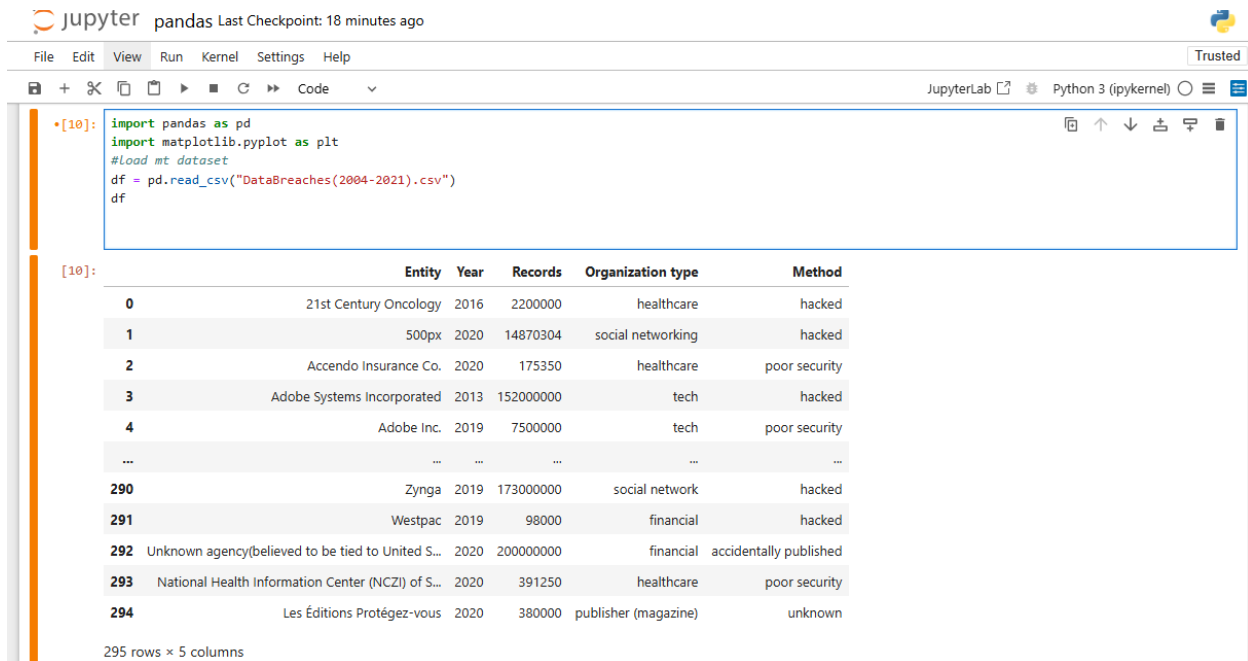


## Pandas Assignment

Lojain Idris

CTEC298

Pandas are python libraries used for data analysis and manipulation. In this assignment I imported pandas as pd in my Jupyter notebook python kernel then I loaded my dataset using pd.read\_csv function which excel file, and it uploaded.



The screenshot shows a JupyterLab interface with a code cell and its output. The code cell contains the following Python code:

```
[10]: import pandas as pd
import matplotlib.pyplot as plt
#Load mt dataset
df = pd.read_csv("DataBreaches(2004-2021).csv")
df
```

The output of the code cell is a preview of the DataFrame, showing the first few rows and the last few rows. The DataFrame has 5 columns: Entity, Year, Records, Organization type, and Method. The first few rows are:

	Entity	Year	Records	Organization type	Method
0	21st Century Oncology	2016	2200000	healthcare	hacked
1	500px	2020	14870304	social networking	hacked
2	Accendo Insurance Co.	2020	175350	healthcare	poor security
3	Adobe Systems Incorporated	2013	152000000	tech	hacked
4	Adobe Inc.	2019	7500000	tech	poor security

The output also shows the last few rows of the DataFrame, including rows 290, 291, 292, 293, and 294. The last row (294) shows a record from Les Éditions Protégez-vous in 2020, with 380,000 records, categorized as a publisher (magazine) with an unknown method.

295 rows x 5 columns

I wrangled my data a little bit causes my data was already clean and organized except I noticed in my method column unknow was not capitalized so I changed through this function/ statement.



```
[18]: df['Method'] = df['Method'].str.capitalize()
      print(df)
```

	Entity	Year	Records	\
0	21st Century Oncology	2016	2200000	
1	500px	2020	14870304	
2	Accendo Insurance Co.	2020	175350	
3	Adobe Systems Incorporated	2013	152000000	
4	Adobe Inc.	2019	7500000	
..	...	...	...	
290	Zynga	2019	173000000	
291	Westpac	2019	98000	
292	Unknown agency(believed to be tied to United S...	2020	200000000	
293	National Health Information Center (NCZI) of S...	2020	391250	
294	Les Éditions Protégez-vous	2020	380000	

	Organization type	Method
0	healthcare	Hacked
1	social networking	Hacked
2	healthcare	Poor security
3	tech	Hacked
4	tech	Poor security
..	...	...
290	social network	Hacked
291	financial	Hacked
292	financial	Accidentally published
293	healthcare	Poor security
294	publisher (magazine)	Unknown

[295 rows x 5 columns]