Forecasting Retail Sales Revenue: A Time Series Approach using ARIMA

Introduction

Daily retail sales fluctuate due to consumer behavior and temporal patterns.

Accurate forecasts help with inventory, supply chain, and financial planning.

This study uses ARIMA to model daily sales trends and make short-term forecasts.

Talking point: Mention you transform the data to stationary before modeling and evaluate accuracy with MAE/RMSE.

OBJECTIVES:

- Analyze daily sales patterns and trends.

- Examine marketing spend vs sales correlations.

- Assess temporal factors like day-of-week effects.

- Build ARIMA forecasts for 30 days.

- Provide insights for inventory and marketing planning.

Methodology

Exploratory Analysis:

- Line plots and summary statistics to visualize trends, variability, and outliers.

- Optional differencing applied if series non-stationary (checked with ADF test).

ARIMA Modeling:

- auto.arima() with seasonal = FALSE to find best ARIMA(p,d,q).

- Residuals checked for autocorrelation, normality, and variance stability.

- Best model selected via AIC/BIC.

Methodology

Dataset Context:

- Daily retail sales data including: sales revenue, marketing spend, visitors, inventory, weather.

- Only sales revenue is modeled (univariate ARIMA).

Data Preparation:

- Chronologically arranged, missing values handled.

- Converted to daily time series (ts(data$Daily_Sales_Revenue, frequency = 1)).

Methodology

Forecasting:

- 30-day sales forecast generated.

- Accuracy evaluated with MAE and RMSE.

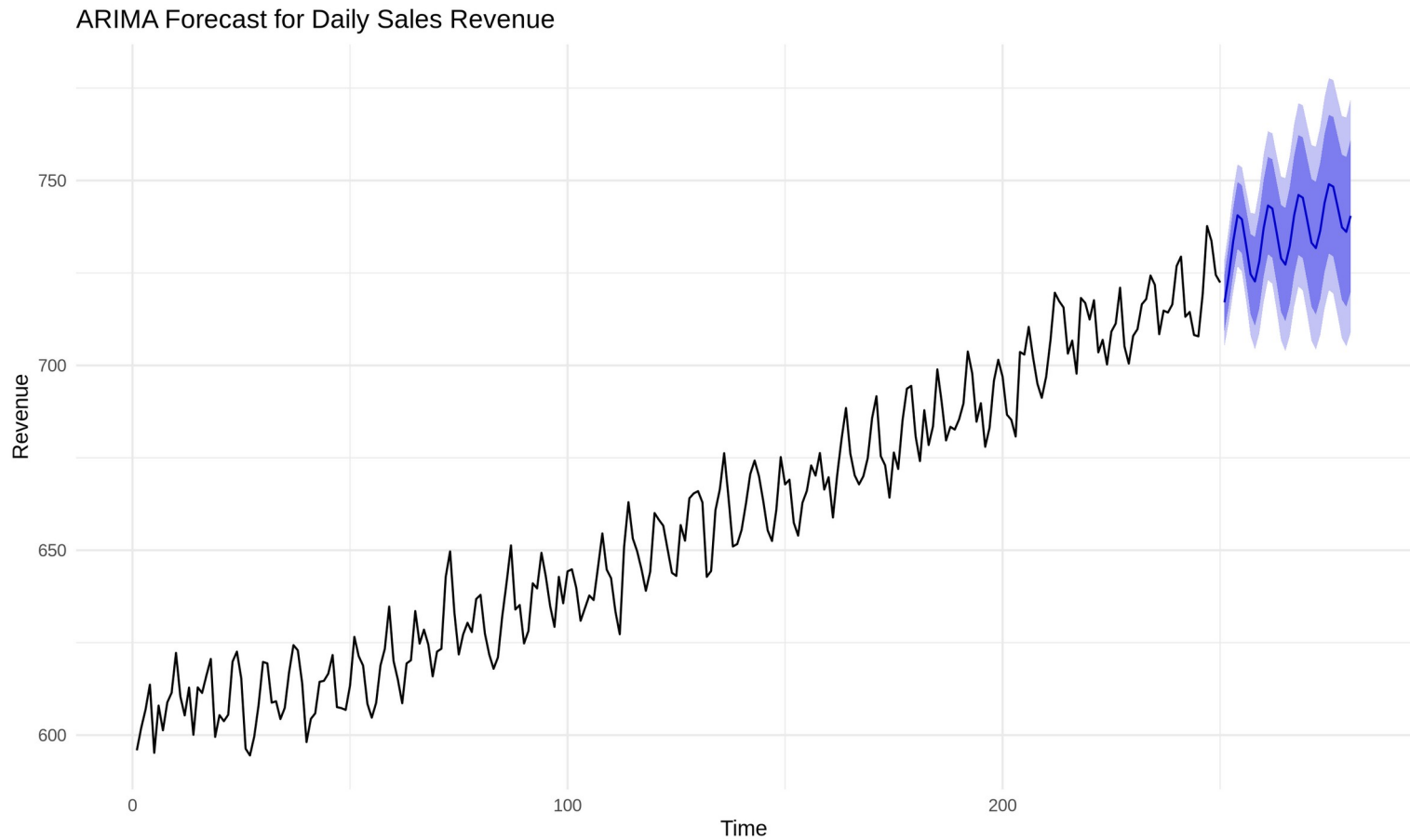- Insights drawn for inventory and marketing planning.

Figure 1;

Daily Sales Revenue Over Time

Findings

- Using auto.arima, the best model selected was ARIMA(3,1,2) with drift.

- AR(3): Sales from the previous three days strongly influence current sales.

- Drift: A positive drift of 0.5223, representing the underlying upward sales trend.

- Model Evaluation:

- The model performed very well on training data. It achieved a MAPE of 0.71%, meaning forecasts were, on average, within about 1% of actual sales.

- The Ljung-Box test showed a significant p-value ($p < 0.05$), meaning there is still some minor autocorrelation left in the residuals. Even so, the extremely low MAPE indicates that the model is reliable for practical forecasting.

Figure 2;

ARIMA Forecast for Daily Sales Revenue

Findings

The forecast indicates that daily revenue is expected to surpass the 750 mark within the next 30 days. Notably, although the model is non-seasonal (Pure ARIMA), it successfully reproduces the historical "sawtooth" volatility pattern, demonstrating that the underlying momentum of the business cycles has been effectively captured by the Auto-Regressive terms.

Summary and Recommendation

Summary:

- Pure ARIMA effectively forecasts daily sales trends.

- Accuracy measured via MAE and RMSE.

Recommendations:

- Use short-term forecasts to manage inventory and marketing budgets.

- Explore multivariate models in the future including visitors and marketing as regressors.

- R codes:

https://github.com/AlieeLinux/stop-statistic-casestudy-geez