

```
# importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv("/content/drive/MyDrive/Data/Iris.csv");
# displaying basic information about datatype
print(df.info())
```

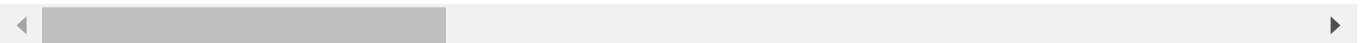
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id               150 non-null   int64
1   SepalLengthCm    150 non-null   float64
2   SepalWidthCm     150 non-null   float64
3   PetalLengthCm    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
```

```
df['Species'].value_counts()
```

```
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
Name: Species, dtype: int64
```

```
# displaying mean value of each column
x = df.mean()
print(x)
```

```
Id               75.500000
SepalLengthCm    5.843333
SepalWidthCm     3.054000
PetalLengthCm    3.758667
PetalWidthCm     1.198667
dtype: float64
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: FutureWarning: Dropping
```



```
#displaying stats about dataset
df.describe()
```

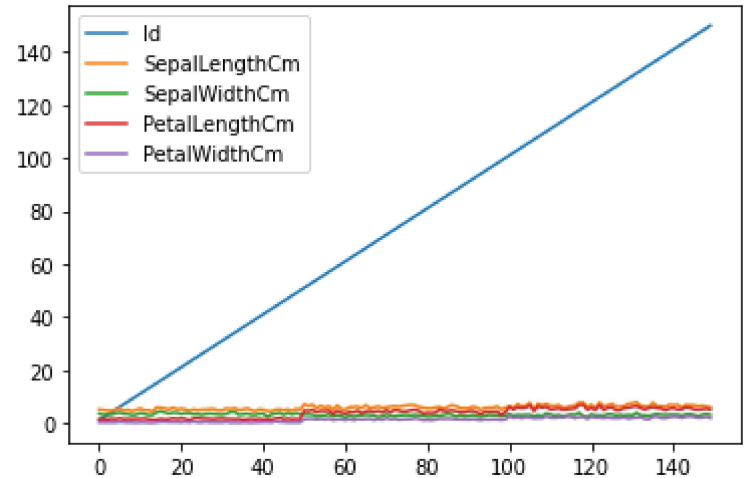
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.200000	5.100000	1.800000

```
# displaying pairwise correlation of all columns
print("correlation")
print(df.corr())
```

correlation		Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	\
Id		1.000000	0.716676	-0.397729	0.882747	
SepalLengthCm		0.716676	1.000000	-0.109369	0.871754	
SepalWidthCm		-0.397729	-0.109369	1.000000	-0.420516	
PetalLengthCm		0.882747	0.871754	-0.420516	1.000000	
PetalWidthCm		0.899759	0.817954	-0.356544	0.962757	
	PetalWidthCm					
Id		0.899759				
SepalLengthCm		0.817954				
SepalWidthCm		-0.356544				
PetalLengthCm		0.962757				
PetalWidthCm		1.000000				

```
# plotting diagrams
df.plot()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f738bb37fd0>



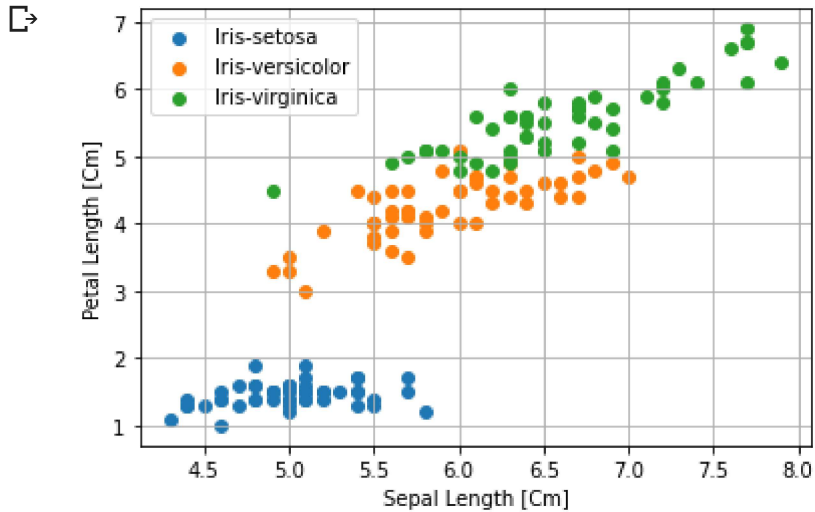
```
species = ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']
for i in range(3):
```

```

for i in range(5):
    x = df[df['Species']==species[i]]
    plt.scatter(x['SepalLengthCm'],x['PetalLengthCm'], label=species[i])

plt.xlabel("Sepal Length [Cm]")
plt.ylabel("Petal Length [Cm]")
plt.legend()
plt.grid()

```



```

# count on y axis and values on x axis
df['SepalLengthCm'].hist()

```

<matplotlib.axes._subplots.AxesSubplot at 0x7f7389bd1c10>

