Capstone Project

Project Title :- Telecom Churn Analysis

Team Members

Pratik Kshirsagar
Ganesh More

Index

- Problem Statement
- Data Summary
- Data Pipeline
- Null value Imputation/ Data Cleaning
- EDA
- Conclusion

Problem Statement

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.

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

Summary of Dataset

The dataset contains 20 features with more than 3000 observations.

Important Features:

STATE: *51 Unique States in United States of America*

Account Length: *Length of The Account*

Area Code: *415 relates to San Francisco,408 is of San Jose and 510 is of City of Okland*

International Plan: *Yes Indicate International Plan is Present and No Indicates no subscription for Internatinal Plan*

Voice Mail Plan: *Yes Indicates Voice Mail Plan is Present and No Indicates no subscription for Voice Mail Plan*

Number vmail messages: *Number of Voice Mail Messages ranging from 0 to 50*

Summary of Dataset

Total day minutes: *Total Number of Minutes Spent By Customers in Morning*

Total day calls: *Total Number of Calls made by Customer in Morning.*

Total day charge: *Total Charge to the Customers in Morning.*

Total eve minutes: *Total Number of Minutes Spent By Customers in Evening*

Total eve calls: *Total Number of Calls made by Customer in Evening.*

Total eve charge: *Total Charge to the Customers in Evening.*

Summary of Dataset

Total night minutes: *Total Number of Minutes Spent By Customers in the Night.*

Total night calls: *Total Number of Calls made by Customer in Night.*

Total night charge: *Total Charge to the Customers in Night*

Total intl minutes: *Total Number of Minutes Spent By Customers Internationally.*

Total intl calls: *Total Number of Calls made by Customer Internationally.*

Total intl charge: *Total Charge to the Customers on International calls.*

Customer service calls: *Total number of Customer service calls.*

Churn: *True means customer churned and False means customer not churned.*

Data Summary

Churn users:- The user which initially connect with the network but after a gap they switch there network, many users do this switch from one plan to other or one network to other network

Purpose:- Purpose of the churn is that the users wants the better features from the network ,when they think the current network doesn't gives better deal for them and the other network have better plan according to there plan the churn there network, many users have want there bucket plan and the network who fills there need the user prefer that network

Data Pipeline

Data processing 1:-in this first part we have check the given data so that there is no null column, and remove unnecessary features.

Data processing 2:- in this part we have check the data and manually go though each features selected from part one and make a better view on the data.

EDA:- this in part we do some exploratory data Analysis on the features selected from part one and two.

Data Visualization :- finally, last but not the last part, we create model to explain that what actually happens with the data and by the help of different model, we cleared all the point that why the given thing is happens, with the help of better performance we gives a exact result.

Null value Imputation/ Data Cleaning

```
print(df.isnull().sum()) # It gives count of null values from each columns of a DataFrame
```

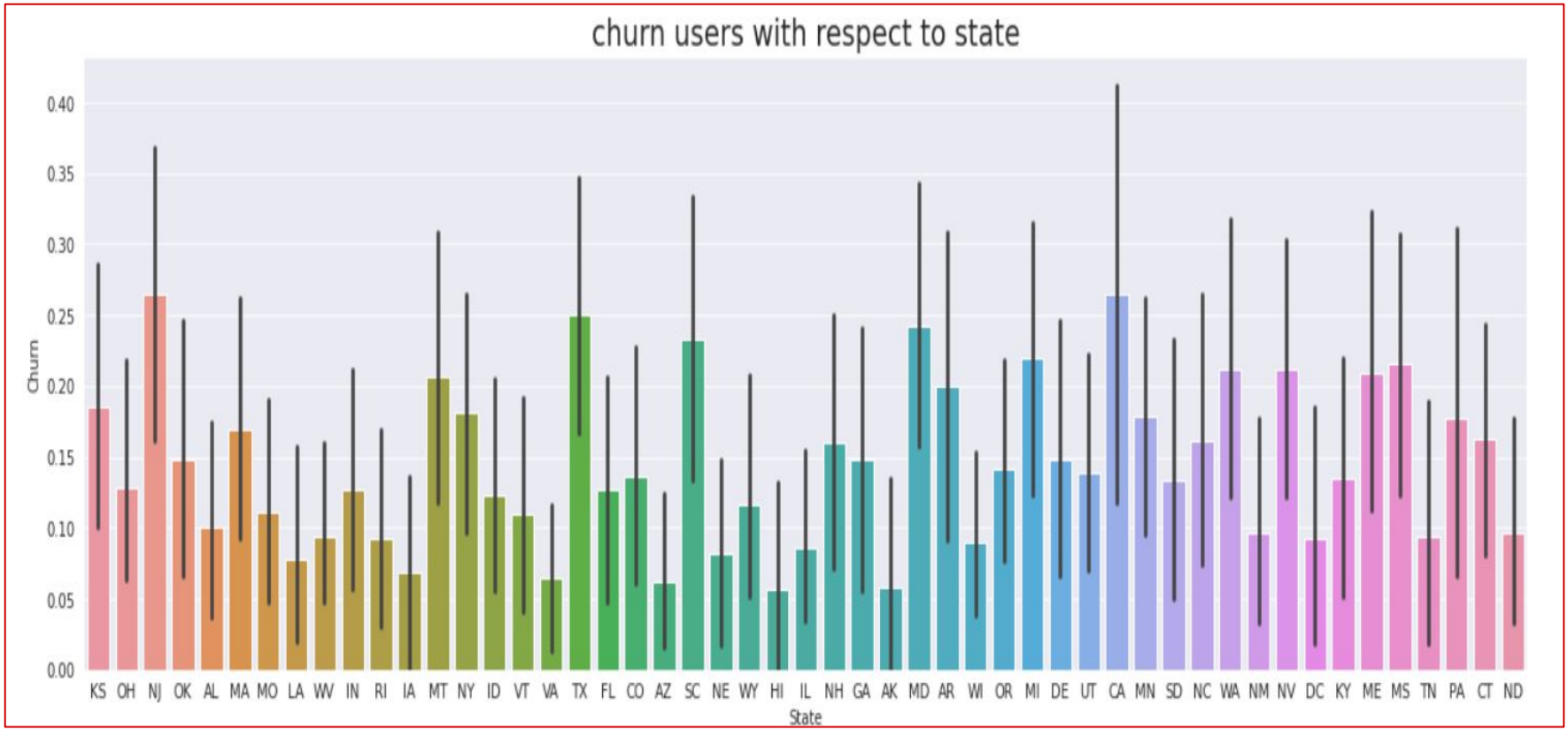
```
State                     0
Account length            0
Area code                 0
International plan        0
Voice mail plan           0
Number vmail messages     0
Total day minutes         0
Total day calls           0
Total day charge          0
Total eve minutes         0
Total eve calls           0
Total eve charge          0
Total night minutes       0
Total night calls         0
Total night charge        0
Total intl minutes        0
Total intl calls          0
Total intl charge         0
Customer service calls    0
Churn                     0
dtype: int64
```

```
[89] print("Missing values:", df.isnull().sum().values.sum())
```

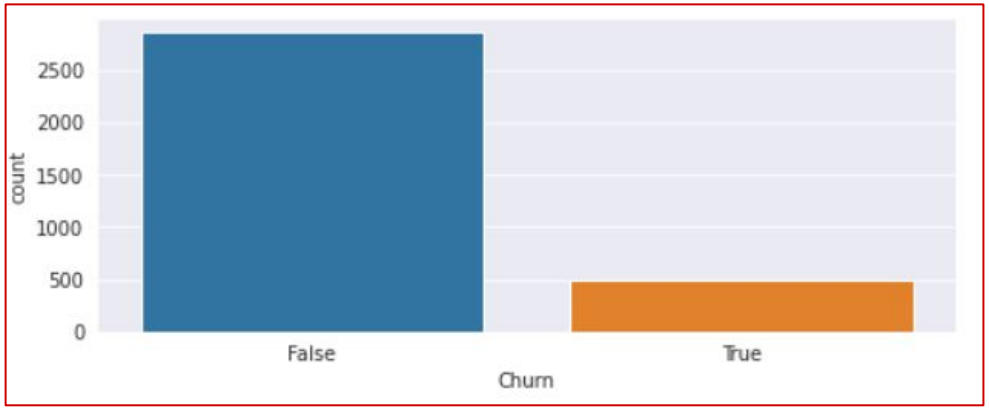
```
Missing values: 0
```

There id no null/NaN/missing values present in the Dataset

EDA (churn users with respect to state)



EDA(churn users count)

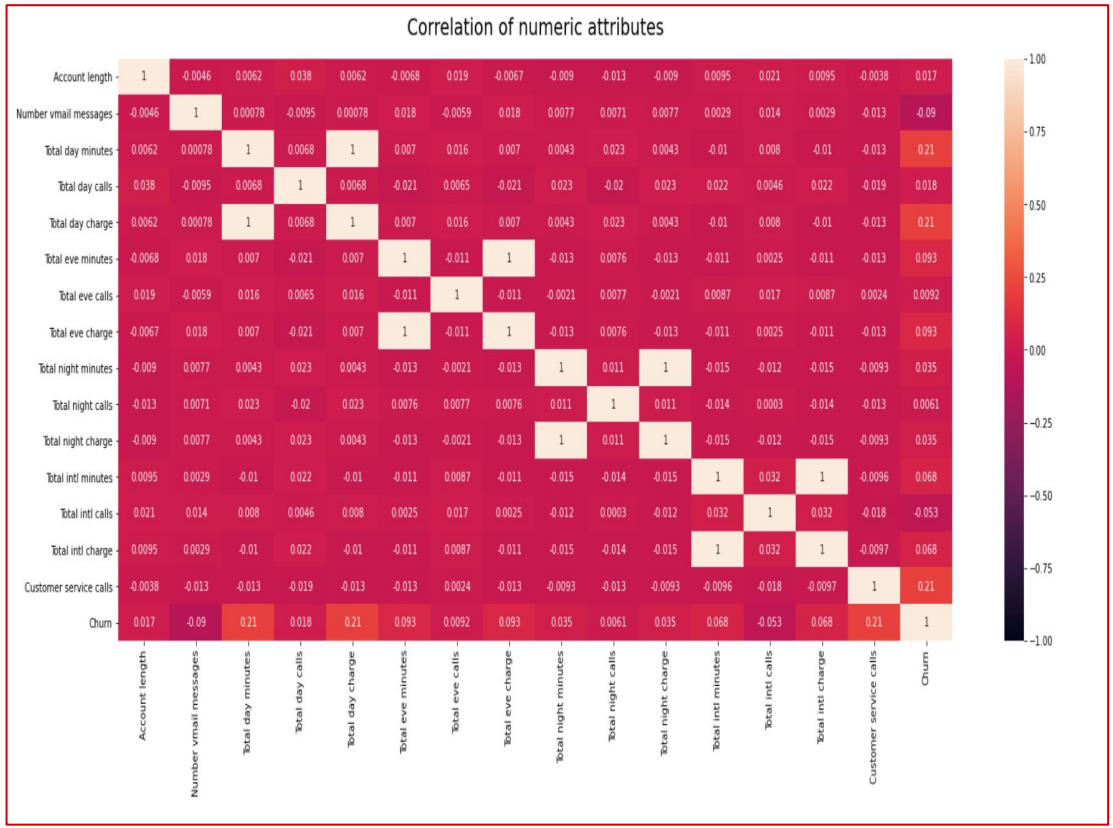



```
False    2850
True      483
Name: Churn, dtype: int64
False    85.508551
True     14.491449
Name: Churn, dtype: float64
```

%
%

2850 (85.50%)users out of 3333 are loyal; their Churn value is 0. To calculate fractions, pass normalize=True to the value counts function.

EDA(Correlation of numeric attributes)

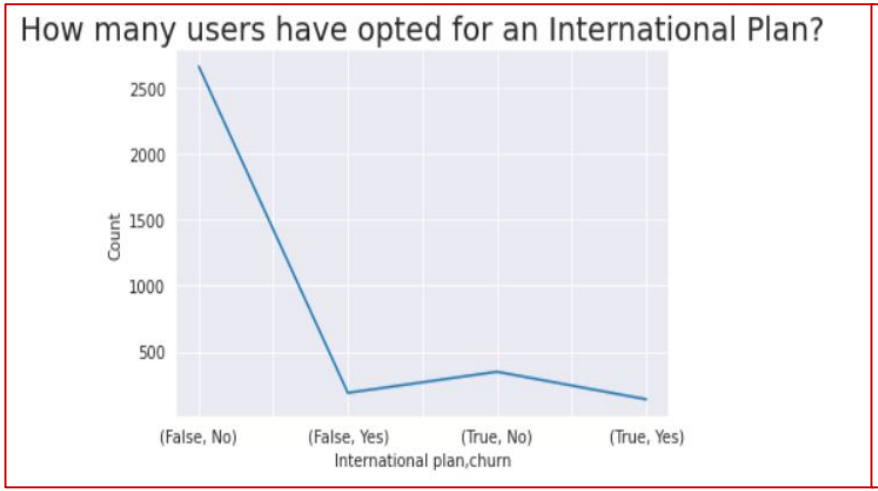


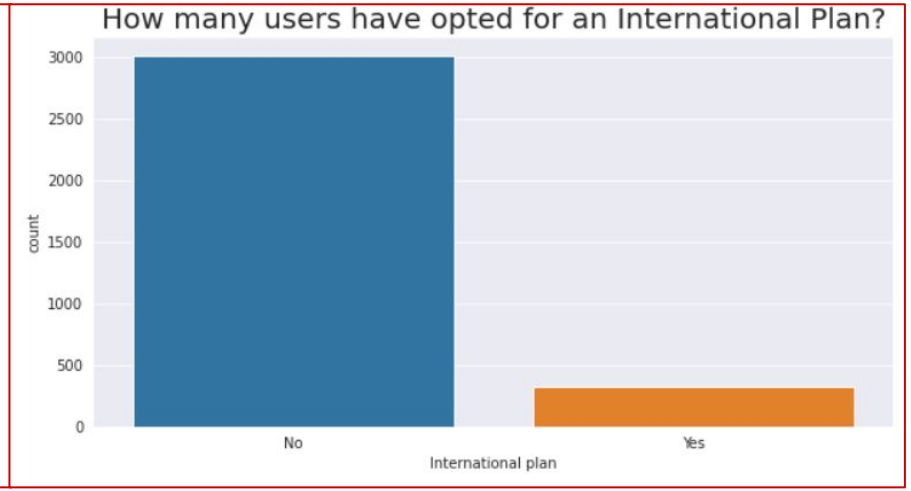
With regard to the heatmap we can see high correlation between the following

variables: With respect to churn

- Total day minutes to Churn at 0.2
- Total day charge to Churn at 0.2
- Customer service calls to Churn at 0.2
- Other relations: Total day charge to Total day minutes at 1
- Total evening charge to total eve minutes at 1
- total night charge to total night minute at 1

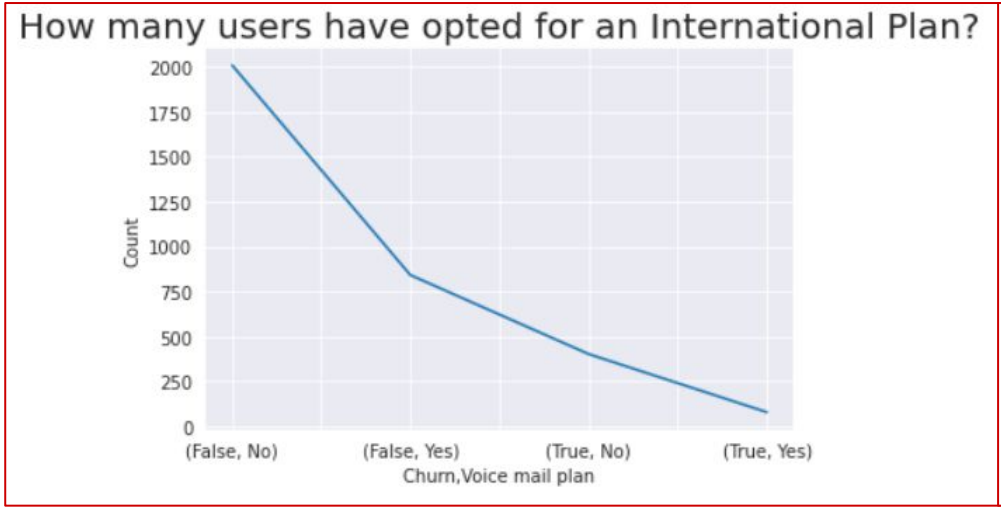
EDA(How many users have opted for an International Plan?)

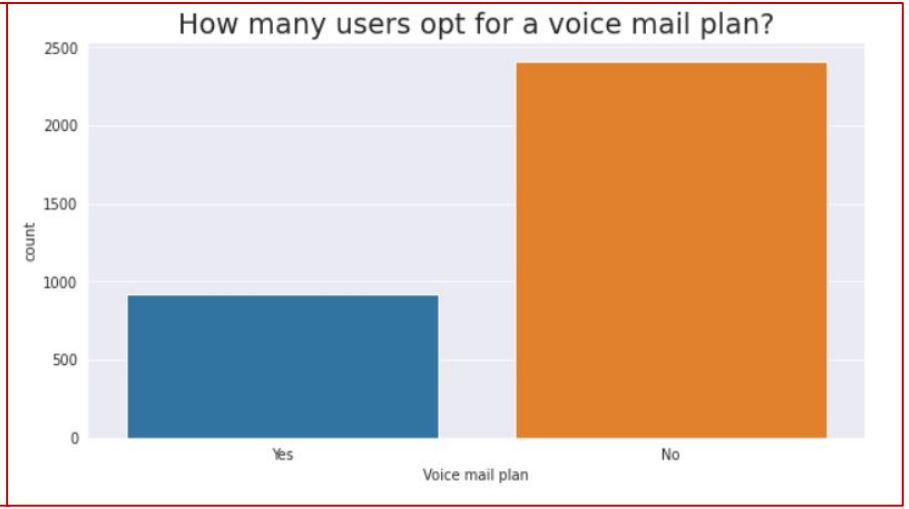




We can see that not many users opt for International plans.

EDA(How many users opt for a voice mail plan?)

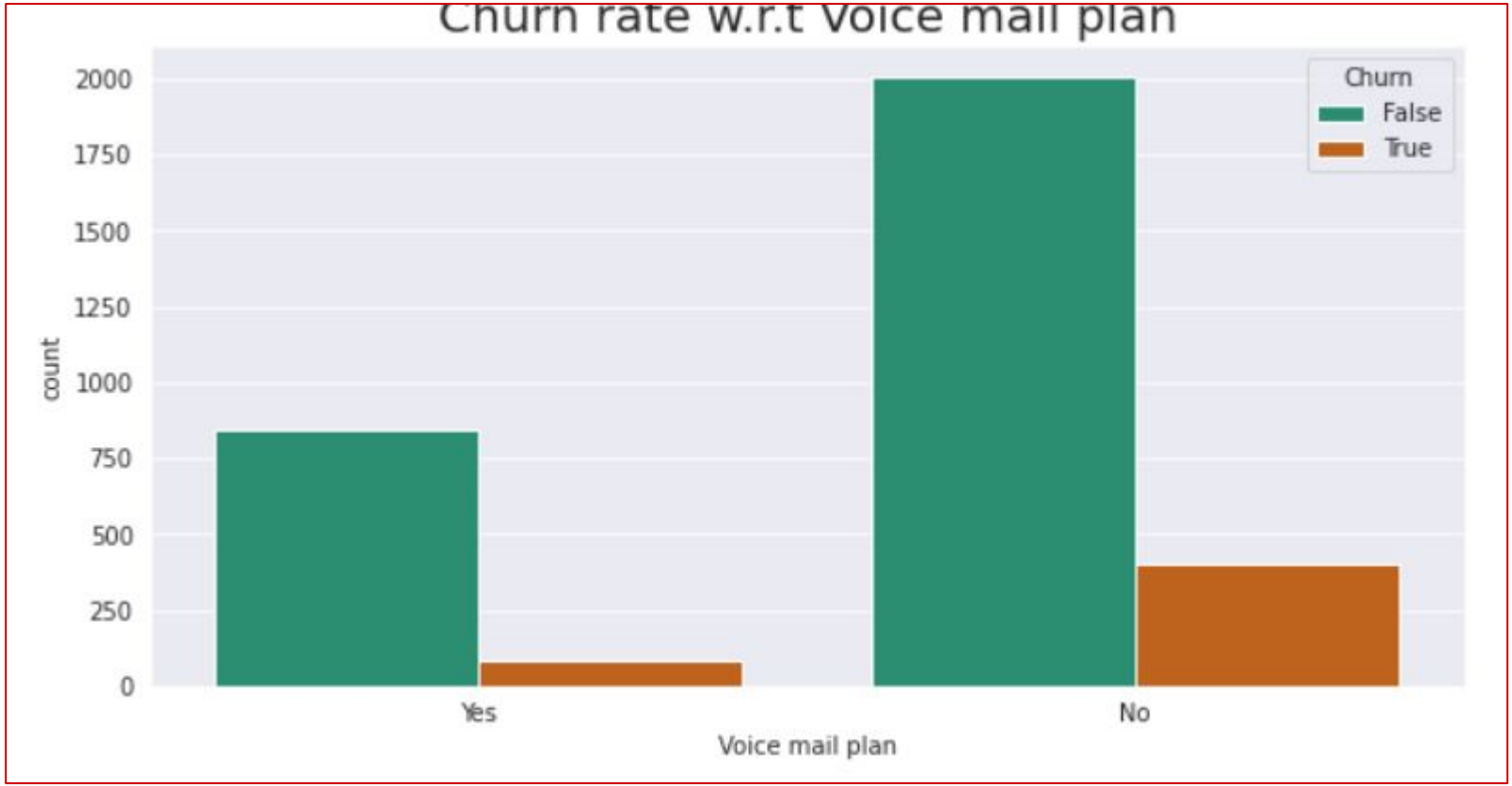




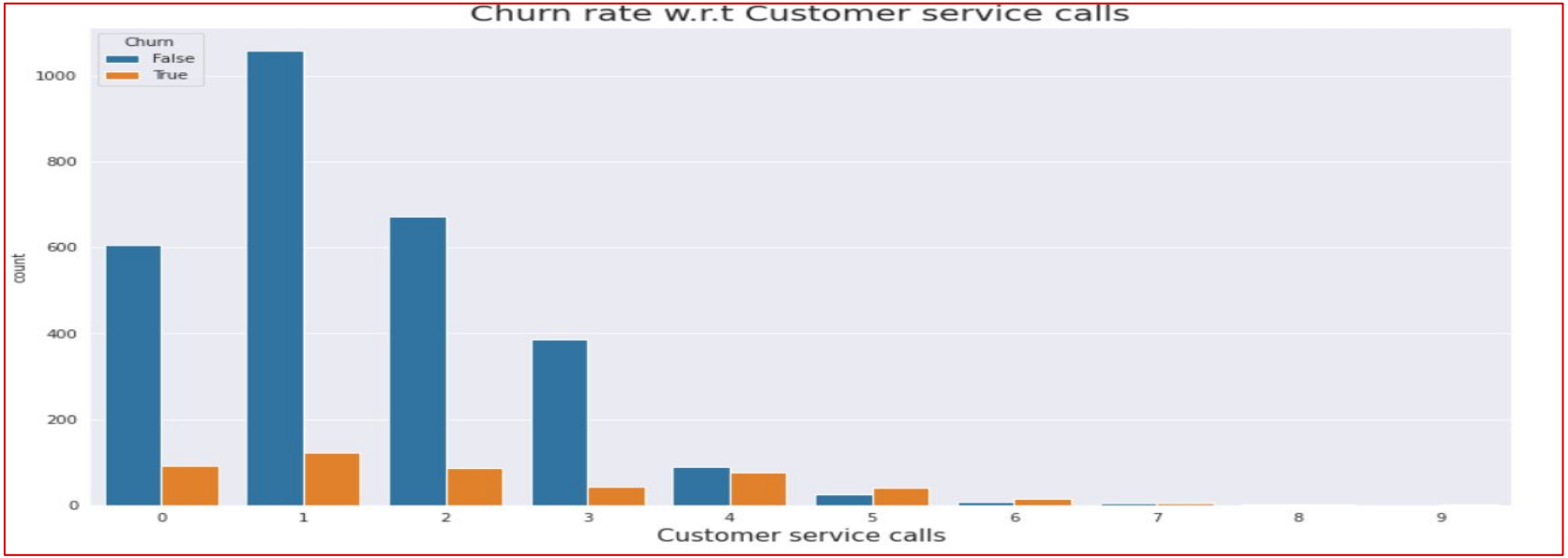
we can see that Many users do not opt for a voice mail plan

EDA(Churn rate w.r.t International Plan)

EDA(Churn rate w.r.t Voice mail plan)

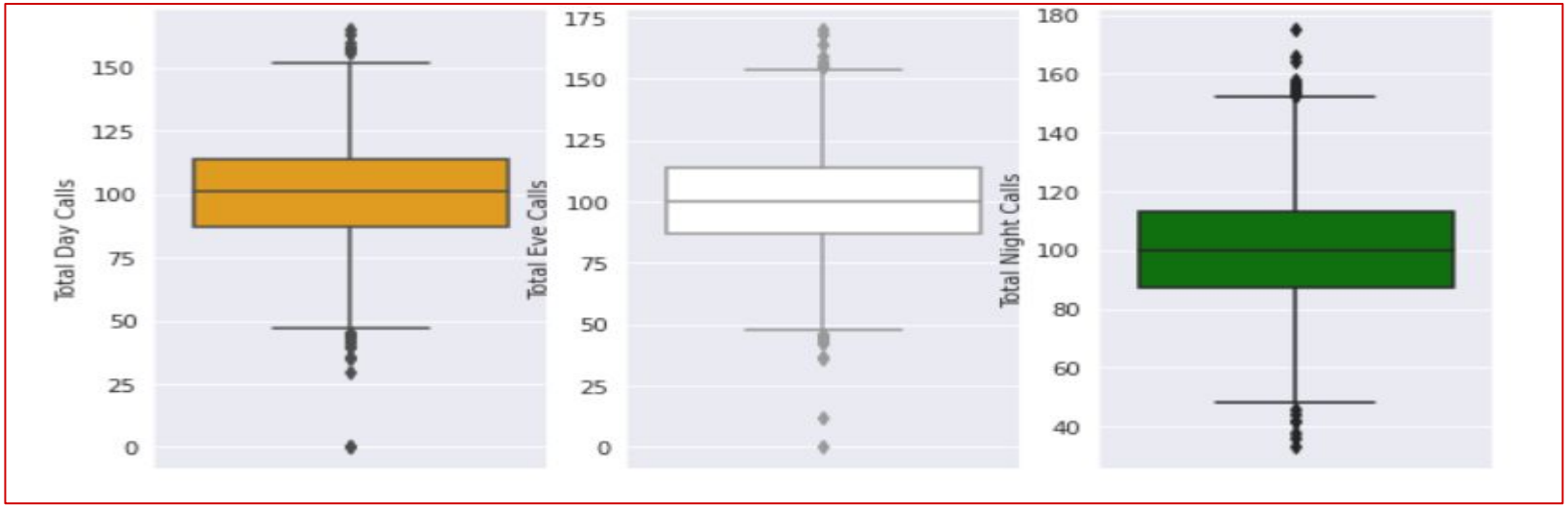


EDA(Churn rate w.r.t Customer service calls)



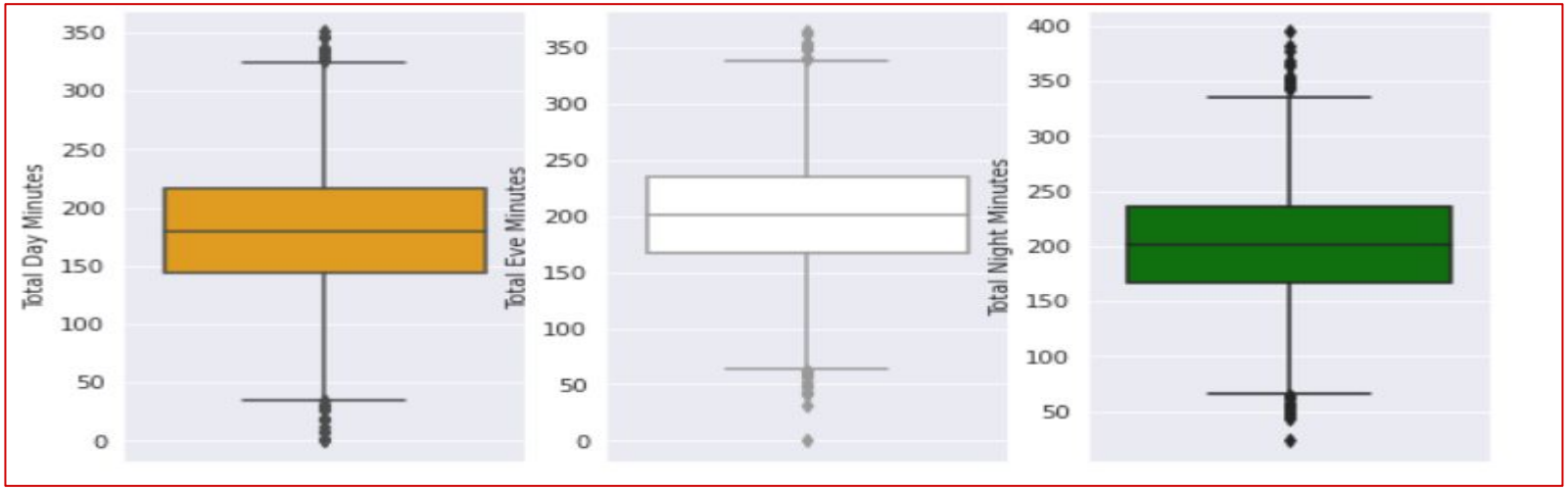
- We can see that if customer service calls are made more than 7 times, then the service is bound to be cancelled. This also comes with high charges imposed on the user.
- At the same time we can see that many users leave the service over 1 to 3 calls made to the customer service when are charged more. Analysis of the total day charge to the minutes spoken over churn rate could give us a better understanding of the same.

EDA(total calls on day, eve, night)



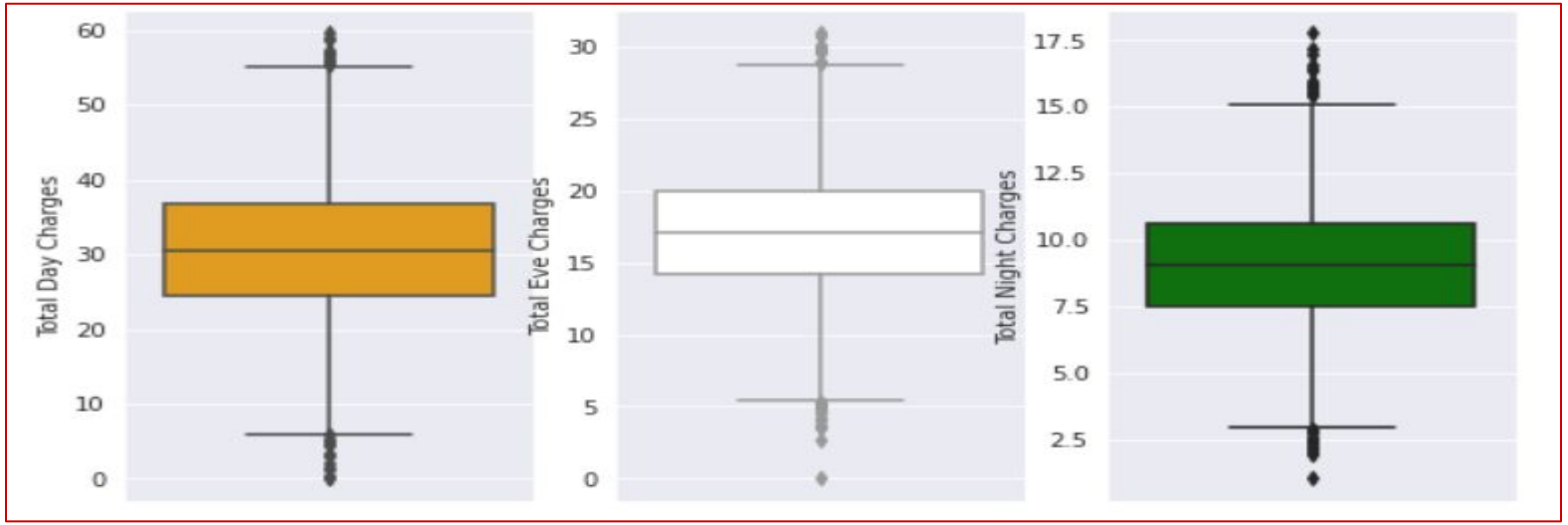
We can see that more calls are made in the morning with respect to the evening and night when it is the lowest

EDA(total Minutes on day, eve, night)



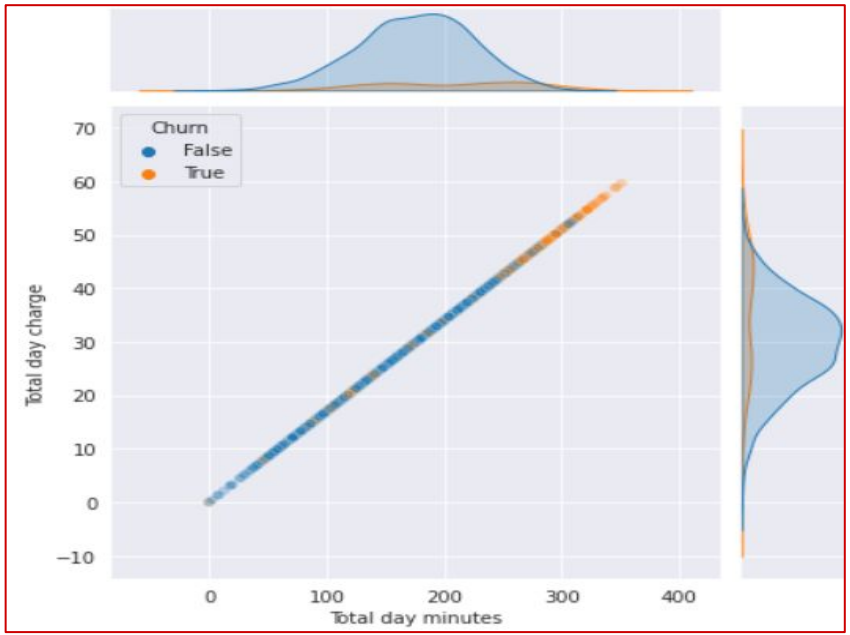
Despite having more calls made during the day, we can see that users do not spend longer time on the calls in the morning. Users tend to talk for longer time in the afternoon than other times.

EDA(total Minutes on day, eve, night)



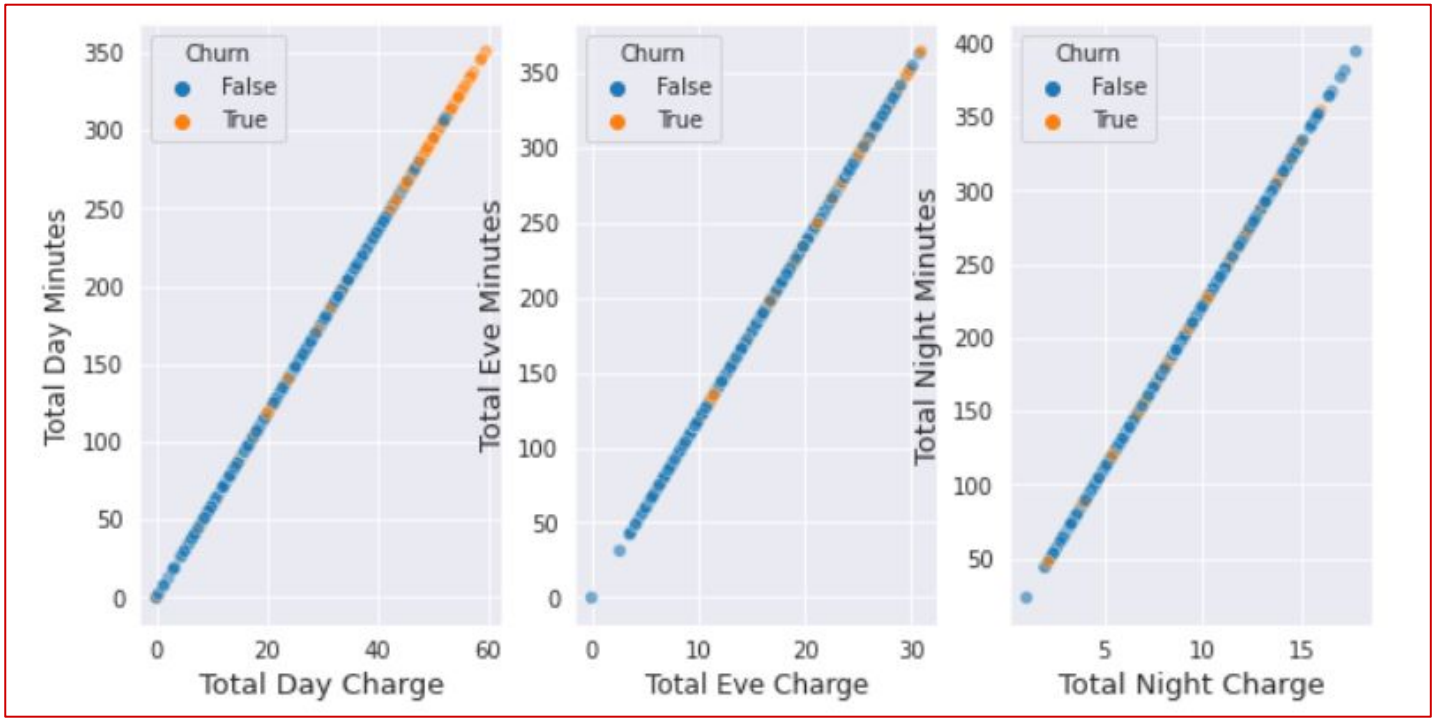
The charges are maximum in the evening time and lowest in the night time. These plots are in conjunction with the minutes spoken.

