Roll No:
Date: __/__/____

St. Claret College

Autonomous, Bengaluru

ST. CLARET COLLEGE LIBRARY Bengaluru-13 AUTONOMOUS

UG END SEMESTER EXAMINATION-MAY 2026
B.Sc. II SEMESTER
DS 225: EXPLORATORY DATA ANALYSIS

TIME: 3 hours. 15 **MAX. MARKS: 80**

This paper contains THREE printed pages and FOUR parts

Instructions:
1. **Verify and ensure that the question paper is completely printed.**
2. **Any discrepancies or questions about the exam paper must be reported to the COE within 1 hour after the examination.**
3. **Students must check the course title and course code before answering the questions.**

PART-A

Answer ALL questions. Each answer carries ONE mark. **[10 x 1 = 10]**

1. Which framework provides a structured process model for data mining and guides EDA from business understanding to deployment?
   (a) Agile (b) CRISP-DM (c) Scrum (d) Waterfall
2. Which of the following data types have an inherent order in them?
   (a) Nominal (b) Ordinal (c) Binary (d) Encoded
3. Which technique replaces missing values with a statistical measure such as mean, median, or mode?
   (a) Deletion (b) Imputation (c) Encoding (d) Normalization
4. Which Python visualization type is best suited to display the frequency distribution of a single continuous numerical variable?
   (a) Pie chart (b) Scatter plot (c) Heatmap (d) Histogram
5. Which language is used in Power BI to create calculated columns and custom measures for data analysis?
   (a) SQL (b) Python (c) DAX (d) MDX
6. Which Power BI visual allows users to interactively filter the report by selecting specific values from a field?
   (a) Tooltip (b) Slicer (c) Matrix (d) Card
7. Which component in Power BI is used to perform data loading, cleaning, and transformation before it is loaded into the data model?
   (a) DAX Editor (b) Report View (c) Power Query (d) Decomposition Tree
8. In Tableau, which pane contains fields that are typically qualitative and used to partition data into discrete categories?
   (a) Rows shelf (b) Measures pane (c) Dimensions pane (d) Analytics pane

9. Which Tableau chart type overlays two different measures on the same view using two independent vertical axes?
   (a) Scatter plot (b) Dual-axis chart (c) Heatmap (d) Treemap
10. What Tableau feature displays contextual summary information when a user hovers the mouse pointer over a data mark?
   (a) Filter (b) Set (c) Tooltip (d) Parameter

PART-B

Answer any TEN questions. Each answer carries TWO marks. **[10 x 2 = 20]**

11. Define EDA. State its purpose and significance in the data analytics pipeline.
12. List and briefly describe any four types of data quality issues commonly encountered during EDA.
13. What are the ethical aspects of EDA? Mention any four of each.
14. Differentiate between IQR and Z-score methods for outlier detection.
15. What is feature encoding?
16. State the difference between Min-Max scaling and Standard scaling.
17. What is DAX? List and briefly explain any two DAX functions used for creating measures in Power BI.
18. Differentiate between a Bar chart visual and a Matrix visual in Power BI with respect to their use case.
19. Explain the role of Power Query in Power BI data preparation.
20. Distinguish between Dimensions and Measures in Tableau. How does Tableau differentiate them by default?
21. What is a Dual-axis chart in Tableau? State one scenario where it provides meaningful insight.
22. Explain the hierarchy in Tableau with one example.

PART-C

Answer any FOUR questions. Each answer carries FIVE marks. **[4 x 5 = 20]**

23. Explain the CRISP-DM framework in detail. Explain how EDA helps in framing business goals and aligning analysis with business context.
24. Describe the various techniques used for handling missing values in Python (imputation, deletion, and interpolation). Write a short Python code using pandas to demonstrate mean imputation and forward-fill on a sample dataset.
25. Explain the steps involved in creating an interactive Power BI dashboard from scratch. Describe the role of at least two visual types in building an insightful report for a sales dataset.
26. Describe the DAX functions SUM, COUNT, AVERAGE, CALCULATE, and IF with syntax and one example each.
27. Write briefly about the pareto principle explaining its applications.
28. Explain the concept of Dashboards and Story Points in Tableau. How can they be used to effectively communicate data insights to business stakeholders?

PART-D

Answer any THREE questions. Each answer carries TEN marks. **[3 x 10 = 30]**

29. A hospital has collected patient records over five years. The dataset contains Age, Gender, Diagnosis, Blood Pressure readings, and Date of Admission. The data has several missing values in Blood Pressure, duplicate patient entries, and a few extreme age values.
    Answer the following based on the scenario:
    (a) Identify the type of each variable (numerical / categorical / time-series). [2]
    (b) List the data quality issues present in this dataset. For each issue, state one method to detect it and one method to fix it. [6]
    (c) Which EDA tool — Python, Power BI, or Tableau — would you recommend for this healthcare dataset and why? [2]
30. You are given the following small student marks dataset in Python:

```
import pandas as pd
data = pd.DataFrame({'Name': ['Alice','Bob','Carol','David','Eve','Frank'],
                     'Marks': [85, 70, 78, 200, 90, 75],
                     'Grade': ['A','B','B','A', NA,'C'],
                     'City': ['Delhi','Mumbai','Delhi','Chennai','Mumbai','Delhi']})
```

    (a) Write Python code to identify and fill the missing values in Grade using imputation. [2]
    (b) The Marks column has an obvious outlier. Write code to detect it using the IQR method. [3]
    (c) Apply one-hot encoding to the City column. Show the resulting DataFrame. [2]
    (d) Apply Min-Max scaling to the Marks column. [3]
31. A supermarket chain has sales data with the following fields: Product Name, Category, Region, Month, Units Sold, Unit Price, and Total Revenue. The data is stored in two tables — Sales and Products — linked by Product ID.
    Answer the following:
    (a) Identify the fact and dimension tables and write about star schema. [3]
    (b) Write DAX expressions to calculate: (i) Total Revenue, (ii) Average Units Sold [4]
    (c) Write briefly about the cardinality. [3]
32. Describe the Tableau interface giving its overview [5]. Compare Power BI with Tableau [5].
33. For the superstore dataset with the below fields:

Product Information	Customer Information	Geospatial Information
Product ID	Customer ID	Country
Product Name	Customer Name	City
Category	Segment	State
Sub-Category		Postal Code
		Region

    Write the steps to create the below mentioned visualizations in Tableau:
    (a) Map Visual [3] (b) Sales Trend and Forecasting [4] (c) Treemap [3]