MKT680 – Project 2
Recommender System

Report on Personalized Soft Drink Promotions

Team 6:

Daniel Byun

Ricky Chen

Jayson Faulds

May Shao

Contents

1. Introduction

Pernalonga, a successful marketing chain operating over four hundred stores across the region of Lunitunia, is looking to implement personalized promotions to offer customers. These promotions are intended to be used as a more efficient manner of offering discounts to customers, rather than the uniform store-wide promotions that we tend to see regularly. We are tasked with developing personalized promotions to potential customers for Pernalonga's soft drink options. More specifically, we must create a full promotion campaign for both Coke and Pepsi Cola, and ultimately decide which campaign should be rolled out for a two-week window. In this report that follows, we will walk through our methodology and assumptions for building these two campaigns, and provide the full personalized promotion in detail: All customers that will be targeted by the campaign, the soft drink product being offered, and the discount price assigned to each customer. We will also generate an estimate of the total discount redemption cost along with the incremental volume as a result of our campaign.

2. Methodology

To tackle this problem, we first develop an understanding of Pernalonga's customer base, and place them into distinct groups. The first partition group is between cherry-pickers and non-cherry pickers (Group 1). In this case, a cherry-picker is a customer who frequently purchases items on discount or promotion. This includes all items, not just soft drinks. The cherry-picker group will be further partitioned into two more groups: cherry-pickers that buy soft drinks (Group 2), and those that do not (Group 3). We will treat those customers who frequently purchase soft drinks on discount differently from cherry pickers that do not have a history of buying soft drinks.

Next, we take the customers that make up Group 2 (Cherry-pickers that purchase soft drinks frequently), and then perform a clustering algorithm to split the group into different segments. These different segments constitute different customers based on their buying patterns and trends. We then pivot to focus on the cherry pickers who don't buy soft drinks. We want to find out which of these customers can be convinced to purchase soft drinks after receiving a personalized promotion that we set. To do this, we will calculate similarity measures between members of Group 2 and the different segments that we constructed in Group 3. If a member is very similar to one of our clusters, their personalized discount (and projected purchasing quantity) will be impacted by that cluster.

Through this method, we will be targeting all of the customers with discounts that we think are likely to utilize them. After creating this list of customers and their associated discounts, we then calculate our forecasted revenue generated from the promotion plan, along with total projected redemption costs on the promotion. Of course, we will engage in this entire process twice, once for Coke products and once for Pepsi.

3. Analysis

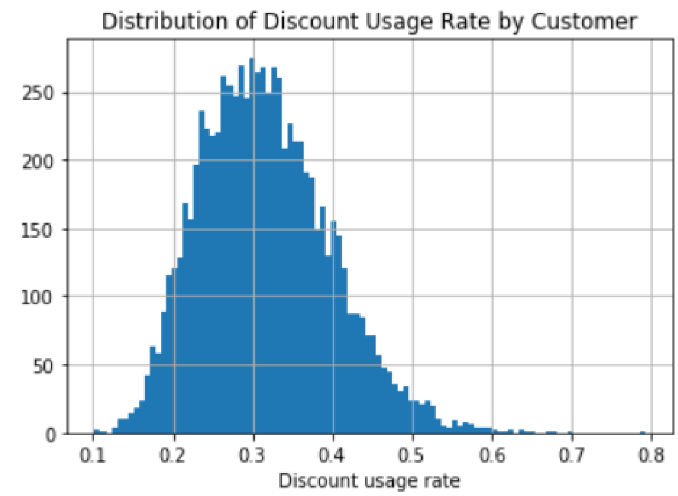
3.1 Target Customer Analysis

3.1.1. Identifying Target Customers

To begin, we filter out those customers who purchase the Pernalonga brand of cola, as we are not interested in these customers. Out of the original 7,920 unique customers of Pernalonga, we are left with 4,432 customers after the filtering.

The next task is to define what a cherry-picker is, so that we may further minimize our customer set. Targeting on cherry-pickers will make the campaign more effective because cherry-pickers are more sensitive to promotion, while non-cherry-pickers are more likely to purchase the product at full price. Pernalonga can avoid unnecessary loss by only giving promotions to the cherry-pickers that would not purchase cola otherwise.

Considering we only have these customers' transaction histories in Pernalonga's stores, we are unable to identify whether a customer tends to purchase items having lower prices at other grocery stores. Instead, we define cherry-pickers by how often they buy products on promotion. For each customer, we calculate their discount usage rate, which is the proportion of items purchased on discount out of the total number of items purchased. This value will range from 0 to 1, with a high value indicating the customer purchases a lot of their items on discount. Looking at the distribution of discount usage rate, we decided to use 0.4 as a threshold value for cherry-pickers. This designation leaves us with 749 cherry-pickers. Using 0.4 as the threshold was arbitrary, however raising the threshold too high severely limits the number of customers we are willing to target with promotion.



