10 minutes to pandas

Written by: M.Danish Azeem Date: 05.12.2023 Email:m danishazeem365@gmail.com

```
import numpy as np
import pandas as pd
import seaborn as sns

df = sns.load_dataset("iris")
df
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             5.1          3.5           1.4          0.2     setosa
1             4.9          3.0           1.4          0.2     setosa
2             4.7          3.2           1.3          0.2     setosa
3             4.6          3.1           1.5          0.2     setosa
4             5.0          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
147           6.5          3.0           5.2          2.0  virginica
148           6.2          3.4           5.4          2.3  virginica
149           5.9          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

Object creation

```
s = pd.Series([1, 3, 5, np.nan, 6, 8])
s
```

```
0    1.0
1    3.0
2    5.0
3    NaN
4    6.0
5    8.0
dtype: float64
```

```
dates = pd.date_range("20130101", periods=6)
dates
```

```
DatetimeIndex(['2013-01-01', '2013-01-02', '2013-01-03', '2013-01-04',
               '2013-01-05', '2013-01-06'],
              dtype='datetime64[ns]', freq='D')
```

```
df1 = pd.DataFrame(np.random.randn(6, 4), index=dates,
columns=list("ABCD"))
df1
```

```
                   A         B         C         D
2013-01-01 -0.242641 -0.356031 -0.274859 -1.259400
2013-01-02 -0.125133 -0.738373  0.835656 -0.421399
2013-01-03  1.691598 -0.383428  1.504235  0.565867
2013-01-04 -0.883170 -1.266696  2.326153  2.455350
2013-01-05 -0.119623  0.946998  2.021825  1.850227
2013-01-06  1.209312 -1.545526 -0.264337 -0.158599
```

```
df2 = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),
        "D": np.array([3] * 4, dtype="int32"),
        "E": pd.Categorical(["test", "train", "test", "train"]),
        "F": "foo",
    }
)


df2
```

```
     A          B    C  D      E    F
0  1.0 2013-01-02  1.0  3   test  foo
1  1.0 2013-01-02  1.0  3  train  foo
2  1.0 2013-01-02  1.0  3   test  foo
3  1.0 2013-01-02  1.0  3  train  foo
```

```
df = sns.load_dataset("iris")
df
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             5.1          3.5           1.4          0.2     setosa
1             4.9          3.0           1.4          0.2     setosa
2             4.7          3.2           1.3          0.2     setosa
3             4.6          3.1           1.5          0.2     setosa
4             5.0          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
147           6.5          3.0           5.2          2.0  virginica
148           6.2          3.4           5.4          2.3  virginica
149           5.9          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

```
df.dtypes
```

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

Viewing data

```
df.head()
```

```
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa
```

```
df.tail()
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
147           6.5          3.0           5.2          2.0  virginica
148           6.2          3.4           5.4          2.3  virginica
149           5.9          3.0           5.1          1.8  virginica
```

```
df.index
```

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

```
df.columns
```

```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

```
df2.dtypes
```

```
A          float64
B    datetime64[s]
C          float32
D            int32
E         category
F           object
dtype: object
```

```
df.dtypes
```

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

```
df.to_numpy()
```

```
array([[5.1, 3.5, 1.4, 0.2, 'setosa'],
       [4.9, 3.0, 1.4, 0.2, 'setosa'],
       [4.7, 3.2, 1.3, 0.2, 'setosa'],
       [4.6, 3.1, 1.5, 0.2, 'setosa'],
       [5.0, 3.6, 1.4, 0.2, 'setosa'],
       [5.4, 3.9, 1.7, 0.4, 'setosa'],
       [4.6, 3.4, 1.4, 0.3, 'setosa'],
       [5.0, 3.4, 1.5, 0.2, 'setosa'],
       [4.4, 2.9, 1.4, 0.2, 'setosa'],
       [4.9, 3.1, 1.5, 0.1, 'setosa'],
       [5.4, 3.7, 1.5, 0.2, 'setosa'],
       [4.8, 3.4, 1.6, 0.2, 'setosa'],
       [4.8, 3.0, 1.4, 0.1, 'setosa'],
       [4.3, 3.0, 1.1, 0.1, 'setosa'],
       [5.8, 4.0, 1.2, 0.2, 'setosa'],
       [5.7, 4.4, 1.5, 0.4, 'setosa'],
       [5.4, 3.9, 1.3, 0.4, 'setosa'],
       [5.1, 3.5, 1.4, 0.3, 'setosa'],
       [5.7, 3.8, 1.7, 0.3, 'setosa'],
       [5.1, 3.8, 1.5, 0.3, 'setosa'],
       [5.4, 3.4, 1.7, 0.2, 'setosa'],
       [5.1, 3.7, 1.5, 0.4, 'setosa'],
       [4.6, 3.6, 1.0, 0.2, 'setosa'],
       [5.1, 3.3, 1.7, 0.5, 'setosa'],
       [4.8, 3.4, 1.9, 0.2, 'setosa'],
       [5.0, 3.0, 1.6, 0.2, 'setosa'],
       [5.0, 3.4, 1.6, 0.4, 'setosa'],
       [5.2, 3.5, 1.5, 0.2, 'setosa'],
       [5.2, 3.4, 1.4, 0.2, 'setosa'],
       [4.7, 3.2, 1.6, 0.2, 'setosa'],
       [4.8, 3.1, 1.6, 0.2, 'setosa'],
       [5.4, 3.4, 1.5, 0.4, 'setosa'],
       [5.2, 4.1, 1.5, 0.1, 'setosa'],
       [5.5, 4.2, 1.4, 0.2, 'setosa'],
       [4.9, 3.1, 1.5, 0.2, 'setosa'],
       [5.0, 3.2, 1.2, 0.2, 'setosa'],
       [5.5, 3.5, 1.3, 0.2, 'setosa'],
       [4.9, 3.6, 1.4, 0.1, 'setosa'],
       [4.4, 3.0, 1.3, 0.2, 'setosa'],
       [5.1, 3.4, 1.5, 0.2, 'setosa'],
       [5.0, 3.5, 1.3, 0.3, 'setosa'],
```

```
       [4.5, 2.3, 1.3, 0.3, 'setosa'],
       [4.4, 3.2, 1.3, 0.2, 'setosa'],
       [5.0, 3.5, 1.6, 0.6, 'setosa'],
       [5.1, 3.8, 1.9, 0.4, 'setosa'],
       [4.8, 3.0, 1.4, 0.3, 'setosa'],
       [5.1, 3.8, 1.6, 0.2, 'setosa'],
       [4.6, 3.2, 1.4, 0.2, 'setosa'],
       [5.3, 3.7, 1.5, 0.2, 'setosa'],
       [5.0, 3.3, 1.4, 0.2, 'setosa'],
       [7.0, 3.2, 4.7, 1.4, 'versicolor'],
       [6.4, 3.2, 4.5, 1.5, 'versicolor'],
       [6.9, 3.1, 4.9, 1.5, 'versicolor'],
       [5.5, 2.3, 4.0, 1.3, 'versicolor'],
       [6.5, 2.8, 4.6, 1.5, 'versicolor'],
       [5.7, 2.8, 4.5, 1.3, 'versicolor'],
       [6.3, 3.3, 4.7, 1.6, 'versicolor'],
       [4.9, 2.4, 3.3, 1.0, 'versicolor'],
       [6.6, 2.9, 4.6, 1.3, 'versicolor'],
       [5.2, 2.7, 3.9, 1.4, 'versicolor'],
       [5.0, 2.0, 3.5, 1.0, 'versicolor'],
       [5.9, 3.0, 4.2, 1.5, 'versicolor'],
       [6.0, 2.2, 4.0, 1.0, 'versicolor'],
       [6.1, 2.9, 4.7, 1.4, 'versicolor'],
       [5.6, 2.9, 3.6, 1.3, 'versicolor'],
       [6.7, 3.1, 4.4, 1.4, 'versicolor'],
       [5.6, 3.0, 4.5, 1.5, 'versicolor'],
       [5.8, 2.7, 4.1, 1.0, 'versicolor'],
       [6.2, 2.2, 4.5, 1.5, 'versicolor'],
       [5.6, 2.5, 3.9, 1.1, 'versicolor'],
       [5.9, 3.2, 4.8, 1.8, 'versicolor'],
       [6.1, 2.8, 4.0, 1.3, 'versicolor'],
       [6.3, 2.5, 4.9, 1.5, 'versicolor'],
       [6.1, 2.8, 4.7, 1.2, 'versicolor'],
       [6.4, 2.9, 4.3, 1.3, 'versicolor'],
       [6.6, 3.0, 4.4, 1.4, 'versicolor'],
       [6.8, 2.8, 4.8, 1.4, 'versicolor'],
       [6.7, 3.0, 5.0, 1.7, 'versicolor'],
       [6.0, 2.9, 4.5, 1.5, 'versicolor'],
       [5.7, 2.6, 3.5, 1.0, 'versicolor'],
       [5.5, 2.4, 3.8, 1.1, 'versicolor'],
       [5.5, 2.4, 3.7, 1.0, 'versicolor'],
       [5.8, 2.7, 3.9, 1.2, 'versicolor'],
       [6.0, 2.7, 5.1, 1.6, 'versicolor'],
       [5.4, 3.0, 4.5, 1.5, 'versicolor'],
       [6.0, 3.4, 4.5, 1.6, 'versicolor'],
       [6.7, 3.1, 4.7, 1.5, 'versicolor'],
       [6.3, 2.3, 4.4, 1.3, 'versicolor'],
       [5.6, 3.0, 4.1, 1.3, 'versicolor'],
       [5.5, 2.5, 4.0, 1.3, 'versicolor'],
```

```
       [5.5, 2.6, 4.4, 1.2, 'versicolor'],
       [6.1, 3.0, 4.6, 1.4, 'versicolor'],
       [5.8, 2.6, 4.0, 1.2, 'versicolor'],
       [5.0, 2.3, 3.3, 1.0, 'versicolor'],
       [5.6, 2.7, 4.2, 1.3, 'versicolor'],
       [5.7, 3.0, 4.2, 1.2, 'versicolor'],
       [5.7, 2.9, 4.2, 1.3, 'versicolor'],
       [6.2, 2.9, 4.3, 1.3, 'versicolor'],
       [5.1, 2.5, 3.0, 1.1, 'versicolor'],
       [5.7, 2.8, 4.1, 1.3, 'versicolor'],
       [6.3, 3.3, 6.0, 2.5, 'virginica'],
       [5.8, 2.7, 5.1, 1.9, 'virginica'],
       [7.1, 3.0, 5.9, 2.1, 'virginica'],
       [6.3, 2.9, 5.6, 1.8, 'virginica'],
       [6.5, 3.0, 5.8, 2.2, 'virginica'],
       [7.6, 3.0, 6.6, 2.1, 'virginica'],
       [4.9, 2.5, 4.5, 1.7, 'virginica'],
       [7.3, 2.9, 6.3, 1.8, 'virginica'],
       [6.7, 2.5, 5.8, 1.8, 'virginica'],
       [7.2, 3.6, 6.1, 2.5, 'virginica'],
       [6.5, 3.2, 5.1, 2.0, 'virginica'],
       [6.4, 2.7, 5.3, 1.9, 'virginica'],
       [6.8, 3.0, 5.5, 2.1, 'virginica'],
       [5.7, 2.5, 5.0, 2.0, 'virginica'],
       [5.8, 2.8, 5.1, 2.4, 'virginica'],
       [6.4, 3.2, 5.3, 2.3, 'virginica'],
       [6.5, 3.0, 5.5, 1.8, 'virginica'],
       [7.7, 3.8, 6.7, 2.2, 'virginica'],
       [7.7, 2.6, 6.9, 2.3, 'virginica'],
       [6.0, 2.2, 5.0, 1.5, 'virginica'],
       [6.9, 3.2, 5.7, 2.3, 'virginica'],
       [5.6, 2.8, 4.9, 2.0, 'virginica'],
       [7.7, 2.8, 6.7, 2.0, 'virginica'],
       [6.3, 2.7, 4.9, 1.8, 'virginica'],
       [6.7, 3.3, 5.7, 2.1, 'virginica'],
       [7.2, 3.2, 6.0, 1.8, 'virginica'],
       [6.2, 2.8, 4.8, 1.8, 'virginica'],
       [6.1, 3.0, 4.9, 1.8, 'virginica'],
       [6.4, 2.8, 5.6, 2.1, 'virginica'],
       [7.2, 3.0, 5.8, 1.6, 'virginica'],
       [7.4, 2.8, 6.1, 1.9, 'virginica'],
       [7.9, 3.8, 6.4, 2.0, 'virginica'],
       [6.4, 2.8, 5.6, 2.2, 'virginica'],
       [6.3, 2.8, 5.1, 1.5, 'virginica'],
       [6.1, 2.6, 5.6, 1.4, 'virginica'],
       [7.7, 3.0, 6.1, 2.3, 'virginica'],
       [6.3, 3.4, 5.6, 2.4, 'virginica'],
       [6.4, 3.1, 5.5, 1.8, 'virginica'],
       [6.0, 3.0, 4.8, 1.8, 'virginica'],
```

```
       [6.9, 3.1, 5.4, 2.1, 'virginica'],
       [6.7, 3.1, 5.6, 2.4, 'virginica'],
       [6.9, 3.1, 5.1, 2.3, 'virginica'],
       [5.8, 2.7, 5.1, 1.9, 'virginica'],
       [6.8, 3.2, 5.9, 2.3, 'virginica'],
       [6.7, 3.3, 5.7, 2.5, 'virginica'],
       [6.7, 3.0, 5.2, 2.3, 'virginica'],
       [6.3, 2.5, 5.0, 1.9, 'virginica'],
       [6.5, 3.0, 5.2, 2.0, 'virginica'],
       [6.2, 3.4, 5.4, 2.3, 'virginica'],
       [5.9, 3.0, 5.1, 1.8, 'virginica']], dtype=object)
```

```
df.describe()
```

```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.057333      3.758000     1.199333
std        0.828066     0.435866      1.765298     0.762238
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000
```

```
df.T
```

```
                 0       1       2       3       4       5       6
7    \
sepal_length     5.1     4.9     4.7     4.6     5.0     5.4     4.6
5.0
sepal_width      3.5     3.0     3.2     3.1     3.6     3.9     3.4
3.4
petal_length     1.4     1.4     1.3     1.5     1.4     1.7     1.4
1.5
petal_width      0.2     0.2     0.2     0.2     0.2     0.4     0.3
0.2
species       setosa  setosa  setosa  setosa  setosa  setosa  setosa
setosa

                 8       9    ...        140        141        142
143  \
sepal_length     4.4     4.9  ...        6.7        6.9        5.8
6.8
sepal_width      2.9     3.1  ...        3.1        3.1        2.7
3.2
petal_length     1.4     1.5  ...        5.6        5.1        5.1
5.9
petal_width      0.2     0.1  ...        2.4        2.3        1.9
2.3
species       setosa  setosa  ...  virginica  virginica  virginica
```

```
virginica

                    144        145        146        147        148
149
sepal_length        6.7        6.7        6.3        6.5        6.2
5.9
sepal_width         3.3        3.0        2.5        3.0        3.4
3.0
petal_length        5.7        5.2        5.0        5.2        5.4
5.1
petal_width         2.5        2.3        1.9        2.0        2.3
1.8
species       virginica  virginica  virginica  virginica  virginica
virginica

[5 rows x 150 columns]
```

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

```
       species  sepal_width  sepal_length  petal_width  petal_length
0       setosa          3.5           5.1          0.2           1.4
1       setosa          3.0           4.9          0.2           1.4
2       setosa          3.2           4.7          0.2           1.3
3       setosa          3.1           4.6          0.2           1.5
4       setosa          3.6           5.0          0.2           1.4
..         ...          ...           ...          ...           ...
145  virginica          3.0           6.7          2.3           5.2
146  virginica          2.5           6.3          1.9           5.0
147  virginica          3.0           6.5          2.0           5.2
148  virginica          3.4           6.2          2.3           5.4
149  virginica          3.0           5.9          1.8           5.1

