# Explanation of Task 2:

This document explains the steps involved in analyzing the Iris dataset using Python. The code performs dataset exploration, visualization, and optional exploratory data analysis (EDA).

## Step 1: Dataset Exploration:

1. The Iris dataset is downloaded from the UCI Machine Learning Repository.

Link : [archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data](https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data)  
2. Column names are assigned to make the dataset more readable.  
3. The first few rows of the dataset are displayed to understand its structure.  
4. Missing values and data types are checked. The 'species' column is converted to a categorical type if needed.  
5. Descriptive statistics (mean, median, standard deviation) are computed for numeric columns.

## Step 2: Data Visualization:

1. Histograms: Display the distribution of individual features using the matplotlib library.  
2. Scatter Plots: Pairwise scatter plots are created using seaborn's pairplot function to analyze relationships between features and observe species groupings.  
3. Box Plots: Box plots are used to identify the range of each feature and detect potential outliers.

## Step 3: Data Analysis :

1. Correlation Matrix: The correlations between numeric features are calculated and displayed as a matrix.  
2. Heatmap: A heatmap of the correlation matrix is created using seaborn to visualize feature relationships.  
3. Outlier Detection: Box plots for each feature help identify outliers.  
4. Observations: Initial observations about correlations, species clustering, and outliers are noted.

Dravid Nagi

nagidravid@gmail.com