SYRIATEL TELECOMMUNICATION FIRM

PROJECT OVERVIEW

- This project aim to build a model that will predict whether a customer is likely to leave SyriaTel soon.

BUSINESS UNDERSTANDING.

▶ SyriaTel, a telecommunication firm, is facing a significant issue of customer churn that may adversely affect its revenue and future. The firm should identify the customers who are likely to discontinue using their services and the reasons why they decide to do so. Through this learning, SyriaTel can implement measures to resolve customer problems and prevent them from leaving.

DATA UNDERSTANDING.

I engaged in the following activities
- I started with importing libraries
- Loaded the dataset
- Checking rows and columns
- Checking the data type
- Checking for the statistics information
- Checking the statistics information with transpose

DATA PREPARATION

▶ **Data cleaning**

• I checked for the missing values

• Duplicates

• Checked for unique values in the categorical values

• Drop unnecessary column

• confirm if the phone number has been removed

• checking the outlier

• identify numerical columns except area code

• Calculating upper Q3 (75th percentile) and lower Q1 (25th percentile)
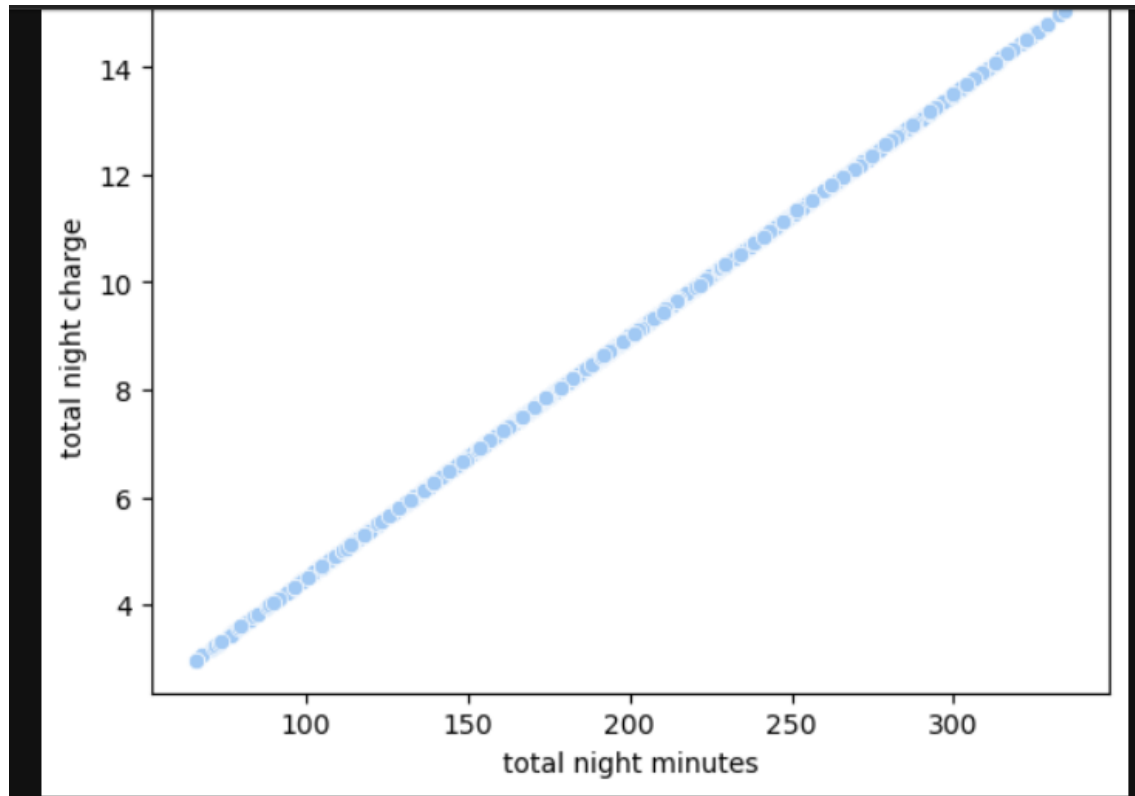
• Calculating interquartile range

EDA

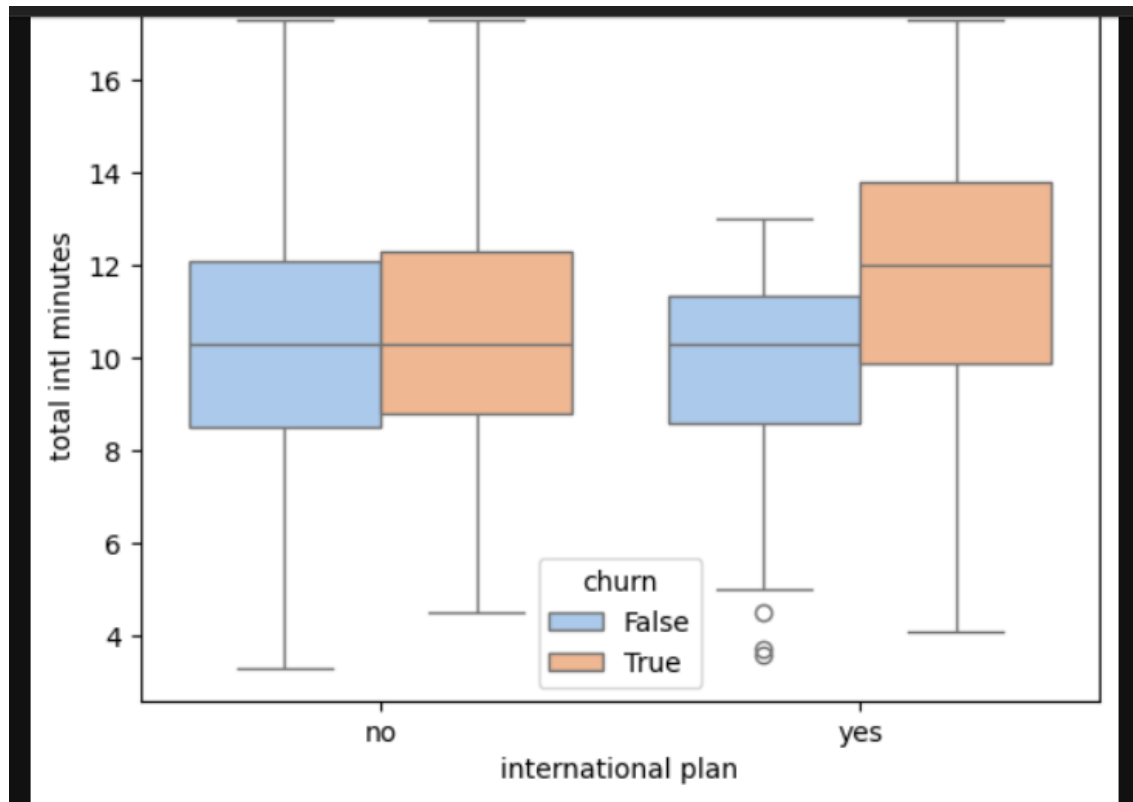▶ **Univariate analysis**

It is the analysis of one variable
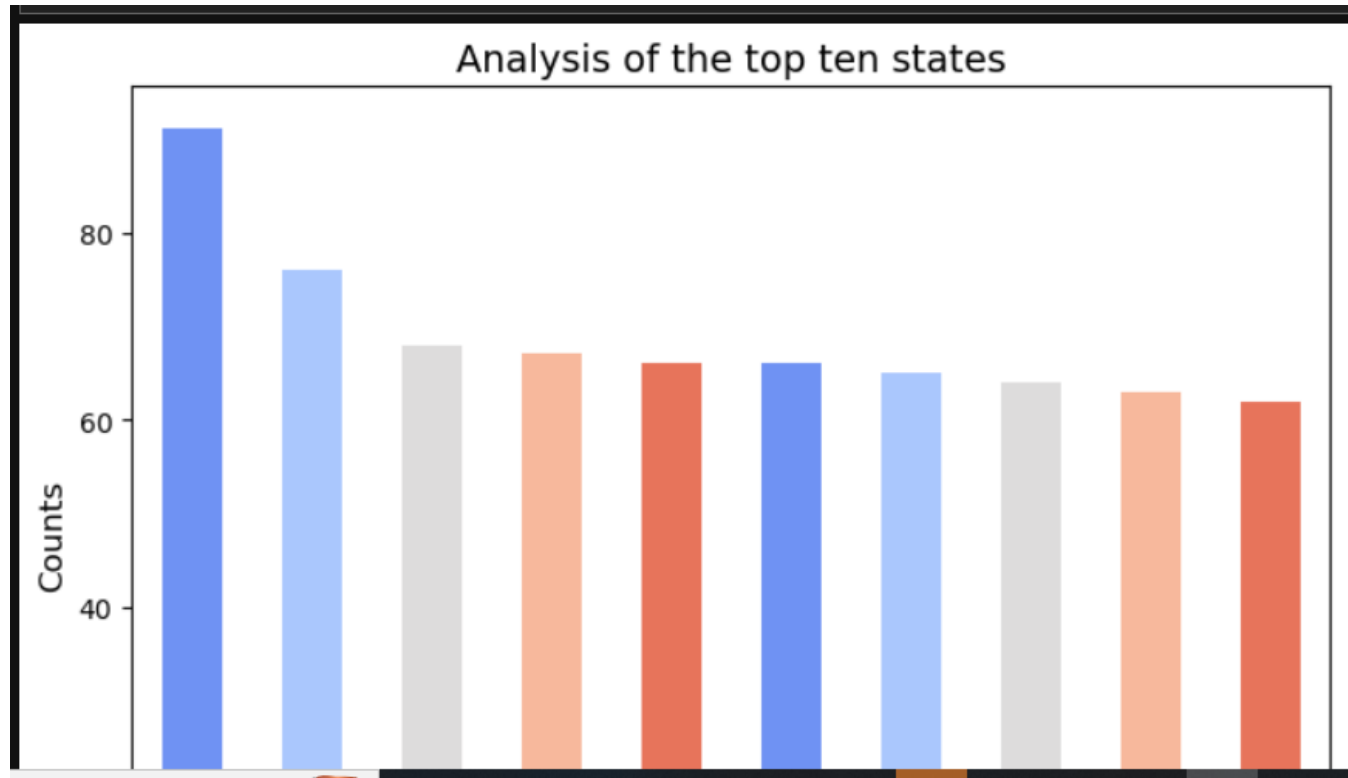
Bivariate analysis

▶ Its the analysis of two variables

Multivariate analysis
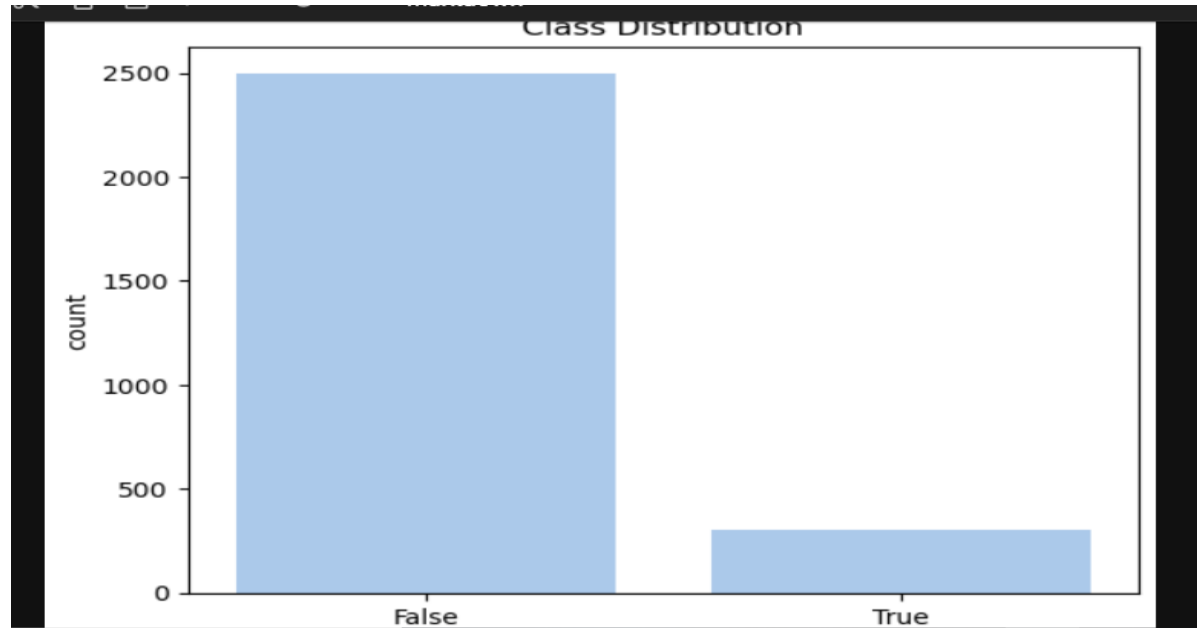
▶ It is the analysis of more than two variables

MODELING

- I used the following models
1. Logistic regression
2. Decision tree

LOGISTIC REGRESSION WHEN THE CLASS IS IMBALANCED

LOGISTIC REGRESSION WHEN THE CLASS IS IMBALANCED

▶ **Preprocessing**

It is the changing of categorical variable to numerical variable

▶ **Splitting data into target and predictors**

▶ **split the data into train, test and split¶**

▶ **Scaling¶**

▶ **Build the model**

▶ **Predict Y**

▶ **The accuracy¶**

The accuracy was 88.59%

LOGISTIC REGRESSION(WITH BALANCED CLASS)/SMOTE TECHNIQUE

▶ **Preprocessing**

It is the changing of categorical variable to numerical variable

▶ **Splitting data into target and predictors**

▶ **split the data into train, test and split**¶

▶ **Scaling**¶

▶ **Build the model**

▶ **Predict Y**

▶ **The accuracy**¶

The accuracy was 68%

DECISION TREE(using hyperparameter)

- **Preprocessing**

It is the changing of categorical variable to numerical variable

- **Splitting data into target and predictors**

- **split the data into train, test and split¶**

- **Scaling¶**

- **Build the model**

- **Predict Y**

- **The accuracy¶**

The accuracy was 94%

EVALUATION

- Decision tree is the best model to use because it has an accuracy of 94% compared to logistic with an accuracy of 68%

RECOMMENDATION

- Use another method such as random forest with a balanced class to avoid bias.