

Chapter 4

Chapter 4 - Exercise 1: Thực hiện những yêu cầu liên quan đến Series

Part 1: Thực hiện các phép toán trên series

1. a) Cho arr_1 là mảng số nguyên chẵn [2, 4, 6, 8, 10], arr_2 là mảng số nguyên lẻ [1, 3, 5, 7, 11]. Tạo biến kiểu Series ser1 từ arr_1, ser2 từ arr_2. In danh sách các phần tử của ser1 và ser2
 - b) Thực hiện phép toán và thể hiện kết quả của: $ser1 + ser2$
 - c) Thực hiện phép toán và thể hiện kết quả của: $ser1 - ser2$
 - d) Thực hiện phép toán và thể hiện kết quả của: $ser1 * ser2$
 - e) Thực hiện phép toán và thể hiện kết quả của: $ser1 / ser2$
2. a) Kiểm tra xem các phần tử của ser1 có $>$ các phần tử của ser2 không?
 - b) Kiểm tra xem các phần tử của ser1 có $<$ các phần tử của ser2 không?
 - c) Kiểm tra xem các phần tử của ser1 có $=$ các phần tử của ser2 không?
3. a) Thêm 2 phần tử [6, 12] vào ser2. In lại danh sách các phần tử của ser2.
 - b) Tạo series ser3 chỉ chứa các phần tử có trong ser1 mà không có trong ser2. In danh sách các phần tử của ser3.
 - c) Tạo series ser4 chỉ chứa các phần tử có trong ser2 mà không có trong ser1. In danh sách các phần tử của ser4.
4. Tạo series ser5 chứa các phần tử chỉ có trong ser1 và chỉ có trong ser2. In danh sách các phần tử của ser5.

Part 2: Truy xuất các phần tử, và thống kê thông tin trên series

1. a) Tạo series ser6 có 35 phần tử số nguyên ngẫu nhiên có giá trị trong khoảng từ 1 đến 9. Cho biết kích thước (shape) của ser6. Xem 5 dòng dữ liệu đầu tiên (head) và 5 dòng dữ liệu cuối cùng (tail) có trong ser6.
 - b) In danh sách các phần tử của ser6 theo dạng array.
 - c) Cho biết thông tin thống kê chung (describe()) của ser6.
 - d) Cho biết tổng của các phần tử có trong ser6.
 - e) Cho biết phần tử có tần suất xuất hiện nhiều nhất trong ser6.
2. Liệt kê các dòng trong ser6 mà giá trị chia hết cho 2 và cho 3.
3. In các phần tử ở vị trí 0, 5, 10, 15 trong ser6.
4. In ra các giá trị unique (array) trong ser6.

5. Tạo series ser7 với mỗi phần tử có giá trị = lập phương của phần tử trong ser6. Xem 5 dòng dữ liệu đầu tiên (head) của ser7.

Part 3: Tạo series từ list, chuỗi và biểu thức điều kiện

1. Cho list sau: `lst = ["abc", "defg", "htlmj", "dfg", "ljsac"]`
 - a) Tạo series ser_chuoi từ lst
 - b) Tạo series ser_dodai với mỗi phần tử có giá trị là chiều dài của mỗi phần tử trong ser_chuoi
2. Cho `ser = pd.Series(np.array([1, 2, 4, 5, 8, 7, 6, 9]))`. Sử dụng biểu thức điều kiện thích hợp để in ra các dòng trong ser có giá trị là số nguyên tố.
3. Cho mẫu email như sau: `pattern = '[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,4}'`
Tạo một series ser_ch, với mỗi phần tử trong ser_ch là một chuỗi. In ra những dòng trong ser_ch thỏa điều kiện chuỗi là email.
4. Cho series:
`ser_names = pd.Series(['Manufacturer', 'Model', 'CarType', 'Min_Price', 'Price', 'Max_Price', 'MPG_city', 'MPG_highway', 'AirBags', 'DriveTrain', 'Cylinders', 'EngineSize', 'Horsepower', 'RPM', 'Rev_per_mile', 'Man_trans_avail', 'Fuel_tank_capacity', 'Passengers', 'Length', 'Wheelbase', 'Width', 'Turn_circle', 'Rear_seat_room', 'Luggage_room', 'Weight', 'Origin', 'Make'])`
Sử dụng biểu thức điều kiện thích hợp để in ra các dòng của ser_names thỏa điều kiện trong chuỗi có chữ 'Price'.

Chapter 4 - Exercise 2: Hãy thực hiện những yêu cầu liên quan tới Data Frame

1. Cho dictionary như sau: `dic_1 = {'X':[78,85,96,80,86], 'Y':[84,94,89,83,86], 'Z':[86,97,96,72,83]}`
Tạo dataframe df1 từ dic_1. In nội dung của dataframe df1.
2. Cho dictionary exam_data như sau:
`exam_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1], 'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'no', 'no', 'yes', 'yes']}`
Cho list labels như sau: `labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']`

- a. Tạo dataframe df2 từ exam_data, với index của dataframe là labels. In nội dung của dataframe df2.
- b. Xem thông tin (info()) của dataframe df2.
3. Tạo dataframe df3 từ df2, chỉ chứa 2 cột là name và score. Xem kiểu dữ liệu (type) và kích thước (shape) của df3. Hiển thị các dòng dữ liệu đầu tiên (head) của df3.
4. Hiển thị thông tin thống kê chung (describe) của dataframe df2.
5. Tạo dataframe df4 từ df2, chỉ chứa 2 cột là name và score, và chỉ có các dòng 1, 3, 5, 6. In nội dung của dataframe df4.
6. Từ dataframe df2, in các dòng có dữ liệu score bị null.
7. Từ dataframe df2, in các dòng có score > 15 và <=20.
8. Cập nhật điểm (score) ở dòng 'd' thành 18.
9. Cho biết điểm (score) nào có tần suất xuất hiện nhiều nhất trong df2, và in ra những dòng có điểm là tần suất xuất hiện nhiều nhất. Cho biết có bao nhiêu dòng thỏa điều kiện?
10. Thêm dòng k có nội dung như sau: ['Suresh', 15.5, 1, 'yes'] vào df2. Hiển thị 5 dòng cuối cùng (tail) của df2.
11. a) Thêm dòng l có nội dung như sau: ['Janny', 12.5, 2, 'yes'] vào df2. Hiển thị 5 dòng cuối cùng (tail) của df2.
b) Xóa bỏ dòng l của df2. Hiển thị lại 5 dòng cuối cùng (tail) của df2.
12. Sắp xếp df2 tăng dần theo điểm (score).
13. Thêm cột result vào df2, dựa vào dữ liệu của cột'score', nếu dòng nào có điểm >=10 thì giá trị của cột result = 1, ngược lại = 0 . In nội dung của dataframe df2.
14. Trong df2, thay tên 'Emily' thành 'Samantha'. In lại nội dung của dataframe df2.
15. Duyệt df2, in name, score, result: nếu nếu giá trị cột result = 1 thì in 'Pass', ngược lại thì in 'Fail'.

Chapter 4 - Exercise 3: Lọc và sắp xếp dữ liệu

UEFA_European_Championship/Euro 2012

Lấy dữ liệu từ đường dẫn [Euro 12](#)

Tạo data frame euro12 từ dữ liệu trên. In euro2: type, shape, danh sách các cột

1. In giá trị cột Goals
2. Có bao nhiêu đội tham gia Euro2012?

3. In thông tin của Euro2012
4. Tạo 1 data frame mới từ euro12 có tên là discipline chỉ chứa 3 cột 'Team', 'Yellow Cards', 'Red Cards'
5. Sắp xếp discipline giảm dần theo 2 cột 'Red Cards', 'Yellow Cards'
6. a) Tính trung bình Yellow Cards
b) Lọc các đội đã ghi hơn 6 bàn thắng
7. In các đội mà tên bắt đầu bằng 'G' (gợi ý: dùng str.startswith())
8. In 7 cột đầu của euro12
9. In tất cả các cột, trừ 3 cột cuối
10. In các cột Team, Goals, Shooting Accuracy, Yellow Cards, Red Cards
11. In các cột chỉ hiển thị 'Team','Shooting Accuracy' từ 'England', 'Italy', 'Russia'