EMORY | Master of Science in Business Analytics MSBA

GOIZUETA BUSINESS SCHOOL

MKT_680_4100:

Pricing System

Group 5

Yi Cai

Sameera Vattikuti

Nicholas Yang

Zeyu Zhu

Price Change Recommendations to Improve Pernalonga's Revenue

Table of Contents

Chapter 1: Introduction

1.1 Business Background

Pernalonga, a leading supermarket chain with 421 stores which sells ~10K products in 430 categories with a consumer base of ~7900. Currently, 30% of their sales are through promotions (mostly in-store) executed in partnership with suppliers.

Pernalonga wants price change recommendations for 100 products from 2 categories in 10 stores that will improve their expected revenue while maintaining overall profitability. They also expect estimated changes in sales quantity, revenue & profitability for each store and overall 10 stores.

1.2 Problem Statement

The business problem could be translated to a data problem statement as 'Build a pricing system to suggest price changes for 100 products from 2 categories across 10 stores. The system should take into consideration factors like shelf price, promoted price, product affinity: substitutes & complements, sales seasonality and elasticity measures.'

Chapter 2: Overview

2.1 Data Overview

For the purpose of our analysis, we have transactional data related to ~2.69 Million unique transactions and 29.6 Million line-items. This data corresponds to 421 stores, selling ~10000 unique products belonging to 429 categories. The stores serviced 7920 unique customers in the given time period.

2.1.1 Assumptions & Considerations

- To uniquely identify each transaction, we created a **unique transaction id** that is a combination of customer id, transaction date and store id with the assumption that a customer visits a store only once in a day.
- **SKU** has been examined and factored into the product level analysis only when there are different SKU sizes within a category.
- Column **Year-Week** has been created based on date data

2.2 Business Overview

2.2.1 Categories Overview

Top Categories (by promoted price)

Pernalonga stocks products from 429 categories and from these, 56 categories (top ~13 % categories) contribute to ~60% of its revenues. Categories like dry salt cod, fine wines, coffees & frozen fish being the top among them in that order, contributing to the top 10% of revenues.

Top Categories (by shelf price)

54 categories (top ~12 % categories) contribute to ~60% of its revenues. Categories like fine wines, dry salt cod, detergents & coffees being the top among them in that order, contributing to the top 10% of revenues.

The following is a list of top 10 categories ranked both by the promoted price & shelf price. While overall there are many differences in promotions – few categories like ready to eat, chicken, pao, specialty milks are heavily promoted, broadly it may be said that some categories contribute heavily to revenues no matter how their value is calculated promoted or shelf price.

The promoted price here can be a result of multiple offers consisting of manufacturer-led discounts, store-led discounts or coupons. Since we are exploring demand-pricing relation to suggest price changes, we do not focus on the number of promotions applied on a product but rather just on the price & its effect on demand.

Category Desc Eng	Category Id	Rank (Promoted Price)	Rank(Shelf Price)
DRY SALT COD	95971	1	2
FINE WINES	95797	2	1
COFFEES AND ROASTED MIXTURES	96026	3	4
FROZEN FISH SERVICE	95974	4	5
WASHING MACHINE DETERGENTS	95788	5	3
FINE WAFERS	95991	6	6
OLIVE OIL	96036	7	7
YOGURT HEALTH	95854	8	8
MINERAL WATERS	95809	9	10
CANNED TUNA	96013	10	9

2.2.1 Store Overview

Top Stores (by promoted price)

Out of its 421 stores, the top 50 stores (top 12% stores) contribute to ~30% of the revenues.

Top Stores (by shelf price)

Out of its 421 stores, the top 50 stores (top 12% stores) contribute to ~30% of the revenues.

The following is the list of stores ranked by their contribution to sales revenue calculated at promoted price & shelf price of products. We can see that the top stores are common among both methods and imply that they may be following similar promotion patterns & strategies across stores.

