Capstone Project:
Exploratory Data Analysis
Project Title:
Telecom Churn Analysis

Anupam Ghosh

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 canceled the subscription.**

❑ **We have to explore and analyze the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.**

Objective of our Project:

❏ Churn Prediction is essentially predicting which clients are most likely to cancel a subscription i.e. 'leave a company' based on their usage of the service.

❏ So the main objective of our project is to detect the key factors responsible for customer churn and find out the ways to ensure customer retention.

Data Summary:

- The Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features). This dataset contain 3333 rows and 20 columns. There is no null value present in our dataset.

Variable names:

- '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'

Variable Breakdown:

- **STATE**: 51 Unique States
- **Account Length**: Length of The Account
- **Area Code:** 3 Unique area code are 415, 408, 510
- **International Plan:** Yes Indicate International Plan is Present and No Indicates no subscription for International Plan
- **Voice Mail Plan:** Yes Indicates Voice Mail Plan is Present and No Indicates no subscription for Voice Mail Plan
- **Number vmail messages:** Number of Voice Mail Messages ranging from 0 to 50
- **Total day minutes:** Total Number of Minutes Spent By Customers in Morning

Variable Breakdown:

- **Total day calls:** Total Number of Calls made by Customer in Morning.
- **Total day charge:** Total Charge to the Customers in Morning.
- **Total eve minutes:** Total Number of Minutes Spent By Customers in Evening
- **Total eve calls:** Total Number of Calls made by Customer in Evening.
- **Total eve charge:** Total Charge to the Customers in Morning.
- **Total night minutes:** Total Number of Minutes Spent By Customers in the Night.
- **Total night calls:** Total Number of Calls made by Customer in Night.
- **Total night charge:** Total Charge to the Customers in Night.
- **Churn:** True Indicate customer cancelled the subscription and False Indicates customer not canceled the subscription.

Defining the Target Variable:

❑ Here the column 'Churn' is our Target variable.

❑ If 'Churn' is True, that indicates the customer had cancelled the subscription, otherwise the customer did not cancel the subscription.
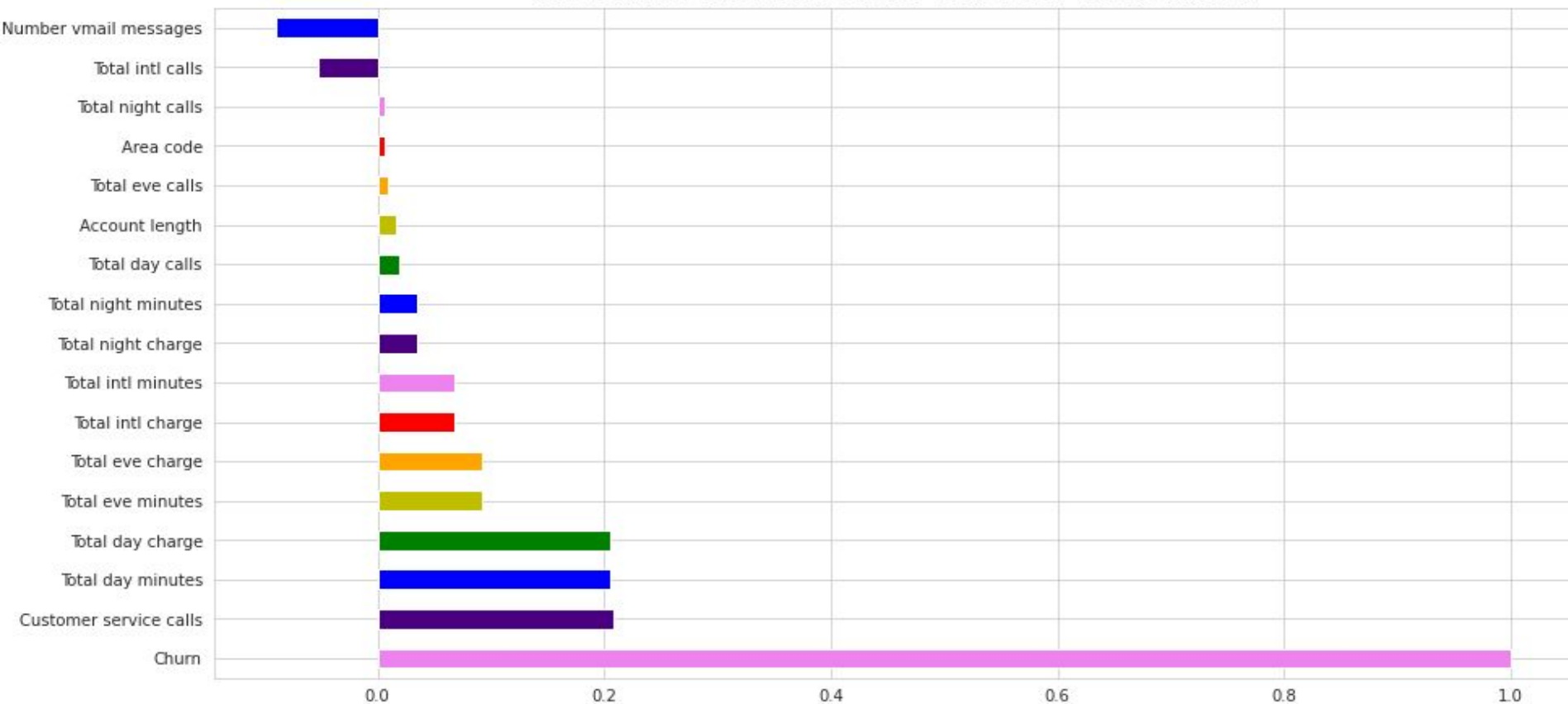
Defining the Target variable:

- In this pie chart we can see that 14.49% customers cancelled the subscription and 85.51% customers did not cancel the subscription.
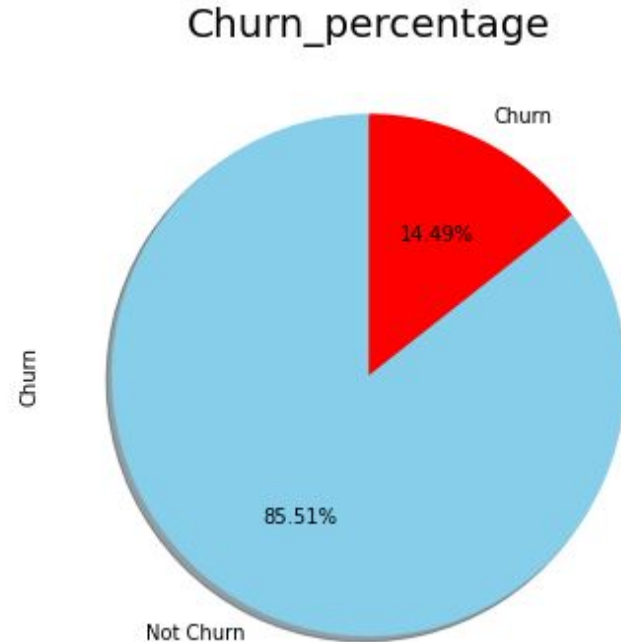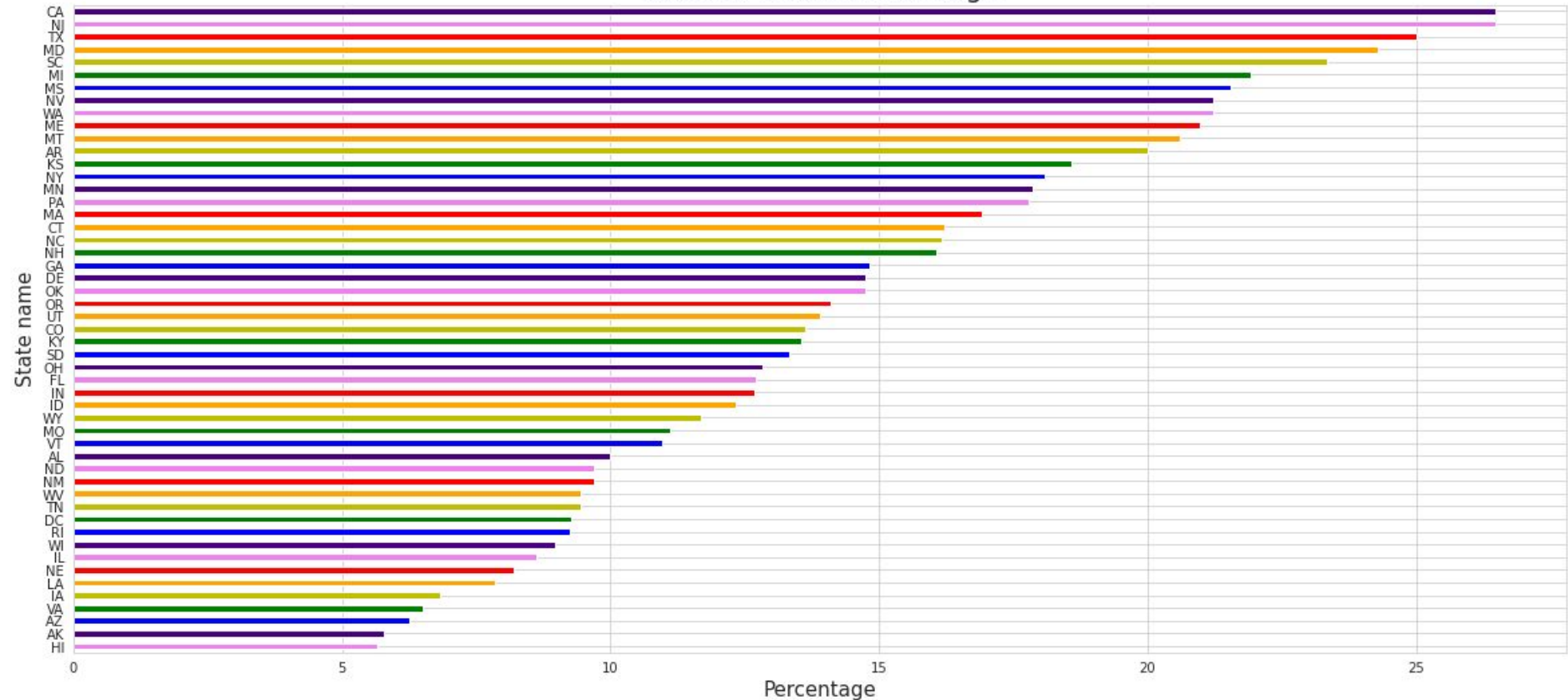

Churn_percentage

Relation with the Target variable:

Correlation between Churn and all of the columns

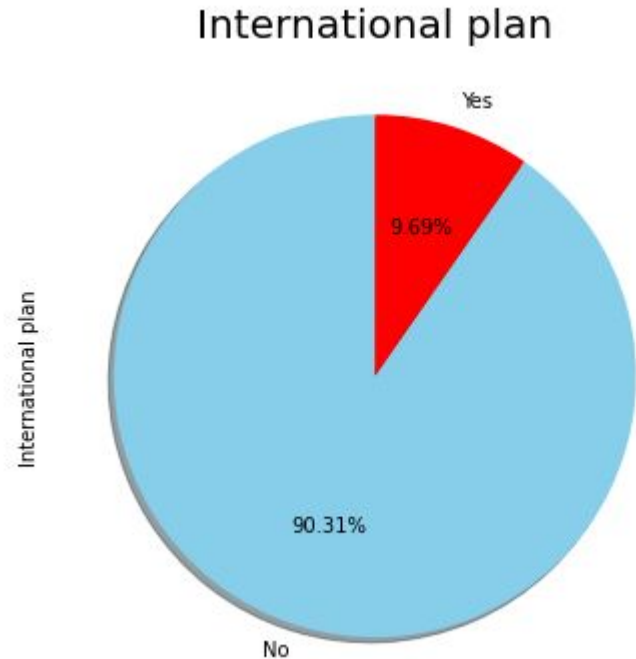# Analysis on column 'State':

State vs Churn Percentage

Analysis on column 'State':

- There are 51 states which have different churn rates.

- CA, NJ, TX, MD, SC are the ones who have higher churn rate (more than 23%) which is even more than 50% of average churn rate (14.49%).

# Analysis on column 'International plan':

- From this pie chart we can see that 9.69% customers do not have International plan and 90.31% customers have International plan.

International plan

Analysis on column 'International plan':

- From this graph we see that customers with the International Plan tend to churn more frequently.
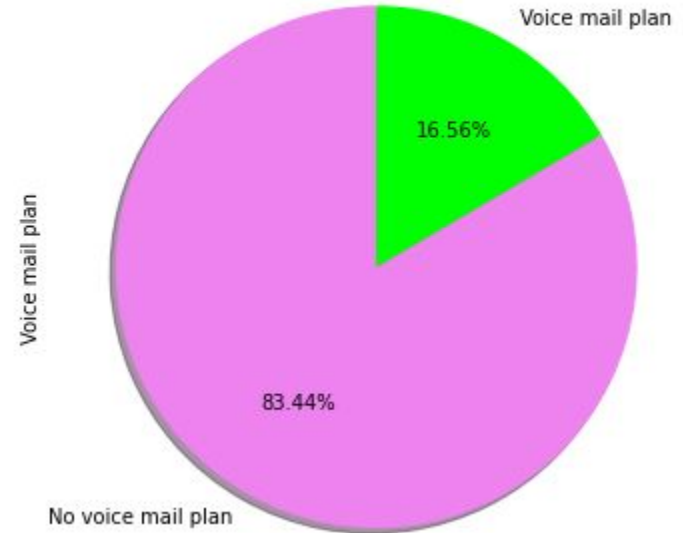

Churn percentage with respect to International plan
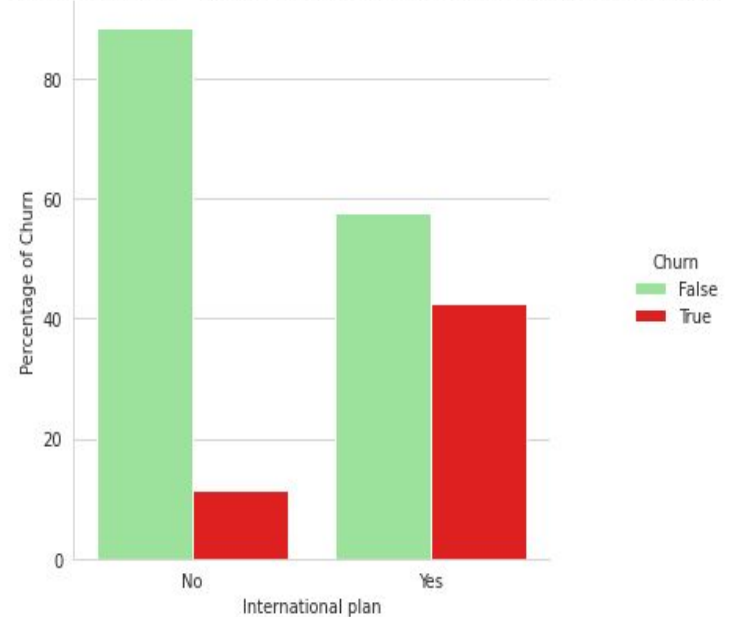
Analysis on column 'Voice mail plan':

- Out of total churned customers, 83.44% did not have an active voice mail plan. Therefore the customers, who did not opt for voice mail plan, are more prone to churn.

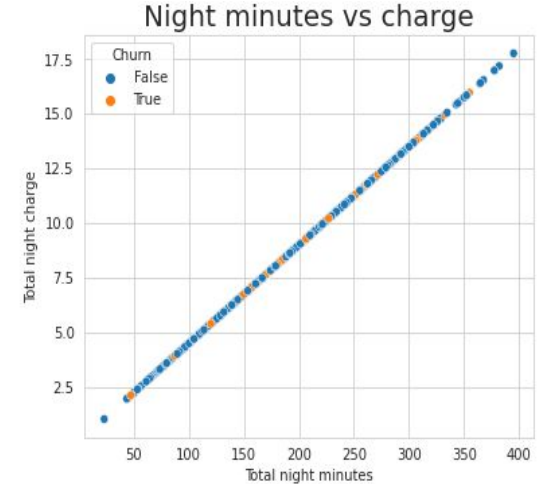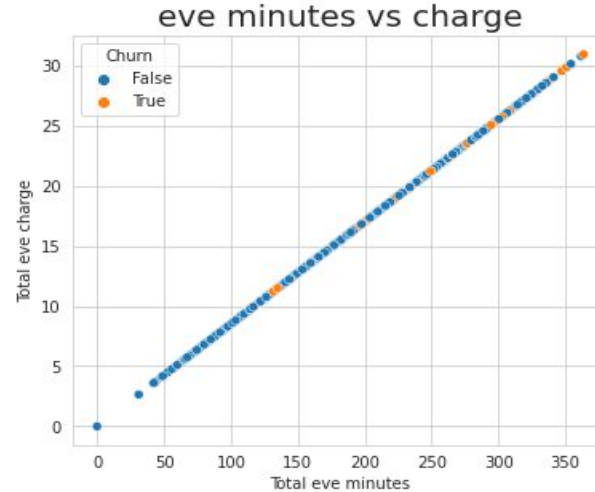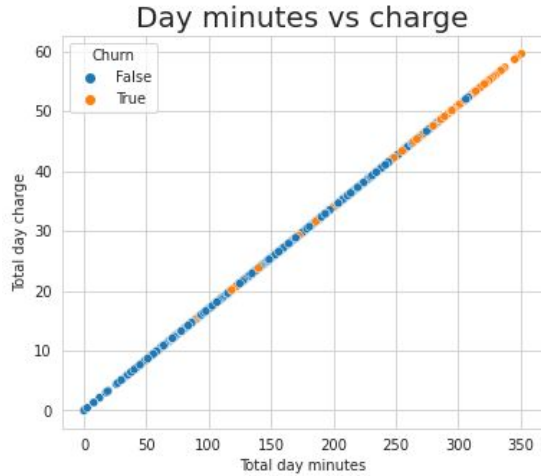voice mail plan dist. for churn customer

Analysis on column 'Area code':

- Since all three area codes have around 14% churn rate, we can conclude area code is not an influential parameter.

Area code vs Churn percentage
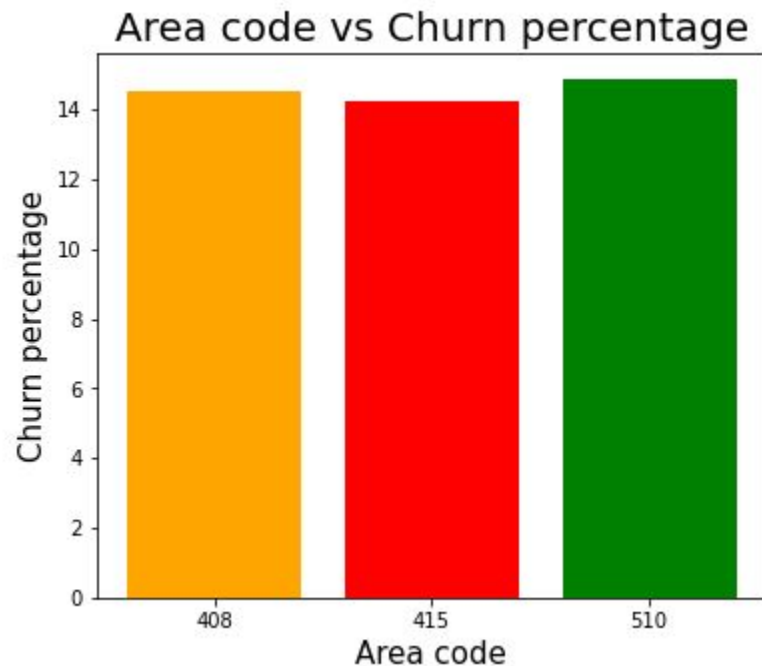
Analysis on 'Total minutes':

Day minutes vs charge | eve minutes vs charge | Night minutes vs charge

❑ From the above three scatter plots, we can see total minutes and total charge are highly positively correlated(almost 1). so we can drop one redundant feature(Total Charge) and carry out our analysis on the feature- Total minutes.

Analysis on column 'Total minutes':

Churn	Total day minutes						Total eve minutes						Total night minutes					
	count	mean	std	min	50%	max	count	mean	std	min	50%	max	count	mean	std	min	50%	max
False	2850.0	175.175754	50.181655	0.0	177.2	315.6	2850.0	199.043298	50.292175	0.0	199.6	361.8	2850.0	200.133193	51.105032	23.2	200.25	395.0
True	483.0	206.914079	68.997792	0.0	217.6	350.8	483.0	212.410145	51.728910	70.9	211.3	363.7	483.0	205.231677	47.132825	47.4	204.80	354.9

❑ We can see that the average minutes of calls for the customers, who cancelled the subscription are more then 200 minutes.

❑ To retain those customers, The telecom company has to come up with some special offer for those customers who spend more then 200 minutes.

