Pandas DataFrame exercises

In [1]:
```
# 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__)
```

```
0.23.0
```

DataFrame creation

Create an empty pandas DataFrame ¶

In [2]:
```
# your code goes here
X = pd.DataFrame()
print(X)
```

```
Empty DataFrame
Columns: []
Index: []
```

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

In [5]:
```
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 [6]:
```
marvel_df = pd.DataFrame(marvel_data)
print(marvel_df)
```

```
                  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]:
```
marvel_df = pd.DataFrame(marvel_data, columns = ['Name', 'Gender', 'Year'])
marvel_df
```

Out[7]:

	Name	Gender	Year
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 [8]:
```
marvel_df.index = marvel_df['Name']
marvel_df
```

Out[8]:

Name	Name	Gender	Year
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 [9]:
```
# your code goes here
marvel_df = marvel_df.drop(['Name'], axis=1)
marvel_df
```

Out[9]:

Name	Gender	Year
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 [10]:
```
marvel_df = marvel_df.drop(['Namor', 'Hank Pym'], axis=0)
marvel_df
```

Out[10]:

Name	Gender	Year
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 [29]:
```
marvel_df.head()
```

Out[29]:

Name	Gender	Year
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 [30]:
```
marvel_df.tail()
```

Out[30]:

Name	Gender	Year
Doctor Strange	male	1963
Scarlet Witch	female	1964
Wasp	female	1963
Black Widow	female	1964
Vision	male	1968

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

In [11]:
```
print(marvel_df.head()['Gender'])
```

```
Name
Spider-Man         male
Captain America    male
Wolverine          male
Iron Man           male
Thor               male
Name: Gender, dtype: object
```

Show the first_appearance of all middle elements on `marvel_df`

In [12]:
```
# your code goes here
marvel_df.iloc[1:-1]['Year']
```

Out[12]:
```
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
Name: Year, dtype: int64
```

Show the first and last elements on `marvel_df`

In [13]:
```
marvel_df.iloc[[0, -1]]
```

Out[13]:

Name	Gender	Year
Spider-Man	male	1962
Vision	male	1968

DataFrame manipulation and operations

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

In [14]:
```
marvel_df.loc['Vision', 'Year'] = 1964
marvel_df
```

Out[14]:

Name	Gender	Year
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 [15]:
```
marvel_df['years_since'] = 2023 - marvel_df['Year']

marvel_df
```

Out[15]:

Name	Gender	Year	years_since
Spider-Man	male	1962	61
Captain America	male	1941	82
Wolverine	male	1974	49
Iron Man	male	1963	60
Thor	male	1963	60
Thing	male	1961	62
Mister Fantastic	male	1961	62
Hulk	male	1962	61
Beast	male	1963	60
Invisible Woman	female	1961	62
Storm	female	1975	48
Hawkeye	male	1964	59
Daredevil	male	1964	59
Doctor Strange	male	1963	60
Scarlet Witch	female	1964	59
Wasp	female	1963	60
Black Widow	female	1964	59
Vision	male	1964	59

DataFrame boolean arrays (also called masks)

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

In [16]:
```
s = marvel_df['Gender'] == 'female'
s
```

Out[16]:
```
Name
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
Name: Gender, dtype: bool
```

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

In [17]:
```
s = marvel_df['Gender'] == 'male'
marvel_df[s]
```

Out[17]:

Name	Gender	Year	years_since
Spider-Man	male	1962	61
Captain America	male	1941	82
Wolverine	male	1974	49
Iron Man	male	1963	60
Thor	male	1963	60
Thing	male	1961	62
Mister Fantastic	male	1961	62
Hulk	male	1962	61
Beast	male	1963	60
Hawkeye	male	1964	59
Daredevil	male	1964	59
Doctor Strange	male	1963	60
Vision	male	1964	59

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

In [18]:
```
s = marvel_df['Year'] > 1970
marvel_df[s]
```

Out[18]:

Name	Gender	Year	years_since
Wolverine	male	1974	49
Storm	female	1975	48

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

In [50]:
```
s = (marvel_df['Gender'] == 'female') & (marvel_df['Year'] > 1970)
marvel_df[s]
```

Out[50]:

Name	Gender	Year
Storm	female	1975

DataFrame summary statistics

Show basic statistics of `marvel_df`

In [19]:
```
marvel_df.describe()
```

Out[19]:

	Year	years_since
count	18.000000	18.000000
mean	1962.888889	60.111111
std	6.720372	6.720372
min	1941.000000	48.000000
25%	1962.000000	59.000000
50%	1963.000000	60.000000
75%	1964.000000	61.000000
max	1975.000000	82.000000

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

In [44]:
```
marvel_df.Year.mean()
```

Out[44]:
```
1962.888888888889
```

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

In [45]:
```
marvel_df.Year.min()
```

Out[45]:
```
1941
```

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

In [47]:
```
s = marvel_df['Year'] == marvel_df.Year.min()
marvel_df[s]
```

Out[47]:

Name	Gender	Year
Captain America	male	1941

Reset index names of `marvel_df`

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

Out[46]:

	Name	Gender	Year
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	Hawkeye	male	1964
12	Daredevil	male	1964
13	Doctor Strange	male	1963
14	Scarlet Witch	female	1964
15	Wasp	female	1963
16	Black Widow	female	1964
17	Vision	male	1964