After identifying the cherry-pickers, we further classify them into two groups: cherry-pickers that buy Coke (or Pepsi) (Group 2), and those that do not (Group 3). Based on past transaction history, we find that 60.75% of the 749 cherry-pickers are loyal to Coke while only 1.07% are loyal to Pepsi.

After we classify the cherry-pickers into those who buy Coke (or Pepsi) (Group 2) and those who don't buy (Group 3), we have the foundation to finally decide our target customers for the personalized promotion. Because members of Group 2 already purchase cola, we will include these customers in the personalized promotion. However, we will stick with the discounts they tend to spend on already; there is no need to give them a promotion for even lower price when they already have a history of purchasing at certain price levels. For example, if a customer tends to purchase Coke when it is on discount for $1, we will not offer them a personalized promotion below $1. Thus, all members of Group 2 will receive a personalized promotion based entirely on their purchasing history.

Next, we focus on the members of Group 3 (Cherry-pickers that don’t buy Coke/Pepsi). These members display cherry-picker tendencies but haven’t yet made the jump to purchase Coke or Pepsi, even when those products may have been on discount. It is possible they did not go to one of our stores during a discount period, so we believe it is possible they may jump on board upon receiving a personalized promotion.

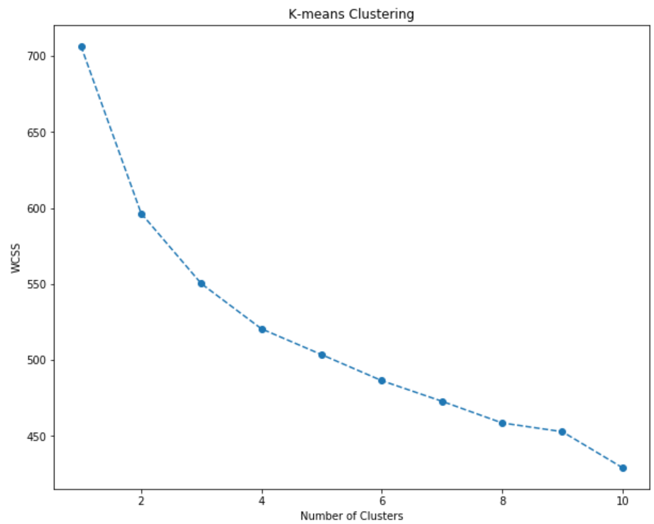
Our Coke and Pepsi buyers may display different types of characteristics, which can lead to different buying patterns. When offering a promotion to the members of Group 3, we want to anticipate what kind of buying patterns they will demonstrate, should they take advantage of the promotion. To capture these different groups and their buying patterns, we run a clustering algorithm to segment the populations of Group 2. If a member of Group 3 resembles one of our cluster groups, we will base their promotion and anticipated quantity purchased on the attributes of that cluster.

To summarize, our target customers consist of two groups: cherry-pickers who already have bought the cola brand in the past (Group 2), and cherry-pickers who we predict will buy this cola brand (Part of Group 3) upon receiving a promotion.

3.1.2 Clustering the Existing Customers

The prediction method we use to identify those with a high purchase tendency is the cosine similarity. The general process is as follows: for a certain cola brand, like Pepsi, we first categorize the Pepsi buyers who are also cherry-pickers (Group 2) into several clusters, based on their purchasing history of non-cola products. Then, we calculate the average purchasing habits of these non-cola products within each cluster. Finally, we calculate the cosine similarity between each member of Group 3 and the average purchasing habits of each cluster. Each member of Group 3, then, will have similarity statistics for each cluster. We use a threshold value for similarity to determine if the member will be placed into that cluster.

For the first step, we make clusters/segments for Group 2, aiming to identify several purchasing patterns inherent within these segments. For Pepsi, we have 175 buyers in Group 2. For these 175 buyers, we identify the thirty non-cola products that are most frequently purchased. We use these thirty products to build a customer-product matrix, where a customer-product combination gets a 1 if the customer has purchased that product, and a 0 otherwise. These 0/1 variables are used as input features for the K-means clustering model. We use the elbow rule to decide the intended number of clusters. According to the scree plot below, we pick 5 as the number of clusters, and the number of customers within each cluster are listed in the table below.



