| Roll No. | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

# PRESIDENCY UNIVERSITY

## BENGALURU

### End - Term Examinations – December 2025

**Date:** 05-12-2025 **Time:** 01:00pm – 04:00pm

| | | |
|---|---|---|
| **School:** All Schools | **Program**: BA,B.COM,BBA LLB HONOR'S | |
| **Course Code:** ECE3807 | **Course Name:** Introduction to Data Analytics | |
| **Semester:** VII,IX | **Max Marks:** 100 | **Weightage:** 50% |

| CO - Levels | CO1 | CO2 | CO3 | CO4 | CO5 |
|---|---|---|---|---|---|
| Marks | 12 | 12 | 24 | 26 | 26 |

**Instructions:**

*(i) Read all questions carefully and answer accordingly.*

*(ii) Do not write anything on the question paper other than roll number.*

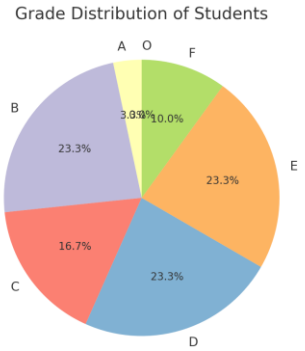
## Part A

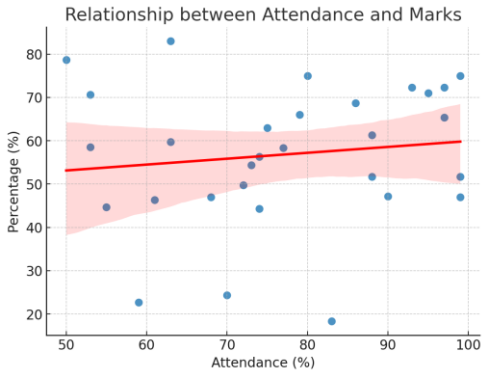
**Answer ALL the Questions. Each question carries 2marks.** **10Q x 2M=20M**

| | | | | |
|---|---|---|---|---|
| **1.** | List the steps used to manipulate the data | **2 Marks** | **L1** | **CO1** |
| **2.** | Write the command to import a CSV file and create a dataframe using Pandas. | **2 Marks** | **L1** | **CO2** |
| **3.** | List any three data transformation functions used in data analysis | **2 Marks** | **L1** | **CO3** |
| **4.** | Name the library used for data visualization. | **2 Marks** | **L1** | **CO3** |
| **5.** | What are the commonly used measures of central tendency? | **2 Marks** | **L1** | **CO4** |
| **6.** | Define Probability Distribution | **2 Marks** | **L1** | **CO4** |
| **7.** | Write the condition to find outliers using Inter-Quartile Range. | **2 Marks** | **L1** | **CO4** |
| **8.** | List any two differences between supervised and unsupervised learning. | **2 Marks** | **L1** | **CO5** |
| **9.** | Which command is used to call linear regression model using sklearn? | **2 Marks** | **L1** | **CO5** |
| **10.** | Name the model evaluation metrics for classification algorithm. | **2 Marks** | **L1** | **CO5** |

## Part B

**Answer the Questions.** **Total Marks 80M**

<table>
<tr><td>11.</td><td>a.</td><td>Explain the different types of data used in data analytics with suitable real-life examples.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>From the given Table, identify the following: Nominal data, Ordinal data, Discrete data, Continuous data
<table>
<tr><th>Mobile Type</th><th>Ward No.</th><th>No. of users</th><th>Rank</th><th>Performance Percentage</th><th>Customer Feedback</th></tr>
<tr><td>Samsung</td><td>1</td><td>84</td><td>4</td><td>74</td><td>Good</td></tr>
<tr><td>Apple</td><td>2</td><td>148</td><td>1</td><td>96</td><td>Outstanding</td></tr>
<tr><td>LG</td><td>3</td><td>147</td><td>2</td><td>94</td><td>Excellent</td></tr>
<tr><td>One Plus</td><td>4</td><td>138</td><td>3</td><td>76</td><td>Very good</td></tr>
</table></td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>12.</td><td>a.</td><td>Outline the stages of the data analytics workflow from data collection to decision-making, and demonstrate the process using a suitable real-life example.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Analyze the following visualizations and make descriptive analysis.



</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
</table>

<table>
<tr><td>13.</td><td>a.</td><td>Summarize the common NumPy operations used in data handling such as indexing, slicing and reshaping with examples.</td><td>10 Marks</td><td>L2</td><td>CO3</td></tr>
<tr><td></td><td>b.</td><td>Describe common data cleaning techniques to handle missing data using Pandas with examples.</td><td>10 Marks</td><td>L2</td><td>CO3</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>14.</td><td>a.</td><td>Describe how filtering and accessing rows/columns work in Pandas with examples.</td><td>10 Marks</td><td>L2</td><td>CO3</td></tr>
<tr><td></td><td>b.</td><td>Describe common transformation functions like apply(), map(), replace() and cut() with examples.</td><td>10 Marks</td><td>L2</td><td>CO3</td></tr>
</table>

<table>
<tr><td>15.</td><td>a.</td><td>Explain five types of outliers commonly observed in data analysis. Describe each type clearly and provide suitable examples</td><td>10 Marks</td><td>L2</td><td>CO4</td></tr>
<tr><td></td><td>b.</td><td>A bank monitors a busy ATM and finds that, on average, 4 customers arrive per hour. Using the Poisson distribution, draw the distribution graph for 10 customers per hour arrival rate. Also calculate the mean and variance.</td><td>10 Marks</td><td>L3</td><td>CO4</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>16.</td><td>a.</td><td>Explain the measures of central tendency (mean, median, mode). Using a sample data, show how are these measures calculated.</td><td>10 Marks</td><td>L2</td><td>CO4</td></tr>
<tr><td></td><td>b.</td><td>A retail shop records daily advertising spend and corresponding sales. Using simple linear model of least squares estimation, calculate the line of best fit. The 7-day data is provided.
<table>
<tr><th>Day</th><th>Ad spend (units of $100)</th><th>Sales (units of $1000)</th></tr>
<tr><td>1</td><td>1</td><td>2.1</td></tr>
<tr><td>2</td><td>2</td><td>2.5</td></tr>
<tr><td>3</td><td>3</td><td>3.7</td></tr>
<tr><td>4</td><td>4</td><td>3.9</td></tr>
<tr><td>5</td><td>5</td><td>5.2</td></tr>
<tr><td>6</td><td>6</td><td>5.8</td></tr>
<tr><td>7</td><td>7</td><td>6.4</td></tr>
</table></td><td>10 Marks</td><td>L3</td><td>CO4</td></tr>
</table>

<table>
<tr><td rowspan="2">17.</td><td>a.</td><td>Explain the different types of Machine Learning (supervised, unsupervised, and reinforcement learning) with suitable real-life examples.</td><td>10 Marks</td><td>L2</td><td>CO5</td></tr>
<tr><td>b.</td><td>Apply the scikit-learn workflow to iris dataset. Perform the following steps:<br>(i) Import iris dataset and map the species to the labels – 0: Setosa, 1: Versicolor, 2: Virginica<br>(ii) Split the dataset into training and testing data with 80% training and 20% testing data<br>(iii) Apply decision tree model to train and fit the data</td><td>10 Marks</td><td>L3</td><td>CO5</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td rowspan="2">18.</td><td>a.</td><td>Summarize the supervised learning workflow. Explain the importance of splitting data into training and testing sets, and define what labels represent in a dataset.</td><td>10 Marks</td><td>L2</td><td>CO5</td></tr>
<tr><td>b</td><td>Apply the scikit-learn workflow to iris dataset. Perform the following steps:</td><td>10 Marks</td><td>L3</td><td>CO5</td></tr>
</table>

| | | | | | |
|---|---|---|---|---|---|
| | | (i) Import iris dataset and map the species to the labels – 0: Setosa, 1: Versicolor, 2: Virginica<br>(ii) Plot pairplot for the dataset with title Histograms of Numeric Features<br>(iii) Find the accuracy score for the trained KNN model. | | | |