Data Collection and Preprocessing Phase

Date	15 April 2024
Team ID	738220
Project Title	Walmart Sales Analysis for Retail Industry with Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

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

Section	Description
Project Overview	The project focuses on sales forecasting for Walmart, a renowned retail corporation operating a chain of hypermarkets. The goal is to accurately estimate future sales using past sales data, store information, and other relevant factors such as holidays and promotions. Accurate sales forecasts enable Walmart to make informed business decisions, optimize operations, and predict short-term and long-term performance.
Data Collection Plan	- Gather data from publicly available repositories, particularly focusing on the Kaggle platform for comprehensive and high-quality datasets. - Prioritize datasets that encompass historical data, detailed store profiles, promotional campaign records, and comprehensive holiday-related data.

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, the popular platform for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as store information, department, size, weekly sales, temperature, holidays, markdowns.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The dataset includes historical sales data from 45 Walmart stores across different regions, with each store having multiple departments. With additional information about promotional markdowns that happen before the prominent holidays.	https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data	CSV	3.22 MB	Public