



Telecom Churn Case Study

By - Ayyub / Harshitha / Nanda Kishore



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 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.

Table of contents

01

Business Objective

02

Solution Methodology

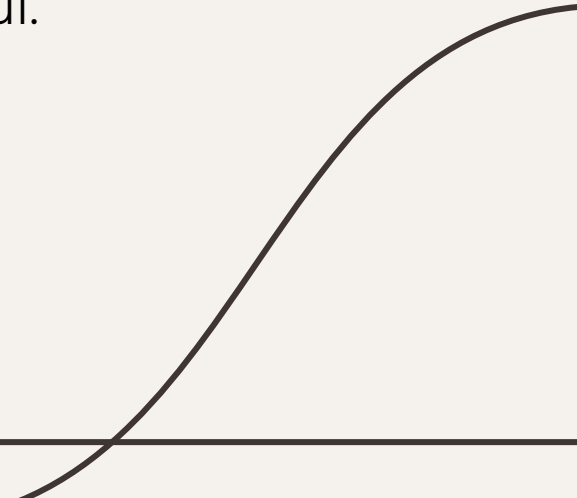
03

Solution Methodology

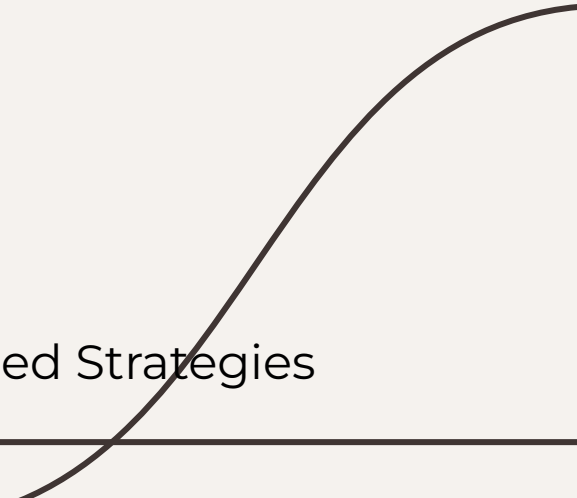
04

Exploratory Data Analysis

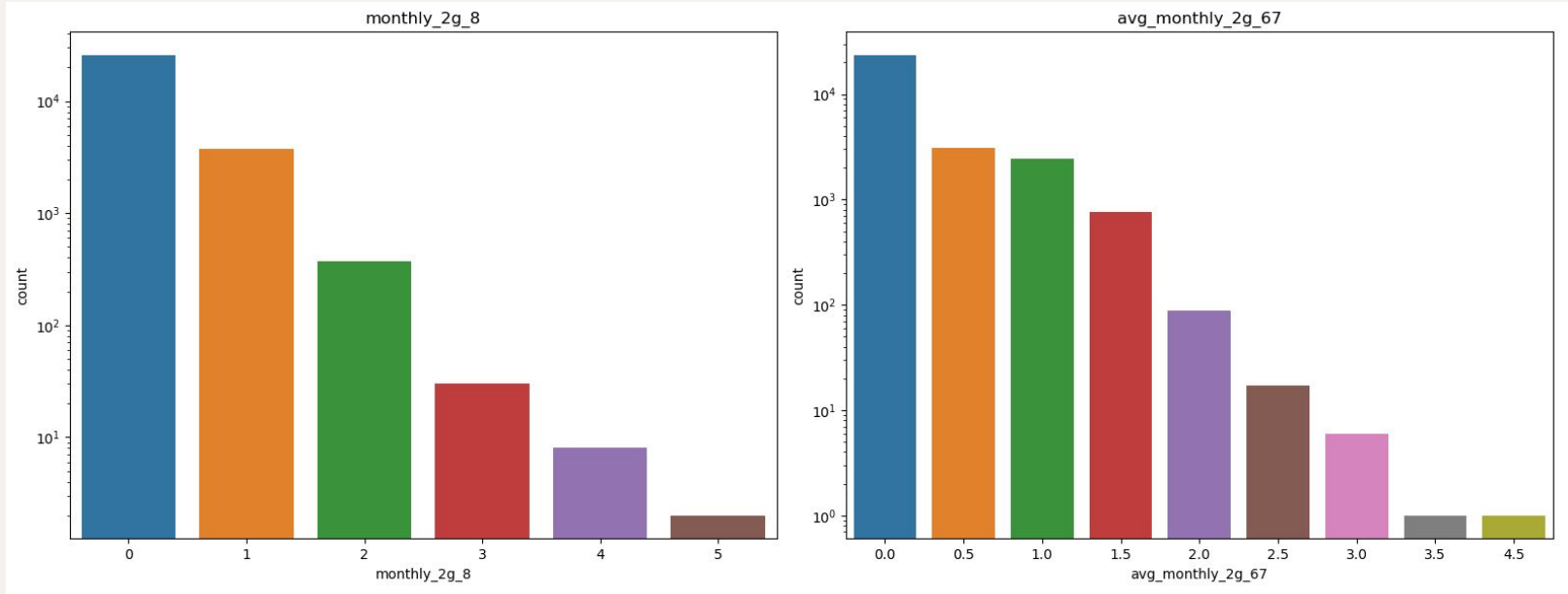
Business Objective

- 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.
- 

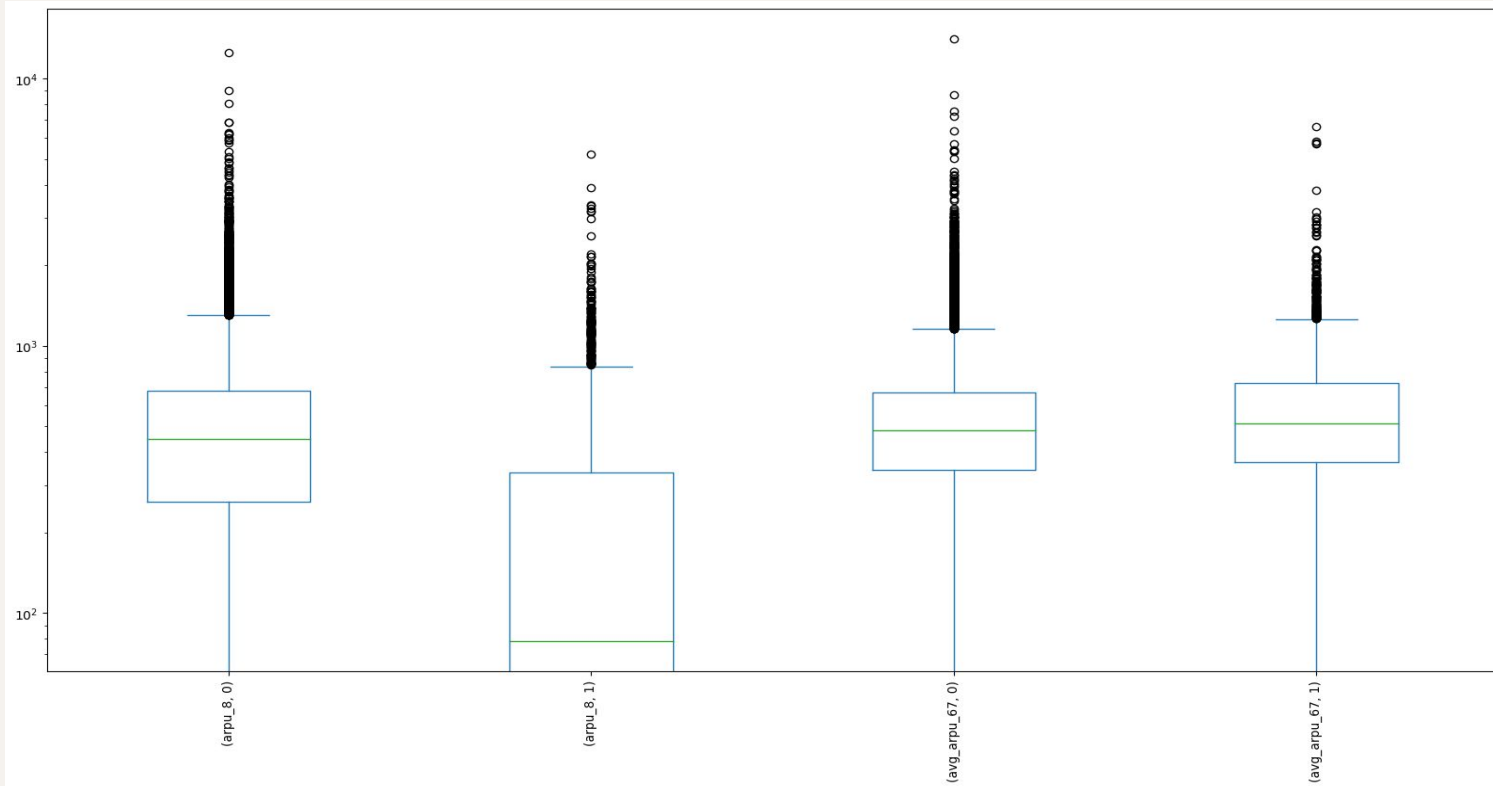
Solution Methodology

- Data cleaning and data manipulation.
 - 1. Check and handle duplicate data.
 - 2. Check and handle NA values and missing values.
 - 3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
 - 4. Imputation of the values, if necessary.
 - 5. Check and handle outliers in data.
 - EDA
 - Model building
 - Model presentation Conclusion & Recommended Strategies
- 

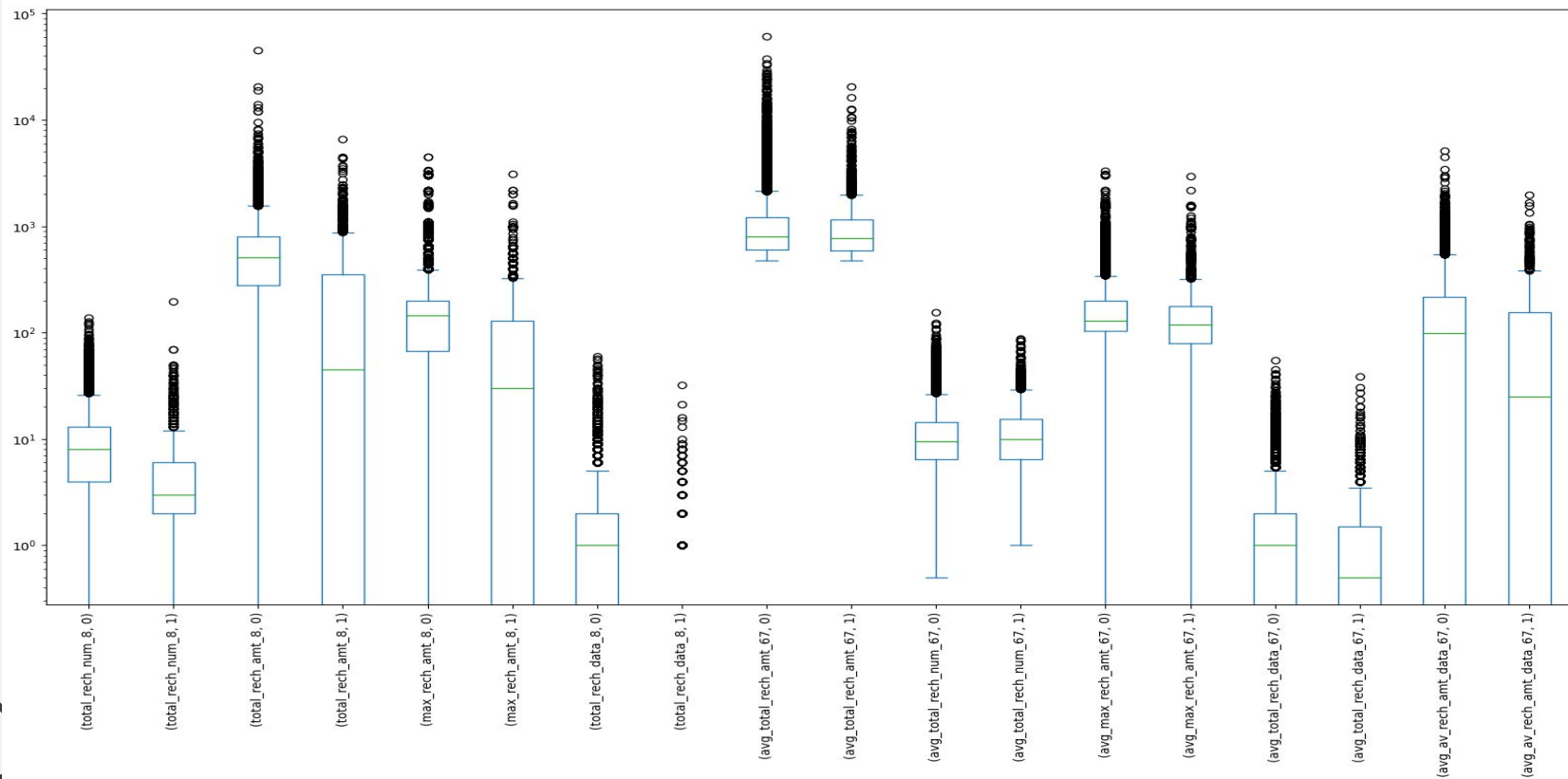
EDA- Bar chart Attributes



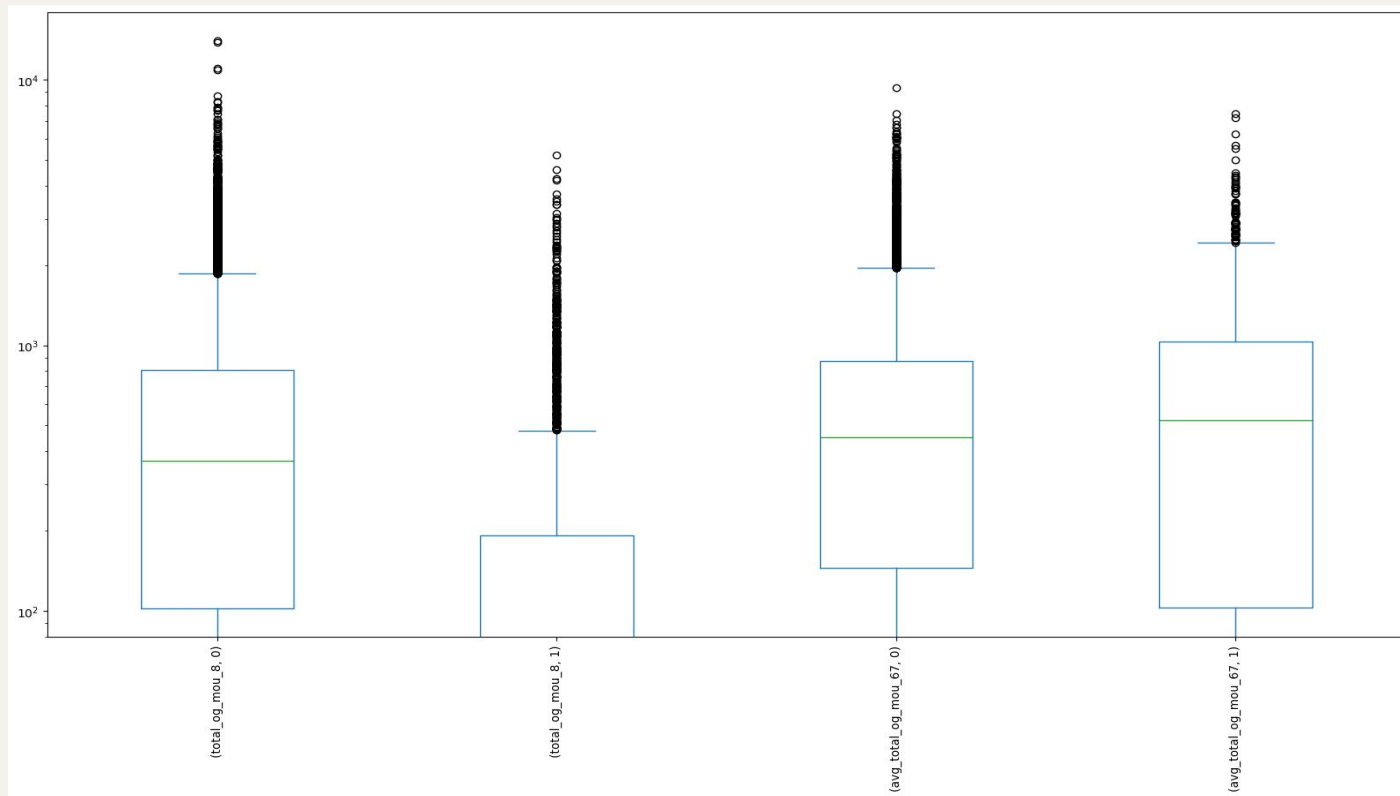
ARPU Distribution



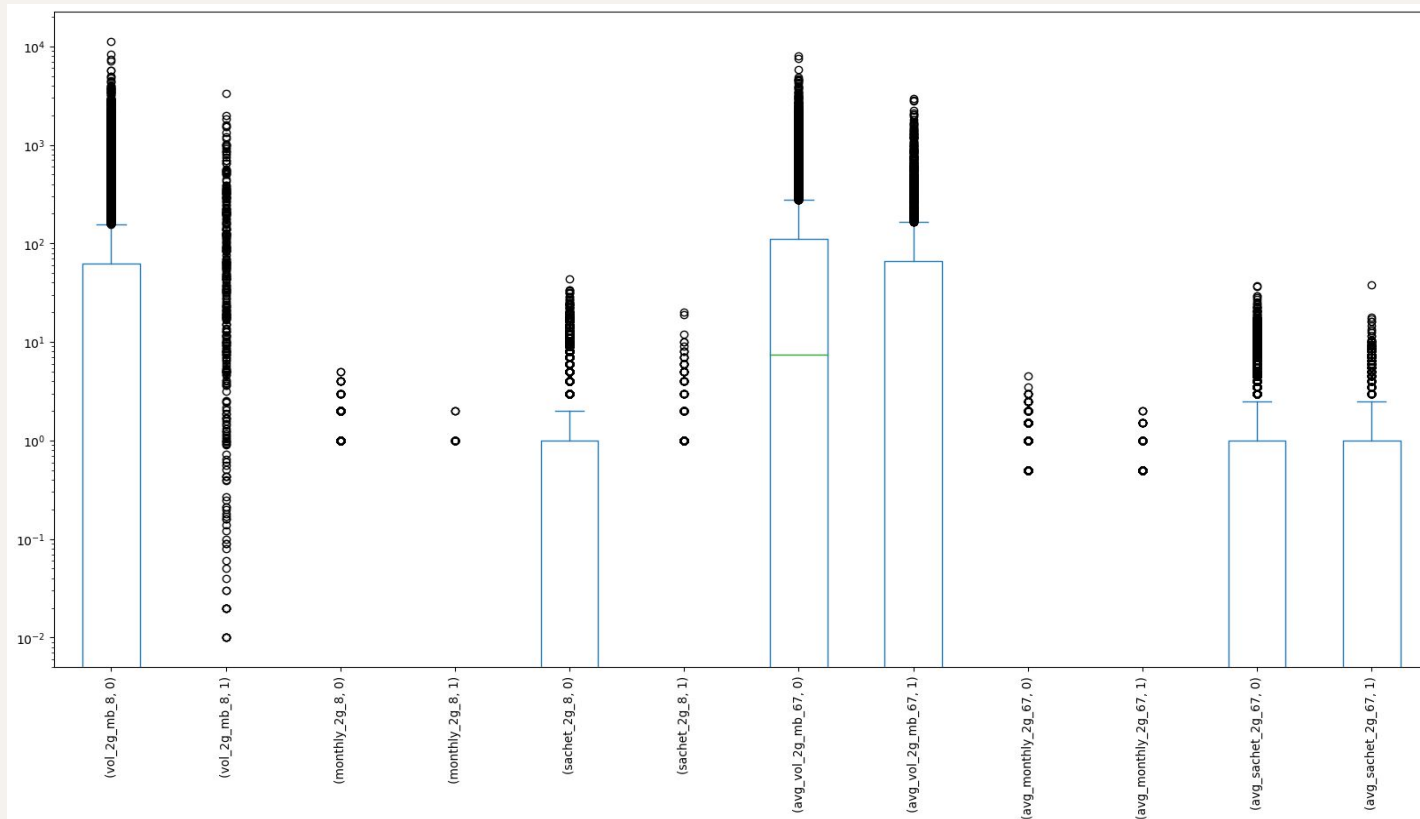
Box Plot



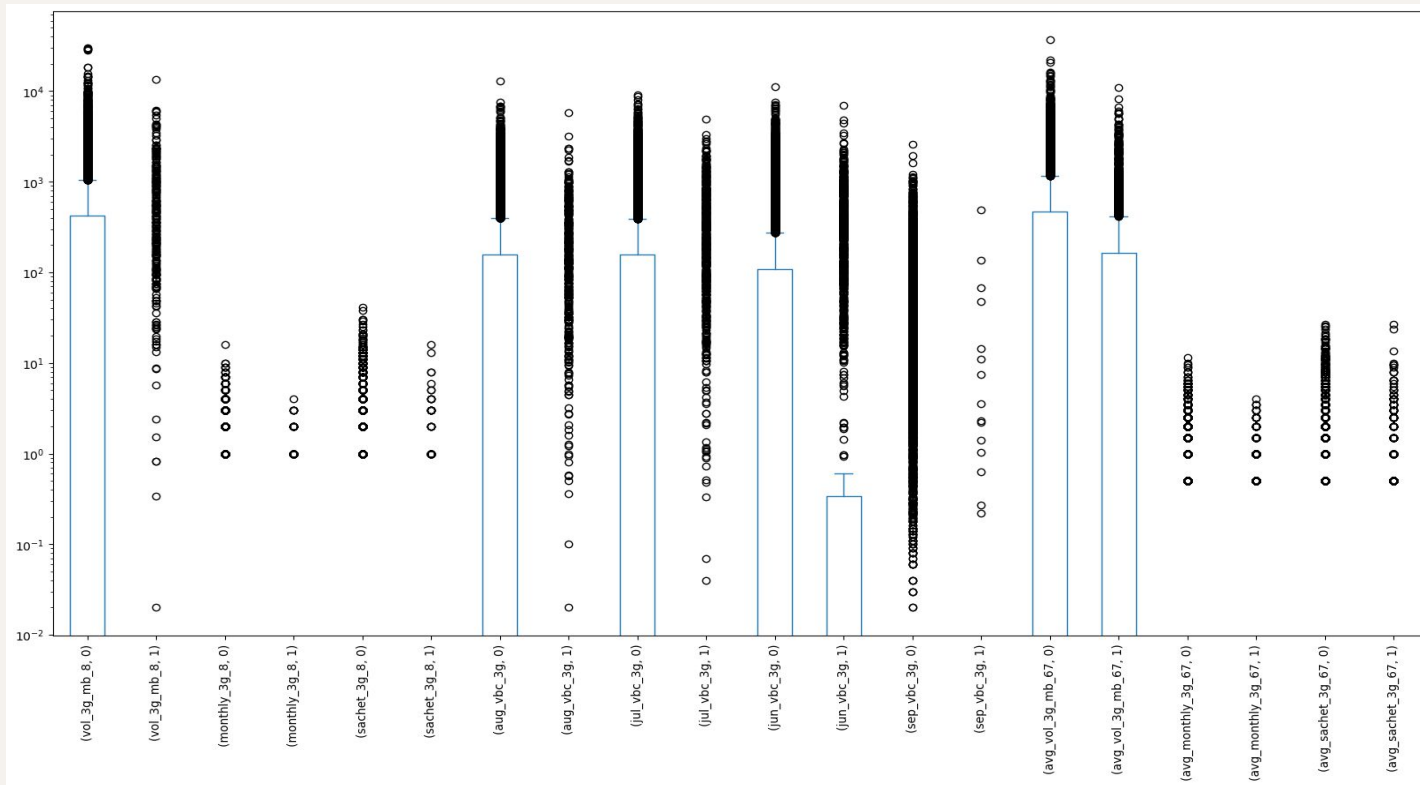
