

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 03 June 2024 |
| Team ID | 739676 |
| Project Title | Harvesting Brilliance: A Taxanomic Tale of Pumpkin Seeds Varieties in detail |
| Maximum Marks | 2 Marks |

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

| Data Source | Data Quality Issue | Severity | Resolution Plan |
|-----------------------|---------------------------|-----------------|--|
| Pumpkin Seeds Dataset | Incomplete Data | Moderate | Use data imputation methods to fill missing values. |
| Pumpkin Seeds Dataset | Duplicate Entries | Moderate | Identify and remove duplicate records |
| Pumpkin Seeds Dataset | Outdated Data | Low | Supplement with recent data or apply temporal analysis. |
| Pumpkin Seeds Dataset | Erroneous Data | High | Validate data entries and correct errors using predefined rules. |
| Pumpkin Seeds Dataset | Inconsistent Data | High | Standardize naming conventions and clean the data. |