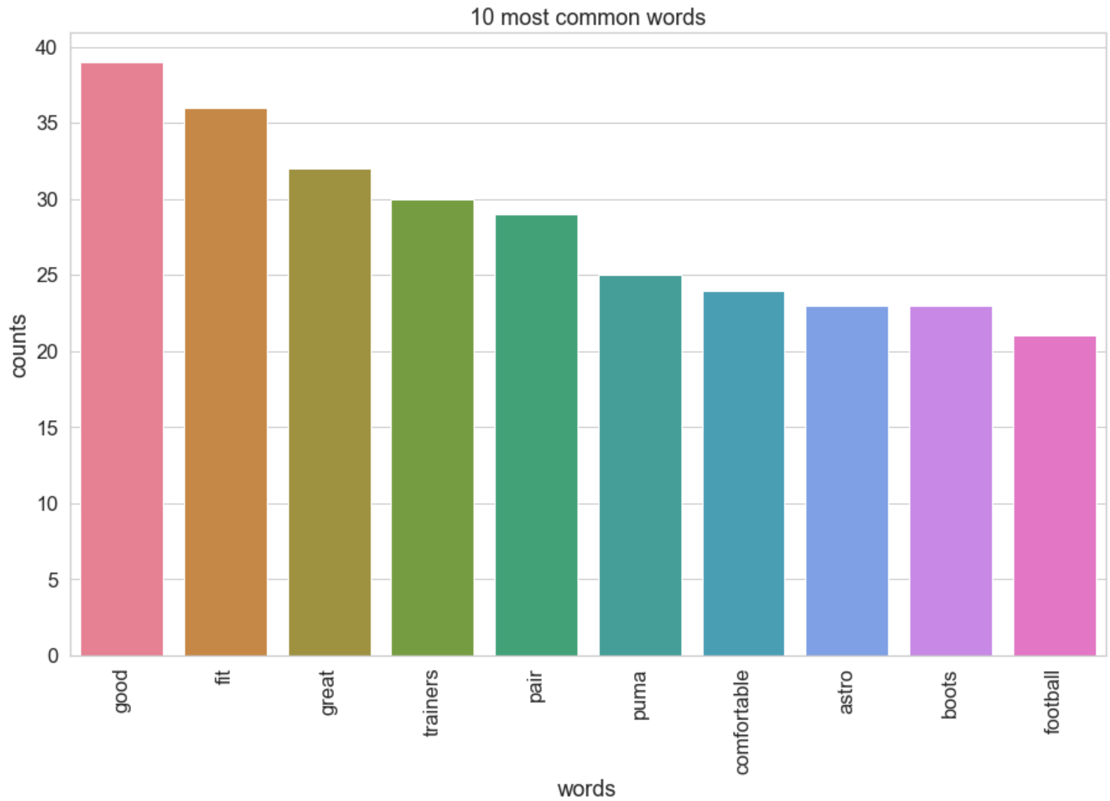


Figure 19: Topic modelling for Puma B00G529SA product

4.3.4 Exploratory data analysis for Timberland product

Figure 20 shows the rating sentiment for Timberland B008H2C6KA product. Overall, positive sentiment is higher for this product than negative and neutral sentiment. It is worth noting that all the 33 Timberland products collected, had much more consumer review data available than Nike and Puma trainers products. This product indicates a frequency of more than 400 consumer reviews collected.

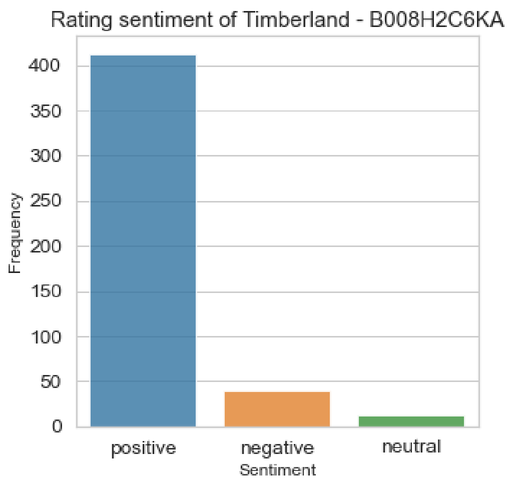
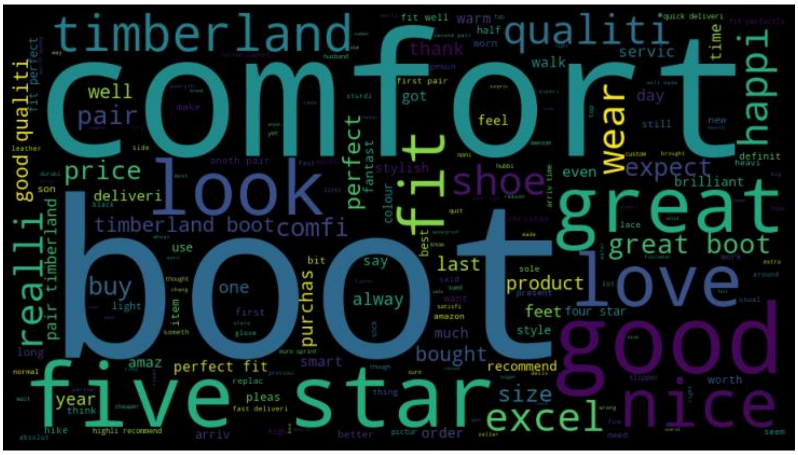


Figure 20: Rating sentiment for Timberland B008H2C6KA product

Figure 21 (a), visualizes a positive wordcloud for the Timberland B008H2C6KA product which illustrates the most frequent positive words as ‘comfort’ and ‘boot.’ From Chapter 3, Timberland prides itself on promoting comfortable products hence, the word comfort displayed here. Likewise, the word ‘boot,’ refers to Timberland’s outdoor footwear products. The words ‘five star’ and ‘timberland’ are also depicted here denoting positive sentiment. Similarly, to the negative wordcloud for the Nike B00CGVQHLM product, Figure 21 (b) displays a negative wordcloud for the Timberland B008H2C6KA product which shows the word ‘fake’ denoting consumers may feel that this product may not be authentic. Following on from this, ‘return’ is also visible here and consumers may wish to return this product due to unsatisfaction. Some of the smaller words portrayed are ‘disappoint’ and ‘bad’ all denoting negative sentiment. Figure 21 (c) shows neutral sentiment for the Timberland product. ‘price,’ ‘boot,’ ‘wear’ and ‘walk’ words are posted by consumers to describe this product and their experience. ‘price’ is mentioned which could refer to the higher cost that Timberland products project and is a factor in consumers purchasing decision making. Price is used as a variable in this research to achieve the objective of whether there is a correlation between price and sentiment which can be seen in Section 4.4.

It is noticeable from Figure 22 that there are considerably more word counts than Nike and Puma products. This is due to a larger volume of data collected for the Timberland B008H2C6KA product. The most common word represented by the red colour is ‘boots’ as Timberland’s footwear products are predominantly called boots and are interchangeable with the word trainers. There are more than 250 word counts for the word ‘boots,’ followed by ‘good’ and ‘great’ both representing positive sentiment. The fourth most popular word is ‘stars’ which relates to the rating and ‘comfortable’ represented by the lighter green colour as the fifth most popular word. ‘Quality’ is referred to as the seventh most recurring word represented by the turquoise colour which depicts what consumers wrote about the quality of the product which is useful for quality management in a business. The next two common words are ‘pair’ and ‘fit’ which are both also mentioned for the Puma product. The tenth most common word is ‘nice’ which again relates to positive sentiment.



(a) Positive sentiment

(b) Negative sentiment

(c) Neutral sentiment

Figure 21: Wordclouds for Timberland B008H2C6KA product visualizing positive, negative and neutral sentiment

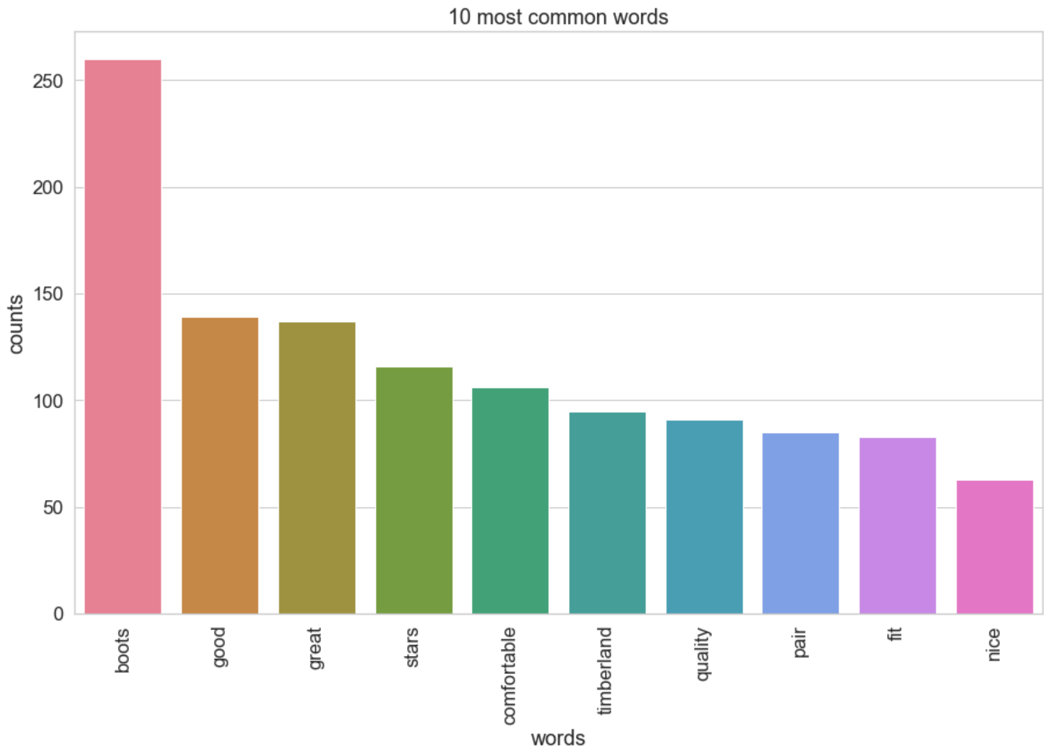


Figure 22: Topic modelling for Timberland B008H2C6KA product

4.4 Discussion

Sentiment analysis was performed for exploratory data analysis for 100 trainers products for the Nike, Puma and Timberland brands. For the purposes of this research, Nike (B00CGVQHLM), Puma (B00G529SA) and Timberland (B008H2C6KA) products were visually presented in the form of positive, negative and neutral wordclouds and topic modelling in this Chapter. The exploratory data analysis highlights the most frequent occurring words in the online reviews for the products mentioned and the top common topics. This was an essential step as it achieves Objective 2 of this research as it identifies, and extracts sentiment derived from online consumer reviews from the Amazon website using a sentiment analysis approach. This stage paves the way for the next step in this research which is to perform correlation analysis between sentiment and price. This can be observed in the next Section.

4.5 Correlation Analysis

This Section will provide descriptive statistics as well as correlation analysis in the form of graphical outputs. Descriptive statistics provides a summary to the dataset collected which consisted of 100 trainer products which are split into 3 brands: 34 products for Nike, 33 products for Puma and 33 products for Timberland.

Table 4 shows a sample of the 100 products collected, for mean price and mean compound score. The full Table including the 100 products can be seen in Appendix 1. Pearson's correlation coefficient was applied to calculate the correlation for each trainers product to indicate the strength of a linear relationship between sentiment and price variables (Drury, 2022). The correlation coefficient for the Nike B00CGVQHLM product is -0.61 indicating the two variables are slightly negatively correlated. Puma B000G529SA product has a high positive correlation of 0.85 and Timberland B008H2C6KA product has a lower positive correlation of 0.15. For the purposes of this thesis, summary statistics for Nike B00CGVQHLM, Puma B000G529SA and Timberland B008H2C6KA products can be observed in Table 4. Refer to Appendix 1 to view summary statistics for the total 100 trainers products collected for this research. When conducting this analysis, it is worth noting that around a third of the data was not suitable for use in this research due to not enough data available to conduct correlation analysis on those products and hence, was disregarded in this study. This can be seen in Table 5 along with the number of positively and negatively correlated products for each brand.

Table 4: Summary statistics for Nike B00CGVQHLM, Puma B000G529SA and Timberland B008H2C6KA products

	Product		
	Nike - B00CGVQHLM	**Puma – B000G529SA**	**Timberland – B008H2C6KA**
Mean Price (£)	93.54	72.47	89.93
Mean compound score	0.27	0.76	0.62
Correlation coefficient	-0.61	0.85	0.15

Table 5: Correlation table for Nike, Puma and Timberland products

Brand	Positive correlation	Negative correlation	Not enough data
Nike	6	11	17
Puma	11	13	9
Timberland	11	10	12

Figure 23 shows the correlation graph of price and polarity and subjectivity scores for the Puma B00G529SA trainers product. The polarity score represented by the orange line starts very high indicating strong positive sentiment in March 2016, which declines to around 0.3 in June 2018. This lower value of positive sentiment increases to a very strong polarity score of 1.0 towards the end of 2018 and declines again to around 0.2 in March 2019. There is a gradual increase after this towards June 2019. This polarity pattern is in contrast to the price trend. For example, as price increases from the beginning of the period in March 2016 to June 2018, polarity score decreases from 1 to just below 0.3 in the same timeframe. There is a rapid increase for polarity score from June 2018 to October 2018 from a polarity score of just under 0.3 to 1.0, which then decreases rapidly to just below 0.2 in March 2019. From June 2018 to November 2018, there is a slight decrease in price from £82 to £79. Subjectivity score increases from November 2018 to March 2019, as well as price steadily increasing by about £1. Polarity and subjectivity scores are good indicators of how subjective or objective a piece of text is. However, as this study is more interested in a total sentiment score, polarity and subjectivity scores will not be used further in this research. As VADER is more optimized for social media data (Athar, 2022) and has an aggregated score of positive, negative and neural metrics which is normalized between +1 and -1, for this reason, compound score is most suitable to use for correlation analysis purposes to determine whether there is a correlation between sentiment and price. The Appendix Section displays graphs on polarity and subjectivity scores for the rest of the trainers products.

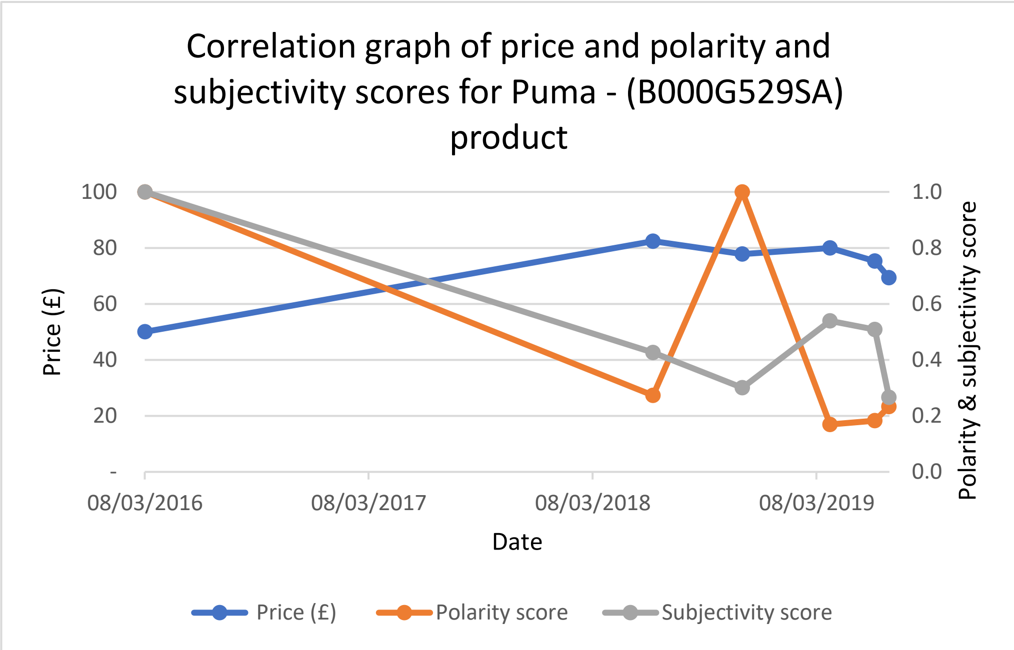


Figure 23: Correlation graph of price and polarity & subjectivity scores for Puma B000G529SA product

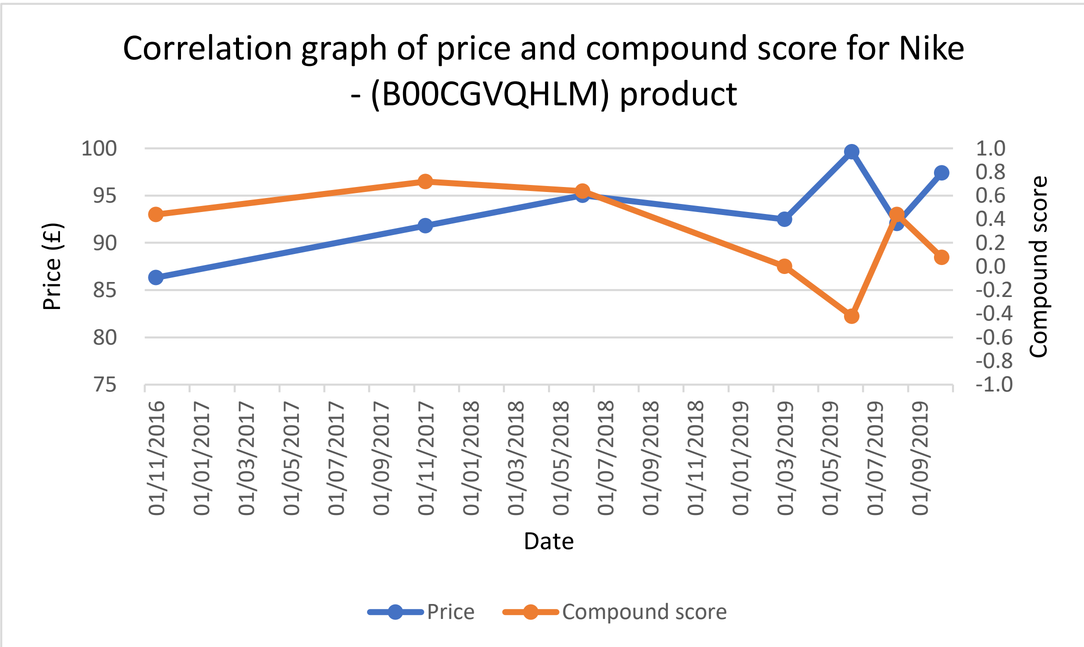


Figure 24: Correlation graph of price and compound score for Nike B00CGVQHLM product

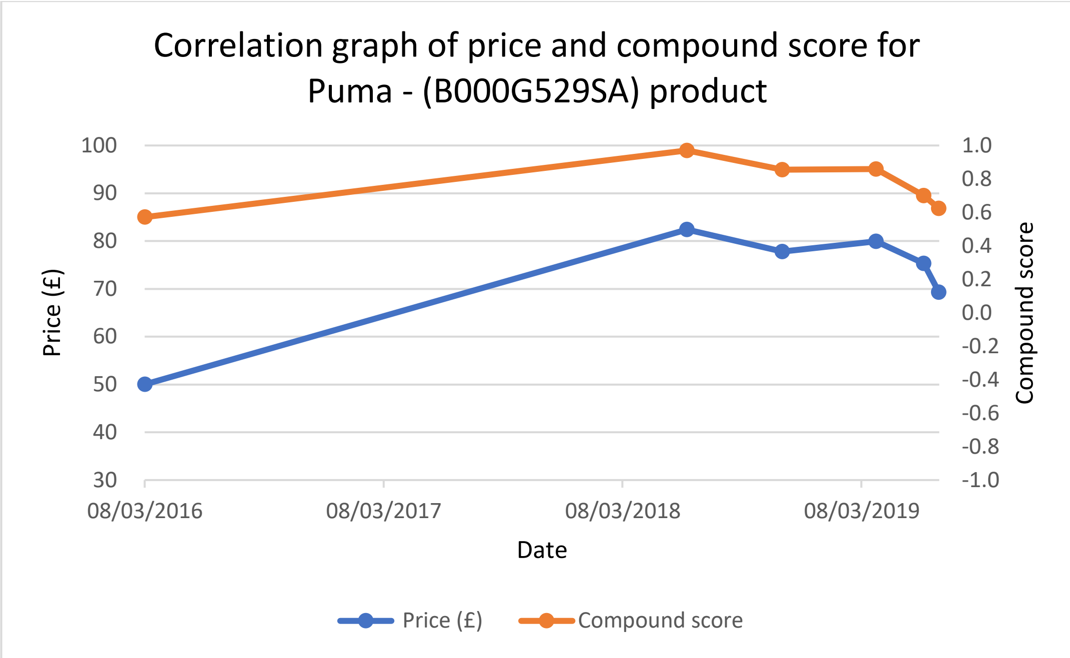


Figure 25: Correlation graph of price and compound score for Puma B000G529SA product

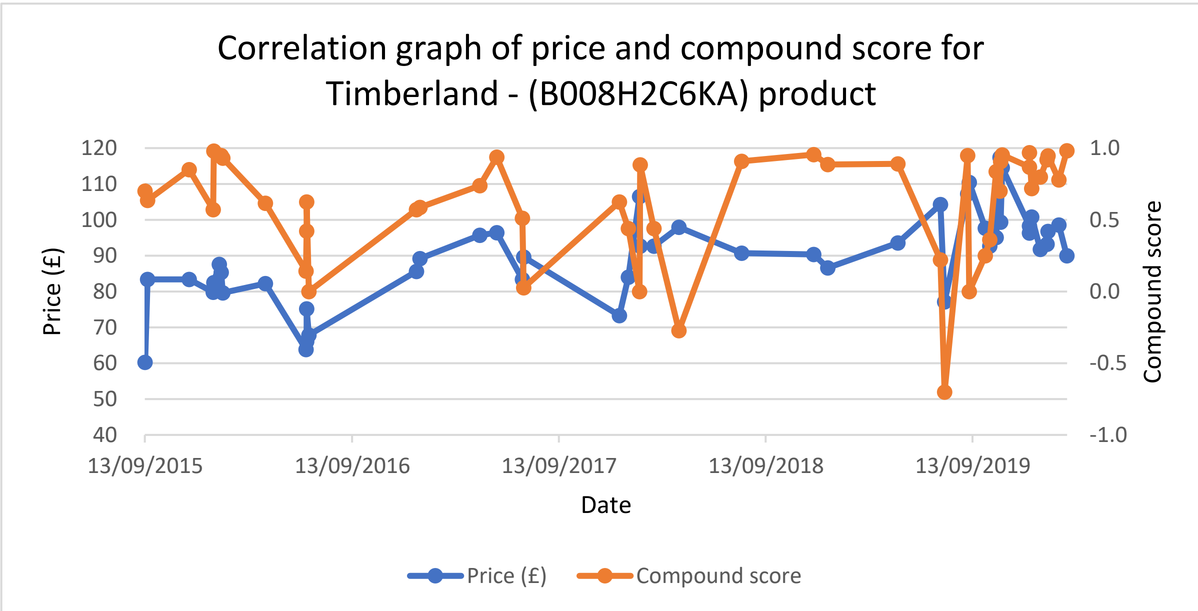


Figure 26: Correlation graph of price and compound score for Timberland B008H2C6KA product

Figure 24 shows correlation graphs between price and compound score. Compound score was plotted here as mentioned in Chapter 2 as the sum of positive, negative and neutral scores (Singh, Tomar and Sangaiah 2020). Price and compound score are plotted over time for the Nike

B00CGVQHLM product. This graph shows both price and compound score starts to increase from when the product was first introduced into the market. The blue line represents price and gradually increases from around £85 in November 2016 and increases by about £10 to £95 in June 2018. Compound score represented by the orange line starts off at 0.4 and increases to 0.7 to match the price line in June 2018. After this point, compound score decreases staggeringly to -0.5 denoting positive sentiment at the initial stages and then dropping to negative sentiment. Price also decreases slightly from £95 to around £92 and then increases to around £100 by June 2019. This shows about a £5 increase in a year. Compound score then starts to increase after July 2019 to 0.4 and drops to 0.1 in September 2019. As compound score increases in this period, price decreases from £100 to £91 displaying negative correlation. Price starts to increase to around £97 by October 2019.

Figure 25 displays a correlation graph of price and compound score for the Puma B000G529SA product. Price represented by the blue line and compound score represented by the orange line follow very similar patterns indicating a positive correlation. As price increases, compound score also increases from March 2016 to June 2018. After this, as price decreases from June 2018 to November 2018, compound score also decreases. Price decreases from £80 by around £1 to £79 and compound score decreases by 0.1 from 0.9 to 0.8. Price slightly increases towards the end of 2018 through to March 2019 and decreases after May 2019. Compound score also follows the same direction decreasing from 0.8 to 0.6.

Compared to Nike and Puma products, Timberland has many more matching data points as price tracker dates match the amazon review date as explained in Section 4.3. Figure 26 shows as there is a decrease in price from £80 to £60 in June 2016, compound score also decreases from 0.6 to 0. After this, as price starts to increase slightly, compound score also starts to increase slightly. Another interesting aspect of the graph is in December 2018, there are 2 data points closely matched where there is a very slight decrease in price and compound score follows. Towards the end of December 2019, patterns for both price and compound score follow the same trajectory resulting in positive correlation. The period of data collection was from September 2015 to December 2019. Price represented by the blue line, start at £60 and increases to £80 within the same month. As compound score, which is represented by the orange line, increases from September 2015 to December 2015 from a score of 0.6 to 0.8, price stays the same. After this, as compound score increases to around a range of 0.9 to 1.0, prices increase slightly to £90. Another noticeable insight is as compound score decreases drastically from 0.9 to 0.1 from February 2016 to June 2016, price also decreases from £80 to £61 in the same period. Compound score then increases within that month to 0.6 and drops again to 0, going from positive to neutral sentiment. Price also follows this pattern as price decreases from £75 to £63. As there is a rise in compound score from 0 or neutral sentiment, to 0.9 (positive sentiment), from the period of July 2016 to May 2017, price also follows from £63 to £97. Another interesting insight from this graph is as price decreases from £90 to £72 from July 2017 to December 2017, compound score increases from 0 to 0.6 denoting a negative correlation for this specific time period. From December 2018 to January 2019 there is a similar pattern which dips for both price and compound score. As price decreases from £90 to around £85, compound score also decreases slightly from slightly lower 1.0 to 0.9. In July 2019, compound score drastically increases from -0.8 denoting negative sentiment to around 0.9 indicating positive sentiment. During this period, price also increases from £88 to £111. Towards the end of 2019 there are multiple fluctuations for both price and compound scores. Overall, the correlation coefficient for the Timberland B008H2CKA product is 0.15 indicating a low positive correlation.

4.6 Chapter summary

This Chapter recognized the importance of analyzing unstructured, textual data that is commonly derived from social media platforms such as Amazon.co.uk website. By creating a sentiment analysis model, this research was able to extract key words from the online reviews according to sentiment. Positive, negative and neutral sentiment have been visualized using wordclouds, as well as applying topic modelling to illustrate the top, common topics associated with online consumer reviews. The next stage explored whether there is a correlation between sentiment and price for the 100 products that were collected in this study. From initial exploratory data analysis, Timberland has a greater volume of data available than Nike and Puma. One reason for this could be that Timberland generally has a higher price point than Nike and Puma and therefore, consumers are more likely to share their opinions and influence others purchasing decision.

The examples shown in this Chapter show that Puma had the highest correlated product compared to Nike and Timberland with a correlation coefficient of 0.85. To conclude this Chapter, sentiment analysis can be implemented as a tool to understand sentiment expressed and how it could be used to influence other potential consumers purchasing decision through opinions and recommendations. This demonstrates a demand for researchers and companies to extract valuable insights into a products positive and negative qualities, subsequently, differentiating themselves from competitors by improving their marketing strategies and customer experience. This Chapter contributed to the objectives of this research which is to determine whether there is a correlation between sentiment and price.

Chapter 5

ABM for Diffusion of Innovation

5.1 Introduction

This Chapter presents a discussion of ABM and its components. This chapter is structured as follows: Section 5.2 introduces the mechanics of diffusion of innovation by providing two examples, spread of disease model (Rand and Wilensky, 2008) and Schelling's segregation Model (Schelling, 1971). Section 5.3 explains the topology of ABM with an emphasis on social network analysis. Section 5.4 explores the design concepts or conceptualization of the agent-based model, Section 5.5 explores implementation of the model, Section 5.6 explains the experimentation and lastly, Section 5.7 concludes with a Chapter summary.

6.2 Mechanics of Diffusion of Innovation

There has been an array of studies on modelling epidemics or spread of diseases. The spread of disease model (Rand and Wilensky, 2008) is a popular basic diffusion model that analyzes "the transmission of infectious disease from one person to another through social connections" (Kumar and Sinha, 2021). This concept can be applied to the spread of information in a social network which aims to observe social relationships by incorporating social parameters. This Section explains the mechanics of diffusion of innovation in more detail by firstly exploring the spread of disease model (Rand and Wilensky, 2008) and Schelling's segregation Model (Schelling, 1971).

An example of applying ABM and examining spread of diffusion can be investigated through the concept of how a virus spreads to others that are in close proximity to the infected individual. When someone contracts a cold virus, this person may spread it along with exhibiting symptoms such as coughs or sneezes thereby infecting others that have come into close contact with the virus particles. People that have been in close contact with this person who has been infected with the virus will spread it to individuals within close proximity. The infected group may then spread to more people consequently infecting tens and hundreds of people which rises the rate of infection exponentially. Visualizing this on a graph such as in Figure 27, developed by Rand and Wilensky (2008), describes the spread of disease model which shows an increase in the rate of infection that flattens out over time as there are fewer people to infect. This can be represented as a simulation model in an ABM environment by creating agents where one agent or individual has the infection and spreads it to other agents or other people that come into contact with the infected individual. As this is relatively quite a simple model, more complex behaviour could be incorporated to explore what-if scenarios. For example, what is the impact on the rate of infection if the modeller increases the population? This study will apply this question to the current research problem in this study as one of the what-if scenarios such as what is the impact of the rate of adoption for trainers products in a population of 50 to 500, which can be observed in Chapter 6.

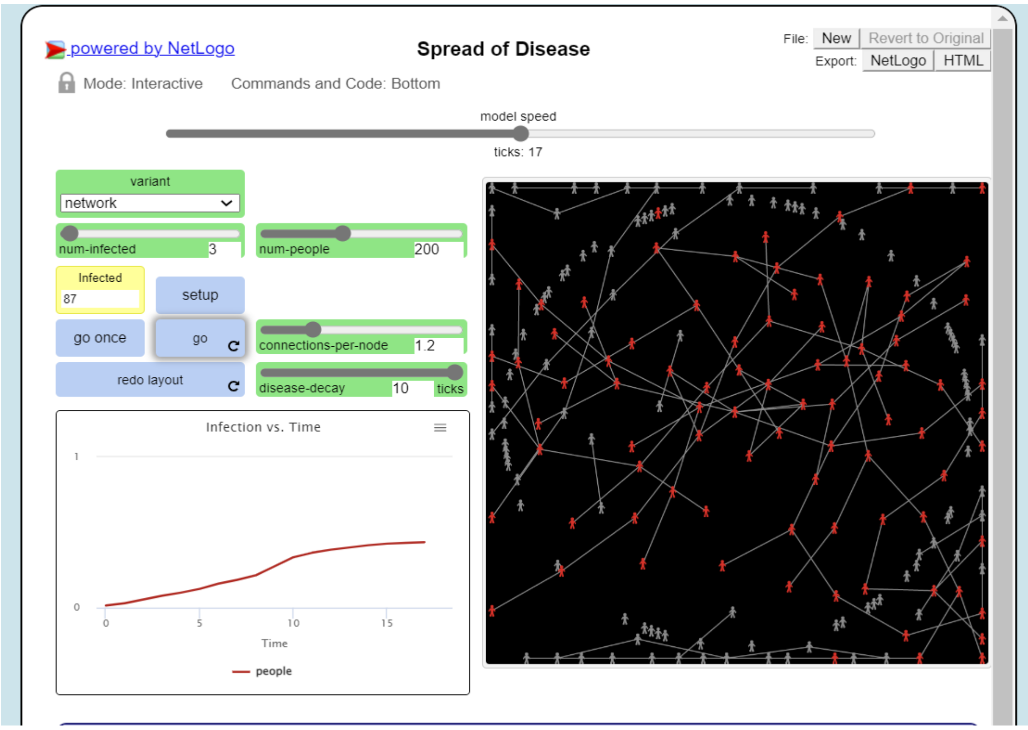


Figure 27: Spread of Disease model in NetLogo (Rand and Wilensky 2008)

As the previous spread of disease model (Rand and Wilensky, 2008) comprises of a social network topology, Schelling's segregation model (Schelling, 1971) is widely considered as developing one of the first agent-based models resulting in social patterns through the investigation of individual level actions. Figure 28, displays Wilensky's segregation model adapted from Schelling (Schelling, 1971), using the NetLogo web platform.

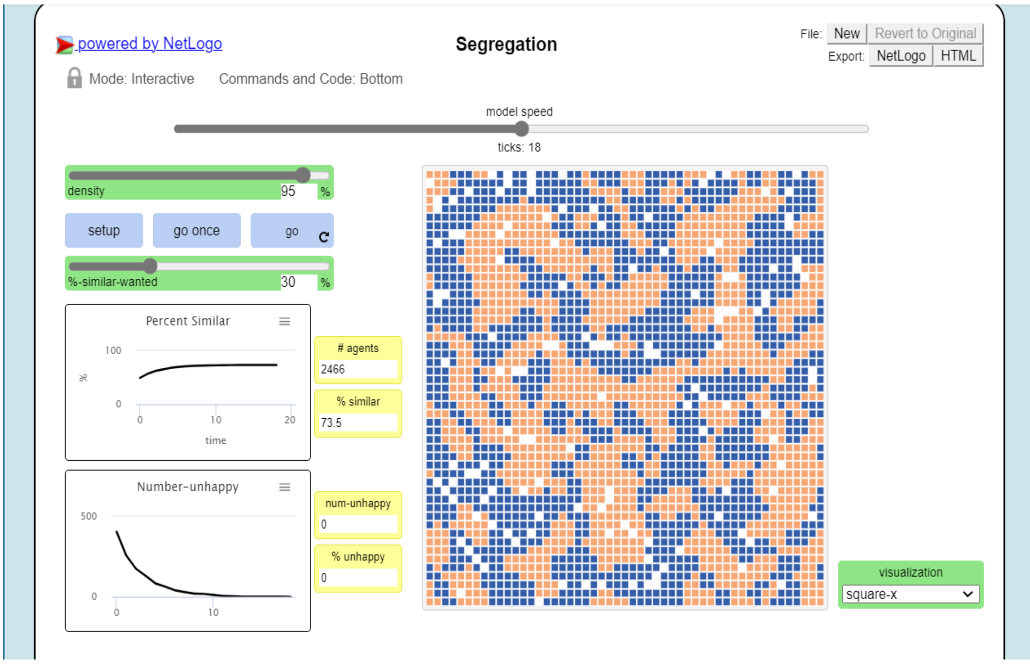


Figure 28: Segregation model based on Schelling's segregation model (Schelling ,1971; Wilensky, 1997)

Schelling's segregation model (Schelling, 1971) places the conceptualization of the model on a checkerboard-like diagram representing a household where each agent occupies a cell. The agents are divided into two groups and are represented by dimes and pennies. In this context, the model was implemented to understand how individuals make relocation decisions based on traits of their immediate neighbourhood depending on tolerance threshold. If the value of the fraction of neighbours which are represented as pennies surrounding each dime exceeds the threshold value, the dime would move to a random empty cell on the checkerboard which leads to segregation within a neighbourhood. The model, therefore, provides a theoretical basis for understanding the ethnic movements of individuals through the perception of residential preferences. However, there are limitations to this model. The simplification of Schelling's segregation model (Schelling 1978) does not take into account the complex origins of people's housing choices. Nevertheless, the model did create awareness for social patterns shown by individuals actions as even if no individual wants to conform to segregation, it will transpire "as long as people have a preference not to be in an extreme minority in their neighborhood" (Wilensky and Rand, 2015). Advances in computational technologies now allow ABM to incorporate more complex simulation approaches which can be applied as an extension to Schelling's segregation model (Schelling, 1978). For example, incorporating different ethnic groups, different demographics and as mentioned earlier, address the limitation of the original model to incorporate individuals housing choices to observe residential patterns (Hatna and Benenson, 2012).

5.3 Social Network Analysis topology for ABM

As this research study aims to observe interactions in a social context, the topology of ABM is beneficial for detailing agent relationships and interactions. In a network topology, the cells surrounding an agent in a grid structure is known as the neighbourhood. Information is shared through close contact interactions to nearby agents (Macal and North, 2009). Similarly, to the spread of disease model (Rand and Wilensky, 2008) as explained earlier, the model infects other people that are in close proximity to the infected agent. This process of infecting others can be applied to a social network analysis context based on the diffusion of innovation theory. For example, considering an agent that has already purchased a product, this agent can share their experience of the product to other agents in a neighbourhood through WOM. In essence, WOM is the process of sharing information to nearby agents. This concept can be applied to this study by incorporating the Bass model (Bass, 1969) as mentioned in Chapter 2, to observe the diffusion of innovation. The agent who owns the product called an innovator, spreads awareness of the product through means of social channels through WOM to their social network. Social network analysis not only detects how agents interact but also who they interact with. Hence, consumers interact with potential consumers in the ABM environment through the effect of social influence. As Simudyne (2019), explains, when agents are represented as people, their interactions are classed as social networks which can refer to the flow of information between millions of interacting individuals. ABM provides an understanding of how complex systems work under certain conditions by creating what-if scenarios. The synergy between ABM and social network analysis stipulates a richer model of a phenomenon by observing patterns of interaction between agents (Wilensky and Rand, 2015). As innovations disperse across social networks, the spread of disease model (Rand and Wilensky, 2008) is a good modelling topology for exploring diffusion of innovation (Valente, 1995). According to Wilensky and Rand (2015), the role of influencing people through social networks are greater at spreading innovations than others.

5.4 Conceptualization

As mentioned in Chapter 2, NetLogo consists of an environment represented by a square lattice. To set parameters, buttons, sliders and switches can be added as well as monitors and graph plots for visualizations. These elements can be moved by clicking the element and dragging to the desired location on the interface. These can also be edited by right clicking and selecting edit to adjust the parameter values. Elements can be deleted by right clicking and selecting 'Delete.' As the environment consists of a 2D view, the size of the gird can be altered and the cells themselves. In NetLogo turtles are mobile agents, patches are stationary agents, and links are connecting agents. A primary difference between patches and turtles is that patches cannot move. Individuals are called agents and in NetLogo they are called turtles which can be inspected throughout the simulation by right clicking on the agent and selecting inspect. They are able to move around the simulated world and their shape, size and colour can be customized. The environment relates to where the agents live, interactions refer to the evolution of the dynamics of the model and the observer/interface is where the modeller can control and extract the data. The schedule tells the model what to do and when to carry the out the run (Wilensky and Rand, 2015). Turtles in this study are potential consumers and innovators. Patches are agents that are stationary and is represented by a blue square in the centre of the world. The observer agent deploys instructions to agents, for example if no turtles are created, putting code into the observer can create new turtles. The same procedure can also be applied to the patch which is represented as the seller agent. A code can be entered when creating a button, for example, to run the simulation once, a 'go' button can be created. This 'go' button can be pressed again to stop the simulation. A 'forever' box can be checked for the simulation to be continuously run until all agents have adopted. As one of the assumptions of the Bass model (Bass, 1969) is all agents eventually adopt, this feature is necessary to be used in this study. It is typically constructed by having a 'go' and 'setup' button where the 'setup' button initializes the simulation and clears the display in order to start running the simulation. The setup button in NetLogo builds the foundation in order for the model to run which creates the agents and initializes the environment. The setup button also clears away data related to the previous run of the model. For example, if parameter values have changed, the setup button needs to be pressed in order to incorporate these new changes such as updating agents actions, or if there have been changes to the environment. Moreover, it is important to note that as the simulations runs are stochastic in nature, replications are needed to get a more represented overview of the data. After the modeller has pressed the setup button, the go button runs the model. The model is measured in ticks or time units. Monitor buttons can be used to keep track of parameter values while the simulation is running. A code such as 'count turtles' with a specific state will be able to monitor those agents in that state. For example, monitoring agents that have adopted the product. Plots have plot pens which can produce multiple graph outputs on the same set of axes. Different colours can be selected for the plot pens for differentiation purposes followed by commands. The code section of the platform tells the agents what actions to undertake, the documentation has a description of the code, and the interface tab allows the modeller to control and monitor the simulation model.

The agent-based model was developed in this chapter to explore and simulate the interactions between the following entities: potential consumers, an Amazon seller, and an innovator. Table 6 shows a description of variables used in this experiment. 'no.CA' represents the number of consumer agents or potential adopters. 'no.SA' represents number of seller agents where for the purposes of this research is the Amazon UK seller. The 'coefficient-of-innovation' parameter is based on the Bass model (Bass, 1969) and relates to consumer agents who adopt the product on their own without influence from others. The 'social-influence' parameter is also based on the Bass model (Bass, 1969) which represents the coefficient of imitation where consumer agents adopt based on influence from others.

These parameters were chosen in order to align with the objectives of this study. As this study investigates online consumer reviews, 'no.CA' refers to the consumers who posted the online reviews for the trainers products. As the online consumer reviews were derived from the Amazon website, 'no.SA' refers to the Amazon seller who is selling the trainers product. As discussed in Chapter 2, the remaining parameters were derived from conducting the literature review. 'no.innovators' refers to the individual who has already adopted the product and is ready to spread the message of adoption at initialization. 'coefficient-of-innovation' and 'social-influence' have been derived from the Bass model (Bass, 1969) to determine how innovation and imitation affects the rate of adoption. These parameters can be adjusted in order to perform parameter sweeping which has been described as an important step in the simulation methodology which can also be referred to in Chapter 2.

Table 7 refers to the model key implemented for the creation of an agent-based model for the purposes of applying it in this chapter. The blue patch is cantered in the middle of the grid of the NetLogo environment and is symbolized as the Amazon UK seller agent. Next, the innovator agent who has adopted and initially spreads the message of adoption to potential adopters is illustrated as a yellow opaque circle in the simulation. Thirdly, white hollow circles are potential adopters or consumer agents that move randomly across the environment. Once a consumer agent receives the message, they change their colour and state from white hollow circles to a red hollow circle. Lastly, consumer agents visually turn a red opaque circle once they start the decision-making process of whether to adopt.

Table 6: Description of parameters

Parameters	Description
no.CA	Potential adopters
no.SA	Amazon seller agent
no. innovators	Innovator agent who has already adopted the product
coefficient-of-innovation	Based on Bass model (Bass, 1969) diffusion of innovation model, consumers who adopt the product on their own initiative without being influenced by others
social-influence	Also based on the Bass model (Bass, 1969) and is more commonly called coefficient of imitation where consumers adopt the product because others have adopted the product through social influence

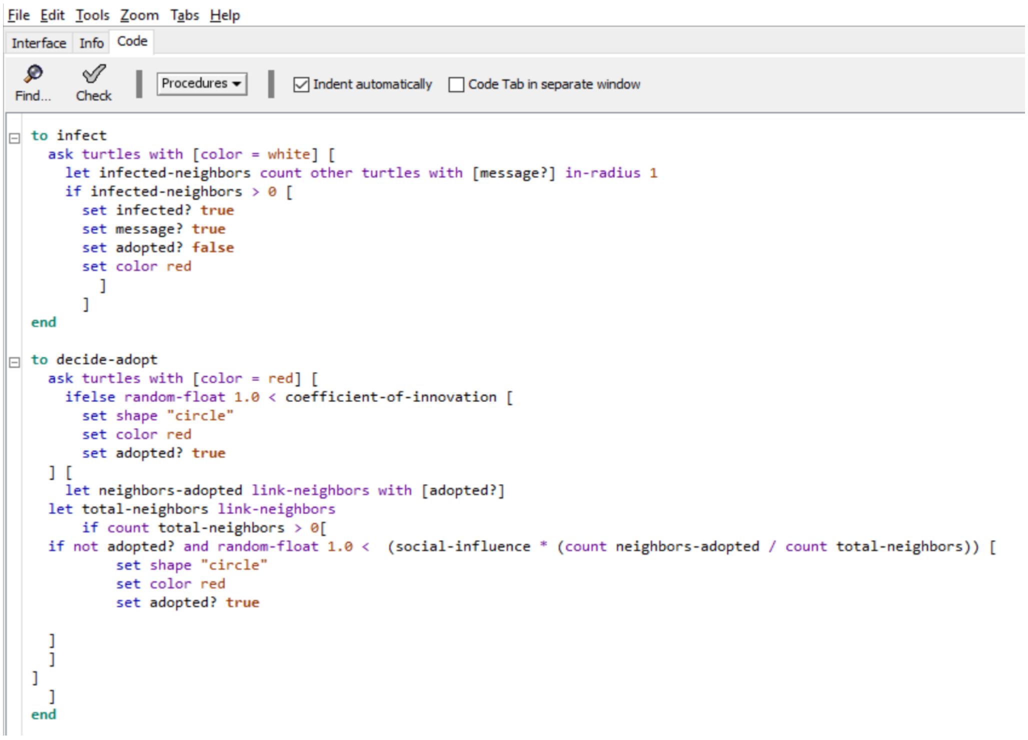
Table 7: Model key

NetLogo components	Description
Blue patch	Amazon UK seller agent
Yellow opaque circle	Innovator agent
White hollow circles	Potential adopters
Red hollow circle	Message received
Red opaque circle	Decision-making

5.5 Implementation

The code tab on the NetLogo interface is where coding for the simulation takes place. The NetLogo platform uses procedures and reporters to combine larger codes and built-in NetLogo commands. The code begins with a procedure name for example, Figure 29, shows 'to infect' as the procedure and the coding in the next few lines are a set of commands. To conclude the piece of code 'end' is entered. Although indenting code is not necessary in NetLogo, it can make the code clearer and easier to read by having a code structure. A reporter is a procedure beginning with 'to-report' that returns a value and includes referencing to another reporter or a variable that returns some value. For example, as shown in Figure 29, 'to infect' is followed by multiple lines of code which includes true and false conditions. It shows a piece of code called 'to infect' in the NetLogo platform which refers to how messages spread to agents. The second line of code represents asking potential consumers represented by the colour white. If this agent receives the message, then they will spread the message onto their neighbours. Once they have received the message they change to a red colour. The attributes the agents now possess are they are infected (an agent has passed on the message to them) represented by the code 'set infected? true,' they have the message represented by the code, 'set message? true,' and the consumer agents haven't yet adopted, represented by the code, 'set adopted? false,' As a result, the consumer agents have changed from a white colour to a red colour representing that they have received the message.

The next block of code, 'to decide-adopt' refers to the adoption process which influences the next stage of the decision-making process of whether to actually purchase or adopt the product. The code illustrates those agents with the colour red, which previously represents agents that have the message, now move on to the adoption threshold process. Firstly, an agent decides to adopt if a random number is less that the 'coefficient-of-innovation' variable albeit not considering social influence. However, if the random number is not less that the 'coefficient-of-innovation,' then the agent takes into account social influence by checking to see if a random number is less than the 'coefficient-of-innovation' variable which in this case, the agent adopts the product regardless of social influence. If the random number is not less than the 'coefficient-of-innovation' variable, then the code checks if there are any neighbours in close proximity to the agent and if there are, the agent adopts, (illustrated by the code 'set adopted? true') if the random number is less than the 'social-influence' variable times the fraction of the neighbours who have adopted. This is based on the Bass model (Bass, 1969), which has been prevalent in the marketing literature as discussed in Chapter 2 (Bass, 2004). In addition, the ‘social-influence’ parameter refers to the coefficient of imitation as presented by Bass (1969). As mentioned in Chapter 2, the coefficient of imitation are consumer agents who are affected by internal influences such as WOM and according to Yaveroglu and Donthu (2002), social influence is a major factor in the diffusion of innovation. This parameter is crucial for capturing the dynamics of consumer agents decisions and actions as it is affected by social interactions and the influence of consumer agents by its immediate neighbours. Consumer agents who have already received the message, will pass on the message to other consumer agents they come into contact with. This can be seen in Figure 29 under the ‘to infect’ code. The strength of influence can be observed in Table 8 and illustrates the measure of influence it has on another consumer agents.




```
to infect
  ask turtles with [color = white] [
    let infected-neighbors count other turtles with [message?] in-radius 1
    if infected-neighbors > 0 [
      set infected? true
      set message? true
      set adopted? false
      set color red
        ]
      ]
end

to decide-adopt
  ask turtles with [color = red] [
    ifelse random-float 1.0 < coefficient-of-innovation [
      set shape "circle"
      set color red
      set adopted? true
  ] [
    let neighbors-adopted link-neighbors with [adopted?]
  let total-neighbors link-neighbors
      if count total-neighbors > 0[
  if not adopted? and random-float 1.0 <  (social-influence * (count neighbors-adopted / count total-neighbors)) [
          set shape "circle"
          set color red
          set adopted? true

  ]
  ]
]
  ]
end
```

Figure 29: Code for receiving the message of adoption in NetLogo

5.6 Experimentation

The user interface which can be seen in Figure 30 displays the ‘setup’ button to run the procedure of the NetLogo commands which have been defined on the ‘Code’ tab of NetLogo and can be seen in Figure 29, ‘step’ button to run the simulation step-by-step and the ‘go’ button which continually runs the steps in the simulation until all conditions are met and thus, the simulation ends. The parameters on the user interface consist of 'no.CA' which refers to the number of consumer agents in the simulation, 'no.SA' which corresponds to the Amazon seller selling the trainers product and 'no. innovators' which is the number of innovators who has already adopted the product at the start of the simulation and possesses the message of adoption, ready to spread to potential adopters or consumer agents. The 'coefficient-of-innovation' parameter which is adapted from the Bass model (Bass, 1969), refers to external influences where consumer agents adopt the product on their own without influence from others and the 'social-influence' parameter is also adapted from the Bass model (Bass, 1969) which represents the coefficient of imitation where consumer agents adopt based on internal influences such as WOM. These parameter values can be changed using the slider to adjust to the desired value.

Table 8 shows the list of parameters used in this chapter to create the agent-based model for this Chapter. The number of consumer agents is set at 500 which represents the number of potential adopters in the simulation. Number of seller agents is set at 1 as this research only focusses on an Amazon UK seller. Number of innovators is set at 1 and consists of an agent that will spread the message of adoption at initialization. According to Liu (2002), the average value of coefficient of innovation lies between 0.01 and 0.03 and the average range of social influence of coefficient of imitation is between 0.3 and 0.5. The coefficient of innovation which is represented by p in the Bass model (Bass, 1969) equation is set at 0.03 for this experiment. The social influence, represented by q in the Bass model (Bass, 1969) equation is set at 0.38. The output of the simulation is visually represented by the monitors which are displayed on the right side of the user interface dashboard. The monitors included for the agent-based model consist of 'Message Received,' 'Innovator Counter,' 'Adopted' and 'Potential adopters' which can be seen in Figure 30. The 'Message Received,' counter represents the number of consumers who have received the message, 'Innovator Counter' represents how many innovators are present while the simulation is running, 'Adopted' represents how many consumers have adopted the product and 'Potential adopters' represent how many consumers are yet to adopt the trainers product. There are two graphs; 'Diffusion Curve' and 'Adopted.' The 'Diffusion Curve' graph includes potential adopters represented by the green line, consumers who have received the message represented by the red line, the grey line represents calculating the threshold probability that both personal and neighbour thresholds have been satisfied in order for the agents to carry the message.

At initialisation of the model, the blue patch which represents the Amazon seller agent does not move and is positioned in the centre of the grid where the grid represents the environment where agents move and interact. Potential consumers represented by hollow white circles move around at random and interact with the innovator represented by the yellow opaque circle. The innovator, as defined by Bass (1969), has already adopted the product and initiates to spread the message. Diffusion of innovation is spread through WOM where if a potential consumer (white hollow circle) interacts with the innovator they will receive the message. Once the potential consumer has received the message, they will change into a hollow red circle. Each agent makes the decision of whether to adopt based on the probability adapted from the Bass model (Bass, 1969). To calculate threshold probability, as mentioned above, two types of thresholds are to be calculated. Firstly, each agent has a personal threshold that is assigned as random during setup. Secondly, neighbour threshold refers to agents that are influenced by their social network. If both thresholds are satisfied, the agents turn into an opaque red colour and move to the next sequential process in the model which is interaction with the seller agent. The monitors which display the output of the simulation model shows a snapshot at a moment in time which refers to ticks. The snapshot of the agent-based model in Figure 30 shows that at tick 134, all 500 agents have received the message of adoption. This is due to the innovator agent initially spreading the message to its neighbours and in turn these consumer agents dispersing the message they possess to their neighbours which can be seen on the 'Message Received' counter. Next, the 'Innovator Counter' is 1 which represents the innovator agent who spreads the message at the start of the simulation. The 'Adopted' counter is 0 and the 'Adopted' graph shows no graphical output of the simulation as no consumer agents have adopted the product. This is because, the agent-based model created in this Chapter shows the NetLogo interface and simulation run **without** sentiment and price variables. Currently, there is no decision-making threshold for agents to justify that consumer agents have adopted the trainers product. This is visible as the simulation only goes as far as meeting the threshold probability represented by the red opaque circles which denote that consumer agents have received the message; but the lack of defining an adoption threshold means agents can not adopt the trainers product at this point in time which would be seen as a change in the colour of consumer agents representing a change in the consumer agents attributes. This is because there are no sentiment and price variables incorporated into the model, hence, there is no adoption threshold stated and therefore, no adoption of the trainers product can take place. The 'Diffusion Curve' graph shows the red line increasing which reflects the number of consumer agents who have received the message, the green line decreases as there are a smaller number of potential adopters left to receive the message as time progresses and the grey line also decreases, visually showing the number of agents who are left to receive the message through social influence.

Figure 30 contributes to the understanding of this research by creating a foundational agent-based model using the simulation methodology as described in Chapter 2. The conceptualization stage shows how to visually represent the different entities in order to distinctly identify the different agents and their corresponding attributes while the simulation is running. The implementation stage denotes the code used to create the agent-based model which includes agent characteristics and how the spread of messages is undertaken using a threshold probability as reported in the literature. The experimentation step details the parameter values to be used for the agent-based model which has been derived from the Bass model (Bass, 1969) through conducting the literature review in Chapter 2. It is evident that no decision to purchase the product had been made as the decision-making process is interrupted. The majority of agents represented by red opaque circles, have received the message, however, when they have approached the seller agent, represented by the blue square in the middle of the environment space, none of the consumer agents change into a different colour which would represent a change in an agents' attribute of adopting the product. As a result, there is no decision-making threshold to influence consumer agents to purchase the product. Another interesting insight is that the innovator agent, which is signified by a yellow opaque circle, will never turn into a different colour to signify that all agents in this simulation have adopted or purchased the product. This contradicts one of the assumptions made by the Bass model (Bass, 1969) which is all agents eventually adopt. Therefore, it is important to add the decision-making threshold consisting of sentiment and price variables to make the significant distinction of agents decision-making process to adopt the trainers product. As Figure 30 displays a snapshot of only messages received by agents, it is an inaccurate representation of reality where adoption is not taking place. Furthermore, as this research has presented sentiment and dynamic pricing in the previous chapter in the form of exploratory data analysis, these variables are significant and need to be incorporated into the agent-based model in order to achieve the aim of this study. Hence, Chapter 6 proposes the SAABM framework which incorporates sentiment and price variables as adoption thresholds for decision-making.

Table 8: Experimental parameters Table

Parameter	Value
no.CA (m)	500
no.SA	1
no. innovators	1
coefficient-of-innovation *(p)*	0.03
social-influence *(q)*	0.38

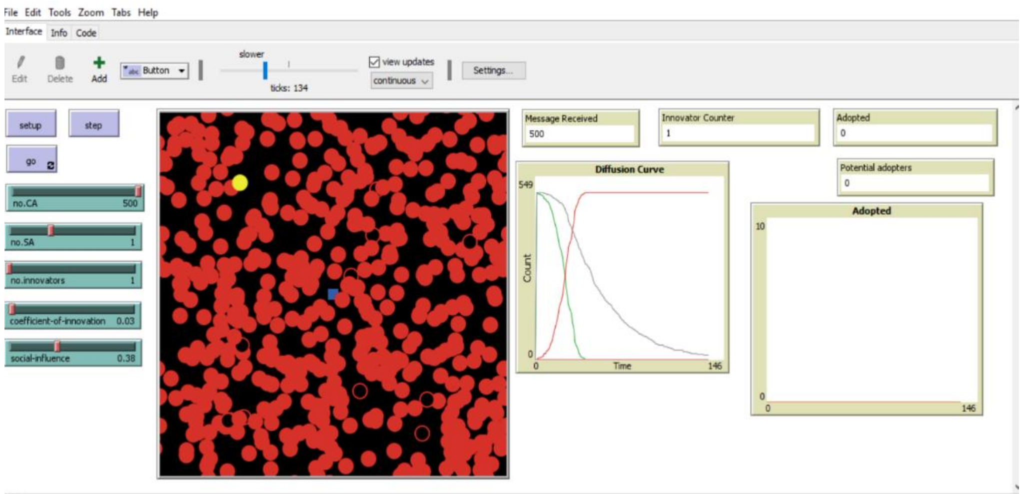


Figure 30: Snapshot of ABM simulation using experimental parameters in NetLogo

5.7 Chapter summary

This Chapter has paved the way for introducing the SAABM framework in the next Chapter by firstly, exploring the mechanics of diffusion of innovation and investigating the social network analysis topology for ABM. By inspecting these concepts, a deeper understanding into creating an agent-based model has been contemplated. The mechanics of diffusion of innovation aids in understanding how the adoption of trainer products are spread through the marketplace environment and allows for predictive insights about diffusion patterns. As social network analysis topology provides insight into the structure and connectivity of social networks, this is important for modelling agent interactions between the consumer agents, innovator agents and the seller agent. It also enables observations of how the innovator agent influences consumer agents in the network. By determining the parameters for this agent-based model through the execution of conducting the literature review in Chapter 2, thus, applying the Bass model (Bass, 1969) parameters, the complexity of modelling agents is achieved. Furthermore, the conceptualization of the model that has been presented in this Chapter builds the foundation for creating the agent-based model for the SAABM framework. This Chapter introduced an agent-based model using the NetLogo platform. As the agent-based model presented in this Chapter did not include sentiment and price variables these two key components will be incorporated in the next Chapter. Observation of the simulation without incorporating sentiment and price variables results in no agents adopting the trainers products. The reason for this is, there is no decision-making process which integrates sentiment and price variables as well as no decision-making threshold that encourages agents to decide whether they want to adopt or purchase the product. The simulation only shows agents receiving the message which is represented by opaque red circles and no adoption decision initialized. This is in contradiction of Bass model (Bass, 1969) where one of the assumptions is that all consumer agents eventually adopt. Therefore, it is of significance to include sentiment and price variables for the development of the SAABM framework to observe the rate of adoption through interactions between consumer agents, innovator agents and a seller agent in a dynamic marketplace environment. As the aim of this study is to develop a framework for combining sentiment analysis and ABM for understanding consumer behaviour in a dynamic marketplace, it is imperative to include sentiment and dynamic pricing for ABM. Thus, the SAABM framework is introduced in the next Chapter which encompasses these two components. The NetLogo code for simulating the agent-based model in this Chapter will be used as the foundation for adding additional code to incorporate sentiment and price for the decision-making process and ultimately the adoption of trainers products in the next Chapter.

Chapter 6

SAABM framework

6.1 Introduction

The research topic of this study is to implement a combination approach of sentiment analysis and ABM to an online consumer review dataset. From Chapter 4, a sentiment analysis model was developed by incorporating time-series to determine whether there is a correlation between sentiment and price fluctuations. An agent-based model was created in Chapter 5 which visualized agents receiving the message of adoption, however, no decision-making process was performed due to not including sentiment and price variables which is a factor to adopt the trainers product. This Chapter will therefore create an agent-based model using sentiment and price variables to influence consumers purchasing decision. The agent-based model will create outputs to visualize market size on adoption rate, diffusion of WOM on market size and remaining number of potential adopters. Section 6.2 describes the evaluation with case study and Section 6.3 presents a conceptualization of the SAABM framework. Section 6.4 comments on the importance of carrying out replication for the simulation, Section 6.5 describes the implementation, followed by Section 6.6. which looks at carrying out the experimentation. Section 6.7 presents what-if scenarios for the SAABM framework which includes investigating scenario’s such as the effect on the simulation model of increasing market size, increasing number of innovators on rate of adoption, observing behaviour for high sentiment and high price, low sentiment and low price and high sentiment and low price. Section 6.8 portrays the output analysis as a result of the agent-based model. Parameter sweeping is performed in Section 6.9 and lastly Section 6.10 highlights the contribution of this chapter and concludes with a Chapter summary.

6.2 Evaluation with case study

As stated in Chapter 1, the aim of the SAABM framework is to investigate consumer behaviour on sentiment and price using a combination approach of sentiment analysis and ABM. Upon reflection of this study so far, Chapter 4 illustrated implementing sentiment analysis as a NLP method to detect how positive, negative, and neutral a review is, and correlation analysis was implemented to determine whether there was a correlation between sentiment and price for each product. Nevertheless, as the present research study is to investigate and observe these variables to better understand consumer behaviour, an agent-based model incorporating these variables into the SAABM framework is proposed in this chapter. As evident from Chapter 5, there were noticeable limitations to the agent-based model produced. Firstly, there was no decision-making process due to no adoption threshold being defined. Hence, no agents could update or change their internal state and continued to stay a red opaque colour signifying staying in the pre-decision-making state. This is contradictory to the Bass model (Bass, 1969) as one of the assumptions states that all consumer agents eventually adopt the trainers product. Sentiment and price variables are necessary to be implemented into the agent-based model in order to progress with the simulation, to observe the interactions between consumer and seller agents in a dynamic marketplace environment through the process of diffusion of innovation and to advance the decision-making stage to the final stage of adoption of trainers products.

6.3 Conceptualization of the SAABM framework

Figure 31 shows the process diagram with the extended SAABM framework and output analysis phases being implemented after the ABM stage that was created in Chapter 5.

- **SAABM framework**

Using an adaptation of the CRISP-DM methodology, after the correlation analysis stage, the SAABM framework aims to incorporate sentiment and price variables. In addition, Chapter 5 showed the generation of ABM **without** sentiment and price variables. The output analysis signified no decision-making process was initialized due to no decision-making threshold being specified. Therefore, to observe the interaction between sentiment and price the SAABM framework is proposed. This framework aims to achieve Objective 4 as previously mentioned in Chapter 1 as a combination approach of sentiment analysis and ABM. The connection of sentiment analysis and ABM is illustrated in Figure 32 linking sentiment analysis and ABM to the SAABM framework. Achieving these objectives should lead to a coherent SAABM framework that is proposed in this research, which aims to combine sentiment analysis and ABM to simulate and observe behaviour on sentiment and price.

- **Output analysis**

The next stage in the process diagram, as shown in Figure 31 is output analysis. Output analysis involves visualizations of the simulation. The agent-based model for the SAABM framework will create outputs using parameter sweeping to visualize market size on adoption rate, diffusion of WOM on market size and the remaining number of potential adopters to adopt the trainers product. What-if scenarios will be undertaken to investigate the effect of increasing market size, increasing number of innovators on rate of adoption, observing behaviour for high sentiment and high price and low sentiment and low price. This can also be adjusted to explore high price, low sentiment and low price and high sentiment. Output analysis can be seen in Section 6.8.

ABM is commonly documented ODD format providing a generic documentation to explain processes and interactions (Alvarez, Brida and London, 2021). Refer to Appendix 70 for the ODD documentation explaining the agent-based model for the SAABM framework created in this study for further detail. The agent-based model follows the basic principles of the Bass model (Bass, 1969), where the model illustrates how new products get adopted in a population and how innovators and potential adopters interact (Charan, 2015). The key adaptive behaviour of agents in the model is the decision to adopt. This is based on personal, neighbour, sentiment and price adoption thresholds being satisfied. Figure 33 shows the conceptualization diagram for the simulation process for the SAABM framework. At initialization, consumer agents interact with an innovator agent who possesses an attribute of carrying a message of sentiment and price they experienced when they adopted the product. Next, if the consumer agent has received the message from the innovator and have satisfied the threshold probability for personal and neighbour threshold as mentioned in Chapter 5, the consumer agents proceed to interact with the seller agent. Calculating the threshold probability is also evident in the literature as mentioned by Čavoški and Marković (2016) which can be referred to in Chapter 2. If neither threshold is satisfied, consumer agents do not proceed to the seller agent. As an extension to the previous model created in Chapter 5, the next stage in the process is after interacting with the seller agent, the adoption probability is calculated from the sentiment and price variables. The consumer agent will adopt the product once the adoption probability has been satisfied. This is represented by the agent turning into an orange smiley, neutral or sad face depending on sentiment and price scores. For the purposes of this research, referring back to Table 4 in Chapter 4, the mean prices for the 3 brands range from £72 to £93, therefore, a price range of £50 to £100 is used for the simulation with £75 referring to neutrality. Prices above are considered as higher prices and prices below this figure are considered a lower price. In addition, the correlation analysis presented a negative correlation for the Nike B00CGVQHLM product, and a positive correlation for Puma B000G529SA and Timberland B008H2CKA products. This simulation assumes as sentiment increases, price also increases and as sentiment decreases, price also decreases indicating positive correlation. The simulation will end once all consumers have adopted the product which is an assumption of the Bass (Bass, 1969) model.

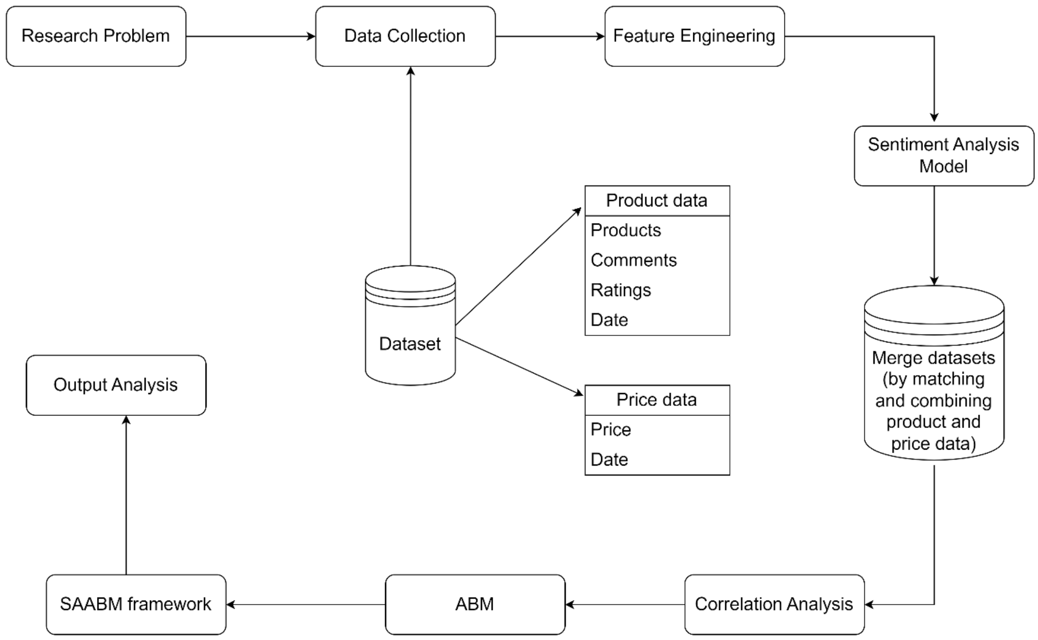


Figure 31: Methodology diagram for developing SAABM framework with output analysis (Adapted from CRISP-DM, 1999)

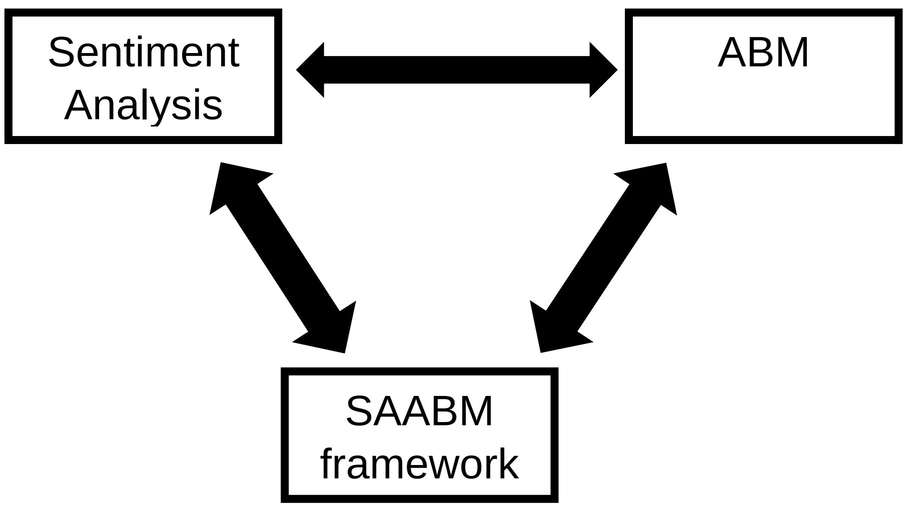


Figure 32: Connection between sentiment analysis and ABM resulting in SAABM framework

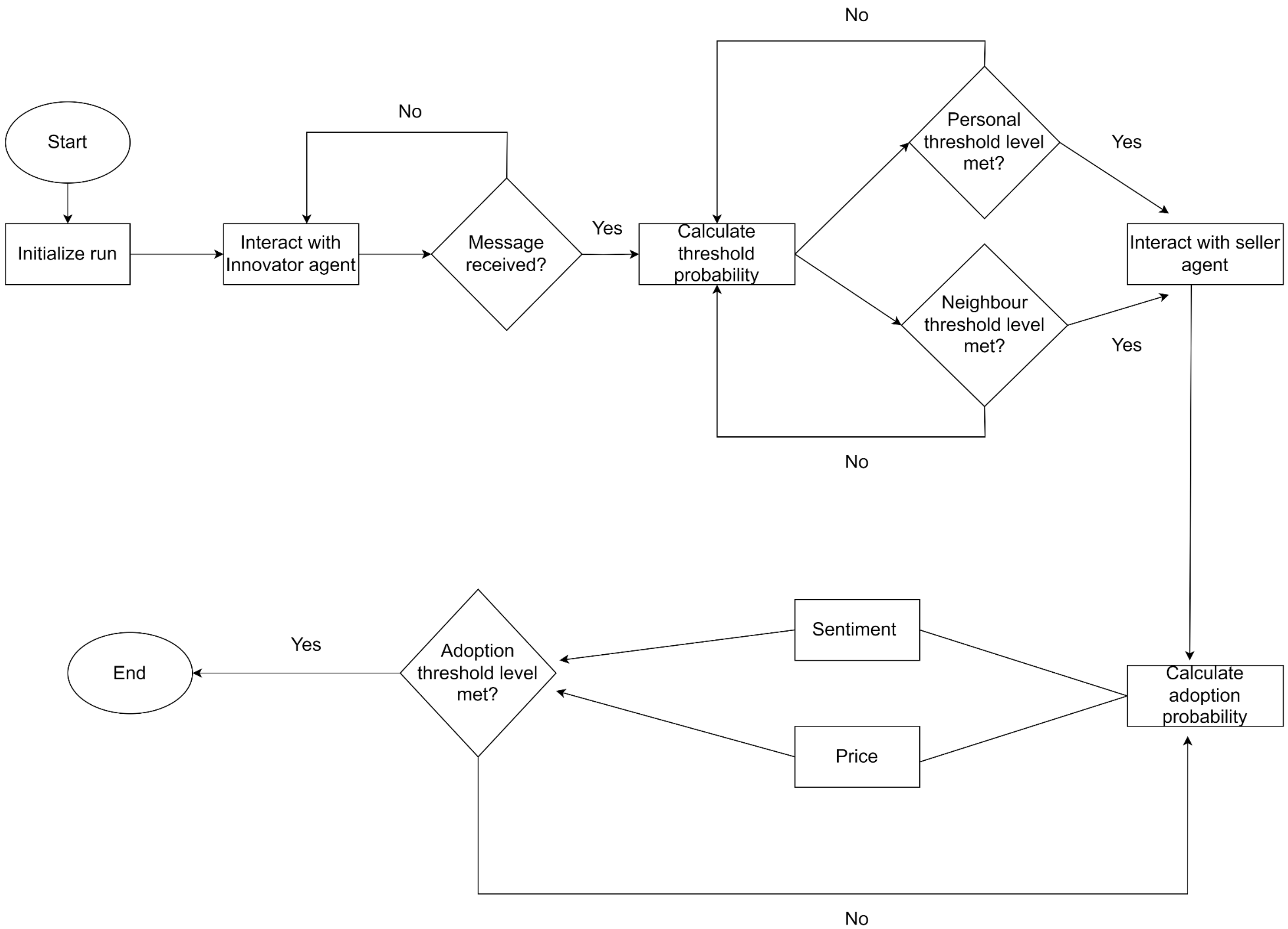


Figure 33: Conceptualisation of adoption processes occurring in sequential order

6.4 Replication

As ABM consists of stochastic components, it is important that the model runs multiple times to characterize the average behaviour of the model (Wilensky and Rand, 2015). Replication across different parameter settings needs to be examined in order to evaluate the data. NetLogo allows the output to be produced as a spreadsheet comprising of a CSV file. In addition, as mentioned in Chapter 5, Schelling's segregation model (Schelling, 1978) repeats the process many times to observe the outcome. The process of replication used in this study was 50 simulation runs to achieve the average values from these runs.

6.5 Implementation

Table 10 describes the parameters used for the SAABM framework model. Similarly, to Table 6 in Chapter 5, there are 2 additions to the Table; 'sentiment' and 'price' are added to the list of parameters. 'no.CA' can have a range of 50, 100 or 500 which is the number of consumer agents in the population. This parameter can be adjusted to examine rate of diffusion on different market sizes. 'no.SA' is set to 1 as explained previously, as this study only considers a UK Amazon seller and does not take into account 3rd party sellers. Next, 'no.innovators' has a range of 1 to 5 which represents the number of innovator agents that have initially adopted the product at the start of the simulation. Similarly, as the 'no.CA' parameter, this can be adjusted within the range to explore the effects of increasing or decreasing the number of innovators on rate of adoption. 'sentiment' can be altered within a range of -1 to +1 where the lower the value, the more negative the sentiment and the higher the value, the more positive the sentiment. A sentiment value of 0 indicates neutral sentiment. 'price' as mentioned earlier in this Section has a range of 50 to 100. The next 2 parameters are adapted from the Bass model (Bass, 1969). 'coefficient-of-innovation' which has a range of 0 to 1 and refers to external influence. 'social-influence' also has a range of 0 to 1 and refers to internal influence such as WOM. In addition, these parameters can be adjusted to explore what-if scenarios to predict adoption rate which can be observed in Section 6.7.

Comparably to Table 7 in Chapter 5, there is an addition to the model key. Table 9 defines that orange opaque circles indicate that a consumer agent has adopted the trainers product. Depending on the value of sentiment and price, the orange circle will express a happy, smiley face for positive sentiment, sad face for negative sentiment and neutral face (neither happy nor sad) portraying neutral sentiment. This decision-making process is not standalone to the sentiment variable as the adoption state also depends on how high or low the price is as stated by the SAABM framework.

Table 9: Model key for SAABM framework

NetLogo components	Description
Blue patch	Amazon seller agent
Yellow opaque circle	Innovator agent
White hollow circles	Potential adopters
Red hollow circle	Message received
Red opaque circle	Decision-making
Orange opaque circle – smiley, sad or neutral face	Agent has adopted the product and the face denotes either positive, negative or neutral sentiment

Table 10: Description of parameters for SAABM framework

Parameters	Value	Description
no.CA (m)	50, 100, 500	Number of consumer agents
no.SA	1	Number of seller agent. As this study only looks at Amazon retailers and not 3rd party retailers this is set to 1.
no. innovators	1 - 5	Number of innovators who have already adopted the product and spreads the message to consumer agents
sentiment	-1 to +1	Slider to adjust positive, neutral or negative sentiment. Sentiment score, -1 to -0.1 is negative, 0 is neutral and 0.1 to 1 is indicated as positive sentiment
price	50 - 100	Price value can be changed via the slider on the Netlogo interface
coefficient-of-innovation (p)	0 - 1	External influence. Adapted from Bass model (Bass, 1969)
social-influence (q)	0 - 1	Internal/WOM (neighbour threshold). Also referred to coefficient of imitation where the message is spread via WOM. Adapted from Bass model (Bass, 1969)

Figure 34 refers to the code where sentiment and price variables are incorporated into the SAABM framework. This is an important part of the decision-making process where data gathered from the sentiment analysis model from Chapter 4 are used for the SAABM framework model. It shows the code and the rules the agents follow. The piece of code ‘to decision-make’ involves consumers who have received the message and as previously stated in Chapter 5 are an opaque red colour that approach the seller agent which is represented by the blue square in the centre of the environment. If sentiment is less than 0, negative sentiment is applied denoting a sad face. In contrast, a sentiment of higher than 0 displays positive sentiment and is represented as a smiley face. Sentiment equal to 0 ascribes to a face that is neutral; neither happy nor sad. As this research is also interested in the price variable, to follow the positively correlated data between sentiment and price variables, if sentiment is higher than 0 and the price is higher than 75, then the shape of the agent changes to a smiley face. In addition, if sentiment is lower than 0 and the price is lower than 75 then the shape reflects a sad face. Lastly, if the sentiment is 0 and the price is equal to 75 the shape represents a neutral face. Moreover, the ‘adopted’ attribute is set to ‘true’ as these agents now possess these two attributes. To differentiate between decision-making and adoption, agents turn into an orange colour when they have adopted the product, providing they satisfy the adoption threshold. The last part of the code is with regards to the processes of how the simulation runs. If all agents adopt, the simulation stops and the simulation is complete. ‘to step’, refers to the processes running in order. A ‘tick’ refers to when a single run has taken place. In this context, a tick represents a day. ‘go’ refers to continually running the simulation until the conditions of adoption have been completed.

```
to decision-make
  ask patches with [pcolor = blue][
    ask turtles-here with [shape = "circle"] [                    ;only consumers who have received the message can decision-make
            if [sentiment] of patch-here > 0 and [price] of patch-here > 75 [set shape "face happy"]
            if [sentiment] of patch-here < 0 and [price] of patch-here < 75 [set shape "face sad"]
            if [sentiment] of patch-here = 0 and [price] of patch-here = 75 [set shape "face neutral"]
      set adopted? true
      set color orange
    ]
  ]
end

to step
  if all? turtles [ adopted? ] [ stop ]
ask turtles [
   rt (random 360)
    fd 1
      ]
  infect
  decide-adopt
  decision-make
 tick
end

to go
  step
end
```

Figure 34: Code for adopting the product in NetLogo

6.6 Experimentation

The 3 what-if scenarios captured by the agent-based model developed in NetLogo are high sentiment and high price, low sentiment and low price and neutral sentiment (sentiment of 0) and neutral price (£75). Figures 35, 36 and 37 shows a snapshot of the NetLogo interface while the simulation is running in each of the three different scenarios. On the left of the NetLogo interface, are sliders which the user can adjust. For example, ‘no. CA’ which is number of consumer agents is represented by hollow white circles, ‘no. SA’ which is the seller agent represented by the blue square patch centred in the middle of the environment, ‘no. innovators’ is represented by the yellow opaque circle where the innovator agent spreads the message to potential adopters. Sentiment and price variables can also be adjusted. As mentioned earlier, the ‘coefficient-of-innovation’ and ‘social-influence’ is derived from the Bass model (Bass, 1969). When the ‘go’ button has been initiated, the agents are moving in a random state in the marketplace environment. To run the simulation, desired measurements are inputted by adjusting the sliders for each of the input parameters. The ‘setup’ button sets the simulation to run based on the modellers input selection. The ‘go’ button is a *forever* button which will run until all conditions are satisfied resulting in completion of the simulation. In this case, the simulation will stop once all consumers have turned an orange colour representing all consumer agents including the innovator agents have adopted the product. The innovator also changes to an orange colour due the assumption of the Bass model (Bass, 1969). The ‘step’ button allows the user to run the simulation and see interactions step-by-step. The speed of the model can be adjusted to view the simulation running slower or faster by dragging the slider for speed. The simulation model is measured in ticks which is a measure of time in the simulation model representing one day. Towards the top right of the NetLogo interface, there are four counters, namely, ‘Innovator Counter,’ ‘No. potential adopters,’ ‘Message Received’ and ‘Adopted’ which are useful for updating and monitoring the change in the numbers while the simulation is running. The ‘Innovator Counter’ monitors the number of innovators in the simulation. The innovator agents will eventually change into an orange colour due to all agents adopting. As the number of potential adopters who have received the message, change their internal state from a potential adopter to decision-making state or symbolized in the Netlogo platform as white hollow circles to a red hollow or opaque circle, the counter number for ‘No. potential adopters' starts to decrease. As this counter decreases, the 'Message Received' counter increases as there is a rise in the number of agents that have been infected with the message and now carry this attribute. Lastly, the 'Adopted' counter increases as the simulation progresses over time as all agents adopt the trainers product.

Below the counters are two graphs which are beneficial to visually observe what is happening during the simulation. The 'Diffusion Curve' graph shows the number of potential adopters decreasing over time, represented by the green line. The grey line also shows a decrease which represents how many agents have changed from an initial state of potential adopters to being infected with agents that possess the message through social influence. As the number of potential adopters decrease, the number of agents who have received the message increases shown by the red line. Lastly, adoption is represented by the orange line on the graphs. On the right of the 'Diffusion Curve' graph, is the 'Adopted' graph which specifically monitors the adoption rate. This is beneficial to this study as it illustrates diffusion of innovation, hence rate of adoption.

6.7 What-if scenarios for the SAABM framework

The parameter values selected for the what-if scenarios for carrying out the SAABM framework are presented in Tables 11, 12 and 13.

Table 11 presents the values for simulating high sentiment and high price. 'no.CA' is set to 100, 'no.SA' is listed as 1 as it refers to an Amazon UK seller. 'no.innovators' is also set to 1. 'sentiment' is positive which is listed as 1, 'price' is also set to high as £100, 'coefficient-of-innovation' is set to 0.03 and 'social-influence' is set to 0.38. Figure 35 shows a snapshot of the what-if scenario of high sentiment and high price. As the simulation initializes the consumer agents are moving at random across the environment. The hollow red circles indicate that they have been infected with the message that has been passed onto them by an agent in close proximity to them that has also received the message. The opaque red circles indicate that the threshold probability has been satisfied and now the consumer agents are in the decision-making process. Once they come across the blue patch which is referred to as the seller agent in the centre of the grid, they start calculating the adoption probability and once this is satisfied through inspecting the sentiment and price values, the consumer agents proceed to adopt the trainers product. As the threshold is met for adoption the consumer agents turn into an orange smiley face indicating both message and adoption probabilities are met and the visualization results in high sentiment and high price. The counters can be seen to the top right of the NetLogo interface. The snapshot was taken at 101 ticks or number of steps where the number of potential adopters referred to as the ‘No. potential adopters’ counter is 41, the number of consumer agents that have received the message is 56 represented by the ‘Message Received’ counter and 4 consumer agents have adopted the Puma trainers product so far, highlighted by the ‘Adopted’ counter. These counters can also be visualized graphically. On the ‘Diffusion Curve’ graph, as the red line indicating message received increases, the number of potential adopters decrease represented by the green line. This is due to more consumer agents becoming infected with receiving the message and are no longer potential adopters as they have moved on to meeting the adoption threshold process. The orange line represents adoption which steadily increases. This is shows clearer in the ‘Adoption’ graph. Lastly, the grey line represents the agents are calculating the adoption probability.

Table 12 lists low sentiment and low price. There are 2 changes to the parameters than the previous table, Table 11. As this scenario is investigating low sentiment and low price, ‘sentiment’ is set at -0.5 indicating negative sentiment and ‘price’ is set at £60 which is at the lower spectrum of the price range. All the other parameters are kept the same. Figure 36 illustrates, low sentiment and low price. The snapshot was also taken at 101 ticks, and it is evident that negative sentiment and negative price is taking place symbolized by the sad orange faces. At the time of the snapshot, the innovator has yet to cross the blue patch and will eventually turn into an orange colour. Most of the consumer agents have progressed to calculating the threshold for receiving the message and there are 2 consumer agents left to infect. This can also be seen on the ‘No. potential adopters’ counter. 94 consumer agents have received the message and 4 consumer agents have adopted the trainers product so far. Similarly, to Figure 35, there is a gradual increase of adoption, which can be seen in the ‘Adoption’ graph. On the ‘Diffusion Curve’ graph, as the red line increases which is consumer agents who have received the message, the number of potential adopters decrease, represented by the green line. The grey line also decreases as consumer agents are ready to move onto the next stage of adopting the product.

Table 13 represents neutral sentiment and neutral price. As the previous Table, all the parameters are kept the same except for sentiment and price.

‘sentiment’ for this scenario is set at 0 indicating neutral sentiment and price is set at ‘£75’ indicating a neutral price point. Figure 37 shows neutral sentiment and neutral price. The snapshot was taken later than the previous 2 snapshots at 272 ticks and one of the observations is that there are no more potential adopters and no hollow red circles indicating that all consumer agents have the attribute of carrying the message. This is shown by the green and grey lines on the ‘Diffusion Curve’ graph. The number of adopters has also increased by double as 8 consumer agents have adopted compared to 4 in the previous snapshots. In addition, there is much more of an adoption progression line which can be seen on the ‘Adopted’ graph. The ‘Message Received’ counter has a value of 92 and this will continue to decrease over the course of the simulation and change to the ‘adopted’ attribute, consequently increasing the ‘Adopted’ counter. The red line which represents message received, increases and is now gradually decreasing as the number of ‘message received’ consumer agents turn into ‘adopted’ consumer agents.

However, if neither adoption thresholds for sentiment and price are satisfied, the simulation output will be similar to Figure 30 in Chapter 5 where agents will not be able to proceed to the next stage of the simulation process to adopt the product. This is due to exploratory data analysis where positive correlation between the 2 variables of sentiment and price are investigated. This is applicable to the following what-if scenarios mentioned below:

- High sentiment, low price
- Low sentiment, high price
- Neutral sentiment, high price
- Neutral sentiment, low price
- High sentiment, neutral price
- Low sentiment, neutral price

Table 11: Parameters for high sentiment and high price

Parameter	Value
no.CA (m)	100
no.SA	1
no. innovators	1
sentiment	1
price	100
coefficient-of-innovation (p)	0.03
social-influence (q)	0.38

Table 12: Parameters for low sentiment and low price

Parameter	Value
no.CA (m)	100
no.SA	1
no. innovators	1
sentiment	-0.5
price	60
coefficient-of-innovation (p)	0.03
social-influence (q)	0.38

Table 13: Parameters for neutral sentiment and neutral price

Parameter	Value
no.CA (m)	100
no.SA	1
no. innovators	1
sentiment	0
price	75
coefficient-of-innovation (p)	0.03
social-influence (q)	0.38

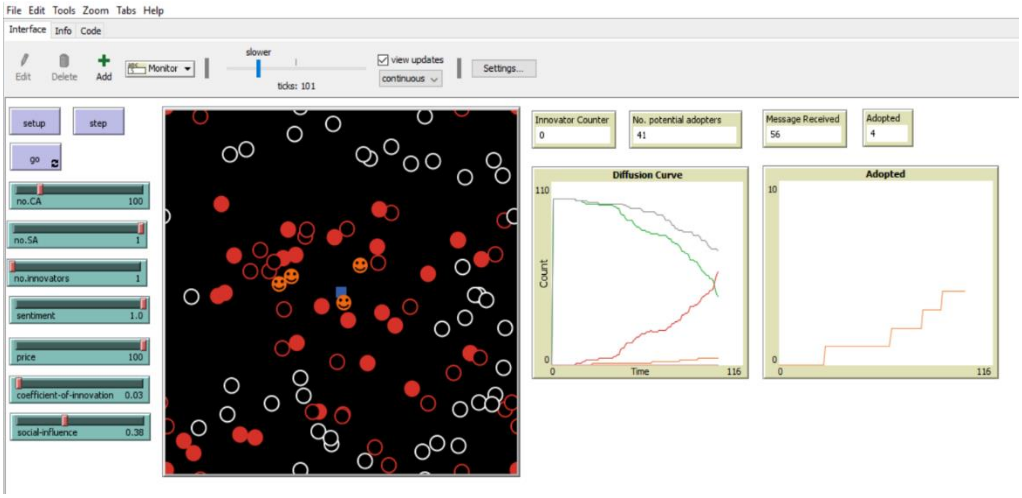


Figure 35: Snapshot of SAABM simulation for high sentiment and high price in NetLogo

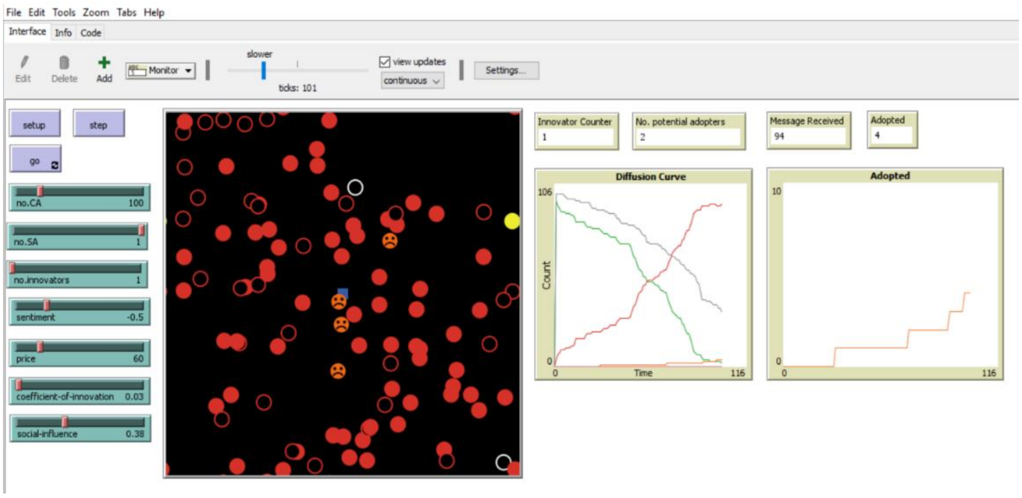


Figure 36: Snapshot of SAABM simulation for low sentiment and low price in NetLogo

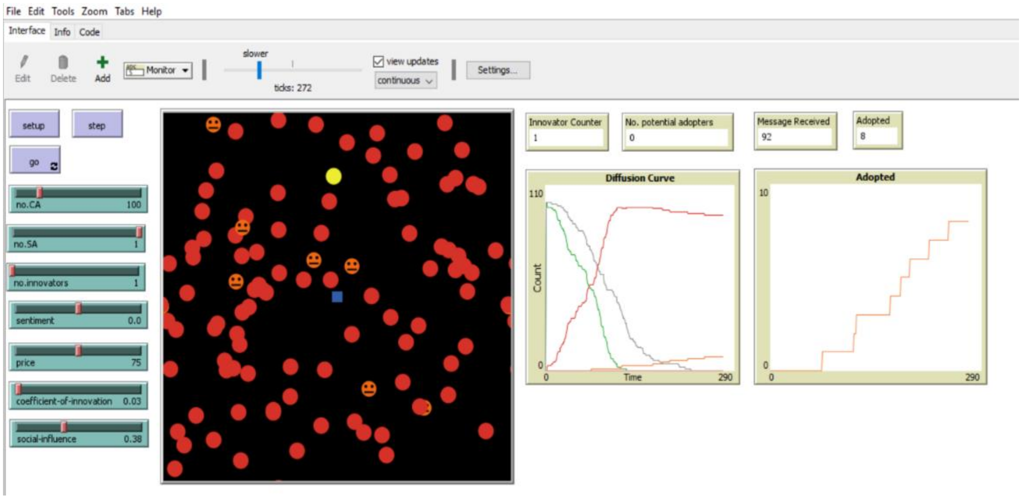


Figure 37: Snapshot of SAABM simulation for neutral sentiment and neutral price in NetLogo

6.8 Output Analysis

As mentioned in Chapter 2, BehaviourSpace is a tool used in NetLogo to conduct experiments. This is particularly useful for parameter sweeping, as multiple experiments can be set up to observe the effect on rate of adoption by adjusting parameter settings. To conduct an experiment, the modeller selects Tools on the menu bar and selects BehaviourSpace. Figure 38 shows the list of experiments with an option to create a new experiment or to edit the current experiment. Pressing the 'New' or 'Edit' button opens the dialogue box presented in Figure 39. The name of the experiment can be entered at the top. The 'Vary variables' section consists of parameters for the experiments which can be set as a list of values. For example, as this study is concerned with rate of adoption on market size, a range of consumer agents, represented by 'no.CA' can be entered as a market size of 50, 100 or 500. This can also be applied to the number of innovators and the coefficient of innovation and social influence. As mentioned in Section 6.4, the number of repetitions used in this study is 50 and each simulation run is processed sequentially as there are a number of processes the agents go through chronologically to achieve adoption. Replication is important as agents move around randomly and therefore no 2 runs will be the same, hence, an average from the 50 runs will be used in this research. The next window box measures each run using the following reporter 'count turtles with [color = orange]' which measures adoption rate by counting the number of consumer agents who have turned into an orange colour referring to their internal state of adopting the product. As this study is interested in visualizing a diffusion curve, the 'Measure runs at every step' is checked. The 'Setup commands' and 'Go commands' are as mentioned in the coding interface of NetLogo where setup is defined as the processes in order and 'Go commands' is to run the simulation. The 'Stop condition' is based on all agents adopting the product. The 'Time limit' is set to 0 as the simulation stops when a condition of all agents adopting has been satisfied. Pressing 'OK' at the bottom of the window saves these settings goes back to the previous window as seen in Figure 38. By clicking 'Run,' a dialog box as shown in Figure 40, allows the modeller to export the data in various formats such as a spreadsheet as a comma-separated values (CSV) file. The simulation will start running when the modeller clicks 'OK.'

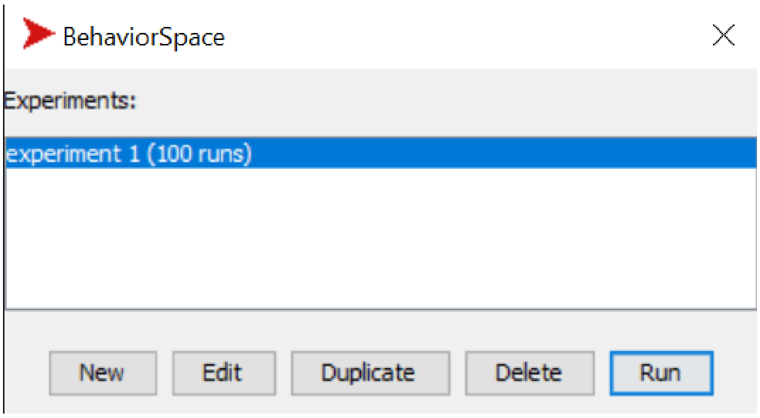


Figure 38: Screenshot of experiments in BehaviourSpace

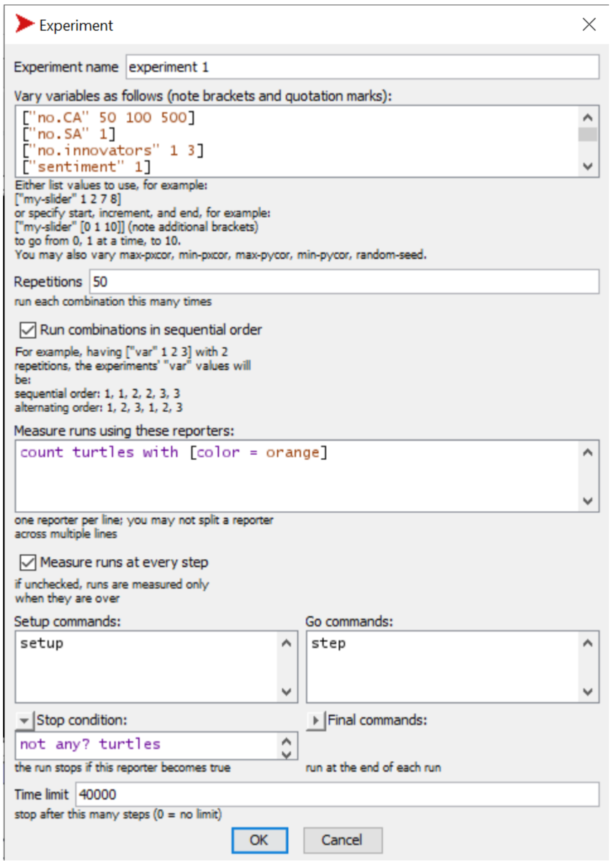


Figure 39: Screenshot of creating experiments in BehaviourSpace

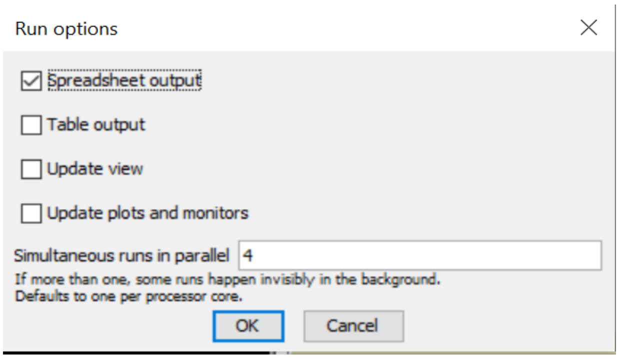


Figure 40: Screenshot of output format of BehaviourSpace experiments

Figure 41 visualizes the output of the simulation run using the experimental parameters shown in Table 14. Table 14 highlights the parameter values used for the Bass model (Bass, 1969) simulation experiment. Using the sliders on the NetLogo interface, as mentioned in Chapter 5, the number of consumer agents can be increased or decreased.

For this experiment, the market size will consist of 50, 100 and 500 consumer agents to visualize the effect of market size on adoption rate. The next parameter is set to 1 for seller agent which represents an Amazon seller agent. The number of innovators can also be adjusted however for this experiment of Figure 41, the number of innovators will be set to 1. The variation in number of innovators can be seen in Figures 45 and 46. This experiment includes high sentiment and high price as the parameter values are '1' for sentiment and '100' for price. The values of 'coefficient-of-innovation *(p)*' and 'social influence *(q)*' are adapted from the Bass model (Bass, 1969) and are set to 0.03 and 0.38 respectively. The graph shows that as market size increases, there is a steeper increase of the number of consumer agents adopting the product over time. This is due to rate of adoption being quicker in a larger population as there are more consumers in close proximity to spread the message to. Another insight into this graph, is the total number of consumer agents adopt quicker in a lower market size of 50 consumer agents where the number of ticks is around 11,900 than the larger market size of 500 consumer agents with the number of ticks around 17,000. This is understandable as there are less agents in a smaller market size to spread the message to and less total number of agents to adopt the trainers product. By observing the number of ticks, more consumer agents adopt the product in a smaller market size than a larger market size.

Table 14: Parameters used for Bass model (Bass, 1969) experiment

Parameter	**Value**
no.CA *(m)*	50, 100, 500
no.SA	1
no.innovators	1, 3
sentiment	1
price	100
coefficient-of-innovation *(p)*	0.03
social-influence *(q)*	0.38

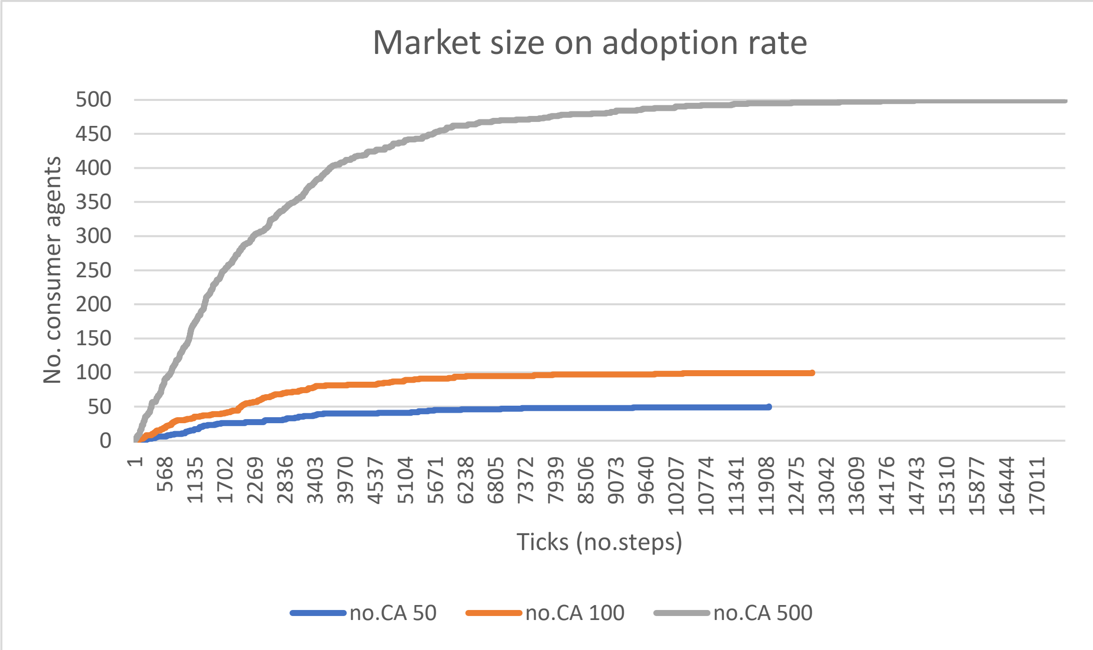


Figure 41: Graph of market size on adoption rate

Figure 42 shows rate of diffusion of WOM on different market sizes using the experimental parameters as listed in Table 14. The results show that in a larger market size, there is a sharp increase in the number of consumer agents who receive the message than in a smaller market size. The initial part of the diffusion curve can be seen clearer in Figure 43. By the 16th tick, there is a steeper incline in a larger market size of 500 consumer agents compared to a market size of 50 and 100. In a market size of 100 consumer agents, the rate of diffusion for WOM starts increasing around 46 ticks. In a smaller market size of 50, there is a slower rate of diffusion WOM as diffusion starts increasing around the 150th tick. A smaller market size shows a more gradual and slower WOM diffusion as there are less consumer agents to infect or spread the message to.

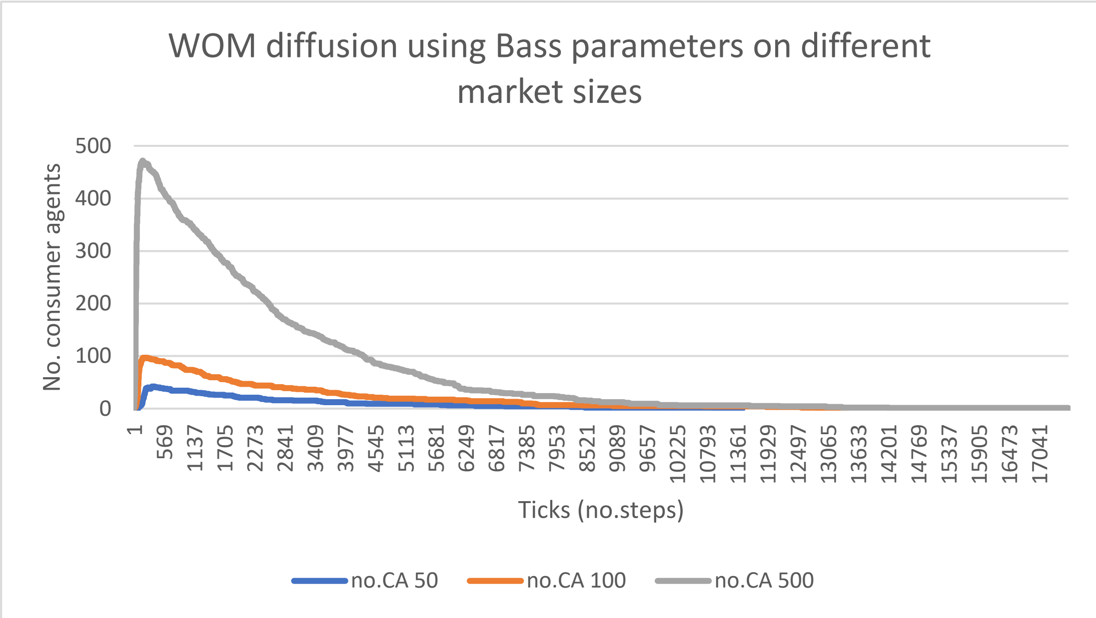


Figure 42: WOM diffusion using Bass parameters on different market sizes

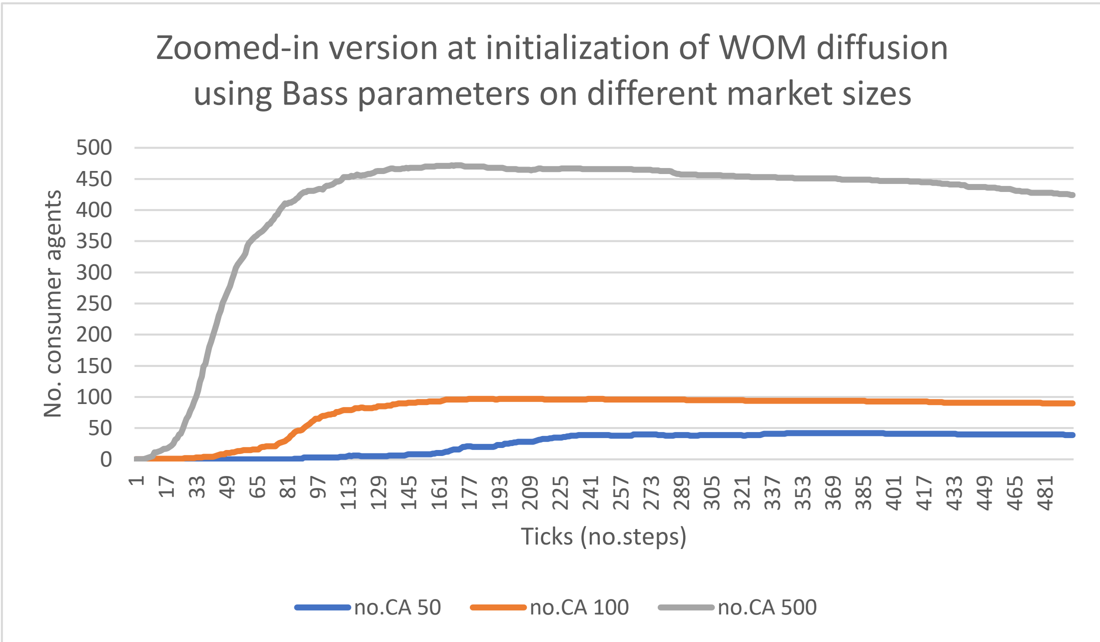


Figure 43: Zoomed-in version at initialization of WOM diffusion using Bass parameters on different market sizes

Table 14 also lists the parameters for conducting a what-if scenario into how changing the number of consumer agents in the environment affects the remaining number of potential adopters. Figure 44 shows the remaining number of potential adopters over time in different market sizes. There is a steeper decline in a larger market size of 500 consumer agents than a market size of 50 or 100. This rapid decrease is due to more agents being in close proximity to other consumer agents in a larger market size in the NetLogo environment. Therefore, as more consumer agents change from their initial potential adopter state to a state where they have received the message, the number of potential adopters decrease quickly. The remaining number of potential adopters that have received the message of adoption has a faster rate of diffusion in a market size of 500 consumer agents which takes around 41 ticks, compared to 121 ticks in a market size of 100 consumer agents and around 169 ticks in a market size of 50 consumer agents.

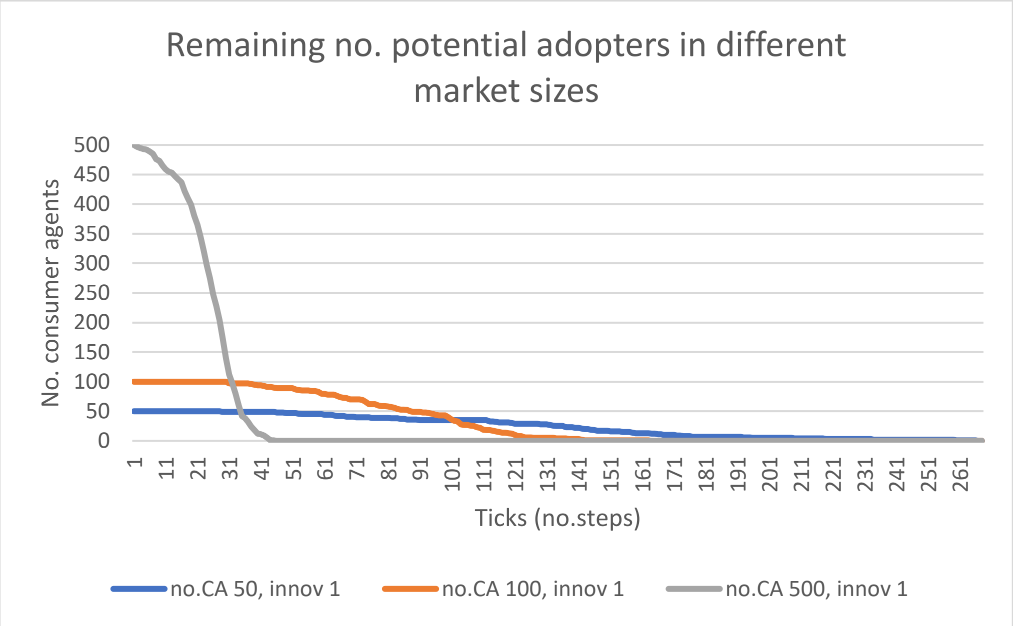


Figure 44: Remaining number of potential adopters in different market sizes

Figure 45 shows remaining number of potential adopters with different number of innovators in a market size of 50 consumer agents. The parameters for this scenario is also listed in Table 14 consisting of values 1 and 3. The rate of diffusion is faster as the number of innovators increase. This is due to 3 innovators that already have the message and spread the message to neighbours resulting in a quicker diffusion of infecting consumer agents with the message. In a market size of 1 innovator, there is a slower rate of diffusion as there is only 1 innovator that can initially spread the message.

Similarly, to Figure 45, Figure 46 illustrates the remaining number of potential adopters in a market size of 100 consumer agents with different number of innovators. This graph also shows that there is a quicker rate of diffusion when there are more innovators to spread the message. The rate of diffusion is also quicker in a larger market size whereby at the 96th tick, the remaining number of potential adopters are very low indicating that all consumer agents have changed from their initial state of potential adopters to consumer agents that have received the message and have satisfied the threshold probability. In comparison, to Figure 45 the remaining number of potential adopters is low at the 150th tick.

The what-if scenarios showed that rate of adoption is faster in a smaller market size compared to a larger market size. Furthermore, by increasing the number of innovators increases the rate of diffusion. The next section will perform parameter sweeping where coefficient of innovation and social influence values can be investigated by adjusting values to observe rate of adoption.

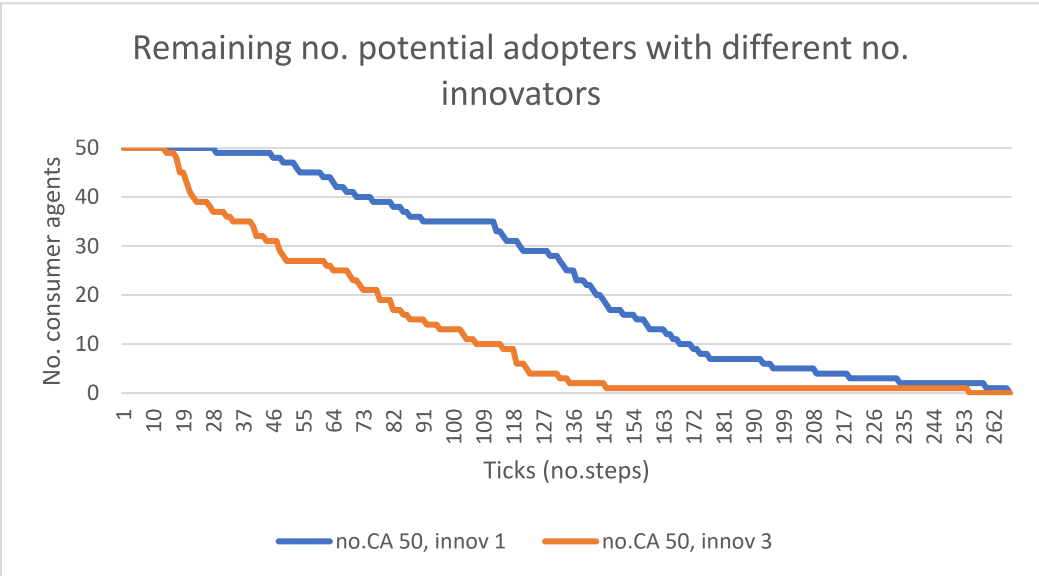


Figure 45: Remaining number of potential adopters in a market size of 50 consumer agents with different number of innovators

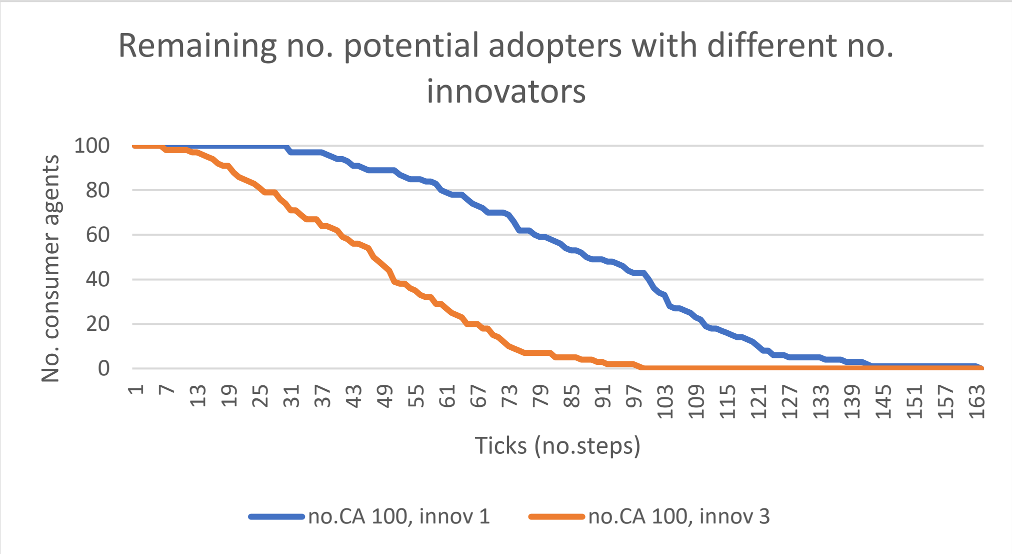


Figure 46: Remaining number of potential adopters in a market size of 100 consumer agents with different number of innovators

6.9 Parameter sweeping

Another function of what-if scenarios includes parameter sweeping. Parameter sweeping refers to examining changes in parameter settings to see how sensitive the model is due to these changes. This is useful to explore the robustness of model behaviour. Table 15 lists the parameters used for parameter sweeping by adjusting the parameters for coefficient of innovation and social influence. The Bass model (Bass, 1969) parameters used for the previous what-if scenarios in Section 6.8 are also included, where the value of coefficient of innovation is 0.03 and social influence is 0.38. As well as incorporating these figures, parameter sweeping will be conducted using values of 0.01 and 0.02 for coefficient of innovation and values of 0.3, 0.4 and 0.5 for social influence to determine the effect on rate of adoption.

Table 15: Parameter sweeping Table

Parameter	Value
no.CA (m)	100
no.SA	1
no.innovators	1
sentiment	1
price	100
coefficient-of-innovation (p)	0.01, 0.02, 0.03
social-influence (q)	0.3, 0.38, 0.4, 0.5

Figure 47 presents a parameter sweeping chart on rate of adoption with varying *p* values (coefficient of innovation) while keeping the value of *q* (social influence), the same. At the start of the simulation, both categories have a similar trend until around 740 ticks where a higher *p* value represented by the orange line has a slightly faster rate of adoption. However, at around the 200th tick, the *p* value of 0.01 represented by the blue line, has a faster rate of adoption than the higher *p* value of 0.03. Nevertheless, the higher *p* value shows a faster rate of adoption at the 7500th tick, resulting in a faster rate of adoption by all 100 consumer agents in comparison to a lower *p* value of all consumer agents adopting by around 12,500 ticks. Conclusively, this chart suggests the higher the *p* value, the faster the rate of adoption.

Figure 48 displays a parameter sweeping chart with differing *p* values while keeping the *q* values the same at 0.4. Similarly, to the parameter sweeping Chart 1, this chart shows that the higher the *p* value, the faster the rate of adoption. At initialization of the simulation, the orange line which has a higher *p* value 0.03 has a higher rate of adoption with 30 consumer agents adopting the trainers product at around the 900th tick, compared to the lower *p* value of 0.01 represented by the blue line where 30 consumer agents have adopted the product around the 1050th tick. There is a higher rate of adoption for the lower *p* value around a time period of around the 2000th tick to around the 4500th tick, and after this, the higher the *q* value of 0.03 overtakes the *p* value of 0.01 resulting in a quicker rate of adoption by around the 11,600th tick compared to a rate of adoption by around the 11,800th tick for the lower *p* value adopted by the market quickly.

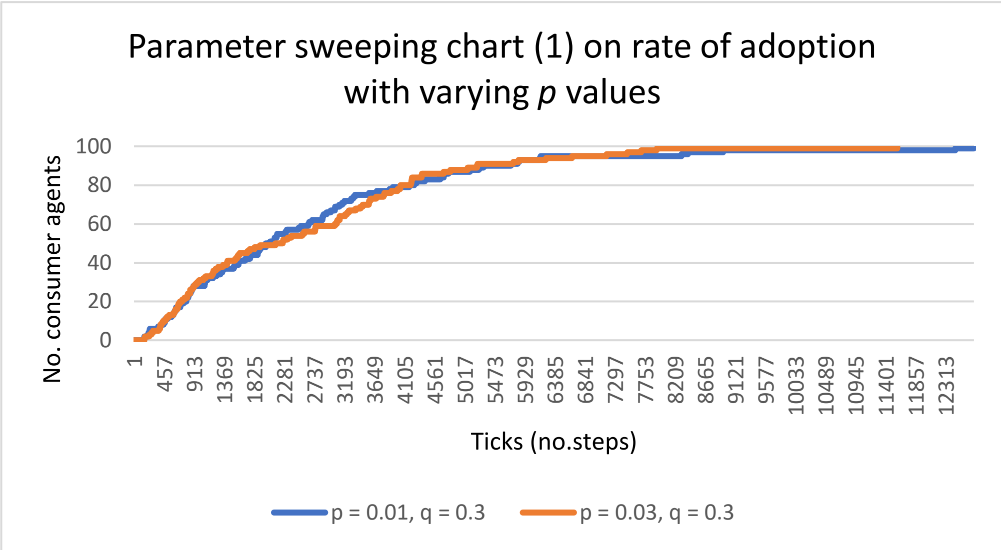


Figure 47: Parameter sweeping chart (1) on rate of adoption with varying *p* values

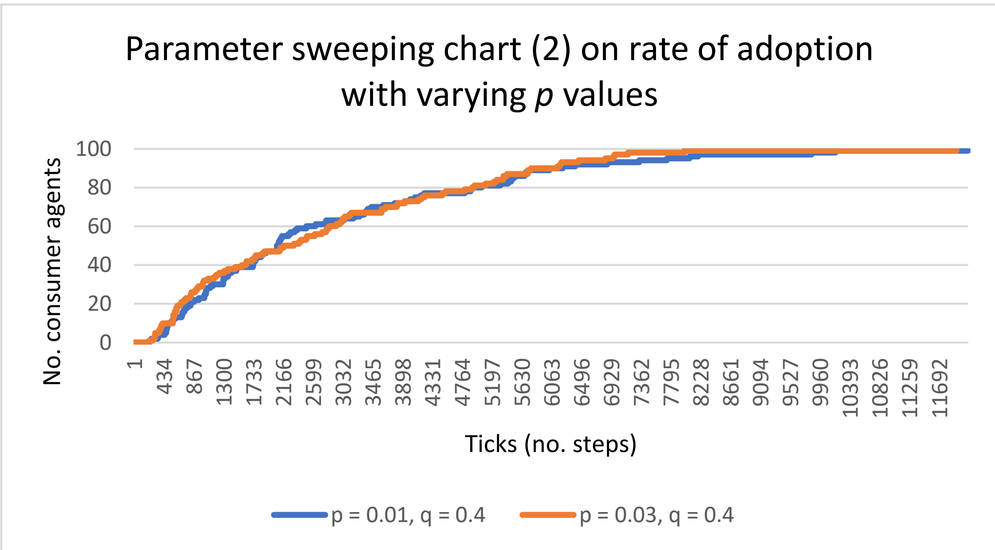


Figure 48: Parameter sweeping chart (2) on rate of adoption with varying *p* values

Figure 49 shows the second parameter chart with altering *q* (social influence) values on rate of adoption. From initialization, a higher *q* value represented by the orange line of 0.5 has a faster rate of adoption compared to a lower *q* value represented by the blue line. All 100 consumer agents have adopted the product by around the 10,000th tick where a higher *q* value has occurred than the lower *q* value where all consumer agents have adopted the trainers product by around the 11,500th tick. A higher social influence results in a faster rate of adoption due to the neighbour threshold being fulfilled through WOM effect.

Figure 50 shows the fourth parameter chart of different *q* values on rate of adoption. The blue line represents a lower *q* value of 0.3 than the higher *q* value of 0.5 shown by the orange line. There is a higher adoption rate when the *q* value is higher. All consumer agents have adopted the product around the 11,300th tick compared to a lower *p* value where all consumer agents adopt the product around the 12,200th tick. Even though, there is a faster rate of diffusion for the lower *p* value, it is evident that due to the *q* value being higher, social influence contributes to all consumer agents adopting quicker. Comparing charts 3 and 4 which can be seen in Figures 49 and 50, shows the sensitivity of adjusting parameters to the model. A higher *p* value of 0.02 in Figure 50 has a higher diffusion rate compared to the *p* value in Figure 49.

The fifth parameter sweeping chart shown in Figure 51, illustrates the effect of rate of adoption with a higher *q* value. A higher *q* value which is represented by the orange line shows a faster rate of adoption, with all 100 consumer agents adopting the product at around the 9800th tick compared to a lower *q* value represented by the blue line where all consumer agents adopt around the 11,600th tick. As the *q* value is higher, there is a greater level of social influence exhibited in the simulation.

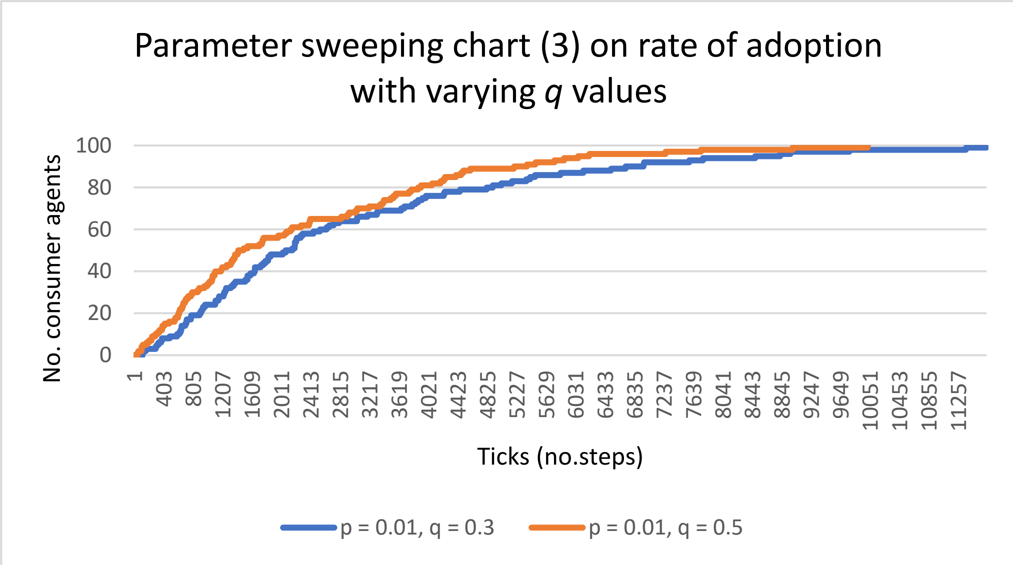


Figure 49: Parameter sweeping chart (3) on rate of adoption with varying *q* values

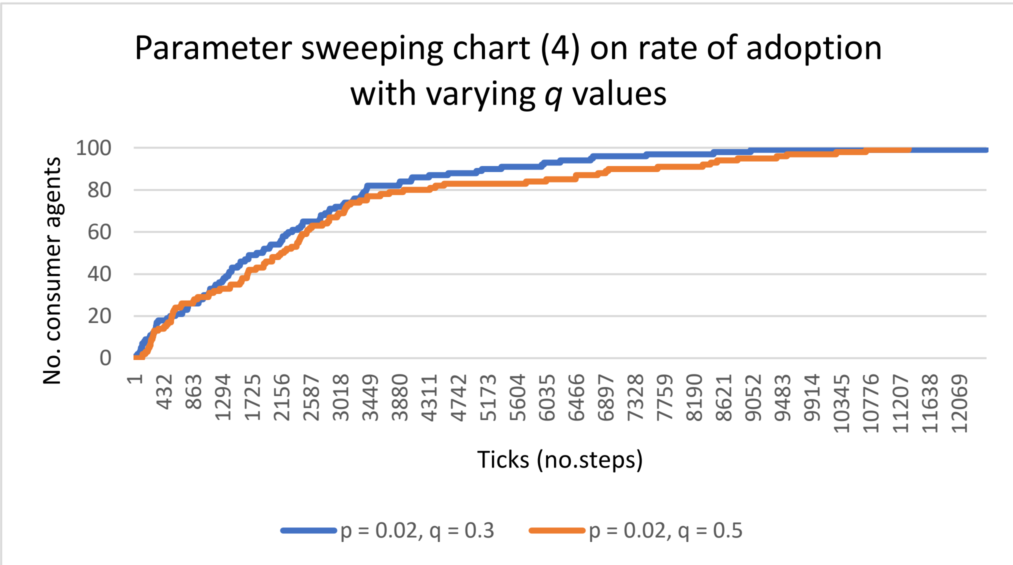


Figure 50: Parameter sweeping chart (4) on rate of adoption with varying *q* values

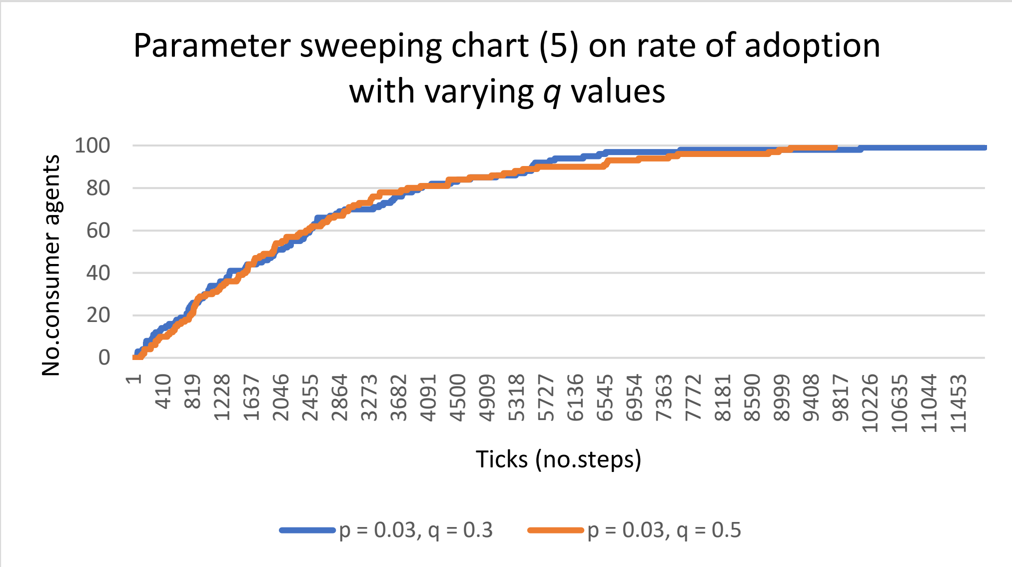


Figure 51: Parameter sweeping chart (5) on rate of adoption with varying *q* values

Figure 52 visualizes a parameter sweeping chart incorporating the Bass parameters as *p* value of 0.3 and *q* value of 0.38 represented by the grey line. The blue line has a *p* value of 0.01 and a *q* value of 0.3 which takes the longest for all 100 consumer agents to adopt the product. The lower the value of *q* which is represented by the blue line takes the longest to adopt followed by the Bass parameters. The orange line which has a *p* value of 0.02 and a *q* value of 0.3 initially has a higher rate of adoption until around the 500th tick and then the Bass parameters accelerates the rate of adoption at around the 2,600th tick. The higher value of *p* suggests the market is quick to adopt the product and the higher value of *q* displays a greater impact of social influence. As the Bass parameters have a higher *p* and *q* value in this simulation, there is a greater rate of adoption. The impact of social influence begins to appear as the simulation progresses resulting in a faster rate of adoption. This chart concludes, the higher the *p* and *q* values, the faster the rate of adoption.

Figure 53 shows a parameter sweeping chart incorporating the Bass parameters implemented in this study represented by the orange line. The blue line represents a *p* value of 0.03 and a *q* value of 0.3. The grey line represents a *p* value of 0.03 and a *q* value of 0.5. The higher the *q* value, the faster the rate of adoption with all consumer agents adopting the trainers product at around the 9600th tick represented by the grey line.

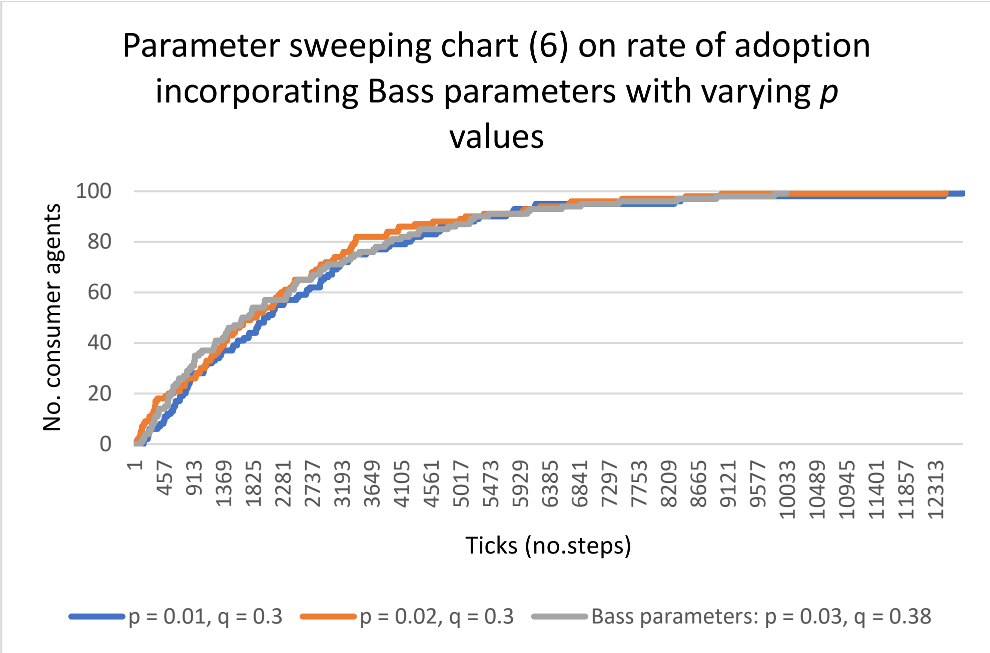


Figure 52: Parameter sweeping chart (6) on rate of adoption incorporating Bass parameters with varying *p* values

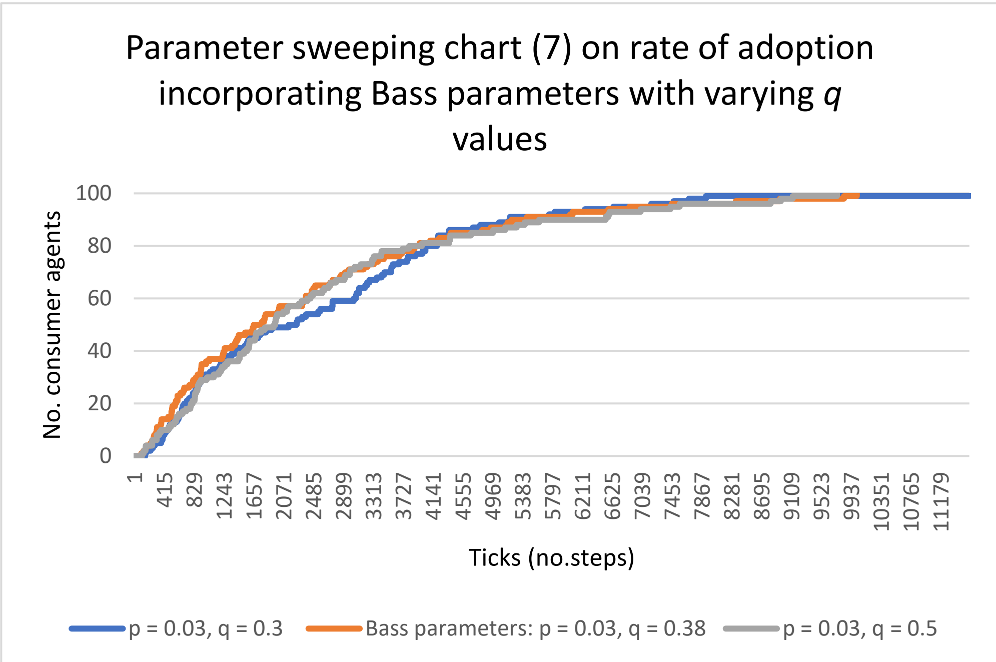


Figure 53: Parameter sweeping chart (7) on rate of adoption incorporating Bass parameters with varying *q* values

6.10 Contribution and Chapter summary

Chapter 5 created the foundation for developing an agent-based model. However, due to the absence of sentiment and price variables, the simulation process was interrupted, and the simulation was not able to advance into the decision-making process of consumer agents and ultimately adopting the trainers products. Therefore, Chapter 5 informed the development of the SAABM framework for this research by observing the effects of the simulation by conducting sentiment analysis as seen in Chapter 4 and ABM separately did not meet the requirements of this study. In order to achieve the aim of this study, this Chapter highlights the development of the SAABM framework for combining sentiment analysis and ABM for understanding consumer behaviour in a dynamic marketplace. The findings from the simulation adhered to this research's expectations that the higher the *p* and *q* values, the faster the rate of adoption. It was also expected that high sentiment and high price, low sentiment and low price, neutral sentiment and neutral price were presented as shown in Figures 35, 36 and 37 that all consumer agents eventually adopt the trainers products as it is one of the assumptions of the Bass model (Bass, 1969). In addition, as one of the assumptions of this research with reference to Chapter 4 that sentiment and price are positively correlated, it is expected that with low sentiment and high price or high sentiment and low price, the output resulted in no adoption taking place as the decision-making threshold is not satisfied to progress the simulation of adopting the trainers product.

This Chapter introduced the SAABM framework using ABM incorporating variables such as sentiment and price. The simulation outputs present what-if scenario's for high sentiment and high price, low sentiment and low price which can be seen in Figures 35, 36 and 37. With regards to agent behaviour, when input values are low sentiment and high price or high sentiment and low price, the adoption thresholds are not satisfied and therefore the adoption does not take place. Bass parameters for coefficient of innovation and coefficient of imitation were applied and parameter sweeping was applied to see what the effect would be on diffusion of innovation. Based on the investigation above, what-if scenario's visualize rate of adoption is faster in a larger population. In addition, the higher the *p* and *q* values, the faster the rate of adoption. Increasing the number of innovators, increases the social influence diffusion rate. Furthermore, by visually observing the simulation while it is running, interactions between agents and the environment can be observed. The SAABM framework can be utilized by companies across different industries to better understand consumer behaviour which can aid in DDD. Some of the research and managerial contributions deriving from this research is to observe the impact of online reviews in a dynamic marketplace with adjusting parameters such as sentiment, price, coefficient of innovation and social influence. This allows for companies such as Amazon to monitor and adjust their pricing strategy. According to Wilensky and Rand (2015), as social complexity is not able to be adequately captured by traditional modelling tools, more advanced computational models such as ABM allow social phenomena to be examined.

As evident from the SAABM framework, the mechanisms such as the presence or absence of sentiment and price variables impact the interactions of entities in the simulation model. As this study aims to simulate how consumer behaviour can be influenced by the environment, by incorporating what-if scenario's through ABM, can aid in better understanding of consumer behaviour and evaluating the impact of such strategic decisions which can be utilised by senior management in order to aid with decision-making for the company (Kiesling *et al.,* 2012). Through what-if scenario's, simulation variables can be adjusted in order to observe the effects this has on interactions of sentiment analysis in a dynamic marketplace. Furthermore, by implementing what-if scenario's, behavioural analysis can be carried out by inspecting the behaviour of a complex system under alternative conditions (Bartolini *et al.,* 2011). As Crooks and Wise (2013), state, there is an increasing awareness and research gap of implementing social network such as sentiment analysis and ABM together rather than separately resulting in a more powerful synergy. Furthermore, Kim *et al.,* (2016); Dani, Li and Liu, (2017) state that the majority of previous studies on social media research have concentrated on one or two techniques Therefore, this research proposes the SAABM framework which integrates sentiment analysis and ABM for understanding consumer behaviour in a dynamic marketplace.

Chapter 7

Conclusion

7.1 Introduction

This Chapter presents the conclusion for this study. Section 7.2 summarizes this research study. Section 7.3 highlights the achievement of the aim and objectives and the contributions derived from this study. Lastly, Section 7.4 expresses the limitations of this research as well as suggestions for possible future work in the area of sentiment analysis and ABM.

7.2 Research summary

This thesis investigated the SAABM framework for combining sentiment analysis and ABM for dynamic marketplace analysis. Due to the abundant amount of data being generated through online social media channels, there is a need to analyze this form of unstructured textual data for analysis which help companies and researchers make data-driven decisions by delving deeper into better understanding consumer behaviour.

In reflection of this study a refinement was made in Chapter 4 while implementing sentiment analysis. Through conducting the literature review, TextBlob and VADER were mentioned as being the most effective for analyzing social media textual data. However, as TextBlob gives an output of 2 scores; polarity and subjectivity this was not suitable for performing correlation analysis. Whereas only compound score implemented by VADER was selected for the purposes of this research. As VADER results in calculating the sum of negative, neutral and positive scores and then normalizing between -1 which indicates a very negative sentiment and +1 which indicates very positive sentiment; this is more suitable for carrying out correlation analysis against price. Through the application of sentiment analysis, this research was able to extract key insights into Amazon consumer reviews of trainers products and identify the most common topics consumers are associating with the products through topic modelling. In addition, correlation analysis was conducted to examine whether there were correlations between sentiment and price. Furthermore, ABM was applied to simulate the parameters of sentiment and price variables between consumers and seller agents in a dynamic marketplace environment. By integrating the coefficient of innovation and coefficient of imitation from the Bass model (Bass, 1969) rate of adoption was explored.

The implications of the results derived from this research contributes to further understanding consumer behaviour and contributes to developing a simulation model that includes what-if scenarios to predict the effects of sentiment on price. On a managerial scale, it allows businesses to be aware of sentiment analysis as a tool to delve deeper into understanding consumer behaviour and allows Amazon sellers to monitor their pricing strategy.
By combining time series analysis with NLP, this study highlights how the sentiment of unstructured text data changes over time, as well as use it to predict future data trends which allows senior stakeholders to make better business decisions (Kim 2016).

7.3 Achievement of aims & objectives

The aim of this thesis was to investigate consumer behaviour on sentiment and price through sentiment analysis and ABM. The following explains achieving each of these objectives:

Objective 1: Analyze literature
Through conducting a literature review, concepts such as diffusion of innovation, sentiment analysis, topic modelling and ABM were explored as the most suitable methods to be applied in this study. The focus of the literature review was to gain a deeper understanding of these methodologies in the area of consumer behaviour. Chapter 2 discusses sentiment and ABM have been applied to many domains individually such as, management sciences, marketing, social sciences and political contexts. In the area of ABM, this has been applied to various simulation contexts such as organizational simulation, market and diffusion simulation (Macal and North, 2010). However, there are limited studies on combining sentiment analysis and ABM in a consumer behaviour context.

Objective 2: NLP application to analyze sentiment analysis
From the literature review, it is evident that as data in the social media space is growing intensively, there is a research gap in analyzing such unstructured, noisy text data which goes beyond what traditional modelling methods are capable of. Therefore, NLP methods were explored in Chapter 2 along with

discussing their limitations. The most popular tools for NLP were speech recognition, POS, NER, machine translation, sentiment analysis and topic modelling. The justification for applying sentiment analysis and topic modelling was stated in Chapter 2 as the most suitable methodology to achieve the research aim and objectives presented in this study. As sentiment analysis and topic modelling have the capability to extract key insights from textual data in the form of online consumer reviews, polarization is able to be carried on to a simulation platform such as ABM.

Objective 3: ABM for diffusion of innovation
As mentioned in Chapter 2, the advantage of applying ABM for diffusion of innovation is to model agent heterogeneity and to observe social interactions. By incorporating what-if scenarios using Bass model (Bass, 1969) parameters, the effect of coefficient of innovation and coefficient of imitation can be explored. However, as the agent-based model developed in Chapter 5, did not define adoption thresholds, and didn't include decision-making parameters, the simulation was not able to advance. Therefore, this led to objective 4 to create the SAABM framework to include adoption parameters.

Objective 4: SAABM framework that combines sentiment analysis and ABM
It is evident from the first objective that there are limited studies which combine sentiment analysis and ABM in an online social media context. Therefore, to address these limitations, Chapter 6 presents the SAABM framework which combines sentiment analysis and ABM to investigate social interactions and behaviour of consumers and seller in a dynamic marketplace environment. The data collected and presented in Chapter 4 for correlation analysis is used for the agent-based model to observe what-if scenarios and to simulate real-world phenomena. Rate of adoption was quicker in a larger population where the higher the p and q values, the faster the rate of adoption. In addition, increasing the number of innovators, increases the social influence (q) diffusion rate. This framework can be utilized by companies and researchers to better understand consumer behaviour through applying a combination approach of sentiment analysis and ABM.

Objective 5: Evaluate SAABM framework through a case study approach
The SAABM framework was evaluated through a case study approach which consisted of Nike, Puma and Timberland trainers products. The justifications of

using these products are explained in Chapter 3 which is largely due to availability of data and price fluctuations. To the best of the authors knowledge, this study is among the first to use these brands of trainers products to evaluate the SAABM framework.

7.4 Research Limitations and Future Work

Although this research presented several research contributions, there are some limitations to this research. Firstly, as this study only looks at trainers products, future studies may want to look at other products in different contexts such as durable goods and different footwear brands. Following on from this, another limitation is this research only focusses on data collected from the Amazon UK website and other online social media platforms such as eBay, sports online shopping websites such as jd.com or asos.com can be used to extract and analyze online consumer reviews. As well as this, the framework can be applied on a global scale using international online shopping sites. By increasing the number of seller agents in the agent-based model, rate of adoption could be observed. As this study only used Amazon sellers, further work could incorporate 3rd party sellers and pricing could be investigated whereas a general rule, 3rd party sellers sell used products at a lower price value. Another major limitation is the sentiment analysis algorithm could be improved. For example, in Chapter 4 some words were picked up being incorrectly spelled such as, 'qualiti,' 'happi,' and 'realli' which should be a 'y' at the end of the word instead of 'i.' Improving this algorithm would improve its accuracy. Furthermore, as this study only implemented the Bass model (Bass, 1969), future work could investigate other methods of diffusion of innovation such as Rogers (2003); to categorize adopters into innovators, early adopters, early majority, late majority and laggards. In terms of simulation modelling, as consumers purchasing decision is complex, more variables could be added to the model. For correlation analysis as seen in Chapter 4, as there was not enough data for some of the correlations, this research could be undertaken on a much larger scale so that there is more availability of data which can improve the accuracy of the sentiment analysis model. Another scope for future work can observe ABM in real-time where the model would be adaptive and behavioural trends can be continuously updated to improve consumer behaviour prediction and observe new interactions. As ABM involves complexity, there are some challenges in validating the model.

Future work can research more on validation and verifiability techniques such as parameter calibration.

Potential limitations or challenges that may arise from combining sentiment analysis and ABM is the lack of available data which was mentioned in Chapter 4 which identified around a third of the data collected for the 100 trainers products were disregarded and not suitable to be used in this study. In addition, a challenge may arise from conducting sentiment analysis on textual data. As social media data may involve slang, informal text and abbreviations, not all sentiment analysis libraries are relevant for analysis. Therefore, this concern was addressed by undertaking the literature review and researching VADER which specifically examines social media text to overcome such noisy, complex data.

The implications of the SAABM framework can lead to more insightful analyses of market dynamics, consumer behaviour and the impact of sentiment on decision-making. It can create enhanced predictive capabilities in both academia and industry in the field of marketplace analysis. It can also improve customer insights by understanding how customer sentiment influences purchasing decisions. Companies can gain valuable insights by conducting sentiment analysis and optimizing product recommendations. There is potential for studies to explore sentiment models to enhance personalization in marketplaces, such as in e-commerce and content recommendation. Businesses can create more personalized user experiences, such as tailoring product recommendations and content suggestions based on a customer's sentiment and preferences. By analyzing how sentiment affects market behaviour, researchers can contribute to the development of economic policies and regulations. The combination of sentiment analysis and ABM can contribute to academia by investigating marketing research by examining the effectiveness of advertising campaigns and consumer responses. Practically, marketers can fine-tune their advertising strategies by understanding how positive and negative sentiment influences customer engagement. For example, they can use sentiment-driven insights to create more effective advertising campaigns. Another potential avenue for applying the SAABM framework is using real-time data to make informed decisions in dynamic market environments. In industry, real-time decision support systems powered by sentiment analysis and ABM can help traders, investors, and businesses adapt to changing market conditions. A combination approach of sentiment analysis and ABM for dynamic marketplace analysis has the potential to drive advances in research and practical applications in both academia and industry. It can provide a more holistic view of market dynamics, consumer behaviour, and decision-making processes, enabling better-informed strategies and actions in a dynamic marketplace.

7.5 Chapter summary

In conclusion, this chapter presented the research summary of this thesis, highlighting the research problem as stated in Chapter 1**.** According to Snelson (2016), social media research is an emerging field, and this is evident from Chapter 2, where many users are using social media platforms to convey their experience of a particular product/service. Through posting online reviews, a large amount of unstructured data consisting of noisy text is often beyond what traditional methods can manage and a need arises to apply advanced tools and technologies for analysis. Chapter 3 examined the case study approach to be the most applicable evaluation tool to be used in this study as well as identifying Amazon trainers products to be considered as the main data to be used in this research. Chapter 4 developed a sentiment analysis model which produced wordclouds and topic modelling as visualizations. Correlation analysis reported 6 Nike products, 11 Puma products and 11 Timberland products were positively correlated from the total of 100 trainers products collected. Chapter 5 explored ABM for diffusion of innovation which presented no decision-making process was able to be performed due to no decision-making variables being defined. Consequently, the SAABM framework was presented in Chapter 6 which incorporated sentiment and price variables into the agent-based model. The results were able to observe the behaviour and interactions of consumer agents and an Amazon UK seller in a dynamic marketplace environment through the application of what-if scenarios. The what-if scenarios visualized the process of adoption with high sentiment and high price resulting in changing the internal attributes of the consumer agents to happy smiley faces, neutral sentiment and neutral price showed a neutral face and low sentiment and low price presented sad faces. Parameter sweeping charts were developed which showed rate of adoption is quicker in a smaller market size, increasing the number of innovators, increases the rate of adoption and the higher the p and q values, the higher the rate of adoption. This chapter presented a summary of this research in Section 7.2. Section 7.3 reflected on achieving the aims and objectives of this research as well as the contributions derived from this thesis and lastly, Section 7.4 states the limitations of this research as well as potential for future work.

Bibliography

Al-Ayyoub, M., Khamaiseh, A., Jararweh, Y., Al-Kabi, M. (2018). A comprehensive survey of arabic sentiment analysis. *Information Processing & Management*. 56.

Alghamdi, A.H. (2013). Adapting Design-Based Research as a Research Methodology in Educational Settings.

Al-Mamary, S., Hasan, Y., Al-Nashmi, M., Hassan, Y., Shamsuddin, A. (2016). A Critical Review of Models and Theories in Field of Individual Acceptance of Technology. *International Journal of Hybrid Information Technology*. 9. pp.143-153.

Altrabsheh, N. (2016). Sentiment analysis on students’ real-time feedback PhD Thesis, University of Portsmouth, United Kingdom.

Alvino, L. (2019). How Can We Improve Consumer Behaviour Research? A Critical Literature Review on the Contributions and the Limitations of Consumer Neuroscience.

Amazon. [Online]. Amazon.co.uk. Available at: https://www.amazon.co.uk/

Amin, N. (2020). Boost Your Exploratory Data Analysis with Pandas Profiling. [online] Medium. Available at: https://am-nazerz.medium.com/boost-your-exploratory-data-analysis-with-pandas-profiling-93174634b9c8 [Accessed 11 April 2021].

Anderson, K.B., Anderson, C.A., Dill, K.E. and Deuser, W.E. (1998), The interactive relations between trait hostility, pain, and aggressive thoughts. *Aggressive Behaviour*., 24 (3) pp.161-171.

Arnson, D. (2022). *Business Innovation*. Routledge.

Bafna, P., Pramod, D., and Vaidya, A. (2016). Document clustering: TF-IDF approach. *International Conference on Electrical, Electronics, and Optimization Techniques*. Available at: https://www.semanticscholar.org/paper/Document-clustering%3A-TF-IDF-approach-Bafna-Pramod/3f45923b74eae06db312008512faf299816e7ab8 [Accessed 5 August 2021].

Barbosa, J and P. Leitão, L. (2011). Simulation of multi-agent manufacturing systems using Agent-Based Modelling platforms. *9th IEEE International Conference on Industrial Informatics*, Lisbon, Portugal. pp. 477-482.

Barney, N. (2023). named entity recognition (NER). [Online]. TechTarget. Last Updated: March 2023. Available at: https://www.techtarget.com/whatis/definition/named-entity-recognition-NER [Accessed 15 February 2022].

Barrett, J. S., Jayaraman, B., Patel, D., & Skolnik, J. M. (2008). A SAS-based solution to evaluate study design efficiency of phase I pediatric oncology trials via discrete event simulation. *Computer methods and programs in biomedicine*, 90 (3). pp. 240–250.

Bartolini, A., Lombardi, M., Milano, M., Benini, L. (2011). Neuron Constraints to Model Complex Real-World Problems. *Principles and Practice of Constraint Programming. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg.

Bass, F. M. (2004) Comments on 'A new product growth for model consumer durables': the bass model. *Management Science*, 50 (12). pp. 1833-1840

Bass, F. M. (1969). A New Product Growth for Model Consumer Durables. *Management Science*, *15*(5), 215–227.

Bass, F. M., Krishnan, T.V., Jain, D.C. (1994). Why the bass model fits without decision variables. *Marketing Science,* 13 (3), pp. 203-223

Bastani, K., Namavari, H., Shaffer, J. (2019). Latent Dirichlet Allocation (LDA) for Topic Modeling of the CFPB Consumer Complaints. *Expert Systems with Applications*. 127.

Bell, D and Mgbemena, C. (2018). Data-driven agent-based exploration of customer behavior. *Simulation: Transactions of the Society for Modeling and Simulation International.* 94(3) pp. 195-212.

Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., Quattrociocchi, W. (2016) Users Polarization on Facebook and Youtube. *PLoS one*. 11(8).

Bhat, A. (2015). Experimental Research: Types of Designs. [Online]. QuestionPro. Available at: https://www.questionpro.com/blog/experimental-research/ [Accessed 6 September 2020].

Bhullar, G., Khullar, A., Kumar, A. and Malhi, A.K. (2022). Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient. [online] ResearchGate. Available at: https://www.researchgate.net/publication/360235192_Time_series_sentiment_analysis_SA_of_relief_operations_using_social_media_SM_platform_for_efficient_resource_management [Accessed 2 January 2023].

Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python. United States: O'Reilly Media.

Boehner, R. and Gold, S. (2012). Modeling the impact of the marketing mix on the diffusion of innovation in the generalized bass model of firm demand. *Developments in Business Simulations & Experiential Exercises*. 39, pp.76-91.

Braithwaite, N. (2021). The history of sneakers: from commodity to cultural icon. [Online]. The Conversation. Available at: https://theconversation.com/the-history-of-sneakers-from-commodity-to-cultural-icon-127268 [Accessed 20 July 2021].

Bryman, A. and Bell, E. 2011. Business Research Methods. 3rd edition. Oxford: Oxford University Press.

Brynjolfsson, E., Hitt, L.M., Kim, H.H. (2011). Strength in Numbers: How Does Data-Driven Decision making Affect Firm Performance?.

Buchmann, C.M., Grossmann, K. and Schwarz, N. (2016). How agent heterogeneity, model structure and input data determine the performance of an empirical ABM – A real-world case study on residential mobility. [online] 75, pp.77–93.

camelcamelcamel.com. [Online]. Available at: https://camelcamelcamel.com/

Cao, L. (2016) Data science: Nature and pitfalls. *IEEE Intelligent Systems*, 31 (5) pp. 66–75.

Carter, R. 2022. The ultimate list of online review statistics for 2022. Available at: https://findstack.com/online-review-statistics/ [Accessed 12 January, 2023].

Čavoški, S., Marković, A. (2016). Agent-based modelling and simulation in the analysis of customer behaviour on B2C e-commerce sites. *Journal of Simulation* pp. 1–11.

Chang, H.C. (2010), A new perspective on Twitter hashtag use: Diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology*. 47 (1) pp.1-4.

Changchit, C., and T. Klaus. 2020. Determinants and impact of online reviews on product satisfaction. *Journal of Internet Commerce* 19 (1), pp.82–102.

Charan, A. (2015). The Bass Diffusion Model. [Online]. MarketingMind. Available at: https://www.ashokcharan.com/Marketing-Analytics/~pv-bass-diffusion-model.php#gsc.tab=0 [Accessed 8 September 2020].

Chauhan, K. (2018). Unsupervised Text Summarization using Sentence Embeddings. Medium. Available at: https://medium.com/jatana/unsupervised-text-summarization-using-sentence-embeddings-adb15ce83db1 [Accessed 14 November 2022].

Chitrakorn, K. (2020). Pop-up power: Why short-term stores are set to thrive. Vogue Business. Available at: https://www.voguebusiness.com/consumers/pop-up-power-why-short-termstores-are-set-to-thrive [Accessed 20 May 2021).

Chow, G.C. (1960) Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28, pp.591-605.

Christie, M.A., Glimm, J., Grove, J.W., Higdon, D.M., Sharp, D.H., Wood-Schultz, M.M (2005). Error Analysis and Simulations of Complex Phenomena. *Los Alamos Science.* 29.

Coussement, K and Poel, D.V.D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating. *Expert Systems with Applications*. 36. pp.6127-6134. [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S095741740800479X [Accessed 9 June 2020].

Crooks, A.T, and Wise, S. (2013). GIS and agent-based models for humanitarian assistance. *Computers, Environment and Urban Systems*, 41. pp.100–111.

Croudace, L. (2022). How our spending has changed since the end of coronavirus (COVID-19) restrictions. [online] Ons.gov.uk. Available at: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/articles/howourspendinghaschangedsincetheendofcoronaviruscovid19restrictions/2022-07-11 [Accessed 7 January 2023].

Dale, R., Moisl, H. and Somers, H. (2000). Handbook of Natural Language Processing (1st ed.). CRC Press.

Dani, H., Li, J., & Liu, H. (2017). Sentiment Informed Cyberbullying Detection in Social Media.

Dearing, J and Cox, J. (2018). Diffusion Of Innovations Theory, Principles, And Practice. *Health Affairs*. 37. pp.183-190.

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., and Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. *Social Networks, 50,* pp. 6–16.

DeVries, E.J. (2004). Epistemology and Methodology in Case Research: A Comparison between European and American IS Journals, PrimaVera Working Paper.

Dennis. (2020) eCommerce Dynamic Pricing: How to Burst Your Profit Margin. [Online]. core dna. Available at: https://www.coredna.com/blogs/dynamic-pricing-ecommerce [Accessed 15 May 2022].

Ding, Z., Gong, W., Li, S.-H. and Wu, Y. (2018). System Dynamics versus Agent-Based Modeling: A Review of Complexity Simulation in Construction Waste Management. [online] 10(7), pp.2484–2484.

Dooley, K.E. (1999). Towards a Holistic Model for the Diffusion of Educational Technologies: An Integrative Review of Educational Innovation. *Educational*

Technology & Society. 2(4), pp.35-45. [Online]. Available at: https://www.jstor.org/stable/jeductechsoci.2.4.35 [Accessed 12 April 2020].

Doyle, G. J., Garrett, B., and Currie, L. M. (2014). Integrating mobile devices into nursing curricula: Opportunities for implementation using Rogers' diffusion of innovation model. *Nurse Education Today*, 34, pp.775–782.

Dubiel, B., and Tsimhoni, O. (2005). Integrating agent based modeling into a discrete event simulation. *Proceedings of the Winter Simulation Conference*, Orlando, FL, USA.

Emiliano, A., Brida, J.G., London, S. (2021). ABM documentation and ODD Protocol in Economics: a bibliometric analysis. *Advances in Complex Systems.* World Scientific Publishing Co. 24(3/4). pp. 1-26.

Feng, X., Fu, S. and Qin, J. (2016). Determinants of consumers' attitudes toward mobile advertising: The mediating roles of intrinsic and extrinsic motivations. [online] 63, pp.334–341.

Ferreira, A.C., Covas, R., Silva, L.R., Esteves, S.C., Duarte, I.F., Fortuna, R., Theron, F., Doutrelant, C., Farine, D.R. (2020). How to make methodological decisions when inferring social networks. *Ecology and Evolution.* 10(17) pp. 9132-9143.

Frederiks, E.R., Stenner, K., Hobman, E.V. (2015). The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review. *Energies*. 8(1) pp.573-609.

Foreshew, J. (2015). Numbers add up for data scientists, the century's 'sexiest' job*.* Available: http://www.theaustralian.com.au/business/technology/numbers-add-up-for-data-scientists-the-centurys-sexiest-job/story-e6frgakx-1227213803103?sv=11f67194e5058cc0660e627d3d3666d0. Last accessed 12 February 2020.

Fournier S and Avery J. (2011), "The Uninvited Brand," *Business Horizons*, 54 (3), pp.193–207.

Gandhi, Y. (2022). What is Latent Semantic Analysis in NLP? Advantages and Disadvantages. [Online]. analyticsteps. Available at:

https://www.analyticssteps.com/blogs/what-latent-semantic-analysis-nlp-advantages-and-disadvantages [Accessed 11 October 2021].

García, D., Abisheva, A., Simon, Schweighofer, Serdült, U., & Schweitzer, F. (2014). Network polarization in online politics participatory media.

Garnett, J., & Bedford, T. (2004). Towards a hybrid modelling system of human performance. *Paper presented at the UK Operational Research Society Simulation Workshop (SW2004), Hornton Grange, Birmingham, UK.* 23-24th March 2004.

Gatti, M.A., Cavalin, P.R., Neto, S.M., Pinhanez, C.S., Santos,C.N., Gribel, D., and Appel, A. (2014). Large-Scale Multi-agent-Based Modeling and Simulation of Microblogging-Based Online Social Network. pp.17-33.

Ghali, H.S., Essawy, Y.A.S., Abdullah, A., Nassar, K. (2022). Simulation of an Earth Embankment Dam in Adverse Rainy Weather. *Construction Research Congress*.

Gilbert, N., Terna, P (2000). How to build and use agent-based models in social science. *Mind & Society 1*, pp. 57–72.

Goyal, C. (2021). Part 1: Step by Step Guide to Master NLP - Introduction. [Online]. Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/06/part-1-step-by-step-guide-to-master-natural-language-pr [Accessed 19 December 2020].

Grignard, A., Taillandier, P., Gaudou, B., Vo, D.A., Huynh, N.Q., Drogoul, A. (2013) GAMA 1.6: Advancing the art of complex agent-based modeling and simulation. *Pacific Rim international Conference on Multi-Agents (PRIMA)* 8291, pp.117-131. Springer.

Gunal, M., Pidd., M. (2005). Simulation modelling for performance measurement in healthcare. *Proceedings - Winter Simulation Conference*. pp. 2663-2668.

Harsha, A. (2022). Understanding Part-of-Speech Tagging in NLP: Techniques and Applications. [Online]. shiksha online. Available at: https://www.shiksha.com/online-courses/articles/pos-tagging-in-nlp/ [Accessed 7 March 2022].

Hatna, E., & Benenson, I. (2012). The Schelling Model of ethnic residential dynamics: Beyond the integrated–segregated dichotomy of patterns. *Journal of Artificial Societies and Social Simulation,* 15 (1).

Holland, J. (1996). Hidden order: How adaptation builds complexity. Basic Books.

Hota, H. S., Sharma, D. K., & Verma, N. (2021). Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad. *Data Science for COVID-19*, pp. 275–295.

Hotz, N. (2023). What is CRISP DM?. [Online]. Data Science Process Alliance. Available at: https://www.datascience-pm.com/crisp-dm-2/ [Accessed 19 November 2022].

Hu, M and Liu B. (2004). Mining and summarizing customer reviews. *Paper presented at the 10th ACM SIGKDD international conference on Knowledge discovery and data mining,* Seattle, WA, USA.

Hutto, C. and Gilbert, E. (2014) VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text, *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), pp. 216-225.

IBM. (1996) Cross-Industry Standard Process for Data Mining.

IBM (2021). IBM Documentation. [online] Available at: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview [Accessed 9 April 2022].

IDC. (2014). Big Data Research. Available: http://www.idc.com/prodserv/4Pillars/bigdata. Last accessed 18 February 2020.

Ismail, Z and Abu, N. (2013). A study on new product demand forecasting based on bass diffusion model. *Journal of Mathematics and Statistics*. 9. pp.84-90.

Jason, P. (2018). Sentiment analysis tutorial in Python: classifying reviews on movies and products. [Online]. tensorscience. Available at: https://www.tensorscience.com/nlp/sentiment-analysis-tutorial-in-python-classifying-reviews-on-movie [Accessed 16 March 2020].

Jiang, K and Lu, X. (2020) Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review. *IEEE 3rd International Conference of Safe Production and Informatization (IICSPI),* Chongqing City, China, 2020, pp. 210-214.

Kalantaryan, S. (2022). WORD OF MOUTH MARKETING [WHAT IT IS, STRATEGIES, TYPES & MORE]. [Online]. Digital Marketers World. Available at: https://digitalmarketersworld.com/word-of-mouth-marketing/#what-is [Accessed 9 January 2023].

Kang, M., Sun, B., Liang, T., and Mao, H. Y. (2022). A study on the influence of online reviews of new products on consumers' purchase decisions: An empirical study on JD.com. *Frontiers in Psychology*, *13*.

Kawamura, Y. (2016). Sneakers: Fashion, Gender, and Subculture. United Kingdom: Bloomsbury Publishing.

Kelton, W. D., Sadowski, R. P., Sturrock, D. T. (2007). Simulation with Arena. United Kingdom: McGraw-Hill Higher Education.

Kerdlap, P., Purnama, A.R., Low, J.S.C., Tan, D.Z.L., Barlow, C.Y., Ramakrishna, S. (2020). Environmental evaluation of distributed versus centralized plastic waste recycling: Integrating life cycle assessment and agent-based modeling. *Procedia CIRP* 90, pp.689-694.

Khanzadi, M., Attar, A. M. and Bagherpour, M. (2016) Finding Optimum Resource Allocation to Optimizing Construction Project Time/Cost through Combination of Artificial Agents CPM and GA. *Periodica Polytechnica Civil Engineering*, 60(2), pp. 169–180.

Kiesling, E., 2011. Planning the market introduction of new products: An agent-based simulation of innovation diffusion.

Kiesling, E., Günther, M., Stummer, C. and Wakolbinger, L. (2012) Agent-Based Simulation of Innovation Diffusion: A Review. *Central European Journal of Operations Research*, 20, pp. 183-230.

Kim E.H.-J., Jeong Y.K., Kim Y., Kang K.Y., Song M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science.* 42 (6), pp.763–781.

Kim, S. (2016). *Understand Customer Data Using Time Series and Sentiment Analysis .* Available: https://algorithmia.com/blog/sentiment-time-series-analysis. Last accessed 5th June 2020.

Kim, Y., Dwivedi, R., Zhang, J., Jeong, S.R. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Information Review: The International Journal of Digital Information, Research and Use*. 40 (1). pp. 42-61.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J.B., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Team, J.D. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. *International Conference on Electronic Publishing*.

Kocaman, V and Talby, D. (2022). Accurate Clinical and Biomedical Named Entity Recognition at Scale. *Software Impacts.* 13.

Kotthoff, F. and Hamacher, T. (2022), Calibrating Agent-Based Models of Innovation Diffusion with Gradients, *Journal of Artificial Societies and Social Simulation*, 25 (3).

Knights, D. and McCabe, D. (1997), How would you measure something like that?. Quality in a Retail Bank. *Journal of Management Studies*, 34. pp. 371-388.

Kruglyak, I. (2020). Natural Language Processing: Tasks and Application Areas. [Online]. avenga. Available at: https://www.avenga.com/magazine/natural-language-processing-application-areas/ [Accessed 21 January 2021].

Kumar, P. and Sinha, A. (2016) Real-time analysis and visualization of online social media dynamics. *IEEE international conference on next generation computing technologies, Dehradun*. pp 362–367.

Kumar, V., and Reinartz, W. (2016). Creating Enduring Customer Value. *Journal of Marketing*, 80(6), pp. 36–68.

Lansdall-Welfare,T., Dzogang, F., and Cristianini, N. (2016). Change-Point Analysis of the Public Mood in UK Twitter during the Brexit Referendum. *IEEE*

16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain. pp. 434-439.

Lateef, F. (2010). Simulation-based learning: Just like the real thing. 3(4), pp.348–348.

Lawani, A., Reed, M.R., Mark, T., Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, Elsevier, 75(C), pp. 22-34.

Lee, B., Collier., P.M., and Cullen, J. (2007). Reflections on the use of case studies in the accounting, management and organizational disciplines. *Qualitative Research in Organizations and Management*, 2 (3), pp. 169-178.

Lengyel, B., Bokányi, E., Riccardo, D.C., Kertész, J., Gonzalez, M.C. (2020). The role of geography in the complex diffusion of innovations. *Scientific Reports.*

Li, S. (2018). Topic Modeling and Latent Dirichlet Allocation (LDA) in Python. [online] Medium. Available at: https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24 [Accessed 15 July 2020].

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

Liu, B., Hu, M., Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. *Paper presented at the Proceedings of the 14th International Conference on World Wide Web*, Chiba, Japan.

Loria, S. (2020). TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation. [online] Available at: https://textblob.readthedocs.io/en/dev/

Luna, Z. (2021). CRISP-DM Phase 2: Data Understanding. [Online]. Medium. Available at: https://medium.com/analytics-vidhya/crisp-dm-phase-2-data-understanding-b4d627ba6b45 [Accessed 8 June 2022].

McAfee, A. and Brynjolfsson, E. (2012) Big Data: The Management Revolution. Harvard Business Review.

Macal, C. and North, M. (2009). Agent-based modeling and simulation. *Proceedings of the 2009 Winter Simulation Conference (WSC)*. 10.

MacDonald, M. (2018). Why online store owners should embrace online reviews. Available online at: www.shopify.com/blog/15359677-why-online-store-owners-should-embrace-online-reviews (accessed March 15, 2018).

Mahajan, V., & Muller, E. (1979). Innovation Diffusion and New Product Growth Models in Marketing. *Journal of Marketing*, 43(4), pp.55–68.

Mahajan, V., Muller, E., Wind, J (2000). New product diffusion models. Kluwer Academic Publishers, New York.

Mahajan, V., Muller, E., & Srivastava, R. K. (1990). Determination of adopter categories by using innovation diffusion models. *Journal of Marketing Research, 27*(1), pp.37–50.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations,* pp. 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Marzukhi, S., Awang, N., Alsagoff, S.N., Mohamed, H. (2021). RapidMiner and Machine Learning Techniques for Classifying Aircraft Data. *Journal of Physics: Conference Series.*

Masters, K. (2019). 89% Of Consumers Are More Likely To Buy Products From Amazon Than Other E-Commerce Sites: Study. [Online]. Forbes. Last Updated: 20 March 2019. Available at: https://www.forbes.com/sites/kirimasters/2019/03/20/study-89-of-consumers-are-more-likely-to-buy-pro [Accessed 14 February 2022].

Matakos, A., Terzi, E. & Tsaparas, P. (2017). Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*. 31, pp. 1480–1505.

Matthews, R., Gilbert, N., Roach, A., Polhill, J. and Gotts, N. (2007) Agent-Based Land-Use Models: A Review of Applications. Landscape Ecology, 22, pp.1447-1459.

Meade, N. and Islam, T. (2006). Modeling and forecasting the diffusion of innovation - A 25-year review. *International Journal of Forecasting*. 22, pp.529-545.

Medhat, W., Hassan, A. and Korashy, H. (2014) Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 5, pp.1093-1113.

Mehta, N., Detroja, P., Agashe, A. (2018). Amazon changes prices on its products about every 10 minutes — here's how and why they do it. [Online]. Business Insider. Available at: https://www.businessinsider.com/amazon-price-changes-2018-8?r=US&IR=T [Accessed 11 June 2021].

Muller, E. and Bass, F.M. (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54. pp. 1-26.

Murakami, A, Thompson, P, Hunston, S and Vajn, D. (2017) 'What is this corpus about?': Using topic modelling to explore a specialised corpus. Corpora, 12 (2). pp. 243-277.

Najlis, R. and North, M.J. (2004), Repast for GIS. *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*, Chicago, IL, pp. 255-260.

Naskar, D., Mokaddem, S., Rebollo, M., & Onaindía, E. (2016). Sentiment Analysis in Social Networks through Topic modeling. *International Conference on Language Resources and Evaluation*.

Negahman, A, and Smith, J.S. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems,* 33(2), pp.241-261.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research*, 43(2), pp.204–211.

Nikolenko, S. I., Koltcov, S., and Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, 43(1), pp. 88–102.

Norton, J.A. and Bass F.M. (1987), A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High Technology Products. *Management Science.* 33. pp.1069-1086.

Oliveira, T., and Martins, M.F. (2011). Literature Review of Information Technology Adoption Models at Firm Level.

Oneaţă, D. (2011). Probabilistic Latent Semantic Analysis.

Pandey, P (2018). Simplifying Sentiment Analysis using VADER in Python (on Social Media Text). [online] Breaking the Jargons. Available at: https://parulpandey.com/2018/09/23/simplifying-sentiment-analysis-using-vader-in-python-on-social-media-text/ [Accessed 18 November 2020].

Pang, B. and Lee, L. (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2, pp. 1-135.

Parsons, R. (2010). Usain Bolt signs “largest ever” athletics deal with Puma. [Online]. MarketingWeek. Available at: https://www.marketingweek.com/usain-bolt-signs-largest-ever-athletics-deal-with-puma/ [Accessed 4 September 2021].

Pascual, F. (2019). Topic Modeling: An Introduction. [Online]. MonkeyLearn. Available at: https://monkeylearn.com/blog/introduction-to-topic-modeling/ [Accessed 6 October 2021].

Pashaeypoor S, Ashktorab T, Rassouli M and Majd, H.A. (2017). Experiences of nursing students of Evidence-Based Practice Education according to Rogers' Diffusion of Innovation Model: A Directed Content Analysis. *Journal of Advances in Medical Education & Professionalism*. 5(4), pp.203-209.

Pidd, M. (2004). Computer simulation in management science (5th ed.): John Wiley & Sons, Inc. New York, NY, USA.

Poel, D.V.D. and Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research.157(1), pp.196–217.

Popescu, A-M. and Etzioni, O. (2005). Extracting Product Features and Opinions from Reviews. *Proceedings of Human Language Technology*

Conference and Conference on Empirical Methods in Natural Language Processing. Vancouver, British Columbia, Canada. Association for Computational Linguistics. pp.339–346

Prabhakaran, V., Hutchinson, B., and Mitchell, M. (2019). Perturbation Sensitivity Analysis to Detect Unintended Model Biases. *Conference on Empirical Methods in Natural Language Processing*.

Provost, F., and Fawcett, T. (2013). Data Science and Its Relationship to Big Data and Data-Driven Decision Making. Big Data, 1, pp.51-59.

Purchase, S., Denize, S., and Olaru, D. (2014). Using Case-Based Research For Agent-Based Modelling. *Field Guide to Case Study Research in Business-to-Business Marketing and Purchasing.* Advances In Business Marketing & Purchasing, Emerald Group Publishing Limited, Bingley, UK. 21. pp. 271-287.

Pylianidis, C, Snow, V, Overweg, H, Osinga, S, Kean, J and Athanasiadis. (2022). Simulation-assisted machine learning for operational digital twins. *Environmental Modelling and Software*, 148.

Python Software Foundation (2019). Python. *Software*. http://www.python.org.

Qian, S., Zhang, T., Xu, C and Shao, J. (2016). Multi-Modal Event Topic Model for Social Event Analysis. *IEEE Transactions on Multimedia*, 18 (2) pp. 233-246.

Qualtrics. (no date). Topic modelling: How a statistical method can increase business results. [Online]. Qualtrics. Available at: https://www.qualtrics.com/uk/experience-management/research/topic-modelling/ [Accessed 22 October 2020].

Raab, R., Lenger, K., Stickler, D., Granigg, W., Lichtenegger, K. (2022). An Initial Comparison of Selected Agent-Based Simulation Tools in the Context of Industrial Health and Safety Management. In *2022 8th International Conference on Computer Technology Applications (ICCTA 2022), May 12–14, 2022, Vienna, Austria.* ACM, New York, NY, USA.

Rand, W., Wilensky, U. (2008). NetLogo Spread of Disease model. http://ccl.northwestern.edu/netlogo/models/SpreadofDisease. Center for

Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

Railsback, S. F., & Grimm, V. (2012). Agent-Based and Individual-Based Modeling: A Practical Introduction. Princeton, NJ: Princeton University Press.

Ratcliff, R., and Doshi, K. (2016). Using the Bass Model to Analyze the Diffusion of Innovations at the Base of the Pyramid. *Business & Society*, 55(2), pp.271–298.

Repast, (2019). Repast Simphony Reference Manual. Version 2.7 October 2019. https://repast.github.io/docs/RepastReference/RepastReference.html.

Roberts K and Harabagiu S.M. (2012). Statistical and Similarity Methods for Classifying Emotion in Suicide Notes. *Biomedical Informatics Insights*.

Robertson, T. S., and Gatignon, H. (1986). Competitive effects on technology diffusion. *Journal of Marketing, 50* (3), pp.1–12.

Robinson, L. (2009). A summary of Diffusion of Innovations. [Online]. Changeology. Available at: https://www.enablingchange.com.au/Summary_Diffusion_Theory.pdf [Accessed 5 April 2022].

Robinson, S. (2013). Conceptual modeling for simulation. *2013 Winter Simulations Conference (WSC),* Washington, DC, USA. pp. 377-388. Ferreira, R., Pottier, L., Coleman, T., Deelman, E. and Casanova, H. (2020). WorkflowHub: Community Framework for Enabling Scientific Workflow Research and Development.

Robinson, S., Brooks, R., Kotiadis, K., & Van Der Zee, D.-J. (2010). Conceptual Modeling for Discrete-Event Simulation (1st ed.). CRC Press.

Rogers, E.M. (2003) Diffusion of Innovations. Free Press, New York.

Rogers, E.M. (1995) Diffusion of Innovations. 4th Edition, the Free Press, New York.

Rogers, E. M. (1962). Diffusion of innovations. New York, Free Press of Glencoe.

Rojewska, K. (2023). 6 ways sentiment analysis will help your business. [Online]. D Labs. Available at: https://dlabs.ai/blog/how-can-sentiment-analysis-will-help-your-business/ [Accessed 21 January 2023].

Ruane, L and Wallace, E. (2015) Brand tribalism and self-expressive brands: social influences and brand outcomes. *The journal of product & brand management,* 24 (4) pp. 333 – 348.

Rust, R. T., and Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing, 69* (2), pp.193–215.

Sahin, I. (2006). Detailed Review of Rogers' Diffusion of Innovations Theory and Educational Technology-Related Studies Based on Rogers' Theory. *The Turkish Online Journal of Educational Technology*, 5, pp.14-23.

Savin, N. (2019). The Amazon Effect: Dynamic Pricing Done Right. [Online]. Pragmatic Institute. Available at: https://www.pragmaticinstitute.com/resources/articles/product/the-amazon-effect-dynamic-pricing-done [Accessed 2 February 2019].

Sayedi, H. (2021). What is Sentiment Analysis? Types and Use Cases. [Online]. TAUS. Available at: https://www.taus.net/resources/blog/what-is-sentiment-analysis-types-and-use-cases [Accessed 4 April 2021].

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. Norton, New York.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2). pp.143–186.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2). pp. 488–493. http://www.jstor.org/stable/1823701

Schock, C., Dumler, J., Doepper, F. (2021). Data Acquisition and Preparation – Enabling Data Analytics Projects within Production. *Procedia CIRP*. 104 pp.636-640.

Schröer, C., Kruse, F., Gómez, J.M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science.* 181, pp.526-534.

Schweitzer, F., Krivachy, T., Garcia, D. (2020). An Agent-Based Model of Opinion Polarization Driven by Emotions. *Complexity*.

Shannon, R.E. (1975) Simulation: A Survey with Research Suggestions, AIIE Transactions, 7 (3), pp. 289-301.

Shannon, R.E. (1977). Simulation modeling and methodology. ACM SIGSIM Simulation Digest, 8, (3), pp. 33–38.

Shin, J.K., and Lorenz, J. (2010). Tipping diffusivity in information accumulation systems: more links, less consensus. *Journal of Statistical Mechanics: Theory and Experiment*.

Shrestha, S. (2016). Influencing Factors on Consumer Buying Behaviour of Smart Phones—A Research on the Buying Behaviour of Young Consumer in Kathmandu, Nepal.

Simudyne (2019). A Complete Guide to Agent-Based Modeling For Financial Services A Bluepaper from Simudyne. Available at: https://www.simudyne.com/wp-content/uploads/2019/08/Guide-to-ABM-new-address.pdf [Accessed 11 October 2022].

Singh, N.K., Tomar, D.S. & Sangaiah, A.K. (2020). Sentiment analysis: a review and comparative analysis over social media. *Journal of Ambient Intelligence and Humanized Computing.*

Snelson, C.L. (2016). Qualitative and Mixed Methods Social Media Research: A Review of the Literature. *International Journal of Qualitative Methods*, 15(1).

Søgaard, A. (2022). Understanding models understanding language. *Synthese*, 200 (6), pp. 443.

Statista. (2022). Topic: Nike. [online] Available at: https://www.statista.com/topics/1243/nike/ [Accessed 3 January 2023].

Sterman, J.D. (2000) Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin McGraw-Hill, Boston.

StockX (2016). World’s First Online Consumer ‘Stock Market of Things’ - StockX - Launches Today. [online] Prnewswire.com. Available at: https://www.prnewswire.com/news-releases/worlds-first-online-consumer-stock-market-of-things---stockx---launches-today-300216849.html [Accessed 20 July 2020].

Sudhira, H.S., Ramachandra, T.V., Wytzisk, A. and Jeganathan, C., (2005). Framework for integration of agent-based and cellular automata models for dynamic geospatial simulations. *Centre for Ecological Sciences, Indian Institute of Science)* Bangalore.

Sun, B., Kang, M., and Zhao, S. (2023). How online reviews with different influencing factors affect the diffusion of new products. *International Journal of Consumer Studies*, 47(4), pp.1377–1396.

Surendra, S.S. (2001). Acceptance of Web technology-based education by professors and administrators of a college of applied arts and technology in Ontario (Doctoral dissertation, University of Toronto, 2001). ProQuest Digital Dissertations. (UMI No. AAT NQ58603).

Swaminathan, V., Schwartz, H. A., Menezes, R., and Hill, S. (2022). The Language of Brands in Social Media: Using Topic Modeling on Social Media Conversations to Drive Brand Strategy. *Journal of Interactive Marketing*, 57(2), pp. 255–277.

Swarnkar, N (2020). Quantitative Finance & Algo Trading Blog by QuantInsti. *VADER Sentiment Analysis: A Complete Guide, Algo Trading and More*. [online] Available at: https://blog.quantinsti.com/vader-sentiment/#:~:text=that%20hot.%E2%80%9D.-,Compound%20VADER%20scores%20for%20analyzing%20sentiment,1%20(most%20extreme%20positive). [Accessed 19 January 2021].

Tacy, A. (2021). Diffusion Of Innovation II – Bass Diffusion, Let's Get Mathematical…. [Online]. SolvInnov. Available at: https://solvinnov.com/innovation-diffusion-part-ii/ [Accessed 10 May 2022].

Tamayo, P., Berger, C., Campos, M.M., Yarmus, J., Milenova, B.L., Mozes, A., Taft, M., Hornick, M.F., Krishnan, R., Thomas, S., Kelly, M., Mukhin, D., Haberstroh, B., Stephens, S., & Myczkowski, J. (2005). Oracle Data Mining.

Tambe, P and Hitt, L.M. (2012). The Productivity of Information Technology Investments: New Evidence from IT Labor Data. *Information Systems Research*, 23(3-1), pp.599-617.

Taylor, S. J., Bogdan, R., and DeVault, M. (2016). Introduction to Qualitative Research Methods: A Guidebook and Resource (4th ed.). Hoboken, NJ: John Wiley & Sons Inc.

Teng, J.T., Grover, V., and Güttler, W. (2002). Information technology innovations: general diffusion patterns and its relationships to innovation characteristics. *IEEE Trans. Engineering Management, 49*, pp.13-27.

Thomas, L.C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16 (2). pp. 149-172.

Torrens, P. M., & O'Sullivan, D. (2001). Cellular Automata and Urban Simulation: Where Do We Go from Here? Environment and Planning B: Planning and Design, 28(2), pp.163–168.

Tsai, Y-F and Jack, L (2015). Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues.

Turing (2022). Which Language Is Useful for NLP and Why? [online] Turing.com. Available at: https://www.turing.com/kb/which-language-is-useful-for-nlp-and-why [Accessed 11 May 2022].

Ulbinaite, A and Moullec, Y. (2010). Towards an ABM-Based Framework for Investigating Consumer Behaviour in the Insurance Industry.

Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*. 21(6).

Valente, T. (2005). Network Models and Methods for Studying the Diffusion of Innovations. Models and Methods in Social Network Analysis. *Structural Analysis in the Social Sciences*, pp. 98-116. Cambridge: Cambridge University Press.

Valente, T.W. (1995) Networks Models of the Diffusion of Innovations. Hampton Press, Inc., Cresskill, NJ, 171.

Vimaladevi, K and Dhanabhakaym, M. (2012). A Study on the Effects of Online Consumer Reviews on Purchasing Decision. *Prestige International Journal of Management & IT* – Sanchayan..

Waldron, P. (2013). Retail Analytics Creating Competitive Advantage, IBM Study Shows. Available: http://insights-on-business.com/retail/retail-analytics-study-use-of-big-data/. Last accessed 8 October 2020.

Wall, M. (2014). Ebola: Can big data analytics help contain its spread?. Available: http://www.bbc.co.uk/news/business-29617831. Last accessed 15 October 2014.

Watts, D.J. (2013). Computational Social Science: Exciting Progress and Future Directions National Academy of Engineering. (Ed). *Frontiers of Engineering*. Washington, DC: National Academies Press. pp.17-24.

Wilensky, U. (2019). *The NetLogo User Manual*. Version 6.1.1. https://ccl.northwestern.edu/netlogo/6.1.1/docs/

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. and Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo.* The MIT Press.

Wilensky, U. and Shargel, B. (2002). Behaviorspace. *Software*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. http://ccl.northwestern.edu/netlogo/behaviorspace.html

Wilensky, U. and Stroup, W. (1999). Hubnet. *Software*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. http://ccl.northwestern.edu/netlogo/hubnet.html

Xiang, X., Kennedy, R., Madey, G. and Cabaniss, S. (2019). Verification and Validation of Agent-based Scientific Simulation Models. [online] Available at: https://www3.nd.edu/~nom/Papers/ADS019_Xiang.pdf.

Yang, Mo, S.Y.K., and Zhu, X. (2014). An empirical study of the financial community network on Twitter. *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, London, UK. pp. 55-62.
Yardi, S., and Boyd, D. (2010). Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society*, *30*(5), pp. 316–327.

Yaveroglu, I., and Donthu, N. (2002). Cultural Influences on the Diffusion of New Products. *Journal of International Consumer Marketing*. 14, pp.49-63.

Zhao, L. (2019). A Dynamic Pricing Mechanism Model Based on Sentiments Analysis. *International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, Changsha, China. pp. 622-625.

Appendix

Appendix 1 – Summary statistics for the 100 trainers products

NIKE	B000ROQJFE	B00BYG3QM8	B00BYG3V1O	B00CGVQHLM	B00MR3W28G	B00PYDTR3Y	B00PYDWHGI	B00PZ8ZLKG	B00PZB1E7W	B00PZB1JB8
MEAN PRICE	29.07	35.16	34.18	93.54	no matching data	47.65	not enough data	46.61	not enough data	47.01
MEAN COMPOUND	0.64	0.47	0.76	0.27		0.39		0.73		0.72
CORRELATION	**-0.06**	**0.16**	**-0.07**	**-0.61**		**-0.33**		**-0.38**		**not enough data**
PUMA	B0783RD7K7	B000G529SA	B00DNY23TQ	B00HPZDM0S	B00ME5WBDS	B01H4CWH1G	B01MZXY8CU	B01N1OWZR6	B004SGJRVA	B004SGJSGY
MEAN PRICE	36.76	72.47	41.01	45.95	33.92	42.36	39.55	36.31	no matching data	32.11
MEAN COMPOUND	0.72	0.76	0.55	0.86	0.51	0.74	0.73	0.67		0.52
CORRELATION	**0.35**	**0.85**	**-0.41**	**0.18**	**-0.42**	**-0.02**	**0.12**	**0.41**		**0.27**
TIMBERLAND	B000G21F42	B000G28J00	B000HVK8JU	B000LEQMF2	B000VMPIZO	B000VQBM6Y	B00BCJY5F4	B00BF0OI16	B00X9CSO9U	B01KHFGDAS
MEAN PRICE	81.31	110.14	58.01	103.53	121.44	89.89	86.47	100.13	115.89	not enough data
MEAN COMPOUND	0.63	0.82	0.71	0.54	0.46	0.86	0.75	0.55	0.79	
CORRELATION	**0.08**	**-1**	**-0.17**	**-0.02**	**0.56**	**0.18**	**0.47**	**0.11**	**0.38**	

NIKE	B00QHYVRLY	B00VDHMQ7Y	B00VDHV9QI	B00VDJGAME	B00XWPX6E2	B00Y13KHPA	B01HZQUBBO	B01JQ3K04W	B07BQS2G49	B07BR2NKPN
MEAN PRICE	no matching data	79.32	not enough data	63.18	0.18	56.04	71.95	79.46	44.46	not enough data
MEAN COMPOUND		0.41		0.63	0.66	0.65	0.80	0.66	0.87	
CORRELATION		**0.02**		**-0.24**	**0.18**	**0.00**	**0.53**	**-0.24**	**not enough data**	
PUMA	B07DBXTX2P	B07DC15G7T	B07DCCCB1T	B07DCDKZG7	B07DCDNVC3	B07JH2RZ6X	B07KFXQ3NV	B07KFY13ZF	B07KFYFBDF	B07KFZSQTC
MEAN PRICE	53.52	25.99	no matching data	no matching data	56.43	38.95	no matching data	32.01	17.10	68.03
MEAN COMPOUND	0.92	0.71			0.51	0.82		-0.28	0.94	0.75
CORRELATION	**0.31**	**-0.99**			**0.60**	**-0.46**		**-0.17**	not enough data	**0.48**
TIMBERLAND	B01KL1NCIO	B01KPCB7SG	B002C62YNQ	B004L3BH54	B005JQS5O8	B06W9K87J1	B007TGA2JC	B07B4K8BN9	B07JM781GR	B07K2KX2P4
MEAN PRICE	not enough data	86.85	97.76	111.25	77.68	83.30	126.94	not enough data	78.01	not enough data
MEAN COMPOUND		0.60	0.68	0.49	0.66	0.75	0.68		0.65	
CORRELATION		**0.32**	**-0.23**	**-0.31**	**-0.13**	**0.44**	**-0.62**		**0.30**	

NIKE	B07D42WLYW	B07H8HYXB1	B07H81GYHG	B07H97D1XM	B07H979BXR	B0059L0GQ8	B075ZZ7HW2	B075ZZHVP7	B075ZZVRSL	B078JM2FXR
MEAN PRICE	not enough data	87.03	49.77	not enough data	not enough data	47.10	89.17	not enough data	94.79	not enough data
MEAN COMPOUND		0.79	0.67			0.88	0.75		0.42	
CORRELATION		**-0.23**	**0.52**			**-1**	**-1**		**1**	
PUMA	B07KG1ZXBM	B07KG6GVTY	B07PTX7HM3	B07R67MSTQ	B015OJSIOQ	B071Z4WKHK	B077J16JNM	B077MKTJSM	B077MKVLX6	B077MLPMX1
MEAN PRICE	49.97	38.47	not enough data	56.04	34.78	not enough data	not enough data	no matching data	47.22	39.72
MEAN COMPOUND	0.79	0.68		0.82	0.73				0.81	0.55
CORRELATION	**not enough data**	**0.29**		**-0.03**	**-0.07**				**0.17**	**-0.20**
TIMBERLAND	B008H2C6KA	B008H22D2Q	B08G21WXJ	B010QRQ3NE	B010QRQT0Q	B014YHPOUA	B015GZOYPA	B0070SJ7X6	B073PZFVQ8	B073X8XDTP
MEAN PRICE	89.93	103.08	116.94	92.29	89.22	76.48	79.25	77.63	73.89	not enough data
MEAN COMPOUND	0.62	0.39	0.72	0.74	0.71	0.66	0.67	0.42	0.60	
CORRELATION	**0.15**	**0.46**	**1.00**	**0.04**	**-0.22**	**0.00**	**-0.18**	**-0.04**	**1**	

NIKE	B078NVK7WX	B078X1CGNP	B0113OCUKY	B001469I4I
MEAN PRICE	not enough data	107.38	50.17	not enough data
MEAN COMPOUND		0.39	0.71	
CORRELATION		**0.26**	**-0.09**	
PUMA	B077MQ6C77	B00756ZHDW	B00756ZIE0	
MEAN PRICE	42.37	32.34	30.85	
MEAN COMPOUND	0.92	0.74	0.79	
CORRELATION	**-0.41**	**-0.07**	**-0.09**	
TIMBERLAND	B075JPKJ35	B0198WKA2S	B01846NEX6	
MEAN PRICE	not enough data	not enough data	not enough data	
MEAN COMPOUND				
CORRELATION				

Appendix 2

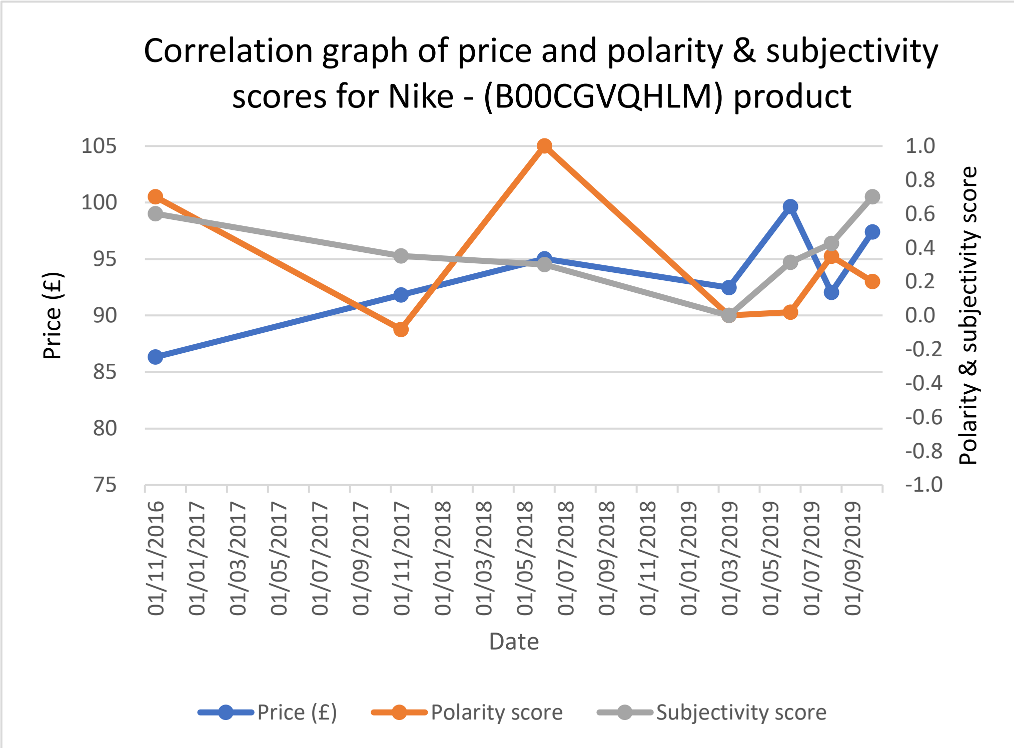


Figure A.1: Correlation graph of price and polarity & subjectivity scores for Nike B00CGVQHLM product

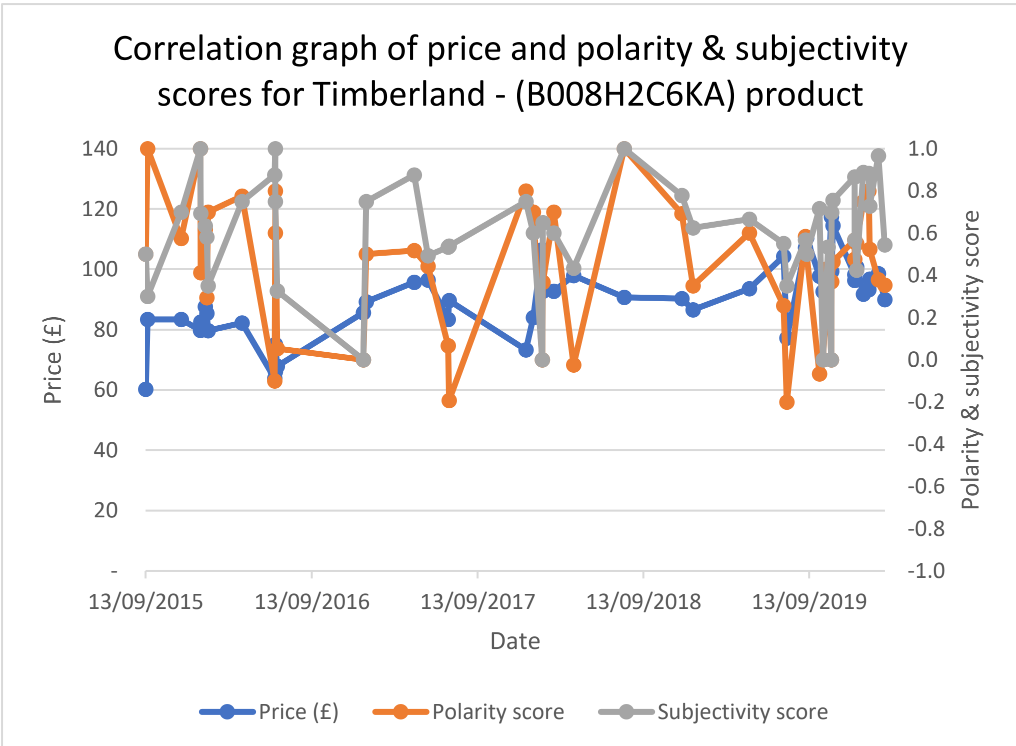


Figure A.2: Correlation graph of price and polarity & subjectivity scores for Timberland B008H2C6KA product

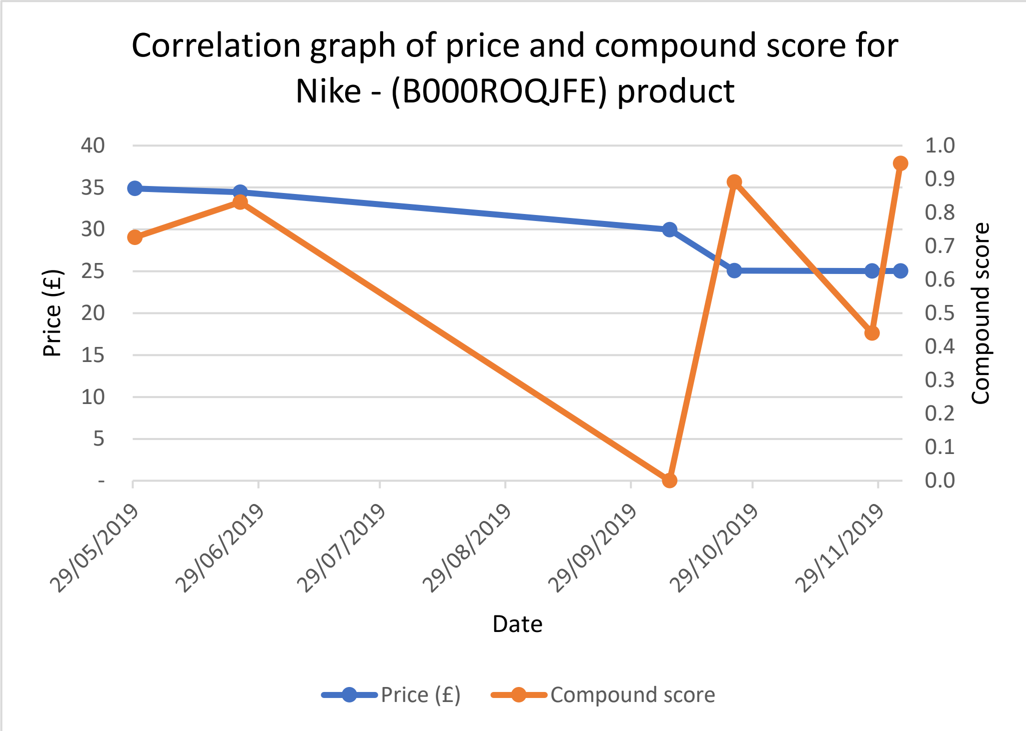


Figure A.3: Correlation graph of price and compound score for Nike B000ROQJFE product

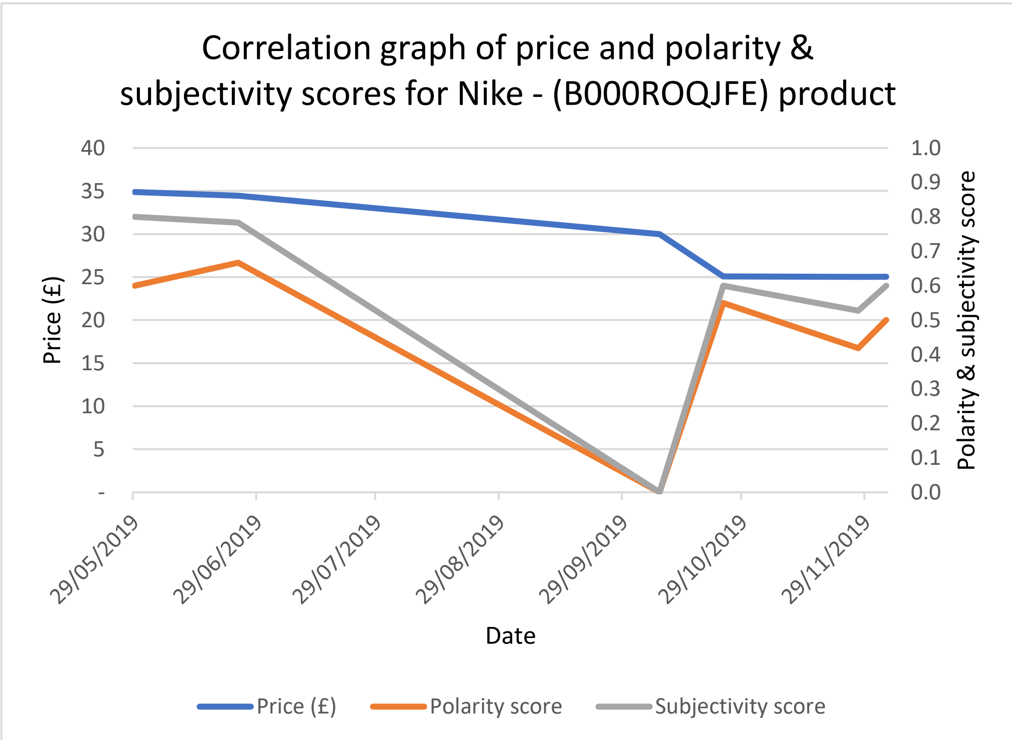


Figure A.4: Correlation graph of price and polarity & subjectivity scores for Nike B000ROQJFE product

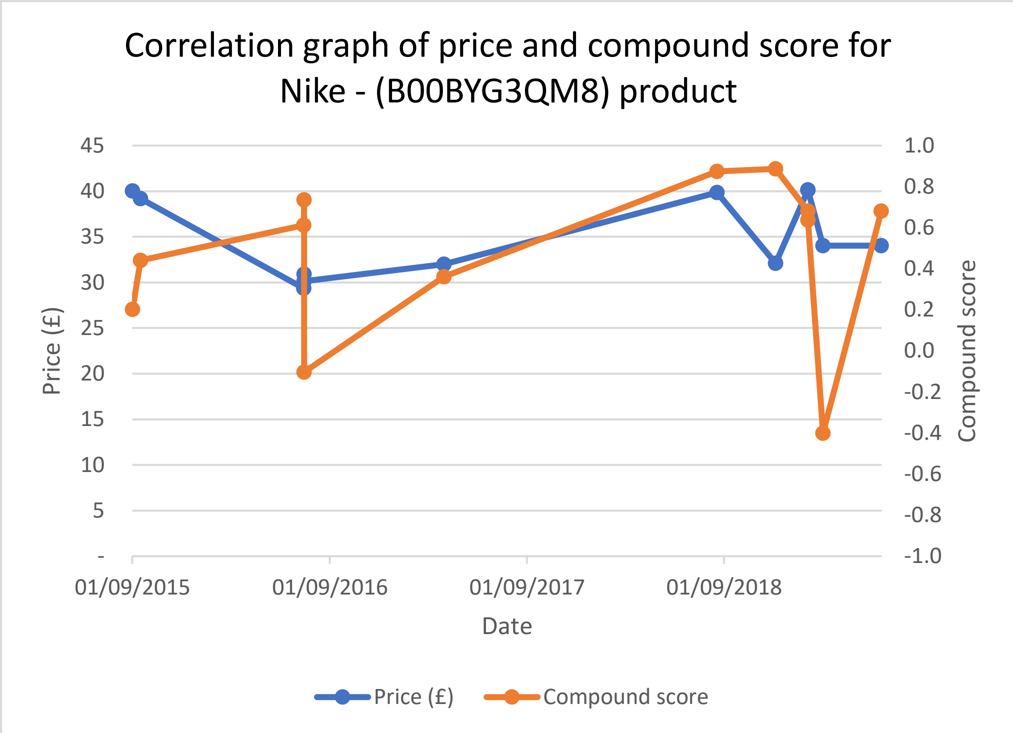


Figure A.5: Correlation graph of price and compound score for Nike B00BYG3QM8 product

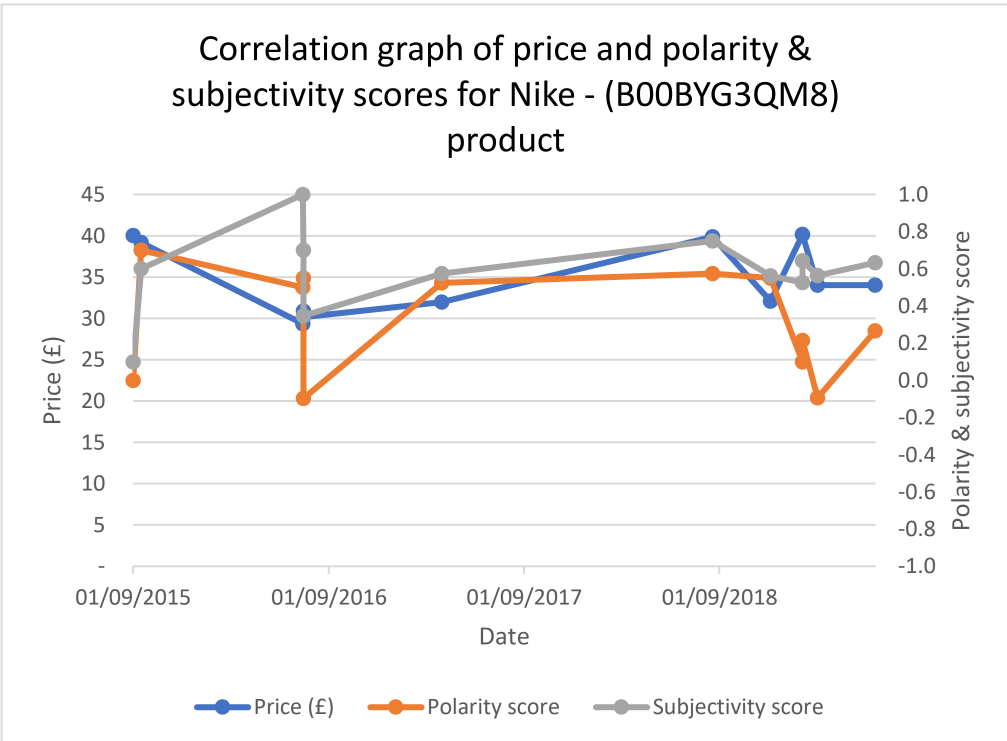


Figure A.6: Correlation graph of price and polarity & subjectivity scores for Nike B00BYG3QM8 product

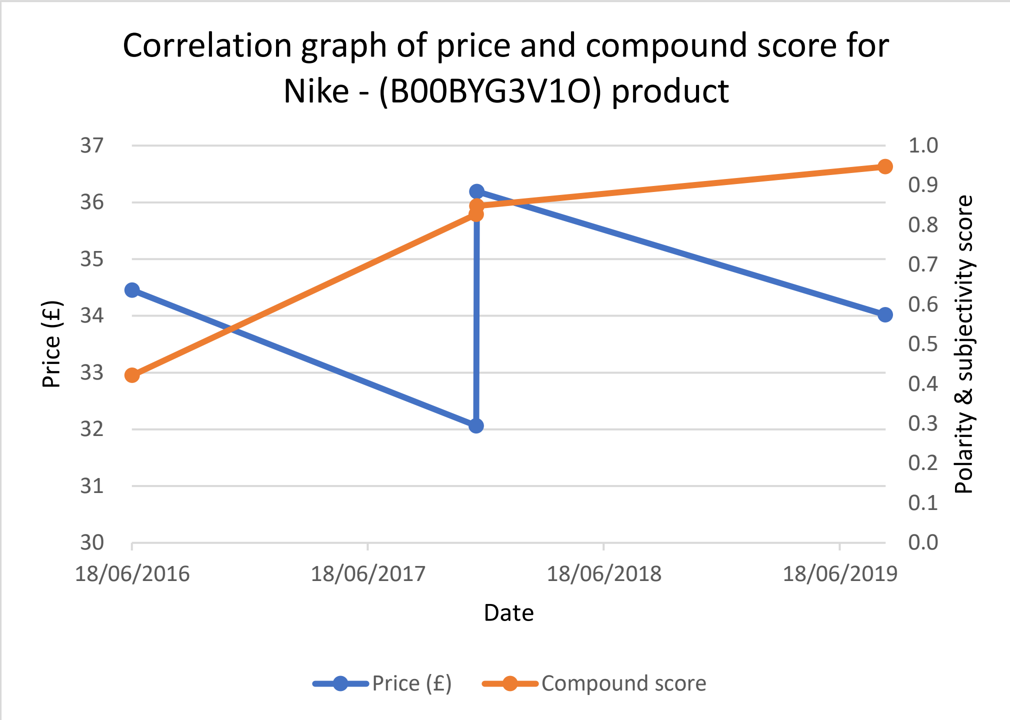


Figure A.7: Correlation graph of price and compound score for Nike B00BYG3V1O product

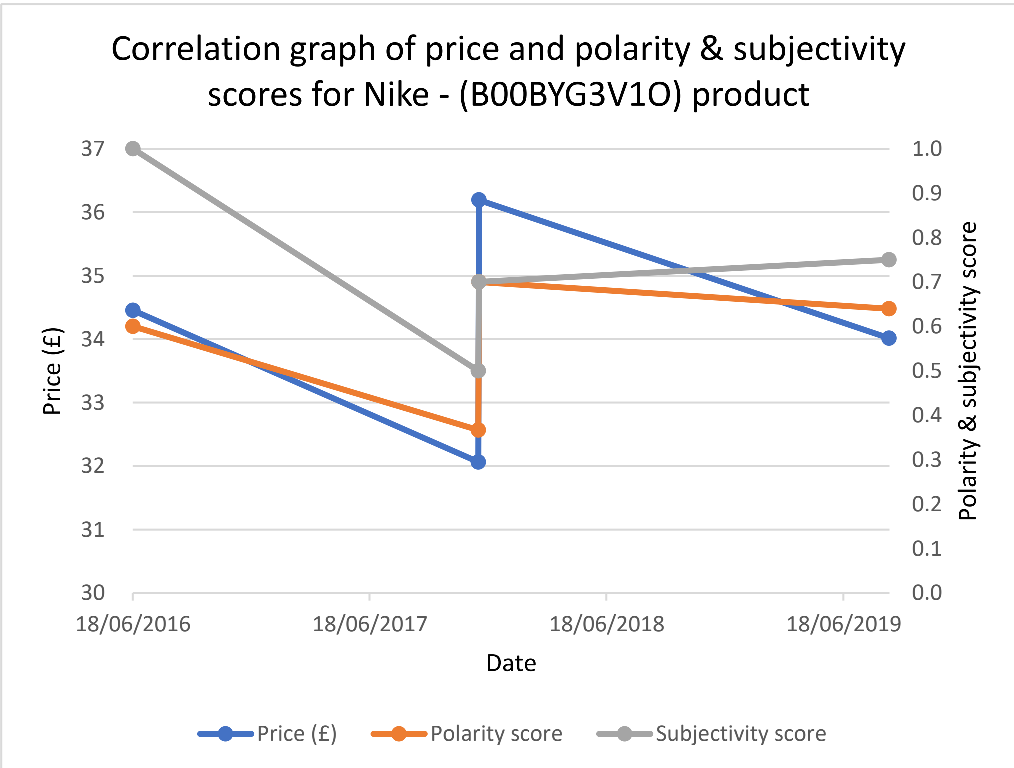


Figure A.8: Correlation graph of price and polarity & subjectivity scores for Nike B00BYG3V1O product

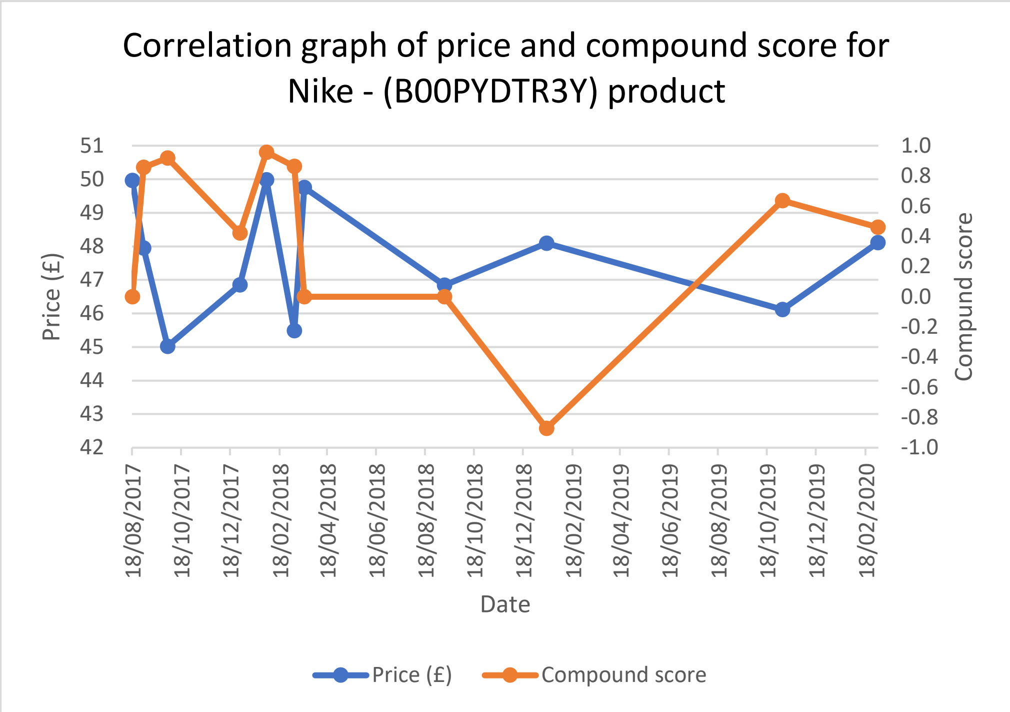


Figure A.9: Correlation graph of price and compound score for Nike B00PYDTR3Y product

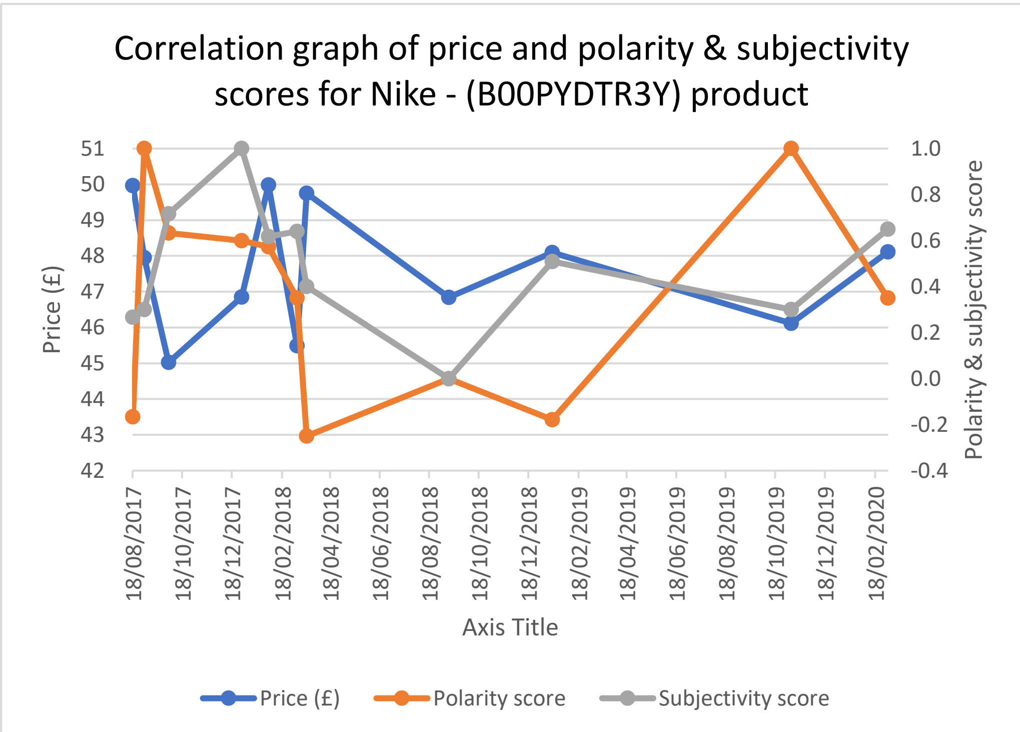


Figure A10: Correlation graph of price and polarity & subjectivity scores for Nike B00PYDTR3Y product

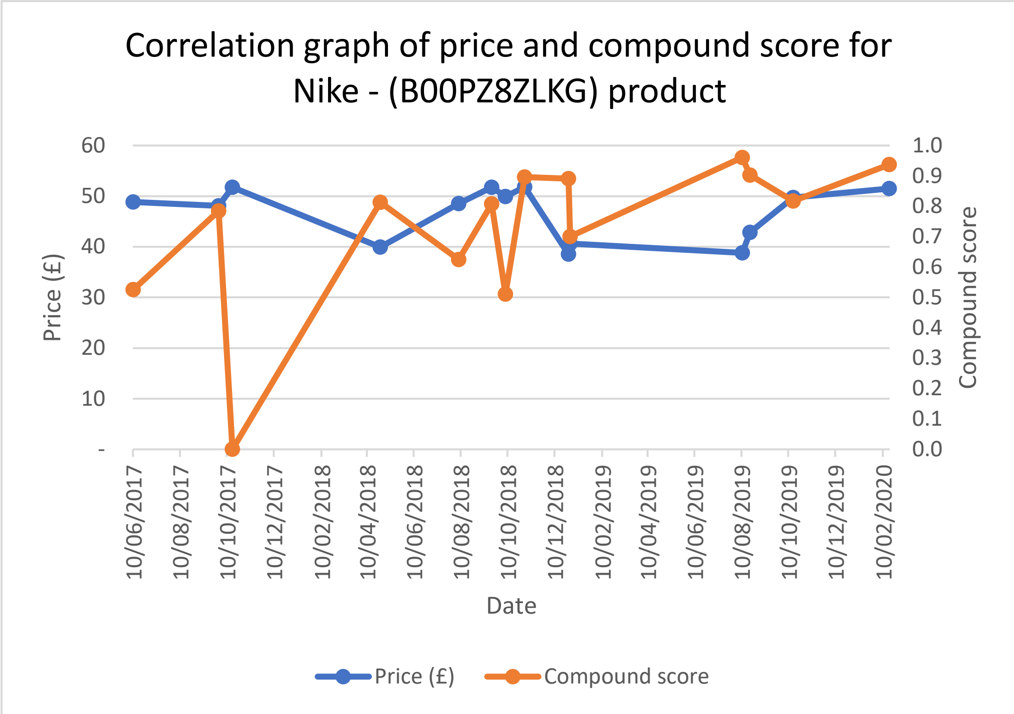


Figure A11: Correlation graph of price and compound score for Nike B00PZ8ZLKG product

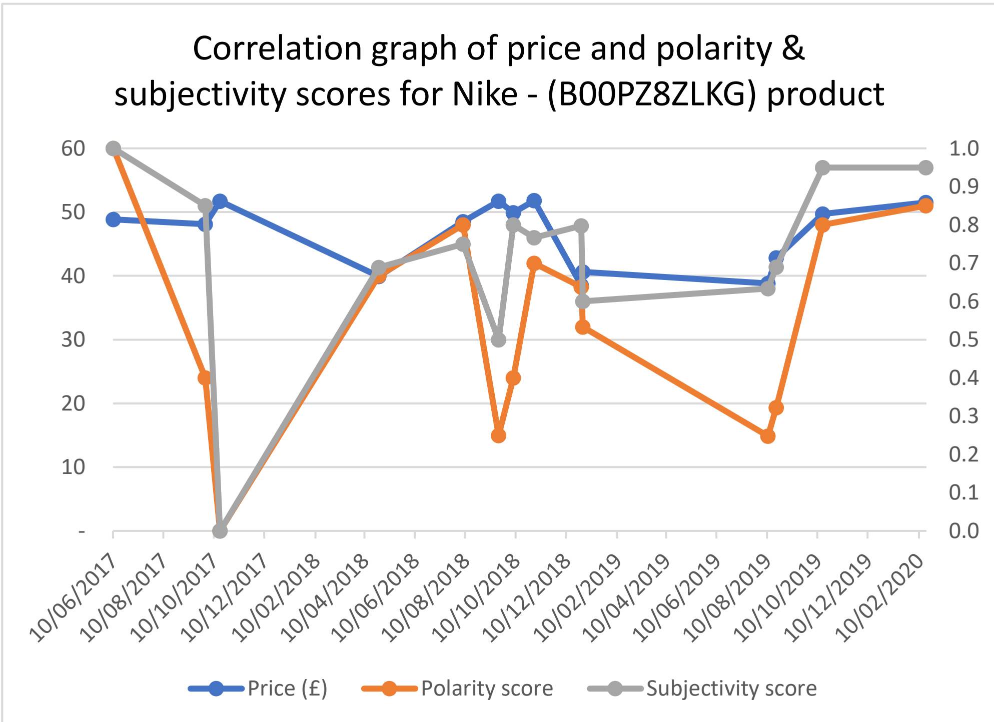


Figure A12: Correlation graph of price and polarity & subjectivity scores for Nike B00PZ8ZLKG product

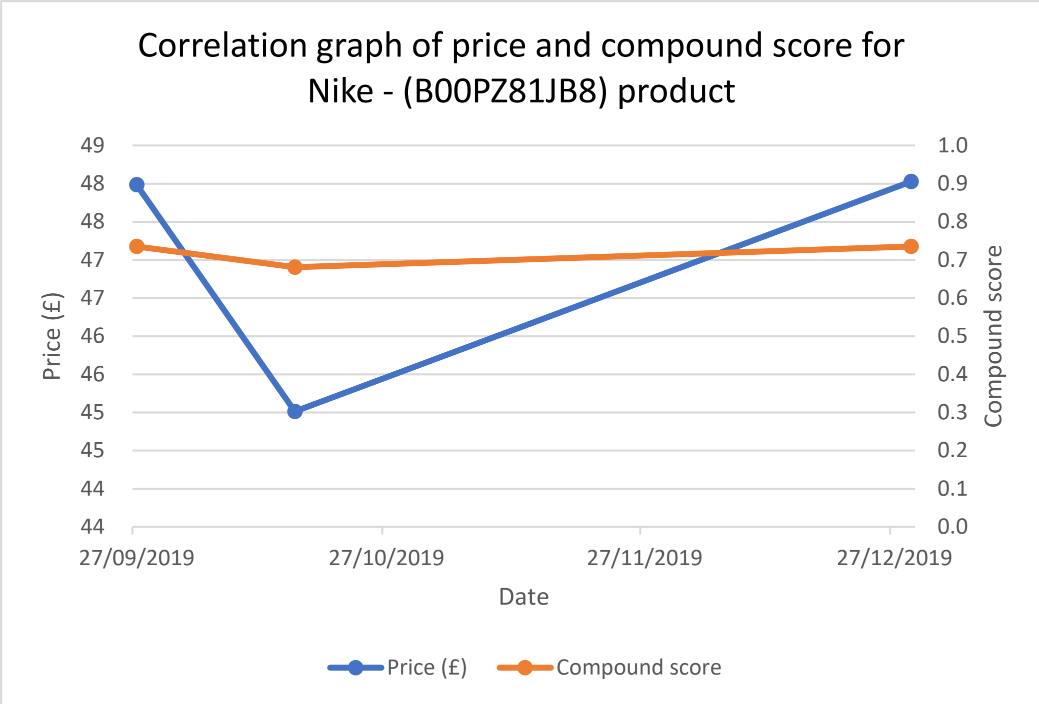


Figure A13: Correlation graph of price and compound score for Nike B00PZ81JB8 product

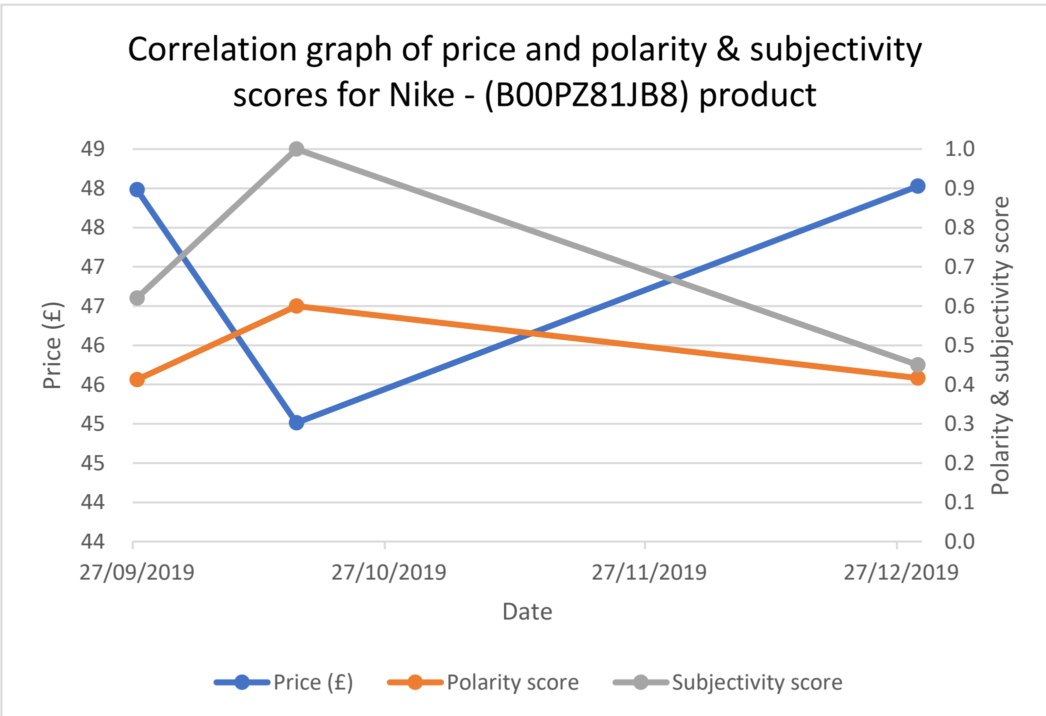


Figure A14: Correlation graph of price and polarity & subjectivity scores for Nike B00PZ81JB8 product

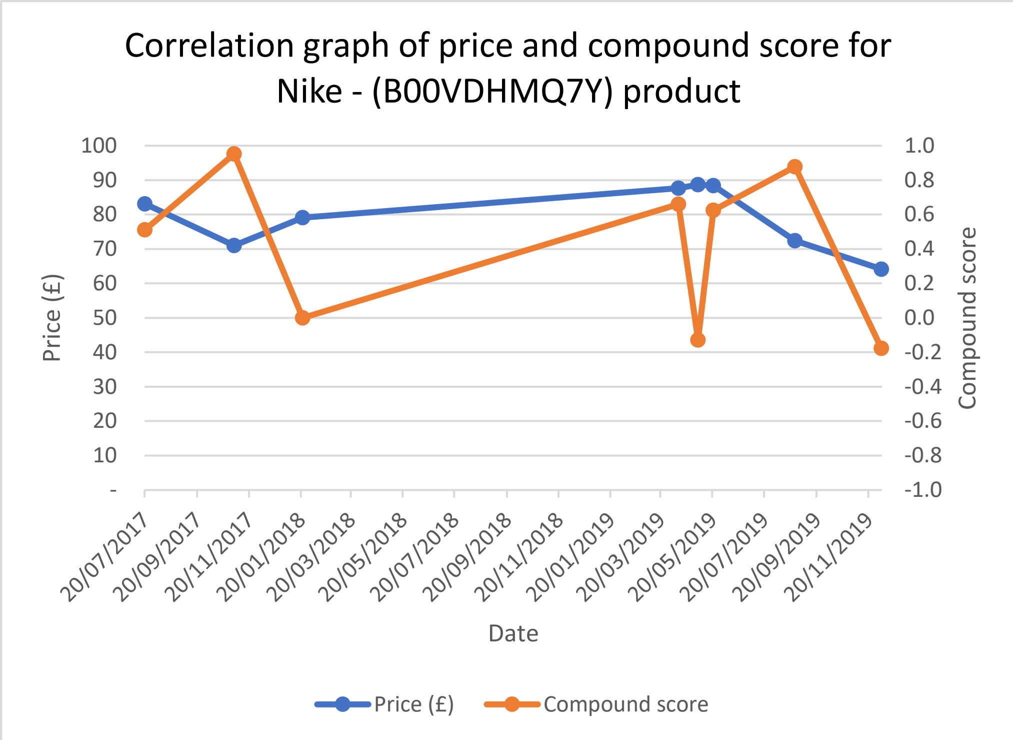


Figure A15: Correlation graph of price and compound score for Nike B00VDHMQ7Y product

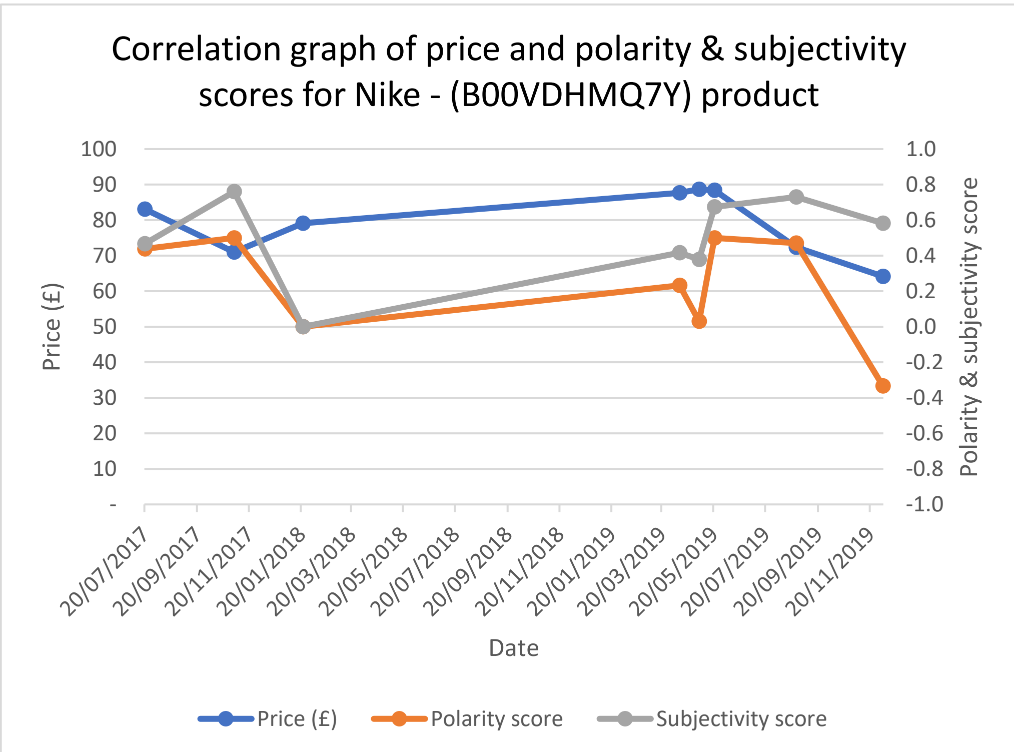


Figure A16: Correlation graph of price and polarity & subjectivity scores for Nike B00VDHMQ7Y product

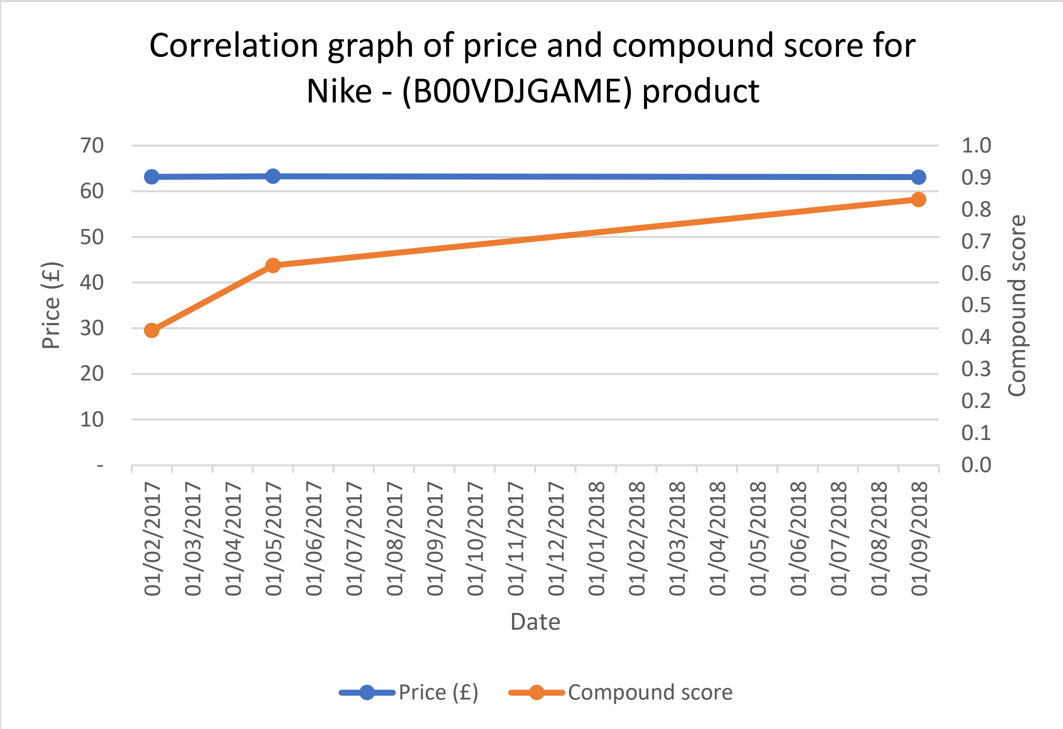


Figure A17: Correlation graph of price and compound score for Nike B00VDJGAME product

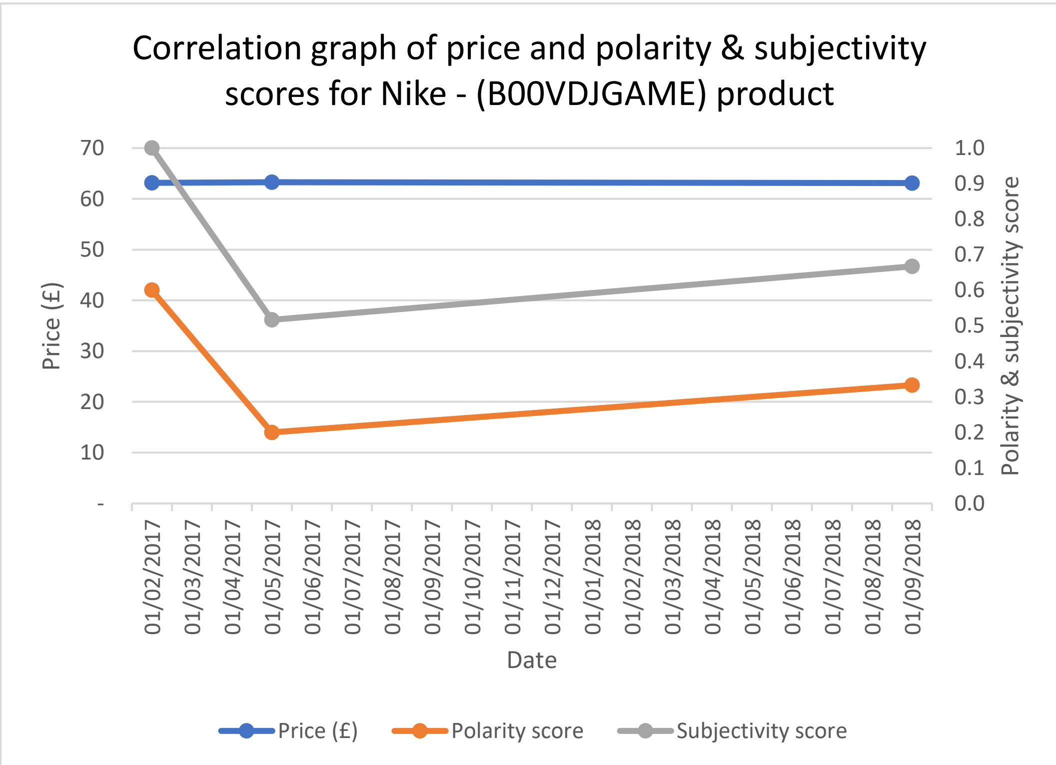


Figure A18: Correlation graph of price and polarity & subjectivity scores for Nike B00VDJGAME product

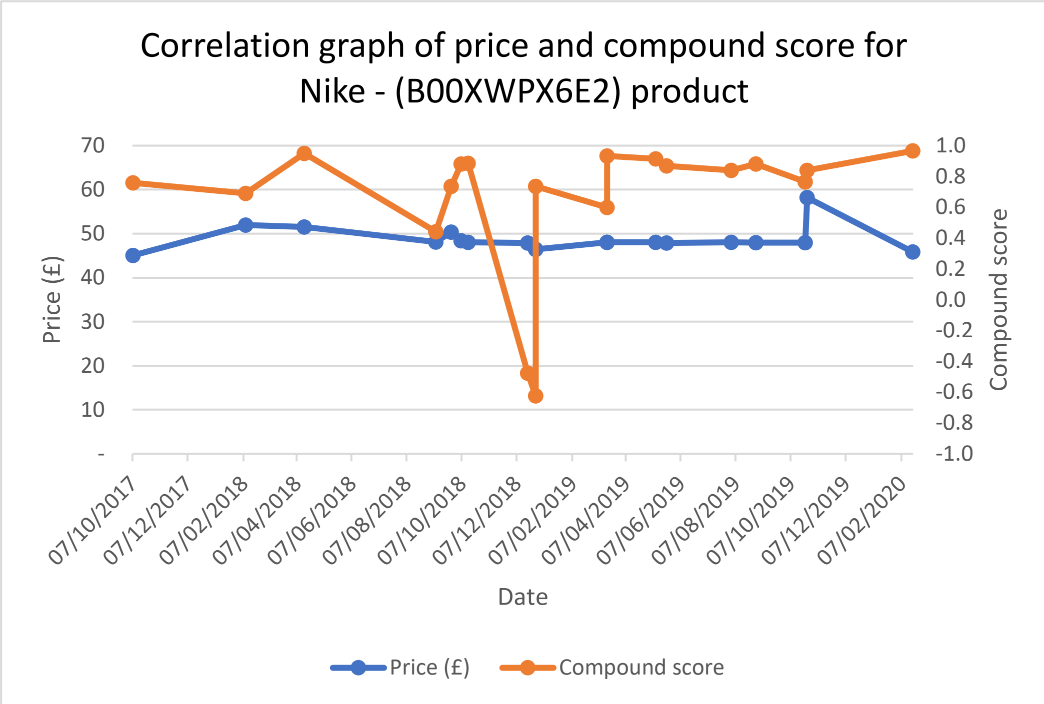


Figure A19: Correlation graph of price and compound score for Nike B00XWPX6E2 product

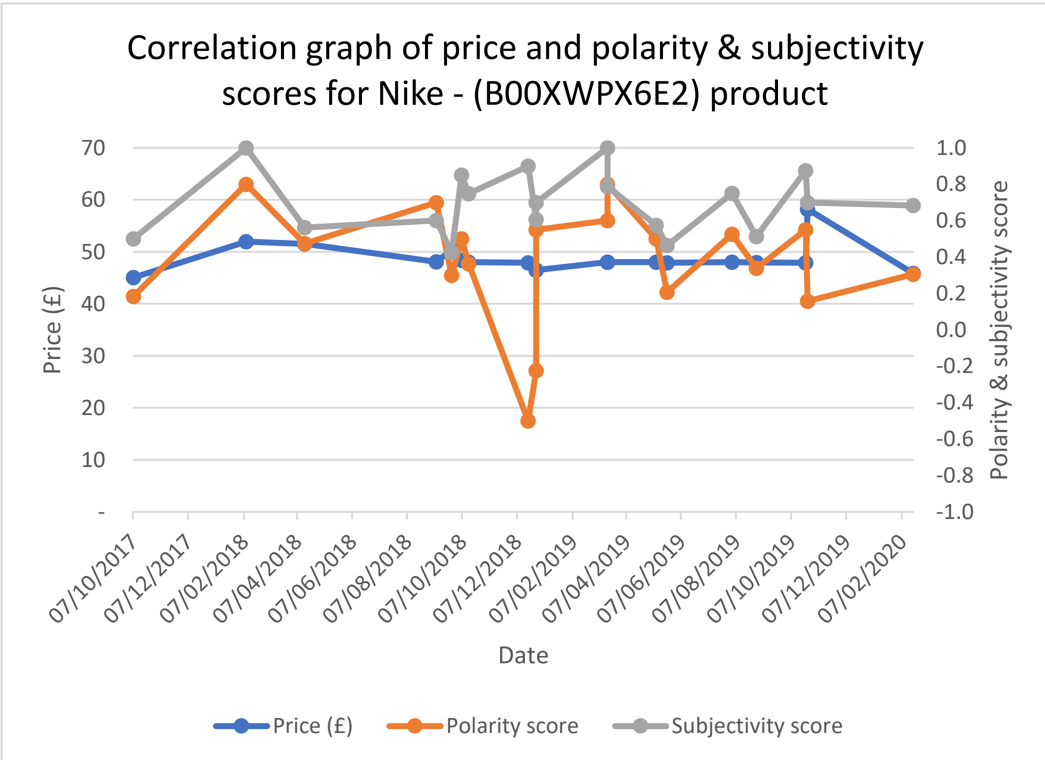


Figure A20: Correlation graph of price and polarity & subjectivity scores for Nike B00XWPX6E2 product

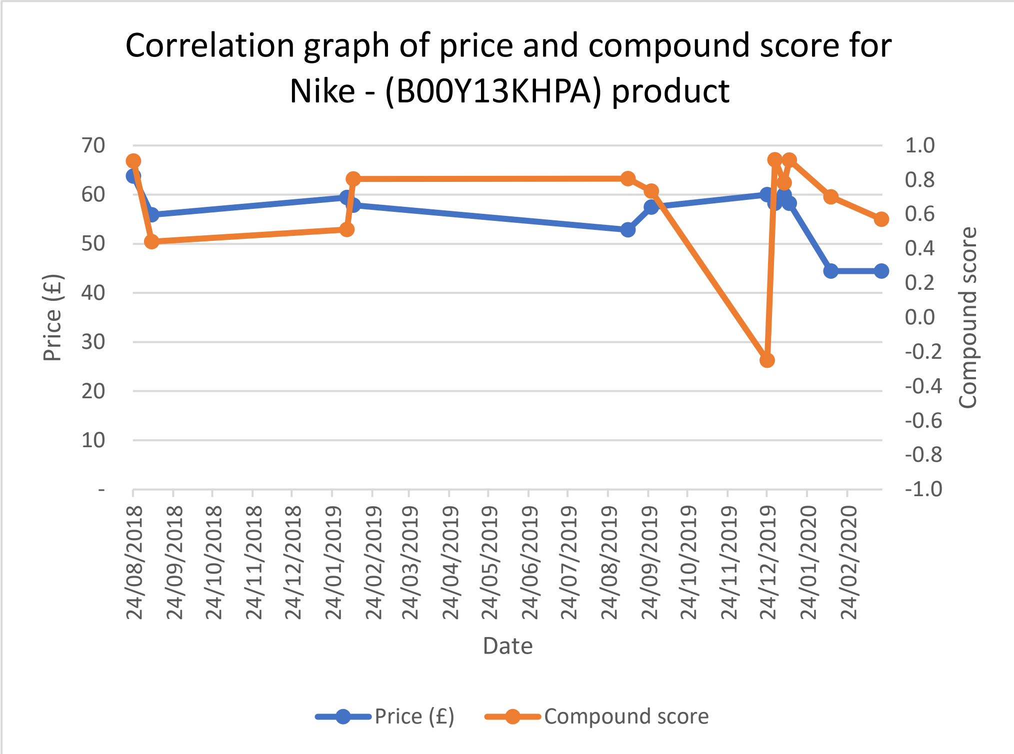


Figure A21: Correlation graph of price and compound score for Nike B00Y13KHPA product

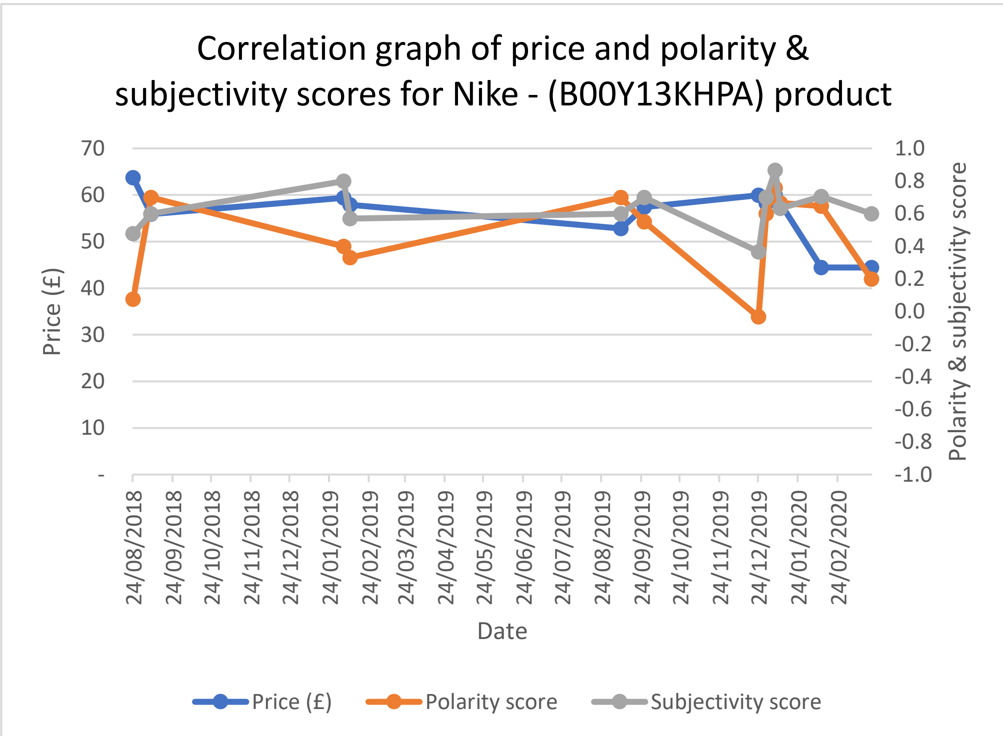


Figure A22: Correlation graph of price and polarity & subjectivity scores for Nike B00Y13KHPA product

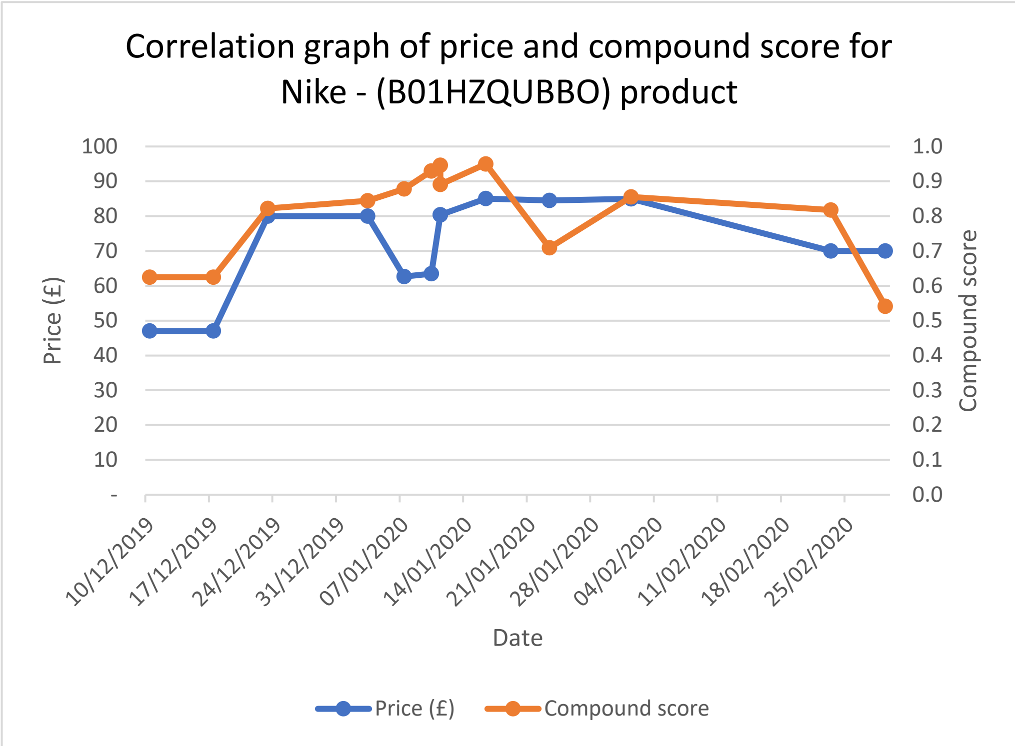


Figure A23: Correlation graph of price and compound score for Nike B01HZQUBBO product

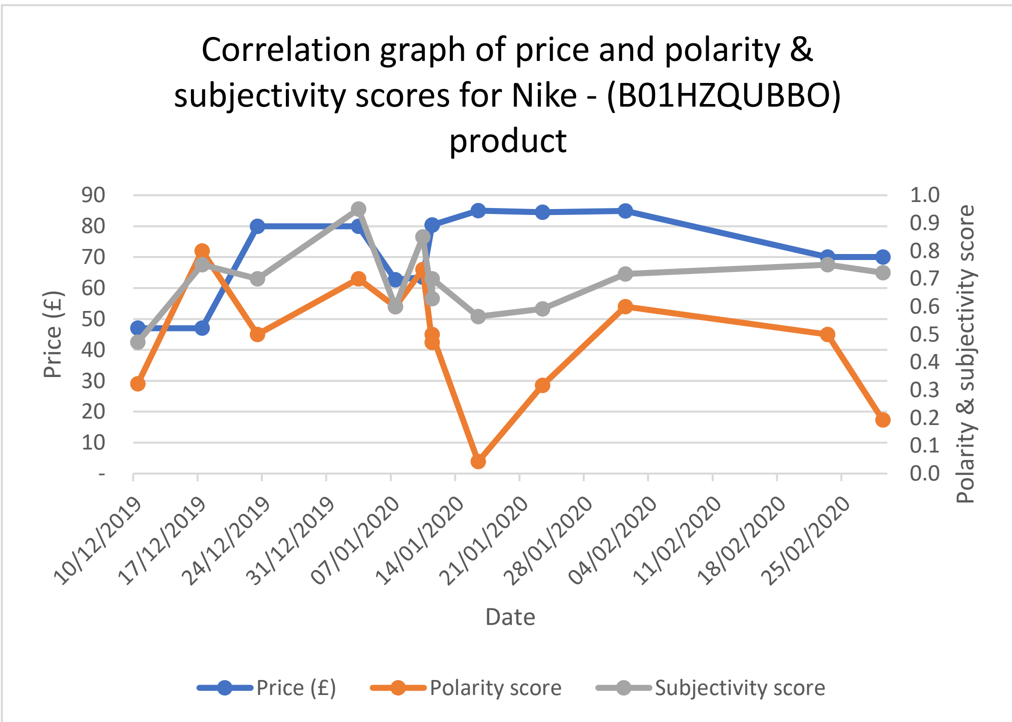


Figure A24: Correlation graph of price and polarity & subjectivity scores for Nike B01HZQUBBO product

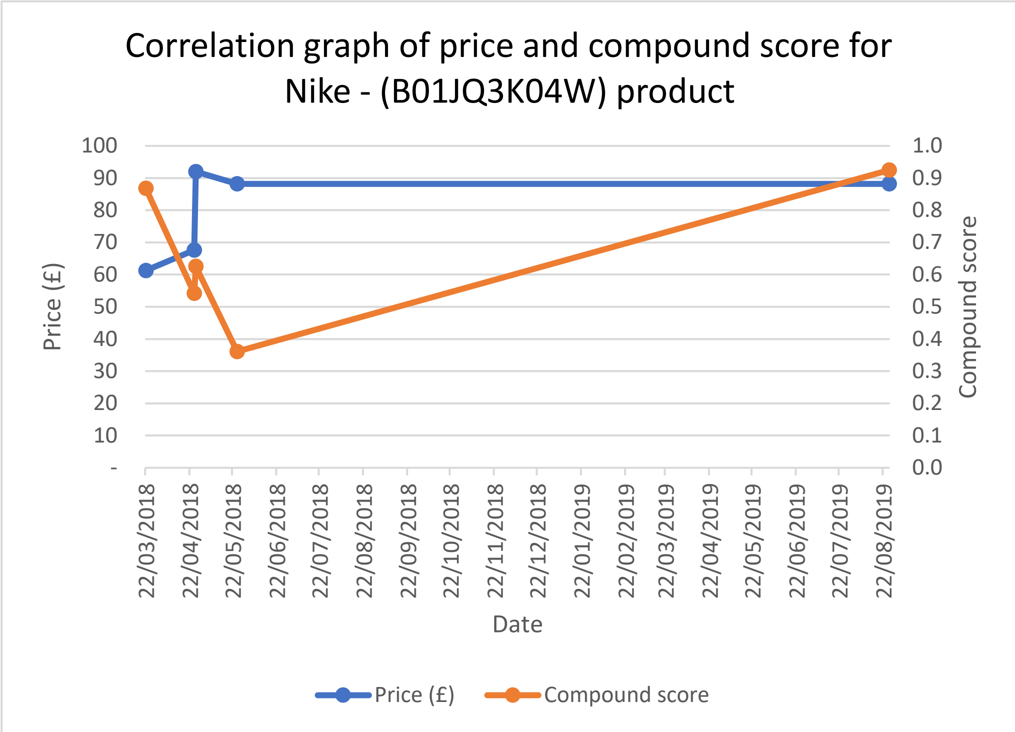


Figure A25: Correlation graph of price and compound score for Nike B01JQ3K04W product

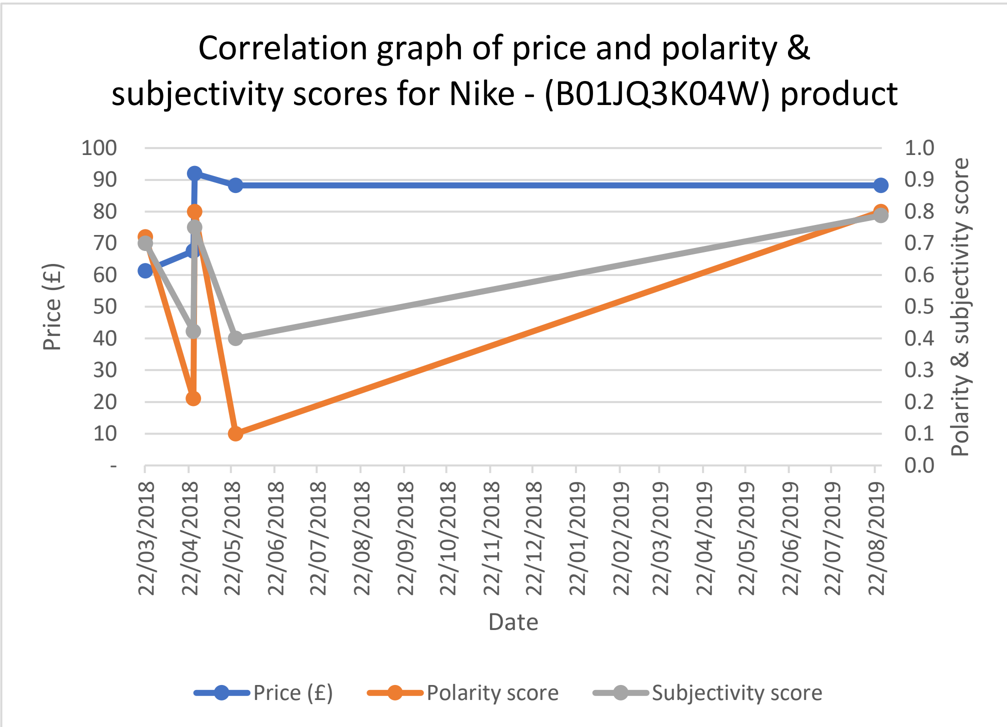


Figure A26: Correlation graph of price and polarity & subjectivity scores for Nike B01JQ3K04W product

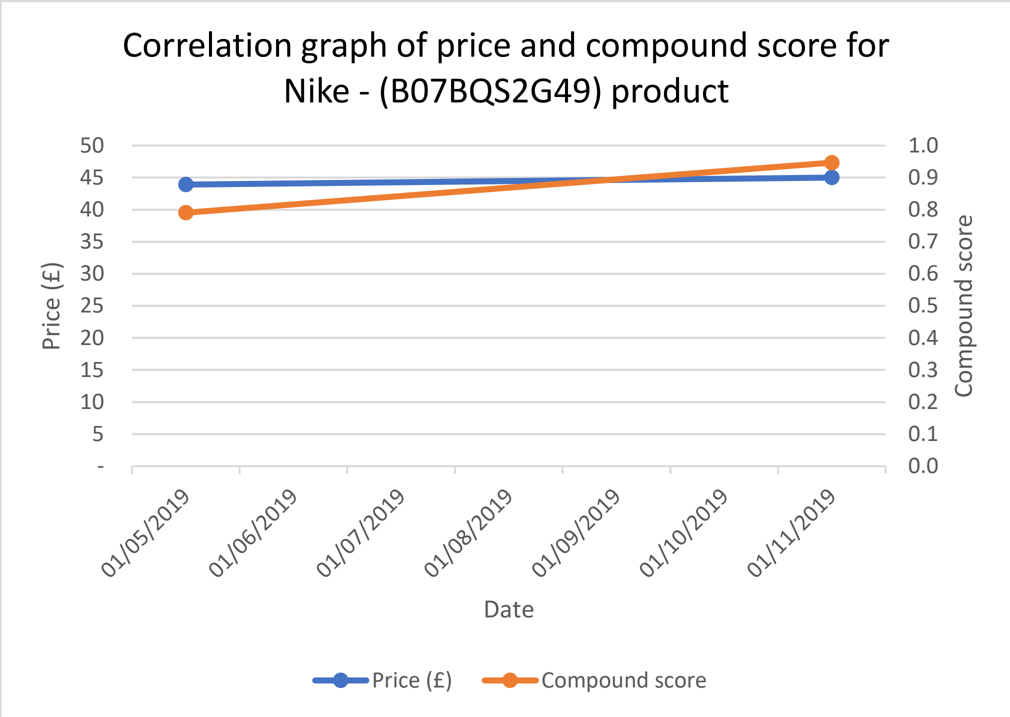


Figure A27: Correlation graph of price and compound score for Nike B07BQS2G49 product

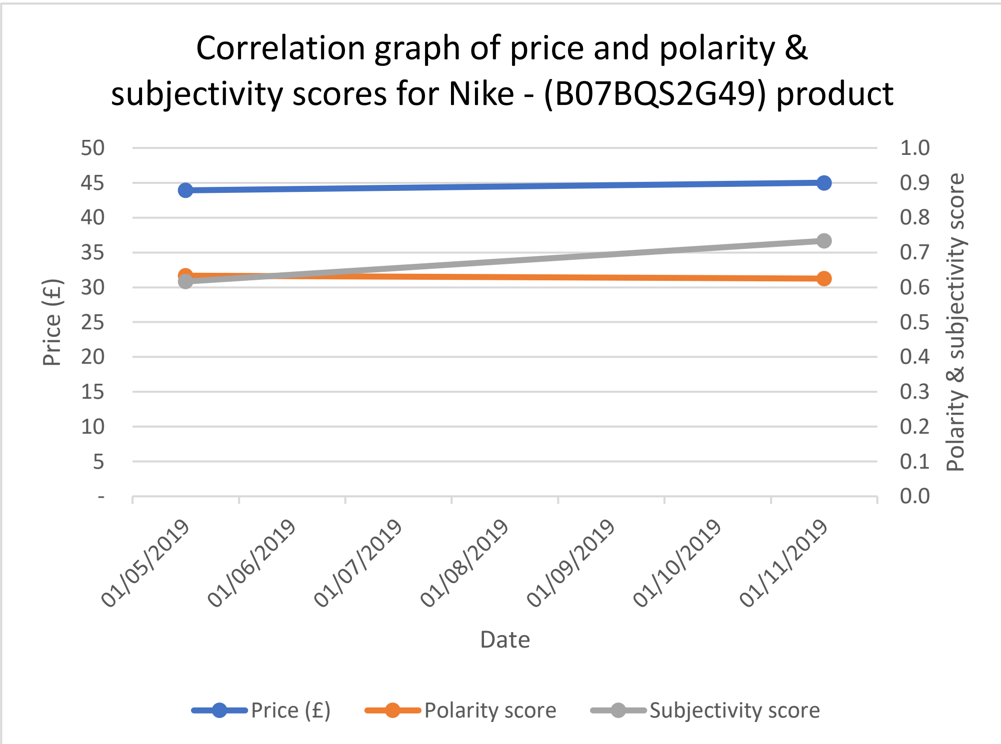


Figure A28: Correlation graph of price and polarity & subjectivity scores for Nike B07BQS2G49 product

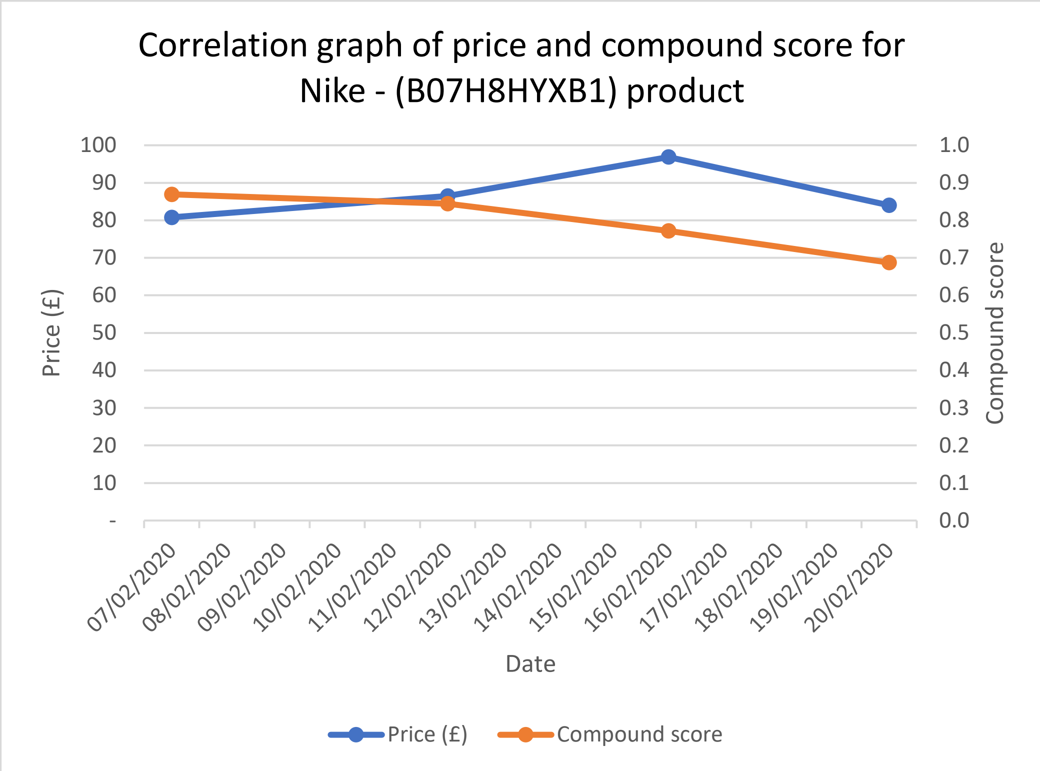


Figure A29: Correlation graph of price and compound score for Nike B07H8HYXB1 product

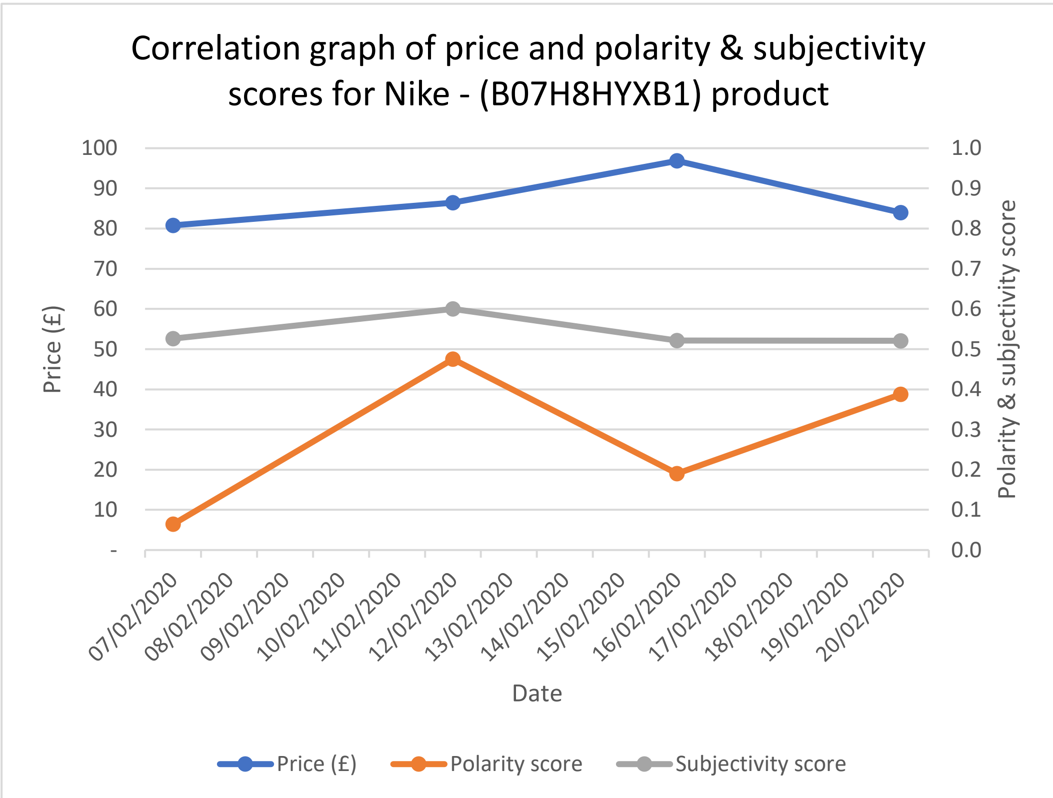


Figure A30: Correlation graph of price and polarity & subjectivity scores for Nike B07H8HYXB1 product

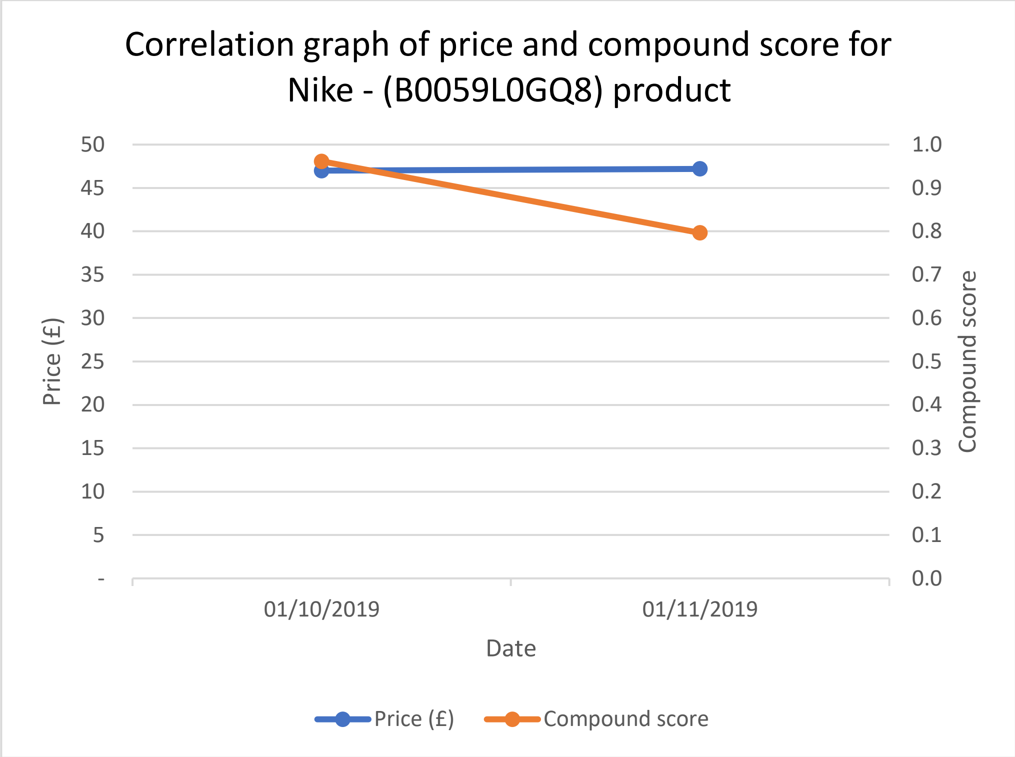


Figure A31: Correlation graph of price and compound score for Nike B0059L0GQ8 product

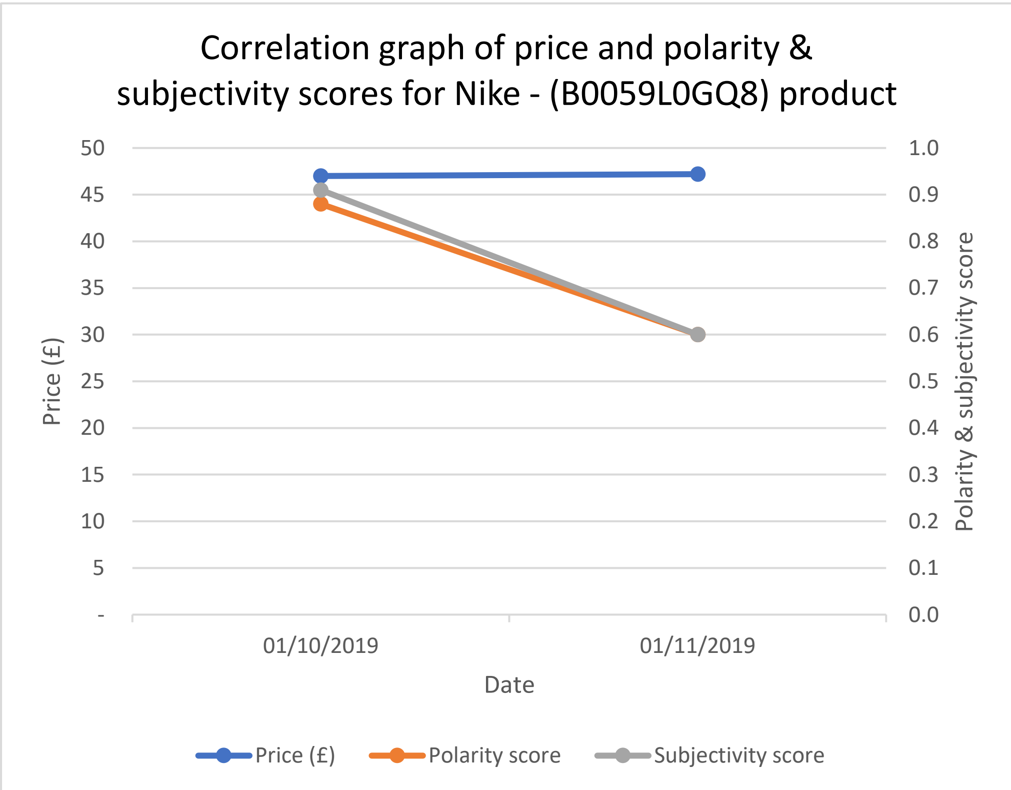


Figure A32: Correlation graph of price and polarity & subjectivity scores for Nike B0059L0GQ8 product

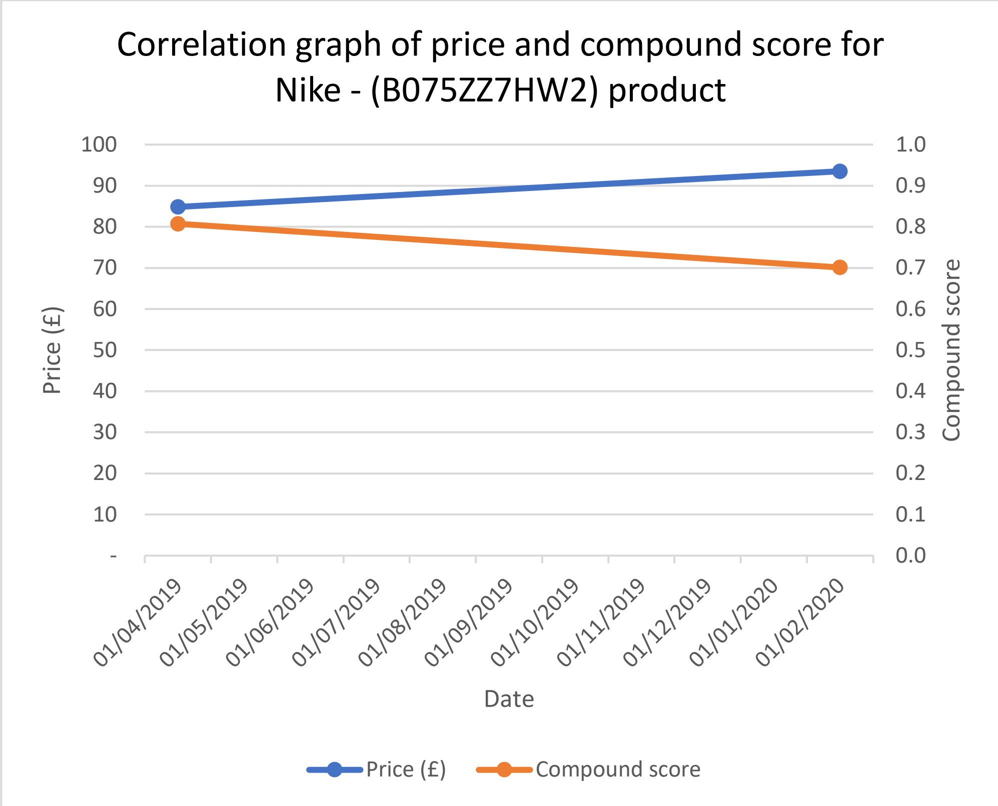


Figure A33: Correlation graph of price and compound score for Nike B075ZZ7HW2 product

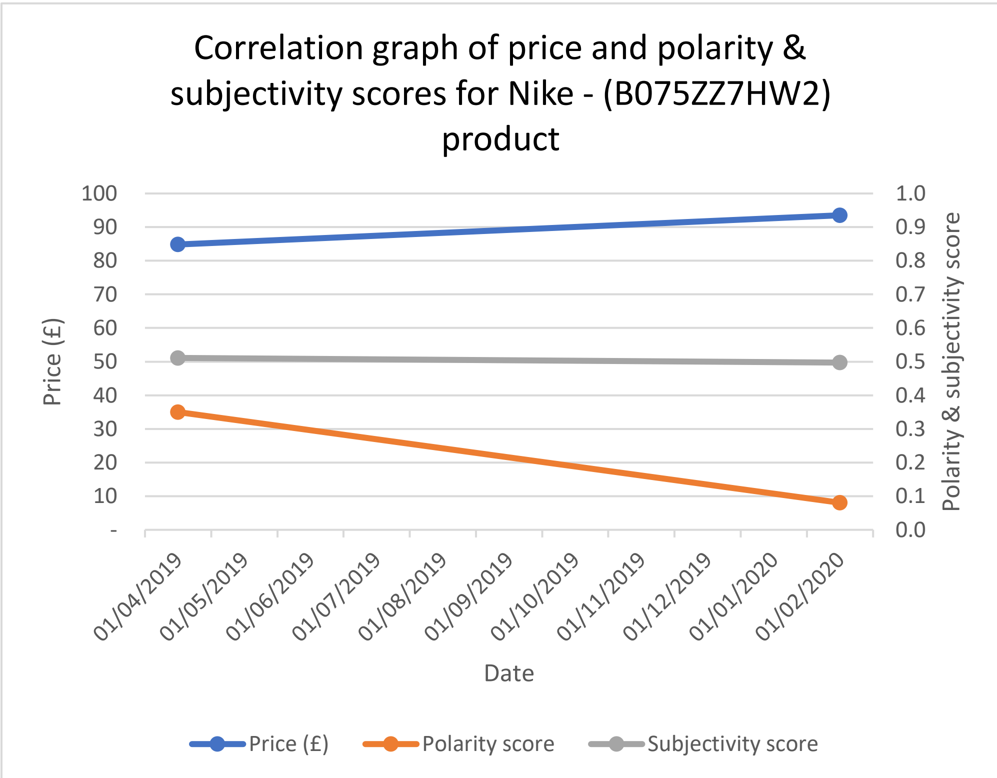


Figure A34: Correlation graph of price and polarity & subjectivity scores for Nike B075ZZ7HW2 product

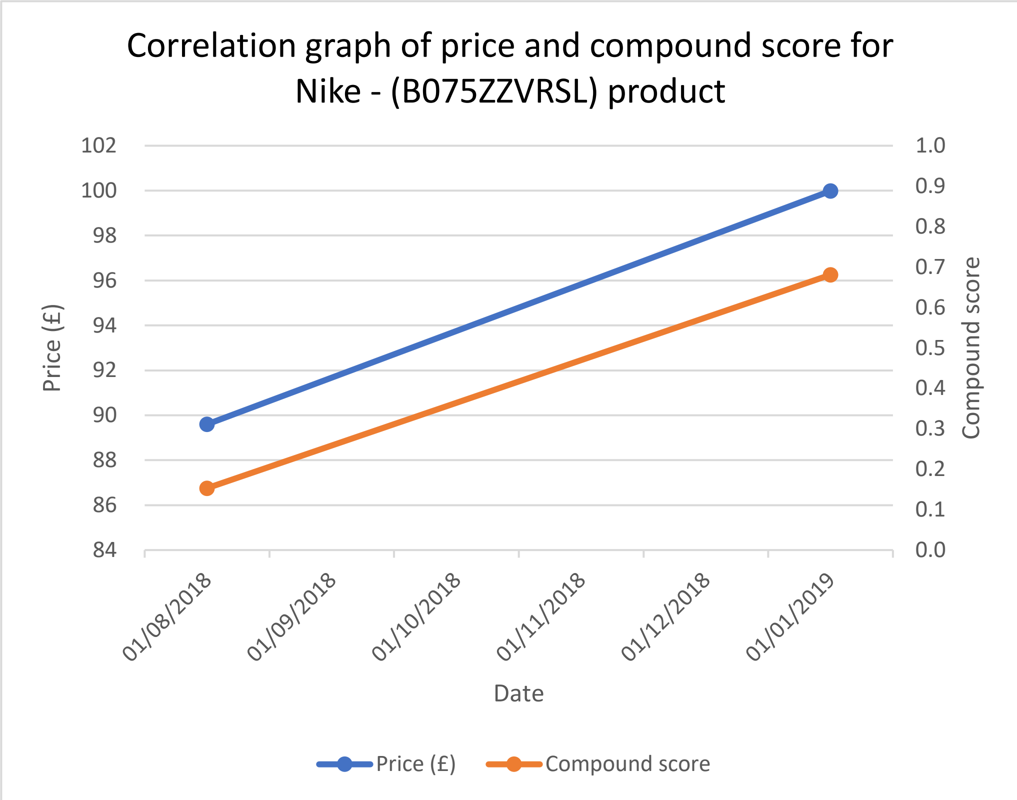


Figure A35: Correlation graph of price and compound score for Nike B075ZZVRSL product

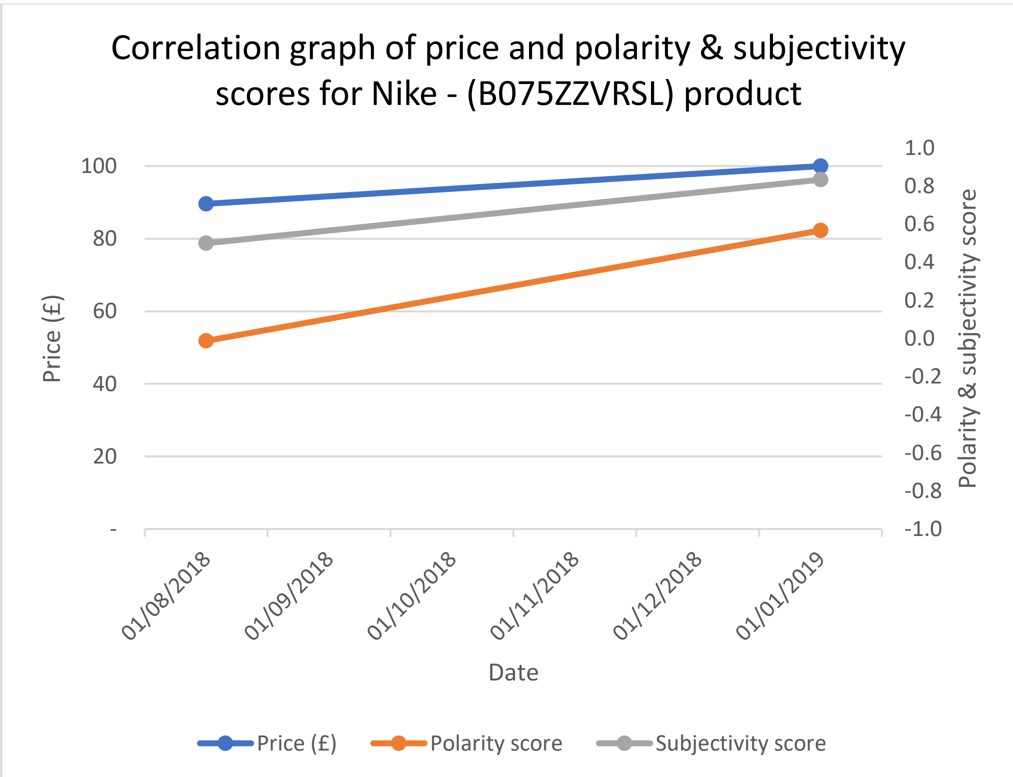


Figure A36: Correlation graph of price and polarity & subjectivity scores for Nike B075ZZVRSL product

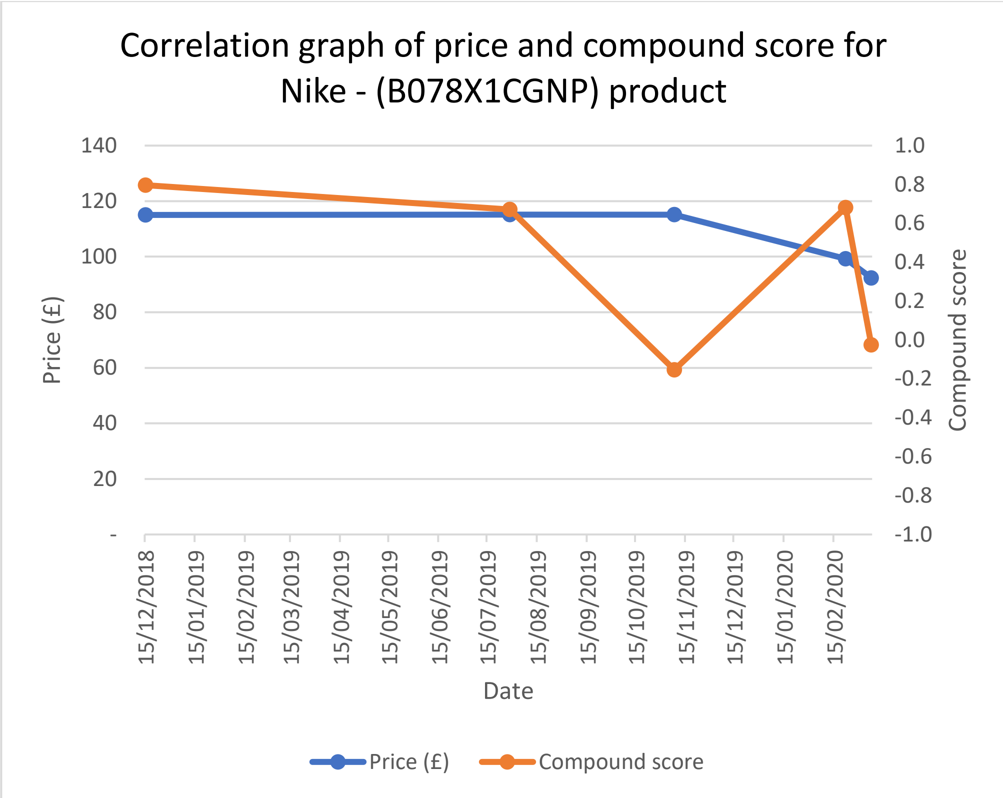


Figure A37: Correlation graph of price and compound score for Nike B078X1CGNP product

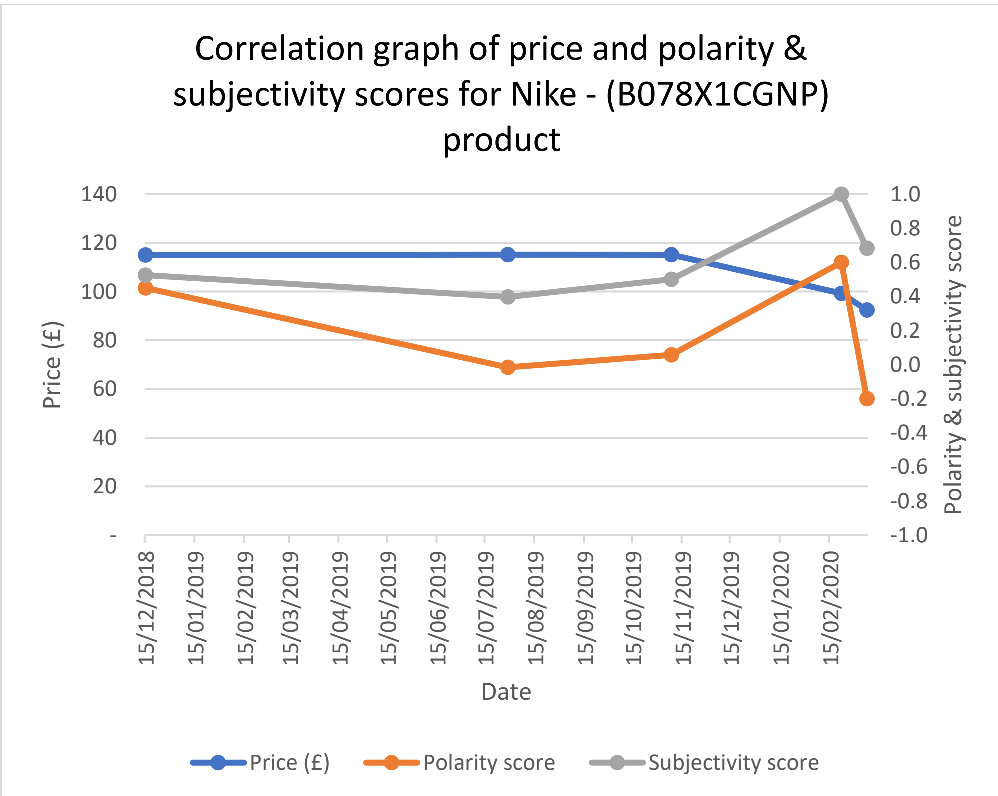


Figure A38: Correlation graph of price and polarity & subjectivity scores for Nike B078X1CGNP product

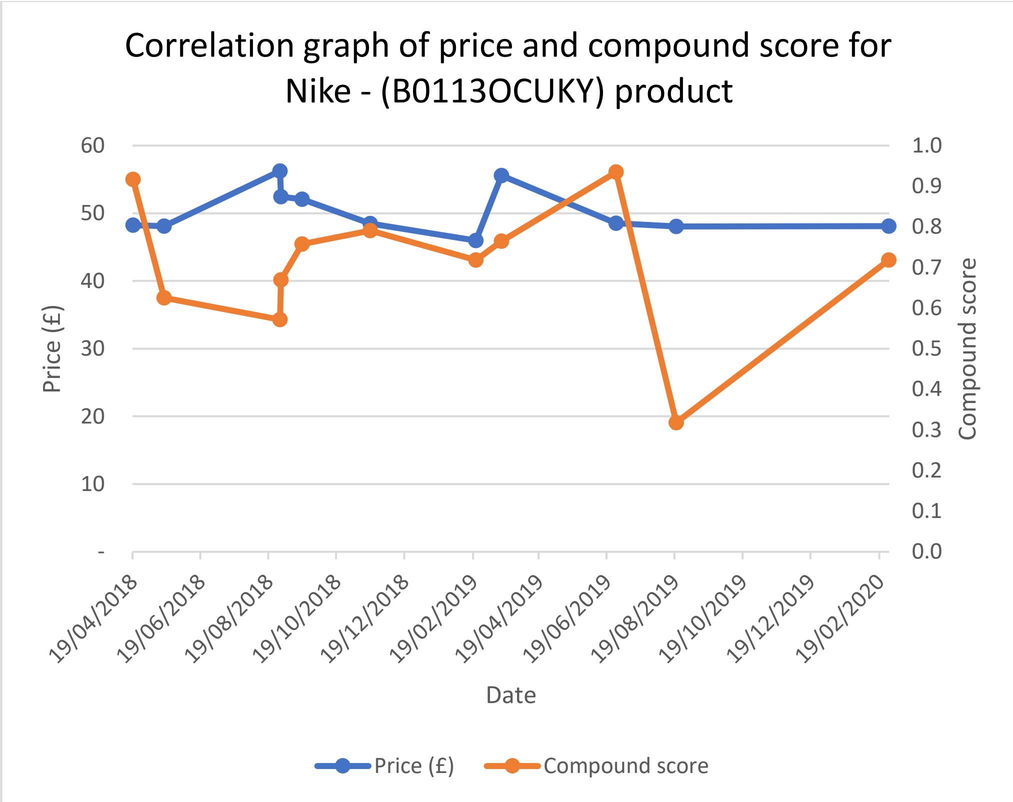


Figure A39: Correlation graph of price and compound score for Nike B0113OCUKY product

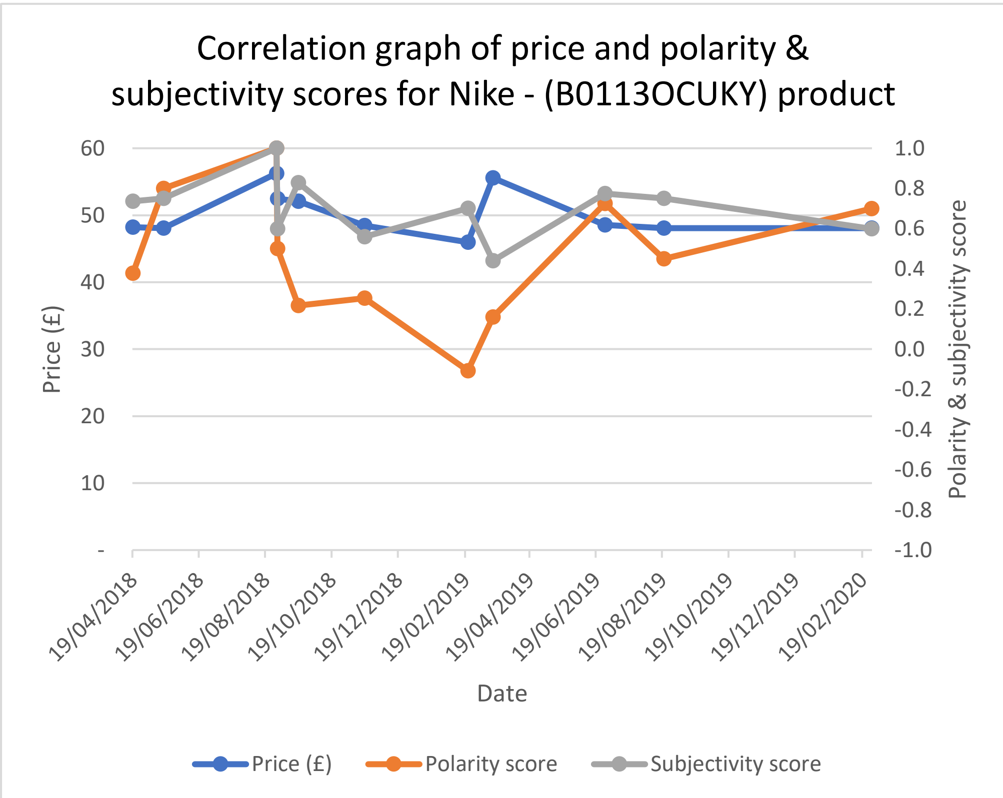


Figure A40: Correlation graph of price and polarity & subjectivity scores for Nike B0113OCUKY product

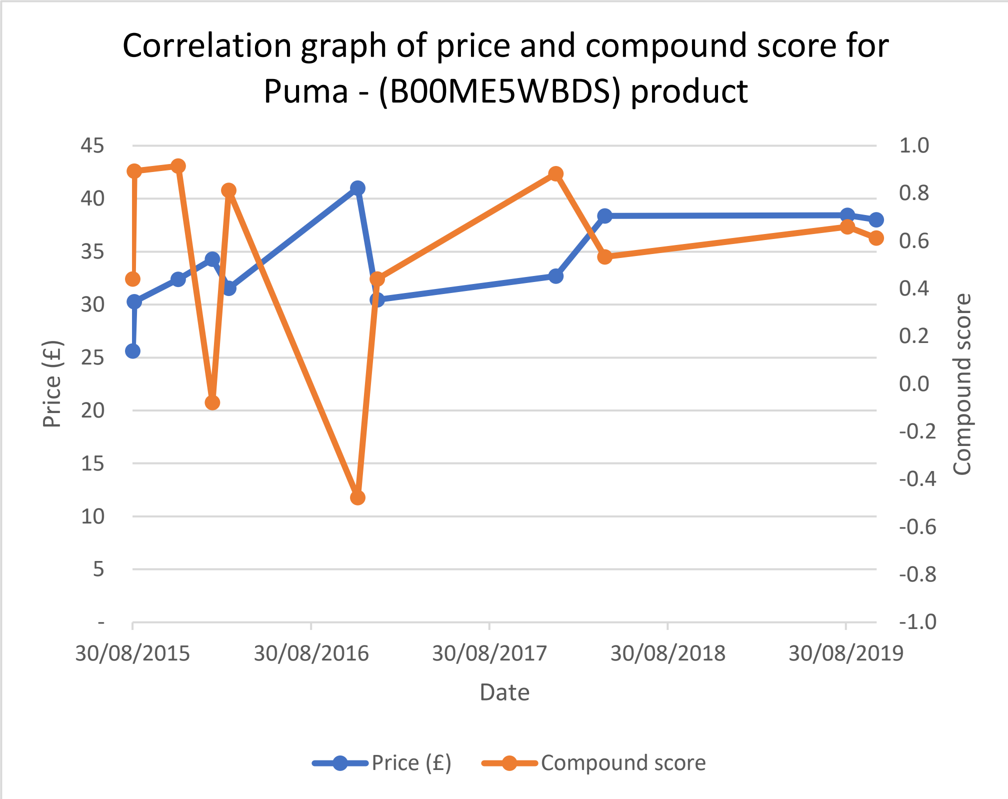


Figure A41: Correlation graph of price and compound score for Puma B00ME5WBDS product

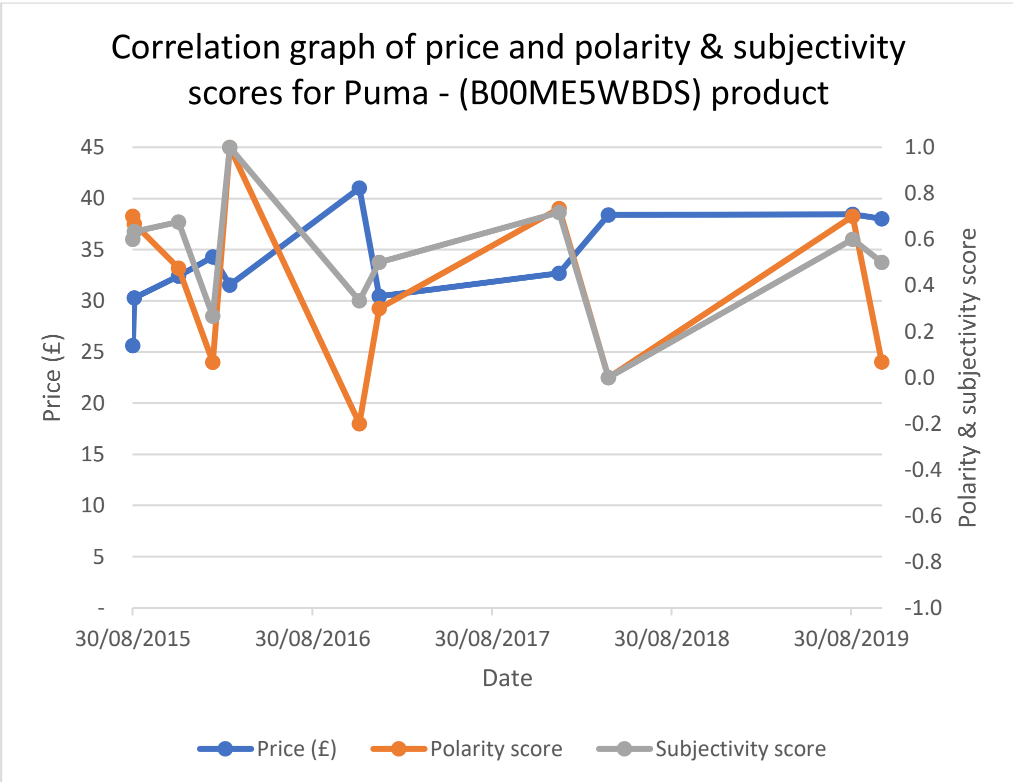


Figure A42: Correlation graph of price and polarity & subjectivity scores for Puma B00ME5WBDS product

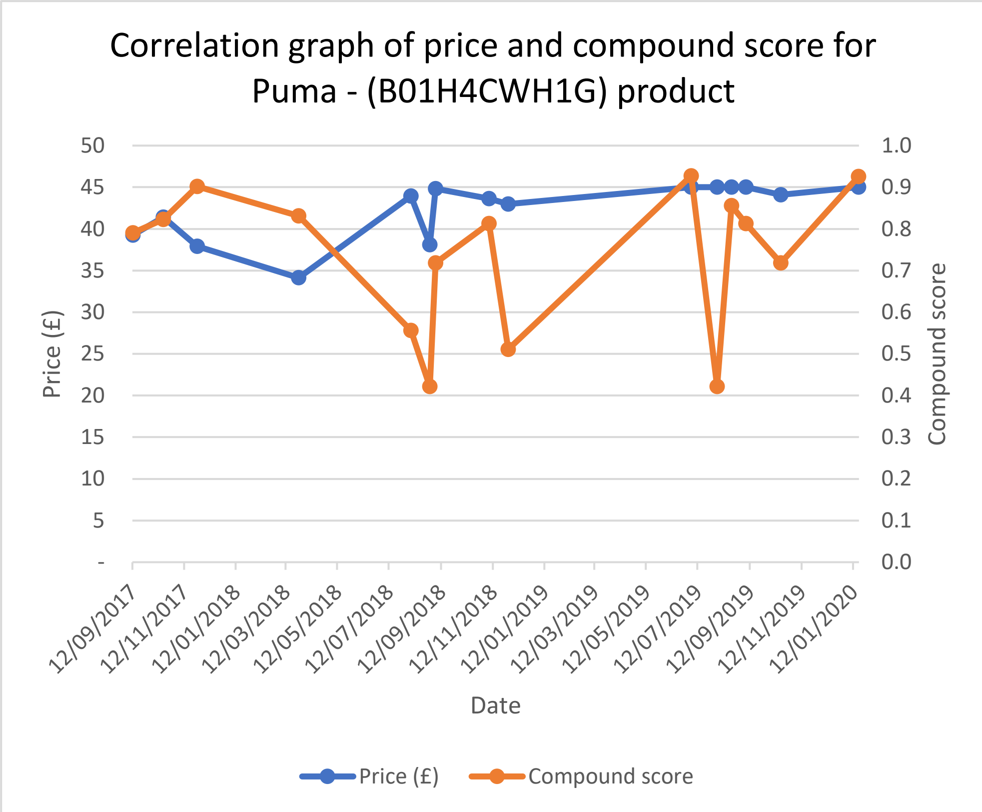


Figure A43: Correlation graph of price and compound score for Puma B01H4CWH1G product

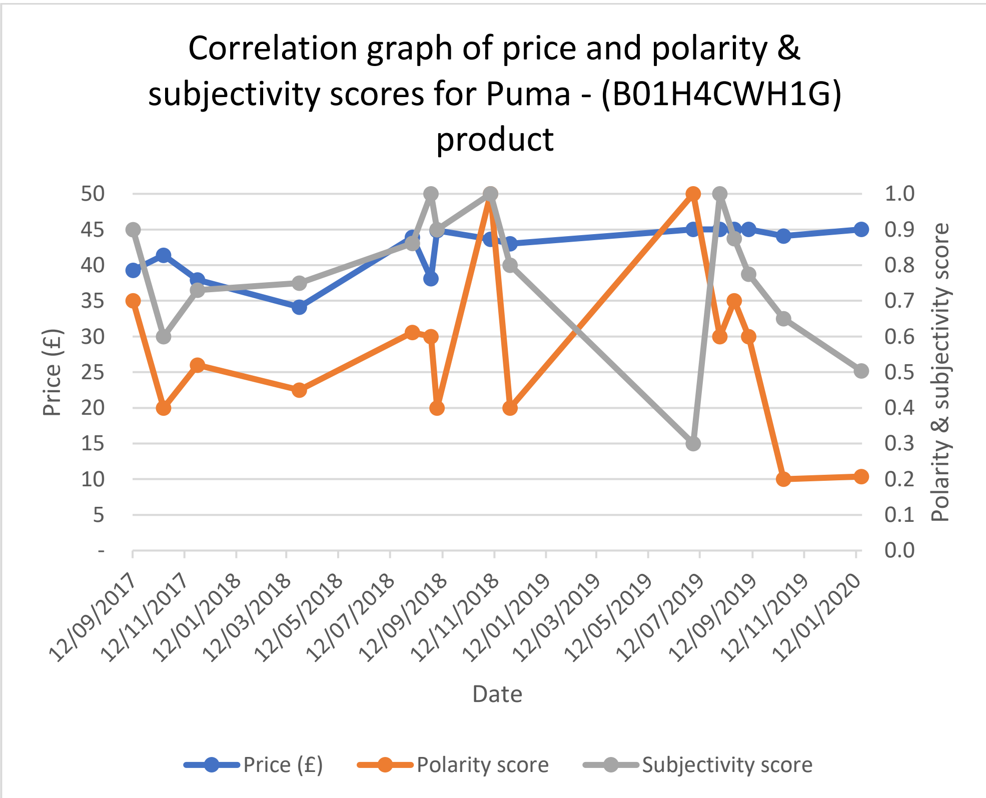


Figure A44: Correlation graph of price and polarity & subjectivity scores for Puma B01H4CWH1G product

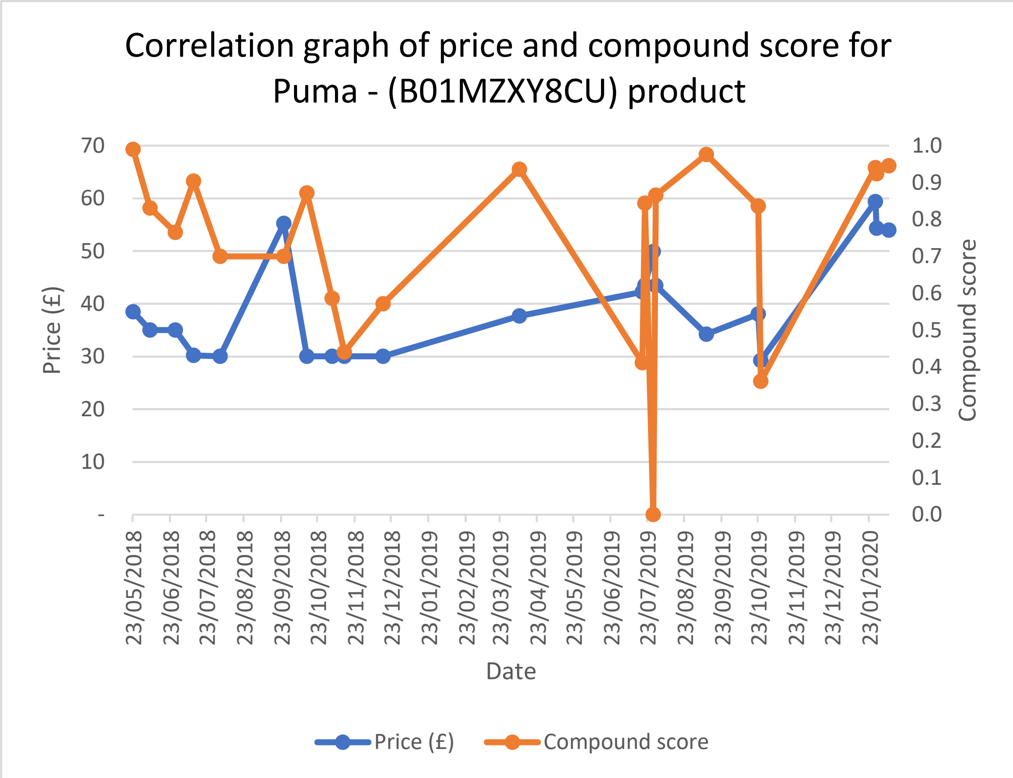


Figure A45: Correlation graph of price and compound score for Puma B01MZXY8CU product

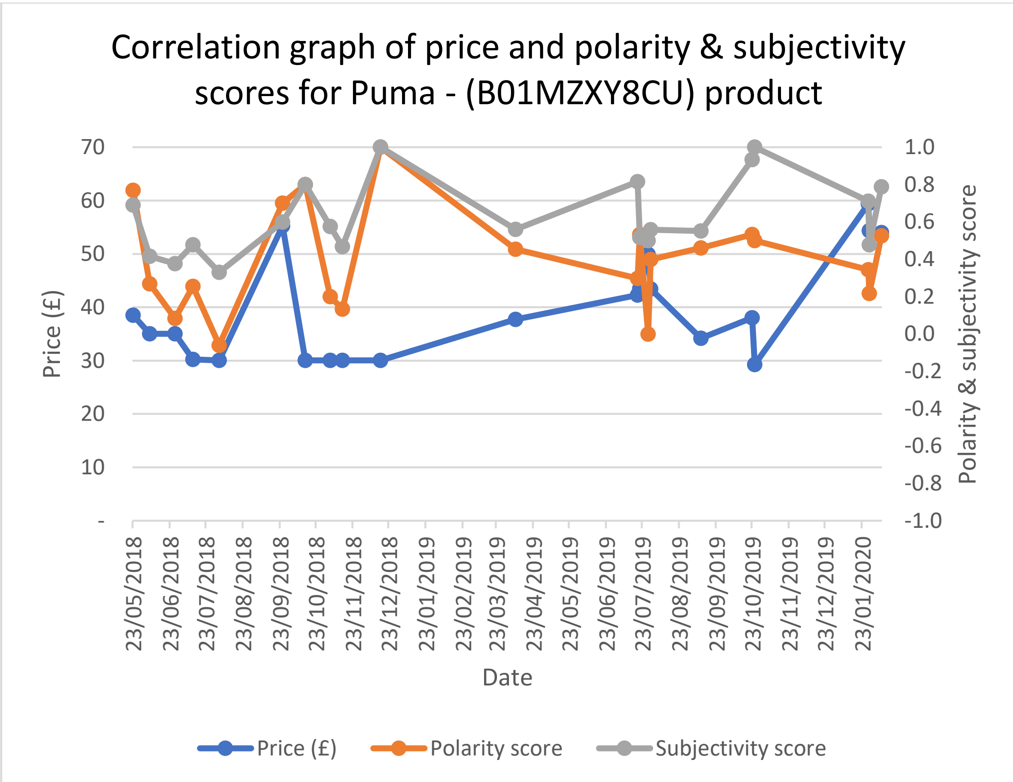


Figure A46: Correlation graph of price and polarity & subjectivity scores for Puma B01MZXY8CU product

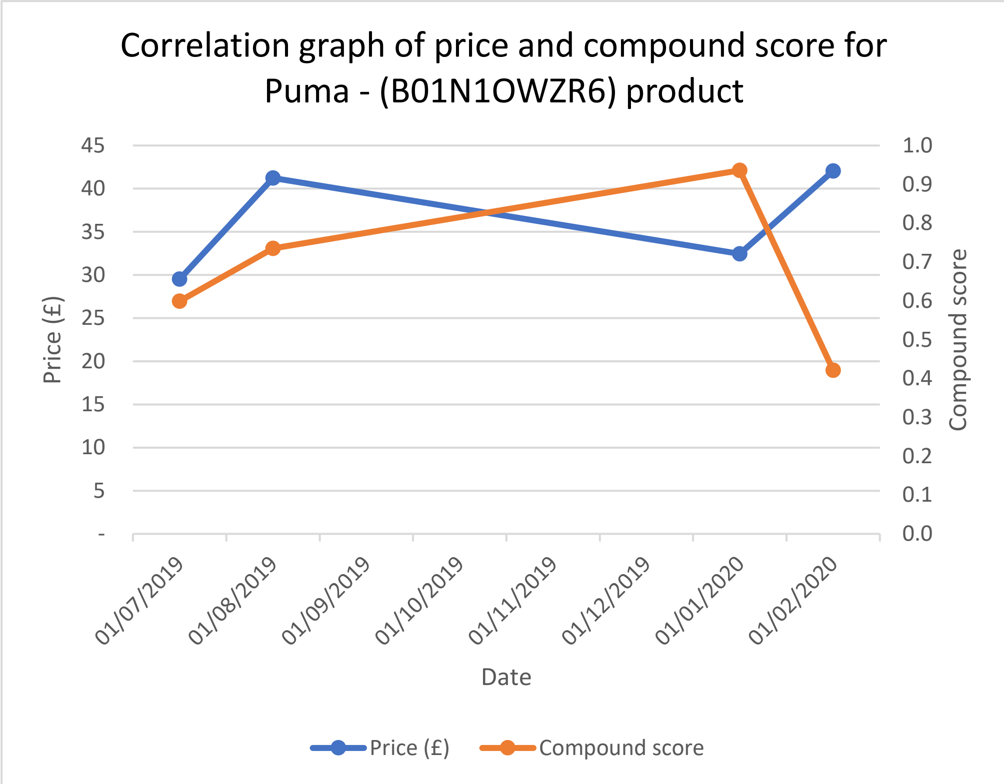


Figure A47: Correlation graph of price and compound score for Puma B01N1OWZR6 product

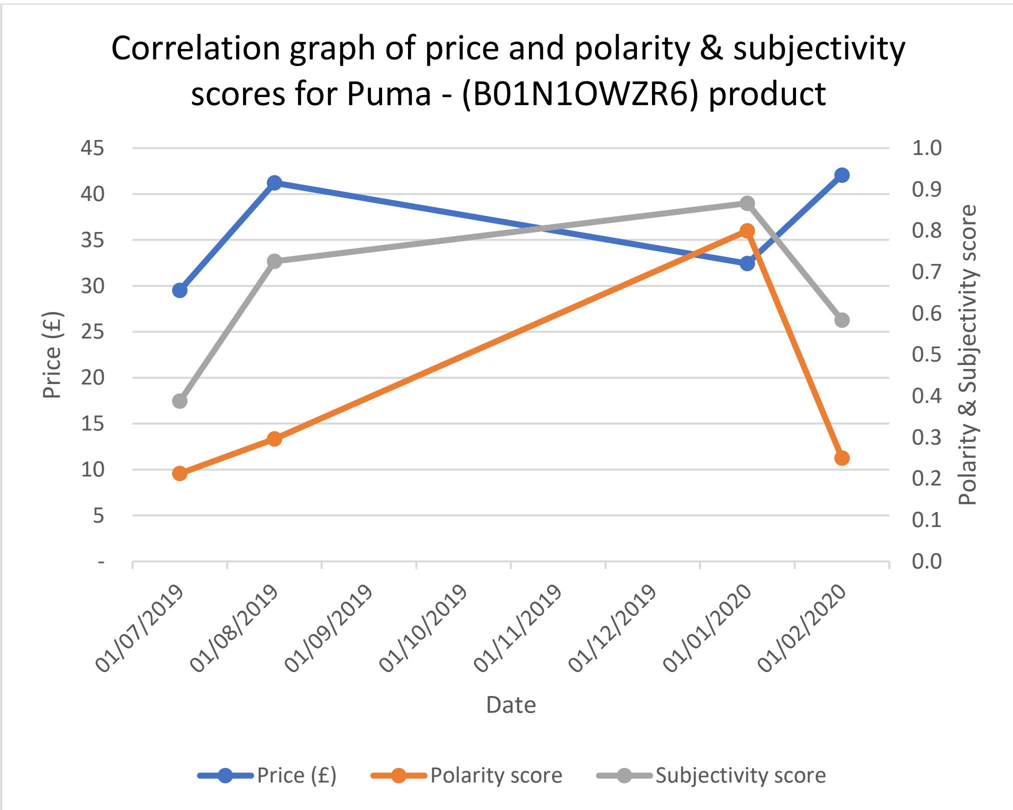


Figure A48: Correlation graph of price and polarity & subjectivity scores for Puma B01N1OWZR6 product

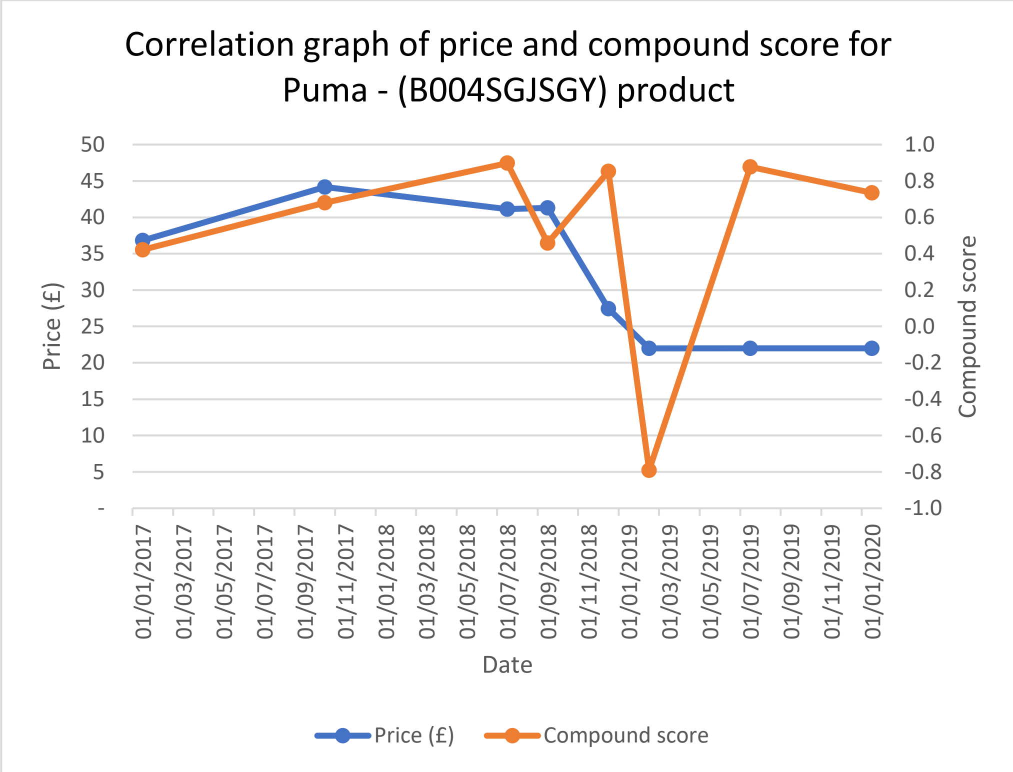


Figure A49: Correlation graph of price and compound score for Puma B004SGJSGY product

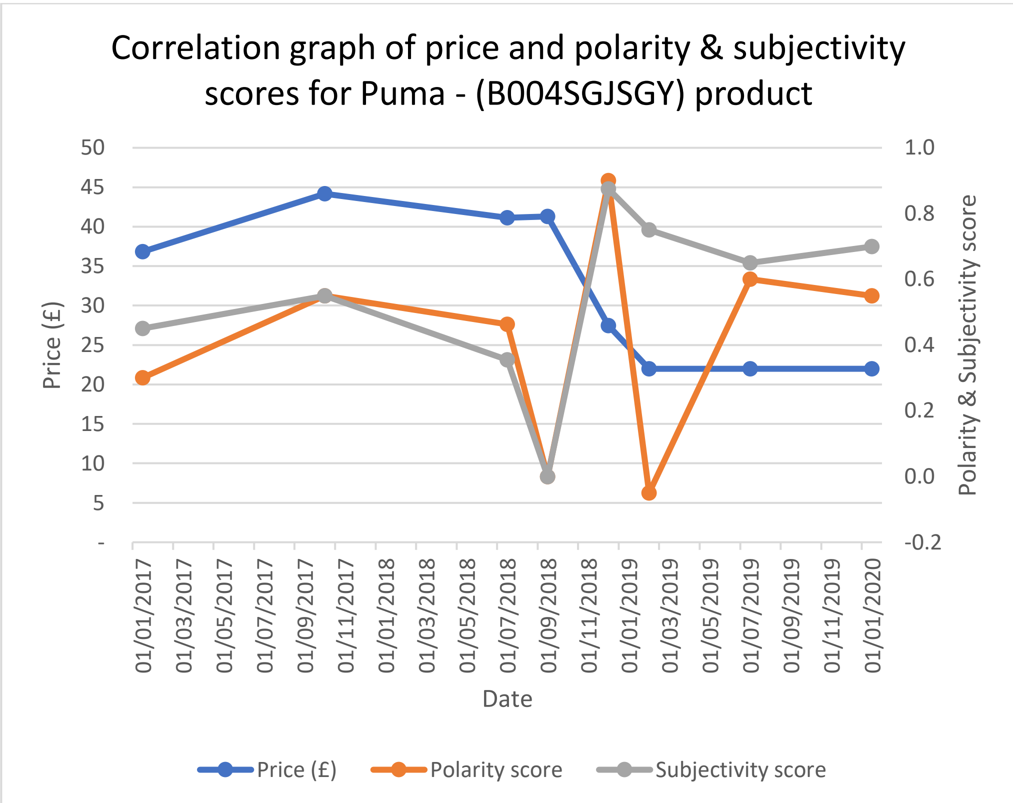


Figure A50: Correlation graph of price and polarity & subjectivity scores for Puma B004SGJSGY product

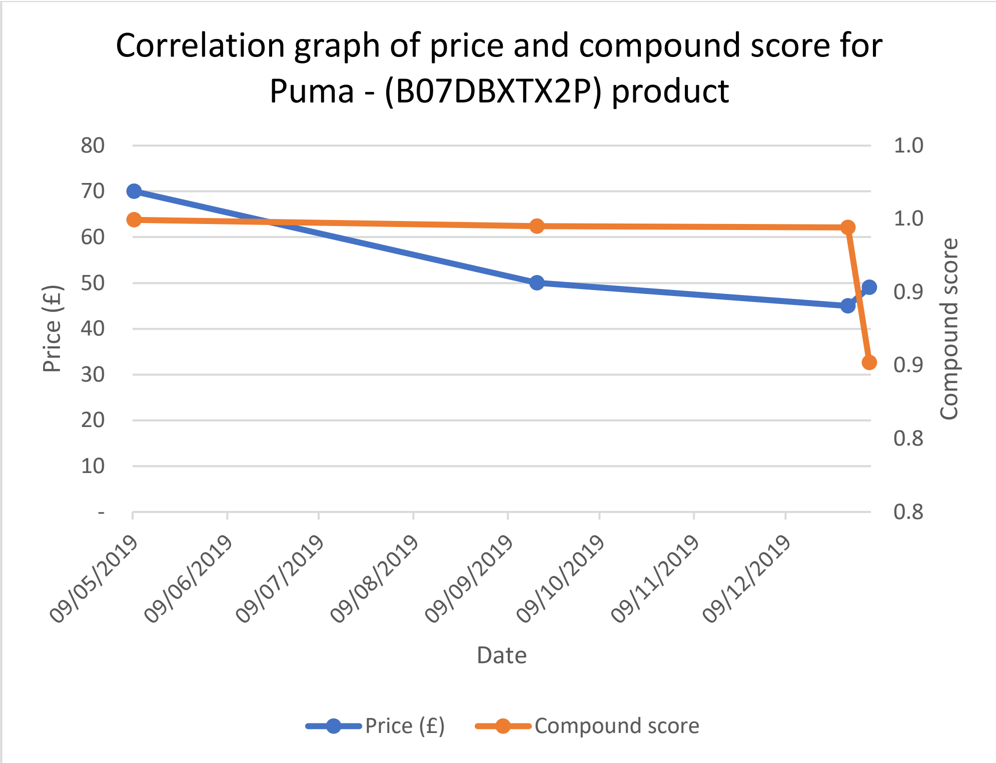


Figure A51: Correlation graph of price and compound score for Puma B07DBXTX2P product

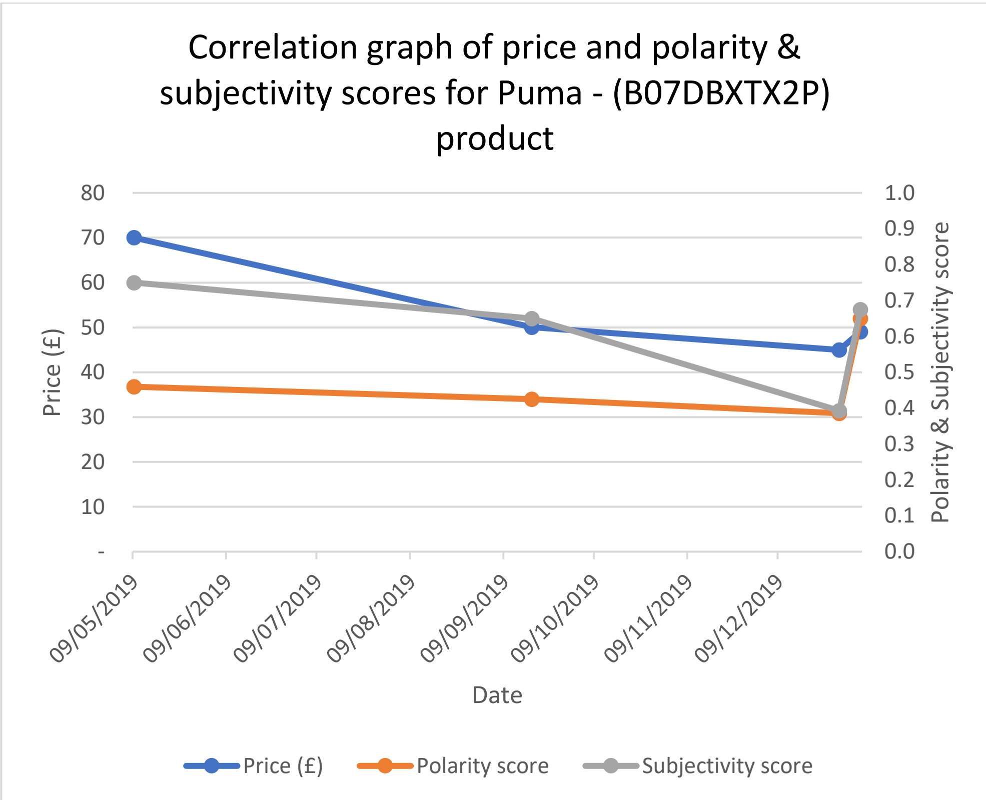


Figure A52: Correlation graph of price and polarity & subjectivity scores for Puma B07DBXTX2P product

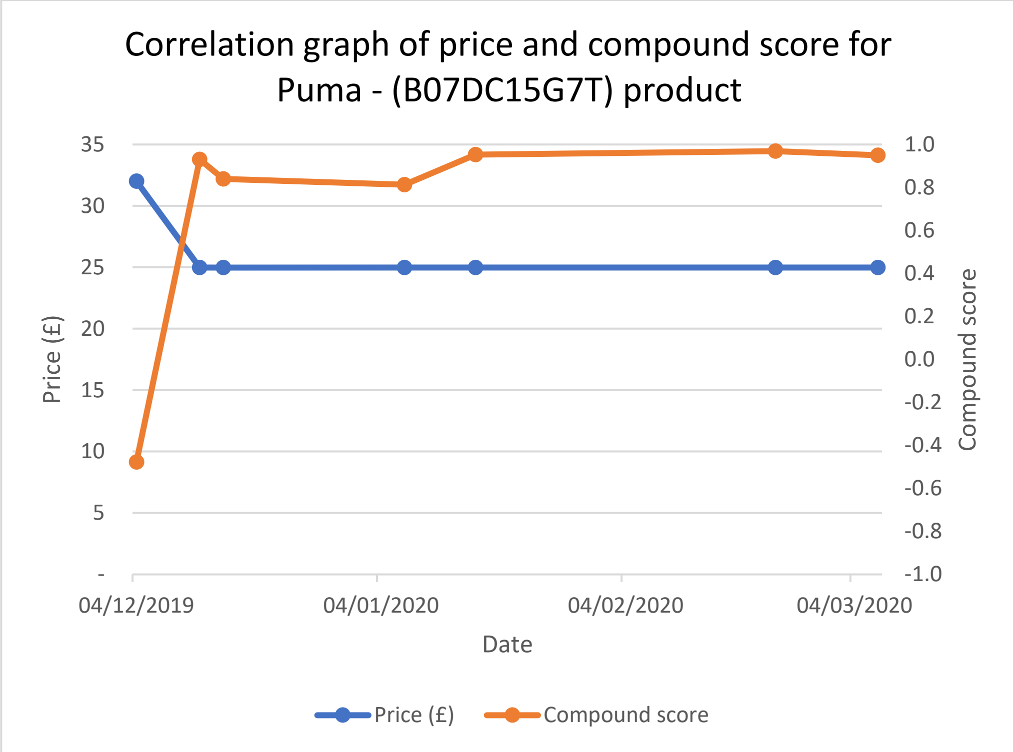


Figure A53: Correlation graph of price and compound score for Puma B07DC15G7T product

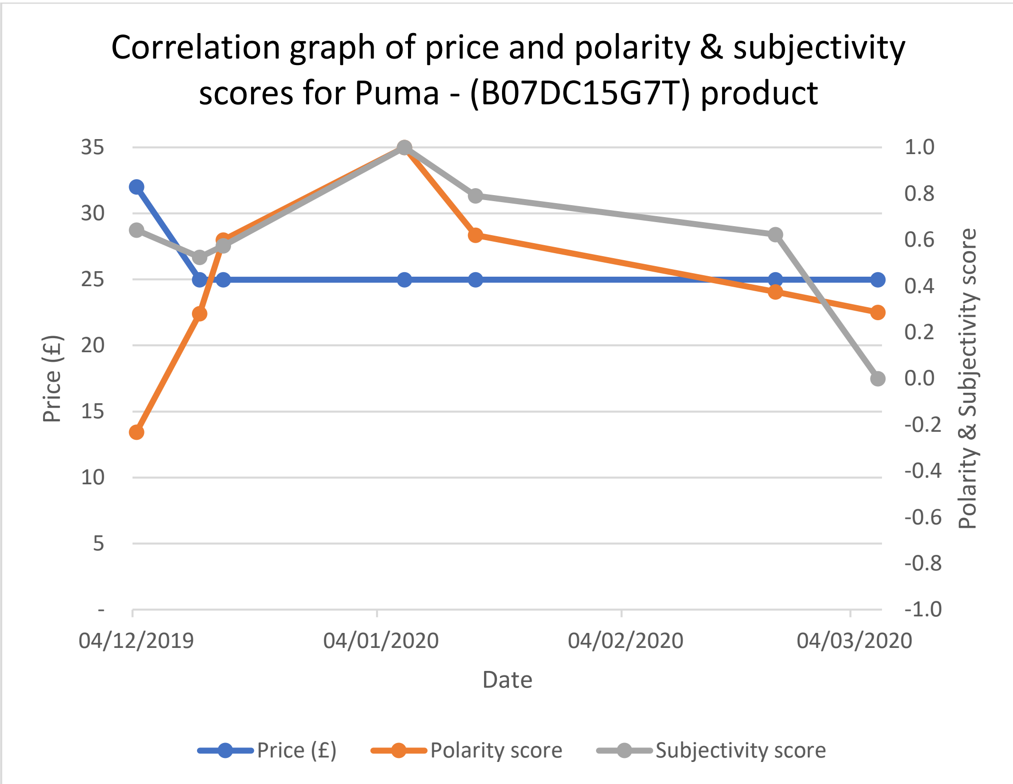


Figure A54: Correlation graph of price and polarity & subjectivity scores for Puma B07DC15G7T product

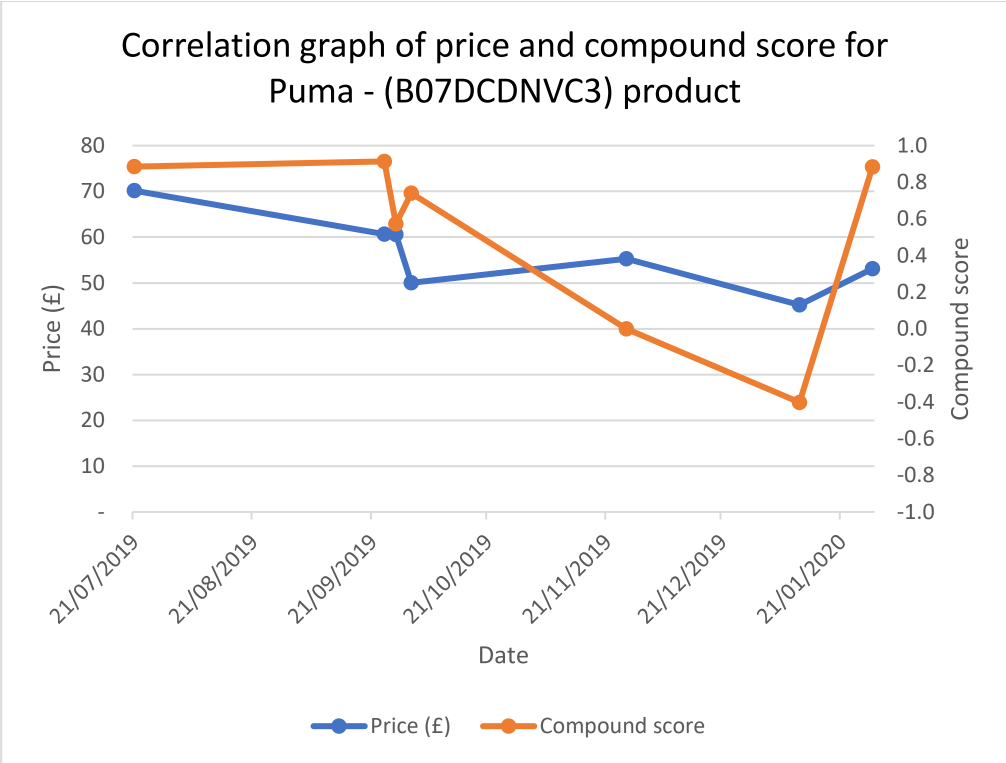


Figure A55: Correlation graph of price and compound score for Puma B07DCDNVC3 product

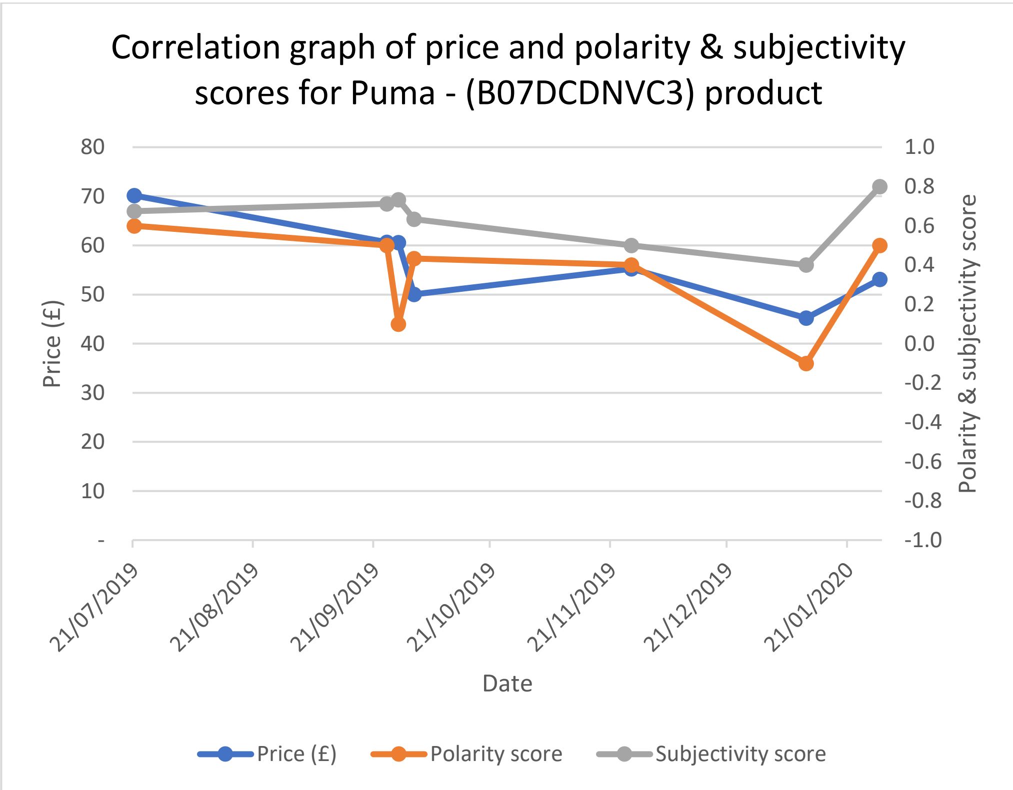


Figure A56: Correlation graph of price and polarity & subjectivity scores for Puma B07DCDNVC3 product

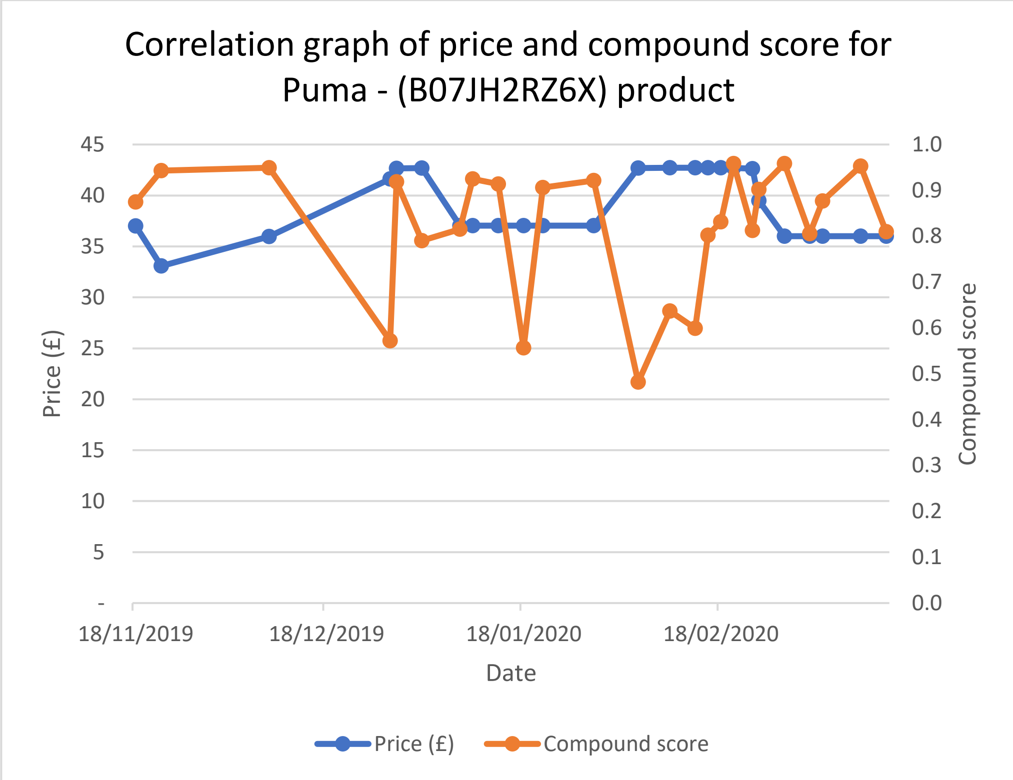


Figure A57: Correlation graph of price and compound score for Puma B07JH2RZ6X product

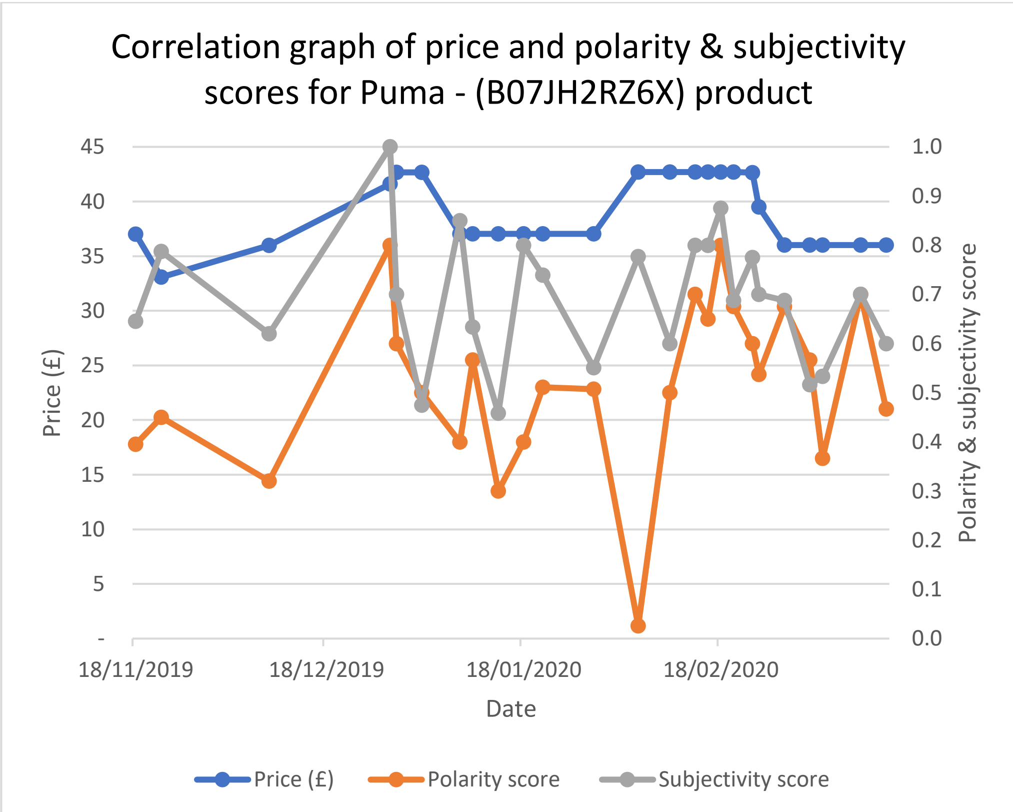


Figure A58: Correlation graph of price and polarity & subjectivity scores for Puma B07JH2RZ6X product

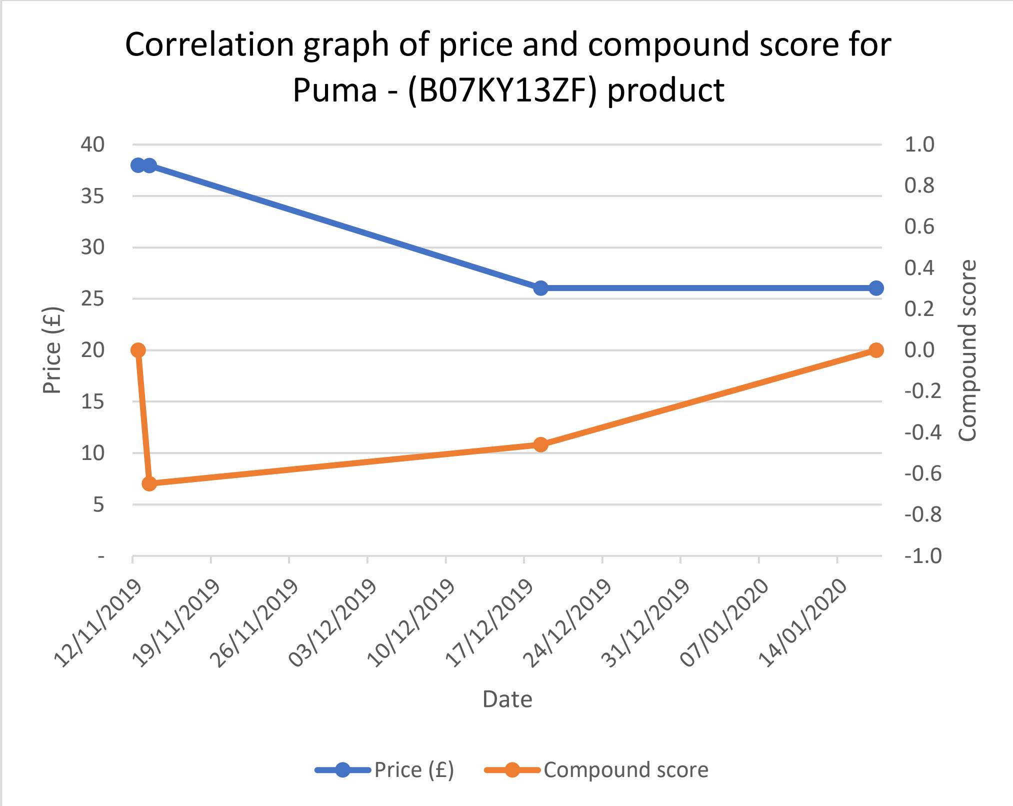


Figure A59: Correlation graph of price and compound score for Puma B07KY13ZF product

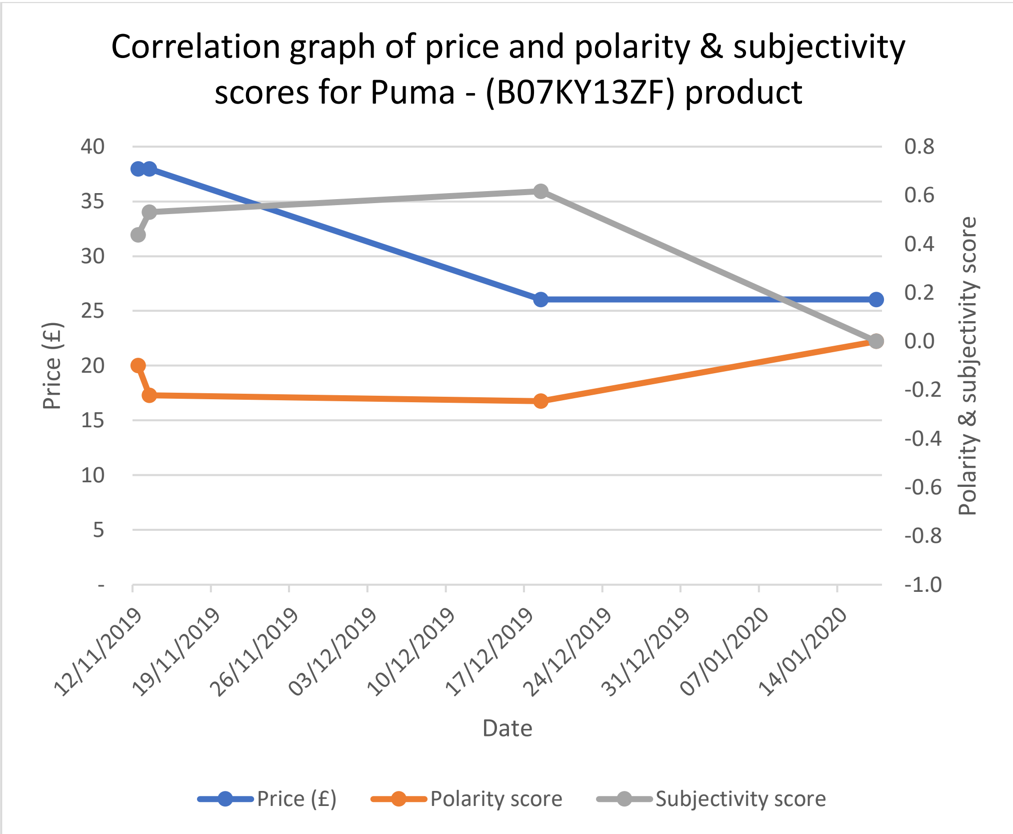


Figure A60: Correlation graph of price and polarity & subjectivity scores for Puma B07KY13ZF product

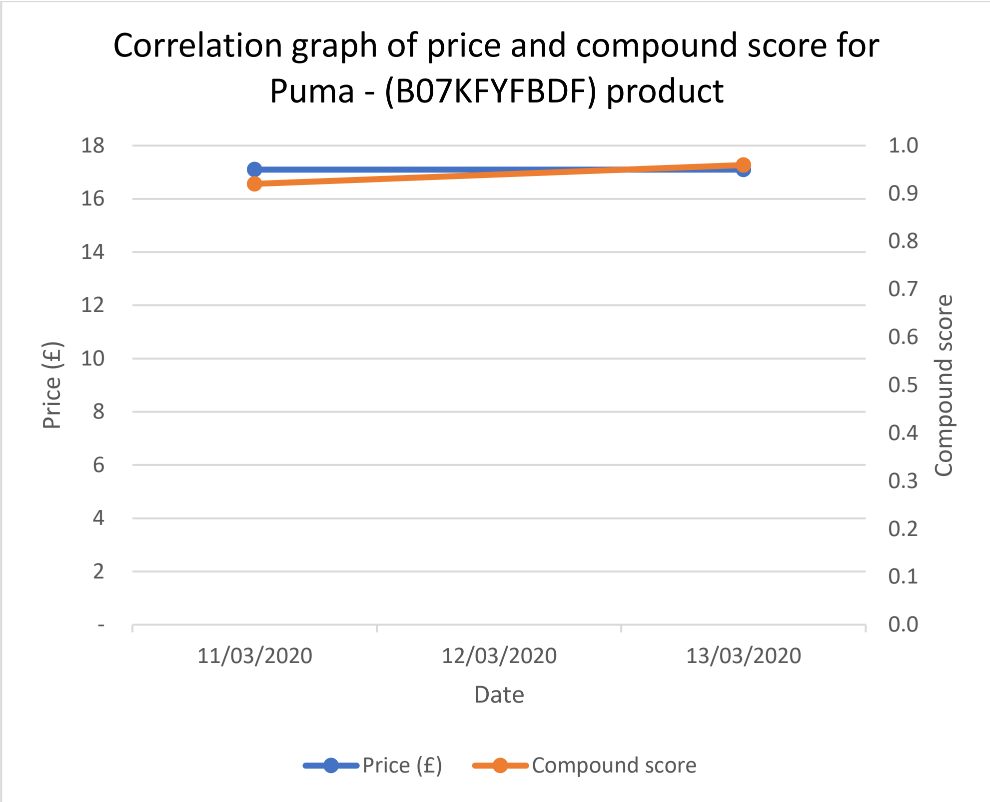


Figure A61: Correlation graph of price and compound score for Puma B07KFYFBDF product

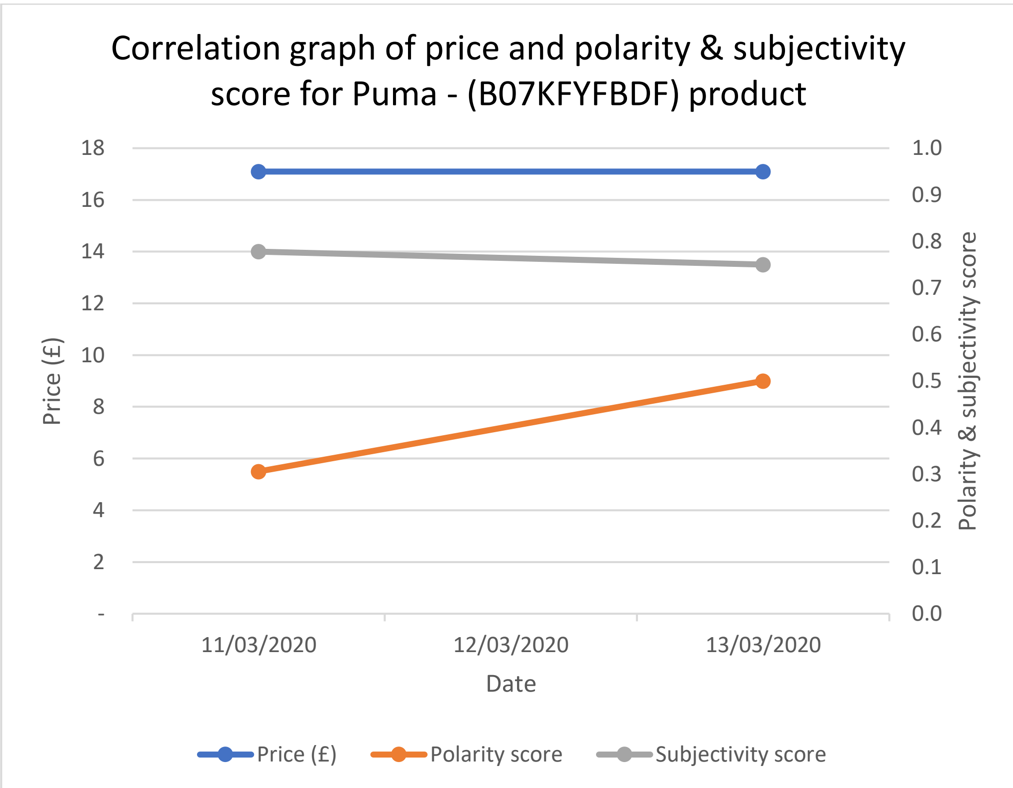


Figure A62: Correlation graph of price and polarity & subjectivity scores for Puma B07KFYFBDF product

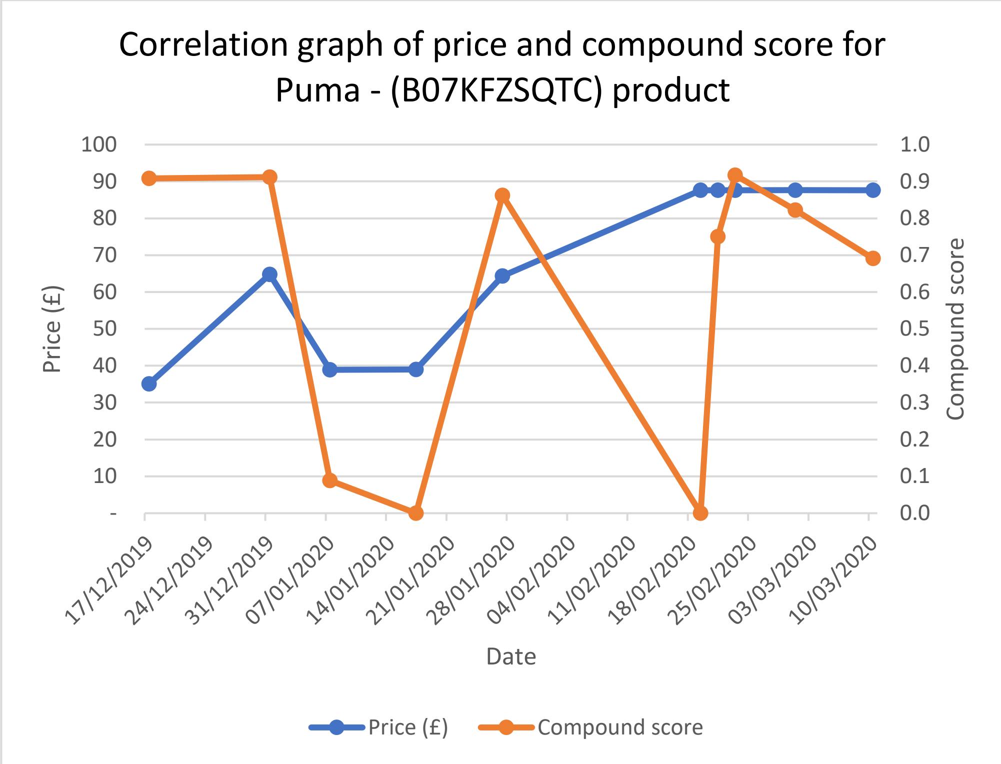


Figure A63: Correlation graph of price and compound score for Puma B07KFZSQTC product

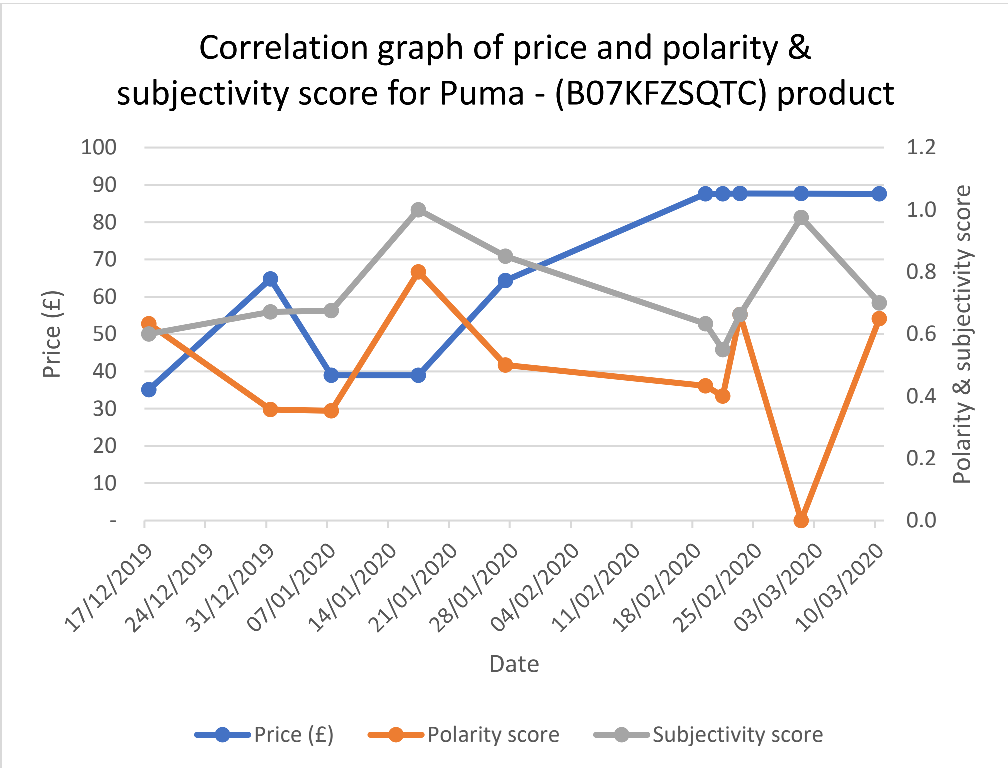


Figure A64: Correlation graph of price and polarity & subjectivity scores for Puma B07KFZSQTC product

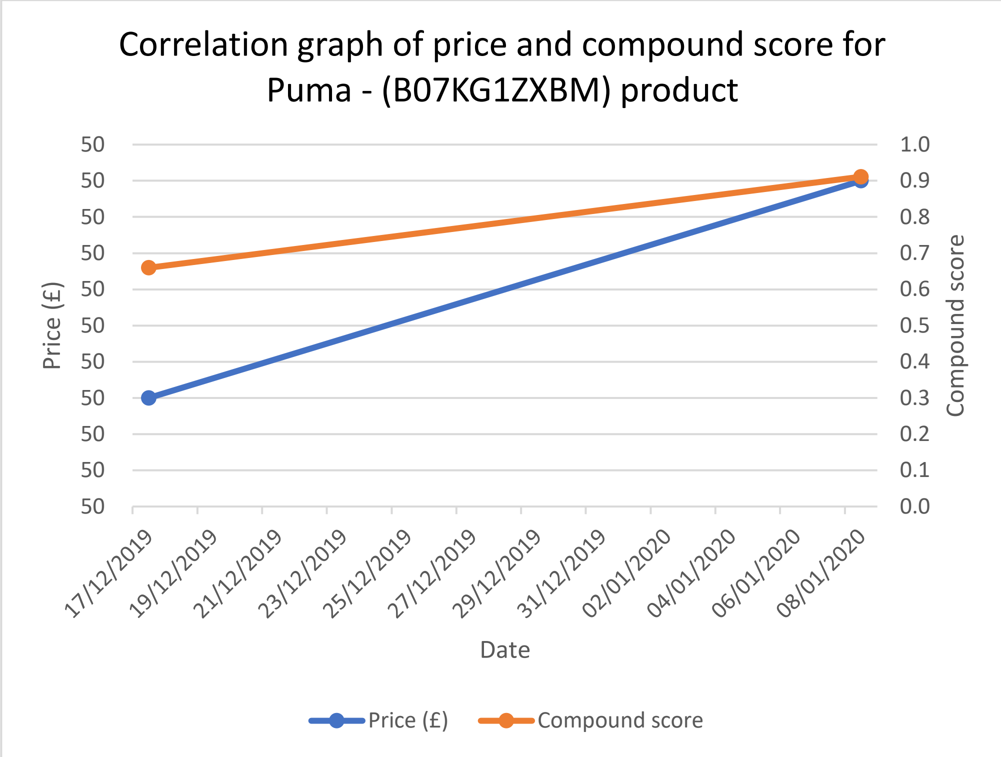


Figure A65: Correlation graph of price and compound score for Puma B07KG1ZXBM product

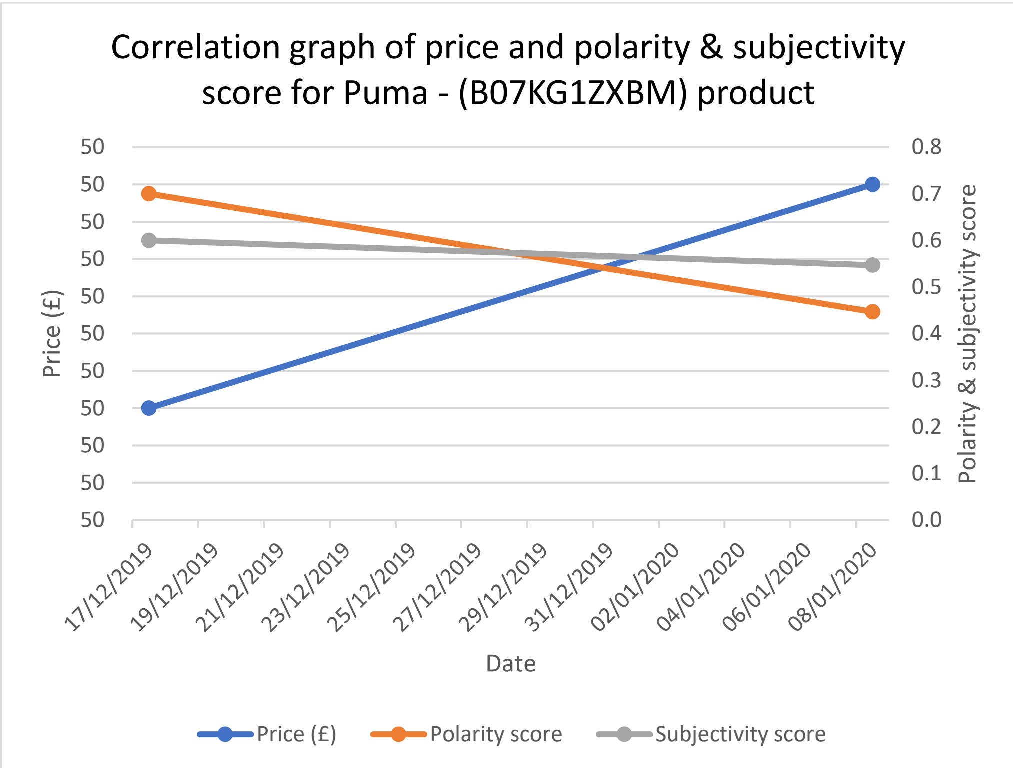


Figure A66: Correlation graph of price and polarity & subjectivity scores for Puma B07KG1ZXBM product

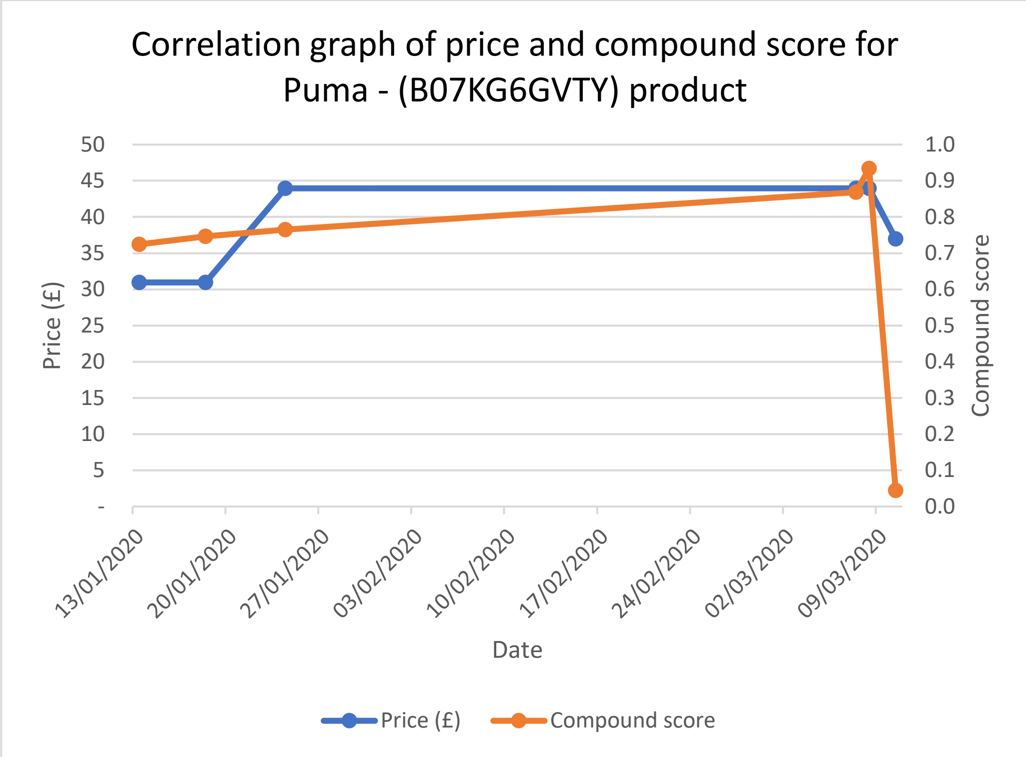


Figure A67: Correlation graph of price and compound score for Puma B07KG6GVTY product

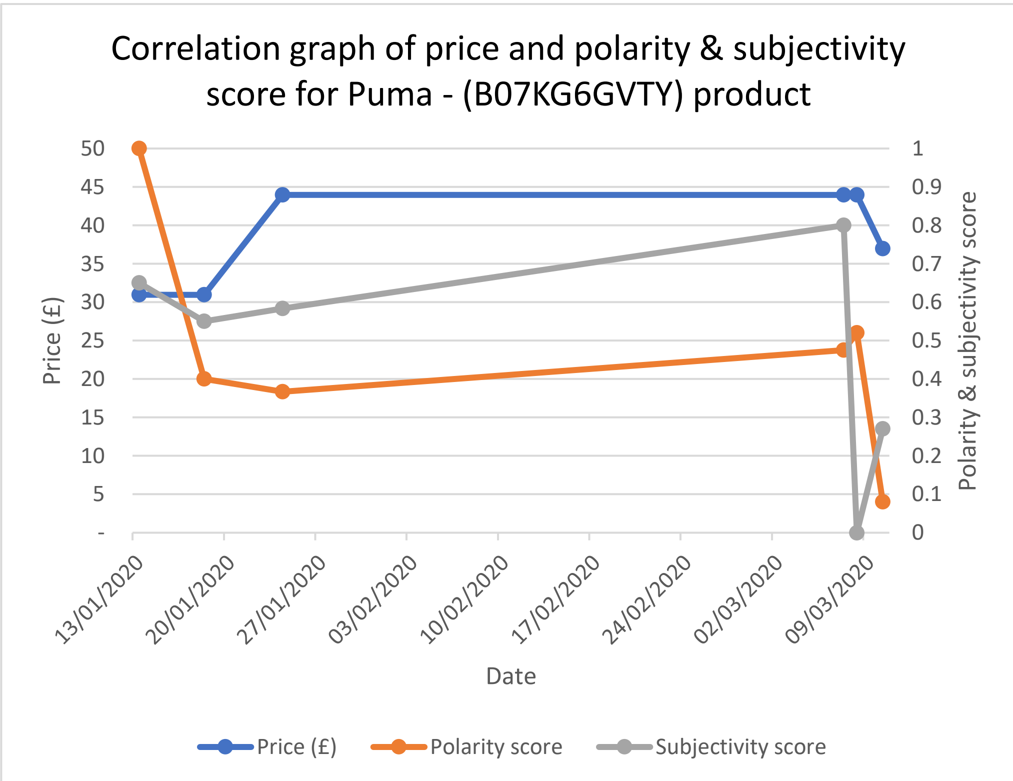


Figure A68: Correlation graph of price and polarity & subjectivity scores for Puma B07KG6GVTY product

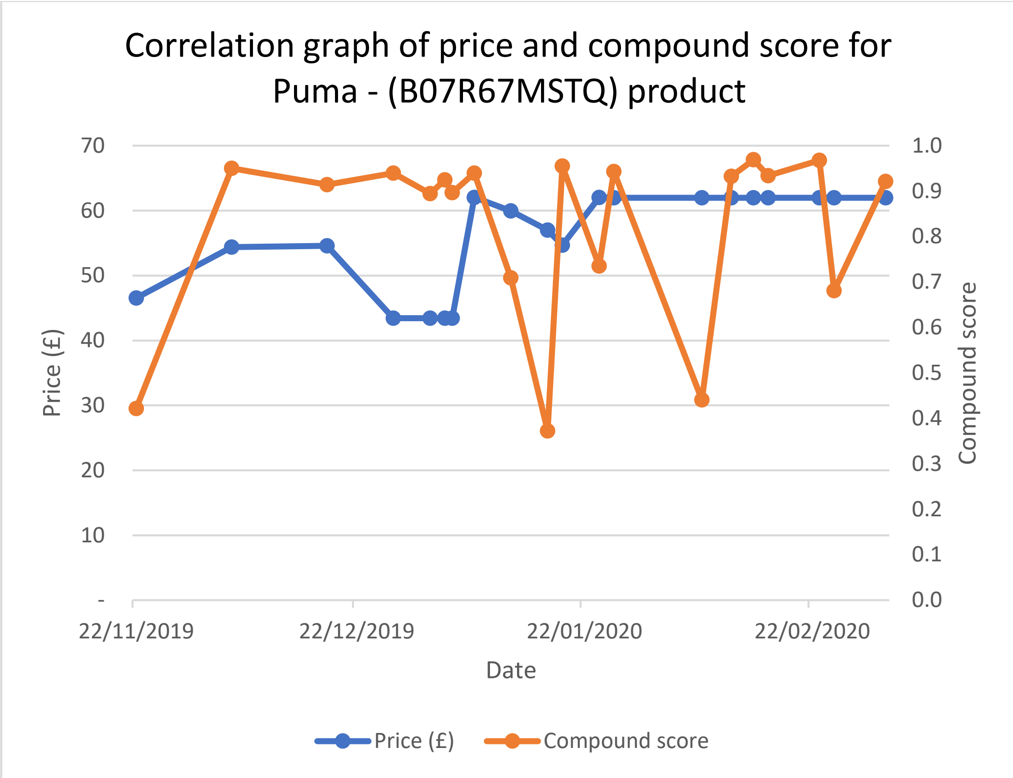


Figure A69: Correlation graph of price and compound score for Puma B07R67MSTQ product

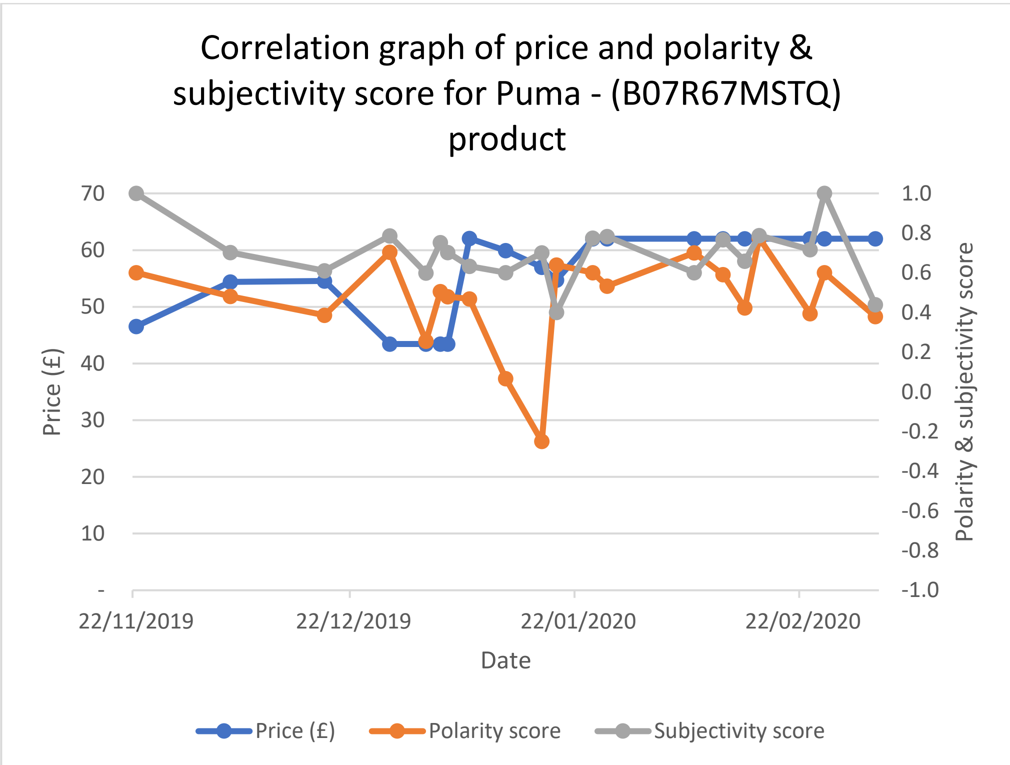


Figure A70: Correlation graph of price and polarity & subjectivity scores for Puma B07R67MSTQ product

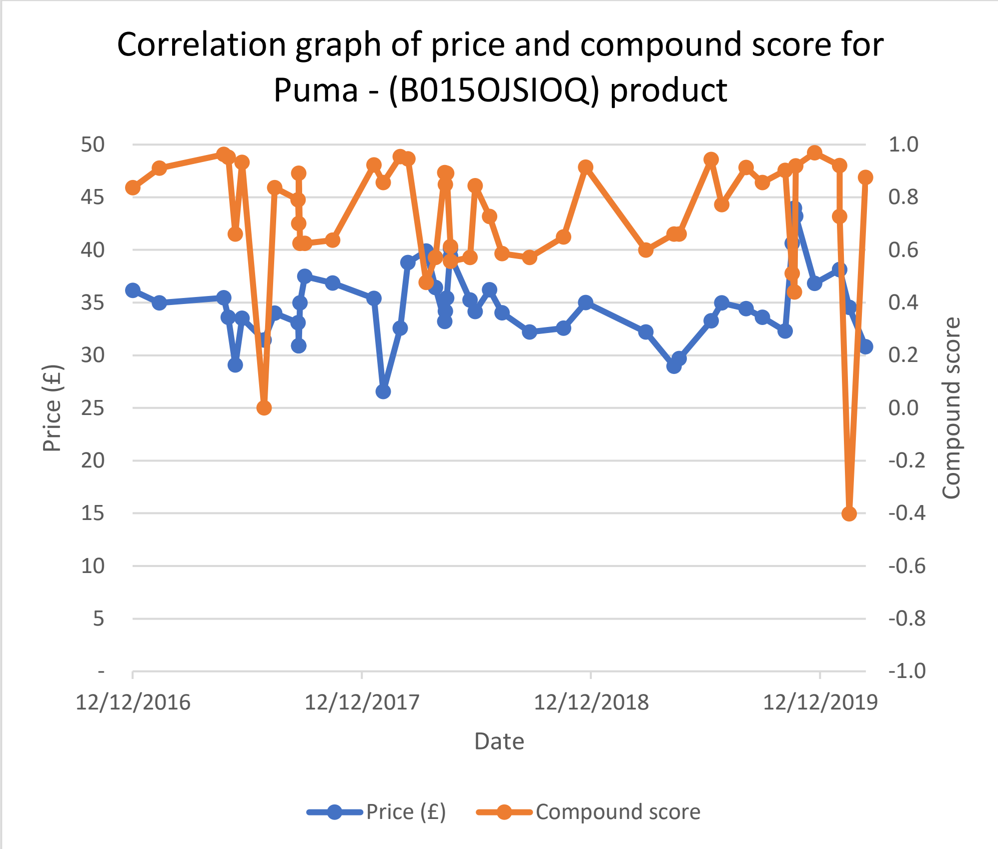


Figure A71: Correlation graph of price and compound score for Puma B015OJSIOQ product

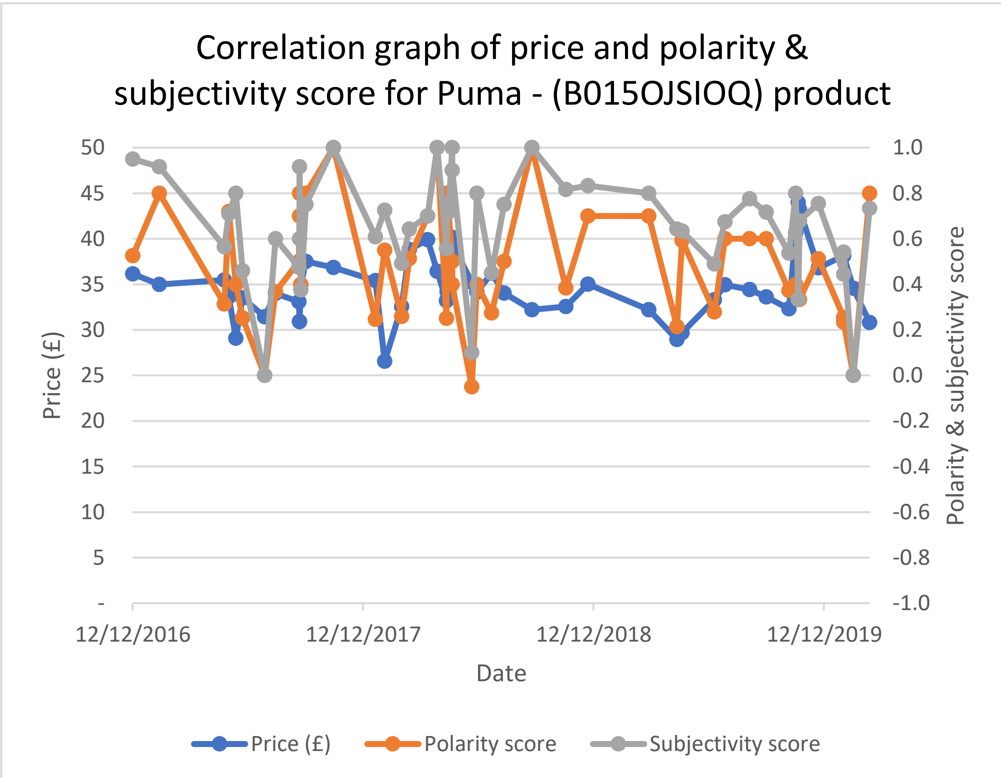


Figure A72: Correlation graph of price and polarity & subjectivity scores for Puma B015OJSIOQ product

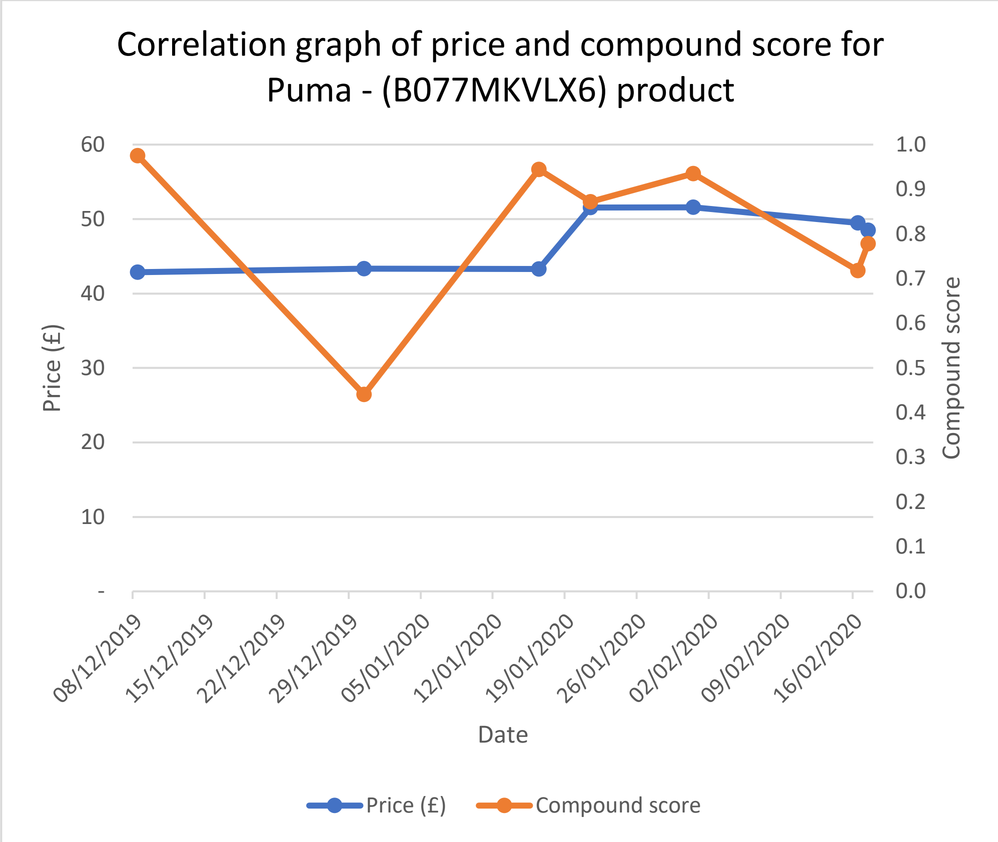


Figure A73: Correlation graph of price and compound score for Puma B077MKVLX6 product

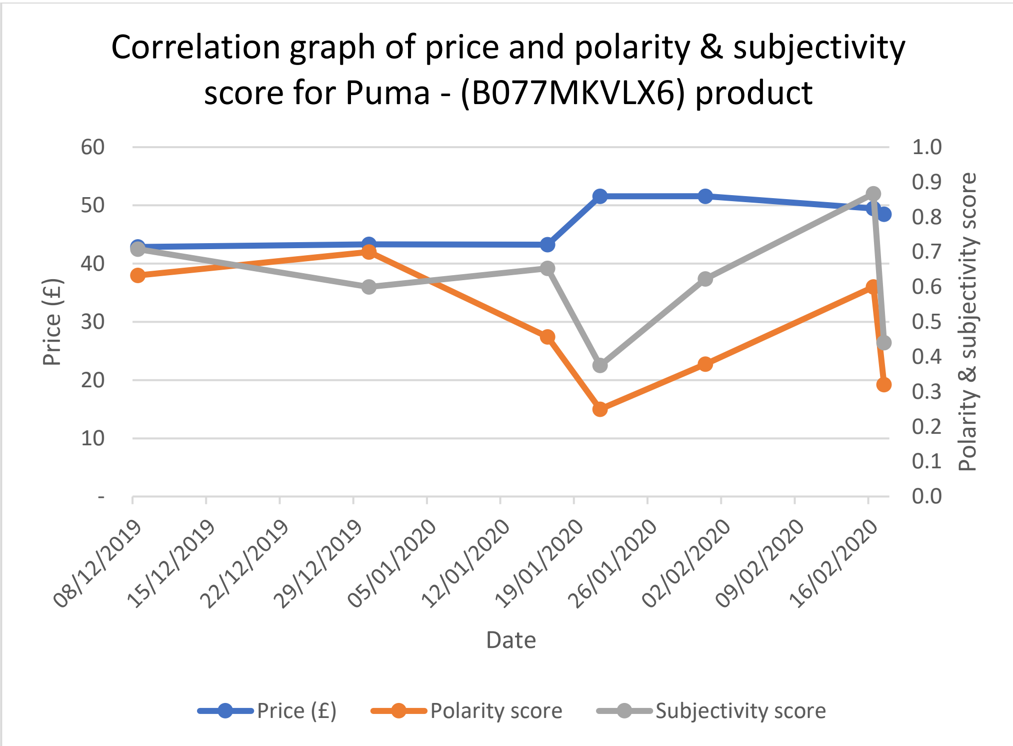


Figure A74: Correlation graph of price and polarity & subjectivity scores for Puma B077MKVLX6 product

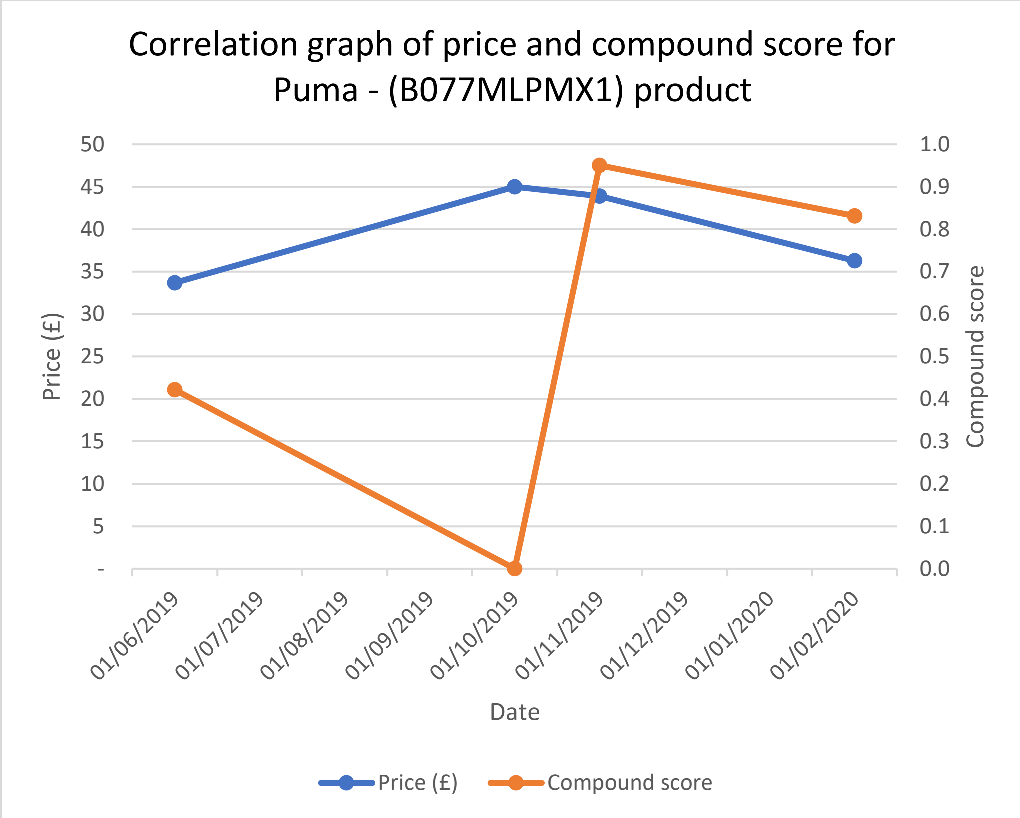


Figure A75: Correlation graph of price and compound score for Puma B077MLPMX1 product

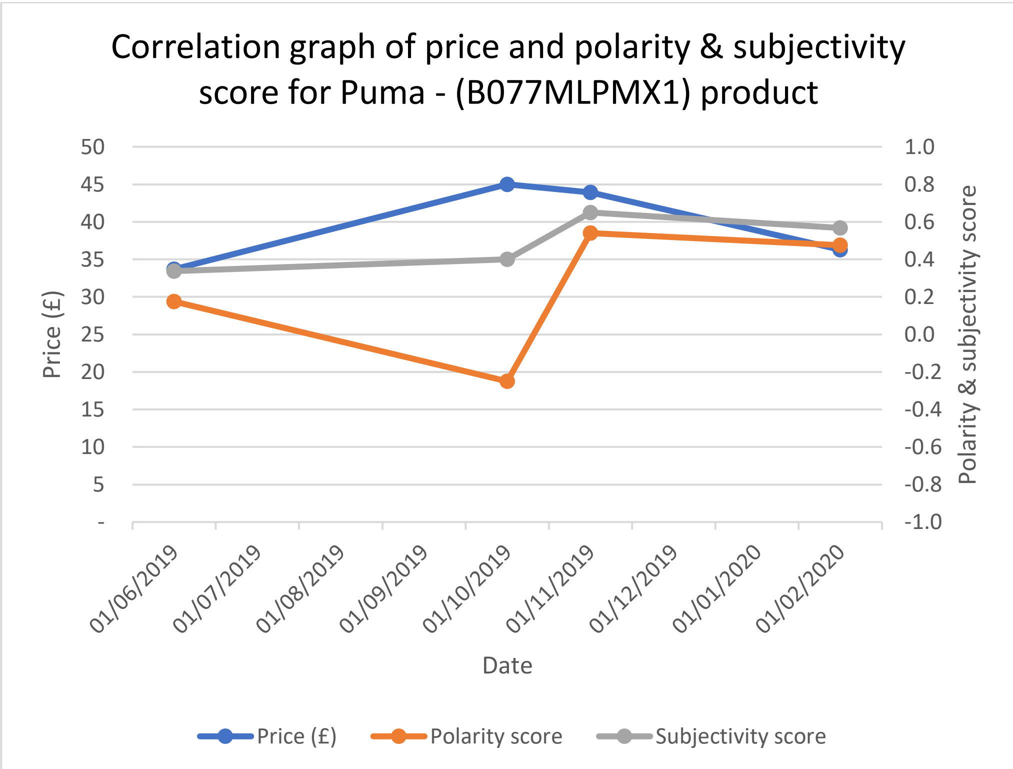


Figure A76: Correlation graph of price and polarity & subjectivity scores for Puma B077MLPMX1 product

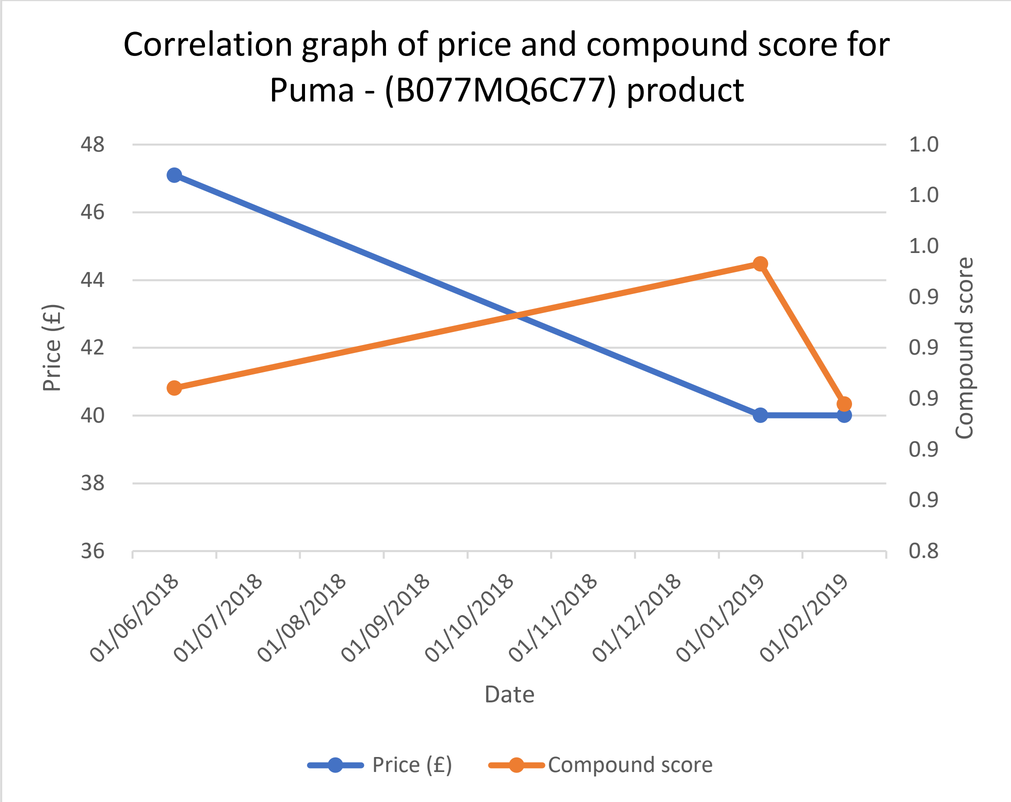


Figure A77: Correlation graph of price and compound score for Puma B077MQ6C77 product

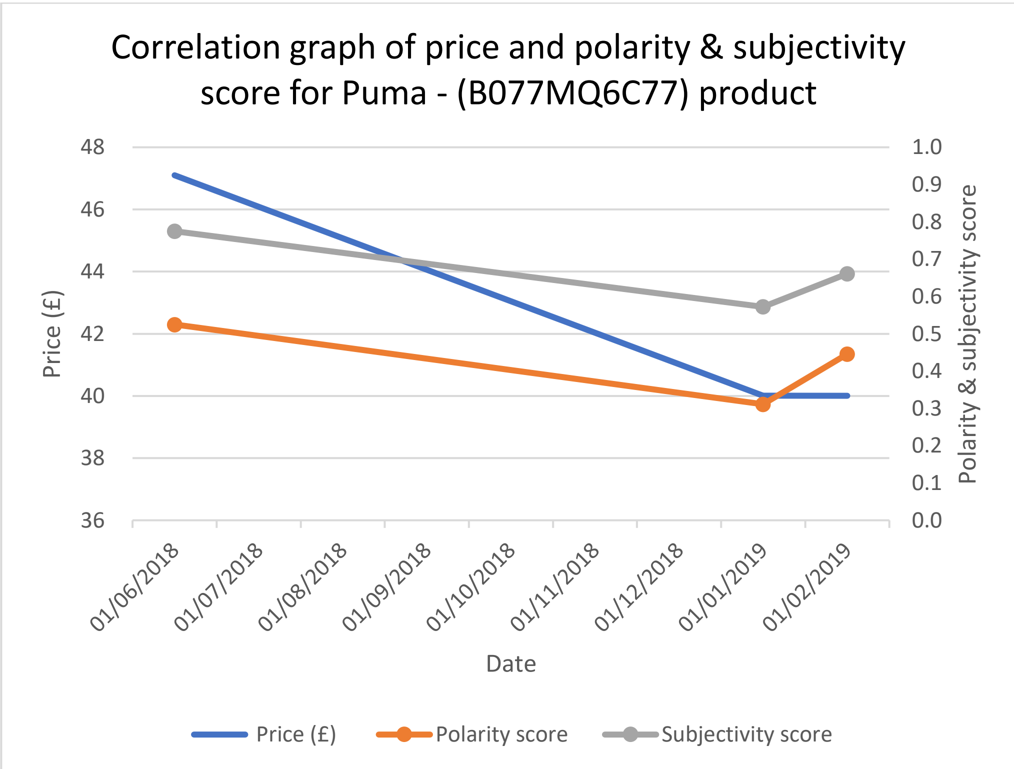


Figure A78: Correlation graph of price and polarity & subjectivity scores for Puma B077MQ6C77 product

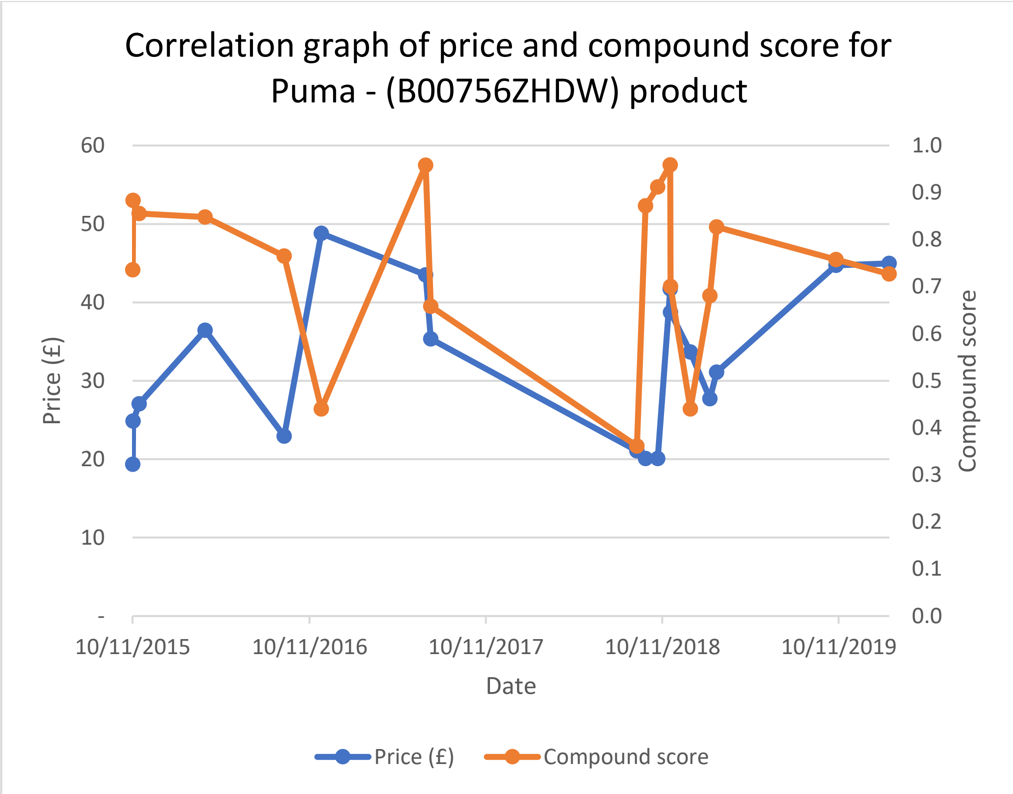


Figure A79: Correlation graph of price and compound score for Puma B00756ZHDW product

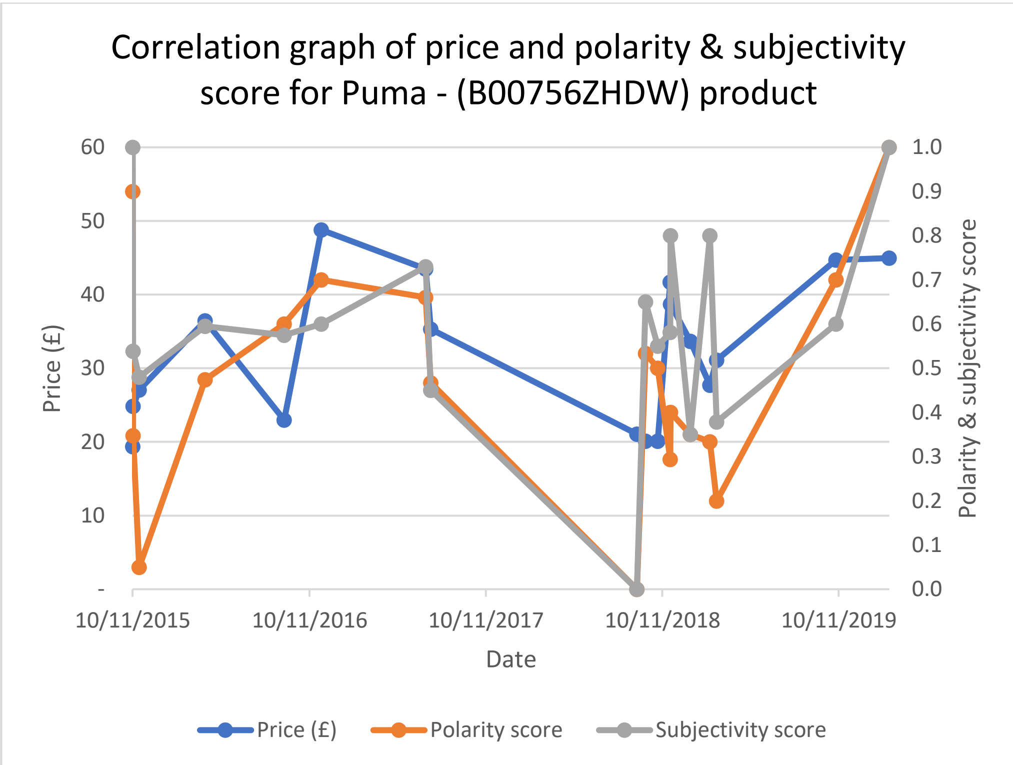


Figure A80: Correlation graph of price and polarity & subjectivity scores for Puma B00756ZHDW product

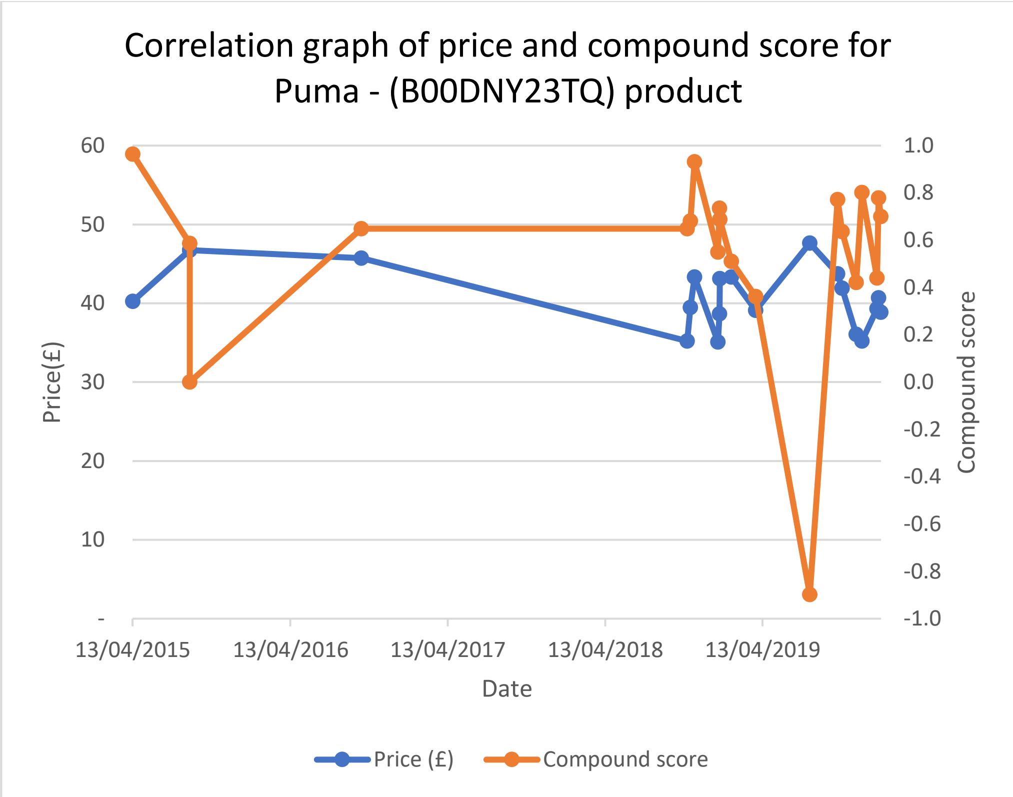


Figure A81: Correlation graph of price and compound score for Puma B00DNY23TQ product

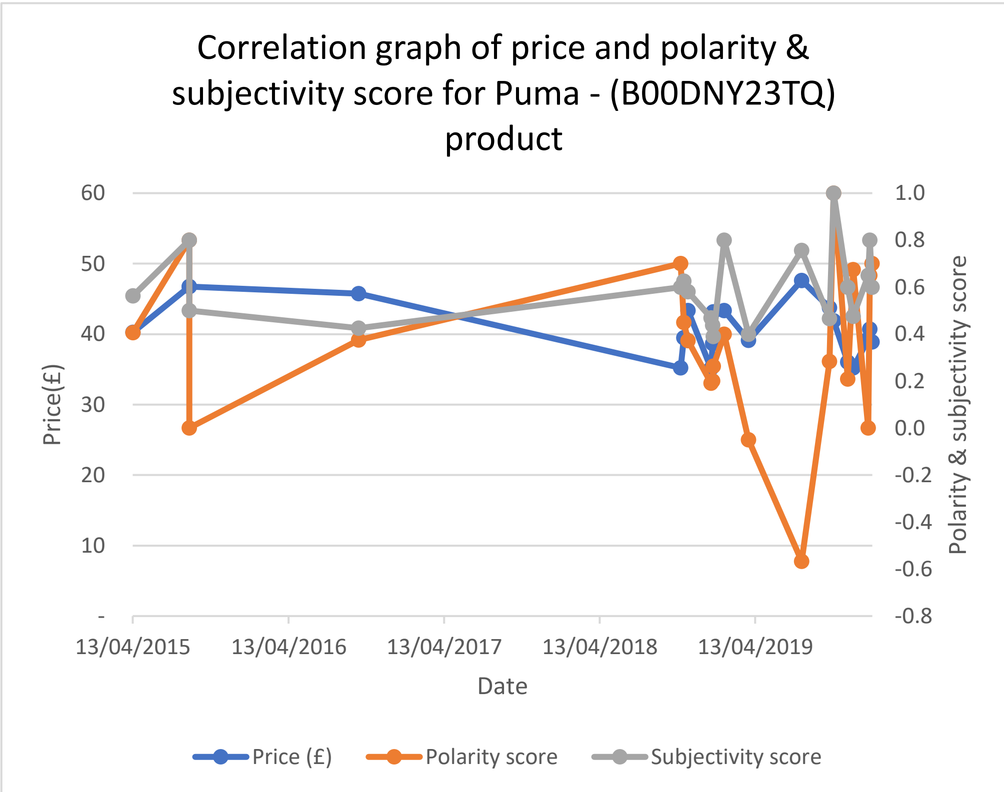


Figure A82: Correlation graph of price and polarity & subjectivity scores for Puma B00DNY23TQ product

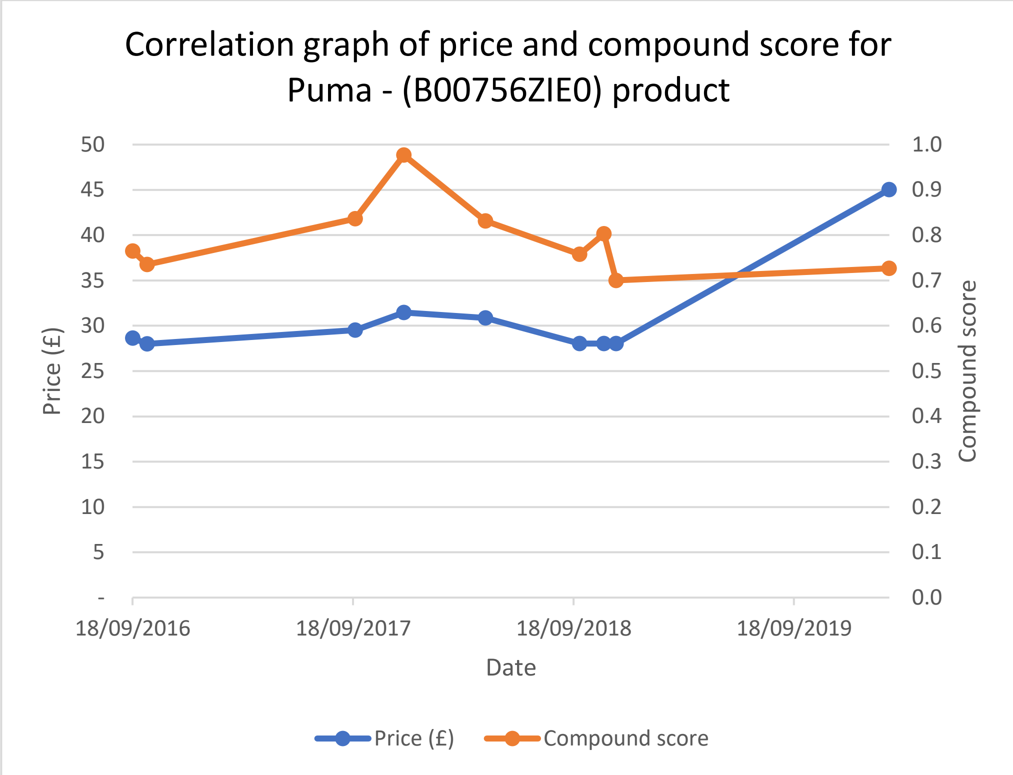


Figure A83: Correlation graph of price and compound score for Puma B00756ZIE0 product

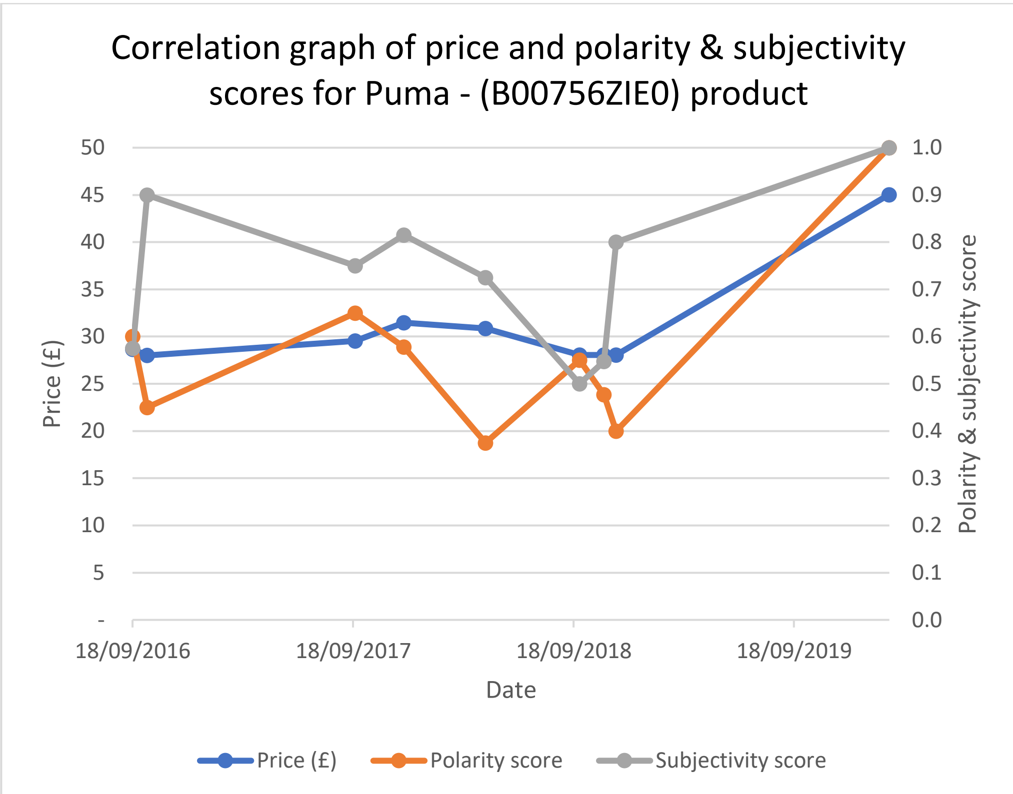


Figure A84: Correlation graph of price and polarity & subjectivity scores for Puma B00756ZIE0 product

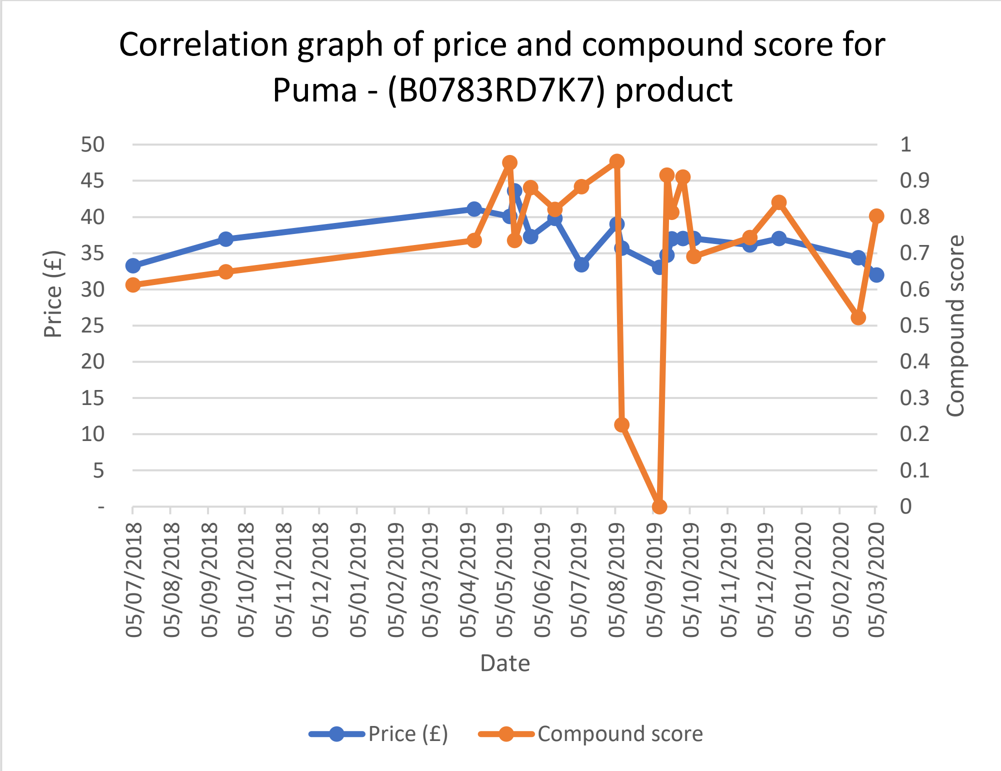


Figure A85: Correlation graph of price and compound score for Puma B0783RD7K7 product

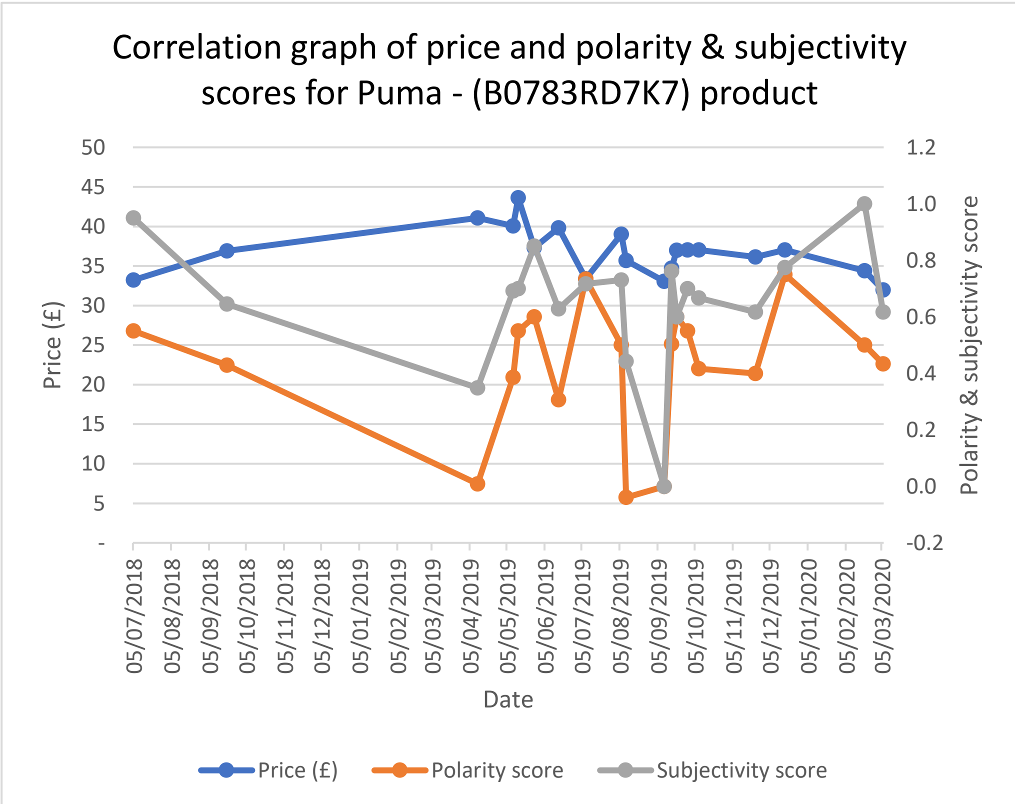


Figure A86: Correlation graph of price and polarity & subjectivity scores for Puma B0783RD7K7 product

