**Telecom Churn Analysis**

Ajinkya Krushnarao Dakhale **,**

Harshjyot singh chawla,

Bhaskar Subanji,

Data Science Trainees,

Alma Better, Bangalore

**Abstract:**

Now a day's telecom sector is experiencing cut throat competition, where customer won't even think twice before porting their sim (service provider) if they are provided with intriguing offer by other telecom company or if they are facing issues with existing telecom service provider.

With this EDA(exploratory Data Analysis) we will try to understand the reason for the churn in Orange S.A., formerly France Telecom S.A., French multinational

telecommunications corporation and come up with ways/recommendation for customer retention. The give Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features), along with a churn label specifying whether a customer cancelled the subscription

**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 analyze the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.

**Introduction**

The telecommunications industry is a dynamic business sector consisting of companies

operating in a subscription-based model. which allows customers to churn easily as

soon as their subscription ends to other rivalry company who so ever is offering better price ,

better network , faster internet services and more secure online experience. This transfer of customer from one company to another is called churn. With the enormous increase in the number of customers using telephone services, the marketing division for a telecom company wants to attract more new customers and avoid contract termination from existing customers (churn rate). Most customers have multiple services with the provider, such as voice, internet, and TV. When the customer cancels one service, the rest of the customer’s bundle of services goes away, too. Through churn rate(churn values) company kind of get their feedback about their standing against the competitor companies, they want to know why the churn happened (i.e. why did the customer switch to other company) and how to plan a strategy to prevent this churn and retain the customers .

The orange telecom company can use our EDA analysis to measure if it is providing a useful product compared with the product provided by its competitors, so as to plan

the strategy they must know the factors responsible churn, for that we will do EDA( exploratory Data Analysis) on the dataset, containing many features such as charges for call made by customer ,which all plan they have active(e.g. international plan, voice mail. plan.) of 3333 customers.

Here is the list of features available in the given dataset of the orange telecom

Data Variable

1. State: States name(in code).

2..Account Length: period for which the 3.Account is active.

4.Area Code: Area code having States

International Plan:

Yes: Means International Plan is subscribed and ,

No: Means the customer has not subscribed for international plan.

5.Voice Mail Plan: Yes: Voice Mail Plan is subscribed, No: voice mail Plan is not subscribed by the customer

6.Number vmail messages: Number of Voice Mail Messages.

7. Total day minutes: Total Number of Minutes Spent in Morning

8.Total day calls: Total Number of Calls made in Morning.

9.Total day charge: Total Charge for all call made in Morning.

10.Total eve minutes: Total Number of Minutes Spent in Evening

11.Total eve calls: Total Number of Calls made in Evening.

12.Total eve charge: Total Charge for all the call made in Evening.

13.Total night minutes: Total Number of Minutes Spent in the Night.

14.Total night calls: Total Number of Calls made in Night.

15.Total night charge: Total Charge for call made in Night.

16.Total intl minutes: Total Number of minutes spent in international calls.

17.Total intl calls: Total Number of call made internationally.

18.Total intl charge: Total charge for all the international call.

19.Customer service calls : Number of customer service calls made by customer

20.Churn Customer Churn,

True :churned customer

False: means retained customer

**Reasons for churn**

below given are some of the reason which could lead to churn :

- better price offered by competitor telecom service provider.

- no or late response in problem solution by customer service.

- Bad network quality, connectivity issue

**Approach:**

**Exploratory Data Analysis:**

Exploratory data analysis is process in which we try to understand the given data in 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. These 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 defining 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 correct format. Data should be handled by 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 :**

A 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 occurs 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 row their 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 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

- Piechart

- 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 variable 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 dataset.

**Conclusion :**

As we analyzed all the column label( variable) with the help of uni, bi and multivariate analysis we can see that some of them had some relation with the churn

rate and no of churned customer while others didn't.

we can say that mostly the reason for churn are network issue, high price and no proper support from customer service.

**References:**

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