Capstone Project
Telecom Churn Analysis

Team members

Anurag Taiskar

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.

Dataset First look

```
# Display first 10 indexes of the dataset
churn_df.head(10)
```

	State	Account length	Area code	International plan	Voice mail plan	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	Churn
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70	1	False
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70	1	False
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29	0	False
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	2	False
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	3	False
5	AL	118	510	Yes	No	0	223.4	98	37.98	220.6	101	18.75	203.9	118	9.18	6.3	6	1.70	0	False
6	MA	121	510	No	Yes	24	218.2	88	37.09	348.5	108	29.62	212.6	118	9.57	7.5	7	2.03	3	False
7	MO	147	415	Yes	No	0	157.0	79	26.69	103.1	94	8.76	211.8	96	9.53	7.1	6	1.92	0	False
8	LA	117	408	No	No	0	184.5	97	31.37	351.6	80	29.89	215.8	90	9.71	8.7	4	2.35	1	False
9	WV	141	415	Yes	Yes	37	258.6	84	43.96	222.0	111	18.87	326.4	97	14.69	11.2	5	3.02	0	False

```
# Display last 10 indexes of the dataset
churn_df.tail(10)
```

	State	Account length	Area code	International plan	Voice mail plan	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	Churn
3323	IN	117	415	No	No	0	118.4	126	20.13	249.3	97	21.19	227.0	56	10.22	13.6	3	3.67	5	True
3324	WV	159	415	No	No	0	169.8	114	28.87	197.7	105	16.80	193.7	82	8.72	11.6	4	3.13	1	False
3325	OH	78	408	No	No	0	193.4	99	32.88	116.9	88	9.94	243.3	109	10.95	9.3	4	2.51	2	False
3326	OH	96	415	No	No	0	106.6	128	18.12	284.8	87	24.21	178.9	92	8.05	14.9	7	4.02	1	False
3327	SC	79	415	No	No	0	134.7	98	22.90	189.7	68	16.12	221.4	128	9.96	11.8	5	3.19	2	False
3328	AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	126	18.32	279.1	83	12.56	9.9	6	2.67	2	False
3329	WV	68	415	No	No	0	231.1	57	39.29	153.4	55	13.04	191.3	123	8.61	9.6	4	2.59	3	False
3330	RI	28	510	No	No	0	180.8	109	30.74	288.8	58	24.55	191.9	91	8.64	14.1	6	3.81	2	False
3331	CT	184	510	Yes	No	0	213.8	105	36.35	159.6	84	13.57	139.2	137	6.26	5.0	10	1.35	2	False
3332	TN	74	415	No	Yes	25	234.4	113	39.85	265.9	82	22.60	241.4	77	10.86	13.7	4	3.70	0	False

3333 Rows and 20 Columns

Data Cleaning

dataset information

```
# Dataset Info
churn_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 20 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   State                   3333 non-null   object 
 1   Account length          3333 non-null   int64  
 2   Area code               3333 non-null   int64  
 3   International plan      3333 non-null   object 
 4   Voice mail plan         3333 non-null   object 
 5   Number vmail messages   3333 non-null   int64  
 6   Total day minutes       3333 non-null   float64
 7   Total day calls         3333 non-null   int64  
 8   Total day charge        3333 non-null   float64
 9   Total eve minutes       3333 non-null   float64
 10  Total eve calls         3333 non-null   int64  
 11  Total eve charge        3333 non-null   float64
 12  Total night minutes     3333 non-null   float64
 13  Total night calls       3333 non-null   int64  
 14  Total night charge      3333 non-null   float64
 15  Total intl minutes      3333 non-null   float64
 16  Total intl calls        3333 non-null   int64  
 17  Total intl charge       3333 non-null   float64
 18  Customer service calls  3333 non-null   int64  
 19  Churn                   3333 non-null   bool   
dtypes: bool(1), float64(8), int64(8), object(3)
memory usage: 498.1+ KB
```

missing value count

```
# Missing Values/Null Values Count
churn_df.isnull().sum()
```

```
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
```

Data Cleaning

Dataset Describe

```
# Dataset Describe
churn_df.describe()
```

	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
count	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000
mean	101.064806	437.182418	8.099010	179.775098	100.435644	30.562307	200.980348	100.114311	17.083540	200.872037	100.107711	9.039325	10.237294	4.479448	2.764581	1.562856
std	39.822106	42.371290	13.688365	54.467389	20.069084	9.259435	50.713844	19.922625	4.310668	50.573847	19.568609	2.275873	2.791840	2.461214	0.753773	1.315491
min	1.000000	408.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	23.200000	33.000000	1.040000	0.000000	0.000000	0.000000	0.000000
25%	74.000000	408.000000	0.000000	143.700000	87.000000	24.430000	166.600000	87.000000	14.160000	167.000000	87.000000	7.520000	8.500000	3.000000	2.300000	1.000000
50%	101.000000	415.000000	0.000000	179.400000	101.000000	30.500000	201.400000	100.000000	17.120000	201.200000	100.000000	9.050000	10.300000	4.000000	2.780000	1.000000
75%	127.000000	510.000000	20.000000	216.400000	114.000000	36.790000	235.300000	114.000000	20.000000	235.300000	113.000000	10.590000	12.100000	6.000000	3.270000	2.000000
max	243.000000	510.000000	51.000000	350.800000	165.000000	59.640000	363.700000	170.000000	30.910000	395.000000	175.000000	17.770000	20.000000	20.000000	5.400000	9.000000

Data Cleaning

Dataset Duplicates

```
[ ]  # Dataset Duplicate Value Count
     churn_df.duplicated().sum()

     0
```

Unique Values

```
# Check Unique Values for each variable.
for i in columns_list:
  print("No. of unique values in",i,"is",churn_df[i].nunique())
```

```
No. of unique values in State is 51
No. of unique values in Account length is 212
No. of unique values in Area code is 3
No. of unique values in International plan is 2
No. of unique values in Voice mail plan is 2
No. of unique values in Number vmail messages is 46
No. of unique values in Total day minutes is 1667
No. of unique values in Total day calls is 119
No. of unique values in Total day charge is 1667
No. of unique values in Total eve minutes is 1611
No. of unique values in Total eve calls is 123
No. of unique values in Total eve charge is 1440
No. of unique values in Total night minutes is 1591
No. of unique values in Total night calls is 120
No. of unique values in Total night charge is 933
No. of unique values in Total intl minutes is 162
No. of unique values in Total intl calls is 21
No. of unique values in Total intl charge is 162
No. of unique values in Customer service calls is 10
No. of unique values in Churn is 2
```

Data Wrangling

Churn by each columns :-

```
[12] # Churn by each columns
churn_count1 = churn_df.groupby('Churn')
churn_count1.mean()
```

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

calculate charges for each kind of min :-

```
[13] #calculate charges for each kind of mins:
day_min_charges = churn_df['Total day charge'].mean()/churn_df['Total day minutes'].mean()
eve_min_charges = churn_df['Total eve charge'].mean()/churn_df['Total eve minutes'].mean()
night_min_charges = churn_df['Total night charge'].mean()/churn_df['Total night minutes'].mean()
int_min_charges= churn_df['Total intl charge'].mean()/churn_df['Total intl minutes'].mean()
print(f' Day_min_charge: {day_min_charges}')
print(f'eve_min_charge: {eve_min_charges}')
print(f'night_min_charge: {night_min_charges}')
print(f'Total intl charge: {int_min_charges}')
```

```
 Day_min_charge: 0.17000300739130672
eve_min_charge: 0.08500104871485778
night_min_charge: 0.04500041448440008
Total intl charge: 0.2700500279887107
```

Churn by state wise :-

states with highest churn rate.

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

```
# percentage of customer churned out of total customer statewise
State_data = pd.crosstab(churn_df["State"],churn_df["Churn"])
State_data['Churn_%'] = State_data.apply(lambda x : x[1]*100/(x[0]+x[1]),
                                         axis = 1)
print(State_data)
```

Churn	False	True	Churn_%
State			
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

Churn by international palm :-

```
[15] # percentage of customer who churned with respect to total no of cusmtomers witn international plan ON & OFF
     International_plan = pd.crosstab(churn_df["International plan"],churn_df["Churn"])
     International_plan['Churn_%'] = International_plan.apply(lambda x : x[1]*100/(x[0]+x[1]),axis = 1)
     print(International_plan)
```

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

- Those who has international plan their churn rate is high.
- Around 42 % customers are churned

Churn by area codewise :-

```
[16] # churn rate area code wise
     churn_area = pd.DataFrame(churn_df.groupby('Area code')['Churn'].value_counts())
     churn_area['churn rate'] = round(churn_area*100/len(churn_df),2)
     churn_area
```

		Churn	churn rate
Area code	Churn		
408	False	716	21.48
	True	122	3.66
415	False	1419	42.57
	True	236	7.08
510	False	715	21.45
	True	125	3.75

Churn by voice mail :-

```
[17] #churn rate for customer with and without voice mail paln
     Voice_mail = pd.crosstab(churn_df["Voice mail plan"],churn_df["Churn"])
     Voice_mail['Churn_%'] = Voice_mail.apply(lambda x : x[1]*100/(x[0]+x[1]),axis = 1)
     print(Voice_mail)
```

