Capstone Project

Starbucks Promotion Recommendation

Sarmilan Sivanesasuntharam

Machine Learning Engineer Nanodegree

Introduction

Project Overview

Retail businesses give offers and promotions to their customers in order to retain their customer base, influencing them to spend more or even to try different products which are not in their typical basket.

According to this case study, Starbucks mainly promotes their products via rewards to its customers through their mobile app. It leads their customers to purchase more or stay in touch with the brand for a longer time. Starbucks sends out offers to users of the mobile app once every few days.

An offer can be any of the following;

- an advertisement for a drink or

- an actual offer such as a discount or

- BOGO (buy one get one free)

Some users might not receive any offers during certain weeks and offers will not be the same for everyone, and offers have validity periods as well.

Problem Statement

Customers receive different types of offers which are not optimised or personalised but these offers influence customers' purchasing pattern and encourage retention with the brand as well. The demographic groups that customers belong to, their earnings, how long they have been customers with the brand and many other factors may affect their choices and purchasing patterns.

This study tries to build a recommender system that suggests best offers to the customers based on their past interactions or responses to the offers that they received. In addition to that this study tries to identify which offers are popular or which groups of people respond positively to the offers for Starbucks to optimise and personalise the offers that they send out.

Success Criteria

Success of the model would be evaluated based on how closely the model predicts customers' preferences. This could be measured using Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) of deviation between actual offer completion rate and predicted score for the offer type at offer level.

Success of the clustering system would be to identify groups of people who respond to the offers positively with a higher rate of response for an offer type than the average response by Starbuck customer base as a whole.

Analysis

Datasets

The data is contained in three files:

- portfolio.json - contains offer id and meta data about each offer (duration, type, etc.)
- profile.json - demographic data for each customer
- transcript.json - records for transactions, offers received, offers viewed, and offers completed

Above datasets captured 17000 unique users' behaviours on offers for a period of a month. Here is the schema and explanation of each variable in the files:

portfolio.json

- id (string) - offer id
- offer_type (string) - type of offer, i.e. BOGO, discount, informational
- difficulty (int) - minimum required spend to complete an offer
- reward (int) - reward given for completing an offer
- duration (int) - time for the offer to be open, in days
- channels (list of strings)

profile.json

- age (int) - age of the customer
- became_member_on (int) - date when customer created an app account
- gender (str) - gender of the customer (note some entries contain 'O' for other rather than M or F)
- id (str) - customer id
- income (float) - customer's income

transcript.json

- event (str) - record description (ie transaction, offer received, offer viewed, etc.)
- person (str) - customer id
- time (int) - time in hours since the start of the test. The data begins at time t=0
- value - (dict of strings) - either an offer id or transaction amount depending on the record

Data Preprocessing

Data preprocessing is done to clean the data and extract meaningful information from the dataset provided, and also to merge the datasets and create a master table.

Functions are created to preprocess the each data; *process_portfolio()*, *process_profile()* and *process_transcript()* are created mainly to do the following processes:

1. Encode some categorical value columns
2. Rename some column names
3. Date format conversion

4. Impute or replace unrealistic values
5. Handle duplicate entries
6. Extract and structure values from dictionary data types

Profile dataset had some missing values and some unrealistic values. About 2175 customer records did not have gender and income details. Also the age column has value 118 for the same customers and this was removed from records as it's unrealistic.

Imputation has not been tried as major customer related information is missing for the mentioned 2175 customers.

Transcript dataset had some duplicate values which were removed from records.

Exploratory Data Analysis

Exploratory data analysis is done to understand the customer behavior and offer types. This includes some summary tables and charts. Important graphs and tables are shared in this part.

General Information

- Starbucks simulated dataset contains 17000 customers records and their transactions for a period of 30 days.
- 8484 of total records contain about 50% male records and 36% female records. Rest of them are missing and others.
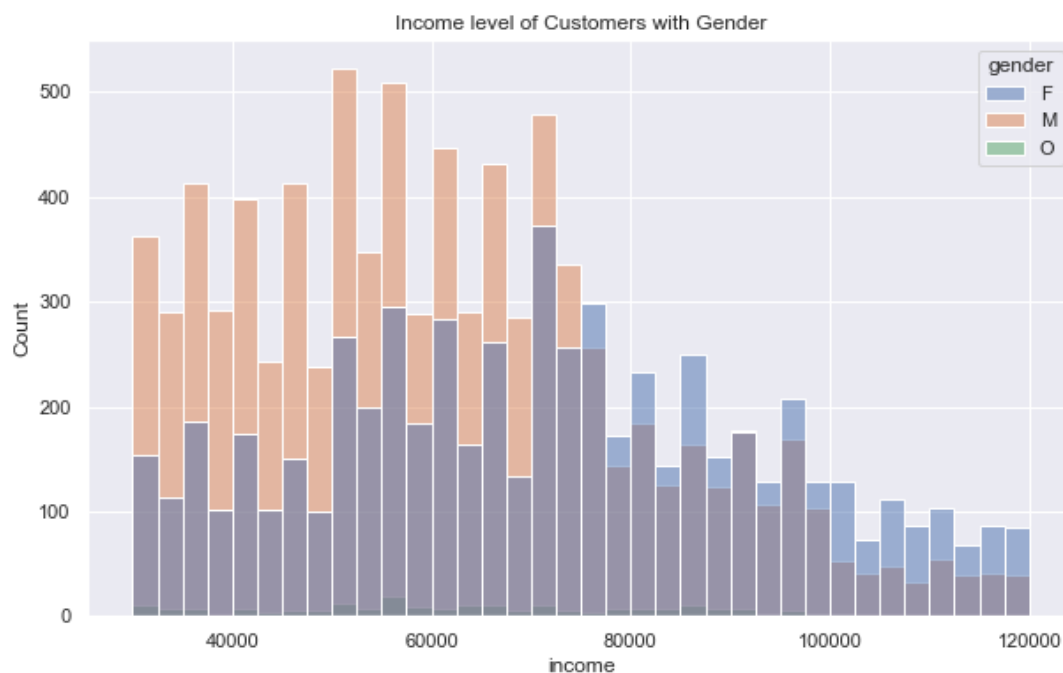
Above histogram indicates that a higher number of male customers have low income compared to female customers, and a higher proportion of higher income customers are female customers.

Offers Information

Ten types of offers are given to the customers and all of them can be categorized as BOGO, Discount and Informational. Offers are analysed at Offer level and ID level to get more understanding about them.

Below tables illustrate how many offers were sent out by the Starbucks and how many of them were utilised by the customers.

| Offer Type	Received	Viewed	Completed	%Viewed	%Completed
BOGO	30,499	25,499	15,501	83%	51%
Discount	30,543	21,445	17,681	70%	58%
Informational	15,235	10,831	-	71%	-

Above table shows that BOGO and Discount types of offers are sent in equal amounts and as twice as Informational offers but BOGO is viewed more than the rest of the offers. Informational and Discount have about 70% views. Discount offers have a higher rate of completion, which is 58%.

Offer id	Received	Viewed	Completed	%Viewed	%Completed
b1	7658	6716	3635	0.88	0.47
b2	7593	7298	3301	0.96	0.43
b3	7677	4171	4303	0.54	0.56
b4	7571	7264	4262	0.96	0.56
d1	7668	2663	3340	0.35	0.44
d2	7646	7337	5112	0.96	0.67
d3	7597	7327	5265	0.96	0.69
d4	7632	4118	3964	0.54	0.52
i1	7617	4144	-	0.54	-
i2	7618	6687	-	0.88	-

Above table shows that each offer was sent approximately in equal amounts but some of them are completed with higher percentages such as **d2** and **d3**. Some offers were viewed in higher percentages. Offers **b2**, **b4**, **d2** and **d3** were viewed more than 95%. Offer **i2** is popular in informational offers.

Transaction Information

This section analyses how transactions are distributed and whether there are any significant differences between democratic variables.


Total Amount of Transactions
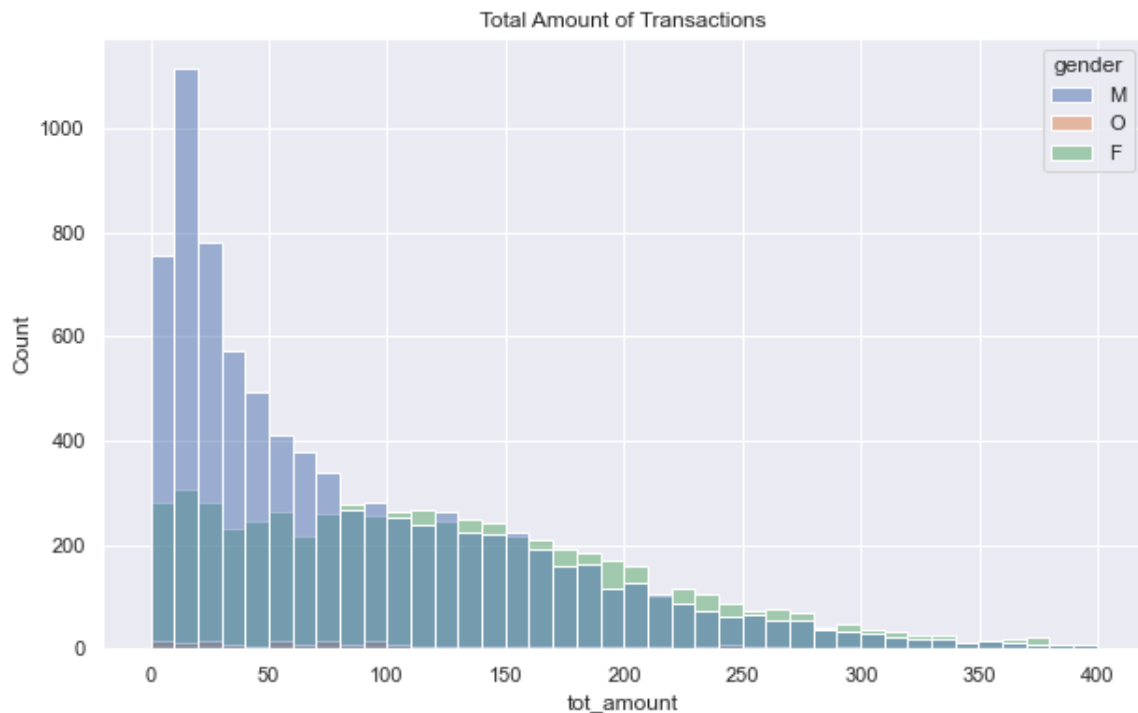
Above histogram shows that the majority of lower amounts of total transactions were done by males and their transactions are skewed towards lower total amounts. Higher amounts of transactions were done by female customers slightly higher than male customers during this one month period.

Very few extreme higher amounts of total transactions were also observed but are not included in this chart.

Average Transaction Amount Vs Customer Income Level
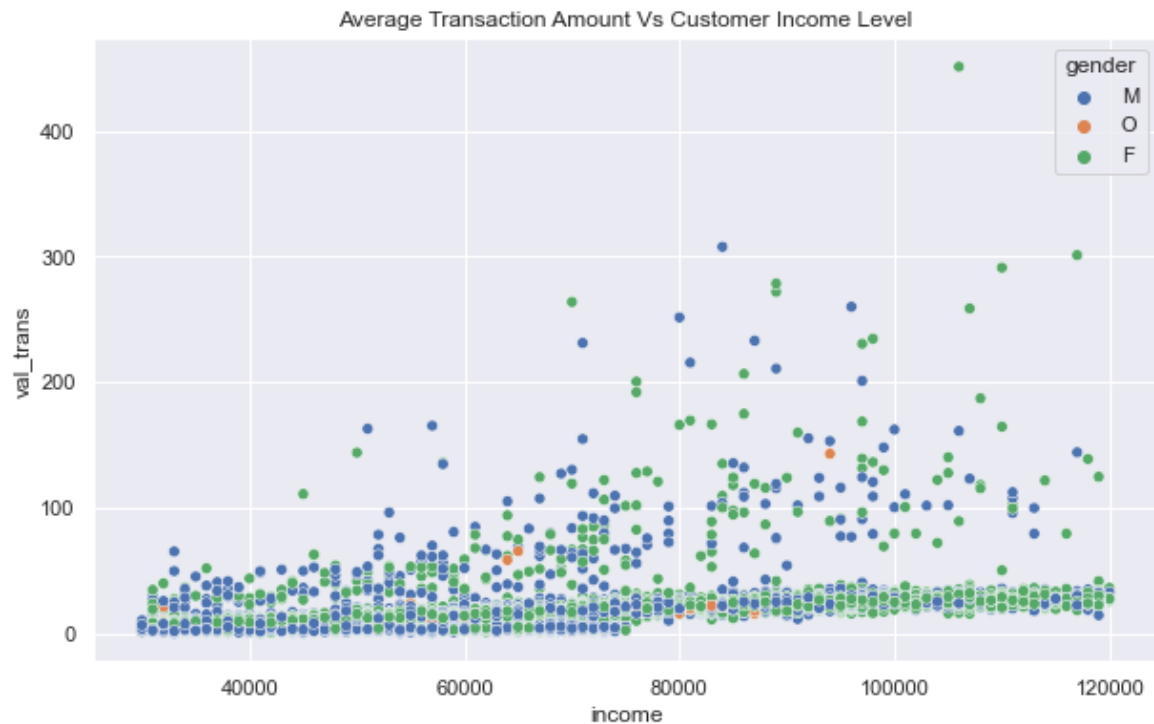
Above chart shows that income slightly influences average value transactions and it could be observed for those who have income higher than $75000. Some customers' average spendings are higher than the rest of the customers and it steadily increases with their income level.


Transaction by Offer ID
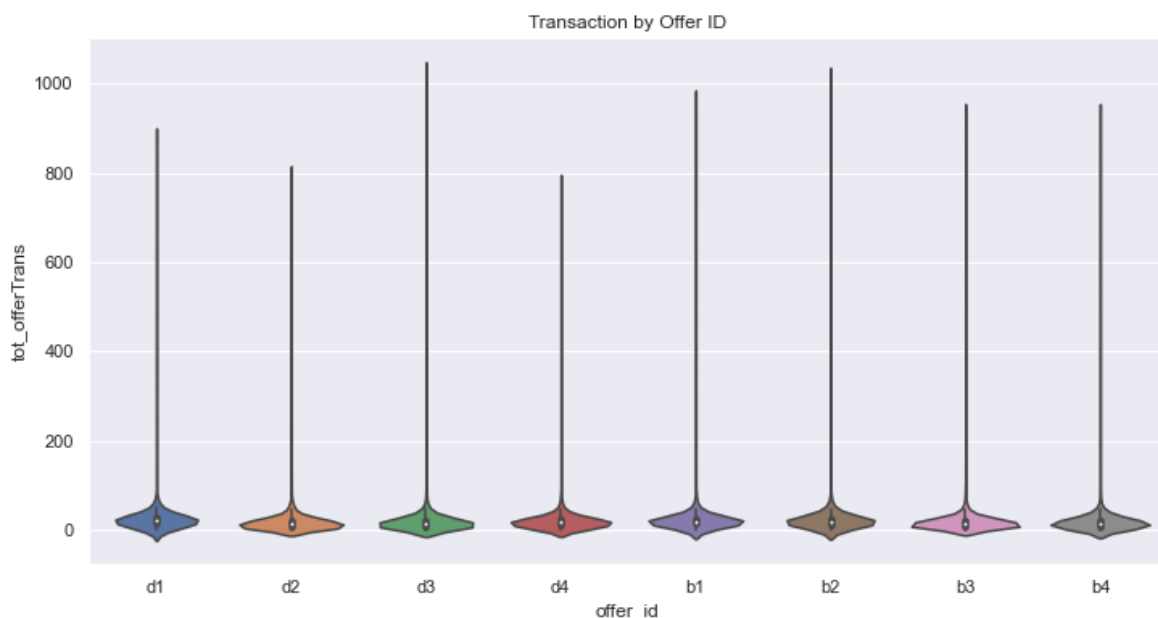
Above Violin plot clearly shows that no individual offers influenced the total transaction amount significantly.

Rewards Information

This section analyses how much rewards were given by the offers.

