

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

1. MEMBACA FILE CSV

INPUT	OUTPUT																																																																																																																																																																								
pwd	<pre>In [5]: pwd Out[5]: 'C:\\Users\\arifs.DESKTOP-EQJN52\\Documents\\unpam\\Semester3\\Pengantar_Big_Data\\KelasBigData'</pre>																																																																																																																																																																								
datacovid = pd.read_csv("datacovid19.csv") datacovid	<pre>In [8]: datacovid = pd.read_csv("datacovid19.csv") datacovid Out[8]:</pre> <table><tr><th></th><th>province_state</th><th>country_region</th><th>date</th><th>latitude</th><th>longitude</th><th>location_geom</th><th>confirmed</th><th>deaths</th><th>recovered</th><th>active</th><th>fips</th><th>admin2</th><th>combined_key</th></tr><tr><td>0</td><td>NaN</td><td>UK</td><td>01-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>0.0</td><td>0.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>1</td><td>NaN</td><td>UK</td><td>18-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>2</td><td>NaN</td><td>UK</td><td>17-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>3</td><td>NaN</td><td>UK</td><td>31-01-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>4</td><td>NaN</td><td>UK</td><td>19-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>995</td><td>Iowa</td><td>US</td><td>21-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>9</td><td>0.0</td><td>0.0</td><td>9.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>996</td><td>Iowa</td><td>US</td><td>21-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>40</td><td>1.0</td><td>0.0</td><td>39.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>997</td><td>Iowa</td><td>US</td><td>26-04-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>4</td><td>0.0</td><td>0.0</td><td>4.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>998</td><td>Iowa</td><td>US</td><td>26-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>45</td><td>1.0</td><td>0.0</td><td>44.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>999</td><td>Iowa</td><td>US</td><td>26-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>13</td><td>0.0</td><td>0.0</td><td>13.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr></table> <p>1000 rows x 13 columns</p>		province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key	0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN	1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN	996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN	997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN	998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN	999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN
	province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key																																																																																																																																																												
0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN																																																																																																																																																												
1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																												
4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
...																																																																																																																																																												
995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN																																																																																																																																																												
996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN																																																																																																																																																												
997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN																																																																																																																																																												
998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN																																																																																																																																																												
999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN																																																																																																																																																												

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT																																																																																																																																																																								
Data yang Hilang datacovid.isnull()	<div><div>In [15]: datacovid.isnull()</div><div>Out[15]:</div><table><tr><th></th><th>province_state</th><th>country_region</th><th>date</th><th>latitude</th><th>longitude</th><th>location_geom</th><th>confirmed</th><th>deaths</th><th>recovered</th><th>active</th><th>fips</th><th>admin2</th><th>combined_key</th></tr><tr><td>0</td><td>True</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>False</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>True</td></tr><tr><td>1</td><td>True</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>False</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>True</td></tr><tr><td>2</td><td>True</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>False</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>True</td></tr><tr><td>3</td><td>True</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>False</td><td>True</td><td>True</td><td>True</td><td>True</td><td>True</td><td>True</td></tr><tr><td>4</td><td>True</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>False</td><td>False</td><td>False</td><td>True</td><td>True</td><td>True</td><td>True</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>995</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>996</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>997</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>998</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>999</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>False</td><td>True</td></tr></table><div>1000 rows x 13 columns</div></div>		province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key	0	True	False	False	True	True	True	False	False	False	True	True	True	True	1	True	False	False	True	True	True	False	False	False	True	True	True	True	2	True	False	False	True	True	True	False	False	False	True	True	True	True	3	True	False	False	True	True	True	False	True	True	True	True	True	True	4	True	False	False	True	True	True	False	False	False	True	True	True	True	995	False	False	False	False	False	False	False	False	False	False	False	False	True	996	False	False	False	False	False	False	False	False	False	False	False	False	True	997	False	False	False	False	False	False	False	False	False	False	False	False	True	998	False	False	False	False	False	False	False	False	False	False	False	False	True	999	False	False	False	False	False	False	False	False	False	False	False	False	True
	province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key																																																																																																																																																												
0	True	False	False	True	True	True	False	False	False	True	True	True	True																																																																																																																																																												
1	True	False	False	True	True	True	False	False	False	True	True	True	True																																																																																																																																																												
2	True	False	False	True	True	True	False	False	False	True	True	True	True																																																																																																																																																												
3	True	False	False	True	True	True	False	True	True	True	True	True	True																																																																																																																																																												
4	True	False	False	True	True	True	False	False	False	True	True	True	True																																																																																																																																																												
...																																																																																																																																																												
995	False	False	False	False	False	False	False	False	False	False	False	False	True																																																																																																																																																												
996	False	False	False	False	False	False	False	False	False	False	False	False	True																																																																																																																																																												
997	False	False	False	False	False	False	False	False	False	False	False	False	True																																																																																																																																																												
998	False	False	False	False	False	False	False	False	False	False	False	False	True																																																																																																																																																												
999	False	False	False	False	False	False	False	False	False	False	False	False	True																																																																																																																																																												
Menghitung Data yang Hilang datacovid.isnull().sum()	<div><div>In [18]: datacovid.isnull().sum()</div><div>Out[18]:</div><table><tr><td>province_state</td><td>40</td></tr><tr><td>country_region</td><td>0</td></tr><tr><td>date</td><td>0</td></tr><tr><td>latitude</td><td>126</td></tr><tr><td>longitude</td><td>126</td></tr><tr><td>location_geom</td><td>126</td></tr><tr><td>confirmed</td><td>0</td></tr><tr><td>deaths</td><td>1</td></tr><tr><td>recovered</td><td>1</td></tr><tr><td>active</td><td>51</td></tr><tr><td>fips</td><td>51</td></tr><tr><td>admin2</td><td>158</td></tr><tr><td>combined_key</td><td>1000</td></tr><tr><td>dtype:</td><td>int64</td></tr></table></div>	province_state	40	country_region	0	date	0	latitude	126	longitude	126	location_geom	126	confirmed	0	deaths	1	recovered	1	active	51	fips	51	admin2	158	combined_key	1000	dtype:	int64																																																																																																																																												
province_state	40																																																																																																																																																																								
country_region	0																																																																																																																																																																								
date	0																																																																																																																																																																								
latitude	126																																																																																																																																																																								
longitude	126																																																																																																																																																																								
location_geom	126																																																																																																																																																																								
confirmed	0																																																																																																																																																																								
deaths	1																																																																																																																																																																								
recovered	1																																																																																																																																																																								
active	51																																																																																																																																																																								
fips	51																																																																																																																																																																								
admin2	158																																																																																																																																																																								
combined_key	1000																																																																																																																																																																								
dtype:	int64																																																																																																																																																																								

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT																																																																																																																																																																								
<p>Perbaiki data dengan menghapus yang kosong</p> <p>menyimpan ke variable baru agar tidak perlu import csv lagi saat memerlukan data yang sama</p> <p>datacovid2=datacovid datacovid2</p>	<div><div>In [21]: datacovid2 = datacovid datacovid2</div><div>Out[21]:</div><table><tr><th></th><th>province_state</th><th>country_region</th><th>date</th><th>latitude</th><th>longitude</th><th>location_geom</th><th>confirmed</th><th>deaths</th><th>recovered</th><th>active</th><th>fips</th><th>admin2</th><th>combined_key</th></tr><tr><td>0</td><td>NaN</td><td>UK</td><td>01-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>0.0</td><td>0.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>1</td><td>NaN</td><td>UK</td><td>18-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>2</td><td>NaN</td><td>UK</td><td>17-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>3</td><td>NaN</td><td>UK</td><td>31-01-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>4</td><td>NaN</td><td>UK</td><td>19-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>995</td><td>Iowa</td><td>US</td><td>21-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>9</td><td>0.0</td><td>0.0</td><td>9.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>996</td><td>Iowa</td><td>US</td><td>21-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>40</td><td>1.0</td><td>0.0</td><td>39.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>997</td><td>Iowa</td><td>US</td><td>26-04-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>4</td><td>0.0</td><td>0.0</td><td>4.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>998</td><td>Iowa</td><td>US</td><td>26-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>45</td><td>1.0</td><td>0.0</td><td>44.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>999</td><td>Iowa</td><td>US</td><td>26-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>13</td><td>0.0</td><td>0.0</td><td>13.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr></table><div>1000 rows × 13 columns</div></div>		province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key	0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN	1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN	996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN	997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN	998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN	999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN
	province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key																																																																																																																																																												
0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN																																																																																																																																																												
1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																												
4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
...																																																																																																																																																												
995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN																																																																																																																																																												
996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN																																																																																																																																																												
997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN																																																																																																																																																												
998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN																																																																																																																																																												
999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN																																																																																																																																																												

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT																																																																																																																																																																															
<p>Perbaiki data dengan menghapus yang kosong</p> <p>menghapus data</p> <pre>print("data sebelum dihapus\n", datacovid2) datacovidCleaned = datacovid2.dropna() print("datasetelah dihapus\n", datacovidCleaned)</pre>	<pre>In [26]: print("data sebelum dihapus\n", datacovid2) datacovidCleaned = datacovid2.dropna() print("datasetelah dihapus\n", datacovidCleaned)</pre> <table><tr><th colspan="7">data sebelum dihapus</th></tr><tr><th></th><th>province_state</th><th>country_region</th><th>date</th><th>latitude</th><th>longitude</th><th>\</th></tr><tr><td>0</td><td>NaN</td><td>UK</td><td>01-02-20</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>1</td><td>NaN</td><td>UK</td><td>18-02-20</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>2</td><td>NaN</td><td>UK</td><td>17-02-20</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>3</td><td>NaN</td><td>UK</td><td>31-01-20</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>4</td><td>NaN</td><td>UK</td><td>19-02-20</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>..</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td></td></tr><tr><td>995</td><td>Iowa</td><td>US</td><td>21-05-20</td><td>42.776443</td><td>-94.207225</td><td></td></tr><tr><td>996</td><td>Iowa</td><td>US</td><td>21-06-20</td><td>42.776443</td><td>-94.207225</td><td></td></tr><tr><td>997</td><td>Iowa</td><td>US</td><td>26-04-20</td><td>42.776443</td><td>-94.207225</td><td></td></tr><tr><td>998</td><td>Iowa</td><td>US</td><td>26-06-20</td><td>42.776443</td><td>-94.207225</td><td></td></tr><tr><td>999</td><td>Iowa</td><td>US</td><td>26-05-20</td><td>42.776443</td><td>-94.207225</td><td></td></tr></table> <table><tr><th></th><th>location_geom</th><th>confirmed</th><th>deaths</th><th>recovered</th><th>active</th><th>\</th></tr><tr><td>0</td><td>NaN</td><td>2</td><td>0.0</td><td>0.0</td><td>NaN</td><td></td></tr><tr><td>1</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td></td></tr><tr><td>2</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td></td></tr><tr><td>3</td><td>NaN</td><td>2</td><td>NaN</td><td>NaN</td><td>NaN</td><td></td></tr><tr><td>4</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td></td></tr><tr><td>..</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td></td></tr><tr><td>995</td><td>POINT(-94.20722537 42.7764426)</td><td>9</td><td>0.0</td><td>0.0</td><td>9.0</td><td></td></tr><tr><td>996</td><td>POINT(-94.20722537 42.7764426)</td><td>40</td><td>1.0</td><td>0.0</td><td>39.0</td><td></td></tr><tr><td>997</td><td>POINT(-94.20722537 42.7764426)</td><td>4</td><td>0.0</td><td>0.0</td><td>4.0</td><td></td></tr><tr><td>998</td><td>POINT(-94.20722537 42.7764426)</td><td>45</td><td>1.0</td><td>0.0</td><td>44.0</td><td></td></tr><tr><td>999</td><td>POINT(-94.20722537 42.7764426)</td><td>13</td><td>0.0</td><td>0.0</td><td>13.0</td><td></td></tr></table> <pre>datasetelah dihapus Empty DataFrame Columns: [province_state, country_region, date, latitude, longitude, location_geom, confirmed, deaths, recovered, active, fips, admin2, combined_key] Index: []</pre>	data sebelum dihapus								province_state	country_region	date	latitude	longitude	\	0	NaN	UK	01-02-20	NaN	NaN		1	NaN	UK	18-02-20	NaN	NaN		2	NaN	UK	17-02-20	NaN	NaN		3	NaN	UK	31-01-20	NaN	NaN		4	NaN	UK	19-02-20	NaN	NaN			995	Iowa	US	21-05-20	42.776443	-94.207225		996	Iowa	US	21-06-20	42.776443	-94.207225		997	Iowa	US	26-04-20	42.776443	-94.207225		998	Iowa	US	26-06-20	42.776443	-94.207225		999	Iowa	US	26-05-20	42.776443	-94.207225			location_geom	confirmed	deaths	recovered	active	\	0	NaN	2	0.0	0.0	NaN		1	NaN	9	0.0	8.0	NaN		2	NaN	9	0.0	8.0	NaN		3	NaN	2	NaN	NaN	NaN		4	NaN	9	0.0	8.0	NaN			995	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0		996	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0		997	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0		998	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0		999	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	
data sebelum dihapus																																																																																																																																																																																
	province_state	country_region	date	latitude	longitude	\																																																																																																																																																																										
0	NaN	UK	01-02-20	NaN	NaN																																																																																																																																																																											
1	NaN	UK	18-02-20	NaN	NaN																																																																																																																																																																											
2	NaN	UK	17-02-20	NaN	NaN																																																																																																																																																																											
3	NaN	UK	31-01-20	NaN	NaN																																																																																																																																																																											
4	NaN	UK	19-02-20	NaN	NaN																																																																																																																																																																											
..																																																																																																																																																																											
995	Iowa	US	21-05-20	42.776443	-94.207225																																																																																																																																																																											
996	Iowa	US	21-06-20	42.776443	-94.207225																																																																																																																																																																											
997	Iowa	US	26-04-20	42.776443	-94.207225																																																																																																																																																																											
998	Iowa	US	26-06-20	42.776443	-94.207225																																																																																																																																																																											
999	Iowa	US	26-05-20	42.776443	-94.207225																																																																																																																																																																											
	location_geom	confirmed	deaths	recovered	active	\																																																																																																																																																																										
0	NaN	2	0.0	0.0	NaN																																																																																																																																																																											
1	NaN	9	0.0	8.0	NaN																																																																																																																																																																											
2	NaN	9	0.0	8.0	NaN																																																																																																																																																																											
3	NaN	2	NaN	NaN	NaN																																																																																																																																																																											
4	NaN	9	0.0	8.0	NaN																																																																																																																																																																											
..																																																																																																																																																																											
995	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0																																																																																																																																																																											
996	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0																																																																																																																																																																											
997	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0																																																																																																																																																																											
998	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0																																																																																																																																																																											
999	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0																																																																																																																																																																											

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT																																																																																																																																																																								
perbaiki data dengan mengisi data yang kosong	<div><div>In [27]:<pre>#simpan variable baru dari data yang utama datacovid3 = datacovid datacovid3</pre></div><div>Out[27]:</div></div>																																																																																																																																																																								
datacovid3=datacovid datacovid3	<table><tr><th></th><th>province_state</th><th>country_region</th><th>date</th><th>latitude</th><th>longitude</th><th>location_geom</th><th>confirmed</th><th>deaths</th><th>recovered</th><th>active</th><th>fips</th><th>admin2</th><th>combined_key</th></tr><tr><td>0</td><td>NaN</td><td>UK</td><td>01-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>0.0</td><td>0.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>1</td><td>NaN</td><td>UK</td><td>18-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>2</td><td>NaN</td><td>UK</td><td>17-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>3</td><td>NaN</td><td>UK</td><td>31-01-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>2</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>4</td><td>NaN</td><td>UK</td><td>19-02-20</td><td>NaN</td><td>NaN</td><td>NaN</td><td>9</td><td>0.0</td><td>8.0</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>995</td><td>Iowa</td><td>US</td><td>21-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>9</td><td>0.0</td><td>0.0</td><td>9.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>996</td><td>Iowa</td><td>US</td><td>21-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>40</td><td>1.0</td><td>0.0</td><td>39.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>997</td><td>Iowa</td><td>US</td><td>26-04-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>4</td><td>0.0</td><td>0.0</td><td>4.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>998</td><td>Iowa</td><td>US</td><td>26-06-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>45</td><td>1.0</td><td>0.0</td><td>44.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr><tr><td>999</td><td>Iowa</td><td>US</td><td>26-05-20</td><td>42.776443</td><td>-94.207225</td><td>POINT(-94.20722537 42.7764426)</td><td>13</td><td>0.0</td><td>0.0</td><td>13.0</td><td>19091.0</td><td>Humboldt</td><td>NaN</td></tr></table> <div>1000 rows x 13 columns</div>		province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key	0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN	1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN	995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN	996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN	997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN	998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN	999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN
	province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key																																																																																																																																																												
0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN																																																																																																																																																												
1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																												
4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN																																																																																																																																																												
...																																																																																																																																																												
995	Iowa	US	21-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	9	0.0	0.0	9.0	19091.0	Humboldt	NaN																																																																																																																																																												
996	Iowa	US	21-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	40	1.0	0.0	39.0	19091.0	Humboldt	NaN																																																																																																																																																												
997	Iowa	US	26-04-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	4	0.0	0.0	4.0	19091.0	Humboldt	NaN																																																																																																																																																												
998	Iowa	US	26-06-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	45	1.0	0.0	44.0	19091.0	Humboldt	NaN																																																																																																																																																												
999	Iowa	US	26-05-20	42.776443	-94.207225	POINT(-94.20722537 42.7764426)	13	0.0	0.0	13.0	19091.0	Humboldt	NaN																																																																																																																																																												

