TELECOM CHURN CASE STUDY

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.

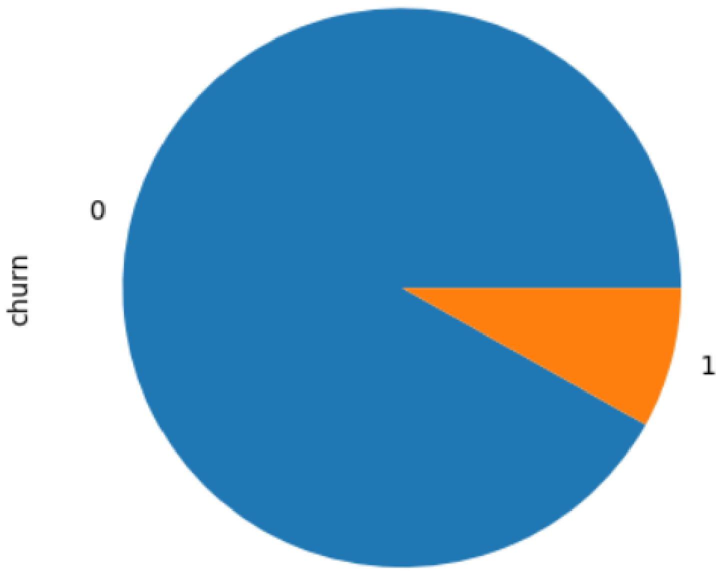
In this project, you will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

DATA VISUALIZATION & ANALYSIS

PIE CHART – Churn %

```
# Creating a pie chart to find the churn percentage

print((churn_df['churn'].value_counts()/len(churn_df))*100)
((churn_df['churn'].value_counts()/len(churn_df))*100).plot(kind="pie")
plt.show()
```

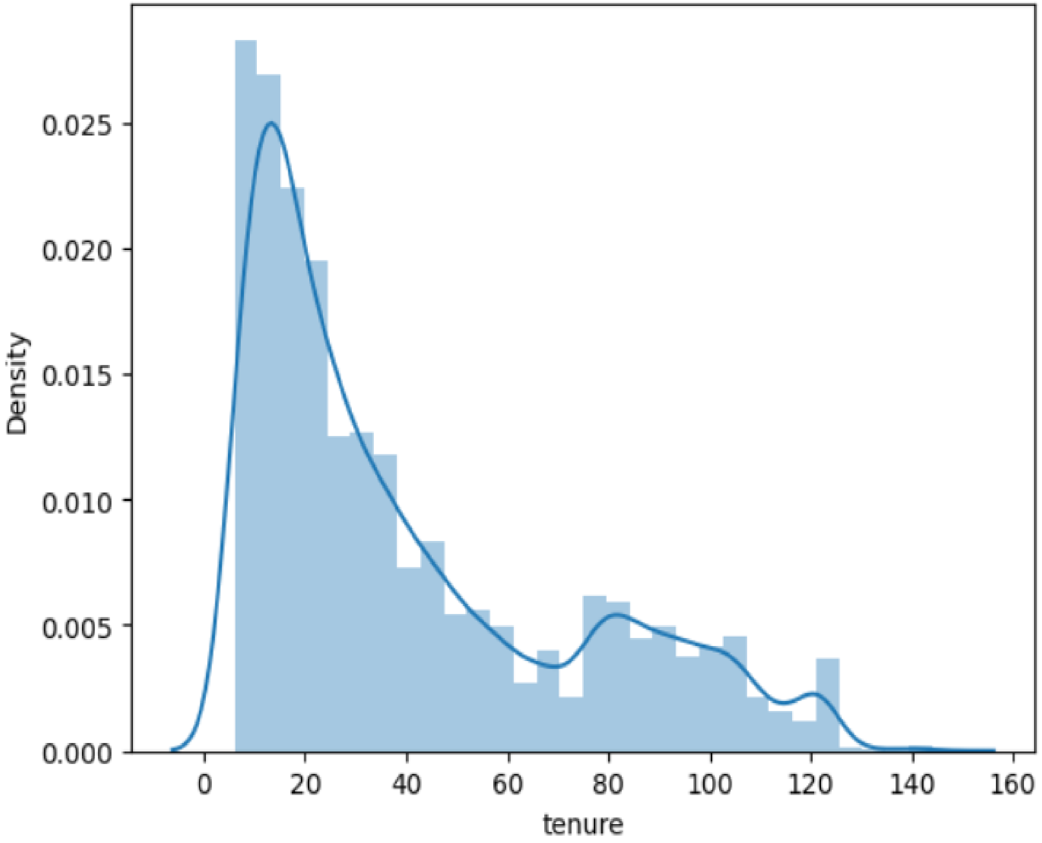
```
0    91.863605
1     8.136395
Name: churn, dtype: float64
```



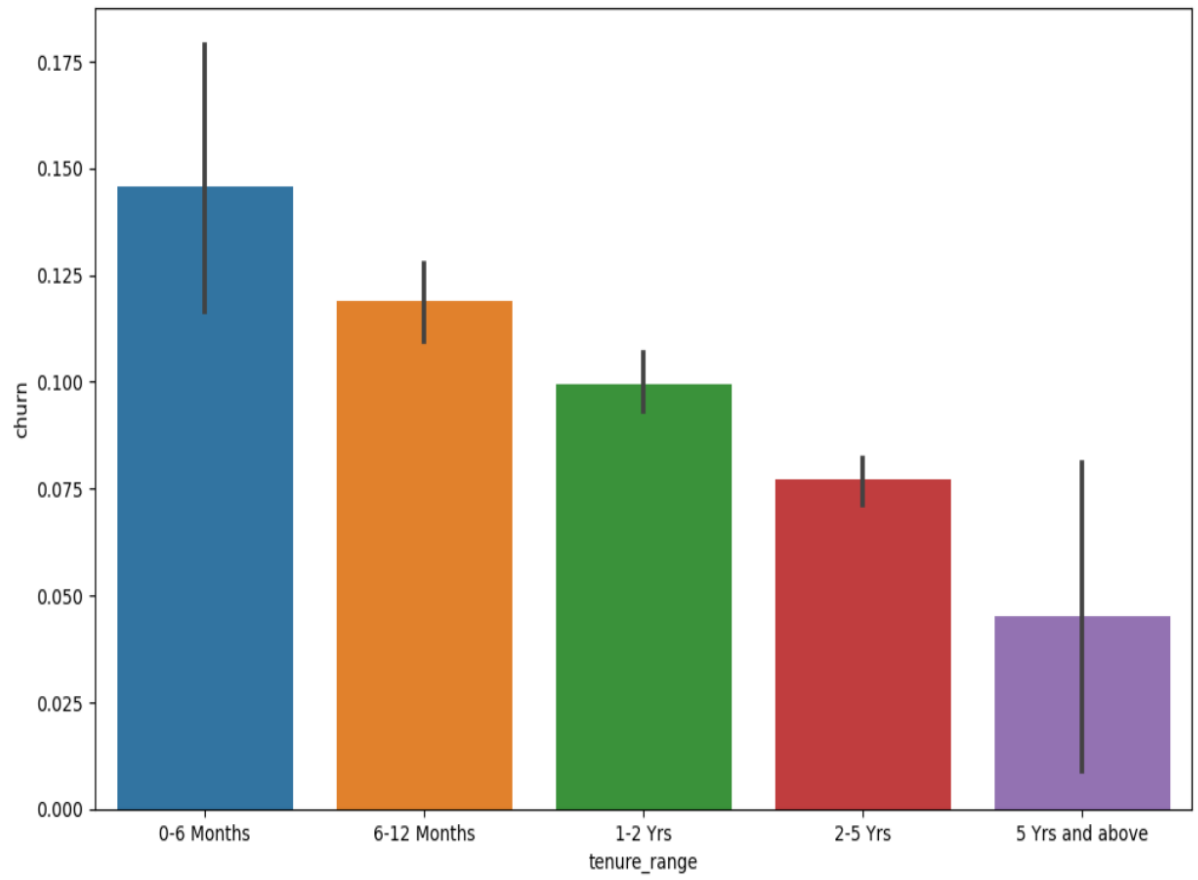
- The blue part in the pie chart shows that 91% of the customers do not churn.
- This indicates a possibility of class imbalance.
- Since the variable 'churn' is the target variable, the related columns can be dropped.

Density vs Tenure

```python
# Creating a dist plot with a bin size of 30

sns.distplot(churn_df['tenure'],bins=30)
plt.show()
```



Graph for Tenure Range



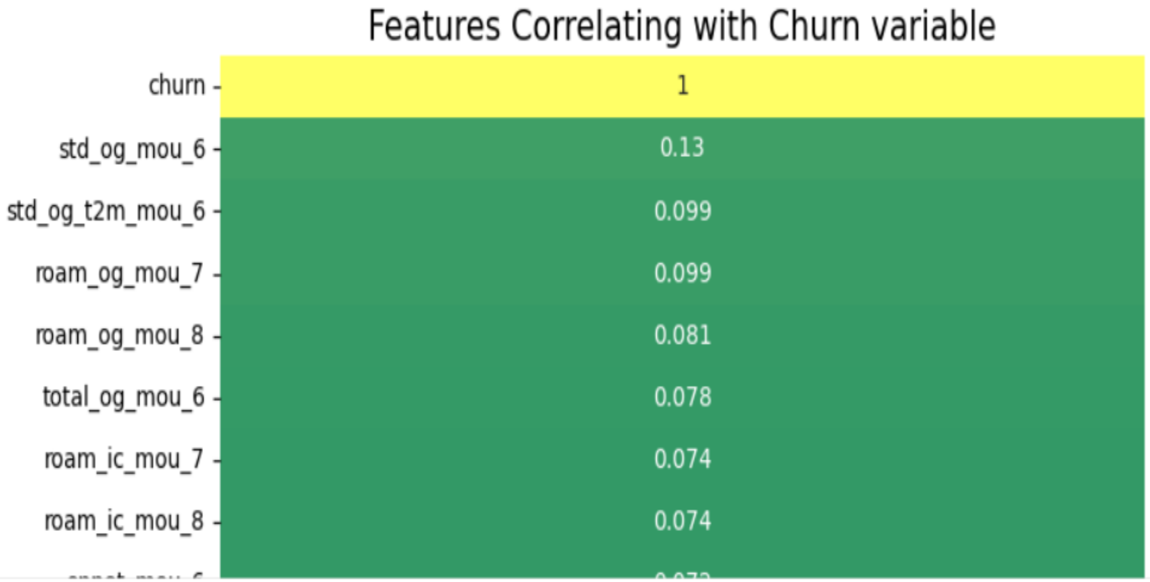
It can be seen that the maximum churn rate happens within 0-6 months and gradually decreases as the customer continues with the network

Heat Map-

Correlation between Target variable and other variables

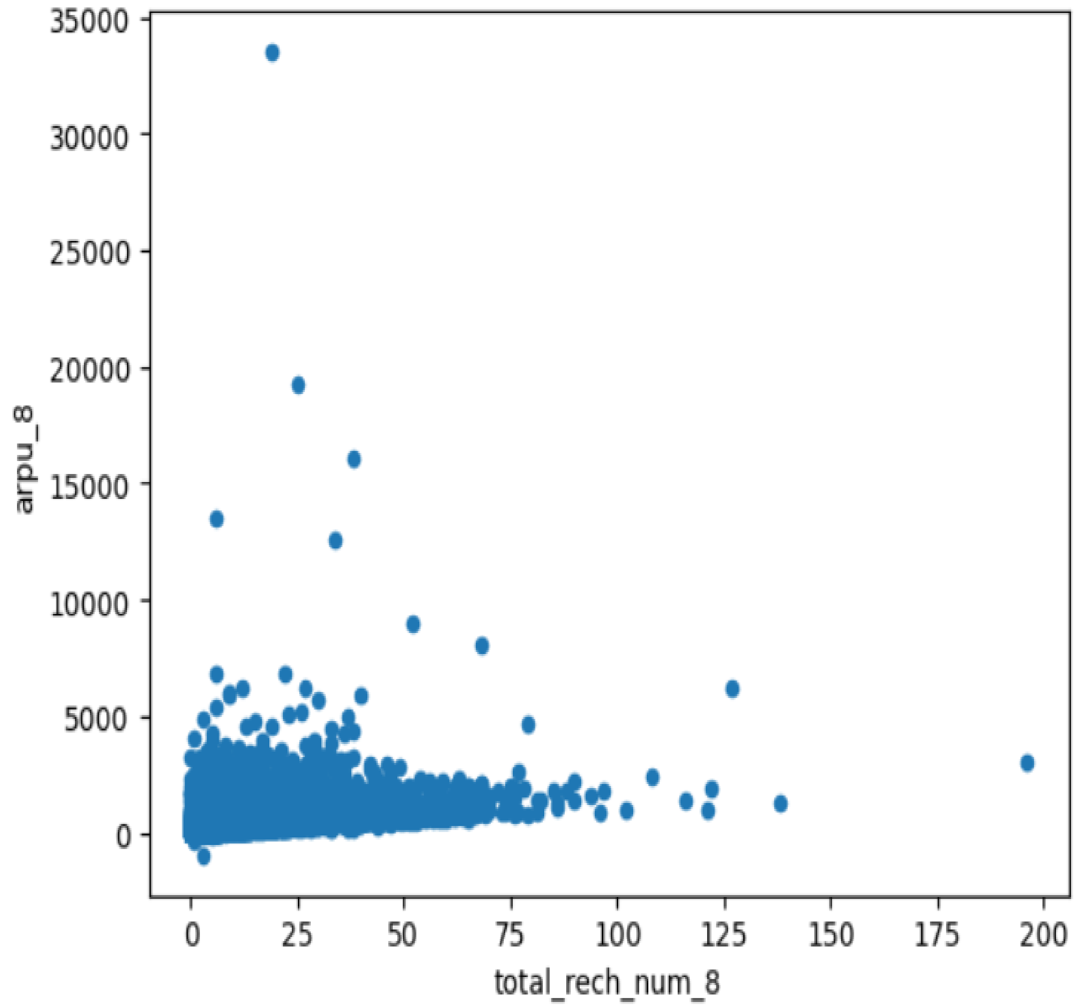
```python
# Correlation between target variable and other variables

plt.figure(figsize=(10,50))
heatmap_churn = sns.heatmap(churn_df.corr()[['churn']].sort_values(ascending=False, by='churn'),annot=True,
                            cmap='summer')
heatmap_churn.set_title("Features Correlating with Churn variable", fontsize=15)
```

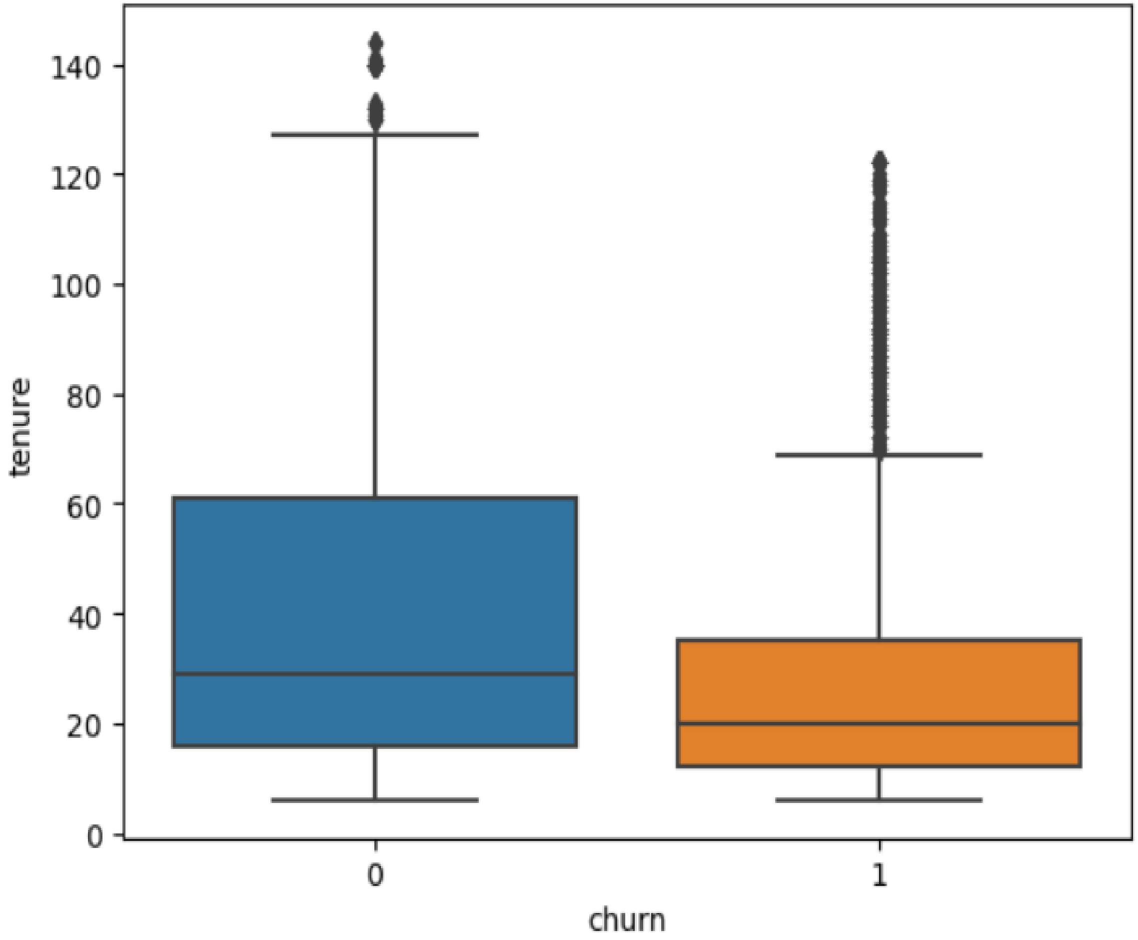
```
Text(0.5, 1.0, 'Features Correlating with Churn variable')
```



Inferences:

- Average Outgoing Calls & calls on roaming for 6 & 7th months are positively correlated with churn.
- Average Revenue, No. Of Recharge for 8th month has negative correlation with churn.

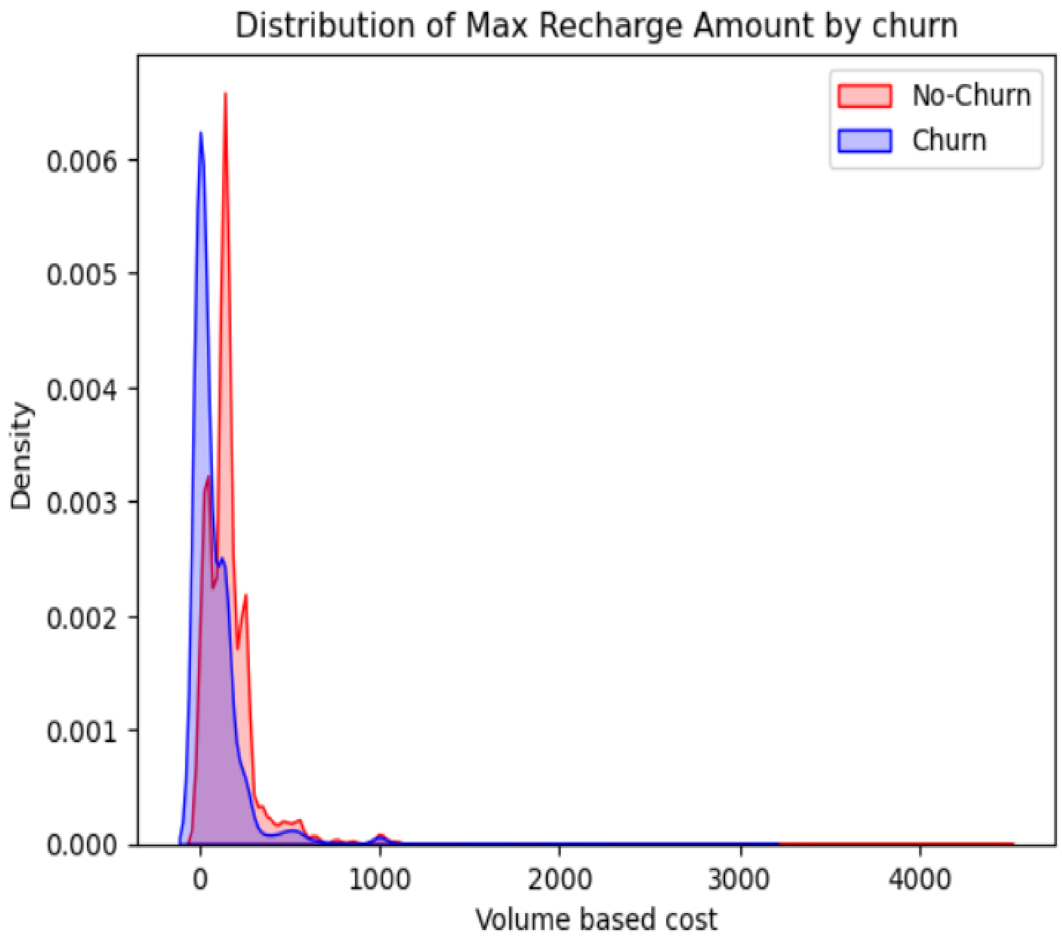

arpu_8
total_rech_num_8
35000
30000
25000
20000
15000
10000
5000
0
0
25
50
75
100
125
150
175
200



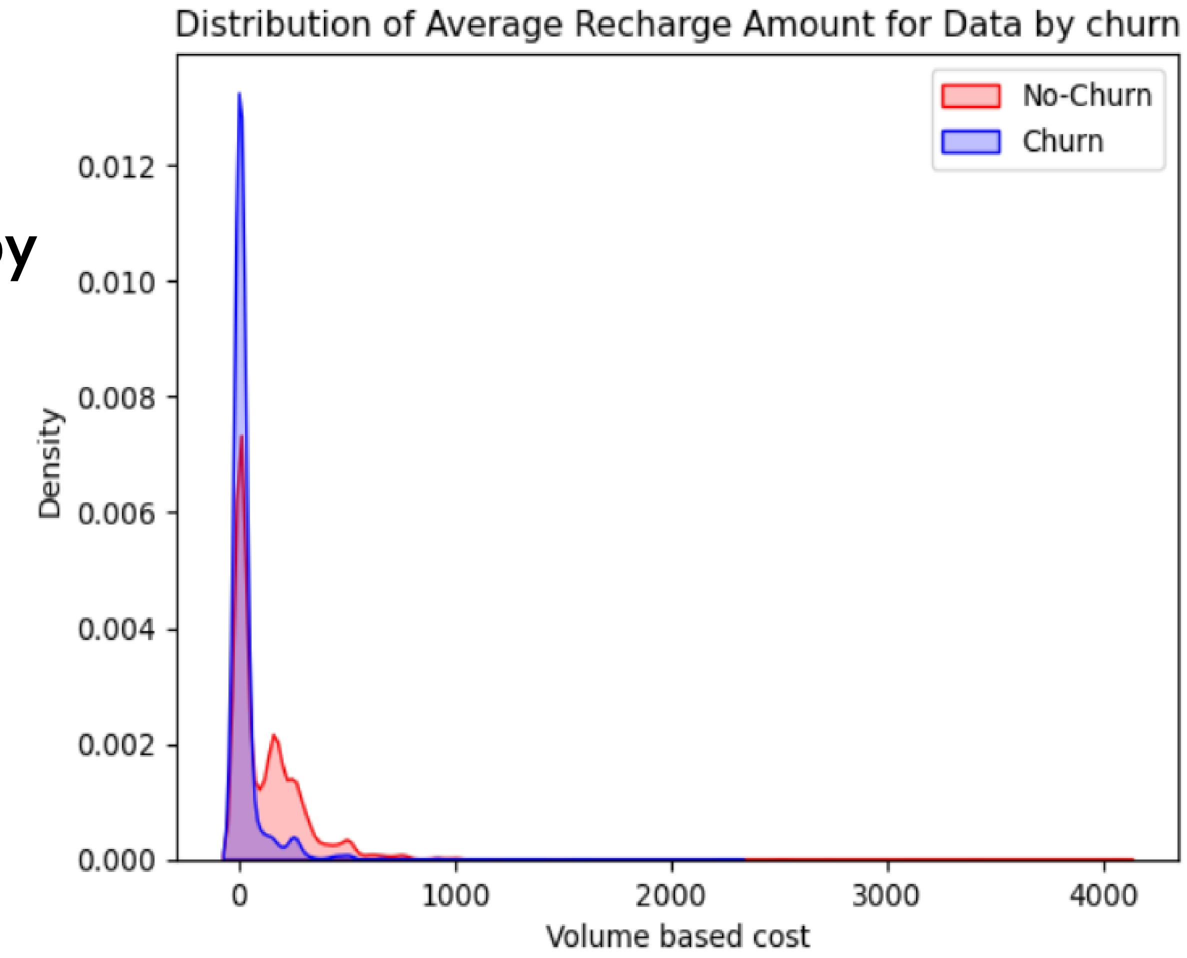
Inferences:

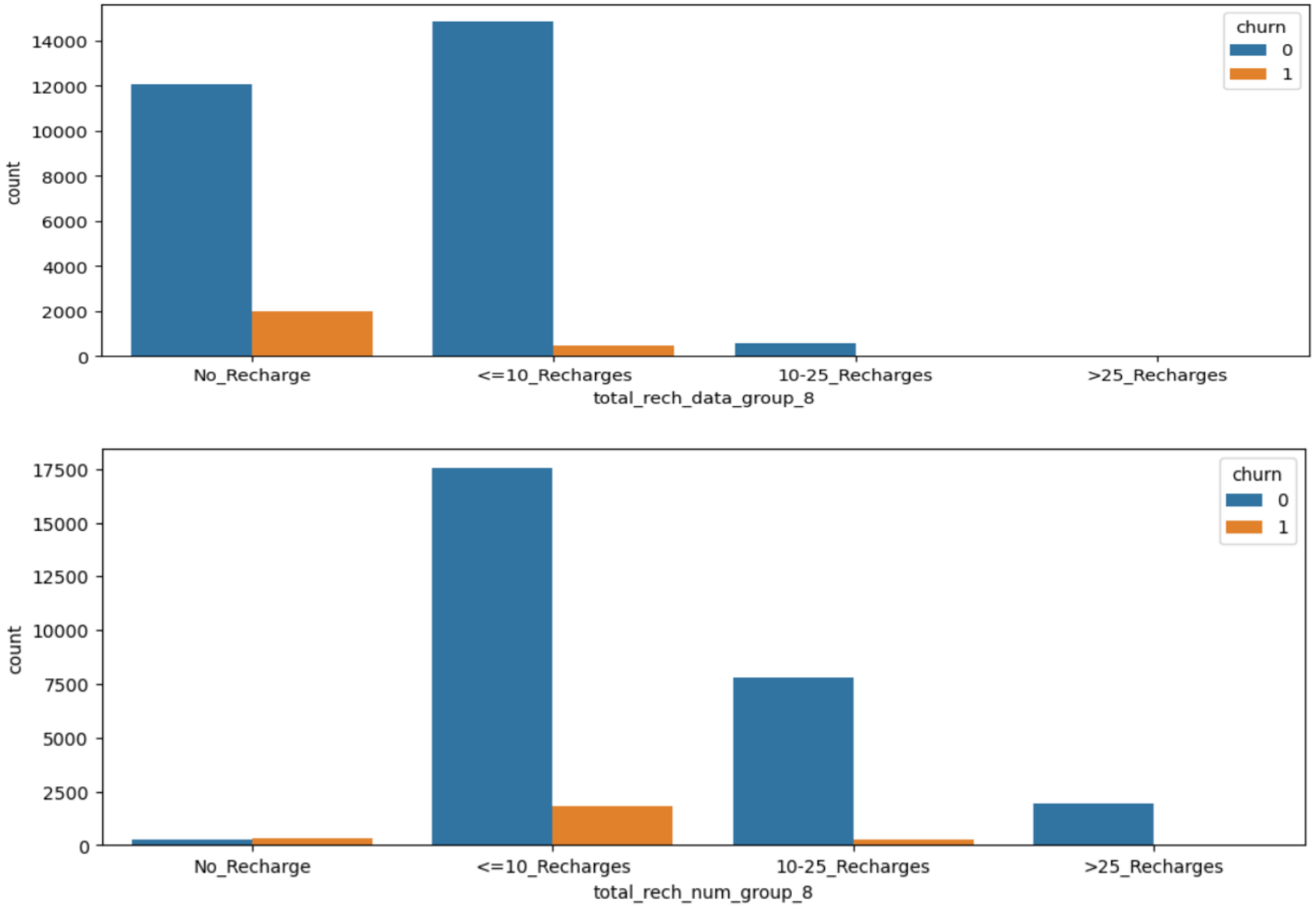
Tenured customers do not show a tendency to churn.

Distribution of Max Recharge Amount by churn



Distribution of Average Recharge Amount for Data by churn

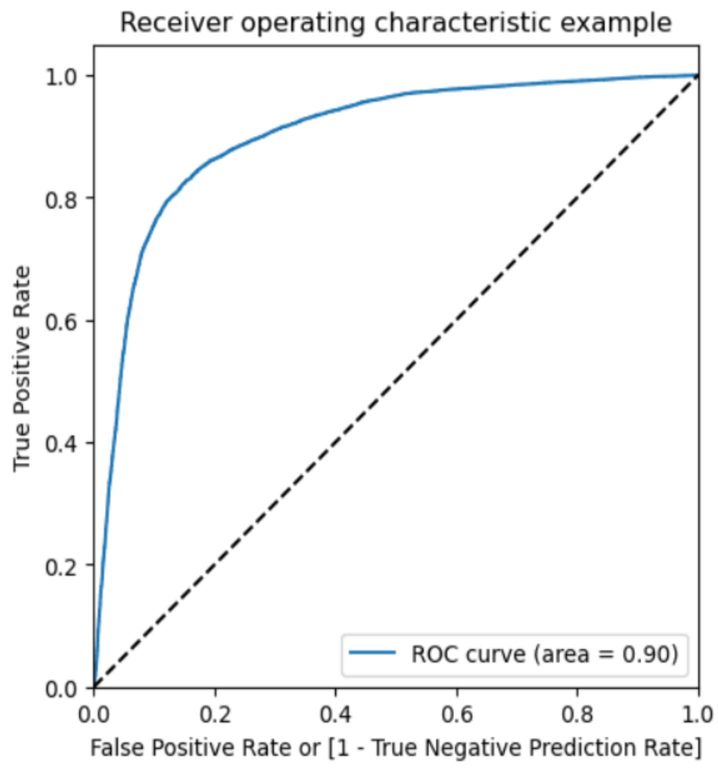




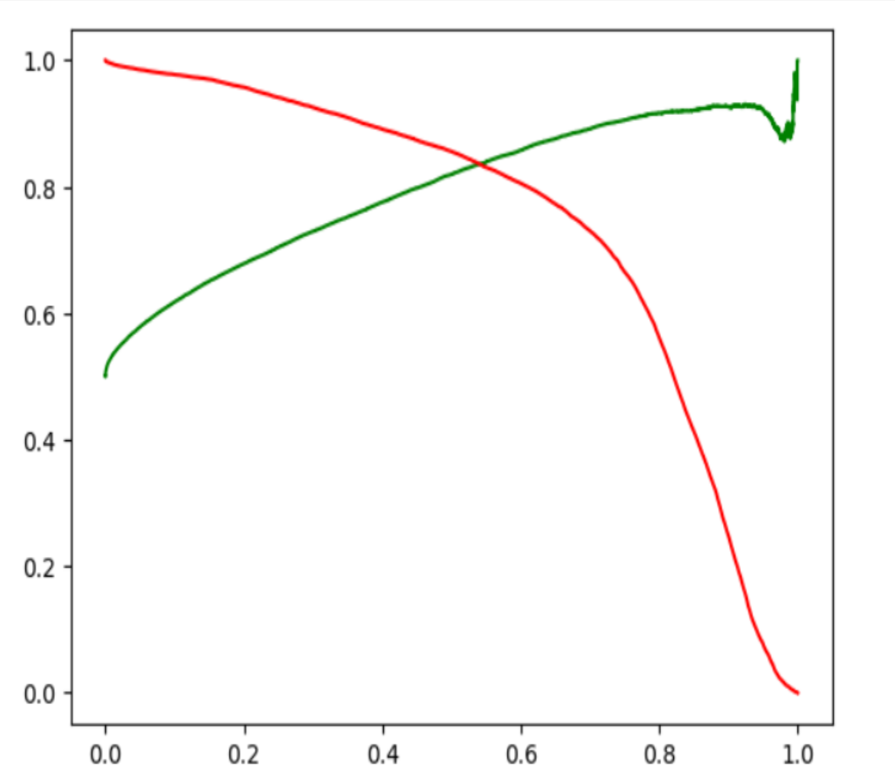
Inference from the above plot:

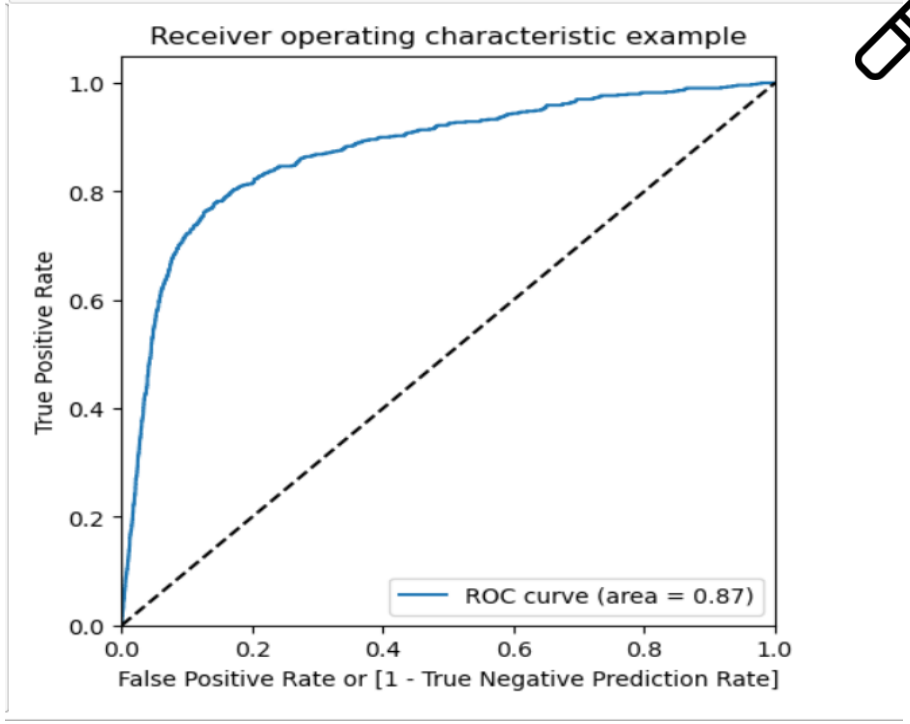
#The churn rate decreases as the rate of recharge increases for prepaid customers.

```
# Curve plotting

draw_roc(y_train_sm_pred_final.Converted, y_train_sm_pred_final.Converted_prob)
```



Curve for Receiver operating characteristic example

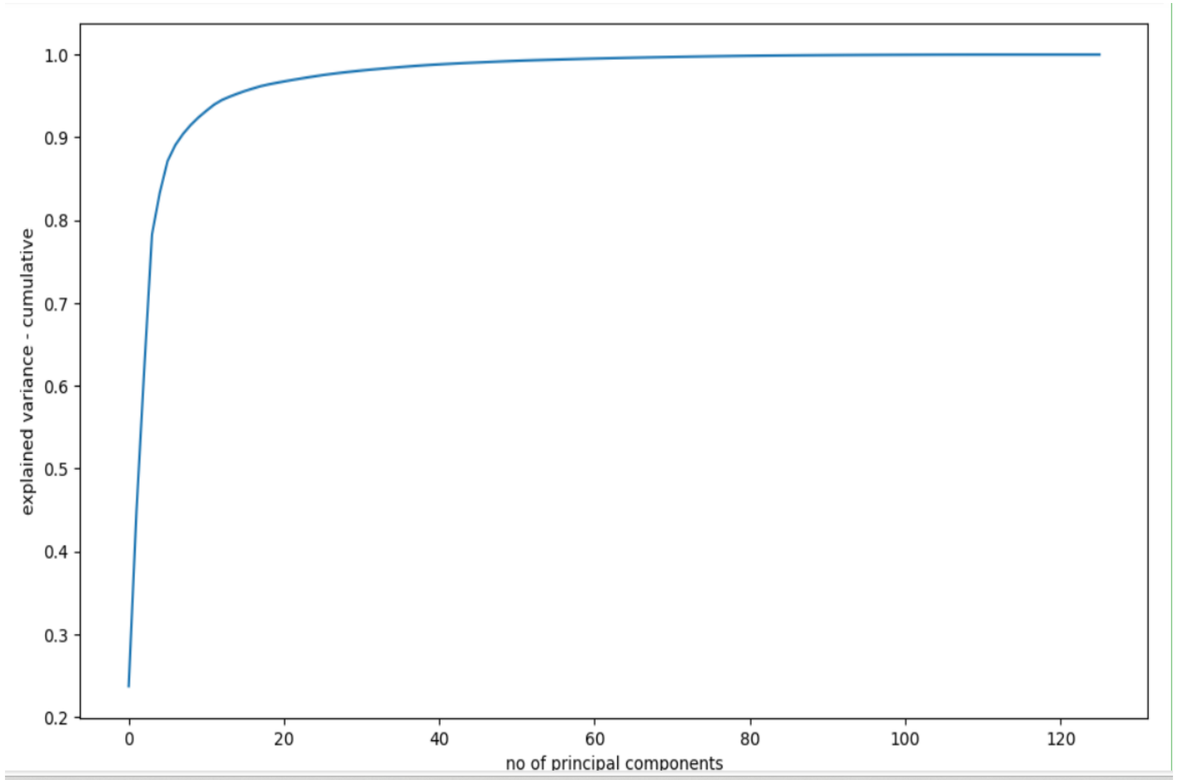




Inferences from the above curve:

The Area Under Curve(AUC) score for train dataset is 0.90 and the test dataset is 0.87.

This model can be considered as a good model.



Inference:

90% of the data can be explained with 8 PCA components

Inference: Accuracy of the Logistic Regression Model with PCA is 75.5%

THANK YOU