

chls6lllj

April 29, 2025

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
```

```
[3]: # Download the Iris dataset from the provided URL
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_df = pd.read_csv(url, header=None, names=['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class'])
```

```
[4]: iris_df.head()
```

```
[4]:   sepal_length  sepal_width  petal_length  petal_width      class
0           5.1           3.5           1.4           0.2  Iris-setosa
1           4.9           3.0           1.4           0.2  Iris-setosa
2           4.7           3.2           1.3           0.2  Iris-setosa
3           4.6           3.1           1.5           0.2  Iris-setosa
4           5.0           3.6           1.4           0.2  Iris-setosa
```

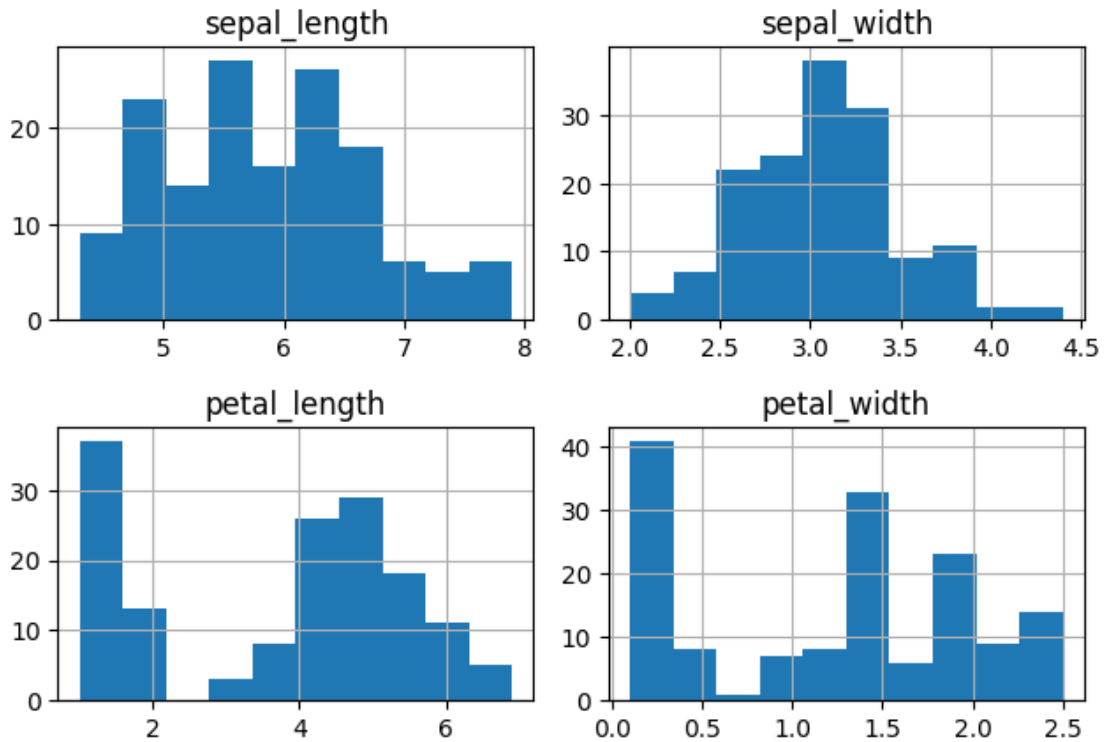
```
[5]: # List down the features and their types
features = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
feature_types = ['numeric', 'numeric', 'numeric', 'numeric']
```

```
[6]: # Print the features and their types
for feature, ftype in zip(features, feature_types):
    print(f"Feature: {feature} - Type: {ftype}")
```

```
Feature: sepal_length - Type: numeric
Feature: sepal_width - Type: numeric
Feature: petal_length - Type: numeric
Feature: petal_width - Type: numeric
```

```
[7]: # Create histograms for each feature
iris_df[features].hist()
plt.suptitle('Histograms of Iris Dataset Features')
plt.tight_layout()
plt.show()
```

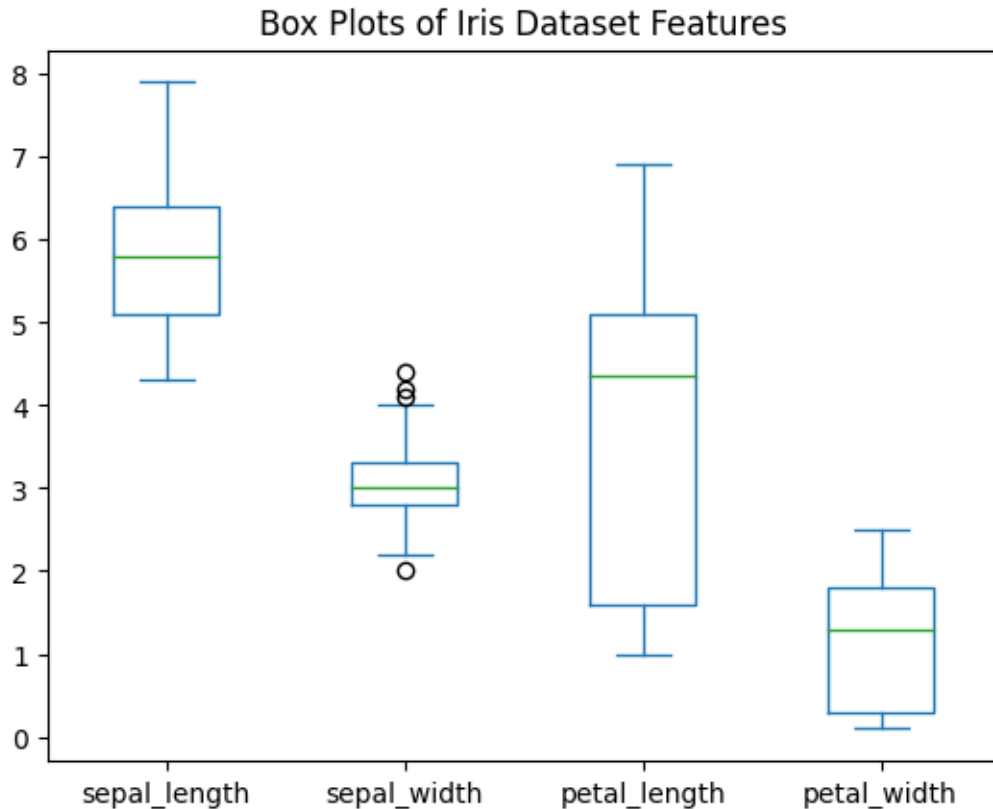
## Histograms of Iris Dataset Features



```
[8]: # Create box plots for each feature
iris_df[features].plot(kind='box')
plt.title('Box Plots of Iris Dataset Features')
plt.show()

# Identify outliers
outliers = []
for feature in features:
    q1 = iris_df[feature].quantile(0.25)
    q3 = iris_df[feature].quantile(0.75)
    iqr = q3 - q1
    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
    feature_outliers = iris_df[(iris_df[feature] < lower_bound) |
    ↪(iris_df[feature] > upper_bound)]
    outliers.append(feature_outliers)

# Print the outliers for each feature
for feature, outlier_df in zip(features, outliers):
    print(f"\nOutliers for feature: {feature}")
    print(outlier_df)
```



Outliers for feature: sepal\_length

Empty DataFrame

Columns: [sepal\_length, sepal\_width, petal\_length, petal\_width, class]

Index: []

Outliers for feature: sepal\_width

	sepal_length	sepal_width	petal_length	petal_width	class
15	5.7	4.4	1.5	0.4	Iris-setosa
32	5.2	4.1	1.5	0.1	Iris-setosa
33	5.5	4.2	1.4	0.2	Iris-setosa
60	5.0	2.0	3.5	1.0	Iris-versicolor

Outliers for feature: petal\_length

Empty DataFrame

Columns: [sepal\_length, sepal\_width, petal\_length, petal\_width, class]

Index: []

Outliers for feature: petal\_width

Empty DataFrame

Columns: [sepal\_length, sepal\_width, petal\_length, petal\_width, class]

Index: []

[ ]: