Gerry William Nemaria / 2440018854 / LB01 / Assignment 01

```
import numpy as np
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
dataset = pd.read_csv('covid_19_indonesia_time_series_all.csv')
```

```
dataset.head()
```

	Date	Location Level	Location ID	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	Total Active Cases	Location	Total Cases per 1Mpop	Total Deaths per 1Mpop	Total Recovered per 1Mpop	Total Active Cases per 1Mpop	Growth Factor of New Cases	Growth Factor of New Deaths
0	2/1/2020	city	0000	0	0	0	0	0	0	0	0	Indonesia	0.0	0.0	0.0	0.0	NaN	NaN
1	2/2/2020	city	0000	0	0	0	0	0	0	0	0	Indonesia	0.0	0.0	0.0	0.0	NaN	NaN
2	2/3/2020	city	0000	0	0	0	0	0	0	0	0	Indonesia	0.0	0.0	0.0	0.0	NaN	NaN
3	2/4/2020	city	0000	0	0	0	0	0	0	0	0	Indonesia	0.0	0.0	0.0	0.0	NaN	NaN
4	2/5/2020	city	0000	0	0	0	0	0	0	0	0	Indonesia	0.0	0.0	0.0	0.0	NaN	NaN

5 rows × 20 columns

```
dataset.shape
```

```
(17156, 20)
```

```
dataset['New Cases'].value_counts()
```

```
0    4127
1    2102
2    1329
3     928
4     716
     ...
2135    1
2136    1
2137    1
2138    1
2139    1
Name: New Cases, Length: 2062, dtype: int64
```

```
dataset['Location'].value_counts()
```

```
Indonesia             340
Jakarta               340
Bali                  340
Jawa Barat            340
Banten                340
West Java             340
Central Java          340
Riau                  340
Kepulauan Riau        340
Jambi                 340
Bangka Belitung       340
East Kalimantan       340
Kalimantan Selatan    340
Kalimantan Timur      340
Lampung               340
Aceh                  340
Kalimantan Tengah     340
Gorontalo             340
Jawa Timur            340
Sumatera Utara        340
Sulawesi Utara        340
Papua                 340
Papua Barat           340
Nusa Tenggara Timur   340
Sulawesi Selatan      340
Sulawesi Barat        340
Sulawesi Tengah       340
Sulawesi Tenggara     340
Maluku                340
Maluku Utara          340
Nusa Tenggara Barat   340
Bengkulu              340
Name: Location, dtype: int64
```

```
dataset['Location ISO Code'].value_counts()
```

```
ID-AE    340
ID-JK    340
ID-BA    340
ID-JB    340
ID-JT    340
ID-JI    340
ID-BT    340
ID-KR    340
ID-RI    340
ID-KepR  340
ID-JA    340
ID-SS    340
ID-KB    340
ID-KT    340
ID-LA    340
ID-SU    340
ID-AC    340
ID-KI    340
ID-KU    340
ID-PA    340
ID-PB    340
ID-NT    340
ID-SN    340
ID-ST    340
ID-SR    340
ID-SG    340
ID-MA    340
ID-MU    340
ID-NB    340
ID-BE    340
Name: Location ISO Code, dtype: int64
```

```
dataset['City or Regency'].value_counts()
```

```
Nirwana], Name City or Regency, dtype: int64
```

```
dataset['Province'].value_counts()
```

```
DKI Jakarta           340
Bali                  340
Jawa Barat            340
Banten                340
Jawa Timur            340
Jawa Tengah           340
Riau                  340
Kepulauan Riau        340
Daerah Istimewa Yogyakarta 340
Jambi                 340
Lampung               340
Kalimantan Tengah     340
Sumatera Utara        340
Sulawesi Utara        340
Papua                 340
Papua Barat           340
Nusa Tenggara Timur   340
Sulawesi Selatan      340
Sulawesi Barat        340
Sulawesi Tengah       340
Sulawesi Tenggara     340
Maluku                340
Maluku Utara          340
Nusa Tenggara Barat   340
Bengkulu              340
Kalimantan Timur      340
Aceh                  340
Gorontalo             340
Name: Province, dtype: int64
```

```
dataset['Country'].value_counts()
```

```
Indonesia    17156
Name: Country, dtype: int64
```

```
dataset['Location Level'].value_counts()
```

```
city    17156
Name: Location Level, dtype: int64
```

```
dataset['Island'].value_counts()
```

```
Sumatera    4420
Jawa        3400
Bali        1360
Kalimantan  3400
Sulawesi    2040
Papua       1360
Maluku      1020
Nusa Tenggara 1020
Name: Island, dtype: int64
```

```
dataset['Special Status'].value_counts()
```

```
Special Province    3400
Special Region      1360
Special Capital     1020
Name: Special Status, dtype: int64
```

```
dataset['Time Zone'].value_counts()
```

```
UTC+07:00    12240
UTC+08:00     3400
UTC+09:00     1360
Name: Time Zone, dtype: int64
```

```
dataset.describe()
```

