

wl9ec4peu

April 29, 2025

```
[17]: import pandas as pd
import numpy as np
```

```
[19]: # Downloading dataset from https://www.kaggle.com/c/titanic/data

df = pd.read_csv('Iris.csv')
```

```
[6]: # First 5 rows
df.head()
```

```
[6]:   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
```

```
[7]: # Last 5 rows
df.tail()
```

```
[7]:   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
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
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

```
[8]: df.isnull()
```

```
[8]:
```

	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 x 6 columns]

```
[9]: df.isnull().sum()
```

```
[9]: Id                0
SepalLengthCm        0
SepalWidthCm         0
PetalLengthCm        0
PetalWidthCm         0
Species              0
dtype: int64
```

```
[10]: df.notnull()
```

```
[10]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
..
145	True	True	True	True	True	True
146	True	True	True	True	True	True
147	True	True	True	True	True	True
148	True	True	True	True	True	True
149	True	True	True	True	True	True

[150 rows x 6 columns]

```
[11]: df.notnull().sum()
```

```
[11]: Id                150
SepalLengthCm        150
SepalWidthCm         150
```

```
PetalLengthCm    150
PetalWidthCm     150
Species          150
dtype: int64
```

```
[12]: df.describe()
```

```
[12]:
```

	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

```
[13]: # Column Names
df.columns
```

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

```
[14]: # Data Types
df.dtypes
```

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

```
[15]: # data type count
df.dtypes.value_counts()
```

```
[15]: float64    4
int64        1
object       1
Name: count, dtype: int64
```

```
[16]: # Shape of data (row, column)
df.shape
```

```
[16]: (150, 6)
```

[]:

[]:

[]:

[]:

[]: