

Internship Project Report

Intern Name: Pavuluri Sai Lithan

College: SR Gudlavalleru Engineering College

Domain: Data Science

Project: Exploratory Data Analysis on Iris Dataset

Organization: Hex Softwares

Submission: Internship Task 1

This report is generated as part of the Data Science Internship Program by Hex Softwares. It includes dataset exploration, data preprocessing, and visualizations for the Iris dataset.

Iris Dataset - Exploratory Data Analysis

Dataset: Iris (from sklearn.datasets)

Rows: 150, Columns: 6

Missing values per column:

sepal length (cm)	0
sepal width (cm)	0
petal length (cm)	0
petal width (cm)	0
target	0
target_name	0

Duplicate rows: 1

Class distribution:

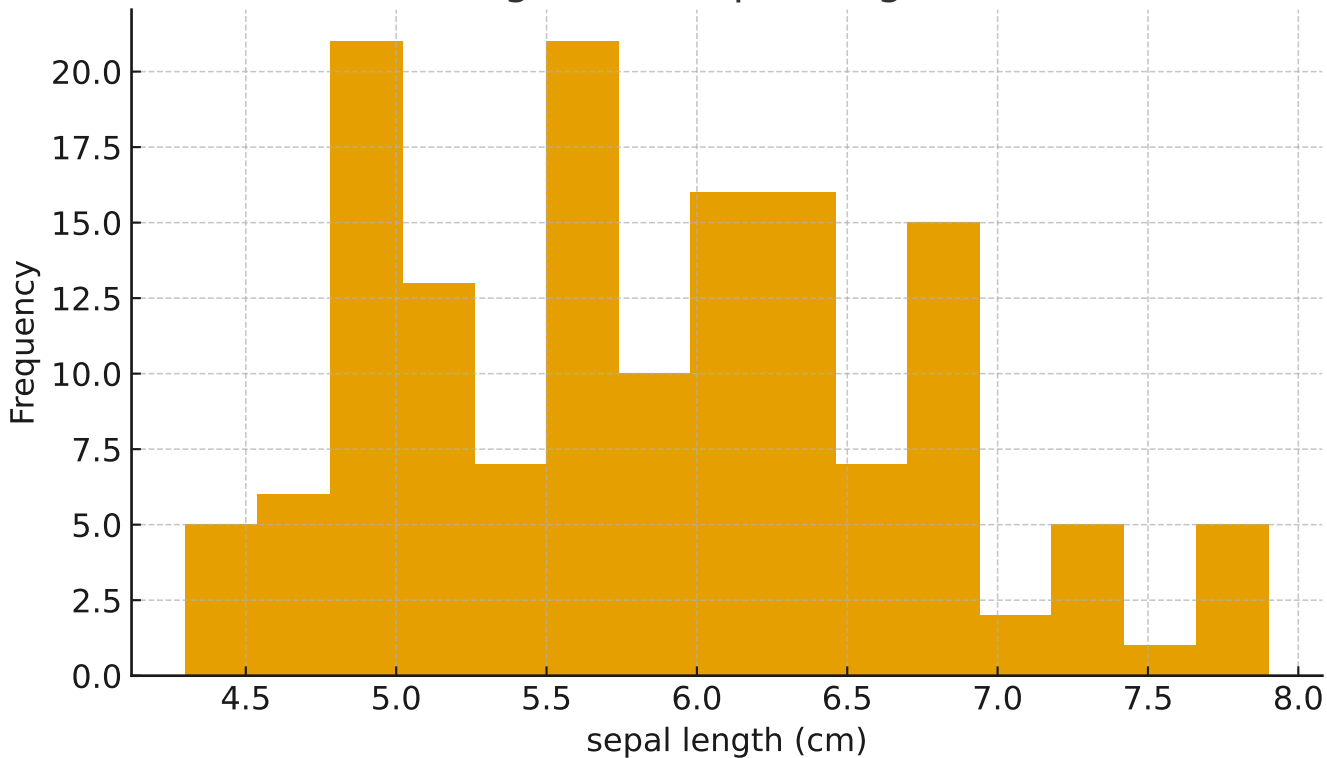
setosa	50
versicolor	50
virginica	50

Top summary statistics (numeric):

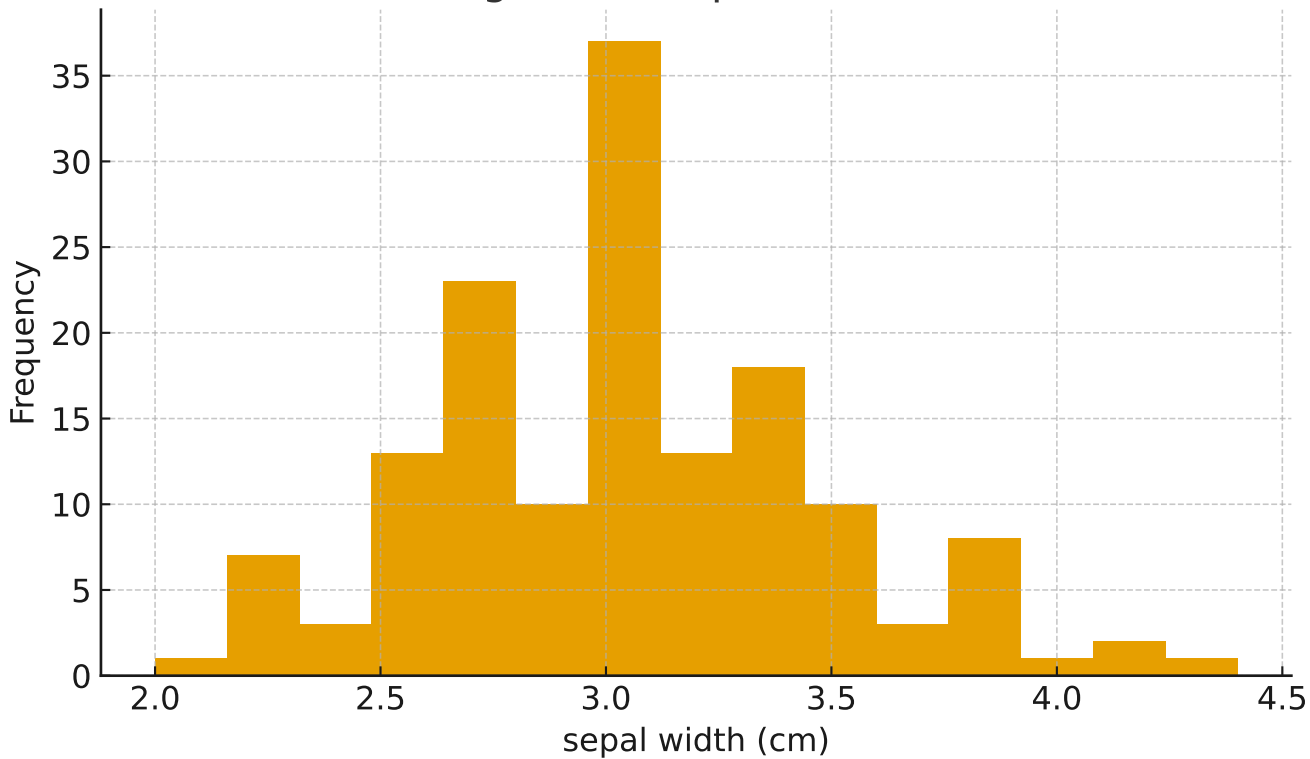
	count	unique	top	freq	mean	std	min	25%	50%	75%	max
sepal length (cm)	150.0	NaN	NaN	NaN	5.843333	0.828066	4.3	5.1	5.8	6.4	7.9
sepal width (cm)	150.0	NaN	NaN	NaN	3.057333	0.435866	2.0	2.8	3.0	3.3	4.4
petal length (cm)	150.0	NaN	NaN	NaN	3.758	1.765298	1.0	1.6	4.35	5.1	6.9
petal width (cm)	150.0	NaN	NaN	NaN	1.199333	0.762238	0.1	0.3	1.3	1.8	2.5

Correlation matrix (saved as a separate page of plots).

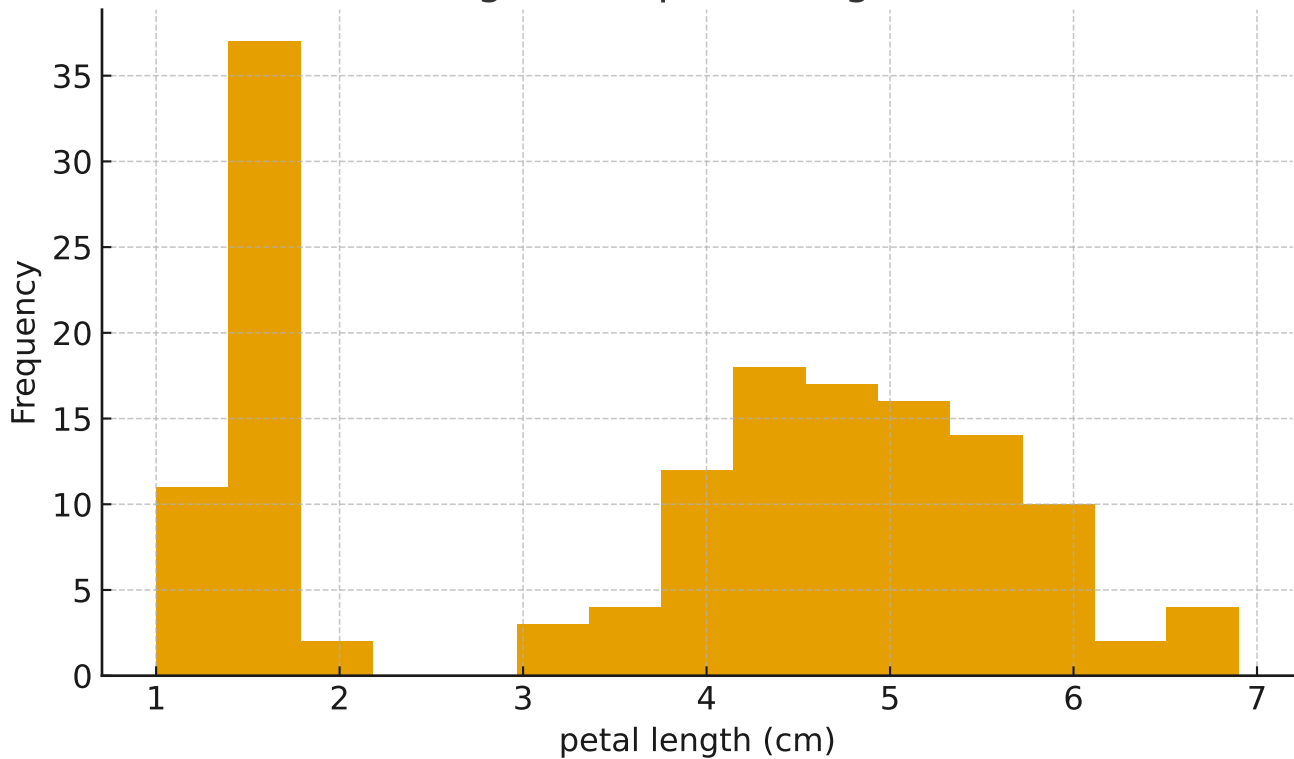
Histogram of sepal length (cm)



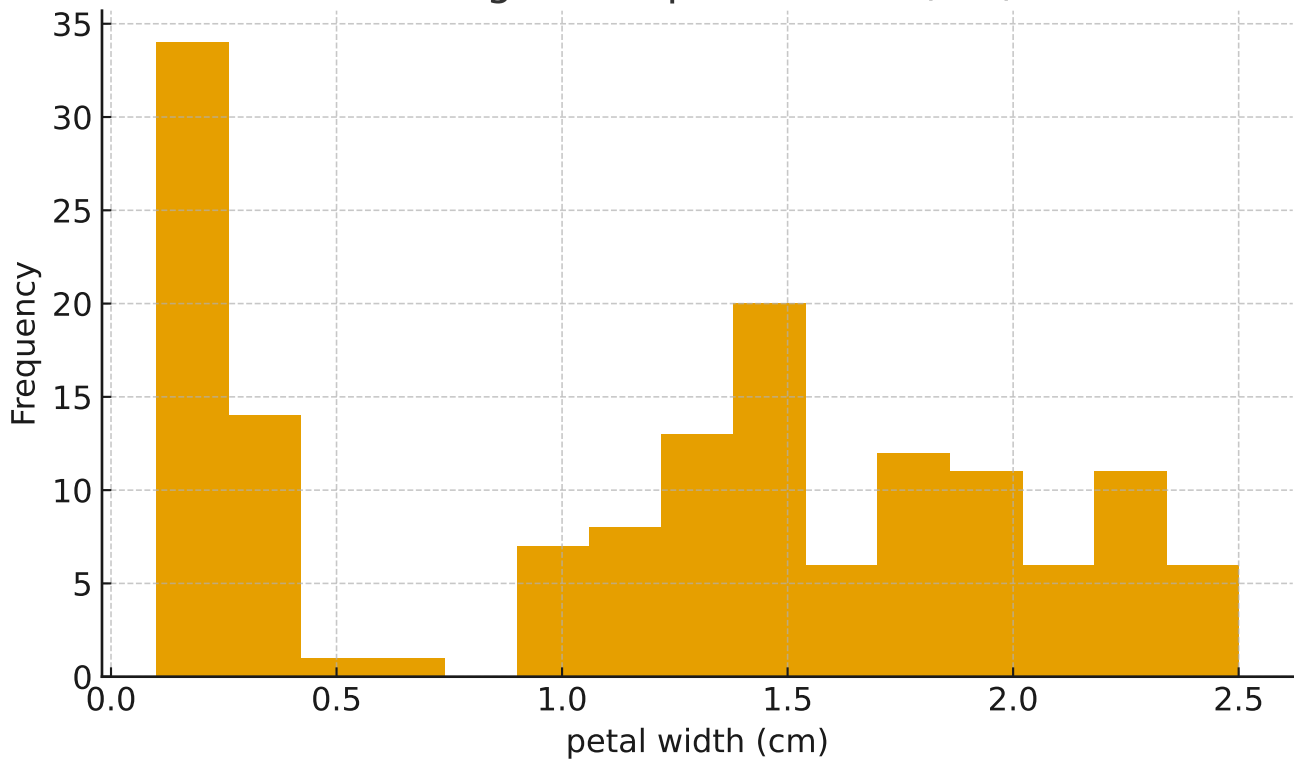
Histogram of sepal width (cm)



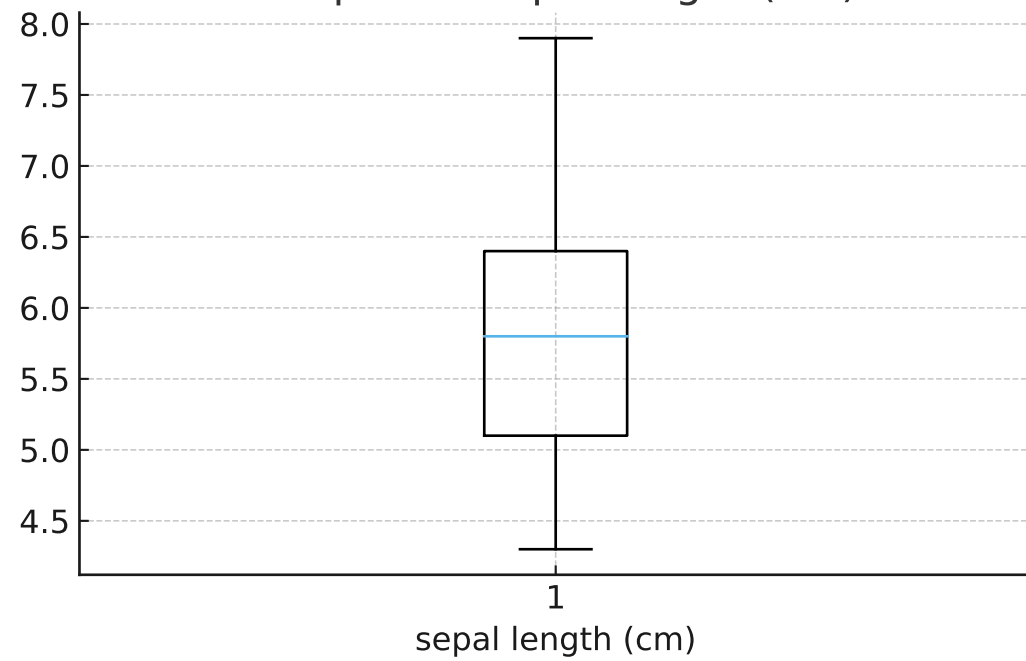
Histogram of petal length (cm)



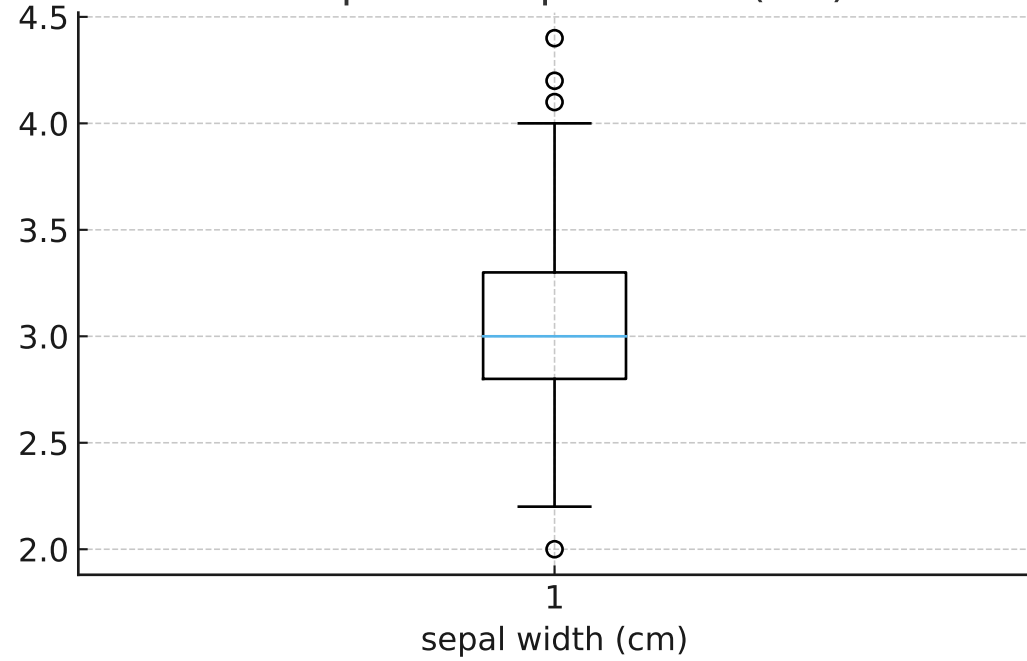
Histogram of petal width (cm)



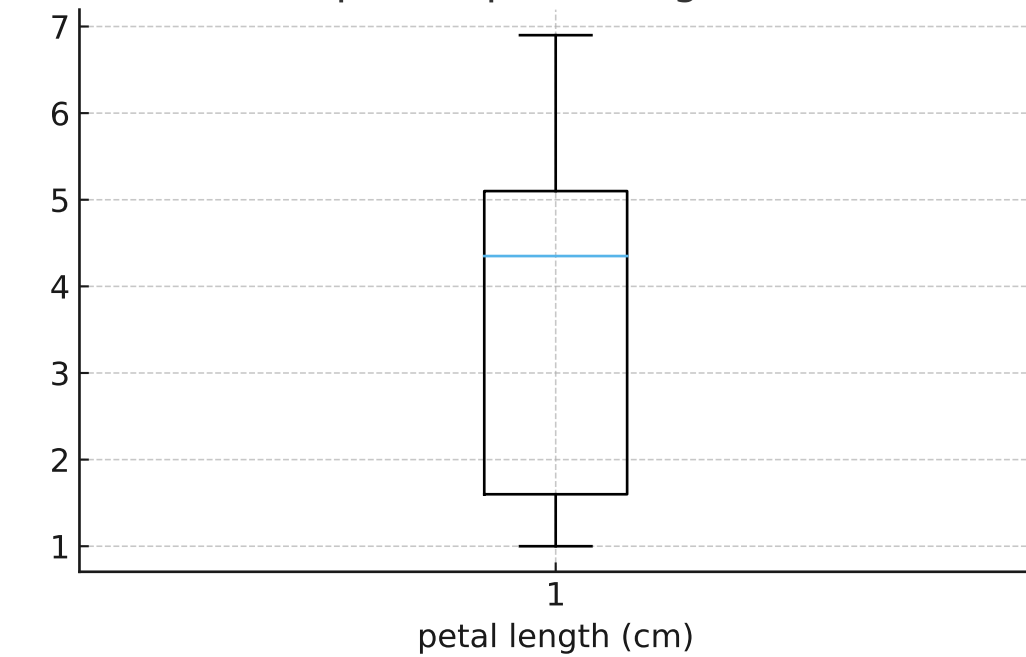
Boxplot of sepal length (cm)



