TELECOM CHURN CASE STUDY

BY

MAYURI ALONI
SREE BADRI
ANADI MITRA

Introduction

Project Objective

Dataset Description

Methodologies

Exploratory Data Analysis

Model selection

Conclusions

INTRODUCTION

- Churn prediction is one of the most popular Big Data use cases in business. It consists of detecting customers who are likely to cancel a subscription to a service.
- Churn is a problem for telecom companies because it is more expensive to acquire a new customer than to keep your existing one from leaving.

PROJECT OBJECTIVE

- TO PREDICT CUSTOMER CHURN.
- HIGHLIGHTING THE MAIN VARIABLES/FACTORS INFLUENCING CUSTOMER CHURN.
- USE VARIOUS ML ALGORITHMS TO BUILD PREDICTION MODELS, EVALUATE THE ACCURACY AND PERFORMANCE OF THESE MODELS.

- FINDING OUT THE BEST MODEL FOR OUR BUSINESS CASE & PROVIDING EXECUTIVE SUMMARY.

DATASET DESCRIPTION

- Source dataset is in csv format.
- There is no missing values for the provided input dataset.

- Churn is the variable which notifies whether a particular customer is churned or not.

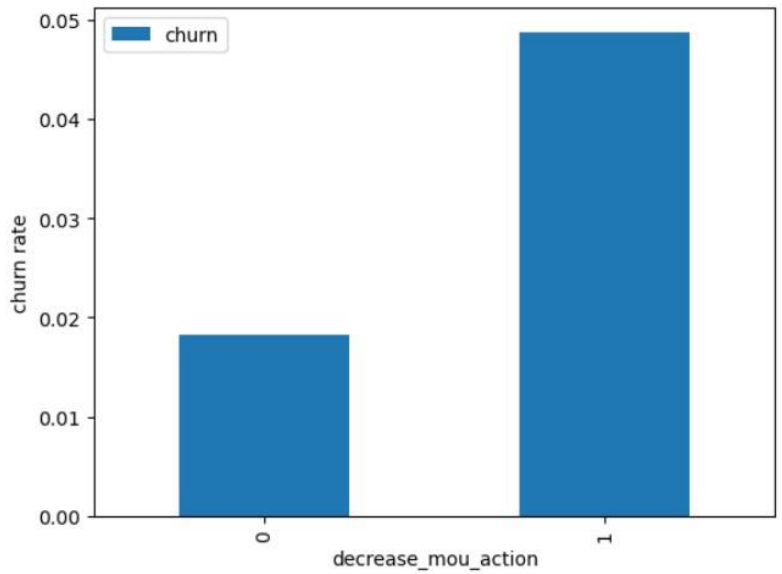
METHODOLOGIES

- EDA(EXPLORATORY DATA ANALYSIS): THE DATASET CONSISTS OF 12 VARIABLES IN ALL. A FEW ARE CONTINUOUS, REST ARE CATEGORICAL. THE CONTROL VARIABLE WAS CUSTOMER.

- MODEL BUILDING WHICH INCLUDES DEFINING THE PURPOSE IF MODEL, DETERMINE THE MODEL BOUNDARY, BUILD THE MODEL, CREATE AN INTERFACE AND EXPORT THE MODEL.

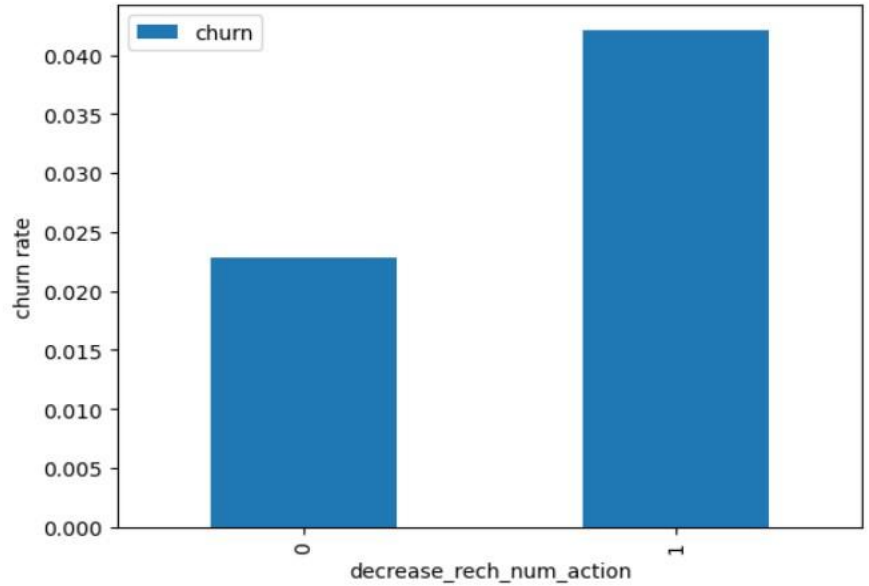
- EVALUATING MACHINE LEARNING ALGORITHM IS AN ESSENTIAL PART OF PROJECT.

