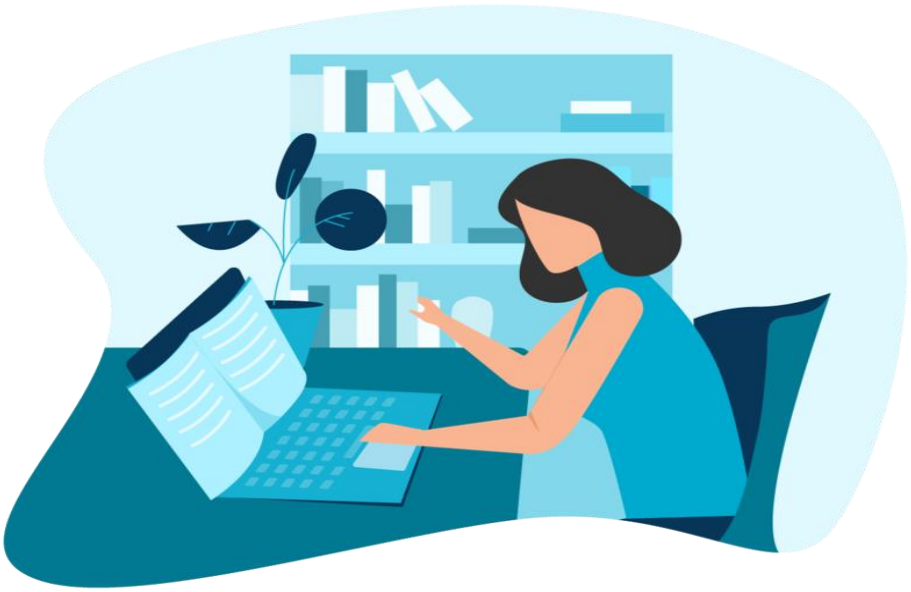
FLIP ROBO

Project Report

Case of Death



Prepared by
Deepika
SME. Khushboo Garg

Acknowledgement

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly thankful to (**flip Robo technologies**) for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

And also thank to **YouTube ,Google** from where I learned a lot

My thanks go to my **SME. Khushboo Garg,** for valuable suggestion provide throughout.

About Dataset

Context

A straightforward way to assess the health status of a population is to focus on mortality – or concepts like child mortality or life expectancy, which are based on mortality estimates. A focus on mortality, however, does not take into account that the burden of diseases is not only that they kill people, but that they cause suffering to people who live with them. Assessing health outcomes by both mortality and morbidity (the prevalent diseases) provides a more encompassing view on health outcomes. This is the topic of this entry. The sum of mortality and morbidity is referred to as the 'burden of disease' and can be measured by a metric called 'Disability Adjusted Life Years' (DALYs). DALYs are measuring lost health and are a standardized metric that allow for direct comparisons of disease burdens of different diseases across countries, between different populations, and over time. Conceptually, one DALY is the equivalent of losing one year in good health because of either premature death or disease or disability. One DALY represents one lost year of healthy life. The first 'Global Burden of Disease' (GBD) was GBD 1990 and the DALY metric was prominently featured in the World Bank's 1993 World Development Report. Today it is published by both the researchers at the Institute of Health Metrics and Evaluation (IHME) and the 'Disease Burden Unit' at the World Health Organization (WHO), which was created in 1998. The IHME continues the work that was started in the early 1990s and publishes the Global Burden of Disease study.

Content

In this Dataset, we have Historical Data of different cause of deaths for all ages around the World. The key features of this Dataset are: Meningitis, Alzheimer's Disease and Other Dementias, Parkinson's Disease, Nutritional Deficiencies, Malaria, Drowning, Interpersonal Violence, Maternal Disorders, HIV/AIDS, Drug Use Disorders, Tuberculosis, Cardiovascular Diseases, Lower Respiratory Infections, Neonatal Disorders, Alcohol Use Disorders, Self-harm, Exposure to Forces of Nature, Diarrheal Diseases, Environmental Heat and Cold Exposure, Neoplasms, Conflict and Terrorism, Diabetes Mellitus, Chronic Kidney Disease, Poisonings, Protein-Energy Malnutrition, Road Injuries, Chronic Respiratory Diseases, Cirrhosis and Other Chronic Liver Diseases, Digestive Diseases, Fire, Heat, and Hot Substances, Acute Hepatitis.

Dataset Glossary (Column-wise)

- 01. Country/Territory - Name of the Country/Territory
- 02. Code - Country/Territory Code
- 03. Year - Year of the Incident
- 04. Meningitis - No. of People died from Meningitis
- 05. Alzheimer's Disease and Other Dementias - No. of People died from Alzheimer's Disease and Other Dementias
- 06. Parkinson's Disease - No. of People died from Parkinson's Disease
- 07. Nutritional Deficiencies - No. of People died from Nutritional Deficiencies
- 08. Malaria - No. of People died from Malaria
- 09. Drowning - No. of People died from Drowning
- 10. Interpersonal Violence - No. of People died from Interpersonal Violence
- 11. Maternal Disorders - No. of People died from Maternal Disorders

- 12. Drug Use Disorders - No. of People died from Drug Use Disorders
- 13. Tuberculosis - No. of People died from Tuberculosis
- 14. Cardiovascular Diseases - No. of People died from Cardiovascular Diseases
- 15. Lower Respiratory Infections - No. of People died from Lower Respiratory Infections
- 16. Neonatal Disorders - No. of People died from Neonatal Disorders
- 17. Alcohol Use Disorders - No. of People died from Alcohol Use Disorders
- 18. Self-harm - No. of People died from Self-harm
- 19. Exposure to Forces of Nature - No. of People died from Exposure to Forces of Nature
- 20. Diarrheal Diseases - No. of People died from Diarrheal Diseases
- 21. Environmental Heat and Cold Exposure - No. of People died from Environmental Heat and Cold Exposure
- 22. Neoplasms - No. of People died from Neoplasms
- 23. Conflict and Terrorism - No. of People died from Conflict and Terrorism
- 24. Diabetes Mellitus - No. of People died from Diabetes Mellitus
- 25. Chronic Kidney Disease - No. of People died from Chronic Kidney Disease
- 26. Poisonings - No. of People died from Poisoning
- 27. Protein-Energy Malnutrition - No. of People died from Protein-Energy Malnutrition
- 28. Chronic Respiratory Diseases - No. of People died from Chronic Respiratory Diseases
- 29. Cirrhosis and Other Chronic Liver Diseases - No. of People died from Cirrhosis and Other Chronic Liver Diseases
- 30. Digestive Diseases - No. of People died from Digestive Diseases
- 31. Fire, Heat, and Hot Substances - No. of People died from Fire or Heat or any Hot Substances
- 32. Acute Hepatitis - No. of People died from Acute Hepatitis

Data Analysis :-

- We have use Jupiter Notebook and many libraries such as pandas ,Numpy for loaded data and fill the data for analysis

- Visualization part we use seaborn and matplotlib for plot chart and visualizes it.

Importing the libraries of the python

```
In [1]: #Importing the important packages  and libraries of the python.

        import pandas as pd
        import numpy as np
        import seaborn as sns
        import warnings
        warnings.filterwarnings('ignore')

        # data visualization
        import matplotlib.pyplot as plt
        import seaborn as sns

        import warnings
        warnings.filterwarnings('ignore')
        %matplotlib inline
```

Dataset:-

```
In [2]: df= pd.read_csv("cause_of_deaths dataset.csv")
        df
```

Out[2]:

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Ma
0	Afghanistan	AFG	1990	2159	1116	371	2087	93	1370	1538	...	2108	3709	338	
1	Afghanistan	AFG	1991	2218	1136	374	2153	189	1391	2001	...	2120	3724	351	
2	Afghanistan	AFG	1992	2475	1162	378	2441	239	1514	2299	...	2153	3776	386	
3	Afghanistan	AFG	1993	2812	1187	384	2837	108	1687	2589	...	2195	3862	425	
4	Afghanistan	AFG	1994	3027	1211	391	3081	211	1809	2849	...	2231	3932	451	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
6115	Zimbabwe	ZWE	2015	1439	754	215	3019	2518	770	1302	...	3176	2108	381	
6116	Zimbabwe	ZWE	2016	1457	767	219	3056	2050	801	1342	...	3259	2160	393	
6117	Zimbabwe	ZWE	2017	1460	781	223	2990	2116	818	1363	...	3313	2196	398	
6118	Zimbabwe	ZWE	2018	1450	795	227	2918	2088	825	1396	...	3381	2240	400	

```
In [8]: Death_cause.shape
Out[8]: (6120, 34)
```

There are total 6120 values and 34 columns in this data there are two columns are same country and code.

Unusual value present in the dataset:-

```
In [4]: Death_cause.tail()
```

Out[4]:

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	M
6115	Zimbabwe	ZWE	2015	1439	754	215	3019	2518	770	1302	...	3176	2108	381	
6116	Zimbabwe	ZWE	2016	1457	767	219	3056	2050	801	1342	...	3259	2160	393	
6117	Zimbabwe	ZWE	2017	1460	781	223	2990	2116	818	1363	...	3313	2196	398	
6118	Zimbabwe	ZWE	2018	1450	795	227	2918	2088	825	1396	...	3381	2240	400	
6119	Zimbabwe	ZWE	2019	1450	812	232	2884	2068	827	1434	...	3460	2292	405	

5 rows × 34 columns

```
In [7]: Death_cause.sample(5)
```

Out[7]:

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Ma
4321	Poland	POL	1991	492	8659	1981	51	0	1843	1436	...	6347	4571	1062	
4600	Saint Vincent and the Grenadines	VCT	2000	3	15	4	8	0	8	15	...	78	22	0	
640	Bhutan	BTN	2000	68	31	11	35	105	37	12	...	64	79	4	
3522	Mongolia	MNG	1999	211	139	43	26	0	161	305	...	80	472	233	
4374	Portugal	PRT	2014	45	5542	1199	144	0	101	171	...	3928	3660	25	

5 rows × 34 columns

```
In [8]: Death_cause.shape
```

Out[8]: (6120, 34)

Checking out presented columns in dataset:-

```
In [11]: #identify which types of data they all belongs

         Death_cause.dtypes
```

```
Out[11]: Country/Territory                          object
         Code                                       object
         Year                                        int64
         Meningitis                                  int64
         Alzheimer's Disease and Other Dementias     int64
         Parkinson's Disease                         int64
         Nutritional Deficiencies                    int64
         Malaria                                     int64
         Drowning                                    int64
         Interpersonal Violence                      int64
         Maternal Disorders                          int64
         HIV/AIDS                                    int64
         Drug Use Disorders                          int64
         Tuberculosis                                int64
         Cardiovascular Diseases                     int64
         Lower Respiratory Infections                int64
         Neonatal Disorders                          int64
         Alcohol Use Disorders                       int64
         Self-harm                                   int64
         Exposure to Forces of Nature                int64
         Diarrheal Diseases                          int64
         Environmental Heat and Cold Exposure        int64
         Neoplasms                                   int64
         Conflict and Terrorism                      int64
         Diabetes Mellitus                           int64
         Chronic Kidney Disease                      int64
         Poisonings                                  int64
         Protein-Energy Malnutrition                 int64
         Road Injuries                               int64
         Chronic Respiratory Diseases                int64
```

we can see there are only 2 numerical columns in whole dataset contains then other categorical columns

EDA:-

Checking the information of the dataset and count the column

```
In [12]: Death_cause.info()

<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
```

Described here about the columns name null value dtypes of columns and memory usage.Here are two types data in dataset int64, only two columns are obejective then other. we count of every column are equal so there are no nan present in dataset,

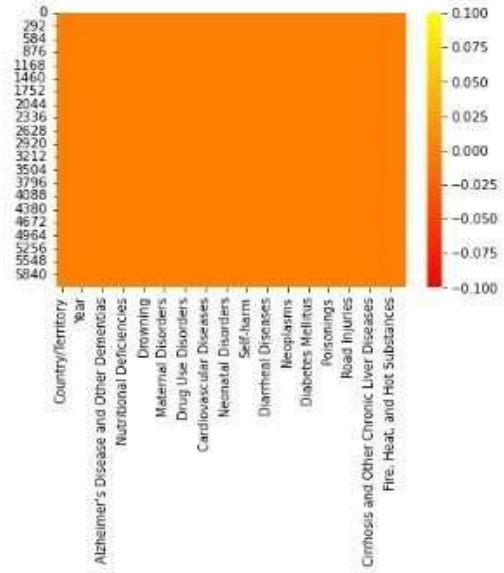
check the Null value:-

```
In [13]: Death_cause.isnull().sum()

Out[13]: Country/Territory                          0
         Code                                       0
         Year                                       0
         Meningitis                                 0
         Alzheimer's Disease and Other Dementias    0
         Parkinson's Disease                        0
         Nutritional Deficiencies                   0
         Malaria                                    0
         Drowning                                   0
         Interpersonal Violence                     0
         Maternal Disorders                         0
         HIV/AIDS                                   0
         Drug Use Disorders                         0
         Tuberculosis                               0
         Cardiovascular Diseases                    0
         Lower Respiratory Infections               0
         Neonatal Disorders                         0
         Alcohol Use Disorders                      0
         Self-harm                                  0
         Exposure to Forces of Nature               0
         Diarrheal Diseases                         0
         Environmental Heat and Cold Exposure       0
         Neoplasms                                  0
         Conflict and Terrorism                     0
         Diabetes Mellitus                          0
         Chronic Kidney Disease                     0
         Poisonings                                 0
         Protein-Energy Malnutrition                0
         Road Injuries                              0
         Chronic Respiratory Diseases               0
         Cirrhosis and Other Chronic Liver Diseases 0
         Digestive Diseases                         0
         Fire, Heat, and Hot Substances             0
         Acute Hepatitis                            0
         dtype: int64
```

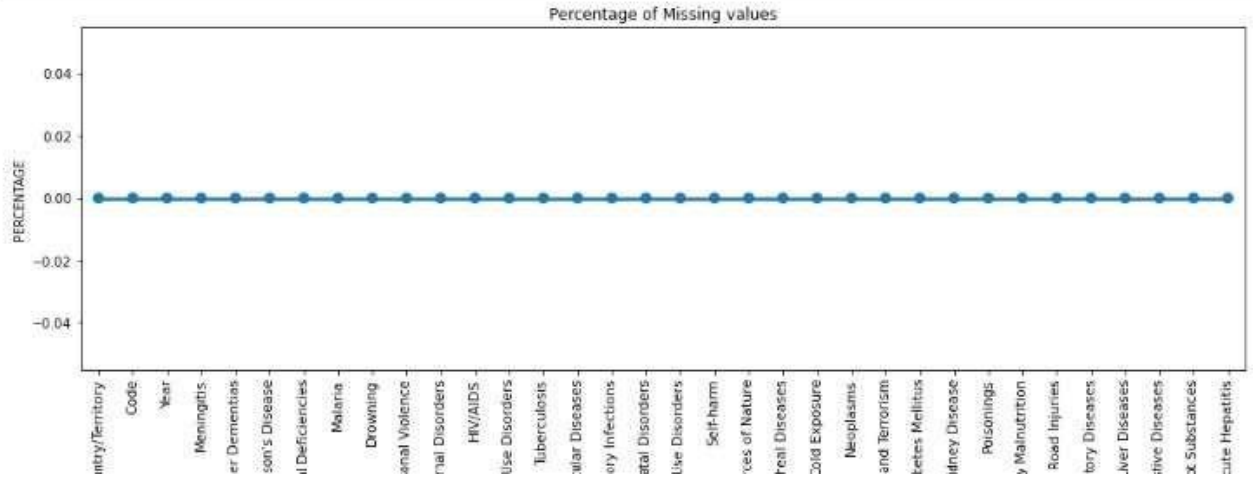
we can see here the nan value is 0 in every column

```
# Let's visualize NaN values
sns.heatmap(df.isnull(),cmap="autumn")
plt.show()
```



dataset is free from nan value

```
missing = pd.DataFrame((df.isnull().sum())*100/df.shape[0]).reset_index()
plt.figure(figsize=(16,5))
ax = sns.pointplot('index',0,data=missing)
plt.xticks(rotation =90,fontsize =11)
plt.title("Percentage of Missing values")
plt.ylabel("PERCENTAGE")
plt.show()
```



we can see Straight line it means here no column have nan value

Separating the Categorical and Numerical Column:-

isolating the categorcal and numerical columns

In [23]:
```
# Counting the categorical columns
categorical_col=[]
for i in df.dtypes.index:
    if df.dtypes[i]=='object':
        categorical_col.append(i)
print("Categorical columns are:\n",categorical_col)
```

```
Categorical columns are:
 ['Country/Territory', 'Code']
```

two types columns are only categorical in dataset

In [24]:
```
# checking  the numerical columns
numerical_col=[]
for i in df.dtypes.index:
    if df.dtypes[i]!='object':
        numerical_col.append(i)
print("Numerical columns are:\n",numerical_col)
```

```
Numerical columns are:
 ['Year', 'Meningitis', "Alzheimer's Disease and Other Dementias", "Parkinson's Disease", 'Nutritional Deficiencies', 'Malari
a', 'Drowning', 'Interpersonal Violence', 'Maternal Disorders', 'HIV/AIDS', 'Drug Use Disorders', 'Tuberculosis', 'Cardiovascul
ar Diseases', 'Lower Respiratory Infections', 'Neonatal Disorders', 'Alcohol Use Disorders', 'Self-harm', 'Exposure to Forces o
f Nature', 'Diarrheal Diseases', 'Environmental Heat and Cold Exposure', 'Neoplasms', 'Conflict and Terrorism', 'Diabetes Melli
tus', 'Chronic Kidney Disease', 'Poisonings', 'Protein-Energy Malnutrition', 'Road Injuries', 'Chronic Respiratory Diseases',
'Cirrhosis and Other Chronic Liver Diseases', 'Digestive Diseases', 'Fire, Heat, and Hot Substances', 'Acute Hepatitis']
```

these numerical column of dataset

We can see above that we have 2 categorical columns out of 34 and rest all 32 columns numerical .

Describe the Dataset:-

```
In [18]: dfcor=df.corr()
         dfcor
```

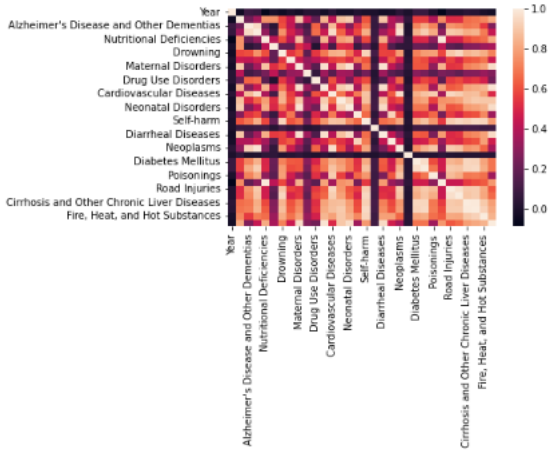
Out[18]:

	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	HIV/AIDS	...	Diabetes Mellitus	Chronic Kidney Disease	P
Year	1.000000	-0.043288	0.083710	0.068756	-0.078266	-0.015964	-0.040910	-0.001122	-0.027460	0.022964	...	0.074292	0.074970	
Meningitis	-0.043288	1.000000	0.216713	0.351668	0.760851	0.755261	0.576347	0.447242	0.899507	0.411881	...	0.561177	0.562013	
Alzheimer's Disease and Other Dementias	0.083710	0.216713	1.000000	0.950785	0.193209	0.031290	0.599403	0.429622	0.200315	0.101628	...	0.723417	0.814841	
Parkinson's Disease	0.068756	0.351668	0.950785	1.000000	0.313033	0.084109	0.753663	0.485528	0.356394	0.145521	...	0.835941	0.911376	
Nutritional Deficiencies	-0.078266	0.760851	0.193209	0.313033	1.000000	0.411149	0.596367	0.407065	0.881740	0.241247	...	0.495049	0.504551	
Malaria	-0.015964	0.755261	0.031290	0.084109	0.411149	1.000000	0.195839	0.184469	0.523581	0.424471	...	0.205493	0.198686	
Drowning	-0.040910	0.576347	0.599403	0.753663	0.596367	0.195839	1.000000	0.539339	0.623558	0.171108	...	0.714079	0.781125	
Interpersonal Violence	-0.001122	0.447242	0.429622	0.485528	0.407065	0.184469	0.539339	1.000000	0.449551	0.315642	...	0.621728	0.589217	
Maternal Disorders	-0.027460	0.899507	0.200315	0.356394	0.881740	0.523581	0.623558	0.449551	1.000000	0.336267	...	0.617109	0.611581	
HIV/AIDS	0.022964	0.411881	0.101628	0.145521	0.241247	0.424471	0.171108	0.315642	0.336267	1.000000	...	0.289675	0.245579	
Drug Use Disorders	0.023917	0.187050	0.641341	0.664385	0.163638	0.011560	0.545004	0.403071	0.162381	0.059646	...	0.492719	0.578125	
Tuberculosis	-0.025297	0.844494	0.273336	0.445504	0.844321	0.423077	0.692165	0.478608	0.968400	0.337561	...	0.714916	0.696064	
Cardiovascular Diseases	0.029813	0.411787	0.860759	0.956667	0.370829	0.107651	0.829424	0.554629	0.420706	0.162285	...	0.826078	0.883884	
Lower Respiratory Infections	-0.027531	0.879827	0.503257	0.638771	0.783957	0.520901	0.840597	0.583966	0.889648	0.354663	...	0.772023	0.803253	
Neonatal Disorders	-0.026949	0.908737	0.270157	0.435508	0.824924	0.501561	0.684802	0.512989	0.970727	0.340600	...	0.688876	0.685575	
Alcohol Use Disorders	0.011315	0.275909	0.529750	0.584208	0.256448	0.073306	0.497126	0.687922	0.300051	0.128654	...	0.526259	0.540952	
Self-harm	-0.004192	0.609952	0.682463	0.826083	0.581602	0.202015	0.929865	0.611616	0.676702	0.234264	...	0.820802	0.879696	
Exposure to Forces of Nature	-0.005178	0.071674	0.066285	0.080246	0.091046	0.025273	0.104622	0.052323	0.094510	0.017443	...	0.098244	0.095391	
Diarrheal Diseases	-0.031911	0.892564	0.189659	0.347421	0.829348	0.551661	0.573699	0.432915	0.972465	0.340339	...	0.629377	0.613615	
Environmental Heat and Cold Exposure	-0.017286	0.390214	0.384578	0.469845	0.356665	0.144287	0.566503	0.623965	0.424968	0.163466	...	0.455900	0.479516	
Neoplasms	0.036753	0.299265	0.925096	0.975203	0.266367	0.051927	0.782617	0.460441	0.277511	0.117547	...	0.752242	0.841755	
Conflict and Terrorism	-0.005941	0.052123	-0.004190	0.000944	0.041965	0.043511	0.012436	0.023094	0.045912	0.021714	...	0.014463	0.014968	

Here we have described the whole dataset by described command .
There count of all the columns that is 6120 which means no Null value is present in the dataset.
All the columns the Min and Max , and the Quartiles over here too
Here we can see the mean and standard deviation of all the Numeric columns in the dataset.