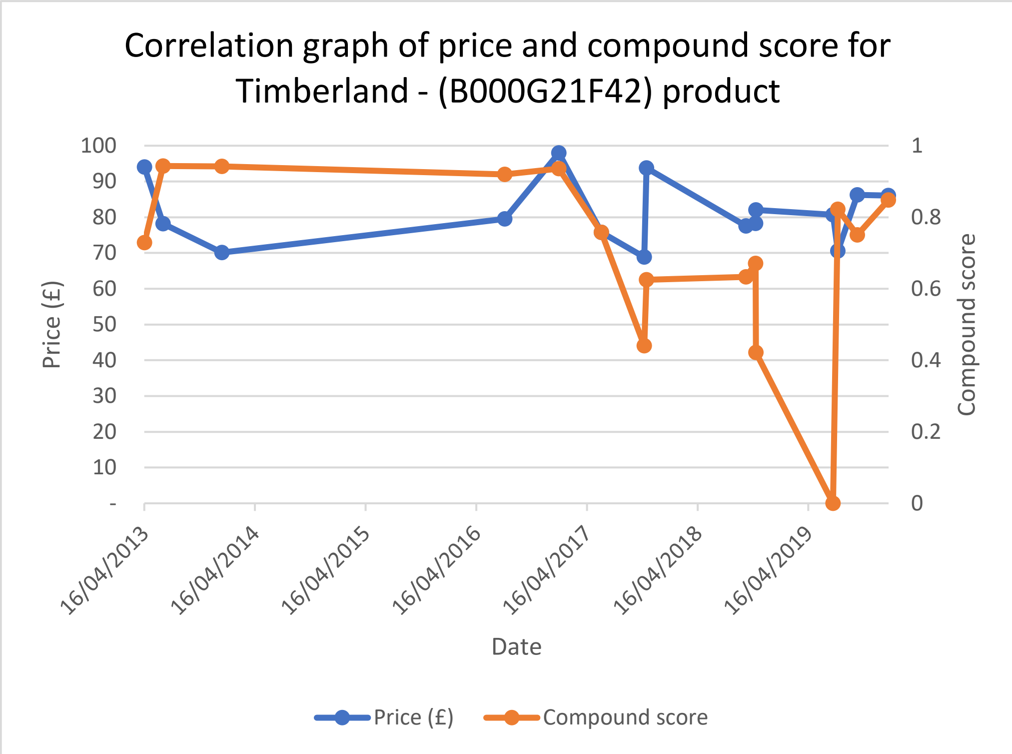


Figure A87: Correlation graph of price and compound score for Timberland B000G21F42 product

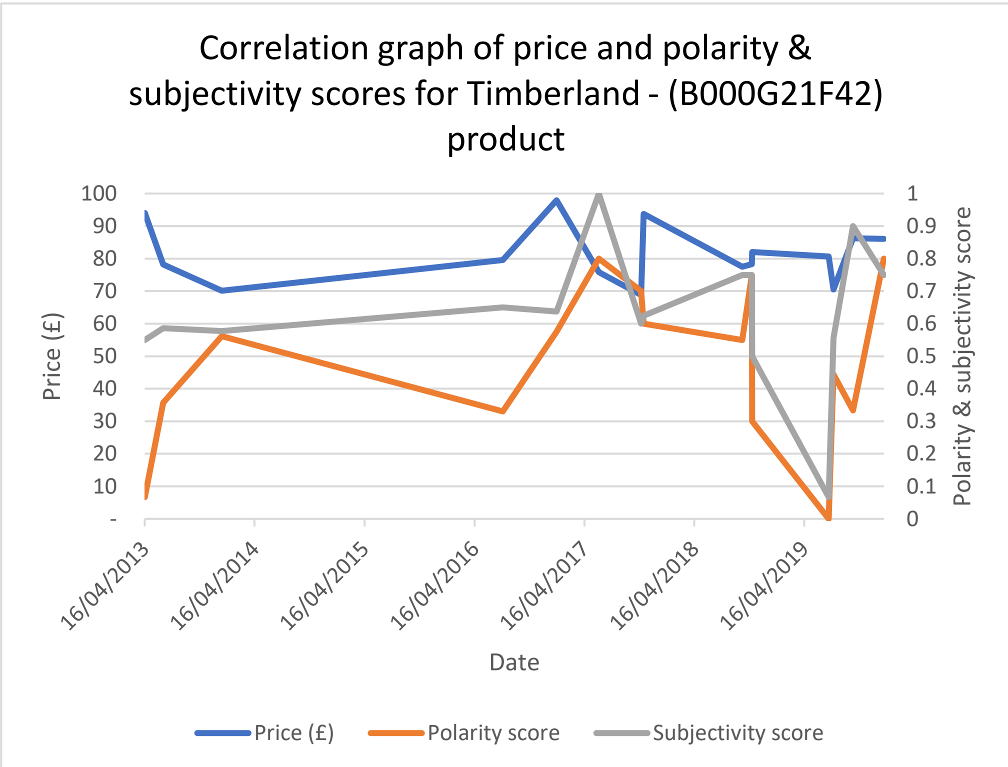


Figure A88: Correlation graph of price and polarity & subjectivity scores for Timberland B000G21F42 product

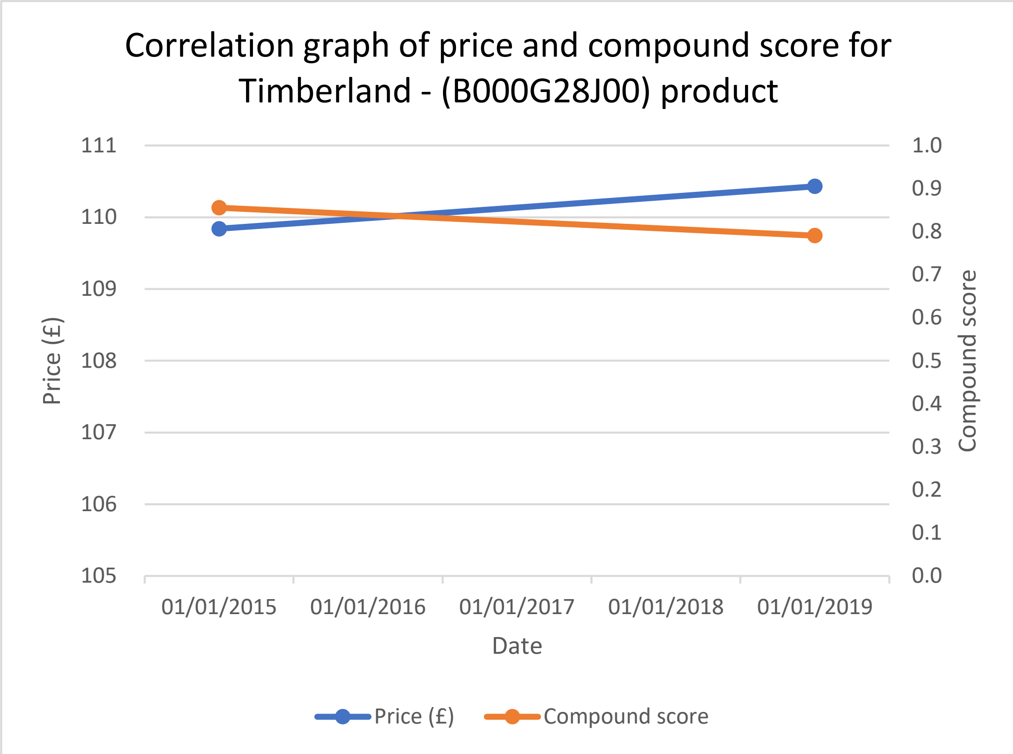


Figure A89: Correlation graph of price and compound score for Timberland B000G28J00 product

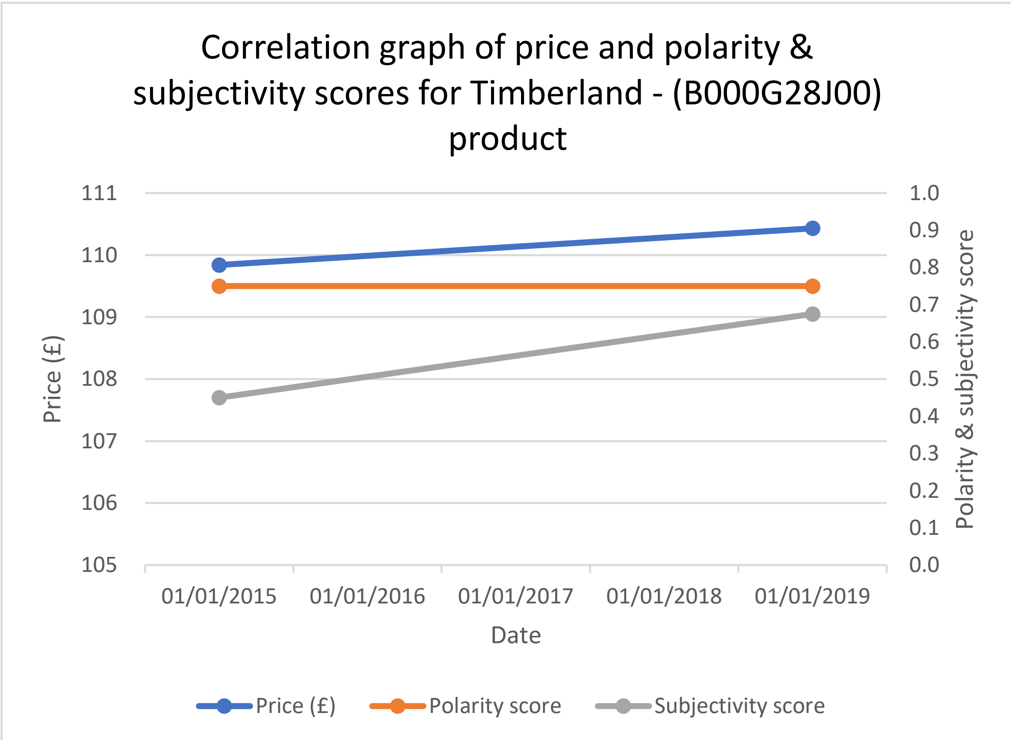


Figure A90: Correlation graph of price and polarity & subjectivity scores for Timberland B000G28J00 product

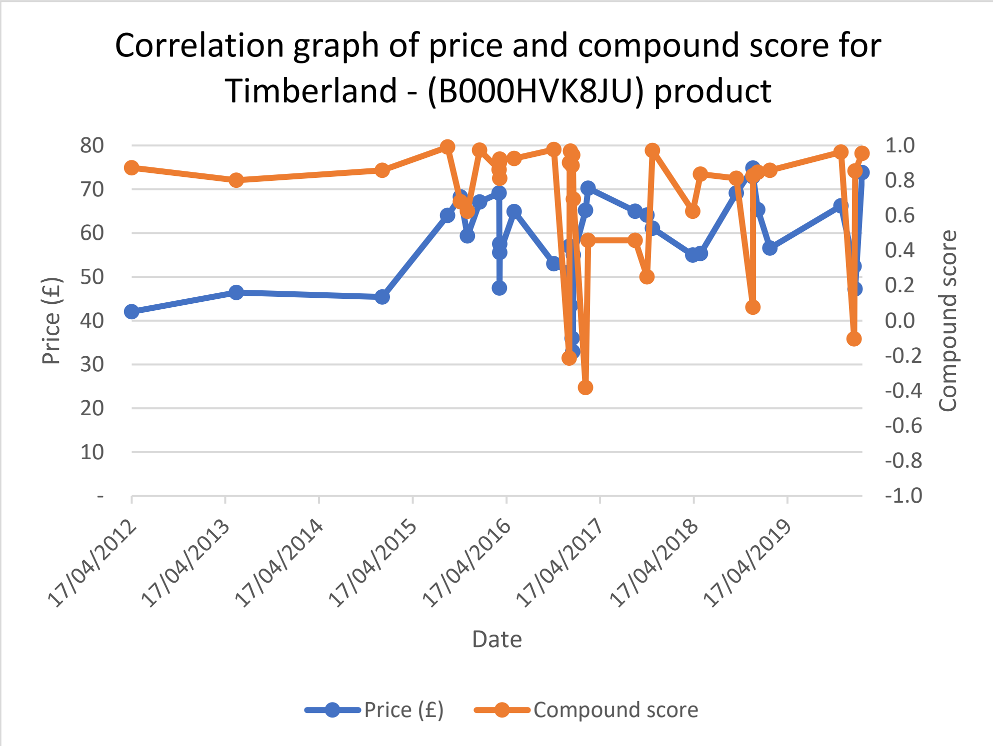


Figure A91: Correlation graph of price and compound score for Timberland B000HVK8JU product

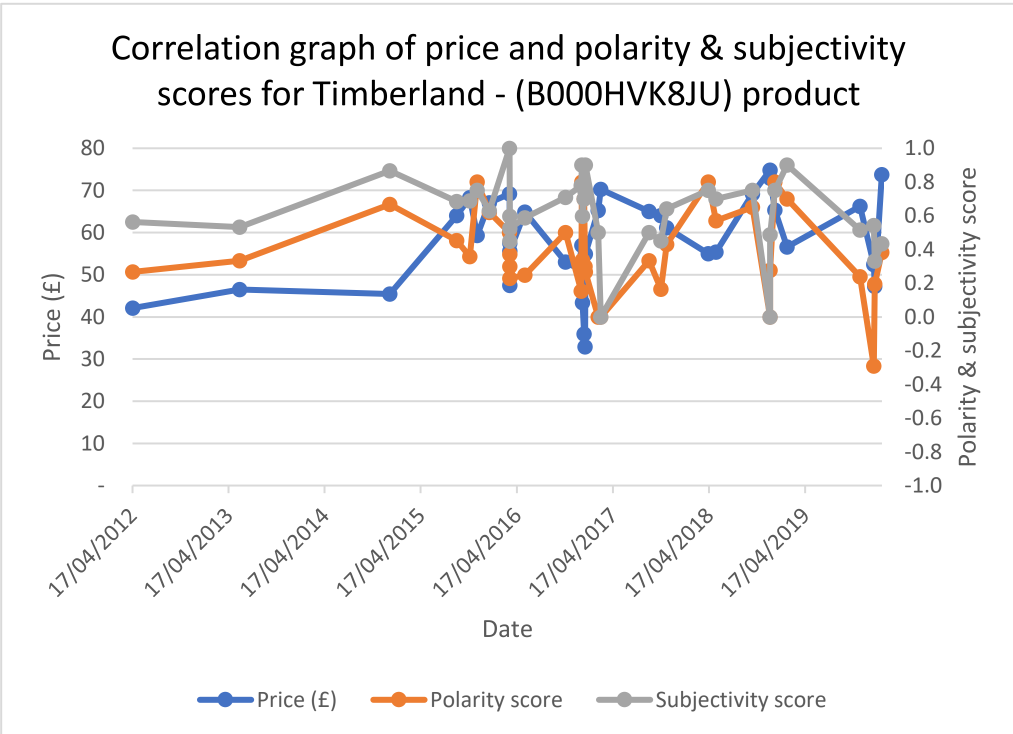


Figure A92: Correlation graph of price and polarity & subjectivity scores for Timberland B000HVK8JU product

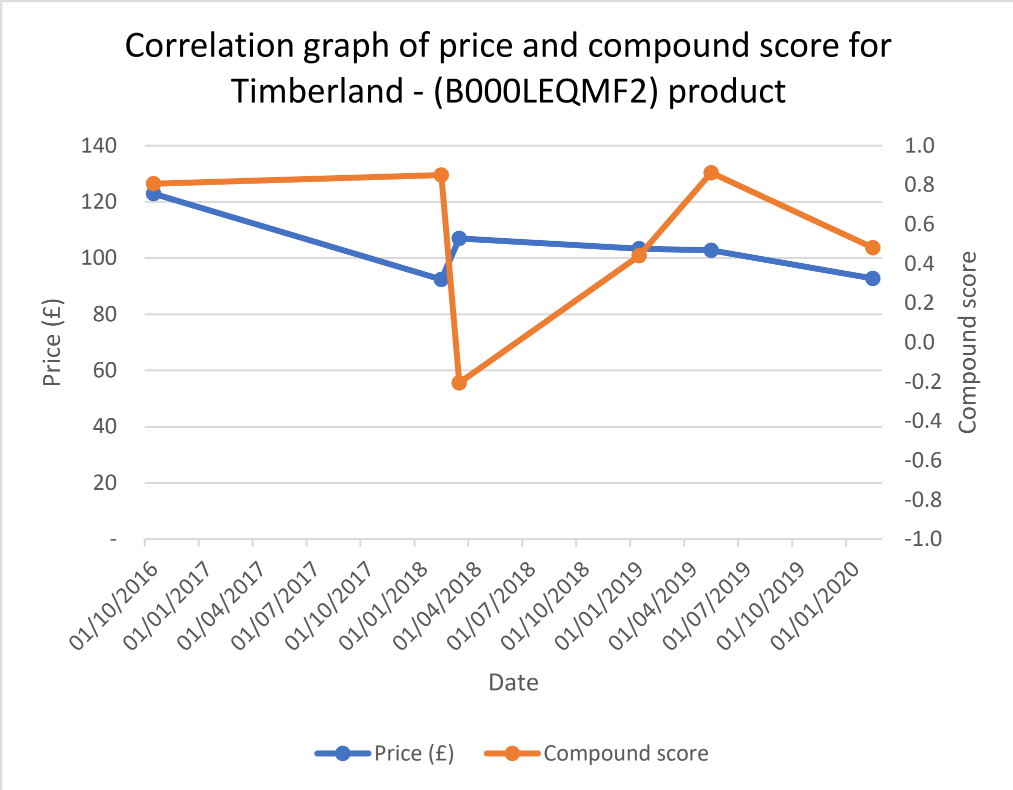


Figure A93: Correlation graph of price and compound score for Timberland B000LEQMF2 product

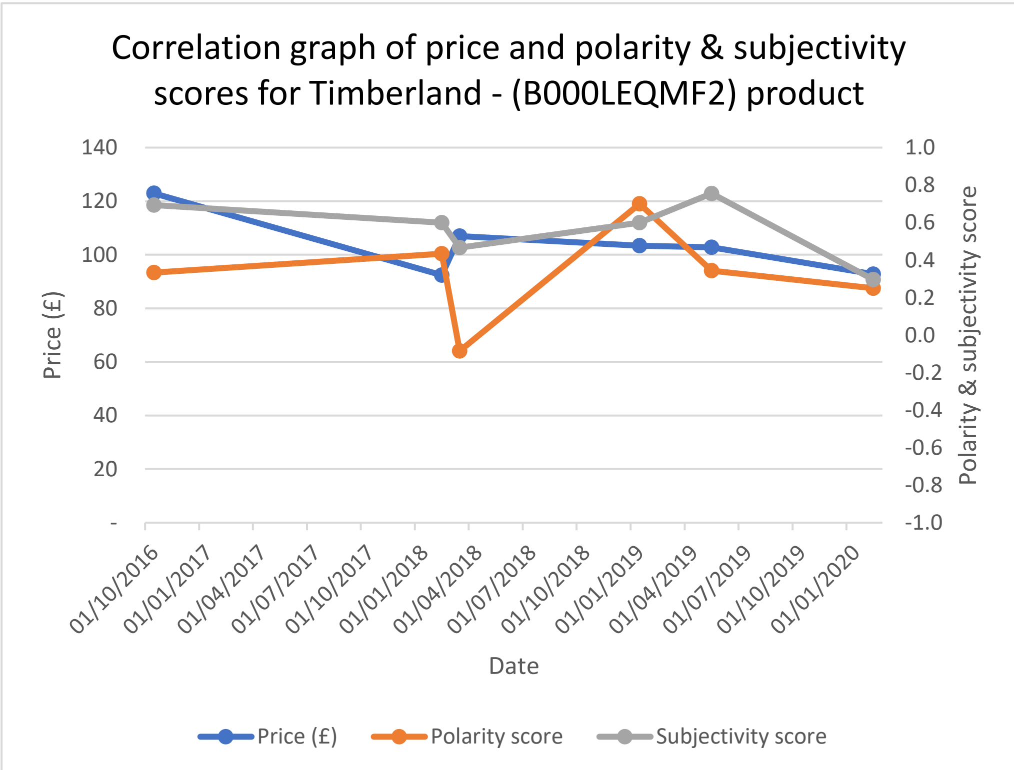


Figure A94: Correlation graph of price and polarity & subjectivity scores for Timberland B000LEQMF2 product

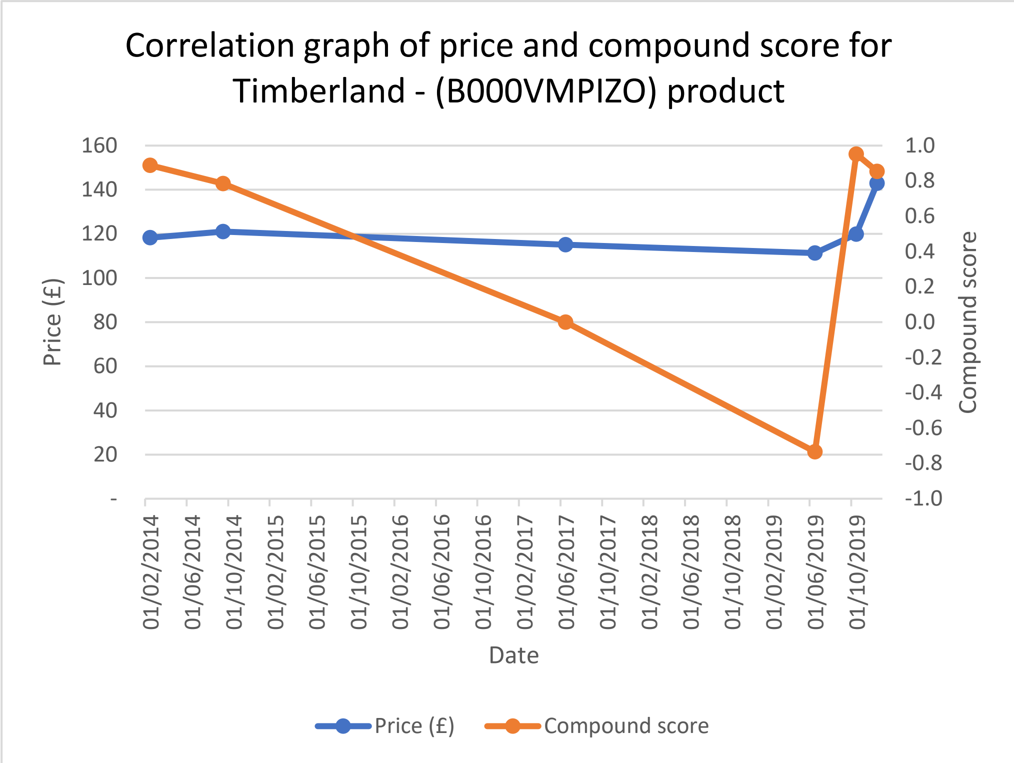


Figure A95: Correlation graph of price and compound score for Timberland B000VMPIZO product

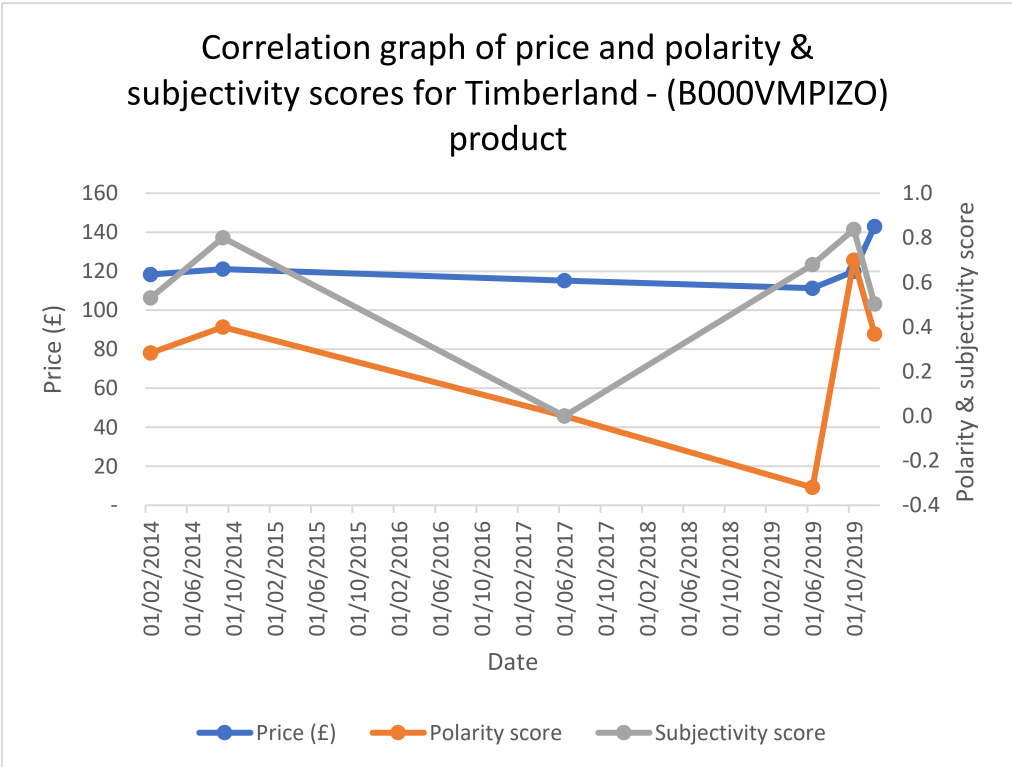


Figure A96: Correlation graph of price and polarity & subjectivity scores for Timberland B000VMPIZO product

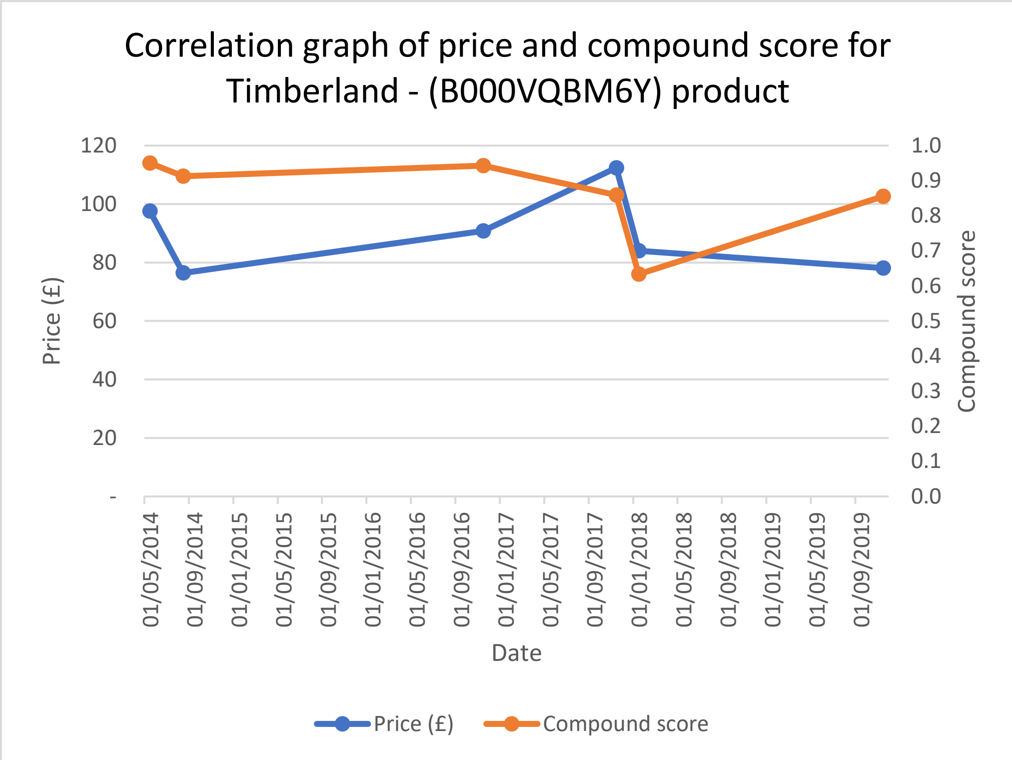


Figure A97: Correlation graph of price and compound score for Timberland B000VQBM6Y product

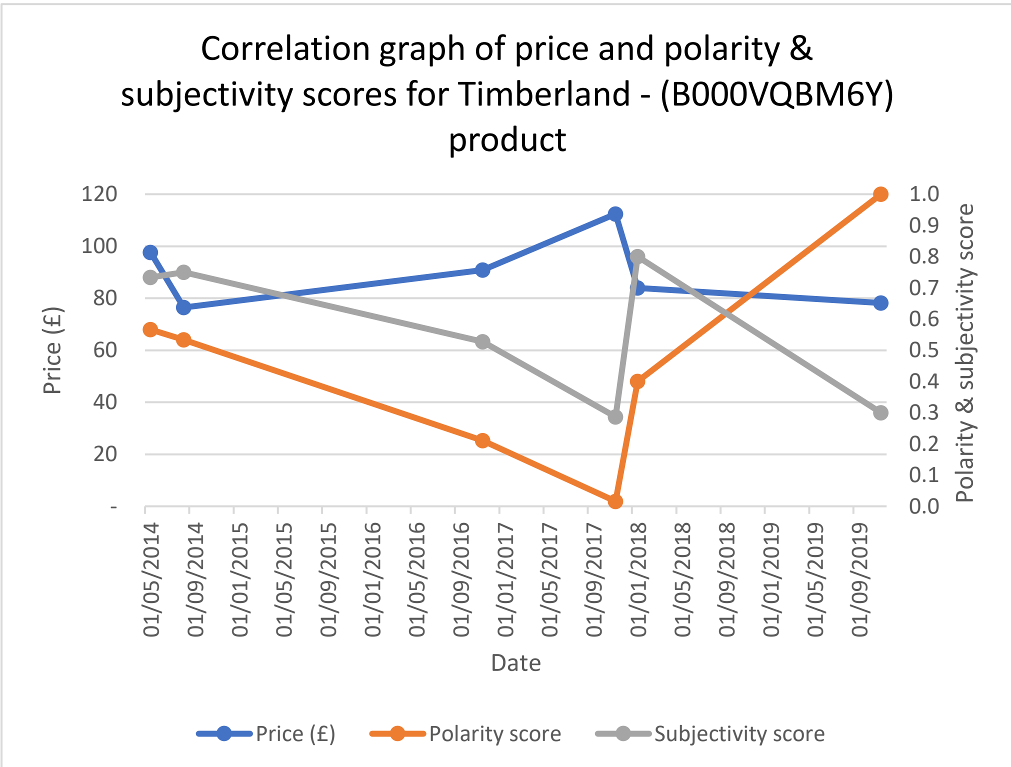


Figure A98: Correlation graph of price and polarity & subjectivity scores for Timberland B000VQBM6Y product

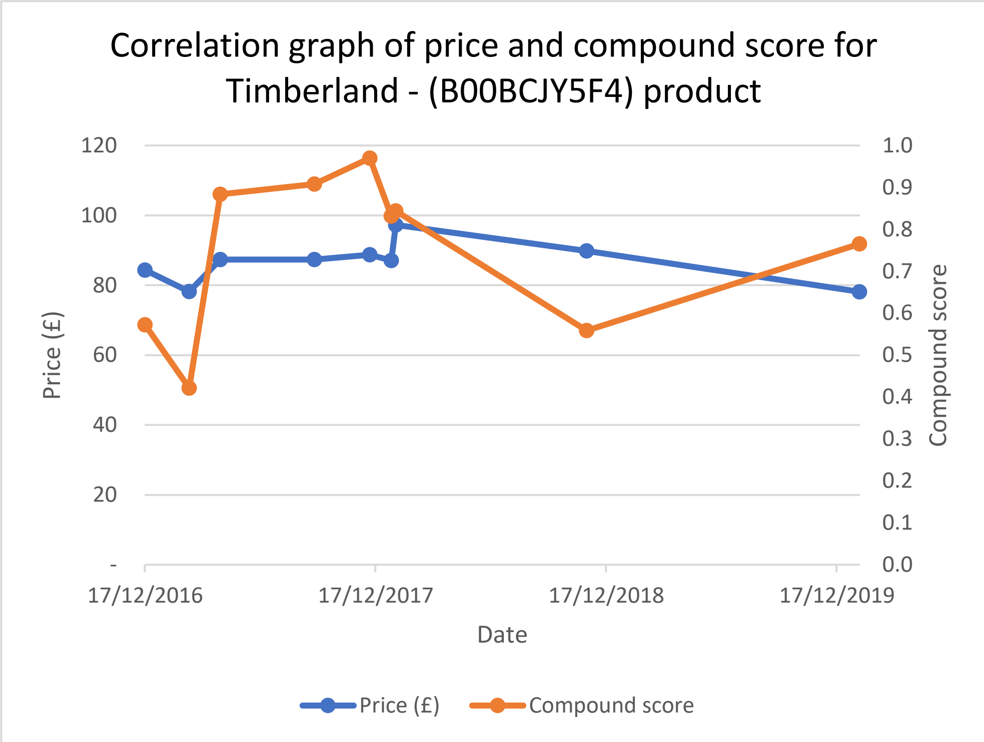


Figure A99: Correlation graph of price and compound score for Timberland B00BCJY5F4 product

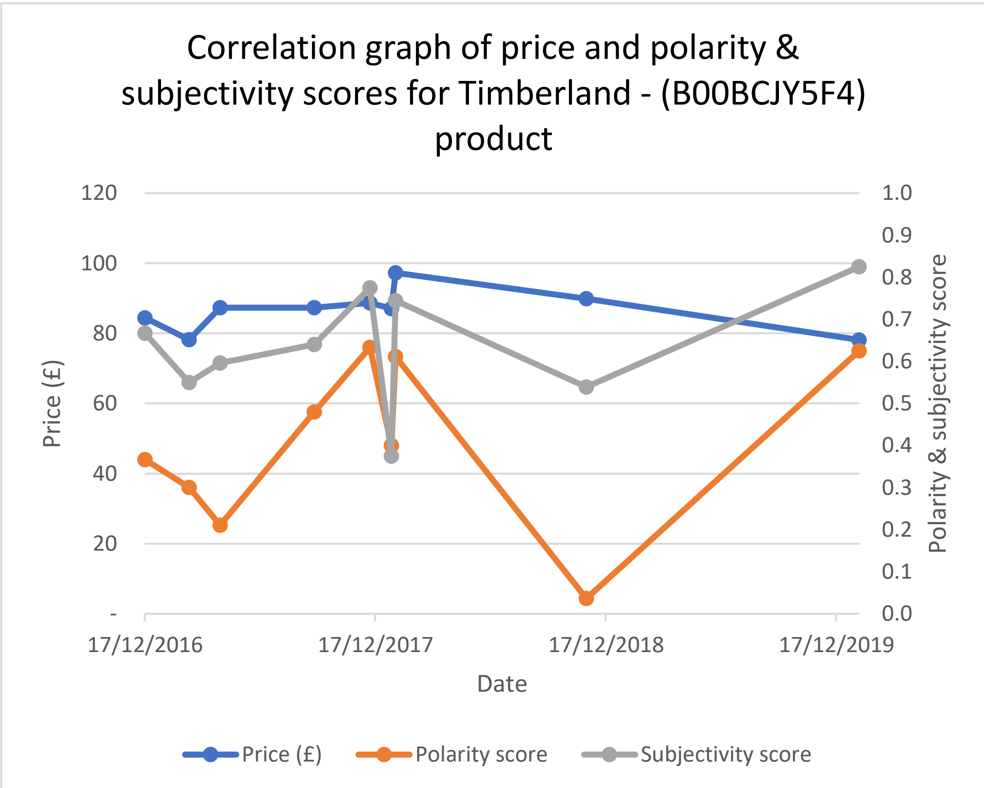


Figure A100: Correlation graph of price and polarity & subjectivity scores for Timberland B00BCJY5F4 product

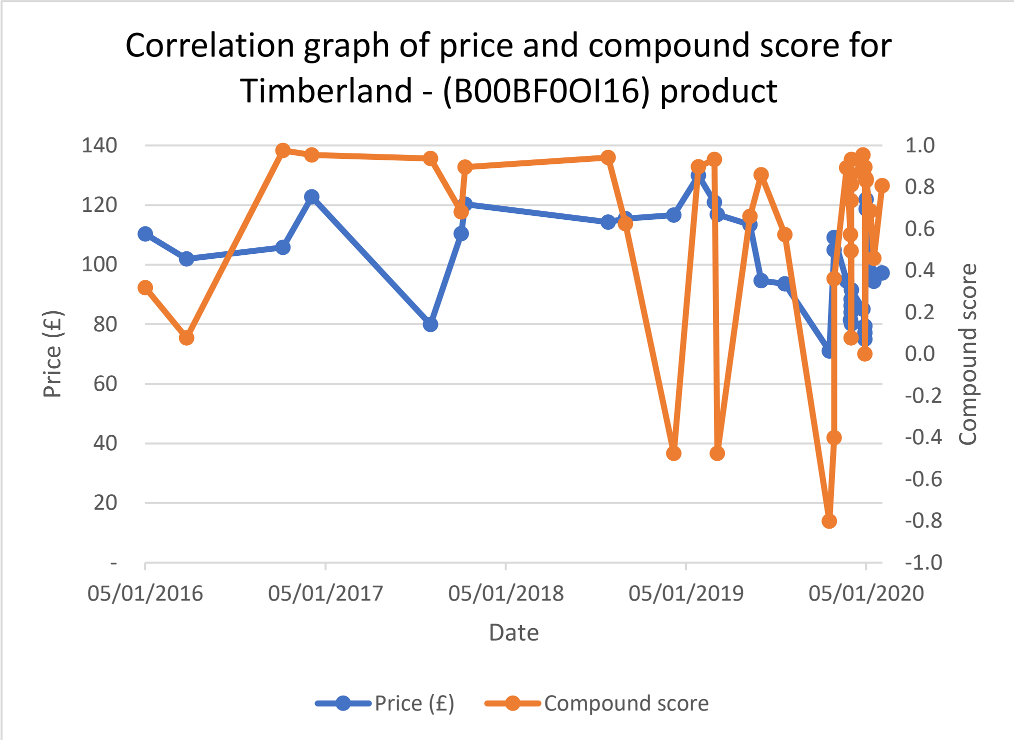


Figure A101: Correlation graph of price and compound score for Timberland B00BF0OI16 product

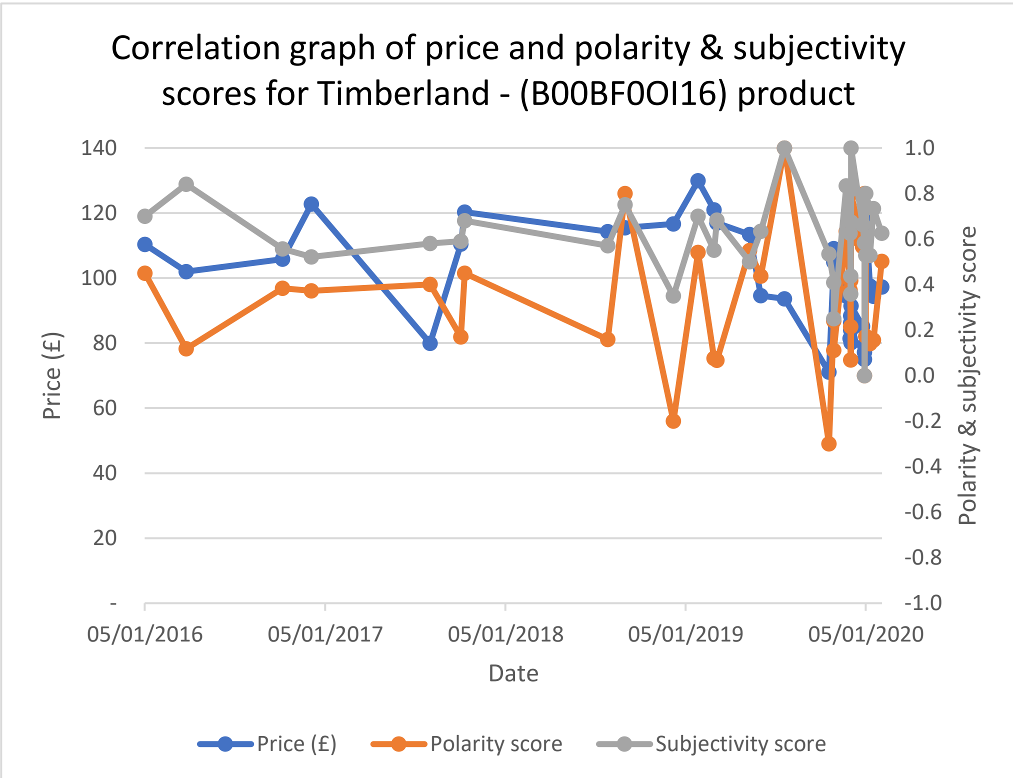


Figure A102: Correlation graph of price and polarity & subjectivity scores for Timberland B00BF0OI16 product

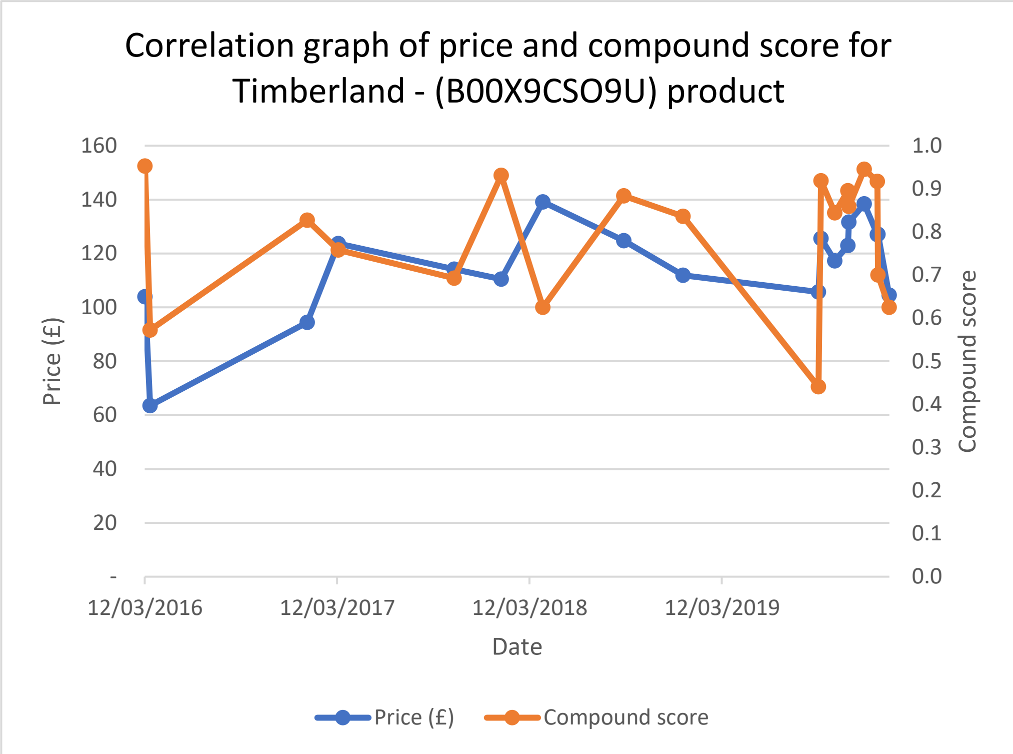


Figure A103: Correlation graph of price and compound score for Timberland B00X9CSO9U product

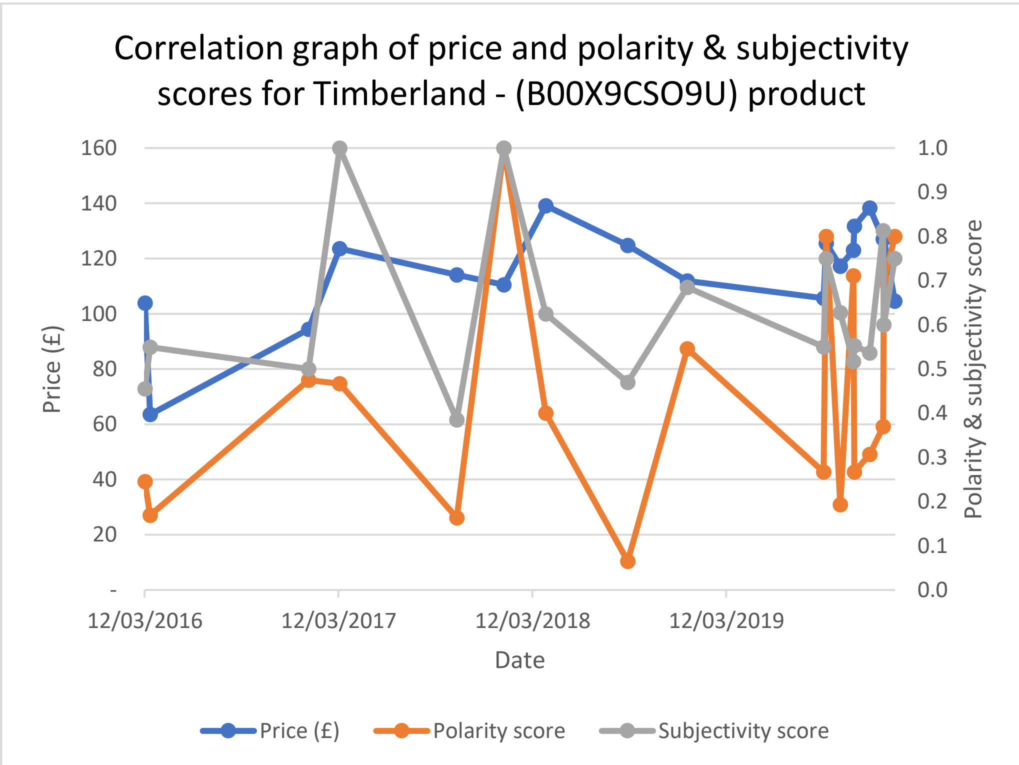


Figure A104 Correlation graph of price and polarity & subjectivity scores for Timberland B00X9CSO9U product

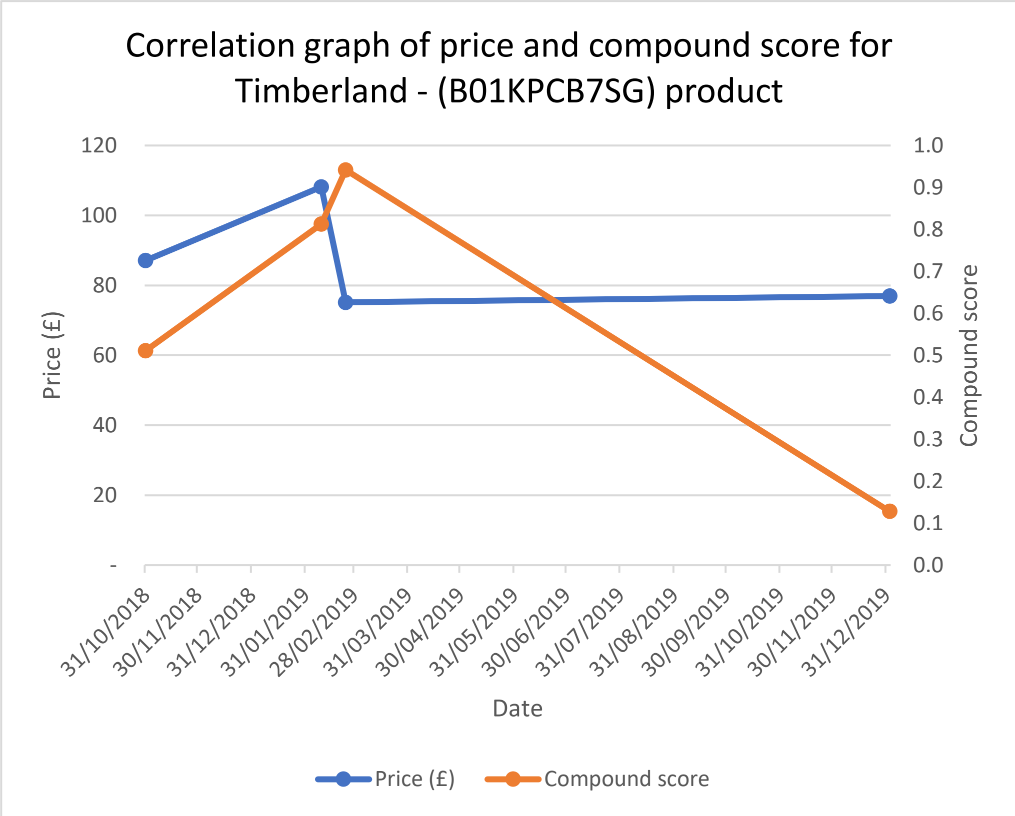


Figure A105: Correlation graph of price and compound score for Timberland B01KPCB7SG product

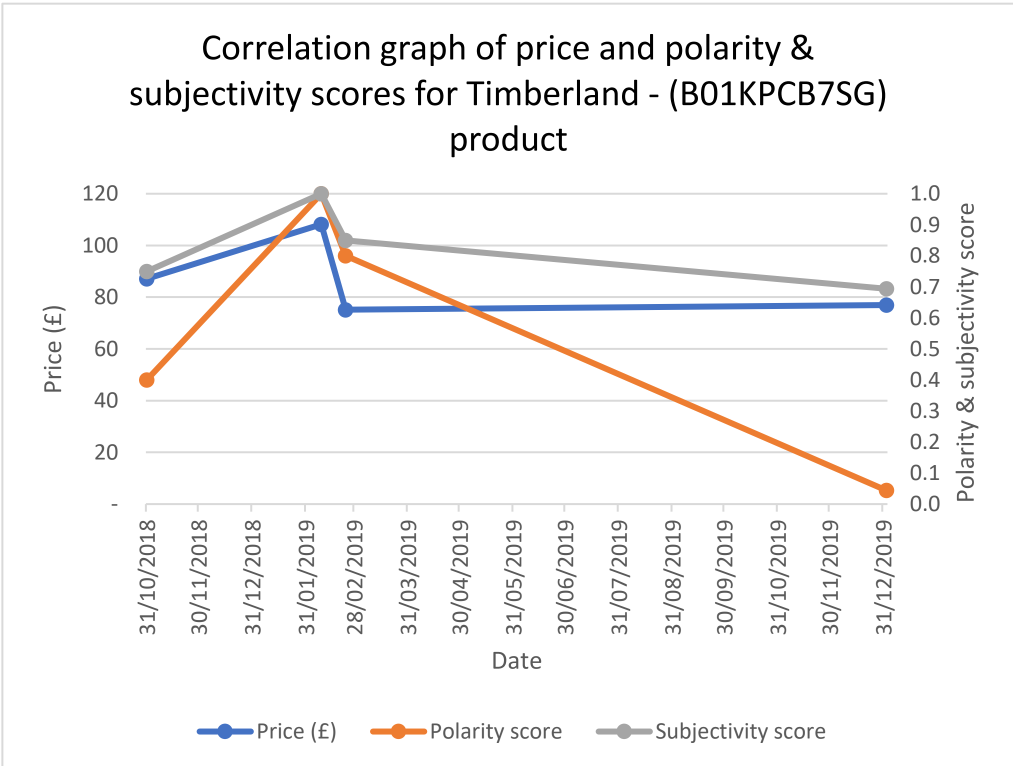


Figure A106: Correlation graph of price and polarity & subjectivity scores for Timberland B01KPCB7SG product

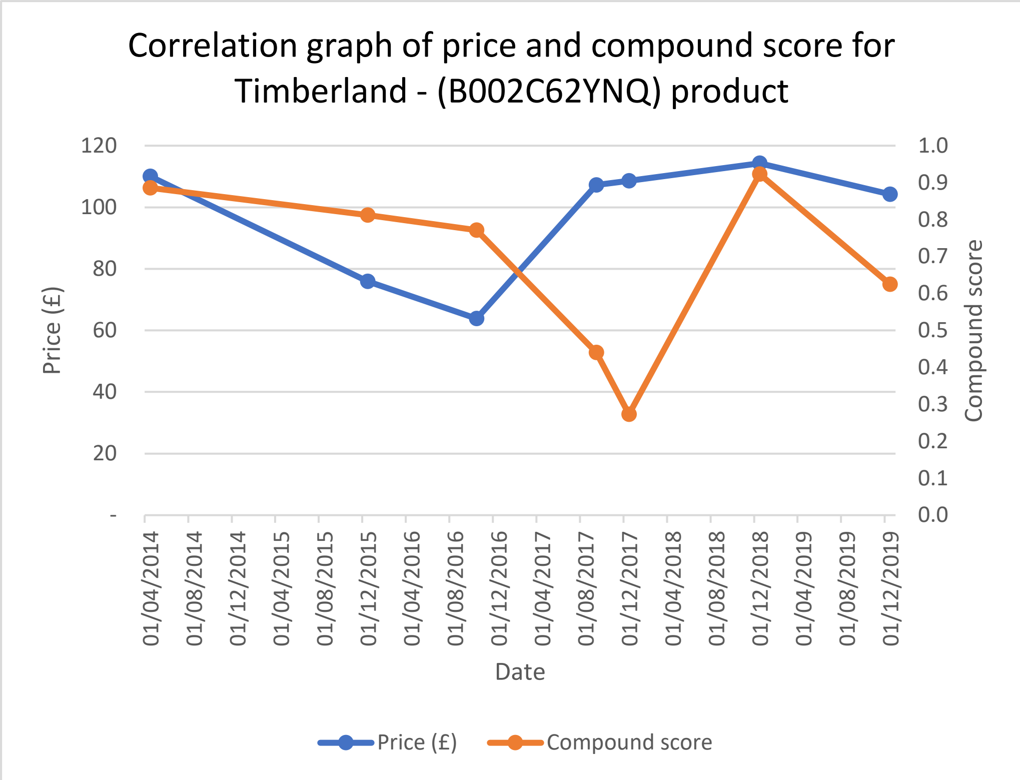


Figure A107: Correlation graph of price and compound score for Timberland B002C62YNQ product

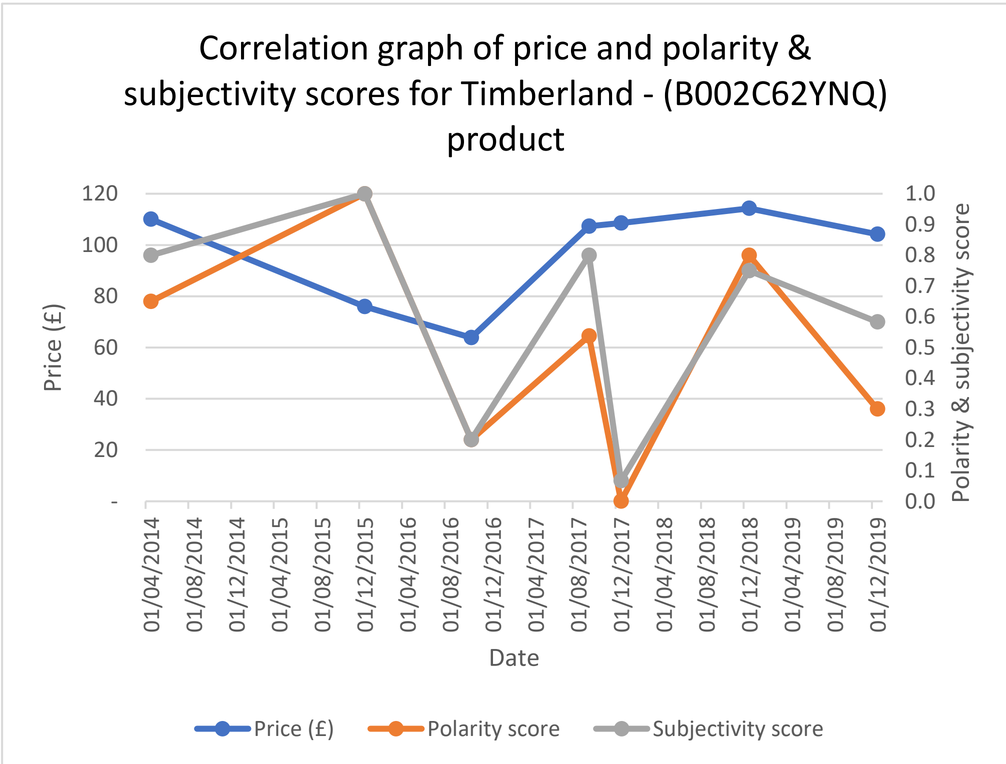


Figure A108: Correlation graph of price and polarity & subjectivity scores for Timberland B002C62YNQ product

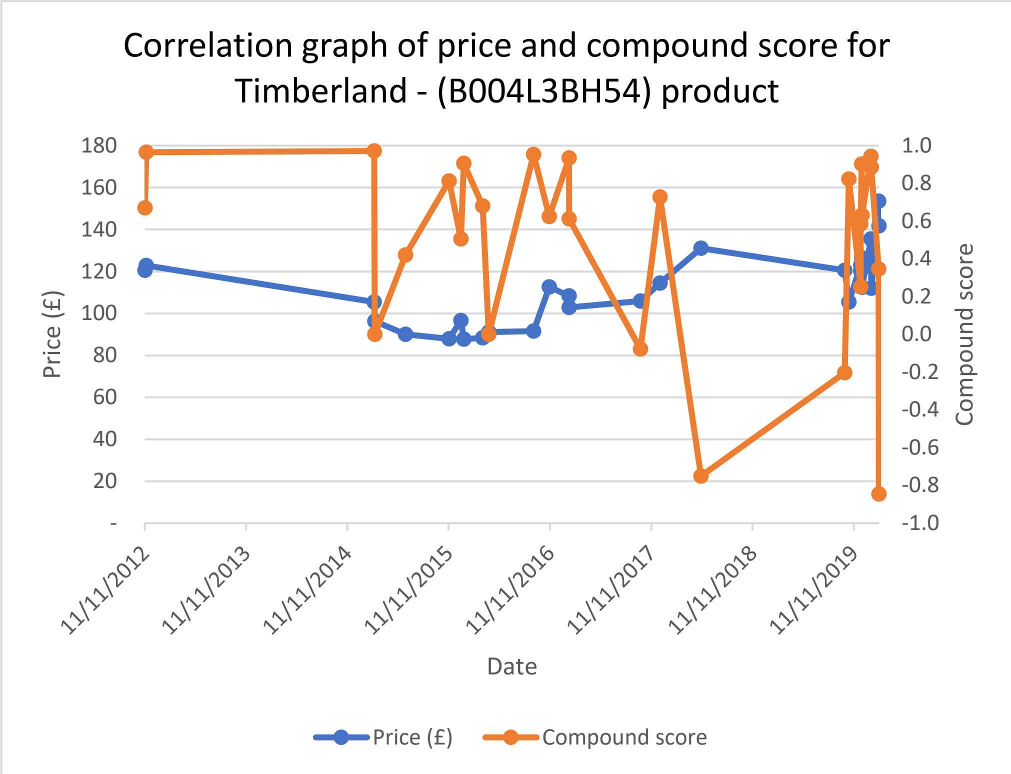


Figure A109: Correlation graph of price and compound score for Timberland B004L3BH54 product

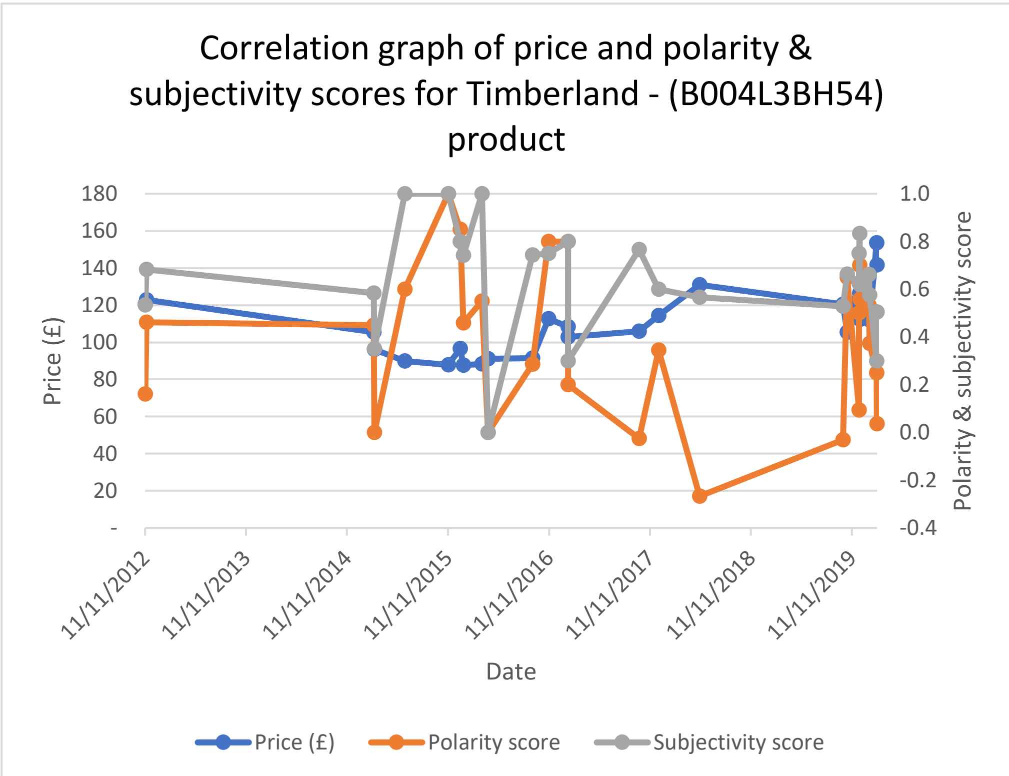


Figure A110: Correlation graph of price and polarity & subjectivity scores for Timberland B004L3BH54 product

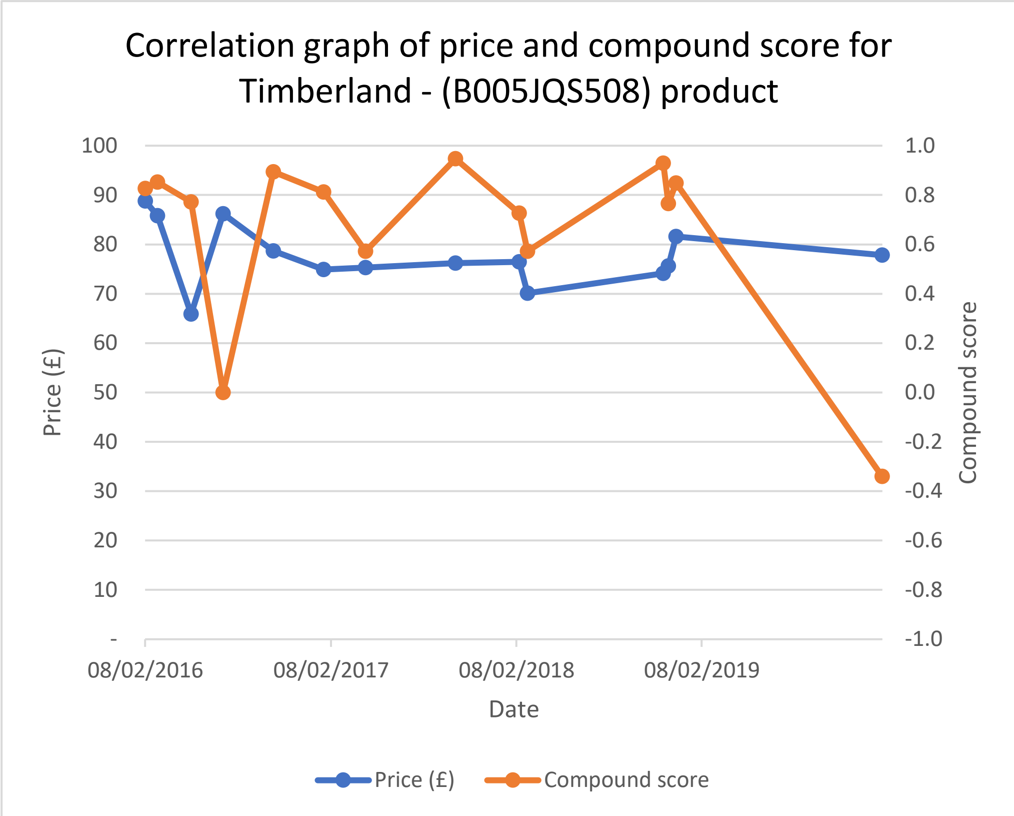


Figure A111: Correlation graph of price and compound score for Timberland B005JQS508 product

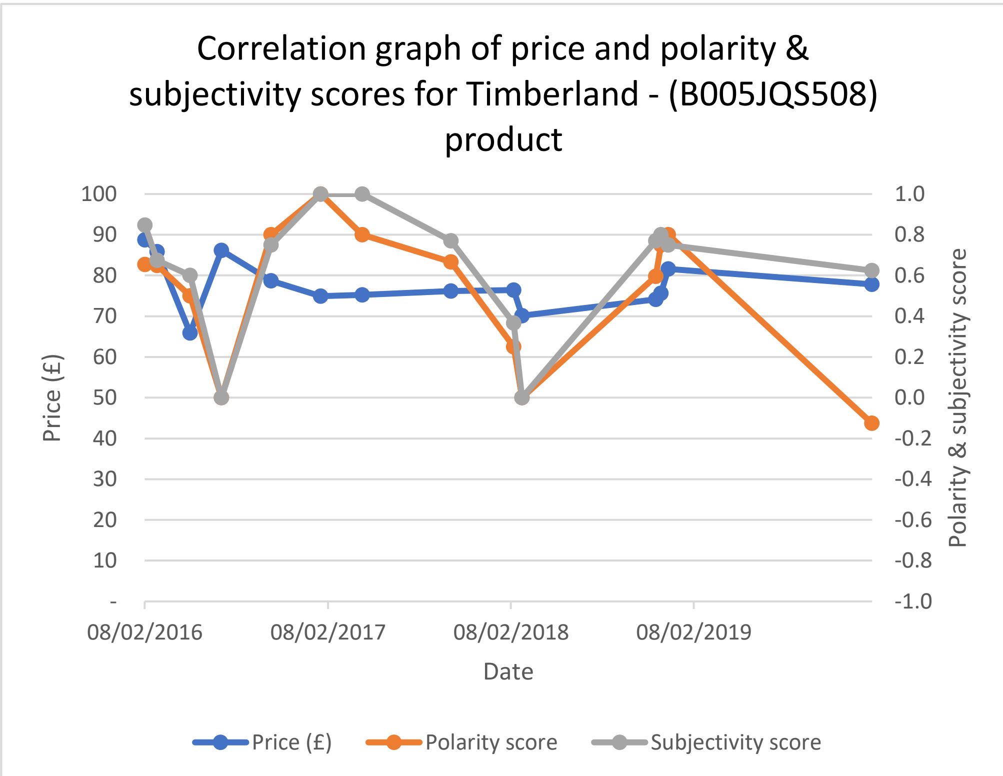


Figure A112: Correlation graph of price and polarity & subjectivity scores for Timberland B005JQS508 product

