****

1. Answer the following questions:
2. What do the prompts C:\>, $ and >>> signify?
3. In which two modes can IDLE be used?
4. What is the purpose of the two programming modes offered by IDLE?
5. How can third party libraries be used in a Python program?
6. Match the following pairs:

| a. Data Ingestion | 1. Replacing missing values with the average value of the feature. |
| --- | --- |
| b. Exploratory Data Analysis (EDA) | 2. Visualizing data to uncover patterns and relationships. |
| c.Feature Engineering | 3. Scaling features to have zero mean and unit variance. |
| d. Mean Imputation | 4. A linear model used for classification tasks. |
| e. StandardScaler | 5. Loading data into the working environment for analysis. |
| f. Logistic Regression | 6. Modifying features to improve model performance. |
| g. Outlier Treatment | 7. Adjusting or removing data points that deviate significantly from others. |
| h.Decision Boundary | 8. The line or surface that separates different classes in a model. |
| **Answers**  1. **Data Ingestion** - **E** 2. **Exploratory Data Analysis (EDA)** - **B** 3. **Feature Engineering** - **F** 4. **Mean Imputation** - **A** 5. **StandardScaler** - **C** 6. **Logistic Regression** - **D** 7. **Outlier Treatment** - **G** 8. **Decision Boundary** - **H** |  |

1. State whether the following statements are True or False:
2. Google Colab allows us to run Python code without needing to install Python or any libraries on our local machine.

**Answer:** **True**

*Google Colab is a cloud-based environment that comes pre-installed with Python and commonly used libraries.*

1. Feature scaling is important to prevent features with larger scales from dominating the learning process.

**Answer:** **True**

*Scaling ensures that all features contribute equally to the model's learning process.*

1. Outliers in the iq column were handled by removing the rows containing these outlier values.

**Answer:** **False**

*Outliers were treated by capping IQ scores above 200 to a maximum value of 200.*

1. Mean imputation is suitable for replacing missing values in the cgpa column when the data is normally distributed.

**Answer:** **True**

*Mean imputation works well with normally distributed data as it doesn't significantly distort the dataset.*

1. A correlation coefficient close to zero indicates a strong linear relationship between two variables.

**Answer:** **False**

*A correlation coefficient close to zero indicates little to no linear relationship between the variables.*

1. Replacing missing values with the mean is always the best strategy for handling missing data.

**Answer:** **False**

*Mean imputation is simple but may not always be the best; other methods might be more appropriate depending on the data.*

1. The placement column had to be converted to numerical values because machine learning algorithms cannot directly handle categorical string values.

**Answer:** **True**

*Machine learning algorithms require numerical input; thus, categorical data must be encoded.*