

```
In [2]: import pandas as pd
import numpy as np

a = pd.read_csv("Iris_missingdata.csv")
a.head()
```

Out[2]:

	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	NaN	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

```
In [3]: a.head(15)
```

Out[3]:

	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	NaN	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	NaN	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	NaN	3.0	1.1	0.1	Iris-setosa
14	15	5.8	NaN	1.2	0.2	Iris-setosa

```
In [4]: a.tail(10)
```

```
Out[4]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
140	141	6.7	3.1	5.6	2.4	Iris-virginica
141	142	6.9	3.1	5.1	2.3	Iris-virginica
142	143	NaN	2.7	5.1	NaN	Iris-virginica
143	144	6.8	3.2	5.9	2.3	Iris-virginica
144	145	6.7	3.3	5.7	2.5	Iris-virginica
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	NaN	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

```
In [5]: a.index
```

```
Out[5]: RangeIndex(start=0, stop=150, step=1)
```

```
In [7]: a.columns
```

```
Out[7]: Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',  
              'Species'],  
              dtype='object')
```

```
In [8]: a.shape
```

```
Out[8]: (150, 6)
```

```
In [10]: a.dtypes
```

```
Out[10]: Id                int64  
SepalLengthCm            float64  
SepalWidthCm             float64  
PetalLengthCm            float64  
PetalWidthCm             float64  
Species                  object  
dtype: object
```

```
In [11]: a['Species']
```

```
Out[11]: 0      Iris-setosa
          1      Iris-setosa
          2      Iris-setosa
          3      Iris-setosa
          4      Iris-setosa
          ...
          145    Iris-virginica
          146    Iris-virginica
          147    Iris-virginica
          148    Iris-virginica
          149    Iris-virginica
          Name: Species, Length: 150, dtype: object
```

```
In [12]: a.iloc[10]
```

```
Out[12]: Id      11
          SepalLengthCm    5.4
          SepalWidthCm     3.7
          PetalLengthCm    1.5
          PetalWidthCm     0.2
          Species      Iris-setosa
          Name: 10, dtype: object
```

```
In [13]: a[0:3]
```

```
Out[13]:
```

	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

```
In [14]: a.loc[:,["Species","Id"]]
```

```
Out[14]:
```

	Species	Id
0	Iris-setosa	1
1	Iris-setosa	2
2	Iris-setosa	3
3	Iris-setosa	4
4	Iris-setosa	5
...
145	Iris-virginica	146
146	Iris-virginica	147
147	Iris-virginica	148
148	Iris-virginica	149
149	Iris-virginica	150

150 rows × 2 columns

```
In [16]: a.iloc[:10,:]
```

```
Out[16]:
```

	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	NaN	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	NaN	0.1	Iris-setosa

```
In [18]: a.iloc[:, :10]
```

```
Out[18]:
```

	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	NaN	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	NaN	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

```
In [19]: a.iloc[:15,:15]
```

```
Out[19]:
```

	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	NaN	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	NaN	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	NaN	3.0	1.1	0.1	Iris-setosa
14	15	5.8	NaN	1.2	0.2	Iris-setosa

```
In [20]: a.sort_values("SepalLengthCm")
```

```
Out[20]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
42	43	4.4	3.2	1.3	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
38	39	4.4	3.0	1.3	0.2	Iris-setosa
41	42	4.5	2.3	1.3	0.3	Iris-setosa
47	48	4.6	3.2	1.4	0.2	Iris-setosa
...
82	83	NaN	2.7	3.9	1.2	Iris-versicolor
99	100	NaN	2.8	4.1	1.3	Iris-versicolor
125	126	NaN	3.2	6.0	1.8	Iris-virginica
130	131	NaN	2.8	6.1	1.9	Iris-virginica
142	143	NaN	2.7	5.1	NaN	Iris-virginica

150 rows × 6 columns

```
In [21]: a.sort_index()
```

Out[21]:

	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	NaN	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	NaN	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

```
In [23]: a.sort_values( by="PetalWidthCm", ascending=False )
```

Out[23]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
109	110	7.2	3.6	6.1	2.5	Iris-virginica
144	145	6.7	3.3	5.7	2.5	Iris-virginica
100	101	6.3	3.3	6.0	2.5	Iris-virginica
140	141	6.7	3.1	5.6	2.4	Iris-virginica
114	115	5.8	2.8	5.1	2.4	Iris-virginica
...
71	72	6.1	2.8	NaN	NaN	Iris-versicolor
97	98	6.2	2.9	4.3	NaN	Iris-versicolor
128	129	6.4	2.8	5.6	NaN	Iris-virginica
139	140	6.9	3.1	5.4	NaN	Iris-virginica
142	143	NaN	2.7	5.1	NaN	Iris-virginica

150 rows × 6 columns

```
In [25]: a.isnull()
```

```
Out[25]:
```

	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	True	False	False	False	False
4	False	False	False	False	False	False
...
145	False	False	False	False	False	False
146	False	False	True	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

```
In [26]: a.fillna(0)
```

```
Out[26]:
```

	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	0.0	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	0.0	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

```
In [4]: a.interpolate(method='linear', limit_direction='forward')
```

```
Out[4]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.10	3.5	1.4	0.2	Iris-setosa
1	2	4.90	3.0	1.4	0.2	Iris-setosa
2	3	4.70	3.2	1.3	0.2	Iris-setosa
3	4	4.85	3.1	1.5	0.2	Iris-setosa
4	5	5.00	3.6	1.4	0.2	Iris-setosa
...
145	146	6.70	3.0	5.2	2.3	Iris-virginica
146	147	6.30	3.0	5.0	1.9	Iris-virginica
147	148	6.50	3.0	5.2	2.0	Iris-virginica
148	149	6.20	3.4	5.4	2.3	Iris-virginica
149	150	5.90	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

```
In [5]: a.describe()
```

```
Out[5]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	139.000000	143.000000	142.000000	141.000000
mean	75.500000	5.841727	3.050350	3.780282	1.198582
std	43.445368	0.820472	0.432789	1.777728	0.763262
min	1.000000	4.400000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.400000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
In [ ]:
```



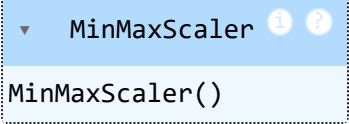
```
In [4]: import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler, StandardScaler

data = pd.read_csv('OneDrive/Documents/Iris_missingdata.csv')
data.dropna(inplace=True)
```

```
In [5]: numerical_features = data.select_dtypes(include=[np.number]).columns
```

```
In [6]: scaler = MinMaxScaler() # for MinMaxScaler
```

```
In [7]: scaler.fit(data[numerical_features])
```

```
Out[7]: 
  ▼ MinMaxScaler ⓘ ?
  MinMaxScaler()
```

```
In [8]: normalized_data = scaler.transform(data[numerical_features])
```

```
In [9]: data[numerical_features] = normalized_data
```

```
In [10]: data.to_csv('normalized_data.csv', index=False)
```

```
In [12]: data.head(150)
```

Out[12]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	0.000000	0.200000	0.625000	0.067797	0.041667	Iris-setosa
1	0.006711	0.142857	0.416667	0.067797	0.041667	Iris-setosa
2	0.013423	0.085714	0.500000	0.050847	0.041667	Iris-setosa
4	0.026846	0.171429	0.666667	0.067797	0.041667	Iris-setosa
5	0.033557	0.285714	0.791667	0.118644	0.125000	Iris-setosa
...
144	0.966443	0.657143	0.541667	0.796610	1.000000	Iris-virginica
145	0.973154	0.657143	0.416667	0.711864	0.916667	Iris-virginica
147	0.986577	0.600000	0.416667	0.711864	0.791667	Iris-virginica
148	0.993289	0.514286	0.583333	0.745763	0.916667	Iris-virginica
149	1.000000	0.428571	0.416667	0.694915	0.708333	Iris-virginica

119 rows × 6 columns

In []: