Exercise- Filtering and Sorting Data- Fictional Army Dataset

Step 1. Import the necessary libraries

Step 2. This is the data given as a dictionary

In [11]:
```
# Create an example dataframe about a fictional army
raw_data = {'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks', 'Nighthaw
            'company': ['1st', '1st', '2nd', '2nd', '1st', '1st', '2nd', '2r
            '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,
            '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',
```

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.

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

In [2]:
```
import pandas as pd

raw_data = {
    'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks', 'Nighthawks', 'Dr
    '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, 1
    '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',
}

# Create the DataFrame
army = pd.DataFrame(raw_data)
```

```
# Set the 'origin' column as the index
army.set_index('origin', inplace=True)

# Display the DataFrame with the updated index
print(army)
```

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

            armored  deserters
origin
Arizona           1          4
California        0         24
Texas             1         31
Florida           1          2
Maine             0          3
Iowa              1          4
Alaska            0         24
Washington        1         31
Oregon            0          2
Wyoming           0          3
Louisiana         1          2
Georgia           1          3
```

Step 5. Print only the column veterans

In [3]:
```
print(army['veterans'])
```

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

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

In [4]:
```
print(army[['veterans', 'deaths']])
```

```
           veterans  deaths
origin
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
Louisiana        63      37
Georgia         345      35
```

Step 7. Print the name of all the columns.

In [5]:
```
print(army.columns)
```

```
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 [6]:
```
selected_columns = ['deaths', 'size', 'deserters']
selected_rows = ['Maine', 'Alaska']

result = army.loc[selected_rows, selected_columns]
print(result)
```

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

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

In [7]:
```
selected_rows = range(2, 7)  # Rows 3 to 7 (0-based indexing)
selected_columns = range(2, 6)  # Columns 3 to 6 (0-based indexing)

result = army.iloc[selected_rows, selected_columns]
print(result)
```

```
         deaths  battles  size  veterans
origin
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 [8]:
```
selected_rows = slice(4, None)  # Rows from the 5th row onwards (0-based inc
selected_columns = slice(None)  # All columns

result = army.iloc[selected_rows, selected_columns]
print(result)
```

```
           regiment company  deaths  battles  size  veterans  readiness  \
origin
Maine       Dragoons     1st      43        4  1592        73          2
Iowa        Dragoons     1st     234        7  1006        37          1
Alaska      Dragoons     2nd     523        8   987       949          2
Washington  Dragoons     2nd      62        3   849        48          3
Oregon        Scouts     1st      62        4   973        48          2
Wyoming       Scouts     1st      73        7  1005       435          1
Louisiana     Scouts     2nd      37        8  1099        63          2
Georgia       Scouts     2nd      35        9  1523       345          3

           armored  deserters
origin
Maine            0          3
Iowa             1          4
Alaska           0         24
Washington       1         31
Oregon           0          2
Wyoming          0          3
Louisiana        1          2
Georgia          1          3
```

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

In [9]:
```
selected_rows = slice(0, 4)  # Rows from the 1st to 4th row (0-based indexir
selected_columns = slice(None)  # All columns

result = army.iloc[selected_rows, selected_columns]
print(result)
```

```
              regiment company  deaths  battles  size  veterans  readiness  \
origin
Arizona     Nighthawks     1st     523        5  1045         1          1
California  Nighthawks     1st      52       42   957         5          2
Texas       Nighthawks     2nd      25        2  1099        62          3
Florida     Nighthawks     2nd     616        2  1400        26          3

            armored  deserters
origin
Arizona           1          4
California        0         24
Texas             1         31
Florida           1          2
```

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

In [10]:
```
selected_columns = army.iloc[:, 2:7]  # Columns from the 3rd to 7th column (
print(selected_columns)
```

```
            deaths  battles  size  veterans  readiness
origin
Arizona        523        5  1045         1          1
California      52       42   957         5          2
Texas           25        2  1099        62          3
Florida        616        2  1400        26          3
Maine           43        4  1592        73          2
Iowa           234        7  1006        37          1
Alaska         523        8   987       949          2
Washington      62        3   849        48          3
Oregon          62        4   973        48          2
Wyoming         73        7  1005       435          1
Louisiana       37        8  1099        63          2
Georgia         35        9  1523       345          3
```

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

In [11]:
```
selected_rows = army[army['deaths'] > 50]
print(selected_rows)
```

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

            armored  deserters
origin
Arizona           1          4
California        0         24
Florida           1          2
Iowa              1          4
Alaska            0         24
Washington        1         31
Oregon            0          2
Wyoming           0          3
```

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

In [12]:
```
selected_rows = army[(army['deaths'] > 500) | (army['deaths'] < 50)]
print(selected_rows)
```

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

            armored  deserters
origin
Arizona           1          4
Texas             1         31
Florida           1          2
Maine             0          3
Alaska            0         24
Louisiana         1          2
Georgia           1          3
```

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

In [13]:
```
selected_rows = army[army['regiment'] != 'Dragoons']
print(selected_rows)
```

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

            armored  deserters
origin
Arizona           1          4
California        0         24
Texas             1         31
Florida           1          2
Oregon            0          2
Wyoming           0          3
Louisiana         1          2
Georgia           1          3
```

Step 16. Select the rows called Texas and Arizona

In [25]:
```
selected_rows = army[army.isin(['Texas', 'Arizona'])]
print(selected_rows)
```

```
            regiment company  deaths  battles  size  veterans  readiness  \
origin
Arizona          NaN     NaN     NaN      NaN   NaN       NaN        NaN
California       NaN     NaN     NaN      NaN   NaN       NaN        NaN
Texas            NaN     NaN     NaN      NaN   NaN       NaN        NaN
Florida          NaN     NaN     NaN      NaN   NaN       NaN        NaN
Maine            NaN     NaN     NaN      NaN   NaN       NaN        NaN
Iowa             NaN     NaN     NaN      NaN   NaN       NaN        NaN
Alaska           NaN     NaN     NaN      NaN   NaN       NaN        NaN
Washington       NaN     NaN     NaN      NaN   NaN       NaN        NaN
Oregon           NaN     NaN     NaN      NaN   NaN       NaN        NaN
Wyoming          NaN     NaN     NaN      NaN   NaN       NaN        NaN
Louisiana        NaN     NaN     NaN      NaN   NaN       NaN        NaN
Georgia          NaN     NaN     NaN      NaN   NaN       NaN        NaN

            armored  deserters
origin
Arizona         NaN        NaN
California      NaN        NaN
Texas           NaN        NaN
Florida         NaN        NaN
Maine           NaN        NaN
Iowa            NaN        NaN
Alaska          NaN        NaN
Washington      NaN        NaN
Oregon          NaN        NaN
Wyoming         NaN        NaN
Louisiana       NaN        NaN
Georgia         NaN        NaN
```

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

In [26]:
```
selected_cell = army.loc['Arizona', 'deaths']
print(selected_cell)
```

```
523
```

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

In [27]:
```
selected_cell = army.iloc[2, army.columns.get_loc('deaths')]
print(selected_cell)
```

```
25
```

In [ ]:
```

```