Subject Code:- ACSAI0512

Roll. No:

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech.

SEM: V - THEORY EXAMINATION (2023 - 2024)

Subject: Data Analytics

Time: 3 Hours   Max. Marks: 100

General Instructions:

IMP: Verify that you have received the question paper with the correct course, code, branch etc.

1. This Question paper comprises of three Sections -A, B, & C. It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.

2. Maximum marks for each question are indicated on right -hand side of each question.

3. Illustrate your answers with neat sketches wherever necessary.

4. Assume suitable data if necessary.

5. Preferably, write the answers in sequential order.

6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

SECTION-A   20

1. Attempt all parts:-

1-a.   A column is a _____ representation of data ? [CO1]   1

    (a)   Horizontal

    (b)   Diagonal

    (c)   Vertical

    (d)   Top

1-b.   Total groups in which data can be characterized is? [CO1]   1

    (a)   1

    (b)   2

    (c)   3

    (d)   4

1-c.   Transactional Data is a information that is captured from? [CO2]   1

    (a)   Transactions

    (b)   Social media

    (c)   Both 1 and 2

    (d)   None of the above

1-d.   The source from which first hand information is collected is known as? [CO2]   1

    (a)   Primary source

    (b)   Secondary source

    (c)   Both 1 and 2

      (d)   None of the above

1-e.   Data Selection is…...........? [CO3]   1

      (a)   The actual discovery phase of a knowledge discovery process

      (b)   The stage of selecting the right data for a KDD process

      (c)   A subject -oriented integrated time-variant non-volatile collection of data in support of management

      (d)   Record oriented classes finding

1-f.   What is the other name one can give to noise in data or noisy data? [CO3]   1

      (a)   Corrupt data

      (b)   Outliers

      (c)   Both Aand B

      (d)   None

1-g.   Time series analysis is helpful in: ? [CO4]   1

      (a)   Rainfall measurements

      (b)   Temperature readings

      (c)   Weather data

      (d)   All of the above

1-h.   Dimensionality reduction visualizes data quickly? [CO4]   1

      (a)   On Large Data

      (b)   On data With Numbers

      (c)   On categorial data

      (d)   All of the Above

1-i.   How do you identify a continuous field in Tableau? [CO5]   1

      (a)   It is identified by blue pill in the visualization

      (b)   It is identified by green pill in the visualization

      (c)   It is preceded by a # symbol in the data window

      (d)   When added to the visualization, it provides distinct values

1-j.   Tableau was introduced in the year of ? [CO5]   1

      (a)   2000

      (b)   2002

      (c)   2003

      (d)   2004

2. Attempt all parts:-

2.a.   What skills do Data Scientists need? [CO1]   2

2.b.   Differentiate between data manipulation and data modification? [CO2]   2

2.c.   What is Sampling? [CO3]   2

2.d.   What type of data is good for PCA? [CO4]   2

2.e.	What is Data Modelling? [CO5]	2

SECTION-B 30

3. Answer any <u>five</u> of the following:-

3-a.	Explain data Analysis. Why python is used for data analysis? [CO1]	6
3-b.	Explain the method to collect and analyse data to use social media to predict the weather condition.? [CO1]	6
3-c.	What are the sources of transactional data? [CO2]	6
3-d.	How do social networks collect data? [CO2]	6
3-e.	What is Binning in noisy data? Explain in detail? [CO3]	6
3-f.	Explain in detail how does python handle missing data and outliers? [CO4]	6
3-g.	Explain in detail about Tableau this code-IF [Container]='Jumbo Box' then 'Big Containers' ? [CO5] ELSEIF [Container]='Jumbo Drum' then 'Big Containers' ELSEIF [Container]='Large Box' then 'Big Containers' ELSEIF [Container]='Medium Box' then 'Midsize Containers' ELSEIF [Container]='Wrap Bag' then 'Midsize Containers' ELSEIF [Container]='Small Box' then 'Little Containers' ELSEIF [Container]='Small Pack' then 'Little Containers'	6

SECTION-C 50

4. Answer any <u>one</u> of the following:-

4-a.	Explain five application areas of data science in detail also explain how Data science is related to AI? [CO1]	10
4-b.	Explain the need of predictive analytics in Data Science? Differentiate between descriptive and prescriptive analytics? [CO1]	10

5. Answer any <u>one</u> of the following:-

5-a.	Explain the term Data Wrangling in Data Analytics, Also explain steps of a Data Analytics Project in detail? [CO2]	10
5-b.	What are the sampling techniques used in a Data Science Project, also describe in detail univariate, bivariate and multivariate analysis? [CO2]	10

6. Answer any <u>one</u> of the following:-

6-a.	Describe the process of binning. How does it help in data visualization and analysis? [CO3]	10
6-b.	Explain the term normal distribution of Data and also explain How can you handle missing values in Data? [CO3]	10

7. Answer any <u>one</u> of the following:-

7-a.	Explain the step by step process to perform univariate analysis for numerical and categorical variables? [CO4]	10
7-b.	Describe PCA in detail and also explain some of the areas where PCA has its applications? [CO4]	10

8. Answer any <u>one</u> of the following:-

8-a.   What is tooltip? How to edit a tooltip and also explain how is overfitting different   10
from underfitting? [CO5]

8-b.   How to use group in calculations field and also explain what is Time Series   10
Analysis? [CO5]

COP . JULY 2024