Telecom Churn Case Study

Submitted By

Karishma Sahay

Contents

Problem Statement

- 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, we will analyze 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.
- In this case study, there are two main models of payment in the telecom industry - postpaid and prepaid. We analyze 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.
- Churn prediction is usually more critical for prepaid customers, and the term ‘churn’ should be defined carefully. Also, prepaid is the most common model in India and Southeast Asia, while postpaid is more common in Europe in North America. This project is based on the Indian and Southeast Asian market. Churn is predicted using two approaches.

Problem Statement

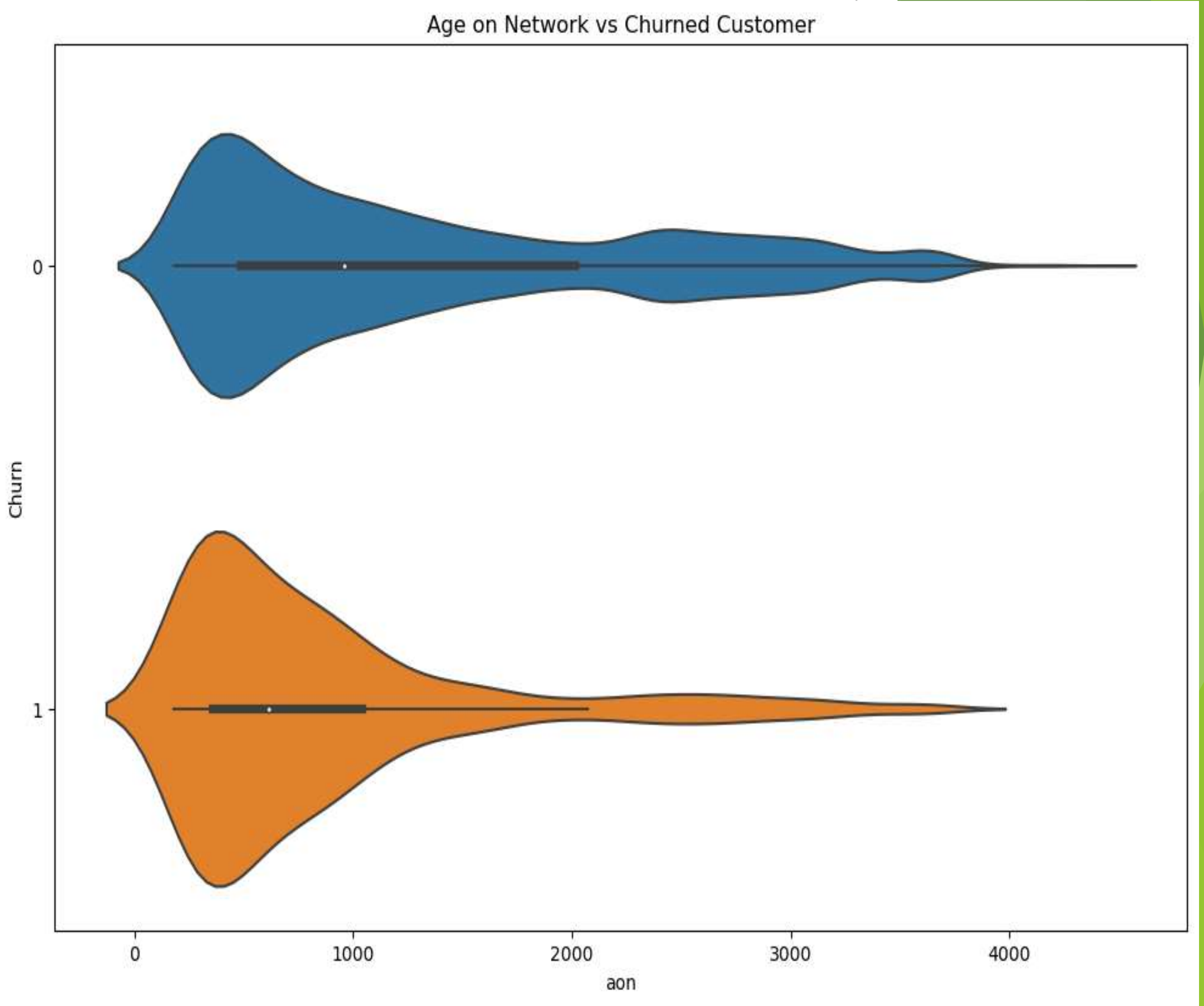
- Usage based churn and Revenue based churn. Usage based churn: Customers who have zero usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time. This case study only considers usage based churn. In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers).
- Thus, if we can reduce churn of the high-value customers, we will be able to reduce significant revenue leakage. Hence, this case study focuses on high value customers only.
- 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.

Problem Approach

- Importing the data and inspecting the
- data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE,VIF and pvalues)
- Model Evaluation (Specificity, Sensitivity,Accuracy)
- Making predictions on test set

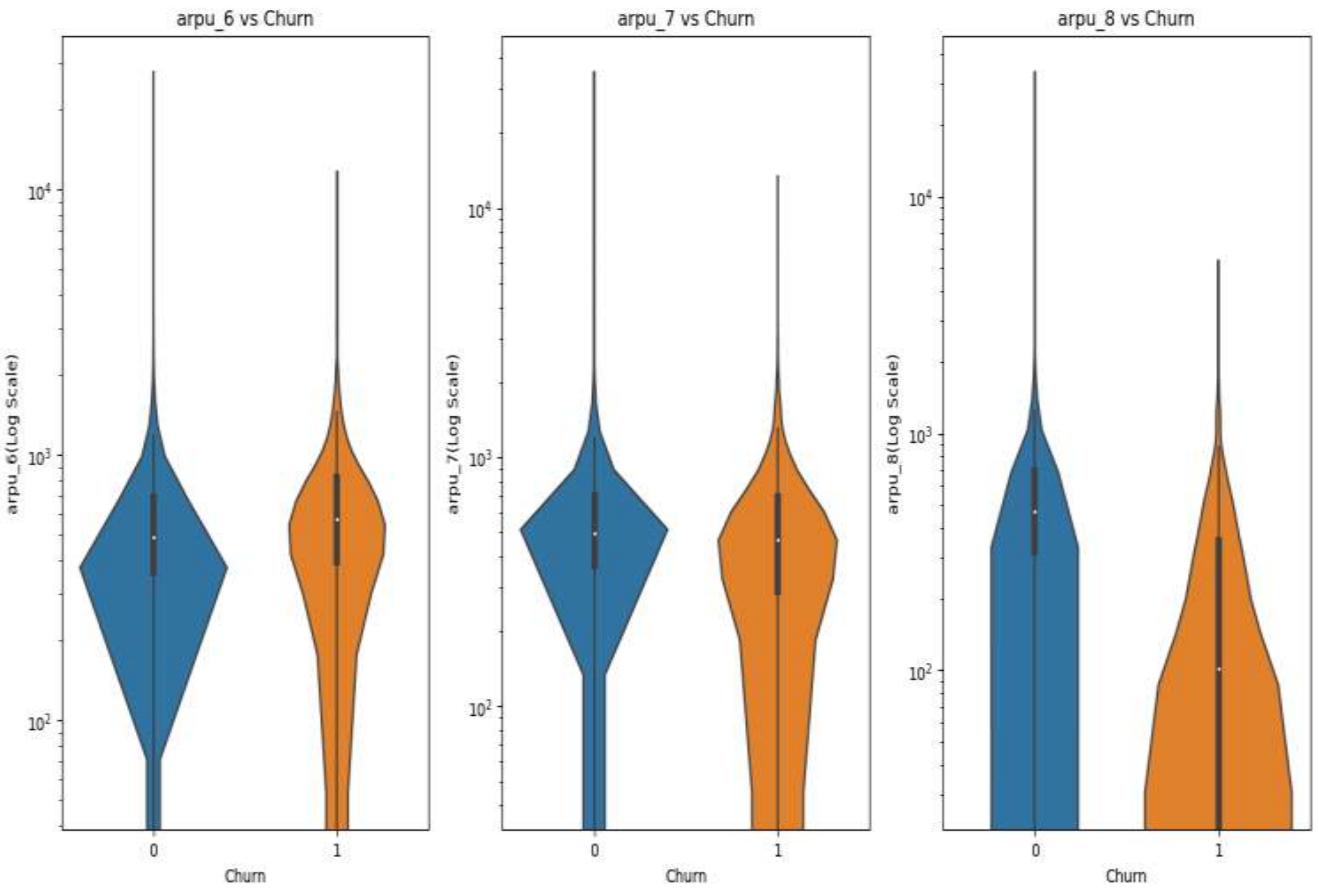
Exploratory Data Analysis

- From the graph the customers with less 'aon' can churn more easily when compared to the customers with high 'aon'



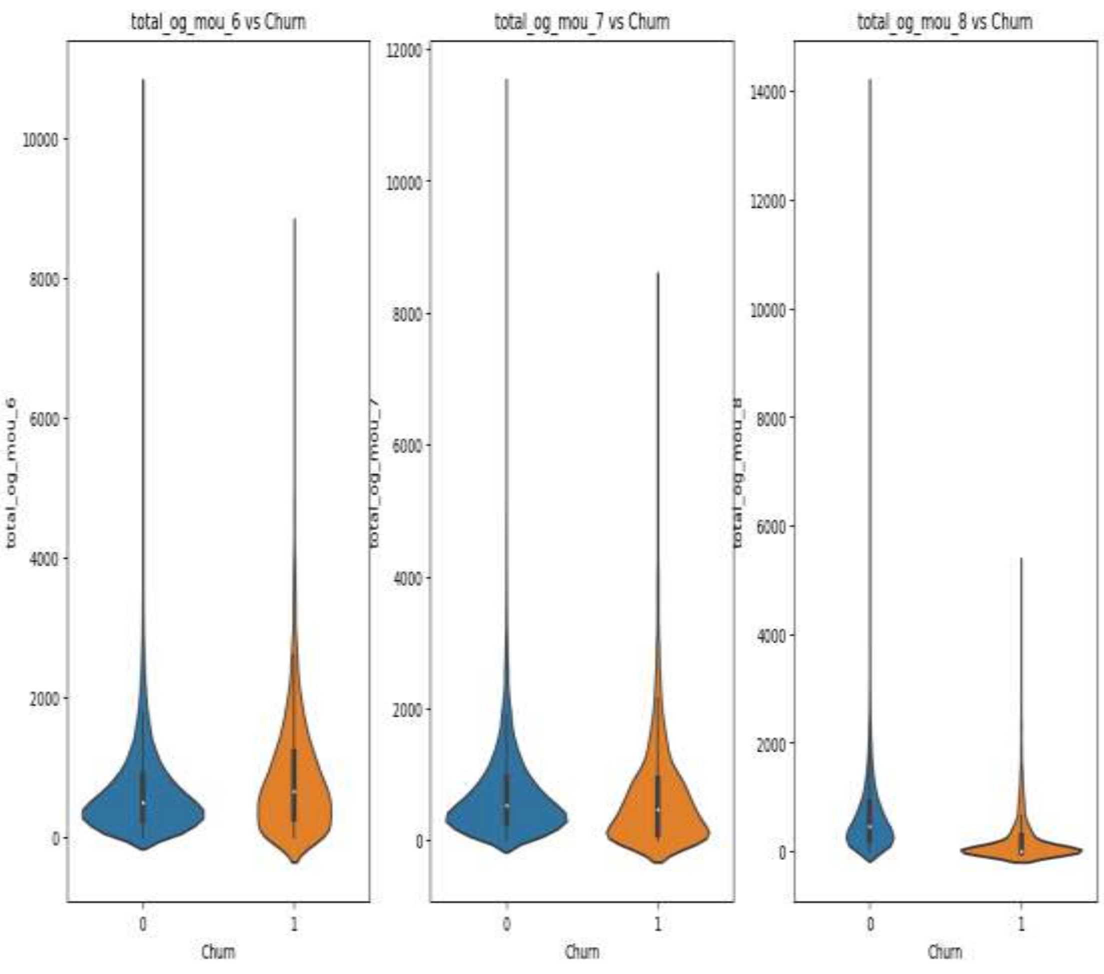
Exploratory Data Analysis

- From the figures we can understand that revenue generated by the customers who are about to churn is very unstable. Those customers whose "arpu" decreases in 7th month are more likely to churn when compared to ones with increase in arpu.



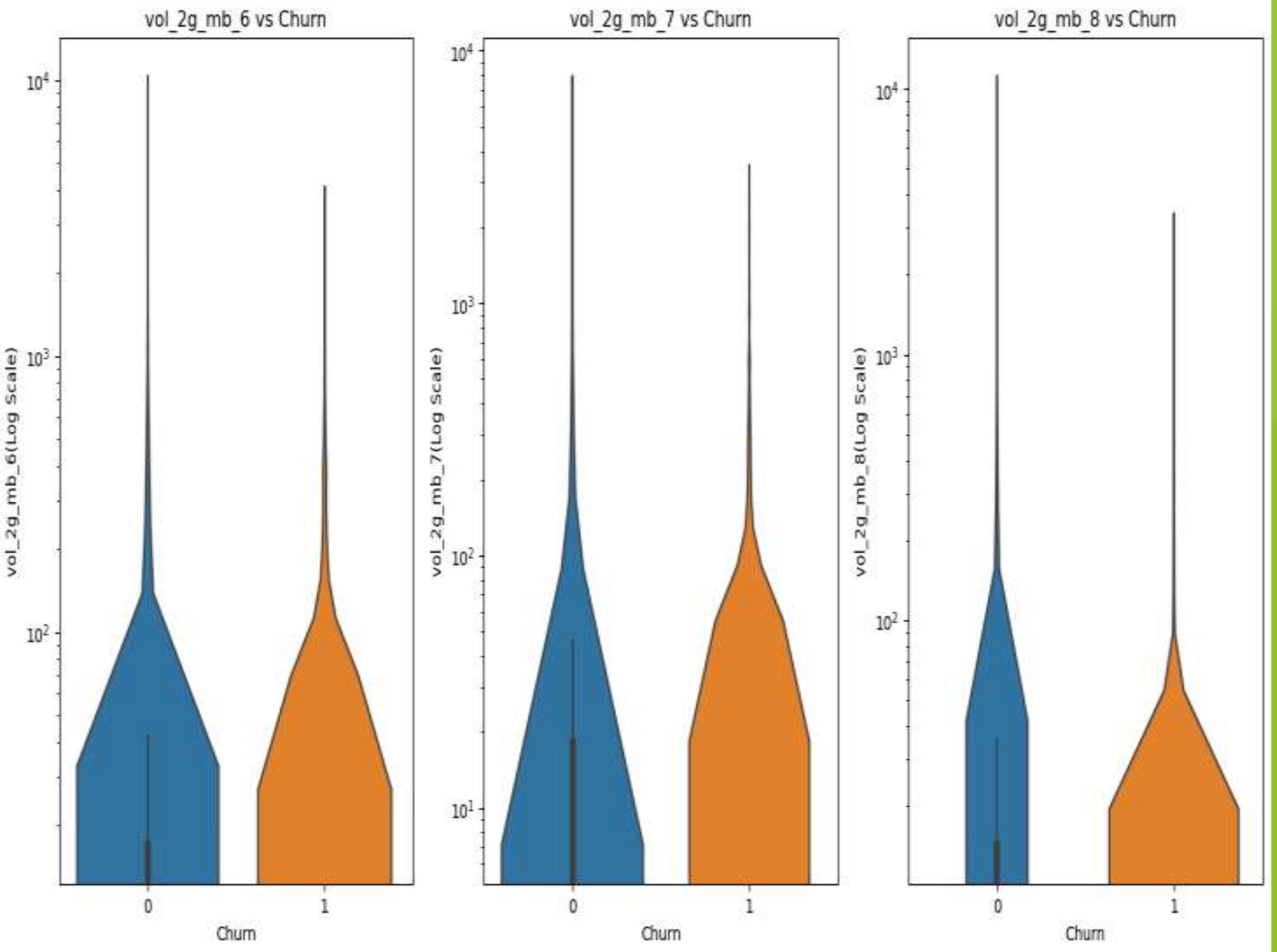
EDA

- Those customers who have high total_og_mou in 6th month and lower total_og_mou in 7th month are more likely to churn compared to the rest.



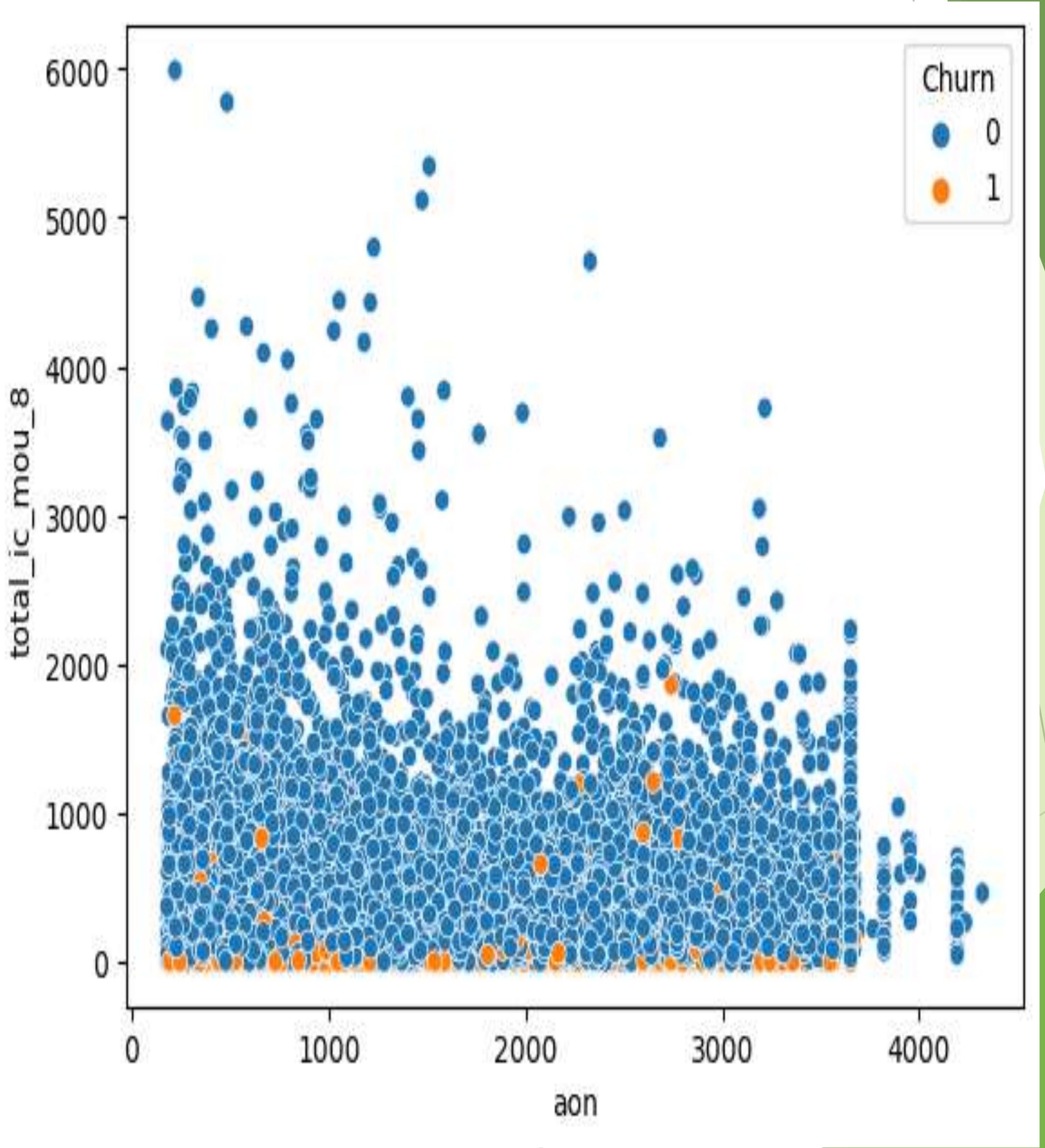
EDA

- Customers who have stable usage of 2g volumes throughout 6 and 7 months are less likely to churn. But the Customers with fall in consumption of 2g volumes in 7th month are more likely to churn.



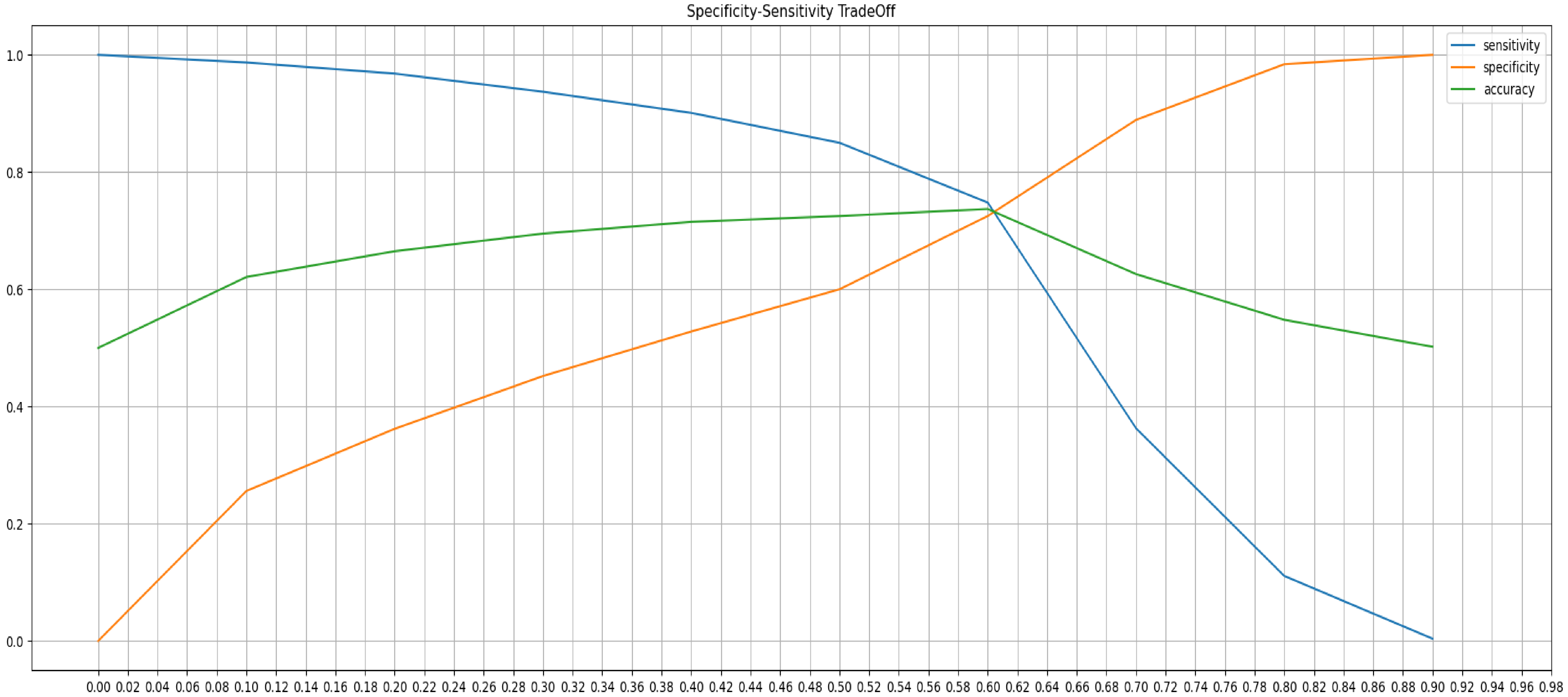
Bivariate Analysis

- Those customers who have less total_ic_mou_8 are more likely to churn irrespective of aon. And the customers who have total_ic_mou_8 > 2000 are very less likely to churn.

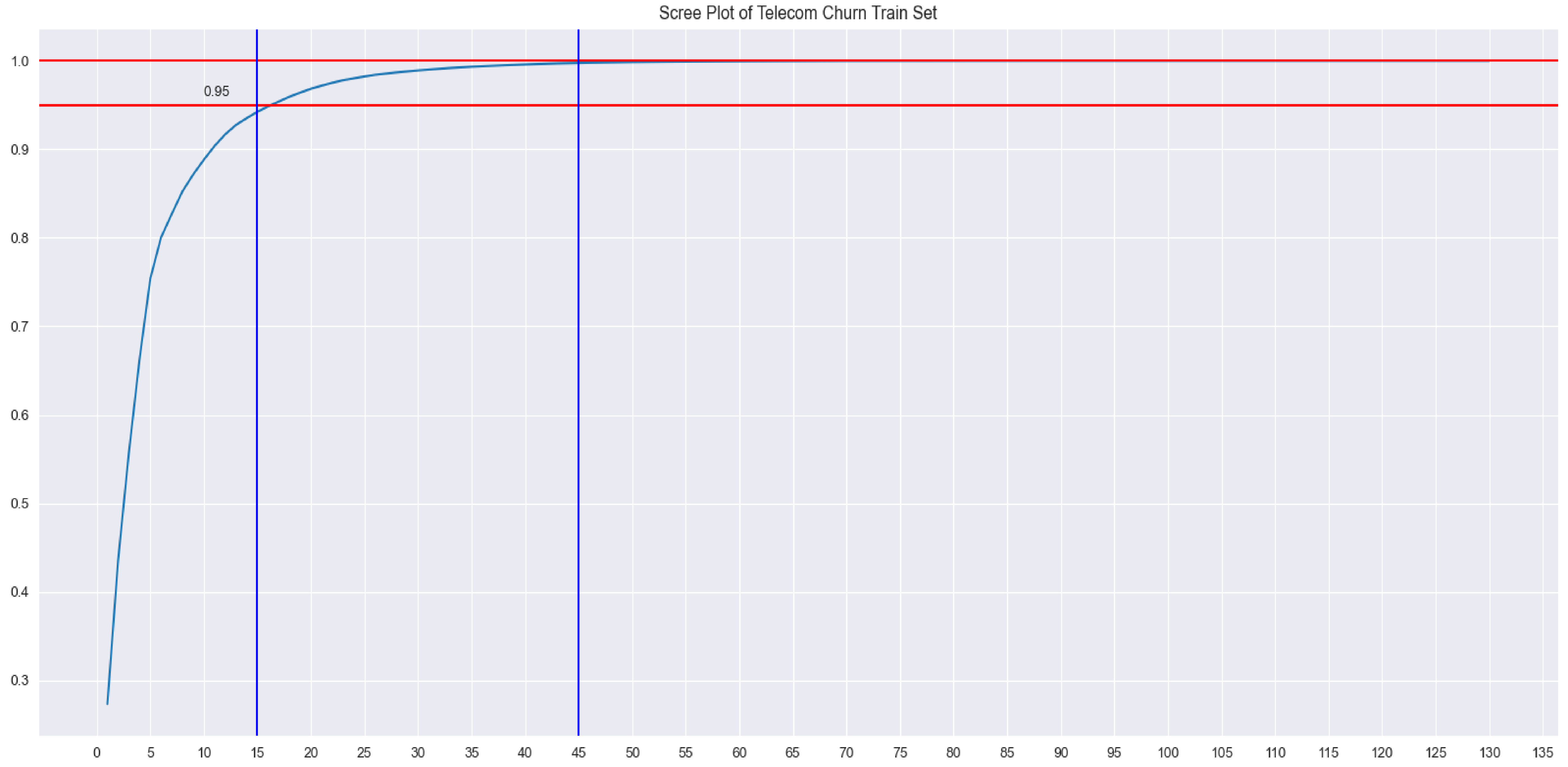


Model Evaluation

- The optimum probability cutoff for Logistic regression model is 0.61



Model Evaluation



Model Evaluation (Logistic Regression)

Conclusion

- ***Strongest Indicators of Churn:***
- Average Monthly Local Incoming Calls from Fixed Line:
- Customers who churn show lower average monthly local incoming calls from a fixed line in the action period by 1.27 standard deviations.
- ***Number of Recharges Done in Action Period:***
- Customers who churn exhibit a lower number of recharges done in the action period by 1.20 standard deviations. Higher Recharge Amount:
- Churning customers have done 0.6 standard deviations higher recharge than non-churn customers. Usage of 'Monthly 2G Package-0 / Monthly 3G Package-0':
- Customers who churn are more likely to be users of 'monthly 2g package-0 / monthly 3g package-0' in the action period (approximately 0.3 standard deviations higher than other packages).

Recommendation

- ***Focus on Customers with Lower Incoming Calls:***
- Concentrate on users with 1.27 standard deviations lower than the average incoming calls from the fixed line. They are most likely to churn.
- ***Target Customers with Fewer Recharges:***
- Concentrate on users who recharge less frequently (less than 1.2 standard deviations compared to the average) in the 8th month. They are the second most likely to churn.
- ***Use Models with High Sensitivity:***
- Models with high sensitivity are recommended for predicting churn. Consider using the PCA + Logistic Regression model, which has an ROC score of 0.87 and a test sensitivity of 100%.
- ***Continuous Monitoring:***
- Implement continuous monitoring to track the effectiveness of the recommendations over time.
- ***Customer Engagement Strategies:***
- Develop targeted customer engagement strategies to retain high-risk customers, such as personalized promotions or loyalty programs.
- ***Communication Channels:***
- Explore communication channels to engage with customers identified as high-risk, providing