Total Rewards
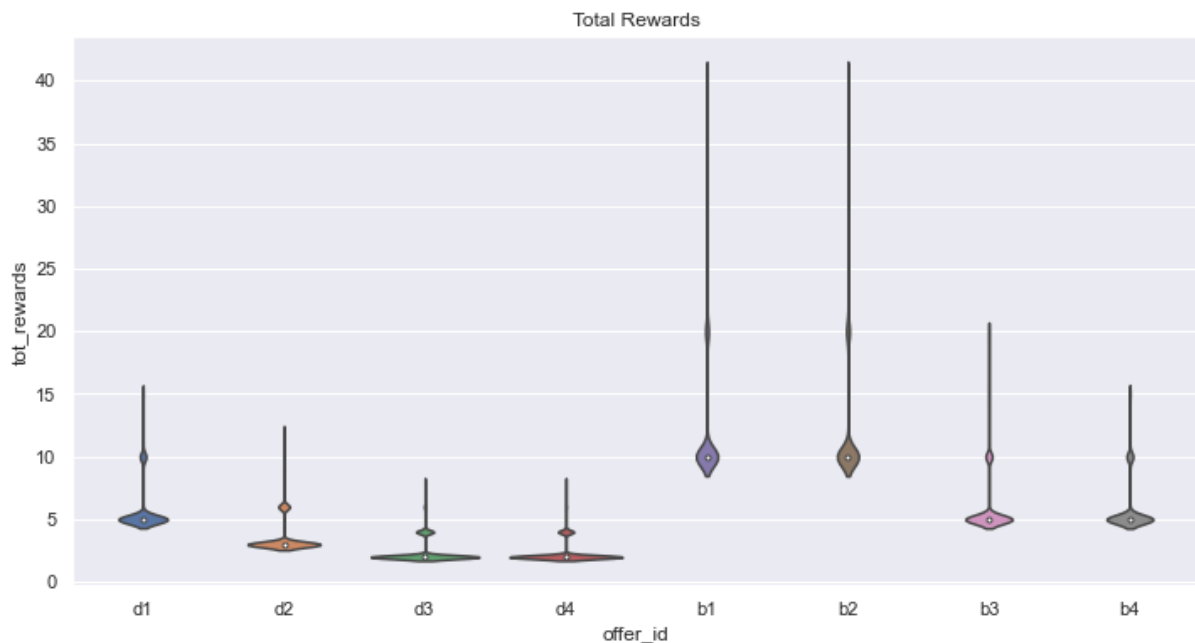
Above plot shows that higher amounts of rewards are given by BOGO offers to the Starbucks customers, while b1 and b2 give comparatively higher amounts of rewards. d1 and d2 are the best offers in the Discount category of offers in terms of rewards it gives.


Total Transaction Vs Total Rewards
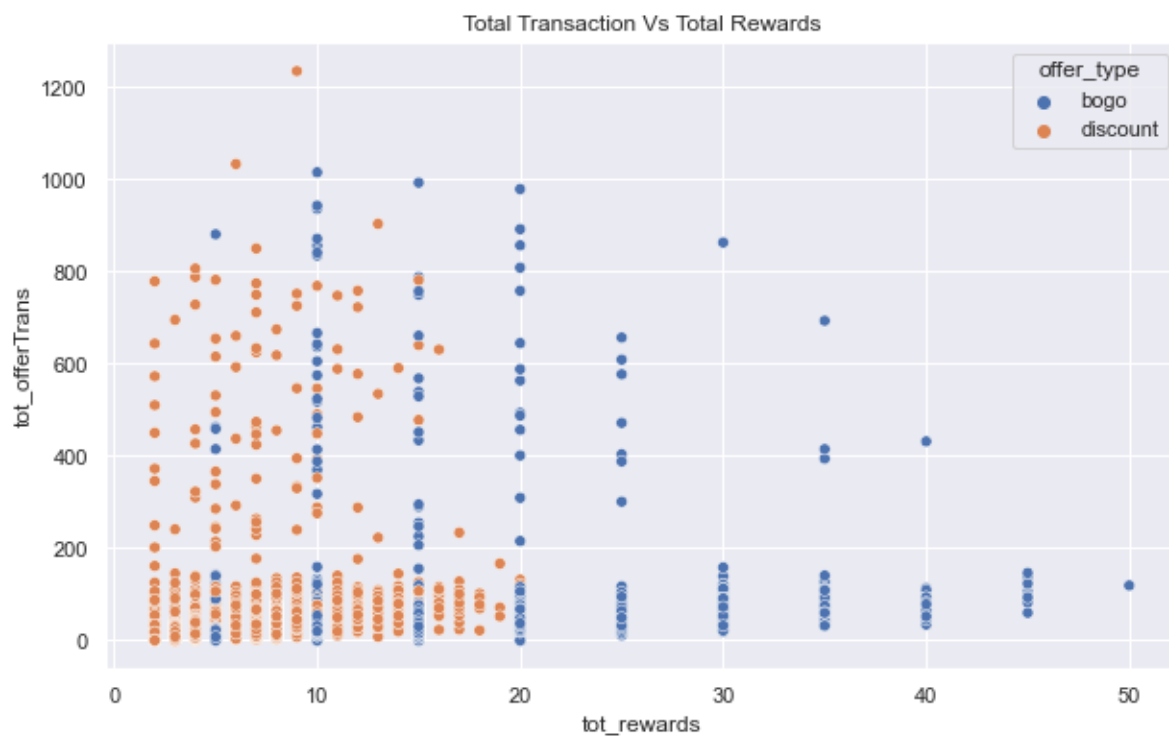
Above scatter plot provides evidence to say rewards don't drive the transaction amount effectively. Some customers have spent a higher amount of money regardless of the reward amount that they received. BOGO offers allow customers to enjoy higher rewards.