Store ID	Rank(Promoted Price)	Rank(Shelf Price)
342	1	1
345	2	2
349	3	3
344	4	4
343	5	5
341	6	7
346	7	6
588	8	8
347	9	9
157	10	10
994	11	11
525	12	12
335	13	13
996	14	14
572	15	15
315	16	16
320	17	17
307	18	18
395	19	19
556	20	22

2.2.1 Assumptions & Considerations

Seasonality

Since the marketing promotions are set to take place in April, we look at a high level to see if any seasonality exists. The above plot gives us a picture of seasonality on a monthly and a weekly basis. There is an effect of seasonality on sales in April for all stores and categories consolidated. We will further examine the effect of seasonality for the relevant products & categories later in the process.

Month
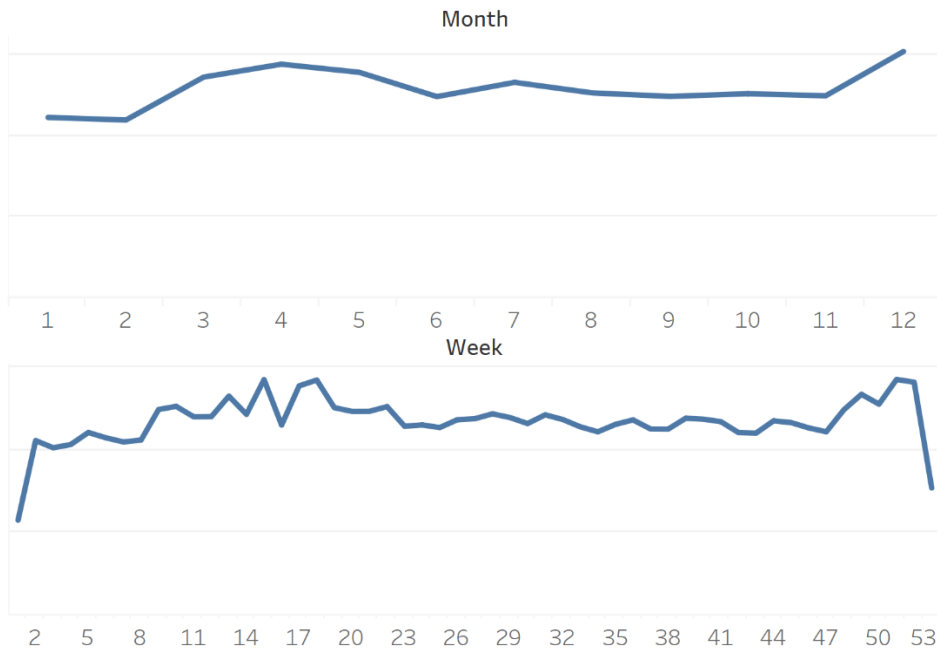
Fresh Produce

There are a lot of categories of fresh produce are present in the data. Fresh produce must be excluded from the pricing analysis. This was done in two steps. One was by filtering out all categories names with fresh included in them. In addition to this, there are other categories like potato, cabbage, tomato etc. which have not been labeled as fresh produce. Such categories have been excluded from analysis in

the next step where we filtered out categories with less than a certain number of products in them.

Unit of time (for analysis)

We take week as a unit of time for analysis. This gives us more detailed idea of trends than when month is taken as a unit and a high-level view than when taken at a daily basis. Weekly data also flattens the weekday vs weekend sales effect while giving a better picture of yearly trends in sales & demand.

2.3 Modeling Overview

Following is a brief description of possible consumers' motivations to purchase and the corresponding approaches we considered while building the recommender models.

2.3.1 Elasticity Measures

We considered different price response functions for the purpose of this project.

Linear Price Response: Simple & convenient to use but may not do a good job of representing the demand-price relationship. It also predicts unrealistic demands at extremely high prices & does not yield realistic results in a competitive market.

Constant Elasticity Price Response: This assumes a point elasticity that is same for all prices. This give unrealistic demands at extremely low prices.

Logit Price Response: Captures shifts in demand with small changes in prices. Gives realistic demands at extreme prices.

The logit price response function would be ideal in this case of price optimization. However, since we do not have the theoretical

maximum demand of a product (as required to use this function), we choose the next best option and use constant elasticity price response function, which is also a reasonably fair representation of most products purchased in a grocery/ supermarket retail. We primarily use the constant PR function. For the few products that have very low prices for which constant PR generates unrealistic demands, we use linear price response function to recommend price changes and estimate revenues.

2.3.2 Category Selection

As the final goal of this exercise is to increase the revenue of Pernalonga through price optimizations, we first look at the categories with highest revenue contribution to begin the analysis. After excluding the categories like fresh produce, we further filter the categories by the number of products they have available in each of the categories.

We use price-response functions at a category level to identify categories with appropriate elasticity & delve further into those categories.

Since we want to identify the relation between demand and price, we base the analysis on the promoted price since that is the price consumers have purchased the product at.

2.3.3 Product Selection

Once we have selected 2 categories from the above step, we use the price response functions to determine elasticity of products. Factors like seasonality, substitutes, complements, discount amount have been included in determining the new price at a product level.

Seasonality has been captured as a categorical variable (-1,0,1) depending on a week registering a 10% spike or dip of sales when compared to the yearly average.

All the products from a category are treated as ***Substitutes*** for a product being analyzed. Their price has been included as a variable in the mode.

Complements have been obtained by building a lift matrix between the subcategories of the 2 chosen categories & all 1410 subcategories which are available at the stores that see an associated increase in sales along with our focus subcategories.

2.3.4 Store Selection

Once again, we identify the top stores by revenue generated from the relevant categories. Depending on a history of products being sold in the outlets, estimates in revenue changes have been captured & list of stores for price change included.

Chapter 3: Recommendations & Modeling

3.1 Pricing & Modeling

The constant-elasticity and logit response functions are the most widely used in retail pricing because they allow interaction of causal factors. In this project, we have limited information to assume a theoretical maximum demand, which is critical in calculating logit response function, on given items, so we choose constant-elasticity response function to evaluate the elasticity of each product and estimate the revenue generated. To validate if the constant-elasticity function can reflect the true elasticity, we take a sample test. The result shows the explained variance score is 0.74, indicating the elasticity generated by the response function align with the real situation.

3.2 Category Level Modeling

Function 1: log(demand) ~ log(price)

Starting on the categorical level, we build the response functions for 18 categories that have more than 100 products. The coefficient of log(price) for each function is the elasticity for each category. We compare all 18 categories and choose the top-two categories with highest elasticity: ice cream (-3.8546) and child food (-1.5029). High elasticity means that the consumers are more sensitive to the price changes in ice cream and child food, which leads to high potential for revenue generation.

3.3 Product Level Modeling

Function 2: log(demand) ~ log(price) + log(price_sub) +
log(price_com) + as.factor(seasonality) + discount

After choosing the category that has the highest potential to generate revenue by changing price, we take a closer look on the product level to choose which products that needs price changes. When build the response function for each product, we include the effects of the prices of the substitutes and complements, the seasonality, and the discount. The selected ice cream and child food categories contain 10 subcategories and 268 products. We build a lift matrix between the 10 subcategories and all 1410 subcategories in the supermarket to find the top-three most associated subcategories for each product. The complementary product price is the average price of product in the top-three categories. We use the average price of products (excluding the product itself) in the same subcategory as the price of substitute product. We built 268 response functions for all product and choose 83 product that are most elastic and 17 products that are most inelastic to be considered for price change.

	Product	Elasticity
81	999245223	-1.0787506
82	999268366	-1.0115713
83	999170820	-1.0005330
84	999155543	-0.9948243
85	999357964	-0.9060737
86	999294112	-0.8595247
87	999224490	-0.8542920
88	999228006	-0.7875351

Because the elasticity for each product is constant, we cannot use elasticity=-1 to find the price that can maximize the revenue. We propose a pricing strategy of increasing 10% of the latest list price for 83 elastic products and reducing 10% of the latest list price for the 17 inelastic products. Comparing the new estimated revenue with the original revenue, the new pricing strategy increase the revenue for all 100 products selected. A closer examination shows that the constant-elasticity response function over predict the demand at small price (as shown in the graph) and leads to extreme values for the revenue

