**20DS505**

**Hall Ticket Number:**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **III/IV B.Tech (Regular) DEGREE EXAMINATION** | | | |
| **February,2023** | **Data Science** | | |
| **Fifth Semester** | **Data Handling & 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) | What effect does scaling have in a visual representation of data? | CO1 | L1 | 1M |
|  | b) | Explain what should be done with suspected or missing data? | CO1 | L2 | 1M |
|  | c) | How can color of an object be defined? | CO1 | L1 | 1M |
|  | d) | When will you use a histogram and when will you use a box plot? | CO2 | L1 | 1M |
|  | e) | Explain what an outlier is? | CO2 | L2 | 1M |
|  | f) | For what type of data is scatter plot usually used for? | CO2 | L3 | 1M |
|  | g) | What is a dirty data record in context of data wrangling? | CO3 | L1 | 1M |
|  | h) | How can we create a copy of the series in Pandas? | CO3 | L1 | 1M |
|  | i) | How to get frequency counts of unique items of a series? | CO3 | L1 | 1M |
|  | j) | How can we convert data Frame into an excel file? | CO3 | L1 | 1M |
|  | k) | Which library would you prefer for plotting in Python language: Seaborn or Matplotlib? | CO4 | L1 | 1M |
|  | l) | What is the use of backward fill? | CO4 | L2 | 1M |
|  | m) | What is the Seaborn function for colouring plots? | CO4 | L2 | 1M |
|  | n) | Illustrate the 2 types of datasets? | CO4 | L2 | 1M |
|  |  | **Unit -I** |  |  |  |
| 2. | a) | Explain the characteristics of structured, semi-structured and unstructured data with examples | CO1 | L1 | 7M |
|  | b) | How do Gestalt principles influence the visual perception? | CO1 | L3 | 7M |
|  |  | **(OR)** |  |  |  |
| 3. | a) | How does an info graphic differ from a data visualization? | CO1 | L2 | 7M |
|  | b) | Explain the benefits of Data Visualization. | CO1 | L1 | 7M |
|  |  | **Unit -II** |  |  |  |
| 4. | a) | Code the different ways in which data can be read from different file types, Dictionary objects and Series objects into a DataFrame. | CO2 | L2 | 7M |
|  | b) | Create a sample XML file and read the data from the file into a DataFrame. | CO2 | L4 | 7M |
|  |  | **(OR)** |  |  |  |
| 5. | a) | Illustrate the difference between count histogram, relative frequency histogram, cumulative frequency histogram and density histogram? | CO2 | L2 | 7M |
|  | b) | What is a JSON? How can we read the JSON Data into a Pandas DataFrame? | CO2 | L1 | 7M |
|  |  | **Unit -III** |  |  |  |
| 6. | a) | Create a sample DataFrame. Explain four different methods of slicing data from the DataFrame with example code, output and description. | CO3 | L4 | 7M |
|  | b) | Create a DataFrame with missing values. Write code to analyse, drop null values and fill the null values with different functions and combination of parameters. List the data after each operation | CO3 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 7. | a) | A Superstore Sales data DataFrame 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 | L1 | 7M |
|  | b) | Explain the syntax and semantics of Pandas pivot\_table() function.  The Stock price DataFrame comprises of day-wise (for a month) indexed rows of StockCode, Open, Close, Min, Max and Volume columns. Write code to create Pivot tables of a) Day-wise Close values for each Stock b)Stock-wise values of total volmes of the month c) Stock-wise mean values of Open, Close, Min and Max. | CO3 | L3 | 7M |
|  |  | **Unit -IV** |  |  |  |
| 8. | a) | Create data for monthly sales of an organization for an year. Write Code to draw a waterfall chart with title, labels, formatting and colours. | CO4 | L2 | 7M |
|  | b) | A Newly Under Graduate joined Students data consisting of Regd. No, Branch, 12th CGPA and I semester CGPA. Select a plot and code the snippet using Seaborn package to draw the plot for comparing 12th CGPA and the I Semester CGPA of each branch | CO4 | L4 | 7M |
|  |  | **(OR)** |  |  |  |
| 9. | a) | Code snippets to draw Line Plot and Subplots using matplotlib with data selected from Iris dataset. Draw the resulting charts with brief descriptions. | CO4 | L3 | 7M |
|  | b) | State the packages available in python for visualizing the data. Explore any 3 of them in detail with a neat diagram for each. | CO4 | L2 | 7M |

****