Marketing Analytics Course Work

Part I

Data Analytics Simulation Report: Strategic Decision Making

K12

01962176

01997207

01924400

01910348

01940804

01940804

01885310

Marketing Analytics

Dr. Stjin Maesen

Imperial College Business School

London, United Kingdom

October 15th, 2020

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Blue, a laundry detergent brand part of Kelsey-White's portfolio, struggled to remain relevant in the US market, with a staggering market share of 11% (Harvard Business Publishing, 2016). According to Harvard Business Publishing (2016), K-W has relied on experience and gut as a basis for their decision making, which distanced Blue from constructing data-based decisions and thrive in the highly competitive laundry detergent market. Until 2018, Blue's strategy was extraordinarily generic and did not consider its target segment's behaviour. In fact, its three competitors were outperforming Blue: Turbo (44.1% market share), Fresh (26.3% market share), and Store (18.6% market share). In an attempt to revive Blue, a team of data-driven brand managers took over the brand's strategy and decision making.

In 2019, Blue decided to concentrate on Blue's liquid detergent, which increased its production costs by 7%. Therefore, the price was raised to offset the increase in costs. They targeted households with low to average income, with three people or more, located in the Southeast, Central, and West regions, that were under the age of 44, in the hopes of attracting a young demographic that was sensitive to price due to its big household and low to average income. This increased profitability and revenue but ignited little change in market share.

In 2020, Blue dropped younger consumers and adopted an older segment, targeting people from 35 – 54 years old, mainly because the team thought they were limiting the potential demand by targeting younger people with big households, which was assumed to be a rare demographic. Further, Blue kept their income and household target segments the same and dropped the Northeast due to poor performance. This year, Blue also dropped the price as it realized it was too high for the target's income demographics. This strategy resulted in a slight increase in market share, but profit and revenue decelerate their growth. Blue estimated this was due to the macroeconomic factors that led other brands to lower their price and make Blue's new lower price not enough to increase demand significantly.

In 2021, Blue brought back younger consumers to its target segment, as they accounted for a large portion of the brand's sales despite not being targeted explicitly. The region and income demographics remained the same within Blue's target segment, but household number was increased to limit the double filtering of younger consumers and larger households. Price was dropped further in hopes that this time it would be sufficient to finally increase demand. This strategy resulted in a minimal increase in market share, paired with diminishing revenue and profit. It was clear to Blue that a more drastic price reduction was needed.

In 2022, Blue's segment remained the same except for age, as the brand included all ages in its target segment to increase potential demand, since its primary goal was to increase market share. Blue was hoping for profit and revenue to increase as a domino effect from increased demand. Blue dropped its price to a record low, and with this strategy, the brand was finally able to significantly increase market share to 17.9%, profit, and revenue.

These four years proved the importance of data-driven decision making for Blue's strategy in the future. Therefore, it is recommended that Blue keeps relying on data using the current information being collected, as well as gathering other types of data and diversifying its analysis. Collecting psychographic data, analysing the return on marketing investment and their elasticities, and introducing in-depth social listening and complementing it with focus groups, are some of the key activities that Blue should focus on in the future.

DECISION 1 – 2019

Segmentation Analysis

Blue conducted a thorough analysis of the whole market to assess its current position and identify new opportunities. Firstly, ethnicities did not seem to impact the detergent market's, revenue, profitability, or market share other than reducing the total demand if any were selected. Therefore, to maximize demand, Blue did not focus on ethnicities.

On another note, the Northeast region is dominated by Turbo, which has 82.3% of the market share, and it is also a strong leader in the Western region. However, Southeast and Central lack a dominant market leader and are more susceptible to being targeted, as seen in **Figure 1**. Moreover, Turbo is also the clear leader for the high-income demographics ($60,000 and over) and is very strong amongst middle earners ($40,000 – $59,999). Still, there is a fiercer competition for lower-middle earners with no clear leader dominating the space **(Figure 2).** For this reason, Blue will be targeting the low-middle income demographics in its strategy. Fresh, which seems to have an erroneous strategy relative to its target segment (Harvard Business Publishing, 2016), leaves room for Blue to accurately target the individuals in the low-middle income segment that Fresh serves and potentially capture their demand.

Moreover, households of 3 and more people are assumed to be more price-sensitive due to the increasing amount of laundry they must take care of compared to those who live alone or with one other person. This can also be supported by the fact that Turbo is a clear and robust leader for smaller households, regarding brand sales and brand demand. Still, a stronger competition is seen as household size increases, and cheaper brands like Fresh and Store get closer to Turbo **(Figure 3)**. Therefore, Blue will target larger households (of three or more people). Finally, the brand will be targeting people aged 44 years and under. This is because Blue has been strongly growing in this segment between 2015 – 2018, as seen in **Figure 4**. This segment is responsible for 70% of sales in 2018. Therefore, by directly targeting them in 2019, Blue can be better positioned to capitalize on their demand.

Segmentation for 2019

Income	Ethnicity	Household Size	Region	Age
Under $20,000 $20,000 – $39,999 $40,000 – $59,999	No ethnicity focus	3 4 5+	Southeast Central West	Under 35 35 - 44

Marketing Mix Analysis

Product

Middle-income earners and the under 44 years segments cannot afford pods but also prefer a modern formula rather than old-fashioned powder (Harvard Business Publishing, 2016). For that reason, liquid is the best suitable formulation. Liquid is the second-highest demand formula at 162 million, therefore, providing a large potential market. Furthermore, odor elimination is the selected brand attribute as it's the preferred formula based on Blue's target segment.

Price

Blue decided to increase the price by 0.99 cents to offset the 7% increase in variable costs due to changing to liquid. Also, by producing a premium formula, a higher price is justified.

Place

The budget for trade channel demand will be mostly distributed proportionally to the units' demand at each location. However, 10% of the budget is deducted from Grocery and transferred to Club and Convenience. This is because Convenience is the fastest-growing channel, and Turbo is not targeting club.

Promotion

The media budget is spent proportionally based on the target audience preferences. Due to the younger target audience, digital ads were prioritized. A 10% increase is added to the digital budget because of the growing consumption rate per year.

Results Analysis

As a result of these changes, Blue experienced considerable revenue and profit growth. Revenue grew by 17.6% to $265.0 million, accompanied by profit growth of 74.3% to $65.9 million. However, Blue could not significantly expand the market share, which increased by 0.3% to 11.3%. Given that the units sold only increased by 3.05%, the profit increase resulted from a higher price rather than gaining demand.

As seen in **Figure 1,** Blue's market share grew from 13.8% to 17.1% within its target segment. This shows that Blue's strategy was fruitful and satisfied its target segment more accurately than previous years. Although profitability and revenue grew for Blue, sales did not heavily increase. Therefore, to ensure long-term market share and unit sales growth, Blue needs to increase demand by enlarging its target segment or lowering its price.

DECISION 2 - 2020

Segmentation Analysis

Based on the results from 2019, Blue altered some aspects of its segmentation and marketing mix. Blue's target income segments will remain unchanged due to them accounting for 89% of Blue's total sales. Moreover, households of 3 people will continue to be the target for this segment because of the assumed price sensitivity of having to buy laundry detergent for a bigger sized household. However, to capture more demand, Blue wants to expand its serviceable addressable market. Having large households paired with a younger segment seems too restrictive as Blue assumes that most households with three or more individuals are older than 35 years old. Therefore, Blue will shift to targeting 35 – 54-year-old consumers. Finally, as another tactic to increase demand, Blue will include the Northeast in its target segment. Although Turbo is potent in the Northeast, as seen in **Figure 1**, Blue has a different positioning that will allow it to stand out regardless.

Segmentation for 2020

Income	Ethnicity	Household Size	Region	Age
Under $20,000 $20,000 – $39,999 $40,000 – $59,999	No ethnicity focus	3 4 5+	Northeast Southeast Central West	35 – 44 45 – 54

Marketing Mix Analysis

Product

Blue maintained its liquid formulation to reduce price and still be profitable.

Price
Despite the revenue and profit growth in 2019, Blue's price will decrease to $7 to remain competitive regarding its target's lower income and price sensitiveness. Selecting $7 Blue can capture two segments of the "price point demand": those that would buy at $5-7 and those at $7-9.

Place
The budget for trade channel demand will be distributed proportionally to the demand in units at each location, prioritizing Convenience and Club.

Promotion
Furthermore, Blue's media channel strategy is proportionally distributed based on the target audience preferences. This year it will be slightly adjusted to align with the older customer segment: TV was prioritized above digital to account for the change in target age.

Results Analysis

2020 was the start of a new market tendency characterized by a decrease in profitability (-$183.3 million in total, which is -40.1% compared to 2019) for all the market players. However, this reduction was not homogeneous, as some companies were way more affected than others. The market can be split into two categories. The first being brands which suffered from a massive shrink in profitability: Fresh (-$92.2 million) and Turbo (-$76.2 million). The second being Store (- $4.1 million) and Blue (- $10.7 million), which have lost profit compared to the previous year, but the difference was much lower. Moreover, Fresh was the only brand that increased its price (+$0.5) the others reduced their price. This move made Fresh lose 7.6% market share, which has been distributed across all the other players. This means that in 2020, price was a key driver of sales.

Moreover, in 2020 profitability decreased by 16.3% ($10.8 million). This happened despite the quantity demanded increasing by 16%. The key driver for the decrease in profitability was the price reduction. As seen in **Figure 5,** this resulted in a 4% decrease in net profit margin**.** This strategy aimed to increase unit sales by lowering the price, positively contributing to Blue's profit, revenue, and market share. However, because other competitors dropped their prices too, fewer customers switched to Blue. This means Blue generated fewer incremental sales from the price decrease than was necessary to increase profitability. Given that the largest contributor to the decline in profitability was price, it is essential to decide on an appropriate pricing strategy with which profit and market share can be maximized. Blue must be careful not to engage in a price war, which will negatively impact category value and all market players.

In 2020, Blue increased its market share by 1.9% to 13.2% and revenue by 3.5% to $274 million. Although Blue did not target people under 35 years old this year, this segment still accounted for the highest sales compared to the other age segments, at 17.8 million units precisely, 45.4% of Blue's total sales. This is assumed to be due to both price sensitivity of people under 35 years old and Blue's heavy investment in digital ads for the year.

On the contrary, 45-55-year olds did not seem to be attracted by Blue's offerings, as for these consumers within Blue's target segment, the brand has the lowest market share at 10.4% and low demand of 3.7 million units, compared to Store, at 12.3 million units (Store seems to be Blue's biggest competitor in its segment). Moreover, Blue's entry into the Northeast was unsuccessful. The region generated only 3.65% (1.4 million units) of the overall sales in 2020, which might be because the Northeast had a clear preference for softness and cold water.

However, this did not correspond with Blue's brand attribute, which focuses on odor elimination.

DECISION 3 - 2021

Segmentation Analysis

As seen in the results for 2020, Blue's strategy was not appealing enough to 45-54-year-old people. Therefore, Blue decided to scratch them off their target segment and add back the under 35 segment, which did show interest in Blue's product. This can be seen in the 19% increase in sales from 2019 to 2020, as seen in **Figure 4**. Moreover, Blue decided to include all households in its target segment due to the assumption that people from 35 - 44 years of age are less likely to have households of three or more people. Targeting that age, in addition to big households, would be detrimental to the serviceable available market. As seen in the previous year's results, the Northeast was not fruitful for Blue. For this reason, the brand decided to exclude it from the target segment. Finally, income remained the same, and Blue's strategy remained aligned to targeting middle to low-income households. Regardless, Blue's main concern for the year ahead will be to optimize price to gain more demand.

Segmentation for 2021

Income	Ethnicity	Household Size	Region	Age
Under $20,000 $20,000 – $39,999 $40,000 – $59,999	No ethnicity focus	1 2 3 4 5+	Southeast Central West	Under 35 35 - 44

Marketing Mix Analysis

Product

Blue maintained its liquid formulation to keep variable costs at bay to reduce the price further and remain competitive on its target segment.

Price

Competitors decreased their price heavily in the previous year leading to Blue not fully benefiting from its price decrease. Taking into account Blue's decrease in profit margin in 2020, the brand will decrease the price further this year. Blue's detergent will be found at stores for $6.50.

Place

According to Blue's target segment, distribution of trade channel investment was made based on the unit demand at each channel, which resulted in an almost equitable distribution between all channels.

Promotion

As a result of targeting younger aged segments, the promotional budget will be increased for digital campaigns and decreased for print campaigns. The rest will remain proportional to the media consumed for Blue's segment.

Results Analysis

In 2021 the total market demand increased by 0.83% (as in previous years), and all companies, except for Fresh (+$0.5) reduced their price, as seen in **Figure 7**. As the total market revenue decreased by 6.93%, all the players, except store brands, suffered from a reduction in revenue and a sharp decline in profits (-56.3%). Store brands performed successfully in 2021 as they earned 5% of market share. Their demand increased by 24.4%, their revenue rose by 6.59%, and their profitability was the least affected among all the players with a 20.1% decrease. This further supports Blue's previous price-reducing tactics, as the market seems to respond positively to price reductions.

Although Blue's performance can be perceived as weak if it is observed individually, Blue is not underperforming in the market, as its financial performance is declining at a slower pace than its competitors. Even if its brand demand and market share increased respectively by 3.42% and 0.3%, Blue's $0.5 reduction of its products selling price led to a 4% decrease in revenue and 28.1% diminution in profitability. Considering the macroeconomic environment, this would mean that the strategy of lowering the price was relevant but not sufficiently aggressive.

Moreover, Blue became the leader in its target segment by owning 29.6% of market share. Blue also managed to attract nearly 10 million customers aged 45 and over within the same region and income categories, even though these people were not explicitly targeted. Therefore, if added to the current target segment, the 45 and older demographics would drive up the market scope in terms of households by 150.6%, as seen in **Figure 7.** This would mean that Blue's potential growth can be significantly higher, expanding its serviceable addressable market and potentially increase market share.

DECISION 4 - 2022

Segmentation Analysis

The most pressing issue for Blue now is to gain more market share. Judging from the previous year's results, Blue is performing well in its current target segment but has not improved its market share. Therefore, the brand decided to expand its possible target consumers to include the 45-55 and 55 and over demographics. It increases the brand's demand prospective, which can be seen in households' increase in **Figure 7**. Furthermore, Blue decided to focus on three people and over households because they seem to be the most price-driven, based on the rise in Store demand for these demographics (Store being the cheapest brand), as seen in **Figure 3**. The change in target demographics is significant because the addition of those two age demographics will push a difference in how trade channels and media channels are targeted, and the focus on large households will impact Blue's pricing strategy. In short, Blue's final target segments are big households with middle to low income that live in the Southeast, Central, and West regions.

Segmentation for 2022

Income	Ethnicity	Household Size	Region	Age
Under $20,000 $20,000 – $39,999 $40,000 – $59,999	No ethnicity focus	3 4 5+	Southeast Central West	Under 35 35 – 44 45 – 54

				55 and Over

Marketing Mix Analysis

Product

As Blue incorporates older demographics, it is most logical to offer liquid. It will not put off the classic older consumers that represent a significant portion of the new demand (Harvard Business Publishing, 2016). Moreover, as Blue lowers its price, switching to pods is not convenient due to its higher production costs. Odor elimination remains the brand attribute to match the segment's preference.

Price

Blue's profit is significantly high, regardless of its low revenue. This indicates that the brand currently has very high margins and can dramatically decrease its price to seize more demand. Moreover, focusing on Store as Blue's main competitor, the price needs to be lowered to capture some of their market share, as their price for 2021 was $4.50, 30.8% lower than Blue's. Therefore, the firm's price was lowered to $5.75.

Place

Trade channel shifted slightly to focus on Convenience and Club, without a vast difference between the other channels. This change accounts for the shift in the data when the older age segments were added to the target. Similarly, to past years, spending was decided proportionally to the segment's behaviour.

Promotion

Accounting for the newly added older demographics, digital becomes less of a priority, and print surfaces as a key promotional channel to attract Blue's new target segment. Therefore, Blue's strategy shifts from investing heavily in digital to a more equitable approach amongst all media channels, reflecting the newly chosen target segment's behaviour.

Results Analysis

Satisfactorily, the strategy of Blue made significant progress in 2022. Market share increased by 4.4% to 17.9% and placed third in the industry. Revenue increased by 17.7% to $310 million, and profitability increased by 46.1% to $57.9 million.

Blue finally recognized the importance of older segments in the market and decided to include them in their segmentation. The combination of 55 years old and over and the price decrease allowed Blue to move up the market share ladder. Additionally, the new price allowed the brand to capture more demand, leading to an excellent profitability level.

Regretfully, Blue was not able to impose itself among its competitors in the revenue race. Even though the brand successfully turned around its market share performance, there was not enough demand to increase its revenue relative to the market. Comparing unit sales and revenue, one can observe how Blue is not at the bottom of the rank regarding demand, in contrast to revenue, where they are notably at the end of the list. This can hint to room for further price reductions that could have been undertaken. This would have likely increased demand and, in turn, revenue.

RECOMMENDATIONS

For Blue to continue using data to structure strategy and make decisions, three main recommendations should be kept in mind. Firstly, more in-depth data about target consumers needs to be collected to allow Blue to have a better understanding of their audience. The current customer data in the simulation mainly focuses on demographic, geographical, and behavioural segmentation. However, this data is not enough to create strong market segments for Blue to target. Therefore, it is beneficial to collect additional data on psychographic segmentation (Maesen, 2020). According to Grigsby (2018), a robust psychographic segmentation includes data on consumer values, tastes, personality, and lifestyle. These factors are essential to include in the segmentation process because individuals in the same demographic group can have varying psychographic profiles (Lin, 2002).

One way in which psychographic data can be collected is through surveys (Lin, 2002). Blue can create a laundry detergent consumers database and distribute surveys to gather information about the values, lifestyle, and which product characteristics are most important to them. The results of this data will allow Blue to create robust and more accurate target segments (Maesen, 2020). A second source for psychographic data collection is customers' social media activities (Maesen, 2020). Blue already has an existing social media presence, and more data can be collected on the customers that are engaging with the brand through these social channels. This data would include the customers' lifestyle and values as it would provide insight into their daily lives and thoughts (Maesen, 2020).

The existing segmentation data can be combined with the newly collected psychographic data to create more accurate and actionable segments. After psychographic data is collected, interpreted, cleaned, and coded, the appropriate segmentation can be carried out through hierarchical clustering, allowing Blue to create relevant clusters to target (Grigsby, 2018). As a result, a more tailored marketing mix can be designed to satisfy those segments' wants and needs (Maesen, 2020). Psychographic data can allow Blue to predict which types of messages a specific segment will respond to (Concordia, 2016). Targeted messages will then translate into an increase in sales for Blue. However, it is crucial for Blue to not only rely on the data but also consider the feasibility of satisfying these segments (Maesen, 2020). In other words, Blue might not have the financial resources to enter all the identified segments, so it should select the ones that are likely to be most profitable.

Secondly, to improve its data-driven decisions, Blue should consider calculating the return on marketing investment. It is vital for any business to understand their marketing mix strategy's effectiveness to determine if it is successful (Maesen, 2020). The simulation currently only provides data on the media channel preferences based on demographics, but there is no data available on the marketing campaigns' outcome. As a result, Blue is blindly investing in marketing without measuring if the campaigns meet the company objectives. However, the return on marketing investment can be calculated based on the available data in the simulation. This is by using the sales data and previous marketing spending on each media channel and running simple or multiple linear regressions to quantify each media channel's effect on sales (Maesen, 2020). Also, Blue can calculate elasticities of its marketing mix to measure the return on marketing investment. With this type of analysis, the brand can predict the percentage increase in sales when the marketing budget for a specific media type increases by a certain percentage. This allows for efficient use of the marketing budget by allocating it to the media that will provide the highest ROI (Maesen, 2020). Finally, by consistently analysing the data, Blue can judge the success of its marketing spending and determine if the strategy should be adjusted in the future by identifying when non-linear ROIs begin to occur. This is beneficial as it shows at which dollar amount marginal returns start to decrease (Maesen, 2020). Therefore, Blue can know when to stop spending money on an ad as it already achieved its maximum customer reach.

Finally, the third recommendation for Blue to improve data-driven decisions is to integrate and use new tools and different data, precisely social listening. By conducting social listening, Blue will collect, track, and interpret a variety of stimuli on social networks to determine people's thoughts on the company, which would be collected from comments and posts made on social media channels (Stewart and Arnold, 2017). Blue is then able to use this data to support psychographic segmentation and other marketing decisions (Stewart and Arnold, 2017). Moreover, Blue should critically listen to the data, meaning evaluating the messages and comments made to determine tone, opinions, and issues that customers are experiencing (Stewart and Arnold, 2017). This data will allow Blue to make better future decisions about the product and the marketing strategy, so it aligns with the target segment. Furthermore, Blue can accurately analyse the return on marketing investment by analysing comments on each launched campaign.

Although the simulation does offer a social media sentiment tab, it only collects data from Twitter and neglects other platforms. This tab allows Blue to see and monitor negative comments towards its brand, those being almost entirely about product quality. Nonetheless, the brand is not able to address product quality concerns in the simulation. Moreover, Blue can carry out an extensive sentiment analysis by exporting data from APIs and inputting it into R or other data analysis software (Maesen, 2020). The textual data will undergo tokenization to identify the most used words regarding the brand and whether those terms are positive or negative (Maesen, 2020). This process helps Blue identify issues and areas of improvement to address the negative comments and maintain the benefits of the product. Blue would be able to use the data from social listening to aid, not only in marketing spending, strategy, and segmentation but also in product development and ensuring that products meet their customers' needs. However, Blue should keep in mind that social listening can only be used to collect data from online platforms but does not capture offline customers' sentiment, likely to be those in older age demographics. Therefore, it is essential to bear in mind that the data will not represent Blue's entire age segment. Blue can further collect psychographic data on the older age demographics by running focus groups.

Focus groups are a qualitative data collection method (Deepak and Jeyakumar, 2019). In focus groups, Blue will understand the consumer's motivation for purchasing the product and their thoughts and attitudes towards the brand (Stewart and Shamdasani, 2016). The brand should focus on conducting focus groups for older age demographics to balance this segment's information gap on social platforms. The benefits identified by customers can then be used as the basis for future marketing messages (Stewart and Shamdasani, 2016). This would create a more targeted advertisement that customers will respond to and encourage them to purchase the product. Lastly, focus groups can suggest improvements for the product or its advertising (Stewart and Shamdasani, 2016); therefore, Blue would have a clear strategy for future product development.

References List

Concordia, 2016. *Cambridge Analytica - The Power Of Big Data And Psychographics*. [video] Available at: <https://www.youtube.com/watch?v=n8Dd5aVXLCc&feature=emb_title> [Accessed 11 October 2020].

Deepak, K. and Jeyakumar, S., 2019. Marketing Management. Educreation Publishing, p.24. Grigsby, M., 2018. *Marketing Analytics: A Practical Guide To Improving Consumer Insights Using Data Techniques*. 2nd ed. London: Kogan Page, Limited, pp.121-135.

