Exercise- Filtering and Sorting Data-Fictional Army Dataset

Step 1. Import the necessary libraries

In [4]:
```
import pandas as pd
```

Step 2. This is the data given as a dictionary

In [5]:
```
# Create an example dataframe about a fictional army
raw_data = {'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks', 'Nighthawks', 'Dragoons', 'Dragoons', 'Dragoons', 'Dragoons',
            'company': ['1st', '1st', '2nd', '2nd', '1st', '1st', '2nd', '2nd','1st', '1st', '2nd', '2nd'],
            '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, 1523],
            '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', 'Alaska', 'Washington', 'Oregon', 'Wyoming'
```

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 [30]:
```
army=pd.DataFrame(raw_data)
army
```

Out[30]:

	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 [33]:
```
army=army.set_index(army.origin)
army
```

Out[33]:

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

Step 5. Print only the column veterans

In [34]:
```
army.veterans
```

Out[34]:
```
origin
Arizona          1
California       5
Texas           62
Florida         26
Maine           73
Iowa            37
Alaska         949
Washington      48
Oregon          48
Wyoming        435
Louisana        63
Georgia        345
Name: veterans, dtype: int64
```

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

In [35]:
```
army[['veterans','deaths']]
```

Out[35]:

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 [41]:
```
army.columns.values
```

Out[41]:
```
array(['regiment', 'company', 'deaths', 'battles', 'size', 'veterans',
       'readiness', 'armored', 'deserters', 'origin'], dtype=object)
```

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

In [45]:
```
#1st Method
army[(army.origin=='Maine')|(army.origin=='Alaska')][['deaths', 'size', 'deserters']]
```

Out[45]:

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

In [48]:
```
#2nd Method
army.loc[['Maine', 'Alaska'], ['deaths', 'size', 'deserters']]
```

Out[48]:

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 [49]:
```
army.iloc[2:7,2:6]
```

Out[49]:

origin	deaths	battles	size	veterans
Texas	25	2	1099	62
Florida	616	2	1400	26
Maine	43	4	1592	73
Iowa	234	7	1006	37
Alaska	523	8	987	949

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

In [55]:
```
army.iloc[5:,:]
```

Out[55]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters	origin
Iowa	Dragoons	1st	234	7	1006	37	1	1	4	Iowa
Alaska	Dragoons	2nd	523	8	987	949	2	0	24	Alaska
Washington	Dragoons	2nd	62	3	849	48	3	1	31	Washington
Oregon	Scouts	1st	62	4	973	48	2	0	2	Oregon
Wyoming	Scouts	1st	73	7	1005	435	1	0	3	Wyoming
Louisana	Scouts	2nd	37	8	1099	63	2	1	2	Louisana
Georgia	Scouts	2nd	35	9	1523	345	3	1	3	Georgia

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

In [52]:
```
army.head()
```

Out[52]:

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

In [53]:
```
army.iloc[:4,:]
```

Out[53]:

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

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

In [57]:
```
army.iloc[:,3:8]
```

Out[57]:

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 [61]:
```
#1st Method
army.loc[army['deaths']>=50,]
```

Out[61]:

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

In [63]:
```
#2nd Method
army[army['deaths']>50]
```

Out[63]:

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

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

In [73]:
```
army[(army['deaths']>500) | (army['deaths']<50)]
```

Out[73]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters	origin
Arizona	Nighthawks	1st	523	5	1045	1	1	1	4	Arizona
Texas	Nighthawks	2nd	25	2	1099	62	3	1	31	Texas
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2	Florida
Maine	Dragoons	1st	43	4	1592	73	2	0	3	Maine
Alaska	Dragoons	2nd	523	8	987	949	2	0	24	Alaska
Louisana	Scouts	2nd	37	8	1099	63	2	1	2	Louisana
Georgia	Scouts	2nd	35	9	1523	345	3	1	3	Georgia

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

In [79]:
```
#1st Methid
army[~(army['regiment']=='Dragoons')]
```

Out[79]:

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

In [80]:
```
#2nd Methid
army[(army['regiment']!='Dragoons')]
```

Out[80]:

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

Step 16. Select the rows called Texas and Arizona

In [93]:
```
#1st Method
army[(army['origin']=='Texas')| (army['origin']=='Arizona')]
```

Out[93]:

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

In [92]:
```
#2nd Method
army.loc[['Texas', 'Arizona']]
```

Out[92]:

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

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

In [98]:
```
#1st Method
army.loc[['Arizona'], ['deaths']]
```

Out[98]:

origin	deaths
Arizona	523

In [99]:
```
#2nd Method
army.iloc[[0], army.columns.get_loc('deaths')]
```

Out[99]:
```
origin
Arizona    523
Name: deaths, dtype: int64
```

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

In [100…
```
#1st Method
army.loc['Texas', 'deaths']
```

Out[100]:
```
25
```

In [101…
```
#2nd Method
army.iloc[[2], army.columns.get_loc('deaths')]
```

Out[101]:
```
origin
Texas    25
Name: deaths, dtype: int64
```

In [ ]: