[This question paper contains 8 printed pages.]

Your Roll No.....

H

Sr. No. of Question Paper : 5025

Unique Paper Code : 2342201202

Name of the Paper : Data Interpretation and Visualization using Python

Name of the Course : B.Sc. (Prog.)

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. **Section A** is compulsory.
3. Attempt any **4 (four)** questions from **Section B.**
4. Parts of a question must be answered together.

Section A
(Compulsory)

1. (a) Briefly explain different measures of central tendency. (3)

(b) With help of an example differentiate between loc and iloc operators. (3)

(c) Write and justify the output of the given script.

```
import pandas as pd
s1 = pd.Series(np.arange(5),
        index=['a', 'c', 'b', 'e', 'd'])
print(s1)
print(s1['b' : 'd'])
```

(3)

(d) Write and justify the output of the given Python statements. (3)

```
import pandas as pd
df = pd.DataFrame(np.arange(20).reshape(4,5))
print(df)
```

(e) Write the formula for calculating range. Calculate the range for the given data [32, 12, 56, 33, 44]. (3)

(f) Write and justify the output of the given Python statements. (3)

```
import pandas as pd
df = pd.DataFrame(np.arange(20).reshape(4,5))
print(df)
```

(g) What is a pivot table? Give one example. (3)

(h) Write the use of randint() function using appropriate example. (3)

(i) What is the purpose of legend in plotting? Explain with appropriate example. (3)

(j) Briefly explain regression using example. (3)

SECTION-B

2. (a) Consider the two given dataframes studf and resdf and write the Python statements to do the following : (10)

studf

rollno	name
101	abc
102	pqr
103	xyz

resdf

rollno	marks
101	66
102	76
101	55
104	82

(i) Display the rollno, name and marks for the students who appear in both dataframes.

(ii) Display the rollno, name and marks of students who appear in either dataframe.

(iii) Count the number of students whose result is not available yet.

(iv) Display the rollno of the students who do not appear in dataframe studf.

(v) Display all the unique rollnos present in either dataframe.

(b) With the help of suitable example, differentiate between is null() and notnull() methods. (5)

3. (a) Consider the given dataframe my df and write the Python statements to perform the following operations : (10)

	age	weight
102	23	52
103	34	67
101	25	60

(i) Set the title of the row index as rollno.

(ii) Add a column 'height' with values 152, 165, 171.

(iii) Sort the dataframe in descending order of the age column.

(iv) Reindex the dataframe in the order 101, 102 and 103.

(v) Drop the row corresponding to row index 102.

(b) What are the different ways to handle missing data? (5)

4. (a) Consider the file sales.csv as given below and answer the following questions. (10)

year;qtr1;qtr2;qtr3;qtr4

2019;3000;3200;;3500

2020;2900;-9999;;

2021;1800;2100;1900;1950

2021;1800;2100;1900;1950

2022;1850;-9999;2900;2550

2023;3400;3200;;

(i) Load the file sales.csv into a dataframe.

(ii) Fill all the NaN values with 0.

(iii) Remove the duplicate rows from the original dataframe.

(iv) Replace all the -9999 values with a 0.

(v) Print the average sales made during qtr1 across all the years.

(b) Differentiate between stack() and unstack() methods. Give examples. (5)

(a) Write the Python statements to do the following using numpy package : (10)

(i) Create an array a = [2,4,5,6,11,12,34,67, 75,60]

(ii) Write numpy functions to find its minimum and maximum value.

(iii) Give output: print (a[2:9]) and print (a[2:])

(iv) Give output: print (a[:7]) and print (a[:])

(v) Differentiate between rand() and nrand()

function using appropriate example.

(b) Write the Python statements to create a dataframe tempdf for the given dataset and then plot a scatter chart with X1 on x-axis and X2 on y-axis. The points should be plotted in red colour. (5)

X1	X2
-2	3
1	1
1	2
3	4

6. (a) Write the Python statements to do the following using numpy package: (10)

(i) Create a 3×3 matrix with values ranging from 12 to 20

(ii) Create a 5×5 identity matrix.

(iii) Create a 2×4 matrix filled with 1s

(iv) Create a 3×2 matrix filled with random values

(v) Change a 4×5 matrix into 5×4 matrix

(b) Write python statements to create the given 3×3 matrix using arange() in numpy and print its transpose.

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \end{bmatrix}$$ (5)

7. (a) Write python statements to perform the following operations using matplotlib package for the given dataset. (10)

Days	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Temperature in Degree Celsius	30	31	40	42	35	36	41

(i) Plot histogram of the given dataset.

(ii) Set title as "7 Days Temperature".

(iii) Set X-axis label as " Weekdays"

(iv) Set Y-axis label as " Temperature"

(v) Save the plotted image as "temp-pick.jpg"

(b) Differentiate between stacked bars and heat map. Also write python commands to plot them. (5)