Bi - Variate Distribution



we used here jointplot from the seaborn library to show relation between Total day minutes and Total day charges

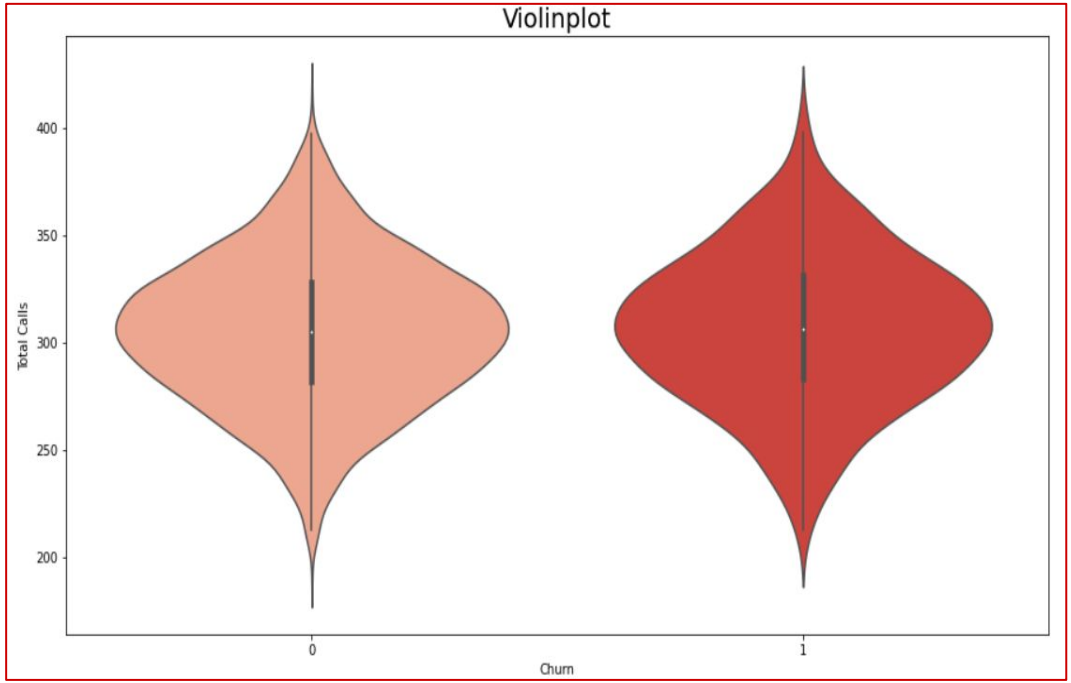
Bi - Variate Distribution (Continued)



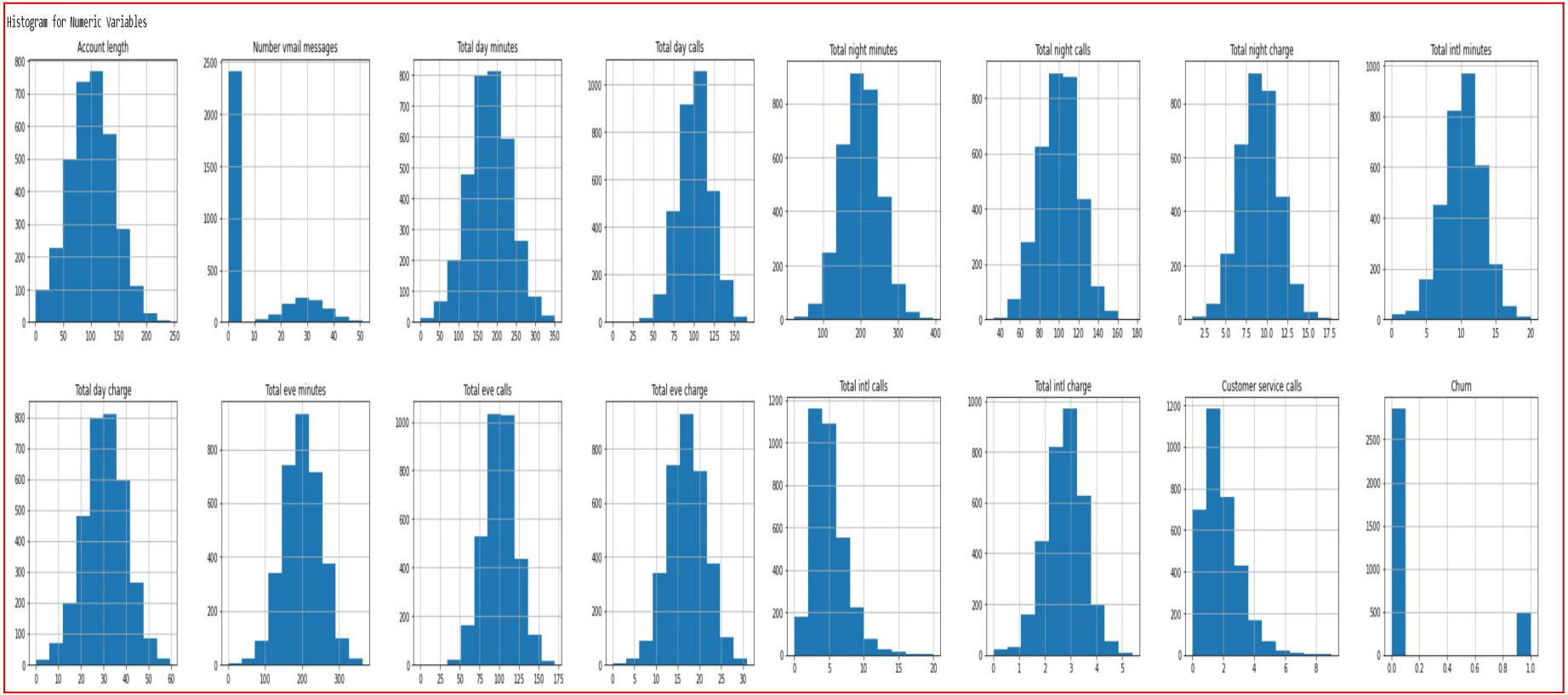
we used here scatterplot from the seaborn library to show relation between Total day minutes and Total day charges

