# Sales Forecasting with Time Series Analysis: Comprehensive Report

## Table of Contents

1. **Executive Summary & Libraries Used**
2. **Data Analysis & Time Series Decomposition**
3. **Forecasting Methodologies & Model Performance**
4. **Key Findings & Recommendations**

## 1. Executive Summary & Libraries Used

This report presents a comprehensive analysis of sales forecasting using multiple time series and machine learning techniques. The analysis covers data from 2015 to 2025, implementing both traditional statistical methods (ARIMA, Simple Exponential Smoothing) and modern machine learning approaches (Random Forest, XGBoost, Linear Regression). The study demonstrates the effectiveness of ensemble methods in improving forecast accuracy, with robust evaluation using time series cross-validation techniques.

**Key Libraries Utilized:**

* **Data Processing**: pandas, numpy, matplotlib, seaborn
* **Time Series Analysis**: statsmodels (seasonal decomposition, ARIMA, ADF test), pmdarima (auto-ARIMA)
* **Machine Learning**: sklearn (Linear Regression, Random Forest, metrics, cross-validation), xgboost
* **Model Persistence**: joblib

## 2. Data Analysis & Time Series Decomposition

**Dataset Overview:** The dataset contains daily sales data spanning from 2015 to 2025. The exploratory analysis reveals a clear upward trend with sales growing from approximately 100-150 units in 2015 to 400-500 units by 2024-2025. Strong seasonal patterns are evident with regular cyclical variations throughout the year, alongside significant daily fluctuations around the underlying trend.

**Time Series Decomposition:** The seasonal decomposition using additive modeling reveals three distinct components: a consistent upward trend indicating business growth, regular cyclical patterns with annual periodicity (365 days), and random fluctuations after removing trend and seasonal components. The Augmented Dickey-Fuller (ADF) test indicates the need for differencing to achieve stationarity, which is addressed in the ARIMA modeling phase.

## 3. Forecasting Methodologies & Model Performance

**Traditional Time Series Models:** Simple Exponential Smoothing (SES) captures underlying trends with exponential weighting of historical observations. ARIMA modeling uses automated parameter selection via auto\_arima function for optimal 30-day forward forecasting.

**Machine Learning Approaches:** Sophisticated feature engineering includes temporal features (day of week, month, weekend indicator), rolling statistics (7-day mean and standard deviation), lag features (previous 7 days' sales), and low-variance feature removal. Four models were implemented: Linear Regression (baseline), Random Forest (non-linear patterns), XGBoost (gradient boosting), and an Ensemble method combining all three.

**Performance Evaluation:** Models were evaluated using MAE, RMSE, and MAPE metrics. Time series cross-validation with 5 folds provided robust generalization assessment. ARIMA showed reasonable accuracy for 30-day forecasts, while the ensemble approach demonstrated superior performance by balancing different model strengths. Random Forest effectively captured seasonal patterns, Linear Regression provided stable baseline predictions, and the ensemble optimized overall performance.

## 4. Key Findings & Recommendations

**Model Performance Insights:** The ensemble methodology demonstrates superior performance by combining Linear Regression, Random Forest, and XGBoost strengths. Lag features and rolling statistics proved crucial for accurate predictions, while both traditional and ML methods successfully captured seasonal patterns. The analysis reveals that ensemble approaches provide better forecast accuracy than individual models.

**Business Implications & Recommendations:** The consistent upward trend indicates healthy business growth, while seasonal patterns inform inventory management and resource allocation decisions. For production deployment, the ensemble model shows strong promise with all models saved via joblib for future use. Regular model retraining should be implemented to maintain accuracy, along with feature monitoring to ensure stability and identify new relevant variables.

**Practical Applications:** The comprehensive analysis successfully addresses sales forecasting complexity by leveraging multiple analytical approaches, comprehensive feature engineering, and rigorous evaluation methodologies. Time series cross-validation ensures reliable model assessment, while the resulting models provide actionable insights for demand planning, business strategy, and operational optimization.

**Conclusion:** This analysis demonstrates the effectiveness of combining traditional time series methods with modern machine learning approaches for sales forecasting. The implementation provides robust predictions suitable for business decision-making, with proper evaluation techniques ensuring model reliability. The methodology successfully captures both trend and seasonal components, offering valuable insights for strategic planning and operational decisions.