

**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 9 July 2024 |
| Team ID | team-739821 |
| Project Title | Precise Coffee Quality Prediction |
| 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** |
| Kaggle  Dataset | Some unwanted columns are  present in the dataset. | Medium | Dropped those columns using  Drop() function. |
| Kaggle  Dataset | Few Null values are present in  the dataset. | Low | Removed the null values from  dataset by using the following  code:  df.dropna(inplace=True) |
| Kaggle  Dataset | Outliers are present in some of  the columns of dataset. | Medium | Identified outliers by drawing  boxplot and removed them from  upper bound and lower bound. |



|  |  |  |  |
| --- | --- | --- | --- |
| Kaggle  Dataset | Some categorical values  consisting columns are present in  the dataset. | High | Converted the sting form of data  into integer by applying Label  Encoding on columns consisting  of categorical values. |