Capstone Project
Telecom Churn Analysis

Team members

Bhaskar Subanji
Ajinkya Dakhale
Harshjyot singh chawla

Let’s begin

1. Defining problem statement
2. Exploratory Data Analysis
3. Conclusion derived from EDA

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 cancelled the subscription. The idea of the project is to identify the factors responsible for the churn .

Data Summary

Churn_df : It contains information regarding the 3333 customer of **The Orange Telecom** (i.e.) state to which they belong, Account length, Area code, whether they have active international plan or not , minutes spent in call and many such information .

Cust_churn_df : It contains same information as of churn_df but only of those Customer who churned.

Cust_not_churn_df : It contains same information as of churn_df but only of those customer who did not churn.

Column Label

- State: States name(in code).
- Account Length: period for which the Account is active.
- Area Code: Area code having States
- International Plan: Yes: Means International Plan is subscribed and,
  No: Means the customer has not subscribed for international plan.
- Voice Mail Plan: Yes: Voice Mail Plan is subscribed,
  No: voice mail Plan is not subscribed by the Customer.
- Number vmail messages: Number of Voice Mail Messages
- Total day minutes: Total Number of Call Minutes Spent by customer during Morning hours.
- Total day calls: Total Number of Calls made by customer during the Morning hours.
- Total day charge: Total Charge for all Calls made by customer during the Morning hours.
- Total eve minutes: Total Number of Call Minutes Spent by customer during Evening hours.

- Total day charge: Total Charge for all Calls made by customer during the Morning hours.
- Total eve minutes: Total Number of Call Minutes Spent by customer during the Evening hours.
- Total eve calls : Total Number of Calls made by customer during the Evening hours.
- Total eve charge: Total Charge for all the Calls made by customer during the Evening hours.
- Total night minutes: Total Number of Call Minutes Spent by customer during the Night hours
- Total night calls: Total Number of Calls made by during the night hours.
- Total night charge: Total Charge for Calls made by customer during the Night hours.
- Total intl minutes: Total Number of Call Minutes spent by customer on international calls.
- Total intl calls: Total Number of International Calls made by customer.
- Total intl charge: Total charge for all the international calls made by customer.
- Customer service calls : Total Number of customer service calls made by customer.
- Churn : True : Churned customer
  False: means retained customer.

Column-wise Stats

We will analyze each column label of the dataframe with respect to churn and try find the relation between them.

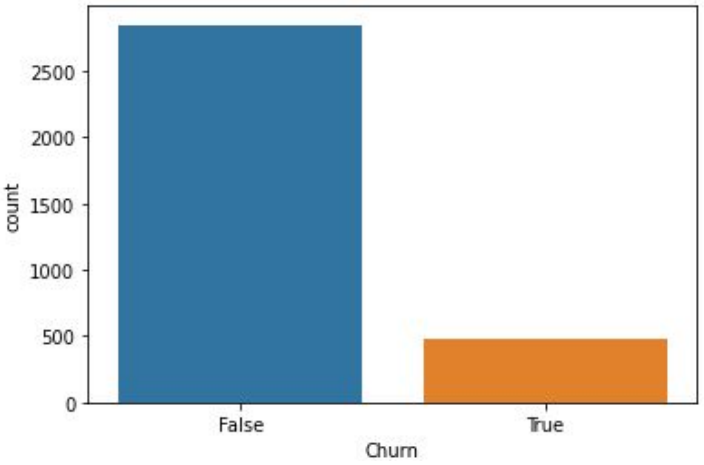
1. State Vs. Churn
2. Account length Vs. Churn
3. Area Vs. Churn
4. International plan Vs. churn

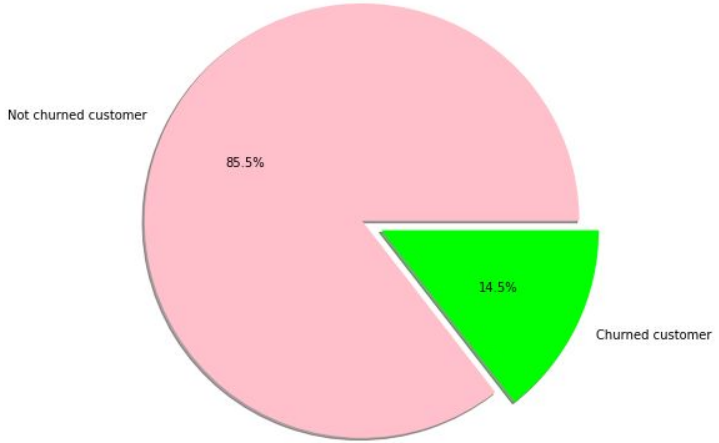
   and so on ……

EDA

Churn

The data consists of record of 3333 customers of the orange telecom customers. Out of which 483 customer churned and the remaining 2850 did not.

