

SyriaTel Customer Churn Project



DENNIS MWANIA



BUSINESS OVERVIEW

- ▶ In this project, we focus on the analysis of customer churn data from SyriaTel, a leading telecommunications provider. Customer churn refers to the phenomenon of customers discontinuing their services, and understanding the factors behind churn is critical for telecom companies like SyriaTel. By predicting churn, companies can take proactive measures to retain customers, improve customer satisfaction, and optimize resources.

BUSINESS UNDERSTANDING

- ▶ SyriaTel, a prominent telecommunications company, seeks to minimize customer churn—the rate at which customers stop doing business with them. High churn rates can be costly, as acquiring new customers is more expensive than retaining existing ones. This project aims to develop a predictive model to identify customers who are at high risk of discontinuing their services (churning) in the near future. By predicting churn, SyriaTel can implement proactive strategies to retain these customers and reduce revenue loss.

OBJECTIVES

Main Objective:

- ▶ The primary goal of this project is to predict customer churn for SyriaTel, a telecommunications company. Customer churn refers to the phenomenon where customers stop using a company's services.

Specific objectives:

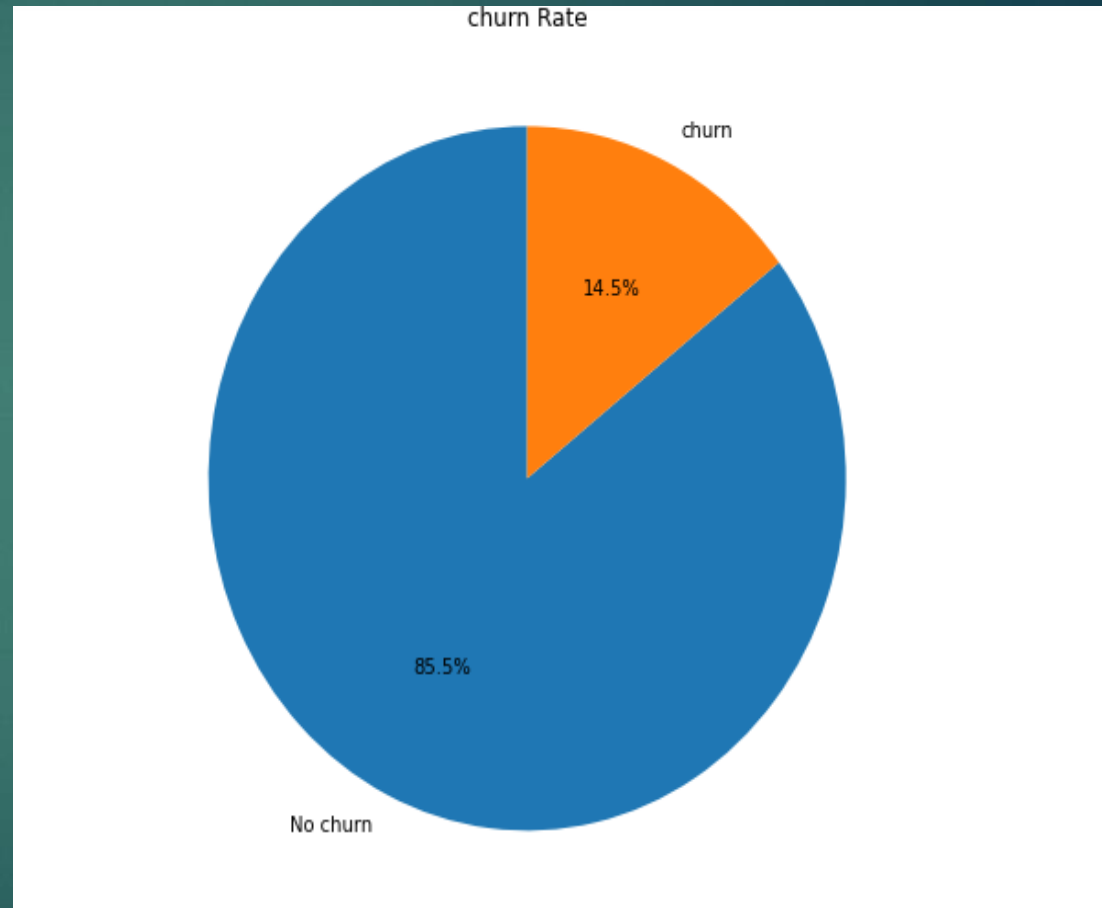
- ▶ To do exploratory data analysis on the data
- ▶ To Identify the key factors or variables that contribute most to customer churn.
- ▶ To fit different classification algorithm models
- ▶ To determine which one works best for churn prediction
- ▶ To Evaluate the effectiveness of churn prediction models using metrics like accuracy, precision, recall, and F1 score.

