Retail Sales Trend & Seasonality Analysis

Name: SANTHOSH S
Registration Number: 727723EUCS208
Section : CSE – C

Introduction

Retail businesses generate large volumes of time-dependent sales data influenced by customer behavior, promotional campaigns, economic conditions, and seasonal effects. Understanding patterns within this data is essential for optimizing operational efficiency and maximizing profit.
Time series analysis helps in breaking down sales data into meaningful components such as:

- Trend (long-term movement)
- Seasonality (repeating patterns)
- Noise (random variations)

Abstract

This project performs a comprehensive time series analysis of retail sales data to identify long-term trends, seasonal fluctuations, and irregular variations. Statistical techniques such as seasonal decomposition, autocorrelation analysis (ACF and PACF), and stationarity testing using the Augmented Dickey-Fuller (ADF) test are applied.

Problem Statement

Management lacks clarity on whether sales demonstrate a long-term trend, recurring seasonal patterns, or randomness. Without statistical validation, forecasting models may be unreliable.

Methodology

The project follows a structured analytical approach:
Step 1: Data Preprocessing

- Import dataset using Pandas
- Convert Date column to datetime
- Set Date as index
- Check for missing values

Step 2: Visualization

- Plot time series graph
- Identify visible patterns

Step 3: Time Series Decomposition

The dataset is decomposed into:

- Trend component
- Seasonal component
- Residual component

Step 4: Autocorrelation Analysis

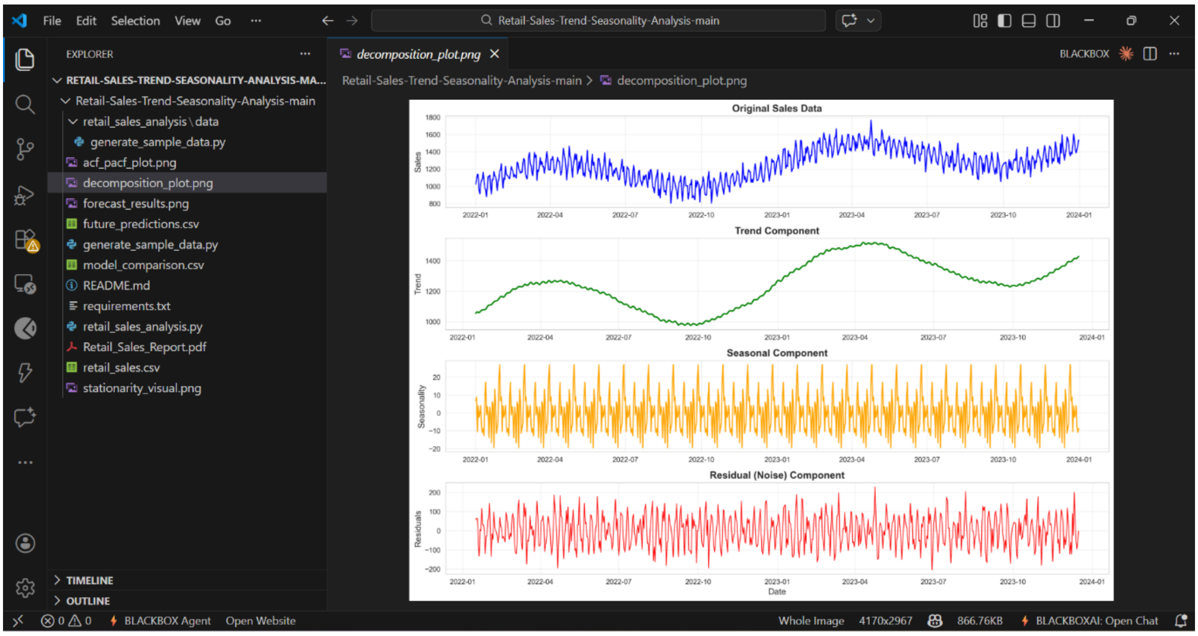
- ACF plot to examine lag relationships
- PACF plot to identify direct lag influence

