DS4220 - Homework Assignment 5

Time Series Forecasting: E-commerce Sales Analysis

Due Date: Thursday, December 4, 2025 by 11:59 PM
Submission: Email your completed R Notebook file (.Rmd) to the instructor
Points: 100 total points

Instructions

This assignment uses the e-commerce daily sales dataset provided in class. All questions can be answered using the R codes and techniques we have covered in our time series lectures (Sessions 1-2). Please complete this assignment in R following the same format as our lecture files.

Important Notes:

- Show all your R code and output

- Provide written interpretations and business insights for each question - computation alone is not sufficient

- Use proper headings and formatting in your R Notebook

- Submit your completed .Rmd file via email by the deadline

Part 1: Exponential Smoothing Methods (25 points)

E-commerce businesses require accurate sales forecasting for inventory management, staffing decisions, and revenue planning. Apply exponential smoothing techniques to model daily sales patterns and generate forecasts for business planning.

1.1 Load the e-commerce sales dataset and create a proper time series object using the (daily_sales) variable. Plot the data and describe the key patterns you observe (trend, seasonality, level changes, any notable business events). **(8 points)**

1.2 Apply Simple Exponential Smoothing to the daily sales data. Generate 30-day forecasts and interpret the results. What type of sales patterns is this method best suited for? **(8 points)**

1.3 Apply Holt-Winters exponential smoothing (both additive and multiplicative seasonality) to capture seasonal patterns in daily sales. Compare the forecasts from both approaches and explain which performs better for this e-commerce dataset. **(9 points)**

Part 2: Stationarity Analysis and Data Preparation (20 points)

ARIMA models require stationary data for reliable parameter estimation. Analyze the stationarity properties of the daily sales series and apply appropriate transformations.

2.1 Test the daily sales series for stationarity using appropriate diagnostic tools. Interpret the results and explain why stationarity matters for ARIMA modeling of sales data. **(8 points)**

2.2 Apply differencing transformations to achieve stationarity. Show both first differencing and seasonal differencing results (consider weekly seasonality for daily data). Which approach is more effective for this sales dataset? Justify your answer with evidence. **(12 points)**

Part 3: ACF/PACF Analysis and ARIMA Parameter Identification (25 points)

Use autocorrelation analysis to identify appropriate ARIMA model parameters for the daily sales data.

3.1 Generate ACF and PACF plots for the stationary daily sales data. Display both the plots and the numerical correlation values for the first 10 lags. **(10 points)**

3.2 Interpret the ACF and PACF patterns to suggest appropriate ARIMA parameters (p, d, q). Use the statistical significance thresholds to justify your parameter recommendations. Show the code that automatically suggests these parameters. **(10 points)**

3.3 Explain the business meaning of the suggested ARIMA parameters in the context of e-commerce sales forecasting. What do the autoregressive and moving average components tell us about sales dependencies and customer behavior patterns? **(5 points)**

Part 4: ARIMA Model Building and Forecasting (20 points)

Build ARIMA models using both automatic and manual parameter selection, then generate forecasts for business planning.

4.1 Fit an ARIMA model using automatic parameter selection. Display the model summary and interpret the selected parameters (p,d,q). What does this tell us about the daily sales time series structure? **(8 points)**

4.2 Generate 30-day forecasts using the fitted ARIMA model. Create forecast plots with confidence intervals and extract the specific forecast values. **(7 points)**

4.3 Interpret the ARIMA forecasts in a business context. How would e-commerce managers use these sales forecasts for inventory planning, marketing budget allocation, and staffing decisions? What are the limitations of these predictions? **(5 points)**

Part 5: Model Comparison and Business Recommendations (10 points)

Compare the forecasting performance of exponential smoothing and ARIMA approaches to provide business recommendations.

5.1 Compare the forecasts from Holt-Winters exponential smoothing and ARIMA models for the same 30-day period. Which method provides more realistic forecasts for daily sales? Consider the business context of e-commerce seasonality. **(5 points)**

5.2 Based on your analysis, recommend the most appropriate forecasting method for e-commerce sales planning. Consider both accuracy and interpretability in your recommendation. Provide specific business

justification for your choice and discuss how this would impact inventory management and marketing decisions. **(5 points)**

Dataset Information:

The e-commerce daily sales dataset contains comprehensive daily business metrics for an online retail company. This dataset includes sales revenue, customer behavior, marketing activities, economic indicators, and operational metrics that influence daily sales performance.

Key Variables for Time Series Analysis:

- `date`: Daily date observations (time series index)

- `daily_sales`: Daily sales revenue in dollars (primary target variable)

- `year`, `month`, `quarter`: Time period indicators for seasonal analysis

- `day_of_week`: Day of week for weekly seasonality patterns

- `is_weekend`: Weekend indicator for business cycle effects

Note: While the dataset contains many external variables (marketing spend, weather, economic indicators), this assignment focuses on univariate time series analysis using only the `daily_sales` variable and time components, consistent with our Sessions 1-2 course material.

Grading Criteria:

- **Code Quality (20%)**: Well-commented, clean R code that runs without errors

- **Technical Analysis (40%)**: Correct implementation of time series methods and proper interpretation of statistical results

- **Business Insights (25%)**: Meaningful interpretation of results in business context with practical recommendations

- **Communication (15%)**: Clear explanations, proper formatting, and professional presentation

Tips for Success:

- Use the `forecast` package for exponential smoothing and ARIMA functions

- Remember to create proper time series objects with correct frequency (daily = 365.25 or weekly = 7)

- Always interpret statistical results in e-commerce business context

- Include forecast plots with appropriate titles and labels

- Explain your recommendations clearly with supporting evidence