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

Column-wise Stats

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

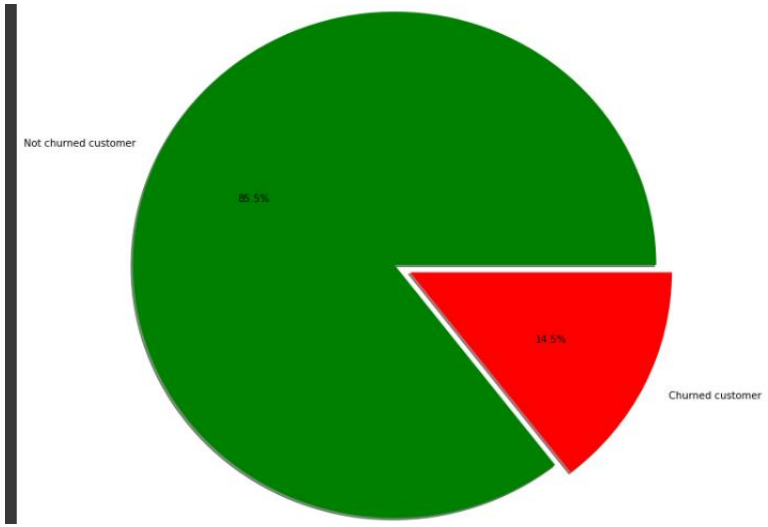
1. State Vs. Churn
2. Account length Vs. Churn
3. Area Vs. Churn
4. International plan Vs. churn

   and so on ……

EDA

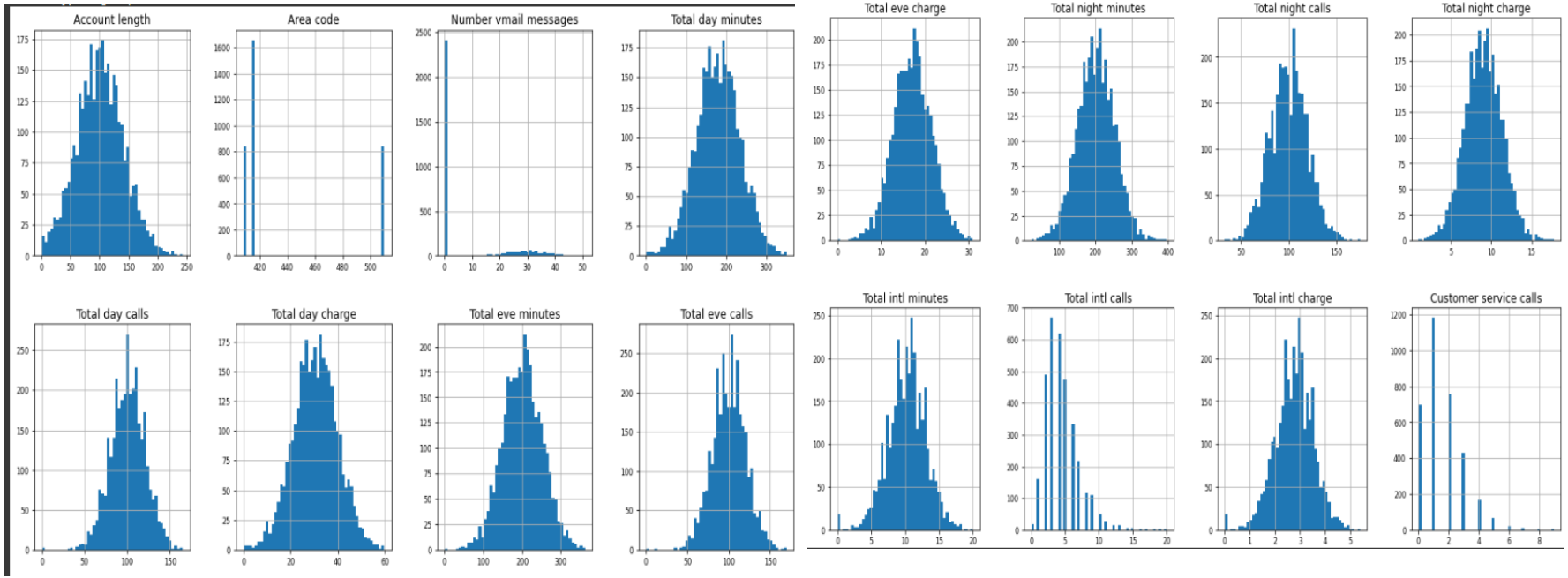
Churn

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.



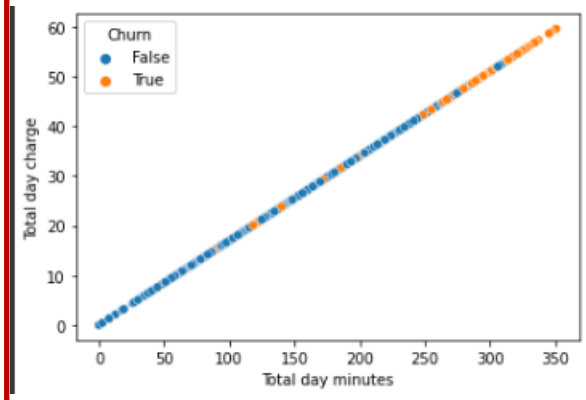
EDA

Churn by each variable

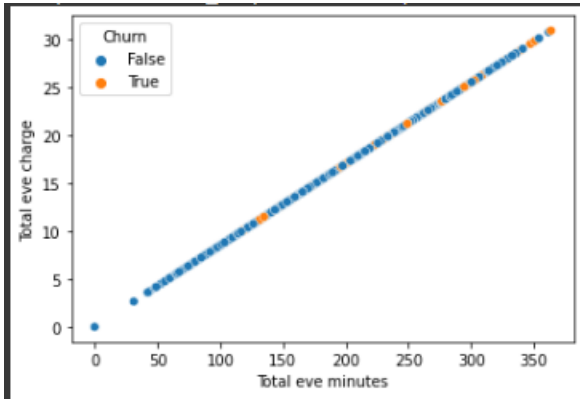


EDA

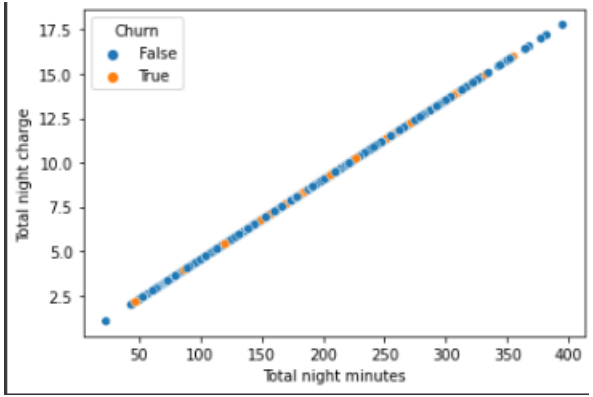
Total day minutes
vs
Total day charge



Total eve minutes
vs
total eve charge



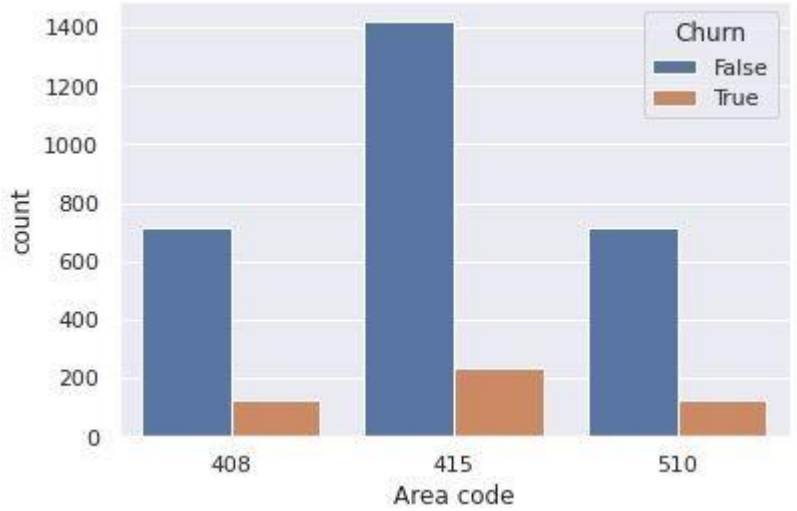
Total night minutes
vs
total night charge



EDA

Area wise churn rate :

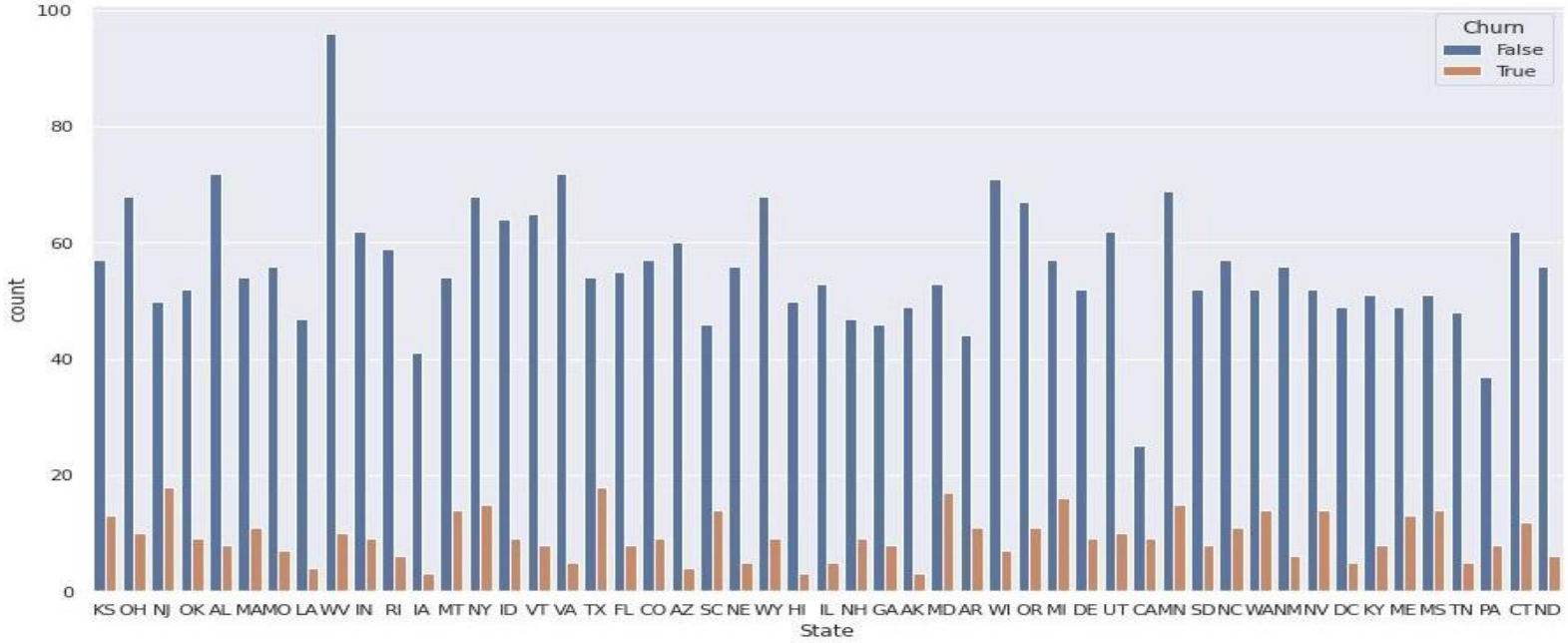
Given dataset consists of 3 unique Area code, under which different state fall.



