

Capstone Project - 1

Telecom Churn Analysis

Gulzar

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**
- **Suggestions**

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

Telecom Data provided by an French multinational telecommunications corporation The Orange S.A.. Their data focuses on United States region. During this data recording period, 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



Problem Statements

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(continued)

- Data inspection is used To ensure that we are dealing with the right dataset and for a clear view of data set at every stage of the transformation process.
- Data inspection is the act of viewing data for verification and debugging purposes, before, during, or after a translation
- During this step we checked shape of data, data types, unique value in column and statistics information



Data Inspection(continued)

- 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, int, float, or bool
- 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 14.5% customers are churned and remaining are not churned.



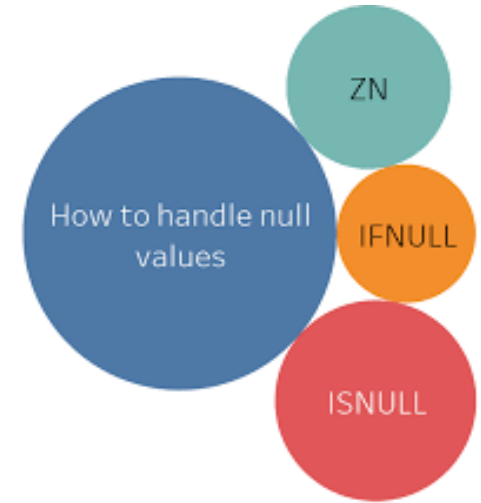
Data Cleaning

Data cleaning is very important process in EDA. because raw data sometimes it consists of null values, missing values, duplicate values and outliers in data set, irregular format due to all this it's very difficult to create insights from dataset.



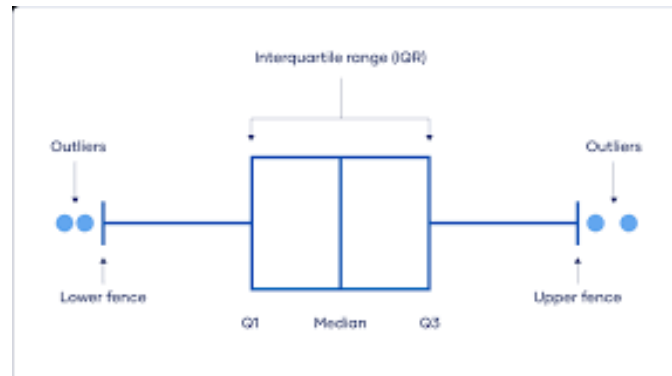
Data Cleaning(continued) : Null values

- 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 Orange S.A. telecommunication dataset, there are no null values.

