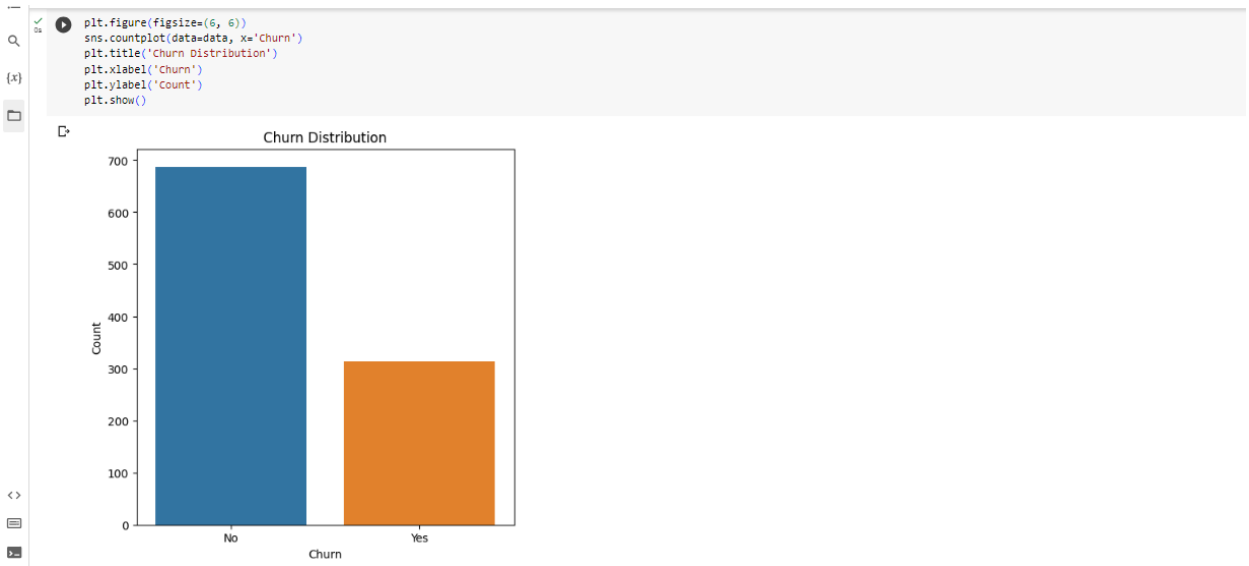
Report

Zunira Sajjad

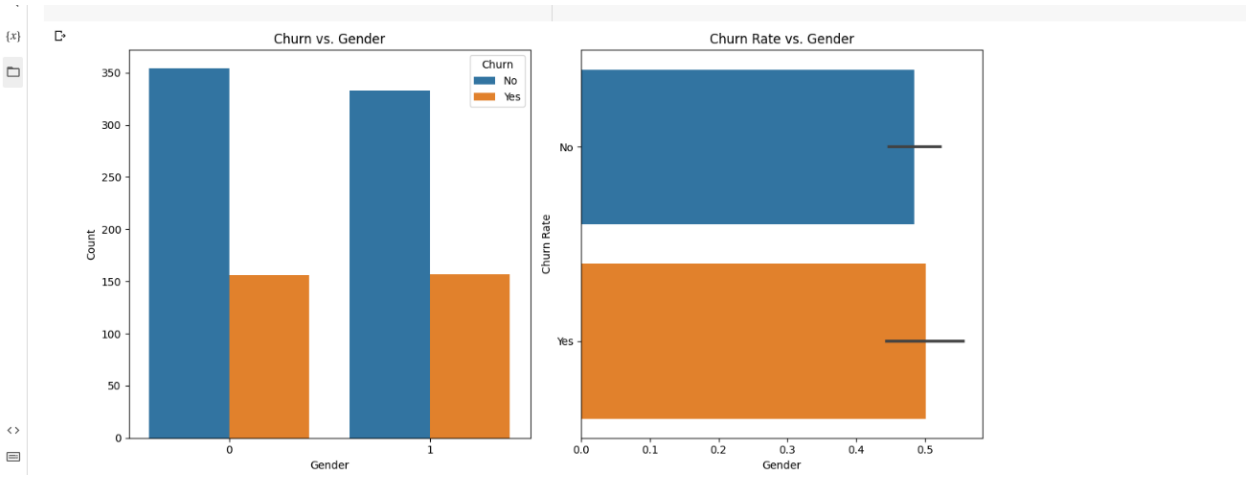
Week_05 Day_02

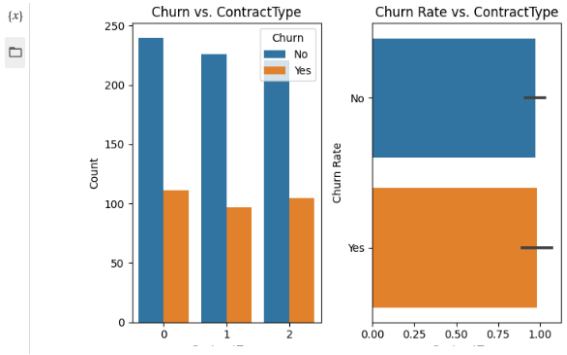
Churn Distribution



Count plot shows the count of each distribution in the form of bar as shown in above picture.

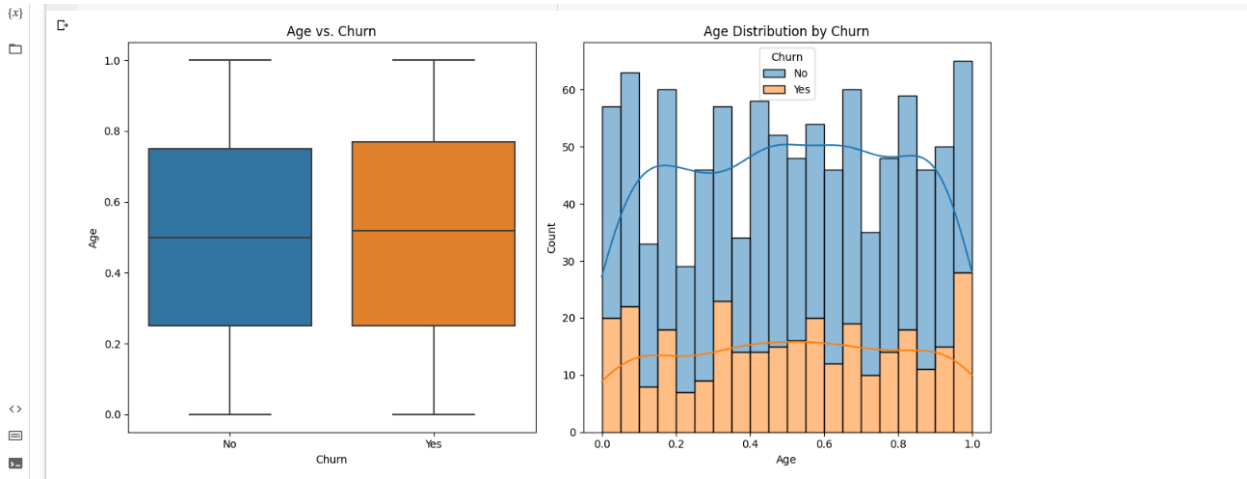
Churn vs Categorical distribution

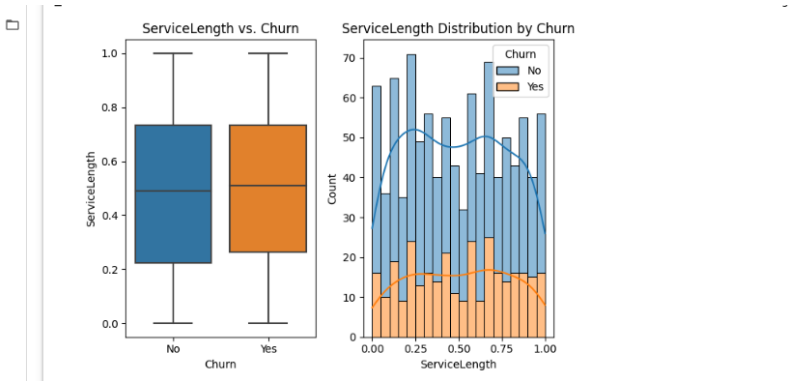


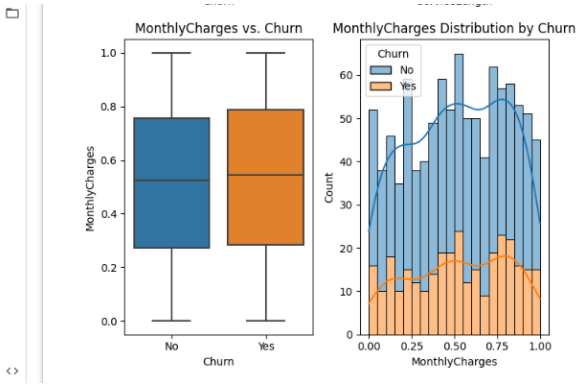


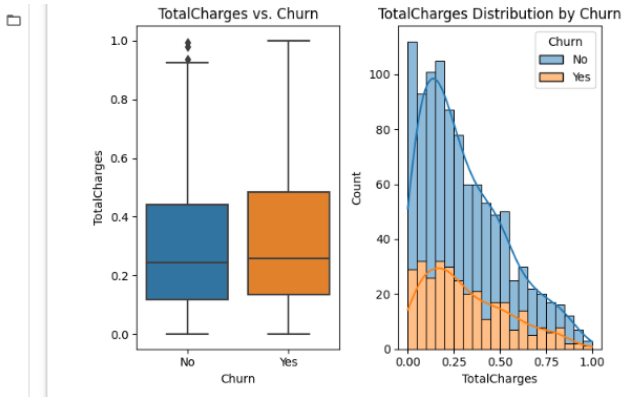
This plot is created using the sns.countplot() function from the Seaborn library. It visualizes the distribution of the 'Churn' outcome within each category of the specified categorical variable. The x axis represents the categorical variable (e.g., 'Gender' or 'ContractType'), and the y axis represents the count of occurrences. This plot is created using the sns.barplot() function from Seaborn. It focuses on visualizing the churn rate for each category of the categorical variable. The x axis represents the categorical variable, and the y axis represents the churn rate (proportion of 'Yes' churns) within each category. The estimator parameter is used to calculate the average churn rate within each category.

