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………………………………………………………………………3
   2. Business Context……………..………………………………………………...3
   3. Data Cleaning and Exploration………………………………………………..4
2. Product Categories and Products…..………………………………………………...5
3. Stores…….…………………………………………………………………………........7
4. Substitutes and Complements………………………………………………………...8
5. Seasonality……….……………………………………………………………………...9
6. Modeling………………………………………………………………………………..10
7. Recommendations and Conclusion………………………………………………….11

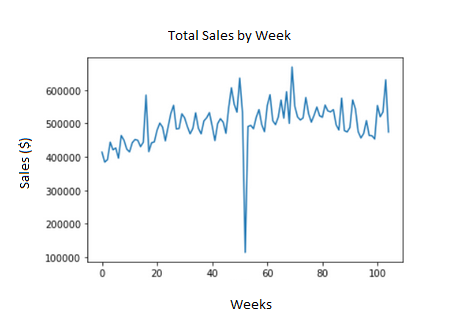
**Executive Summary**

We performed a price optimization analysis for Pernalonga based on 2016 and 2017 sales data. The report includes suggested price changes for 100 products in two categories in 10 stores starting on the week of April 13-19, 2021. The two categories for which we have recommended price changes are Beer with Alcohol and Standard Wines, and these changes will be implemented in the following stores: 320, 335, 342, 343, 345, 346, 347, 349, 395 and 588. On top of a detailed explanation of our methodology and results, recommendations and next steps are also provided at the end of this report.

**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. While Pernalonga still has its seasonal offers, based on the fluctuations in the sales plot by week, we can see that Pernalonga makes most of its revenue by constantly offering promotions.. Pernalonga recognizes, as a data-driven business, that a potential opportunity in optimizing list prices is raised to boost sales contribution. Pernalonga plans to adjust the list price of a select number of its products starting on the week of April 13-19, 2021, in order to increase overall sales and profitability. Using historical Pernalonga transactional data, we hope to develop a pricing model that recommends price adjustments for 100 items across two product categories, optimizing overall profit across ten stores.

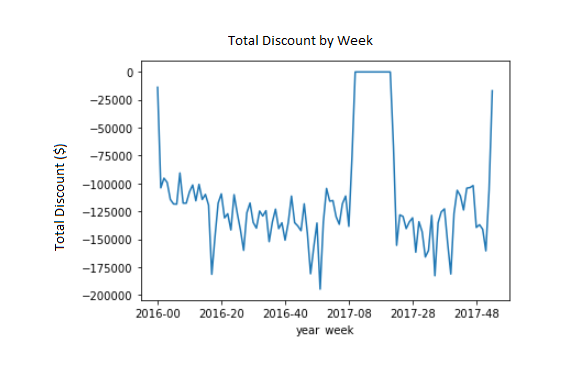


(Week 53 contains only two days of data, which presents a “deep drop” of sales)

**1.2 Business Context**

We wanted to understand Pernalonga's historical pricing and customer behavioral responses to it in order to generate the best possible price recommendation. We assumed Pernalonga will have similar patterns of promotions and seasonality impacts during the 2021 target implementation period as 2017, so this will serve as a good benchmark for optimizing our profit for this year's target period. And by controlling those factors, we can see how list price changes impact the consumer demand level.

Another important thing to note is that no items were sold at a discount during 12 weeks in 2017 - from week 10 to week 21. This range includes the second and third weeks of April 2017. Pernalonga appears to employ "everyday low price" strategies mainly based on the fluctuations in the discount plot below, however, it still provides seasonal promotions at holidays such as Christmas. We also looked at any changes in demand, as this could be a better indicator of a profit increase over the two weeks. However, the data revealed a similar story, indicating that there was no significant change. These price and demand changes are minor in the grand scheme of things. Although it varies by product, we can conclude that between the second and third weeks of April in 2017, Pernalonga did not make any significant price changes. Therefore, Pernalonga has the opportunity to improve its pricing strategy in order to increase overall sales and revenue.



**1.3 Data Cleaning and Exploration**

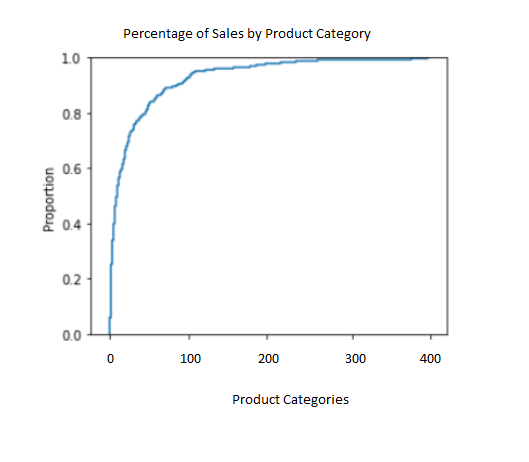
We were provided two files:

* 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.

There are many products in a typical supermarket to which customers may be very subjective to seasonality and other factors, such as many price changes for these products are usually markdowns due to the item no longer being as fresh as it once was. As a result, all products that are considered “fresh” have been excluded, including fresh meats, bakery items, and others. To simplify the manipulation process, we assumed all “KG” goods are likely to be fresh and thus we removed them from our analysis. Any products that have never had a price change are not considered. In addition, products that only have a few price points are not considered because calculating elasticity can be less meaningful in this case due to the lack of enough instances. 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.

We investigated the distribution of sales among the resulting product categories once our data was clean and ready for analysis. We discovered that 83 product categories account for roughly 75% of total sales. These product categories account for approximately 20% of the total product categories, which is in line with the Pareto principle, which states that 80% of sales should come from 20% of products.



**2. Product Categories, Stories and Products**

We select our target products which we will perform the price change on, we make our choice for stores and categories based on a systematic approach of setting up conditional constraints:

1. We consider whether an object is generating a lot of sales.
2. We select categories that contain a larger number of products.
3. We pick products that are sold at more stores.
4. We choose products with larger price variance.

We filter data and only include the ones that fulfill all constraints above. By doing this, we can select the best products for our analysis with relatively high amount of sales and price variance while large amounts of data are well distributed among the stores. Thus, there is a lower possibility of instances with missing values or price elasticity, giving us enough basis for the analysis.

As a result, The 10 stores chosen for this analysis are 320, 335, 342, 343, 345, 346, 347, 349, 395 and 588, and the top two categories are Standard Wine and Beer with Alcohol. Besides the reasoning we already explained, the two categories are more similar to each other, where we found that similar categories are more likely to be sold at the same store.

|  |  |
| --- | --- |
| StoreID | Total Revenue Across Selected 100 Products ($) |
| 320 | 42,016.11 |
| 335 | 46,388.14 |
| 342 | 90,887.08 |
| 343 | 71,779.74 |
| 345 | 83,831.58 |
| 346 | 69,889.59 |
| 347 | 64,542.18 |
| 349 | 42,016.11 |
| 395 | 38,996.88 |
| 588 | 64,869.62 |

We were able to ensure that these product categories would drive revenue and profit for Pernalonga by using a combination of product elasticities and sales to determine our product categories. Both product categories have a wide range of products, with 83 different products in Beer with Alcohol and 61 in Standard Wines over the last two years. This suggests that Pernalonga's product offerings in these two categories appeal to a wide range of customer preferences and tastes, and that these products are in high demand.

Pernalonga has sold 853,630 units of Beer with Alcohol at an average list price of $8.39 over the last two years, and 267,531 units of Standard Wines at an average list price of $2.88. Both of these prices are low, but Beer with Alcohol has a much wider range, with prices ranging from $0.22 to $810. In contrast, the price range for Standard Wines is from $0.39 to $175.23. Both of these categories have historically low discounts, with an average discount of $2.93 for Beer with Alcohol and $0.47 for Standard Wines, so keep that in mind as we move forward with our pricing recommendations. Finally, we sorted all products within those categories to obtain the top 100 products based on each product's elasticity once we determined our two product categories, Beer with Alcohol and Standard Wines. We were able to incorporate different product aspects into our response function and calculate each product's optimal price using these 100 products from Beer with Alcohol and Standard Wines.

**3. Model Attributes**

We created a pricing response model that incorporated various attributes for each product, such as list price, promoted price, product substitute and complement prices, and seasonality. In order to properly analyze Pernalonga's historical data and recommend price changes for our 100 products, we obtained the noted attributes for each product on a weekly basis, resulting in one observation for each week in our dataset, along with the corresponding price, complement, substitute, and seasonality details for each product. The following information was collected for each product using the methods explained:

**Demand** - Demand records the total number of units sold each product in the corresponding week. In our model, this is used as the response/target variable.

**Week** - From the first week of our data to the last week, we created a variable to indicate the week of each transaction. This variable was used to roll up all of our variables to the weekly level. We also combined week and year to investigate the week from each year separately, where we found week 53 contains only two days and removed it to ensure data consistency.

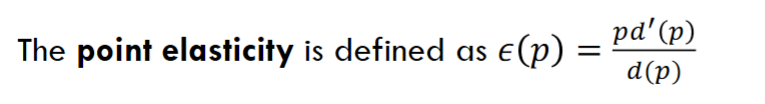
**List Price** - We calculate List Price by taking an average of all unit prices for the same product at a single store in the same week.

**Promoted Price** - We evaluate Promoted Price on a discount level, calculating the weekly total discount amount divided by total weekly transaction sale amount for each product.

**Seasonality** - To account for seasonality in the data, we recorded the percentage that the sales is deviating from average sales across two years on a weekly basis.

**Substitutes and Complements** - The price of complementary goods and substitutional goods is explained in the next section.

As explained earlier, the two product categories we choose have large price variances where we see a significant relationship between prices and demand. To calculate elasticity, we had to make the assumption that products have a constant elasticity. The elasticity of a product can vary a lot from store to store and perform very differently on a store-level or market-level. The following formula was used to calculate point elasticity for each product in this case.



Here we used a slightly different approach, where we used the average list price (p), average demand (d(p)), and performed log transformation for each product to calculate its average price elasticity during the period. Moreover, we use elastic constant linear regression to calculate price elasticity. Also, We used data from both 2016 and 2017 to calculate all of these figures.

Last but not the least, we included the cost of goods sold for each product on the store-level to calculate revenue and profit for our recommendations, which would be very valuable while optimizing the profit and giving the best price changes.

**4. Substitutes and Complements**

The prices and availability of a product's substitutes and complements have an impact on a customer's decision to buy or not buy that product. Price increases in complementary products can potentially discourage customers from buying the target goods. Meanwhile, the decrease in sales of substitutional goods will increase the likelihood customers buy the target product, implying a negative relationship between target products and substitute sales. Tea and coffee are two examples of products that are unlikely to be purchased together. Complements, on the other hand, are items that are frequently purchased together across all categories. The sales of a target product and its complement usually have a positive relationship. Burger and burger buns are an example, as most people prefer to eat both items together.

We needed to account for the effects of substitutes and complements in order to better predict demand for our 100 targeted products. We first identified complementary goods using association rules by dividing the frequency of co-purchases divided by the frequency of target products, which produce the probability of conditional probability in terms of confidence level. By ranking the confidence level, we can figure out the top complementary goods for a particular product. However, we didn’t use them individually because some complementary goods are not sold at many stores, which will produce empty values. Thus, we used the average price of top 5 complementary products instead.

For substitutional products, we used Pearson correlation based on the number of co-purchases with complementary goods which we counted for in the last step. The methodology is if the products share similar co-purchase patterns, then should be more substitutable to each other. For the same reason, we used the average price of the top 3 substitutable to avoid empty values.

**5. Modeling**

We used the logit price response function to solve our profit optimization problem as we progressed with our modeling approach.

1. Response function

The average weekly demand is our Y-variable, indicating the product's weekly demand. We created a dataset containing various attributes, such as price, complement, substitute, and seasonality details for each product at each store on a weekly basis. To solve the logit price response for this retail price optimization problem, we did first a logit transformation to weekly average demand, using weekly average demand dividing by the subtraction of theoretically maximum demand and weekly average demand. We then used a linear regression model to solve the response function.

1. Model interpretation

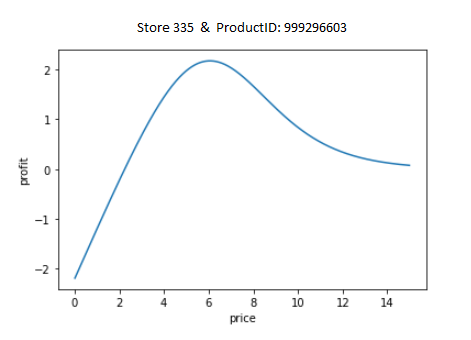
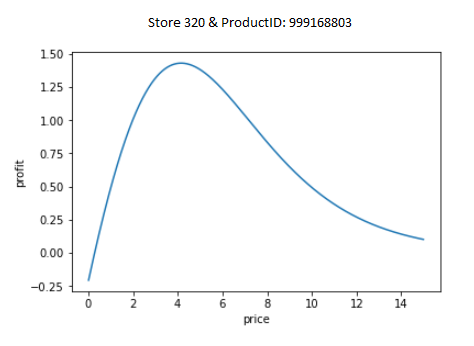
We modeled three different product and store combinations, the coefficients made sense. For all three models, the coefficients for price are negative, and they indicate that the higher the price, the lower the demand. Also, discounts all have negative coefficients, where discounts are all negative numbers, meaning that the higher the discount the higher the demand is.

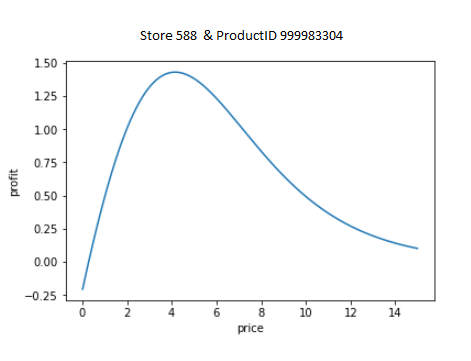
|  |  |  |  |
| --- | --- | --- | --- |
| Store ID | Product ID | Price | Coefficient |
| 320 | 999168803 | 4.15 | -0.41 |
| 335 | 999296603 | 6.06 | -0.61 |
| 588 | 999983304 | 6.87 | -0.26 |

1. Model application and profit optimization

We apply the model with profit curves, where we calculate projected profit based on our response function and calculated coefficients for each product at each store, for price $0.01 - $15, ticking at $0.01 (price cannot be lower than 0, and the average price is around 4 or 5, so the value too large wouldn’t make any sense).

Based on the profit curve we produced we found there are existing maximized profits for each bell-shaped curve. For product 999168803 at store 320, the profit is maximized at $4.15. The optimal profit is the calculated by pmax \* exp(d)/(dmax + exp(d)), which is the inverse of logit transformation multiplying the price. The result for the other two sample cases are shown in the other two bell-shape curves and can also be seen in the result table.





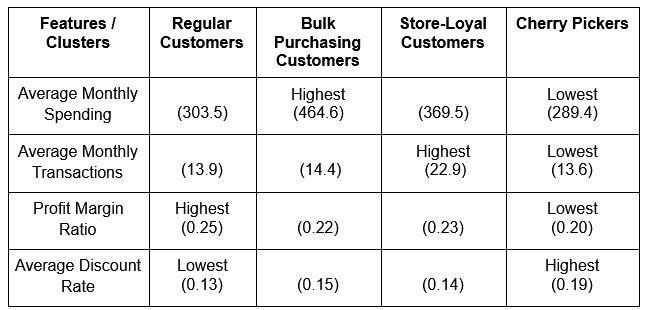
**6. How the optimal prices affects customers from different segments**

In order to understand how this price optimization would affect customers, we calculated the updated sales quantity, revenue and profit based on the coefficients of each model from the last section, and adjusted everything according to the difference between the optimal price level and the original prices. The methodology is to calculate the changes in demand by the coefficient (price elasticity) from the model and multiply it by the changes in prices. And then calculate the revenue and profit by multiplying the new prices and subtracting the costs.

Based on the result, we see that Cluster 2 and 3 are less affected by the price changes. Cluster 3 which is the store-loyal customer who doesn't care much about the price changes and discounts, and Cluster 2 is the regular customers who are also not so sensitive to those promotions. While the Cluster 0 is the cherry picker who chases for price changes and promotions all the time, getting affected the most. It also reflects that Cluster 2 and 3 might buy less alcohol, although we can not confirm that due to the small sample sizes.

In addition, from the updated table, we observe that for all clusters our optimal prices increase the revenue level and profit level in some degrees, and for cluster 1, prod\_id 999168803 and store 320, the profit is increased the most by $110.52.





**6. Recommendations and Conclusion**

Although we only optimize on three products, the approach we designed is able to select an optimal list price for each product in each store utilizing the price response function above. If the dataset is large enough, we can even account for different customer segmentation or various time periods and perform more detailed optimization to capture more incremental profits. The products whose prices did not change were already being sold at their predicted optimal prices. However, during the target week, only a subset of products were sold in each of the selected stores. Below is a list of a subset of the products in each store, along with their optimized list prices.

For evaluation, we could implement at 5 of those 10 stores for the targeted products, and have a control set consisting of the other 5 stores. Since they sell similar assortments, we can first do a A/A testing first on them for a month, and then perform a A/B testing to test for the significance of the average profit changes.