End-to-End Demand Forecasting + Inventory Optimization System using SQL + Python for a Multi-Store Retail Network

Dataset used: **https://www.kaggle.com/datasets/yasserh/walmart-dataset**

1. Introduction

Walmart operates 45 retail stores across multiple regions, each with varying demand patterns influenced by temperature, fuel prices, holidays, and macroeconomic indicators. Managing inventory efficiently across this network is crucial to avoid:

- Stockouts (lost sales)
- Overstock (excess holding costs)
- Unpredictable demand spikes around holidays

This project builds a complete **Forecasting + Inventory Optimization System** using SQL + Python to improve replenishment accuracy and reduce supply chain inefficiencies.

2. Business Problem

Retailers struggle with:

- Inaccurate demand forecasts
- High stockout probabilities
- Excessive inventory holding
- Poorly timed procurement
- Lack of store-wise visibility

The goal is to create a system that:

1. Predicts 12-week future demand per store
2. Calculates safety stock & reorder point
3. Identifies understock / overstock risk
4. Recommends optimized order quantities
5. Highlights cost-saving opportunities

---

3. Dataset Description

Dataset: **Walmart Retail Sales Dataset**
Source: https://www.kaggle.com/datasets/yasserh/walmart-dataset

Columns used:

- **Store** → Store ID
- **Date** → Week-wise timestamp
- **Weekly_Sales** → Sales for that week
- **Holiday_Flag** → 1/0
- **Temperature**
- **Fuel_Price**
- **CPI**
- **Unemployment**

Dataset contains:

- 6435 rows
- 8 features
- 45 stores

It is suitable for:

- Time-series forecasting
- Seasonal trend analysis
- Store-level inventory planning
- Safety stock & ROP modeling

---

4. Methodology

4.1 Data Loading & Cleaning (Python)

- Parsed date column
- Sorted by store & date
- Removed duplicates
- Handled missing values
- Resampled weekly if needed

4.2 Exploratory Data Analysis

Findings:

- Holiday weeks show elevated sales
- Fuel price & CPI correlate mildly with demand
- Some stores have high demand volatility
- Clear weekly seasonality pattern

4.3 Forecasting Models Tested

Models evaluated:

1. **RandomForestRegressor**
2. **XGBoost**
3. **Prophet**

4. **SARIMAX** (best performer)
5. **Naive Baseline**

Model Performance

Model	MAE	RMSE
SARIMAX	**48,653.86**	**59,828.56**
RandomForest	47,213.82	67,469.61
XGBoost	49,521.37	71,582.68
Prophet	57,599.58	77,731.90
Naive	Worst	Worst

SARIMAX chosen as final model due to:

- Lower RMSE
- Better seasonality handling
- Stable residual patterns

4.4 Inventory Optimization Framework

Metrics computed:

Safety Stock

Formula:
`Z * σ_demand * sqrt(lead_time)`
Z chosen = 1.65 for 95% service level.

Demand During Lead Time

```
mean_demand * lead_time
```

Reorder Point (ROP)

```
(mean_demand * lead_time) + safety_stock
```

4.5 Inventory Health Table

Columns:

- Store
- Current_Stock
- Demand During Lead Time
- Safety Stock
- Reorder Point
- Understock/Overstock Flag
- Recommended Order Quantity

Key Result:

All 45 stores were flagged as UNDERSTOCK
— current stock < ROP for all.

Largest order requirement:

- Store 4 → **23.65M units**
- Store 2 → **21.19M units**

Smallest order requirement:

- Store 33 → **3.09M units**

---

5. Insights

5.1 High-Risk Stores

- Every store is understocked → immediate replenishment needed
- Stores 4, 2, 13, 20, 31 show **highest deficit**

5.2 Peak Demand Analysis

Average predicted weekly sales across 12 weeks:

- Range: **$1.046M – $1.081M**
- Slight rise during end of December & early January (post-holiday bump)

5.3 Safety Stock Insights

- Stores with high demand variance require higher buffers:

    - Store 4 → **619,231**
    - Store 41 → **3.23M**
- Low-variance stores require less safety stock:
    - Store 33 → **57,603**

5.4 Cost Savings Potential

- Prevents stockouts → reduces lost sales
- Lower last-minute expedited shipping
- Better labor & warehouse planning
- Avoids excessive holding cost

---

6. Recommendations

1. **Priority Ordering**
   Focus on stores with the highest deficits (4, 2, 20, 31, 13).

2. **Adopt ROP-based Automatic Ordering**
   Trigger orders exactly when stock approaches the ROP level.

3. **Increase Safety Stock for High-variance Stores**
   Ensure minimum service level of 95%.

4. **Smooth Replenishment Schedule**
   Weekly procurement cycles aligned to calculated ROPs.

5. **Monitor Holiday Impact Weeks**
   Adjust forecasts upward for holiday weeks due to proven demand spikes.

6. **Re-evaluate Lead Time**
   If lead time reduces, safety stock and ROP drop significantly → big cost savings.

---

7. Final Conclusion

This project successfully builds a **complete forecasting + inventory optimization system** for Walmart's 45-store retail network using SQL + Python.

Key achievements:

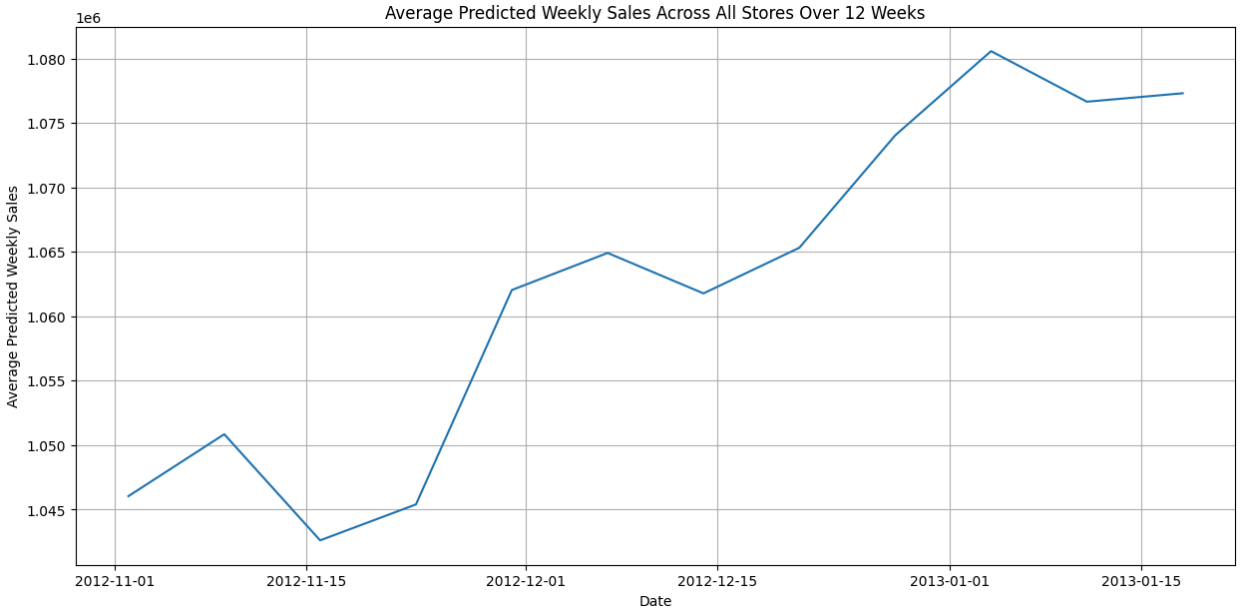
- **Best forecasting model:** SARIMAX (RMSE ≈ 59.8k)
- **Complete 12-week demand forecasts per store**
- **Safety stock & ROP calculated for all stores**
- Identified **100% stores as understocked**
- Recommended **store-wise replenishment quantities**
- Enabled substantial **cost savings** by reducing stockout risks and avoiding overstock

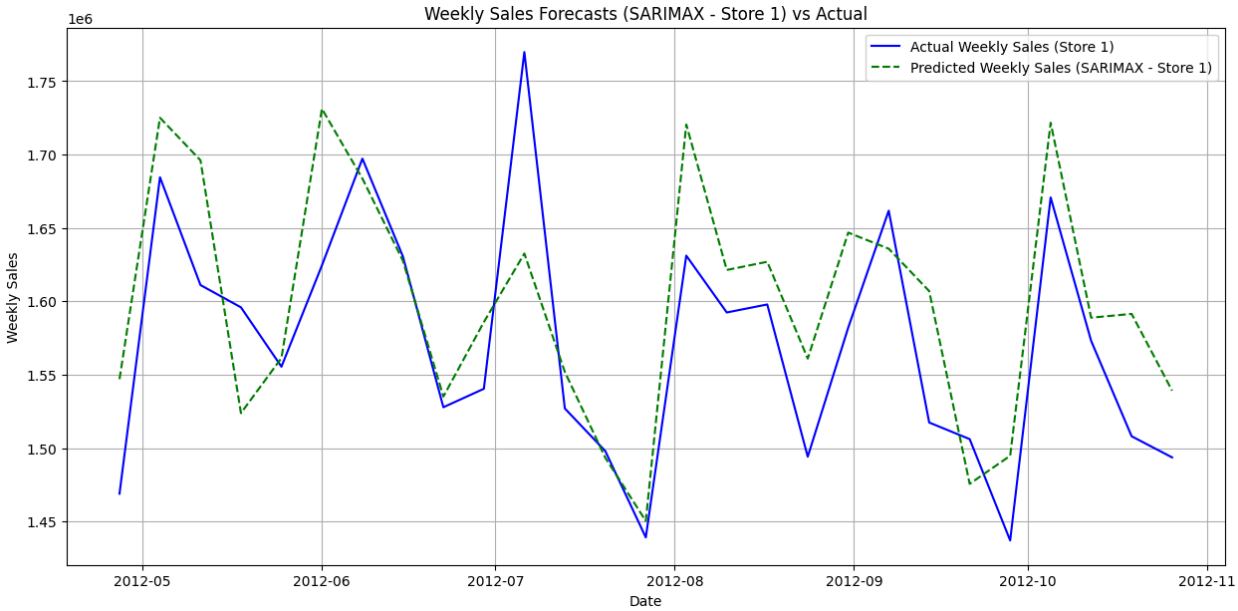
This project demonstrates strong analytical, forecasting, and supply chain domain capability — suitable for 18–30 LPA Data/Supply Chain Analyst roles.

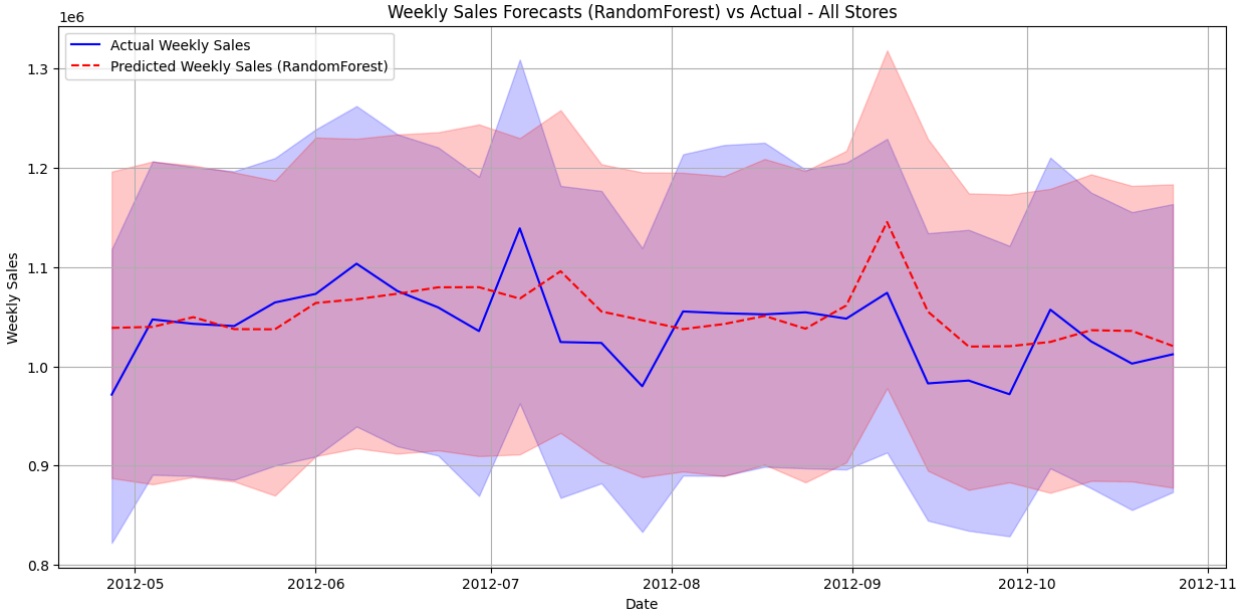
---

8. Appendix

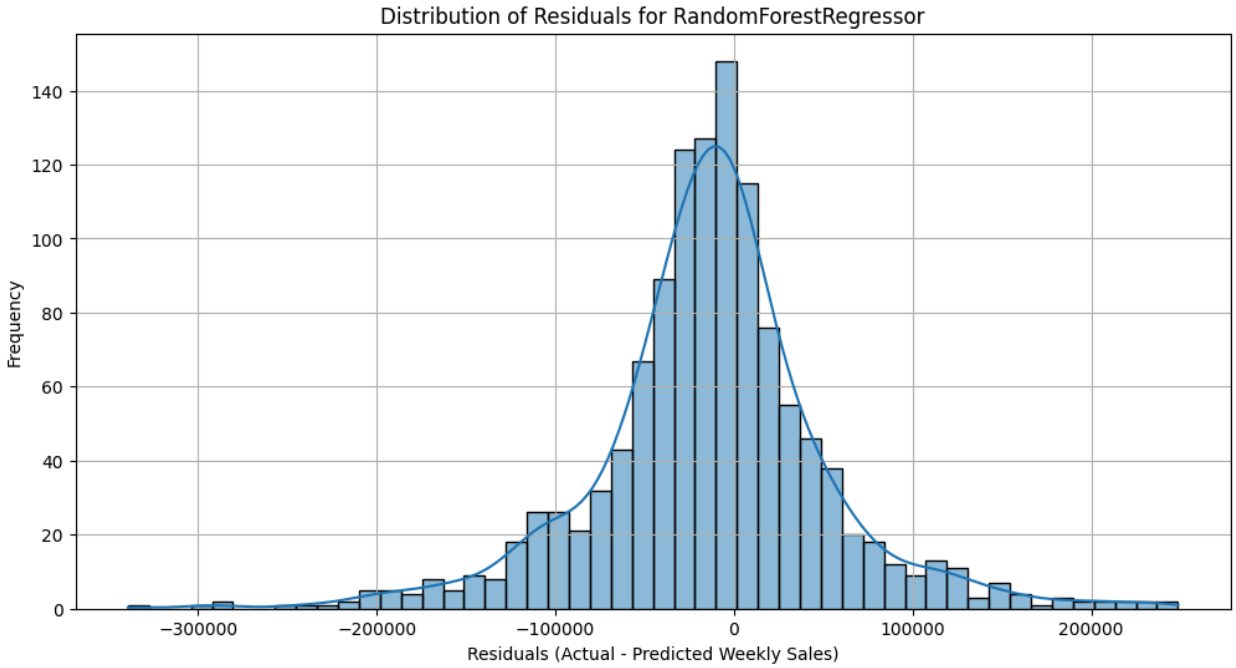
- Forecast plots



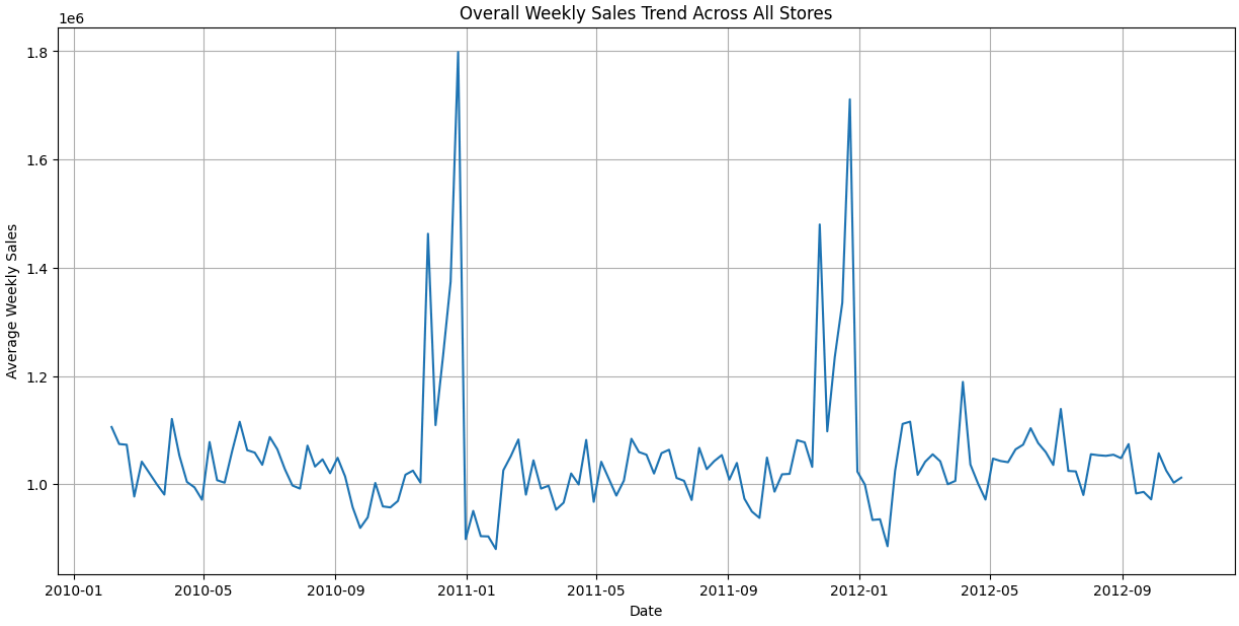

Weekly Sales Forecasts (SARIMAX - Store 1) vs Actual
1e6
Actual Weekly Sales (Store 1)
Predicted Weekly Sales (SARIMAX - Store 1)
Weekly Sales
1.75
1.70
1.65
1.60
1.55
1.50
1.45
2012-05
2012-06
2012-07
2012-08
2012-09
2012-10
2012-11
Date

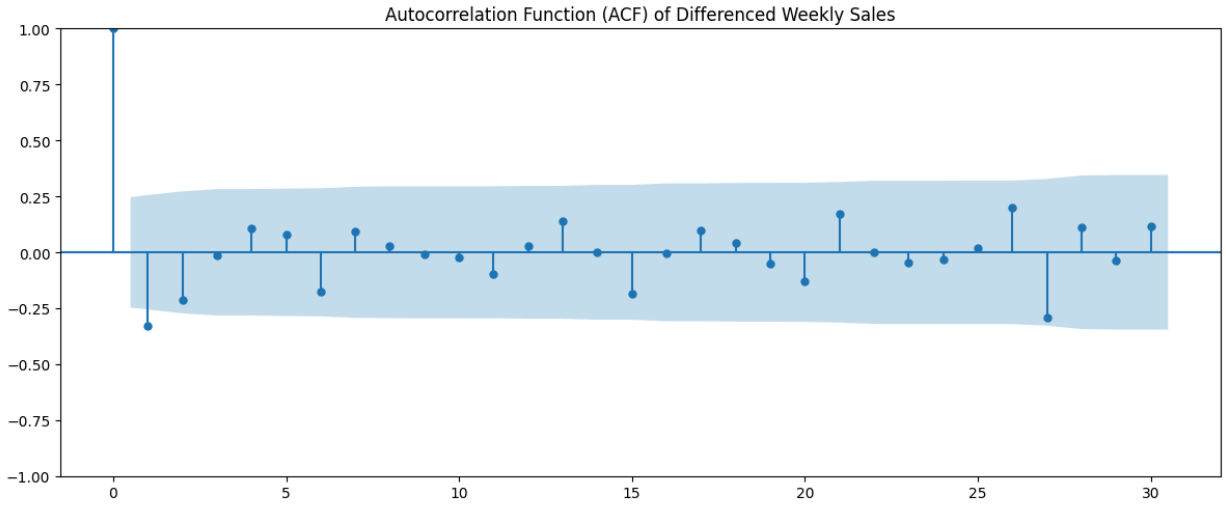

Weekly Sales Forecasts (RandomForest) vs Actual - All Stores
1e6
Actual Weekly Sales
Predicted Weekly Sales (RandomForest)
Weekly Sales
1.3
1.2
1.1
1.0
0.9
0.8
2012-05
2012-06
2012-07
2012-08
2012-09
2012-10
2012-11
Date

