**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | SWTID1720077079 |
| Project Title | Wild Blueberry Yield Prediction |
| Maximum Marks | 2 Marks |

**Data Quality Report**

The dataset, provided as a CSV file, is of high quality and well-suited for our blueberry yield prediction project. It contains no missing values or duplicate entries, ensuring data integrity. Additionally, the dataset does not require label encoding or one-hot encoding, simplifying the preprocessing steps. Overall, it is an excellent dataset for our intended purpose.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Kaggle | Outlier issue in honeybee feature | Moderate | Fitting the feature based on interquartile range found from boxplot. |