VELALAR COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution, Affiliated to Anna University, Chennai)

Continuous Assessment Test - I   QP Set   1   Regulations-2018 ver. 4

Programme   B.E. - CSE   Semester: 5   Max. Marks: 50   Duration   1 ½ Hrs.
Course Code & Title: **21CSE11 - EXPLORATORY DATA ANALYSIS**

Class:21CS5A&B   Date: 09.09.2023   **Time:** 11.00 am – 12.30 pm

| Knowledge | K1 – Remembering | K3 – Applying | K5 – Evaluating |
| Levels (KL) | K2 - Understanding | K4 – Analysing | K6 – Creating |

Part A – 10 x 2 = 20 Marks

1.	What is Exploratory Data Analysis?	CO1	K1
2.	List out the stages of EDA.	CO1	K1
3.	Define the term 'Data Wrangling' in Data Analytics.	CO1	K1
4.	What are the common problems that data analysts encounter during analysis?	CO1	K2
5.	Which are the technical tools that you have used for analysis and presentation purposes?	CO1	K1
6.	What is the output of the following python codes?	CO2	K3

i)   pd.Series([2, 4, 6])

ii)   df = pd.DataFrame({
        'Income': [15000, 1800, 120000, 10000],
        'Age': [25, 18, 42, 51],
        'Department': ['HR','Legal','Marketing','Management'] })

7.	data = pd.Series([1, np.nan, 'hello', None]) In the above data how to detect null values?	CO2	K3
8.	Outline the difference between .loc and .iloc with an example?	CO2	K2
9.	How to transform the below table from left side to right side view in python code?	CO2	K3

10.	Assume, you are given two lists : a=[1,2,3,4,5], b=[6,7,8,9]. The task is to create a list which has all the elements of a and b in one dimension. How will you implement in python?	CO2	K3

Part B – 2 x 15 = 30 Marks

No.			Question	Marks	CO	KL
11.	(a)	(i)	Outline the comparison of EDA with classical and Bayesian analysis.	7	CO1	K2
		(ii)	Summarize the purpose of different types of charts for visual aids in EDA.	8	CO1	K2

OR

	(b)		Explain the transforming techniques in EDA with an example.	15	CO1	K2
12.	(a)	(i)	How to combine the different data sources using python packages and explain the operations in briefly?	5	CO2	K2
		(ii)		10	CO2	K3

Consider the following DataFrames:

StudentID	ScoreSE
9	22
11	66
13	31
15	51
17	71
--	...
--	...
27	73
29	92

StudentID	ScoreSE
2	98
4	93
6	44
8	77
10	69
--	...
--	...
28	56
30	27

StudentID	ScoreML
1	39
3	49
5	55
7	77
9	52
--	--
--	--
27	23
29	49

StudentID	ScoreML
2	98
4	93
6	44
8	77
10	69
--	--
--	--
28	56
30	27

Perform the following cases
  i. How many students who are not taking the software engineering exam?
  ii. How many students who are not taking the machine learning exam?
  iii. How many students who appeared in both courses?
  iv. How many students appeared for the exams in total?
  v. How many students only appeared for the Machine Learning course?

OR

| | (b) | (i) | Write the python implementation of Outlier detection and filtering in transformation techniques | 5 | CO2 | K3 |
| | | (ii) | Explain the steps to perform EDA on wine Quality Data set. | 10 | CO2 | K2 |

31/08/23
Faculty Incharge

31/8/23
HOD/CSE

VELALAR COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution, Affiliated to Anna University, Chennai)

Continuous Assessment Test – II	QP Set	1	Regulations-2018 ver. 4

Programme **B.E. - CSE**   Semester: **5** Max. Marks: **50** Duration   1 ½ Hrs.

Course Code & Title: **21CSE11 - EXPLORATORY DATA ANALYSIS**

Class:21CS5A&B   **Date: 18.10.2023**   **Time: 11.00 am – 12.30 pm**

Knowledge Levels (KL)	K1 – Remembering	K3 – Applying	K5 – Evaluating
	K2 - Understanding	K4 – Analysing	K6 – Creating

Part A – 10 x 2 = 20 Marks

No.	Question	CO	KL
1.	What is the purpose of Univariate Analysis? Give an example.	CO3	K1
2.	What are the two techniques for reducing the number of digits?	CO3	K1
3.	State the important aspects of histograms during inspection of data.	CO3	K2
4.	Define standardized variable.	CO3	K1
5.	Write the formula for Gini coefficient?	CO3	K1
6.	Show the anatomy of a contingency table.	CO4	K4
7.	List out types of graphs are used to depict the bivariate analysis?	CO4	K3
8.	Show some examples of bivariate analysis?	CO4	K2
9.	Define the degree of freedom.	CO4	K1
10.	List out the importance of scatterplots.	CO4	K1

Part B – 2 x 15 = 30 Marks

No.			Question	Marks	CO	KL
11.	(a)	(i)	Compare Univariate , Bivariate and Multivariate Analysis.	8	CO3	K4
		(ii)	The dataset below shows the gross earnings in pounds per week of twenty men and twenty women drawn randomly from the 1979 New Earnings Survey (see appendix to this chapter on the accompanying website). The respondents are all full-time adult workers. Men are deemed to be adult when they reach age 21, women when they reach age 18.	7	CO3	K3

Men		Women	
150	58	90	39
55	122	76	47
82	120	87	80
107	83	58	42
102	115	50	40
78	69	46	99
154	99	63	77
85	94	68	67
123	144	116	49
66	55	60	54

Calculate the mean and standard deviation of the male earnings of the data. Compare them with the median and midspread you calculated. Why do they differ?

OR

(b)   What is measure of dispersion? Explain the following
i) Skewness
ii) Kurtosis   15   CO3   K2
iii) Calculating percentiles
iv) Quartiles

12.   (a)   (i)   From the following given table find   7   CO4   K3
i) Total percentage table
ii) Row percentage table
iii) Column percentage table

	a	b	c	d
x	663	102	529	85
y	158	69	330	55
z	149	19	144	56

(ii)   Explain the various types and importance of bivariate analysis.   8   CO4   K4

OR

(b)   How to perform Bivariate analysis for Numerical-numerical, Categorical-Categorical, and Numerical-Categorical variables?   15   CO4   K3

VELALAR COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution, Affiliated to Anna University, Chennai)

Continuous Assessment Test – III   QP Set   1   Regulations-2018 ver. 4

Programme B.E. - CSE   Semester: 5   Max. Marks: 50   Duration   1 ½ Hrs.
Course Code & Title: **21CSE11 - EXPLORATORY DATA ANALYSIS**

Class:21CS5A&B   Date: 18.11.2023   **Time:** 11.00 am – 12.30 pm

Knowledge   K1 – Remembering   K3 – Applying   K5 – Evaluating
Levels (KL)   K2 - Understanding   K4 – Analysing   K6 – Creating

Part A – 10 x 2 = 20 Marks

No.	Question	CO	KL
1.	List the components of multivariate analysis.	CO5	K1
2.	Outline the different causal relationships between variables.	CO5	K2
3.	Examine the unique aim for Principal Component Analysis (PCA).	CO5	K4
4.	What is the use of time series data analysis in organizations?	CO5	K1
5.	Define Time Series Analysis.	CO5	K1
6.	What do you meant by trend in Time Series Analysis?	CO5	K1
7.	Name two methods of measuring trend.	CO5	K1
8.	List the uses of analysis of Time series.	CO5	K1
9.	What do you mean by moving averages?	CO5	K1
10.	Name the methods of measurement of trend in time series analysis.	CO5	K1

Part B – 2 x 15 = 30 Marks

No.			Question	Marks	CO	KL
11.	(a)	(i)	Explain multivariate analysis. Illustrate the causal path models for three variables.	8	CO5	K2
		(ii)	Summarize the levels of the techniques in multivariate analysis.	7	CO5	K2

OR

| | (b) | | Explain how will you perform Time Series Analysis and apply the various Visualization techniques? Explain it with an example to plot trend over time. | 15 | CO5 | K2 |

| 12. | (a) | (i) | Calculate 3 years moving averages for the given data. | 7 | CO5 | K3 |

2014	2015	2016	2017	2018
15	18	17	20	23
2019	2020	2021	2022	2023
25	29	33	36	40

| | | (ii) | Identify the following in Time Series Analysis. | 8 | CO5 | K3 |

a. Secular Trend   b. Seasonal variation
c. Cyclical variation d. Irregular variation

OR

| | (b) | | Apply Multivariate Analysis of variance (MANOVA) in one case study. | 15 | CO5 | K3 |