SALES FORECASTING USING TIME-SERIES ANALYSIS IN PYTHON

1. Background and Problem Statement

A U.S.-based e-commerce startup, operating in the lifestyle and home décor segment, observed substantial volatility in its monthly sales volume. Despite collecting structured transaction data for the last three years, the company faced regular issues with understocking during promotions and overstocking in low seasons. Their existing forecasting method, based on simple averages and visual trend observation, failed to capture the temporal patterns in customer demand.

To support accurate inventory planning and better marketing calendar alignment, the company decided to build a systematic time-series forecasting pipeline using Python. The goal was to predict future sales and compare performance across various time-series models.

2. Objectives

- Analyze the monthly sales dataset for trends, seasonality, and cyclic patterns.
- Apply multiple forecasting techniques: ARIMA, Holt-Winters Exponential Smoothing, and Naïve Forecasting.
- Compare model performance using error metrics such as MAE, RMSE, and MAPE.
- Visualize actual vs predicted sales in an interactive dashboard.
- Generate a Python-based forecast report to guide business decisions for the next 12 months.

3. Methodology

3.1 Data Overview

- Format: Monthly aggregated sales (Jan 2021 – Dec 2023)
- Variables: Date (Month-Year), Total Sales (USD), Orders Count
- Data Source: Internal MySQL database, exported to CSV

3.2 Preprocessing in Python

- Missing values in some months imputed using linear interpolation
- Outlier months (e.g., Black Friday sales) retained with annotation

- Series converted to DateTime format with monthly frequency

3.3 Model 1: Naïve Forecast

- Next month's forecast = This month's actual
- Serves as the baseline model for performance comparison

3.4 Model 2: Holt-Winters Exponential Smoothing

- Captures level, trend, and seasonality components
- Parameters optimized using grid search
- Used both additive and multiplicative seasonal models

3.5 Model 3: ARIMA (AutoRegressive Integrated Moving Average)

- Stationarity checked using Augmented Dickey-Fuller (ADF) test
- Auto_arima used for parameter tuning (p, d, q)
- Residuals analyzed for autocorrelation using ACF/PACF plots

3.6 Forecast Evaluation

- Rolling cross-validation for error metric estimation
- Metrics used:
    - MAE: Mean Absolute Error
    - RMSE: Root Mean Squared Error
    - MAPE: Mean Absolute Percentage Error

3.7 Visualization & Dashboarding

- Created with Plotly Dash
- Interactive sliders to compare time windows
- Tabs to switch between raw data, forecasts, and error summaries

4. Results

Model	MAE	RMSE	MAPE (%)
Naïve Forecast	5,210	6,012	17.3
Holt-Winters (Additive)	3,420	4,118	11.2

ARIMA (2,1,2)	2,985	3,800	9.6

- ARIMA performed best on all error metrics.
- The dashboard clearly showed that ARIMA captured the post-holiday sales dip better than Holt-Winters.

5. Interpretation and Insights

- Strong monthly seasonality detected (peak in November and dip in February).
- Holt-Winters was more interpretable but slightly less accurate than ARIMA.
- Naïve model performed poorly, validating the need for statistical forecasting.
- Forecasts for 2024 suggested a 12% YoY increase in Q4 sales.

6. Recommendations

- Adopt ARIMA-based forecasting for monthly inventory planning.
- Use Holt-Winters as a secondary method for interpretability and triangulation.
- Automate the pipeline using scheduled Python scripts and export monthly forecasts to Google Sheets for stakeholder access.
- Integrate promotion calendar and external economic data in future iterations.

7. Future Work

- Incorporate exogenous variables (Google Trends, ad spend) into ARIMAX or SARIMAX models
- Explore ensemble approaches combining ARIMA, Prophet, and LSTM
- Add SKU-level forecasting for granular inventory optimization

8. Stakeholder Relevance

Academic Use

- A practical case for time-series forecasting using Python
- Useful for teaching model evaluation, error metrics, and decomposition techniques

Corporate Use

- Actionable sales forecast report for inventory, marketing, and finance teams
- Supports decision-making in seasonal pricing and restocking cycles