Capstone Project Submission

My Name, Email and Contribution:

Gaurav Yogeshwar (yogesh.grv9@gmail.com)

- Upload dataset to Google colab..
- Data cleaning.
- Correction of data types
- Data wrangling
- Data Visualizations
- Dimensionality reduction
- Technical Write up
- PowerPoint presentation
- video explanation
- Project summary

Github Link-

https://github.com/Gaurav2912/Telecom-Churn-Analysis---Capstone-Project.git

Short summary of Capstone project and its components.

While customer acquisition and retention is a major concern for many, with the rapid growth of the telecoms industry, service providers are keen to expand the customer base. To meet the need to survive in the competitive environment, retaining existing customers has become a major challenge, it is cheaper to retain a loyal customer then acquire a new one.
Therefore, gathering insights from telecoms industries can help predict customer engagement, whether they leave the company or not.

Goal -

Explore and analyze the data to discover key factors responsible for customer churn in Orange Telecom's Churn data set and come up with recommendations to ensure customer retention.

This project focuses on exploratory data analysis to identify potential customers, classify them based on usage patterns, and analyze the data to discover key factors responsible for customer churn and come up with ways to ensure customer retention., so that the telecoms industry can stabilize its market value to acquire its associated customers to take necessary steps.

In this EDA project we were provided datasets, which have 20 columns

At first, we break down the datasets by importing necessary library classes, followed by checking unique values, getting a statistical summary of the numeric columns, converting the data types to similar objects, doing feature engineering and making the entire dataset ready for analyzing & plotting actionable insights.
After examining null & duplicate values from the dataset we directly went deep into the visualization steps.

Some insights on which we worked are as follows:

- Display the balance of the class labels (Churn) with a Count Plot and Donut Chart.
- Explore the average call rate and our customer's talking behavior for different-different time-zone.
- To understand the distribution of each numerical attribute, I plot their histogram.
- Correlations check among all features & visualize on heat map and bar plot.
- Calculated churn rate per category for categorical variables.
- Visualization using box plot of Churn behavior with all numerical features.
- Separately visualize using histograms, box-plot, violin-plot, bar-plot, scatter-plot and count-plots with features that have no correlation, then for those which have correlation with the target variable .
- I made a comprehensive grouping based on total minutes of the day for in-depth analysis.
- I did data pre-processing to analyze churn rate with different states of the USA, then analyze it with all states in the Choropleth map in Plotly.
- Visualization of correlation of all features with Churn, for states having high churn rate.
- Visualize the dataset and mark the patterns in two dimensions using t-SNE technique.
- After visualization we came to some conclusions.
- and finally we came up with certain recommendations that can reduce the churning behavior of Orange Telecom's customers.

Recommendations

- We need a very attractive **international plan**, which can provide satisfaction to the customers making international calls.
- Need to introduce better **tariff plans** for the day as well as evening calling which is specially designed for users talking too much in the day. Day tariff will be the first priority.
- Need to improve **feedback systems** that don't ignore customer problems.
- Need a new voicemail plan along with the old one which is specially designed keeping in mind the more voicemail senders.