

# Intro to Pandas

Tags	Pandas	Python
Class		
Finished Yet?	<input checked="" type="checkbox"/>	
Knowledge	 <a href="#">The Ninth Sprint: Essential Python for Data Science</a>	

\*Sprint นี้มีเนื้อหาต่อจาก Sprint 8 (Intro to Python) และเนื้อหาทุกอย่างใน Sprint นี้กับ Sprint ก่อนหน้าเชื่อมต่อกัน จึงขอนับ Lesson ของ Sprint นี้ เริ่มตั้งแต่ 11 เป็นต้นไป

-pdf:

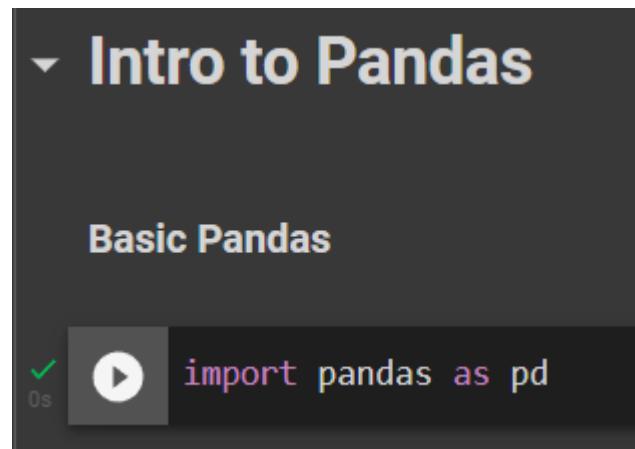
[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/cd03533c-4193-49be-8523-70d46b702cf0/Pandas\\_Foundation\\_Slide.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/cd03533c-4193-49be-8523-70d46b702cf0/Pandas_Foundation_Slide.pdf)

-Pandas documentation: [https://pandas.pydata.org/docs/getting\\_started/index.html](https://pandas.pydata.org/docs/getting_started/index.html)

## Lesson 11: Intro to Pandas

-Pandas ใช้ในการจัดการกับ Data Frame

-Pandas สามารถสร้าง Data Frame ได้ด้วย pd.DataFrame



---

## Lesson 12: Create Data Frame

-เราสามารถสร้าง Data Frame ได้ด้วย pd.DataFrame

```
#Create DataFrame from scratch by creating a dictionary

raw_data = {
    "name" : ["Cyno", "Nilou", "Candace", "Kaveh", "Amogus"],
    "age" : [33, 25, 20, 22, 31],
    "gender" : ["M", "F", "F", "M", "F" ]
}

df = pd.DataFrame(raw_data)

df
```

The screenshot shows the output of the code in Google Colab. It displays a Pandas DataFrame with columns 'name', 'age', and 'gender'. The data is as follows:

	name	age	gender
0	Cyno	33	M
1	Nilou	25	F
2	Candace	20	F
3	Kaveh	22	M
4	Amogus	31	F

\*ใน Google Colab เราสามารถเปลี่ยน Data Frame ที่เราสร้างเป็น Interactive Table ได้ด้วยการกดคลิกไอคอน Magic Wand ที่อยู่ข้าง ๆ Data Frame ทำให้เราสามารถกรองข้อมูลในตารางด้วยค่าต่าง ๆ ที่เป็น Key ในตารางได้

The screenshot shows an interactive table interface. At the top, there are search fields for 'index', 'name', 'age', and 'gender', along with a 'Search by all fields:' input field. Below the search fields is a table with the same data as the previous screenshot:

index	name	age	gender
0	Cyno	33	M
1	Nilou	25	F
2	Candace	20	F
3	Kaveh	22	M
4	Amogus	31	F

At the bottom left, there are buttons for 'Show 25' and 'per page'. A note at the bottom says 'Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.'

## Lesson 13: Add a new column

-เราสามารถเพิ่ม column ใหม่ได้ด้วยการเพิ่บค่า Key และ Value

-Function DataFrame.shape สามารถใช้เช็คจำนวน row และ column ของ Data Frame ได้

```
[13] #Adding a new column  
df['city'] = ['Sumeru', 'Sumeru', 'Sumeru', 'Inazuma', 'Mondstadt']  
  
df
```

	name	age	gender	city
0	Cyno	33	M	Sumeru
1	Nilou	25	F	Sumeru
2	Candace	20	F	Sumeru
3	Kaveh	22	M	Inazuma
4	Amogus	31	F	Mondstadt

```
df.shape  
(5, 4)
```

## Lesson 14: Drop column

- การใช้งาน Function DataFrame.drop จะต้องกำหนด Labels และ Axis ในการลบ column ที่ต้องการ
- เราสามารถใช้ drop ใน การลบ row ออกได้เช่นกัน
- เราสามารถใช้ DataFrame.reset\_index ในการรีเซ็ตค่า index ของ Data Frame เพื่อให้เลข index กลับมาเรียงเหมือนเดิม

```
[25] #Dropping a column  
df = df.drop("city", axis = 1)  
  
df
```

	name	age	gender
0	Cyno	33	M
1	Nilou	25	F
2	Candace	20	F
3	Kaveh	22	M
4	Amogus	31	F

---

```
[26] #Remove index = 2  
df = df.drop(2, axis = 0)  
  
df
```

	name	age	gender
0	Cyno	33	M
1	Nilou	25	F
3	Kaveh	22	M
4	Amogus	31	F

The screenshot shows a Jupyter Notebook cell with the following code:

```
#Reset index
df = df.reset_index(drop = True)
```

Below the code, the variable `df` is displayed as a DataFrame:

	name	age	gender
0	Cyno	33	M
1	Nilou	25	F
2	Kaveh	22	M
3	Amogus	31	F

---

## Lesson 15: Rename columns

- เราสามารถเรียกคุณสมบัติ Attribute ของ Data Frame ได้ด้วย `DataFrame.columns`
- เราจะสร้าง List เพื่อเก็บค่าของ `DataFrame.columns` และพิมพ์ค่าใหม่กับค่าเดิม

```
✓ [28] #Rename columns  
  
#Column names  
df.columns  
  
Index(['name', 'age', 'gender'], dtype='object')  
  
✓ [29] list(df.columns)  
  
['name', 'age', 'gender']  
  
✓ 0s df.columns = ['nickname', 'age', 'sex']  
  
df  
  


|   | nickname | age | sex |
|---|----------|-----|-----|
| 0 | Cyno     | 33  | M   |
| 1 | Nilou    | 25  | F   |
| 2 | Kaveh    | 22  | M   |
| 3 | Amogus   | 31  | F   |


```

## Lesson 16: Series VS Data Frame

- แต่ละ column ใน Data Frame จะถูกมองว่าเป็น Series
- เราสามารถสร้าง Series ใหม่ได้ด้วย pd.Series
- เราสามารถใช้ DataFrame.append ในการเพิ่ม Series เข้าสู่ Data Frame
- เราสามารถใช้หลักการของ Series เพิ่ม column เข้าสู่ Data Frame ได้เช่นกัน

```
✓ [33] #Series VS DataFrame
0s type(df['nickname'])

pandas.core.series.Series

✓ [34] type(df)
0s

pandas.core.frame.DataFrame

✓ [36] #Create a new Series
0s s1 = pd.Series(["Candace", 20, "F"], index = ["nickname", "age", "sex"])
print(s1)
print(type(s1))

nickname    Candace
age          20
sex           F
dtype: object
<class 'pandas.core.series.Series'>
```

```
[37] #Append Series to Data Frame
df = df.append(s1, ignore_index = True)

df
```

	nickname	age	sex
0	Cyno	33	M
1	Nilou	25	F
2	Kaveh	22	M
3	Amogus	31	F
4	Candace	20	F

```
s2 = pd.Series(['Sumeru', 'Sumeru', 'Sumeru', 'Inazuma', 'Mondstadt'])
df["city"] = s2

df
```

	nickname	age	sex	city
0	Cyno	33	M	Sumeru
1	Nilou	25	F	Sumeru
2	Kaveh	22	M	Sumeru
3	Amogus	31	F	Inazuma
4	Candace	20	F	Mondstadt

## Lesson 17: Write CSV File

- เราสามารถเชฟไฟล์ Data Frame ที่เราทำใน Python (Pandas) ได้ด้วย  
DataFrame.to\_csv("ชื่อไฟล์ที่เราอยากตั้งให้.csv")

- เราสามารถเช็คไฟล์ที่ถูกเขียนขึ้นมาได้ที่ Iocon Files และเราสามารถ Download ไฟล์นำมายังคอมของเราต่อได้

The screenshot shows a Jupyter Notebook interface. On the left, the 'Files' sidebar displays a folder structure with 'sample\_data' and 'my\_data.csv'. The main area shows a table with columns: nickname, age, sex, and city. The data consists of five rows:

	nickname	age	sex	city
0	Cyno	33	M	Sumeru
1	Nilou	25	F	Sumeru
2	Kaveh	22	M	Sumeru
3	Amogus	31	F	Inazuma
4	Candace	20	F	Mondstadt

Below the table, a code cell contains the command: `[39] #Write .csv file  
df.to_csv("my_data.csv")`.

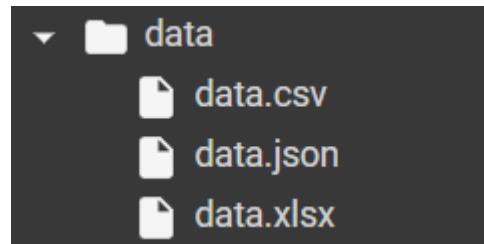
-Ի՞նչ Փա՞ն Excel

The screenshot shows a Microsoft Excel spreadsheet titled 'my\_data'. The data is identical to the one shown in the Jupyter Notebook:

	nickname	age	sex	city
0	Cyno	33	M	Sumeru
1	Nilou	25	F	Sumeru
2	Kaveh	22	M	Sumeru
3	Amogus	31	F	Inazuma
4	Candace	20	F	Mondstadt

## Lesson 18: Upload data to Cloud

-ใน Google Colab ให้คลิกขวา แล้วเลือก New Folder จากนั้น คลิก Upload เพื่ออัปโหลดไฟล์ที่เราต้องการขึ้น Google Colab



-เราสามารถนำไฟล์ที่เรา Upload แล้วมาจัดการใน Google Colab ได้ด้วยการ Read ไฟล์นั้นด้วย pd.read\_csv

The screenshot shows a Jupyter Notebook cell with the following code:

```
[39] #Write .csv file  
df.to_csv("my_data.csv")  
  
#Import .csv file  
df2 = pd.read_csv("data/data.csv")  
  
df2
```

Below the code, the resulting DataFrame is displayed:

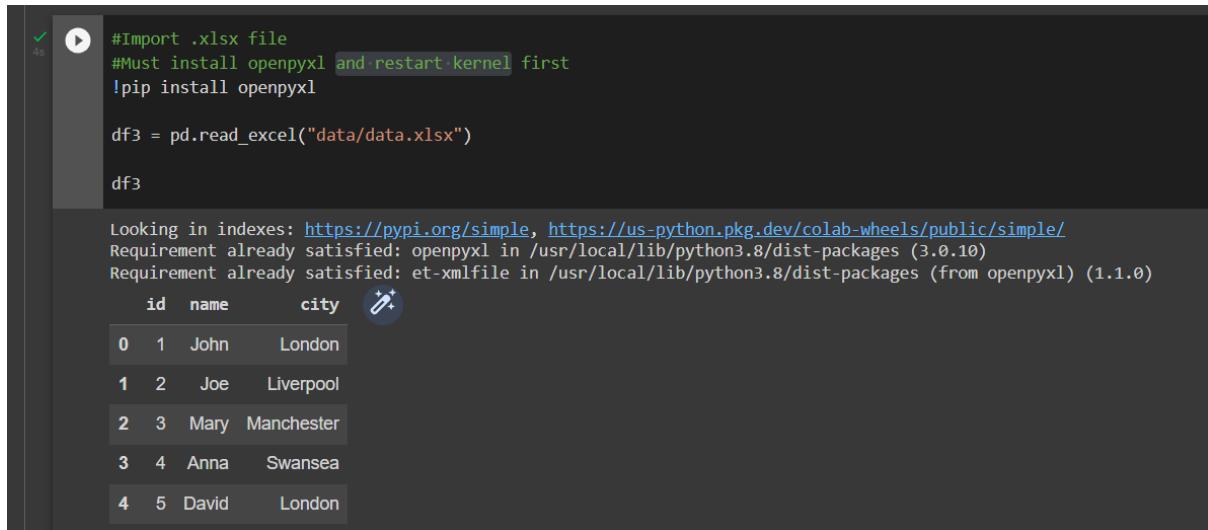
	<b>id</b>	<b>name</b>	<b>city</b>
<b>0</b>	1	John	London
<b>1</b>	2	Joe	Liverpool
<b>2</b>	3	Mary	Manchester
<b>3</b>	4	Anna	Swansea
<b>4</b>	5	David	London

\*ใน DataLore ก็ทำคล้าย ๆ กัน

## Lesson 19: Read Excel Data (.xlsx)

- เราจะใช้ pd.read\_excel ในการอ่านไฟล์ .xlsx (ไฟล์ Excel)

\* จำเป็นต้อง Download library ชื่อ openpyxl มาก่อน ให้โหลดด้วย !pip install openpyxl และ Restart Kernel หลังจากนั้น



The screenshot shows a Jupyter Notebook cell with the following code:

```
#Import .xlsx file
#Must install openpyxl and restart kernel first
!pip install openpyxl

df3 = pd.read_excel("data/data.xlsx")

df3
```

Output:

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: openpyxl in /usr/local/lib/python3.8/dist-packages (3.0.10)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.8/dist-packages (from openpyxl) (1.1.0)
```

	<code>id</code>	<code>name</code>	<code>city</code>
0	1	John	London
1	2	Joe	Liverpool
2	3	Mary	Manchester
3	4	Anna	Swansea
4	5	David	London

## Lesson 20: JSON (JavaScript Object Notation)

- JSON เป็น Format ของไฟล์ที่นิยมใช้กันในหมู่ API Developer มีความคล้ายคลึงกับ Dictionary ใน Python เก็บค่าด้วย Key : Value เมื่อบอกกัน

- ใน DataLore เราสามารถแก้ไขไฟล์ JSON ที่เรา Upload ขึ้น Cloud ก่อน Read ไฟล์ได้

- ค่าความจริง (Boolean) ในไฟล์ JSON จะเป็นตัวพิมพ์เล็กกึ่งหนด (true) แต่ค่าความจริงของ Dictionary จะมีตัวพิมพ์ใหญ่ข้างหน้า (True)

```
[45] #Import .json file  
df4 = pd.read_json("data/data.json")  
  
df4
```

	ebook	language	amazonRating	edit
0	Getting started with Python	python	4.89	
1	Introduction to R	r	4.88	
2	SQL for Beginners	sql	4.75	

```
df4[ "ebook" ].dtype  
  
dtype('O')
```

My file:

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/e5740317-3e35-46b4-b245-c4aa379190e3/Intro\\_to\\_Pandas\\_Poorin.ipynb](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/e5740317-3e35-46b4-b245-c4aa379190e3/Intro_to_Pandas_Poorin.ipynb)