

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

School of Computing Science and Engineering

Master of Science in Computer Science

Mid Term Examination - May 2024

Duration : 90 Minutes

Max Marks : 50

Sem II - E1PY205B - Introduction to Data Analytics

General Instructions

Answer to the specific question asked

Draw neat, labelled diagrams wherever necessary

Approved data hand books are allowed subject to verification by the Invigilator

- 1) How does the classification of data impact data analytics? K2 (2)
- 2) Explain the tools used in data analytics. K1 (3)
- 3) Explain how the introduction of Big Data platforms impacted data analytics. K2 (4)
- 4) Analysis and reporting made for each other in the context of data analytics. Justify this statement. K2 (6)
- 5) List and explain 5 'V's of Big data. Provide examples of real-world applications of data analytics across industries. K3 (6)
- 6) Company X is considering implementing a data analytics strategy. Explain the difference between structured and unstructured data, and provide examples of each. How might the company leverage both types of data in its analytics efforts? K3 (9)
- 7) Discuss the impact of missing values on statistical analysis and decision-making processes. Explain two common techniques used to handle missing values during data cleaning. K4 (8)
- 8) Explore the concept of data transformation in the context of data cleaning. Provide two examples of common transformation techniques and discuss their significance in preparing data for analysis. Consider a dataset with 200 records. After data cleaning, 20% of the records were found to be duplicates and were removed. If the dataset originally had 240 records, how many records contained duplicate values? K4 (12)

OR

Assess the impact of data volume, complexity, and accessibility on the choice between reading and writing data from CSV, Excel, and SQL databases in Pandas. Provide specific examples to support your assessment. K4 (12)