Starbucks_Capstone_notebook

May 23, 2020

1 Starbucks Capstone Challenge

1.0.1 Introduction

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.

Every offer has a validity period before the offer expires. As an example, a BOGO offer might be valid for only 5 days. For example, if an informational offer has 7 days of validity, you can assume the customer is feeling the influence of the offer for 7 days after receiving the advertisement.

We are given transactional data showing user purchases made on the app including the timestamp of purchase and the amount of money spent on a purchase. This transactional data also has a record for each offer that a user receives as well as a record for when a user actually views the offer. There are also records for when a user completes an offer.

Someone using the app might make a purchase through the app without having received an offer or seen an offer.

1.0.2 Example

To give an example, a user could receive a discount offer buy 10 dollars get 2 off on Monday. The offer is valid for 10 days from receipt. If the customer accumulates at least 10 dollars in purchases during the validity period, the customer completes the offer.

However, there are a few things to watch out for in this data set. Customers do not opt into the offers that they receive; in other words, a user can receive an offer, never actually view the offer, and still complete the offer. For example, a user might receive the "buy 10 dollars get 2 dollars off offer", but the user never opens the offer during the 10 day validity period. The customer spends 15 dollars during those ten days. 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.

1.0.3 Project Overview

In this project, an attempt has been mae to analyse how Starbucks customers use the app. The promotion for starbucks using various offers and are these offers actually helpful. The data sets used in this project contains simulated data that mimics customer behavior on the Starbucks rewards mobile app. The customer's behaviour has been attempted to analyse based on the dataset.

1.0.4 Cleaning

Cleaning data is especially important and tricky. From a business perspective, if a customer is going to make a 10 dollar purchase without an offer anyway, you wouldn't want to send a buy 10 dollars get 2 dollars off offer. It is important to assess what a certain demographic group will buy when not receiving any offers.

1.0.5 Analysis

Business understanding and data exploration The dataset should be well used and all aspects of the data need to be considered to build the model, but first the data should be explored, the relationship between various variables need to be extracted to form an intuition.

2 Data Sets

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

Here is the schema and explanation of each variable in the files:

portfolio.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)

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

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

Note: If you are using the workspace, you will need to go to the terminal and run the command `conda update pandas` before reading in the files. This is because the version of pandas in the workspace cannot read in the transcript.json file correctly, but the newest version of pandas can. You can access the termnal from the orange icon in the top left of this notebook.

You can see how to access the terminal and how the install works using the two images below. First you need to access the terminal:

Then you will want to run the above command:

Finally, when you enter back into the notebook (use the jupyter icon again), you should be able to run the below cell without any errors.

DATA LOADING

```
In [1]: import pandas as pd
        import numpy as np
        import math
```

```
import json
% matplotlib inline


# read in the json files
portfolio = pd.read_json('data/portfolio.json', orient='records', lines=True)
profile = pd.read_json('data/profile.json', orient='records', lines=True)
transcript = pd.read_json('data/transcript.json', orient='records', lines=True)
```

In [ ]:

In [ ]:

```
In [2]: from sklearn.preprocessing import OneHotEncoder,MultiLabelBinarizer
        import matplotlib.pyplot as plt

        import seaborn as sns
```

DATA EXPLORATION The below cells contain the exploration of data and an attempt has been made to analyse data and derive relation out of it using plots

```
In [3]: portfolio.head()
```

```
Out[3]:                          channels  difficulty  duration  \
        0       [email, mobile, social]          10         7
        1  [web, email, mobile, social]          10         5
        2          [web, email, mobile]           0         4
        3          [web, email, mobile]           5         7
        4                  [web, email]          20        10

                                         id     offer_type  reward
        0  ae264e3637204a6fb9bb56bc8210ddfd           bogo      10
        1  4d5c57ea9a6940dd891ad53e9dbe8da0           bogo      10
        2  3f207df678b143eea3cee63160fa8bed  informational       0
        3  9b98b8c7a33c4b65b9aebfe6a799e6d9           bogo       5
        4  0b1e1539f2cc45b7b9fa7c272da2e1d7       discount       5
```

```
In [4]: portfolio.shape
```

```
Out[4]: (10, 6)
```

```
In [5]: profile.head()
```

```
Out[5]:    age  became_member_on gender                                id    income
        0  118          20170212   None  68be06ca386d4c31939f3a4f0e3dd783       NaN
        1   55          20170715      F  0610b486422d4921ae7d2bf64640c50b  112000.0
        2  118          20180712   None  38fe809add3b4fcf9315a9694bb96ff5       NaN
        3   75          20170509      F  78afa995795e4d85b5d9ceeca43f5fef  100000.0
        4  118          20170804   None  a03223e636434f42ac4c3df47e8bac43       NaN
```

In [6]: profile.shape

Out[6]: (17000, 5)

In [7]: transcript.head()

Out[7]:
```
            event                            person  time  \
0  offer received  78afa995795e4d85b5d9ceeca43f5fef     0
1  offer received  a03223e636434f42ac4c3df47e8bac43     0
2  offer received  e2127556f4f64592b11af22de27a7932     0
3  offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
4  offer received  68617ca6246f4fbc85e91a2a49552598     0

                                              value
0  {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
1  {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
2  {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
3  {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
4  {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
```

In [8]: transcript.shape

Out[8]: (306534, 4)

In [9]: portfolio.dtypes,transcript.dtypes,profile.dtypes

Out[9]:
```
(channels      object
 difficulty     int64
 duration       int64
 id            object
 offer_type    object
 reward         int64
 dtype: object, event     object
 person    object
 time       int64
 value     object
 dtype: object, age                   int64
 became_member_on      int64
 gender               object
 id                   object
 income              float64
 dtype: object)
```

In [ ]:

In [10]: portfolio['offer_type']

Out[10]:
```
0             bogo
1             bogo
```

```
2    informational
3             bogo
4         discount
5         discount
6         discount
7    informational
8             bogo
9         discount
Name: offer_type, dtype: object
```

In [11]:
```
data=pd.get_dummies(portfolio['offer_type'])
```

In [ ]:

In [12]:
```
portfolio = pd.concat([portfolio,data],axis=1)
```

In [ ]:

In [13]:
```
portfolio
```

Out[13]:
```
                       channels  difficulty  duration  \
0       [email, mobile, social]          10         7
1  [web, email, mobile, social]          10         5
2          [web, email, mobile]           0         4
3          [web, email, mobile]           5         7
4                  [web, email]          20        10
5  [web, email, mobile, social]           7         7
6  [web, email, mobile, social]          10        10
7       [email, mobile, social]           0         3
8  [web, email, mobile, social]           5         5
9          [web, email, mobile]          10         7

                                 id     offer_type  reward  bogo  discount  \
0  ae264e3637204a6fb9bb56bc8210ddfd           bogo      10     1         0
1  4d5c57ea9a6940dd891ad53e9dbe8da0           bogo      10     1         0
2  3f207df678b143eea3cee63160fa8bed  informational       0     0         0
3  9b98b8c7a33c4b65b9aebfe6a799e6d9           bogo       5     1         0
4  0b1e1539f2cc45b7b9fa7c272da2e1d7       discount       5     0         1
5  2298d6c36e964ae4a3e7e9706d1fb8c2       discount       3     0         1
6  fafdcd668e3743c1bb461111dcafc2a4       discount       2     0         1
7  5a8bc65990b245e5a138643cd4eb9837  informational       0     0         0
8  f19421c1d4aa40978ebb69ca19b0e20d           bogo       5     1         0
9  2906b810c7d4411798c6938adc9daaa5       discount       2     0         1

   informational
0              0
1              0
2              1
3              0
```

```
4              0
5              0
6              0
7              1
8              0
9              0
```

DATA WRANGLING

```
In [14]: mlb = MultiLabelBinarizer() #channels column is exploited to break it into various colu
         data1 = pd.DataFrame(mlb.fit_transform(portfolio['channels']),
                           columns=mlb.classes_,
                           index=portfolio['channels'].index)

In [15]: portfolio = portfolio.drop('channels', axis=1)  # channels column is dropped and rest d
         portfolio = pd.concat([portfolio, data1], axis=1, sort=False)

In [16]: portfolio

Out[16]:    difficulty  duration                                id     offer_type  \
         0          10         7  ae264e3637204a6fb9bb56bc8210ddfd           bogo
         1          10         5  4d5c57ea9a6940dd891ad53e9dbe8da0           bogo
         2           0         4  3f207df678b143eea3cee63160fa8bed  informational
         3           5         7  9b98b8c7a33c4b65b9aebfe6a799e6d9           bogo
         4          20        10  0b1e1539f2cc45b7b9fa7c272da2e1d7       discount
         5           7         7  2298d6c36e964ae4a3e7e9706d1fb8c2       discount
         6          10        10  fafdcd668e3743c1bb461111dcafc2a4       discount
         7           0         3  5a8bc65990b245e5a138643cd4eb9837  informational
         8           5         5  f19421c1d4aa40978ebb69ca19b0e20d           bogo
         9          10         7  2906b810c7d4411798c6938adc9daaa5       discount

            reward  bogo  discount  informational  email  mobile  social  web
         0      10     1         0              0      1       1       1    0
         1      10     1         0              0      1       1       1    1
         2       0     0         0              1      1       1       0    1
         3       5     1         0              0      1       1       0    1
         4       5     0         1              0      1       0       0    1
         5       3     0         1              0      1       1       1    1
         6       2     0         1              0      1       1       1    1
         7       0     0         0              1      1       1       1    0
         8       5     1         0              0      1       1       1    1
         9       2     0         1              0      1       1       0    1

In [17]: profile.isna().sum()

Out[17]: age                    0
         became_member_on       0
         gender              2175
         id                     0
```

```
income              2175
dtype: int64
```

In [18]:
```
profile['gender'].fillna('NA', inplace=True)
profile['income']=profile['income'].fillna(profile['income'].mean())
```

In [19]:
```
profile.isna().sum()
```

Out[19]:
```
age                 0
became_member_on    0
gender              0
id                  0
income              0
dtype: int64
```

In [20]:
```
transcript.isna().sum()
```

Out[20]:
```
event     0
person    0
time      0
value     0
dtype: int64
```

In [21]:
```
portfolio.isna().sum()
```

Out[21]:
```
difficulty       0
duration         0
id               0
offer_type       0
reward           0
bogo             0
discount         0
informational    0
email            0
mobile           0
social           0
web              0
dtype: int64
```

In [22]:
```
transcript.shape
```

Out[22]:
```
(306534, 4)
```

In [23]:
```
profile.shape
```

Out[23]:
```
(17000, 5)
```

In [24]:
```
transcript['value'][306510]
```

Out[24]:
```
{'amount': 14.31}
```

In [25]:
```
type_ = [] # a loop to extract different types of values in events column
for idx, row in transcript.iterrows():
    for k in row['value']:
        if k in type_:
            continue
        else:
            type_.append(k)

type_
```

Out[25]:
```
['offer id', 'amount', 'offer_id', 'reward']
```

In [26]:
```
for idx, row in transcript.iterrows():
    print(row['value'])
    for k,i in row['value'].items():
        print(k)
        print(i)
        break
    break
```

```
{'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
offer id
9b98b8c7a33c4b65b9aebfe6a799e6d9
```

The values from value column in trnascript is extracted and put into three columns for better analysis

In [27]:
```
transcript['offer_id']=''
transcript['reward']=0
transcript['amount']=0
```

In [28]:
```
transcript
```

Out[28]:
```
              event                            person  time  \
0    offer received  78afa995795e4d85b5d9ceeca43f5fef     0
1    offer received  a03223e636434f42ac4c3df47e8bac43     0
2    offer received  e2127556f4f64592b11af22de27a7932     0
3    offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
4    offer received  68617ca6246f4fbc85e91a2a49552598     0
5    offer received  389bc3fa690240e798340f5a15918d5c     0
6    offer received  c4863c7985cf408faee930f111475da3     0
7    offer received  2eeac8d8feae4a8cad5a6af0499a211d     0
8    offer received  aa4862eba776480b8bb9c68455b8c2e1     0
9    offer received  31dda685af34476cad5bc968bdb01c53     0
10   offer received  744d603ef08c4f33af5a61c8c7628d1c     0
11   offer received  3d02345581554e81b7b289ab5e288078     0
12   offer received  4b0da7e80e5945209a1fdddfe813dbe0     0
13   offer received  c27e0d6ab72c455a8bb66d980963de60     0
```

14	offer received	d53717f5400c4e84affdaeda9dd926b3	0
15	offer received	f806632c011441378d4646567f357a21	0
16	offer received	d058f73bf8674a26a95227db098147b1	0
17	offer received	65aba5c617294649aeb624da249e1ee5	0
18	offer received	ebe7ef46ea6f4963a7dd49f501b26779	0
19	offer received	1e9420836d554513ab90eba98552d0a9	0
20	offer received	868317b9be554cb18e50bc68484749a2	0
21	offer received	f082d80f0aac47a99173ba8ef8fc1909	0
22	offer received	102e9454054946fda62242d2e176fdce	0
23	offer received	4beeb3ed64dd4898b0edf2f6b67426d3	0
24	offer received	9f30b375d7bd4c62a884ffe7034e09ee	0
25	offer received	25c906289d154b66bf579693f89481c9	0
26	offer received	6e014185620b49bd98749f728747572f	0
27	offer received	02c083884c7d45b39cc68e1314fec56c	0
28	offer received	c0d210398dee4a0895b24444a5fcd1d2	0
29	offer received	8be4463721e14d7fa600686bf8c8b2ed	0
...	...	...	...
306504	transaction	8524d450673b4c24869b6c94380006de	714
306505	transaction	b895c57e8cd047a8872ce02aa54759d6	714
306506	offer completed	b895c57e8cd047a8872ce02aa54759d6	714
306507	offer viewed	8dda575c2a1d44b9ac8e8b07b93d1f8e	714
306508	transaction	8431c16f8e1d440880db371a68f82dd0	714
306509	offer completed	8431c16f8e1d440880db371a68f82dd0	714
306510	transaction	ba620885e51c4b0ea64a4f61daad494f	714
306511	transaction	a1a8f40407c444cc848468275308958a	714
306512	transaction	8d80970192fa496f99d6b45c470a4b60	714
306513	transaction	bde275066f3c4fa0bff3093e3b866a2c	714
306514	transaction	f1e4fd36e5a0446f83861308bddf6945	714
306515	transaction	0b64be3b241c4407a5c9a71781173829	714
306516	transaction	86d03d35d7e0434b935e7743e83be3a0	714
306517	transaction	3408fd05c781401f8442fb6dbaaea9c7	714
306518	transaction	1593d617fac246ef8e50dbb0ffd77f5f	714
306519	transaction	f1b31d07b5d84f69a2d5f1d07843989e	714
306520	transaction	2ce987015ec0404a97ba333e8e814090	714
306521	transaction	2e33545f0a764d27b2ccff95fc8d72c4	714
306522	transaction	d1c4500ace2e45e9a45d3cd2fccac8d8	714
306523	transaction	b65affd9e07346a1906364a396950e3d	714
306524	transaction	d613ca9c59dd42f497bdbf6178da54a7	714
306525	transaction	eec70ab28af74a22a4aeb889c0317944	714
306526	transaction	24f56b5e1849462093931b164eb803b5	714
306527	offer completed	24f56b5e1849462093931b164eb803b5	714
306528	transaction	5ca2620962114246ab218fc648eb3934	714
306529	transaction	b3a1272bc9904337b331bf348c3e8c17	714
306530	transaction	68213b08d99a4ae1b0dcb72aebd9aa35	714
306531	transaction	a00058cf10334a308c68e7631c529907	714
306532	transaction	76ddbd6576844afe811f1a3c0fbb5bec	714
306533	transaction	c02b10e8752c4d8e9b73f918558531f7	714

```
                                                  value offer_id  reward  \
0        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}              0
1        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
2        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}              0
3        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}              0
4        {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}              0
5        {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}              0
6        {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}              0
7        {'offer id': '3f207df678b143eea3cee63160fa8bed'}              0
8        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
9        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
10       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
11       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
12       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}              0
13       {'offer id': '3f207df678b143eea3cee63160fa8bed'}              0
14       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
15       {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}              0
16       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
17       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}              0
18       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}              0
19       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}              0
20       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}              0
21       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}              0
22       {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}              0
23       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}              0
24       {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}              0
25       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}              0
26       {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}              0
27       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}              0
28       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}              0
29       {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}              0
...                                                   ...      ...     ...
306504                                {'amount': 4.89}              0
306505                                {'amount': 4.48}              0
306506  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...              0
306507   {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}              0
306508                                {'amount': 1.19}              0
306509  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...              0
306510                               {'amount': 14.31}              0
306511                                {'amount': 2.37}              0
306512                                {'amount': 6.92}              0
306513                               {'amount': 12.73}              0
306514                                 {'amount': 8.2}              0
306515                                 {'amount': 2.6}              0
306516                                 {'amount': 9.2}              0
306517                                {'amount': 11.7}              0
306518                               {'amount': 40.67}              0
306519                               {'amount': 31.13}              0
```

```
306520                     {'amount': 1.6400000000000001}                  0
306521                                 {'amount': 17.35}                   0
306522                                  {'amount': 4.42}                   0
306523                                 {'amount': 18.35}                   0
306524                                 {'amount': 25.14}                   0
306525                                 {'amount': 43.58}                   0
306526                                 {'amount': 22.64}                   0
306527  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...                  0
306528                                   {'amount': 2.2}                   0
306529                     {'amount': 1.5899999999999999}                  0
306530                                  {'amount': 9.53}                   0
306531                                  {'amount': 3.61}                   0
306532                     {'amount': 3.5300000000000002}                  0
306533                                  {'amount': 4.05}                   0

        amount
0            0
1            0
2            0
3            0
4            0
5            0
6            0
7            0
8            0
9            0
10           0
11           0
12           0
13           0
14           0
15           0
16           0
17           0
18           0
19           0
20           0
21           0
22           0
23           0
24           0
25           0
26           0
27           0
28           0
29           0
...        ...
306504       0
```

```
306505        0
306506        0
306507        0
306508        0
306509        0
306510        0
306511        0
306512        0
306513        0
306514        0
306515        0
306516        0
306517        0
306518        0
306519        0
306520        0
306521        0
306522        0
306523        0
306524        0
306525        0
306526        0
306527        0
306528        0
306529        0
306530        0
306531        0
306532        0
306533        0

[306534 rows x 7 columns]
```

In [29]:
```
# Iterate over transcript table, check value column and update it, put each key in sepe
for idx, row in transcript.iterrows():
    for k_y,v_l in row['value'].items():
        if k_y == 'offer_id' or k_y == 'offer id':
            transcript.at[idx, 'offer_id']= v_l
        if k_y == 'amount':
            transcript.at[idx, 'amount']= v_l
        if k_y == 'reward':
            transcript.at[idx, 'reward']= v_l
```

In [30]:
```
transcript['offer_id'].loc[1]
```

Out[30]:
```
'0b1e1539f2cc45b7b9fa7c272da2e1d7'
```

In [31]:
```
transcript
```

```
Out[31]:                   event                            person  time  \
        0        offer received  78afa995795e4d85b5d9ceeca43f5fef     0
        1        offer received  a03223e636434f42ac4c3df47e8bac43     0
        2        offer received  e2127556f4f64592b11af22de27a7932     0
        3        offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
        4        offer received  68617ca6246f4fbc85e91a2a49552598     0
        5        offer received  389bc3fa690240e798340f5a15918d5c     0
        6        offer received  c4863c7985cf408faee930f111475da3     0
        7        offer received  2eeac8d8feae4a8cad5a6af0499a211d     0
        8        offer received  aa4862eba776480b8bb9c68455b8c2e1     0
        9        offer received  31dda685af34476cad5bc968bdb01c53     0
        10       offer received  744d603ef08c4f33af5a61c8c7628d1c     0
        11       offer received  3d02345581554e81b7b289ab5e288078     0
        12       offer received  4b0da7e80e5945209a1fdddfe813dbe0     0
        13       offer received  c27e0d6ab72c455a8bb66d980963de60     0
        14       offer received  d53717f5400c4e84affdaeda9dd926b3     0
        15       offer received  f806632c011441378d4646567f357a21     0
        16       offer received  d058f73bf8674a26a95227db098147b1     0
        17       offer received  65aba5c617294649aeb624da249e1ee5     0
        18       offer received  ebe7ef46ea6f4963a7dd49f501b26779     0
        19       offer received  1e9420836d554513ab90eba98552d0a9     0
        20       offer received  868317b9be554cb18e50bc68484749a2     0
        21       offer received  f082d80f0aac47a99173ba8ef8fc1909     0
        22       offer received  102e9454054946fda62242d2e176fdce     0
        23       offer received  4beeb3ed64dd4898b0edf2f6b67426d3     0
        24       offer received  9f30b375d7bd4c62a884ffe7034e09ee     0
        25       offer received  25c906289d154b66bf579693f89481c9     0
        26       offer received  6e014185620b49bd98749f728747572f     0
        27       offer received  02c083884c7d45b39cc68e1314fec56c     0
        28       offer received  c0d210398dee4a0895b24444a5fcd1d2     0
        29       offer received  8be4463721e14d7fa600686bf8c8b2ed     0
        ...                 ...                               ...   ...
        306504      transaction  8524d450673b4c24869b6c94380006de   714
        306505      transaction  b895c57e8cd047a8872ce02aa54759d6   714
        306506  offer completed  b895c57e8cd047a8872ce02aa54759d6   714
        306507     offer viewed  8dda575c2a1d44b9ac8e8b07b93d1f8e   714
        306508      transaction  8431c16f8e1d440880db371a68f82dd0   714
        306509  offer completed  8431c16f8e1d440880db371a68f82dd0   714
        306510      transaction  ba620885e51c4b0ea64a4f61daad494f   714
        306511      transaction  a1a8f40407c444cc848468275308958a   714
        306512      transaction  8d80970192fa496f99d6b45c470a4b60   714
        306513      transaction  bde275066f3c4fa0bff3093e3b866a2c   714
        306514      transaction  f1e4fd36e5a0446f83861308bddf6945   714
        306515      transaction  0b64be3b241c4407a5c9a71781173829   714
        306516      transaction  86d03d35d7e0434b935e7743e83be3a0   714
        306517      transaction  3408fd05c781401f8442fb6dbaaea9c7   714
        306518      transaction  1593d617fac246ef8e50dbb0ffd77f5f   714
        306519      transaction  f1b31d07b5d84f69a2d5f1d07843989e   714
```

```
306520       transaction  2ce987015ec0404a97ba333e8e814090   714
306521       transaction  2e33545f0a764d27b2ccff95fc8d72c4   714
306522       transaction  d1c4500ace2e45e9a45d3cd2fccac8d8   714
306523       transaction  b65affd9e07346a1906364a396950e3d   714
306524       transaction  d613ca9c59dd42f497bdbf6178da54a7   714
306525       transaction  eec70ab28af74a22a4aeb889c0317944   714
306526       transaction  24f56b5e1849462093931b164eb803b5   714
306527   offer completed  24f56b5e1849462093931b164eb803b5   714
306528       transaction  5ca2620962114246ab218fc648eb3934   714
306529       transaction  b3a1272bc9904337b331bf348c3e8c17   714
306530       transaction  68213b08d99a4ae1b0dcb72aebd9aa35   714
306531       transaction  a00058cf10334a308c68e7631c529907   714
306532       transaction  76ddbd6576844afe811f1a3c0fbb5bec   714
306533       transaction  c02b10e8752c4d8e9b73f918558531f7   714

                                                     value  \
0         {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
1         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
2         {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
3         {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
4         {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
5         {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
6         {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
7         {'offer id': '3f207df678b143eea3cee63160fa8bed'}
8         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
9         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
10        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
11        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
12        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
13        {'offer id': '3f207df678b143eea3cee63160fa8bed'}
14        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
15        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
16        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
17        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
18        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
19        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
20        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
21        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
22        {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
23        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
24        {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
25        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
26        {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
27        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
28        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
29        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
...                                                    ...
306504                                  {'amount': 4.89}
```

```
306505                                        {'amount': 4.48}
306506  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306507   {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
306508                                        {'amount': 1.19}
306509  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306510                                       {'amount': 14.31}
306511                                        {'amount': 2.37}
306512                                        {'amount': 6.92}
306513                                       {'amount': 12.73}
306514                                         {'amount': 8.2}
306515                                         {'amount': 2.6}
306516                                         {'amount': 9.2}
306517                                        {'amount': 11.7}
306518                                       {'amount': 40.67}
306519                                       {'amount': 31.13}
306520                          {'amount': 1.6400000000000001}
306521                                       {'amount': 17.35}
306522                                        {'amount': 4.42}
306523                                       {'amount': 18.35}
306524                                       {'amount': 25.14}
306525                                       {'amount': 43.58}
306526                                       {'amount': 22.64}
306527  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306528                                         {'amount': 2.2}
306529                          {'amount': 1.5899999999999999}
306530                                        {'amount': 9.53}
306531                                        {'amount': 3.61}
306532                          {'amount': 3.5300000000000002}
306533                                        {'amount': 4.05}
```

	offer_id	reward	amount
0	9b98b8c7a33c4b65b9aebfe6a799e6d9	0	0
1	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
2	2906b810c7d4411798c6938adc9daaa5	0	0
3	fafdcd668e3743c1bb461111dcafc2a4	0	0
4	4d5c57ea9a6940dd891ad53e9dbe8da0	0	0
5	f19421c1d4aa40978ebb69ca19b0e20d	0	0
6	2298d6c36e964ae4a3e7e9706d1fb8c2	0	0
7	3f207df678b143eea3cee63160fa8bed	0	0
8	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
9	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
10	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
11	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
12	ae264e3637204a6fb9bb56bc8210ddfd	0	0
13	3f207df678b143eea3cee63160fa8bed	0	0
14	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0
15	fafdcd668e3743c1bb461111dcafc2a4	0	0
16	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0

```
17      2906b810c7d4411798c6938adc9daaa5       0        0
18      9b98b8c7a33c4b65b9aebfe6a799e6d9       0        0
19      ae264e3637204a6fb9bb56bc8210ddfd       0        0
20      2906b810c7d4411798c6938adc9daaa5       0        0
21      9b98b8c7a33c4b65b9aebfe6a799e6d9       0        0
22      4d5c57ea9a6940dd891ad53e9dbe8da0       0        0
23      2906b810c7d4411798c6938adc9daaa5       0        0
24      2298d6c36e964ae4a3e7e9706d1fb8c2       0        0
25      2906b810c7d4411798c6938adc9daaa5       0        0
26      f19421c1d4aa40978ebb69ca19b0e20d       0        0
27      ae264e3637204a6fb9bb56bc8210ddfd       0        0
28      9b98b8c7a33c4b65b9aebfe6a799e6d9       0        0
29      fafdcd668e3743c1bb461111dcafc2a4       0        0
...                                  ...     ...      ...
306504                                         0        4
306505                                         0        4
306506  fafdcd668e3743c1bb461111dcafc2a4       2        0
306507  0b1e1539f2cc45b7b9fa7c272da2e1d7       0        0
306508                                         0        1
306509  fafdcd668e3743c1bb461111dcafc2a4       2        0
306510                                         0       14
306511                                         0        2
306512                                         0        6
306513                                         0       12
306514                                         0        8
306515                                         0        2
306516                                         0        9
306517                                         0       11
306518                                         0       40
306519                                         0       31
306520                                         0        1
306521                                         0       17
306522                                         0        4
306523                                         0       18
306524                                         0       25
306525                                         0       43
306526                                         0       22
306527  fafdcd668e3743c1bb461111dcafc2a4       2        0
306528                                         0        2
306529                                         0        1
306530                                         0        9
306531                                         0        3
306532                                         0        3
306533                                         0        4

[306534 rows x 7 columns]
```

In [32]:
```
transcript.loc[306528]
```

Out[32]:
```
event                             transaction
person       5ca2620962114246ab218fc648eb3934
time                                      714
value                        {'amount': 2.2}
offer_id
reward                                      0
amount                                      2
Name: 306528, dtype: object
```

In [33]:
```
trans=transcript.drop('value',axis=1)
```

In [34]:
```
trans.head()
```

Out[34]:
```
             event                            person  time  \
0  offer received  78afa995795e4d85b5d9ceeca43f5fef     0
1  offer received  a03223e636434f42ac4c3df47e8bac43     0
2  offer received  e2127556f4f64592b11af22de27a7932     0
3  offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
4  offer received  68617ca6246f4fbc85e91a2a49552598     0

                           offer_id  reward  amount
0  9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0
1  0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0
2  2906b810c7d4411798c6938adc9daaa5       0       0
3  fafdcd668e3743c1bb461111dcafc2a4       0       0
4  4d5c57ea9a6940dd891ad53e9dbe8da0       0       0
```

In [35]:
```
trans['time']
```

Out[35]:
```
0            0
1            0
2            0
3            0
4            0
5            0
6            0
7            0
8            0
9            0
10           0
11           0
12           0
13           0
14           0
15           0
16           0
17           0
18           0
19           0
```

```
20          0
21          0
22          0
23          0
24          0
25          0
26          0
27          0
28          0
29          0
          ...
306504    714
306505    714
306506    714
306507    714
306508    714
306509    714
306510    714
306511    714
306512    714
306513    714
306514    714
306515    714
306516    714
306517    714
306518    714
306519    714
306520    714
306521    714
306522    714
306523    714
306524    714
306525    714
306526    714
306527    714
306528    714
306529    714
306530    714
306531    714
306532    714
306533    714
Name: time, Length: 306534, dtype: int64
```

In [36]:
```
profile.describe()
```

Out[36]:

	age	became_member_on	income
count	17000.000000	1.700000e+04	17000.000000
mean	62.531412	2.016703e+07	65404.991568

```
std        26.738580      1.167750e+04   20169.288288
min        18.000000      2.013073e+07   30000.000000
25%        45.000000      2.016053e+07   51000.000000
50%        58.000000      2.017080e+07   65404.991568
75%        73.000000      2.017123e+07   76000.000000
max       118.000000      2.018073e+07  120000.000000
```

In [37]: trans.describe()

```
Out[37]:
                   time         reward         amount
count   306534.000000  306534.000000  306534.000000
mean       366.382940       0.537219       5.570133
std        200.326314       1.805208      21.266669
min          0.000000       0.000000       0.000000
25%        186.000000       0.000000       0.000000
50%        408.000000       0.000000       0.000000
75%        528.000000       0.000000       7.000000
max        714.000000      10.000000    1062.000000
```

In [38]: portfolio.describe()

```
Out[38]:
        difficulty   duration     reward       bogo   discount  informational  \
count    10.000000  10.000000  10.000000  10.000000  10.000000      10.000000
mean      7.700000   6.500000   4.200000   0.400000   0.400000       0.200000
std       5.831905   2.321398   3.583915   0.516398   0.516398       0.421637
min       0.000000   3.000000   0.000000   0.000000   0.000000       0.000000
25%       5.000000   5.000000   2.000000   0.000000   0.000000       0.000000
50%       8.500000   7.000000   4.000000   0.000000   0.000000       0.000000
75%      10.000000   7.000000   5.000000   1.000000   1.000000       0.000000
max      20.000000  10.000000  10.000000   1.000000   1.000000       1.000000

        email     mobile     social        web
count    10.0  10.000000  10.000000  10.000000
mean      1.0   0.900000   0.600000   0.800000
std       0.0   0.316228   0.516398   0.421637
min       1.0   0.000000   0.000000   0.000000
25%       1.0   1.000000   0.000000   1.000000
50%       1.0   1.000000   1.000000   1.000000
75%       1.0   1.000000   1.000000   1.000000
max       1.0   1.000000   1.000000   1.000000
```

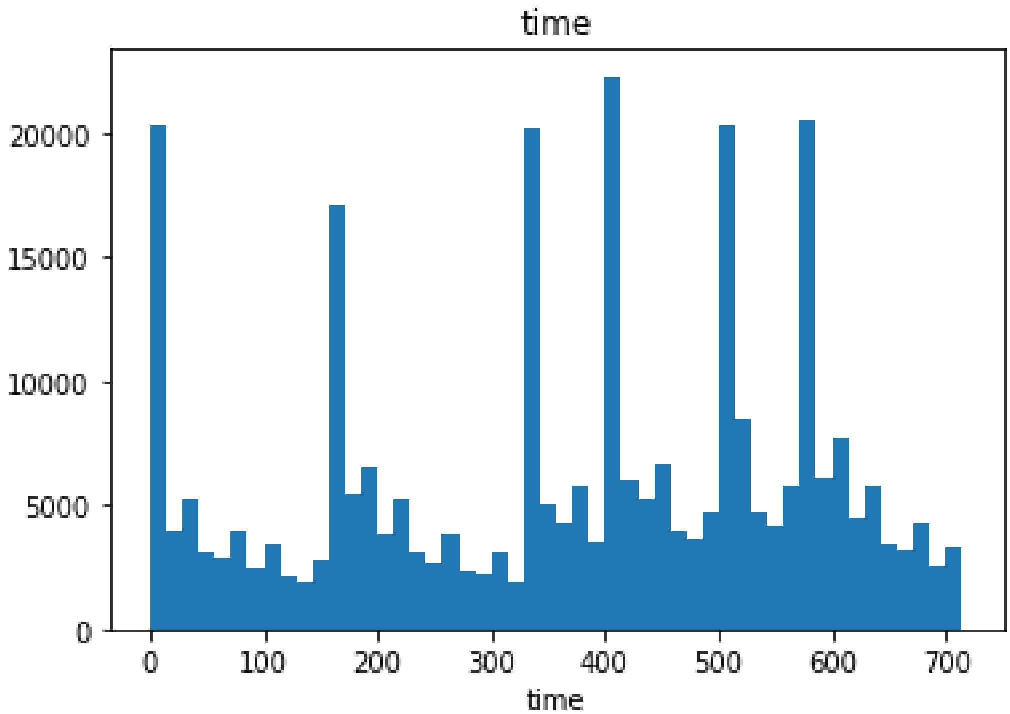
ANALYZING DATA In the following cells usage of plots has been made to understand the data as well as visualize the data and thus answer some of the questions

Q1 How is the data distributed in various columns? Q2 What is the average income for the customers? Q3 What is the age age of Starbucks customers> Q4 What are the various types of offers being laid out? Q5 What is the distribution of these offers? Q6 How were the offers received, viewed and completed? Q7 Which offer_id is connected to which offer type? Q8 For all the unique customers how much amount do they spend and who all are the loyal customers? Q9 what is the distribution of events in transcript? Q10 What is the most common promotion among various age groups? Q11 how are the people among various genders taking these offers and how effective the offers are?

Q1) How is the data distributed in various columns? The distribution of variables in colums is exploited to extract relationship amoong them.

In [39]:
```
plt.hist(trans['time'],bins=50)
plt.xlabel("time")

plt.title("time")
plt
```

Out[39]: `<module 'matplotlib.pyplot' from '/opt/conda/lib/python3.6/site-packages/matplotlib/pyp`



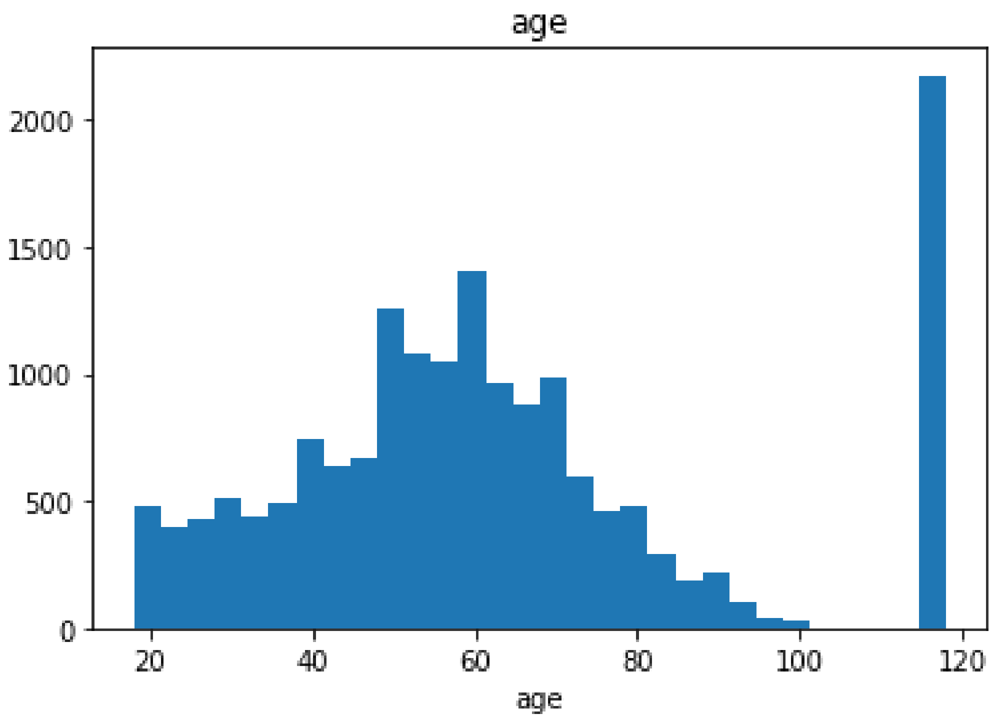
In [ ]:

In [40]:
```
profile.columns
```

Out[40]: `Index(['age', 'became_member_on', 'gender', 'id', 'income'], dtype='object')`

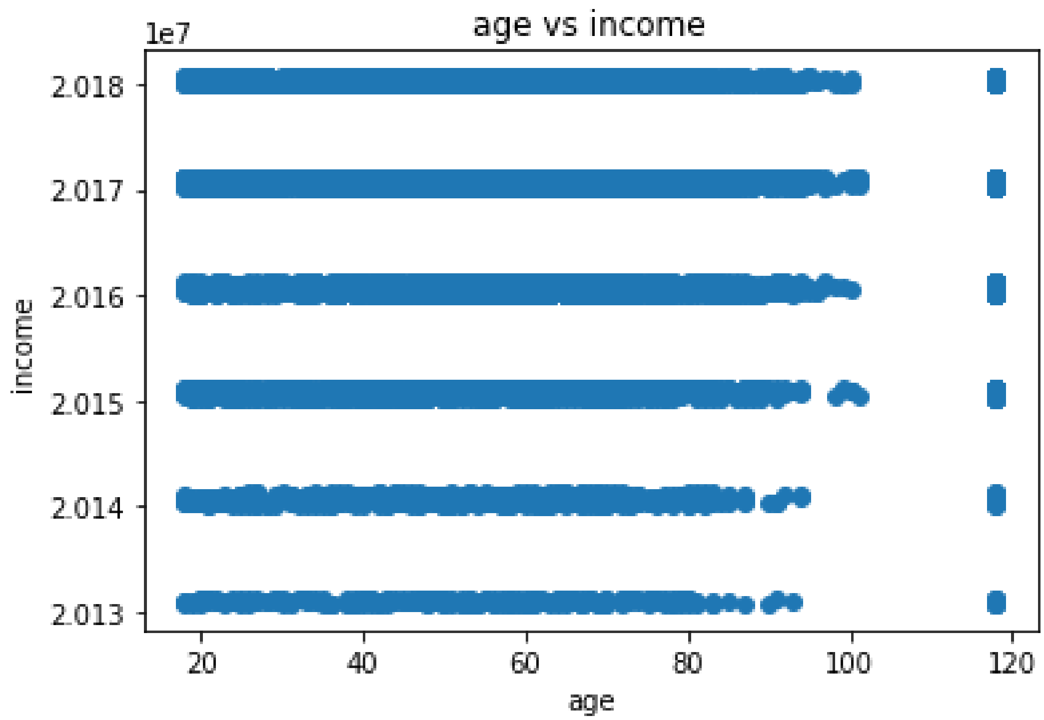
The graph shows histogram plot for age in profile dataset

```
In [41]: plt.hist(profile['age'],bins=30)
         plt.xlabel("age")

         plt.title("age")

Out[41]: Text(0.5,1,'age')
```



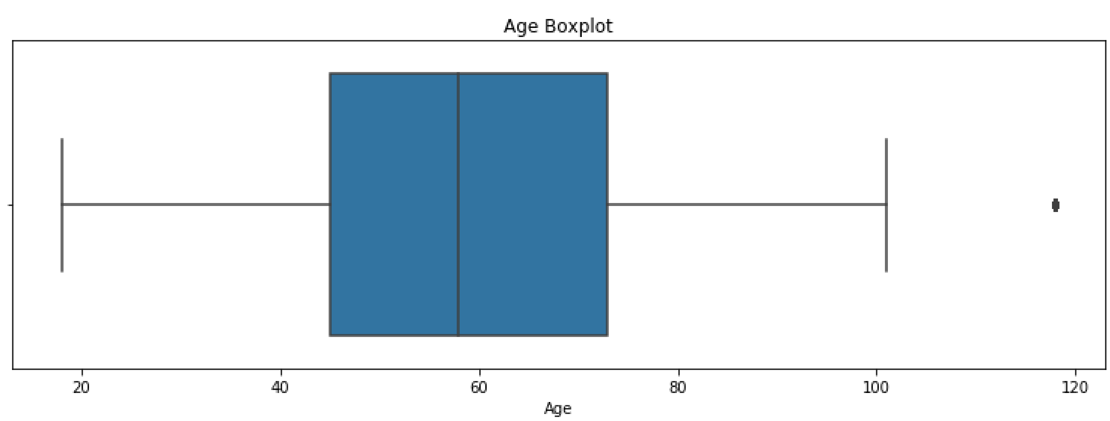
The graph shows scatter plot for income vs age in profile dataset

```
In [42]: plt.scatter(profile['age'],profile['became_member_on'])
         plt.xlabel("age")
         plt.ylabel("income")
         plt.title("age vs income")

Out[42]: Text(0.5,1,'age vs income')
```



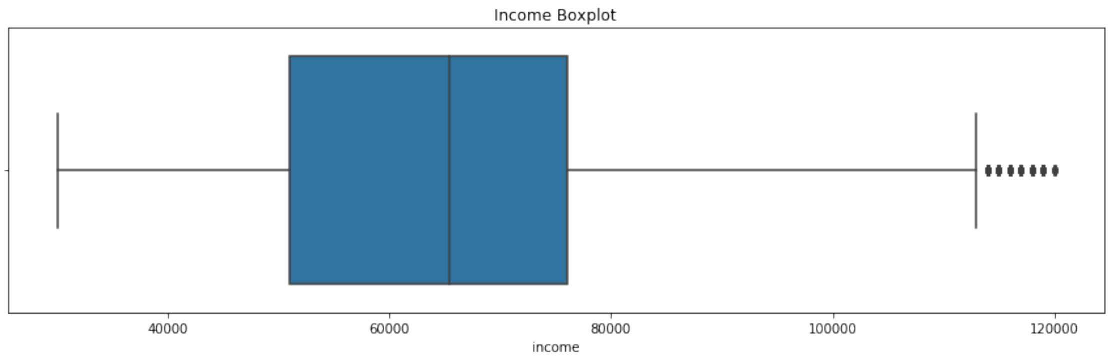
The graph shows box plot for income in pofile dataset

In [43]:
```
plt.figure(figsize=(13, 4))
sns.boxplot(profile['age'])
plt.title('Age Boxplot')
plt.xlabel('Age')
plt.xticks(rotation = 0)
plt.show();
```



Here we can see an outlier in the graph
The graph shows box plot for income in pofile dataset

```
In [44]: plt.figure(figsize=(15,4))
         sns.boxplot(profile["income"])
         plt.title('Income Boxplot')
         plt.xlabel('income')
         plt.xticks(rotation = 0)
         plt.show();
```

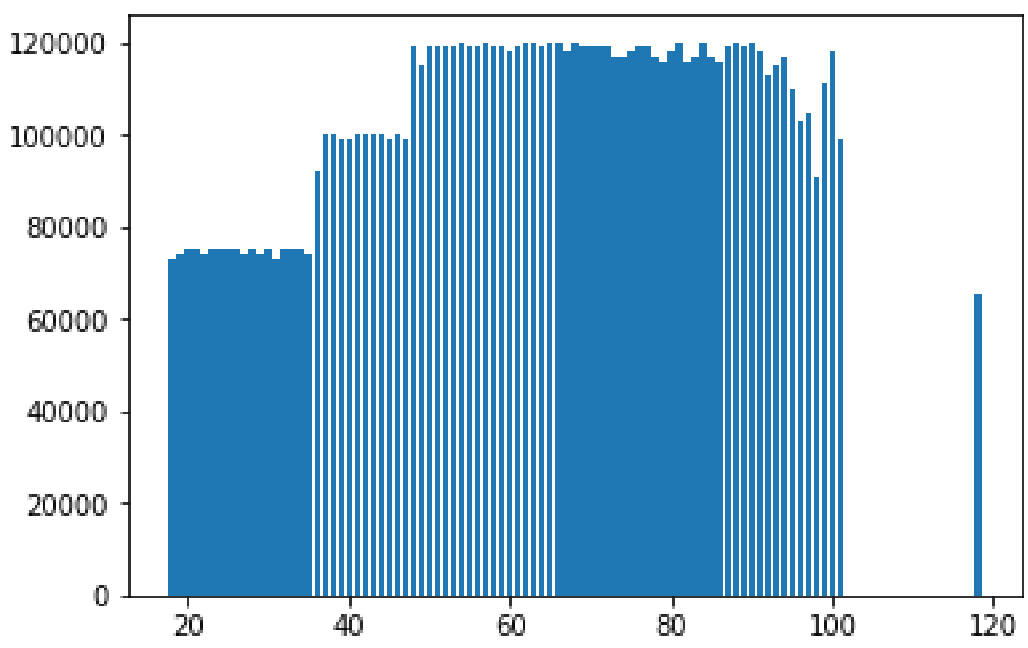


Here again we can see outlier in the graph for distribution of income

```
In [ ]:
```

The following graph shows the distribution of age and income in profile dataset

```
In [45]: plt.bar(profile['age'],profile['income'])

Out[45]: <Container object of 17000 artists>
```



```
In [46]: count_age = profile['age'].value_counts()
```

```
In [47]: count_age.shape
```

```
Out[47]: (85,)
```

This shows 85 different values in age column

```
In [48]: profile['age'].describe()
```

```
Out[48]: count    17000.000000
         mean        62.531412
         std         26.738580
         min         18.000000
         25%         45.000000
         50%         58.000000
         75%         73.000000
         max        118.000000
         Name: age, dtype: float64
```

The following distributes the age column in different age categories

```
In [49]: profile['age_by_decade'] = pd.cut(x=profile['age'], bins=[10,19, 29, 39, 49,59,69,79,89
```

```
In [50]: profile['age_by_decade']
```

```
Out[50]: 0        110s
         1         50s
         2        110s
         3         70s
         4        110s
         5         60s
         6        110s
         7        110s
         8         60s
         9        110s
         10       110s
         11       110s
         12        50s
         13        60s
         14        20s
         15        60s
         16        40s
         17       110s
         18        50s
         19        60s
```

```
20        40s
21        60s
22        70s
23       110s
24        40s
25        50s
26       110s
27        30s
28        40s
29        50s
         ...
16970     60s
16971     50s
16972     40s
16973     30s
16974     50s
16975     60s
16976     30s
16977    110s
16978     20s
16979     60s
16980    110s
16981     80s
16982    110s
16983     70s
16984     70s
16985     20s
16986     50s
16987     50s
16988     60s
16989    110s
16990     70s
16991    110s
16992     20s
16993     60s
16994    110s
16995     40s
16996     60s
16997     40s
16998     80s
16999     60s
Name: age_by_decade, Length: 17000, dtype: category
Categories (11, object): [10s < 20s < 30s < 40s ... 80s < 90s < 100s < 110s]
```

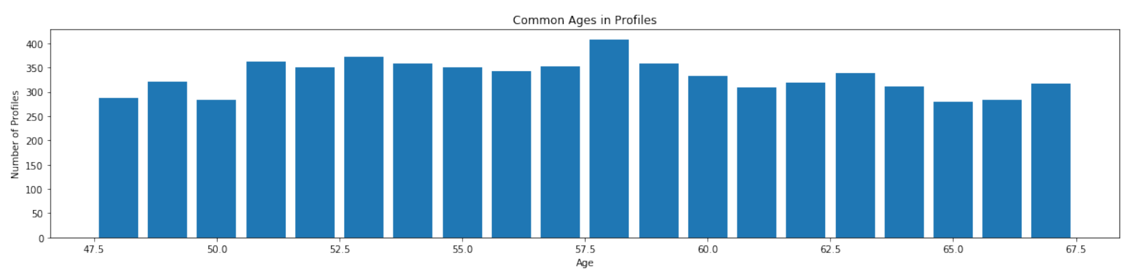
```
In [51]: max_count_ages = profile['age'].value_counts()[1:].head(20).reset_index()

In [52]: max_count_ages['index']

Out[52]: 0     58
```

```
1      53
2      51
3      54
4      59
5      57
6      52
7      55
8      56
9      63
10     60
11     49
12     62
13     67
14     64
15     61
16     48
17     50
18     66
19     65
Name: index, dtype: int64
```

Age number for the customers we are leaving the first column with age 118

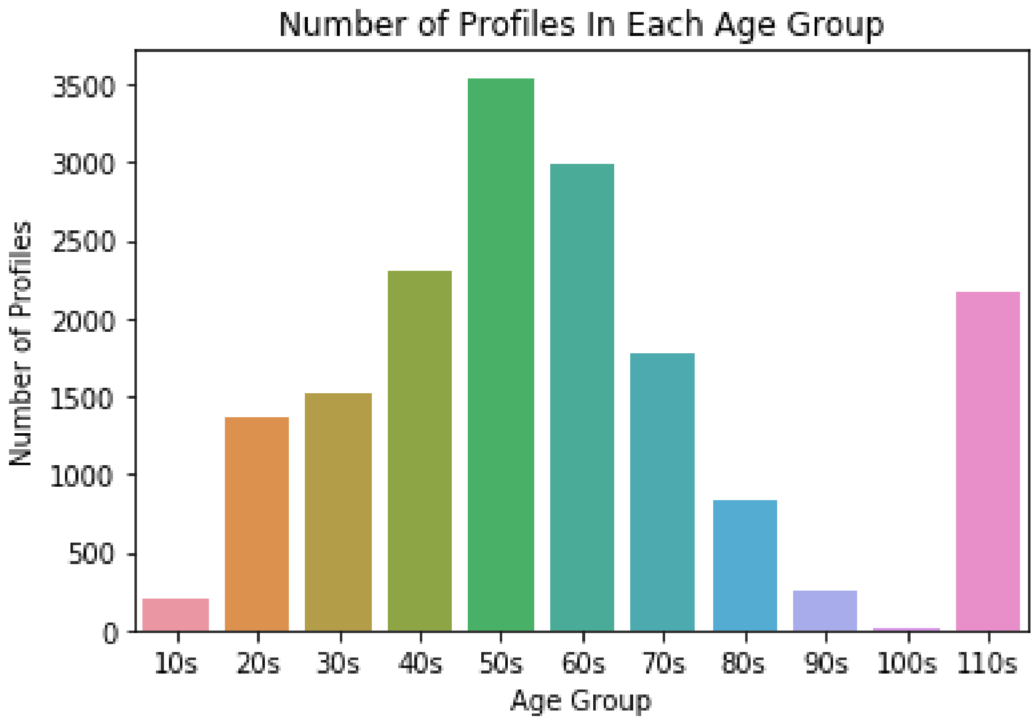
```
In [53]: plt.figure(figsize=(20,4))
         plt.bar(max_count_ages['index'],max_count_ages['age'])
         plt.title('Common Ages in Profiles')
         plt.ylabel('Number of Profiles')
         plt.xlabel('Age')
```

```
Out[53]: Text(0.5,0,'Age')
```



Here we can see that the common age group acting as out biggest customer is 50

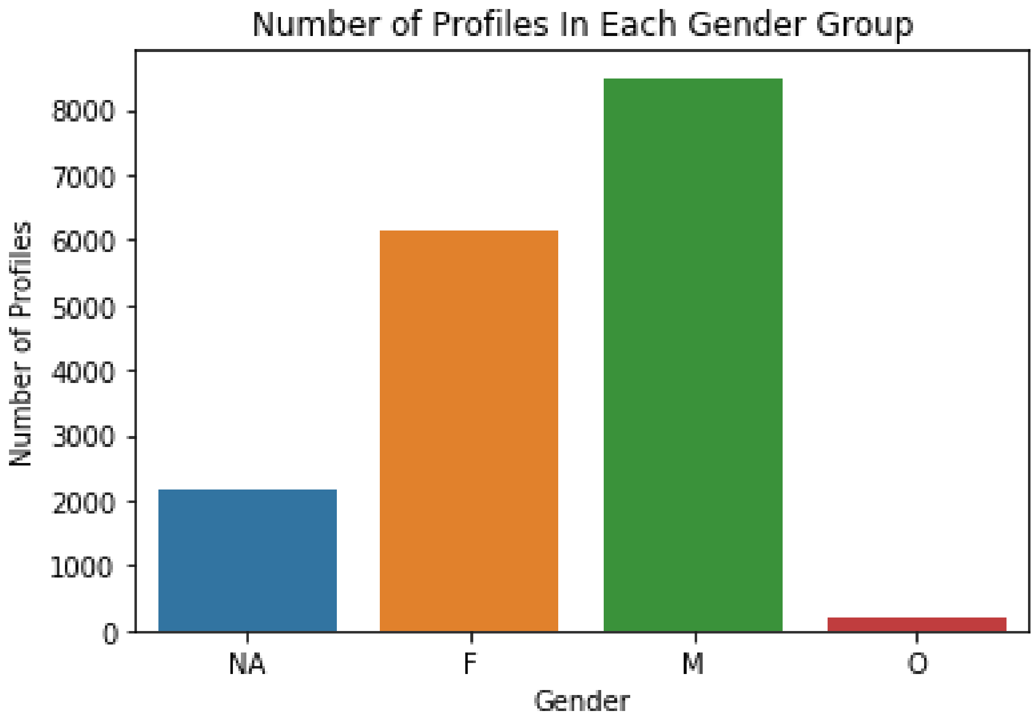
```
In [54]: sns.countplot(x=profile['age_by_decade'])
         plt.title('Number of Profiles In Each Age Group')
         plt.ylabel('Number of Profiles')
         plt.xlabel('Age Group')
         plt.show();
```



In [ ]:

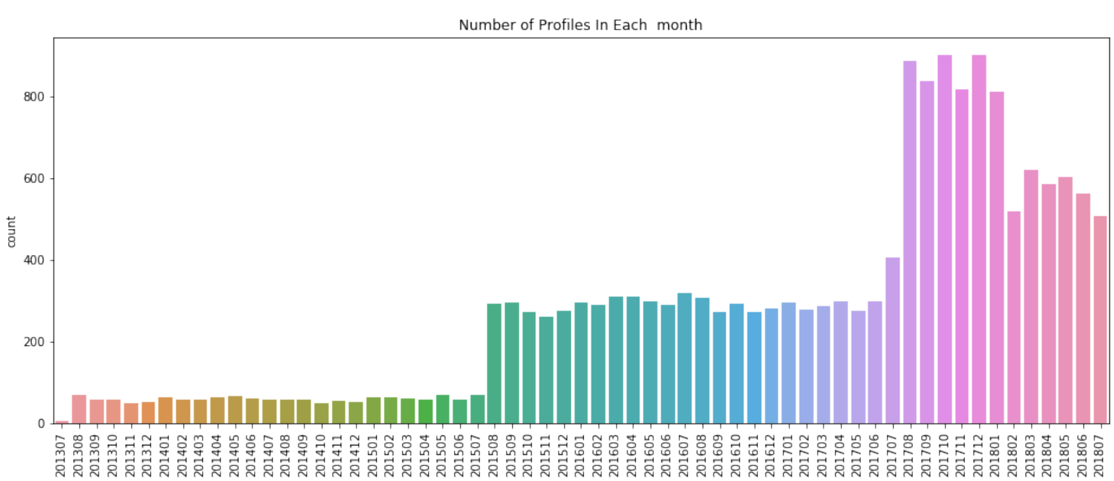
Male are buying more than male and a good portion of database contains 'NA'

```
In [55]: sns.countplot(x=profile['gender'])
         plt.title('Number of Profiles In Each Gender Group')
         plt.ylabel('Number of Profiles')
         plt.xlabel('Gender')
         plt.show();
```




```
In [56]: profile['profile_per_month']=profile['became_member_on'].astype(str).str[:6]
```

```
In [57]: plt.figure(figsize=(16, 6))
         sns.countplot(x=profile['profile_per_month'])
         plt.title('Number of Profiles In Each  month')
         plt.xlabel('month')
         plt.xticks(rotation=90)
         plt.show();
```



In [58]: transcript.head()

```
Out[58]:             event                            person  time  \
         0  offer received  78afa995795e4d85b5d9ceeca43f5fef     0
         1  offer received  a03223e636434f42ac4c3df47e8bac43     0
         2  offer received  e2127556f4f64592b11af22de27a7932     0
         3  offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
         4  offer received  68617ca6246f4fbc85e91a2a49552598     0

                                                       value  \
         0  {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
         1  {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
         2  {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
         3  {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
         4  {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}

                                     offer_id  reward  amount
         0  9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0
         1  0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0
         2  2906b810c7d4411798c6938adc9daaa5       0       0
         3  fafdcd668e3743c1bb461111dcafc2a4       0       0
         4  4d5c57ea9a6940dd891ad53e9dbe8da0       0       0
```

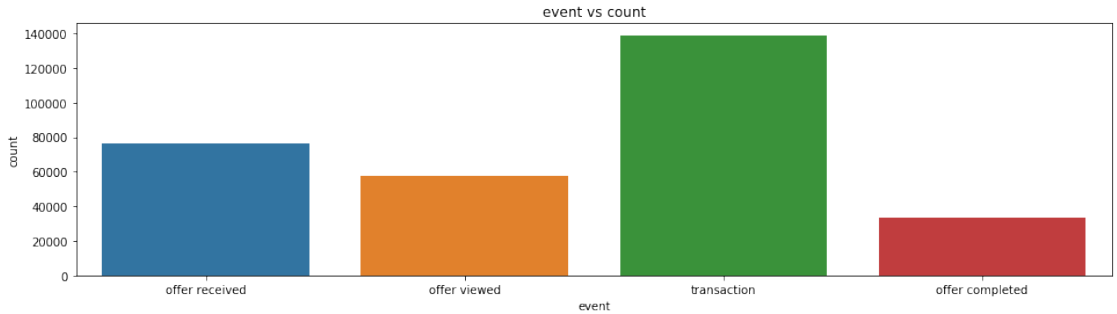
We can see an increase in count in 2017 october and 2017 december

In [59]: transcript['event'].value_counts()

```
Out[59]: transaction        138953
         offer received      76277
         offer viewed        57725
         offer completed     33579
         Name: event, dtype: int64
```

In [60]:
```
plt.figure(figsize=(16,4))
sns.countplot(x=transcript['event'])
plt.title("event vs count")
plt.xlabel("event")
plt.ylabel("count")
```

Out[60]: Text(0,0.5,'count')



Here we can see that most people gor for transaction and there is a good amount of peoplr who view and dont make use of the offer and some people make use of the offers

```
In [61]: transcript['person'].value_counts()

Out[61]: 94de646f7b6041228ca7dec82adb97d2    51
         8dbfa485249f409aa223a2130f40634a    49
         5e60c6aa3b834e44b822ea43a3efea26    48
         79d9d4f86aca4bed9290350fb43817c2    48
         d0a80415b84c4df4908b8403b19765e3    48
         bd2cdd691aca4bb0a0e039979ee5de5c    46
         28681c16026943e68f26feaccab0907f    46
         a42ed50acc4d4b25bca647c9e0b916ad    46
         b1f4ece7d49342628a9ed77aee2cde58    46
         ab25fd6fbd5040f880751921e4029757    44
         9ae56116908640fc83477982da0aaec4    43
         d087fd0166404163b7d1e1e7cf2a9ac7    43
         ca265792e65949d79b2b0e91bdd31c57    43
         86e9d338b85b4177b369fe6b0ad4fed3    43
         4142f5e23db741b1af4be0287dc91c1c    43
         0ebc3c4c39234ab6a2701fe2525705a9    42
         cd9bac9e8aea4609929a55b9b468c88e    42
         edc7b04392144da9979f3077095f268a    42
         8e7d398d4bd948e397e201ad2bd5cce8    42
         40ce078d5b2a43d19138a788754520be    42
         0d74b166a5e54b269795dbaf38c6dfae    42
         5da599f0f0ca40a6916e28487a55e655    41
         1d755c218f714559a57ee7df7e6b1ca0    41
         f0a3a3c05e3c4e2e84929a49a6b5488c    41
         2ddc7d63c32d4606a7a45e3e70439b44    41
         a6fce370a2ce4df995fc4899bfeb3b6a    41
         6cd32ababc644c6c8a2644368c795728    41
         562b9efc809b4bd897ce0381347aaea2    41
         417c8e42cba54dd0ba1ea7ee079ef87a    41
         81a263ee0b8544b6a8910ee690cc6edd    41
                                             ..
```

```
c6e579c6821c41d1a7a6a9cf936e91bb    4
0a9fe790648c471bb8ee7bd554ce3122    4
11e2d9f6ef474d13bcf1db21f5c7bb62    4
1b978feee51d4afe83ecde1ae78907cc    3
d727102ac242449ab15f1bd1af28e6ff    3
fafcd6ee168140fbbb5da43be1d3daa7    3
76341c0cc6684b3eb23661e195dfc9a3    3
19a0510b9ce24b9da44618f7161ae72d    3
50f26d1787444bb6a1e1b8be501f6877    3
83abd84070344617824 83fb32d3d5f5c    3
ce1579c557c14f7785869dc80638bc0f    3
12ede229379747bd8d74ccdc20097ca3    3
bc0c484263b94b0896f20c5e4fdf3585    3
08e7c2a166ff44e4a009a18b5e8e4b81    3
3a4874d8f0ef42b9a1b72294902afea9    3
f67a6524092d48a788a415c453bd2e00    3
af63cf0ed6ad4c458a03cb321927b463    3
2e9660f6e83b49bbb9d533b9317359c4    3
3045af4e98794a04a5542d3eac939b1f    2
1bfe13d2453c4185a6486c6817e0d568    2
22617705eec442e0b7b43e5c5f56fb17    2
cae5e211053f4121a389a7da4d631f7f    2
3a4e53046c544134bb1e7782248631d1    2
fccc9279ba56411f80ffe8ce7e0935cd    2
912b9f623b9e4b4eb99b6dc919f09a93    2
df9fc9a86ca84ef5aedde8925d5838ba    2
afd41b230f924f9ca8f5ed6249616114    2
e63e42480aae4ede9f07cac49c8c3f78    2
7ecfc592171f4844bdc05bdbb48d3847    2
da7a7c0dcfcb41a8acc7864a53cf60fb    1
Name: person, Length: 17000, dtype: int64
```

Q2 Average income for starbucks customer

```
In [62]: profile['income'].mean()
```

```
Out[62]: 65404.991568296799
```

Q3 Average age of the customers

```
In [63]: profile['age'].describe()['mean']
```

```
Out[63]: 62.531411764705879
```

Q4 Various types of offers being laid out

```
In [64]: transcript['offer_id'].value_counts()
```

```
Out[64]:                                      138953
         fafdcd668e3743c1bb461111dcafc2a4     20241
         2298d6c36e964ae4a3e7e9706d1fb8c2     20139
         f19421c1d4aa40978ebb69ca19b0e20d     19131
         4d5c57ea9a6940dd891ad53e9dbe8da0     18222
         ae264e3637204a6fb9bb56bc8210ddfd     18062
         9b98b8c7a33c4b65b9aebfe6a799e6d9     16202
         2906b810c7d4411798c6938adc9daaa5     15767
         5a8bc65990b245e5a138643cd4eb9837     14305
         0b1e1539f2cc45b7b9fa7c272da2e1d7     13751
         3f207df678b143eea3cee63160fa8bed     11761
         Name: offer_id, dtype: int64
```

```
In [65]: transcript['offer_id'].value_counts().index.tolist()
```

```
Out[65]: ['',
          'fafdcd668e3743c1bb461111dcafc2a4',
          '2298d6c36e964ae4a3e7e9706d1fb8c2',
          'f19421c1d4aa40978ebb69ca19b0e20d',
          '4d5c57ea9a6940dd891ad53e9dbe8da0',
          'ae264e3637204a6fb9bb56bc8210ddfd',
          '9b98b8c7a33c4b65b9aebfe6a799e6d9',
          '2906b810c7d4411798c6938adc9daaa5',
          '5a8bc65990b245e5a138643cd4eb9837',
          '0b1e1539f2cc45b7b9fa7c272da2e1d7',
          '3f207df678b143eea3cee63160fa8bed']
```

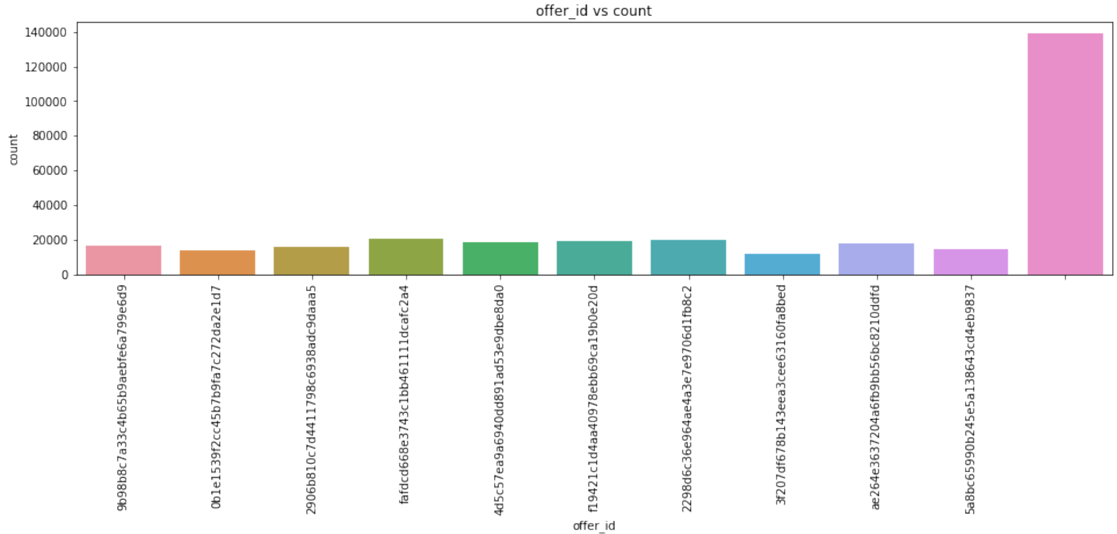
```
In [66]: transcript['offer_id'].dtype
```

```
Out[66]: dtype('O')
```

Q5 Distribution of these offers

```
In [67]: plt.figure(figsize=(16,4))
         sns.countplot(x=transcript['offer_id'])
         plt.title("offer_id vs count")
         plt.xlabel("offer_id")
         plt.ylabel("count")
         plt.xticks(rotation=90)
```

```
Out[67]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10]),
          <a list of 11 Text xticklabel objects>)
```



This shows the most columns dont have offer id

In [68]: transcript.columns

Out[68]: Index(['event', 'person', 'time', 'value', 'offer_id', 'reward', 'amount'], dtype='obje

In [69]: id_offer = transcript['offer_id'].value_counts().index.tolist()[1:]

In [70]: transcript['event'].value_counts()

Out[70]:
```
transaction        138953
offer received      76277
offer viewed        57725
offer completed     33579
Name: event, dtype: int64
```

In [71]:
```
offer_type=transcript['event'].value_counts().index.tolist()
offer_type
```

Out[71]: ['transaction', 'offer received', 'offer viewed', 'offer completed']

In [72]: transcript['event'][transcript['offer_id']=='fafdcd668e3743c1bb461111dcafc2a4'].value_c

Out[72]: 7597

In [73]: new_df=pd.DataFrame(id_offer)

In [74]: new = pd.DataFrame(columns=['transaction', 'offer_received', 'offer_viewed', 'offer_com

new_df.join(new)

In [75]: new_df=new_df.join(new)

In [76]: new_df.columns

Out[76]: Index([0, 'transaction', 'offer_received', 'offer_viewed', 'offer_completed'], dtype='o

Q6 Distribution of offers into received viewd and completed and how many times transaction was done? A new dataframe is created to view the distribution

```
In [77]: def get_count_offer_received():
             '''
             Input
             Though not explicitly taken but the input is transcript['event']
             Output
             Classificationof all the offers based on the no of times they were received by cust
             '''
             o=0
             for i in id_offer:
                 new_df.offer_received[o]=transcript['event'][transcript['offer_id']==i].value_c
                 print(transcript['event'][transcript['offer_id']==i].value_counts()['offer rece
                 o=o+1
```

```
In [78]: get_count_offer_received()
```

```
7597
7646
7571
7593
7658
7677
7632
7618
7668
7617
```

```
In [79]: def get_count_transaction():
             '''
             Input
             Though not explicitly taken but the input is transcript['event']
             Output
             Classificationof all the offers based on the no of times transaction was done
             '''
             o=0
             for i in id_offer:
                 try:
                     new_df.transaction[o]=transcript['event'][transcript['offer_id']==i].value_
                     o=o+1
                 except:
                     new_df.transaction[o]=0
                     o=o+1
```

```
In [80]: get_count_transaction()

In [ ]:

In [83]: def get_count_offer_viewed():
             '''
             Input
             Though not explicitly taken but the input is transcript['event']
             Output
             Classificationof all the offers based on the no of times they were viewed by custom
             '''
             o=0
             for i in id_offer:
                 new_df.offer_viewed[o]=transcript['event'][transcript['offer_id']==i].value_cou
                 print(transcript['event'][transcript['offer_id']==i].value_counts()['offer rece
                 o=o+1

In [87]: get_count_offer_viewed()

7597
7646
7571
7593
7658
7677
7632
7618
7668
7617

In [88]: def get_count_offer_completed():
             '''
             Input
             Though not explicitly taken but the input is transcript['event']
             Output
             Classificationof all the offers based on the no of times they were completed by cus
             '''
             o=0
             for i in id_offer:
                 try:
                     new_df.offer_completed[o]=transcript['event'][transcript['offer_id']==i].va
                     o=o+1
                 except:
                     new_df.offer_completed[o]=0
                     o=o+1

                 #print(transcript['event'][transcript['offer_id']==i].value_counts()['offer rec
```

In [89]: get_count_offer_completed()

In [90]: new_df

```
Out[90]:                                    0  transaction  offer_received  offer_viewed  \
         0  fafdcd668e3743c1bb461111dcafc2a4            0            7597          7327
         1  2298d6c36e964ae4a3e7e9706d1fb8c2            0            7646          7337
         2  f19421c1d4aa40978ebb69ca19b0e20d            0            7571          7264
         3  4d5c57ea9a6940dd891ad53e9dbe8da0            0            7593          7298
         4  ae264e3637204a6fb9bb56bc8210ddfd            0            7658          6716
         5  9b98b8c7a33c4b65b9aebfe6a799e6d9            0            7677          4171
         6  2906b810c7d4411798c6938adc9daaa5            0            7632          4118
         7  5a8bc65990b245e5a138643cd4eb9837            0            7618          6687
         8  0b1e1539f2cc45b7b9fa7c272da2e1d7            0            7668          2663
         9  3f207df678b143eea3cee63160fa8bed            0            7617          4144

            offer_completed
         0             5317
         1             5156
         2             4296
         3             3331
         4             3688
         5             4354
         6             4017
         7                0
         8             3420
         9                0
```

'The above dataframe shows the response of various of customers towards the various offer_ids

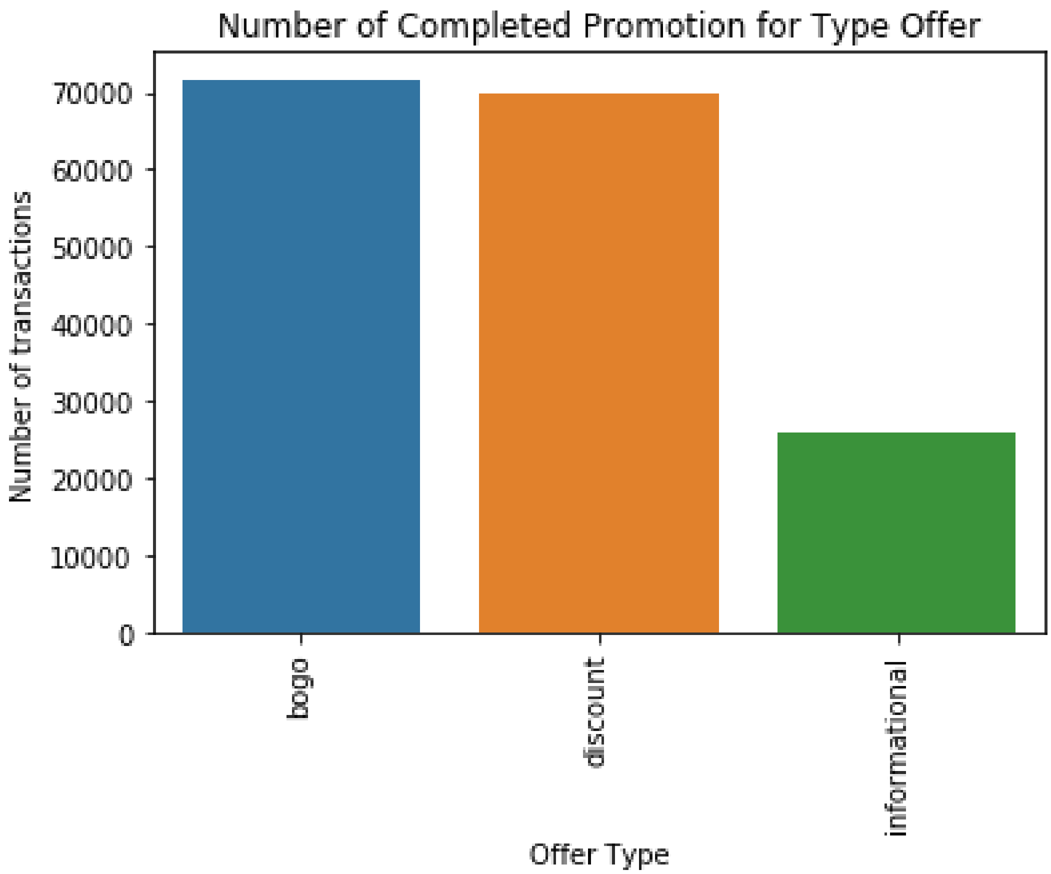
This shows that offer_id with least completion are 5a8bc65990b245e5a138643cd4eb9837 and 3f207df678b143eea3cee63160fa8bed

Q7 Classification of offer_ids into offer type

```
In [91]: def get_offer_type(offer_id):
             try:
                 offer_type = portfolio[portfolio['id'] == offer_id]['offer_type'].values[0]
                 return offer_type
             except:
                 offer_type = 'NA'
                 return offer_type

         transcript['offer_type_map'] = transcript.apply(lambda x: get_offer_type(x['offer_id'])
```

```
In [92]: sns.countplot(transcript[transcript['offer_type_map'] != 'NA']['offer_type_map'])
         plt.title('Number of Completed Promotion for Type Offer')
         plt.ylabel('Number of transactions')
         plt.xlabel('Offer Type')
```

```
plt.xticks(rotation = 90)
plt.show();
```



We can see that bogo and discount offers are offeres in a lot compared to informational offers

In [93]: `transcript`

Out[93]:

	event	person	time	\
0	offer received	78afa995795e4d85b5d9ceeca43f5fef	0	
1	offer received	a03223e636434f42ac4c3df47e8bac43	0	
2	offer received	e2127556f4f64592b11af22de27a7932	0	
3	offer received	8ec6ce2a7e7949b1bf142def7d0e0586	0	
4	offer received	68617ca6246f4fbc85e91a2a49552598	0	
5	offer received	389bc3fa690240e798340f5a15918d5c	0	
6	offer received	c4863c7985cf408faee930f111475da3	0	
7	offer received	2eeac8d8feae4a8cad5a6af0499a211d	0	
8	offer received	aa4862eba776480b8bb9c68455b8c2e1	0	
9	offer received	31dda685af34476cad5bc968bdb01c53	0	
10	offer received	744d603ef08c4f33af5a61c8c7628d1c	0	
11	offer received	3d02345581554e81b7b289ab5e288078	0	
12	offer received	4b0da7e80e5945209a1fdddfe813dbe0	0	

13	offer received	c27e0d6ab72c455a8bb66d980963de60	0
14	offer received	d53717f5400c4e84affdaeda9dd926b3	0
15	offer received	f806632c011441378d4646567f357a21	0
16	offer received	d058f73bf8674a26a95227db098147b1	0
17	offer received	65aba5c617294649aeb624da249e1ee5	0
18	offer received	ebe7ef46ea6f4963a7dd49f501b26779	0
19	offer received	1e9420836d554513ab90eba98552d0a9	0
20	offer received	868317b9be554cb18e50bc68484749a2	0
21	offer received	f082d80f0aac47a99173ba8ef8fc1909	0
22	offer received	102e9454054946fda62242d2e176fdce	0
23	offer received	4beeb3ed64dd4898b0edf2f6b67426d3	0
24	offer received	9f30b375d7bd4c62a884ffe7034e09ee	0
25	offer received	25c906289d154b66bf579693f89481c9	0
26	offer received	6e014185620b49bd98749f728747572f	0
27	offer received	02c083884c7d45b39cc68e1314fec56c	0
28	offer received	c0d210398dee4a0895b24444a5fcd1d2	0
29	offer received	8be4463721e14d7fa600686bf8c8b2ed	0
...	...	...	...
306504	transaction	8524d450673b4c24869b6c94380006de	714
306505	transaction	b895c57e8cd047a8872ce02aa54759d6	714
306506	offer completed	b895c57e8cd047a8872ce02aa54759d6	714
306507	offer viewed	8dda575c2a1d44b9ac8e8b07b93d1f8e	714
306508	transaction	8431c16f8e1d440880db371a68f82dd0	714
306509	offer completed	8431c16f8e1d440880db371a68f82dd0	714
306510	transaction	ba620885e51c4b0ea64a4f61daad494f	714
306511	transaction	a1a8f40407c444cc848468275308958a	714
306512	transaction	8d80970192fa496f99d6b45c470a4b60	714
306513	transaction	bde275066f3c4fa0bff3093e3b866a2c	714
306514	transaction	f1e4fd36e5a0446f83861308bddf6945	714
306515	transaction	0b64be3b241c4407a5c9a71781173829	714
306516	transaction	86d03d35d7e0434b935e7743e83be3a0	714
306517	transaction	3408fd05c781401f8442fb6dbaaea9c7	714
306518	transaction	1593d617fac246ef8e50dbb0ffd77f5f	714
306519	transaction	f1b31d07b5d84f69a2d5f1d07843989e	714
306520	transaction	2ce987015ec0404a97ba333e8e814090	714
306521	transaction	2e33545f0a764d27b2ccff95fc8d72c4	714
306522	transaction	d1c4500ace2e45e9a45d3cd2fccac8d8	714
306523	transaction	b65affd9e07346a1906364a396950e3d	714
306524	transaction	d613ca9c59dd42f497bdbf6178da54a7	714
306525	transaction	eec70ab28af74a22a4aeb889c0317944	714
306526	transaction	24f56b5e1849462093931b164eb803b5	714
306527	offer completed	24f56b5e1849462093931b164eb803b5	714
306528	transaction	5ca2620962114246ab218fc648eb3934	714
306529	transaction	b3a1272bc9904337b331bf348c3e8c17	714
306530	transaction	68213b08d99a4ae1b0dcb72aebd9aa35	714
306531	transaction	a00058cf10334a308c68e7631c529907	714
306532	transaction	76ddbd6576844afe811f1a3c0fbb5bec	714
306533	transaction	c02b10e8752c4d8e9b73f918558531f7	714

```
                                                     value  \
0        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
1        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
2        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
3        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
4        {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
5        {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
6        {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
7        {'offer id': '3f207df678b143eea3cee63160fa8bed'}
8        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
9        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
10       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
11       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
12       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
13       {'offer id': '3f207df678b143eea3cee63160fa8bed'}
14       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
15       {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
16       {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
17       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
18       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
19       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
20       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
21       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
22       {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
23       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
24       {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
25       {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
26       {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
27       {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
28       {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
29       {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
...                                                    ...
306504                                {'amount': 4.89}
306505                                {'amount': 4.48}
306506  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306507   {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
306508                                {'amount': 1.19}
306509  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306510                               {'amount': 14.31}
306511                                {'amount': 2.37}
306512                                {'amount': 6.92}
306513                               {'amount': 12.73}
306514                                 {'amount': 8.2}
306515                                 {'amount': 2.6}
306516                                 {'amount': 9.2}
306517                                {'amount': 11.7}
306518                               {'amount': 40.67}
```

```
306519                            {'amount': 31.13}
306520               {'amount': 1.6400000000000001}
306521                            {'amount': 17.35}
306522                             {'amount': 4.42}
306523                            {'amount': 18.35}
306524                            {'amount': 25.14}
306525                            {'amount': 43.58}
306526                            {'amount': 22.64}
306527  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306528                              {'amount': 2.2}
306529               {'amount': 1.5899999999999999}
306530                             {'amount': 9.53}
306531                             {'amount': 3.61}
306532               {'amount': 3.5300000000000002}
306533                             {'amount': 4.05}

                                offer_id  reward  amount offer_type_map
0       9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
1       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
2       2906b810c7d4411798c6938adc9daaa5       0       0       discount
3       fafdcd668e3743c1bb461111dcafc2a4       0       0       discount
4       4d5c57ea9a6940dd891ad53e9dbe8da0       0       0           bogo
5       f19421c1d4aa40978ebb69ca19b0e20d       0       0           bogo
6       2298d6c36e964ae4a3e7e9706d1fb8c2       0       0       discount
7       3f207df678b143eea3cee63160fa8bed       0       0  informational
8       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
9       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
10      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
11      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
12      ae264e3637204a6fb9bb56bc8210ddfd       0       0           bogo
13      3f207df678b143eea3cee63160fa8bed       0       0  informational
14      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
15      fafdcd668e3743c1bb461111dcafc2a4       0       0       discount
16      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
17      2906b810c7d4411798c6938adc9daaa5       0       0       discount
18      9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
19      ae264e3637204a6fb9bb56bc8210ddfd       0       0           bogo
20      2906b810c7d4411798c6938adc9daaa5       0       0       discount
21      9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
22      4d5c57ea9a6940dd891ad53e9dbe8da0       0       0           bogo
23      2906b810c7d4411798c6938adc9daaa5       0       0       discount
24      2298d6c36e964ae4a3e7e9706d1fb8c2       0       0       discount
25      2906b810c7d4411798c6938adc9daaa5       0       0       discount
26      f19421c1d4aa40978ebb69ca19b0e20d       0       0           bogo
27      ae264e3637204a6fb9bb56bc8210ddfd       0       0           bogo
28      9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
29      fafdcd668e3743c1bb461111dcafc2a4       0       0       discount
...                                  ...     ...     ...            ...
```

```
306504                                      0       4          NA
306505                                      0       4          NA
306506  fafdcd668e3743c1bb461111dcafc2a4    2       0    discount
306507  0b1e1539f2cc45b7b9fa7c272da2e1d7    0       0    discount
306508                                      0       1          NA
306509  fafdcd668e3743c1bb461111dcafc2a4    2       0    discount
306510                                      0      14          NA
306511                                      0       2          NA
306512                                      0       6          NA
306513                                      0      12          NA
306514                                      0       8          NA
306515                                      0       2          NA
306516                                      0       9          NA
306517                                      0      11          NA
306518                                      0      40          NA
306519                                      0      31          NA
306520                                      0       1          NA
306521                                      0      17          NA
306522                                      0       4          NA
306523                                      0      18          NA
306524                                      0      25          NA
306525                                      0      43          NA
306526                                      0      22          NA
306527  fafdcd668e3743c1bb461111dcafc2a4    2       0    discount
306528                                      0       2          NA
306529                                      0       1          NA
306530                                      0       9          NA
306531                                      0       3          NA
306532                                      0       3          NA
306533                                      0       4          NA

[306534 rows x 8 columns]
```

In [94]:
```
transcript['person'][transcript['event']=='offer completed'].value_counts()
```

Out[94]:
```
e22206bf65234c0e9f273bef859c02a1    6
a20c4cf47deb418f8e995c9015f15fda    6
8c410d84af08408fb41f953c93ffac27    6
9c569de8e9f747bcb5c8f78c0a014fc5    6
516e91869c8546d8a885401e850ce66c    6
f1d65ae63f174b8f80fa063adcaa63b7    6
cc328665db7d4f77babe3cb7c87e2e2f    6
c94c15ddd30845faaf072d20b111aa1e    6
94ce14b4e3774ecd908575ba7ab32f85    6
f31e27cdcdac4b0a9a384c0567a7ae43    6
9196c2bfb739494f902a13bab46199d2    6
1e5d6388b5214fefb090a9a2a4d21983    6
41eb66dfb4824a029f7abb81c140c267    6
```

7055b5fa5f8647618aa14e220b7c6b5c	6
07e29d02d0074da28b430e1dc40cc1fc	6
06d7f5abc31b4a02836349333ac02f33	6
aa445cee24ba41389eadabcf2fc95bac	6
d80f9e3f974448ec902c42818097ebf3	6
0494aa6671414fab9837fa3cd45e72bc	6
759c2269abc44dfdb8e6d140d4060856	6
adc7ef85f1714699b4f341a7e12c632f	6
a0db7a1fb55e402cb7f5b8ddc10b0818	6
82980813d7094cf5878dd906a6734307	6
bdfba75f0e0c4533b4b6b97fa37daa4f	6
325da16cf16c4d72972f28281f752201	6
acdf9a19b60d4c13ad594ebcb6b23eb9	6
cb45be0c399a4b4fa13dda5c89c442f4	6
4357cb4e15f04f988045c5a7e37c5775	6
76018bc0dfd64ccc8f2e7959bf627ba7	6
5c2a03a9d03145cc9bf953faa2c51fc1	6
	..
9c692f9c82634eae9b881367b008c242	1
f7977eccb81b49b9ac6d2bd84cd64bb0	1
ee6d71acc74a4b1183ef4a2822e652e5	1
99b18b650c2745489bf81a306852a7c5	1
0e18338db3524564870a52ee5cfe68a4	1
332aa1f51a7643199d9d60f885230511	1
3b65578368194a428d38390dd27cd447	1
32fedc747de04e508c6a538f78b130f0	1
62dd66d3bb1c4153b6126caa0a9f8235	1
c06029a8eea747d69a13b52d58f4ccbb	1
cca43d4a61554c78abb7b62c3dc9927c	1
af046d91eec74ac79e40b6c12004ecab	1
61e81e0f321d49e7a62c44a2595c3bab	1
65daaf0f413744789903a355d6370b9d	1
5139d8d43e674e5d9dcd6773d9684462	1
e6590c7f954f4ff482c069b45070b351	1
7bf785bb8b6c4a5bb14815ca85ceb87b	1
85a4b98da3254e0986fc5a1ee9d5c1cc	1
3ecd6653a2d145a4b2bc3b35e8aece7a	1
a305233f5c944df5bcaacc029658e085	1
2814e3e30f2749a1bab4c14ee51de542	1
8aa2f57d85934578b6a5da5682500d08	1
fd64a3758d0b49549df03a5c25ba89c9	1
7651ccd65fa7484d972f75f2f2307010	1
bcfc432fc0be4938a9a688c5c2873a49	1
1dbb8d4481aa4b0c8c9b8912fb340c76	1
edfe76e1510841d79f0ecd0472e4680d	1
e6d3335fdf2c4cd2ac8edcd2bb06b289	1
58ad7189fe54463fa221873e553f3173	1
b1a6c6b97e114a75adbd008e9222397e	1

```
Name: person, Length: 12774, dtype: int64
```

In [95]:
```
heavy_buying_customer =transcript[(transcript['event']=='offer_completed') | (transcrip
```

In [96]:
```
unique_customer=transcript[(transcript['event']=='offer_completed') | (transcript['even
```

In [97]:
```
customer_df=pd.DataFrame(unique_customer)
```

In [98]:
```
amount=pd.DataFrame(columns=['amount'])
```

In [99]:
```
customer_df=customer_df.join([amount])
customer_df
```

Out[99]:

	0	amount
0	02c083884c7d45b39cc68e1314fec56c	NaN
1	9fa9ae8f57894cc9a3b8a9bbe0fc1b2f	NaN
2	54890f68699049c2a04d415abc25e717	NaN
3	b2f1cd155b864803ad8334cdf13c4bd2	NaN
4	fe97aa22dd3e48c8b143116a8403dd52	NaN
5	629fc02d56414d91bca360decdfa9288	NaN
6	bbeb54e861614fc7b22a8844f72dca6c	NaN
7	a97e6f33219c432db82acfa0d19c602d	NaN
8	676506bad68e4161b9bbaffeb039626b	NaN
9	8f7dd3b2afe14c078eb4f6e6fe4ba97d	NaN
10	4cbe33c601a5407f8202086565c55111	NaN
11	b432b74402bb4981a4651c8df1670365	NaN
12	a04fcfd571034456aaa6d56c0a3fd9b6	NaN
13	227f2d69e46a4899b70d48182822cff6	NaN
14	bb0f25e23a4c4de6a645527c275cd594	NaN
15	c2c72ce6038644c797208046d1e3498a	NaN
16	7ca349e55ff544c7a13addfdea2e2c06	NaN
17	d72d201be5794279aa716d8ad82b8d90	NaN
18	ad80753fc9e0485c9e6b1cc9478d827f	NaN
19	73ffefd41e9a4ca3ab26b2b3697c6eb7	NaN
20	3e621194f72e40d7a0b695ee9b7c38b7	NaN
21	1f5c961416e64c5d88098b02b1bdf246	NaN
22	3bcc51fdde354eb1949c813dbc905182	NaN
23	ed46fca6de7042478b411690878dc069	NaN
24	b860d355ef6e4c66b5d5a837c56ef32d	NaN
25	4ad3748475204cf99571183f05b5e2f7	NaN
26	99297ea01107436fa8c2e2bc86f55d89	NaN
27	24115a61df25473e84a8a03f3c98de1a	NaN
28	7195944f0cc34115b0a5e7b4a62055f2	NaN
29	afce4cf8194f4e90a3e92da941a23601	NaN
...	...	...
16548	99ffa82129904cd6adabcdfe02072904	NaN
16549	d448fa911f3f4b4aafe87396d220fad6	NaN
16550	cfefa4c2c2aa40a196f392e43eb5eeed	NaN
16551	6b214a99748949608e302160c8845f7f	NaN

```
16552  47e28072fc414f54bf1aa27be3a19d32    NaN
16553  25015650d5fb4a8b96e6d289e76cc892    NaN
16554  54d60cf626294a1685867c5f31dbf7b8    NaN
16555  3be15bacb2064198874cbf898ca5cf82    NaN
16556  2dbb5a86750d4f589ef19c4bf0dd1b3a    NaN
16557  8578196a074a4f328976e334fa9383a3    NaN
16558  4d3e33d680e04a57a565c847948bf1ca    NaN
16559  db1805ed333844978e8b46ed3e4643ae    NaN
16560  5ebe9f31969a453da31dd70fedd6c1eb    NaN
16561  39145fee7c1240469a46bf939fb8370c    NaN
16562  2ad98e767e5c472cbdf9e44a87366ef7    NaN
16563  2a6bb506b41540378b1d373985502233    NaN
16564  00ee2ca6421c4af0aeca60a1b3e00f6c    NaN
16565  695d4738abe7406da6966df59a1cd4fc    NaN
16566  77961f4298bd4340a3988bde6754d19d    NaN
16567  dc9d14e5d3be406b99e0d7e9a859bec0    NaN
16568  f6581b347e914638a6e04e3a5285a31d    NaN
16569  fef4f14201c44ddd915f504b7d88bf08    NaN
16570  fe285265878948309a32b47b216d2a65    NaN
16571  aabc3d56a2ea4446bef4c3f9eec8ab72    NaN
16572  999338157fd54f639b089cef0038e06a    NaN
16573  542c41f5afc049e7ae7d4721ace9d286    NaN
16574  448dabde725040978b8a247a20bac126    NaN
16575  7718656997f3453db0f5aeca9cd35240    NaN
16576  54463e5d95124b7fb3133fc1eae71952    NaN
16577  0c027f5f34dd4b9eba0a25785c611273    NaN

[16578 rows x 2 columns]
```

Q8 For all the unique customers how much amount do they spend and who all are theloyal customers

```
In [100]: op=0 # for all unique customers the amount they spend for various offers is added
          for customer in heavy_buying_customer:
              customer_df.amount[op]=customer[1]['amount'].sum()
              op=op+1
```

```
In [101]: customer_df=customer_df.sort_values(['amount'],ascending=False)
```

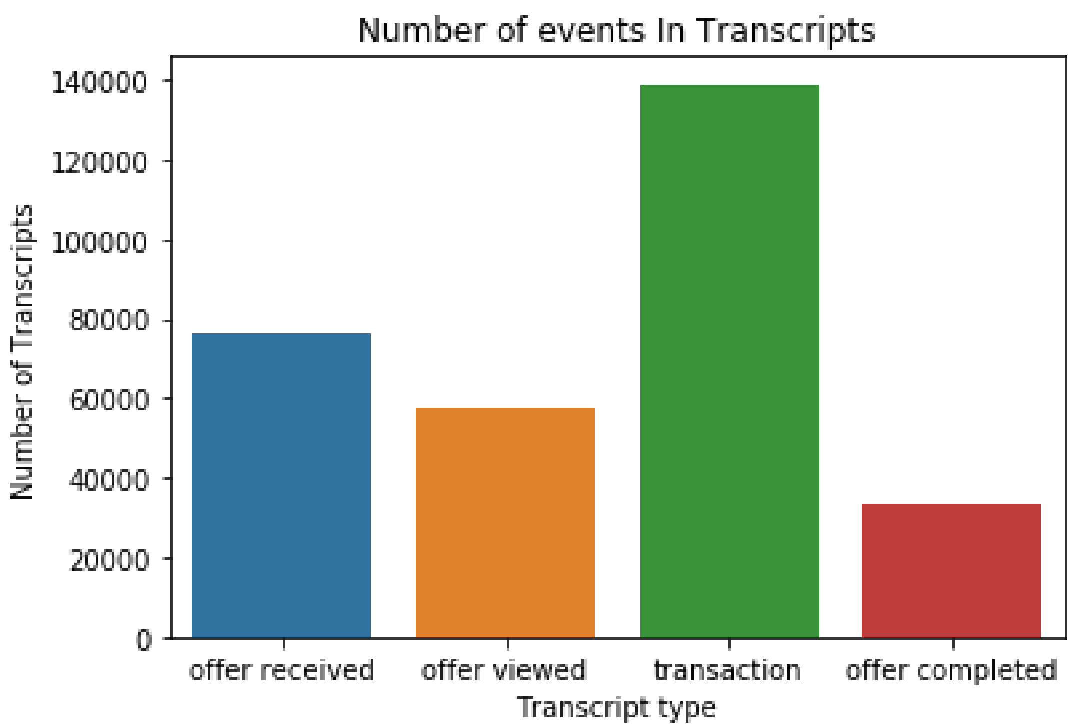
```
In [102]: customer_df.head(10)
```

```
Out[102]:                                       0  amount
          3929   e2f8376c32084327b1991706b124c8ed   1606
          15693  ed127151cf1c4c70bcb2c9a50ce749c8   1360
          11422  431a9ec074b44e80886ba3dfed6dfe2f   1320
          7492   8571868385524806bcacd7e73c1ae5e1   1314
          6366   6118ca1b782a4df0a852406b194219b2   1314
          5358   408574b846484c9f97f458b678a01a0e   1285
          11334  9407a3a0c9114b15b05f7704d0b40ef0   1256
```

```
13672   c2df0a0925684591975861d7622ac296    1244
9673    7784652beb5d4a26955c561b63f3099b    1224
6484    ae5f3528d8324fcbbdbc51cf8a46cffa    1206
```

These are the top spenders
Q9 distribution of events in transcript

```
In [103]: sns.countplot(transcript['event'])
          plt.title('Number of events In Transcripts')
          plt.ylabel('Number of Transcripts')
          plt.xlabel('Transcript type')
          plt.xticks(rotation = 0)
          plt.show();
```




```
In [ ]:
```

Q10 What is the most common promotion among various age groups?

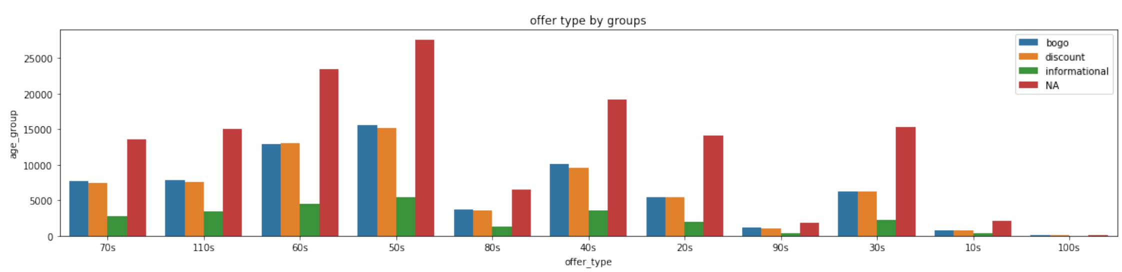
```
In [104]: #the person id from profile dataset is mapped to transcript dataset
          def age_offer(id):
              age_group = profile[profile['id']==id]['age_by_decade'].values[0]
              return age_group

In [105]: transcript['age_offer']=transcript.apply(lambda x:age_offer(x['person']),axis=1)
```

```
In [106]: transcript.head()
```

```
Out[106]:             event                            person  time  \
          0  offer received  78afa995795e4d85b5d9ceeca43f5fef     0
          1  offer received  a03223e636434f42ac4c3df47e8bac43     0
          2  offer received  e2127556f4f64592b11af22de27a7932     0
          3  offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
          4  offer received  68617ca6246f4fbc85e91a2a49552598     0

                                                       value  \
          0  {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
          1  {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
          2  {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
          3  {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
          4  {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}

                                     offer_id  reward  amount offer_type_map age_offer
          0  9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo       70s
          1  0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount      110s
          2  2906b810c7d4411798c6938adc9daaa5       0       0       discount       60s
          3  fafdcd668e3743c1bb461111dcafc2a4       0       0       discount      110s
          4  4d5c57ea9a6940dd891ad53e9dbe8da0       0       0           bogo      110s
```

```
In [107]: plt.figure(figsize=(20,4))
          sns.countplot(x="age_offer",hue="offer_type_map",data=transcript)
          plt.xlabel("offer_type")
          plt.ylabel("age_group")
          plt.title("offer type by groups")
          plt.legend()
```

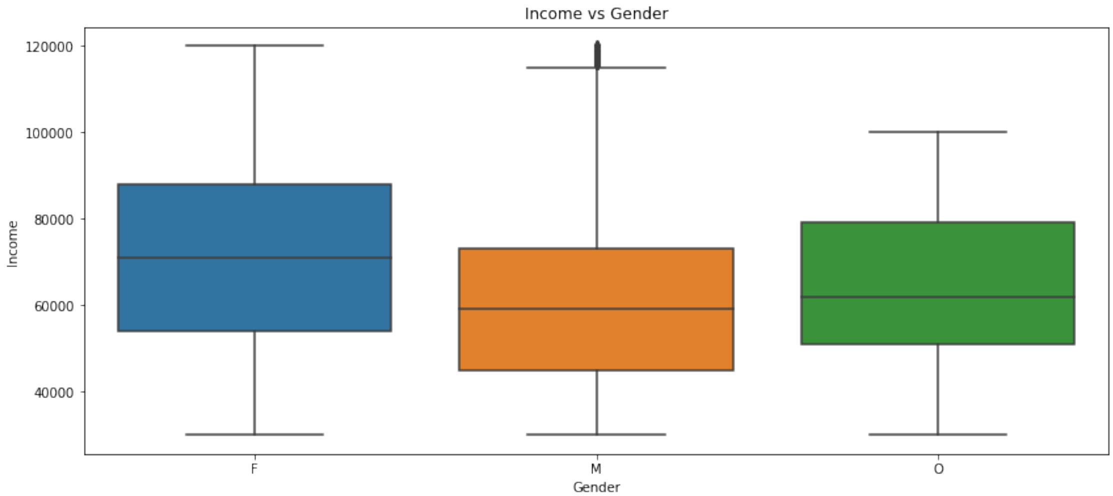
```
Out[107]: <matplotlib.legend.Legend at 0x7fc8c3f1d208>
```



The graph describes the distribution of offers among various age groups and also the unavailability of data in this respect

```
In [108]: plt.figure(figsize=(14, 6))
          sns.boxplot(x=profile[profile['gender'] != 'NA']['gender'], y=profile['income'])
          plt.title('Income vs Gender')
```

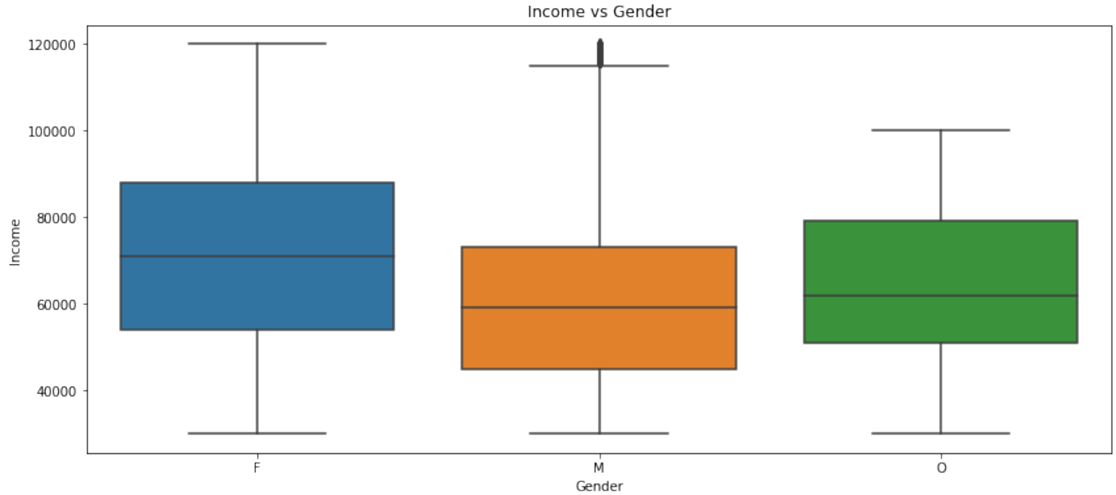
```
plt.ylabel('Income')
plt.xlabel('Gender')
plt.xticks(rotation = 0)
plt.show();
```



The above graph represents the income distribution among genders

```
In [109]: def gender_offer(id):
              gender_of=profile[profile['id']==id]['gender'].values[0]
              return gender_of
          transcript['map_gender_offer']=transcript.apply(lambda x:gender_offer(x['person']),axi
```

```
In [110]: plt.figure(figsize=(14, 6))
          sns.boxplot(x=profile[profile['gender'] != 'NA']['gender'], y=profile['income'])
          plt.title('Income vs Gender')
          plt.ylabel('Income')
          plt.xlabel('Gender')
          plt.xticks(rotation = 0)
          plt.show();
```



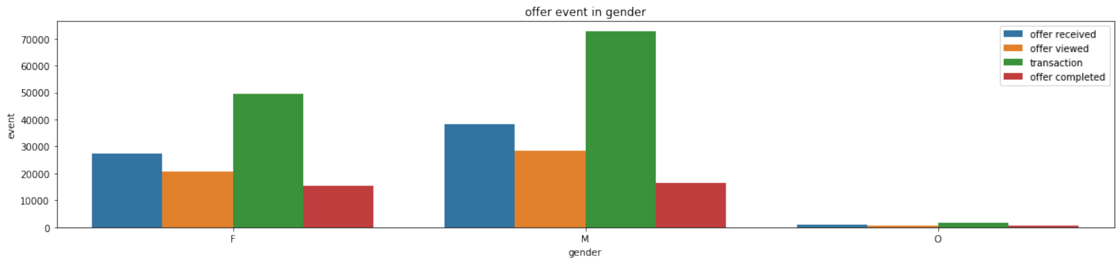

```
In [111]: plt.figure(figsize=(20,4))
          sns.countplot(x="map_gender_offer",hue="offer_type_map",data=transcript)
          plt.xlabel("gender")
          plt.ylabel("total")
          plt.title("offers in gender")
          plt.legend()
```

```
Out[111]: <matplotlib.legend.Legend at 0x7fc8c3e49a90>
```

The graph shows the distribution of offers among genders

```
In [112]: plt.figure(figsize=(20,4))
          sns.countplot(x=transcript[transcript["map_gender_offer"]!='NA']["map_gender_offer"],h
          plt.xlabel("gender")
          plt.ylabel("event")
          plt.title("offer event in gender")
          plt.legend()
```

```
Out[112]: <matplotlib.legend.Legend at 0x7fc8c3d8cc88>
```




```
In [113]: transcript.columns
```

```
Out[113]: Index(['event', 'person', 'time', 'value', 'offer_id', 'reward', 'amount',
                 'offer_type_map', 'age_offer', 'map_gender_offer'],
                dtype='object')
```

From the above two graphs it is conclusive that males received more offers compared to woman and hence offer completion is also more for them

```
In [114]: offer_gender_type = transcript[transcript["map_gender_offer"] != 'NA'].groupby(['map_g
          offer_gender_event = transcript[transcript["map_gender_offer"] != 'NA'].groupby(['map_
```

```
In [115]: offer_gender_type,offer_gender_event
```

```
Out[115]: (                                  event  person   time  value  offer_id  \
           map_gender_offer offer_type_map
           F                NA              49382   49382  49382  49382     49382
                            bogo            27619   27619  27619  27619     27619
                            discount        26652   26652  26652  26652     26652
                            informational    9448    9448   9448   9448      9448
           M                NA              72794   72794  72794  72794     72794
                            bogo            35301   35301  35301  35301     35301
                            discount        34739   34739  34739  34739     34739
                            informational   12856   12856  12856  12856     12856
           O                NA               1781    1781   1781   1781      1781
                            bogo              914     914    914    914       914
                            discount          920     920    920    920       920
                            informational     356     356    356    356       356

                                            reward  amount  age_offer
           map_gender_offer offer_type_map
           F                NA               49382   49382      49382
                            bogo             27619   27619      27619
                            discount         26652   26652      26652
                            informational     9448    9448       9448
           M                NA               72794   72794      72794
                            bogo             35301   35301      35301
```

```
                  discount         34739   34739      34739
                  informational    12856   12856      12856
O                 NA                1781    1781       1781
                  bogo               914     914        914
                  discount           920     920        920
                  informational      356     356        356  ,
                                 person   time  value  offer_id  reward  \
map_gender_offer event
F                offer completed   15477  15477  15477     15477   15477
                 offer received    27456  27456  27456     27456   27456
                 offer viewed      20786  20786  20786     20786   20786
                 transaction       49382  49382  49382     49382   49382
M                offer completed   16466  16466  16466     16466   16466
                 offer received    38129  38129  38129     38129   38129
                 offer viewed      28301  28301  28301     28301   28301
                 transaction       72794  72794  72794     72794   72794
O                offer completed     501    501    501       501     501
                 offer received      916    916    916       916     916
                 offer viewed        773    773    773       773     773
                 transaction        1781   1781   1781      1781    1781

                                 amount  offer_type_map  age_offer
map_gender_offer event
F                offer completed   15477           15477      15477
                 offer received    27456           27456      27456
                 offer viewed      20786           20786      20786
                 transaction       49382           49382      49382
M                offer completed   16466           16466      16466
                 offer received    38129           38129      38129
                 offer viewed      28301           28301      28301
                 transaction       72794           72794      72794
O                offer completed     501             501        501
                 offer received      916             916        916
                 offer viewed        773             773        773
                 transaction        1781            1781       1781  )
```

Q11 how are the people among various genders taking these offers and how effective the offers are? the no of females completing the offer

```
In [116]: transcript[(transcript["map_gender_offer"]=="F") & (transcript["event"]=="offer comple
```

```
Out[116]: 15477
```

the no of males completing the offer

```
In [117]: transcript[(transcript["map_gender_offer"]=="M") & (transcript["event"]=="offer comple
```

```
Out[117]: 16466
```

the no of females receiving the offer

```
In [118]: transcript[(transcript["map_gender_offer"]=="F") & (transcript["event"]=="offer receiv
```

```
Out[118]: 27456
```

the no of males receiving the offer

```
In [119]: transcript[(transcript["map_gender_offer"]=="M") & (transcript["event"]=="offer receiv
```

```
Out[119]: 38129
```

the no of females viewing the offer

```
In [120]: transcript[(transcript["map_gender_offer"]=="F") & (transcript["event"]=="offer viewed
```

```
Out[120]: 20786
```

the no of males viewing the offer

```
In [121]: transcript[(transcript["map_gender_offer"]=="M") & (transcript["event"]=="offer viewed
```

```
Out[121]: 28301
```

the no of females completing the transaction

```
In [122]: transcript[(transcript["map_gender_offer"]=="F") & (transcript["event"]=="transaction"
```

```
Out[122]: 49382
```

the no of males completing the transaction

```
In [123]: transcript[(transcript["map_gender_offer"]=="M") & (transcript["event"]=="transaction"
```

```
Out[123]: 72794
```

Hence no of females who completed the offer is 15477 and no of males who completed the offer is 16466 and total number of females receiving the offer is 27456 and that of males is 38129 WITH 20786 females viewing the offer and 28301 males viewing the offer. Hence in proportion 56.37 % woman completed the offer and 43.6% woman only viewed the offer and in males 43.18% completed the offer and 56.81% viewed the offer with no of males receiving the offer more than no. of females. For transaction no of females doing transaction is 49382 and that of males is 72794

So we come to a conclusion that females are more in number for completing offers and male for transaction

adding income column into transcript data

```
In [124]: #taking people_id from profile dataset and mapping it to transcript to get their incom
          def income_pro(id):
              inc=profile[profile['id']==id]['income'].values[0]
              return inc
          transcript['income']=transcript.apply(lambda x:income_pro(x['person']),axis=1)
```

converting all the datsets into csv file and saving

```
In [125]: transcript.to_csv("new_transcript.csv")
          portfolio.to_csv("new_portfolio.csv")
          profile.to_csv("new_profile.csv")
```

checking if there are any null values in the columns

```
In [126]: transcript.isna().sum()

Out[126]: event              0
          person             0
          time               0
          value              0
          offer_id           0
          reward             0
          amount             0
          offer_type_map     0
          age_offer          0
          map_gender_offer   0
          income             0
          dtype: int64

In [127]: portfolio.isna().sum()

Out[127]: difficulty      0
          duration        0
          id              0
          offer_type      0
          reward          0
          bogo            0
          discount        0
          informational   0
          email           0
          mobile          0
          social          0
          web             0
          dtype: int64
```

MODELLING Let the variable to be predicted be event in transcript

Following cells involve removal of some information and convertinf of some columns into dummies

```
In [128]: transcript

Out[128]:                  event                            person  time  \
          0       offer received  78afa995795e4d85b5d9ceeca43f5fef     0
          1       offer received  a03223e636434f42ac4c3df47e8bac43     0
          2       offer received  e2127556f4f64592b11af22de27a7932     0
```

3	offer received	8ec6ce2a7e7949b1bf142def7d0e0586	0
4	offer received	68617ca6246f4fbc85e91a2a49552598	0
5	offer received	389bc3fa690240e798340f5a15918d5c	0
6	offer received	c4863c7985cf408faee930f111475da3	0
7	offer received	2eeac8d8feae4a8cad5a6af0499a211d	0
8	offer received	aa4862eba776480b8bb9c68455b8c2e1	0
9	offer received	31dda685af34476cad5bc968bdb01c53	0
10	offer received	744d603ef08c4f33af5a61c8c7628d1c	0
11	offer received	3d02345581554e81b7b289ab5e288078	0
12	offer received	4b0da7e80e5945209a1fdddfe813dbe0	0
13	offer received	c27e0d6ab72c455a8bb66d980963de60	0
14	offer received	d53717f5400c4e84affdaeda9dd926b3	0
15	offer received	f806632c011441378d4646567f357a21	0
16	offer received	d058f73bf8674a26a95227db098147b1	0
17	offer received	65aba5c617294649aeb624da249e1ee5	0
18	offer received	ebe7ef46ea6f4963a7dd49f501b26779	0
19	offer received	1e9420836d554513ab90eba98552d0a9	0
20	offer received	868317b9be554cb18e50bc68484749a2	0
21	offer received	f082d80f0aac47a99173ba8ef8fc1909	0
22	offer received	102e9454054946fda62242d2e176fdce	0
23	offer received	4beeb3ed64dd4898b0edf2f6b67426d3	0
24	offer received	9f30b375d7bd4c62a884ffe7034e09ee	0
25	offer received	25c906289d154b66bf579693f89481c9	0
26	offer received	6e014185620b49bd98749f728747572f	0
27	offer received	02c083884c7d45b39cc68e1314fec56c	0
28	offer received	c0d210398dee4a0895b24444a5fcd1d2	0
29	offer received	8be4463721e14d7fa600686bf8c8b2ed	0
...	...	...	...
306504	transaction	8524d450673b4c24869b6c94380006de	714
306505	transaction	b895c57e8cd047a8872ce02aa54759d6	714
306506	offer completed	b895c57e8cd047a8872ce02aa54759d6	714
306507	offer viewed	8dda575c2a1d44b9ac8e8b07b93d1f8e	714
306508	transaction	8431c16f8e1d440880db371a68f82dd0	714
306509	offer completed	8431c16f8e1d440880db371a68f82dd0	714
306510	transaction	ba620885e51c4b0ea64a4f61daad494f	714
306511	transaction	a1a8f40407c444cc848468275308958a	714
306512	transaction	8d80970192fa496f99d6b45c470a4b60	714
306513	transaction	bde275066f3c4fa0bff3093e3b866a2c	714
306514	transaction	f1e4fd36e5a0446f83861308bddf6945	714
306515	transaction	0b64be3b241c4407a5c9a71781173829	714
306516	transaction	86d03d35d7e0434b935e7743e83be3a0	714
306517	transaction	3408fd05c781401f8442fb6dbaaea9c7	714
306518	transaction	1593d617fac246ef8e50dbb0ffd77f5f	714
306519	transaction	f1b31d07b5d84f69a2d5f1d07843989e	714
306520	transaction	2ce987015ec0404a97ba333e8e814090	714
306521	transaction	2e33545f0a764d27b2ccff95fc8d72c4	714
306522	transaction	d1c4500ace2e45e9a45d3cd2fccac8d8	714
306523	transaction	b65affd9e07346a1906364a396950e3d	714

```
306524      transaction  d613ca9c59dd42f497bdbf6178da54a7   714
306525      transaction  eec70ab28af74a22a4aeb889c0317944   714
306526      transaction  24f56b5e1849462093931b164eb803b5   714
306527  offer completed  24f56b5e1849462093931b164eb803b5   714
306528      transaction  5ca2620962114246ab218fc648eb3934   714
306529      transaction  b3a1272bc9904337b331bf348c3e8c17   714
306530      transaction  68213b08d99a4ae1b0dcb72aebd9aa35   714
306531      transaction  a00058cf10334a308c68e7631c529907   714
306532      transaction  76ddbd6576844afe811f1a3c0fbb5bec   714
306533      transaction  c02b10e8752c4d8e9b73f918558531f7   714

                                                    value  \
0         {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
1         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
2         {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
3         {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
4         {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
5         {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
6         {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
7         {'offer id': '3f207df678b143eea3cee63160fa8bed'}
8         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
9         {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
10        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
11        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
12        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
13        {'offer id': '3f207df678b143eea3cee63160fa8bed'}
14        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
15        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
16        {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
17        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
18        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
19        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
20        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
21        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
22        {'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}
23        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
24        {'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}
25        {'offer id': '2906b810c7d4411798c6938adc9daaa5'}
26        {'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}
27        {'offer id': 'ae264e3637204a6fb9bb56bc8210ddfd'}
28        {'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
29        {'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
...                                                   ...
306504                                {'amount': 4.89}
306505                                {'amount': 4.48}
306506  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306507   {'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
306508                                {'amount': 1.19}
```

```
306509  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306510                             {'amount': 14.31}
306511                              {'amount': 2.37}
306512                              {'amount': 6.92}
306513                             {'amount': 12.73}
306514                               {'amount': 8.2}
306515                               {'amount': 2.6}
306516                               {'amount': 9.2}
306517                              {'amount': 11.7}
306518                             {'amount': 40.67}
306519                             {'amount': 31.13}
306520                 {'amount': 1.6400000000000001}
306521                             {'amount': 17.35}
306522                              {'amount': 4.42}
306523                             {'amount': 18.35}
306524                             {'amount': 25.14}
306525                             {'amount': 43.58}
306526                             {'amount': 22.64}
306527  {'offer_id': 'fafdcd668e3743c1bb461111dcafc2a4...
306528                               {'amount': 2.2}
306529                 {'amount': 1.5899999999999999}
306530                              {'amount': 9.53}
306531                              {'amount': 3.61}
306532                 {'amount': 3.5300000000000002}
306533                              {'amount': 4.05}

                                offer_id  reward  amount offer_type_map  \
0       9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
1       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
2       2906b810c7d4411798c6938adc9daaa5       0       0       discount
3       fafdcd668e3743c1bb461111dcafc2a4       0       0       discount
4       4d5c57ea9a6940dd891ad53e9dbe8da0       0       0           bogo
5       f19421c1d4aa40978ebb69ca19b0e20d       0       0           bogo
6       2298d6c36e964ae4a3e7e9706d1fb8c2       0       0       discount
7       3f207df678b143eea3cee63160fa8bed       0       0  informational
8       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
9       0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
10      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
11      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
12      ae264e3637204a6fb9bb56bc8210ddfd       0       0           bogo
13      3f207df678b143eea3cee63160fa8bed       0       0  informational
14      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
15      fafdcd668e3743c1bb461111dcafc2a4       0       0       discount
16      0b1e1539f2cc45b7b9fa7c272da2e1d7       0       0       discount
17      2906b810c7d4411798c6938adc9daaa5       0       0       discount
18      9b98b8c7a33c4b65b9aebfe6a799e6d9       0       0           bogo
19      ae264e3637204a6fb9bb56bc8210ddfd       0       0           bogo
20      2906b810c7d4411798c6938adc9daaa5       0       0       discount
```

```
21      9b98b8c7a33c4b65b9aebfe6a799e6d9      0      0        bogo
22      4d5c57ea9a6940dd891ad53e9dbe8da0      0      0        bogo
23      2906b810c7d4411798c6938adc9daaa5      0      0    discount
24      2298d6c36e964ae4a3e7e9706d1fb8c2      0      0    discount
25      2906b810c7d4411798c6938adc9daaa5      0      0    discount
26      f19421c1d4aa40978ebb69ca19b0e20d      0      0        bogo
27      ae264e3637204a6fb9bb56bc8210ddfd      0      0        bogo
28      9b98b8c7a33c4b65b9aebfe6a799e6d9      0      0        bogo
29      fafdcd668e3743c1bb461111dcafc2a4      0      0    discount
...                                  ...    ...    ...         ...
306504                                        0      4          NA
306505                                        0      4          NA
306506  fafdcd668e3743c1bb461111dcafc2a4      2      0    discount
306507  0b1e1539f2cc45b7b9fa7c272da2e1d7      0      0    discount
306508                                        0      1          NA
306509  fafdcd668e3743c1bb461111dcafc2a4      2      0    discount
306510                                        0     14          NA
306511                                        0      2          NA
306512                                        0      6          NA
306513                                        0     12          NA
306514                                        0      8          NA
306515                                        0      2          NA
306516                                        0      9          NA
306517                                        0     11          NA
306518                                        0     40          NA
306519                                        0     31          NA
306520                                        0      1          NA
306521                                        0     17          NA
306522                                        0      4          NA
306523                                        0     18          NA
306524                                        0     25          NA
306525                                        0     43          NA
306526                                        0     22          NA
306527  fafdcd668e3743c1bb461111dcafc2a4      2      0    discount
306528                                        0      2          NA
306529                                        0      1          NA
306530                                        0      9          NA
306531                                        0      3          NA
306532                                        0      3          NA
306533                                        0      4          NA

       age_offer map_gender_offer         income
0            70s                F  100000.000000
1           110s               NA   65404.991568
2            60s                M   70000.000000
3           110s               NA   65404.991568
4           110s               NA   65404.991568
5            60s                M   53000.000000
```

6	110s	NA	65404.991568
7	50s	M	51000.000000
8	60s	F	57000.000000
9	60s	F	71000.000000
10	110s	NA	65404.991568
11	110s	NA	65404.991568
12	60s	M	100000.000000
13	70s	F	71000.000000
14	80s	F	53000.000000
15	40s	M	69000.000000
16	50s	F	88000.000000
17	110s	NA	65404.991568
18	50s	M	41000.000000
19	20s	M	70000.000000
20	90s	F	89000.000000
21	40s	M	33000.000000
22	60s	F	57000.000000
23	110s	NA	65404.991568
24	20s	F	63000.000000
25	110s	NA	65404.991568
26	70s	M	40000.000000
27	20s	F	30000.000000
28	40s	M	33000.000000
29	110s	NA	65404.991568
...	...	...	...
306504	70s	M	75000.000000
306505	110s	NA	65404.991568
306506	110s	NA	65404.991568
306507	60s	F	64000.000000
306508	30s	M	39000.000000
306509	30s	M	39000.000000
306510	30s	F	68000.000000
306511	60s	M	63000.000000
306512	30s	M	62000.000000
306513	30s	M	65000.000000
306514	40s	M	56000.000000
306515	60s	F	42000.000000
306516	50s	F	53000.000000
306517	20s	M	48000.000000
306518	100s	F	82000.000000
306519	60s	M	84000.000000
306520	30s	M	61000.000000
306521	40s	M	64000.000000
306522	80s	M	59000.000000
306523	40s	M	69000.000000
306524	40s	F	68000.000000
306525	50s	M	95000.000000
306526	40s	F	80000.000000

```
306527       40s                  F   80000.000000
306528      110s                 NA   65404.991568
306529       60s                  M   47000.000000
306530       50s                  M   62000.000000
306531       60s                  F   52000.000000
306532       50s                  M   40000.000000
306533      110s                 NA   65404.991568

[306534 rows x 11 columns]
```

Removing all 'NA' in transcript.event

```
In [129]: tran=transcript[transcript['event']!='NA']
```

```
In [ ]:
```

converting event column into different columns as per the various events

```
In [130]: data=pd.get_dummies(transcript['event'])

In [131]: tran=tran.join(data) #joining data and tran

In [132]: tran=tran.drop('value',axis=1)

In [133]: data=pd.get_dummies(transcript['offer_type_map']) #dumming offer_type_map column

In [134]: tran=tran.join(data) #joining data and tran

In [135]: data=pd.get_dummies(transcript['offer_id']) #dumming offer_id column

In [136]: tran=tran.join(data) #joining data and tran

In [137]: tran
```

```
Out[137]:                    event                            person  time  \
          0         offer received  78afa995795e4d85b5d9ceeca43f5fef     0
          1         offer received  a03223e636434f42ac4c3df47e8bac43     0
          2         offer received  e2127556f4f64592b11af22de27a7932     0
          3         offer received  8ec6ce2a7e7949b1bf142def7d0e0586     0
          4         offer received  68617ca6246f4fbc85e91a2a49552598     0
          5         offer received  389bc3fa690240e798340f5a15918d5c     0
          6         offer received  c4863c7985cf408faee930f111475da3     0
          7         offer received  2eeac8d8feae4a8cad5a6af0499a211d     0
          8         offer received  aa4862eba776480b8bb9c68455b8c2e1     0
          9         offer received  31dda685af34476cad5bc968bdb01c53     0
          10        offer received  744d603ef08c4f33af5a61c8c7628d1c     0
          11        offer received  3d02345581554e81b7b289ab5e288078     0
          12        offer received  4b0da7e80e5945209a1fdddfe813dbe0     0
          13        offer received  c27e0d6ab72c455a8bb66d980963de60     0
          14        offer received  d53717f5400c4e84affdaeda9dd926b3     0
```

```
15        offer received  f806632c011441378d4646567f357a21     0
16        offer received  d058f73bf8674a26a95227db098147b1     0
17        offer received  65aba5c617294649aeb624da249e1ee5     0
18        offer received  ebe7ef46ea6f4963a7dd49f501b26779     0
19        offer received  1e9420836d554513ab90eba98552d0a9     0
20        offer received  868317b9be554cb18e50bc68484749a2     0
21        offer received  f082d80f0aac47a99173ba8ef8fc1909     0
22        offer received  102e9454054946fda62242d2e176fdce     0
23        offer received  4beeb3ed64dd4898b0edf2f6b67426d3     0
24        offer received  9f30b375d7bd4c62a884ffe7034e09ee     0
25        offer received  25c906289d154b66bf579693f89481c9     0
26        offer received  6e014185620b49bd98749f728747572f     0
27        offer received  02c083884c7d45b39cc68e1314fec56c     0
28        offer received  c0d210398dee4a0895b24444a5fcd1d2     0
29        offer received  8be4463721e14d7fa600686bf8c8b2ed     0
...                  ...                               ...   ...
306504       transaction  8524d450673b4c24869b6c94380006de   714
306505       transaction  b895c57e8cd047a8872ce02aa54759d6   714
306506   offer completed  b895c57e8cd047a8872ce02aa54759d6   714
306507      offer viewed  8dda575c2a1d44b9ac8e8b07b93d1f8e   714
306508       transaction  8431c16f8e1d440880db371a68f82dd0   714
306509   offer completed  8431c16f8e1d440880db371a68f82dd0   714
306510       transaction  ba620885e51c4b0ea64a4f61daad494f   714
306511       transaction  a1a8f40407c444cc848468275308958a   714
306512       transaction  8d80970192fa496f99d6b45c470a4b60   714
306513       transaction  bde275066f3c4fa0bff3093e3b866a2c   714
306514       transaction  f1e4fd36e5a0446f83861308bddf6945   714
306515       transaction  0b64be3b241c4407a5c9a71781173829   714
306516       transaction  86d03d35d7e0434b935e7743e83be3a0   714
306517       transaction  3408fd05c781401f8442fb6dbaaea9c7   714
306518       transaction  1593d617fac246ef8e50dbb0ffd77f5f   714
306519       transaction  f1b31d07b5d84f69a2d5f1d07843989e   714
306520       transaction  2ce987015ec0404a97ba333e8e814090   714
306521       transaction  2e33545f0a764d27b2ccff95fc8d72c4   714
306522       transaction  d1c4500ace2e45e9a45d3cd2fccac8d8   714
306523       transaction  b65affd9e07346a1906364a396950e3d   714
306524       transaction  d613ca9c59dd42f497bdbf6178da54a7   714
306525       transaction  eec70ab28af74a22a4aeb889c0317944   714
306526       transaction  24f56b5e1849462093931b164eb803b5   714
306527   offer completed  24f56b5e1849462093931b164eb803b5   714
306528       transaction  5ca2620962114246ab218fc648eb3934   714
306529       transaction  b3a1272bc9904337b331bf348c3e8c17   714
306530       transaction  68213b08d99a4ae1b0dcb72aebd9aa35   714
306531       transaction  a00058cf10334a308c68e7631c529907   714
306532       transaction  76ddbd6576844afe811f1a3c0fbb5bec   714
306533       transaction  c02b10e8752c4d8e9b73f918558531f7   714

                                   offer_id  reward  amount offer_type_map  \
```

0	9b98b8c7a33c4b65b9aebfe6a799e6d9	0	0	bogo
1	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
2	2906b810c7d4411798c6938adc9daaa5	0	0	discount
3	fafdcd668e3743c1bb461111dcafc2a4	0	0	discount
4	4d5c57ea9a6940dd891ad53e9dbe8da0	0	0	bogo
5	f19421c1d4aa40978ebb69ca19b0e20d	0	0	bogo
6	2298d6c36e964ae4a3e7e9706d1fb8c2	0	0	discount
7	3f207df678b143eea3cee63160fa8bed	0	0	informational
8	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
9	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
10	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
11	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
12	ae264e3637204a6fb9bb56bc8210ddfd	0	0	bogo
13	3f207df678b143eea3cee63160fa8bed	0	0	informational
14	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
15	fafdcd668e3743c1bb461111dcafc2a4	0	0	discount
16	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
17	2906b810c7d4411798c6938adc9daaa5	0	0	discount
18	9b98b8c7a33c4b65b9aebfe6a799e6d9	0	0	bogo
19	ae264e3637204a6fb9bb56bc8210ddfd	0	0	bogo
20	2906b810c7d4411798c6938adc9daaa5	0	0	discount
21	9b98b8c7a33c4b65b9aebfe6a799e6d9	0	0	bogo
22	4d5c57ea9a6940dd891ad53e9dbe8da0	0	0	bogo
23	2906b810c7d4411798c6938adc9daaa5	0	0	discount
24	2298d6c36e964ae4a3e7e9706d1fb8c2	0	0	discount
25	2906b810c7d4411798c6938adc9daaa5	0	0	discount
26	f19421c1d4aa40978ebb69ca19b0e20d	0	0	bogo
27	ae264e3637204a6fb9bb56bc8210ddfd	0	0	bogo
28	9b98b8c7a33c4b65b9aebfe6a799e6d9	0	0	bogo
29	fafdcd668e3743c1bb461111dcafc2a4	0	0	discount
...	...	...	...	...
306504		0	4	NA
306505		0	4	NA
306506	fafdcd668e3743c1bb461111dcafc2a4	2	0	discount
306507	0b1e1539f2cc45b7b9fa7c272da2e1d7	0	0	discount
306508		0	1	NA
306509	fafdcd668e3743c1bb461111dcafc2a4	2	0	discount
306510		0	14	NA
306511		0	2	NA
306512		0	6	NA
306513		0	12	NA
306514		0	8	NA
306515		0	2	NA
306516		0	9	NA
306517		0	11	NA
306518		0	40	NA
306519		0	31	NA
306520		0	1	NA

```
306521                                          0     17           NA
306522                                          0      4           NA
306523                                          0     18           NA
306524                                          0     25           NA
306525                                          0     43           NA
306526                                          0     22           NA
306527  fafdcd668e3743c1bb461111dcafc2a4        2      0     discount
306528                                          0      2           NA
306529                                          0      1           NA
306530                                          0      9           NA
306531                                          0      3           NA
306532                                          0      3           NA
306533                                          0      4           NA

       age_offer map_gender_offer         income  \
0            70s                F  100000.000000
1           110s               NA   65404.991568
2            60s                M   70000.000000
3           110s               NA   65404.991568
4           110s               NA   65404.991568
5            60s                M   53000.000000
6           110s               NA   65404.991568
7            50s                M   51000.000000
8            60s                F   57000.000000
9            60s                F   71000.000000
10          110s               NA   65404.991568
11          110s               NA   65404.991568
12           60s                M  100000.000000
13           70s                F   71000.000000
14           80s                F   53000.000000
15           40s                M   69000.000000
16           50s                F   88000.000000
17          110s               NA   65404.991568
18           50s                M   41000.000000
19           20s                M   70000.000000
20           90s                F   89000.000000
21           40s                M   33000.000000
22           60s                F   57000.000000
23          110s               NA   65404.991568
24           20s                F   63000.000000
25          110s               NA   65404.991568
26           70s                M   40000.000000
27           20s                F   30000.000000
28           40s                M   33000.000000
29          110s               NA   65404.991568
...          ...              ...            ...
306504       70s                M   75000.000000
306505      110s               NA   65404.991568
```

```
306506       110s              NA   65404.991568
306507        60s               F   64000.000000
306508        30s               M   39000.000000
306509        30s               M   39000.000000
306510        30s               F   68000.000000
306511        60s               M   63000.000000
306512        30s               M   62000.000000
306513        30s               M   65000.000000
306514        40s               M   56000.000000
306515        60s               F   42000.000000
306516        50s               F   53000.000000
306517        20s               M   48000.000000
306518       100s               F   82000.000000
306519        60s               M   84000.000000
306520        30s               M   61000.000000
306521        40s               M   64000.000000
306522        80s               M   59000.000000
306523        40s               M   69000.000000
306524        40s               F   68000.000000
306525        50s               M   95000.000000
306526        40s               F   80000.000000
306527        40s               F   80000.000000
306528       110s              NA   65404.991568
306529        60s               M   47000.000000
306530        50s               M   62000.000000
306531        60s               F   52000.000000
306532        50s               M   40000.000000
306533       110s              NA   65404.991568

                     ...            0b1e1539f2cc45b7b9fa7c272da2e1d7  \
0                    ...                                           0
1                    ...                                           1
2                    ...                                           0
3                    ...                                           0
4                    ...                                           0
5                    ...                                           0
6                    ...                                           0
7                    ...                                           0
8                    ...                                           1
9                    ...                                           1
10                   ...                                           1
11                   ...                                           1
12                   ...                                           0
13                   ...                                           0
14                   ...                                           1
15                   ...                                           0
16                   ...                                           1
17                   ...                                           0
```

```
18                                 ...                                0
19                                 ...                                0
20                                 ...                                0
21                                 ...                                0
22                                 ...                                0
23                                 ...                                0
24                                 ...                                0
25                                 ...                                0
26                                 ...                                0
27                                 ...                                0
28                                 ...                                0
29                                 ...                                0
...                                ...                              ...
306504                             ...                                0
306505                             ...                                0
306506                             ...                                0
306507                             ...                                1
306508                             ...                                0
306509                             ...                                0
306510                             ...                                0
306511                             ...                                0
306512                             ...                                0
306513                             ...                                0
306514                             ...                                0
306515                             ...                                0
306516                             ...                                0
306517                             ...                                0
306518                             ...                                0
306519                             ...                                0
306520                             ...                                0
306521                             ...                                0
306522                             ...                                0
306523                             ...                                0
306524                             ...                                0
306525                             ...                                0
306526                             ...                                0
306527                             ...                                0
306528                             ...                                0
306529                             ...                                0
306530                             ...                                0
306531                             ...                                0
306532                             ...                                0
306533                             ...                                0

        2298d6c36e964ae4a3e7e9706d1fb8c2  2906b810c7d4411798c6938adc9daaa5  \
0                                      0                                 0
1                                      0                                 0
2                                      0                                 1
```

3	0	0
4	0	0
5	0	0
6	1	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	1
18	0	0
19	0	0
20	0	1
21	0	0
22	0	0
23	0	1
24	1	0
25	0	1
26	0	0
27	0	0
28	0	0
29	0	0
...	...	...
306504	0	0
306505	0	0
306506	0	0
306507	0	0
306508	0	0
306509	0	0
306510	0	0
306511	0	0
306512	0	0
306513	0	0
306514	0	0
306515	0	0
306516	0	0
306517	0	0
306518	0	0
306519	0	0
306520	0	0
306521	0	0
306522	0	0
306523	0	0

306524	0	0
306525	0	0
306526	0	0
306527	0	0
306528	0	0
306529	0	0
306530	0	0
306531	0	0
306532	0	0
306533	0	0

	3f207df678b143eea3cee63160fa8bed	4d5c57ea9a6940dd891ad53e9dbe8da0 \
0	0	0
1	0	0
2	0	0
3	0	0
4	0	1
5	0	0
6	0	0
7	1	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	1	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	1
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
...	...	...
306504	0	0
306505	0	0
306506	0	0
306507	0	0
306508	0	0

306509	0	0
306510	0	0
306511	0	0
306512	0	0
306513	0	0
306514	0	0
306515	0	0
306516	0	0
306517	0	0
306518	0	0
306519	0	0
306520	0	0
306521	0	0
306522	0	0
306523	0	0
306524	0	0
306525	0	0
306526	0	0
306527	0	0
306528	0	0
306529	0	0
306530	0	0
306531	0	0
306532	0	0
306533	0	0

	5a8bc65990b245e5a138643cd4eb9837	9b98b8c7a33c4b65b9aebfe6a799e6d9	\
0	0	1	
1	0	0	
2	0	0	
3	0	0	
4	0	0	
5	0	0	
6	0	0	
7	0	0	
8	0	0	
9	0	0	
10	0	0	
11	0	0	
12	0	0	
13	0	0	
14	0	0	
15	0	0	
16	0	0	
17	0	0	
18	0	1	
19	0	0	
20	0	0	

21	0	1
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	1
29	0	0
...	...	...
306504	0	0
306505	0	0
306506	0	0
306507	0	0
306508	0	0
306509	0	0
306510	0	0
306511	0	0
306512	0	0
306513	0	0
306514	0	0
306515	0	0
306516	0	0
306517	0	0
306518	0	0
306519	0	0
306520	0	0
306521	0	0
306522	0	0
306523	0	0
306524	0	0
306525	0	0
306526	0	0
306527	0	0
306528	0	0
306529	0	0
306530	0	0
306531	0	0
306532	0	0
306533	0	0

	ae264e3637204a6fb9bb56bc8210ddfd	f19421c1d4aa40978ebb69ca19b0e20d	\
0	0	0	
1	0	0	
2	0	0	
3	0	0	
4	0	0	
5	0	1	

6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	1	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	1	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	1
27	1	0
28	0	0
29	0	0
...	...	...
306504	0	0
306505	0	0
306506	0	0
306507	0	0
306508	0	0
306509	0	0
306510	0	0
306511	0	0
306512	0	0
306513	0	0
306514	0	0
306515	0	0
306516	0	0
306517	0	0
306518	0	0
306519	0	0
306520	0	0
306521	0	0
306522	0	0
306523	0	0
306524	0	0
306525	0	0
306526	0	0

306527	0	0
306528	0	0
306529	0	0
306530	0	0
306531	0	0
306532	0	0
306533	0	0

	fafdcd668e3743c1bb461111dcafc2a4
0	0
1	0
2	0
3	1
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	1
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	1
...	...
306504	0
306505	0
306506	1
306507	0
306508	0
306509	1
306510	0
306511	0

```
306512                                      0
306513                                      0
306514                                      0
306515                                      0
306516                                      0
306517                                      0
306518                                      0
306519                                      0
306520                                      0
306521                                      0
306522                                      0
306523                                      0
306524                                      0
306525                                      0
306526                                      0
306527                                      1
306528                                      0
306529                                      0
306530                                      0
306531                                      0
306532                                      0
306533                                      0

[306534 rows x 29 columns]
```

Dropping some columns and selecting others as features

```
In [138]: tran.columns
```

```
Out[138]: Index(['event', 'person', 'time', 'offer_id', 'reward', 'amount',
                 'offer_type_map', 'age_offer', 'map_gender_offer', 'income',
                 'offer completed', 'offer received', 'offer viewed', 'transaction',
                 'NA', 'bogo', 'discount', 'informational', '',
                 '0b1e1539f2cc45b7b9fa7c272da2e1d7', '2298d6c36e964ae4a3e7e9706d1fb8c2',
                 '2906b810c7d4411798c6938adc9daaa5', '3f207df678b143eea3cee63160fa8bed',
                 '4d5c57ea9a6940dd891ad53e9dbe8da0', '5a8bc65990b245e5a138643cd4eb9837',
                 '9b98b8c7a33c4b65b9aebfe6a799e6d9', 'ae264e3637204a6fb9bb56bc8210ddfd',
                 'f19421c1d4aa40978ebb69ca19b0e20d', 'fafdcd668e3743c1bb461111dcafc2a4'],
                dtype='object')
```

The following columns are selected as target variables

```
In [139]: targets = tran[["offer completed","offer received","offer viewed","transaction"]]

In [ ]:

In [140]: features=tran.drop(["person","event","offer_id","offer_type_map","offer completed","of

In [141]: features=features.drop(['',"NA"],axis=1)
```

In [142]:
```
data=pd.get_dummies(transcript['map_gender_offer'])
```

In [143]:
```
features=features.join(data)
```

In [144]:
```
features=features.drop(["map_gender_offer"],axis=1)
```

In [ ]:

In [145]:
```
remove_s=lambda x:x[:-1]
features['age_offer']=features["age_offer"].apply(remove_s)
```

In [146]:
```
features['age_offer']=features['age_offer'].astype(int)
```

In [147]:
```
features
```

Out[147]:

	time	reward	amount	age_offer	income	bogo	discount	\
0	0	0	0	70	100000.000000	1	0	
1	0	0	0	110	65404.991568	0	1	
2	0	0	0	60	70000.000000	0	1	
3	0	0	0	110	65404.991568	0	1	
4	0	0	0	110	65404.991568	1	0	
5	0	0	0	60	53000.000000	1	0	
6	0	0	0	110	65404.991568	0	1	
7	0	0	0	50	51000.000000	0	0	
8	0	0	0	60	57000.000000	0	1	
9	0	0	0	60	71000.000000	0	1	
10	0	0	0	110	65404.991568	0	1	
11	0	0	0	110	65404.991568	0	1	
12	0	0	0	60	100000.000000	1	0	
13	0	0	0	70	71000.000000	0	0	
14	0	0	0	80	53000.000000	0	1	
15	0	0	0	40	69000.000000	0	1	
16	0	0	0	50	88000.000000	0	1	
17	0	0	0	110	65404.991568	0	1	
18	0	0	0	50	41000.000000	1	0	
19	0	0	0	20	70000.000000	1	0	
20	0	0	0	90	89000.000000	0	1	
21	0	0	0	40	33000.000000	1	0	
22	0	0	0	60	57000.000000	1	0	
23	0	0	0	110	65404.991568	0	1	
24	0	0	0	20	63000.000000	0	1	
25	0	0	0	110	65404.991568	0	1	
26	0	0	0	70	40000.000000	1	0	
27	0	0	0	20	30000.000000	1	0	
28	0	0	0	40	33000.000000	1	0	
29	0	0	0	110	65404.991568	0	1	
...	...	...	...	...	...	...	...	
306504	714	0	4	70	75000.000000	0	0	
306505	714	0	4	110	65404.991568	0	0	

306506	714	2	0	110	65404.991568	0	1
306507	714	0	0	60	64000.000000	0	1
306508	714	0	1	30	39000.000000	0	0
306509	714	2	0	30	39000.000000	0	1
306510	714	0	14	30	68000.000000	0	0
306511	714	0	2	60	63000.000000	0	0
306512	714	0	6	30	62000.000000	0	0
306513	714	0	12	30	65000.000000	0	0
306514	714	0	8	40	56000.000000	0	0
306515	714	0	2	60	42000.000000	0	0
306516	714	0	9	50	53000.000000	0	0
306517	714	0	11	20	48000.000000	0	0
306518	714	0	40	100	82000.000000	0	0
306519	714	0	31	60	84000.000000	0	0
306520	714	0	1	30	61000.000000	0	0
306521	714	0	17	40	64000.000000	0	0
306522	714	0	4	80	59000.000000	0	0
306523	714	0	18	40	69000.000000	0	0
306524	714	0	25	40	68000.000000	0	0
306525	714	0	43	50	95000.000000	0	0
306526	714	0	22	40	80000.000000	0	0
306527	714	2	0	40	80000.000000	0	1
306528	714	0	2	110	65404.991568	0	0
306529	714	0	1	60	47000.000000	0	0
306530	714	0	9	50	62000.000000	0	0
306531	714	0	3	60	52000.000000	0	0
306532	714	0	3	50	40000.000000	0	0
306533	714	0	4	110	65404.991568	0	0

	informational	0b1e1539f2cc45b7b9fa7c272da2e1d7	\
0	0	0	
1	0	1	
2	0	0	
3	0	0	
4	0	0	
5	0	0	
6	0	0	
7	1	0	
8	0	1	
9	0	1	
10	0	1	
11	0	1	
12	0	0	
13	1	0	
14	0	1	
15	0	0	
16	0	1	
17	0	0	

```
18                 0                          0
19                 0                          0
20                 0                          0
21                 0                          0
22                 0                          0
23                 0                          0
24                 0                          0
25                 0                          0
26                 0                          0
27                 0                          0
28                 0                          0
29                 0                          0
...              ...                        ...
306504             0                          0
306505             0                          0
306506             0                          0
306507             0                          1
306508             0                          0
306509             0                          0
306510             0                          0
306511             0                          0
306512             0                          0
306513             0                          0
306514             0                          0
306515             0                          0
306516             0                          0
306517             0                          0
306518             0                          0
306519             0                          0
306520             0                          0
306521             0                          0
306522             0                          0
306523             0                          0
306524             0                          0
306525             0                          0
306526             0                          0
306527             0                          0
306528             0                          0
306529             0                          0
306530             0                          0
306531             0                          0
306532             0                          0
306533             0                          0

       2298d6c36e964ae4a3e7e9706d1fb8c2 ...  \
0                                     0 ...
1                                     0 ...
2                                     0 ...
```

3	0	...
4	0	...
5	0	...
6	1	...
7	0	...
8	0	...
9	0	...
10	0	...
11	0	...
12	0	...
13	0	...
14	0	...
15	0	...
16	0	...
17	0	...
18	0	...
19	0	...
20	0	...
21	0	...
22	0	...
23	0	...
24	1	...
25	0	...
26	0	...
27	0	...
28	0	...
29	0	...
...	...	...
306504	0	...
306505	0	...
306506	0	...
306507	0	...
306508	0	...
306509	0	...
306510	0	...
306511	0	...
306512	0	...
306513	0	...
306514	0	...
306515	0	...
306516	0	...
306517	0	...
306518	0	...
306519	0	...
306520	0	...
306521	0	...
306522	0	...
306523	0	...

```
306524                                   0  ...
306525                                   0  ...
306526                                   0  ...
306527                                   0  ...
306528                                   0  ...
306529                                   0  ...
306530                                   0  ...
306531                                   0  ...
306532                                   0  ...
306533                                   0  ...

        4d5c57ea9a6940dd891ad53e9dbe8da0  5a8bc65990b245e5a138643cd4eb9837  \
0                                      0                                 0
1                                      0                                 0
2                                      0                                 0
3                                      0                                 0
4                                      1                                 0
5                                      0                                 0
6                                      0                                 0
7                                      0                                 0
8                                      0                                 0
9                                      0                                 0
10                                     0                                 0
11                                     0                                 0
12                                     0                                 0
13                                     0                                 0
14                                     0                                 0
15                                     0                                 0
16                                     0                                 0
17                                     0                                 0
18                                     0                                 0
19                                     0                                 0
20                                     0                                 0
21                                     0                                 0
22                                     1                                 0
23                                     0                                 0
24                                     0                                 0
25                                     0                                 0
26                                     0                                 0
27                                     0                                 0
28                                     0                                 0
29                                     0                                 0
...                                  ...                               ...
306504                                 0                                 0
306505                                 0                                 0
306506                                 0                                 0
306507                                 0                                 0
306508                                 0                                 0
```

306509	0	0
306510	0	0
306511	0	0
306512	0	0
306513	0	0
306514	0	0
306515	0	0
306516	0	0
306517	0	0
306518	0	0
306519	0	0
306520	0	0
306521	0	0
306522	0	0
306523	0	0
306524	0	0
306525	0	0
306526	0	0
306527	0	0
306528	0	0
306529	0	0
306530	0	0
306531	0	0
306532	0	0
306533	0	0

	9b98b8c7a33c4b65b9aebfe6a799e6d9	ae264e3637204a6fb9bb56bc8210ddfd \
0	1	0
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	1
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	1	0
19	0	1
20	0	0

```
21                                        1                                    0
22                                        0                                    0
23                                        0                                    0
24                                        0                                    0
25                                        0                                    0
26                                        0                                    0
27                                        0                                    1
28                                        1                                    0
29                                        0                                    0
...                                     ...                                  ...
306504                                    0                                    0
306505                                    0                                    0
306506                                    0                                    0
306507                                    0                                    0
306508                                    0                                    0
306509                                    0                                    0
306510                                    0                                    0
306511                                    0                                    0
306512                                    0                                    0
306513                                    0                                    0
306514                                    0                                    0
306515                                    0                                    0
306516                                    0                                    0
306517                                    0                                    0
306518                                    0                                    0
306519                                    0                                    0
306520                                    0                                    0
306521                                    0                                    0
306522                                    0                                    0
306523                                    0                                    0
306524                                    0                                    0
306525                                    0                                    0
306526                                    0                                    0
306527                                    0                                    0
306528                                    0                                    0
306529                                    0                                    0
306530                                    0                                    0
306531                                    0                                    0
306532                                    0                                    0
306533                                    0                                    0

        f19421c1d4aa40978ebb69ca19b0e20d  fafdcd668e3743c1bb461111dcafc2a4  F  \
0                                      0                                 0  1
1                                      0                                 0  0
2                                      0                                 0  0
3                                      0                                 1  0
4                                      0                                 0  0
5                                      1                                 0  0
```

6	0	0	0
7	0	0	0
8	0	0	1
9	0	0	1
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	1
14	0	0	1
15	0	1	0
16	0	0	1
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	1
21	0	0	0
22	0	0	1
23	0	0	0
24	0	0	1
25	0	0	0
26	1	0	0
27	0	0	1
28	0	0	0
29	0	1	0
...	...	...	..
306504	0	0	0
306505	0	0	0
306506	0	1	0
306507	0	0	1
306508	0	0	0
306509	0	1	0
306510	0	0	1
306511	0	0	0
306512	0	0	0
306513	0	0	0
306514	0	0	0
306515	0	0	1
306516	0	0	1
306517	0	0	0
306518	0	0	1
306519	0	0	0
306520	0	0	0
306521	0	0	0
306522	0	0	0
306523	0	0	0
306524	0	0	1
306525	0	0	0
306526	0	0	1

306527	0	1	1
306528	0	0	0
306529	0	0	0
306530	0	0	0
306531	0	0	1
306532	0	0	0
306533	0	0	0

	M	NA	O
0	0	0	0
1	0	1	0
2	1	0	0
3	0	1	0
4	0	1	0
5	1	0	0
6	0	1	0
7	1	0	0
8	0	0	0
9	0	0	0
10	0	1	0
11	0	1	0
12	1	0	0
13	0	0	0
14	0	0	0
15	1	0	0
16	0	0	0
17	0	1	0
18	1	0	0
19	1	0	0
20	0	0	0
21	1	0	0
22	0	0	0
23	0	1	0
24	0	0	0
25	0	1	0
26	1	0	0
27	0	0	0
28	1	0	0
29	0	1	0
...	..	..	..
306504	1	0	0
306505	0	1	0
306506	0	1	0
306507	0	0	0
306508	1	0	0
306509	1	0	0
306510	0	0	0
306511	1	0	0

```
306512  1   0  0
306513  1   0  0
306514  1   0  0
306515  0   0  0
306516  0   0  0
306517  1   0  0
306518  0   0  0
306519  1   0  0
306520  1   0  0
306521  1   0  0
306522  1   0  0
306523  1   0  0
306524  0   0  0
306525  1   0  0
306526  0   0  0
306527  0   0  0
306528  0   1  0
306529  1   0  0
306530  1   0  0
306531  0   0  0
306532  1   0  0
306533  0   1  0

[306534 rows x 22 columns]
```

Normalising all the columns having large values so that they do not dominate in the model and all rows are equal

In [148]:
```python
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
to_normalize = ['time', 'amount', 'reward', 'income','age_offer']

features[to_normalize] = scaler.fit_transform(features[to_normalize])
features.head()
```

Out[148]:
```
   time  reward  amount  age_offer    income  bogo  discount  informational  \
0   0.0     0.0     0.0        0.6  0.777778     1         0              0
1   0.0     0.0     0.0        1.0  0.393389     0         1              0
2   0.0     0.0     0.0        0.5  0.444444     0         1              0
3   0.0     0.0     0.0        1.0  0.393389     0         1              0
4   0.0     0.0     0.0        1.0  0.393389     1         0              0

   0b1e1539f2cc45b7b9fa7c272da2e1d7  2298d6c36e964ae4a3e7e9706d1fb8c2 ...  \
0                                 0                                 0 ...
1                                 1                                 0 ...
2                                 0                                 0 ...
3                                 0                                 0 ...
```

```
4                                  0                                  0 ...

   4d5c57ea9a6940dd891ad53e9dbe8da0  5a8bc65990b245e5a138643cd4eb9837  \
0                                 0                                 0
1                                 0                                 0
2                                 0                                 0
3                                 0                                 0
4                                 1                                 0

   9b98b8c7a33c4b65b9aebfe6a799e6d9  ae264e3637204a6fb9bb56bc8210ddfd  \
0                                 1                                 0
1                                 0                                 0
2                                 0                                 0
3                                 0                                 0
4                                 0                                 0

   f19421c1d4aa40978ebb69ca19b0e20d  fafdcd668e3743c1bb461111dcafc2a4  F  M  \
0                                 0                                 0  1  0
1                                 0                                 0  0  0
2                                 0                                 0  0  1
3                                 0                                 1  0  0
4                                 0                                 0  0  0

   NA  O
0   0  0
1   1  0
2   0  0
3   1  0
4   1  0

[5 rows x 22 columns]
```

```
In [149]: from sklearn.model_selection import train_test_split, GridSearchCV

In [150]: X_train, X_test, y_train, y_test = train_test_split(features, targets, random_state=0)

In [151]: print('Training Features Shape:', X_train.shape)
          print('Training Labels Shape:', y_train.shape)
          print('Testing Features Shape:', X_test.shape)
          print('Testing Labels Shape:', y_test.shape)
```

```
Training Features Shape: (229900, 22)
Training Labels Shape: (229900, 4)
Testing Features Shape: (76634, 22)
Testing Labels Shape: (76634, 4)
```

Model using Decision tree classifier gets an accuracy of 95.5% on test set

```
In [157]: from sklearn.tree import DecisionTreeClassifier
          from sklearn.metrics import accuracy_score
          dt = DecisionTreeClassifier(criterion='gini', splitter='best', max_depth=None, min_sam
                                      class_weight=None, presort='deprecated')

          dt.fit(X_train, y_train)
          yt_pred=dt.predict(X_train)
          print(f'Accuracy of Decision Tree classifier on training set: {accuracy_score( y_train
          y_pred=dt.predict(X_test)
          print(f'Prediction Accuracy: {accuracy_score(y_test, y_pred)*100}%')
```

```
Accuracy of Decision Tree classifier on training set: 95.64984775989561%.
Prediction Accuracy: 95.56071717514419%
```

Model using RandomForestClassifier gets an accuracy of 91.3% on test set

```
In [158]: from sklearn.ensemble import RandomForestClassifier
          from sklearn.metrics import accuracy_score
          rc = RandomForestClassifier(n_estimators=200, criterion='gini', max_depth=None,min_sam

          rc.fit(X_train, y_train)
          yt_pred=rc.predict(X_train)
          print(f'Accuracy of Decision Tree classifier on training set: {accuracy_score( y_train
          y_pred=rc.predict(X_test)
          print(f'Prediction Accuracy: {accuracy_score(y_test, y_pred)*100}%')
```

```
building tree 1 of 200


[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.3s remaining:    0.0s


building tree 2 of 200
building tree 3 of 200
building tree 4 of 200
building tree 5 of 200
building tree 6 of 200
building tree 7 of 200
building tree 8 of 200
building tree 9 of 200
building tree 10 of 200
building tree 11 of 200
building tree 12 of 200
building tree 13 of 200
building tree 14 of 200
building tree 15 of 200
building tree 16 of 200
building tree 17 of 200
```

```
building tree 18 of 200
building tree 19 of 200
building tree 20 of 200
building tree 21 of 200
building tree 22 of 200
building tree 23 of 200
building tree 24 of 200
building tree 25 of 200
building tree 26 of 200
building tree 27 of 200
building tree 28 of 200
building tree 29 of 200
building tree 30 of 200
building tree 31 of 200
building tree 32 of 200
building tree 33 of 200
building tree 34 of 200
building tree 35 of 200
building tree 36 of 200
building tree 37 of 200
building tree 38 of 200
building tree 39 of 200
building tree 40 of 200
building tree 41 of 200
building tree 42 of 200
building tree 43 of 200
building tree 44 of 200
building tree 45 of 200
building tree 46 of 200
building tree 47 of 200
building tree 48 of 200
building tree 49 of 200
building tree 50 of 200
building tree 51 of 200
building tree 52 of 200
building tree 53 of 200
building tree 54 of 200
building tree 55 of 200
building tree 56 of 200
building tree 57 of 200
building tree 58 of 200
building tree 59 of 200
building tree 60 of 200
building tree 61 of 200
building tree 62 of 200
building tree 63 of 200
building tree 64 of 200
building tree 65 of 200
```

```
building tree 66 of 200
building tree 67 of 200
building tree 68 of 200
building tree 69 of 200
building tree 70 of 200
building tree 71 of 200
building tree 72 of 200
building tree 73 of 200
building tree 74 of 200
building tree 75 of 200
building tree 76 of 200
building tree 77 of 200
building tree 78 of 200
building tree 79 of 200
building tree 80 of 200
building tree 81 of 200
building tree 82 of 200
building tree 83 of 200
building tree 84 of 200
building tree 85 of 200
building tree 86 of 200
building tree 87 of 200
building tree 88 of 200
building tree 89 of 200
building tree 90 of 200
building tree 91 of 200
building tree 92 of 200
building tree 93 of 200
building tree 94 of 200
building tree 95 of 200
building tree 96 of 200
building tree 97 of 200
building tree 98 of 200
building tree 99 of 200
building tree 100 of 200
building tree 101 of 200
building tree 102 of 200
building tree 103 of 200
building tree 104 of 200
building tree 105 of 200
building tree 106 of 200
building tree 107 of 200
building tree 108 of 200
building tree 109 of 200
building tree 110 of 200
building tree 111 of 200
building tree 112 of 200
building tree 113 of 200
```

```
building tree 114 of 200
building tree 115 of 200
building tree 116 of 200
building tree 117 of 200
building tree 118 of 200
building tree 119 of 200
building tree 120 of 200
building tree 121 of 200
building tree 122 of 200
building tree 123 of 200
building tree 124 of 200
building tree 125 of 200
building tree 126 of 200
building tree 127 of 200
building tree 128 of 200
building tree 129 of 200
building tree 130 of 200
building tree 131 of 200
building tree 132 of 200
building tree 133 of 200
building tree 134 of 200
building tree 135 of 200
building tree 136 of 200
building tree 137 of 200
building tree 138 of 200
building tree 139 of 200
building tree 140 of 200
building tree 141 of 200
building tree 142 of 200
building tree 143 of 200
building tree 144 of 200
building tree 145 of 200
building tree 146 of 200
building tree 147 of 200
building tree 148 of 200
building tree 149 of 200
building tree 150 of 200
building tree 151 of 200
building tree 152 of 200
building tree 153 of 200
building tree 154 of 200
building tree 155 of 200
building tree 156 of 200
building tree 157 of 200
building tree 158 of 200
building tree 159 of 200
building tree 160 of 200
building tree 161 of 200
```

```
building tree 162 of 200
building tree 163 of 200
building tree 164 of 200
building tree 165 of 200
building tree 166 of 200
building tree 167 of 200
building tree 168 of 200
building tree 169 of 200
building tree 170 of 200
building tree 171 of 200
building tree 172 of 200
building tree 173 of 200
building tree 174 of 200
building tree 175 of 200
building tree 176 of 200
building tree 177 of 200
building tree 178 of 200
building tree 179 of 200
building tree 180 of 200
building tree 181 of 200
building tree 182 of 200
building tree 183 of 200
building tree 184 of 200
building tree 185 of 200
building tree 186 of 200
building tree 187 of 200
building tree 188 of 200
building tree 189 of 200
building tree 190 of 200
building tree 191 of 200
building tree 192 of 200
building tree 193 of 200
building tree 194 of 200
building tree 195 of 200
building tree 196 of 200
building tree 197 of 200
building tree 198 of 200
building tree 199 of 200
building tree 200 of 200
```

```
[Parallel(n_jobs=1)]: Done 200 out of 200 | elapsed:  1.2min finished
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.1s remaining:    0.0s
[Parallel(n_jobs=1)]: Done 200 out of 200 | elapsed:   13.4s finished
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.0s remaining:    0.0s
```

```
Accuracy of Decision Tree classifier on training set: 91.99869508481949%.
```

```
Prediction Accuracy: 91.2480100216614%
```

```
[Parallel(n_jobs=1)]: Done 200 out of 200 | elapsed:    4.1s finished
```

Model built using Kneighbours Classifier rreached an accuracy og 86% on test set

```
In [159]: from sklearn.neighbors import KNeighborsClassifier
          from sklearn.metrics import accuracy_score
          kn = KNeighborsClassifier()

          kn.fit(X_train, y_train)
          yt_pred=kn.predict(X_train)
          print(f'Accuracy of Decision Tree classifier on training set: {accuracy_score( y_train
          y_pred=kn.predict(X_test)
          print(f'Prediction Accuracy: {accuracy_score(y_test, y_pred)*100}%')
```

```
Accuracy of Decision Tree classifier on training set: 90.72422792518486%.
Prediction Accuracy: 86.0049064383955%
```

We can see that the best accuracy is achieved with decison tree classifier of 95.5% on test set
Following cells depict the use of GridSearchCV for improving model

```
In [163]: from sklearn.model_selection import GridSearchCV
          min_samples_leaf=[10,20,40,50,100]
          min_samples_split=[10,30,50,40,100]
          min_weight_fraction_leaf=[0.0,0.1,0.4]
          min_impurity_decrease=[0.0,0.2]
          max_features=["auto","sqrt"]
          param_grid = dict( min_weight_fraction_leaf=min_weight_fraction_leaf,min_samples_leaf=

          grid = GridSearchCV(estimator=dt, param_grid=param_grid,cv=3,n_jobs=-1)
```

```
In [ ]: grid_result = grid.fit(X_train, y_train)
```

```
In [ ]: print(f'Best Score: {grid_result.best_score_}')
        print(f'Best params: {grid_result.best_params_}')
```

Conclusion

An attempt has been made to analyze the dataset of Starbucks and understand the data and make model to predict how will a customer take the offer whether thee offer will be received or not and whether it will be completed. Firstly the data is analysed and explored and relation between different variables is understood with the help of graphs. The distribution of variables in the data =set, what types of customer base is there and gender wise distribution of the customer base based on their income and how much offers do they receive do they complete the offer or not is analysed.It is found that Females complete offers compared to males with 56% completion of what they receive compared to 43.18% among males but males receive more offers than females.

The scenario completely changes for transactions which is 72794 among males and 49382 among females. Further exploration was made on which offer is being received better by the customer. Hence females could be given more offers further offer id fafdcd668e3743c1bb461111dcafc2a4 should be more distributed compared to others since it is completed more number of times.

'Improvements

Improvement can be made on how NAN values are filled and compeleting the NAN values, further location based data could be used for location specific recommmendations.All these could help us to give more personalised recommendation to customers

```
In [ ]:
```