KEY_Practice15_Pandas-Reading

January 2, 2020

1 Practice: Reading Data with Pandas

Let's start by importing pandas. Remember to give it its special nickname!

[1]:
```
import pandas as pd
```

We have another dataset called `iris` located in the GWC GitHub. Let's load it into a DataFrame called `iris`:

[2]:
```
# this is where the file is located
path = 'https://raw.githubusercontent.com/GWC-DCMB/ClubCurriculum/master/'
filename = path + 'SampleData/iris.csv'

# load the iris dataset into a DataFrame
iris = pd.read_csv(filename)
```

Explore the `DataFrame`; take a look at the beginning and end:

[3]:
```
# View the first few rows

iris.head()
```

[3]:

	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

[4]:
```
# View the last few rows

iris.tail()
```

[4]:

	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
```

After taking a close look at the data, what do you think each row in the DataFrame represents?

Answer: Each row represents an individual flower.

How many rows are in the `DataFrame`? Use `len` to find out:

[5]:
```
# number of rows

len(iris)
```

[5]:
```
150
```

How many columns are in the `DataFrame`? Use `columns` and `len` to find out:

[7]:
```
# number of columns

len(iris.columns)
```

[7]:
```
5
```

How many data points are in the `DataFrame`?

[8]:
```
# calculate the number of data points

len(iris)*len(iris.columns)
```

[8]:
```
750
```