Capstone Project
Telecom Churn Data – EDA

Team - stars
1. Sai Krishna Vamshi Devarasetty (krishnavamshidevarasetty@gmail.com)
2. Gangadhar Palle (pallegangadhar156@gmail.com)
3. Abhishek Sharma (abhisheksharmatrio@gmail.com)
4. Nitish Rao – (nitishrao1896@gmail.com)

AI

Telecom Churn Data – Intro

- **Orange S.A., 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.**

Data Description

State – The state the telecom service user belongs to.

Account Length – How long account has been active.

Area Code – Identifier for a geographic region.

International Plan – Did user opted for International Plan ?

Voice mail Plan – Is Voice mail plan activated?

Number vmail messages – Number of voice mail messages sent.

Total day minutes–Total minutes used in day.

Total day calls – Total day calls made.

Total day charge – Total price user was charged for day calls.

Total eve minutes – Total minutes used for calls in evening.

Total eve calls – Total evening calls made.

Total eve charge – Total price user was charged for evening calls.

Total night minutes – Total minutes used for calls in night.

Total night calls – Total night calls made.

Total night charge – Total price user was charged for night calls.

Total intl minutes – Total minutes used for international calls.

Total intl calls – Total calls made for international numbers

Total intl charge – Total price user was charged for international calls

Customer service calls – Number of Customer service calls user made.

Churn – Did user churned the services?

About Data

- There are no missing values in any of the columns.
- Except State, International plan, Voice mail plan and Churn all other columns are numerical(either int64 or float64)
- Even though Area Code is numerical, it is not ordinal.
- State is the only variable which is a string.
- International plan and Voice mail plan are also given as type string(Yes/No) but we need to considered them as Boolean for analysis
- Churn is bool type, True meaning users have churned and False implies who didn't churned
- Out of 3333 users 483 users Churned and 2850 didn't churned

Problem Statements

- Do Churned user speaking more during any specific duration (Day, Evening or night) compared to Non Churned ?

- Does the user calls at particular duration has any impact on his calls of other durations ?

- Do Area has any significant impact on Churned users ?

- Do State has any impact on Churned users ?

- Did opting for International plan resulted in Churning of users ?

- Did opting for Voicemail plan resulted in Churning of Users ?

- Did customers making more service calls churned more than those who didn't ?

Frequency Distribution(Histogram Plot) of Churned vs Non Churned Users based Total Minutes users used during day, evening and night
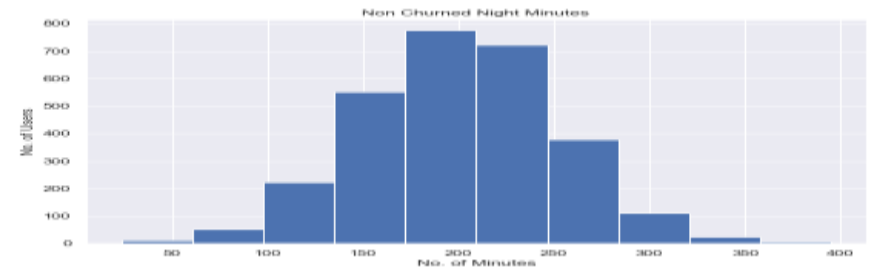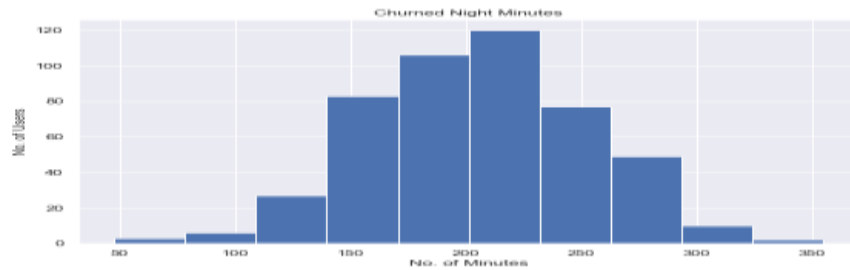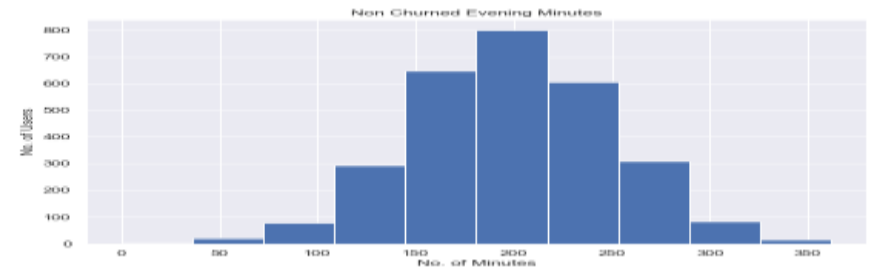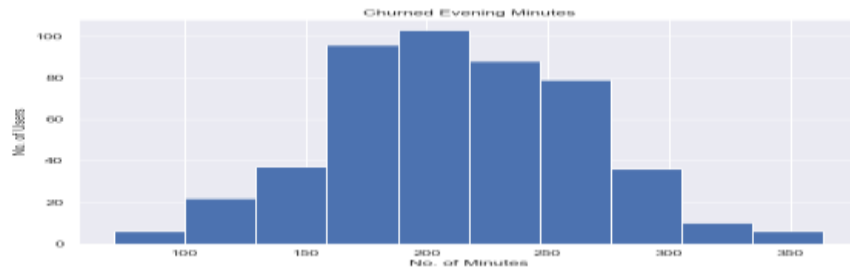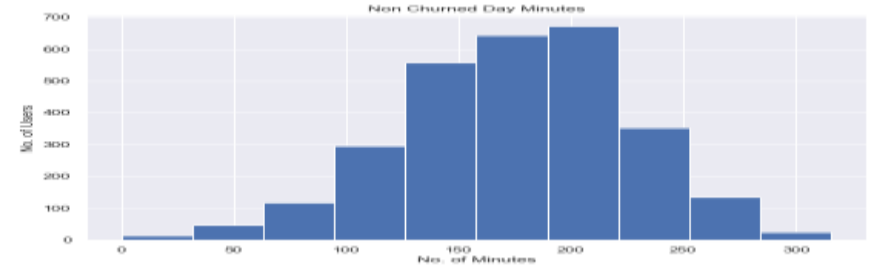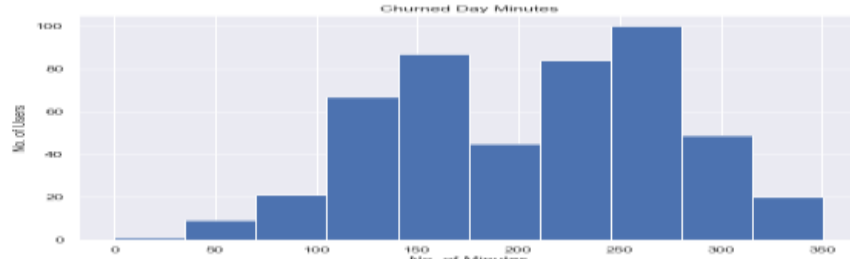
Frequency Distribution(Histogram Plot) of Churned vs Non Churned Users based Total calls users made used during day, evening and night
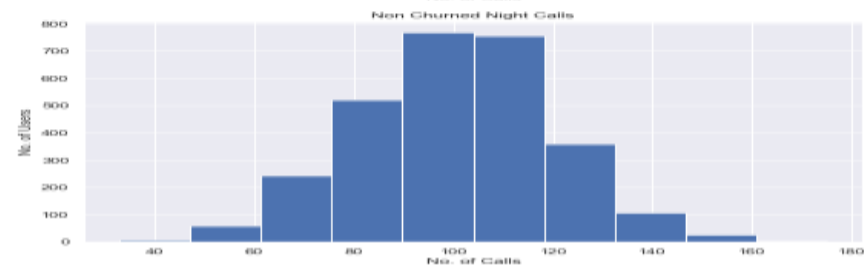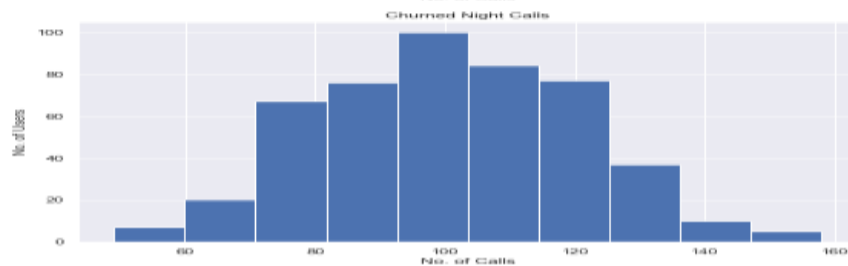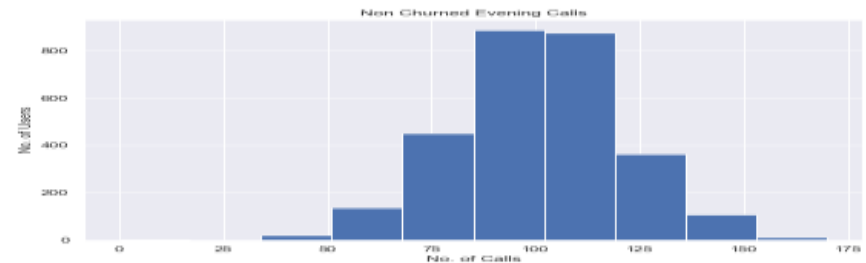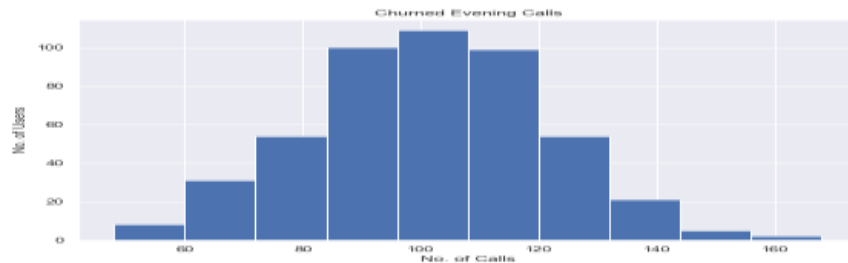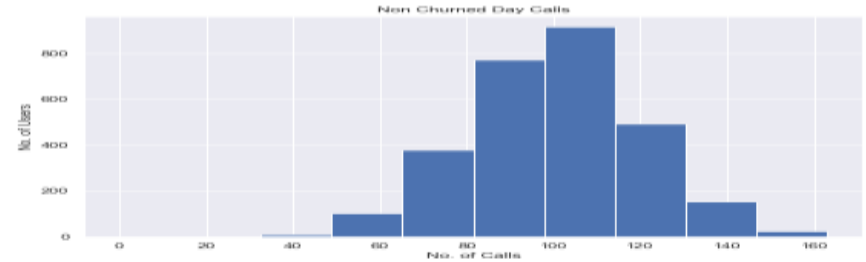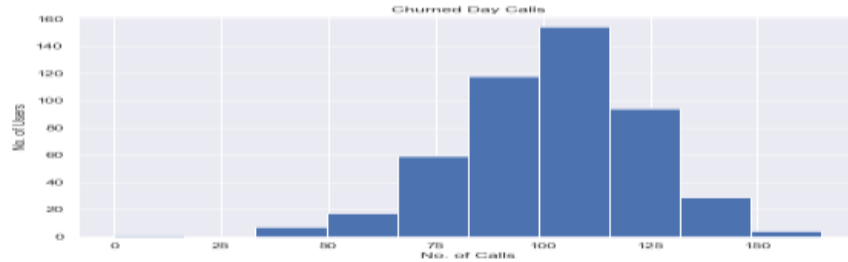
Frequency Distribution(Histogram Plot) of Churned vs Non Churned Users based Total price users were charged used during day, evening and night
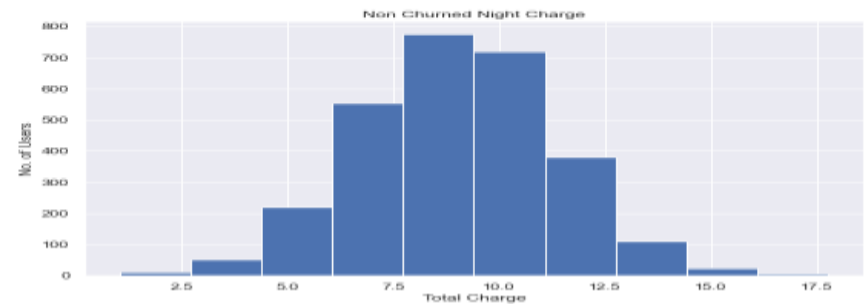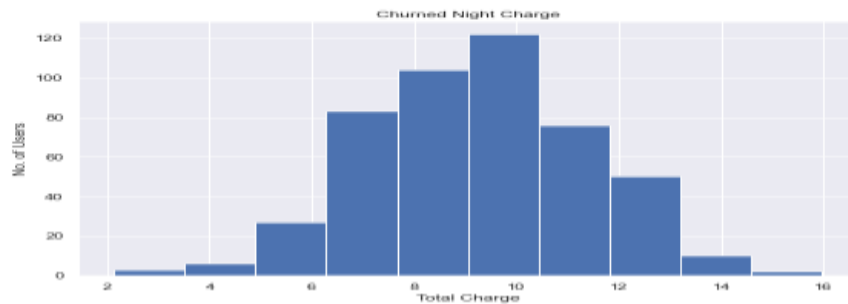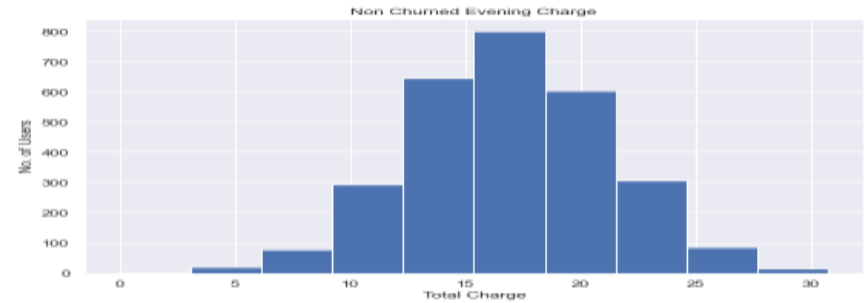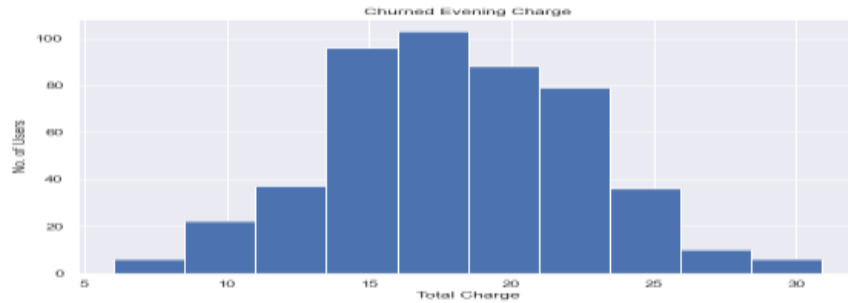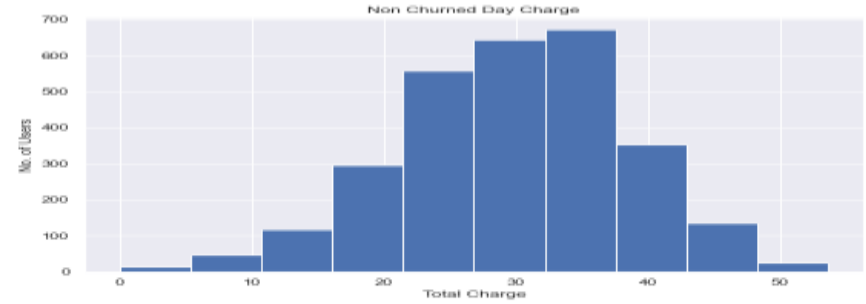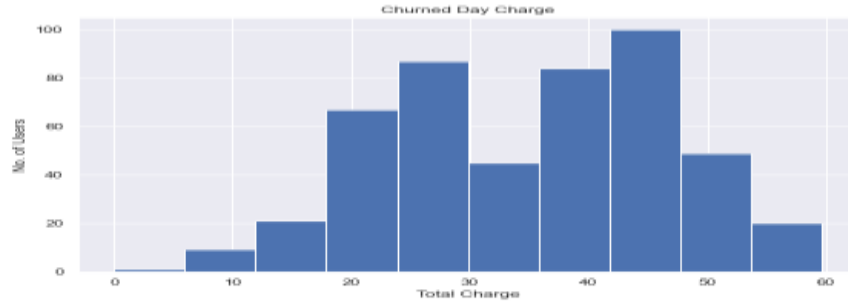
# Correlation among Day, Evening and Night Times for Minutes, Calls and Charges

	Total day minutes	Total eve minutes	Total night minutes
Total day minutes	1.000000	0.007043	0.004323
Total eve minutes	0.007043	1.000000	-0.012584
Total night minutes	0.004323	-0.012584	1.000000
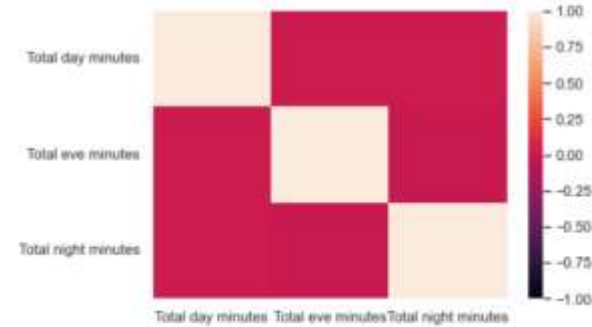
	Total day calls	Total eve calls	Total night calls
Total day calls	1.000000	0.006462	-0.019557
Total eve calls	0.006462	1.000000	0.007710
Total night calls	-0.019557	0.007710	1.000000
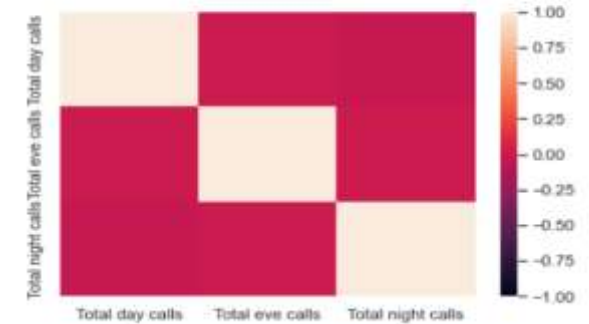
|	Total day charge	Total eve charge	Total night charge
Total day charge	1.000000	0.007036	0.004301
Total eve charge	0.007036	1.000000	-0.012601
Total night charge	0.004301	-0.012601	1.000000
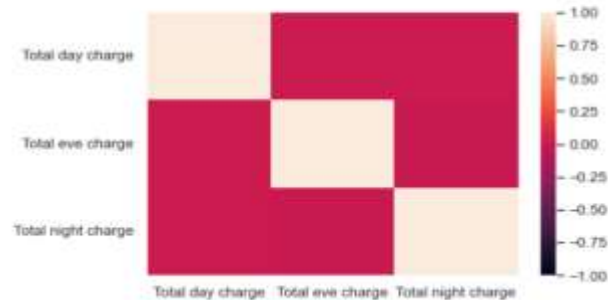
Area Code Based Analysis

We figured out from Data that there are total 3 Area Codes(415,408,510) to which users belongs to. To make analysis easier we created a nested python dictionary of the following type:

```
AreaCodeWiseChurnAndNonChurn = {
    'Churn': {
        415: DataFrame of 415 Area Code with Churn equal to True,
        408: DataFrame of 408 Area Code with Churn equal to True,
        510: DataFrame of 510 Area Code with Churn equal to True
    },
    'NonChurn': {
        415: DataFrame of 415 Area Code with Churn equal to False,
        408: DataFrame of 408 Area Code with Churn equal to False,
        510: DataFrame of 510 Area Code with Churn equal to False
    }
}
```
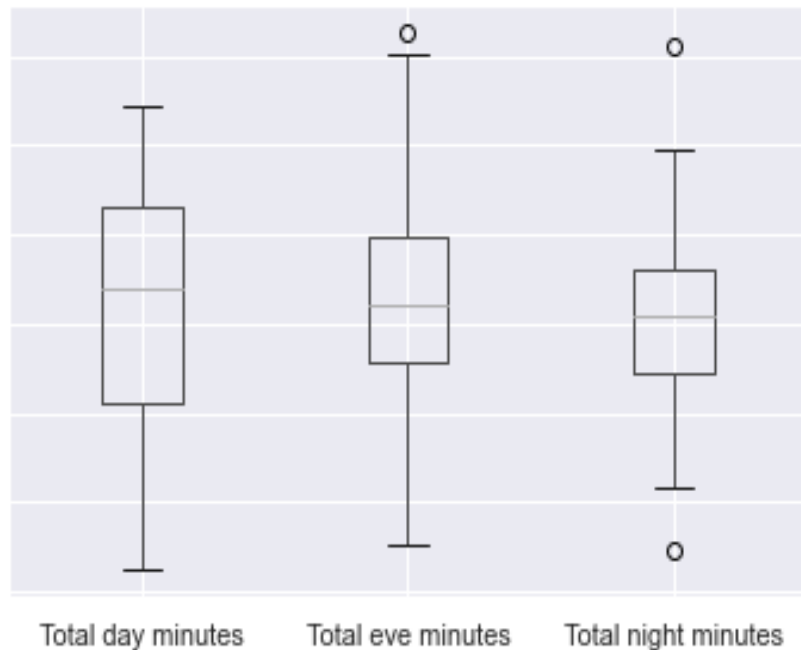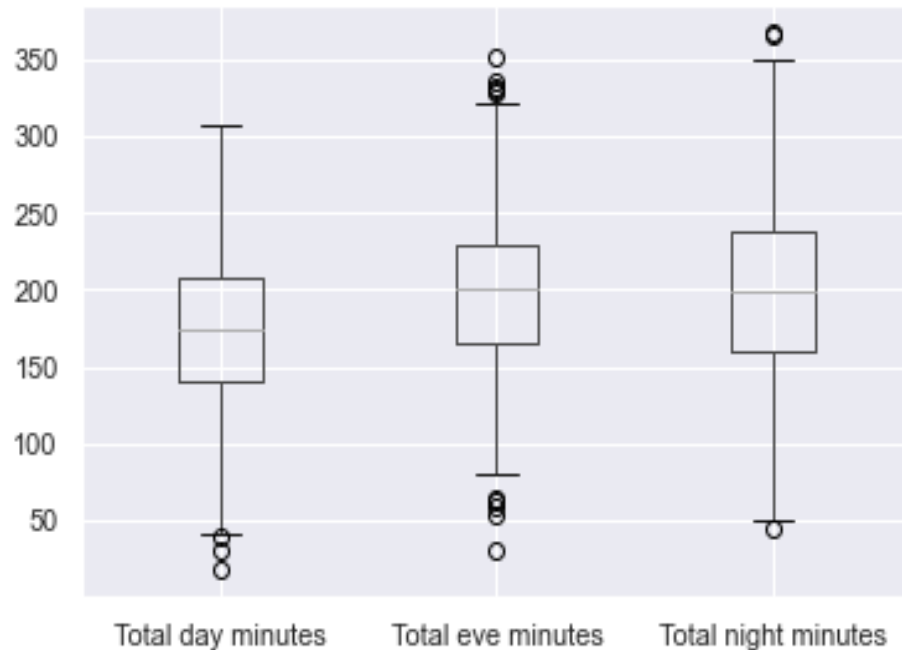
Area Code - 415
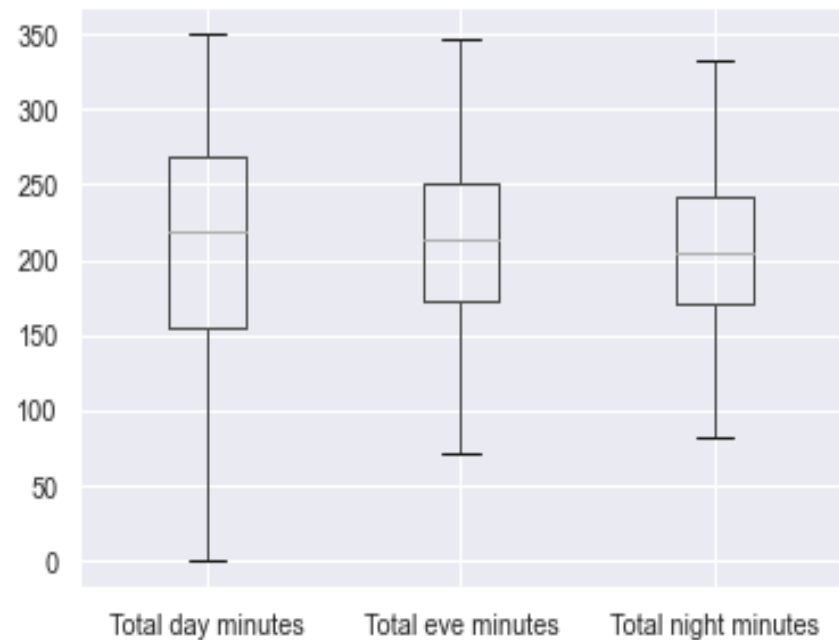
Churned Users
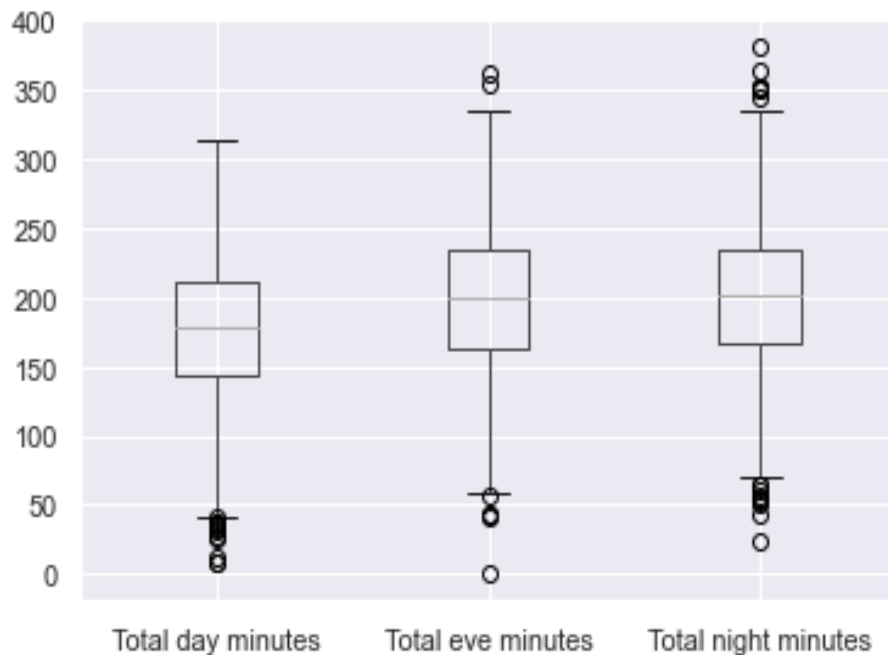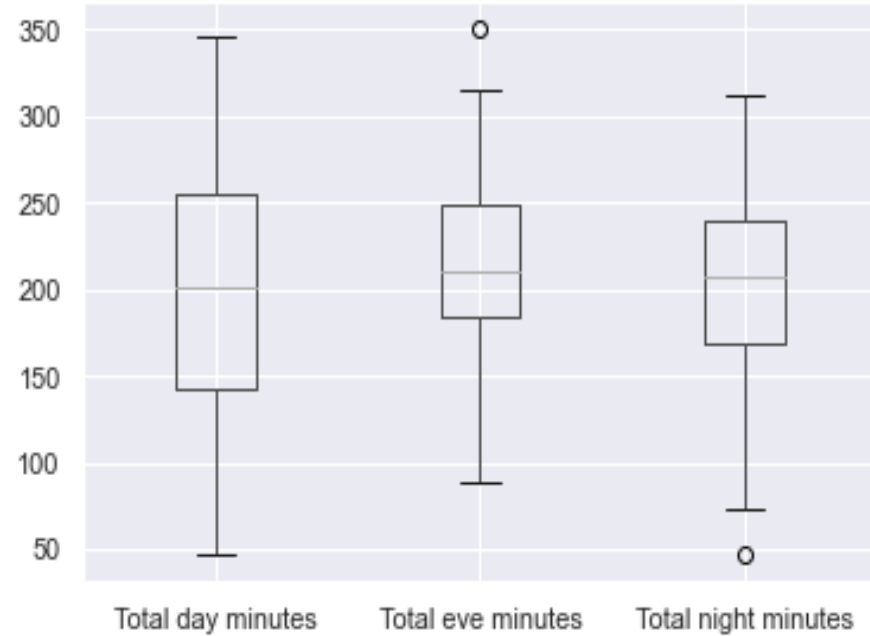
Non Churned Users
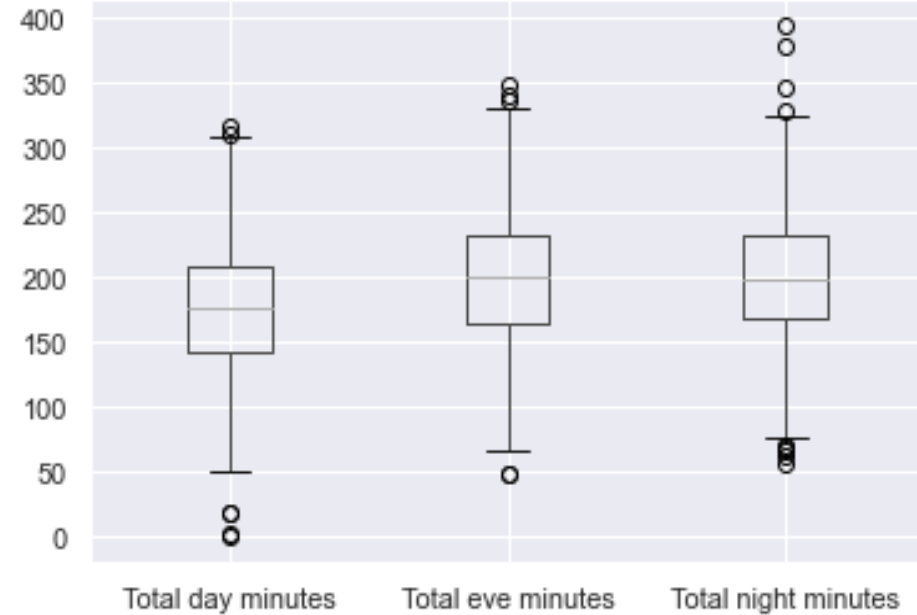
Area Code - 408

Churned Users

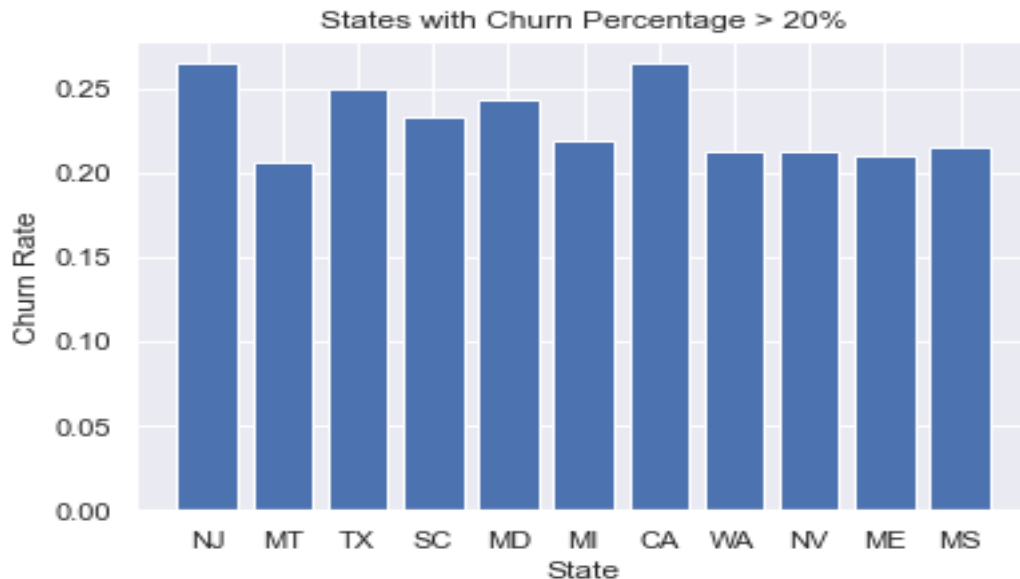Non Churned Users

Area Code - 510

Churned Users

Non Churned Users

State Based Analysis

There are 51 states in total in the data set. Out of which 11 states have more than 20% Churned Rate.

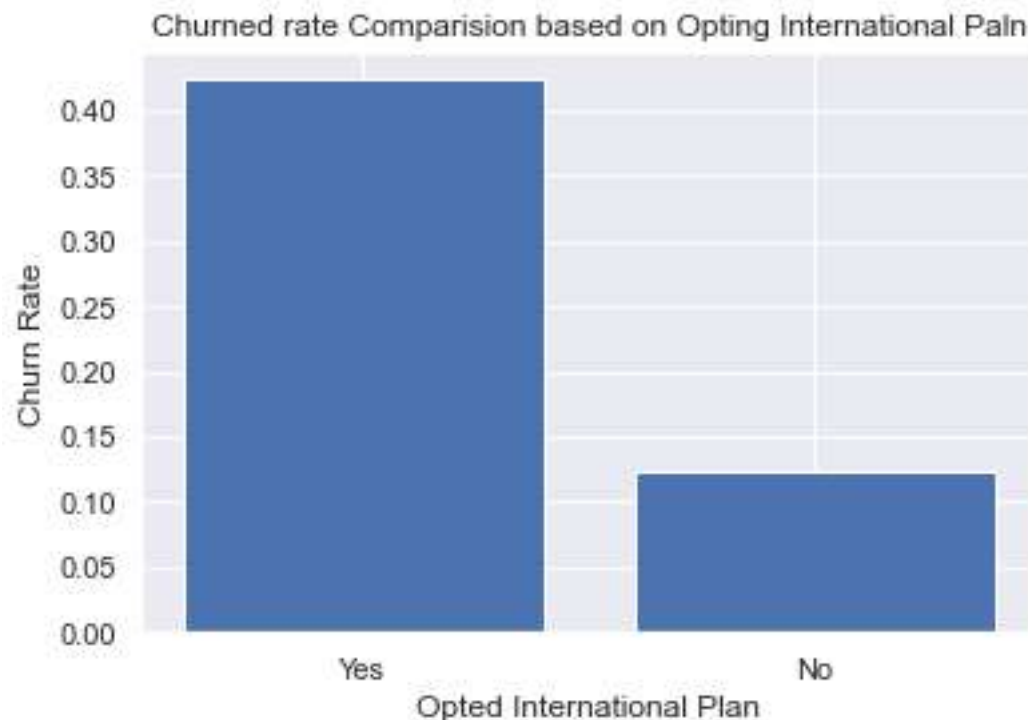States with Churn Percentage > 20%

Out of these 11 states, NJ and CA have highest churn rate of around 26.47%

International Plan Based Analysis
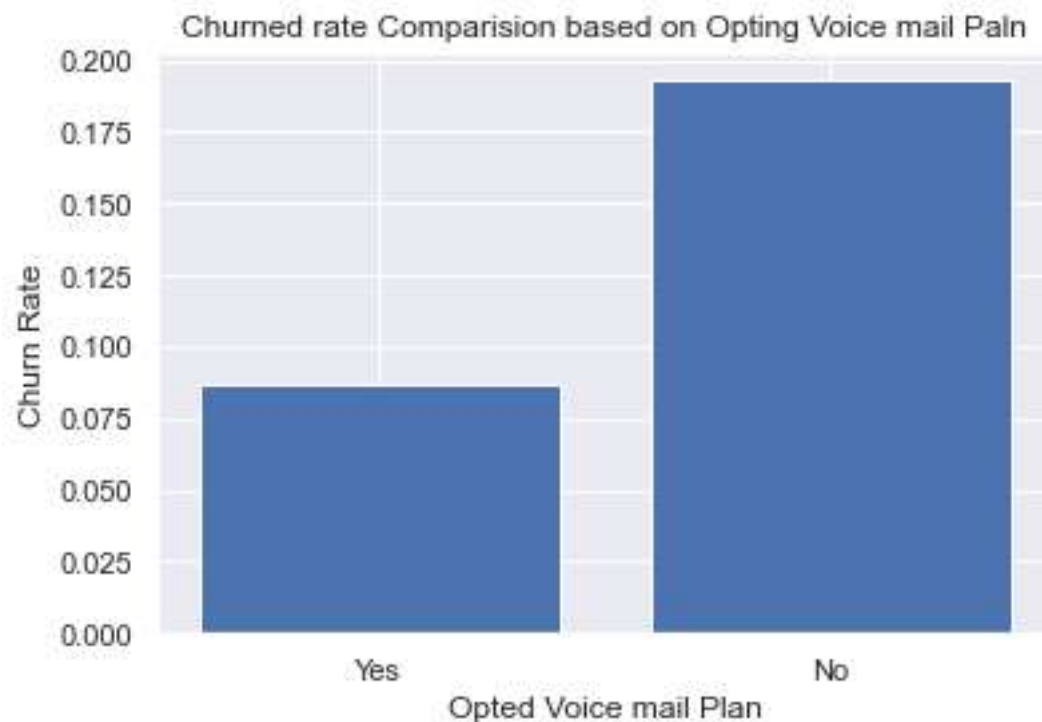


```
International plan  Churn
No                 False    2664
                   True      346
Yes                False     186
                   True      137
Name: Churn, dtype: int64
```

Churned rate Comparision based on Opting International Paln

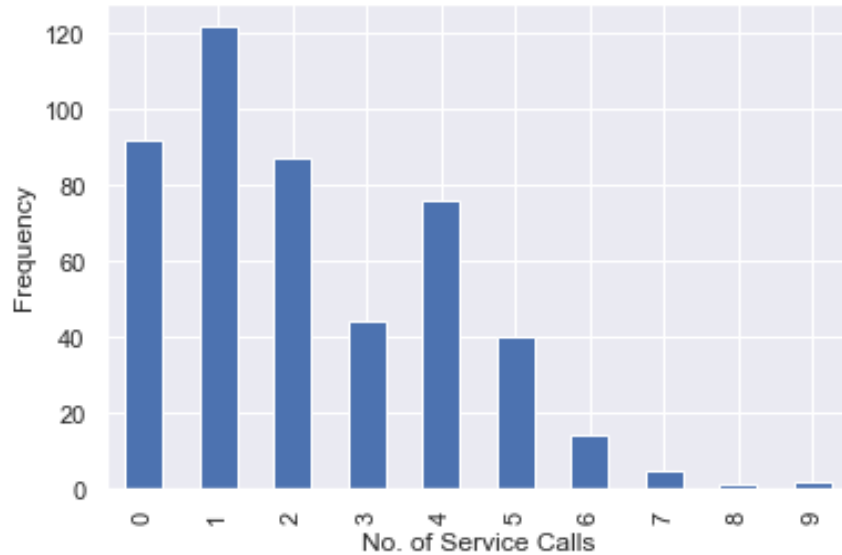# Voice mail Plan Based Analysis
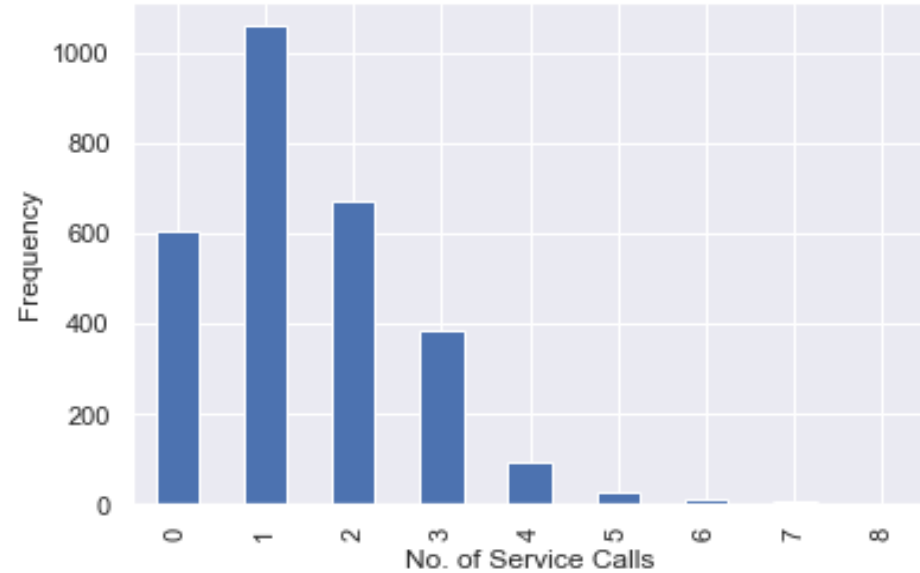
```
Voice mail plan  Churn
No               False    2008
                 True      403
Yes              False     842
                 True       80
Name: Churn, dtype: int64
```


Churned rate Comparision based on Opting Voice mail Paln

Analysis Based on Number of Customer Service Calls

Churned Users

Non Churned Users

Observations

- Users with International Plan tend to Churn more frequently.
- 11 States out of 51 States have more than 20% Churn rate indicating the Telecom service is not up to expectations in those states.
- Uses with Voice mail plan have very low Churn rate than others which implies that Voice mail plan of Telecom service is good.
- There's almost no(near to zero) correlation among day, evening and night time.
- Users with 4 or more customer service calls Churned more than 4 times as often as other users

References

- Numpy, Pandas, Matplotlib and Seaborn documentation.
- Alma Better Recorded Classes.
- Articles on Towards Data Science

THANK YOU