Capstone Project

EDA on Telecom Churn

Team Anonymous

1.Sumit Jadhav

2.Rushikesh Shinde

3.Akash Jadhav

4.Aniket Okate

Problem Statement

Formerly France Telecom S.A., is a French multinational telecommunications corporation. The Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features), along with a churn label specifying whether a customer canceled the subscription.

Explore and analyze the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.

Objectives

1.To get summary of dataset.
2.To clean the dataset.
3.To perform data preprocessing and make final dataset.
4.To Visualization to find Behavior of Data.
(Number of churners, Distribution of Numerical variables by density plot, Distribution of Categorical variables by count plot, Check outliers in the dataset, To check correlation with target variable)
5. Visualization for Other Analysis
-Checking minutes, calls and charge by day, evening and night
-Impact of customer service calls on Churn
-Churned and non churned data in specific Areas using their area codes
-In which state have more number of churners
-To find churn count according to their plans

Data Description

1.State: Names of various states

2.length: Area length

3.Code: Area code

4.International plan: YES or NO

5.Voice mail plan: YES or NO

6.Number vmail messages: Number of voice mails in area

7.Total day minutes: Average minutes in area (day)

8.Total day calls: Average number of calls in area (day)

9.Total day charge: Average charge of calls in area (day)

10.Total day minutes: Average minutes in area (Evening)

11. Total day calls: Average number of calls in area (Evening)
12. Total day charge: Average charge of calls in area (Evening)
13. Total day minutes: Average minutes in area (Night)
14. Total day calls: Average number of calls in area (Night)
15. Total day charge: Average charge of calls in area (Night)
16. Total day minutes: Average minutes in area (International)
17. Total day calls: Average number of calls in area (International)
18. Total day charge: Average charge of calls in area (International)
19. Customer service calls: Number of service calls
20. Churn: True or False (target variable)

Data Summary and Cleaning

-To check Nan or null values.

-If Nan values are present then delete these nan value rows.

-To get information about numerical columns and categorical columns.

-To find unique values in each feature.

-After summary print all the information as overview.

Visualization

Visualization to find behavior of data

-For this dataset we select “Churn” as a target variable. This contains Boolean values (TRUE/FALSE).

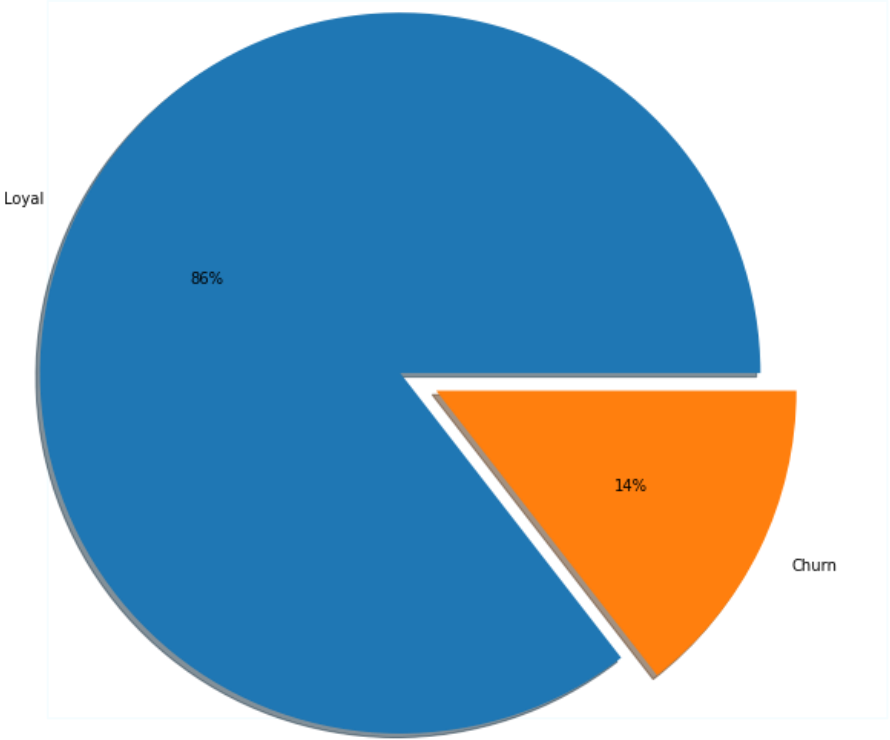
-As well as we have 19 independent variables which can used for find out behavior of dataset.

1.Number of churners

This chart shows percentages of loyal and churn customers.

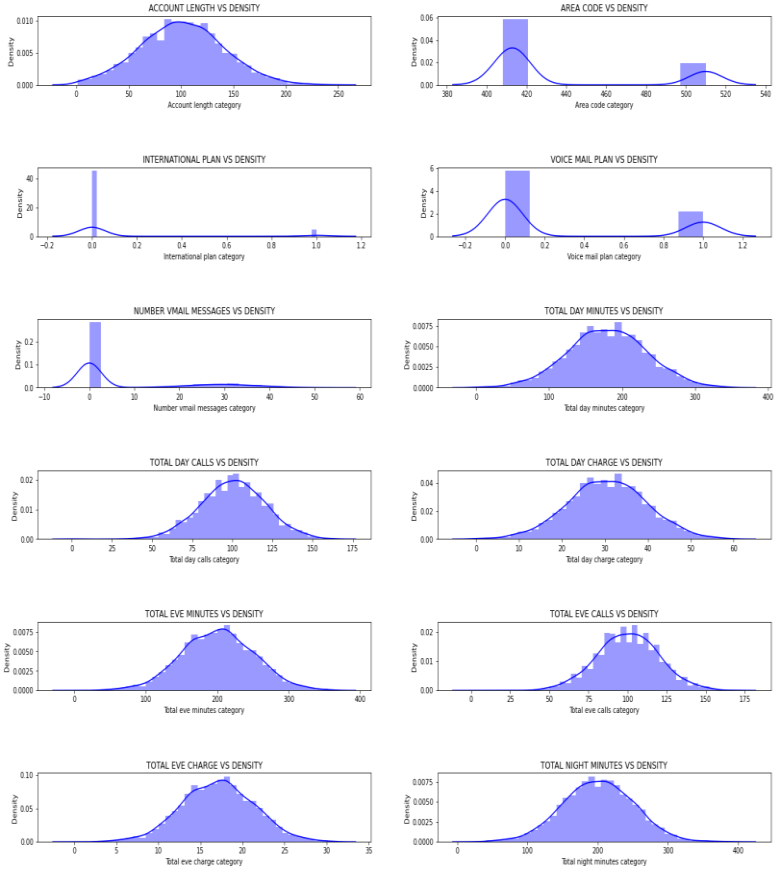
Loyal Customers- 86%

Churn Customers- 14%

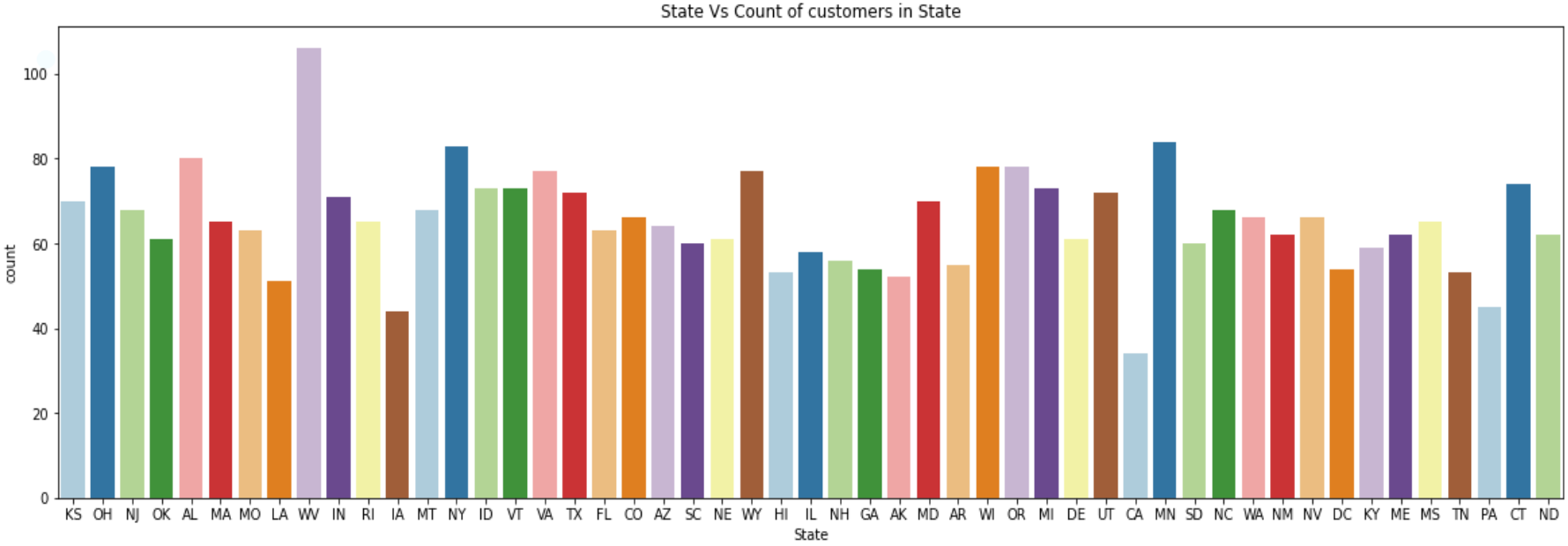


2.Distribution of Numerical variables by density plot

Most of the plot show normal distribution curve. So we can say that, maximum number of features can show normal distributed data.



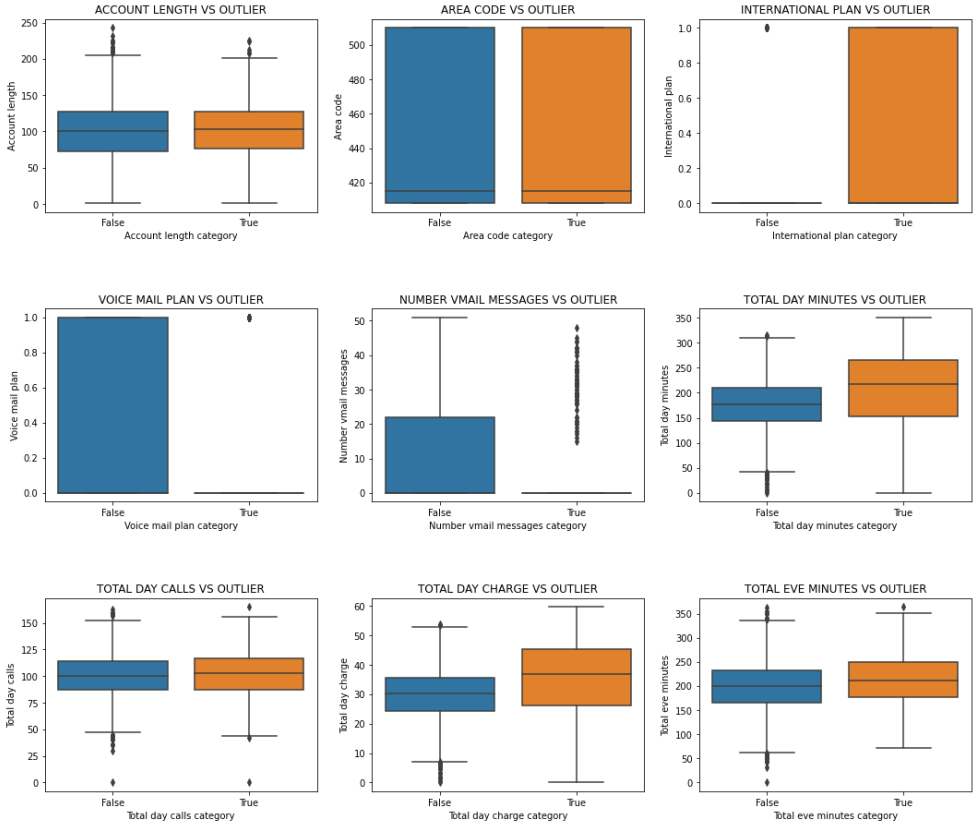
3. Distribution of Categorical variables by count plot



This plot can shows the information about each state contains how many number of customers are present

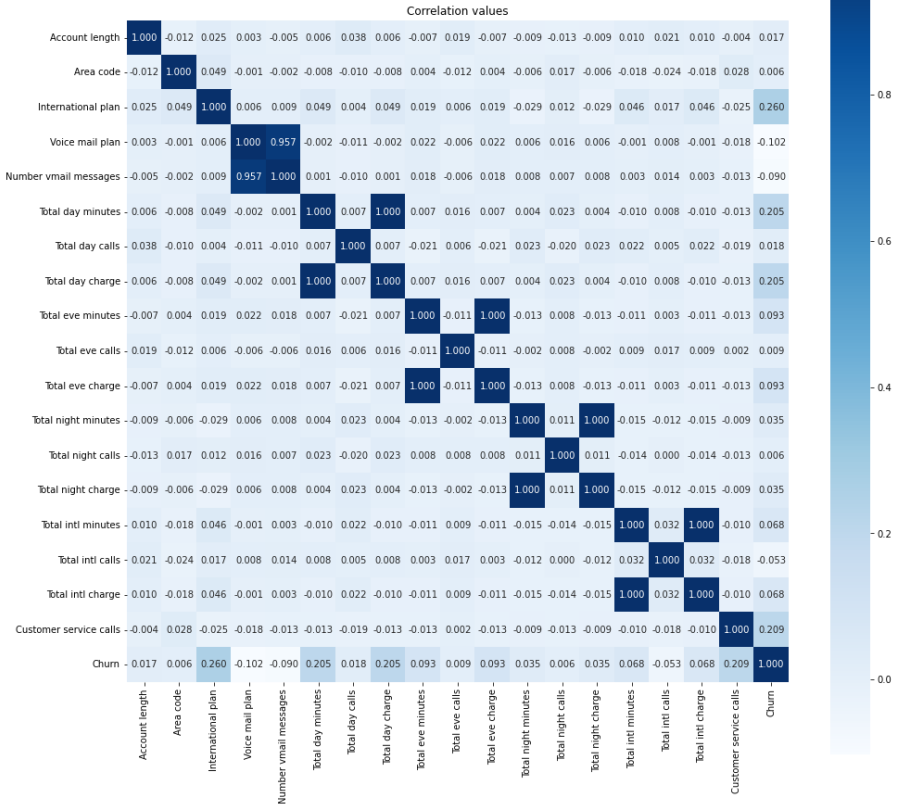
4.Check outliers in the dataset

This boxplot show us, In some cases outliers are less and in some cases number of outliers are more.



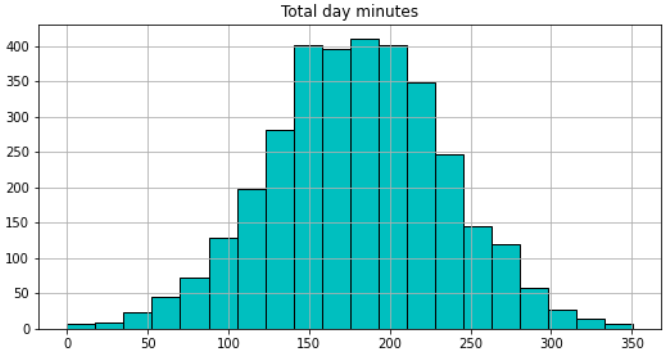
5.To check correlation with target variable

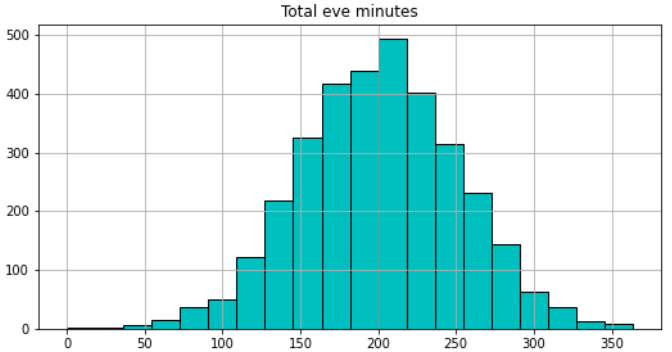
There is no as such correlation in dependent and independent features. International plan have high positive correlation value(0.26) and voice mail plan have high negative correlation value(-0.102)

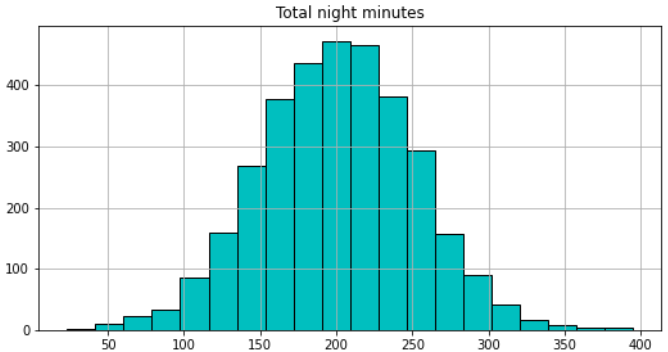


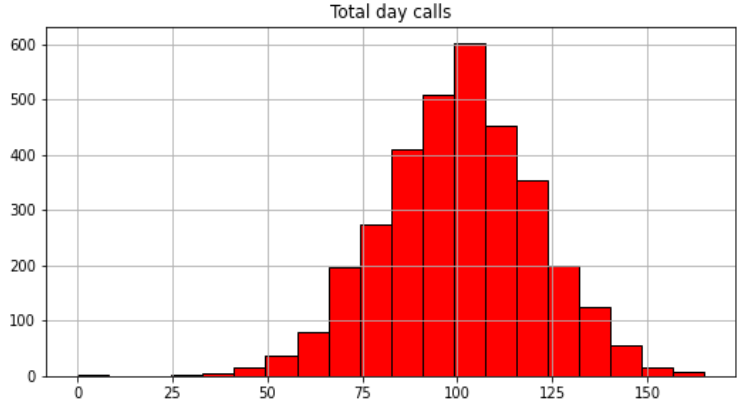
Visualization for Other Analysis

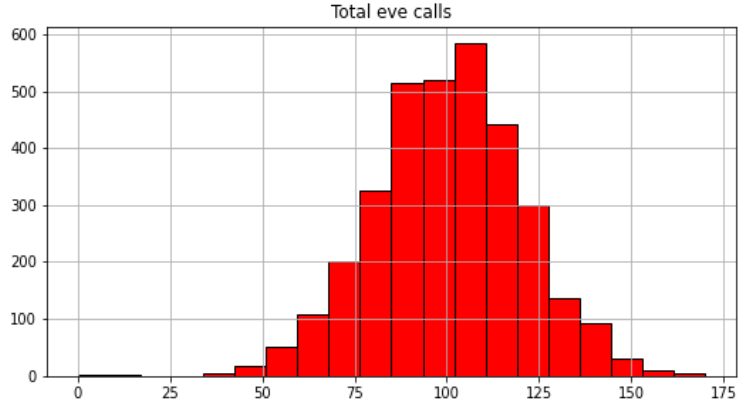
1.Checking minutes, calls and charge by day, evening and night

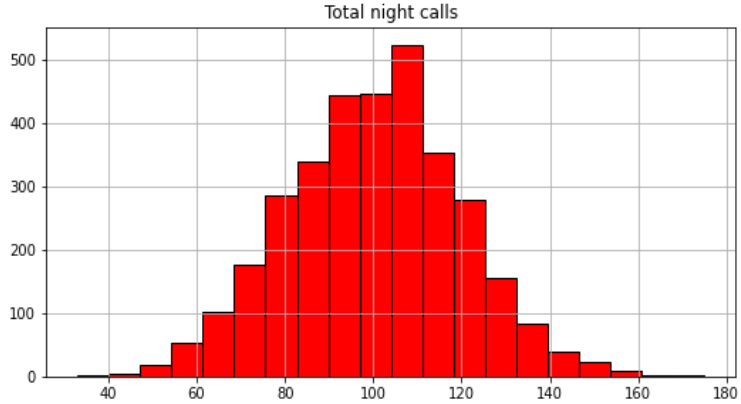


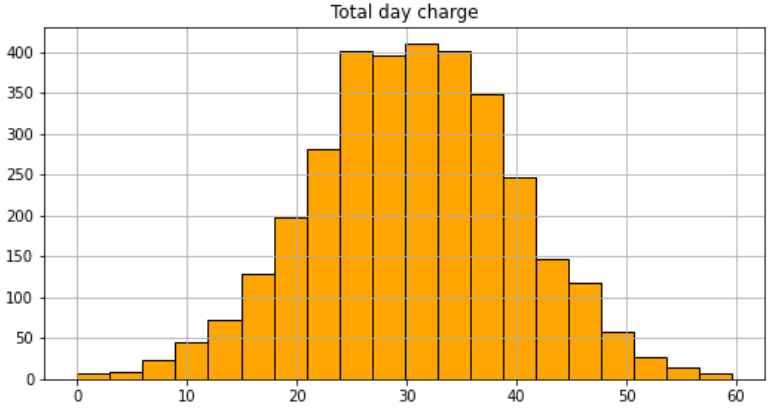


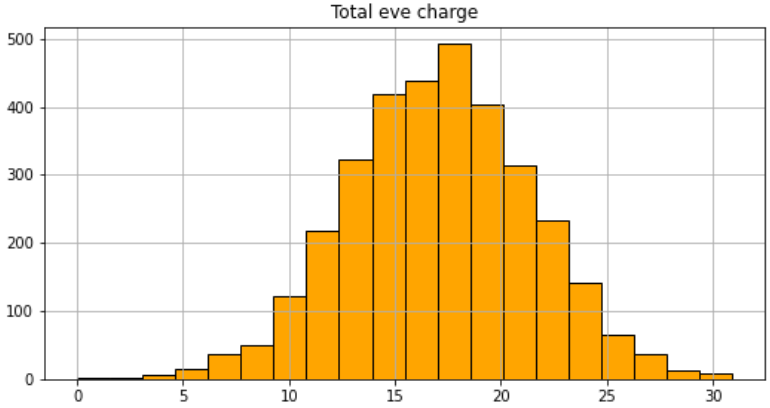


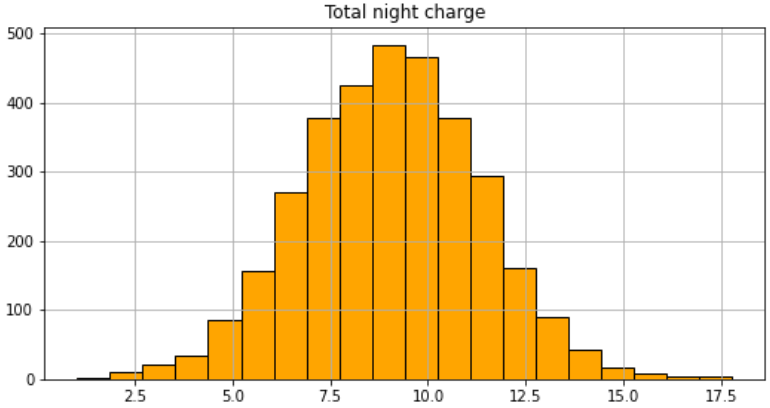

Total day calls
600
500
400
300
200
100
0
0
25
50
75
100
125
150


Total eve calls
600
500
400
300
200
100
0
0
25
50
75
100
125
150
175

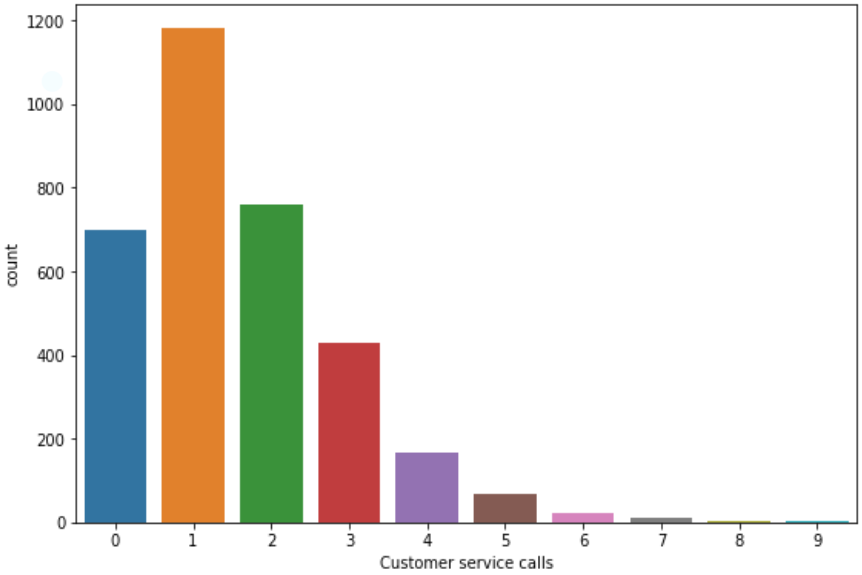

Total night calls
500
400
300
200
100
0
40
60
80
100
120
140
160
180

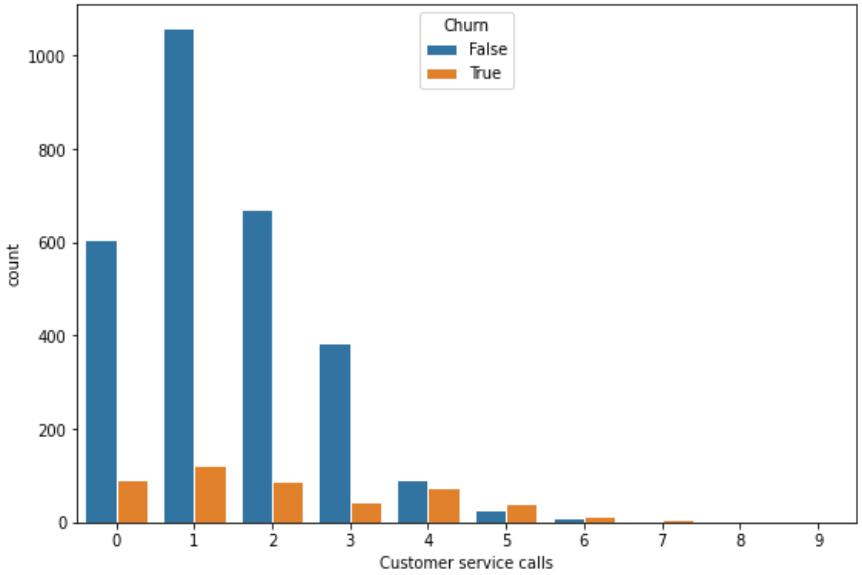

Total day charge
0
50
100
150
200
250
300
350
400
0
10
20
30
40
50
60


Total eve charge
0
100
200
300
400
500
0
5
10
15
20
25
30


Total night charge
0
100
200
300
400
500
2.5
5.0
7.5
10.0
12.5
15.0
17.5

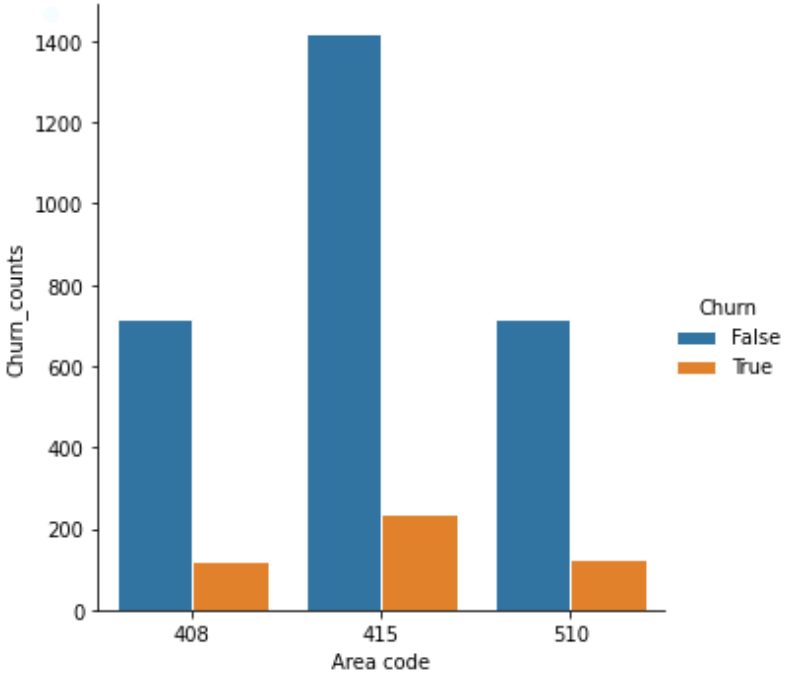
2.Impact of customer service calls on Churn

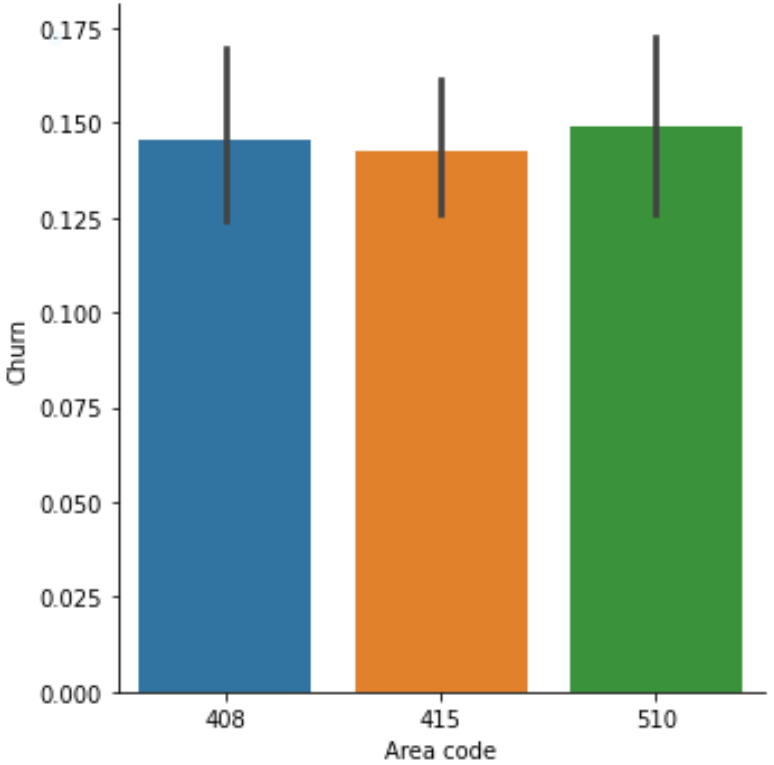




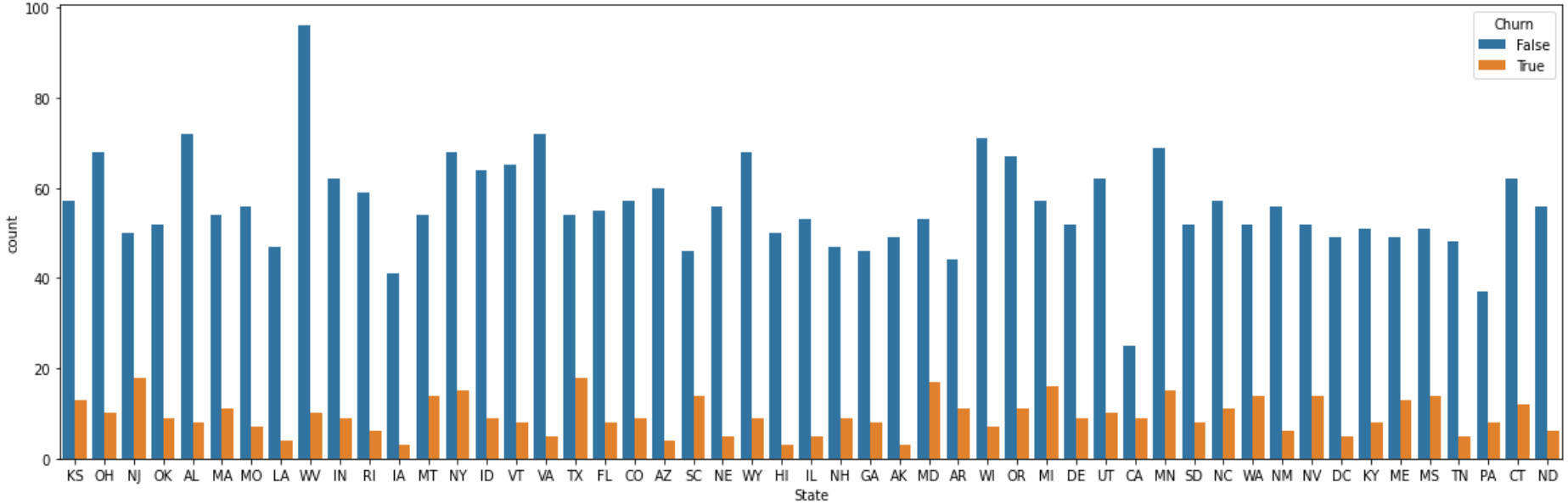
As customer increases customer service calls might they are not satisfying with the solution given by the resolution of the problem so customers frustrate and getting churn in this case while Some customers are lazy and hence without resolving the issue they have jumped to other network operator, while the customers who have called once also have high churn rate indicating their issue was not solved in first attempt

3.Churned and non churned data in specific Areas using their area codes



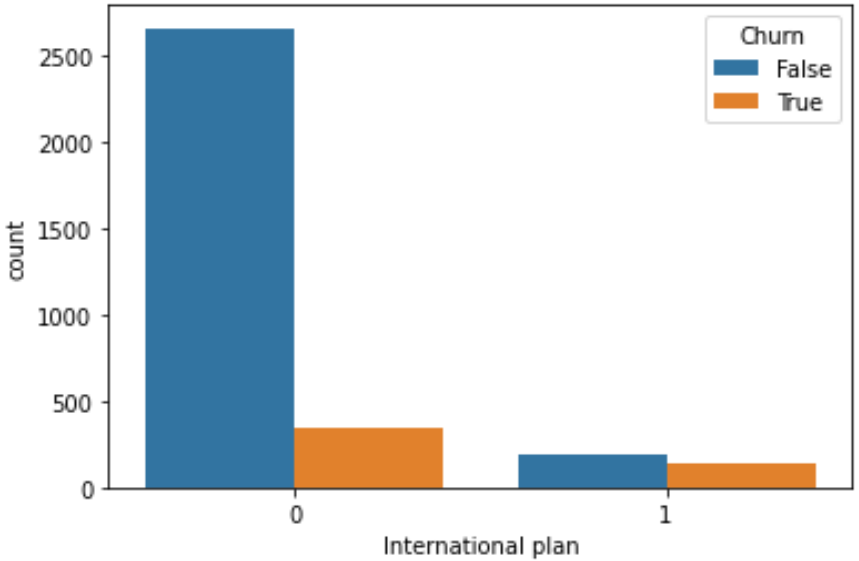


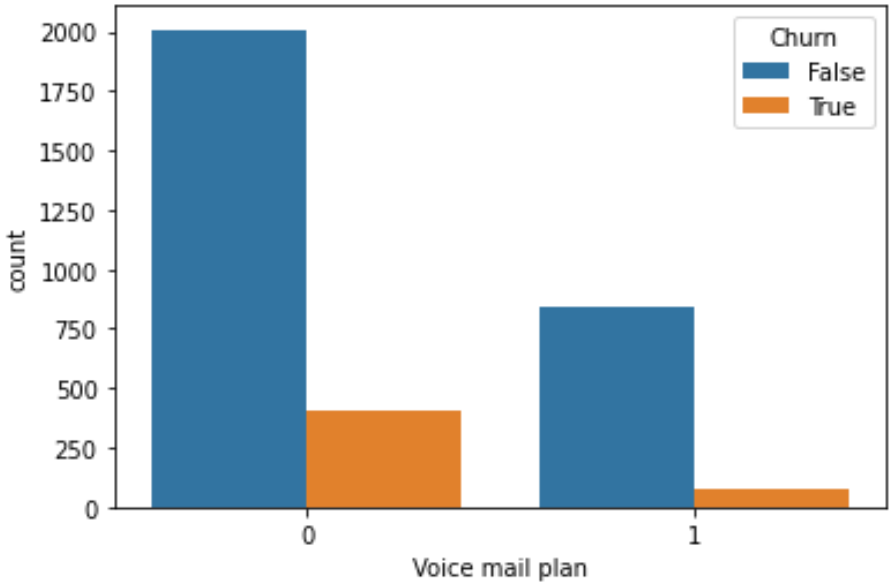
4.In which state have more number of churners



"WV" state has more number non churn(Loyal) customers present and "TX" & "NJ" state has more number of churn customers.

5.To find churn count according to their plans





Conclusion

-Find summary and overview of the dataset.

-No null values are present then the dataset is cleaned.

-From visualization we find out behavior of dataset that includes heatmap, boxplot for find outliers, density plot etc.

-As per day, night and evening density plot we conclude that the data is almost normally distributed.

-As customer increases customer service calls might they are not satisfying with the solution given by the resolution of the problem so customers frustrate and getting churn in this case.

-We conclude that area code 415 have more number of customers and number of churn customers are also more.
-"WV" state has more number non churn(Loyal) customers present and "TX" & "NJ" state has more number of churn customers.
-The customer having International plans are getting churned as compared to the customers having no international plan so it might be due to high charges on international calls so company should try to minimize the international calls charges.
-International plans and voice mail plan gives complex output. Means customers having international plan or voice mail plan both can either churn or non churn according to there convinence.

Thank You.....