

**In this lab, we will work with one of the toy datasets that you will be able to directly import and use in your notebook. This dataset is about the iris flower classification.**

```
In [16]: from sklearn.datasets import load_iris  
import numpy as np
```

## The Shape of the Data

```
In [17]: iris=load_iris()  
print(iris.data.shape)  
  
(150, 4)
```

## Features names

```
In [18]: print(iris.feature_names)  
  
['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
```

## Target names

```
In [19]: print(iris.target_names)  
  
['setosa' 'versicolor' 'virginica']
```

## Data

```
In [20]: print (iris.data)
[[5.1 3.5 5.5 2.1]
 [5.7 2.5 5. 2. ]
 [5.8 2.8 5.1 2.4]
 [6.4 3.2 5.3 2.3]
 [6.5 3. 5.5 1.8]
 [7.7 3.8 6.7 2.2]
 [7.7 2.6 6.9 2.3]
 [6. 2.2 5. 1.5]
 [6.9 3.2 5.7 2.3]
 [5.6 2.8 4.9 2. ]
 [7.7 2.8 6.7 2. ]
 [6.3 2.7 4.9 1.8]
 [6.7 3.3 5.7 2.1]
 [7.2 3.2 6. 1.8]
 [6.2 2.8 4.8 1.8]
 [6.1 3. 4.9 1.8]
 [6.4 2.8 5.6 2.1]
 [7.2 3. 5.8 1.6]
 [7.4 2.8 6.1 1.9]
 [7.9 3.8 6.4 2. ]
 ...]
```

## Target

```
In [7]: print(iris.target)
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2
2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2
2 2]
```

**Write a logic to print all the rows of iris data that has petallength > 1.5 and sepallength < 5.0**

```
In [11]: for i in range(len(iris.target)) :  
        if(iris.data[i][0]<5.0 and iris.data[i][2]>1.5):  
            print(iris.data[i])
```

```
[4.8 3.4 1.6 0.2]  
[4.8 3.4 1.9 0.2]  
[4.7 3.2 1.6 0.2]  
[4.8 3.1 1.6 0.2]  
[4.9 2.4 3.3 1. ]  
[4.9 2.5 4.5 1.7]
```

**Calculate the mean, median and standard deviation of each of the four columns. Understand the importance of these values in statistics.**

```
In [21]: data=iris.data  
print("mean =",np.mean(data,axis=0))  
print("median =",np.median(data,axis=0))  
print("stand deviation =",np.std(data,axis=0))
```

```
mean = [5.84333333 3.05733333 3.758      1.19933333]  
median = [5.8  3.    4.35 1.3 ]  
stand deviation = [0.82530129 0.43441097 1.75940407 0.75969263]
```

**Use the petal length (3rd) column of iris data to form a text array with categorical data, such that if petal length is:**

- a. Less than 3 --> 'Small'**
- b. Between 3 to 5 --> 'Medium'**
- c. Greater than 5 --> 'Large'**

**Print the newly created categorical array.**

In [4]: # arr= np.empty(150, dtype=str)

```
arr=[]
for i in range(len(iris.target)):
    if(iris.data[i][2]<3.0) :
        arr.append("small")
    elif(iris.data[i][2]>=3.0 and iris.data[i][2]<=5.0):
        arr.append("medium")
    if(iris.data[i][2]>5.0) :
        arr.append("large")
arr=np.array(arr)
print(arr)
```

```
['small' 'small' 'small' 'small' 'small' 'small' 'small' 'small' '
small'
'small' 'small' 'small' 'small' 'small' 'small' 'small' 'small' '
small'
'small' 'small' 'small' 'small' 'small' 'small' 'small' 'small' '
small'
'small' 'small' 'small' 'small' 'small' 'small' 'small' 'small' '
small'
'small' 'small' 'small' 'small' 'small' 'small' 'small' 'small' '
small'
'small' 'small' 'small' 'small' 'small' 'medium' 'medium' 'medium'
,
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'm
edium'
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'm
edium'
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'm
edium'
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'large' 'me
dium'
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'm
edium'
'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'medium' 'l
arge'
'large' 'large' 'large' 'large' 'large' 'medium' 'large' 'large'
'large'
'large' 'large' 'large' 'medium' 'large' 'large' 'large' 'large'
'large'
'medium' 'large' 'medium' 'large' 'medium' 'large' 'large' 'mediu
m'
'medium' 'large' 'large' 'large' 'large' 'large' 'large' 'large'
'large'
'large' 'large' 'medium' 'large' 'large' 'large' 'large' 'large'
'large'
'large' 'medium' 'large' 'large' 'large']
```

