Capstone Project – 1

EXPLORATORY DATA ANALYSIS

Topic - TELECOM CHURN ANALYSIS

BY – Avilash Srivastava

Contents of the presentation

- Company Intro
- Problem Statement
- Data Summary
- Anomaly Detection
- Numerical Variable Analysis
- Categorical Variable Analysis
- Conclusion

Company Intro

Orange S.A., formerly **France Télécom S.A.**, is a French multinational telecommunications corporation. It has 266 million customers worldwide and employs 89,000 people in France, and 59,000 elsewhere. It is the 11th largest mobile network operator in the world and the 3rd largest in Europe after Vodafone, Telefónica . In 2015, the group had revenue of €40 billion. The company's head office is located in the 15th arrondissement of Paris.

Orange has been the company's main brand for mobile, landline, internet and Internet Protocol television (IPTV) services since 2006. It originated in 1994 when Hutchison Whampoa acquired a controlling stake in Microtel Communications during the early 1990s and rebranded it as "Orange". It became a subsidiary of Mannesmann in 1999 and was acquired by France Télécom in 2000. The company was rebranded as Orange on 1 July 2013.

Problem Statement

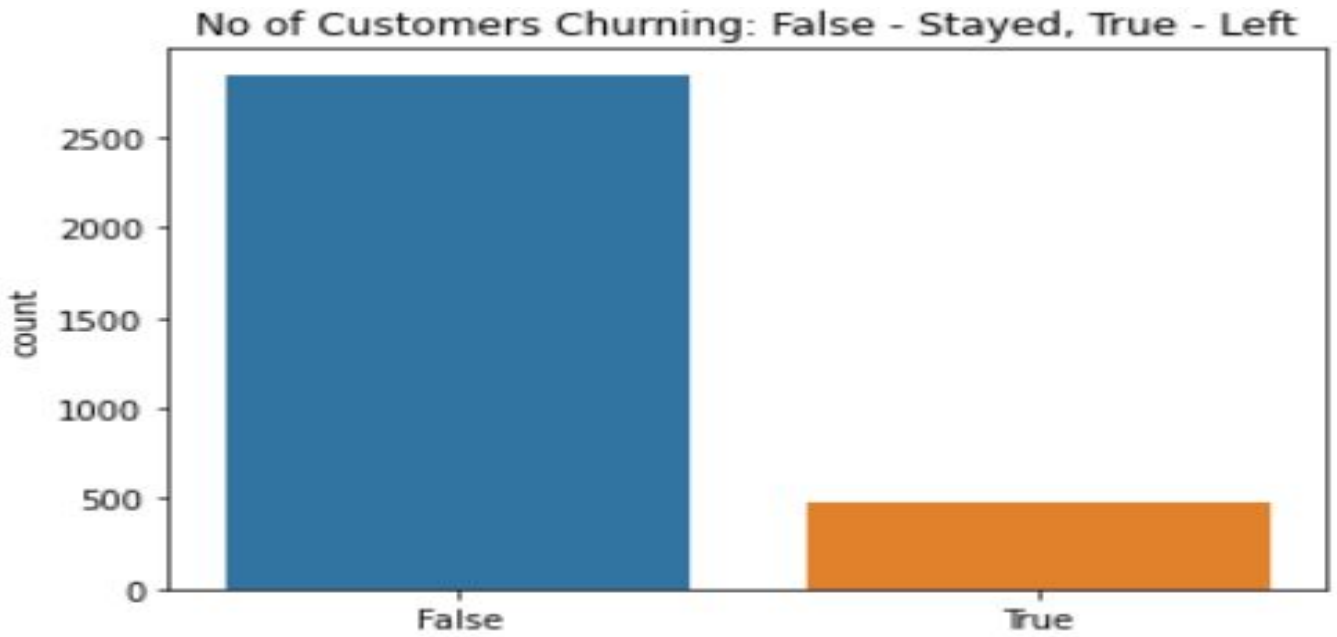
- The most important part of a business is to maintain a healthy relationship with its customers, because without them the company simply won't survive.
- Hence, it becomes an utmost priority to fulfill customer's needs, provide satisfaction and form a strong bond.
- But even after all efforts people tend to move on and adopt new changes. Thats life. But what can be done?
- Obvious way to keep customers loyal towards the company, is to keep checking the factors that are causing more customer dissatisfaction, and to improve on it.
- Thus it is always reasonable to keep records and try to find some patterns, and come up with positive solutions to ensure customer retention.

Data Summary

The data used for analysis had these features:

1. **State:** This variable tells us from where the customer belongs. Might be useful because different areas might have different price labels for using the service.
2. **Account Length:** This tells us the duration for which the customer has been using the services. Very useful to track loyal customers as obviously higher the no. the more likely the customer likes the service.
3. **Area code:** Again the idea of keeping track of the location of the customer
4. **International plan, Voice mail plan:** Having these plan may influence the prices that needs to be paid by customers.
5. **Total day minutes, Total day calls, Total eve minutes, Total eve calls, Total night minutes, Total night calls, Total intl minutes, Total intl calls:** These features give us an idea about how a customer utilizes the services i.e. how may calls he makes in a day, evenings and night. How much minutes he spends totally, which part of the day he is more active, does he make international calls etc. Clearly these have a positive correlation with the charges.
6. **Total day charge, Total eve charge, Total night charge, Total intl charge:** These features tell us the charges incurred by customers. Clearly dependent on the minutes spent, also very important data to keep track of because higher prices may bring customer dissatisfaction.
7. **Customer service calls:** Also a noteworthy variable because customers frequent calls may show dissatisfaction.
8. **Churn:** Finally the variable that tells whether the customer stayed with the company or moved on.

Data Summary (continued)



Total - 3333

False - 2850, True - 483

Anomaly Detection

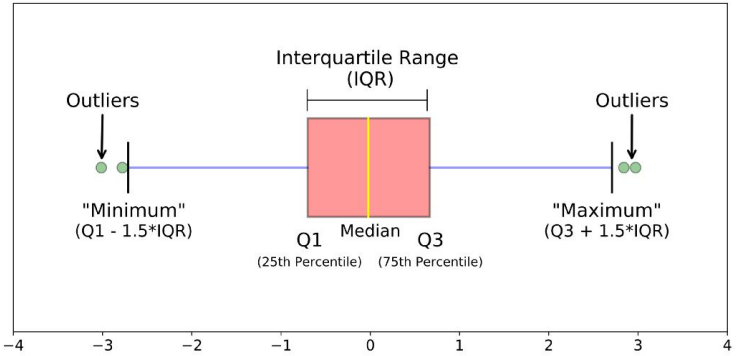
- It is always advisable to check for any missing values in our data.
- Depending on the strength and use cases we can either try to drop them, or replace with new values such as mean, median etc
- It seems here we don’t have any null values to take care of, hence we can proceed further.

State	0.0
Account length	0.0
Area code	0.0
International plan	0.0
Voice mail plan	0.0
Number vmail messages	0.0
Total day minutes	0.0
Total day calls	0.0
Total day charge	0.0
Total eve minutes	0.0
Total eve calls	0.0
Total eve charge	0.0
Total night minutes	0.0
Total night calls	0.0
Total night charge	0.0
Total intl minutes	0.0
Total intl calls	0.0
Total intl charge	0.0
Customer service calls	0.0
Churn	0.0

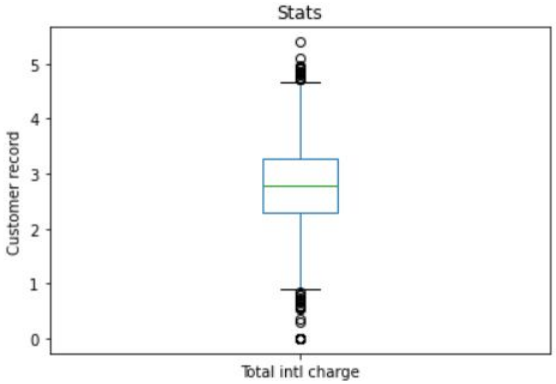
Fig- No of null values in each column

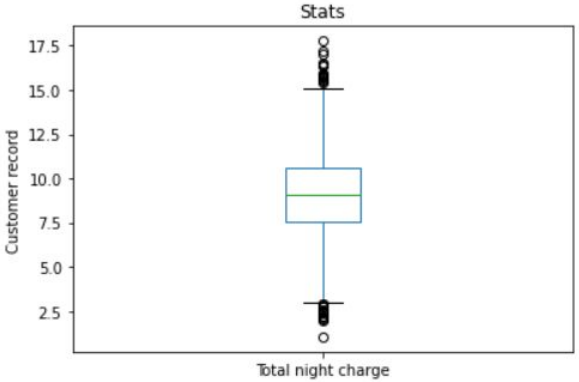
Anomaly detection (outliers)

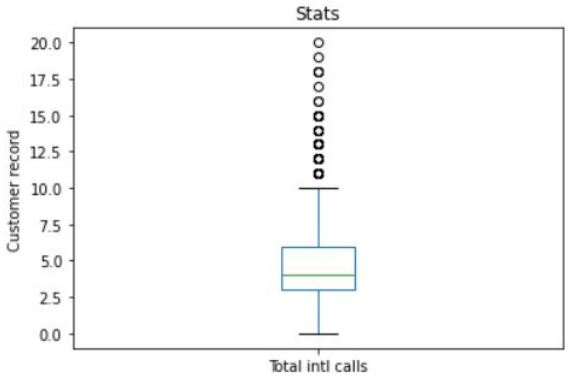
- Outliers are those values which do not comply with the normal range of values of a sequence of numbers.
  Eg. 2, 6, 134, 145, 152, 138, 141, 165, 900, 1478…
  Clearly 2, 3, 900, 1478 are outliers in the above example.
- These values need to be taken care of as they can dramatically affect the outcomes of an experiment.
- There is a very clear and easy way of detecting outliers with the help of box plots. Let's look at some of those.

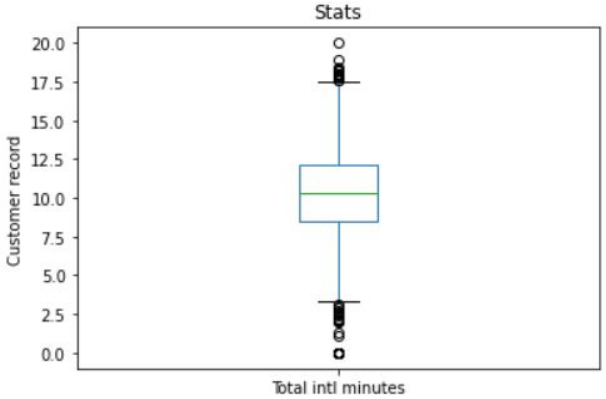


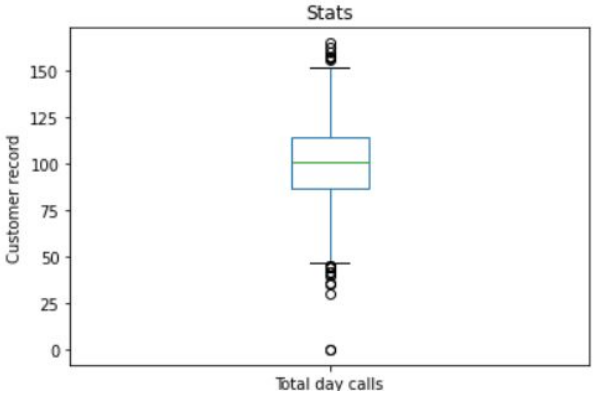
Anomaly Detection (continued)

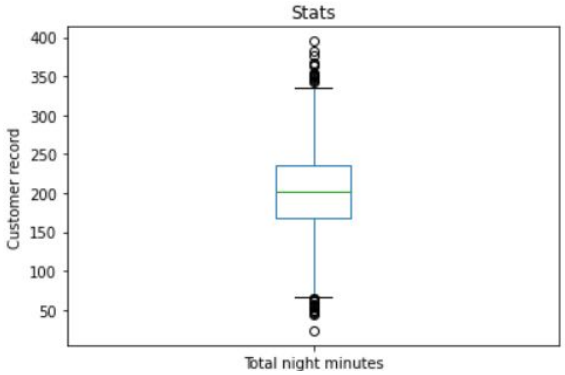




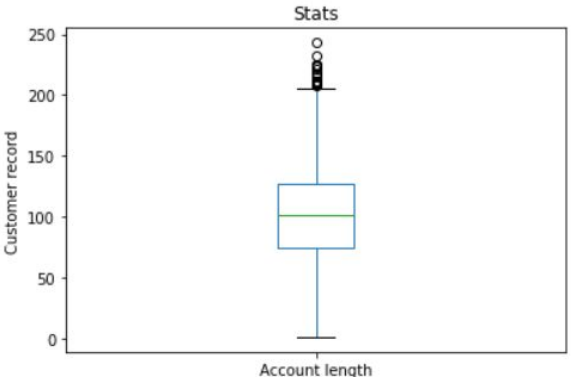


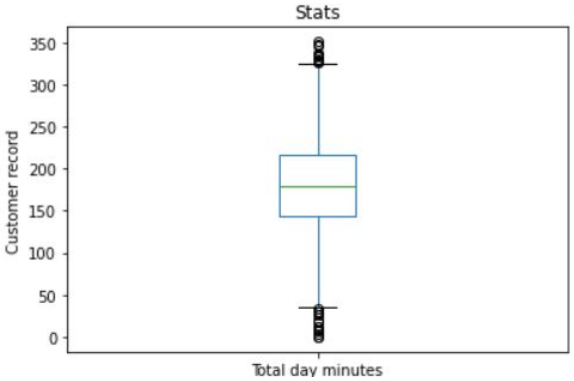


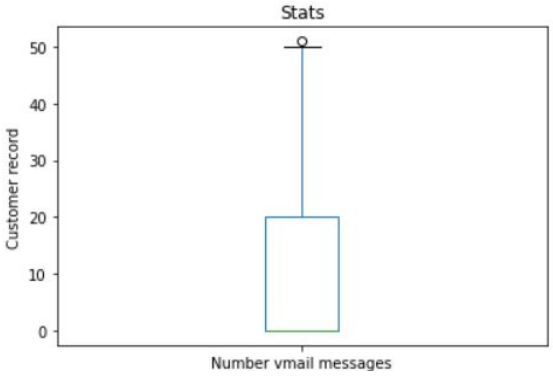


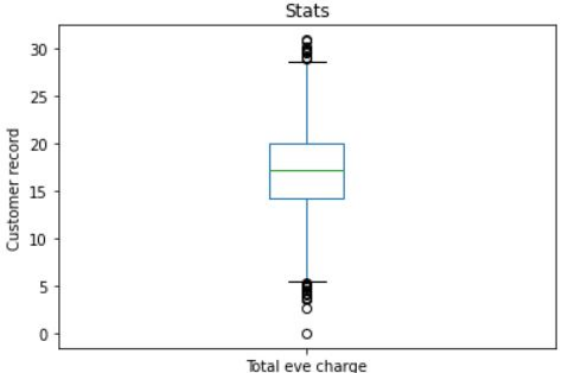


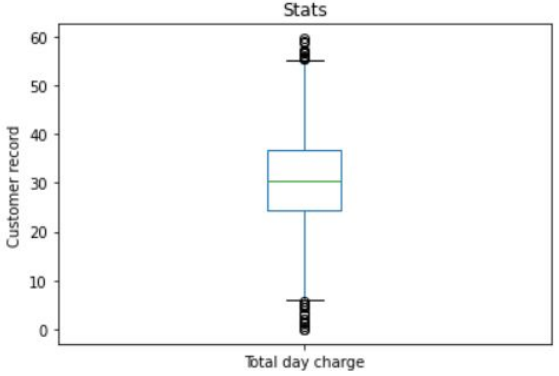
Anomaly Detection (continued)





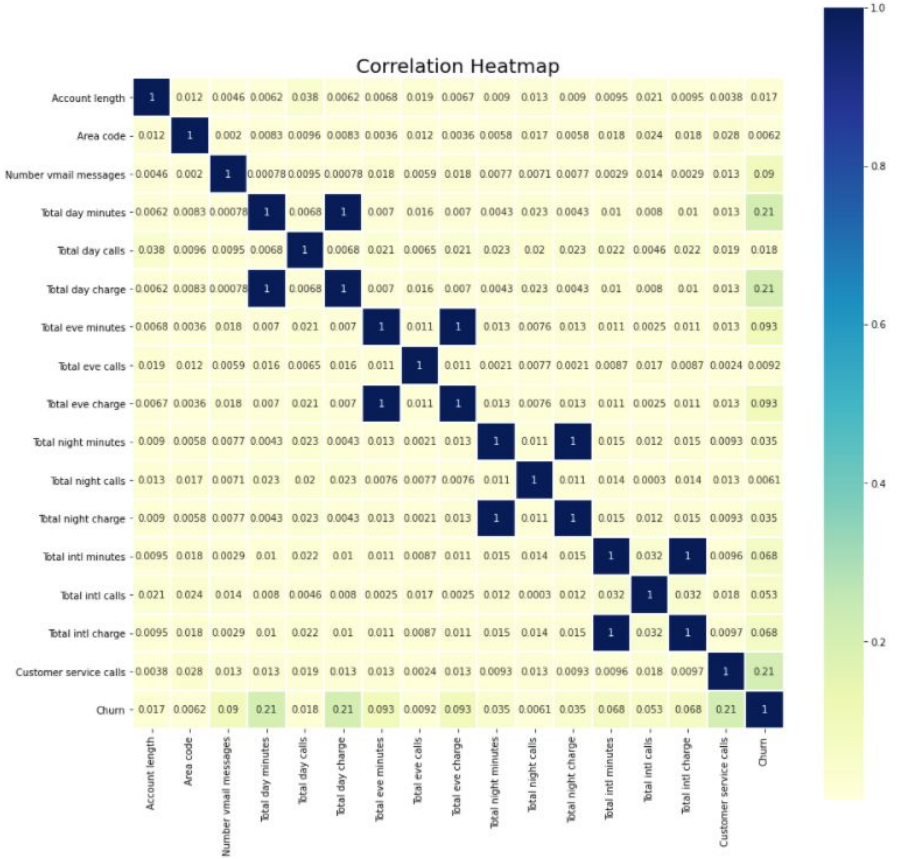




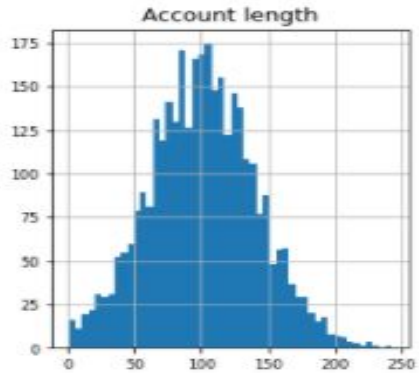


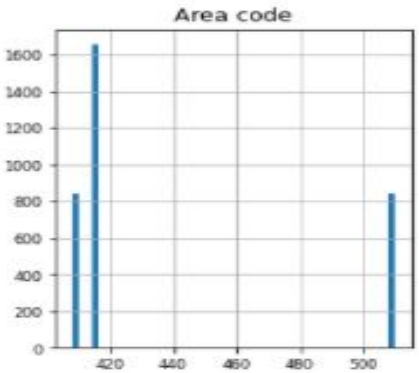
Numerical Variable Analysis

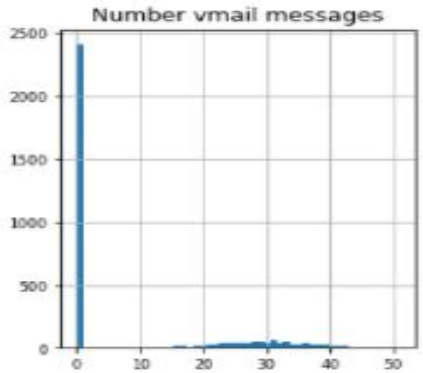
- Heatmap is a beautiful representation to show correlation between each variables.
- Darker the colour, the more correlated they are.
- Here we can see, minutes and charges data are almost perfectly correlated.
- Another way to examine is by using density or distribution plots, as correlated variables have similar kind of distribution, also useful to detect outliers. Any value, greater or smaller than mean plus 3 times standard deviation is treated as outliers.

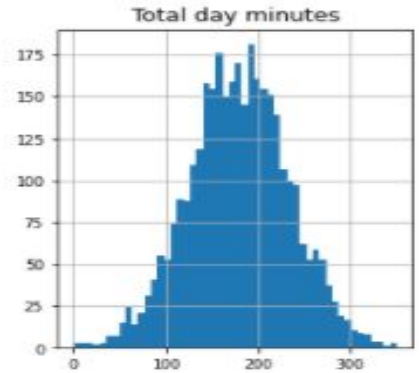


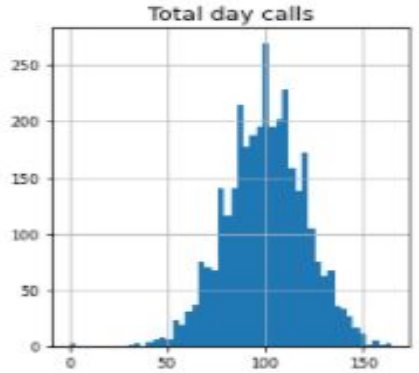
Numerical Variable Analysis (continued)

