

S.No. of Question Paper : 6498

L

Unique Paper Code : 2514001202

Name of the Course : B. Sc.(H) Electronics (GE)

Name of the Paper : Data Visualization Techniques

Semester : II

Duration : 03 hours

Maximum Marks : 90

Instruction for Candidates

1. There are seven questions in all, out of which you have to attempt any five questions.
2. All questions carry equal marks.
3. Question no. 1 is compulsory.

1.
 - a. What is Data Visualization? Mention any two objectives of data visualization in data analysis.
 - b. What is Exploratory Data Analysis (EDA)? State any two visualization techniques commonly used in EDA.
 - c. Define Word Cloud. How does the size and frequency of words represent information in a word cloud?
 - d. What is a Box Plot? List any two statistical parameters represented by a box plot.
 - e. Describe the process of data cleaning and preparation. Highlight the techniques used for handling missing data.
 - f. Define Data Story in Tableau. How does it help in communicating data insights effectively to decision makers?
(3 x 6 = 18)
2.
 - a. Explain the concept of Descriptive Analytics. How does it help in understanding and presenting data? Illustrate with the help of a suitable example.
 - b. Write a program code to:
 - i. Create a Pandas dataframe for employee data shown below
 - ii. Set Employee_ID as index and sort the data according to Experience
 - iii. Display a scatter plot for Salary vs. Experience and compute average salary by Department.

Employee_ID	Name	Age	Salary	Department	Experience
101	Asha	25	35000	IT	2
102	Ravi	30	48000	HR	5
103	Neha	28	42000	Finance	4
104	Kiran	35	60000	IT	8
105	Jatin	28	44000	HR	3
106	Priya	32	36000	Finance	1

- c. Differentiate between Histogram and Bar Chart. Plot a histogram for Salary distribution.
(6+6+6=18)

3. a. Explain how Plotly provides more advanced and interactive visualization features as compared to Matplotlib. Discuss at least three major differences between the two libraries in terms of interactivity, visualization capabilities, and usability.
- b. Write a program code to:
- To store the student dataset shown below in a Pandas dataframe
 - Add two new columns 'Total Marks' and 'Average Marks'
 - Add a new row showing mean values of all the numerical columns

Student	Study Hours	Attendance (%)	Physics	Chemistry	Maths
S1	3	70	55	60	58
S2	5	80	65	68	70
S3	7	85	75	78	80
S4	4	75	60	62	65

- c. Write a program code to:
- Display a scatter plot of Average Marks vs. Study Hours
 - Display a bar chart of Total Marks for all students
- (6+6+6=18)
4. a. Explain the Seaborn library used for data visualization in Python. Discuss its key features and advantages over Matplotlib. Also explain any two types of plots that can be created using Seaborn with suitable examples.
- b. What is a Heatmap? Mention two practical applications of heatmaps in data visualization? Explain how heatmaps help in understanding patterns, correlations, and intensity of data values.
- c. The following dataset represents marks obtained by students in three subjects:

Student	Physics	Chemistry	Maths
A	72	65	70
B	80	78	85
C	68	70	72
D	90	88	92

Write a program code to:

- Create a Pandas DataFrame for the given dataset.
 - Generate a heatmap representing marks of students in different subjects.
 - Display annotations and a suitable colormap in the heatmap.
- (6+6+6=18)
5. a. Describe the main components of a violin plot and how it combines the features of box plot and kernel density estimation (KDE). Write a Python code to generate a violin plot for the following dataset: [45, 50, 60, 55, 70, 65, 58, 62, 68].
Briefly interpret what information about the data distribution can be obtained from the

generated violin plot.

- b. Describe at least three application areas where geospatial visualization is widely used and discuss any two Python libraries used for geospatial visualization.
- c. Mention two practical applications where bubble charts are commonly used. Differentiate between Bubble chart and Choropleth Map with respect to data representation, visualization method, and type of data used.

(6+6+6=18)

6.
 - a. Discuss the role of visualization in simplifying complex datasets. Explain how effective visualization improves decision making in business and healthcare sectors. Differentiate between numeric data and categorical data in terms of their analysis and visualization techniques.
 - b. Discuss the key features of Tableau. What is a dashboard? Describe the process of creating a dashboard in Tableau.
 - c. The following dataset shows number of visitors in a museum for one week:

Date	Visitors
01-05-2026	120
02-05-2026	150
03-05-2026	180
04-05-2026	200
05-05-2026	140
06-05-2026	170
07-05-2026	190

Write a program code to:

- i. Rename column 'Visitors' to 'Local Visitors'
- ii. Add two new columns, 'Foreign Visitors' and 'Student Visitors'
- iii. Plot time Series graphs for all the three columns

(6+6+6=18)

7.
 - a. Differentiate between unstructured, semi-structured and structured data in terms of the complexity involved in analysis and visualisation.
 - b. What are Regression plots? Explain how regression plots help in detecting correlation and prediction patterns with the help of a suitable example.
 - c. The following dataset represents monthly electricity consumption of different households in a locality.

Month	Household_A	Household_B	Household_C
Jan	210	190	180
Feb	220	200	195
Mar	250	230	210
Apr	270	245	225
May	300	275	250

Write a program code to perform the following tasks:

- i. Create a Pandas dataframe to store the above dataset.
- ii. Calculate the average electricity consumption for each household.
- iii. Plot a line graph to visualize electricity consumption trend for all households across months.
- iv. Identify the month with the highest electricity consumption for Household_A.

(6+6+6 =18)