

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)*

UG/PG DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL/MAY 2026

*Third Semester**Information Technology*

UCST24303 / CS3352 - FOUNDATIONS OF DATA SCIENCE

*(Common to Computer Science and Engineering)**Regulations - 2024 & 2021**Duration : 3 Hrs**Maximum Marks: 100***PART - A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
1.	Define data warehousing in the context of data science.	L1	1	2
2.	What are the facets of data that impact the choice of a data science model?	L2	1	2
3.	Differentiate between discrete and continuous variables.	L2	2	2
4.	State the usefulness of z-scores in determining outliers in a dataset.	L2	2	2
5.	Identify the slope of a regression line indicate about the relationship between two variables?	L2	3	2
6.	How would you apply the concept of regression towards the mean in predicting the performance of students based on their previous test scores?	L1	3	2
7.	What are the differences between a NumPy array and a traditional Python list?	L2	4	2
8.	How would you apply aggregation functions to summarize data in a Pandas DataFrame?	L2	4	2
9.	What factors influence the choice of colors in data visualization with Matplotlib?	L1	5	2
10.	Define annotations in the context of data visualization. Why are they important?	L1	5	2

PART - B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
11.	a i Discover the steps of data preparation and exploratory data analysis to a sample dataset and explain how these processes help in building an accurate model.	L4	1	8
	ii Analyze the relationship between data mining and data warehousing in enabling data-driven decision-making for businesses.	L4	1	8
	Or			
b	Analyze the role of data mining in supporting the effectiveness of a data warehousing system for large-scale business applications.	L4	1	16

12.	a	List and explain the different types of data and types of variables used in data analysis. Provide examples of each, and describe how these types impact the methods used for describing data with tables and graphs.	L2	2	16	
	Or					
	b	Relate how z-scores help standardize data in a normal distribution. Provide examples to illustrate how they are used to compare data points from different datasets.	L2	2	16	
13.	a	i	Explain how a regression line is used to make predictions, and describe its significance in understanding the relationship between variables.	L3	3	8
		ii	Analyze the difference between correlation and causation. Use examples to explain why a strong correlation does not necessarily mean that one variable causes the other.	L4	3	8
	Or					
	b	i	What is the standard error of estimate, and how is it used in the context of regression analysis?	L3	3	8
		ii	Analyze how multiple regression equations differ from simple linear regression. Provide an example of when multiple regression would be used.	L4	3	8
14.	a	i	List the basic types of aggregations that can be performed on NumPy arrays.	L3	4	8
		ii	Explain the concept of fancy indexing in NumPy. Provide an example of how it can be used to select elements from an array.	L3	4	8
	Or					
	b	Apply the concept of pivot tables in Pandas to analyze a dataset. Create a pivot table and explain its significance in summarizing and restructuring data.	L3	4	16	
15.	a	Design a comprehensive data visualization report using Matplotlib that includes various plot types (e.g., line plots, scatter plots, histograms). Include elements like legends, colors, and subplots, and discuss how each visualization aids in understanding the data.	L3	5	16	
	Or					
	b	Elaborate the features of Seaborn to create a visually appealing scatter plot with regression lines. Explain how Seaborn simplifies the process of creating complex visualizations compared to Matplotlib.	L3	5	16	