



SYRIA

SYRIA TEL CUSTOMER CHURN

Outline

- 01 Project Overview
- 02 Business Understanding
- 03 Objectives
- 04 Data Understanding
- 05 Data Visualization
- 06 Data Modeling
- 07 Conclusions
- 08 Recommendations



Project Overview

SyriaTel faces customer churn, impacting revenue and market standing.

To adapt, the company must understand attrition reasons, differentiate by offering superior services, and use predictive machine learning models.

Build a classifier to predict whether a customer will ("soon") stop doing business with SyriaTel, a telecommunications company.



Business Understanding

SyriaTel is developing a machine learning classifier to predict customer churn, aiming to optimize revenue and customer satisfaction.

The classifier will analyze historical data, including call patterns and interactions, to provide insights into factors influencing churn.

This predictive model will enable SyriaTel to address customer needs, reduce attrition, and foster long-term loyalty.



Objectives

1. Identifying common characteristics and behaviors associated with customers who have churned
2. Evaluate the importance of different features
3. Develop models and validate their predictive capabilities



Data Understanding

Data science involves understanding a dataset to gain insights into its structure, contents, and characteristics.

This dataset has 3333 rows and 21 columns, describing each customer with 21 attributes. The 'churn' column is the target variable for a predictive model.



Data Understanding



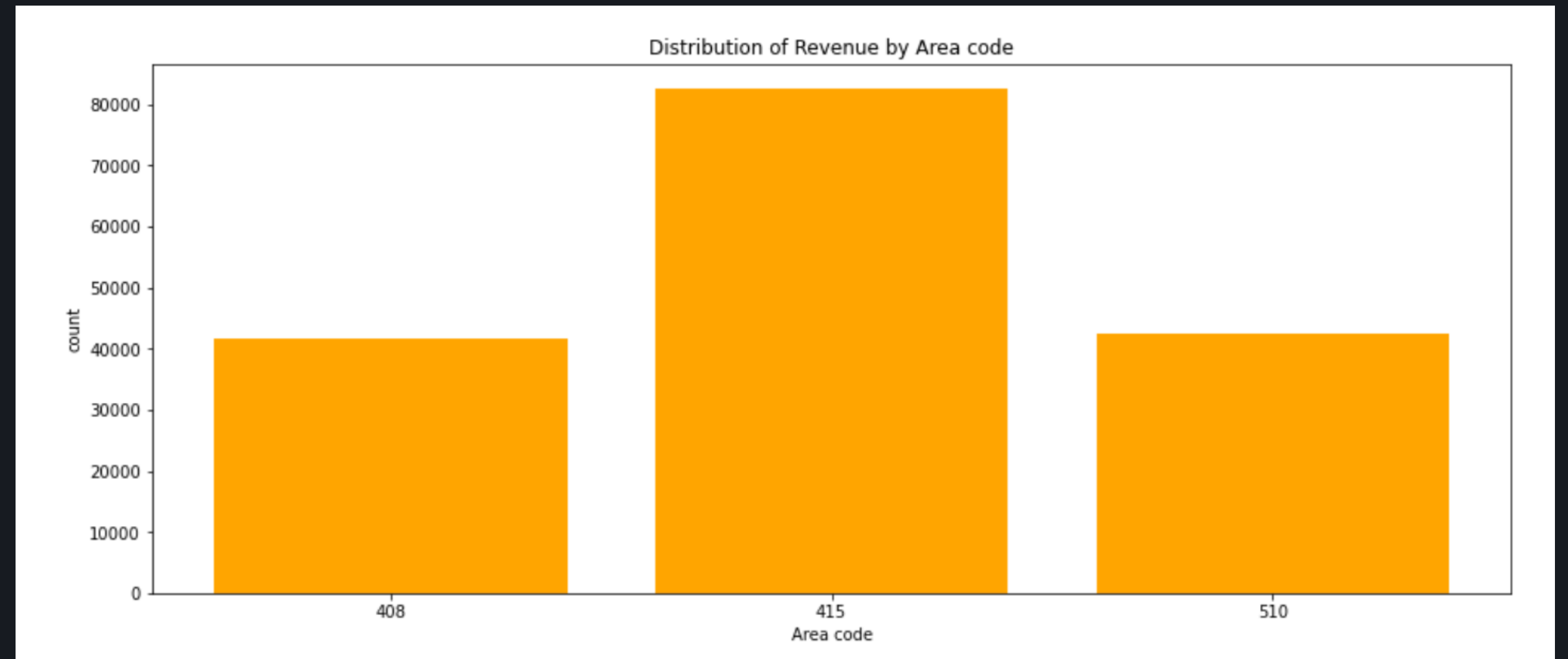
The columns include;

- state
- account length
- area code
- phone number
- international plan voice mail plan
- number vmail messages
- total day minutes, calls, charge
- total eve minutes, calls, charge
- total night minutes, calls, charge
- total intl minutes, calls, charge
- customer service calls
- churn - our target

Exploratory Data Analysis:

The distribution of expenditure in area codes

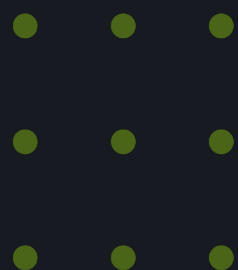
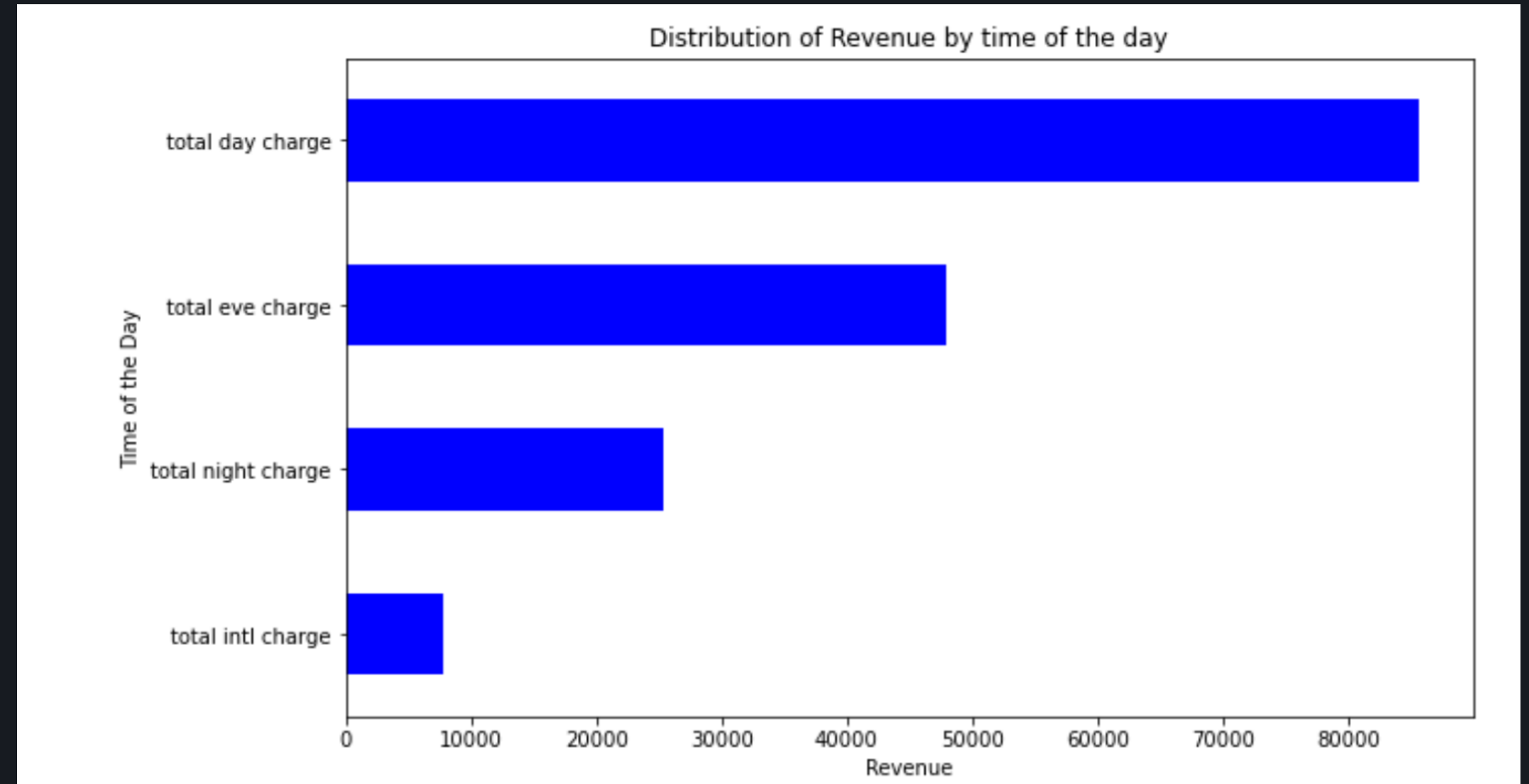
- The area code 415 has the highest expenditure



