Name: Andrew Kalil
Subject: Capstone Project Proposal

Starbucks Capstone Challenge

Project's domain background

Due to the recent advancements in technology, the world is causing us to adapt extremely fast to the changes. It is thanks to the technology that I am able to write this proposal and email it to a mentor to be revised. With this being said, it is no secret that large enterprises and multi millionaire companies use this amazing invention called "the internet" to reach a larger audience and grow their business.

Starbucks is one of these many businesses that take advantage of the internet to obtain new customers daily. The following are some of the ways that Starbucks uses technology to reach new customers.

1) Free WiFi: What a better way to attract buyers and tempt them to buy your product than giving them free access to WiFi almost any time of the day?
2) Mobile or online purchase: Studies show that 11% of starbucks sales come from mobile purchases.
3) Discounts and Offers: Like any other store, Starbucks sends their customers offers which can be very tempting but rewarding. This therefore helps to create loyal customers and increase sales.

Therefore, for this project we research the effects that some of the technological advancements have had over the last decade for Starcbucks. We will specifically look at some of the offers made to a selected number of customers affects purchase behaviors and how this information can help to increase sales.

Problem statement:

This 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.

Keep in mind as well that someone using the app might make a purchase through the app without having received an offer or seen an offer.

Datasets and inputs:

The datasets used for this project are the following:
profile.json

- id (string) - offer id
- offer_type (string) - type of offer ie BOGO, discount, 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)

Portfolio.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

Transcript.json

- event (str) - record description (ie 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

Benchmark model

As researched by the Sardar Patel Institute of Technology, we will assimilate a systematic approach which is a combination of clustering and recommender systems. It offers a more improved approach to customer segmentation.

Evaluation metrics

The planned metrics for this project are:

1. <u>Elbow method</u>: For determining the optimal number k-means clusters by plotting the value of the cost function produced by k values.
2. <u>Silhouette value</u>: Measures similarity between a point and its own cluster (cohesion) compared to other clusters (separation).
3. <u>Davies Bouldin metric</u>: Average similarity measure of each cluster to its most similar cluster. In this case similarity is the ratio of distances within the cluster to distances

between clusters. The minimum score is zero, and lower values indicate better clustering.

Project design

The project for the problem is as following:
1. Establish a workspace in a jupyter environment
2. Download the data provided into my jupyter notebook
3. Initial data cleaning
4. Perform exploratory analysis on the data
5. Cleaning up the data as needed for modeling
6. Experimenting to determine most appropriate unsupervised learning model to use for the data, whether that be k-means clustering, DBSCAN, or some other model
7. Leveraging our benchmark model and evaluation metric(s) to ensure sanity
8. Summarizing our findings writing a blog post

References
1) https://www.nytimes.com/2010/06/15/technology/15starbux.html
2) https://www.qsrmagazine.com/consumer-trends/5-ways-starbucks-innovating-customer-experience