Task 1 Report: Exploring and Visualizing the Iris Dataset

Objective

The objective of this task was to understand how to load, explore, summarize, and visualize a dataset using Python libraries such as Pandas, Matplotlib, and Seaborn.

Dataset Description

The Iris dataset contains 150 observations and 5 columns. The features include measurements of sepal length, sepal width, petal length, petal width, and the species of the flower.

Steps Performed

- Loaded the dataset using pandas and seaborn libraries.
- Examined the dataset structure using df.shape to determine rows and columns.
- Displayed column names using df.columns.
- Viewed the first five records using df.head() to understand data format.
- Created a scatter plot to analyze the relationship between petal length and petal width.
- Generated a histogram to examine the distribution of sepal length values.
- Constructed a box plot to compare petal length across different species and detect outliers.

Observations

The scatter plot showed that the Setosa species is clearly separated from the other species based on petal measurements. The histogram indicated that most sepal length values fall within a central range. The box plot demonstrated differences in median values and spread among species, with minimal outliers observed.

Conclusion

This task provided practical understanding of basic Exploratory Data Analysis (EDA). It demonstrated how to inspect dataset structure and apply fundamental visualization techniques to interpret relationships, distributions, and variability within the data.