**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 8 June2024 |
| Team ID | 739763 |
| Project Title | Air Quality Index Analyzer using ML |
| 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 | Missing values in  PM2.5 4593  PM10 11140  NO 3582  NO2 3585  NOx 4185  NH3 10328  CO 2059  SO2 3854  O3 4022  Benzene 5623  Toluene 8041  Xylene 18109  AQI 4681  AQI\_Bucket 4681 | Moderate | Using KNN imputer  To fill the missing values |
| Kaggle | Categorical values ‘City’,’AQI\_Bucket | Moderate | Used Label Encoder to convert the categorical to numerical |