Data Understanding

- ▶ In this project, we will work with a customer churn dataset from the telecom industry sourced from [Kaggle](#). The dataset contains information about customers, their usage patterns, and whether they have churned or not. The dataset contains 3,333 records and spans 21 columns.

ANALYSIS

The chart clearly shows that the majority of customers (85.5%) have not churned, while a smaller portion (14.5%) has churned. This suggests a relatively low churn rate, which is generally positive for a business.

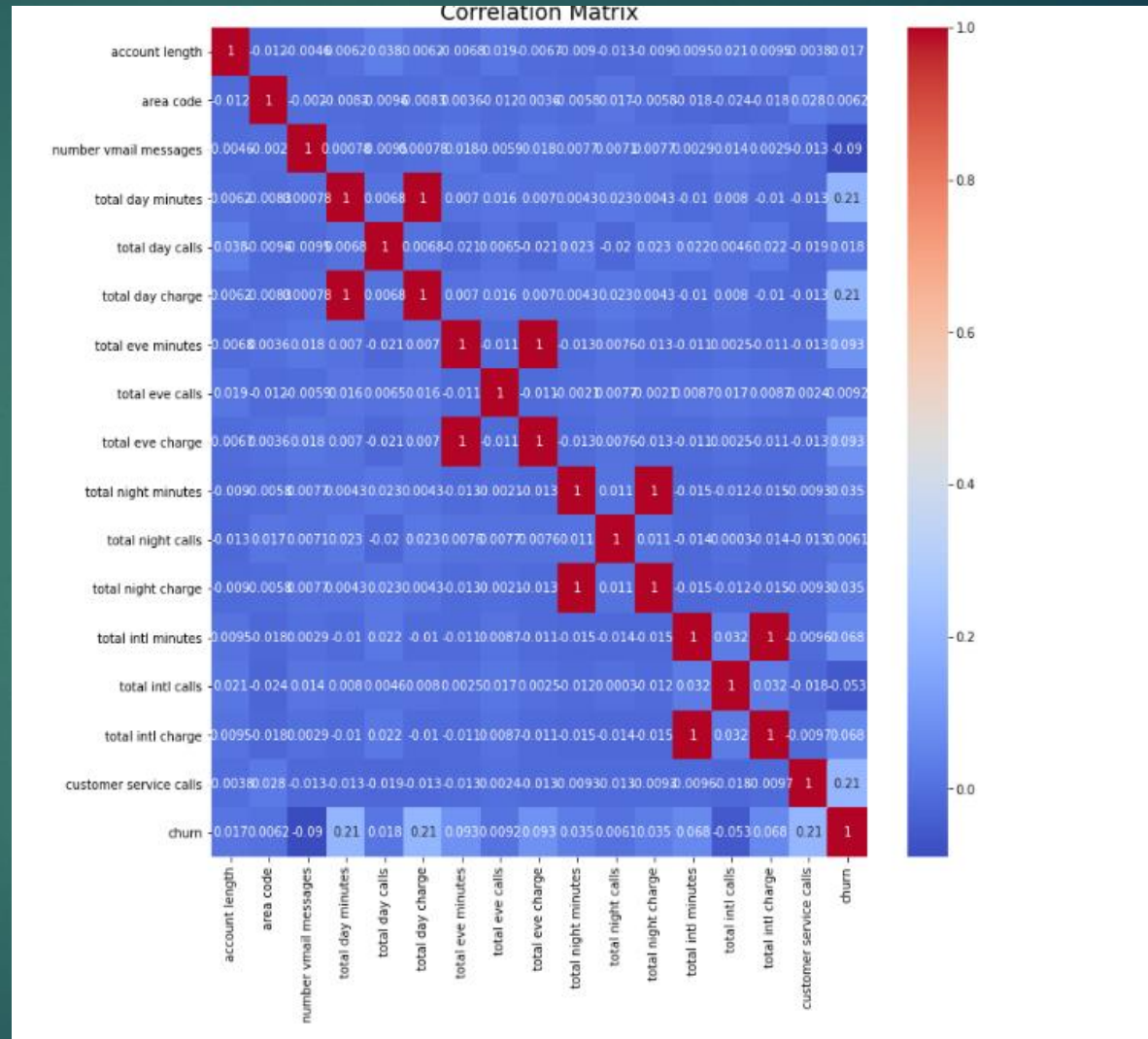


Correlation Matrix

The matrix examines correlations between variables like account length, call usage (day, evening, night, international), customer service calls, and churn (whether a customer left).

