PROJECT REPORT

IRIS DATASET

Objective:
The objective of this analysis is to evaluate whether Iris flower species can be distinguished using available physicall features and to assess the suitability of these features for classification modeling.
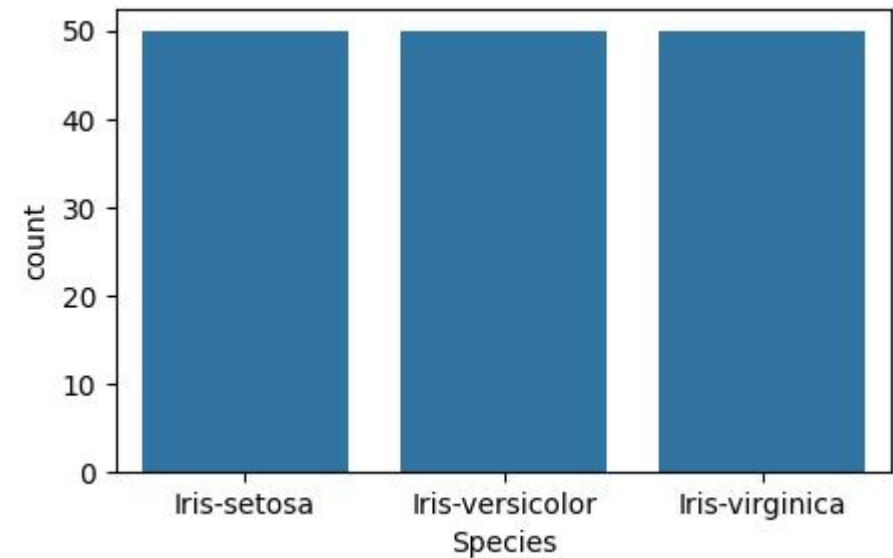
Dataset Overview:
The dataset consists of 150 observations across three Iris species.
Non-informative identifiers were removed to ensure the analysis focuses solely on predictive features.

Statistical Summary:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

The summary statistics indicates that petal length and petal width exhibit substantially higher variability, as reflected by larger standard deviations and wider ranges compared to sepal measurements, indicating stronger potential for species differentiation.
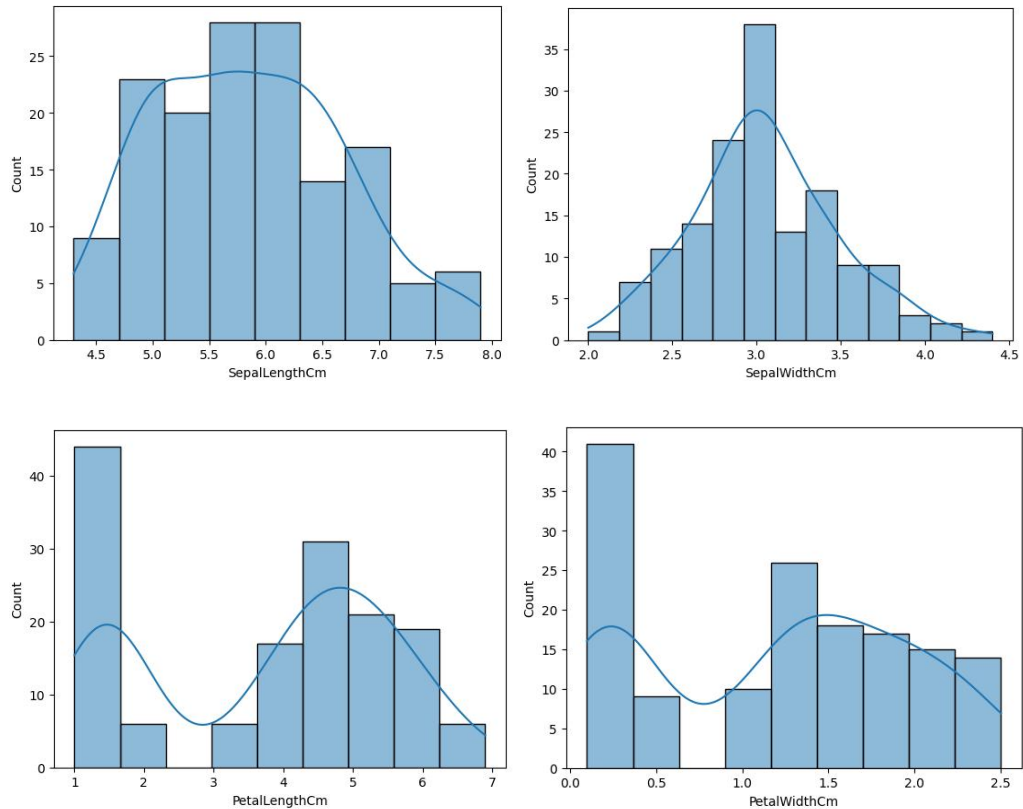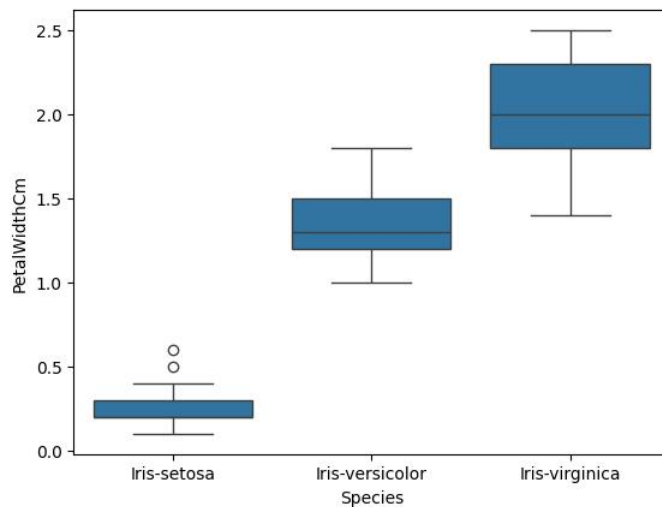
Class Distribution:

