Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering

End Sem (Odd) Examination Dec-2022

OE00075 Exploratory Data Analytics

Programme: B.Tech. Branch/Specialisation: All

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

Q.1 i. Which of the following gave rise to need of graphs in data analysis? **1**

(a) Data visualization (b) Communicating results

(c) Decision making (d) All of these

ii. Which of the following is characteristic of exploratory graph? **1**

(a) Made slowly

(b) Axes are not cleaned up

(c) Color is used for personal information

(d) All of these

iii. Which of the following graph can be used for simple summarization of data? **1**

(a) Scatterplot (b) Overlaying

(c) Barplot (d) All of these

iv. Which of the following statement is TRUE? **1**

(a) Outliers should be identified and removed always from a dataset.

(b) Outliers can never be present in the testing dataset.

(c) Outliers is a data point that is significantly close to other data points.

(d) The nature of our business problem determines how outliers are used.

v. Data analysis is a process of- **1**

(a) Inspecting data (b) Cleaning data

(c) Transforming data (d) All of these

P.T.O.

vi. Data analytics uses _____ to get insights from data. **1**
(a) Statistical figures (b) Numerical aspects
(c) Statistical methods (d) None of these

vii. Amongst which of the following is / are the types of linear regression? **1**
(a) Simple linear regression
(b) Multiple linear regression
(c) Both (a) and (b)
(d) None of these

viii. Least square method uses _____. **1**
(a) Linear polynomial (b) Linear regression
(c) Linear sequence (d) None of these

ix. A value that is much higher or much lower than the other values in a set of data- **1**
(a) Range (b) Outlier (c) Box plot (d) Histogram

x. A sector graph is known as- **1**
(a) Bar diagram (b) Histogram
(c) Pie diagram (d) Historigram

Q.2 Attempt any two:
i. Explain qualitative and quantitative data with suitable example. **5**
ii. Explain autocorrelation in great detail. **5**
iii. What is frequency polygon & how does it differs from histogram? **5**

Q.3 Attempt any two:
i. What is box cox linearity plot and how does it vary from box cox normality plot? **5**
ii. Differentiate between mean plot and contour plot. **5**
iii. Goals of EDA and role of graphics in EDA explain in great detail. **5**

Q.4 Attempt any two
i. What do you mean by assumptions in EDA and Why it is needed? With refer to this explain the 4 plot interpretation. **5**
ii. Explain types of data based on factors. **5**
iii. Explain youden plot in detail. **5**

Q.5 i. Ram and 7 of his classmates took the ACT test, a college entrance exam. The student's test scores were as follows: 26, 28, 32, 19, 23, 24, and 31. Using the data provided, make a stem-and-leaf plot. Explain the steps you used. **4**
ii. Differentiate heat map and tree map. **6**
OR iii. What is the difference between univariate, bivariate, and multivariate analysis? **6**

Q.6 Write short note on any two:
i. Explain Chi-square calculations. **5**
ii. Phi coefficient, scatter plot. **5**
iii. EDA techniques. **5**

******

Scheme of Marking

MEDI-CAPS UNIVERSITY Knowledge is Power	Faculty of Engineering End Sem (Odd) Examination Dec-2022 Exploratory data analysis-OE00075(T)
	Programme: B.Tech. | Branch/Specialisation:

Note: The Paper Setter should provide the answer wise splitting of the marks in the scheme below.

Q.1	i)	Which of the following gave rise to need of graphs in data analysis? a) Data visualization b) Communicating results c) Decision making d) **All of the mentioned**	1
	ii)	Which of the following is characteristic of exploratory graph? a) Made slowly b) Axes are not cleaned up c) **Color is used for personal information** d) All of the mentioned	1
	iii)	Which of the following graph can be used for simple summarization of data? a) Scatterplot b) Overlaying c) **Barplot** d) All of the mentioned	1
	iv)	Which of the following statement is TRUE? a) Outliers should be identified and removed always from a dataset. b) Outliers can never be present in the testing dataset. c) Outliers is a data point that is significantly close to other data points. **d) The nature of our business problem determines how outliers are used.**	1
	v)	Data Analysis is a process of? a)Inspecting data B)cleaning data C. transforming data **D. All of the above**	1
	vi)	Data Analytics uses ___ to get insights from data. a)Statistical figures b)Numerical aspects **c)Statistical methods** d)None of the mentioned above	1
	vii)	Amongst which of the following is / are the types of Linear Regression, A. Simple Linear Regression B. Multiple Linear Regression **C. Both A and B** D.None of the mentioned above	1
	viii)	Least Square Method uses ___ A.Linear polynomial **B.Linear regression** C Linear sequence D.None of the mentioned above	1
	ix)	A value that is much higher or much lower than the other values in a set of data A Range B **Outlier** C Box plot D Histogram	1
	x)	a sector graph is known as A Bar diagram B. Histogram **C. Pie diagram** D. Historigram	1
Q.2	i.	Why EDA is high on demand than classical analysis?	
	ii.	Explain Autocorrelation in great detail. 5	5
	iii.	What is frequency polygon 3 how does it differs from histogram? 2	5
OR	iv.	What is box cox linearity plot 3 and how does it vary from box cox normality plot. 2	5

		.	
Q.3	i.	Differentiate between mean plot and contour plot. 5	**5**
	ii.	Goals of EDA 2 and Role of Graphics in EDA explain in great detail 3.	**5**
OR	iii.	~~What do you mean by assumptions in Eda and Why it is needed?With refer to this explain the 4 plot interpretation. 5~~	**5**
Q.4	i.	What do you mean by assumptions in Eda 3 and Why it is needed?With refer to this explain the 4 plot interpretation. 2	**5**
	ii.	Explain types of data based on factors. 5	**5**
OR	iii.	Explain youden plot in detail5	**5**
Q.5	i.	7 of his classmates took the ACT test, a college entrance exam. The student's test scores were as follows: 26, 28, 32, 19, 23, 24, and 31. Using the data provided, make a stem-and-leaf plot. Explain the steps you used.	**4**
	ii.	Differentiate heat map and tree map 6	**6**
OR	iii.	What is the Difference between Univariate, Bivariate, and Multivariate analysis? (2 marks each)	**6**
s			
Q.6		Attempt any two: (Shorts Notes)	
	i.	Explain Chi-Square calculations 5	**5**
	ii.	Short Notes on Phi Coefficient, scatter plot 5	**5**
	iii.	EDA techniques 5	

what is boxcox linearity plot & how does it vary from boxcox normality plot.

******