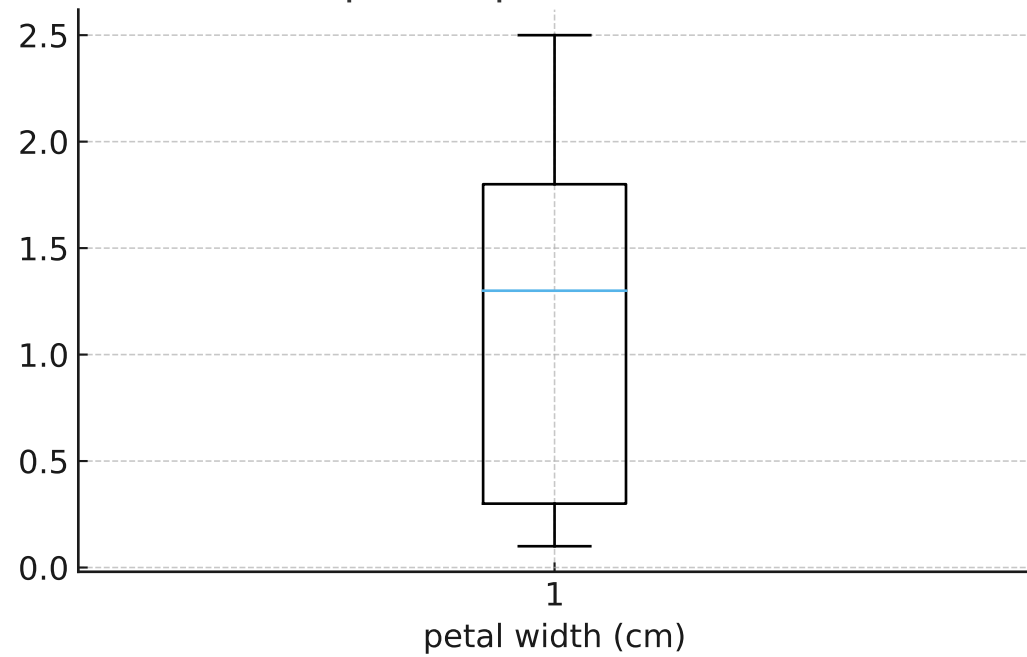
Boxplot of sepal width (cm)



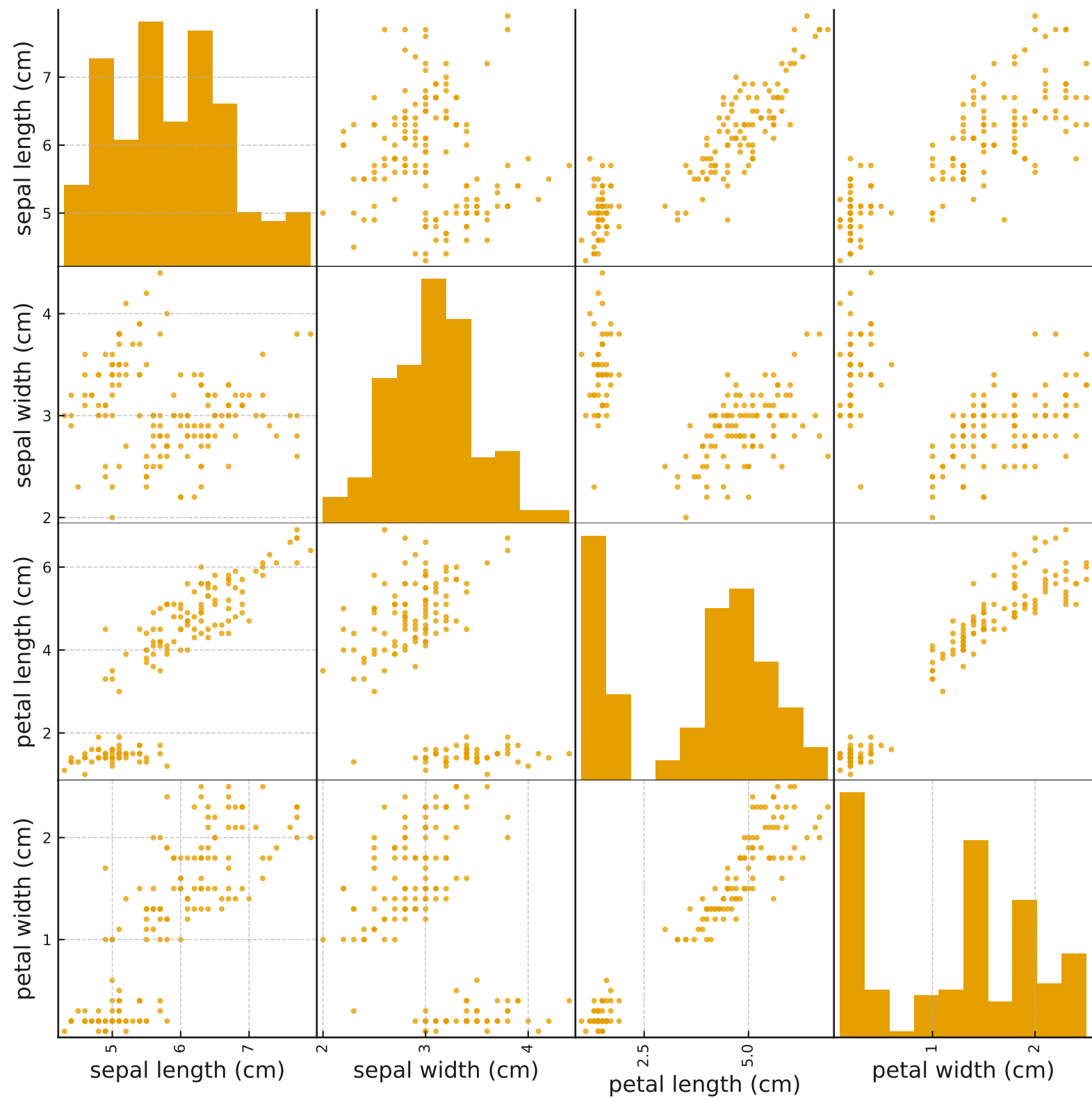
Boxplot of petal length (cm)



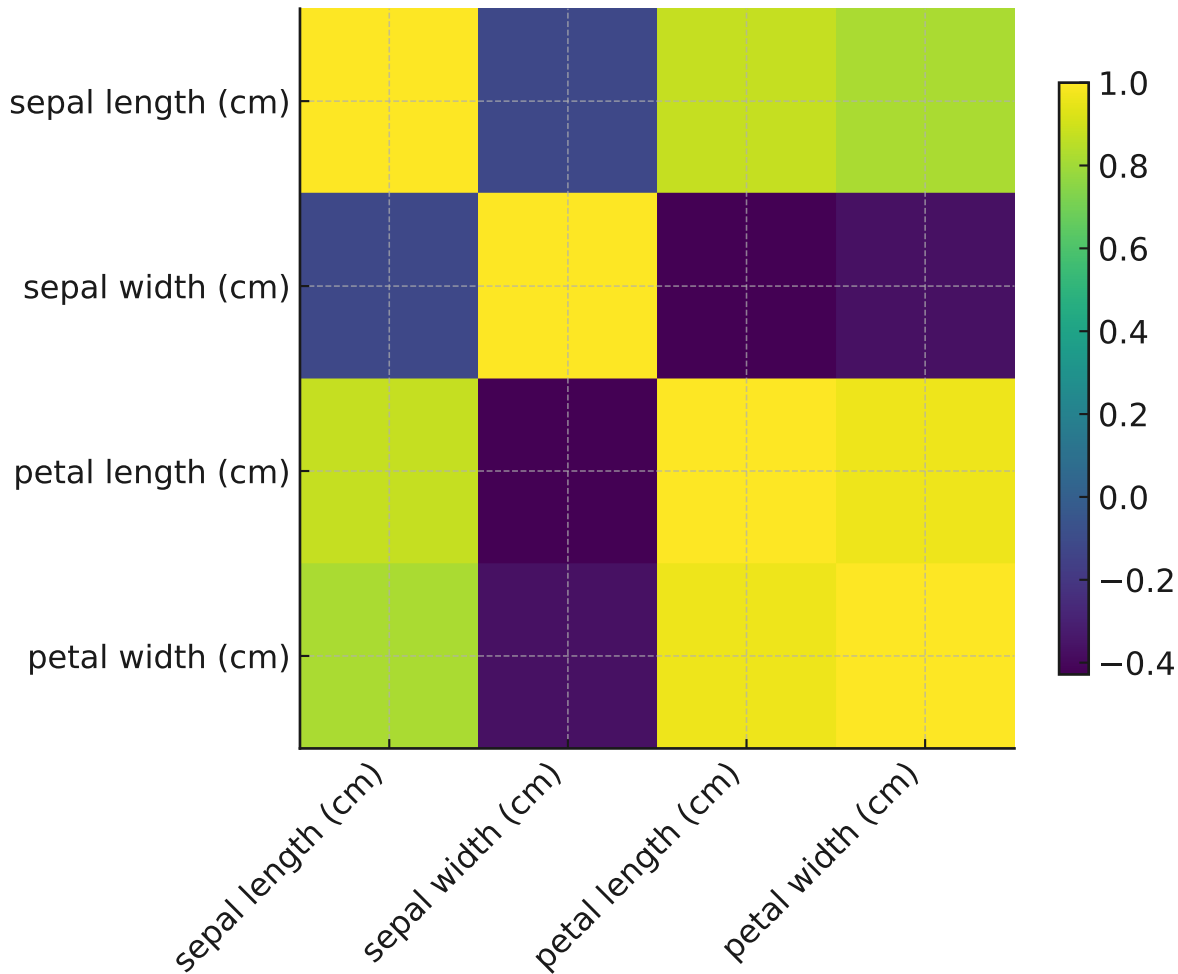
Boxplot of petal width (cm)



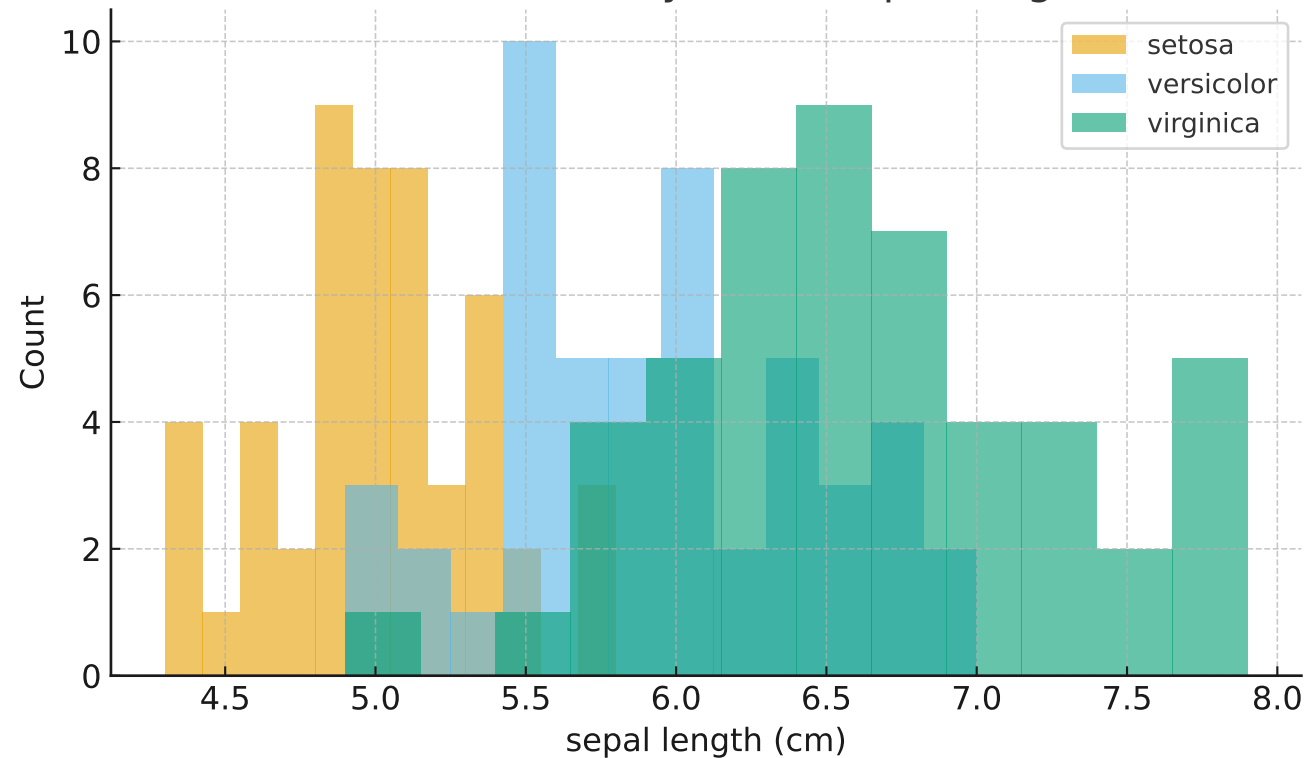
Scatter Matrix (pairwise relationships)



