Practical Questions (Code + Output)

Write Python scripts for each task. Submit both your code and output screenshots.

1. CSV Data Exploration & Visualization

- Download any small CSV dataset (e.g., Titanic, Iris, Wine Quality). Using pandas and matplotlib:

1. Load & Inspect
    - Load CSV file into a DataFrame.
    - Show the first 10 rows.
    - Display shape, column names, and data types.
2. Summary Statistics
    - describe() for numeric columns.
    - Count missing values per column.
    - Fill missing numeric values with the mean.
3. Filter & Sort
    - Filter rows by a numeric condition (e.g., Age > 30).
    - Sort the dataset by a column in descending order.
4. Group & Aggregate
    - Group by a categorical column, calculate mean of a numeric column.
5. Visualiza
    - Create a histogram for a numeric column.
    - Create a bar chart of group averages.
6. Extra Challenge: Save the cleaned dataset as processed_data.csv.

2. Decision Tree Classifier

1. Use the Iris dataset (load_iris from scikit-learn).
2. Train a DecisionTreeClassifier.
3. Print the accuracy score.
4. Plot the tree using plot_tree.

3. Text Processing with NLTK

1. Take a short paragraph from a news article.
2. Tokenize it.
3. Remove stopwords.
4. POS tag the remaining words.
5. Count how many nouns, verbs, and adjectives are in the text.
    a. Hint: Explore nltk.pos_tag and collections.Counter.

4. K-Means Clustering with Visualization

1. Generate a random 2D dataset using make_blobs (scikit-learn).
2. Apply KMeans with 3 clusters.
3. Plot results using matplotlib, with each cluster in a different color.

5. Confusion Matrix Plot

1. Train any classifier (Decision Tree, Logistic Regression, etc.) on the Iris dataset.
2. Predict on test set.
3. Plot confusion matrix using matplotlib or ConfusionMatrixDisplay.

Submission Guidelines:

1. Write answers for Section A in a single .pdf. (should be handwritten)
2. Submit Section B as .py files and screenshots of output. (attach it to your github repo)
3. Mention your dataset source in each practical question. (V.V. Important)