Covid-19 Project

Data cleaning and transformation

Importing libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Loading Data

```
raw data confirmed =
pd.read csv('C:/Users/as355/Downloads/Covid-19/time series covid19 con
firmed global.csv')
raw data deaths =
pd.read csv('C:/Users/as355/Downloads/Covid-19/time series covid19 dea
ths global.csv')
raw data recovered =
pd.read csv('C:/Users/as355/Downloads/Covid-19/time series covid19 rec
overed global.csv')
print("The shape of confirmed:",raw data confirmed.shape)
print("The shape of deaths:",raw_data_deaths.shape)
print("The shape of recovered:", raw data recovered.shape)
The shape of confirmed: (271, 350)
The shape of deaths: (271, 350)
The shape of recovered: (256, 350)
raw data recovered
    Province/State
                        Country/Region
                                               Lat
                                                          Long 1/22/20
/
0
               NaN
                           Afghanistan 33.939110
                                                     67.709953
                                                                      0
1
               NaN
                               Albania
                                        41.153300
                                                     20.168300
                                                                      0
2
               NaN
                               Algeria 28.033900
                                                      1.659600
                                                                      0
3
               NaN
                               Andorra 42.506300
                                                      1.521800
                                                                      0
               NaN
                                Angola -11.202700
                                                     17.873900
                                                                      0
251
                               Vietnam 14.058324 108.277199
               NaN
```

252		NaN Wes	st Bank and	Gaza 31.9	952200 35	.233200	Θ
253		NaN	,	Yemen 15.5	552727 48	.516388	0
						.310300	J
254		NaN	Za	ambia -13.1	133897 27	.849332	0
255		NaN	Ziml	babwe -19.0	915438 29	. 154857	0
1/	23/20	1/24/20	1/25/20 1,	/26/20 1/2	27/20	12/23/20	
12/24/2		0	0	0	0	20602	
0 40359	0	0	0	0	0	39692	
1	0	Θ	Θ	0	0	29799	
30276	0	0	0	0	0	64401	
2 64777	0	0	0	0	0	64401	
3	0	0	0	0	0	7106	
7171	0	0	0	0	0	0720	
4 9729	ีย	U	U	0	0	9729	
 251	0	0	0	0	0	1281	
1281	U	U	U	U	0	1201	
252	0	0	0	0	Θ	102942	
105233 253	0	0	0	0	0	1384	
1384	U	U	U	U	0	1304	
254	0	0	0	0	0	17969	
18036 255	0	Θ	0	0	0	10259	
10314	U	U	O	U	0	10233	
10	/25 /20	12/26/20	12/27/20	12/20/20	12/20/20	12/20/20	
12/31/2	25/20	12/26/20	12/2//20	12/28/20	12/29/20	12/30/20	
0	40444	40784	41096	41441	41543	41612	
41727 1	30790	31181	31565	32122	32700	33185	
33634	30790	3110	31303	32122	32700	33103	
2	65144	65505	65862	66214	66550	66855	
67127 3	7203	7252	7288	7318	7360	7384	
7432	7203	7232	7200	7510	7500	7504	
4	9921	9976	10354	10354	10627	10859	
11044							
251	1303	1303	3 1303	1303	1319	1323	
1325							

```
252
       106718
                  108118
                              109377
                                         110927
                                                    113082
                                                               115127
117183
253
          1384
                     1384
                                1384
                                           1384
                                                      1387
                                                                 1392
1394
                                                                18530
254
        18127
                    18200
                               18210
                                          18296
                                                     18380
18660
255
        10468
                    10519
                               10593
                                          10705
                                                     11067
                                                                11154
11250
     1/1/21
0
      41727
1
      33634
2
      67395
3
       7463
4
      11146
. .
251
       1325
252
     118926
253
       1396
254
      18773
255
      11347
[256 rows x 350 columns]
```

Unpivoting the data

```
raw_data_confirmed1 =
pd.melt(raw_data_confirmed,id_vars=['Province/State','Country/Region',
'Lat','Long'],var_name='Date')
raw_data_deaths1 =
pd.melt(raw_data_deaths,id_vars=['Province/State','Country/Region','La
t','Long'],var_name='Date')
raw_data_recovered1 =
pd.melt(raw_data_recovered,id_vars=['Province/State','Country/Region',
'Lat','Long'],var_name='Date')

print("The shape of confirmed1:",raw_data_confirmed1.shape)
print("The shape of deaths1:",raw_data_deaths1.shape)
print("The shape of recovered1:",raw_data_recovered1.shape)

The shape of confirmed1: (93766, 6)
The shape of recovered1: (88576, 6)
```

Converting the new column date in the dataframes to 'datetime' format

```
raw data confirmed1['Date'] =
pd.to datetime(raw_data_confirmed1['Date'])
raw data confirmed1
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\2916553968.py:1:
UserWarning: Could not infer format, so each element will be parsed
individually, falling back to `dateutil`. To ensure parsing is
consistent and as-expected, please specify a format.
  raw data confirmed1['Date'] =
pd.to datetime(raw data confirmed1['Date'])
      Province/State
                           Country/Region
                                                  Lat
                                                             Long
Date
                                                        67.709953 2020-
                 NaN
                              Afghanistan 33.939110
0
01 - 22
                                  Albania 41.153300
                                                        20.168300 2020-
                 NaN
01 - 22
2
                                                         1.659600 2020-
                 NaN
                                  Algeria 28.033900
01-22
                                  Andorra 42.506300
                                                         1.521800 2020-
                 NaN
01 - 22
                                   Angola -11.202700
                                                        17.873900 2020-
4
                  NaN
01-22
. . .
. . .
                                                       108.277199 2021-
93761
                 NaN
                                  Vietnam 14.058324
01 - 01
                      West Bank and Gaza 31.952200
                                                        35.233200 2021-
93762
                 NaN
01 - 01
                                    Yemen 15.552727
93763
                 NaN
                                                        48.516388 2021-
01 - 01
93764
                 NaN
                                   Zambia -13.133897
                                                        27.849332 2021-
01-01
93765
                 NaN
                                 Zimbabwe -19.015438
                                                        29.154857 2021-
01 - 01
        value
0
            0
1
            0
2
            0
3
            0
4
            0
93761
         1474
93762
       139223
93763
         2101
93764
        20997
```

```
93765
        14084
[93766 rows x 6 columns]
raw data deaths1['Date'] = pd.to datetime(raw data deaths1['Date'])
raw data deaths1
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\407412485.py:1:
UserWarning: Could not infer format, so each element will be parsed
individually, falling back to `dateutil`. To ensure parsing is
consistent and as-expected, please specify a format.
  raw data deaths1['Date'] = pd.to datetime(raw data deaths1['Date'])
      Province/State
                          Country/Region
                                                 Lat
                                                             Long
Date
                                                       67.709953 2020-
                 NaN
                             Afghanistan 33.939110
0
01 - 22
                                                       20.168300 2020-
                 NaN
                                  Albania
                                           41.153300
1
01 - 22
                                                        1.659600 2020-
                 NaN
                                  Algeria 28.033900
01-22
                 NaN
                                  Andorra 42.506300
                                                        1.521800 2020-
01-22
                 NaN
                                   Angola -11.202700
                                                       17.873900 2020-
01-22
                                  Vietnam 14.058324
                                                      108.277199 2021-
93761
                 NaN
01-01
                      West Bank and Gaza 31.952200
93762
                 NaN
                                                       35.233200 2021-
01 - 01
                                    Yemen 15.552727
                                                       48.516388 2021-
93763
                 NaN
01 - 01
93764
                 NaN
                                   Zambia -13.133897
                                                       27.849332 2021-
01 - 01
93765
                 NaN
                                 Zimbabwe -19.015438
                                                       29.154857 2021-
01-01
       value
```

U	U
1	0
2	0
3	0
4	0
93761 93762 93763 93764 93765	35 1418 610 390 369

```
[93766 rows x 6 columns]
raw data recovered1['Date'] =
pd.to datetime(raw data recovered1['Date'])
raw data recovered1
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\1902036656.py:1:
UserWarning: Could not infer format, so each element will be parsed
individually, falling back to `dateutil`. To ensure parsing is
consistent and as-expected, please specify a format.
  raw data recovered1['Date'] =
pd.to datetime(raw data recovered1['Date'])
      Province/State
                          Country/Region
                                                 Lat
                                                             Long
Date
                              Afghanistan 33.939110
                                                       67.709953 2020-
                 NaN
01-22
                 NaN
                                  Albania 41.153300
                                                       20.168300 2020-
01 - 22
2
                 NaN
                                  Algeria 28.033900
                                                         1.659600 2020-
01 - 22
3
                 NaN
                                  Andorra 42.506300
                                                         1.521800 2020-
01-22
                                                       17.873900 2020-
4
                 NaN
                                   Angola -11.202700
01 - 22
. . .
                                  Vietnam 14.058324
                                                      108.277199 2021-
88571
                 NaN
01 - 01
88572
                 NaN
                      West Bank and Gaza 31.952200
                                                       35.233200 2021-
01 - 01
88573
                 NaN
                                    Yemen 15.552727
                                                       48.516388 2021-
01-01
                                   Zambia -13.133897
                                                       27.849332 2021-
88574
                 NaN
01-01
                                 Zimbabwe -19.015438
                                                       29.154857 2021-
88575
                 NaN
01 - 01
        value
0
            0
1
            0
2
            0
3
            0
4
            0
88571
         1325
88572
       118926
88573
         1396
88574
        18773
```

```
88575 11347
[88576 rows x 6 columns]
```

Renaming columns of dataframes

```
raw data confirmed1.columns =
raw data confirmed1.columns.str.replace('value','Confirmed')
raw_data_deaths1.columns =
raw data deaths1.columns.str.replace('value','Deaths')
raw data recovered1.columns =
raw data recovered1.columns.str.replace('value','Recovered')
raw data recovered1
      Province/State
                           Country/Region
                                                 Lat
                                                             Long
Date
                              Afghanistan 33.939110
                                                       67.709953 2020-
                 NaN
01 - 22
                 NaN
                                  Albania 41.153300
                                                       20.168300 2020-
1
01 - 22
                 NaN
                                  Algeria 28.033900
                                                         1.659600 2020-
01-22
                 NaN
                                  Andorra 42.506300
                                                         1.521800 2020-
3
01 - 22
                                   Angola -11.202700
                                                        17.873900 2020-
4
                 NaN
01 - 22
. . .
                                  Vietnam 14.058324
                                                      108.277199 2021-
88571
                 NaN
01 - 01
                      West Bank and Gaza 31.952200
88572
                 NaN
                                                        35.233200 2021-
01-01
                                                       48.516388 2021-
88573
                 NaN
                                    Yemen 15.552727
01 - 01
                                                       27.849332 2021-
88574
                 NaN
                                   Zambia -13.133897
01-01
                                 Zimbabwe -19.015438
                                                       29.154857 2021-
88575
                 NaN
01 - 01
       Recovered
0
1
               0
2
               0
3
               0
4
               0
88571
            1325
88572
          118926
88573
            1396
```

```
88574 18773
88575 11347
[88576 rows x 6 columns]
```

Checking null values

```
raw data confirmed1.isnull().sum()
Province/State
                   65048
Country/Region
                       0
                     346
Lat
                     346
Long
Date
                       0
Confirmed
                       0
dtype: int64
raw data deaths1.isnull().sum()
Province/State
                   65048
Country/Region
                       0
                     346
Lat
                     346
Long
Date
                       0
                       0
Deaths
dtype: int64
raw data recovered1.isnull().sum()
Province/State
                   65394
Country/Region
                       0
                       0
Lat
                       0
Long
                       0
Date
                       0
Recovered
dtype: int64
```

Dealing with null values

```
raw_data_confirmed1['Province/
State'].fillna(raw_data_confirmed1['Country/Region'],inplace=True)
raw_data_deaths1['Province/State'].fillna(raw_data_deaths1['Country/
Region'],inplace=True)
raw_data_recovered1['Province/State'].fillna(raw_data_recovered1['Country/Region'],inplace=True)

C:\Users\as355\AppData\Local\Temp\ipykernel_15976\370297445.py:1:
FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values
```

```
always behaves as a copy.
For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.
raw data confirmed1['Province/State'].fillna(raw data confirmed1['Coun
try/Region'],inplace=True)
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\370297445.py:2:
FutureWarning: A value is trying to be set on a copy of a DataFrame or
Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.
For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.
raw data deaths1['Province/State'].fillna(raw data deaths1['Country/
Region'],inplace=True)
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\370297445.py:3:
FutureWarning: A value is trying to be set on a copy of a DataFrame or
Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.
For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.
raw data recovered1['Province/State'].fillna(raw data recovered1['Coun
try/Region'],inplace=True)
raw data recovered1.isnull().sum()
Province/State
                  0
Country/Region
                  0
                  0
Lat
                  0
Long
```

0

Date

```
Recovered
dtype: int64
print("The shape of confirmed1:", raw data confirmed1.shape)
print("The shape of deaths1:", raw data deaths1.shape)
print("The shape of recovered1:", raw data recovered1.shape)
The shape of confirmed1: (93766, 6)
The shape of deaths1: (93766, 6)
The shape of recovered1: (88576, 6)
raw data recovered1
           Province/State
                               Country/Region
                                                      Lat
Long \
                                   Afghanistan 33.939110
                                                            67.709953
0
              Afghanistan
1
                  Albania
                                       Albania 41.153300
                                                            20.168300
2
                  Algeria
                                       Algeria 28.033900
                                                             1.659600
3
                  Andorra
                                       Andorra 42.506300
                                                             1.521800
                   Angola
                                        Angola -11.202700
                                                            17.873900
                                       Vietnam 14.058324
                                                           108.277199
88571
                  Vietnam
88572
       West Bank and Gaza
                           West Bank and Gaza 31.952200
                                                            35.233200
88573
                    Yemen
                                         Yemen 15.552727
                                                            48.516388
88574
                   Zambia
                                        Zambia -13.133897
                                                            27.849332
88575
                 Zimbabwe
                                      Zimbabwe -19.015438
                                                            29.154857
                  Recovered
            Date
0
      2020-01-22
                          0
      2020-01-22
                          0
1
2
                          0
      2020-01-22
3
      2020-01-22
                          0
4
      2020-01-22
                          0
88571 2021-01-01
                       1325
88572 2021-01-01
                     118926
88573 2021-01-01
                       1396
88574 2021-01-01
                      18773
88575 2021-01-01
                      11347
```

Joining Dataframes to form a new dataframe

```
full join =
raw_data_confirmed1.merge(raw_data_deaths1[['Province/State','Country/
Region','Date','Deaths']],how='left',left on=['Province/
State', 'Country/Region', 'Date'], right on=['Province/State', 'Country/
Region','Date'])
full join.head()
  Province/State Country/Region
                                                            Date
                                       Lat
                                                 Long
Confirmed \
     Afghanistan
                    Afghanistan 33.93911 67.709953 2020-01-22
0
1
         Albania
                        Albania 41.15330
                                           20.168300 2020-01-22
0
2
                                             1.659600 2020-01-22
         Algeria
                        Algeria 28.03390
0
3
         Andorra
                        Andorra 42.50630
                                            1.521800 2020-01-22
0
4
          Angola
                         Angola -11.20270 17.873900 2020-01-22
0
   Deaths
0
        0
1
        0
2
        0
3
        0
full_join =
full join.merge(raw data recovered1[['Province/State','Country/Region'
,'Date','Recovered']],how='left',left_on=['Province/State','Country/
Region','Date'],right on=['Province/State','Country/Region','Date'])
full join.head()
  Province/State Country/Region
                                      Lat
                                                 Long
                                                            Date
Confirmed \
0
     Afghanistan
                    Afghanistan 33.93911 67.709953 2020-01-22
0
1
         Albania
                        Albania 41.15330
                                           20.168300 2020-01-22
0
2
                        Algeria 28.03390
                                            1.659600 2020-01-22
         Algeria
0
3
                                            1.521800 2020-01-22
         Andorra
                        Andorra 42.50630
0
4
                         Angola -11.20270 17.873900 2020-01-22
          Angola
0
```

D										
0 1 2	0 0 0	ecovere 0. 0. 0.	0 0 0							
3 4	0 0	0. 0.	_							
full_	join.is	null().	sum()							
Count Lat Long Date Confi Deaths Recove	S	on 5	0 0 346 346 0 0 0 536							
full_	join									
Long	_	rovince	/State	(Count	ry/Region		Lat		
Long 0	\	Afgha	nistan		Afo	ghanistan	33.9	39110	67.709	953
1		А	lbania			Albania	41.1	.53300	20.168	300
2		А	lgeria			Algeria	28.0	33900	1.659	600
3		А	ndorra			Andorra	42.5	06300	1.521	.800
4			Angola			Angola	-11.2	202700	17.873	900
93761		V	'ietnam			Vietnam	14.0	58324	108.277	199
93762	West	Bank an	d Gaza	West	Bank	and Gaza	31.9	52200	35.233	200
93763			Yemen			Yemen	15.5	552727	48.516	388
93764			Zambia			Zambia	-13.1	.33897	27.849	332
93765		Zi	mbabwe			Zimbabwe	-19.0	15438	29.154	857
0 1 2	2020 - 0 2020 - 0 2020 - 0	1-22 1-22		d Dea 0 0 0	aths 0 0	Recovere 0. 0. 0.	0 0			

```
3
      2020-01-22
                                              0.0
      2020-01-22
4
                            0
                                    0
                                              0.0
93761 2021-01-01
                                           1325.0
                        1474
                                   35
93762 2021-01-01
                      139223
                                 1418
                                        118926.0
93763 2021-01-01
                        2101
                                  610
                                           1396.0
93764 2021-01-01
                       20997
                                  390
                                          18773.0
93765 2021-01-01
                       14084
                                  369
                                          11347.0
[93766 rows x 8 columns]
```

Adding a month and year column

```
full join['Month-year'] = full join['Date'].dt.strftime('%b-%Y')
full join.head()
  Province/State Country/Region
                                       Lat
                                                 Long
                                                            Date
Confirmed \
     Afghanistan
                    Afghanistan 33.93911 67.709953 2020-01-22
0
1
         Albania
                        Albania 41.15330 20.168300 2020-01-22
0
2
         Algeria
                        Algeria 28.03390
                                             1.659600 2020-01-22
0
3
         Andorra
                        Andorra 42.50630
                                             1.521800 2020-01-22
0
4
          Angola
                         Angola -11.20270 17.873900 2020-01-22
0
           Recovered Month-year
   Deaths
0
        0
                 0.0
                       Jan-2020
1
        0
                 0.0
                       Jan-2020
2
                       Jan-2020
        0
                 0.0
3
        0
                 0.0
                       Jan-2020
4
        0
                       Jan-2020
                 0.0
```

Creating a new date column - 1 on a new df

```
full join2 = full_join.copy()
full join2['Date-1'] = full join['Date']+pd.Timedelta(days=1)
full_join2.rename(columns={'Confirmed':'Confirmed-1', 'Deaths':'Deaths-
1', 'Recovered': 'Recovered-1', 'Date': 'Date Minus 1'}, inplace=True)
full join2.head()
  Province/State Country/Region
                                       Lat
                                                 Long Date Minus 1 \
     Afghanistan
                    Afghanistan
                                 33.93911
                                            67.709953
                                                        2020-01-22
1
         Albania
                        Albania
                                            20.168300
                                                        2020-01-22
                                 41.15330
```

2 3 4	Algeria Andorra Angola	a A	lgeria 28.0 ndorra 42.5 Angola -11.2	0630 1.52	1800 2020	-01-22 -01-22 -01-22
0 1 2 3 4	Confirmed-1 0 0 0 0 0	Deaths-1 0 0 0 0	Recovered-1 0.0 0.0 0.0 0.0	Jan-2020 Jan-2020 Jan-2020 Jan-2020	Date-1 2020-01-23 2020-01-23 2020-01-23 2020-01-23 2020-01-23	

Creating a yet another new df by joining prev df and recently created df

```
full join3 =
full_join.merge(full_join2[['Province/State','Country/Region','Date
Minus 1', 'Confirmed-1', 'Deaths-1', 'Recovered-1', 'Date-
1']],how='left',left_on=['Province/State','Country/
Region', 'Date'], right_on=['Province/State', 'Country/Region', 'Date-1'])
full join3.head()
  Province/State Country/Region
                                        Lat
                                                  Long
                                                              Date
Confirmed \
     Afghanistan
                     Afghanistan 33.93911 67.709953 2020-01-22
0
1
         Albania
                         Albania
                                  41.15330
                                             20.168300 2020-01-22
0
2
         Algeria
                         Algeria 28.03390
                                              1.659600 2020-01-22
0
3
                         Andorra 42.50630
                                              1.521800 2020-01-22
         Andorra
0
4
                          Angola -11.20270 17.873900 2020-01-22
          Angola
0
   Deaths
           Recovered Month-year Date Minus 1
                                                Confirmed-1
                                                              Deaths-1 \
0
                        Jan-2020
        0
                  0.0
                                           NaT
                                                         NaN
                                                                   NaN
1
        0
                  0.0
                        Jan-2020
                                           NaT
                                                         NaN
                                                                   NaN
2
        0
                        Jan-2020
                  0.0
                                           NaT
                                                         NaN
                                                                   NaN
3
        0
                  0.0
                        Jan-2020
                                           NaT
                                                         NaN
                                                                   NaN
4
        0
                        Jan-2020
                  0.0
                                           NaT
                                                         NaN
                                                                   NaN
   Recovered-1 Date-1
0
           NaN
                   NaT
1
           NaN
                   NaT
2
           NaN
                   NaT
3
           NaN
                   NaT
4
           NaN
                   NaT
```

Creating columns in df3

```
full join3['Confirmed Daily'] = full join3['Confirmed'] -
full join3['Confirmed-1']
full join3['Deaths Daily'] = full join3['Deaths'] -
full ioin3['Deaths-1']
full join3['Recovered Daily'] = full join3['Recovered'] -
full join3['Recovered-1']
full join3.shape
(93766, 17)
full join3.head()
  Province/State Country/Region
                                   Lat
                                                       Date
                                            Long
Confirmed \
                  Afghanistan 33.93911 67.709953 2020-01-22
    Afghanistan
0
1
        Albania
                      Albania 41.15330 20.168300 2020-01-22
0
2
        Algeria
                      Algeria 28.03390
                                        1.659600 2020-01-22
0
3
        Andorra
                      Andorra 42.50630
                                        1.521800 2020-01-22
0
4
         Angola
                       Angola -11.20270 17.873900 2020-01-22
0
  Deaths
          Recovered Month-year Date Minus 1
                                          Confirmed-1
                                                      Deaths-1 \
                     Jan-2020
0
       0
               0.0
                                      NaT
                                                  NaN
                                                           NaN
                     Jan-2020
1
       0
               0.0
                                      NaT
                                                  NaN
                                                           NaN
2
                                                  NaN
       0
               0.0
                     Jan-2020
                                                           NaN
                                      NaT
3
       0
               0.0
                     Jan-2020
                                      NaT
                                                  NaN
                                                           NaN
4
                     Jan-2020
       0
               0.0
                                      NaT
                                                  NaN
                                                           NaN
  Recovered-1 Date-1 Confirmed Daily
                                     Deaths Daily Recovered Daily
0
          NaN
                                NaN
                                             NaN
                                                             NaN
                NaT
          NaN
                NaT
                                NaN
                                             NaN
                                                             NaN
1
2
          NaN
                                NaN
                                             NaN
                                                             NaN
                NaT
3
                                NaN
                                             NaN
                                                             NaN
          NaN
                NaT
                                             NaN
                                                             NaN
          NaN
                NaT
                                NaN
##############################
#################### Braking the numbers by Day
```

```
#############################
#creating a new df
full join2 = full join.copy()
#creating a new date columns - 1
full join2['Date - 1'] = full join2['Date'] + pd.Timedelta(days=1)
full_join2.rename(columns={'Confirmed': 'Confirmed - 1', 'Deaths':
'Deaths - 1', 'Recovered': 'Recovered - 1',
                         'Date': 'Date Minus 1'}, inplace=True)
#Joing on the 2 DFs
full_join3 = full_join.merge(full_join2[['Province/State',
'Country/Region', 'Confirmed - 1', 'Deaths - 1',
                           'Recovered - 1', 'Date - 1', 'Date Minus
1']], how = 'left',
                           left on =
['Province/State','Country/Region','Date'],
                           right on = ['Province/State',
'Country/Region', 'Date - 1'1)
#minus onedf.rename(columns={'Confirmed': 'Confirmed - 1', 'Deaths':
'Deaths - 1', 'Recovered': 'Recovered - 1'}, inplace=True)
full join3.head()
# Additional Calculations
full join3['Confirmed Daily'] = full join3['Confirmed'] -
full join3['Confirmed - 1']
full join3['Deaths Daily'] = full join3['Deaths'] - full join3['Deaths']
- 1']
full join3['Recovered Daily'] = full join3['Recovered'] -
full join3['Recovered - 1']
print(full join3.shape)
(93766, 17)
full join3.head()
  Province/State Country/Region
                                    Lat
                                              Long
                                                        Date
Confirmed \
    Afghanistan
                   Afghanistan 33.93911 67.709953 2020-01-22
0
0
1
        Albania
                       Albania 41.15330 20.168300 2020-01-22
0
2
                       Algeria 28.03390
                                          1.659600 2020-01-22
        Algeria
0
3
                       Andorra 42.50630 1.521800 2020-01-22
        Andorra
```

0											
4	F	Angola	a .	Д	ngola	-11	.20270	17.87	73900 2	020-0	1-22
0											
D	L. I	D		M 1-1-		C -		. 1	D 1	-	D
Dea ⁻	tns	Recov	/erea i	Month	-year	CC	ntirmed) - I	νeatns	- 1	Recovered
0	0		0.0	lan	-2020			NaN		NaN	
NaN	Ū		0.0	Juli	2020			Itali		itait	
1	0		0.0	Jan	-2020			NaN		NaN	
NaN											
2	0		0.0	Jan	-2020			NaN		NaN	
NaN 3	0		0 0	lan	2020			NaN		NaN	
s NaN	0		0.0	Jan	-2020			NaN		NaN	
4	0		0.0	Jan	-2020			NaN		NaN	
NaN			0.0	Juli	2020			Hall		···	
	- 1	Date	Minus	1 (onfirm	ned	Daily	Deaths	Daily	Rec	overed
Daily	N-T		NI.	- T			NI - NI		NI - NI		
0 NaN	NaT		Na	aT			NaN		NaN		
1	NaT		Na	аT			NaN		NaN		
NaN							Hall		itait		
2	NaT		Na	аT			NaN		NaN		
NaN											
3	NaT		Na	аT			NaN		NaN		
NaN	NaT		NI.	ς Т			NaN		NI ~ NI		
4 NaN	NaT		Na	аТ			NaN		NaN		

manually adding values for first date

```
full join3['Confirmed Daily'].loc[full join3['Date']=='2020-01-22'] =
full join3['Confirmed']
full join3['Deaths Daily'].loc[full join3['Date']=='2020-01-22'] =
full join3['Deaths']
full join3['Recovered Daily'].loc[full join3['Date']=='2020-01-22'] =
full join3['Recovered']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:1:
FutureWarning: ChainedAssignmentError: behaviour will change in pandas
3.0!
You are setting values through chained assignment. Currently this
works in certain cases, but when using Copy-on-Write (which will
become the default behaviour in pandas 3.0) this will never work to
update the original DataFrame or Series, because the intermediate
object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a
DataFrame, like:
```

```
df["col"][row indexer] = value
Use `df.loc[row indexer, "col"] = values` instead, to perform the
assignment in a single step and ensure this keeps updating the
original `df`.
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
returning-a-view-versus-a-copy
  full join3['Confirmed Daily'].loc[full join3['Date']=='2020-01-22']
= full join3['Confirmed']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  full_join3['Confirmed Daily'].loc[full_join3['Date']=='2020-01-22']
= full join3['Confirmed']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:2:
FutureWarning: ChainedAssignmentError: behaviour will change in pandas
You are setting values through chained assignment. Currently this
works in certain cases, but when using Copy-on-Write (which will
become the default behaviour in pandas 3.0) this will never work to
update the original DataFrame or Series, because the intermediate
object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a
DataFrame, like:
df["col"][row indexer] = value
Use `df.loc[row indexer, "col"] = values` instead, to perform the
assignment in a single step and ensure this keeps updating the
original `df`.
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
returning-a-view-versus-a-copy
  full join3['Deaths Daily'].loc[full join3['Date']=='2020-01-22'] =
full join3['Deaths']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
```

```
returning-a-view-versus-a-copy
  full join3['Deaths Daily'].loc[full join3['Date']=='2020-01-22'] =
full join3['Deaths']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:3:
FutureWarning: ChainedAssignmentError: behaviour will change in pandas
3.0!
You are setting values through chained assignment. Currently this
works in certain cases, but when using Copy-on-Write (which will
become the default behaviour in pandas 3.0) this will never work to
update the original DataFrame or Series, because the intermediate
object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a
DataFrame, like:
df["col"][row indexer] = value
Use `df.loc[row indexer, "col"] = values` instead, to perform the
assignment in a single step and ensure this keeps updating the
original `df`.
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
returning-a-view-versus-a-copy
  full join3['Recovered Daily'].loc[full join3['Date']=='2020-01-22']
= full join3['Recovered']
C:\Users\as355\AppData\Local\Temp\ipykernel 15976\891303359.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
returning-a-view-versus-a-copy
  full join3['Recovered Daily'].loc[full join3['Date']=='2020-01-22']
= full join3['Recovered']
```

