

# Starbucks Project Capstone Proposal

---

## Background:

Starbucks is a global organization that has over 32,000 coffeehouses, stores and roastery reserves. To improve the user experience to buy their products at their stores, Starbucks has mobile apps through which customers can enroll in to rewards program and place orders on various products to pick them up in the store. Starbucks has rolled out various campaigns to either reward the loyal customers or boost the sales or promote the new products. All customers may not respond in same manner and optimizing the process of offer rollout to maximize the response is important. Their current goal is to improve the process of offering these offers to their customers. So, for this project, Starbucks has provided a simulated data set from their mobile app to develop an algorithm that can recommend optimal offers to their customers that maximizes the response to the sent offer.

## Problem Statement:

Starbucks have various types of promotional offers for its customers, but do not know the type of offer to offer to its customer. They want to avoid sending incorrect offer which might lead to negative user experience. They want to develop an algorithm that sends the promotional offer to the appropriate user and maximize the response rate.

## Datasets and inputs:

There are three types of offers that can be sent: buy-one-get-one (BOGO), discount, and informational. In a BOGO offer, a user needs to spend a certain amount to get a reward equal to that threshold amount. In a discount, a user gains a reward equal to a fraction of the amount spent. In an informational offer, there is no reward, but neither is there a requisite amount that the user is expected to spend. Offers can be delivered via multiple channels.

There are 3 data sets available to develop the algorithm-

**profile.json:-** Data set containing the list of customers who are members of rewards program. This data set contains user specific details such as gender, age, income and age of membership

**portfolio.json:-** Data set contains details about the offers. It contains information like the reward of the offer, channels through which the offers are presented, conditions to use the offer, duration of the offer, and type of the offer

**transcript.json:-** This data set contains the activity details of the customers of Starbucks who are enrolled in to rewards program. The data set contains the details such as amount of the transaction, type of offer used, duration of the offer use before expiry, and reward obtained for the offer

## Solution statement:

Develop a recommender engine to obtain the recommended type of offer that maximizes the response rate. Based on the available information about users and their behavior for various

offers, will develop multiple types of recommender systems and evaluate the best one based on the accuracy of the model. User-offer recommender system using Funk SVD can be used to develop the model to predict the optimal recommendations to the user.

### **Benchmark model:**

As a benchmark model, we can use a simple model of choosing the most popular offers in cohort (age group) as recommendation and compare it with a more complex models developed as specified in the solution above. Since the recommendation would be same for everyone belonging to a particular age group, it will be a good benchmark model.

### **Evaluation metrics:**

I will use precision and recall as the success metrics to evaluate the developed model. To be more precise, I will calculate precision@k and recall@k where k being the number of recommendations. Below are the formulas to calculate the success metrics-

$$recall@k = \frac{\# \text{ of true offers predicted}}{\# \text{ of true offers}}$$

$$precision@k = \frac{\# \text{ of true offers predicted}}{\# \text{ of offers predicted}}$$

Precision and recall must be balanced, and we can use a metric called F1 Score, it is a harmonic mean of recall and precision. It can be calculated by the following formula-

$$F1@k = \frac{2}{\frac{1}{precision@k} + \frac{1}{recall@k}}$$

### **Project Design:**

- Data ingestion- Ingest the data
- Data cleaning- Clean the data to use it to develop the model. Merge different data sets and identify the missing data, and engineer features necessary for the model
- Data exploration- Explore the patterns in the data set, and derive insights about features
- Train the models- Identify the appropriate model after testing various algorithms and hyperparameters
- Test the models- Test the developed model on the test data and measure the generalization of the model
- Summarize the results- Write the project report summarizing the details and the results of the project