

Generate data set

In [1]:

```
import pandas as pd
import numpy as np

dict = {
    "id": [1,2,3,4,5],
    "age": [20,34,76,58,44],
    "gender": ['F', 'M', 'M', 'M', 'F']
}

df = pd.DataFrame(dict)
df
```

Out[1]:

	id	age	gender
0	1	20	F
1	2	34	M
2	3	76	M
3	4	58	M
4	5	44	F

Add new variable

In [2]:

```
df["height"] = [158,164,170,165,161]
df
```

Out[2]:

	id	age	gender	height
0	1	20	F	158
1	2	34	M	164
2	3	76	M	170
3	4	58	M	165
4	5	44	F	161

Add rows

In [3]:

```
df.append(pd.Series([6, 32, "F", 173], index=df.columns), ignore_index=True)
```

Out[3]:

	id	age	gender	height
0	1	20	F	158
1	2	34	M	164
2	3	76	M	170
3	4	58	M	165
4	5	44	F	161
5	6	32	F	173

Combine data frames based on id

In [4]:

```
dict2 = {  
    "id": [1,2,3,5],  
    "weight": [60,64,80,82],  
}  
  
df2 = pd.DataFrame(dict2)  
df2
```

Out[4]:

	id	weight
0	1	60
1	2	64
2	3	80
3	5	82

In [5]:

```
df = df.merge(df2, on="id", how="left")
df
```

Out[5]:

	id	age	gender	height	weight
0	1	20	F	158	60.0
1	2	34	M	164	64.0
2	3	76	M	170	80.0
3	4	58	M	165	NaN
4	5	44	F	161	82.0

Filter rows

In [6]:

```
df[df["age"] >= 40]
```

Out[6]:

	id	age	gender	height	weight
2	3	76	M	170	80.0
3	4	58	M	165	NaN
4	5	44	F	161	82.0

Select columns

In [7]:

```
df3 = df[["id", "age", "height"]]
df3
```

Out[7]:

	id	age	height
0	1	20	158
1	2	34	164
2	3	76	170
3	4	58	165
4	5	44	161

Drop variables

In [8]:

```
df4 = df.drop(["age", "gender"], axis=1)
df4
```

Out[8]:

	id	height	weight
0	1	158	60.0
1	2	164	64.0
2	3	170	80.0
3	4	165	NaN
4	5	161	82.0

Batch modify

Add 1 year to everybody's age

In [9]:

```
df["age"] = df["age"] + 1
df
```

Out[9]:

	id	age	gender	height	weight
0	1	21	F	158	60.0
1	2	35	M	164	64.0
2	3	77	M	170	80.0
3	4	59	M	165	NaN
4	5	45	F	161	82.0

Add 2 kg of weight to all males

In [10]:

```
df.loc[(df["gender"]=="M"), "weight"] = df.loc[(df["gender"]=="M"), "weight"] + 2
df
```

Out[10]:

	id	age	gender	height	weight
0	1	21	F	158	60.0
1	2	35	M	164	66.0
2	3	77	M	170	82.0
3	4	59	M	165	NaN
4	5	45	F	161	82.0

Drop rows

Drop row where id = 4. First generate an index of the rows to be dropped. Then pass it to df.drop

In [11]:

```
idx = df[df["id"]==4].index  
df5 = df.drop(idx)  
df5
```

Out[11]:

	id	age	gender	height	weight
0	1	21	F	158	60.0
1	2	35	M	164	66.0
2	3	77	M	170	82.0
4	5	45	F	161	82.0

In []: