Iris Dataset EDA Report

1. Basic Information

- The dataset contains 150 samples and 5 columns.

- Columns represent 4 features (sepal length, sepal width, petal length, petal width) and 1 target label (species).

- No missing values.

- Dataset is balanced with 50 samples each of Setosa, Versicolor, and Virginica.

2. Summary Statistics

- Sepal Length ranges roughly between 4.3 to 7.9 cm.

- Petal Length and Petal Width vary significantly between species.

- Setosa has notably smaller petals compared to the other two species.

3. Key Visualizations & Observations

Pairplot:

- Petal length and petal width clearly separate species.

- Setosa is completely separable.

Heatmap:

- Strong positive correlation (~0.96) between petal length and petal width.

Histograms:

- Petal measurements show clear differences across species.

Boxplots:

- Setosa has smallest petals; Virginica has largest.

- Overlap between Versicolor and Virginica.

Scatterplots:

- Petal Length vs Petal Width: Clear separation.

Iris Dataset EDA Report

- Sepal features show overlap.

4. Overall Findings

- Dataset is clean, balanced, and suitable for classification.

- Petal length and petal width are most important for distinguishing species.

- Setosa is easily separable; Versicolor and Virginica have slight overlap.

- No missing or abnormal values were found.