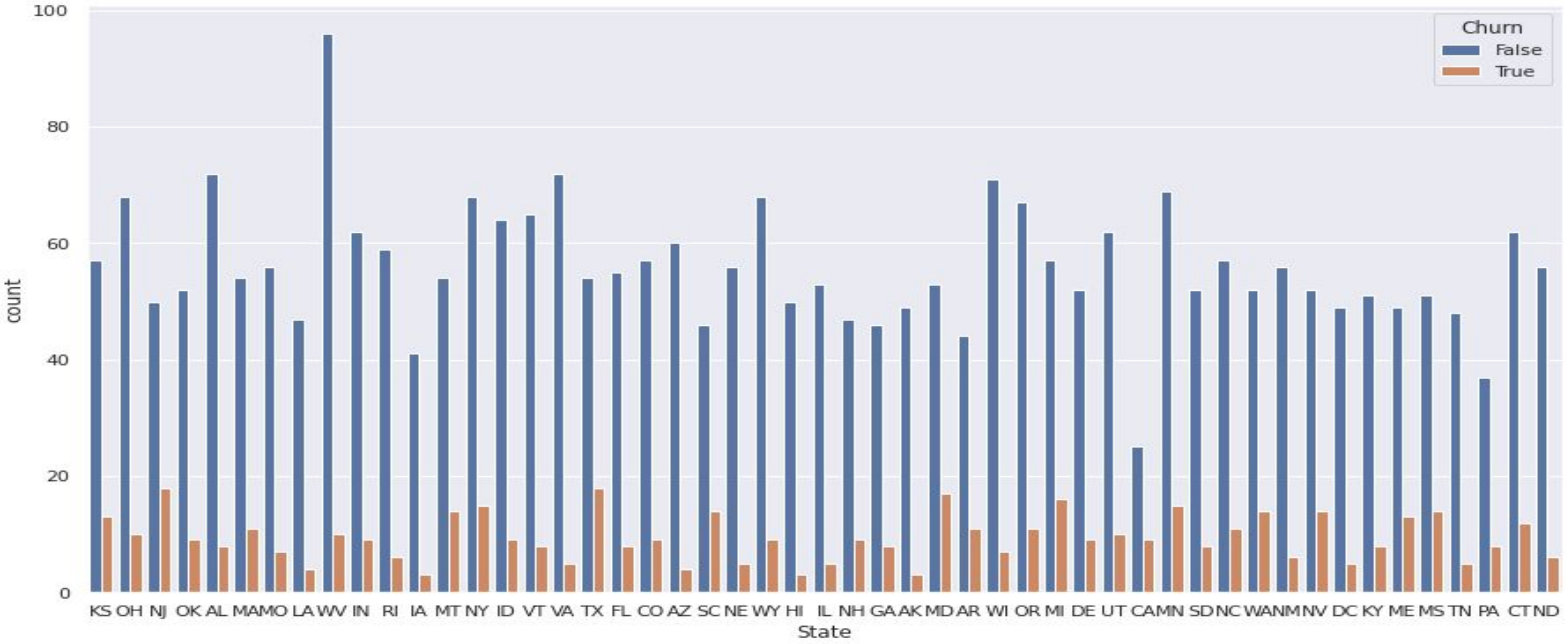




EDA

State wise churn

The data consists of us customers from 51 different states.



TOP 5 states with highest churn rate.

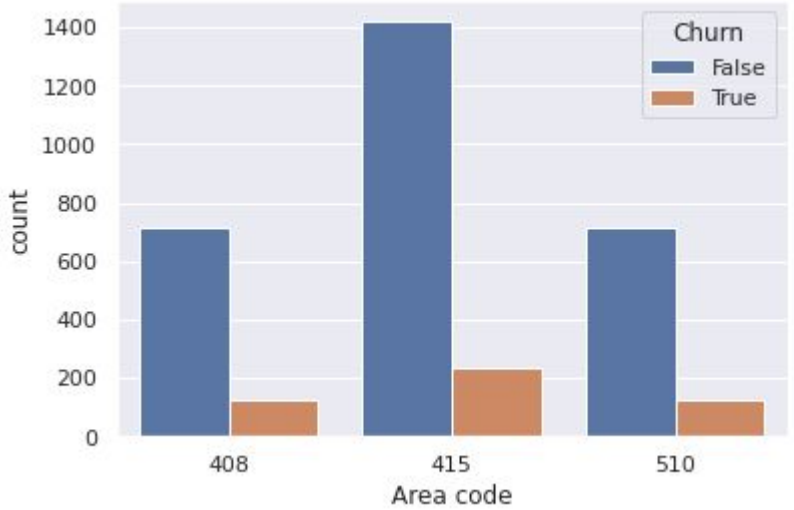
1. CA
2. NJ
3. TX
4. MD
5. SC

Churn State	False	True	%_Churn
AK	49	3	5.769231
AL	72	8	10.000000
AR	44	11	20.000000
AZ	60	4	6.250000
CA	25	9	26.470588
CO	57	9	13.636364
CT	62	12	16.216216
DC	49	5	9.259259
DE	52	9	14.754098
FL	55	8	12.698413
GA	46	8	14.814815
HI	50	3	5.660377
IA	41	3	6.818182
ID	64	9	12.328767
IL	53	5	8.620690
IN	62	9	12.676056
KS	57	13	18.571429
KY	51	8	13.559322
LA	47	4	7.843137
MA	54	11	16.923077
MD	53	17	24.285714
ME	49	13	20.967742
MI	57	16	21.917808
MN	69	15	17.857143
MO	56	7	11.111111
MS	51	14	21.538462
MT	54	14	20.588235
NC	57	11	16.176471
ND	56	6	9.677419
NE	56	5	8.196721
NH	47	9	16.071429
NJ	50	18	26.470588
NM	56	6	9.677419
NV	52	14	21.212121
NY	68	15	18.072289
OH	68	10	12.820513
OK	52	9	14.754098
OR	67	11	14.102564
PA	37	8	17.777778
RI	59	6	9.230769
SC	46	14	23.333333
SD	52	8	13.333333
TN	48	5	9.433962
TX	54	18	25.000000
UT	62	10	13.888889
VA	72	5	6.493506
VT	65	8	10.958904
WA	52	14	21.212121
WI	71	7	8.974359
WV	96	10	9.433962
WY	68	9	11.688312

EDA

Area wise churn rate :

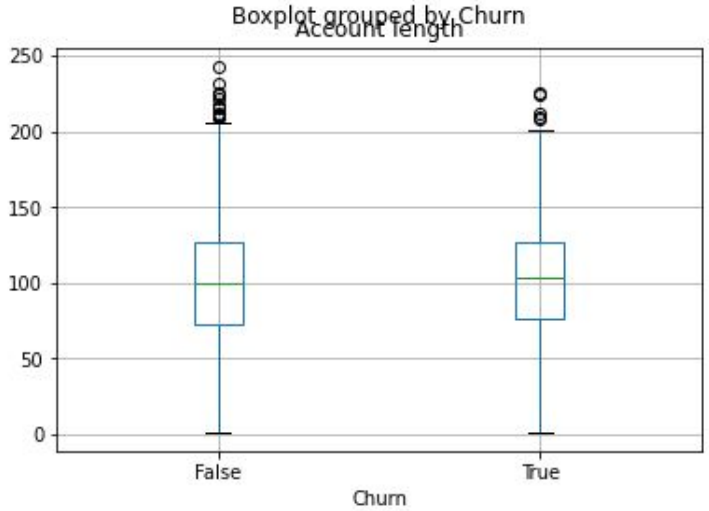
Given dataset consists of 3 unique Area code, under which different state fall.



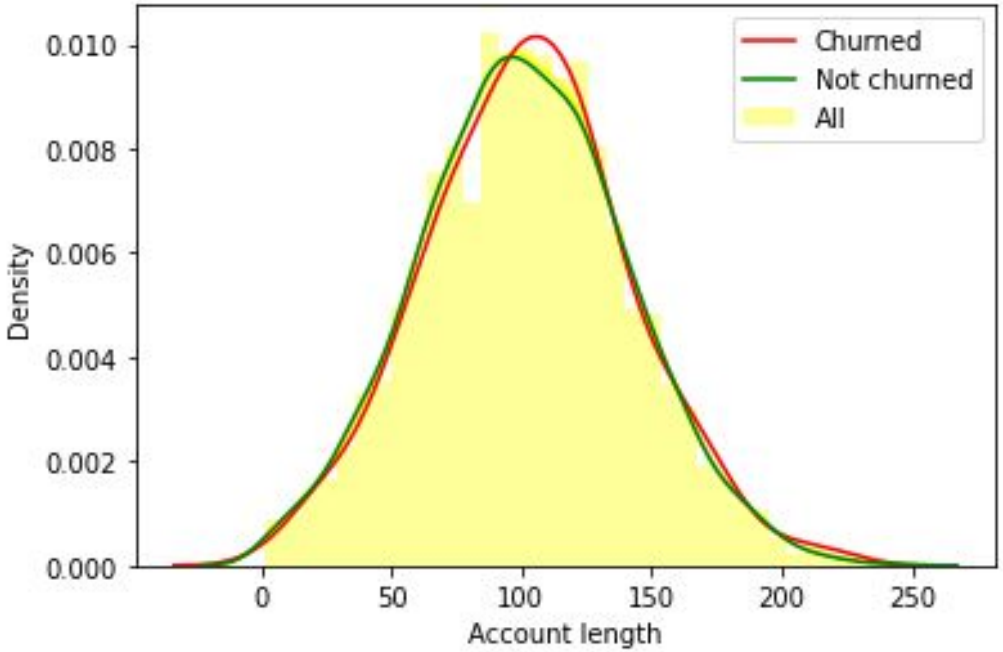
EDA

Account length

Account length is the time period between the day on the sim was activated and the day on which the data was recorded. From the boxplot plotted below, we can observe that the median(of account length for both churned and those who did not churn is approximately same.



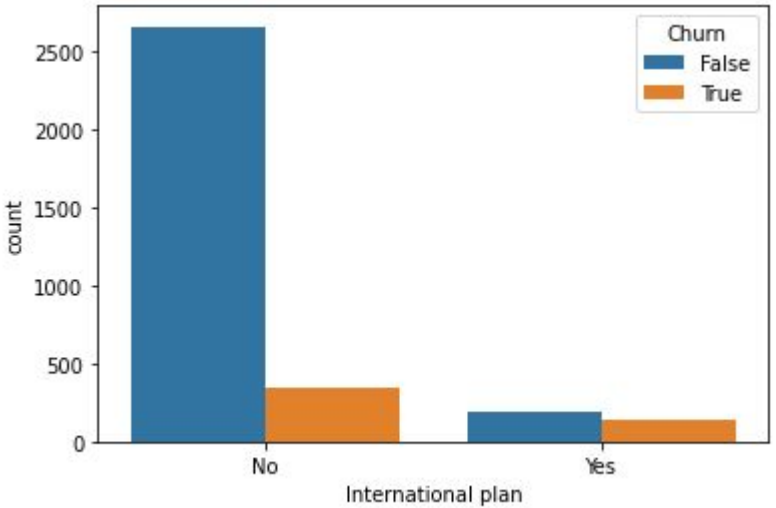
EDA

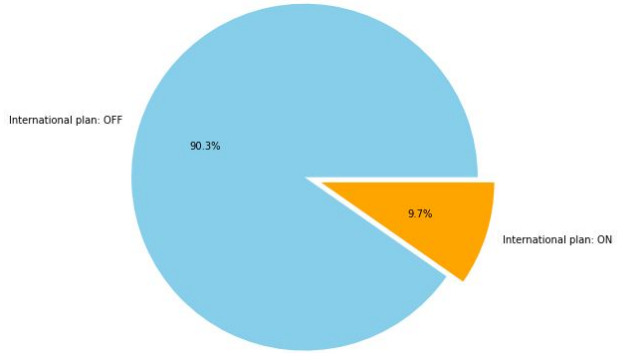

Churned
Not churned
All
Density
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
250
Account length

EDA

International plan

International plan allows customers to make call to the person from other country. It should be noted that the tariff of international are always high as compared to domestic, not many people opt for plan . Out of 3333 customers only 323 had active international plan.





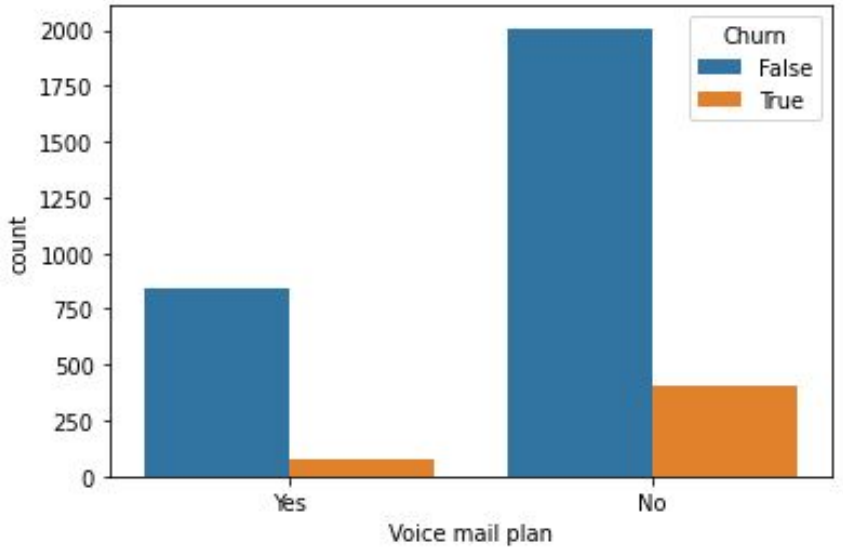
Churn	False	True	Churn_%
International plan			
No	2664	346	11.495017
Yes	186	137	42.414861

EDA

Voice mail plan

A **voicemail** is a computer-based system that allows users and subscribers to Exchange personal voice mail messages.

Churn	False	True	Churn_%
Voice mail plan			
No	2008	403	16.715056
Yes	842	80	8.676790



EDA

Number vmail messages

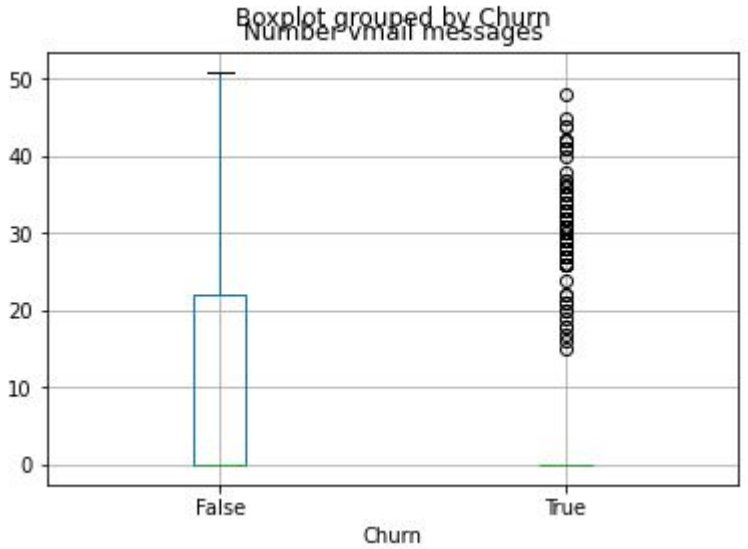
The number of voice mail exchanged by the customer.

Churned Customer

Count : 483.000000
mean :5.115942
std :11.860138
min :0.000000
25% :0.000000
50% :0.000000
75% :0.000000
max :48.000000

Customers who did not churn

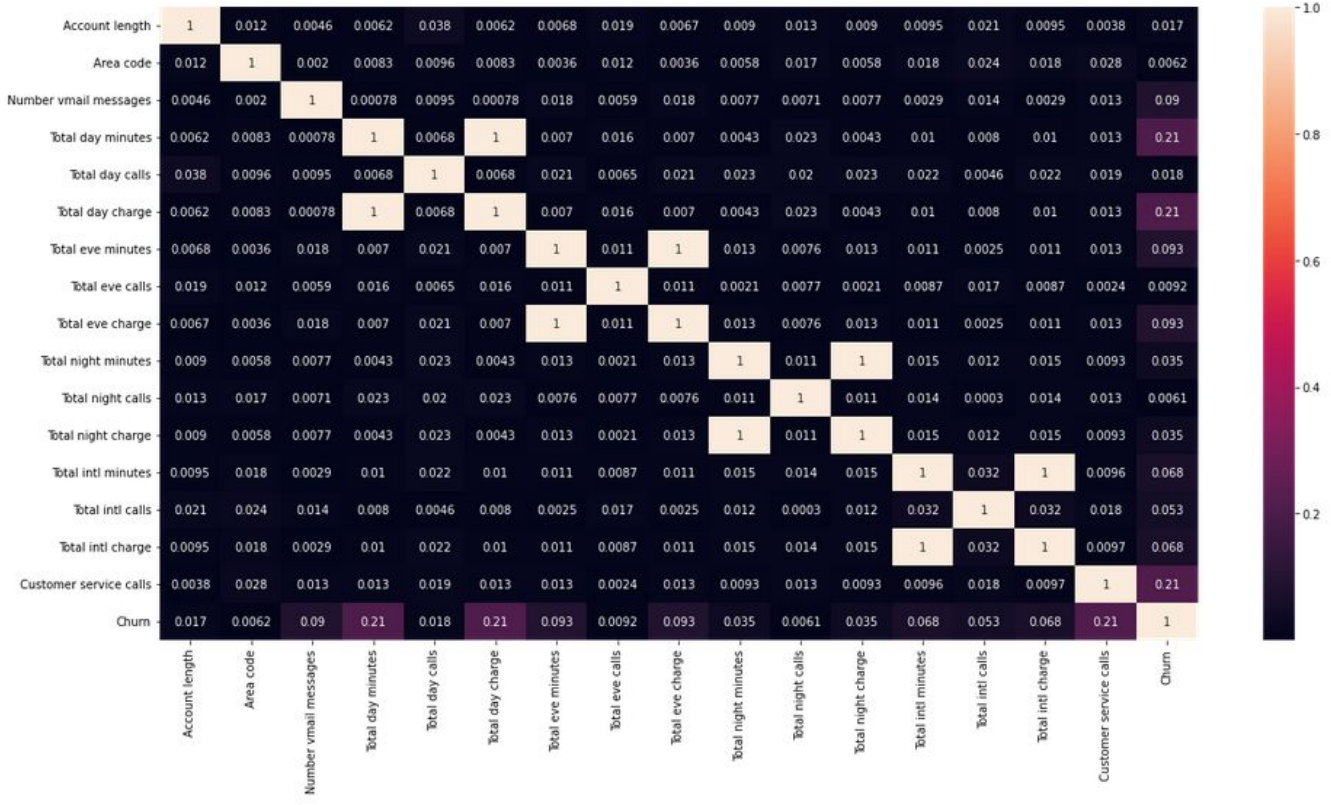
Count : 2850.000000
mean :8.604561
std :13.913125
min :0.000000
25% :0.000000
50% :0.000000
75% :22.000000
max :51.000000



EDA

Call Minutes and Call Charges

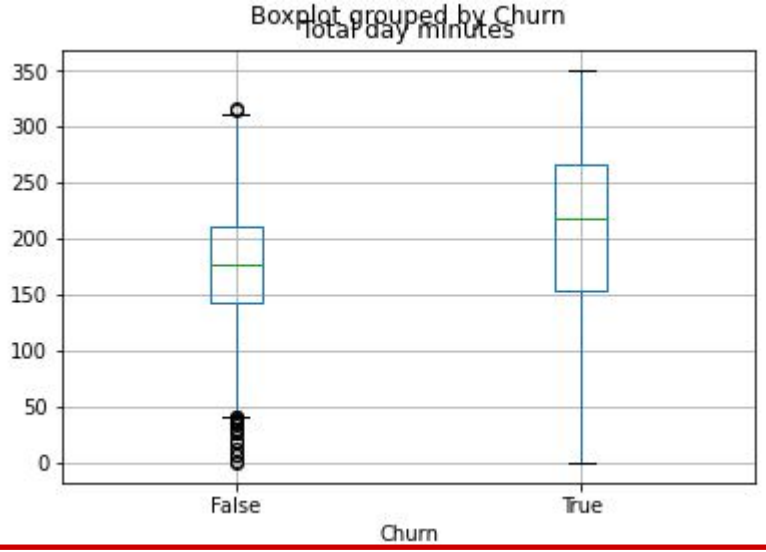
Below table shows the correlation between the columns of the dataframe.

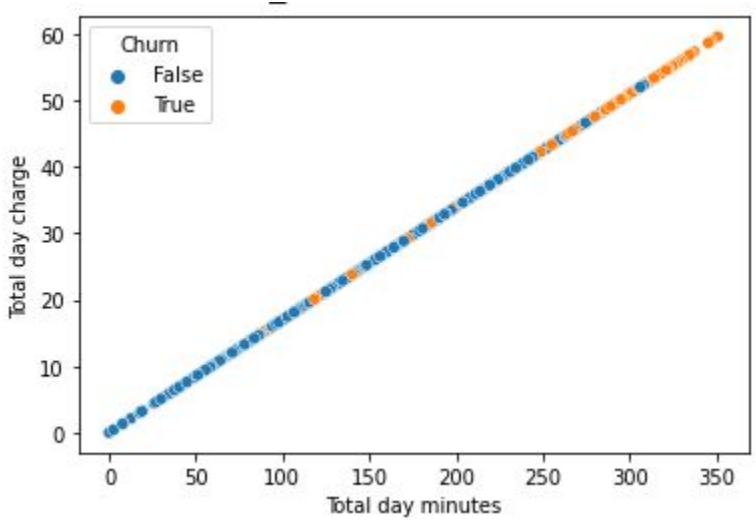


EDA

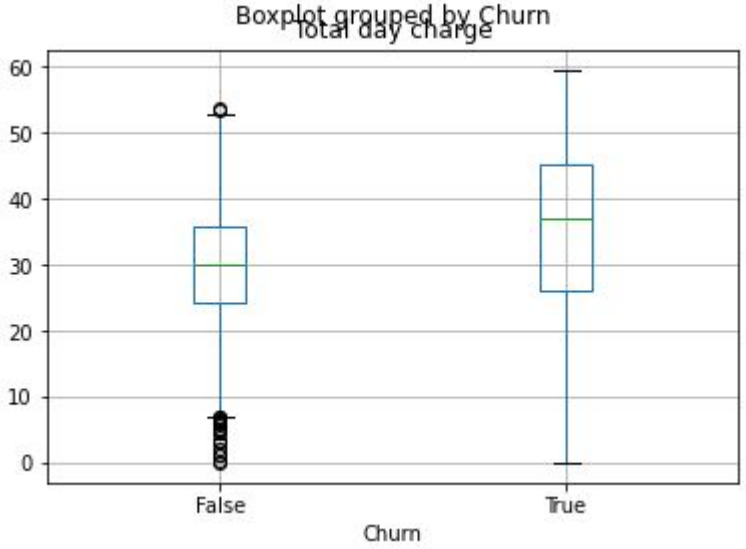
The given dataset consists of (number of calls made, duration of those calls In minutes, total charge charged for those calls) divided in different phase such as Daytime, Evening and night time.

For call made during day time:



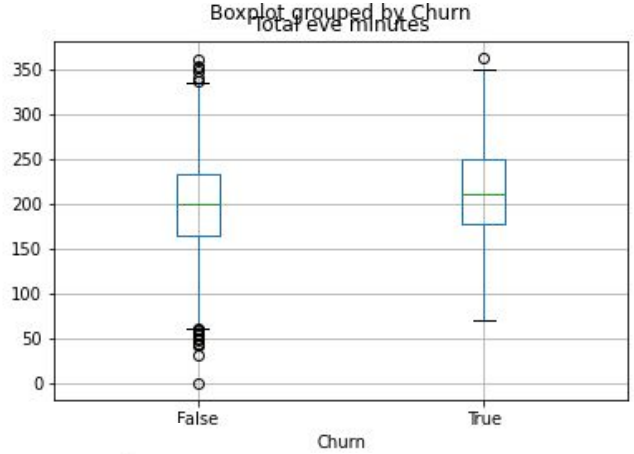


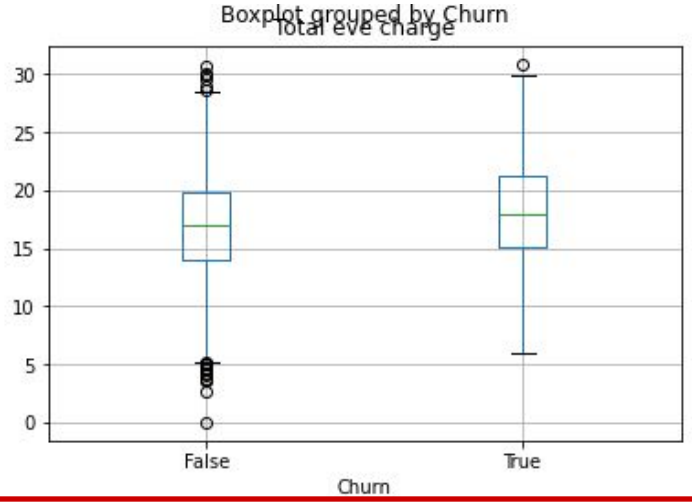
EDA

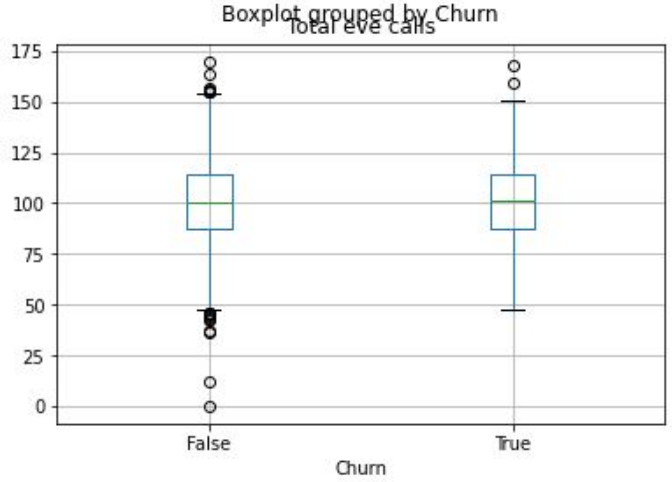

Boxplot grouped by Churn
total day charge
60
50
40
30
20
10
0
False
True
Churn

EDA

For evening calls:

