

Roll No



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SET A**

**SCHOOL OF MANAGEMENT  
END TERM EXAMINATION - JAN 2024**

**Semester :** Semester III - 2022

**Course Code :** BBB3003

**Course Name :** Essentials Statistics for Business Analytics

**Program :** BBA

**Date :** 09-JAN-2024

**Time :** 1:00 PM - 4:00 PM

**Max Marks :** 100

**Weightage :** 50%

**Instructions:**

- (i) Read all questions carefully and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and non-programmable calculator are permitted.
- (iv) Do not write any information on the question paper other than Roll Number.

**PART A**

**ANSWER ALL THE QUESTIONS**

**5 X 2M = 10M**

1. Explain the difference between the loc[] and iloc[] methods in Pandas.  
(CO1,CO2) [Knowledge]
2. How do you create a 3x3 matrix with values ranging from 0 to 8 in NumPy?  
(CO3,CO2) [Knowledge]
3. Write a Python script to print "Hello, World!".  
(CO4,CO3) [Knowledge]
4. What is a hypothesis in research methodology?  
(CO3,CO4) [Knowledge]
5. What is the purpose of a simple table, and how is it typically used in data presentation?  
(CO5,CO4) [Knowledge]

**PART B**

**ANSWER ALL THE QUESTIONS**

**5 X 10M = 50M**

6. Scenario: You are preparing a report on the average rainfall in five different cities over the past year. Question: Create a simple table to display this data effectively. Explain why a simple table is the most suitable choice for this scenario, and how it aids in the analysis and presentation of the data.  
(CO2,CO1) [Comprehension]

7. A school counselor wants to investigate whether there's an association between students' preferred learning styles (visual, auditory, or kinesthetic) and their chosen extracurricular activities (sports, arts, or academics). They survey a random sample of 60 students and obtain the following data:

Learning Style	Sports	Arts	Academics	Total
Visual	15	10	5	30
Auditory	8	12	10	30
Kinesthetic	7	8	15	30
Total	30	30	30	60

drive\_spreadsheetExport to Sheets

**Conduct a chi-square test of independence to determine if there's a significant association between learning styles and extracurricular activity choices.**

**Critical value from the table (for a significance level of 0.05):**

**$\chi^2$  critical = 5.991 (with 2 degrees of freedom)**

(CO2,CO3) [Comprehension]

8. You are provided with a dataset 'data.csv' containing two columns: 'Variable\_A' and 'Variable\_B'. These columns represent measurements from two different sensors during an experiment. Perform the following tasks using Python and the Pandas library:

1. Load the dataset into a Pandas DataFrame. (2 marks)
2. Calculate the Pearson correlation coefficient between 'Variable\_A' and 'Variable\_B'. (4 marks)
3. Interpret the meaning of the correlation coefficient in the context of this data. What does it tell you about the relationship between the two variables? (2 marks)
4. Visualize the relationship between 'Variable\_A' and 'Variable\_B' using an appropriate plot. Briefly describe what the plot shows. (2 marks)

(CO3,CO2) [Comprehension]

9. Demonstrate how to create a DataFrame using Pandas with custom data. Perform basic operations like adding a new column, deleting a column, and renaming columns. Illustrate each step with an example.

(CO4,CO3) [Comprehension]

10. Explain how to create a multi-plot layout using Matplotlib's subplot functionality. Illustrate with an example that includes at least three different plots on the same canvas.

(CO5,CO4) [Comprehension]

## PART C

**ANSWER ALL THE QUESTIONS**

**2 X 20M = 40M**

11. You have a dataset 'market\_data.csv' with three independent variables: 'Advertising\_Spend', 'Market\_Share', and 'Location\_Score', and one dependent variable: 'Sales\_Revenue'. Using Python, Pandas, and sklearn, perform a multiple linear regression analysis with the following steps:

1. Load the dataset into a Pandas DataFrame. (2 mark)
2. Conduct any necessary data cleaning or preprocessing. (2 mark)
3. Split the data into training and testing sets, using the three independent variables to predict 'Sales\_Revenue'. (4 marks)
4. Implement a multiple linear regression model using sklearn and train it on the training data. (4 marks)
5. Predict 'Sales\_Revenue' for the testing data and calculate the model's accuracy. (4 marks)
6. Interpret the coefficients of the model, discussing the influence of each independent variable on 'Sales\_Revenue'. (4 marks)

(CO5,CO4,CO3) [Application]

12. Answer the following questions in the context of the graphic representation of data.

1. Define the terms "diagram" and "graph" in the context of data visualisation. Briefly explain the differences and similarities between them.
2. What are the advantages and disadvantages of using diagrams and graphs compared to simply presenting data in tabular format? Explain your answer.

Choose one of the following type of diagram:

- **Simple Bar Diagram:** Represent the monthly sales of three different product categories in a grocery store.
- **Multiple Bar Diagram:** Compare the performance of two marketing campaigns across different target demographics.

3. Clearly explain the chosen diagram type and its appropriateness for the scenario you selected.

**4. Manually draw the chosen diagram on the paper, ensuring proper labeling and scaling.** Briefly explain the key insights conveyed by your diagram.

(CO5,CO4,CO3) [Application]