MKT 680

Project Report

Product Recommendation and Cost-benefit Analysis of the Nivea Personalized Campaign for Pernalonga

Students:

Meng Cheng

Feifan Gu

Sidi Liu

Under supervision of

Prof. Alvin Lim   Prof. David Sackin

Emory University

Goizueta Business School

CONTENTS

1. Introduction

Pernalonga, a leading supermarket chain of over 400 stores in Lunitunia, sells over 10 thousand products in over 400 categories. Pernalonga regularly partners with suppliers to fund promotions and derives about 30% of its sales on promotions. While a majority of its promotion activities are in-store promotions, it recently started partnering with select suppliers to experiment on personalized promotions.

This project aims to help Pernalonga, our client to develop a marketing campaign focusing on Shampoo and Hair Conditioner categories to experiment on personalized promotions. Specifically, our team will propose the brand to invite, the customers to target with and the specific products to promote. This marketing campaign aims to maximize the opportunity for incremental sales and profits for Pernalonga, thus we also include additional cost and benefit analysis in the end for reference.

2. Data Preparation and Exploratory Analysis

To better understand Pernalonga's customers and the shampoo and hair conditioner brands and products in a precise way, we first conducted data cleaning and feature engineering.
We received two datasets from our client, covering the transaction history with 29,617,585 observations and product information with 10,767 observations. We drop 8 transaction observations with negative payment amounts and recreate the unique transaction id based on customer, transaction and store information. Our clean datasets include purchasing data with 2,830,564 transaction records from 7,920 customers with 10,770 products in 421 stores.

To better serve our client's goal, maximizing the opportunity for incremental sales and profits for Pernalonga, we need information about the profit and cost. In general, the supermarket industry has a profit margin range from 1% to 6%. We assume that profit margin should be negatively correlated with sales volume and assign different profit margin levels according to different total sales volume levels for products with sales units of count and kilogram respectively.

Level	Total Sales Volume unit in count	unit in kilogram	Assigned Profit Margin
below 25%	less than 354	less than 218	6%
25% to 50%	345 - 925	218 - 710	4.50%
50% to 75%	925 - 2731	710 - 3444	2.50%
above 75%	more than 2731	more than 3444	1%

Table1. Profit Margin Assumption Rule.

We notice that products in the BAG category have significantly high sales volume. After checking the subcategories under this category and related transaction history, we find this category includes products such as shopping bags, freezing bags and shopping carts with an average price lower than 1 dollar. We decide to drop all related transactions (28,852,307 observations) as customers do not come to Pernalonga to buy those things.

Since this project focuses on the shampoo and hair conditioner categories, our team takes further steps to understand the brands and products in both categories. Overall, shampoo and hair conditioner products bring $517,020.5 in revenue and $21,727.1 in profit with 158,334 counts of product from 7,666 customers' 116,088 visits to 418 stores. The average discount on shampoo and hair conditioner products is 0.36. There are 254 customers, which consists of 3% of the total customers, who never purchase shampoo or hair conditioner products from Pernalonga. Those two categories contribute to 0.8% of the total revenue, 2% of the total profit of Pernalonga, and have an average higher discount level compared to the average discount level across all products in Pernalonga, which is 0.16. It suggests that shampoo and conditioner are ideal categories for marketing campaigns since they can not only bring profits from their own sales increase, but also generate halo effect and benefit the store by improving customers' overall impression on Pernalonga.

At the product level, we find all top products belong to the shampoo category. Surprisingly, three products specially designed for men pop up with top volume, visit count, customer count, store count and total discount. This may be due to the limited choice customers have in this subcategory. At subcategory level, common shampoo products contribute the most with 166 products. At the category level, shampoo products offer more choices to customers and contribute more monetarily.

Measurement	Top Product				Value
	id	sub category	category	brand	
Volume	999441917	CHAMPO HOMEM	SHAMPOO	LINIC	2,172
Revenue	999273570	CHAMPO COMUM	SHAMPOO	TRESEMME	7,193
Profit	999180573	CHAMPO COMUM	SHAMPOO	ELVIVE	205
Visit Count	999441917	CHAMPO HOMEM	SHAMPOO	LINIC	1,909
Customer Count	999442459	CHAMPO HOMEM	SHAMPOO	LINIC	1,127
Store Count	999441917	CHAMPO HOMEM	SHAMPOO	LINIC	355
Total Discount	999184254	CHAMPO HOMEM	SHAMPOO	HEAD&SHOULDERS	0.49

Table2. Top Shampoo and Hair Conditioner Products

Sub Category Desc	Category Desc Eng	Volume	Profit	Revenue	Prod Cnt
CHAMPO COMUM	SHAMPOO	87,130	11,783	291,430	166
AMACIAD CABELO NORMA	HAIR CONDITIONERS	27,854	4,139	86,921	68
MASCARAS P/ CABELO	HAIR CONDITIONERS	9,880	1,943	38,785	30
CHAMPO HOMEM	SHAMPOO	12,951	1,144	33,024	16
TRATAMENTOS CABELO	HAIR CONDITIONERS	8,718	1,357	30,862	19
CHAMPO ANTI-CASPA NR	SHAMPOO	7,618	772	20,645	12
OUT AMACIADOR CABELO	HAIR CONDITIONERS	2,221	367	9,189	5
CHAMPO 2 EM 1	SHAMPOO	1,962	222	6,164	3

Table3. Sales Summary by Sub Category

	SHAMPOO	HAIR CONDITIONERS
Volume	109,661	48,673
Profit	13,921	7,806
Revenue	351,264	165,757
Prod Cnt	197	122
Sub Category Cnt	28	30
Brand Cnt	16	14

Table4. Sales Summary by Category

At the brand level, ELVIVE, ULTRA SUAVE and PANTENE are the top three brands from a monetary perspective as well as traffic-driving perspective. Most of the brands offer products in both categories and we notice that the diversified product line is indeed related to larger customer groups while the discount level is not significantly correlated with the size of the customer group.

