Walmart Sales Analysis

Data visualization domain internship

AMAL SALILAN

Walmart Sales Analysis

Our project analyzes Walmart's sales data to uncover customer buying patterns and trends. The goal is to enhance Walmart's strategic planning, marketing efforts, and product management while improving the shopping experience for customers.

A G E N D A

- **Introduction**

- **Dataset outline**

- **Tech Stack**

- **Exploratory Data Analysis (EDA)**

- **Modeling**

- **Deployment**

- **Challenges**

- **Result and the future scope**

Overview

The project is structured around three key objectives:

1. Understanding Customer Groups

Identifying different types of customers based on their purchasing behavior, such as high-spending or frequent buyers. Using these insights to offer personalized deals, discounts, and services that cater to specific customer needs.

2. Predicting Future Sales

Analyzing past sales trends to estimate future sales and profit figures. Helping Walmart anticipate customer demand and prepare for peak shopping periods.

3. Improving the Customer Experience

Utilizing data-driven insights to ensure products are available when needed. Providing tailored promotions and a seamless shopping experience.

Methodology:

Preparing the Data
Collected Walmart's sales data, including transaction dates, sales amounts, profits, and regional performance. Cleaned and structured the data for effective analysis.

Grouping Customers
Categorized customers based on shopping habits to identify spending patterns. Recognized trends in customer behavior, such as high-value shoppers and frequent buyers.

Predicting Sales
Studied historical sales and profit trends to forecast future performance. Provided insights to help Walmart plan for seasonal demand and promotional events.

Visualizing Results
Created clear and simple charts to illustrate customer segments and sales trends.

DATASET OUTLINE...

About the Data
- Time period covered by the data : 4years
- Key metrics/headline figures :
- Sales, Profit, discount data
- Customer demographics &order details
- Product hierarchy (category/subcategory)
- Shipping & logistics information

Why Walmart Dataset?
- Business Relevance
- Data Quality & Scope
- Analysis Potential
- Operational Metrics
- Decision Support:
- Learning Value

Business Value:
- Track sales performance
- Analyze customer behaviour
- Monitor operational efficiency
- Guide pricing strategies

TECH STACK

Visualization

Power BI,Matplotlib, Seaborn,Plotly

Data Processing

Pandas,Numpy,Scikit learn

Forecasting

Time Series Models (SARIMA, Prophet)
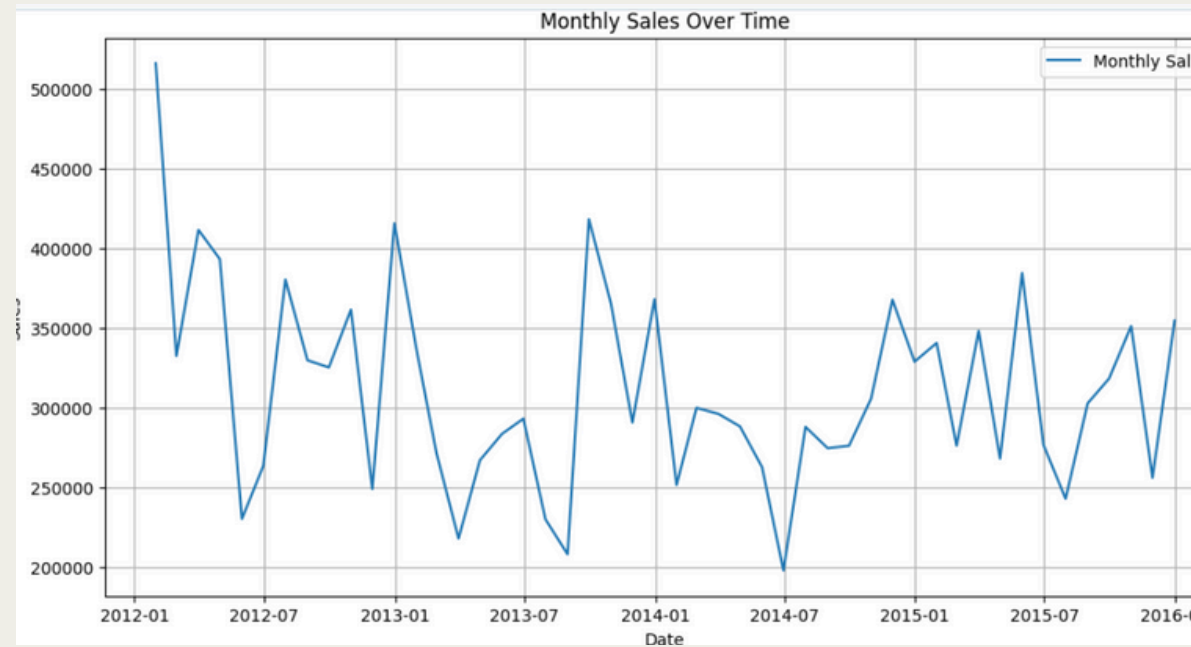
ML models

XGBoost,Random Forest

Model evaluation

Evaluate the performance of the selected models with appropriate metrics by using Scikit learn

Deployment

A Streamlit app for real-time predictions with an interactive interface.