All three species are equally represented in the dataset.
This balanced class distribution reduces the risk of model bias toward any single species.

Feature Distributions:
Sepal measurements exhibit largely unimodal distributions with limited spread, indicating lower variability across observations. In contrast, petal length and petal width display pronounced bimodal density patterns, suggesting natural grouping within the data and stronger potential for species-level separation.
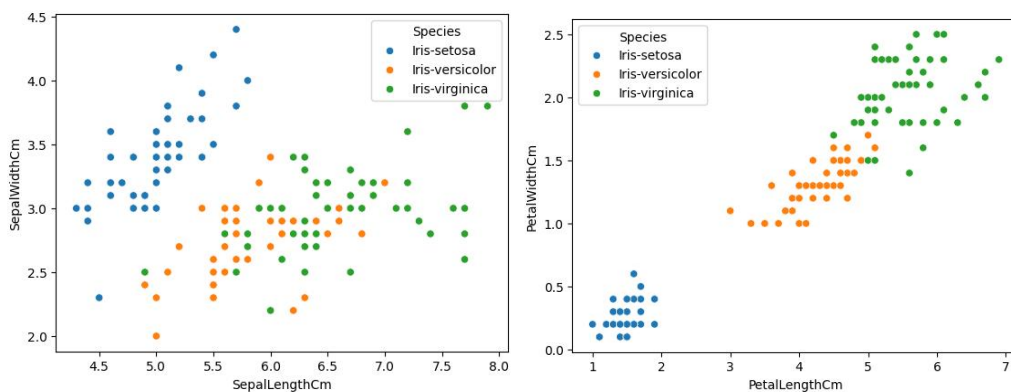
Variability Across Species:

Iris-setosa exhibits lower variability compared to the other species, while Iris-virginica shows wider dispersion.
Overlapping interquartile ranges between Iris-versicolor and Iris-virginica indicate potential classification challenges.

Feature Relationships

Sepal-based scatter plots reveal clear clustering for Iris-setosa, whereas significant overlap exists between Iris-versicolor and Iris-virginica.
This suggests that sepal features alone may be insufficient for robust classification.

Conclusion

The exploratory analysis indicates that while sepal features provide limited separability, petal measurements offer strong discriminatory power across Iris species.
As a result, classification models incorporating petal features are expected to achieve higher predictive accuracy.