Reg. No.										

MANIPAL ACADEMY OF HIGHER EDUCATION, MANIPAL

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL

THIRD SEMESTER MASTER OF SCIENCE – M.Sc. (INFORMATION SCIENCE)

DEGREE EXAMINATION – NOVEMBER 2021

SUBJECT: MIS 609.8 (ELECTIVE 2) – PRINCIPLES OF DATA VISUALIZATION

Monday, November 29, 2021

Time: 10.00 – 13.00 Hrs. Max.Marks:100

All questions carry 10 marks each

1. Write short notes on the following (2x5= 10 Marks)
   a) Discuss basic terms used in web scrapping. b). Briefly discuss type of parsers

2. Name and explain 4 WEB SCRAPING TOOLS? Discuss the steps involved in Web scrapping (2x5= 10 Marks)

3. Briefly explain Analysis of HTML Tags using BeautifulSoup (10 Marks)

4. Consider L(r) = {a}; L(s)={b}. Explain What do the following regular expression represent? (2.5x4=10 Marks)
   i). r* ii) (r s)* iii) (r|s)* iv) (r|ss)*

5. Write a Selenium script to extract summary information from the search "test" results in web page https://techwithin.net (10 Marks)

6. Discuss whether python is an interpreter or compiled language? List and explain string methods in Python? Explain with an example the creation of Mutable and immutable objects (3+4+3=10 Marks)

7. Consider the following in python (2x5=10 Marks)

a) Write output of the following statement in python

In [11]: obj = pd.Series([4, 7, -5, 3])

In [12]: obj

b) Write python program snippet to create a DataSeries from a dictionary {'Ohio': 35000, 'Texas': 71000, 'Oregon': 16000, 'Utah': 5000} and display the object created.

c) Write python program snippet to create a DataFrame from a dictionary given below of equal length that of NumPy array or a list and show the output of the frame created.

```
{'state': ['Ohio', 'Ohio', 'Ohio', 'Nevada', 'Nevada', 'Nevada'],
 'year': [2000, 2001, 2002, 2001, 2002, 2003],
 'pop': [1.5, 1.7, 3.6, 2.4, 2.9, 3.2]}
```

d) Write python program snippet to create DataFrame containing the data

```
(data, columns=['year', 'state', 'pop', 'debt'],
 index=['one', 'two', 'three', 'four',
        'five', 'six'])
```

e) Display the output of this DataFrame and give the commands to print the columns "state", "year", and show the output of frame2.loc['three']

8. Explain with examples Control Flow - for loops, while loops in python (10 Marks)

9. With a suitable dataset and a code snippet (use seaborn) describe four types of plots a) Distribution Plots b) Joint Regression, Histogram, KDE Plots c) Pair Plot d) Bar Plots used to analyze data (2.5x4=10 Marks)

10. Write Python program to create the following dataset as Data Frame "df" and display the frame created (2x5= 10 Marks)

a) ({'key1' : ['a', 'a', 'b', 'b', 'a'],

....: 'key2' : ['one', 'two', 'one', 'two', 'one'],

....: 'data1' : np.random.randn(5),

....: 'data2' : np.random.randn(5)})

a) Give the output of

grouped = df['data1'].groupby(df['key1'])

grouped

b) Discuss mean, unstack operations with the above data set

******