**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 26 September 2024 |
| Team ID | LTVIP2024TMID24973 |
| Project Title | Detection of Phishing Websites from URLs  Using Machine learning |
| Maximum Marks | 6 Marks |

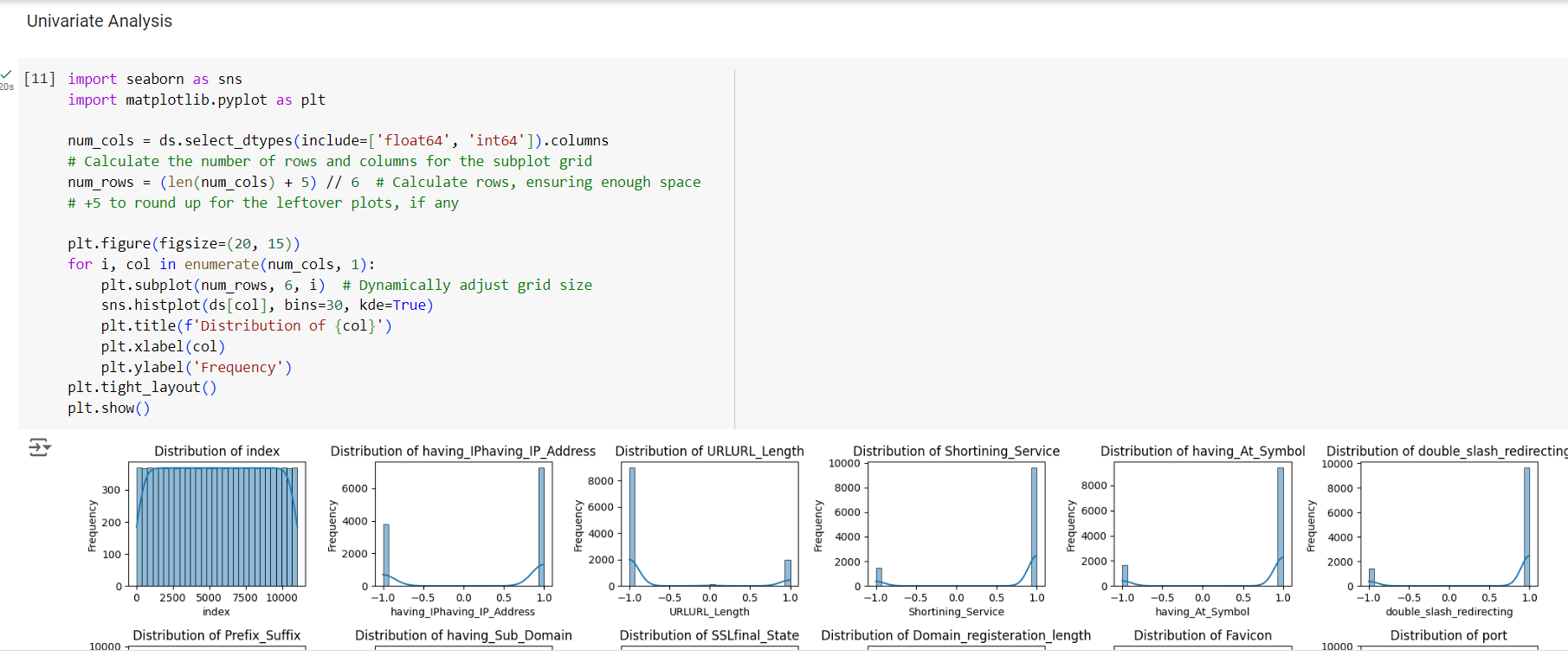
**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