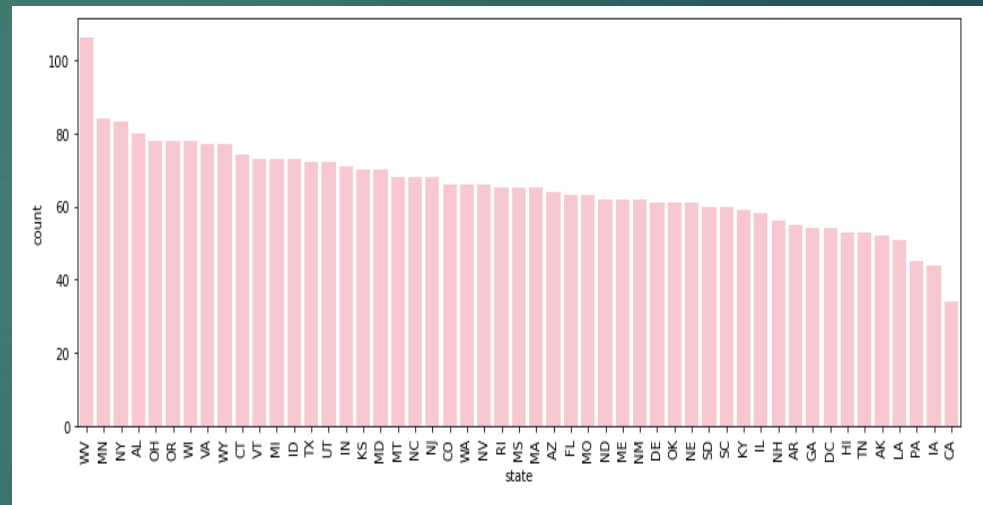
Strong Correlations: Expectedly, there are strong positive correlations between call minutes and corresponding charges (day minutes vs. day charge, etc.).

Churn Indicators: The matrix suggests a moderate positive correlation between customer service calls and churn, indicating that customers who call customer service more often are more likely to churn. There's also a weak negative correlation between account length and churn, meaning longer-term customers are slightly less likely to leave



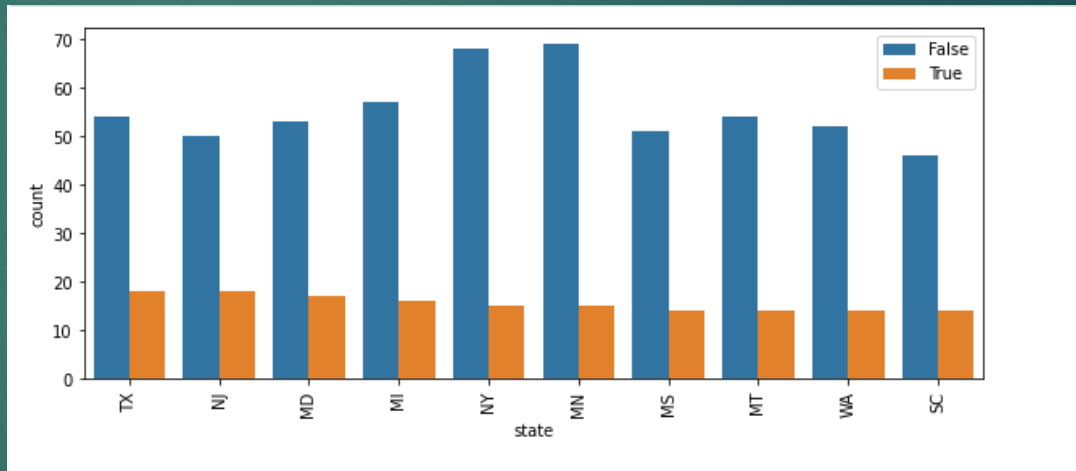
Categorical distribution 'state'

Most customers are from West
Virginia, Minnesota, New York, Alabama
and Alabama



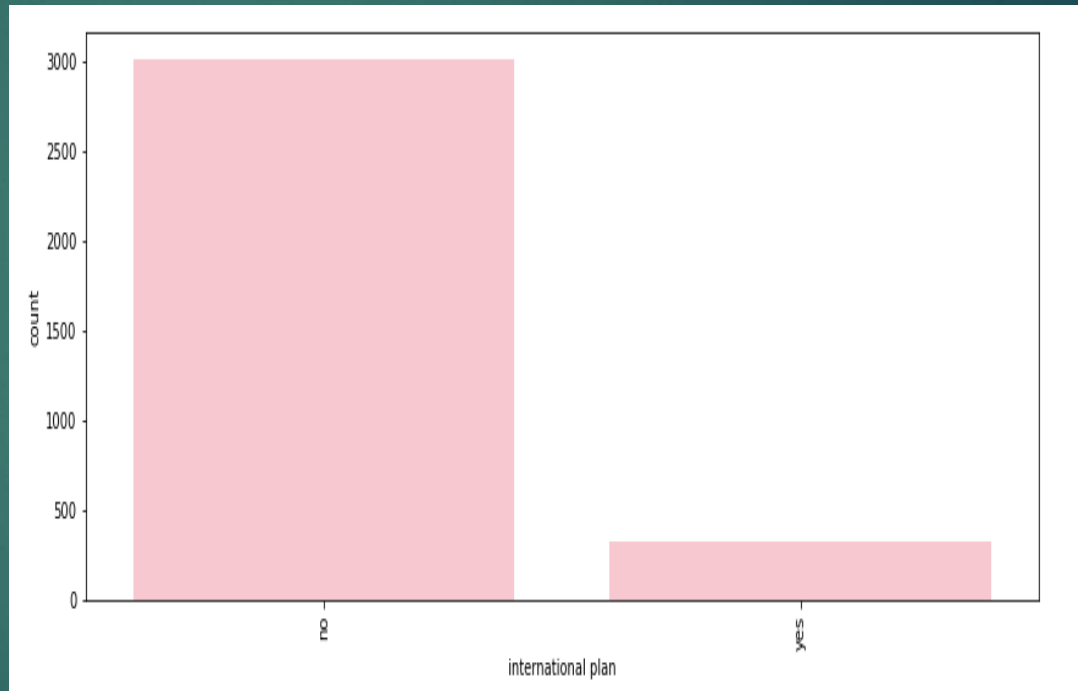
Categorical Distribution 'state'

Texas, New Jersey, Maryland, Miami and New York has the highest churn rate among the top 10 states, as indicated by the highest orange bar



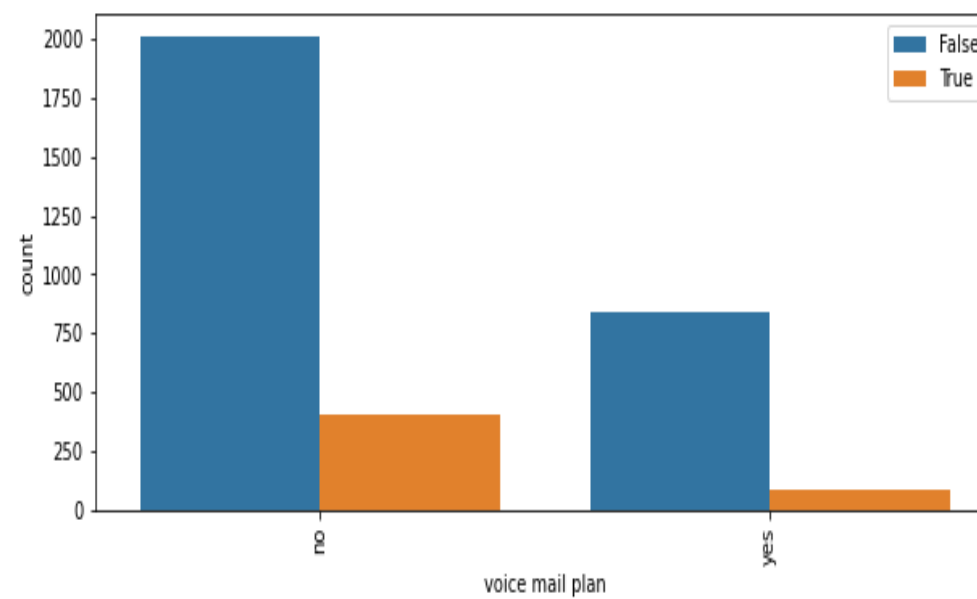
international plan

Dominance of "no": The chart clearly shows that the majority of customers in the dataset do not have an international plan ("no" bar is significantly taller than the "yes" bar).



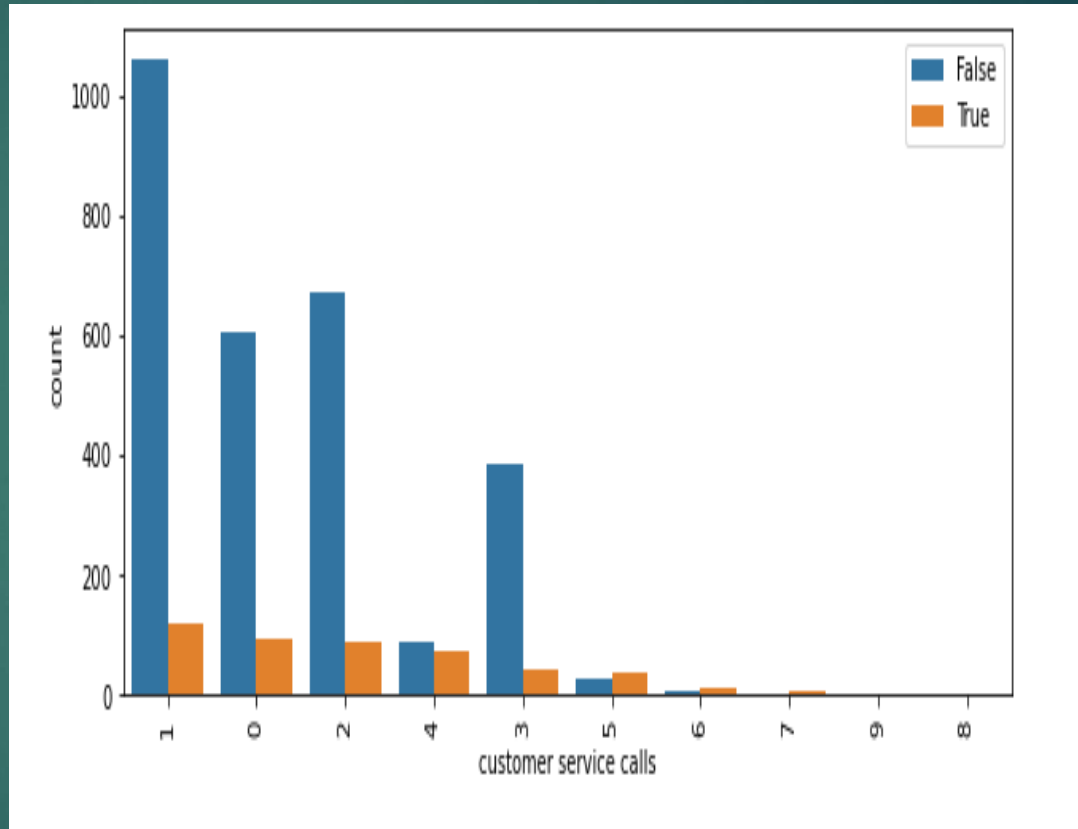
Voicemail Plan

most customers who churned did not have an voicemail plan.



Customer Service Plan

the chart suggests that the number of customer service calls is a potential predictor of customer churn. Customers who make more service calls are more likely to churn, while those who make fewer calls are more likely to remain customers. Additionally, the absence of a voicemail plan is observed to be associated with churn



MODELING

ALGORITHMS USED:

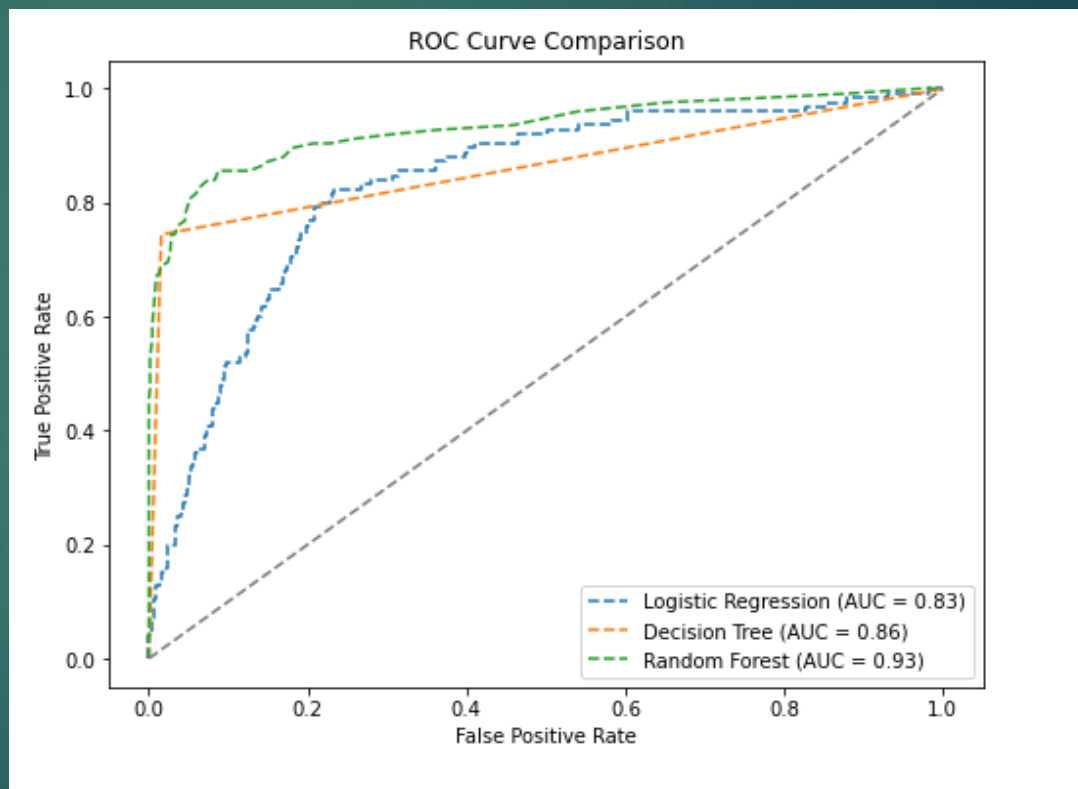
- ▶ Logistic Regression
- ▶ Decision Tress
- ▶ Random Forest

EVALUATION

AUC (Area Under the Curve) Scores:

Logistic Regression (AUC = 0.83): Performs well but is outperformed by tree-based models.
Decision Tree (AUC = 0.86): Better than logistic regression but still not the best.
Random Forest (AUC = 0.93): The best-performing model, with the highest AUC, meaning it has the strongest ability to differentiate between classes. Model Performance:

The Random Forest model (green curve) is closest to the top-left corner, meaning it has the highest True Positive Rate (TPR) while keeping a low False Positive Rate (FPR). The Decision Tree model (orange curve) performs better than Logistic Regression but is still lower than Random Forest. Logistic Regression (blue curve), while decent, has a more gradual rise, meaning it struggles more with classification. Best Model: Since higher AUC values indicate better performance, the Random Forest model (AUC = 0.93) is the best model for this dataset.



RECOMMENDATIONS

- ▶ -Most customers do not have an international plan or voicemail plan. The business should think about promoting these services to customers who might benefit from them but do not currently have them. -increase marketing in states with higher churn rates, such as Texas, New Jersey, Maryland, Miami, and New York. -Total day minutes, total eve charge, total night minutes, and total international calls and charges were identified as influential predictors the company can

Next steps

- ▶ **Monitor and Continuously Improve**
- ▶ Set up **real-time monitoring dashboards** to track model performance over time.
- ▶ Regularly **retrain models** with new data to maintain accuracy and relevance.
- ▶ Gather **customer feedback** to understand churn drivers beyond the available data.

Thank You!

