

6630250435\_HW1\_870.ipynb

+ Code

+ Markdown

Run All

Restart

Clear All Outputs

Variables

Outline

Python 3.12.3

WorkShop (Download Data from URL)

import requests

download\_url = "https://media.geeksforgeeks.org/wp-content/uploads/employees.csv"

target\_csv\_path = "all\_testdata.csv"

response = requests.get(download\_url)

response.raise\_for\_status() # Check that the request was successful

with open(target\_csv\_path, "wb") as f:

f.write(response.content)

print("Download ready.")

[1] ✓ 1.9s Python

Download ready.

ทำการอ่านไฟล์ที่ Download

from pandas import DataFrame, read\_csv

employee = read\_csv("all\_testdata.csv")

[2]

6630250435\_HW1\_870.ipynb

+ Code

+ Markdown

Run All

Restart

Clear All Outputs

Variables

Outline

Python 3.12.3

ทำการอ่านไฟล์ที่ Download

from pandas import DataFrame, read\_csv

employee = read\_csv("all\_testdata.csv")

len(employee)

employee.shape

[2] ✓ 3.1s Python

(1000, 8)

ทำการแสดงข้อมูลทั้งหมด

# เรียกใช้ function display โดยส่ง object employee เข้าไปเป็น argument

display(employee)

[3] ✓ 0.0s Python

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance

6630250435\_HW1\_870.ipynb X

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.3

ทำการแสดงข้อมูลทั้งหมด

▶

# เรียกใช้ function display โดยส่ง object employee เข้าไปเป็น argument  
display(employee)

[3] ✓ 0.0s Python

...

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
...	...	...	...	...	...	...	...	...
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

1000 rows × 8 columns

6630250435\_HW1\_870.ipynb X

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.3

ทำการสำรวจข้อมูลที่ download มา ด้วยการ count จำนวน

▶

# เรียกใช้ method count ในการแสดงผลการนับของข้อมูลแต่ละคอลัมน์ โดยจะไม่นับแถวของข้อมูลที่มีค่าเป็น NaN  
print(employee.count())

[4] ✓ 0.0s Python

...

First Name 933  
Gender 855  
Start Date 1000  
Last Login Time 1000  
Salary 1000  
Bonus % 1000  
Senior Management 933  
Team 957  
dtype: int64

ทำการสร้าง DataFrame

[5] ✓ 0.0s Python

# สร้าง object ของ dataframe  
df = DataFrame(employee)

6630250435\_HW1\_870.ipynb X

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.3

ดึงข้อมูลชื่อ และเงินเดือน

# เรียกใช้ function display โดยส่ง object แล้วเราเลือกแสดงผลคอลัมน์เฉพาะ ชื่อ และ เงินเดือน ก็จะได้ข้อมูล

# ทุกแถวของคอลัมน์ First Name และ Salary

display(df[["First Name", "Salary"]])

[6] ✓ 0.0s Python

...

	First Name	Salary
0	Douglas	97308
1	Thomas	61933
2	Maria	130590
3	Jerry	138705
4	Larry	101004
...	...	...
995	Henry	132483
996	Phillip	42392
997	Russell	96914
998	Larry	60500
999	Albert	129949

1000 rows × 2 columns

6630250435\_HW1\_870.ipynb X

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.3

ดึงข้อมูลของพนักงานคนที่ 5 ถึงคนที่ 10

# เอาข้อมูลของตารางตั้งแต่แถวที่ 5 ถึงแถวที่ 10 โดยระบุเลข index ตามนี้คือ เริ่มที่แถวที่ 5 และจบที่แถวที่ 10

# แต่ต้องใส่เป็นเลข 11 เพราะต้องใส่ +1 เสมอนั้นจะจบจนถึงเลขสุดท้าย

employees = df[5:11]

display(employees)

[7] ✓ 0.0s Python

...

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal
6	Ruby	Female	8/17/1987	4:20 PM	65476	10.012	True	Product
7	NaN	Female	7/20/2015	10:43 AM	45906	11.598	NaN	Finance
8	Angela	Female	11/22/2005	6:29 AM	95570	18.523	True	Engineering
9	Frances	Female	8/8/2002	6:51 AM	139852	7.524	True	Business Development
10	Louise	Female	8/12/1980	9:01 AM	63241	15.132	True	NaN

ข้อมูลของพนักงานที่ได้เงินเดือนเกิน 10000

# เขียน condition ว่าให้เช็คคอลัมน์ Salary ว่าค่าเงินเดือนมีค่าเกิน 10000 ให้เอาข้อมูลแถวที่มีค่าเงินเดือนมีค่าน้อยกว่า 10000 จะไม่เอา

# แล้วก็จะได้อาจารย์ข้อมูลที่ได้เงินเดือนเกิน 10000

[8] ✓ 0.0s Python

ข้อมูลของพนักงานที่ได้เงินเดือนเกิน 10000

```
# เขียน condition ว่าให้เช็คคอลัมน์ Salary ว่าค่าเงินเดือนมีค่าเกิน 10000 ให้เอาข้อมูลแถวที่ค่าเงินเดือนมีค่าน้อยกว่า 10000 จะไม่เอามา
# แล้วก็จะได้อาไรข้อมูลที่ได้เงินเดือนเกิน 10000
employees2 = df[df["Salary"] > 10000]
display(employees2)
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
...	...	...	...	...	...	...	...	...
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

1000 rows x 8 columns

ดึงข้อมูลของพนักงานที่อยู่ใน ทีม Finance กับ Sales

```
# ตรง condition ใช้ method isin เพื่อเช็คค่าของข้อมูลในคอลัมน์ Team ว่าค่ามีค่าเป็น Finance หรือ Sales
# ก็จะเลือกข้อมูลของแถวที่มาแล้วก็จะได้อาไรข้อมูลที่มี Team เป็น Finance กับ Sales
employees3 = df[df["Team"].isin(["Finance", "Sales"])]
display(employees3)
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
7	NaN	Female	7/20/2015	10:43 AM	45906	11.598	NaN	Finance
13	Gary	Male	1/27/2008	11:40 PM	109831	5.831	False	Sales
14	Kimberly	Female	1/14/1999	7:13 AM	41426	14.543	True	Finance
...	...	...	...	...	...	...	...	...
975	Susan	Female	4/7/1995	10:05 PM	92436	12.467	False	Sales
987	Gloria	Female	12/8/2014	5:08 AM	136709	10.331	True	Finance
992	Anthony	Male	10/16/2011	8:35 AM	112769	11.625	True	Finance
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

6630250435\_HW1\_870.ipynb

Code

Markdown

Run All

Restart

Clear All Outputs

Variables

Outline

Python 3.1

ดึงข้อมูลชื่อของคนที่ได้เงินเดือนมากกว่า 100000

# เขียน condition ให้เช็คคอลัมน์ Salary ว่าถ้าเงินเดือนมีค่ามากกว่า 100000 ให้เอาข้อมูลแถวนั้นมาถ้าเงินเดือนมีค่าน้อยกว่า 100000 จะไม่เอามา

# แล้วก็จะได้อาไรข้อมูลที่ได้เงินเดือนมากกว่า 100000

employees4 = df[df["Salary"] >= 100\_000]

display(employees4)

[10] ✓ 0.0s

Python 3.1

...

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal
9	Frances	Female	8/8/2002	6:51 AM	139852	7.524	True	Business Development
...	...	...	...	...	...	...	...	...
990	Robin	Female	7/24/1987	1:35 PM	100765	10.982	True	Client Services
991	Rose	Female	8/25/2002	5:12 AM	134505	11.051	True	Marketing
992	Anthony	Male	10/16/2011	8:35 AM	112769	11.625	True	Finance
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

409 rows x 8 columns