

7-Day Sales Forecasting for Superstore

...

Optimizing Inventory Through Time Series Modeling
Maria Suarez

The Problem

- Superstore often runs out of items or over-orders.
- Goal: Forecast daily sales for the next **7 days** to reduce stockouts and overstocking.



Why it matters:

- Support data-driven supply chain strategy
- Reduce stockouts & overstocking
- Optimize inventory decisions
- Enable proactive regional response to demand shifts

Who might care?

- Regional Manager
- Inventory & Supply Chain Teams
- Finance & Strategy Teams
- Merchandising & Sales Teams
- Data is & Analytics Team

Data Overview

Source

Kaggle Superstore Dataset

<https://www.kaggle.com/datasets/rohithsahoo/sales-for-ecasting>

Time Range

- 2015-2019
- Granularity:Daily

Focus:

Sales per Category

- Office Supplies
- Technology
- Furniture

Processing Steps

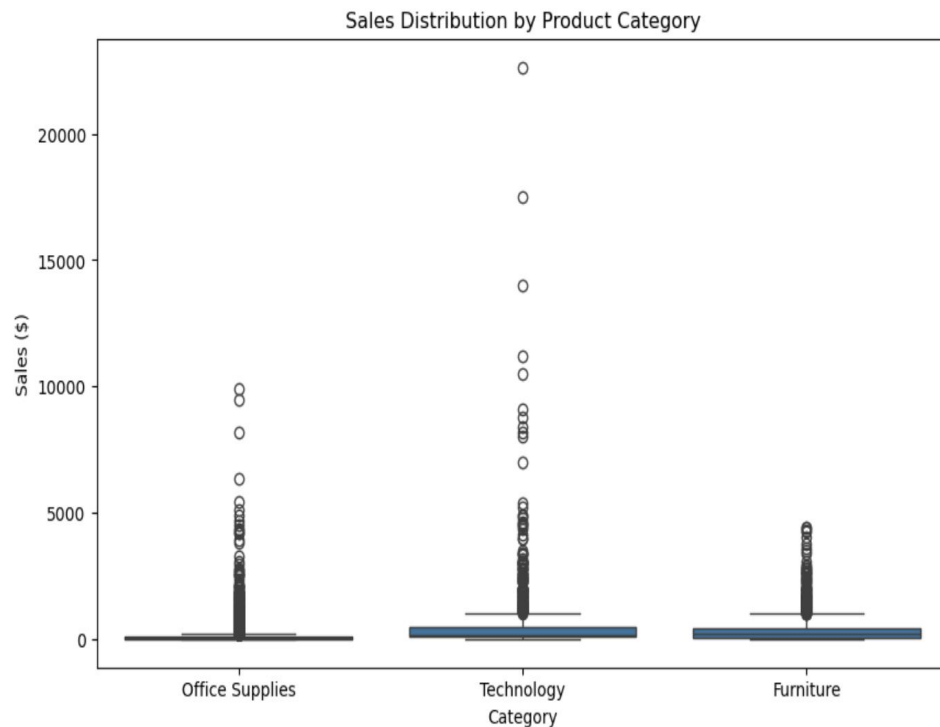
	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	S
0	CA-2017-152156	2017-11-08	2017-11-11	Second Class	CG-12520	C
1	CA-2017-152156	2017-11-08	2017-11-11	Second Class	CG-12520	C
2	CA-2017-138688	2017-06-12	2017-06-16	Second Class	DV-13045	C
3	US-2016-108966	2016-10-11	2016-10-18	Standard Class	SO-20335	C
4	US-2016-108966	2016-10-11	2016-10-18	Standard Class	SO-20335	C

Parsed and set 'Order Date' as datetime index

Feature Engineering

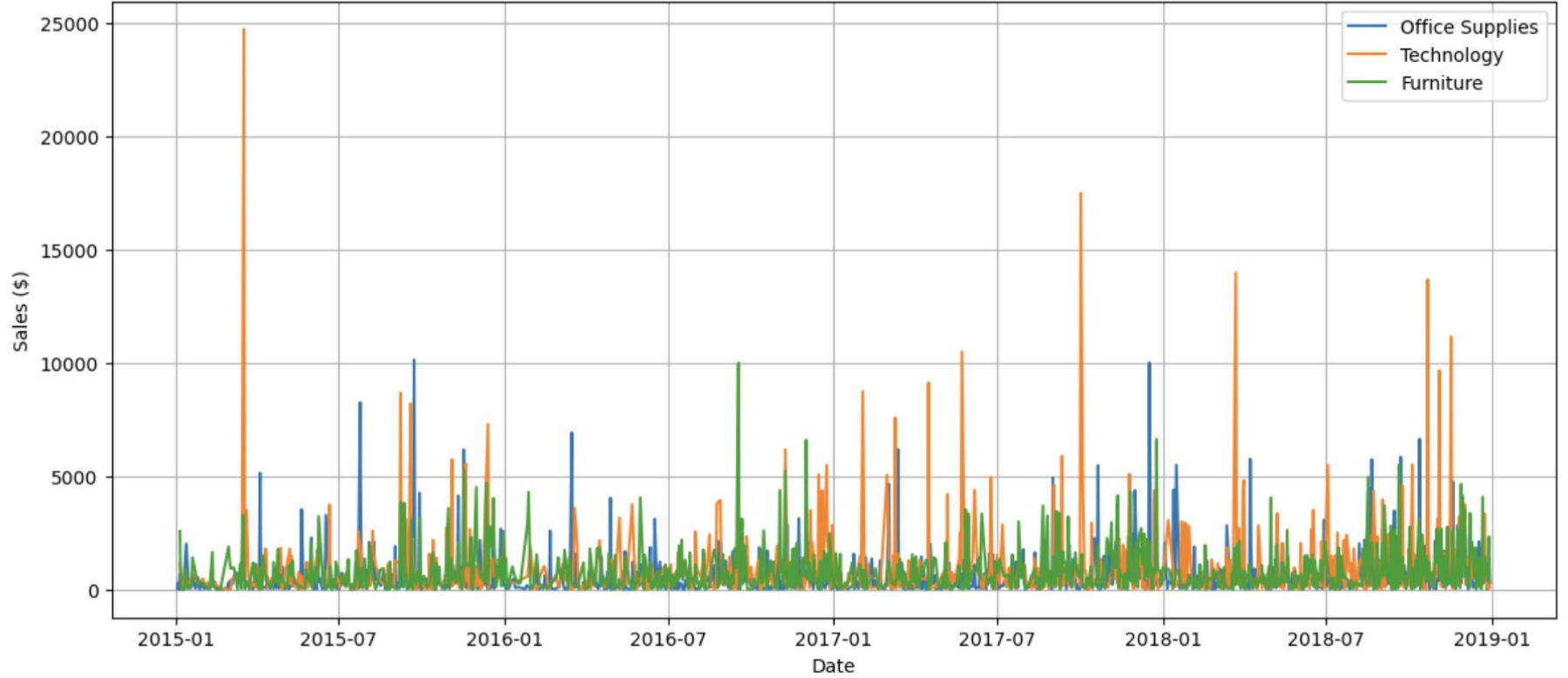
- Extract day, month year from Order Date
- Aggregate sales per day or region for time series modeling
- Filtering for top product categories

Exploratory Analysis



- **Office Supplies:** Stable daily sales with few outliers
- **Technology:** High volatility; sales spikes from big-ticket items (e.g., Copiers)
- **Furniture:** Moderate volume; weekly patterns observed

Sales by Category Over Time



Modeling Approach:

Model: Seasonal ARIMA (SARIMA)

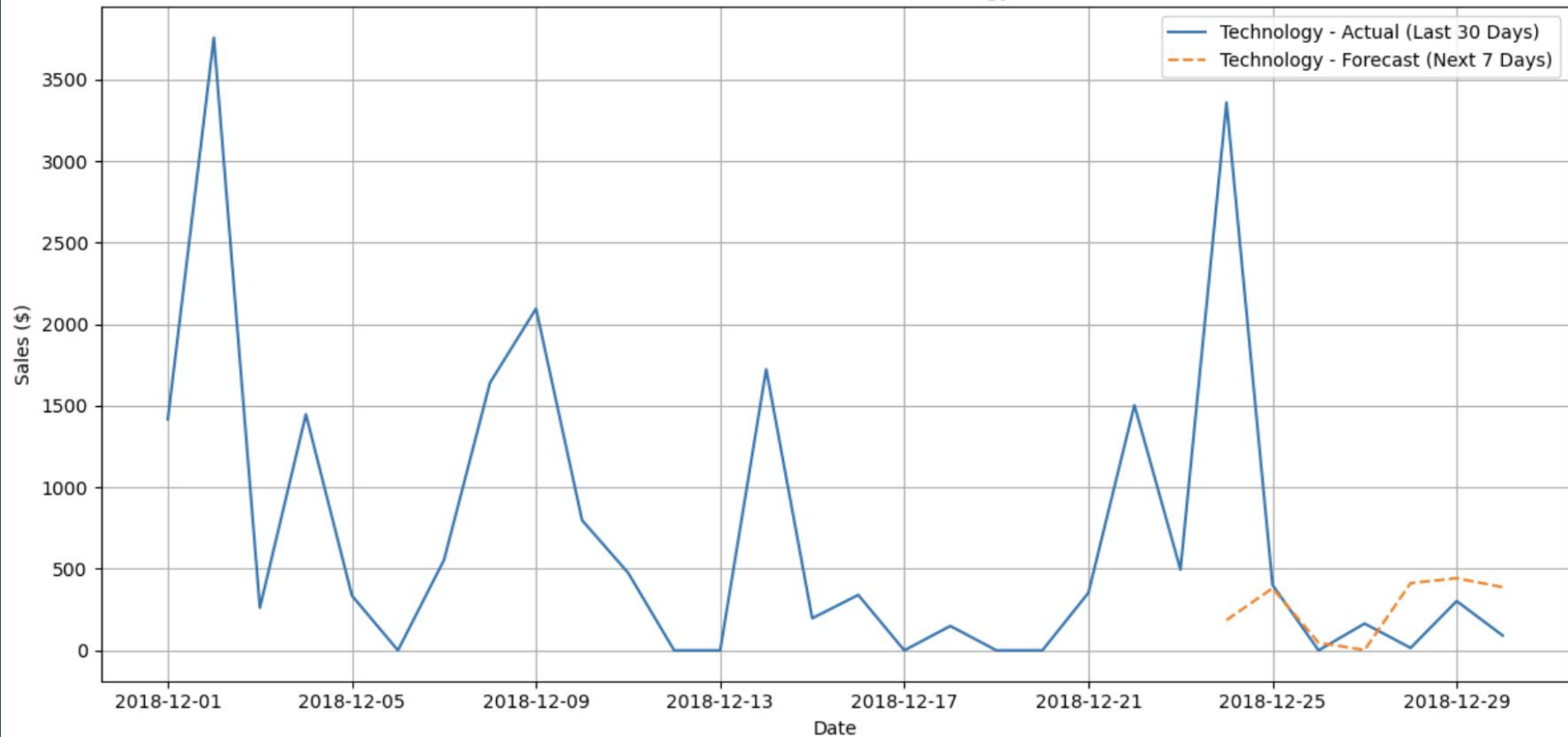
Parameters: $(1,1,1)(1,1,1,7)$

Applied separately for each category

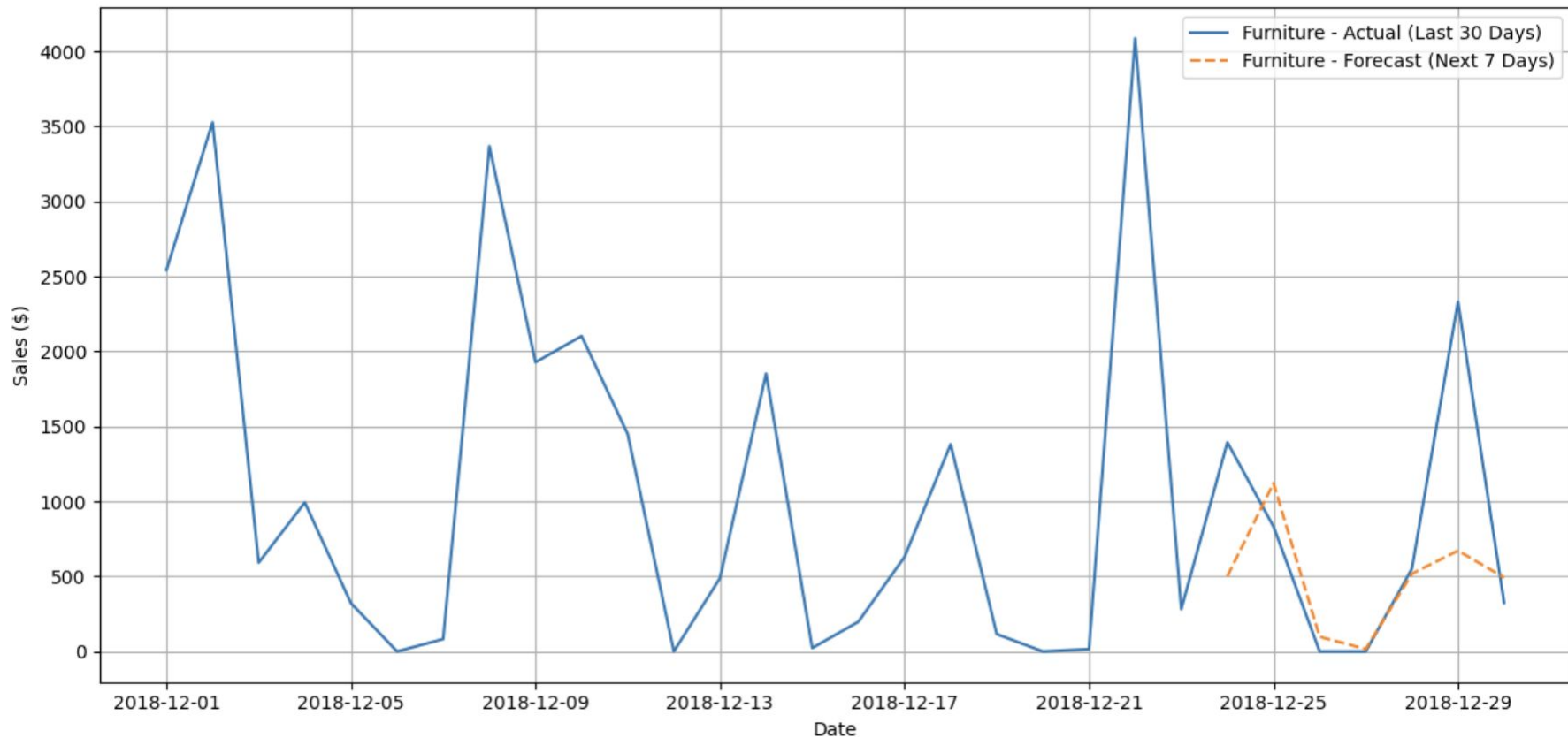
Used log transformation to stabilize
variance

Forecasted sales for next 7 calendar days

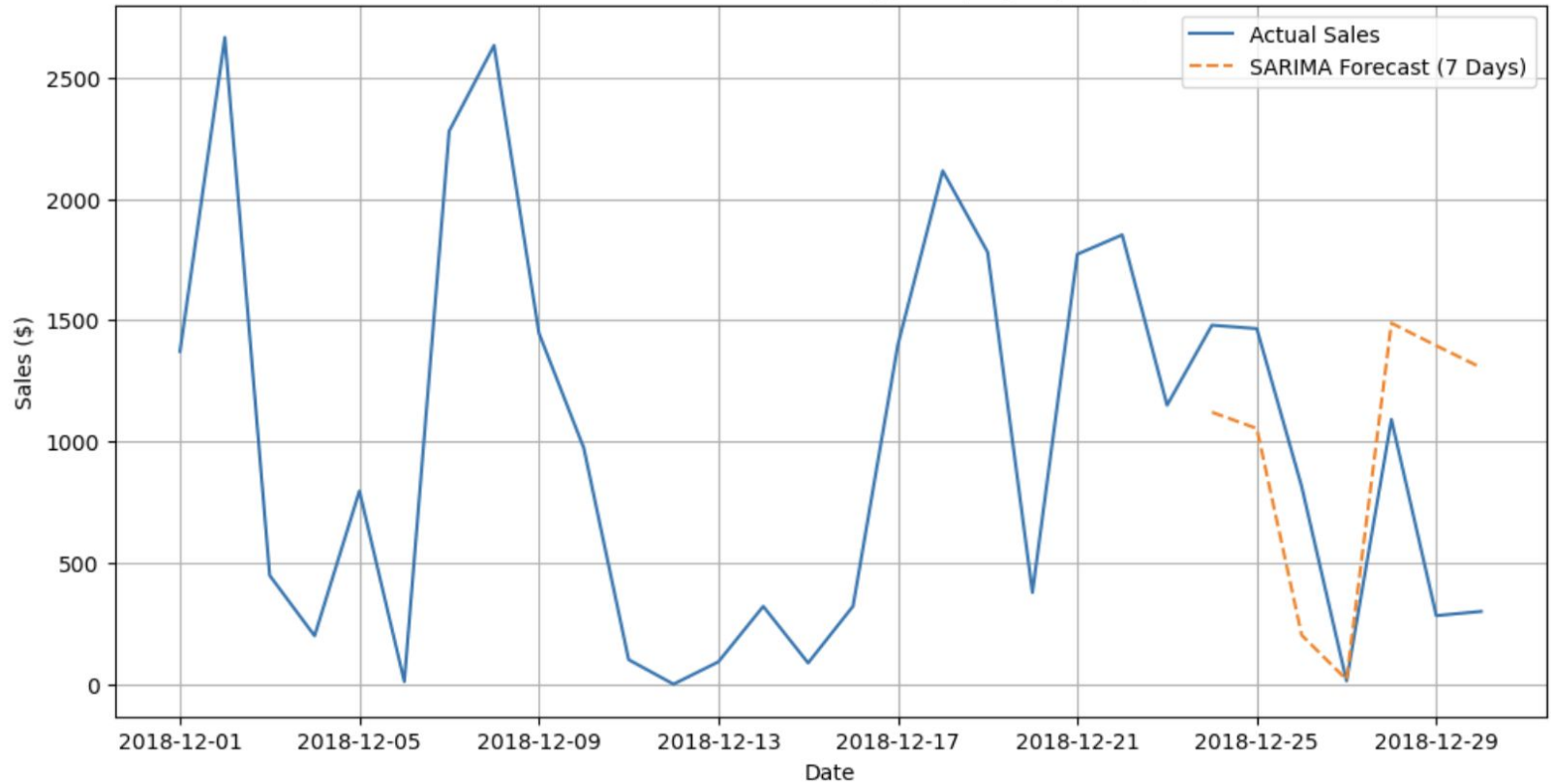
SARIMA Forecast for Technology



SARIMA Forecast for Furniture



SARIMA Forecast - Office Supplies (Log-Adjusted)



Model Candidates

Model

Description

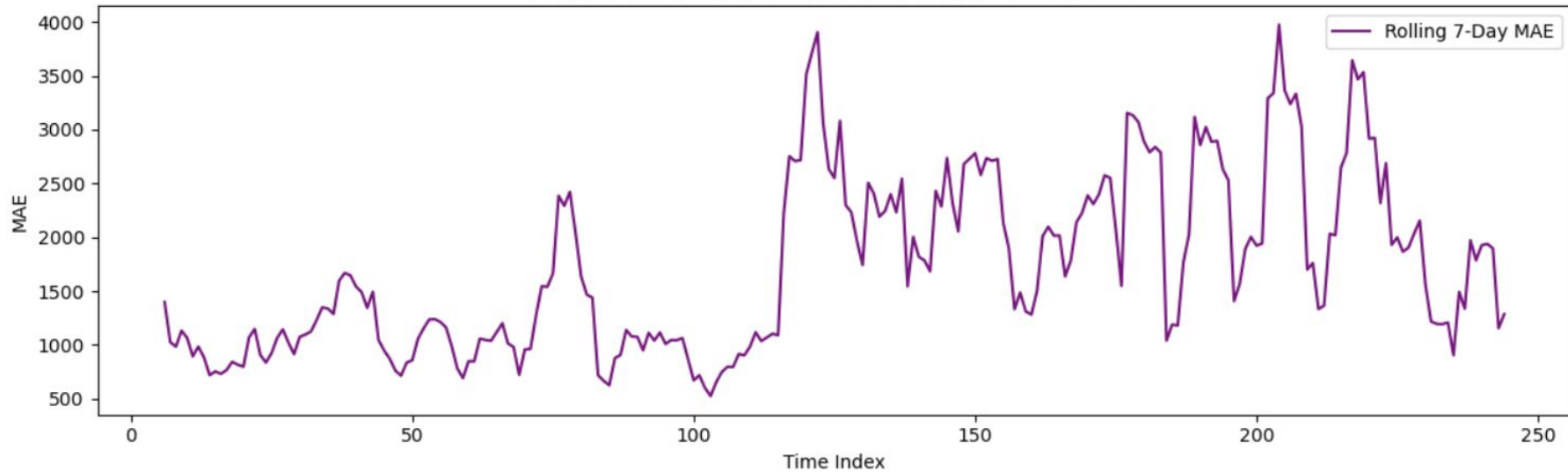
Lag-Based Linear Regression

Uses past 1-day and 7-day sales to predict the next day (simple + interpretable)

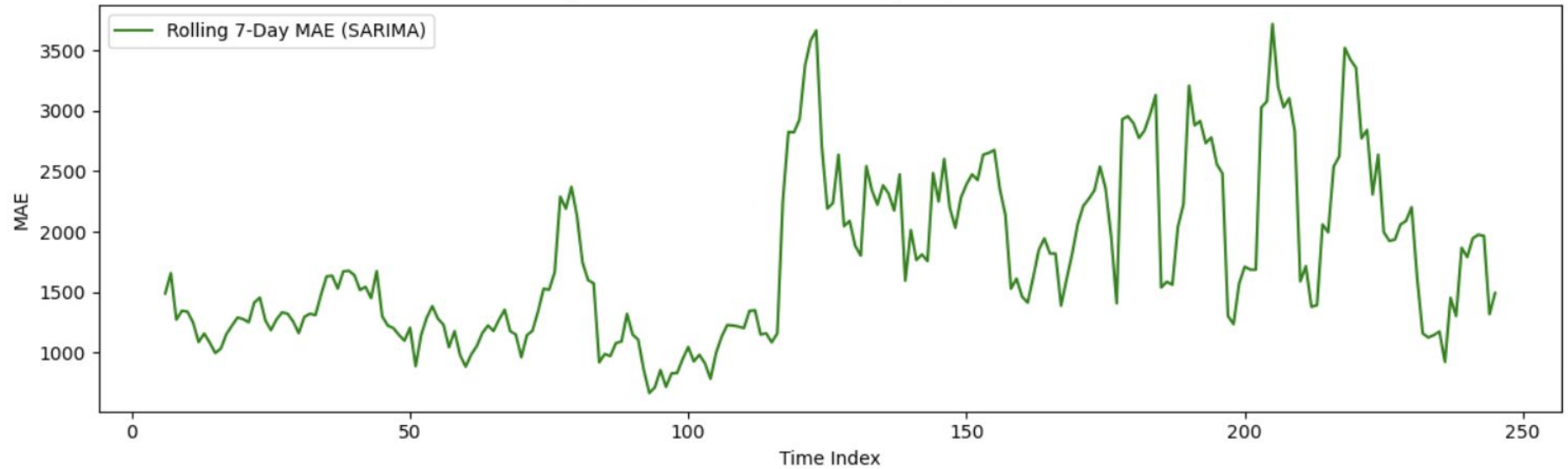
SARIMA

Captures both trend and weekly seasonality using statistical time series modeling

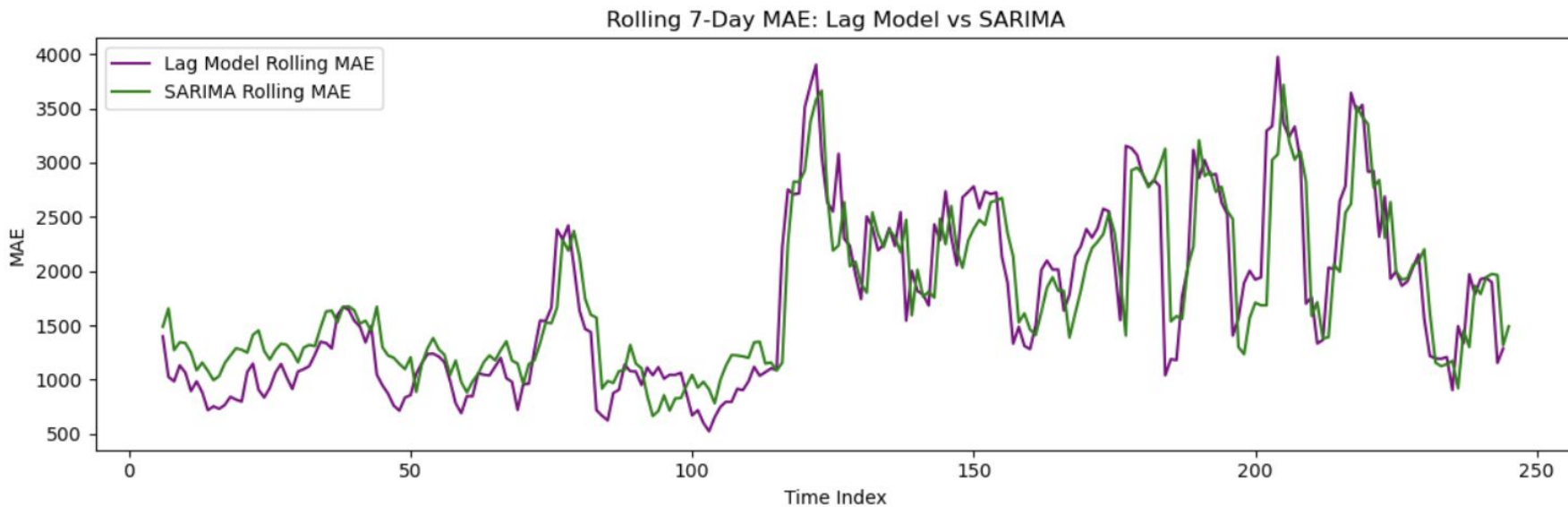
Rolling 7-Day Mean Absolute Error (MAE) for Lag Model



Rolling 7-Day Mean Absolute Error (MAE) for SARIMA Model



Model	MAE	RMSE	Strengths
Lag Regression	1,716.42	2,509.28	Simple, accurate for daily planning
SARIMA	1,761.55	2,452.65	Better at handling sales spikes



Final Model Justification


Despite SARIMA's strong performance on RMSE and its ability to capture weekly patterns:

- We selected the **Lag-Based Regression model** as the **final forecasting solution** due to:
 - **Lowest MAE** — key for routine inventory decisions
 - **Simplicity and interpretability**
 - **Ease of daily updates** with minimal compute or retraining


3 Simple Recommendations

- Run weekly forecasts
- Order based on predictions
- Train teams to act on trends


Future Scope




Add more helpful data like promotions, holidays, and weather to improve predictions



Make forecasts more detailed by going down to the product level instead of just categories



Build a dashboard so store managers can easily see trends and forecasts in real time



Set up an automated system that keeps updating the model with new sales data every day

Final Conclusions

1. **Lag-Based Regression was the best overall model**
It had the lowest average error (MAE), making it reliable for daily sales forecasting.
2. **SARIMA performed well during sales spikes**
It captured weekly patterns but was more complex and less practical for fast updates.
3. **Daily sales in Superstore show short-term trends**
Using recent sales (lags) is an effective way to forecast the next 7 days.
4. **Rolling MAE revealed stability differences**
The Lag model had more consistent performance over time, while SARIMA fluctuated.
5. **The chosen model supports smarter inventory decisions**
It helps reduce stockouts and overstocking, enabling managers to act proactively.