

23U258

(Pages: 2)

Name:

Reg.No:

SECOND SEMESTER B.Voc. DEGREE EXAMINATION, APRIL 2024

(CBCSS - UG)

(Regular/Supplementary/Improvement)

CC21U SDC2 DS06 - DATA SCIENCE WITH PYTHON

(Information Technology)

(2021 Admission onwards)

Time : 2.5 Hours

Maximum : 80 Marks

Credit : 4

Part A (Short answer questions)

Answer *all* questions. Each question carries 2 marks.

1. What is big data?
2. Explain why data cleaning is important in analysis.
3. What is stratified sampling?
4. Explain how the K - Means algorithm works.
5. List the skill set needed for a data scientist
6. What is Fancy Indexing?
7. How do you create a 3D array?
8. What is broadcasting for Numpy arrays?
9. What is the use of iloc indexer attribute?
10. What is the use of how parameter?
11. What is the use of Numexpr library?
12. What is the use of get_dummies()?
13. What is the use of plt.style directive?
14. What is a scatter plot?
15. What is the difference between plt.scatter and plt.plot?

(Ceiling: 25 Marks)

Part B (Paragraph questions)

Answer *all* questions. Each question carries 5 marks.

16. Explain Exploratory Data Analysis.

17. Explain the creation of arrays from python lists.
18. Define structured arrays.
19. What are Pandas Groupby? What are the steps involved in groupby operation?
20. How to concatenate two or more pandas' data frames? Explain.
21. What is the difference between merge, join and concatenate? Explain.
22. Mention the different types of Data Structures in Pandas.
23. Explain about the hierarchical indexing.

(Ceiling: 35 Marks)

Part C (Essay questions)

Answer any *two* questions. Each question carries 10 marks.

24. Explain k-Nearest Neighbors (k-NN) algorithm with suitable example.
25. Explain k-means clustering algorithm.
26. Explain the differences between univariate, bivariate and multivariate analysis with suitable example.
27. What is array and explain the different ways to create a numpy array?

(2 × 10 = 20 Marks)
