## Import all the required Python Libraries.

```
import pandas as pd
import numpy as np
```

## Load the Dataset into pandas dataframe.

```
csv_url="/content/Iris.csv"
data=pd.read_csv(csv_url)
data
```

| | 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 |
| **...** | ... | ... | ... | ... | ... | ... |
| **145** | 146 | 6.7 | 3.0 | 5.2 | 2.3 | Iris-virginica |
| **146** | 147 | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |
| **147** | 148 | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
| **148** | 149 | 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
| **149** | 150 | 5.9 | 3.0 | 5.1 | 1.8 | Iris-virginica |

150 rows × 6 columns

## Data Preprocessing

```
data.isnull()
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species |
|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False |
| **1** | False | False | False | False | False | False |
| **2** | False | False | False | False | False | False |
| **3** | False | False | False | False | False | False |
| **4** | False | False | False | False | False | False |
| **...** | ... | ... | ... | ... | ... | ... |
| **145** | False | False | False | False | False | False |
| **146** | False | False | False | False | False | False |
| **147** | False | False | False | False | False | False |
| **148** | False | False | False | False | False | False |
| **149** | False | False | False | False | False | False |

150 rows × 6 columns

```
dataset=pd.DataFrame(data)
dataset
```

| | 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 |
| **...** | ... | ... | ... | ... | ... | ... |

```
data.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 | 3.300000 | 5.100000 | 1.800000 |
| **max** | 150.000000 | 7.900000 | 4.400000 | 6.900000 | 2.500000 |

```
dataset.head(n=5)
```

| | 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.tail(n=5)
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species |
|---|---|---|---|---|---|---|
| **145** | 146 | 6.7 | 3.0 | 5.2 | 2.3 | Iris-virginica |
| **146** | 147 | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |
| **147** | 148 | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
| **148** | 149 | 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
| **149** | 150 | 5.9 | 3.0 | 5.1 | 1.8 | Iris-virginica |

```
dataset.index
```

```
RangeIndex(start=0, stop=150, step=1)
```

```
dataset.columns
```

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

```
dataset.shape
```

```
(150, 6)
```

```
dataset.dtypes
```

```
Id                int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

```
dataset.columns.values
```

```
array(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm',
       'PetalWidthCm', 'Species'], dtype=object)
```

```
dataset.describe(include='all')
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species |
|---|---|---|---|---|---|---|
| **count** | 150.000000 | 150.000000 | 150.000000 | 150.000000 | 150.000000 | 150 |
| **unique** | NaN | NaN | NaN | NaN | NaN | 3 |
| **top** | NaN | NaN | NaN | NaN | NaN | Iris-setosa |
| **freq** | NaN | NaN | NaN | NaN | NaN | 50 |
| **mean** | 75.500000 | 5.843333 | 3.054000 | 3.758667 | 1.198667 | NaN |
| **std** | 43.445368 | 0.828066 | 0.433594 | 1.764420 | 0.763161 | NaN |
| **min** | 1.000000 | 4.300000 | 2.000000 | 1.000000 | 0.100000 | NaN |
| **25%** | 38.250000 | 5.100000 | 2.800000 | 1.600000 | 0.300000 | NaN |
| **50%** | 75.500000 | 5.800000 | 3.000000 | 4.350000 | 1.300000 | NaN |
| **75%** | 112.750000 | 6.400000 | 3.300000 | 5.100000 | 1.800000 | NaN |
| **max** | 150.000000 | 7.900000 | 4.400000 | 6.900000 | 2.500000 | NaN |

```
dataset['PetalWidthCm']
```

```
0      0.2
1      0.2
2      0.2
3      0.2
4      0.2
      ...
145    2.3
146    1.9
147    2.0
148    2.3
149    1.8
Name: PetalWidthCm, Length: 150, dtype: float64
```

```
dataset.sort_index(axis=1,ascending=False)
```

| | Species | SepalWidthCm | SepalLengthCm | PetalWidthCm | PetalLengthCm | Id |
|---|---|---|---|---|---|---|
| **0** | Iris-setosa | 3.5 | 5.1 | 0.2 | 1.4 | 1 |
| **1** | Iris-setosa | 3.0 | 4.9 | 0.2 | 1.4 | 2 |
| **2** | Iris-setosa | 3.2 | 4.7 | 0.2 | 1.3 | 3 |
| **3** | Iris-setosa | 3.1 | 4.6 | 0.2 | 1.5 | 4 |
| **4** | Iris-setosa | 3.6 | 5.0 | 0.2 | 1.4 | 5 |
| **...** | ... | ... | ... | ... | ... | ... |
| **145** | Iris-virginica | 3.0 | 6.7 | 2.3 | 5.2 | 146 |
| **146** | Iris-virginica | 2.5 | 6.3 | 1.9 | 5.0 | 147 |
| **147** | Iris-virginica | 3.0 | 6.5 | 2.0 | 5.2 | 148 |
| **148** | Iris-virginica | 3.4 | 6.2 | 2.3 | 5.4 | 149 |
| **149** | Iris-virginica | 3.0 | 5.9 | 1.8 | 5.1 | 150 |

150 rows × 6 columns

```
dataset.sort_values(by='PetalWidthCm')
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species |
|---|---|---|---|---|---|---|
| **32** | 33 | 5.2 | 4.1 | 1.5 | 0.1 | Iris-setosa |
| **13** | 14 | 4.3 | 3.0 | 1.1 | 0.1 | Iris-setosa |
| **37** | 38 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| **9** | 10 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| **12** | 13 | 4.8 | 3.0 | 1.4 | 0.1 | Iris-setosa |
| **...** | ... | ... | ... | ... | ... | ... |

```
data.isna().sum()
```

```
Id              0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64Id              0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

## Data Formatting and Normalizations

```
data.dtypes
```

```
Id                int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

```
data['SepalLengthCm']=data['SepalLengthCm'].astype('int')
print(data.dtypes)
```

```
Id                int64
SepalLengthCm     int64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

```
dataset['Species'].replace(['Iris-setosa','Iris-virginica'],[0,1],inplace=True)
dataset
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species |
|---|---|---|---|---|---|---|
| **0** | 1 | 5 | 3.5 | 1.4 | 0.2 | 0 |
| **1** | 2 | 4 | 3.0 | 1.4 | 0.2 | 0 |
| **2** | 3 | 4 | 3.2 | 1.3 | 0.2 | 0 |
| **3** | 4 | 4 | 3.1 | 1.5 | 0.2 | 0 |
| **4** | 5 | 5 | 3.6 | 1.4 | 0.2 | 0 |
| **...** | ... | ... | ... | ... | ... | ... |
| **145** | 146 | 6 | 3.0 | 5.2 | 2.3 | 1 |
| **146** | 147 | 6 | 2.5 | 5.0 | 1.9 | 1 |
| **147** | 148 | 6 | 3.0 | 5.2 | 2.0 | 1 |
| **148** | 149 | 6 | 3.4 | 5.4 | 2.3 | 1 |
| **149** | 150 | 5 | 3.0 | 5.1 | 1.8 | 1 |

150 rows × 6 columns

## Turn categorical variables into quantitative variables

```
data=pd.get_dummies(data,columns=['Species'])
data
```

| | Id | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm | Species_0 | Specie |
|---|---|---|---|---|---|---|---|
| **0** | 1 | 5 | 3.5 | 1.4 | 0.2 | 1 | |
| **1** | 2 | 4 | 3.0 | 1.4 | 0.2 | 1 | |
| **2** | 3 | 4 | 3.2 | 1.3 | 0.2 | 1 | |
| **3** | 4 | 4 | 3.1 | 1.5 | 0.2 | 1 | |
| **4** | 5 | 5 | 3.6 | 1.4 | 0.2 | 1 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **145** | 146 | 6 | 3.0 | 5.2 | 2.3 | 0 | |
| **146** | 147 | 6 | 2.5 | 5.0 | 1.9 | 0 | |
| **147** | 148 | 6 | 3.0 | 5.2 | 2.0 | 0 | |
| **148** | 149 | 6 | 3.4 | 5.4 | 2.3 | 0 | |
| **149** | 150 | 5 | 3.0 | 5.1 | 1.8 | 0 | |



✓ 0s completed at 10:26 PM