

Total No. of Questions : 5]

SEAT No. :

PC3251

[6383]-1008

[Total No. of Pages :3

F.Y. M.C.A. (Management)

EC - 11 - 3 : FUNDAMENTAL OF DATA SCIENCE

(2024 Pattern) (Semester- I)

Time : 2½ Hours]

[Max. Marks : 50

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Each question carry equal marks.

Q1) Solve any 2 (two): [10]

- a) Explain the various method of data collection and data cleaning.
- b) Explain the social and economic impact of data science in the field of health care industries.
- c) Describe the data Discretization with an example.
- d) What are the key components of data science.

Q2) Solve any 1 (one): [10]

- a) To find a root of the equation $f(x) = x^3 - x - 1$ using the Bisection Method.
- b) Consider the following matrix A:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$$

- i) Find the inverse of the matrix A.
- ii) Verify your result by demonstrating that : $AA^{-1} = I$ where I is the identity matrix.

Q3) Solve any 2 (two): [10]

- a) Propose a data processing model for the Financial Fraud Detection, detailing how data is collected, clean.
- b) Explore challenges in data processing.
- c) Describe key step involve in data validatin in data science.
- d) Demonstrate the process of data transformation on a given dataset.

P.T.O.

Sales Performance Dataset

Salesperson-ID	Age	Years-of-Experience	Total-Sales (Rs.)	Region	Number-of-Clients
301	25	2	35,000	North	15
302	30	5	50,000	South	25
303	40	10	70,000	West	30
304	28	4	45,000	East	20
305	35	8	60,000	North	28
306	22	1	30,000	South	12
307	45	20	90,000	West	40
308	50	25	1,00,000	East	50
309	38	15	80,000	North	35
310	29	6	55,000	South	22

Q4) Solve any 1 (one):

[10]

- a) Sales data : rows represent products, columns represent months
(Jan to Dec)

Month	Product A	Product B	Product C
1	1,500	1,200	900
2	1,800	1,300	950
3	1,700	1,250	900
4	1,600	1,350	1,000
5	1,900	1,400	1,100
6	2,000	1,450	1,200
7	2,100	1,500	1,150
8	2,200	1,550	1,300
9	2,300	1,600	1,400
10	2,400	1,650	1,450
11	2,500	1,700	1,500
12	2,600	1,750	1,600

For above sales data answer the following.

- Read data as NumPy array.
- Calculate the total sales for each product over the year.
- Determine the average monthly sales for each product.
- Identify the month with the highest sales for each product.
- Find the overall best - selling product for the year.

- b) Analyse customer purchase behaviour using Pandas to derive insights on sales trends and customer preferences.

Given data set : Customer Purchase Data

Transaction ID	Date	Customer-ID	Product-ID	Quantity	Price	Total Sale
1	01-01-2023	C001	P001	2	15	30
2	01-01-2023	C002	P002	1	25	25
3	02-01-2023	C001	P003	3	10	30
4	02-01-2023	C003	P001	1	15	15
5	03-01-2023	C002	P001	5	15	75
6	04-01-2023	C001	P002	2	25	50
7	04-01-2023	C003	P003	4	10	40
8	05-01-2023	C004	P002	1	25	25
9	05-01-2023	C004	P001	2	15	30
10	06-01-2023	C005	P003	1	10	10

For the above given data set give answer of the following:

- Read data as data frame using pandas.
- Calculate the total sales from the dataset.
- Determine the total sales for each product.
- Identify the top 3 customers based on their total spend.
- Identify the product with the highest quantity sold.

Q5) Solve any 1 (one):

[10]

Dataset : Iris Flower

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	3.5	1.4	0.2	Setosa
4.9	3	1.4	0.2	Setosa
3.7	2	1.4	0.2	Setosa
7	3.2	4.7	1.4	Versicolor
6.4	3.2	4.5	1.5	Versicolor
6.3	3.3	6	2.5	Virginica
5.8	2.7	5.1	1.9	Virginica

- Write a python code to draw Boxplot for visualizing above the data frame using Matplotlib.
- Write a python code to draw Pie plot for Setosa species for above the data frame using Matplotlib.