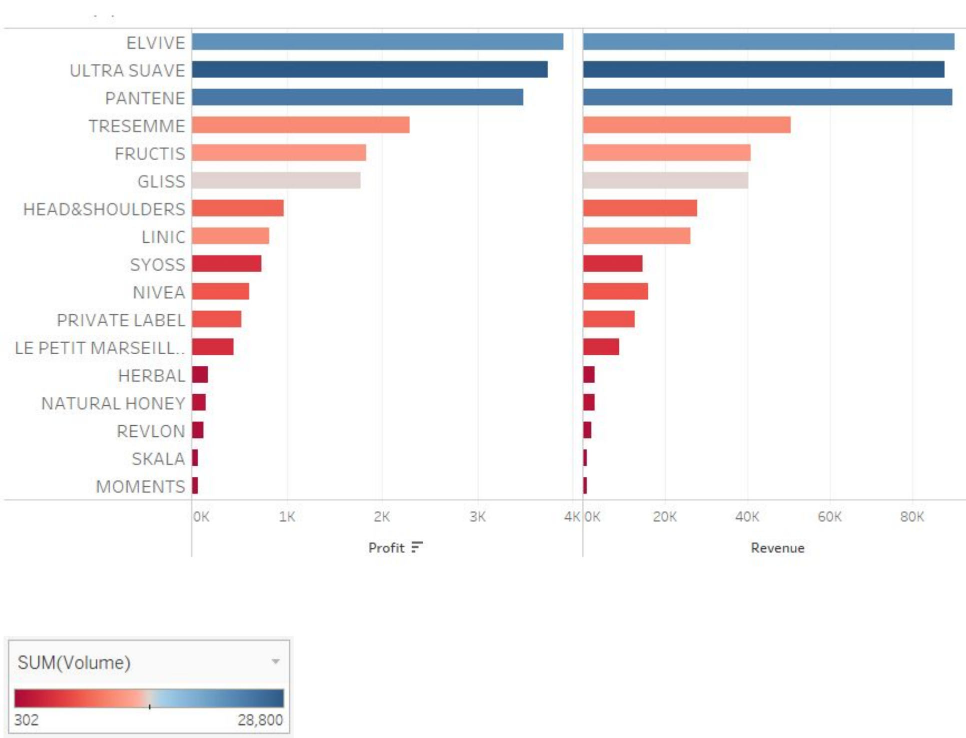

Fig5. Sales Summary by Brand I

	Category Cnt	Prod Cnt	Sub Category Cnt	Customer Cnt	Visit Cnt	Totaldiscount
MOMENTS	1	2	1	357	419	0.44
TRESEMME	2	28	5	2,824	8,180	0.42
SYOSS	2	12	3	1,786	3,537	0.42
GLISS	2	33	4	3,574	11,385	0.41
LE PETIT MARSEILLAIS	2	8	2	1,638	3,039	0.40
HEAD&SHOULDERS	1	13	4	2,509	6,328	0.39
LINIC	1	11	3	3,086	9,212	0.39
NIVEA	2	12	4	2,608	6,043	0.38
PANTENE	2	42	6	4,637	18,066	0.37
REVLON	2	3	2	426	578	0.35
ELVIVE	2	45	5	4,550	17,815	0.34
FRUCTIS	2	28	3	3,411	9,625	0.33
ULTRA SUAVE	2	59	5	5,091	22,193	0.32
HERBAL	2	4	2	547	801	0.31
NATURAL HONEY	2	4	2	560	1,358	0.03
PRIVATE LABEL	2	13	6	2,348	6,250	0.02
SKALA	1	2	1	157	296	0.01

Table6. Sales Summary by Brand II

As for the seasonality and time trend for each brand, we notice that there was a temporary boost in profit and revenue for most of the brands in 2017 Quarter two. There was not any significant boost in the promotion or a higher discount level. For ELVIVE and ULTRA SUAVE, both brands show decreasing trends in sales volume. This may indicate that those brands are becoming less popular and less attractive to our customers. Our team takes this into consideration when choosing the brand to invite.