EDA (Total calls, Churn)

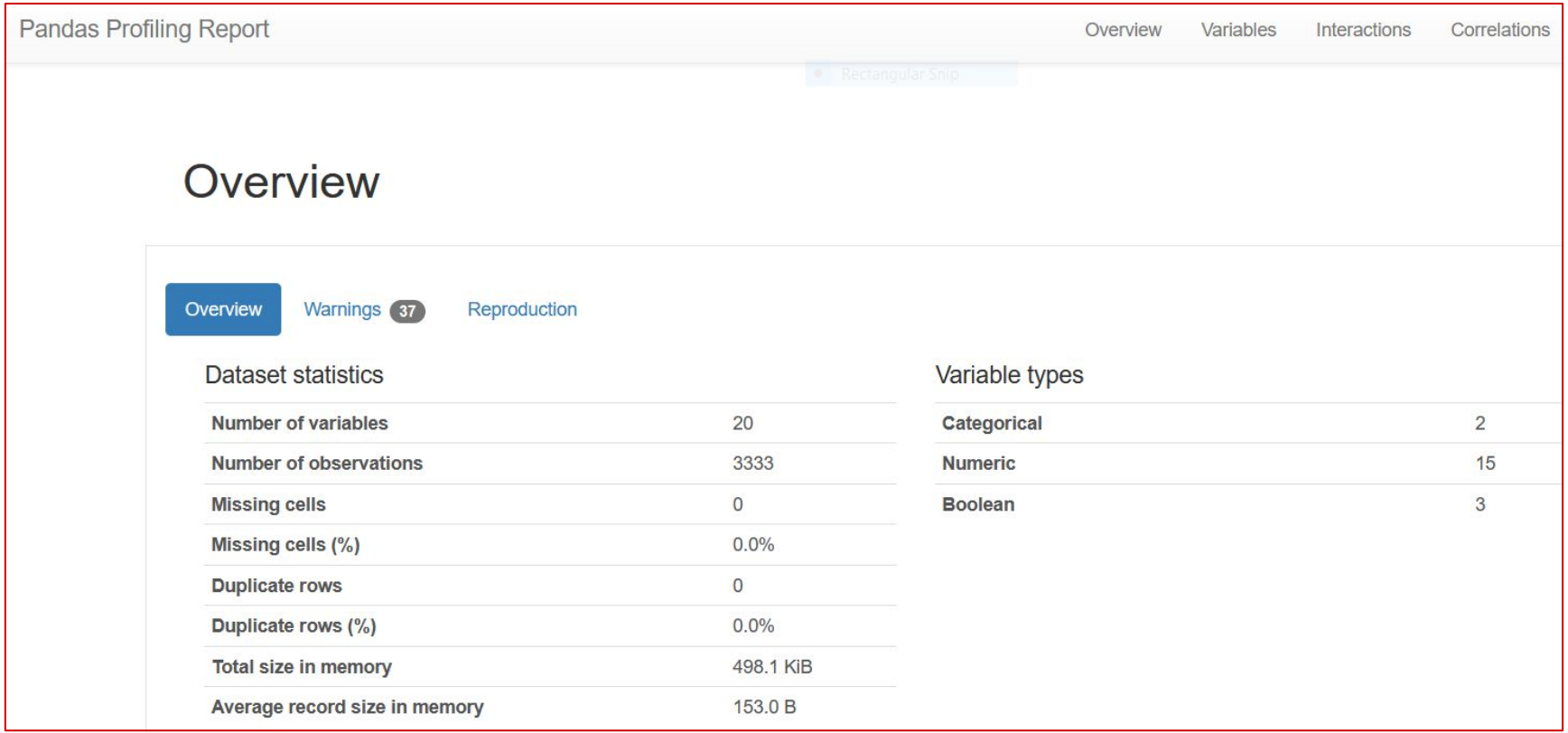
Here we use Violinplot to represent relationship between Total calls and Churn



EDA (Histogram for Numeric variables)



EDA Report



Pandas Profiling Report

Overview Variables Interactions Correlations

Overview

Overview | Warnings 37 | Reproduction

Dataset statistics

Number of variables	20
Number of observations	3333
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	498.1 KiB
Average record size in memory	153.0 B

Variable types

Categorical	2
Numeric	15
Boolean	3

Conclusions:

1. Based on our observation we can see that most people who leave the service are the ones who use the service in the day/morning.
2. It can also be observed that most people who use the service in the morning speak for shorter amounts of time but make more calls.
3. International plan users are more consistent with their churn w.r.t the ones who do not have the service.

We can definitely suggest to prevent churn by:

1. Upgrading network to improve services for long duration users.
2. Updating Pricing Strategies.
3. Updating and Optimizing Internationall Call Rates.
4. Implmenting a better network infrastructure in Maryland and Texas Areas where there is more Churn Rate.
5. Upgrading their services when in emegency only in evening period as low network traffic.
6. Introducing plans which minimize costs for more number of calls can be used.
7. Decreasing the prices as the talk-time increases can be an effective way to reduce the churn.
8. Improvement in the customer service can be done to reduce the number of calls which cause the churn.

CodeText

Q & A