Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720007638
Project Title	Predicting Co2 Emission By Countries Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	Our project's focus is the development of a machine learning model that can forecast CO2 emissions in different countries. We intend to create a comprehensive predictive framework by utilizing important factors like population, GDP, energy consumption, and industrial activities.
Data Collection Plan	The Kaggle website has been our main source for the project's primary dataset. To be more specific, we are using the World Development Indicators dataset available at Kaggle World Development Indicators. It is essential for our analysis and for predicting the modeling of CO2 emissions by countries. This dataset contains a broad array of indicators related to population, GDP, energy consumption, and industrial activities.

Raw Data Sources Identified	The main raw data source used in our project is Indicators.csv. This type of file creates a robust framework for analyzing CO2 emissions by providing essential contextual, methodological, and metric-related information.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Indicators	It lists various indicators used in the study, potentially including their codes, descriptions, and units of measurement.	https://www.kaggle.com/datasets/kaggle/world-development-indicators?resource=download&select=Indicators.csv	CSV	500MB	Public