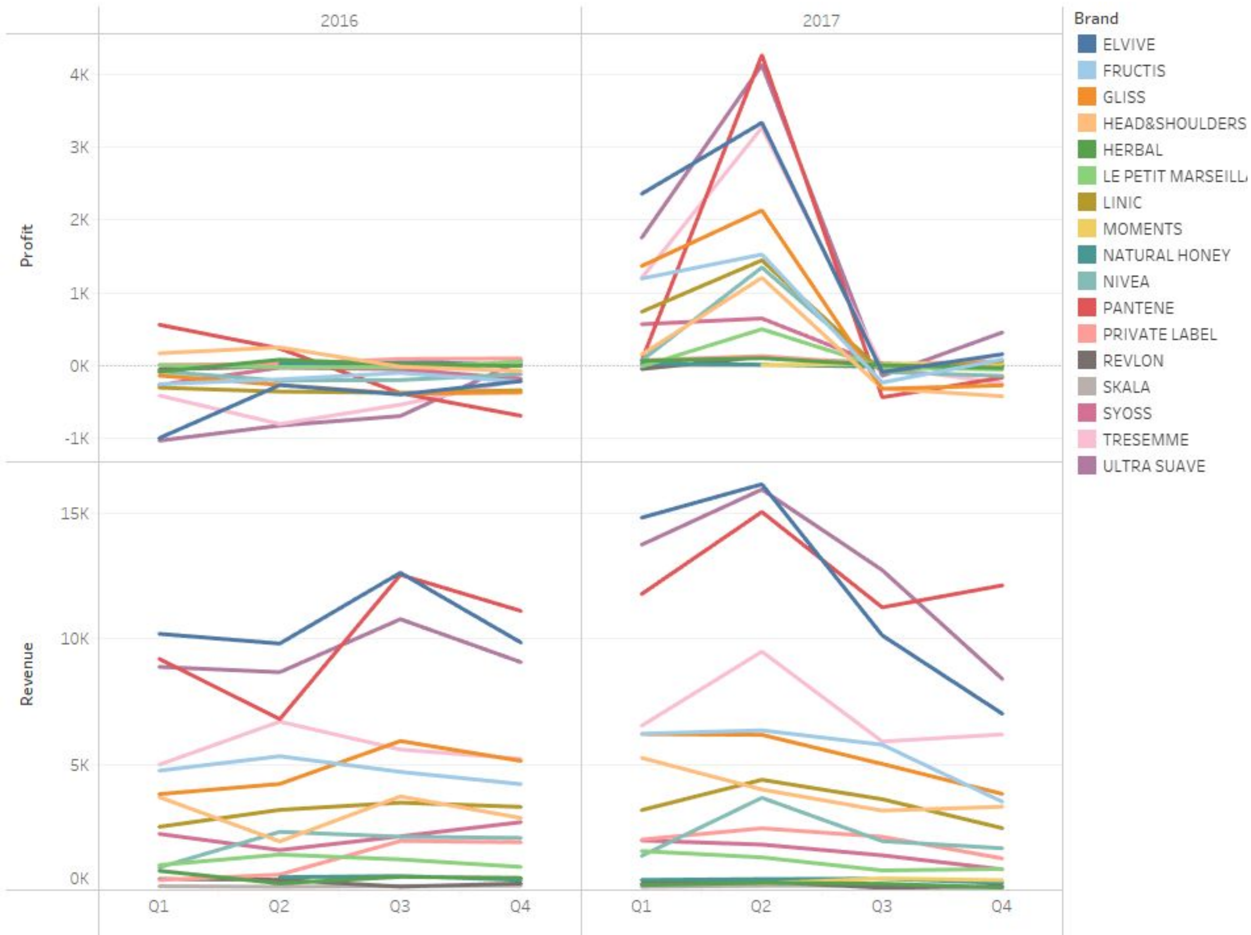

Fig7. Time Trend in Profit and Revenue by Brand

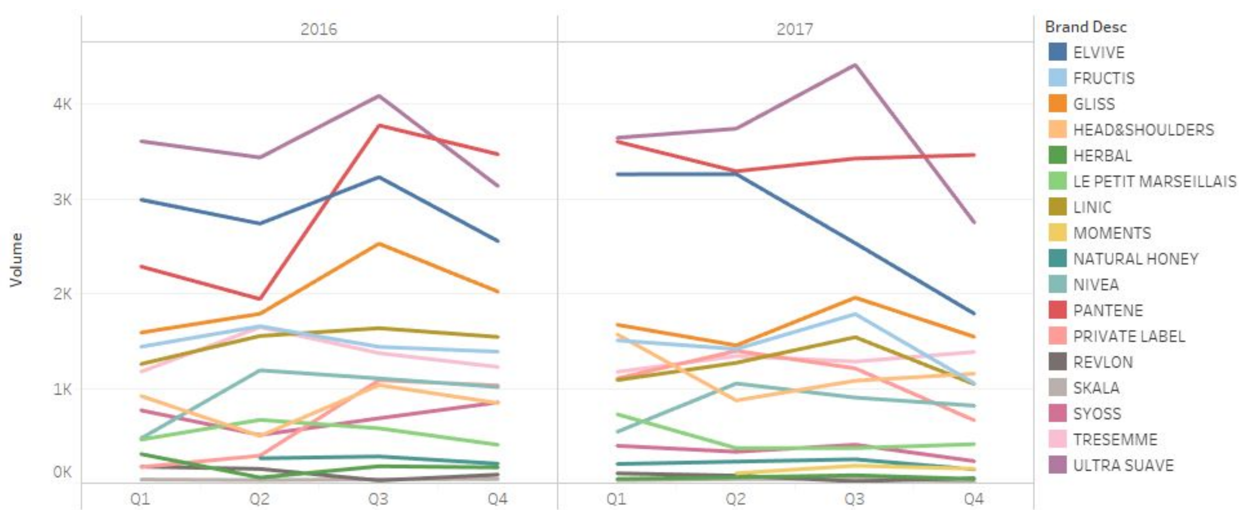

Fig8. Time Trend in Volume by Brand

To better understand the customer behaviours of shampoo and hair conditioner, we want to know the number of brands and products purchased by a customer. The majority of customers have purchasing history with 4 to 12 products offered by 4 to 5 brands. It suggests that customers, generally, would not focus on one favourite brand and it is feasible to recommend products offered by other brands with personalization recommender.

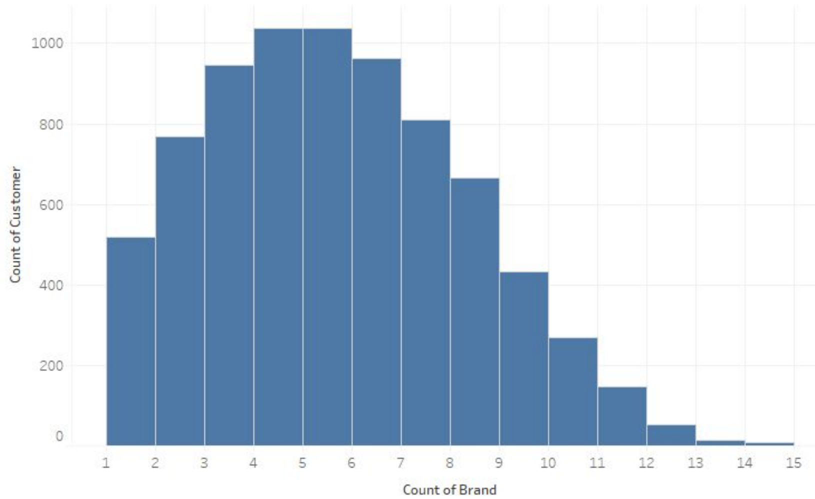

Fig9. Distribution of Brand Purchased Count

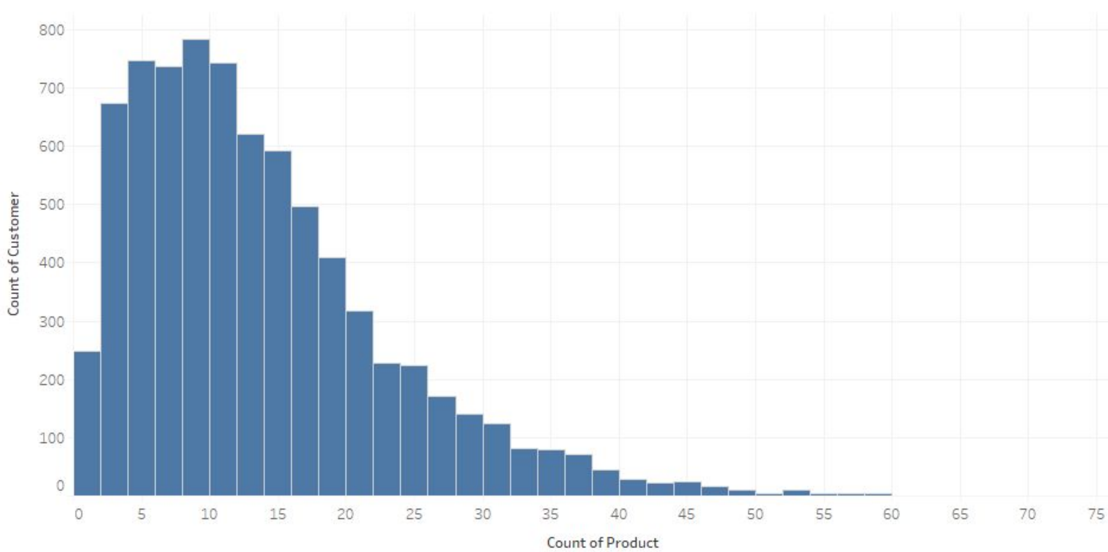

Fig10. Distribution of Product Purchased Count

3. Customer Segmentation and Target Customers Choice

In this part, we would like to segment the customers based on "shampoo/hair conditioners"-related features. After filtering all the transactions regarding the shampoo and hair conditioners categories, we calculated the total revenue, total profit, transaction times, quantity of goods purchased, unique goods purchased count, store visited count, brand purchased count, average discount, discounted transaction percentage, private label goods count, and non private label goods count for each customer, and use all those features to do K-means clustering after scaling them.

Out of 7920 customers of Pernalonga, 7666 of them have shampoo or hair conditioners purchasing records, which could generate those features. The remaining 254 customers have never purchased any shampoo or hair conditioners goods. Since they only contain 3% of the customers and never touch these categories throughout such a long time, it's hard for us to determine which product in the two categories is suitable for them, neither can we reasonably predict the sales. Therefore, in the following part we will cluster them remaining 97% of customers who have at least purchased shampoo or hair conditioners once.

Using the elbow method as is shown in the following graph, we determined that using four clusters in this segmentation makes the most sense.

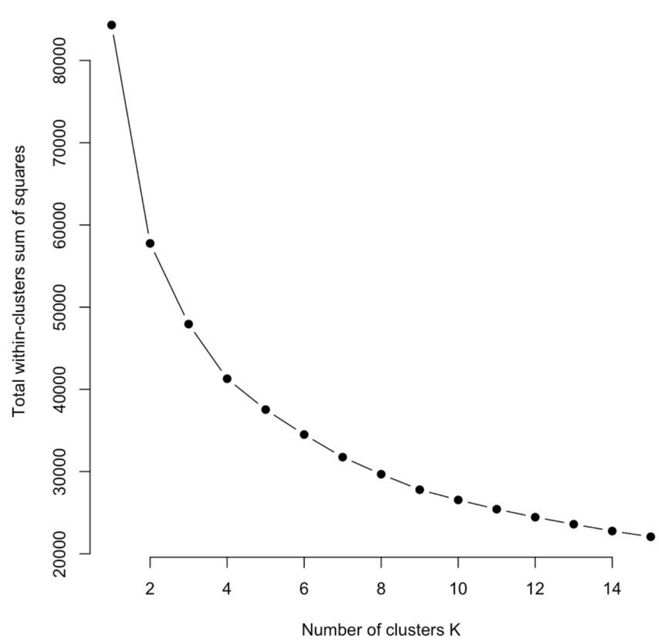

Fig11. Total within-clusters sum of squares ~ number of clusters K. The elbow is at k=4.

When running a K-means model with K=4, we separated the customers into four groups as is shown in the graph below. The number of customers in each cluster is rather balanced.

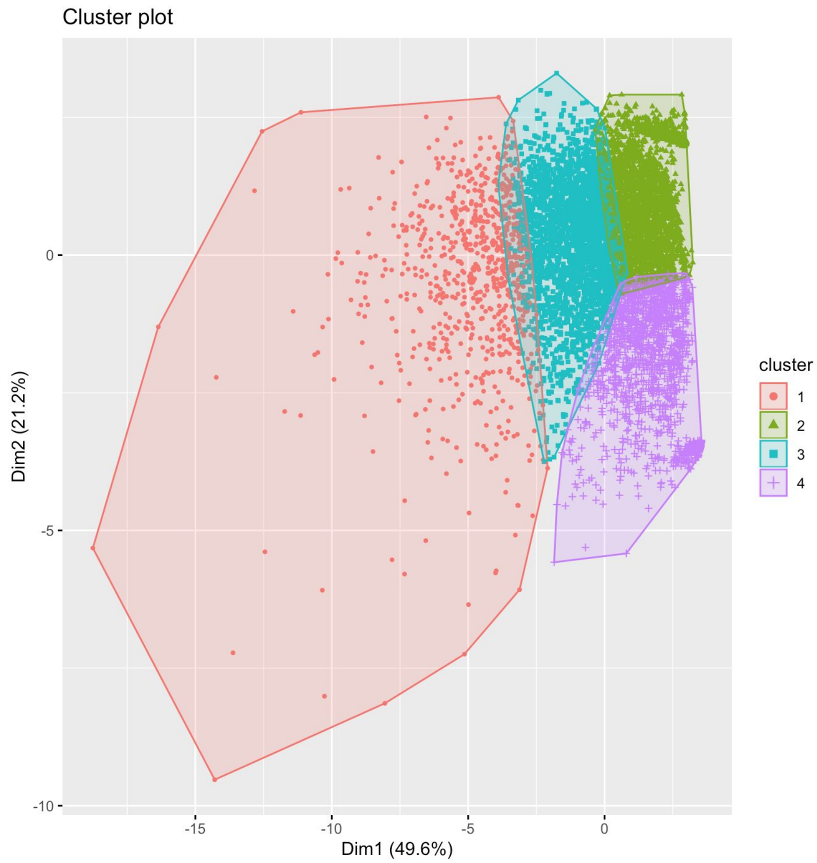

Fig12. 2-D visualization of customer segmentation using PCA.

To explore what characteristics each cluster has, we took a look at the centers of the four clusters.

	x_revenue	x_profit	x_visit_cnt	x_qty	x_prod_cnt	x_store_cnt	x_brand_cnt	x_totaldiscount	x_time_discount	qty_private	qty_nonprivate
1	2.1192871	0.71600733	2.0424596	2.1475650	1.9322589	0.4429629	1.2363033	0.07639839	0.03997655	0.74518806	2.1084600
2	-0.6392279	-0.41581817	-0.6476725	-0.6227477	-0.6366956	-0.2787634	-0.5629976	0.63693128	0.65118385	-0.26994567	-0.6037766
3	0.3460796	-0.04999683	0.4319830	0.3735492	0.4889983	0.3617790	0.6236099	0.19784515	0.16032602	0.09721487	0.3713388
4	-0.5600834	0.43275755	-0.6522892	-0.6487081	-0.7119998	-0.3590394	-0.7156848	-1.45616086	-1.39876240	-0.07873032	-0.6576356

Fig13. All the scaled features of the centers of the four customer clusters.

More intuitively, below is the bar chart of the important features of the four centers.

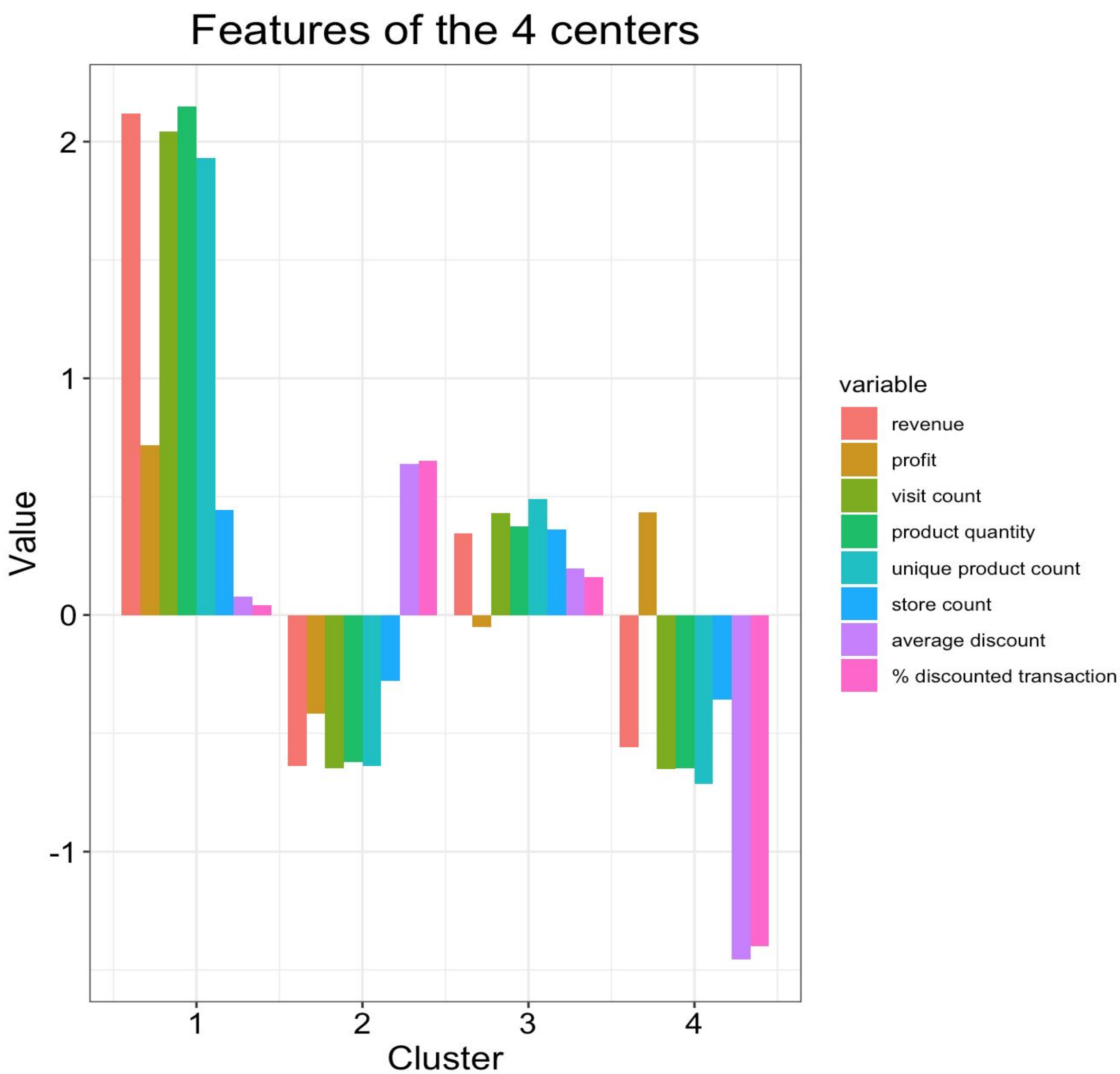

Fig14. Bar chart of the features of the 4 centers, showing that each cluster distinguishes in terms of those key shampoo/hair conditioner-related features.

From the graph, cluster 1 has the most revenue, profit, and all means of transaction frequency of shampoo/hair conditioners, while cluster 2 has the highest average discount and percentage of discounted transaction. Cluster 3's performance is rather modest and cluster 4 is indifferent about shampoo/hair conditioners, no matter if the product is in discount or not. Therefore, we decided to target the first two clusters of customers. Cluster 1 consists of 786 shampoo/hair conditioners lovers and cluster 2 consists of 2666 shampoo/hair conditioners discount lovers. 3452 customers in all out of the 7920 customers of Pernalonga, about 43%.

4. Brand Choice

In this part, we will decide which brand from the shampoo or hair conditioner category to invite to participate in a personalized promotion campaign. Since we have 2 targeted clusters of customers, we will find suitable brands for each of them and choose an optimal one as our final brand. In each situation, we will not only consider the benefit of selling the brand itself, but also their halo effect. The halo effect is the phenomenon that a person tends to see a person's performance as all good or all bad. In the context of retail, it refers to a consumer's favoritism toward a line of products due to positive experiences with other products by this maker. Retailers can usually create the halo effect by capitalizing on their existing strengths. Therefore, we would like to promote a brand that is already popular among our customers, and has not only been generating considerable profits of its own, but also driving profits for other products.

The following four graphs will show you the effect of selling the brand itself and the halo effect for the two clusters. Each graph is sorted according to the profit in the descending order, which we care about the most. We excluded the brand "PRIVATE LABEL", because it is sold by Pernalonga itself. It definitely will not be the brand to invite.

Cluster 1: shampoo/hair conditioners lovers.

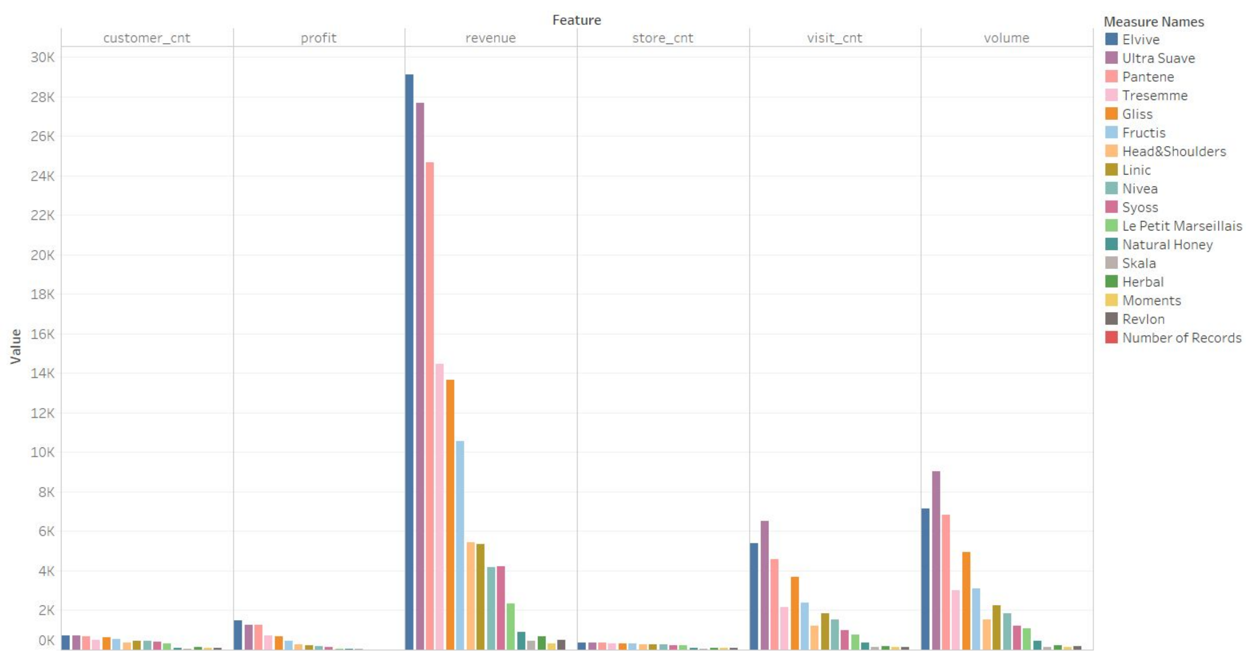

Fig15. Brand priority of shampoo/hair conditioners lovers in terms of the benefit of selling the brand itself.

