EDA Capstone Project

Telecom Churn Analysis

By- Mitesh Bhanushali

Points for Discussion

- Problem Statement
- Introduction
- Data Summary
- Data cleaning
- Visualizing columns
- Relationships of different column and target column
- Bi-variant and Multi-variant analysis
- Conclusion

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.

Explore and analyse the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.

Introduction:

Before jumping into our data set let’s understand what a churn or churn rate is,
In the telecommunications industry, the customer is the most valuable asset for the company to sustain the business process. Churn rate is a measure of the number of individuals or items moving out of a collective group over a specific period. Churn rate is also called attrition rate.

The formula for churn rate is defined as (Total number of customers leaving/ Total number of customers) x 100.

Data Summary

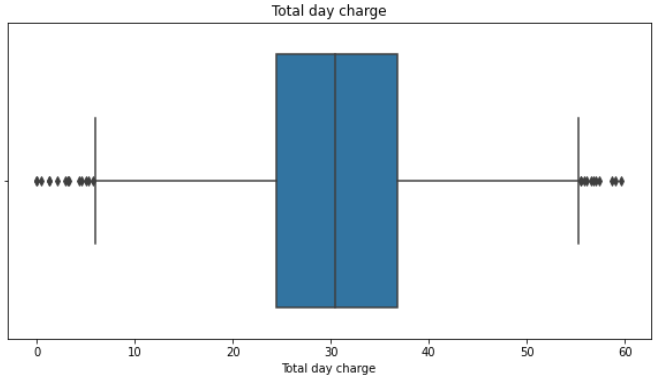
telecom_data: This is a data frame which consists of all the required data to understand why the customer is cancelling the subscription plan. It contains about 3000 rows of data and 20 columns including churn which has a Boolean datatype .

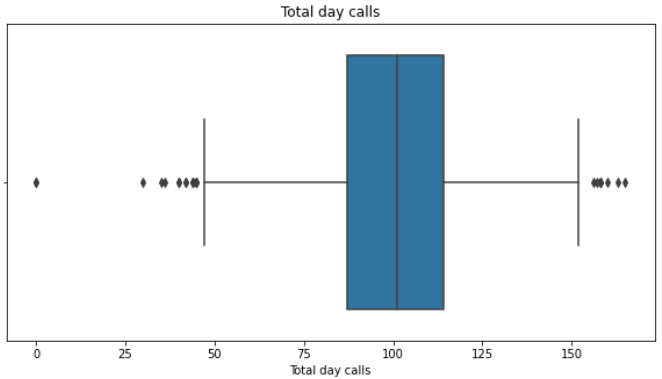
True: Customer cancelled the subscription.
False: Customer didn’t cancelled the subscription.

Data Cleaning

There where no null values present in the dataset but there are some outliers that we need to address. A box plot gave us the rough idea where the outliers lied. Using IQR (Inter Quartile Range) we found the minimum and maximum of a column if the data point lies outside this two values we replaced it with a mean of that value.

We also changed the datatype of few columns such as Customer service call and Total day call

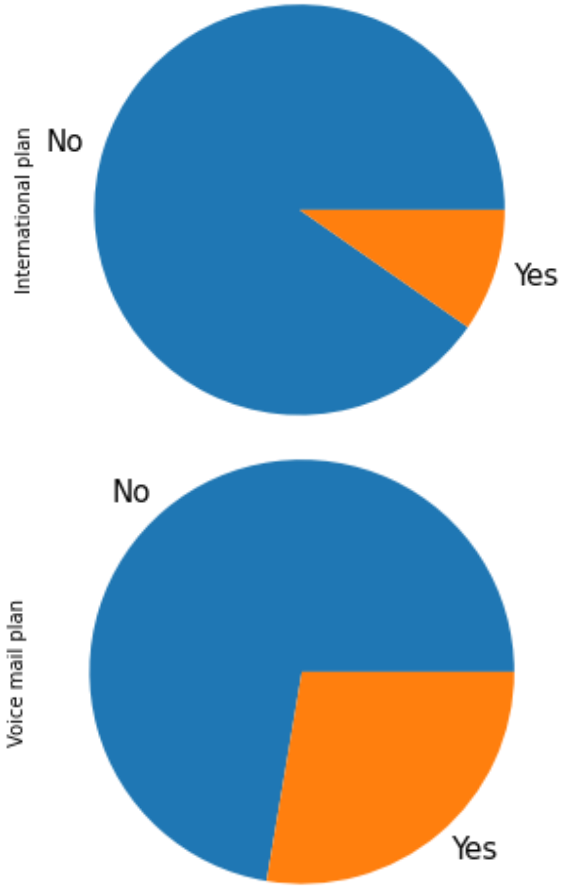




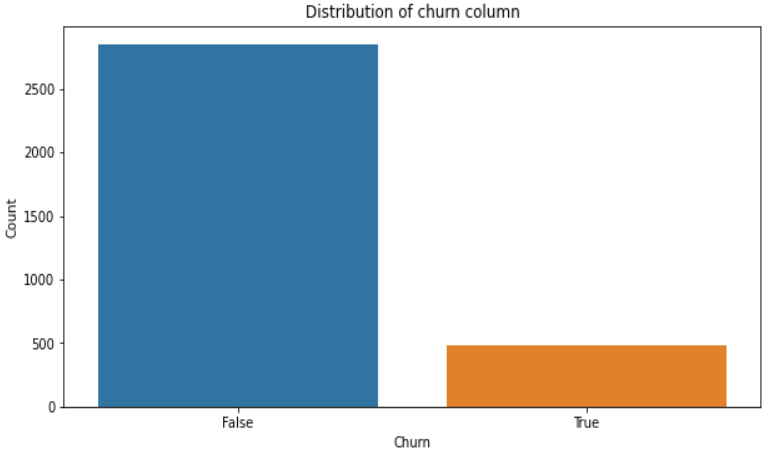
Visualizing Columns

- 90% of customer didn’t opt for international plan i.e. 3010 where not subscribed to international plan and only 323 subscribed the were subscribed to international plan

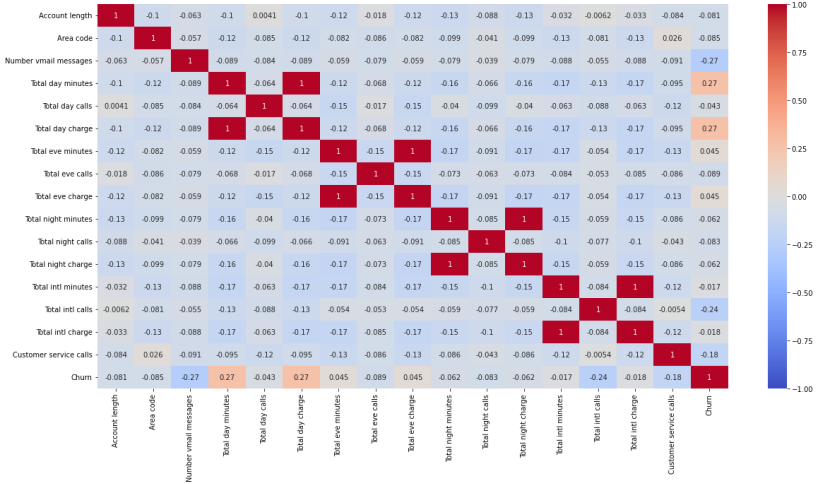
- Whereas for Voice mail plan 72% didn’t subscribe and 28% subscribe to voice mail plan



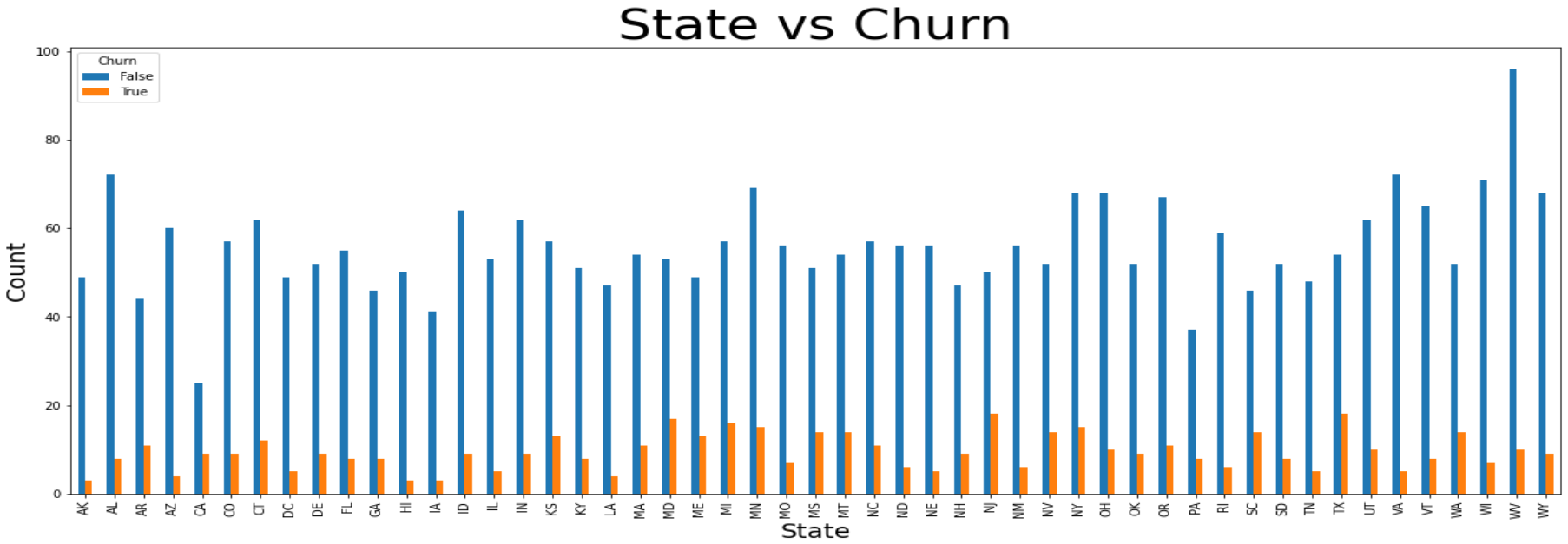
This is our target column Churn. We can see that the data is not properly distributed it has 2850 rows for FALSE value and only 483 for TRUE value. 85% of values are False and 15% are True.



We can see the corelation of target column with different column. It has highest corelation with column Total day charge with value of 0.27.

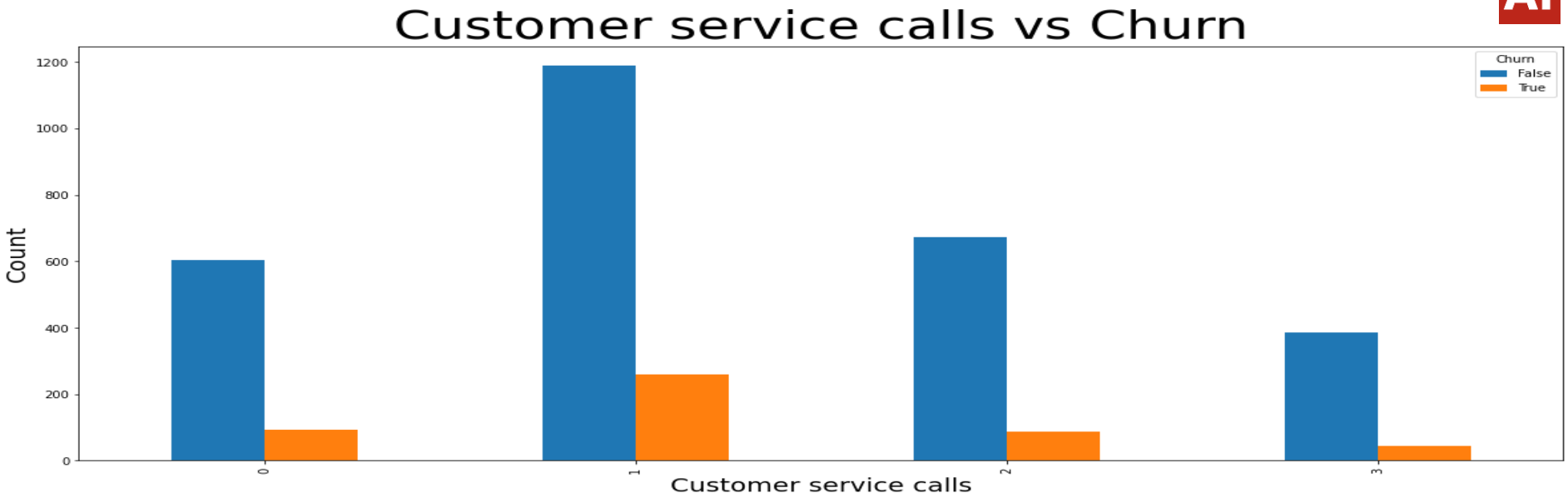


Target vs Independent variable



The state that has highest number of customer not cancelling the subscription is WV and

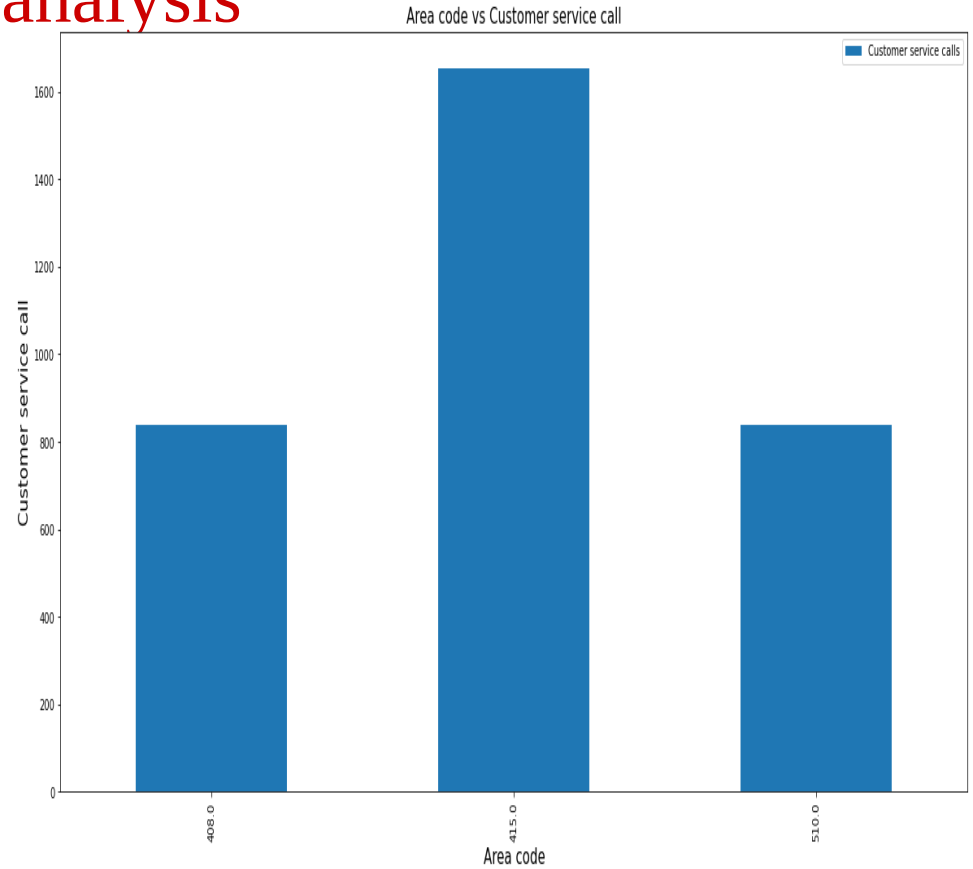
The state that has highest number of customer cancelling the subscription is TX

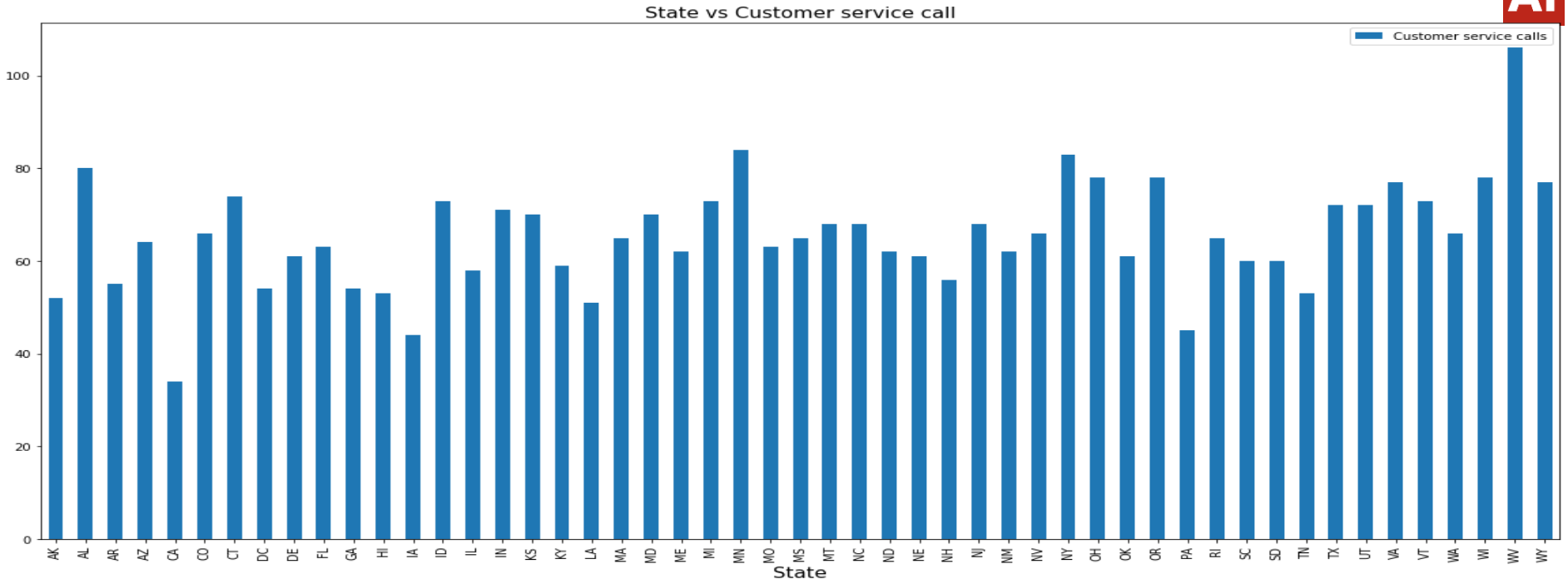


Out of total data provided 1448 customers called customer service at least once. We can also see that if customer is calling for first time for service it has high change of customer will cancel the subscription or not, based on the service given to customer.

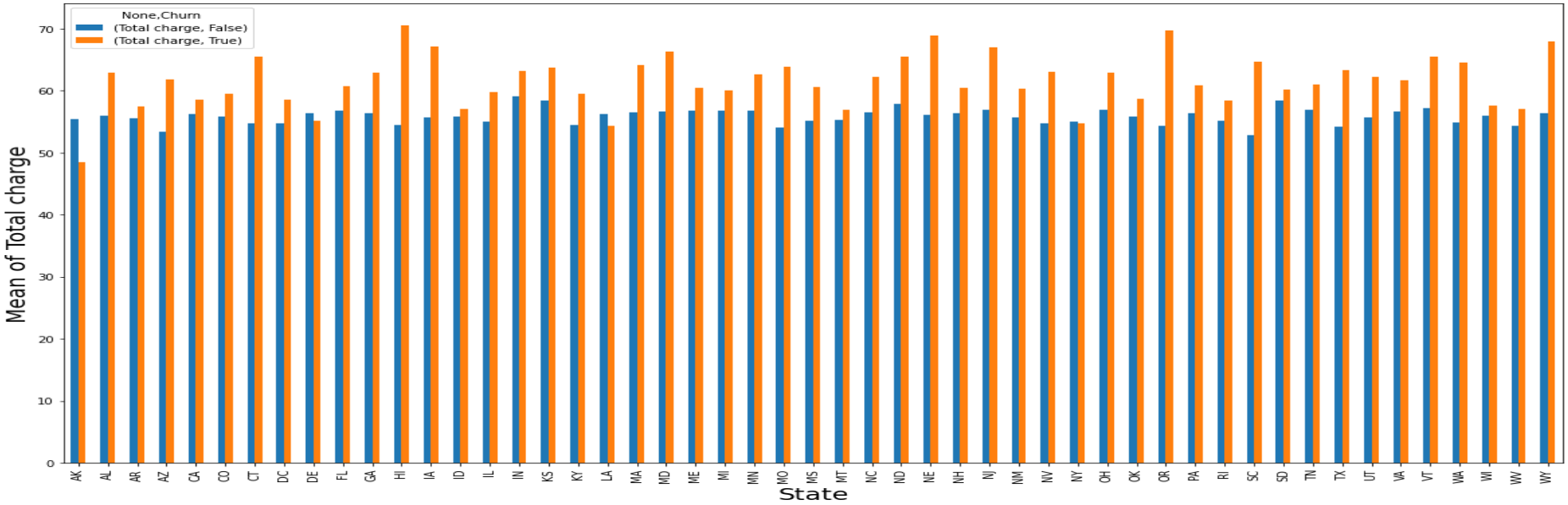
Bi-variant and Multi-variant analysis

Graph of Area code vs customer service call. The Area code which had highest number of customer service call is 415





