

Data Collection and Preprocessing Phase

Date	7 July 2024
Team ID	SWTID1720104754
Project Title	Cereal Analysis Based On Rating By Using Machine Learning Techniques
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	"Development of a machine learning model for regression, including training, testing, and evaluating with R-squared, RMSE, and MAPE metrics, and saving using pickle."
Data Collection Plan	SKILLWALLET
Raw Data Sources Identified	The dataset for the " Nutrition data on 80 cereal products " project on Kaggle is approximately 2.17 KB in size. It contains information about different types of cereal products, including attributes such as calories, protein, fat, sodium, fiber, carbohydrates, and sugars.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
KAGGLE	The nutritional data on 80 cereal products in the dataset offers comprehensive information about the calories, protein, fat, sodium, fiber, and sugars content, enabling in-depth analysis of breakfast cereals' nutritional characteristics.	https://www.kaggle.com/crawford/80-cereals	CSV	5GB	Public