

Data Collection and Preprocessing Phase

Date	15 JULY 2024
Team ID	740033
Project Title	Car Performance Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Identified raw data sources for car performance prediction include vehicle manufacturer specifications, fuel consumption records, user reviews, sales data, insurance claims, maintenance records, environmental data, sensor data, and market surveys.

Data Collection Plan Template

Section	Description
Project Overview	The project aims to predict car performance by analyzing various factors such as engine specifications, fuel type, and maintenance records, using machine learning techniques to provide accurate, data-driven insights for manufacturers and consumers.
Data Collection Plan	Gather data from manufacturers, maintenance records, fuel consumption logs, user reviews, sales data, insurance claims, sensor outputs, and environmental sources to create a comprehensive dataset for accurate car performance prediction using machine learning models.
Raw Data Sources Identified	Identified raw data sources for car performance prediction include vehicle manufacturer specifications, fuel consumption records, user reviews, sales data, insurance claims, maintenance records, environmental data, sensor data, and market surveys.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Historical data on Air Quality Data in India Air Quality	https://www.kaggle.com/datasets/rohanrao/airquality-data-inindia	CSV	76MB	Public

	Index (AQI) and hourly data across stations and cities in India				
--	--------------------------------------------------------------------------	--	--	--	--