Harvard Business Publishing. (2016) *Data-Driven Management of Blue Detergent*. [Online case study provided by faculty of the Imperial College Business School.]

Lin, C., 2002. Segmenting customer brand preference: demographic or psychographic. *Journal of Product & Brand Management*, [online] 11(4), pp.249-268. Available at: <https://www.emerald.com/insight/content/doi/10.1108/10610420210435443/full/html> [Accessed 10 October 2020].

Maesen,S., 2020. *Return on Marketing Investment*. [lecture] Imperial College London, 28 September 2020

Maesen,S., 2020. *Text Analytics*. [lecture] Imperial College London, 12 October 2020

Maesen,S., 2020. *Understanding consumers using segmentation analysis*. [lecture] Imperial College London, 21 September 2020

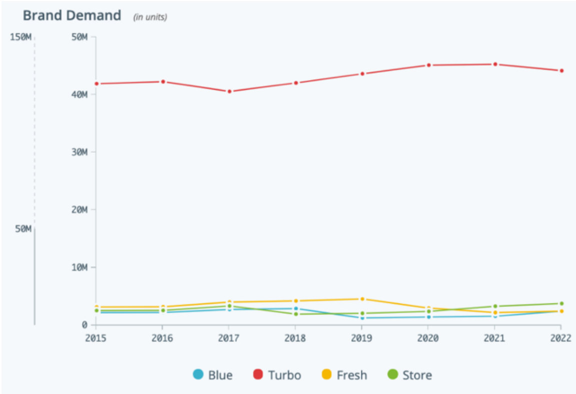
Stewart, D. and Shamdasani, P., 2016. Online Focus Groups. *Journal of Advertising*, [online] 46(1), pp.48-60. Available at: <https://www.tandfonline.com/doi/full/10.1080/00913367.2016.1252288?casa_token=j0PxT4tUrzcAAAAA%3Aflyo7HdNRiQZ8hO6tu2glu8taCsQgRP8kWs9HkIkycTQ9j7WCfQ43s3PdrlkFWQg63hPajjIP2cRNA> [Accessed 12 October 2020].

Stewart, M. and Arnold, C., 2017. Defining Social Listening: Recognizing an Emerging Dimension of Listening. *International Journal of Listening*, [online] 32(2), pp.85-100. Available at: <https://www.tandfonline.com/doi/full/10.1080/10904018.2017.1330656?casa_token=zGtX7Fr3l7kAAAAA%3AzjrRPKSLnGlztL1czK33yYO_gaxOzrW97F0pbeCy6SCL1FFUFKo3zWGycnfcFETuFWb8U0khXAc> [Accessed 11 October 2020].

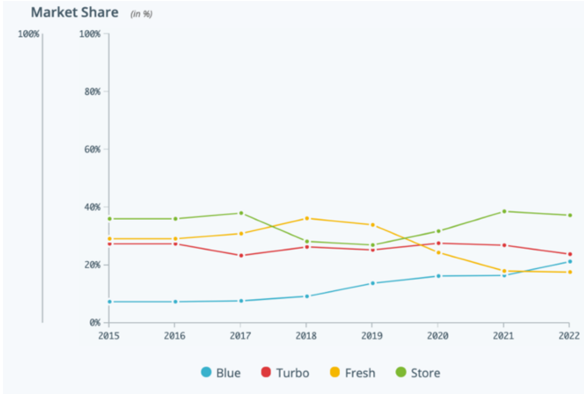
Appendix

Figure 1 – Market Share: Southeast and Central

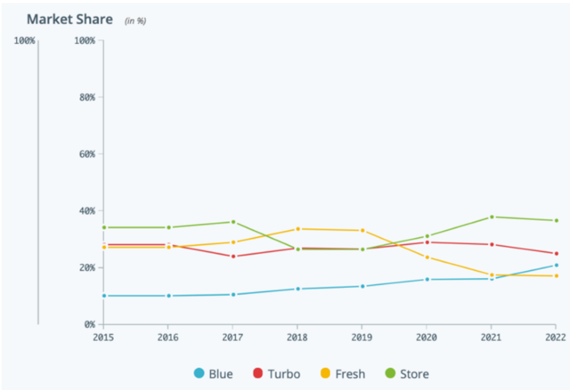
Northeast



Southeast



Central



West

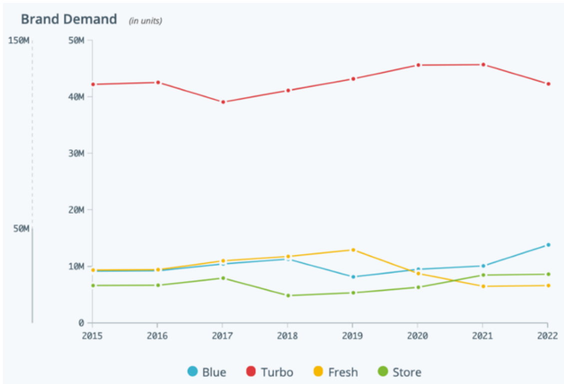
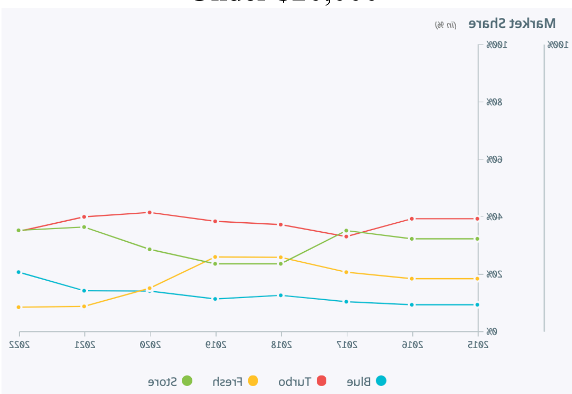


Figure 2 – Market Share: Lower to Average Income

Under $20,000



$20,000 - $59,999

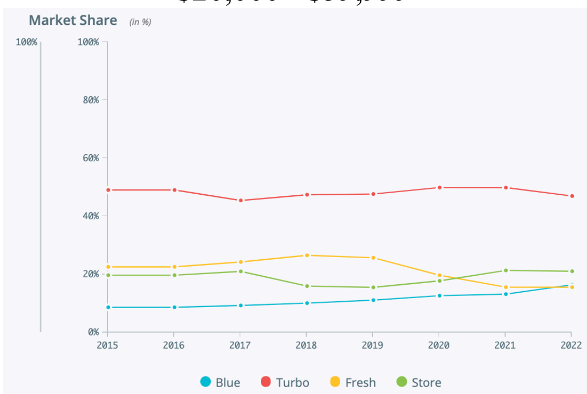
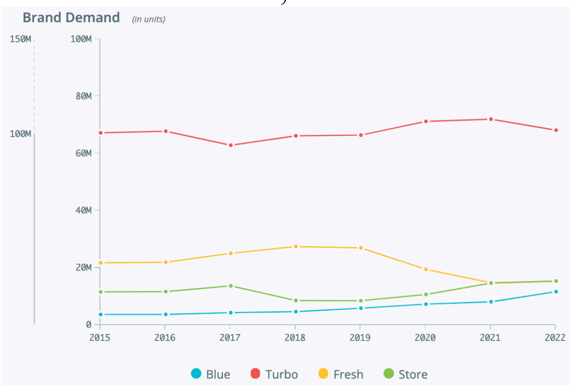


Figure 3 – Brand Demand and Brand Sales by Household Size (demand = sales)

1, 2



3, 4, 5+

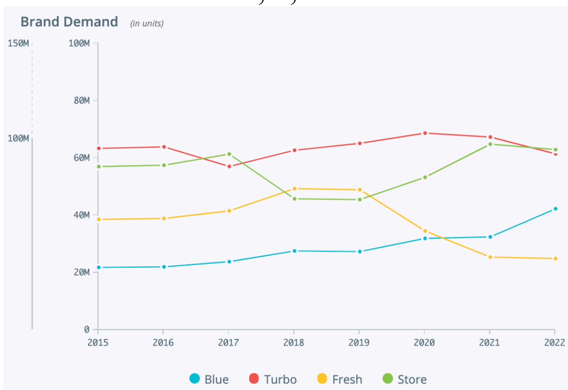
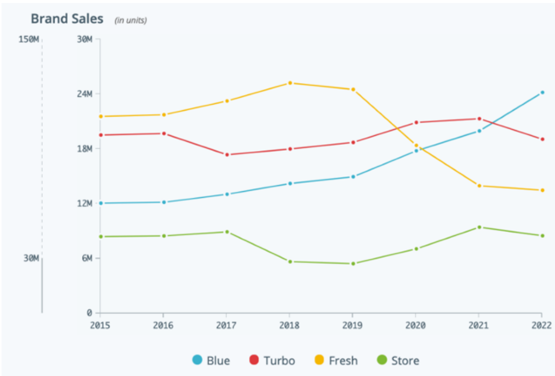
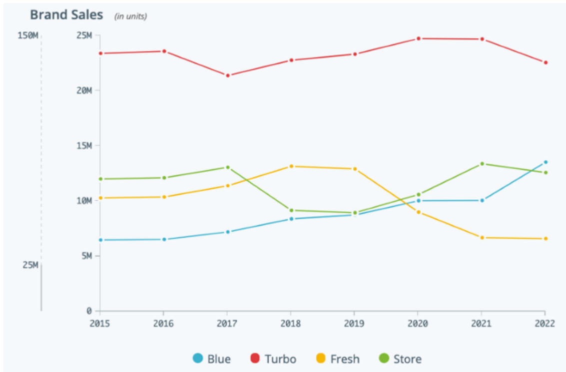


Figure 4 – Brand Sales by Age

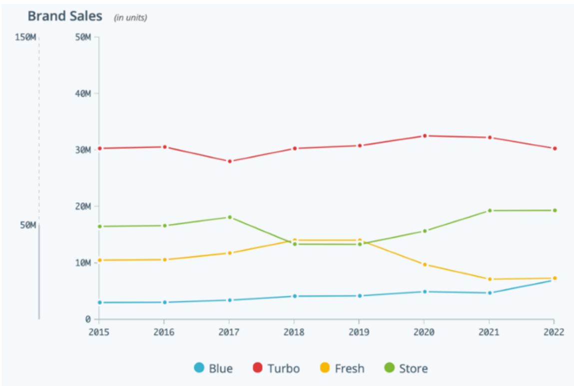
Under 35



35-44



45-54



54 and Over

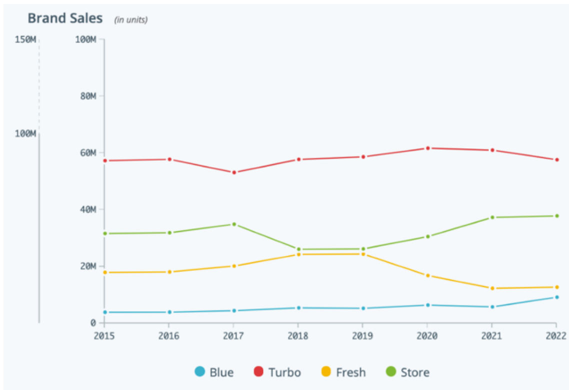


Figure 5 – Profit Margin: 2019 and 2020

	in millions of dollars	in millions of dollars	
Year	**Operating Profit**	**Revenue**	**Profit Margin**
2019	65.9	269	24%
2020	55.1	274	20%

Figure 6 – Price by Brand

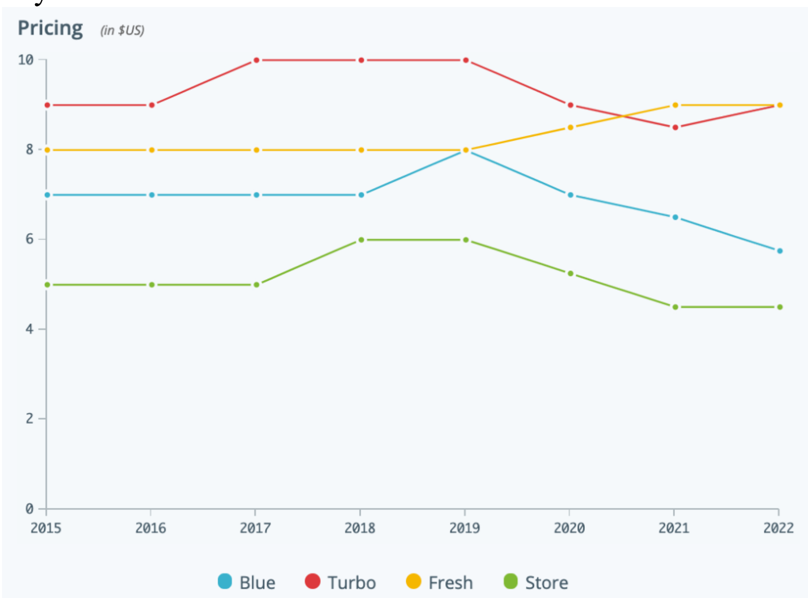
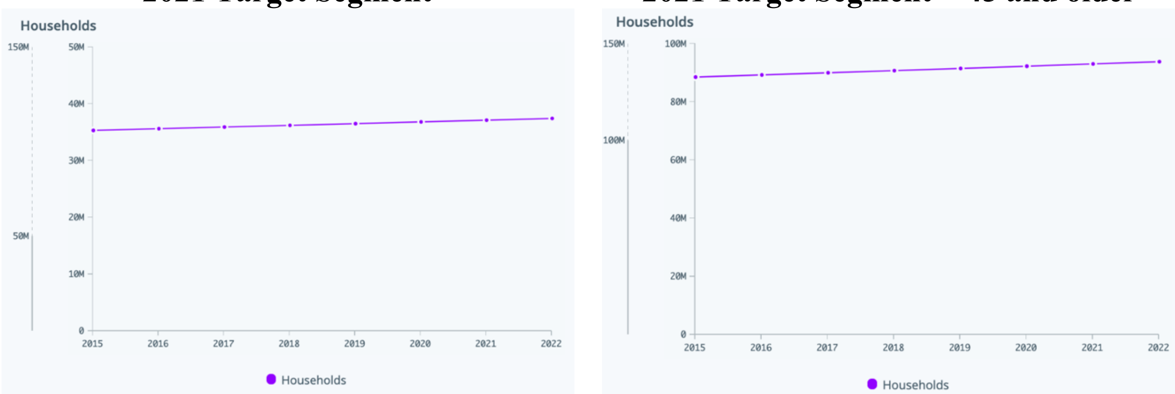


Figure 7 – Households: 2021 and 2022

2021 Target Segment **2021 Target Segment + 45 and older**



Marketing Analytics Course Work

Part II
Marketing Mix Model: Deospray

K12
01962176
01997207
01924400
01910348
01940804
01940804
01885310

Marketing Analytics
Dr. Stjin Maesen
Imperial College Business School
London, United Kingdom
October 15th, 2020

TABLE OF CONTENTS

MARKETING MIX SPECIFICATION

Describing and Motivating the Model

The aim of the created model is to gain a better understanding into what drives the sales of Brand 3 within Chain 5. Here, chain 5 represents the *place* within the marketing mix. The 5 brands that are presented in the data each represent a *product* within the marketplace. Within this model, the dependent variable is labeled as the sales level of a brand within a certain week within a certain year. This variable is denoted by "sales_brandX", where X stands for the relevant brand number.

