Assign-1

February 27, 2023

[7]:
```
import pandas as pd

csv_url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.
 ↪data'
data = pd.read_csv(csv_url)
data
```

[7]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
..	…	…	…	…	…
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

[149 rows x 5 columns]

[37]:
```
dataSet = pd.DataFrame(data)
dataSet
```

[37]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
..	…	…	…	…	…
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

```
[149 rows x 5 columns]
```

[19]:
```
dataSet.isnull()
```

[19]:
```
       5.1    3.5    1.4    0.2  Iris-setosa
0    False  False  False  False        False
1    False  False  False  False        False
2    False  False  False  False        False
3    False  False  False  False        False
4    False  False  False  False        False
..     …    …    …    …          …
144  False  False  False  False        False
145  False  False  False  False        False
146  False  False  False  False        False
147  False  False  False  False        False
148  False  False  False  False        False

[149 rows x 5 columns]
```

[20]:
```
dataSet.isna().sum()
```

[20]:
```
5.1            0
3.5            0
1.4            0
0.2            0
Iris-setosa    0
dtype: int64
```

[21]:
```
dataSet.describe()
```

[21]:
```
              5.1         3.5         1.4         0.2
count  149.000000  149.000000  149.000000  149.000000
mean     5.848322    3.051007    3.774497    1.205369
std      0.828594    0.433499    1.759651    0.761292
min      4.300000    2.000000    1.000000    0.100000
25%      5.100000    2.800000    1.600000    0.300000
50%      5.800000    3.000000    4.400000    1.300000
75%      6.400000    3.300000    5.100000    1.800000
max      7.900000    4.400000    6.900000    2.500000
```

[22]:
```
type = dataSet.dtypes
print(type)
```

```
5.1            float64
3.5            float64
1.4            float64
```

```
0.2             float64
Iris-setosa      object
dtype: object
```

[23]:
```
print(dataSet.shape)
```

```
(149, 5)
```

[34]:
```
dataSet['5.1'] = dataSet['5.1'].astype("str")
print(data.dtypes)
```

```
5.1              object
3.5             float64
1.4             float64
0.2             float64
Iris-setosa      object
dtype: object
```

[42]:
```
dataSet['Iris-setosa'].replace(['Iris-setosa', 'Iris-virginica'],
                        [0, 1], inplace=True)
```

[43]:
```
dataSet
```

[43]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	0
1	4.7	3.2	1.3	0.2	0
2	4.6	3.1	1.5	0.2	0
3	5.0	3.6	1.4	0.2	0
4	5.4	3.9	1.7	0.4	0
..	…	…	…	…	…
144	6.7	3.0	5.2	2.3	1
145	6.3	2.5	5.0	1.9	1
146	6.5	3.0	5.2	2.0	1
147	6.2	3.4	5.4	2.3	1
148	5.9	3.0	5.1	1.8	1

[149 rows x 5 columns]