Recommending Shelf Price Changes for 100 Products Across 2 Categories to Boost Pernalonga's Revenue

MKT 680 - Pricing Group Project
Team 3- Gary Gao, Miller Page, Abhishek Saha & Alana Zhao

1. Introduction

Pernalonga, a large supermarket chain in Lunitunia that operates over 400 stores and sells over 10,000 unique products, is seeking to improve its revenue by adjusting shelf prices of certain products. As it is an organization that understands the benefits of making data-driven decisions, they are interested in making price changes on a select number of products that will improve their revenue while maintaining overall profitability. By analyzing two years of transactional data, we were able to generate price changes that will meet their needs by coming up with recommended price changes for 100 products that will boost their revenue across 10 of their stores.

2. Business Context

As our objective is to recommend price changes for the 100 products in the 1st week of April 2018 and we were given that Pernalonga implemented a similar promotion schedule this time last year, our first step is to look at how price changes were performing in the 1st week of April in 2017 to understand the past behaviors/trends. The first thing we notice is that there are no discounts being offered in any transaction between Apr 1st - 7th, 2017, that is to say. the promoted price and the shelf price are the same.

Thereafter, we analyze the permanent price modifications that were conducted in the 1st week of April, i.e. changes in the shelf price. We notice that on average there were no significant price changes that were done (The range of price changes was between -0.6% and 1.5% with a median of 0% & mean of 0.2%).

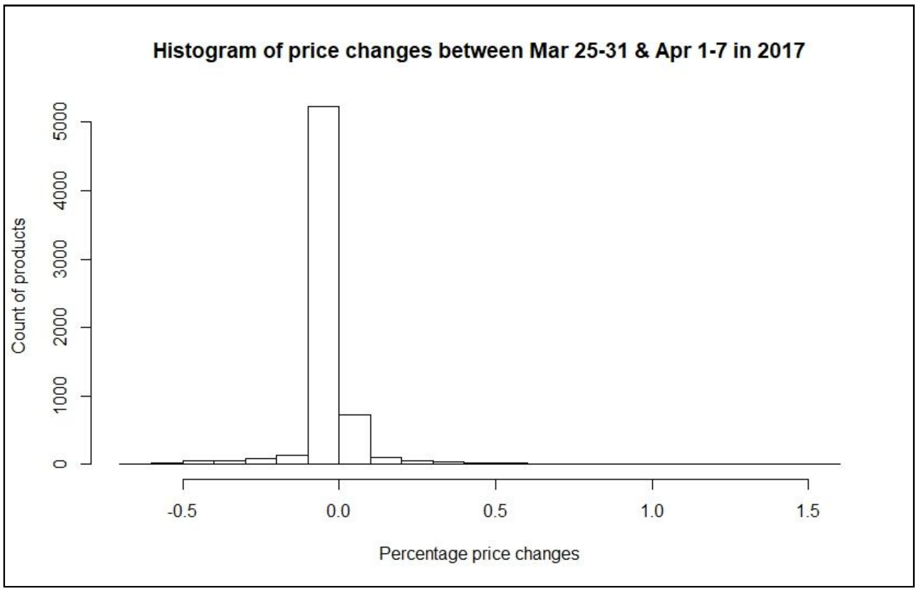

Figure 1. *Number of products by percentage price changes in two weeks*

Thus, Pernanloga did not make many significant changes in price for the week of Apr 1-7, 2017.

Subsequently, as we analyze the change in demand between the 2 weeks, we find that there hasn't been any significant change in demand between the 2 weeks (The median of the difference in quantity sold by each product is 0% and the mean is 0.96%).

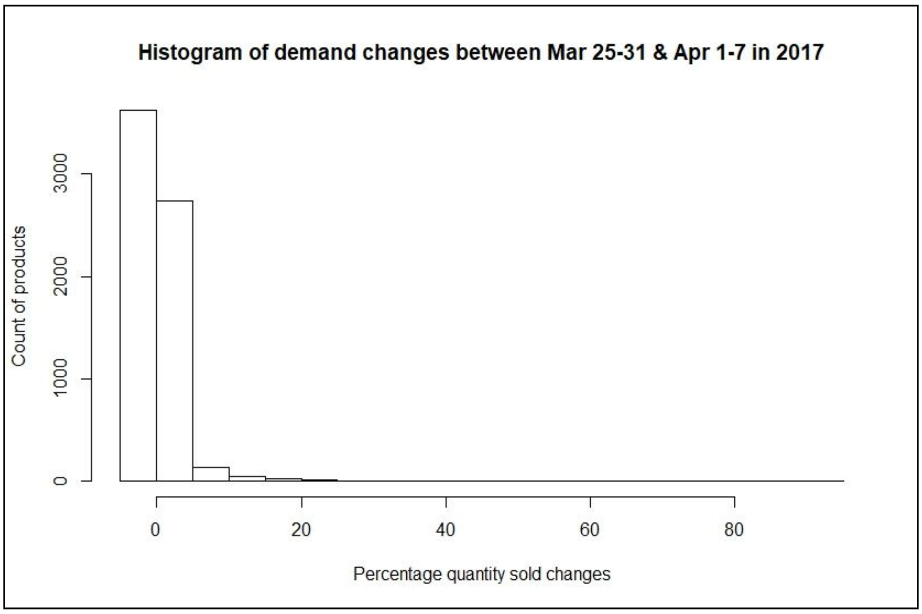

Figure 2. *Number of products by percentage quantity sold changes in two weeks*

Thus, there was no significant change in demand between the week of Mar 25-31 & Apr 1-7, 2017. This leads us to believe that Pernalonga could look at modifying prices of the products in order to increase sales and thus benefit from the increase in overall revenue.

3. Data Preparation and Exploration

For this project, we are looking at two previous years worth of data, from the beginning of 2016 to the end of 2017. Before proceeding, it is important to first prepare the data by removing the products that won't be involved in our potential price changes. There are two main changes that we made in this regard: fresh products, and those involving data limitations. First, we removed all products whose category included the word "Fresh" (such as fresh fish, fresh beef, etc), as well as products that were sold by weight (i.e., anything whose units were KG). As we are not changing any prices on fresh products, they do not need to be considered at this stage.

Secondly, we made two changes due to the data present (or lack thereof). The first deals with the condition that we making price change recommendations for only ten stores. As such, the products we consider for price changes must be sold in a minimum of ten stores. If they did not meet this criteria, they were removed from the data. The other change involves cleaning data in regard to the number of products in a category. As we are limited to making changes for 100 products in only 2 categories, we decided that in order for this to be balanced (i.e., not 99 products from 1 category and only from the other) we established a minimum cutoff of 10 as the number of products a category must contain. The products that did not meet this criteria were removed, and we are left with ~20.5 million transactions (down from ~29 million) to proceed with in the project.

After cleaning the data, we now want to take a more indepth look into the products and transactions we have left. And as we are dealing mainly with transactions of products, we first decided to take a look at the long-tail graph to get a sense of the breakdown of sales of different products:

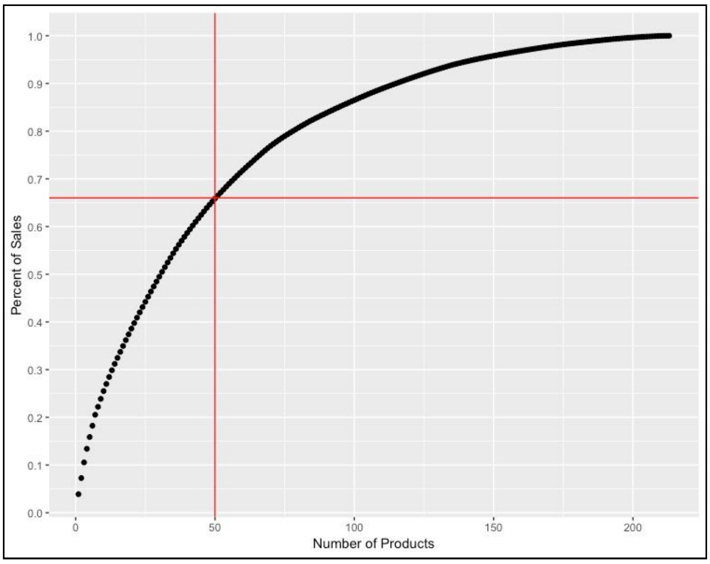

Figure 3. *Long tail plot across products*

From here we see that ~25% of the product categories account for over 65% of the sales. Thus, as we move forward, we want to make sure that the products we make changes for are included in the top selling products.

4. Choosing our Products

After finishing our data exploration, we set out to choose the products that we will adjust prices for. To do so, we decided to look at products that are elastic, as they will be the ones that are most responsive to changes we are looking to implement. To find the elasticity values, we first assumed constant elasticity, allowing us to use the following formula:

$$\boxed{\epsilon = p * d'(p)/d(p)}$$

Figure 4. *Constant-Elasticity function for elasticity*

To find these elements, we looked at the different shelf prices of each products for the past two years, and found the average price (*p*), the average demand (*d(p)*), and the slope of the price vs demand curve. The resulting values give us an elasticity for each product. A sample of the output is shown below:

Product ID	Elasticity
145519008	-0.6895245599
145519009	-0.1417936883
145519010	-0.0574287896
...	...

Table 1. *Elasticity by Product ID*

From here, we established an elasticity cutoff of -0.3 as a way to filter down the list of all the products. We chose -0.3 as it provided a good balance between quantity of products that met the cutoff while remaining relatively inelastic. After finding all the products that have an elasticity value less than -0.3, we were able to group by product category, finding those that have the most products present on the filtered list. Below are the top 5 categories with the most products found on this list:

Category Description	Count of Products
Fine Wines	64
Yogurt Health	44
Fine Wafers	39
Fruit Juices	35
Children's Food	29

Table 2. *Top 5 categories with most products sold*

As we can see, "Fine Wines" and "Yogurt Health" are our top two categories, and between them have 108 products with an elasticity of at most -0.3. Of these, the 8 products with the highest elasticity were removed, leaving us with 100 products that we will recommend price changes for. From here, we can set out to calculate the different aspects of the response function to be able to find these products' optimal price.

5. Finding Substitute and Complementary Goods

Constructing the response function requires the historical prices for a product's substitutes and complements during a certain time period. Substitutes is defined as items within the same subcategory that are rarely bought together, and it follows that the more a substitute product gets sold, the less the product of interest that will be sold. Hence, substitutes products have negative effect on each other. Conversely, complementary products are defined as the items that are commonly purchased together across all categories. We used association rules to find the substitutes and complements for each of the 100 products we chose using the package "arules" in R, which utilized the APRIORI algorithm. The parameters used in the APRIORI function allows us to define the minimum support and minimum confidence values, giving us flexibility in finding the substitutes and complements.

To get the substitutes of a product, we applied APRIORI function with support and confidence at 0, so it can generate all of the co-purchased items with in the same subcategory with the lift. Lift gives the correlation between item A and item B in the rule A=>B (how purchasing A affects purchasing B). If the rule had a lift of 1,then A and B are independent and no rule can be derived from them. If the lift is greater than 1, item A and B are dependent on each other. If the lift is smaller than 1, the presence of A will have negative effect on B.[1] In this way, we listed the substitutes for the 100 products by looking at the pairs of products generated with APRIORI function that have lift value smaller than 1.

As for complements, instead of looking at the lift value, we can directly use the rules generated with the algorithm by setting a threshold for confidence. Complements are usually treated at category level where co-purchasing behaviors are commonly found. But since our goal is to find complementary products at different brand or package quantity levels (i.e., a 4-pack of an item vs a single, etc) based on product ID, the probabilities of co-purchasing two products are relatively low. For example, someone may buy a bottle of wine and some cheese one time, and next time buy the same bottle of wine and a different type of cheese. While we could conclude that cheese and wine are complementary products, we don't know the exact complement for that specific bottle. Hence, the complementary product values were relatively low. To adjust for this, we set the confidence value to 0.25, meaning that the probability of purchasing both product A and B over the probability of purchasing product A is 25%. Then, the products found in the A=>B rules generated were extracted out as complements.

6. Accounting for Seasonality

One of the factors which impact the demand is seasonality, as demand can increase because it is a holiday season or because of a festival or weather changes. Hence, to factor that in, we looked at category weekly average sales across years and created an index of the same. The sales at weekly levels reveal the trend of the uncontrollable factor with respect to a certain product in that time frame. Thus, we can take weekly sales of the category that a product belongs to as the seasonality factor in the response function. This will allow us to control for this factor when looking at the historical and future sales values.

7. Building the Response Function and Finding Optimal Price

To determine the optimal prices of the 100 products in two categories is to maximize the revenue, we need to build a demand response function for each product. Other than the

[1] Jabeen, Hafsa. "Market Basket Analysis Using R." DataCamp Community, Datacamp, 1 Aug. 2018, www.datacamp.com/community/tutorials/market-basket-analysis-r.

price of the product itself, several additional factors are considered, including discount amount of the product, prices of its substitutes, prices of its complements, and seasonality. Since the price changes are going to be implemented during a specific week, we decided to fit the model based on weekly data.

One thing worth noticing is that although we have 2 years of data, the 100 products we chose are not necessarily sold in every single week. Hence, for each product, we only fitted the model with the weekly data when the product is actually sold, and therefore the prices of this product's substitutes and complements that are sold in that week. More specifically, the price of each substitute or complementary product is calculated based on the average unit price of this product sold across all stores in that week. Similarly, the price of the product of interest is the average unit price looking at all stores. The discount amount of the product of interest is the average unit discount amount, which is calculated using transaction discount amount divided by transaction quantity. The seasonality index, as discussed above, is created on a weekly basis so it can be directly added to a specific product. The product demand is simply the total quantity of this product being sold in this week across all stores.

Once we constructed all the attributes needed, an example table can be generated for building the response function, such as the one shown below in Table 3, in which "avg_sales" represents the seasonality index, and the columns started with "s_#" are substitutes' prices (where the substitute's product number follows "s_") while the columns started with "c_#" are complements' prices (same as substitutes about product number).

demand	**year_week**	**s_999274511**	**c_999231999**	**...**	**prod_price**	**discount_amount**	**avg_sales**
14	2016, 17	1.79	0.1	...	1.79	-0.31	7123.5
5	2016, 18	1.74	0.1	...	1.92	0	4604.0
18	2016, 19	1.79	0.1	...	1.99	0	4678.5
4	2016, 20	1.79	0.1	...	1.99	0	5156.5
....	...	...	...	...	...	...	...

Table 3. *Attributes table for fitting a linear response function for one product*

For simplicity, we used linear regression to build the response function. The data in the table shown above for each product were fitted into the linear regression to calculate the coefficient of each variable. Then, to calculate the demand of one product in the week of April 1st to 7th in 2018, we assumed that prices of substitutes and complements, discount values, and seasonality index stay the same as the data in 2017. In this way, the only independent variable in the response function would be the price of the product of interest. The ultimate goal of this step is to find the optimal price of each product that helps maximize the overall revenue. Because the response function is built on product level across all stores, the overall revenue of this product can be calculated based on the product of optimal price and corresponding demand from the response function. Yet, in some occasional instances, the coefficient of the price in the response function is positive, which means that the increase in price will lead to higher demand of certain product, so we set the price of this kind of product as the highest unit shelf price in history. Most of the products show the regular tendency that when the price increase, the demand decreases. For those products, we can determine their optimal prices that result in the highest revenues with the optimization function in R. It is noted that one constraint we put on the optimal price is that the new price for 2018 cannot exceed 1.5 times of the highest historical price for one product as it is unusual to boost a product price by over 150% just for one week, which can potentially lead to unexpected customer purchasing behavior.

8. Optimizing Revenue and Choosing Stores

The goal of the project is to maximize revenue. But first, we must limit our products to only 10 stores, as directed by Pernalonga. To do so, we filtered our original data to calculate the sales revenue and sales quantity of only our 100 products. This enabled us to focus on the stores where our products sold the best, which will help us achieve the goal of maximizing revenue. An initial exploration of the list of stores revealed that the top stores accounted for significantly more revenue than the others. Below we can see a graph of the stores' revenue in 'Fresh Wine' and 'Yogurt Health' categories.

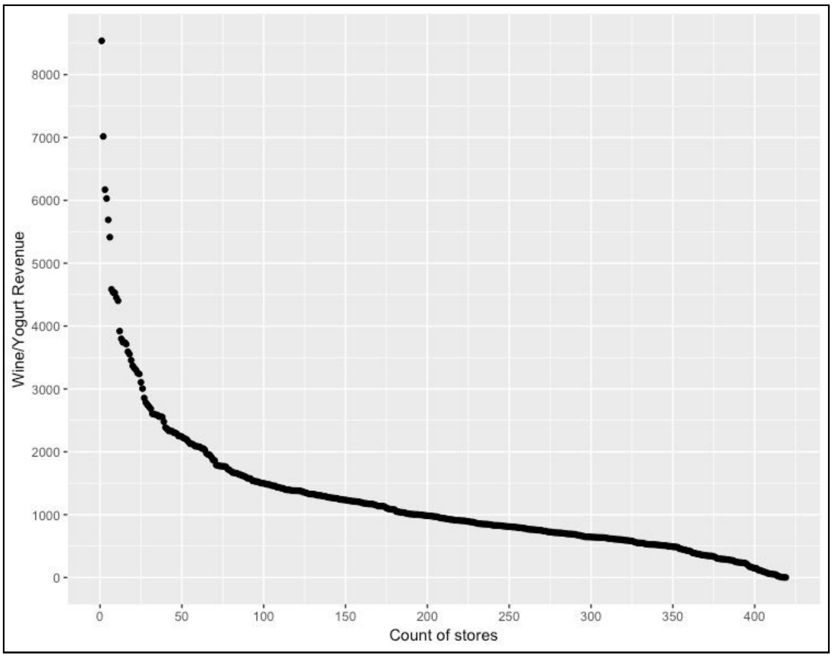

Figure 5. *Plot of Wine/Yogurt revenue by stores, decreasing by revenue*

From here, we see that there is a significant drop off after the first 15-25 stores. Then, taking a more in-depth look at the actual stores, we are able to come up with a list of the best 10 stores to use. The list we chose was a composite of those that sold the most in terms of average quantity and in terms of average revenue, and the stores and their sales revenue quantities can be found below.

Store	Quantity Sold	Revenue
349	8,051	$8,537
346	6,095	$7,017
343	5,953	$6,029
342	5,535	$5,690
345	5,702	$6,170
588	4,819	$5,414
335	2,861	$4,585
344	4,207	$4,534
347	4,081	$3,920
398	3,412	$4,452

Table 4*. Top 10 stores used*

Once we had our 100 products (across 2 categories) and 10 stores, we proceeded to find the weekly sales for those products at those stores along with the mean weekly price. Post that, we chose the last available sales weekly sales information for each product at each store. This was done to compute our 'current demand' and 'current price'.

The demand function of each product was then used to scale the slope of the product at a store level. This was done using the ratio of the demand of the product at that store to the demand of the product overall. We then used the slope, 'current demand', 'current price' and 'optimal price' to calculate the new demand at the optimal price. Change in sales quantity was the difference between new demand and current demand. Change in revenue was new revenue (new demand * optimal price) - current revenue (current demand * current price).

9. Results

The change in sales quantity, revenue & profitability was then aggregated at a store level to come up with the table below. From our data exploration, we had found that during the week of Apr 1-7, 2017, Pernanlonga had not provided any 'discounts'. Hence, there were no investments proposed for the week of Apr 1-7, 2018. Thus, change in revenue was equal to profitability.

On an overall basis across stores, we see an increase in sales of 11,452 units and increase in revenue of $18,090 by implementing these price changes. This increase is ~32% of the average revenue Pernanlonga earned from these 100 products at these 10 stores.

Store	Change in sales quantity	Change in revenue	Profitability
335	870	$1,572	$1,572
342	1,144	$1,575	$1,575
343	898	$1,313	$1,313
344	1,181	$1,830	$1,830
345	2,271	$3,821	$3,821
346	1,148	$1,768	$1,768
347	1,058	$1,587	$1,587
349	795	$1,204	$1,204
398	345	$360	$360
588	1,743	$3,060	$3,060
Overall	11,452	$18,090	$18,090

Table 5. *Effects of adjusting prices of 100 products in 10 stores*

Note, for space and readability purposes, the full list of products and recommended price changes was excluded from this report. However, the list can be found in the excel file submitted along with this report under the "optimal_price" tab.

10. Conclusion

By constructing demand response functions and optimization studies on prices at product level, we conclude that changing prices for appropriate products would lead to significant increase in revenue and profitability at both store and overall levels. The 100 products in two categories are chosen based on their high elasticities and relatively large sales, and the 10 stores we chose to implement the new pricing strategy are based on the sales of those products at store level. The revenue lift from pricing optimization proves that in future, Pernalonga could continue this exercise at larger scale and in a more frequent fashion. In addition, Pernalonga could integrate its shelf price changing strategy along with personalized promotions to further boost overall revenue and profitability.