Capstone Project - 1
Telecom Churn Analysis

Individual Project
Annamaneni Ajay – Cohort Himalaya

Contents...

1. Defining problem statement.
2. Summary regarding the data set
3. Explore the data set
4. Analyse the data set
5. Insights obtained from Analysis and also recommendations.

Problem Statement

- With the rapid development of telecommunication industry, the service providers are inclined more towards expansion of the subscriber base. To meet the need of surviving in the competitive environment, the retention of existing customers has become a huge challenge. In the survey done in the Telecom industry, it is stated that the cost of acquiring a new customer is far more that retaining the existing one.

Problem Statement

Objective :

- Orange S.A., formerly France Telecom S.A, is a French multi national telecommunications corporation. We are given with 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 or not. The idea of this project is to identify why users are churning to another Telecom service and recommend the measures to be taken to stop user churn. In order to complete this task I will initially proceed with knowing about features in the data set and then analysing and exploring the dataset.

Data Summary...

- There are a total of 20 columns and 3333 rows in our data set. Out of 20 columns/features 19 are independent variables and ‘Churn’ is the only dependent variable. So lets get to know about the features in data set

State: This depicts us to which State a user belong to.

Account length: This might depict us the validity period chosen by the users.

Area code: This depicts to which area the user belongs to.

International Plan: This depicts whether the user has chosen for international plan or not(either Yes or NO).

Data Summary…

Voice mail plan: This depicts us whether the user has chosen for voice mail plan or not (either Yes or No).

Number vmail messages: This depict us the total number of voice mail messages sent or received by user.

Total Day minutes: This depicts us the total number of day minutes the user have called to others through our service.

Total Day calls: This depicts us the total number of day calls the user have called to others through our service.

Data Summary...

Total Day charge: It is the bill for which the user has to pay for day calls

Total eve minutes: This depicts us the total number of evening minutes the user have called to others through our service.

Total eve calls: This depicts us the total number of evening calls the user have called to others through our service.

Total eve charge: It is the bill for which the user has to pay for evening calls.

Data Summary...

Total night minutes: This depicts us the total number of night minutes the user have called to others through our service.

Total night calls: This depicts us the total number of night calls the user have called to others through our service.

Total night charge: It is the bill for which the user has to pay for night calls.

Total intl minutes: This depicts us the total number of international minutes the user have called to others through our service.

Data Summary...

Total intl calls: This depicts us the total number of intl calls the user have called to others through our service.

Total intl charge: It is the bill for which the user has to pay for international calls.

Customer Service calls : This depicts us the total number of times a user have called to customer service for complaining about the problem he faced.

Churn: This depicts us whether the user has churned to another network or not(either True or False)

Data set exploration...

```
df.shape
```

```
(3333, 20)
```

- The Telecom churn data set consists of 3333 rows and 20 columns

- Also, the data set do no contain any null/NAN values.
- The data type of various columns is also displayed on right.
- The data set contain 1 column which is Boolean, 8 columns which are float64, 8 columns which are int64, 3 columns which are object type.

```
<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
```

Data set exploration...

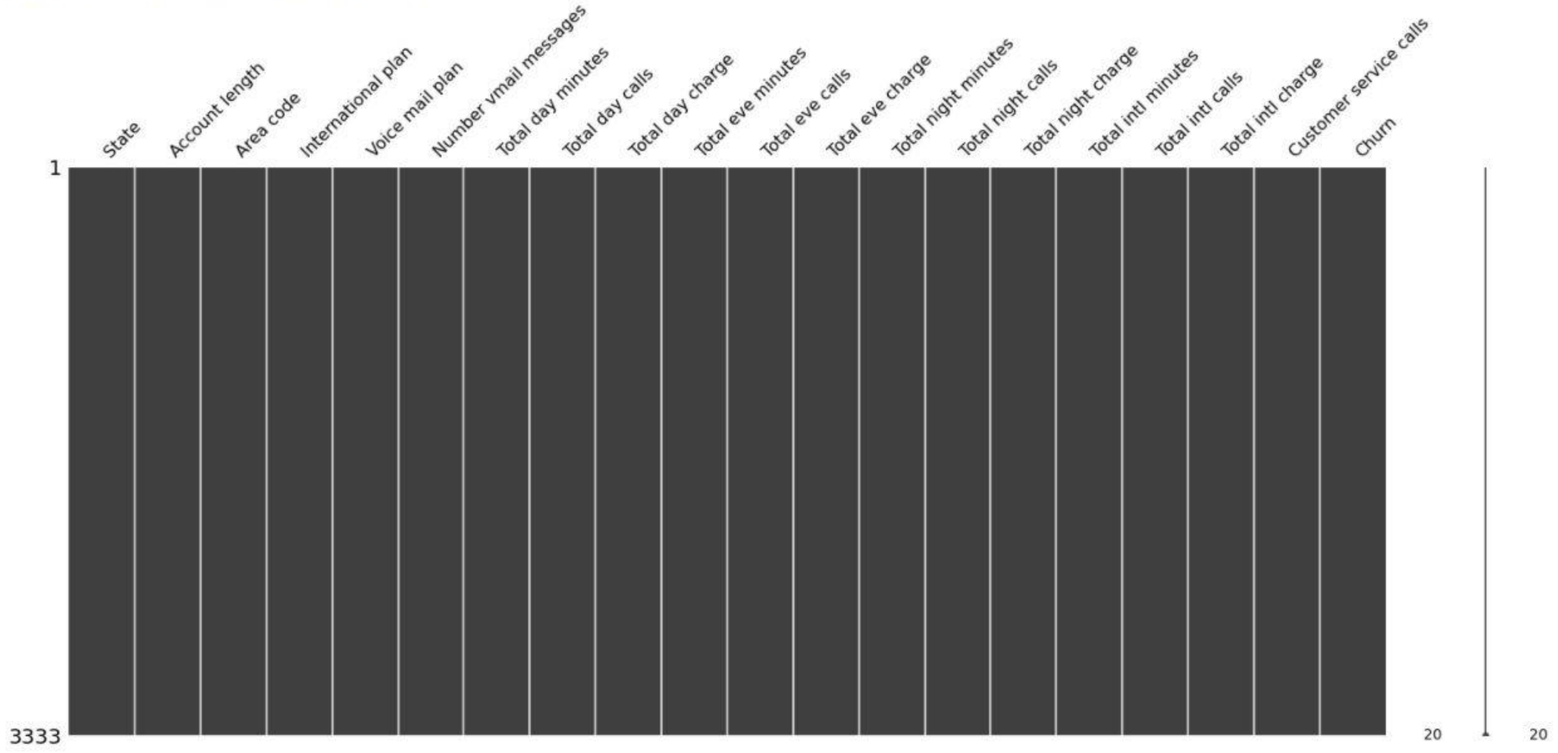
- NULL Values : Missingno is a Python library that provides the ability to understand whether the columns contain missing values through informative visualizations. The visualizations can be in the form of heat maps or bar charts. With this library, it is possible to observe where the missing values have occurred. Since the above data set doesn't contain any null values the visualization is uniform across all columns but if we consider data set which has null values and if we apply missingo library then the we can see few lines in the columns where null values are present.

.

Data set exploration...

```
import missingno as msno
msno.matrix(df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f144c4a3790>
```



Data set exploration...

- Lets us check the corelation between the specific columns by plotting a heatmap. Heat map depicts how a column is related to other column. If it is completely dependent then the value is 1 else it decreases up to -1

- for example :Total night charge and Churn are very less related since the value is 0.035.

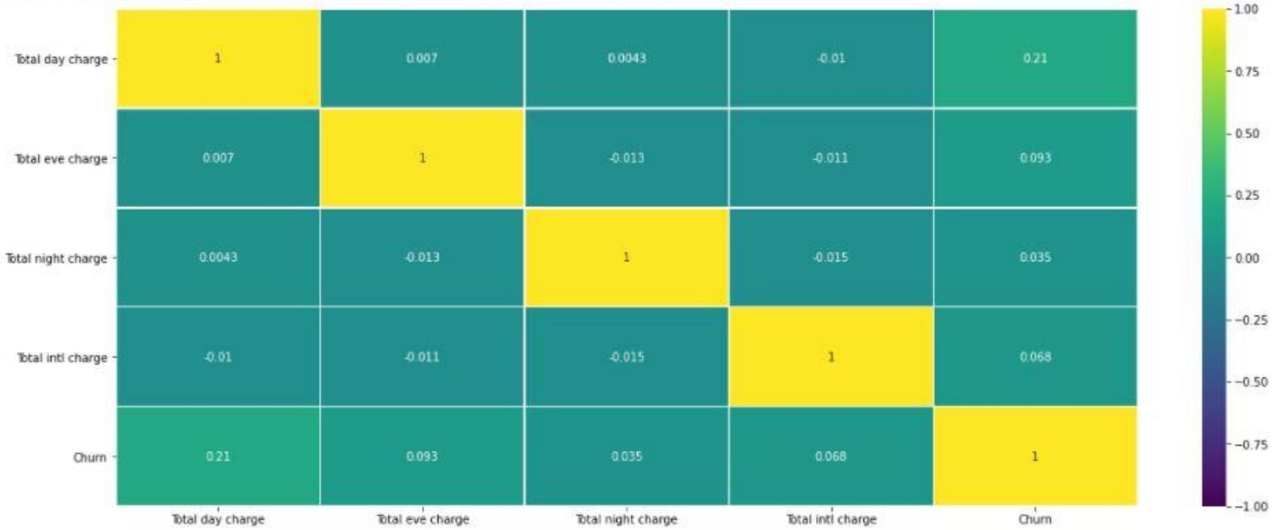
.

Data set exploration...

```
[ ] corr_df = df[['Total day charge','Total eve charge','Total night charge','Total intl charge','Churn']] # Storing the required columns into another data frame
```

```
plt.figure(figsize=(20,8)) #Setting the length and bredth of the visualization.
sns.heatmap(corr_df.corr(), vmin=-1, cmap ='viridis', linewidths = 0.30, annot = True)  #vmin=-1 sets the minimum value for the color scale at -1
                                                                                        #cmap is the color of the visualization
                                                                                        #annot=True means it Annotates values
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f144c3d9b90>
```



Data set exploration...

- Let us check for the unique State codes present in data set.

```
df['State'].unique() #gives the unique state codes present in data set.
```

```
array(['KS', 'OH', 'NJ', 'OK', 'AL', 'MA', 'MO', 'LA', 'WV', 'IN', 'RI',
       'IA', 'MT', 'NY', 'ID', 'VT', 'VA', 'TX', 'FL', 'CO', 'AZ', 'SC',
       'NE', 'WY', 'HI', 'IL', 'NH', 'GA', 'AK', 'MD', 'AR', 'WI', 'OR',
       'MI', 'DE', 'UT', 'CA', 'MN', 'SD', 'NC', 'WA', 'NM', 'NV', 'DC',
       'KY', 'ME', 'MS', 'TN', 'PA', 'CT', 'ND'], dtype=object)
```

- Describe method returns various attributes(like count, mean,..)for integer and float data type columns.

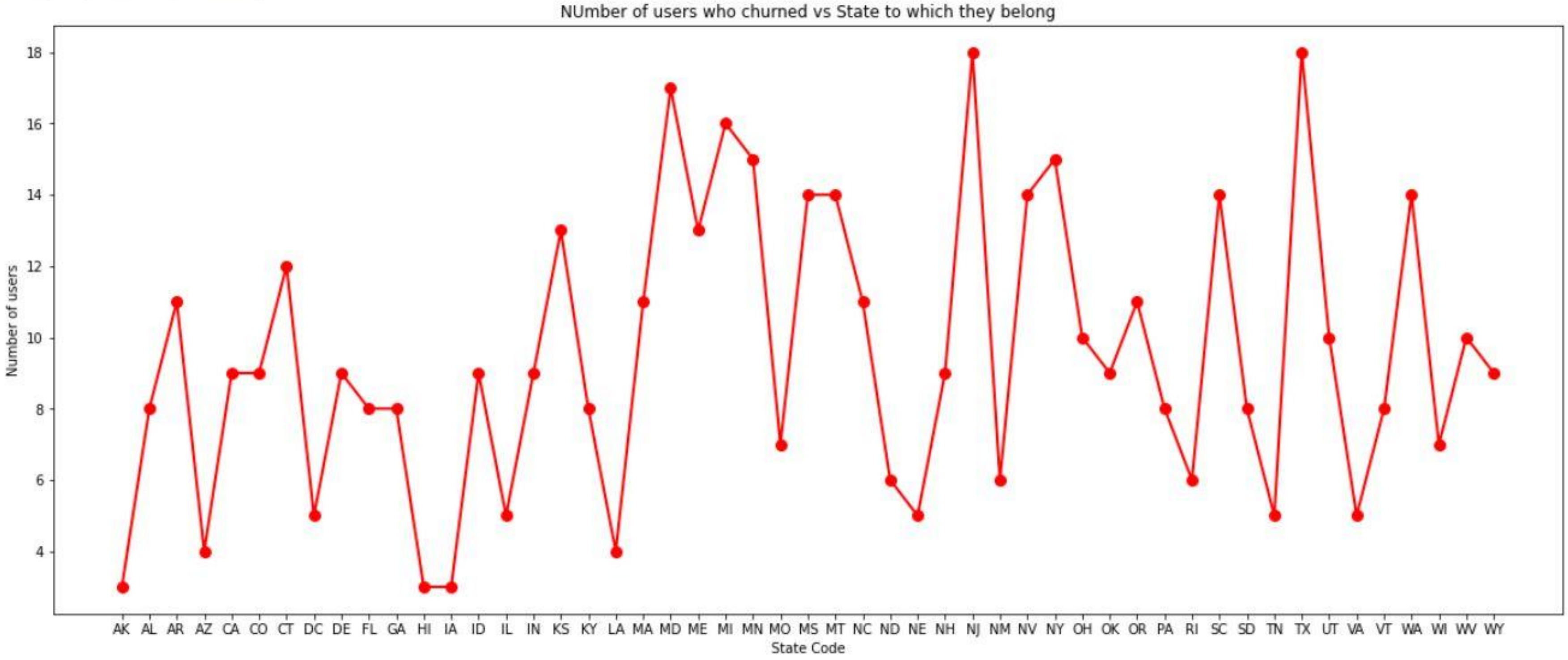
```
df.describe() #describe method return various attributes(like count, mean,min,max,...) for integer and float data type columns.
```

	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

Analysing Data

1. Let us analyse which state has maximum people churning to another network.

```
Text(0.5, 0, 'State Code')
```



Analysing Data

1. Let us analyse which state has maximum people churning to another network.

Insight-1 : People in the states (NJ and TX) are more churning to another company compared with another states. This might be because the people from those areas are facing low network issues.

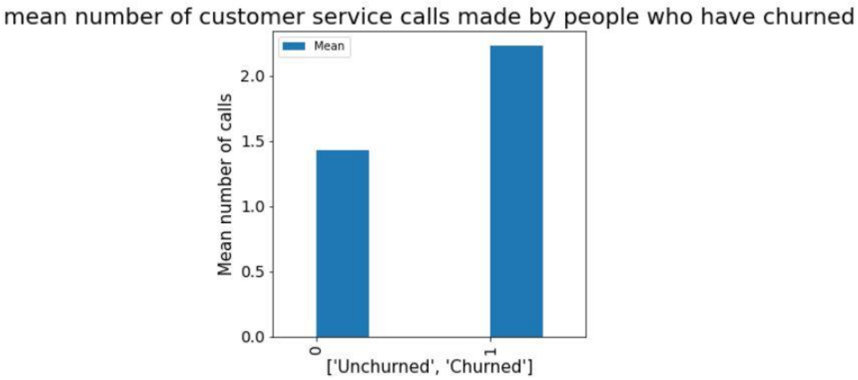
Insight-2 : States (AK, HI, IA) have the least number of people who churn. This might be because the people might not be facing any issues.

Recommendation-1 : .The Telecom company should focus more on areas NJ and TX and improve the infrastructure in those areas to stop churn.

Analysing Data

2. Let us check how many times the churned people called to customer service and compare it with number times the churned people called to customer service.

Customer Service Calls : This depicts us the total number of times a user have called to customer service for complaining about the problem he faced.