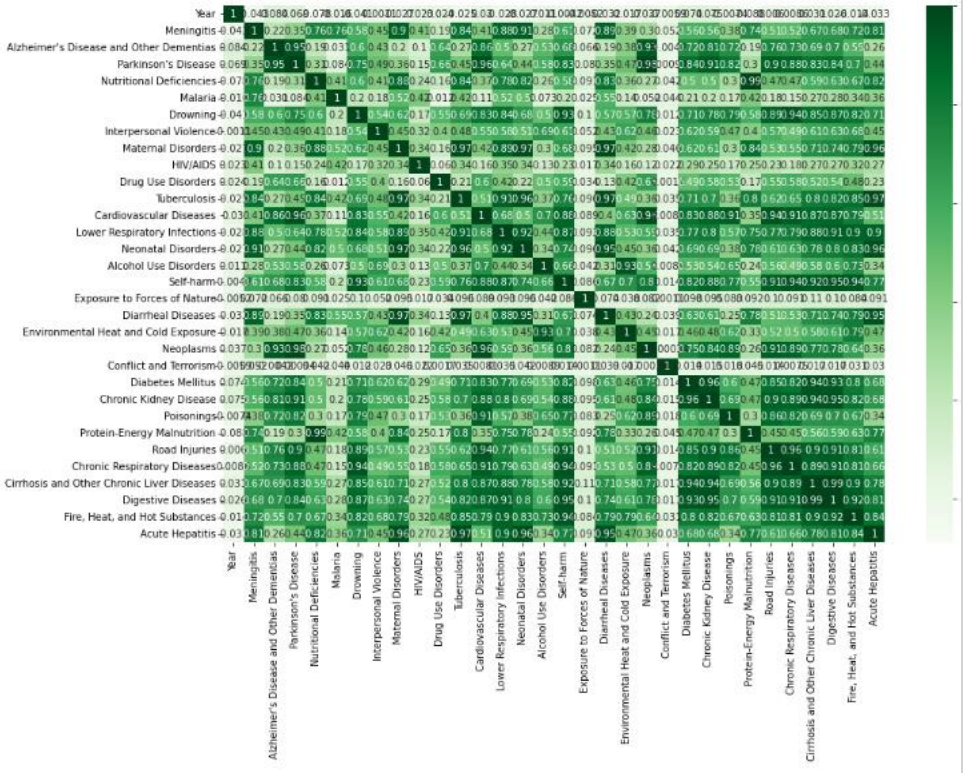
```
In [19]: sns.heatmap(dfcor)
```

Out[19]: <AxesSubplot:>



```
In [20]: plt.figure(figsize=(15,10))
         sns.heatmap(dfcor,cmap='Greens', annot=True)
```

```
Out[20]: <AxesSubplot:>
```



Feature Engineering:-

Feature Engineering

```
In [33]: e_of_deaths=['Meningitis','Alzheimer\'s Disease and Other Dementias','Parkinson\'s Disease','Nutritional Deficiencies',
                      'Malaria','Drowning','Interpersonal Violence','Maternal Disorders','HIV/AIDS','Drug Use Disorders',
                      'Tuberculosis','Cardiovascular Diseases','Lower Respiratory Infections','Neonatal Disorders',
                      'Alcohol Use Disorders','Self-harm','Exposure to Forces of Nature','Diarrheal Diseases',
                      'Environmental Heat and Cold Exposure','Neoplasms','Conflict and Terrorism','Diabetes Mellitus',
                      'Chronic Kidney Disease','Poisonings','Protein-Energy Malnutrition','Road Injuries','Chronic Respiratory Diseases',
                      'Cirrhosis and Other Chronic Liver Diseases','Digestive Diseases','Fire, Heat, and Hot Substances','Acute Hepatitis',
                      ]
```

```
In [34]: #Creating a new column for 'Total no of Deaths' for individual Country and Year
```

In [34]:
```
#Creating a new column for 'Total_no_of_Deaths' for individual Country and Year
df['Total_no_of_Deaths']=df[cause_of_deaths].sum(axis=1)
```

In [35]:
```
# Top 10 Total_no_of_Deaths
top10_Total_no_of_Deaths=df.sort_values(by='Total_no_of_Deaths',ascending=False)[:10][['Total_no_of_Deaths','Country/Territory']]
top10_Total_no_of_Deaths
```

Out[35]:

	Total_no_of_Deaths	Country/Territory
1139	10442561	China
1138	10163943	China
1137	9978653	China
1119	9814213	China
1118	9591222	China
1117	9503904	China
1116	9411928	China
1114	9366974	China
1115	9364587	China
1113	9284664	China

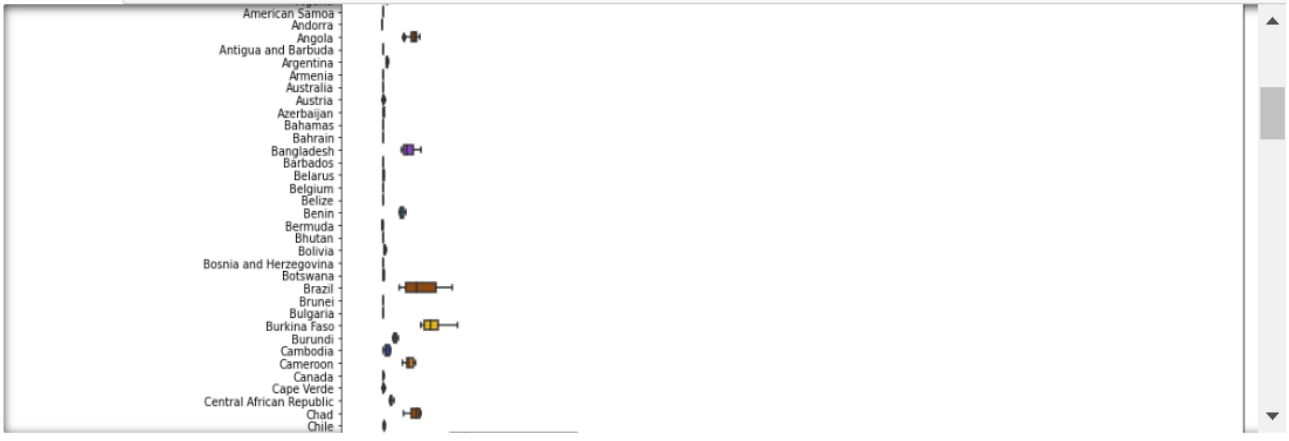
From the Above table, the "top 10-Total_no_of_Deaths" belongs to 'China' and the "top60-Total_no_of_Deaths" belongs to 'China' and 'India' combined, followed by 'USA' and 'RUSSIA'. This because of the fact "China" and "India" are the countries that stand in top 2 in terms of population

Visulation :-

In [36]:
```
# Comparing the next day latitude & temperature

plt.figure(figsize=(15,40))
plt.suptitle('DEATH BY MENINGITIS',fontsize=20)

sns.boxplot(x='Meningitis',y='Country/Territory',data=df,palette="bright")
```

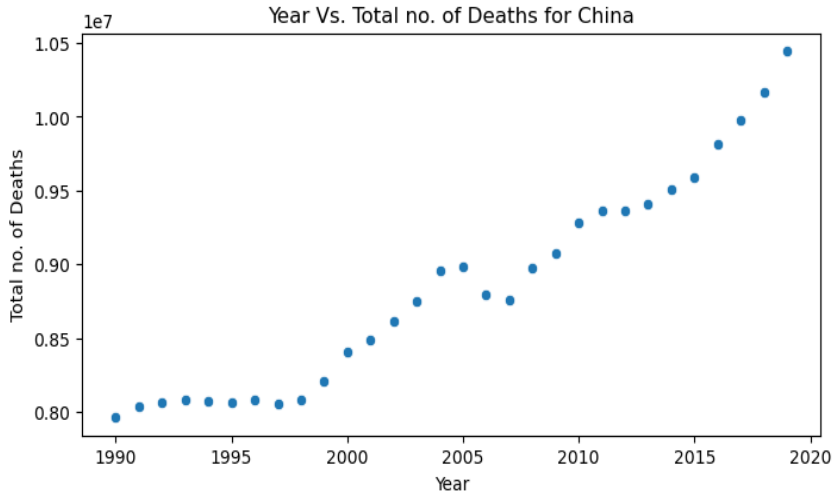


Here i have tried doing Univariate Analysis of Meningitis column and Country columns but here we can see that it now clearly visible but we will try it in other Way .. Little conclusion we can derive from it is there are many country which are

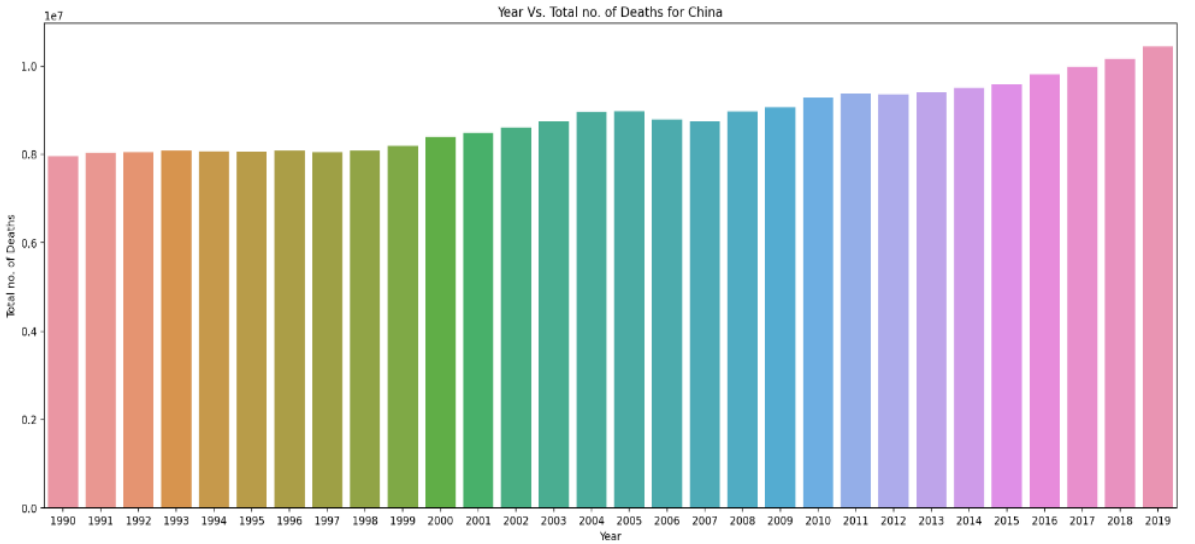
Visualization of China dataset:-

In [37]:
```
# china - "Total_no_of_Deaths" against "Year"
China_Total_no_of_Deaths_df=df[df['Country/Territory']=='China'].sort_values(by='Total_no_of_Deaths',ascending=False)
```

In [38]:
```
# China - "Total_no_of_Deaths" against "year"
plt.figure(figsize=(8,4),dpi=100)
sns.scatterplot(data=China_Total_no_of_Deaths_df, x='Year', y='Total_no_of_Deaths')
plt.xlabel("Year")
plt.ylabel("Total no. of Deaths")
plt.title("Year Vs. Total no. of Deaths for China")
plt.show()
```



In [39]:
```
plt.figure(figsize=(20,8), dpi=200)
sns.barplot(data=China_Total_no_of_Deaths_df, x='Year', y='Total_no_of_Deaths')
plt.xlabel("Year")
plt.ylabel("Total no. of Deaths")
plt.title("Year Vs. Total no. of Deaths for China")
plt.show()
```



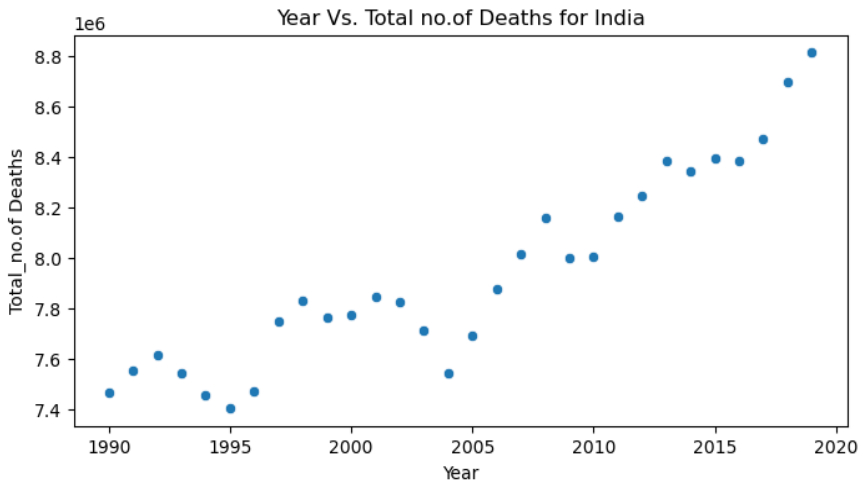
---

we can see is a clear raise in Total No. of Deaths recorded with each year for China

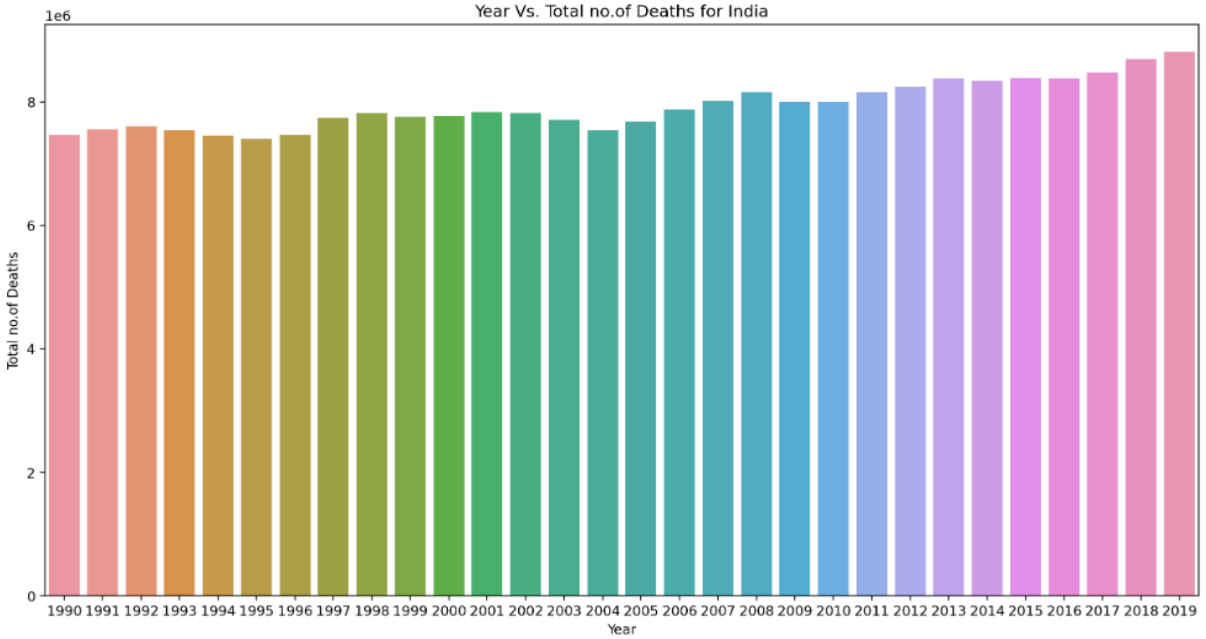
cause of India:-

In [54]:
```
#India-"Total_no_of_Deaths" against "Year"
India_Total_no_of_Deaths_df=df[df['Country/Territory']=='India'].sort_values(by='Total_no_of_Deaths',ascending=False)
```

In [55]:
```
#India-"Total_no_of_Deaths" against "Year"
plt.figure(figsize=(8,4), dpi=100)
sns.scatterplot(data=India_Total_no_of_Deaths_df, x='Year', y='Total_no_of_Deaths')
plt.xlabel("Year")
plt.ylabel("Total_no.of Deaths")
plt.title("Year Vs. Total no.of Deaths for India")
plt.show()
```



In [56]:
```
plt.figure(figsize=(15,8), dpi=200)
sns.barplot(data=India_Total_no_of_Deaths_df, x='Year', y='Total_no_of_Deaths')
plt.xlabel("Year")
plt.ylabel("Total no.of Deaths")
plt.title("Year Vs. Total no.of Deaths for India")
plt.show()
```



Overall there is a raise in Total No. of Deaths recorded with each year for India, even though there are fluctuations in between

Total Cause of Death across 30 year:-

In [72]:
```
# Total causes of death across 30 years
Countries_Total_no_of_Deaths_noyear_df=df.groupby('Country/Territory').sum()
Countries_Total_no_of_Deaths_noyear_df.drop('Year',axis=1, inplace=True)
```

In [73]:
```
# Top 5 Countries interms of "Total no. of Deaths"-For All the Years
Countries_Total_no_of_Deaths_noyear_df.sort_values(by='Total_no_of_Deaths', ascending=False)[:5]
```

Out[73]:

Country/Territory	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	HIV/AIDS	Drug Use Disorders	...	Chronic Kidney Disease	Poisonings
China	480899	5381846	1533092	584236	13418	2873619	776275	243257	433709	626914	...	4195276	770140
India	2008944	1707561	756832	3290569	2439244	2110438	1237163	2292449	2454374	168928	...	4556172	170119
United States	40032	3302609	661288	133044	0	114752	596818	25206	528417	800798	...	2018497	40259
Russia	60519	972305	236367	15906	0	423044	1215179	15028	350679	259452	...	325433	298954
Indonesia	337724	487566	145752	604467	74664	237902	81342	376966	74981	12980	...	964478	27837

