asmita-1-1

April 28, 2024

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
df=pd.read_csv("C:/Users/pavla/Downloads/Iris.csv")
df
```

[3]:
```
       Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
0       1            5.1           3.5            1.4           0.2
1       2            4.9           3.0            1.4           0.2
2       3            4.7           3.2            1.3           0.2
3       4            4.6           3.1            1.5           0.2
4       5            5.0           3.6            1.4           0.2
..    ...            ...           ...            ...           ...
145   146            6.7           3.0            5.2           2.3
146   147            6.3           2.5            5.0           1.9
147   148            6.5           3.0            5.2           2.0
148   149            6.2           3.4            5.4           2.3
149   150            5.9           3.0            5.1           1.8

            Species
0       Iris-setosa
1       Iris-setosa
2       Iris-setosa
3       Iris-setosa
4       Iris-setosa
..              ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 6 columns]
```

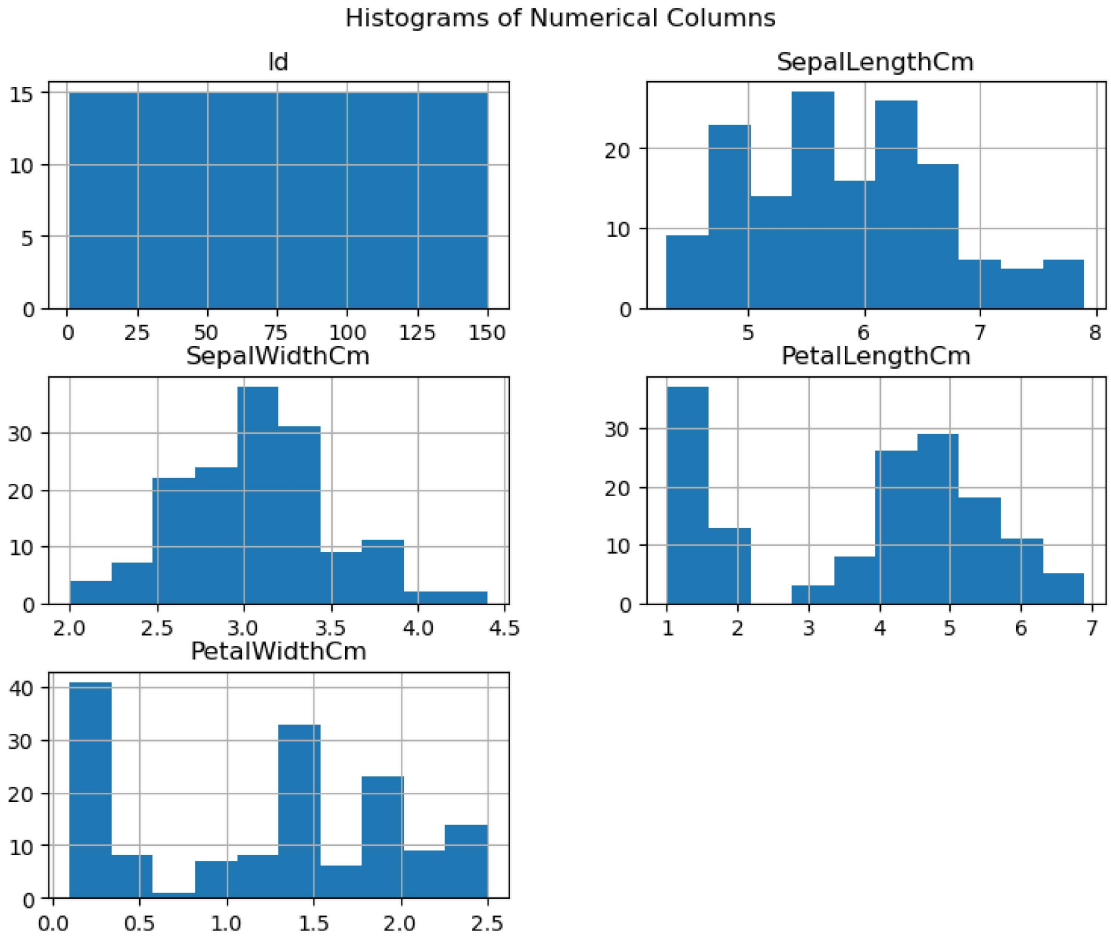
[4]:
```
# Display the first few rows of the dataset
print(df.head())
```

	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

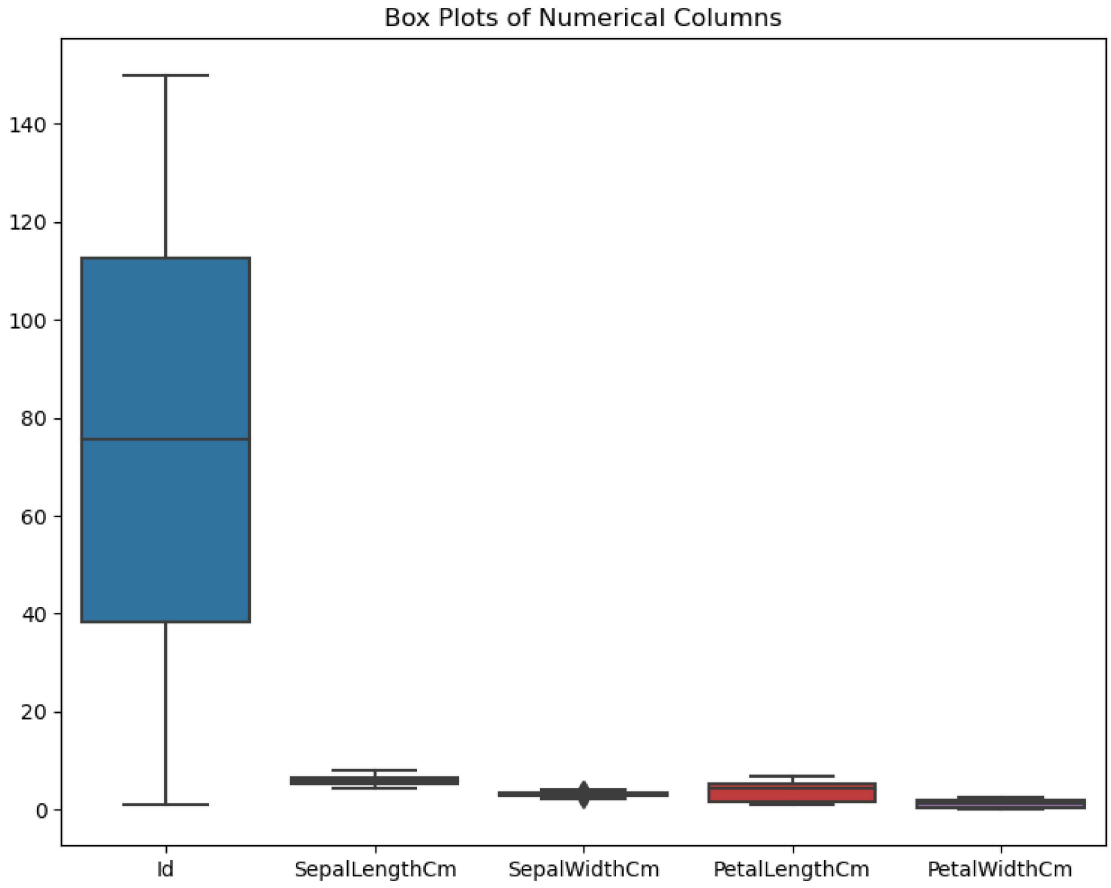
[5]:
```
# Summary statistics
summary_stats = df.describe()
print("\nSummary Statistics:\n", summary_stats)
```

```
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
```

[6]:
```
# Histograms for numerical columns
df.hist(figsize=(9, 7))
plt.suptitle("Histograms of Numerical Columns", y=0.95)
plt.show()
```



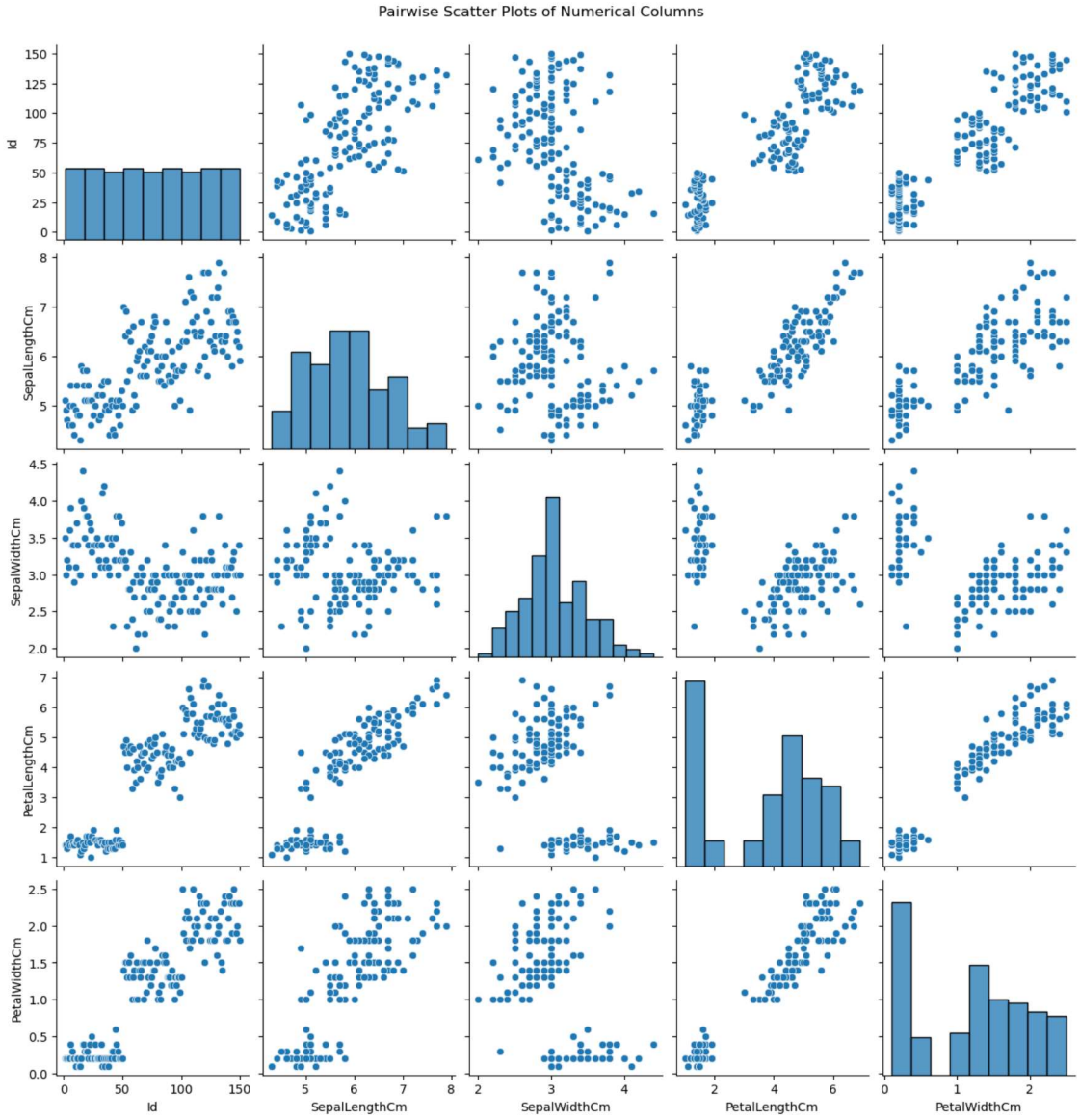
[7]:
```
# Box plots for numerical columns
plt.figure(figsize=(9, 7))
sns.boxplot(data=df)
plt.title("Box Plots of Numerical Columns")
plt.show()
```



[8]:
```
# Scatter plots for numerical columns
plt.figure(figsize=(9, 7))
sns.pairplot(df, kind="scatter")
plt.suptitle("Pairwise Scatter Plots of Numerical Columns", y=1.02)
plt.show()
```

```
C:\Users\pavla\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning:
The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

```
<Figure size 900x700 with 0 Axes>
```


Pairwise Scatter Plots of Numerical Columns
Id
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm

[ ]: