

Capstone Project

Telecom Churn Analysis

From:-
Prateek Gupta

Content

- **Understanding the business problem**
- **Objective**
- **Data Overview**
- **Exploratory data analysis (EDA)**
- **Recommendations**
- **Conclusion**

Understanding the business problem

- Customer churn in the telecom industry poses one of the most significant risks of loss of revenue.
- Since the cost of acquiring new customers is higher than the cost of retaining them, retaining customer is a key.
- Company targets these customers with special plans and offers and if churning predictions are inaccurate, incurs huge loss.
- With competitive customer plans and quality of service provided, the rate of customer churn increased.

Objective

- To predict customer churn
- Highlighting the main variables/factors influencing customer churn.

Business constraints:

- Improving service quality and offering discounted prices (if required) without affecting the profits.

Dataset Description

Numerical Variables

- Account length
- 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

Categorical Variables

- State
- Area code
- International plan
- Voice mail plan

Decision Variable

- Churn

Exploratory Data Analysis

First 5 Rows of the dataset to get the overview

```
[5] # inspecting loaded data
df.head()
```

	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

Shape of Dataset and Unique Values

We have 3333 row and 20 columns

```
#shape of the dataset  
df.shape  
(3333, 20)
```

And there are 51 states, 3 area codes

```
# inspecting unique values of all the columns  
df.nunique()  
  
State          51  
Account length 212  
Area code      3  
International plan 2  
Voice mail plan 2  
Number vmail messages 46  
Total day minutes 1667  
Total day calls 119  
Total day charge 1667  
Total eve minutes 1611  
Total eve calls 123  
Total eve charge 1440  
Total night minutes 1591  
Total night calls 120  
Total night charge 933  
Total intl minutes 162  
Total intl calls 21  
Total intl charge 162  
Customer service calls 10  
Churn          2  
dtype: int64
```

Churn Percentage

Out of 3333 customers 2850 not churned or taking the services of company and 483 customers churned or stopped taking companies services.

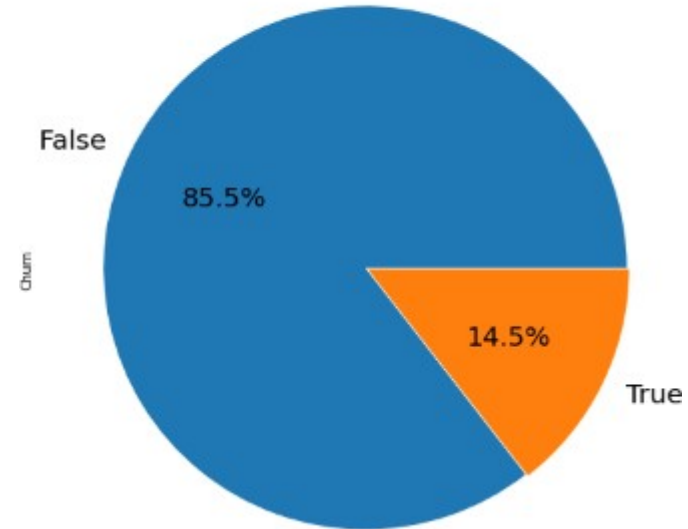
Percentage of churn

Not Churned: 85.5%

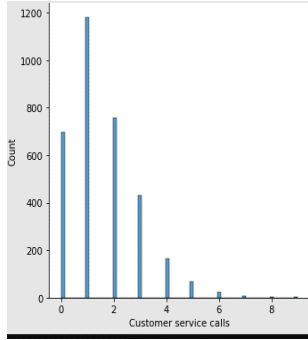
Churned: 14.5%

So from the pie chart we can see that 14.5% of customer who were taking companies services have churned or in other words stopped taking companies services this issue needs to be addressed as customer churn impacts company's growth and profits negatively.

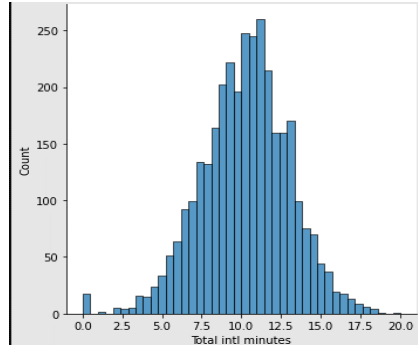
Total Customer Churn Percentage



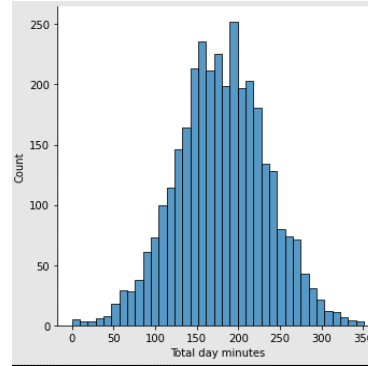
Distribution Plot



Customer Service Call



Total International Minutes



Total Day Minutes

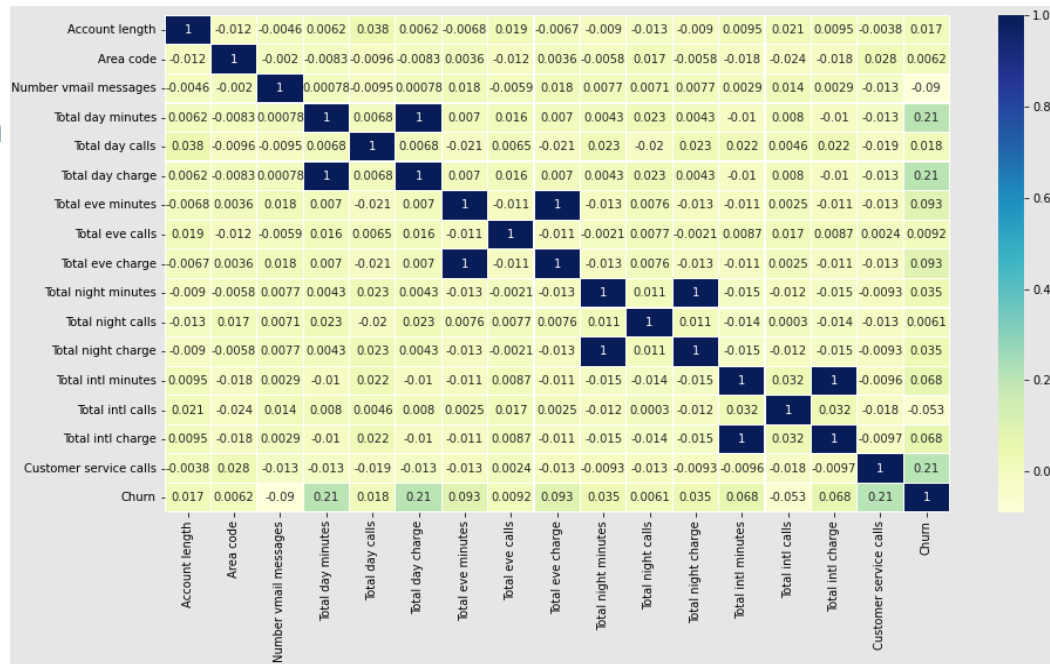
Distribution plot of all columns were made and it was observed that data is normally distributed

Correlation

Using heat map we can see the correlation between Churn and other independent variables

From the heat map we can see correlation following variables have high correlation with churn-

- Total day minutes
- Total day charge
- Customer service calls

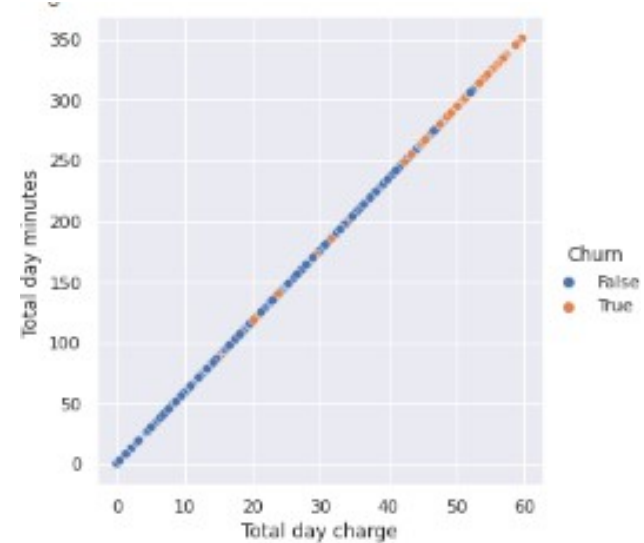
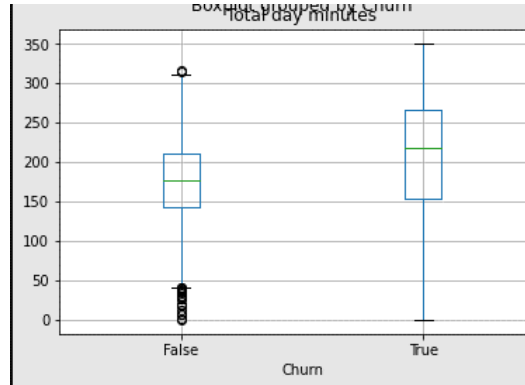


Relation B/w Total day minutes and churn

As we can see, the graph between Total day minutes and Total day charge is linear and as total day charge or total day minutes increases, the churn increases.

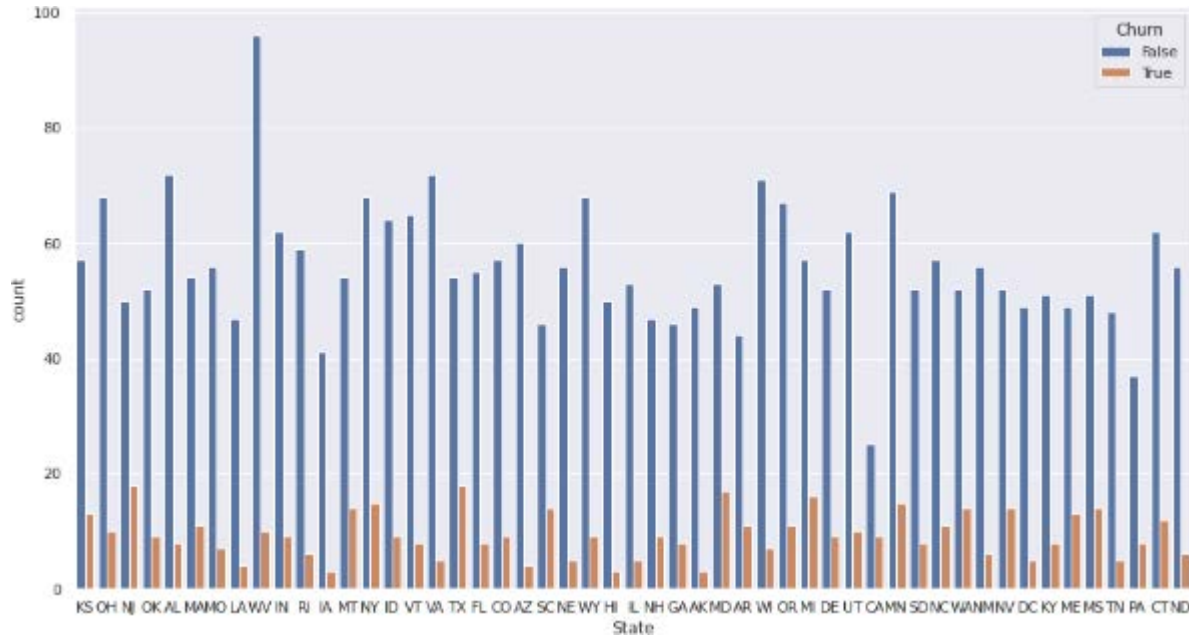
From the graph, we can see, if total day minutes is more than 250 minutes or the day charge is above 40 dollars the customer churn has been increased strongly.

From box plot, Total Day Minutes mean is different for churned and non churned users.



State wise churn

To know if there is more churn in some states we plot the graph between churn and states



States with high percentage churn

Top 5 States with more then 20% customer churn

States with highest churn rate:

1. State NJ (New Jersey) has highest churn rate.
2. State CA (California) is having 2nd highest churn rate.
3. TX, MD and SC are also in the list of high churn states.

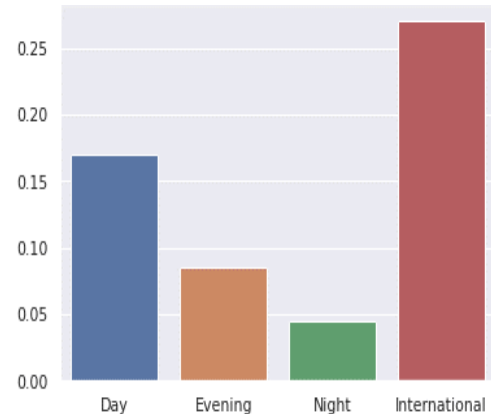
High churn of customers in these states indicates that either these states have high competition (other companies offering cheaper rates or better network coverage) or in these states company have poor network coverage.

Churn	False	True	Percentage_churn
State			
NJ	50	18	26.470588
CA	25	9	26.470588
TX	54	18	25.000000
MD	53	17	24.285714
SC	46	14	23.333333

International Calls Vs Churn

Data shows that customer who opted for international plan churned 200% more than customer without international plan.

42.41% of customers who have international calling plan have churned and 11.49 % of customers churned who don't have international calling plan.

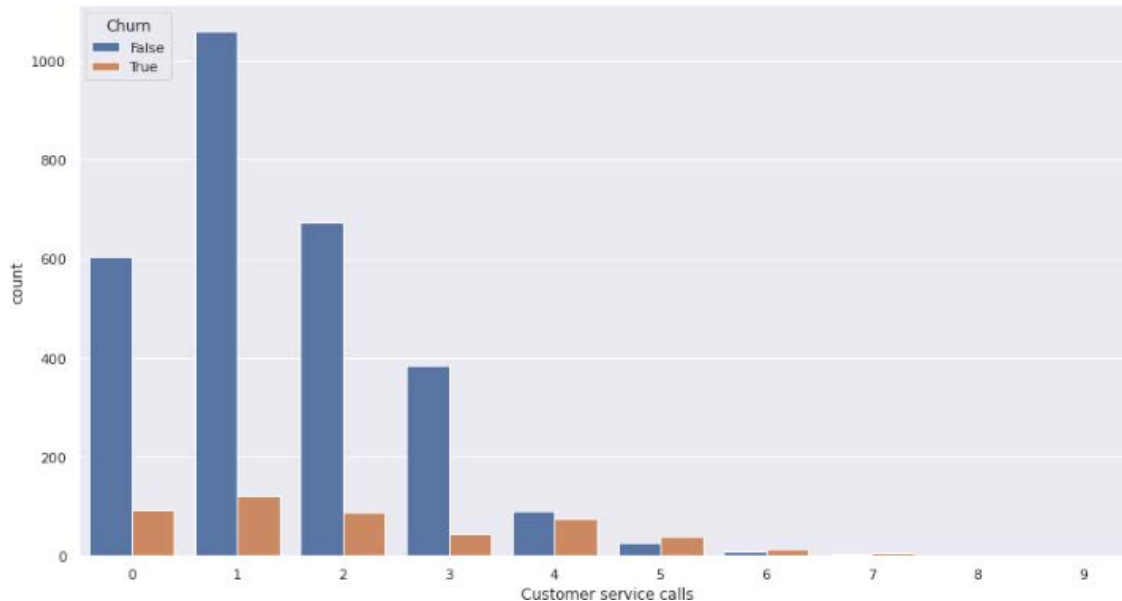


	Churn		Percentage_churn
International plan	False	True	
Yes	186	137	42.414861
No	2664	346	11.495017

Customer service calls vs. churn

To know, if there is relation between Number of Customer service calls and churn ,we plot the graph between churn and Customer service calls.

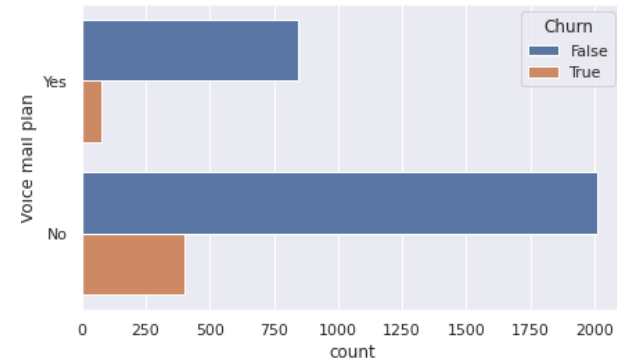
This shows, churning rate is rapidly increasing if more than 3 Customer service calls are made.



Voice Mail Plan vs. Churn

- To know, if there is relation between Voice Mail Plan and Churn ,we plot the graph between churn and Voice mail plan.
- We can see that Customers who are not subscribed to Voice mail plan are churning more.

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



Recommendations

- ❖ As per observation, churn is high when the customer's Total Calling minutes per day is high. There may be some competitors who are offering cheaper calling rates so company should come up with optimal calling rate or launch a calling plan targeting customer with high total calling minutes.
- ❖ In states (NJ, CA, TX, MD, SC) where churn is high company needs to inspect if there is low network penetration or competitor are offering cheaper prices.
- ❖ Customer with international calling plan churn 200% more than normal customers so this needs to be addressed with optimal international calling rate.
- ❖ Customer service quality needs to be improved if there are more than 3 customer services calls the customer churn increases.
- ❖ Company can focus on selling more Voice mail plans as churning is less with this feature.

Thank You