

Data Collection and Preprocessing Phase

Date	14 June 2024
Team ID	SWTID1749709340
Project Title	Predicting Co2 Emission 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
Exploring Co2 emission	Large Dataset with Shape: (5656458, 6).	High	Extracted a random subset of 1 lakh sample from the original dataset to reduce training time, and implemented advanced machine learning models like Random Forest for good accuracy
Exploring Co2 emission	Large number of categorical columns	Moderate	Used One-Hot Encoding for encoding categorical columns like CountryName, CountryCode, IndicatorName and IndicatorCode