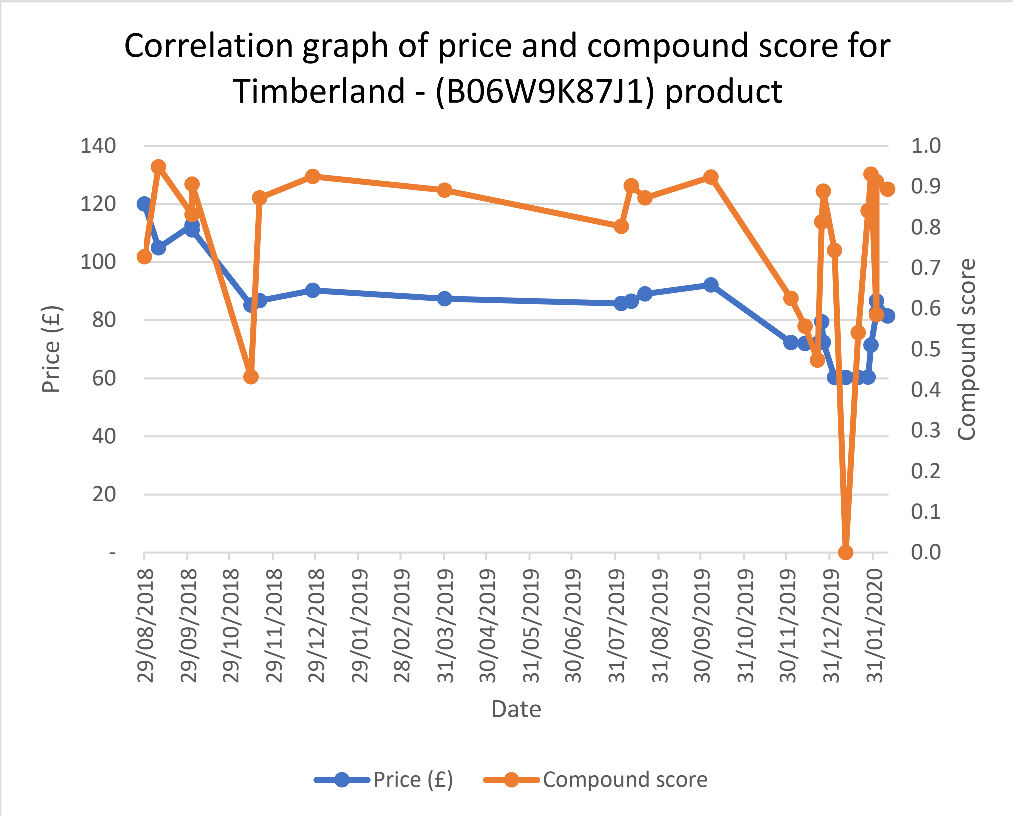


Figure A113: Correlation graph of price and compound score for Timberland B06W9K87J1 product

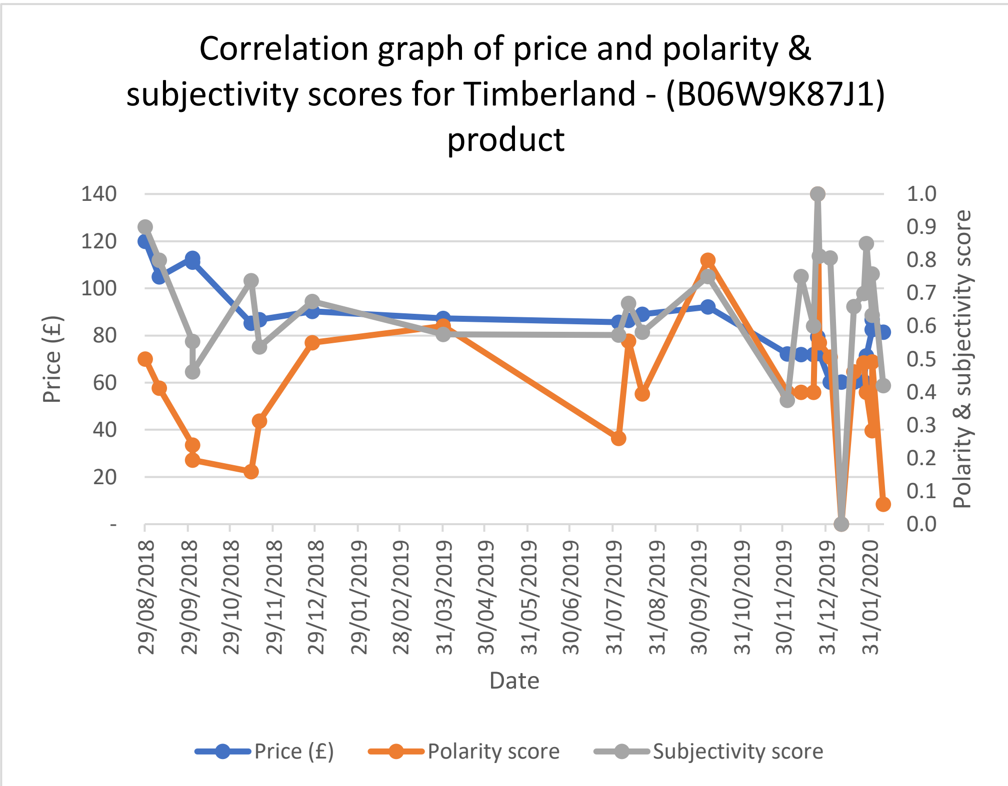


Figure A114: Correlation graph of price and polarity & subjectivity scores for Timberland B06W9K87J1 product

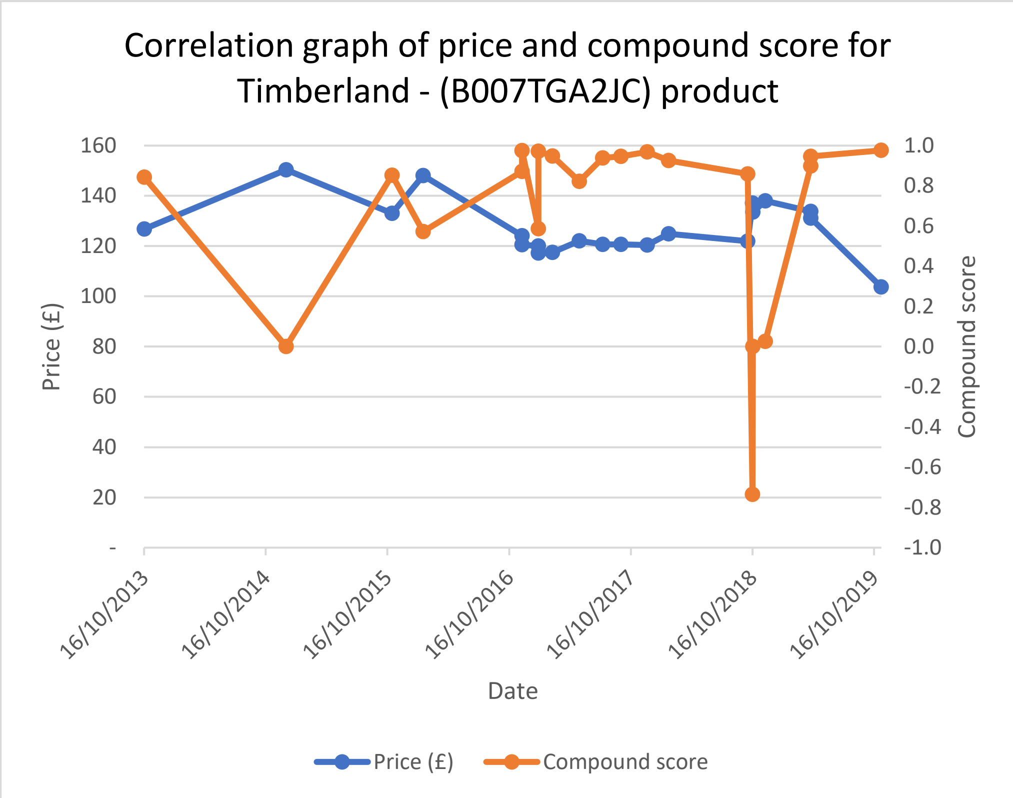


Figure A115: Correlation graph of price and compound score for Timberland B007TGA2JC product

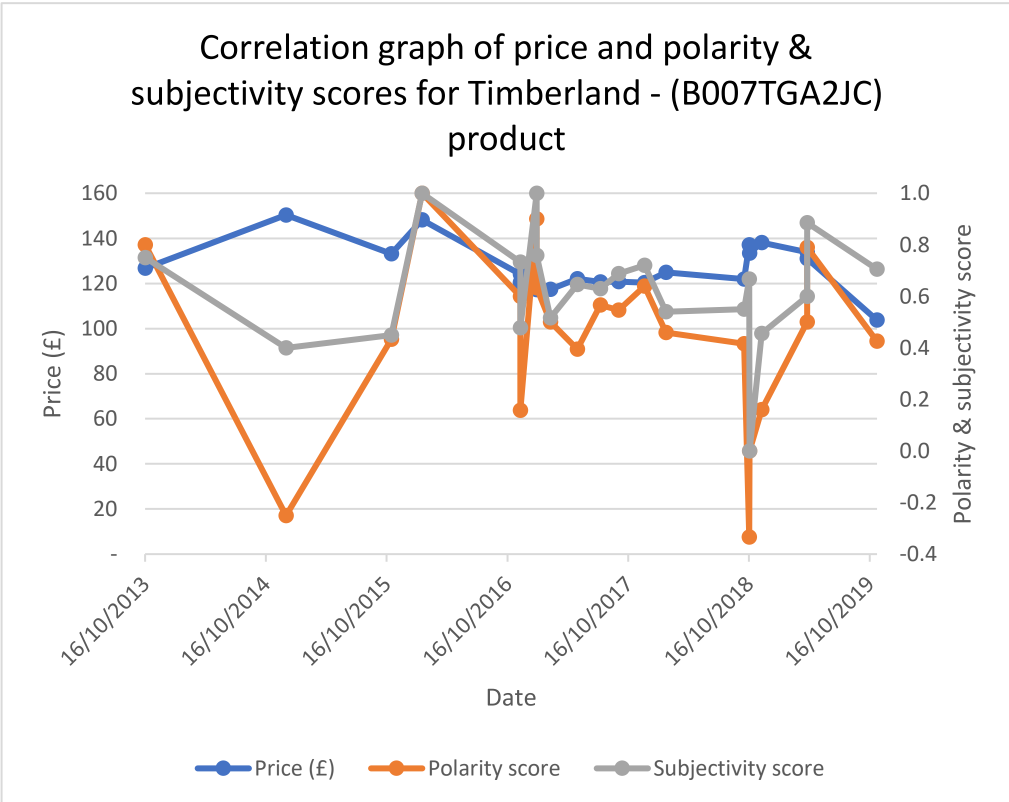


Figure A116: Correlation graph of price and polarity & subjectivity scores for Timberland B007TGA2JC product

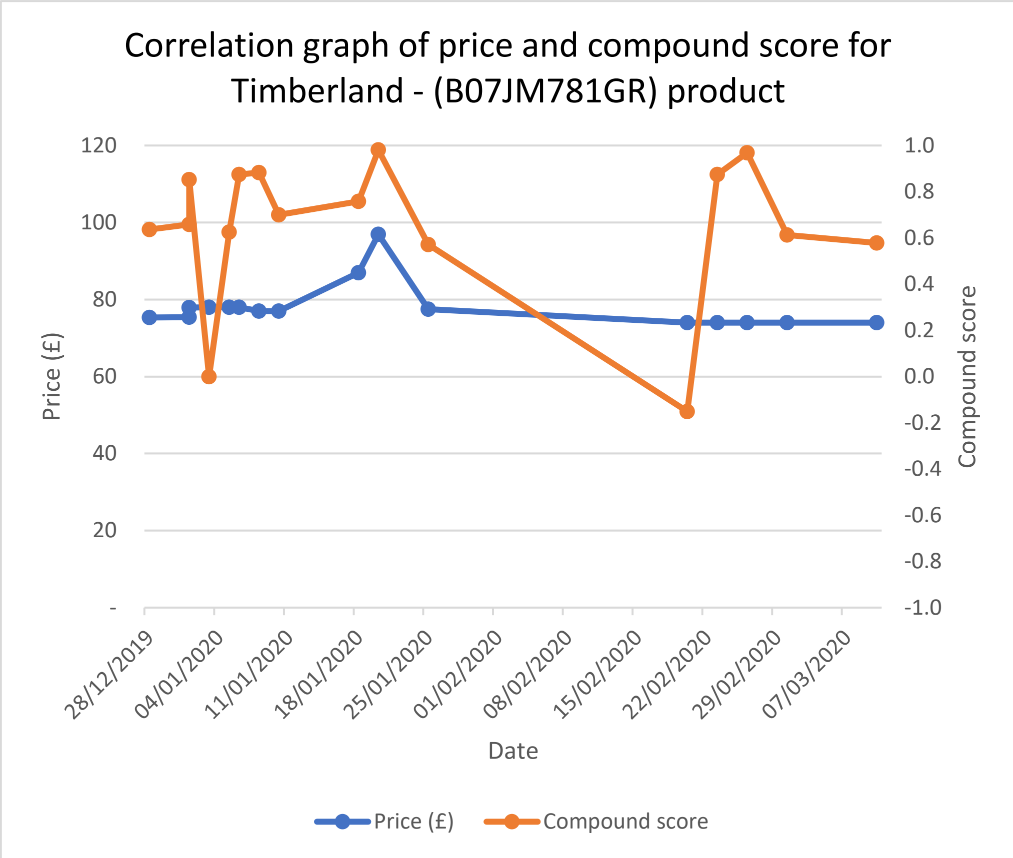


Figure A117:Correlation graph of price and compound score for Timberland B07JM781GR product

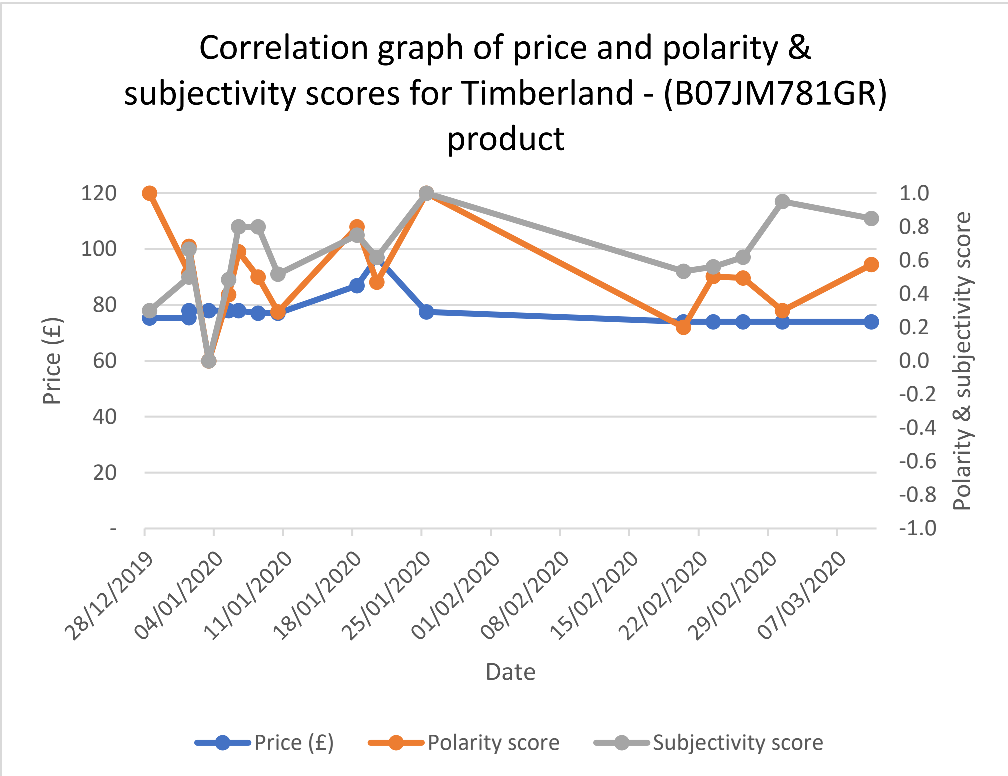


Figure A118: Correlation graph of price and polarity & subjectivity scores for Timberland B07JM781GR product

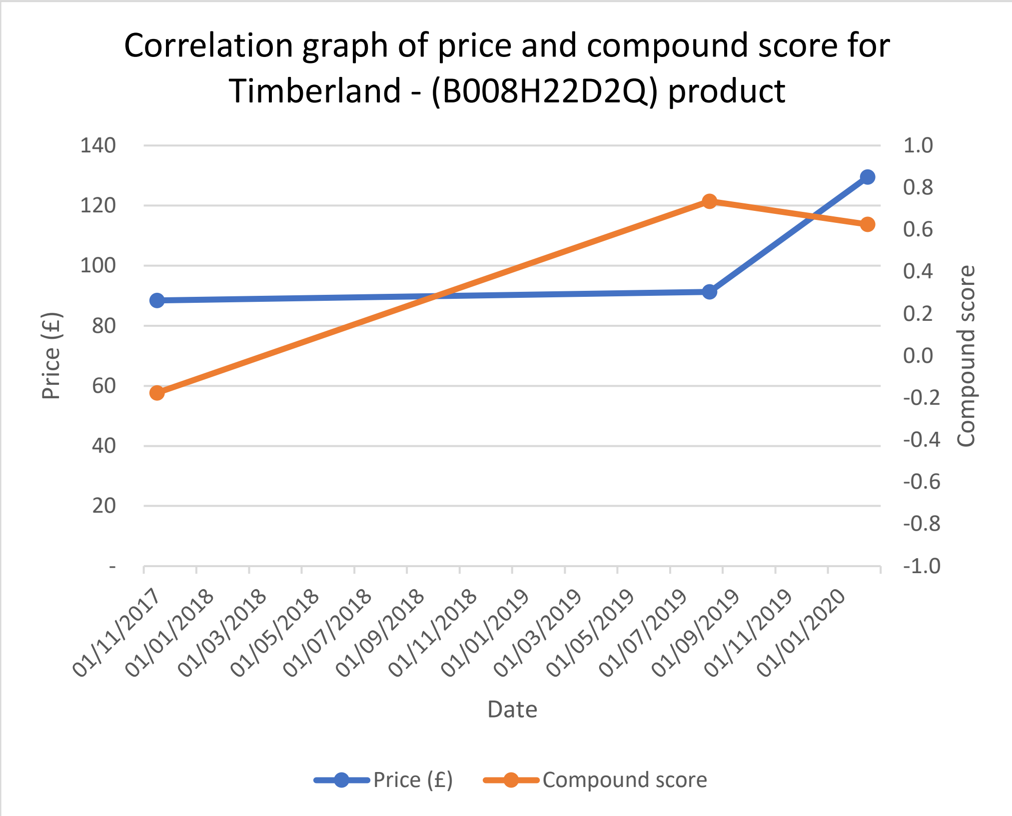


Figure A119: Correlation graph of price and compound score for Timberland B008H22D2Q product

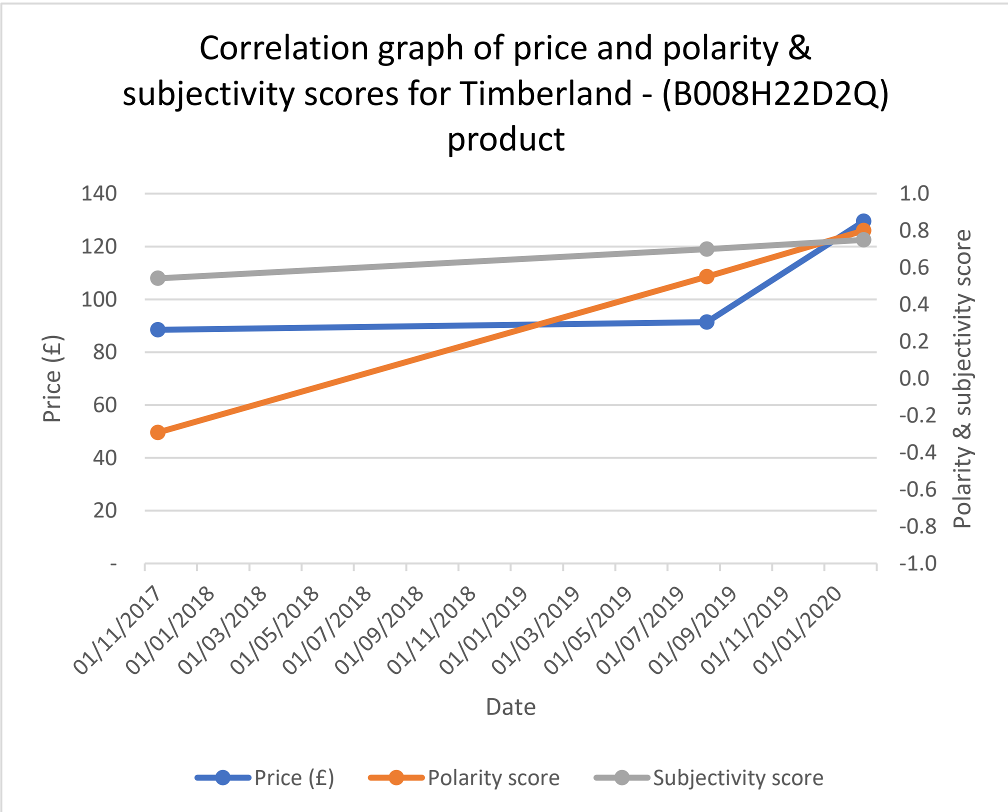


Figure A120: Correlation graph of price and polarity & subjectivity scores for Timberland B008H22D2Q product

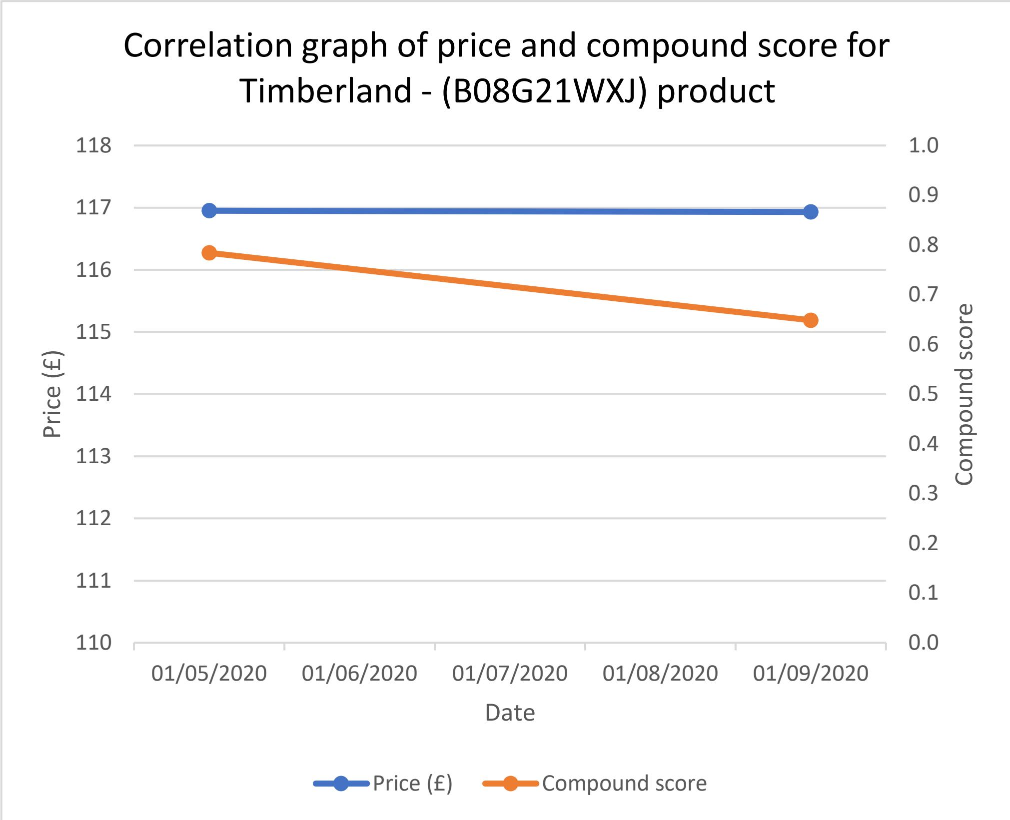


Figure A121: Correlation graph of price and compound score for Timberland B08G21WXJ product

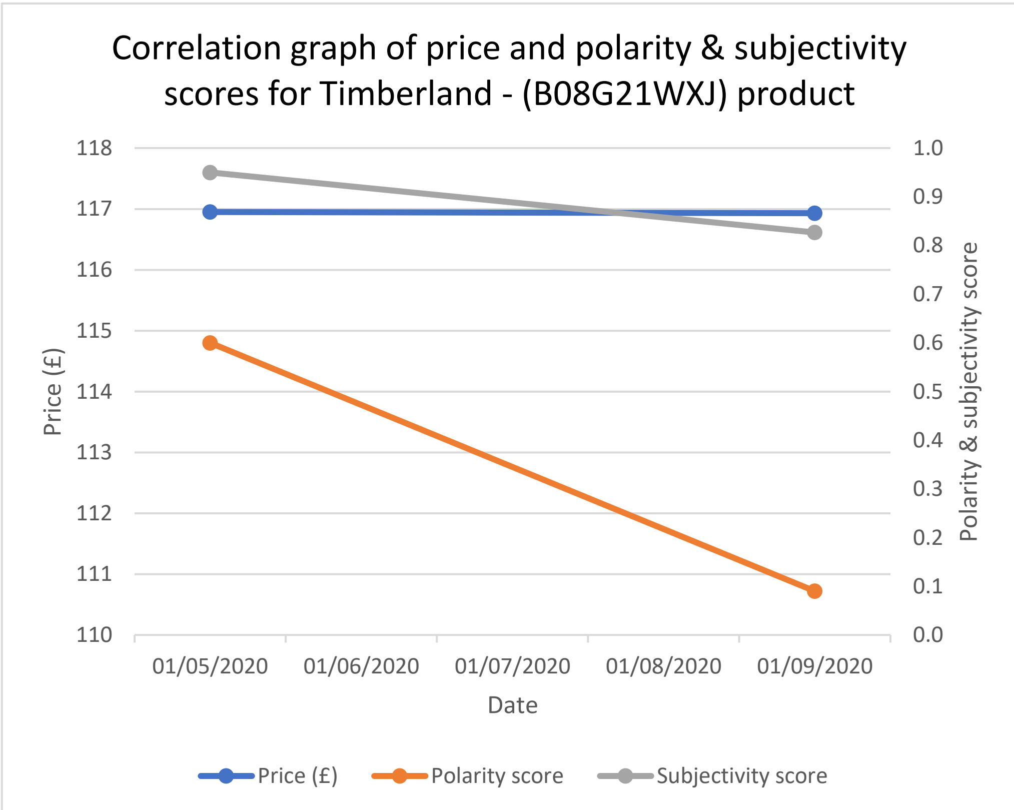


Figure A122: Correlation graph of price and polarity & subjectivity scores for Timberland B08G21WXJ product

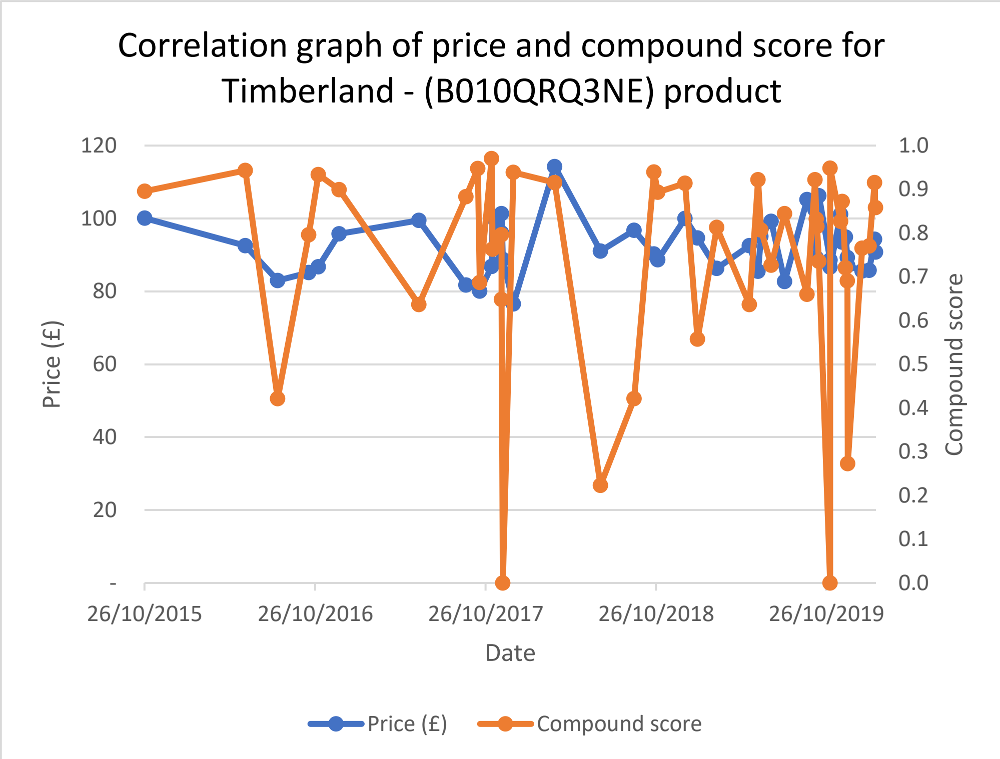


Figure A123: Correlation graph of price and compound score for Timberland B010QRQ3NE product

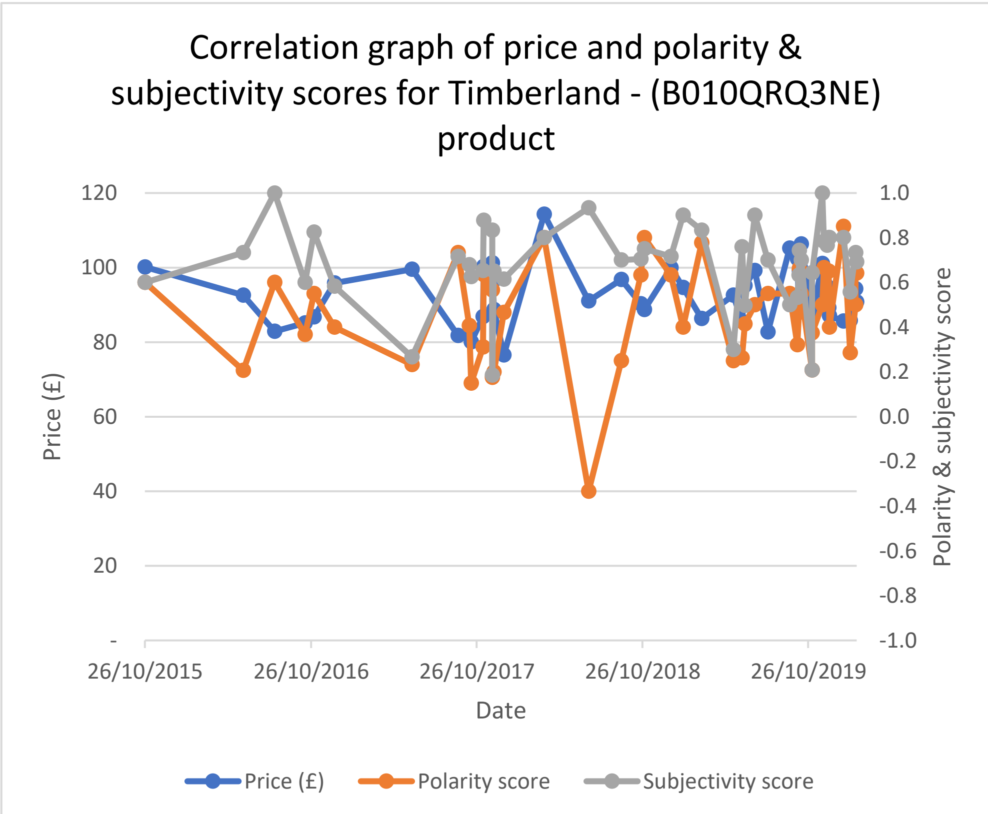


Figure A124: Correlation graph of price and polarity & subjectivity scores for Timberland B010QRQ3NE product

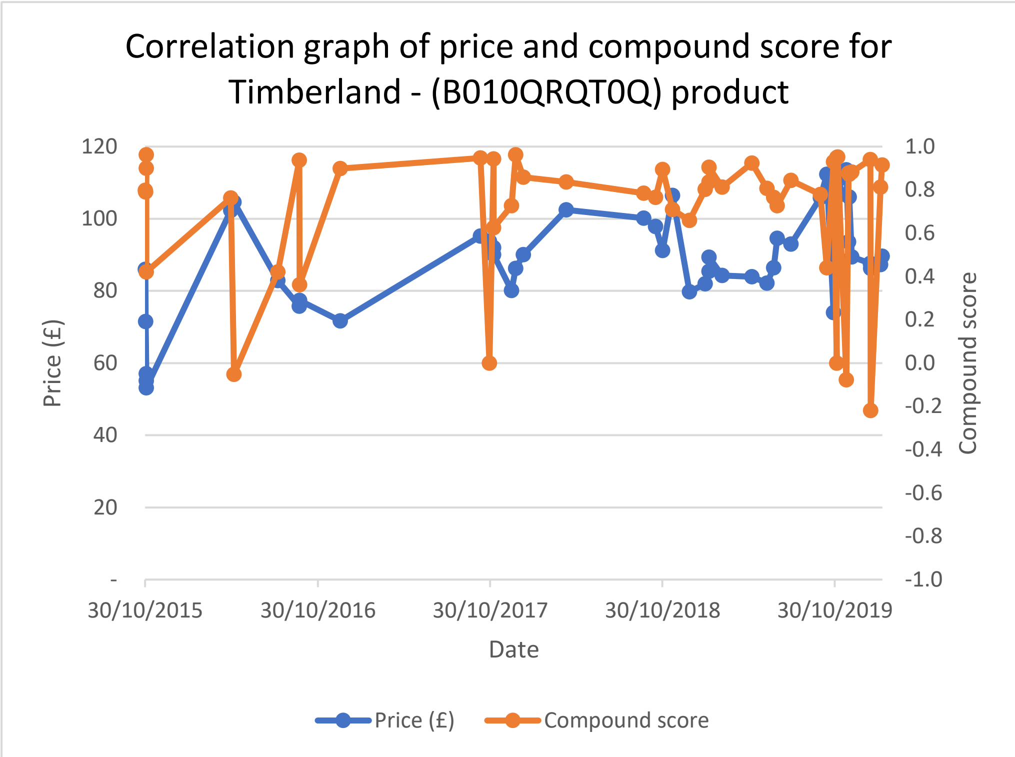


Figure A125: Correlation graph of price and compound score for Timberland B010QRQT0Q product

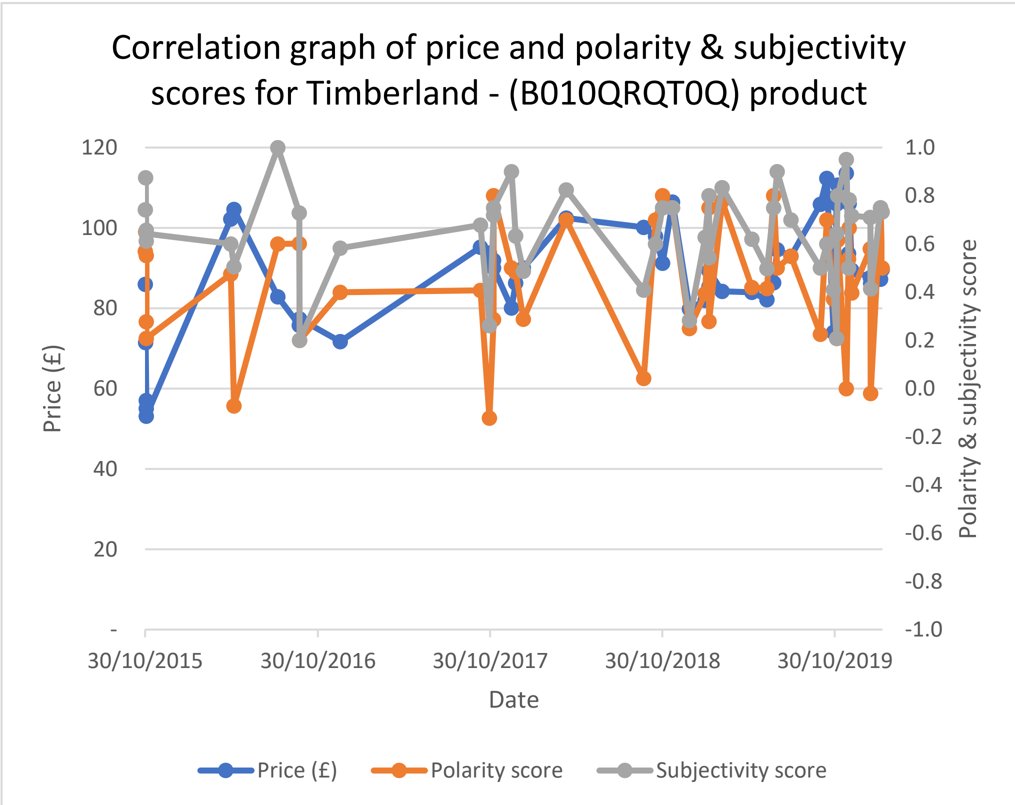


Figure A126: Correlation graph of price and polarity & subjectivity scores for Timberland B010QRQT0Q product

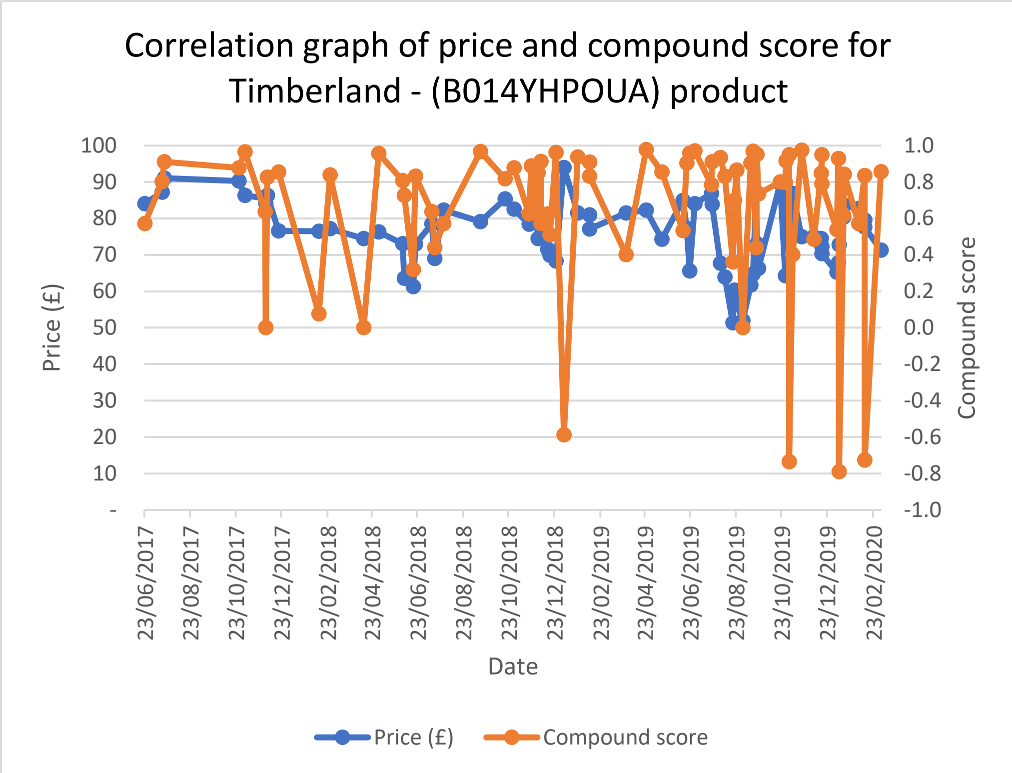


Figure A127: Correlation graph of price and compound score for Timberland B014YHPOUA product

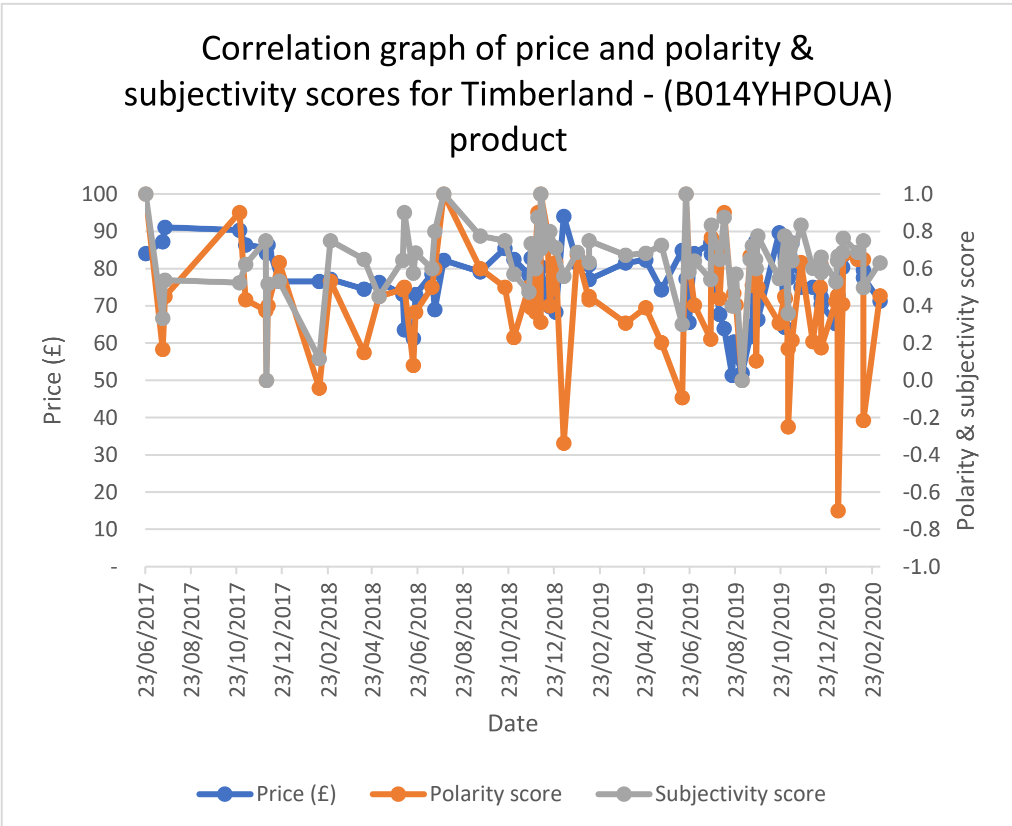


Figure A128: Correlation graph of price and polarity & subjectivity scores for Timberland B014YHPOUA product

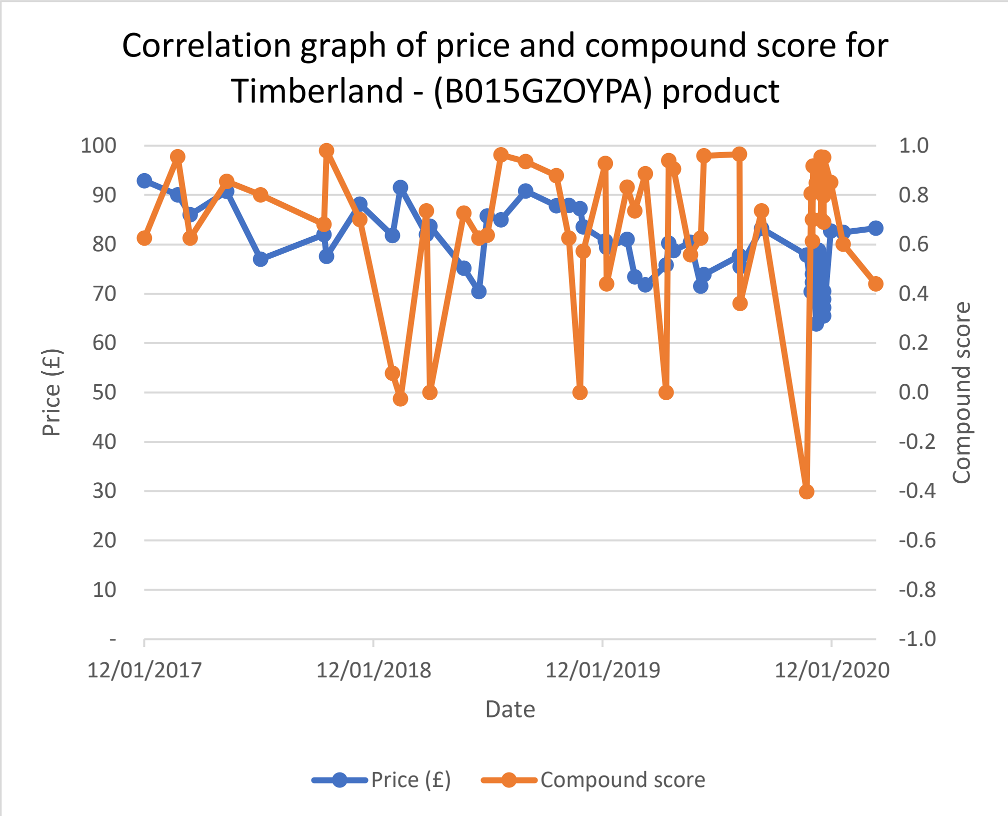


Figure A129: Correlation graph of price and compound score for Timberland B015GZOYPA product

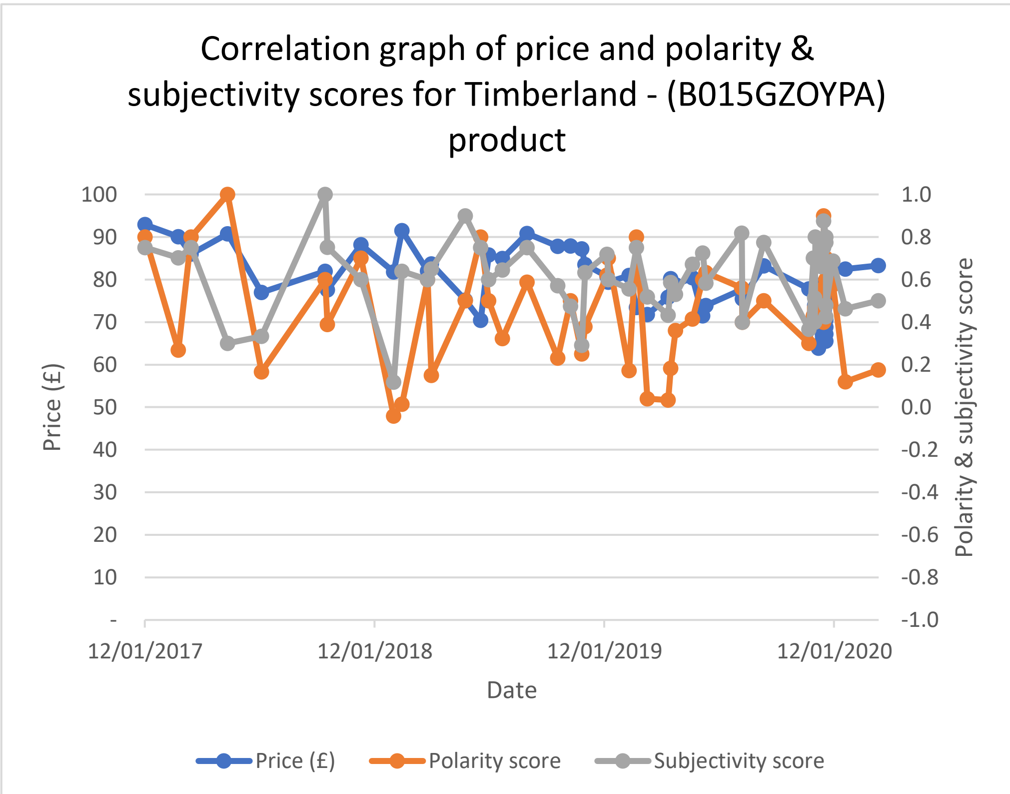


Figure A130: Correlation graph of price and polarity & subjectivity scores for Timberland B015GZOYPA product

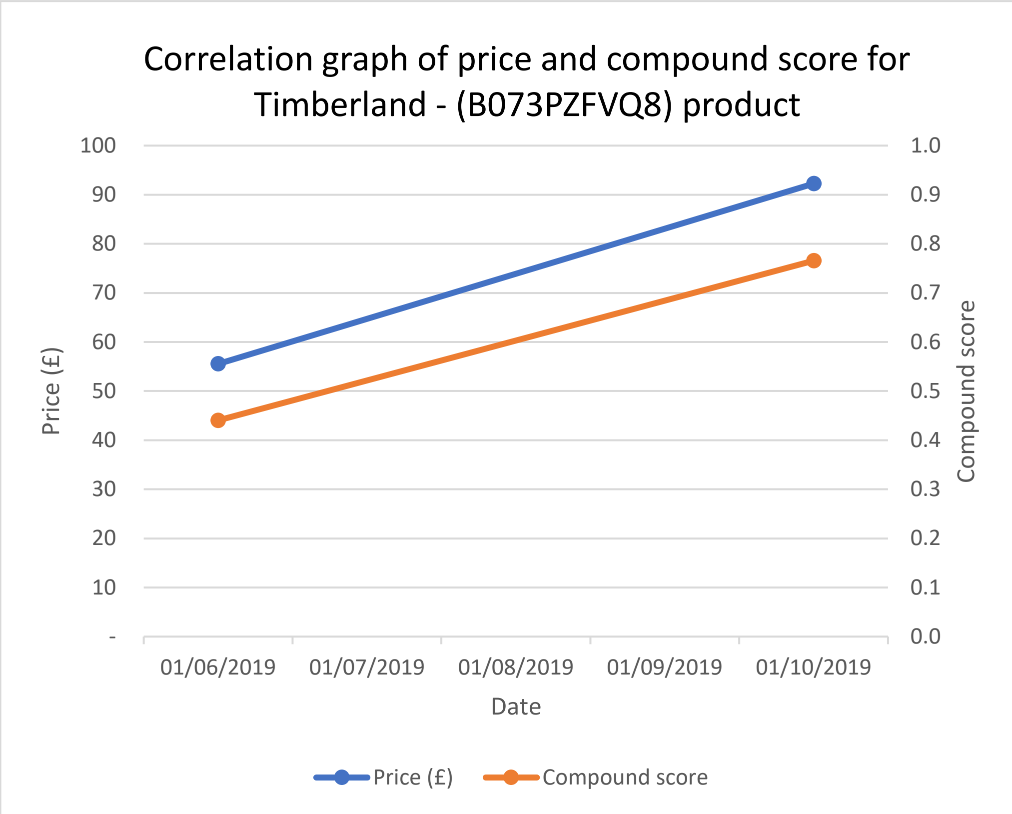


Figure A131: Correlation graph of price and compound score for Timberland B073PZFVQ8 product

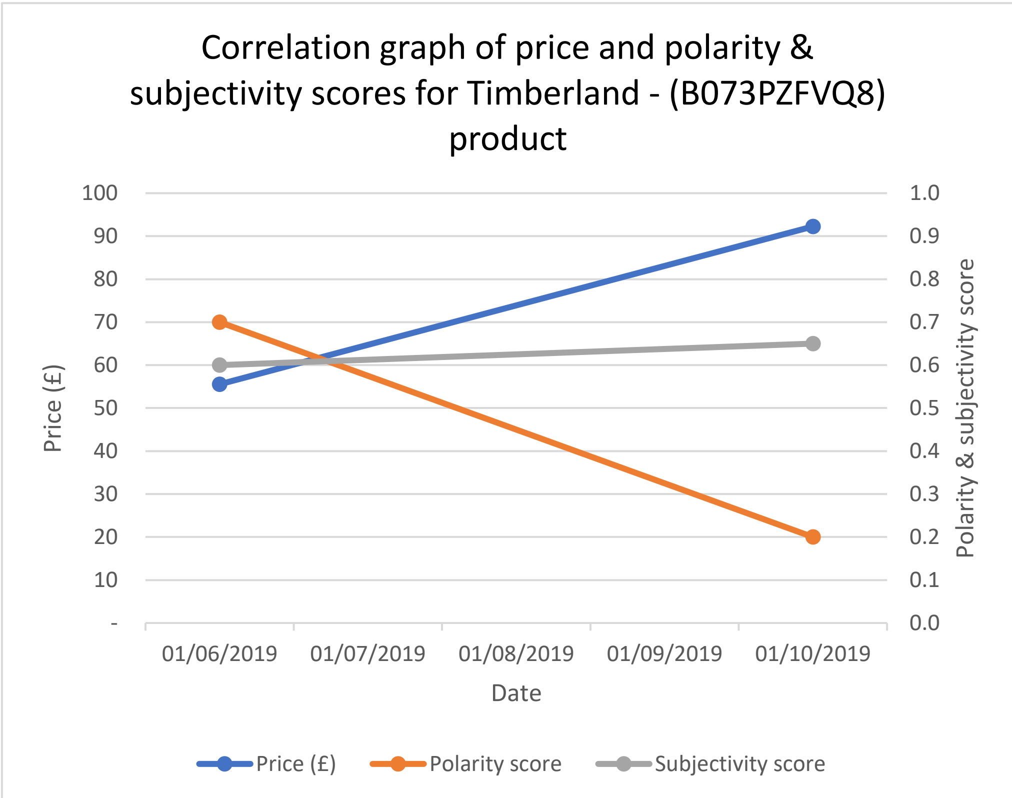


Figure A132: Correlation graph of price and polarity & subjectivity scores for Timberland B073PZFVQ8 product

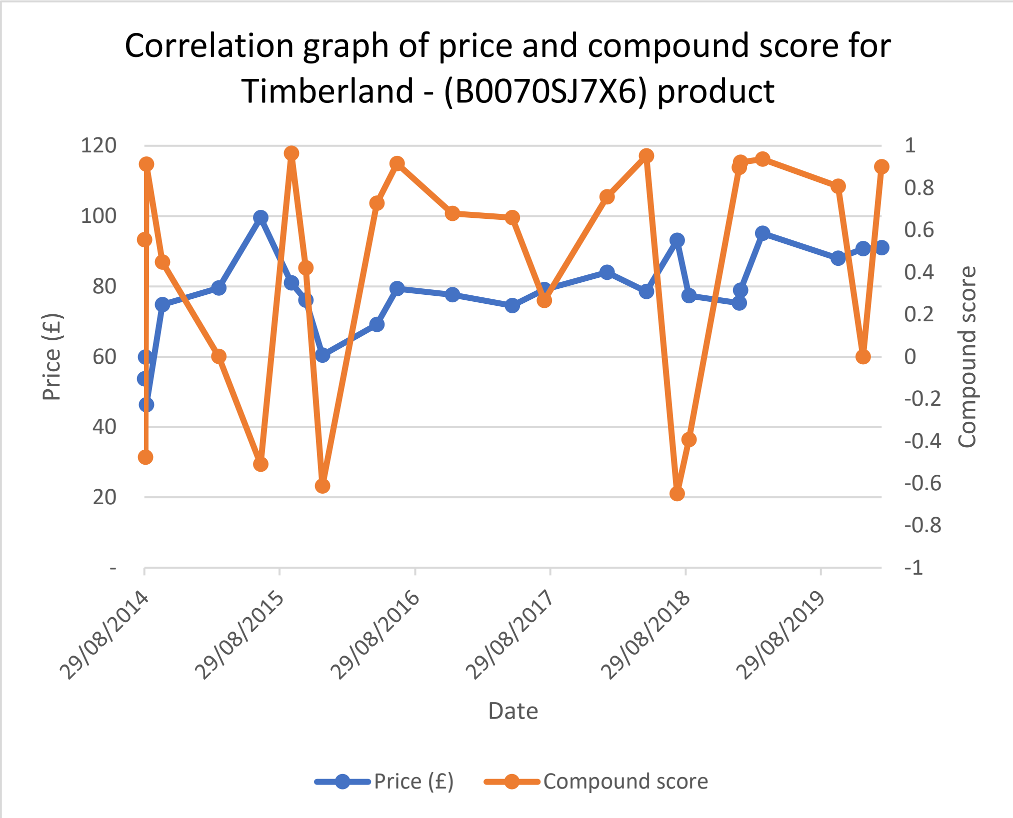


Figure A133: Correlation graph of price and compound score for Timberland B0070SJ7X6 product

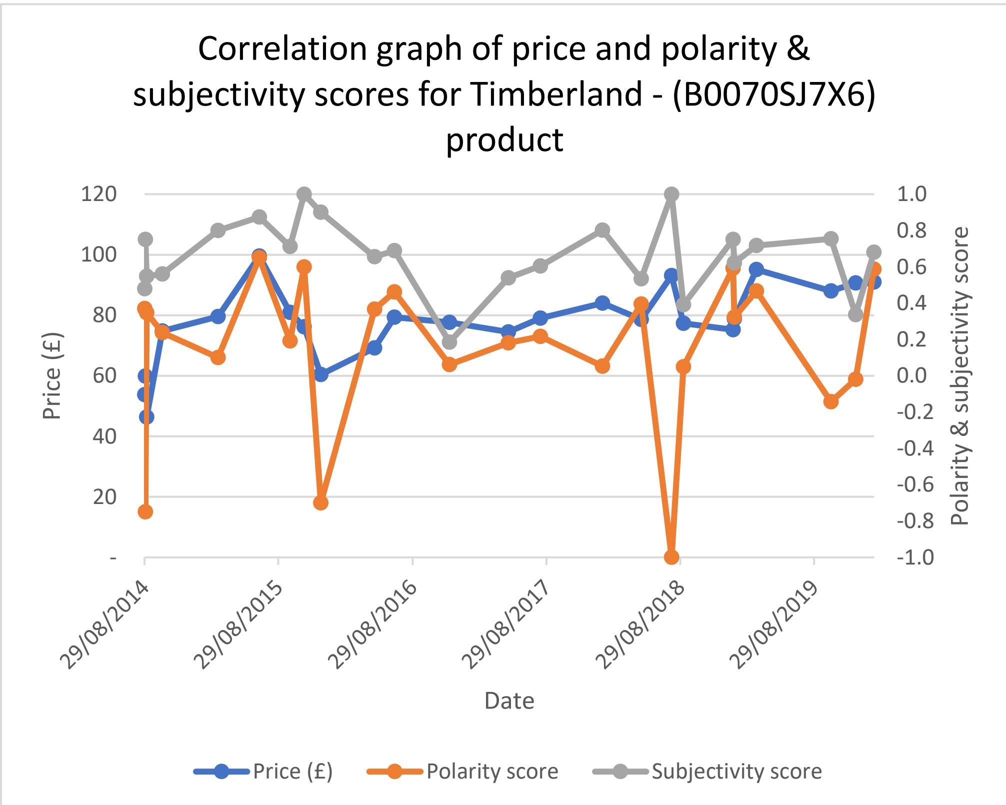


Figure A134: Correlation graph of price and polarity & subjectivity scores for Timberland B0070SJ7X6 product

Appendix 70 – ODD document

<u>ABM model Overview, Design concepts and Details (ODD) protocol</u>

An agent-based simulation approach to sentiment analysis in a dynamic marketplace: ABM model

Habiba Daroge, Anastasia Anagnostou, Crina Grosan

Full model description

Overview

Purpose

Entities

Process overview

Design concepts

Basic Principles

Emergence

Adaptation

Interaction

Stochasticity

Observation

Details

Initialisation

Input data

Sub-models

<u>Overview</u>

Purpose:

The agent-based model was developed to explore and simulate the interactions between consumers, sellers and innovators in a dynamic marketplace environment where there are constant price fluctuations. It will also run experiments for what-if scenario's and predict product adoption rate.

Entities:

The entities in the ABM model are consumers, sellers and innovators (agents).

Model key:

Blue patches = seller agents which represent Amazon sellers

White hollow circles = potential consumers

Yellow opaque circles = innovators

Red hollow circle = message received

Red opaque circle = decision-making

Orange opaque smiley, sad or neutral face = adopted the product

At initialisation of the model, blue patches do not move. Potential consumers represented by hollow white circles move around and interact with innovators represented by yellow opaque circles. The innovators as defined by Bass (1969) have already adopted the product and spread the message. Word-of-mouth plays a role here as whoever interacts with the innovators will receive the message. Once the consumer has received the message they will change into a hollow red circle. Each individual potential consumer possesses 2 types of thresholds. Firstly, if the random number is less than 0.01 for personal threshold, the agent adopts the product. However, if not, based on the Bass model (1969), if the agent has any neighbours which is indicated by social influence, and the random number is less than 0.5 times the fraction of the neighbours who have adopted, then the agent adopts the product. (Wilensky and Rand, 2015).

If both thresholds are satisfied, the consumer decides to adopt and changes into an orange smiley, sad or neutral face depending on input parameters.

The input parameters are as follows:

- No. potential consumers
- No. sellers
- No. innovators
- Sentiment
- Price

The input parameters can be modified by the user by dragging a slider to their preferred number in the NetLogo interface.

To run the simulation, the user inputs their desired measurements by adjusting the sliders for each of the input parameters. The setup button sets the simulation to run. The go button is a forever button which will run until all conditions are satisfied. In this case, the simulation will stop once all consumers have turned an orange colour representing all have adopted the product. The step button allows the user to run the simulation and see interactions step-by-step. The user can adjust at which speed the model runs by dragging the slider for speed. The simulation model is measured in ticks which is a measure of time in the simulation model.

The graph visualizes the count of potential consumers, message received and adoption. There is also a graph which looks specifically at adoption rate. This is useful as it shows the rate at which consumers adopt a product over time.

Counters may be helpful to measure the numbers of the different parameters.

As the model runs, message received, decision-making and adoption status are updated. The decision-making stage involves agent thresholds.

All agents are characterised as having adopted or not adopted. The model will run until all agents have adopted the product. One of the assumptions of the Bass model (1969) is all consumers will eventually adopt the product.

Process overview:

Figure A1 shows the conceptualisation of the processes occurring in sequential order.

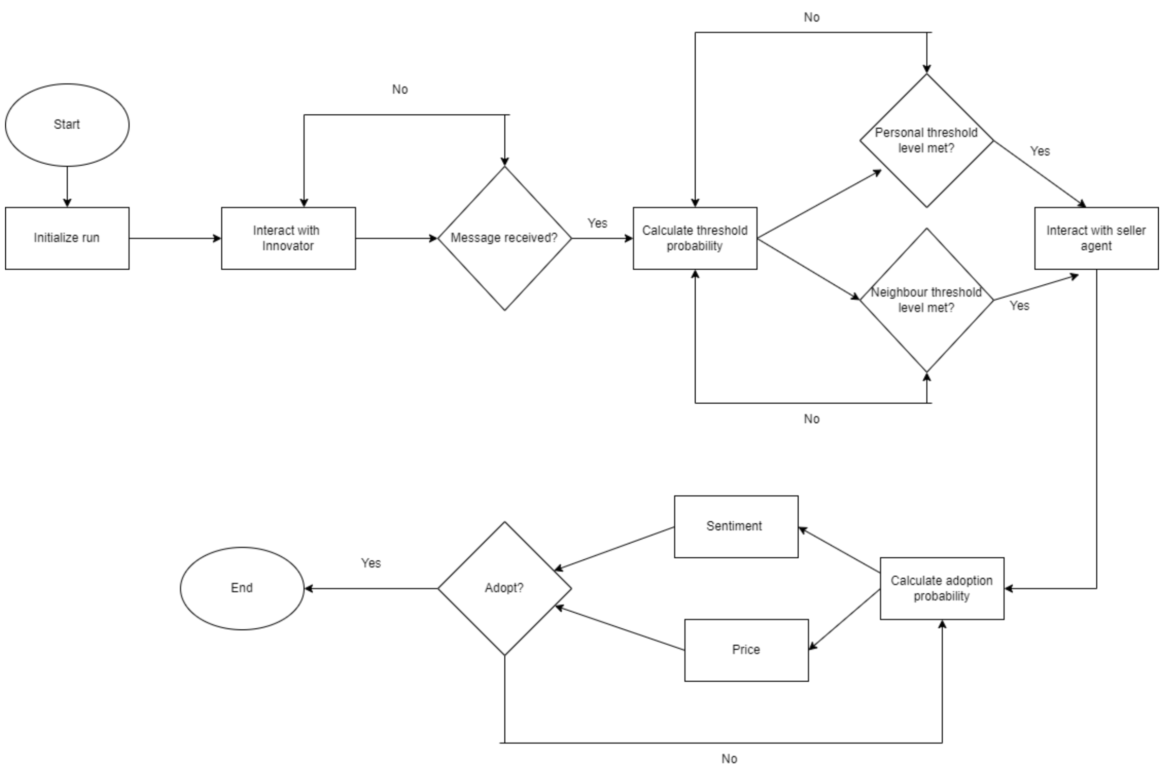


Figure **A1.** Conceptualisation of the ABM model.

Interaction with Innovator:

At initialization, innovators interact with potential consumers, spreading the message of sentiment from online reviews. Once a potential consumer has received the message they turn into a hollow red colour.

Calculate threshold probability

There are 2 types of threshold to be calculated:

- Personal threshold:
  Each agent has an individual threshold that is random at setup
- Neighbour threshold:
  Agents are influenced by their social network

If both thresholds are satisfied, the agents turn into an opaque red colour and move to the next sequential process in the model which is interaction with the seller agent.

Interaction with seller agents:

Once agents have met both threshold levels, they come across the seller agent which is represented as a blue patch in the NetLogo platform. The seller agent represents the Amazon retailer in this context. Only consumers who have received the message and have exceeded the threshold level are able to interact with sellers and go on to the next process which is decision-making.

Calculate adoption probability:

The adoption probability is based on sentiment and price which are both input parameters to the ABM model defined by the user. A sentiment score of -0.1 to -1 is represented as negative, a score of 0 is neutral and a score of 0.1 to 1 is represented as positive. Price is also determined by the user and the model will be able to explore adoption rate based on low or high prices. Once an agent turns an opaque red colour, they are actively deciding whether to adopt the product bearing in mind these parameters.

Adopt:

When adoption probability has been satisfied, the consumer will adopt the product. This is represented as the agent turning into an orange smiley, neutral or sad face depending on sentiment and price scores. The simulation will end once all consumers have adopted the product which is an assumption of the Bass (1969) model.

Design concepts

Basic Principles:

The ABM model follows the basic principles of the Bass Model (1969), where it illustrates how new products get adopted in a population and how innovators and potential adopters interact. (Charan, 2015)

There are several assumptions of the Bass (1969) model:

- There are two types of behaviour investigated which are innovation and imitation. Innovation is represented as the innovator agents in the ABM model who already know about the product and have adopted it. They are also described as a broadcast or external influence as they spread the message (Wilensky and Rand, 2015). Imitation is often described as the social influence as where others decide to adopt because of the recommendation of those who already have it. Internal influence as through word-of-mouth.

- All potential adopters eventually purchase

- Repeat purchases are not accounted for.

According to Charan (2015), the basic Bass model (1969), is represented by the following equations:

$$n(t) = adopters\ via\ innovation + adopters\ via\ imitation$$

$$adopters\ via\ innovation = p * remaining\ potential = p[M - N(t)]$$

$$adopters\ via\ imitation = q * proportion\ of\ adopters * remaining\ potential$$

(Eq.1)

Where:

n(t) = number of adopters at time t.

N(t) = n(0) + n(1) + n(2)… n (t).

M = total number of potential adopters.

p = coefficient of innovation

q = coefficient of imitation.

Adaptation:

The key adaptive behaviour of agents in the model is the decision to adopt. This is based on personal and neighbour thresholds as well as sentiment and price.

Interaction:

The following interactions take place during the simulation. Firstly, potential consumers interact with innovators. Depending on their interactions and the threshold of individual agents they continue to interact with seller agents.

Stochasticity:

In order to determine adoption probability, if personal threshold which at setup is a random number generated to all individual potential consumer agents is less than 0.01, the agent will adopt the product regardless of its social situation. However, if the agent has a social influence, the agent adopts if the random number is less than 0.5 times the fraction of neighbours who have adopted (Wilensky and Rand, 2015).

Observation:

The user interface displays the setup button to setup the parameters, step button to run the next step and the go button which continually runs the simulation. The parameters on the user interface are displayed which consist of no. consumer agents, no. seller agents, no. innovators, sentiment and price which can be changed by the user by dragging the slider to the desired number. The output of the simulation can also be observed on the user interface dashboard. The output relevant to this study is a diffusion curve which includes how many people have received the message and how many people have adopted the product. There is also a separate graph which shows number of adoptions over time and number of potential adopters over time.

Details

Initialisation:

Parameters	Value	Description
No.CA *(m)*	50, 100, 500	No. agents representing consumers
No.SA	1	No. retailers. As this study only looks at Amazon retailers and not 3$^{rd.}$ party retailers this is set to 1.
No. innovators	1, 5	Spread message to potential adopters
Sentiment	-1 to +1	Sentiment score, -1 to -0.1 is negative, 0 is neutral 0.1 to 1 is indicated as positive sentiment
Price (£)	50 - 100	Price value can be changed via the slider on the Netlogo interface
Coefficient-of-innovation *(p)*	0.03	Innovator who already knows about the product and has adopted it
Social-influence *(q)*	0.38	Coefficient of imitation where the message is spread via word-of-mouth

References:

Bass, F.M (1969). *“A New Product Growth Model for Consumer Durables”,* Management Science, Vol. 15, No. 5, pp. 215– 227.

Bass, F. M., T. V. Krishnan and D. C. Jain (1994). "Why the Bass Model Fits without Decision Variables." Marketing Science 13(3), pp. 203-223.

Charan, A. (2015). The Bass Diffusion Model. [Online]. MarketingMind. Available at: https://www.ashokcharan.com/Marketing-Analytics/~pv-bass-diffusion-model.php#gsc.tab=0 [Accessed 8 September 2020].

Wilensky, U and Rand, W (2015). *An introduction to agent-based modelling*. United States of America: Massachusetts Institute of Technology.