Dogs Data

June 2, 2020

0.1 Slicing and Indexing

0.1.1 Explicit indexes

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[2]:
```
dogs = pd.read_csv('dogs.csv')
```

[7]:
```
dogs.columns
```

[7]:
```
Index(['name', 'breed', 'color', 'height_cm', 'weight_kg'], dtype='object')
```

[8]:
```
dogs.index
```

[8]:
```
RangeIndex(start=0, stop=7, step=1)
```

0.1.2 Setting a column as the index

[9]:
```
dogs
```

[9]:

	name	breed	color	height_cm	weight_kg
0	Bella	Labrador	Brown	56	25
1	Charlie	Poodle	Black	43	23
2	Lucy	Chow Chow	Brown	46	22
3	Cooper	Schnauzer	Grey	49	17
4	Max	Labrador	Black	59	29
5	Stella	Chihuahua	Tan	18	2
6	Bernie	St. Bernard	White	77	74

[11]:
```
dogs_ind = dogs.set_index("name")
dogs_ind
```

[11]:

name	breed	color	height_cm	weight_kg

```
Bella       Labrador  Brown         56         25
Charlie       Poodle  Black         43         23
Lucy       Chow Chow  Brown         46         22
Cooper     Schnauzer   Grey         49         17
Max         Labrador  Black         59         29
Stella     Chihuahua    Tan         18          2
Bernie   St. Bernard  White         77         74
```

0.1.3 Removing an index

[12]:
```
dogs_ind.reset_index()
```

[12]:

	name	breed	color	height_cm	weight_kg
0	Bella	Labrador	Brown	56	25
1	Charlie	Poodle	Black	43	23
2	Lucy	Chow Chow	Brown	46	22
3	Cooper	Schnauzer	Grey	49	17
4	Max	Labrador	Black	59	29
5	Stella	Chihuahua	Tan	18	2
6	Bernie	St. Bernard	White	77	74

0.1.4 Dropping/discarding an index

[16]:
```
dogs_ind.reset_index(drop = True) # Entirely removes dogs names
```

[16]:

	breed	color	height_cm	weight_kg
0	Labrador	Brown	56	25
1	Poodle	Black	43	23
2	Chow Chow	Brown	46	22
3	Schnauzer	Grey	49	17
4	Labrador	Black	59	29
5	Chihuahua	Tan	18	2
6	St. Bernard	White	77	74

0.1.5 Indexes make subsetting simpler

[17]:
```
#subsetting for the rows where the dog is called Bella or Stella
dogs[dogs["name"].isin(["Bella", "Stella"])]
```

[17]:

	name	breed	color	height_cm	weight_kg
0	Bella	Labrador	Brown	56	25
5	Stella	Chihuahua	Tan	18	2

[27]:
```
#Equivalent code when the names are in the index
# loc for selecting rows
#df have a subsetting method called "loc", which filters on index values
#Passing dog names to loc as list
dogs_ind.loc[["Bella", "Stella"]]
```

[27]:

name	breed	color	height_cm	weight_kg
Bella	Labrador	Brown	56	25
Stella	Chihuahua	Tan	18	2

0.1.6 Index values don't need to be unique

[30]:
```
dogs_ind2 = dogs.set_index("breed")
dogs_ind2
# Here there are two labradors in the index
```

[30]:

breed	name	color	height_cm	weight_kg
Labrador	Bella	Brown	56	25
Poodle	Charlie	Black	43	23
Chow Chow	Lucy	Brown	46	22
Schnauzer	Cooper	Grey	49	17
Labrador	Max	Black	59	29
Chihuahua	Stella	Tan	18	2
St. Bernard	Bernie	White	77	74

0.1.7 Subsetting on duplicated index values

[31]:
```
# Subset on "Labrador" using loc, all the Labrador data is returned
dogs_ind2.loc["Labrador"]
```

[31]:

breed	name	color	height_cm	weight_kg
Labrador	Bella	Brown	56	25
Labrador	Max	Black	59	29

0.1.8 Multi-level indexes a.k.a. hierarchical indexes

[34]:
```
# We can include multiple columns in the index by
# passing a list of column names to set index
dogs_ind3 = dogs.set_index(["breed", "color"])
dogs_ind3
```

[34]:
```
                      name  height_cm  weight_kg
breed       color
Labrador    Brown    Bella         56         25
Poodle      Black  Charlie         43         23
Chow Chow   Brown     Lucy         46         22
Schnauzer   Grey    Cooper         49         17
Labrador    Black      Max         59         29
Chihuahua   Tan     Stella         18          2
St. Bernard White   Bernie         77         74
```

0.1.9 Subset the outer level i.e breed with a list

[38]:
```
dogs_ind3.loc[["Labrador", "Chihuahua"]]
# To take a subset of rows at the outer level index, you pass a list
# of index values to loc
# Here the list contains Labrador and Chihuahua, and the resulting
# subset contains all dogs from both breeds
```

[38]:
```
                   name  height_cm  weight_kg
breed     color
Labrador  Brown   Bella         56         25
          Black     Max         59         29
Chihuahua Tan    Stella         18          2
```

0.1.10 Subset inner levels with a list of tuples

[40]:
```
# To subset on inner levels, you need to pass a list of tuples
# Here, the first tuple specifies Labrador at the outer level and
# Brown at the inner level. The resulting rows have to match all
# conditions from a tuple for eg. the black labrador wasn't returned
# because the brown condition wasn't matched
dogs_ind3.loc[[("Labrador", "Brown"), ("Chihuahua", "Tan")]]
```

[40]:
```
                   name  height_cm  weight_kg
breed     color
Labrador  Brown   Bella         56         25
Chihuahua Tan    Stella         18          2
```

0.1.11 Sorting by index values

[42]:
```
# We can also sort by index values using sort_index
# By default it sorts all index levels from outer to inner in
# ascending order
```

```
dogs_ind3.sort_index()
```

[42]:

breed	color	name	height_cm	weight_kg
Chihuahua	Tan	Stella	18	2
Chow Chow	Brown	Lucy	46	22
Labrador	Black	Max	59	29
	Brown	Bella	56	25
Poodle	Black	Charlie	43	23
Schnauzer	Grey	Cooper	49	17
St. Bernard	White	Bernie	77	74

[43]:
```
dogs_ind3
```

[43]:

breed	color	name	height_cm	weight_kg
Labrador	Brown	Bella	56	25
Poodle	Black	Charlie	43	23
Chow Chow	Brown	Lucy	46	22
Schnauzer	Grey	Cooper	49	17
Labrador	Black	Max	59	29
Chihuahua	Tan	Stella	18	2
St. Bernard	White	Bernie	77	74

0.1.12 Controlling sort_index

[44]:
```
dogs_ind3.sort_index(level = ["color", "breed"], ascending = [True, False])
```

[44]:

breed	color	name	height_cm	weight_kg
Poodle	Black	Charlie	43	23
Labrador	Black	Max	59	29
	Brown	Bella	56	25
Chow Chow	Brown	Lucy	46	22
Schnauzer	Grey	Cooper	49	17
Chihuahua	Tan	Stella	18	2
St. Bernard	White	Bernie	77	74

[ ]:
```

```