Data Scientist Project

Starbucks Capstone Report

1- **Project Definition:**

In reality, the company often make a promotion to attract new users and maintain the long- term customer. What the main promotion strategies are: 1) Discounting, deduct an amount from the total; 2) Buy one get one method; 3) Informational.

The Starbuck simulate the data that mimics customer behavior on the Starbucks rewards mobile app. There are three main conversion paths:

1. The BOGO(Buy one get one) And the discount type
    - Offer Received => Offer Viewed => Offer Completed => Promotion Effect
    - Offer Received => Offer Completed => Promotion Effect
2. The information type:
    - Offer Received => Offer Viewed => Promotion Effect

All the files store the transaction record, the users information record, and the pormotion information record, which tend to figure out How is the relationship between the promotion and the users.

2- **Problem Statement**

The Starbucks provides different promotion type. At the customer side, maybe they make a different choice to respond to the different promotion. And there is another question that whether the users go to purchase the product more frequently. Right now, the data set contains the simulated data that mimics customer behavior on the Starbucks rewards mobile app that offer three different promotion. Now we have the appropriate data, and we are curiouse about the consumer behavior. So we put one question: What is the consumed amount after the promotion?

3- **Analysis**

The data is contained in three files:
- portfolio.json, containing offer ids and meta data about each offer
  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) - running time
  channels (list of strings)

- profile.json - demographic data for each customer
  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

- ranscript.json - records for transactions, offers received, offers viewed, and offers completed
event (str) - record description (ie transaction, offer received, offer viewed, etc.)
person (str) - customer id
time (int) - time in hours. The data begins at time t=0
value - (dict of strings) - either an offer id or transaction amount depending on the record

4-   Data Wrangling

we find that there are some data quantity problem. Now we need to preprocess the data, so that we can continue to analyze.

5-   Data Visualization

we explore what the trend is that the users subscribe the membership.Before 2017 July, almost 300 users join in the Starbucks membership in every 30 days. Then the new member trend increase at the top that is almost 900 users, between 2017 July and 2018 February. After 2018 February, the trend goes down to almost 600 users.