Price Response Function
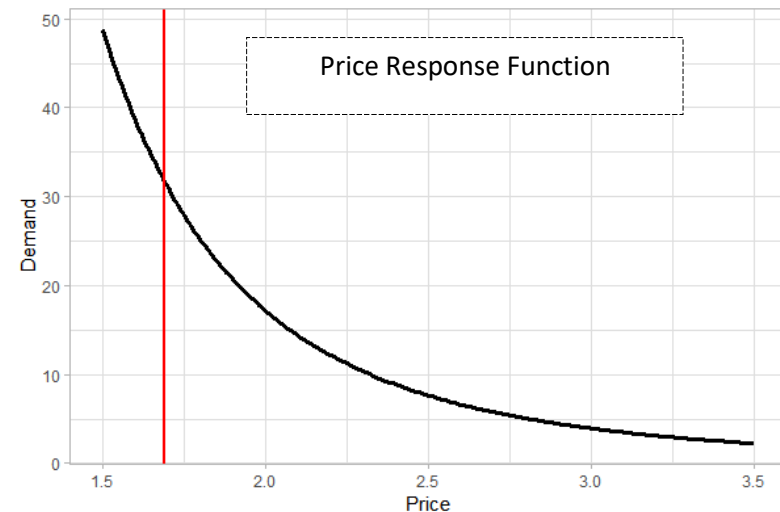
increased. To adjust the extreme estimate revenues, we apply a linear price response function at small prices to predict the expected demand and revenue. The increase in sales quantity and revenue for each 100 products in each store is shown in the table below.

prod_id	store	quant_change	quant_change_ratio	rev_increase	rev_increase_ratio
999528656	307	631.0563375	132.88%	278.6045576	109.60%
999231779	343	0.795265214	66.64%	4.563034068	83.31%
999180039	588	0.610865089	54.46%	3.91272766	69.91%
999180096	344	0.507496002	51.65%	3.275901271	66.82%
999441670	342	0.880847286	48.65%	3.093501931	63.51%
999712436	157	1.823662679	81.51%	6.506633655	63.36%
999712437	347	0.44504864	37.87%	2.786493884	51.65%
999361208	342	1.391956149	66.37%	2.910070988	49.73%
999180096	588	0.354377095	35.22%	2.447286057	48.74%
999261327	346	0.299054179	27.36%	0.872163912	40.09%
999228002	343	0.420295321	19.14%	1.902659346	31.05%
999356555	342	0.242745763	18.96%	1.576268886	30.86%
999528656	342	287.4230617	42.68%	100.5380044	29.71%
999266082	345	0.204570433	15.21%	0.888040088	26.73%

By comparing the new estimated revenue with the original revenue in the week of April 1, 2017, the new pricing strategy generated an average of 22.6% increase in the revenue across all stores. The increase in sales will tentatively affect the cost per product sold, but we assume no change in total cost in our analysis when estimating the profitability, so the increase in profitability is comparable with the increase in revenue.

Chapter 4: Conclusion

The new pricing strategy will generate an average of 22.6% increase in store revenue. Going forward, few adjustments should be applied to improve the analysis. First, more cost data will help us to better evaluate the profitability driven by the new pricing strategy. Second, several products still show an estimated increase of revenue of over 100% after the adjustment. More analysis is needed to bridge the gap between the original constant-elasticity function and the linear response function for a more accurate prediction. Third, with more information on the theoretical maximum demand, we can build the

logistic response functions to simulate the correlation between price and demand, find out the optimal price for each product, and compare the pricing strategy with the current one.

The work so far suggests price recommendations for 100 products across 2 categories along with the list of (10) stores that these price changes could be implemented in.

Reiterating on the reason for selection for selecting the constant elasticity price response function – the objective of this exercise is to recommend price changes that could lead to revenue increases and not necessarily to recommend the 100% optimal price for a product. And the function we chose serves this purpose well. It helps us recommend changes without making too many assumptions (eg. Theoretical maximum demand of a product) about the data.

Secondly, the model we use has an explained variance score of .7 which indicates that our model is good.