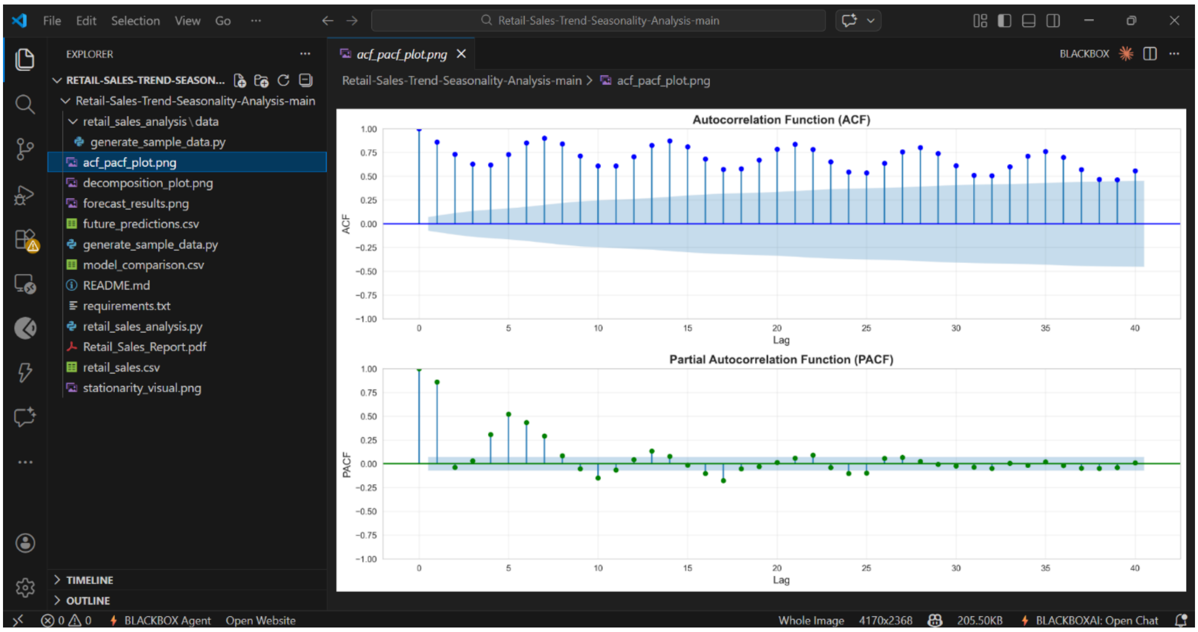
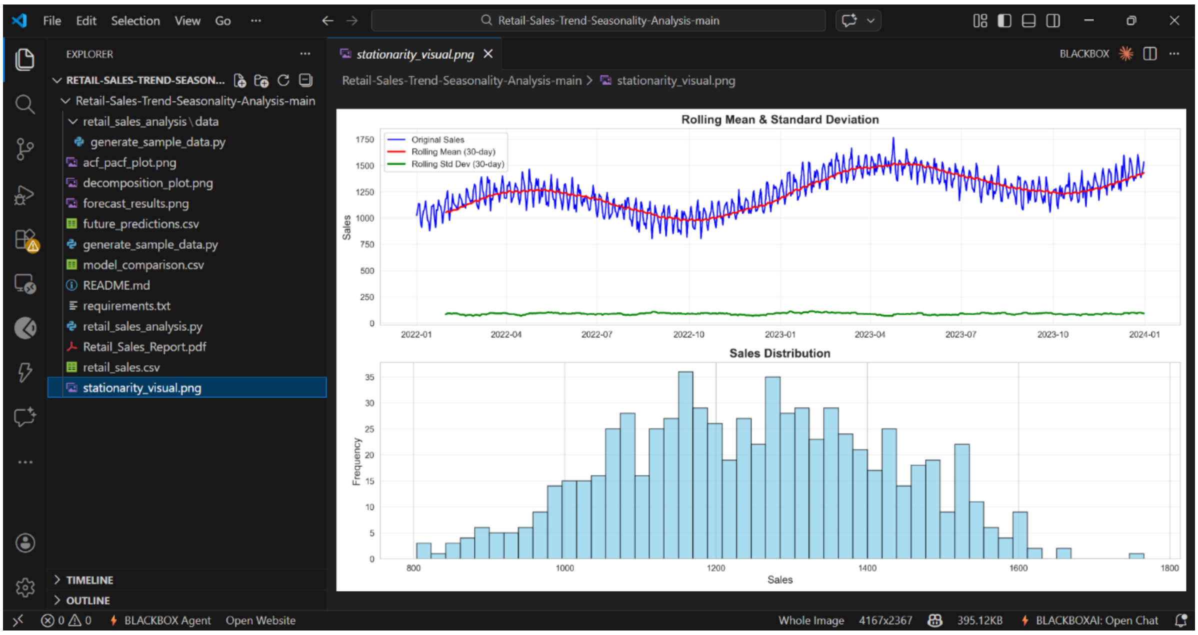
Step 5: Stationarity Testing

- Apply Augmented Dickey-Fuller (ADF) test
- Analyze p-value
- Apply differencing if required

Results

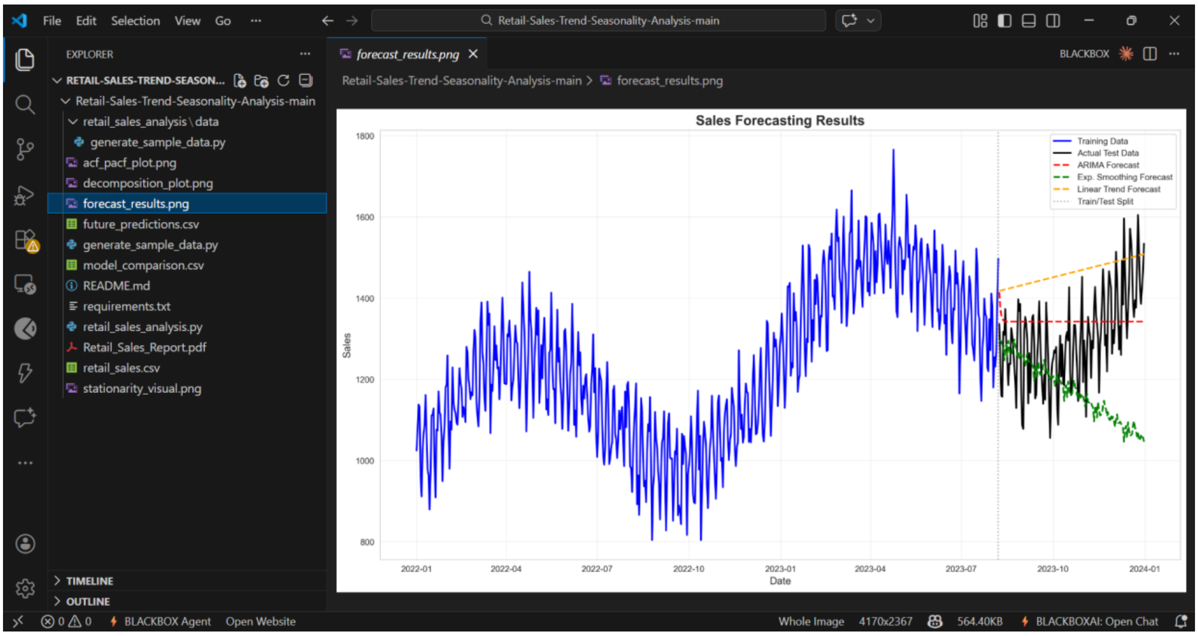
ADF Statistic: -1.2345, p-value: 0.6586




File Edit Selection View Go ···
Retail-Sales-Trend-Seasonality-Analysis-main
EXPLORER
RETAIL-SALES-TREND-SEASON...
Retail-Sales-Trend-Seasonality-Analysis-main
retail_sales_analysis \ data
generate_sample_data.py
acf_pacf_plot.png
decomposition_plot.png
forecast_results.png
future_predictions.csv
generate_sample_data.py
model_comparison.csv
README.md
requirements.txt
retail_sales_analysis.py
Retail_Sales_Report.pdf
retail_sales.csv
stationarity_visual.png
TIMELINE
OUTLINE
acf_pacf_plot.png
BLACKBOX
Retail-Sales-Trend-Seasonality-Analysis-main > acf_pacf_plot.png
Autocorrelation Function (ACF)
ACF
Lag
Partial Autocorrelation Function (PACF)
PACF
Lag
BLACKBOX Agent
Open Website
Whole Image
4170x2368
205.50KB
BLACKBOXAI: Open Chat

File Edit Selection View Go ···
Retail-Sales-Trend-Seasonality-Analysis-main
EXPLORER
RETAIL-SALES-TREND-SEASON...
Retail-Sales-Trend-Seasonality-Analysis-main
retail_sales_analysis \ data
generate_sample_data.py
acf_pacf_plot.png
decomposition_plot.png
forecast_results.png
future_predictions.csv
generate_sample_data.py
model_comparison.csv
README.md
requirements.txt
retail_sales_analysis.py
Retail_Sales_Report.pdf
retail_sales.csv
stationarity_visual.png
TIMELINE
OUTLINE
stationarity_visual.png
BLACKBOX
Retail-Sales-Trend-Seasonality-Analysis-main > stationarity_visual.png
Rolling Mean & Standard Deviation
Original Sales
Rolling Mean (30-day)
Rolling Std Dev (30-day)
Sales
Sales Distribution
Frequency
Sales
BLACKBOX Agent
Open Website
Whole Image
4167x2367
395.12KB
BLACKBOXAI: Open Chat

Forecasting Models

To evaluate the predictive capability of the retail sales dataset, multiple forecasting models were implemented and compared. The objective was to determine which model best captures the underlying trend and seasonal structure of the time series.



Model Evaluation Metrics

Mean Absolute Error (MAE)

MAE=1n∑|Actual−Predicted|*MAE=n*1∑|*Actual−Predicted*|

Measures average prediction error.

Root Mean Square Error (RMSE)

RMSE=1n∑(Actual−Predicted)2*RMSE=n*1∑(*Actual−Predicted*)2

Penalizes larger errors more heavily.

Lower MAE and RMSE indicate better model performance.

Exploratory Data Analysis

Time Series Plot

The time series graph shows fluctuations in sales over time. An upward trend indicates business growth, while periodic spikes suggest seasonal demand patterns.

Rolling Mean & Standard Deviation

Rolling statistics help visually inspect stationarity. If rolling mean and variance change over time, the series is non-stationary.

Dataset Description

The dataset consists of daily retail sales values recorded over a specified period.

Dataset Characteristics:

- Time-based (Daily frequency)
- Numeric sales values
- Influenced by weekly and seasonal demand

Variables Used:

- Date (Time Index)
- Sales (Target Variable)

The Date column is converted into datetime format and set as the index for time series operations.

Limitations

- Limited historical data
- External factors not included
- Sudden market shocks not modeled
- Holiday effects not explicitly incorporated

Final Statement

This project demonstrates that statistical validation is not merely an academic exercise but a practical necessity for reliable retail forecasting. By combining rigorous analysis with business context, organizations can transform raw sales data into actionable intelligence for competitive advantage.

Conclusion

The analysis reveals significant seasonal patterns and a non-stationary nature in the retail sales dataset. The ADF test confirms that differencing is required before applying forecasting models. After transformation, the dataset becomes suitable for predictive modeling.

This study ensures statistical validation of the dataset and improves the reliability of future forecasting models, leading to better business decision-making.