**Cause of deaths dataset**

By looing the dataset two column are object data type , we have to delete them one of them

Year is column which is having value from 1990 to 2019 and they are repetitive in nature so they are also similar to somewhat label , so either we can sum them like 30 years of otherwise delete them. Converting the year to sum of 30 years of each country will make the whole data set shrink to 204 X 34.

No target column here which we need to predict so this problem comes under unsupervised machine learning .

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 6120 entries, 0 to 6119

Data columns (total 34 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Country/Territory 6120 non-null object

1 Code 6120 non-null object

2 Year 6120 non-null int64

3 Meningitis 6120 non-null int64

4 Alzheimer's Disease and Other Dementias 6120 non-null int64

5 Parkinson's Disease 6120 non-null int64

6 Nutritional Deficiencies 6120 non-null int64

7 Malaria 6120 non-null int64

8 Drowning 6120 non-null int64

9 Interpersonal Violence 6120 non-null int64

10 Maternal Disorders 6120 non-null int64

11 HIV/AIDS 6120 non-null int64

12 Drug Use Disorders 6120 non-null int64

13 Tuberculosis 6120 non-null int64

14 Cardiovascular Diseases 6120 non-null int64

15 Lower Respiratory Infections 6120 non-null int64

16 Neonatal Disorders 6120 non-null int64

17 Alcohol Use Disorders 6120 non-null int64

18 Self-harm 6120 non-null int64

19 Exposure to Forces of Nature 6120 non-null int64

20 Diarrheal Diseases 6120 non-null int64

21 Environmental Heat and Cold Exposure 6120 non-null int64

22 Neoplasms 6120 non-null int64

23 Conflict and Terrorism 6120 non-null int64

24 Diabetes Mellitus 6120 non-null int64

25 Chronic Kidney Disease 6120 non-null int64

26 Poisonings 6120 non-null int64

27 Protein-Energy Malnutrition 6120 non-null int64

28 Road Injuries 6120 non-null int64

29 Chronic Respiratory Diseases 6120 non-null int64

30 Cirrhosis and Other Chronic Liver Diseases 6120 non-null int64

31 Digestive Diseases 6120 non-null int64

32 Fire, Heat, and Hot Substances 6120 non-null int64

33 Acute Hepatitis 6120 non-null int64

dtypes: int64(32), object(2)

memory usage: 1.6+ MB

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 6120 entries, 0 to 6119

Data columns (total 31 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Meningitis 6120 non-null int64

1 Alzheimer's Disease and Other Dementias 6120 non-null int64

2 Parkinson's Disease 6120 non-null int64

3 Nutritional Deficiencies 6120 non-null int64

4 Malaria 6120 non-null int64

5 Drowning 6120 non-null int64

6 Interpersonal Violence 6120 non-null int64

7 Maternal Disorders 6120 non-null int64

8 HIV/AIDS 6120 non-null int64

9 Drug Use Disorders 6120 non-null int64

10 Tuberculosis 6120 non-null int64

11 Cardiovascular Diseases 6120 non-null int64

12 Lower Respiratory Infections 6120 non-null int64

13 Neonatal Disorders 6120 non-null int64

14 Alcohol Use Disorders 6120 non-null int64

15 Self-harm 6120 non-null int64

16 Exposure to Forces of Nature 6120 non-null int64

17 Diarrheal Diseases 6120 non-null int64

18 Environmental Heat and Cold Exposure 6120 non-null int64

19 Neoplasms 6120 non-null int64

20 Conflict and Terrorism 6120 non-null int64

21 Diabetes Mellitus 6120 non-null int64

22 Chronic Kidney Disease 6120 non-null int64

23 Poisonings 6120 non-null int64

24 Protein-Energy Malnutrition 6120 non-null int64

25 Road Injuries 6120 non-null int64

26 Chronic Respiratory Diseases 6120 non-null int64

27 Cirrhosis and Other Chronic Liver Diseases 6120 non-null int64

28 Digestive Diseases 6120 non-null int64

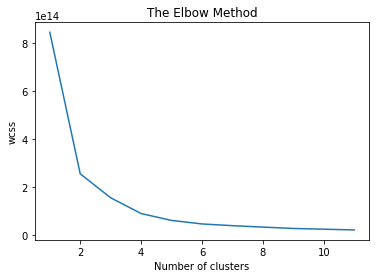
29 Fire, Heat, and Hot Substances 6120 non-null int64

30 Acute Hepatitis 6120 non-null int64

dtypes: int64(31)

memory usage: 1.4 MB

By applying the Elbow method we get the silhouette score nearly to 0.793 which is the good one



silhouette\_score(x,y\_kmeans)

0.7930635535847472