

Computer Science Advanced

LAB 10. TRUY XUẤT DỮ LIỆU VỚI PANDAS

Các bài tập bên dưới sử dụng dữ liệu trong file *provinces.xls* được tạo ở **Lab 9**, đọc vào *DataFrame df*.

Bài 1. Danh Sách Tỉnh Thành

Trích xuất tên các tỉnh và thành phố trong **df** để in ra màn hình theo định dạng sau:

List of provinces and cities: Thành phố Hà Nội, Tỉnh Hà Giang, Tỉnh Cao Bằng, <tên các tỉnh còn lại>

Gợi ý: Sử dụng vòng lặp **for** để lặp qua *Series* tương tự như với *list*.

Bài 2. Danh Sách Thành Phố

Tìm tên các **thành phố** và vùng tương ứng trong **df**.
Kết quả mong đợi như ảnh bên.

	Name	Region
0	Thành phố Hà Nội	Đồng bằng sông Hồng
19	Thành phố Hải Phòng	Đồng bằng sông Hồng
31	Thành phố Đà Nẵng	Nam Trung Bộ
49	Thành phố Hồ Chí Minh	Đông Nam Bộ
58	Thành phố Cần Thơ	Đồng bằng sông Cửu Long

Bài 3. Các Vùng Địa Lý

Từ **df**, hãy tìm:

- Số vùng địa lý ở Việt Nam.
- Số tỉnh thành trong mỗi vùng địa lý.

Bài 4. Dân Số

Tìm tỉnh/thành phố có dân số đông nhất và tỉnh/thành phố có dân số ít nhất.
In ra tên tỉnh/thành phố và dân số theo đơn vị người.

Bài 5. Diện Tích Vùng

Tính tổng diện tích trong các vùng địa lý, sau đó tính tỉ lệ diện tích của từng vùng theo công thức:

Tỉ lệ diện tích một vùng =
Diện tích vùng đó / Tổng diện tích các vùng

Kết quả mong đợi như ảnh bên.

	Region	Total Area (km2)	Area Proportion (%)
0	Đồng bằng sông Cửu Long	40816.4	12.332886
1	Đồng bằng sông Hồng	15081.1	4.556832
2	Đông Bắc Bộ	50802.8	15.350328
3	Nam Trung Bộ	44542.0	13.458595
4	Tây Bắc Bộ	50575.7	15.281708
5	Đông Nam Bộ	23518.7	7.106296
6	Bắc Trung Bộ	51110.8	15.443392
7	Tây Nguyên	54508.3	16.469964