Moreover, the data provided three other variables which serve as potential drivers for the sales level that a brand achieves. These three variables are labeled as, "price_brandX", "display_brandX" and "feature_brandX". The first of these denotes the average *price* level of a brand in a specific chain within a specific week. The following two variables refer to the levels of advertising that the brand may use in a specific period, these would symbolize the *promotion* of the product in certain periods. In addition to the provided variables, certain variables were manipulated and merged. These variables were identified as other potential sales drivers. These consist of the average price level of brand 3 in chains 1 to 4, the total sales of deodorant spray in chain 5 and the display & feature levels of brand 3 across the chains.

The first significant part of the model involved running a corrplot (line 13-16**, Figure 1**) to understand whether any variables were correlated to each other. If they were, it would be evaluated whether or not it is logical to remove them from the model. Only two variables were found to have a correlation score greater than 0.70. These two variables were the price of brand 4 and the price of brand 1. However, as these are two different brands which are likely ran by two different management teams, there is little reason to believe the prices are causally linked. Thus, these variables were not removed from the data set. The next step involved cleaning and organizing the data set into relevant columns, groups and specific combinations of variables as well as checking for missing values. This occurred in lines 22 to 76 of the model. The missing values output graph is in **Figure 2**.

Subsequently, in lines 78-82 all values were converted into logarithmic values. The first reason why this was done was because the data set was extremely skewed, and all values had different scales. Through using a logarithmic model, the data was normalized. The second reason is because in real life contexts, it is unlikely that the relationship between a sale and its driver is deterministic and constant across all levels. Using a logarithmic function allows for the possibility that the relationship is not linear and has diminishing returns or exponentiality.

Afterwards, the first regression was formed. This regression ("reg", line 85 & 86) included all of the initial variables and the variables that were created. From this regression, all statistically and marginally significant variables were identified, these would be used in the following models. All statistically insignificant variables were removed from the future, more fine-tuned models. This was done to prevent false inflation of the adjusted R-squared value, prevent a decrease in mean average error percentage (MAPE) and to simplify the model. Simplifying the model entails lowering the number of total variables to focus on their effects as well as making the model more manageable and actionable. A large model with a high number of variables can end up being too complex and vague to have any real-world value or meaning. This was followed by centering all of the logarithmic variables to ensure the normalization of the data.

Having completed this, the carryover effects of promotions were accounted for in the data set. This is because without accounting for them, the model will assume that the advertising campaigns will only affect the sales within the same period that the campaign takes place in. This does not happen in the real world, and hence must be dealt with. Otherwise the effect of advertising on sales will be undervalued and the elasticity values will be inaccurate. This is observed in lines 103 to 113 in the code.

After having prepared and split the data into a training and hold out group (lines 114 – 120), the model that only included the statistically significant variables was developed in lines 122 to 125. This model had the name, “full”. This model not only included independent variables, but also the interaction effects between all independent variables.

Next, having only selected the significant variables and interactions effects, a new model, “reg_adj” was created. In this model, a VIF test was conducted to see whether any remaining independent variables were correlated. From this, three variables were found to have VIF scores that were greater than 5, indicating significant levels of correlation. From this, 3 variables were found to have VIF scores that were greater than 5, indicating significant levels of correlation. Two of them were removed because if the variables had remained in the model, the R squared will be falsely inflated which means the model’s validity is overstated. The subsequent, and final model was called "reg_final". As a result of removing 2 correlated variables, the VIF score of c5_sales became acceptable and fell below 5 **(**see **Figure 3** and **Figure 4).**

Subsequently the mean average percentage error (MAPE) of the final 3 models (“full”, “reg_adj” and “reg_final”) were calculated to understand the impact of removing insignificant variables. As a result of these tests and the removal of insignificant variables, reg_final is a model with high predictive accuracy (low MAPE), a high explanatory power (high adjusted R-squared) and a low likelihood of being overfitted.

Final Equation

The final result of this model leads to the following equation:

$$\log(c5\_sales\_brand3) = 3.37161 - 0.25026 * \log(c5\_sales\_brand1) - 0.09632 * \log(c5\_sales\_brand2) - 3.08756 * \log(c5\_price\_brand3) + 0.71245 * \log(c5\_price\_brand) + 0.28799 * \log(c5\_sales) + 3.12577 * \log(price\_brand3) + 0.26100 * \log(c5\_sales\_brand1{:}c4\_sales\_brand4) + 1.09014 * \log(c5\_sales\_brand1{:}c5\_sales\_brand3) - 2.14668 * \log(c5\_price\_brand3{:}c5\_sales)$$

RESULTS

Results Table

Table 1. Summary of Model Estimation Results

Coefficient:	**Estimate**	**Std. Error**	**T value**	**Pr(>|t|)**	**95% Conf. Interval**
Intercept	3.37161	0.06965	48.409	<2e-16	(3.23322, 3.51000)
c5_sales_brand1	-0.25026	0.07865	-3.182	0.002016	(-0.40654, -0.09397)
c5_sales_brand2	-0.09632	0.05746	-1.676	0.097215	(-0.21010, 0.01786)
c5_price_brand3	-3.08756	0.22810	-13.536	<2e-16	(-3.54078, -2.63432)

c5_price_brand5	0.71245	0.16639	-4.282	4.67e-05	(0.38183. 1.04307)
c5_sales	0.28799	0.09044	3.184	0.002000	(0.10829, 0.46769)
Price_brand3	3.12577	0.83844	3.728	0.000339	(1.45981, 4.79173)
C5_sales_brand1:c4_sales_brand4	0.26100	0.19107	1.366	0.175391	(-0.11866, 2.82229687
c5_sales_brand1:c5_sales_brand3	1.09014	0.87176	1.251	0.214392	(-0.64202, 2.82230)
C5_price_brand3:c5_sales	-2.14668	0.83700	-2.565	0.012001	(-3.80977, - 0.48358)

**R-squared: 0.7673, Adjusted R-squared: 0.7438*
**T-statistic: 32.61 on 9 and 89 DF, p-value: < 2.2e-16*

Model Assumption

Statistical **Assumptions** about the probability distribution of ϵ of modeling:

1. The mean of the probability distribution of is 0.
2. The variance, of the probability distribution of is constant for all settings of x.
3. The probability distribution of is normal.
4. The errors associated with any two different observations are independent.

Model Accuracy

In the hypothesis test,

$$H_0: \beta_{1,}\ \beta_{2,}\ \beta_3 \ldots \beta_{11} = 0$$
$$H_0: \text{at least one } \beta_{1,}\ \beta_{2,}\ \beta_3 \ldots \beta_{11} \neq 0$$

Given that p ≈ 0 (T = 32.61, p = <2.09e-16) > α = 0.05, there is sufficient evidence to reject the null hypothesis and conclude that at least one of the effects is useful for predicting sales of brand 3 in chain 5.

Adjusted R square: The adjusted r squared, is 0.7438. About 74.38% of the variation in sales of brand 3 in chain 5 can be explained by the model after adjusting for the sample size and number of predictors.

MAPE: The mean absolute percent error between predictive sales of chain 5 in brand 3 and actual values is 4.87% which is appropriate.

Variable Interpretation – First Order Variables and Interaction Effects

Intercept: When all the variables equals to 0, the log of estimate average sales of chain 5 in brand 3 is 32.79021 units. The interpretation is invalid because it is impossible for all variables equals to 0.

On average, a 1% percent increase in sales of Brand 1 in Chain 5 will lead to a statistically significant 1.2844% (T= -3.182, p=.02) decrease in the sales of brand 3 in chain 5, holding all other variables constant.

On average, a 1% percent increase in sales of Brand 2 in Chain 5 will lead to a marginally significant 1.1011% (T= -1.676, p=.0972) decrease in the sales of brand 3 in chain 5, holding all other variables constant.
On average, a 1% percent increase in price of Brand 3 in Chain 5 will lead to a statistically significant 21.9235% (T= -13.536, p ≈ 0) decrease in the sales of brand 3 in chain 5, holding all other variables constant.
On average, a 1% percent increase in price of Brand 5 in Chain 5 will lead to a statistically significant 2.0390% (T= -4.282, p ≈ 0) increase in the sales of brand 3 in chain 5, holding all other variables constant.
On average, a 1% percent increase in the sales of all brands in Chain 5 will lead to a statistically significant 1.3337% (T= 3.184, p =.002) increase in the sales of brand 3 in chain 5, holding all other variables constant.
On average, a 1% percent increase in the average price of Brand 3 in chains 1 to 4 will lead to a statistically significant 22.7774% (T= 3.728, p ≈ 0) increase in the sales of brand 3 in chain 5, holding all other variables constant.

In the final model, there were three interaction effect variables. Of these, the interaction effects between “brands 1 and 4 sales in chain 5” and “brands 1 and 3 in chain 5” were found to be statistically significant. The remaining interaction variable, “price of brand 3 in chain 5 and total deodorant spray sales in chain 5” were found to have a statistically significant impact on sales (T= 2.565, p = 0.01). If both the price of brand 3 and the total sales level of deodorant spray in chain 5 were to increase by 1%, there would be an additional decrease in the sales of brand 3. This decrease would be greater than the sum of the decreases caused by the two variables separately.

DISCUSSION

Summary

In the following section, the main findings, their implications, the limitations of the analysis and the model’s validity are presented. This will allow for a deeper understanding of brand 3's sales drivers in chain 5. These variables can be used for further strategy development when modified. Moreover, unless stated otherwise, all variables mentioned below are assumed to be within chain 5.

The final model is a regression analysis which tests the effect of a change in one predictor variable on the response variable, *ceteris paribus* (holding all other variables equal). The final model only includes statistically significant first order variables to ensure that all the included variables are actual sales drivers. Starting with the first order variables, the regression analysis indicated that three of the six first order variables had a negative relationship with sales. These three variables were the sales of brand 1 and brand 2, and the price of brand 3. However, it must be noted that the sales of brand 2 are only marginally significant whilst the other two are statistically significant. All of these variables have an exponential effect on the sales of brand 3 because their elasticity percentages are greater than 1%. This means that as the price of brand 3 increases, the sales of brand 3 fall at an increasing rate. The three remaining variables all had positive relationships to the sales level of brand 3. These variables also have an exponential marginal effect on the sales of brand 3. The variables with the greatest impact on the sales of brand 3 were related to brand 3’s pricing. The effect of brand 3’s price, both within and outside chain 3, on sales dwarfed the effect of other dependent variables.

Out of the interaction variables, only the interaction between the price of brand 3 in chain 5 and the total deodorant spray of sales within chain 5 were found to be statistically significant.

This relationship was negative which implies that as these two variables increase in value, their combined effect on sales is greater than the sum of their effects as single order variables.

In addition, this regression (reg_final) was found to have an explanatory power of 74.38%. This is a high value when considering the endless number of variables that can impact the sales of a brand. It also performed strongly when predicting the sales value of the holdout data set. It was able to predict the sales value with a mean average percentage error (MAPE) of 4.87% which indicates our model has a relatively high level of accuracy due to it greatly exceeding the 20% minimum threshold.

Implications and Recommendations

In terms of product offering, the greatest competitor and substitute to brand 3 is brand 5. This is shown through the cross elasticity of demand between brand 3 and brand 5 which is 2.04 ($XED = \frac{\%\, change\ in\ sales\ of\ brand\ 3}{\%\, change\ in\ price\ of\ brand\ 5} = \frac{2.04}{1}$). Cross elasticity of demand measures the responsiveness of quantity demanded in one product relative to a price change in another product, *ceteris paribus* (Blink and Dorton, 2012). Brand 5 was the only brand whose price had a statistically significant impact on brand 3's sales level. It is likely that both brand's products target similar customers and satisfy similar consumer needs. This would explain why consumers show a high level of price awareness between the two products (Blink and Dorton, 2012). In the future, the brand 3's pricing strategy should consider the potential pricing strategy of brand 5. This would avoid underpricing which leads to a price war and loss of category value, or overpricing which would lead to a lower cost-to-value ratio and achieved sales value. Brands 1, 2 and 4 are also shown to be substitutes and competitors to brand 3. However, the substitutability of brands 1, 2 and 4 is not driven by price. From this, the likely conclusion is that these brands are positioned differently. This is because brands with similar offerings tend to compete on price. If they do not compete on price, as is the case here, they tend to compete on their offering and key product attributes

Moreover, the data indicates the price of brand 3 in chains 1 – 4 has the greatest statistical effect on the sales value of brand 3 in store 5. This would indicate that consumer loyalty to a chain is not outweighed by the benefits of a lower price. In other words, consumers are willing to switch where they shop to realize the temporary financial benefits of a chain specific price promotion. To ensure that consumers are not solely switching between stores to buy at the lowest price, which dilutes the profit margin, price promotions should be coordinated between stores so that they occur in the same week.

Finally, growing the total sales of deodorant spray in chain 5 is very beneficial to helping brand 3 grow its market share and own revenue. As the sales in chain 5 increase by 1 %, the sales of brand 3 increase by 1.33%. This would indicate that a large part of new demand within chain 5 goes to brand 3 as opposed to going to competitor brands.

Validity

Validity refers to the extent to which the model accurately measures the construct it is intending to measure (Shantikumar, 2018). The first measure that can indicate the final model's validity is the Adjusted R-squared value which indicates what percentage of the dependent variable's variance is explained by the independent variables (Maesen, 2020). The final model was able to account for 74.38% of the variance in sales levels. This value is high and would indicate a high level of validity in our model. However, a criticism of the Adjusted R-Square is that the value can be falsely inflated simply by adding more variables into the model, some of which may not have a strong or relevant correlation (Maesen, 2020). This is not the case in our model. First, we ensured that all variables that were not statistically significant were removed from the regression equation (reg_full). Next, all variables in our penultimate model (reg_adj) were tested for multi collinearity using the VIF function. This is done to remove correlated independent variables who have similar effects and falsely inflate the adjusted R-squared variables (Petrie, n.d). Three variables were found to have VIF scores above 5 (see **Figure 3**). These were the interaction variable, "c5_sales_brand4:price_brand3" and first order variables, "c5_sales_brand4" and "c5_sales." The first two variables were removed, which resulted in a new and final regression model in which there were only low correlations between the independent variables (see **Figure 4**). Doing so caused the adjusted R Squared to fall from an initial 84.86% (reg_full) to 74.38% (reg_final). Despite the lower R squared value, the model is still highly accurate and more accurately reflects its ability to predict sales.

The second measure of validity is the MAPE of the model which is a measure of prediction accuracy. It compares the accuracy of the model's prediction for the hold out data to the actual values (Maesen, 2020). In this case, the model's predictions had a MAPE of 4.87% which indicates that the model has a high level of predictive ability and validity, this can be seen in **Figure 5** by comparing the predicted and actual sales values. It is unlikely that we have over fitted our model as numerous measures were taken to remove insignificant and non-correlated variables without hurting it's predictive ability. Moreover, as rule of thumb for deciding on the number of variables ~~is to~~ one should have no less than 20 observations per variable (Maesen, 2020). The final model adheres to this by having 13.8 observations per variable. In conclusion, our model has a high level of validity and all relevant precautions were taken to ensure this.

Limitations and Further Research

The model took sufficient steps to ensure that it had a high level of validity. However, the assumptions of the model were not tested. If the assumptions mentioned in Section 2 are violated, the model may have limited levels of validity(Duke University, n.d). This is because the model may contain certain biases, be misleading or could be inefficient. Moreover, an outlier analysis could have been done to ensure that the data is more representative of actual sales patterns observed in the real world. This would increase the generalizability of the model. Finally, many other sales drivers are neglected in the model. Examples of these would be advertisement levels in other media channels such as TV and digital, the percentage share of total shelf space and the effect of complementary products & bundles. Missing these variables could lead way to an omitted variable bias.

In addition to adding other sales drivers into a model, a quick glance at the data would show that the sales of all deodorant sprays are higher in summer months (estimated to be at weeks 23 – 36). This would indicate that there are non-numerical and binary variables which impact sales too. Future analysis could look into the impact of these variables as well as the effectiveness of doing campaigns in summer months vs winter months.

Further research should also be done to find out the implications of the interaction effects of independent variables and how these can be harnessed to achieve greater overall returns. The theory behind this is that interaction effects can result in better ROI of investments compared to investment in solely first order variables.

Finally, given the extremely high impact of price changes of brand 3 on sales, further research here is recommended to help devise optimal and accurate pricing strategies as poor pricing has the greatest effect on sales.

Reference List

Blink, J. and Dorton, I., 2012. *IB Diploma Programme: Economics Course Companion*. 2nd ed. Oxford: Oxford University Press.

Duke University, n.d. *Testing The Assumptions Of Linear Regression*. [online] People.duke.edu. Available at: <http://people.duke.edu/~rnau/testing.htm> [Accessed 13 October 2020].

Maesen,S., 2020. Return on Marketing Investment. [lecture] Imperial College London, 28 September 2020

Petrie, A., n.d. *VIF Function | R Documentation*. [online] RDocumentation. Available at: <https://www.rdocumentation.org/packages/regclass/versions/1.6/topics/VIF> [Accessed 13 October 2020].

Shantikumar, S., 2018. *Validity, Reliability And Generalisability*. [online] Health Knowledge. Available at: <https://www.healthknowledge.org.uk/content/validity-reliability-and-generalisability> [Accessed 13 October 2020].

Appendix

Figure 1 – Corrplot

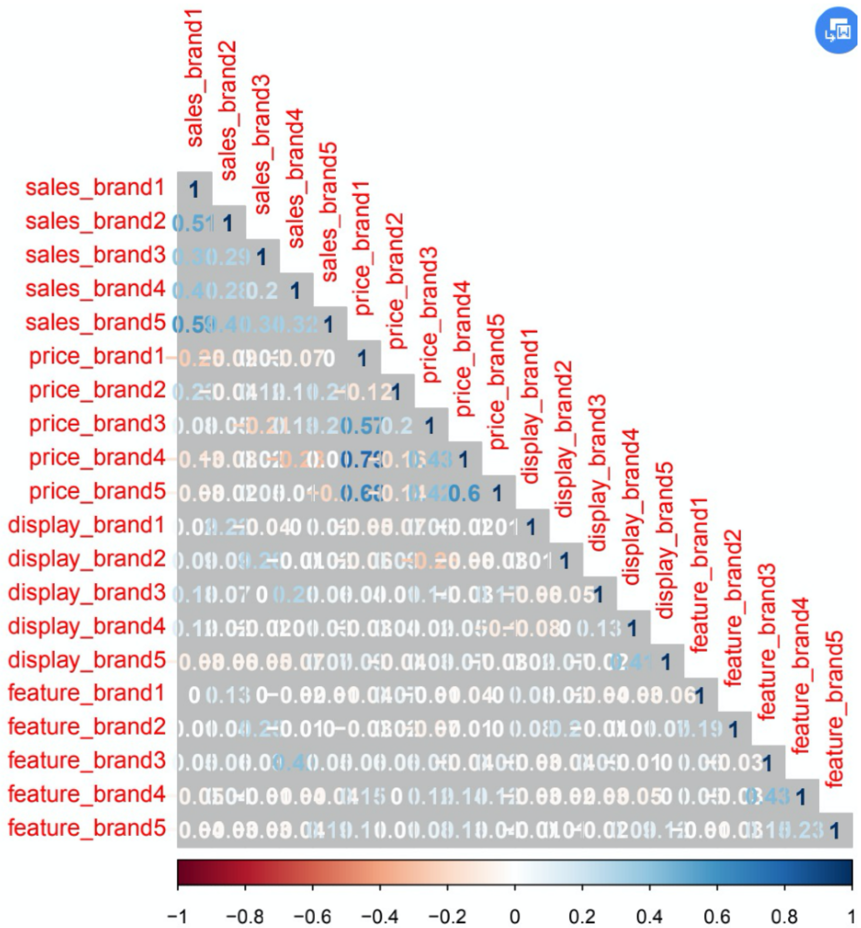


Figure 2 – Missing Values Output Graph

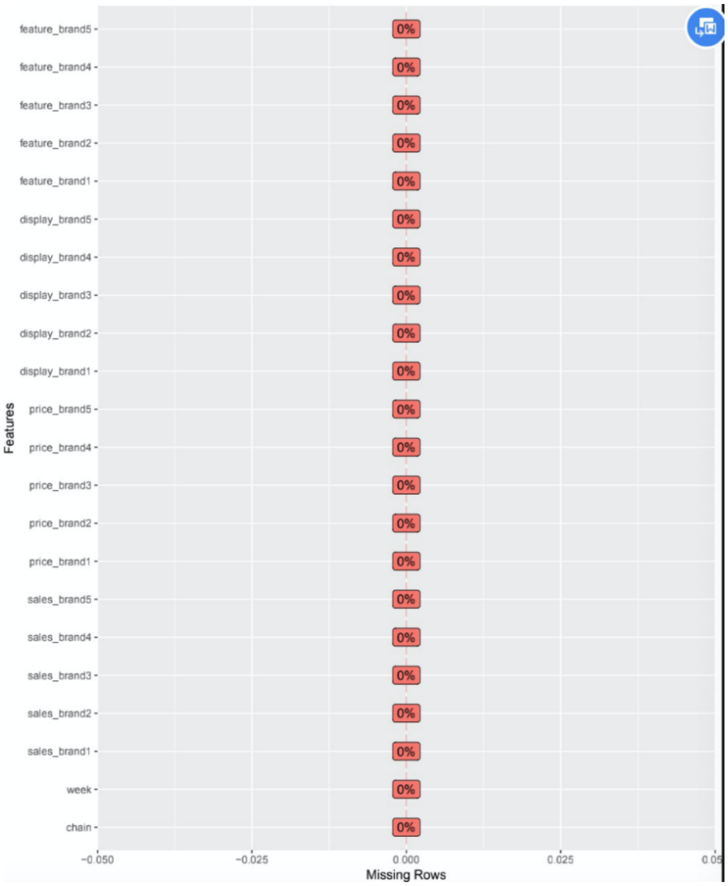


Figure 3 – VIF Score for Penultimate Model (reg_adj)

```
                 c5_sales_brand1                  c5_sales_brand2
                        1.754517                         2.590427
                 c5_sales_brand4                  c5_price_brand3
                       28.746913                         2.299516
                 c5_price_brand5                         c5_sales
                        1.721061                         6.959099
                    price_brand3  c5_sales_brand1:c5_sales_brand4
                        2.912444                         1.826978
 c5_sales_brand1:c5_price_brand3     c5_sales_brand4:price_brand3
                        1.974232                        25.166457
          c5_price_brand3:c5_sales
                        2.426480
```

Figure 4 – VIF Score for Final Model (reg_final)

```
           c5_sales_brand1                    c5_sales_brand2
                  1.590303                           2.163975
           c5_price_brand3                    c5_price_brand5
                  2.149084                           1.268885
                  c5_sales                       price_brand3
                  1.905906                           2.369198
c5_sales_brand1:c5_sales_brand4 c5_sales_brand1:c5_price_brand3
                  1.209100                           1.850970
      c5_price_brand3:c5_sales
                  2.266935
```

Figure 5 – Visualization on Comparing Actual and Predicted Values