

**Starbucks Capstone Project Proposal**  
**Deepika Vijayakumar | June 30, 2021**  
**Udacity – Machine Learning Engineer Nanodegree**

**Domain Background:**

Marketeers can achieve the best results by planning their campaigns with more insights about their potential customers. These insights can be obtained from the data present in their site, current campaigns and social media. Data science projects help marketers target the right customers and thereby enable profit maximization.

This capstone project is one such project where on analyzing the dataset, we get insights that help the marketing team to perform well. Thereby achieving business objectives. In this project, we take Starbucks which gives various promotions to its customers. 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.

These promotional offers use multiple channels and they are e-mail, social media, on the web, or via the Starbucks's app. One of the goals of every marketing campaign is to bring in more profit, that is, the profit generated must be higher than marketing costs. The campaign aims to attract the customers that would eventually buy the product. Targeting people who are not likely to buy Starbucks drinks is not ideal. We need to find people who possess a high probability to buy Starbucks products by using promotions.

**Problem statement:**

The data provided must be analyzed to find patterns between various features and find out which offer is appropriate to give to which kind of customers. That way, the offer leads that customer to make a purchase at Starbucks.

**Datasets and inputs:**

The data set that is going to be used in this project is provided by Udacity and Starbucks as part of the Machine Learning Engineer Nanodegree program. It contains simulated data that mimics customer behavior on the Starbucks mobile app. The program used to create the data simulates how people make purchasing decisions and how those decisions are influenced by promotional offers. Each person in the simulation has some hidden traits that influence their purchasing patterns and are associated with their observable traits. People produce various events, including receiving offers, opening offers, and making purchases. Only the amounts of each transaction or offer are recorded.

## Data Dictionary

profile.json

Rewards program users (17000 users x 5 fields)

- gender: (categorical) M, F, O, or null
- age: (numeric) missing value encoded as 118
- id: (string/hash)
- became\_member\_on: (date) format YYYYMMDD
- income: (numeric)

portfolio.json

Offers sent during 30-day test period (10 offers x 6 fields)

- reward: (numeric) money awarded for the amount spent
- channels: (list) web, email, mobile, social
- difficulty: (numeric) money required to be spent to receive reward
- duration: (numeric) time for offer to be open, in days
- offer\_type: (string) bogo, discount, informational
- id: (string/hash)

transcript.json

Event log (306648 events x 4 fields)

- person: (string/hash)
- event: (string) offer received, offer viewed, transaction, offer completed
- value: (dictionary) different values depending on event type
  - offer id: (string/hash) not associated with any "transaction"
  - amount: (numeric) money spent in "transaction"
  - reward: (numeric) money gained from "offer completed"
- time: (numeric) hours after start of test

## **Solution Statement:**

To find a solution to the above stated problem, in this project we apply machine learning techniques to understand customers' behavior by analyzing their previous transactions with Starbucks. To find out which offer to send to a specific kind of customer, we perform Exploratory Data Analysis and find information such as which offer the customers are most interested in, demographics details of those customers that make the purchase using the offer, and others. To find out the appropriate response of a customer to an offer, we will use models such as Logistic regression, Decision Tree classifier and Random Forest classifier to determine the data that best represents our data.

## **Benchmark Model:**

A quick and fairly accurate model can be considered as a benchmark. we will use the KNeighborsClassifier to build the benchmark, as it is a fast and standard method for binary classification machine learning problems and evaluate the model result using accuracy as the evaluation metric.

## **Evaluation Metrics:**

We use accuracy in this project as an evaluation metric. We use accuracy since it is one of the common evaluation metrics in classification problems, that is the total number of correct predictions divided by the total number of predictions made for a dataset.

**Project Design:**

The following are the steps we will take to perform this project:

- ☐ Loading the data in the Udacity workspace
- ☐ Cleaning up the data to understand the dataset and for better modeling.
- ☐ Do exploratory analysis on the three data frames that will be created
- ☐ Building different models to determine the most appropriate one for the data
- ☐ Perform benchmark model and evaluation metric.
- ☐ Put forth the findings and the working in the project in a blog post.