Telecom Churn Prediction By Jeet Mehta

Objective: Predict customer churn and identify key factors influencing churn to improve customer retention strategies.
Methods Used: Univariate and Bivariate Analysis, PCA, Logistic Regression.

Business Problem

- **Telecommunications Industry Challenges**:
  - With growing competition, customer churn has surged to an annual rate of 15-25%. This high turnover directly impacts profitability, as the cost of acquiring new customers is 5-10 times more than retaining existing ones.
  - For established telecom operators, the priority is to **retain high-value, profitable customers** to sustain long-term growth.
- **Project Objective**:
  - To tackle this issue, telecom companies must accurately predict which customers are likely to churn. This project aims to analyze customer data from a leading telecom provider and build predictive models to **identify customers at risk of leaving** and highlight the **key drivers behind their decision to churn**.

Objective

In-Depth Customer Data Analysis:

- Perform a detailed examination of customer-level data from a major telecom provider to reveal trends and patterns that indicate potential churn.

Build Predictive Models:

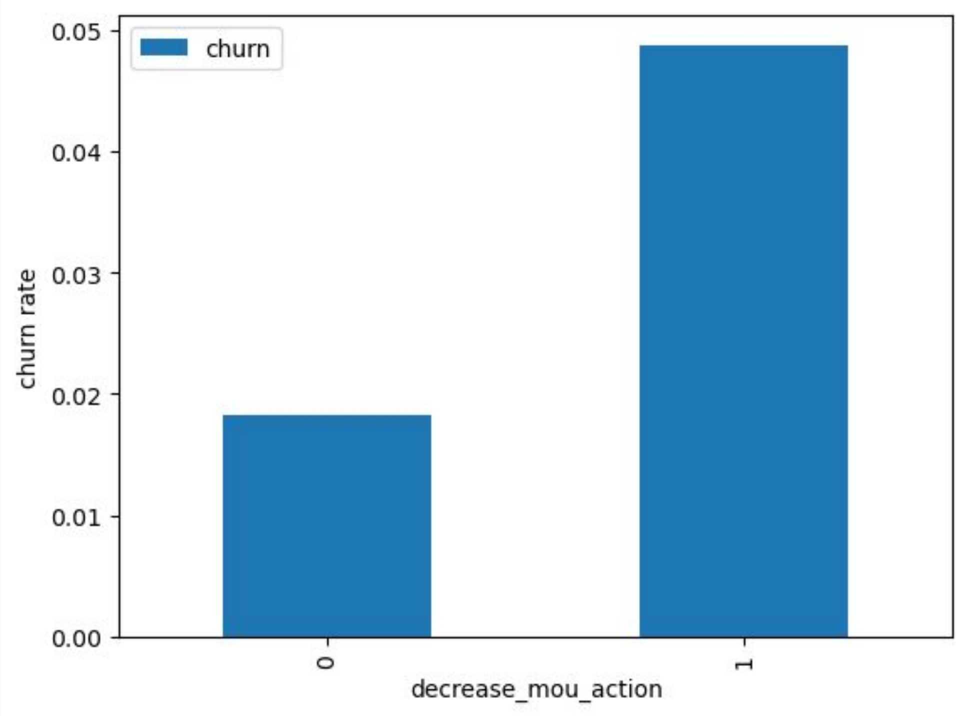
- Develop advanced predictive models to identify customers at **high risk of leaving**, empowering the company to implement effective retention strategies.

Uncover Key Drivers of Churn:

- Pinpoint the main factors leading to customer churn, offering actionable insights to focus retention efforts on the most critical areas.

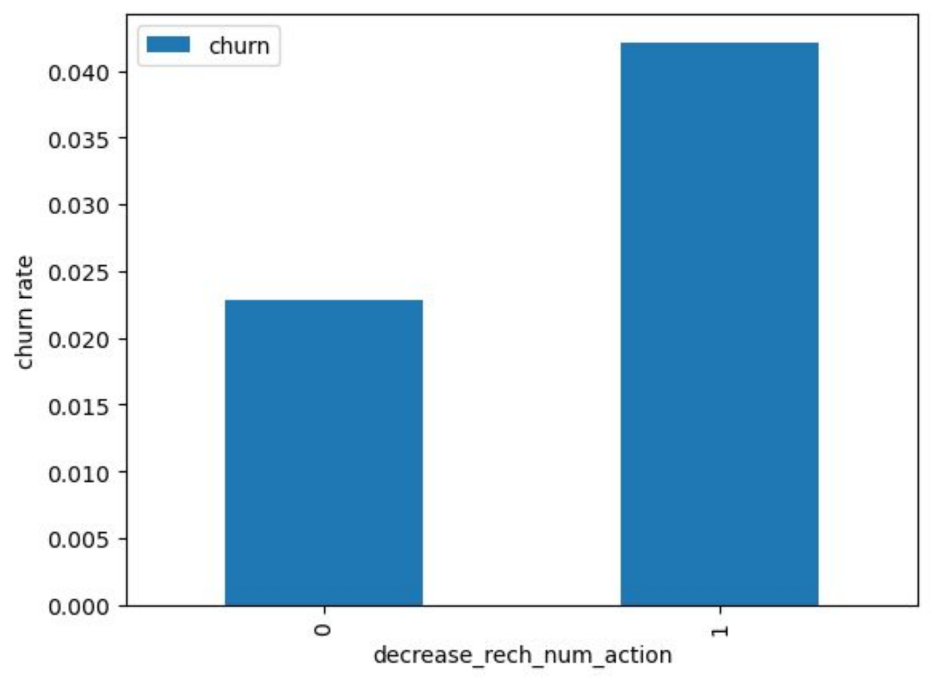
Univariate Analysis - Churn Based on Minutes of Usage

Insight: Customers with decreased **minutes of usage (MOU)** in the action phase show a higher churn rate compared to the good phase.



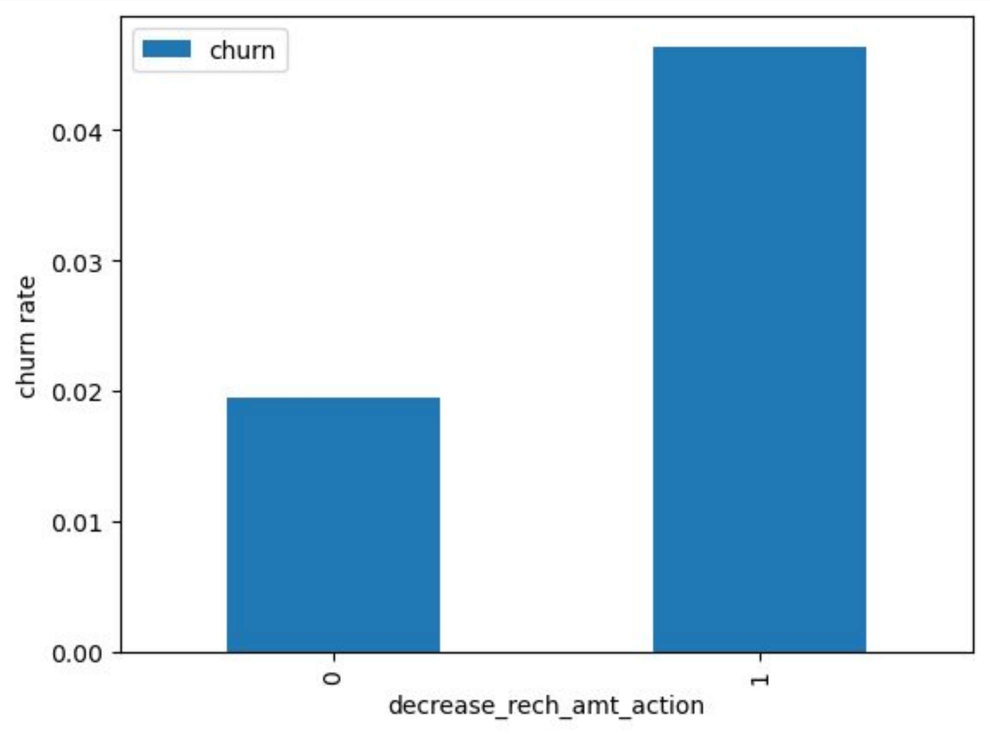
Univariate Analysis - Churn Based on Number of Recharges

Insight: Customers with a reduced **number of recharges** in the action phase are more likely to churn.



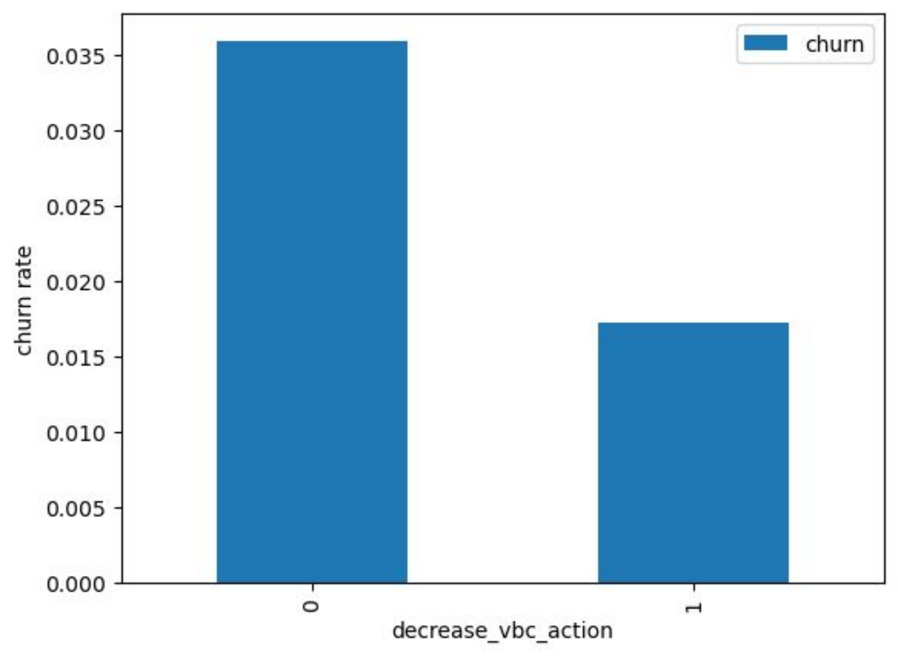
Univariate Analysis - Churn Based on Recharge Amount

Insight: A drop in the **recharge amount** during the action phase corresponds to a higher churn rate.



Univariate Analysis - Churn Based on Volume-Based Cost

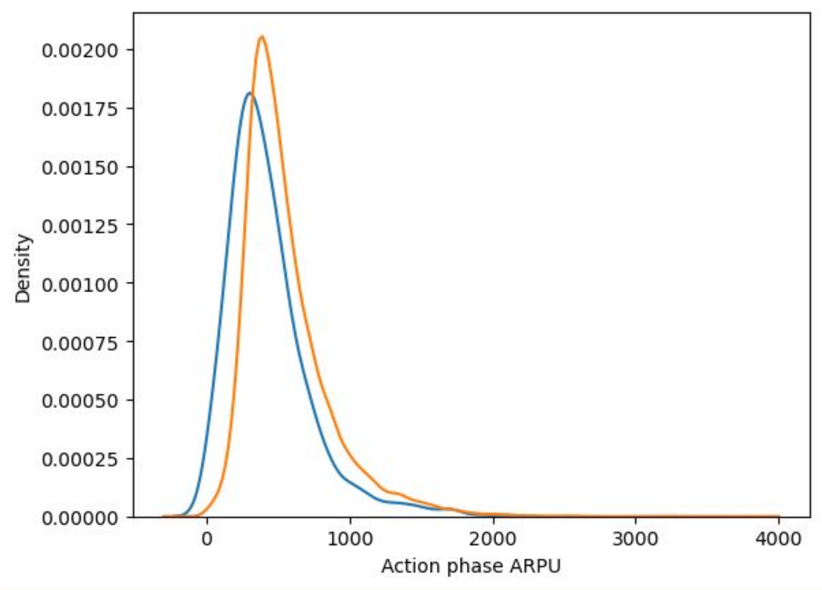
Customers with increased **volume-based cost** (VBC) in the action phase show a higher churn rate. They tend not to recharge monthly when in the action phase.



Analysis of ARPU (Average Revenue Per User)

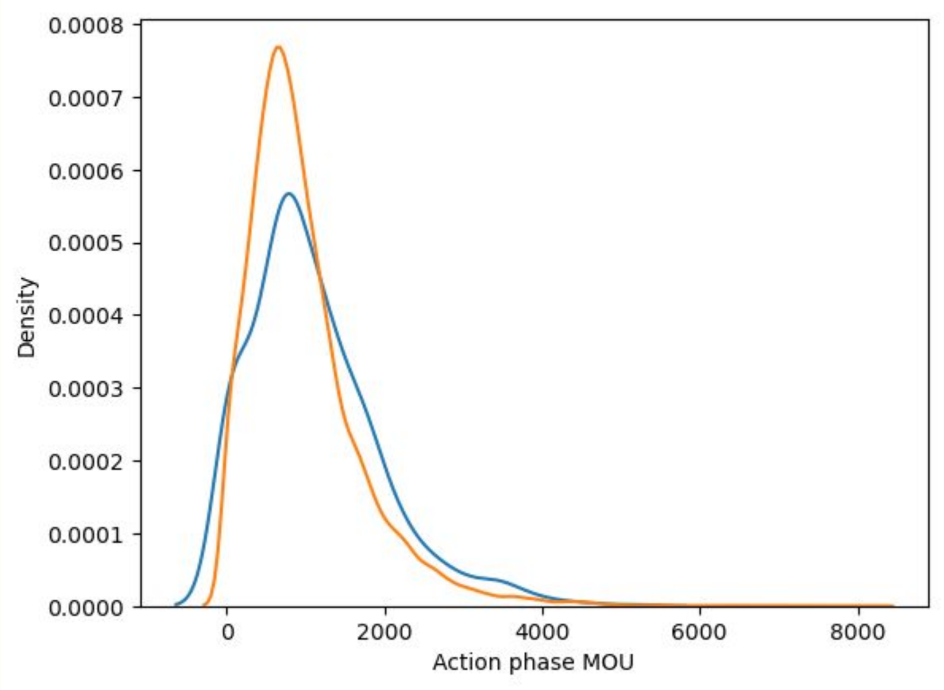
Churned Customers: ARPU for churned customers is densely concentrated in the **0 to 900** range.

Non-Churned Customers: ARPU for non-churned customers is mostly in the **0 to 1000** range.



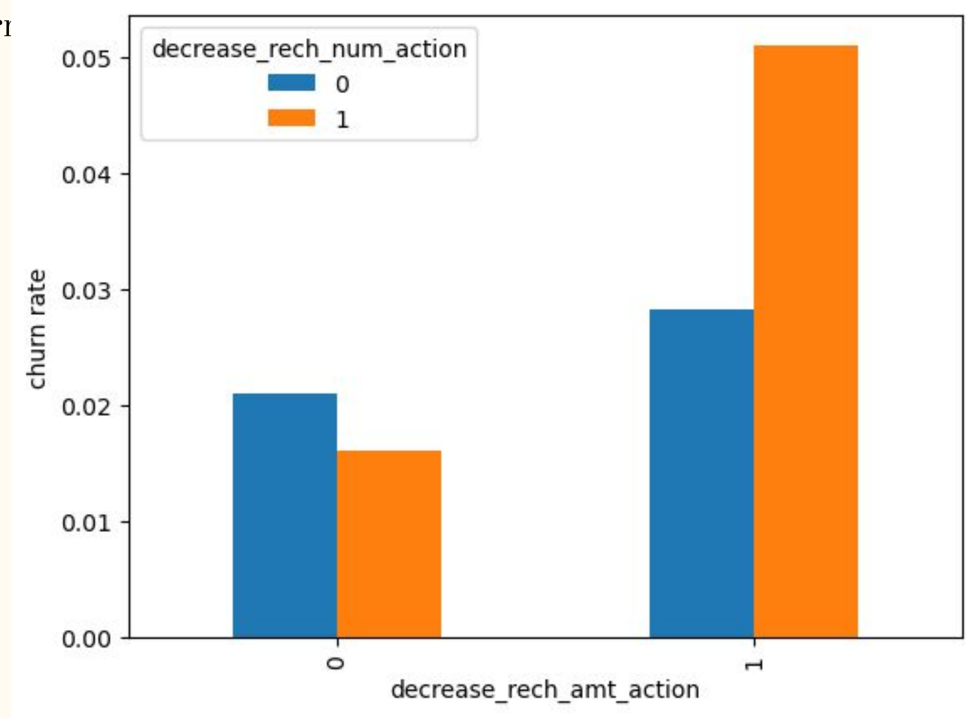
Analysis of Minutes of Usage (MOU)

Insight: Higher MOU is associated with a lower churn rate. Churned customers typically have MOU concentrated in the **0 to 2500** range.



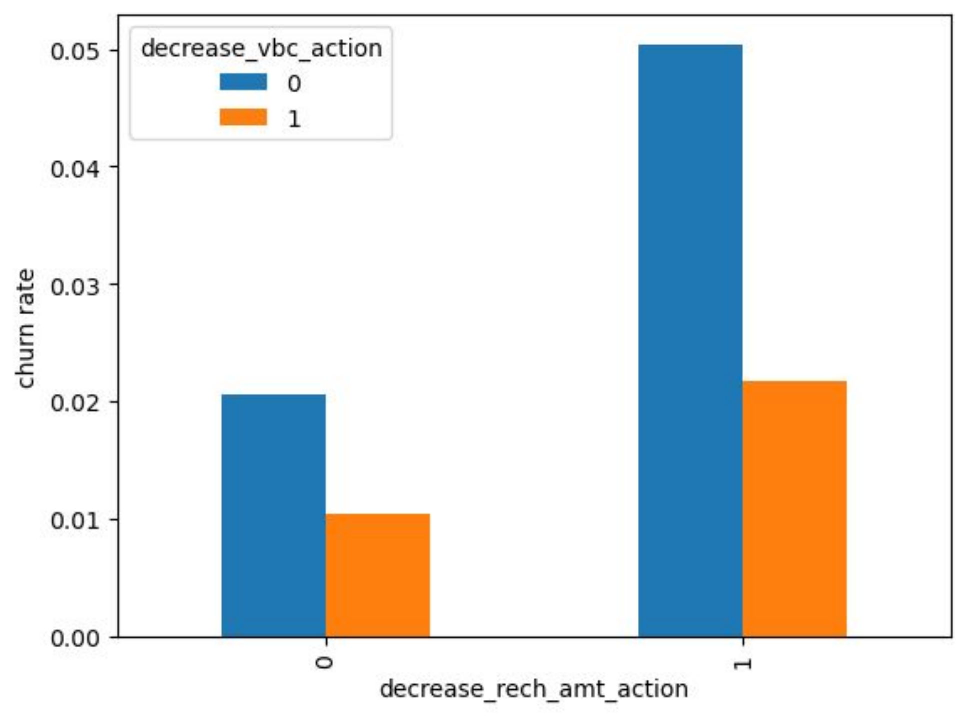
Bivariate Analysis - Recharge Amount & Number of Recharges

Insight: Customers with both a **decrease in recharge amount** and **recharge number** in the action phase show a significantly higher churn



