Telecom Churn Case Study

retaining high profitable customers is the number one business goal.

Business Problem Overview

- In the highly competitive telecom industry, customer churn is a significant challenge. Telecom companies face an average annual churn rate of 15-25%. Given the high cost of acquiring new customers compared to retaining existing ones, reducing churn has become a top priority.

- The Case Study focuses on analysing customer-level data from a leading telecom company to build predictive models for identifying high-risk churn customers and identifying key indicators of churn.

Understanding and Defining Churn

- Churn is the phenomenon of customers switching from one telecom service provider to another. It can be defined differently depending on the payment model:
- Postpaid Model: Customers inform the existing operator when switching, making it easy to identify churn.
- Prepaid Model: Customers can stop using services without notice, making churn prediction more challenging.
- This project is primarily based on the Indian and Southeast Asian markets, where the prepaid model is prevalent.

- Churn can be defined in various ways:
  - Revenue-Based Churn: Customers not utilizing revenue-generating facilities for a specified time.
  - Usage-Based Churn: Customers with zero usage (incoming or outgoing) for a certain period.
- This project uses the usage-based definition to identify churn.
- High-Value Churn : In the Indian and Southeast Asian markets, approximately 80% of revenue comes from the top 20% of customers (high-value customers). Reducing churn among high-value customers can significantly reduce revenue loss.

Project Goal

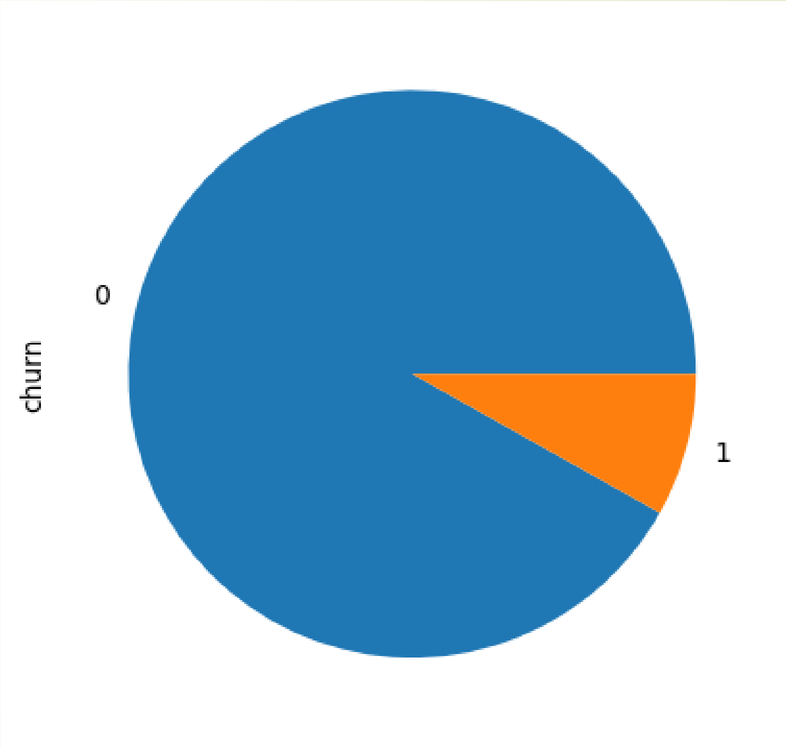
- The primary goal of the project is to develop predictive models that can:
- Identify high-value customers at risk of churn.
- Determine the key factors influencing churn.
- Help telecom companies take proactive measures to retain high-value customers and reduce revenue leakage.

Project Steps

- 1. **Data Preprocessing:** Cleaning and preparing the dataset for analysis.
- 2. **Exploratory Data Analysis (EDA):** Exploring the data to understand customer behavior and patterns.
- 3. **Feature Engineering:** Creating relevant features for modeling.
- 4. **Model Building:** Developing predictive models to identify high-value churn customers.
- 5. **Model Evaluation:** Assessing the performance of the models using appropriate metrics.
- 6. **Interpretation:** Identifying key indicators of churn.
- 7. **Recommendations:** Providing actionable insights to reduce churn among high-value customers.

EDA:
Churn Percentage

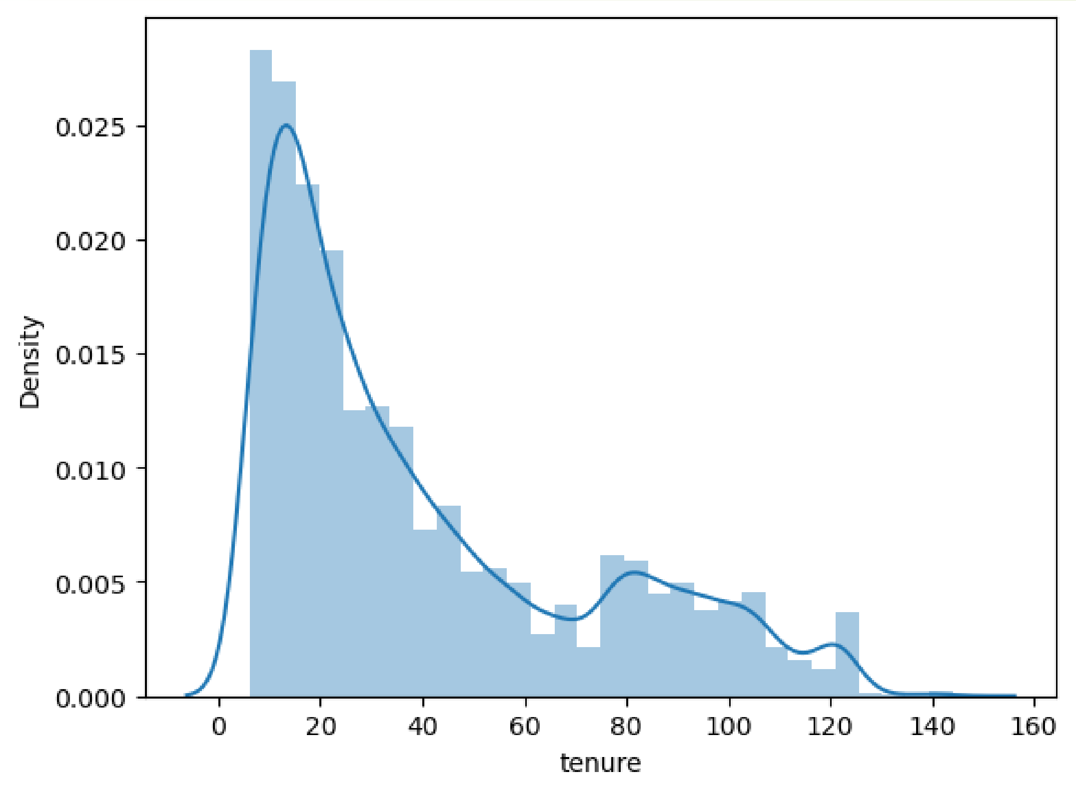
1. As can be observed from the above pie chart and results obtained that around 92% of the Customers Do Not churn, which might mean that there might be a class imbalance.



