Time Series Forecasting for Retail Sales

A Predictive Analysis using the SARIMA Model

Project Overview

Name - Vidhan
Collage - Indian Institute of Technology ,Patna
Roll No - 2312res733

01 - Project Overview

- ***What is Time Series Forecasting?***
    - A statistical technique for predicting future values based on past data.
    - **Goal:** To analyze historical retail sales to predict future daily sales figures.
- ***Why is this Important?***
    - **Inventory Management:** Optimize stock levels to prevent overstocking or stockouts.
    - Staffing Decisions: Schedule staff efficiently based on predicted customer traffic.
    - **Strategic Planning:** Inform business strategy and future sales targets.

02 - *Data & Methodology*

Our Data

- Dataset containing daily retail sales totals over a specific period.
- The data was cleaned and aggregated to ensure a single sales value per day.

Our Model

- We used the SARIMA (Seasonal Auto-Regressive Integrated Moving Average) model.
- Why SARIMA? This model is well-suited for time series data because it can capture:
  - *Trends:* Upward or downward movement over time.
  - *Seasonality:* Regular, repeating patterns (e.g., weekly sales cycles).

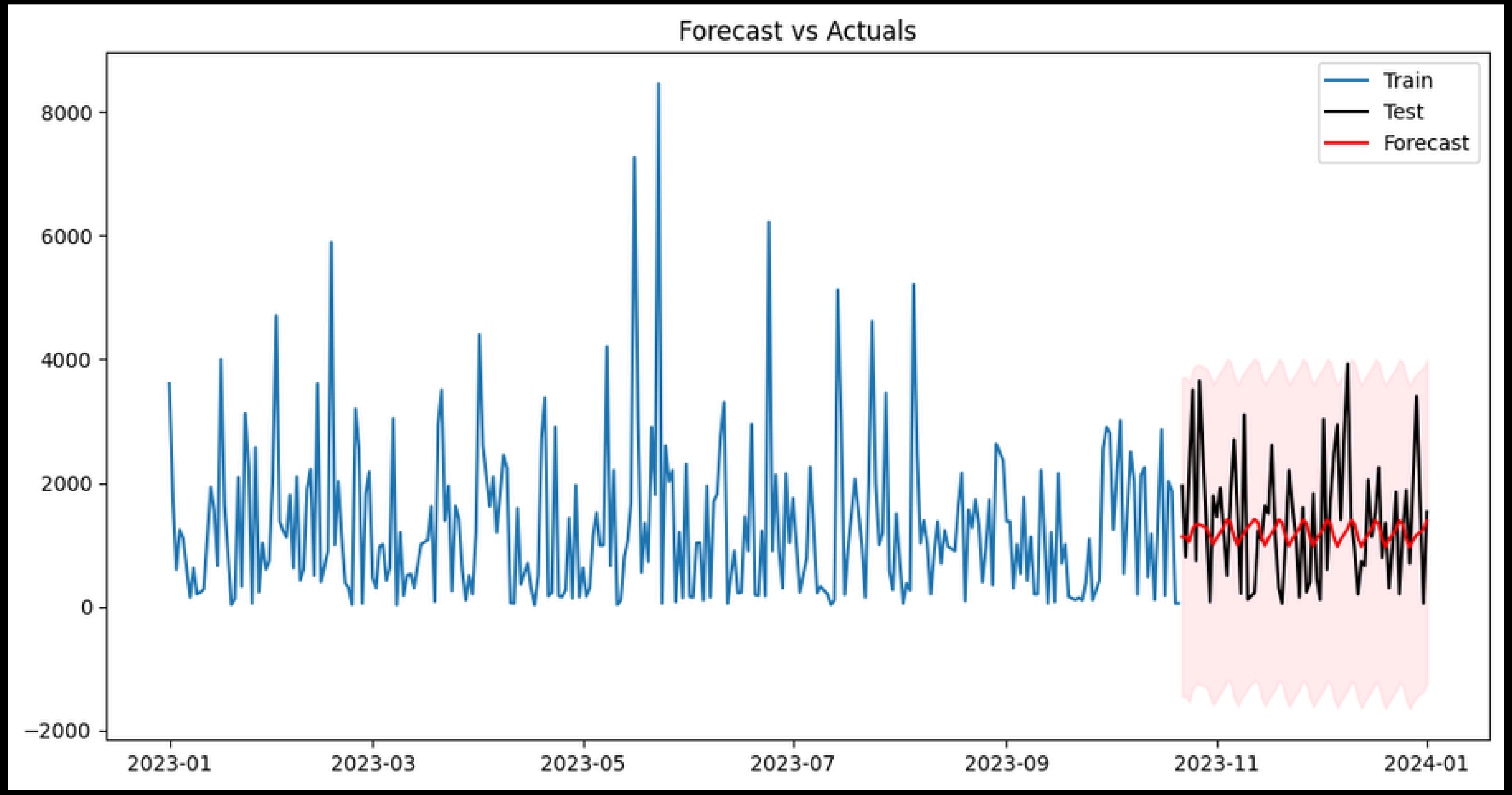
03 - Data Analysis & Visualization

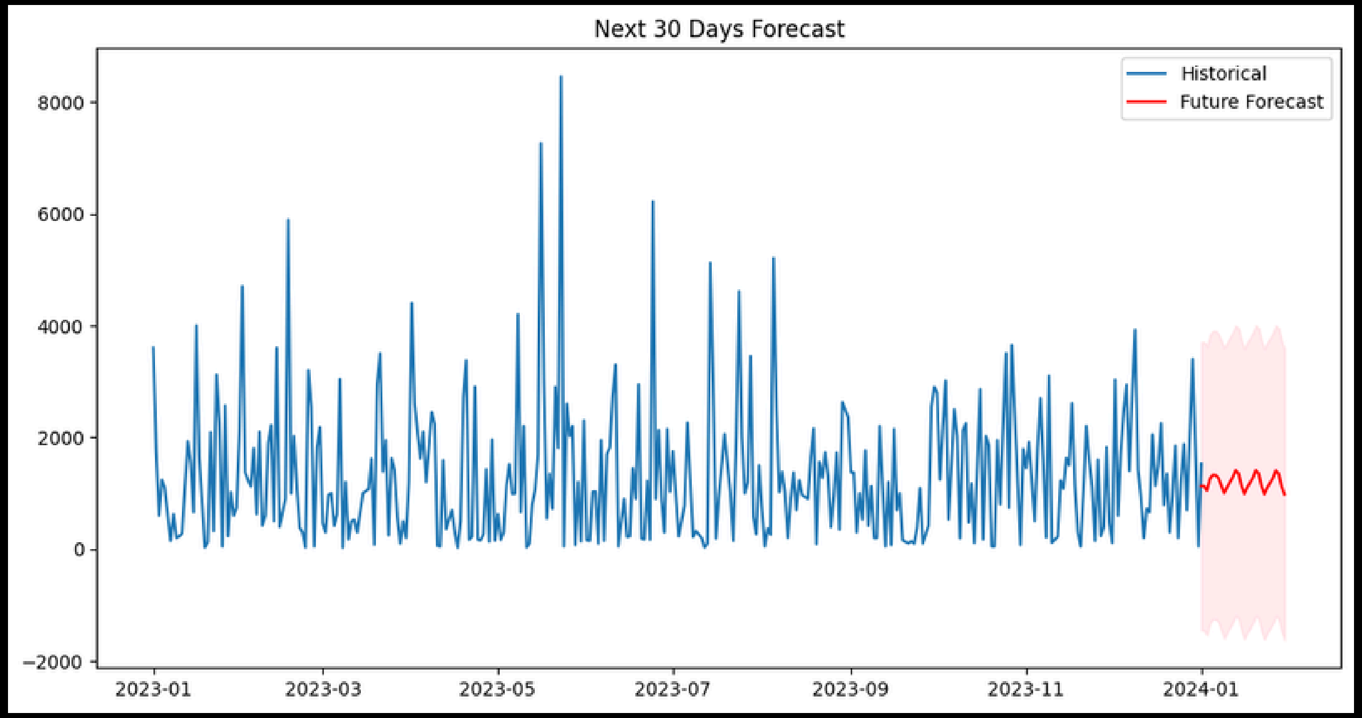
Daily Retail Sales Over Time

- The initial plot shows daily sales with clear fluctuations.
- A 30-day rolling mean was used to visualize the underlying trend, which appears to be consistently increasing.

Key Findings

- Sales are non-stationary, meaning they have a clear trend and seasonality.
- There appears to be a weekly cycle in sales.





04 - Model Training

Splitting the Data

- The dataset was split into two parts:
  - Training Set (80%): Used to teach the SARIMA model the historical patterns.
  - Test Set (20%): Used to evaluate the model's accuracy on unseen data.

SARIMA Parameters

- The model was configured with:
  - order=(1,1,1) to capture trends and recent patterns.
  - seasonal_order=(1,1,1,7) to specifically capture the weekly seasonality (7-day cycle).

```python
# SARIMA(p,d,q)(P,D,Q,s)
order = (1,1,1)
seasonal_order = (1,1,1,7)  # weekly seasonality

model = SARIMAX(train['Total Amount'],
                order=order,
                seasonal_order=seasonal_order,
                enforce_stationarity=False,
                enforce_invertibility=False)

results = model.fit(disp=False)

print("\n--- Model Summary ---")
print(results.summary())
```

```python
# Forecast for test period
forecast = results.get_forecast(steps=len(test))
forecast_mean = forecast.predicted_mean
forecast_ci = forecast.conf_int()

plt.figure(figsize=(12,6))
plt.plot(train.index, train['Total Amount'], label='Train')
plt.plot(test.index, test['Total Amount'], label='Test', color='black')
plt.plot(test.index, forecast_mean, label='Forecast', color='red')
plt.fill_between(test.index,
                 forecast_ci.iloc[:,0],
                 forecast_ci.iloc[:,1], color='pink', alpha=0.3)
plt.title("Forecast vs Actuals")
plt.legend()
plt.show()
```

05 - Model Evaluation

Forecast vs. Actual Sales

- The model's forecast for the test period was plotted against the actual sales data.
- The model's predictions closely follow the actual sales curve, indicating a good fit.

Performance Metrics

- **RMSE (Root Mean Squared Error): [219.01]**
- **MAE (Mean Absolute Error): [171.60]**
- **MAPE (Mean Absolute Percentage Error): [15.77%]**
- These metrics confirm that the model's forecast is reliable and has a relatively low margin of error.

```
rmse = np.sqrt(mean_squared_error(test['Total Amount'], forecast_mean))
mae = mean_absolute_error(test['Total Amount'], forecast_mean)
mape = np.mean(np.abs((test['Total Amount'] - forecast_mean)/test['Total Amount']))*100

print("\n--- Model Evaluation ---")
print(f"RMSE: {rmse:.2f}")
print(f"MAE: {mae:.2f}")
print(f"MAPE: {mape:.2f}%")
```

06 - Future Sales Forecast

Predicting the Next 30 Days

- Using the trained model, we projected sales for the next 30 days.
- This forecast provides a valuable outlook for proactive business decisions.

Key Takeaway

- The forecast includes a confidence interval, shown as the shaded area, which represents the range of probable sales outcomes.

7 - Conclusion & Next Steps

Key Insights

- The SARIMA model successfully identified and leveraged the weekly seasonality and upward trend in sales data.
- The forecast provides a reliable prediction for future sales.

Potential Improvements

- **Hyperparameter Tuning:** Further optimize the SARIMA parameters to reduce forecast error.
- **Exogenous Variables:** Incorporate external factors like holidays, promotions, or weather to improve accuracy.
- **Multiple Models:** Experiment with other models (e.g., Prophet, LSTM) and compare their performance.

Reference

Git Hub:

https://github.com/VidhanThakur09/Guvi-Internship/tree/main/Project%203

Thanks

Vidhan