**Telecom Churn Analysis**

Puneet subanji,

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

The telecom churn dataset includes information about customers with various plans, their call durations, call types, and area information, as well as calls made to the customer service centre, etc. Exploratory data analysis has been carried out on these datasets to understand the cognizance of the data by understanding the corresponding relationship between the features and identifying the crucial performance aspects.

Use visualizations techniques to obtain some insights and make informed judgments to  retain more customers by identifying the issues that cause them to churn from the service and recommending ways for the company to improve its service and retain the customers.

*Keywords- Churn analysis, customer churn, exploratory data analysis, Retention, EDA.*

**1. INTRODUCTION:**

The telecommunications sector is large industry and offers a variety of services to customers through a service subscription model. These companies struggle with problems like high service costs, network coverage concerns in a specific area, and customer churn.The primary factor or explanation for a significant portion of a company's losses or a decline in earnings is customer churn.

A consumer choosing a rival or different company's service because of a low tariff or for any other reason is referred to as customer churn.Thus, some companies take steps to identify the causes of customer churn and make innovative services and products to retain consumers from leaving.

Customer management is a tactical approach to overseeing customer service with the goal of retaining customers. Some companies investigate and examine customer data to understand consumer behaviour and get vital information that aids in enhancing customer care.

There are many causes of churn, but the three main ones are  service quality, tariff rate, and aggressive competition from other companies..

**BUSINESS OVERVIEW**

It is essential to understand the business in order to understand the project's goal. The service offered to clients with their plans and features is where the majority of the profit in the  telecommunication industry. In order to translate their data into analysis and problem description, we first need to understand the business.These types of businesses prioritise customer retention, and customer retention entails engaging the customer with

their business and to retain them from churn. Businesses must recognize the factors influencing their consumers and pushing them to churn from their services. Knowing the churn ratio will provide you a better understanding of the percentage of churned customers as well as a general concept of the reasons why they left. Companies can forecast customer behaviour and make plans based on these characteristics, therefore having an appropriate churn analysis model is crucial for them to retain their consumers.

**CUSTOMER CHURN:**

If a consumer unsubscribes a membership with one company and becomes a customer of another company, this customer is known as a Churn customer. Rate reductions are not the only strategy to increase customer loyalty, as evidenced by the present business trend. additionally, providing value-added service to the products has become an industry trend or rule to have gain and retain a customer. The major goal of this study is to identify a client who will probably churn and calculate the cost of retaining them.While performing the analysis, the most essential point is to define the churned customer.

For companies, losing a customer is a major issue. Companies that provide services, such as those in finance, banking, insurance, and telecommunications,etc are likely to lose clients. The cost of acquiring a new consumer base is significantly rising for organizations. Therefore, keeping existing consumers is essential and much more crucial than acquiring new ones.

**REASONS FOR CUSTOMER CHURNING:**

We can observe some of the main reasons or causes for customer churn.

**Tariff rate:** comparatively high tariff or low tariff from competitors affects customers to churn due to intense competition in the telecom industry, companies continually work to introduce new products with lower tariff or customized services to attract  customers.

**Service quality:** Any business needs to provide excellent customer service in order to keep customers and attract new ones.Lack of network coverage or network with issues.Also, the service quality of the company's products is impacted by the slow resolution of these issues, which may result in customer churn.

**customer service:** Providing excellent customer service will always result in increase in the customers, and vice versa. Sometimes not speedily resolving customer issues may also lead to churn.

**Plans or offers**: Customers leave services in droves as a result of competitive plans and remarkable alternatives offered by other companies. Therefore, it's important to give customers more options and offers so they can pay for and use the service. Otherwise, users would cancel their subscriptions to the services and switch to another subscription where they may pay less money or receive the same features in their Plans at a lower price.

**2. Problem Statement**

Identifying the elements that cause customers to churn from the service and recommending the causes to the company to improve its service and retain customers will help you identify the factors to retain customers.

The main focus of the project is to: Identifying the factors that lead to customer churn. Retain customers by taking the proper actions while taking into account of impacting factors and using derived factors, This document focuses on data analysis and identifying factors that are associated with churn., which will give insights about customer, who may churn.

**ABOUT DATA:**

Collected dataset from telecom operator Orange SA, which includes information on how customers use its services as well as records of their locations and plans. It consists of 3333rows and 20 columns.The variable 'Churn' displays customer churn or non-churn based on existing conditions. Approximately 14.5% are churn and the remaining are non-churn.

**VARIABLES IN DETAIL**:

**State**: States name in code.

**Account Length**: Active period of Account

**Area Code:** Area code having States

**International Plan:** Yes: indicates active international plan user and, No: indicates inactive international plan user.

**Voice Mail Plan**: Yes: indicates Active voice mail plan user, No: indicates inactive voice mail, plan user.

**Number of vmail messages**: Number of voice mail Messages

**Total day minutes**: Total number of minutes usage in the morning

**Total day calls:** Total number of calls made in the morning.

**Total day charge**: Total charges for all calls made in a day.

**Total eve minutes:** Total number of minutes usage in the evening

**Total eve calls**: Total number of calls made in an evening.

**Total eve charge:** Total charge for all the calls made in an evening.

**Total night minutes**: Total number of minutes usage in the night.

**Total night calls**: Total number of calls made at night.

**Total night charge**: Total charge for the call made in the night.

**Total intl minutes:** Total Number of minutes usage in international calls.

**Total intl calls:** Total number of calls made internationally.

**Total intl charge**: Total charge for all the international calls.

**Customer service calls**: Number of customer service calls made by the customer

**Churn customer:-**

True : churned customer

False: retained customer

* **APPROACH:**

**EXPLORATORY DATA ANALYSIS:**

Exploratory data analysis is a procedure where we attempt to comprehend the provided data in the best way possible in order to gain some insights from it. Using the exploratory data analysis, we may discover the crucial elements or features, such as the average, mean, and standard deviations, etc also to verify the missing values or null values and outliers.

Exploratory data analysis is the process of examining the data set that is already available to identify patterns, anomalies, test hypotheses, and validate assumptions.It is simple to understand and gain insights by using Python in the exploratory data analysis process and visually comparison between the variables.

**ANALYSIS OF DATA:**

The step of data analysis, which involves both descriptive statistics and data analysis, is crucial.In this step, the data are summarised and unknown relationships and effects between the various datasets are found, assisting in the development and prediction of the models, With great accuracy, evaluate them and specify the factors.Utilizing summary columns, graphs, descriptive and inferential statistics, correlation statistics, searching, grouping, and mathematical models are some of the stages or procedures used to summarise data.

**SOURCING OF DATA**

The process of locating and storing data in our devices or systems is known as data sourcing. There are numerous ways to find the data, which must be treated carefully and presented in the right format.A knowledgeable and authorised employee of the relevant company should handle data. There are numerous tools available to locate, gather, and save data.

**DATA PROCESSING & CLEANING:**

Pre-processing of data is essential to improve the quality of the data because raw data can occasionally contain noise, null values, inconsistent format, and inconsistent values. Only then would it be simple for us to complete the data cleaning, which is crucial for identifying inconsistencies (such as NullValues, incorrect format & Headers Anomalies/Outliers etc). After that, cleaning the data to gain some insightful information is simple.

**TRANSFORMATION OF DATA:**

Data transformation is the process through which the data is further enhanced to increase performance and clarity. Data can occasionally have duplicate rows and values. To enhance the quality of the dataset, duplicate values must be removed.

**MISSING VALUES:**

Missing values can occur in datasets for a number of reasons, including errors or handling errors in data. There are occasions when a client does not subscribe to all of their services and plans, and as a result, some product representation columns may have missing values in the corresponding columns. There are no Null values in this Orange SA telecommunication dataset. Therefore, before performing any kind of analysis, we need to deal with any missing values that may be present in the dataset.

**DROPPING OF MISSING VALUES:**

We can employ a variety of strategies and then put them to use to eliminate the missing values. For instance, the Panda library's IsNull() and not-null() functions can be used to identify null values and get rid of them using dropna. Also outliers are also handled based on how they affect our dataset.So that we can eliminate the missing values from the dataset, we can continue with exploratory analysis.

**UNIVARIATE ANALYSIS, BIVARIATE ANALYSIS & MULTIVARIATE ANALYSIS:**

**Univariate analysis**

In a process called as univariate analysis, we looked at data from a dataset that only had one variable or column. In Univariate analysis, we take one feature at a time. Where a feature is analysed individually, the main goal of the analysis is typically to determine the distribution of its values (range) and to ignore other features in the dataset. The simplest type of data analysis is called a univariate analysis. We must conduct analysis on the data, which must only contain one sort of variable. Univariate analysis is mostly used to acquire data, summarise that data, and identify trends among the results. It doesn't discuss the reasons behind the values or their relationships.

Graphical methods we used for this are

- Donutchart

- Distplot

**Bivariate analysis**

Bivariate Analysis is the process of analysing data by considering two variables or columns from a dataset. In this case, the majority of the time, we kept the Churn variable constant and altered another variable for each column label.

Graphical methods we used for this are

- Count plot

- Boxplot

- Scatterplot

**Multivariate analysis**

In Multivariate analysis, we analyze three or more variables. This enables us to examine correlations. (that is, how one variable changes concerning another) and allows us to more accurately identify the correlation and between each other and their behaviour than using with bivariate analysis.Making a pair plot is one frequent method of visualising multivariate data. Here, we used a Heat map to determine whether any features were correlated (column label) present in the dataset

**CORRELATION** :

Correlation aids in establishing the relationships between the variables and provides crucial information, such as how the variables interact and what are the factors that are having a high impact on each other as well churn of the customer and helps the company to work on these factors. Also, It aids in making wise decisions and accurate predictions and is highly helpful for future or visionary prediction of company products.

**GRAPHICAL REPRESENTATION OF THE RESULTS:**

In this step, the intended audience is presented with the studied dataset using visualisation techniques such as graphs, summary tables, plots, drawings, and histograms.As the primary responsibility of EDA, this phase is crucial and simple for business stakeholders to understand. Most of the graphical analysis techniques include charts like Table charts, Polar charts, histograms, Line charts, Bar charts, Scatter plots, Area plots, Pie charts, Donut charts etc.

**CONCLUSIONS:**

Some states have higher churn rate than other, for which network issues could be the reason because if the contender company had low tariff for calls also utmost of the states would have shown the appprox same churn rate. Customers with international plan ON has higher churn rate compared to customerswith international plan OFF , this could be because the customer could be unhappy with the high tariff cost or network issues. Area and Account length has no relation with churn rate, hence this columns can be neglected or it can be said that the data is superfluous. Customers with higher day call mins has higher churn rate compared to other, could be because of the higher charges which is relatively obvious, frequent caller might have introduce some other company offering low tariff. With other varaibles like as evening, night calls no relation could be found. It could be seen that customers with vmails more than 20 (approx.) has higher churn rate. The churn rate increases as the call to the service center increases. Customers who have called customer service three or lesser times have a markedly lesser churn rate than that of customers who have called customer service four or more times.

**REFERENCES:**

<https://medium.com/analytics-vidhya/telecom-churn-prediction>

<https://journalofbigdata.springeropen.com/articls/10.1186/s40537-019-0191-6>

Hands-On Exploratory Data Analysis with Python Perform EDA techniques to understand, summarize, and investigate your data by Suresh Kumar Mukhiya, Usman Ahmed (z-lib.org)

<https://bunker2.zlibcdn.com/dtoken/01c5fc197a94283bfb0c0943bd5b2d0c>