FLIP ROBO

ANALYSIS OF CAUSE OF DEATH

SUBMITTED BY

LAVINA VAIDYA

ACKNOWLEDGMENT

We would like to thank Flip Robo for giving us the opportunity to work on this project. We are grateful for the trust and support that you have placed in us and for the resources that you have provided to help us complete this project. We would also like to thank our project manager for their dedication, hard work, and commitment to the project. Without their help, this project would not have been possible. Finally, we would like to thank all of the stakeholders involved in this project for their valuable input and feedback. Their contributions have been invaluable in helping us to deliver a successful project. Thank you all for your support.

INTRODUCTION

- ## CAUSE OF DEATH

  Cause of Death is a term used to indicate the medical cause of death. It lists the disease(s) or injuries that caused death. Specific cause of death information is recorded on the death certificate and is entered into the Vital Statistics System.

  The **cause of death** is defined as the disease or injury which started the train (sequence) of   morbid (disease-related) events which led directly to death, or the circumstances of the accident or violence which produced the fatal injury.

  This definition is derived from the International classification of diseases (ICD) of the World Health Organization. Although international definitions are harmonised, the statistics may not be fully comparable among countries, as classifications may vary when the cause of death is multiple or difficult to evaluate, and because of different notification procedures

- ## BUSINESS PROBLEM FRAMING

  The business problem of cause of death is to identify the root causes of death and develop strategies to prevent it. This includes understanding the factors that contribute to death, such as lifestyle, environmental, and medical conditions. It also involves understanding how different populations are affected by specific causes of death and developing strategies to reduce morbidity and mortality. This could include identifying disparities in access to care, developing preventive health initiatives, and improving public health infrastructure.

- **CONCEPTUAL BACKGROUND OF THE DOMAIN PROBLEM**

The main concept behind the domain problem of cause of death is the identification of the underlying factors that contribute to the death of an individual. This involves an analysis of the medical history, lifestyle, environment, and other factors to determine the most likely cause of death.

It is important to accurately identify the cause of death so that appropriate measures can be taken to prevent similar deaths in the future. Additionally, the identification of the cause of death can help inform public health policies and strategies to reduce the risk of death from certain causes.

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.

- **REVIEW OF LITERATURE**

The literature on cause of death analysis provides insights into many aspects of mortality, from the leading causes of death. The literature also covers the underlying causes of death, such as genetic and environmental factors, as well as the impact of different diseases.
To review this literature following steps are done to analyse
1. Importing required libraries
2. Loading dataset
3. Checking null values
4. Checking for white space
5. Visualization
6. conclusion

- **MOTIVATION FOR THE PROBLEM UNDERTAKEN**

The cause of death is a critical public health issue that can inform disease prevention, health promotion, and research into potential treatments. Additionally, accurately determining the cause of death can assist public health practitioners in identifying potential risk factors and planning interventions to reduce the mortality from certain causes. Determining the cause of death can also help inform public health policies and strategies for reducing the mortality rate from particular causes.

ANALYTICAL PROBLEM FRAMING

- ## MATHEMATICAL/ ANALYTICAL MODELING OF THE PROBLEM

  In this Dataset, we have Historical Data of different cause of deaths for all ages around the World. We analyse given dataset and observe that in given dataset target variable is absent so it is a case of unsupervised learning.

  For mathematical modelling first we load dataset. And studied about the variables/columns present in dataset.

- ## DATA SOURCES AND THEIR FORMATS

  We have a dataset of cause of death in different country from world.

  First I load a dataset

```
In [2]:  # loading dataset
         df=pd.read_csv('cause_of_death.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	
6119	Zimbabwe	ZWE	2019	1450	812	232	2884	2068	827	1434	...	3460	2292	405	

6120 rows × 34 columns

There are 6120 rows and 34 columns present in dataset

Description of columns are as follows:

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. HIV/AIDS - No. of People died from HIV/AIDS

13. Drug Use Disorders - No. of People died from Drug Use Disorders

14. Tuberculosis - No. of People died from Tuberculosis

15. Cardiovascular Diseases - No. of People died from Cardiovascular Diseases

16. Lower Respiratory Infections - No. of People died from Lower Respiratory Infections

17. Neonatal Disorders - No. of People died from Neonatal Disorders

18. Alcohol Use Disorders - No. of People died from Alcohol Use Disorders

19. Self-harm - No. of People died from Self-harm

20. Exposure to Forces of Nature - No. of People died from Exposure to Forces of Nature

21. Diarrheal Diseases - No. of People died from Diarrheal Diseases

22. Environmental Heat and Cold Exposure - No. of People died from Environmental Heat and Cold Exposure

23. Neoplasms - No. of People died from Neoplasms

24. Conflict and Terrorism - No. of People died from Conflict and Terrorism

25. Diabetes Mellitus - No. of People died from Diabetes Mellitus

26. Chronic Kidney Disease - No. of People died from Chronic Kidney Disease

27. Poisonings - No. of People died from Poisoning

28. Protein-Energy Malnutrition - No. of People died from Protein-Energy Malnutrition

29. Road Injuries - No. of People died from Road Injuries

30. Chronic Respiratory Diseases - No. of People died from Chronic Respiratory Diseases

31. Cirrhosis and Other Chronic Liver Diseases - No. of People died from Cirrhosis and Other Chronic Liver Diseases

32. Digestive Diseases - No. of People died from Digestive Diseases

33. Fire, Heat, and Hot Substances - No. of People died from Fire or Heat or any Hot Substances

34. Acute Hepatitis - No. of People died from Acute Hepatitis

- **DATA PREPROCESSING DONE**

  After loading a dataset

  ➢ check for shape of data set

```
In [3]:   # checking shape of dataset
          df.shape

Out[3]:   (6120, 34)
```

➢ Check for datatype

```
In [8]:  # checking datatype of dataframe
         df.dtypes

Out[8]:  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
         Cirrhosis and Other Chronic Liver Diseases     int64
         Digestive Diseases                             int64
         Fire, Heat, and Hot Substances                 int64
         Acute Hepatitis                                int64
         dtype: object
```

Only 2 columns are of categorical type and others are integer type

➢ Checking for null values

```
In [11]: #visualization of null values
         plt.figure(figsize=(20,18))
         plt.title('visualization of null values')
         sns.heatmap(df.isnull())
         plt.show()
```


visualization of null values

we don't have any null value in dataset
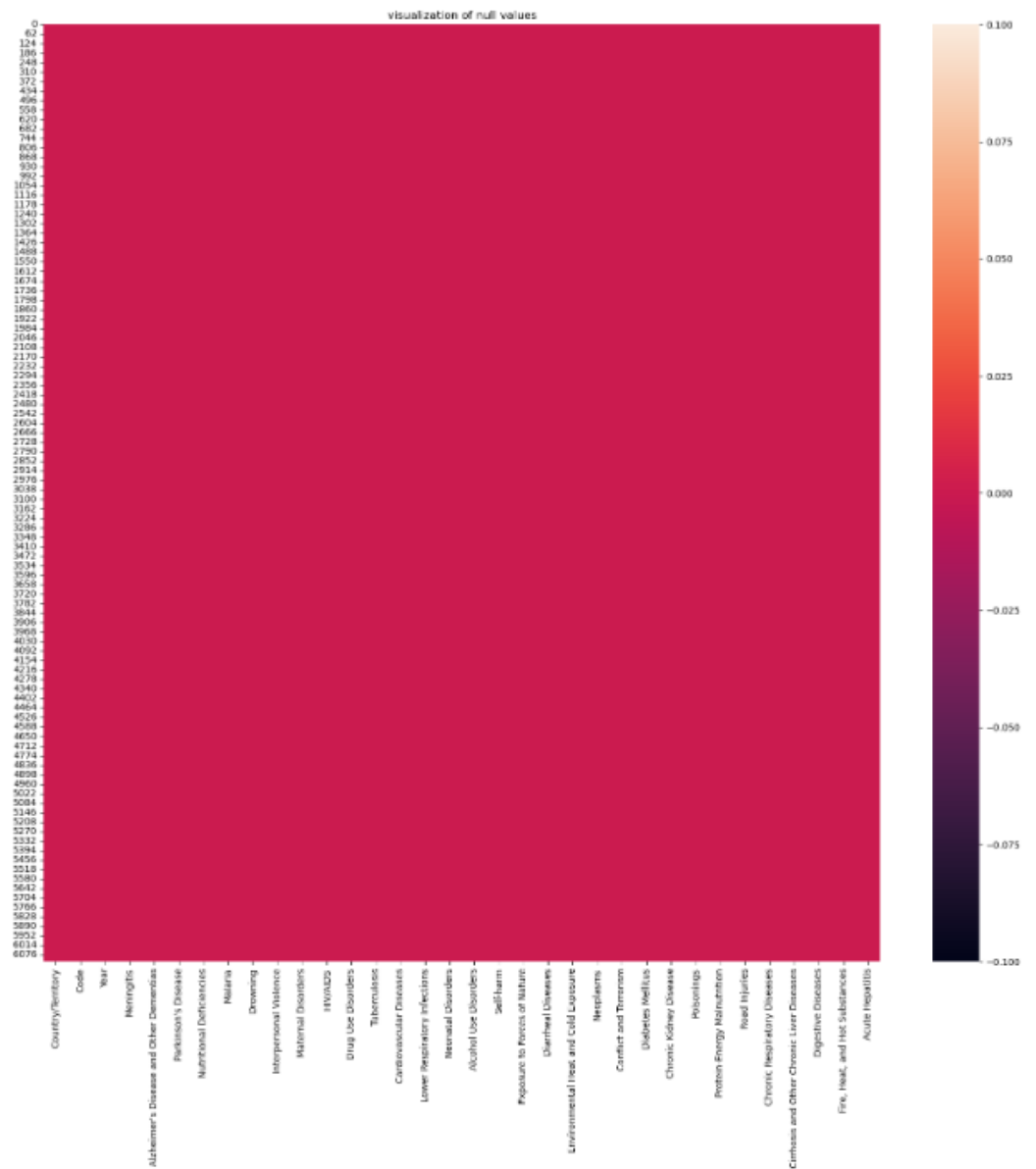
➢ Remove whitespaces if present in dataset with the help of
   following function

```
In [3]: # Creating a function which will remove extra leading
        # and tailing whitespace from the data.
        # pass dataframe as a parameter here
        def whitespace_remover(df):
            # iterating over the columns
            for i in df.columns:
                # checking datatype of each columns
                if df[i].dtype == 'object':
                    # applying strip function on column
                    df[i] = df[i].map(str.strip)
                else:
                    # if condn. is False then it will do nothing.
                    pass

        # applying whitespace_remover function on dataframe
        whitespace_remover(df)

        # printing dataframe
        df
```

Now data is ready for visualization

- **HARDWARE AND SOFTWARE REQUIREMENTS AND TOOLS
  USED**

To analyze death of cause projects it is important to have the following
hardware and software requirements and tools.

- **Hardware required:**
  - Processor: core i5 or above
  - RAM: 8 GB or above
  - ROM/SSD: 250 GB or above
- **Software required:**
  - Anaconda 3- language used Python3
- **Library used**
  import pandas as pd
  import seaborn as sns
  import matplotlib.pyplot as plt
  import warnings
  warnings.filterwarnings('ignore')

MODEL/S DEVELOPMENT AND EVALUATION

- ## VISUALIZATIONS

  Some visualization of input variable are given which is very usefull for interpretation of our conclusion.
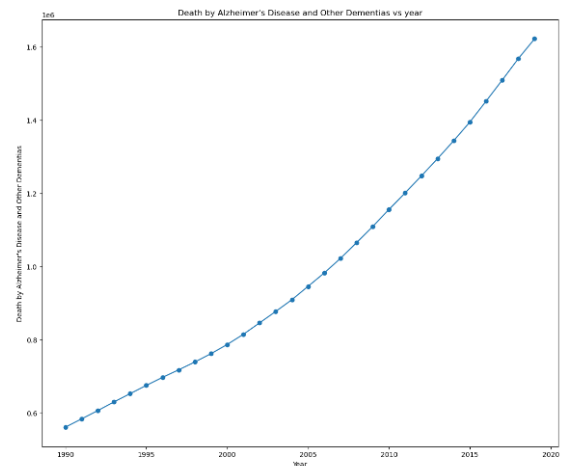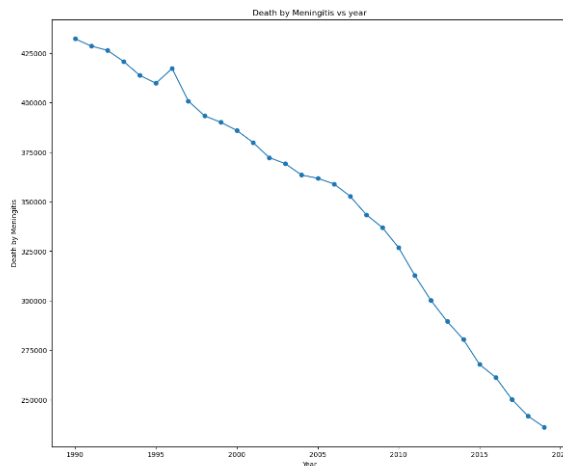
observation:

   1. Except for 1996, when the death rate climbed over the year before, meningitis-related deaths are declining. in 1990 more than 42500 death were recorded which is high death record.

   2. Death by Alzheimer's Disease and Other Dementias death rates are rising year over year.
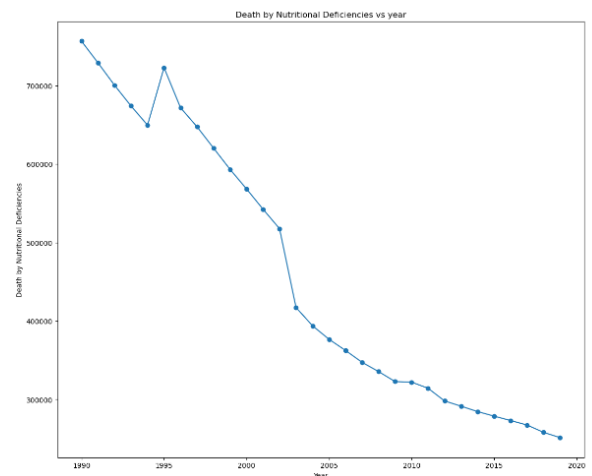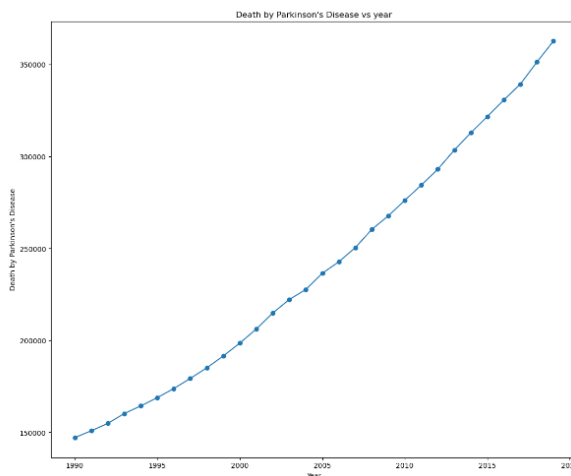
observation:

1. Parkinson's disease death rates are rising year over year. more than 350000 deaths were recorded in 2019 which is high death record.

2. Deaths caused by nutritional deficiencies are on the decline, although overall death rates are rising slowly in 1995. and in 1990 high record in death recorded which is more than 700000.
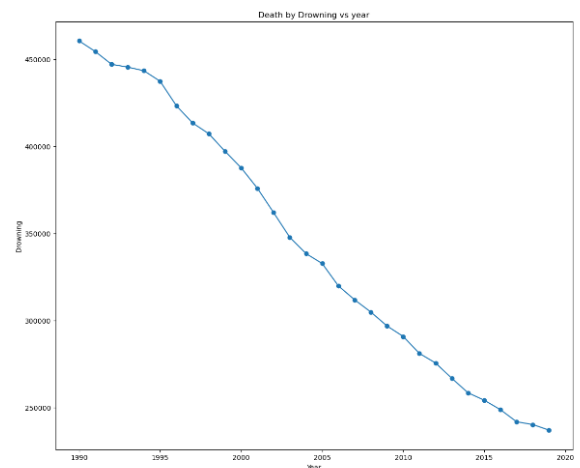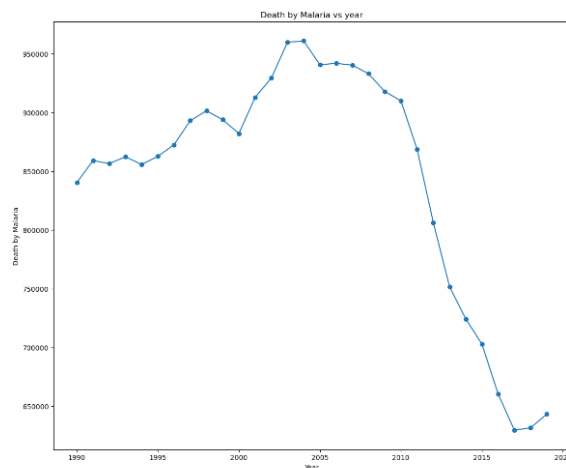
observation:

1. in year 2003 and 2004 deaths by malaria is increased,which is reported by highest deaths of malaria after that deaths are gradually decreased but from 2017 death rate is increasing

2. death by drowing is decreases by year over year and highest death recorded in 1990
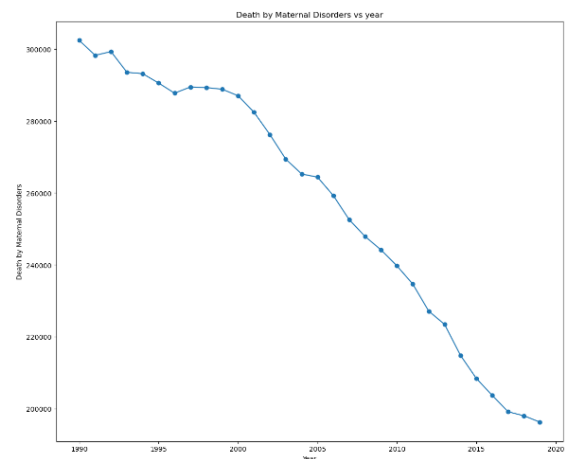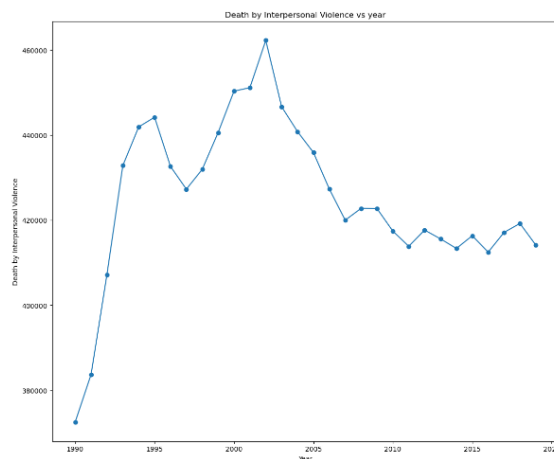
observation:

1. Death by Interpersonal Violence is high in 2003

2. Death by Maternal Disorders is decreases by year over year highest death recorded in 1990.
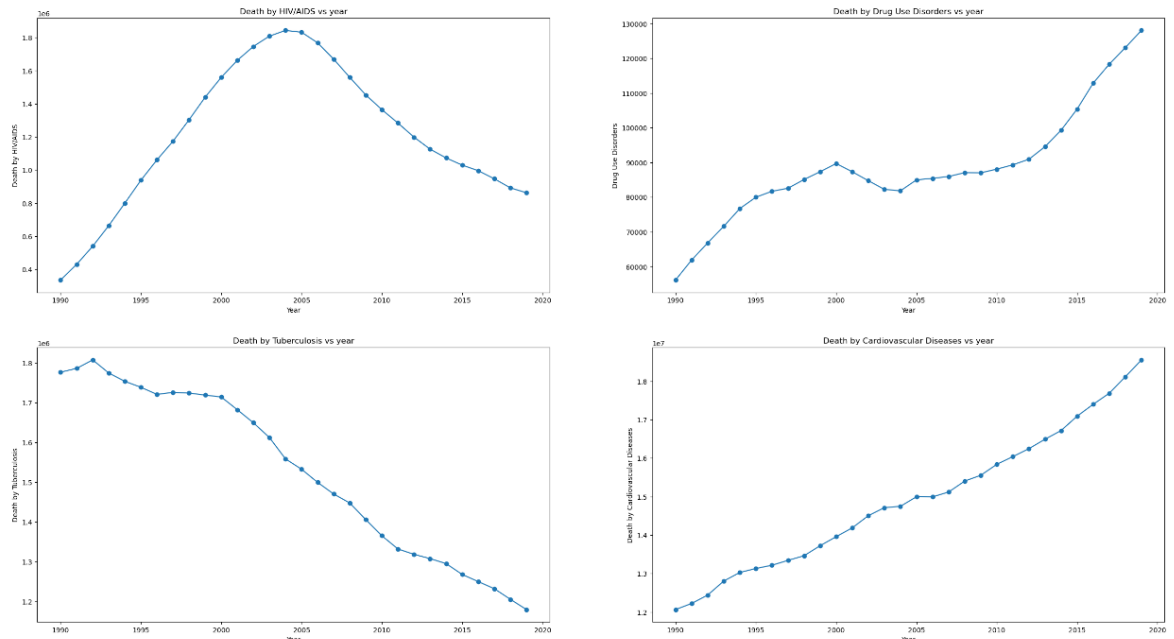
observation:

1. Death by HIV/AIDS is increases from 1990 year over year, but we observed that deaths are decreases from 2004 which recorded as highest death year.

