Telecom churn analysis

Anurag Taiskar,
Data science trainees,
AlmaBetter, Bangalore

Abstract:

Telecom churn dataset contains details related to the customer with different plans and their call hours, call types and area details also calls made to customer service center etc, so exploratory data analysis has been performed on this datasets to understand the cognizance of the data by knowing the corresponding relation between the features and finding the important performance aspects. Utilize visualization methods to get some interpretations to take the accurate decisions to retain more customers by defining the factors affecting them to churn from the service and suggesting the factors to company in order to improve its service and retain the customers.

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

1.Problem Statement

Identifying the factors to retain customers by defining the factors affecting them to churn from the service and suggesting the factors to company in order to improve its service and retain the customers.

The main focus of the project is to: Finding factors and cause those influence customers to churn. Retain customers by taking appropriate steps considering affecting factors with the use of derived factors, this document focus onto analyse the data and determine variables that are correlated

with churn, which will gives insights about customer, who may churn.

ABOUT DATA:

Collected dataset from the telecom company called Orange SA, which contains the detailes of customer usage of its services and records about thier location and plans. It consist of 3333rows and 20 columns.
The variable ‘Churn’ displays customer churn or non-churn based on existing conditions. Approximately 14.5% are churn and remaining are no 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 morning
Total day calls: Total number of calls made in morning.
Total day charge: Total charges for all call made in a day.
Total eve minutes: Total number of minutes usage in evening
Total eve calls: Total number of calls made in evening.
Total eve charge: Total charge for all the call made in evening.
Total night minutes: Total number of minutes usage in the night.
Total night calls: Total number of calls made in night.
Total night charge: Total charge for call made in night.
Total intl minutes: Total Number of minutes usage in international calls.
Total intl calls: Total number of call made internationally.
Total intl charge: Total charge for all the international call.
Customer service calls : Number of customer service calls made by customer
Churn customer :-
True : churned customer
False: retained customer

2.INTRODUCTION:

Telecommunication industry is vast industry with different kind of its service providing to customer in service subscription model. These companies face challenges such as Customer churn, network coverage issue in particular area or the charges of their service. Customer churn is main cause or reason behind the large chunk loss to companies or causes reduction in profits.

Customer churn means a customer opting to rival or different company service due to low tariff or for any other reasons.
Thus, some companies take measures in finding the reasons that why the customers churn and try to come up with innovative products and service in order to retain the customers from churn.

Customer management is a strategic process of managing customer service with the aim of customer retention.
Some companies explore and analyse customers' data to understand the behaviour of their customers and gain important inputs that help improve customer service.
There are different factors for churn; and the major reasons are service quality, tariff rate, competitive offers from other companies.

BUSINESS OVERVIEW

In order to understand the objective of project it is necessary to understand the business. In telecommunication industry the main profit comes from service providing to customers with their plans and features. so we have to understand the business and then converting their data into analysis problem definition. Customer retention is the main focus for these kind of business and customer retention is to engage the customer with their business and to retain them from churn. Business needs to understand the factors that are affecting the customers and causing them churn from their service. Identifying the churn ration will help to understand the percentage of churn customer and also gives broad

idea about factors that made them to churn.by using these factors companies can predict customer behaviour and plan accordingly,so its much needed to have proper churn analysis model for companies to retain their customers.

REASONS FOR CUSTOMER CHURNING:

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

Tariff rate: comparatively high tariff or low tariff from competitors affects customers to churn because telecom industry is very competetitve and different companies in order to attract customers always thrive to launch different kind of products with lower tariff or customized services.

Service quality: service is the essential part of any business and that is very major factor to retain the customer to get business.

Lack of network coverage or network with issues.

Also slow resolving of these problems may also cause the customer churn and affects the service quality of company products.

customer service: providing a good customer service will always leads to increase in number of customers and vice versa.sometimes not resolving customer issues in speedy manner may also leads to churn.

Plans or offers: Competitive plans and better alternatives from the rival companies leads to churning of the customers from services. So its always key factor to provide them more flexibility and offers to afford and benefit from the service otherwise customer will churn or unsubscribe from the services and switch to different subscription where they can afford at low cost or getting the same benefits at lesser price in their Plans.

3. 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 maths models.

SOURCING OF DATA

Data sourcing is the method of finding and storing it into our machines or systems. Their 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. Their 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 thier 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.

DROPPING OF MISSING VALUES:

In order to remove the missing values we can use different approaches and then applying those to remove the missing values.for example isnull() and notnull() functions from the pandas library to determine null values and remove using dropna.
Also handling of outliers based on their impact on our dataset.
In this way we can remove the missing values from the dataset to proceed further in exploratory 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

CORRELATION :

Correlation helps to establish the relations between the variables and gives the important insights such as how variable affecting each other and what are the factors that are having high impact on each other as well churn of the customer and helps company to work on these factors. Also it very useful for future or visionary prediction of company products and helps to make good decisions and prediction.

GRAPHICAL REPRESENTATION OF THE RESULTS:

In this step portraying the analyzed dataset to the target audience in the form of visualization methods like graphs, summary tables, plots, drawings, histograms.

This step is very essential and easy to interpret by the business stakeholders, as this one is the main task of EDA.

Most of the graphical analysis techniques include charts like Table chart, Polar chart, Histogram, Line chart, Bar chart, Scatter plot, Area plot, Pie chart etc.

CONCLUSIONS:

Some states have higher churn rate compared to other, for which network issues might be the reason so it is recommended that company should increase the coverage area and solve the network problems in local as well as international level.

Area, account length, evening & night calls has no relation with churn rate but the churn rate increases as the number of call to the customer service center increases so its evident that company must provide good customer service and resolution to customers to prevent churn by gathering feedback.

Customers with international plan has higher churn rate compared to customers without international plan the reason could be high tariff cost or network issues so company should work on these factors and provide lower or competitive international plans and coverage.

Customers with voicemails more than 20 (approx.) and customers with higher day call min`s has higher churn rate compared to others so company should focus on providing discounts or offers, also they can offer customized service with different plans based on customer needs.

Telecommunication industry can avoid churn by such analysis and predicting factors for customer churn.

These type of EDA research in the telecom segment helps companies to gain more profit. Predicting churn is very important factor for telecom companies Hence, this project aimed to build a system that predicts the possibility of customer churn.

REFERENCES:

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

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