

Price Optimization for Pernalonga

Report by Group 10 for Pricing Project: 03/28/2019

* Chengzhang Jiang
* Eric Greenberg Goldy
* Haoyang Sang
* Pooja Adiseshu

# Executive Summary

In this report you will find the details of a price optimization exercise performed for Pernalonga. Pernalonga is a leading supermarket chain of over 400 stores in Lunitunia and sells over 10 thousand products in over 400 categories.

For the week of April 1-7 in 2018, by changing the prices for 100 products across 2 categories, the following results can be achieved

* **Increase in Profit: 2,629 Price Units**
* **Increase in Revenue: 5,557 Price Units**
* **Increase in Sales: 1,108 Units**

The categories chosen to get the above results are **Beer with Alcohol** and **Coffees and Roasted Mixtures** for the following stores – **342, 344, 343, 349, 395, 346, 341, 348, 345, 588**.

The details of the products like their optimized prices, estimated revenue, etc. are present in a flat file.

The methodology, reasoning and assumptions used to arrive at this result are also provided. To explain the flow of the project, it is explained in the following steps:

1. Important considerations and assumptions
2. Narrow down the target Categories and Products
3. Narrow down the target Stores and Products
4. Calculate Product affinity
5. Estimate Demand and Revenue for the filtered Products
6. Finalize the Stores, Categories and Products that are the most profitable

Some improvements to this price optimization are suggested at the end of the report.

# Important Observations and Assumptions

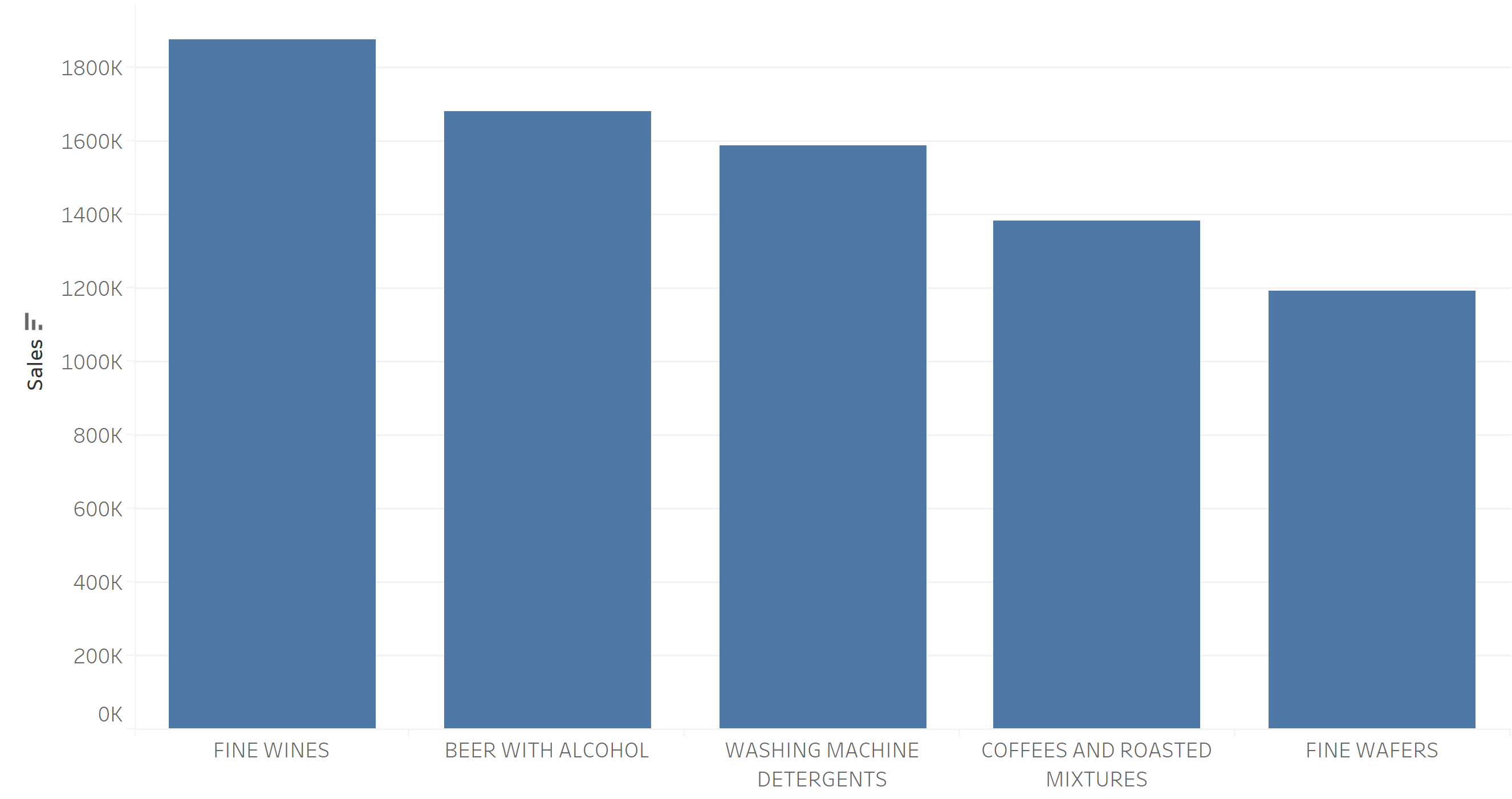
The following are some of the important considerations and assumptions we made for our analysis.

1. The week prior to April 1-7 has holidays, the retail staff required to change the prices on the shelves will be limited
2. Manually examined the categories and classified them as fresh or not
3. Excluded fresh products categories for price optimization but considered their effect on the products whose price we optimized
4. Shelf price does not vary on a daily basis but more likely on a weekly basis.
5. The price change will be permanent until the next (unscheduled) shelf price change.
6. Price elasticity for the same product varies across stores
7. The promotional details will be the same according to the corresponding period last year
8. Grocery margins are typically less than 5% (assuming the cost of a product to be 95% of its lowest shelf price)
9. This supermarket is not a typical Hi-Lo market as Walmart but more like a hybrid: most of its products' price change with week but some others remain constant

# Narrow Down the Target Products and Categories

By removing the fresh categories (such as fruits, vegetables and meat), reduced the number of categories by 15% and thereby the number of products by 11%.

Further, we look at the sales of the products and compare it across categories and select the top 5 categories by sales reducing the number of products to 10% of the original. They account for a big share of the sales and implementing price change on these products is likely to affect revenue to a large extent. This criterion still considers those products with lack of sales in the historical data but have high potential growth in the future. Below we can see the sales of the top 5 chosen categories.



# Narrow Down the Target Stores and Products

Since the shelf price and price elasticity does not vary daily but more likely weekly. Also, there might not be enough demand (thereby data points) for each product every day, and thereby the noises that arise in demand could potentially affect model performance. For the analysis, we considered data aggregated at a week level to accommodate these constraints.

The price elasticity is different in different stores. To calculate the price elasticity, the price level of a store-product combination is calculated by dividing sum of the sales and sum of total quantities sold in that week. Instead of simply averaging the price of all the products, we also consider how many units are sold at each price as a weight in order to compute a reasonable price of a specific product in a store within a week.

We filter out the store-product combinations without price change on a weekly basis because we can neither compute their pricing elasticity nor use them as complements or substitutes. 34% of the store-products were eliminated this way. In addition, we removed store-product combinations that have less than 10 observations on a weekly basis.

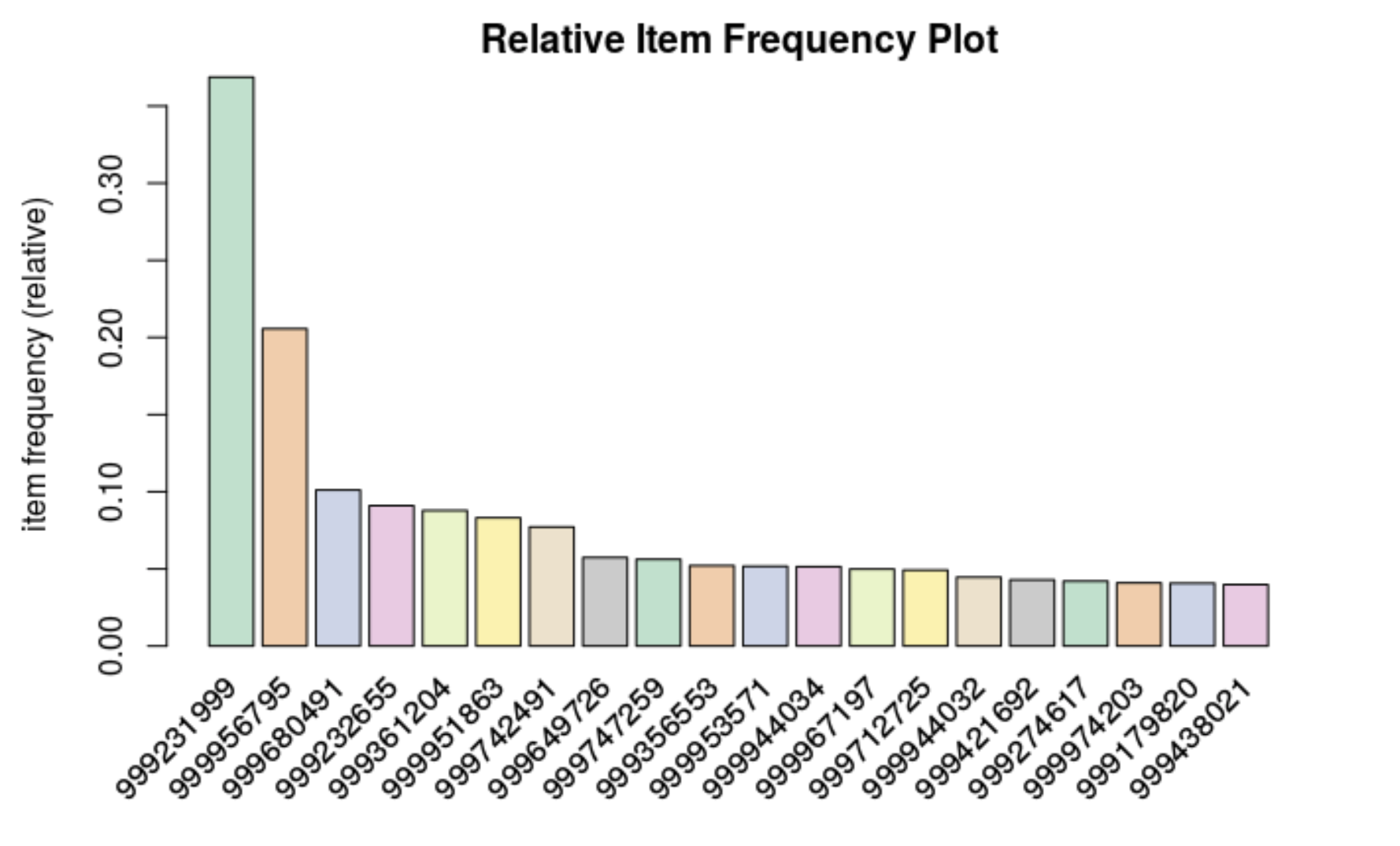
By the end of this exercise, we are left with 20 stores (from 421 stores) and 553 products (5% of the original number).

# Calculate Product Affinity

Compliments and substitutes are an important factor to consider when we are doing price optimization. They are crucial in estimating the change in revenue and demand. They are used as attributes to better fit price response function. We tried 2 methods to find out the compliments and substitutes. Both the methods are explained using a store 342, the store with the greatest number of transactions.

Method 1: Detecting association rules using basket analysis

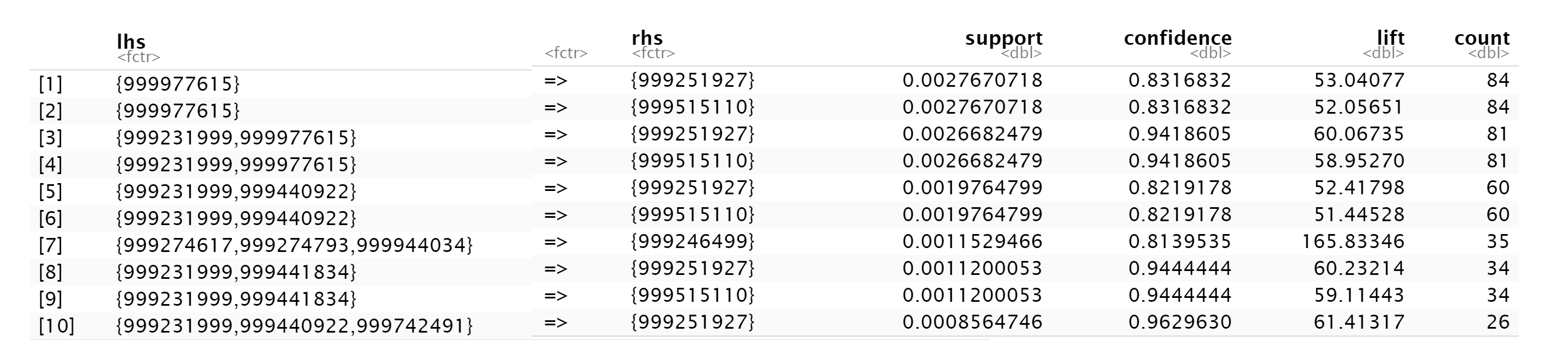
When we observe the store with the greatest number of transactions, we notice that apart from the top 2 selling items all other items are pretty much evenly distributed, so the association rule is easier to stand.



We used the APRIORI algorithm to mine the association rules. We set the rules such for an event to be considered, it much have occurred a minimum of 10 times in the store with a minimum confidence of 0.8. Also, we set a limit of 5 items for computational efficiency.

For each store selected, we are finding the complimentary and substitute products only for target products. The target products were not considered to find complimentary and substitute products for other target products since this will complicate the optimization.

Below is an example of a table that shows the probable complimentary and substitute products (rhs) for target products (lhs). The metrics needed to decide if they belong to the same basket such as support, confidence, lift and counts are also present alongside. If the confidence is high enough given enough support, then according to economic definition they are complementary goods.

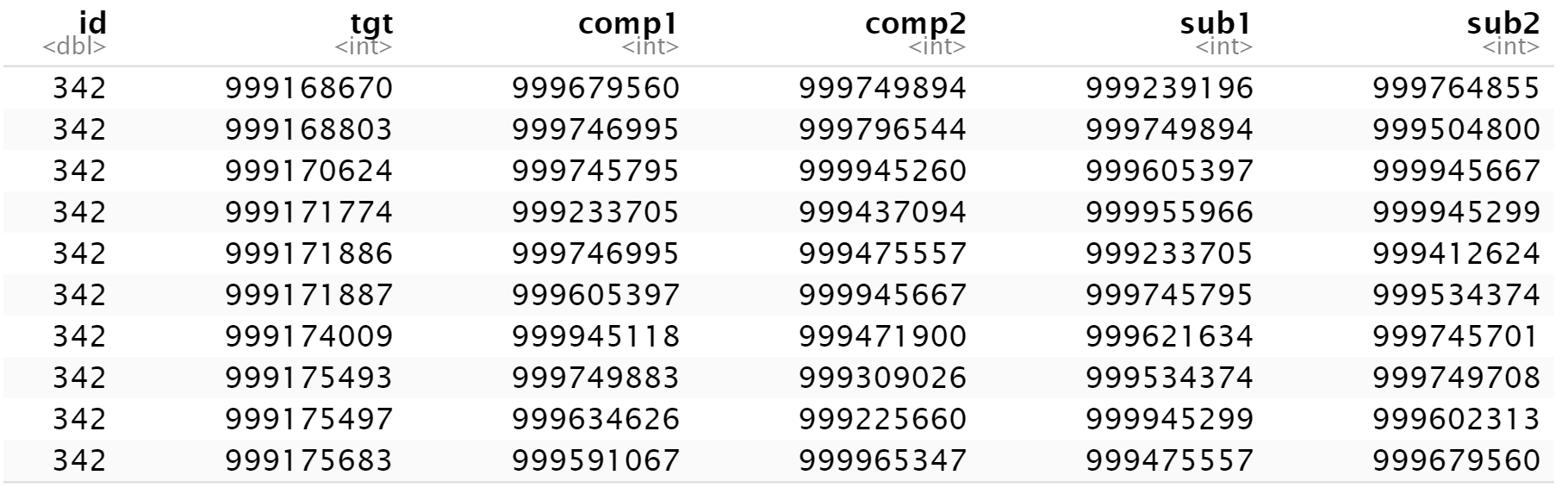


APRIORI is good at calculating complementary goods, because it is filtering out association rules based on confidence and support. If a combination happens frequently in transactions and the conditional probability of B given A (confidence) is high enough, we are confident to say that increase in demand for A increases demand for B, thus A and B are complementary goods. However, APRIORI cannot determine which goods are substitute goods as a low confidence doesn't necessarily mean that they are substitute goods, they are more likely independent from each other.

Method 2: Detecting complements and substitutes by calculating cross product elasticity

Since we are changing shelf price, we need to know how shelf price affects volume. So, elasticity was calculated based on shelf price. Also, while calculating elasticity, we assumed that we expect that the good A responds to price change in good B within the same week. We remove products which have no price changes.

Let’s consider the same store used in the previous example in the table below. For each target product we suggest 2 complementary and substitute products.



By restricting the number of complementary and substitute products to 2, we were able to find them for the target products.

# Estimate Demand and Revenue for the Filtered Products

Logit response function is chosen for estimation since it generally predicts revenue and profits better compared to the other two (especially in extreme cases) and allows elasticity to vary with different prices.

Some aspects we considered

* Weekly price: Vary this to get the best price which maximizes revenue
* Weekly discount: Used as a control variable to control for promotions
* Product affinity: The price variation of complements and substitutes
* Seasonality: We thought of three aspects to prevent or control the effects of seasonality.
  + Decomposition: We considered decomposition but since the quantities sold of a product may skip a few weeks and the data is not continuous in that regard, it could be hard to extract seasonality from the data
  + Weekly seasonality: We decided to introduce weekly index to the model to capture seasonal trends
  + Holidays: As holidays will have a big influence on demand, we account for this effect in the regressors

Some assumptions:

* In logit transformation of demand, the theoretical maximum volume is assumed to be 10% more than the maximum historical volume
* The promotional details will be the same according to the corresponding period last year
* It might not be appropriate to vary the price a lot. If the price is reduced to a large extent, this will even fall below cost, hurting profitability; if the price is increased to a large degree, this will affect customers' price perception of the brand ('expensive'), which could damage its brand image. The idea is to search for the price within a range around its recent price. (low: 80%, high: 120%, step: 2%)

Considering all the points mentioned above, we built the logit response function to estimate revenue and demand for the target products.

# Finalize the Stores, Categories and Products that are the most Profitable

To do this, we systematically eliminated combinations until we got the desired results which is the most profitable.

1. Discard those that will cause loss
2. Get the combinations with price specified that will bring positive incremental revenue compared to their original price
3. Restrict to 10 stores based on incremental revenue
4. Restrict to 2 categories based on average incremental revenue of each product within that category
5. Finalize top 100 products from their incremental revenue

This resulted in new and optimized prices for 100 products across 2 categories which would enable the following results to be achieved

* Increase in Profit: 2,629 Price Units
* Increase in Revenue: 5,557 Price Units
* Increase in Sales: 1,108 Units

The categories chosen to get the above results are Beer with Alcohol and Coffees and Roasted Mixtures for the following stores – 342, 344, 343, 349, 395, 346, 341, 348, 345, 588.

# Improvements and Short Comings to our Model

1. We could improve product selection method, for now we are selecting the category based on average revenue contribution for each product, and we could solve with optimization
2. although all cross-product elasticities are calculated on a weekly basis and are only calculated if transactions for both products happen in the same week, we are loose on restricting the seasonality effect on elasticity, and we suggest the supplement and substitute based on cross elasticity neglecting week number as we don't have enough weekly data. This could be improved if we have data with more granularity in time and quantity
3. We must limit the number of items in a basket either in basket analysis or in cross-elasticity analysis, because too many features and little data could lead to bad fitting of the model
4. Because of the complexity we can handle, we don't allow target products to appear as complements and substitutes or it might greatly increase workload
5. In the final response model, we only fit with 1 complement and 1 substitute good because the reliability of the model will decrease given the number of features (10+) compared with limited data
6. For some items in so stores, only less than 10 are sold out every week, making elasticity calculation harder because of not accurate enough delta in quantity. Customizing the range of prices from which we find optimized price for different categories and brands may be more appropriate. For example, the price of beer is relatively high, if we increase its price by 20% that could result in a $5 increase in its price. Therefore, to set different price ranges from which we aim for best price will be more reasonable. We thought of using the range between the minimum historical shelf price and the maximum historical shelf price. But due to time constraint, we were not able to implement our idea. We expect to improve the model in this respect with more time.
7. This supermarket is not a typical Hi-Lo market as Walmart but more like a hybrid: most of its products' price change with week but some others remain constant. However, while calculating elasticity we only calculate products which change prices, thus we are using the Hi-Lo response function
8. We might need to integrate competitors' prices to get a more meaningful response functions which control for competition factors
9. We should be aware of how the price change will affect customers' perception of the company. This analysis would require basket penetration calculations. If a product is a key product of consumers' perception, it might damage brand image by making them believe this company sells expensive products.