In [ ]:

```
import numpy as np
import pandas as pd
import torch
import matplotlib.pyplot as plt
from google.colab import drive
drive.mount('/content/gdrive/')
PATH = 'gdrive/My Drive/'
!ls 'gdrive/My Drive/'
```

```
Drive already mounted at /content/gdrive/; to attempt to forcibly remount, call drive.mount("/content/gdrive/", force_remount=True).
'Colab Notebooks'   cut_Dataset.csv   pa2_data
'comp4211 tuto'     Oxford3000.csv    Title.csv
```

In [ ]:

```
df = pd.read_csv(PATH+"cut_Dataset.csv")
df
```

Out[ ]:

	total_cases	people_fully_vaccinated_per_hundred	Unnamed: 2	standard_vaccination_rate	string
0	1.0	NaN	NaN	NaN	
1	1.0	NaN	NaN	NaN	
2	1.0	NaN	NaN	NaN	
3	1.0	NaN	NaN	NaN	
4	1.0	NaN	NaN	NaN	
...	...	...	...	...	
99592	81168.0	4.28	NaN	2573.0	
99593	82613.0	4.30	NaN	2823.0	
99594	83619.0	4.31	NaN	2827.0	
99595	85732.0	4.33	NaN	2843.0	
99596	88415.0	NaN	NaN	NaN	

99597 rows × 21 columns

In [ ]:

```
df.describe()
```

Out[ ]:

	total_cases	people_fully_vaccinated_per_hundred	Unnamed: 2	standard_vaccination_rate	stri
count	9.549500e+04	13525.000000	0.0	31233.000000	
mean	3.731543e+05	12.456047	NaN	3389.378766	
std	1.949306e+06	16.024510	NaN	4599.277888	
min	1.000000e+00	0.000000	NaN	0.000000	
25%	1.242000e+03	1.390000	NaN	400.000000	
50%	1.297000e+04	5.700000	NaN	1803.000000	
75%	1.296355e+05	17.480000	NaN	4928.000000	

	total_cases	people_fully_vaccinated_per_hundred	Unnamed: 2	standard_vaccination_rate	stri
max	3.417477e+07	115.840000	NaN	118759.000000	

In [ ]:
```
df.isna().sum() #check missing
```

Out[ ]:
```
total_cases                          4102
people_fully_vaccinated_per_hundred    86072
Unnamed: 2                           99597
standard_vaccination_rate            68364
stringency_index                     12397
population                             695
population_density                    3690
median_age                            7517
aged_65_older                         8555
aged_70_older                         8028
gdp_per_capita                        7137
extreme_poverty                      37473
cardiovasc_death_rate                 7170
diabetes_prevalence                   4675
female_smokers                       27446
male_smokers                         28505
handwashing_facilities               53440
hospital_beds_per_thousand           15482
life_expectancy                       1497
human_development_index               7045
excess_mortality                     95933
dtype: int64
```

In [ ]:
```
def normalize(df): # Standardization function
    result = df.copy()
    for feature_name in df.columns:
        result[feature_name] = (df[feature_name] -df[feature_name].min())/(df[feature_
    return result
```

In [ ]:
```
std_input = normalize(df)
std_input
```

Out[ ]:

	total_cases	people_fully_vaccinated_per_hundred	Unnamed: 2	standard_vaccination_rate	string
0	0.000000	NaN	NaN	NaN	
1	0.000000	NaN	NaN	NaN	
2	0.000000	NaN	NaN	NaN	
3	0.000000	NaN	NaN	NaN	
4	0.000000	NaN	NaN	NaN	
...	...	...	...	...	
99592	0.002375	0.036948	NaN	0.021666	
99593	0.002417	0.037120	NaN	0.023771	
99594	0.002447	0.037206	NaN	0.023805	

	total_cases	people_fully_vaccinated_per_hundred	Unnamed: 2	standard_vaccination_rate	string
99595	0.002509	0.037379	NaN	0.023939	
99596	0.002587	NaN	NaN	NaN	

99597 rows × 21 columns

In [ ]:

```
std_input.to_csv(PATH+"new_std.csv")
```