

# Capstone Proposal: Starbucks Capstone Challenge

## Domain Background

The project is about creating an offer recommendation engine for the Starbucks marketing team. Every few days, Starbucks is sending out offers to users of the mobile app. There are mainly three main types of offers: discount, BOGO (buy one get one) and informational offers. Not all Starbucks customers receive the same offer.

## Problem Statement

The problem of the project is to find a solution that will suggest which customers should get which type of offers.

## The Datasets and Inputs

The dataset is the simplified version of the real Starbucks app data. The the dataset contains three files: portfolio.json, profile.json, and transcript.json. The task is to combine transaction data with the demographic and offer data to identify which type of customers fit well to which type of offers.

## A Benchmark Model

As a benchmark model, a Logistic Regression will be used. Logistic regression is the go-to model for problems with two class values. Therefore the dataset will be combined and cleaned for binary classification approach. Mainly the goal will be to predict if the user should receive the BOGO or the discount offer.

## Evaluation metrics

As the main evaluation metric, the accuracy score will be used. For this type of problem, the accuracy score is perfectly fine. There is no need to use f1 or any other metrics because in case of false positives or false negatives there will be no damage to customers or the company. It's ok if in some cases the customer will receive the wrong right offer.

## Project Design

There will be 5 steps in the project.

1. Exploratory Analysis: During the EDA phase there will be the expoloration of data and some basic metrics calculation. The metrics calculation step is needed to get some understand of the different customer type and the performance.
2. Data Cleaning: During the data cleaning step NaN values, other values that create noise, and values that make no sense will be removed from the datasets.

3. Feature Engineering: The main goal of feature engineering is to prepare the features for the algorithms. During this step label encoding, one hot encoding will be used to encode values into the proper data structure.
4. Algorithm Selection: As the baseline Logistic Regression algorithm was chosen. The algorithm selection will start with some basic algorithms and will gradually move towards more advanced algorithms.
5. Model Training & Tuning: During the training and tuning phase the models will be trained and the best algorithm will be chosen. Grid search will be used to fine-tune the models.