

Executive Summary

This project extends the Superstore Sales Analysis (2019–2022) by developing a predictive sales forecasting model aimed at anticipating monthly revenue trends and supporting strategic decision-making.

The analysis leverages historical sales data from the Superstore dataset to understand seasonality, growth patterns, and business performance. Using time-series forecasting techniques, the model provides actionable insights that can help optimize inventory management, marketing strategies, and revenue planning.

Objectives

Forecast monthly sales for the upcoming 12 months  
Capture seasonality and trend patterns in sales data  
Compare traditional statistical and machine learning forecasting methods  
Evaluate model accuracy and provide business recommendations

Methods Used

1. Data Preparation:  
Cleaned and standardized historical data  
Aggregated daily transactions into monthly totals  
Created new time-based features (year, month, lag variables)
2. Forecasting Techniques:  
SARIMAX Model: Captured trend and seasonality using statistical time-series modeling  
Random Forest Regressor: Leveraged machine learning to predict sales using lag and rolling features  
Baseline Models: Naïve and seasonal-naïve forecasts for comparison
3. Model Evaluation:  
Metrics used: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)  
Models tested and validated using the last 12 months of data

Key Findings

Consistent upward trend in sales over the years, with notable seasonal peaks during Q4 (holiday months). SARIMAX captured seasonality effectively and provided stable short-term forecasts. Random Forest handled non-linear patterns better in recent months, offering slightly improved predictive accuracy. Baseline models performed reasonably well but lacked predictive precision beyond repeating past values.

Results Summary

Model	MAE (↓)	RMSE (↓)	Key Strength
Naïve Forecast	Moderate	High	Simple benchmark

Seasonal Naïve	Lower	Moderate	Good seasonality reflection
SARIMAX	Low	Low	Best at capturing trend + seasonality
Random Forest	Low	Low	Captures non-linear patterns

## Business Insights

Sales are seasonally cyclical, peaking around November–December due to holiday demand. Q2 and Q3 show steady growth, suggesting potential for targeted mid-year promotions. Forecasting models enable better inventory and staffing optimization ahead of peak demand. Combining SARIMAX and ML models can yield robust hybrid forecasts for business planning.

## Recommendations

1. Increase inventory and marketing budgets during Q4 to align with seasonal sales surges.
2. Use forecast outputs for demand planning and revenue budgeting.
3. Continue retraining models quarterly with new data for more accurate predictions.
4. Explore additional factors (discounts, region, category) to build multi-variate forecasting models.

## Conclusion

The Superstore Sales Forecasting project demonstrates how predictive analytics can transform historical sales data into forward-looking business intelligence. By combining statistical models (SARIMAX) and machine learning techniques (Random Forest), this analysis provides a reliable foundation for data-driven planning, revenue forecasting, and strategic decision-making.