Capstone Project Submission

Team Member's Name, Email, and Contribution:

Team Member's Role:-

- **Sanjaya Kumar Khadanga**
  eMail- skhadanga38@gmail.com
  - Data Understanding
  - Feature Analysis
    - Calls Data, international call, voice mail call, No. of voice mail
  - Feature Engineering
    - Null value check
  - Data Visualization
    - Donut plot, box plot
  - Multivariate Analysis
    - Correlation matrix
  - Research Analytics
    - Technical documentation

- **Bibhuti Bhusan Sahu**
  eMail- sahubibhuti45@gmail.com
  - Data Understanding
  - Feature Analysis
    - Area code, account length, state
  - Feature Engineering
    - Missing value
  - Data Visualization
    - Count plot
  - Multivariate Analysis
    - Heat map
  - Research Analytics
    - Technical documentation

- **Balaram panigrahy**
  eMail- balarampanigrahy42@gmail.com
  - Data Understanding
  - Feature Analysis
    - Customer service call, churn
  - Feature Engineering
    - Duplicate value
  - Data Visualization
    - Scatter plot

- o Multivariate Analysis
- o Research Analytics
  - Technical documentation

Please paste the GitHub Repo link.

Github Link:- https://github.com/sanjaykhadanga/CAPSTONE-PROJECT---ONE

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 our telecom data set there are 20 columns/variables and 3333 rows present out of which 3 objective, integer, and float both have 8 and 1 Boolean data type. First, we 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. we compare each column to the dependent variable i.e. churn. After that, we analyzed the correlation between the variables and found that there are 4 to 5 variables present which is more effective to the churn rate. Then we plot some graphs for better understanding of the data like count plot, display, box plot, scatter plot, donut plot, etc. after analyzing the vast data we make some conclusions, which are below;

CONCLUSION:

There are some states where the churn rate is high as compared to others may be due to low network coverage. Area code and account length do not play any kind of role regarding the churn rate. In the international plan those customers who have this plan are churn 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 then there is a churn so it means that 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.