Deleting certain columns

1 Albania Albania 41.153300 20.168300 2 Algeria Algeria 28.033900 1.659600 3 Andorra Andorra 42.506300 1.521800 4 Angola Angola -11.202700 17.873900 93761 Vietnam Vietnam 14.058324 108.277199 93762 West Bank and Gaza West Bank and Gaza 31.952200 35.233200 93763 Yemen Yemen 15.552727 48.516388 93764 Zambia Zambia -13.133897 27.849332 93765 Zimbabwe Zimbabwe -19.015438 29.154857 Date Confirmed Deaths Recovered Month-year Confirmed Daily \ 0 2020-01-22 0 0 0 0.0 Jan-2020 0.0 0.0 Jan-2020 0.0 0 0.0 Jan-2020 0.0 0.0 Jan-2020 0.0 0 0.0 Jan-2020 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0 Jan-2020 0.0 0.0 Jan-2020 0.0 0.0 0.0 Jan									
3 Andorra Andorra 42.506300 1.521800 4 Angola Angola -11.202700 17.873900	1		Albania			Albania	41.153300	20.168300	
4 Angola Angola -11.202700 17.873900	2		Algeria			Algeria	28.033900	1.659600	
	3		Andorra			Andorra	42.506300	1.521800	
93761	4		Angola			Angola	-11.202700	17.873900	
93762 West Bank and Gaza West Bank and Gaza 31.952200 35.233200 93763 Yemen Yemen 15.552727 48.516388 93764 Zambia Zambia -13.133897 27.849332 93765 Zimbabwe Zimbabwe -19.015438 29.154857 Date Confirmed Deaths Recovered Month-year Confirmed Daily \ 0 2020-01-22 0 0 0 0.0 Jan-2020 0.0 1 2020-01-22 0 0 0 0.0 Jan-2020 0.0 2 2020-01-22 0 0 0 0.0 Jan-2020 0.0 2 2020-01-22 0 0 0 0.0 Jan-2020 0.0 3 2020-01-22 0 0 0 0.0 Jan-2020 0.0 4 2020-01-22 0 0 0 0.0 Jan-2020 0.0 4 2020-01-22 0 0 0 0.0 Jan-2020 0.0 9.0 Jan-2020 0.0 9.0 Jan-2020 0.0 9.0 Jan-2020 0.0 0 0 0.0 Jan-2020 0.0 0 0.0 Jan-2020 0.0 0 0 0.0 Jan-2021 0.0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0									
93763 Yemen Yemen 15.552727 48.516388 93764 Zambia Zambia -13.133897 27.849332 93765 Zimbabwe Zimbabwe -19.015438 29.154857 Date Confirmed Deaths Recovered Month-year Confirmed Daily \ 0 2020-01-22 0 0 0 0.0 Jan-2020 0.0 0 0.0 Jan-2021 0.0 0 0.0 Jan-2021 0.0 0.0 0.0 Jan-2021 0.0 0.0 0.0 Jan-2021 0.0 0.0 0.0 Jan-2021 0.0 0.0 0.0 0.0 Jan-2021 0.0 0.0 0.0 0.0 0.0 0.0 Jan-2021 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	93761		Vietnam			Vietnam	14.058324	108.277199	
93764 Zambia Zambia -13.133897 27.849332 93765 Zimbabwe Zimbabwe -19.015438 29.154857 Date Confirmed Deaths Recovered Month-year Confirmed Daily \ 0	93762	West Bank	and Gaza	West	Bank	and Gaza	31.952200	35.233200	
Date Confirmed Deaths Recovered Month-year Confirmed	93763		Yemen			Yemen	15.552727	48.516388	
Date Confirmed Deaths Recovered Month-year Confirmed Daily \ 0	93764		Zambia			Zambia	-13.133897	27.849332	
Daily \ 0	93765		Zimbabwe			Zimbabwe	-19.015438	29.154857	
Daily \ 0									
0 2020-01-22 0 0 0 0.0 Jan-2020 0.0 1 2020-01-22 0 0 0 0.0 Jan-2020 0.0 2 2020-01-22 0 0 0 0.0 Jan-2020 0.0 3 2020-01-22 0 0 0 0.0 Jan-2020 0.0 4 2020-01-22 0 0 0 Jan-2020 0.0 93761 2021-01-01 1474 35 1325.0 Jan-2021 9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 1219.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 2.0 93765 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 277.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	Dailv		e Confirme	d De	aths	Recovered	Month-year	Confirmed	
1 2020-01-22 0 0 0 0.0 Jan-2020 0.0 2 2020-01-22 0 0 0 0.0 Jan-2020 0.0 3 2020-01-22 0 0 0 0.0 Jan-2020 0.0 4 2020-01-22 0 0 0 0.0 Jan-2020 0.0 93761 2021-01-01 1474 35 1325.0 Jan-2021 9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 2.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	0			0	0	0.0	Jan-2020		
2 2020-01-22 0 0 0 0.0 Jan-2020 0.0 3 2020-01-22 0 0 0 0.0 Jan-2020 0.0 4 2020-01-22 0 0 0 0.0 Jan-2020 0.0	1	2020-01-22		0	0	0.0	Jan-2020		
3	2	2020-01-22		0	0	0.0	Jan-2020		
4 2020-01-22 0 0 0 0.0 Jan-2020 0.0 93761 2021-01-01 1474 35 1325.0 Jan-2021 9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 0.0	3	2020-01-22		0	Θ	0.0	Jan-2020		
93761 2021-01-01 1474 35 1325.0 Jan-2021 9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0		2020-01-22		0	0	0.0	Jan-2020		
93761 2021-01-01 1474 35 1325.0 Jan-2021 9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	0.0								
9.0 93762 2021-01-01 139223 1418 118926.0 Jan-2021 1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0		2021-01-01	147	4	35	1325.0	lan-2021		
1219.0 93763 2021-01-01 2101 610 1396.0 Jan-2021 2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	9.0								
2.0 93764 2021-01-01 20997 390 18773.0 Jan-2021 272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	1219.	9							
272.0 93765 2021-01-01 14084 369 11347.0 Jan-2021 217.0 Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	2.0								
Deaths Daily Recovered Daily 0 0.0 0.0 1 0.0 0.0	272.0								
$egin{array}{cccccccccccccccccccccccccccccccccccc$		2021-01-01	. 1408	4	369	11347.0	Jan-2021		
$egin{array}{cccccccccccccccccccccccccccccccccccc$		Deaths Da	ilv Recov	ered	Dailv				
	0 1		0.0		0.0				
2 0.0 0.0	2		0.0		0.0				

