

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from google.colab import files
uploaded = files.upload()
```



↗ Choose files Iris.csv

- **Iris.csv**(text/csv) - 5107 bytes, last modified: 19/09/2019 - 100% done

Saving Iris.csv to Iris (1).csv

```
df=pd.read_csv("Iris (1).csv")
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	

Next steps:

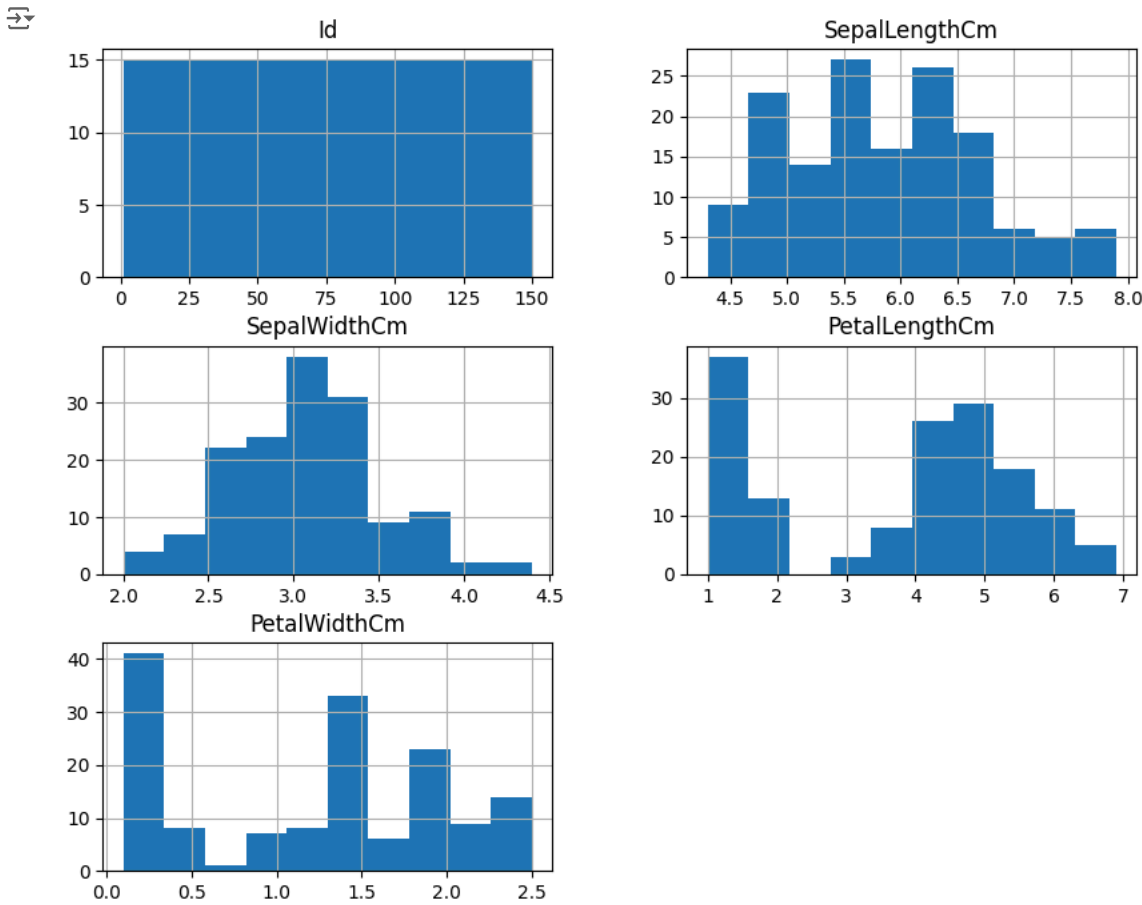
[Generate code with df](#)

 [View recommended plots](#)

```
df.isnull().sum()

# Check for duplicate rows
df.duplicated().sum()

# Check the distribution of each variable
df.hist(figsize=(10, 8), bins=10)
plt.show()
```



```
import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

numerical_df = df.select_dtypes(include=['float', 'int'])

# Calculate the correlation matrix
corr_matrix = numerical_df.corr()
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm")
plt.show()
# Explore the relationship between sepal length and petal length
sns.scatterplot(x="SepalLengthCm", y="PetalLengthCm", data=df)
plt.show()

# Group the data by species and calculate the mean of each variable
df.groupby("Species").mean()

# Create a boxplot of petal length for each species
sns.boxplot(x="Species", y="PetalLengthCm", data=df)
plt.show()
```

