

**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Project Title | Harvesting Brilliance: A Taxanomic Tale of  Pumpkin Seeds Varieties |
| 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.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | This section provides an overview of the pumpkin seed varieties dataset. It includes basic statistics such as the number of varieties, dimensions of the dataset (e.g., number of rows and columns), and the general structure of the data (e.g., types of variables, data types) |
| Univariate Analysis | This section focuses on analyzing individual variables within the pumpkin seed varieties dataset. It involves calculating and interpreting descriptive statistics like mean, median, mode, and standard deviation for each variable. |
| Bivariate Analysis | This section examines the relationships between two variables in the pumpkin seed varieties dataset. It includes techniques like correlation analysis and scatter plots to understand how different variables interact with each other. |
| Multivariate Analysis | This section investigates patterns and relationships involving multiple variables simultaneously. It involves more complex statistical methods to understand how different variables collectively influence certain outcomes. |
| Outliers and Anomalies | This section focuses on identifying and treating outliers and anomalies within the pumpkin seed varieties dataset. Outliers are data points that deviate significantly from the rest of the data, which can affect the analysis. |
| **Data Preprocessing Code Screenshots** | |



|  |  |
| --- | --- |
| Loading Data |  |
| Handling Missing Data |  |