Exercise- Filtering and Sorting Data-Fictional Army Dataset

Step 1. Import the necessary libraries

In [1]:

```
import pandas as pd
```

Step 2. This is the data given as a dictionary

In [2]:

```
# Create an example dataframe about a fictional army
raw_data = {'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks', 'Nighthawks', 'Drag
            'company': ['1st', '1st', '2nd', '2nd', '1st', '1st', '2nd', '2nd','1st',
            'deaths': [523, 52, 25, 616, 43, 234, 523, 62, 62, 73, 37, 35],
            'battles': [5, 42, 2, 2, 4, 7, 8, 3, 4, 7, 8, 9],
            'size': [1045, 957, 1099, 1400, 1592, 1006, 987, 849, 973, 1005, 1099, 152
            'veterans': [1, 5, 62, 26, 73, 37, 949, 48, 48, 435, 63, 345],
            'readiness': [1, 2, 3, 3, 2, 1, 2, 3, 2, 1, 2, 3],
            'armored': [1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1],
            'deserters': [4, 24, 31, 2, 3, 4, 24, 31, 2, 3, 2, 3],
            'origin': ['Arizona', 'California', 'Texas', 'Florida', 'Maine', 'Iowa',
```

Step 3. Create a dataframe and assign it to a variable called army.

Don't forget to include the columns names in the order presented in the dictionary ('regiment', 'company', 'deaths'...) so that the column index order is consistent with the solutions. If omitted, pandas will order the columns alphabetically.

In [3]:

```
dataframe = pd.DataFrame(raw_data)
dataframe
```

Out[3]:

	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters	origin
0	Nighthawks	1st	523	5	1045	1	1	1	4	Arizona
1	Nighthawks	1st	52	42	957	5	2	0	24	California
2	Nighthawks	2nd	25	2	1099	62	3	1	31	Texas
3	Nighthawks	2nd	616	2	1400	26	3	1	2	Florida
4	Dragoons	1st	43	4	1592	73	2	0	3	Maine
5	Dragoons	1st	234	7	1006	37	1	1	4	Iowa
6	Dragoons	2nd	523	8	987	949	2	0	24	Alaska
7	Dragoons	2nd	62	3	849	48	3	1	31	Washington
8	Scouts	1st	62	4	973	48	2	0	2	Oregon
9	Scouts	1st	73	7	1005	435	1	0	3	Wyoming
10	Scouts	2nd	37	8	1099	63	2	1	2	Louisana
11	Scouts	2nd	35	9	1523	345	3	1	3	Georgia

Step 4. Set the 'origin' colum as the index of the dataframe

In [4]:

```
dataframe = dataframe.set_index('origin')
```

In [5]:

```
dataframe
```

Out[5]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
California	Nighthawks	1st	52	42	957	5	2	0	24
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Maine	Dragoons	1st	43	4	1592	73	2	0	3
Iowa	Dragoons	1st	234	7	1006	37	1	1	4
Alaska	Dragoons	2nd	523	8	987	949	2	0	24
Washington	Dragoons	2nd	62	3	849	48	3	1	31
Oregon	Scouts	1st	62	4	973	48	2	0	2
Wyoming	Scouts	1st	73	7	1005	435	1	0	3
Louisana	Scouts	2nd	37	8	1099	63	2	1	2
Georgia	Scouts	2nd	35	9	1523	345	3	1	3

Step 5. Print only the column veterans

In [6]:

```
dataframe[['veterans']]
```

Out[6]:

origin	veterans
Arizona	1
California	5
Texas	62
Florida	26
Maine	73
Iowa	37
Alaska	949
Washington	48
Oregon	48
Wyoming	435
Louisana	63
Georgia	345

Step 6. Print the columns 'veterans' and 'deaths'

In [7]:

```
dataframe[['veterans','deaths']]
```

Out[7]:

origin	veterans	deaths
Arizona	1	523
California	5	52
Texas	62	25
Florida	26	616
Maine	73	43
Iowa	37	234
Alaska	949	523
Washington	48	62
Oregon	48	62
Wyoming	435	73
Louisana	63	37
Georgia	345	35

Step 7. Print the name of all the columns.

In [8]:

```
print(dataframe.columns.values)
```

```
['regiment' 'company' 'deaths' 'battles' 'size' 'veterans' 'readiness'
 'armored' 'deserters']
```

Step 8. Select the 'deaths', 'size' and 'deserters' columns from Maine and Alaska

In [9]:

```
dataframe.iloc[[4,6],[2,4,8]]
```

Out[9]:

origin	deaths	size	deserters
Maine	43	1592	3
Alaska	523	987	24

Step 9. Select the rows 3 to 7 and the columns 3 to 6

In [10]:

```
dataframe.iloc[[3,4,5,6,7],[3,4,5,6]]
```

Out[10]:

origin	battles	size	veterans	readiness
Florida	2	1400	26	3
Maine	4	1592	73	2
Iowa	7	1006	37	1
Alaska	8	987	949	2
Washington	3	849	48	3

Step 10. Select every row after the fourth row and all columns

In [11]:

```
dataframe.iloc[[3,4,5,6,7],[3,4,5,6]]
```

Out[11]:

origin	battles	size	veterans	readiness
Florida	2	1400	26	3
Maine	4	1592	73	2
Iowa	7	1006	37	1
Alaska	8	987	949	2
Washington	3	849	48	3

Step 11. Select every row up to the 4th row and all columns

In [12]:

```
dataframe.iloc[[0,1,2,3,4]]
```

Out[12]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
California	Nighthawks	1st	52	42	957	5	2	0	24
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Maine	Dragoons	1st	43	4	1592	73	2	0	3

Step 12. Select the 3rd column up to the 7th column

In [13]:

```
dataframe.iloc[:,[3,4,5,6,7]]
```

Out[13]:

origin	battles	size	veterans	readiness	armored
Arizona	5	1045	1	1	1
California	42	957	5	2	0
Texas	2	1099	62	3	1
Florida	2	1400	26	3	1
Maine	4	1592	73	2	0
Iowa	7	1006	37	1	1
Alaska	8	987	949	2	0
Washington	3	849	48	3	1
Oregon	4	973	48	2	0
Wyoming	7	1005	435	1	0
Louisana	8	1099	63	2	1
Georgia	9	1523	345	3	1

Step 13. Select rows where df.deaths is greater than 50

In [14]:

```
dataframe[dataframe['deaths']>50]
```

Out[14]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
California	Nighthawks	1st	52	42	957	5	2	0	24
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Iowa	Dragoons	1st	234	7	1006	37	1	1	4
Alaska	Dragoons	2nd	523	8	987	949	2	0	24
Washington	Dragoons	2nd	62	3	849	48	3	1	31
Oregon	Scouts	1st	62	4	973	48	2	0	2
Wyoming	Scouts	1st	73	7	1005	435	1	0	3

Step 14. Select rows where df.deaths is greater than 500 or less than 50

In [15]:

```
dataframe[dataframe['deaths']>500]
```

Out[15]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Alaska	Dragoons	2nd	523	8	987	949	2	0	24

In [16]:

```
dataframe[dataframe['deaths']<50]
```

Out[16]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31
Maine	Dragoons	1st	43	4	1592	73	2	0	3
Louisana	Scouts	2nd	37	8	1099	63	2	1	2
Georgia	Scouts	2nd	35	9	1523	345	3	1	3

Step 15. Select all the regiments not named "Dragoons"

In [17]:

```
dataframe[dataframe['regiment'] != 'Dragoons']
```

Out[17]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
California	Nighthawks	1st	52	42	957	5	2	0	24
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Oregon	Scouts	1st	62	4	973	48	2	0	2
Wyoming	Scouts	1st	73	7	1005	435	1	0	3
Louisana	Scouts	2nd	37	8	1099	63	2	1	2
Georgia	Scouts	2nd	35	9	1523	345	3	1	3

Step 16. Select the rows called Texas and Arizona

In [18]:

```
dataframe.iloc[[0,2]]
```

Out[18]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31

Step 17. Select the third cell in the row named Arizona

In [19]:

```
dataframe.iloc[[0],[2]]
```

Out[19]:

origin	deaths
Arizona	523

Step 18. Select the third cell in the column named deaths

In [20]:

```
dataframe.iloc[[2],[2]]
```

Out[20]:

origin	deaths
Texas	25