UNIRAZAK
UNIVERSITI TUN ABDUL RAZAK

FINAL EXAMINATION

MARCH 2024

COURSE TITLE	DATA SCIENCE FOR SOCIAL SCIENCES
COURSE CODE	TISC3113
DATE/DAY	27 JUNE 2024 / THURSDAY
TIME/DURATION	02:00 PM - 04:00 PM / 02 Hour(s) 00 Minute(s)

INSTRUCTIONS TO CANDIDATES :

1. Please read the instruction under each section carefully.
2. Candidates are reminded not to bring into examination hall/room any form of written materials or electronic gadget except for stationery that is permitted by the Invigilator.
3. Students who are caught breaching the Examination Rules and Regulation will be charged with an academic dishonesty and if found guilty of the offence, the maximum penalty is expulsion from the University.

(This Question Paper consists of 3 Printed Pages including front page)

DO NOT OPEN THE QUESTION PAPER UNTIL YOU ARE TOLD TO DO SO

There are FOUR (4) questions in this paper. Answer ALL questions in the answer booklet provided.

[100 MARKS]

QUESTION 1 (25 Marks)

You are managing a market research study and want to assess the effectiveness of different advertising strategies among consumers. A survey is designed with multiple-choice questions to gather information on advertising platform preferences, frequency of exposure to ads, and perceived impact on purchasing decisions. The collected data will be subjected to descriptive statistical analysis.

a) Define descriptive statistics and explain its importance in analyzing survey data. How can descriptive statistics help in summarizing and interpreting the collected information?

(8 marks)

b) Explain the different types of data that can be collected in the survey.

(7 marks)

c) Discuss the role of measures of central tendency (mean, median, mode) in analyzing the survey data.

(10 marks)

QUESTION 2 (25 Marks)

Imagine you're working with a dataset that records stock market information, including columns for date, stock price, trading volume, and dividend yield. Answer the following questions related to data manipulation using pandas:

a) Discuss the method to load the sales data from a CSV file into a pandas DataFrame.

(8 marks)

b) Explain how to check the basic information about the DataFrame, such as the number of rows, columns, and data types of each column.

(8 marks)

c) Describe Pandas function that will return a statistical summary of the data.

(4 marks)

d) Discuss the significance of the 'groupby' function in pandas and provide an example scenario where it can be applied to extract meaningful insights from the weather dataset.

(5 marks)

QUESTION 3 (25 Marks)

A government agency is aiming to enhance agricultural productivity in a region using data science techniques. The agency has access to a diverse set of data, including weather patterns, soil quality, crop yields, and historical agricultural practices. The agency plans to leverage Python for data science to address these challenges and provide actionable insights to farmers. They are looking for data scientists to develop algorithms, visualize data, and create a user-friendly system for farmers to access relevant information.

a) Determine TWO popular Python libraries that can be used for the data visualization.

(8 marks)

b) Explain the concept of scatter plot and how it can be used to depict the relationship between two variables in agricultural productivity.

(10 marks)

c) Describe the concept of box plots for data visualization.

(7 marks)

QUESTION 4 (25 Marks)

You are provided with a hypothetical dataset containing the following economic indicators for a country:

- GDP Growth Rate: Quarterly GDP growth rates over the past five years.
- Inflation Rate: Monthly inflation rates over the past five years.
- Unemployment Rate: Quarterly unemployment rates over the past five years.
- Consumer Confidence Index: Annual consumer confidence index over the past five years.
- Trade Balance: Annual trade balance data over the past five years.

a) Describe the steps for dealing with missing data.

(8 marks)

b) Explain the descriptive statistics to summarize each economic indicator.

(5 marks)

c) Describe plots or charts would you use to visualize the trends in GDP growth rates over time?

(6 marks)

d) Discuss the importance detecting outliers in economic data, and how can outliers affect the analysis?

(6 marks)

*** END OF QUESTION PAPER ***