

EXPLORATORY DATA ANALYSIS ON TELECOM CHURN

COHORT AZAADI, ALMABETTER

BY: - POOJA RATHORE

Abstract :

Orange S.A. is a telecommunication company consists that transmit and receive the data in voice, audio across the globe. Telecom equipment, telecom services, and wireless communication are the three basic sub-sectors of telecommunications.

EDA of France Telecom helps us to investigate the data to understand the customers of Telecommunication, spot anomalies, cleaning the data, & check the assumption with the help of the summary statics & graphical representation.

1. Problem Statement

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the Orange S.A. telecommunications company experiences an average of 15% churn rate

To reduce customer churn, orange S.A. Company need to predict which customers are at high risk of churn. In this EDA project, we will analyse customer-level data of a leading telecom firm, and identify the reason behind it.

Below are the info that is available in given dataset-

- State- Having abbreviation name of USA states only
- International plan - A check for international plan
- Voice mail plan - A check for voicemail plan.
- Number vmail messages - Number of voicemail messages sent by customers who opted for voice mail plan.
- Total day minutes - Having total of minutes which a customer spent in a day-time
- Total day calls - Having total number of calls of a customer in a day-time

- Total day charge - Having total of charges of a customer's spending in a day-time
- Total eve minutes - Having total of minutes which a customer spent in a evening-time
- Total eve calls - Having total number of calls of a customer in a evening-time
- Total eve charge - Having total of charges of a customer's spending in a evening-time
- Total night minutes - Having total of minutes which a customer spent in a night-time
- Total night calls - Having total number of calls of a customer in a night-time
- Total night charge - Having total of charges of a customer's spending in a night-time
- Total intl minutes - Having total of minutes which a customer spent on international calls
- Total intl calls - Having total number of international calls of a customer
- Total intl charge - Having total of charges of a customer's spending on international calls
- Customer service calls - Having number of calls made by a particular customer to customer service centre
- Churn - Having churned and non-churned status of customers

2. Introduction

Orange S.A. dataset has information about State, Account length, Area code, International plan, Voice mail plan, Number voice mail messages, Total day minutes, Total day calls, Total day charge, Total evening minutes, Total evening calls, Total evening charge, Total night minutes, Total night calls, Total night charge, Total international minutes, Total international calls, Total international charge, Customer service calls and Churn. So, with Exploratory Data Analysis we are analysing the patterns, changes, fluctuations in duration and charges, state wise distribution of churn, priority and relevancy of the people, and many things can be taken out as a conclusion from this dataset which can be used for the future development of the company.

Our goal here is to provide future-oriented conclusions for Orange S.A. So that they can take decisions based on that conclusions and earn fruitful results.

3. Challenges Faced

- Understanding the dataset specially it's columns
- Understanding the business model of Orange S.A.
- Removing Null values, changing datatype, identify outliers.
- Designing the presentation to visualize the summarization such that the data is easily accessible to the reader

4. Libraries Used in EDA

- Pandas: Pandas is a python library used for data manipulation & statistical analysis. It is a fast & easy way to use open-source library that enables several data manipulation tasks. These include merging, reshaping, wrangling, statistical analysis and much more.
- NumPy: NumPy can be used to perform a wide variety of mathematical operations on arrays. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.
- Matplotlib: Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits.
- Seaborn: Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with Pandas data structures. Seaborn helps you explore and understand your data. Its plotting functions operate on DataFrames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

5. EDA Procedure

- Importing Libraries
- Analysing the Dataset
 - ✚ Understanding the dataset
- Data Wrangling
 - ✚ Removing null values
 - ✚ Changing Dtype of International Plan, voice mail plan and Churn.
 - ✚ Removing the duplicates
 - ✚ Dropping unnecessary columns like Account length
 - ✚ Outliers handling
- Plotting graph using matplotlib & Seaborn
- Observing the graphs & plots
- Final Conclusion

6. Findings

- Total day minutes, total day charge, number of customer service calls, total evening minutes and international plan are the most useful features in the data set.
- Approximately 14% of the customers in the data base are classified as churn.
- Customers having no international plan out of which 10% are classified as churn and Customers having international plan out of which approximately 40% are classified as churn.
- Customers having no voice mail plan out of which 15% are classified as churn and Customers having voice mail plan out of which 8% are classified as churn.
- 50% of the customer who called the company more than 3 times are classified as churn.

7. Conclusion

Beginning with loading the data so far we have done Exploratory Data Analysis, null values treatment, encoding of categorical columns, plots and charts selection and then visualization/predictions building.

So according to our analysis the major cause of churning is-

- Same Pricing Strategies for all the Customers

- There are some Network Disturbance/Network related problem in some States
- Bad/Poor Quality of Customer Service. Company make sure to improve the services
- Need to Increase the International Call Rates and their related services

8. References

- Stackoverflow
- GeeksforGeeks
- Kaggle