

Capstone Project

Telecom Churn Analysis

TEAM MEMBERS	EMAIL
Pranav Rajmane	rajmanepuranav19@gmail.com
Sayali Mandhare	saileemandhare98@gmail.com
Prthamesh Kalambe	Prathmeshkalambe122@gmail.com
Dayanand Katkade	dayanandkatkade3@gmail.com
Pavan Malvay	pmalvay@gmail.com

Table of Content

- **Data Summary**
- **Problem Statement**
- **Objective**
- **Data Inspection**
- **Data Cleaning(null and missing values ,outliers etc.)**
- **Data Analysis and Visualization :**
 - a. **Top States for Churned Users**
 - b. **Relation Between factors and Churning**
 - c. **International plan and Churning**
 - d. **Customer Service and Churning**
 - e. **Correlation**
- **Conclusion and summary**
- **Suggestions**

Summary(Basic terms used in the Dataset)

- State - State in which the customer resides, it is indicated by a two-letter abbreviation
- Account length- the number of active accounts
- Area code - the three-digit area code of the corresponding customer's phone number**
- International plan- whether the customer has an active international calling plan**
- Voice mail plan - whether the customer has an active voice mail feature**
- Number vmail message - the average number of voice mail messages per month**
- Total day minutes - the total number of calling minutes used during the day**
- Total day calls - the total number of calls placed during the day**
- Total day charge - the billed cost/charges of daytime calls**
- Total eve minutes - the total number of calling minutes used in the evening**
- Total eve calls - the total number of calls placed during the evening**
- Total eve charge - the billed cost/charges of calls placed in evening
- Total night minutes - the total number of calling minutes used in the night**
- Total night calls - the total number of calls placed during the night **
- Total night charge- the billed cost/charges of night time calls**
- Total intl minutes - the total number of minutes during an international call
- Total intl calls - the total number of international calls**
- Total intl charge - the billed cost/charges of international calls placed**
- Customer service calls - the number of calls placed to Customer Service**
- Churn - the number of customers or employees who canceled the services

Problem statement



Telecom Data provided by an French multinational telecommunications corporation The Orange S.A.. Their data focuses on United States region. During this data recording period, they have noticed that some of their users left their company because of some reason. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. Our analysis can help in knowing the reason why users will leave that telecom service and what should be the perfect strategy for customer retention

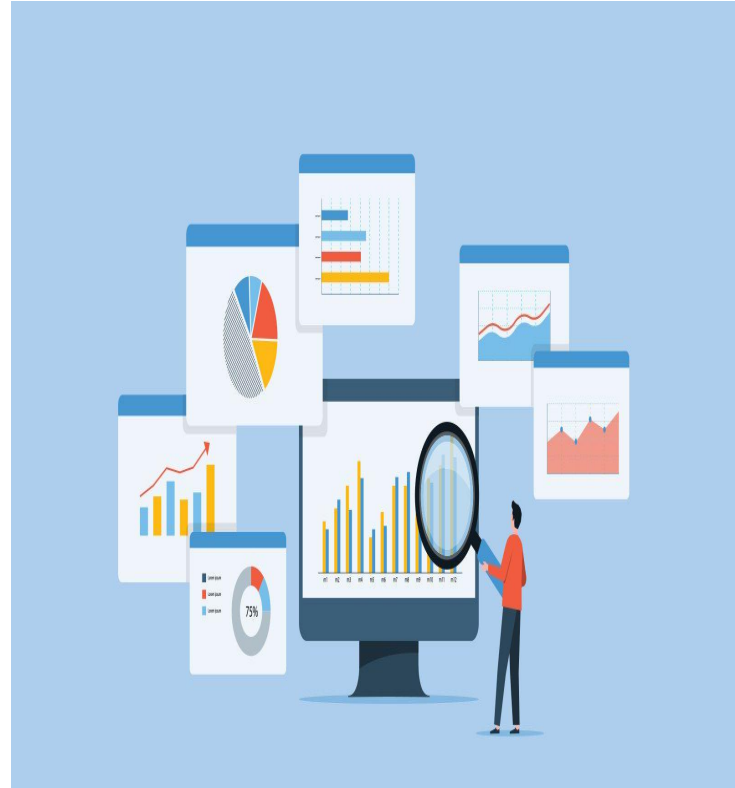
Objective

- The main objective is to do some analysis, which could help them in findings the key factors responsible for customer churn, identifying churn behaviour and validate the reasons for customer churn with the help of EDA .
- Based on that we can recommend some suggestions for customer retention to business team based on analysis of telecom churn data set for reducing customer churn and increasing profit of the company.



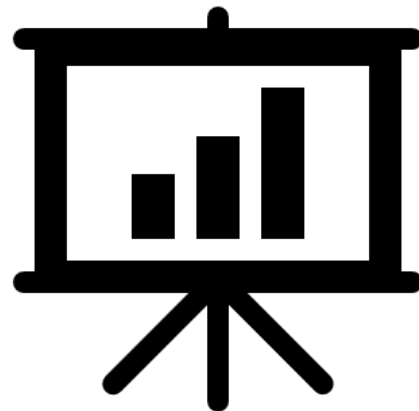
Data Inspection

- Data inspection is used To ensure that we are dealing with the right dataset and for a clear view of data set at every stage of the transformation process.
- Data inspection is the act of viewing data for verification and debugging purposes, before, during, or after a translation
- During this step we checked shape of data, data types, unique value in column and statistics information



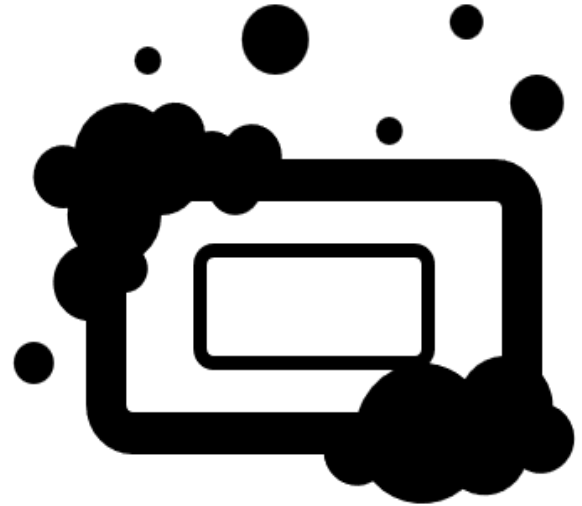
Data Inspection(continued)

- In order to understand our data, we can look at each variable and try to understand their meaning and relevance to this problem.
- Telecom churn dataset have 3333 rows and 20 columns having all the columns with data type of object, int, float, or bool.
- target column is Churn and telecom dataset consider only 3 categorical data columns those are state, international plan, voice mail plan and remaining columns are numerical dataset.
- the output column churn has 2 variables false or true based on dataset 14.5% customers are churned and remaining are not churned.



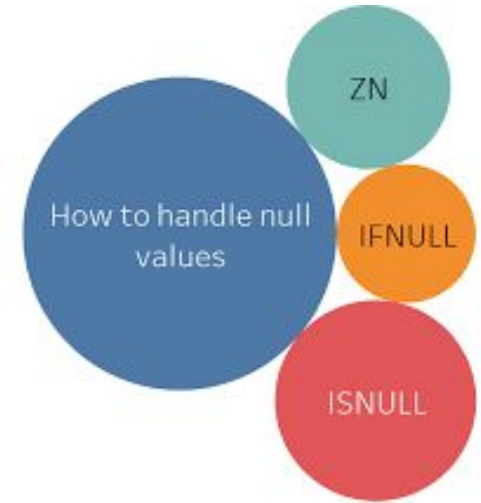
Data Cleaning

- Data cleaning is very important process in EDA. because raw data sometimes it consists of null values, missing values, duplicate values and outliers in data set, irregular format due to all this it's very difficult to create insights from dataset.
- Because of the huge data we are unable to go on main topic to solve it so we go through data and make needed cleaning



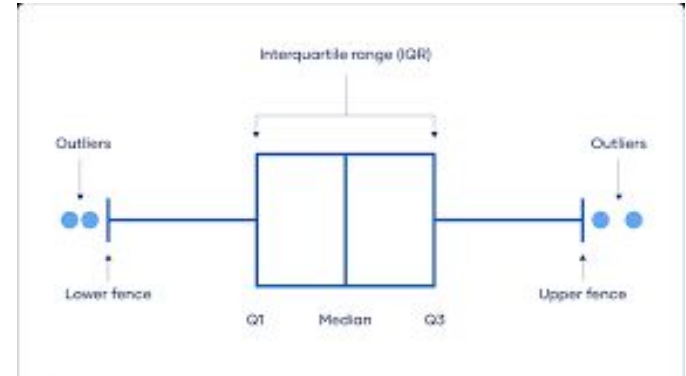
Data Cleaning(continued) : Null values

- Dealing with null values is very important because Missing data in the training data set can reduce usefulness of a result or can lead to a biased results. It can lead to wrong prediction also.
- In Orange S.A. telecommunication dataset, there are no null values.



Data Cleaning(continued): Outliers

- Checking outlier in the dataset is necessary because Outliers is also something that we should be aware of. Why? Because outliers can markedly affect our results and can be a valuable source of information. It is also providing us the insights of specific behaviours.
- Outliers is a complex subject and it deserves more attention. Here, we'll just do a quick analysis on data set to find outliers



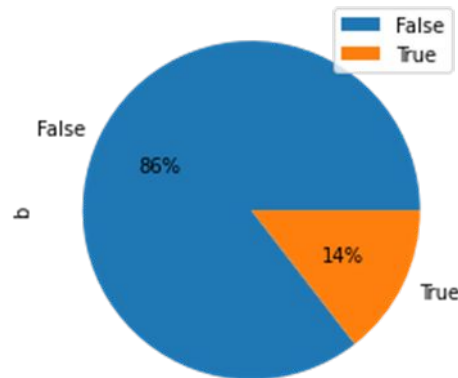
- **Data Visualization** represents the text or numerical data in a visual format, which makes it easy to grasp the information the data express. We, humans, remember the pictures more easily than readable text, so Python provides us various libraries for data visualization like matplotlib, seaborn etc.
- In this telecom churn EDA we will use Matplotlib and seaborn for performing various techniques to explore data using various plots.



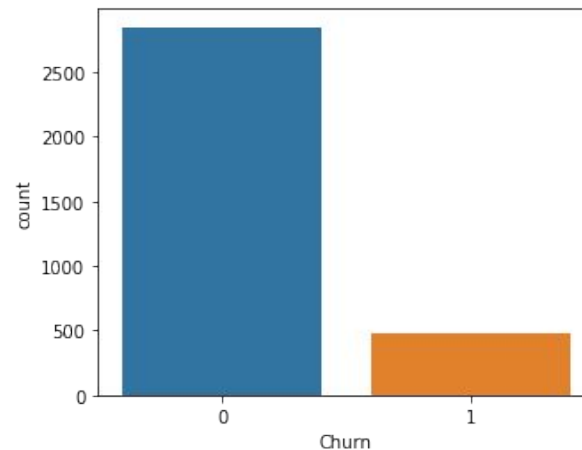
EDA (continued): Target variable

- Target variable in telecom churn dataset has 2 categories False or True .
- From pie chart and churned column we can see that our almost 14% customer has been churned. We can definitely do something about that after analysing the data.
- The graphs are as follows-

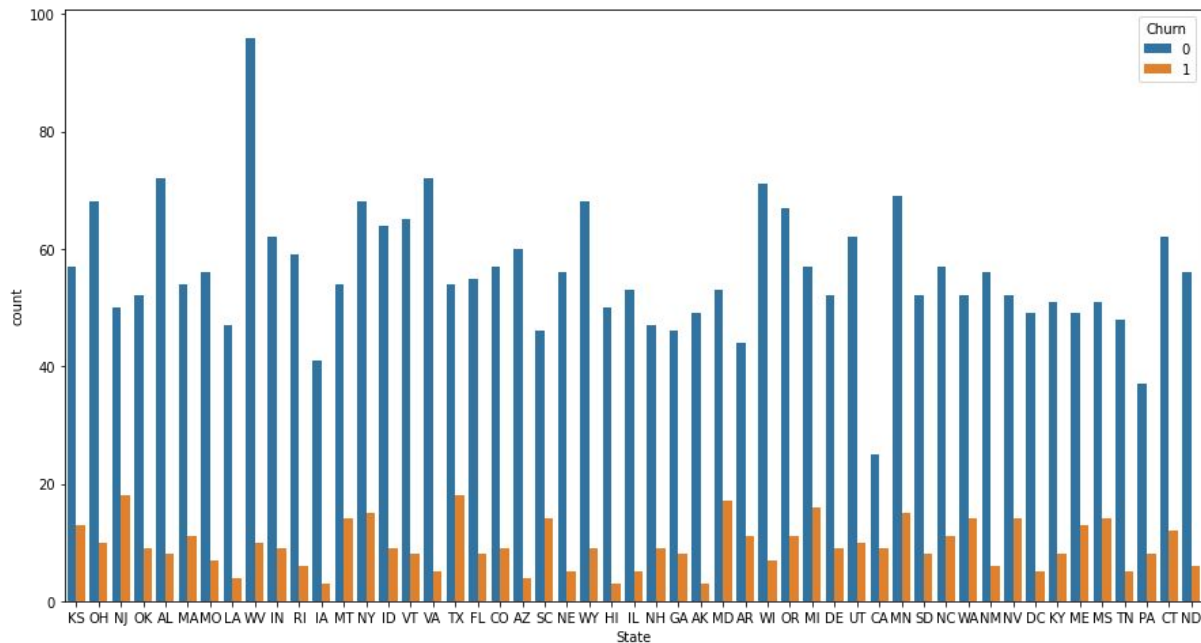
Percentage of churned and non-churned users



