

[This question paper contains 12 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 1095

C

Unique Paper Code : 32347507

Name of the Paper : Data Analysis and Visualisation

Name of the Course : **B.Sc. (Hons.) Computer Science**

Semester : V

Duration : 3 Hours

Maximum Marks : 75

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Question No. 1 is compulsory.
3. Attempt any **four** questions out of Q. 2 to Q. 7.
4. Parts of a question must be answered together.

1. (a) Give output of the following code.

(i) `import pandas as pd`

(2)

P.T.O.

```
obj3 = pd.Series(['wow', 'good', 'great'],
                 index=[0, 2, 4])
```

```
obj3.reindex(range(6), method='ffill')
```

```
obj3
```

```
(ii) matrix = [[j for j in range(3)] for i in range(3)]
print(matrix) (2)
```

```
(iii) import pandas as pd
df=pd.DataFrame([1,1,1], [2,2,2], [1,2,1],
[2,1,1],index=['one','two','three','four'],
columns=pd.Index(['A','B','C'],name='MyPlot'))
```

Give the output for df.plot.bar(). (2)

(b) What is a pivot table? Give one example. (2)

(c) Provide the output of following codes. (3)

Given the value of string object s=3.1456 and

```
c="This is a long string
that spans multiple lines"
```

```
(i) fval= float(s)
type(fval)
```

```
(ii) bool(s)
```

```
(iii) c.count('\n')
```

(d) Consider a list seq= [1, 2, 0, 4, 6, 5, 2, 1]. Write a code to find the sum of elements of the value till element 5. (2)

(e) Consider the given arr = [1,2,8,9,3,4,7,5,10,6]. What will be the resulting array if these operations are performed arr[2:5], arr[-5: -1] and arr[::2]. (3)

(f) Create a dataframe with four rows and three columns and populate it with random values. Index of the rows are 'Utah', 'Ohio', 'Texas', 'Oregon' and column indexes are 'b', 'd', 'e'. Write a lambda function to compute the difference between the maximum and minimum of each column. (3)

(g) Create an array *num* of size 2×3 filled with all zeros then insert [[1,2,3], [4,5,6]] into array. Identify the shape of the array *num*. (3)

(h) Write a code to read a CSV file with new delimiter as ';' and line terminator as '\n'. (3)

(i) Consider following piece of code and give the output. (3)

```
import pandas as pd
```

```
a = pd.DataFrame({'id': [1, 2, 9, 10],  
                  'val': ['a', 'b', 'c', 'd']})
```

```
b = pd.DataFrame({'id': [1, 7, 10, 12, 13, 7],  
                  'val': ['p', 'q', 'r', 's', 't', 'u']})
```

```
c = pd.merge(a, b, on='id', how='right')
```

(i) How many 'NaN' values are in the dataframe 'c'?

(ii) Drop duplicate values from dataframe 'b' and keep the last duplicated value.

(j) Generate DateTimeIndex of length 20 where each index will be Tuesday of the third week of a month starting from 10-Jan-2022. (3)

(k) Consider dataframe df (4)

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.DataFrame({'key': ['a', 'b', 'c'] * 4,  
                  'value': np.arange(12.0)})
```

What will be the output of the following statements?

(i) Print the dataframe *df*.

(ii) Write a code to group the dataframe using key.

(iii) Multiply each group value by 2.

2. (a) Consider a dataframe df as (6)

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.DataFrame({'key1': ['a', 'a', 'b', 'b', 'a'],  
                  'key2': ['one', 'two', 'one', 'two', 'one'],  
                  'data1': np.random.randn(5),  
                  'data2': np.random.randn(5)})
```

Provide the output for the following :

(i) print(df)

(ii) m1 = df['data1'].groupby([df['key1'],
df['key2']]).mean()
print(m1)

(iii) m2 = df['data1'].groupby([df['key1']]).mean()

(iv) pieces = dict(list(df.groupby('key1')))
pieces['b']


```
(v) for(k1,k2),group in
    df.groupby(['key1','key2']):
        print ((k1, k2))
        print(group)
```

(b) Give output of the following code. Justify.

```
(i) val=['foo', 2, [4,2]] (2)
    val[2] = (5, 4)
    print(val)
```

```
(ii) var=(3, 5, (4,5)) (2)
     var[1] = 'two'
     .print(var)
```

3. (a) Given the following list of strings (5)

```
List1 = ['Amazon', 'Amazing Amazon', 'Apple',
'Microsoft', 'Apple is good for health', 'I like
Microsoft'].
```

Using 'List1', generate the following dictionary 'Anydict' where key is the count of words in a string and value is the list of strings having that

```
count. Anydict={1:['Amazon', 'Apple', 'Microsoft'],
2: ['Amazing Amazon'], 3: ['I like Microsoft'],
4: ['Apple is good for health']}.
```

(b) Write a code to read the data from a csv file. Find the number of rows and columns in the data, replace missing values with zero, and remove duplicate values. Write the modified data back to the original file. (5)

4. (a) What is the use of generator function? Write a generator function to print square of first n natural numbers where n is user input. (4)

(b) Write a code program to draw a scatter plot comparing marks of Mathematics= [88, 92, 80, 89, 100, 80, 60, 100, 80, 34] and Science = [35, 79, 79, 48, 100, 88, 32, 45, 20, 30] subjects.

Import the necessary libraries.

Title the plot as 'Marks Comparison' and label y-axis as 'Marks Scored'.

Assign red color to mathematics marks points and blue color to science marks points. (6)

5. (a) Consider the following data frame Family containing a family name, gender of the family member and her/his monthly income and expenditure in each record.

Name	Gender	Monthly Income	Expenditure
Shahin	Male	114000.00	58000.00
Vimal	Male	65000.00	32000.00
Vimala	Female	69500.00	38500.00
Vimala	Female	155000.00	70000.00
Karan	Male	103000.00	52000.00
Shahin	Male	55000.00	18000.00
Seema	Female	112400.00	60000.00
Seema	Female	81030.00	25000.00
Vimal	Male	71900.00	30000.00

- (i) Find correlation between *Monthly Income* and *Expenditure*. (1)
- (ii) Use map function to convert each value of *Name* into uppercase. (2)
- (iii) Create a new data frame Info having a hierarchical index on columns *Name* and *Gender*. (2)

- (b) Consider the data array = [0.9296, 0.3164, 0.1839, 0.2046, 0.5677, 0.5955, 0.9645, 0.6532, 0.7489, 0.6536] of 10 floating-point values. Write code for following :

- (i) Create 5 bins of the array using the cut method. (1)
- (ii) Create 5 bins of the array using the qcut method. (1)
- (iii) Create 5 bins of the array with precision = 2 using cut method. Also explain the usage of parameter precision. (3)

6. (a) Consider the following code :

```
import pandas as pd
left = pd.DataFrame({'key1': ['foo', 'foo', 'bar'],
                     'key2': ['one', 'two', 'one'], 'lval': [1,2,3]})
right = pd.DataFrame({'key1': ['foo', 'foo', 'bar',
                              'bar'], 'key2': ['one', 'one', 'one', 'two'],
                      'rval': [4,5,6,7]})
```

Provide output of the following :

(i) `pd.merge (left, right, on=['key1'])` (2)

(ii) `prop_cumsum=left.sort_values (by='key2',
ascending=False).1val.cumsum()
print(prop_cumsum)` (2)

(iii) `left.append (right)` (2)

(b) Consider a data given below :

EMP ID	EMP NAME	SALARY
1	Satish	5000
2	Vani	7500
3	Ramesh	10000
4	Rajesh	8000
5	Virat	9500

Write a code for the following :

(i) Create a dataframe for the above data. (2)

(ii) Print elements of 2nd to 4th column of 3rd to 5th row. (1)

(iii) Print elements of all the columns for first two rows. (1)

7. (a) Consider the code given below :

```
import pandas as pd
from datetime import datetime
dates = [datetime(2011,1,2),datetime(2011,1,5),
datetime(2011,1,7),datetime(2011,1,8),
datetime(2011,1,10),datetime(2011,1,12)]
ts = pd.Series(np.random.randn(6), index=dates)
```

Provide output for the following code:

(i) `print (ts)` (1)

(ii) `print(ts + ts[::-1])` (1)

(iii) `print (ts.index[0])` (1)

(b) Write a code to convert string of date '2022-10-20' to string of date '20/10/2022'. (3)

(c) Provide output of the following code : (4)

```
rng=pd.date_range('2010-01-01',periods=12,freq='T')
```



```
ts= pd.Series(np.arange(12), indexing=rng)
print(ts)
print(ts.resample('5min', closed= 'right').sum())
print(ts.resample('5min', closed= 'right', label=
'right', loffset= '-1s').sum())
print(ts.resample('5min').ohlc())
```