Data Collection and Preprocessing Phase

Date	21 June 2024
Team ID	739954
Project Title	Cereal analysis based on ratings 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	The project aims to analyze and predict cereal ratings using machine learning techniques. Objectives include identifying top-rated cereals, understanding key factors influencing ratings, and developing a predictive model for future ratings.
Data Collection Plan	Data will be collected from various sources such as online cereal reviews, consumer surveys, manufacturer data, market research reports, retail sales data, social media feedback, expert reviews, and nutritional information.
Raw Data Sources Identified	**Online Cereal Reviews:** Websites that aggregate user reviews and ratings for cereals. **Consumer Surveys:** Direct feedback from consumers about their cereal preferences and ratings. **Manufacturer Data:** Official ratings and information provided by cereal manufacturers. **Retail Sales Data:** Sales figures and ratings data from major retail stores. **Expert Reviews:** Ratings and reviews from food critics and nutritio

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Access Permissions
Dataset 1	In this project we have used .csv data. This data is downloaded from kaggle.com.	https://www.kaggle.com/crawford/80-cereals	CSV	Public
Dataset 2	Access Google Drive with a Google account (for personal use) or Google Workspace account	https://www.kaggle.com/crawford/80-cereals	Excel	Private (with access)