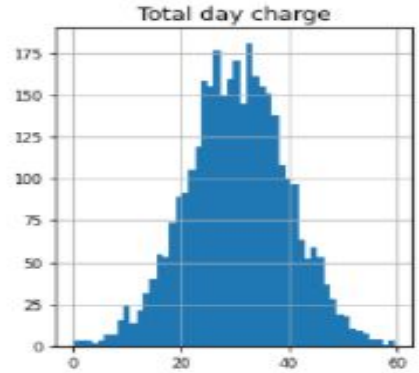


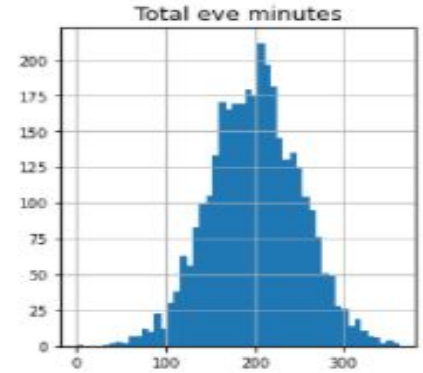


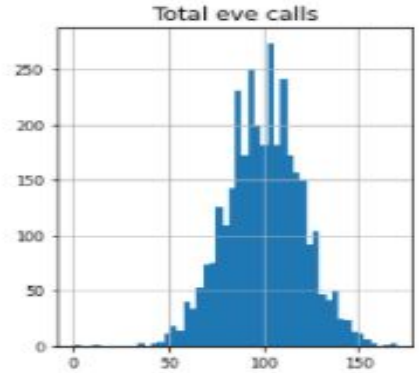




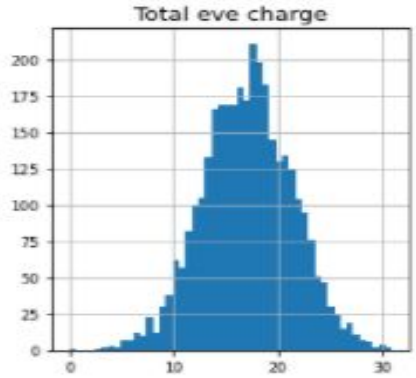


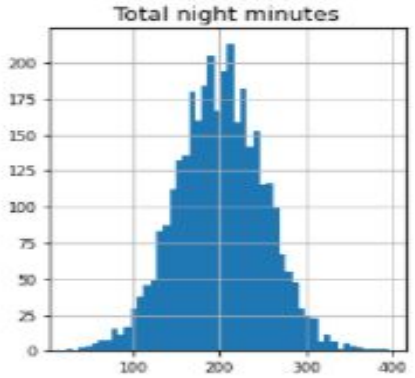


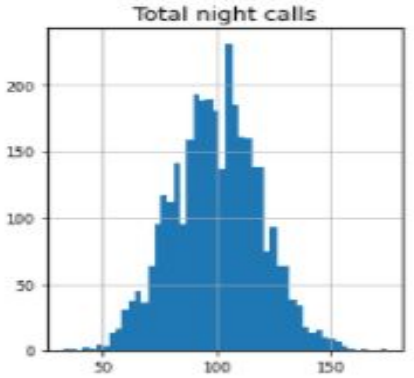


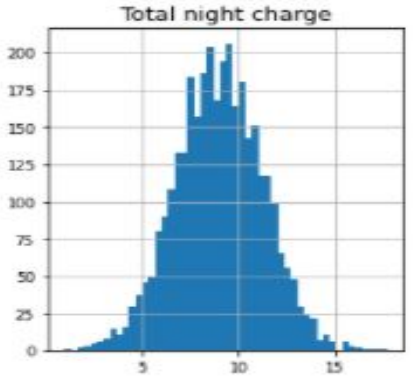


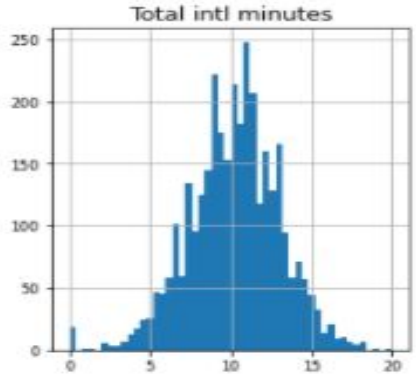
Numerical Variable Analysis (continued)

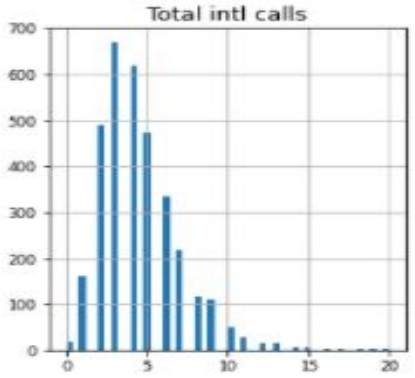


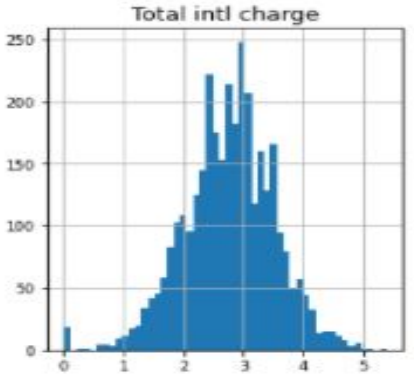










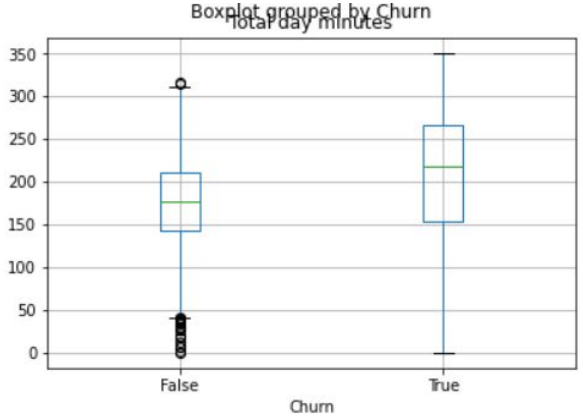


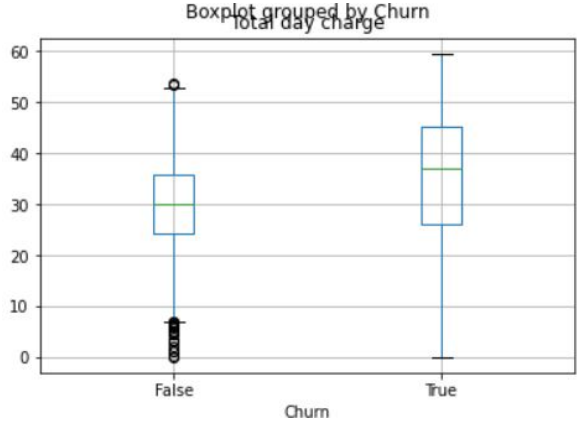
Numerical Variable Analysis (continued)

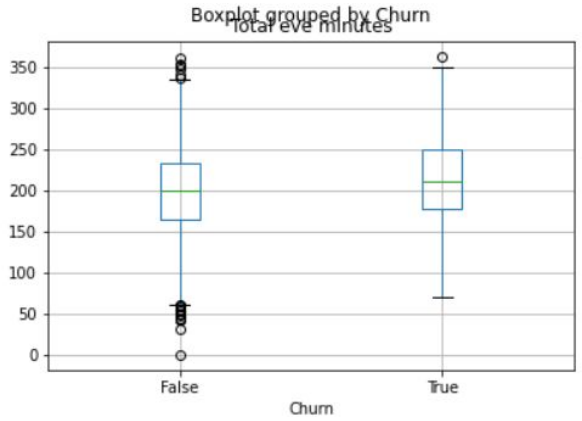
Churn	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls
False	100.793684	437.074737	8.604561	175.175754	100.283158	29.780421	199.043298	100.038596	16.918909	200.133193	100.058246	9.006074	10.158877	4.532982	2.743404	1.449825
True	102.664596	437.817805	5.115942	206.914079	101.335404	35.175921	212.410145	100.561077	18.054969	205.231677	100.399586	9.235528	10.700000	4.163561	2.889545	2.229814

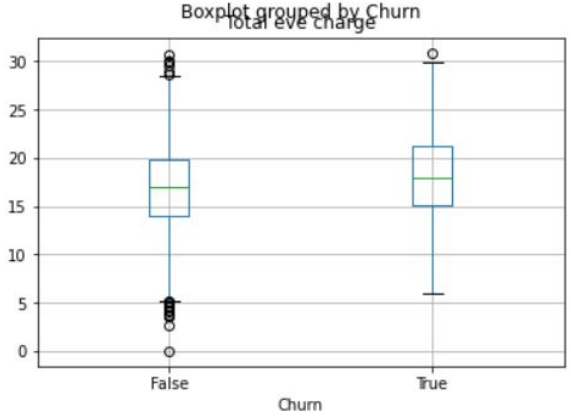
- Above figure shows the mean values of each variable grouped by Churn.
- Here we can observe a significant difference, for variables such as:
  - ➔ Number Vmail messages
  - ➔ Total day minutes
  - ➔ Total day charge
  - ➔ Total eve minutes
  - ➔ Total eve charge
  - ➔ Total night minutes
  - ➔ Customer service calls

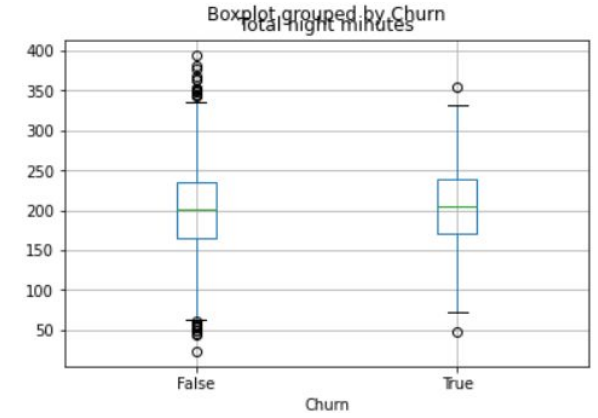
Numerical Variable Analysis (continued)

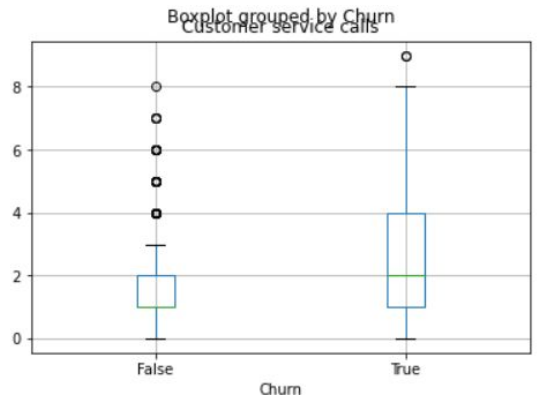




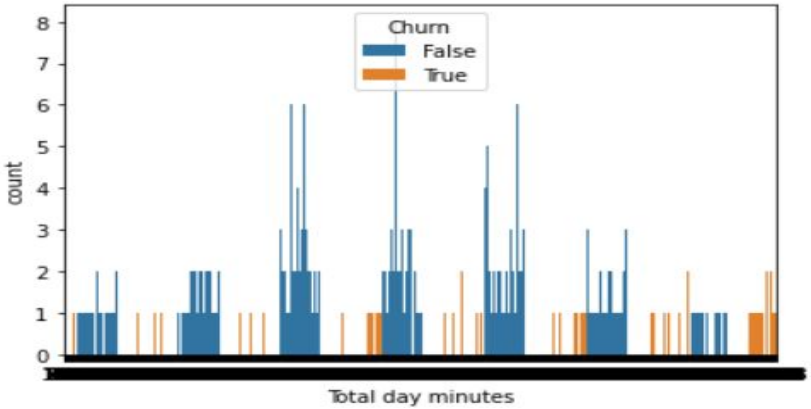


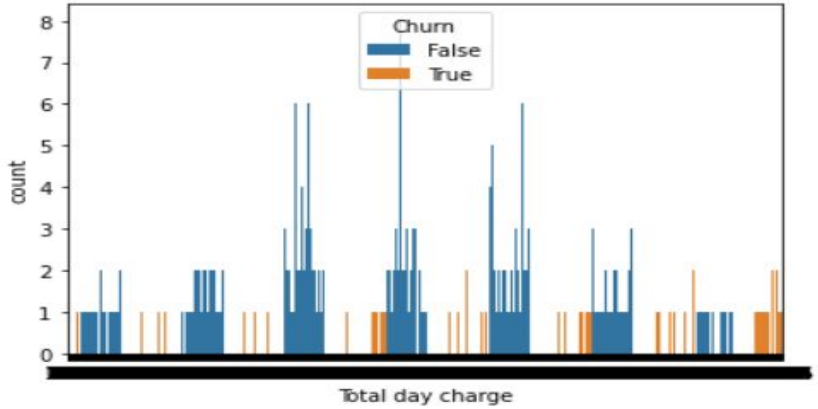


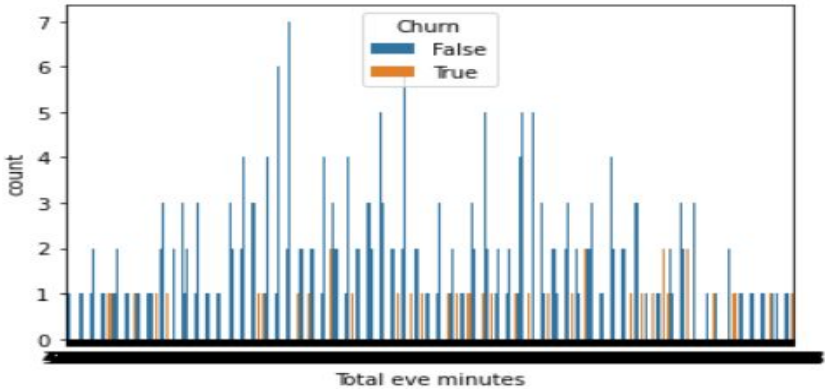


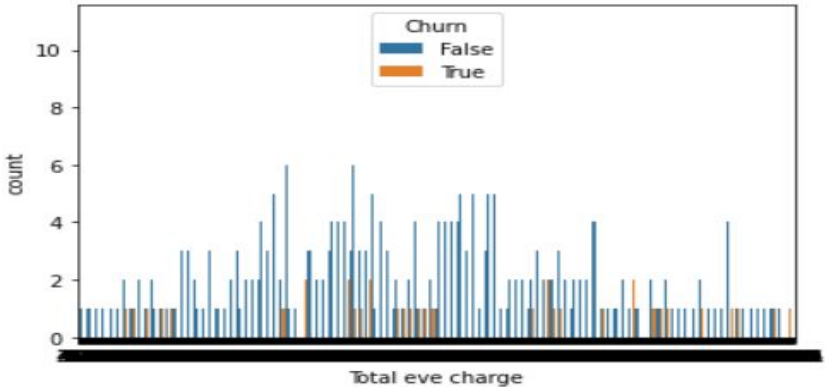


Numerical Variable Analysis (continued)

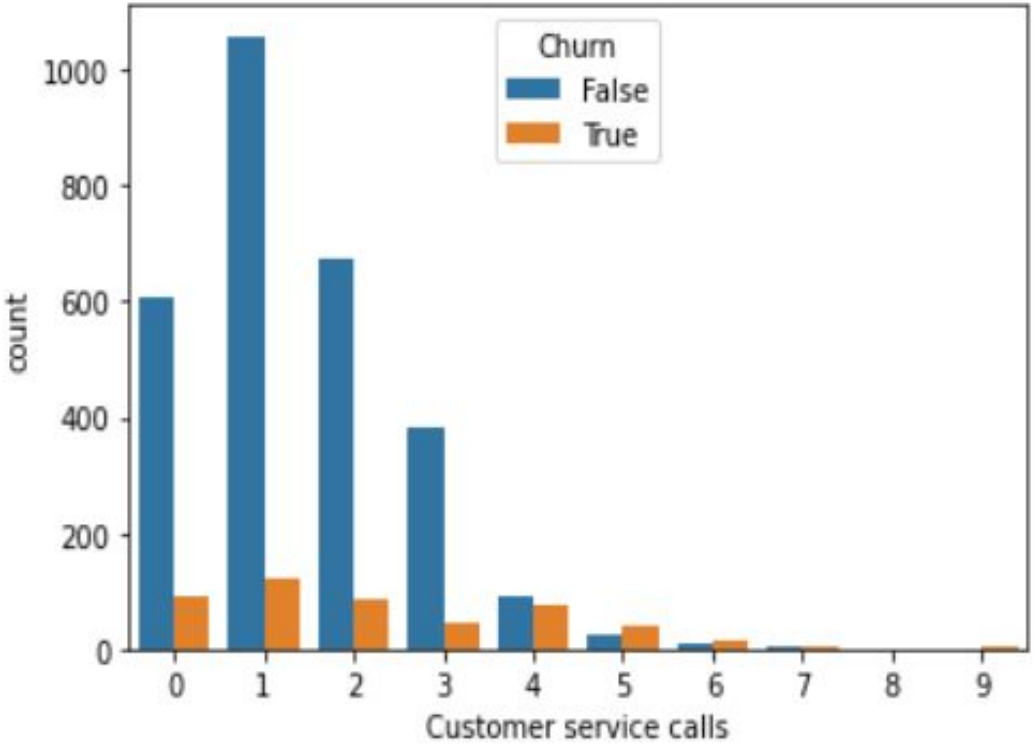






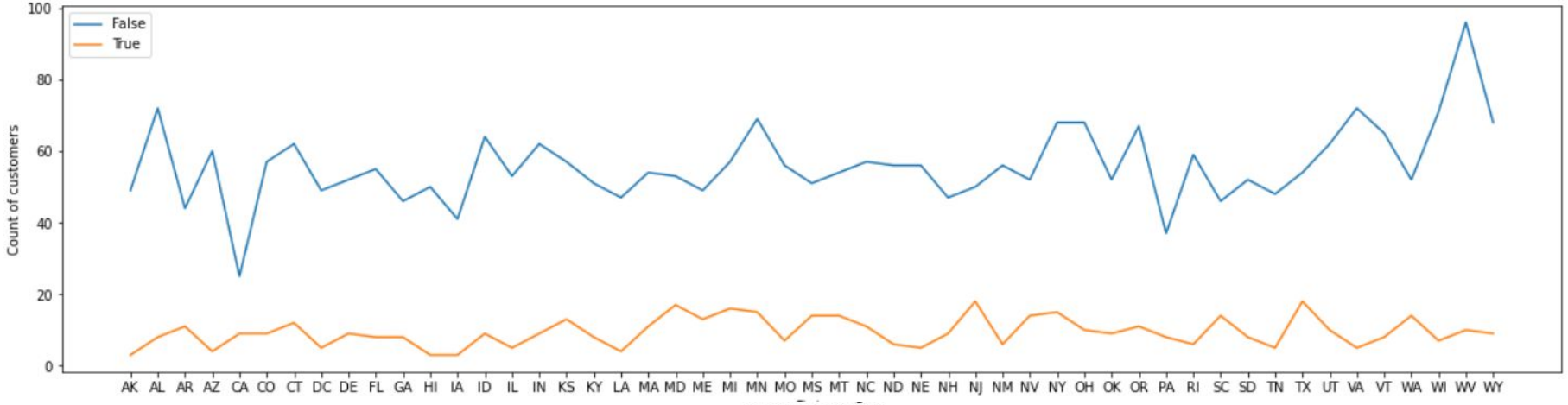


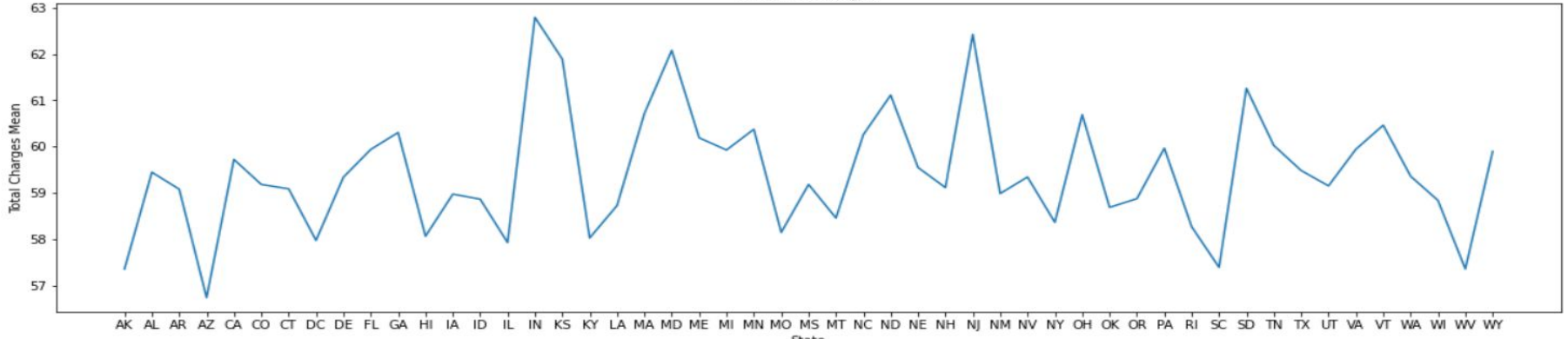
Numerical Variable Analysis (continued)



