Telecom Churn Case Study

Satishkumar Sersia(ssersia@gmail.com)
Boosa Madhan Mohan Reddy(mohan.madhan14@gmail.com)
Yaswanth(yaswantsaigampa@gmail.com)

Business problem

- 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*.

Objective

- To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

- 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.

Definitions of churn

- **Revenue-based churn**: Customers who have not utilised any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as ‘customers who have generated less than INR 4 per month in total/average/median revenue.
- **Usage-based churn**: Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.

- In this project, you will use the **usage-based definition** to define churn.

Understanding customer behaviour during churn

There are **three phases** of the customer lifecycle :

- The ‘good’ phase: In this phase, the customer is happy with the service and behaves as usual.

- The ‘action’ phase: The customer experience starts to sore in this phase, for e.g. he/she gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behaviour than in the ‘good’ months. Also, it is crucial to identify high-churn-risk customers in this phase, since some corrective actions can be taken at this point (such as matching the competitor’s offer/improving the service quality etc.)

- The ‘churn’ phase: In this phase, the customer is said to have churned. You **define churn based on this phase**. Also, it is important to note that at the time of prediction (i.e. the action months), this data is not available to you for prediction. Thus, after tagging churn as 1/0 based on this phase, you discard all data corresponding to this phase.

High-value churn

- In the Indian and Southeast Asian markets, approximately 80% of revenue comes from the top 20% of customers (called high-value customers). Thus, if we can reduce the churn of high-value customers, we will be able to reduce significant revenue leakage.
- 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).

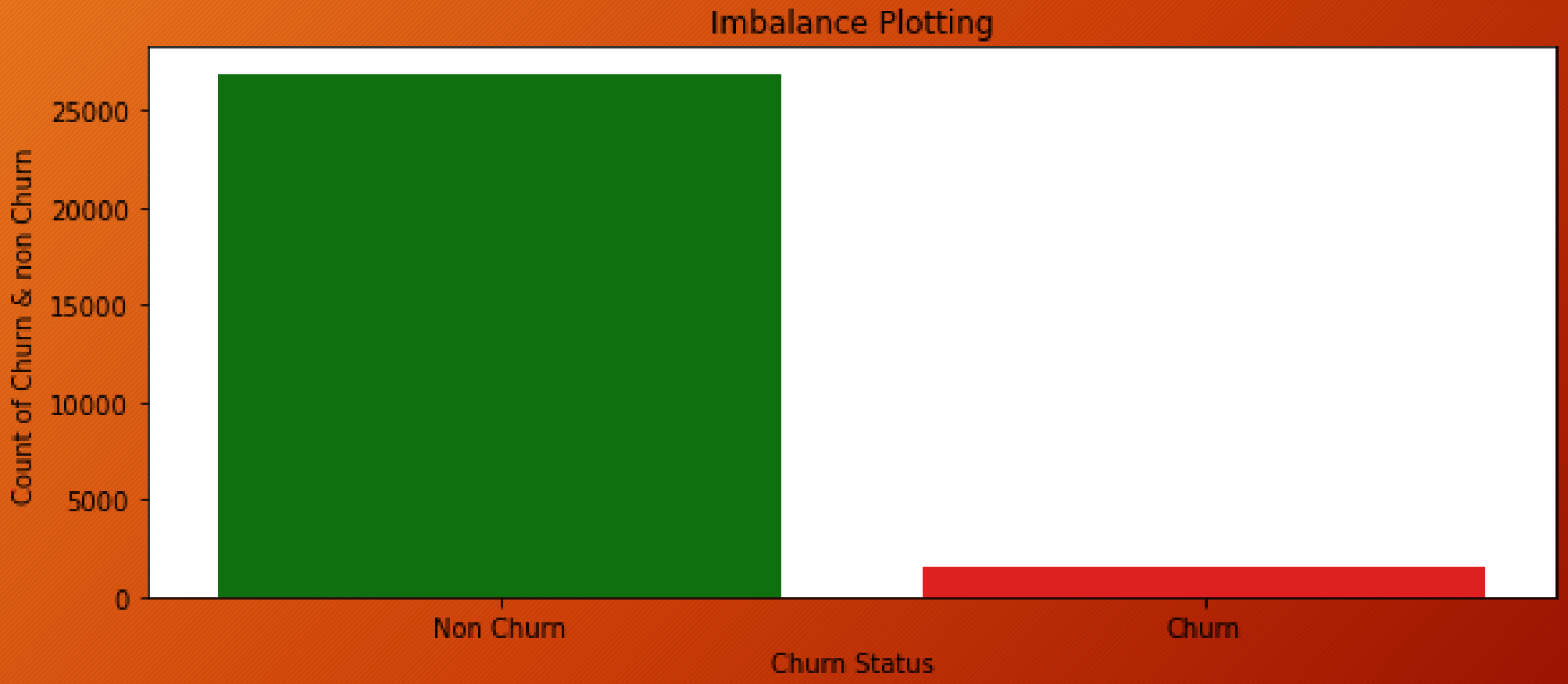
Tag churners

- 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

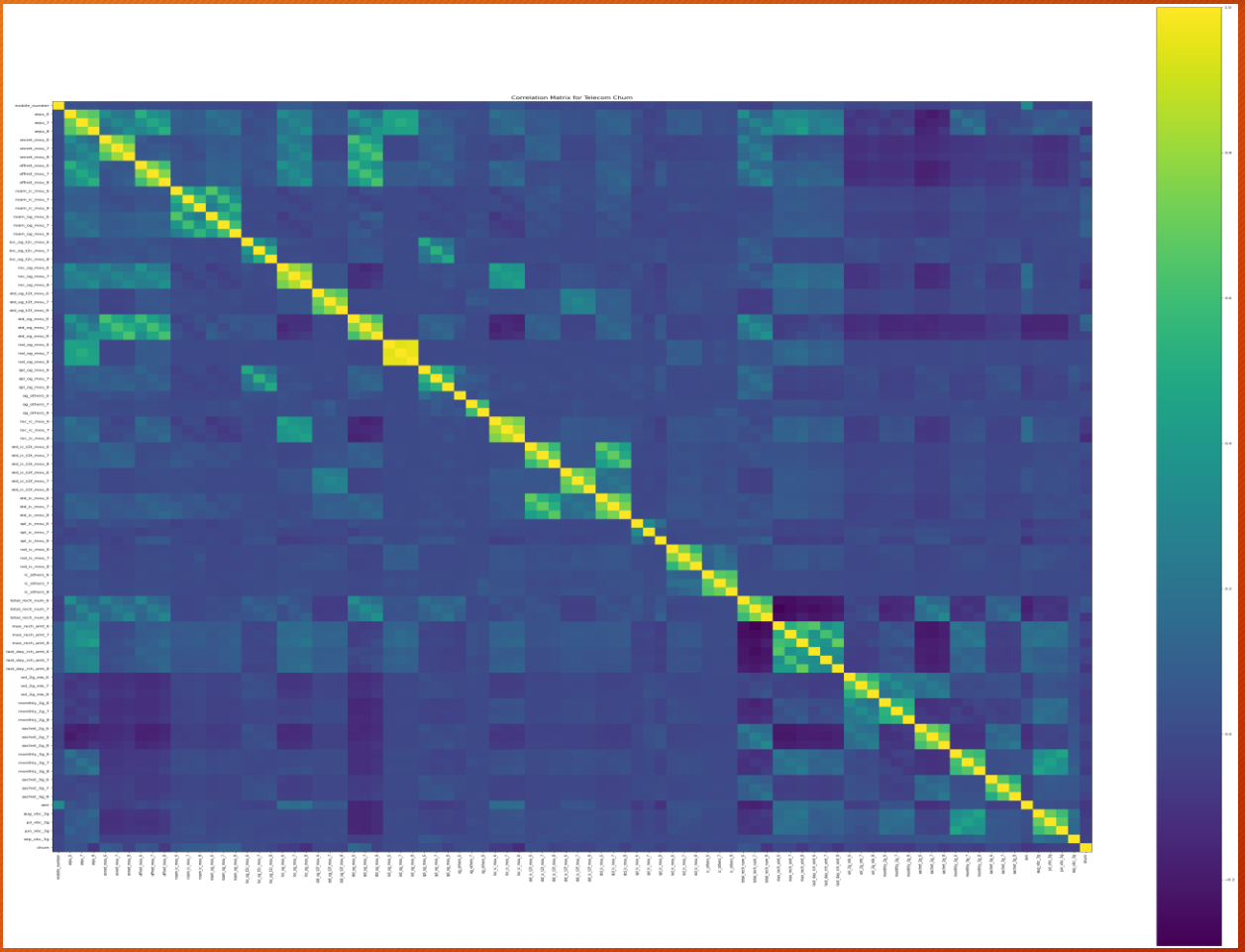
Exploratory Data Analysis

- Data visulaisation using seaborn and matplotlib
- EDA is an approach to analyse data set & to summarize their main characteristics, often with visual methods.
- A Statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis.

Class Imbalance

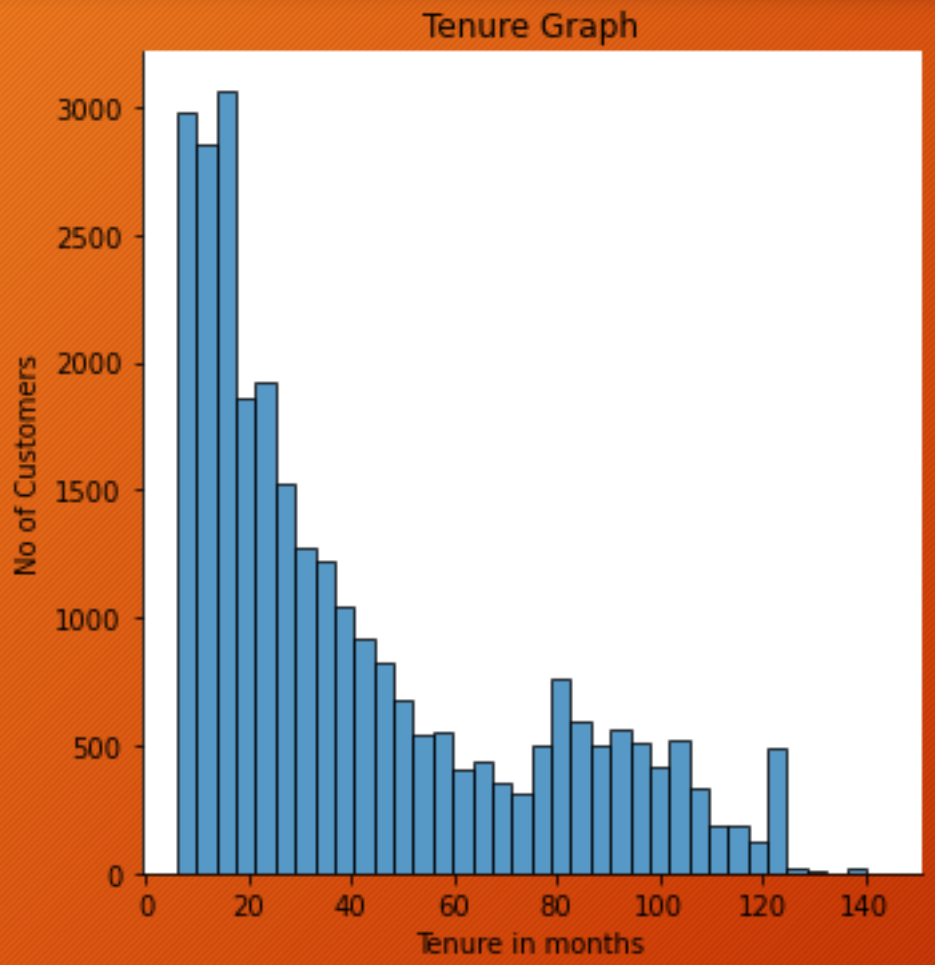


Correlation



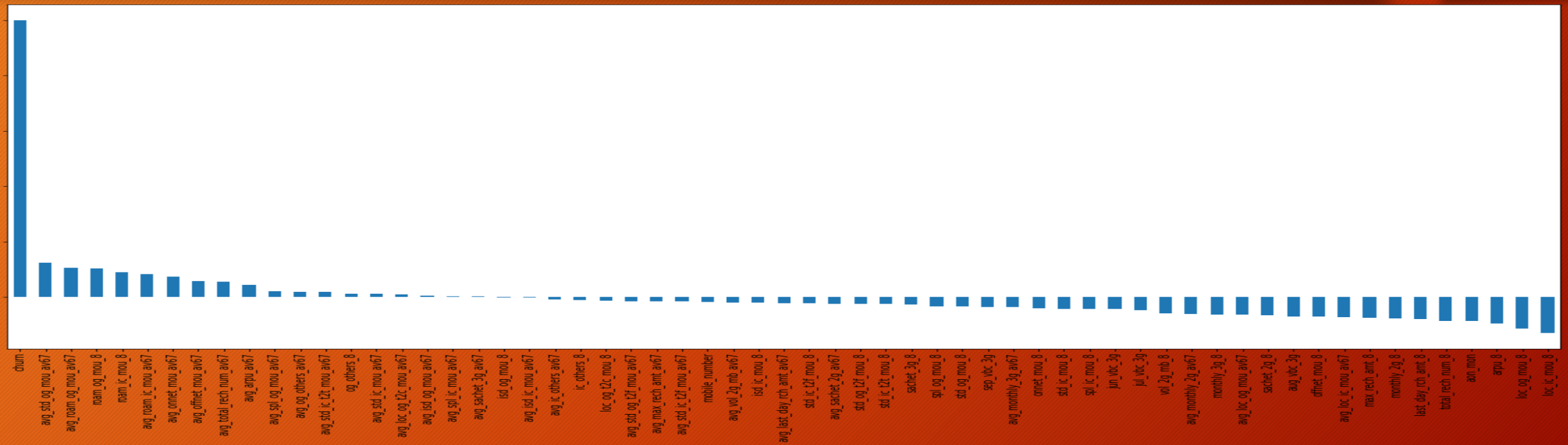
The heatmap on the left shows that there are highly correlated variables in the data

Tenure Vs No Of Customer



The graph show the relationship between Tenure and No of Customers

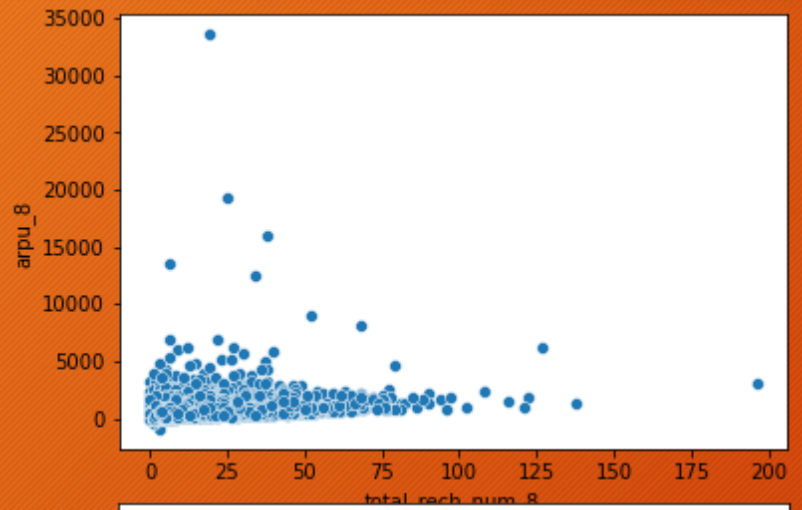
Correlation With Churn

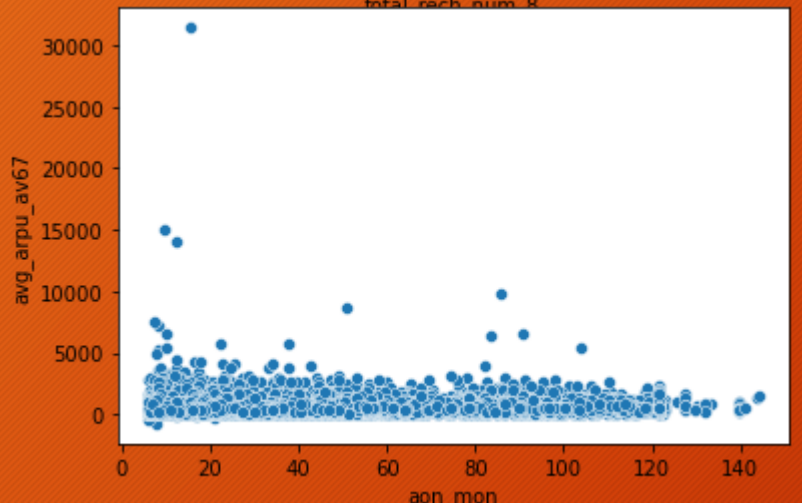


- Observations :

1. Average outgoing calls and outbound calls for the sixth and seventh months have a favourable correlation with churn.
2. Average Revenue and Recharge Number for the Eighth Month Have a Negative Relationship with Churn.

Scatter Plot

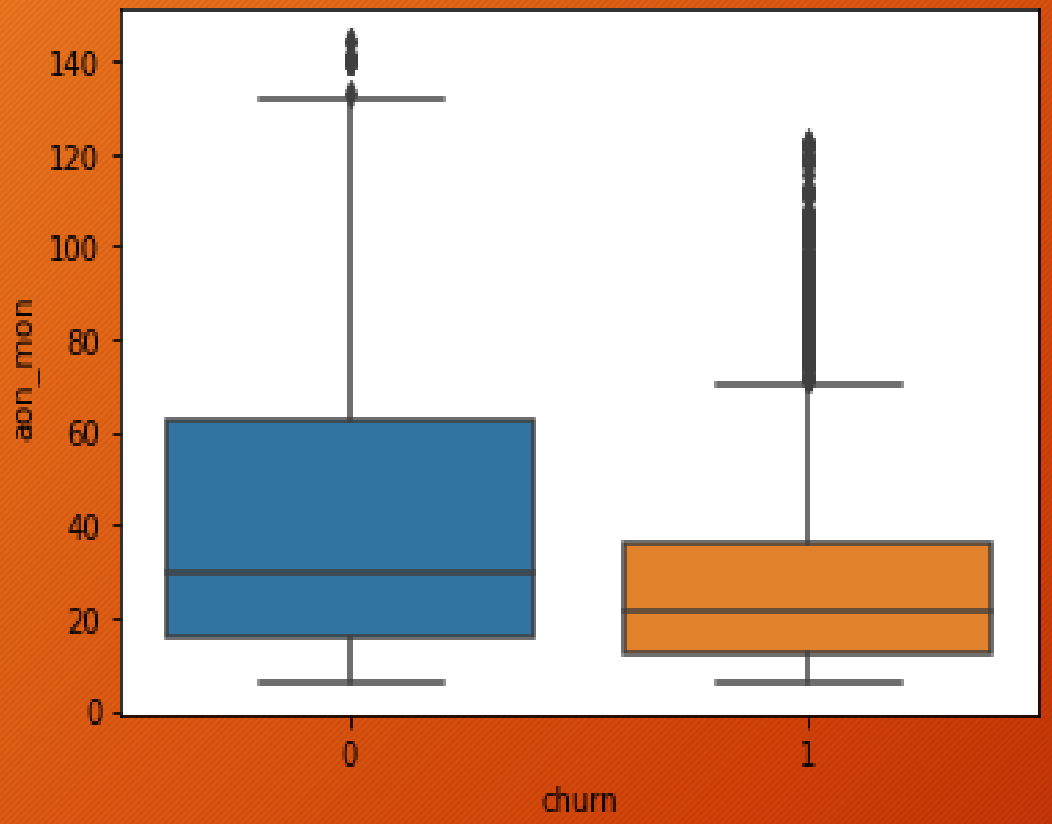




There seems to be no relationship between this variables.

- Average Vs Total recharge for 8 month
- Tenure Vs Revenue

Scatter Plot



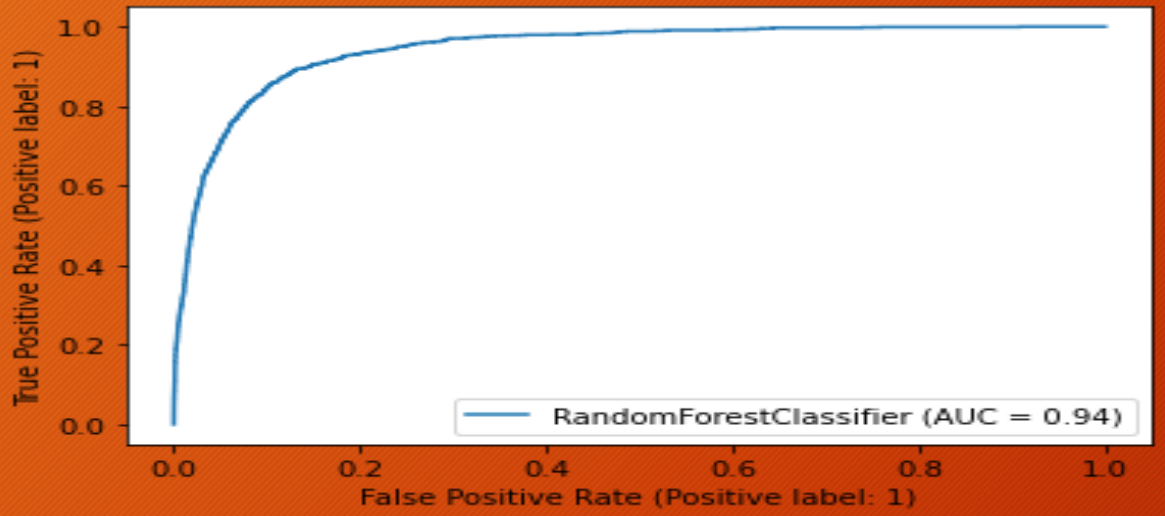
From the plot , its clear tenured customers do not churn and they keep availing telecom services

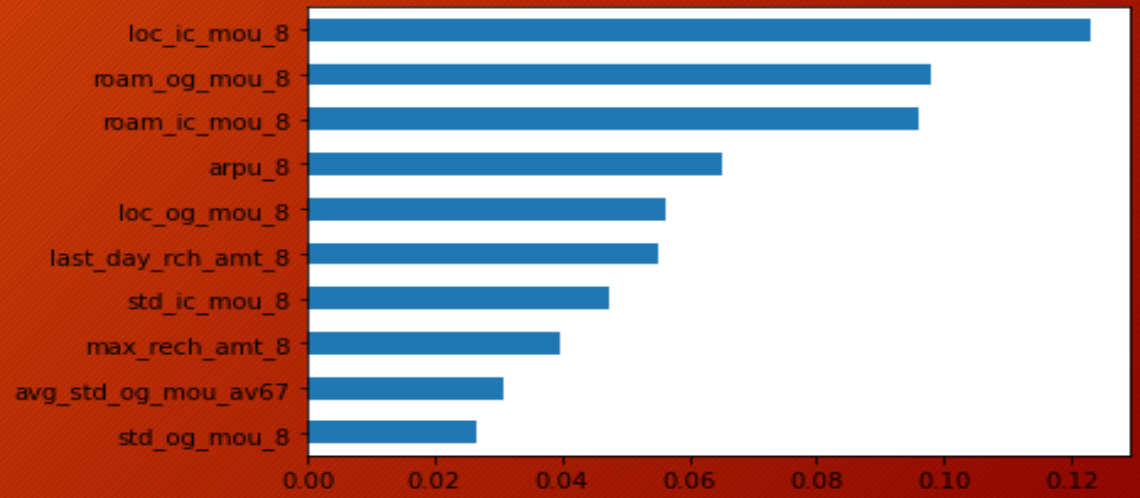
Accuracy Of Various Model

Sr. No	Models	Dimensionality Reduction(PCA)	Accuracy(%) (Test Set)
1	Logistic Regression	No	78
2	Logistic Regression(All Components)	Yes	80
3	Logistic Regression(Top 33 Components)	Yes	80
4	Decision Tree Classifier(Cross Validation)	No	94
5	Random Forest	No	95

Best Model

- As from the Last Slide Random forest is the best Model with 95 % accuracy
- The most significant predictor variables for predicting churn are Local Incoming for Month 8, Average Revenue Per Customer for Month 8, and Max Recharge Amount for Month 8.





Overall Findings

1. Revenue Per Customer and Standard Outgoing Calls are reliable measures of Churn.

2. The most crucial columns for predicting churn are local incoming and outgoing calls for the eighth month and average revenue for the eighth month.

3. Clients are more prone to churn if their duration is under 4 years.

4. Max Recharge Amount is a reliable indicator of churn.