EXPLORATORY DATA ANALYSIS



the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.

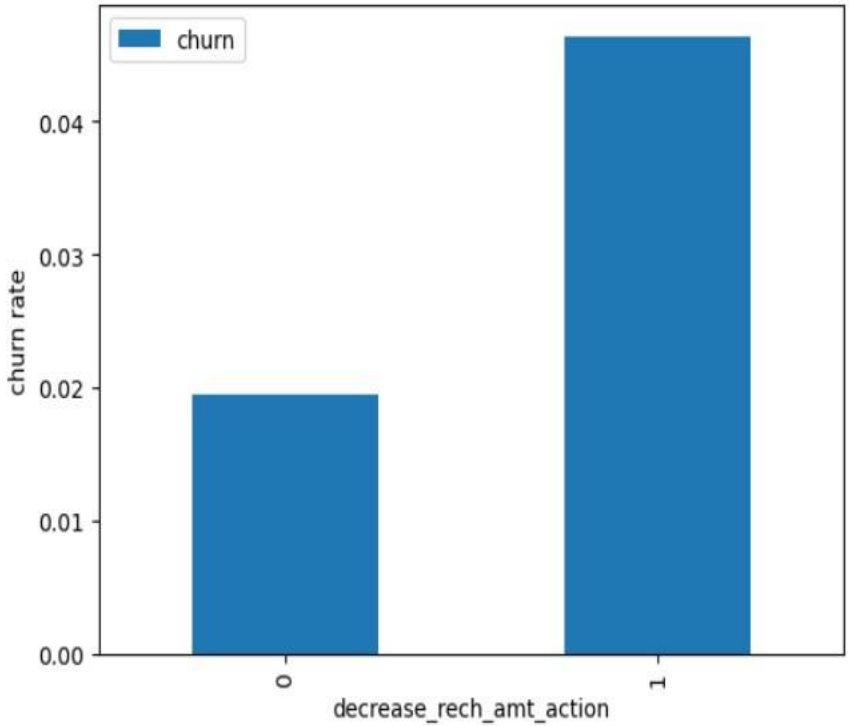
Churn rate on the basis whether the customer decreased her/his number of recharge in action month

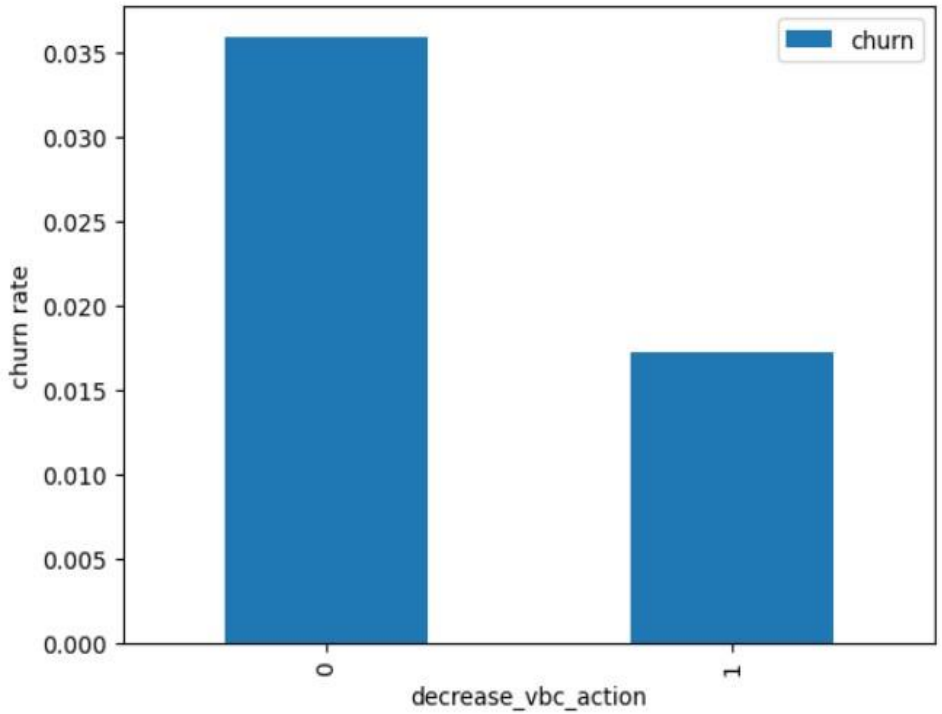


Analysis

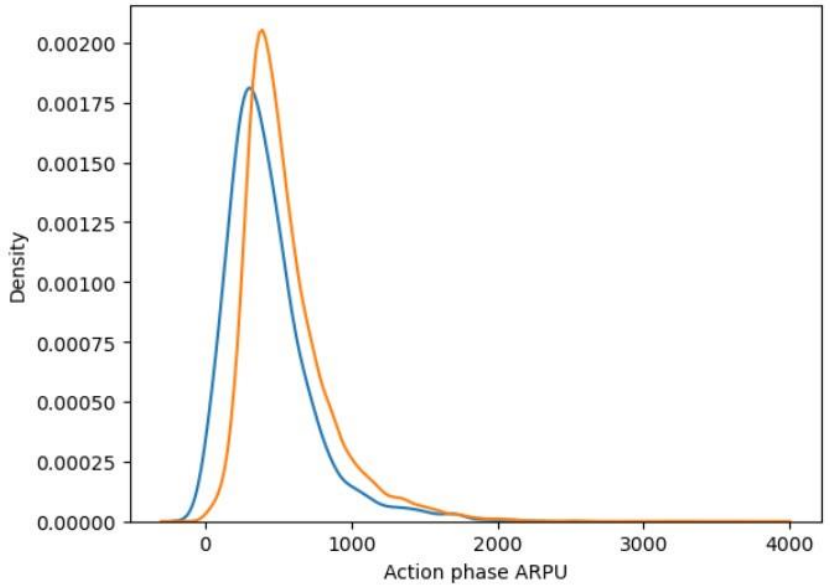
As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

Churn rate on the basis whether the customer decreased her/his amount of recharge in action month



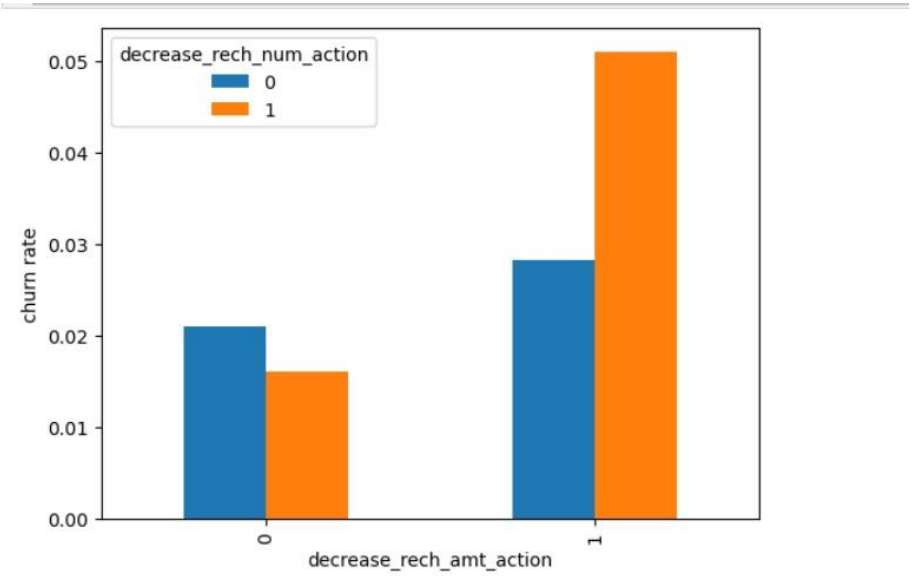


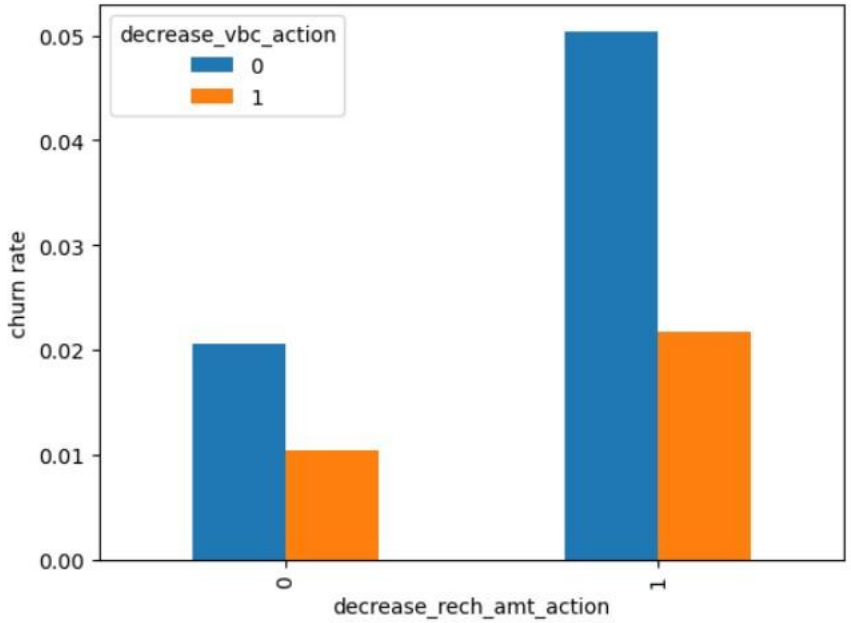
Analysis:
Here we see the expected result. The churn rate is more for the customers, whose volume-based cost in action month is increased. That means the customers do not do the monthly recharge more when they are in the action phase. Analysis of the average revenue per customer (churn and not churn) in the action phase.

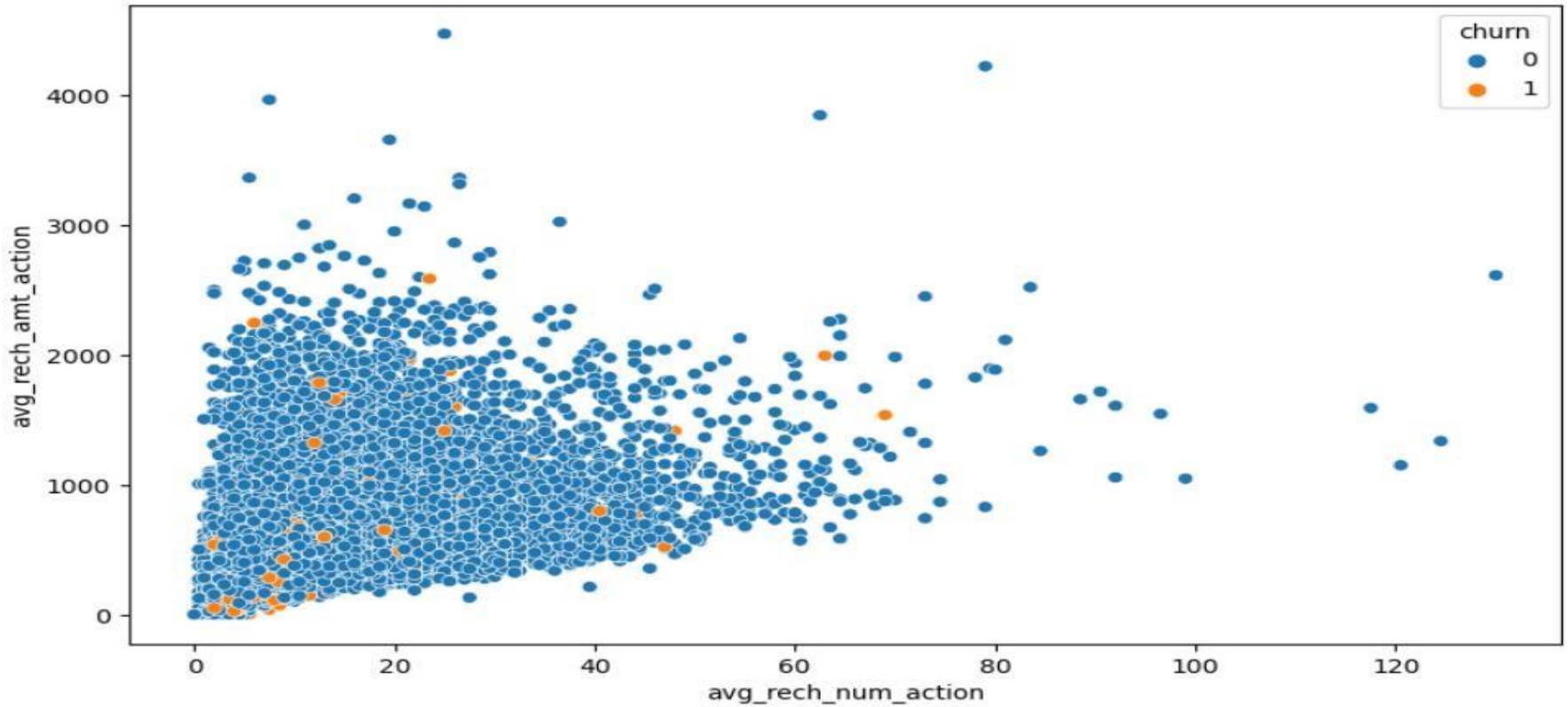


- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.
- ARPU for the not churned customers is mostly densed on the 0 to 1000.
- Analysis of the minutes of usage MOU (churn and not churn) in the action phase

BI-VARIATE ANALYSIS



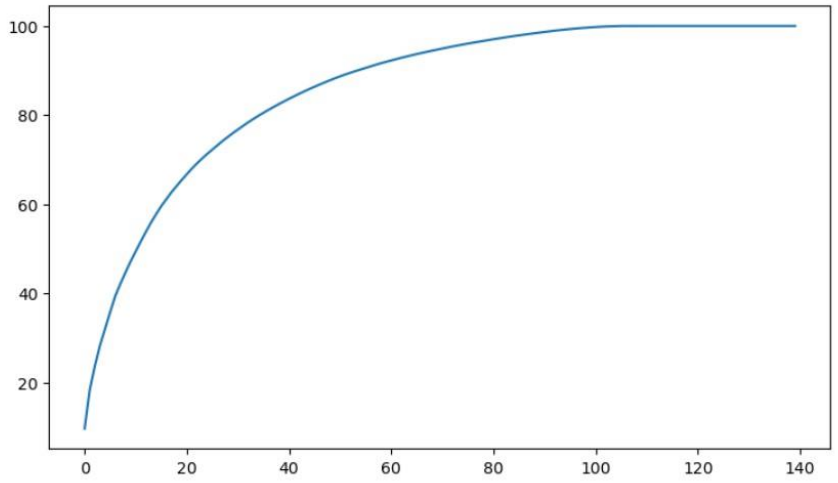



churn
0
1
avg_rech_amt_action
4000
3000
2000
1000
0
0
20
40
60
80
100
120
avg_rech_num_action

Analysis

We can see from the above pattern that
the recharge number and the recharge amount are mostly proportional.
More the number of recharge, more the amount of the recharge.

MODELLING – LOGISTIC REGRESSION WITH PCA



Sensitivity: 0.33
Specificity: 0.99
AUC: 0.97

Poor sensitivity. The best model is PCA along with Logistic regression.

AUC for Logistics regression

```
# print best hyperparameters
print("Best AUC: ", model.best_score_)
print("Best hyperparameters: ", model.best_params_)
```