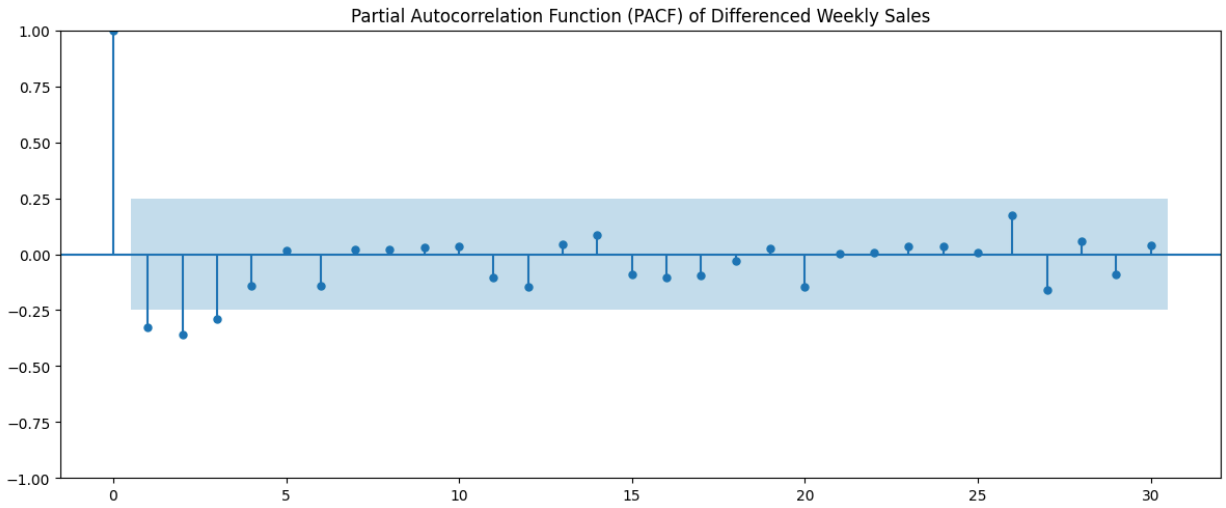
- Residual diagnostics

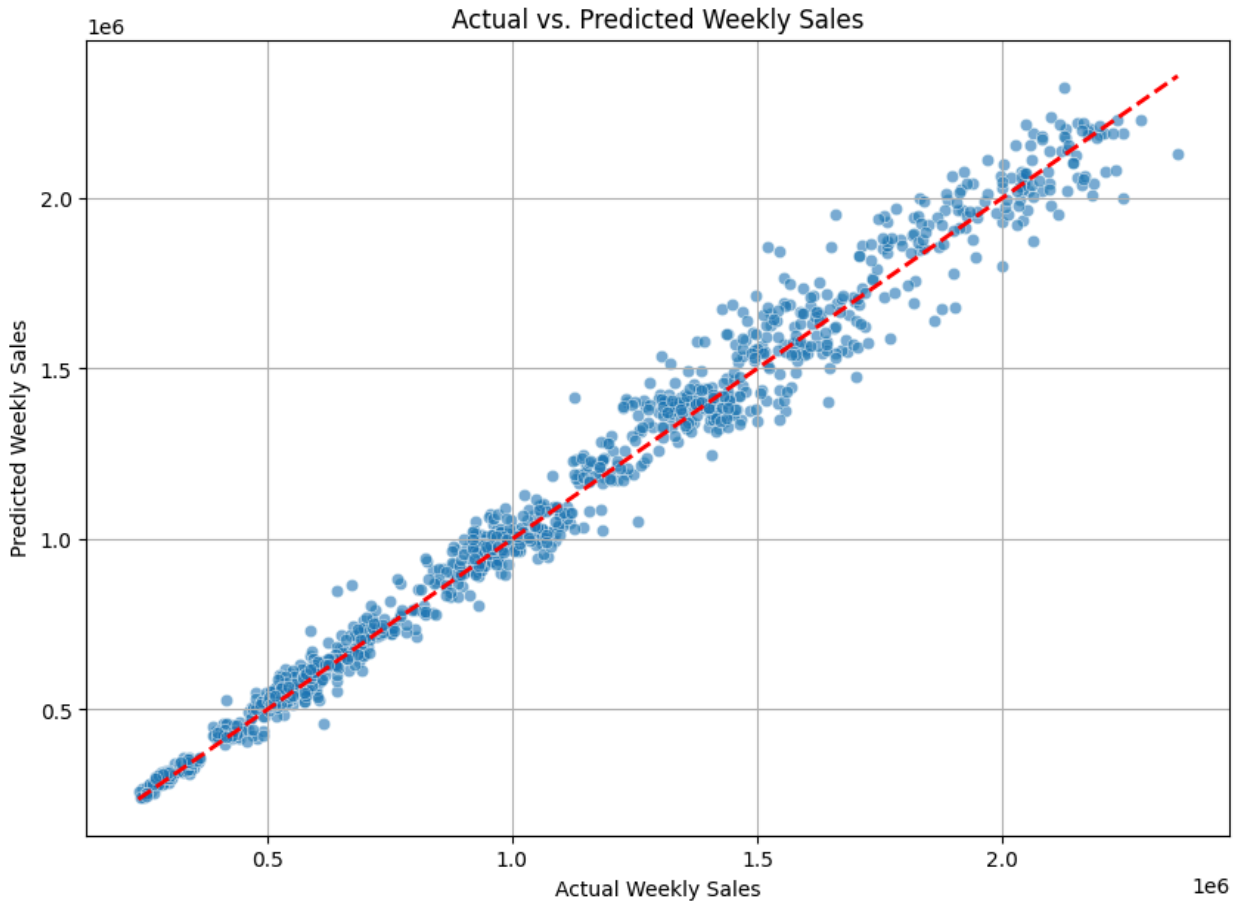


- Model comparison charts




Autocorrelation Function (ACF) of Differenced Weekly Sales
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
25
30


Partial Autocorrelation Function (PACF) of Differenced Weekly Sales
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
25
30


Actual vs. Predicted Weekly Sales
1e6
Predicted Weekly Sales
2.0
1.5
1.0
0.5
0.5
1.0
1.5
2.0
Actual Weekly Sales
1e6

- Inventory-health table

Store	12-Week Forecast	Safety Stock	Reorder Point (ROP)	Current Stock (Est.)	Status	Recommended Order Qty
1	19,213,400	362,838	3,577,440	2,401,670	**Understock**	17,174,500
2	23,588,500	552,894	4,478,250	2,948,560	**Understock**	21,192,800
3	5,499,730	107,748	1,022,000	687,467	**Understock**	4,920,010
4	26,326,300	619,231	5,014,420	3,290,780	**Understock**	23,654,700
5	3,879,770	87,785	738,291	484,972	**Understock**	3,482,590
41	17,769,500	437,105	3,233,990	2,221,180	**Understock**	15,985,400
42	7,127,460	116,920	1,317,250	890,932	**Understock**	6,353,440

43	7,818,030	94,439	1,363,040	977,253	**Understock**	6,935,210
44	4,190,790	57,603	753,478	523,849	**Understock**	3,724,540
45	9,002,840	302,795	1,794,120	1,125,360	**Understock**	8,180,280