

# Computer Science Advanced

## LAB 11. XỬ LÝ DỮ LIỆU TRONG PANDAS

Dữ liệu cho các bài tập bên dưới được lưu trong thư mục *VietNam Data*.

### Bài 1. Mật Độ Dân Số

Cho dữ liệu các tỉnh thành ở Việt Nam trong file *provinces.xls* tương tự như ở *Lab 9* và *Lab 10*.  
Hãy đọc dữ liệu vào *DataFrame* **df** và tính mật độ dân số của từng tỉnh thành theo công thức:

$$\text{Mật độ dân số} = \text{Tổng dân số} / \text{Tổng diện tích}$$

lấy đơn vị người / km<sup>2</sup>. Biết đơn vị của cột **Population** là nghìn người và cột **Area** là km<sup>2</sup>.

Kết quả mong đợi:

	Name	Division	Region	Area	Population	PopulationDensity
0	Thành phố Hà Nội	Thành phố Trung ương	Đồng bằng sông Hồng	3358.6	8093.9	2409.902936
1	Tỉnh Hà Giang	Tỉnh	Đông Bắc Bộ	7929.5	858.1	108.216155
2	Tỉnh Cao Bằng	Tỉnh	Đông Bắc Bộ	6700.3	530.9	79.235258
3	Tỉnh Bắc Kạn	Tỉnh	Đông Bắc Bộ	4860.0	314.4	64.691358
4	Tỉnh Tuyên Quang	Tỉnh	Đông Bắc Bộ	5867.9	786.3	134.000239
...	...	...	...	...	...	...

**Gợi ý:** Pandas hỗ trợ sử dụng các toán tử toán học trên toàn *Series*.

**Ví dụ:** `population_in_persons = df['Population'] * 1000.`

Sau khi hoàn tất tính toán, tìm tỉnh/thành có mật độ dân số cao nhất và tỉnh/thành có mật độ dân số thấp nhất.

### Bài 2. Bờ Biển

Cho dữ liệu về các tỉnh thành giáp biển trong file *coastlines.xls*.

Thực hiện các yêu cầu sau:

1. Đọc dữ liệu từ file trên vào *DataFrame*. Tìm số tỉnh thành giáp biển ở Việt Nam.
2. Nhập dữ liệu về độ dài đường bờ biển (km) của từng tỉnh/thành theo một cột mới vào *DataFrame* **df** ở *Bài 1*. Tỉnh/thành không giáp biển xem như có đường bờ biển bằng 0. Kết quả mong đợi như ảnh trên.
3. Lưu dữ liệu trong **df** vào file *provinces\_updated.xls*.

	Name	Division	Region	Area	Population	PopulationDensity	Coastline
0	Thành phố Hà Nội	Thành phố Trung ương	Đồng bằng sông Hồng	3358.6	8093.9	2409.902936	0
1	Tỉnh Hà Giang	Tỉnh	Đông Bắc Bộ	7929.5	858.1	108.216155	0
2	Tỉnh Cao Bằng	Tỉnh	Đông Bắc Bộ	6700.3	530.9	79.235258	0
3	Tỉnh Bắc Kạn	Tỉnh	Đông Bắc Bộ	4860.0	314.4	64.691358	0
4	Tỉnh Tuyên Quang	Tỉnh	Đông Bắc Bộ	5867.9	786.3	134.000239	0
...	...	...	...	...	...	...	...
58	Thành phố Cần Thơ	Thành phố Trung ương	Đồng bằng sông Cửu Long	1439.0	1236.0	858.929812	0
59	Tỉnh Hậu Giang	Tỉnh	Đồng bằng sông Cửu Long	1621.7	732.2	451.501511	0
60	Tỉnh Sóc Trăng	Tỉnh	Đồng bằng sông Cửu Long	3311.9	1199.5	362.178810	72
61	Tỉnh Bạc Liêu	Tỉnh	Đồng bằng sông Cửu Long	2669.0	908.2	340.277257	56
62	Tỉnh Cà Mau	Tỉnh	Đồng bằng sông Cửu Long	5221.2	1194.3	228.740519	254

63 rows x 7 columns

## Bài 3. Biên Giới

Cho dữ liệu về các tỉnh biên giới trong file *borders.xls*, với các sheet *CHN*, *LAO*, *KHM* tương ứng với các tỉnh có biên giới với Trung Quốc, Lào và Campuchia.

Hãy kết hợp dữ liệu trong các sheet và đưa vào *DataFrame df* với định dạng như hình bên dưới, với:

- Cột *Name* chứa tên tỉnh.
- Cột *BorderWith* chứa tên các nước có biên giới giáp tỉnh đó.
- Cột *BorderCount* chứa số nước có biên giới giáp tỉnh đó.

Sau khi hoàn tất, lưu dữ liệu trong *df* vào file *borders\_updated.xls*.

Dựa vào dữ liệu đã xử lý, hãy tìm các tỉnh có biên giới giáp với hai nước.

	Name	BorderWith	BorderCount
0	Điện Biên	China, Laos	2
1	Lai Châu	China	1
2	Lào Cai	China	1
3	Hà Giang	China	1
4	Cao Bằng	China	1
...	...	...	...