EDA

State wise churn

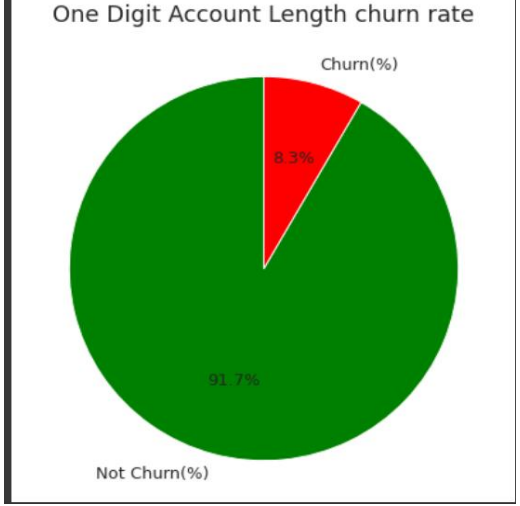
The data consists of us customers from 51 different states.



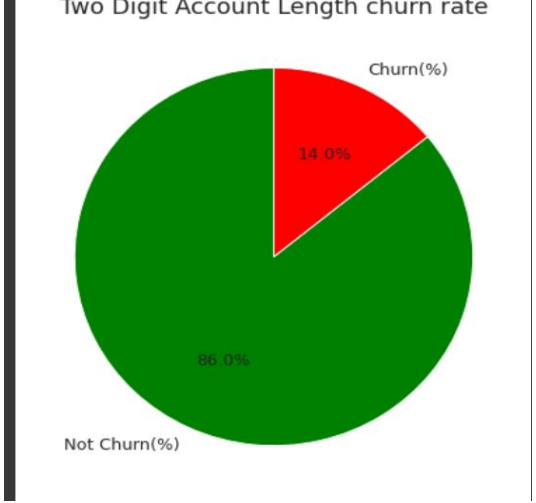
EDA

Account Length Pie Chart

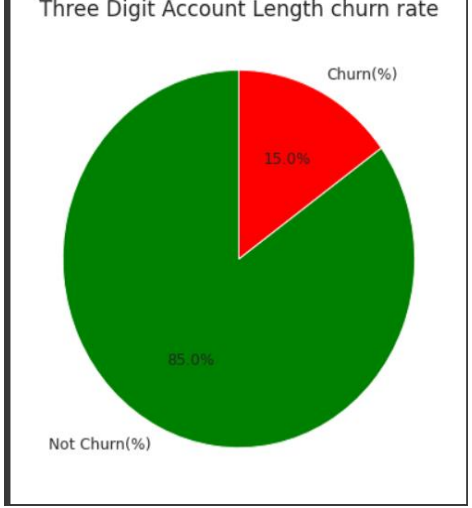
```
Churn
False    22
True      2
dtype: int64
```




```
Churn
False    1378
True      225
dtype: int64
```

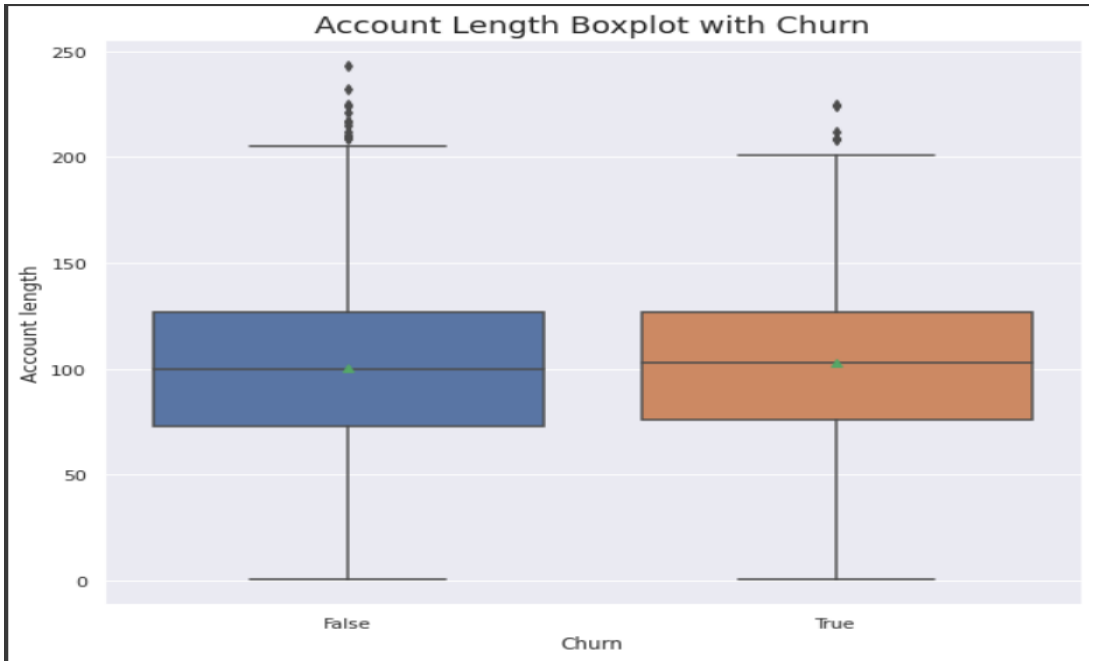



```
Churn
False    1450
True      256
dtype: int64
```



EDA

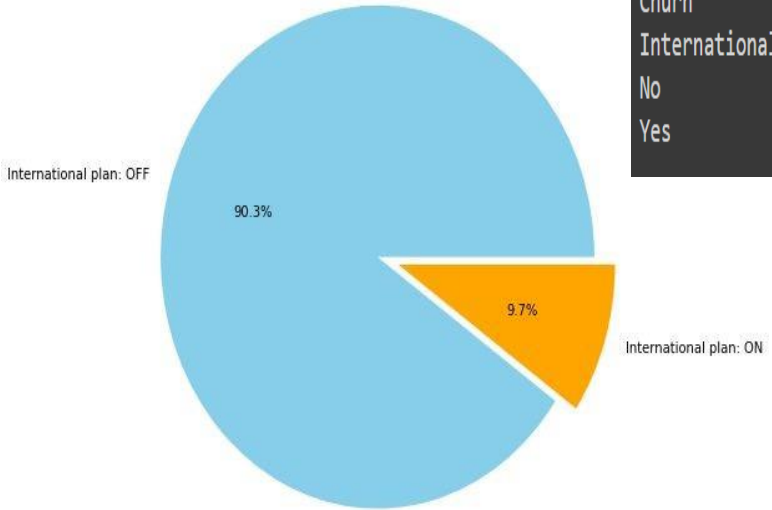
Account Length with Churn Box Plot



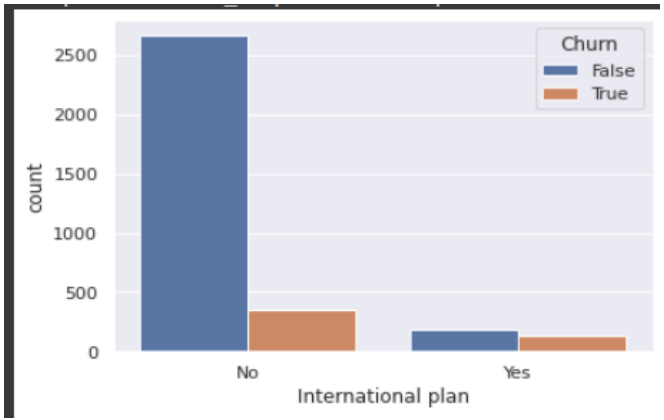
EDA

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

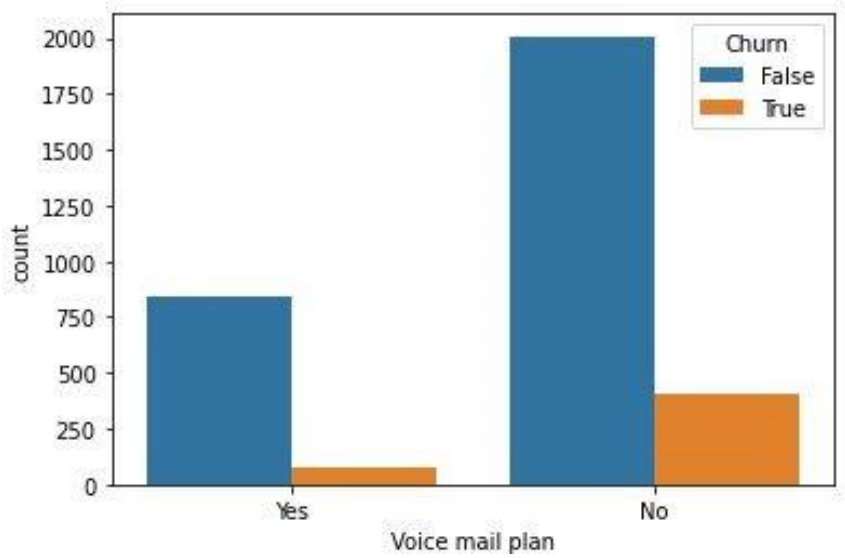


EDA

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

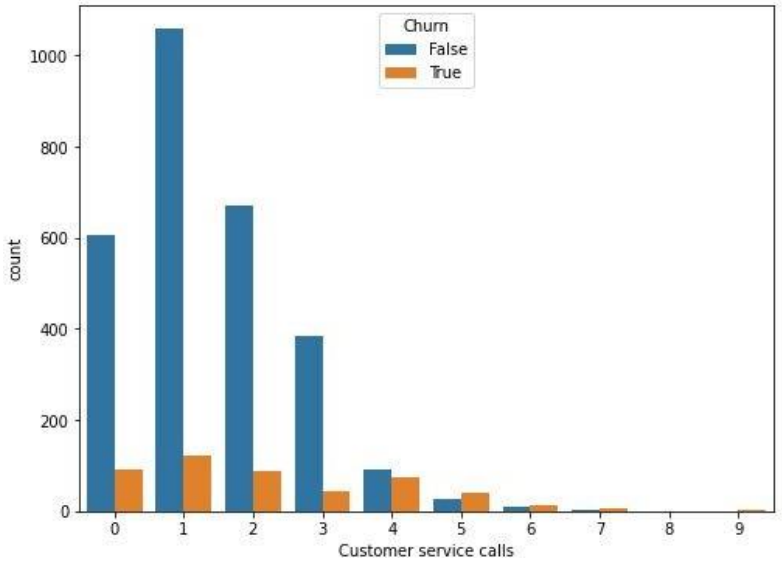


EDA

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.

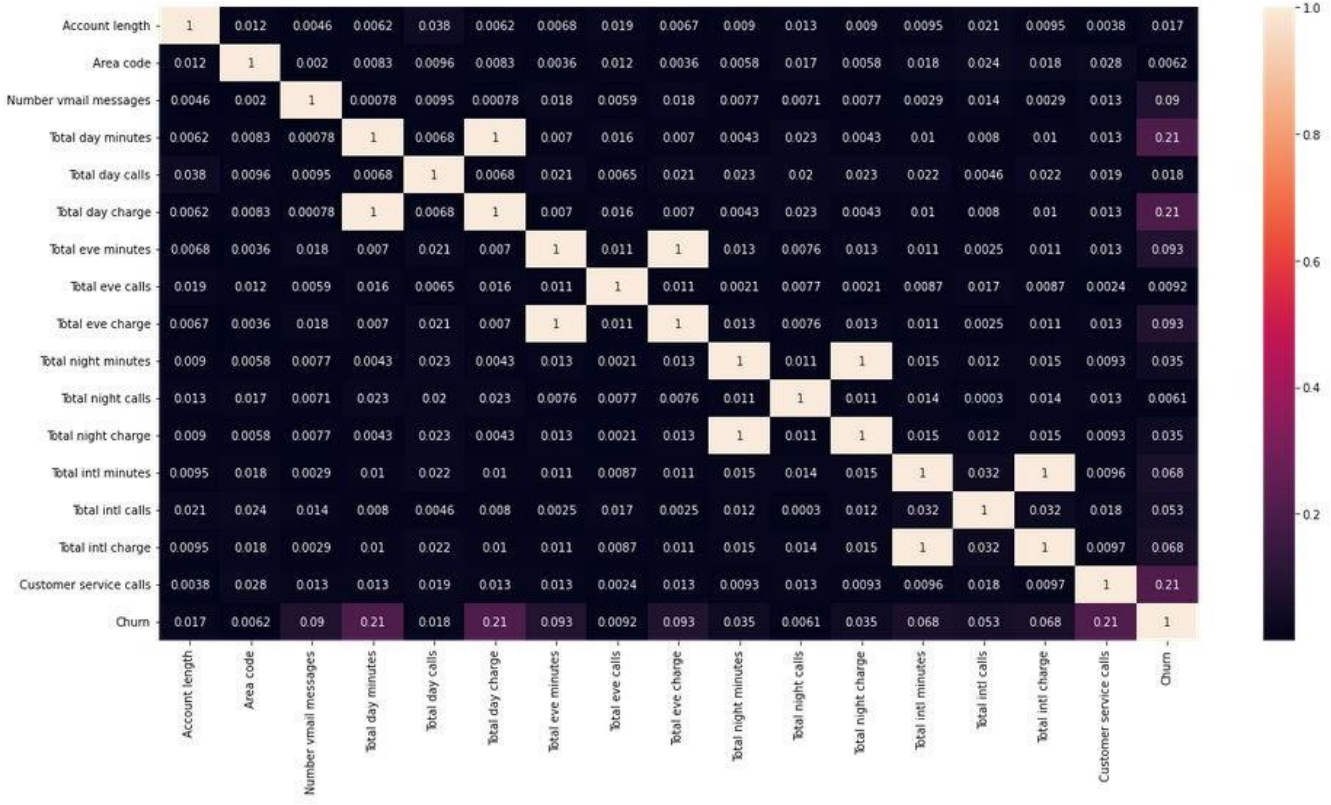
Customer service calls	0	1	2	3	4	5	6	7	8	9	All
Churn											
False	605	1059	672	385	90	26	8	4	1	0	2850
True	92	122	87	44	76	40	14	5	1	2	483
All	697	1181	759	429	166	66	22	9	2	2	3333



