Pandas DataFrame exercises

In [ ]:

```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__)
```

DataFrame creation

Create an empty pandas DataFrame

In [ ]:

```python
# your code goes here
```

In [ ]:

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

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

In [ ]:

```
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 [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df = pd.DataFrame(data=marvel_data)

marvel_df
```

Add column names to the `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

```
col_names = ['name', 'sex', 'first_appearance']

marvel_df.columns = col_names
marvel_df
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df.index = marvel_df['name']
marvel_df
```

Drop the name column as it's now the index

In [ ]:

```
# your code goes here
```

In [ ]:

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

Drop 'Namor' and 'Hank Pym' rows

In [ ]:

```
# your code goes here
```

In [ ]:

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

DataFrame selection, slicing and indexation

Show the first 5 elements on `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

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

Show the last 5 elements on `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

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

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
#marvel_df.iloc[:5,]['sex'].to_frame()
marvel_df.iloc[:5,].sex.to_frame()
#marvel_df.head().sex.to_frame()
```

Show the first_appearance of all middle elements on `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

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

Show the first and last elements on `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

```
#marvel_df.iloc[[0, -1],][['sex', 'first_appearance']]
marvel_df.iloc[[0, -1],]
```

DataFrame manipulation and operations

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df.loc['Vision', 'first_appearance'] = 1964

marvel_df
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df['years_since'] = 2018 - marvel_df['first_appearance']

marvel_df
```

DataFrame boolean arrays (also called masks)

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
mask = marvel_df['sex'] == 'female'

mask
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
mask = marvel_df['sex'] == 'male'

marvel_df[mask]
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
mask = marvel_df['first_appearance'] > 1970

marvel_df[mask]
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
mask = (marvel_df['sex'] == 'female') & (marvel_df['first_appearance'] > 1970)

marvel_df[mask]
```

DataFrame summary statistics

Show basic statistics of `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df.describe()
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
#np.mean(marvel_df.first_appearance)
marvel_df.first_appearance.mean()
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
#np.min(marvel_df.first_appearance)
marvel_df.first_appearance.min()
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

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

DataFrame basic plottings

Reset index names of `marvel_df`

In [ ]:

```
# your code goes here
```

In [ ]:

```
marvel_df = marvel_df.reset_index()

marvel_df
```

Plot the values of `first_appearance`

In [ ]:

```
# your code goes here
```

In [ ]:

```
#plt.plot(marvel_df.index, marvel_df.first_appearance)
marvel_df.first_appearance.plot()
```

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

In [ ]:

```
# your code goes here
```

In [ ]:

```
plt.hist(marvel_df.first_appearance)
```