Mì "Python"

Mì AI Training

Bài số 06

Nội dung khóa học

- Bài 1. Python cơ bản A
- Bài 2. Python cơ bản B
- Bài 3. Python với OpenCV
- Bài 4. Python với Keras
- Bài 5. Python với Keras 2
- Bài 6. Python vơi Pandas
- Bài 7. Xây dựng Backend Server với Python

Bài 6

- Pandas là gì?
- Các khái niệm: Series, DataFrame
- Các thao tác đọc cơ bản với DataFrame
- Các thao tác sửa/xóa
- Các thao tác thống kê/sắp xếp
- Vẽ biểu đồ với Pandas

Pandas là gì?

- Thư viện pandas trong python là một thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu, phân tích và xử lý dữ liệu mạnh mẽ
- Thư viện này được sử dụng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu.

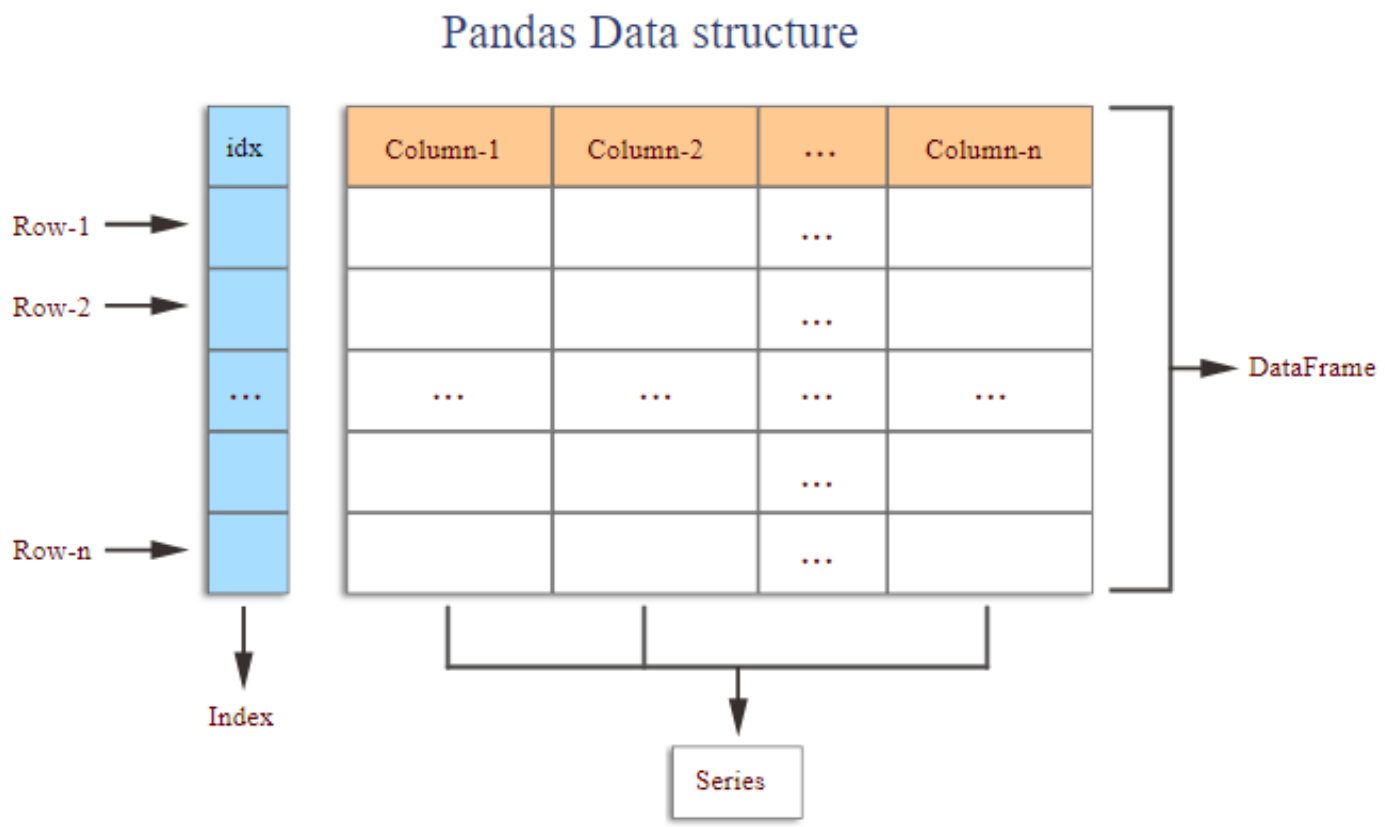
Ưu điểm của Pandas?

