Machine Learning Engineer Nanodegree
Udacity

Starbucks Capstone Project Report

Peter Potros Nassim

Thursday, July 8, 2021

I. Overview

I came across the article '*AI for humanity: How Starbucks plans to use technology to nurture the human spirit*' and I was very much motivated by influence of *AI* not just to achieve profits but also help care for our customers and make them feel comfortable and cosy through building a connection which is to be determined through use of models to understand our preferences. I think that's what makes able to boom nowadays. If we look around, we will be seeing many businesses working in same domain but we could easily notice the variations and outliers in the market come from what is different and what can build a relation with the customer with least error as much as possible!

Starbucks was established in March 31st ,1971 by three local businessmen to sell high quality whole beans coffee. Starbucks focus on consumer habits and share its speciality of coffee with the buyers.

As we are now living in a new era where tons of data are generated every day which grow and scale dramatically. Applying businesses analytics concepts enable us to build **Descriptive Reporting, Predictive Reporting** and **Prescriptive reporting.** Additionally, we can enhance the business throughout different stages through the application of Machine learning.

One of the most applications of ML takes place in marketing and sales domain in order to provide more personalized, effective promotions, recommend products, increase profitability, studying new products and its effect on markets.

Customer targeting has changed the whole idea of marketing whereby you can be able to reach the most likely customer through patterns generated from search history. We are having many case studies applying this concept in their businesses including real estate, telecom, e-commerce also many research papers are discussing how can we make better use of customer data to reach better results and launching successful campaigns.

II. Problem Introduction

Our project aims at identifying a criterion upon which Starbucks will be sending offers to its customers and predicting how they will respond to different offers depending on hidden traits of customers and which cluster is most likely to accept each offer from the following: buy-one-get-one (BOGO), discount, and informational. This helps perform better prediction of which offer is most likely to excite each of our clients individually.

Let’s start by understanding the journey of our customer. It starts by receiving an offer out of three:

- BOGO (Buy One Get One free)
- Discount
- Informational

Then there will be two possibilities whether or not our customer is going to view the offer. In case our client viewed the offer and took an action, he will be going to Starbucks, at this point our offer is considered effective and our client is a successful target. There is another possibility that the offer is not that attractive and so he wouldn’t be taking action, here our offer isn’t effective and it would be preferred if offer was not sent or at least reduce the probability of sending unusable offers.

When dealing with transactions which translate whether or not our offer was effective, we must take into account validity time of our offer as if a transaction takes place after validity time then our offer was not effective and this transaction had happened with no effect of the offer sent. Also, informational offer is considered to be effective if a transaction took place within certain time of our offer after receiving the advertisement.

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 7 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.

All of the previous scenarios must be taken into account when building our model in order to be able to extract the correct features able to come up with a model that better describes our customers and be able to send the most desirable offer to the customer leading to building a relation with him.

III. Explore & Process Data

This data set contains simulated data that mimics customer behavior on the Starbucks rewards mobile app whereby it simulates how people decide which product to purchase. The data is contained in three files:

portfolio.json: includes 10 offers with the following features:

- 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: includes 17000 users with the following features:

- age (int) - age of the customer, missing value encoded as 118.
- became_member_on (int) - date when customer created an app account
- gender (str) - gender of the customer (note some entries contain 'O' for other and rather than M or F)
- id (str) - customer id
- income (float) - customer's income

transcript.json: includes 306648 observations with the following features:

- 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 - (dictionary) different values depending on event type
    - ✓ offer id: (string/hash) not associated with any "transaction"
    - ✓ amount: (numeric) money spent in "transaction"
    - ✓ reward: (numeric) money gained from "offer completed"

Step 1 : After exploring our data and understand each of the features given, we start by importing the libraries to be used pandas, numpy, matplotlib and seaborn for visualisation in addition to libraries related to building models including keras, sklearn, XGBoost, etc.

Step 2 : We then take each of the datasets and start dealing with nan and missing data through imputation of the average into these values or removing observation in case it doesn't affect the output, checking incorrect data formats and duplicates.

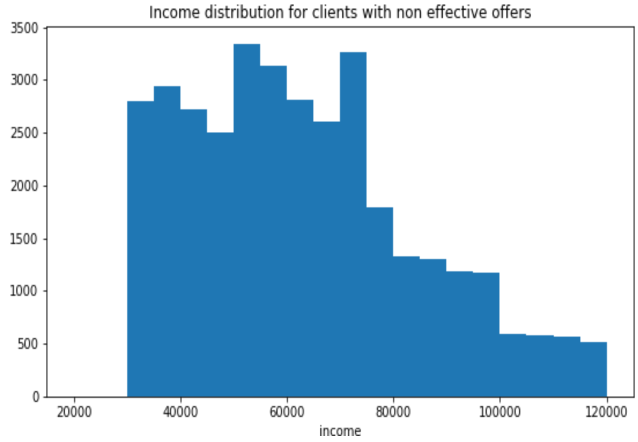
Step 3 : Splitting transcript into 2 datasets, one including transactions and another one including offers with events occurring throughout time. We then start merging our datasets together in order to come up with a data frame carrying all observations with all information about our customers from their profile and all information about offers from portfolio dataset.

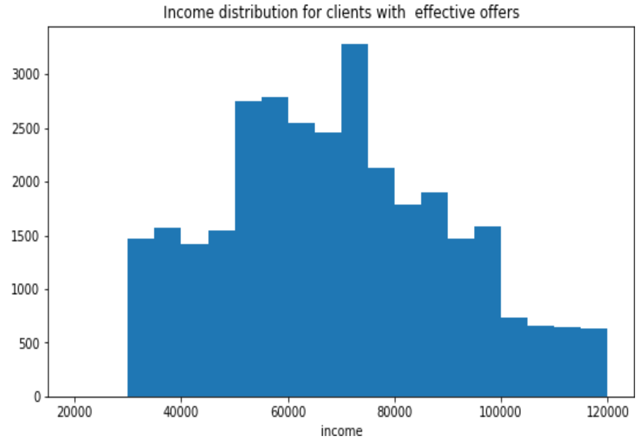
Step 4 : Applying some feature transformation on our data by calculating expiry time of the offer and extracting each of the events whether received, viewed or completed in separate datasets.

Step 5 : looping through our distinct customers to determine what are the offers do they receive and which of these offers is completed to determine whether or not it is effective and we will be generating a new dataframe.

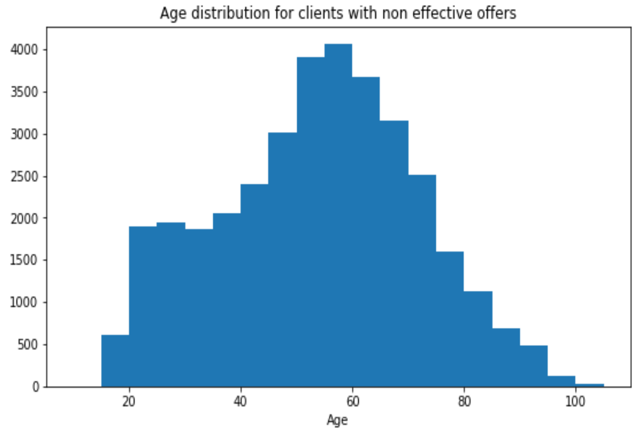
IV. Data Visualization and analysis

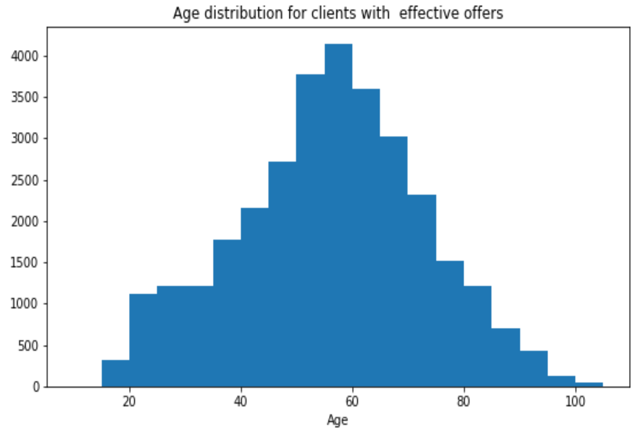
Among the visualizations that we have constructed is one studying income distribution for clients with effective and non-effective offers and we can see that clients with lower income range are less likely to be effective than middle range income.


Income distribution for clients with non effective offers

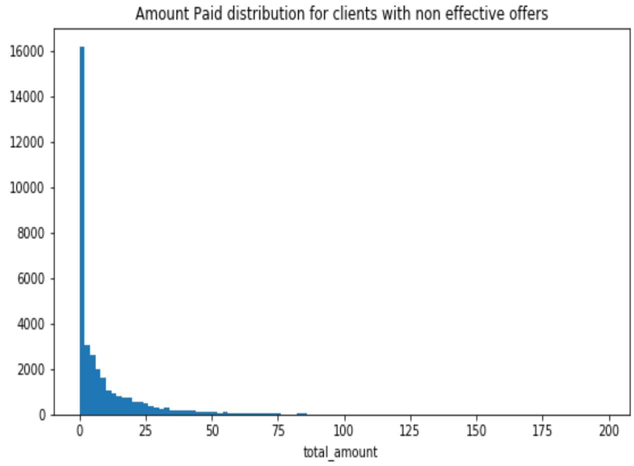

Income distribution for clients with effective offers

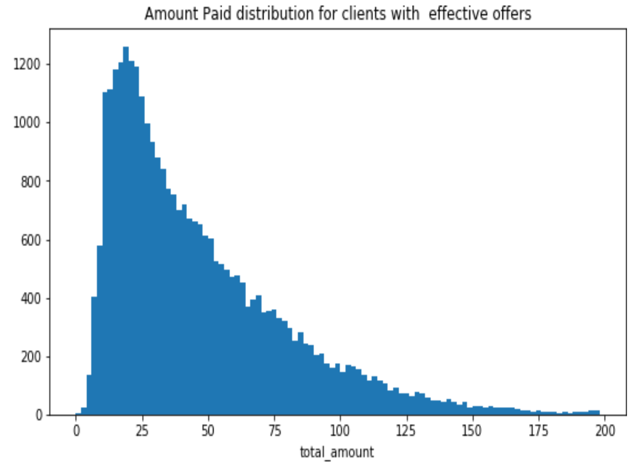
Another one studying age distribution for both effective and non-effective clients and it shows that non-effective distribution is somehow skewed to the right whereby youth are less likely to some extent to proceed towards completion of offer.


Age distribution for clients with non effective offers

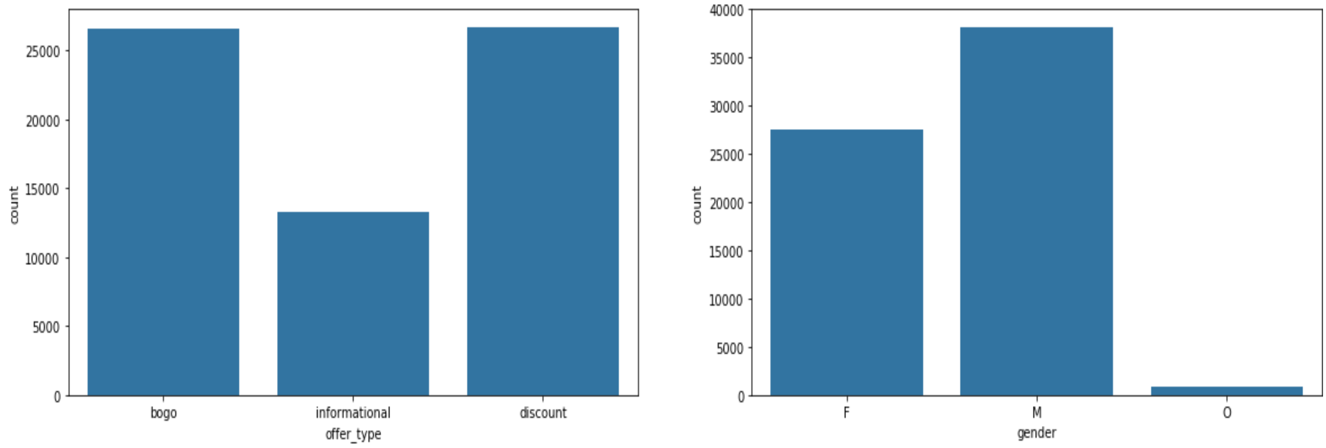

Age distribution for clients with effective offers

It makes sense that completed offers are associated with higher amount paid whereas amount paid out of offer is low compared to that in offer. This may be due to motivation achieved by the offer.

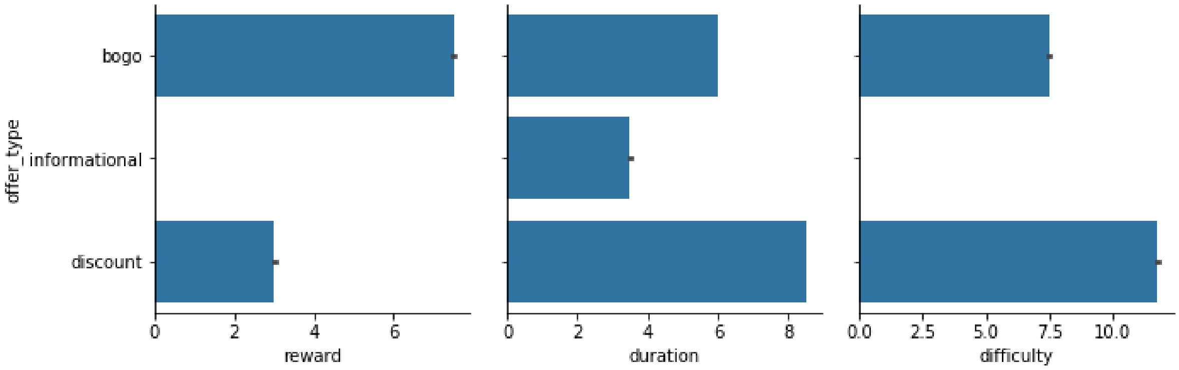




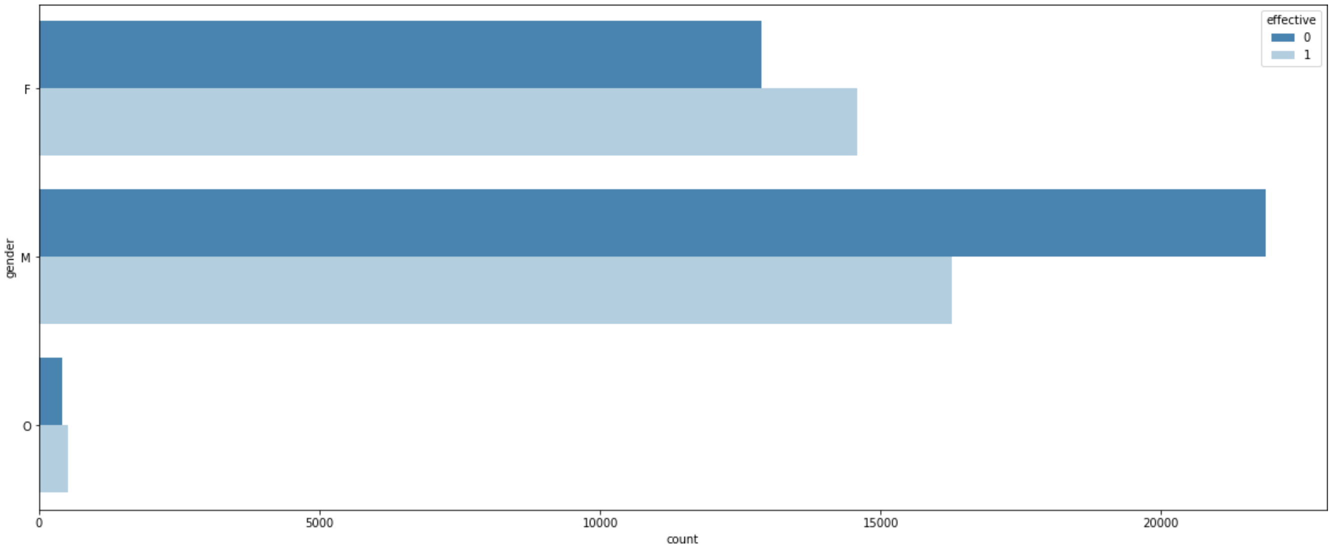
This figure illustrates statistical analysis of the different offers and gender



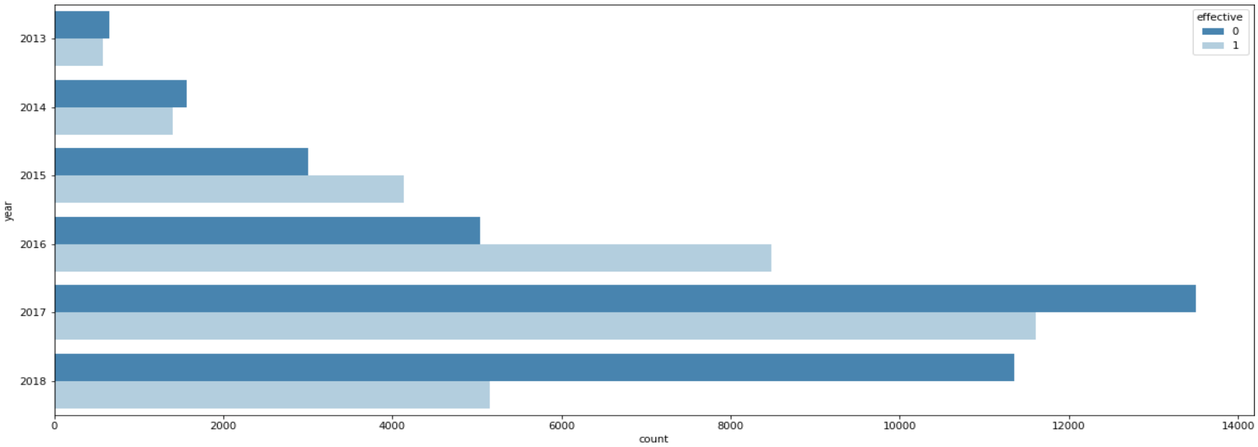
This figure illustrates some statistics about different offer type comparing between three of them in terms of average duration, reward and difficulty.



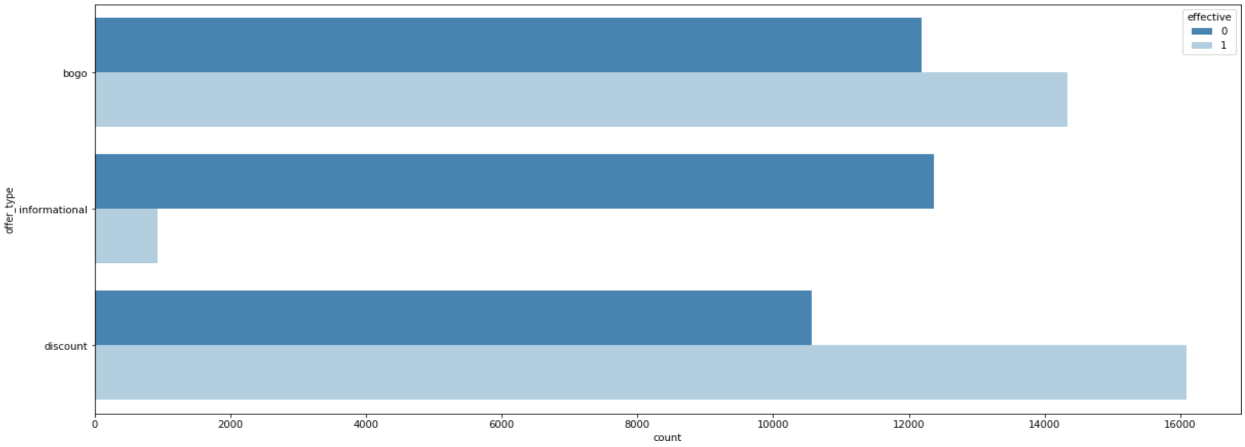
This figure shows that females achieve higher possibility for completing an offer compared to males. Although number of males is higher but overall, males are less likely to proceed.