5 rows × 32 columns

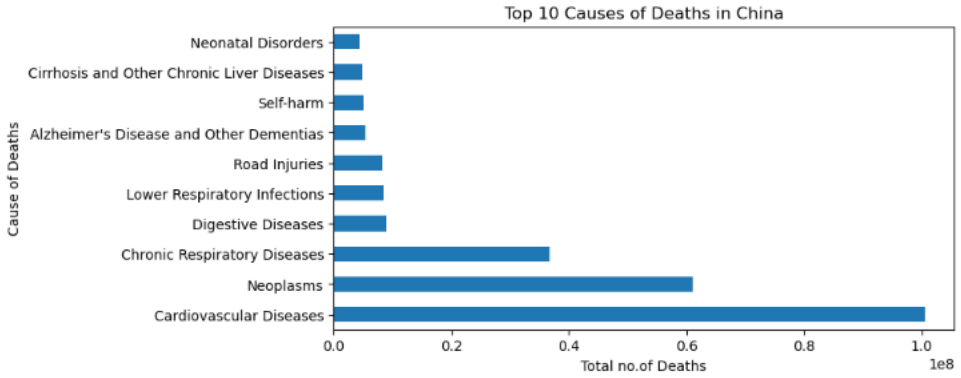
China-Top 10 Causes of Deaths:-

In [74]:
```
china_10=Countries_Total_no_of_Deaths_noyear_df.sort_values(by='Total_no_of_Deaths', ascending=False)[:1]
```

In [75]:
```
china_10.iloc[0].sort_values(ascending=False)[1:11]
```

Out[75]:
```
Cardiovascular Diseases                    100505973
Neoplasms                                   61060527
Chronic Respiratory Diseases                36676826
Digestive Diseases                           8924906
Lower Respiratory Infections                 8525819
Road Injuries                                8350399
Alzheimer's Disease and Other Dementias      5381846
Self-harm                                    5078550
Cirrhosis and Other Chronic Liver Diseases   4918899
Neonatal Disorders                           4353666
Name: China, dtype: int64
```

In [76]:
```
plt.figure(figsize=(8,4), dpi=100)
china_10.iloc[0].sort_values(ascending=False)[1:11].plot(kind='barh')
plt.xlabel("Total no.of Deaths")
plt.ylabel("Cause of Deaths")
plt.title("Top 10 Causes of Deaths in China")
plt.show()
```



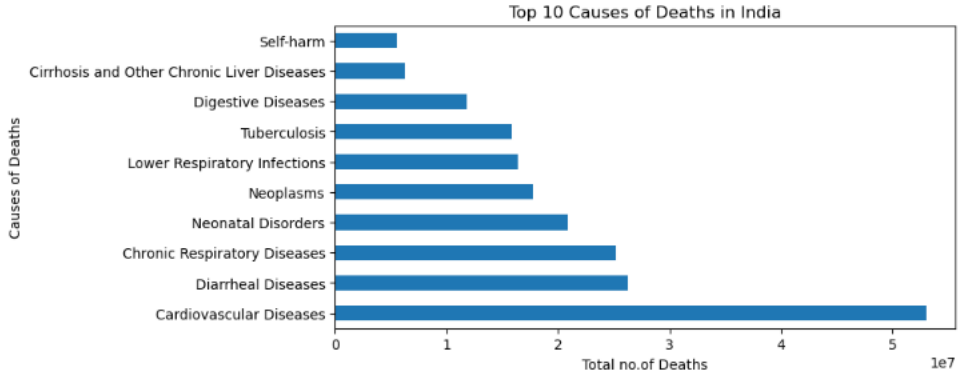
India-Top 10 Causes of Deaths:-

```
In [77]: India_10=Countries_Total_no_of_Deaths_noyear_df.sort_values(by='Total_no_of_Deaths', ascending=False)[1:2]
```

```
In [78]: India_10.iloc[0].sort_values(ascending=False)[1:11]
```

```
Out[78]: Cardiovascular Diseases                    52994710
         Diarrheal Diseases                         26243547
         Chronic Respiratory Diseases               25232974
         Neonatal Disorders                         20911570
         Neoplasms                                  17762703
         Lower Respiratory Infections               16419404
         Tuberculosis                               15820922
         Digestive Diseases                         11804380
         Cirrhosis and Other Chronic Liver Diseases  6294910
         Self-harm                                   5543395
         Name: India, dtype: int64
```

```
In [79]: plt.figure(figsize=(8,4), dpi=100)
         India_10.iloc[0].sort_values(ascending=False)[1:11].plot(kind='barh')
         plt.xlabel("Total no.of Deaths")
         plt.ylabel("Causes of Deaths")
         plt.title("Top 10 Causes of Deaths in India")
         plt.show()
```



United States-Top 10 Causes of Deaths:-

```
In [80]: usa_10=Countries_Total_no_of_Deaths_noyear_df.sort_values(by='Total_no_of_Deaths',ascending=False)[2:3]
```

```
In [81]: usa_10.iloc[0].sort_values(ascending=False)[1:11]
```

```
Out[81]: Cardiovascular Diseases                    26438346
         Neoplasms                                  18905315
         Chronic Respiratory Diseases                4949052
         Alzheimer's Disease and Other Dementias     3302609
         Digestive Diseases                          3026943
         Lower Respiratory Infections                2248625
         Diabetes Mellitus                           2030631
         Chronic Kidney Disease                      2018497
         Cirrhosis and Other Chronic Liver Diseases  1514325
         Road Injuries                               1359744
         Name: United States, dtype: int64
```

```
In [82]: plt.figure(figsize=(8,4), dpi=100)
         usa_10.iloc[0].sort_values(ascending=False)[1:11].plot(kind='barh')
         plt.xlabel("Total no.of Deaths")
         plt.ylabel("Causes of Deaths")
         plt.title("Top 10 Causes of Deaths in United States")
         plt.show()
```