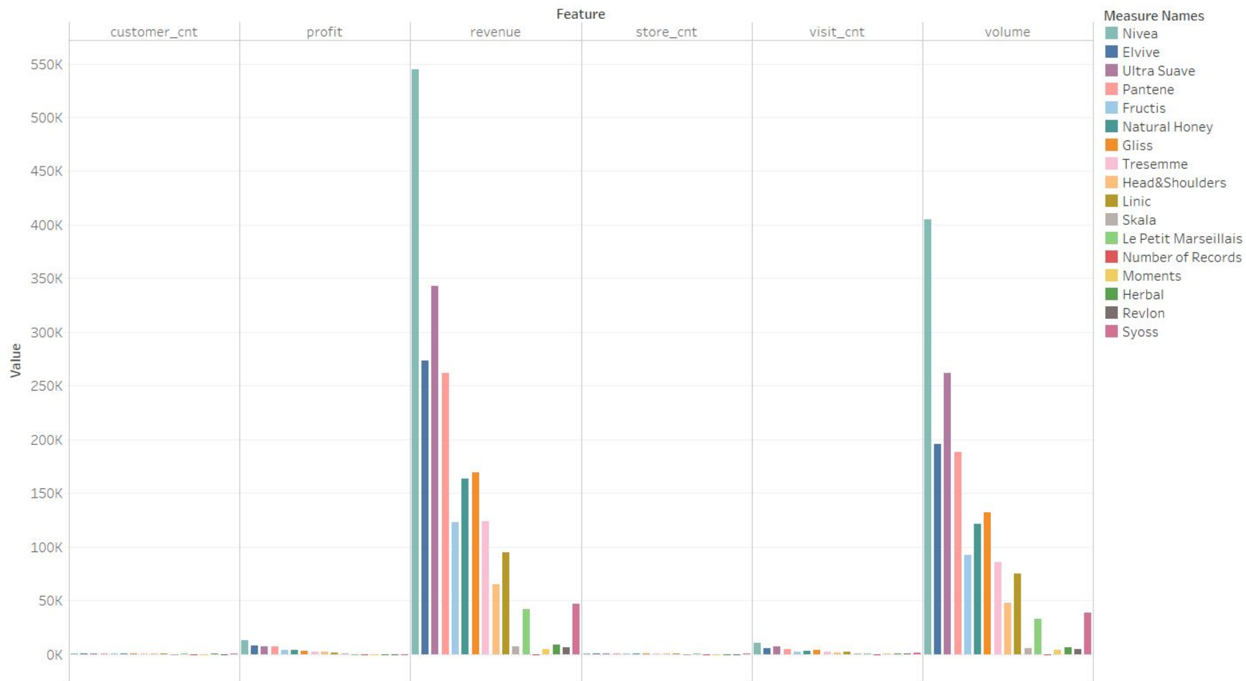

Fig16. Brand priority of shampoo/hair conditioners lovers in terms of the halo effect.

For cluster 1, in terms of the profit of selling the brand itself, Elvive is the best brand; whereas considering the halo effect, Nivea is the best one.

Cluster 2: shampoo/hair conditioners discount lovers.

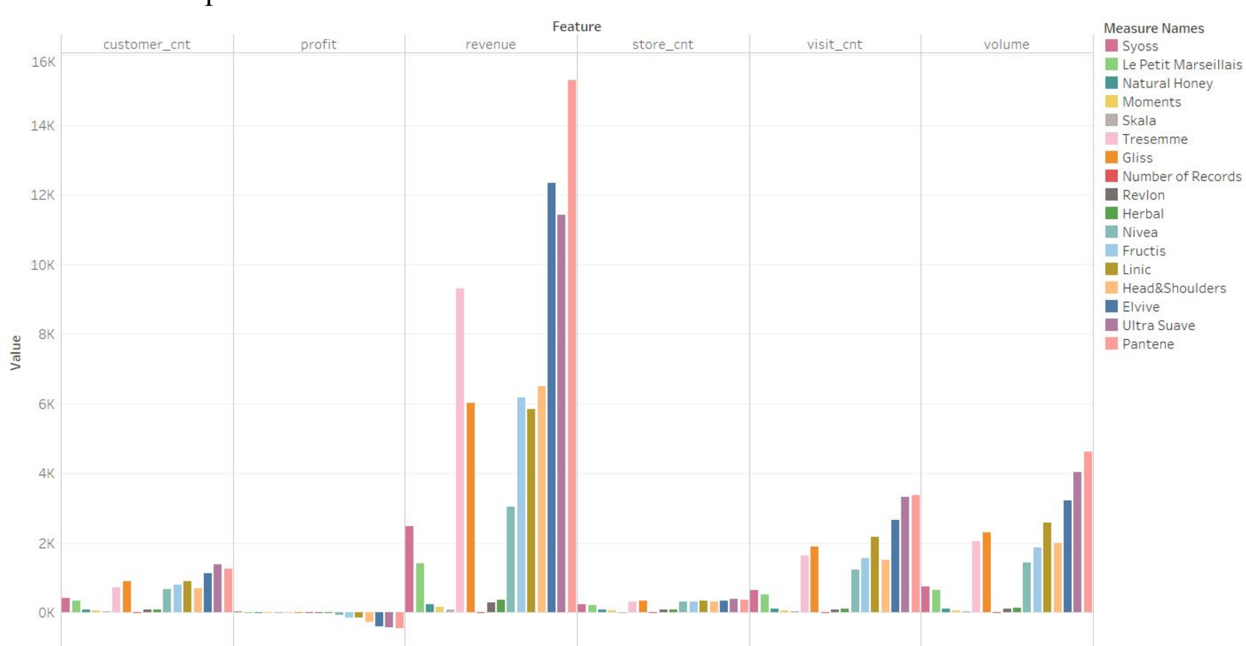

Fig17. Brand priority of shampoo/hair conditioners discount lovers in terms of the benefit of selling the brand itself.

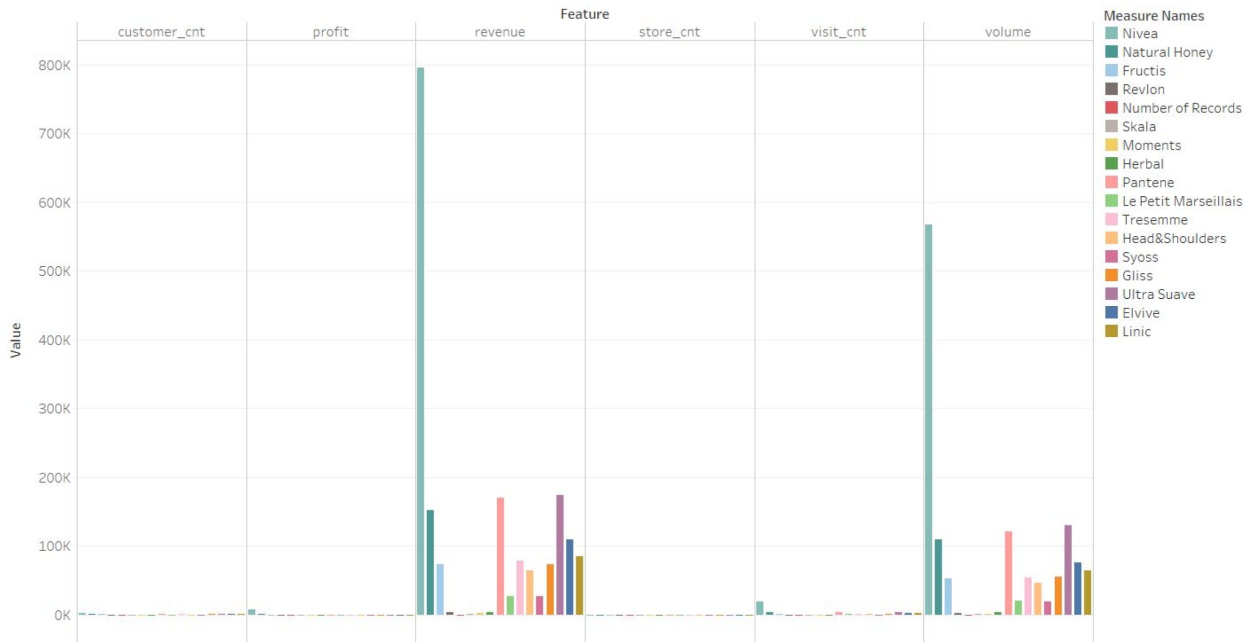

Fig18. Brand priority of shampoo/hair conditioners discount lovers in terms of the halo effect.

