

BÀI 6: Series – DataFrame

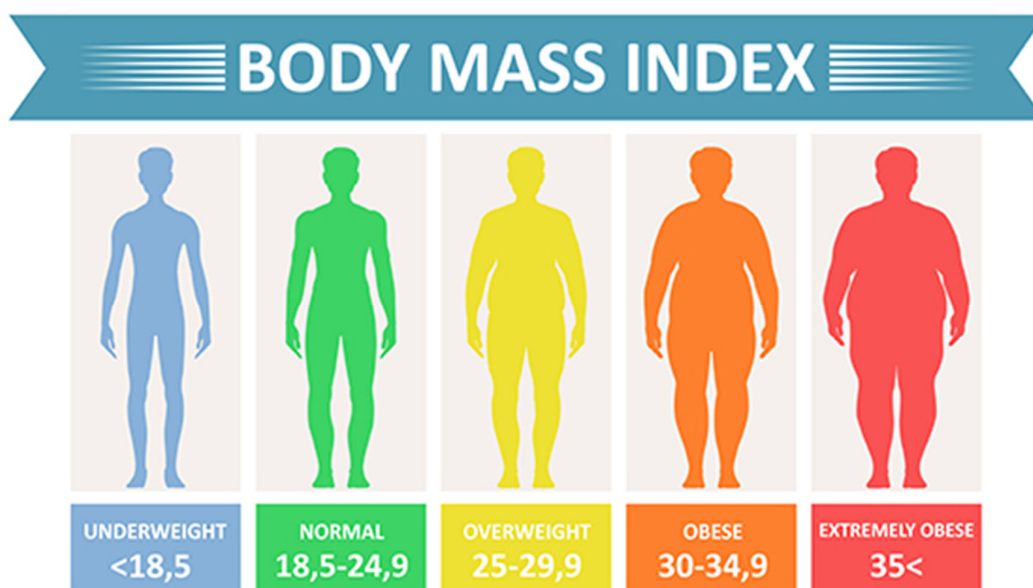


Mục tiêu chính: Cung cấp cho học viên kiến thức và kỹ năng sử dụng:

- *Series*
- *DataFrame*

6.1. Body Mass Index (BMI)

- ✓ **Dữ liệu:** Dữ liệu **heights** (tính theo inches) và **weights** (tính theo pounds) là chiều cao và cân nặng của các cầu thủ có tham gia 1 số giải của MLB (Dữ liệu **heights.txt** và **weights.txt** được đính kèm trong file **data**).



✓ **Yêu cầu:**

1. Tạo Series **ser_height** từ list height.
 - a. In danh sách các phần tử của **ser_height**
 - b. Cho biết kích thước (shape) của **ser_height**
 - c. Cho biết kiểu dữ liệu của các phần tử trong **ser_height**
2. Tạo Series **ser_weight** từ list weight.
 - a. In danh sách các phần tử của **ser_weight**
 - b. Cho biết kích thước (shape) của **ser_weight**

- c. Cho biết kiểu dữ liệu của các phần tử trong **ser_weight**
3. Cho hệ số quy đổi từ inch sang m là *0.0254*
 - a. Tạo Series **height_m** dựa trên công thức: *ser_height * hệ số quy đổi*, làm tròn đến 2 chữ số thập phân
 - b. In danh sách các phần tử của **height_m**
4. Cho hệ số quy đổi từ pound sang kg là *0.453592*
 - a. Tạo Series **weight_kg** dựa trên công thức: *ser_weight * hệ số quy đổi*
 - b. Làm tròn đến hàng đơn vị
 - c. In danh sách các phần tử của **weight_kg**
5. Tính giá trị BMI (Body Mass Index) của **height_m** và **weight_kg** theo công thức **BMI = weight_kg / (height_m)²** và lưu vào **bmi**. In ra danh sách các phần tử của Series **bmi**.
6. Cho biết giá trị cân nặng của VĐV ở vị trí index = 50 trong **weight_kg**
7. Lọc ra chiều cao của các VĐV ở vị trí index từ 100 đến 110 (lấy cả index 110) trong **height_m**
8. Tạo và in kết quả của biểu thức điều kiện dùng để lấy ra các cầu thủ bóng chày có **bmi < 21**. Áp dụng biểu thức điều kiện để in ra các cầu thủ bóng chày có bmi < 21 trong **bmi**
9. Cho biết chiều cao và cân nặng trung bình của các VĐV
10. Cho biết mức chiều cao và cân nặng có nhiều VĐV nhất
11. Cho biết chiều cao và cân nặng lớn nhất của các VĐV
 - a. Cho biết các cầu thủ có chiều cao lớn nhất
 - b. Vị trí index các cầu thủ có chiều cao lớn nhất
 - c. Cho biết có bao nhiêu VĐV cao từ 2m trở lên
12. In ra chiều cao tăng dần
 - a. In ra 5 VĐV thấp nhất
13. In ra chiều cao giảm dần
 - a. In ra 5 VĐV cao nhất
14. Thống kê chung và cân nặng và chiều cao của các VĐV

6.2. Pokemon

- ✓ **Dữ liệu:** Bộ dữ liệu Pokemon cung cấp thông tin về 800 loài Pokemon, bao gồm các thuộc tính như tên (Name), các chỉ số cơ như điểm lực chiến (HP), chỉ số tấn công (Attack), chỉ số phòng thủ (Defense), ... và nhiều thông tin khác (Dữ liệu **pokemon.csv** và **combats.csv** được đính kèm trong file **data**).



✓ Yêu cầu:

1. Đọc file pokemon và combats đã được cung cấp
 - a. In ra dataframe của **pokemon** và **combats**
 - b. Cho biết thông tin chung (info) của **pokemon** và **combats**
2. Tạo cột dữ liệu mới - Tỷ lệ chiến thắng:

$$\text{Win_Percent} = (\text{Total_Wins} / \text{Total_Fights}) * 100$$

- a. Tạo một DataFrame mới có tên **WinStatsDF** bao gồm các cột dữ liệu: *Pokedex*, *Total_Fights*, *Total_Wins*, *Win_Percent*
- b. Tạo cột dữ liệu **Total_Fights** (Tổng số trận thi đấu) bằng cách sử dụng `.groupby()` để tổng hợp tất cả trận đấu của từng Pokemon
- c. Tạo cột dữ liệu **Total_Wins** (Tổng số trận thắng) bằng cách đếm số trận thắng của từng Pokemon ở file **combats.csv** và gán vào cột dữ liệu **Total_Wins**
- d. Thực hiện kết nối 2 bảng dữ liệu **pokemon** và **WinStatsDF** thông qua phương thức `.merge()` và lưu vào biến **data**: thực hiện kết nối bằng tham

số ***left_on*** = "#", ***right_on*** = ["Pokedex"]. Sau đó xóa cột **Pokedex** khỏi dữ liệu **data**

3. Cho biết Pokemon có chỉ số lực chiến (**HP**) cao nhất
4. Cho biết Pokemon có chỉ số lực chiến (**HP**) thấp nhất
5. In ra top 10 Pokemon có chỉ số tấn công (**Attack**) cao nhất
6. In ra top 10 Pokemon có chỉ số phòng thủ (**Defense**) cao nhất
7. In ra top 10 Pokemon có chỉ số tốc độ (**Speed**) nhanh nhất
8. Truy xuất thông tin những Pokemon huyền thoại (**Legendary**)
9. Cho biết thuộc tính (**Type 1**) nào có nhiều Pokemon huyền thoại (**Legendary**)
10. In ra những Pokemon huyền thoại (**Legendary**) có tỷ lệ chiến thắng cao (> 90%)
11. Cho biết có bao nhiêu Pokemon có hai thuộc tính (hệ) (**Type 1 và Type 2**)