

# **IIIT - BANGALORE TELECOM CHURN CASE STUDY**

By:

Shubham Gupta, Manisha Vinulu and Sumaiyya Khan

# OBJECTIVE



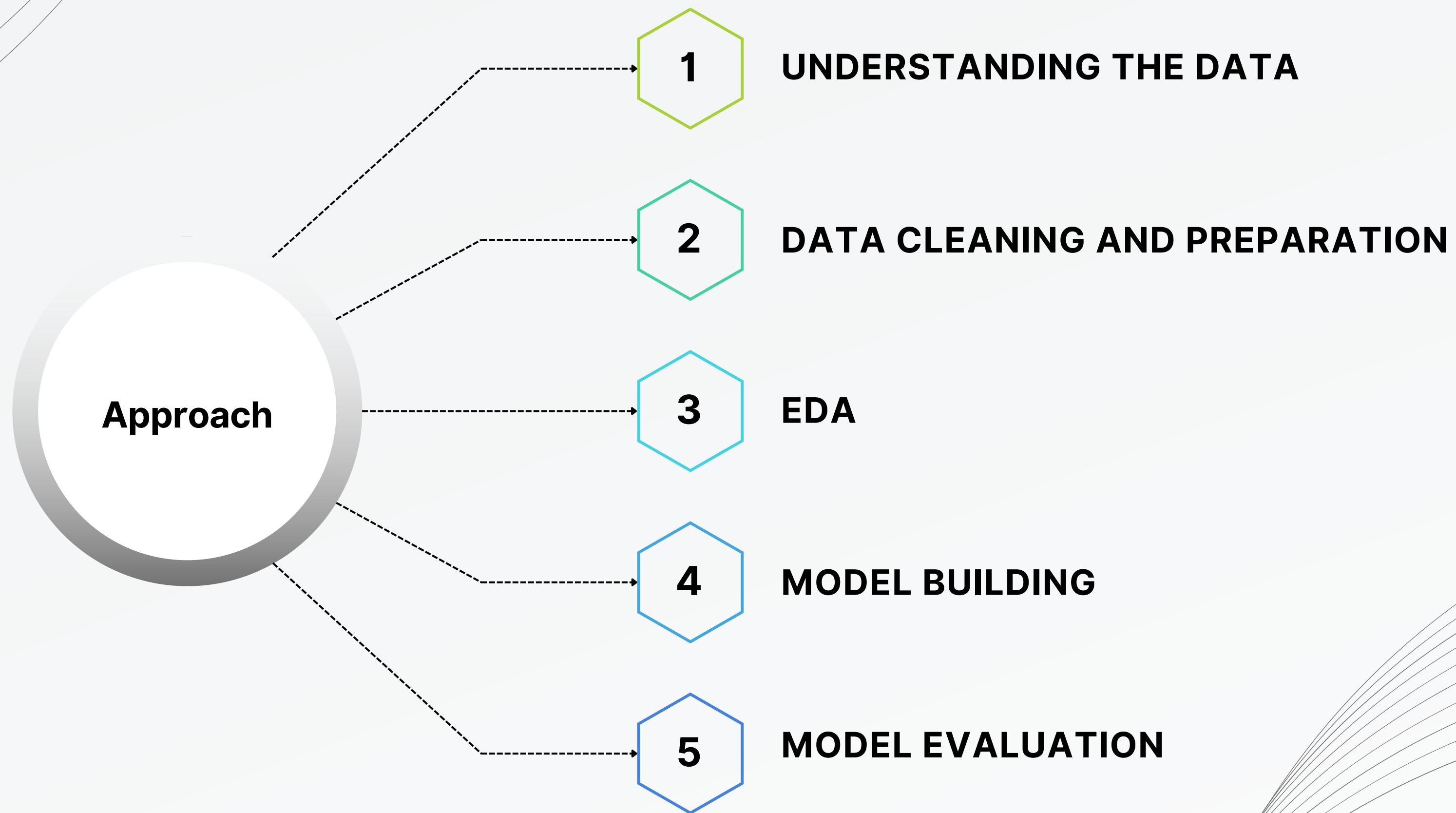
In the telecom industry, customers switch between multiple service providers which leads to rise in churn rate.



To reduce customer churn, we need to predict which customers are at high risk of churn.



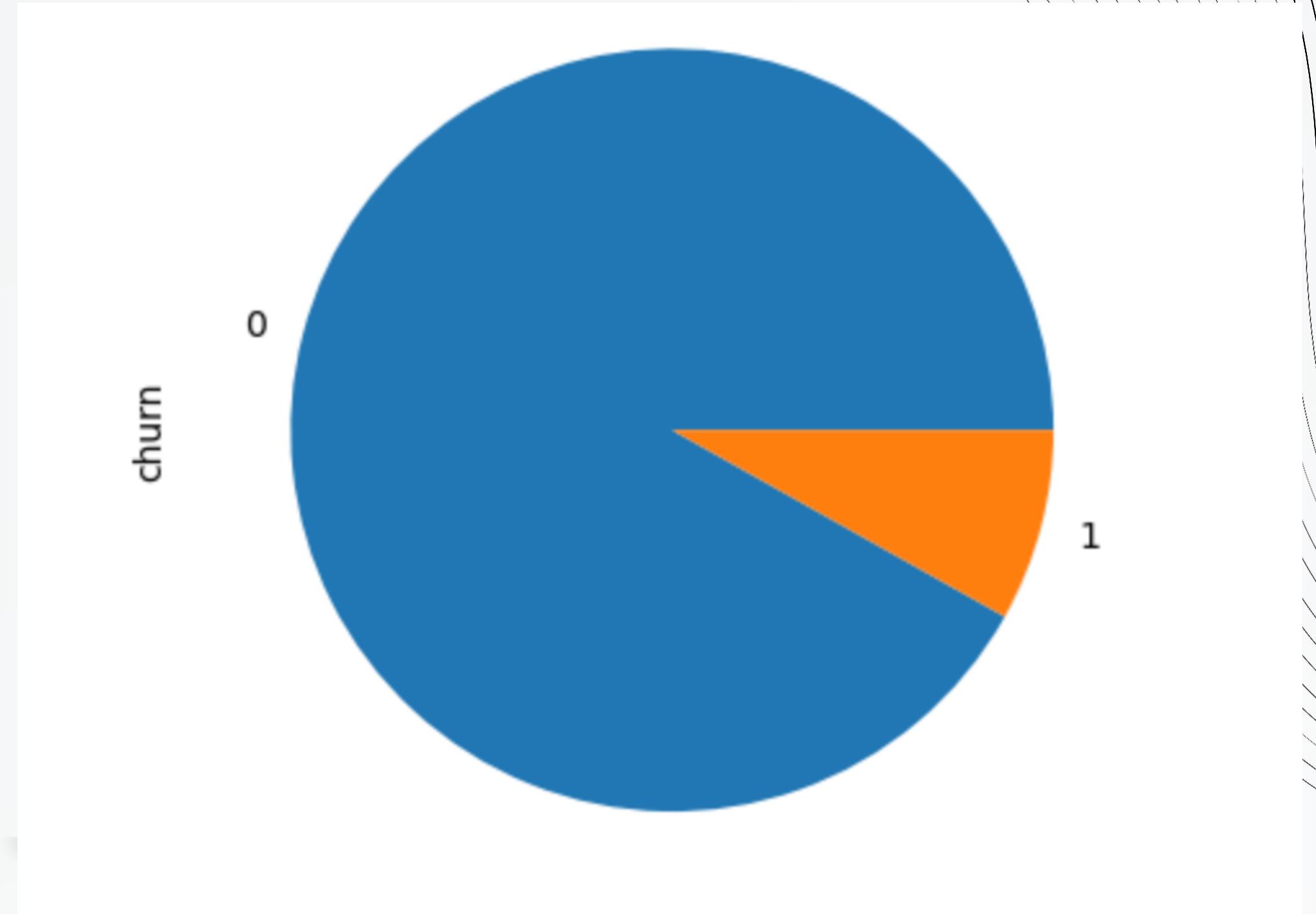
# APPROACH FOR CASE STUDY



# DATA IMBALANCE

# DATA IMBALANCE

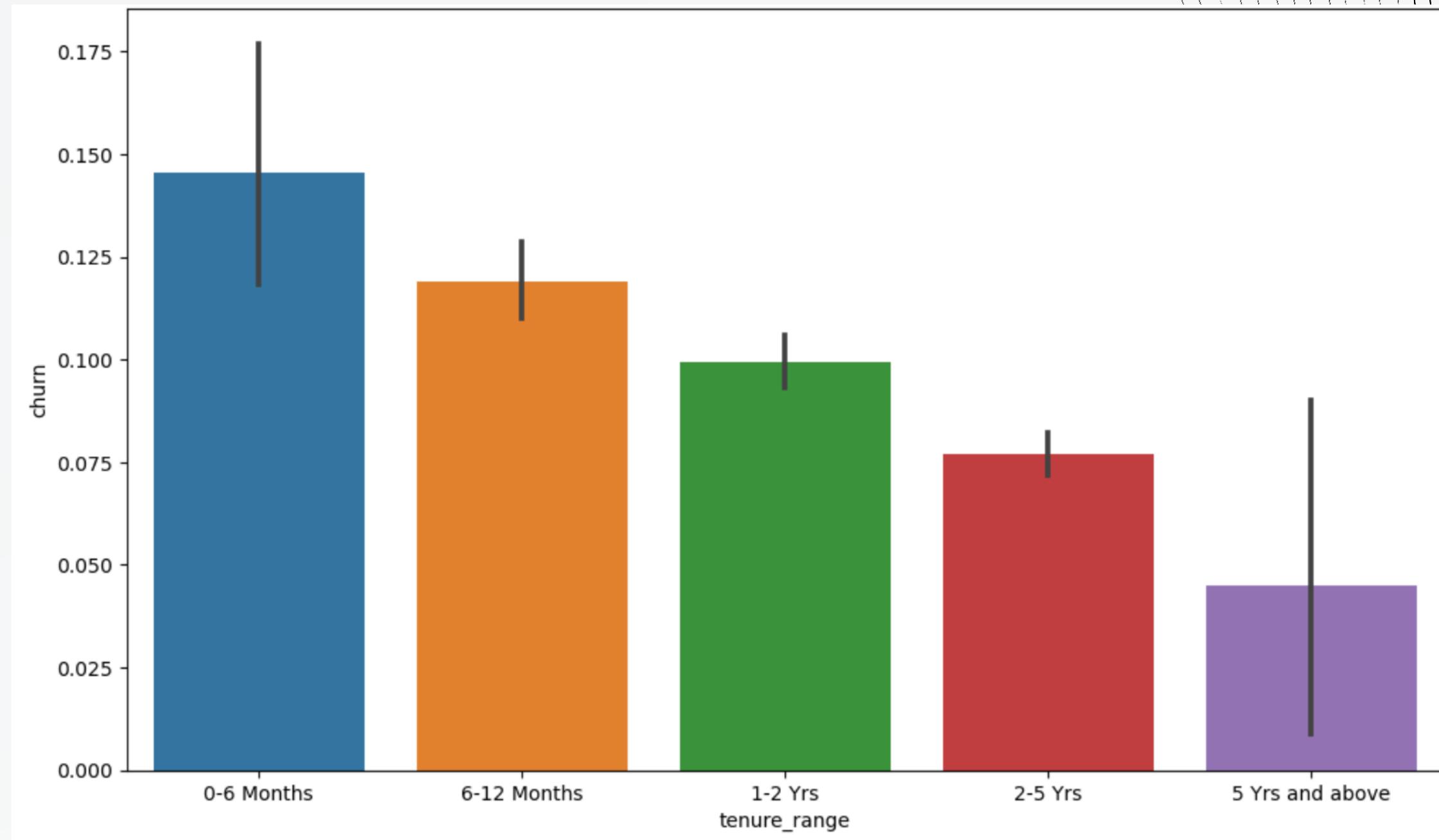
- 91% of the customers here do not churn.
- The churn rate is around 8.13%



# DATA VISUALIZATION

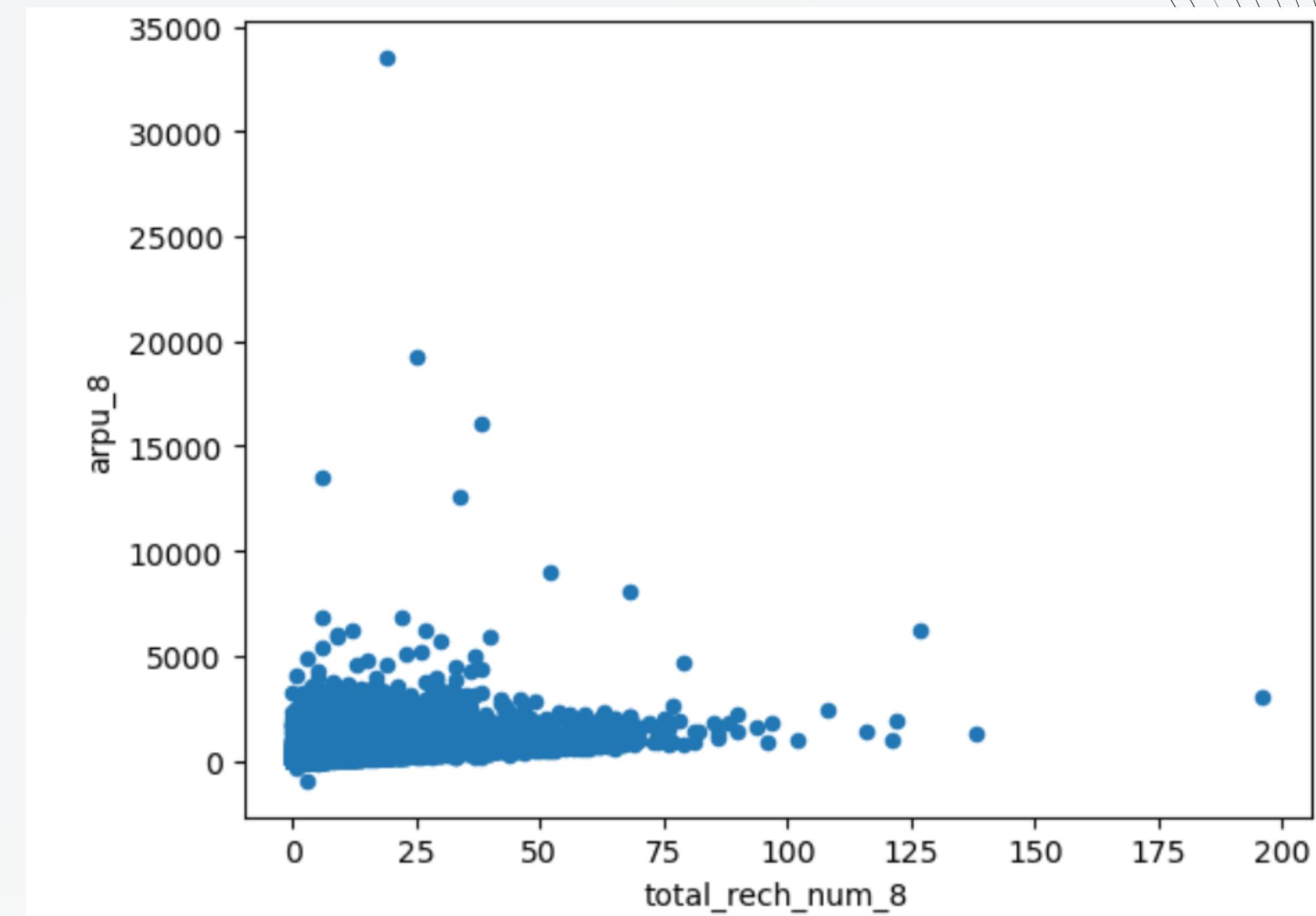
# TENURE VS CHURN

- Maximum churn is between 0-6 months.
- The churn rate decreases with an increase in tenure.



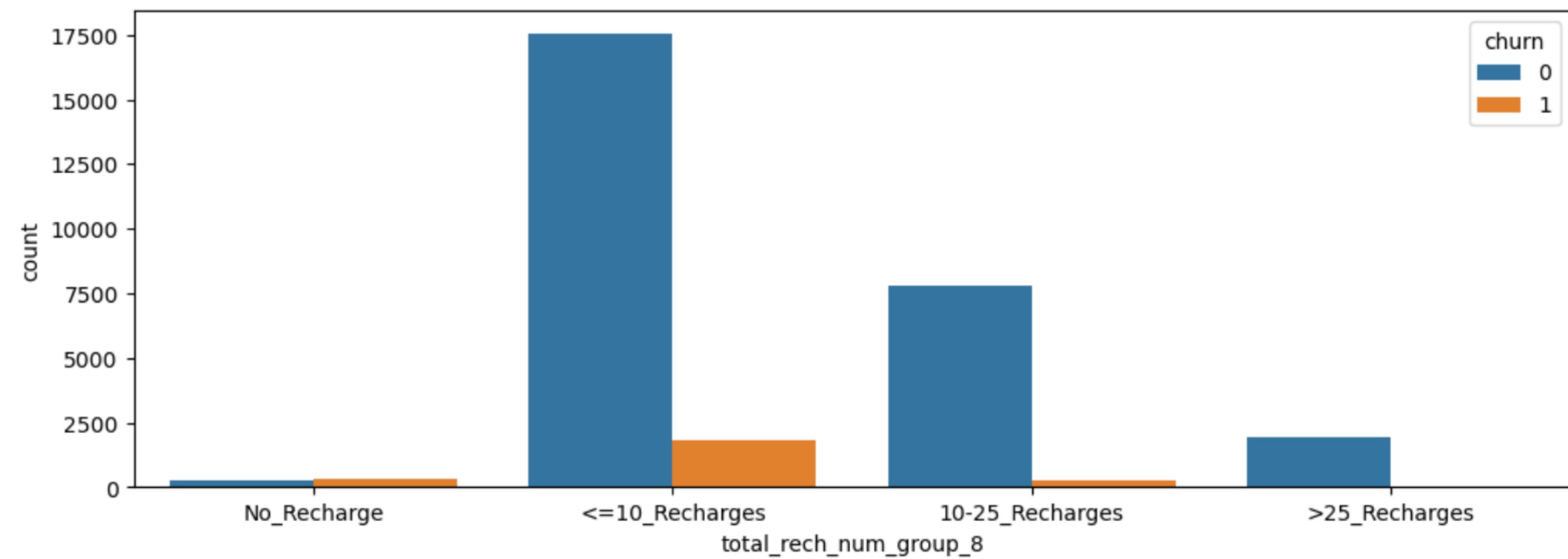
# TOTAL RECHARGE VS ARPU

- The volume of recharge for 8th month is high at lower recharge rates.
- This volume is generating more revenue than the customers with high recharge amount.



# TOTAL RECHARGE VS COUNT

- Increase in total recharge leads to decrease in churn
- The highest number of customers who do not churn are customers with less than or equal to 10 recharges

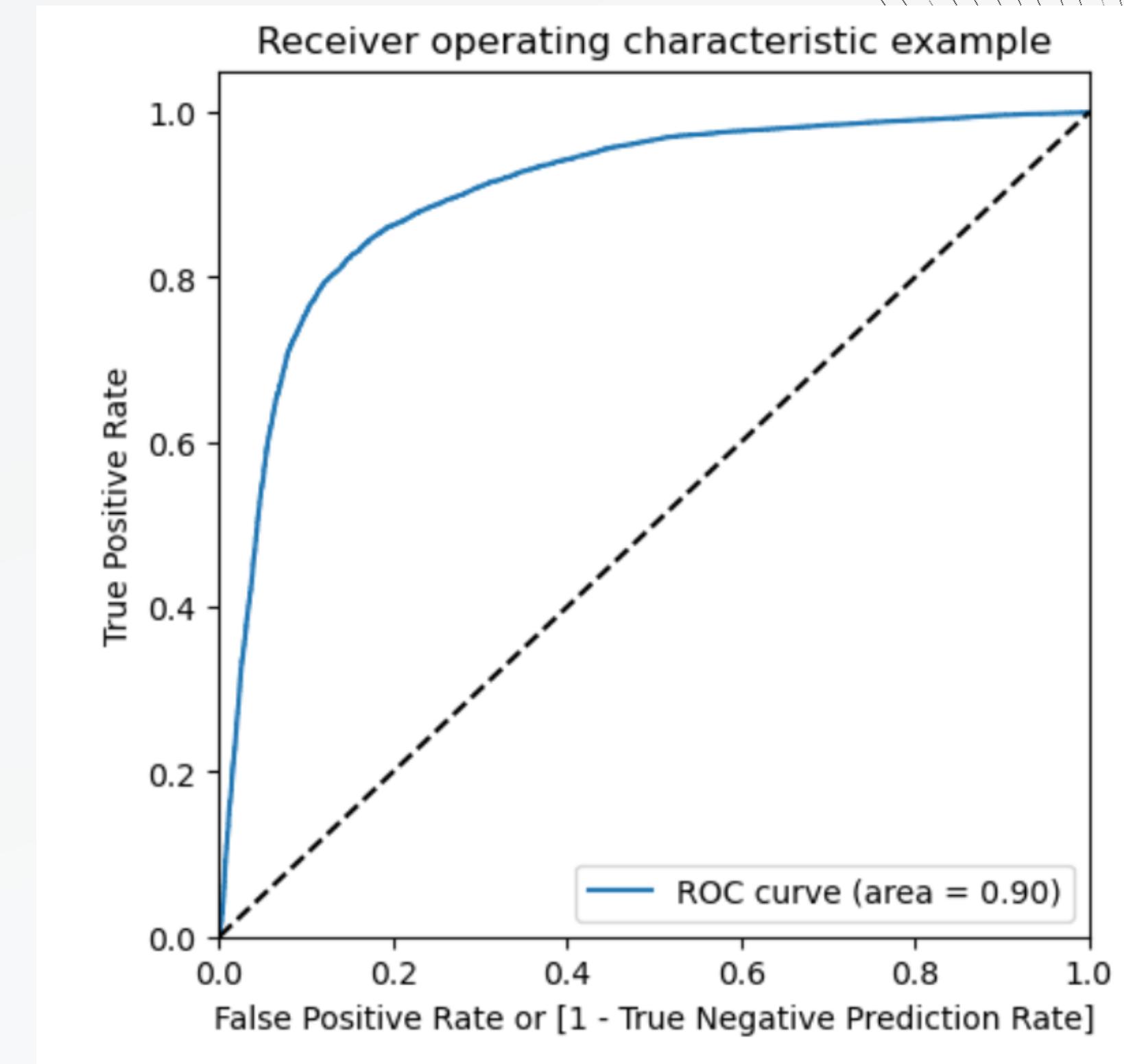




# **MODEL BUILDING & EVALUATION**

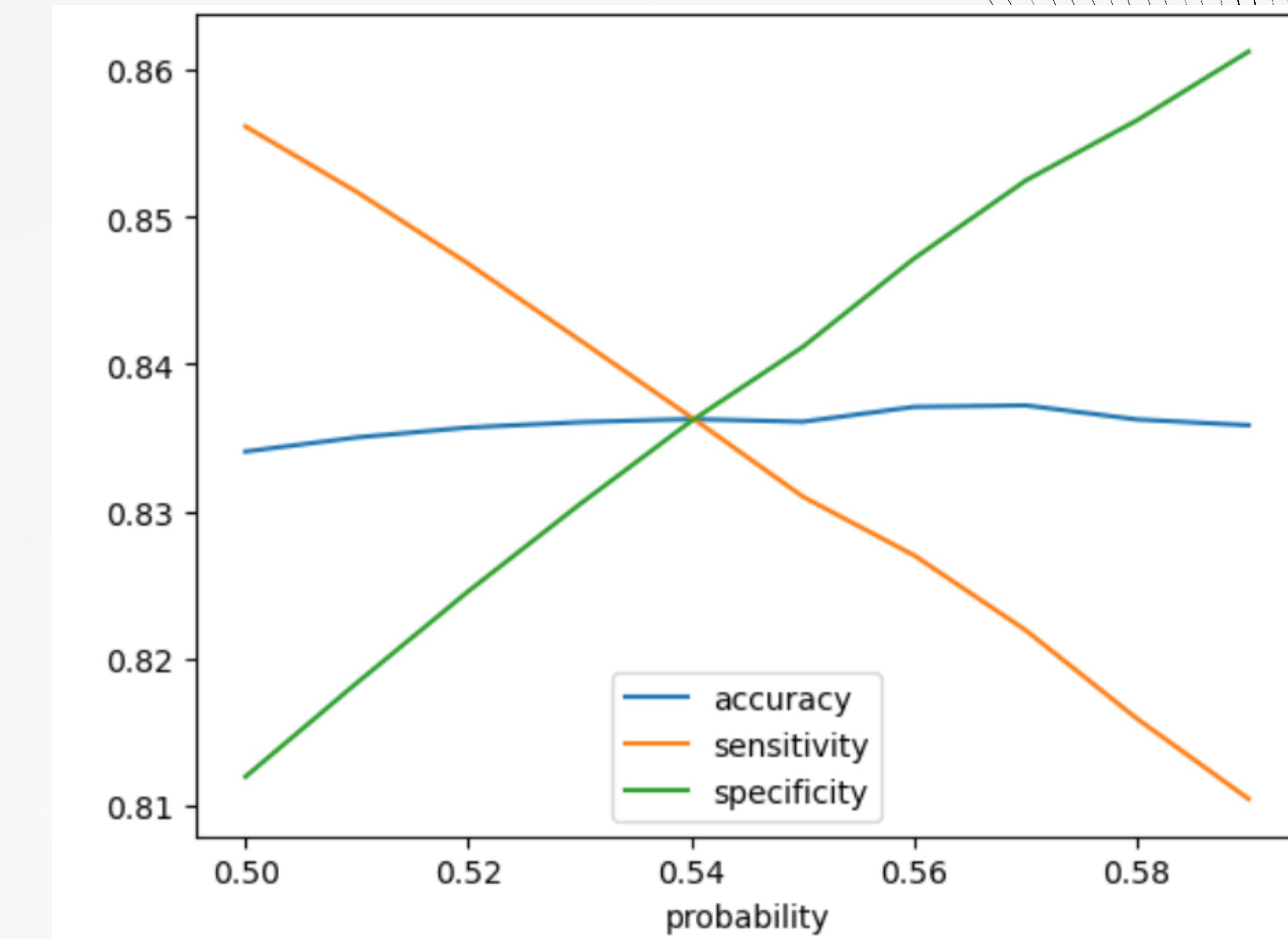
# ROC CURVE (TRAIN SET)

- The AUC Score is 90%, which is quite high and good with an accuracy of 83% (approx) on the train set
- That means the model is good at distinguishing between the positive and negative classes.



# OPTIMAL CUTOFF POINT

- Initial optimum point of classification was 0.5
- From the plot, the optimal cutoff point in the probability to define the predicted churn variable converges at 0.54



# SENSITIVITY & SPECIFICITY (TRAIN SET)

- The sensitivity and specificity is high.
- The model is correctly identifying most of the positive results and the negative results.

---

Sensitivity = 0.8363749481542928  
Specificity = 0.8361157196184156  
False Positive Rate = 0.1638842803815844  
Precision = 0.8361581920903954  
True Negative Prediction Rate = 0.8363325208733081

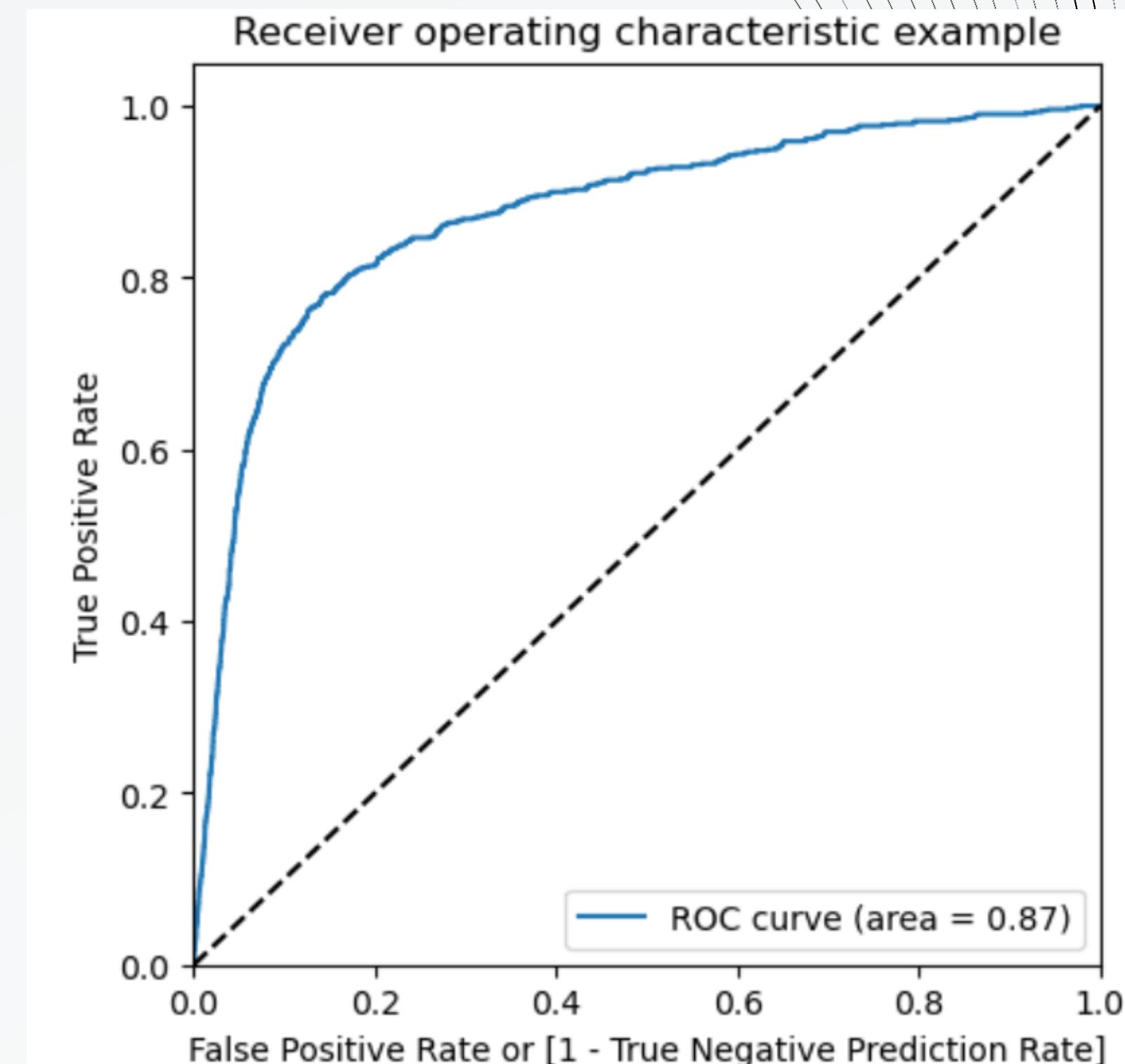
# SENSITIVITY & SPECIFICITY (TEST SET)

- The sensitivity and specificity is high as the train set.
- The model is correctly identifying most of the positive results and the negative results on the test set as well.

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

# ROC CURVE (TEST SET)

- The AUC Score is 87% on the test set.
- That means the model is good at distinguishing between the positive and negative classes similar to the train set.



# RECOMMENDATION

- Logistic Regression explained the important predictor variables as well as the significance of each variable.
- The company should focus on larger tenure range.
- Customers within 0-6 months of tenure are likely to churn.
- The recharge rates should be low as these are the customers generating revenue for the company.
- Customers with less than or equal to 10 recharges should be focussed on by the company.



THANK YOU!