

## Data Collection and Preprocessing Phase

Date	12 June 2025
Team ID	SWTID1749709340
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	This project involves predicting CO <sub>2</sub> emissions by country using machine learning techniques on a global development dataset. It includes trend analysis, feature encoding, and model-driven estimation of emission values based on country and year.
Data Collection Plan	The dataset was sourced from publicly available global development databases and includes country-level indicators across multiple years. Relevant features such as CO <sub>2</sub> emissions, country, and year were selected and imported in CSV format for analysis.

### Raw Data Sources Template

Source Name	Description	Location/ URL	Format	Size	Access Permissions
Kaggle- World Development Indicators	The World Development Indicators from the World Bank contain over a thousand annual indicators of economic development from hundreds of countries around the world.	<a href="#">Exploring Co2 emission</a>	CSV	574.31 MB	Public