

(b) What are Regression Plots? What information can be obtained from such Plots?

(c) The dataset shown below represents an International Movie Database

S.No.	Movie Name	Year of Release	Rating	Duration(In hrs)
1	Movie1	2022	3.9	1.268888889
2	Movie2	2021	3.5	1.218888889
3	Movie3	2022	3.2	2.517222222
4	Movie4	2020	2.8	1.708333333
5	Movie5	2020	2.8	1.423888889
6	Movie5	2021	3.8	1.481388889

Write a program :

- To store the given dataset in a pandas dataframe.
- Group the given dataset according to 'Rating'.
- Apply data aggregation `agg()` function to determine total number of movies released each year. (6+6+6=18)

(1000)

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 6041

H

Unique Paper Code : 2514001202

Name of the Paper : Data Visualization Techniques

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

Semester : II

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

- Write your Roll No. on the top immediately on receipt of this question paper.
 - There are **seven** questions in all, out of which you have to attempt any **five** questions.
 - All** questions carry equal marks.
 - Question no. 1 is compulsory.
-
- (a) What is Raw Data? Mention atleast two data transformation techniques which are used for preparing raw data for visualization.

P.T.O.

- (b) Explain the features of 'pywaffle' library available in python which is used for data visualization.
- (c) Differentiate between Univariate and Multivariate Visualization with reference to the IRIS dataset shown below:

S.No.	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	7	3.2	4.7	1.4	Iris-versicolor
4	6.4	3.2	4.5	1.5	Iris-versicolor
5	6.3	3.3	6	2.5	Iris-virginica
6	5.8	2.7	5.1	1.9	Iris-virginica

- (d) What are 'Heat Maps'? What type of insight into Data do they convey?
- (e) What is the role of 'cloud computing' in Data Visualization? Mention any one cloud data repository which is available.
- (f) How many types of 'Filters' are used in Tableau?
(3×6=18)
2. (a) Explain 'Exploratory Data Analysis (EDA)' using any of Data Visualization Techniques? Discuss the significance of 'scatter plot' in EDA.
- (b) Write a program code :

- (c) The table below shows record of 'Number of Admissions' taking place in a college, monitored across one week from 07-03-2023 to 13-03-2023

Date	Number of Admissions
03-04-2023	20
04-04-2023	60
05-04-2023	35
06-04-2023	70
07-04-2023	100
08-04-2023	55
09-04-2023	120
10-04-2023	70

Write program code to :

- (i) Change column 'Number of Admissions' to 'Admissions in BSc'.
- (ii) Add columns 'Admissions in BA' & 'Admissions in BCom'.
- (iii) Create a Time Series Plots for all three columns.
(6+6+6=18)
7. (a) Mention some of the features of 'Matplotlib' library which makes it very useful for Data Visualization.

- (ii) What kind of a plot will illustrate the relation between 'Date' vs 'Product1' and 'Product2'? (6+6+6=18)

5. (a) Illustrate the significance of a 'Box' plot in Data Visualization.

Use the following Data : 1,24,20,18,25,22,32,30, 29,35,30,28,24,38,50 to explain the various parameters of 'Box' plot.

- (b) What are the various graphical visualization techniques / plots which are used for Geospatial Data Visualization? Discuss the use of 'Folium' package in Python in Visualizing and Analyzing Geo-Spatial Data.

- (c) Discuss what are 'Choropleth' plots?

(6+6+6=18)

6. (a) Compare the features of the following python libraries which are used for 'Data Visualization' (mention any two points)

(i) Seaborn

(ii) Plotly

- (b) What is Tableau? How can a 'Data Story' be created using Tableau?

- (i) To create a Pandas DataFrame which stores employee database as shown below:

Name	Age	Salary	Department	Status
John	21	30000	Sales	Junior
Reeya	23	40000	Marketing	Junior
Linda	40	100000	Operations	Senior
Vishal	35	75000	Marketing	Senior
Mohan	26	45000	Finance	Junior
Alia	30	60000	Finance	Senior

- (ii) Sort the given dataset according to 'Salary' and display the name of the employee having highest 'Salary'

- (c) Display a Bubble plot for 'Salary' vs. 'Age'. Bubble size should vary according to 'Status' and Bubble colour should indicate 'Department'.

(6+6+6=18)

3. (a) What are the main sources of Big Data? Why is Data Cleaning and Data Preparation required prior to Visualization of Big Data?

- (b) Data related to marks obtained by students in three subjects is tabulated as shown below :

Student Name	No. of Study Hours	Attendance (%)	Physics	Chemistry	Maths
Student 1	2	85	78	80	88
Student 2	1.5	70	67	77	95
Student 3	3	90	85	85	100
Student 4	1	65	96	90	90

Write a program code:

- (i) To create a pandas dataframe to store the above dataset.
 - (ii) To add a new column containing 'Percentage' marks obtained.
 - (iii) To sort the Dataset according to column 'Percentage'.
- (c) After adding 'Percentage' marks column in part (b), write Python code (using Matplotlib Library) to :
- (i) Display a 'LINE' plot of 'Percentage' vs. 'No. of Study Hours'.
 - (ii) Display a 'BAR' plot comparing 'Percentage' marks of all the students. (6+6+6=18)
4. (a) What are the various data pre-processing steps which should be performed in order to transform the given data-set in a form suitable for visualization?
- (b) The table below shows sales record of two products, Product1 and Product2, monitored across one week from 07-03-2023 to 13-03-2023

Date	Product1	Product2
07-03-2023	100	20
08-03-2023	70	78
09-03-2023	NaN	65
10-03-2023	125	100
11-03-2023	65	NaN
12-03-2023	30	80
13-03-2023	50	55

Write a program code to :

- (i) Create a Data Frame to store the sales record data shown in the table above.
 - (ii) Transform all the dates in 'Date' column into format of type 'Month Day, Year'
 - (iii) Detect presence of any missing or invalid entries in the dataset. Replace/Fill missing/invalid entries in 'Product1 and 'Product2' column with value any aggregate function related to that column.
- (c) Write the program code to perform the following functions on the modified dataset.
- (i) Analyze sales of 'Product1' across one week through a Pie Chart