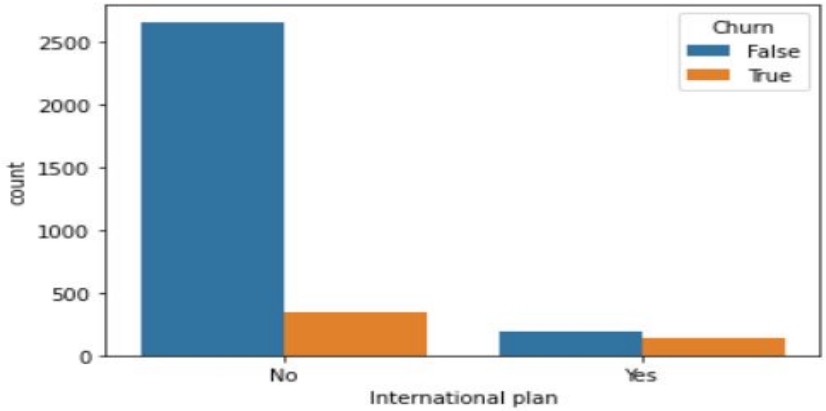
- Here we see barplots for all customers and how many calls they made both for false and true.
- We can observe after 3 calls, the bars have similar heights.
- This tells us the churn rate is increasing.
- Similar observations for previous 2 slides

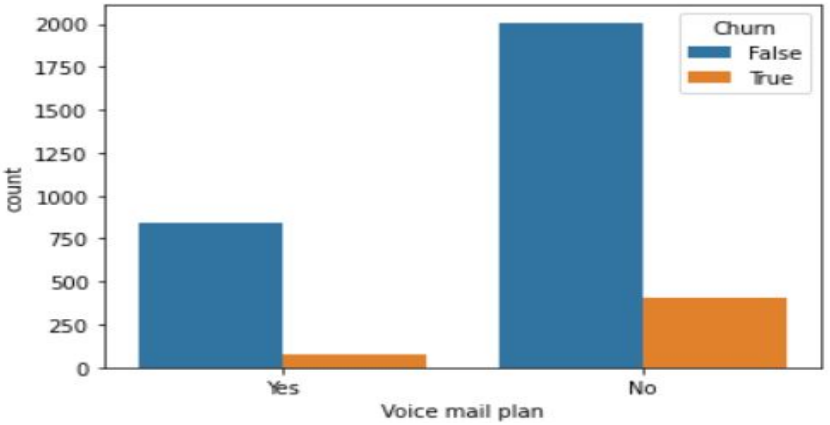
Categorical Variable Analysis

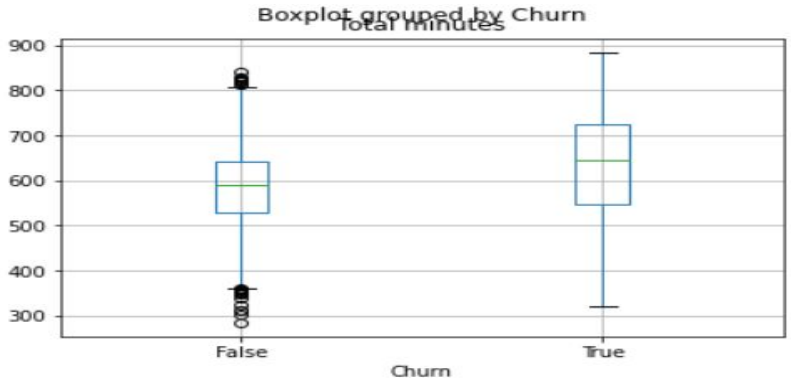


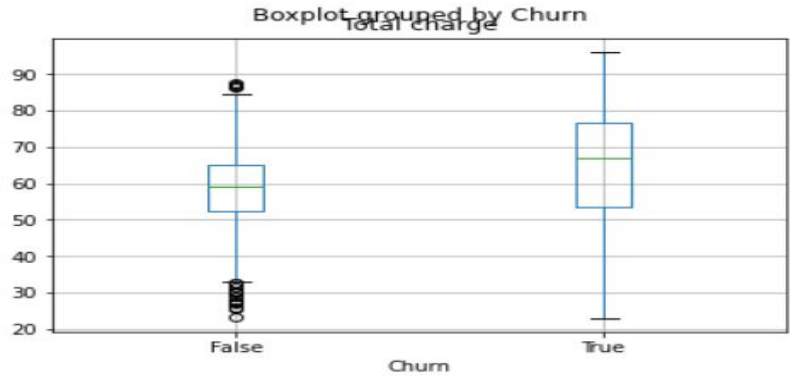


Categorical Variable Analysis (continued)









Conclusion

Insights drawn from the analysis are:

1. We saw minutes and charges are strongly correlated. Customers with high charges and minutes were more probable to churn.
2. Customers who made 4 or more service calls tend to churn more.
3. Churn rate was high for customers having international plan.
4. States having high overall charges had more customers who churned

Some recommendations:

1. We can try keeping a low correlation between the minutes spent and charges. Maybe with increasing minutes, decrease charge by some weights, because longer time spent shows good and uninterrupted services and if the services are good people will tend to talk more. Thus decreasing rates would help to retain more people. And if there are losses, it can be recovered because if customer stays, he will talk more. Its better than losing a customer completely due to high charges.
2. Keep a healthy relationships with the customers. People who are calling customer services more than often, shows that their problems are not solved or overlooked. Thus monitoring their problems and providing quick solutions can help to decrease churn rates.
3. There is clearly some issues with international plan which is causing dissatisfactions among customers. Might be due to heavy charges or poor connectivity and various other reasons. Therefore it would be better to re check the plan benefits and improve.
4. Although price may vary depending on the area but reducing the price fluctuations range from 6 to lesser can help in retaining customers