Regiment

Introduction:

Special thanks to: http://chrisalbon.com/ for sharing the dataset and materials.

Step 1. Import the necessary libraries

In [1]:
```
import pandas as pd
import numpy as np
```

Step 2. Create the DataFrame with the following values:

In [2]:
```
raw_data = {'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks', 'Nighthawks', 'Dragoons', 'Dragoons', 'Dragoons', 'Dragoons',
        'company': ['1st', '1st', '2nd', '2nd', '1st', '1st', '2nd', '2nd','1st', '1st', '2nd', '2nd'],
        'name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze', 'Jacon', 'Ryaner', 'Sone', 'Sloan', 'Piger', 'Riani', 'Ali'],
        'preTestScore': [4, 24, 31, 2, 3, 4, 24, 31, 2, 3, 2, 3],
        'postTestScore': [25, 94, 57, 62, 70, 25, 94, 57, 62, 70, 62, 70]}
```

Step 3. Assign it to a variable called regiment.

Don't forget to name each column

In [3]:
```
re = pd.DataFrame(raw_data,columns=['regiment', 'company', 'name', 'battles', 'preTestScore', 'postTestScore'])
re
```

Out[3]:

	regiment	company	name	battles	preTestScore	postTestScore
0	Nighthawks	1st	Miller	NaN	4	25
1	Nighthawks	1st	Jacobson	NaN	24	94
2	Nighthawks	2nd	Ali	NaN	31	57
3	Nighthawks	2nd	Milner	NaN	2	62
4	Dragoons	1st	Cooze	NaN	3	70
5	Dragoons	1st	Jacon	NaN	4	25
6	Dragoons	2nd	Ryaner	NaN	24	94
7	Dragoons	2nd	Sone	NaN	31	57
8	Scouts	1st	Sloan	NaN	2	62
9	Scouts	1st	Piger	NaN	3	70
10	Scouts	2nd	Riani	NaN	2	62
11	Scouts	2nd	Ali	NaN	3	70

Step 4. What is the mean preTestScore from the regiment Nighthawks?

In [7]:
```
re[re['regiment'] == 'Nighthawks'].groupby('regiment').mean()
```

Out[7]:

regiment	preTestScore	postTestScore
Nighthawks	15.25	59.5

Step 5. Present general statistics by company

In [12]:
```
re.groupby("company").describe()
```

Out[12]:

	preTestScore								postTestScore							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
company																
1st	6.0	6.666667	8.524475	2.0	3.00	3.5	4.00	24.0	6.0	57.666667	27.485754	25.0	34.25	66.0	70.0	94.0
2nd	6.0	15.500000	14.652645	2.0	2.25	13.5	29.25	31.0	6.0	67.000000	14.057027	57.0	58.25	62.0	68.0	94.0

Step 6. What is the mean of each company's preTestScore?

In [24]:
```
re.groupby("company").preTestScore.mean()
```

Out[24]:
```
company
1st     6.666667
2nd    15.500000
Name: preTestScore, dtype: float64
```

Step 7. Present the mean preTestScores grouped by regiment and company

In [19]:
```
re.groupby(["company","regiment"]).preTestScore.mean()
```

Out[19]:
```
company  regiment
1st      Dragoons       3.5
         Nighthawks    14.0
         Scouts         2.5
2nd      Dragoons      27.5
         Nighthawks    16.5
         Scouts         2.5
Name: preTestScore, dtype: float64
```

Step 8. Present the mean preTestScores grouped by regiment and company without heirarchical indexing

In [18]:
```
re.groupby(["company","regiment"]).preTestScore.mean().unstack()
```

Out[18]:

regiment	Dragoons	Nighthawks	Scouts
company			
1st	3.5	14.0	2.5
2nd	27.5	16.5	2.5

Step 9. Group the entire dataframe by regiment and company

In [20]:
```
re.groupby(["company","regiment"]).mean()
```

Out[20]:

		preTestScore	postTestScore
company	regiment		
1st	Dragoons	3.5	47.5
	Nighthawks	14.0	59.5
	Scouts	2.5	66.0
2nd	Dragoons	27.5	75.5
	Nighthawks	16.5	59.5
	Scouts	2.5	66.0

Step 10. What is the number of observations in each regiment and company

In [21]:
```
re.groupby(["company","regiment"]).size()
```

Out[21]:
```
company  regiment
1st      Dragoons      2
         Nighthawks    2
         Scouts        2
2nd      Dragoons      2
         Nighthawks    2
         Scouts        2
dtype: int64
```

Step 11. Iterate over a group and print the name and the whole data from the regiment

In [23]:
```
for name, group in re.groupby('regiment'):
    print(name)
    print(group)
```

```
Dragoons
   regiment company    name battles  preTestScore  postTestScore
4  Dragoons     1st   Cooze     NaN             3             70
5  Dragoons     1st   Jacon     NaN             4             25
6  Dragoons     2nd  Ryaner     NaN            24             94
7  Dragoons     2nd    Sone     NaN            31             57
Nighthawks
     regiment company      name battles  preTestScore  postTestScore
0  Nighthawks     1st    Miller     NaN             4             25
1  Nighthawks     1st  Jacobson     NaN            24             94
2  Nighthawks     2nd       Ali     NaN            31             57
3  Nighthawks     2nd    Milner     NaN             2             62
Scouts
   regiment company   name battles  preTestScore  postTestScore
8    Scouts     1st  Sloan     NaN             2             62
9    Scouts     1st  Piger     NaN             3             70
10   Scouts     2nd  Riani     NaN             2             62
11   Scouts     2nd    Ali     NaN             3             70
```