Capstone Project
Telecom Churn Analysis

By:- Om Prakash Pradhan & Ruchika Nayak

Key Points

- ❑Introduction
- ❑Project Objectives
- ❑Data Summary
- ❑Feature Description
- ❑Univariate Analysis
- ❑Influence of Numerical Features on Churn
- ❑Influence of Categorical Features on Churn
- ❑Correlation matrix & Outlier detection
- ❑Suggestions
- ❑Conclusions

Introduction

- A churn is defined as the **percentage of subscribers** moving from a specific service **in a given period of time**.
- Churn is a problem for telecom companies because it is more expensive to acquire a new customer than to keep existing one from leaving.
- Telecom companies apply **machine learning models** to predict churn on an individual customer basis and take counter measures such as discounts, special offers or other gratifications to keep their customers.

Project Objectives

- To visualize and manipulate the data without any assumptions in order to help assessing the quality of the data.
- To find the hidden insights of the data by different visualization tools.
- To highlight the main variables/factors influencing customer churn.
- To come up with ways/recommendations to ensure customer retention.

Data Summary

- Source dataset is in csv format.
- Dataset contains total 3333 rows and 20 columns.
- The dataset compose of categorical and numerical features.
- Each row indicates the data of a single customer whereas columns indicates different features of customers.
- Churn is the variable which shows whether a customer is churned or not.
- There are no missing values and duplicate values in the dataset.

Feature Description

- **State:** Indicates the state where the customer resides.
- **Account length:** Indicates duration of account (in days).
- **International plan/Voice mail plan:** Indicates whether a customer opted for these plans or not.
- **Total minutes:** There are total 4 columns which indicates time spent on calls during day, evening, night and on international calls.
- **Total calls:** There are total 4 columns which indicates number of calls during day, evening, night and international calls.
- **Total charges:** There are total 4 columns which indicate charges on calls during day, evening, night and on international calls.
- **Customer service calls:** Indicates how many times a customer contacted customer service.
- **Churn:** Indicates whether a customer churned or not.

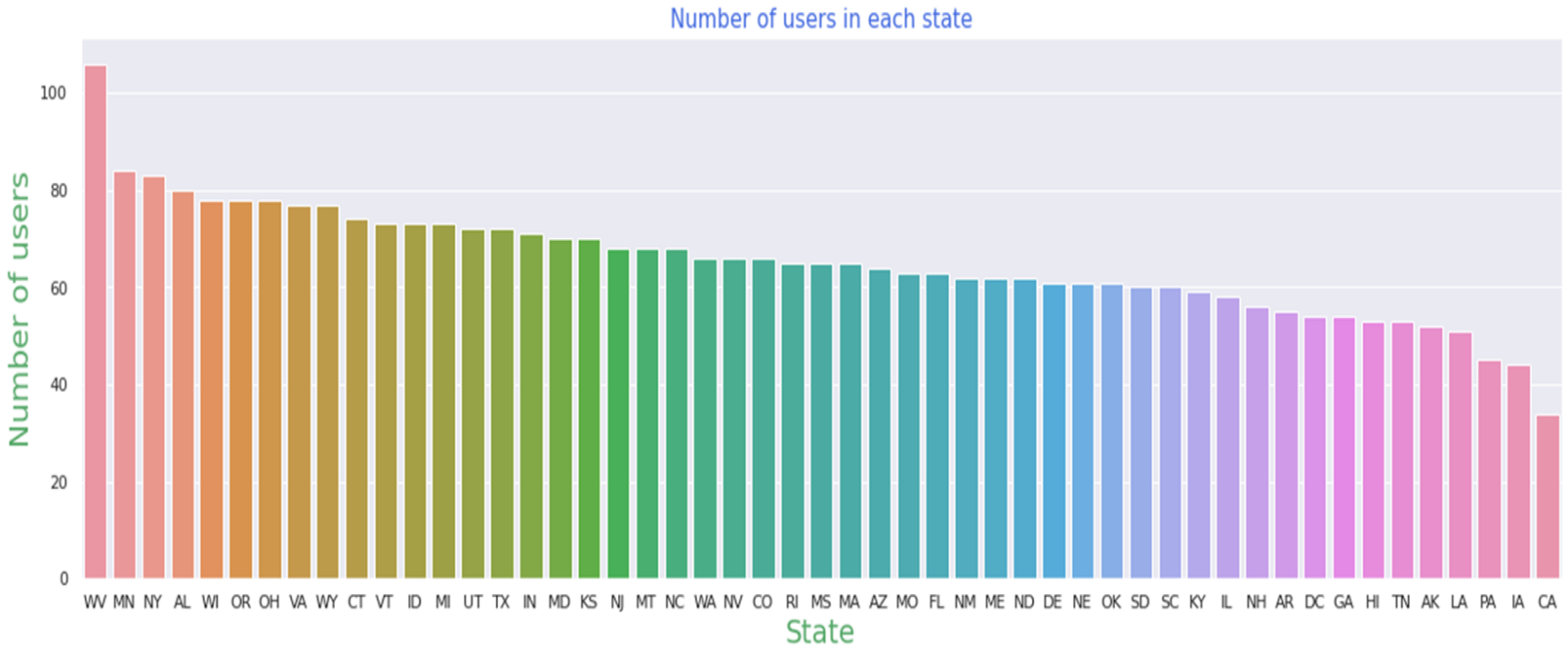
Columns	Data Types
State	Object
Account length	Int64
Area code	Int64
International plan	Object
Voice mail plan	Object
Number vmail messages	Int64
Total day minutes	Float64
Total day calls	Int64
Total day charge	Float64
Total eve minutes	Float64
Total eve calls	Int64
Total eve charge	Float64
Total night minutes	Float64
Total night calls	Int64
Total night charge	Float64
Total intl minutes	Float64
Total intl calls	Int64
Total intl charge	Float64
Customer service calls	Int64
Churn	Bool

Univariate Analysis

Analysis of state column

Bar plot

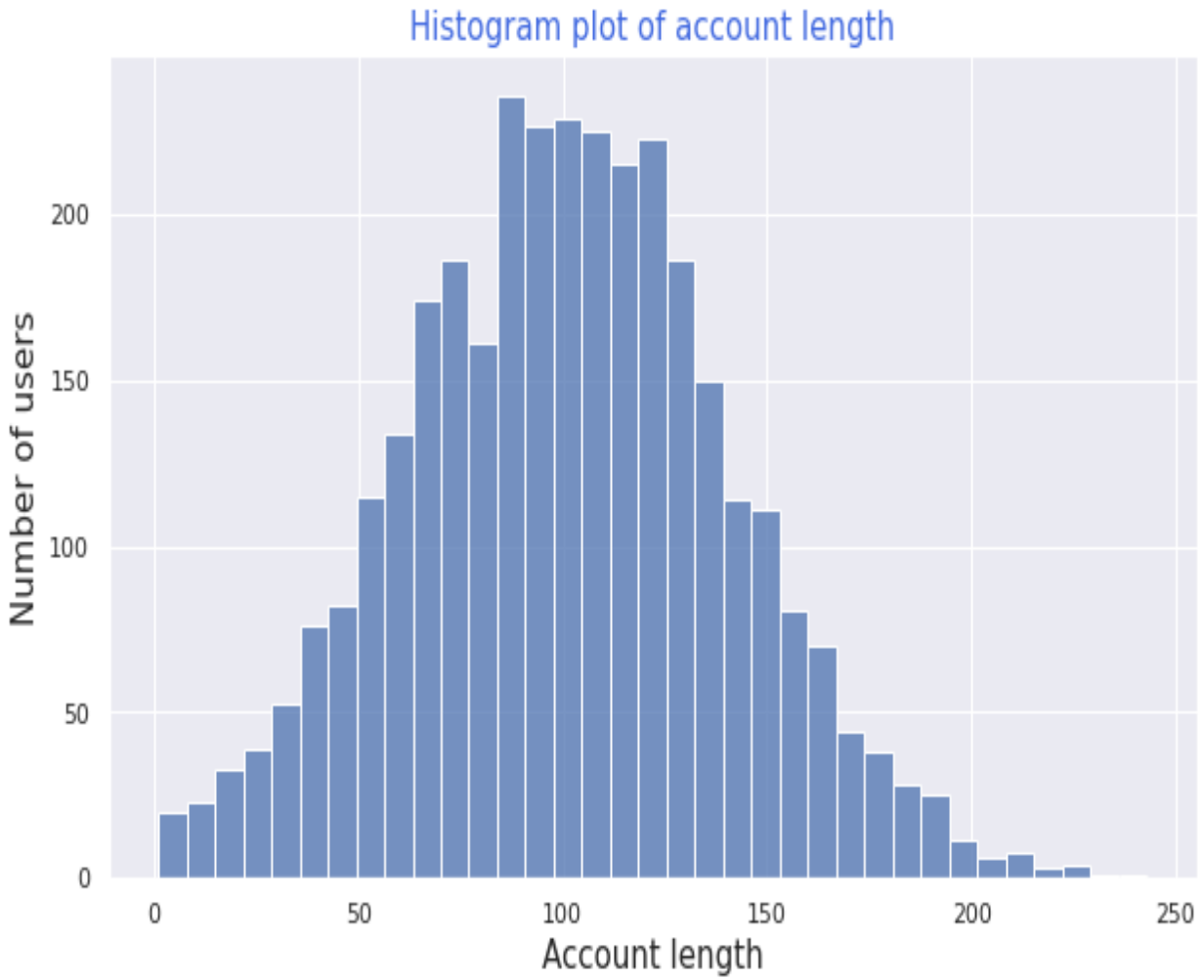
- This plot shows number of users in each state.
- There are more customers in WV, MN and NY states and less customers in CA, LA and PA states.



Analysis of Account length

Histogram

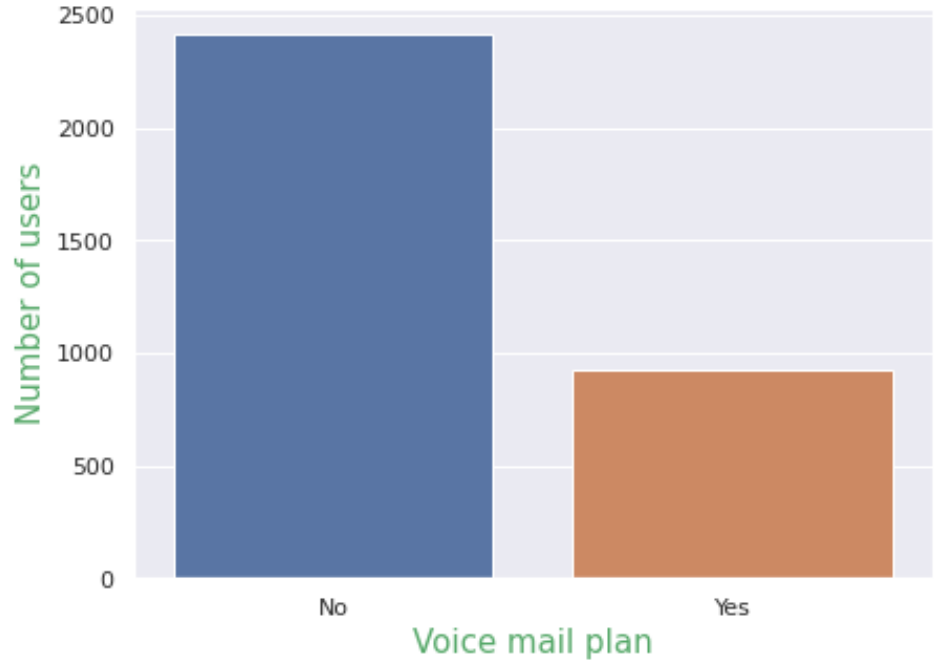
- This plot shows the distribution of account length of customers.
- Most of the customers have account for a duration of 100 to 130 days.

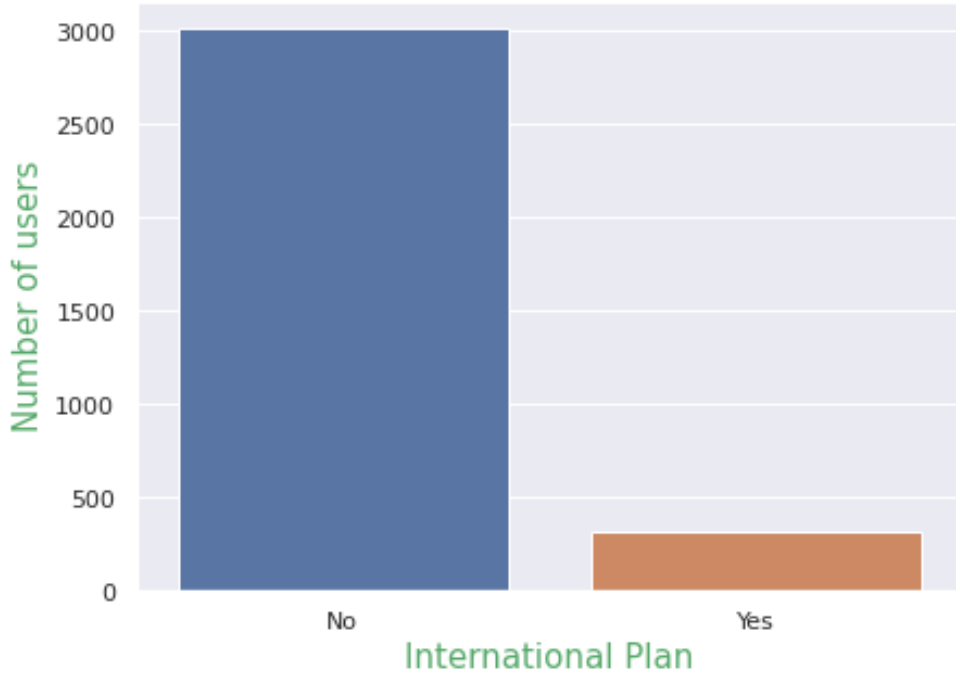


Analysis of Voice mail plan and International plan columns

Bar plot

- This plot shows number of customers opted for Voice mail plan and International plan.
- There are very few customers who have opted for international plan. And also there are not many customers who have opted for voice mail plan.

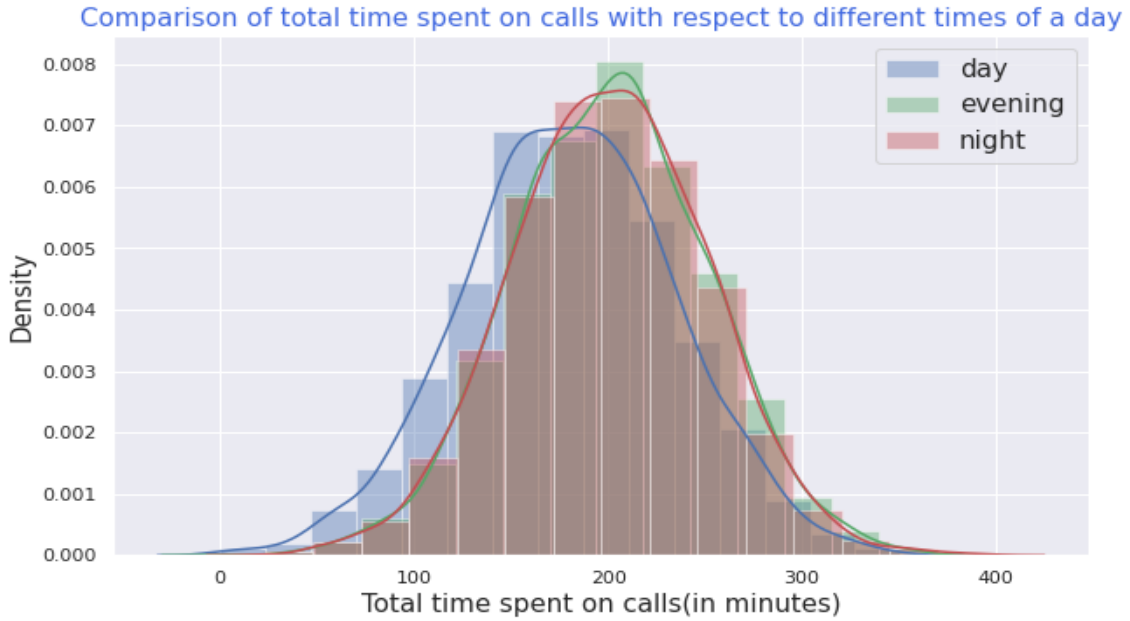


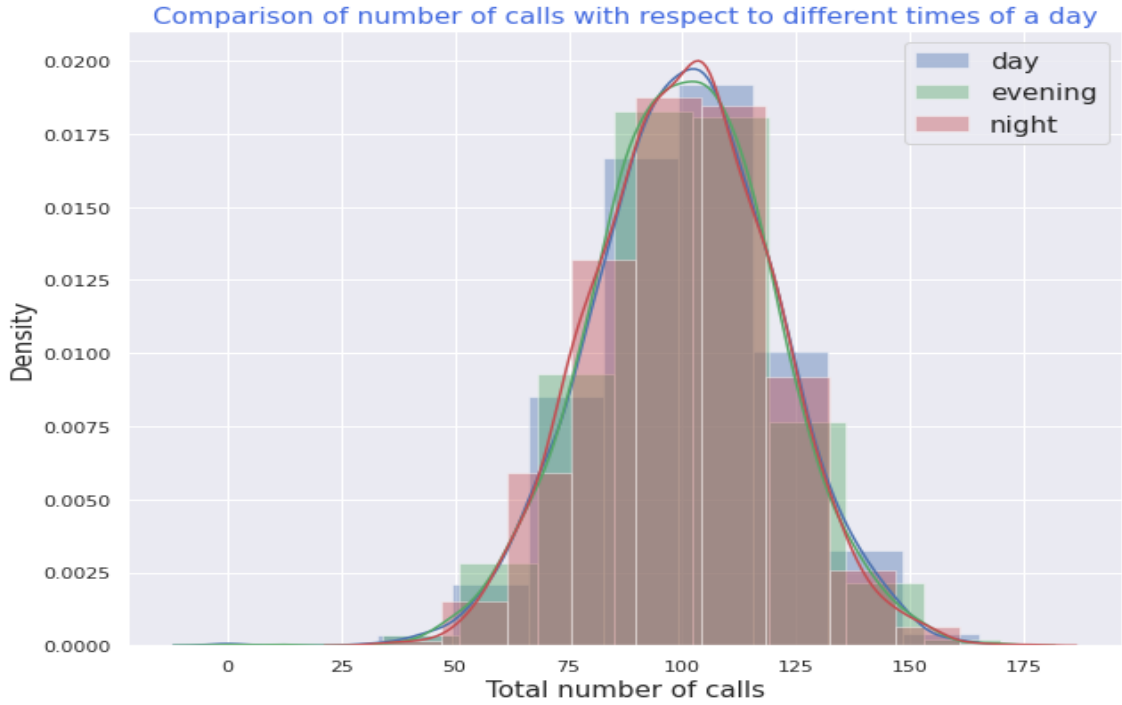


Analysis of Total time spent and total number of calls column

Distribution plot

- These plots show the distributions of total time spent on calls and total number of calls.
- Total time spent on calls in evening and night are a bit higher and Total number of calls are nearly same for all 3 times of a day.

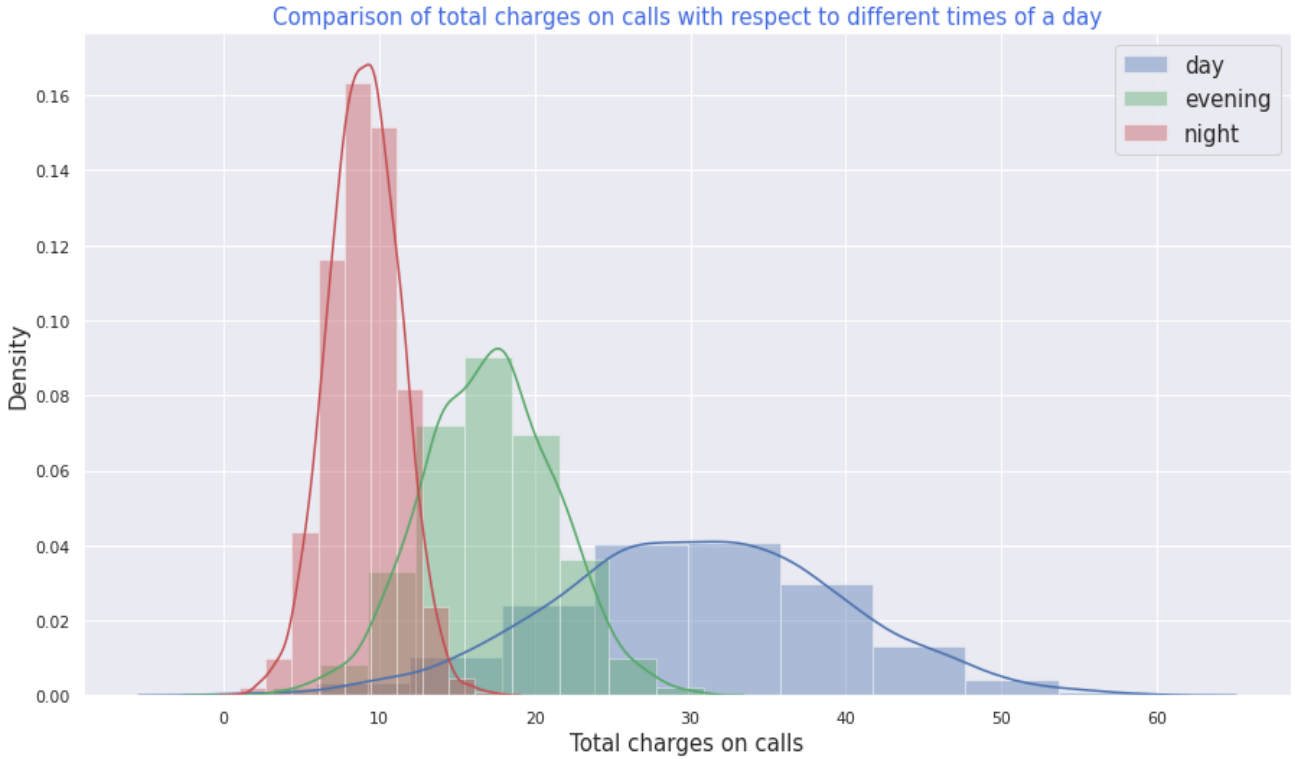




Analysis of total charges on calls column

Distribution plot

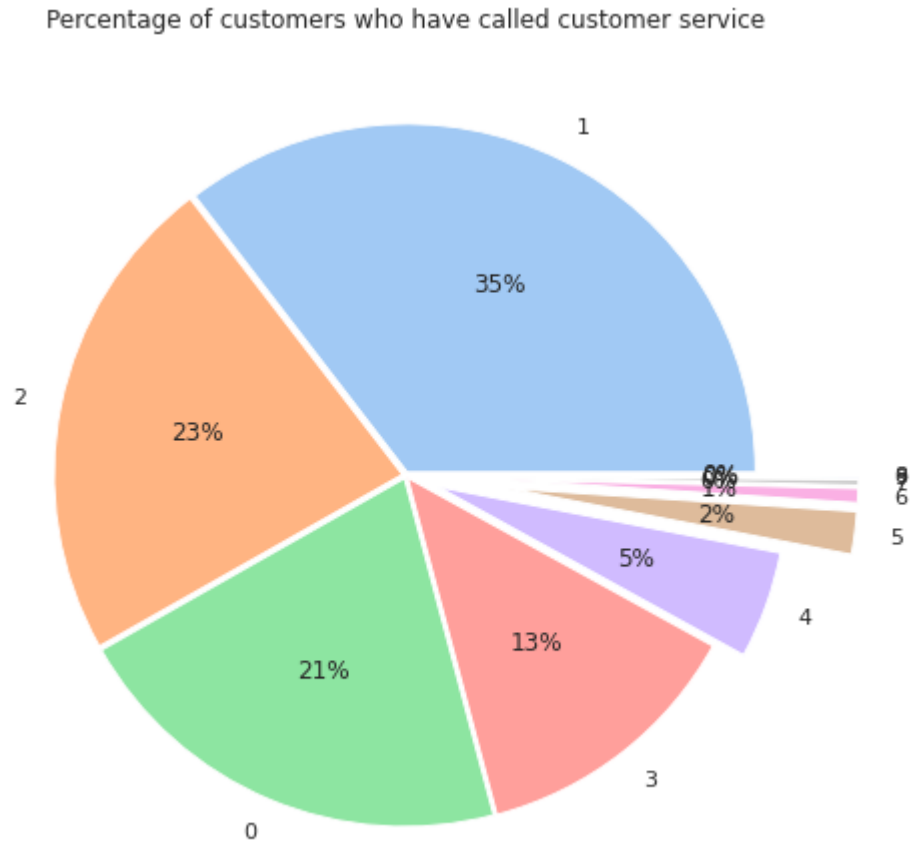
- This plot shows the distributions of total charges on calls.
- The charges in night are less as compared to other times of the day.



Analysis of customer service calls column

pie plot

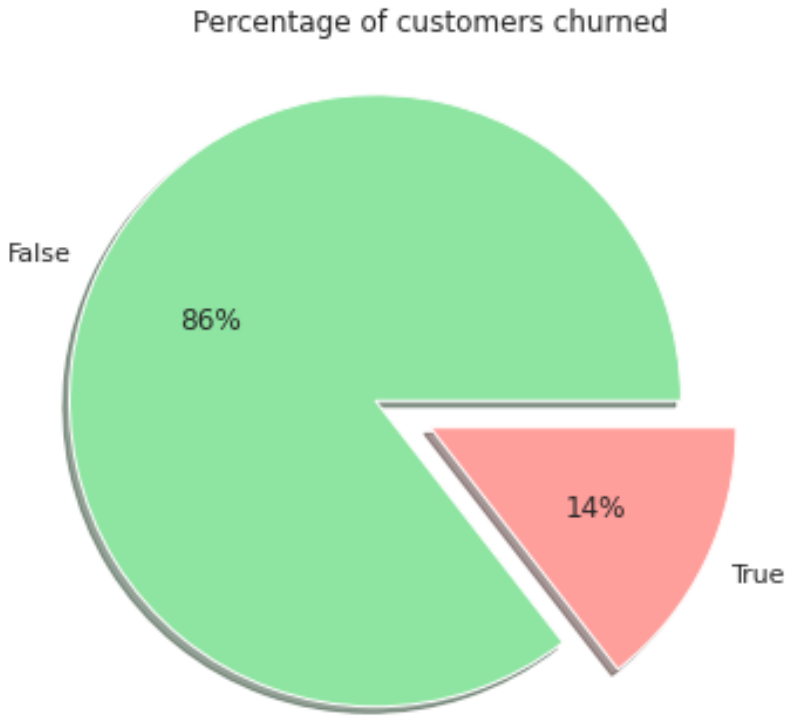
- This plot shows the percentage of customers who have called customer service.
- There are very few customers(around 8%) who have called customer service more than 3 times.



Analysis of churn column

pie plot

- This plot shows percentage of customers who have churned.
- There are 14% customers who have churned and 86% have not churned from this telecom company.

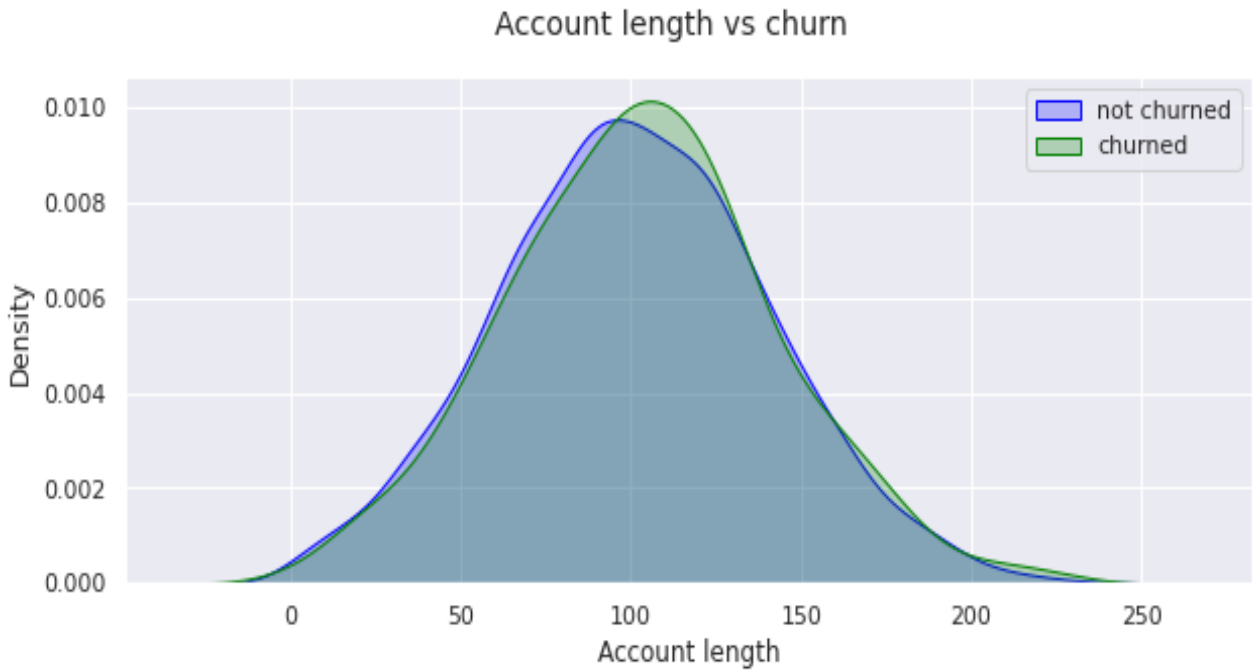


Influence of numerical features on customer churn

Account length vs. Churn

kernel density estimate (KDE) plot

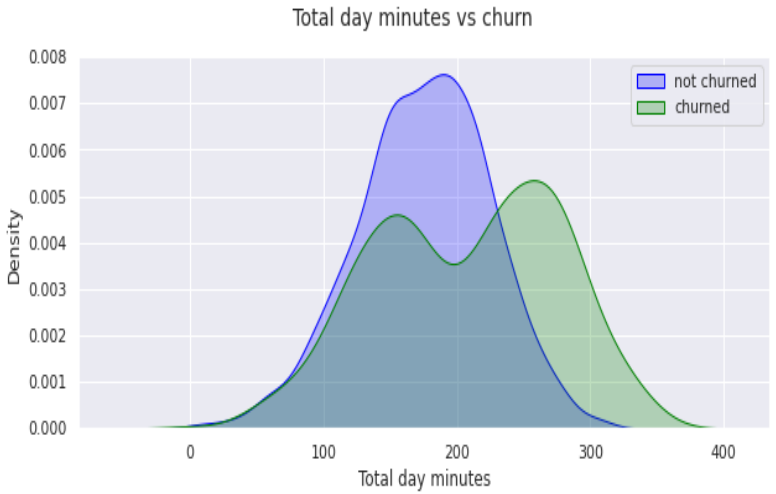
- This plot shows the density plot of account length vs. churn.
- Clearly there is no effect of account length on customer churn.

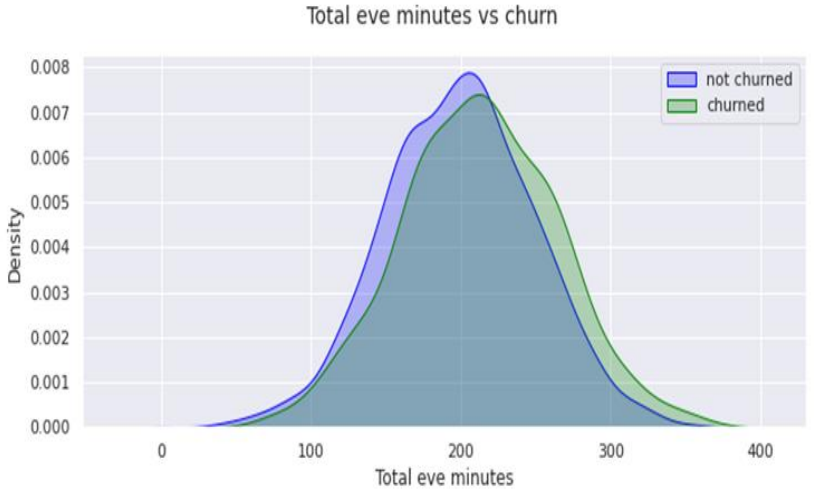


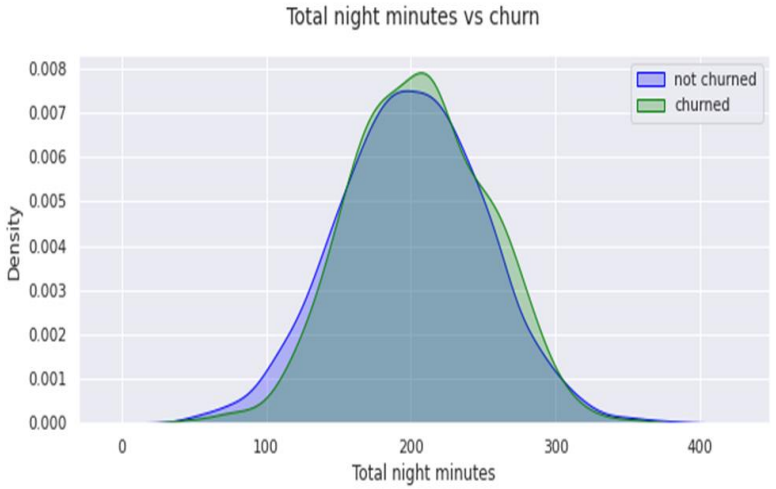
Total minutes vs. Churn

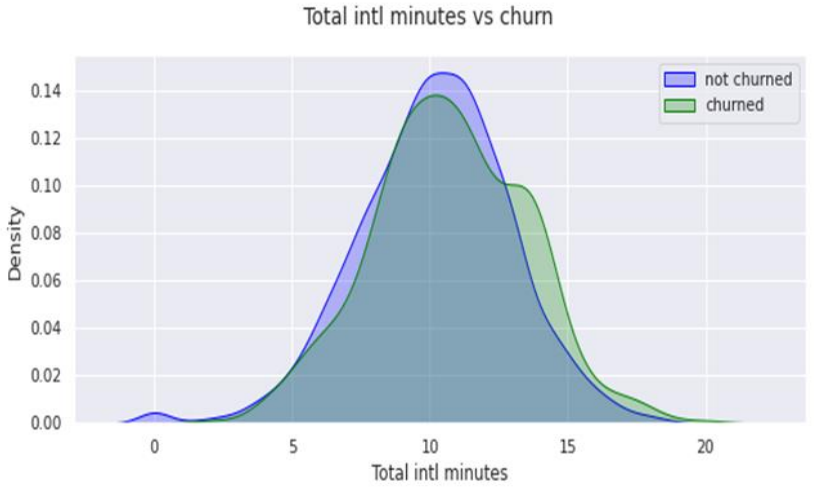
kernel density estimate (KDE) plot

- Churn is high when total day minute is high.
- Whereas there is no significant effect of total evening minutes, total night minutes and total international minutes on churn.





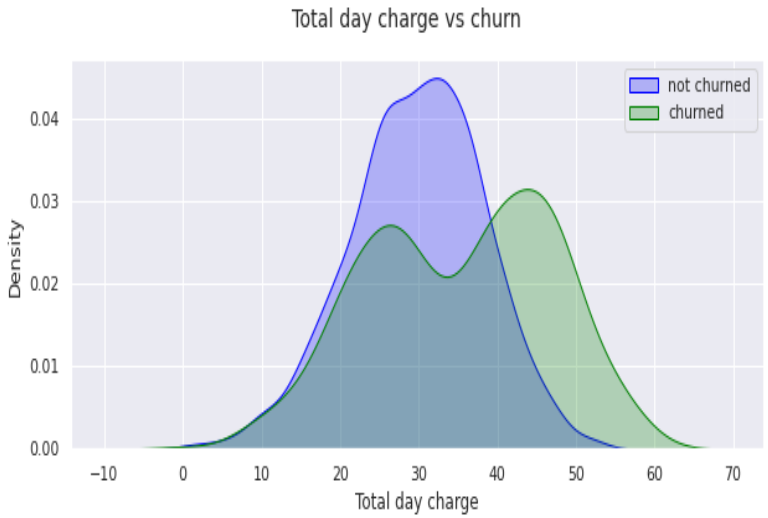


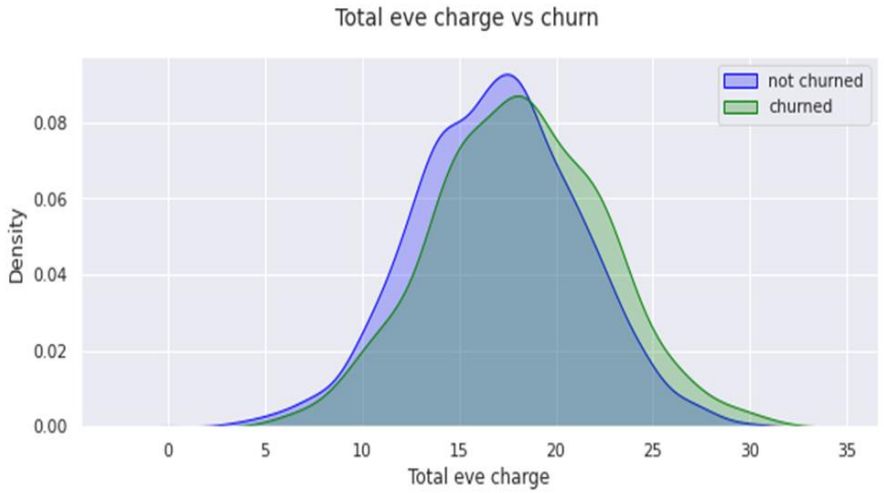


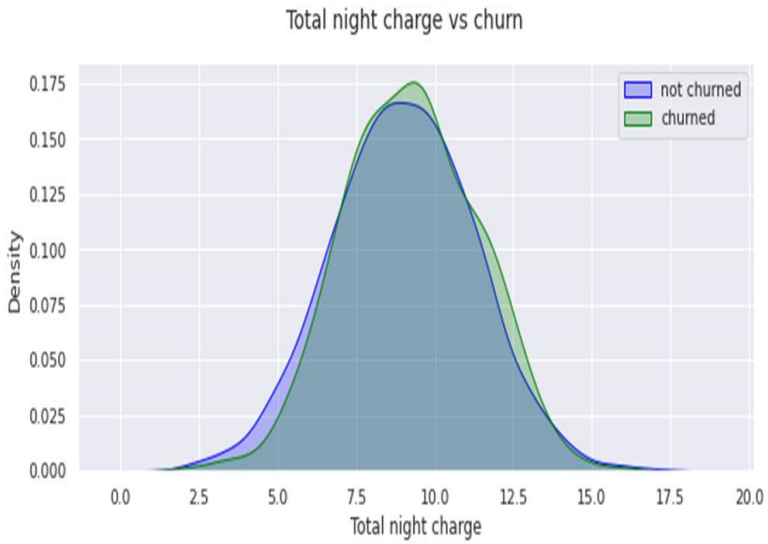
Total charge vs. Churn

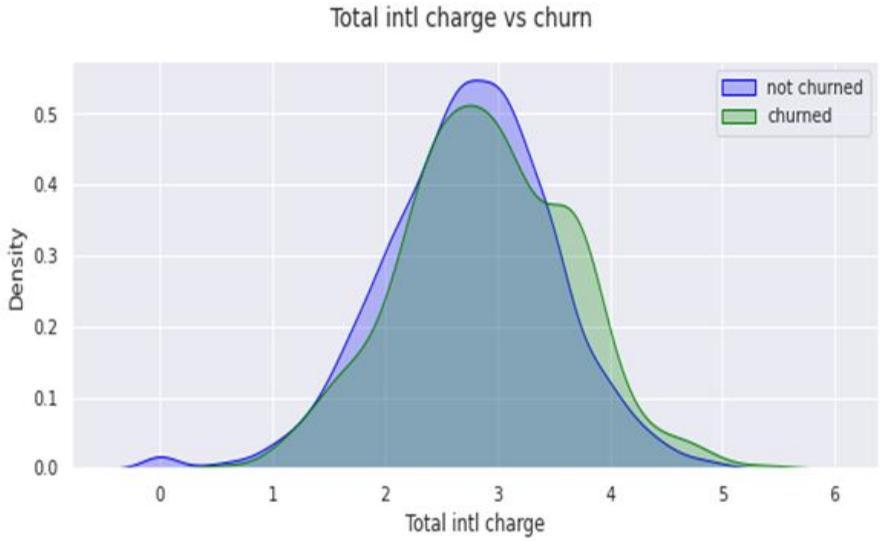
kernel density estimate (KDE) plot

- Churn is high when total day charge is high.
- Whereas there is no significant effect of total evening charge, total night charge and total international charge on churn.





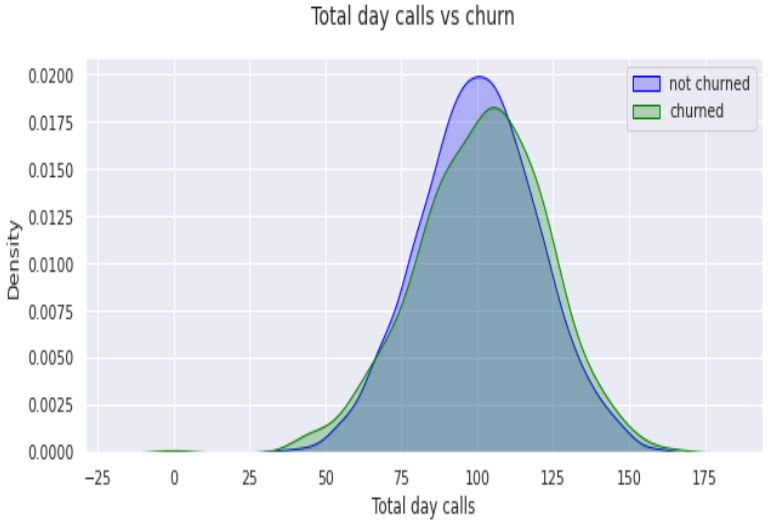


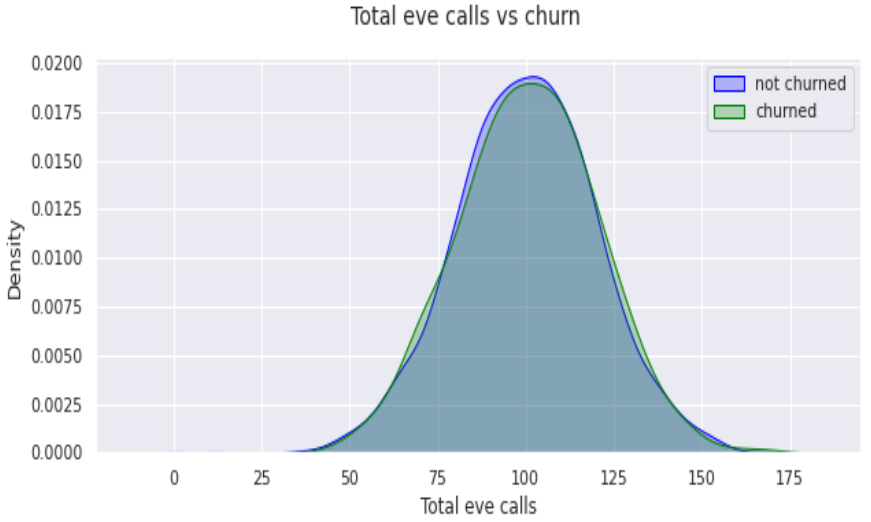


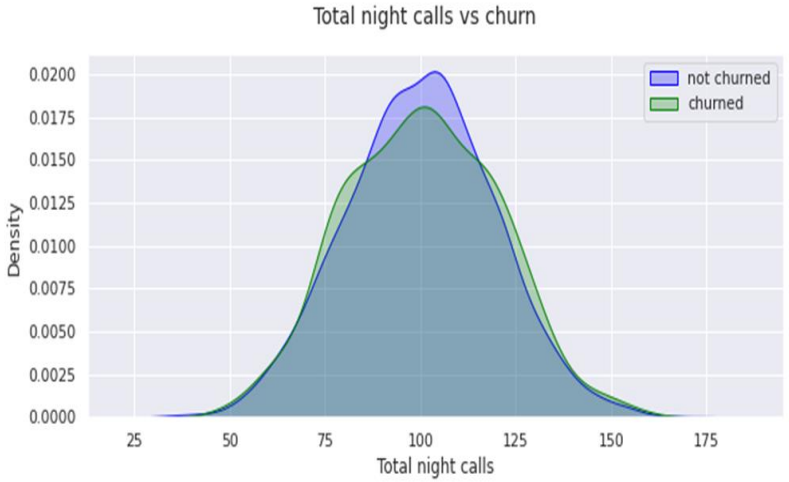
Total number of calls vs. Churn

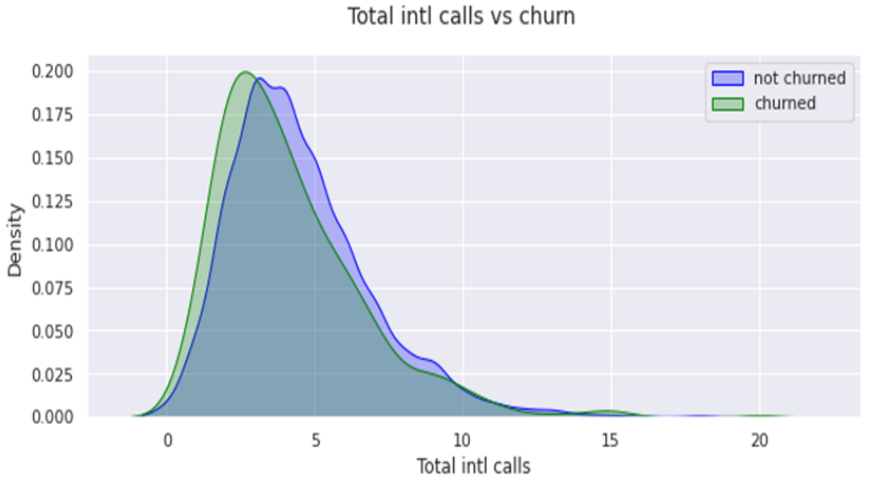
kernel density estimate (KDE) plot

- Churn is high when total international calls are less.
- Whereas there is no significant effect of total day calls, total evening calls and total night calls on churn.







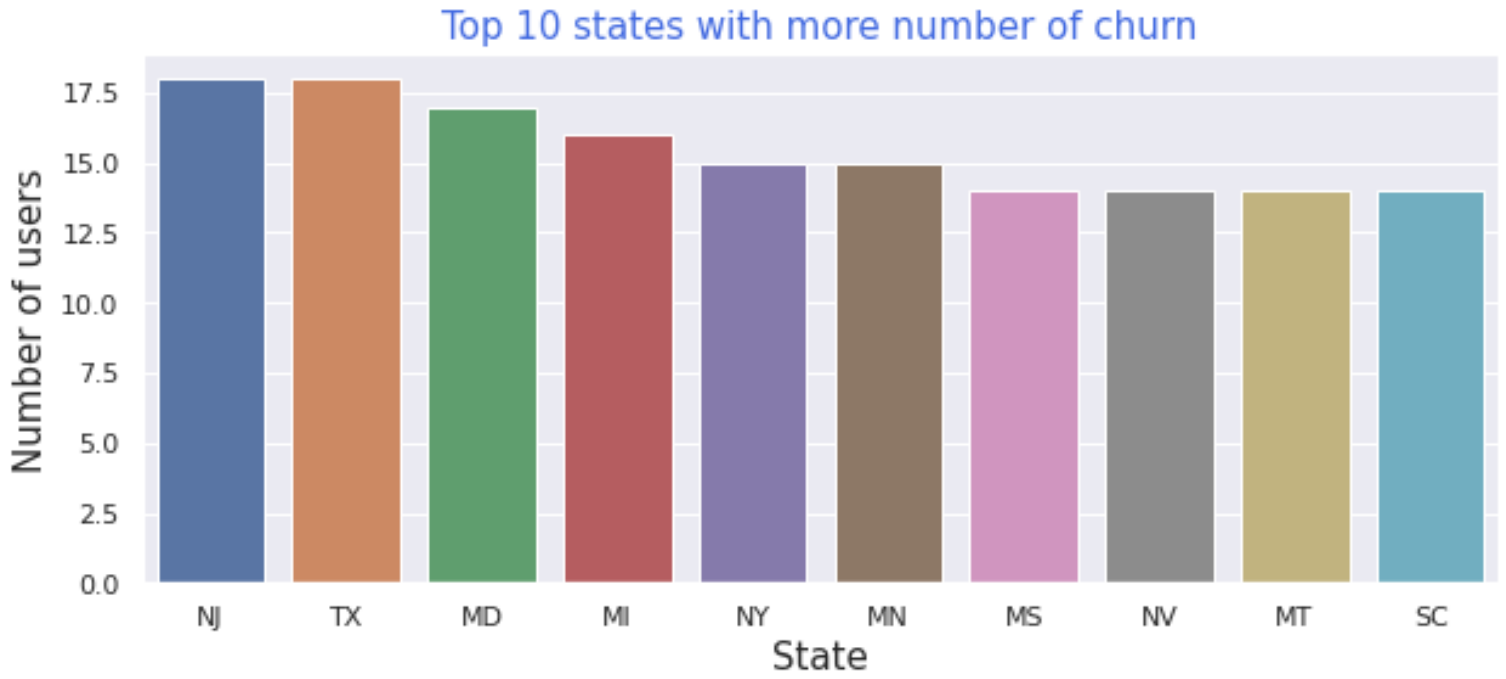


Influence of categorical features on customer churn

State vs. Churn

Bar plot

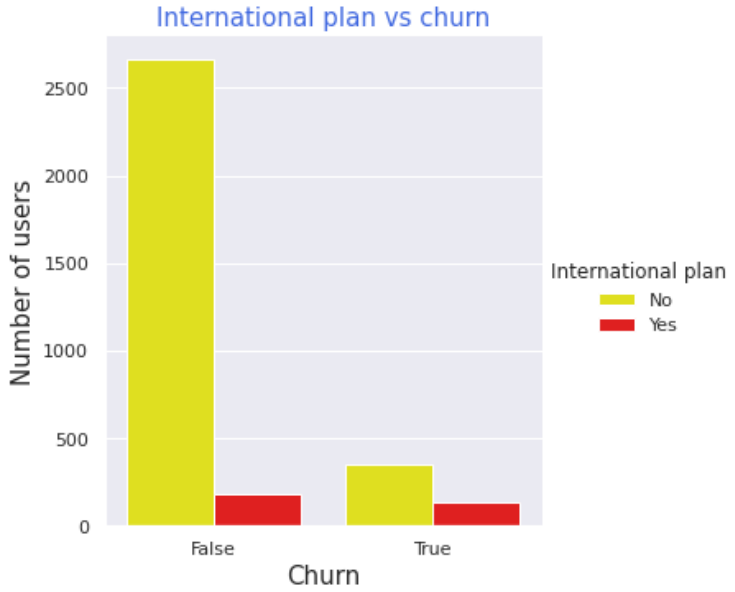
- This plot shows top 10 states where the churn is more.
- NJ, TX and MD are the states where customer churn is relatively high.

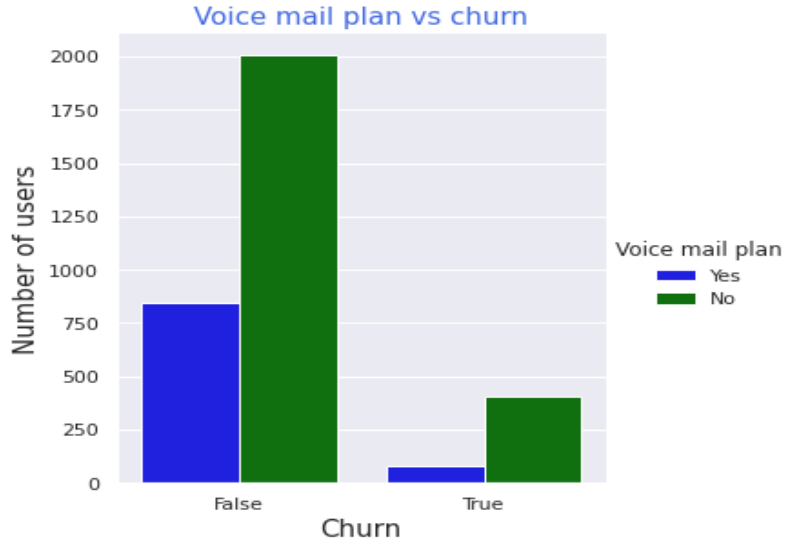


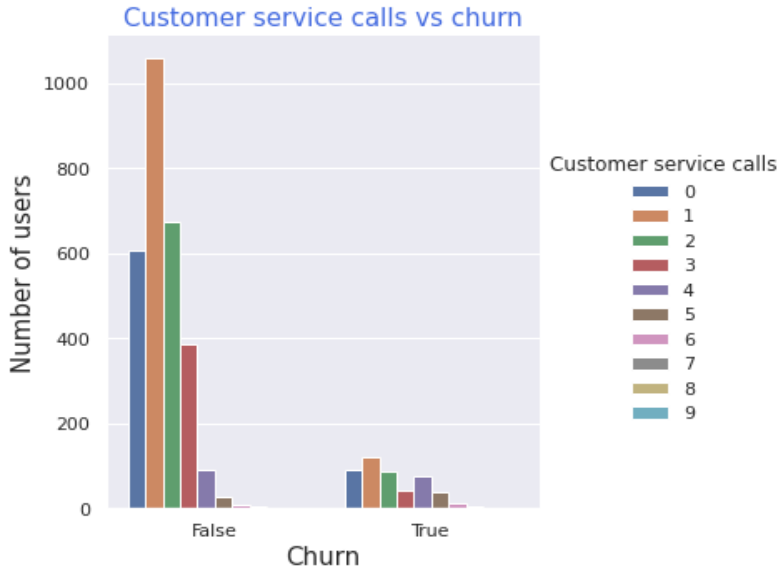
Other categorical features vs. churn

Cat plot

- Customers who have churned among those most of them have not opted for international plan.
- Customers who have churned among those most of them have not opted for voice mail plan.
- Customers who have churned among those most of them have called customer service once.





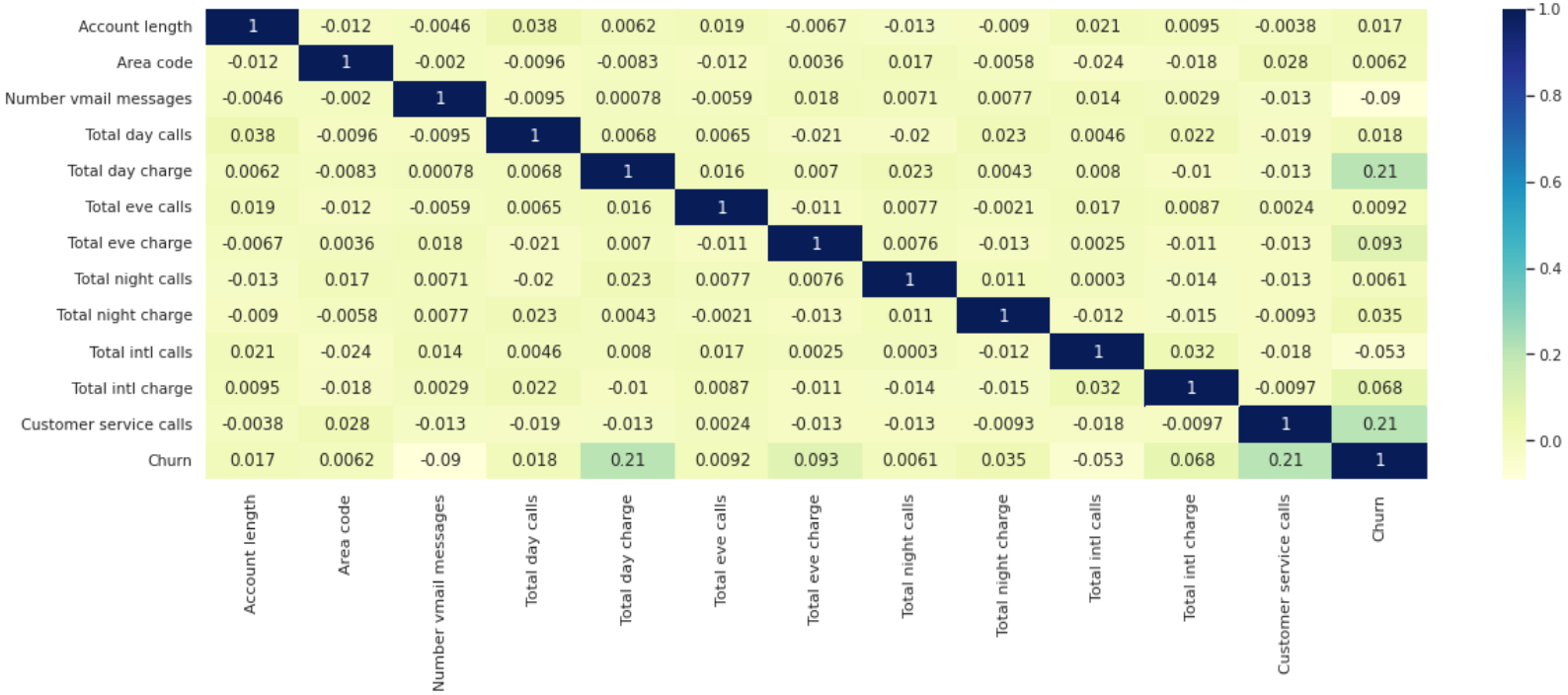


Correlation matrix
&
Outlier detection

Correlation matrix

Heat map

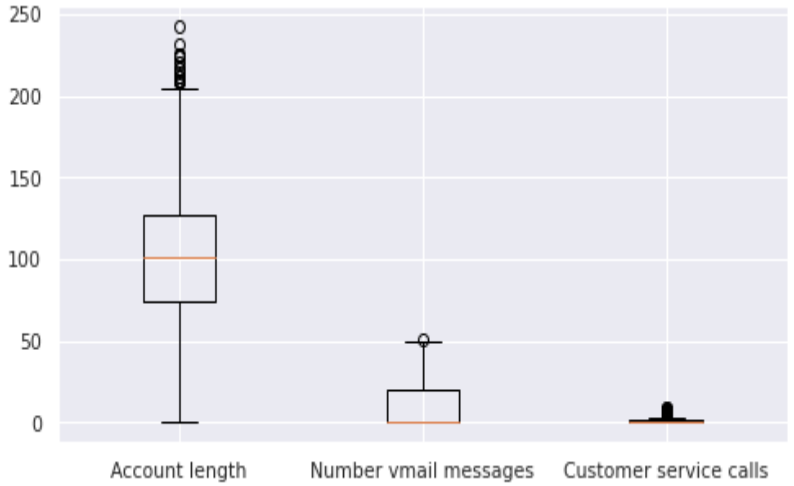
- These plots show the interdependency between the numerical variables.
- Here total charge and total minutes for day, evening, night and international calls are highly correlated.

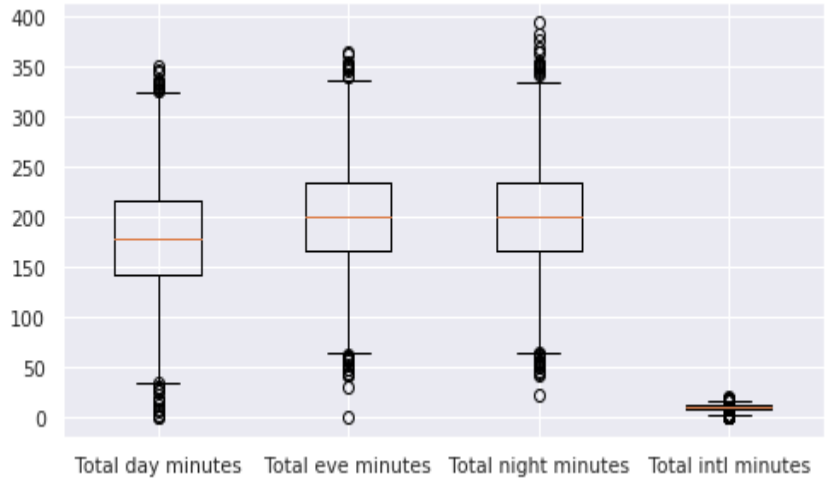


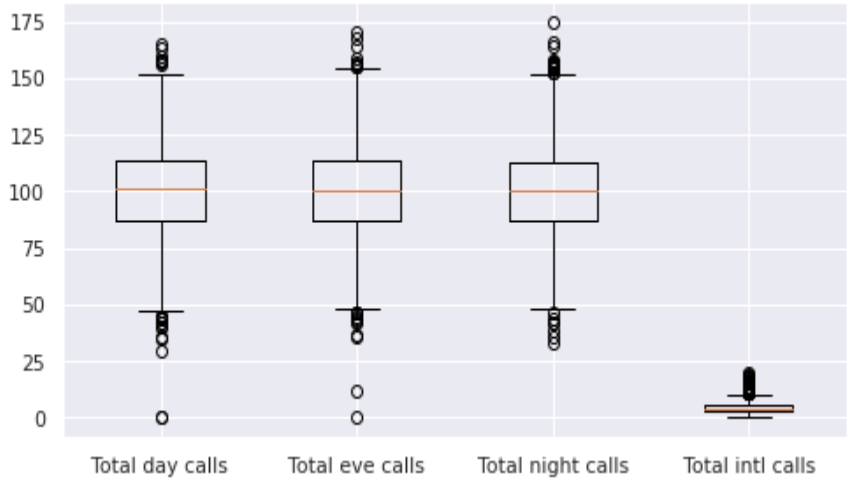
Outlier detection

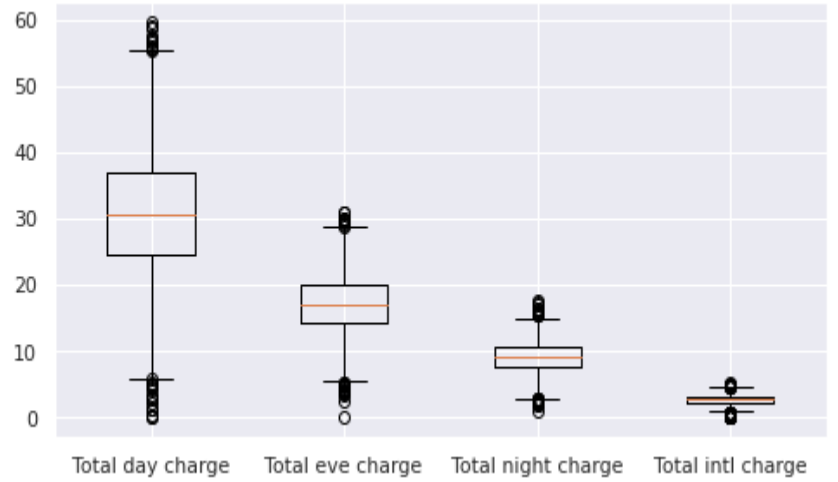
boxplot

- There are some outliers present in both sides of maximum variables. We can remove or replace these values with suitable statistical metric for further analysis to get better results.
- It can be concluded from the plot that the charges in days are higher.









Suggestions

- The company need to focus on some factors to reduce churn because now a days, an unsatisfied customer has many more channels to distribute his/her discontent about telecom services. This poor experience influences others in a wider digital community, which then leads to higher churn.
- It need to answer constantly the feedback of customers on all channels, especially on social media.
- It need to reduce charges in order to reduce customer churn.
- It need to increase the quality of customer service to reduce churn.
- It should try to offer the better service for the churn customers and see how much this impact before and later.
- It could Anticipate churn by investing in better technology.

Conclusions

- 14% of customers are churning.
- There are some states like New Jersey, Texas etc. where the churn is more. The company need to grab more customers from states like California, Louisiana etc.
- Most of the customers have account for duration of 3-5 months. There are not a single customer having account length of more than a year.
- The charges in night are less as compared to other times of the day.
- There are very few customers(around 8%) who have called customer service more than 3 times.
- Churn is high when total international calls are less. But there are very few customers who have opted for this plan.
- As higher day charge leads to more churn, the company need to reduce day charges in order to reduce customer churn.
- Customers who have churned among those most of them have called customer service once.
- Total day charge and customer service calls are correlated with churn with correlation coefficient of 0.21.
- There are some outliers present in both sides of maximum variables.


Thank
you