SALES FORECASTING WITH TIME SERIES ANALYSIS IN PYTHON

1. Background and Problem Statement

A U.S.-based chain of retail stores needed to enhance its monthly sales forecasting capabilities to improve inventory planning and reduce overstocking or missed sales. The finance team previously relied on spreadsheets and heuristic estimates, which led to stockouts during peak months and excessive inventory in others. They sought a reliable forecasting model using Python to analyze sales trends, seasonality, and growth, enabling proactive decision-making.

2. Objectives

- To develop a time series forecasting model for monthly sales using Python
- To detect and quantify seasonal and trend components in the sales data
- To evaluate the forecast model accuracy using appropriate metrics
- To produce a reporting dashboard that visualizes past sales and future predictions

3. Methodology

3.1 Data Description

- Source: Internal POS system (Jan 2019 – Dec 2023)
- Frequency: Monthly aggregated sales by store
- Total Observations: 60 months × 12 stores = 720 records
- Variables Used:
    - date (monthly)
    - store_id
    - monthly_sales

3.2 Preprocessing and Transformation

- Converted date column to datetime format and set it as index
- Checked for missing values and imputed with linear interpolation
- Applied log transformation to stabilize variance

- Differencing used to remove trend and achieve stationarity

3.3 Exploratory Time Series Analysis

- Decomposition using seasonal_decompose() (trend, seasonality, residual)
- ACF and PACF plots used to determine AR and MA terms
- Seasonal trend identified: peaks in November–December, dips in February

3.4 Modeling Techniques Applied

- Compared the following models:
    - ARIMA (Auto Regressive Integrated Moving Average)
    - SARIMA (Seasonal ARIMA)
    - Prophet (Facebook's time series forecasting model)
- Model selection based on:
    - RMSE
    - MAE
    - MAPE
    - Visual inspection of forecast alignment
- Final model: **SARIMA(1,1,1)(1,1,1,12)** selected based on lowest RMSE and best seasonality capture

3.5 Forecast Output

- Forecast generated for next 12 months (Jan–Dec 2024)
- Forecast interval: 95% confidence bounds included
- Interactive dashboard created using Plotly and Dash for internal stakeholders

4. Results and Accuracy

- RMSE: $12,870
- MAPE: 6.3%
- Strong seasonality captured, especially during holidays
- SARIMA model outperformed ARIMA and Prophet for this dataset

5. Strategic Recommendations

- Use monthly forecasts to guide procurement and store-level inventory planning
- Align marketing campaigns with seasonal peaks
- Use forecast intervals for risk-adjusted financial planning

6. Deliverables

- Python Jupyter notebook with end-to-end time series workflow
- .pkl file of final SARIMA model
- Monthly forecast CSV and visualization dashboard
- PDF report summarizing analysis, model, forecast, and recommendations

7. Stakeholder Relevance

Academic:

- Practical example of time series forecasting using SARIMA
- Demonstrates real-world implementation of model selection and diagnostics

Corporate:

- Applicable to retail, FMCG, or any business requiring monthly demand forecasting
- Facilitates proactive inventory, sales, and operations planning