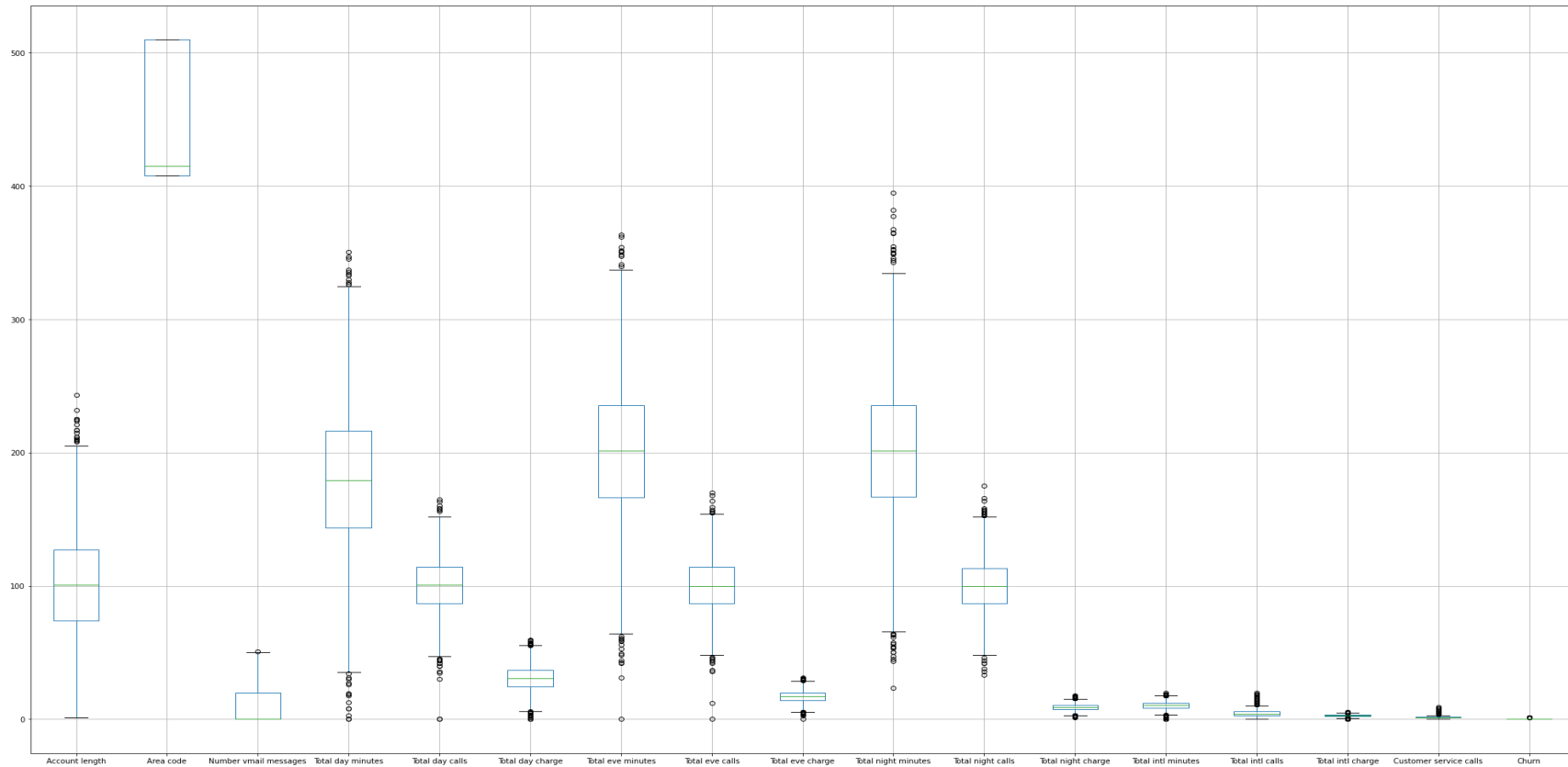


Data Cleaning(continued): Outliers

- 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 behaviours.
- 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 telecom churn data set all the columns having outliers.



Data Cleaning(continued): Outliers

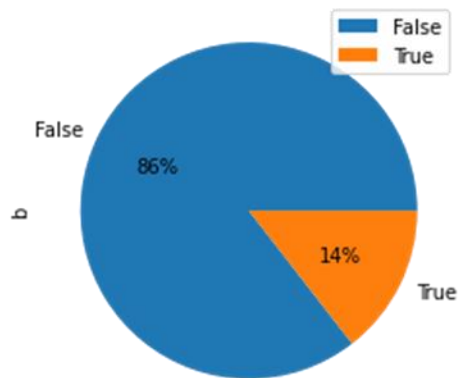
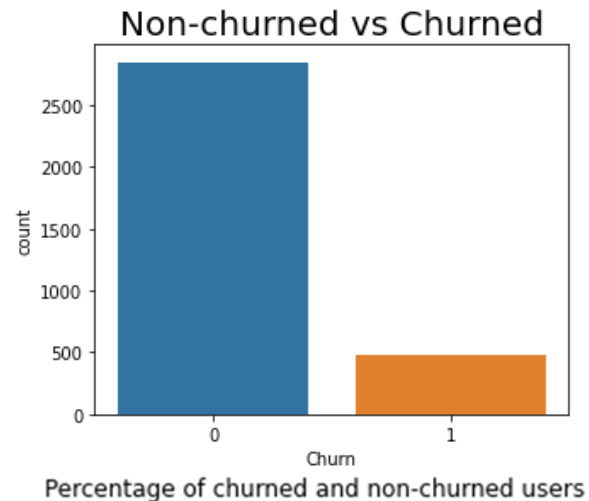


- **Data Visualization** represents the text or numerical data in a visual format, which makes it easy to grasp the information the data express. We, humans, remember the pictures more easily than readable text, so Python provides us various libraries for data visualization like matplotlib, seaborn etc.
- In this telecom churn EDA we will use Matplotlib and seaborn for performing various techniques to explore data using various plots.

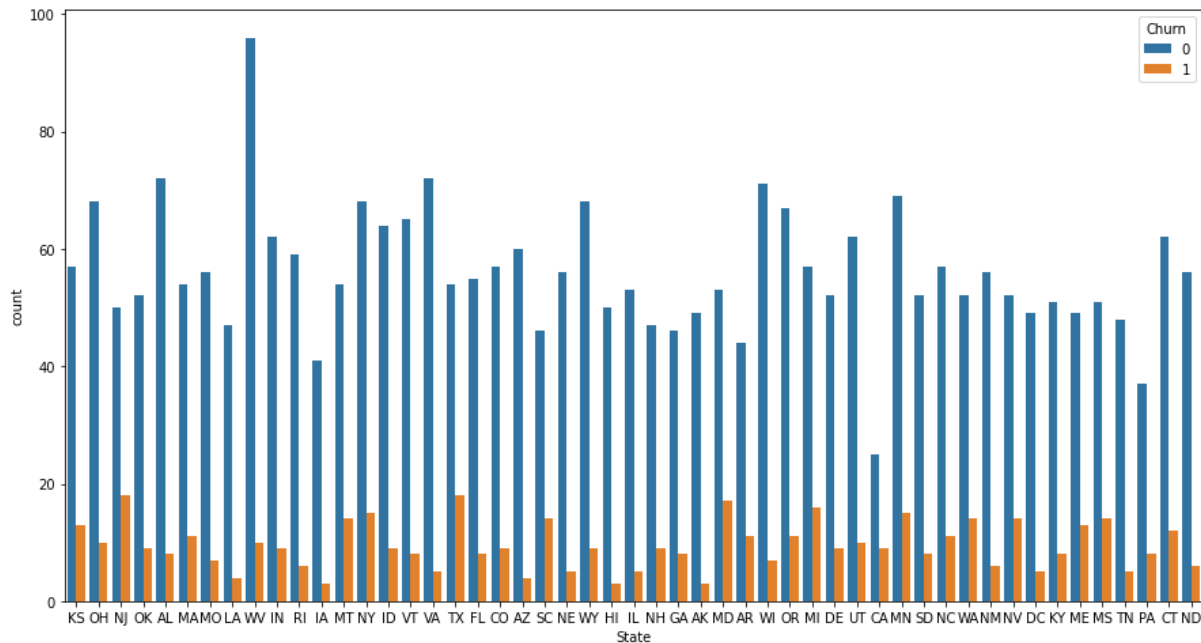


EDA(continued): Target variable

- 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 analysing the data.



EDA(continued): Comparing target column with states

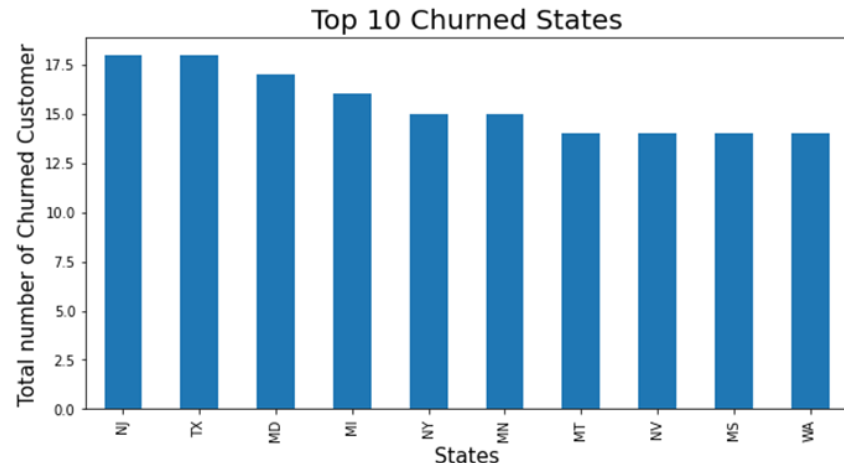


EDA(continued) : Top 10 States for Churned Users

From bar plot we can observe that users of these states are prone to churn.

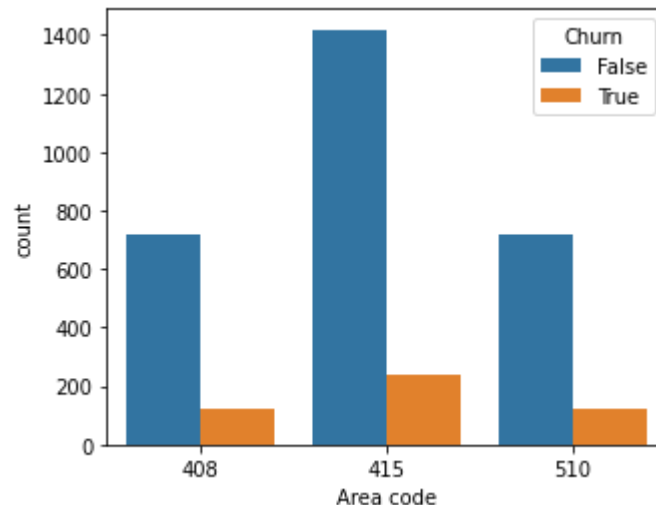
According to my hypothesis, following would be the factors that should be implemented for stopping churn in these states:

1. In top three states New Jersey, Texas and Maryland we can definitely choose aggressive pricing strategies and network upgradation 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.



EDA(continued) :Relation between Churn vs area code

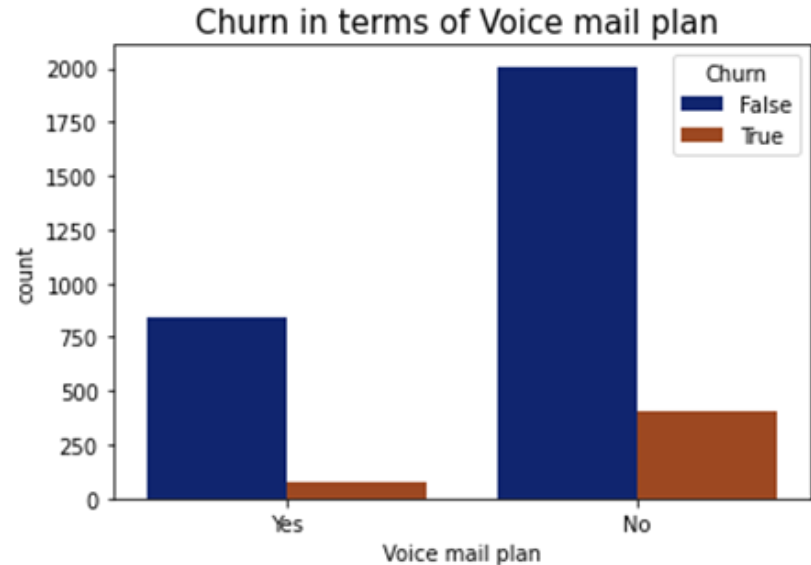
Churn			
Area code	False	True	Perc_Churn
408	716	122	14.558473
415	1419	236	14.259819
510	715	125	14.880952



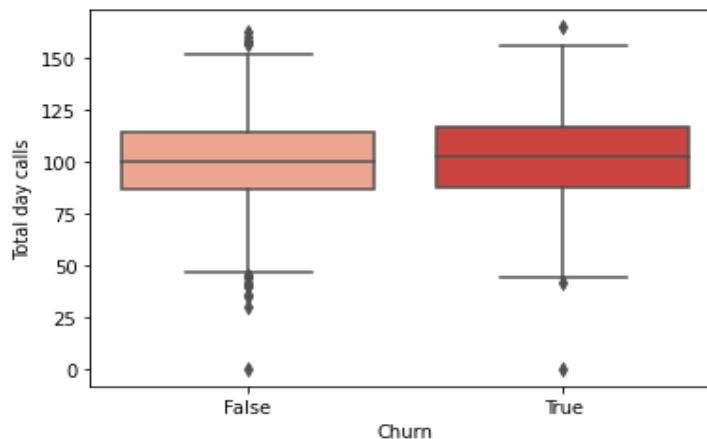
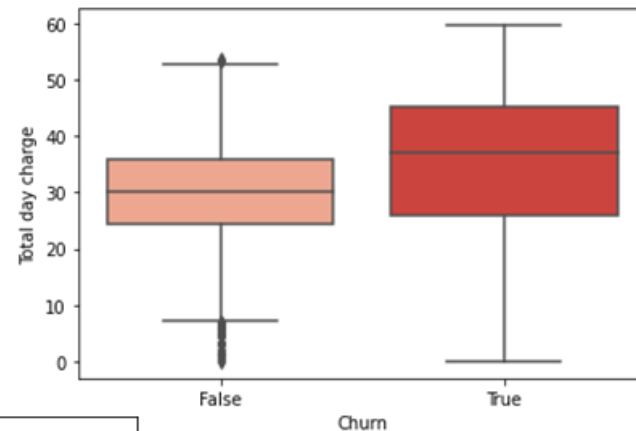
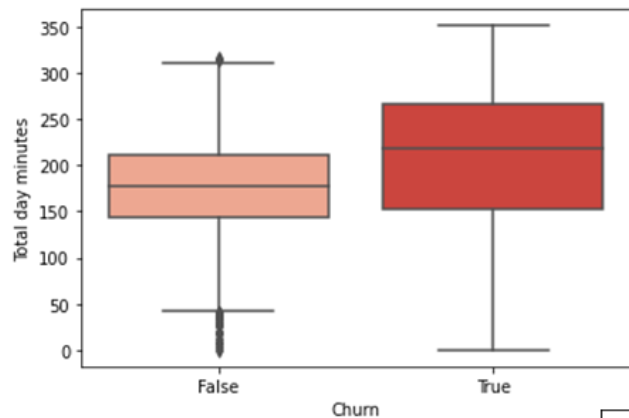
- From above plot we can see that the number of churned users percentage from each area code is same and area 415 has more number of churn .

EDA(continued): Churn distribution vs voice mail plan

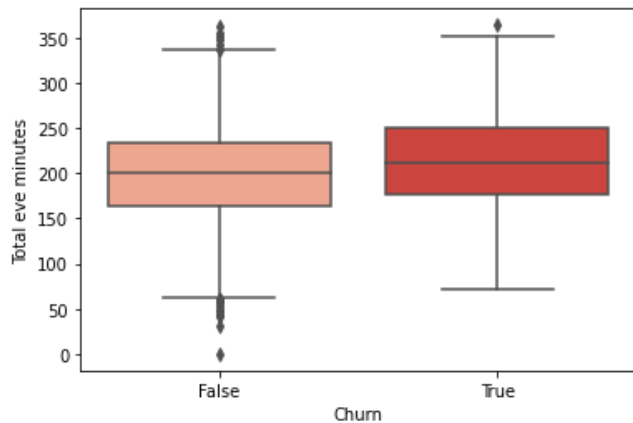
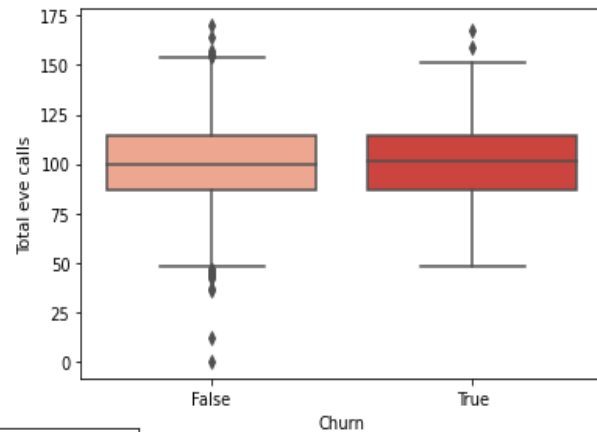
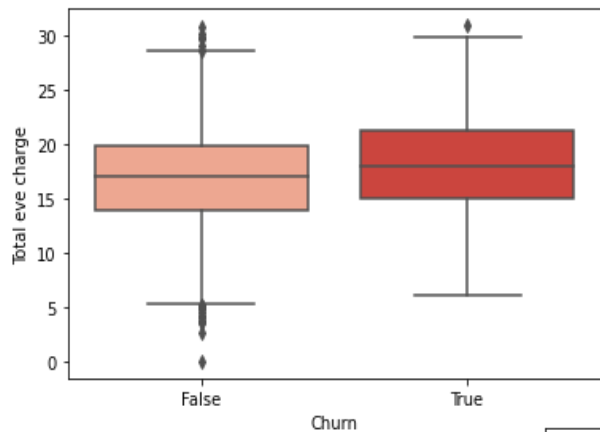
From count plot we can see that the number of churned users who opt for voice mail plan is lesser than compare to those who did not opt for that, so we can predict that our voice mail plan is doing good hence we can improve it further.



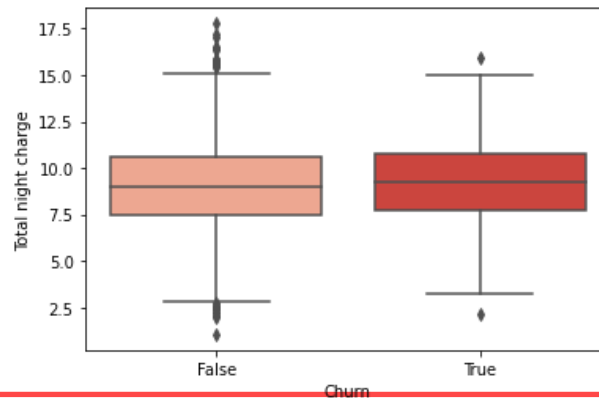
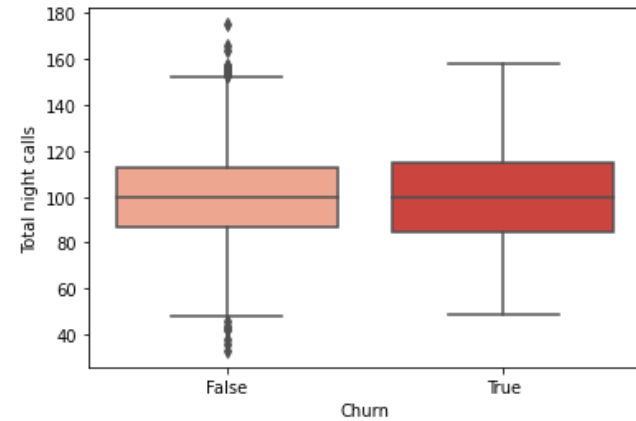
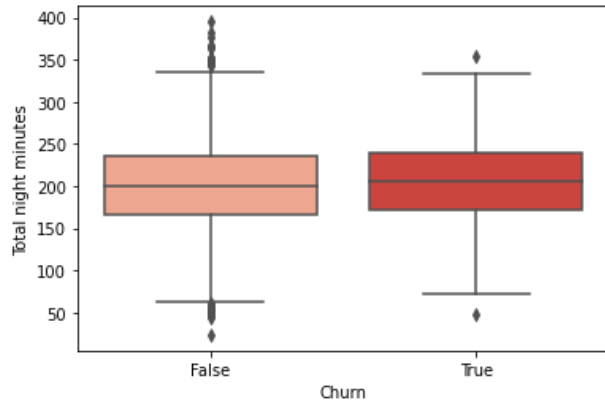
EDA(continued): Relation between Churn and Total day minutes, calls and charge



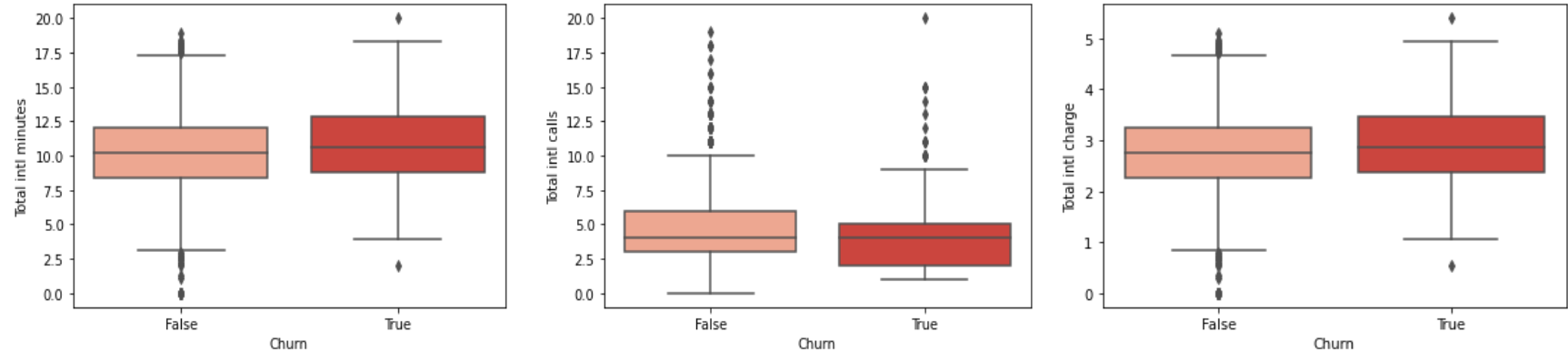
EDA(continued): Relation between Churn and Total eve minutes, calls and charges



EDA(continued): Relation between Churn and Total night minutes, calls and charge



EDA(continued): Relation between Churn and Total Intl minutes, calls and charge

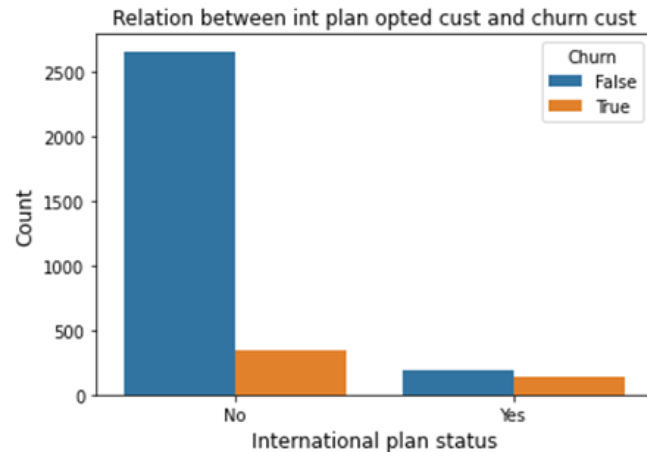


According to our findings from our above box-plot we can observe that users who are spending more minutes are tend to switch to other operator so following would be the factors that should be implemented:

- **Implementing Different Pricing Strategy**
- **Network Upgradation .**
- **Implementing international Calling Rate Optimization would need to lower churn rate.**

EDA(continued): Relation between Churn vs International plan

	Churn-	False	True	Perc_ Churn
International plan				
No		2664	346	11.49
Yes		186	137	42.41



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.

According to our findings after monitoring the voice quality of international calls, following would be the factors that should be implemented :

1. Need to Upgrade or make smarter use of technologies like VoLTE for improvement of Voice Quality during calls.
2. Network Upgradation for international calls.

