**20DS505/JO**

**Hall Ticket Number:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **III/IV B.Tech (Regular/Supplementary) DEGREE EXAMINATION** | | | |
| **January, 2024** | **Data Science** | | |
| **Fifth Semester** | **Data Handling and Visualization** | | |
| **Time:** Three Hours | | **Maximum: 7**0 Marks | |
| *Answer Question No. 1 Compulsorily.* | | | (14X1 = 14 Marks) |
| *Answer* ***ANY ONE*** *question from each Unit.* | | | (4X14=56 Marks) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. | a) | Draw the classification of digital data. | CO1 | L1 | 1M |
|  | b) | What are the limitations of structured data? | CO1 | L2 | 1M |
|  | c) | Define data visualization. | CO1 | L1 | 1M |
|  | d) | Which chart do you choose to depict the relation between two continuous variables? | CO2 | L4 | 1M |
|  | e) | Write two best practices for drawing a Pie chart. | CO2 | L2 | 1M |
|  | f) | Differentiate between bar chart and grouped bar chart. | CO2 | L2 | 1M |
|  | g) | Name any two plots suitable to visualize a data distribution. | CO2 | L4 | 1M |
|  | h) | What is the difference between pivot and pivotable operations? | CO3 | L2 | 1M |
|  | i) | What is the difference between accessing a column of a dataframe with dot and indexing notations? | CO3 | L2 | 1M |
|  | j) | Differentiate between long-form and wide-form dataframes. | CO3 | L2 | 1M |
|  | k) | What is the role of hue parameter in Seaborn? | CO4 | L1 | 1M |
|  | l) | What are row and col parameters used for in Seaborn? | CO4 | L1 | 1M |
|  | m) | How is Seaborn superior to matplotlib? | CO4 | L2 | 1M |
|  | n) | Write any two merits of plotly API. | CO4 | L1 | 1M |
| **Unit -I** | | | | | |
| 2. | a) | Explain the characteristics and applications of structured, semi-structured and unstructured data in data analysis with examples. | CO1 | L1 | 7M |
|  | b) | What are the differences between Data Visualization and Infographics? Explain with sample sketches. | CO1 | L1 | 7M |
| **(OR)** | | | | | |
| 3. | a) | How are different principles of Gestalt’s theory of visual perception related to Data Visualization? Explain the principles with example charts. | CO1 | L2 | 7M |
|  | b) | How is Data Visualization beneficial in business analytics? Explain each of them with a real time example. | CO1 | L2 | 7M |
| **Unit -II** | | | | | |
| 4. | a) | Code the different ways in which data can be read from different file formats, Dictionary objects and Series objects into a dataframe. | CO2 | L3 | 7M |
|  | b) | What is the applicability of the following charts. Draw sample charts and explain them.  a) Waterfall chart b) Variable width bar chart c) Connected dot chart. | CO2 | L2 | 7M |
| **(OR)** | | | | | |
| 5. | a) | Present appropriate data and sketch grouped Bar Chart, Bubble Chart and a Treemap Map from the data. Describe the charts and analyse the inferences that can be drawn from each. | CO2 | L3 | 7M |
|  | b) | Compute the five number summary for the following data and sketch the resulting Boxplot. Data : 35 42 25 12 8 76 85 56 43 32 17 20 35 40 15. | CO2 | L3 | 7M |
| **Unit -III** | | | | | |
| 6. | a) | Create a sample dataframe. Code and explain four different methods of slicing data from the dataframe with example code, output and description. | CO3 | L3 | 7M |
|  | b) | Create a dataframe with NaN values. Write code to analyse, drop and handle the NaN values with different functions and combination of parameters. List the resulting data after each operation. | CO3 | L3 | 7M |
| **P.T.O**  **20DS505/JO** | | | | | |
| **(OR)** | | | | | |
| 7. | a) | A Superstore Sales data consists of OrderID, Region, SalesAmt columns. Code a snippet without using groupby() function to compute the Region-wise total sales. Code the same functionality with groupby() function. Iterate over the groups and print the rows in each group. | CO3 | L3 | 7M |
|  | b) | Create two dataframes with common columns. Code all possible combinations of merge operations and join operations. Explain the parameters and outputs. | CO3 | L3 | 7M |
| **Unit -IV** | | | | | |
| 8. | a) | Code the appropriate plots using matplot library and datasets to visualize the following. Explain the dataset selected and role of each parameter used and draw the resulting output. a) The whole and part relation b) Distribution of continuous data over intervals c) Relationship between two continuous variables. | CO4 | L4 | 7M |
|  | b) | Code the snippets for different ways of creating subplots using matplotlib. | CO4 | L3 | 7M |
| **(OR)** | | | | | |
| 9. | a) | Choose and Code the appropriate plots using plotly library to display the following. Draw the resulting output and explain the role of each parameter.  a) The relation between three columns of gapminder dataset b) Day-wise total bill of male and female groups in tips data c) Treemap of three levels using tips data. | CO4 | L4 | 7M |
|  | b) | Code the following charts choosing appropriate data columns in the given datasets. Draw the resulting output and explain the role of each parameter.  a) Strip plot using iris dataset b) Box plot in seaborn using tips c) Grouped bar chart in seaborn using superstore sales. | CO4 | L3 | 7M |

****