Faculty of Business and Law

Department of Operations, Technology, Event and Hospitality Management (OTEHM)

Measuring retail brand performance in the UK: A quantitative analysis into the impact of COVID-19 pandemic on the Home Fragrance Candle category and the business of Company A

A dissertation submitted in partial fulfilment of the requirements for the award of Master of Science in Business Analytics by the Manchester Metropolitan University Business School

Khanh Ha Tran

20040495

Word count: 13,080

Supervised by Dr Xin Shi

2021

ACKNOWLEDGEMENTS

I would like to express deep gratitude to my dissertation supervisor, Dr Xin Shi for the invaluable guidance and continuous support during the formation, writing and finalisation of the research project, and also for the understanding of my personal circumstances during the writing process.

I would like to also thank company A for the provision of data and research questions that form the basis and inspiration for this research project. I hope that the research findings and implications would provide useful insights for their home fragrance business.

I would like to thank my family and friends for unwavering emotional support and encouragement when I was feeling down, for reminding me to believe in myself and take small steps towards the ultimate goal. This project could not be completed without your support, for that I eternally grateful.

Thank you.

DECLARATIONS

The author has not, whilst being registered for the degree of Master of Science in Business Analytics, been a candidate for another degree or qualification of this or any other university or other institute of learning.

I declare that the work in this dissertation was carried out in accordance with the Regulations of the Manchester Metropolitan University. The paper is original and no portion of the work referred to in the dissertation has been copied from elsewhere and that all sources have been properly acknowledged. I consent to the submission being subjected to electronic analysis for the purpose of detecting plagiarism, collusion and other types of unfair mean.

ABSTRACT

The COVID-19 pandemic in 2020-2021 has affected multi facets of life and businesses in unprecedented proportion, facilitating temporal and permanent changes in consumer behaviours.

This study focuses on highlighting the pandemic impact on candle sales at category, brands, and SKU levels using sales data provided from Company A and the benchmark measurements for public health during COVID-19 outbreak from the UK Government. In addition to investigating the relationship between COVID-19 and candle sales, the study also seeks to detect changes in consumer behaviours toward product premiumization and ratings on e-commerce sites during the pandemic.

Conclusively, the statistical analyses confirm previous findings in the robust retail research literature that COVID-19 pandemic has positively impacted candle sales. Furthermore, signs of premiumization in the period 2020 – 2021 is relatively marginal as consumers gravitate more toward lower priced products. Analysis of star ratings on e-commerce sites also shows that the loss of smell - a common COVID-19 symptom – is an increasingly common reason for negative ratings during the pandemic.

Keywords: COVID-19, sales performance, retail sector, consumer behaviours, candle industry, home fragrance

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

1.1. Background of the study

As a consequence of the COVID-19 pandemic of 2020, the entire world encountered socio-economic hardship, with consumers from all corners of the globe experiencing limitations in their daily lives. The diminished likelihood of physical store visits combined with less engagement outside the home has resulted in several significant changes for the UK Retail sector and Home Fragrance category in particular:

- A significant drop in footfall on high streets due to the closure of non-essential retailers.
- In the Food, Drug & Mass (FDM) channel, supermarkets have reaped large returns owing to stockpiling and wider retail closings.
- A rapid surge in e-commerce growth.
- Scented candles and home fragrances have become more relevant as consumers spending more time at home. When the same area serves multiple purposes, home fragrance is increasingly being employed to distinguish between home and work.

This changing landscape causes perilous uncertainty and raises many critical concerns for the home fragrance business of Company A[1]. The Company had purchased the compiled weekly candle sales data in the UK from 30 March 2019 to 27 March 2021 and provided a briefing of its needs for data analysis and market insights. The research scope aligned strictly with the company's briefing and subsequent analyses were applied to the provided datasets. Additional data was scraped from a major e-commerce website to address the research question concerning e-commerce.

1.2. Research questions

[1] The company name and its products mentioned in this paper have been anonymised to comply with the non-disclosure agreement signed by both parties.

(1) Which brands and SKUs were leading the category as in beginning of 2021 in terms of Value Share and Rates of Sale? Which product sizes excelled? Was there any evidence of premiumization?

(2) How had sales values shifted since the beginning of the pandemic at both brand and category levels?

(3) With a specific focus on changes since the pandemic, how did e-commerce reviews of Company A's candles trend over time in terms of star ratings? Were there signs of "loss of scent" review in accordance with COVID-19 symptoms? How would this compare with other leading brands?

(4) What was/were driving these trends?

Brands of interests: 169 (SKU 335 – 918), 120, 42, 3.

1.3. Research objectives

The first research objective was to help Company A understanding the pandemic COVID-19 impacts on the sales performance of its Home Fragrance category. To this end, the following sub-objectives would be investigated:

(1) Determine the leading brands and SKUs of the category, in terms of value share and rates of sale as in the beginning of 2021;
(2) Determine the best performing product sizes of the category and each brand of interests;
(3) Finding evidence for/against the premiumization;
(4) Exploring the trend of sales values since the beginning of the pandemic, at brand and category levels;
(5) Considering how e-commerce reviews of company A's candles trended over time in terms of star ratings (focusing on changes since the pandemic) in contrast with other leading brands;
(6) Finding whether there were signs of "loss of scent" reviews;
(7) Identifying factors driving these trends.

The second research objective was to extract relevant insights from the analyses to suggest ways for candle companies to cope and thrive with the derivative impacts of the pandemic, specifically in terms of e-commerce and category management. Even though the adverse consequences of COVID-19 on the retail sector were extensively investigated and analysed since the emergence of the disease, to this end only a few studies focused on the home fragrance candle business.

1.4. Significance of the study

The research was one of the very few to study COVID-19 impact on retail sales volume using authentic weekly data. The resulting conclusion and insights can be applicable not only to businesses in the candle industry – overcrowded and fragmented by multinational FMCG (Fast Moving Consumer Goods) brands as well as small medium design labels (Deshmukh, 2019) – but also other home fragrance product categories including air fresheners and air sprays. Its investigation on consumers reaction to product sizes, premiumization, and e-commerce based on sales data illuminated conclusive insights for businesses looking to adapt with the shifting consumer trends during and after the pandemic.

1.5. Scope and expected limitations

The research scope aligned strictly with Company A's brief. It investigated sales variables extracted from the dataset dating 30 March 2019 to 27 March 2021 that were crucial to answering the research questions and other variables not included in the dataset – reviews and product ratings of different candle brands on a major e-commerce site. With the above exception, the research would not concern variables not found or extractable from the dataset. Expected limitations included 1) lack of benchmarking studies and published data on the topic, 2) incomplete data insufficient to thoroughly address the research questions – in which case supplement data would be applied at best, and 3) overgeneralisation of findings.

1.6. Research structure

This dissertation consists of 7 chapters in total and organised as follows:

Chapter 1. Introduction – Presented the background of the study, research objectives and questions, research significance, scope and expected limitations, and dissertation structure;

Chapter 2. Literature Review – Discussed the literature review of relative research and studies that set out the theoretical background for the dissertation;

Chapter 3. Methodology – Described the research philosophy, approach to theory development, process of data collection, ethical considerations and research methods used in the dissertation;

Chapter 4. Empirical findings – Presented and analysed the research results and their implications;

Chapter 5. Summary of findings – Encapsulated key findings of the research;

Chapter 6. Discussion and Conclusion – Evaluated and summarised the research relevance to the literature and the present state of the business;

Chapter 7. Recommendations and Research Limitations – Provided recommendations for the business and highlighted research limitations for further research.

2. Literature Review

2.1. The impact of COVID-19 pandemic on retail sector

COVID-19 was a contagious disease caused by a novel coronavirus (SARS-CoV-2) began in Wuhan, Hubei Province, China and had since widely spread across the world. The virus mutated rapidly, causing serious clinical complications like severe lung infections for elders, those with underlying respiratory conditions, and increasingly even younger people (Huang et *al*, 2020; Li et *al*, 2020; WHO, 2021). Due to its massive infectibility and deadly prognosis for the vulnerable population, governments had been restricting movements and activities from slight to severe to contain the virus spread (Bloomberg, 2021). Unprecedented fear and chaos in healthcare coupled with systematic social restrictions had upturned the socio-economic landscape worldwide and in the UK. Since the beginning of the outbreak, the global economy declined by 3.5%, a 7% drop from the 3.4% growth projected back in October 2019 (International Monetary Fund, 2021). The global recession caused by COVID-19 was unprecedented in scale since the Great Depression in 1930. Many countries in the Organisation for Economic Co-operation and Development (OECD) area suffered a GDP shrink of more than 20% and a rise in unemployment. Amongst this group, the UK marked the steepest drop of minus 21.7% (OECD, 2020).

The retail industry suffered with sales volumes plummeted throughout 2020. This included volume from businesses (e.g., supermarkets, door-to-door salespeople, online retailers, etc.) and individuals involved in selling products directly to end consumers (Hutton and Rhodes, 2021). The shockwaves began when Prime Minister Boris Johnson announced the first lockdown of non-essential stores on 23 March 2020 (UK Government, 2020). As a vital part of the economy, in 2020, the retail economic output was £97 billion, accounting for nearly 5.1% of UK GDP and a 2.5% drop in 2019 (ONS, 2021). Regarding the total sales volume, Figure 1 shows that the pandemic caused a 1.9% decline compared to 2019, marking the most dramatic annual fall on record (Source: ONS, 2021: online), as the imposition of lockdown restrictions, social distancing rules and persistent limitations on non-essential retail affecting the sector altogether.

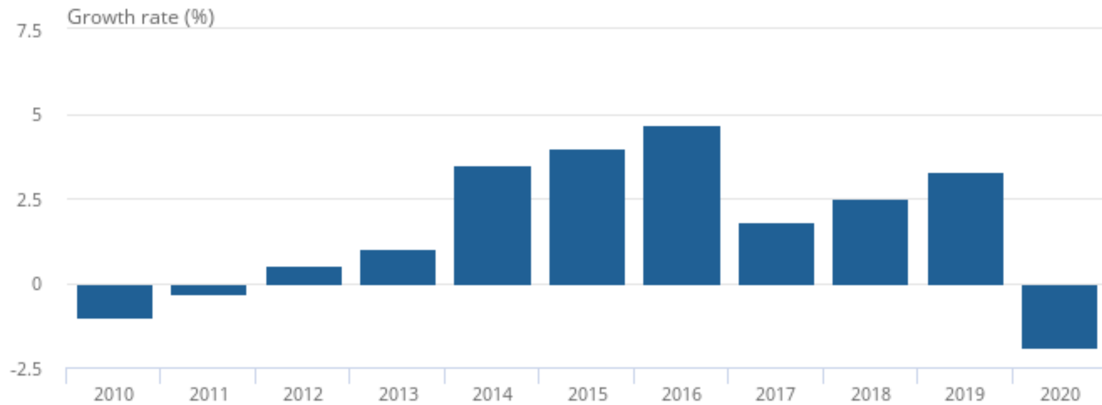

Figure 1. Total retail sales volume, seasonally adjusted, 2010 to 2020 (Source: ONS, 2021: online)

Footfall volumes revealed the impact of these restrictions, especially on the high street, as this figure began to plummet dramatically by 89.86% in March 2020 (Mumford *et al.,* 2020). The catastrophic impact was also reflected on retail parks and shopping centres index, as illustrated in Figure 2 (Source: ONS, 2021: online). By the time lockdown constraints were lifted and some non-essential stores were allowed to re-open in June, total footfall volume increased considerably. This upward trend was presumably due to pent-up demand, a common consequence during such crisis time of COVID-19 when consumers must postpone purchasing and consuming discretionary goods (Sheth, 2020).

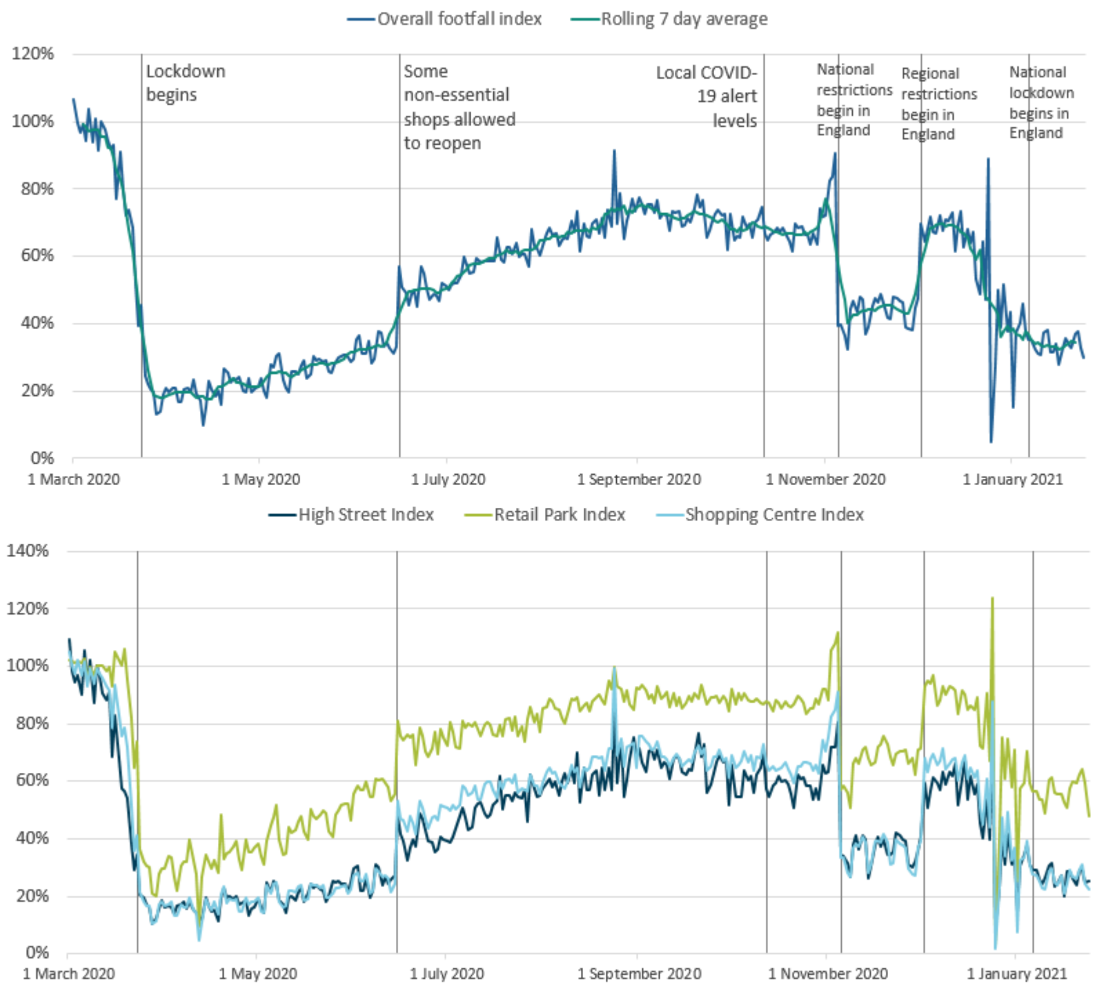

Figure 2. Volume of footfall, percentage change from the same day the previous year, UK, 1 March 2020 to 24 January 2021 (Source: ONS, 2021: online)

As access to retail was severed due to lockdown restrictions, fear of scarcity altered the pre-COVID perception of price elasticities for necessity items and powered panic stockpiling. UK traditional brick-and-mortar retailers witnessed a "queuing" crisis in the beginning of the outbreak, where customers braved hours of waiting just to enter the store – many horded food and necessities proceeding their successful entry. This prompted retailers to set purchase limits, implement online services, and priority service for the elderly. However as reactive measures were infrequently practice, they were often poorly executed, resulting in low customers' satisfaction (Pantano *et. al.,* 2020). As observed in Figure 3 (Source: ONS, 2021: online), while most sectors saw a decrease in volume sales, except for food stores that remained at a consistent sales level throughout the year, with a strong surge in March at 9.4% compared to February due to stockpiling of essentials in concerning disrupted supply chain disruptions (Khalid, 2021). Despite warnings and purchase limitation policy introduced by supermarkets on essential products, this trend of stockpiling continued, leading to an increase of 14.3% in mass groceries volume sales over three months from March (Jahshan, 2020). Figures published by Kantar (2020) revealed that most Food/drug/mass (FDM) channel retailers experienced immense growth in March, with Sainsbury's leading the top 4 well-known grocers at 7.4%, followed by Tesco, Asda and Morrisons at 5.5%, 4.9% and 4.6%, respectively.

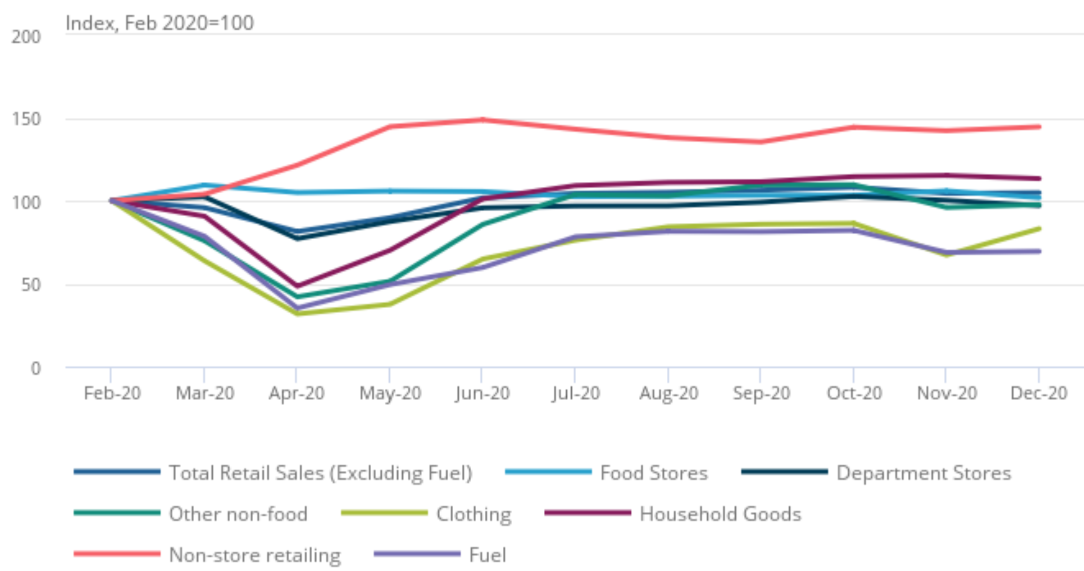

Figure 3. Value of UK retail sales at current prices, by retail sector, seasonally adjusted, Great Britain, index February 2020=100 (Source: ONS, 2021: online)

Academic research also aligned with statistical data released by the national government and leading business intelligence institutes. O'Connell, de Paula, U., & Smith (2021) studied the panic stockpiling during the first wave of the pandemic prior to the first lockdown on 23 March 2020. They found that the initial spike in volume sales of storable staple food products and non-food household items (including the notoriously famous toilet paper) was powered by real demand increase across households and across socio-economic groups, rather than outlying individual cases. Compared to pre-COVID frequency, these products were purchased more frequently in larger quantity during the two weeks of panic buying prior to the first lockdown. Consumer spending study conducted by Chronopoulos, Lukas, & Wilson on the UK population also confirmed this trend of constant increase in groceries spending throughout the pandemic in 2020, regardless of changes in official notice about movement restrictions. Other spending categories including discretionary and restaurant dining fluctuated depending on the official notice, with both categories experiencing the greatest increase during the transition from “lockdown” to “stay alert”.

Retail volume began to recover in 2021 in conjunction with massive vaccination schedule and partial opening scheme. Illustrated in Figure 4 (Source: ONS, 2021: online), full V-shape recovery was observed in May 2020 when the retail sales index rose to 62 point and surged above the 2018 benchmark of 100 point in June 2020 after a race toward the index bottom throughout the first four months of the year. Coming to 2021, sales volume buoyance by pent-up purchase impulse and positive consumer confidence as the UK emerged from lockdown with restrictions went from eased to fully lifted since 2021 Easter holidays (Cherry & Anghel, 2021; Romei, 2021).

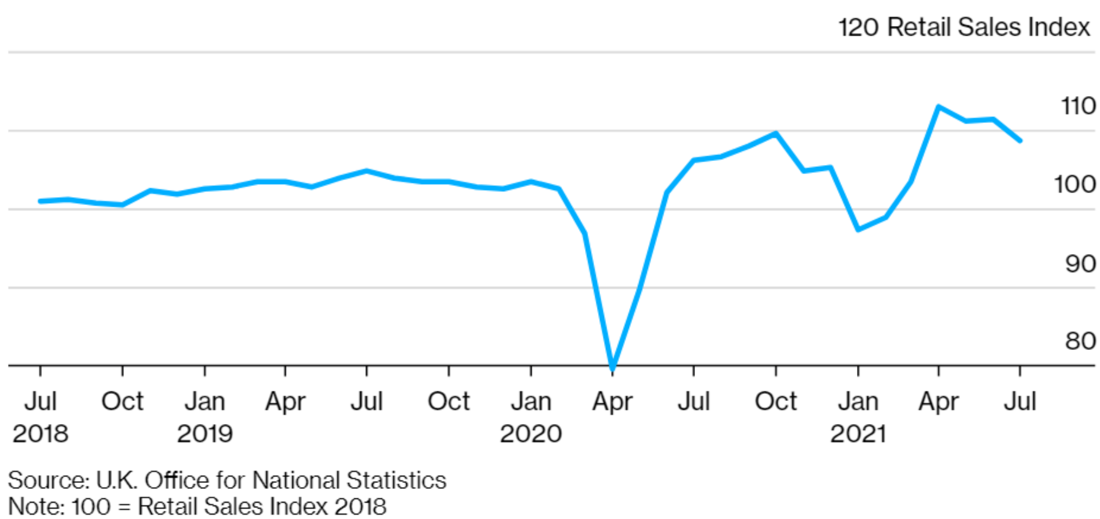

Figure 4. Value of UK retail sales at current prices, by retail sector, seasonally adjusted, Great Britain, index 2018=100 (Source: Bloomberg, 2021: online)

As seen in figure 5, Household goods stores consistent outperform other sectors in volume recovery during the period May 2020 – July 2021. The sector was the first to rise beyond the 2018 pre-COVID's benchmark of 100 point and had bypassed it since March 2021 (Bloomberg, 2021). The high-volume contribution of household goods stores illuminated the anticipated shift in consumer purchase from non-essential to essential goods (Boston Consulting Group, 2020). Non-essential products, for example clothing, mid-end and luxury products retailed through department stores also recovered after the all-time-low floor of April 2020, however they struggled to reach full recovery at the 2018 pre-COVID benchmark. Clothing stores, which contributed the second largest retail volume prior to February 2020, was hit the hardest by pandemic impacts out of all four channels.

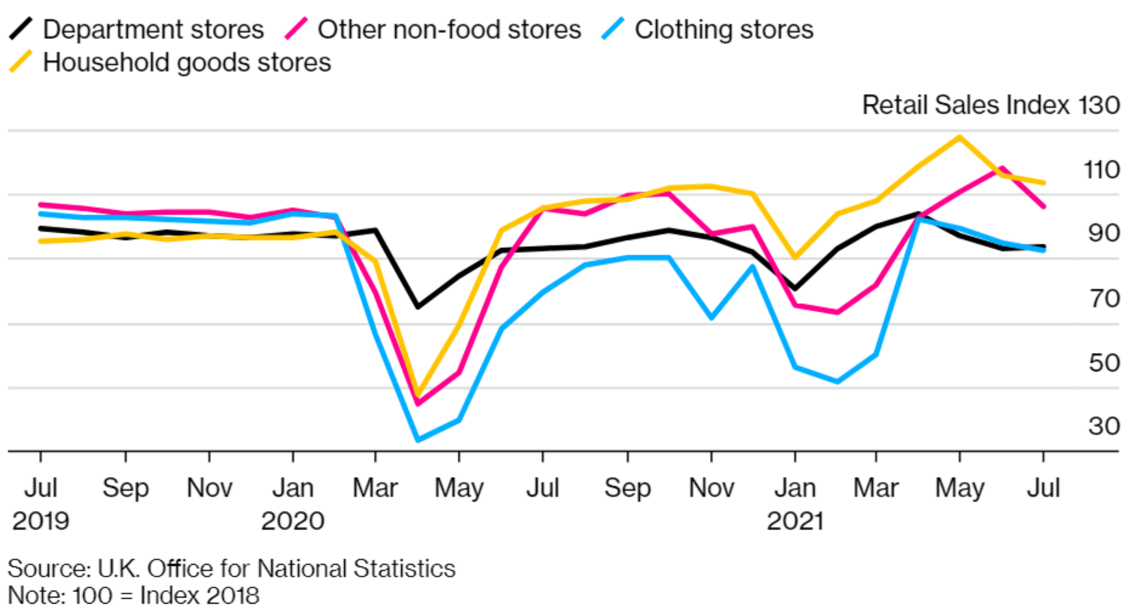

Figure 5. Value of UK retail sales at current prices, by retail sector, seasonally adjusted, Great Britain, index 2018=100 (Source: Bloomberg, 2021: online)

The impact of COVID-19 on the retail sector highlighted through sales volume in the two-year period from July 2019 to July 2021 was of critical important to the analysis of research question 2: Identify shifted trends in market shares and sales values since the beginning of the pandemic at both brand and category levels. As the candle business also fell under the home goods category, the data was expected to mirror UK retail volume trend with a steep decline in April 2020 to full V-shape recovery in June 2020, at which point volume should pick up the pre-COVID normal trend. Since the dataset included covered weekly sales of this period, candle sales graph at industry level was expected to mirror the V-shape in figure 4 and 5. Furthermore, graphs of candle sales at brand levels which deviated from the expected trend could signify the impact of experimental variables such as industry seasonal, differentiating and Covid-19 factors.

The stockpiling behaviour at the start of the outbreak, which was unanimously confirmed by governmental statistics and academic research, could also have direct impact on research question 1: Identify leading brands and SKUs, optimal product sizes and signs of premiumization, beginning 2021 by retail sales indicator (i.e., Value Share, Rates of Sale). The key word here was "product sizes" and "premiumization". Research had identified product size as an important factor in the purchasing decision of price conscious consumers (Ordabayeva & Chandon, 2013). With the dataset including product weights, it was possible to monitor the impact of sizes, particularly in candles of large sizes from minimum 600g, during the stockpiling situation under COVID-19 and normal condition pre-COVID. In terms of premiumization, positive attributes were often associated with premium pricing. Consumers were also less responsive to price changes for premium products (Huang et. al., 2017). According to Nielsen premium report in 2016, home care was among the sectors experiencing the highest growth in premiumization, with consumers reporting the two most valued premium features being exceptional quality and superior performance. In the candle business, these two qualities respectively corresponded to quality-related features such as natural ingredients, unique and/or natural scents, and performance-related features such as scent penetration and lasting duration. Company A's brand was particularly known for its product differentiation strategy and premium pricing (Wilson, 2005). As expected during the pandemic for consumers to prefer stronger performing candles, the effect of premiumization on retail volume could be observed through unit prices. The dataset did not provide specific premium variables.

2.2. Online retail growth and changes in consumer behaviour

Prior to the pandemic, online shopping was already an indispensable trend resulting from internet adoption, emerging digitalisation, and consumers' fast-paced lifestyle, especially the young generation. Tighe (2021) reported that 40% of UK consumers purchased online more in March to avoid going to brick-and-mortar stores, further explaining the acceleration of E-Commerce growth. Whilst in-store retail sales volume experienced a drastic reduction, online retail sales, in contrast, saw massive growth as consumers migrated to online shopping. In 2020, online sales accounted for 27.9% of total retail sales, increasing 8.7% compared to the previous year (Figure 6).

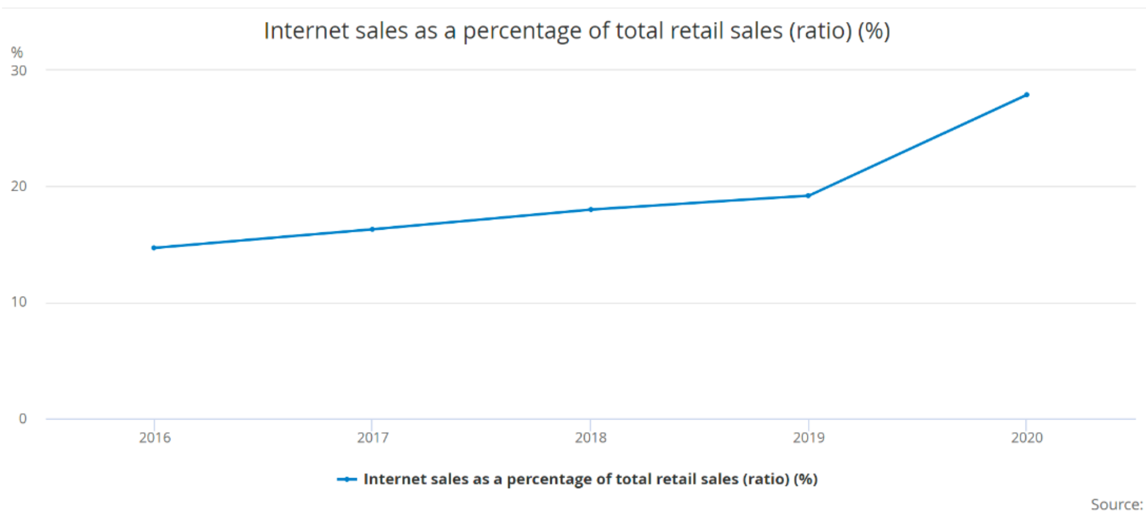

Figure 6. Internet sales as a percentage of total retail sales (ratio) (%) (Source: ONS, 2021: online)

The primary rationale for this significant change in consumer behaviour was that people work, study and relax at home, blurring the boundaries between work and personal life. As consumers adapting to this circumstance, they were more likely to inhabit alternative ways of making work, study and consumption more convenient for them, even when things are back to normality (Sheth, 2020). Furthermore, the pandemic had encouraged online shopping habits in the most unexpected consumer segment: the elderly. Compared to younger age groups, senior over 65 had relatively lower digital penetration and internet usage, both effectively lowered online shopping's attractiveness for this age group. However, as the pandemic raged on and elders found it extremely difficult to queue for brick-and-mortar store entrant, they began exploring and enjoying the convenient perks of online shopping, including versatile receiving options and cashless payment (Pantano *et. al.*, 2020). Figure 7 illustrated the permanency of this trend, as online retail index volume consistently doubled that of in-store retail, even as the latter stumbled back to partial operations.

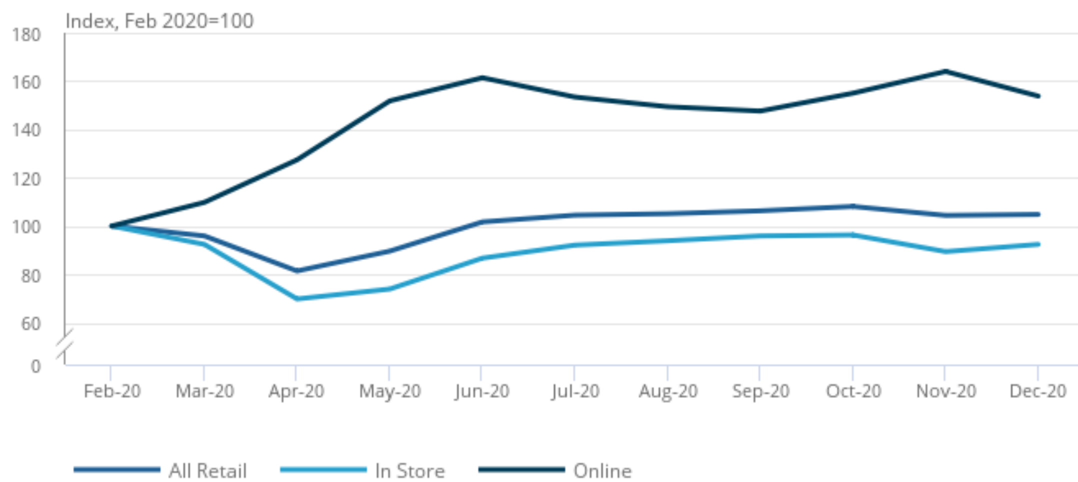

Figure 7. Value of retail sales at current prices, seasonally adjusted, Great Britain, Index Feb 2020=100 (Source: ONS, 2021: online)

The dominance of online shopping during the pandemic manifested globally. In the US, the latest Census Bureau report showed a 7.7% down in revenue of retail and food services compared to 2019's number, whereas there was a substantial rise in E-Commerce and other online services revenue, which marked an increase of 16% and 14.8% compared to the same period in 2019 respectively. In the EU, there was an explosion of 30% increase in E-Commerce activities whereas traditional retail sector revenue diminished by approximately 18% (OECD, 2020). The degree of competition between traditional brick-and-mortar retail and e-commerce platform varies across countries. For example, Pérez-Sánchez, Velasco-Fernández and Giampietro (2021) found a similar pattern in the development of e-commerce activities in the US and UK. In particular, the share of E-Commerce in retail sector in the United State rose from 9.6% in the first quarter of 2018 to 16.1% as of the end of 2020. In the same vein, shares of E-Commerce activities surge nearly 11% during 8 months of the pandemic from March to December 2020, from 20.3% to 31.5%. In China, where the share of online retail in total accumulated retail sales between January and August 2020 reached 24.6 percent, up from 19.4 percent in August 2019 and 17.3 percent in August 2018.

In 2020, the Boston Consulting Group expected the pandemic to accelerate and cement the retail migration from store to online. Initially, because consumers were confined to their residence and unable to purchase at physical stores, retailers must inevitably adjust to this new habit by shifting towards online avenues. However, as seen through historical examples in Figure 8, toward the end of every global crisis adaptation became long lasting habits. As consumers explored and became used to the alternative options, prolonged crisis established “new normal” in every facet of life. Emergence of these “new normals” - predicted by the greater changes in attitude, policy, and experiences - were most prominent in three areas of remote working, social welfare, and global dependency in the supply chain. Once policy was established and operating procedures were put into place, it would be unwise and difficult for governments and companies to go back to pre-crisis practices in these areas (Reeves *et. al*., 2020). For example, at the peak of the pandemic, several local governmental had designated online shopping and home deliveries as “essential services” (Pantano *et. al*., 2020).

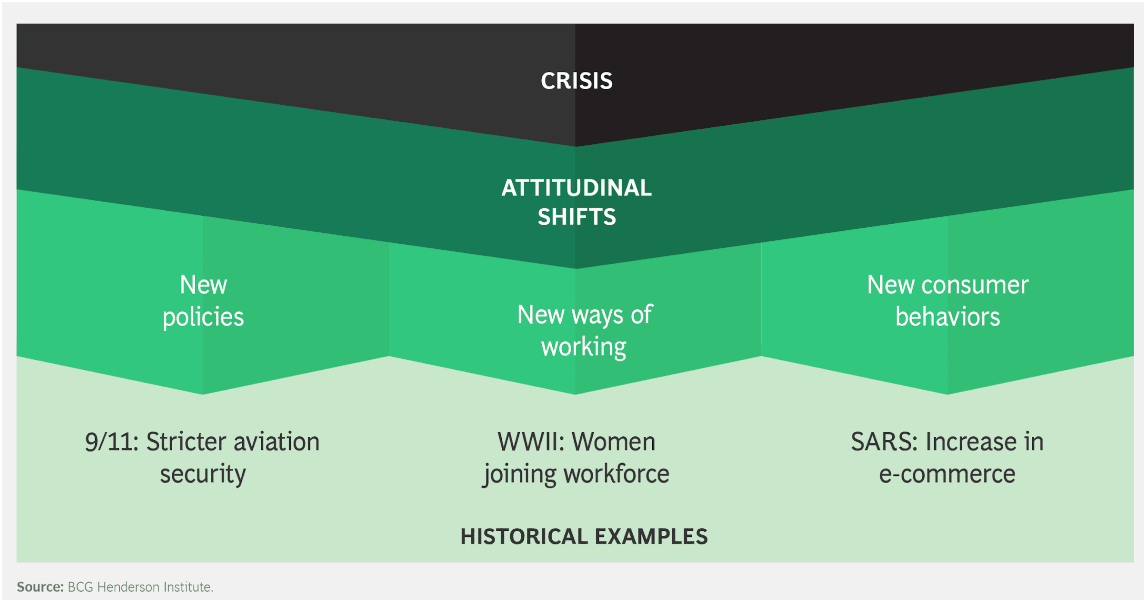

Figure 8. Historical Examples of Crisis Leading to Long Lasting Changes (Source: Boston Consulting Group, 2020: online)

Academic research conducted by Sayyida, Hartini, Gunawan, & Husin in 2021 argued the dominancy of online shopping in the post-COVID world. With secondary data compiled from the US, UK, Germany, France, and Latin America they found that two dominant retail methods during the pandemic were webrooming and pure online shopping, with purchase volume from physical retailers still outnumbered the online sales volume at 70 to 30 percent. This was first due to the disparity in online shopping adaptation and supporting infrastructures between countries. Secondly, even though consumers used the internet to check prices, they still completed the entire purchase in-store, as the sensory desire to experience the product remained a strong purchase motivator. However, Sayyida, Hartini, Gunawan, & Husin (2021) did not deny that consumers were accepting online shopping at rapid rate. They also raised the need for traditionally offline retailers to migrate online and integrate their management system for smooth operations in both environments.

Online retail dominance could be here to stay, not only because consumers preferred shopping competitive prices conveniently from home over the full package sensory experiences at brick-and-mortar store, but also because retailers had largely adapted the online operating model. Every operator in the retail value chain could realize the benefits of "online" integration, from global sourcing, order processing, warehousing, payment processing to logistics and last mile delivery. Most prominently, the online system integration powered by Internet of Things (IoT) management technology and digital transformation of the retail ecosystem had allowed concrete cost reduction benefits: cheaper, 24/07 accessible automated processes, more precise forecasting, and more adaptive inventory flow and faster turnover at a time of supply chain disruption – when holding mass inventory proved unsustainably expensive (Ivanov, 2021). In the post COVID world, more than at any point in the history of online retail, adaptation and online migration would be a "do or die" matter for retailers.

The shift toward online environment would reshape the fragrance market in general and scented candle in particular. Apart from changing business model, companies would experience an unprecedented change in marketing approach. Consumer journey with online shopping would be very different compared to the shopping path in traditional brick-and-mortar stores. Online shopping eliminated the first basic motivation for purchase and subsequent opportunities for customer to test their purchasing decision: the smell. In her cutting edge paper of sensory marketing, Krishna (2012) contended smells trigger various gamut of consumer's subconscious, which appealed to the basic senses may be a more efficient way to engage consumers. Experiment on sensory marketing of Willander and Larsson (2006) suggested that olfactory information had direct impacts on memory. Subsequent studies of Nobbs, Moore and Sheridan (2012) related the implications of sensory marketing to fragrance market, especially in perfumes and beauty care products. In a major advance in 2017, Nierobisch *et al.*(2017) found that smell is the first cornerstone of consumer journey in the fragrance market. Further, according to Huang and Cai (2015), marketers could alter brand equity via impacting olfactory system.

Perhaps most conclusive to the impact of smelling on retail experience was Spangenberg, Crowley, & Henderson findings in 1996, which indicated consumers behaved and evaluated products differently in a scented store environment as supposed to an unscented store environment. Without the overwhelming support from the olfactory sense, scented candle retailers must re-evaluate their marketing and service strategy for online consumers. As the consumer purchase preference shifted toward online environment, it rendered the effectiveness of sensory marketing in the fragrance candles market. Vision superseded olfactory to receive the first impression of candle products. If the product's name represented sufficient sensory cues, the recognition over recall heuristic could aid consumers to recognize the smell (Gigerenzer & Goldstein, 2011). However, the lack of research in this area presented substantial challenges for marketers to adopt new marketing strategies catered to the online market.

The impact of online shopping and changing consumer behaviour discussed in this section had direct implication to the analyses of research question 3: Illuminate the changes if any in E-Commerce reviews of Company A candles trended over time, focusing on pandemic factors and compare with other brands' review performance, and 4: Identify the driving forces of these trends. For question 3, this research would partially address the impact of the olfactory loss from two directions. The first was organic olfactory loss as an integral part of the online shopping experience. Sales data through online retail activities prominent during COVID-19 lockdown would be analysed and compared with pre-COVID sales data, of which brick-and-mortar retail contributed the majority. The second was inorganic olfactory loss as a symptom of COVID-19. Scraped data from a major E-Commerce site was analysed to evaluate consumer's ratings of product at the peak of the pandemic in the UK. Hypothetically the increased number of consumers losing their sense of smell would reflect in an increased number of unfavourable product rating.

2.3. Introduction to the candle industry: product usage and market size

By product specification, home fragrance had received much attention from scholars in the past decade, many focused on the negative effects of fragrance production on environment and on personal wellbeing through emission absorbed into the respiratory system. Recent research by Steinemann (2016) identified fragrance consumer products, despite their claims of natural ingredients, were prominent triggers of adverse wellbeing consequences and further suggested policy intervention for fragrance-free public spaces to effectively reduce health risks and strengthen air quality. Even though scent candle was not singled out as a trigger in Steinnemann 2016 study, as its usage grew, so did public concerns about potential health impacts from exposure to candle emissions and to fragrance. Specifically, the generation of candle soot had been identified as a likely source of an adverse health phenomena known as black soot deposition. Scented candles, jar candles, and oil candles generated more soot than conventional wax candles. Other harmful pollutants from burning candles include carbon monoxide (CO), carbon dioxide (CO2), nitrogen oxides (NOx), aldehydes, and unburned/partially burned hydrocarbons, particularly polycyclic aromatic hydrocarbons (PAH), which are well known for their cancer-causing properties (Lau *et al*., 1997; United States Environmental Protection Agency (USEPA), 2001; Lee and Wang, 2006; Orecchio, 2011; Derudi *et al.,* 2013). Excessive and regular burning of candles in enclosed indoor environments could also expose people to harmful amounts of organic chemicals (USEPA, 2001). These precaution about candle usage could have considerable impact on candle sales during the COVID-19 pandemic, as respiratory vulnerabilities came into the disease prevention spotlight. However, surprisingly the adverse impact of home fragrance and candles were not well-known in the general public, as marketers and manufacturers were not required to disclose all ingredients on product label. Marketing efforts for the candle industry, on the contrary, focused on its health benefits derived from aromatherapy. Based on the industry's rising sales, these efforts had paid off handsomely.

In recent years, the home fragrance market in general and the candles market in particular experienced tremendous growth due to rising consumer demands. The global home fragrance market value was worth $5,628.9 million in 2018 and forecasted to reach $9,122.6 million by 2026, at a CAGR of 6.3% for period 2019 – 2026 (Deshmukh, 2019). In terms of distribution channels, supermarkets and hypermarkets accounted for more than 51.5% in 2018 (Grand View Research, 2019), while the rest was shared between convenience stores and online channel. Online shopping had picked up momentum due to the technological advances (e.g., scanning QR/bar code) and store staff to assist consumers with purchase. The market was segmented into three main categories, including candles, reed diffusers, and room sprays. Regarding regional markets, Europe was the top consumption destination in 2018 with a revenue share of 32.3%. It represented two-thirds of the global candle market for recent years. The UK was the second-largest importer, accounted for 16% of the European market for candles (CBI, 2020). A report published by Vend, summarised by Dunsby (2019), concluded that each British individual bought an average of 6 candles per year at £7.40 each, resulting in a total spend of £1.6 billion annually for the whole population. The product was also highly seasonal with purchasing demand increases during festive seasons (e.g., Mother's Day, Christmas) for both consumption and gifting purposes.

Nowadays across the UK, candles were used not only in private homes but also public spaces to conceal unpleasant odours and aid stress reduction. The latter played a strong role in making candles a regular consumption product which many believed to have positive health benefits. Previous studies showed that *aromatherapy* – a holistic, therapeutical treatment using scented products – could heal well-being imbalances, reducing stress and anxiety (La Torre, 2003, Keville and Green, 2009). A study conducted by Warrenburg in 2005 confirmed the positive effect of relaxing fragrances in reducing stress-induced muscle tension and that suitable fragrances could generate a "physio shield" to defend the body from experiencing bodily impacts of stress. Since smell was undeniably a powerful sense that could easily trigger memories and feelings, its applications in spa, self-care, psychological therapies, and even retail environments had been widely accepted and practiced.

During the pandemic, as people stay at home, many suffered from mental distress and struggled to maintain a healthy mental state (Shevlin *et al.,* 2020). Due to its therapeutic impact previously mentioned, scented candles attracted both existing and new consumers for stress reduction and wellness purposes. Another factor boosting candles demand was that the product itself was highly accessible for consumers of different market segments, from low-end to high-end. Regarding demographics, candle consumers came from different genders, age groups and occupations. As the product had a seasonal attribute, its popularity remained peaked during holiday seasons (Grand View Research, 2019). This trend was to be questioned and tested in the prolonged pandemic of 2019 - 2021. Preeminent to this research, candle usage as a stress-reduction product would provide crucial insights for the analyses of research question 4: Identify the driving forces of these trends. If the dataset showed increase retail volume during the pandemic compared to normal period, the therapeutic characteristic of candle and its usage as a wellness product could explain the data peaks. Similarly, should the data reflect a decrease in consumption, the health concerns from candle usage, which had remained largely dormant in public perception, and the assumption that candle was only a seasonal gifting product could both be attributable to the pandemic volume drop.

2.4. Significant industry media coverage during COVID-19

Media coverage of the candle industry had reported a sharp rise in volume sales during the pandemic. Specifically, shortly after the first lockdown implemented in May 2020, articles surfaced reporting the industry sales figures collected and consolidated by NPD Group – one of the world's largest consumer research companies headquartered in the US (NPD, 2021). The report showed increased sales in all three major product categories of home fragrance: compared to March 2019, room fresheners sales soared 37% in March 2020, followed by increased candles sales of 6%, and 3% in reed diffusers sales. The best-selling items represented natural flowers blossomed in British countryside during spring while candles infused with essential oils known for mood improvement were also popular (Wells, 2020). Using data collected by Kantar research, media reports also indicated an increased month-on-month sales of 29% from October to November 2020 for scented candles and essential oils, following changes in alert levels and the cold weather kicking in (Woods, 2020). Based on these media reports, it was difficult to deduce the driving factors behind the increased sales of scented candles, as it could be consumer demands for stress reduction methods and/or the cyclic seasonal trends exacerbated by pent-up demands. These media report had direct implications for the analysis of research question

2: Identify shifted trends in sales values since the beginning of the pandemic at both brand and category levels, as the retail volume trend seemed to shift upward.

The second major media coverage of the candle industry during COVID-19 concerned popular Twitter of US based science illustrator Terri Nelson and data scientist Kate Petrova. Nelson first posted on Twitter on November 2020 about the surge in negative review for a popular candle brand during the pandemic. Petrova became intrigued with the post and subsequently analysed the star rating of top 5 best-selling candles on a popular E-Commerce platform. Her illustration of the result by figure 9 below had led public opinion to deduce a significant relationship between the loss of olfactory sense by COVID-19 and consumers' dissatisfaction in the quality of scented candles (Millard & Laube, 2020; Petrova, 2020). It was difficult to predict if scraped data for research question 3 would confirm Petrova's study, since this dataset had significant sample size constraint. Furthermore, the assumption that COVID-positive consumers tended to leave more negative candle reviews required extensive variable testing not accommodable in the scope of this research. However, it would be interesting to observe the changes in rating trend from 2020 to 2021 if any, as vaccination significantly reduced infection rates and subsequent symptoms derivatives.

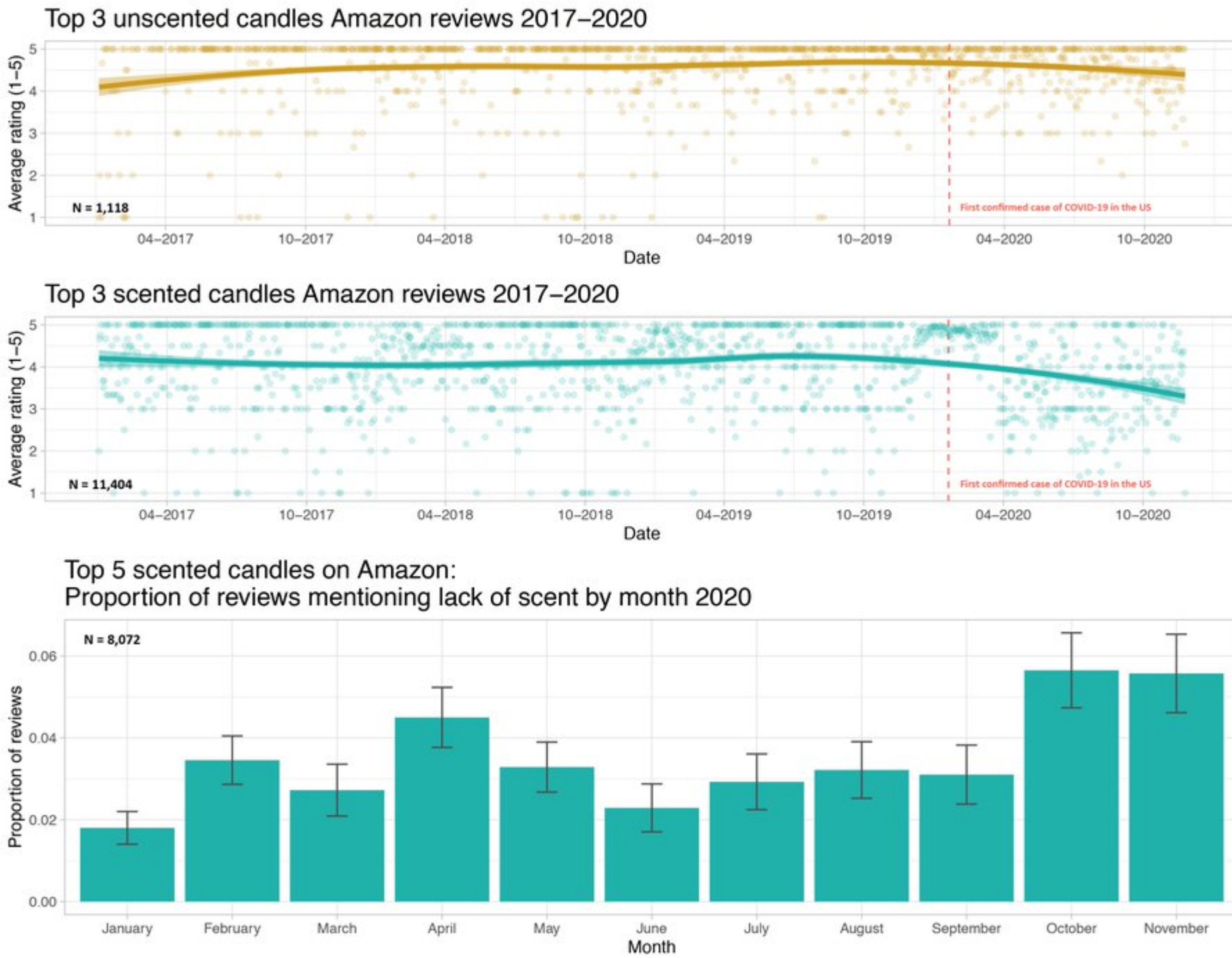

Figure 9. Star ratings and reviews of top candles on a major E-Commerce platform 2017-2020

(Source: Kate Petrova on Twitter, 2020: online)

3. Methodology

3.1. Research philosophy

This empirical research utilized annual sales data provided by company A combined with scraped star ratings and product reviews from a popular e-commerce website to consider the potential impact of COVID 19 on sales and product ratings. The research philosophy was positivist. In terms of ontological assumptions, the research worked with real and independent data that were categorized in order (Saunders *et. al.,* 2016). Based on epistemological assumptions, it obeyed the scientific method in analysing qualitative data to draw causal explanation and prediction. Concerning axiological values, the author was detached and preserve an objective stance of the research topic. Lastly, based on methodology, the study is deductive, structured, with large sample size drawn from real business activities (Saunders *et. al.,* 2016).

3.2. Approach to theory development

The study's approach to theory development was deductive, with rigorous variable testing conducted by descriptive and inferential statistics to affirm or reject a variety of hypotheses concerning data trends (Saunders *et. al.,* 2016). Through hypothesis testing, the author was able to draw inferences and replicate the model for dataset with similar characteristics.

3.3. Methodological choice

Statistical processes were employed to analyse the three sets of quantitative data and draw inferences about the variables. The methodology was well-structured, utilizing descriptive and inferential statistics, as well as graphical illustration, and retail sales indicators as described in the subsequent sections (Saunders *et. al.,* 2016).

3.4. Ethical considerations

Research ethics were strictly observed according to the Manchester Metropolitan University's guidelines, with approved research proposal by university's faculty member and proper consent form signed by Company A and the author. All data used in the research was provided by company A, disclosed by the UK Government, and obtained directly by the author from a popular E-Commerce source that is accessible to the public.

3.5. Dataset and variables

Data provided by Company A was the primary source of data for the study. The dataset A contained weekly tracking of aggregated supermarket data in the UK from 30 March 2019 to 27 March 2021 (105 weeks). Directly extractable variables from the dataset included brand, SKU, weight, weekly units sold, weekly sales value, all commodity volume (ACV). Indirectly extracted variables included monthly sales value calculated by adding weekly sales values, average unit price calculated by sales value over weekly units sold, premiumization determined by assessing the relationship between sales value/units sold and average unit price, size determined by range of weight, total distribution points (TDPs) calculated by adding ACV.

The time variable in dataset A, represented as week in year, corresponded to the time variable in dataset B – which contained secondary data of weekly cumulative COVID-19 cases obtained from the published Coronavirus information website by the UK Government.[2] Weekly cases were calculated to match the time frame in dataset A.

Lastly, dataset C used in the study contained secondary data of customer's star ratings and reviews for 5 popular candle brands dating from 2013 to 2021, scraped from the largest operating e-commerce website in the UK. The main purpose of dataset C was to evaluate research question 3. It pertained two new distinctive variables: star ratings and mentioning loss of smell.

3.6. Choice of software and tools

3.6.1. STATA

[2] COVID-19 daily cases obtained from https://coronavirus.data.gov.uk/details/cases, cumulative weekly cases calculated by author

STATA has been used extensively in several scientific disciplines, including social sciences, as a powerful graphical and statistical tool capable of handling univariate and multivariate data through a combination of comprehensive built-in features as well as user-developed functions (Mehmetoglu and Jakobsen, 2017). The author employed STATA to conduct the majority of significance analyses in this research, most notably the sign rank test and Pearson's correlation and regression analyses.

3.6.2. R programming language

R is a highly extensible, open-source programming language and environment developed by John Chambers at Bell Laboratories that is widely used for statistical computing and graphics (The R Foundation, n.d.). This study applied R to mirror Kate Petrova's experiment to answer research question 3. A set of codes can be found in the appendix.

3.6.3. Tableau

Tableau is interactive business intelligence and visual analytics platform which is well-used by industrial practitioners to analyse and report vast volumes of data for better decision making (Tableau Software, n.d.). Tableau was used in this dissertation to conduct graphical representations of the research results.

3.7. Descriptive statistics

In order to address the four research questions, analyses were organized into three main categories: retail indicator analyses, descriptive statistics, and inferential statistics. Descriptive statistical testing was primarily employed to provide comprehensive and succinct overviews of datasets (Laerd Statistics, n.d.). It was carried out using mean, median, percentage, and sum calculations for the key variables addressing research question 1, 2, and 3, including sales value, weights, and star ratings. Results of descriptive statistics were recorded in summary statistics tables, transferred into graphical illustrations and rankings to determine leading brands and SKUs, optimal product sizes and identifying shifted trends in sales values since the beginning of the pandemic at both brands of interest and category levels.

3.8. Inferential statistics

Inferential statistical testing was employed for hypothesis testing to study the primary relationships and interactions between the key variables in research questions 1, 2, 3, and provide explanations for question 4 by identifying the variables that posit significant different on data trends (Laerd Statistics, n.d.).

3.8.1. Shapiro-Wilk test for normality distributions

Considering that it is crucial to determine whether data are deviating from the normal distribution (Elliott and Woodward, 2007) which is assumed in various statistical procedures, the Shapiro-Wilk test is widely used in scientific research to evaluate normality. The validity of the test is determined by the correlation between the data and the corresponding normal scores (Peat and Barton, 2008; Ghasemi and Zahediasl, 2012). The author employed Shapiro-Wilk test in conjunction with visual methods (density histograms, quantile – quantile plots) and descriptive statistics results to inspect distribution of all variables (sales value during both normal and pandemic periods)

3.8.2. Two-sample paired sign test

The two-sample paired sign test is a non-parametric significance test used to compare the medians of two related continuous variables introduced by Dixon and Mood (1946). It was used to ascertain if there was a statistically significance difference in sales value between the non-COVID 19 period and the COVID -19 period for the category as a whole and for each brand of interests. Prior to conducting the test, the following assumptions must be considered:

(1) The dependent variable should be measured on either continuous or ordinal level;

(2) There should be two categorical "matched pairs" for the independent variable, meaning that the same subjects are present in both groups and have been measured at two separate time points regarding the same dependent variable;

(3) The paired observations of each participant must be independent, i.e., each participant's value cannot affect another participant's value;

(4) The difference scores must be calculated from a continuous distribution (Laerd Statistics, n.d)

3.8.3. Spearman's rank-order correlation and local polynomial regression

3.8.3.1. Spearman's rank-order correlation

Spearman's rank-order correlation was performed to determine the relationship between cumulative COVID-19 cases and sales value and between average unit price and units sold/sales value. The main assumptions required for this test were satisfied in that all variables (cumulative COVID-19 cases, average unit price, weekly units sold and sales value) were continuous, the paired variables had a monotonic relationship, which meant that either the variables increased in value together, or as one variable value increased, the other variable value degreased (Laerd Statistics, n.d.). Specifically, as a rank-based correlation measure, this method is non-parametric and does not rely upon normality assumption, making it suitable for testing the correlation coefficient of non-normal bivariate populations and rendering more robust results as opposed to parametric measures such as Pearson's correlation (Schober *et al*, 2018). A phenomenon resulted from market dominance by leading brands which would be discussed in subsequent sections.

The hypotheses for Spearman's correlation in the analysis part were formed as following:

H_0: There is no monotonic association between cumulative weekly COVID-19 cases and sales value of category and each brand of interests

H_A: There is a monotonic association between cumulative weekly COVID-19 cases and sales value of category and each brand of interests

and

H_0: There is no monotonic association between units sold/ sales value and average unit price of category and each brand of interests

H_A: There is a monotonic association between units sold/ sales value and average unit price of category and each brand of interests

Spearman's correlation coefficient r_s ranges in value from -1 to +1 and as the absolute value of r_s increases, the monotonic relationship becomes stronger (Schobe *et al,* 2018). For the purpose of this report, the ranges of assessed values in the following table were used as a rule of thumb in order to gauge the strength of r_s.

Absolute magnitude of the observed correlation coefficient	**Interpretation**
0.90 – 1.00	Very strong correlation
0.70 – 0.89	Strong correlation
0.40 – 0.69	Moderate correlation
0.10 – 0.39	Weak correlation
0.00 – 0.10	Negligible correlation

Figure 10. Rule of Thumb of a conventional approach to interpreting a correlation coefficient (Source: adapted from Schober et al., 2018)

3.8.3.2. Local polynomial regression (LOWESS)

Local polynomial regression, also known as LOWESS (locally weighted scatter plot smoother), is a method proposed by Cleverland (1979) for fitting a smooth curve between the dependent variable and one or up to four predictor variables. It is a non-parametric regression method that relaxes the linearity assumptions of conventional regression, thus allowing the regression coefficient to vary with the value of the explanatory variable (Ledolter, 2013). In the latter analysis of the paper, local polynomial regression was conducted in conjunction with Spearman's correlation to identify the relationships of focused variables (units sold, sales value, cumulative COVID-19 cases, average unit price).

3.9. Retail sales indicators

Retail sales indicators used in the analyses were explained in detail below. Conclusively, these indicators were used to evaluated important factors such as distribution and rate of sales for any given product. These factors consequently reflected the supply and demand vectors that were highly anticipated to be under severe pressure during the pandemic.

3.9.1. Distribution

In FMCG market research, All Commodity Volume (ACV) of a store is the total cash volume of the accumulated merchandise scanning through the check-out counter in a given period. The ACV for a specific channel is the summation of all stores' ACVs within that channel. %ACV weighted distribution is the measure of a product's availability, which signifies the spread of that product across all stores in a particular channel or the entire market. %ACV weighted distribution can be measured at the lowest SKU level, as given in the dataset. This measure is a *non-addictive* fact (Martin, 2012).

Total distribution points (TDPs) is another distribution measure that considers the *depth* factor, meaning TDPs of a brand is the summation of %ACV distribution figures of every SKUs under that brand (Simon, 2013). As %ACV distribution for each brand is not supplied, TDPs measure was calculated to determine sales velocity at brand level.

3.9.2. Velocity

Sales velocity, or Rate of Sales (ROS), is an essential retail performance metric to indicate the speed by which a product is sold when it is available for consumers on store shelves (Martin, 2014).

Sales per point of distribution (SPPD) is one of the most intuitive methods to measure velocity within a single market, in the given dataset, supermarket channel. It incorporates %ACV weighted distribution measure to quantify how quickly an SKU may be sold *from each distribution point*. The mathematical formula is:

$$Sales\ per\ point\ of\ distribution\ (SPPD) = \frac{Sales}{\%\ ACV\ Distribution}$$

Sales per total distribution points (SPTDP) for each brand can be measure based on the total sales value and TDPs:

$$Sales\ per\ total\ distribution\ points\ (SPTDPs) = \frac{Total\ Sales}{TDPs}$$

Sales velocity is also a *non-addictive fact* (Martin, 2012). For the rest of the dissertation, SPPD and SPTDP will be used as the method to calculate ROS.

3.10. Year over year calculation

Year-over-year (YoY) calculation is a frequently used financial analysis to determine performance growth for a business. It compares statistics for the same period (e.g., week, month, quarter) of two consecutive years and calculates the percentage change over 12 months between those two periods (Novotna, 2021). The formula to calculate is simplified as follows:

$$\%YoY\ change = \left[\frac{This\ year - Last\ year}{Last\ year}\right] \times 10$$

3.11. Data visualisation

Data visualisation is the process of displaying data in a meaningful and effective way to generate insights that would aid in making better decisions (Evans, 2016). In this research, the author employed STATA graphical tools, Tableau and R to illustrate statistical findings and identify potential trends.

4. Empirical findings

4.1. Impact of COVID-19 on home fragrance category

4.1.1. Category level

Prior to assessing the impact of COVID-19 on sales, it was necessary to investigate the aggregated data wherein potential pandemic impact was assumed to be erased, as data spread included the normal period of 2019 – 2020.

As seen in figure 11 below, aggregated data for total category sales during the whole period March 2019 – March 2021 indicated the seasonality attribute of candles. There was a high peak in March and an increasing slope towards the winter holiday season from October to December, potentially due to lower temperature and increasing gift buying demand. The data trend further confirmed the previously mentioned industry report and literature review on the seasonal cyclic behaviour of candle sales as a major gifting product that peaked before major shopping seasons such as before Easter (March or April) and Christmas (December).

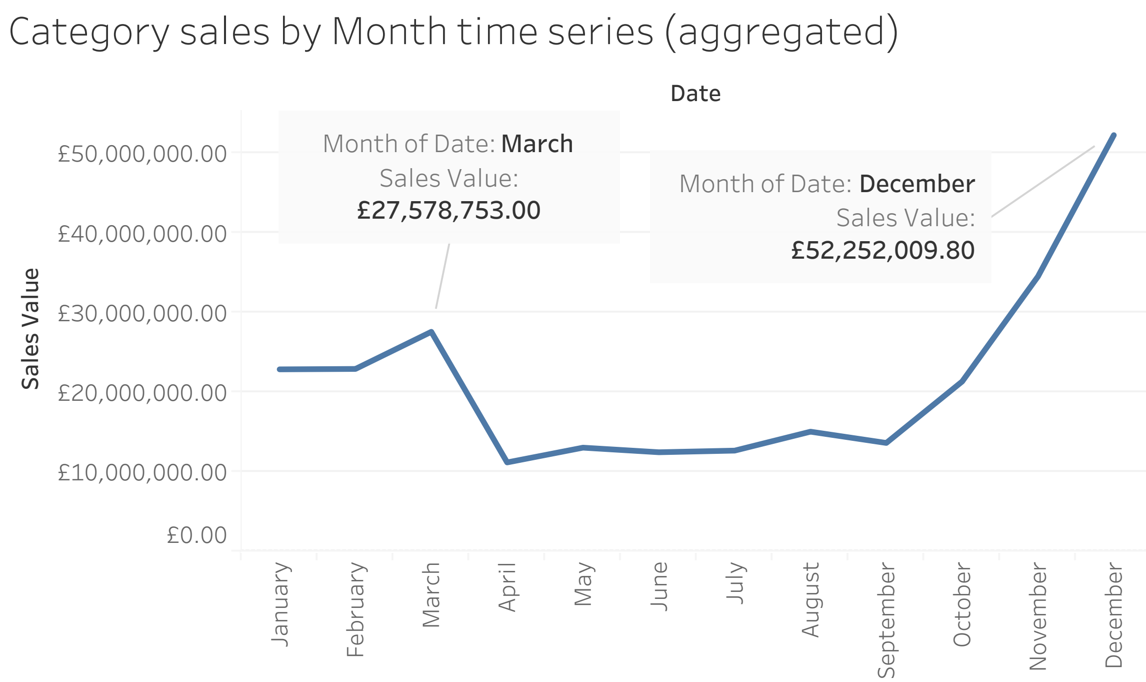

Figure 11. Total category sales by Month (March 2019 - March 2021) (Source: original, conducted by author)

When COVID-19 was factored, category sales before the pandemic also indicated similar seasonal peaks as seen in Figure 12 below. The chosen non-COVID 19 period is 06 April 2019 – 28 March 2020 (52 weeks) and COVID-19 period is 04 April 2020 – 27 March 2021 (52 weeks). Overall, the period after the pandemic has seen an increase of £28,941,100.30 (approximately 26% year-over-year change) in sales value when comparing to the period before the pandemic. It was observable that the sales gaps between the two major seasonal peaks Christmas and Easter was aggressively intensified during the pandemic compared to that during the normal period.

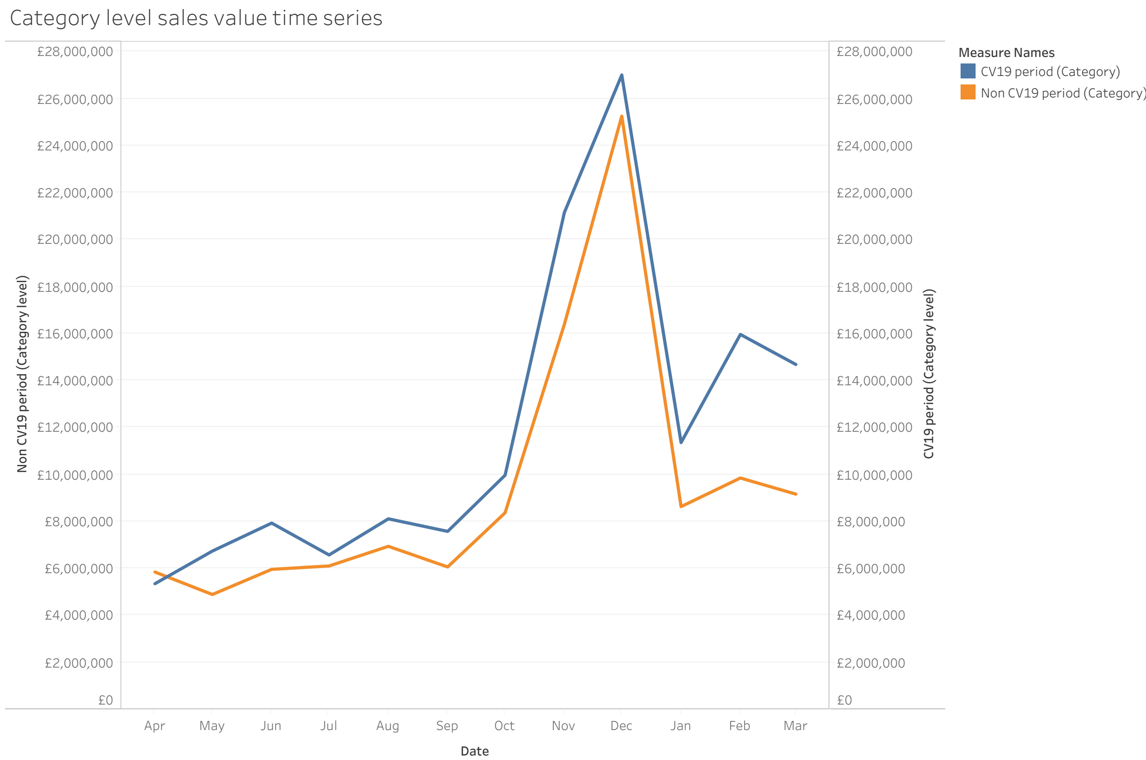

Figure 12. Category level sales value time series comparison (Source: original, conducted by author)

stats	Non_CV19	CV19_era
min	1086841	1231018
max	7798566	7247148
mean	2181408	2737967
p50	1813332	2196169
p75	2295720	3302031
skewness	2.287783	1.478957
kurtosis	8.0856	4.293541
sd	1412094	1569905

Figure 13. Descriptive statistics summary table – Category level sales (Source: original, conducted by author)

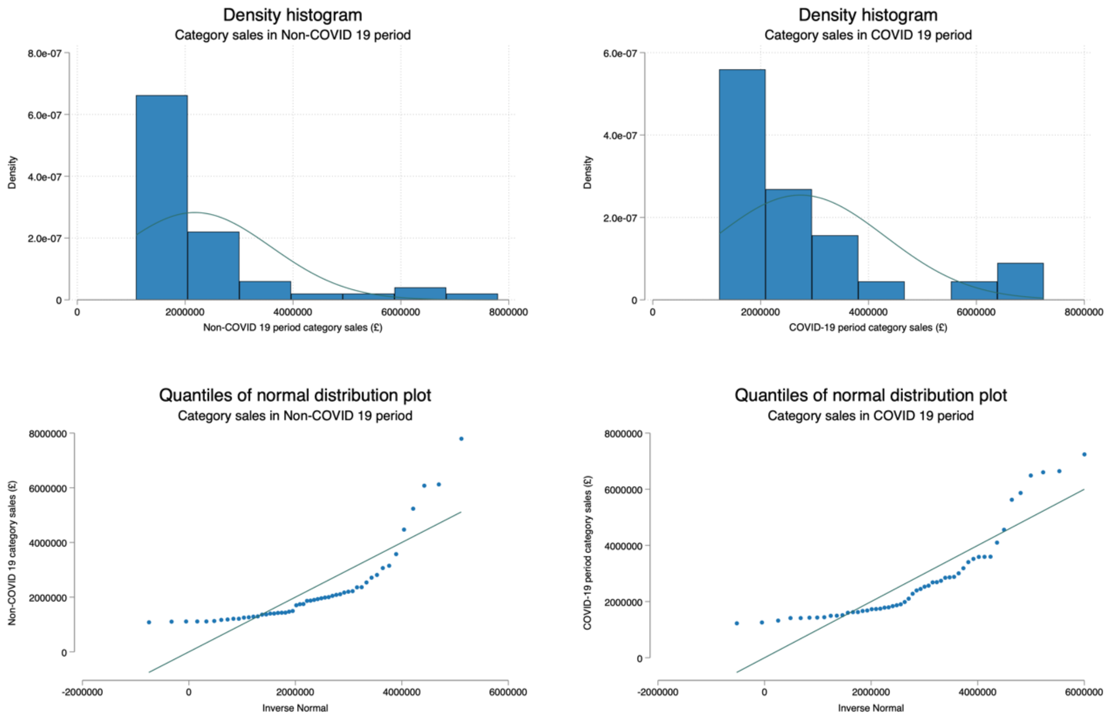

Figure 14. Distributional graphs - Category sales (Source: original, conducted by author)

Inspecting the descriptive statistics table (Figure 13) and the distributional graphs, it can be quite certain that both periods did not posit a normal distribution. Between the mean and median, there was a large discrepancy, which indicates that the distributions of both periods were quite different from a perfectly symmetrical distribution with a skewness and kurtosis value of 0. Skewness coefficient for Non_CV19 period's sales value was 2.7783 and for COVID-19 period's sales value was 1.478959. Kurtosis coefficient for both periods were 8.0856 and 4.293541, respectively. Data was right-skewed. Shapiro-Wilk normality test further confirmed the illustrations as the results showed that the distribution significantly departed from normality for both periods Non_CV19 ($W = 0.70293$, p-value < 0.01) and CV19_era ($W = 0.79973$, p-value < 0.01).

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Non_CV19	52	0.70293	14.410	5.703	0.00000
CV19_era	52	0.79973	9.715	4.860	0.00000

Figure 15. Shapiro-Wilk test for normality - Category sales (Source: Original, conducted by author)

In light of the above, a two-sample paired sign test was employed to ascertain whether the difference between aggregated sales value of non-COVID 19 period and COVID 19 period was statistically significant for the category as a whole. The following hypothesis was formulated:

H_0: The median of the difference between two samples is equal to 0

H_A: The median of the difference between two samples is not equal to 0

The resulting p-value was lower than the chosen significance level of 0.05. Therefore, the null hypothesis can be rejected and there was a statistically difference of the median of two samples.

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           8          26
    negative |          44          26
        zero |           0           0
-------------+------------------------
         all |          52          52

One-sided tests:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era > 0
      Pr(#positive >= 8) =
         Binomial(n = 52, x >= 8, p = 0.5) =  1.0000

  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era < 0
      Pr(#negative >= 44) =
         Binomial(n = 52, x >= 44, p = 0.5) =  0.0000

Two-sided test:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era != 0
      Pr(#positive >= 44 or #negative >= 44) =
         min(1, 2*Binomial(n = 52, x >= 44, p = 0.5)) =  0.0000
```

Figure 16. Sign Test results - Category sales (Source: original, conducted by author)

To further inspect the impact of COVID-19 pandemic on category sales, a Spearman's correlation was run on full date (105 weeks) to justify the correlation between cumulative COVID-19 cases and category sales. There was a statistically significant, moderate positive relationship between the two variables, $r_s = 0.4472$, p-value < 0.01. Specifically, the scatter plot with a fitted local polynomial regression (LOWESS) line showed that sales value increased with an increase in cases up to about 200,000, after which it declined.

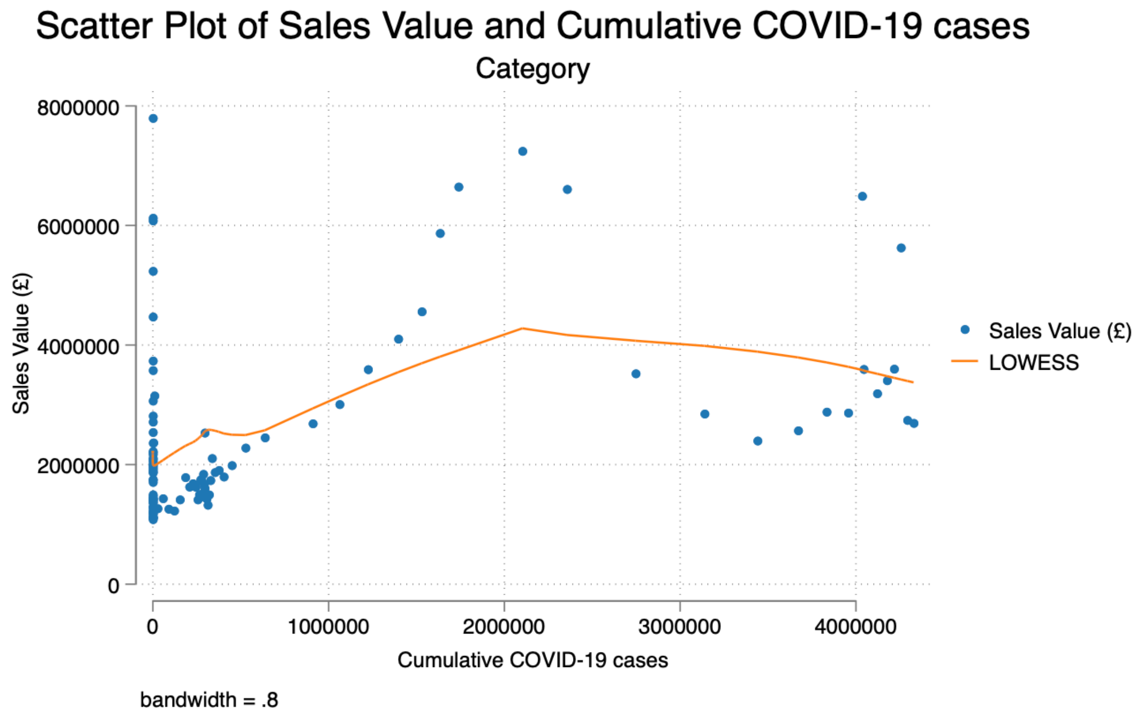

Figure 17. Category overall - Scatter Plot of Sales Value and Cumulative weekly COVID-19 cases (Source: original, conducted by author)

4.1.2. Brand 169 (SKU 335 – 918)

When comparing to the non-COVID 19 period, the COVID 19 period recorded an increase of £7,839,736 in sales, approximately 25% year-over-year. As illustrated in the following histograms and Q-Q plots, both periods had a right-skewed distribution. Skewness value was 2.293997 for non-COVID-19 period and for COVID-19 period was 1.789476. Kurtosis values for non-COVID-19 period was 8.179875. As both skewness and kurtosis values were greater than ±1.0, the distribution clearly cannot be considered normal. Furthermore, Shapiro-Wilk tests showed neither variable Non_CV19 (W = 0.69038, p-value < 0.01) and variable CV19_era (W = 0.71681, p-value < 0.01) were normally distributed.

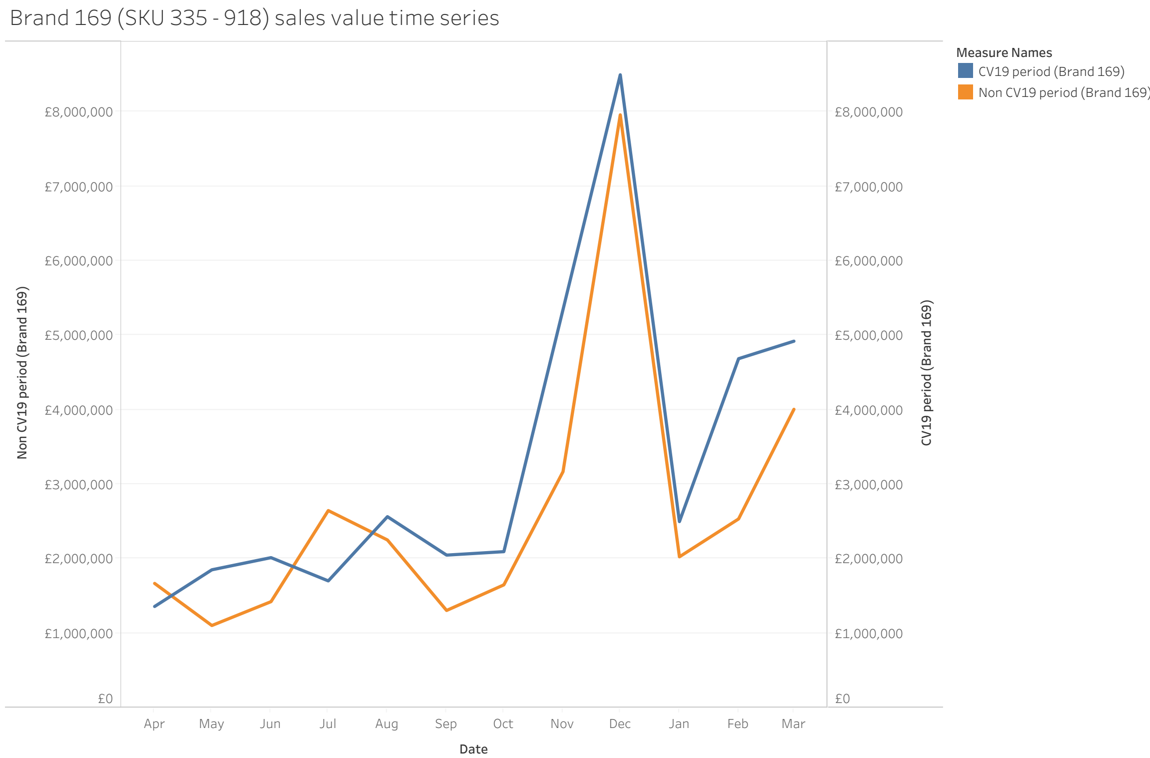

Figure 18. Brand 169 (SKU 335 - 918) sales value time series comparison (Source: original, conducted by author)

stats	Non_CV19	CV19_era
min	241849.9	278431
max	2609234	2249910
mean	610145.4	760909.6
p50	414249.5	536171.7
p75	672625.1	871390.1
skewness	2.293997	1.789476
kurtosis	8.179875	5.159208
sd	493736.7	537142.9

Figure 19 Descriptive statistics summary table - Brand 169 (SKU 336 – 918) sales value (Source: original, conducted by author)

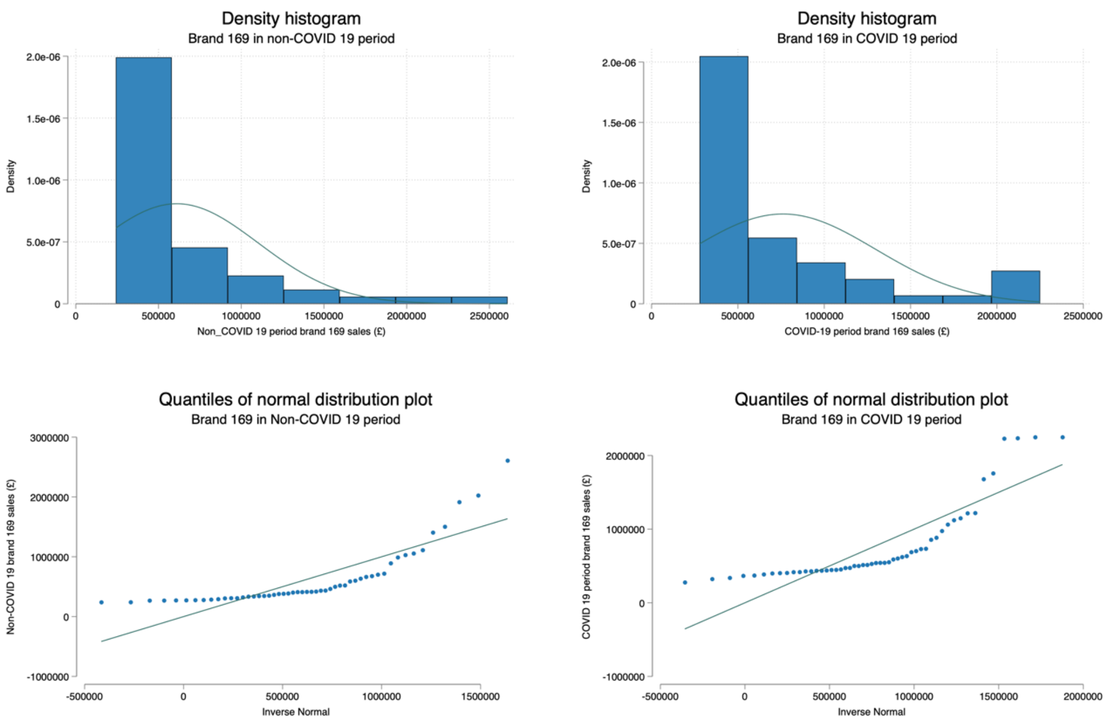

Figure 20. Distributional graphs – Brand 169 (Source: original, conducted by author)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Non_CV19	52	0.69038	15.019	5.791	0.00000
CV19_era	52	0.71681	13.737	5.601	0.00000

Figure 21. Shapiro-Wilk test for normality – Brand 169 sales (Source: Original, conducted by author)

A two-sample paired sign test (two-sided) was used to determine whether there was a significant difference between the median value of non-Covid period and that of COVID period for Brand 169. The following hypothesis was formulated:

H_0: The median of the difference between two samples is equal to 0

H_A: The median of the difference between two samples is not equal to 0

The resulting p-value was lower than the chosen significance level of 0.05. Therefore, the null hypothesis can be rejected and there was a statistically difference of the median of two samples.

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |          10          26
    negative |          42          26
        zero |           0           0
-------------+------------------------
         all |          52          52

One-sided tests:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era > 0
      Pr(#positive >= 10) =
         Binomial(n = 52, x >= 10, p = 0.5) =  1.0000

  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era < 0
      Pr(#negative >= 42) =
         Binomial(n = 52, x >= 42, p = 0.5) =  0.0000

Two-sided test:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era != 0
      Pr(#positive >= 42 or #negative >= 42) =
         min(1, 2*Binomial(n = 52, x >= 42, p = 0.5)) =  0.0000
```

Figure 22. Sign Test results – Brand 169 sales (Source: original, conducted by author)

When incorporating COVID-19 data to test how it impacted total sales of Brand 169, the scatterplot with LOWESS line showed a similar pattern to that of the category overall. Sales increased up to the point of around 200,000 cases. The relationship was moderately positive: $r_s = 0.4705$, p-value < 0.01.

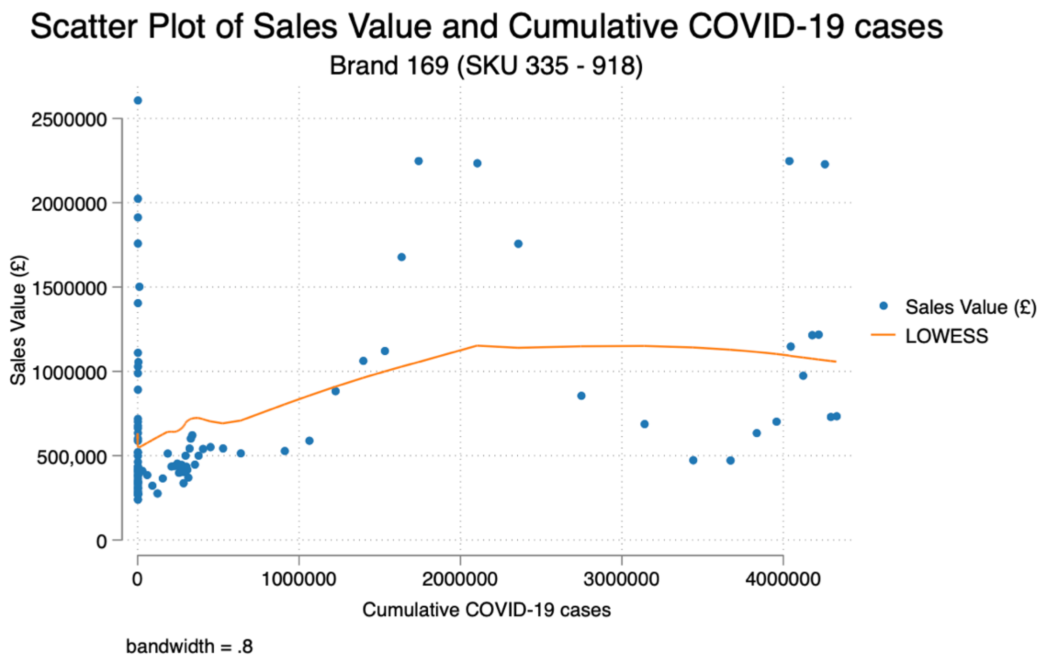

Figure 23. Brand 169 - Scatter Plot of Sales Value and Cumulative weekly COVID-19 cases (Source: original, conducted by author)

4.1.3. Brand 120

Sales value during the COVID-19 period was 30% year-over-year higher than the non-COVID period, totalling £14,522,647.60. The density histograms and Q-Q plots showed highly right-skewed distributions of both periods. A substantial difference exists between the mean and median of the data, and the skewness and kurtosis values fall outside the range of ±1.0, which indicates that the distribution of the data is not normal. Results of Shapiro-Wilk test at 99% confidence interval also showed that the distribution statistically deviated from the normality for both period Non_CV19 ($W = 0.73265$, $p\text{-value} < 0.01$) and period CV19_era ($W = 0.86835$, $p\text{-value} < 0.01$).

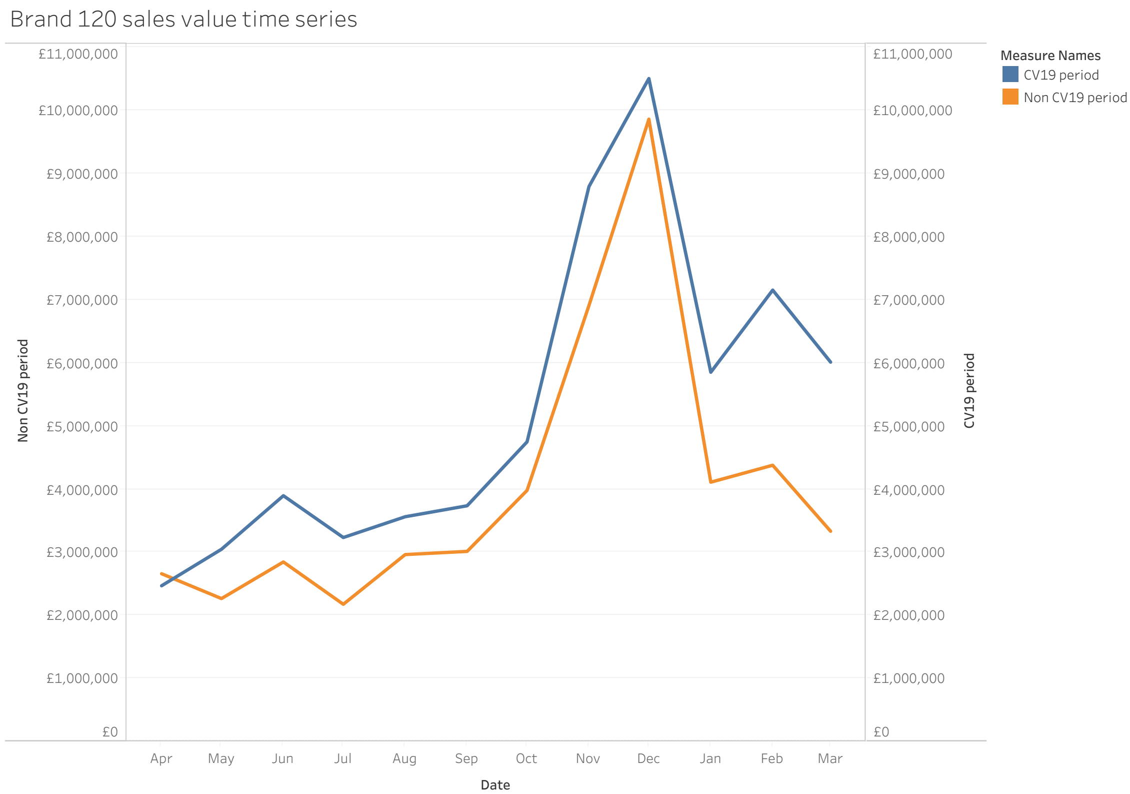

Figure 24. Brand 120 sales value time series comparison (Source: original, conducted by author)

stats	Non_CV19	CV19_era
min	493248.9	566620.2
max	3074460	2869072
mean	932332	1211614
p50	800411.9	1066090
p75	1022432	1422659
skewness	2.170396	1.225207
kurtosis	7.719522	3.962246
sd	537113.7	583148.5

Figure 25. Descriptive statistics summary table - Brand 120 sales (Source: original, conducted by author)

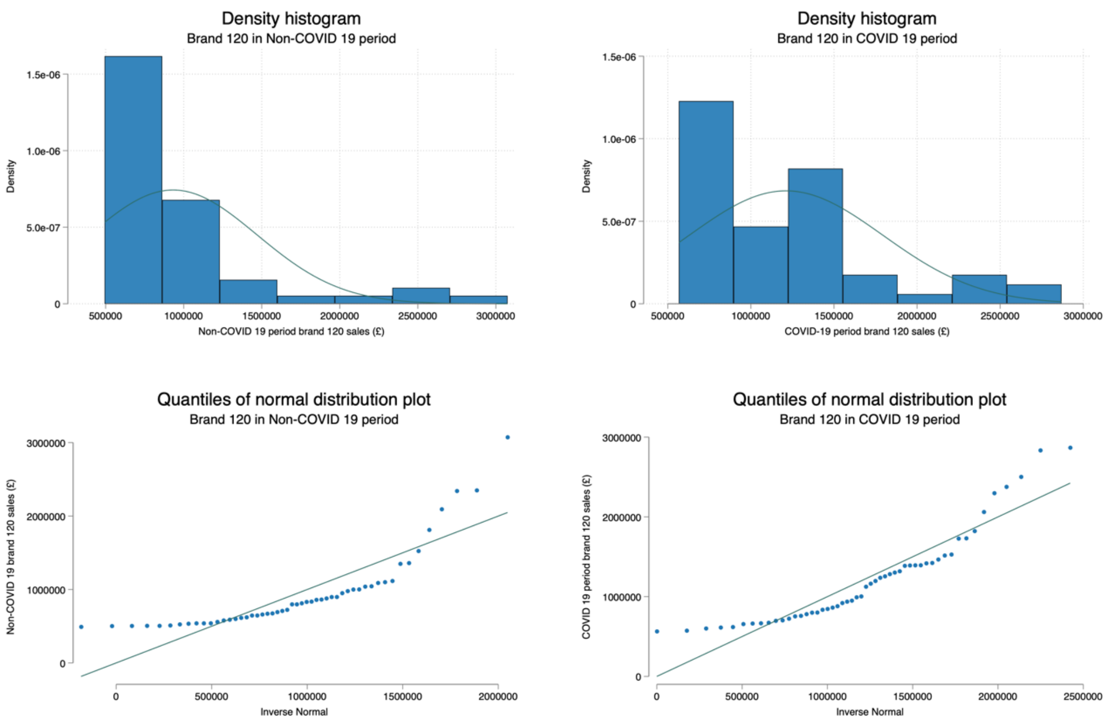

Figure 26. Distributional graphs – Brand 120 (Source: original, conducted by author)

Shapiro–Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Non_CV19	52	0.73265	12.968	5.478	0.00000
CV19_era	52	0.86835	6.386	3.963	0.00004

Figure 27. Shapiro-Wilk test for normality – Brand 120 sales (Source: Original, conducted by author)

Consequently, a paired sample sign test was carried out to ascertain whether there was a significant difference between the median value of non-Covid period and that of COVID period for Brand 120. The following hypothesis was formulated:

H_0: The median of the difference between two samples is equal to 0

H_A: The median of the difference between two samples is not equal to 0

The resulting p-value was lower than the chosen significance level of 0.05. Thus, the null hypothesis can be rejected and there was a statistically difference of the median of two samples.

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           6          26
    negative |          46          26
        zero |           0           0
-------------+------------------------
         all |          52          52

One-sided tests:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era > 0
      Pr(#positive >= 6) =
         Binomial(n = 52, x >= 6, p = 0.5) =  1.0000

  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era < 0
      Pr(#negative >= 46) =
         Binomial(n = 52, x >= 46, p = 0.5) =  0.0000

Two-sided test:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era != 0
      Pr(#positive >= 46 or #negative >= 46) =
         min(1, 2*Binomial(n = 52, x >= 46, p = 0.5)) =  0.0000
```

Figure 28. Sign Test results – Brand 120 sales (Source: original, conducted by author)

A Spearman's correlation was performed to test how reported COVID-19 cases statistically affected Brand 120's total sales. The results showed a moderate positive relationship between our two variables ($r_s = 0.4856$, p-value < 0.01). The following scatter plot showed an identical pattern compared to that of the whole category as sales witnessed a surge up to approximately 200,000 – 250,000 cases, then started to decline.

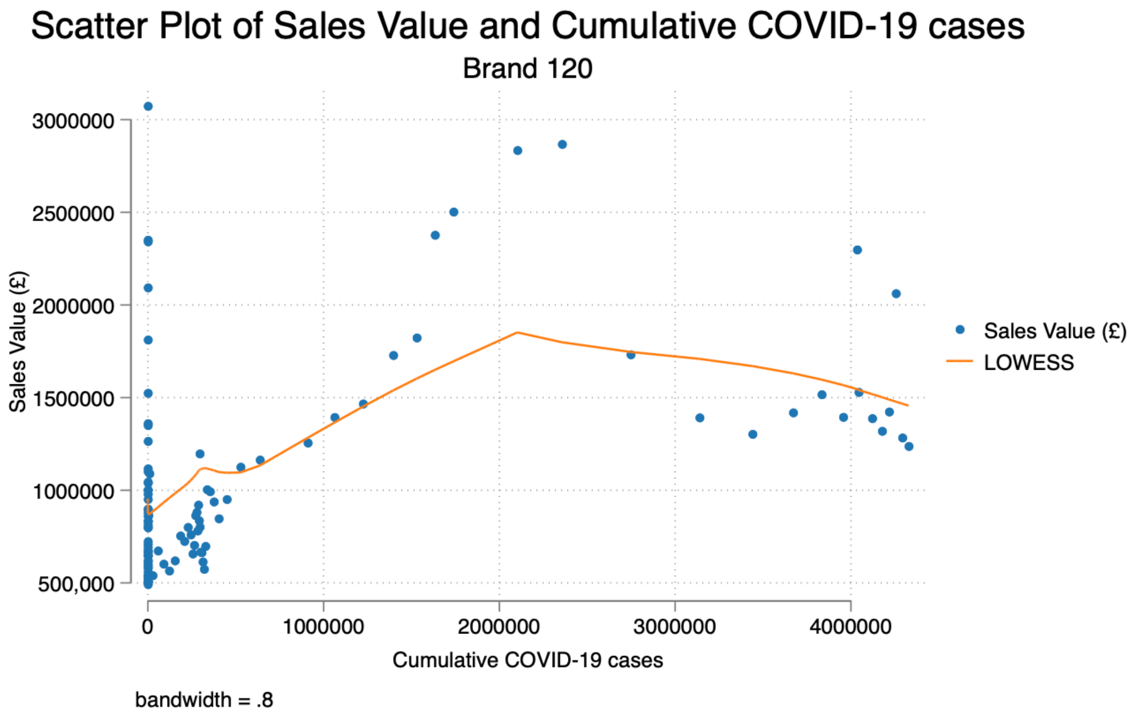

Figure 29. Brand 120 - Scatter Plot of Sales Value and Cumulative weekly COVID-19 cases (Source: original, conducted by author)

4.1.4. Brand 42

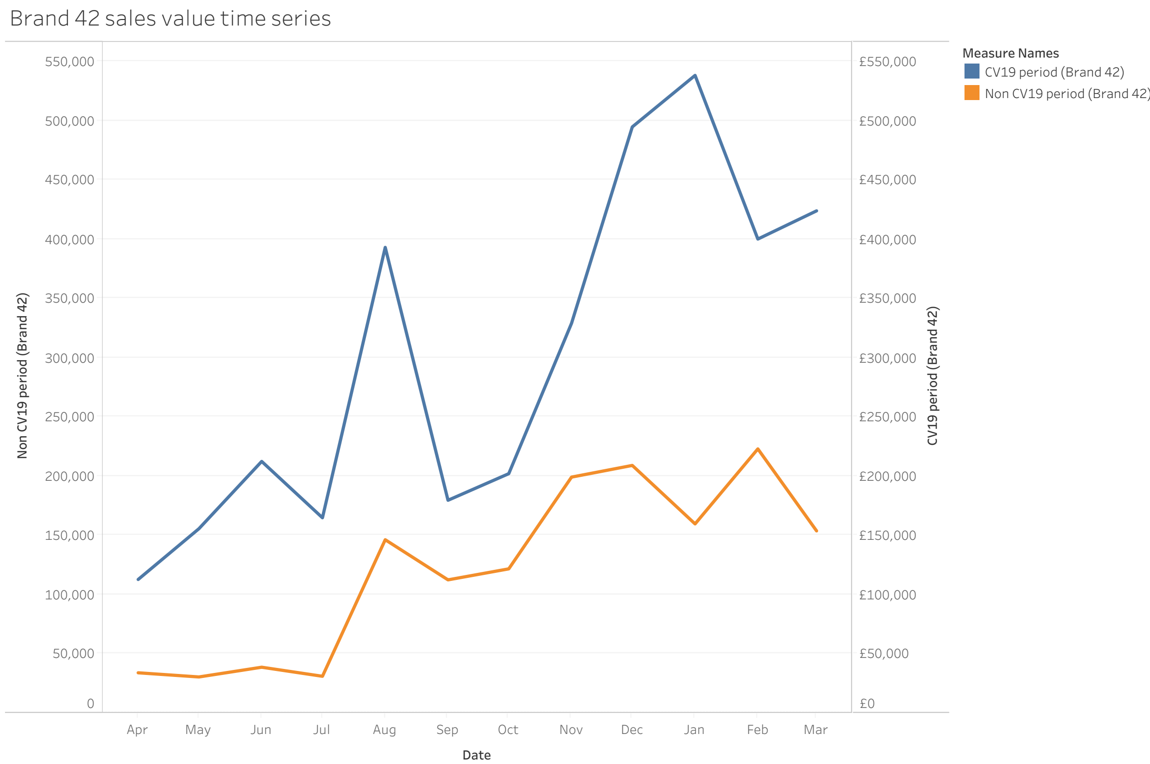

Figure 30. Brand 42 sales value time series comparison (Source: original, conducted by author)

Brand 42 experienced a phenomenon increase at 147% year-over year (£2,147,834.90) of sales value during the COVID-19 period compared to the pre-COVID-19 periods. In figure 32, the graphical representation of both periods illustrated a positively skewed distribution. The implication drawn from Shapiro-Wilk tests was that both periods did not have a normal distribution (Non_CV19: W = 0.89910, p-value < 0.01; CV19_era: W = 0.76874, p-value = 0.01).

stats	Non_CV19	CV19_era
min	6061.6	26201.4
max	67258.4	235600.2
mean	27883.77	69358.15
p50	31644.7	54004.9
p75	40854.6	78271.95
skewness	.1688979	1.955988
kurtosis	1.920198	6.611546
sd	17267.12	45065.3

Figure 31. Descriptive statistics summary table - Brand 42 sales (Source: original, conducted by author)

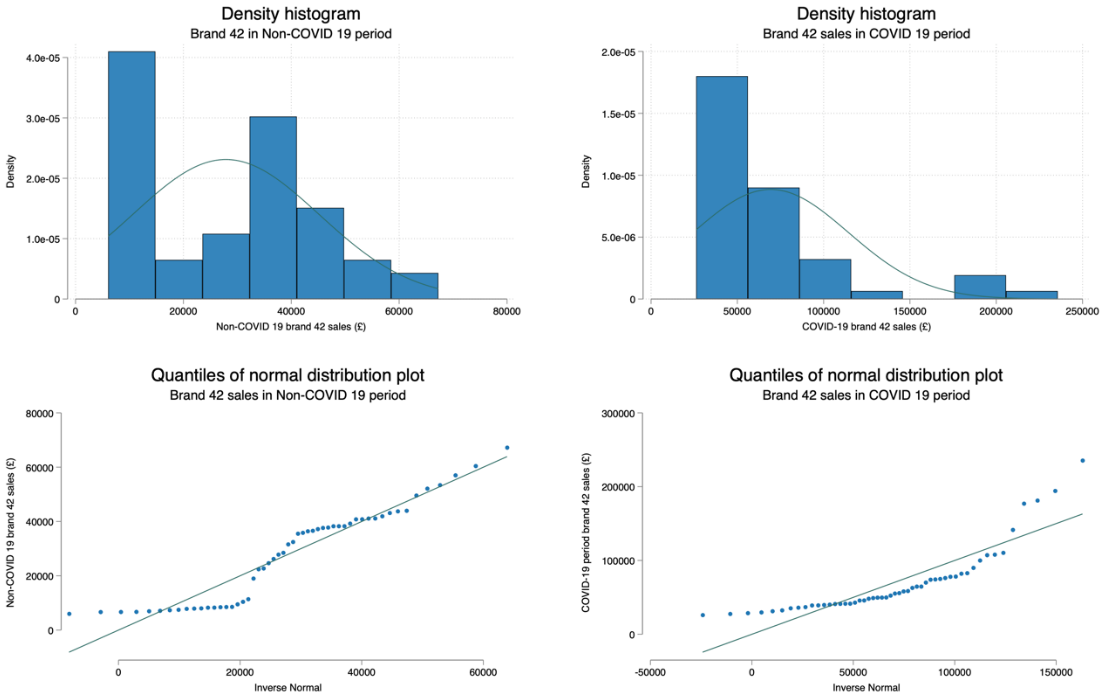

Figure 32. Distributional graphs – Brand 42 (Source: original, conducted by author)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Non_CV19	53	0.89910	4.969	3.431	0.00030
CV19_era	52	0.76874	11.218	5.168	0.00000

Figure 33. Shapiro-Wilk test for normality – Brand 42 sales (Source: Original, conducted by author)

A paired sample sign test was carried out to determine if there were a statistically significant difference median value of non-Covid period and that of COVID period for Brand 42. The following hypothesis was formulated:

H_0: The median of the difference between two samples is equal to 0

H_A: The median of the difference between two samples is not equal to 0

The resulting p-value was lower than the chosen significance level of 0.05. Therefore, the null hypothesis can be rejected and there was a statistically difference of the median of two samples.

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           0          26
    negative |          52          26
        zero |           0           0
-------------+------------------------
         all |          52          52

One-sided tests:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era > 0
      Pr(#positive >= 0) =
         Binomial(n = 52, x >= 0, p = 0.5) =  1.0000

  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era < 0
      Pr(#negative >= 52) =
         Binomial(n = 52, x >= 52, p = 0.5) =  0.0000

Two-sided test:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era != 0
      Pr(#positive >= 52 or #negative >= 52) =
         min(1, 2*Binomial(n = 52, x >= 52, p = 0.5)) =  0.0000
```

Figure 34. Sign Test results – Brand 42 sales (Source: original, conducted by author)

When including COVID-19 cases as a factor, Spearman's correlation results clearly concluded that there was a strong positive relationship between cumulative COVID-19 cases and Brand 42's total sales value ($r_s = 0.7696$, p-value < 0.01). The scatter plot below illustrated that sales increased as the COVID-19 cases surged, up to the point of around 250,000 cases then it decreased and a negative relationship was developed.

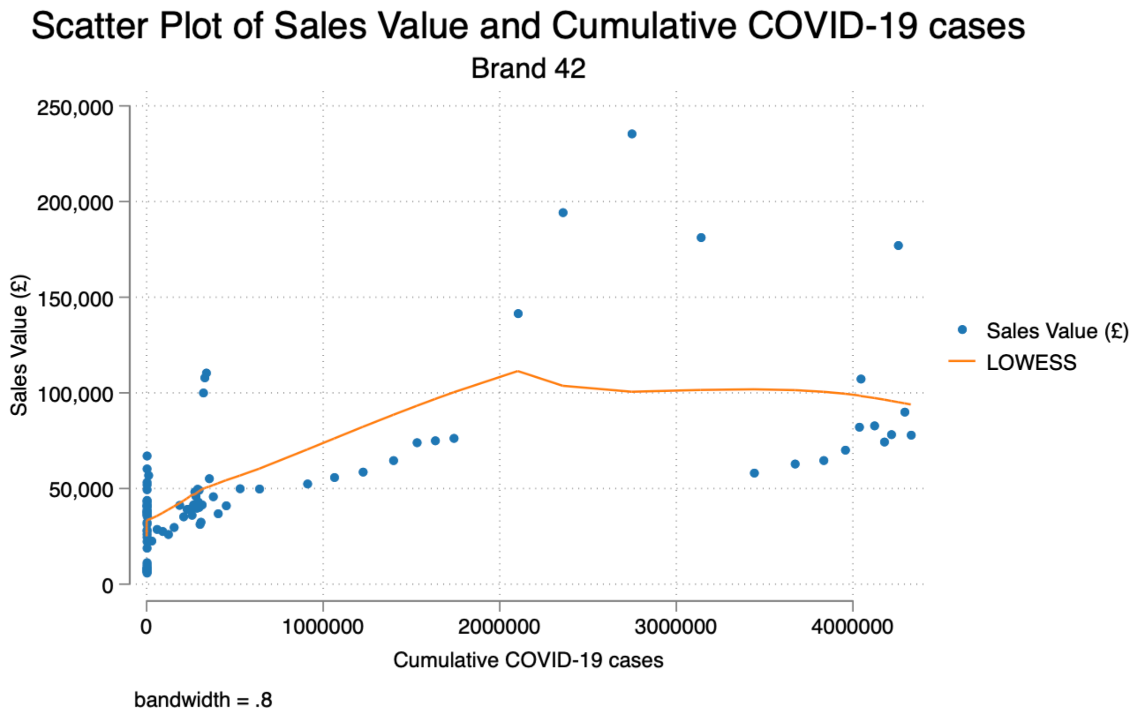

Figure 35. Brand 42 - Scatter Plot of Sales Value and Cumulative weekly COVID-19 cases (Source: original, conducted by author)

4.1.5. Brand 3

The COVID-19 period witnessed an increase of £1,301,895.50 (around 31% year-over-year change) in sales comparing to the non-COVID-19 period. There was evidence found in the descriptive statistics summary table and graphical illustrations that the data of both periods marginally deviated from the normality. Skewness score was 0.6858409 for non-COVID-19 period and was 0.8473334 for COVID-19 period. Kurtosis value for both variables were greater than the range ±1.0 to be considered as normally distributed.

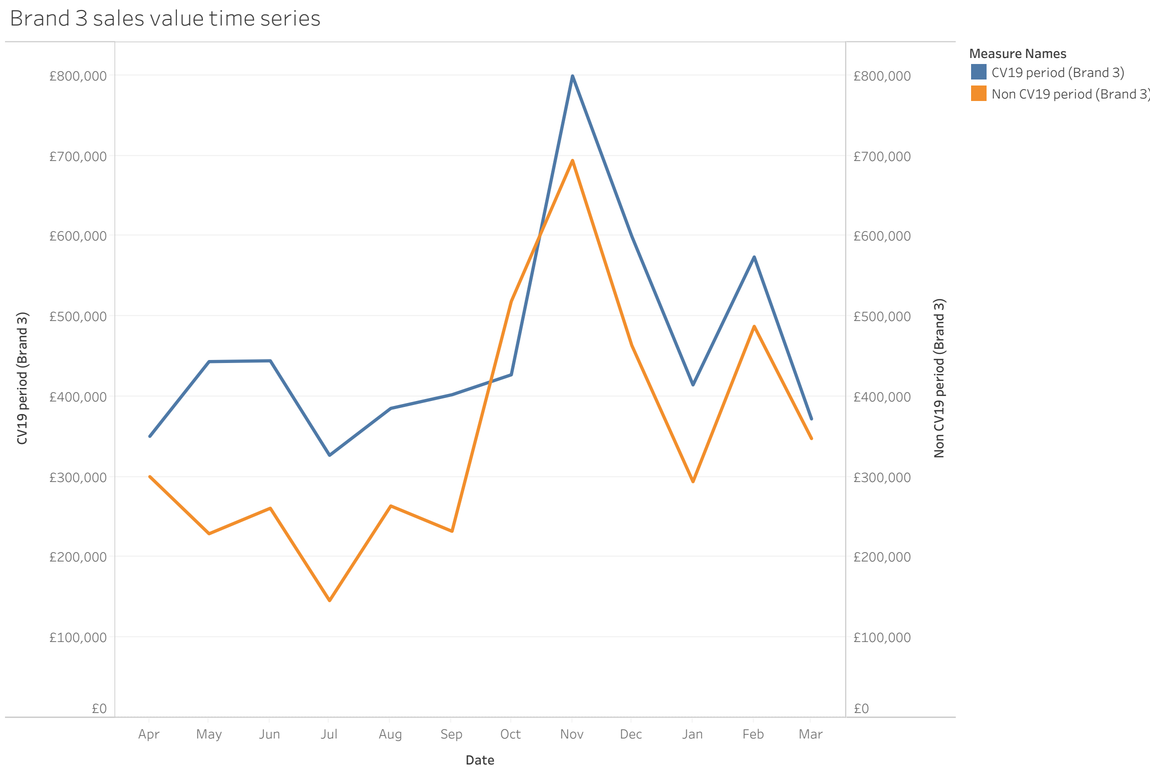

Figure 36. Brand 3 sales value time series comparison (Source: original, conducted by author)

stats	Non_CV19	CV19_era
min	31894.5	72593.3
max	160945	167446.1
mean	81506.3	106542.8
p50	72520.4	102051.5
p75	107308.1	118586.1
skewness	.6858409	.8473334
kurtosis	2.444796	2.788284
sd	35114.73	27069.36

Figure 37. Descriptive statistics summary table - Brand 3 sales (Source: original, conducted by author)

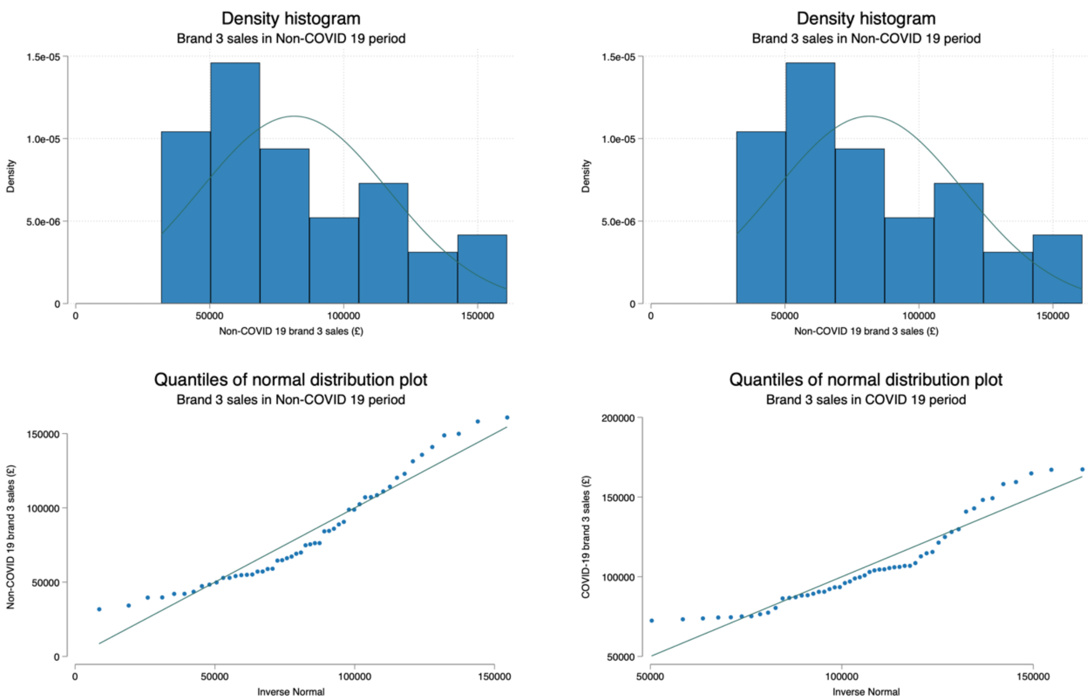

Figure 38. Distributional graphs – Brand 3 (Source: original, conducted by author)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Non_CV19	52	0.92705	3.538	2.701	0.00345
CV19_era	52	0.90194	4.757	3.334	0.00043

Figure 39. Shapiro-Wilk test for normality – Brand 3 sales (Source: Original, conducted by author)

Results derived from Shapiro-Wilk normality tests showed that the distributions were significantly non-normal for both Non_CV19 (W = 0.92705, p-value < 0.01) and CV19_era (W = 0.90194, p-value < 0.01). Thus, non-parametric significance test was chosen instead of parametric one. A two-sample paired sign test was carried out to determine if there were a statistically significant difference median value of non-Covid period and that of COVID period for Brand 3. The following hypothesis was formulated:

H_0: The median of the difference between two samples is equal to 0

H_A: The median of the difference between two samples is not equal to 0

The resulting p-value was lower than the chosen significance level of 0.05. Therefore, the null hypothesis can be rejected and there was a statistically difference of the median of two samples.

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |          10          26
    negative |          42          26
        zero |           0           0
-------------+------------------------
         all |          52          52

One-sided tests:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era > 0
      Pr(#positive >= 10) =
         Binomial(n = 52, x >= 10, p = 0.5) =  1.0000

  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era < 0
      Pr(#negative >= 42) =
         Binomial(n = 52, x >= 42, p = 0.5) =  0.0000

Two-sided test:
  Ho: median of Non_CV19 - CV19_era = 0 vs.
  Ha: median of Non_CV19 - CV19_era != 0
      Pr(#positive >= 42 or #negative >= 42) =
         min(1, 2*Binomial(n = 52, x >= 42, p = 0.5)) =  0.0000
```

Figure 40. Sign Test results – Brand 3 sales (Source: original, conducted by author)

To better scrutinize the impact of the pandemic on Brand 3’s sales, a Spearman’s correlation was conducted and the result showed that there was a moderate positive relationship between cumulative weekly COVID-19 cases and Brand 3’s total sales value ($r_s = 0.4692$, p-value < 0.01). The scatter plot below illustrated that sales increased as the COVID-19 cases surged, up to the point of around 180,000 - 250,000 cases then it started to decrease.

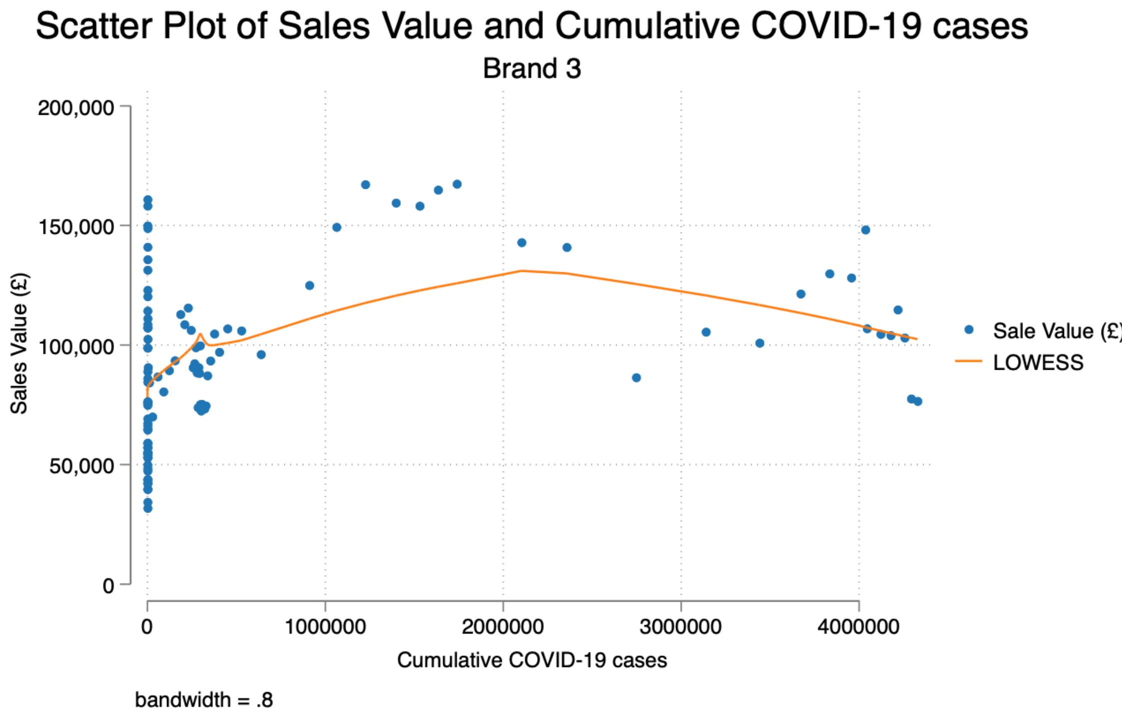

Figure 41. Brand 3 - Scatter Plot of Sales Value and Cumulative weekly COVID-19 cases (Source: original, conducted by author)

4.2. Leading brands and SKUs – Value Share[3]

4.2.1. Category level

Figure 42 showed the brands with largest sales volume and category percentage shares in the period 2019 – 2021. Brand 120 accounted for the most significant share at nearly half of category sales (43.44%), followed by brand 169 at 31.99%. Together with brand 3, these brands are also the brands of interests, ranked first, second and fourth at 43.44%, 31.99% and 3.78%, respectively, in sales volume. The fact that brands 120 and 169 contribute to over 70% of category share posit their tremendous sway in statistical analyses. Total sample's sales behaviours are forecasted to mirror the pattern of brands 120 and 169. Brand was a significant factor in total sales and subsequent analyses should focus on SKU differentiations within these two brands and how other brands differ from them.

[3] Details of exact sales value for each brand/SKU can be found in the Appendices

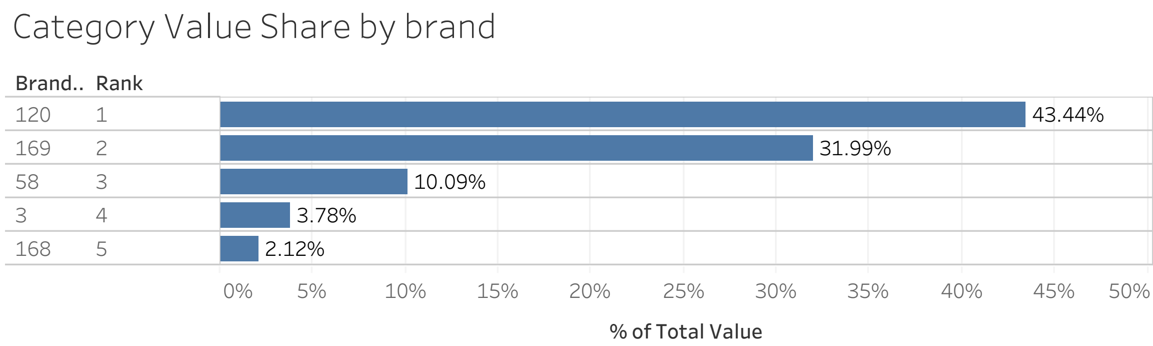

Figure 42. Top 5 leading brands (% sales value share) (March 2019 – March 2021) (Source: original, conducted by author)

Top 10 most significant SKUs and their brand groups are summarised in Figure 43. Most of those belong to brands of interests group (169, 120, 3). However, none of Brand 42's SKUs was significant enough at the category level. While two specific brands capture over 70% of category share, SKUs sales dominance are fractured with each top 10 SKU occupied from 0.78% to 1.04% of total value. The combined category share % of top 10 SKUs is under 10%. This indicates wide product range with diverse product differentiation.

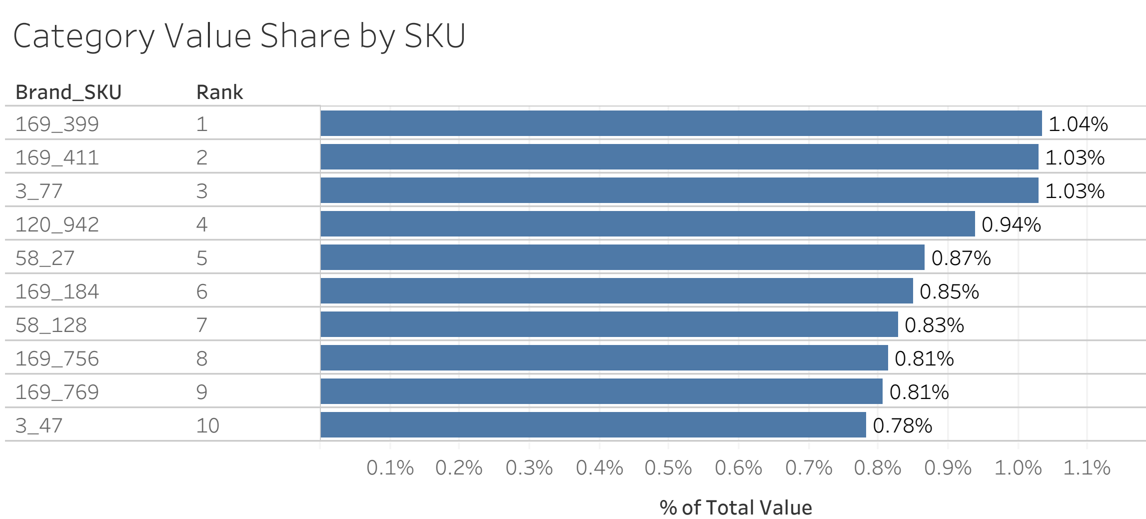

Figure 43. Top 10 leading SKUs at category level (% sale value share) (Source: Original, conducted by author)

4.2.2. Brand 169 (SKU 335 – 918)

As seen in Figure 44, for Brand 169 (SKU 335 – 918), the top 5 best-selling SKUs contributed to over 10% of total sales. No SKUs had significant dominance in sales contribution and top SKUs competed closely with average value gaps between each ranking less than £100,000. Considered that brand 169 was the second-best selling brand, this indicated a diversified and well-segmented product range wherein individual SKUs were able to target the diverse customers profiles, resulting in equally distributed SKUs sales contribution.

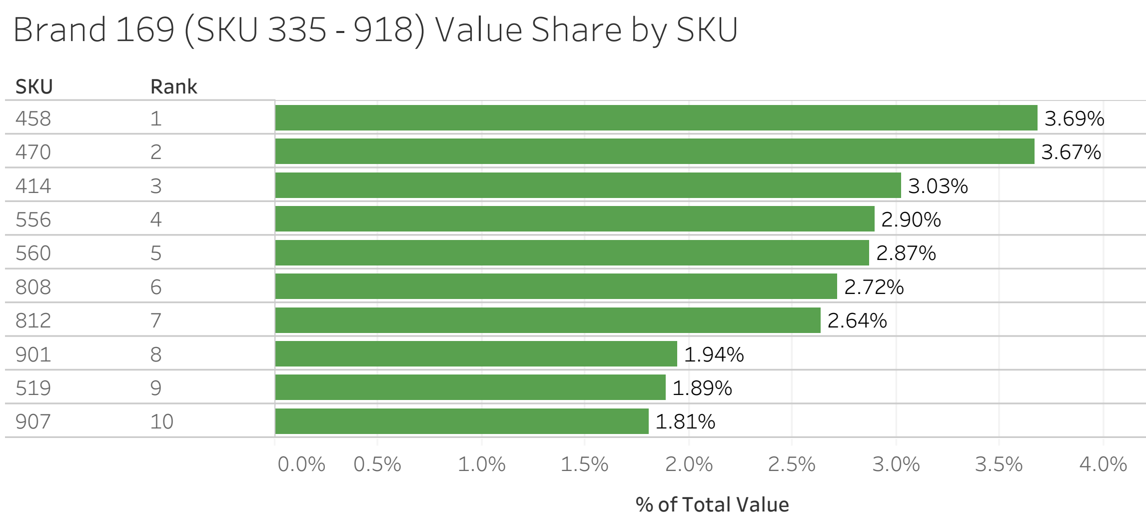

Figure 44. Brand 169 Top 10 best-selling SKUs (Range 335 - 918) (Source: original, conducted by author)

4.2.3. Brand 120

For Brand 120, the top 5 leading SKUs contributed to less than 8% of total sales. No SKUs has more than 2.5% of sales value share within the brand and only SKU 1161 has a value share larger than 2%. Given that brand 120 and brand 169 occurred to have the highest value share of the category, this indicated a situation related to brand 169. Here, the customer base was extremely diverse, and the sales volume generated was primarily made up of many strong products rather than a few isolated SKUs.

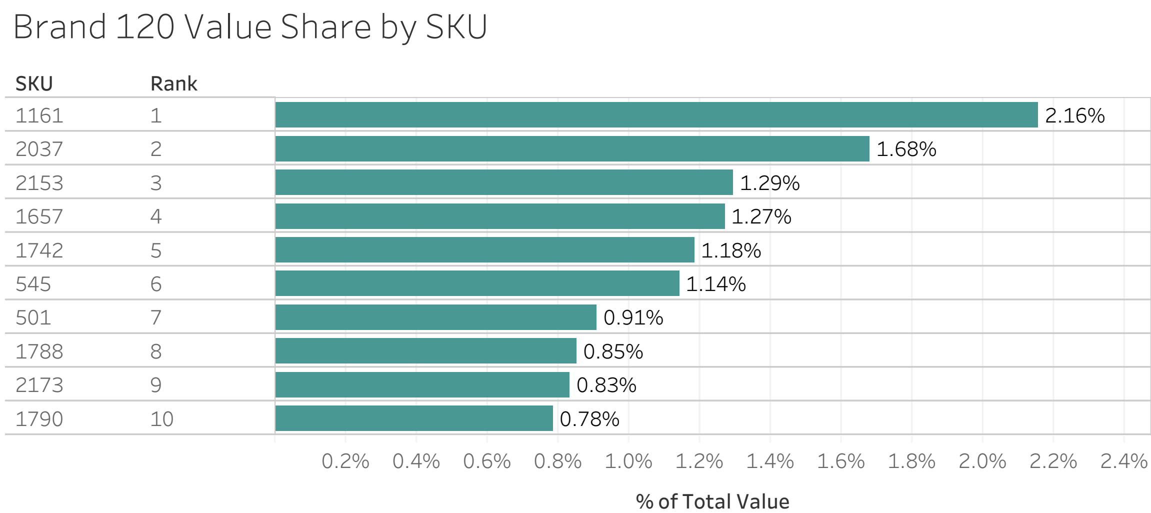

Figure 45. Brand 120 Top 10 best-selling SKUs (Source: original, conducted by author)

4.2.4. Brand 42

In the case of Brand 42, the top five best-selling SKUs contributed to nearly 40% of total sales. The top three SKUs each contributed nearly 10% of sales value share, more than triple the tenth-ranked SKU 27. The situation of brand 42 was dramatically different from that of brands 120 and 169, where a broad variety of SKUs shared equally in the value of sales. Only a few strong SKUs contributed to most of brand 42's sales, indicating a smaller product range.

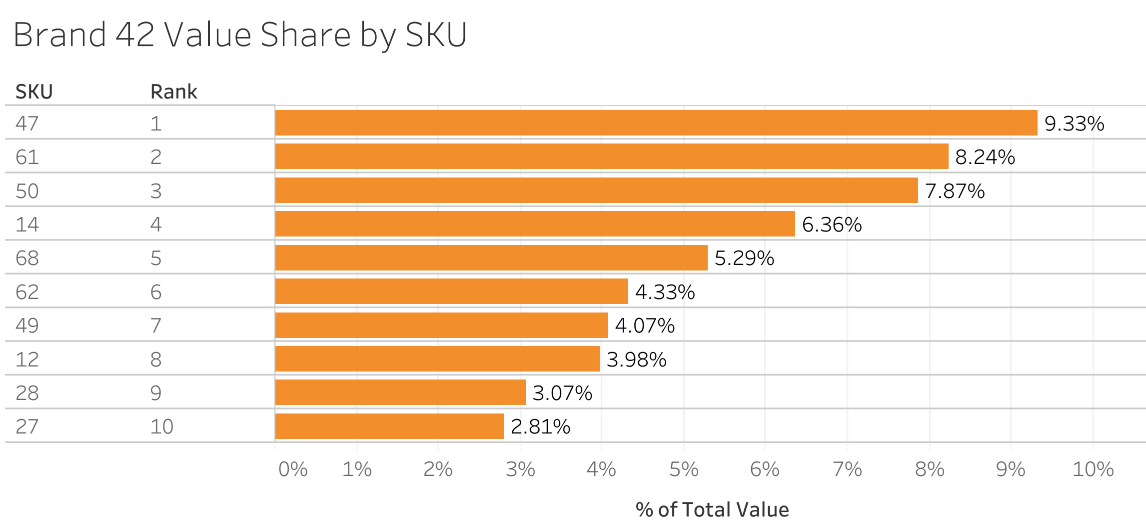

Figure 46. Brand 42 Top 10 best-selling SKUs (Source: original, conducted by author)

4.2.5. Brand 3

For Brand 3, the top 5 best-selling SKUs contributed to over 70% of total sales. The best-selling SKU 46 contributed to nearly 30% of sales value share, which together with second-best selling SKU 24 at around 20% dominated the shares of sales. The domination of few SKUs in total sales value shares was even more drastic than that of brand 42. This indicated that brand 3 relied strongly on the performance of its top SKUs for overall brand performance. The product range was expected to be small.

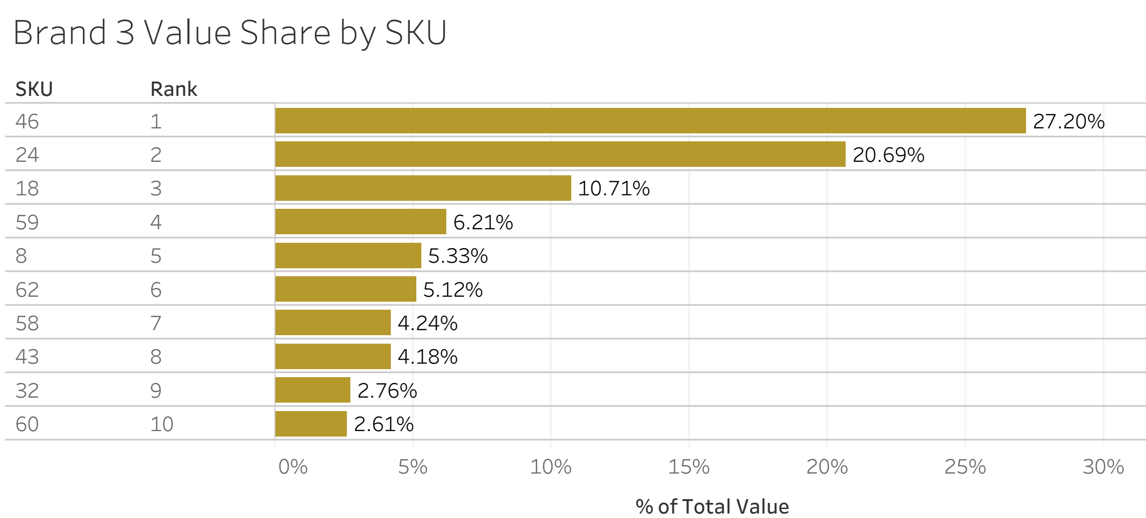

Figure 47. Brand 3 Top 10 best-selling SKUs (Source: original, conducted by author)

4.3. Leading Brands and SKUs – Rate of sale (ROS)

4.3.1. Category level

ROS for each SKU and brand at category level was calculated following the Sales per point of distribution and Sales per total distribution points formula in section 3.9. However, as stated, velocity is a non-addictive fact; therefore, instead of summing up the measure for the whole period, which would return incorrect results, the average value was used to determine the fastest-selling brands and SKUs.

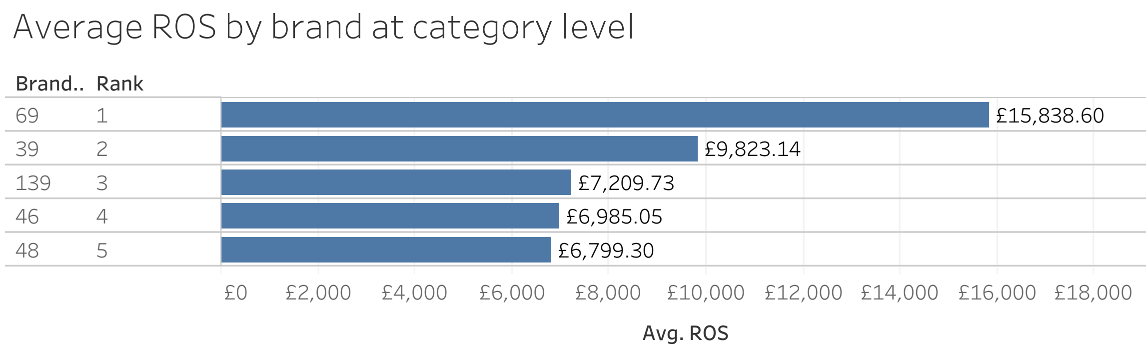

Figure 48. Average ROS by brand at category level (Source: original, conducted by author)

Figure 48 showed that among leading brands, there was a tremendous gap between the best performer – Brand 69 and the rest. Brand 69, despite having a limited value share, was sold at the highest rate with nearly 15 million units in total distribution points, outperforming Brand 39 with approximately 10 million units. Average ROS varied little between the top 10 SKUs and most of them belong to Brand 169, 120 and 168, as reported in Figure 49 below.

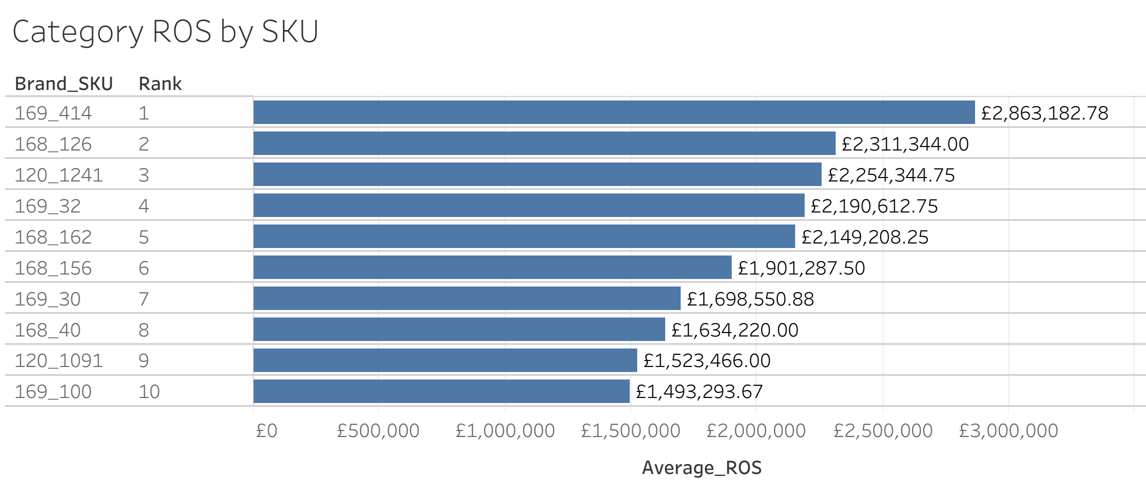

Figure 49. Average ROS by SKU at category level (Source: original, conducted by author)

4.3.2. Brand 169 (SKU 335 – 918)

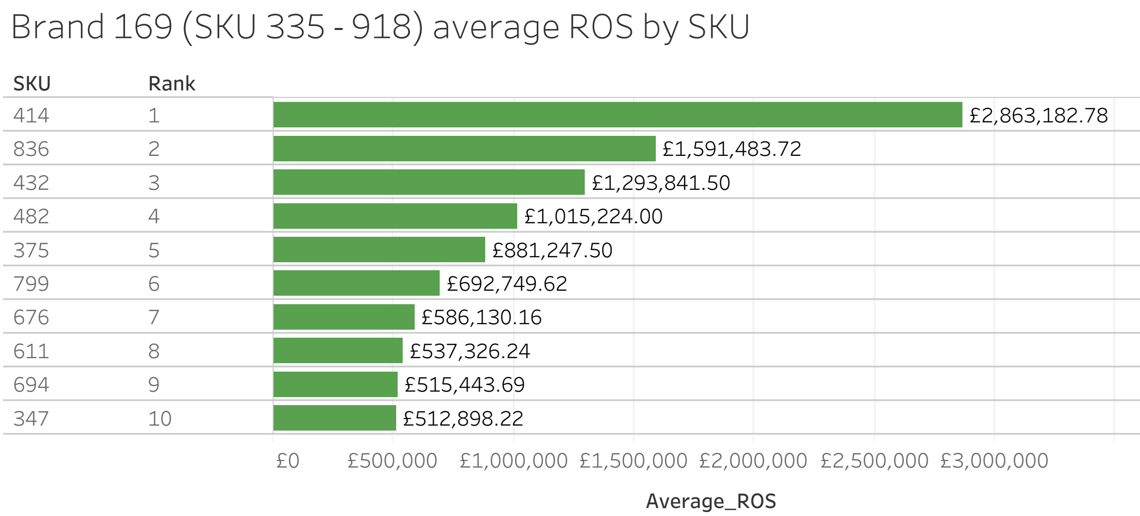

Figure* 50. *Brand 169 (SKU 335 - 918) average ROS by SKU (Source: original, conducted by author)

As seen in Figure 50, ROS for top selling SKUs in brand 169 varied greatly, with SKU 414 posited the highest average ROS that more than tripled those of the remaining SKUs in top ten. This indicated the immense consumer appetite for SKU 414 when it was available at stores. Recalled that this SKU were ranked the third in terms of total sales value for brand 169, it was deducible that increased availability could still further boost sales for this SKU.

4.3.3. Brand 120

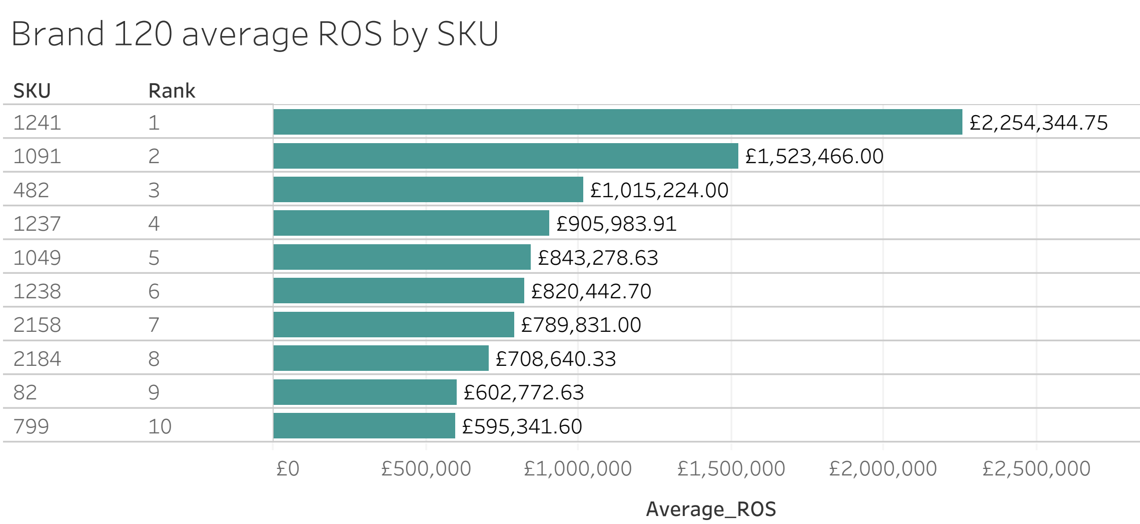

Figure 51. Brand 120 average ROS by SKU (Source: original, conducted by author)

Recalled that brand 120 had the most extensive product range, the average ROS further indicated approximately equal speed of purchase for top SKUs, with small discrepancies between SKUs, except for the SKU 1241. As seen in Figure 51, the average of ROS insinuated that brand 120's availability did not vary widely for its wide range of SKUs. However, maintaining a wide range of product with high availability at store could underline heavy logistic costs. This risk could be alleviated by the general stockpiling behaviour of consumers during the pandemic.

4.3.4. Brand 42

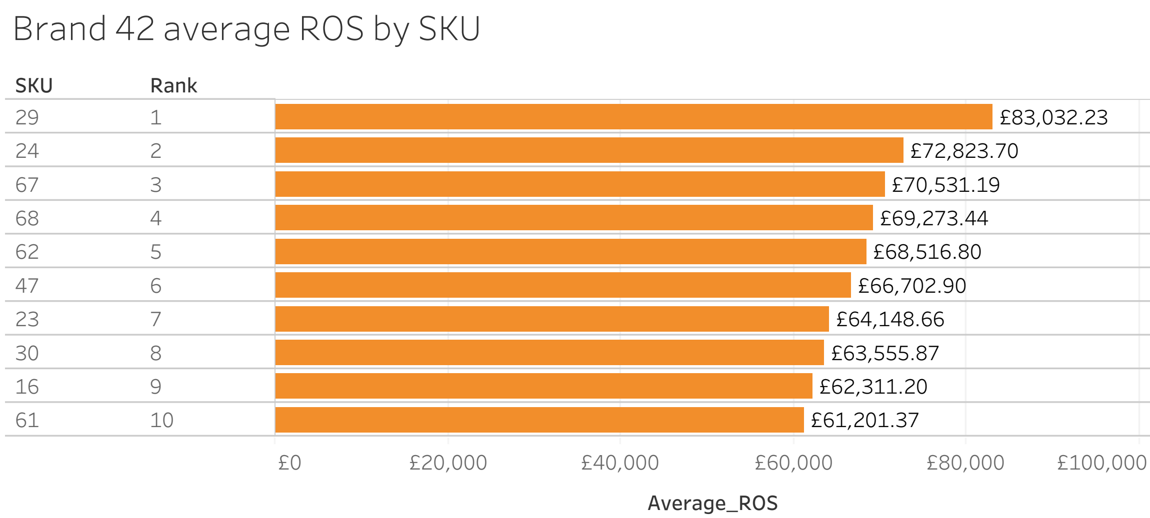

Figure 52. Brand 42 average ROS by SKU (Source: original, conducted by author)

Similar to brand 120, brand 42's average of ROS did not vary greatly among top SKUs. However, there were several SKUs with high speed of sales that ranked lower in brand 42's sales values – namely SKUs 29 and 24, indicating that consumers snatched up these items as soon as they were available at stores and their stock availability could further be improved to maximize sales values.

4.3.5. Brand 3

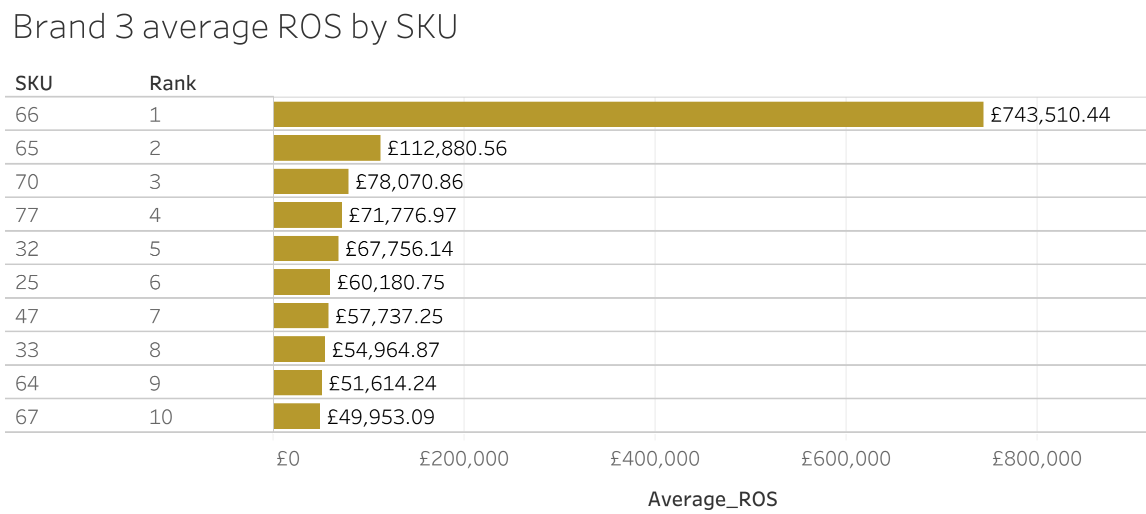

Figure 53. Brand 3 average ROS by SKU (Source: original, conducted by author)

ROS of brand 3 indicated that it was facing a severe logistical problem, where the speed of purchase for SKU 56 was nearly seven times those of the remaining SKUs. Most critically, SKU 66 was not the number one best seller for brand 3 with highest sales values, indicating that it was not available for purchase to maximize sales. Recalled that brand 3 and brand 42 had the smallest product ranges among the brands of interests, it would be highly recommendable for brand 3 to strengthen its product availability at stores, as it could not afford to lose sales on top-selling products.

4.4. Product size

Specifically for this section, the weight data points denoted as "Not applicable" and "Estimated weight" in the original dataset were treated as outliers and removed. As the weight range varies for the category overall and most of the focused brands, grouping was executed to determine which weight group was performing the best.

4.4.1. Category level

As seen in figure 54 and at category level, the smallest weight group (3 – 200 grams) generated the most sales value, totally over £111 million in the period, followed by 205 - 400 grams with approximately £33 million sales. These two categories contributed to over 80% of total sales for the period, signifying a clear preference for small sizes, with products weighting 405 – 600 grams trailing behind, accounting for just over 10% of total sales. Larger size groups, 609.5 – 800 grams and 805 – 1134 grams, both underperformed, with the latter group contributing an insignificant amount of £111,143.30.

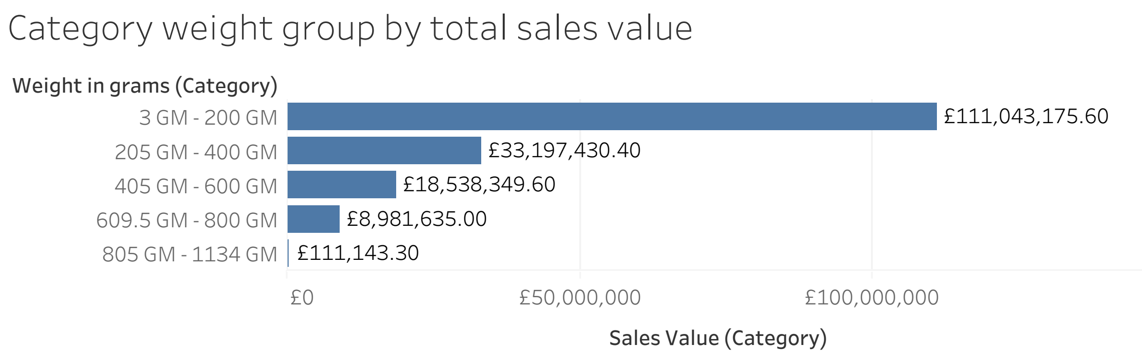

Figure 54. Category weight group by sales value (Source: original, conducted by author)

In terms of ROS, however, the most successful products ranged from 609.5 up to 800 grams, followed closely by the smallest weight group, which fell between 3 and 200 grams, providing ROS of around 630 and 500, respectively. In line with the comparison of sales value, the largest weight group (805 - 1134 grams) performed poorly at £128.68 ROS, suggesting consumers hesitate when purchasing larger candles. Possibly this is because of the price difference per unit, as large sizes tend to be more costly, or perhaps this is because consumers prefer small to medium size bottles for experimenting with various scents and or to give as gifts at an affordable price.

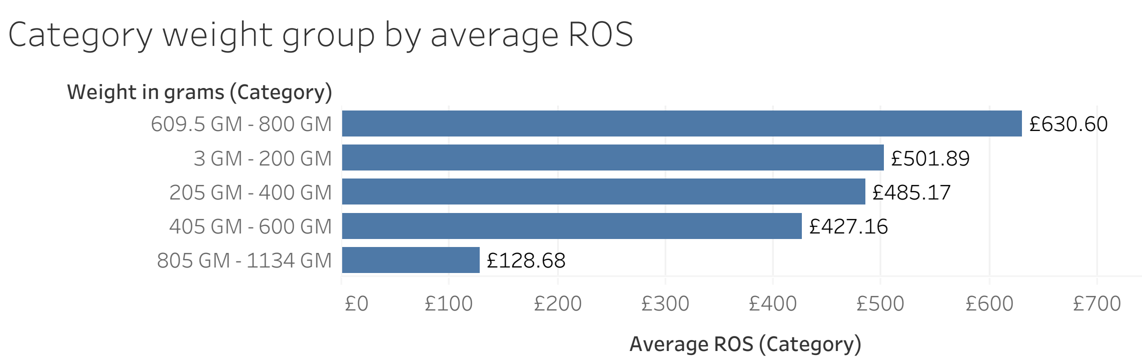

Figure 55. Category weight group by average ROS (Source: original, conducted by author)

4.4.2. Brand 169 (SKU 335 – 918)

The revenue pattern of brand 169 was different from those of the category overall with two dominant weights, 340 and 538 grams, belonging to medium range and accumulating 76% of the brand's total. An enormous gap existed between these sizes and those that floundered, for instance, size 198 and 159 grams accounted for less than £100 thousand in sales. ROS statistics from Figure 57 showed that products weighing 340 grams lost their prominence in favour of those weighing 538 grams and 104 grams. Hence, product weight did not significantly affect consumer decision-making when purchasing brand 169 products, but it is possible that some other factors played a role.

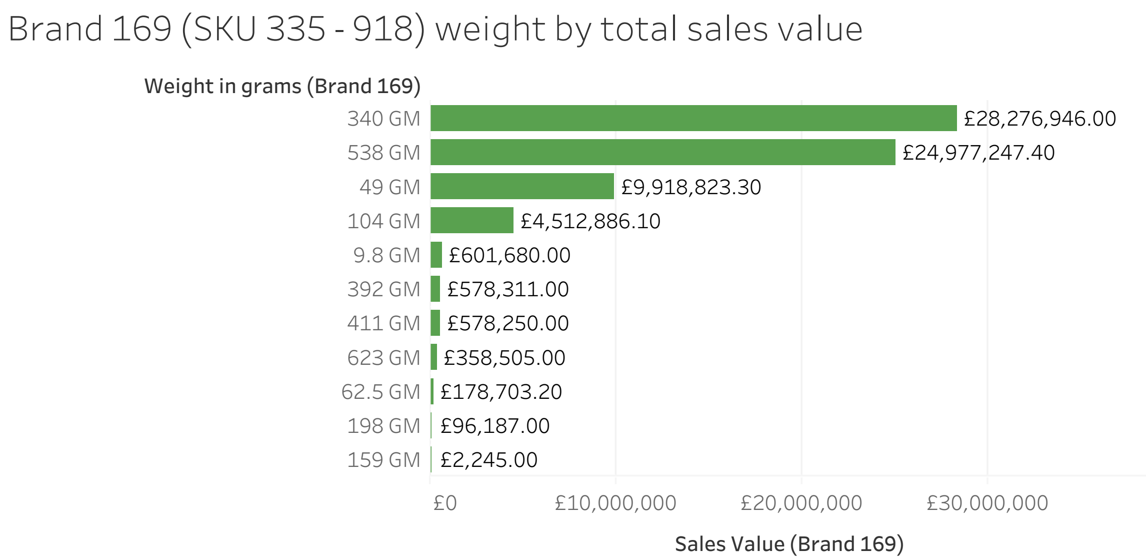

Figure 56. Brand 169 (SKU 335 - 918) weight group by sales value (Source: original, conducted by author)

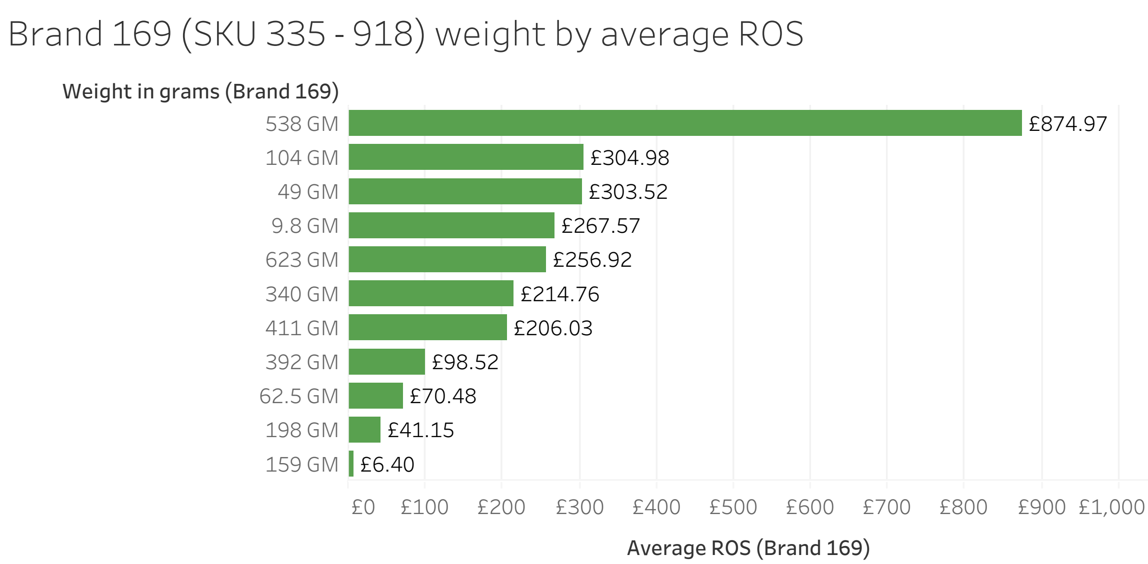

Figure 57. Brand 169 (SKU 335 - 918) weight group by average ROS (Source: original, conducted by author)

4.4.3. Brand 120

Brand 120's revenue was mainly derived from the following weight groups: 510 - 571 grams, 311 - 357 grams, 120 - 200 grams, and 120 - 200 grams, as illustrated in figure 58. In terms of ROS, buyers preferred the smallest weight group, followed by 510 – 571 grams at around £960 and £320, respectively. Product weights in the largest weight groups (700-3750 grams) almost did not sell at all, indicating that these sizes were insufficient for distribution in stores, as they also had TDP values of zero. Details of ROS statistics for each weight group were showed in Figure 59.

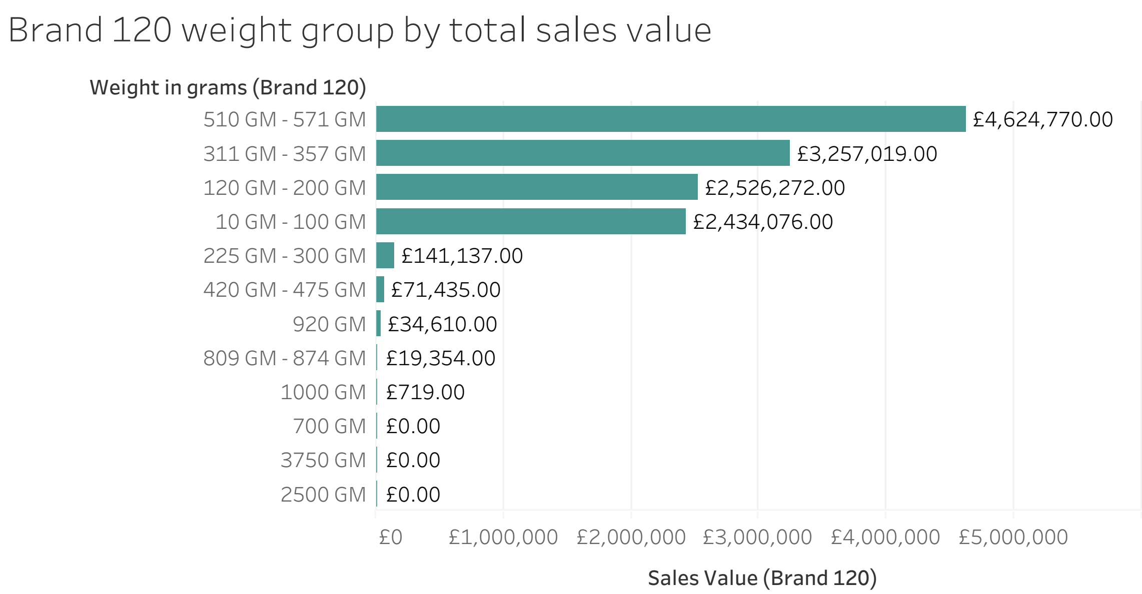

Figure 58. Brand 120 weight group by sales value (Source: original, conducted by author)

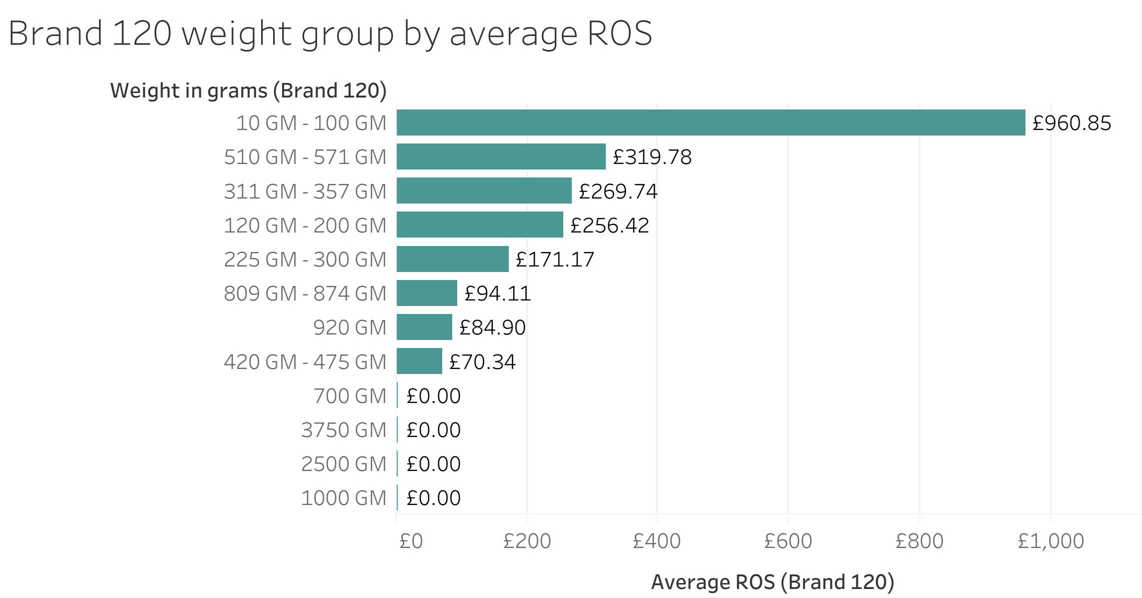

Figure 59. Brand 120 weight group by average ROS (Source: original, conducted by author)

4.4.4. Brand 42

For brand 42, after removing outliers, the only size remaining is 425 grams, which contributed to £1,473,342 in sales value. It is noteworthy that Brand 42's ROS for the period under review was £412.40, quite close to the category ROS for the weight group 405 - 600 grams reported earlier.

4.4.5. Brand 3

Figure 60 demonstrates how different weight groups of Brand 3 performed regarding sales value. As is evident, consumers prefer small sizes as products weighing 105 grams and under account for 80% of brand revenue. It may appear that the second-best group has a similar weighting, but its revenue makes up only one-fourth of that of the best performer. Neither of the larger sizes, 265 and 480 grams, were very popular. Regarding ROS, Brand 3's weight groups did not differ in ranking from those based on revenue contribution. Figure 61 shows a detailed breakdown of average ROS by size.

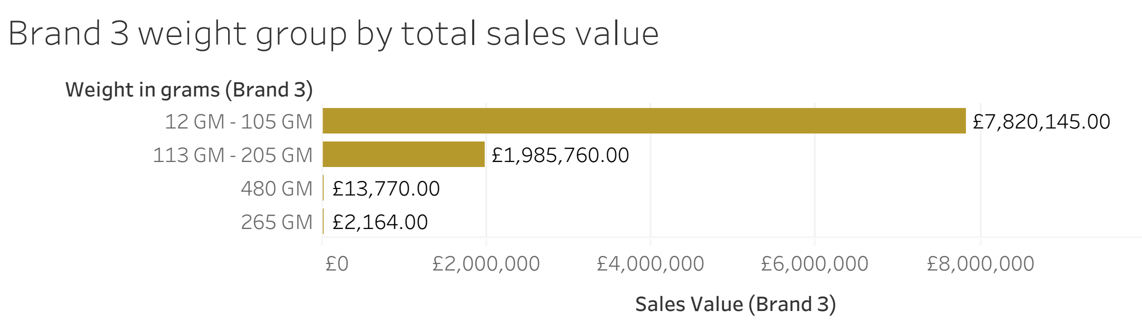

Figure 60. Brand 3 weight group by sales value (Source: original, conducted by author)

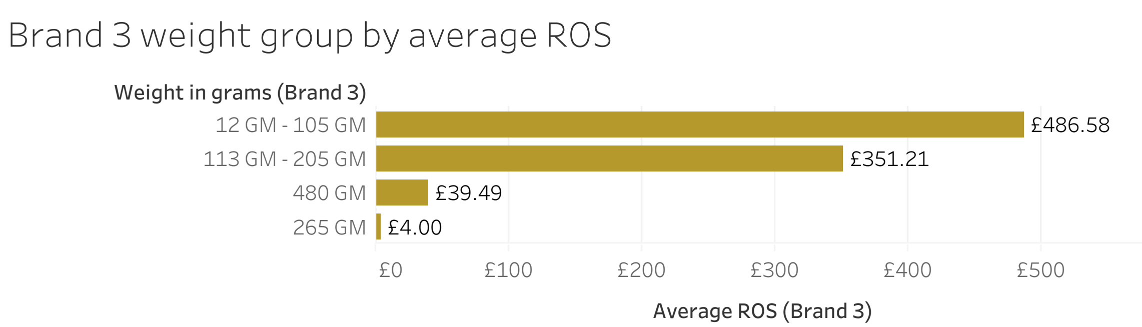

Figure 61. Brand 3 weight group by average ROS (Source: original, conducted by author)

4.5. Premiumization

In order to investigate the signs of premiumization in sales performance of four brands of interests, Spearman's rank-order correlation was carried out to determine the relationship between average unit price with either units sold and total sales value of these brands. Local polynomial regression (LOWESS) line was fitted to better mapping data points and illustrate the relationship in the scatter plots. Table results in STATA can be found in Appendix 3.

4.5.1. Brand 169 (SKU 335 – 918)

For brand 169, there was a moderate positive correlation between average unit price and sales, which was statistically significant at the 99% of confidence, $r_s = 0.6736$, $p < 0.01$. Likewise, the relationship between average unit price and unit sold was also moderate positive, $r_s = 0.3979$, $p < 0.01$. The relationship was non-linear. Figures below illustrated that as the price increased, units sold and sales value decreased with the £5.5 - £7 price range resulted in highest sales performance on records (439,916 units sold; £2,609,234 sales value).

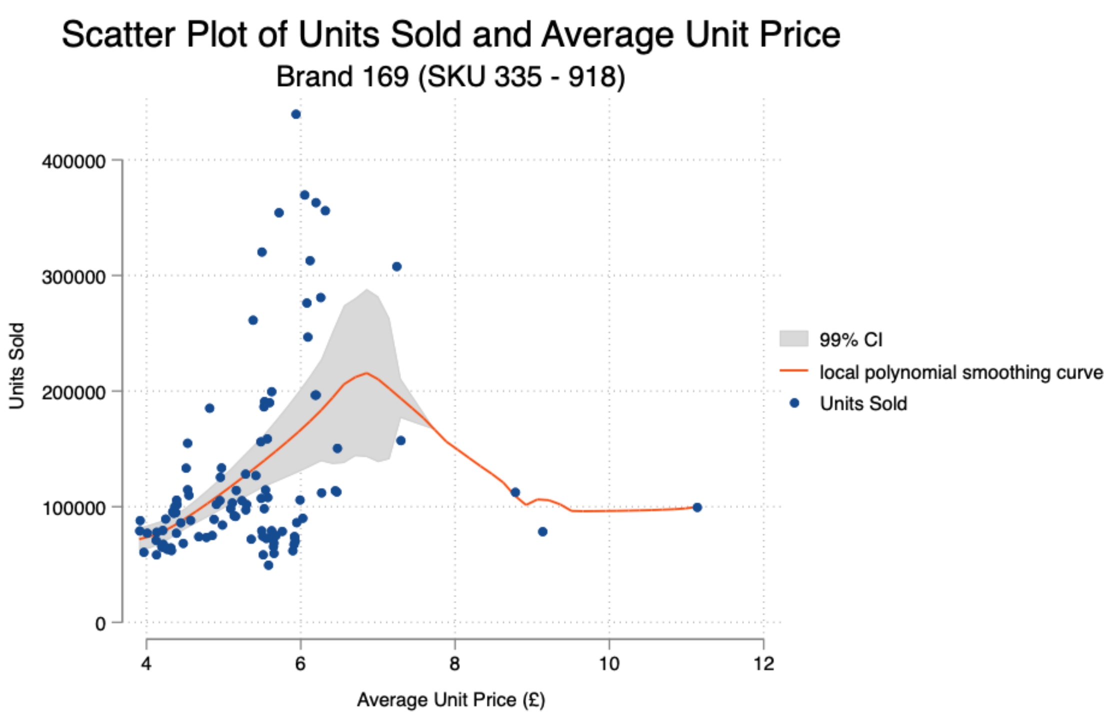

Figure 62. Brand 169 - Scatter Plot of Units Sold and Average Unit Price (Source: original, conducted by author)

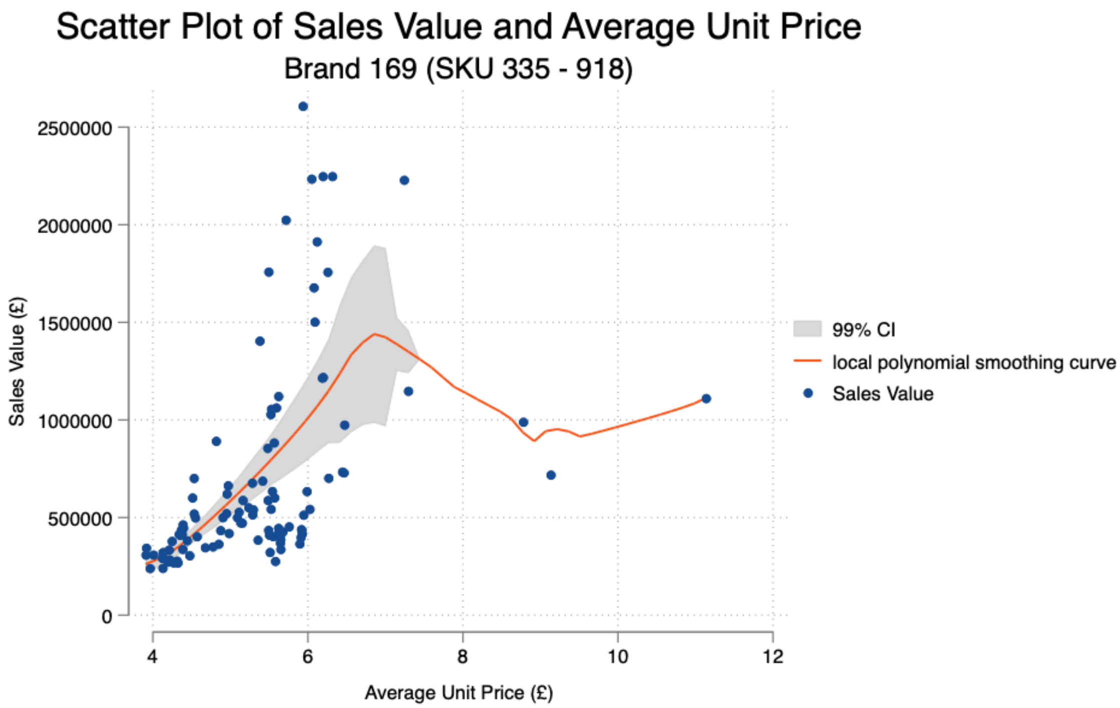

Figure 63. Brand 169 - Scatter Plot of Sales Value and Average Unit Price (Source: original, conducted by author)

4.5.2. Brand 120

Spearman's correlation has been utilized in order to determine the relationship between the number of units sold and the price per unit, and between the sales value and the price per unit. There was a moderate positive relationship between units sold and average price, $r_s = 0.5706$, $p < 0.01$. Likewise, sales value and average unit cost were found to be positively moderately correlated, $r_s = 0.6712$, $p < 0.01$. Figures below illustrate that sales performance and sales value increased as the average unit price rose, with the highest performance (1,086,517 units sold; £3,074,460 sales value) falling into the £2.8 - £3.1 unit price range.

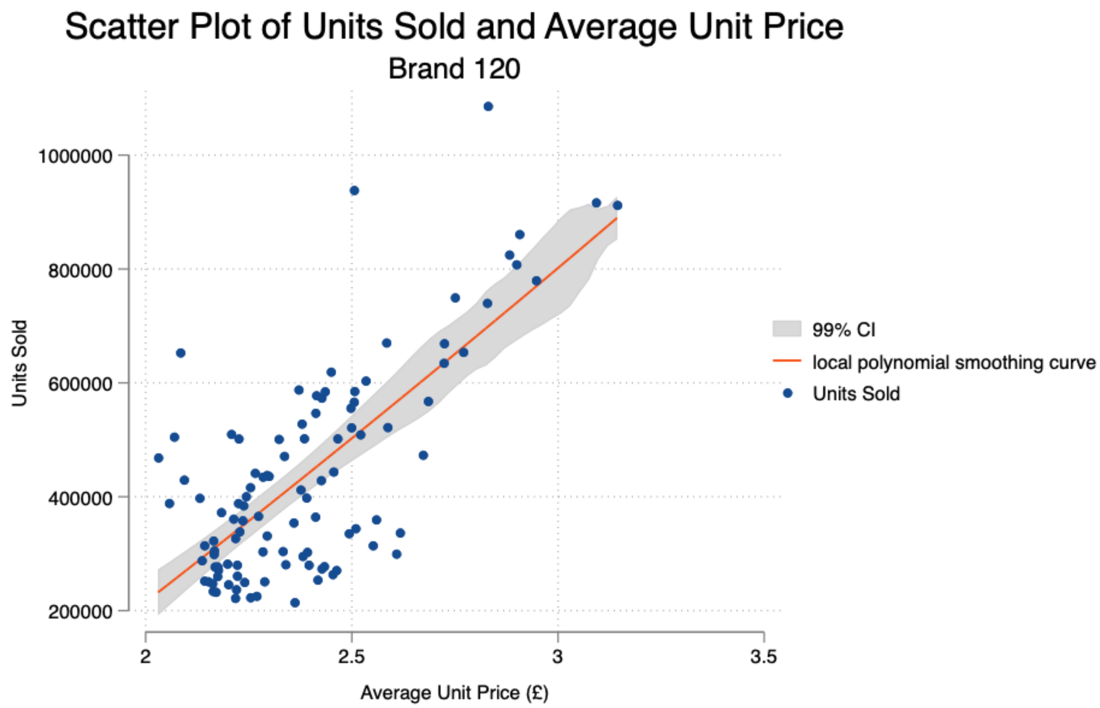

Figure 64. Brand 120 - Scatter Plot of Units Sold and Average Unit Price (Source: original, conducted by author)

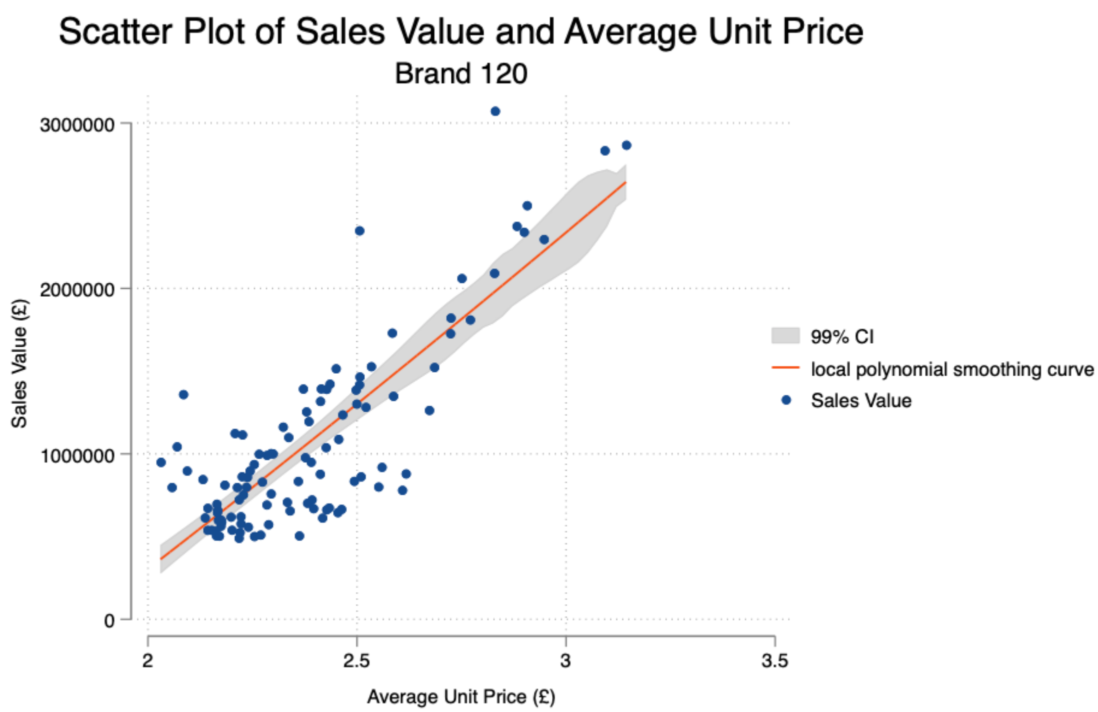

Figure 65. Brand 120 - Scatter Plot of Sales Value and Average Unit Price (Source: original, conducted by author)

4.5.3. Brand 42

Spearman's correlation results showed that there was a moderate negative correlation between unit sold and average unit price, $r_s = -0.4259$, $p < 0.01$. Sales value and average unit price had a weak negative relationship, $r_s = -0.3149$, $p < 0.01$. Local polynomial regression was used to smoothed out the data points in the following scatterplots. It can be seen that both the amount of unit sold and total sales value decreased tremendously as unit price increased, with highest sales performance (74,859 units sold; £235,600 total sales) on records at approximately £3 - £3.7 unit price range.

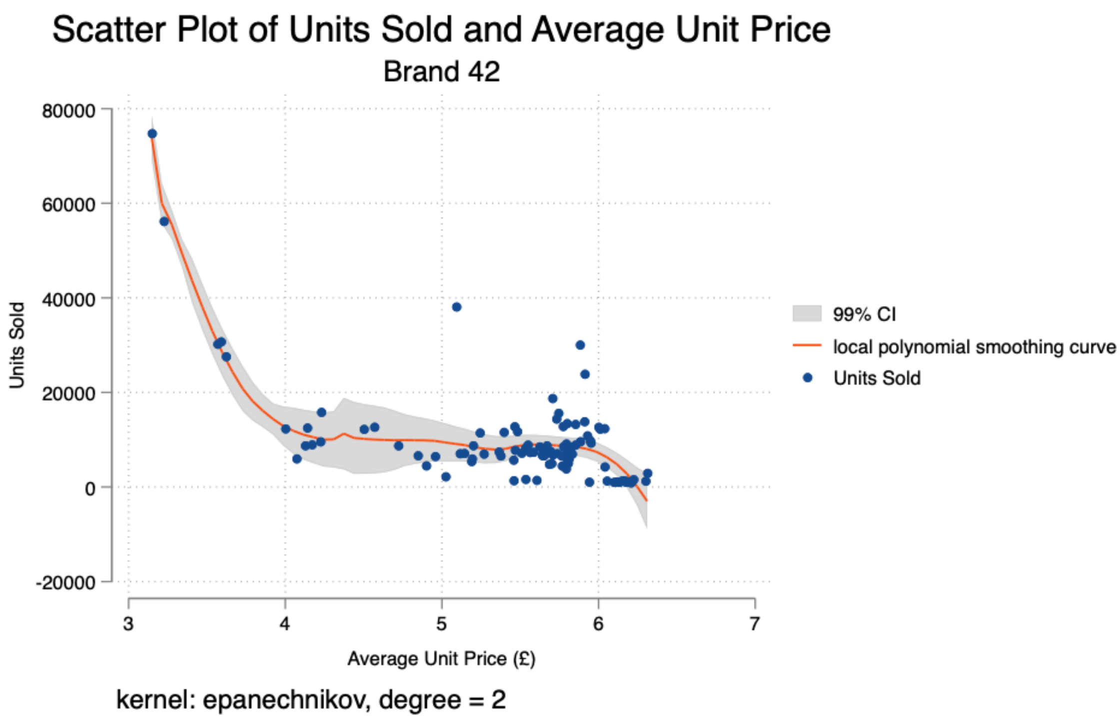

Figure 66. Brand 42 - Scatter Plot of Units Sold and Average Unit Price (Source: original, conducted by author)

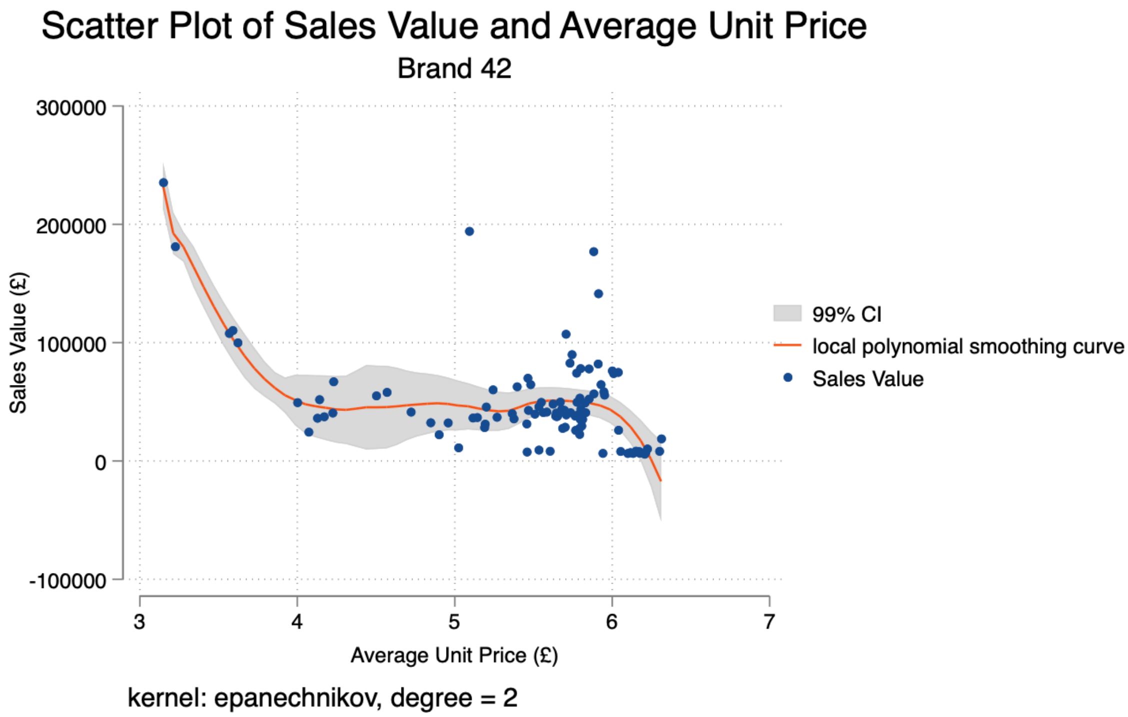

Figure 67. Brand 42 - Scatter Plot of Sales Value and Average Unit Price (Source: original, conducted by author)

4.5.4. Brand 3

For Brand 3, units sold and average unit price were found to have a strong negative relationship, $r_s = -0.0773$, $p = 0.443$ which indicates there was not a statistically significant correlation between these two variables at $\alpha = 0.01$. The correlation between the unit price and sales value was positively weak, $r_s = 0.3099$, $p = 0.0013$. Similarly, a local polynomial smoother was run to fit data points into a curvature. Units sold was recorded at the highest price around £2 - £2.5 and decreased as the cost per unit rose.

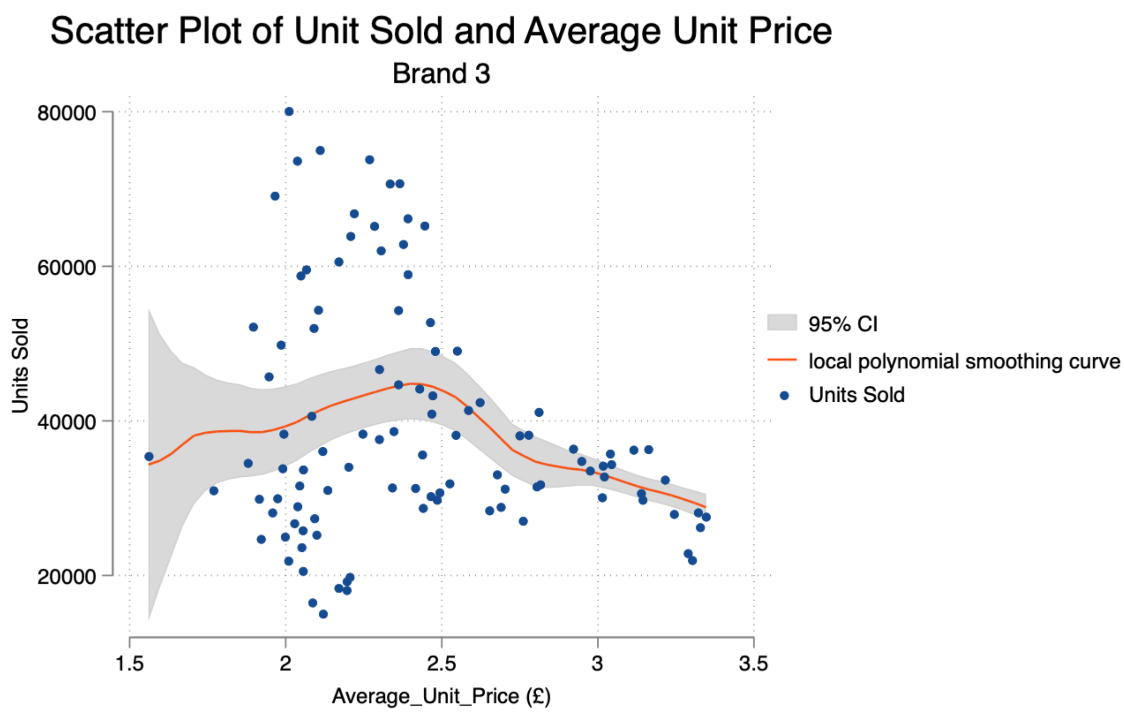

Figure 68. Brand 3 - Scatter Plot of Units Sold and Average Unit Price (Source: original, conducted by author)

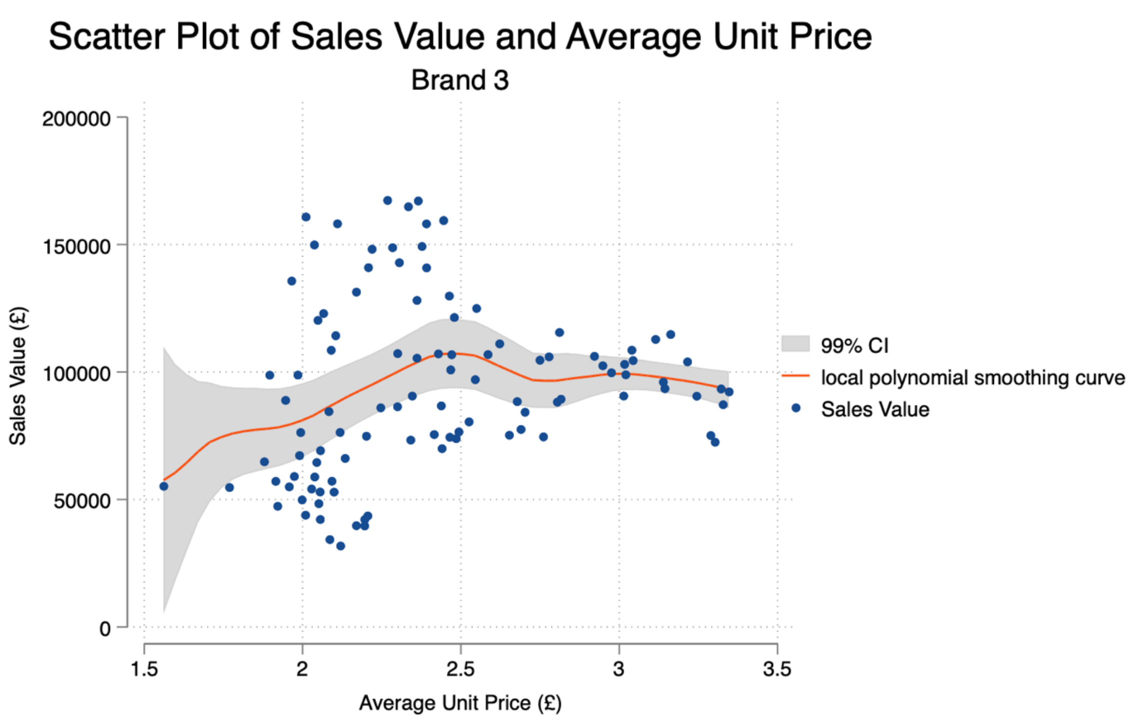

Figure 69. Brand 3 - Scatter Plot of Sales Value and Average Unit Price (Source: original, conducted by author)

To conclude, premiumization signs was the most prevalent in brand 120, dimishing in brand 169 and was not found at all in brand 42 and brand 3. Recalled that brand 120 had the largest product range and was most impacted by the perceived risks of COVID-19, the pricing strategy and range diversification of these brands had benefited its sales during the pandemic, as premium pricing generated higher sales volume.

4.6. E-commerce reviews

Similar analyses to that conducted by Kate Petrova were performed on dataset C, containing scraped data of star ratings and reviews for five major candle brands on the largest UK e-commerce site (2021). For confidentiality, the brands were labelled L, J, S, W, and Y respectively; no inferences would be drawn between these brands and the brands numbered in previous sections.

The star ratings and loss of smell charts – generated from review contents mentioning lack of scent - aimed to determine a uniformed pattern of consumers ranking candles worse during the pandemic than during the normal period due to increased lack of scent, which was a derivative of COVID-19's common symptom lack of smell.

4.6.1. Star ratings

As seen in figure 70, brand L ratings were significantly lowered during the pandemic outbreak compared to the normal period. The ratings increased for a short period from March to September 2020 before taking a deep in October 2020. As of June 2021, brand L's star ratings had not fully recovered to its pre-pandemic highs of five star and hovered at a historic low of three stars.

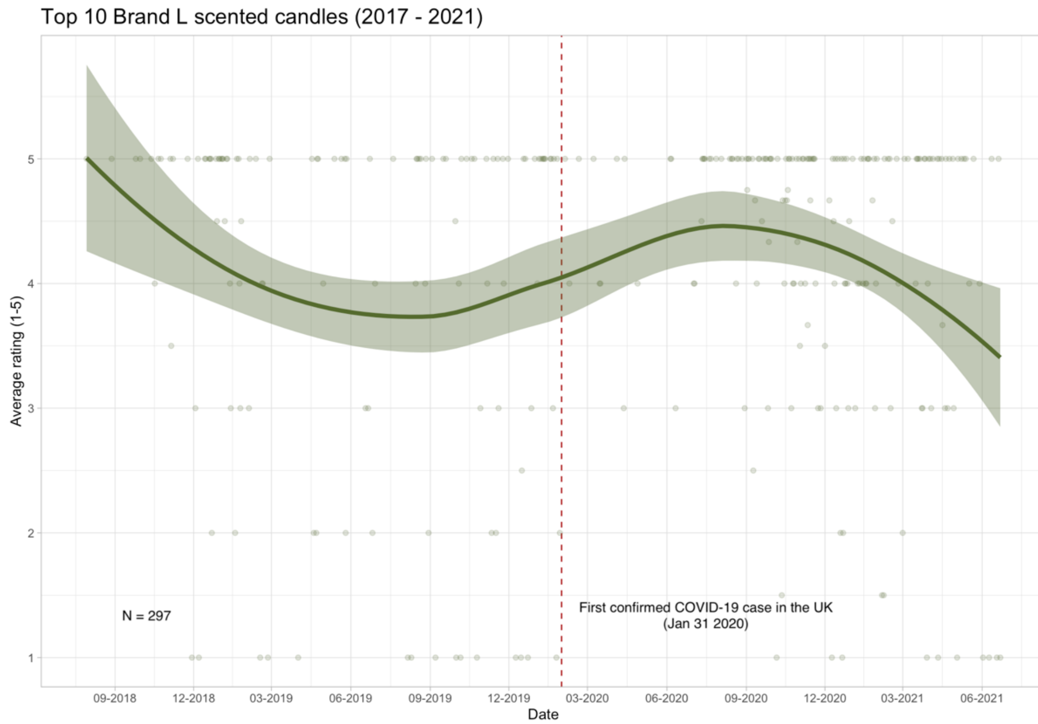

Figure 70. Average star ratings of top 10 brand L scented candles (2017 - 2020)
(Source: Original, conducted by author)

A similar trend was observed for brand J and brand S with sample sizes nearly quadrupled that of brand L. As seen in the figures below, star ratings for both brands have dropped consistently since the start of the pandemic and failed to recover to the pre-pandemic benchmarks. However, it would be difficult based on graphical evidence alone to conclude that COVID-19 had an impact on star ratings of brands J, S, and L. This was because the downward trend was previously observed in brand L and brand J prior to the pandemic outbreak, hence there could be pre-existing factors impacting the trends.

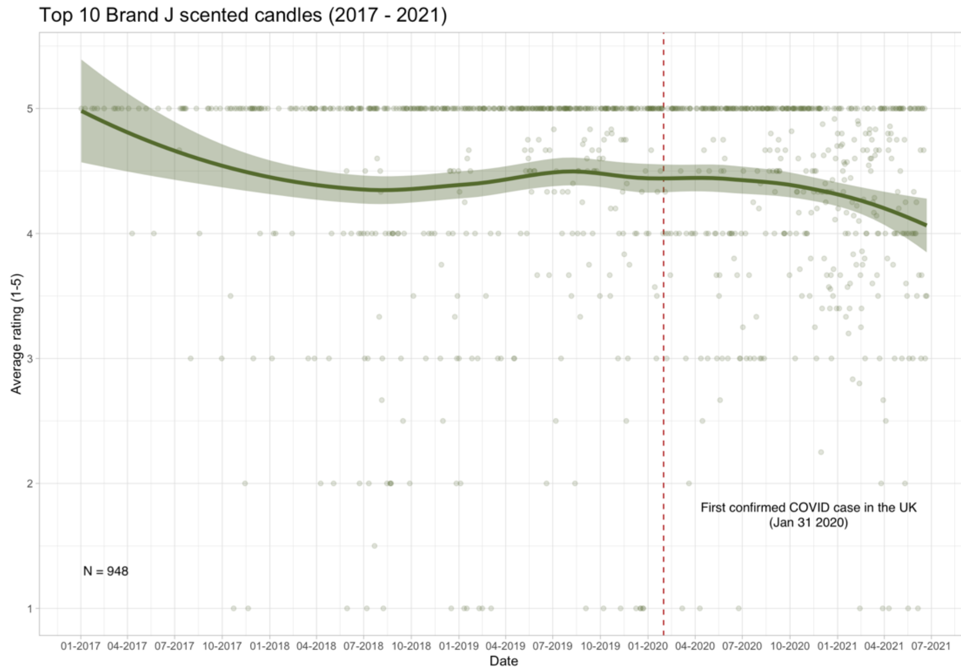

Figure 71. Average star ratings of top 10 brand J scented candles (2017 - 2020)
(Source: Original, conducted by author)

Further investigation should be conducted on brand S and brand Y, however, as they both posited an upward star rating trend before the pandemic that were disrupted more or less at the beginning of the outbreak. Furthermore, the speed of decreasing scores in star ratings for these two brands were significantly faster than that of brand L and J.

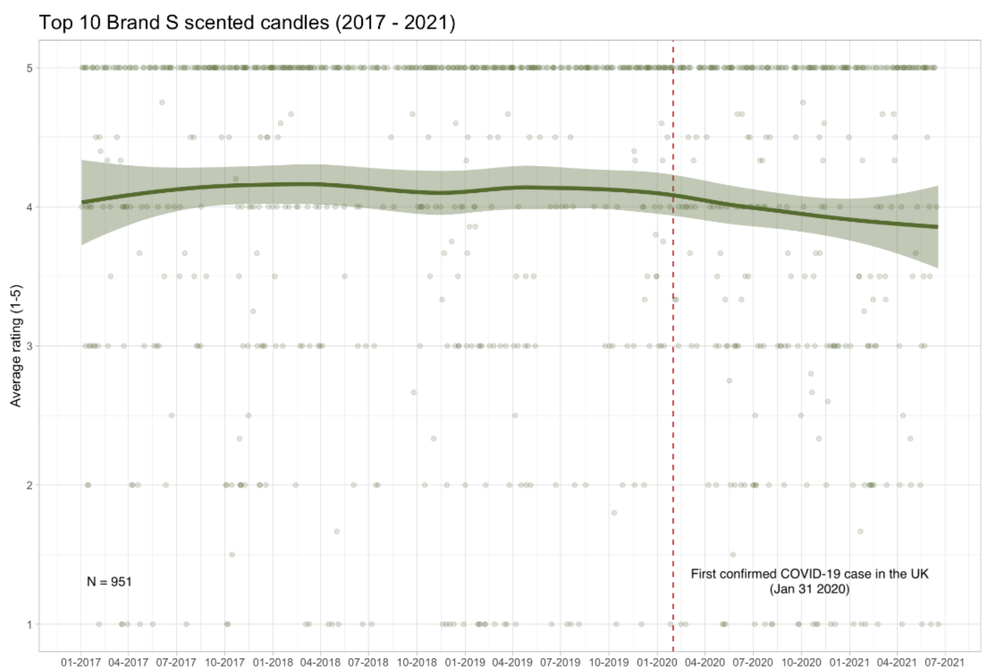

Figure 72. Average star ratings of top 10 brand S scented candles (2017 - 2020)

(Source: Original, conducted by author)

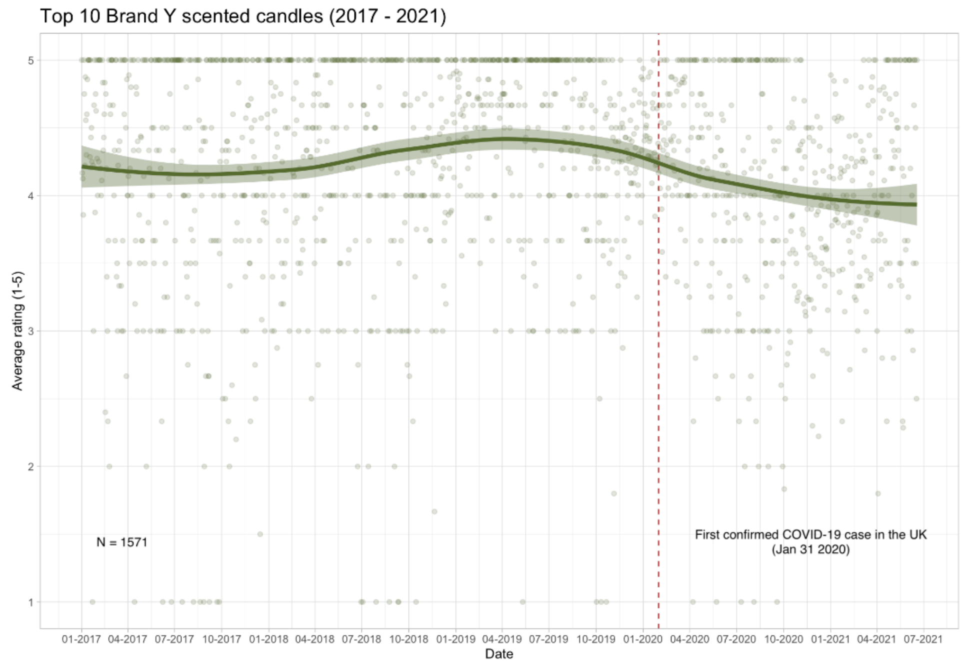

Figure 73. Average star ratings of top 10 brand Y scented candles (2017 - 2020)
(Source: Original, conducted by author)

Last but not least, as seen in figure 74, brand W defied the observed trend as its star ratings since the pandemic outbreak had risen. However, this was also a trend on continuum prior to the pandemic, hence it was not conclusive based on graphical evidence alone if COVID-19 had an impact on the improved star ratings of brand W.

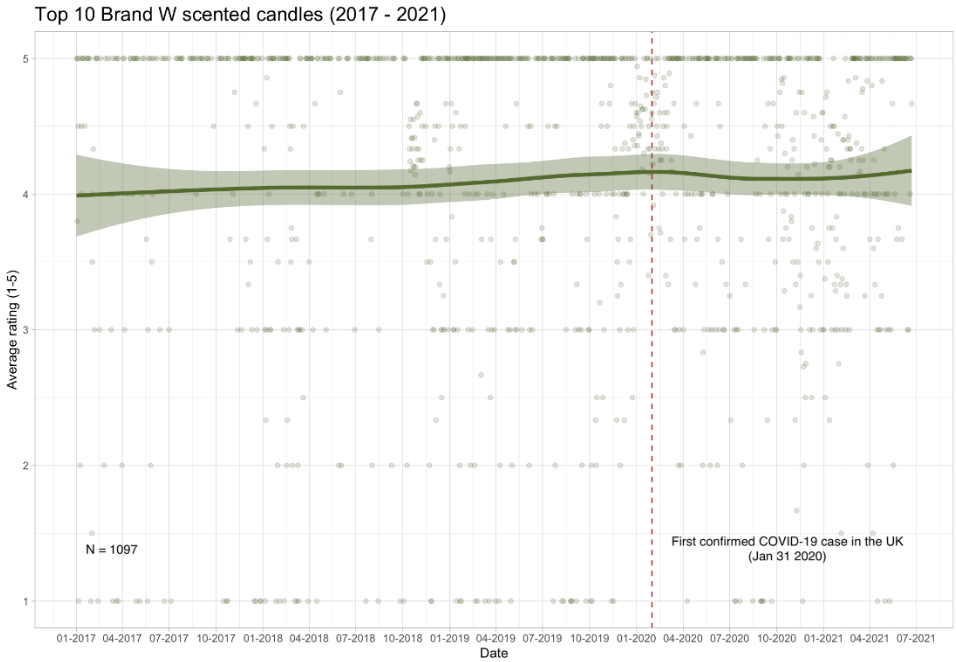

Figure 74. Average star ratings of top 10 brand W scented candles (2017 - 2020)
(Source: Original, conducted by author)

4.6.2. Lack of scent (LOS) mentioned

Similar to Kate Petrova's study, scraped data for mentioning LOS in the written reviews could not serve singularly as evident of COVID-19 impact on candles' online reviews, due to the lack of sufficient benchmark data before the pandemic for contrast. Furthermore, the scraped data analyses below also showed inconsistency with the star ratings. For example, as seen in figure 75, for brand L, the month with highest mentioning of LOS – September 2020 – was also the month with the best star ratings. LOS was not mentioned in any other months of 2020. For brand S, LOS was record in February, before the official outbreak of the pandemic – this was also the month with highest mentioning of LOS. This trend failed to show the progressive LOS exacerbation, should there be an impact of COVID-19.

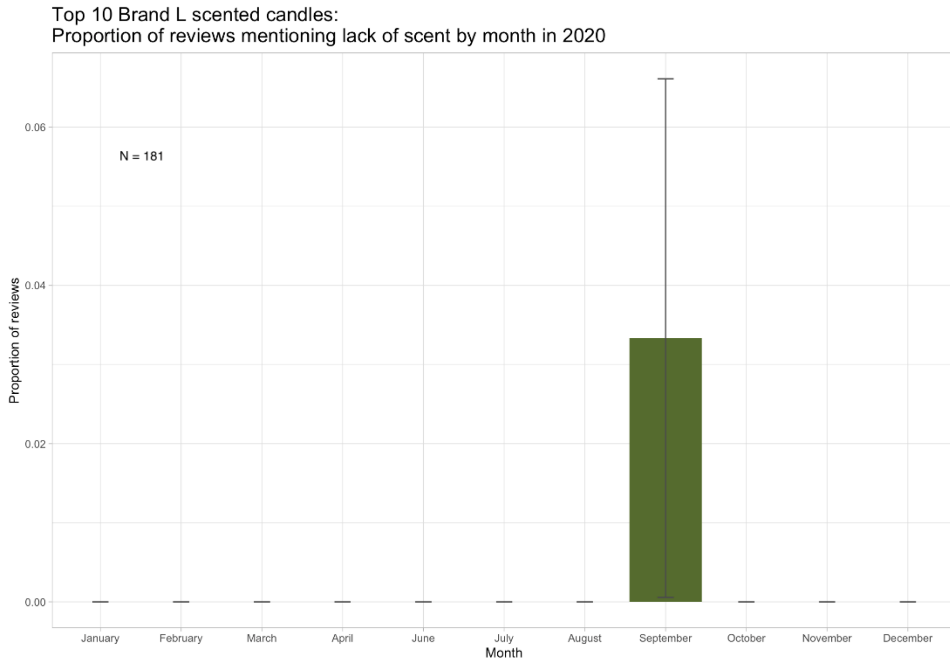

Figure 75*. Top 10 *Brand L scented candles: Proportion of reviews mentioning lack of scent by month in 2020 (Source: original, conducted by author)

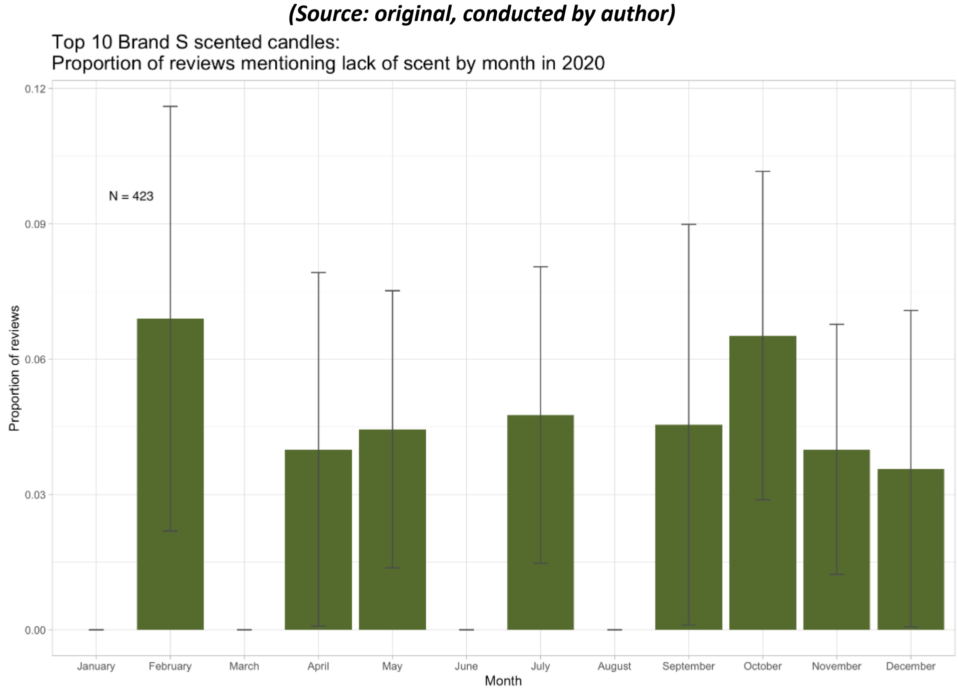

Figure 76. Top 10 Brand S scented candles: Proportion of reviews mentioning lack of scent by month in 2020 (Source: original, conducted by author)

For brand J and brand W, even though LOS was mentioned progressively more since March 2020, decreased presence of LOS did not correlate with higher star ratings. Furthermore, there were intermittent months wherein no LOS was mentioned.

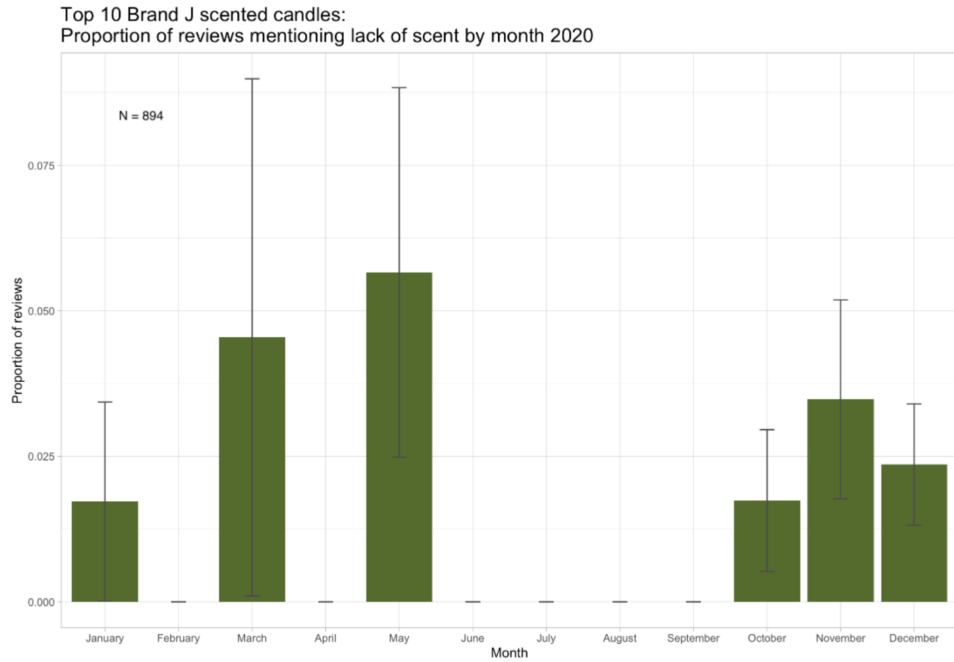

Figure 77. Top 10 Brand J scented candles: Proportion of reviews mentioning lack of scent by month in 2020 (Source: original, conducted by author)

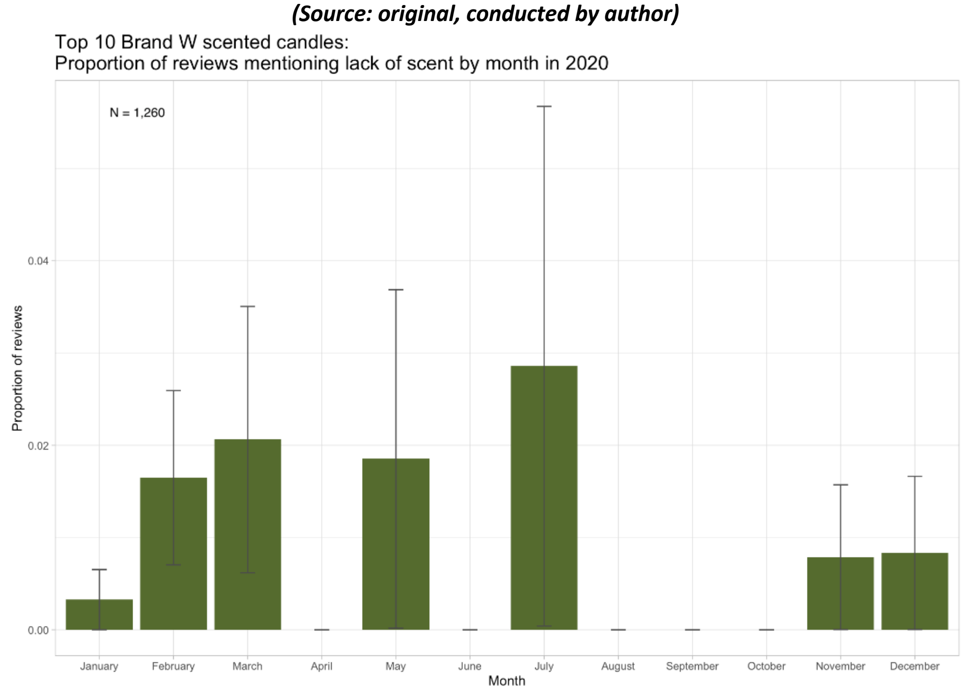

Figure 78. Top 10 Brand W scented candles: Proportion of reviews mentioning lack of scent by month in 2020 (Source: original, conducted by author)

Lastly, brand Y showed the strongest hypothetical link to COVID-19 impacts on reviews, with a large sample size more than double those of others. Since the pandemic outbreak, LOS was mentioned with increased intensity on average and peaked on December 2020, presented in nearly 7% of all reviews for brand Y scented candles. In conjunction with the star ratings evidence, this graph showed that further investigation should be conducted on brand Y reviews to review the actual relationship between increased mentioning of LOS and COVID-19 progress.

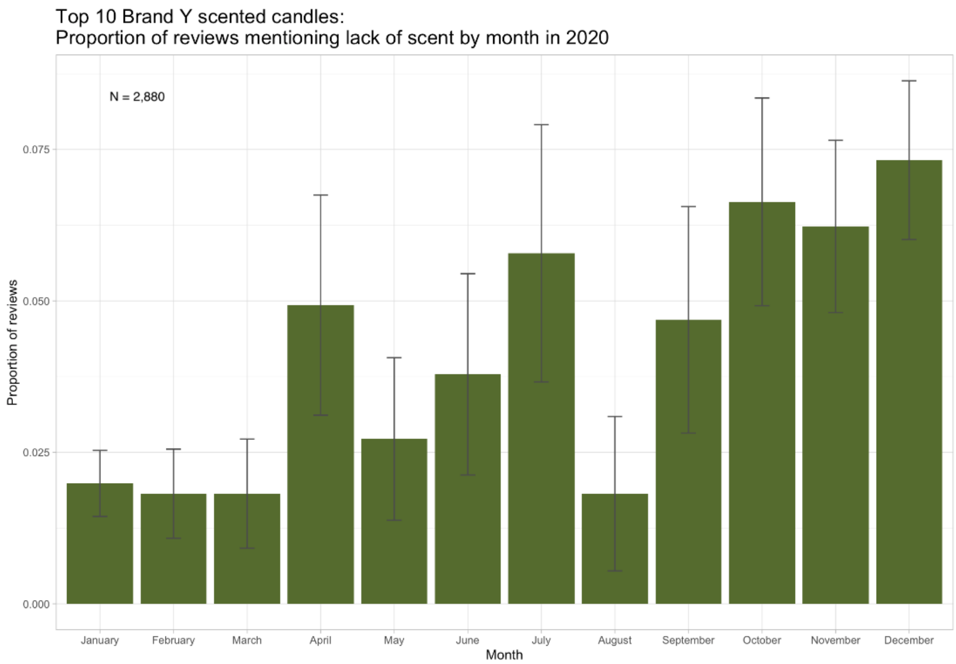

Figure 79. Top 10 Brand Y scented candles: Proportion of reviews mentioning lack of scent by month in 2020 (Source: original, conducted by author)

5. Summary of findings

5.1. Product Size

The findings concluded that there was a clear preference for small to medium product sizes and for the category overall and some brands of interests, which could be due to these sizes tend to be suitable for gift giving and daily use in households. Meanwhile, extremely large sizes were not sold at all, indicating that bulk-buying for money mentality was not observed for this type of product. Brand 120 dominated the heaviest and lightest categories of product weights while the remaining brands of interests concentrated in the 200-500 grams range, indicating different diversification strategy.

5.2. Premiumization

Premiumization was observed for brand 120, as a moderate positive correlation existed between average unit price and either units sold or sales values, which indicated that consumers were willing to purchase at a higher cost. For brand 169, as the unit price increased, there was clear evidence that sales value and units sold only increased to a certain point then started to decrease. This further illustrated the different diversification strategy of these two leading brands: whereas brand 120 focused on product diversification for premium pricing, brand 169 focused more on cost leadership. In contrast, a reverse indication was found for brand 3 and brand 42, as the consumers were reluctant to purchase when the price increased, causing a substantially low sales value for higher unit prices.

5.3. COVID impact

At category and brands of interest levels, the candle industry benefited from the COVID-19 pandemic with stronger sales performance compared to the normal, non-COVID period. Upon analysis of COVID-19 cases, it was determined that total sales in the category and brands of interest increased as the number of infected cases increased, at the very least, up to a point of around 200,000 cases. However, brand 120 thrived the most during the pandemic compared to its competitors. This was hypothetically due to its large product range with premium pricing strategy that not only offered more similar buffer choices to consumers during product shortages caused by lockdowns and supply chain disruptions but also reaped gains from the above-average pricing when consumers stockpiled goods.

5.4. Online retail during COVID-19 pandemic

While 4/5 brands studied suffered from worse star ratings during the pandemic compared to normal, non-COVID period, only 2/5 brands experienced a sudden downward trend that began at the pandemic outbreak. For 1/5 brand, this sudden downward trend could be strongly related to consumers’ loss of smell as a common COVID’s symptoms.

6. Discussion and research conclusion

6.1. Potential drivers of the sales performance trend during COVID-19 pandemic

Concrete evidence of increased sales found in the empirical findings chapter, together with discussed scenarios in the literature review chapter, suggested that during the pandemic, consumers showed a tendency to buy more candles. First, this change in behaviour could be due to the need to escape from normality and “travel” to another place. In lockdowns, when people were not permitted to leave their homes and travel was prohibited, scented products provided ambience and prompted reminiscences of favourite holiday spots. In a sense, this provided an escape from the daily life of quarantine and partly fulfilled the wish of people to travel.

Furthermore, the therapeutic properties of home fragrance products may account for this phenomenon. Due to the intensification of the wellness trend during the pandemic, more and more consumers were motivated to purchase these products in order to create a relaxing and stress-free environment at home. In addition to this, another driver could be the decorating effects of candles which further aid their calming benefits. Also, during the pandemic, remote working became a more common practice, which led to the increased use of candles as home decorations. In the aftermath of the pandemic, flexible working policies have been promoted continuously by employers as the way of the future. This left the door open for home fragrance products to flourish in the near future.

6.2. Pandemic challenges and threats

The most critical challenges revealed through this research is to provide adequate stock at stores. Brands of interests contained SKUs with high average ROS that did not translate to high sales values ranking, indicating that these SKUs were not available for purchase despite ample consumers’ appetite. For brands 42 and 3, because of their small range and dependency on a few SKUs to generate the majority of sales, this could be an unsustainable problem that threatened brand’s compatibility. Despite their surge of popularity during the pandemic, securing supply chain stability in a post-COVID world would be one of the most pressing challenges for candle companies.

At category level, there were overwhelming advantages for dominant market leaders with the majority of market shares captured by strong performing brands. However, there were presence of new and upcoming players in the top best performing brands with strong product differentiation that garnered good average ROS. These new commers could not exploit sales values due to their limited capability to get products to stores. With the growing presence of e-commerce and digital retail, these barriers could be removed at relatively low costs, creating competitive threats for market leader such as Company A in the future.

6.3. Opportunities and industry adaptation

The research has shown that the industry's seasonal trends were clearly forecastable and remained theoretically stable year on year. This indicated that the annual supply crunch due to seasonal factors could be predicted with certainty. Even though the COVID-19 pandemic had showed that to every forecasting model there would be exceptions and outliers, companies needed to be agile in adapting to crises by diversification and operating strategies. These included continuous product differentiation for better pricing and fastening the adoption to digital retail – as the trend of shopping online and death of physical retail were more or less quickened by the pandemic.

7. Recommendations and Research limitations

7.1. Recommendations for category management and operating strategy

As the industry leader, company A is well positioned to continuously improve its diversification and operating strategy. This implied strengthening supply chain capacity to ensure that products would be available to consumers during and after the pandemic. Even though it would be difficult to develop forecasting models for a global force majeure crisis such as COVID-19, it was possible to build effective product portfolio and extensive online and offline retail channel to absorb risks in case of supply chain disruption.

7.2. Research limitations and recommendations for further research

The first limitation to be considered was the data quality. The research was a combination of three datasets, one of which was authentic and contained a significant number of outliers. This resulted in the non-normal distribution of data and, almost always, non-parametric statistical methods (the sign test, Spearman's rank-order correlation) were employed instead of more preferred parametric techniques.

In addition, the other two datasets were obtained from public resources, which raised concerns in data processing accuracy. The dataset B concerned COVID-19 cases were collected from open Governement statistics thus the author had no control over the original data collection process. Furthermore, limitation inherent to the scraped dataset C and their connection with the primary dataset A also prevented the author from addressing the research questions to the extent desired.

Another limitation was that although the data related to COVID-19 cases was adequately used to analyse the impact on the home fragrance category sales, the author did not concern the moderating effects from other potential existing factors. To be more specific, as the collected COVID-19 cases data was for the UK as a whole, the research analysis did not reflect the pandemic influence on candle sales in different countries and regions within the UK. On top of this, the lack of analysis on the impact of the newly rolled-out vaccination program by the NHS was also considered to be a notable limitation.

