Telecom churn – Case Study

Telecom churn – Case Study

Problem Statement

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.

For many incumbent operators, retaining high profitable customers is the number one business goal.

To reduce customer churn, telecom companies need to <u>predict which customers are at high risk of churn</u>.

In this project, you will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Understanding the business objective and the data

The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

Data preparation

The following data preparation steps are crucial for this problem:

1. **Filter high-value customers**

As mentioned above, you need to predict churn only for high-value customers. Define high-value customers as follows: Those who have recharged with an amount more than or equal to X, where X is the **70th percentile** of the average recharge amount in the first two months (the good phase).

After filtering the high-value customers, you should get about 30k rows.

2. **Tag churners and remove attributes of the churn phase**

Now tag the churned customers (churn=1, else 0) based on the fourth month as follows: Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase. The attributes you need to use to tag churners are:

total_ic_mou_9

total_og_mou_9

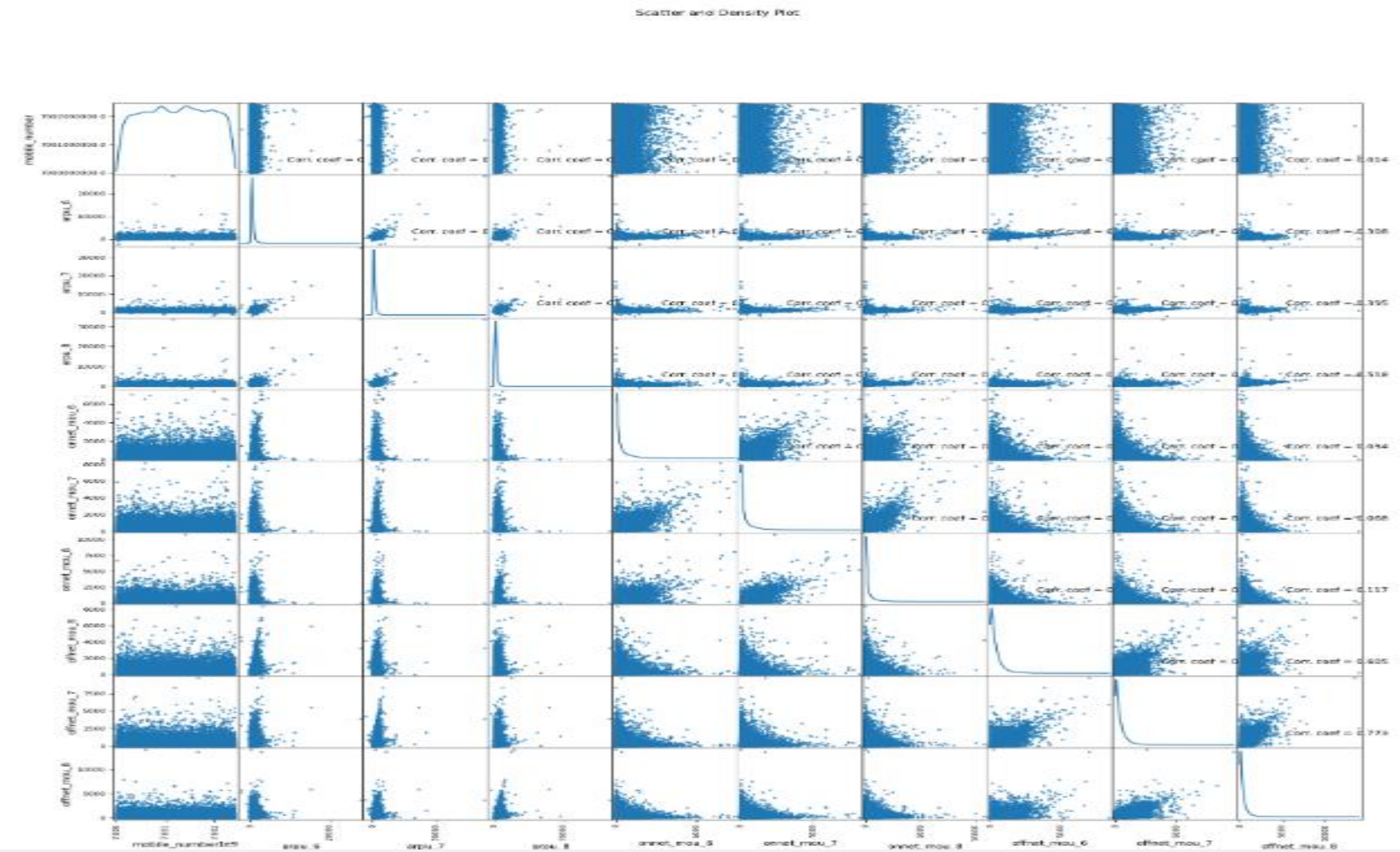
vol_2g_mb_9

vol_3g_mb_9

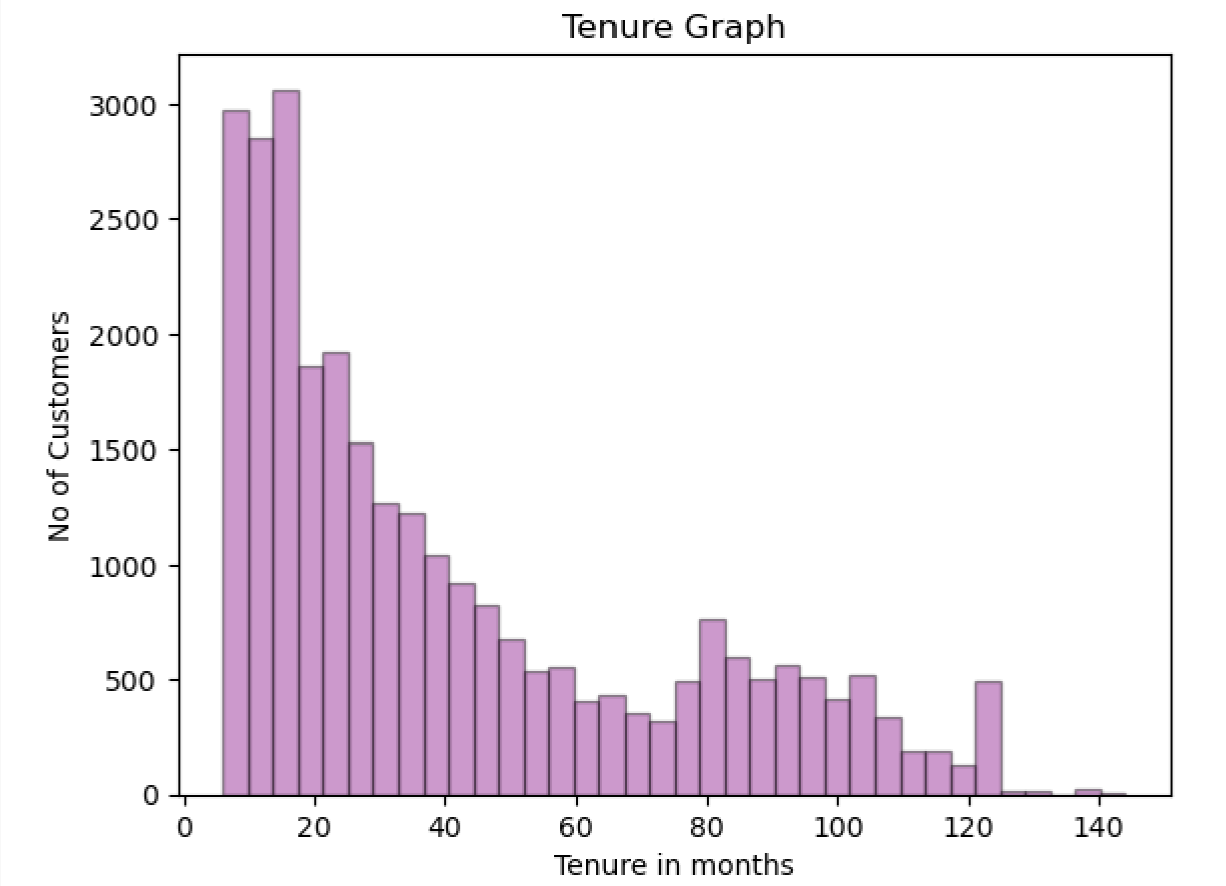
After tagging churners, **remove all the attributes corresponding to the churn phase** (all attributes having ‘ _9’, etc. in their names).

Telecom churn – Case Study

Exploratory Data Analysis



Exploratory Data Analysis



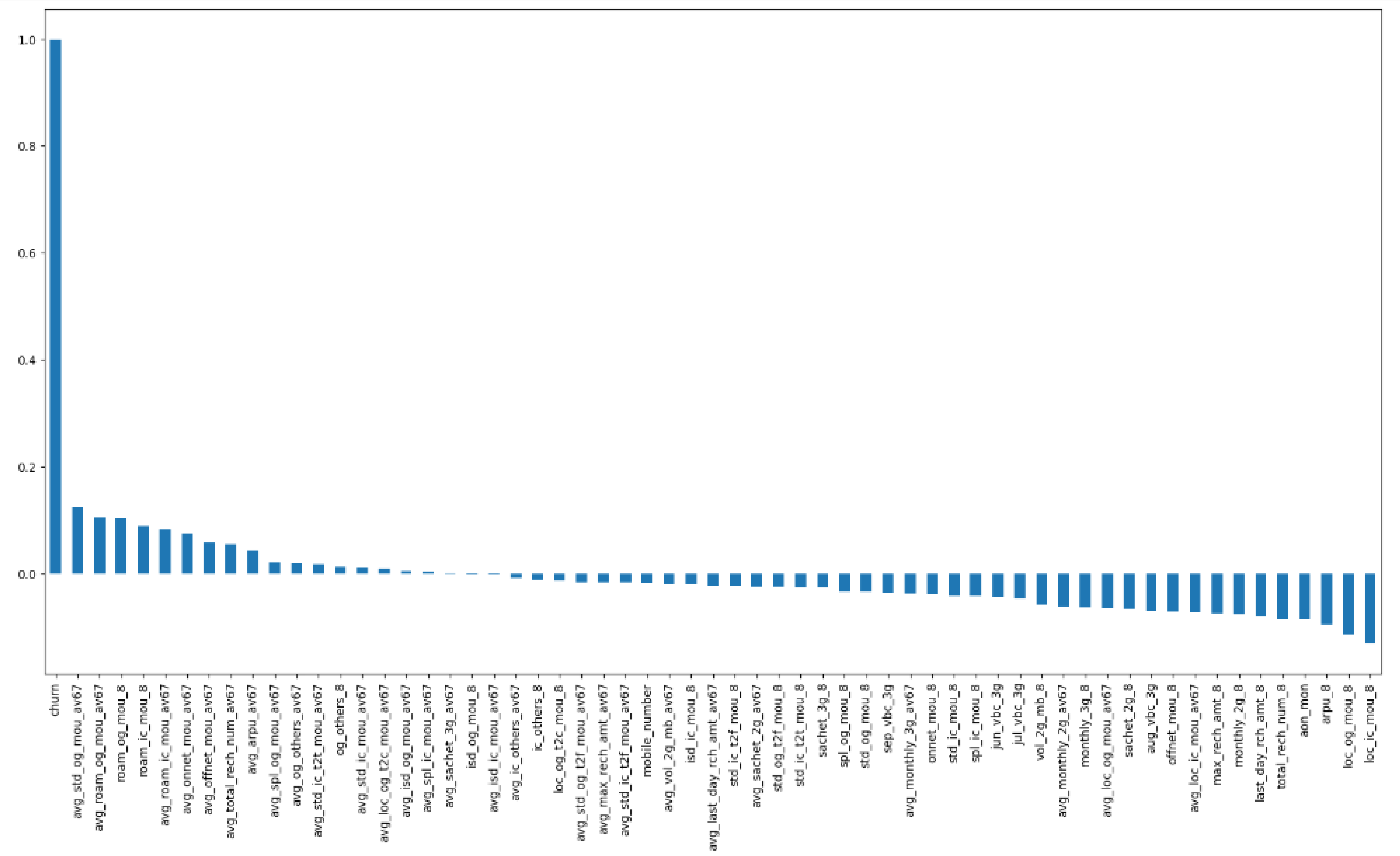
This graph simply shows the tenure of the customers

T el ecom ch urn – Case Study

Exploratory Data Analysis

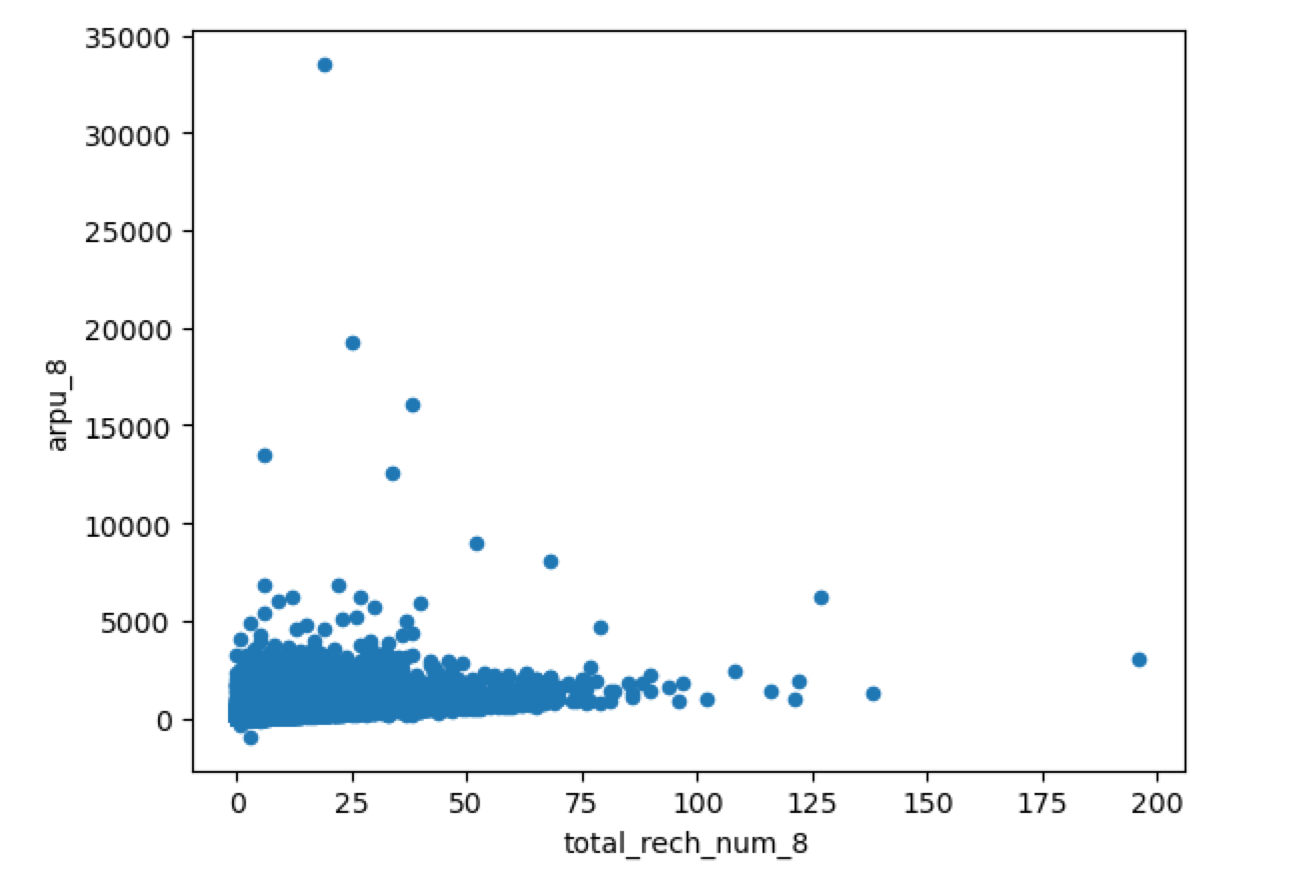
Observation:

1. Avg Outgoing Calls & calls on romaning for 6 & 7th months are positively correlated with churn.
2. Avg Revenue, No. Of Recharge for 8th month has negative correlation with churn.



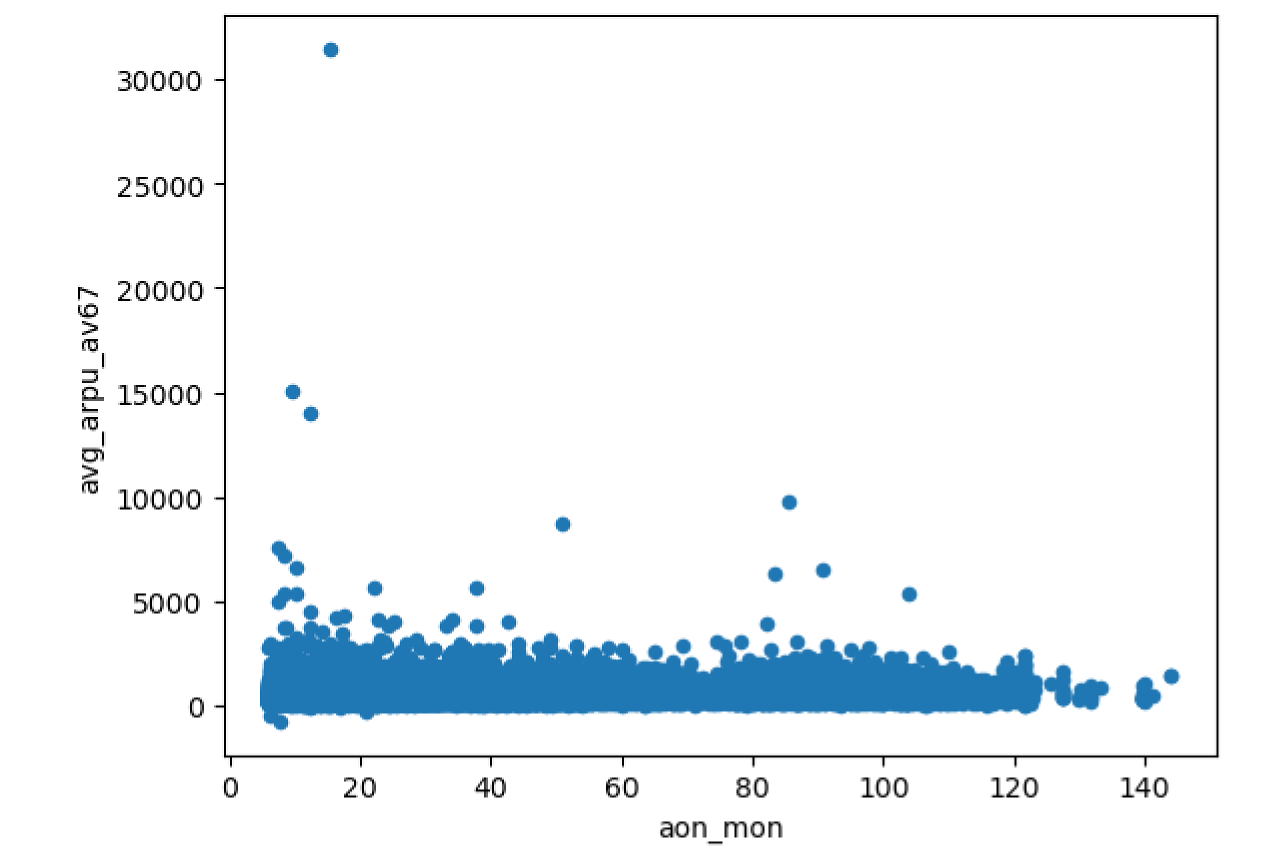
Exploratory Data Analysis

Plot between total recharge and avg revenue for the 8th month



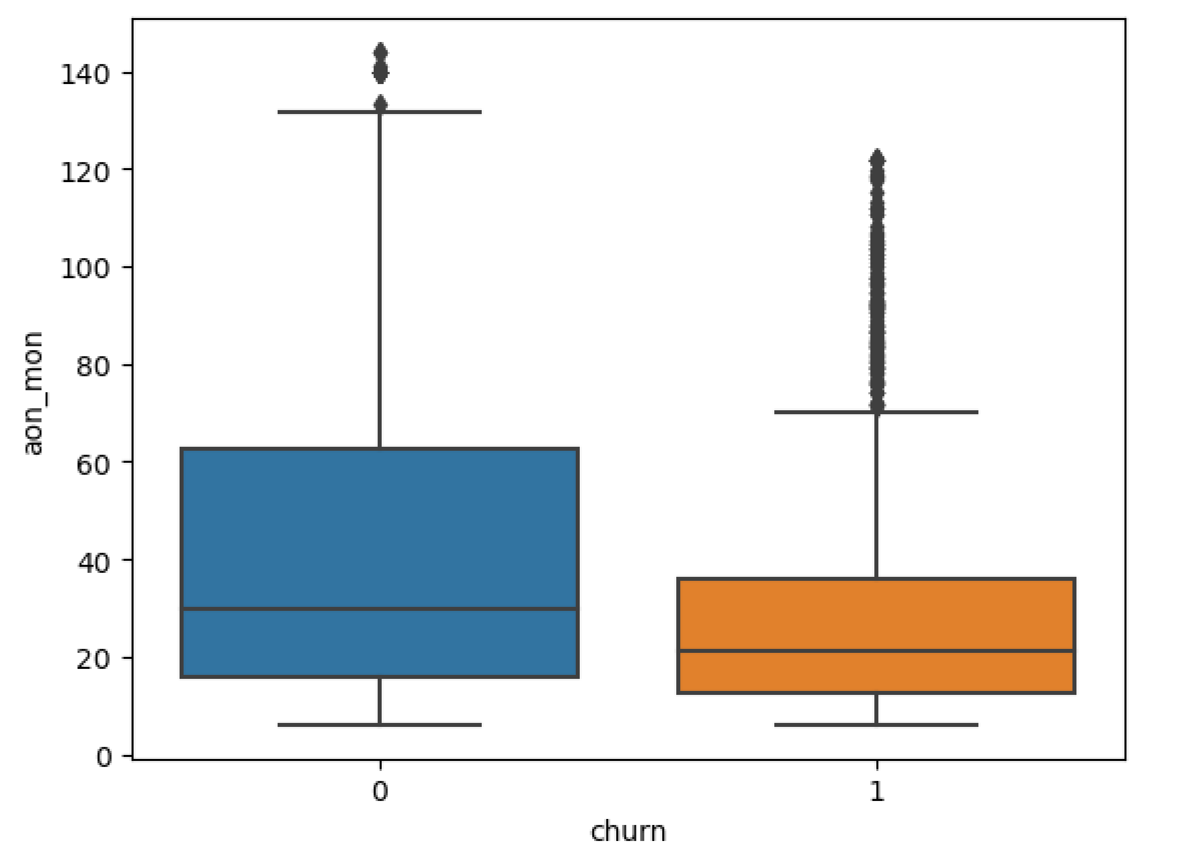
Exploratory Data Analysis

Plot between tenure and revenue

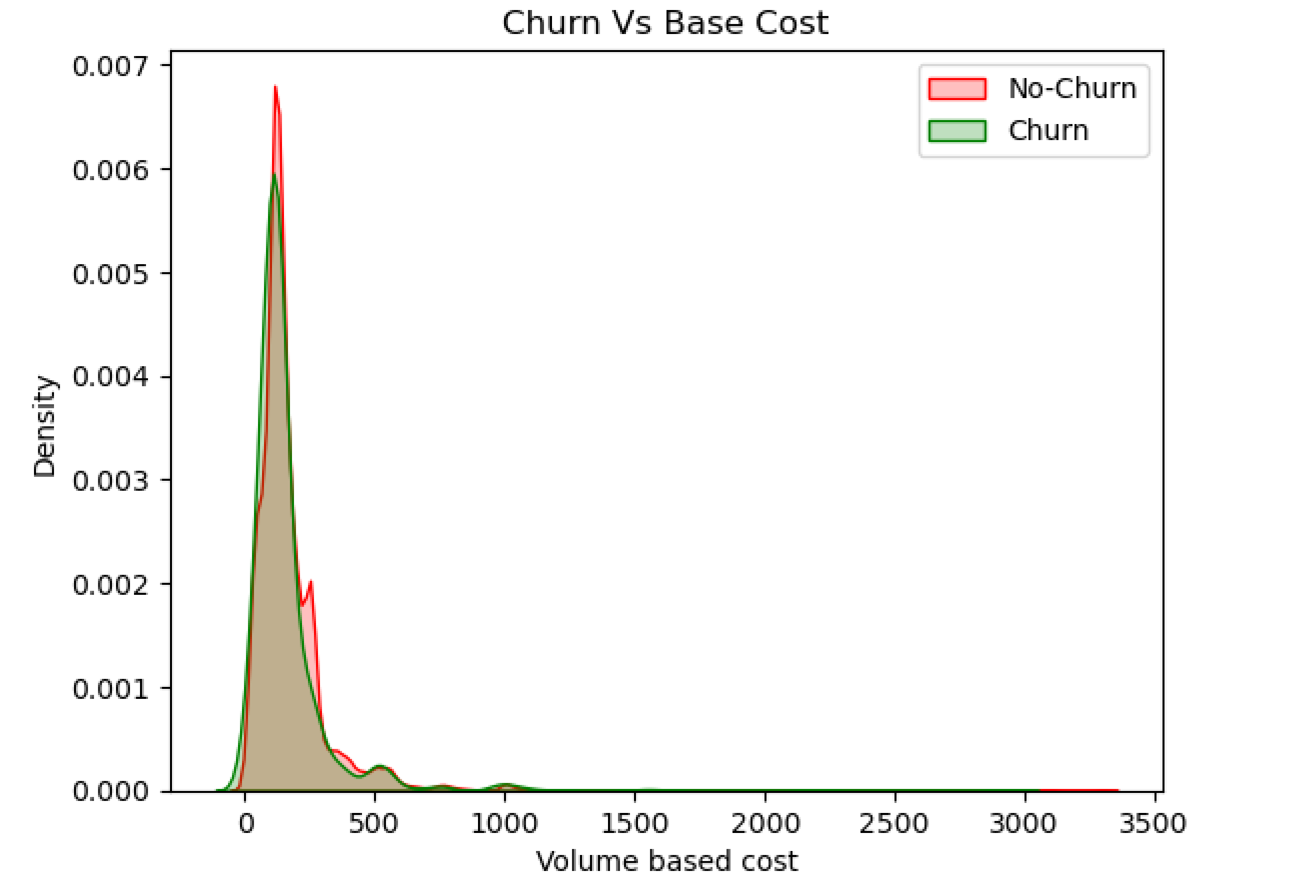


Exploratory Data Analysis

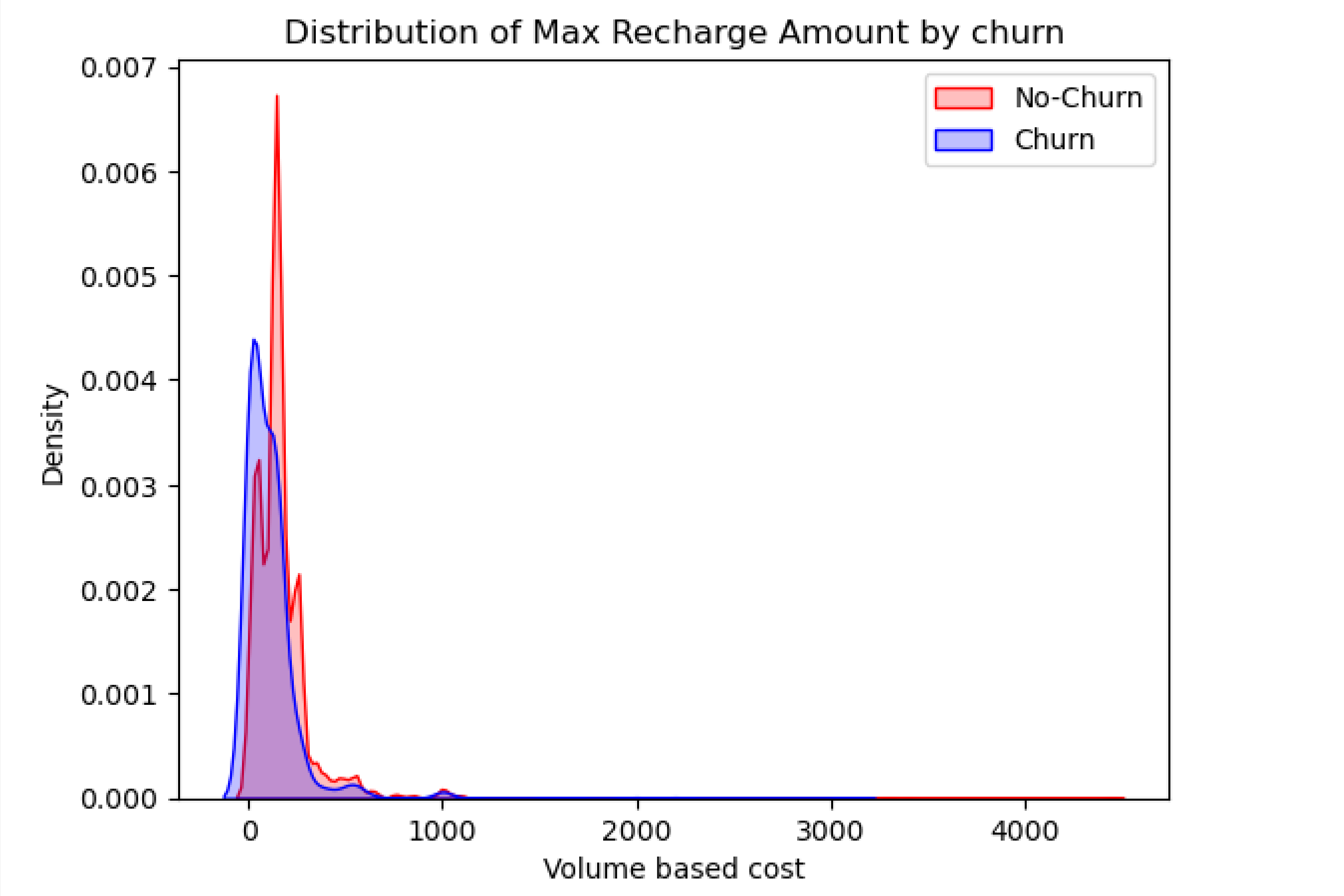
From this plot, its clear tenured customers do no churn and they keep availing telecom services



Exploratory Data Analysis



Exploratory Data Analysis



Conclusion

Conclusions from Random Forest

1. Local Incoming for Month 8, Average Revenue Per Customer for Month 8 and Max Recharge Amount for Month 8 are the most important predictor variables to predict churn.

Link code.

Overall Conclusions

1. Std Outgoing Calls and Revenue Per Customer are strong indicators of Churn.
2. Local Incoming and Outgoing Calls for 8th Month and avg. revenue in 8th Month are the most important columns to predict churn.
3. customers with tenure less than 4 years are more likely to churn.
4. Max Recharge Amount is a strong feature to predict churn.
5. Random Forest produced the best prediction results followed by SVM.