Project Proposal - Starbucks Capstone Project

Arya Suneesh

April 2023

Domain Background

The project is in the field of marketing analytics. It involves analyzing customer transaction data to identify the most popular offers and their effectiveness. Marketing analytics is a critical area of research for companies to improve their customer acquisition, retention, and revenue generation.

The given data set contains simulated data that mimics customer behavior on the Starbucks rewards mobile app. Once every few days, Starbucks sends out an offer to users of the mobile app. An offer can be merely 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 offer during certain weeks.

Not all users receive the same offer, and that is the challenge to solve with this data set.

Problem Statement

The problem being investigated is to identify the most popular offers among customers and understand their effectiveness in increasing customer engagement and driving sales. The analysis will focus on using customer data to identify customer segments that respond well to specific offers, and developing a strategy to tailor offers to those segments. This will involve analyzing customer demographics, purchase history, and offer interaction data to gain insights into what types of offers are most effective for different segments of customers. The ultimate goal of the project is to improve customer engagement and increase sales by delivering more personalized and effective offers to customers.

Datasets and Inputs

The input data is customer transaction data, which includes information about the offers they received, viewed, and completed. The dataset also includes demographic information such as age, gender, and income.

The data used for this project is contained in three files: `portfolio.json`, `profile.json`, and `transcript.json`.

`portfolio.json` contains offer ids and metadata about each offer. It includes the following variables:

- `id` (string) - offer id
- `offer_type` (string) - type of offer, such as BOGO, discount, or informational
- `difficulty` (int) - minimum required spend to complete an offer
- `reward` (int) - reward given for completing an offer
- `duration` (int) - time for offer to be open, in days
- `channels` (list of strings) - channels used to distribute the offer

`profile.json` contains demographic data for each customer. It includes the following variables:

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

`transcript.json` contains records for transactions, offers received, offers viewed, and offers completed. It includes the following variables:

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

Solution Statement

The proposed solution is to use data analysis and visualization techniques to identify the most popular offers among customers and their effectiveness. Additionally, clustering algorithms will be used to identify customer segments that respond well to specific offers. The insights generated from this analysis will enable the company to create targeted marketing campaigns that are more effective in acquiring and retaining customers.

Benchmark Model

The benchmark model will be a simple rule-based approach that recommends the most popular offers to all customers. This approach assumes that all customers have similar preferences and behavior, which may not be accurate.

Evaluation Metrics

The evaluation metrics for the proposed solution will be the revenue generated from the targeted marketing campaigns and the customer retention rate. Additionally, the accuracy of the clustering algorithms will be evaluated using metrics such as silhouette score and within-cluster sum of squares.

Project Design

The project will involve the following steps:

1. Data cleaning and preprocessing to prepare the dataset for analysis.
2. Exploratory data analysis and visualization to identify the most popular offers and their effectiveness.
3. Clustering analysis to identify customer segments that respond well to specific offers.
4. Developing targeted marketing campaigns based on the insights generated from the analysis.
5. Evaluation of the effectiveness of the targeted marketing campaigns using the defined metrics.
6. Comparison of the proposed solution with the benchmark model to measure its effectiveness.

Conclusion

In this project, we aim to develop a customer segmentation and offer targeting strategy using transactional data. The proposed solution will use unsupervised and supervised learning techniques to identify customer segments and design targeted offers for each segment. The effectiveness of the model will be evaluated using conversion rate, incremental revenue, and customer satisfaction.