Submitted By: Muhammad Arham Adeel

Ex3 - Filtering and Sorting Data - Fictional Army

Introduction:

This exercise was inspired by this page

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 [11]:
```
army = pd.DataFrame(raw_data)
army
```

Out[11]:

	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters	origi
0	Nighthawks	1st	523	5	1045	1	1	1	4	Arizon
1	Nighthawks	1st	52	42	957	5	2	0	24	Californi
2	Nighthawks	2nd	25	2	1099	62	3	1	31	Texa
3	Nighthawks	2nd	616	2	1400	26	3	1	2	Florid
4	Dragoons	1st	43	4	1592	73	2	0	3	Main
5	Dragoons	1st	234	7	1006	37	1	1	4	Iow
6	Dragoons	2nd	523	8	987	949	2	0	24	Alask
7	Dragoons	2nd	62	3	849	48	3	1	31	Washingto
8	Scouts	1st	62	4	973	48	2	0	2	Orego
9	Scouts	1st	73	7	1005	435	1	0	3	Wyomin
10	Scouts	2nd	37	8	1099	63	2	1	2	Louisan
11	Scouts	2nd	35	9	1523	345	3	1	3	Georgi

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

In [12]:
```
army.set_index("origin",inplace = True)
```

In [13]:
```
army
```

Out[13]:

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 [14]:
```
army[["veterans"]]
```

Out[14]:

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 [15]:
```
army[["veterans","deaths"]]
```

Out[15]:

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 [16]:
```
army.columns
```

Out[16]:
```
Index(['regiment', 'company', 'deaths', 'battles', 'size', 'veterans',
       'readiness', 'armored', 'deserters'],
      dtype='object')
```

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

In [21]:
```
maine_alska = army.groupby("origin")["deaths","size","deserters"].sum()
maine_alska
```

```
C:\Users\eAgLe\AppData\Local\Temp\ipykernel_1888\347496838.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
  maine_alska = army.groupby("origin")["deaths","size","deserters"].sum()
```

Out[21]:

origin	deaths	size	deserters
Alaska	523	987	24
Arizona	523	1045	4
California	52	957	24
Florida	616	1400	2
Georgia	35	1523	3
Iowa	234	1006	4
Louisana	37	1099	2
Maine	43	1592	3
Oregon	62	973	2
Texas	25	1099	31
Washington	62	849	31
Wyoming	73	1005	3

In [22]:
```
## Select only Alska and maine

maine_alska.loc[["Alaska","Maine"],:]
```

Out[22]:

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

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

In [23]:
```
army.iloc[3:8,3:7]
```

Out[23]:

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 [24]:
```
army.iloc[4 : , :]
```

Out[24]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
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 11. Select every row up to the 4th row and all columns

In [25]:
```
army.iloc[ : 5 , :]
```

Out[25]:

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 [28]:
```
army.iloc[ : , 3 : 8]
```

Out[28]:

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 [26]:
```
army[army["deaths"] > 50]
```

Out[26]:

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 [31]:
```
army[(army["deaths"] > 500) | (army["deaths"] < 50)]
```

Out[31]:

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
Florida	Nighthawks	2nd	616	2	1400	26	3	1	2
Maine	Dragoons	1st	43	4	1592	73	2	0	3
Alaska	Dragoons	2nd	523	8	987	949	2	0	24
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 [32]:
```
army[army["regiment"] != "Dragoons" ]
```

Out[32]:

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 [33]:
```
army.loc[["Texas","Arizona"], :]
```

Out[33]:

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

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

In [34]:
```
army.loc["Arizona"]
```

Out[34]:
```
regiment     Nighthawks
company             1st
deaths              523
battles               5
size               1045
veterans              1
readiness             1
armored               1
deserters             4
Name: Arizona, dtype: object
```

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

In [38]:
```
army.iloc[-3 , :]
```

Out[38]:
```
regiment     Scouts
company         1st
deaths           73
battles           7
size           1005
veterans        435
readiness         1
armored           0
deserters         3
Name: Wyoming, dtype: object
```

In [39]:
```
army.iloc[-3: , :]
```

Out[39]:

origin	regiment	company	deaths	battles	size	veterans	readiness	armored	deserters
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