



# CAPSTONE PROJECT PROPOSAL

Machine Learning Engineer  
Nanodegree



# INFO

Toni Magdy

15<sup>th</sup> August, 2020

Starbucks Capstone Challenge

# DOMAIN BACKGROUND

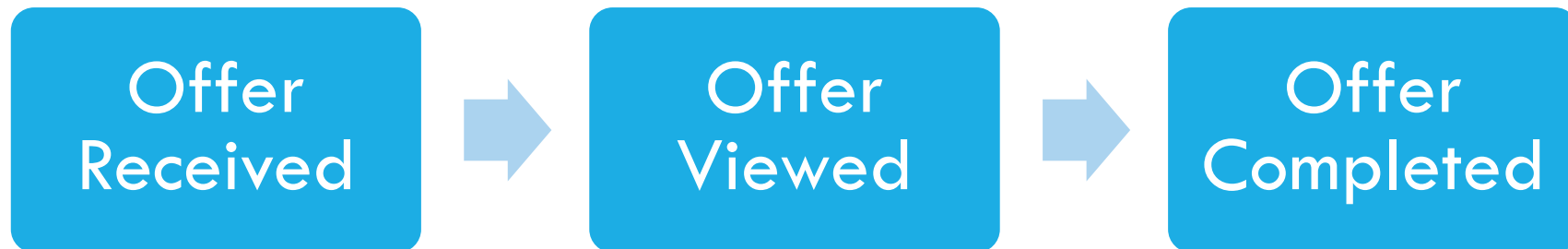
- This project from the marketing system, An application by Starbucks to keep in touch with its costumers and to make orders online.
- Once every few days, Starbucks sends out an offer to users of the mobile app. An offer can be merely an advertisement for a drink or an actual offer such as a discount or BOGO (buy one get one free).
- 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.

# PROBLEM STATEMENT

- Starbucks needs a way to send to each customer the right offer.
- Our goal is to analyze historical app data to find most appropriate offer for each one of the customers.
- The appropriate offer when the customer sees the offer and buy the products under the offer influence.
- The user can receive an offer, never actually view the offer, and still complete the offer, there will be an offer completion record in the data set, however, the customer was not influenced by the offer because the customer never viewed the offer.

# PROBLEM STATEMENT (CONT.)

- The offer lifecycle



# DATASETS

## The data is contained in three files:

- **portfolio.json:** containing offer ids 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.

# DATASETS (CONT.)

## **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

# DATASETS (CONT.)

## **transcript.json**

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



# SOLUTION STATEMENT

- In order to solve this problem, we will build a machine learning model to study costumers behaviors.
- The model should predict that the costumer will complete each of the offers or not.
- The model will choose the best offer to send to the user

# BENCHMARK MODEL

- We will use Logistic Regression as a simple machine learning algorithm to compare the results with.
- Logistic Regression is simple and easy to implement.

# EVALUATION METRICS

- Accuracy
- Precision/Recall
- $F_1$  Score

# PROJECT DESIGN

- Data loading and exploration: load the dataset and present some data visualization in order to understand the data.
- Data Cleaning: clean the dataset and fix any issues.
- Feature Engineering: prepare the data to be suitable for the model.
- Split Data: split the data into training and test sets.
- Train the model: train the machine learning model.
- Train the benchmark model: train the benchmark model to compare the results with.
- Evaluate the models: Test the models and compare the results.



**THANK YOU!**