SRM Institute of Science and Technology
College of Engineering and Technology
School of Computing

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamil Nadu

Academic Year: 2024-25 (ODD)

Mode of Exam
OFFLINE

Test: CLA-T1 **Date: 15-06-2024**
Course Code & Title: 18CSE490T & Big Data Visualization **Duration:** 50 minutes
Year & Sem: IV Year / VII Sem **Max. Marks: 25**

Course Articulation Matrix:

S. No	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	CO1	1	2	2	2	1	-	-	-	2	1	2	2	-	-	-
2	CO2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	CO3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	CO4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	CO5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	CO6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

CO-1: Design and exploring the result with data visualizations.

Part – A (10 x 1 = 10 Marks)
Instructions: Answer all

Q. No	Question	Marks	BL	CO	PO	PI Code
1	Structured data format include: a. Created, provoked, compiled, transactional data b. Created, provoked, captured, transactional data c. Created, transactional, compiled, Experimental data d. Created, submitted, compiled, transactional data i. A and B ii. B and C iii. C and D iv. A and C	1	L1	1	1	1.6.1
2	Methods to deal with outliers are a. Moving outliers to another file b. Removing outliers with scrubbing c. Replacing outliers with appropriate value i. A and B ii. A and C iii. A, B, and C	1	L1	1	1	1.6.1
3	Representation of information perceived as casual, funny, or frivolous is termed as a. Data visualization b. Infographics c. Exploratory visualization d. Visual art	1	L1	1	1	1.6.1
4	"A person's habitual reading patterns will determine their default eye movements over a page, and the order in which they will encounter the various visual elements in your design." – Name the facet of reader's mindset. a. Labels, titles and tags b. Compatibility with reality c. Patterns and Consistency d. Directional Orientation	1	L1	1	1	1.6.1
5	Which graph is best used when there are few relevant fractions and precisions isn't required. a. Bar graph b. Pie graph c. Time series d. Histograms	1	L1	1	1	1.6.1

Part – B (2 x 4 = 8 Marks) Answer any TWO Questions						
Q. No	**Question**	**Marks**	**BL**	**CO**	**PO**	**PI Code**
6	Enumerate the visualization philosophies with respect to the 3V's of big data.	**4**	**L2**	**1**	**1**	**2.6.2**
7	State and explain any four facets of reader's mindset that has to be considered while choosing the visualization encoding.	**4**	**L2**	**1**	**1**	**2.6.2**
8	Justify the role of placement and its proximity in data visualization.	**4**	**L2**	**1**	**1**	**2.6.2**

Part – C (1 x 12 = 12 Marks)						
Q. No	**Question**	**Marks**	**BL**	**CO**	**PO**	**PI Code**
9	(a). Compare and Contrast infographics and data visualization in detail. (b). How exploratory data visualization is different from explanatory data visualization.	**10** **2**	**L2**	**1**	**2** **1**	**2.6.4** **2.6.4**
10	(a). Interpret trinity relationship with exploratory data visualization component with neat sketch. (b). List the key factors to be considered in choosing the appropriate visual encoding.	**10** **2**	**L2**	**1**	**2** **1**	**2.6.3** **2.6.3**

Test: CLA-T2 **Date: 26-06-2024**
Course Code & Title: 18CSE490T & Big Data Visualization **Duration:** 1 hr. 40 mins
Year & Sem: IV Year / VII Sem **Max. Marks: 50**

Course Articulation Matrix:

Course Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3
CO2	2	3	2	2	2	-	-	-	2	1	2	2	-	-	-
CO3	2	2	3	3	2	2	-	-	2	2	3	3	-	-	-

CO-2: Conducting exploratory data analysis using visualization techniques and tools.
CO-3: Visual presentations of data for effective communication.

Part – A (10 x 1 = 10 Marks) Instructions: Answer all							
Q. No	**Question**	**Marks**	**BL**	**CO**	**PO**	**PI Code**	**Mark**
1	Which profiling technique involves mathematical calculations like mean, variance, standard deviation etc in it. i. Definitions and explanations ii. Comparisons and Contrasts iii. Tendencies iv. Dispersion **a. i and iii b. ii and iii c. iii and iv d. i and iv**	**1**	**L1**	**2**	**1**	**1.6.1**	
2	The process of adding an opposite to a data point to see if it perhaps determines different perspective. It is termed as a. Definitions & explanations b. Contrasts c. Tendencies d. Dispersion	**1**	**L1**	**2**	**1**	**1.6.1**	
3	'Adding context' to the data is not an easy job. However, it can be done by the following process a. Data Profiling b. Data munging c. Data Wrangling d. Data Cleaning	**1**	**L1**	**2**	**1**	**1.6.1**	
4	Which function in R programming is similar to the SQL group by command **a.** Table **b.** data.frame **c.** summarise **d.** aggregate	**1**	**L1**	**2**	**1**	**1.6.1**	

5	To execute external commands in Data Manager, we can use a. Java Script b. Visual Basic Script c. Python d. jQuery	**1**	**L1**	**2**	**1**	**1.6.1**	
6,7	Match the following: A. Trifacta – i. Visual Query Language B. Tableau – ii. Design and address big data quality C. R program – iii. Data Wrangler D. Data Manager – iv. Statistical Computation a. A-ii, B-iv, C-i, D-iii b. A-iv, B-ii, C-iii, D-i c. A-ii, B-iii, C-ii, D-iv d. A-iii, B-i, C-iv, D-ii	**2**	**L1**	**3**	**1**	**1.6.1**	
8	What file formats are supported by Trifacta in generating results: a. CSV, JSON, TSV b. CSV, JSON, TDE c. CSV, TSV, TDE d. CSV, DATA, TSV	**1**	**L1**	**3**	**1**	**1.6.1**	
9	"A collection of information in a single place, so you can compare and monitor simultaneously rather than clicking through individual sheets" – Describes?. a. Dashboard b. Workspace c. Containers d. Worksheet	**1**	**L1**	**3**	**1**	**1.6.1**	
10	"The ideal visualization combines science with art. With formatting, dashboards, and stories you can make your data discoveries clearer, more persuasive, and beautiful." The above product documentation describes the tool **a.** Data Manager **b.** R.program **c.** D3.js **d.** Tableau	**1**	**L1**	**3**	**1**	**1.6.1**	

Part – B (4 x 4 = 16 Marks)
Answer any FOUR Questions

Q. No	Question	Marks	BL	CO	PO	PI Code	Mark
11	Explain any four data profiling technique for understanding the dataset before visualization.	**4**	**L2**	**2**	**2**	**2.6.2**	
12	List and elaborate are the different categories of data quality that has to be addressed to enrich the data quality.	**4**	**L2**	**2**	**2**	**2.6.2**	
13	Sketch and Illustrate data profiling process.	**4**	**L3**	**2**	**2**	**2.6.2**	
14	Discuss about the various file format for an input/output file for Tableau.	**4**	**L2**	**3**	**2**	**2.6.2**	
15	Compare and contrast: Data profiling with Data wrangling.	**4**	**L2**	**3**	**2**	**2.6.4**	
16	Demonstrate the following visualization tool of Tableau: 1. Adding calculation formula 2. Changing the type of visualization 3. Filtering 4. resizing	**4**	**L3**	**3**	**3**	**2.7.2**	

Part – C (2 x 12 = 24 Marks)

Q. No	Question	Marks	BL	CO	PO	PI Code	Mark
17.a	Given the patient dataset with following attributes: patientID, sex, height, weight, blood_type, heart_rate, blood_pressure, no_hospitalvisits and year_ofvisit. Understand and apply data profiling using R programming and visualize the same. 1. Definition and explanation: Add Body Mass Index (BMI) 2. Comparison: Calculate No of Female vs Male patients and visualize 3. Dispersion: Find no of patient visited hospital monthly and visualize.	**12**	**L3**	**2**	**3**	**2.7.2**	
	OR						
17.b	Elaborate the procedure to read the date field from a dataset and reformat the date field with 2-character length to 4-character length, analyze the dataset for missing data and fill them using Data Manger tool.	**12**	**L4**	**2**	**3**	**2.7.1**	

18.a	Illustrate the means of profiling and manipulating your raw dataset to a format that can be easily consumed and visualized using Trifacta Wrangler. Illustrate with your own dataset.	**12**	**L4**	**3**	**3**	**2.7.1**	
	OR						
18.b	Design a business visualization dashboard for 'Business Sales and Promotions' using Tableau for the following: 1. Monthly product sales dollars with an average indicator 2. Sales versus spend by month 3. Trend of promotional spend as a percent of total product sales	**12**	**L6**	**3**	**3**	**2.7.2**	

Test: CLA-T3 **Date: 13-06-2024**

Course Code & Title: 18CSE490T & Big Data Visualization **Duration: 1 hr. 40 mins**

Year & Sem: IV Year / VII Sem **Max. Marks: 50**

Course Articulation Matrix:

Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PO 12	PSO1	PSO2	PSO3
CO4	2	2	2	3	3	-	-	-	2	2	3	3	-	-	-
CO5	2	3	3	3	3	-	-	-	2	2	3	3	-	-	-

CLO-4 : Designing and evaluating color palettes for visualization based on principles of perception

CLO-5 : Using the knowledge of perception and cognition to evaluate visualization design alternatives.

Part – A (10 x 1 = 10 Marks)
Instructions: Answer all

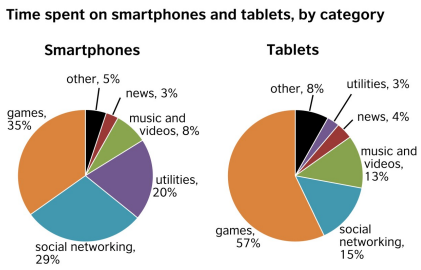
Q. No	Question	Marks	BL	CO	PO	PI Code	Mark
1	**Which of the following is TRUE about D3.js?** **a. Combines powerful visualization and interaction techniques with a data-driven approach to DOM manipulation** **b. Combines a Structured Query Language (SQL) with a descriptive language, to form Visual Query Language (VizQL)** **c. Uses prebuilt JavaScript functions to select elements, create SVG objects, style them, or add transitions** **d. Easily bind or use your large datasets to common scalable vector graphics objects** **i. A, B and C ii. A, C and D** **iii. A, B and D iv. A, B, C and D**	**1**	**L1**	**4**	**1**	**1.6.1**	
2	**The process of capturing the production data. stores it in memory and periodically sends it to a log file or data repository. Which program supports this process?** **i. Data Aggregator ii. Data logger** **iii. Data Handler iv. Data Manager**	**1**	**L1**	**4**	**1**	**1.6.1**	
3	**Which is the predecessor for D3.js?** **a. R Program b. Tableau** **c. Protovis d. Data Manager**	**1**	**L1**	**4**	**1**	**1.6.1**	

4	Identify the file formats that are supported by D3.js. a. CSV, JSON, TSV b. CSV, JSON, TDE c. CSV, TSV, geoJSON d. CSV, DATA, TSV i. A and B ii. B and C iii. A and C iv. D and D	1	L1	4	1	1.6.1	
5	"Separate the lines spatially" is an alternative strategy for a. Pie graph b. Slope graph c. Bar graph d. Spaghetti graph	1	L1	4	1	1.6.1	
6	An appropriate visual display should avoid the following types of charts. a. Pie, Donut, 3D, Secondary y-axis b. Pie, Line, Bar, primary y-axis c. Spaghetti, Donut, Bar, primary x-axis d. Line, Bar, Pie, Spaghetti	2	L1	5	1	1.6.1	
7	Identify the appropriate visuals to use when a single view of the data is used for both presentation and report. i. With animation ii. Contrast colors iii. Logic in order iv. Use Pie chart	1	L1	5	1	1.6.1	
8	Which chart is used to better visualize the data value consisting of both positive and negative values? a. Pie chart b. Donut Chart c. Bar chart d. Line chart	1	L1	5	1	1.6.1	
9	Which of the following are alternatives to PIE graph. a. Simple bar graph b. Slope graph c. Spaghetti graph d. Show numbers directly i. A, B and C ii. A, C and D iii. A, B and D iv. A, B, C and D	1	L1	5	1	1.6.1	
10	We have a scenario of multiple features to be visualized, which visualization better suits? i. With animation ii. Contrast colors iii. Logic in order iv. Use Pie chart	1	L1	5	1	1.6.1	

Part – B (4 x 4 = 16 Marks)
Answer any FOUR Questions

Q. No	Question	Marks	BL	CO	PO	PI Code	Mark
11	Create a webpage and add circle, rectangle, line, eclipse in SVG canvas using d3.js.	4	L6	4	2	2.6.2	
12	Write code to show how to attach data with the SVG element rectangle.	4	L6	4	2	2.6.2	
13	Utilizing manufacturing plant dataset, visualize machines parts count broken out by shift using D3.js.	4	L6	4	2	2.6.2	
14	Briefly explain the role of colors in visualization and how does it impact the visuals while communicating the data.	4	L4	5	2	5.4.2	
15	What are the strategies to be followed for creating more visual sense of the data using spaghetti graph?	4	L4	5	2	5.4.2	

16	Illustrate “logic in order” concept and how it plays an important role in displaying information.	4	L4	5	3	5.4.2	

Part – C (2 x 12 = 24 Marks)							
Q. No	Question	Marks	BL	CO	PO	PI Code	Mark
17.a	Demonstrate the procedure for runtime reconfiguration by switching the visualization view from stacked bar format to multiples bar format with suitable dataset and code using d3.js	12	L6	4	5	5.4.2	
OR							
17.b	Which chart from D3 libraries is used for handling negative values. Given the data, write the relevant code and explain the process to visualize the data using the chart in detail. name value machine 001 550 machine 002 -200 machine 003 -220 machine 004 800 machine 006 2000	12	L6	4	5	5.4.2	
18.a	The following pie chart shows the time spent on smart phones and tablets by category using pie chart.  Discuss about the other alternatives for displaying the above data.	12	L5	5	2	2.6.5	
OR							
18.b	Describe about the visuals of the data to be utilized when a single view of the data to be used for both presentation and report.	12	L5	5	2	2.6.5	