

Data Collection and Preprocessing Phase

Date	9 th July 2024
Team ID	SWTID1720435231
Project Title	Walmart Sales Analysis For Retail Industry With Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

The Data Collection Plan and the Raw Data Sources report enable complete data curation and integrity, allowing for informed decision-making in all analyses and decision-making endeavours.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project will employ data analysis to identify the optimal prices for products, classify customers based on their purchasing preferences, and forecast which products Walmart should carry.
Data Collection Plan	<ul style="list-style-type: none"> ▪ Search for datasets involving the sales data of Walmart. ▪ Assign datasets with a range of demographic data a higher priority. ▪ The dataset must have data from both holidays and non-holidays.
Raw Data Sources Identified	The Raw data for this project was collected from Kaggle, a trusted and reliable platform for data collection and repositories. The

	provided dataset has information regarding the date, the store data was collected from and whether or not a particular day is a holiday
--	---

Raw Data Sources Template

Source Name	Description	Location/ URL	Format	Size	Access Permission
train.csv	<p>This is the historical training data, which covers from 2010-02-05 to 2012-11-01. Within this file you will find the following fields:</p> <ul style="list-style-type: none"> • Store - the store number • Dept - the department number • Date - the week • Weekly_Sales - sales for the given department in the given store • IsHoliday - whether the week is a special holiday week 	https://www.kaggle.com/competition/s/walmart-recruiting-store-sales-forecasting/data?select=train.csv.zip	CSV	2.59 MB	Public
Features.csv	<p>This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:</p> <ul style="list-style-type: none"> • Store - the store number • Date - the week • Temperature - average temperature in the region • Fuel_Price - cost of fuel in the region • Markdown1-5 - anonymized data related to promotional markdowns that Walmart is running. 	https://www.kaggle.com/competition/s/walmart-recruiting-store-sales-forecasting/data?select=features.csv.zip	CSV	161.7 kB	Public

	<p>MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA.</p> <ul style="list-style-type: none">• CPI - the consumer price index• Unemployment - the unemployment rate• IsHoliday - whether the week is a special holiday week				
--	---	--	--	--	--