Chapter 6 - Exercise 2: Phân tích dữ liệu thế giới qua các năm

Cho các dữ liệu gdp_cap, life_exp, pop, color từ tập tin gdp_life.csv

- gdp_cap, life_exp là thu nhập bình quân đầu người và tuổi thọ trung bình của một số quốc gia
- pop là dân số thế giới thực tế và dự đoán tương ứng với năm (year)
- color là màu các quốc gia, danh sách màu tương ứng với các nhóm quốc gia theo châu lục:
- 'Asia':'red', 'Europe':'green', 'Africa':'blue', 'Americas':'yellow', 'Oceania':'black'

Yêu cầu

Câu 1:

- Đọc tập tin gdp_life.csv và xem thông tin: shape, head, ...
- Kiểm tra NULL
- Kiểm tra duplicate

Câu 2:

- Cho biết thu nhập bình quân đầu người và tuổi thọ trung bình của item đầu (làm tròn 2 số lẻ)
- Cho biết thu nhập bình quân đầu người và tuổi thọ trung bình của item cuối (làm tròn 2 số lẻ)

Câu 3:

- Thử vẽ biểu đồ line giữa gdp_cap và life_exp với x-axis: gdp_cap, y-axis: life_exp.

Câu 4:

- Vẽ biểu đồ histogram của life_exp, màu cột xanh, viền đỏ (mặc định là 10 bins)
- Bạn nhận xét gì qua biểu đồ vừa vẽ

Câu 5:

- Vẽ biểu đồ histogram của life_exp, màu cột xanh dương, viền đỏ, với bins = 5, 15, 20

Câu 6:

- Tạo scatter plot của gdp_gap và life_exp nhưng sử dụng plt.xscale('log').
- Thang đo logarit plt.xscale('log') cho phép chúng ta hình dung các thay đổi một cách trực quan hơn.

Câu 7:

- Tạo Scatter plot của gdp_gap và life_exp, sử dụng plt.xscale('log').
- Thiết lập xlabel, ylabel, title
- Với: tick_val = [1000,10000,100000] và tick_lab = ['1k','10k','100k'] => plt.xticks(tick_val, tick_lab)

Câu 8:

- Vẽ scatter plot của gdp_cap và life_exp, với s = pop * 2, màu magenta

Câu 9:

- Vẽ scatter plot của gdp_cap và life_exp, với s = pop*2,
- Màu c = color (giá trị color trong file dữ liệu) , alpha=0.8

Câu 10:

- Vẽ scatter plot của gdp_cap, life_exp,
- với s = pop*2, màu c = color, alpha=0.8
- Thêm text cho 2 nơi là India và China:
- plt.text(1550, 71, 'India'), plt.text(5700, 80, 'China')

In [1]:
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
```

In [2]:
```
1 # Câu 1: đọc tập tin gdp_life.csv và xem thông tin: shape, head, ...
2
```

Nhấn vào đây để xem kết quả!

	gdp_cap	life_exp	pop	color
0	974.580338	43.828	31.889923	red
1	5937.029526	76.423	3.600523	green
2	6223.367465	72.301	33.333216	blue
3	4797.231267	42.731	12.420476	blue
4	12779.379640	75.320	40.301927	yellow

In [3]:
```
1 # kiểm tra Null
2
```

In [4]:
```
1 # kiểm tra duplicate
2
```

In [5]:
```
1 # Câu 2: Cho biết thu nhập bình quân đầu người và tuổi thọ trung bình
2 # của item đầu (làm tròn 2 số lẻ)
3
```

In [6]:
```
# Câu 2: Cho biết thu nhập bình quân đầu người và tuổi thọ trung bình
# của item cuối

```

In [7]:
```
# Câu 3: Thử vẽ biểu đồ line liên hệ giữa gdp_cap và life_exp
# với x-axis: gdp_cap, y-axis: life_exp.

# Biểu đồ này có thể xem được không?
# Nếu không thì bạn hãy đề xuất một loại biểu đồ phù hợp?

```

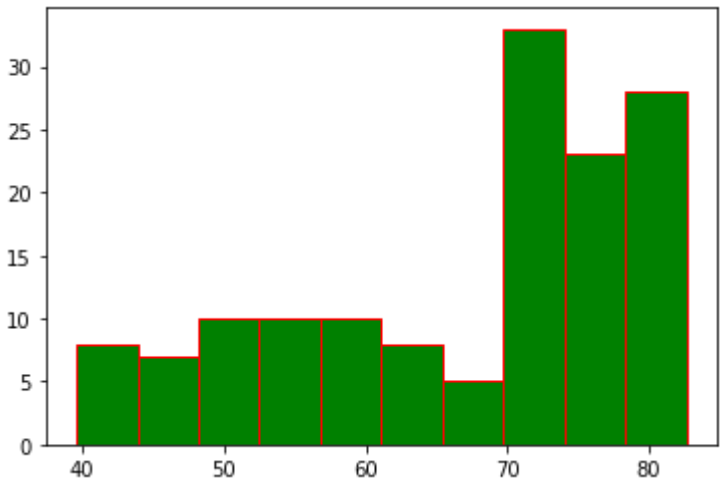
In [8]:
```
# Có thể thay biểu đồ line thành biểu đồ scatter plot

```

In [9]:
```
# Câu 4: Vẽ biểu đồ histogram của life_exp, màu cột xanh, viền đỏ (mặc định là 10 b
# Bạn nhận xét gì qua biểu đồ vừa vẽ

```

Nhấn vào đây để xem kết quả!

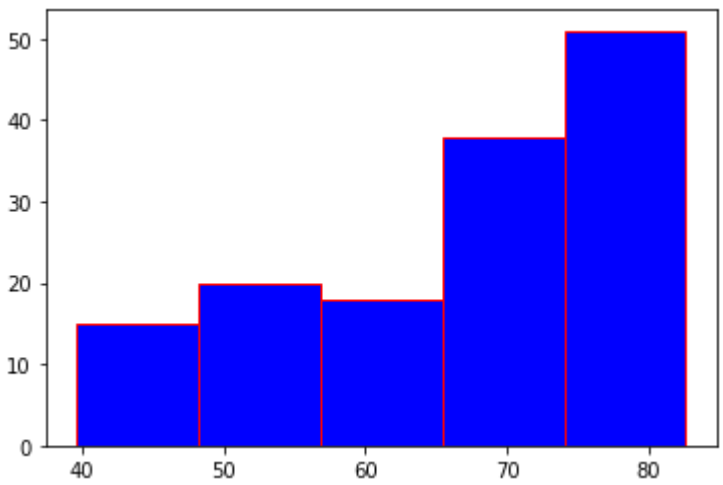


In [10]:
```
# Câu 5: Vẽ biểu đồ histogram của life_exp, màu cột xanh dương, viền đỏ,
# với bins = 5, 15, 20
# Với bins = 5

# Bạn nhận xét gì qua các biểu đồ vừa vẽ ?

```

Nhấn vào đây để xem kết quả!



In [11]:
```
# Với bins = 15

```

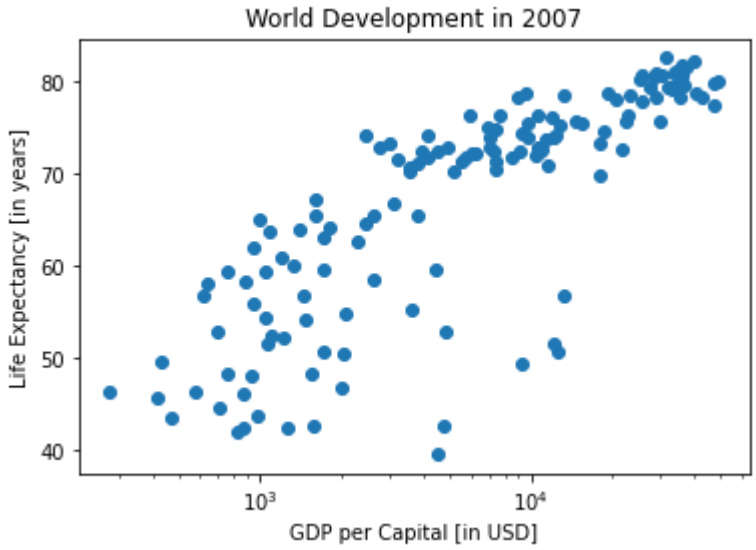
In [12]:
```
# Với bins = 20

```

In [13]:
```
# Câu 6: Tạo scatter plot của gdp_gap và life_exp nhưng sử dụng plt.xscale('log').
# Khi trực quan hóa dữ liệu thay đổi trong phạm vi rất rộng,
# thang đo logarit plt.xscale('log') cho phép
# chúng ta hình dung các thay đổi một cách trực quan hơn.

```

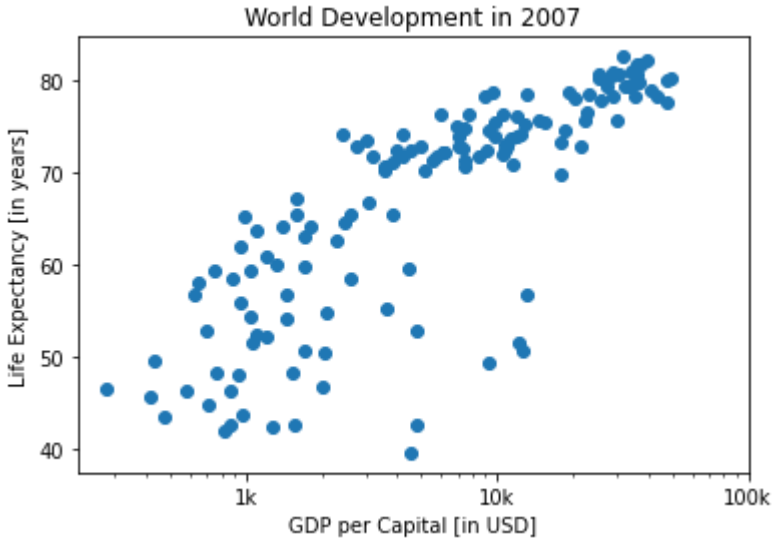
Nhấn vào đây để xem kết quả!



In [14]:
```
# Câu 7: Tạo Scatter plot của gdp_gap và life_exp, sử dụng plt.xscale('log').
# Thiết lập xlabel, ylabel, title
# Với: tick_val = [1000,10000,100000] và
# tick_lab = ['1k','10k','100k'] => plt.xticks(tick_val, tick_lab)

```

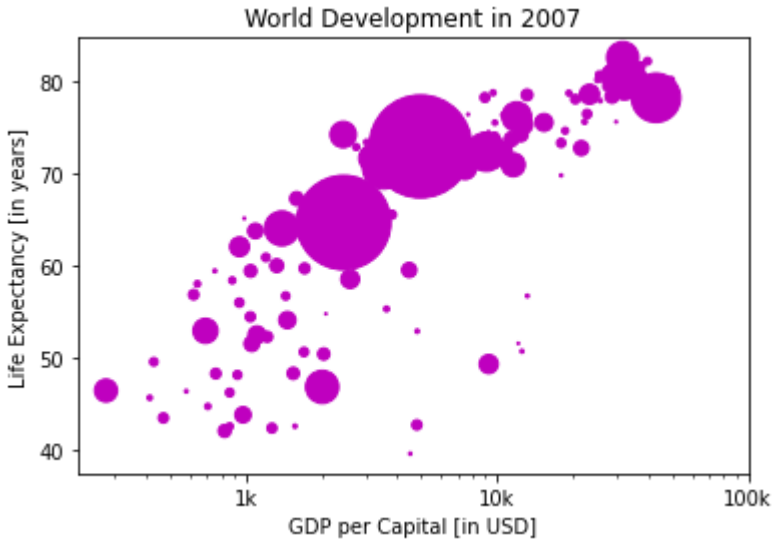
Nhấn vào đây để xem kết quả!



In [15]:
```
# Câu 8: Vẽ scatter plot của gdp_cap và life_exp, với s = pop * 2, màu magenta

```

Nhấn vào đây để xem kết quả!

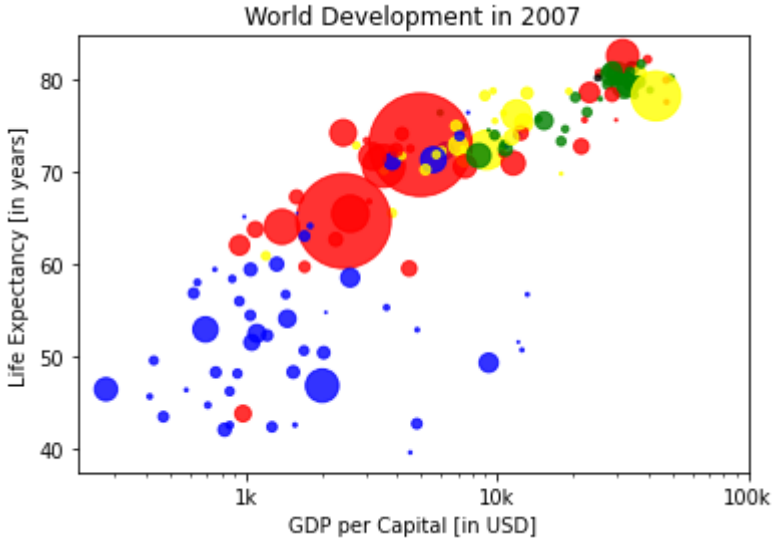


In [16]:
```
# Câu 9: Vẽ scatter plot của gdp_cap và life_exp, với s = pop*2,
# màu c = color (giá trị color trong file dữ liệu) , alpha=0.8

# Bạn nhận xét gì về biểu đồ vừa vẽ

```

Nhấn vào đây để xem kết quả!

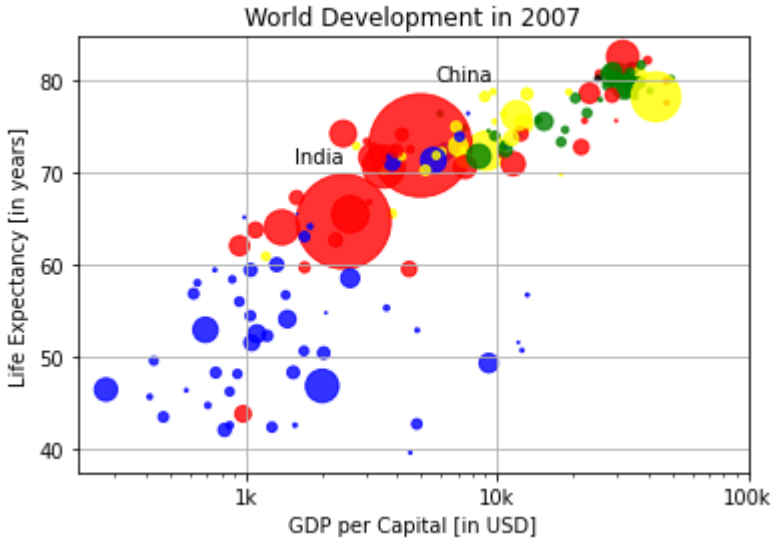


In [17]:
```
# Câu 10: Vẽ scatter plot của gdp_cap, life_exp,
# với s = pop*2, màu c = color, alpha=0.8

# Thêm text cho 2 nơi là India và China:
# plt.text(1550, 71, 'India'), plt.text(5700, 80, 'China')

```

Nhấn vào đây để xem kết quả!


World Development in 2007
China
India
Life Expectancy [in years]
80
70
60
50
40
1k
10k
100k
GDP per Capital [in USD]

In [ ]:
```
1
```