



Exploratory Data Analysis - Telecom Churn Analysis

Technical documentation

Mohammed Akifuddin Kashif
Akifkashif007@gmail.com



Introduction


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.

Overview of the data

1. State: Name of the State
2. Account length: Length of the account
3. Area Code: Area where customer resides
4. International Plan: If customer opted for International Plan
5. Voicemail plan: If customer opted for Voicemail Plan
6. Number vmessages: Number of voicemails received
7. Total day minutes: minutes spent on call during day time
8. Total day calls : Number of calls made during day time
9. Total day charge : Amount charged for calls during day time
10. Total eve minutes: minutes spent on call during evening time

- 
11. Total eve calls :Number of calls made during evening time
 12. Total eve charge : Amount charged for calls during evening time
 13. Total night minutes : minutes spent on call during night time
 14. Total night calls : Number of calls made during night time
 15. Total night charge : Amount charged for calls during night time
 16. Total night charge : Amount charged for calls during night time
 17. Total intl minutes : minutes spent on international calls
 18. Total intl calls : Number of calls made internationally
 19. Total intl charge : Amount charged for international calls
 20. Customer service calls : Number of customer service calls received
 21. Churn : If customer churned

Performing EDA (Performing Exploratory Analysis)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses and to see assumptions with the assistance of summary statistics and graphical representations.

1. Extracting head and tail of the dataset
2. Extracting info of the dataset which tells the type of data present in different columns
3. Producing description of data
4. Checking the unique count of column values.

5. Checked for null values and there weren't any.
6. Creating dummies for data which aren't numeric.
7. Plotting relevant graphs to extract information from them.
8. Detecting outliers in columns using IQR method and visualizing them using BoxPlots.
9. Removing them by defining upper and lower limits using the IQR method.
10. Plotting correlation plots for numeric features to get additional information

Plots used

Plots are a way to describe the data given to us in a visual manner which is more understandable and convenient to draw insights from , compared to raw data. In This project the plots we used are:

1. Countplot - provides the count of required instances from the data.
2. Boxplot - provides information on the outliers present in the data.
3. Histplot - also provides count of instances (seaborn based).
4. Pieplot - provides info in the form of a pie chart.
5. Barplot - BarPlot uses bar to represent the numeric values on the X and Y axis.



Conclusions from data

- 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.



Challenges Faced

1. Pre-processing the data was one of the challenges we faced, such as detecting outliers and removing them.
2. Manipulating the data was difficult as it could affect and alter important information provided by the dataset.
3. As the dataset has more columns it was quite difficult to extract relevant information from all the columns.

Final Conclusion

- We are finally at the conclusion of our project!
- We performed exploratory data analysis from the given dataset and were able to draw important and relevant information which explains why in a particular instance there are more customers who were loyal or have churned compared to other instances.
- We also used different types of plots to better visualize and present the data in a more appealing manner.