

KEY_Practice17_Pandas-Subsetting-II

January 2, 2020

1 Practice: Subsetting Pandas DataFrames II

For this practice, let's use the iris dataset:

```
[22]: # import the pandas package
import pandas as pd
# set the path
path = 'https://raw.githubusercontent.com/GWC-DCMB/ClubCurriculum/master/'
# this is where the file is located
filename = path + 'SampleData/iris.csv'
# load the iris dataset into a DataFrame
iris = pd.read_csv(filename)
```

Take a look at the dataset:

```
[23]: # take a look at the beginning

iris.head()
```

```
[23]:
```

| | 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 |

```
[24]: # subset the first 5 rows from iris
# save it to a variable called subset1
subset1 = iris.iloc[:5]
```

```
[25]: # subset a few columns from the subset1 dataframe
# save it to a variable called subset 2
columns = ['sepal_length', 'sepal_width']
subset2 = subset1[columns]
```

Let's try subsetting both rows and columns at the same time!

```
[26]: # create a new subset from iris that's identical to subset2
# but write only one line of code
# save it to a variable called subset3
subset3 = iris.iloc[:5][['sepal_length', 'sepal_width']]
```

```
[27]: # check your work -- how does subset2 compare to subset3?
subset2 == subset3
```

```
[27]:      sepal_length  sepal_width
0             True             True
1             True             True
2             True             True
3             True             True
4             True             True
```

```
[28]: # subset rows 20 through 30 and columns petal_length & petal width
# write only one line of code
iris.iloc[20:31][['petal_length', 'petal_width']]
```

```
[28]:      petal_length  petal_width
20             1.7             0.2
21             1.5             0.4
22             1.0             0.2
23             1.7             0.5
24             1.9             0.2
25             1.6             0.2
26             1.6             0.4
27             1.5             0.2
28             1.4             0.2
29             1.6             0.2
30             1.6             0.2
```

Now let's subset using query:

```
[29]: # subset rows where the species is not setosa

iris.query('species != "setosa"')
```

```
[29]:      sepal_length  sepal_width  petal_length  petal_width  species
50             7.0           3.2           4.7           1.4  versicolor
51             6.4           3.2           4.5           1.5  versicolor
52             6.9           3.1           4.9           1.5  versicolor
53             5.5           2.3           4.0           1.3  versicolor
54             6.5           2.8           4.6           1.5  versicolor
..            ...            ...            ...            ...  ...
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 |

[100 rows x 5 columns]

```
[30]: # subset rows where sepal_width is greater than 4
```

```
iris.query('sepal_width > 4')
```

```
[30]:      sepal_length  sepal_width  petal_length  petal_width  species
15           5.7         4.4         1.5         0.4    setosa
32           5.2         4.1         1.5         0.1    setosa
33           5.5         4.2         1.4         0.2    setosa
```

```
[31]: # subset rows where sepal_width is between 2 and 3
```

```
iris.query('2 < sepal_width < 3')
```

```
[31]:      sepal_length  sepal_width  petal_length  petal_width  species
8           4.4         2.9         1.4         0.2    setosa
41          4.5         2.3         1.3         0.3    setosa
53          5.5         2.3         4.0         1.3  versicolor
54          6.5         2.8         4.6         1.5  versicolor
55          5.7         2.8         4.5         1.3  versicolor
57          4.9         2.4         3.3         1.0  versicolor
58          6.6         2.9         4.6         1.3  versicolor
59          5.2         2.7         3.9         1.4  versicolor
62          6.0         2.2         4.0         1.0  versicolor
63          6.1         2.9         4.7         1.4  versicolor
64          5.6         2.9         3.6         1.3  versicolor
67          5.8         2.7         4.1         1.0  versicolor
68          6.2         2.2         4.5         1.5  versicolor
69          5.6         2.5         3.9         1.1  versicolor
71          6.1         2.8         4.0         1.3  versicolor
72          6.3         2.5         4.9         1.5  versicolor
73          6.1         2.8         4.7         1.2  versicolor
74          6.4         2.9         4.3         1.3  versicolor
76          6.8         2.8         4.8         1.4  versicolor
78          6.0         2.9         4.5         1.5  versicolor
79          5.7         2.6         3.5         1.0  versicolor
80          5.5         2.4         3.8         1.1  versicolor
81          5.5         2.4         3.7         1.0  versicolor
82          5.8         2.7         3.9         1.2  versicolor
83          6.0         2.7         5.1         1.6  versicolor
87          6.3         2.3         4.4         1.3  versicolor
89          5.5         2.5         4.0         1.3  versicolor
```

| | | | | | |
|-----|-----|-----|-----|-----|------------|
| 90 | 5.5 | 2.6 | 4.4 | 1.2 | versicolor |
| 92 | 5.8 | 2.6 | 4.0 | 1.2 | versicolor |
| 93 | 5.0 | 2.3 | 3.3 | 1.0 | versicolor |
| 94 | 5.6 | 2.7 | 4.2 | 1.3 | versicolor |
| 96 | 5.7 | 2.9 | 4.2 | 1.3 | versicolor |
| 97 | 6.2 | 2.9 | 4.3 | 1.3 | versicolor |
| 98 | 5.1 | 2.5 | 3.0 | 1.1 | versicolor |
| 99 | 5.7 | 2.8 | 4.1 | 1.3 | versicolor |
| 101 | 5.8 | 2.7 | 5.1 | 1.9 | virginica |
| 103 | 6.3 | 2.9 | 5.6 | 1.8 | virginica |
| 106 | 4.9 | 2.5 | 4.5 | 1.7 | virginica |
| 107 | 7.3 | 2.9 | 6.3 | 1.8 | virginica |
| 108 | 6.7 | 2.5 | 5.8 | 1.8 | virginica |
| 111 | 6.4 | 2.7 | 5.3 | 1.9 | virginica |
| 113 | 5.7 | 2.5 | 5.0 | 2.0 | virginica |
| 114 | 5.8 | 2.8 | 5.1 | 2.4 | virginica |
| 118 | 7.7 | 2.6 | 6.9 | 2.3 | virginica |
| 119 | 6.0 | 2.2 | 5.0 | 1.5 | virginica |
| 121 | 5.6 | 2.8 | 4.9 | 2.0 | virginica |
| 122 | 7.7 | 2.8 | 6.7 | 2.0 | virginica |
| 123 | 6.3 | 2.7 | 4.9 | 1.8 | virginica |
| 126 | 6.2 | 2.8 | 4.8 | 1.8 | virginica |
| 128 | 6.4 | 2.8 | 5.6 | 2.1 | virginica |
| 130 | 7.4 | 2.8 | 6.1 | 1.9 | virginica |
| 132 | 6.4 | 2.8 | 5.6 | 2.2 | virginica |
| 133 | 6.3 | 2.8 | 5.1 | 1.5 | virginica |
| 134 | 6.1 | 2.6 | 5.6 | 1.4 | virginica |
| 142 | 5.8 | 2.7 | 5.1 | 1.9 | virginica |
| 146 | 6.3 | 2.5 | 5.0 | 1.9 | virginica |

```
[32]: # subset rows where sepal_width is less than 3.5 and the species is virginica
iris.query('sepal_width < 3.5 and species == "virginica"')
```

```
[32]:   sepal_length  sepal_width  petal_length  petal_width  species
100           6.3         3.3         6.0         2.5  virginica
101           5.8         2.7         5.1         1.9  virginica
102           7.1         3.0         5.9         2.1  virginica
103           6.3         2.9         5.6         1.8  virginica
104           6.5         3.0         5.8         2.2  virginica
105           7.6         3.0         6.6         2.1  virginica
106           4.9         2.5         4.5         1.7  virginica
107           7.3         2.9         6.3         1.8  virginica
108           6.7         2.5         5.8         1.8  virginica
110           6.5         3.2         5.1         2.0  virginica
111           6.4         2.7         5.3         1.9  virginica
112           6.8         3.0         5.5         2.1  virginica
```

| | | | | | |
|-----|-----|-----|-----|-----|-----------|
| 113 | 5.7 | 2.5 | 5.0 | 2.0 | virginica |
| 114 | 5.8 | 2.8 | 5.1 | 2.4 | virginica |
| 115 | 6.4 | 3.2 | 5.3 | 2.3 | virginica |
| 116 | 6.5 | 3.0 | 5.5 | 1.8 | virginica |
| 118 | 7.7 | 2.6 | 6.9 | 2.3 | virginica |
| 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 |
| 122 | 7.7 | 2.8 | 6.7 | 2.0 | virginica |
| 123 | 6.3 | 2.7 | 4.9 | 1.8 | virginica |
| 124 | 6.7 | 3.3 | 5.7 | 2.1 | virginica |
| 125 | 7.2 | 3.2 | 6.0 | 1.8 | virginica |
| 126 | 6.2 | 2.8 | 4.8 | 1.8 | virginica |
| 127 | 6.1 | 3.0 | 4.9 | 1.8 | virginica |
| 128 | 6.4 | 2.8 | 5.6 | 2.1 | virginica |
| 129 | 7.2 | 3.0 | 5.8 | 1.6 | virginica |
| 130 | 7.4 | 2.8 | 6.1 | 1.9 | virginica |
| 132 | 6.4 | 2.8 | 5.6 | 2.2 | virginica |
| 133 | 6.3 | 2.8 | 5.1 | 1.5 | virginica |
| 134 | 6.1 | 2.6 | 5.6 | 1.4 | virginica |
| 135 | 7.7 | 3.0 | 6.1 | 2.3 | virginica |
| 136 | 6.3 | 3.4 | 5.6 | 2.4 | virginica |
| 137 | 6.4 | 3.1 | 5.5 | 1.8 | virginica |
| 138 | 6.0 | 3.0 | 4.8 | 1.8 | virginica |
| 139 | 6.9 | 3.1 | 5.4 | 2.1 | virginica |
| 140 | 6.7 | 3.1 | 5.6 | 2.4 | virginica |
| 141 | 6.9 | 3.1 | 5.1 | 2.3 | virginica |
| 142 | 5.8 | 2.7 | 5.1 | 1.9 | virginica |
| 143 | 6.8 | 3.2 | 5.9 | 2.3 | virginica |
| 144 | 6.7 | 3.3 | 5.7 | 2.5 | virginica |
| 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 |

```
[33]: # subset rows where the petal width is 0.3 or the species is versicolor
```

```
iris.query('petal_width == 0.3 or species == "versicolor"')
```

| [33]: | sepal_length | sepal_width | petal_length | petal_width | species |
|-------|--------------|-------------|--------------|-------------|---------|
| 6 | 4.6 | 3.4 | 1.4 | 0.3 | setosa |
| 17 | 5.1 | 3.5 | 1.4 | 0.3 | setosa |
| 18 | 5.7 | 3.8 | 1.7 | 0.3 | setosa |
| 19 | 5.1 | 3.8 | 1.5 | 0.3 | setosa |
| 40 | 5.0 | 3.5 | 1.3 | 0.3 | setosa |
| 41 | 4.5 | 2.3 | 1.3 | 0.3 | setosa |

| | | | | | |
|----|-----|-----|-----|-----|------------|
| 45 | 4.8 | 3.0 | 1.4 | 0.3 | setosa |
| 50 | 7.0 | 3.2 | 4.7 | 1.4 | versicolor |
| 51 | 6.4 | 3.2 | 4.5 | 1.5 | versicolor |
| 52 | 6.9 | 3.1 | 4.9 | 1.5 | versicolor |
| 53 | 5.5 | 2.3 | 4.0 | 1.3 | versicolor |
| 54 | 6.5 | 2.8 | 4.6 | 1.5 | versicolor |
| 55 | 5.7 | 2.8 | 4.5 | 1.3 | versicolor |
| 56 | 6.3 | 3.3 | 4.7 | 1.6 | versicolor |
| 57 | 4.9 | 2.4 | 3.3 | 1.0 | versicolor |
| 58 | 6.6 | 2.9 | 4.6 | 1.3 | versicolor |
| 59 | 5.2 | 2.7 | 3.9 | 1.4 | versicolor |
| 60 | 5.0 | 2.0 | 3.5 | 1.0 | versicolor |
| 61 | 5.9 | 3.0 | 4.2 | 1.5 | versicolor |
| 62 | 6.0 | 2.2 | 4.0 | 1.0 | versicolor |
| 63 | 6.1 | 2.9 | 4.7 | 1.4 | versicolor |
| 64 | 5.6 | 2.9 | 3.6 | 1.3 | versicolor |
| 65 | 6.7 | 3.1 | 4.4 | 1.4 | versicolor |
| 66 | 5.6 | 3.0 | 4.5 | 1.5 | versicolor |
| 67 | 5.8 | 2.7 | 4.1 | 1.0 | versicolor |
| 68 | 6.2 | 2.2 | 4.5 | 1.5 | versicolor |
| 69 | 5.6 | 2.5 | 3.9 | 1.1 | versicolor |
| 70 | 5.9 | 3.2 | 4.8 | 1.8 | versicolor |
| 71 | 6.1 | 2.8 | 4.0 | 1.3 | versicolor |
| 72 | 6.3 | 2.5 | 4.9 | 1.5 | versicolor |
| 73 | 6.1 | 2.8 | 4.7 | 1.2 | versicolor |
| 74 | 6.4 | 2.9 | 4.3 | 1.3 | versicolor |
| 75 | 6.6 | 3.0 | 4.4 | 1.4 | versicolor |
| 76 | 6.8 | 2.8 | 4.8 | 1.4 | versicolor |
| 77 | 6.7 | 3.0 | 5.0 | 1.7 | versicolor |
| 78 | 6.0 | 2.9 | 4.5 | 1.5 | versicolor |
| 79 | 5.7 | 2.6 | 3.5 | 1.0 | versicolor |
| 80 | 5.5 | 2.4 | 3.8 | 1.1 | versicolor |
| 81 | 5.5 | 2.4 | 3.7 | 1.0 | versicolor |
| 82 | 5.8 | 2.7 | 3.9 | 1.2 | versicolor |
| 83 | 6.0 | 2.7 | 5.1 | 1.6 | versicolor |
| 84 | 5.4 | 3.0 | 4.5 | 1.5 | versicolor |
| 85 | 6.0 | 3.4 | 4.5 | 1.6 | versicolor |
| 86 | 6.7 | 3.1 | 4.7 | 1.5 | versicolor |
| 87 | 6.3 | 2.3 | 4.4 | 1.3 | versicolor |
| 88 | 5.6 | 3.0 | 4.1 | 1.3 | versicolor |
| 89 | 5.5 | 2.5 | 4.0 | 1.3 | versicolor |
| 90 | 5.5 | 2.6 | 4.4 | 1.2 | versicolor |
| 91 | 6.1 | 3.0 | 4.6 | 1.4 | versicolor |
| 92 | 5.8 | 2.6 | 4.0 | 1.2 | versicolor |
| 93 | 5.0 | 2.3 | 3.3 | 1.0 | versicolor |
| 94 | 5.6 | 2.7 | 4.2 | 1.3 | versicolor |
| 95 | 5.7 | 3.0 | 4.2 | 1.2 | versicolor |

| | | | | | |
|----|-----|-----|-----|-----|------------|
| 96 | 5.7 | 2.9 | 4.2 | 1.3 | versicolor |
| 97 | 6.2 | 2.9 | 4.3 | 1.3 | versicolor |
| 98 | 5.1 | 2.5 | 3.0 | 1.1 | versicolor |
| 99 | 5.7 | 2.8 | 4.1 | 1.3 | versicolor |

Bonus: Try to subset with both `query` and square brackets `[]` on the same line:

```
[34]: # pick any query and any columns to subset with

iris.query('species != "setosa"')[['sepal_length', 'species']] # answers may
    ↪ vary
```

```
[34]:      sepal_length      species
50          7.0  versicolor
51          6.4  versicolor
52          6.9  versicolor
53          5.5  versicolor
54          6.5  versicolor
..          ...          ...
145         6.7   virginica
146         6.3   virginica
147         6.5   virginica
148         6.2   virginica
149         5.9   virginica
```

[100 rows x 2 columns]