CapstoneProject-1 Telecom Churn Analysis

Submitted To

Alma Better

Submitted By

Pooja Rathore

Prashant Dube

Cohort - Azaadi

Table of Content

- Data Summary
- Problem Statement
- Objective
- Data Inspection
- Data Cleaning(null and missing values ,outliers etc.)
- Data Analysis and Visualization :
  - a. Top States for Churned Users
  - b. Relation Between factors and Churning
  - c. International plan and Churning
  - d. Customer Service and Churning
  - e. Correlation
- Conclusion
- Suggestion

Summary

▶ State- Having abbreviation name of USA states only .

▶ Area code- Having area code of customers.

▶ International plan - A check for international plan.

▶ Voice mail plan - A check for voicemail plan.

▶ Number vmail message-Number of voicemail messages sent by customers who opted for voice mail plan.

▶ Total day minutes - Having total of minutes which a customer spent in a day-time.

▶ Total day calls - Having total number of calls of a customer in a day-time.

▶ Total day charge - Having total of charges of a customer's spending in a day-time.

▶ Total eve minutes - Having total of minutes which a customer spent in a evening-time.

▶ Total eve calls - Having total number of calls of a customer in a evening-time.

▶ Total eve charge - Having total of charges of a customer's spending in a evening-time.

- Total night minutes - Having total of minutes which a customer spent in a night-time.

- Total night calls - Having total number of calls of a customer in a night-time.

- Total night charge - Having total of charges of a customer's spending in a night-time.

- Total intl minutes - Having total of minutes which a customer spent on international calls.

- Total intl calls - Having total number of international calls of a customer .

- Total intl charge - Having total of charges of a customer's spending on international calls.

- Customer service calls - Having number of calls made by a particular customer to customer service care.

- Churn - Having churned and non-churned status of customers.

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 French multinational telecommunication company Orange S.A. telecommunications company experiences churn. they have noticed that some of their users left their company because of some reason. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. Our analysis can help in knowing the reason why users will leave that telecom service and what should be the perfect strategy for customer retention.

Objective

► The main objective is to do some analysis, which could help them in findings the key factors responsible for customer churn, identifying churn behaviour and validate the reasons for customer churn with the help of EDA .

► Based on that we can recommend some suggestions for customer retention to business team based on analysis of telecom churn data set for reducing customer churn and increasing profit of the company

Data Inspection

▶ Data Inspection is **the act of viewing data for verification and debugging purposes, before, during, or after a translation**.

▶ It is an open source library in python in which we can make sure that we are dealing with the right dataset and for a clear view of data set at every stage of the transformation process

▶ During this step we checked shape of data, data types, unique value in column and statistics information i.e. the summary of given data set.

▶ In order to understand our data, we can look at each variable and try to understand their meaning and relevance to this problem.

▶ Telecom churn dataset have 3333 rows and 20 columns having all the columns with data type of object, integer, float, and Boolean.

▶ In the given data set our target column is Churn and telecom dataset consider only 3 categorical data columns those are state, international plan, voice mail plan and remaining columns are numerical dataset

▶ the output column churn has 2 variables false or true based on dataset 14.49% customers are churned and remaining are the loyal towards the company.

Data Cleaning

- **Data cleansing** or **data cleaning** is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.[

Data Cleaning – null and missing values

- Dealing with null values is very important because Missing data in the training data set can reduce usefulness of a result or can lead to a biased results. It can lead to wrong prediction also.

- In the given data set of Orange S.A. telecommunication dataset, there are no null values, and missing values is present

Data Cleaning - Outliers

▶ Checking outlier in the dataset is necessary because Outliers is also something that we should be aware of. Why? Because outliers can markedly affect our results and can be a valuable source of information. It is also providing us the insights of specific behaviors.

▶ Outliers is a complex subject and it deserves more attention. Here, we'll just do a quick analysis on data set to find outliers.

▶ In the given data set of telecom churn all the columns having outliers.

▶ With the help of Box plot we can determine the outliers of every columns

Outliers
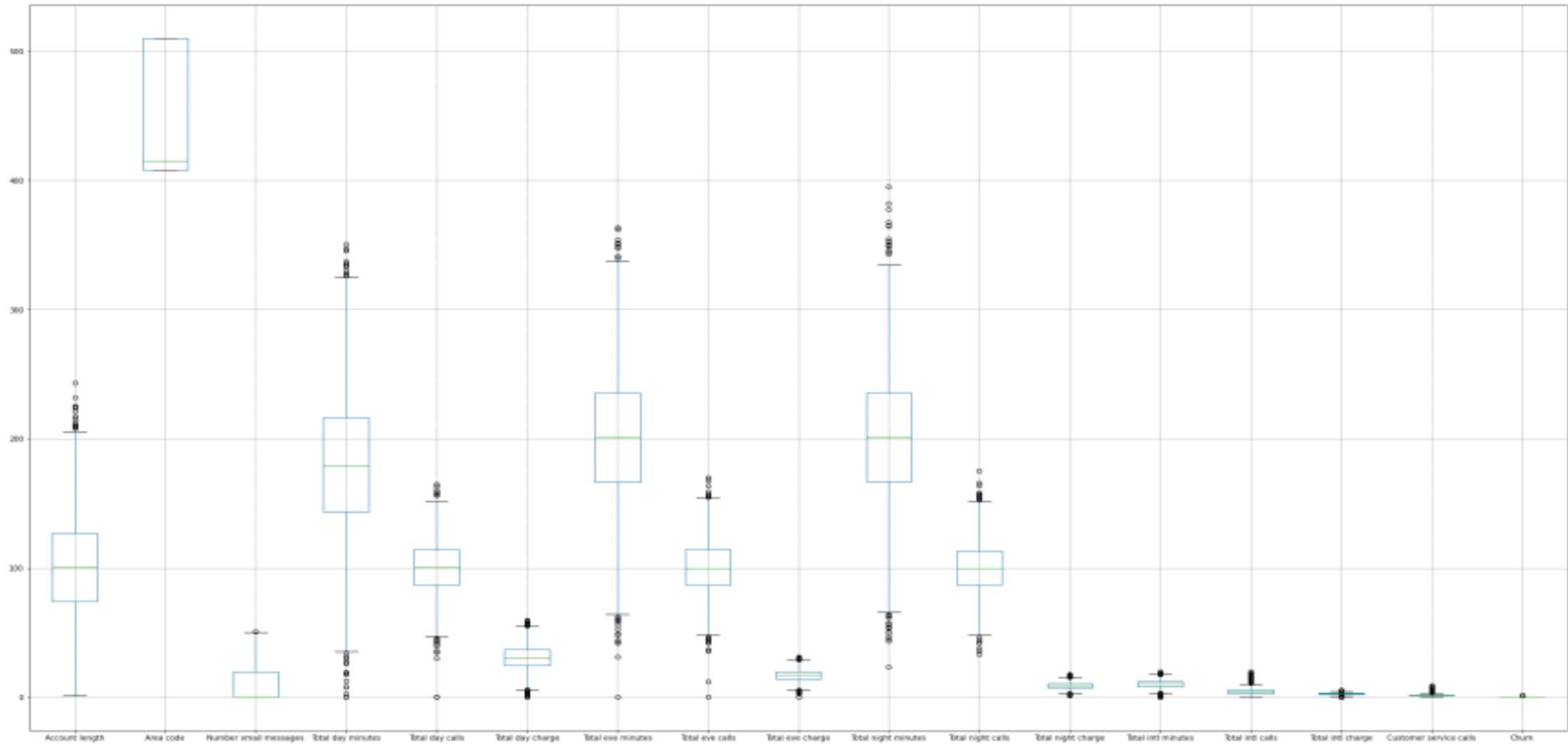
Exploratory Data Analysis

- Data Visualization represents the text or numerical data in a visual format, which makes it easy to grasp the information the data express. We, humans, remember the pictures more easily than readable text, so Python provides us various libraries for data visualization like matplotlib, seaborn etc.

- In this telecom churn EDA we will use Matplotlib and seaborn for performing various techniques to explore data using various plots

- Target variable in telecom churn dataset has 2 categories False or True

- From pie chart we can see that our almost 14% Customer has been churned. We can definitely do something about that after analyzing the data

- In the given pie chart False is replaced with 0 and True is replaced with 1

Count of TARGET Variable per category
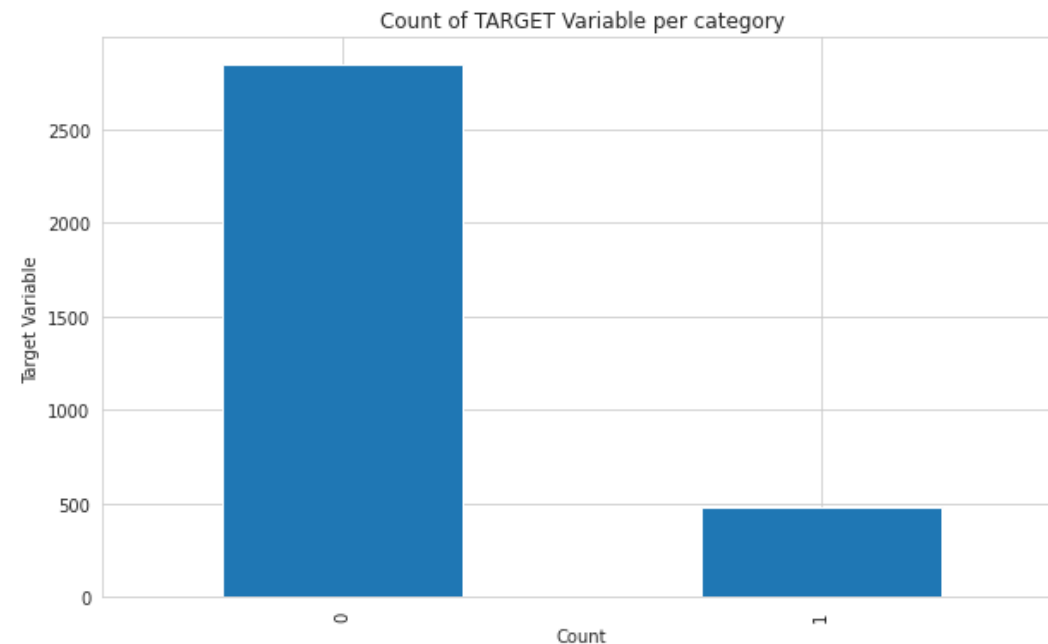
Percentage of churned and non-churned users
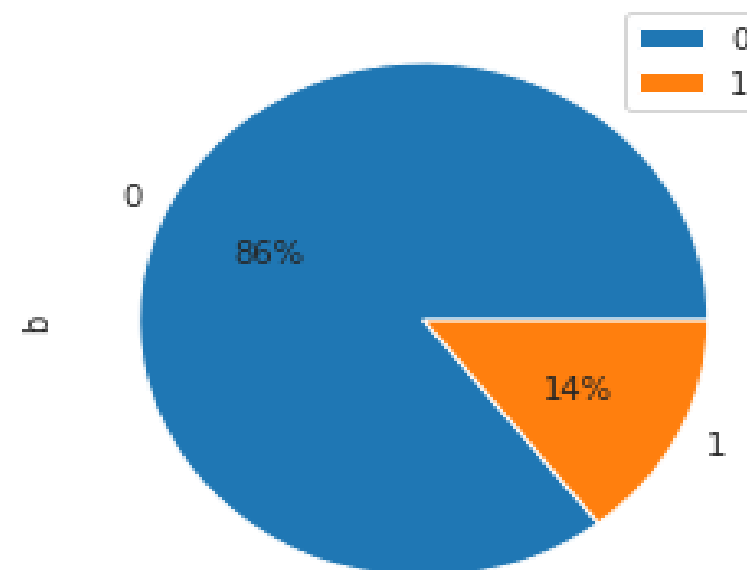
Comparing Churn variable across the states

Churn Distribution Across the States
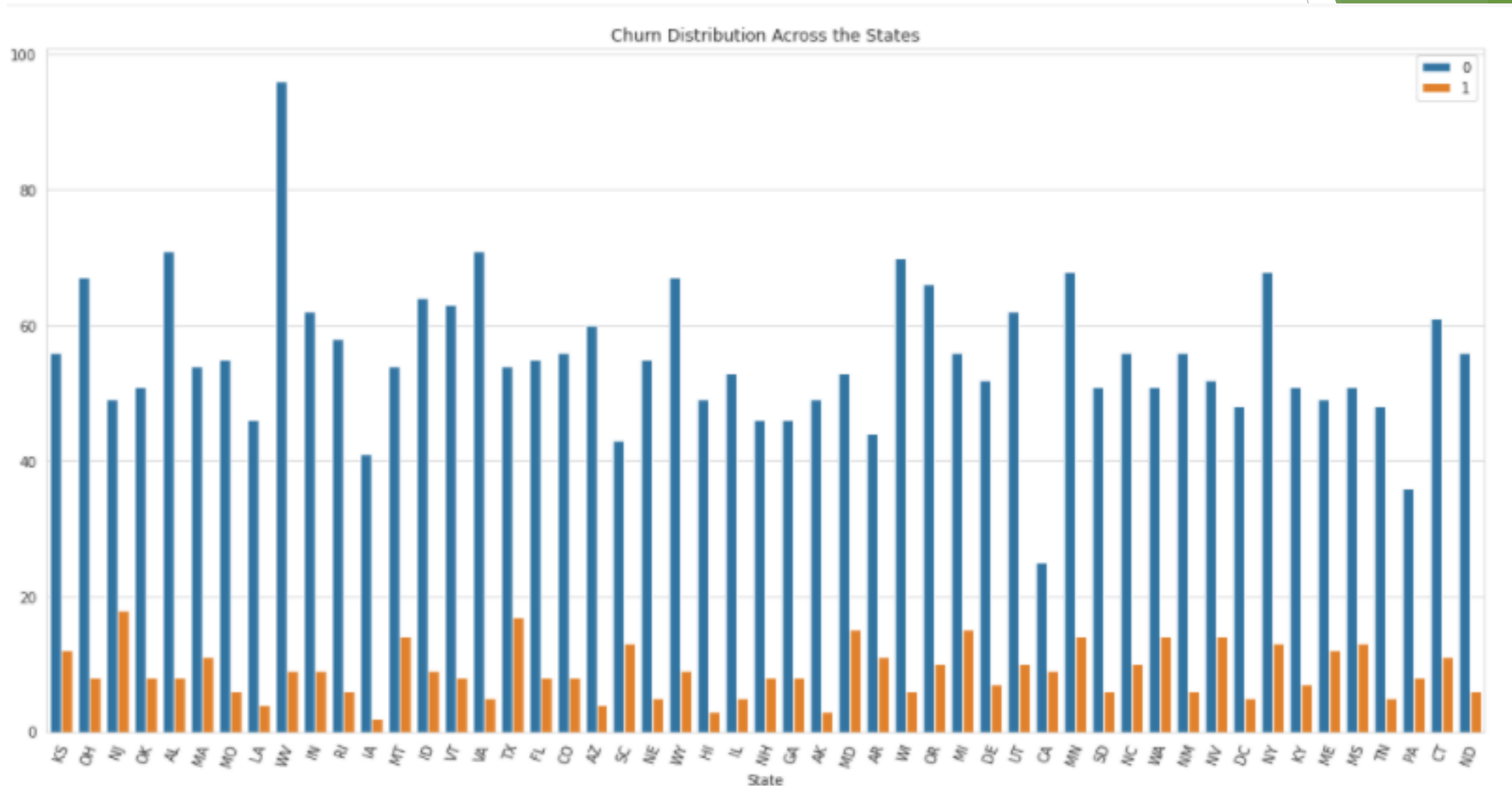
Top 5 States for Churned Users

▶ From bar plot we can observe that users of these states are prone to churn. According to my hypothesis, following would be the factors that should be implemented for stopping churn in these states:

▶ 1. In top three states New Jersey, Texas and Maryland we can definitely choose aggressive pricing strategies and network up gradation as well as improving voice quality of calls in these states.

▶ 2. We have to look out for our competitor in these states also as observation says that there may be strong competition.

Top 5 Churned States
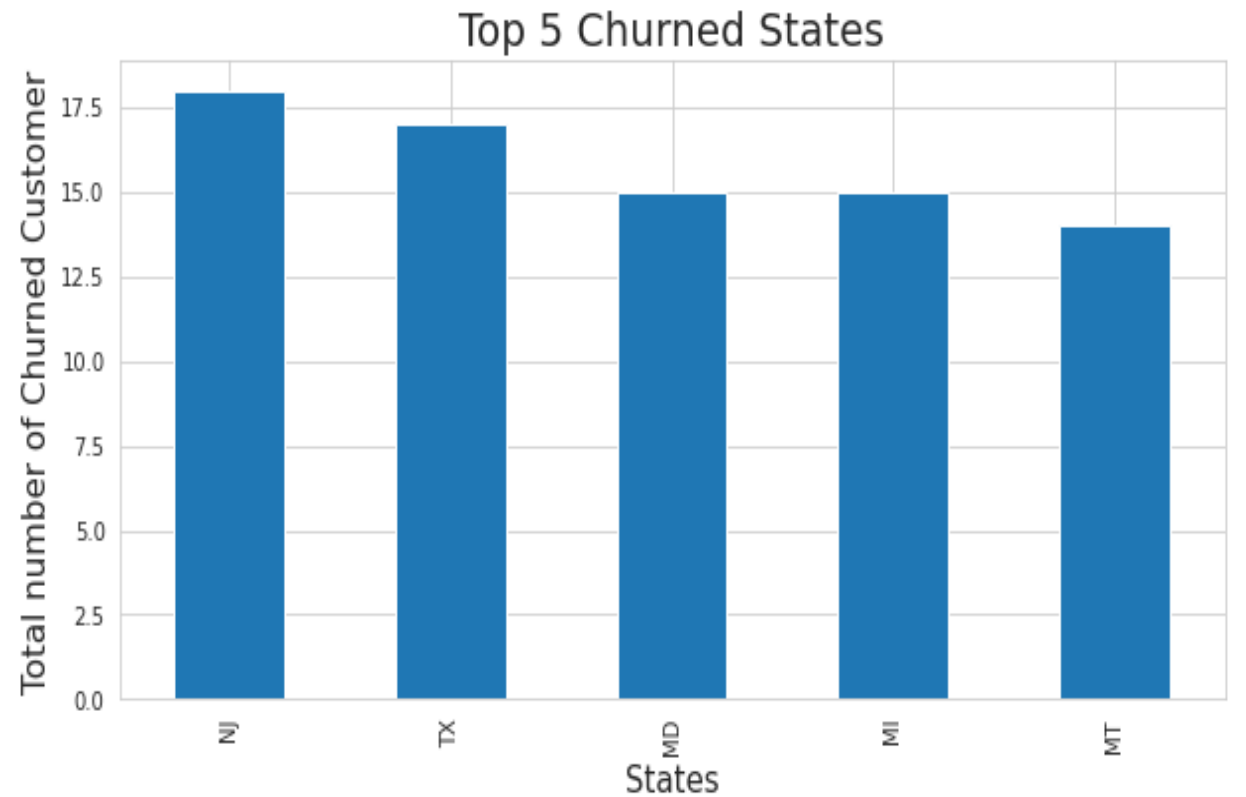
Relation between Churn vs area code

Area code	Churn	
408	0	709
	1	117
415	0	1404
	1	221
510	0	708
	1	117

▶ From above plot we can see that the number of churned users percentage from each area code is same and area 415 has more number of churn .
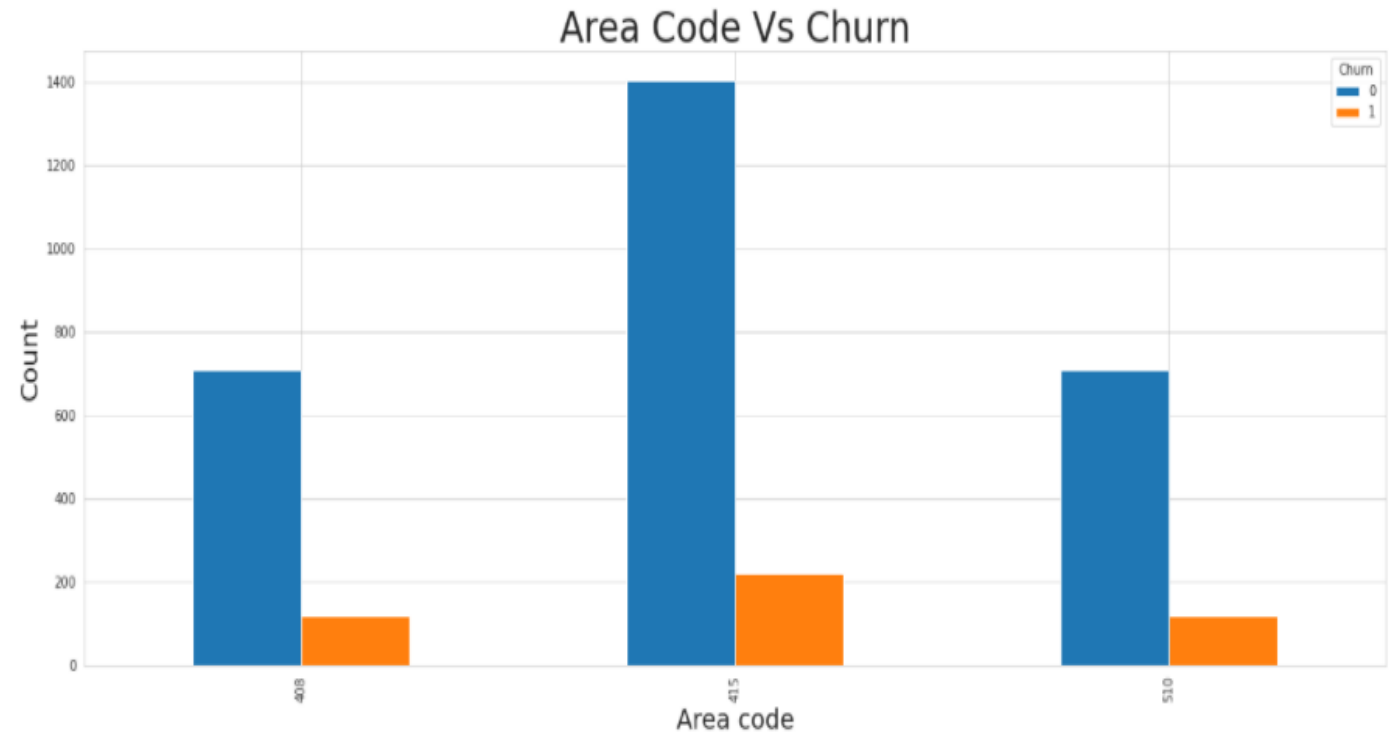
Churn distribution vs voice mail plan

▶ From count plot we can see that the number of churned users who opt for voice mail plan is lesser than compare to those who did not opt for that, so we can predict that our voice mail plan is doing good hence we can improve it further
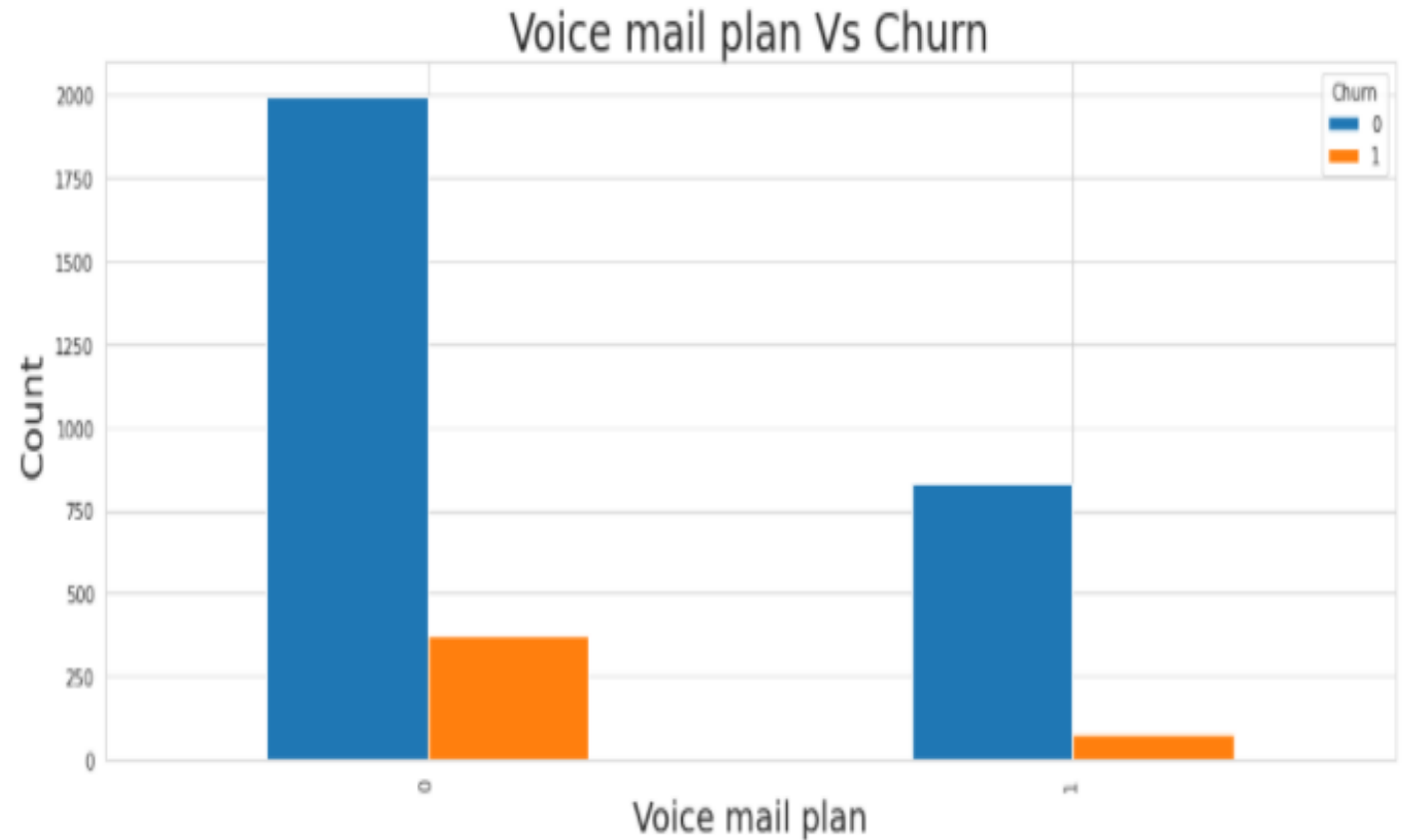
Relation between Churn and Total day minutes, calls and charge
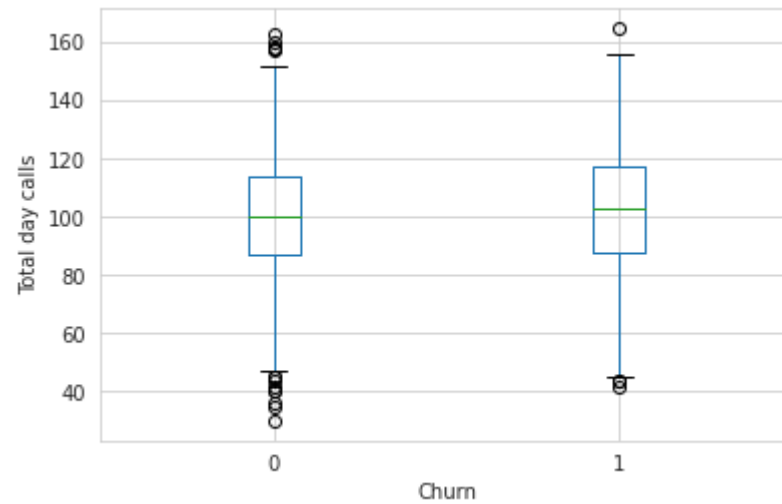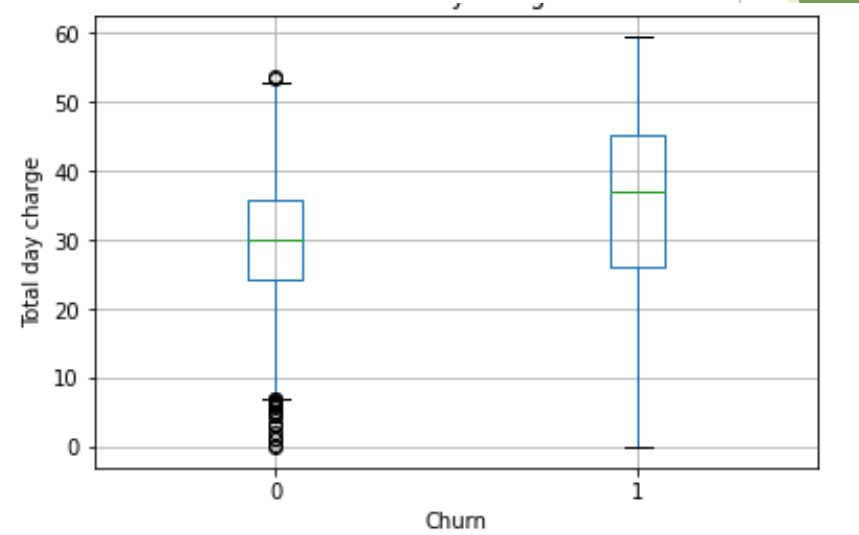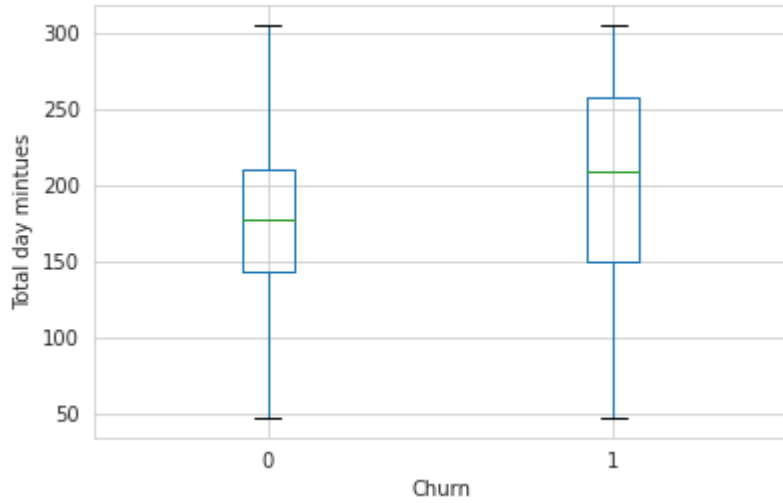
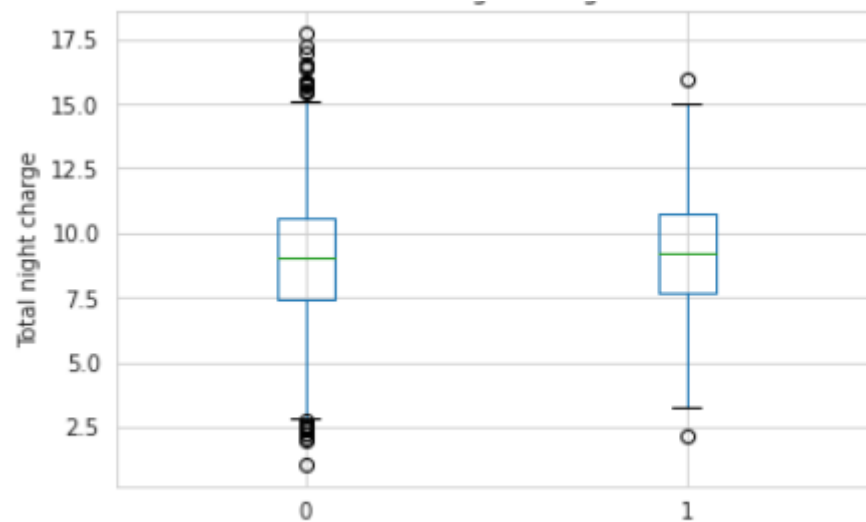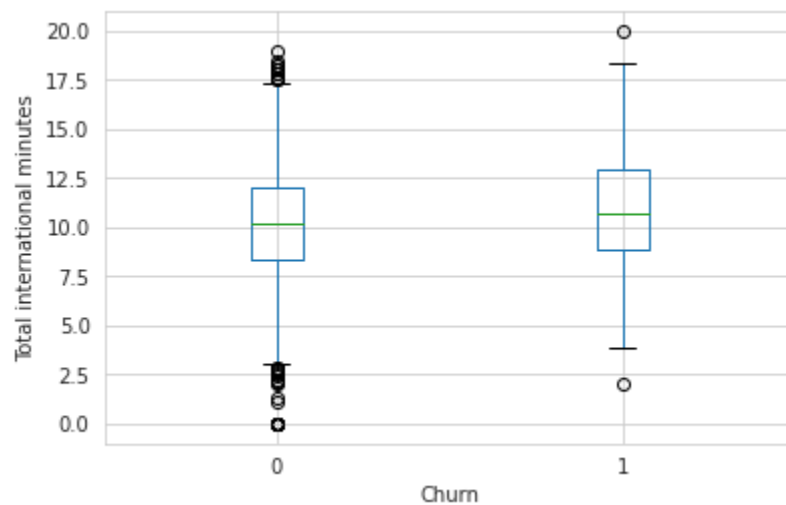# Relation between Churn and Total eve minutes, calls and charges
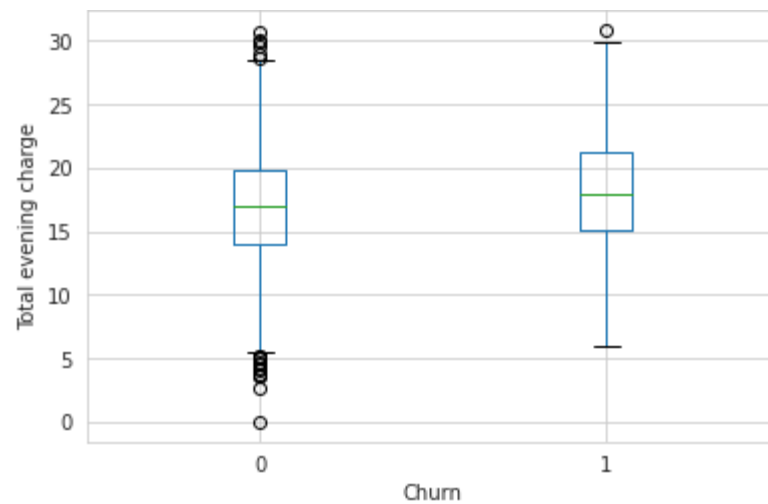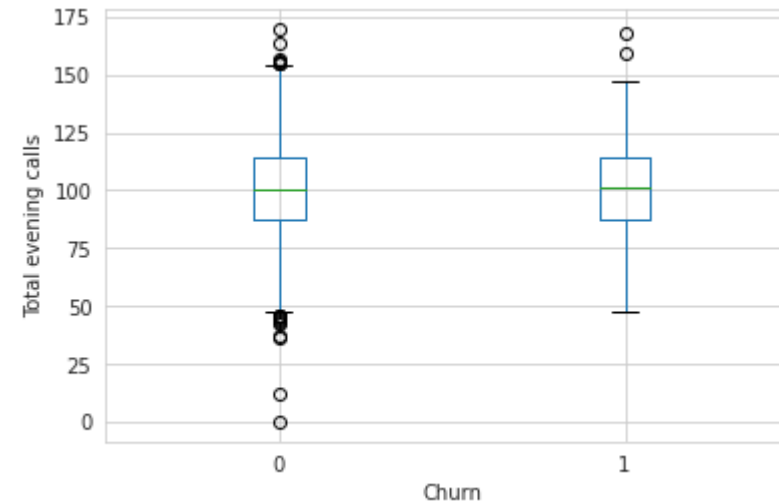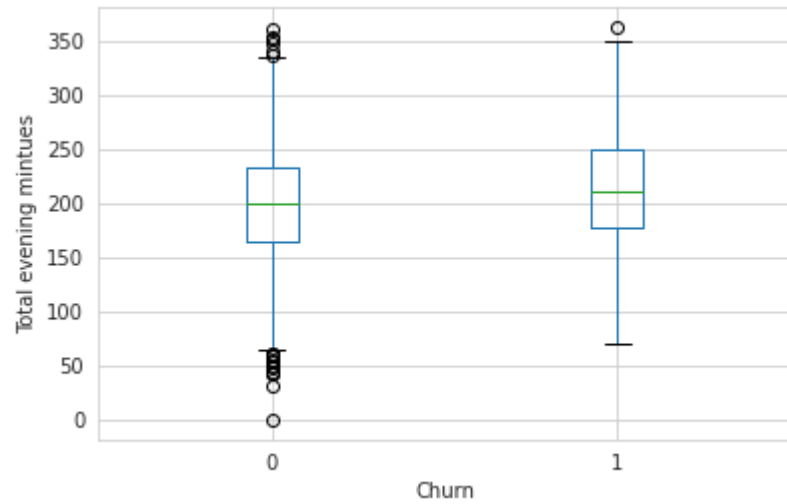
Relation between Churn and Total night minutes, calls and charge
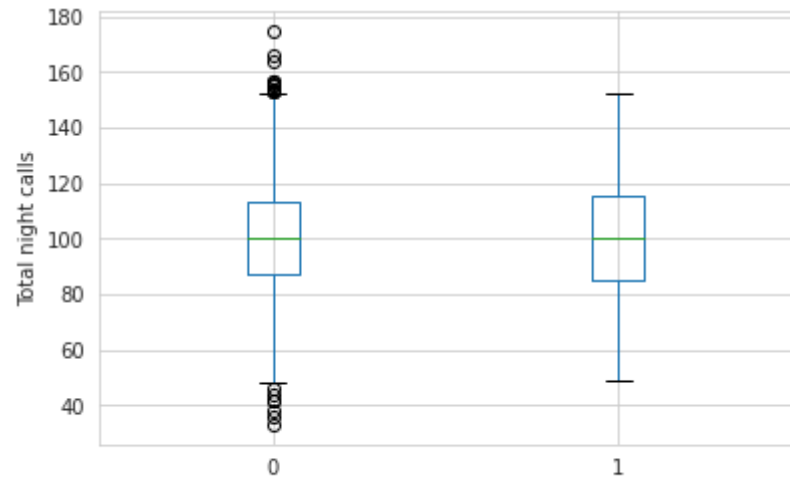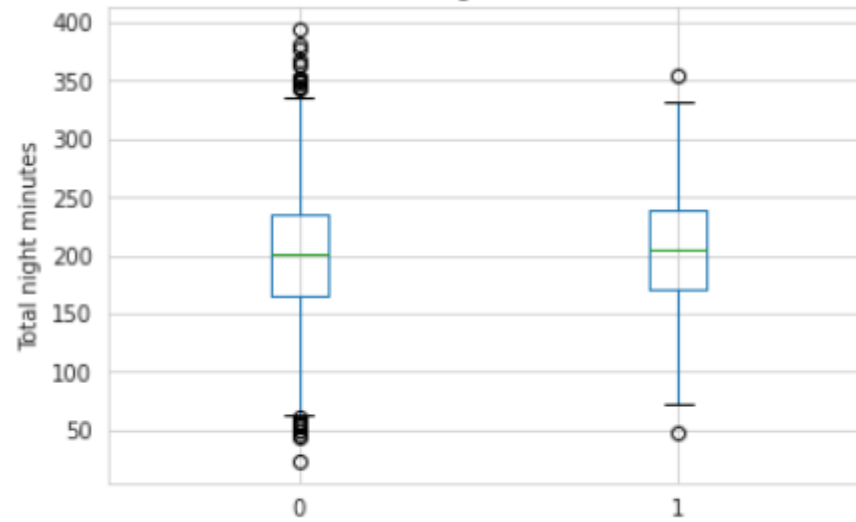
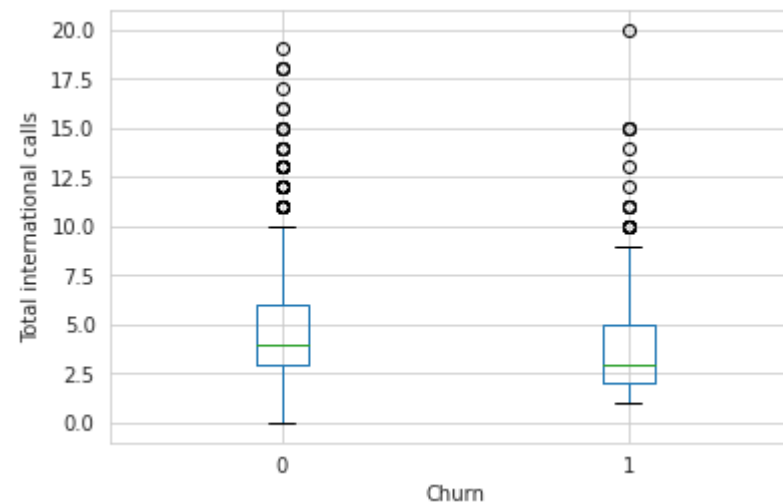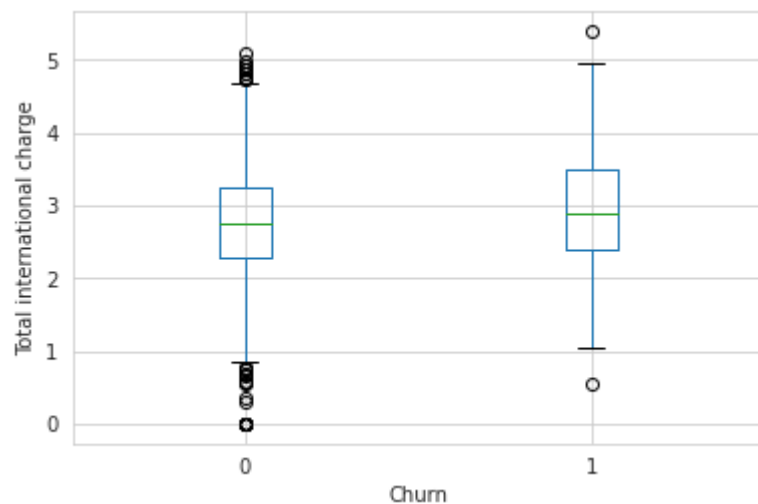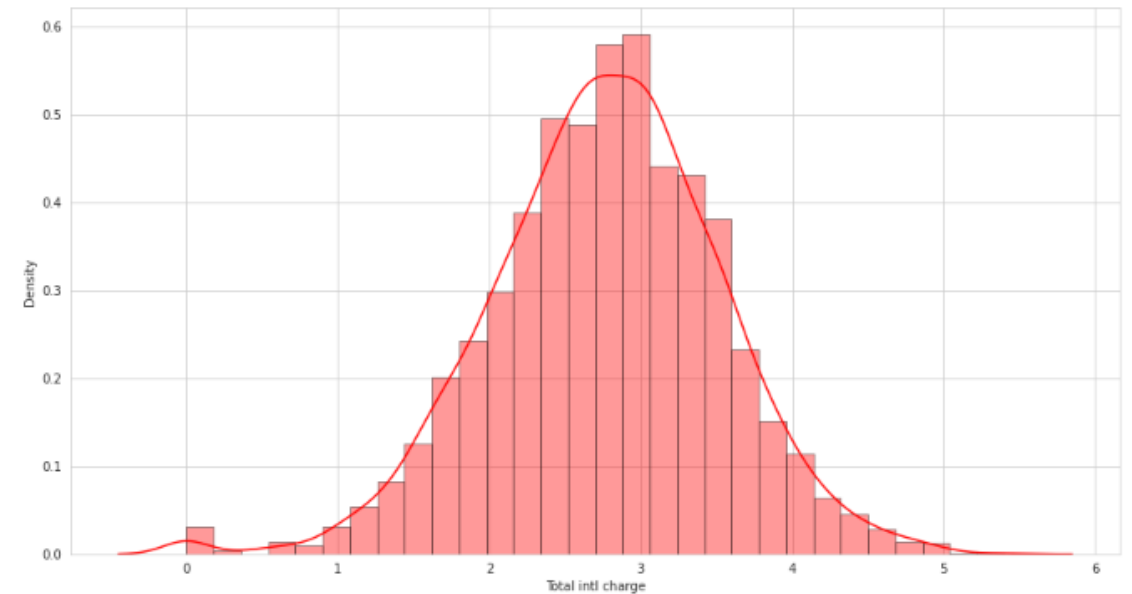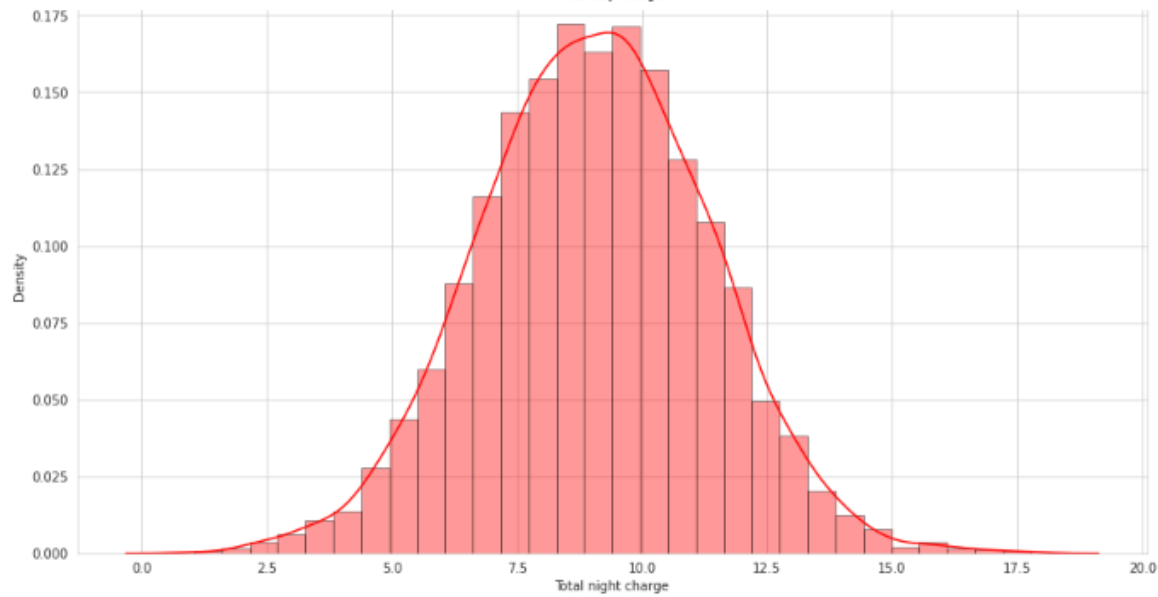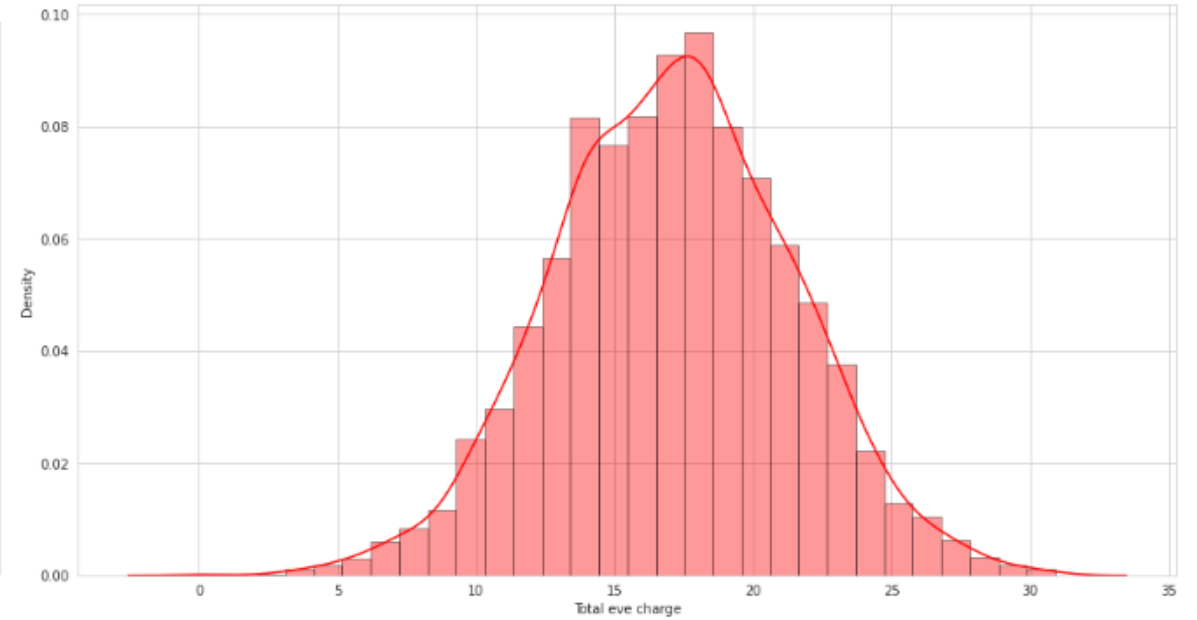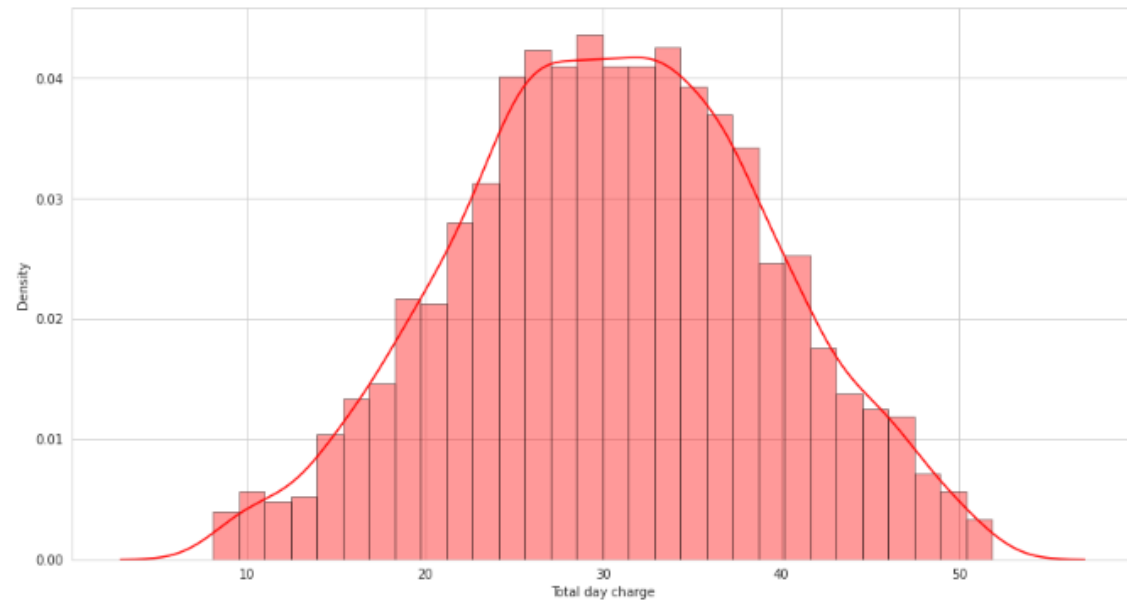# Relation between Churn and Total Intl minutes, calls and charge

