# EDA & Data Visualization of the Iris Dataset

## Introduction

The purpose of this analysis is to perform Exploratory Data Analysis (EDA) on the Iris dataset using Python and visualize the identified patterns using Power BI or Tableau. The Iris dataset consists of 150 observations of iris flowers, with four features (sepal length, sepal width, petal length, and petal width) and one target variable (species).

## Approach and Methodologies

### A) Exploratory Data Analysis (EDA) with Python

1. **Loading the Dataset:**
   - The Iris dataset is loaded using the Seaborn library in Python.
2. **Basic Statistics:**
   - Descriptive statistics are computed to understand the central tendency, dispersion, and overall distribution of the data.
3. **Checking for Missing Values:**
   - The dataset is checked for any missing values to ensure data completeness.
4. **Visualizations:**
   - **Histograms:** Created to visualize the distribution of each feature.
   - **Box Plots:** Used to identify outliers and understand the spread of the data.
   - **Scatter Plots:** Utilized to explore relationships between pairs of features.
   - **Pair Plot:** Generated to observe interactions between all pairs of features.
   - **Correlation Matrix:** Computed and visualized to understand the relationships between features.

## B) Data Visualization with Power BI/Tableau

1. **Importing the Dataset:**
   - The Iris dataset is imported into Power BI or Tableau.
2. **Creating Visualizations:**
   - **Histograms:** Represent the distribution of each feature.
   - **Box Plots:** Visualize the spread and identify outliers.
   - **Scatter Plots:** Explore the relationships between pairs of features, with color coding for species.
   - **Correlation Heatmap:** A matrix visual to show correlation values with conditional formatting.

## C) Patterns Identified in the Iris Dataset

1. **Feature Distributions:**
   - **Sepal Length:** Most values range between 4.3 and 7.9 cm, with a peak around 5.8 cm.
   - **Sepal Width:** Values range from 2.0 to 4.4 cm, with a peak around 3.0 cm.
   - **Petal Length:** Values range from 1.0 to 6.9 cm, with a distinct peak around 1.5 and 5.5 cm.
   - **Petal Width:** Values range from 0.1 to 2.5 cm, with peaks around 0.2 and 1.8 cm.
2. **Species Differences:**
   - Setosa species are distinct with shorter petal lengths and widths.
   - Versicolor and Virginica species overlap more but can still be distinguished based on petal measurements.
3. **Correlations:**
   - Strong positive correlation between petal length and petal width.
   - Moderate positive correlation between sepal length and petal length.
   - Weak correlation between sepal width and the other features.

## D) Implementation Details

Python ipynb Files and Power BI .pbix files are attached along the the pdf.

### Power BI/Tableau Visualizations:

1. **Import Dataset:**
   - Load the CSV file containing the Iris dataset into Power BI or Tableau.
2. **Create Visuals:**
   - **Histograms:** Use the Histogram chart type for each feature.
   - **Box Plots:** Use the Box and Whisker chart type to visualize the spread.
   - **Scatter Plots:** Create scatter plots and color by species to observe relationships.
   - **Correlation Heatmap:** Create a matrix visual and apply conditional formatting to show correlation values.

## E) Conclusion

The analysis of the Iris dataset revealed distinct patterns in the distribution of features and correlations between them. Visualizing these patterns in Power BI or Tableau further highlighted the differences between species and relationships among features, providing valuable insights for further exploration and analysis.