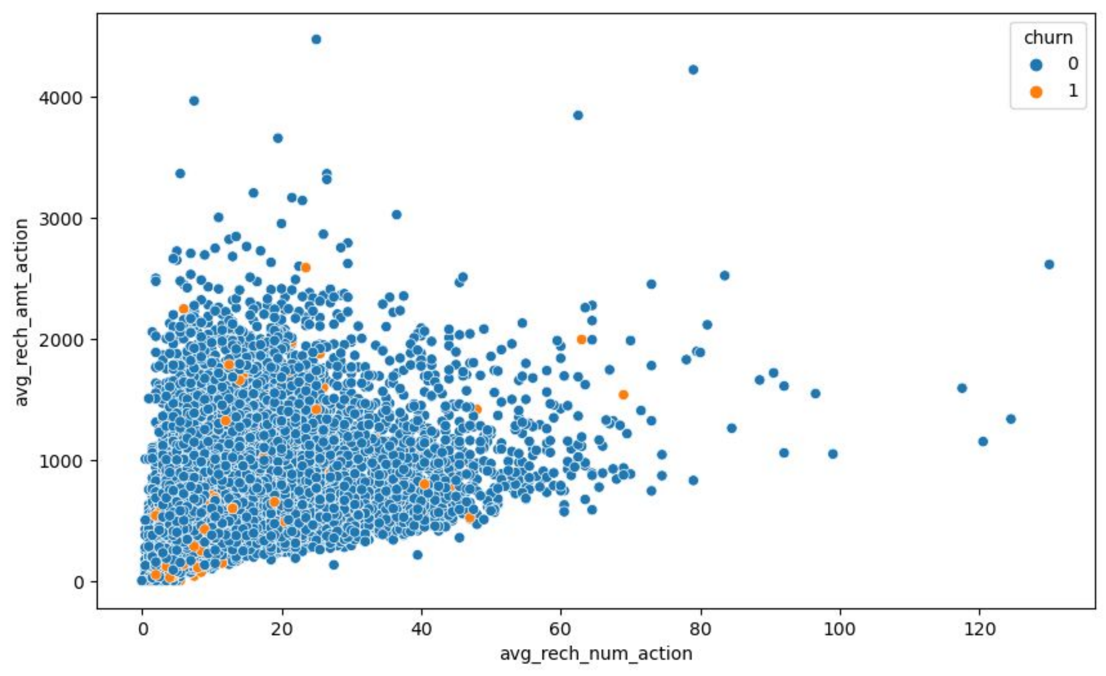
Bivariate Analysis - Recharge Amount & Volume-Based Cost

Insight: A higher churn rate is seen when the **recharge amount decreases** and the **volume-based cost increases** in the action phase.



Bivariate Analysis - Recharge Patterns

Insight: The number of recharges and the recharge amount are proportional. More recharges lead to a higher recharge amount.



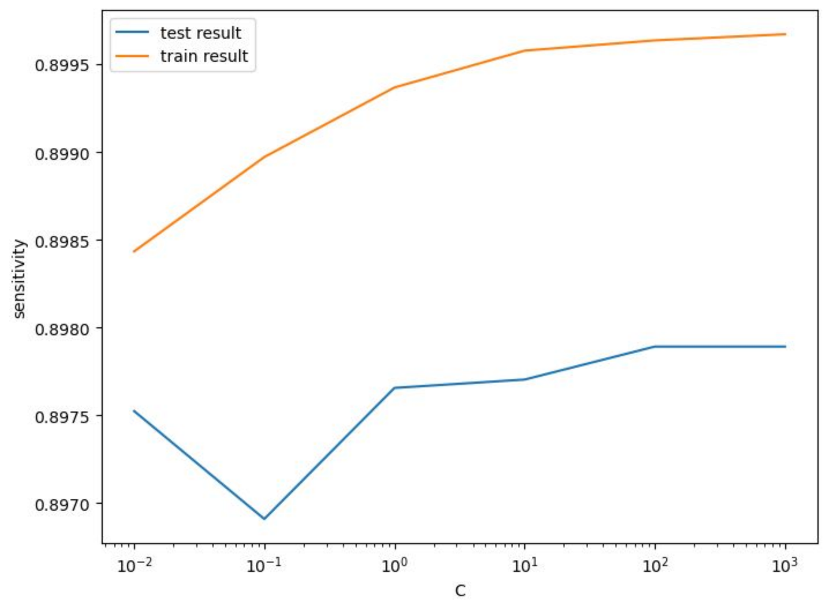
PCA Model - Emphasizing Sensitivity/Recall

Model Goal: Focused on increasing **Sensitivity/Recall** over Accuracy.

Rationale: Retaining high-risk churn customers is the goal. Missing some non-churn customers is acceptable if it helps identify potential churners.

Logistic Regression Model (Without PCA)

Insight: Some features have high p-values, making them insignificant. Feature elimination was performed using Recursive Feature Elimination (RFE) and manual tuning.



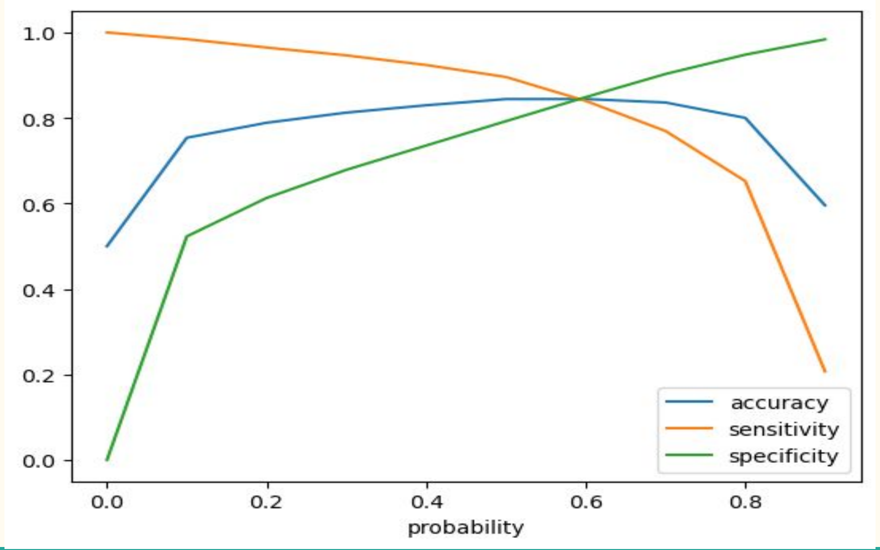
Conclusion

Key Findings:

- High churn rate is associated with reductions in usage, recharge number, and recharge amount in the action phase.
- Volume-based cost increases also drive churn.
- Higher ARPU and MOU customers are less likely to churn.

Without PCA

- **Model Performance**:
  - The **logistic regression model without PCA** shows good sensitivity and accuracy, comparable to models with PCA.
  - This simpler model effectively identifies important predictor variables while maintaining clarity.
- **Relevance**:
  - The logistic model without PCA highlights which variables are significant and can be acted upon for churn prediction.
  - It offers better interpretability, making it more suitable for business decision-making.



Recommendations

1. **Target Customers with Lower Usage**:
    - Customers with reduced **minutes of usage** for **incoming local calls** and **outgoing ISD calls** in the action phase (August) are more likely to churn.
2. **Focus on Recharge and Usage Behavior**:
    - Customers with a decrease in **outgoing others charge in July** and **incoming others usage in August** should be targeted.
    - Customers with an **increased volume-based cost in the action phase** are at higher risk of churn.
3. **3G and 2G Data Usage**:
    - **High 3G recharge amounts in August** correlate with higher churn rates.
    - Customers with a **decrease in 2G usage in August** are more likely to churn.
4. **Roaming and STD Usage**:
    - Customers with **decreasing STD incoming minutes** to fixed lines in August are probable churners.
    - **Roaming outgoing minutes of usage (roam_og_mou_8)** has a positive coefficient, indicating that increased roaming usage leads to a higher likelihood of churn.

Key Focus Areas

- **Provide Offers** to:
  - Customers with declining usage in key services (local calls, ISD, 2G/3G data).
  - Customers showing increased volume-based cost during the action phase.
- **Preemptive Targeting**:
  - Proactively offer special plans to retain customers who display patterns of potential churn.

Thank you