FEATURE ENGINEERING:
Checking the distribution of the tenure variable

1. The High tenure variable is more concentrated in the 0 to 25 tenure range.

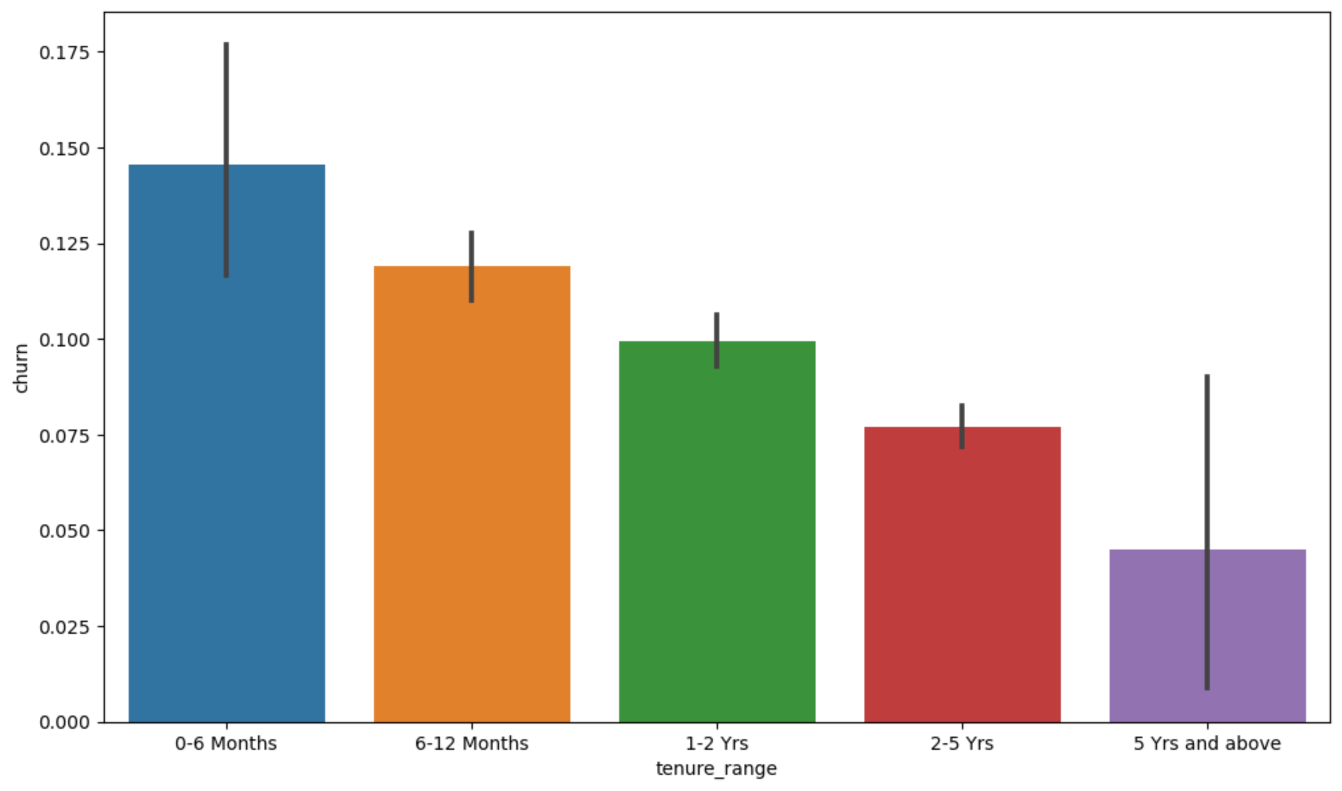
2. The density of the Customers churning at the tenure greater than 80 is significantly less than the density at 0 to 20.



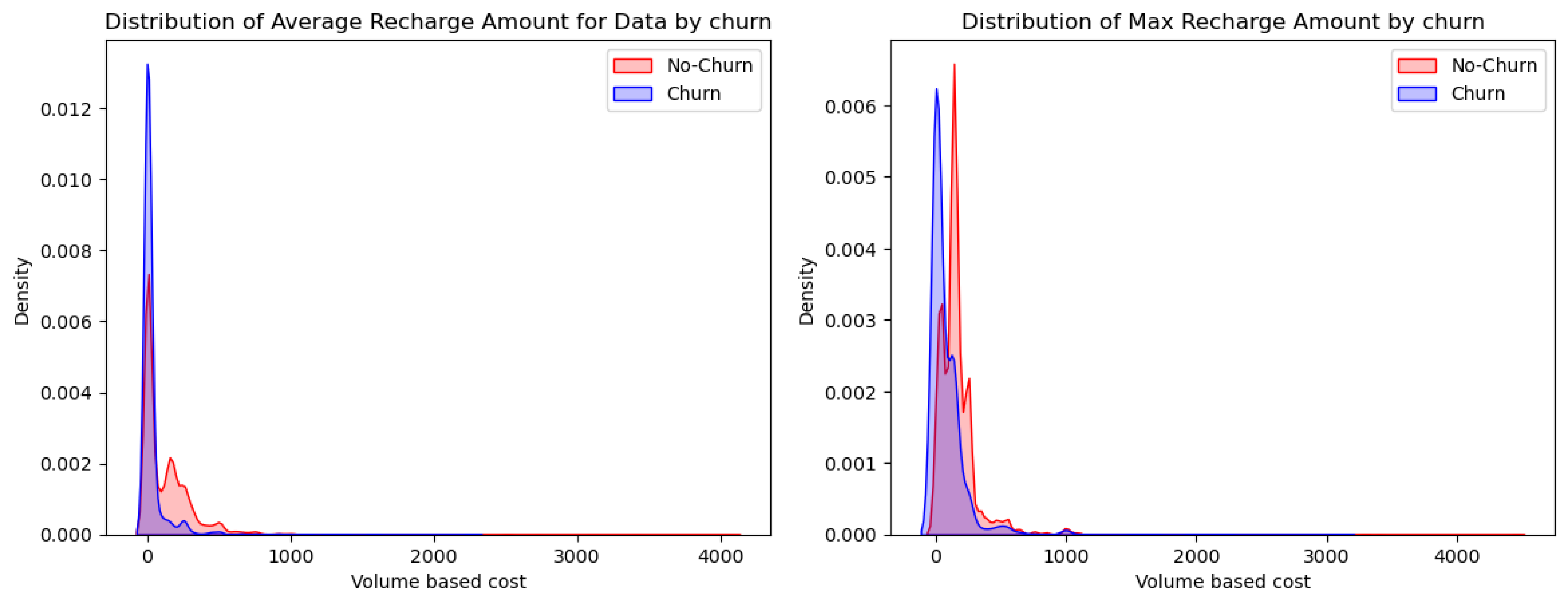
From the Bar Graph we can determine that:
1. The Highest Customer Churn is between 0 to 6 months.

2. The Lowest Customer Churn if for the period 5 years and above.

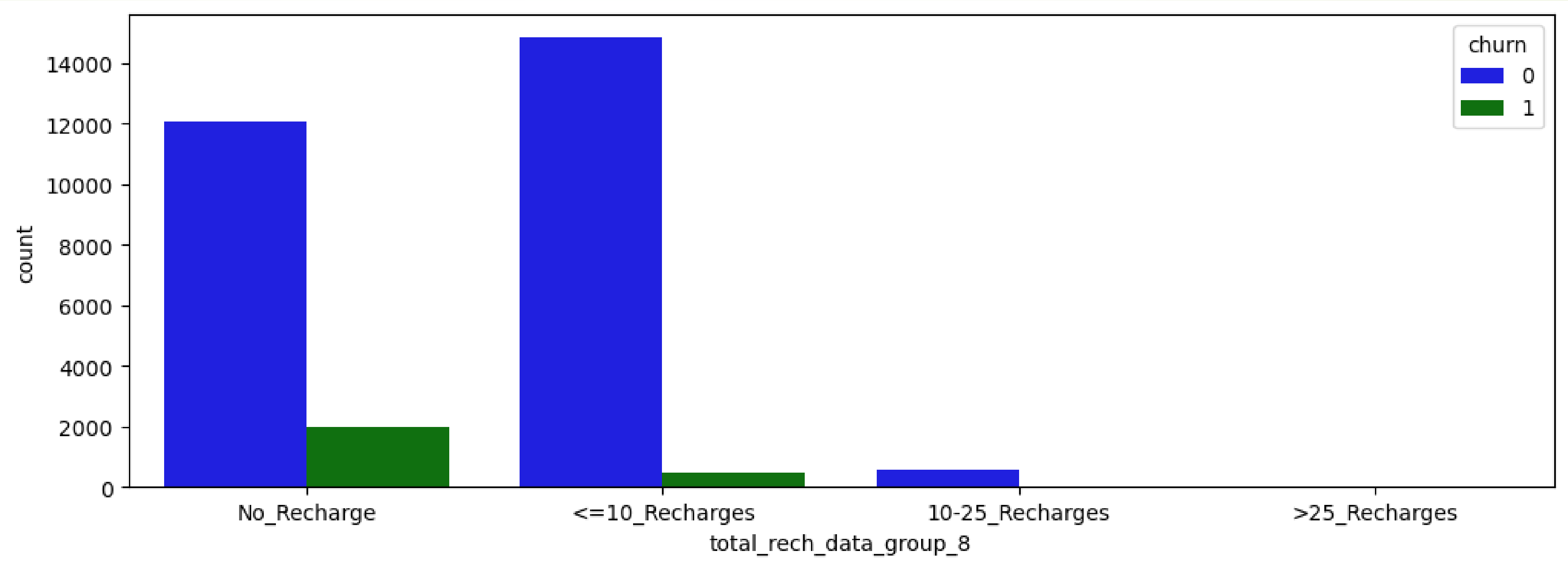
3. The Customer Churn slowly decreases as the time period for which they have been a customer of the Network



