**Telecom Churn Analysis**

**Sampanna Mishra**

**Data Science Trainee**

**AlmaBetter, Bengaluru**

**Abstract:**

Customer churn has emerged as one of the major issues in Telecom Industry. Due to the direct effect on the revenues of the companies in the telecom field, companies are seeking to develop means to understand potential cause of customer churn. Telecom providers need to know the reasons of churn, which can be realized through the knowledge extracted from Telecom data. Orange S.A., formerly France Télécom S.A., is a French multinational telecommunications corporation. The Orange Telecom's Churn Dataset has been considered here for churn analysis.

I have carried out exploratory data analysis on the dataset and analysing various impacts of feature variables on churn rate.

***Keywords: Telecom industry, customer churn, exploratory data analysis***

**1. Problem Statement**

The Orange Telecom's Churn Dataset, used for analysis consists of cleaned customer activity data (features), along with a churn label specifying whether a customer cancelled the subscription. The main objective is to:

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

**2. Introduction**

The deregulation of the telecom industry has increased competition and the situation is only made worse by the fact that customers have more choices than ever. Thus, telecommunication companies should understand their customers’ needs and meet them in order to prevent their flee to the competition. There are many factors that influence Customer to churn.

* Unlike post-paid customers, prepaid customers are not bound by service contracts and they often churn for simplest reasons. Thus, it is quite difficult to predict their churn rate.
* Another factor is customer loyalty that may be determined by customer service and product quality offered by the service providers. Issues like network coverage issues and reception quality may influence customers to move to the competitor with broader reach and better reception quality.
* Other factors that increase probability of customers defecting to the competition include slow or inadequate response to complaints

and billing errors.

* Factors such as packaging prices, inadequate features, and older technology may also cause customers to defect to the competition.

Customers often compare their providers with others and churn to whoever they feel provides better overall value.

**3. Data Exploration**

The dataset consists of 3333 rows and 20 columns. The independent features are as follows:

* State : 51 unique states have been used
* Account Length : Length of the account
* Area code : Code number of areas of some states
* International plan : Yes indicates international plan is present and No indicates no subscription for international plan.
* Voice mail plan: Yes indicates subscription is present and No indicates no subscription for Voice mail.
* Number vmail messages : No.of voice mail measages ranging from 0 to 50.
* Total day minutes :Total number of minutes spent in the morning.
* Total day calls : Total no.of calls made in morning.
* Total day charge: Total charges to the customer in the morning.
* Total eve minutes: Total no.of minutes spent in the evening.
* Total eve calls: Total no.of calls made in evening.
* Total eve charge: Total charges to the customer in the evening.
* Total night minutes: Total number of minutes spent in night.
* Total night calls: Total no.of calls made in night.
* Total night charge: Total charges to the customer in the night.
* Total intl minutes: Total no.of minutes spent in international calls.
* Total intl calls: Total number of calls made to people outside the country.
* Total intl charge: Total charges to the customer making international calls.
* Customer service calls: Number of customer service calls made by customers.

The target variable considered here in analysis is as follows:

* Churn: Denoted by True i.e. customer churned and False i.e. customer retained.

The features can be divided into following data types.

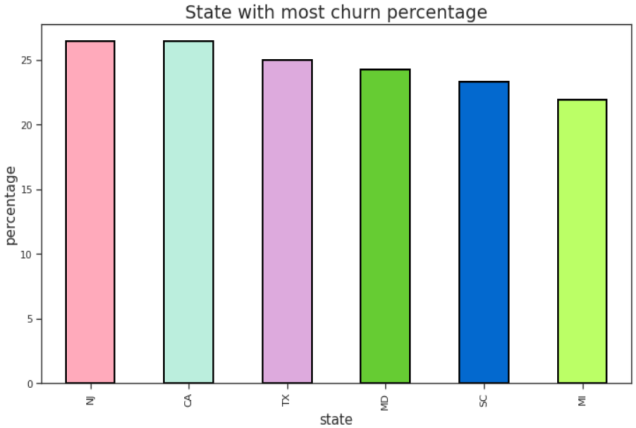
* One Boolean Datatype
* Eight Float Datatype
* Eight Integer Datatype
* Three Object Datatype i.e categorical values are present

**Missing Value:** There was no trace of missing values found.

**Duplicate Value:** There are no duplicate values present in the dataset.

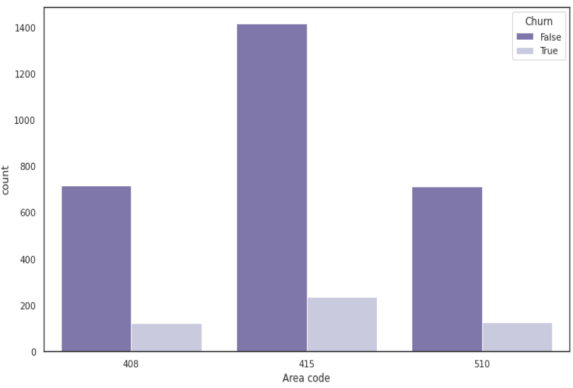
**4. Data Visualization**

* **State vs Churn**

****

A bar plot was graphed between state vs percentage churn. Top five states with highest churn rate is shown in the above graph. The states are NJ, CA, TX, MD, SC, MI.

* **Area Code vs Churn**

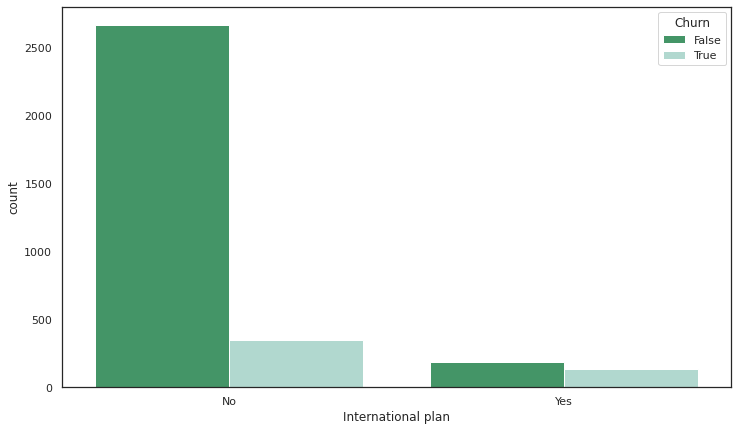


The whole data set has been divided into 3 area codes namely 408, 415, 510. The percentage churn in these area codes is almost similar. This feature doesn’t contribute to churn.

* **Account length vs Churn**

While comparing account length with churn there was no meaningful relation established between the two. Hence account length doesnot impact customer churn.

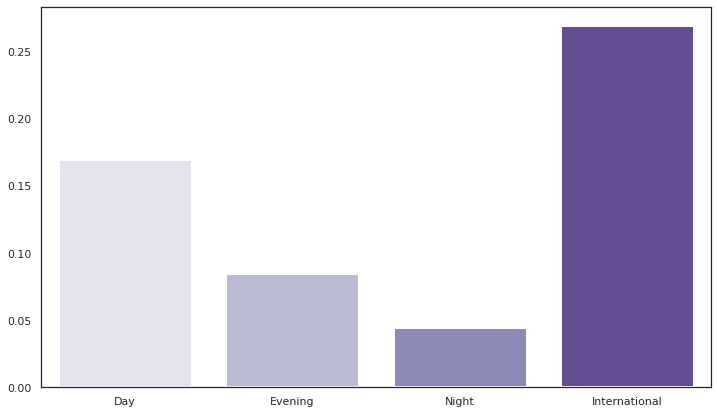
* **International Plan vs Churn**

****

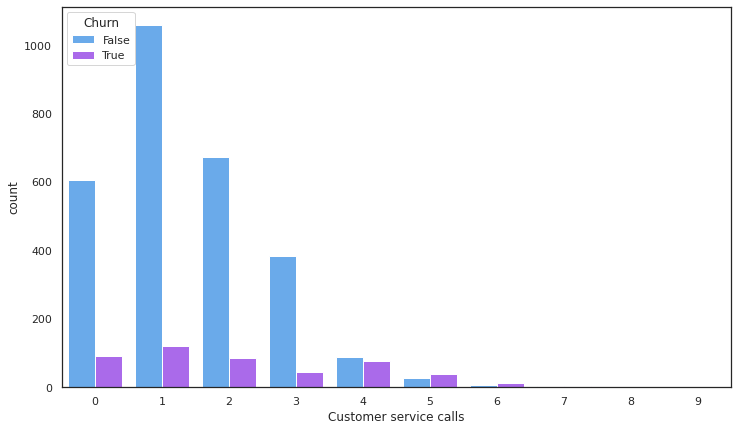
From the above analysis we observe that customers having international plans among those 42.4% churn. Customer who do not have this plan among those only 11.4% churn. So customers who bought international plan churned in big numbers.

* **Total Minutes vs Total Charges**

While analysing Total minutes vs total charges for day, evening, night, international categories it was found that the features share a linear relationship. Further the charges rate was compared among the categories. It can be concluded that charges are highest for international calls.

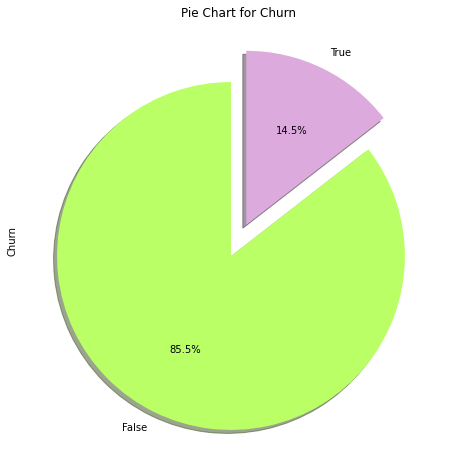


* **Customer service calls vs churn percentage**



From the above grapg it can be observed that with increase in customer center calls the churn percentage increases. There is a higher probability of customer churn when it crosses the 4 customer service calls.

* **Percentage Churn**

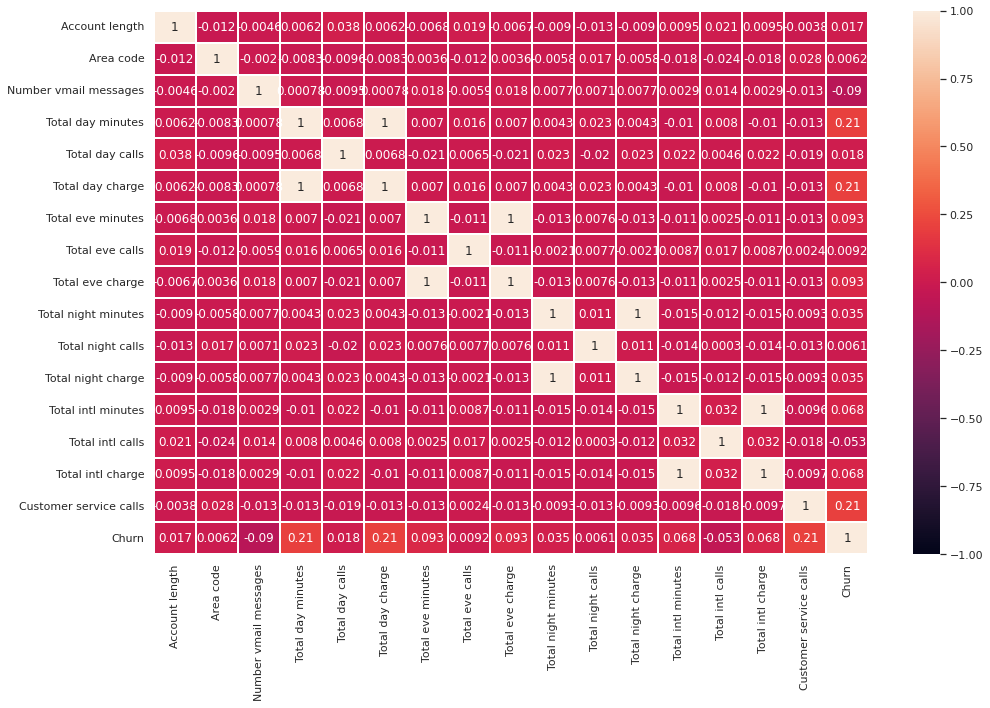


After analysing the Churn column its evident that approximately 15% of customers have churned.

* **Correlation Matrix**

The graph shows correlation heatmap.

It depicts the positive or negative correlation of features with Churn column.



**5. Conclusion**

* Those customers who have International plan churn more and also the international calling charges are also high so the customer who has the plan unsatisfied with network high calls charges.
* In Customer service calls data shows us that whenever an unsatisfied customer called the service centre the churn rate is high, which means the service centre didn't resolve the customer issue.
* Total day call 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, these columns didn't play any kind of role regarding the churn rate.
* There are some states where the churn rate is high as compared to others may be due to low network coverage.
* Area code and Account length do not play any kind of role regarding the churn rate so, it’s redundant data columns.

**6. References**

* towardsdatascience
* analyticsvidhya
* simplilearn