**Telecom churn Prediction**

**Manoj Patil M**

**Gulzar**

**Parikshit Purohit**

**Piyush Mishra**

**Siddhant Mehra**

**Data science trainees,**

**Alma Better, Bangalore**

**Abstract:**

In this project, we working on the customer churn dataset of Orange S.A., formerly France Telecom S.A. in this telecom churn dataset consists of details related to telecom subscription plan of customer, day, evening and night time call charge of the customer, call types, area details, customer service call and customer is churned or not churned details.

First, we did inspection on our dataset, and cleaning of dataset like finding duplicate rows, null values and outlier and did exploratory data analysis on the data set got the observation like correlation between each feature in dataset, how all the features dependent on target column, draw some insights from data visualization and observe why customer from orange S.A. company will cancelled the subscription of telecom service. Suggesting company to what are the factors need to do for control churns.

***Keywords: customer churn, data leaning, data inspection, data visualization***

**1.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

**2.Dataset info:**

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.

**Features in details:**

**State:** this column consists of states details.

**Account length:** active days of account.

Area code: this column having area code details.

**International plan:** it has 2 variables yes – indicate active international plan, no- indicate inactive international plan.

**Voice mail plan:** it has 2 variables yes- indicate active voice mail plan, no- indicate inactive voice mail plan.

**Number of vmail messages:** it shows details of number of voice mail messages of customer.

**Total day minutes:** it shows details of number of minutes usage in day time.

**Total day calls:** it shows details of number of calls made in day time.

**Today day charge:** it shows details of charges of all the calls made in day time.

**Today eve minutes:** it shows details of number of minutes usage in eve time.

**Today eve calls:** it shows details of number of calls made in eve time.

**Today eve charge:** it shows details of charges of all the calls made in eve time.

**Today night minutes:** it shows details of number of minutes usage in night time.

**Today night calls:** it shows details of number of calls made in eve time.

**Today night charge:** it shows details of charges of all the calls made in night time.

**Total intl minutes**: it shows details of intl number minutes usage in intl calls.

**Total intl calls:** it shows details of number of calls made internationally.

**Total intl charge:** it shows details of charges of all the calls made internationally.

**Customer service calls:** it shows details of customer service calls made by customer.

**Customer churn:** it has 2 variables False- customer not churned, True-customer churned.

**2. Introduction:**

The telecommunications industry is a large type of industry with providing different types of services to customers.

These companies face challenges such as customer disruptions., network coverage issues some places and fees for their service. Customer is out of business as the primary cause or reason for the loss or decrease in profits of

The company.

Customer churn means a  
customer who chooses a service that competes with or differs from another  
company because of low prices or for any other 1reason.

So, some companies are taking action to find out why customers unsubscribe from current plan and trying to offer innovative products and services to keep customers

unsubscribing.

Customer management is a strategic customer service management process  
with the goal of building customer loyalty  
Some companies analyze customer data to understand their customer behavior and

Obtain important insights that improve customer service  
There are different factors; and the main

reasons are price of calls, network quality, competition with other companies.

**Enterprise overview:**

To understand the purpose of the project  
it is necessary to understand the business  
Features.

so, we need to understand business  
and then convert its data into an analytical problem definition. Customer retention is the main goal of this type of business and customer retention is all about attracting customers into their business and retaining them in the company.

The company must understand the  
factors that influence customers and cause

them to refuse their service. Determining

the drop outrate will help understand the percentage of customers that give up and also give a general idea of ​​the  
factors that cause them to give up

By using these factors, businesses can predict customer behavior and plan  
accordingly, so having a proper  
churn analytical model is a must for  
businesses to build customer loyalty.

**Reasons for customer churning:**

Main reasons for customer to churned are

1. Competitors: in telecom business we can see high competition will be there it affects customer to churn because telecom industry is highly competitive and companies will try to attract customer by offering discount, attractive telecom plan etc.
2. Service: providing good service very essential for hold the customer in company. Due to poor service, network issues because of this customer will churn and it affects business profit of the company.
3. Customer service: customer service plays important role in holding customer in company, if customer service is good customer churn rate is very less.

**3.EDA:**

Exploratory data analysis is a method with help of this we can understand dataset, we can create some insights from data. We can understand statistics part of the data like mean, mode etc.

With eda we can find null values, missing values, duplicate in the data set and outliers. We can find correlation between features in dataset. In eda we can perform various data visualization method on data and observe what is happening in data.

In our dataset we used 3 steps do eda Data inspection, data cleaning and data visualization.

**Data inspection:**

Before doing data inspection we first import telecom churn dataset into Panda’s data frame. We used 4 libraries in our dataset Pandas for handling data, NumPy for numerical calculation, matplotlib and seaborn for data visualization.

### In data inspection we checked shape of data set it is 3333\*20,Checking all the data types of telecom churn dataset holds. In telecom churn dataset object, float, int, bool data types are present, checking info dataset it gives summary of data including data types, null values count and memory usage.

in telecom churn dataset 2 column types are present numerical and categorical features. And target column in the dataset is churn column.to know more about the characteristics of the dataset used describe method it will give all the stats information of data

### from the statistics part of our data, we can observe

### \* Avg length of account of customer is 101

### \* Avg total time spend on calls in even is highest compare to night and day

### \* International call, mins and cost is very less compared to other calls

### \* There are 51 states in our dataset and WV has highest number of customers

### \* Around 90% customers don’t prefer international plan

### \* Around 72% customers don’t prefer voice mail plan

### \* Around 85% customers don’t cancel subscription of telecom service

**Data cleaning:**

Data cleaning is very important process in EDA. because raw data sometimes it consists 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.

### In our dataset first check the null values, dealing with null values is very important because Missing data in the training data set can reduce the power / fit of a model or can lead to a biased model because we have not analyzed the behavior and relationship with other variables correctly. It can lead to wrong prediction.

### Missing values are the unknown values in the dataset. the concept of missing values is important to understand in order to successfully manage data. first step is to detect the missing value in the dataset and then treat them using the appropriate method.

In Orange SA telecommunication dataset, there is no null values.

Checking duplicate values important thing in EDA.in our data set there is no duplicate values.

### Checking outlier in the dataset because Outliers is also something that we should be aware of. Why? Because outliers can markedly affect our models and can be a valuable source of information, providing us insights about specific behaviors. Outliers is a complex subject and it deserves more attention. Here, we'll just do a quick analysis on data set to find outliers. in telecom churn data set all the columns having outliers.

we need few more columns for better observing data. We created total charge, mins, calls of the whole day

**Data visualization:**

### Now it's time to visualize the data through matplotlib and seaborn library. we can able to see how the data appears and what kind of relation between the data with help of data visualization.

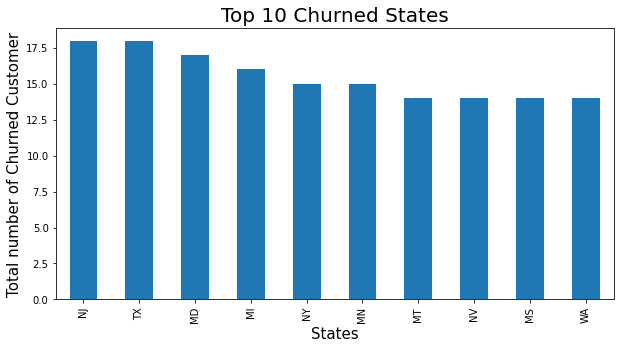
## 1.Target variable

### target variable in telecom churn dataset has 2 categories False or True

### This plot shows a no of churners and non-churners we can observe count of churners around 500, and non-churners more than 2500.

### From above calculations we can see that our almost 14% customer has been churned. We can definitely do something about that after analyzing the data.

# **2.comparing Target column with states**



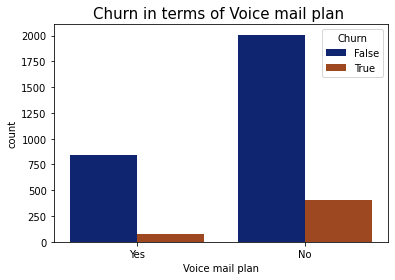
### From above bar plot we can observe that users of these states are prone to churn.

### 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.

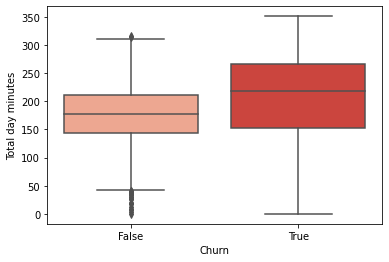
### 2. We have to look out for our competitor in these states also as observation says that there may be strong competition.

**3.churn distribution vs voice mail plan**



### From above 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.

# **4.Relation between Churn and Total day minutes**



### From above box-plot we can observe that users who are spending 225 minutes or more (approx. 4hrs) tend to switch to another operator.

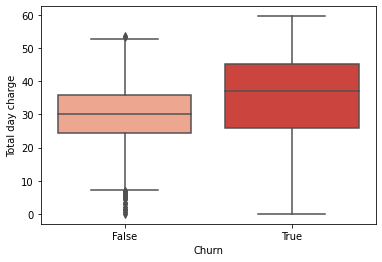
### According to my hypothesis, following would be the factors that should be implemented:

### 1. Network Disturbance during a Call.

### 2. Need to Upgrade or make smarter use of technologies like VoLTE for improvement of Voice Quality during calls.

### 3. Network Upgradation.

# **5.Relation between Churn and Total day charge**



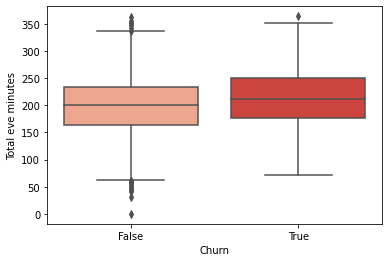
### From observing above boxplot, we can clearly indicate a good strategy to be implemented. We can say that Customers having more day minutes spent on the network tend to leave its subscription and from the above box-plot it clearly indicates that there is defect in the pricing strategy of the company.

### According to my Hypothesis:

### 1. Strategy of pricing needs to be re-evaluated.

### 2. we can give discounts for the customers who have high spending of day minutes.

# **6.Relation between Churn and Total eve minutes**



### From above box-plot we can observe that users who are spending 210 minutes or more tend to switch to another operator.

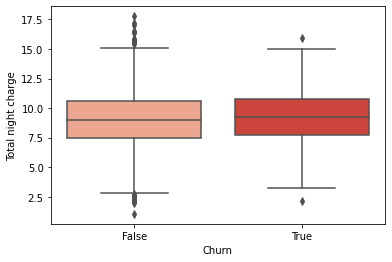
### According to my hypothesis, again following would be the factors that should be implemented:

### 1. Network Disturbance during a Call.

### 2. Need to Upgrade or make smarter use of technologies like VoLTE for improvement of Voice Quality during calls.

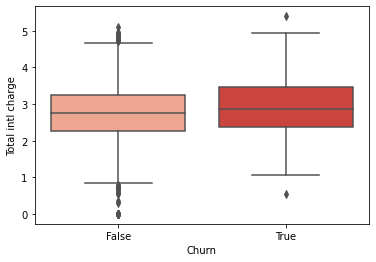
### 3. Network Upgradation.

# **7.Relation between Churn and Total night charge**



### From observing boxplots of night charges, we can say that Customers having more night minutes spent on the network and most night charge tend to leave its subscription and from the above box-plot it clearly indicates that there is defect in the pricing strategy of the company and we can give discounts for the customers who have high night call minutes spending.

# **8.Relation between Churn and Total international charge**



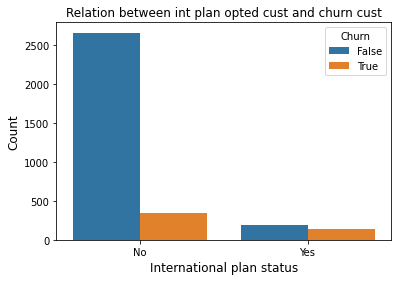
### From above box plot we can say that customer who are churned having more intl calls

### According to my hypothesis, following would be the factors that should be implemented:

### 1.Implementing Different Pricing Strategy

### 2.Implementing international Calling Rate Optimization would need to lower churn rate.

# **9.Relation between Churn and International plan**



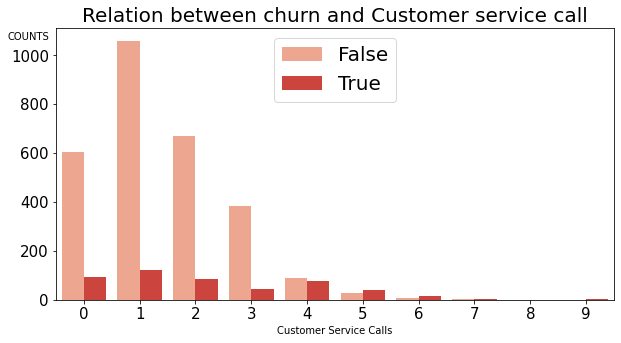
### The ratio between churned user who opted for international plan and who don't is almost 50-50. It means there is some problem with the pricing or voice call quality for international plan opted users.

### According to my hypothesis 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.

# **10.Relation between Churn and Customer service**



### After observing the above plot, we can say that some customers are lazy and hence without resolving the issue they have jumped to other network operator, while the customers who have called once also have high churn rate indicating their issue was not solved in first attempt.

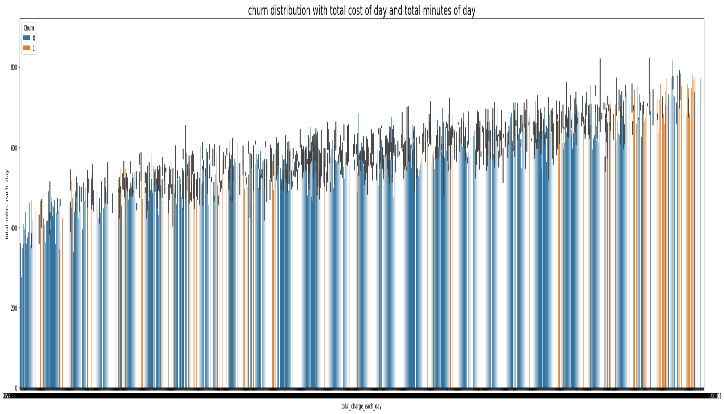
### According to my hypothesis, following would be the factors that should be implemented:

### 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.

# **11.churn distribution with total cost of day and total minutes of day**



### we can clearly observe that with increases in total minutes directly correlate to increases in total cost, majority of customer having higher total cost of each day will churn

### According to my hypothesis, following would be the factors that should be implemented:

### 1. Strategy of pricing needs to be re-evaluated.

### 2. we can give discounts for the customers who have high spending of each day minutes.

# **12.Checking correlation of all the features in our dataset**



### we can observe from heat map all the charge, mins, calls columns of day, eve, night and each day total are highly correlated with each other

### customer service calls, total mins and charge of each day is highly correlated to churn feature

# **Conclusion:**

### In this project, we tried to analyze customer churn. First, we did inspection of dataset on a basic level. We looked for missing values and check the outlier.

### Then we used the matplotlib and seaborn to do Exploratory Data Analysis on sample data by plotting different graphs like count plot, pie chart, lmplot, bar plot, boxplot, subplot and heat map from this we got useful insights like: customer having more daily charge will be more chances of churn, states like New Jersey, Texas and Maryland have higher churn rate, customer having international plan have more churn rate, customer having less customer service call have more churn rate and 14% percentage of customer who has been churned

Here are some suggestions to prevent churn

We can clearly see some states having very high churn rate might be because of network issues or very high competition for other telecom company so implementing better networks, upgrading network in these areas and providing offers for customer will leads to reduce churn rate.

Customer having higher day mins, day charge has more churn rate maybe because of poor network quality, or high cost so providing discount or implementing new pricing strategy and improving network quality will reduce customer churn rate.

Customer with international plan has more churn rate maybe because of cost and network fluctuation, network quality so implementing new pricing strategy, providing discount and upgrading network will reduce churn rate.

We can observe poor customer service because of this high churn rate. So, providing better customer service, promoting customer service, collecting feedback from customer these are the factors company can implement to reduce customer churn.

Orange S.A telecom company has to implement these are the ideas and business suggestion to reduce churn rate and it will increase profit of the company also.

Further more Improvement we can suggest and discuss more strategies to the company by collecting other data and through a domain expert

**Reference:**

1. Analytics Vidhya
2. Medium
3. Kaggle
4. GitHub