



STRATHMORE UNIVERSITY
SCHOOL OF COMPUTING & ENGINEERING SCIENCES
Master of Science in Sustainable Energy Transitions
Final Examination
MSET 8502: Data Science Concepts

DATE: 28th April 2026

TIME: 18.00-20.30

Instructions: Answer question one and any other two questions

Read the following case study and use it to answer the questions that follow.

CASE STUDY: RetailAppliancesGlobal Supermarket Chain

Background

RetailAppliancesGlobal is a mid-sized supermarket chain operating across several cities. The company has experienced declining profits despite increasing customer traffic. Management suspects inefficiencies in pricing, inventory management, and customer targeting.

You have been hired as part of its data science team to:

- I. Understand customer purchasing behavior
- II. Optimize pricing and promotions
- III. Improve inventory forecasting
- IV. Build a predictive model to estimate whether a customer will respond to a promotion

Available Data Sources

- Transaction data (sales, timestamps, product IDs)
- Customer data (age, location, loyalty status)
- Product data (category, price, supplier)
- Promotion data (discount %, campaign type)
- External data (holidays, weather)

Key Issues Identified

- Missing customer demographic values
- Duplicate transaction entries

- Incorrect price formats (e.g., strings instead of numeric)
- Outliers in purchase quantities
- Imbalanced dataset for promotion response (few responders)

Goal

Develop a machine learning model that predicts whether a customer will respond to a promotion and recommend actionable business strategies.

Question 1 (20 Marks)

- Define mutually exclusive and collectively exhaustive statements. Provide example(s) from the case study. **(4 Marks)**
- Compare design thinking and the scientific method in the context of this case. **(6 Marks)**
- Write pseudocode to determine if a customer qualifies for a promotion. **(3 Marks)**
- Describe a flowchart for (a model for) predicting promotion response. **(4 Marks)**
- Provide any three types of data associated with the case study. Include example(s) for each type of data. **(3 Marks)**

Question 2 (15 Marks)

- Convert this statement into a logic tree. **(4 Marks)**

“If discount > 20% and customer is loyal, then high response probability.”
- Describe how you will fix any four of the five issues described in the case study. **(4 Marks)**
- Describe any three Measures of Central Tendency that can be applied in the case study- explain each measure using example from the case study. **(3 Marks)**
- Write a query to find total sales per product. **(4 Marks)**

Question 3 (15 Marks)

- a) Differentiate Univariate vs Multivariate Analysis in relation to the provided case study. **(2 Marks)**
- b) In what proportions would you split the data, and why? **(3 Marks)**
- c) Describe any three Model Performance Metrics that can be applied to the case study. **(6 Marks)**
- d) Differentiate between Bias vs Variance. **(4 Marks)**

Question 4 (15 Marks)

- a) Describe the full data science pipeline in the case study. **(5 Marks)**
- b) Write a query to show the sum total of all product categories. **(2 Marks)**
- c) Create a **Logic Tree** or **Flowchart** for the following logic. **(4 Marks)**

If a bin's fill level is over 80%, mark for 'Immediate Collection'. If it's between 50% and 80% AND the battery is low, mark for 'Maintenance Check'. Otherwise, 'Skip'.
- d) Write a Python Program to print numbers 1–10. **(2 Marks)**
- e) Write a Python Program to check if a number is even or odd. **(2 Marks)**