

Data Collection and Preprocessing Phase

Date	4 June, 2024
Team ID	SWTID1719938571
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	This project aims to enhance Walmart's sales performance using machine learning and historical data. It focuses on forecasting demand, segmenting customers, and optimizing pricing with algorithms like Random Forest, Decision Tree, XGBoost, and ARIMA. By analyzing sales trends and holiday impacts, the project will develop predictive models integrated into a user-friendly Flask application. This approach is expected to improve inventory management, customer satisfaction, and revenue.
Data Collection Plan	<ul style="list-style-type: none"> Walmart dataset (Walmart store sales prediction) from Kaggle Retail analysis with Walmart sales data from Kaggle Walmart Recruiting – Store Sales Forecasting from Kaggle

Raw Data Sources Identified	The raw data sources for this project includes the dataset obtained from Kaggle, the popular platform for data science competitions and repositories. The provided data predicts department-wide sales for 45 Walmart stores using historical data. Model the impact of promotional markdowns preceding major holidays, which are weighted five times higher in the evaluation.
-----------------------------	---

Raw Data Sources:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Walmart Recruiting – Store Sales Forecasting Predict sales for 45 Walmart stores, focusing on department-wide performance and the impact of promotional markdowns before major holidays, which have higher evaluation weights.	https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data	CSV	3.22 MB	Public