How it works - Pandas, data selection

In [1]:

```
import pandas as pd
```

In [83]:

```
list1 = ["Jane", "John", "June", "Jim", "Jay"]
list2 = ["Ford", "BMW", "Mini", "Mercedes", "Toyota"]
list3 = ["Blue", "Grey", "Red", "White", "White"]
list4 = ["1.6l", "2.0l", "1.6l", "2.2l", "1.2l"]
df = pd.DataFrame({"Make":list2, "Color":list3, "Capacity":list4}, index = list1
)
```

In [84]:

```
df
```

Out[84]:

	Capacity	Color	Make
Jane	1.6l	Blue	Ford
John	2.0l	Grey	BMW
June	1.6l	Red	Mini
Jim	2.2l	White	Mercedes
Jay	1.2l	White	Toyota

Selection by column

In [85]:

```
df["Capacity"] # Selection of the SERIES by the column name (note the single bra
cket)
```

Out[85]:

```
Jane     1.6l
John     2.0l
June     1.6l
Jim      2.2l
Jay      1.2l
Name: Capacity, dtype: object
```

In [86]:

```python
df[["Capacity"]] # Selection of the DATAFRAME by column name (note the double br
acket)
```

Out[86]:

	Capacity
Jane	1.6l
John	2.0l
June	1.6l
Jim	2.2l
Jay	1.2l

In [87]:

```python
df[["Color", "Capacity"]] # Selection by multiple column names (note the new lis
ted order)
```

Out[87]:

	Color	Capacity
Jane	Blue	1.6l
John	Grey	2.0l
June	Red	1.6l
Jim	White	2.2l
Jay	White	1.2l

In [88]:

```python
df.loc[:,["Color", "Capacity"]] # Selection by column names using the .loc metho
d (columns is the 2nd argument)
```

Out[88]:

	Color	Capacity
Jane	Blue	1.6l
John	Grey	2.0l
June	Red	1.6l
Jim	White	2.2l
Jay	White	1.2l

In [89]:

```
df.iloc[:,[1, 0]] # Selection by column indices using the .iloc method (columns
  is the 2nd argument)
```

Out[89]:

	Color	Capacity
Jane	Blue	1.6l
John	Grey	2.0l
June	Red	1.6l
Jim	White	2.2l
Jay	White	1.2l

Selection by row

In [90]:

```
df.loc[["Jane", "June"],:] # Selection by row names using the .loc method (rows
  is the 1st argument)
```

Out[90]:

	Capacity	Color	Make
Jane	1.6l	Blue	Ford
June	1.6l	Red	Mini

In [91]:

```
df.iloc[[0,2],:] # Selection by row indices using the .iloc method (rows is the
  1st argument)
```

Out[91]:

	Capacity	Color	Make
Jane	1.6l	Blue	Ford
June	1.6l	Red	Mini

Selection by row and column

In [92]:

```
df.loc[["Jane", "June"],["Capacity", "Make"]] # Selection using the .loc method
 would be preferred!
```

Out[92]:

	Capacity	Make
Jane	1.6l	Ford
June	1.6l	Mini

In [93]:

```
df.iloc[[0,1],[0,2]] # Selection using the .iloc method would be preferred!
```

Out[93]:

	Capacity	Make
Jane	1.6l	Ford
John	2.0l	BMW

Selection by filter

In [94]:

```
df[df["Color"] == "White"] # Notice that we essentially filter on a series and t
hen apply the result to the df
                          # df["Color] is the series
                          # we look for values == "White" in that series
                          # apply the result to the df
                          # When in doubt... build the code from the inside ou
t!
```

Out[94]:

	Capacity	Color	Make
Jim	2.2l	White	Mercedes
Jay	1.2l	White	Toyota