📊 Retail Sales Forecasting Report

# 1. Objective

The goal of this project is to develop a time series forecasting model to predict monthly sales for a retail store over the next 6 months.  
Accurate forecasts support inventory management, marketing campaigns, and financial planning.

# 2. Data Description

Dataset: Synthetic retail sales data (4 years, monthly frequency).  
Variables:  
 1. Date → Monthly timestamps (2020–2023).  
 2. SalesAmount → Monthly sales revenue.  
 3. Promotion → Binary flag for promotional campaigns (1 if promotion active).  
 4. HolidayMonth → Binary flag for December (holiday effect).  
  
Sales were generated with:  
 1. Upward trend (growth over time).  
 2. Seasonality (Q4 higher, Q1 lower).  
 3. Promotional and holiday boosts.  
 4. Random noise.

# 3. Data Preparation & Exploration

1. Datetime handling: Converted Date to datetime and set as index.  
 2. Missing values: Checked — none found.  
 3. Feature engineering: Added Promotion and HolidayMonth as potential regressors.  
  
**Time Series EDA:**  
 1. Trend: Clear upward trajectory in sales.  
 2. Seasonality: Strong peaks around December (holiday effect).  
 3. Stationarity: Augmented Dickey-Fuller test indicated non-stationarity   
 4. ACF/PACF plots: Suggested potential AR and MA terms.

# 4. Models Considered

Three models were tested:  
1. **ARIMA** (AutoRegressive Integrated Moving Average)  
 - Captures autoregressive patterns, differencing (trend), and moving averages.  
 - Tuned using auto-ARIMA (p,d,q).  
  
2. **SARIMA** (Seasonal ARIMA)  
 - Extension of ARIMA with seasonal terms (P,D,Q,s).  
 - Tuned using seasonal auto-ARIMA.  
  
3**. Exponential Smoothing** (Holt-Winters)  
 - Captures level, trend, and seasonality explicitly.  
 - Seasonal period = 12 (monthly data).

# 5. Model Training & Validation

- Train-Test Split:  
 - Training: First 42 months.  
 - Validation: Last 6 months.  
- Evaluation Metrics:  
 - MAE (Mean Absolute Error)  
 - MSE (Mean Squared Error)  
 - RMSE (Root Mean Squared Error)  
 - MAPE (Mean Absolute Percentage Error)

# 6. Results

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MAE | RMSE | MAPE |
| ARIMA | 2427.40 | 2769.81 | 21.03 |
| SARIMA | 1193.97 | 1534.42 | 9.37 |
| Exponential Smoothing | 1046.97 | 1383.19 | 8.03 |

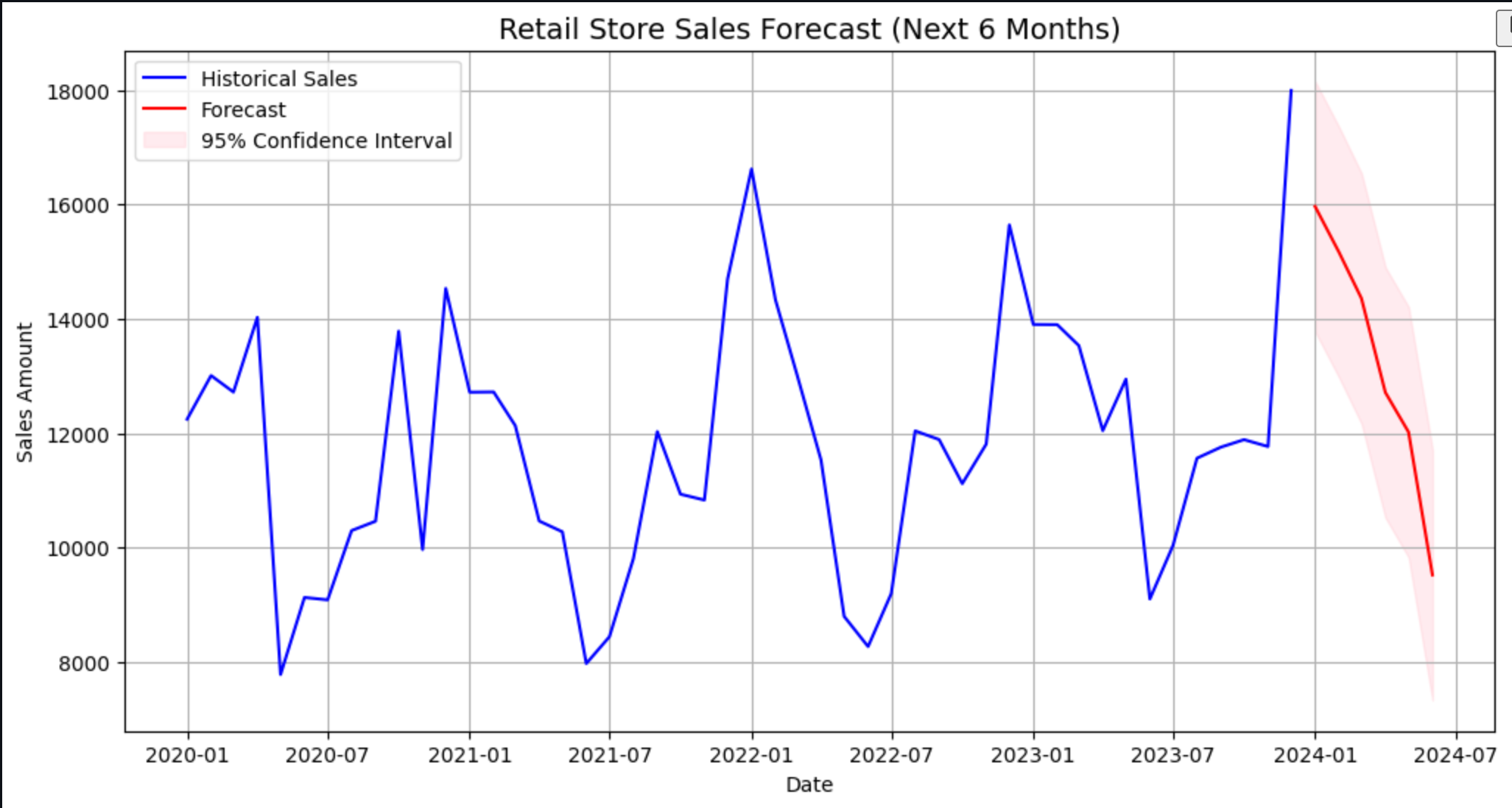
Best Model (based on RMSE): **Exponential Smoothing**  
Retrained this model on the entire dataset (2020–2023).

# 7. Forecasting (Next 6 Months)

**Forecast Table**

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Forecast | Lower CI | Upper CI |
| 2024-01-01 | 15966.66 | 13778.57 | 18154.77 |
| 2924-02-01 | 15165.60 | 12977.64 | 17353.72 |
| 2024-03-01 | 14360.76 | 12172.65 | 16548.88 |
| 2024-04-01 | 12716.09 | 10527.97 | 14904.20 |
| 2024-05-01 | 12023.30 | 9835.19 | 14211.21 |
| 2024-06-01 | 9525.57 | 7337.46 | 11713.69 |

**Visualization:**  
- Historical data shown in blue.  
- Forecast shown in red.  
- 95% confidence interval shaded in pink.



# 8. Conclusion & Business Implications

- The chosen model Exponential Smoothing successfully captured both trend and seasonality in retail sales.  
- Promotions and holidays significantly increase sales.  
- The 6-month forecast provides actionable insights for:  
 - Inventory planning → stock increases before holiday season.  
 - Marketing strategy → schedule promotions strategically.  
 - Financial budgeting → more accurate revenue projections.  
  
**Future work:**  
- Incorporate external regressors (e.g., inflation, competitor sales).  
- Explore machine learning models (XGBoost, LSTM, Transformer).  
- Automate forecast updates with live sales data.