Capstone Project Submission

Team Member's Name, Email and Contribution:

Name	Email ID
Rishabh Kumar	kumarrishubh0@gmail.com
Nitesh Singh	niteshrajput.edu@gmail.com
Sanjog Mishra	sanjogmishra1997@gmail.com
Ankit Kumar	ankitiitd221@gmail.com
Nilanjan Chandra	nilanjan.chandra85@gmail.com

Contributor roles:

1. **Nitesh Singh**:
   i. Data wrangling of Telecom Churn Analysis dataset
   ii. Creation of two new features for analysis, Total Charges and Total Minutes
   iii. Heatmap analysis to find correlation
   iv. Churn customer percentage across total call hours analysis

2. **Sanjog Mishra :**
   i. Data wrangling of Telecom Churn Analysis dataset.
   ii. Univariate Analysis of the features.
   iii. Churn customer percentage across month analysis
   iv. Customer distribution along area code analysis

3. **Ankit Kumar:**
   i. Data wrangling of Telecom Churn Analysis dataset.
   ii. Customer distribution across states
   iii. Finding top 10 states where churn percentage is high
   iv. Analysis based on international plan

4. **Rishabh Kumar:**
   i. Data wrangling of Telecom Churn Analysis dataset.
   ii. Analysis based on Customer service calls
   iii. Analysis based on Voice mail service
   iv. Total international minutes vs Churn analysis

5. **Nilanjan Chandra:**
   i. Data wrangling of Telecom Churn Analysis dataset
   ii. Total Charges vs Churn analysis using Box plot
   iii. Total day minutes vs churn analysis using Box plot
   iv. Analysis of Total day minutes with Churn when Customer service calls increases

Please paste the GitHub Repo link.

Github Link:- https://github.com/iamtheone07/Power--Telecom-Churn-Analysis

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Customer churn or Attrition is a crucial problem and one among the foremost principal concerns for giant telecom companies. We were given the data set of a French multinational telecommunications corporation which contains cleaned customer activity data (features), along with a churn label specifying whether a customer cancelled the subscription.

Our objective was to perform Exploratory data analysis on the dataset to identify the major possible factors that triggered the attrition. At first, we checked for the presence of any Null value. There was no Null value in the dataset. We also analyzed the raw dataset and looked for any garbage values , outliers and the important and unnecessary features. In this step we manually went through each feature and segregated the categorical and numeric/continuous data. Also added two extra columns for analysis named total charges and total minutes. In this step we have done some exploratory data analysis on the selected features to find the various relations amongst the features with the targeted column. After that we have plotted the models using Univariate and Bivariate analysis to draw some insights out of that.

In the first step of creating the models, we found the percentage of the customer base that has churned out using Pie chart and we also found the top five states with attrition ratio in the significant range. After that, heatmap analysis was done to find the correlation amongst various features in order to select the crucial features. Next, we did some univariate analysis and found the percentage of international plans and voicemail users.

By plotting the distribution of account length, we could identify the time duration after which most of the customers left the service. The distribution of Total day, evening and night charges were Normal distribution and of similar type. We did not find any significant relationship of area code. From the analysis of churn vs the customer service calls, a significant spike in churn was found after a certain customer service calls. Next, we did some box plot analysis to find the effect of different call charges on churn and found the result.

We also tried to find if there is any effect of international plan users on Churn by plotting a Box plot. We also analysed if there is any advantage of having an international plan for the customers. After that we did Bivariate analysis using a box plot to recognize how the total minutes vary for the Churn and Non-Churn users as customer service calls increase. Based on the data wrangling, we drew some insights and factors responsible for the customer churn. Based on this analysis we recommended key strategies to the companies to prevent the customer attrition.