For cluster 2, Syoss is the best in terms of the profit of selling the brand itself, while Nivea is the best in terms of the halo effect, no matter in terms of driving profit or revenue.

So it is necessary to compare which effect we should care more, selling the brand itself or the halo effect. Let's take a look at only the profit of the four graphs.

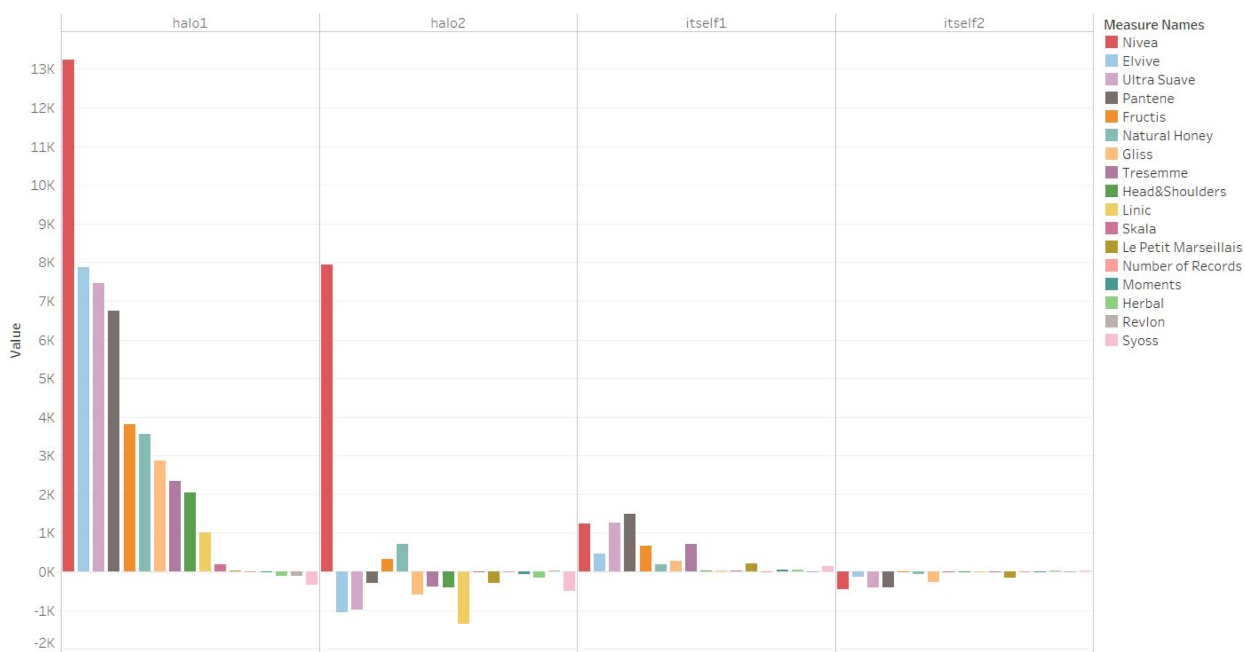

Fig19. Comparison of the four kinds of profits: brand itself of cluster1, halo effect of cluster 1, brand itself of cluster 2, halo effect of cluster 2.

As you can see, the profit of the halo effect is on average much larger than that of selling the brand itself. Of course it is because the profit of the halo effect consists of all the products purchased along with the brand. In the sense of the variance inside each kind of profit, for the two halo ones, Nivea outperforms other brands a lot in terms of the halo effect for both cluster 1 and 2, creating a large variance, whereas most of the profit of the brands themselves are pretty similar (and on a small scale). Especially considering cluster 2 which consists of discount lovers, obviously we can hardly make any profit out of them. Most of the profits of the halo effect for those customers are even negative. So it is reasonable that we value the halo effect of Nivea outperforming other brands. Taking a step back, shampoo and hair conditioners are never great profit makers as they are necessities and are usually the traffic drivers for the stores.

In conclusion, we will invite Nivea to participate in a personalized promotion campaign for our target customers decided in the previous section.

5. Personalized Product Recommendation

Before building a recommender system for our target customers, we have to exclude those customers who purchased Pernalonga's private label more than Nivea in the two target clusters. We cannot do so in Chapter 2 because we only want to exclude those who love the private label more than our target brand, which is determined in Chapter 3. It is like a loop where we came up with baseline customers to do clustering and determine the related brand, and then go back to adjust the customers. With this being done, we excluded 173 private-label-loving customers from the 3452 target customers, and kept 663 customers remaining in the first cluster, and 2616 customers in the second. The amount of target customers we have left is therefore 3279.

After segmenting out our target customers and picking out Nivea to be the brand we are going to invite, we can now build our recommender system in order to promote a Nivea product to each of the target customers. When building the recommender system, we would provide a recommendation of Nivea products based on a customer's previous transactions regarding shampoo and hair conditioner products. We need to identify the specific Nivea product in the shampoo and hair conditioner category that each of our target customers is more likely to purchase by evaluating their transaction history.

We have done some data-preprocessing in order to implement the recommender system. We first created a dataframe containing all the customer ID of our target customers and the corresponding product ID of the products they purchased. These two columns need to be converted into factors. The dataframe also contains the corresponding quantity they purchased, which is a numeric column. We then convert the dataframe into a 'realRatingMatrix', which is similar to an affiliation matrix, in order to build the recommender model. The matrix has customer ID as row names and product ID as column names, with transaction quantity being the value. All the products that have been purchased by a customer have a corresponding non-zero value.

The recommender algorithm we picked is Matrix factorization with LIBMF (LIBMF). The model will generate a list of ratings for each of the 12 Nivea products within the category for each of the target customers. Since the ratings reflect the customer's individual preference on those different products, we will then pick out the highest rated product for each customer in our target set to be the recommended product. In the following section, we will discuss the discount rate we are going to give to each of the target customers and the validity of our personalized promotion.

Discount Rates for Each Customer

We plan to give different target customers different discount rates during our personalized promotion campaign. As illustrated in the previous section, there are two clusters of customers that we are targeting. Cluster 1 contains all the ‘Shampoo Lovers’, where they are likely to buy shampoo disregarding whether there is a promotion or not. Cluster 2 contains all the ‘Shampoo Discount Lovers’, where they tend to buy shampoo and hair conditioners under discount. For all the customers in cluster 1, since they rarely care about the level of promotion, we will provide them with the minimal discount rate they ever received in their previous transactions related to shampoo and hair conditioners. On the other hand, for all the customers in cluster 2, since they tend to buy shampoo and hair conditioners under promotion, we will provide them with the average discount rate that they received previously in the purchase of similar products.

6. Cost-Benefit Analysis

Since the suppliers may not pay for the promotion, we need to perform a cost-benefit analysis to justify the practicability of our personalized promotion campaign. We consider the cost to be the expected total redemption cost for the promotion, which is the same as the total discounts redeemed. The benefit would be estimated by calculating the expected profits generated by our personalized promotion campaign.

Before calculating the extra profits we are going to gain from the campaign, we need to project the expected sales. There are two scenarios: expected sales if we do not have the promotion and expected sales if we do. For the first scenario, we identified if Nivea is their most frequently purchased brand among other brands of shampoo and air conditioners. If the answer is yes, we presume they will continue to buy Nivea even if we do not hold the campaign. We could get the expected sales without the promotion by calculating the sales brought by those customers. For the second scenario, we built two generalised linear models, one for each cluster, using the discount rate as the independent variable and expected sales quantity as the dependent variable. These models will predict what the expected sales for each cluster will be after we launched the campaign. The incremental volume is then calculated by using the expected sales quantity with the promotion minus the expected sales without the promotion. The expected incremental volume for each product is shown below:

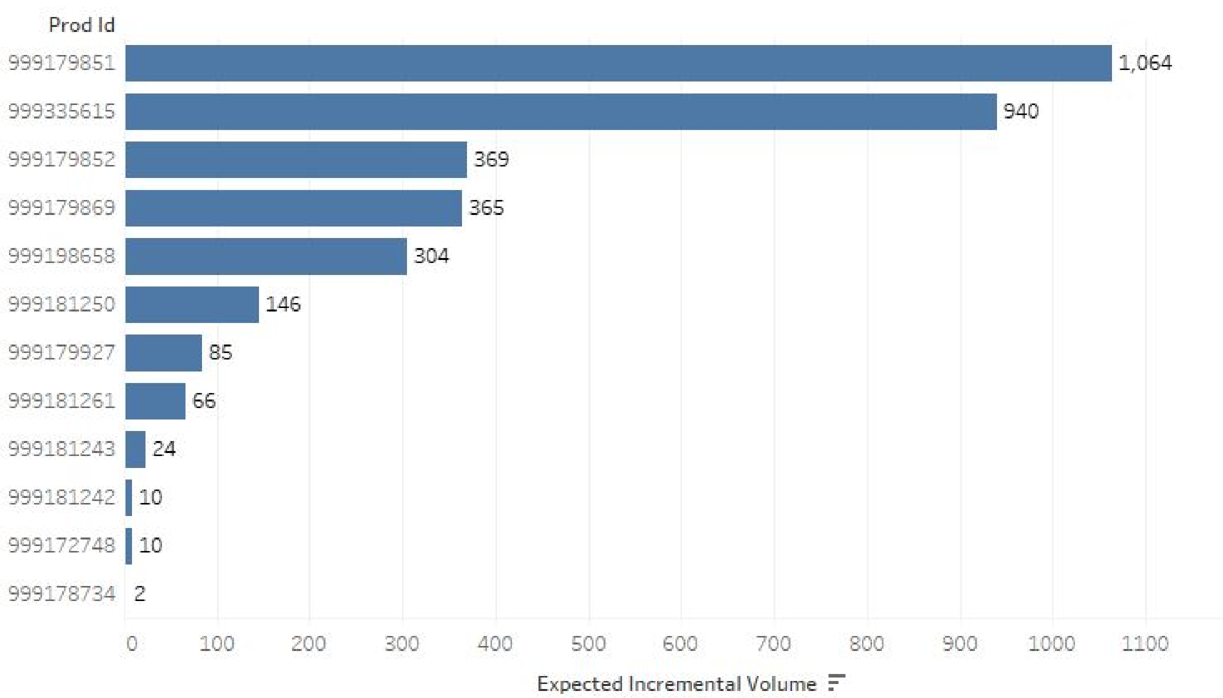

Fig20. Expected Incremental Volume by Produce

Product ID	Incremental Volume
999179851	1064.33129
999335615	940.441471
999179852	369.182384
999179869	365.035516
999198658	304.120609
999181250	146.04423
999179927	84.698533
999181261	65.989659
999181243	23.652817
999181242	9.696146
999172748	9.694639
999178734	2.148865

Based on our assumption, if we do not hold this campaign, we are going to sell only 159 Nivea products. However, after the personalized promotion campaign, we will be able to sell 3541 Nivea products based on the projection of the model. This will bring us a profit of $1144.287 from selling Nivea itself. In addition, we should consider the halo effect Nivea would bring to our sales. Based on the transaction history of our target customers, the average profits we gained from their purchasing of other products along with Nivea products are about 70 times more than the profits that brought by the selling of Nivea product itself. As a result, we will generate a profit of approximately 1144*71 = $80,000 from the two-week campaign.

We can also calculate the expected total redemption cost for the promotion, which is the total discounts redeemed, by summing up the discounts we are giving out to each individual target customer. The total cost for the campaign is $4326.637 (consider only the redemption cost for the promotion), which is far lower than the expected profits we are going to generate for the stores. Therefore, this idea of inviting Nivea from the Shampoo and Hair Conditioner categories to participate in a personalized promotion campaign is not only viable but also lucrative.

7. Conclusion

Summary

Let's recall the whole process again. First, two groups of customers are targeted by the campaign: 663 shampoo or hair conditioner product lovers and 2616 shampoo or hair conditioner discount lovers, 3279 customers in total. Second, we decided to invite Nivea based on the two clusters of customers, and Nivea attracted us with its excellent halo effect in terms of profit. Third, we recommended each customer one of the 12 Nivea products sold in Pernalonga using a "Parallel Matrix Factorization" recommender system based on the customers' transaction records, and gave each of them a personalized discount based on his cluster characteristics and his personal discount preference. Then, we estimated the expected purchasing amount of the product for each customer using 2 glm models, one for each cluster. Finally, we estimated that the campaign will cost $4,327 in redeemed discount, and achieve $80,000 overall in profits consisting of both selling Nivea products themselves and the halo effect.

In conclusion, we advise Pernalonga to invite Nivea to participate in the personalized promotion campaign, representing the Shampoo and Hair Conditioner categories.

Future steps

To further develop the project, we will consider the following aspects:

1. Consider finer clustering, and consider those who have never purchased shampoo or hair conditioner before;
2. Test more recommender systems;
3. Estimate more accurately regarding how many times customers will purchase our products within the two weeks of the promotion campaign.

Those ideas do not influence our conclusion much since we have already considered most of the aspects regarding the project. However, they could potentially offer a more thoughtful plan, provide a more careful targeting of customers, products, and expectations of cost and benefit. As a result, we would like to consider them in the future.