KEY_Practice12_Pandas-Subsetting

July 17, 2019

1 Practice: Subsetting Pandas DataFrames

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

```python
[4]: # mount Google Drive
from google.colab import drive
drive.mount('/content/gdrive')
path = '/content/gdrive/My Drive/SummerExperience-master/'
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).

```python
[0]: # import pandas package
import pandas as pd
```

```python
[0]: # this is where the file is located
filename = path + 'SampleData/iris.csv'
# load the iris dataset into a DataFrame
iris = pd.read_csv(filename)
```

Refamiliarize yourself with the dataset:

```python
[7]: # take a look at the beginning

iris.head()
```

```
[7]:    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
```

Try subsetting on columns:

```python
[8]: # subset the species column

iris['species']
```

```
[8]: 0          setosa
    1          setosa
```

```
2         setosa
3         setosa
4         setosa
5         setosa
6         setosa
7         setosa
8         setosa
9         setosa
10        setosa
11        setosa
12        setosa
13        setosa
14        setosa
15        setosa
16        setosa
17        setosa
18        setosa
19        setosa
20        setosa
21        setosa
22        setosa
23        setosa
24        setosa
25        setosa
26        setosa
27        setosa
28        setosa
29        setosa
          ...
120     virginica
121     virginica
122     virginica
123     virginica
124     virginica
125     virginica
126     virginica
127     virginica
128     virginica
129     virginica
130     virginica
131     virginica
132     virginica
133     virginica
134     virginica
135     virginica
136     virginica
137     virginica
```

```
138    virginica
139    virginica
140    virginica
141    virginica
142    virginica
143    virginica
144    virginica
145    virginica
146    virginica
147    virginica
148    virginica
149    virginica
Name: species, Length: 150, dtype: object
```

[9]:
```python
# subset the sepal_length and sepal_width columns

iris[ ['sepal_length','sepal_width']]
```

[9]:
```
    sepal_length  sepal_width
0            5.1          3.5
1            4.9          3.0
2            4.7          3.2
3            4.6          3.1
4            5.0          3.6
5            5.4          3.9
6            4.6          3.4
7            5.0          3.4
8            4.4          2.9
9            4.9          3.1
10           5.4          3.7
11           4.8          3.4
12           4.8          3.0
13           4.3          3.0
14           5.8          4.0
15           5.7          4.4
16           5.4          3.9
17           5.1          3.5
18           5.7          3.8
19           5.1          3.8
20           5.4          3.4
21           5.1          3.7
22           4.6          3.6
23           5.1          3.3
24           4.8          3.4
25           5.0          3.0
26           5.0          3.4
27           5.2          3.5
28           5.2          3.4
```

```
29               4.7          3.2
..               ...          ...
120              6.9          3.2
121              5.6          2.8
122              7.7          2.8
123              6.3          2.7
124              6.7          3.3
125              7.2          3.2
126              6.2          2.8
127              6.1          3.0
128              6.4          2.8
129              7.2          3.0
130              7.4          2.8
131              7.9          3.8
132              6.4          2.8
133              6.3          2.8
134              6.1          2.6
135              7.7          3.0
136              6.3          3.4
137              6.4          3.1
138              6.0          3.0
139              6.9          3.1
140              6.7          3.1
141              6.9          3.1
142              5.8          2.7
143              6.8          3.2
144              6.7          3.3
145              6.7          3.0
146              6.3          2.5
147              6.5          3.0
148              6.2          3.4
149              5.9          3.0

[150 rows x 2 columns]
```

Try subsetting on rows:

```
[10]: # subset the 2nd column

iris[iris.columns[1]]
```

```
[10]: 0        3.5
      1        3.0
      2        3.2
      3        3.1
      4        3.6
      5        3.9
      6        3.4
      7        3.4
```

```
8       2.9
9       3.1
10      3.7
11      3.4
12      3.0
13      3.0
14      4.0
15      4.4
16      3.9
17      3.5
18      3.8
19      3.8
20      3.4
21      3.7
22      3.6
23      3.3
24      3.4
25      3.0
26      3.4
27      3.5
28      3.4
29      3.2
        ...
120     3.2
121     2.8
122     2.8
123     2.7
124     3.3
125     3.2
126     2.8
127     3.0
128     2.8
129     3.0
130     2.8
131     3.8
132     2.8
133     2.8
134     2.6
135     3.0
136     3.4
137     3.1
138     3.0
139     3.1
140     3.1
141     3.1
142     2.7
143     3.2
```

```
144    3.3
145    3.0
146    2.5
147    3.0
148    3.4
149    3.0
Name: sepal_width, Length: 150, dtype: float64
```

[11]:
```python
# subset the first 5 rows

iris.loc[:4]
```

[11]:

	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

[12]:
```python
# subset rows 10 through 20

iris.loc[10:20]
```

[12]:

	sepal_length	sepal_width	petal_length	petal_width	species
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa
12	4.8	3.0	1.4	0.1	setosa
13	4.3	3.0	1.1	0.1	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
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
20	5.4	3.4	1.7	0.2	setosa

[13]:
```python
# subset rows 6, 9, and 12

iris.loc[[6,9,12]]
```

[13]:

	sepal_length	sepal_width	petal_length	petal_width	species
6	4.6	3.4	1.4	0.3	setosa
9	4.9	3.1	1.5	0.1	setosa
12	4.8	3.0	1.4	0.1	setosa

Now do both!

[14]:
```python
# subset the first 3 rows and the first 3 columns

iris.loc[:2][iris.columns[:3]]
```

```
[14]:    sepal_length   sepal_width   petal_length
     0          5.1           3.5            1.4
     1          4.9           3.0            1.4
     2          4.7           3.2            1.3
```

```
[15]: # subset row 20 and the species column

      iris.loc[20]['species']
```

```
[15]: 'setosa'
```

Now let's subset using query:

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

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

```
[16]:     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
```

```
[17]: # subset rows where sepal_width is less than 3.5 and the species is `virginica`.

      iris.query('sepal_width < 3.5 and species=="virginica"')
```

```
[17]:      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
```

[18]:
```python
# subset rows where the pedal width is 0.3 or the species is `versicolor`.

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

[18]:
```
     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

[0]: