☐ Sales Data Analysis & Forecasting

Planning & Design Documentation

1. Planning & Design

☐ *Project Objective*

The objective of this project is to analyze historical sales data and forecast future sales trends using machine learning and time-series forecasting techniques. The system aims to provide insights into historical performance and generate reliable future predictions.

Key System Functions:

- Data preprocessing and cleaning
- Sales trend analysis
- Linear Regression-based prediction
- ARIMA time-series forecasting
- Performance evaluation using MAE (Mean Absolute Error)
- Graphical visualization of results

2. Design of Program Structure

The project follows a **modular and layered architecture** to ensure maintainability and scalability.

☐ *Directory Structure*

The project is organized into the following files and directories:

- `data/` : Contains the raw dataset (e.g., `walmart_sales.csv`).
- `sales_analysis.py` : The main logic for data analysis.
- `Test.py` : Unit testing for system validation.
- `requirements.txt` : List of necessary Python libraries.
- `README.md` : Project documentation and setup instructions.

☐ *Module Design*

The system is divided into five logical modules:

Module	Responsibility	Key Functions
Data Handling	Dataset reading, cleaning, and aggregation	load_data(), aggregate_sales()

Machine Learning	Training regression models and predicting trends	train_linear_regression()
Time Series	ARIMA-based future forecasting	arima_forecast()
Visualization	Plotting sales trends and model results	plot_sales_trend(), plot_arima_forecast()
Testing	Unit testing and output validation	Managed in Test.py

3. Algorithms

☐ *Algorithm 1: Data Loading & Preprocessing*

- **Objective:** Prepare dataset for modeling.
- **Steps:** Read CSV via pandas, convert dates to datetime format, sort by date, and aggregate Weekly_Sales using sum().

☐ *Algorithm 2: Linear Regression*

- **Objective:** Predict sales trend over time.
- **Mathematical Model:** $y = b_0 + b_1 x$
  - ○ y = Predicted sales
  - ○ x = Time (Date converted into numerical ordinal format)

Steps:

1. Convert the Date column into numerical ordinal format.
2. Define the feature variable XX as the numerical representation of Date.
3. Define the target variable yy as Weekly_Sales.
4. Fit the Linear Regression model using the training data.
5. Generate predicted sales values.
6. Compute Mean Absolute Error (MAE) to evaluate model performance.

☐ *Algorithm 3: ARIMA Forecasting*

- **Objective:** Forecast future weekly sales using **ARIMA(5,1,0)**.

Steps:

1. Select the historical Weekly_Sales time series.
2. Apply first-order differencing ($d = 1$d=1) to make the series stationary.
3. Fit the ARIMA(5,1,0) model to the data.
4. Forecast sales for the next 12 weeks.
5. Plot and analyze the forecasted values.

4. Pseudocode

☐ Main Program logic
START

  Load dataset and Convert Date column

  Aggregate weekly sales

  Plot historical sales trend

  Train Linear Regression model & Calculate MAE

  Train ARIMA model & Forecast next 12 weeks

  Plot all visualizations

END

5. Object-Oriented Design (Proposed)

While currently function-based, the system is designed to be structured into the following classes:

- **DataProcessor**: Handles `load_data()` and `aggregate_sales()`.
- **SalesPredictor**: Manages regression training and evaluation.
- **TimeSeriesForecaster**: Handles ARIMA training and forecasting.
- **Visualizer**: Generates all graph outputs.

Inheritance Design: A `BaseModel` abstraction can be used where both `LinearRegression` and `ARIMAModel` implement `train()` and `predict()` methods.

6. Data Flow Summary

The flow starts with a **CSV File**, which passes through the **DataProcessor** to create **Processed Sales Data**. This data is fed into both the **Linear Regression** and **ARIMA** models to generate **Predictions** and **Forecasts**, respectively. Finally, the **Visualizer** produces the **Graph Output**.

☐ Key Objects Used

Object	Description
df	Raw dataset
sales_data	Aggregated dataset

model	Trained regression model
predictions	Predicted sales
forecast	Future ARIMA forecast