Telecom Churn Case Study – By Nishant Kumar, Kuldeep Singh & Kushal Sugur

UPGRAD **2024**

Problem Statement

- Telecommunication industry is facing 15-25% of Annual churn rate on an average since it's a highly competitive market customers can choose from multiple service providers and actively choose from one operator to another.
- Also, it costs 5-10 times more to acquire new customers than to retain an existing one

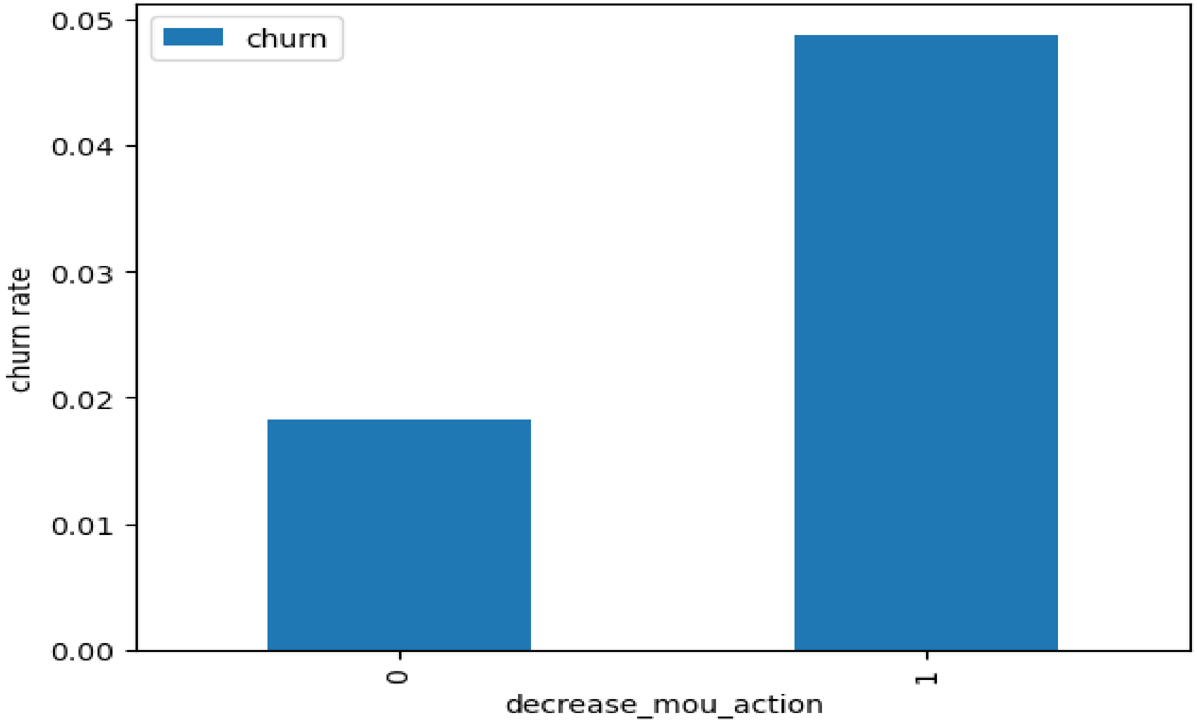
Business Objectives

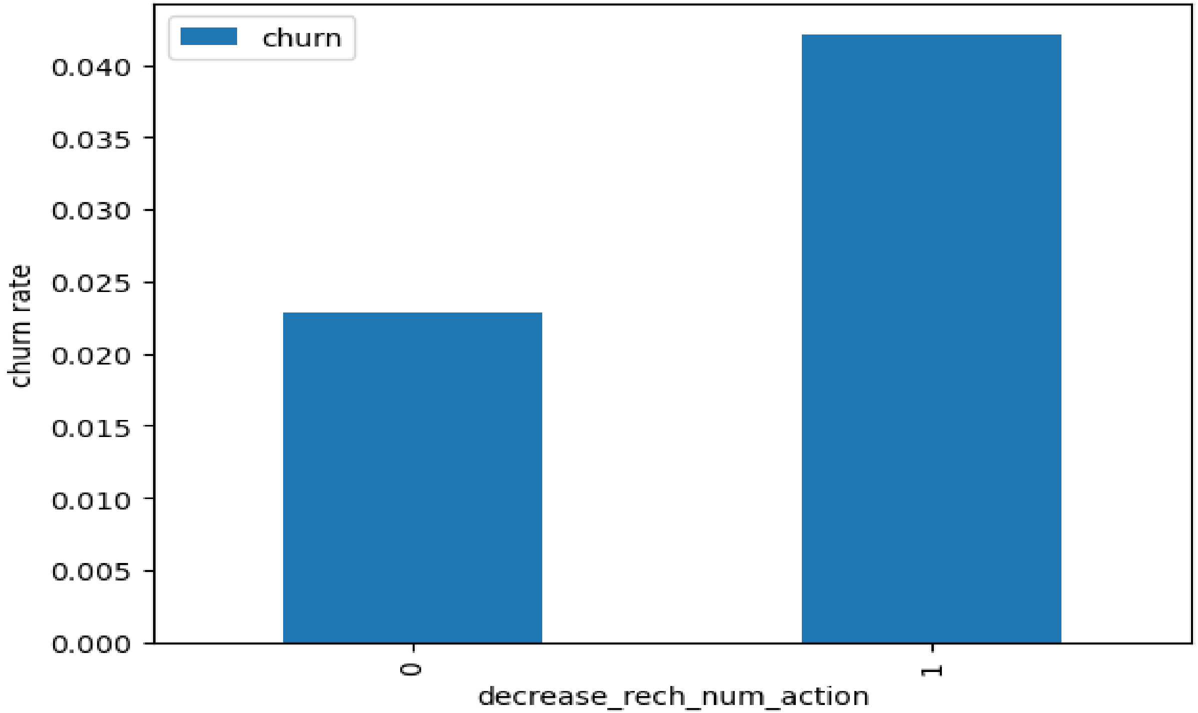
- The main objective is to retain customers than to acquire new ones.
- Reducing Customer churn rate for high profitable customers is to be focused here.
- To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

Problem Approach

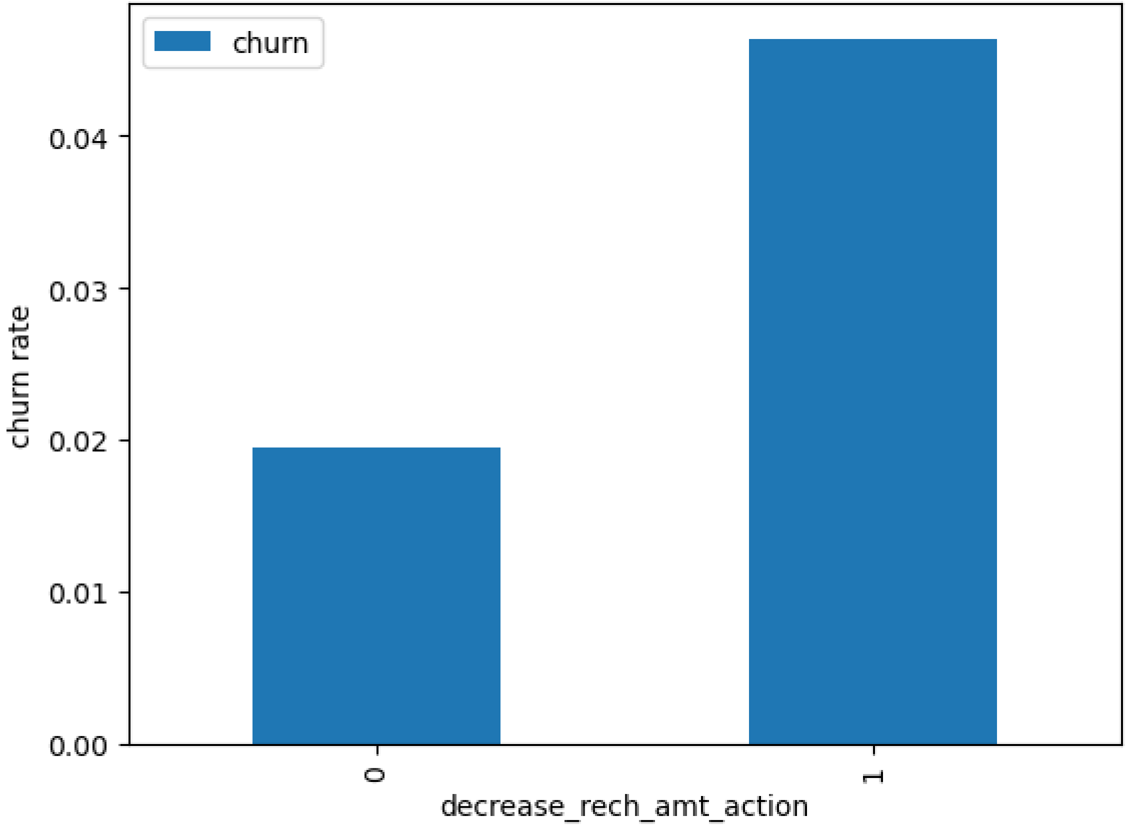
- Importing the data and inspecting the data frame
- Data Preparation (Handling missing values, Creating new features and Churn column with High values customers.)
- EDA
- Train-Test data split
- Handing Data imbalance
- Feature Scaling
- Model Building (With & without PCA)
- Model Evaluation(Accuracy, Sensitivity & Accuracy)
- Features selection using RFE (VIF & P-value)
- Making prediction on test set

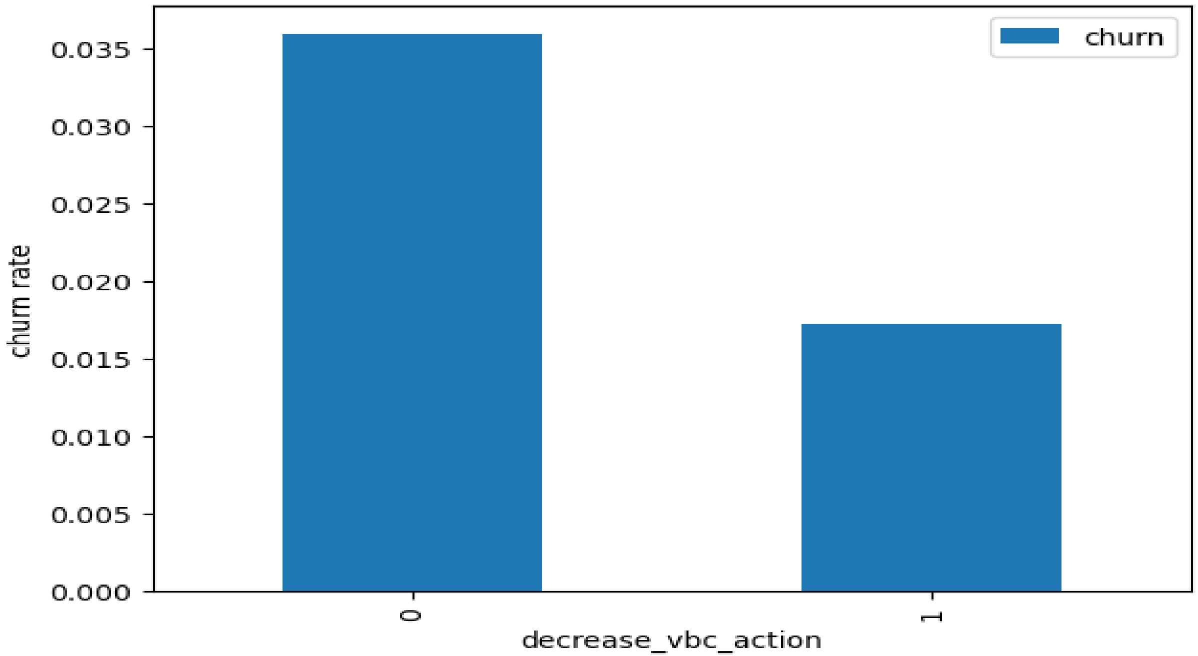
EDA - Univariate Analysis



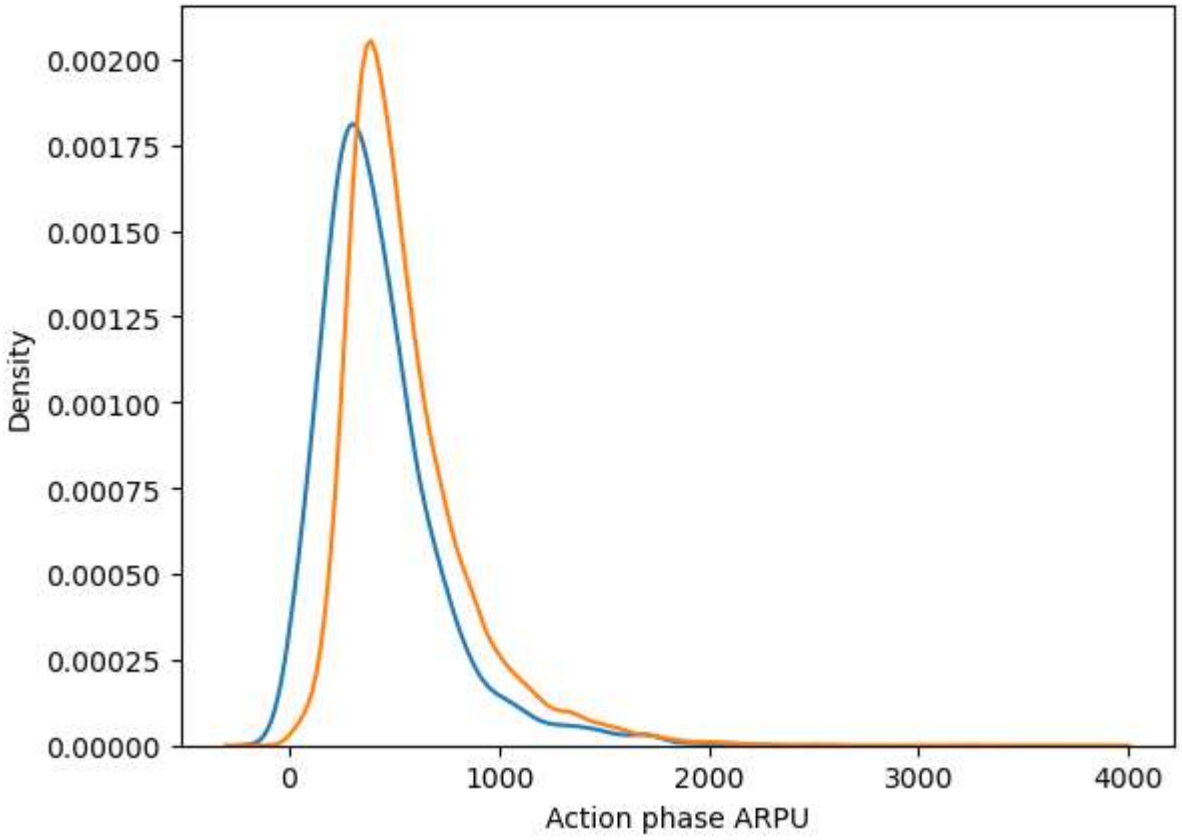


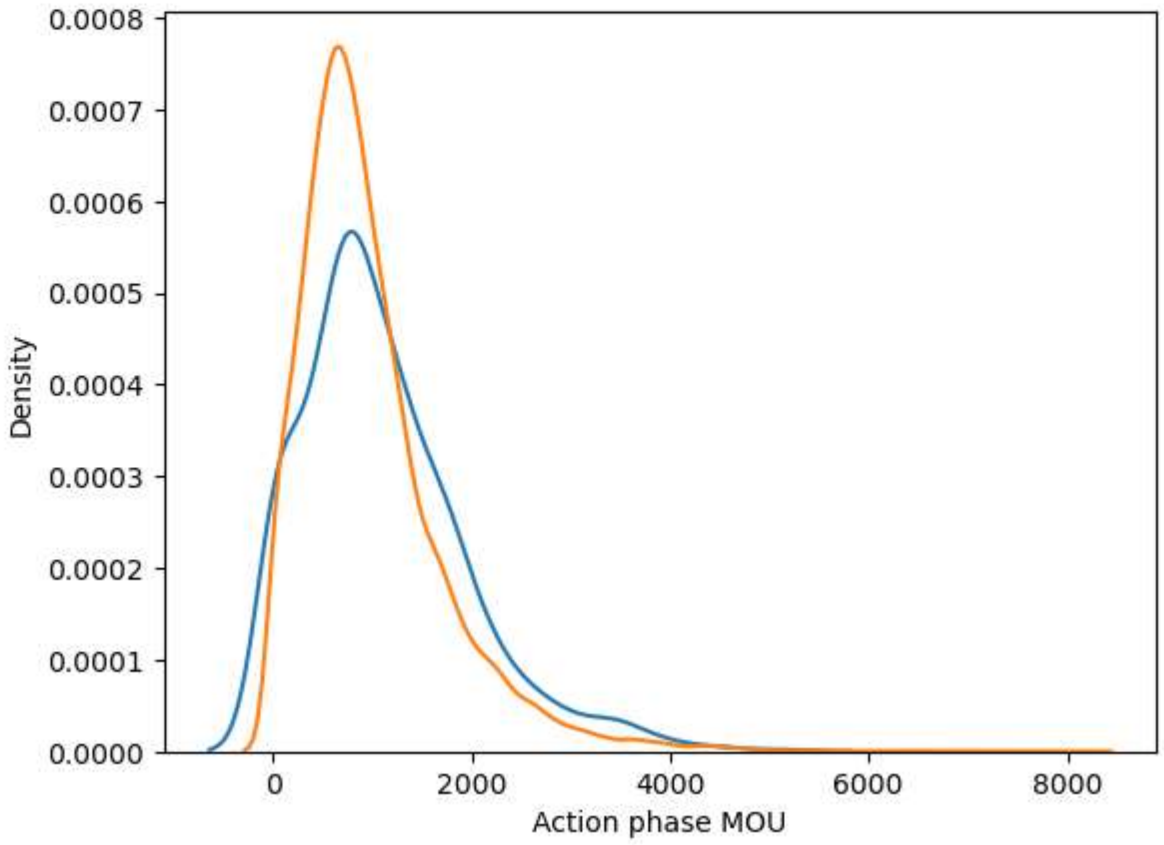
1. Customers who experienced a decrease in their minutes of usage (mou) during the action phase had a higher churn rate compared to the good phase.
2. As anticipated, customers who had a lower number of recharges during the action phase had a higher churn rate compared to the good phase.





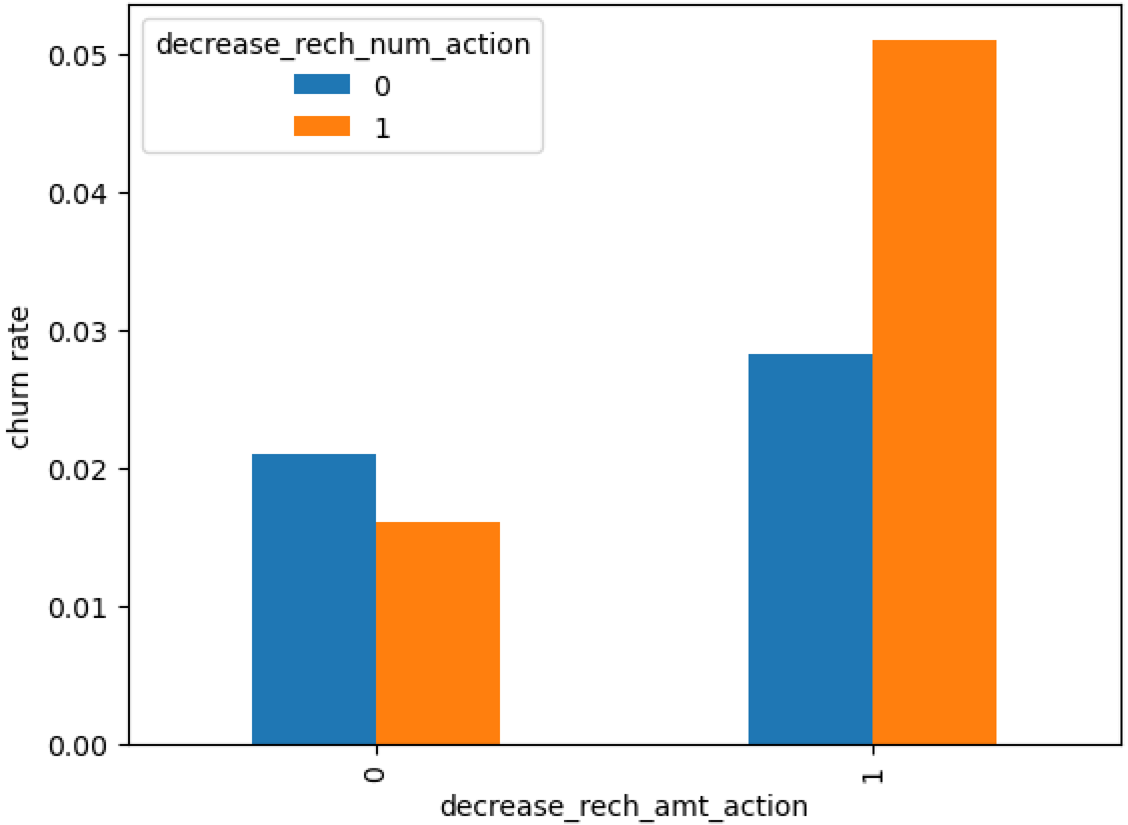
1. Similarly, we observe the same pattern: customers who recharged with a lower amount during the action phase had a higher churn rate in comparison to the good phase.
2. As anticipated, the churn rate is higher for customers whose volume-based cost increased during the action month. This suggests that customers are less inclined to perform monthly recharges during the action phase.

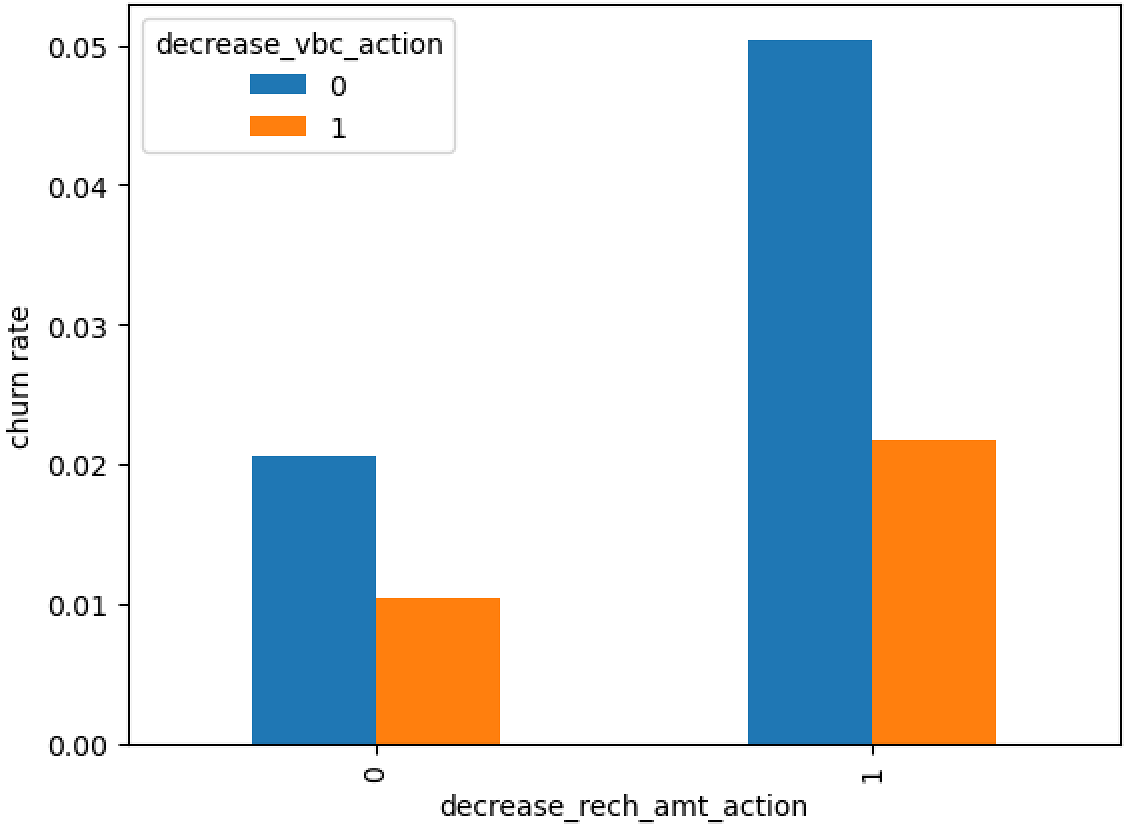




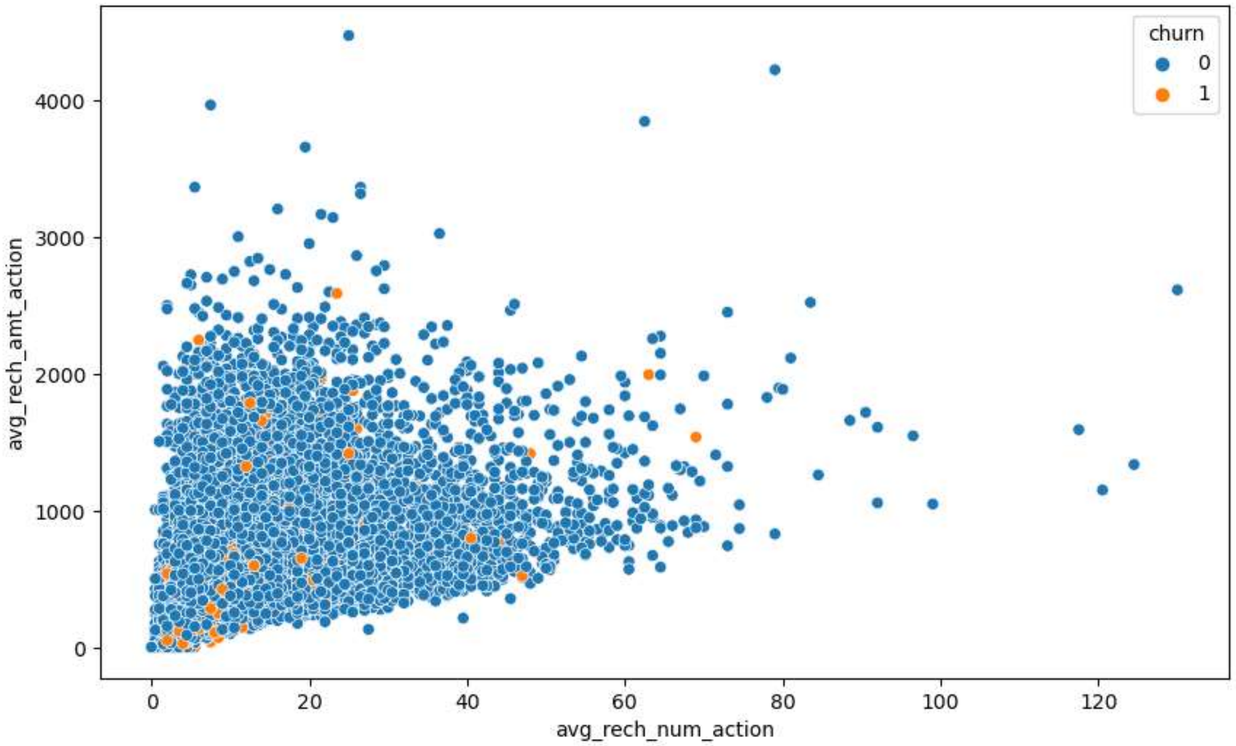
- For churned customers, most of the average revenue per user (ARPU) values are concentrated in the range of 0 to 900, indicating that customers with lower ARPU are more likely to churn.
- Conversely, not churned customers exhibit denser ARPU values in the range of 0 to 1000, suggesting that customers with higher ARPU are less likely to churn.
- Minutes of usage(MOU) of the churn customers is mostly populated on the 0 to 2500 range. Higher the MOU, lesser the churn probability.

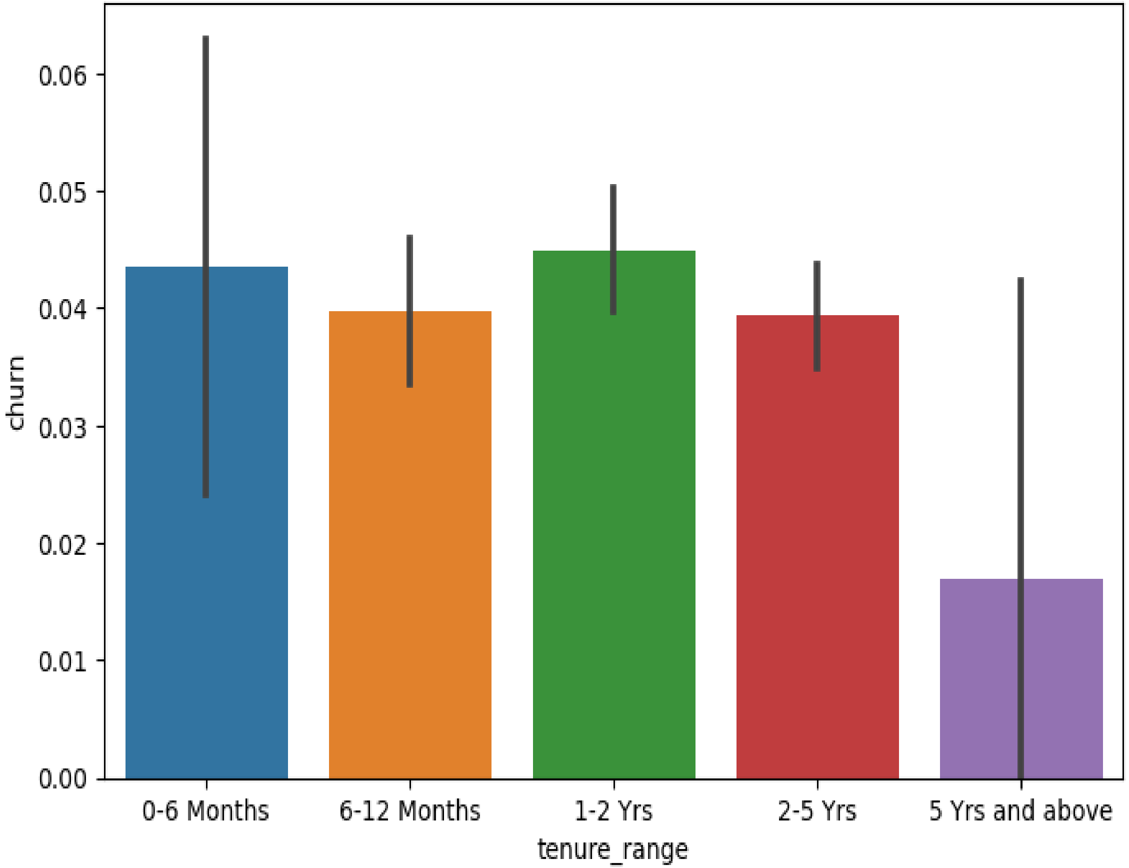
Bivariate Analysis





- We can see from the above plot, that the churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.
- In Analysis, customers are more likely to churn when their recharge amount decreases, and volume-based costs increase during the action month.





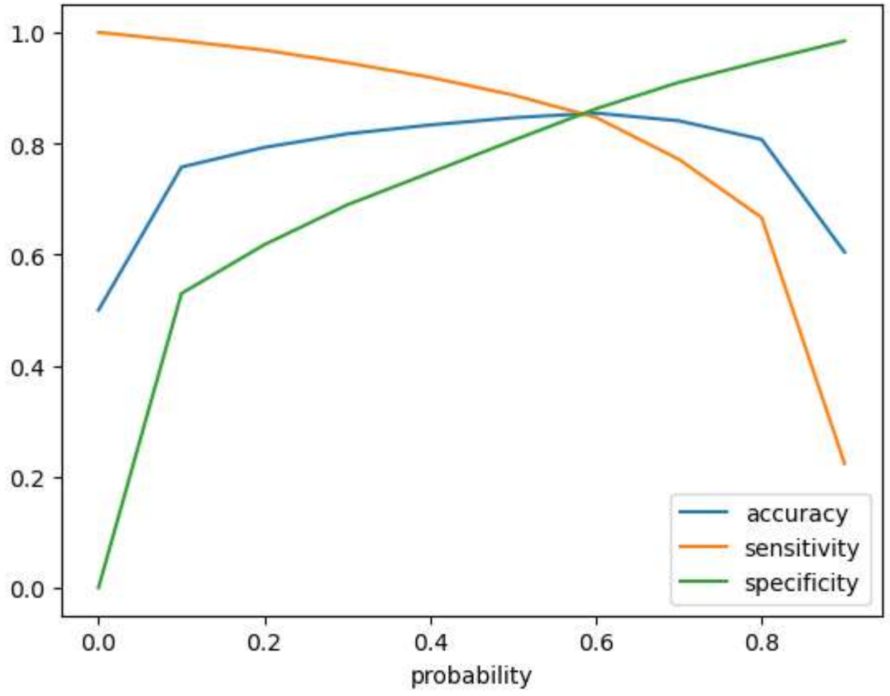
- In the first graph we can see that the recharge number and the recharge amount are mostly proportional. More the number of recharge, more the amount of the recharge.
- In the second graph, The highest churn rate is observed in the first 0-6 months of customer subscription, after which it gradually decreases as customers stay with the network longer.

Model Evaluation with PCA

Types of Model (60 components)	Train Datasets			Test Datasets		
	Accuracy Score	Sensitivity	Specificity	Accuracy Score	Sensitivity	Specificity
Logistic Regression (Best C score – 0.01)	0.86	0.90	0.83	0.82	0.81	0.83
Support Vector Machine (C – 100, gamma – 0.0001)	0.89	0.93	0.85	0.85	0.81	0.85
Decision Trees (max_depth=10, min_samples_leaf=50, min_samples_split=50, random_state=100)	0.90	0.91	0.88	0.85	0.64	0.86

- In summary, both Logistic Regression and SVM models consistently achieve an excellent sensitivity rate of around 81%, along with a strong overall accuracy of approximately 85%. These models excel in identifying the target outcome effectively.

Logistic Regression Model Evaluation without PCA



- Accuracy - Becomes stable around 0.6
- Sensitivity - Decreases with the increased probability.
- Specificity - Increases with the increasing probability.
- At point 0.6 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.
- Here we are intended to achieve better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the optimum probability cutoff, we are taking ***0.6*** for achieving higher sensitivity, which is our main goal.
- Also, the area of the ROC curve is closer to 1, which is the Gini of the model.

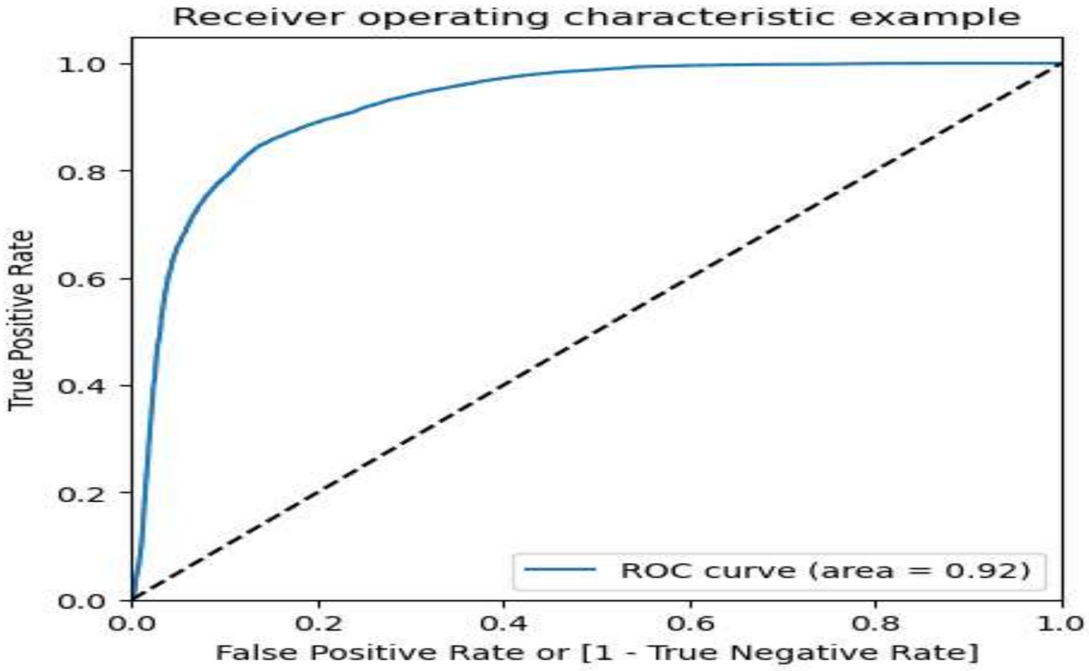
Model summary

•Train set

- Accuracy = 0.84
- Sensitivity = 0.88
- Specificity = 0.80

•Test set

- Accuracy = 0.80
- Sensitivity = 0.81
- Specificity = 0.80



Final conclusion with no PCA

- The Logistic Regression model without PCA performs well in terms of sensitivity and accuracy.
- Its performance is comparable to models that utilize PCA for dimensionality reduction.
- Choosing the Logistic Regression model with PCA simplifies the model while retaining the ability to explain important predictor variables and their significance.
- This model is valuable for identifying the variables that influence decisions regarding potentially churned customers.
- Its simplicity makes it a more business-friendly choice and enhances its relevance for explaining insights to stakeholders.

Business Recommendation

1. We should focus on customers exhibiting reduced usage of incoming local calls and outgoing ISD calls during the action phase, especially in August.
2. Customers with decreased charges for outgoing others in July and lower incoming others charges in August should be targeted.
3. Those customers experiencing an increase in value-based costs during the action phase are at a higher risk of churning, making them prime candidates for targeted offers.
4. Customers who significantly increase their monthly 3G recharge in August are more likely to churn and should be addressed.
5. Churn is more probable for customers with a decrease in STD incoming minutes of usage for operators T to fixed lines of T in August.
6. A decrease in monthly 2G usage for August is a significant indicator of potential churn.
7. A decrease in incoming minutes of usage for operators T to fixed lines of T in August suggests a higher likelihood of churn.
8. Customers with increased roaming outgoing minutes of usage (roam_og_mou_8) are more likely to churn, as indicated by their positive coefficients.

In essence, focusing on these specific customer behaviors and variables can help in identifying and targeting potential churners effectively.