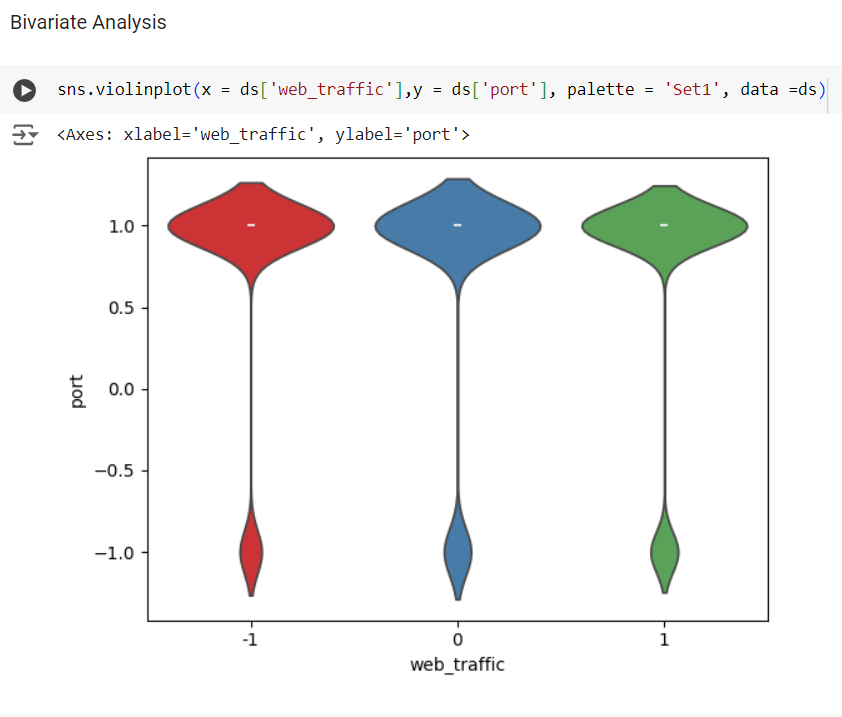
**1. Data Overview**

* **Basic Statistics**:
  + Calculate the mean, median, and mode of numeric features (e.g., URL length, number of subdomains).
  + Count the number of phishing vs. legitimate URLs.
  + Check for missing values in important fields like URL and labels.

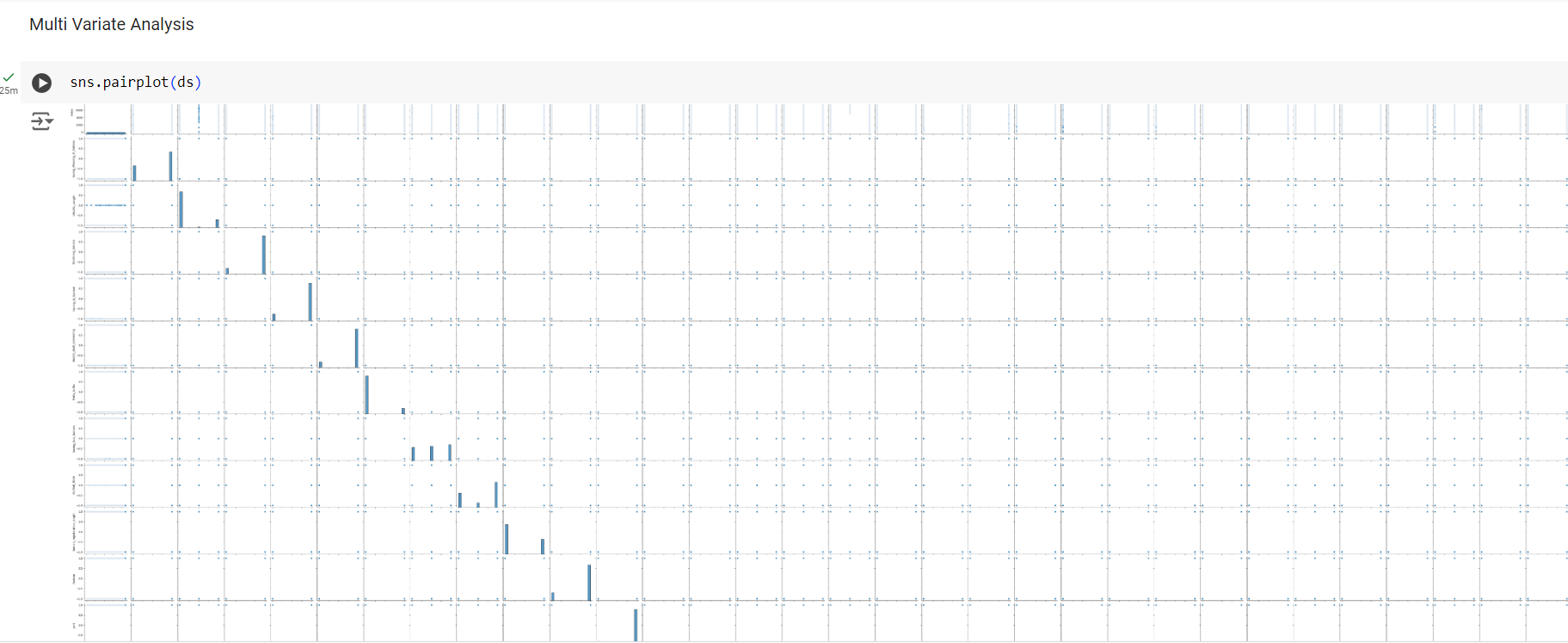
**2. Univariate Analysis**

****

**3. Bivariate Analysis**

****

**4. Multivariate Analysis**.



* **Visualizations**:
  + Generate heatmaps to show correlations among multiple variables.
  + Use 3D scatter plots if applicable to visualize interactions among three features.
* 

|  |  |
| --- | --- |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Checking Null Values |  |
| Feature Extraction |  |