Non-churned vs Churned



EDA (continued): Comparing target column with states

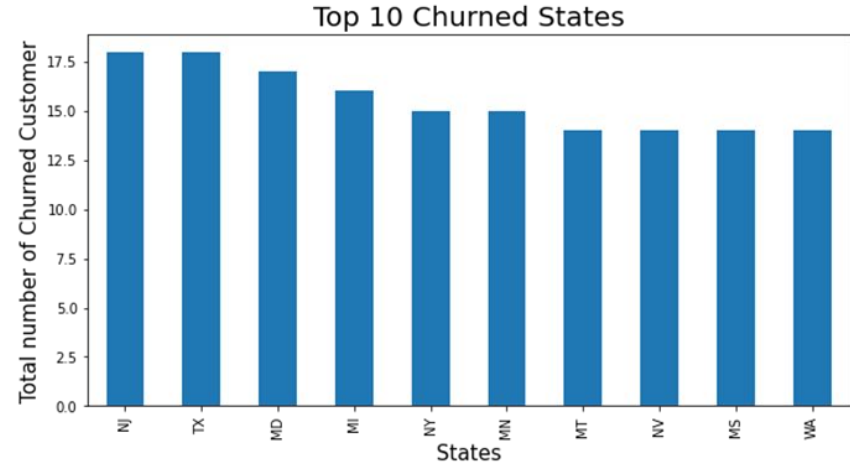


EDA (continued) : Top States for Churned Users

From bar plot we can observe that users of these states are prone to churn.

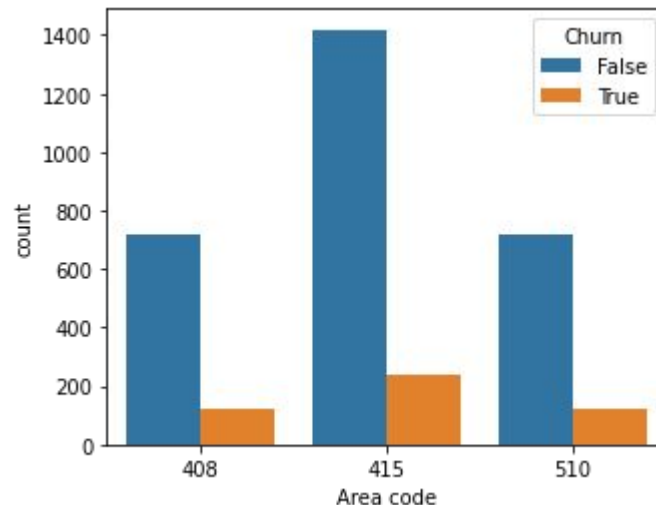
According to my hypothesis, following would be the factors that should be implemented for stopping churn in these states:

1. In top three states New Jersey, Texas and Maryland we can definitely choose aggressive pricing strategies and network upgradation as well as improving voice quality of calls in these states.
2. We have to look out for our competitor in these states also as observation says that there may be strong competition.



EDA(continued) :Relation between Churn vs area code

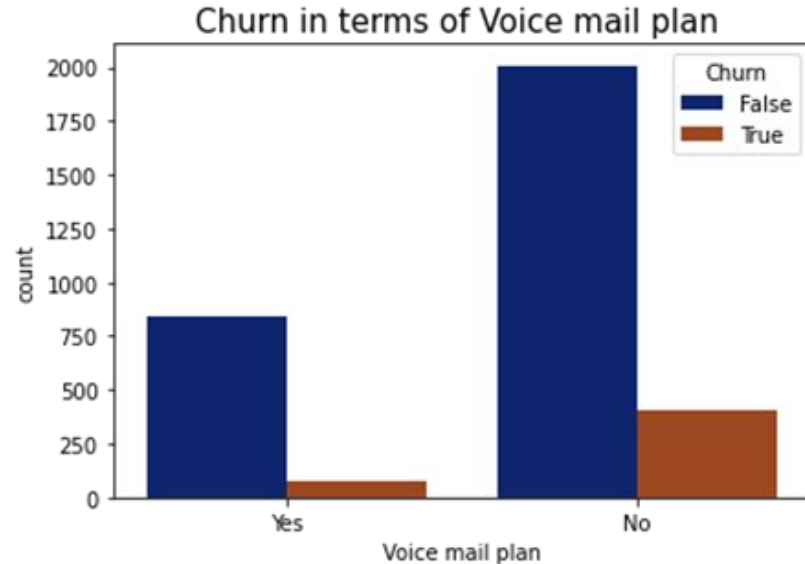
Churn			
Area code	False	True	Perc_Churn
408	716	122	14.558473
415	1419	236	14.259819
510	715	125	14.880952



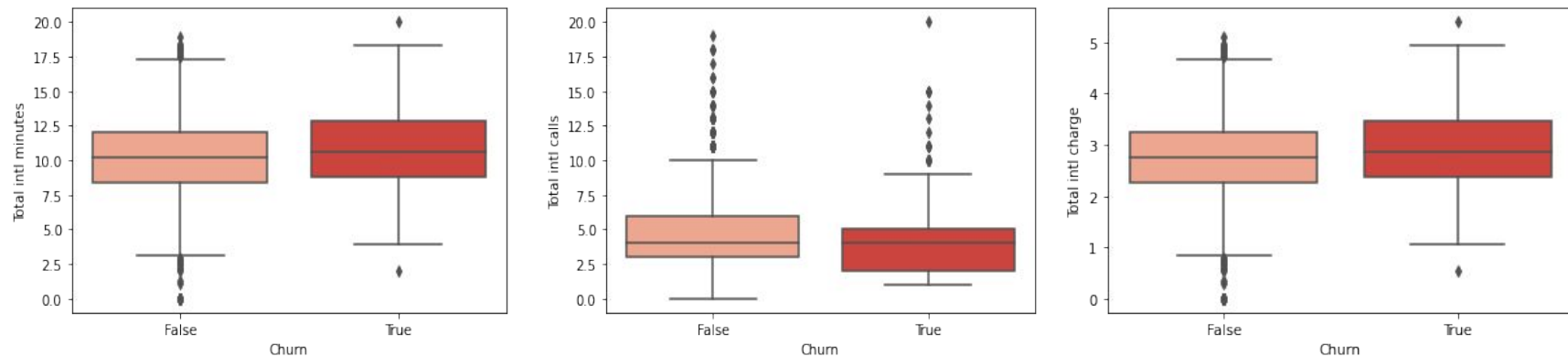
- From above plot we can see that the number of churned users percentage from each area code is same and area 415 has more number of churn .

EDA(continued): Churn distribution vs voice mail plan

From count plot we can see that the number of churned users who opt for voice mail plan is lesser than compare to those who did not opt for that, so we can predict that our voice mail plan is doing good hence we can improve it further.



EDA(continued): Relation between Churn and Total intl minutes, calls and charge

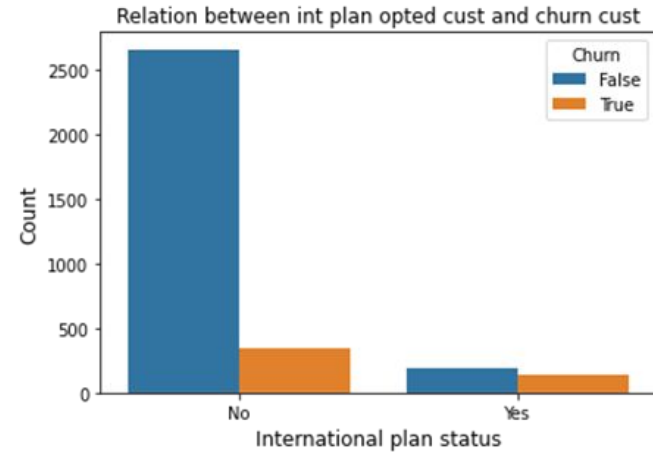


According to our findings from our above boxplot we can observe that users who are spending more minutes are tend to switch to other operators following would be the factors that should be implemented:

- **Implementing Different Pricing Strategy**
- **Network Upgradation.**
- **Implementing international Calling Rate Optimization would need to lower churn rate.**

EDA(continued): Relation between Churn vs International plan

	Churn-	False	True	Perc_ Churn
International plan				
No		2664	346	11.49
Yes		186	137	42.41



The ratio between churned user who opted for international plan is 42%. It means there is some problem with the pricing or voice call quality for International plan opted users.

According to our findings after monitoring the voice quality of international calls, following would be the factors that should be implemented :

1. Need to Upgrade or make smarter use of technologies like VoLTE for improvement of Voice Quality during calls.
2. Network Upgradation for international calls.

EDA(continued): Relation between Churn and Customer service



According to our findings, following would be the factors that should be implemented:

- 1. We have to promote our customer services.**
- 2. Strengthening the customer service department can be helpful also.**
- 3. We can take Feedback from customer who called customer services.**

Conclusion and Summary

- In the telecom churn analysis, we have 51 states and 3 Area codes. In this EDA capstone project, we were provided various data sets like ('State', 'Account length', 'Area code', 'International plan', 'Voice mail plan', 'Number email 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',
- As the first step, perform data wrangling over raw data. We divided the project into seven parts i.e 1. the churn column-wise analysis 2. the state-wise, and 3. Area code-wise, 4. International plan-wise, 5. customer service calls-wise, and 6. Total day, night, and evening calls, and Total day, evening, and night call charges.
- In the churn-wise analysis, after analyzing the churn column, we can say that almost 15% of customers have churned. Now let's see how other features of our data are related to churn.
- In the state-wise analysis, After Analyzing the state column, we can say that CA, NJ, TX, MD, SC, and MI states have the most churn rate of more than 21. The reason for this churn rate from a particular state can be due to the low coverage of the cellular network.

Summary (continued)

- ❑ In the 'Area code' analysis, In the data, we can see there is only 3 common area code, and their churn rate is almost the same. we don't think there is any relation between area code and churn due to the customer leaving the operator.
- ❑ In the 'International plane' analysis, as a conclusion from this analysis, we can see that the number of customers with international plans is 325, however, when 137 of those customers cancel the service, the churn rate is very high at 42.4148614%. Maybe this is the reason why customers leave the service.
- ❑ In the 'customer service calls' analysis, it is observed from the analysis that, mostly because of bad customer service, people tend to leave the operator. The data indicates that for those customers who called the service center 5 times or above the customer churn percentage is higher than 60%, And customers who have called once also have a high churn rate indicating their issue was not solved in the first attempt. So, the operator should work to improve the service call.
- ❑ In Total day, evening, and night calls and their charges analysis, if we try to understand the graphs the x-axis is the total charges and the y-axis is the number of calls during each time of the day. Comparing the charges for calls during the day, evening, and night it can be observed that the charges for calls during the day are very high followed by evening and night. The average rate at which customers are placing calls during the day ranges between 20 - 40, for the evening it is approx. 15 - 19 and for the night is 5.5 - 11.

Summary (continued)

- The rates might be one of the possible reasons for customer churn or let's due to which the customers might opt for companies that offer the same service at a cheaper rate.
- If the company already has a wide range of plans then it might be another possibility that the customers are not aware of those plans.

Some Suggestions;

our suggestions to prevent churn :

- I. Maintaining low charges
- II. Upgrading network to improve services for long duration users.
- III. Improving Pricing Strategies.
- IV. Optimizing and Updating International Call Rates.
- V. Implementing a better network infrastructure in New Jersey, Texas and Maryland Areas where there is more Churn Rate.
- VI. Improvement in the customer service can be done to reduce the factors which cause the churn.
- VII. Decreasing the prices as the talk-time increases can be an effective way to reduce the churn.

From our analysis these are the some points which will help in further improvement in business

THANK YOU