EXPLORATORY DATA ANALYSIS (EDA)

Line Plot

- **The x-axis of the plot represents the order date and the y-axis representing the sales amount.**
- **The plot shows an overall downward trend in monthly sales over the period from 2012 to 2016.**
- **Higher sales generally occurring in the first half of each year and lower sales in the second half .**
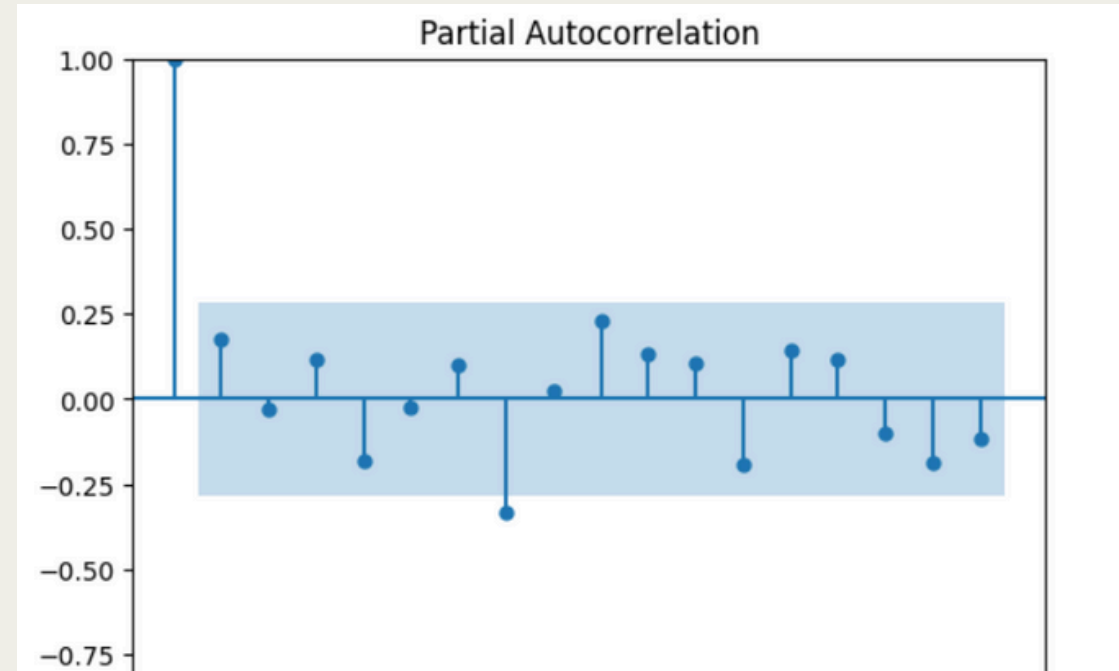
Partial autocorrelation plot

- **The x-axis of the plot represents the lag between the current value and the past value.**
- **The y-axis represents the partial autocorrelation coefficient.**
- **1 indicates perfect positive correlation, -1 indicates perfect negative correlation, and 0 indicates no correlation.**
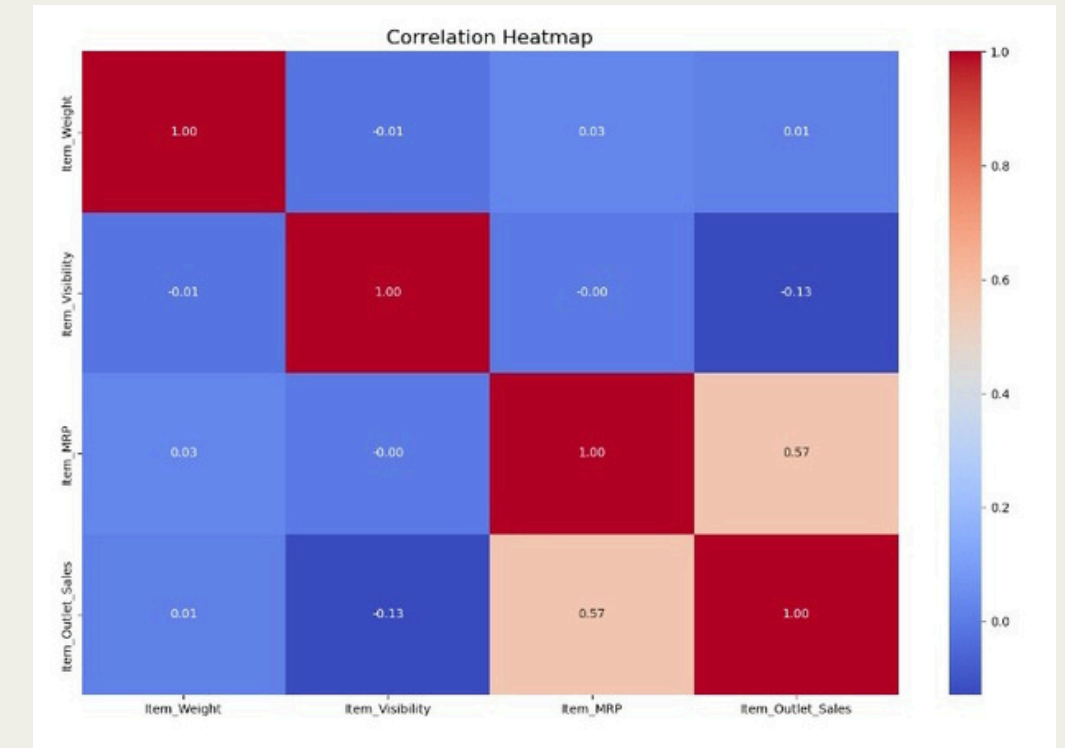
Correlation Heatmap

- **Item MRP and Item_Outlet_Sales have a high positive correlation.**
- **Item Weight and Item_Outlet_Sales have a negligible positive correlation.**
- **Item Visibility and Item_Outlet_Sales have a negative correlation.**

Sum of Sales by Seasonal


Sum of Unit Price by Product Category
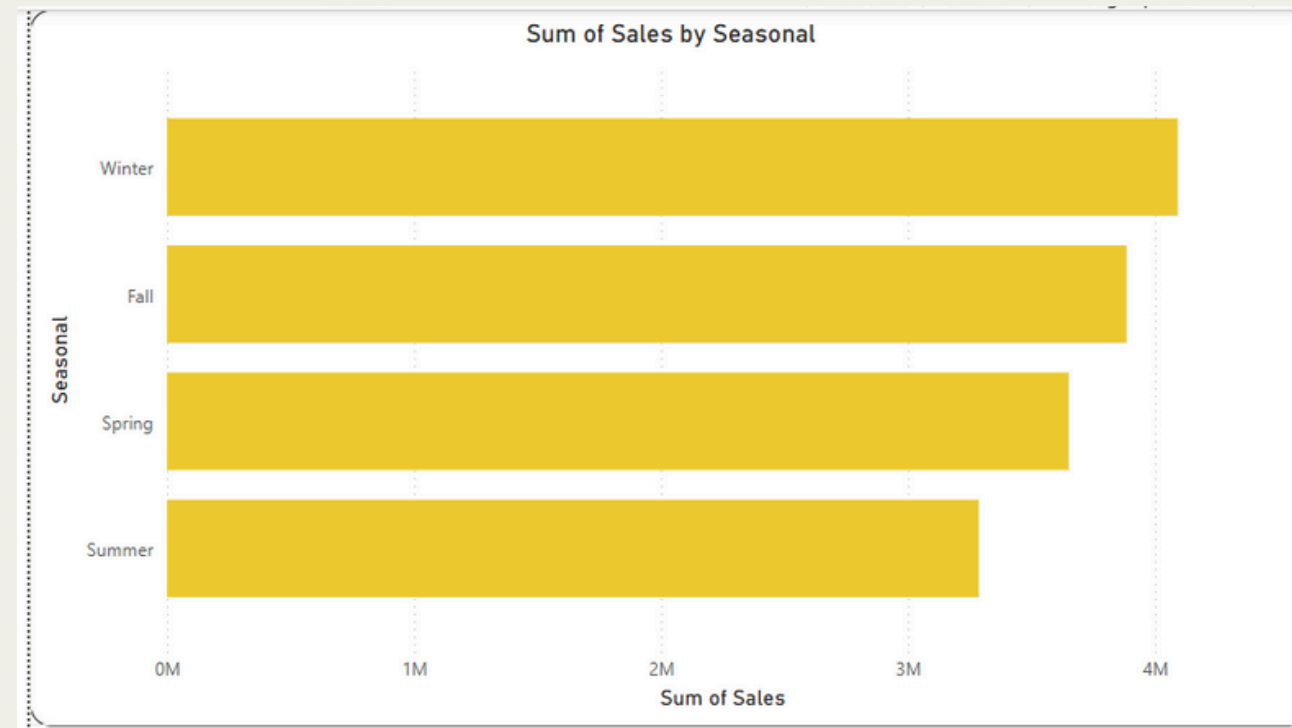
- **Winter has the highest sales:** Winter has the highest sales compared to other seasons, with sales of 40,92,497.12.
- **Summer has the lowest sales:** Summer has the lowest sales, at 32,86,978.20.
- **Winter sales are 24.51% higher than Summer sales:** Winter sales are significantly higher than Summer sales, with a difference of 24.51%.
- **Sales range across all seasons:** Sales range from 32,86,978.20 to 40,92,497.12 across all four seasons.
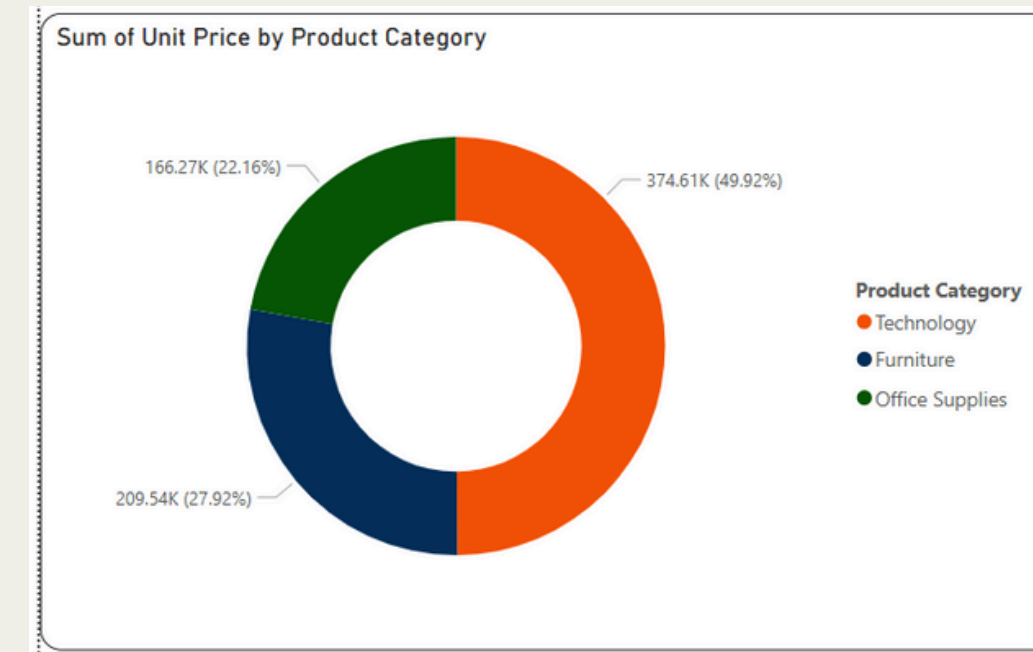
- **Technology dominates the market:** Technology has the highest sum of unit prices (49.92%) followed by furniture (27.92%) and office supplies (22.16%).
- **Technology has a significant contribution:** The sum of unit prices for Technology is almost double the sum of unit prices for furniture.
- **Office supplies have the lowest contribution:** The plot suggests office supplies have the lowest contribution to the total sum of unit prices.

