



EMORY

GOIZUETA
BUSINESS
SCHOOL

ISOM 680 - Marketing Analytics

Project Report 2: Recommender Systems

Team 6

Shuwen Huang, Firat Geyik, Gaows Mohammad, Cecilia Wang

1. Introduction

1.1 Background

ACSE Supermarket, a company that sells everything, has over 40 stores in Lunitunia and sells over 100 thousand products in over 100 categories. ACSE customers can opt to join the Lunie Rewards program to avail of weekly sales and promotions. ACSE regularly partners with suppliers to fund promotions and derives a significant portion 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. In theory, personalized promotions are more efficient as offers are only made to targeted individuals who require an offer to purchase a product. In contrast, most in-store promotions make temporary price reductions on a product available to all customers whether or not a customer needs the incentive to purchase the product. The efficiency of personalized promotion comes from an additional analysis required on customer transaction data to determine which customers are most likely to purchase a product to be offered in order to maximize the opportunity for incremental sales and profits.

1.2 Business Problem

Our team was selected by ACSE (the client) to develop a marketing campaign to experiment on personalized promotions. While the details of specific partnerships with suppliers to fund the experimental personalized promotions are still being negotiated, you have received data from the client. Our team was asked to analyze the data to support the proposed marketing campaign below:

Coca-Cola and Pepsi Cola are both planning personalized promotion campaigns, but ACSE can only pick one. Which brand and CSD product to promote?

For the above campaign, ACSE forbids the targeting of customers who currently buy an ACSE branded competing product. A minimum requirement is to provide (with justification) a personalized promotion plan with the following:

- The customers that will be targeted
- For each targeted customer, which product is being promoted with an offer
- The expected total redemption cost for the promotion (total discounts redeemed)
- The expected incremental volume for each product

To answer the question of which brand and CSD product to promote, we analyze respectively from product perspective and customer perspective.

2. Data Understanding

2.1 Data Preparation

There are two datasets, transactions dataset and products dataset:

- 1.transactions.csv contains transaction history from June, 2017 to December, 2020 for over 9 million customers
2. products.csv contains the product to subcategory and category mapping and descriptions for over 100,000 products

For efficiency purposes, we decided to use a sampled transaction dataset with 12,061,594 transaction records. Removing negative sales amount and negative quantity results in a dataset containing 11,907,126 transaction records and 8 columns.

Given the scope of our business problem, we would like to use two separate datasets for Coca Cola and Pepsi Cola, respectively, by filtering out product descriptions(prod_desc) for Pepsi and Coca.

In order to investigate seasonality trends, we created three new columns namely weekday, month, and year by splitting the original column "trans_dt". These datasets are ready for product-level seasonality analysis.

2.2 Product-Level Analysis

2.2.1 Seasonality Analysis

We analyzed and visualized seasonality in sales trends from July, 2017 to December, 2020 and found that there are some trends for both Pepsi Cola and Coca Cola as months progress.

Figure 1 shows seasonality trends for Pepsi Cola. As we can see from the graph, the sales fluctuates around between the start of each year till September/October, followed by an increase in sales and an obvious high peak of sales in December, which might be due to winter holidays boosting sales. However, the year of 2020 did

not follow this general trend. One potential reason for this might be the influence of COVID-19.

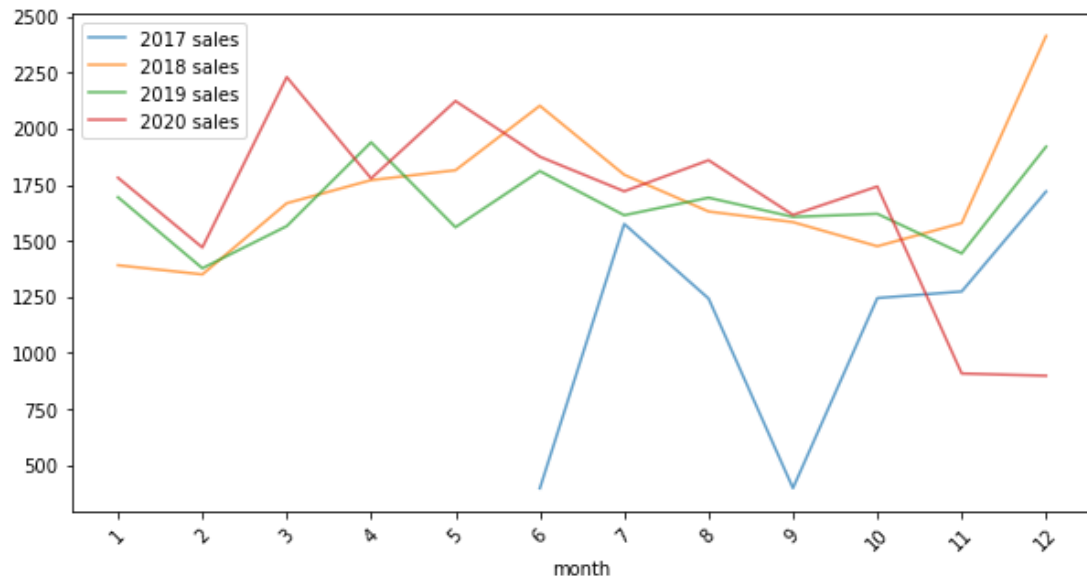


FIGURE 1: Pepsi Seasonality Sales Trends

Figure 2 shows seasonality trends for Coca-Cola. Trends are similar to what Pepsi Cola displays. However, 2020 was a relatively better year in terms of higher sales for Coca-Cola with the same dip in Q4 like we had seen for Pepsi.

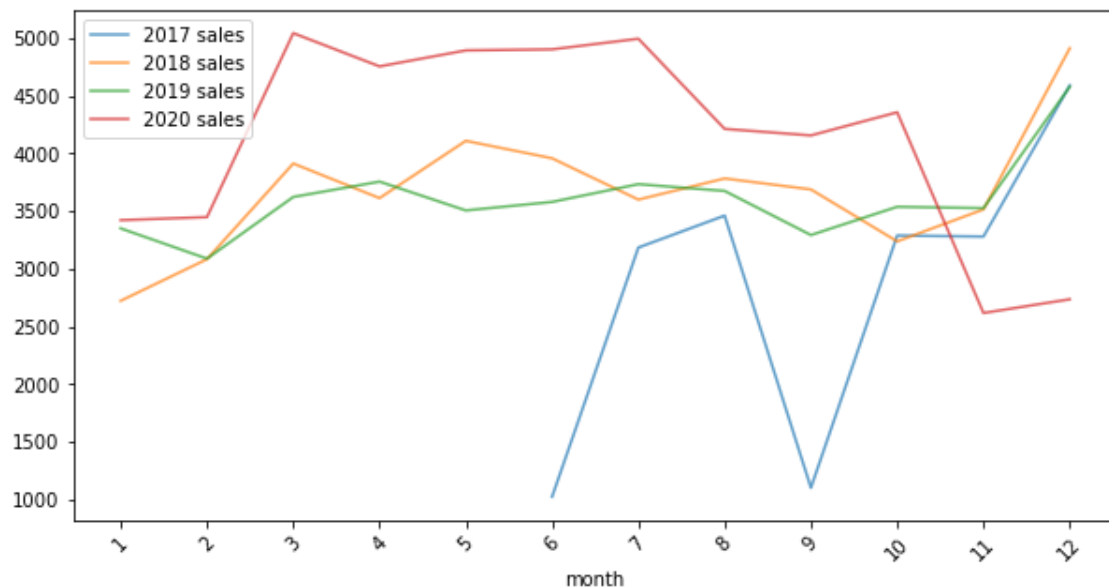


FIGURE 2: Coca Seasonality Sales Trends

After comparing those seasonality sales, we believe there are indeed trends in sales for both Coca and Pepsi, and the trends have similarities. However, for each year, the sales for Coca Cola is nearly twice of Pepsi.

2.2.2 Product Category Analysis

At the first glance, we would like to look into the specific product categories for Pepsi Cola and Coca Cola and their summary statistics. After *researching*¹ we found out that some product types are not in production anymore in the year of 2022, therefore those products would not take into account our analysis and should be removed. As Figure 3 and Figure 4 displays, there are 28 types of Pepsi Cola products (Appendix 1) and 45 types of Coca Cola products (Appendix 2). One thing to notice here is that some of the products' names are the same, however, they have different product id and product unit quantity count, therefore, at this stage, we decide to display them as different products.

In the product descriptions, there was a mix of camel case and capitalization for the same product. Some instances of the same product were hyphenated and some were not. We merged all the variations of a product description into a single category hitherto.

2.2.3 Market Share

The next step in our product analysis is to look into the market share of each product for both Pepsi Cola and Coca Cola. We would like to make a modification here, which is to categorize products by unique names. For example, product PEPSI with product unit quantity equal to 1 and PEPSI with product unit quantity equal to 12 will be categorized into the same market share proportion.

Figure 5 displays the market share for each pepsi product. The products that have the TOP 3 market shares are: Pepsi, Pepsi Diet, Pepsi Cola Diet Caffeine Free, which together accounts for around 75% of market share. Therefore, when considering which CSD products to promote, these 3 categories are the ones that we should focus on in our later analysis.

¹ Research references: <https://www.pepsicopartners.com/pepsico/en/USD/BEVERAGES/Soft-Drinks/c/softDrinks?root=beverages>
<https://www.buzzfeednews.com/article/venessawong/diet-pepsi-is-having-a-serious-identity-crisis>
https://the-soda.fandom.com/wiki/Pepsi_Berry

