|  |
| --- |
| The orange S.A. |
| **Telecom Churn Analysis** |
| **Gulzar** |
| **Data science trainees,**  **AlmaBetter, Bangaluru** |
|  |
| **5/7/2022** |

|  |
| --- |
| In this project, we will be working on The Orange Telecom's Churn Dataset. The Orange S.A. is an a French multinational telecommunications corporation. Here we will be doing a complete EDA process to determine the major causes of customer churning from particular telecom company meanwhile we will draw some insights from data visualization and analysis, so that we could get the factors which will affect the output i.e. churn of the customer because customer churn is a major problem and one of the most important concerns for large companies. |

**1.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

The main objective is to do some analysis, which could help them in findings the key factors responsible for customer churn. This would in turn help them to ensure customer retention quickly and efficiently.

Below are the info that is available in given dataset-

* State- Having abbreviation name of USA states only
* International plan - A check for international plan
* Voice mail plan - A check for voicemail plan.
* Number vmail messages - Number of voicemail messages sent by customers who opted for voice mail plan.
* Total day minutes - Having total of minutes which a customer spent in a day-time
* Total day calls - Having total number of calls of a customer in a day-time
* Total day charge - Having total of charges of a customer's spending in a day-time
* Total eve minutes - Having total of minutes which a customer spent in a evening-time
* Total eve calls - Having total number of calls of a customer in a evening-time
* Total eve charge - Having total of charges of a customer's spending in a evening-time
* Total night minutes - Having total of minutes which a customer spent in a night-time
* Total night calls - Having total number of calls of a customer in a night-time
* Total night charge - Having total of charges of a customer's spending in a night-time
* Total intl minutes - Having total of minutes which a customer spent on international calls
* Total intl calls - Having total number of international calls of a customer
* Total intl charge - Having total of charges of a customer's spending on international calls
* Customer service calls - Having number of calls made by a particular customer to customer service centre
* Churn - Having churned and non-churned status of customers

**2. Introduction**

The telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues:

1. Acquire new customers,

2. Upsell the existing customers,

3. Increase the retention period of customers.

However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy, proves that retaining an existing customer costs much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer’s churn, known as “the customer movement from one provider to another” .

Our goal here is to find the key factors responsible for customer churn, based on that we can make some suggestions for customer retention.

**3. Steps involved**

* **Observing and Exploring Dataset**

After loading the dataset we performed some basic functions and methods for knowing the data type. This process helped us figuring out various aspects and relationships among the users and the data factors. It gave us a better idea of which factor behaves in which manner compared to the churner.

After observing the data we would say that there are 20 columns and 3333 rows. According to churn column there are 86% non churned and 14% churned users. We can definitely make some suggestions for churned users for their retention.

* **Null, Missing and Duplicate Values Treatment**

It is an important aspect of Data Cleaning because there can be some null, missing and duplicate values in our dataset. But our dataset doesn't contains a null or missing values which might tend to disturb our accuracy, if it has null or missing values then we have to drop them at the beginning of our project in order to get a better results.

* **Dropping unneccessary Column**

Our dataset does contains some unnecessary column like Account length etc. we didn't need them for analysis so we can drop them out from dataset.

* **Encoding of categorical columns**

We changed no and yes to 0 and 1 to encode our categorical column because categorical column that are in string format cannot be understood by the machine and needs to be converted to numerical format for visualization and good analysis.

* **Analyzing and Visualization of data**

In these steps we used plots like count plot to check the results of each churned customer and we used box plot for observing which column is more important compared to churned column and which is of less importance.

Next we used bar plot for finding top states in which customer have churned a lot.We are using pie plot and cat plot also as well as heat map for correlation.

* **Using different visualization methods**

For visualization we tried various plots like:

* **Count Plot**
* **Bar Plot**
* **Pie Plot**
* **Box Plot**
* **Cat Plot**
* **Heat map**

**4. Conclusion**

That's it! We have reached the end of our exercise.

Starting with loading the data so far we have done EDA , null values treatment, encoding of categorical columns, plots and charts selection and then visualization/predictions building.

So according to our analysis the major cause of churning is-

* Same Pricing Strategies for all
* Network Disturbance in some States
* Bad Quality of Customer Service
* Higher International Call Rates

**References-**

* Stackoverflow
* GeeksforGeeks
* Kaggle