EDA(continued): Relation between Churn and Customer service

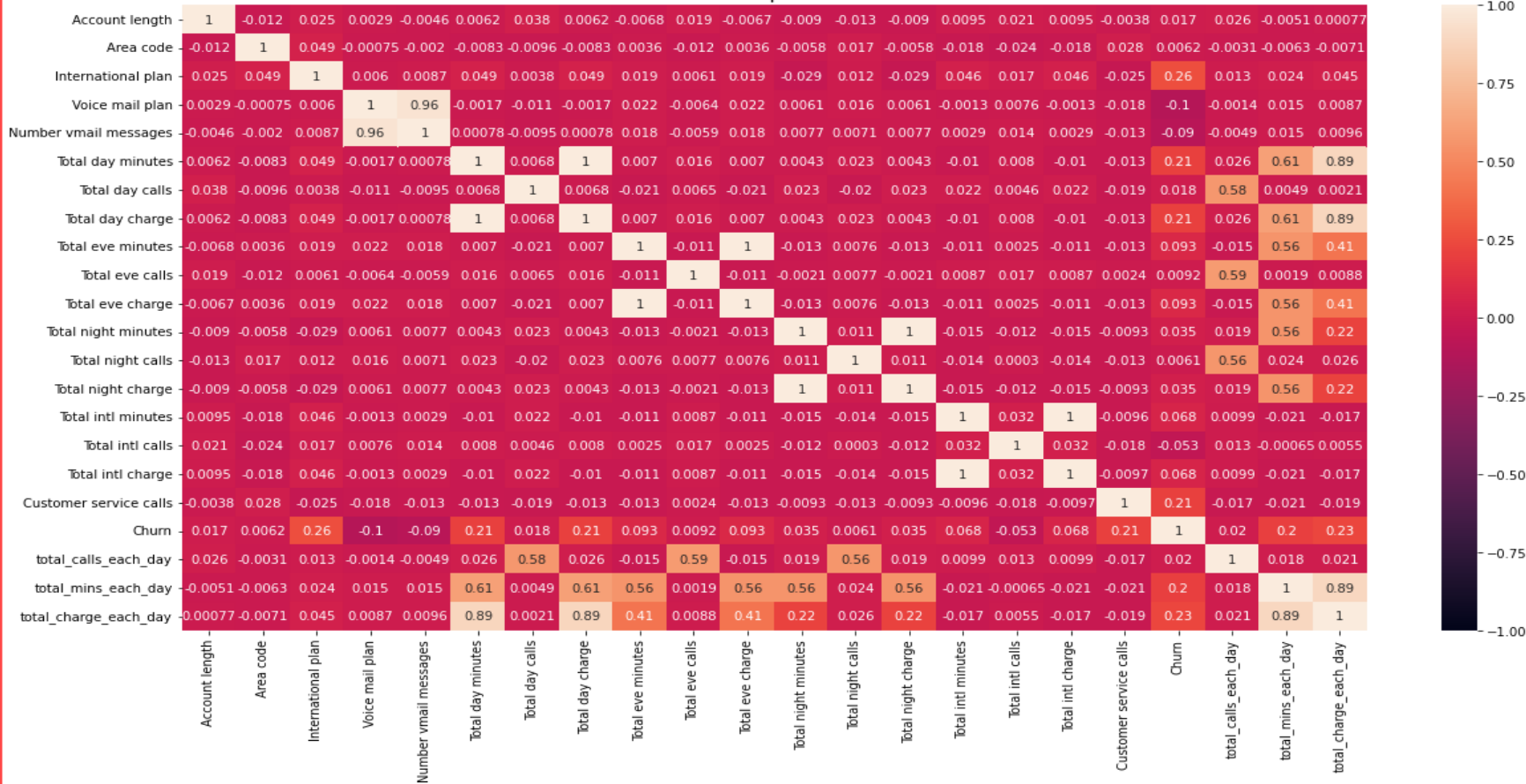


According to our findings, following would be the factors that should be implemented:

- 1. We have to promote our customer services.**
- 2. Strengthening the customer service department can be helpful also.**
- 3. We can take Feedback from customer who called customer services.**

EDA(continued): Checking correlation of all the features in our dataset

Heatmap of data



Conclusion

- In this project ,we tried to analyse 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(EDA) 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

Here is our suggestions to prevent churn :

- 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.



Further more Improvement we can suggest and discuss more strategies to the company by collecting other data and through a domain expert.

THANK YOU