21CSS303T 08 - Data scienece QUstion paper

Computer science (SRM University)

Scan to open on Studocu

B.Tech/ M.Tech (Integrated) DEGREE EXAMINATION, JULY 2024

Sixth Semester

21CSS303T – DATA SCIENCE

(For the candidates admitted from the academic year 2022-2023 onwards)

Note:
(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
(ii) **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours

Max. Marks: 75

PART – A (20 × 1 = 20Marks)
Answer ALL Questions

		Marks	BL	CO	PO

1. In pandas, which object is used for representing one dimensional labelled data? — Marks 1, BL 2, CO 1, PO 1
   - (A) Dataframe
   - (B) Array
   - (C) Series
   - (D) Index

2. What does the 'eye( )' function in NumPy do? — Marks 1, BL 1, CO 1, PO 1
   - (A) Generates an identity matrix
   - (B) Performs element-wise multiplication
   - (C) Calculates the dot product of two arrays
   - (D) Splits an array into sub_arrays

3. What is the purpose of reindexing in pandas? — Marks 1, BL 1, CO 1, PO 1
   - (A) To sort the data in ascending order
   - (B) To reset the index of a dataframe
   - (C) To align data from multiple series of dataframes
   - (D) To drop missing values from a dataframe

4. Which of the following approach should be used to ask Data Analysis question? — Marks 1, BL 2, CO 1, PO 1
   - (A) Find only one solution for particular problem
   - (B) Find out the question which is to be answered
   - (C) Find out answer from dataset without asking question
   - (D) Finding out solution from dataset by asking question

5. What does the term "Data Alignment" refer to in Pandas? — Marks 1, BL 1, CO 1, PO 1
   - (A) Matching data based on common indices
   - (B) Combing data from different sources
   - (C) Arranging data in ascending order
   - (D) Formatting data for visualization

6. Which of the following input can be accepted by dataframe? — Marks 1, BL 2, CO 1, PO 1
   - (A) Structured ndarray
   - (B) Series
   - (C) Both (A) and (B)
   - (D) Unstructured view

7. Which of the following step is performed by data scientist after acquiring the data?
   (A) Data cleaning (B) Data integration
   (C) Data replication (D) Data modelling

8. We write _____ infront of mean to let python know that we want to activate the mean function from NumPy library.
   (A) npm (B) np
   (C) ng (D) ngm

9. Which types of algorithms is well-suited for handling large data without needing to load the entire dataset into memory?
   (A) Parallel algorithms (B) Sequential algorithms
   (C) Online learning algorithms (D) Offline learning algorithms

10. What is the advantage of using online learning algorithms for handling large data?
    (A) They require loading entire dataset into memory (B) They support parallelized calculations
    (C) They are only suitable for offline processing (D) They can be trained using one observation at a time without needing to load all data into memory

11. What does data transformation involve in context of data analysis?
    (A) Converting text data into numeric format (B) Removing outliers from dataset
    (C) Deleting missing values (D) Adding noise to the data

12. Which of the following techniques is used for summarizing data in statistical analysis?
    (A) Aggregation (B) Summation
    (C) Stat (D) Amplification

13. In data preprocessing, what is the purpose of binning or classing data?
    (A) To split data into different categories or groups (B) To remove outliers from dataset
    (C) To convert categorial data into numerical data (D) To summarize data using statistical measures

14. What does standardization aim to achieve in data preprocessing?
    (A) To make data distribution more uniform (B) To increase the size of dataset
    (C) To reduce computational complexity (D) To scale the data to a common range

15. Which of the following is a common approach to identify and handle outliers or anomalies in a dataset?
    (A) Replacing them with mean values (B) Ignoring them during analysis
    (C) Detecting and removing them (D) Treating them as missing data

08JF6-21CSS303T

16. What parameter of the plt.scatter ( ) function controls the size of markers in a scatterplot?   1   1   4   1
   (A) marker_size   (B) sizes
   (C) Markers   (D) S

17. Which seaborn function is used to create a heatmap?   1   2   4   1
   (A) seaborn.heatmap ( )   (B) seaborn.plot_heatmap ( )
   (C) seaborn.heat_map ( )   (D) seaborn.plot.heatmap ( )

18. Which parameter of the plt.boxplot ( ) function controls the orientation of box plot?   1   2   4   1
   (A) Vert   (B) Orientation
   (C) Vertical   (D) Horiz

19. In matplotlib, which function is used to save a plot to a file?   1   2   5   1
   (A) plt.save ( )   (B) plt.export ( )
   (C) plt.saveplot ( )   (D) plt.savefig ( )

20. Which python library is commonly used for data visualization?   1   2   5   1
   (A) NumPy   (B) Pandas
   (C) Seaborn   (D) SciPy

PART – B (4 × 10 = 40 Marks)
Answer ANY FOUR Questions

	Marks	BL	CO	PO

21. Illustrating the concept of facets of data and their significance in exploring complex datasets.   10   2   1   1

22. Discuss general techniques and programming tips used to effectively handle large datasets in data science projects.   10   2   2   1

23. Demonstrate the merging process on index labels and implications of merging for data analysis and manipulation tasks with appropriate python code.   10   3   3   1

24. Elucidate the concept of web APIs and provide a python code snippet illustrating how to interact with a web API to retrieve data.   10   2   4   1

25. Discuss common formatting options available for ticks, labels and legends in matplotlib and provide examples of when and how to use them effectively.   10   2   5   1

26. Describe the process of creating a 3D plot for a surface using python and potential application of this visualization technique in various fields.   10   2   5   1

PART – C (1 × 15 = 15 Marks)
Answer ANY ONE Question

	Marks	BL	CO	PO

27. Discuss the complete lifecycle of data science process for e-commerce company to enhance its sales and customer experience through data-driven insights.   15   2   1   1

28.i. Explain various data wrangling techniques by illustrating them with suitable python code.   8   2   4   1

ii. Explain string manipulation with python code example.   7   3   4   1

* * * * *