# Data Collection and Preprocessing Phase

| Date | 03 july 2024 |
| --- | --- |
| Team ID | 739647 |
| Project Title | Predicting co2-emissions by countries using machine learning |
| 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 | Dataset is too large it contains 55lakhs rows . Problem faced while loading the dataset. | Moderate | Used Kaggle commands in Google Colab to load the dataset |
| Kaggle Dataset | Categorical data in the dataset. | Moderate | Encoding has to be done in the data |