Data Collection and Preprocessing Phase

Date	5 May 2024
Team ID	737906
Project Title	Walmart Sales Analysis For Retail Industry
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 Walmart sales retail industry aims to techniques to analyze and forecast sales data within the retail sector and empower Walmart with data-driven insights and predictive capabilities to optimize operations, improve customer satisfaction, and drive business growth within the retail sector.
Data Collection Plan	Data will be collected from various sources to build comprehensive datasets for analysis and modeling. the sources from which the data will be collected is Walmart point of sale(pos) ,walmart store attritube ,product information ,external factors ,customer data.
Raw Data Sources Identified	the raw data sources identified for the Walmart sales retail industry , along with relevant details:

1.Walmart Point of Sale (POS) Systems:

Description: Transactional data captured at Walmart's POS systems, providing details of sales transactions, including sales amounts, product SKUs, quantity sold, transaction timestamps, and store IDs.

2.Walmart Store Attributes Database:

Description: Database containing attributes of Walmart stores, such as store location (address, city, state, zip code), store type (supercenter, neighborhood market), store size (square footage), number of checkout lanes, and departmental information.

3.Product Information Database:

Description: Database containing information on product attributes and characteristics sold at Walmart stores, including product categories and subcategories, brand names, product descriptions, pricing information, and dimensions.

4.Promotional Calendars:

Description: Records of promotional activities and special offers run by Walmart, including promotional start and end dates, types of promotions (e.g., BOGO, percentage discounts), and associated products and categories.

5.External Data Providers (Economic Indicators and Weather Data):

Description: External sources providing data on economic indicators (e.g., GDP, unemployment rate) and weather conditions (temperature, precipitation, humidity) that may influence sales performance at Walmart stores.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The Walmart Point of Sale (POS) Systems data source contains transactional data captured at Walmart's POS systems across various store locations. It provides detailed information about sales transactions, including sales amounts, product SKUs, quantity sold, transaction timestamps, and store IDs.	https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data	CSV	3.22 MB	Public
Dataset 2	Description of the data in this source.	Link of Dataset 2	Excel	YY GB	Private (with access)
…	…	…	…	…	…