Pandas DataFrame exercises

```python
# Import the numpy package under the name np
import numpy as np

# Import the pandas package under the name pd
import pandas as pd

# Import the matplotlib package under the name plt
import matplotlib.pyplot as plt
%matplotlib inline

# Print the pandas version and the configuration
print(pd.__version__)
```

2.2.2

DataFrame creation

Create an empty pandas DataFrame

In [2]:
```python
# your code goes here
```

In [3]:
```python
pd.DataFrame(data=[None],
             index=[None],
             columns=[None])
```

Out[3]:

	None
None	None

Create a `marvel_df` pandas DataFrame with the given marvel data

```
In [4]: marvel_data = [
    ['Spider-Man', 'male', 1962],
    ['Captain America', 'male', 1941],
    ['Wolverine', 'male', 1974],
    ['Iron Man', 'male', 1963],
    ['Thor', 'male', 1963],
    ['Thing', 'male', 1961],
    ['Mister Fantastic', 'male', 1961],
    ['Hulk', 'male', 1962],
    ['Beast', 'male', 1963],
    ['Invisible Woman', 'female', 1961],
    ['Storm', 'female', 1975],
    ['Namor', 'male', 1939],
    ['Hawkeye', 'male', 1964],
    ['Daredevil', 'male', 1964],
    ['Doctor Strange', 'male', 1963],
    ['Hank Pym', 'male', 1962],
    ['Scarlet Witch', 'female', 1964],
    ['Wasp', 'female', 1963],
    ['Black Widow', 'female', 1964],
    ['Vision', 'male', 1968]
]
```

```
In [5]: # your code goes here
```

```
In [6]: marvel_df = pd.DataFrame(data=marvel_data)
```

```
marvel_df
```

Out[6]:

	0	1	2
0	Spider-Man	male	1962
1	Captain America	male	1941
2	Wolverine	male	1974
3	Iron Man	male	1963
4	Thor	male	1963
5	Thing	male	1961
6	Mister Fantastic	male	1961
7	Hulk	male	1962
8	Beast	male	1963
9	Invisible Woman	female	1961
10	Storm	female	1975
11	Namor	male	1939
12	Hawkeye	male	1964
13	Daredevil	male	1964
14	Doctor Strange	male	1963
15	Hank Pym	male	1962
16	Scarlet Witch	female	1964
17	Wasp	female	1963
18	Black Widow	female	1964
19	Vision	male	1968

Add column names to the `marvel_df`

In [7]:
```python
# your code goes here
```

In [8]:
```python
col_names = ['name', 'gender', 'first_appearance']

marvel_df.columns = col_names
marvel_df
```

Out[8]:

	name	gender	first_appearance
0	Spider-Man	male	1962
1	Captain America	male	1941
2	Wolverine	male	1974
3	Iron Man	male	1963
4	Thor	male	1963
5	Thing	male	1961
6	Mister Fantastic	male	1961
7	Hulk	male	1962
8	Beast	male	1963
9	Invisible Woman	female	1961
10	Storm	female	1975
11	Namor	male	1939
12	Hawkeye	male	1964
13	Daredevil	male	1964
14	Doctor Strange	male	1963
15	Hank Pym	male	1962
16	Scarlet Witch	female	1964
17	Wasp	female	1963
18	Black Widow	female	1964
19	Vision	male	1968

Add index names to the `marvel_df` (use the character name as index)

In [9]:
```python
# your code goes here
```

In [10]:
```python
marvel_df.index = marvel_df['name']
marvel_df
```

name	name	gender	first_appearance
Spider-Man	Spider-Man	male	1962
Captain America	Captain America	male	1941
Wolverine	Wolverine	male	1974
Iron Man	Iron Man	male	1963
Thor	Thor	male	1963
Thing	Thing	male	1961
Mister Fantastic	Mister Fantastic	male	1961
Hulk	Hulk	male	1962
Beast	Beast	male	1963
Invisible Woman	Invisible Woman	female	1961
Storm	Storm	female	1975
Namor	Namor	male	1939
Hawkeye	Hawkeye	male	1964
Daredevil	Daredevil	male	1964
Doctor Strange	Doctor Strange	male	1963
Hank Pym	Hank Pym	male	1962
Scarlet Witch	Scarlet Witch	female	1964
Wasp	Wasp	female	1963
Black Widow	Black Widow	female	1964
Vision	Vision	male	1968

Drop the name column as it's now the index

In [11]:
```python
# your code goes here
```

In [12]:
```python
#marvel_df = marvel_df.drop(columns=['name'])
marvel_df = marvel_df.drop(['name'], axis=1)
marvel_df
```

name	gender	first_appearance
Spider-Man	male	1962
Captain America	male	1941
Wolverine	male	1974
Iron Man	male	1963
Thor	male	1963
Thing	male	1961
Mister Fantastic	male	1961
Hulk	male	1962
Beast	male	1963
Invisible Woman	female	1961
Storm	female	1975
Namor	male	1939
Hawkeye	male	1964
Daredevil	male	1964
Doctor Strange	male	1963
Hank Pym	male	1962
Scarlet Witch	female	1964
Wasp	female	1963
Black Widow	female	1964
Vision	male	1968

Drop 'Namor' and 'Hank Pym' rows

In [13]:
```python
# your code goes here
```

In [14]:
```python
marvel_df = marvel_df.drop(['Namor', 'Hank Pym'], axis=0)
marvel_df
```

name	gender	first_appearance
Spider-Man	male	1962
Captain America	male	1941
Wolverine	male	1974
Iron Man	male	1963
Thor	male	1963
Thing	male	1961
Mister Fantastic	male	1961
Hulk	male	1962
Beast	male	1963
Invisible Woman	female	1961
Storm	female	1975
Hawkeye	male	1964
Daredevil	male	1964
Doctor Strange	male	1963
Scarlet Witch	female	1964
Wasp	female	1963
Black Widow	female	1964
Vision	male	1968

DataFrame selection, slicing and indexation

Show the first 5 elements on `marvel_df`

In [15]:
```python
# your code goes here
```

In [16]:
```python
#marvel_df.loc[['Spider-Man', 'Captain America', 'Wolverine', 'Iron Man', 'Tho
#marvel_df.loc['Spider-Man': 'Thor', :]
#marvel_df.iloc[0:5, :]
#marvel_df.iloc[0:5,]
marvel_df.iloc[:5,]
#marvel_df.head()
```

	gender	first_appearance
name		
Spider-Man	male	1962
Captain America	male	1941
Wolverine	male	1974
Iron Man	male	1963
Thor	male	1963

Show the last 5 elements on `marvel_df`

In [17]:
```python
# your code goes here
```

In [18]:
```python
#marvel_df.loc[['Hank Pym', 'Scarlet Witch', 'Wasp', 'Black Widow', 'Vision'],
#marvel_df.loc['Hank Pym':'Vision', :]
marvel_df.iloc[-5:,]
#marvel_df.tail()
```

Out[18]:

	gender	first_appearance
name		
Doctor Strange	male	1963
Scarlet Witch	female	1964
Wasp	female	1963
Black Widow	female	1964
Vision	male	1968

Show just the gender of the first 5 elements on `marvel_df`

In [19]:
```python
# your code goes here
```

In [20]:
```python
#marvel_df.iloc[:5,]['gender'].to_frame()
marvel_df.iloc[:5,].gender.to_frame()
#marvel_df.head().gender.to_frame()
```

name	gender
Spider-Man	male
Captain America	male
Wolverine	male
Iron Man	male
Thor	male

Show the first_appearance of all middle elements on `marvel_df`

In [21]:
```python
# your code goes here
```

In [22]:
```python
marvel_df.iloc[1:-1,].first_appearance.to_frame()
```

	first_appearance
name	
Captain America	1941
Wolverine	1974
Iron Man	1963
Thor	1963
Thing	1961
Mister Fantastic	1961
Hulk	1962
Beast	1963
Invisible Woman	1961
Storm	1975
Hawkeye	1964
Daredevil	1964
Doctor Strange	1963
Scarlet Witch	1964
Wasp	1963
Black Widow	1964

Show the first and last elements on `marvel_df`

In [23]:
```python
# your code goes here
```

In [24]:
```python
#marvel_df.iloc[[0, -1],]][['gender', 'first_appearance']]
marvel_df.iloc[[0, -1],]
```

Out[24]:

	gender	first_appearance
name		
Spider-Man	male	1962
Vision	male	1968

DataFrame manipulation and operations

Modify the `first_appearance` of 'Vision' to year 1964

In [25]:
```python
# your code goes here
```

In [26]:
```python
marvel_df.loc['Vision', 'first_appearance'] = 1964

marvel_df
```

Out[26]:

name	gender	first_appearance
Spider-Man	male	1962
Captain America	male	1941
Wolverine	male	1974
Iron Man	male	1963
Thor	male	1963
Thing	male	1961
Mister Fantastic	male	1961
Hulk	male	1962
Beast	male	1963
Invisible Woman	female	1961
Storm	female	1975
Hawkeye	male	1964
Daredevil	male	1964
Doctor Strange	male	1963
Scarlet Witch	female	1964
Wasp	female	1963
Black Widow	female	1964
Vision	male	1964

Add a new column to `marvel_df` called 'years_since' with the years since `first_appearance`

In [27]: `# your code goes here`

In [28]:
```python
marvel_df['years_since'] = 2018 - marvel_df['first_appearance']

marvel_df
```

Out[28]:

name	gender	first_appearance	years_since
Spider-Man	male	1962	56
Captain America	male	1941	77
Wolverine	male	1974	44
Iron Man	male	1963	55
Thor	male	1963	55
Thing	male	1961	57
Mister Fantastic	male	1961	57
Hulk	male	1962	56
Beast	male	1963	55
Invisible Woman	female	1961	57
Storm	female	1975	43
Hawkeye	male	1964	54
Daredevil	male	1964	54
Doctor Strange	male	1963	55
Scarlet Witch	female	1964	54
Wasp	female	1963	55
Black Widow	female	1964	54
Vision	male	1964	54

DataFrame boolean arrays (also called masks)

Given the `marvel_df` pandas DataFrame, make a mask showing the female characters

In [29]: `# your code goes here`

In [30]:
```python
mask = marvel_df['gender'] == 'female'

mask
```

Out[30]:

name	gender
Spider-Man	False
Captain America	False
Wolverine	False
Iron Man	False
Thor	False
Thing	False
Mister Fantastic	False
Hulk	False
Beast	False
Invisible Woman	True
Storm	True
Hawkeye	False
Daredevil	False
Doctor Strange	False
Scarlet Witch	True
Wasp	True
Black Widow	True
Vision	False

dtype: bool

Given the `marvel_df` pandas DataFrame, get the male characters

In [31]: `# your code goes here`

In [32]:
```python
mask = marvel_df['gender'] == 'male'

marvel_df[mask]
```

Out[32]:

name	gender	first_appearance	years_since
Spider-Man	male	1962	56
Captain America	male	1941	77
Wolverine	male	1974	44
Iron Man	male	1963	55
Thor	male	1963	55
Thing	male	1961	57
Mister Fantastic	male	1961	57
Hulk	male	1962	56
Beast	male	1963	55
Hawkeye	male	1964	54
Daredevil	male	1964	54
Doctor Strange	male	1963	55
Vision	male	1964	54

Given the `marvel_df` pandas DataFrame, get the characters with `first_appearance` after 1970

In [33]: `# your code goes here`

In [34]:
```python
mask = marvel_df['first_appearance'] > 1970

marvel_df[mask]
```

Out[34]:

name	gender	first_appearance	years_since
Wolverine	male	1974	44
Storm	female	1975	43

Given the `marvel_df` pandas DataFrame, get the female characters with `first_appearance` after 1970

In [35]:
```python
# your code goes here
```

In [36]:
```python
mask = (marvel_df['gender'] == 'female') & (marvel_df['first_appearance'] > 19
marvel_df[mask]
```

Out[36]:

name	gender	first_appearance	years_since
Storm	female	1975	43

DataFrame summary statistics

Show basic statistics of `marvel_df`

In [37]:
```python
# your code goes here
```

In [38]:
```python
marvel_df.describe()
```

Out[38]:

	first_appearance	years_since
count	18.000000	18.000000
mean	1962.888889	55.111111
std	6.720372	6.720372
min	1941.000000	43.000000
25%	1962.000000	54.000000
50%	1963.000000	55.000000
75%	1964.000000	56.000000
max	1975.000000	77.000000

Given the `marvel_df` pandas DataFrame, show the mean value of `first_appearance`

In [39]:
```python
# your code goes here
```

In [40]:
```python
#np.mean(marvel_df.first_appearance)
marvel_df.first_appearance.mean()
```

Out[40]: np.float64(1962.888888888889)

Given the `marvel_df` pandas DataFrame, show the min value of `first_appearance`

In [41]:
```python
# your code goes here
```

In [42]:
```python
#np.min(marvel_df.first_appearance)
marvel_df.first_appearance.min()
```

Out[42]: 1941

Given the `marvel_df` pandas DataFrame, get the characters with the min value of `first_appearance`

In [43]:
```python
# your code goes here
```

```
In [44]:  mask = marvel_df['first_appearance'] == marvel_df.first_appearance.min()
          marvel_df[mask]
```

Out[44]:

name	gender	first_appearance	years_since
Captain America	male	1941	77

DataFrame basic plottings

Reset index names of `marvel_df`

```
In [45]:  # your code goes here
```

```
In [46]:  marvel_df = marvel_df.reset_index()

          marvel_df
```
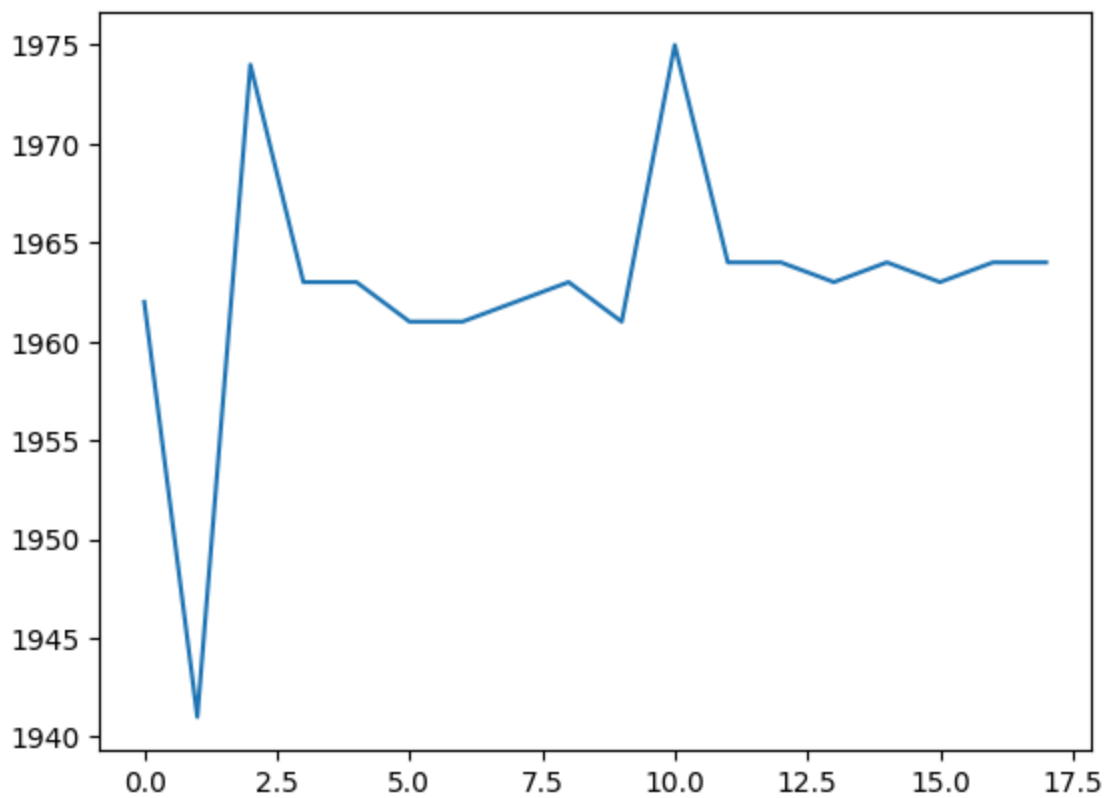
	name	gender	first_appearance	years_since
0	Spider-Man	male	1962	56
1	Captain America	male	1941	77
2	Wolverine	male	1974	44
3	Iron Man	male	1963	55
4	Thor	male	1963	55
5	Thing	male	1961	57
6	Mister Fantastic	male	1961	57
7	Hulk	male	1962	56
8	Beast	male	1963	55
9	Invisible Woman	female	1961	57
10	Storm	female	1975	43
11	Hawkeye	male	1964	54
12	Daredevil	male	1964	54
13	Doctor Strange	male	1963	55
14	Scarlet Witch	female	1964	54
15	Wasp	female	1963	55
16	Black Widow	female	1964	54
17	Vision	male	1964	54

Plot the values of `first_appearance`

In [47]:
```python
# your code goes here
```

In [48]:
```python
#plt.plot(marvel_df.index, marvel_df.first_appearance)
marvel_df.first_appearance.plot()
```

Out[48]: <Axes: >

Plot a histogram (plot.hist) with values of `first_appearance`

```
In [49]: # your code goes here
```

```
In [50]: plt.hist(marvel_df.first_appearance)
```

```
Out[50]: (array([ 1.,  0.,  0.,  0.,  0.,  3., 12.,  0.,  0.,  2.]),
 array([1941. , 1944.4, 1947.8, 1951.2, 1954.6, 1958. , 1961.4, 1964.8,
        1968.2, 1971.6, 1975. ]),
 <BarContainer object of 10 artists>)
```