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT
<p>perbaiki data dengan mengisi data yang kosong</p> <p># melakukan pengecekan tipe data column sebelum fullfillment datacovid3.info()</p>	<pre>In [40]: # melakukan pengecekan tipe data column sebelum fullfillment datacovid3.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 1000 entries, 0 to 999 Data columns (total 13 columns): # Column Non-Null Count Dtype --- - 0 province_state 1000 non-null object 1 country_region 1000 non-null object 2 date 1000 non-null object 3 latitude 1000 non-null float64 4 longitude 1000 non-null float64 5 location_geom 1000 non-null object 6 confirmed 1000 non-null int64 7 deaths 1000 non-null float64 8 recovered 1000 non-null float64 9 active 1000 non-null float64 10 fips 1000 non-null float64 11 admin2 1000 non-null object 12 combined_key 0 non-null float64 dtypes: float64(7), int64(1), object(5) memory usage: 101.7+ KB</pre>

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT
<p>perbaiki data dengan mengisi data yang kosong</p> <p># mengisi data yang hilang berdasarkan tipe data kolomnya</p> <pre>for col in datacovid3.columns: dtype = datacovid3[col].dtype if dtype == 'float64': datacovid3[col].fillna(datacovid3[col].mean(), inplace=True) elif dtype == 'int64': datacovid3[col].fillna(datacovid3[col].median(), inplace=True) elif dtype == 'object': datacovid3[col].fillna('x', inplace=True) if col == 'combined_key': datacovid3[col].fillna('y', inplace=True)</pre>	<pre>In [49]: # mengisi data yang hilang berdasarkan tipe data kolomnya for col in datacovid3.columns: dtype = datacovid3[col].dtype if dtype == 'float64': datacovid3[col].fillna(datacovid3[col].mean(), inplace=True) elif dtype == 'int64': datacovid3[col].fillna(datacovid3[col].median(), inplace=True) elif dtype == 'object': datacovid3[col].fillna('x', inplace=True) if col == 'combined_key': datacovid3[col].fillna('y', inplace=True)</pre>

LAPORAN MEMBACA FILE DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT
perbaiki data dengan mengisi data yang kosong	
print("menampilkan data setelah fullfillment\n", datacovid3)	<pre> In [51]: print("menampilkan data setelah fullfillment\n", datacovid3) menampilkan data setelah fullfillment province_state country_region date latitude longitude \ 0 x UK 01-02-20 38.426810 -60.738179 1 x UK 18-02-20 38.426810 -60.738179 2 x UK 17-02-20 38.426810 -60.738179 3 x UK 31-01-20 38.426810 -60.738179 4 x UK 19-02-20 38.426810 -60.738179 995 Iowa US 21-05-20 42.776443 -94.207225 996 Iowa US 21-06-20 42.776443 -94.207225 997 Iowa US 26-04-20 42.776443 -94.207225 998 Iowa US 26-06-20 42.776443 -94.207225 999 Iowa US 26-05-20 42.776443 -94.207225 location_geom confirmed deaths recovered \ 0 x 2 0.000000 0.000000 1 x 9 0.000000 8.000000 2 x 9 0.000000 8.000000 3 x 2 4.672673 0.768769 4 x 9 0.000000 8.000000 995 POINT(-94.20722537 42.7764426) 9 0.000000 0.000000 996 POINT(-94.20722537 42.7764426) 40 1.000000 0.000000 997 POINT(-94.20722537 42.7764426) 4 0.000000 0.000000 998 POINT(-94.20722537 42.7764426) 45 1.000000 0.000000 999 POINT(-94.20722537 42.7764426) 13 0.000000 0.000000 active fips admin2 combined_key 0 192.571128 24880.724974 x y 1 192.571128 24880.724974 x y 2 192.571128 24880.724974 x y 3 192.571128 24880.724974 x y 4 192.571128 24880.724974 x y 995 9.000000 19091.000000 Humboldt y 996 39.000000 19091.000000 Humboldt y 997 4.000000 19091.000000 Humboldt y 998 44.000000 19091.000000 Humboldt y 999 13.000000 19091.000000 Humboldt y [1000 rows x 13 columns] </pre>