Recommender Systems For Coca-Cola vs Pepsi-Cola At Pernalonga

Team 6

Joyce Zhang, Nicole Santolalla, Justin Wood

Table of Contents

Introduction……..2

Background……...2

Proposed approach……3

Methodology……3

Data exploration……3

Identify eligible products……..5

Identify target customers……...7

Recommender System……7

Conclusion & Next Steps……8

Introduction

Background

Pernoalonga, a leading supermarket chain with over 400 stores in Lunitunia, sells over 10 thousand products throughout 400 categories. Our goal is to experiment on performing personalized promotions from data given on two products, Coca-Cola and Pepsi-Cola. While the details of specific partnerships with the suppliers to fund the experimental personalized promotions are being negotiated, we will use recommender systems to determine if we should choose Coca-Cola or Pepsi-Cola for personalized promotions at Pernalonga.

In order to answer this question, we first need to understand the market and our customers. We conducted market research on both Coca-Cola and Pepsi-Cola in Portuguese speaking countries. We found that Coca-Cola dominates the carbonated beverage category worldwide. In 2019, Coca-Cola's global market share was 48.6% compared to Pepsi-Cola's 20.5%[1]. If we wanted to compare this statistic to our Pernalonga dataset, the market share of Coca-Cola and Pepsi-Cola over the two years provided is 76% and 5%, respectively. We divided the total sales of each product by the "tran_prod_paid_amt", or *revenue,* of the "Cola Drink" category. This initial research shows signs that Coca-Cola is more popular than Pepsi-Cola and has a high brand value in Portguese speaking countries. In Brazil (Portugeuse speaking country), Coca-Cola is ranked no.2 as brand favorites over no.5 Pepsi-Cola[2]. Brazilian's search "Coca-Cola" 43,163 times on an average day compared to 6,074 US Americans searching for "Pepsi-Cola"[3].

We also wanted to see what the average price of a soda was in Portugal because price sensitivity will play a major role in deciding which product will be chosen for personalized promotions. The average price of a carbonated soft drink in Portugal in 2020 was $1.49, and both Coca-Cola and Pepsi-Cola were at the core of that price level[4].

---

[1] Market Share. (2015) *Market share of leading carbonated beverage companies worldwide 2015.* Retrieved from Stat Investor: https://statinvestor.com/data/5496/coca-cola-beverage-market-share/#:~:text=The%20carbonated%20soft%20drink%20market,estimated%20at%20341.6%20billion%20USD

[2] Gartner. (2016 2 8) *Top 10 Beverage Brands in Brazil.* Retrieved from Gartner Daily Insights: https://www.gartner.com/en/marketing/insights/daily-insights/top-10-beverage-brands-in-brazil

[3] Kalinax. (2015) *Pepsi vs Coca-Cola.* Retrieved from Kalinax: https://www.kalinax.com/pepsi-coca-cola.html

[4] Statista. (2021) *Consumer Markets, Soda in Portugal.* Retrieved from Statista: https://www-statista-com.proxy.library.emory.edu/outlook/20020000/147/soft-drinks/portugal

Proposed approach

To decide which brand Pernalonga should pick to implement personalized promotion campaigns, Coca-Cola or Pepsi-Cola, we will look to compare at the frequency of promotion offers from Coca-Cola and Pepsi-Cola received by target customers. We will identify the products Pernalonga should target for promotion for Coca-Cola and Pepsi-Cola based on business understanding, such as great potential in revenue and sale quantity growth if the promotions are offered. The target customers will be identified as the group who shopped in the "Cola Drink" category, sharing similar purchasing behavior as the Coca-Cola and Pepsi-Cola customers, but who have not purchased Coca-Cola or Pepsi-Cola yet. The top 1 product to promote for each customer in our target segment will be suggested using the user-based recommender system we developed from Pernalonga's data.

Methodology

Data Exploration

We explored both the transactions and products datasets, eventually merging them together for more constructive analysis. Next, we wanted to explore the "Cola Drinks" category including Coca-Cola and Pepsi-Cola. One discrepancy we found was that Coca-Cola was also included in the "NUTS" category. We omitted "NUTS" and continued our comparative analysis. We found 4 brands, "COCA-COLA", "KENDY", "PEPSI-COLA", and "PRIVATE LABEL" as seen in *figure 1*.

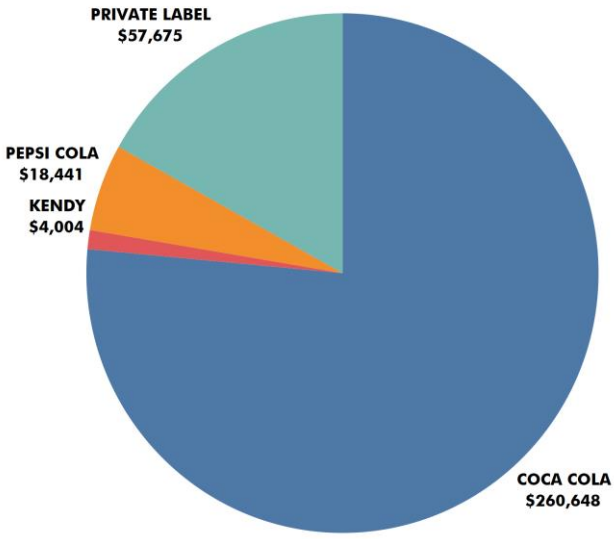

Figure 1. Brand table of cola categories by Revenue

We compared Coca-Cola and Pepsi-Cola against the whole "Cola Drink" category. Compared to the other two brands, "KENDY" and "PRIVATE LABEL", Coca-Cola leads the category with 76.5% of the revenue in the category, with Private Label in second place with 16.9% of the revenue, Pepsi-Cola third with 5.4%, and Kendy last with 1.2% of revenue. Pepsi-Cola leads the most promo proportion with 50.7% and Coca-Cola third at 17.3%. From this comparison, we can assume that Coca-Cola is not only dominant and popular in terms of revenue, but is also purchased on low promotion, which means it is purchased by some of the most loyal customers, who aren't necessarily sensitive to promotions. Pepsi-Cola's revenues and promotions were relatively low, and it can be assumed that this brand leveraged on strong promotions to acquire market share. Coca-Cola also leads the category in volume at 66% while Pepsi-Cola is much lower in comparison at 4.4%.

To get a better sense of when these products are purchased the most, we explored seasonality. December was the top revenue producer for soda beverages at 10.9%. This could be due to the holiday season with gatherings and social events that include food and beverages so there is a natural uptick in the sale of soda products. Other than December, the highest revenue comes from the months during the summer season. Especially from June, July and August with revenue at 28.6%. It can be assumed that soda provides relief from the hot weather as the soda beverages tend

to be cold, and carbonated. According to Talking Retail in Europe, "For every one-degree temperature change, soft drink sales increase by 1.6%"[5]. January and February are the two months that have the least amount of revenue at a combined 11.5%. These two months could potentially be selected for personalized promotions as the products are not producing much revenue as compared to other months.

Identifying target products

An effective approach to compare campaign results is to start with selecting target products with the same criteria for both Coca-Cola and Pepsi-Cola that benefit Pernalonga the most. We expect to see that the promotion campaigns on those products will drive greater sale quantity, revenue, and traffic for Pernalonga. Thus, we segmented the products based on four metrics.

- *Sales quantity*: the total sale quantity of the product
- *Revenue*: the total revenue generated by the product
- *Number of Transactions*: the count of frequency the product occurs in each transaction
- *Promotion Discount*: the average discount of each product

Based on the Pernalonga transaction data, Coca-Cola has 21 unique products while Pepsi-Cola only has 5. First, we applied our product segmentation technique to Coca-Cola. The output resulted in 4 different clusters. Our cluster analysis showcased that cluster 1 and 3 were the best to target as seen in *figure 3*.

[5] Talking Retail. (2019 3 17) *In Focus: Summer Soft Drinks.* Retrieved from Talking Retail: https://www.talkingretail.com/advice/category-management/focus-summer-soft-drinks-17-03-2019/

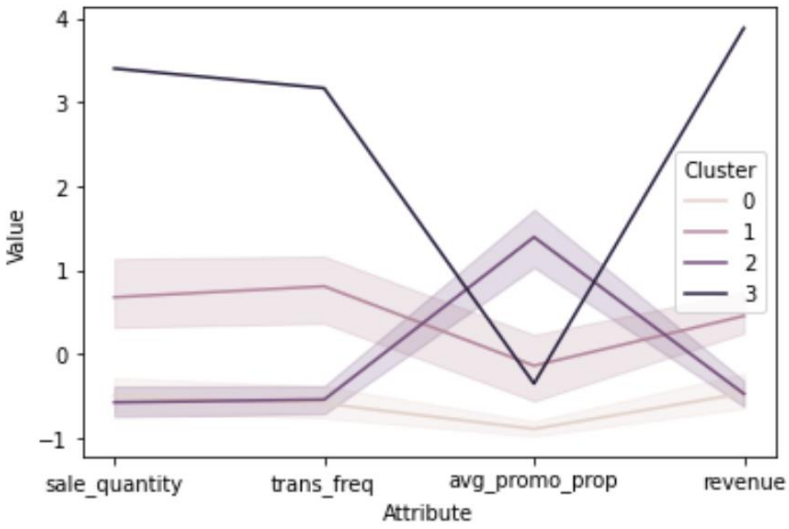

Figure 3. Coca-Cola Product Segmentation

Cluster 1 included 6 products that were high on transaction frequency, even though it showcased a relatively low average promotion rate, still indicated a strong potential on high revenue growth. Cluster 3 had only 1 product, with high volume, frequency, and profit, but was low on promotion. This is the "golden" product of Coca-Cola in that it was not heavily promoted, but it was the traffic driver and generated the most revenue. It indicated that we could offer promotions on this product to further increase traffic to the store. We labeled this item as "premium" since it was a highly frequent bought item that usually was not on promotion[6]. These two clusters are the ideal target products for personalized promotions.

Next, we applied the product segmentation to Pepsi-Cola. However, Pepsi-Cola only had 5 different products and using K-means clustering for segmentation was not possible. Instead, we analyzed these 5 products with our 4 criteria metrics to find which products to choose. We ended up using 2 products from Pepsi-Cola[7]. These two products share the similar characteristics as the target Coca-Cola products.

Identify target customers

[6] Coca-Cola Product IDs: 999153856, 999153857, 999165976, 999261787, 999329592, 999329593, 888422862

[7] Pepsi-Cola Product IDs: 999349816, 999590416

With the target products defined, we further identified our target customers. We want to apply the target products promotion to customers who have shopped in the "Cola Drink" category but who have not purchased Coca-Cola or Pepsi-Cola before. This group constitutes 49.2% or 3,561 of customers within the "Cola Drink" category. Our non-target customers will be the group that purchased Coca-Cola or Pepsi-Cola before, and we will identify them as our "baseline" customers. We evaluated three metrics to narrow down our target customer list

- *Medium to high price sensitivity*: have an average promotion discount rate over 0.155 (75% percentile) across all the transaction of each customer
- *Loyal to Cola Drinks based on revenue*: have a total revenue of $5 (75% percentile) in cola drinks
- *Share similar historical purchase behavior as the "baseline" customers through cosine similarity*: cosine similarity is calculated based on the number of quantities purchased for each product by each customer, target customers need to reach a cosine similarity score of 0.6 with "baseline" customers

We found that our 6,500 "baseline" customers contributed to 84,706 transactions in Pernalonga within the two years period. With the metrics implementation, we drilled down the number of our target customers from 3,561 to 314.

Recommender Systems

Generally, there are three types of recommender systems: content-based filtering, collaborative filtering, and model-based. Since there is not enough information about the product, we lack the characteristics of the products to perform content-based filtering. Hybrid recommender systems are less intuitive to interpret even though they prove to be robust in certain scenarios, including algorithms such as Singular Value Decomposition (SVD) and Bayesian Personalized Ranking. We decided to move forward with user-based collaborative filtering that focuses on predicting target customers' promotion acceptance on target products based on their similar customers' preference.

Based on the target products and target customers that we discovered above, we took the following steps to recommend target products to new customers:

1) Find the top 3 “baseline” customers that share similar purchasing behavior as the current target customer based on cosine similarity
2) Calculate the average ratings of similar “baseline” customers for only the target products that the current target customer has not purchased
3) Grab the top 1 product based on the average rating of the similar “baseline” customers and recommend it to the current target customer

Since we only have a total of 8 target products to implement, we only promoted the top 1 product to our target customers as it more effectively compared the campaign results between Coca-Cola and Pepsi-Cola. We chose customer id **99169981** from our target customer list and implemented the recommender system, as the top 1 product result shown in *figure 4*.

cust_id	prod_id	subcategory_id	category_id	brand_desc	category_desc_en
99169981	9992611787	93719	95807	COCA COLA	COLA DRINK

Figure 4. Product Recommendation for customer 99169981

Over the 314 target customers, we discovered that over 90% of the products through the recommender system is a Coca-Cola product. This result is anticipated as Coca-Cola is dominating the cola drink market in Portuguese speaking countries, and Coca-Cola has a much bigger profit margin and broader product choice based on Pernalonga’s transaction data.

Conclusion & Next Steps

In conclusion, choosing personalized promotions for Coca-Cola instead of Pepsi-Cola will potentially increase revenue at Pernalonga. The ideal testing phase should start during the month of January, when the sales are at their lowest and continue the campaign until the end of summer. In this way we can analyze trends throughout several months. Depending on the success of the metrics used in the campaign, such as revenue, frequency, and volume, we can scale this promotion to stores with similar metrics to increase Pernolonga’s profit.

Possible improvements

There are some concerns and considerations to address for future campaigns. Having a low sample of Pepsi-Cola data to cluster was concerning. The goal was to have appropriate comparisons for both products and recommend one. We did not eliminate the bias when choosing the target products for both campaigns, as there were much less products we could choose from Pepsi-Cola. We did notice that Private Label wasn't massive but was the next biggest competitor to Coca-Cola. If we were able to consider using Private Label instead, we might have had more compelling insights and recommendations. Also, location is an important factor that could change with different customer behaviors, purchases, and price sensitivity. Certain stores may be in more centralized areas that aren't high in household income, while others are in very populated tourist areas. Prices may vary across stores and promotion sensitivity may be different as well. Testing out personalized promotion campaigns may be good for several stores that deem identical to get similar results. However, these are some factors that would need to be addressed before scaling this promotion widespread to multiple locations. Most concerns can be addressed while we continually iterate this personalized promotion process and enhance the campaign as needed.