

# Telecom churn analysis

BAISHNAVEE MAHATO

Data science trainee,  
AlmaBetter, Bangalore

## Abstract:

Telecom churn dataset includes details regarding customers with different plans and their call hours, call types and area details, as well as calls made to the customer service centre, etc., so exploratory data analysis has been conducted on these datasets to fully comprehend the cognizance of the data by knowing the corresponding relationship between the features and finding the important performance aspects.

Utilize visualisation approaches to obtain some interpretations in order to make correct judgments to retain more consumers by defining the variables causing them to churn from the service and recommending the factors to the organisation in order to enhance its service and retain the customers.

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

## 1. INTRODUCTION:

The telecommunications sector is extensive, with several types of services offered to customers via a service subscription model. These businesses encounter obstacles like as client attrition, network coverage issues in a certain location, or service prices. The primary cause or reason for the huge is customer turnover.

Companies suffer a loss of revenue or profits.

Customer churn occurs when a customer switches to a competitor's or a different company's service owing to a cheap tariff or for any other cause.

As a result, some businesses take steps to identify the causes of customer churn and develop creative goods and services to keep consumers from leaving.

Customer management is a deliberate method of managing customer service with the goal of retaining customers. Some businesses investigate and analyze customer data in order to better understand their customers' behavior and gather valuable information that can be used to improve customer care. Churn is caused by a variety of variables, the most important of which are service quality, tariff rate, and competing offers from other organizations.

## OVERVIEW OF BUSINESS

To understand the objective of the project, it is necessary to understand the business.

The major source of profit in the telecommunications sector is the service offered to clients through their plans and features. So we must first comprehend the business before converting their data into an analytical problem statement. Customer retention is the key priority for these types of firms, and customer retention is the process of engaging customers in their business and keeping them from leaving. Businesses must understand the variables influencing clients and prompting them to abandon their services. Identifying the churn ratio will assist in understanding the percentage of turnover customers as well as providing a general concept of the circumstances that caused them to churn. Companies may forecast client behavior and prepare appropriately by employing these elements, thus having a proper churn analysis model for companies to retain their customers.

## **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.

The present business climate demonstrates that rate reductions are not the only method to win consumer loyalty. Furthermore, adding value to products has become an industry trend or norm for gaining and retaining customers. The primary goal of this research is to identify a client who is likely to leave and estimate the cost of bringing them back. The most important aspect of the analysis is defining the churned consumer.

Losing a customer is a major issue for businesses. Finance, banks, insurance, and telecommunications industries, among others, are at risk of losing consumers.

For businesses, the effort to acquire a new customer base can be overwhelming. Retaining customers is therefore necessary and this is much more important than attracting a new customer base.

## **REASONS FOR CHURN:**

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

**Tariff rate:** Telecommunications industry is highly competitive, various companies are always trying to launch different kinds of products with lower tariffs or customized services to attract customers, so relatively high tariffs from competitors. Or low tariffs cause customer churn.

**Service quality:** Service is an integral part of any business and this is a very important factor in keeping your customers in business.

Poor network coverage or network problems.

**customer service:** Good customer service always leads to more customers and vice versa.

In some cases, the inability to resolve customer issues quickly can also lead to churn.

**Plans or offers:** Competitive plans and better alternatives by competing companies lead to the loss of service customers. So offering more flexibility and offers to help you buy and benefit

from our services is always a key factor. Otherwise, the customer can either unsubscribe from the service, unsubscribe and switch to another subscription that can be purchased more cost-effectively, or get the same benefits at a lower price within the plan. increase.

## 2. Problem Statement

Identify the factors that keep customers away by defining the factors that drive them away from your service and suggesting factors for your company to improve its service and retain customers.

The focus of the project is:  
Find the factors and find out what causes those customers to churn. Retention Clients focus on analyzing data and determining correlated variables through rational procedures that consider influencing factors using derived factors. This gives you insight into which customers are likely to churn.

## INFORMATION OF DATA:

A data set aggregated by telecommunications company Orange SA that contains details of customer service usage and records of customer locations and plans. It consists of 3333 rows and 20 columns.

The churn variable indicates customer churn or non-churn based on existing conditions. About 14.5% are churn companies and the rest are non-churn companies.

## FEATURES 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

### **3. PROCEDURE:**

#### **(EDA)EXPLORATORY DATA ANALYSIS:**

EDA is the process of trying to understand data in the ways possible in order to derive insights from it.

Use exploratory data analysis to understand important factors or characteristics such as Avg, Std Mean Deviations, and also check for missing or null values and outliers.

Exploratory data analysis is the process of looking at available data sets to identify patterns and anomalies, test hypotheses, and validate assumptions using statistical means.

Using Python in exploratory data analysis processes and visual comparisons between variables is easy to understand and insightful.

#### **DATA ANALYSIS:**

Data analysis is an essential step, which involves descriptive statistics and data analysis.

This step involves summarizing data and uncovering unseen relationships and effects between different data sets, helping to develop and predict models, evaluate them, and identify factors with relative precision. high precision.

The procedures/steps used to aggregate data are the use and application of summary columns, graphs, descriptive and inferential statistics, correlation statistics, research, clustering as well as models. mathematical figure.

### **DATA SOURCING**

Data sourcing is the method of finding and storing it in our machines or systems. There are many ways to find data and it needs to be handled properly and in the right format.

The data must be processed by a competent and authorized person of the company concerned. There are many tools for finding, collecting, and storing data.

### **PROCESSING & CLEANING:**

The Raw data sometimes includes noise, nulls, formatting, and inconsistent values, so preprocessing the data is important for improving data quality.

Only then can we do data cleanup easily, which is important for finding outliers (such as

NullValues, format and title are incorrect

Anomalies/Exceptions, etc.)

The data can then be easily cleaned to get useful insights.

### **DATA TRANSFORMATION :**

Data transformation is the process in which data is further enhanced for performance and clarity. Sometimes the data contains duplicate rows and values. The removal of duplicate values is very important to improve the quality of the data set.

### **NULL/MISSING VALUES:**

In the data set, missing values occur due to many causes such as errors or data processing errors.

Sometimes, some customers don't subscribe to all of their services and plans, so in the respective columns or rows, there may be missing values in some columns representing the products.

In this Orange SA telecom dataset we have no nulls.

So we have to deal with the missing values, if any, present in the data set before doing any kind of analysis.

## **TREATMENT OF MISSING VALUES:**

To take off missing values, we can use different approaches and then apply them to remove missing values. For example, the pandas library has `IsNull()` and `not-null()` functions to determine null values and remove them with `dropna`.

Also deal with outliers based on their impact on our data set.

In this way we can remove the missing values from the data set to continue the exploratory analysis.

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

### **Univariate analysis**

We analyzed single variable/column data from the dataset, also known as univariate analysis. In univariate analysis, we take one feature at a time.

When we analyze a feature independently, the main motto of this analysis is usually to find the distribution of values (ranges) of that feature and ignore other features in the data set. .

Univariate analysis is the simplest form of data analysis. The data should consist of only one type of variable and we perform analysis on it. The main purpose of univariate analysis is to get data, summarize data, and find patterns between values. It does not address causes or relationships between values.

Graphical methods we used for this are

- Piechart
- Distplot

### **Bivariate analysis**

we analyze data by looking at two variables/columns from a data set, which is called two-variable analysis. Here, we mainly keep one constant, Churn, and change another variable for each column label.

Graphical methods we used for this are

- count plot
- Scatterplot
- Boxplot

### **Multivariate analysis**

In multivariate analysis, we analyze three or more variables. This allows us to look at correlations (i.e. how one variable changes relative to another) and allows us to understand the correlation between them and their behavior more accurately than a two-variable analysis. .

A common way to plot multivariate data is to create a pairwise histogram. Here, we used heatmap to find correlation

between all features (column labels) present in the dataset

## **FINDING CORRELATION**

Correlation helps establish relationships between variables and provides important insights such as how the variables affect each other and which factors have a high impact on each other, as well as customer interest and help the company work on these factors. In addition, it is very useful for future predictions or foresight about the company's product and helps in making the right decisions and predictions.

## **UNDERSTANDING GRAPHICAL REPRESENTATION:**

This step presents the analyzed data set to the target audience in the form of visual methods such as graphs, summary tables, charts, figures, charts.

This step is essential and easy to understand for business stakeholders, as it is the primary mission of EDA.

Most graphical analysis techniques include charts such as tables, polar charts, histograms, line charts, bar charts, scatter charts, area charts, pie charts, and more.

## **FINAL CONCLUSIONS:**

Some states have higher churn rates than others, which could be due to network issues. Therefore, the company should increase coverage and solve domestic and international network problems. Region, account length, evening and night calls have no relationship with each other, but the call rate increases as the number of calls to the customer

service center increases, so it is clear that the Companies must provide good customer service and resolve to customers to avoid being unsubscribed by collecting feedback.

Customers with international packages have a longer usage time than customers without international packages, which may be due to high fees or network problems. Therefore, the company must work on these factors and offer lower or more competitive international plans and coverage.

Customers with more than 20 voicemail messages (approximately) and customers with a higher number of daily calls have higher call interruption rates than other customers. So the company should focus on offering discounts or offers. It can also provide a personalized service with different packages according to customer needs.

The telecommunications industry can avoid disruption through such customer inactivity analytics and predictions. This type of EDA research in the telecom sector helps companies generate more profits. Anticipation of attractiveness is a very important factor for telecom companies. Therefore, this project aims to build a system that predicts the likelihood of customers giving up.

## **USEFUL REFERENCES:**

<https://medium.com/analytics-vidhya/telecom-churn-prediction>  
<https://bunker2.zlibcdn.com/dtoken/01c5fc197a94283bfb0c0943bd5b2d0c>