

BÀI THỰC HÀNH 4

THỰC HÀNH THAO TÁC DỮ LIỆU VỚI PANDAS

1. Cài đặt pandas.

Cách 1 sử dụng PIP: `pip install pandas`

Cách 2 sử dụng Anaconda: `conda install pandas`

2. Xem version của pandas đang sử dụng.

Cách 1 sử dụng PIP: `pip list`

Cách 2 sử dụng Anaconda: `conda list pandas`

3. Trong Jupyter Notebook, có thể xem version của pandas bằng lệnh sau:

```
import pandas as pd  
print(pd.__version__)
```

4. Tạo một vector với các giá trị như sau: 20, 17, 32, 19, 65

- Tạo một series từ vector ở trên.
- Thêm tên cho các dòng của series vừa tạo.
- Lấy giá trị ở dòng thứ 2.
- Lấy giá trị 3 dòng cuối của serie trên.

5. Các cặp dữ liệu có giá trị như sau:

```
111 Hong 20  
222 Tai 19  
333 Tam 23  
444 Phuong 22  
555 Huy 21
```

- Tạo ra một list để lưu trữ các dữ liệu trên.
- Tạo ra một dataframe từ list ở trên.

- Thêm tên dòng cho dataframe vừa tạo: SV1, SV2, SV3, SV4, SV5. (Gợi ý: sử dụng tham số index)
- Thêm tiêu đề cho dataframe: MSSV, HoTen, Tuoi (Gợi ý: sử dụng tham số columns)
- Hiển thị toàn bộ thông tin của sinh viên 2.
- Hiển thị tên của tất cả sinh viên.
- Tạo một dataframe mới với thông tin bao gồm: MSSV và HoTen.

6. Danh sách sinh viên có thông tin như sau:

2301	Ly Thi Hong	2000	HCM
2302	Nguyen Van Hai	2002	VT
2401	Le Phuong Anh	2001	DN

- Sử dụng một dictionary để lưu danh sách sinh viên như trên.
- Tạo một dataframe để lưu trữ danh sách viên trên (datafram được tạo ra từ dictionary).
- Đặt tên cột cho danh sách trên lần lượt là: MSSV, HoTen, NamSinh, DiaChi
- Thêm một cột dữ liệu “GioiTinh” vào danh sách trên
- Thêm dữ liệu bên dưới vào danh sách sinh viên?

2306	Nguyen Van Tai	2002	NT	Nam
2404	Le Van Nam	2001	VT	Nam
2304	Phan Hong Tham	2001	HCM	Nu

- Tạo một dataframe là danh sách các bạn có giới tính Nam từ danh sách trên?
- Lọc ra danh sách các bạn sinh viên Nữ có địa chỉ ở HCM. Cho biết có bao nhiêu bạn sinh viên trong danh sách này?
- Lưu danh sách các bạn nữ có địa chỉ ở HCM vào file .CSV?

7. Load dữ liệu từ file `sinhvien.csv`

- Cho biết số sinh viên Nam và sinh viên Nữ.
- Cho biết số sinh viên Nam và sinh viên Nữ trong từng lớp.
- Đổi tên của các thuộc tính lần lượt thành: MS, Ten, NgaySinh, GT, Diem, Lop.
- Cho biết danh sách các bạn Nữ có điểm số > 7?
- Cho biết danh sách sinh viên chưa có điểm số, sắp xếp theo MS tăng dần.
- Thêm điểm số cho các sinh viên chưa có điểm. Điểm thêm vào là giá trị nhỏ nhất của các bạn sinh viên.
- Thêm một cột dữ liệu là "XepLoai". Với kết quả là: Giỏi nếu Diem ≥ 8 , Khá nếu $8 > \text{Diem} \geq 6$, TrungBinh nếu Diem < 6.
- Cho biết tên, lớp, điểm của 2 sinh viên có điểm số cao nhất của từng lớp.
- Cho biết điểm số lớn nhất, nhỏ nhất, trung bình của các bạn sinh viên Nam.

8. Sinh viên tìm hiểu hàm và tham số sau:

- Sử dụng hàm nào để kiểm tra số lượng dữ liệu trùng?
- Hàm **drop()** và **dropna()** dùng để làm gì?
- Tham số inplace trong một số hàm của Pandas có ý nghĩa như thế nào?

Ví dụ hàm có tham số inplace: `dropna()`, `drop()`, `rename()`

Viết lệnh tạo ra dataframe có sử dụng tham số inplace và không sử dụng tham số inplace sau đó so sánh kết quả?