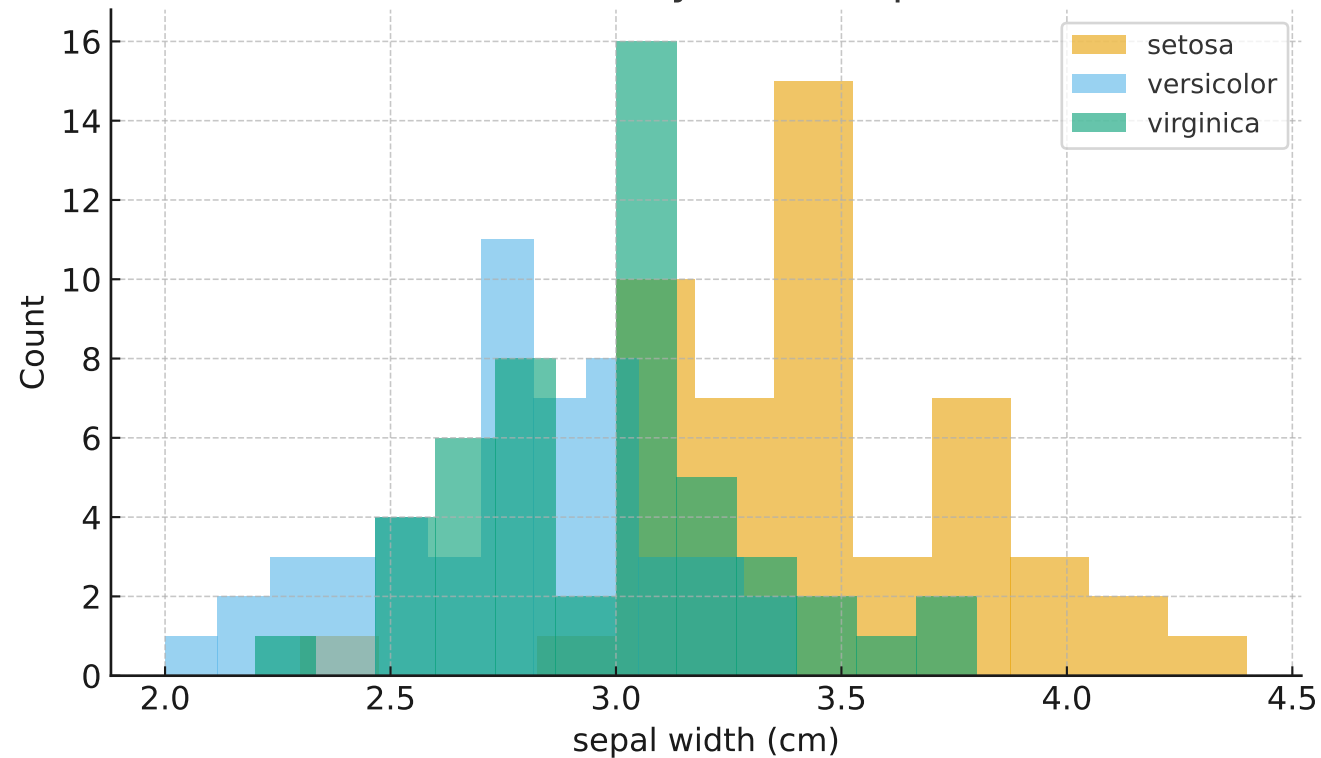
Correlation Matrix (features)



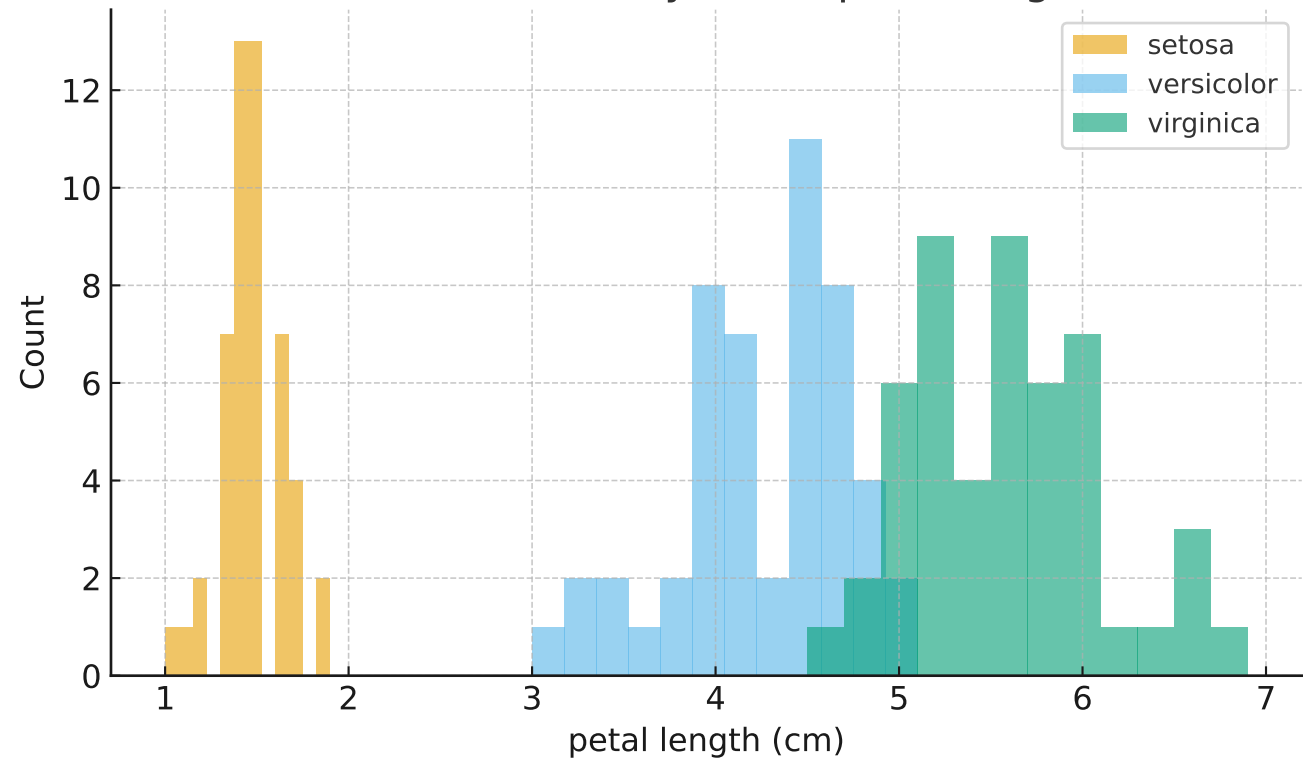
Feature distribution by class: sepal length (cm)



Feature distribution by class: sepal width (cm)



Feature distribution by class: petal length (cm)



Feature distribution by class: petal width (cm)

