# **Data Science Nanodegree Program: A case study for Starbucks**

As a final step in the Data Science Nanodegreee program of Udacity. We are given an artificial data supposedly mimicing the customer behaviour for a spcific product type seld in Sturbucks stores. As a promotion prolicy Sturbucks send offer to their customers via different channels like web,email,mobile and social for each offer type. As a company policy each customer take different offers in different days based on the demographic of the customer.

We have tree different offer type fort the product given in the data as follows:

1. Informational
2. Discount
3. BOGO

Informatiol offer does not provide any reward for the customers , it just informs the customers about the product. In discount offer type customers can get reward if the transaction amount for the purchases exceeds the predefined amount in a given time horizon like 5-7 days. BOGO offer type is the abbreviaton of “ buy one get one”. Customers receiving BOGO offer can buy two same product by paying just for one product.

In this project we are provided three datasets. These datasets provide information about the offers, customer demographics and transactions of each customer int he given time frame. We will be analysing each dataset belove.

# **Problem Statement**

In this project we will be analysing the performance of each ogger type for different customer groups and make suggestion for a feasible promotion policy by sending ideal offer to each customer category so as to increase customer purchases and make profit. Besides, we will try to predict the reaction of the customer to a specific offer type by acknowleging the customer demographics and attributes of the offer.

**Data Exploration**

* **Portfolio data**: contains atributes of each offer type (10 rows)

1. **id** – Identity of the offer type
2. **offer\_type** – category of the offer (discount, informational or bogo)
3. **channels** – which promotion channels is used (web, email,mobile and/or social)
4. **duration** – how long (days) the offer is valid after receiving
5. **difficulty** – minumum purchase quantity needed to be eligible for the offer
6. **reward** – the amount the the customer will get as a discount when the difficulty criteria is met in duration period.

Table

Description automatically generated

* **Profile data:** demograchics of customers (17000 rows)

1. gender – gender of the customer (M: Male, F:Female, O: Other)

2. age – age of the customer

3. id – customer id

4. become\_member\_on – the date customer created an account

5. income – annual income of the customer

Table

Description automatically generated

* transcript data – contains all the user activity on a time table (306534 rows)

1. event — event description ( transaction, offer received, offer viewed or offer completed)
2. person — customer id
3. time (int) — time in hours since start of test. The data begins at time t=0
4. value — (dict of strings) — details of each record. If a customer received or viewed an offer then it contains the offer id. If a customer completed an offer it contains offer id that the customer completed. And lastly if the record is a purchase record, it contsins the transaction amount.

We need to get familiar with the transcript data to obtain seasonable analysis from the data. You might think of a scenario , how much a pustomer purchased to get reward for the given offer. We cannot reach this info by just looking at one line of the data. This scenario is recorded as two seperate row in the data. First it records which offer the customer used and then create another record for purchase amount. Here is a glimple of the record for a customer;

Table

Description automatically generated

At time=408 te customer receive the offer “ae264e3637204a6fb9bb56bc8210ddfd” and see it. 102 hours later the customer make a purchase of 21.72 $ and get 10 $ reward. We can see that this offer is a BOGO type offer and the threshold value is 10 $. Since the customer made a purchase above 20 $ , the customer become eligible to this offer. You might have noticed that there is another offer completed at time t=510. Since the customer made a purchase above 20 $ the customer still satisfy the threshold for the seond BOGO offer and so the customer get 5 $ reward at =510.

We covered the input datasets and now in the next chapters we will be playing with the data by cleaning the datasets to make our analysis easier and understandible.

**Data Cleaning & Processing**

In this aort we will change the columns names , create new columns from the exisiting columns ans drop outlier rows.

profile DataFrame 🡪 new\_profile DataFrame

Change “id” column to “customer\_id”

Label endoce the “channel” column

Change duration column values from days to hours and rename it as “offer\_duration\_hour”