# Frequency distribution of all the charges features in datset

- According to our findings from our above box-plot we can observe that users who are spending more minutes are tend to switch to other operator so following would be the factors that should be implemented:

- Implementing Different Pricing Strategy

- Network Up gradation .

- Implementing international Calling Rate Optimization would need to lower churn rate

Relation between Churn vs International plan

| International plan	Churn	
0	0	2637
	1	322
1	0	184
	1	133

International plan Vs Churn
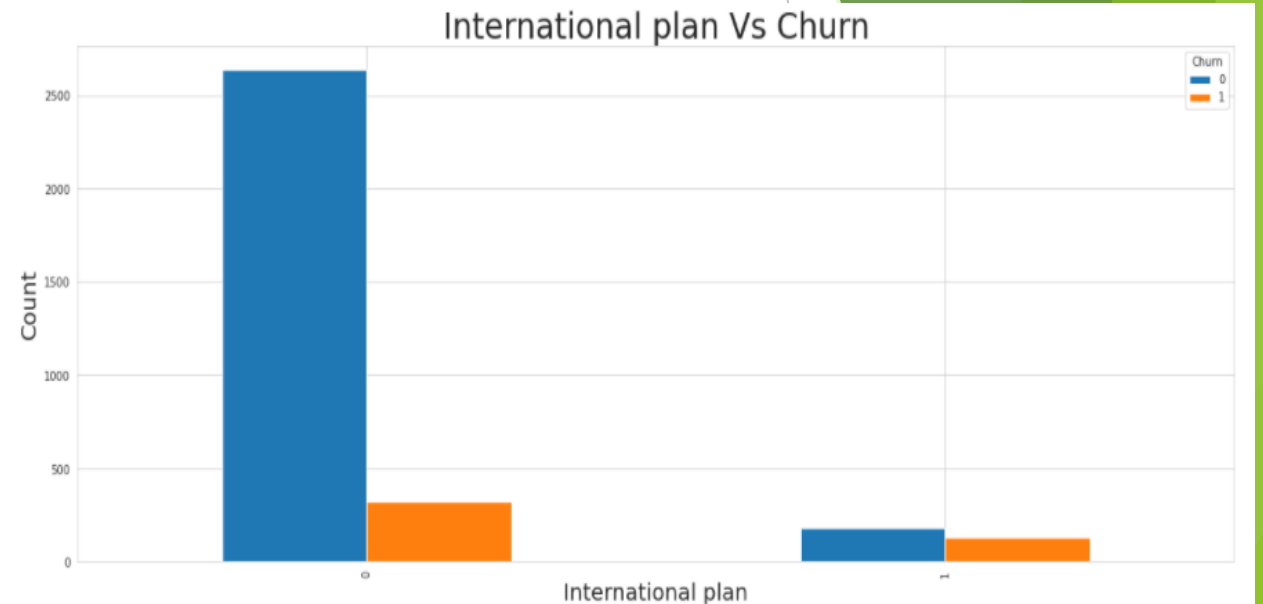
The percentage of churned user who opted for international plan is 42%. It means there is some problem with the pricing or voice call quality for International plan opted users

- Need to improvement of Voice Quality during calls.

- Network Upgradation for international calls.
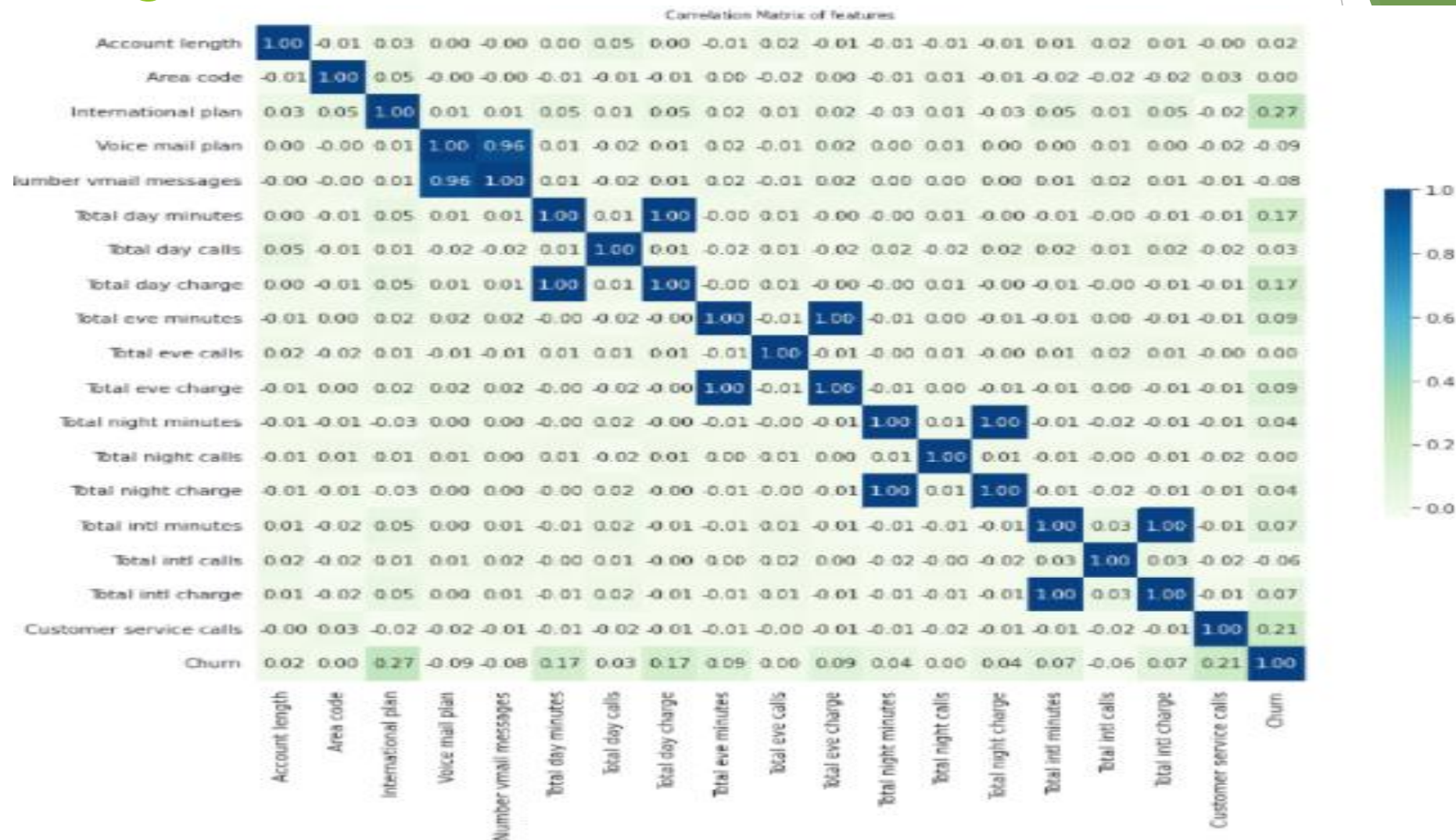
Relation between Churn and Customer service

Customer service calls Vs Churn
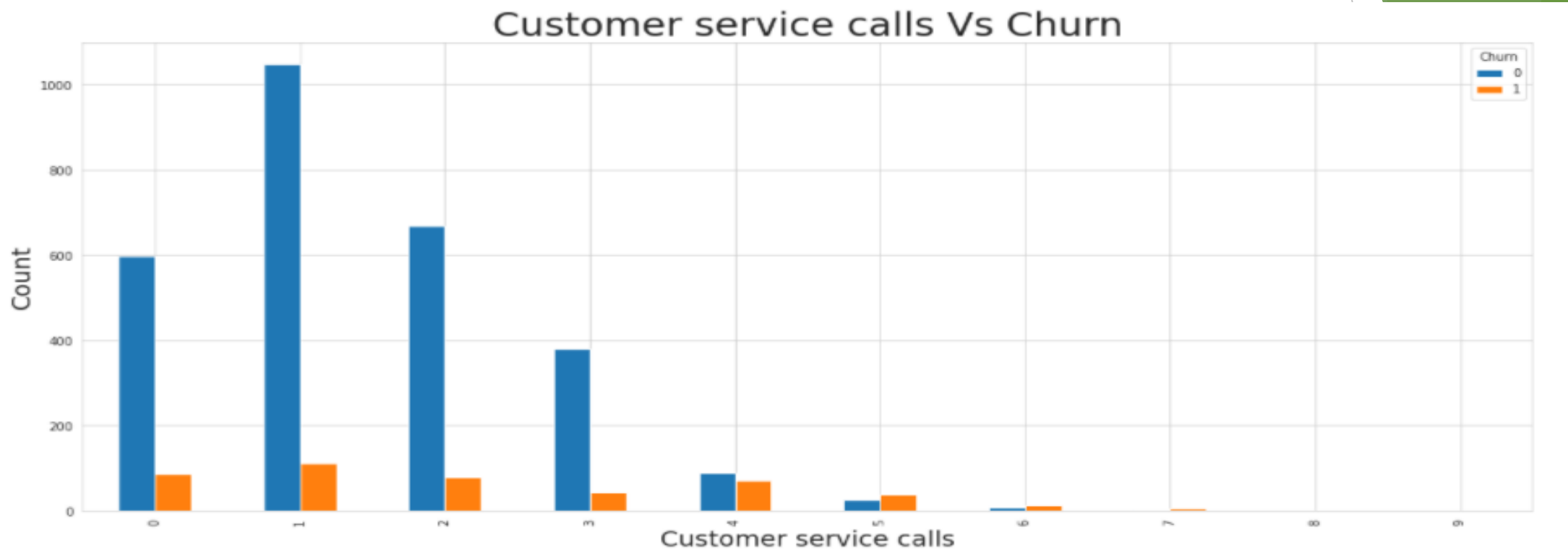
- We have to promote our customer services.

- Strengthening the customer service department can be also be helpful.

- We can take Feedback from customer who called customer services

Checking correlation of all the features in our dataset

Correlation Matrix of Features

Conclusion

▶ In this project ,we tried to analyze the factor of customer churn. First we did inspection of dataset on a basic level. We looked for missing values and check the outlier.

▶ Then we used the matplotlib and seaborn to do Exploratory Data Analysis on given data by plotting different graphs like count plot, pie chart, bar plot, boxplot, heat map from this then we got useful insights like: churn percentage of customer is 14%, customer having more daily charge will be more chances of churn, states like New Jersey, Texas and Maryland have higher churn rate, customer having international plan have more churn rate, customer having less customer service call have more churn rate

Suggestions

- Upgrading network to improve services for long duration users.

- Improving Pricing Strategies.

- Optimizing and Updating International Call Rates.

- Implementing a better network infrastructure in New Jersey, Texas and Maryland Areas where there is more Churn Rate.

- Improvement in the customer service can be done to reduce the factors which cause the churn.

- Decreasing the prices as the talk-time increases can be an effective way to reduce the churn.

Thank You