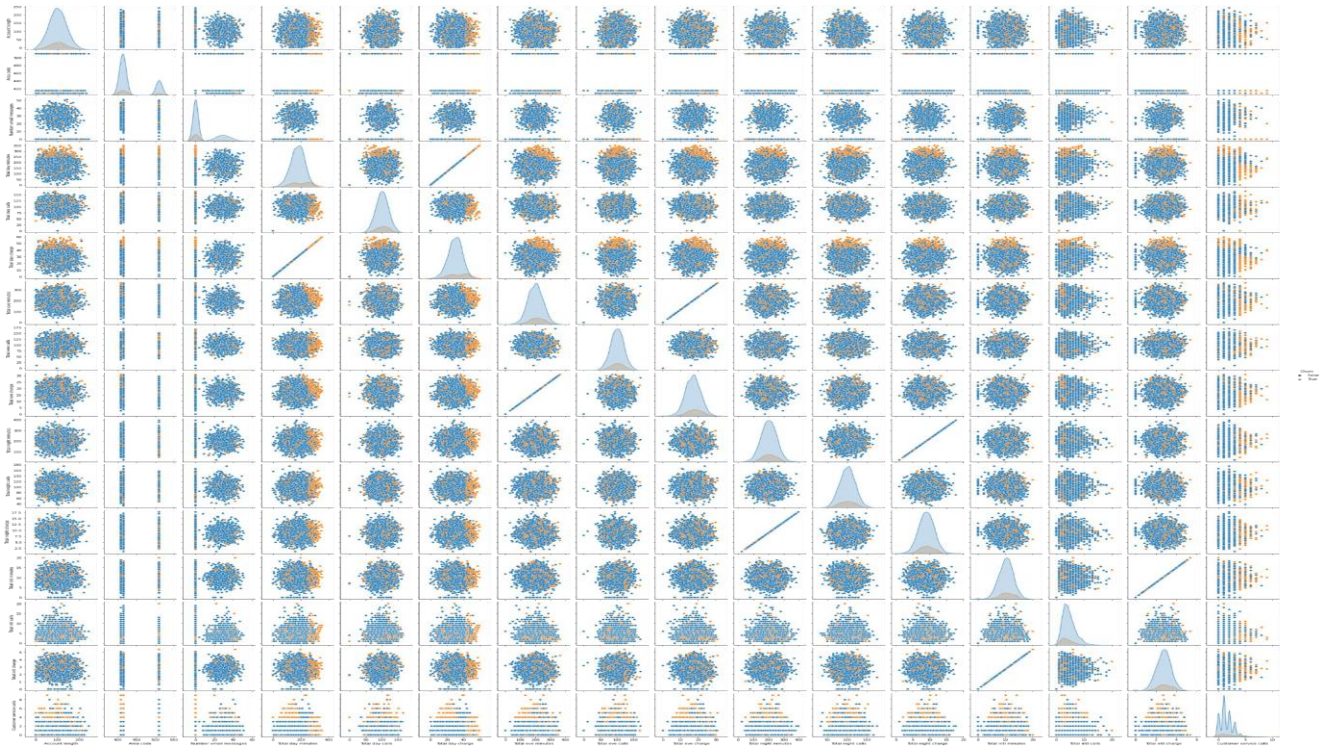
EDA

Call Minutes and Call Charges

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



Pair plot



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.

Recommendation

- Modify International Plan as the charge is same as normal one.
- Be proactive with communication.
- Ask for feedback often.
- Periodically throw Offers to retain customers.
- Look at the customers facing problem in the most churning states.
- Lean into best customers.
- Regular Server Maintenance.
- Solving Poor Network Connectivity Issue.
- Define a roadmap for new customers.
- Analyze churn when it happens.