	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	Total Active Cases	City or Regency ID	Population	Longitude	Latitude	New Cases per 1Mpop	New Deaths per 1Mpop
count	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000
mean	292.350000	13.129441	277.624888	1.575327	3.928904E+04	2844.390952	27126.844144	9959.676802	NaN	24.180256	124.869754	1.223488E01	2.710490	35.369263
std	1874.150176	76.402	1868.924241	187.113566	3.891161E+04	3173.114147	44399.574149	27431.223341	NaN	68.780928	214.109718	9.457000	3.550000	49.941005
min	0.000000	0.000000	0.000000	-3372.000000	1.000000E+00	0.000000	0.000000E+00	-2383.000000	NaN	1.249000	-0.000000	36.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	1.000000E+04	1.000000	1.000000E+03	0.000000	NaN	1.000000	0.000000	1.000000E01	0.000000	0.000000
50%	11.000000	1.000000	31.000000	0.000000	2.270000E+04	1.000000	5.700000E+03	0.000000	NaN	2.000000	101.000000	1.211700E01	2.400700	2.000000
75%	75.000000	5.000000	140.000000	17.000000	3.640000E+04	3.000000	2.350000E+04	1.000000	NaN	10.000000	202.750000	131.000000	3.210537	35.000000
max	34270.000000	2004.000000	48052.000000	6530.000000	1.675700e+06	34500.000000	1.000000e+06	674100.000000	NaN	418.000000	1034.000000	1.941000E01	2.000000	2383.000000

8 rows × 25 columns

```
dataset.isnull()
```

```
Date                   False
Location               False
Location ISO Code      False
Location Level         False
City or Regency        False
Province               False
New Cases              False
New Deaths             False
New Recovered          False
New Active Cases       False
Total Cases            False
Total Deaths           False
Total Recovered        False
Total Active Cases     False
City or Regency ID     False
Island                 False
Country                False
Population             False
Sovereign              False
Time Zone              False
Special Status         False
Total Cases per 1Mpop  False
Total Deaths per 1Mpop False
Total Districts        False
Total Active Villages  False
Total Cases per Village False
Population Density     False
Longitude              False
Latitude               False
New Cases per Million  False
New Deaths per Million False
Total Deaths per Million False
Total Deaths per 100K  False
Growth Factor of New Cases False
Growth Factor of New Deaths False
dtype: bool
```

```
dataset.drop('City or Regency', axis=1, inplace=True)
```

```
dataset.shape
```

```
(17156, 37)
```

```
dataset.isnull.sum()
```

```
Date                   0
Location               0
Location ISO Code      0
Location Level         0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Island                 0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset.drop('Island', axis=1, inplace=True)
```

```
dataset.shape
```

```
(17156, 36)
```

```
dataset.isnull.sum()
```

```
Date                   0
Location               0
Location ISO Code      0
Location Level         0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset.drop('Location', axis=1, inplace=True)
```

```
dataset.shape
```

```
(17156, 35)
```

```
dataset.isnull.sum()
```

```
Date                   0
Location Level         0
Location ISO Code      0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset.drop('Location Level', axis=1, inplace=True)
```

```
dataset.shape
```

```
(17156, 34)
```

```
dataset.isnull.sum()
```

```
Date                   0
Location Level         0
Location ISO Code      0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset.drop('Location Level', inplace=True)
```

```
dataset.shape
```

```
(17156, 33)
```

```
dataset.isnull.sum()
```

```
Date                   0
Location ISO Code      0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset['Location ISO Code'].value_counts()
```

```
ID-AE    340
ID-JK    340
ID-BA    340
ID-JB    340
ID-JT    340
ID-JI    340
ID-BT    340
ID-KR    340
ID-RI    340
ID-KepR  340
ID-JA    340
ID-SS    340
ID-KB    340
ID-KT    340
ID-LA    340
ID-SU    340
ID-AC    340
ID-KI    340
ID-KU    340
ID-PA    340
ID-PB    340
ID-NT    340
ID-SN    340
ID-ST    340
ID-SR    340
ID-SG    340
ID-MA    340
ID-MU    340
ID-NB    340
ID-BE    340
Name: Location ISO Code, dtype: int64
```

```
dataset['Province'].value_counts()
```

```
DKI Jakarta           340
Bali                  340
Jawa Barat            340
Banten                340
Jawa Timur            340
Jawa Tengah           340
Riau                  340
Kepulauan Riau        340
Daerah Istimewa Yogyakarta 340
Jambi                 340
Lampung               340
Kalimantan Tengah     340
Sumatera Utara        340
Sulawesi Utara        340
Papua                 340
Papua Barat           340
Nusa Tenggara Timur   340
Sulawesi Selatan      340
Sulawesi Barat        340
Sulawesi Tengah       340
Sulawesi Tenggara     340
Maluku                340
Maluku Utara          340
Nusa Tenggara Barat   340
Bengkulu              340
Kalimantan Timur      340
Aceh                  340
Gorontalo             340
Name: Province, dtype: int64
```

```
dataset['Country'].value_counts()
```

```
Indonesia    17156
Name: Country, dtype: int64
```

```
dataset['Sovereign'].value_counts()
```

```
Indonesia    17156
Name: Sovereign, dtype: int64
```

```
dataset['Time Zone'].value_counts()
```

```
UTC+07:00    12240
UTC+08:00     3400
UTC+09:00     1360
Name: Time Zone, dtype: int64
```

```
dataset['Special Status'].value_counts()
```

```
Special Province    3400
Special Region      1360
Special Capital     1020
Name: Special Status, dtype: int64
```

```
dataset.describe()
```