Outgoing calls Distribution



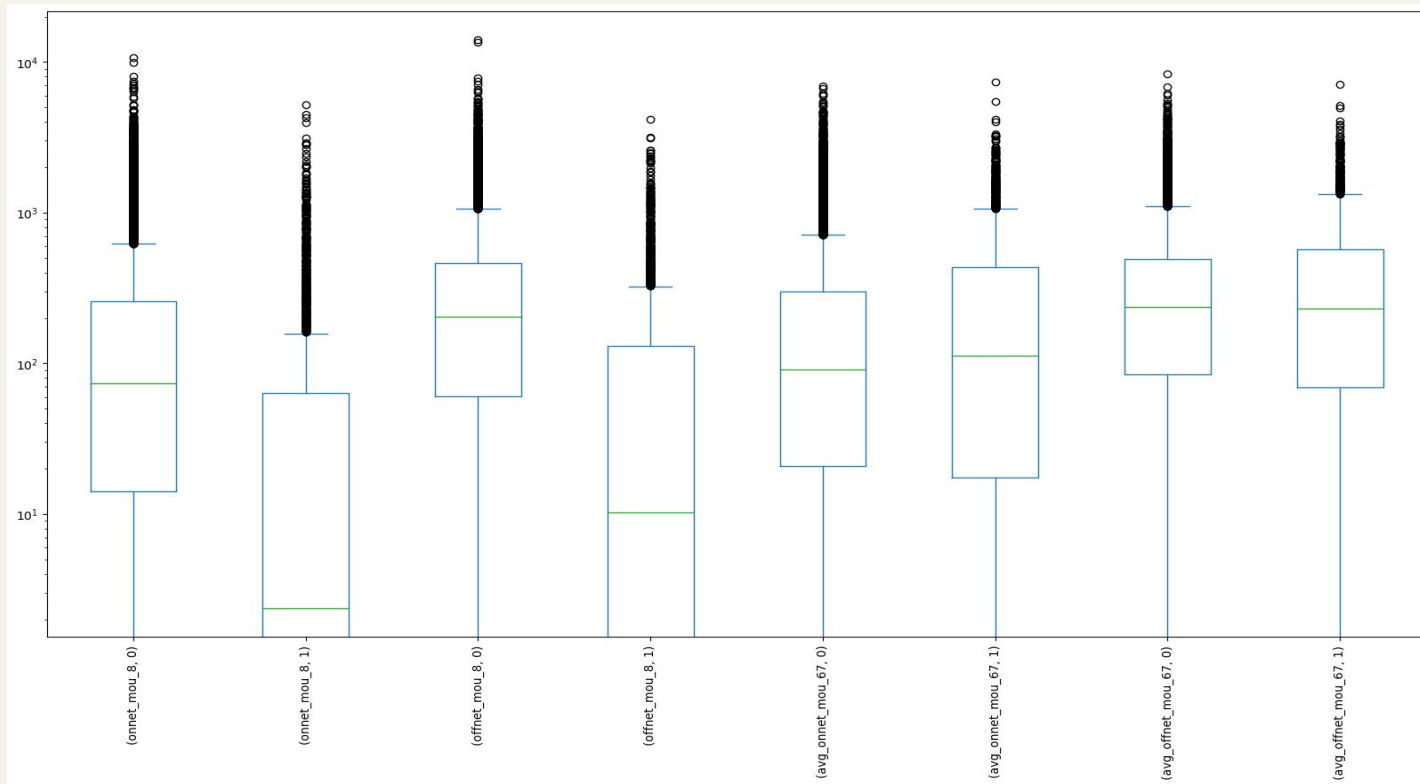
2G Data Distribution



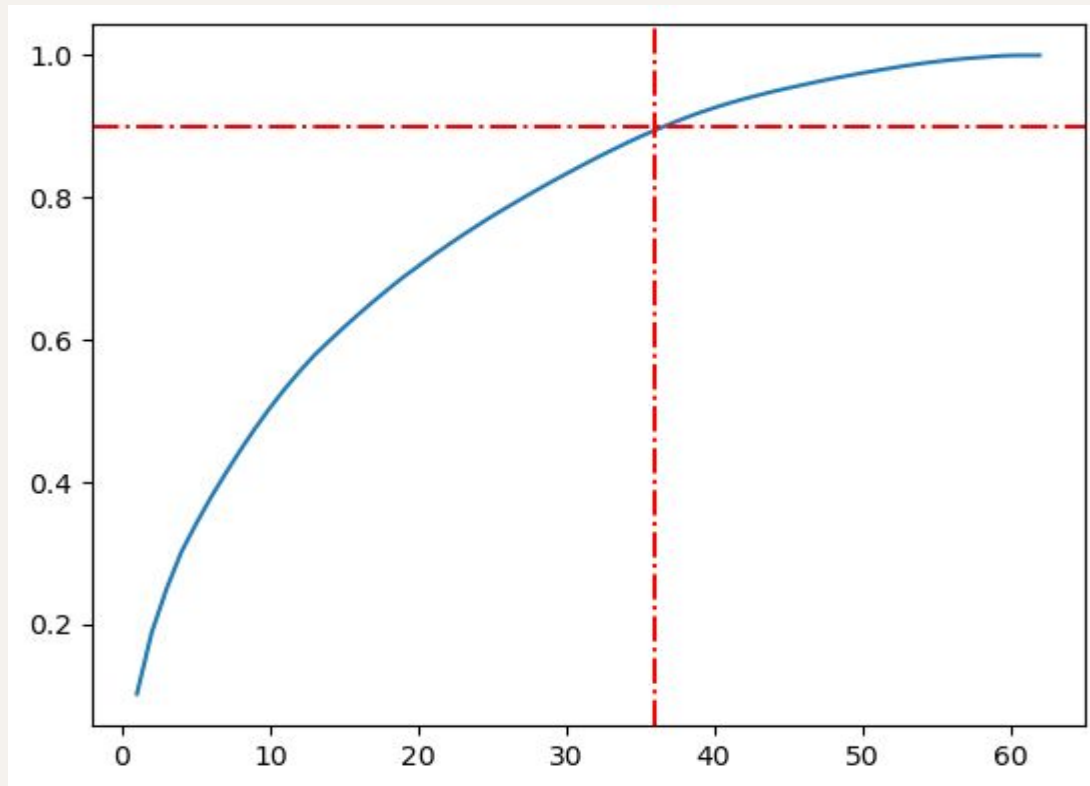
3G Data Distribution



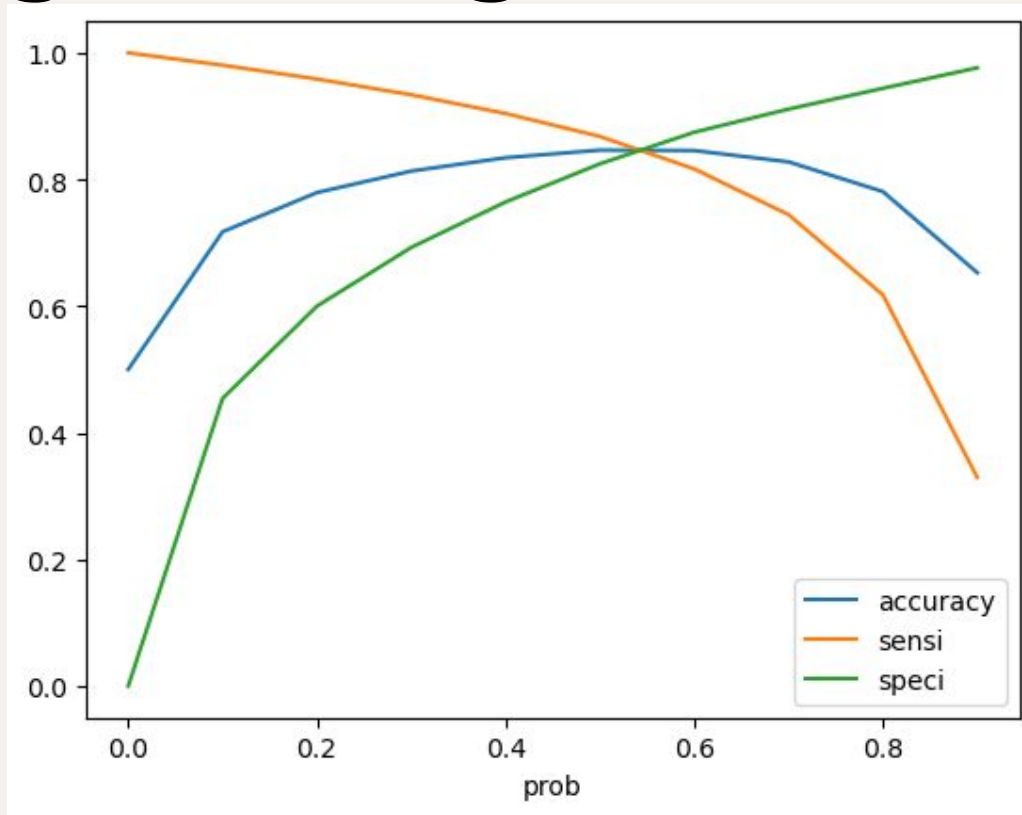
Onnet & Offnet Distribution



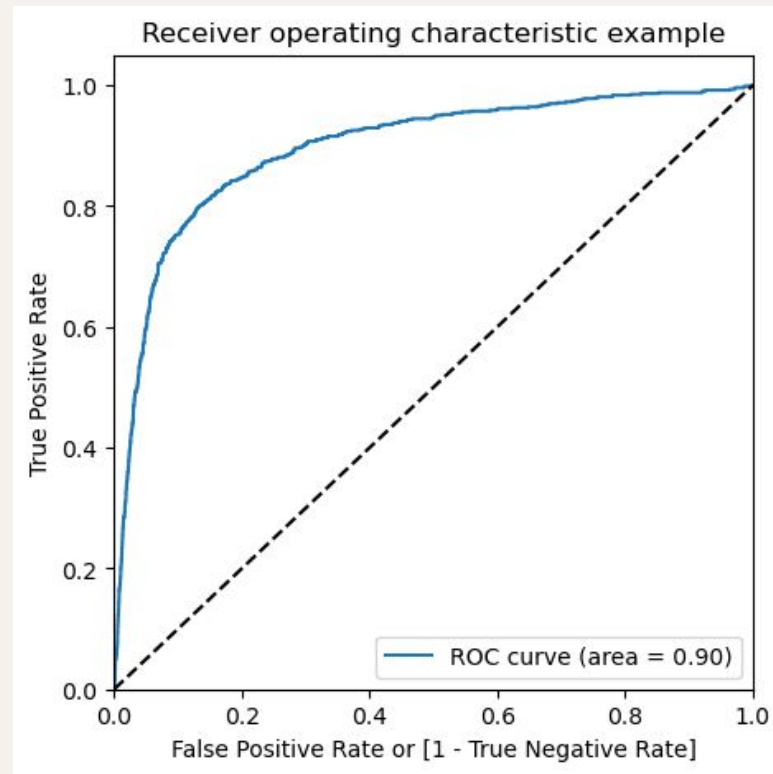
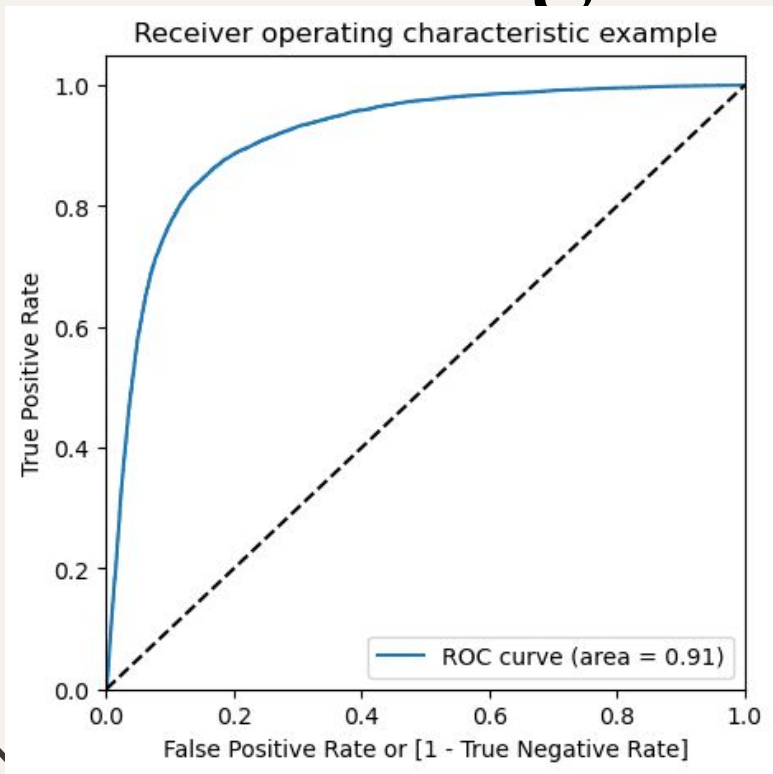
Principal Component Analysis



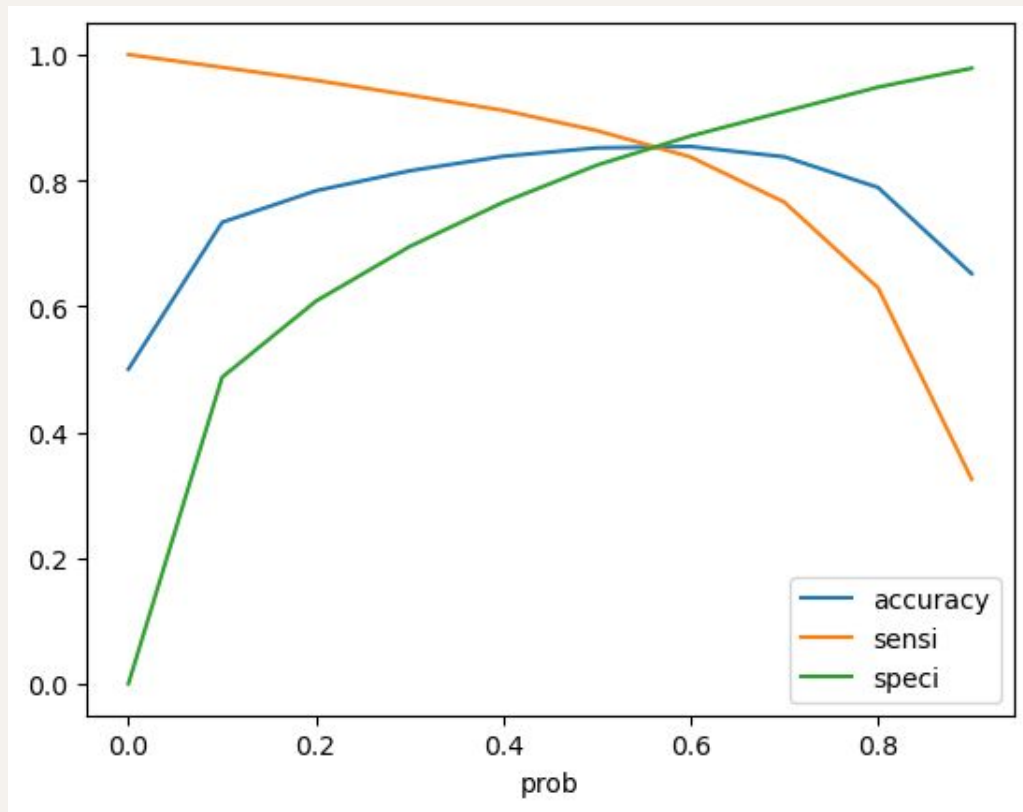
Logistic Regression PCA



Evaluating Model Performance



Logistic Regression RFE



Model results

- Model Accuracy is better along with Sensitivity and Specificity while using Logic Regression along with PCA. Model which is built with probability of 0.6 or greater yielded Better results.
- Area under the curve also resulted in 0.92

Important Predictive variables:

- total_og_mou_8
- sep_vbc_3g
- onnet_mou_8
- loc_ic_t2m_mou_8
- std_og_t2m_mou_8
- avg_total_ic_mou_67
- loc_ic_t2t_mou_8
- og_others_8

Recommended Strategy

- Based on the above variables, reduced usage of the mobile either for call or data is a clear indication of churn.
 - Hence on any month, if the usage of customer reduces or go below the average usage in previous month, then we do the following:
 - Assign a dedicated customer service specialist to understand the reason for leaving and bring the issue to closure in case of any pending problems reported by customers.
 - Roll out offers to retain customers.
-

The image features two thin, dark horizontal lines. The top line starts with a curved segment on the left and then continues straight to the right edge. The bottom line starts straight from the left edge and ends with a curved segment on the right.

Thanks