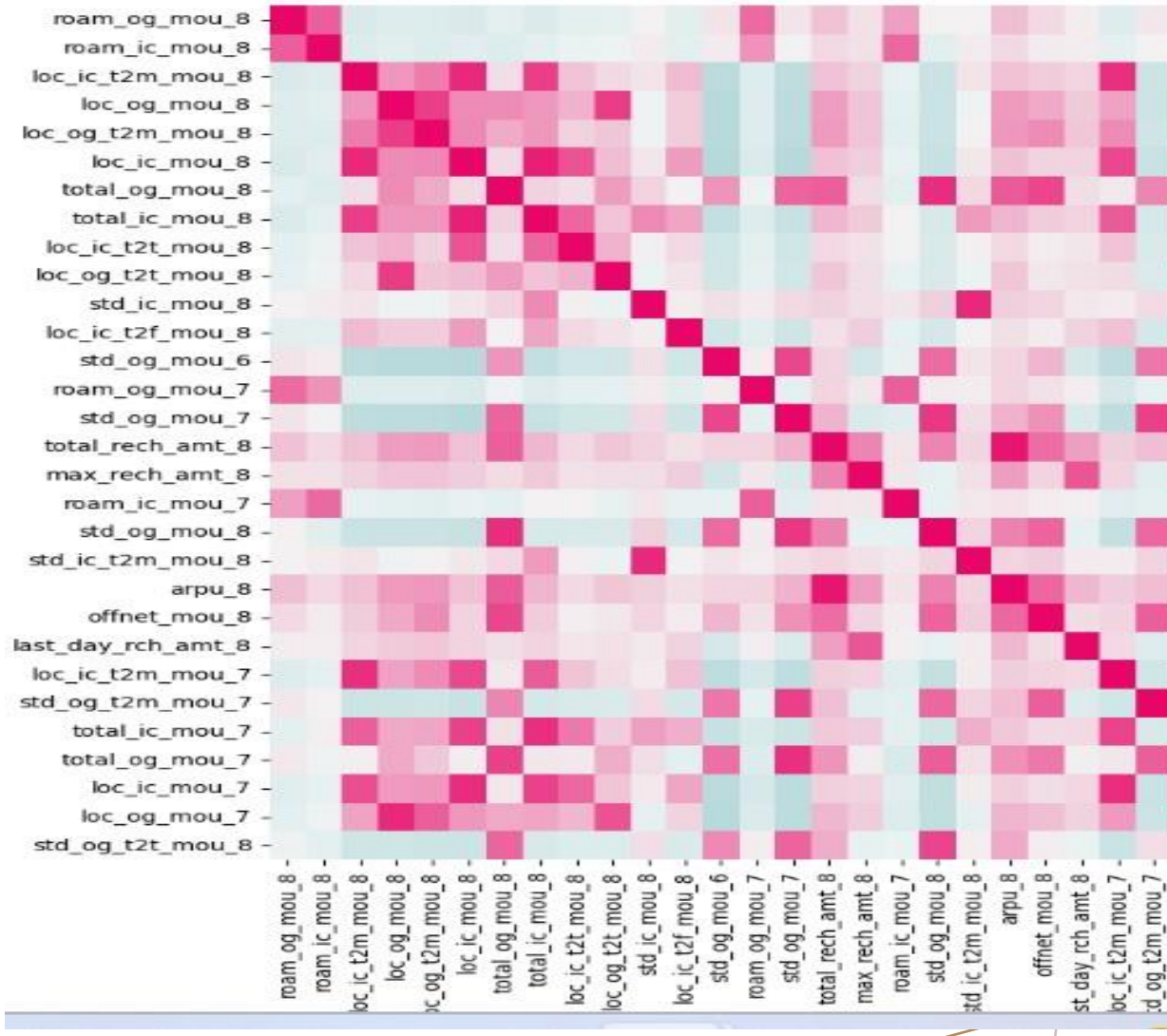
```
Best AUC:  0.89131004098945O8
Best hyperparameters:  {'logistic__C': 0.1, 'logistic__penalty': 'l2', 'pca__n_components': 80}
```

AUC for Random Forest

```
# print best hyperparameters
print("Best AUC: ", model.best_score_)
print("Best hyperparameters: ", model.best_params_)
```

```
Best AUC:  0.9590258118140058
Best hyperparameters:  {'criterion': 'entropy', 'max_features': 0.4}
```

Extraction of 30 features


roam_og_mou_8
roam_ic_mou_8
loc_ic_t2m_mou_8
loc_og_mou_8
loc_og_t2m_mou_8
loc_ic_mou_8
total_og_mou_8
total_ic_mou_8
loc_ic_t2t_mou_8
loc_og_t2t_mou_8
std_ic_mou_8
loc_ic_t2f_mou_8
std_og_mou_6
roam_og_mou_7
std_og_mou_7
total_rech_amt_8
max_rech_amt_8
roam_ic_mou_7
std_og_mou_8
std_ic_t2m_mou_8
arpu_8
offnet_mou_8
last_day_rch_amt_8
loc_ic_t2m_mou_7
std_og_t2m_mou_7
total_ic_mou_7
total_og_mou_7
loc_ic_mou_7
loc_og_mou_7
std_og_t2t_mou_8

CONCLUSION

- Telecom company needs to pay attention to the roaming rates.
- They need to provide good offers to the customers who are using services from a roaming zone. The company needs to focus on the STD and ISD rates.
- Perhaps, the rates are too high.
- Provide them with some kind of STD and ISD packages.
- To look into both issues stated above, it is desired that the telecom company collects customer query and complaint data and work on their services according to the needs of customers.