

# dsbda10

May 4, 2025

```
[1]: # Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[3]: # Load the dataset
df = pd.read_csv("C:/Users/aasth/OneDrive/Desktop/DSBDA/Iris.csv")
```

```
[5]: # 1. Display basic information about the dataset
print("First 5 records:\n", df.head())
print("\nDataset Info:")
print(df.info())

print("\nSummary Statistics:")
print(df.describe())
```

First 5 records:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 150 entries, 0 to 149

Data columns (total 6 columns):

#	Column	Non-Null Count	Dtype
0	Id	150 non-null	int64
1	SepalLengthCm	150 non-null	float64
2	SepalWidthCm	150 non-null	float64
3	PetalLengthCm	150 non-null	float64
4	PetalWidthCm	150 non-null	float64
5	Species	150 non-null	object

dtypes: float64(4), int64(1), object(1)

memory usage: 7.2+ KB

None

Summary Statistics:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
[7]: # 2. Identify feature types
print("\nFeature types:")
for col in df.columns:
    dtype = df[col].dtype
    if dtype == 'object':
        print(f"{col}: Nominal")
    else:
        print(f"{col}: Numeric")
```

Feature types:

Id: Numeric

SepalLengthCm: Numeric

SepalWidthCm: Numeric

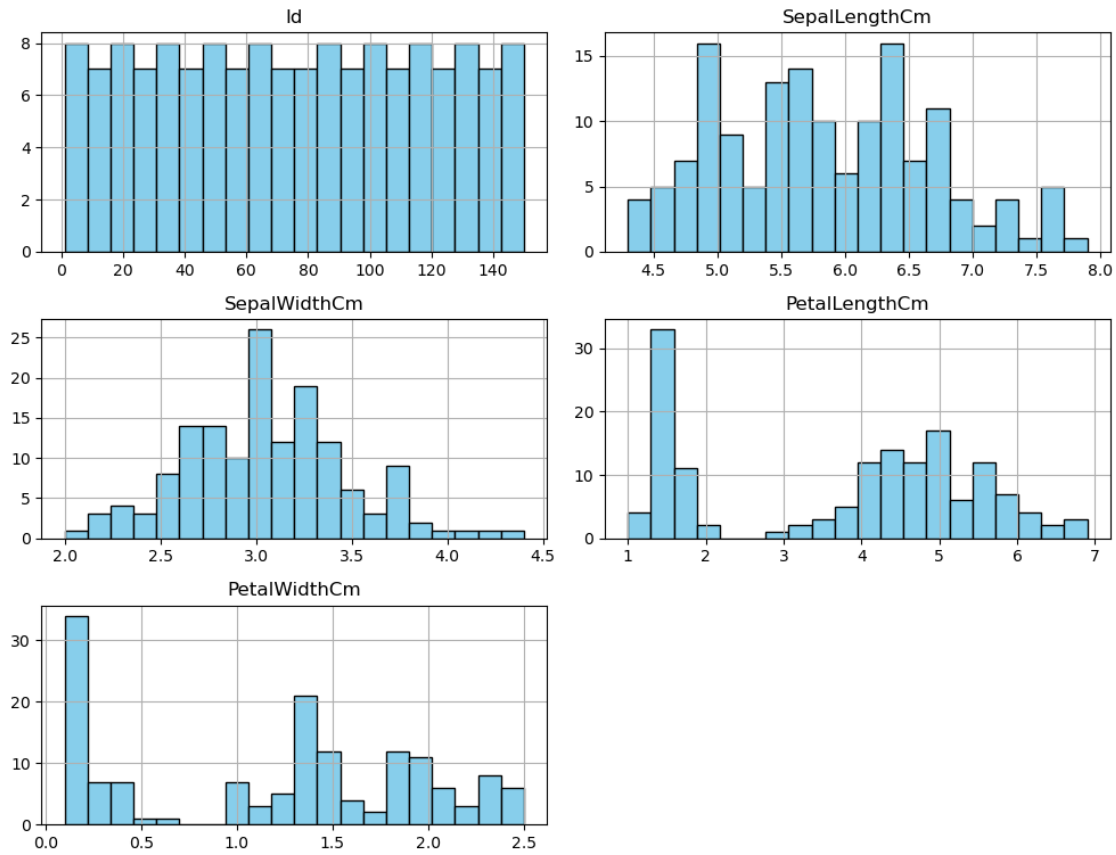
PetalLengthCm: Numeric

PetalWidthCm: Numeric

Species: Nominal

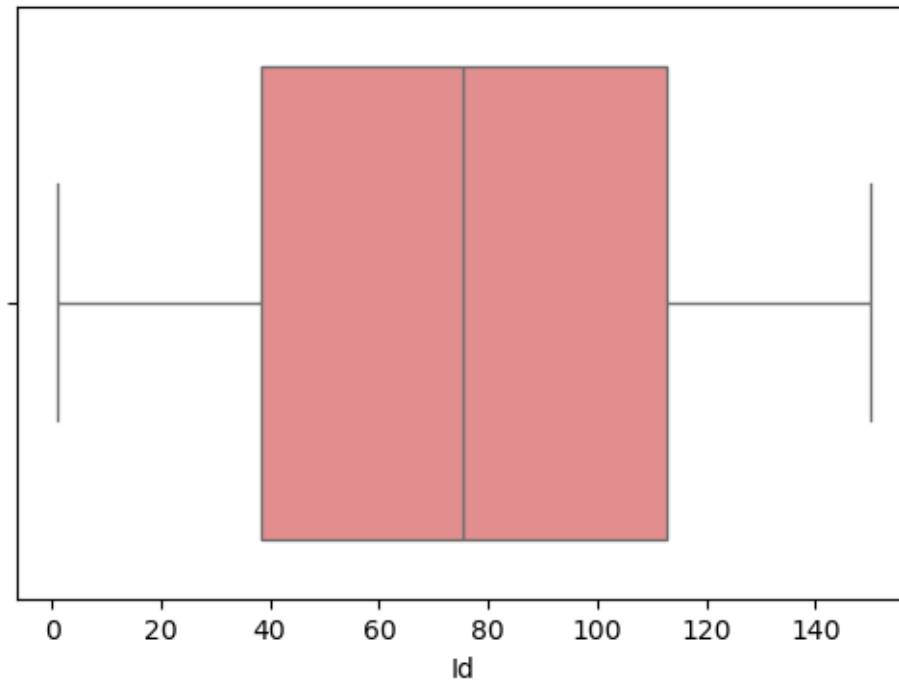
```
[9]: # 3. Histograms for each numeric feature
df.hist(figsize=(10, 8), bins=20, color='skyblue', edgecolor='black')
plt.suptitle("Histograms of Iris Dataset Features")
plt.tight_layout()
plt.show()
```

Histograms of Iris Dataset Features

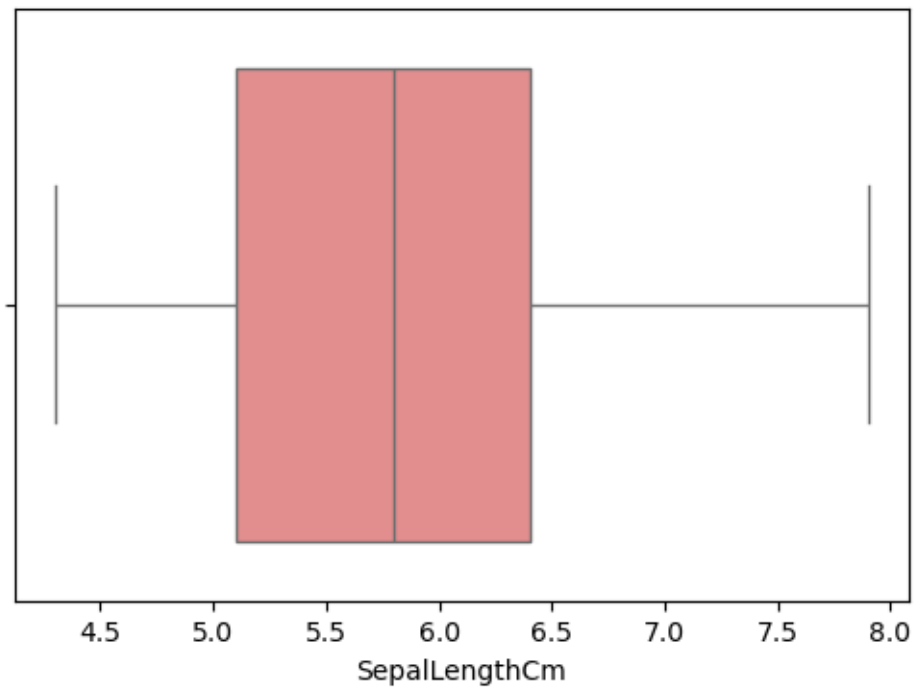


```
[11]: # 4. Box plots for each numeric feature
numeric_features = df.select_dtypes(include=['float64', 'int64']).columns
for feature in numeric_features:
    plt.figure(figsize=(6, 4))
    sns.boxplot(x=df[feature], color='lightcoral')
    plt.title(f"Boxplot of {feature}")
    plt.show()
```

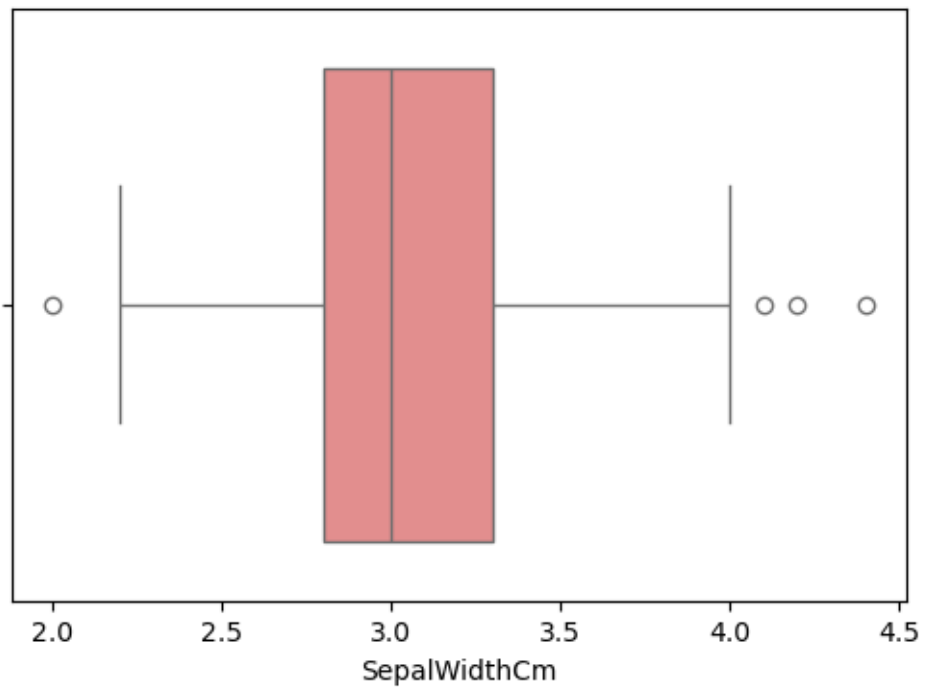
Boxplot of Id



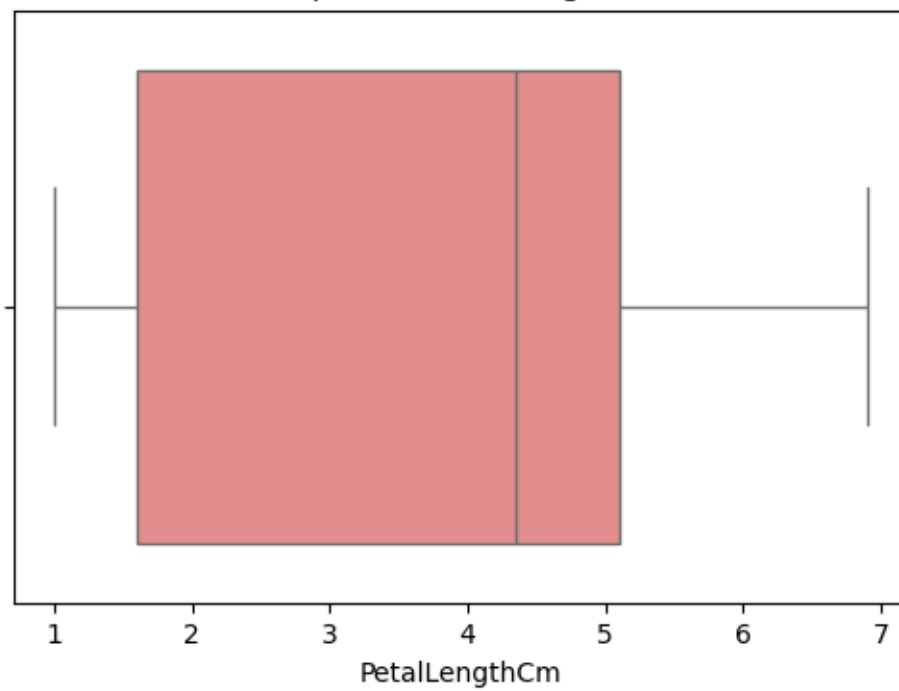
Boxplot of SepalLengthCm

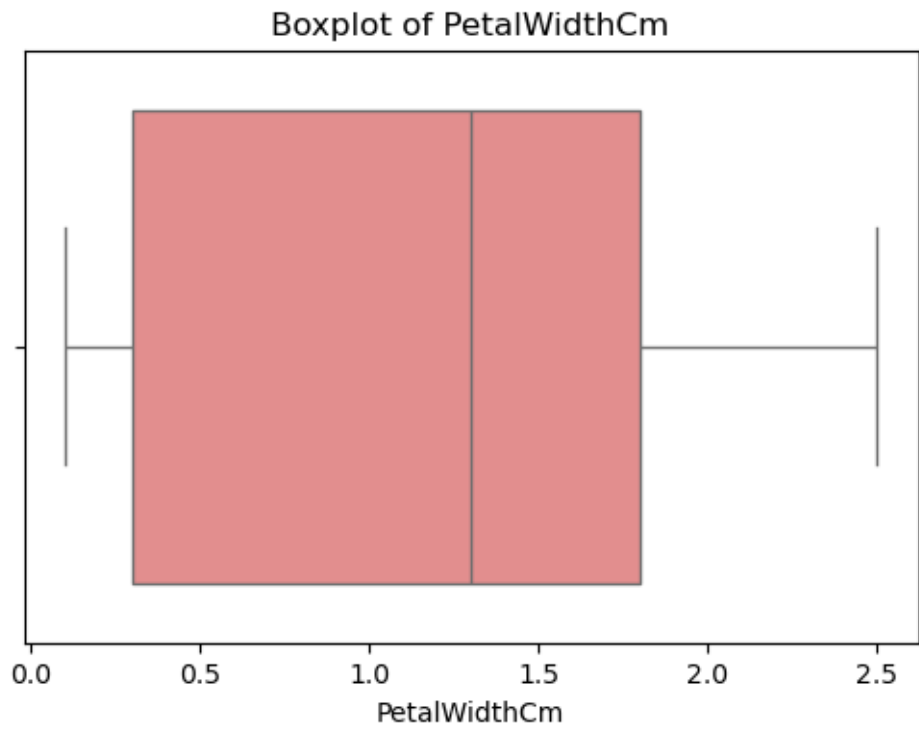


Boxplot of SepalWidthCm



Boxplot of PetalLengthCm





[ ]: