Emory University

Master of Science in Business Analytics Program

Data-Driven Marketing Insights for Pernalonga

Alice Jiang, Cassio Salge, Qiyu Wang

MKT 680 Marketing Analytics

Professor David Sackin and Alvin Lim

2021/2/13

**Table of Contents**

Executive Summary…………………………………………………………………….3

1. Introduction………………………………………………………………………………3
   1. Background………………………………………………………………………4
   2. Data Exploration and Cleaning………………………………………………...4
2. Exploratory Data Analysis……………………………………………………………...4
   1. Overview………………………………………………………………………....4
   2. Competitor Analysis…………………………………………………………….4
   3. Colgate-Palmolive toothpaste analysis (product level)...............................7
   4. Colgate-Palmolive toothpaste analysis (customer level)............................8
3. Segmentation…………………………………………………………………………....9
   1. Overview………………………………………………………………………....9
   2. Calculation of margin and considerations…………………………………….9
   3. Segmentation: Store Level……………………………………………………10
   4. Segmentation: Toothpaste Level……………………………………………..12
   5. Segmentation: Colgate Toothpaste Level…………………………………...13
4. Building Similarity Model……………………………………………………………...13
   1. Identifying complementary and substitutional goods……………………....13
   2. Finding customers to target…………………………………………………..15
   3. Evaluation.. …………………………………………………………………….18
5. Recommendations…………………………………………………………………….19

**Executive Summary**

This project was commissioned to develop a marketing campaign that utilizes data analysis to provide insights and recommendations for creating personalized promotions for Colgate-Palmolive toothpaste products at Pernalonga stores. According to our findings, Colgate-Palmolive should market their toothpaste to customers who are already loyal to Colgate toothpaste, as well as a mix of customers who frequently buy other Colgate products and customers who primarily buy other brands of toothpaste. We will offer Colgate toothpaste promotions to our target customers because we know Colgate is interested in a promotional campaign to boost sales of Colgate toothpaste.

We believe that targeting these customers will increase the likelihood of purchasing Colgate toothpaste by 16 percent. As mentioned in our report, promotions would differ depending on the customer. Pernalonga can expect this 14-day promotional program to generate an additional sales of close to $4041.7 by providing 10% discounts through personalized Colgate toothpaste promotions, according to our analysis. Please note that all outcome values are approximate and are based on certain assumptions and past transactional data.

**1.Introduction**

**1.1 Background**

Pernalonga is an undisputed leader in the retail space of Lunitunia, with over 10,000 products in 400+ categories. In order to drive sales, Pernalonga finds itself in a place where it depends on promotions, as over 30 percent of sales come from promotions. In-store promotions are currently the majority of the promotions at Pernalonga. The issue with in-store promotions is that they offer customers temporary price reductions irrespective of their needs. Therefore, by offering promotions to those who would be willing to buy the items at full price, Pernalonga is losing part of its revenue. The company is interested in partnering with Colgate-Palmolive to provide personalized promotions to specific target customers in order to boost sales of Colgate toothpaste. Pernalonga is looking for targeting customers on the basis of how likely they are to cash in on the promotion. We were able to solve this problem and create a personalized promotion marketing campaign for Colgate toothpaste by examining Pernalonga's transaction-level data.

**1.2 Data Exploration and Cleaning**

We were provided two files:

* transaction table which contains 2,961,785 observations and 12 variables covering transaction data from 2016 to 2017.
* product table which contains 10767 observations and 7 variables covering product information.

Some discrepancies were found in the data (i.e. stores with $5 in sales, transactions with random sales volumes and discounts applied incorrectly to products). These observations were removed. Another anomaly was that the transaction IDs were captured incorrectly resulting in only 753 unique values. In order to fix this issue, we used a combination of customer ID, store ID and transaction date to define a new transaction ID. This approach resulted in over 2.8 million unique records for transaction ID with a varying number of products purchased across customers. Once we cleaned and then understood the structure of the data, we started looking more in-depth into toothpaste products and current / potential customers to explore whether there are any trends we could find. Last but not the least, we notice there are products like plastic bags that are not products for generating profit, so we remove those from analysis too.

**2. Exploratory Data Analysis**

**2.1 Overview**

Colgate, Private Label, Aquafresh, Sensodyne, and Oral B are the five toothpaste brands available at Pernalonga. We analyzed products from all five brands in order to understand Colgate’s products, competitors’ products and complementary products.

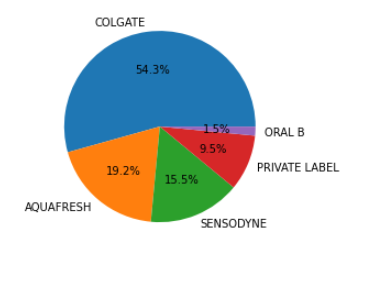
**2.2 Competitor Analysis**

Revenue Analysis:

Pernalonga carries 77 distinct toothpaste products across five brands (Colgate, Sensodyne, Aquafresh, Oral B and Private Label). These products brought in $157,758.99 in 2017 which represents an increase of almost 16% when compared to 2016 ($136,175.58). Colgate is the current market leader with 54.34% of the market followed by Aquafresh, Sensodyne, Private Label and Oral B. In addition, Colgate was also the market leader for total units sold. The company sold more than twice as much Private Label products.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **BRAND** | **2016**  **Revenue** | **2017**  **Revenue** | **TOTAL REVENUE** | **MARKET SHARE** |
| COLGATE | $ 71,328.87 | $ 88,385.78 | $ 159,714.65 | 54.34% |
| AQUAFRESH | $ 28,068.36 | $ 28,259.99 | $ 56,328.35 | 19.16% |
| SENSODYNE | $ 21,418.62 | $ 24,179.70 | $ 45,598.32 | 15.51% |
| PRIVATE LABEL | $ 13,543.79 | $14,276.14 | $ 27,819.93 | 9.46 % |
| ORAL B | $1,815.94 | $ 2,657.38 | $ 4,473.32 | 1.52% |
| **TOTAL** | **$157,758.99** | **$ 136,175.58** | **$ 293,934.57** | **100%** |

*Table 1: Market Share and revenue for each toothpaste brand*

****

*Figure 1: Market share in pie chart*

Discount Analysis

The toothpaste category was heavily discounted in 2016 when compared to 2017. Total discounts in 2016 summed up $68,682.76 against $33,397.08 in 2017.

|  |  |  |
| --- | --- | --- |
| **BRAND** | **TOTAL DISCOUNTS**  **2016** | **TOTAL DISCOUNTS**  **2017** |
| COLGATE | $ 44,043.85 | $ 22,648.35 |
| AQUAFRESH | $ 17,333.98 | $ 6,433.18 |
| SENSODYNE | $ 7,000.69 | $ 3,948.80 |
| PRIVATE LABEL | $ 179.78 | $ 216.61 |
| ORAL B | $ 124.46 | $ 150.14 |
| **TOTAL** | **$68,682.76** | **$ 33,397.08** |

*Table 2: Discounts in 2016 and 2017 for all brands*

Even though there were less discounts in 2017 when compared to 2016, some companies such as Colgate actually sold more units in 2017.

|  |  |  |  |
| --- | --- | --- | --- |
| **BRAND** | **TOTAL UNITS SOLD**  **2016** | **TOTAL UNITS SOLD**  **2017** | **PERCENTAGE**  **CHANGE** |
| COLGATE | 34,974 | 38,120 | 9% |
| AQUAFRESH | 14,638 | 12,119 | * 17.21% |
| SENSODYNE | 6,564 | 6,265 | * 4.56% |
| PRIVATE LABEL | 15,123 | 15,879 | 5% |
| ORAL B | 520 | 781 | 50.19% |
| **TOTAL** | **71,819** | **73,164** |  |

*Table 3: Total units sold in 2016 and 2017 and the percentage change*

We calculated the discount ratio (product discount amount / product sales amount) across brands in order to understand whether there is a relationship between discount and the purchase patterns of customers.

|  |  |
| --- | --- |
| **BRAND** | **DISCOUNT RATIO** |
| AQUAFRESH | 29.67% |
| COLGATE | 29.46% |
| SENSODYNE | 19.36% |
| ORAL B | 7.08% |
| PRIVATE LABEL | 1.12% |

*Table 4: Brands and discount ratio*

We noticed that there is no particular relationship between higher discount ratio and higher purchase quantity. For example, the discount ratio of Private Label is 1.12% but 98% of customers bought 1 or 2 toothpastes per transaction. On the other hand, Aquafresh, Colgate and Sensodyne have a higher discount ratio but most people still purchased 1 or 2 toothpastes per transaction. Thus, we believe that’s the common buying pattern for the toothpaste section, and it’s quite inelastic to discounts. So, we recommend promoting bundles with no more than 2 toothpastes.

**2.3 Colgate-Palmolive toothpaste analysis (Product Level)**

There are 4 sub-categories of toothpaste from Colgate: pasta dentifr tradic, pasta dentifr branq, pasta dentifr medici, pasta dentifr infant. In the table below, we can see that pasta dentifr tradic toothpaste takes a major lead (83.9%) in the most lucrative toothpaste subcategory of Colgate.

|  |  |  |  |
| --- | --- | --- | --- |
| **COLGATE TOOTHPASTE** | **TOTAL SALES** | **% in SALES** | **PROFIT MARGIN** |
| PASTA DENTIFR TRADIC | $133,958.46 | 83.9% | 29.80% |
| PASTA DENTIFR BRANQ | $ 20,160.51 | 12.6% | 29.88% |
| PASTA DENTIFR MEDICI | $ 2,964.07 | 1.9% | 17.60% |
| PASTA DENTIFR INFANT | $ 2,631.61 | 1.6% | 18.34% |

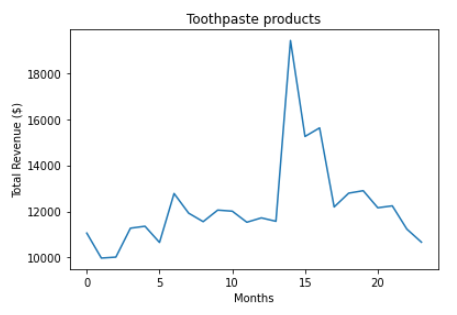
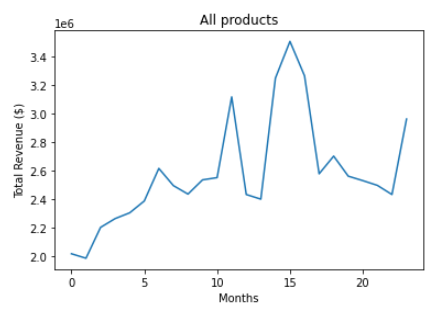
*Table 5: Colgate subcategories sales and profit margin*

We also noticed that the Colgate toothpaste products follow the 80/20 rule. The top 20% of the products generate over 80% of the total sales.

**2.4 Colgate-Palmolive toothpaste analysis (customer level)**

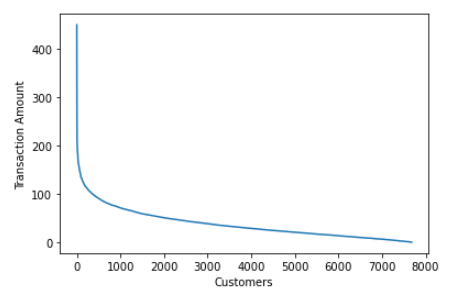
There are 7920 unique customers at Pernalonga. 97% of them have bought a toothpaste product and 87% have bought a Colgate toothpaste product. We also found that 82% of customers buy a quantity of 1 toothpaste at a time and 13% buy 2 at a time. In addition, more popular brands could sell more toothpastes, up to 12 toothpastes at a time while less popular brands could only sell 4. This indicates selling two toothpastes in a bundle sale could be profitable.

In the two figures below, we can see that there is a similarity in the trend of the sales of toothpaste and the sales of all products at Pernalonga. However, the difference in two trends is the range of sales. Sales of all products have a larger variation in trends ($2,000,000 ~ $3,500,000) throughout a year while toothpaste sales ($10,000 ~ $19,000) do not differ less from month to month. Particularly, in month 12, the general sales skyrocket likely because of Christmas where the sales for toothpaste stays below average.

****

*Figure 2:Revenue distribution across 2-years Figure 3: Revenue distribution across 2-years*

Our most valuable customers in buying toothpaste spent $542, $330, $321 toothpaste, where the most valuable customers in buying Colgate toothpaste spent $511, $322, $288. However, most people spend much less than half of the most valuable customers. The lowest spending on toothpaste is $0.84, and on Colgate toothpaste is $1.34. We also noticed that a small percentage of customers contributed to a high percentage of total sales.

****

*Figure 4: Transaction amount on customers*

**3. Segmentation**

**3.1 Overview**

We segmented customers on three different levels:

**Store Level** - segmentation on different purchasing behavior of customers and how they fall into different groups (store-loyal customers, regular customers, cherry pickers and bulk purchasing customers).

**Toothpaste Level** - segmentation based on purchase of all toothpaste brands and subcategories. Provides a clear picture of toothpaste purchasing behavior and identifies which customer groups to target in selling Colgate toothpastes.

**Colgate Toothpaste Level** - segmentation based on purchase of Colgate toothpaste products to identify whether different Colgate customer groups might have preferences on specific product types.

**3.2 Calculation of profit margin and considerations**

Before any segmentation was performed, we estimated the profit margin ratio for each customer. This information will not only help us identify the most profitable customers for Pernalonga, but also be used as a feature in our segmentation. The formula used is shown below:

Definition of components:

Base Unit Price - the average price paid for the product.

Minimum Price - for products having “KG” as units, the minimum of unit prices of the same season in two different years. For products having “CT” as units where these products are less likely to have a huge fluctuation in prices, the minimum price is the minimum unit price in the two-year period.

Transaction Sale Quantity - quantity sold for each product.

Transactions Paid Amount - sum of actual amount paid in each transaction.

Considerations about Minimum price (Cost of Goods Sold)

The minimum price is a price we assigned the products based on their past prices in order to calculate a general margin taking the variability of prices into account. We first observed and graphed the monthly distribution of paid amounts across the two-year period. The products with “KG” units are mostly fresh vegetables, fruits and meat that have a variability in prices. Thus, we took the unit prices paid from the same seasons in two years as base unit price and extracted for each store the lowest as our minimum price for products with “KG” units for that store. For products with “CT” units which are consumer packaged goods in most cases, which is less sensitive to seasonality, we decided to use the unit prices paid as base unit price for the entire two years period and extracted the lowest as our minimum price for each store as the minimum price for that particular store. Then we consider those minimum prices as the costs of goods sold.

As analyzing the prices, we noticed different products have different margin levels, so there’s no perfect way to decide the cost. For example, Oral-B barely has any discount, so their minimum prices stay high. But it doesn’t mean their margin is low. Also, for most goods, there are still margins even at their lowest price. However, we don’t have access to that information with the data given. We might assume there’s a certain percentage of margin for goods at minimum price and take that into account, but since it’s happening to all goods it will become indifferent for our analyses. So, we end up deciding to stick to the minimum prices and use it as our costs to calculate margin.

**3.3 Segmentation: Store-Level**

Feature Selection

We began our analysis by segmenting customers on a store level so that we could explore Pernalonga’s customers general purchase behavior. After tuning on different variables, we decided to use average discount rate, average monthly spending, average monthly transactions and profit margin ratio. We used all these normalized measures to rule out some unexpected influence. For example, some customers might be new but loyal but not represented well by the raw measures.

* Average discount rate: the average percentage of discount a customer applied
* Average monthly spending: the monthly average spending by each customer
* Average monthly transactions: number of transactions by each customer each month
* Profit margin ratio: the level of margin a customer generated at a store in ratio

Results

The segmentation shows that Pernalonga has a group of Regular customers that is not as persuasive by discounts. There is also a Store-Loyal group that has the highest number of monthly transactions and a relatively high profit margin ratio. Additionally, Store-Loyal customers are particularly interested in Private-branding products. Bulk Purchasing Customers are the ones that do not go to stores often, but each time buy a large bulk of a mixture of discounted and undiscounted goods. Finally, there is a cherry picker group that buys on discount. We compared the results of this section with results from toothpaste level segmentation to get more detailed insights, which is in the next section.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features /**  **Clusters** | **Regular Customers** | **Bulk Purchasing Customers** | **Store-Loyal**  **Customers** | **Cherry Pickers** |
| Average Monthly Spending | (303.5) | Highest  (464.6) | (369.5) | Lowest  (289.4) |
| Average Monthly Transactions | (13.9) | (14.4) | Highest  (22.9) | Lowest  (13.6) |
| Profit Margin Ratio | Highest  (0.25) | (0.22) | (0.23) | Lowest  (0.20) |
| Average Discount Rate | Lowest  (0.13) | (0.15) | (0.14) | Highest  (0.19) |

*Table 6: Store-Level segmentation clusters*

**3.4 Segmentation: Toothpaste Level**

Feature Selection

For this segmentation, we only included data on toothpaste transactions. Also, instead of using average monthly transactions, we used the average monthly units of toothpaste purchased. In addition, we added a variable called colgate loyalty which is calculated by the frequency of colgate toothpaste purchased divided by the frequency of toothpaste of any brand purchased.

Results

Customers were divided into four groups where Loyal Colgate Customers are the loyal customers for Colgate and Store-Loyal Customers are the loyal customers for Private Label. Combining these results with the analysis from segmentation on the store level, we noticed that Pernalonga has a group of loyal customers, and these customers (Store-Loyal) have a 60% probability in buying a Pernalonga’s Private Label toothpaste and only a 13% probability in buying a Colgate toothpaste. Thus, we have decided not to target this group. On the other hand, High Volume Customers and Cherry Pickers both have 53 ~ 55 % of customers buying Colgate toothpaste. Thus, we can focus on these two groups where there’s a lot of room for improvements when sending personalized promotions. Loyal Colgate Customers are not the ideal ones to target because they are already very attached to Colgate and they are not so sensitive to discounts. Offering discounts to Loyal Colgate Customers might lose a large portion of our margin without gaining too much market shares.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features**  **/ Clusters** | **High Volume Customers** | **Store-Loyal**  **Customers** | **Cherry Picker - Toothpaste** | **Loyal Colgate Customers** |
| Average Monthly Units | Highest  (2.6) | (1.4) | (1.5) | Lowest  (1.4) |
| Average Monthly Spending | Highest  (5.9) | Lowest  (2.4) | (2.9) | (3.3) |
| Colgate Loyalty | (0.54) | Lowest  (0.15) | (0.56) | Highest  (0.73) |
| Profit Margin Ratio | (0.20) | Lowest  (0.12) | (0.13) | Highest  (0.28) |
| Average Discount Rate | (0.26) | Lowest  (0.11) | Highest  (0.34) | (0.18) |

*Table 7: Toothpaste Level customer clusters*

**3.5 Segmentation: Colgate Toothpaste Level**

We wanted to see if different groups/clusters of Colgate customers have different purchasing behaviors for specific Colgate products. In other words, if the clusters would purchase a specific Colgate toothpaste, we might have some insights on what items to recommend to them. Customers were distributed into three clusters. However, when we analyzed the cluster information, we noticed that the purchase pattern for all three clusters are rather similar. In all three groups, 85 ~ 87% of customers buy the pasta dentifr tradic; 11 ~ 12% of customers buy pasta dentifr branq; about 1% buy pasta dentifr medici, and about 1% buy pasta dentifr infant. Thus, we believe that there is no particular clear purchasing pattern on selecting various Colgate products that could be identified based on different Colgate customer groups.

**4. Building Similarity Model**

**4.1 Identifying complementary and substitutional goods**

Identifying Complementary Goods

One effective way to promote goods is simply recommend them to people who are buying their complementary goods or substitutions. In order to identify complementary goods, we first removed duplicates on transactions that customers buy multiple toothpaste products together since we already know they are substitutions. Then, we selected all possible pairs of products that were bought together in one transaction with a toothpaste. Then we aggregated and calculated how many times unique products were purchased with a toothpaste to find out the most purchased or the most complementary products with Colgate toothpastes.

We identified the top items that were frequently purchased with Colgate toothpastes.

Items purchased with pasta dentifr tradic :

|  |  |
| --- | --- |
| Product name | Count |
| PRIVATE LABEL\_AGUA LISA | 11,653 |
| PRIVATE LABEL\_BEB ICE TEA S/GAS | 8,373 |
| NO LABEL\_PORCO STANDARD FR | 7,866 |

*Table 8: items purchased most with pasta dentifr tradic*

Items purchased with pasta dentifr branq:

|  |  |
| --- | --- |
| Product name | Count |
| PRIVATE LABEL\_AGUA LISA | 1,611 |
| PRIVATE LABEL\_BEB ICE TEA S/GAS | 1,223 |
| NO LABEL\_PORCO STANDARD FR | 1,154 |

*Table 9: items purchased most with pasta dentifr branq*

Items purchased with pasta dentifr infant:

|  |  |
| --- | --- |
| Product name | Count |
| PRIVATE LABEL\_AGUA LISA | 290 |
| PRIVATE LABEL\_BEB ICE TEA S/GAS | 210 |
| NO LABEL\_PORCO STANDARD FR | 181 |

*Table 10: items purchased most with pasta dentifr infant*

Items purchased with pasta dentifr medici:

|  |  |
| --- | --- |
| Product name | Count |
| PRIVATE LABEL\_AGUA LISA | 123 |
| NO LABEL\_PORCO STANDARD FR | 75 |
| PRIVATE LABEL\_BEB ICE TEA S/GAS | 63 |

*Table 11: items purchased most with pasta dentifr medici*

We found that bananas are the top bought together goods for all toothpaste, but we believe that’s because of the stores’ particular layout. They might have bananas close to cashiers, which made bananas the top sold item. Thus, we removed it from our top complementary list, since it’s being bought together with all goods rather than only with toothpastes. Moreover, we can find a trend here that most of the top 3 complementary items are very similar for different subcategories of Colgate toothpastes, which is a clear pattern that could potentially help us to design our promotion plans.

Finding the most substitutable toothpaste (between brands)

In order to promote a specific subcategory of Colgate toothpaste to customers that are already buying toothpaste from another brand that’s highly similar to Colgate toothpaste we used cosine similarity to find out the similarity score for the most similar products to Colgate products. We used a 95% similarity as cut-off point which means we will recommend a Colgate toothpaste to the customer if he/she has purchased a toothpaste more than 95% similar to the Colgate toothpaste (subcategory). In addition, we took out any substitutable toothpaste that is also a Colgate subcategory to avoid competition within Colgate.

**Substitutes for Colgate Toothpastes Traditional:** (Mentioning only Top 5 here) Aquafresh Toothpaste Traditional (99.4%), Aquafresh Toothpaste White (99.1%), Aquafresh Toothpaste Medical (98.3%), Sensodyne Sensib (98.0%), Private Label 2 EM 1 (97.7%).

**Substitutes for Colgate Toothpastes White:** (Mentioning only Top 5 here) Aquafresh Toothpaste Traditional (98.9%), Aquafresh White (98.5%), Aquafresh Medical (98.3%), Private Label 2 EM 1 (97.1%), Sensodyne Sensib (97.1%).

**Substitutes for Colgate Toothpastes Medical:** Sensodyne Sensib (95.3%)

**Substitutes for Colgate Toothpastes Child:** Aquafresh Child (96.0%)

**4.2 Finding customers to target**

We plan to implement cosine similarity to find the best customers to target:

1. Easy to implement and computational efficient.
2. Suitable in this case to find similar purchasing patterns.
3. Can be used in further studies to predict sales or demand transference

In order to identify the best customers we wish to target with our promotion, we need to select a group of the most Colgate-loyal customers as our base-customers for further implementation. Then, we used cosine similarity to find out customers that are most similar to those Colgate-loyal customers as our targets.

Select base-customers

We first used the 95 percentile to find the top 5% of customers based on the Colgate loyalty scores which is calculated by frequency purchasing Colgate toothpaste divided by frequency to purchase a toothpaste. While this selection gives us about 669 customers, we wanted to narrow this group down more, because there are customers who bought Colgate for only once and account for 100% Colgate Loyal Score. Another reason we want to further narrow it down is that for customers with too few visits to store we might not have enough transaction data to determine their buying pattern, where there can be a lot of variance in buying patterns for those who only came into the store a few times. So we selected only customers purchased more than 30 transactions at Pernalonga and bought more than 39 toothpastes to narrow the group of most Colgate-loyal customers down to 29 people. These 29 people are our base-customers, where those two numbers were selected based on 95 percentile of summary statistics.

Calculate similarity score between customers and our base-customers

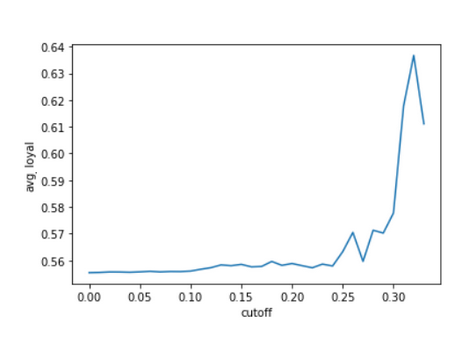
With our base-customers, we mapped each unique customer in our dataset to the 29 most loyal customers. Then, we calculated similarity scores between each customer and the 29 base-customers based on the frequencies they have purchased each product. In order to get a robust similarity score for each customer, we decided to average the 29 similarity scores of each customer. Combining similarity scores and the clusters he/she belongs to (according to the Toothpaste Level segmentation), a table with top 5 in similarity score is shown below (where we see that the Colgate Loyal Score are significantly higher than average customers):

|  |  |  |  |
| --- | --- | --- | --- |
| Customer ID | Similarity Score | Colgate Loyal Score | Cluster (he/she belongs to) |
| 30209822 | 0.350 | 0.636 | Cherry Picker - Toothpaste |
| 13889806 | 0.343 | 0.826 | Cherry Picker - Toothpaste |
| 8409701 | 0.337 | 0.556 | High Volume Customers |
| 39389624 | 0.337 | 0.333 | Cherry Picker - Toothpaste |
| 39289826 | 0.334 | 0.667 | High Volume Customers |

*Table 12:customer similarity table*

After observing the table above and combining with our knowledge on the segmentation on the Toothpaste Level, we filtered out customers in Store-Loyal Customers cluster and Colgate-Loyal Customers cluster. Store-Loyal customers are loyal to Private Label products which are not as persuasive to Colgate promotions. Colgate-Loyal customers are already Colgate loyal customers who don’t necessarily need to be promoted to. Especially when the goal is to enhance on incremental profits, promoting to those customers reduces the margin and won’t increase the sales by much.

Then, we need to decide a cut-off point for the similarity score for our customers as the final step to identify our target customers. We determined the cut-off point by observing the graph below:



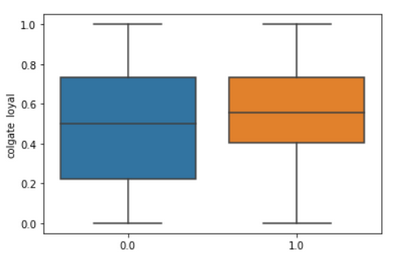
*Figure 5: Determine cutoff point and loyalty score*

We noticed a sharp Colgate loyal score drop at 0.26 and stayed roughly flat from 0.26 to 0. Thus, we chose 0.26 as the cut-off score, and we have 623 customers that we are targeting.

**4.3 Evaluation**

Higher Loyalty score

In order to assess our similarity score model, we conducted some analysis on our target customer group. We found that our target customer group has a mean Colgate loyal score of 58%, which is 16% higher than the mean Colgate Loyal Scores of all customers of 49%. Additionally, the variance of our target customer group is much smaller which gives a more steady behavioral evidence for our model performance. This pattern is also seen in the box plot below comparing all customers and the target customer group.



*Figure 6: Boxplot on targeted customer loyalty and all customer loyalty*

Projected sales increase

Since most customers only buy 1 toothpaste for a certain period, it’s hard to determine the price elasticities and how discounts would impact sales. So we applied a simple rule to project it. As we discussed, there is a 16% increase in Colgate Loyal Score, so we assume here there will be 16% more sales redistributed to Colgate from other brands.

By using this rule, we calculated a $ 28871.32 projected sales and $ 17043.38 projected paid revenue compared to $ 24829.62 sales and $ 16286.08 paid revenue in the past, where there’s a $ 757.3 potentially realized increase in paid revenue.

Evaluation after Implementation

Evaluating the promotional plan with A/B testing would be more preferable than project revenue. We plan to select 10 stores to perform the testing, consisting of 5 pairs of stores with similar demographic characteristics. To begin with the process, we should first perform an A/A testing for each pair. Then we implement the promotional plan to 5 of the stores to see whether it impacts sales in an expected way.

**5. Recommendation and Implementation**

Based on our analysis, we believe utilizing the segmentations and similarity scores based purchase history could select our targeting group in a refined way and therefore boost sales for Colgate’s toothpaste products:

1. Based on the Store-level segmentation and Toothpaste-level segmentation, we notice a large segment of customers are store-loyal and purchase the store’s private label most of the time. And the Colgate-loyal group uses much less discounts than average customers. So, discounts are not as persuasive for them. Moreover, targeting Colgate-loyal groups might hurt our margin more than the benefits from sales increase.
2. Utilizing the complementary and substitutional lists, we can target our customers with more diversified methods other than direct mails and in-store discounts. Such as placing advertisements on or even directly adding it next to the shelf of Colgate’s toothpastes top complementary and substitutional products. Therefore when customers browse on those products they might look into Colgate’s toothpastes as well.
3. By implementing a Cosine Similarity approach, we can realistically determine how likely a customer is to buy Colgate toothpaste, letting us to more precisely personalize promotions to capture the incremental margin, where we don’t have to target all customers, instead we target only the customers with high similarity scores and not already a Colgate-loyal customer. Furthermore, we hope to help Pernalonga use these insights to appropriately design a pricing model that we can then determine the level of discounts of Colgate’s toothpaste offerings to even more precisely capture those incremental margins, and therefore maximize profit by optimizing the discount amounts at an individual level.

We recommend using various methods to do the promotions including coupons, direct mails, in-store discount and on-shelf advertisements that allow us to maximize the exposure of Colgate toothpaste for the best campaign's results. We determined a 10% discount rate for the introduction of this promotional plan, but if Colgate decides to continue the promotion in the future, we recommend using a variable discount rate at the individual level that takes each customer's purchase history and similarity scores into account. For our current approach, based on historical transactional data, incremental revenue and discounts redeemed were calculated at the toothpaste sub-category and customer segment levels. This will enable us to calculate the lowest discount rate that will keep that customer buying Colgate toothpaste and maximize the incremental profit.