

# Computer Science Advanced

## LAB 9. GIỚI THIỆU THƯ VIỆN PANDAS

### Bài 1. Tạo DataFrame

Cho dữ liệu về các tỉnh và thành phố trực thuộc Trung ương ở Việt Nam, bao gồm các thông tin sau:

Tên	Cấp	Vùng	Diện tích (km <sup>2</sup> )	Dân số (nghìn người)
Thành phố Hà Nội	Thành phố Trung ương	Đồng bằng sông Hồng	3358.6	8093.9
Tỉnh Hà Giang	Tỉnh	Đông Bắc Bộ	7929.5	858.1
Tỉnh Cao Bằng	Tỉnh	Đông Bắc Bộ	6700.3	530.9
...	...	...	...	...

Dữ liệu hoàn chỉnh được lưu trong biến `col_list` và `province_list` đính kèm ở cuối bài.

- **`col_list`** : một *list* chứa các *string* là **danh sách tên các cột**
- **`province_list`** : một *list* chứa các *tuple* là **thông tin trên từng dòng**

**Yêu cầu:** Từ hai biến trên, hãy dùng thư viện **pandas** để khởi tạo **DataFrame** **df** chứa đầy đủ dữ liệu đã cho.

Kết quả mong đợi  
trên *Jupyter*  
*Notebook*:

df					
	Name	Division	Region	Area	Population
0	Thành phố Hà Nội	Thành phố Trung ương	Đồng bằng sông Hồng	3358.6	8093.9
1	Tỉnh Hà Giang	Tỉnh	Đông Bắc Bộ	7929.5	858.1
2	Tỉnh Cao Bằng	Tỉnh	Đông Bắc Bộ	6700.3	530.9
3	Tỉnh Bắc Kạn	Tỉnh	Đông Bắc Bộ	4860.0	314.4
4	Tỉnh Tuyên Quang	Tỉnh	Đông Bắc Bộ	5867.9	786.3
...	...	...	...	...	...
58	Thành phố Cần Thơ	Thành phố Trung ương	Đồng bằng sông Cửu Long	1439.0	1236.0
59	Tỉnh Hậu Giang	Tỉnh	Đồng bằng sông Cửu Long	1621.7	732.2
60	Tỉnh Sóc Trăng	Tỉnh	Đồng bằng sông Cửu Long	3311.9	1199.5
61	Tỉnh Bạc Liêu	Tỉnh	Đồng bằng sông Cửu Long	2669.0	908.2
62	Tỉnh Cà Mau	Tỉnh	Đồng bằng sông Cửu Long	5221.2	1194.3
63 rows × 5 columns					

**Gợi ý:** Trong hàm khởi tạo **DataFrame**, **tuple** được xử lý tương tự như **list**.

## Bài 2. Kiểm Tra DataFrame

Từ *DataFrame* **df** đã tạo, dựa vào các thuộc tính và phương thức trên *DataFrame* trong bài học để tìm:

- Số dòng và số cột trong **df**
- Kích thước lưu trữ của **df** trong bộ nhớ
- Kiểu dữ liệu của từng cột trong **df**

## Bài 3. Lưu và Đọc File

**Yêu cầu 1:** Lưu dữ liệu trong **df** đã tạo vào file *provinces.xls* theo định dạng Excel.

Ghi chú: *Pandas* mặc định lưu các giá trị *index* vào file. Để tắt tính năng này, ta truyền tham số **index=False** vào phương thức **to\_excel()** trên *DataFrame*.

**Yêu cầu 2:** Sau khi hoàn tất lưu dữ liệu, kiểm tra lại bằng cách đọc dữ liệu từ file *provinces.xls* đã lưu vào biến **province\_df**. Dữ liệu trong **province\_df** cần khớp hoàn toàn với **df**.

Kết quả mong đợi:

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

## Dữ Liệu: Các Tỉnh Thành ở Việt Nam

```
col_list = ['Name', 'Division', 'Region', 'Area', 'Population']
province_list = [
    ('Thành phố Hà Nội', 'Thành phố Trung ương', 'Đồng bằng sông Hồng', 3358.6, 8093.9),
    ('Tỉnh Hà Giang', 'Tỉnh', 'Đông Bắc Bộ', 7929.5, 858.1),
    ('Tỉnh Cao Bằng', 'Tỉnh', 'Đông Bắc Bộ', 6700.3, 530.9),
    ('Tỉnh Bắc Kạn', 'Tỉnh', 'Đông Bắc Bộ', 4860.0, 314.4),
    ('Tỉnh Tuyên Quang', 'Tỉnh', 'Đông Bắc Bộ', 5867.9, 786.3),
    ('Tỉnh Lào Cai', 'Tỉnh', 'Tây Bắc Bộ', 6364.0, 733.3),
    ('Tỉnh Điện Biên', 'Tỉnh', 'Tây Bắc Bộ', 9541.3, 601.7),
    ('Tỉnh Lai Châu', 'Tỉnh', 'Tây Bắc Bộ', 9068.8, 462.6),
    ('Tỉnh Sơn La', 'Tỉnh', 'Tây Bắc Bộ', 14123.5, 1252.7),
    ('Tỉnh Yên Bái', 'Tỉnh', 'Tây Bắc Bộ', 6887.5, 823.0),
    ('Tỉnh Hoà Bình', 'Tỉnh', 'Tây Bắc Bộ', 4590.6, 855.8),
    ('Tỉnh Thái Nguyên', 'Tỉnh', 'Đông Bắc Bộ', 3526.6, 1290.9),
    ('Tỉnh Lạng Sơn', 'Tỉnh', 'Đông Bắc Bộ', 8310.1, 782.8),
    ('Tỉnh Quảng Ninh', 'Tỉnh', 'Đông Bắc Bộ', 6178.2, 1324.8),
    ('Tỉnh Bắc Giang', 'Tỉnh', 'Đông Bắc Bộ', 3895.6, 1810.4),
    ('Tỉnh Phú Thọ', 'Tỉnh', 'Đông Bắc Bộ', 3534.6, 1466.4),
    ('Tỉnh Vĩnh Phúc', 'Tỉnh', 'Đồng bằng sông Hồng', 1235.9, 1154.8),
    ('Tỉnh Bắc Ninh', 'Tỉnh', 'Đồng bằng sông Hồng', 822.7, 1378.6),
    ('Tỉnh Hải Dương', 'Tỉnh', 'Đồng bằng sông Hồng', 1668.2, 1896.9),
    ('Thành phố Hải Phòng', 'Thành phố Trung ương', 'Đồng bằng sông Hồng', 1561.8, 2033.3),
    ('Tỉnh Hưng Yên', 'Tỉnh', 'Đồng bằng sông Hồng', 930.2, 1255.8),
    ('Tỉnh Thái Bình', 'Tỉnh', 'Đồng bằng sông Hồng', 1586.4, 1862.2),
    ('Tỉnh Hà Nam', 'Tỉnh', 'Đồng bằng sông Hồng', 861.9, 854.5),
    ('Tỉnh Nam Định', 'Tỉnh', 'Đồng bằng sông Hồng', 1668.6, 1780.9),
    ('Tỉnh Ninh Bình', 'Tỉnh', 'Đồng bằng sông Hồng', 1386.8, 984.5),
    ('Tỉnh Thanh Hóa', 'Tỉnh', 'Bắc Trung Bộ', 11114.6, 3645.8),
    ('Tỉnh Nghệ An', 'Tỉnh', 'Bắc Trung Bộ', 16481.4, 3337.2),
    ('Tỉnh Hà Tĩnh', 'Tỉnh', 'Bắc Trung Bộ', 5990.7, 1290.3),
    ('Tỉnh Quảng Bình', 'Tỉnh', 'Bắc Trung Bộ', 8000.0, 896.6),
    ('Tỉnh Quảng Trị', 'Tỉnh', 'Bắc Trung Bộ', 4621.7, 633.4),
    ('Tỉnh Thừa Thiên Huế', 'Tỉnh', 'Bắc Trung Bộ', 4902.4, 1129.5),
    ('Thành phố Đà Nẵng', 'Thành phố Trung ương', 'Nam Trung Bộ', 1284.9, 1141.1),
    ('Tỉnh Quảng Nam', 'Tỉnh', 'Nam Trung Bộ', 10574.7, 1497.5),
    ('Tỉnh Quảng Ngãi', 'Tỉnh', 'Nam Trung Bộ', 5155.8, 1231.9),
    ('Tỉnh Bình Định', 'Tỉnh', 'Nam Trung Bộ', 6066.2, 1487.8),
    ('Tỉnh Phú Yên', 'Tỉnh', 'Nam Trung Bộ', 5023.4, 873.2),
    ('Tỉnh Khánh Hòa', 'Tỉnh', 'Nam Trung Bộ', 5137.8, 1232.8),
    ('Tỉnh Ninh Thuận', 'Tỉnh', 'Nam Trung Bộ', 3355.3, 591.0),
    ('Tỉnh Bình Thuận', 'Tỉnh', 'Nam Trung Bộ', 7943.9, 1232.3),
    ('Tỉnh Kon Tum', 'Tỉnh', 'Tây Nguyên', 9674.2, 543.4),
    ('Tỉnh Gia Lai', 'Tỉnh', 'Tây Nguyên', 15511.0, 1520.2),
    ('Tỉnh Đắk Lắk', 'Tỉnh', 'Tây Nguyên', 13030.5, 1872.6),
    ('Tỉnh Đắk Nông', 'Tỉnh', 'Tây Nguyên', 6509.3, 625.8),
    ('Tỉnh Lâm Đồng', 'Tỉnh', 'Tây Nguyên', 9783.3, 1299.3),
    ('Tỉnh Bình Phước', 'Tỉnh', 'Đông Nam Bộ', 6876.8, 997.8),
    ('Tỉnh Tây Ninh', 'Tỉnh', 'Đông Nam Bộ', 4041.3, 1171.7),
    ('Tỉnh Bình Dương', 'Tỉnh', 'Đông Nam Bộ', 2694.6, 2456.3),
    ('Tỉnh Đồng Nai', 'Tỉnh', 'Đông Nam Bộ', 5863.6, 3113.7),
    ('Tỉnh Bà Rịa - Vũng Tàu', 'Tỉnh', 'Đông Nam Bộ', 1981.0, 1152.2),
    ('Thành phố Hồ Chí Minh', 'Thành phố Trung ương', 'Đông Nam Bộ', 2061.4, 9038.6),
    ('Tỉnh Long An', 'Tỉnh', 'Đồng bằng sông Cửu Long', 4494.9, 1695.1),
    ('Tỉnh Tiền Giang', 'Tỉnh', 'Đồng bằng sông Cửu Long', 2510.6, 1766.3),
    ('Tỉnh Bến Tre', 'Tỉnh', 'Đồng bằng sông Cửu Long', 2394.8, 1289.1),
    ('Tỉnh Trà Vinh', 'Tỉnh', 'Đồng bằng sông Cửu Long', 2358.3, 1009.3),
    ('Tỉnh Vĩnh Long', 'Tỉnh', 'Đồng bằng sông Cửu Long', 1525.7, 1022.6),
    ('Tỉnh Đồng Tháp', 'Tỉnh', 'Đồng bằng sông Cửu Long', 3383.8, 1598.8),
    ('Tỉnh An Giang', 'Tỉnh', 'Đồng bằng sông Cửu Long', 3536.7, 1907.4),
    ('Tỉnh Kiên Giang', 'Tỉnh', 'Đồng bằng sông Cửu Long', 6348.8, 1723.7),
    ('Thành phố Cần Thơ', 'Thành phố Trung ương', 'Đồng bằng sông Cửu Long', 1439.0, 1236.0),
    ('Tỉnh Hậu Giang', 'Tỉnh', 'Đồng bằng sông Cửu Long', 1621.7, 732.2),
    ('Tỉnh Sóc Trăng', 'Tỉnh', 'Đồng bằng sông Cửu Long', 3311.9, 1199.5),
    ('Tỉnh Bạc Liêu', 'Tỉnh', 'Đồng bằng sông Cửu Long', 2669.0, 908.2),
    ('Tỉnh Cà Mau', 'Tỉnh', 'Đồng bằng sông Cửu Long', 5221.2, 1194.3)
]
```