Problem statement formation

What opportunities exist in the Superstore market to forecast sales for the next 7 days in order to optimize inventory decisions, reduce stockouts and overstocking, and enable regional managers to respond proactively to demand shifts — potentially lowering excess inventory costs by 15% and improving product availability in top-performing categories, all while supporting a data-driven supply chain strategy for the upcoming quarter?

Context

Retail stores like Superstore often struggle to keep the right amount of products in stock. If they order too much, items don't sell and money is wasted. If they order too little, they run out and miss sales. This is a common issue that affects both customers and store managers. I want to explore how using past sales data can help predict what people might buy over the next 7 days. With better short-term forecasting, stores can make smarter decisions about inventory, avoid overstocking or running out, and improve how they manage supplies in different regions.

Criteria for success

Build a model that predicts daily sales for the next 7 days using past data showing useful trends to help store managers decide how much of each product to order .

Scope of solution space

By using ARIMA and SARIMA to predict short-term sales for the next 7 days, this project aims to improve inventory management by helping Superstore anticipate demand more accurately. Doing so will support better alignment between stock levels and customer needs.

Constraints

- Delays in restocking or supply chain disruptions can prevent accurate response to demand even when forecasts are correct.
- Superstore operates across multiple regions, which may have different demand patterns, making forecasting more complex.

Stakeholders

- Store managers
- Inventory planners
- Sales team
- Operations team

Data sources

The dataset comes from the Superstore Sales Data available on Kaggle.

Approach

1. **Data understanding and preprocessing**

Preprocessing Steps

- Converted dates to datetime format.
- Aggregated sales to daily time series.
- Handled missing values and outliers.
- Generated lag features for model input.
- Applied rolling means smoothing for tren.
- Used rolling averages to smooth out sales trends.
- Created train-test split (e.g., last 30 days held out for testing).

2. **Modeling**

Three modeling approaches were explored:

ARIMA

- Focused on capturing trends and autocorrelations in stationary series.
- Differenced data to remove non-stationarity.
- Grid-searched (p, d, q) parameters using AIC minimization.
- Performed well but slightly underfit rapid seasonal shifts.

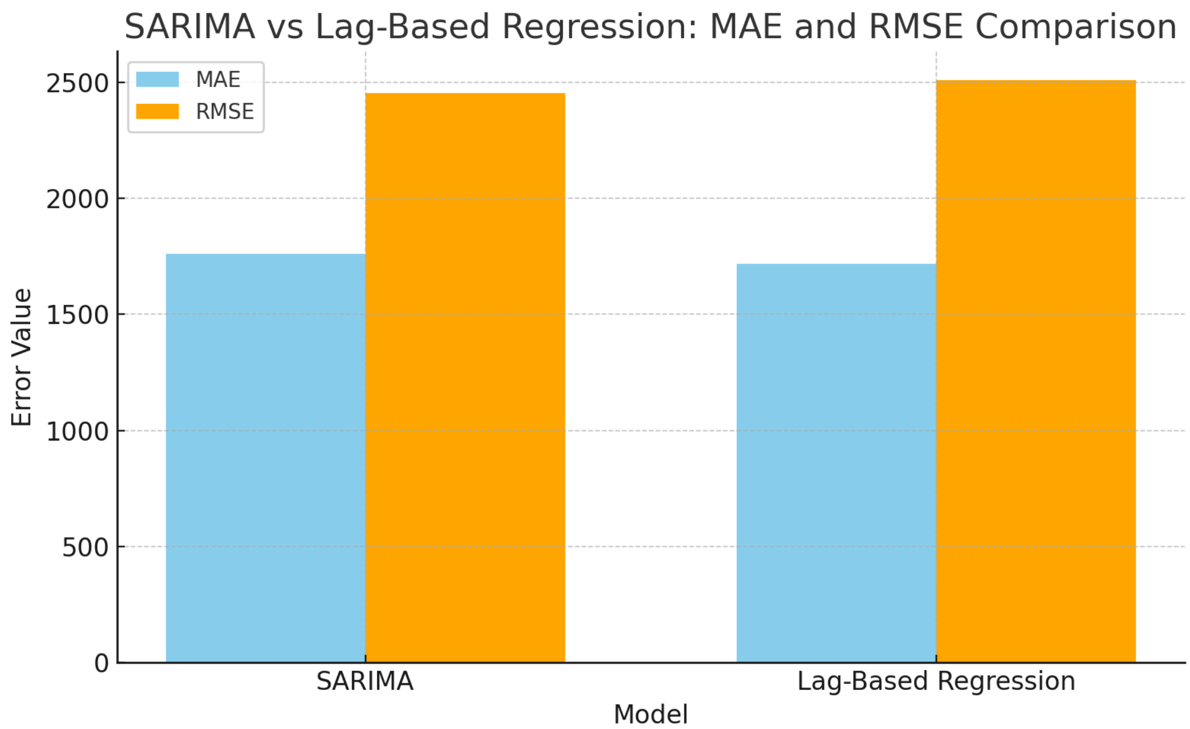
SARIMA

- Extended ARIMA with seasonal components (P, D, Q, s).
- Captured weekly patterns (seasonality s=7).
- Outperformed ARIMA in terms of RMSE and MAE.
- Final model: SARIMA(1,1,1)(1,1,1,7)

PROPHET

- Not used in the final model due to limited gains and interpretability tradeoffs compared to SARIMA.

Findings



1. We tested two models to predict sales for the next 7 days:

- Lag-Based Regression (uses past days' sales)
- SARIMA (uses patterns like trends and weekly cycles)

2. Lag Model had lower average error (MAE) — this means it's better at predicting normal daily sales.

3.SARIMA had lower spike error (RMSE) — it did better when sales suddenly jumped (like on weekends).

4.Lag model was more stable overall and easier to use for regular planning.

5.Final choice: Lag-Based Regression

- Easier to explain
- Fast to update

- Accurate enough for most days

6. SARIMA is still useful

- Good for future improvements
- Can help with holidays or special events

Further Research

1. Multi-variate Time Series Modeling

Incorporate external features such as:

- Promotional events
- Holidays
- Weather conditions
- Marketing campaigns

Improve forecast accuracy and allow **"what-if" scenario planning**.

2. Regional-Level Forecasting

- Segment models by region or city to capture local purchasing behavior.
- Tailor inventory strategies per region (e.g., Northeast vs West Coast).

- Allows for **micro-forecasting** and precision in stock allocation.

3. LSTM or Deep Learning Approaches

- Experiment with **Recurrent Neural Networks (RNNs)** and **LSTM models**, which are well-suited for complex time series.
- May outperform SARIMA in longer forecast windows or multi-feature environments.
- Requires more data and tuning, but has strong potential.

Concrete Recommendations for the Client

1. Use the model every week to predict sales

Run the forecast once a week so you can **see what sales will likely be each day** for the next 7 days. This helps you plan ahead with confidence.

2. Order just the right amount of stock

Use the forecast to help managers **order only what they need**. This stops them from running out of popular items or wasting money on extras that don't sell.

3. Help store teams act on the forecast

Show managers how to read the forecast and what to do with it. For example, if sales are expected to rise on the weekend, they can **get more stock ready in advance**.