## Inference Document Exploratory Data Analysis (EDA) of Loan Sanction Dataset

This Expirement conducts exploratory data analysis (EDA) on a loan sanction dataset using the Python libraries pandas, numpy, seaborn, and matplotlib. The analysis includes univariate and bivariate visualizations to understand the distribution and relationships within the dataset.

**Data Loading and Preparation:**

The code loads the loan sanction dataset from the specified CSV file and creates a pandas Data Frame.

The 'Loan\_ID' column is dropped from the Data Frame.

1. **Univariate Analysis:**

**Numerical Variables:**

* For each numerical variable, a histogram is plotted to visualize its distribution. This helps in understanding the central tendency and spread of each numerical feature.
* Box plots are also generated to identify the presence of outliers in numerical variables.

**Categorical Variables:**

* For each categorical variable, a count plot is created to show the distribution of categories. This provides insights into the frequency of each category in the dataset.
* Pie charts are used to display the proportion of each category within categorical variables.

1. **Bivariate Analysis:**

**Numerical-Numerical Relationships:**

* Pairwise scatter plots are generated for numerical variables to explore relationships between pairs of features.

**Categorical-Categorical Relationships:**

* Count plots are created for each pair of categorical variables to visualize their relationships.

**Numerical-Categorical Relationships:**

* Scatter plots are generated for each pair of numerical variables, with points differentiated by categories of a third categorical variable. This allows for a multivariate analysis, providing insights into how the numerical variables relate to each other within different categories.

1. **Multivariate Analysis:**

**Numerical-Numerical-Categorical Relationships:**

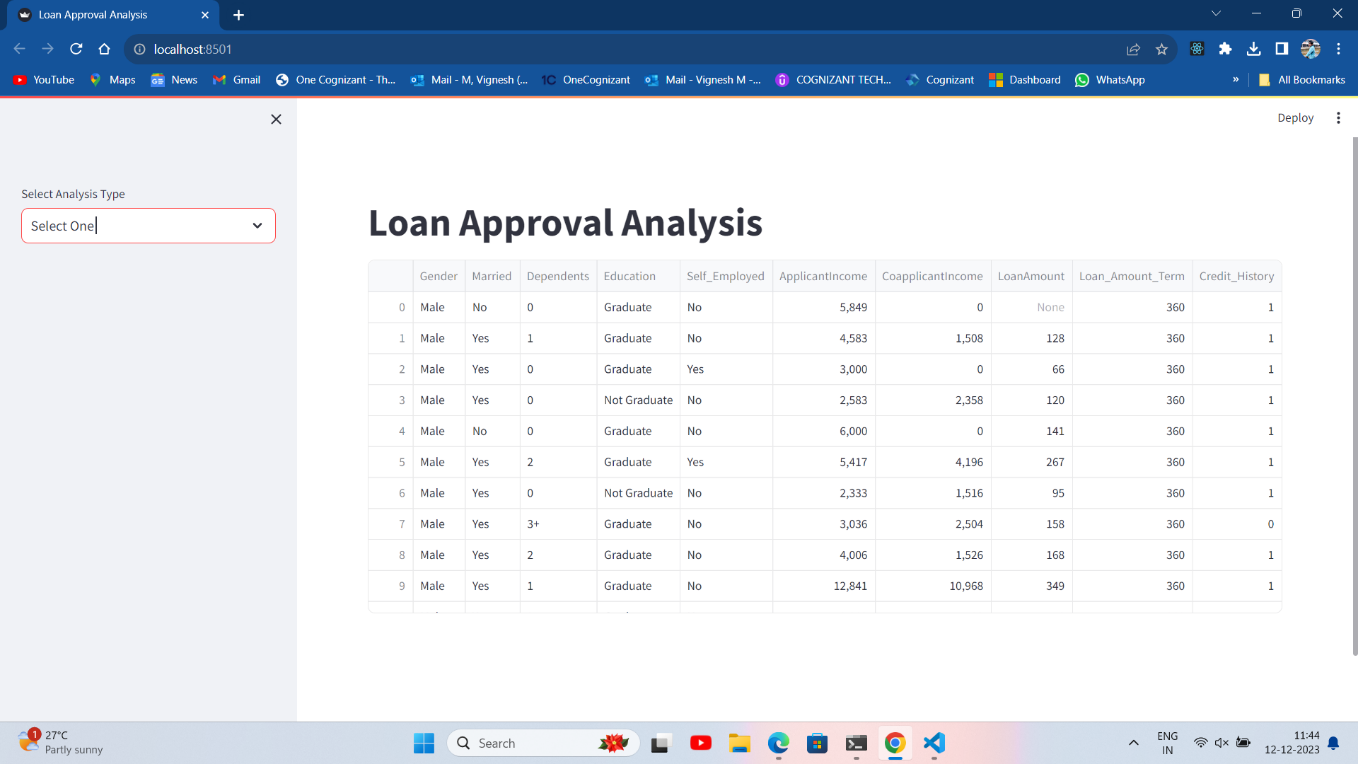
* Scatter plots are created for combinations of three variables: two numerical variables on the axes and a categorical variable represented by different colours. This helps in understanding how the relationship between numerical variables varies across different categories.

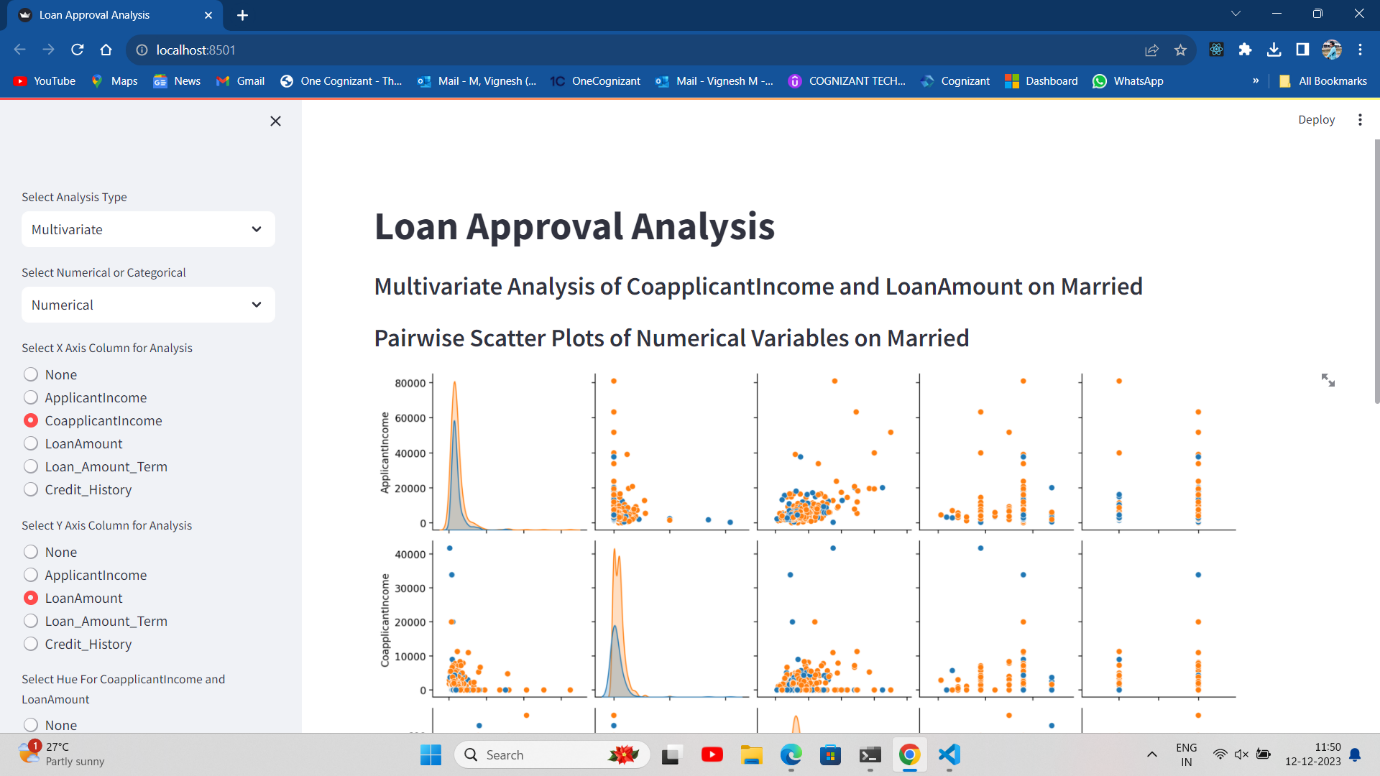
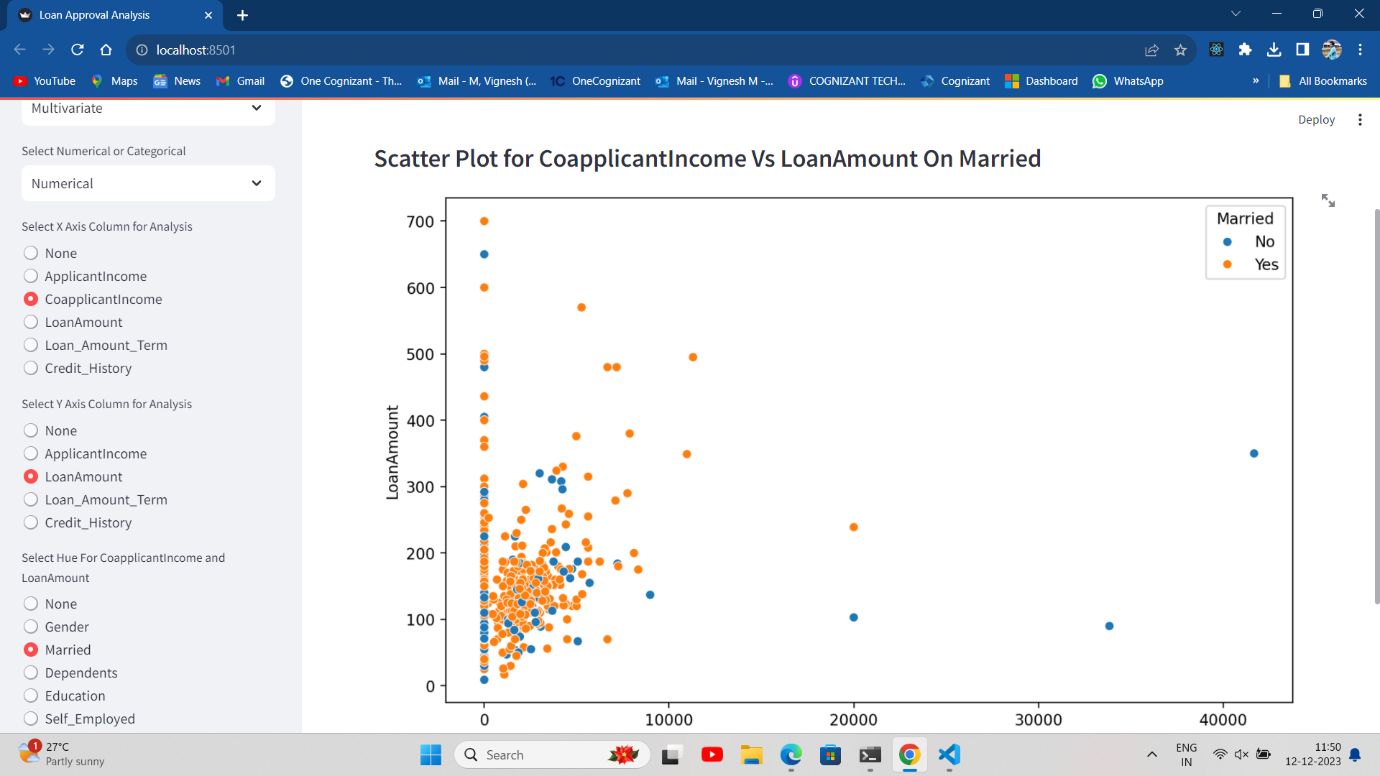
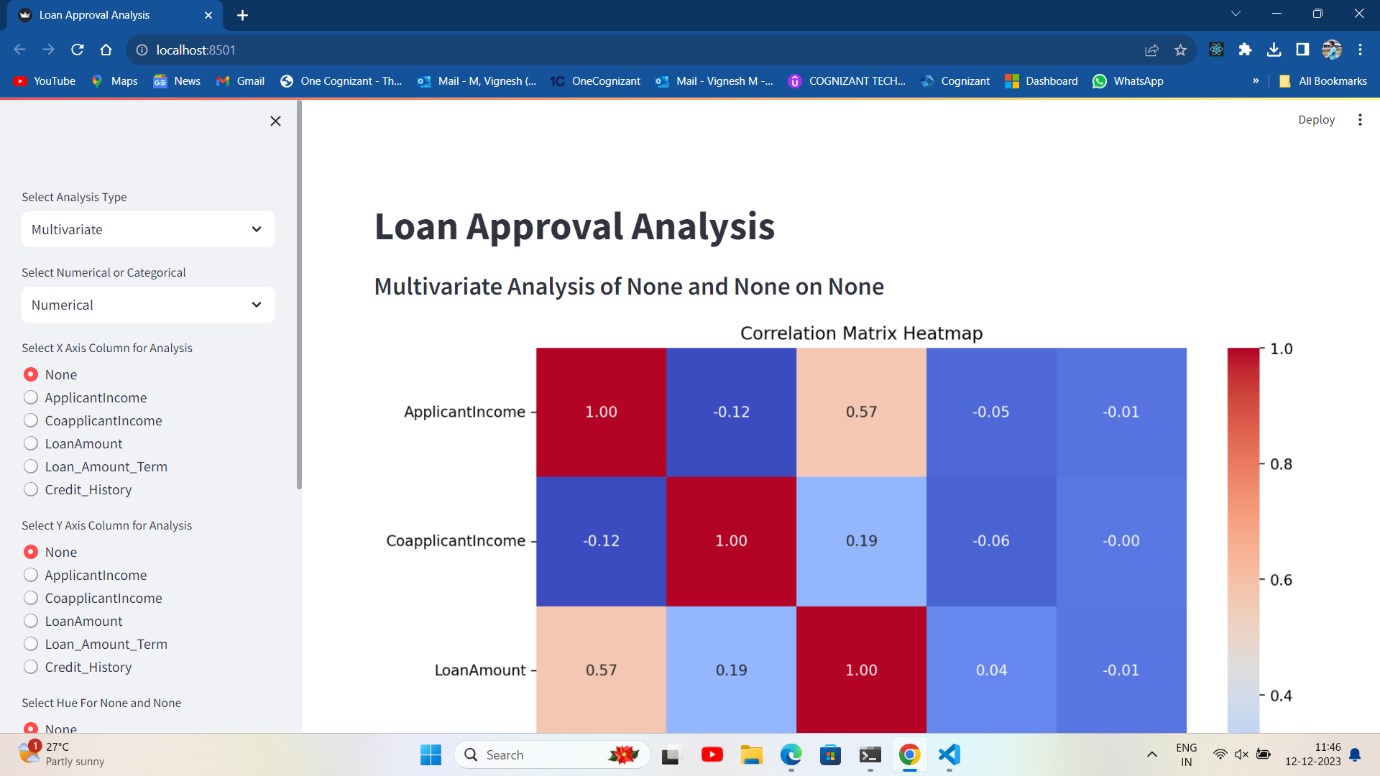
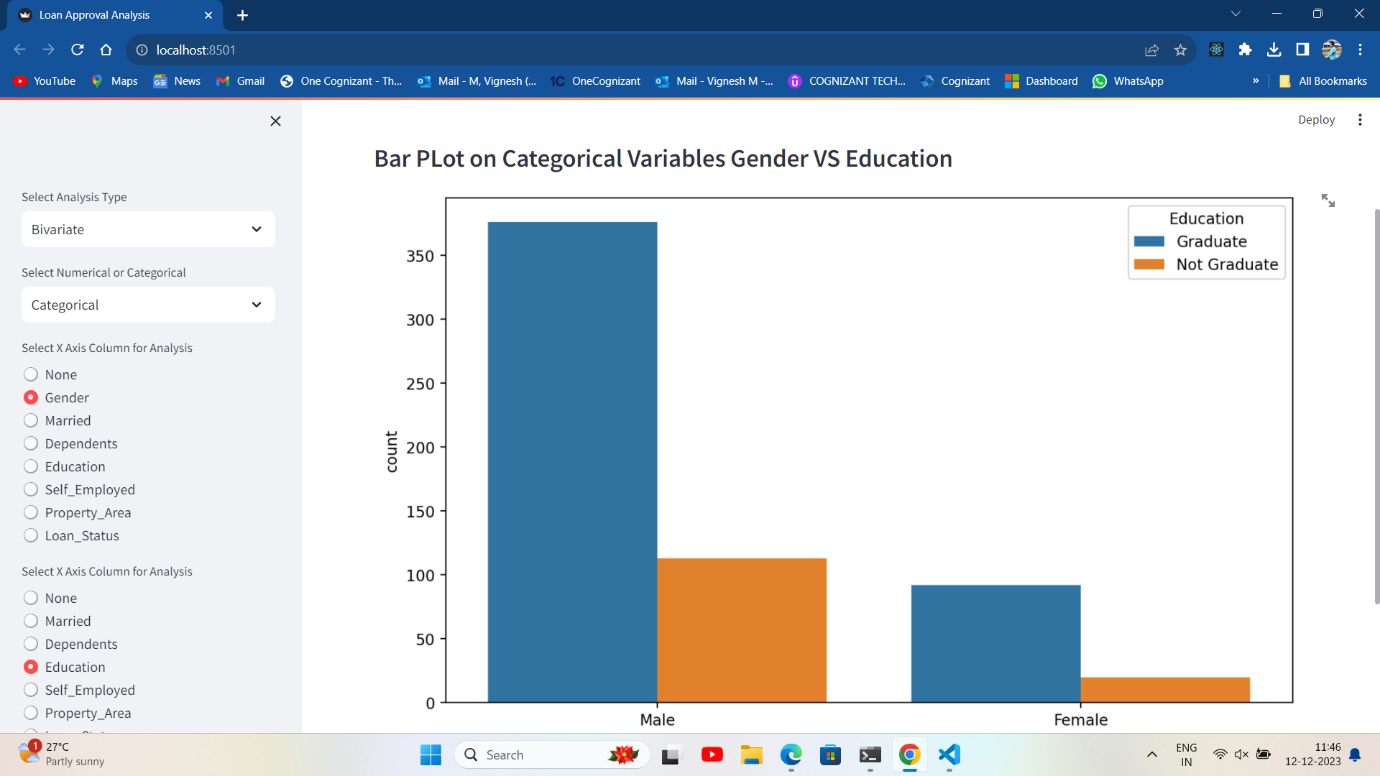
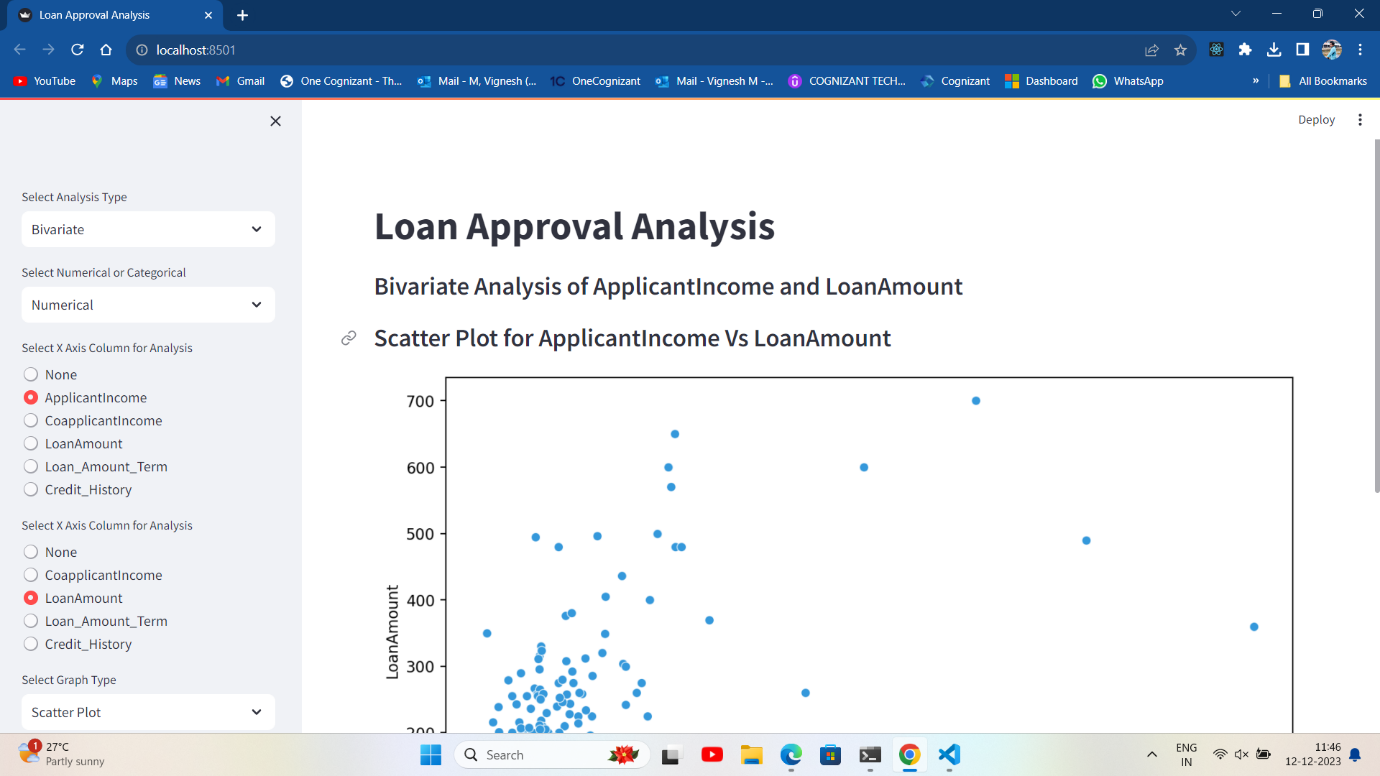
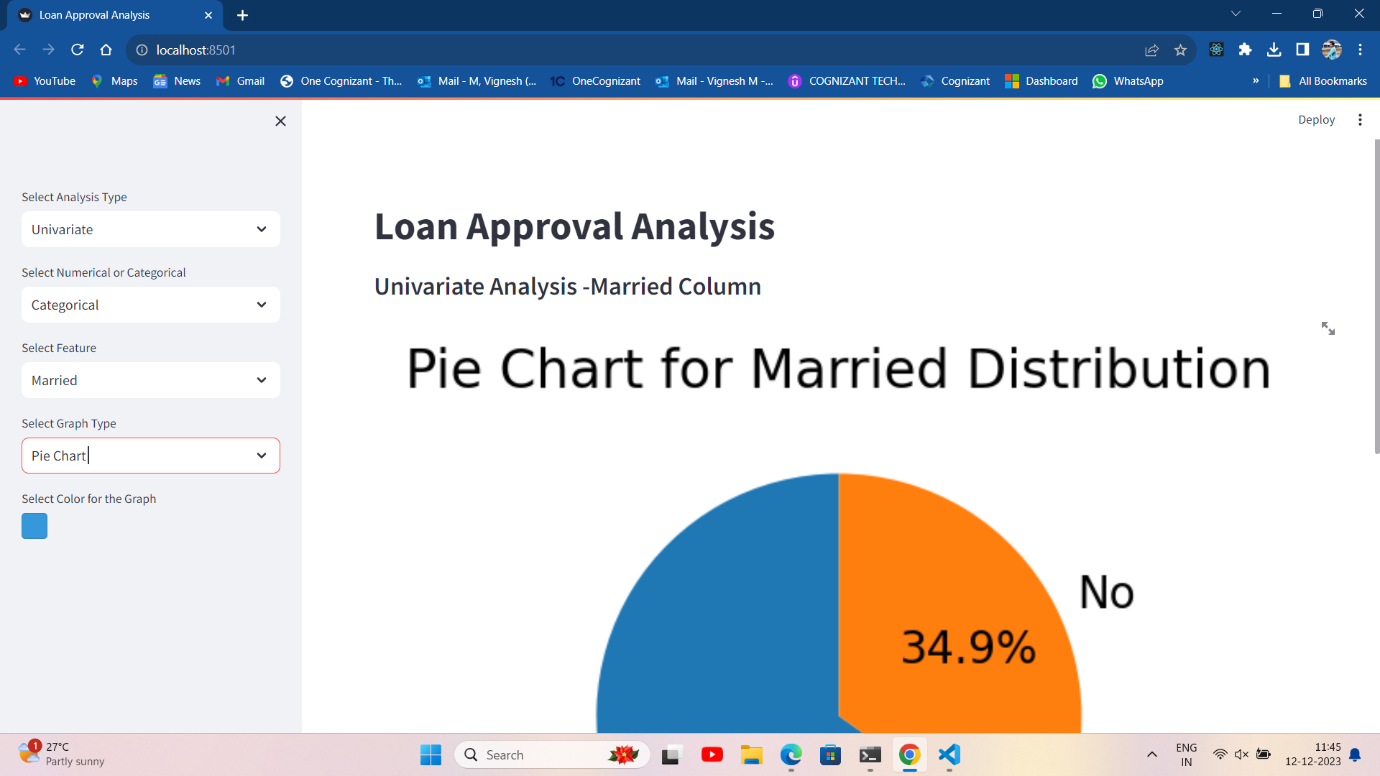
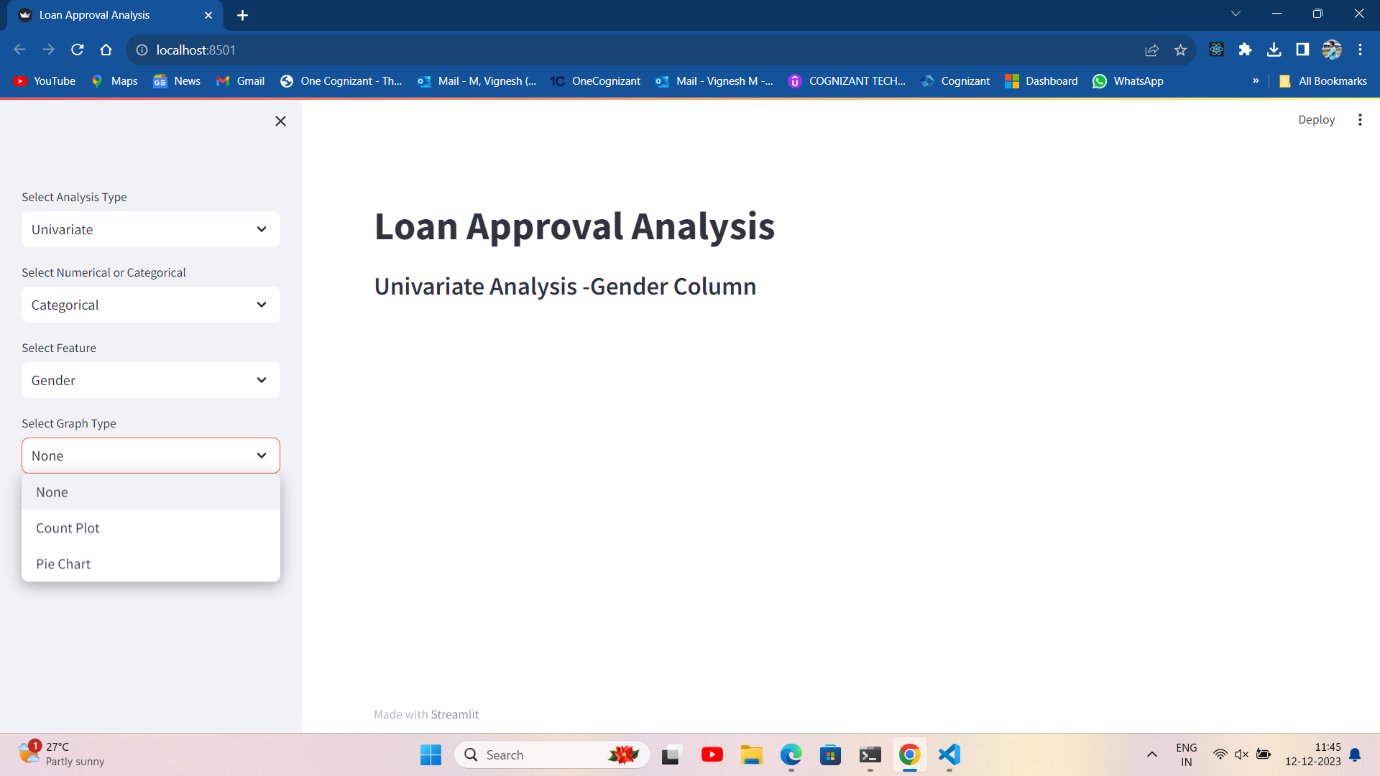
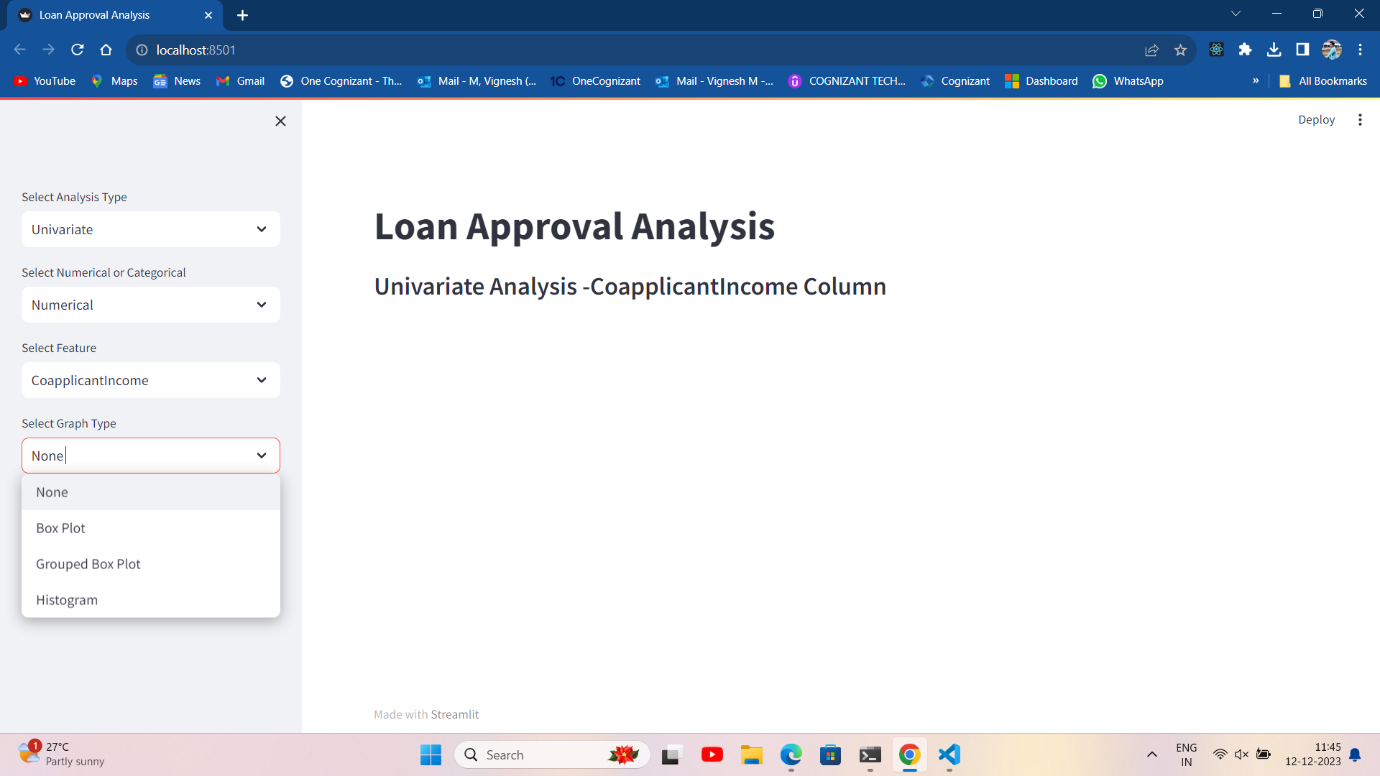
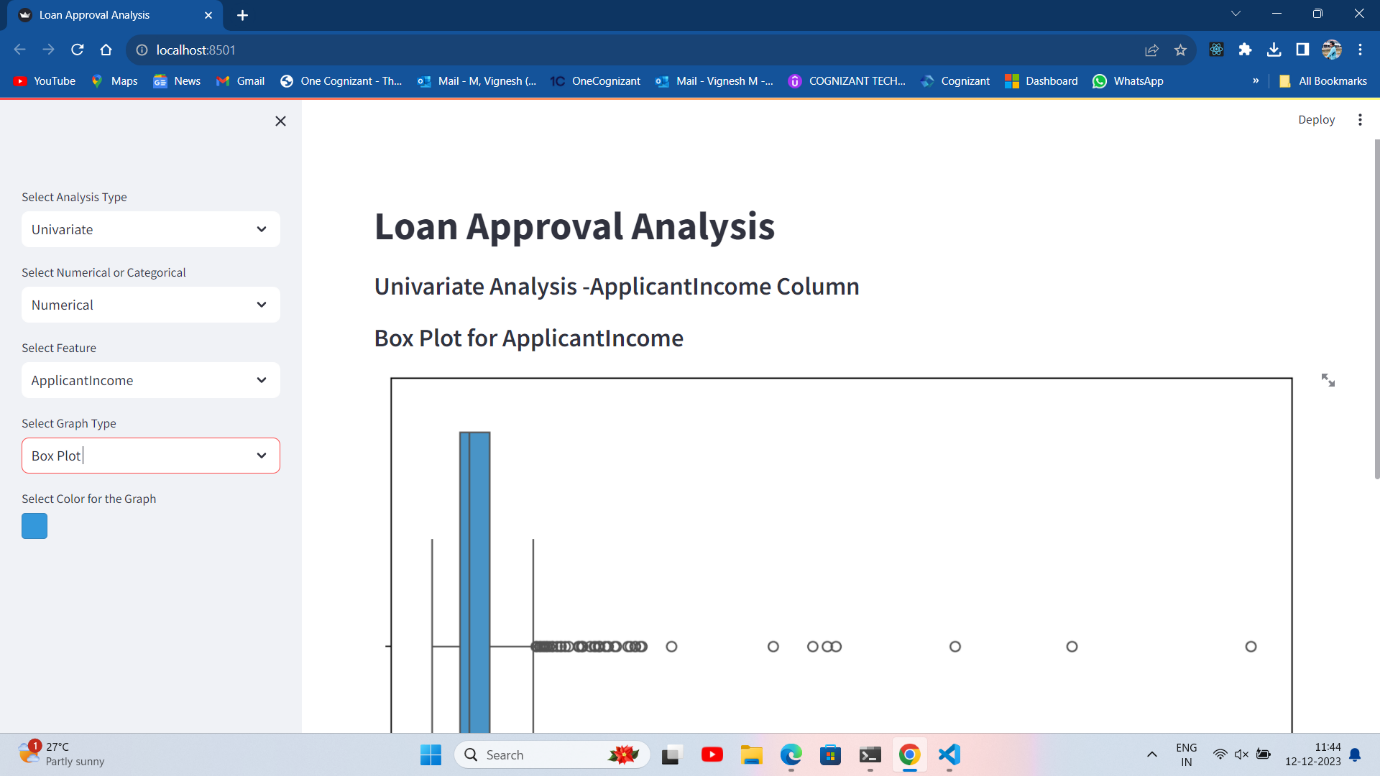
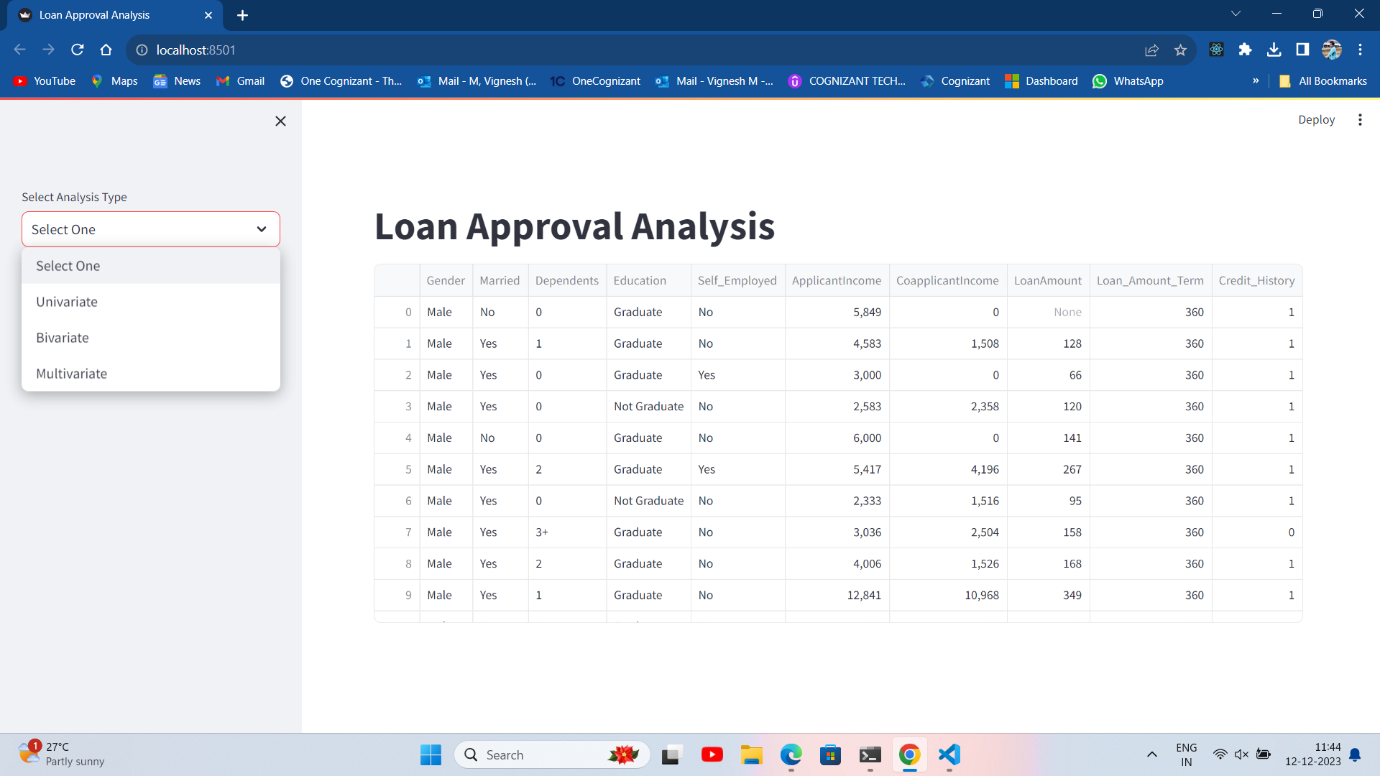
**Correlation Matrix Heatmap:**

* A heatmap of the correlation matrix is generated for numerical variables. This visualization helps identify linear relationships between numerical features, with correlation coefficients annotated on the heatmap.

**Key Findings and Recommendations:**

1. The distribution of numerical variables provides insights into their central tendency and spread.
2. Outliers in numerical variables can be identified from the box plots.
3. Categorical variables exhibit varying frequencies of categories, and pie charts offer a visual representation of their proportions.
4. Pairwise scatter plots provide insights into potential correlations or patterns between numerical variables.
5. Count plots for categorical variables reveal relationships between different categories.
6. Multivariate scatter plots with numerical variables and a categorical hue allow for a deeper understanding of how numerical features relate to each other in the context of different categories.
7. The correlation matrix heatmap shows the strength and direction of linear relationships between numerical variables.

**OUTPUT OF STREAM LIT : **

****