Advanced Analysis

Advanced Statistical and Machine Learning methodologies are used to address the problems identified.

Cluster Analysis

Hierarchical Clustering method was employed to group the customers who have similar behavior or engagement patterns. Customer democratic information, transaction details, offer utilisation at Offer type level and many other factors were feeded to the clustering algorithm for this purpose.
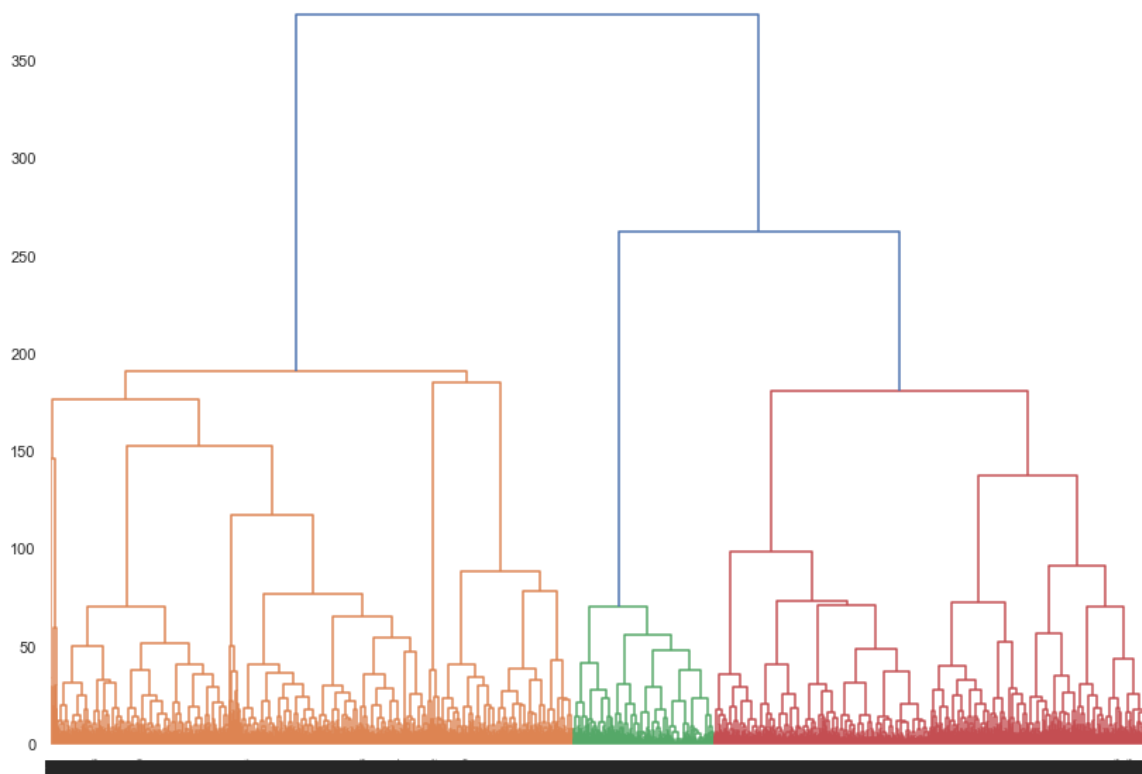
Above dendrogram shows the groups of Starbucks customers considering their purchasing behavior and their democratic information. Three distinct groups are observable in the dendrogram.

Agglomerative Hierarchical Clustering is used to cluster customers into 03 groups and a summary table is derived to see if there are any clear patterns in customers' features or behaviours which lead them to spend differently or respond to offers better than rest of the customers.

cluster	0	1	2
# Customers	8094	6731	2175
Transaction Amount	167.19	56.71	18.63
# Transaction	8.64	8.03	6.89
Rewards Amount	16.55	3.66	1.86
# Rewards	3.03	1.12	0.51
BOGO completion	0.85	0.12	0.11
Discount completion	0.79	0.47	0.18
Age	56.79	51.51	-
Income	71,446.87	58,139.65	-
Gender ratio F:M	1.0	0.5	-

Above table shows 3 natural clusters with distinct behaviors and feature sets. Cluster 2 represents the customer with missing information but shows very significant deviation from the rest of the clusters such as small total transaction amount, less offer engagements and lower rewards.

Cluster 0 looks like it contains premium customers. Their total transaction is thrice as high as cluster 2. They are very active to offers and earn higher rewards. Their income level and their age both are slightly higher than cluster 2.

In contrast, cluster 2 does the same number of transactions as cluster 1 but their amount spent on transactions is not as high as cluster 1. Cluster 1 has twice the number of male population.

Recommender System

A recommender system will help Starbucks to identify best offer types for its customers based on similar customers' choices in the past. If a customer utilized an offer type that means they like that offer more than the rest of the offers. Offer utilization is defined as the percentage of completion of an offer type out of the number of offers they received.

Information is processed and features are created at Offer ID level to recommend offers at Offer ID level to the customers.

Collaborative Filtering models are tested as it takes customers' interactions towards offers and suggests best offers to them. Matrix factorization models such as SVD, SVD++, and Slope One and Co Clustering were tested on this offer recommendation problem.

Co Clustering method gave a good response on 05 fold cross validation and which was chosen to make recommendations after a careful refinement process.

Benchmark

Average completion rates of BOGO and Discount offers are 50% and 60% respectively. But the clustering algorithm identified customer groups which respond to the above offer types with higher rate of completion i.e. 85% and 80% respectively. In addition to that, their transaction amount is also high which will be beneficial for Starbucks.

Model Evaluation and Validation

Five fold cross validation method is used to train and test the model accuracy against RMSE and MAE matrices.

RMSE for all models are between 0.39 - 0.41 but Co Clustering with default parameters got approximately RMSE 0.43.

Co Clustering was fine tuned to get the best parameters. Number of epochs, Item cluster number and User cluster number were adjusted to find the best value combination and which is used to fit the best model.

Recommendation function is built to make offer suggestions to the customers. The function looks like the following;

get_recommendations(customer_id = '0011e0d4e6b944f998e987f904e8c1e5', model=algo, n_rec=3)

There is no evaluation mechanism available for Agglomerative Hierarchical Clustering but it identified customers with higher potentials.

Findings and Discussion

- Agglomerative Hierarchical Clustering identified customer groups which respond to the offers very positively and have higher purchasing capacity.

- RMSE matrix would not be suitable to identify the best model due to many reasons. Deviation will be high if a customer likes an offer which he never received. Recommendation score would be high in this scenario but the actual is zero as he didn't try that offer.

- Recommendations can be used as an experiment in the future to measure customers' reaction.

- Informational Offers are not included in modeling as they don't have offer completion therefore measuring customer behavior towards Informational Offers is difficult.

- It's believed that Collaborative filtering captures customer's behavior and makes recommendations based on their or similar users' behavior but a hybrid approach can be tried for this problem.