KEY_Practice16_Pandas-Subsetting-I

January 11, 2020

1 Practice: Subsetting Pandas DataFrames I

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

```
[1]: # 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)
```

Get familiar with the dataset:

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

     iris.head()
```

```
[2]:    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:

```
[3]: # subset the species column

     iris['species']
```

```
[3]: 0        setosa
     1        setosa
     2        setosa
     3        setosa
     4        setosa
            …
```

```
145    virginica
146    virginica
147    virginica
148    virginica
149    virginica
Name: species, Length: 150, dtype: object
```

[4]:
```python
# subset the sepal_length and sepal_width columns

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

[4]:
```
     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
..            ...          ...
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]
```

[5]:
```python
# here's a list of columns we want to keep
columns = ['species', 'petal_width']
# now subset the iris dataframe using the list
iris[columns]
```

[5]:
```
       species  petal_width
0       setosa          0.2
1       setosa          0.2
2       setosa          0.2
3       setosa          0.2
4       setosa          0.2
..         ...          ...
145  virginica          2.3
146  virginica          1.9
147  virginica          2.0
148  virginica          2.3
149  virginica          1.8

[150 rows x 2 columns]
```

Try subsetting on rows:

```
[6]: # select the first 10 rows
     iris.iloc[:10]
     # iris.head(10) # alternative method
```

```
[6]:    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
     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
     8           4.4          2.9           1.4          0.2  setosa
     9           4.9          3.1           1.5          0.1  setosa
```

```
[7]: # think of another way to select the first 10 rows
     iris.head(10)
     # iris.iloc[:11] # alternative method
```

```
[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
     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
     8           4.4          2.9           1.4          0.2  setosa
     9           4.9          3.1           1.5          0.1  setosa
```

```
[8]: # select rows 15 through 20
     iris.iloc[15:21]
```

```
[8]:     sepal_length  sepal_width  petal_length  petal_width species
     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
```

```
[9]: # select row 30
     iris.iloc[30]
```

```
[9]: sepal_length         4.8
     sepal_width          3.1
     petal_length         1.6
     petal_width          0.2
     species           setosa
     Name: 30, dtype: object
```

Bonus: Do you think it's possible to subset with both rows AND columns at the same time? Why or why not?

Feel free to experiment in the space below:

```
[ ]:
```