

Code No.: AI742PE

R20

H.T.No.

8 R

**CMR ENGINEERING COLLEGE: : HYDERABAD  
UGC AUTONOMOUS**

**IV-B.TECH-I-Semester End Examinations (Regular) - November- 2023  
DATA ANALYTICS AND VISUALIZATION  
(CSM)**

[Time: 3 Hours]

[Max. Marks: 70]

**Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 20 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

**PART-A**

**(20 Marks)**

1. a) Explain about any four factors which help measuring the data quality. [2M]
- b) List and explain any two types of outliers. [2M]
- c) Explain about Discrete, Contiguous variables. [2M]
- d) List the different types of data modeling Techniques. [2M]
- e) Where does a Logistic Regression concept Applied. [2M]
- f) What is a Confusion matrix? [2M]
- g) What do you mean by overfitting? [2M]
- h) Describe about ARIMA. [2M]
- i) Write the purpose of Data visualization. [2M]
- j) Write the purpose of using Marimekko chart. [2M]

**PART-B**

**(50 Marks)**

2. Discuss the process of handling duplicate values in organizational data. [10M]
- OR**
3. Explain about the different types of data sources for data analysis. [10M]
4. Distinguish SQL and NoSQL Databases. [10M]
- OR**
5. Explain data imputation and how can repeated imputations enormously improve the quality of data. [10M]
6. With a neat diagram explain the Model Building Life Cycle in Data Analytics. [10M]
- OR**
7. Explain about Linear Regression Algorithm. [10M]
8. Explain the concept of pruning and complexity in tree building. [10M]
- OR**
9. Discuss in detail the steps involved in ETL process and tools available for this process. [10M]
10. Explain the challenges in visualizing complex data and relations and suggest suitable mechanisms to address them. [10M]
- OR**
11. Elaborate in detail about Icon-Based Visualization Techniques. [10M]

\*\*\*\*\*