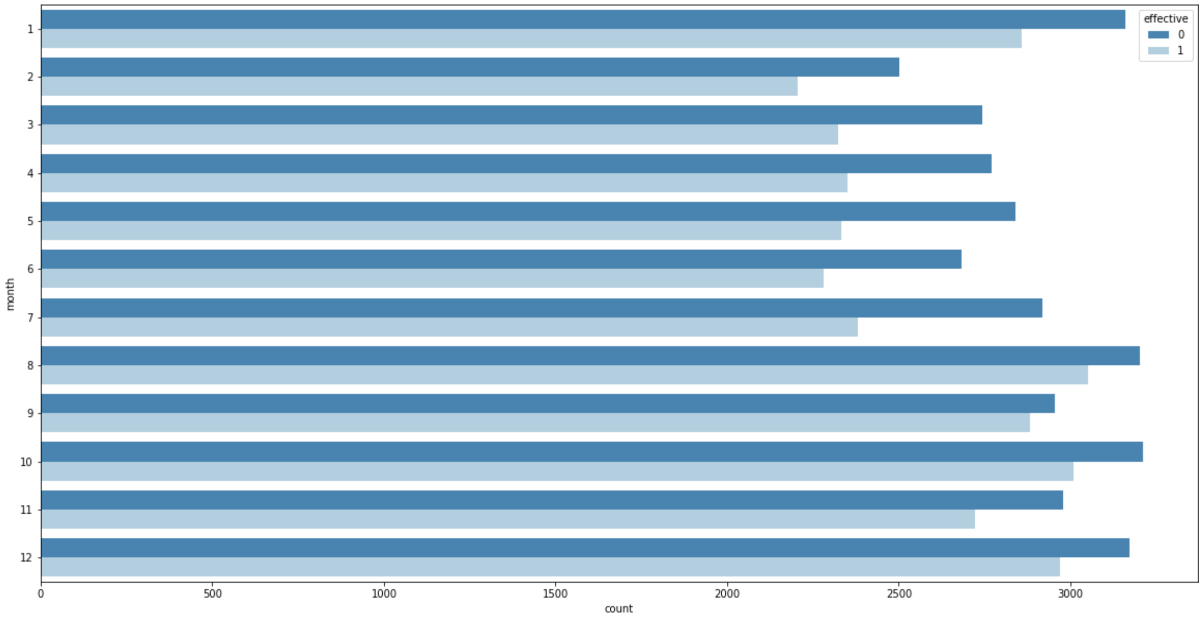
This figure shows that 2016 clients achieve better results followed by 2015 whilst 2017 represents year with greatest number of customers joining starbucks.

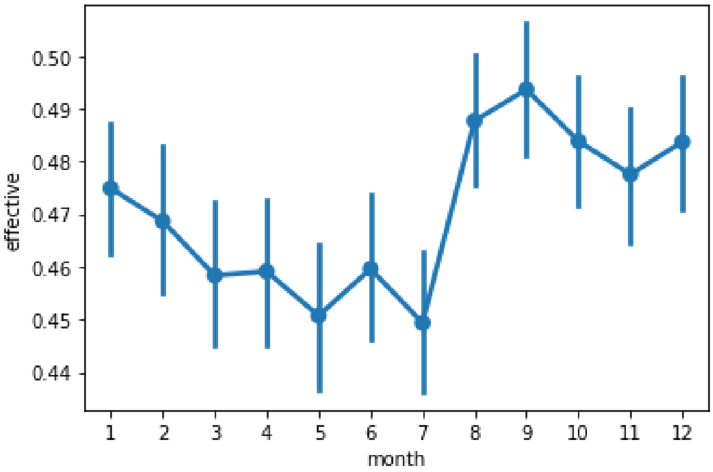


This shows that discount offers are most preferred among all other offers.

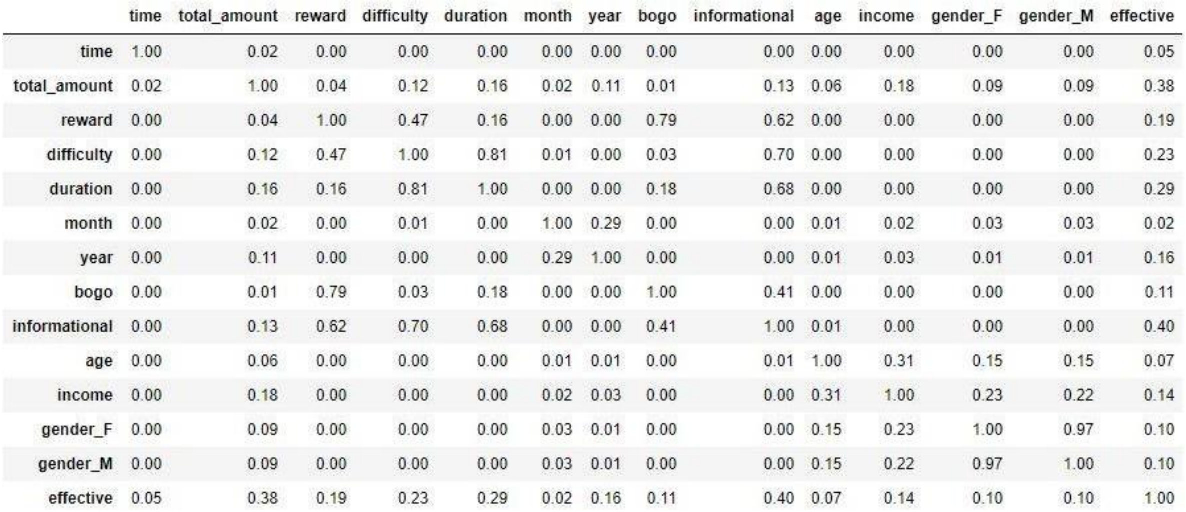


These 2 figures show that clients who became members in months august, September and October are more effective and respond to the offers sent to them whilst those who became members in may, june and july are least effective.





We then create a correlation matrix to identify if any of the features is correlated to another.



	time	total_amount	reward	difficulty	duration	month	year	bogo	informational	age	income	gender_F	gender_M	effective
time	1.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
total_amount	0.02	1.00	0.04	0.12	0.16	0.02	0.11	0.01	0.13	0.06	0.18	0.09	0.09	0.38
reward	0.00	0.04	1.00	0.47	0.16	0.00	0.00	0.79	0.62	0.00	0.00	0.00	0.00	0.19
difficulty	0.00	0.12	0.47	1.00	0.81	0.01	0.00	0.03	0.70	0.00	0.00	0.00	0.00	0.23
duration	0.00	0.16	0.16	0.81	1.00	0.00	0.00	0.18	0.68	0.00	0.00	0.00	0.00	0.29
month	0.00	0.02	0.00	0.01	0.00	1.00	0.29	0.00	0.00	0.01	0.02	0.03	0.03	0.02
year	0.00	0.11	0.00	0.00	0.00	0.29	1.00	0.00	0.00	0.01	0.03	0.01	0.01	0.16
bogo	0.00	0.01	0.79	0.03	0.18	0.00	0.00	1.00	0.41	0.00	0.00	0.00	0.00	0.11
informational	0.00	0.13	0.62	0.70	0.68	0.00	0.00	0.41	1.00	0.01	0.00	0.00	0.00	0.40
age	0.00	0.06	0.00	0.00	0.00	0.01	0.01	0.00	0.01	1.00	0.31	0.15	0.15	0.07
income	0.00	0.18	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.31	1.00	0.23	0.22	0.14
gender_F	0.00	0.09	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.15	0.23	1.00	0.97	0.10
gender_M	0.00	0.09	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.15	0.22	0.97	1.00	0.10
effective	0.05	0.38	0.19	0.23	0.29	0.02	0.16	0.11	0.40	0.07	0.14	0.10	0.10	1.00

V. Pre-processing Data

We start by extracting features and targets and putting them in 2 arrays preparing data for train test split and stratifying data so that test set carried data with same distribution as train then we scale our data so none of the features reduce the importance of another using standard scaler.

VI. Evaluation Metrics

We will evaluate our classification models using precision, recall and f1-score after calculation of each of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) and then construct confusion matrix. Additionally, we will look for accuracy which will be a valid choice in case of well-balanced and no skewed data. Precision is calculated through the formula: (TP)/ (TP+FP) and the recall through the formula: (TP)/ (TP+FN) and finally, the f1-score through: (2*((precision*recall)/(precision+recall))

VII. Training and Testing Model

Benchmark Model ***: Logistic Regression***

We would refer to logistic regression model as benchmark model which can be used to give us an indication of the minimum score that our model can achieve.

```
Confusion Matrix:
[[8003  781]
 [1346 6496]]
 Accuracy:  0.8720678455431252
 F1_score:  0.8593160923341491
```

Naive Bayes:

```
[[4804 3980]
 [ 874 6968]]
Accuracy:  0.7080476362324071
F1_score:  0.7416711016498138
```

Neural Network Using Keras

```
[[7910  874]
 [ 655 7187]]
Accuracy:  0.9080356068807891
F1_score:  0.9038546186254166
```

Support Vector Machine (svm)

```
[[8084  700]
 [ 933 6909]]
Accuracy:  0.9017803440394563
F1_score:  0.8943110478286195
```

Random Forest Classifier

```
[[7946  838]
 [ 576 7266]]
Accuracy:  0.9149524840611091
F1_score:  0.9113257243195786
```

XGBoost

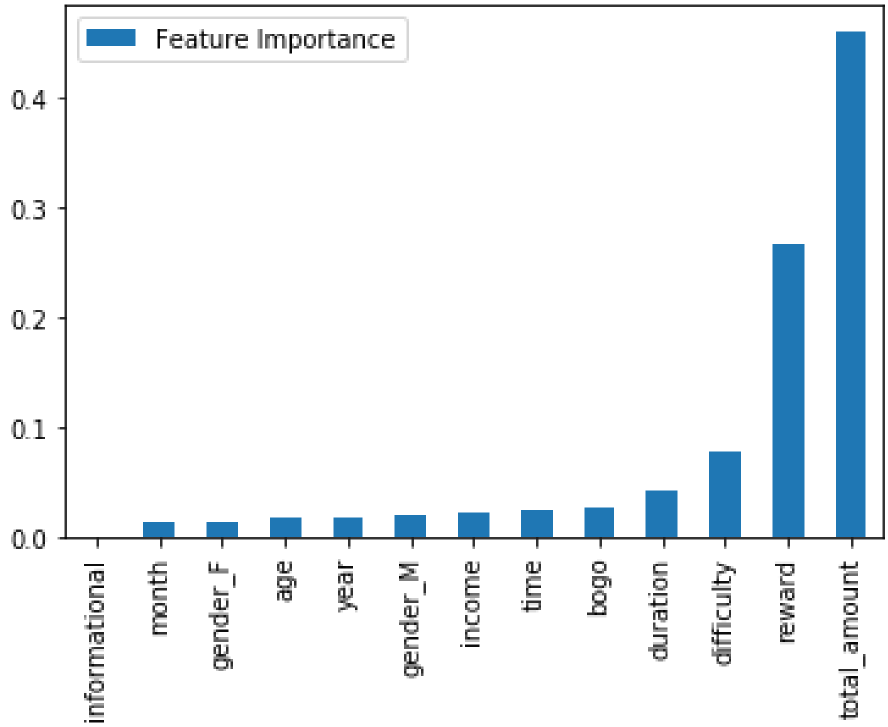
```
[[7982  802]
 [ 593 7249]]
Accuracy:  0.9160952724648141
F1_score:  0.9122255080853205
```

XGBoost after hyperparameter tuning

```
[[7954  830]
 [ 582 7260]]
Accuracy:  0.9150727775772886
F1_score:  0.9113733366808938
```

Tuned XGBoost and Random forest classifier gave the highest results with least f1_score, in my opinion, tuned XGBoost is somehow complicated and random forest yields good results and is considered reliable.

After tuning hyperparameters, we can see that total_amount and reward are the most important features that help perform better predictions


Feature Importance
informational
month
gender_F
age
year
gender_M
income
time
bogo
duration
difficulty
reward
total_amount