Finally, as mentioned in the introduction chapter, the shortage of previous studies in the industry also resulted in the lack of benchmarking and overgeneralisation of interpretation.

However, to some extent, it should be noted that this research has suggested some interesting findings and evidences in regards to the performance of home fragrance industry and consumer behaviours during the pandemic. Further research could perhaps evaluate the impact of the pandemic concerning other factors as mentioned above.

REFERENCE

Bloomberg Intelligence (2021, August). *The Covid Resilience Ranking* [Online] Available at: https://www.bloomberg.com/graphics/covid-resilience-ranking/ [Accessed on 25 August 2021]

CBI (2020). *The European market potential for candles* [Online] Available at: https://www.cbi.eu/market-information/home-decoration-home-textiles/candles/market-potential [Accessed on 20 August 2021]

Cherry, L., & Anghel, I. (2021, August 20). *U.K. Retail Sales Fall in Signal Recovery Losing Momentum*. Bloomberg [Online] Available at: https://www.bloomberg.com/news/articles/2021-08-20/u-k-retail-sales-fall-in-a-signal-recovery-is-losing-momentum [Accessed on 15 June 2021]

Chronopoulos, D. K., Lukas, M., & Wilson, J. O. S. (2020). Consumer Spending Responses to the COVID-19 Pandemic: An Assessment of Great Britain. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.3586723

Cleveland, W. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74: pp 829 – 836.

Derudi, M., Gelosa, S., Sliepcevich, A., Cattaneo, A., Cavallo, D., Rota, R., Nano, G. (2013) Emission of air pollutants from burning candles with different composition in indoor environments, *Environmental science and pollution research international, 21*(6) http://doi.org/10.1007/s11356-013-2394-2.

Deshmukh, P.R. (2019). *Home Fragrance Market by Product (Diffusers, Spray, Scented Candles, and Others) and Distribution Channel (Supermarkets/Hypermarkets, Online Stores, Convenience Stores, and Others): Global Opportunity Analysis and Industry Forecast, 2019 – 2026.* [Online] Available at: https://www.alliedmarketresearch.com/home-fragrance-market [Accessed on 17 June 2021]

Dixon, W.J., & Mood, A.M. (1946). The Statistical Sign Test. *Journal of the American Statistical Association,* 41(236), pp. 557 – 566. http://doi.org/10.2307/2280577

Dunsby, S. (2019). *The UK candle economy is worth £1.9bn* [Online] Available at: https://londonlovesbusiness.com/the-uk-candle-economy-is-worth-1-9bn/ [Accessed on 18 June 2021]

Elliott, A.C., & Woodward W.A. (2007). *Statistical analysis quick reference guidebook with SPSS examples.* 1st ed. London: Sage Publications

Evans, J. (2016). *Business Analytics: Methods, Models and Decisions*. Global ed. Harlow: Pearson Education. pp. 686 – 687

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of endocrinology and metabolism,* 10(2), pp.486 – 489. http://doi.org/10.5812/ijem.3505

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. Judgment and Decision Making, 6(1), pp. 100–121.

Grand View Research (n.d.) (2019). *Home Fragrance Market Size, Share & Trends Analysis Report by Product (Sprays, Candles), By Distribution Channel (Hypermarket & Supermarkets, Online), By Region, And Segment Forecasts, 2019 – 2025* [Online] Available at: https://www.grandviewresearch.com/industry-analysis/home-fragrance-market [Accessed on 18 June 2021]

Huang, A., Dawes, J., Lockshin, L., & Greenacre, L. (2017). Consumer response to price changes in higher-priced brands. *Journal of Retailing and Consumer Services*, *39*, pp. 1–10. https://doi.org/10.1016/j.jretconser.2017.06.009

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y. et al. (2020). *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.* The Lancet. 395(10223), pp. 397-506

Hutton, G. and Rhodes, C. (2021). *Briefing paper: Retail sector in the UK.* House of Commons Library. SN06186. [Online] Available from: https://commonslibrary.parliament.uk/research-briefings/sn06186/ [Accessed on 27 June 2021]

International Monetary Fund (2021). World Economy Outlook. [Online]. Available at: https://www.imf.org/en/Publications/WEO/Issues/2021/03/23/world-economic-outlook-april-2021. [Accessed on 27 June 2021]

Ivanov, D. (2021). Supply Chain Viability and the COVID-19 pandemic: a conceptual and formal generalisation of four major adaptation strategies. *International Journal of Production Research*, *59*(12), 3535–3552. https://doi.org/10.1080/00207543.2021.1890852

Jahshan, E. (2020). *UK Grocery enjoys biggest sales jump in 26 years* [Online] Available from: https://www.retailgazette.co.uk/blog/2020/05/uk-grocery-enjoys-biggest-sales-jump-in-26-years/ [Accessed on 30 June 2021]

KANTAR (2020). Record grocery sales in UK as shoppers prepare for lockdown. [Online] Available at: https://www.kantar.com/Inspiration/FMCG/Record-grocery-sales-in-UK-as-shoppers-prepare-for-lockdown [Accessed on 30 June 2021]

Kate Petrova on Twitter. (2020, November 25). Twitter. [Online] Available at: https://twitter.com/kate_ptrv/status/1332398737604431874 [Accessed on 28 August 2021]

Keville, K. and Green, M. (2009). Aromatherapy: A Complete Guide to the Healing Art. Second ed. New York: Crossing Press

Khalid, M. (2021, January). Impact of the coronavirus (COVID-19) pandemic on retail sales in 2020. UK Office of National Statistics (ONS) [Online] Available from: https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/impactofthecoronaviruscovid19pandemiconretailsalesin2020/2021-01-28#retail-sales-around-the-world [Accessed on 18 June 2021]

KPMG. (2020, November). *Responding to consumer trends in the new reality*. https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/11/consumers-new-reality.pdf

Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgement and behavior. *Journal of Consumer Psychology.* 22(3), pp. 332-351.

La Torre, M.A. (2003). Aromatherapy and the use of scents in psychotherapy. *Perspectives in Psychiatric Care. 9*(1).

Lau, C., Fiedler, H., Hutzinger, O., Schwind, K.H., Hosseinpour, J. (1997). Levels of selected organic compounds in materials for candle production and human exposure to candle emissions, *Chemosphere 34*, pp. 1623 – 1630.

Laerd Statistics, n.d. *Descriptive and Inferential Statistics.* [Online] Available at: https://statistics.laerd.com/statistical-guides/descriptive-inferentialstatistics.php [Accessed on 4 September 2021]

Laerd Statistics, n.d. *Spearman's Correlation using STATA* [Online] Available at: https://statistics.laerd.com/stata-tutorials/spearmans-correlation-using-stata.php [Accessed on 4 September 2021]

Laerd Statistics, n.d. *Sign Test using SPSS Statistics* [Online] Available at: https://statistics.laerd.com/spss-tutorials/sign-test-using-spss-statistics.php [Accessed on 4 September 2021]

Lee, S.C., Wang, B. (2006). Characteristics of emissions of air pollutants from mosquito coils and candles burning in a large environmental chamber, *Atmospheric Environment 40*, pp. 2128 – 2138.

Ledolter, Johannes (2013). 'Chapter 4. Local Polynomial Regression: A Nonparametric Regression Approach'*. Data mining and business analytics with R.* New Jersey: John Wiley & Sons, Inc. pp. 55 – 66. Doi: 10.1002/9781118596289.ch4

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med, 382*(13), pp. 1199-1207.

Martin, S. (2014). Velocity: How Well Your Product REALLY Sells [Online] Available at: https://www.cpgdatainsights.com/measure-sales/velocity-how-prod-really-sells/ [Accessed on 25 June 2021]

Mehmetoglu, M., & Jakobsen, T.G. (2017). Applied statistics using Stata: A guide for the social sciences. First ed. London: SAGE, pp. 16

Millard, E., & Laube, J., MD. (2020, December 17). *Scented Candles and COVID-19 | Everyday Health*. EverydayHealth.Com. [Online] Available at: https://www.everydayhealth.com/coronavirus/angry-that-your-scented-candle-has-no-smell-the-reason-could-be-covid/ [Accessed on 4 September 2021]

Mumford, C., Lorono-Leturiondo, M., Ntounis, N., Parker, C., Quin, S., Saga, R.S., Jeelani, A., Glass, V. (2020). Review of high street footfall July 2019 - June 2020. High Street Task Force. [Online] Available at: https://www.highstreetstaskforce.org.uk/media/b5dnkp4z/hstf-footfall-report-2020-for-publication.pdf [Accessed on 30 June 2021]

Nielsen. (2016, December). *Moving on up premium products are in high demand around the world*. https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/global-premiumization-report-december-2016.pdf

Nierobisch, T., Toporowski, W., Dannewald, T., Jahn, S. (2017). Flagship Stores for FMCG National Brands: Do They Improve Brand Cognitions and Create Favorable Consumer Reactions? *Journal of Retailing and Consumer Services,* 34(1), pp. 117–37

Nobbs, K., Moore, C. M., Sheridan, M. (2012). The flagship format within the luxury fashion market. International. *Journal of Retail & Distribution Management*, 40, pp. 920–934.

Novotna, E. (2021). Year-Over-Year (YoY): Complete Guide + Examples [Online] Available at: https://www.financetuts.com/year-over-year-yoy/ [Accessed on 25 June 2021]

NPD (2021, July 28). *What is ? | About , Inc*. NPD Group. https://www.npd.com/about-npd/ [Accessed on 30 August 2021]

O’Connell, M., de Paula, U., & Smith, K. (2021). Preparing for a pandemic: spending dynamics and panic buying during the COVID-19 first wave. *Fiscal Studies*, *42*(2), pp. 249–264. https://doi.org/10.1111/1475-5890.12271

Office for National Statistics (ONS) (2021). Dataset: GDP output approach - low-level aggregate. [Online] Available at: https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/ukgdpolowlevelaggregates [Accessed on 18 June 2021]

Office for National Statistics (ONS) (2021). Internet sales as a percentage of total retail sales (ratio) (%). [Online] Available at: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/timeseries/j4mc/drsi [Accessed on 18 June 2021]

Office for National Statistics (ONS) (2021) Retail sales, Great Britain: December 2020. (Release date: 22 January) [Online] Available at: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/bulletins/retailsales/december2020#annual-fall-in-total-retail-sales [Accessed on 18 June 2021]

Ordabayeva, N., & Chandon, P. (2013). Predicting and Managing Consumers' Package Size Impressions. *Journal of Marketing*, *77*(5), 123–137. https://doi.org/10.1509/jm.12.0228

Organisation for Economic Co-operation and Development (OECD) (2020), *OECD Economic Outlook, Volume 2020 Issue 1*, OECD Publishing, Paris, https://doi.org/10.1787/0d1d1e2e-en.

Organisation for Economic Co-operation and Development (OECD). (2020). OECD GDP growth: Quarterly National Accounts., (March), pp.1–2. [Online]. Available at: https://www.oecd.org/sdd/na/GDP-Growth-Q220.pdf.

Organisation for Economic Co-operation and Development (OECD) (2020), E-commerce in the time of COVID-19, *OECD Policy Responses to Coronavirus (COVID-19)* [Online] Available at: https://www.oecd.org/coronavirus/policy-responses/e-commerce-in-the-time-of-covid-19-3a2b78e8/ [Accessed on 20 June 2021]

Orecchio, S., (2011). Polycyclic aromatic hydrocarbons (PAHs) in indoor emission from decorative candles, *Atmospheric Environment* 45, pp. 1888 – 1895

Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020). Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research*, *116*, 209–213. https://doi.org/10.1016/j.jbusres.2020.05.036

Peat, J., & Barton, B. (2005). *Medical Statistics: A guide to data analysis and critical appraisal.* Blackwell Publishing

Reeves, M., Carlsson-Szlezak, P., Whitaker, K., & Abraham, M. (2020, April). *Sensing and shaping the post-covid era*. Boston Consulting Group. https://image-src.bcg.com/Images/BCG-Sensing-and-Shaping-the-Post-COVID-Era-Apr-2020-rev_tcm9-244426.pdf

Romei, V. (2021, May 21). UK set for 'fast and full' recovery after strong economic data. Financial Times. [Online] Available at: https://www.ft.com/content/a1ee290c-851d-41fb-8205-5615d211cd87 [Accessed on 2 September 2021]

Saunders, M., Lewis, P. & Thornhill, A., (2016). Understanding research philosophy and approaches to theory development. In: T. Hill, ed. *Research Methods for Business Students.* Harlow: Pearson Education Limited, pp. 128-170.

Savills Research. (2020, September). *Impact of Covid-19 on European retail*. [Online] Available at: https://pdf.euro.savills.co.uk/portugal/spotlight-european-retail-sep-20.pdf [Accessed on 2 September 2021]

Sayyida, S., Hartini, S., Gunawan, S., & Husin, S. N. (2021). The Impact of the Covid-19 Pandemic on Retail Consumer Behavior. *Aptisi Transactions on Management (ATM)*, *5*(1), 79–88. https://doi.org/10.33050/atm.v5i1.1497

Schober, P., Boer, C. & Schwarte, L. A., (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia,* 126(5), pp. 1763 – 1768. doi: 10.1213/ANE.0000000000002864

Sheth, J. (2020). 'Impact of Covid-19 on consumer behavior: Will the old habits return or die?' *Journal of Business Research*, 117, pp. 280–283.

Shevlin, M., McBride, O., Murphy, J., Gibson Miller, J., Hartman, T. K., Levita, L., Mason, L., Martinez, A. P., McKay, R., Stocks, T. V. A., Bennett, K. M., Hyland, P., Karatzias, T., & Bentall, R. (2020). Anxiety, Depression, Traumatic Stress, and COVID-19 Related

Anxiety in the UK General Population During the COVID-19 Pandemic. BJPsych Open, 6(125), pp. 1–9. https://doi.org/10.31234/osf.io/hb6nq

Simon, R. (2013). *Total Distribution Points: Master of all distribution measures* [Online] Available at: https://www.cpgdatainsights.com/distribution/total-distribution-points-post/ [Accessed on 25 June 2021]

Spangenberg, E., Crowley, A., Henderson, P. (1996). Improving the Store Environment: Do Olfactory Cues Affect Evaluations and Behaviors? Author(s). *Journal of Marketing*, 60(2), http://doi.org/10.2307/1251931

Steinemann, A. (2016). Fragranced consumer products: exposures and effects from emissions. *Air Quality, Atmosphere & Health*, *9*(8), pp. 861–866. https://doi.org/10.1007/s11869-016-0442-z

Tableau Software LLC, n.d. *What is Tableau?* [Online] Available at: https://www.tableau.com/en-gb/why-tableau/what-is-tableau [Accessed on 4 September 2021]

Tighe, D. (2021) *Percentage change in online purchases due to the coronavirus (COVID-19) pandemic in the United Kingdom from March 2020 to February 2021*. Statista. [Online] Available at: https://www.statista.com/statistics/1230225/changes-in-online-buying-among-uk-consumers-since-covid-19/ [Accessed on 25 June 2021]

The R Foundation, n.d. *What is R*? [Online] Available at: https://www.r-project.org/about.html [Accessed on 4 September 2021]

UK Government (2020) 'Prime Minister's statement on coronavirus (COVID-19): 23 March 2020', (23 March). [Online] Available at: https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020 [Accessed on 17 June 2021]

UK Government (2021) 'Coronavirus (COVID-19) in the United Kingdom' [Online] Available at: https://coronavirus.data.gov.uk/details/cases [Accessed on 17 June 2021]

United States Environmental Protection Agency (USEPA) (2001). Candles and Incense as Potential Sources of Indoor Air Pollution: Market Analysis and Literature Review, USEPA-600/R-01-001.

Warrenburg, S. (2005). Effects of Fragrance on Emotions: Moods and Physiology. *Chemical Senses*, *30*(Supplement 1), i248–i249. https://doi.org/10.1093/chemse/bjh208

Wells, L. (2020, May 1). *Coronavirus leads to rise in home scents sales*. Talking Retail. [Online] Available at: https://www.talkingretail.com/news/industry-news/coronavirus-leads-rise-home-scents-sales-01-05-2020/ [Accessed on 31 August 2021]

WHO (2021). Coronavirus Overview [Online] Available at: https://www.who.int/health-topics/coronavirus#tab=tab_1 [Accessed on 18 June 2021]

Willander, J., & Larsson, M. (2006). Smell your way back to childhood: Autobiographical odour memory. *Psychonomic Bulletin & Review*, 13(2), pp. 240–244.

Wilson, E. J. (2005). The nose knows: new product development at Yankee Candle Company. *Journal of Business Research*, *58*(7), 989–994. https://doi.org/10.1016/j.jbusres.2003.12.008

Woods, Z. (2020, November 10). *UK sales of scented candles soar as Covid restrictions tighten*. The Guardian. [Online] Available at: https://www.theguardian.com/business/2020/nov/10/uk-sales-scented-candles-covid-restrictions-kantar. [Accessed on 20 June 2021]

APPENDICES

Appendix 1. Spearman's correlation results – Relationship between sales value and cumulative weekly COVID-19 cases

Appendix 1.1. Category overall

```
. spearman Value Cum_Cases, stats(rho obs p)

 Number of obs =      105
Spearman's rho =       0.4472

Test of Ho: Value and Cum_Cases are independent
    Prob > |t| =       0.0000
```

Appendix 1.2. Brand 169

```
. spearman Value cum_cases, stats(rho obs p)

 Number of obs =      105
Spearman's rho =       0.4705

Test of Ho: Value and cum_cases are independent
    Prob > |t| =       0.0000
```

Appendix 1.3. Brand 120

```
. spearman Value cum_cases, stats(rho obs p)

 Number of obs =      105
Spearman's rho =       0.4856

Test of Ho: Value and cum_cases are independent
    Prob > |t| =       0.0000
```

Appendix 1.4. Brand 42

```
. spearman Value cum_cases, stats(rho obs p)

 Number of obs =      105
Spearman's rho =       0.7696

Test of Ho: Value and cum_cases are independent
    Prob > |t| =       0.0000
```

Appendix 1.5. Brand 3

```
. spearman Value cum_cases, stats(rho obs p)

 Number of obs =      105
Spearman's rho =       0.4692

Test of Ho: Value and cum_cases are independent
    Prob > |t| =       0.0000
```

Appendix 2. Total sales value and value share in details

Appendix 2.1. Category level – Top 5 brands

Brand	**Total Sales Value (£)**	**Category overall value share (%)**	**Rank**
120	£112,751,517.60	43.44%	1
169	£83,039,774	31.99%	2
58	£26,189,176.70	10.09%	3
3	£9,822,540.20	3.78%	4
168	£5,495,505.60	2.12%	5

Appendix 2.2. Category level – Top 5 SKU

Brand - SKU	Total Sales Value (£)	Category overall value share (%)	Rank
169 – 399	£2,688,421.20	1.04%	1
169 – 411	£2,671,921.2	1.03%	2
3 – 77	£2,671,470.30	1.03%	3
120 – 942	£2,433,709.50	0.94%	4
58 – 27	£2,249,085	0.87%	5
169 – 184	£2,208,741	0.85%	6
58 – 128	£2,148,113.50	0.83%	7
169 – 756	£2,110,273.90	0.81%	8
169 – 769	£2,092,566.10	0.81%	9
3 - 47	£2,031,801.20	0.78%	10

Appendix 2.3. Brand 169 (SKU 335 – 918) – Top 10 SKUs

Brand	SKU	Total sales value (£)	Value share at brand level (%)	Rank
169	458	£2,688,421.20	3.69%	1
	470	£2,671,921.20	3.67%	2
	414	£2,208,741.00	3.03%	3
	556	£2,110,273.90	2.87%	4
	560	£2,092,566.10	2.72%	5
	808	£1,978,621.80	2.64%	6
	812	£1,922,447.00	2.63%	7
	901	£1,414,705.90	1.94%	8
	519	£1,375,873.10	1.88%	9
	907	£1,317,721.80	1.80%	10

Appendix 2.4. Brand 120 – Top 10 SKUs

Brand	SKU	Total sales value (£)	Value share at brand level (%)	Rank
120	1161	£2,433,709.50	2.16%	1
	2037	£1,897,347.20	1.68%	2
	2153	£1,459,855.60	1.29%	3
	1657	£1,434,746.20	1.27%	4
	1742	£1,334,232	1.18%	5
	545	£1,289,537.60	1.14%	6
	501	£1,023,394.50	0.91%	7
	1788	£959,553.20	0.85%	8
	2173	£937,872	0.83%	9
	1790	£883,582.40	0.78%	10

Appendix 2.5. Brand 42 – Top 10 SKUs

Brand	SKU	Total sales value (£)	Value share at brand level (%)	Rank
42	47	£474,543.50	9.33%	1
	61	£419,083.70	8.24%	2
	50	£400,039.70	7.87%	3
	14	£323,482.50	6.36%	4
	68	£269,176.80	5.29%	5
	62	£219,906.30	4.33%	6
	49	£207,130.70	4.07%	7
	12	£202,324.20	3.98%	8
	28	£156,251.60	3.07%	9
	27	£142,807.80	2.81%	10

Appendix 2.6. Brand 3 – Top 10 SKUs

Brand	SKU	Total sales value (£)	Value share at brand level (%)	Rank
3	46	£2,671,470.30	27.20%	1
	24	£2,031,801.20	20.69%	2
	18	£1,052,184.70	10.71%	3
	59	£610,411.70	6.21%	4
	8	£523,096.40	5.33%	5
	62	£503,216	5.12%	6
	58	£416,377.90	4.24%	7
	43	£410,793.80	4.18%	8
	32	£271,470.70	2.76%	9
	60	£256,596.50	2.61%	10

Appendix 3. Spearman's correlation results – Relationship between sales value and average unit price and between unit sold and average unit price

Appendix 3.1. Brand 169 (SKU 335 – 918)

```
. spearman Value Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.6736

Test of Ho: Value and Avg_Unit_Price are independent
    Prob > |t| =       0.0000

. spearman Unit Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.3979

Test of Ho: Unit and Avg_Unit_Price are independent
    Prob > |t| =       0.0000

. spearman Value Cum_Cases, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.4472

Test of Ho: Value and Cum_Cases are independent
    Prob > |t| =       0.0000
```

Appendix 3.2. Brand 120

```
. spearman Value Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.6712

Test of Ho: Value and Avg_Unit_Price are independent
    Prob > |t| =       0.0000

. spearman Unit Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.5706

Test of Ho: Unit and Avg_Unit_Price are independent
    Prob > |t| =       0.0000
```

Appendix 3.3. Brand 42

```
. spearman Value Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =      -0.3149

Test of Ho: Value and Avg_Unit_Price are independent
    Prob > |t| =       0.0011

. spearman Unit Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =      -0.4259

Test of Ho: Unit and Avg_Unit_Price are independent
    Prob > |t| =       0.0000
```

Appendix 3.4. Brand 3

```
.  spearman Value Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =       0.3099

Test of Ho: Value and Avg_Unit_Price are independent
    Prob > |t| =       0.0013

. spearman Unit Avg_Unit_Price, stats(rho obs p)

 Number of obs =     105
Spearman's rho =      -0.0773

Test of Ho: Unit and Avg_Unit_Price are independent
    Prob > |t| =       0.4334
```

Appendix 4. E-commerce reviews and star ratings codes to mirror Kate Petrova's study

Brand W

```
#### NO SCENT FUNCTION ####
no_scent <- function(x){
  case_when(
    str_detect(x, "[Nn]o scent") ~ "1",
    str_detect(x, "[Nn]o smell") ~ "1",
    str_detect(x, "[Dd]oes not smell like") ~ "1",
    str_detect(x, "[Dd]oesn't smell like") ~ "1",
    str_detect(x, "[Cc]an't smell") ~ "1",
    str_detect(x, "[Cc]annot smell") ~ "1",
    str_detect(x, "[Ff]aint smell") ~ "1",
    str_detect(x, "[Ff]aint scent") ~ "1",
    str_detect(x, "[Dd]on't smell") ~ "1",
    str_detect(x, "[Ll]ike nothing") ~ "1",
    TRUE ~ x
  )

}

#### REVIEWS MENTIONING LACK OF SCENT 2020 - 2021 ####
w5 <- Scented_W %>%
  arrange(Date) %>%
  filter(Date >= "2020-01-01", Date <= "2020-12-31") %>%
  mutate(noscent = no_scent(Review)) %>%
  mutate(noscent = ifelse(noscent != 1, 0, 1)) %>%
  mutate(month = reorder(format(Date, '%B'), Date)) %>%
  group_by(month) %>%
  add_tally() %>%
  summarise(n =n, noscent = sum(noscent)) %>%
  mutate(nsprop = noscent/n) %>%
  mutate(se = sqrt((nsprop*(1-nsprop))/n)) %>%
  summarise(n=mean(n), se=mean(se), nsprop=mean(nsprop))

w5r <- ggplot(w5, aes(x=as.factor(month), y = nsprop, group = month))+
  geom_bar(stat = "identity", fill = "darkolivegreen")+
  geom_errorbar(aes(ymin = (nsprop-se), ymax = (nsprop+se)), width=0.2, colour = "gray30")+
  labs(x = "Month", y = "Proportion of reviews", title = "Top 10 Brand W scented candles:
\nProportion of reviews mentioning lack of scent by month in 2020")+
  theme_light()+
  theme(plot.title = element_text(size=16))
w5r
```

```
#### SETUP ####
library(readxl)
library(tidyverse)
library(dplyr)
library(ggplot2)

Scented_W <- read_excel("Scented_W.xlsx")

#### SCENTED CANDLES BRAND W ####
w <- Scented_W %>%
  arrange(Date) %>%
  filter(Date >= "2017-01-01") %>%
  group_by(Date) %>%
  summarise(Rating=mean(Rating))

w1721 <- ggplot(w, aes(x = (as.Date(Date)), y = Rating)) +
  geom_vline(xintercept = as.numeric(as.Date("2020-01-31")), colour = "firebrick", linetype =
"dashed")+
  geom_smooth(method = "loess", size = 1.5, colour = "darkolivegreen", fill = "darkolivegreen") +
  geom_point(alpha = 0.2, colour = "darkolivegreen")+
  labs(x = "Date", y = "Average rating (1-5)", title = "Top 10 Brand W scented candles (2017 -
2021)")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "3 month")
w1721
```

Brand Y

```
#### SETUP ####
library(readxl)
library(tidyverse)
library(dplyr)
library(ggplot2)

Scented_Y <- read_excel("Scented_Y.xlsx")

#### SCENTED CANDLES BRAND Y ####
y <- Scented_Y %>%
  arrange(Date) %>%
  filter(Date >= "2017-01-01") %>%
  group_by(Date) %>%
  summarise(Rating=mean(Rating))

y1721 <- ggplot(y, aes(x = (as.Date(Date)), y = Rating)) +
  geom_vline(xintercept = as.numeric(as.Date("2020-01-31")), colour = "firebrick", linetype =
"dashed")+
  geom_smooth(method = "loess", size = 1.5, colour = "darkolivegreen", fill = "darkolivegreen") +
  geom_point(alpha = 0.2, colour = "darkolivegreen")+
  labs(x = "Date", y = "Average rating (1-5)", title = "Top 10 Brand Y scented candles (2017 -
2021)")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "3 month")
y1721

 #### NO SCENT FUNCTION ####
 no_scent <- function(x){
   case_when(
     str_detect(x, "[Nn]o scent") ~ "1",
     str_detect(x, "[Nn]o smell") ~ "1",
     str_detect(x, "[Dd]oes not smell like") ~ "1",
     str_detect(x, "[Dd]oesn't smell like") ~ "1",
     str_detect(x, "[Cc]an't smell") ~ "1",
     str_detect(x, "[Cc]annot smell") ~ "1",
     str_detect(x, "[Ff]aint smell") ~ "1",
     str_detect(x, "[Ff]aint scent") ~ "1",
     str_detect(x, "[Dd]on't smell") ~ "1",
     str_detect(x, "[Ll]ike nothing") ~ "1",
     TRUE ~ x
   )

 }

 #### REVIEWS MENTIONING LACK OF SCENT 2020 ####
 y5 <- Scented_Y %>%
   arrange(Date) %>%
   filter(Date >= "2020-01-01", Date <= "2020-12-31") %>%
   mutate(noscent = no_scent(Review)) %>%
   mutate(noscent = ifelse(noscent != 1, 0, 1)) %>%
   mutate(month = reorder(format(Date, '%B'), Date)) %>%
   group_by(month) %>%
   add_tally() %>%
   summarise(n =n, noscent = sum(noscent)) %>%
   mutate(nsprop = noscent/n) %>%
   mutate(se = sqrt((nsprop*(1-nsprop))/n)) %>%
   summarise(n=mean(n), se=mean(se), nsprop=mean(nsprop))

 y5r <- ggplot(y5, aes(x=as.factor(month), y = nsprop, group = month))+
   geom_bar(stat = "identity", fill = "darkolivegreen")+
   geom_errorbar(aes(ymin = (nsprop-se), ymax = (nsprop+se)), width=0.2, colour = "gray30")+
   labs(x = "Month", y = "Proportion of reviews", title = "Top 10 Brand Y scented candles:
 \nProportion of reviews mentioning lack of scent by month in 2020")+
   theme_light()+
   theme(plot.title = element_text(size=16))
 y5r
```

Brand L

```
#### SETUP ####
library(readxl)
library(tidyverse)
library(dplyr)
library(ggplot2)

Scented_L <- read_excel("Scented_L.xlsx")

#### SCENTED CANDLES BRAND Y ####
l <- Scented_L %>%
  arrange(Date) %>%
  filter(Date >= "2017-01-01") %>%
  group_by(Date) %>%
  summarise(Rating=mean(Rating))

l1721 <- ggplot(l, aes(x = (as.Date(Date)), y = Rating)) +
  geom_vline(xintercept = as.numeric(as.Date("2020-01-31")), colour = "firebrick", linetype =
"dashed")+
  geom_smooth(method = "loess", size = 1.5, colour = "darkolivegreen", fill = "darkolivegreen") +
  geom_point(alpha = 0.2, colour = "darkolivegreen")+
  labs(x = "Date", y = "Average rating (1-5)", title = "Top 10 Brand L scented candles (2017 -
2021)")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "3 month")
l1721

  #### NO SCENT FUNCTION ####
  no_scent <- function(x){
    case_when(
      str_detect(x, "[Nn]o scent") ~ "1",
      str_detect(x, "[Nn]o smell") ~ "1",
      str_detect(x, "[Dd]oes not smell like") ~ "1",
      str_detect(x, "[Dd]oesn't smell like") ~ "1",
      str_detect(x, "[Cc]an't smell") ~ "1",
      str_detect(x, "[Cc]annot smell") ~ "1",
      str_detect(x, "[Ff]aint smell") ~ "1",
      str_detect(x, "[Ff]aint scent") ~ "1",
      str_detect(x, "[Dd]on't smell") ~ "1",
      str_detect(x, "[Ll]ike nothing") ~ "1",
      TRUE ~ x
    )

  }

  #### REVIEWS MENTIONING LACK OF SCENT 2020 ####
  l5 <- Scented_L %>%
    arrange(Date) %>%
    filter(Date >= "2020-01-01") %>%
    mutate(noscent = no_scent(Review)) %>%
    mutate(noscent = ifelse(noscent != 1, 0, 1)) %>%
    mutate(month = reorder(format(Date, '%B'), Date)) %>%
    group_by(month) %>%
    add_tally() %>%
    summarise(n =n, noscent = sum(noscent)) %>%
    mutate(nsprop = noscent/n) %>%
    mutate(se = sqrt((nsprop*(1-nsprop))/n)) %>%
    summarise(n=mean(n), se=mean(se), nsprop=mean(nsprop))

  l5r <- ggplot(l5, aes(x=as.factor(month), y = nsprop, group = month))+
    geom_bar(stat = "identity", fill = "darkolivegreen")+
    geom_errorbar(aes(ymin = (nsprop-se), ymax = (nsprop+se)), width=0.2, colour = "gray30")+
    labs(x = "Month", y = "Proportion of reviews", title = "Top 10 Brand L scented candles:
  \nProportion of reviews mentioning lack of scent by month 2020")+
    theme_light()+
    theme(plot.title = element_text(size=16))
  l5r
```

Brand J

```
#### SETUP ####
library(readxl)
library(tidyverse)
library(dplyr)
library(ggplot2)

Scented_J <- read_excel("Scented_J.xlsx")

#### SCENTED CANDLES BRAND J ####
j <- Scented_J %>%
  arrange(Date) %>%
  filter(Date >= "2017-01-01") %>%
  group_by(Date) %>%
  summarise(Rating=mean(Rating))

j1721 <- ggplot(j, aes(x = (as.Date(Date)), y = Rating)) +
  geom_vline(xintercept = as.numeric(as.Date("2020-01-31")), colour = "firebrick", linetype =
"dashed")+
  geom_smooth(method = "loess", size = 1.5, colour = "darkolivegreen", fill = "darkolivegreen") +
  geom_point(alpha = 0.2, colour = "darkolivegreen")+
  labs(x = "Date", y = "Average rating (1-5)", title = "Top 10 Brand J scented candles (2017 -
2021)")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "3 month")
j1721

 #### NO SCENT FUNCTION ####
 no_scent <- function(x){
   case_when(
     str_detect(x, "[Nn]o scent") ~ "1",
     str_detect(x, "[Nn]o smell") ~ "1",
     str_detect(x, "[Dd]oes not smell like") ~ "1",
     str_detect(x, "[Dd]oesn't smell like") ~ "1",
     str_detect(x, "[Cc]an't smell") ~ "1",
     str_detect(x, "[Cc]annot smell") ~ "1",
     str_detect(x, "[Ff]aint smell") ~ "1",
     str_detect(x, "[Ff]aint scent") ~ "1",
     str_detect(x, "[Dd]on't smell") ~ "1",
     str_detect(x, "[Ll]ike nothing") ~ "1",
     TRUE ~ x
   )

 }

 #### REVIEWS MENTIONING LACK OF SCENT 2020 ####
 j5 <- Scented_J %>%
   arrange(Date) %>%
   filter(Date >= "2020-01-01", Date <= "2020-12-31") %>%
   mutate(noscent = no_scent(Review)) %>%
   mutate(noscent = ifelse(noscent != 1, 0, 1)) %>%
   mutate(month = reorder(format(Date, '%B'), Date)) %>%
   group_by(month) %>%
   add_tally() %>%
   summarise(n =n, noscent = sum(noscent)) %>%
   mutate(nsprop = noscent/n) %>%
   mutate(se = sqrt((nsprop*(1-nsprop))/n)) %>%
   summarise(n=mean(n), se=mean(se), nsprop=mean(nsprop))

 j5r <- ggplot(j5, aes(x=as.factor(month), y = nsprop, group = month))+
   geom_bar(stat = "identity", fill = "darkolivegreen")+
   geom_errorbar(aes(ymin = (nsprop-se), ymax = (nsprop+se)), width=0.2, colour = "gray30")+
   labs(x = "Month", y = "Proportion of reviews", title = "Top 10 Brand J scented candles:
 \nProportion of reviews mentioning lack of scent by month 2020")+
   theme_light()+
   theme(plot.title = element_text(size=16))
 j5r
```

Brand S

```r
#### SETUP ####
library(readxl)
library(tidyverse)
library(dplyr)
library(ggplot2)

Scented_S <- read_excel("Scented_S.xlsx")

#### SCENTED CANDLES BRAND S ####
s <- Scented_S %>%
  arrange(Date) %>%
  filter(Date >= "2017-01-01") %>%
  group_by(Date) %>%
  summarise(Rating=mean(Rating))

s1721 <- ggplot(s, aes(x = (as.Date(Date)), y = Rating)) +
  geom_vline(xintercept = as.numeric(as.Date("2020-01-31")), colour = "firebrick", linetype =
"dashed")+
  geom_smooth(method = "loess", size = 1.5, colour = "darkolivegreen", fill = "darkolivegreen") +
  geom_point(alpha = 0.2, colour = "darkolivegreen")+
  labs(x = "Date", y = "Average rating (1-5)", title = "Top 10 Brand S scented candles (2017 -
2021)")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "3 month")
s1721

#### NO SCENT FUNCTION ####
no_scent <- function(x){
  case_when(
    str_detect(x, "[Nn]o scent") ~ "1",
    str_detect(x, "[Nn]o smell") ~ "1",
    str_detect(x, "[Dd]oes not smell like") ~ "1",
    str_detect(x, "[Dd]oesn't smell like") ~ "1",
    str_detect(x, "[Cc]an't smell") ~ "1",
    str_detect(x, "[Cc]annot smell") ~ "1",
    str_detect(x, "[Ff]aint smell") ~ "1",
    str_detect(x, "[Ff]aint scent") ~ "1",
    str_detect(x, "[Dd]on't smell") ~ "1",
    str_detect(x, "[Ll]ike nothing") ~ "1",
    TRUE ~ x
  )

}

#### REVIEWS MENTIONING LACK OF SCENT 2020 ####
s5 <- Scented_S %>%
  arrange(Date) %>%
  filter(Date >= "2020-01-01", Date <= "2020-12-31") %>%
  mutate(noscent = no_scent(Review)) %>%
  mutate(noscent = ifelse(noscent != 1, 0, 1)) %>%
  mutate(month = reorder(format(Date, '%B'), Date)) %>%
  group_by(month) %>%
  add_tally() %>%
  summarise(n =n, noscent = sum(noscent)) %>%
  mutate(nsprop = noscent/n) %>%
  mutate(se = sqrt((nsprop*(1-nsprop))/n)) %>%
  summarise(n=mean(n), se=mean(se), nsprop=mean(nsprop))

s5r <- ggplot(s5, aes(x=as.factor(month), y = nsprop, group = month))+
  geom_bar(stat = "identity", fill = "darkolivegreen")+
  geom_errorbar(aes(ymin = (nsprop-se), ymax = (nsprop+se)), width=0.2, colour = "gray30")+
  labs(x = "Month", y = "Proportion of reviews", title = "Top 10 Brand S scented candles:
\nProportion of reviews mentioning lack of scent by month in 2020")+
  theme_light()+
  theme(plot.title = element_text(size=16))
s5r
```