Exploratory Data Analysis:

How the Expenses are allocated according to the time of the day

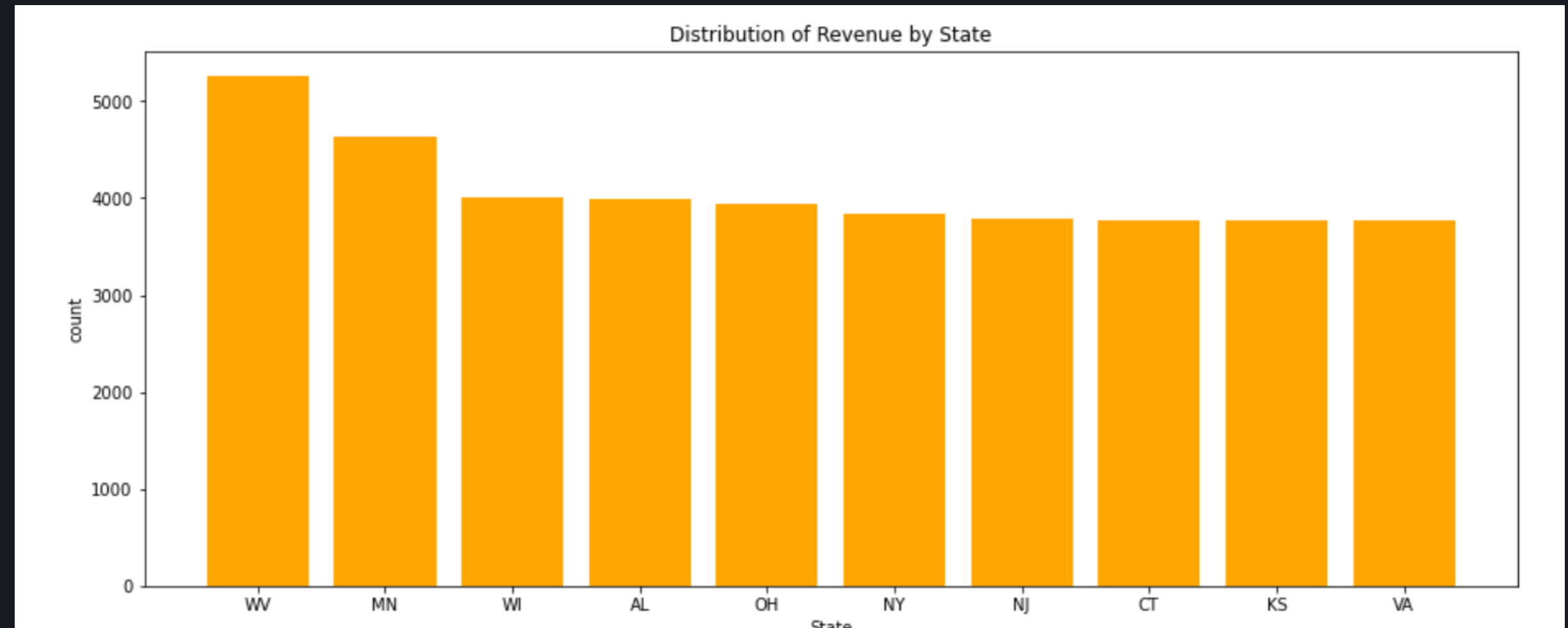
- According to our distribution , total day charge brings in the most revenue more than 80,000 while the rest of the day is less than 80,000



Exploratory Data Analysis:

Distribution of Revenue by State

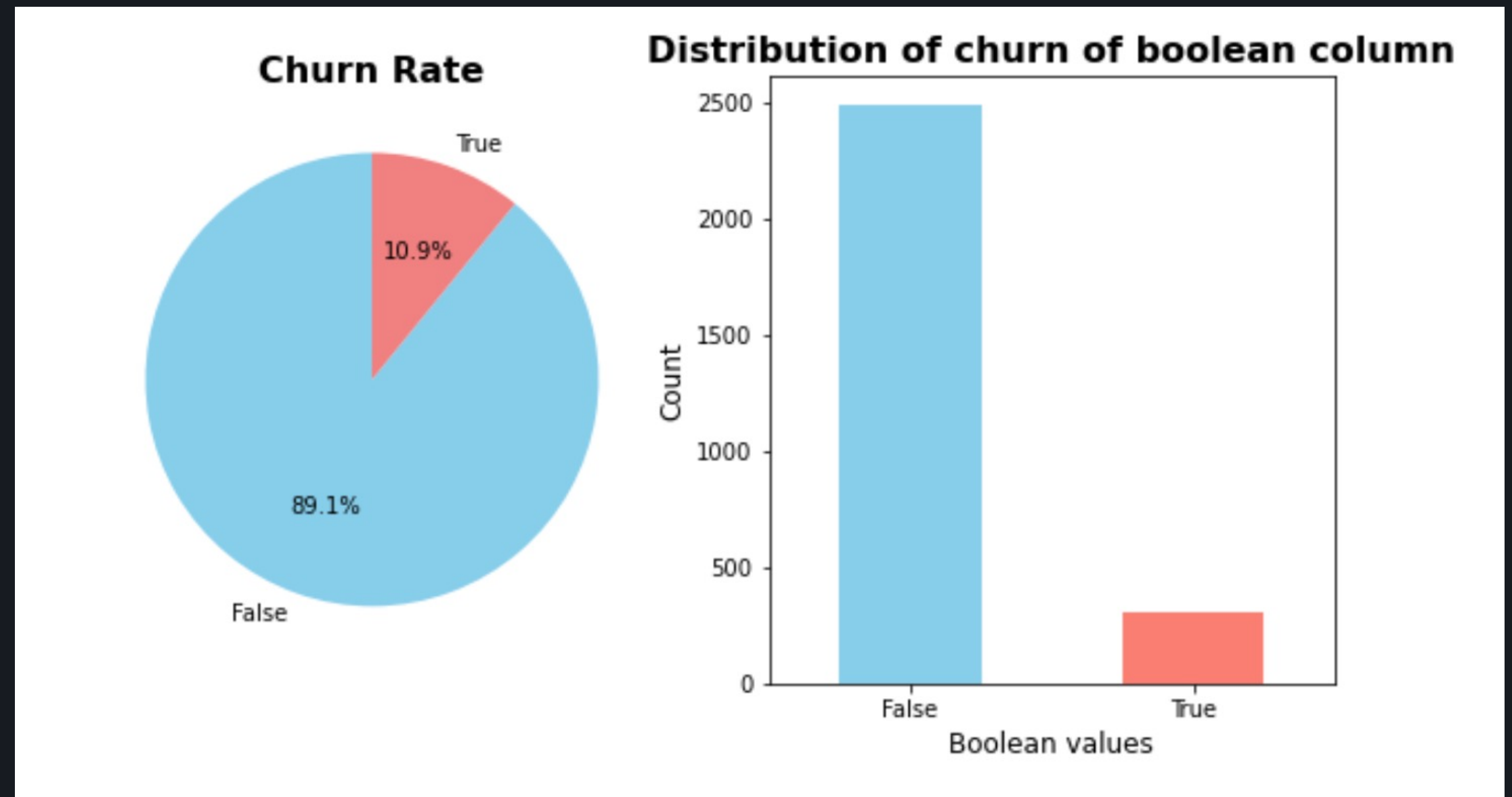
- West Virginia is the state with the highest revenue with over 6000



Exploratory Data Analysis:

The rate at which the customers stopped doing business with the company

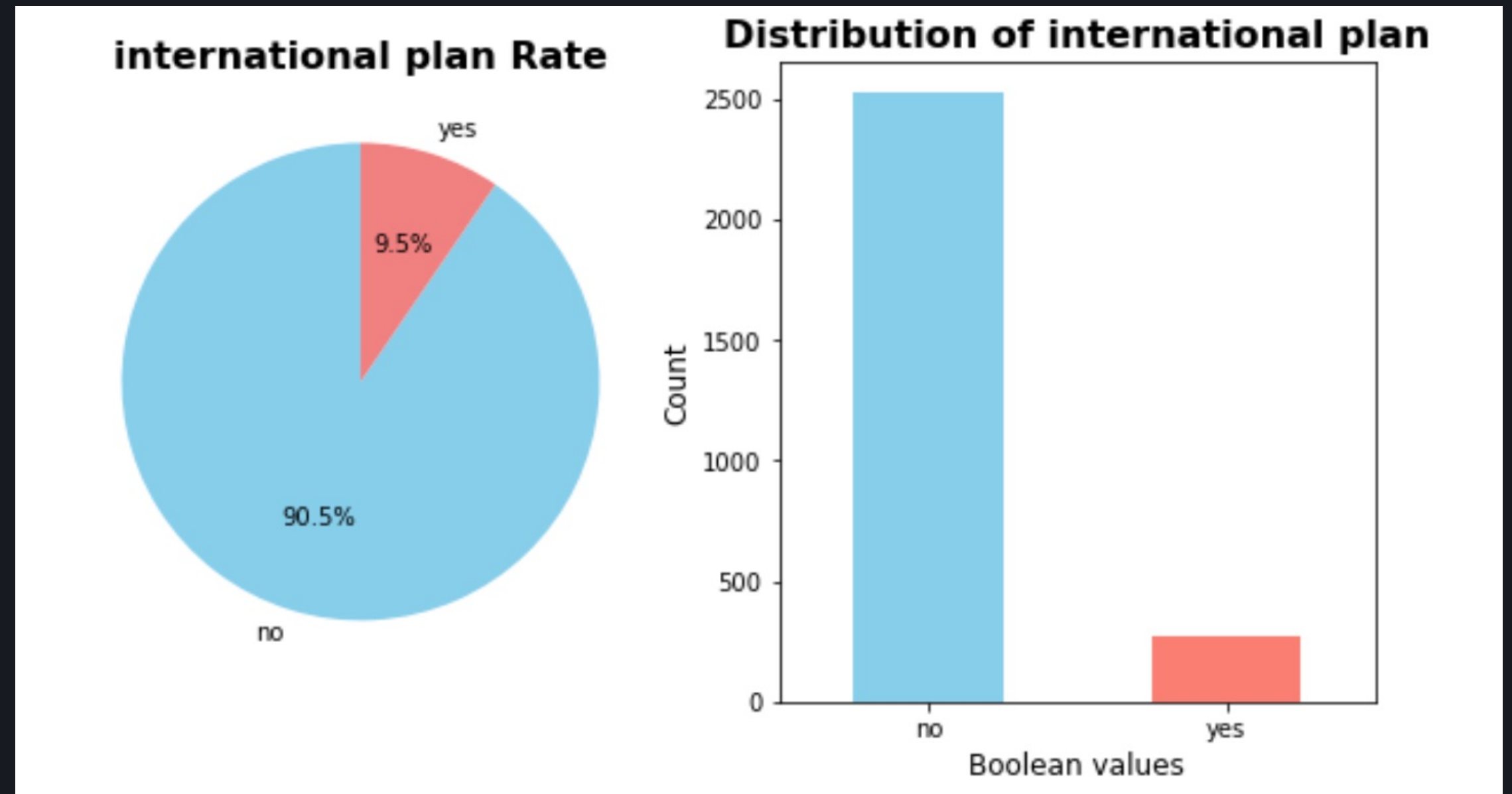
- The churn rate is 10.9%



Exploratory Data Analysis:

The distribution of the 'international plan' feature with counts for each Boolean value.

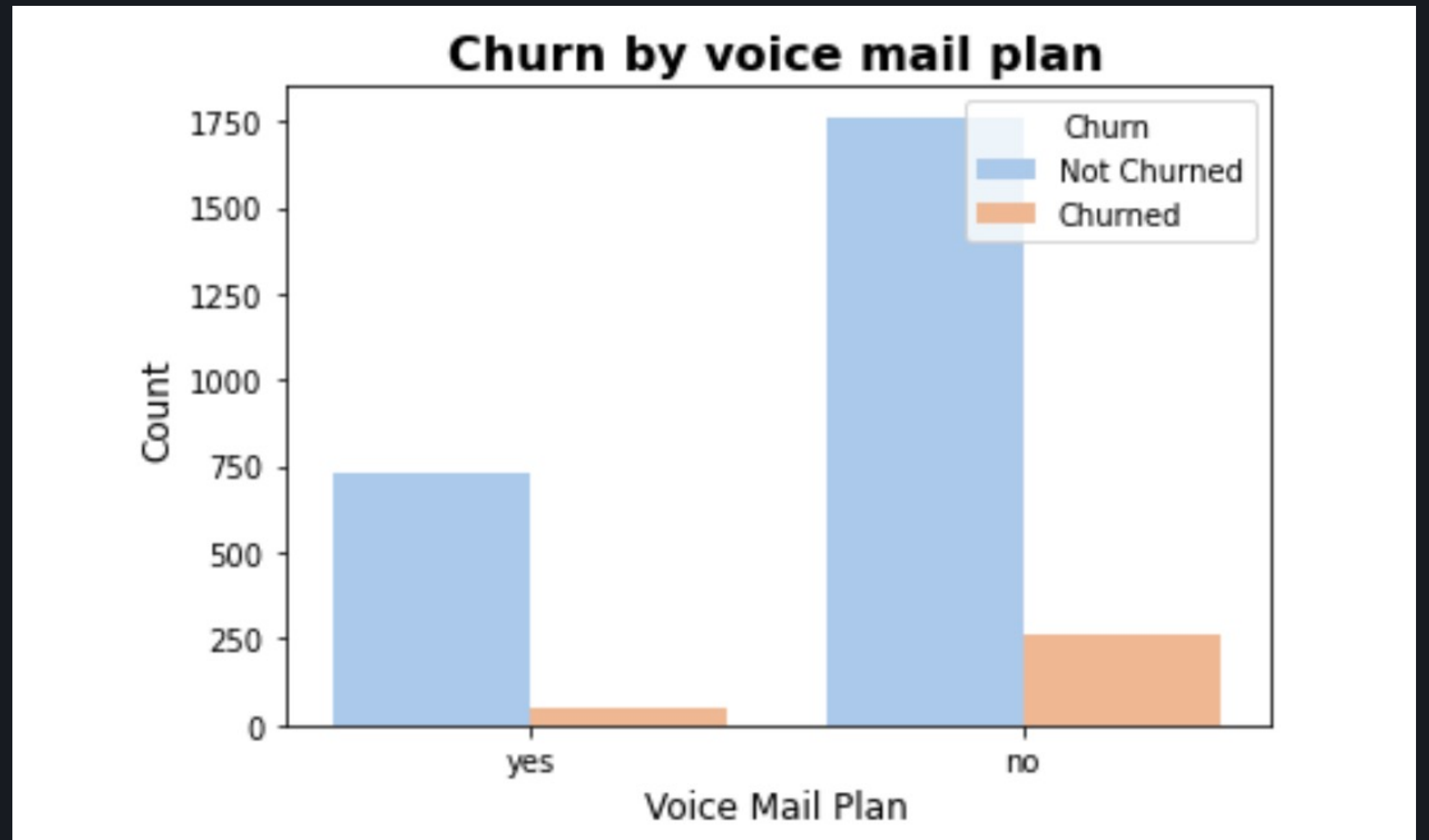
- We can see that most of the customers do not have international plan



Exploratory Data Analysis:

Visualize the distribution of churn based on the 'voice mail plan' feature

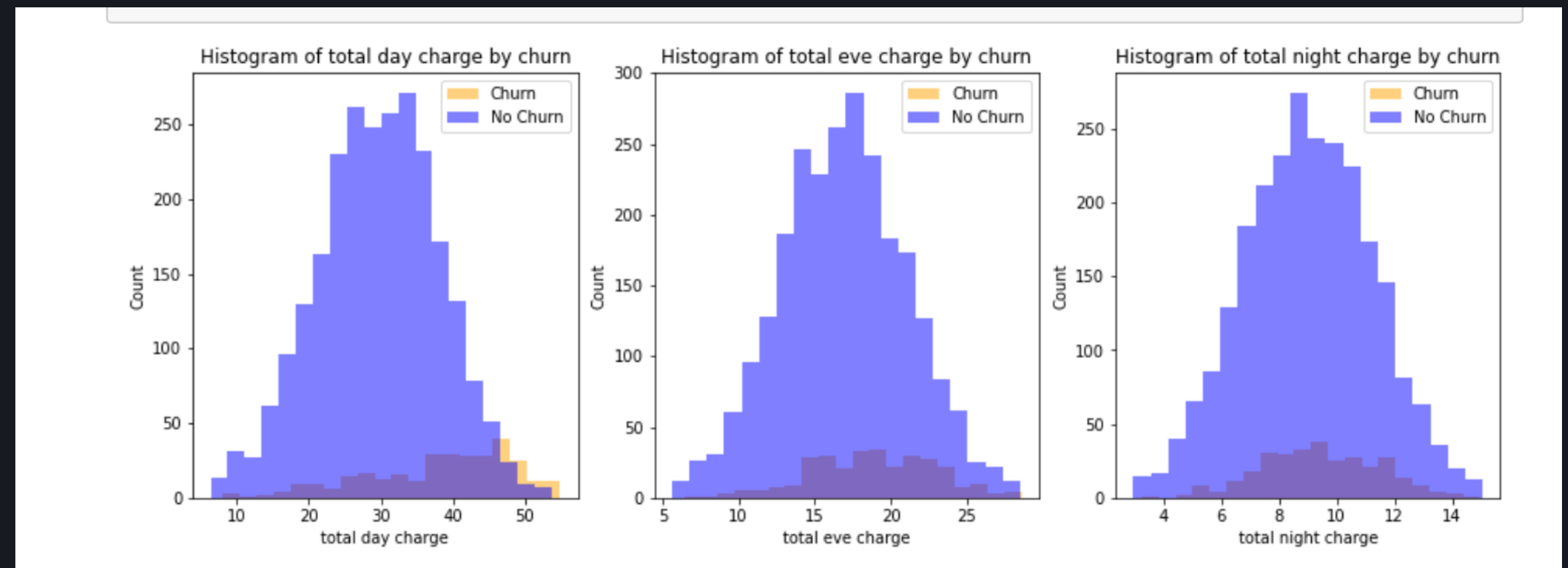
The customers who do not have voice mail plan have a higher churn rate compared to the ones who have the voice mail plan



Exploratory Data Analysis:

Visualize the distribution of specific numeric features

- The trend of lower churn rates suggests that customers with higher charges are less likely to leave .



Data Modeling

1. Logistic Regression

Model results

Precision score :0.6857142857142857

Recall score :0.22641509433962265

Accuracy score :0.8892857142857142

F1 score :0.3404255319148936

Model evaluation

The model has high precision, predicting positive instances 69% of the time, but only captures 23% of actual positive instances.

Its accuracy is 88.9%, but may not be sufficient in imbalanced datasets. The F1 score balances precision and recall, with a score around 0.34.

Data Modeling

2. Decision Tree Model

Model results

Precision score
:0.9107142857142857

Recall score
:0.4811320754716981

Accuracy score
:0.9285714285714286

F1 score :0.6296296296296295

Model evaluation

- The new model shows improved precision and recall, with a 92.9% accuracy rate and a better balance between precision and recall.

Data Modeling

3. Random Forest Regression Model

Model results

Precision score: 0.9792

Recall score: 0.4434

Accuracy score: 0.9286

F1 score: 0.6104

Model evaluation

The Random Forest model has a high precision score of 0.9792, indicating a low false positive rate. It correctly identified 44.34% of all positive instances, with an overall accuracy of 92.86%.

The model's F1 score, which measures the harmonic mean of precision and recall, is approximately 61.04%, indicating a good balance between precision and recall.

Data Modeling

4. Gradient Boost Model

Model results

Precision score: 0.9828

Recall score: 0.5377

Accuracy score: 0.9405

F1 score: 0.6951

Model evaluation

The Gradient Boosting model's high precision minimizes false positives, making it effective in high-cost scenarios. Moderate recall suggests room for improvement in capturing actual positives, balancing precision and recall..

Conclusion

Highest contributors to Customer Churn include;
Having International Plan ,customers with international plans are more likely to churn.

Total number of day minutes , higher usage of daytime minutes is associated with higher churn

High number of customer service calls, customers who make frequent customer service calls are more likely to churn

High number of evening minutes , is correlated to higher churn

Lower contributors include:

Number of voice mail messages, total international calls,state,area code



Recommendations

- Investigate the reasons behind the correlation. It could be related to pricing, service satisfaction, or specific features of international plans. Consider targeted retention strategies for customers with international plans.
- Assess the pricing or service satisfaction related to daytime usage. Consider offering plans that cater to high daytime users or communicating the value proposition of existing plans.
- Identify the root causes of frequent service calls—these could indicate dissatisfaction or unresolved issues. Focus on improving customer service quality and addressing common pain points.



Next steps

- To develop targeted retention strategies based on the insights gained from high contributors to churn





THANK YOU