MODELING
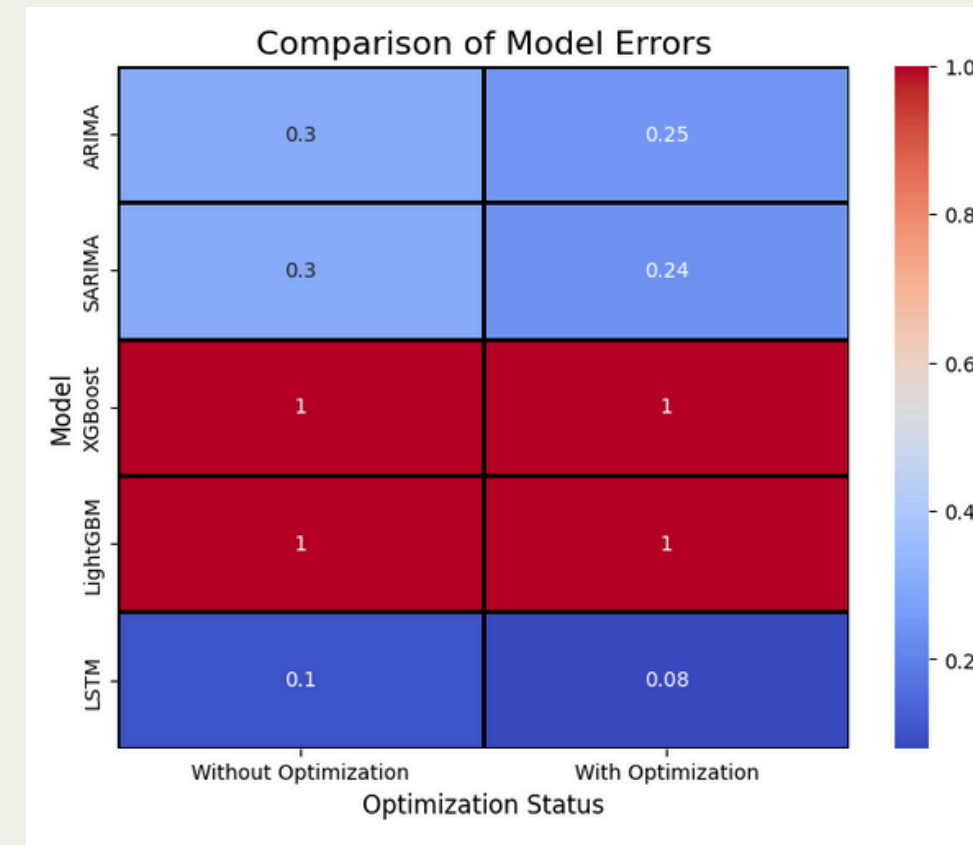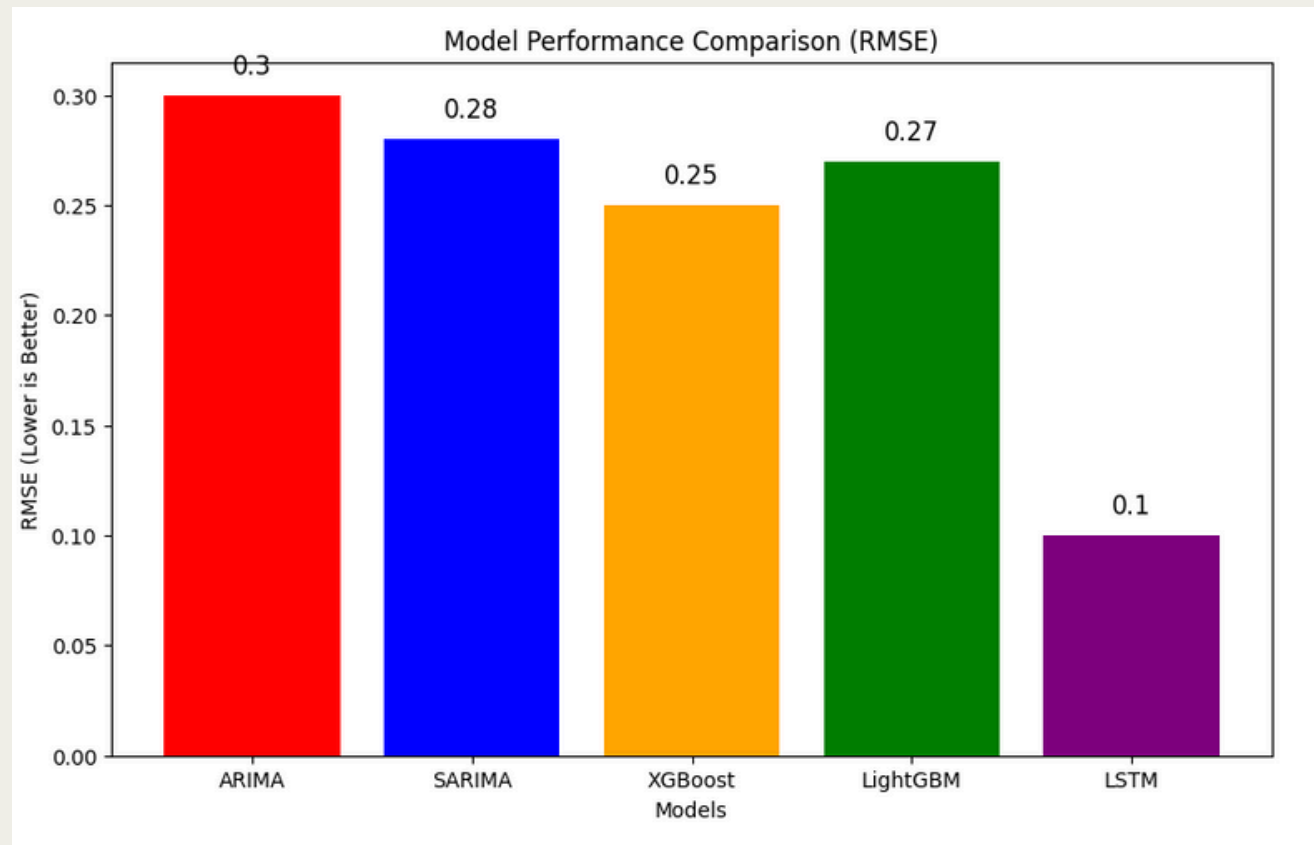
✔ LSTM shows the best improvement with optimization.
✔ SARIMA improves slightly, benefiting from seasonality handling.
✔ Boosting models (XGBoost & LightGBM) need evaluation, expected to handle large datasets efficiently.
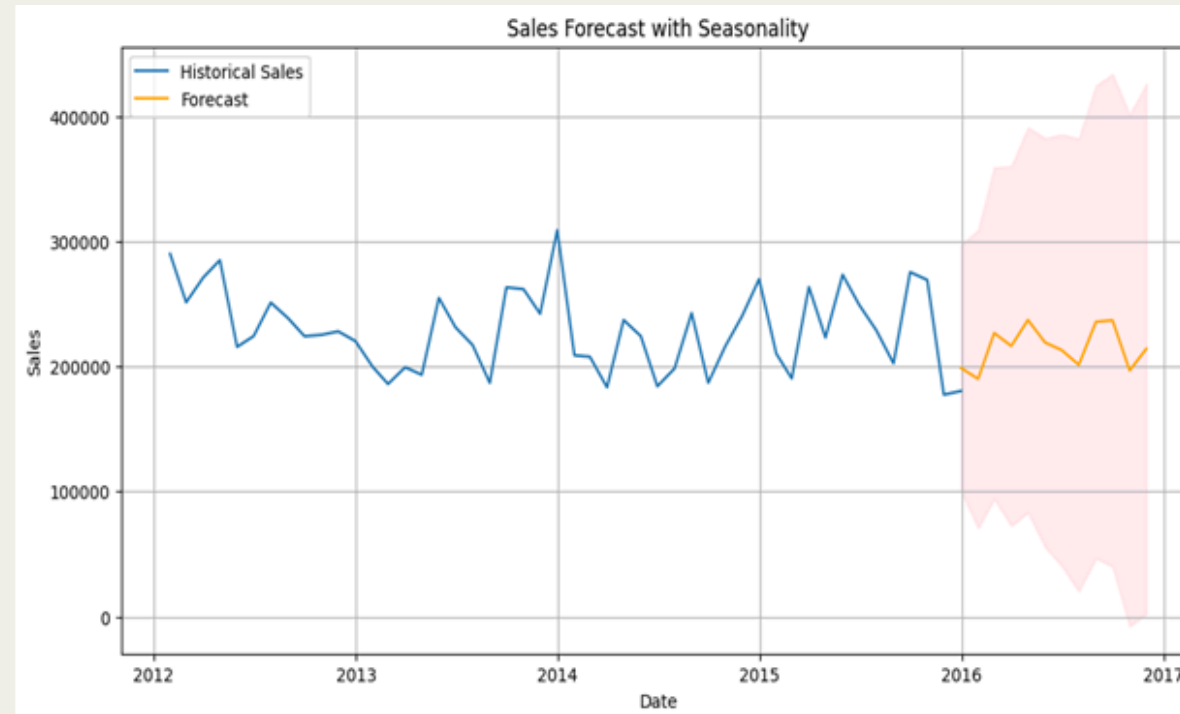✔ Traditional models (ARIMA, SARIMA) show moderate gains, but struggle with complex data patterns.

Time Series Forecasting Models
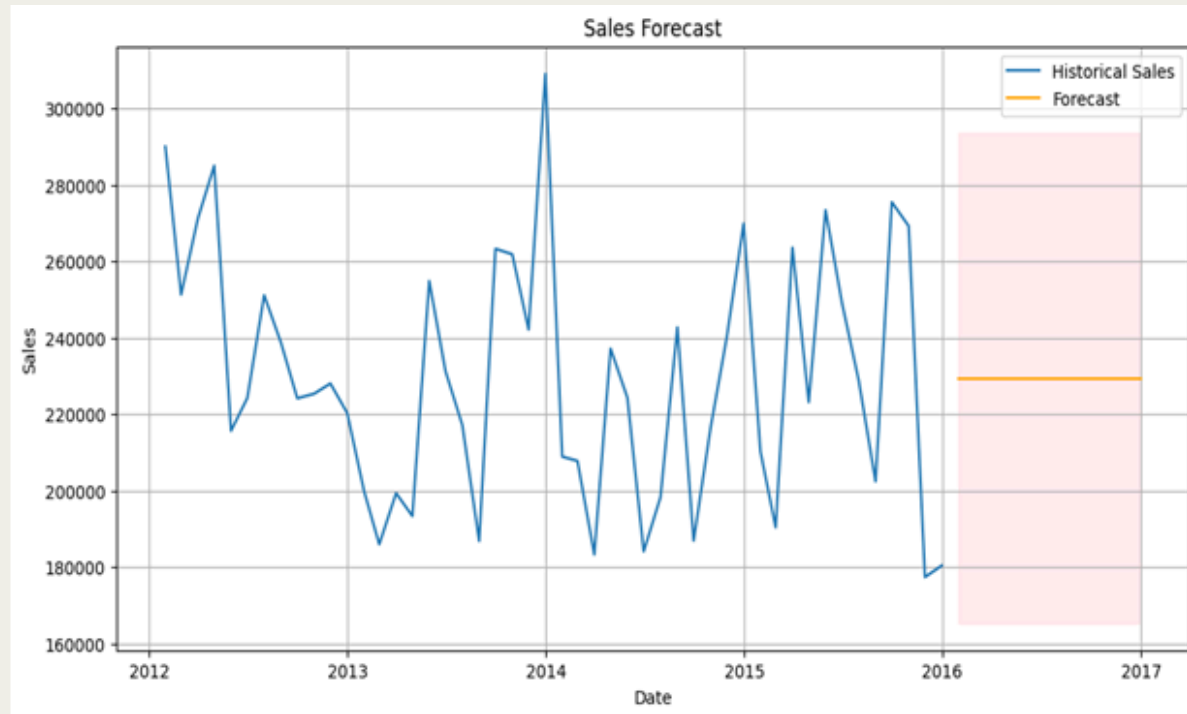
Auto Regressive Integrated Moving Average   (ARIMA)

Uses past values (**AR**), differencing (**I**), and
past errors (**MA**) for forecasting.
Assumes stationary data (requires differencing if not).

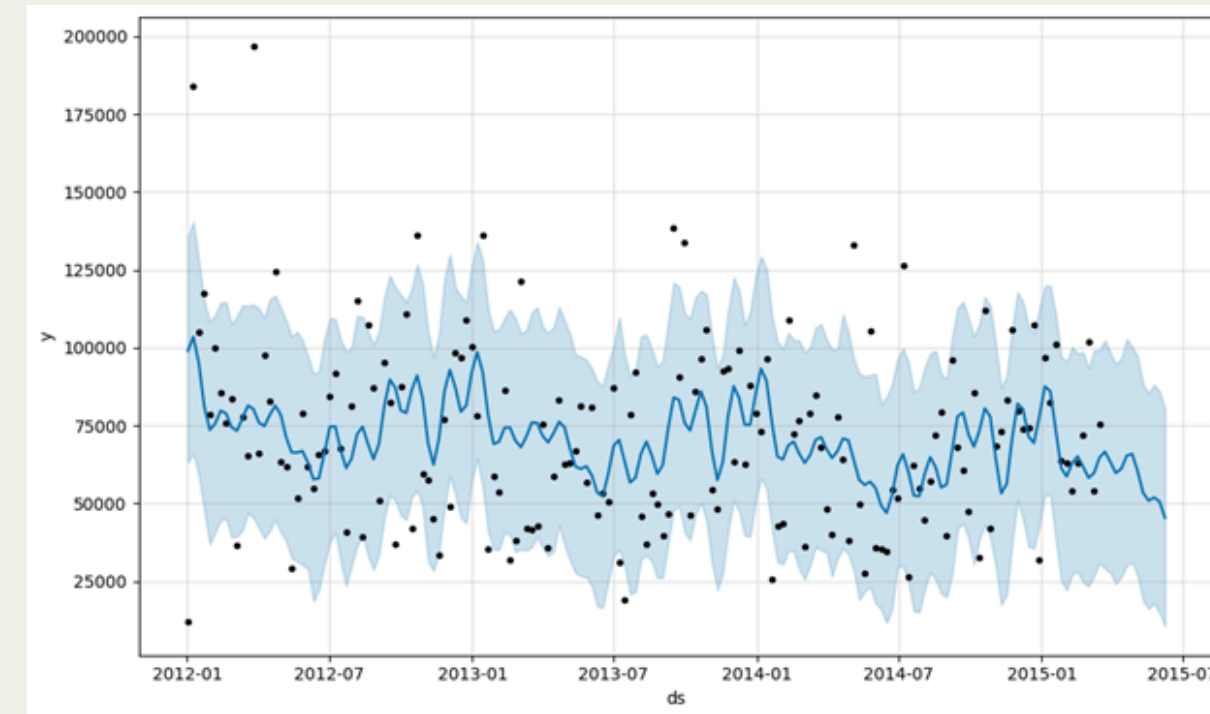**For non-seasonal, simple trends ARIMA is more appropriate.**

Seasonal ARIMA   (SARIMA)

Extends ARIMA by adding seasonal components (**SAR, SI, SMA**)
Parameter tuning can be complex.

For seasonal, fixed-pattern data SARIMA Is best.

Prophet (Facebook)

Prophet was developed by **Facebook's Core Data Science team**
Handles **daily, weekly, yearly** seasonality automatically.
Includes external factors like holidays.

For business & general use Prophet Model is best (easy & robust).

User Interface for Shipping Delay Prediction

Input Features

Order Priority

Low

Ship Mode

Delivery Truck

Region

West

Shipping Cost

100.00

Order Quantity

2

Discount

0.50

0.00   1.00

Profit

-100.00

Predict Shipping Status

📦 **Shipping Delay Prediction &** 📈 **Sales Forecasting**
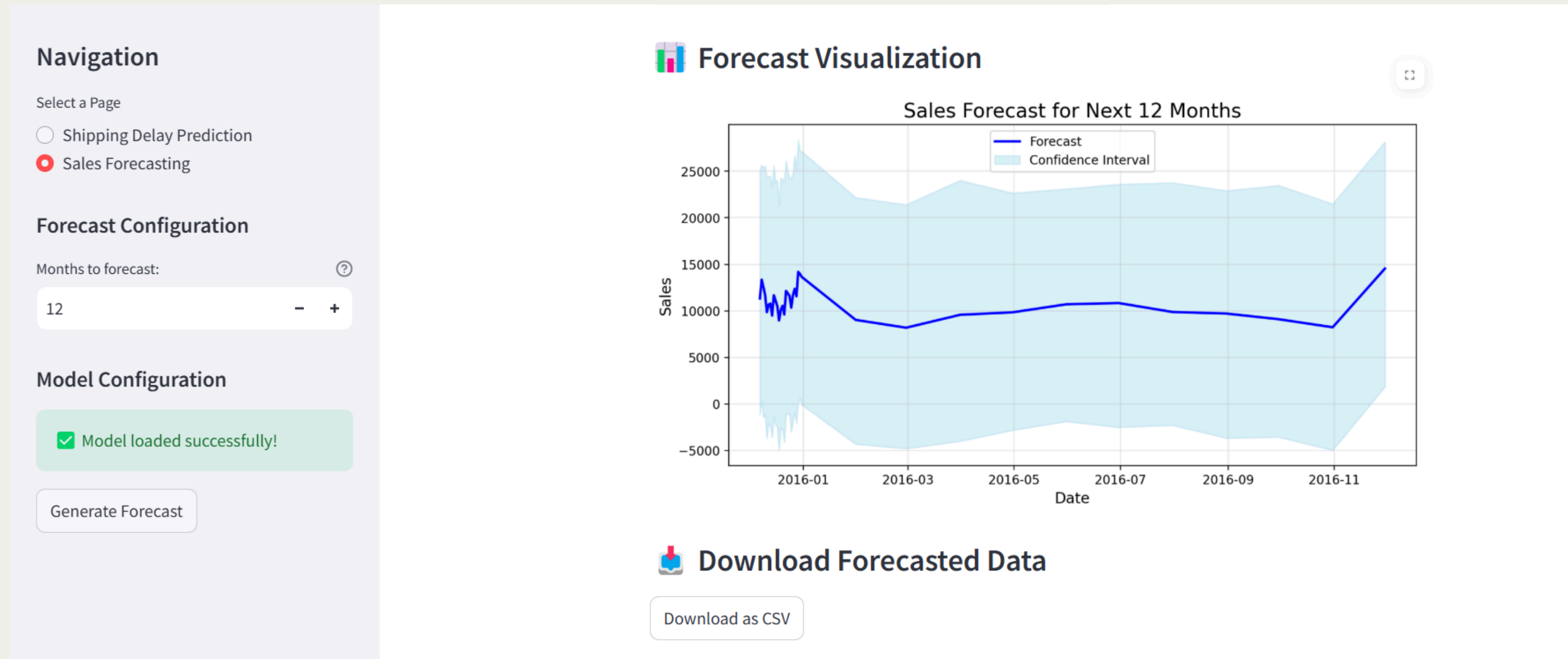
Shipping Delay Prediction

Prediction

Predicted Shipping Status: Delayed

Created an interactive Streamlit interface that allows users to input shipping parameters and receive real-time predictions on potential delays.

User Interface for Sales forecasting

Developed a user-friendly interface with Streamlit, allowing users to input data, generate real-time sales forecasts, and explore trends through interactive visualizations.

Results

- Forecasted sales trends using SARIMA to anticipate demand.
- Predicted shipping delays with machine learning for better logistics.
- Identified customer segments for personalized marketing.
- Created interactive visualizations in Power BI and Streamlit.
- Provided insights to optimize inventory and improve efficiency.

Future Scope

Enhance models with deep learning, enable real-time predictions, automate decisions, analyze broader data, and scale to other retailers.

Thank you!