[150 rows x 5 columns]
```

```
df.sort_values(by="species")
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             5.1          3.5           1.4          0.2     setosa
27            5.2          3.5           1.5          0.2     setosa
28            5.2          3.4           1.4          0.2     setosa
29            4.7          3.2           1.6          0.2     setosa
30            4.8          3.1           1.6          0.2     setosa
..            ...          ...           ...          ...        ...
119           6.0          2.2           5.0          1.5  virginica
120           6.9          3.2           5.7          2.3  virginica
121           5.6          2.8           4.9          2.0  virginica
111           6.4          2.7           5.3          1.9  virginica
149           5.9          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

Getitem ([])

```
df["petal_length"]
```

```
0      1.4
1      1.4
2      1.3
3      1.5
4      1.4
      ...
145    5.2
146    5.0
147    5.2
148    5.4
149    5.1
Name: petal_length, Length: 150, dtype: float64
```

```
df[0:4]
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa

```
df[4:8]
```

	sepal_length	sepal_width	petal_length	petal_width	species
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa

Selection by label

```
df.loc[[0]]
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa

```
df
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
..	...	...	...	...	...

```
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
147           6.5          3.0           5.2          2.0  virginica
148           6.2          3.4           5.4          2.3  virginica
149           5.9          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

```
df.loc[:, ["sepal_length", "species"]]
```

```
     sepal_length    species
0             5.1     setosa
1             4.9     setosa
2             4.7     setosa
3             4.6     setosa
4             5.0     setosa
..            ...        ...
145           6.7  virginica
146           6.3  virginica
147           6.5  virginica
148           6.2  virginica
149           5.9  virginica

[150 rows x 2 columns]
```

```
df.loc["1":"3", ["sepal_length", "species"]]
```

```
    sepal_length species
1            4.9  setosa
2            4.7  setosa
3            4.6  setosa
4            5.0  setosa
5            5.4  setosa
6            4.6  setosa
7            5.0  setosa
8            4.4  setosa
9            4.9  setosa
10           5.4  setosa
11           4.8  setosa
12           4.8  setosa
13           4.3  setosa
14           5.8  setosa
15           5.7  setosa
16           5.4  setosa
17           5.1  setosa
18           5.7  setosa
19           5.1  setosa
20           5.4  setosa
21           5.1  setosa
22           4.6  setosa
```

```
23            5.1  setosa
24            4.8  setosa
25            5.0  setosa
26            5.0  setosa
27            5.2  setosa
28            5.2  setosa
29            4.7  setosa
```

```
df.loc[[5], "species"]
```

```
5    setosa
Name: species, dtype: object
```

```
# df.at["sepal_length"[0], "sepal_length"]
```

```
df.at[df.index[4], "sepal_length"]
```

```


5.0
```

Selection by position

```
df
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             5.1          3.5           1.4          0.2     setosa
1             4.9          3.0           1.4          0.2     setosa
2             4.7          3.2           1.3          0.2     setosa
3             4.6          3.1           1.5          0.2     setosa
4             5.0          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
147           6.5          3.0           5.2          2.0  virginica
148           6.2          3.4           5.4          2.3  virginica
149           5.9          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

```
df.iloc[3]
```

```
sepal_length       4.6
sepal_width        3.1
petal_length       1.5
petal_width        0.2
species         setosa
Name: 3, dtype: object
```

df: This is the DataFrame you're working with.

.iloc: This is a pandas DataFrame attribute that is used for integer-location based indexing.

[4:5, 0:3]: This is the selection part. It consists of two slices separated by a comma. The first slice (4:5) refers to rows, and the second slice (0:3) refers to columns.

4:5: This indicates that you want to select rows starting from index 4 up to (but not including) index 5. In Python, indexing starts from 0, so this is selecting the fifth row of the DataFrame.

0:3: This indicates that you want to select columns starting from index 0 up to (but not including) index 3. It selects the columns at positions 0, 1, and 2.

Putting it all together, the code is selecting a specific subset of your DataFrame, specifically the fifth row and the columns at positions 0, 1, and 2.

```
df.iloc[4:5, 0:3]
```

```
   sepal_length  sepal_width  petal_length
4           5.0          3.6           1.4
```

```
df.iloc[1:5, 0:3]
```

```
   sepal_length  sepal_width  petal_length
1           4.9          3.0           1.4
2           4.7          3.2           1.3
3           4.6          3.1           1.5
4           5.0          3.6           1.4
```

```
df.iloc[[1, 2, 4], [0, 2]]
```

```
   sepal_length  petal_length
1           4.9           1.4
2           4.7           1.3
4           5.0           1.4
```

[1:3, :]: This is the selection part. It consists of two slices separated by a comma. The first slice (1:3) refers to rows, and the second slice (:) refers to all columns.

1:3: This indicates that you want to select rows starting from index 1 up to (but not including) index 3. In Python, indexing starts from 0, so this is selecting the second and third rows of the DataFrame.

: This indicates that you want to select all columns.

```
df.iloc[1:3, :]
```

```
   sepal_length  sepal_width  petal_length  petal_width species
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
```

```
df.iloc[:, 1:3]
```

```
     sepal_width  petal_length
0            3.5           1.4
1            3.0           1.4
2            3.2           1.3
3            3.1           1.5
4            3.6           1.4
..           ...           ...
145          3.0           5.2
146          2.5           5.0
147          3.0           5.2
148          3.4           5.4
149          3.0           5.1

[150 rows x 2 columns]
```

df: This is the DataFrame you're working with.

.iloc: This is a pandas DataFrame attribute that is used for integer-location based indexing.

[1, 1]: This is the selection part. It consists of two indices separated by a comma. The first index (1) refers to the row, and the second index (1) refers to the column.

Putting it all together, the code is selecting the element at the second row and second column of your DataFrame.

```
df.iloc[1, 1]
```

```
3.0
```

```
df.iat[1, 1]
```

```
3.0
```

Boolean indexing

```
df[df["sepal_width"] > 3.5]
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
4             5.0          3.6           1.4          0.2     setosa
5             5.4          3.9           1.7          0.4     setosa
10            5.4          3.7           1.5          0.2     setosa
14            5.8          4.0           1.2          0.2     setosa
15            5.7          4.4           1.5          0.4     setosa
16            5.4          3.9           1.3          0.4     setosa
18            5.7          3.8           1.7          0.3     setosa
19            5.1          3.8           1.5          0.3     setosa
21            5.1          3.7           1.5          0.4     setosa
22            4.6          3.6           1.0          0.2     setosa
32            5.2          4.1           1.5          0.1     setosa
```

```
33            5.5           4.2            1.4           0.2     setosa
37            4.9           3.6            1.4           0.1     setosa
44            5.1           3.8            1.9           0.4     setosa
46            5.1           3.8            1.6           0.2     setosa
48            5.3           3.7            1.5           0.2     setosa
109           7.2           3.6            6.1           2.5  virginica
117           7.7           3.8            6.7           2.2  virginica
131           7.9           3.8            6.4           2.0  virginica
```

```
print(df.dtypes)
```

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

```
df['sepal_length'] = pd.to_numeric(df['sepal_width'], errors='coerce')
df['sepal_length']
```

```
0      3.5
1      3.0
2      3.2
3      3.1
4      3.6
      ...
145    3.0
146    2.5
147    3.0
148    3.4
149    3.0
Name: sepal_length, Length: 150, dtype: float64
```

```
# df[df > 0]
selected_columns = ['sepal_length', 'sepal_width']
filtered_df = df["sepal_length" ][df["sepal_width"] > 4]
filtered_df
```

```
15    4.4
32    4.1
33    4.2
Name: sepal_length, dtype: float64
```

The code df2 = df.copy() creates a copy of the DataFrame df and assigns it to the variable df2. This is a common practice when you want to work with a copy of a DataFrame, leaving the original DataFrame unchanged.

Here's what happens in this line:

df: This is the original DataFrame.

.copy(): This is a method in Pandas that creates a deep copy of the DataFrame. A deep copy means that a new copy of the data and the index is created, and changes made to the copy do not affect the original DataFrame, and vice versa.

df2 = ...: The result of the .copy() operation is assigned to the variable df2, so now df2 is an independent copy of df.

?

```
df2 = df.copy()
df2
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	3.5	3.5	1.4	0.2	setosa
1	3.0	3.0	1.4	0.2	setosa
2	3.2	3.2	1.3	0.2	setosa
3	3.1	3.1	1.5	0.2	setosa
4	3.6	3.6	1.4	0.2	setosa
..	...	...	...	...	...
145	3.0	3.0	5.2	2.3	virginica
146	2.5	2.5	5.0	1.9	virginica
147	3.0	3.0	5.2	2.0	virginica
148	3.4	3.4	5.4	2.3	virginica
149	3.0	3.0	5.1	1.8	virginica

[150 rows x 5 columns]

The code df2[df2["petal_width"].isin(["two", "four"])] is using boolean indexing to filter rows in the DataFrame df2 based on whether the values in the "petal_width" column are either "two" or "four".

Let's break it down:

df2["petal_width"]: Selects the "petal_width" column from the DataFrame df2.

.isin(["two", "four"]): Checks whether each value in the "petal_width" column is either "two" or "four". This creates a boolean Series where each element is True if the condition is met and False otherwise.

df2[...]: Uses boolean indexing to filter rows from the DataFrame df2. Only the rows where the condition is True will be included in the result.

This boolean Series is then used for boolean indexing, filtering only the rows in the DataFrame where the condition is True. In the context of your original code (df2[df2["petal_width"].isin(["two", "four"])]), only the rows with "petal_width" values of "two" or "four" will be included in the result.

```
df2[df2["petal_width"].isin(["two", "four"])]
# filtered_df = df2[df2["petal_width"].isin(["two", "four"])]
```

```
# filtered_df
df2
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             3.5          3.5           1.4          0.2     setosa
1             3.0          3.0           1.4          0.2     setosa
2             3.2          3.2           1.3          0.2     setosa
3             3.1          3.1           1.5          0.2     setosa
4             3.6          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           3.0          3.0           5.2          2.3  virginica
146           2.5          2.5           5.0          1.9  virginica
147           3.0          3.0           5.2          2.0  virginica
148           3.4          3.4           5.4          2.3  virginica
149           3.0          3.0           5.1          1.8  virginica

[150 rows x 5 columns]
```

```
df.info
```

```
<bound method DataFrame.info of      sepal_length  sepal_width
petal_length  petal_width    species
0             3.5          3.5           1.4          0.2     setosa
1             3.0          3.0           1.4          0.2     setosa
2             3.2          3.2           1.3          0.2     setosa
3             3.1          3.1           1.5          0.2     setosa
4             3.6          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           3.0          3.0           5.2          2.3  virginica
146           2.5          2.5           5.0          1.9  virginica
147           3.0          3.0           5.2          2.0  virginica
148           3.4          3.4           5.4          2.3  virginica
149           3.0          3.0           5.1          1.8  virginica

[150 rows x 5 columns]>
```

Setting

```
s1 = pd.Series([1, 2, 3, 4, 5, 6], index=pd.date_range("20130102",
periods=6))
```

```
s1
```

```
2013-01-02    1
2013-01-03    2
2013-01-04    3
2013-01-05    4
2013-01-06    5
```

```
2013-01-07    6
Freq: D, dtype: int64
```

```
df["sepal_length"] = s1
s1
```

```
2013-01-02    1
2013-01-03    2
2013-01-04    3
2013-01-05    4
2013-01-06    5
2013-01-07    6
Freq: D, dtype: int64
```

```
df.at["sepal_width"[0], "sepal_length"] = 0
df
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	NaN	3.5	1.4	0.2	setosa
1	NaN	3.0	1.4	0.2	setosa
2	NaN	3.2	1.3	0.2	setosa
3	NaN	3.1	1.5	0.2	setosa
4	NaN	3.6	1.4	0.2	setosa
..	...	...	...	...	...
146	NaN	2.5	5.0	1.9	virginica
147	NaN	3.0	5.2	2.0	virginica
148	NaN	3.4	5.4	2.3	virginica
149	NaN	3.0	5.1	1.8	virginica
s	0.0	NaN	NaN	NaN	NaN

```
[151 rows x 5 columns]
```

```
df.iat[0, 1] = 0
df
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	NaN	0.0	1.4	0.2	setosa
1	NaN	3.0	1.4	0.2	setosa
2	NaN	3.2	1.3	0.2	setosa
3	NaN	3.1	1.5	0.2	setosa
4	NaN	3.6	1.4	0.2	setosa
..	...	...	...	...	...
146	NaN	2.5	5.0	1.9	virginica
147	NaN	3.0	5.2	2.0	virginica
148	NaN	3.4	5.4	2.3	virginica
149	NaN	3.0	5.1	1.8	virginica
s	0.0	NaN	NaN	NaN	NaN

```
[151 rows x 5 columns]
```

```
df.loc[:, "petal_width"] = np.array([5] * len(df))
df
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
0             NaN          0.0           1.4          5.0     setosa
1             NaN          3.0           1.4          5.0     setosa
2             NaN          3.2           1.3          5.0     setosa
3             NaN          3.1           1.5          5.0     setosa
4             NaN          3.6           1.4          5.0     setosa
..            ...          ...           ...          ...        ...
146           NaN          2.5           5.0          5.0  virginica
147           NaN          3.0           5.2          5.0  virginica
148           NaN          3.4           5.4          5.0  virginica
149           NaN          3.0           5.1          5.0  virginica
s             0.0          NaN           NaN          5.0        NaN

[151 rows x 5 columns]
```

```
df2 = df.copy()
df2[df2 > 0] = -df2
df2

# # Assuming "column_name" is a column with numerical values
# df2["sepal_width"] = pd.to_numeric(df2["sepal_width"],
errors="coerce")
# df2[df2 > 0] = -df2

# Assuming "sepal_width" is the column with mixed data types
# df2["sepal_width"] = pd.to_numeric(df2["sepal_width"],
errors="coerce")

# # Now perform the negation operation
# df2[df2 > 0] = -df2


# Assuming "sepal_width" is the column with mixed data types
df2["sepal_width"] = pd.to_numeric(df2["sepal_width"],
errors="coerce")

# Filter rows where "sepal_width" is greater than 0 and perform
negation
mask = df2["sepal_width"] > 0
df2.loc[mask, "sepal_width"] = -df2.loc[mask, "sepal_width"]
df2
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
data1  \
0             NaN          0.0           1.4          5.0     setosa
NaN
```

```
1              NaN          -3.0           1.4           5.0     setosa
NaN
2              NaN          -3.2           1.3           5.0     setosa
NaN
3              NaN          -3.1           1.5           5.0     setosa
NaN
4              NaN          -3.6           1.4           5.0     setosa
NaN
..             ...           ...           ...           ...        ...
...
146            NaN          -2.5           5.0           5.0  virginica
NaN
147            NaN          -3.0           5.2           5.0  virginica
NaN
148            NaN          -3.4           5.4           5.0  virginica
NaN
149            NaN          -3.0           5.1           5.0  virginica
NaN
s              0.0           NaN           NaN           5.0        NaN
0.0

     species1  column_name
0         0.0          0.0
1         3.0          3.0
2         3.2          3.2
3         3.1          3.1
4         3.6          3.6
..        ...          ...
146       2.5          2.5
147       3.0          3.0
148       3.4          3.4
149       3.0          3.0
s         NaN          NaN

[151 rows x 8 columns]
```

Missing data

```
df1 = df.reindex(index=dates[0:4], columns=list(df.columns) + ["E"])

df1.loc[dates[0] : dates[1], "E"] = 1

df1
```

```
            sepal_length  sepal_width  petal_length  petal_width
species    E
2013-01-01           NaN          NaN           NaN          NaN
NaN  1.0
```

```
2013-01-02            NaN           NaN           NaN           NaN
NaN  1.0
2013-01-03            NaN           NaN           NaN           NaN
NaN  NaN
2013-01-04            NaN           NaN           NaN           NaN
NaN  NaN
```