**Telecom Churn Analysis**

**Harshjyot Singh Chawla,**

**Ajinkya Dakhale,**

**Bhaskar Subanji,**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

The base for any business to succeed is the customer and that is why firms know the importance of achieving the satisfaction of customers. Customer churn is an essential issue and it is considered as one of the most important concerns among firms because of the increased significance of marketing strategies, increasing competition among firms, and customer's conscious behaviour in present years.

Customers are migrating from one service provider to another provider because of good services or rates or various advantages provided by rival firms. Gaining new customers requires a huge amount of cost therefore prediction of customer churn has become an important part of the planning and strategy process in the telecom sector.

The main aim of our project is to initiate a customer churn analysis using exploratory data analysis which will help in analyzing Telecom’s dataset and how these results are essential in reducing customer churn and improving customer service.

***Keywords- Churn prediction, customer churn, exploratory data analysis, Retention, Telecommunication***

**1.Problem Statement:**

French multinational telecommunications corporation, Orange S.A. Telecom's Churn Dataset has been evaluated to find out the reasons for customers to churn. The dataset contains customer activity data as well as churn variable stating retained customers. Factors responsible for churning, variables correlated to churning, recognizing customers that might churn will be analyzed in this project.

The goal of the project is to maximize the company's profit by retaining customer and minimize customer churn by identifying the key cause of the problem.

**Exploring dataset:**

Exploring the dataset phase begins with an initial data collection and proceeds with activities in order to get familiar with the data. Identifying data quality problems, discovering first insights into the data and detecting interesting subsets to form hypotheses from hidden information are activities of this step. Data which is collected from a telecommunication company to get analysed, involves usage details of customers from.

**Data Description:**

The data was taken from Orange Telecom Company. It had 3333 rows and 20 columns. Most columns are related to subscriber personnel. Other column was indicative of service usage by the subscriber. Based on the business understanding of the data all the columns were taken into consideration to analyse the data The feature ‘Churn’ shows customer churn or non-churn based on existing conditions. Approximately 14.5% are churn and 84.5% are non churn.

**Dataset Variables:**

* **STATE**: 51 Unique States
* **Account Length:** Duration of length customer use their The Account
* **Area Code:** There are 3 unique area code present 415, 408, and 510
* **International Plan:** Yes Indicate International Plan is Present and No Indicates no subscription for International Plan
* **Voicemail Plan:** Yes Indicates Voicemail Plan is Present and No indicates no subscription for VoiceMail Plan
* **Number vmail messages:** Number of Voicemail Messages ranging from 0 to 50
* **Total day minutes:** Total Number of Minutes Spent by Customers in Morning
* **Total day calls:** Total Number of Calls made by Customer in Morning.
* **Total day charge:** Total Charge to the Customers in Morning.
* **Total eve minutes:** Total Number of Minutes Spent by Customers in Evening
* **Total eve calls:** Total Number of Calls made by Customer in Evening.
* **Total eve charge:** Total Charge to the Customers in Morning.
* **Total night minutes:** Total Number of Minutes Spent by Customers in the Night.
* **Total night calls:** Total Number of Calls made by Customer at Night.
* **Total night charge:** Total Charge to the Customers in Night.
* **Churn:** Churning status of the customer

**2.Introduction:**

The telecommunications industry across the world is becoming one of the major sectors Telecom firms are making an effort to survive in this rivalry market and some measures have been formulated to bring in huge amounts of revenues. To enhance the retention time of customers it is important for the companies to lessen the possibility of churn of customers, referred to as “the movement of customers from one service provider to another service provider. The churn of customers’ is considered a major issue in service fields with increased cutthroat services.

The underlying principle of customer churn prediction in terms of telecom industry is to  
calculate subscribers approximately who literally feel like to leave from a company they used so far and suggest solutions to prevent considerable churns.

For acquiring every new customer, the cost associated with this (CAC) tends to be far more than that associated with keeping a customer into the system of a company. Moreover, by keeping overall customer management costs low, a company can focus it’s time and effort into improving the features of their product and the quality of their services, which will ultimately improve the health of the company financially.The major role that the telecom industry plays made it all the more significant to build prediction mechanisms alongside the lines of churn prediction.

**Business Overview:**

The starting point of data analysis focuses on understanding the objectives of the project and requirements from a business point of view, and then converting this knowledge into a data analysis problem definition. Customer retention consists of “Identifying which customers are likely to Churn, determining which customers should retain and developing strategies to retain profitable customers”. The main thing in the retention process is identifying the Churn ratio which is a very meaningful and vital determination for many companies. Determination of Churn ratio indicators is also very important. By using those indicators, firms can make predictions on future behaviour of new customers and can develop new strategies much before customers start to think about churn. Thus, it is vital to build a very successful and accurate Churn model during the retention studies.

**Customer Churn:**

Customer churn is the percentage of customers that stopped using your company's product or service during a certain time frame. The churn rate, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity. It is most commonly expressed as the percentage of service subscribers who discontinue their subscriptions within a given time period

A high churn rate could adversely affect [profits](https://www.investopedia.com/terms/p/profit.asp) and impede growth. Churn rate is an important factor in the telecommunications industry. In most areas, many of these companies compete, making it easy for people to transfer from one provider to another.

The churn rate not only includes when customers switch carriers but also includes when customers terminate service without switching. This measurement is most valuable in subscriber-based businesses in which subscription fees comprise most of the revenues.

**Reasons for customer churning**:

* **Service quality** - Lack of network coverage may make a customer go to another company with good network coverage.
* **Features and content availability-** It occurs in marketing or sales when a customer signs up for a plan that isn’t best suited to their needs. Unfortunately, such a decision most often ends up in terminating the contract. Therefore, it’s crucial to measure your customer churn rate for specific services or bundles to identify an acquisition problem.
* **Lower cost substitutes from competitors-** Telcos are always in a state of instability where even the slightest change, such as pricing changes or new product launches, can cause customers to try another service provider. In addition, from the awarding of new telco licenses introducing new competitors to higher taxes borne by the consumer, these factors also contribute to customer churn in the telecom sector.
* **Negative customer service experiences**-Slow or no response to customer complaints makes a customer more likely to churn.
* **Price-** comparatively high Pricing leads the customers to flee from one carrier to a competitor.
* **Billing disputes**
* **Competitors introducing new products or technology.**

**3.Approach:**

**Exploratory Data Analysis:**

Exploratory data analysis is a process in which we try to understand the given data in a possible way, so that we can get some insights out of it. Using the exploratory data analysis we understand the important factors or characteristics such as Avg, mean Std deviations etc also to verify the missing values or null values and outliers. Exploratory data analysis is a process of verifying the available data set to determine patterns, anomalies, test hypotheses, and check assumptions using statistical measures. Using python in exploratory data analysis process and visualization comparison between the variables is easy to understand and get the insights.

**Analysis of Data :**

Analysis of data is an essential step, which deals with descriptive statistics and analysis of the data. This step involves summarizing the data and detecting the unseen relation and effects in between the different dataset, which helps to develop and predict the models, evaluate them and define the factors with much accuracy. Procedure/steps used for data summarization are using and application of summary columns, graphs, descriptive and inferential statistics, correlation statistics, searching, grouping them as well as math's models.

**Sourcing of Data**

Data sourcing is the method of finding and storing it into our machines or systems. There are many ways to find the data and it must be handled properly and in the correct format. Data should be handled by a skilled and authorized person of the respective company. There are many tools to find the data and to collect and store them.

**Data Processing & Cleaning :**

Raw data sometimes consists of noise, null values, also inconsistent format and values, so pre-processing of data is important to enhance the quality of data. Then only we can perform the Data cleaning easily which is very important to find the irregularities (such as NullValues, incorrect format & Headers Anomalies/ Outliers etc)Then it is easy to clean the data in order to get some useful insights.

**Transformation of Data :**

Transformation of data is the process in which the data will be further improved to gain the performance and clarity of data. Sometimes data contains duplicate rows and values. Deleting duplicate values is important to improve the quality of the dataset.

**Missing values:**

In datasets missing values occur due to many reasons such as errors, or handling errors in data. Sometimes some customer is not subscribed to all of its service and plans so in respective columns or rows there may be the possibilities of missing values in some product representation columns. In this Orange SA telecommunication dataset we don’t have any Null values. So we need to handle the missing values, if any present in the dataset before conducting any sort of analysis.

**Univariate analysis, Bivariate analysis & Multivariate analysis:**

**Univariate analysis**

We analyzed data of a single variable/column from a dataset, also known as Univariate Analysis. In Univariate analysis we take one feature at a time. Where we analyse a feature independently, usually primary motto of this analysis is to find the distribution of its values(range) and ignore other features in the dataset

Univariate analysis is the simplest form of data analysis. The data must consist of only one type of variable and that we perform analysis over it. The main purpose of univariate analysis is to take data, summarize that data, and find patterns among the values. It doesn't deal with causes or relationships between the values.

Graphical methods we used for this are

- Pie Chart

- Distplot

**Bivariate analysis**

We analyze data by taking two variables/columns into consideration from a dataset, known as Bivariate Analysis. Here most of the time we kept one variable constant that is Churn and changed other variables for each column label.

Graphical methods we used for this are

- countplot

- Boxplot

- Scatterplot

**Multivariate analysis**

In Multivariate analysis we analyse three or more variables. This allows us to look at correlations (that is, how one variable changes with respect to another) and enables us to understand the correlation and amongst each other and their behaviour more accurately than with bivariate analysis.

One common way of plotting multivariate data is to make a pair plot. Here we used Heat map to find correlation among all features (column label) present in the dataset.

**4.Conclusion :**

In the highly competitive market of the telecom industry, customer retention becomes the most important concern for every telecom company. Service standardization and good quality customer policies allow customers to change their preference from one service provider to another service provider.It has become economically feasible to retain existing customer then to acquire new customers.

We have performed univariate,bivariate and multivariate analysis on each variable of Telecom's dataset and realized some variables are correlated to churn variable whereas some variables do not share a strong relation with the churn variable.

On the basis of our analysis we can conclude that the reasons for churn are network coverage issues, high tariff and no proper solution provided by customer service calls to the customer.

**5.REFERENCES:**

* Data science for business : what you think about data mining
* GeeksforGeeks
* Analytics Vidhya