

# LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443

## 1. MEMBACA FILE CSV

INPUT	OUTPUT																																																																																																												
pwd	<div>In [1]: <code>pwd</code></div> <div>Out[1]: <code>'C:\\Users\\arifs.DESKTOP-EQJN52\\Documents\\unpam\\Semester3\\Pengantar_Big_Data'</code></div>																																																																																																												
pd.read_csv("covid_19_data.csv")	<div>In [3]: <code>pd.read_csv("covid_19_data.csv")</code></div> <div>Out[3]:</div> <table><thead><tr><th></th><th>SNo</th><th>ObservationDate</th><th>Province/State</th><th>Country/Region</th><th>Last Update</th><th>Confirmed</th><th>Deaths</th><th>Recovered</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>01/22/2020</td><td>Anhui</td><td>Mainland China</td><td>1/22/2020 17:00</td><td>1.0</td><td>0.0</td><td>0.0</td></tr><tr><td>1</td><td>2</td><td>01/22/2020</td><td>Beijing</td><td>Mainland China</td><td>1/22/2020 17:00</td><td>14.0</td><td>0.0</td><td>0.0</td></tr><tr><td>2</td><td>3</td><td>01/22/2020</td><td>Chongqing</td><td>Mainland China</td><td>1/22/2020 17:00</td><td>6.0</td><td>0.0</td><td>0.0</td></tr><tr><td>3</td><td>4</td><td>01/22/2020</td><td>Fujian</td><td>Mainland China</td><td>1/22/2020 17:00</td><td>1.0</td><td>0.0</td><td>0.0</td></tr><tr><td>4</td><td>5</td><td>01/22/2020</td><td>Gansu</td><td>Mainland China</td><td>1/22/2020 17:00</td><td>0.0</td><td>0.0</td><td>0.0</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>306424</td><td>306425</td><td>05/29/2021</td><td>Zaporizhia Oblast</td><td>Ukraine</td><td>2021-05-30 04:20:55</td><td>102641.0</td><td>2335.0</td><td>95289.0</td></tr><tr><td>306425</td><td>306426</td><td>05/29/2021</td><td>Zeeland</td><td>Netherlands</td><td>2021-05-30 04:20:55</td><td>29147.0</td><td>245.0</td><td>0.0</td></tr><tr><td>306426</td><td>306427</td><td>05/29/2021</td><td>Zhejiang</td><td>Mainland China</td><td>2021-05-30 04:20:55</td><td>1364.0</td><td>1.0</td><td>1324.0</td></tr><tr><td>306427</td><td>306428</td><td>05/29/2021</td><td>Zhytomyr Oblast</td><td>Ukraine</td><td>2021-05-30 04:20:55</td><td>87550.0</td><td>1738.0</td><td>83790.0</td></tr><tr><td>306428</td><td>306429</td><td>05/29/2021</td><td>Zuid-Holland</td><td>Netherlands</td><td>2021-05-30 04:20:55</td><td>391559.0</td><td>4252.0</td><td>0.0</td></tr></tbody></table> <div>306429 rows x 8 columns</div>		SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered	0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0	1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0	2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0	3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0	4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0	...	...	...	...	...	...	...	...	...	306424	306425	05/29/2021	Zaporizhia Oblast	Ukraine	2021-05-30 04:20:55	102641.0	2335.0	95289.0	306425	306426	05/29/2021	Zeeland	Netherlands	2021-05-30 04:20:55	29147.0	245.0	0.0	306426	306427	05/29/2021	Zhejiang	Mainland China	2021-05-30 04:20:55	1364.0	1.0	1324.0	306427	306428	05/29/2021	Zhytomyr Oblast	Ukraine	2021-05-30 04:20:55	87550.0	1738.0	83790.0	306428	306429	05/29/2021	Zuid-Holland	Netherlands	2021-05-30 04:20:55	391559.0	4252.0	0.0
	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered																																																																																																					
0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0																																																																																																					
1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0																																																																																																					
2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0																																																																																																					
3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0																																																																																																					
4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0																																																																																																					
...	...	...	...	...	...	...	...	...																																																																																																					
306424	306425	05/29/2021	Zaporizhia Oblast	Ukraine	2021-05-30 04:20:55	102641.0	2335.0	95289.0																																																																																																					
306425	306426	05/29/2021	Zeeland	Netherlands	2021-05-30 04:20:55	29147.0	245.0	0.0																																																																																																					
306426	306427	05/29/2021	Zhejiang	Mainland China	2021-05-30 04:20:55	1364.0	1.0	1324.0																																																																																																					
306427	306428	05/29/2021	Zhytomyr Oblast	Ukraine	2021-05-30 04:20:55	87550.0	1738.0	83790.0																																																																																																					
306428	306429	05/29/2021	Zuid-Holland	Netherlands	2021-05-30 04:20:55	391559.0	4252.0	0.0																																																																																																					

```
DataCovid =  
pd.read_csv("covid_19_  
data.csv")  
Data Covid
```

```
In [4]: DataCovid = pd.read_csv("covid_19_data.csv")
```

```
In [5]: DataCovid
```

```
Out[5]:
```

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0
2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0
3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...
306424	306425	05/29/2021	Zaporizhia Oblast	Ukraine	2021-05-30 04:20:55	102641.0	2335.0	95289.0
306425	306426	05/29/2021	Zeeland	Netherlands	2021-05-30 04:20:55	29147.0	245.0	0.0
306426	306427	05/29/2021	Zhejiang	Mainland China	2021-05-30 04:20:55	1364.0	1.0	1324.0
306427	306428	05/29/2021	Zhytomyr Oblast	Ukraine	2021-05-30 04:20:55	87550.0	1738.0	83790.0
306428	306429	05/29/2021	Zuid-Holland	Netherlands	2021-05-30 04:20:55	391559.0	4252.0	0.0

306429 rows x 8 columns

```
firstData =  
DataCovid.head()  
firstData
```

```
In [6]: firstData = DataCovid.head()
```

```
In [7]: firstData
```

```
Out[7]:
```

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0
2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0
3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0

```
lastData =  
DataCovid.tail()  
lastData
```

```
In [9]: lastData = DataCovid.tail()
```

```
In [10]: lastData
```

Out[10]:

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
306424	306425	05/29/2021	Zaporizhia Oblast	Ukraine	2021-05-30 04:20:55	102641.0	2335.0	95289.0
306425	306426	05/29/2021	Zeeland	Netherlands	2021-05-30 04:20:55	29147.0	245.0	0.0
306426	306427	05/29/2021	Zhejiang	Mainland China	2021-05-30 04:20:55	1364.0	1.0	1324.0
306427	306428	05/29/2021	Zhytomyr Oblast	Ukraine	2021-05-30 04:20:55	87550.0	1738.0	83790.0
306428	306429	05/29/2021	Zuid-Holland	Netherlands	2021-05-30 04:20:55	391559.0	4252.0	0.0

```
randData =  
DataCovid.sample(10)  
randData
```

```
In [11]: randData = DataCovid.sample(10)
```

```
In [12]: randData
```

Out[12]:

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
90240	90241	08/18/2020	NaN	Zimbabwe	2021-04-02 15:13:53	5378.0	141.0	4105.0
102523	102524	09/03/2020	Noord-Brabant	Netherlands	2021-04-02 15:13:53	12400.0	1550.0	0.0
90580	90581	08/18/2020	New South Wales	Australia	2021-04-02 15:13:53	3966.0	52.0	0.0
301626	301627	05/23/2021	Okinawa	Japan	2021-05-24 04:20:53	15139.0	149.0	12884.0
253118	253119	03/21/2021	Bremen	Germany	2021-04-02 15:13:53	19711.0	392.0	18184.0
70052	70053	07/22/2020	NaN	Seychelles	2021-04-02 15:13:53	108.0	0.0	27.0
221628	221629	02/08/2021	NaN	Niger	2021-04-02 15:13:53	4621.0	166.0	3983.0
112616	112617	09/17/2020	NaN	Vietnam	2021-04-02 15:13:53	1066.0	35.0	940.0
100870	100871	09/01/2020	Hiroshima	Japan	2021-04-02 15:13:53	455.0	3.0	439.0
26850	26851	05/19/2020	NaN	South Sudan	2021-04-02 15:13:53	290.0	4.0	4.0

DataCovid.info()

In [15]: DataCovid.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306429 entries, 0 to 306428
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   SNo              306429 non-null  int64
1   ObservationDate  306429 non-null  object
2   Province/State  228326 non-null  object
3   Country/Region  306429 non-null  object
4   Last Update     306429 non-null  object
5   Confirmed       306429 non-null  float64
6   Deaths         306429 non-null  float64
7   Recovered       306429 non-null  float64
dtypes: float64(3), int64(1), object(4)
memory usage: 18.7+ MB
```

dataSample =  
DataCovid[['Country/Region', 'Confirmed',  
'Deaths', 'Recovered']]  
dataSample

In [16]: dataSample = DataCovid[['Country/Region', 'Confirmed', 'Deaths', 'Recovered']]

In [17]: dataSample

Out[17]:

	Country/Region	Confirmed	Deaths	Recovered
0	Mainland China	1.0	0.0	0.0
1	Mainland China	14.0	0.0	0.0
2	Mainland China	6.0	0.0	0.0
3	Mainland China	1.0	0.0	0.0
4	Mainland China	0.0	0.0	0.0
...	...	...	...	...
306424	Ukraine	102641.0	2335.0	95289.0
306425	Netherlands	29147.0	245.0	0.0
306426	Mainland China	1364.0	1.0	1324.0
306427	Ukraine	87550.0	1738.0	83790.0
306428	Netherlands	391559.0	4252.0	0.0

306429 rows x 4 columns

```
dataSample.to_csv("data_19_covid_baru.csv",  
index=False)
```

FilesRunningClusters

Select items to perform actions on them.

0

/ Documents / unpam / Semester3 / Pengantar\_Big\_Data

Name

Last Modified

File size

..

seconds ago

snapshot

3 minutes ago

myfirstJupyter.ipynb

Running a minute ago

49.3 kB

03SIFE003 (7 September 2023).pdf

2 months ago

406 kB

Arif-Frima-Ari-Suwadji\_03SIFE003\_Pertemuan8.xlsx

17 hours ago

104 kB

covid\_19\_data.csv

19 hours ago

22.5 MB

Data Excel BPJS.xlsx

19 hours ago

9.91 MB

data\_19\_covid\_baru.csv

seconds ago

8.62 MB

```
readExcel =  
pd.read_excel("Data  
Excel BPJS.xlsx",  
sheet_name='Sheet2')  
readExcel
```












```
In [24]: readExcel = pd.read_excel("Data Excel BPJS.xlsx", sheet_name='Sheet2')
```

```
In [26]: readExcel
```

Out[26]:

	Tanggal/Jam	Trace Code	Reference Code	Client	Kode Produk	Nama Produk	Tipe Transaksi	No. Pelanggan	Status Transaksi	#
0	23-01-10 11:59:51	488a9df322c74c439342e4fe3f1a4b33	3505159491	Sepulsa BPJS	ASRBPKJSKES	ASURANSI BPJS KESEHATAN	PASCABAYAR	0003175208076 - 01	0	T
1	23-01-10 11:59:47	99d22edbe57b47c4ac3de491724ad1c3	2898155588	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888801824224837	0	T
2	23-01-10 11:59:46	08afb333fa8441c5b137cb7e2e01d91a	9067592	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802051450752	0	T
3	23-01-10 11:59:46	c66f18e055344e57aa99ac3330bc6686	231AC0335001	Mitracomm Ekasarana	ASRBPKJSKES	ASURANSI BPJS KESEHATAN	PASCABAYAR	8888801816570697 - 01	0	T
4	23-01-10 11:59:45	1a8c34eec5114ceebe28c81459c41063	2898155430	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802440666214	0	T
...	...	...	...	...	...	...	...	...	...	...
24979	23-01-09 12:00:14	867696794ea746889d72e3c1c715bd3a	8790223	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802282326187	0	T
24980	23-01-09 12:00:13	16c53efdc7ce42698a6be9a043f01041	2897503082	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802481624551	0	T
24981	23-01-09 12:00:11	928ba352e2bd403a80392f81a36cfc3f	2319C0359001	Mitracomm Ekasarana	ASRBPKJSKES	ASURANSI BPJS KESEHATAN	PASCABAYAR	0003136310381 - 01	0	T

```
readExcel.to_excel("Data  
a Excel BPJS Baru.xlsx",  
index=False,  
sheet_name="Detail")
```

Files			Running		Clusters	
Duplicate			Rename	Move	Download	View Edit 
1			Upload			New 
 / Documents / unpam / Semester3 / Pengantar_Big_Data			Name 		Last Modified	File size
 ..					seconds ago	
<input type="checkbox"/>  screenshot					6 minutes ago	
<input type="checkbox"/>  myfirstJupyter.ipynb			Running		seconds ago	69.3 kB
<input type="checkbox"/>  03SIFE003 (7 September 2023).pdf					2 months ago	406 kB
<input type="checkbox"/>  Arif-Frima-Ari-Suwadji_03SIFE003_Pertemuan8.xlsx					18 hours ago	104 kB
<input type="checkbox"/>  covid_19_data.csv					19 hours ago	22.5 MB
<input type="checkbox"/>  Data Excel BPJS Baru.xlsx					seconds ago	4.47 MB
<input checked="" type="checkbox"/>  Data Excel BPJS.xlsx					19 hours ago	9.91 MB
<input type="checkbox"/>  data_19_covid_baru.csv					10 minutes ago	8.62 MB
<input type="checkbox"/>  MODUL DIGITAL 1-20230905.zip					2 months ago	391 kB

pd.read\_excel("Data  
Excel BPJS Baru.xlsx",  
sheet\_name="Detail")

In [28]: pd.read\_excel("Data Excel BPJS Baru.xlsx", sheet\_name="Detail")

Out[28]:

	Tanggal/Jam	Trace Code	Reference Code	Client	Kode Produk	Nama Produk	Tipe Transaksi	No. Pelanggan	Status Transaksi	#
0	23-01-10 11:59:51	488a9df322c74c439342e4fe3f1a4b33	3505159491	Sepulsa BPJS	ASRBPKJSKES	ASURANSI BPJS KESEHATAN	PASCABAYAR	0003175208076 - 01	0	T
1	23-01-10 11:59:47	99d22edbe57b47c4ac3de491724ad1c3	2898155588	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888801824224837	0	T
2	23-01-10 11:59:46	08afb333fa8441c5b137cb7e2e01d91a	9067592	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802051450752	0	T
3	23-01-10 11:59:46	c66f18e055344e57aa99ac3330bc6686	231AC0335001	Mitracomm Ekasarana	ASRBPKJSKES	ASURANSI BPJS KESEHATAN	PASCABAYAR	8888801816570697 - 01	0	T
4	23-01-10 11:59:45	1a8c34eec5114ceebe28c81459c41063	2898155430	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802440666214	0	T
...	...	...	...	...	...	...	...	...	...	...
24979	23-01-09 12:00:14	867696794ea746889d72e3c1c715bd3a	8790223	Arthasera	BPJSKS	BPJS KESEHATAN BRIDGE ORCHESTRA	PASCABAYAR	8888802282326187	0	T



```
readHtml =  
pd.read_html("https://w  
ww.fdic.gov/resources/r  
esolutions/bank-  
failures/failed-bank-  
list/")  
readHtml
```

```
In [29]: readHtml = pd.read_html("https://www.fdic.gov/resources/resolutions/bank-failures/failed-bank-list/")
```

```
In [30]: readHtml[0]
```

```
Out[30]:
```

	Bank NameBank	CityCity	StateSt	CertCert	Acquiring InstitutionAI	Closing DateClosing	FundFund
0	Citizens Bank	Sac City	IA	8758	Iowa Trust & Savings Bank	November 3, 2023	10545
1	Heartland Tri-State Bank	Elkhart	KS	25851	Dream First Bank, N.A.	July 28, 2023	10544
2	First Republic Bank	San Francisco	CA	59017	JPMorgan Chase Bank, N.A.	May 1, 2023	10543
3	Signature Bank	New York	NY	57053	Flagstar Bank, N.A.	March 12, 2023	10540
4	Silicon Valley Bank	Santa Clara	CA	24735	First-Citizens Bank & Trust Company	March 10, 2023	10539
...	...	...	...	...	...	...	...
563	Superior Bank, FSB	Hinsdale	IL	32646	Superior Federal, FSB	July 27, 2001	6004
564	Malta National Bank	Malta	OH	6629	North Valley Bank	May 3, 2001	4648
565	First Alliance Bank & Trust Co.	Manchester	NH	34264	Southern New Hampshire Bank & Trust	February 2, 2001	4647
566	National State Bank of Metropolis	Metropolis	IL	3815	Banterra Bank of Marion	December 14, 2000	4646
567	Bank of Honolulu	Honolulu	HI	21029	Bank of the Orient	October 13, 2000	4645

```
568 rows x 7 columns
```