VASAVI COLLEGE OF ENGINEERING *(AUTONOMOUS)*, HYDERABAD
Accredited by NAAC with A++ Grade
B.E. (I.T.) VII-Semester Main & Backlog Examinations, Dec.-23/Jan.-24
Data Analytics with Visualization (PE-II)

Time: **3 hours**   Max. Marks: **60**

*Note: Answer all questions from **Part-A** and any **FIVE** from **Part-B***

Part-A (10 × 2 = 20 Marks)

Q. No.	Stem of the question	M	L	CO	PO
1.	What is data-driven decision-making?	2	1	1	1
2.	Identify NOIR categorized for the following data i) Duration of a class(in minutes) ii) Rating a movie with a maximum score of 5	2	2	1	1
3.	Write syntax to install and attach the package in R. Give an example.	2	2	2	1
4.	Given data D=84, 81, 55, 52, 46, 41, 28, 97, 91, 54, 41, 49. Find the five-number summary of D using R.	2	3	2	1
5.	Write any three applications of the boxplot.	2	1	3	1
6.	Define the significance of the heatmap?	2	1	3	2
7.	Given data D = 24, 31, 32, 41, 43, 51, 55, 55, 57, 58, 62, 67, 67, 69, 70. Find the interquartile range.	2	3	4	1
8.	Discuss one tail and two- tail in hypothesis testing	2	2	4	1
9.	How is visual analytics useful in understanding the supply chain?	2	1	5	1
10.	Discuss the applications of data analytics	2	2	5	1

Part-B (5 × 8 = 40 Marks)

Q. No.	Stem of the question	M	L	CO	PO
11. a)	Briefly explain the NOIR in the data analysis? Discuss with an example.	4	2	1	1
b)	What are the merits and demerits of data-driven decision-making?	4	2	1	1
12. a)	Discuss the data import and Export in R.	4	2	2	1
b)	How do you create the matrix, array and list data structures in R?	4	1	2	1
13. a)	What is an outlier? Write R code for creating a student dataset S=(RNo, Suject1-Marks, Suject2-Marks, Suject3-Marks,), and create a boxplot to identify the outliers in S.	4	3	3	2

		M	L	CO	PO
b)	Write R code for creating data frame D= (Day, Wind, Rainfall, Temperature) and plot a scatterplot between the Wind and Rainfall.	4	3	3	2
14. a)	Discuss the chi- square test with an example.	4	2	4	2
b)	A given vector v= [3, 5, 4, 7, 5, 3, 6, 9, 2, 8] and perform a sample t-test at 95% confidence interval. Given that the population mean is 2, analyze the result.	4	3	4	2
15. a)	Describe the role of visual analytics in healthcare.	4	2	5	2,3
b)	What is feature engineering? Discuss the feature engineering techniques.	4	1	5	1
16. a)	What is the difference between exploration and explanation?	4	2	1	1
b)	Apply Z-scale transformation for the following data X ={20,30,40,50} Y={1,7,8,9,21,35}	4	3	2	2
17.	Answer any *two* of the following:				
a)	Write R code for creating data frame *df=(ID, Income, Age, Gender)*. Plot a single scatter plot that describes the relationship among all the variables in the *df*, what do you infer from the plot?	4	4	3	2
b)	Perform one sample t-test for a vector with values 55, 83, 74, 61, 91, 52, 54, 62, 79, 81. Find the t-value. What is the significance of the t-value?	4	2	4	2
c)	Consider you are given the role of Data Analyst at a corporate hospital and asked to come up with a dashboard design involving patient admissions and patient history details. Listdown your assumptions related to data requirements Listdown various queries or objectives you consider for Exploratory data analysis.	4	3	5	2

M : Marks;   L: Bloom's Taxonomy Level;   CO; Course Outcome;   PO: Programme Outcome

i)	Blooms Taxonomy Level – 1	20%
ii)	Blooms Taxonomy Level – 2	40%
iii)	Blooms Taxonomy Level – 3 & 4	40%

*****