Cluster	# of customers within the cluster
Cluster 1	23
Cluster 2	24
Cluster 3	41
Cluster 4	12
Cluster 5	75

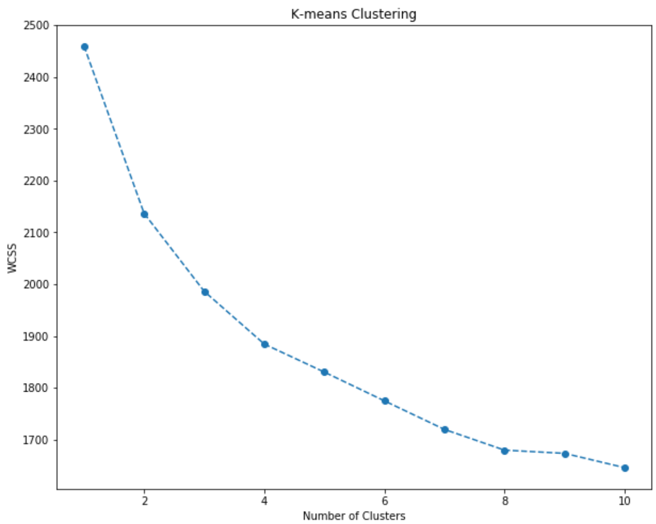
Size of each cluster for Pepsi cherry-pickers

From the table above, we can see that when we split the existing customers into 5 clusters, each cluster has enough customers to generate a purchasing pattern for the 30 high-frequency products we selected. The final step for the clustering is to calculate the average purchasing preference for the 30 products within these 5 clusters, which is shown below. For products with values close to 1.0, this means that the product is purchased by a large percentage of customers within that particular segment.

Segment K-means	**999231999**	**999956795**	**999361204**	**999746519**	**999401572**	**999401500**	**999455829**	**...**
0	1	1	1	0.478261	0.652174	0.652174	0.869565	...
1	0.958333	1	0.916667	0.791667	0.541667	0.708333	0.75	...
2	0.95122	0.97561	1	0.560976	0.902439	0.902439	0.902439	...
3	1	0.333333	0.25	0.166667	0.666667	0.416667	0.583333	...
4	0.986667	1	0.986667	0.746667	0.92	0.813333	1	...

Each cell refers to the proportion of segment members purchasing that product

For Coke, the process is very similar to what we just did for Pepsi. Coke seems to have a bigger following than Pepsi, as we have 622 customers that we place into Group 3. We then classify these 622 customers into clusters. The scree plot below is used to determine the cluster number, which ultimately leads to five different clusters as well. Alongside the scree plot, we also show the corresponding cluster size table and purchasing patterns table for Coke products.



Cluster	# of customers within the cluster
Cluster 1	124
Cluster 2	94
Cluster 3	301
Cluster 4	73
Cluster 5	30

Size of each cluster for Pepsi cherry-pickers

Segment K-means	999956795	999231999	999361204	999746519	999401500	999455829	999953571	...
0	0.991935	0.975806	1	0.629032	0.693548	0.903226	0.959677	...
1	1	0.978723	0.989362	0.680851	0	0.93617	0.861702	...
2	0.993355	0.973422	0.990033	0.700997	1	0.973422	0.92691	...
3	0.986301	0.958904	0.958904	0.465753	0.589041	0.410959	0.794521	...
4	0.066667	0.966667	0.033333	0.1	0.433333	0.533333	0.9	...

Each cell refers to the proportion of segment members purchasing that product

3.1.3 Appoint potential cola brand customers to an appropriate cluster

Now that we have our clusters for both Pepsi and Coke cherry-pickers, our next task is to calculate cosine similarity statistics between each cluster and the members of Group 3 that don't purchase cola. The general process is:

1. Calculate the cosine similarity between each customer and the purchase preferences of each cluster (Shown in the tables above)
2. Find the cluster to which each customer is most similar
3. Compare the similarity measure with some threshold to determine whether the customer is assigned to the cluster.

Cosine similarity was chosen for a few reasons, the first of which is purely due to the ease of using this method to calculate the similarity statistics. Additionally, it can take every feature into account (Products purchased, in this case), as opposed to Pearson correlation similarity, which omits those features that have missing entries in certain cells. In our case, not buying some product is related to the similarity between a potential customer and the average purchase tendency of existing customers' clusters. Additionally, we assume that a purchase of a specific item made by different people will have similar attributes and purchase tendencies. Therefore, we found it to be superfluous to subtract the average purchase tendency for each customer. Based on these attributes, we use cosine similarity to determine whether a potential customer belongs to a Group 3 cluster.

For Pepsi, we have 574 potential customers to be placed into a cluster. By comparing these customers' purchasing histories with the cluster's average purchase preference, we generate 5 similarity scores for each customer, shown in the table below. We then find the one cluster to which they are most similar and identify which cluster this is in the final column. Something important to note here is that the purchasing history we are using is not the full product portfolio for each customer. These data were pulled from a two-week sample from April of the previous year. This is to simulate as close as possible the time and seasonal effects for soft drink (And other products) purchases during the month of April. Because our intended promotion is taking place in April, it makes sense to pull data from that month to make our estimates.

cust_id	Cluster 0 Similarity	Cluster 1 Similarity	Cluster 2 Similarity	Cluster 3 Similarity	Cluster 4 Similarity	max_value	max_index
229838	0.936721	0.902339	0.953076	0.75041	0.9385	0.953076	2
339665	0.915231	0.925803	0.924487	0.75464	0.953112	0.953112	4
399566	0.827075	0.773985	0.742957	0.553399	0.744052	0.827075	0
529915	0.958866	0.97395	0.978314	0.785196	0.998101	0.998101	4
749556	0.926583	0.91168	0.955072	0.744208	0.932868	0.955072	2
789747	0.877895	0.861405	0.881887	0.729412	0.908609	0.908609	4
819740	0.936862	0.953452	0.959131	0.74577	0.981228	0.981228	4
1129755	0.84504	0.856096	0.855228	0.751694	0.872081	0.872081	4
1369928	0.9054	0.864041	0.907221	0.679162	0.883513	0.907221	2
1539975	0.954969	0.930362	0.947088	0.688393	0.931332	0.954969	0

Potential Pepsi customers' similarity scores to each cluster

We set a threshold for each cluster and filter out those who are not similar enough to current customers. 432 customers are left after the filtering, and placed into their subsequent clusters. We combine these customers with the other Pepsi buyers from Group 2; this is our list of customers that we would target for the Pepsi promotion. The total number of our target customers for Pepsi is 607.

For Coke, we follow the same process just described above for Pepsi. We have 127 cherry-pickers who do not buy Coke currently. After we filter by their cosine similarity to existing buyers, 93 customers are included as our target customers. Along with these 93 customers, we also target the Group 3 members for Coke, for a total of 715 customers to be targeted for Coke promotion. An excerpt of the two tables are shown below:

cust_id	Segment K-means	pepsi-buyer
1109886	0	1
1249961	4	1
1609938	1	1
1639780	3	1
2009681	2	1
2689745	2	1
2899928	4	1
3259598	2	1
4229650	0	1
5529802	4	1

Pepsi target customers

cust_id	Segment K-means	coke-buyer
229838	1	1
339665	1	1
529915	2	1
749556	2	1
819740	2	1
1109886	0	1
1129755	3	1
1249961	2	1
1539975	0	1
1609938	2	1

Coke target customers

3.2 Promotion Design

To design the personalized promotion to assign to each of our targeted customers, we take a cluster-level approach. We begin by focusing on the Group 3 customers that were placed into clusters based on cosine similarity in the previous section. The first task is to decide which product to recommend/promote to those Group 3 customers placed in a cluster. There are 21 unique Coke products, and 5 unique Pepsi products. For each cluster, we identify the most frequent-purchased product, and this is the product we will promote to the new customers.

After selecting the product to recommend, we calculate the cluster-level quantity and discount price by aggregating and finding the averages for the cluster. This is the "average person" within that cluster. We believe the newly-placed customers are similar to this average person.

Ideally, we would give each customer this average discount price as the promotion, and we would expect them to purchase roughly the same quantity as the average. However, if this were truly the case, all of these newly-placed customers would have been part of the cluster to begin with. In other words, these customers would be in Group 2 instead of Group 3. Thus, these Group 3 customers need a slightly bigger push than those customers that already purchase Coke/Pepsi on discount. Otherwise, they would have purchased Coke/Pepsi in the past already.

To make our calculations a bit more realistic and conservative, then, we dampen the discount and expected quantity for these Group 3 customers. This involved taking an extra 10% off of the discount offered, and reducing the expected purchase quantity of each customer by 25%.

For the Group 2 customers that have a previous history purchasing soft drinks of either brand, they receive the promotions that they have used in the past. A customer who consistently purchased Coke for $1 on discount in the past will continue to receive these same promotions, the only change is that they will be personalized rather than store-wide. Because we are holding their discounts constant, we expect the revenue generated from these customers to remain consistent during the two-week promotion.

This is how the Group 2 and Group 3 customers will be targeted differently by our personalized promotion. Another difference that we held between the two groups is the products offered on promotion. If you recall, we are only offering the most-frequently purchased product in each cluster to the Group 3 customers getting placed into that cluster. Thus, each Group 3 member is getting offered one product on promotion. On the other hand, Group 2 members that have a history purchasing soft drinks at discount will be offered the same products they usually buy on discount. So if a customer has purchased three different Coke products, we will have those three products available with a discount for their personalized promotion.

With all of this foundation established, it is now time to generate our estimates of performance during the two-week promotion period. This involves a simple calculation, as we already have found the expected quantity for each customer along with the discount price being offered. We expect the total revenue generated from the promotion applied to Coke to be approximately $23,000. As a form of comparison, the store-wide promotions used for Coke products made approximately $22,700. Thus, the incremental revenue provided from the Coke promotion is estimated to be around $300 ($263 to be specific) over the two-week period. We expect the total redemption cost and incremental volume sold to be $6,714 and 209 Cokes respectively.

After applying the same revenue calculation process from the Coke example, we were able to calculate Pepsi's new revenue value. The total revenue for Pepsi came out to be ~$2145.50 with an incremental revenue value of ~$607.90. Although the total revenue is abysmal compared to Coke's numbers, the incremental value for Pepsi is a little less than 3X of Coke's. In other words, Pepsi brands will increase its original revenue by around $600 through personalized promotional efforts for this brand. This is $300 greater than what Coke will experience from personalized promotions and this can also be explained when looking at the redemption costs as Pepsi would have to expend more on discounts for more purchases. This then relates to the volume increase as well. However, it is also paramount that we do not double count the revenue contributed to Pepsi and the revenue loss to Coke in our total revenue generation calculation. This is because there is a strong chance that a lot of the incremental revenue is a result of customer churn from one soda brand to the other, and not an increase in new soft-beverage consumers.

Therefore, we also calculated the overlap between the two strategies by calculating the opportunity costs for each promotion. In our discoveries, the opportunity cost for Coke promotions comes out to be $19.45, whereas the opportunity cost for Pepsi promotions is $9970.55. Hence, the total incremental revenue for Coke promotions is ($263-$19.45) or $243.55 and for Pepsi promotions is ($600-$9970.55) or -$9370.55. The reason Pepsi's numbers are negative is that while they are increasing revenue from Pepsi purchases, the discount may also be purging Coke customers, cancelling out this revenue from lost Coke sales. We believe this number for Pepsi is indeed inflated, as we posit customers simply purchase more Coke than Pepsi at bulk. Regardless, once increased revenues and opportunity costs are considered, it is evident that Coke is the better brand to promote partly because it would cannibalize less of Pepsi's original revenue stream than vice versa.

	Total Revenue	**Incremental Revenue**	**Total Redemption Cost**	**Incremental Volume**	**Overlap Value**
Coke	$22927.02	$263.43	$6714	209	19.45
Pepsi	$2145.50	$607.90	$815.59	503.79	9370.55

4. Conclusion & Recommendations

To recap our progress and methodology, we began with the full list of unique customers. We focused on those customers that were cherry-pickers who do not purchase the Pernalonga brand cola. We split these cherry-pickers into the soft drink-buyers (Group 2) and the non soft-drink buyers (Group 3). We perform clustering on Group 2, and made five different groups for both Pepsi and Coke. Members of Group 3 were placed into these clusters based on their cosine similarity. Their placement into a cluster had a significant impact on the discount price that was offered to them. The members of Group 2 received their promotions a little differently, in an effort to maintain the volume purchased from them. At this point, we were able to calculate the expected incremental revenue for both promotion plans and select the best option.

Ultimately, our final conclusion is that Pernalonga should go with Coke as its soft drink of choice for personalized promotions. To reach the numbers that we got above, we did make an important assumption: while the personalized promotion is going on, the regular, store-wide promotions will still persist for the other brand of soda. So, by picking Coke, the Coke products will be on sale via personalized promotion, while Pepsi products will receive the regular promotion accessible to all customers that attend the store.

We reached this conclusion by considering both revenue increases for both scenarios and the opportunity costs associated with each. What we found is that although Pepsi will boast an increase in revenue by three times as large and a volume that is more than twice as large as Coke would, the opportunity cost to promote Pepsi is far too expensive to find this strategy lucrative. In fact, Pernalonga will only gain a net positive in incremental revenues if it follows through with Coke's personalized promotions, unless we restrict our promotion to only non-coke-drinkers. Our analysis on the values generated from our model has the potential to lead us to many different conclusions, but after looking at margins between the incremental revenue and the opportunity costs, Coke is the ideal candidate for the personalized promotions.