

## Contents

PYTHON PANDAS .....	1
Pandas Introduction.....	1
Pandas Getting Started .....	1
Pandas Series .....	1
DataFrames .....	1
Pandas Read CSV .....	2
Pandas Read JSON.....	2

## PYTHON PANDAS

### Pandas Introduction

It has functions for analyzing, cleaning, exploring, and manipulating data.

Nó có các chức năng phân tích, làm sạch, khám phá và thao tác dữ liệu.

### Pandas Getting Started

#Import Pandas

```
import pandas
```

#thường được nhập dưới bí danh pd

```
import pandas as pd
```

```
mydataset = {
```

```
    'cars': ["BMW", "Volvo", "Ford"],
```

```
    'passings': [3, 7, 2]
```

```
}
```

```
myvar = pd.DataFrame(mydataset)
```

```
print(myvar)
```

#Kiểm tra phiên bản

```
print(pd.__version__)
```

```
cars passings
```

```
0 BMW 3
```

```
1 Volvo 7
```

```
2 Ford 2
```

```
0.24.2
```

### Pandas Series

Chuỗi gấu trúc giống như một cột trong bảng.

Nó là một mảng một chiều chứa dữ liệu thuộc bất kỳ kiểu nào.

```
import pandas as pd
```

```
a = [1, 7, 2]
```

```
myvar = pd.Series(a)
```

```
print(myvar)
```

#tạo nhãn : dùng đối số index

```
myvar = pd.Series(a, index = ["x", "y", "z"])
```

```
print(myvar)
```

```
print(myvar[0])#truy cập giá trị
```

```
print(myvar["x"])
```

#có thể dùng key như từ điển khi tạo series

```
calories = {"day1": 420, "day2": 380, "day3": 390}
```

```
myvar = pd.Series(calories)
```

```
print(myvar)
```

```
print(myvar[0])
```

```
print(myvar["day1"])
```

```
0 1
```

```
1 7
```

```
2 2
```

```
dtype: int64
```

```
x 1
```

```
y 7
```

```
z 2
```

```
dtype: int64
```

```
1
```

```
1
```

```
day1 420
```

```
day2 380
```

```
day3 390
```

```
dtype: int64
```

```
420
```

```
420
```

### DataFrames

Pandas DataFrame là một cấu trúc dữ liệu 2 chiều, giống như một mảng 2 chiều hoặc một bảng có các hàng và cột.

Series giống như một cột, DataFrame là toàn bộ bảng.

```
import pandas as pd
```

```
data = {
```

```
    "calories": [420, 380, 390],
```

```
    "duration": [50, 40, 45]
```

```
}
```

```
df = pd.DataFrame(data)
```

```
print(df)
```

#định vị hàng : sử dụng df.loc[]

```
print(df.loc[0])
```

print(df.loc[[0, 1]])#Trả lại hàng 0 và 1:

#tạo nhãn : dùng đối số index

```
df = pd.DataFrame(data, index = ["day1", "day2",
```

```
"day3"])
```

```
print(df)
```

```
print(df.loc["day1"])
```

#Tải tệp vào DataFrame

```
#df = pd.read_csv('data.csv')
```

```
#print(df)
```

```
calories duration
```

```
0 420 50
```

```
1 380 40
```

```
2 390 45
```

```
calories 420
```

```

duration    50
Name: 0, dtype: int64
   calories duration
0    420         50
1    380         40
   calories duration
day1    420         50
day2    380         40
day3    390         45
calories    420
duration    50
Name: day1, dtype: int64

```

### Pandas Read CSV

Một cách đơn giản để lưu trữ tập dữ liệu lớn là sử dụng tệp CSV (tệp được phân tách bằng dấu phẩy).

```

import pandas as pd
df = pd.read_csv('data.csv')
#Mẹo: sử dụng to_string() để in toàn bộ DataFrame.
print(df.to_string())
#In một mẫu giảm:<sẽ có ... ..
print(df)
# Theo mặc định, khi bạn in DataFrame, bạn sẽ chỉ
nhận được 5 hàng đầu tiên và 5 hàng cuối cùng:

```

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0
5	60	102	127	300.5
6	60	110	136	374.0
7	45	104	134	253.3
8	30	109	133	195.1
9	60	98	124	269.0
10	60	103	147	329.3

  

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0
..	...	...	...	...
6	60	110	136	374.0
7	45	104	134	253.3
8	30	109	133	195.1
9	60	98	124	269.0
10	60	103	147	329.3

[11 rows x 11 columns]

### Pandas Read JSON

Tập hợp dữ liệu lớn thường được lưu trữ hoặc trích xuất dưới dạng JSON.( văn bản thuần túy, nhưng có định dạng của một đối tượng)

#Tải tệp JSON vào DataFrame:

```

import pandas as pd
df = pd.read_json('data.json')
#Mẹo: sử dụng to_string() để in toàn bộ DataFrame.
print(df.to_string())

```

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0
5	60	102	127	300.5

#Tư điển dưới dạng JSON

```

import pandas as pd

```

```

data = {
    "Duration":{
        "0":60,"1":60,
        "2":60,"3":45,
        "4":45,"5":60
    },
    "Pulse":{
        "0":110,"1":117,
        "2":103,"3":109,
        "4":117,"5":102
    },
    "Maxpulse":{
        "0":130,"1":145,
        "2":135,"3":175,
        "4":148,"5":127
    },
    "Calories":{
        "0":409.1,"1":479.0,
        "2":340.0,"3":282.4,
        "4":406.0,"5":300.5
    }
}
df = pd.DataFrame(data)
print(df)

```

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0
5	60	102	127	300.5

