Give the output of following commands :

(i) Score[['Name', 'Class']]

(ii) Score[Score['Class'] ==1] ['Name']

(iii) Score[Score['Score3'] < 80]

(iv) Score['Class'].value_counts().sort_index()

(v) Score.sum(axis="columns")

Write a function diff to compute the difference between the maximum and minimum of each column of dataframe Score and apply it to dataframe Score. (10)

[This question paper contains 12 printed pages.]

Your Roll No..............

Sr. No. of Question Paper : 6060 **H**

Unique Paper Code : 2344001201

Name of the Paper : Data Analysis and Visualization Using Python

Name of the Course : **Computer Science: Generic Elective (G.E.)**

(NEP-UGCF-2022)

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 **4** 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.

RAMANUJAN COLLEGE LIBRARY
RAMANUJAN COLLEGE
EMPOWER TRANSFORM DISCOVER
UNIVERSITY OF DELHI
Kalkaji, New Delhi-110019

P.T.O.

Section A

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

1. (a) Consider the following numpy arrays : (5)

arr1 = np.array([[4,3,2], [1,9,6]])

arr2 = np.array([[3,7,5], [2,9,8], [5,1,6]])

Give the output of the following commands :

(i) arr2 [1] [1]

(ii) arrl [: 2, –1]

(iii) arrl * 3

(iv) arrl > 5

(v) arr2 [2] =4

(b) List and describe different types of sampling of data. (5)

(c) Consider the Series object Company having 'Company_Name' as index and Profit (in Crores) as values: (3)

(ii) Assign rank in descending order.

(iii) Retrieve all values except NaN. (6)

7. (a) Write Numpy commands to perform the following operations on array num : (5)

(i) Create an array num containing values from 31 to 4 6.

(ii) Convert datatype of array num to floating type data.

(iii) Reshape array num to an array of size 4×4.

(iv) Replace the diagonal elements of array num to 0.

(v) To create an array of 1's with the same shape and type as the given array num.

(b) Consider the dataframe Score given below :

Name	Class	Score1	Score2	Score3
A	1	85	90	88
B	2	74	86	80
C	1	83	71	92
D	2	64	68	73
E	2	77	62	72
F	1	90	87	92

6. (a) Consider the pandas series s2 = pd.Series ([2, 4, 6, 8, 10, 12]).

Write python code to plot cumulative sum of s2. Set the x limit to [0, 10] and y limit to [0,50]. Set the style of line graph to dot(.) pattern and marker to star shape. Set appropriate values for xticks and yticks. (5)

(b) Consider dataframe df given below : (4)

Number	**One**	**Two**	**Three**
State			
Ohio	0	1	2
Colorado	3	4	5

Provide the output of following commands.

(i) df.stack()

(ii) df.unstack(level=0)

(c) Consider the series a given below and write commands to perform the following operations :

a = pd.Series([6,np.nan,–4,np.nan,3,8,np.nan,5])

(i) Sort the values and keep NaN in initial positions.

Company_Name	**Profit**
TCS	350
Reliance	200
L&T	800
Wipro	150

Write the python commands to perform the following operations :

(i) To display the Company_Name having profit > 250.

(ii) To display the index.

(iii) To assign name 'Company_Name' to index.

(d) Write a python code to draw a scatter plot comparing monthly revenue (in Crores) and monthly expenditure (in Crores) of a company for year 2021. (5)

revenue = [581, 684, 739, 563, 856, 716, 589, 820, 792, 695, 770, 812]

expenditure = [631, 545, 435, 532, 688, 540, 485, 679, 709, 535] .

Import necessary libraries. Assign the title of the plot as ‘Revenue vs Expenditure’ and label y-axis as ‘Expenditure’. Assign red color to ‘Expenditure’ data points and green color to ‘Revenue’ data points.

(e) Define correlation and covariance. Outline the difference between the two. (5)

(f) Create a DataFrame having five rows and four columns and populate it with random values in the range 1 to 100. Set the index of the rows as [‘L’, ‘M’, ‘N’, ‘O’, ‘P’] and column indexes as [‘Col1’, ‘Col2’, ‘Col3’, ‘Col4’]. (4)

(g) Give the output of the following code : (3)

```
import Pandas as pd
s1 = pd.Series([‘Certificate’, ‘Bachelor’,
‘Master’, ‘Doctorate’],index = [2,4,6,8])
s1.reindex(range(10), method = ‘ffill’)
print(s1)
```

5. (a) Define categorical and interval data. Give example of each. (4)

(b) What is hierarchical Indexing? Why do we use hierarchical indexing in pandas? Which pandas feature enables you to have multiple index levels on an axis? Give an example of hierarchical indexing. (6)

(c) Consider the data fame df 2 given below : (5)

	Name	Age
0	Rohit	10
1	Amit	13
2	Ankur	12

Write python commands to perform following operations :

(i) Create a new object df 3 by reindexing df 2 row index as [0, 1, 2, 3, 4] and column index as [‘x’, ‘y’].

(ii) Delete the entry of ‘Amit’ from df3.

(iii) Rename index of df 2 as [1, 2, 3].

(iv) Check if the entry ‘Rohit’ exists in df 2.

(v) Modify Age of ‘Ankur’ to 15 usings loc command.

(i) Read the file test.csv into a DataFrame data.

(ii) Print the first 10 rows of data.

(iii) Display the 5 summary statistics for each column of data.

(iv) Remove the rows with all null values.

(v) Identify duplicate values in data.

(c) Consider the following piece of code and give the output : (5)

```
import pandas as pd
df1 = pd.DataFrame({‘id’ : [1,3,6,7], ‘val’ : [‘a’, ‘b’, ‘c’, ‘d’]})
df2 = pd.DataFrame({‘id’ : [1,2,3,5,6,8], ‘val’ : [‘p’, ‘q’, ‘r’, ‘s’, ‘t’, ‘u’]})
df3 = pd.merge(df1, df2, on = ‘id’, how = ‘outer’)
print(df3)
```

How many NaN values are there in the data frame df 3? Write pandas command to replace NaN with the last known valid value in df3.

Section B

2. (a) Consider the following DataFrame House_Rent given below : (10)

Rooms	Area	Bathroom	Furnishing_Status	Rent
2	1100	2	Unfurnished	10000
2	800	1	Semi-Furnished	16000
2	900	2	Furnished	22000
1	250	1	Unfurnished	5000
2	1000	2	Semi-Furnished	23000
3	1200	2	Semi-Furnished	25000
1	400	1	Unfurnished	7000
1	250	1	Furnished	6500
1	375	1	Unfurnished	6000
3	900	2	Unfurnished	8500
3	1286	2	Furnished	35000
2	600	1	Semi-Furnished	8000
2	800	1	Unfurnished	12000

Write python commands to perform the following operations :

(i) Find the index of house with maximum rent.

(ii) Sort the dataframe House_Rent on “Area”.

(iii) Calculate total Area and total rent.

(iv) Compute the count of houses having rooms 1, 2, 3 etc.

(v) Create a new DataFrame df having a hierarchical index on columns "Rooms" and "Furnishing Status".

(b) Refer to DataFrame House_Rent given in question 2(a), Write a python code to plot a bar plot displaying no of Furnished, Unfurnished, Semi-Furnished houses. Import appropriate libraries. The title of graph should be "House Data". Give appropriate labels for x and y axis. Save the figure with name "house.jpg". (5)

3. (a) Write python code to create a numpy array a1 containing 50 floating points values in the range 0 to 1. Put the data of numpy array a1 into 5 bins. Set the precision to 4. Assign names to bins as ['Small', 'Medium', 'Large', 'x-Large', 'xx-Large']. (5)

(b) Write a numpy code to create a 3D array a3 of size 4 × 5 × 3 of random numbers in range 1 to 60 and swap axis 1 with axis 2. Identify the number of matrices in the array a 3, dimension of a matrix in array a3 and the datatype of array a3. (5)

(c) Consider numpy array arr given below : (5)

```
arr = [ [0, 1, 2, 3],
        [4, 5, 6, 7],
        [8, 9, 10, 11],
        [12, 13, 14, 15],
        [16, 17, 18, 19],
        [20, 21, 22, 23] ]
```

Write numpy commands to retrieve following elements :

(i) (1, 4), (3, 1), (5, 0), and (2, 3)

(ii) Retrieve 0, 2, 4 rows (use positive index)

(iii) Retrieve 1, 3, 5 rows (use negative index)

(iv) Retrieve values greater than 10

(v) Retrieve rows 1 to 4.

4. (a) What is data wrangling? Identify the possible issues that can arise in data wrangling process? (5)

(b) Consider a csv file test.csv having 3 columns and 50 rows. Write python command to perform following operations : (5)

P.T.O.