Analysis on column 'Customer service calls':

Customer service calls vs Churn percentage
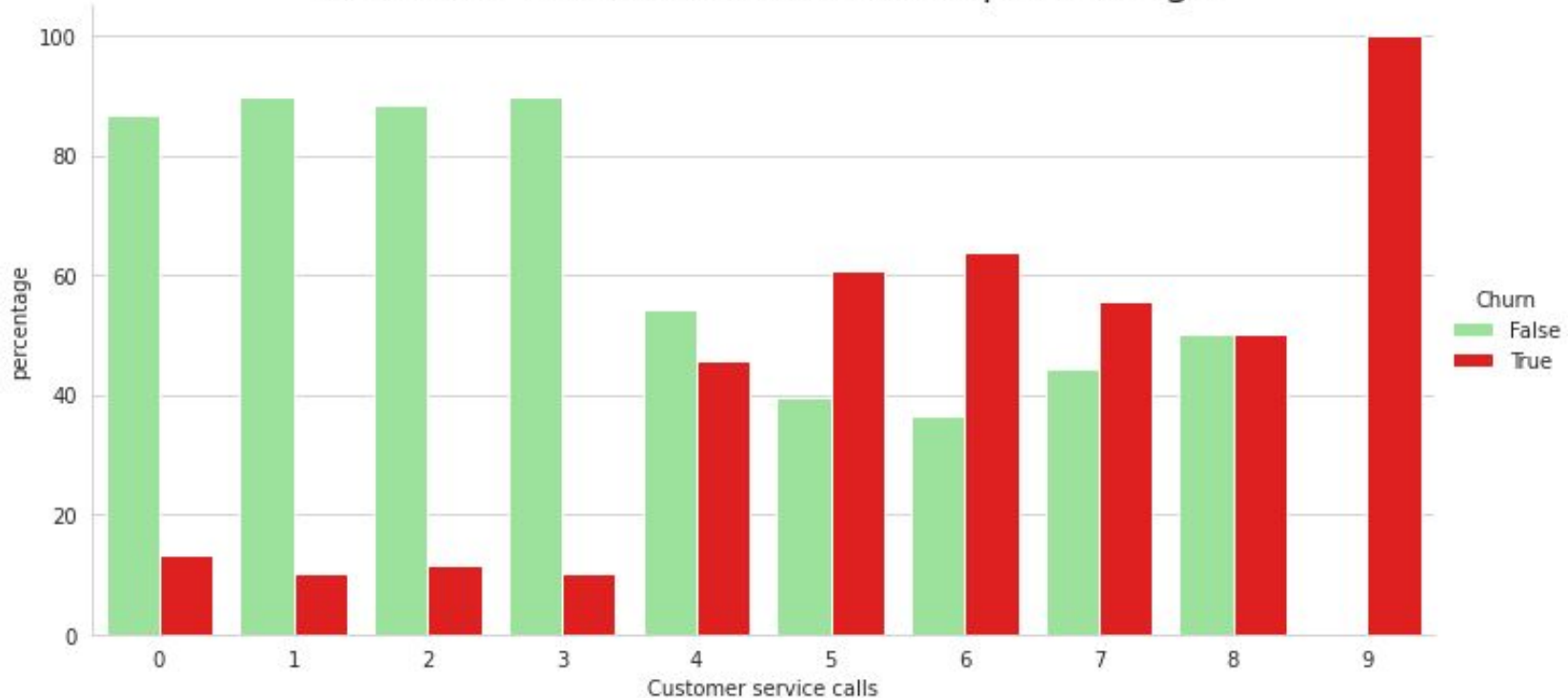
Analysis on column 'Customer service calls':

- **Observation:** The service calls of customers varies from 0 to 9 .
- Those customers who make more service calls, are very dissatisfied with their current service provider and more likely to churn.
- We can see from graph, for the customers who have more then 4 service calls, probability of leaving is more then 50% (Aprx.)

- **Suggestion:** Hence for the customers who make more then 4 service calls, the queries should be resolved immediately and they should be given better service so that they won't leave the company.

Conclusion:

❏ After doing all the analysis we can conclude that high call charge on International calls, no special offer for those customers who make more calls, less active voice mail plan and after making several customer service calls, customer's problem still not resolved are the key factors **responsible for customer churn.**

❏ **For customer retention:**

☐ International call charge should be minimized

☐ The company should provide some special offers to those customers who spend more than 200 minutes on calls

☐ Encouraging the customers to buy more voice mail plans

☐ Customer's problem should be resolved before 4 customer service calls

Challenges:

- Limited number of instances

- Imbalanced dataset (Churn: 14.49% and non-churn: 85.51%)

Thank you