End-to-End Project Report: Time Series Forecasting Using ARIMA & SARIMAX

1. Executive Summary

This project focuses on forecasting future sales using classical time series models in Python, specifically **ARIMA** and **SARIMAX**. The objective was to analyze historical sales data, identify underlying patterns such as trend and seasonality, build reliable forecasting models, and generate future sales predictions to support business planning and decision-making.

---

2. Business Objective

- To forecast future sales based on historical data
- To understand sales trends and seasonal behavior
- To support stakeholders in demand planning, inventory management, and strategic decision-making

---

3. Dataset Overview

- Input data consisted of historical time-based sales records
- Key column used: **Date / Time Index** and **Sales (Value)**
- Data frequency: Daily / Monthly (as per dataset)
- Target variable: **Sales**

---

4. Tools & Technologies Used

- Programming Language: **Python**
- Libraries:
- Pandas & NumPy – data handling and preprocessing
- Matplotlib & Seaborn – visualization
- Statsmodels – ARIMA and SARIMAX modeling
- Scikit-learn – evaluation metrics

---

5. Step-by-Step Workflow Performed

Step 1: Data Loading

- Imported the dataset into Python using Pandas
- Converted date column into datetime format

• Set the date column as the time index

Step 2: Data Understanding & Exploration

• Checked data shape, data types, and missing values
• Visualized sales over time to identify:
• Overall trend
• Seasonality
• Fluctuations and irregular patterns

Step 3: Data Preprocessing

• Handled missing values (if any)
• Ensured consistent time frequency
• Sorted data by time index

Step 4: Stationarity Check

• Used visual inspection and statistical methods
• Applied differencing where required to make the series stationary
• Confirmed stationarity before model building

Step 5: Train-Test Split

• Split the dataset into:
• Training data (historical period)
• Testing data (recent period for validation)

Step 6: ARIMA Model Development

• Identified suitable (p, d, q) parameters
• Built the ARIMA model using training data
• Fitted the model and generated forecasts

Step 7: SARIMAX Model Development

• Extended ARIMA to SARIMAX to capture seasonality
• Defined seasonal parameters (P, D, Q, m)
• Trained the SARIMAX model
• Generated seasonal-aware forecasts

Step 8: Model Evaluation

• Compared actual vs predicted values
• Used evaluation metrics such as:
• Mean Absolute Error (MAE)
• Root Mean Squared Error (RMSE)
• Assessed model performance visually and numerically

Step 9: Future Forecasting

- Generated future date ranges
- Forecasted sales for upcoming periods
- Combined historical data with future predictions for clarity

Step 10: Visualization & Insights

- Plotted:
- Historical sales
- Forecasted sales
- Confidence intervals
- Highlighted trend continuation and seasonal patterns

---

6. Key Insights

- Sales data shows a clear trend and seasonal behavior
- SARIMAX performed better in capturing seasonality compared to ARIMA
- Forecasts provide a reliable estimate for short- to mid-term planning

---

7. Business Impact

- Enables proactive demand forecasting
- Helps in inventory and resource planning
- Reduces uncertainty in decision-making
- Provides data-driven support to stakeholders

---

8. Limitations

- Forecast accuracy depends on historical data quality
- External factors (market changes, promotions, economic events) are not included
- Long-term forecasts may have higher uncertainty

---

9. Future Scope & Improvements

- Incorporate external variables (holidays, promotions, pricing)
- Automate model selection and tuning
- Deploy the model using dashboards or APIs
- Compare with machine learning and deep learning models

---

10. Conclusion

This end-to-end forecasting project successfully demonstrates the use of ARIMA and SARIMAX models to analyze historical sales data and predict future trends. The structured approach, from data preprocessing to forecasting and evaluation, ensures transparency and reliability, making the results suitable for business and stakeholder decision-making.

Prepared by: Jwel Aktar
Role: Junior Data Scientist / Data Analyst