Author’s Name: Zeeshan Ahmed

Course Title: Telecom Churn Prediction

(EDA) 10 January 2023

Keywords: Data Science, Data Analysis, EDA

Telecom Churn Analysis

ABSTRACT

This data set comprises information regarding the churning of a customer for a Telecom Company. It includes information such as number of minutes spent on a call, amount charged, total number of calls, whether or not it was an international call at different times of a day, namely day, evening(eve), or night. It shows for how long an individual has been loyal to the company in the form of account length, customer choice on voice plan and number of customer service calls received. It also includes data pertaining to customers location i.e., State and Area code.

PROBLEM STATEMENT

We are tasked with performing exploratory data analysis on the given dataset to get relevant insights from the data and understanding the key factors responsible for customer churn given particular parameters.

INTRODUCTION

1. What is EDA?

Exploratory data analysis (EDA) is employed by data scientists to research and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps

determine how best to control data sources to urge the answers you would like , making it easier for data scientists to get patterns, spot anomalies, test a hypothesis, or check assumptions.

EDA primarily wants to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a far better understanding of knowledge set variables and therefore the relationships between them. It also can help determine if the statistical techniques you're considering for data analysis are appropriate. Originally developed by American mathematician John Tukey within the 1970s, EDA techniques are still a widely used method within the data discovery process today.

2. Why is EDA important in Data Science?

The main purpose of EDA is to assist check out data before making any assumptions. It can help identify obvious errors, also better understand patterns within the info , detect outliers or anomalous events, and find interesting relations among the variables.

Data scientists can use exploratory analysis to make sure the results they produce are valid and applicable to any desired business outcomes and goals. EDA also helps stakeholders by confirming they're asking the proper questions. EDA can help answer questions on standard deviations, categorical variables, and confidence intervals. Once EDA is complete and insights are drawn, its features can then be used for more sophisticated data analysis or modeling, including machine learning.

3. Types Of EDA

There are four primary sorts of EDA:

- Univariate non-graphical. This is often the simplest sort of data analysis, where the info being analyzed consists of only one variable. Since

it’s one variable, it doesn’t affect causes or relationships. The main purpose of univariate analysis is to explain the info and find patterns that exist within it.

- Univariate graphical. Non-graphical methods don’t provide a full picture of the info . Graphical methods are therefore required. Common sorts of univariate graphics include:

- Stem-and-leaf plots, which show all data values and therefore the shape of the distribution.

- Histograms, a bar plot during which each bar represents the frequency (count) or proportion (count/total count) of cases for a variety of values.
- Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

- Multivariate non graphical: Multivariate data arises from quite one variable. Multivariate non-graphical EDA techniques generally show the connection between two or more variables of the info through cross-tabulation or statistics.

- Multivariate graphical: Multivariate data uses graphics to display relationships between two or more sets of knowledge. The foremost used graphic may be a grouped bar plot or bar graph with each group representing one level of 1 of the variables and every bar within a gaggle representing the amount of the opposite variable.

Other common sorts of multivariate graphics include:

- Scatter plot, which is employed to plot data points on a horizontal and a vertical axis to point out what proportion one variable is suffering from another.

- Multivariate chart, which may be a graphical representation of the relationships between factors and a response.

- Heat map, which may be a graphical representation of knowledge where values are depicted by colour.

PROCEDURE FOR DATA ANALYSIS:

1. Importing packages
2. To check and treat Null Values
3. Outlier Detection
4. Data Manipulation
5. Data Visualisation
6. Insights Types of plots used in analysis:

- Barplot
- Line Diagram
- Count Plot
- Histogram
- PieChart

CONCLUSIONS:

- Most of the customers don't prefer international plans and voicemail plans.

- Most of the customers spend more time on calls during day time.

- Business is booming in Area Code 415 specifically in the East Coast Region of the US.
- Customers on average are using the services for a period of around 8 years.

- Most of the customers are located in West Virginia.

- The factors that are affecting the most for customer churn are International Plan, Total day charge, Customer service calls.

To retain customers :

- The company should give incentives who spend more time on calls during day time

- The company should concentrate more on prices than features.

- The company should lower prices for international plans and its hourly call rates.

- The company should decrease the frequency of customer service calls received by customers.