**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 29 April 2024 |
| Team ID | 737820 |
| Project Title | EcoForecast: AI-Powered Prediction of Carbon Monoxide Levels |
| 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** |
| Dataset-1 | Missing Values | Low | Imputation: Replace missing values with a suitable estimate. This could be the mean, median, mode, or a value predicted using machine learning algorithms. |
| Dataset-1 | Outliers | Moderate to high | Transformations: Apply transformations such as logarithmic or power transformations to make the data more normally distributed and reduce the impact of outliers. |
| Dataset-1 | Duplicate Records | Moderate | Remove Duplicates: Delete duplicate records from the dataset, keeping only one instance of each unique record. |