- Có thể xử lý tập dữ liệu khác nhau về định dạng: chuỗi thời gian, bảng không đồng nhất, ma trận dữ liệu
- Khả năng import dữ liệu từ nhiều nguồn khác nhau như CSV, DB/SQL
- Có thể xử lý vô số phép toán cho tập dữ liệu: subsetting, slicing, filtering, merging, groupBy, re-ordering, and re-shaping,..
- Xử lý dữ liệu mất mát theo ý người dùng mong muốn: bỏ qua hoặc chuyển sang 0
- Xử lý, phân tích dữ liệu tốt như mô hình hoá và thống kê
- Tích hợp tốt với các thư viện khác của python

Cài đặt Pandas

pip install pandas

Series và DataFrame



Đọc dữ liệu vào DF

Ghi dữ liệu ra CSV

Đọc dữ liệu DF

- Xem thông tin: Info
- Lấy theo cột/ nhiều cột
- Lấy theo dòng, nhiều dòng: index, tail, head, iloc, loc
- Lấy theo điều kiện

	0	1	2	3
	name	**region**	**sales**	**expenses**
0	William	East	50000	42000
1	Emma	North	52000	43000
2	Sofia	East	90000	50000
3	Markus	South	34000	44000
4	Edward	West	42000	38000
5	Thomas	West	72000	39000
6	Ethan	South	49000	42000
7	Olivia	West	55000	60000
8	Arun	West	67000	39000
9	Anika	East	65000	44000
10	Paulo	South	67000	45000

Edit dữ liệu DF

- Thêm cột
- Sửa giá trị cột
- Xóa cột/dòng

	0	1	2	3
	name	**region**	**sales**	**expenses**
0	William	East	50000	42000
1	Emma	North	52000	43000
2	Sofia	East	90000	50000
3	Markus	South	34000	44000
4	Edward	West	42000	38000
5	Thomas	West	72000	39000
6	Ethan	South	49000	42000
7	Olivia	West	55000	60000
8	Arun	West	67000	39000
9	Anika	East	65000	44000
10	Paulo	South	67000	45000

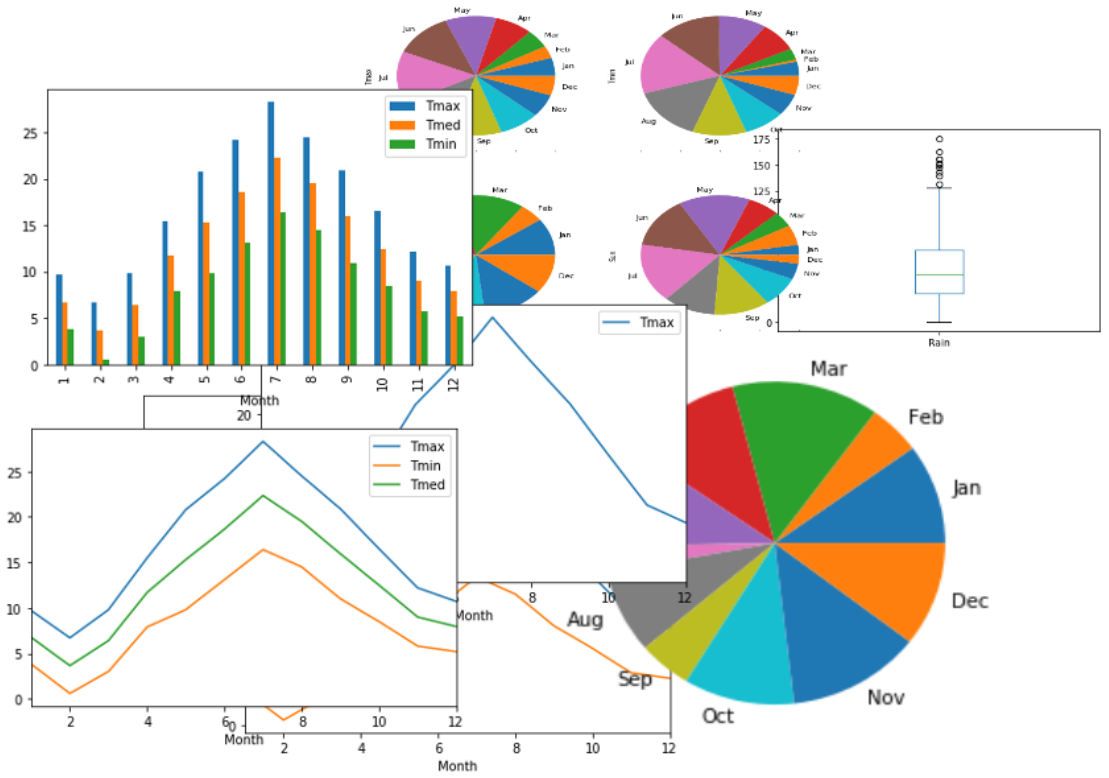
Thống kê/sắp xếp

- Describe dataframe
- Sắp xếp Dataframe
- Nối 2 DataFrame

	0	1	2	3
	name	**region**	**sales**	**expenses**
0	William	East	50000	42000
1	Emma	North	52000	43000
2	Sofia	East	90000	50000
3	Markus	South	34000	44000
4	Edward	West	42000	38000
5	Thomas	West	72000	39000
6	Ethan	South	49000	42000
7	Olivia	West	55000	60000
8	Arun	West	67000	39000
9	Anika	East	65000	44000
10	Paulo	South	67000	45000

Vẽ biểu đồ

- Biểu đồ line
- Biểu đô bar
- Biểu đồ Pie



Làm sạch dữ liệu

- Kiểm tra kiểu dữ liệu
- Kiểm tra tính đúng đắn qua describe()
- Kiểm tra NULL bằng isna()
- Xóa/fill dữ liệu thiếu: fillna, bfill, ffill

https://codelearn.io/sharing/xu-ly-missing-data-trong-python-the-nao