Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Name, Email and Contribution:

Role:-

- **Mahefuj Husen Khan**

  EMail- mahefujhusenkhan552@gmail.com

  - Data Understanding.
  - Feature Analysis.
    - Calls Data,
    - International call.
    - Voice mail call.
    - No. of voice mail.
    - Area code.
    - Account length.
    - State.
    - Customer service call.
    - Churn.
  - Feature Engineering.
    - Null value check.
    - Missing value.
    - Duplicate value
  - Data Visualization.
    - Donut plot.
    - Box plot.
    - Count plot.
    - Scatter plot.
  - Multivariate Analysis.
    - Correlation matrix.
    - Heat map.
  - Research Analytics.
    - Technical documentation.

Please paste the GitHub Repo link.

Github Link:- https://github.com/mahefujh/Capstone_project_1

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

PROBLEM STATEMENT:

Customer churn in the telecom industry poses one of the most significant risks to loss of revenue To reduce customer churn, telecom companies do data analysis of customer data to predict who is most likely subject to churn, and what to do to retain the most valuable customer.

APPROACH:

In my telecom data set, there are 20 columns/variables and 3333 rows present, of which 3 objective, integer, and float both have 8 and 1 Boolean data types. First, I clean the data set by checking missing values and duplicate values. Thankfully there are no missing values present then we analyze the data each column-wise first, take some insight into what our variables said to us then proceed with bivariate analysis i.e. I compare each column to the dependent variable i.e. churn. After that, I analyzed the correlation between the variables and found that there are 4 to 5 variables present which is more effective for the churn rate. Then I plot some graphs for a better understanding of the data like count plot, display, box plot, scatter plot, donut plot, etc. after analyzing the vast data I make some conclusions, which are below;

CONCLUSION:

There are some states where the churn rate is high compared to others due to low network coverage. Area code and account length do not play any kind of role in the churn rate. In the international plan, those customers who have this plan are churning more, and also the international calling charges are also high so the customer who has the plan unsatisfied with network issues and high call charges. In the voicemail section, when there are more than 20 voice-mail messages, there is a churn, so the quality of voice mail is not good. The customer who has high day call minutes also has a high call price this customer tends to churn. In customer service calls data shows us that whenever an unsatisfied customer called the service center the churn rate is high, which means the service center didn't resolve the customer issue.