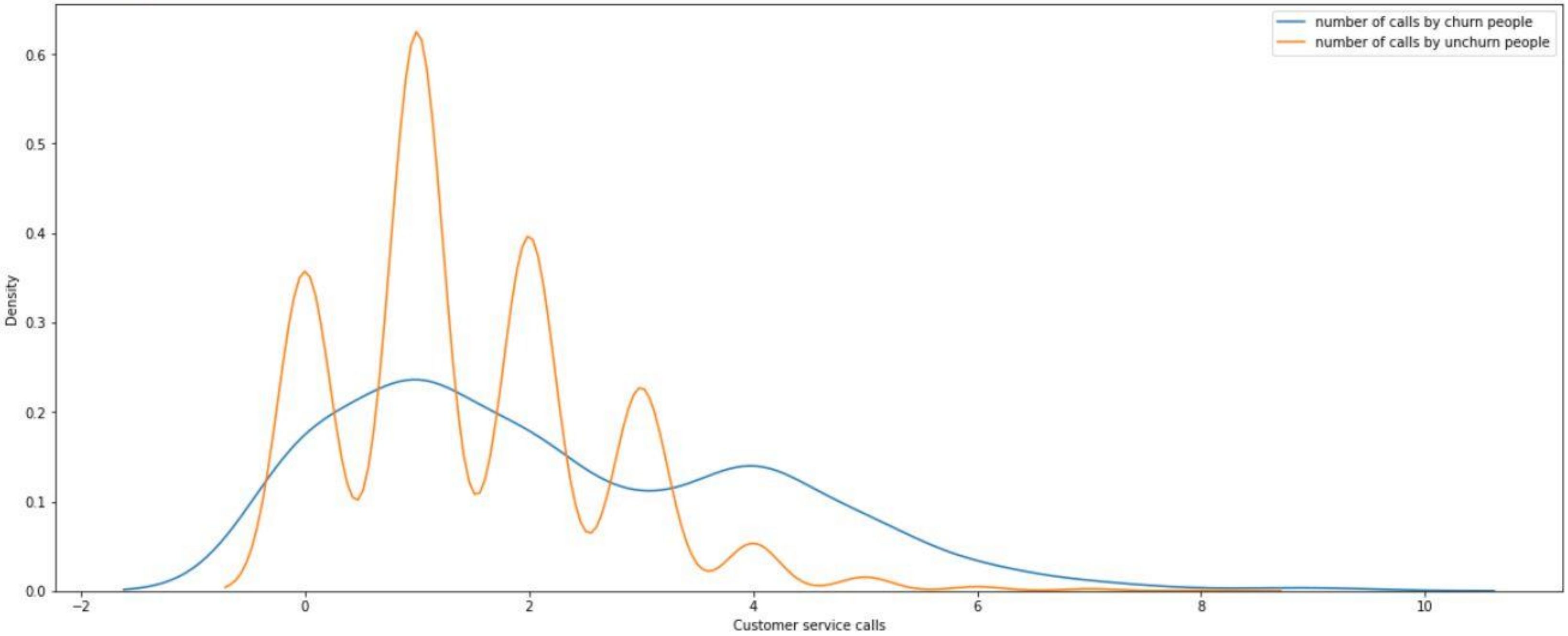
Analysing Data

2. Let us check how many times the churned people called to customer service and compare it with number times the churned people called to customer service.

Insight-3 :

From the above graph it is clear that people before churning have made more number of calls to customer Service(greater than 2 on an average), may be for complaining regarding an issue. Since the problem isn’t resolved even after complaining they might have churned to another network.

Analysing Data



- Blue line depicts the number of calls made to customer service for complaining by churned
  people.
- Orange line depicts the number of calls made to customer service for complaining by un-churn people

Analysing Data

2. Let us check how many times the churned people called to customer service and compare it with number times the churned people called to customer service.

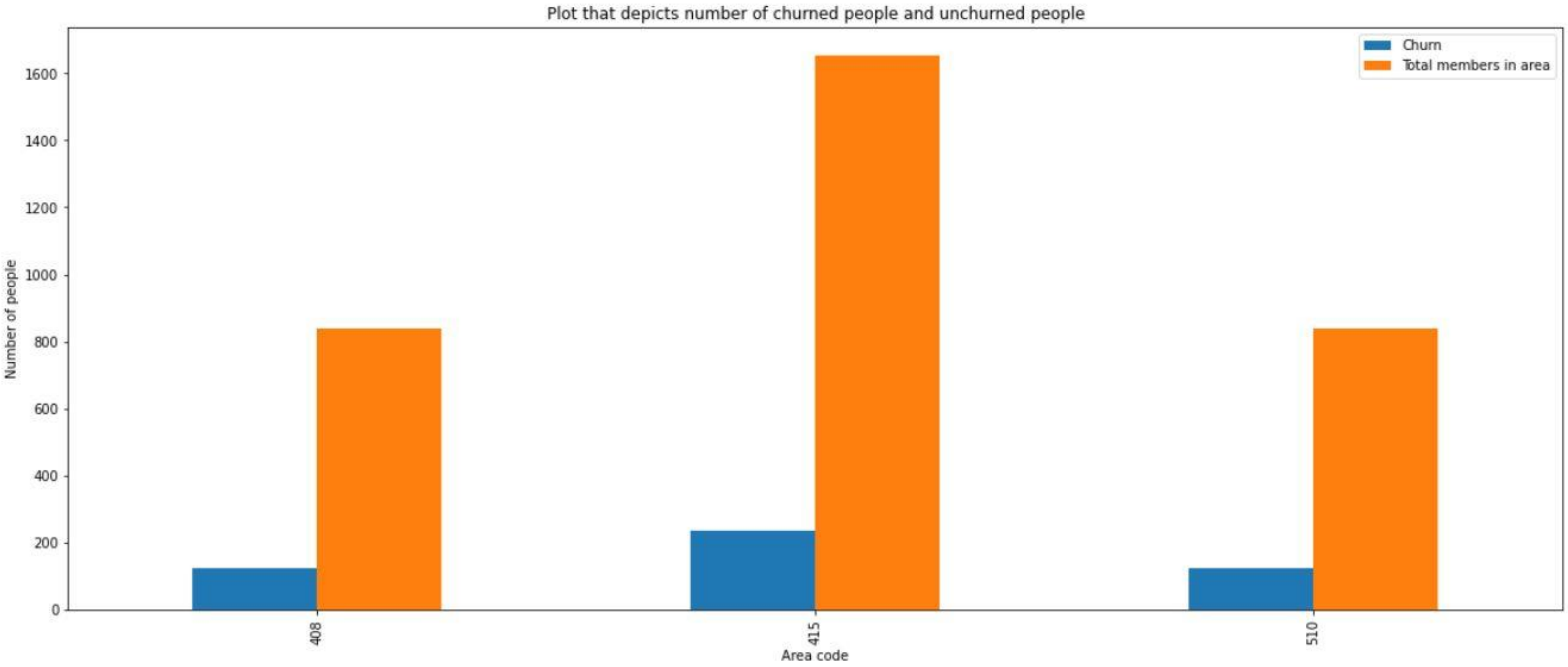
Insight-4 : From the above dist plot it is clear there are people call customer care once or twice and get there problem solved but we can see that even after people calling customer care for 4 to 8 times their problems are not getting fixed as a result they are churning.

- Blue line depicts churn people Orange line depicts un-churn people

Recommendation-2 : Improving customer service and taking quick action on complaints filed might stop people churning to another network.

Analysing Data

3. Analyze the customers who have churned based on area code.



This plot depicts the churned people and tool number of people in specific area code

Analysing Data

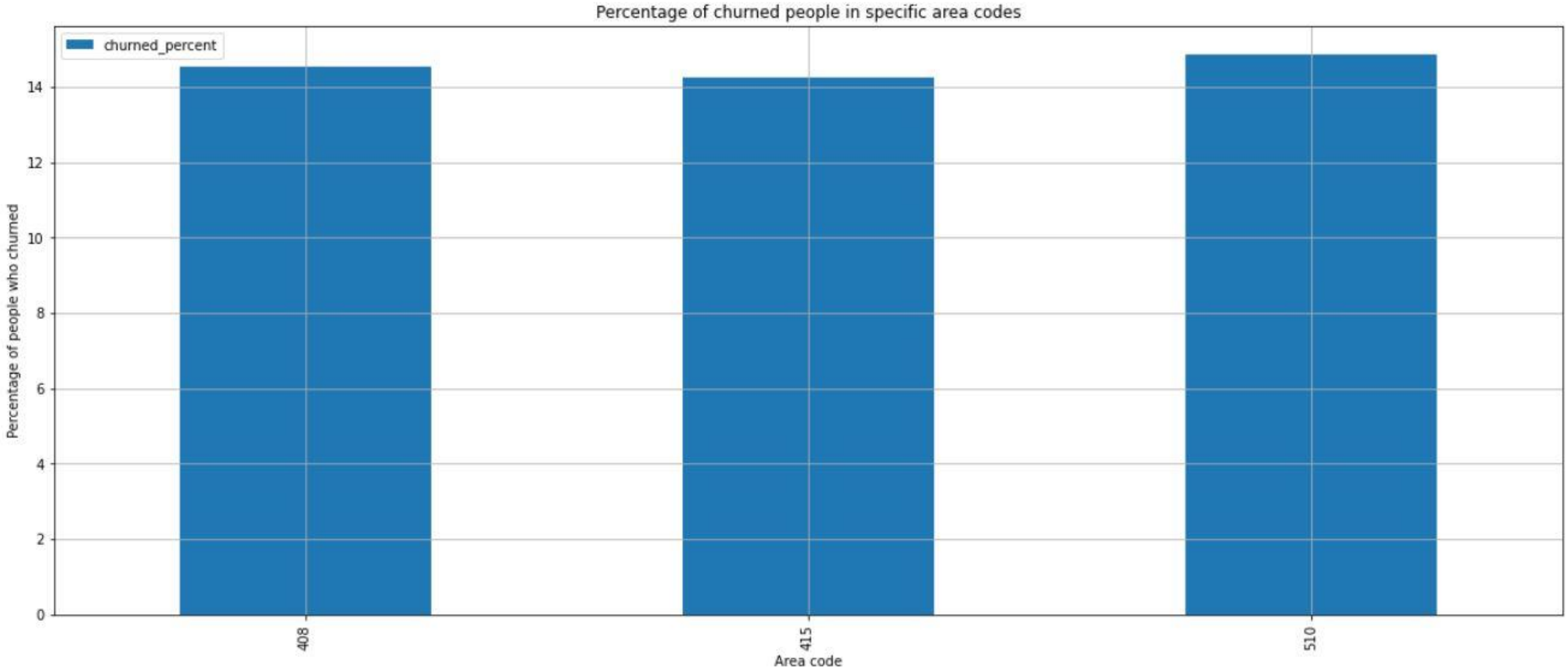
3. Analyze the customers who have churned based on area code.

Insight-5:

- This plot depicts the churned people and total number of people in specific area code.
- In area code 415 there are highest number users and less number of people who churned.

Analysing Data

3. Analyze the customers who have churned based on area code.



Analysing Data

3. Analyze the customers who have churned based on area code.

Insight-6 :

- From the above two graphs it is clear that area code 415 has more number of users then followed by 510 and 408.
- But More percentage of people who churned are from area code 510 and then followed by 408 and 415.

Recommendation-3 : The company has to more focus on area codes 408 and 510 and solve problems faced by people so that people do not churn from network.

Challenges

- Had to refer various blogs for making the visualization look colorful and easily understandable.
- Had to go through documentations of various functions for syntax.

Conclusion..

- People in the states (NJ and TX) are more churning to another company compared with another states. This might be because the people from those areas are facing low network issues.
- States (AK, HI, IA) have the least number of people who churn. This might be because the people might not be facing any issues.
- People before churning have made more number of calls to the customer service, may be for complaining regarding an issue. Since the problem was not resolved even after calling many times, they might have churned.
- In area code 415 there are highest number users and less number of people who churned.

Conclusion..

- The area code 415 has more number of users then followed by 510 and 408.
- But More percentage of people who churned are from area code 510 and then followed by 408 and 415.