[This question paper contains 8 printed pages.]

Sr. No. of Question Paper : 1669   C

Unique Paper Code   : 42343307

Name of the Paper   : Data Analysis using Python Programming (SEC)

Name of the Course   : **B.Sc. (P) Physical Science with Computer Science (LOCF)**

Semester   : III

Duration : 2 Hours   Maximum Marks : 25

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 3** questions from **Section B**.

4. Parts of a question must be answered together.

P.T.O.

SECTION A

(Compulsory)

1. Consider the following Arrays : (5)

arr1 = np.array([[1, 1, 0], [9, 7, 8], [6, 8, 4]])

arr2 = np.array([[[5, 2, 1], [2, 1, 8]], [[1, 2, 3], [4, 5, 6]]])

Find output of the following :

(a) arr2[0]

(b) arr2[1] [1]

(c) arr1[:2, -1]

(d) arr1[:,2:]

(e) arr1[-1,:]

2. Consider the following Data Frame : (1+2+1+1=5)

Data = {'State': ['Delhi', 'Mumbai', 'Bangalore', 'Delhi', 'Kolkata', 'Pune']}

'Year': [2001, 1992, 1995, 2005, 1992, 2001]

'Population': [5.2, 1.8, 1.5, 7.8, 1.2, 2.5]

  }

df = pd.DataFrame(Data)

Answer the following :

(a) Rename the indexes as "One", "Two", "Three", "Four", Five", "Six"

(b) Add another column in the Data Frame "Extra" having values True corresponding to state 'Delhi' else False.

(c) Name all the states with Population more than 2.

(d) Drop rows 'Three' and 'Six'.

SECTION B

(Attempt any 3 questions)

(Parts of a question must be answered together.)

3. Consider the Dataset given below : (1+2+1+1=5)

Name	Nationality	Sex	Age	Handedness
Neha	India	Female	24	Left
Rahul	India	Male	28	Right
Katherine	USA	Female	32	Right
Zenitsu	Japan	Male	16	Left
Linda	USA	Female	28	Left
Mikasa	Japan	Female	22	Right
Aryan	India	Male	42	Right
Gabriel	USA	Female	32	Right
Kiran	India	Female	20	Left
Shishir	USA	Male	40	Right

Answer the following using Contingency Table:

(a) Find frequency distribution between Sex and Handedness

(b) Find relationship between Sex, Nationality, and Handedness. Also, display total.

(c) Normalize the table in part 'b' over each column.

(d) Find the average of age in part 'a'.

4.   (a) Explain a few ways to handle missing values.

(2)

(b) Consider the following DataFrame :   (3)

df = pd.DataFrame([('bird', 2, 2),

('mammal', 4, np.nan),

('arthropod', 8,0),

('bird', 2, np.nan)],

index=('falcon', 'horse', 'spider', 'ostrich'),

columns=('species', 'legs', 'wings'))

What will be the output of the following statements?

(i) df.mode()

(ii) df.mode(drop_NA = False)

(iii) df.mode(axis = 'columns', numeric_only = True)

5. Consider the following Series :   (5)

SR = pd.Series([8, -2, 6, 4, 3, 0, 6, -1], index = [0, 3, 4, 6, 7, 10, 12, 14]) Find the output of the following :

(a) SR.rank()

(b) SR.rank(method = 'first')

(c) SR.rank(ascending = False, method = 'min')

(d) SR.reindex(range(6), method = 'ffill')

(e) SR.replace({-2: np.nan, 6: 0})

6. Consider the following (df):   (1+2+2=5)

S. No	Restaurant Name	Location	User Rating	User Review
1	Sky Bar	Hauz Khas	3.4	Good
2	Diggin	Chanakyapuri	4.2	Very Good
3	Cafe Dori	Nehru Place	4.0	Very Good
4	Art of Dumplings	Connaught Place	4.6	Excellent
5	Flying Saucer	Connaught Place	4.0	Very Good
6	Imperfecto	Hauz Khas	4.2	Very Good
7	Sutra	Chanakyapuri	4.4	Excellent
8	Dunkin Donuts	Nehru Place	3.8	Good
9	Wrangler	Connaught Place	4.8	Excellent
10	Cha Bar	Connaught Place	4.2	Very Good

(a) Group the Restaurants according to their locations and find the restaurant with the highest rating.

✗ (b) Make a world cloud of column User Review. Also use 'Bad' and 'Very Bad' for Stopwords.

(c) Plot Box Plot of User Rating in Data Frame df.

7. Consider the following list : (5)

$x1 = [32, 21, 65, 89, 33, 23, 22, 40, 21, 32, 56, 45, 44, 38, 52]$

(a) From 'x1' create an ndarray 'x2' with 5 rows and 3 columns.

(b) Find Mean, Variance, Median of data in x2 along the rows.

(c) Find cumulative sum of rows in 'x2'.

(d) Sort the data in 'x2' along the columns.

(e) Replace values greater than 40 in 'x2' with 0 using 'where' function.