MODEL BUILDING AND EVALUATION: CHURN VS MAX RECHARGE AMT

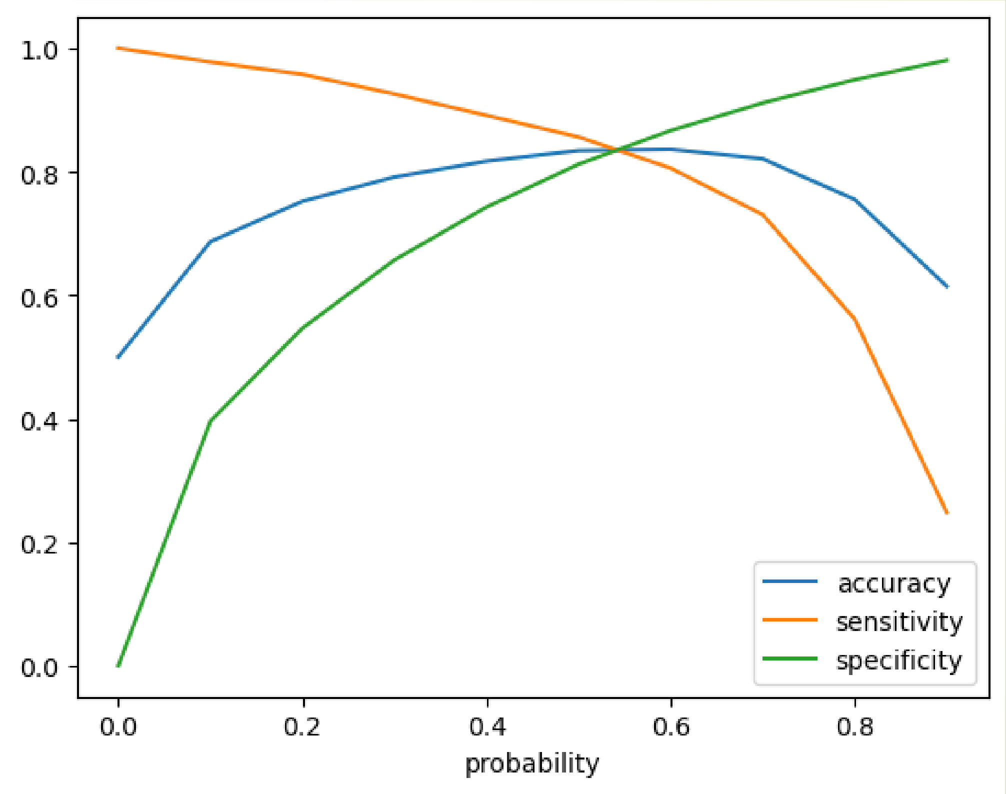


Plotting the first countplot



From the PLOT LINE we can observe that the optimum point of Classification lies between 0.5 and 0.6(nearby 0.6).

	probability	accuracy	sensitivity	specificity
0.0	0.0	0.500000	1.000000	0.000000
0.1	0.1	0.686696	0.977603	0.395790
0.2	0.2	0.751996	0.957538	0.546454
0.3	0.3	0.791321	0.925653	0.656989
0.4	0.4	0.816881	0.891176	0.742586
0.5	0.5	0.834042	0.856128	0.811956
0.6	0.6	0.836116	0.805682	0.866549
0.7	0.7	0.820795	0.730350	0.911240
0.8	0.8	0.755003	0.561230	0.948776
0.9	0.9	0.614294	0.248185	0.980402



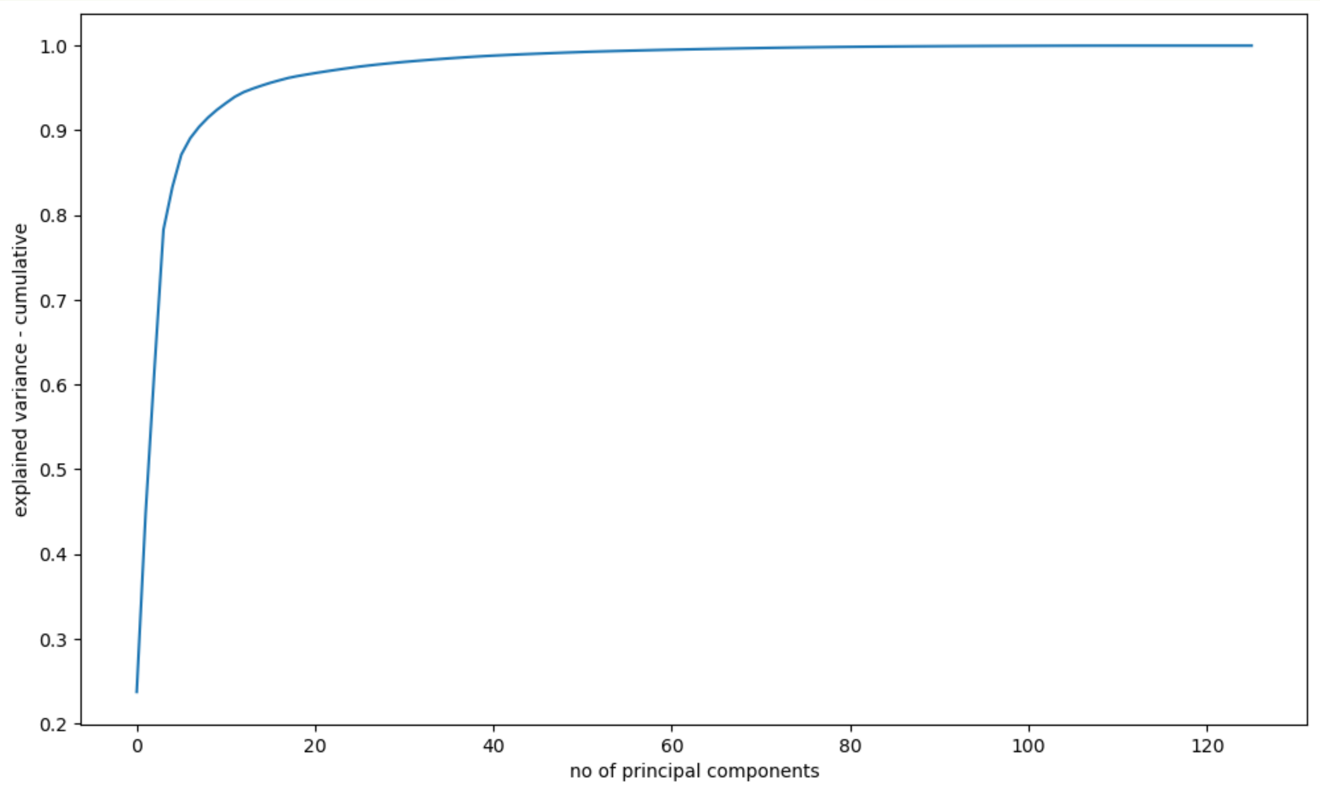
Values of Logistic Regression Model

```
Sensitivity =  0.8010973936899863
Specificity =  0.8293036750483559
False Positive Rate =  0.1706963249516441
Precision =  0.2925851703406814
True Negative Prediction Rate =  0.979300499643112
```

Tested Accuracy of the logistic regression model with PCA of the given values: 0.8179091212087546

The accuracy of the predicted model is: 83.0 % The sensitivity of the predicted model is: 80.0 %.

As the model created is based on a senstivity model, i.e. the True positive rate is given more importance as the actual and prediction of churn by a customer



Recommendations:

1. Improve Model Precision: The logistic regression model without PCA achieved an accuracy of approximately 83% but had a precision of only around 29%.
    1. This means that many customers predicted as churned were not actual churners. To reduce customer churn, focus on improving the precision of the model.
    2. This can be done by adjusting the classification threshold or using techniques like feature engineering to create more informative features.

2. Identify False Positives: Analyse the false positives (customers predicted as churned but didn't actually churn).

3. Investigate common characteristics among churning customers which can provide insights into areas where improvements can be made to retain these customers.

- Segmentation and Personalization based on usage patterns, demographics, or customer behaviour.
- Offer Incentives and Promotions with targeted offers, discounts, or promotions to customers at risk of churning.
- Customer Engagement by reaching out to customers with personalized messages, reminders, and recommendations based on their usage and preferences.
- Customer Feedback Monitoring
- Loyalty Programs and Quality of Service.