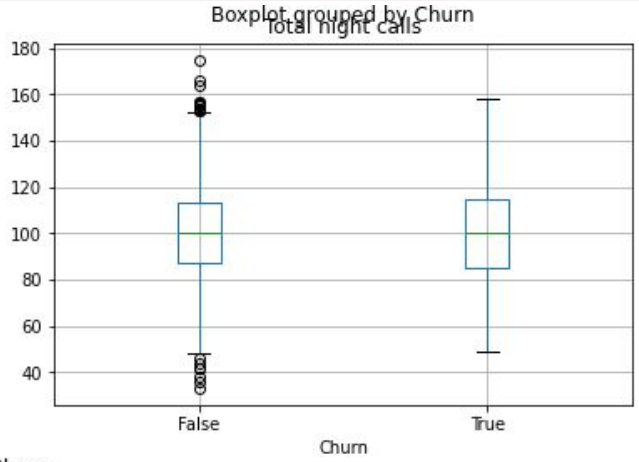


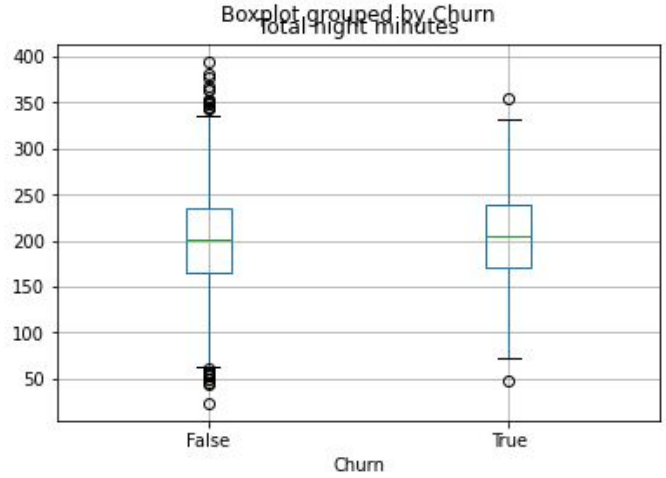


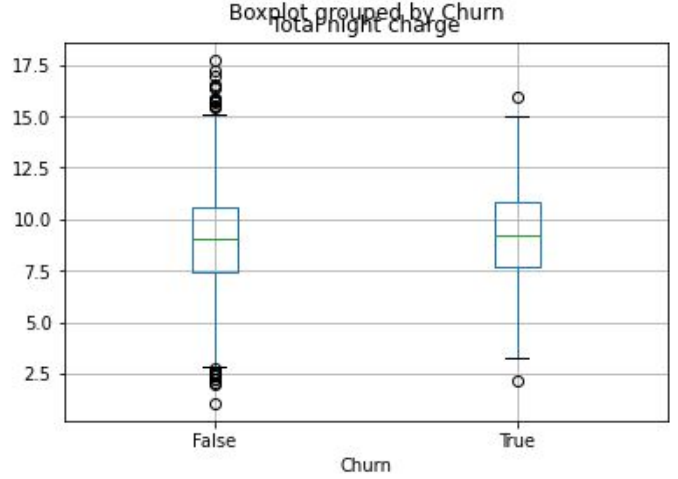


EDA

For Night calls :





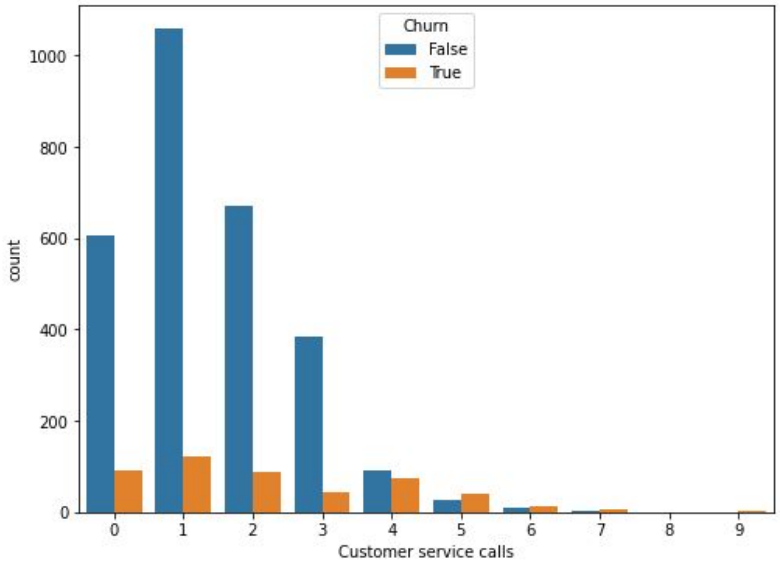


EDA

Customer service calls

This column present us with the data about the number of times a customer called customer service, mostly calls made to customer service are primarily for the reasons such as network issue, info regarding charged deducted etc.

Churn	False	True	Churn_%
Customer service calls			
0	605	92	13.199426
1	1059	122	10.330229
2	672	87	11.462451
3	385	44	10.256410
4	90	76	45.783133
5	26	40	60.606061
6	8	14	63.636364
7	4	5	55.555556
8	1	1	50.000000
9	0	2	100.000000



CONCLUSION

- Some states have higher churn rate than other, for which network issues could the reason because if the competitor company had low tariff for calls then most of the states would have shown the approximately same churn rate.
- Area and Account length has no relation with churn rate, hence this columns can be omitted or it can be said that the data is redundant.
- Customers with international plan ON has higher churn rate compared to customers with international plan OFF , this could be because the customer could be unhappy with the high tariff cost or network issues.
- It seems that most of the churned customer had 20 or more number of voicemails, the reason for churn might be that the quality of voicemails was not good and customers were not satisfied with the quality of voice mails.
- Customers with higher day call minutes has higher churn rate compared to other , could be because of the higher charges which is quite obvious, frequent caller might have found some other company offering low tariff .
- With other variables such as evening ,night calls no relation could be found.
- The churn rate increases as the call to the customer service increases. Customers who have called customer service three or fewer times have a markedly lower churn rate than that of customers who have called customer service four or more times. More customer service call means mostly likely the customer had some issue that needed to be resolved ,can which also be said as no of customer call = no of times problem faced by customer.

Recommendation

- They should improve in coverage area and solving network issues (both domestic as well as internat ional).
- Company can give discount or create a plan in which as the day call minutes crosses certain higher range the customer can be given some relaxation over tariff(i.e. the charge per min)
- lower the International plan tariff or provide customer with some discounts/offers.
- They can improve their customer service and provide better problem solution, also take their feedback and work on the feedback suggested by the customers.