

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 6495

L

Unique Paper Code : 2344001201

Name of the Paper : Computer Science: Generic Elective

Name of the Course : **Data Analysis and Visualization using Python**

Semester : II

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. This question paper has **two** Sections **A** and **B**.
3. Question **1** in **Section A** is compulsory.
4. Attempt any **four** questions from **Section B**.
5. Parts of a question must be attempted together.
6. **Section A** carries **30** marks and each question in **Section B** carries **15** marks.
7. Use of Calculator is not allowed.

Assume numpy has been imported as np and pandas has been imported as pd.

Section A

(Compulsory)

1. (a) Given an array marks : (3)

marks = np.array ([55, 72, 88, 45, 67, 90, 34, 76, 81])

Write Python code for the following :

- (i) Print marks greater than 60.

P.T.O.

- (ii) Print marks that are odd numbers.
- (iii) Reshape the array into a 3×3 array named marks_2D.
- (b) Write Python commands to create a Series named S with values [5, 6, 7, 8] and index [1, 2, 3, 4] and then rename index 1 to 'one'. (3)
- (c) Write Python statements to create a 3 x 3 matrix using arange() in NumPy such that the matrix contains numbers from 1 to 9 arranged row-wise. Also, print its transpose. (3)
- (d) What is the purpose of legend in plotting? Explain with an appropriate example. (3)
- (e) Differentiate between loc() and iloc() with the help of a suitable example. (3)
- (f) Create a DataFrame having 6 rows and 3 columns and populate it with random values in the range 10 to 50. Set the row index as ['A', 'B', 'C', 'D', 'E', 'F'] and column names as ['X', 'Y', 'Z']. (3)
- (g) Write a command to merge the two given DataFrames df1 and df2 based on a common column Student_ID. (3)

```
df1 = pd.DataFrame({  
    'Student_ID': [101, 102, 103, 104],  
    'Name': ['Amit', 'Neha', 'Ravi', 'Sara']  
})  
  
df2 = pd.DataFrame({  
    'Student_ID': [101, 102, 103, 104],  
    'Marks': [85, 90, 78, 88]  
})
```

- (h) Differentiate between a series and a DataFrame. (3)
- (i) A retail store maintains a customer database containing fields such as customer name, membership type, number of purchases, annual spending, and customer satisfaction level. Classify each field as either qualitative or quantitative data. (3)
- (j) What is correlation? How does it help in understanding relationships between two variables? Mention the name of the command used to calculate correlation. (3)

Section B

- 2. (a) What is hierarchical indexing (Multiindex) in Pandas? Explain how to create and access data using hierarchical indexing. (5)
- (b) List and describe the steps involved in process of Data Analysis. (5)
- (c) You are given the following DataFrame : (5)

```
data = {  
    'Product': ['A', 'A', 'B', 'B', 'A', 'B'],  
    'Month': ['Jan', 'Feb', 'Jan', 'Feb', 'Mar', 'Mar'],  
    'Sales': [100, 120, 80, 90, 110, 95]  
}  
  
df = pd.DataFrame(data)
```

Write the code for the following :

- (i) Create a pivot table showing total sales of each product.

- (ii) Add total (margins) to the pivot table.
- (iii) Create a cross-tabulation showing the frequency of Product across different Months.

3. (a) Write the code to create a line plot using the following data: (5)

```
import matplotlib.pyplot as plt  
  
months = ['Jan', 'Feb', 'Mar', 'Apr']  
  
sales = [50, 70, 40, 90]
```

Use Specifications :

- Line color: blue and style: dashed (--)
- Markers: circle (o) with size 8
- Line width: 2
- Title: “Monthly Sales Trend”
- X-axis label: “Months”, Y-axis label: “Sales”
- Display the plot

(b) You are given the following DataFrame: (5)

```
data = {  
    'Math': [90, 85],  
    'Science': [88, 92]  
}  
  
df = pd.DataFrame(data, index=['Alice', 'Bob'])
```

Answer the following :

- (i) Write the output of: `df.stack()`
 - (ii) Write the output of: `df.unstack()`
 - (iii) Briefly explain the functioning of `stack()` and `unstack()`.
- (c) Define Categorical and interval data. Give examples of each. (5)
4. (a) Differentiate between descriptive statistics and inferential statistics with examples. (5)
- (b) You are given the following DataFrame: (5)

```
data = {  
    'Name': ['A', 'B', 'C', 'D'],  
    'Marks': [80, np.nan, 75, np.nan],  
    'Grade': ['A', 'B', None, 'C']  
}  
  
df = pd.DataFrame(data)
```

Write Python code for the following :

- (i) Identify and display the total number of missing values in the DataFrame.
- (ii) Fill missing values in the Marks column with 0.
- (iii) Drop all rows containing any missing values.

5. (a) Given the following data : (5)

```
import matplotlib.pyplot as plt
```

```
data = [4, 6, 7, 4, 5, 6, 8, 9]
```

```
categories = ['A', 'B', 'C']
```

```
values = [15, 25, 20]
```

Write Python code to:

- (i) Create 2 subplots (2 rows, 1 column)
 - (ii) Plot a histogram of data in the first subplot
 - (iii) Plot a bar chart of categories vs values in the second subplot
- (b) Given the dataset: [15, 22, 35, 40, 50], answer the following : (5)
- (i) Identify the minimum and maximum values of the dataset.
 - (ii) Write the formula used for Min-Max scaling and perform scaling of the dataset to the range [0,1].
 - (iii) Write any two advantages of using Min-Max scaling in data analysis or machine learning.
6. (a) Write NumPy commands to perform the following operations on an array num: (5)
- (i) Create an array num containing values from 31 to 46.
 - (ii) Convert the data type of array num to floating type.
 - (iii) Reshape the array num into a 4×4 array.
 - (iv) Replace the diagonal elements of the array with 0.

(v) Create a new array of 1 's having the same shape and data type as num.

(b) What is data normalization? Why is it important in data preprocessing? (5)

(c) You are given the following DataFrames : (5)

```
import matplotlib.pyplot as plt
```

```
df1 = pd.DataFrame({  
    'Product': ['A', 'B', 'C'],  
    'Sales': [120, 80, 150]  
})
```

```
df2 = pd.DataFrame({  
    'Product': ['D', 'E'],  
    'Sales': [90, 170]  
})
```

Write Python code for the following :

- (i) Concatenate df1 and df2 into a new DataFrame df.
- (ii) Create a bar plot of Product vs Sales with different bar colors.
- (iii) Add title, axis labels, and display the plot clearly.

7. (a) What are missing values? Explain different methods used in Pandas to handle missing data 5 with examples.

- (b) You are given the following list of student marks: (5)

marks = [35, 67, 89, 45, 72, 91, 58, 40, 76, 82]

Write Python code using Pandas to perform the following :

- (i) Create a Pandas Series named s using the given marks.
 - (ii) Divide the marks into the following categories using binning:
 - 0–40 → "Low"
 - 41–70 → "Medium"
 - 71–100 → "High"
 - (iii) Display the category assigned to each mark.
- (c) Explain unimodal, bimodal and multimodal distribution with the help of examples. (5)