DataCollectionandPreprocessingPhase

Date	10-july-2024
TeamID	739969
ProjectTitle	WalmartSalesAnalysisForRetailIndustryWith Machine Learning
MaximumMarks	2Marks

DataCollectionPlan&RawDataSourcesIdentificationReport:

ElevateyourdatastrategywiththeDataCollectionplanandtheRawDataSourcesreport,ensuringmeticulousdata curation and integrity for informed decision-making in every analysisand decision-making endeavour**.**

DataCollectionPlan:

Section	**Description**
ProjectOverview • •	The machine learning project aims to predict sales for Walmart stores based on historical sales data andvariousinfluencingfactors.Usingadatasetwith features such as store , department, holiday, weekend days, promotional events, temperature, unemployment,Weekly_Salesandhistorical sales; the objective is to build a model that accurately forecasts sales, facilitating efficientand informed decision-making in inventory management, staffing, and marketing strategies.
DataCollectionPlan	Searchfordatasetsrelatedtoretailsales, holidays, and promotional events. Prioritize datasets with detailed sales informationanddiverseinfluencingfactors.
RawDataSources Identified	The raw data sources for this project include datasetsobtained fromWalmart'sinternal sales records, Kaggle, and other relevant data repositories. The provided sample data represents a subset of the collectedinformation,encompassingvariablessuc has storelocation,salesfigures,holidayperiods,and promotional events for machine learninganalysis.

RawDataSourcesReport:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	1.storedatasetcontainsstore,type ,size2.train dataset contains store,Dept,Date,Weekly_Sales,IsHoliday 3.feature dataset contains store,date,temperature,Fuel_Price,MarkDown1,MarkDown2,MarkDown3,MarkDown4,MarkDown5,CPI,unemployment, IsHoliday	https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data	csv	3.2 MB	Public