Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name - V Bhavya Reddy Email – bhavya.reddy0711@gmail.com Contribution – Whole project. 1. LOOKING FOR AND HANDLING NaN/Null/Missing DATA 2. DATA CLEANING 3. ANALYSING THE DATA. - 1. The correlation of various parameters is visualized in a heat map. 2. Stacked bar graph demonstrating relation between non-numerical attributes and churn ratio. 3. Scatter plot for state and the churn rate. 4. RELATION WITH CHARGES AND CHURNS. 5. INTERNATIONAL PLAN AND VOICE MAIL PLAN FOR CHURNED AND ACTIVE ACCOUNTS. 6. ANALYSING ACCOUNT LENGTH AND NUMBER OF VOICE MAIL MESSAGES ON CHURN. 7. CUSTOMER RETENTION: HOW IMPORTANT IS IT? WAYS IT CAN BE DONE?
Please paste the GitHub Repo link.
Github Link:- https://github.com/coolphotography/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)

Problem Statement:
The Churn dataset is provided by a French multinational telecommunications corporation, Orange. The Orange Telecom's Churn Dataset consists of 3333 rows and 20 columns. Every row corresponds to a unique account and in each column, there is a unique attribute indicating several services, account information, churn, etc.
The main objective is to analyze these attributes, their dependence on customer churn, and ways to increase customer retention.

Approach:
As a first step, I checked for Nan as well as incorrect data, followed by converting white spaces to underscores (_) in column names.

The next step was to find correlations between the attributes and plot it on a Heatmap. However, this method only works for numerical columns. In the non-numerical data, the churn ratio was calculated by grouping the data and then calculating the mean churn, which was later plotted as a stacked bar graph. I then plotted the state vs churn ratio scatter plot.

From the above steps, I have been able to identify the key elements for churn analysis as charges and international plans. A slight ambiguity existed regarding the voice mail plan. A demonstration was required that the account length and the number of voice mail messages are not correlated with churn.

In order to understand charges and churn on a deeper level, a new attribute called 'Total_charge' has been introduced where Total Charge = Total day charge + Total eve charge + Total night charge + Total international charge. The data was sorted by Total Charge and consecutive 101 charges were grouped. Cumulative churn, day charges, evening charges, night charges, and international call charges were then calculated.

I plotted several pie charts to determine the impact of International and Voice mail plans on churn for the active and churned accounts. As it turns out, Voice plans don't affect churn.

The last graph shows that account length and voice mail messages do not influence churn.

Later, a small calculation has shown this:
For a churn rate of 14.49% the loss percent is 15.93% and loss is 31566.93.

Then, I explained the reason for customer retention and ways to increase the customer retention period.

Conclusions:

1. One is charged based on the amount of time one spends on the call.

2. The attributes which are not contributing to the churn are -
    1. Area code
    2. Account length
    3. Voice Mail Plan
    4. Number of voice messages.

3. Either the international call services are poor, or the prices are higher than those of other operators.

4. It would make sense to offer lucrative incentives to customers who make more than four service calls, such as free limited data or OTT services.

5. Of the 51 states, 23 states have churn rates higher than the average. This requires more attention to understanding local customer needs.

6. The churn rate was highest when the total charge (day charges + night charges + evening charges + international charges) exceeded 75 dollars.

7. In the daytime, more calls are made than at night or in the evening, which means more traffic, which results in call drops, which ultimately leads to poor service and churn.