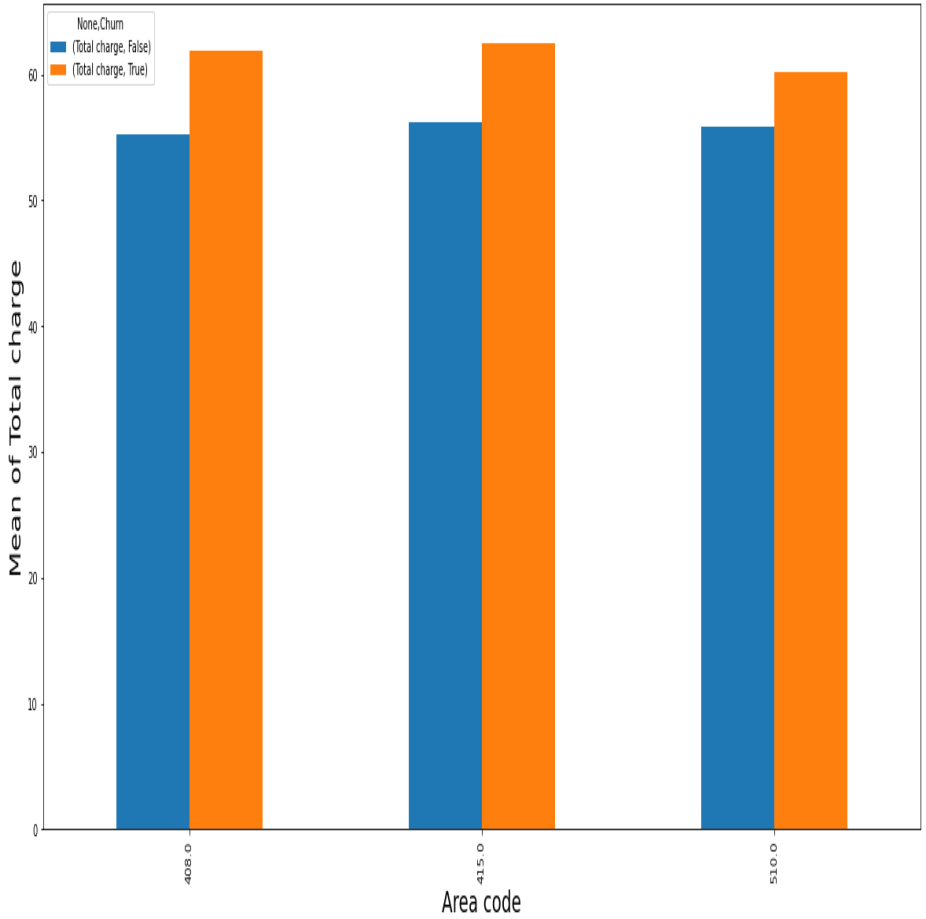
Maximum of customer service call is done by state WV and minimum customer service call is done by state CA.



We can see that the customer which are cancelling the subscription are paying more than average of total charge than the customer that are not cancelling the subscription. State with highest average total charge for costumer cancelling the subscription is HI with charge of 70.47. State with highest average total charge for costumer not cancelling the subscription is IN with charge of 63.16.

Top 3 states with highest average charge and cancelling the subscription are 'HI', 'OR' and 'NE'.

We can see the same thing for area code. The average charge paid by the customer that are cancelling the subscription is higher than the customer that are not.



Conclusion:

So at the end we can see that the customer that are cancelling the subscription are being charged more than the customer that are not cancelling the subscription.

So to address this issue and for customer retention we can introduce a new plan with a charge less than the current average charge and we can also lower the charge of current average charge.

Thank You