Churn vs numerical distribution





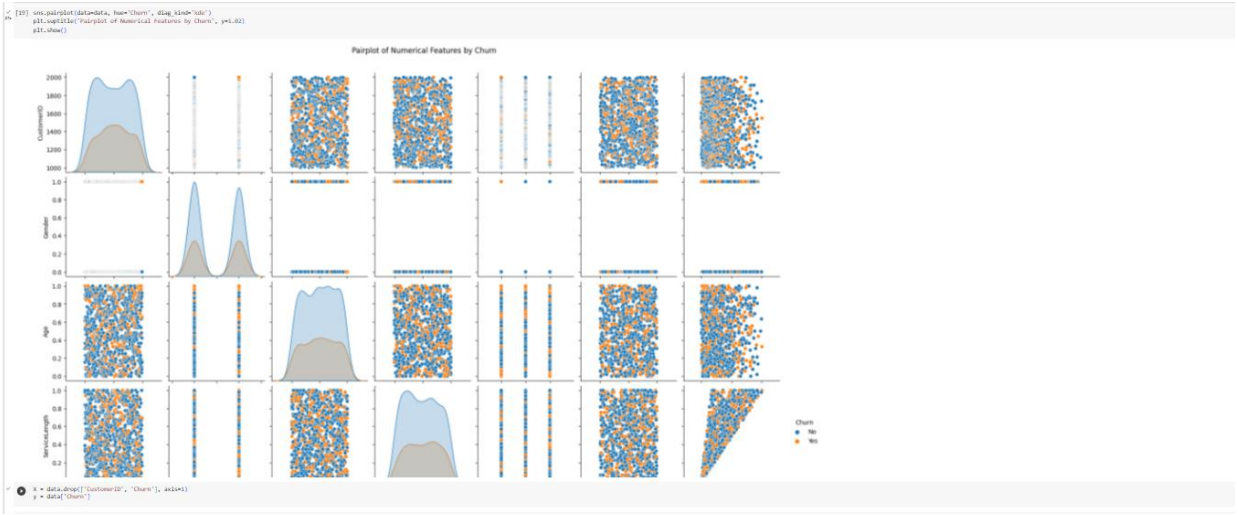


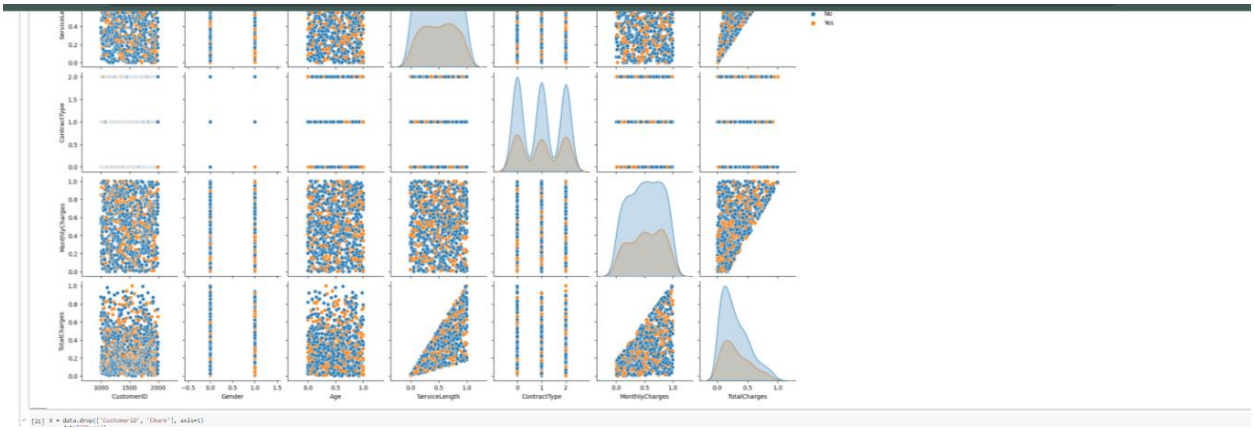


The left side of the code snippet creates a boxplot for each numerical variable grouped by the 'Churn' outcome. A boxplot summarizes the distribution of a numerical variable by displaying its quartiles (25th, 50th, and 75th percentiles) along with any potential outliers. The x axis represents the 'Churn' status ('Yes' or 'No'), and the y axis represents the values of the numerical variable. This plot helps you understand how the central tendency and spread of each numerical variable differ between churned and non-churned customers. The right side of the code snippet creates a histogram for each numerical variable, with two distributions stacked on top of each other based on

the 'Churn' outcome ('Yes' or 'No'). The x axis represents the values of the numerical variable, and the y axis represents the count of occurrences. The colors of the bars indicate the proportion of customers who churned ('Yes') and those who did not ('No') at different ranges of the numerical variable. The kde=True argument adds a kernel density estimate to the histogram, which provides a smooth representation of the data's distribution. This plot allows you to compare how the distribution of numerical variables differs between churned and non-churned customers.

Numerical Features distribution





The pairplot() function creates a matrix of scatter plots, where each scatter plot represents the relationship between two numerical features. The diagonal axes of the matrix display kernel density estimate (KDE) plots for each feature separately, showing the distribution of each feature. The scatter plots in the matrix visualize how the data points are distributed in relation to each other.