

Exam Code: 0461
Sub. Code: 26047

2125
M.Sc. (IT), Third Semester
MS-75 [Opt (ii)]: Data Analytics-III

Time allowed: 3 Hours

Max. Marks: 80

NOTE: Attempt five questions in all, including Question No. IX (Unit-V) which is compulsory and selecting one question each from Unit I-IV.

x-x-x

Unit-I		
I	What are the challenges involved in data preparation using different datasets such as multimedia, social media, biological, and sensor data?	(16)
II	a. Differentiate between various types of data visualization techniques such as histograms, box plots, and scatter plots? b. Explain the concept of dashboard layout techniques and their significance in data visualization?	(16)
Unit-II		
III	a. Explain the significance of descriptive statistics in data analysis and list some common graphical representations? b. Discuss the measures of central tendency and how they help in understanding data distribution?	(16)
IV	a. Explain the concepts of correlation and linear regression? Provide an example scenario? b. What is the significance of understanding measures of skewness and kurtosis before data modeling?	(16)
Unit-III		
V	Discuss the importance of quality data in machine learning models? Illustrate with examples how anomalies and outliers can affect model performance?	(16)
VI	Describe the concept of dimensionality reduction? How does it improve machine learning model accuracy? Outline common techniques used for dimensionality reduction?	(16)
Unit-IV		
VII	Explain the significance of data science tools in modern data analytics? Discuss how MS-Excel, Tableau, Matplotlib, and TensorFlow are used in data analysis workflows?	(16)
VIII	Describe the importance of data ethics and privacy in data science? Discuss the challenges involved and future trends related to data privacy?	(16)
Unit-V		
IX	a. Why are filters useful in visualization? b. What does a heat map represent? c. What does a positive skewness indicate? d. What is the least squares method used for? e. Why are properties like data velocity and volume important in data engineering? f. What are outliers, and why is their detection important? g. What is imputation in data preprocessing? h. What is the goal of data exploration?	(16)

x-x-x