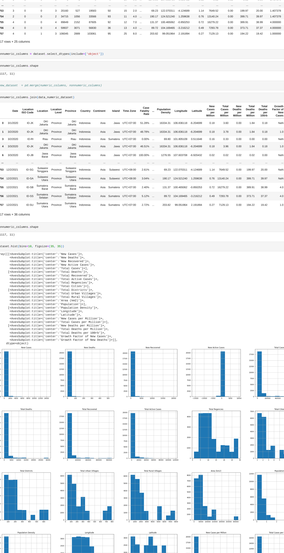
	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	Total Active Cases	City or Regency ID	Population	Longitude	Latitude	New Cases per 1Mpop	New Deaths per 1Mpop
count	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000	17156.000000
mean	292.350000	13.129441	277.624888	1.575327	3.928904E+04	2844.390952	27126.844144	9959.676802	NaN	24.180256	124.869754	1.223488E01	2.710490	35.369263
std	1874.150176	76.402	1868.924241	187.113566	3.891161E+04	3173.114147	44399.574149	27431.223341	NaN	68.780928	214.109718	9.457000	3.550000	49.941005
min	0.000000	0.000000	0.000000	-3372.000000	1.000000E+00	0.000000	0.000000E+00	-2383.000000	NaN	1.249000	-0.000000	36.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	1.000000E+04	1.000000	1.000000E+03	0.000000	NaN	1.000000	0.000000	1.000000E01	0.000000	0.000000
50%	11.000000	1.000000	31.000000	0.000000	2.270000E+04	1.000000	5.700000E+03	0.000000	NaN	2.000000	101.000000	1.211700E01	2.400700	2.000000
75%	75.000000	5.000000	140.000000	17.000000	3.640000E+04	3.000000	2.350000E+04	1.000000	NaN	10.000000	202.750000	131.000000	3.210537	35.000000
max	34270.000000	2004.000000	48052.000000	6530.000000	1.675700e+06	34500.000000	1.000000e+06	674100.000000	NaN	418.000000	1034.000000	1.941000E01	2.000000	2383.000000

8 rows × 25 columns

```
dataset.isnull.sum()
```

```
Date                   0
Location ISO Code      0
New Cases              0
New Deaths             0
New Recovered          0
New Active Cases       0
Total Cases            0
Total Deaths           0
Total Recovered        0
Total Active Cases     0
City or Regency ID     0
Province               0
Country                0
Population             0
Sovereign              0
Time Zone              0
Special Status         0
Total Cases per 1Mpop  0
Total Deaths per 1Mpop 0
Total Districts        0
Total Active Villages  0
Total Cases per Village 0
Population Density     0
Longitude              0
Latitude               0
New Cases per Million  0
New Deaths per Million 0
Total Deaths per Million 0
Total Deaths per 100K  0
Growth Factor of New Cases 0
Growth Factor of New Deaths 0
dtype: int64
```

```
dataset.hist(bins=50, figsize=(40, 30))
```




```
plt.show()
```