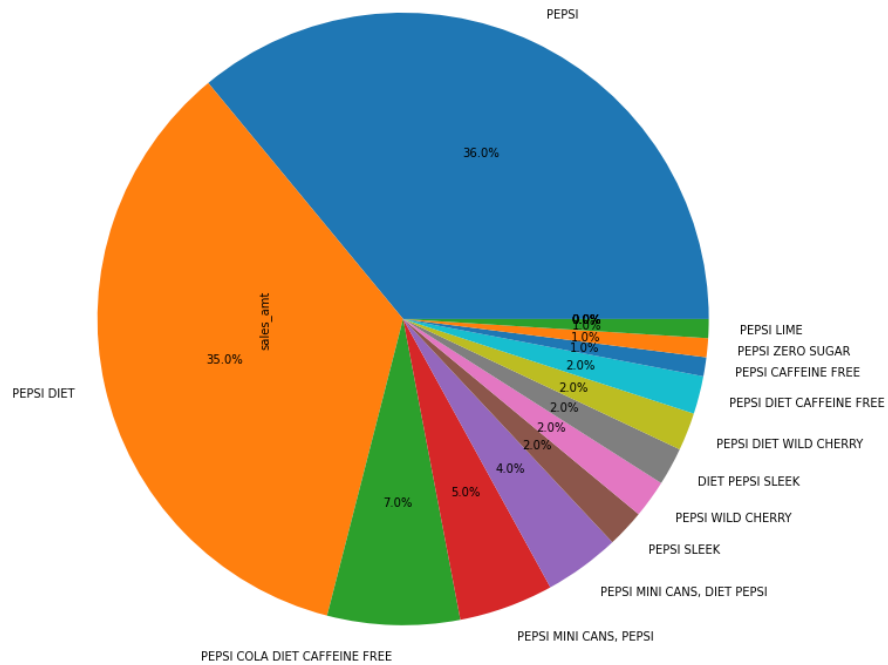


FIGURE 5: Pepsi Products Market Share

Figure 6 displays the market share for each Coca product. The 4 products Coca Cola Fridgemate, Coca Cola Zero, Coca Cola Diet and Coca Cola occupy around 85% percent of Coca Cola total market share.

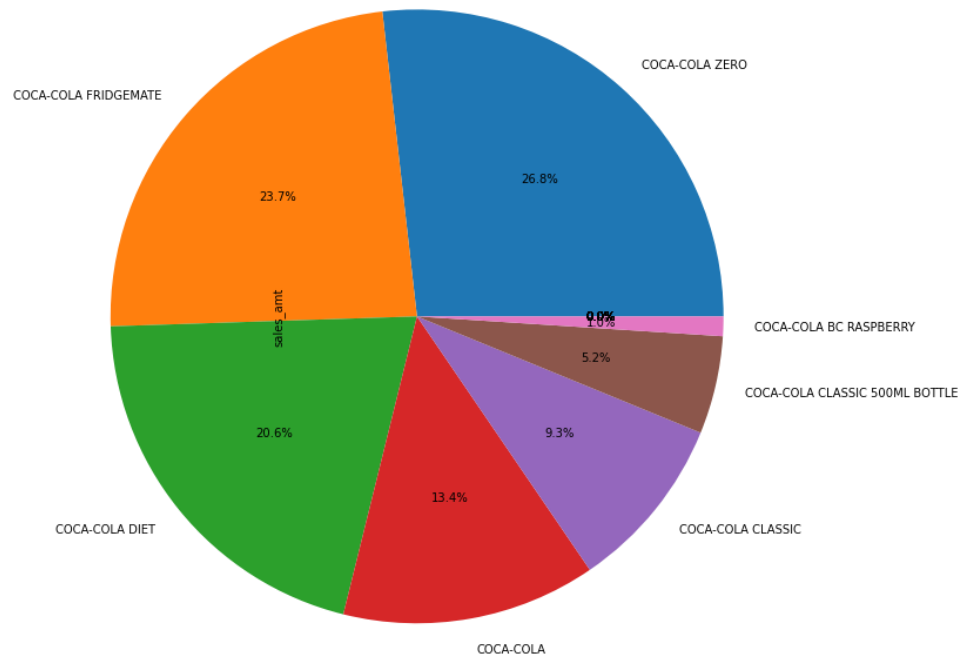


FIGURE 6: Coca Products Market Share

2.2.4 Sales by Product Category

After looking into market share of Pepsi Cola and Coca Cola products, we gained business sense of which CSD products to promote for each brand. The next step in our product analysis is to look into sales contribution by product categories for Pepsi Cola and Coca Cola. Using the pareto principle we found out that 4 categories contribute to 80% sale of Pepsi and surprisingly enough, 4 categories contribute to 80% of sales of Coca-Cola too. (Appendix 3: Pareto Analysis)

It would be interesting to explore the reasons why sales of Coca-Cola largely outperform sales of Pepsi Cola. According to our *research*², we found out that geographical preferences, and location of key markets are the main reasons.

Figure 9 visualizes global interest for Pepsi Cola and Coca Cola. Red regions represent higher interest in Coca Cola, and blue regions indicate higher interest in Pepsi Cola. Obviously, in a global perspective, Coca Cola outperforms Pepsi Cola. However, some regions show a strong interest in Pepsi Cola, such as the United States, where Pepsi Cola also has its key markets. Given that our analysis of ACSE stores are located in “Lunitunia”, and that our analysis shows that sales of Coca cola far more outweigh sales of Pepsi, therefore, we believe “Lunitunia” might be a place where people have strong preferences for Coca Cola, that is, “Lunitunia” is located within those “red regions”.

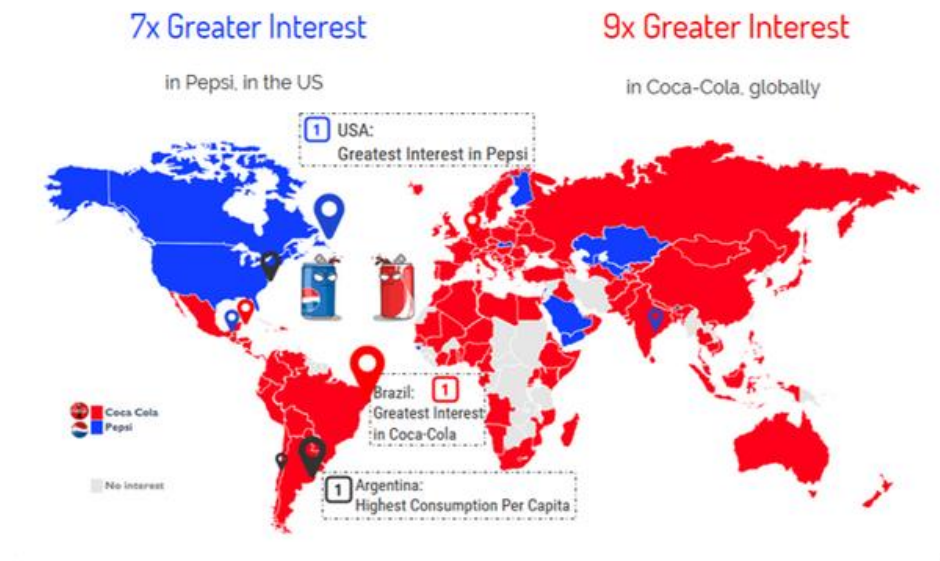


FIGURE 9: Pepsi vs Coca-Cola Global Interest Report

² Research Reference: <https://www.kalinux.com/pepsi-coca-cola.html>

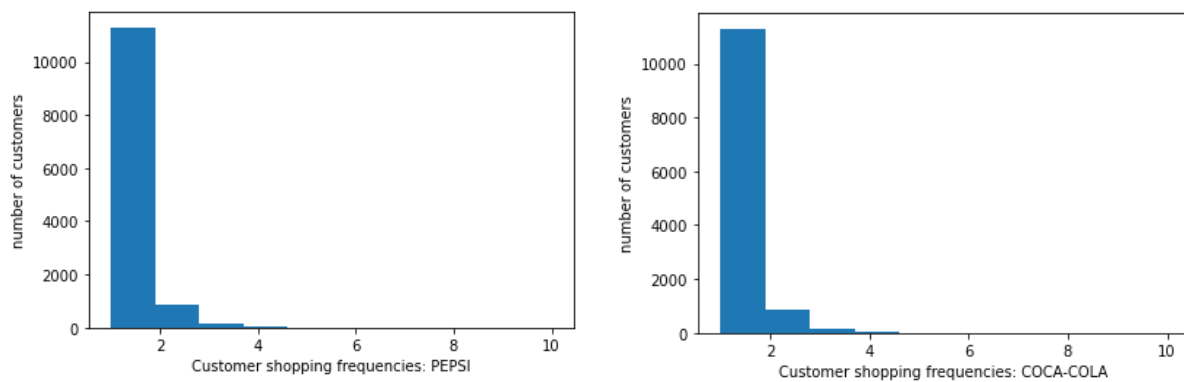
2.2.5 ACSE branded CSD

For the Carbonated Soft Drink industry, ACSE also has its own products. This is important because we do not want to cannibalize the sales of ACSE branded products. We found out that there are 85 different brands that ACSE carry in this category of products. In our sampled dataset, these 85 product SKUs had 19241 customers and accounted for 69,495.85 sales over the years. Relative information can be found in appendix.

2.3 Customer-Level Analysis

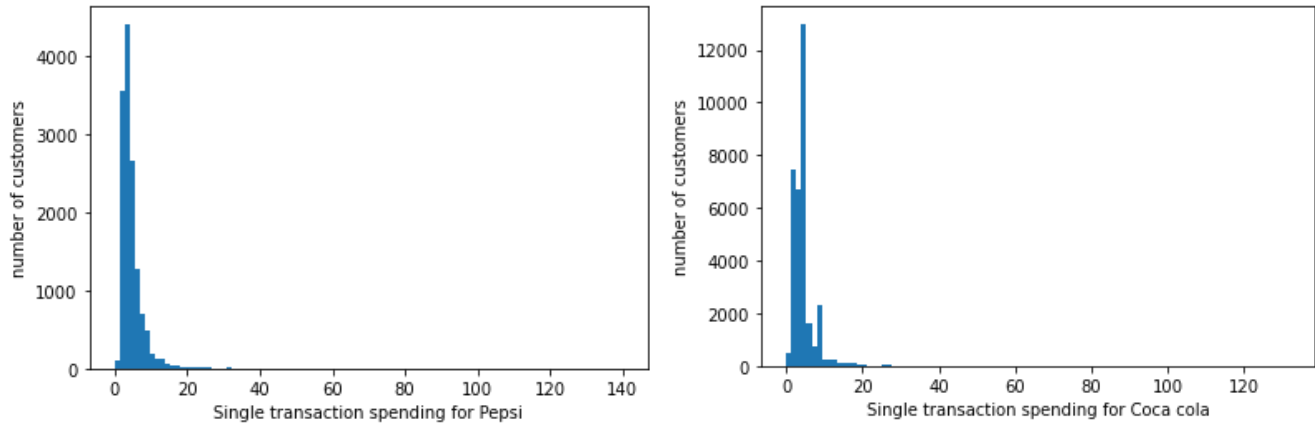
In this part, we further study the shopping patterns of customers for Pepsi and Coca Cola respectively. Based on project 1 customer grouping, we consider customer shopping frequency, single transaction spending, and customer growth by year as comparison criteria.

- Shopping Frequency



The majority of Pepsi and Coca Cola customers only purchase once in the 3 year dataset. Specifically, 11,298 customers shop once for Pepsi and 28,051 customers shop once for Coca Cola. The highest frequency for Pepsi customers is 10 times, compared with 9 times for Coca Cola. Also we can notice that the customer size has huge differences for the two brands. There was a total of 13,877 customers for Pepsi compared to 33,681 of Coca-Cola. Customers numbers are not mutually exclusive between the brands and years in this case.

- Single Transaction Spending



Next, we want to compare the customer spending per transaction for the two drinks. From the distribution plot, we can see majorities of both drinks spend less than 20 dollars for each transaction. The average price for Coca-Cola is lower than that of Pepsi.

	Pepsi(\$)	Coca Cola(\$)
Mean	4.75	4.52
Median	3.99	3.99
Max	139.70	131.76
Min	0.00	0.00

Table 1: single transaction spending for Pepsi and Coca Cola

- Customer growth by year

Last but not least, we compared customers' growth by year to as another measure of their market size. We aggregated our customer size by year to have an idea of how many unique customers have purchased Pepsi and Coca Cola from 2018 to 2020. And in our previous study, over 90% of customers only shop once for both Pepsi and Coca Cola, this number is a great proxy for the number of new customer size of each year. We can also find that Coca-Cola has a larger customer group.

Year	Pepsi	Coca Cola
------	-------	-----------

2017 (Starts from July)	1782	4535
2018	4086	9585
2019	3871	9098
2020	3488	9100

Table 2: customer size from 2017 to 2020

3. Modeling

For Modeling, we have first used an algorithmic approach based on Personal Recommendation which would use the user's prior history to model which item is ideal as a recommendation. On the second part, we have made an experimental algorithm that uses collaborative filtering to recommend products to customers.

3.1 Personalized Recommender

We first wanted to identify which of the two brands sells more while on discount. We wanted to find the discounted or promotional sales contribution of both the brands.

We first considered the price sensitivity of the different CSDs of Coca-Cola. Given only four CSD contribute to 80% of sales, we considered the four CSDs only ('COCA-COLA ZERO', 'COCA-COLA FRIDGEMATE', 'COCA-COLA DIET', 'COCA-COLA'). We found out among these items, 90% of sales is generated from discounted prices and hence promotions

We considered the four CSDs in Pepsi that contribute to 80% of Pepsi sales ('PEPSI', 'PEPSI DIET', 'PEPSI COLA DIET CAFFEINE FREE', 'PEPSI MINI CANS, PEPSI'), and found out that 77% of its sales are promotional i.e. on discounted price.

In terms of sales, Coca-Cola was almost thrice the size than that of Pepsi. This data is summarized in the table below:

	Coca-Cola	Pepsi
Total Sales	152,071	54,736
Sales with Promotional price	137,733	42,203
Promotional Sale %	91%	77%

Given the higher sales share for Coca-Cola and the larger market size, we decided to promote Coca-Cola other than Pepsi.

Among the 4 SKUs inside Coca-Cola, we found out that Coca-Cola FridgeMate yields the highest contribution of discounted sales, i.e., almost 25% of it.

Next, we compared the customer sizing of the two brands. To find the customers we wanted to find the ones who are more prone to buying items on promotion. We have looked to find customers who have at least 3 purchase transactions and more than 50% of these transactions include discounted products of Coca-Cola. In these criteria we found 245 customers.

We have found out that each of these customers come to ACSE at least 0.17 times a month and 98% of their transactions include discounted sales.

Promoting a discounted Coca-Cola FridgeMate to these customers would yield a total of USD 1011.87 from the campaign.

3.2 Experimental Collaborative Filtering

This portion of the deck is beyond the demands from ACSE and is not limited to only Coca-Cola and Pepsi. The team developed a recommender based on Collaborative Filtering which would publish the user – item recommendations.

We first made a Pareto analysis on the total assortment of products that ACSE sells. We found out that 8638 different product SKUs contribute to 80% of ACSE sales. We added the 80% contributing SKUs from Coca-Cola and Pepsi to this list of products. We will be using only these products for further evaluation.

We will be using the 245 ideal customers we had selected as the ideal customer and find collaborative filtering association within these 245 and 8711 products.

One issue in using collaborative filtering is that for customers without any prior transactions, it is hard to provide recommendations for them. To combat this we have designed an algorithm to find similar products for any new customers.

For the collaborative filtering to work in our case, we need to ensure that sparsity level in our data remains below 99.5%. With our subset of most sold products and ideal customers, the sparsity of the data was 99.2%. Using alternative least squares we measured the similarity of users and items.

ACSE can update the transactions data to this model and it will update the recommendations and similar items for users. Ideally, this customer and product recommendation needs to be stored somewhere and constantly updated.

An instance of the recommender working is shared below:

User Input:
6
User purchased:
20060696_NESTLE PURE LIFE
20306510_ACSE POTATO CHIPS REGULAR
20318694001_Coca-Cola Zero
20974528_ACSE EYE RND STEAK BF
20111346_ACSE FF AC CHKN BRST CUTLET TP
20310858_MINUTE RICE
20500252_HL MARKET CUTS - SALMON LOINS
Recomendations:
'20338720_ACSE MUFFINS CHOCOLATE CHIP'
'21204739_CASHMERE BT 12=24'
'20336823_ACSE MUFFINS BLUEBERRY'
'20419715005_Danone Oikos 2% - Rasp/Pom'
'20354288001_COCA- COLA ZERO'
'21097012001_AVOCADO BULK'
'21031672_ACSE CHICKEN BLT SALAD MINI'
'20168304001_ONION YELLOW'

4. Conclusion:

In the experimental section, we basically shared our personal ambition of making a RS which is beyond the requirements of ACSE. In terms of takeaways, there were a lot of stages in this exercise, where the team needed to converge to a direction unlike the first assignment. This open-ended approach meant, the team had to research externally to find the best practices and incorporate them. As a next step, we want to beta test our personalized recommender and the experimental CF recommender. We want to measure the performance between the recommenders which will help us fine tune the models and enable us to support ACSE further.

APPENDIX

1. Product Categories of Pepsi

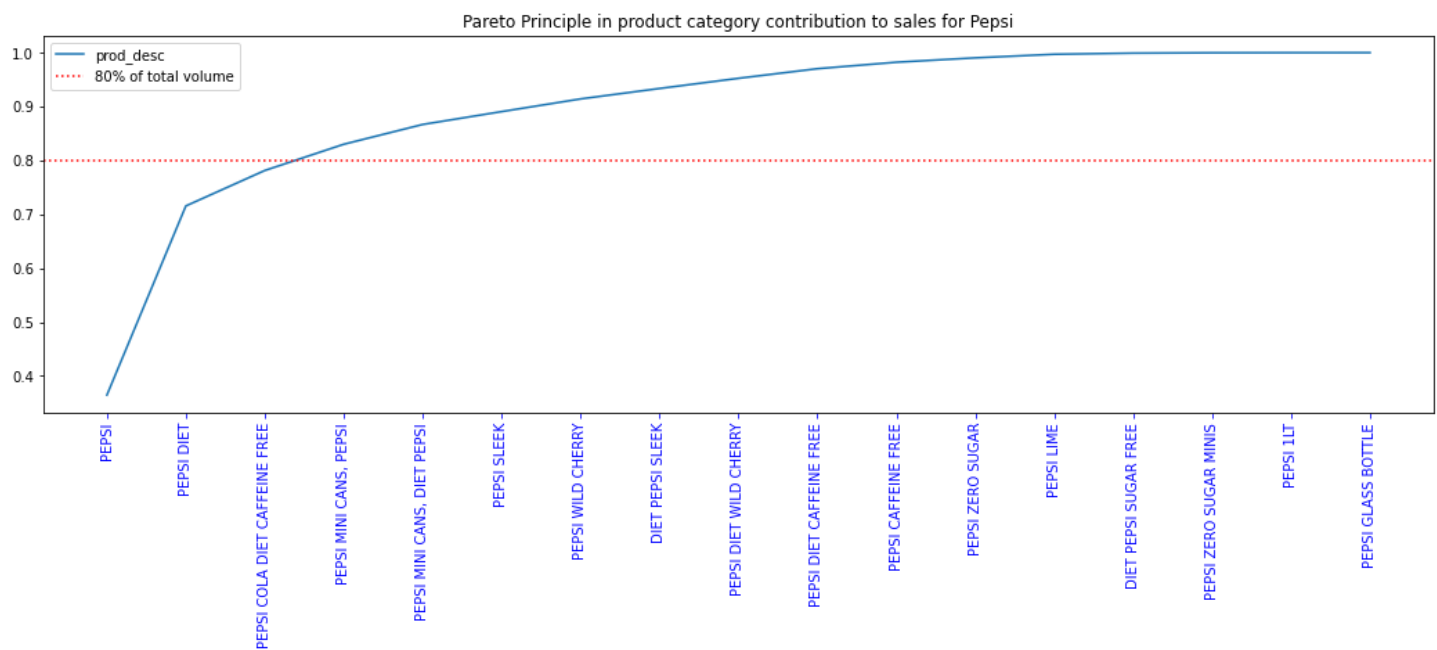
prod_id	prod_desc	prod_unit_qty_count	prod_count_uom	max	min	mean
20696250	PEPSI GLASS BOTTLE	4	ML	3.74	3.74	3.74
21020300	DIET PEPSI SLEEK	15	ML	5.99	4.99	5.97454
21020440	PEPSI SLEEK	15	ML	5.99	4.49	5.97326
21096820	PEPSI DIET WILD CHERRY	1	ML	2.49	2	2.22359
21164080	PEPSI ZERO SUGAR MINIS	6	ML	3.29	3.29	3.29
21189960	PEPSI ZERO SUGAR	6	ML	3.29	2.5	2.82318
20036150000	PEPSI	1	L	1.99	0.94	1.465
20040200000	PEPSI 1LT	1	L	1.49	1.49	1.49
20296930000	PEPSI	1	L	2.49	0.99	1.89644
20296930000	PEPSI DIET	1	L	2.49	0.88	1.88678
20296930000	PEPSI DIET CAFFEINE FREE	1	L	2.49	0.88	1.92702
20306690000	PEPSI DIET	24	ML	12	4.44	7.37379
20306690000	PEPSI	24	ML	11	3.99	7.07011
20310550000	PEPSI	6	ML	4.99	3.33	4.72332
20310550000	PEPSI DIET	6	ML	4.99	1.49	4.75386
20321600000	PEPSI	1	ML	2.49	2	2.24265
20321600000	PEPSI DIET	1	ML	2.49	1	2.24685
20321600000	DIET PEPSI SUGAR FREE	1	ML	2.49	2	2.25125
20321600000	PEPSI WILD CHERRY	1	ML	2.49	2	2.25432
20322360000	PEPSI	12	ML	6.35	0.69	4.22146
20322360000	PEPSI DIET	12	ML	5.99	0.69	4.27647
20322360000	PEPSI CAFFEINE FREE	12	ML	5.99	3.17	4.07112
20322360000	PEPSI COLA DIET CAFFEINE FREE	12	ML	5.99	3.17	4.2179
20322360000	PEPSI LIME	12	ML	5.99	3.25	4.20558
20322360000	PEPSI WILD CHERRY	12	ML	5.99	3.17	4.1216
20322360000	PEPSI DIET WILD CHERRY	12	ML	5.99	3.79	4.4675
20857560000	PEPSI MINI CANS, PEPSI	6	ML	3.29	0.94	2.97374
20857560000	PEPSI MINI CANS, DIET PEPSI	6	ML	3.29	1.39	2.99307

2. Product Categories of Coca-Cola

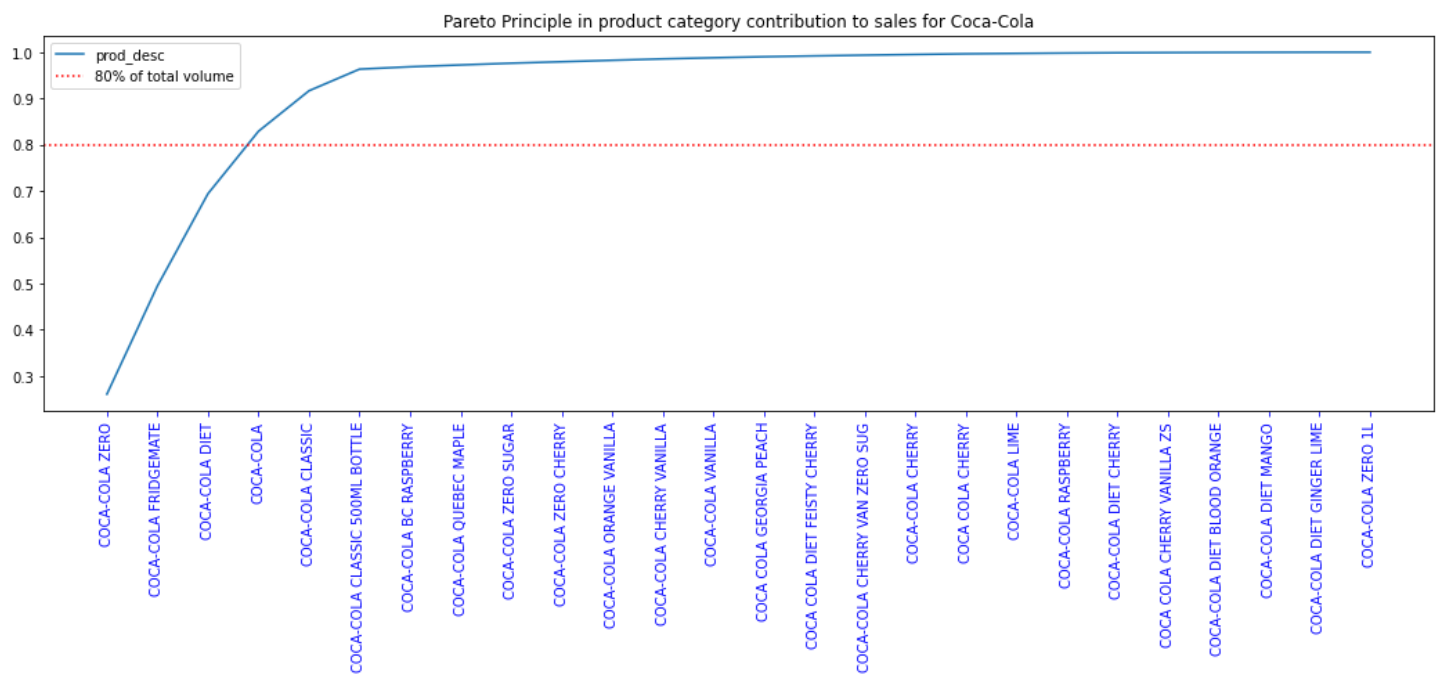
prod_id	prod_desc	prod_unit_qty_count	prod_count_uom	max	min	mean
20299050	COCA-COLA	1	L	1.99	2	1.99
20659780	COCA-COLA ZERO 1L	1	L	1.99	1.3	1.64
20783410	COCA-COLA DIET	1	L	2.19	2	2.10429
20827290	COCA-COLA CLASSIC 500ML BOTTLE	1	ML	2.49	1	2.24071
20895310	COCA COLA CHERRY	1	ML	2.49	2	2.28133
20941140	COCA-COLA CHERRY	24	ML	3.99	4	3.99
20968250	COCA-COLA VANILLA	1	ML	2.49	2	2.19459
20970980	COCA-COLA CLASSIC	1	L	1.79	1.5	1.63902
21021560	COCA-COLA ZERO CHERRY	12	ML	5.79	3.8	4.11453
21022000	COCA-COLA ZERO CHERRY	1	ML	2.49	2	2.24585
21032440	COCA-COLA CHERRY	6	ML	3.29	2.5	3.07175
21032900	COCA-COLA CLASSIC	4	ML	2.99	2.2	2.98793
21041150	COCA-COLA ZERO SUGAR	4	ML	2.99	2.5	2.98017
21059770	COCA-COLA ZERO	10	ML	4.49	3	4.21877
21059790	COCA-COLA DIET	10	ML	4.49	3.5	4.2901
21059890	COCA-COLA ZERO CHERRY	10	ML	4.49	4.5	4.49
21060040	COCA-COLA CLASSIC	10	ML	4.49	3	4.24974
21101390	COCA-COLA RASPBERRY	1	ML	2.29	2	2.22556
21101400	COCA-COLA DIET GINGER LIME	1	ML	1.29	1	1.03615
21101400	COCA-COLA DIET BLOOD ORANGE	1	ML	1.29	0.9	1.11955
21101400	COCA-COLA DIET MANGO	1	ML	1.29	1	1.05774
21101400	COCA-COLA DIET CHERRY	1	ML	1.29	1	1.19735
21109540	COCA-COLA ORANGE VANILLA	1	ML	2.49	2	2.19272
21184050	COCA COLA DIET FEISTY CHERRY	1	ML	2.29	0.5	2.14206
21185430	COCA COLA GEORGIA PEACH	4	ML	5.49	3.9	5.37125
21185430	COCA-COLA BC RASPBERRY	4	ML	5.49	5	5.42077
21189250	COCA-COLA LIME	1	ML	2.29	1.8	2.16585
21190120	COCA-COLA CHERRY VANILLA	1	ML	2.29	2	2.18218
21240620	COCA-COLA ORANGE VANILLA	12	ML	4.99	4.5	4.63192
21240670	COCA-COLA QUEBEC MAPLE	4	ML	5.49	5	5.35869
21284160	COCA-COLA CHERRY VANILLA	12	ML	6.99	4.5	4.72457
21284700	COCA-COLA CHERRY VAN ZERO SUG	12	ML	6.99	4.5	4.75231
21295930	COCA COLA CHERRY VANILLA ZS	1	ML	2.29	2	2.18316
20308200000	COCA-COLA	24	ML	11	0.7	7.21051
20308200000	COCA-COLA ZERO	24	ML	11	6.4	7.77552
20313740000	COCA-COLA ZERO	6	ML	4.99	2	4.53227
20313740000	COCA-COLA	6	ML	4.99	1.5	4.48991
20316030000	COCA-COLA	1	L	2.49	1	1.89667

20316030000	COCA-COLA ZERO	1	L	2.49	1	1.90834
20318690000	COCA-COLA ZERO	12	ML	6.35	0.8	4.30016
20318690000	COCA-COLA FRIDGEMATE	12	ML	6.47	1.7	4.22298
20318690000	COCA-COLA DIET	12	ML	6.35	0.7	4.27526
20354290000	COCA-COLA ZERO	6	ML	3.29	2.2	2.77173
20354290000	COCA-COLA CLASSIC	6	ML	3.29	1.4	2.78784
20354290000	COCA-COLA DIET	6	ML	3.29	1	2.79093

3. Pareto Analysis:
a. Pepsi:



b. Coca-Cola:



4. Table for Product Category wise sales of Pepsi:

prod_desc	Sales
PEPSI	24039.78
PEPSI DIET	23145.39
PEPSI COLA DIET CAFFEINE FREE	4341.21
PEPSI MINI CANS, PEPSI	3209.47
PEPSI MINI CANS, DIET PEPSI	2424.61
PEPSI SLEEK	1573.86
PEPSI WILD CHERRY	1552.84
DIET PEPSI SLEEK	1277.86
PEPSI DIET WILD CHERRY	1245.85
PEPSI DIET CAFFEINE FREE	1176.64
PEPSI CAFFEINE FREE	792.1
PEPSI ZERO SUGAR	535.83
PEPSI LIME	439.37
DIET PEPSI SUGAR FREE	140.07
PEPSI ZERO SUGAR MINIS	42.77
PEPSI 1LT	11.92
PEPSI GLASS BOTTLE	3.74

5. Table for Product Category wise sales of Coca-cola:

prod_desc	Sales
COCA-COLA ZERO	39681.68
COCA-COLA FRIDGEMATE	35579.74
COCA-COLA DIET	30296.54
COCA-COLA	20527.26
COCA-COLA CLASSIC	13313.69
COCA-COLA CLASSIC 500ML BOTTLE	7109.63
COCA-COLA BC RASPBERRY	779.06
COCA-COLA QUEBEC MAPLE	599.88
COCA-COLA ZERO SUGAR	574.87
COCA-COLA ZERO CHERRY	481.5
COCA-COLA ORANGE VANILLA	474.52
COCA-COLA CHERRY VANILLA	461.54
COCA-COLA VANILLA	370.27
COCA COLA GEORGIA PEACH	337.67
COCA COLA DIET FEISTY CHERRY	278.16
COCA-COLA CHERRY VAN ZERO SUG	237.33
COCA-COLA CHERRY	233.86
COCA COLA CHERRY	213.1
COCA-COLA LIME	126.59
COCA-COLA RASPBERRY	125.89
COCA-COLA DIET CHERRY	90.15
COCA COLA CHERRY VANILLA ZS	57.48
COCA-COLA DIET BLOOD ORANGE	53.42
COCA-COLA DIET MANGO	34.08
COCA-COLA DIET GINGER LIME	29.91
COCA-COLA ZERO 1L	3.28

6. ACSE CSD comparison with Pepsi and Coca-Cola

	ACSE CSD	Coca-Cola	Pepsi
SKU count	85	45	28
Number of customers	19,241	11,298	28,051
Total Sales	69,495.85	152,071.10	65,953.31