

Question Paper

Exam Date & Time: 12-Feb-2021 (10:00 AM - 01:15 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL
FIRST SEMESTER MASTER OF ENGINEERING- ME (BIG DATA AND DATA ANALYTICS / HEALTHCARE DATA
ANALYTICS / MACHINE LEARNING) DEGREE EXAMINATION - FEBRUARY 2021

Principles of Data Visualization -Elective - 1 [BDA-622]

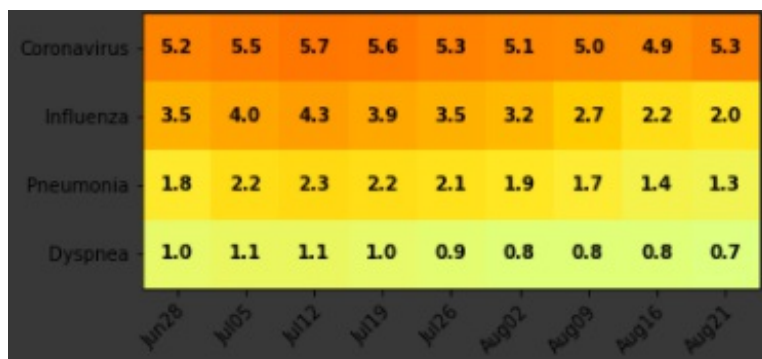
Marks: 100

Duration: 180 mins.

FRIDAY, FEBRUARY 12, 2021

Answer all the questions.

- 1) Explain five types of web scraping techniques. (TLO: 1.1) (2x5 Marks) (10)
- 2) Explain various steps involved in Web scraping framework and illustrate it using Beautiful Soup. (URI: www.quotestoscrape.com) (TLO: 1.2) (10 Marks) (10)
- 3) Illustrate scrapy framework. Write a spyder script to extract information about a product having attributes: name, cost, manufacturer, rating and store it in a Jason file. URI: www.shopping-TV.com. Note: website is not allowed to scrape using bots or script, mention the required steps to overcome this problem. (TLO: 1.2) (10 Marks) (10)
- 4) Outline the importance of heatmap. Write a python script for heatmap Figure 1. (10)



(TLO: 3.2) (10 Marks)

- 5) With a suitable dataset and a code snippet (use seaborn) describe four types of plots used to analyze categorical data. (TLO: 3.1) (2.5x4 Marks) (10)
- 6) What is Exploratory Data Analysis (EDA)? Illustrate various steps performed in EDA with a reasoning for the same? Specify at least 5 EDA library. (TLO 2.3) (2+6+2 Marks) (10)
- 7) Write a program to create 3x4 numpy array, delete 3rd column and insert new column in the 3rd column. Print the intermediate results. (TLO: 2.1) (10 Marks) (10)
- 8) Write a regular expression for matching passwords under following consideration: (10)
 - a. 6 to 12 characters in length
 - b. Must have at least one uppercase letter
 - c. Must have at least one lower case letter
 - d. Must have at least one digit

e. Should contain other characters (TLO 1.1) (2x5 Marks)

9) Explain categories of explanatory visualizations based on the relationships between the three necessary players. (TLO 3.1) (10 Marks) (10)

10) Assume that "power.csv" contains information about power production. Fields in csv are "Type of fuel", "Country", "Place", "Mega Watt", and "Cost per Unit". Write the script using pandas, to perform following tasks. (10)

a. List different types of fuels used.

b. Which country generates more power using 'Coal' as fuel used.

c. Which country generates maximum power (including all types of fuels)

d. List the place name that produces maximum power. (TLO 2.1) (2.5x4 Marks)

-----End-----