University Examinations 2023/2024

THIRD YEAR SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE DATA SCIENCE

CDS 3352: TIME SERIES DATA ANALYSIS

DATE: APRIL 2024   **TIME: 2 HOURS**

INSTRUCTIONS: *Answer question **one** and any other **two** questions*

QUESTION ONE (30 MARKS)

a. Consider a time series dataset representing monthly sales data for a retail store from 2015 to 2023, which exhibits strong seasonal patterns and a gradually increasing trend. How would you apply an additive model to forecast sales for the year 2024? Describe the process and the reasoning behind choosing an additive model over a multiplicative model.   (5 Marks)

b. Write a Python function to apply a moving average filter to a time series dataset. The function should take a pandas Series object and the size of the moving window as input and return the smoothed time series as output.   (5 Marks)

c. Given a dataset with missing values, outliers, and multiple time series variables of different frequencies, outline a strategy for cleaning and preprocessing this data for analysis. What specific techniques would you use to handle each of these issues?

(5 Marks)

d. Describe how you would use line plots and histograms to analyze the trend and distribution of monthly unemployment rates over a decade. What insights can these visualizations provide regarding economic health? (5 Marks)

e. Describe how you would use time series decomposition to analyze monthly retail sales data, identifying seasonal patterns to plan stock levels. What would indicate the presence of an additive versus a multiplicative model? (5 Marks)

f. Illustrate how to use a moving average to smooth a time series dataset in Python, aiming to reduce volatility and trend. Compare the original and smoothed series through plotting. (5 Marks)

QUESTION TWO (20 MARKS)

a. Using the statsmodels library in Python, demonstrate how to perform a seasonal decomposition of a time series dataset. Include code to plot the observed, trend, seasonal, and residual components. (7 Marks)

b. You have daily stock price data for multiple companies over the last decade. Explain how you would perform EDA to uncover trends, seasonality, and correlation between different stocks. (6 Marks)

c. Given daily stock price data, discuss how decomposition can help identify underlying trends and seasonal effects. How might anomalies in the data affect the decomposition process? (7 Marks)

QUESTION THREE (20 MARKS)

a. Provide a Python script that identifies missing values in a time series dataset and imputes them using forward fill and backward fill methods. Explain the choice of method under different scenarios. (7 Marks)

b. Provide Python code for conducting the KPSS test on a time series to evaluate stationarity. Discuss how the KPSS test results complement the findings from the ADF test. (7 Marks)

c. Discuss different transformations (e.g., logarithmic, square root) that can be applied to make a non-stationary time series stationary. Provide examples of when each transformation might be appropriate. (6 Marks)

QUESTION FOUR (20 MARKS)

a. Given daily trading volume data for a stock, explain how scatter plots could help identify patterns or anomalies in trading activity. What would you look for in these plots?
(6 Marks)

b. Write Python code to perform the Augmented Dickey-Fuller (ADF) test on a time series dataset. Interpret the test results to assess stationarity. (8 Marks)

c. Explain why feature scaling might be necessary when preparing time series data for predictive modeling. Give an example where not scaling features could adversely affect model performance. (6 Marks)

QUESTION FIVE (20 MARKS)

a. Illustrate how to plot multiple time series data on a single graph for comparison, detailing the steps to differentiate each series visually. (7 Marks)

b. Evaluate the potential use of machine learning models for weather forecasting. Compare their potential advantages and disadvantages to traditional time series forecasting methods. (7 Marks)

c. Discuss how exponential smoothing models could be used to forecast monthly sales in a retail business. Include how you would account for seasonality and trends in the sales data. (6 Marks)