3	0.0	0.0
4	0.0	0.0
93761	0.0	0.0
93762	18.0	1743.0
93763	0.0	2.0
93764	2.0	113.0
93765	6.0	97.0
[93766	rows x 12 columns]	

Removing negative values
full join3[full join3['Deaths Daily']<0]

full_join3[full_jo	oin3['Deaths	Daily']<0]				
	ovince/State		Country/	'Region		Lat	
Long \ 14778	Iceland]	celand	64.9	963100	-
19.020800 15653	Philippines		Phili	ppines	12.8	379721	
121.774017 15862	Iceland]	celand	64.9	963100	-
19.020800 16134	India			India	20.	593684	
78.962880 16311	Quebec			Canada	52.9	939900	-
73.549100							
 87140	Ireland]	reland	53.	142400	-
7.692100 87259	Yemen			Yemen	15.	552727	
48.516388 88203	France			France	46.2	227600	
2.213700 88585	Tajikistan		Taji	kistan	38.8	361000	
71.276100	Herzegovina	Bosnia	a and Herze			915900	
17.679100		2002		901=			
Date Daily \	Confirmed [Deaths	Recovered	Month-y	ear	Confir	med
14778 2020-03-16 9.0	180	0	0.0	Mar-2	020		
15653 2020-03-19 15.0	217	17	8.0	Mar-2	020		
15.60 15862 2020-03-20 79.0	409	0	5.0	Mar-2	020		
16134 2020-03-21 86.0	330	4	23.0	Mar-2	020		

16311 38.0	2020-03-22	219	4	NaN	Mar-2020	
87140 214.0	2020-12-08	74682	2097	23364.0	Dec-2020	
87259	2020-12-08	2078	606	1382.0	Dec-2020	-
	2020-12-12	2350793	57210	152555.0	Dec-2020	-
	2020-12-13	12704	88	12133.0	Dec-2020	
80.0 92439 220.0	2020-12-28	109911	3942	76121.0	Dec-2020	
14778 15653 15862 16134 16311 87140 87259 88203 88585 92439	-1.0 -1.0 -1.0 -2.0		ed Daily -8.0 3.0 0.0 3.0 NaN 0.0 -172.0 0.0 83.0 404.0			
[83 rd	ows x 12 colum	ns]				
	join3['Deaths ']< <mark>0</mark> ,0,full_jo				3['Deaths	
<pre>full_join3['Confirmed Daily'] = np.where(full_join3['Confirmed Daily']<0,0,full_join3['Confirmed Daily'])</pre>						
	join3['Recover ']< <mark>0,0</mark> ,full_jo				oin3['Recov	ered
full_	join3					
Long	Province	/State	Countr	y/Region	Lat	
0	•	nistan	Afg	hanistan 3	33.939110	67.709953

Albania

Algeria

Andorra

41.153300

28.033900

42.506300

20.168300

1.659600

1.521800

Albania

Algeria

Andorra

1

2

3

4		Angola			Angola	-11.202700	17.873900	
93761		Vietnam			Vietnam	14.058324	108.277199	
93762	West Bank	and Gaza	West	Bank	and Gaza	31.952200	35.233200	
93763		Yemen			Yemen	15.552727	48.516388	
93764		Zambia			Zambia	-13.133897	27.849332	
93765		Zimbabwe			Zimbabwe	-19.015438	29.154857	
	Date	Confirme	d De	aths	Recovered	l Month-year	Confirmed	
Daily		CONTINUE	u be	a chis	Recovered	i Hollen year	CONTINUE	
0 0.0	2020-01-22		0	0	0.0	Jan-2020		
1 0.0	2020-01-22		0	0	0.0	Jan-2020		
2	2020-01-22		0	0	0.0	Jan-2020		
0.0	2020-01-22		0	0	0.0	Jan-2020		
0.0	2020-01-22		0	0	0.0	Jan-2020		
0.0								
			•					
93761 9.0	2021-01-01	147	4	35	1325.0	Jan-2021		
	2021-01-01	13922	3	1418	118926.0	Jan-2021		
	2021-01-01	210	1	610	1396.0	Jan-2021		
	2021-01-01	2099	7	390	18773.0	Jan-2021		
93765	2021-01-01	1408	4	369	11347.0	Jan-2021		
217.0								
0 1 2 3 4	(ily Recov 0.0 0.0 0.0 0.0 0.0	ered	Daily 0.0 0.0 0.0 0.0				
93761 93762 93763	18	0.0 0.0 0.0 0.0	1	0.0 743.0 2.0				

93764	2.0	113.0
93765	6.0	97.0

[93766 rows x 12 columns]

full_join3.pd.to_csv('C:\Users\as355\Downloads/prepared_Covid_19_date