Pricing Optimization Analysis for Pernalonga

Harshit Arora, Patrick He, Halle Steinberg, Sean Yu

Team 1

MKT 680 Marketing Analytics

Emory University

April 6th, 2020

Executive Summary

In this report, we have detailed the results of a price optimization analysis for Pernalonga supermarket chain. The details in this report contain recommended price changes for 100 products across 2 categories in 10 stores for the week of April 13-19, 2020. The 2 categories we have recommended price changes for are **“Fine Wines”** and **“Fine Wafers”**, and these changes will be made in the following stores: **342, 588, 341, 349, 346, 395, 345, 344, 343,** and **994**.

Based on our recommended changes, Pernalonga should expect approximately the following results:

- A total increase in revenue of **$1,407**
- A total increase in demand of **1,314 units**
- An average change in price of **$0.85 per unit for Fine Wines and $0.11 for Fine Wafers**

The details of each product and its optimized price are provided in the attached Excel file. In addition to a detailed explanation of our methodology and results, recommendations and next steps are also provided at the end of this report.

Background

Pernalonga is a leading supermarket chain with over 400 stores in Lunitunia that offers a vast variety of over 10 thousand products among more than 400 categories. Its business is currently heavily reliant on promotions, both in-store and through third-party suppliers, with nearly 30% of sales being driven by promotions. Due to its heavy reliance on promotions, Pernalonga has two methods of making sales - through a product’s list price or through its promoted price. As a data-driven company, Pernalonga understands that list prices are increased to drive contribution and lowered to drive sales. Pernalonga is looking to make list price changes to a select number of its products during the week of April 13-19, 2020, in order to maximize its overall revenue and profitability. Using historical Pernalonga transactional data, our goal is to create a pricing model that suggests price changes for 100 products across 2 product categories and will boost overall revenue across 10 stores.

Business Context

In order to generate the best recommended price changes possible, we first wanted to understand Pernalonga’s historical pricing and promotional behavior. We understand that Pernalonga offered similar promotions during the target week in 2017, so this will act as a good comparison for that week this year. Since schedules may be slightly different this year, as holidays move around and buying patterns may change slightly, we took a look at patterns and trends in both the second week of April (4/6-4/12) and the third week of April (4/13-4/19) in 2017, which is our target week for 2020. By analyzing these two weeks, we can see how prices change from one week to the next, indicating whether one week may have shown a significant price change from the other.

The third week of April 2017 saw 27,151 transactions, and the second week of April saw 28,504. To understand how prices might have changed during these weeks, we calculated the average list price for each product sold during each of these two weeks and found the percent change in the prices. On average, product prices across the two weeks increased by 0.04%, with some decreasing as much as 0.60% and some increasing as much as 2.01%. This implies that some products may have been on sale in the second week of April 2017, while others may have been on sale in the third week. Another interesting point to make is that all products sold in the second and third week of April 2017 were sold at list price, meaning that no products were sold at a discount. All of this indicates that Pernalonga is not a typical “hi-lo” retailer, as most of its products’ prices change week-to-week while others remain the same. Pernalonga appears to utilize both “hi-lo” and “everyday low price” strategies, making it somewhat of a hybrid between the two types of retailers.

We also examined any changes in demand, as this may be more indicative of a resulting increase in profit across the two weeks. However, we found a similar story in the data, as it showed that, on average, there was no significant change. On average, we see a percent increase in demand of about 0.52%, with a maximum change of 7.30% and a minimum change of -0.99%.

As a whole, these changes in both price and demand are not significant. Although it depends on the product, we can generally conclude that Pernalonga did not make any significant changes in price between the second week and the third week of April in 2017. That said, there is an opportunity for Pernalonga to optimize its pricing strategy in order to maximize its overall sales and revenue potential.

Data Preparation and Exploration

To begin our analysis, we first merged the two datasets to create a file that contains all product and transaction information. This gave us a dataset containing nearly 30 million observations across 421 stores and 429 product categories. In total, our data consists of 10,767 unique products. We know from our previous analysis that there are some anomalies in the data that must be accounted for. One of the biggest errors in the data comes from the fact that there are only 753 unique transaction ID’s assigned to the data. To account for the errors in these ID’s, we again used a combination of customer ID, store ID, and transaction date to correctly define the transaction ID’s, which resulted in 2.83 million unique transactions with a varying number of products purchased across all consumers.

In a typical grocery store or supermarket, there are many products to which customers may not react as well to changes in price. For this reason, we have excluded all products that are considered "fresh" - this includes all produce, fresh meats, and bakery items, among others. This makes sense because most price adjustments for these types of products are often markdown prices due to the fact that the item is no longer as fresh as before. This type of price adjustment is not the same as that which we are looking to optimize, so we have disregarded these products. We also excluded any products that had never had any price changes before. Any products that only had one price point in our data were removed because, in this case, calculating elasticity can be very difficult due to lack of price changes. We have also excluded "BAGS" since a large portion of transactions include this item, and we don't want it to affect our analysis of sales of products in those transactions.

Once our data was clean and ready for analysis, we explored the distribution of sales among the resulting product categories. We could see in our resulting data that 72 of the product categories make up about 80% of sales. These product categories make up about 22% of the resulting product categories, which is consistent with the Pareto principle that states that 80% of sales should come from 20% of products.

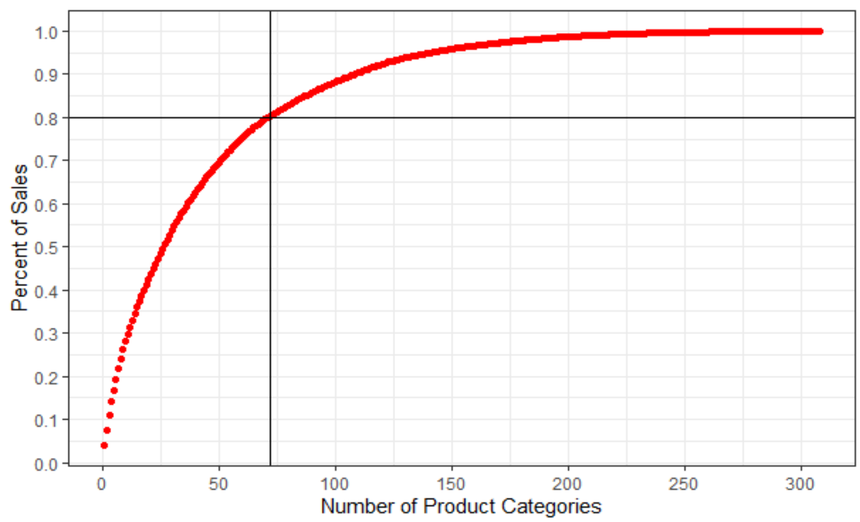

Figure 1: Percent of Sales of Each Product Category

Choosing Products and Product Categories

Moving forward, it was important that any product we selected to target would be able to adapt its demand well to changes in price. If we selected products that do not see positive changes in demand when price changes, we will very likely not make additional revenue. That being said, we wanted to choose highly elastic products, as these should be the products that are most responsive to changes. If we change the price of these products, demand should also change significantly. These are products that consumers are more price-sensitive towards, so a small decrease in these products will likely yield a significant change in demand. The same is also the case in the opposite direction - if we make too big of an increase in price, demand could decrease significantly. This is something we needed to be wary of as well. One assumption we had to make to calculate elasticity is that products have a constant elasticity. Generally, product elasticity may vary slightly over longer periods of time, and also possibly from store-to-store. In this case, point elasticity $\boldsymbol{\epsilon}$ for each product was calculated using the following formula:

$$\frac{pd'(p)}{d(p)} = \epsilon$$

To properly use this formula and calculate each product's elasticity, we needed each product's average list price (*p*), average demand (*d(p)*), and change in demand (*d'(p)*). For all of these values, we used all data from both 2016 and 2017. For example, each product's average list price was calculated as an average over all list prices the product has ever sold at, and so on for the other two metrics. We used each product's calculated elasticity to determine which product categories contained the most highly elastic products, ensuring that both categories were among Pernalonga's top selling categories and had at least a total of 100 elastic products. Finally, one key business requirement is to make recommended price changes for our two selected categories across 10 stores. Given this, we wanted to confirm that both of our chosen categories were being sold across a minimum of 10 stores.

Based on our determination of each product's price elasticity and ensuring each product is sold in a minimum of 10 stores, we were able to find Pernalonga's top 5 product categories based on the number of highly elastic products in each category, as seen below. Here, we defined "elastic" as those products with a calculated price elasticity of -0.7 or less. Realistically, we would classify a product as "elastic" if its elasticity is less than -1 and "inelastic" if it is greater than -1. However, to ensure we have enough products to choose from in each of the selected product categories, we used an elasticity cutoff of -0.7 instead.

Product Category	Number of Elastic Products
Fine Wines	67
Fine Wafers	53
Dry Food Animals	45
Ice Cream	31
Olive Oil	25

Figure 2: Top Product Categories by Count of Elastic Products

To choose our top 2 categories from the top 5 above, we calculated the sum of all historical sales in each product category and chose the two with the highest total sales in 2017. Our top two categories ended up being "Fine Wines" and "Fine Wafers". We can see from the chart below that both of these categories fall in the top 10 for sales and are the top 2 categories based on those included in the table above. For the average wine drinker, when debating between two wines of the same type, he or she will likely buy whatever is cheapest, and the same is usually the case for wafers or crackers that will go with the wine. That being said, it makes sense that these two product categories would be highly elastic.

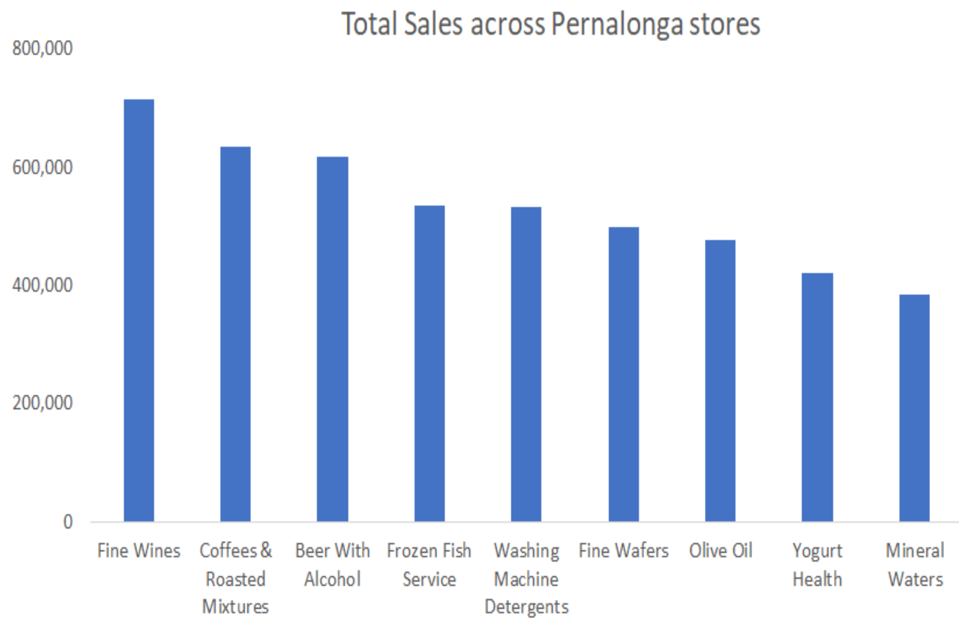

Figure 3: Total Sales ($) by Product Category Across All Stores

By using a combination of product elasticities and sales to determine our product categories, we were able to ensure that these product categories would drive revenue and profit for Pernalonga. Both product categories have a wide variety of product offerings, with "Fine Wines" having 393 different products over the past two years and "Fine Wafers" having 282. This likely implies that Pernalonga's product offerings in these two categories appeal to a wide range of preferences and tastes of customers and that there is high demand for these products. Over the past two years, Pernalonga has sold 557,945 units of "Fine Wine" at an average list price of $3.34 and 844,768 units of "Fine Wafers at an average list price of $1.46. Both of these prices are relatively low, but "Fine Wines" has a much wider range, with a minimum of $0.49 and a maximum of $19.98. On the other hand, "Fine Wafers" only ranges from $0.14 to $9.99, indicating that there is a much wider variety of wine types and qualities than there is for wafers. Both of these categories have had historically low discounts, with "Fine Wines" having an average discount amount of $1.03 and "Fine Wafers" of $0.31, so this is something to keep in mind as we move forward with our pricing suggestions.

Once we determined our 2 product categories, “Fine Wines” and “Fine Wafers”, we sorted all products within those categories to obtain the top 100 products based on each product’s elasticity. Using these 100 products from “Fine Wines” and “Fine Wafers”, we were able to incorporate different product aspects into our response function and calculate each product’s optimal price.

Choosing Stores

In order to maximize revenue from this week-long campaign, it was important that the 10 selected stores have the highest potential of generating incremental revenue. To achieve this, we shortlisted a subset of 20 stores that had the highest combined sales across our 100 selected products.

Historically, not all of our targeted 100 products were sold across all stores in the week of April 13th, 2017. By knowing that the timing and temporary price change levels are planned to be the same for corresponding weeks in 2017 and 2020, it was possible for us to identify the stores that can drive the highest incremental revenue. Through our pricing model predictions, we were able to calculate expected incremental revenue for each of the 20 stores in focus. Thus, our 10 selected stores not only had high historical sales, but they also have the potential to drive the highest incremental revenue.

Store ID	Total Revenue Across Selected 100 Products ($)
349	$24,628
346	$22,456
343	$22,051
342	$19,986
988	$16,943

Figure 4: Top Stores by Total Revenue of 100 Target Products

The 10 stores chosen for this analysis (in order of expected incremental revenue) were 588, 395, 349, 346, 345, 344, 343, 994, 342, 341

Product/Week-Level Data Creation

In order to properly analyze Pernalonga’s historical data and recommend price changes for our 100 products, we created a pricing model that incorporated various attributes for each product, such as list price, promoted price, product substitute and complement prices, and seasonality. For each product, we obtained the noted attributes at a weekly level, so each product would have one observation for each week in our dataset, along with the corresponding price, complement, substitute, and seasonality details. For each product, the following data was collected for each product using the following approaches:

- **Week** - We created a variable to indicate the week of each transaction, starting with the first week of our data until the last week. We used this variable to roll up our variables to the weekly level.
- **List Price** - Collected using each product's weekly average unit price from the original data. Average is calculated using all unit prices across all stores for the corresponding product in the corresponding week.
- **Promoted Price** - Collected using each product's weekly average transaction sale amount per unit quantity. Average is calculated using all sale amounts across stores for the corresponding product in the corresponding week.
- **Demand** - This is the total quantity sold for each product in the corresponding week. This is used as the response in our model.
- **Seasonality** - To account for seasonality in the data, each week was denoted as either being a holiday week or not. A variable that accounted for the month (1-12) in which the transaction occurred was also used as a categorical variable.

Finding Substitutes and Complements

A customer's decision of whether or not to make a purchase of a product is also impacted by the prices and availability of that product's substitutes and complements. Substitutes are defined as items that are rarely bought together. That said, the more often a substitute item is sold, the less the target product is likely being sold, indicating a negative relationship between the sales of a target product and its substitute. Examples include Coca-Cola and Pepsi, as these two items are most likely not purchased together. On the other hand, complements are items that are commonly purchased together across all categories. There is typically a positive relationship between the sales of a target product and its complement. Examples include hot dogs and buns, as most people usually choose to eat both items together.

To better predict the demand for our 100 targeted products, we needed to ensure we were accounting for the effects of substitutes and complements. We first determined a substitute and complement for each of our 100 products by performing a basket analysis on the transactions from our target stores to detect association rules (support, confidence, and lift) between products that were purchased in the same transaction. We understand that realistically each target product could have more than one substitute and one complement, but our model is built around the one product that is the “best” substitute or complement based on our computations below.

Using the “arules” package in R, we applied the APRIORI algorithm on the selected transactions to determine the association rules in co-purchased products. As a minimum threshold, an association event’s occurrence (two products sold in a single transaction) must at least be twice.

In order to better understand how association rules were used to find substitutes and complements, consider two products A and B. A lift value greater than 1 means that product A is likely to be bought if product B is bought, while a value less than 1 means that product A is unlikely to be bought if product B is bought. So, the higher the lift, the higher the chance of A and B being bought together[1]. In this case, we use 1 as a threshold to distinguish substitutes and complements, and the associated item with a lift above 1 is considered as the complement of the target product. While selecting the final substitute or complement for a product, we preferred the one with the high confidence (precision) and count over ones with high support (recall). Based on these rules, we selected complements and substitutes for our target products, and the following table shows a few examples.

Target Product Category	Target Product	Substitute Product Category	Substitute Product	Complement Product Category	Complement Product
FINE WINES	999653192	SUGAR	999356553	FROZEN BREAD	999305477
FINE WINES	999327358	MINERAL WATERS	999401500	OLIVE OIL	999345410
FINE WAFERS	999254355	FRESH UHT MILK	999401500	SUGAR	999356553
FINE WAFERS	999337215	STANDARD WINES	999356553	MINERAL WATER	999401500

Figure 5: Sample of Target Products and Corresponding Substitutes and Complements

[1]Source: https://www.dataminingapps.com/2017/04/what-is-the-lift-value-in-association-rule-mining/

Accounting for Seasonality

As with most consumer goods, there is a seasonal component that tends to impact the sales of products in focus. We saw that sales tend to vary across months and tend to get affected during holidays as well. To account for these factors, we introduced a month variable in our model which acted as a categorical variable to control for seasonality within the data. Further, while holidays tend to be on a specific day, their effect is often seen a few days before or after as well. To account for this, we introduced a holiday identifier (Yes vs. No) depending on whether the week had a holiday or not.

Modeling Approach

Moving forward with our modeling approach, we used the Logit Price Response Function to calculate predicted demand. In general, every price response function follows this formula structure:

$$\mathsf{T}(d_{it}) = \alpha + \beta_i \mathsf{G}(p_{it}) + \sum_l \sigma_{il} \mathsf{H}_l(s_{ilt}) + \sum_m \chi_{im} \mathsf{J}_m(k_{imt}) + \sum_n \phi_{in} \mathsf{L}_n(v_{int}) + \sum_r \eta_{ir} \mathsf{Q}_r(u_{irt})$$

where

d_{it} is the historical demand for item i in time period t

p_{it} is the historical price for item i in time period t

s_{ilt} is the historical price for item i's substitute item l in time period t

k_{imt} is the historical price for item i's complement item m in time period t

v_{int} is the condition on the controllable factor n with respect to item i in time period t

u_{irt} is the condition on the uncontrollable factor r with respect to item i in time period t

In our case, the Logit Price Response Function was chosen due to its many advantages in demand prediction. We know that it can easily be solved as a linear regression problem, and it is best used when optimal prices are needed. The function makes reasonable predictions, even for products with extreme prices. Since Pernalonga sells products across a wide variety of price points, this allows for flexibility in choosing our products for which we will make price changes.

Based on the formula above, $T(d_{it})$ is our Y-variable indicating the product's weekly demand. In addition to each product's weekly demand, we created a dataset containing various attributes for each product in each week that we were then able to put into a linear regression model and predict demand. The model used the following independent variables, corresponding to the right-hand side parameters in the formula above:

- **Target Product List Price** - weekly product average
- **Target Product Promoted Price** - weekly product average
- **Complement Product List Price** - weekly product average
- **Complement Product Promoted Price** - weekly product average
- **Substitute Product List Price** - weekly product average
- **Substitute Product Promoted Price** - weekly product average
- **Seasonality Indicator** - Holiday Week Yes/No

One thing to note is that not every one of our 100 target products is sold in every store in every single week. Therefore, we had to impute the weeks for any product-week combination that was missing any sales information. To do this, we used the average price for the corresponding store in that week, but if there has never been a sale at that store, the price was simply set to 0.

We chose to use a linear regression model because our prediction task is numeric (quantity demanded), and it is a parameterized model that will give us coefficients for each independent variable. Linear regression models are also among some of the simplest to understand and interpret, so for simplicity, we decided it was the best option. Once we trained the linear regression model on our data, we received an output of coefficients to be used in determining each product's optimal price.

Our ultimate goal was to predict sales for the week of April 13-19, 2020, and to do this we used historical details from the same week in 2017 combined with our model coefficients to predict demand by optimizing price. Since our goal is price optimization, we found the optimal list price for each product in each store based on its maximum possible demand. To do this, we selected a range of list prices for each of our 100 products in each store. Using our coefficients derived from our linear regression model, we predicted demand based on each list price in the range. The range for each product was generated using the historical maximum and minimum list price that each product was ever sold at in the corresponding store, with steps of 0.05 in between. For example, if a certain "Fine Wine" product in Store 1 has historically sold at a minimum of $14 and a maximum of $15, there would be steps of $14, $14.05, $14.10, and so on, all the way up to $15.

The list and promoted prices for complements and substitutes were determined as the average from the same week in 2017 for each product-store combination, and the promoted price for the target product was also determined as the average for each product-store combination from the same week in 2017. We then used all of these historical product details, each product's possible list price step, and the coefficients from the linear regression model to calculate each possible product price's predicted demand, in each store. The demand output from the model was then multiplied by the given list price to calculate the total predicted sales for that product-store combination. We then chose the optimal list price for each product-store combination based on which one generated the highest total sales.

Price Change Recommendations and Conclusion

Upon completion of predicting demand using our price response model, we were able to select an optimal list price for each product in each store. We see that, on average, a product's list price was changed by $0.81. For our target week, of the *theoretically possible* 1,000 possible price changes we could have made (100 products in each of our 10 stores), 64.3% of the products incurred a price change, and 35.7% did not. The products that did not see a price change were already being sold at their predicted optimal prices. That said, only a subset of products was sold at every selected store in the target week. A subset of the products and their corresponding optimized list price in each store is detailed below.

Store ID	Product ID	Past List Price (week of April 13, 2017)	Recommended List Price	% Change
588	999979366	1.69	1.59	5.9%
395	999979366	1.69	1.59	5.9%
349	999979366	1.69	1.69	0%
588	999329006	5.29	4.99	5.6%
343	999329006	5.29	4.99	5.6%

Figure 6: Example of Products and Recommended Price Changes

Across our 10 stores, we see a total increase in demand of 1,314 units and a total increase in revenue of $1,407 for the week of April 13-19, 2020. We can see that store 994 saw the biggest percent increase in both revenue and quantity demanded, but previously, it was not selling much to begin with. We also see that store 341 saw an increase in revenue even though it saw an increase in demand. This could be due to the fact that it has a wealthier customer base who is used to paying higher prices for premium goods. If we decrease the list price too much for these customers, they may think the quality of the products has decreased as well. This store was likely already getting these wealthier customers in the door and buying their products, so they may see a lower total revenue for this week compared to others.

Store ID	Change in Revenue ($)	% Change in Revenue	Change in Quantity Demanded (units)	% Change in Quantity Demanded	Change in Profit ($) *(Assuming 5% profit on old shelf price)*
342	$46	2.86%	138	25.00%	1.38
588	$258	48.70%	195	95.59%	7.74
341	-$80	-2.83%	70	7.27%	-2.4
349	$181	111.14%	167	231.94%	5.43
346	$169	17.88%	162	32.73%	5.07
395	$256	29.14%	163	54.88%	7.68
345	$147	10.91%	154	29.84%	4.41
344	$149	17.69%	109	28.17%	4.47
343	$181	45.44%	136	76.84%	5.43
994	$100	2,616.67%	20	333.33%	3

Figure 7: Our 10 Stores and Expected Change in Revenue and Demand

We can conclude from our analysis that finding the correct products and store combinations for which to optimize prices can significantly increase revenue and profitability for

Pernalonga. In the future, Pernalonga should incorporate a similar pricing strategy to ensure list prices are appealing to customers, while also implementing the personalized promotion strategy outlined in our previous report in order to keep customers in the stores and making purchases. Our methodology and analysis only uses 100 products and 10 stores, but it's possible to expand this to more product offerings and stores. We recommend that Pernalonga try different numbers of products and different combinations of stores in the future as they start to gather more data.

One potential drawback to our approach is the fact that we are only using two years of data from 2016 and 2017, while trying to make recommendations for 2020. In the future, as Pernalonga continues to grow and profit, we hope to be able to re-visit our analysis using more data. By having more data, we would be able to more appropriately address issues such as seasonality and different trends in the data and incorporate them into our model. Similarly, having data on the stores specifically could give us insight into what products might be best to optimize. Finally, having data on competitor pricing over time can also provide insights on changes in demand of Pernalonga products, especially for products where prices have been stagnant and elasticity measures were not possible to accurately measure.