Basic EDA Questions on the Iris Dataset

Instructions:

Answer the following questions using Python and basic EDA techniques such as info(), describe(), isnull(), value_counts(), histograms, boxplots, pairplots, and correlation analysis.

Questions:

1. How many rows and columns are there in the Iris dataset?

2. What are the names and data types of the features (columns) in the dataset?

3. Are there any missing values in the dataset?

4. What is the mean, median, and standard deviation of the 'sepal_length' column?

5. Which species has the longest average 'petal_length'?

6. Create a histogram of 'sepal_width'. What does the distribution look like (e.g., normal, skewed)?

7. Use a boxplot to compare the 'petal_length' across different species. What do you observe?

8. Plot a pairplot of all the numerical columns. Do you observe any clear separation between species?

9. What is the correlation between 'sepal_length' and 'petal_length'? Interpret the result.

Basic EDA Questions on the Iris Dataset

10. Based on your analysis, which feature(s) do you think are most useful in distinguishing between the three species? Why?