2. we observed that Death by Drug Use Disorders is increased upto 2000 then fall in record observed upto 2005 and then again high increase record observed and highest death record in 2019.

3. Death by Tuberculosis decrease year over year and highest death record in 1992 .

4. Death by Cardiovascular Diseases increases year over year and highest death record in 2019.
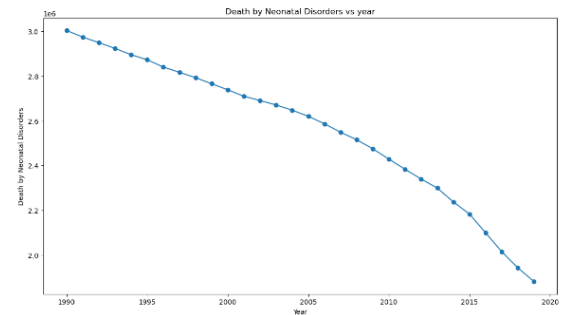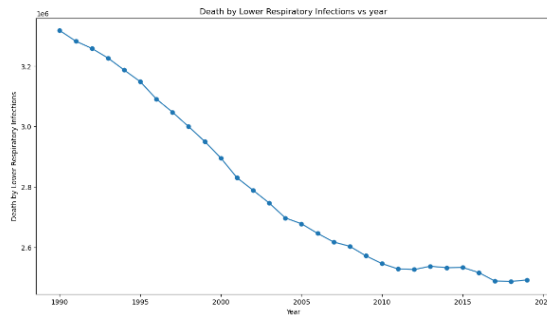
observation:

1. Death by Lower Respiratory Infections decrease year over year and highest death record in 1990.

2. Death by Neonatal Disorders decrease year over year and highest death record in 2019.
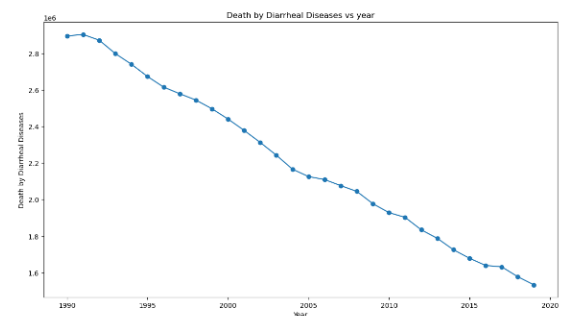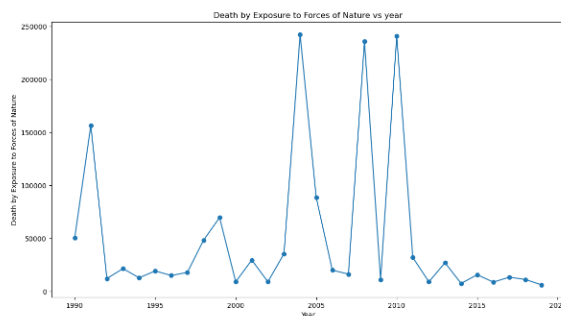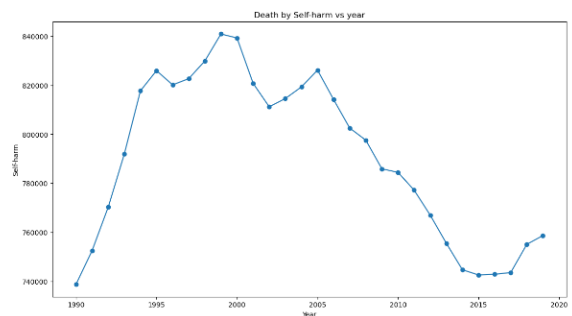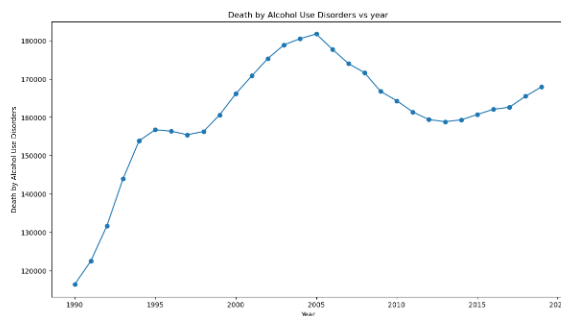
Observation:

1. Death by Alcohol Use Disorders increases from 1990 and scores high in 2005 but after 2005 there is decrease in death rate

2. Death by Self-harm is highly unstable and recorded high in 1999

3. Death by Exposure to Forces of Nature is highly unstable and recorded high in 2004, 2008 and 2010

4. Death by Diarrheal Diseases decreses gradually and recorded high in 1991
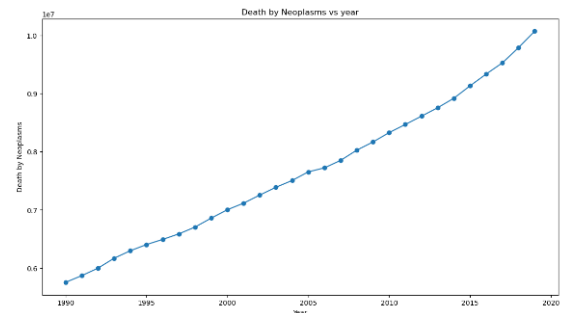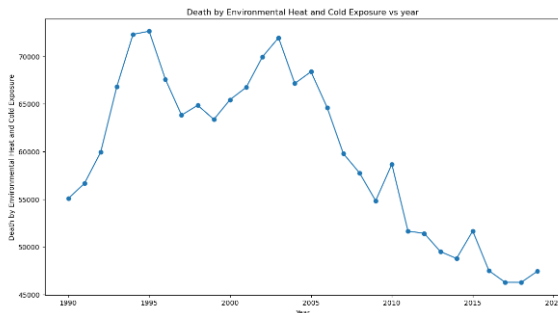
Observation:

1. Death by Environmental Heat and Cold Exposure is unstable and recorded high in 1995

2. Death by Neoplasms increases year over year and recorded high in 2019.
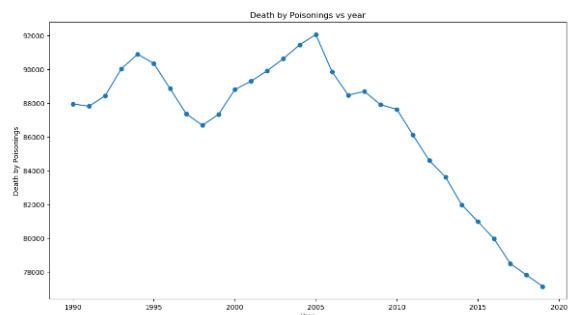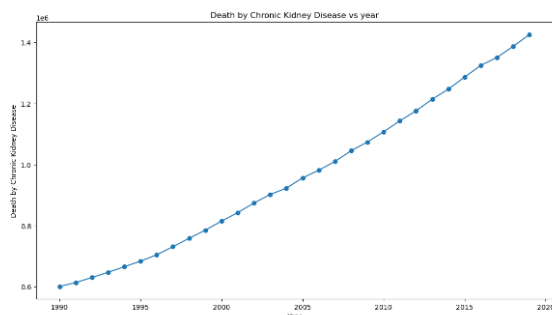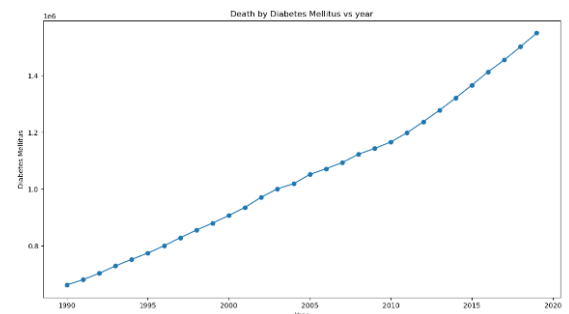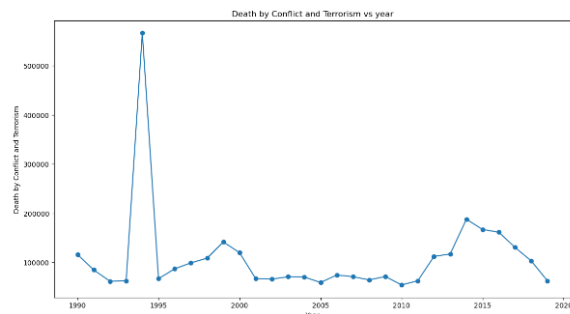
Observation:

1. Death by Conflict and Terrorism is highly unstable and recorded very high rate in 1994

2. Death by Diabetes Mellitus is increases year over year and recorded high in 2019

3. Death by Chronic Kidney Disease is increases year over year and recorded high in 2019

4. Death by Poisonings is unstable and recorded high in 2005
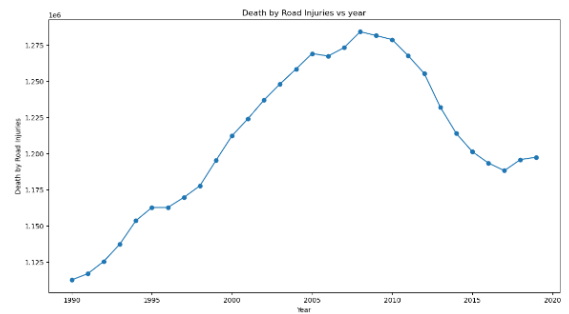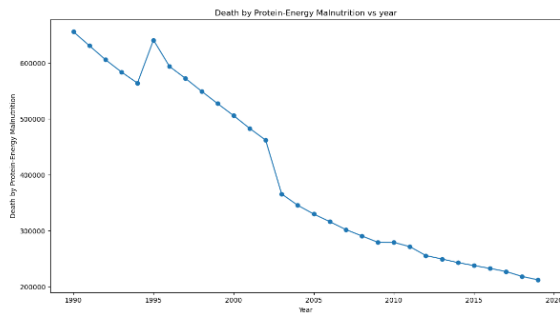
Observation:

1. Death by Protein-Energy Malnutrition is decreases and recorded high in 1990

2. Death by Road Injuries increases year over year upto 2008 which recorded as high rate after that it decreases gradually
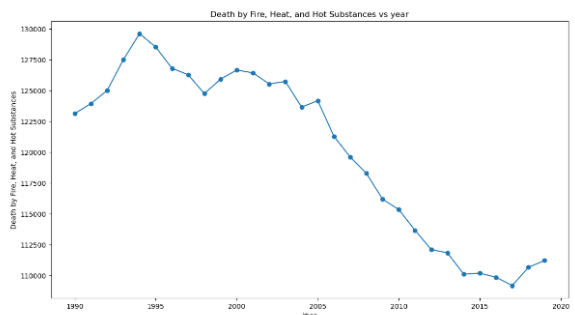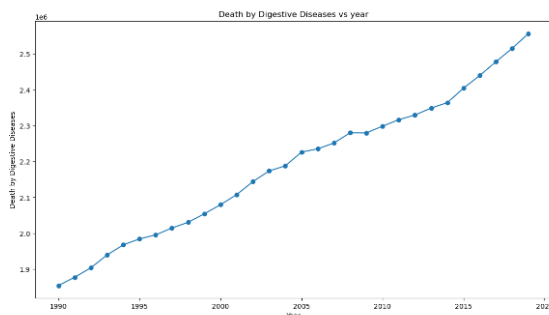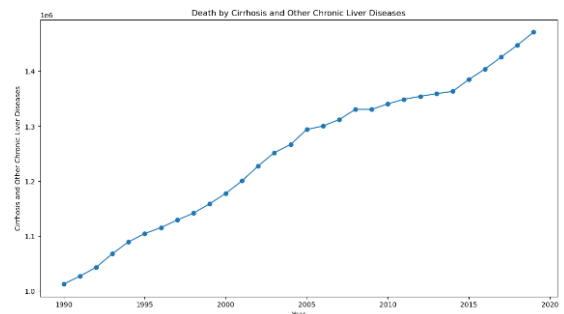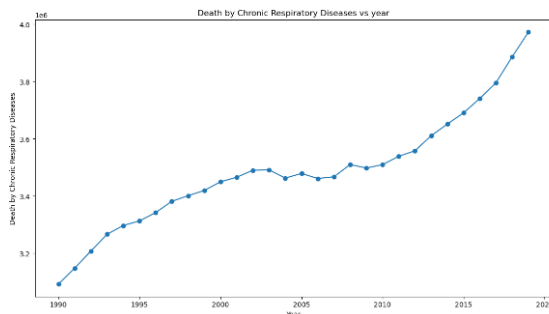
Observation:

1. Death by Chronic Respiratory Diseases increases year over year and recorded high in 2019
2. Death by Cirrhosis and Other Chronic Liver Diseases increases year over year and recorded high in 2019
3. Death by Digestive Diseases increases year over year and recorded high in 2019
4. Death by Fire, Heat, and Hot Substances is highly unstable and recorded high in 1994

Death by Acute Hepatitis vs year
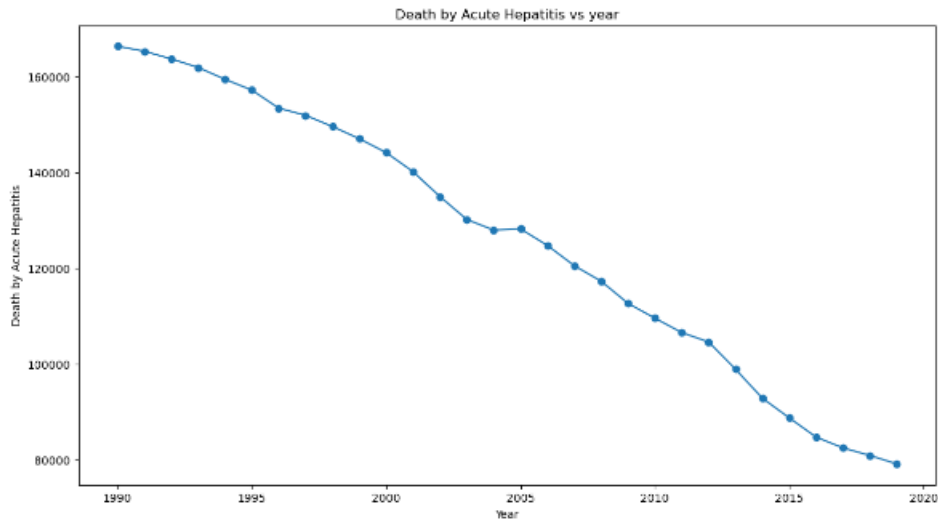
Observation:

1. Death by Acute Hepatitis decreses year over year and recorded high in 1990.

VISUALIZATION OF TOP 10 COUNTRY WITH HIGH DEATH RATE
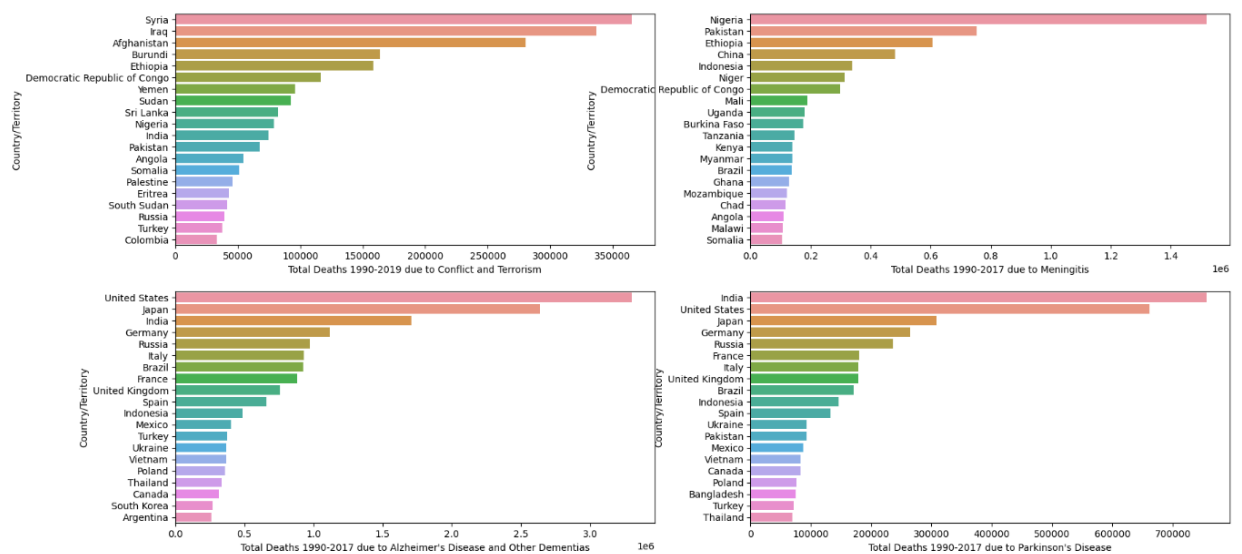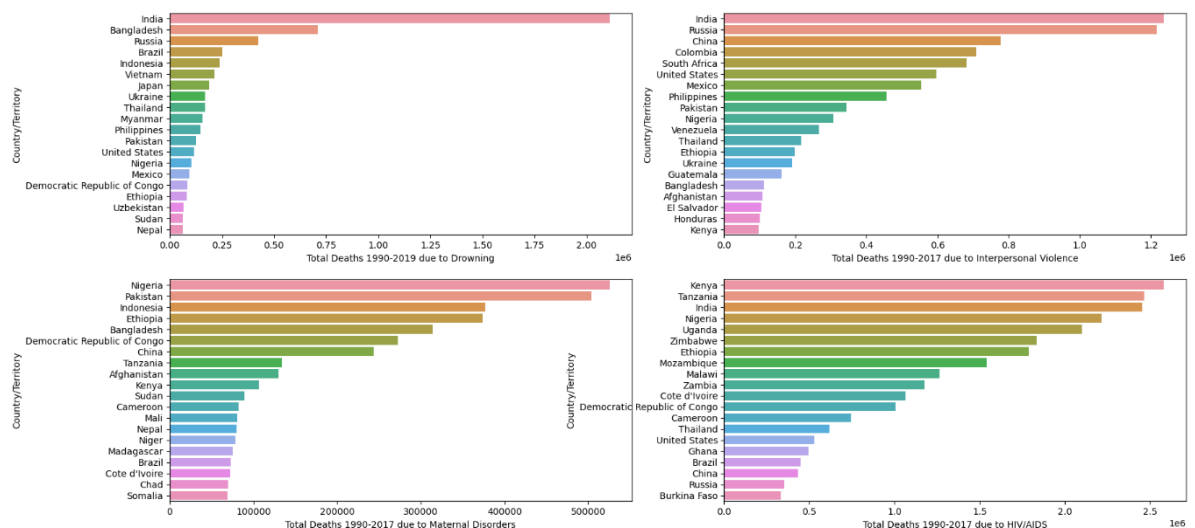

Observation:

1. Total Death due to Conflict and Terrorism is high in Syria followed by Iraq

2. Total Death due to Meningitis is high in Nigeria followed by Pakistan

3. Total Death due to Alzheimer's Disease and Other Dementias is high in United State followed by Japan

4. Total Death due to Parkinson's Disease is high in India followed by United State
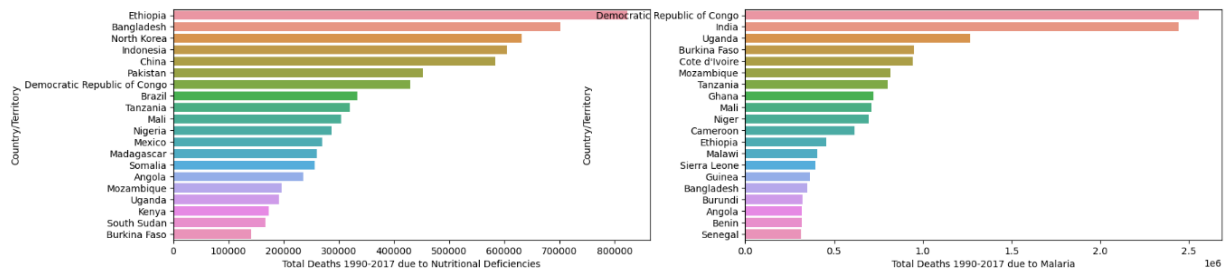
Observation:

1. Total Death due to Nutritional Deficiencies is high in Ethopia followed by Bangladesh

2. Total Death due to Malaria is high in Republic of congo followed by India

Observation:

1. Total Death due to Drowning is high in India followed by Bangladesh

2. Total Death due to Interpersonal Violence is high in India followed by Russia

3. Total Death due to Maternal Disorders is high in Nigeria followed by Pakistan

4. Total Death due to HIV/AIDS is high in Kenya followed by Tanzaia and India
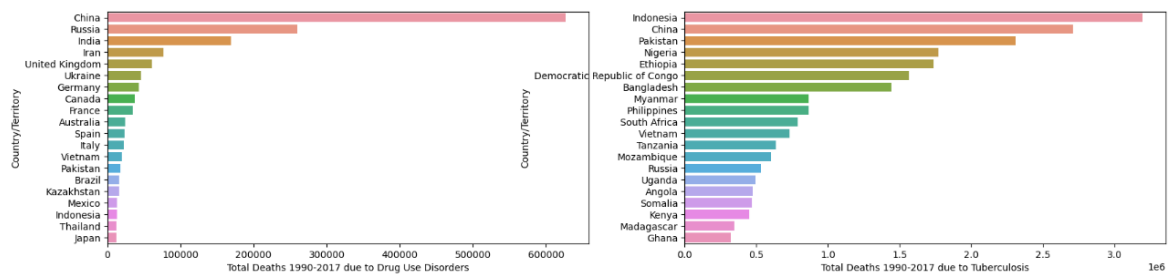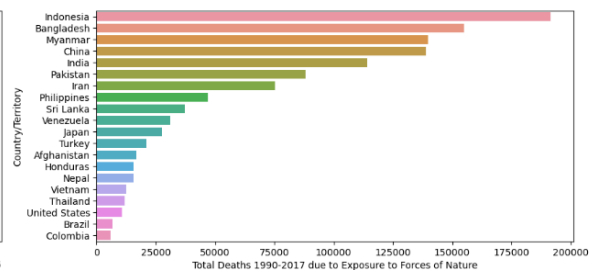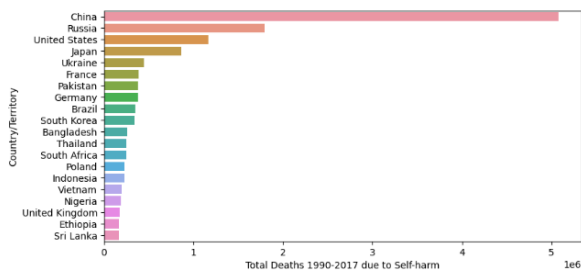
Observation:

1. Total Death due to Drug Use Disorders is high in China followed by Russia

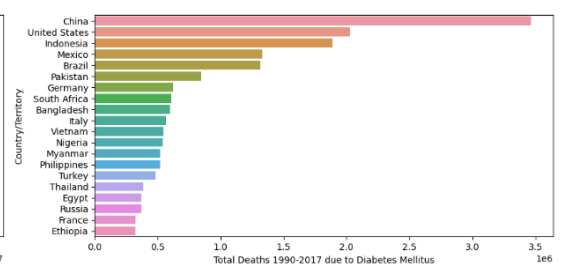2. Total Death due to Tuberculosis is high in Indonesia followed by China
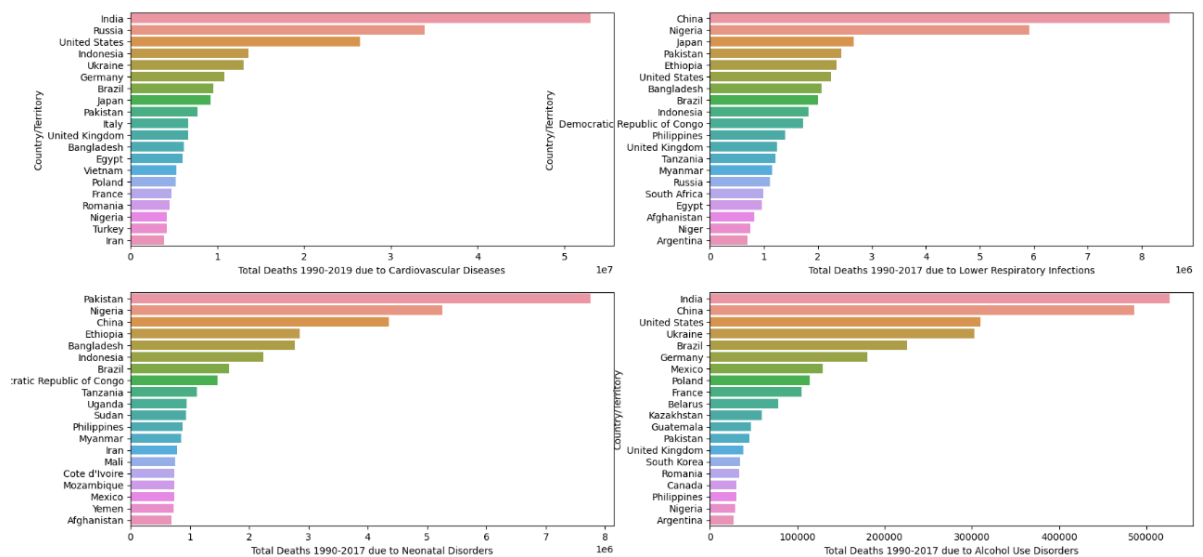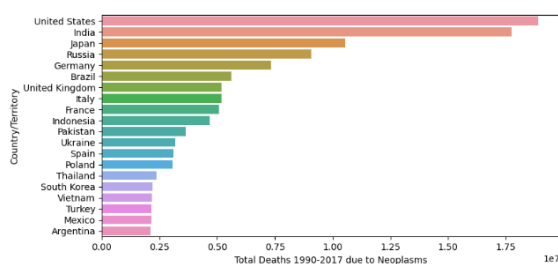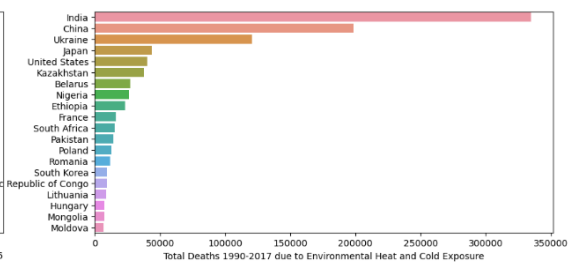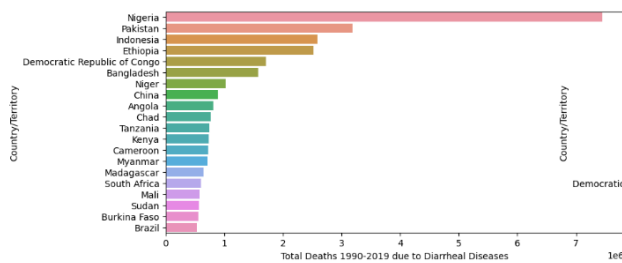

Observation:

1. Total Death due to Cardiovascular Diseases is high in India followed by Russia

2. Total Death due to Lower Respiratory Infections is high in China followed by Nigeria

3. Total Death due to Neonatal Disorders is high in Pakistan followed by Nigeria

4. Total Death due to Alcohol Use Disorders is high in India followed by China, United State
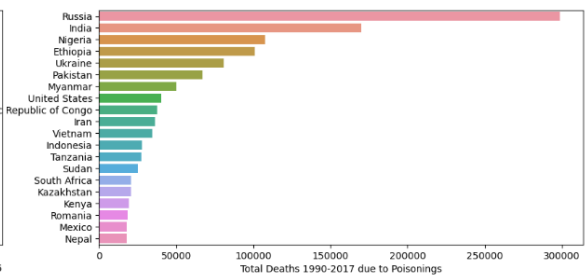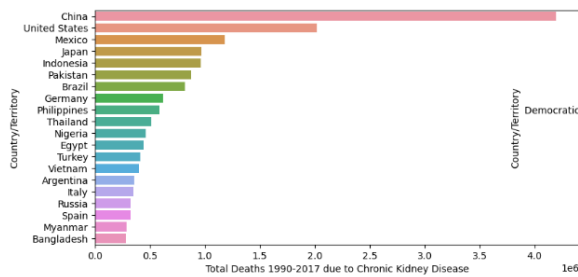
Observation:

1. Total Death due to Self-harm is high in China followed by Russia

2. Total Death due to Exposure to Forces of Nature is high in Indonesia followed by Bangladesh
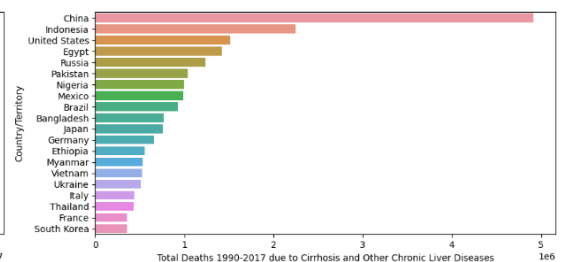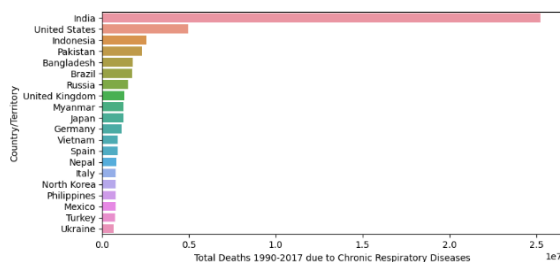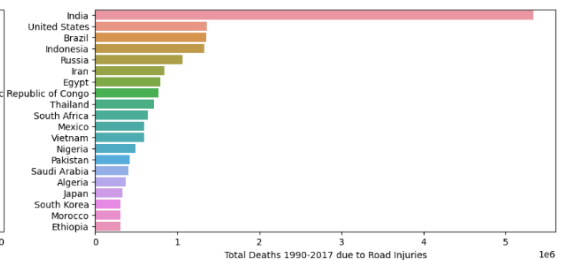

Observation:

1. Total Death due to Diarrheal Diseases is high in Nigeria followed by Pakistan

2. Total Death due to Environmental Heat and Cold Exposure is high in India followed by China

3. Total Death due to Neoplasms is high in United State followed by Japan

4. Total Death due to Diabetes Mellitus is high in China followed by United State

Observation:

1. Total Death due to Chronic Kidney Disease is high in China followed by United State

2. Total Death due to Poisonings is high in Russia followed by India
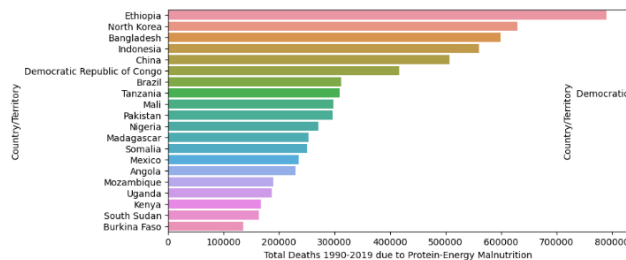
Observation:

1. Total Death due to Protein-Energy Malnutrition is high in Ethopia followed by North korea

2. Total Death due to Road Injuries is high in India followed by United State

3. Total Death due to Chronic Respiratory Diseases is high in India followed by United State

4. Total Death due to Cirrhosis and Other Chronic Liver Diseases is high in China followed by Indonesia
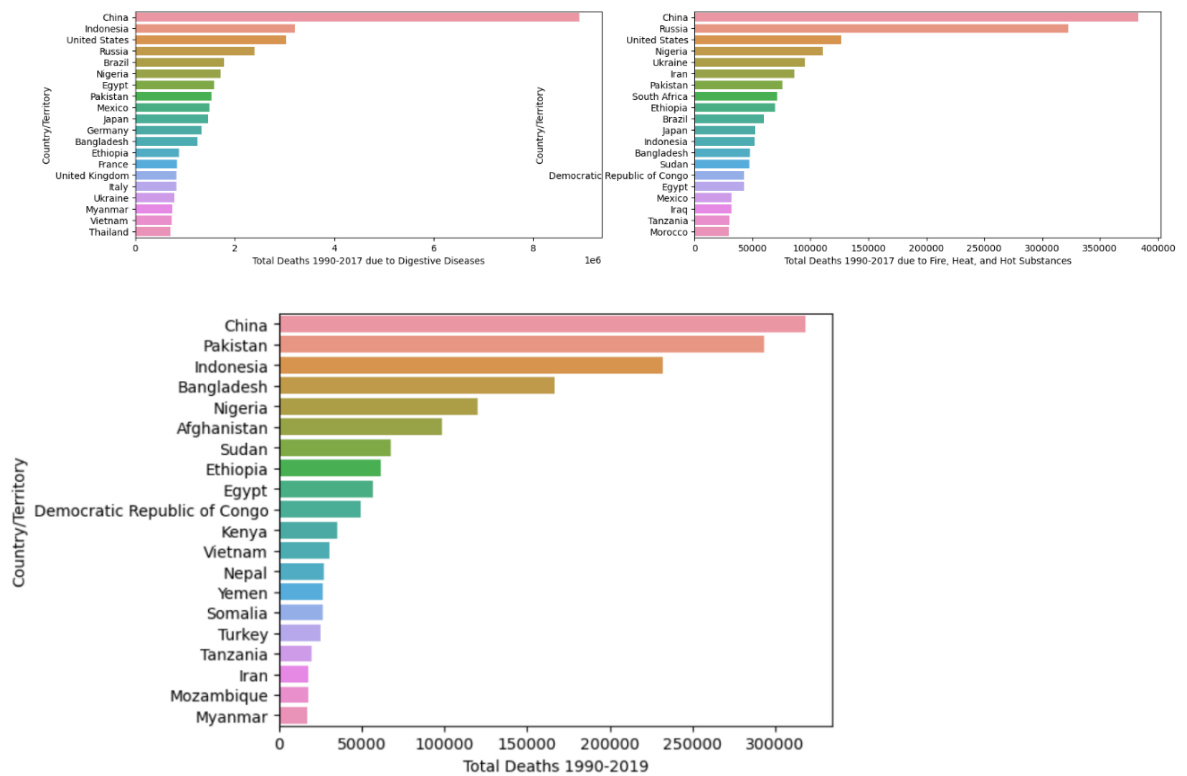
Observation:

1. Total Death due to Digestive Diseases is high in China followed by Indonesia

2. Total Death due to Fire, Heat, and Hot Substances is high in China followed by Russia

3. total death due to Total Deaths is high in china followed by pakistan, indonesia

CONCLUSION

1. **India** : Death rate is high due to Chronic Respiratory Diseases, Environmental Heat and Cold Exposure, Neoplasms, Cardiovascular Diseases, Alcohol Use Disorders, Parkinson's Disease, Malaria, Drowning, Interpersonal Violence, HIV/AIDS.

2. **Pakistan** : Death rate is high due to Acute Hepatitis, Diarrheal Diseases, Neonatal Disorders, Maternal Disorders, Tuberculosis, Meningitis.

3. **Indonesia** : Death rate is high due to Acute Hepatitis, Cirrhosis and Other Chronic Liver Diseases, Digestive Diseases, Diarrheal Diseases, Exposure to Forces of Nature, Tuberculosis, Nutritional Deficiencies.

4. **China** : Death rate is high due to Acute Hepatitis, Cirrhosis and Other Chronic Liver Diseases, Digestive Diseases,Fire, Heat, and Hot Substances, Chronic Kidney Disease, Environmental Heat and Cold Exposure, Diabetes Mellitus, Lower Respiratory Infections, Alcohol Use Disorders, Self-harm, Nutritional Deficiencies, Tuberculosis, Drug Use Disorders, Interpersonal Violence.

5. **United state** : Death rate is high due to Alzheimer's Disease and Other Dementias, Parkinson's Disease, Cardiovascular Diseases, Self-harm, Neoplasms, Diabetes Mellitus, Road Injuries, Chronic Respiratory Diseases, Cirrhosis and Other Chronic Liver Diseases, Digestive Diseases, Fire, Heat, and Hot Substances

6. **Ethopia** : Death rate is high due to Protein-Energy Malnutrition, Diarrheal Diseases, Neonatal Disorders, Tuberculosis, Maternal Disorders, Nutritional Deficiencies, Meningitis, Conflict and Terrorism.

7**. Russia** : Death rate is high due to Fire, Heat, and Hot Substances, Digestive Diseases, Cirrhosis and Other Chronic Liver Diseases, Road Injuries, Neoplasms, Poisonings, Self-harm, Cardiovascular Diseases, Alzheimer's Disease and Other Dementias, Parkinson's Disease, Interpersonal Violence, Drowning

8. **North Korea** : Death rate is high due to Nutritional Deficiencies, Protein-Energy Malnutrition.

9. **Nigeria** : Death rate is high due to Fire, Heat, and Hot Substances, Digestive Diseases, Diarrheal Diseases, Poisonings, Meningitis, HIV/AIDS, Neonatal Disorders.

10**. Bangladesh** : Death rate is high due to Acute Hepatitis, Nutritional Deficiencies, Drowning, Maternal Disorders, Neonatal Disorders, Exposure to Forces of Nature, Protein-Energy Malnutrition, Chronic Respiratory Diseases

11. **Mexico** : Death rate is high due to Chronic Kidney Disease, Diabetes Mellitus.

12. **Japan** : Death rate is high due to Chronic Kidney Disease, Neoplasms, Self-harm, Lower Respiratory Infections, Alzheimer's Disease and Other Dementias, Parkinson's Disease, Drowning, Cardiovascular Diseases.

13. **Ukraine** : Death rate is high due to Cardiovascular Diseases, Alcohol Use Disorders, Drug Use Disorders, Environmental Heat and Cold Exposure, Self-harm.

14. **Kenya** : Death rate is high due to HIV/AIDS

15. **South Africa** : Death rate is high due to Interpersonal Violence.

16. **Democratic republic of congo** : Death rate is high due to Nutritional Deficiencies, Maternal Disorders, Conflict and Terrorism, Tuberculosis, Poisonings, Protein-Energy Malnutrition, Diarrheal Diseases.

17. **Syria** : Death rate is high due to Conflict and Terrorism,

18. **Tanzania** : Death rate is high due to HIV/AIDS, Malaria, Acute Hepatitis

- **LIMITATIONS OF THIS WORK**

  The main limitation of course of death analysis is that it does not take into account other factors that can contribute to death, such as lifestyle and environmental factors. Additionally, it does not consider the potential for misclassification of cause of death. Furthermore, it does not consider the effect of different treatments on mortality.

- **SCOPE FOR FUTURE WORK**

  The scope for future work in this area could include further analysis of the role of lifestyle and environmental factors in mortality, as well as the potential for misclassification of cause of death. Additionally, further research could be done on the effects of various treatments on mortality. Additionally, more research could be done to understand the impact of social determinants of health on mortality.

  Finally, further research could be done to better understand the impact of specific conditions and treatments on mortality.

STEPS/TECHNIQUES CAN BE FOLLOWED TO FURTHER EXTEND THIS STUDY AND IMPROVE THE RESULTS.

1. Perform Autopsy: Autopsy is the most detailed examination of a body after death. It provides an in-depth analysis of the cause of death and can provide a more accurate diagnosis.

2. Obtain Medical History: Medical records, such as hospital and doctor visits, can provide additional information about the deceased. This information can help to rule out certain conditions or illnesses that may have contributed to the death.

3. Investigate the Scene: It is important to investigate the scene of death to determine if the death was caused by an accident or intentional means.

4. Collect Evidence: Evidence such as DNA samples, clothing, and other objects found at the scene of death should be collected and analyzed to help determine the cause of death.

5. Perform Toxicology Tests: Toxicology tests can help to determine if the deceased had any drugs or alcohol in their system at the time of death.

6. Interview Witnesses: Witnesses can provide valuable information about the deceased, including any signs or symptoms they exhibited prior to their death.

7. Review Medical Records: Medical records can provide additional information about the deceased's previous medical history and any treatments they had prior to their death.

8. Consult Path

REFERENCES

- https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Cause_of_death#:~:text=The%20cause%20of%20death%20is,which%20produced%20the%20fatal%20injury.
- https://snohomishcountywa.gov/806/Cause-Manner-of-Death
- https://www.kaggle.com/code/spscientist/a-simple-tutorial-on-exploratory-data-analysis
- https://en.wikipedia.org/wiki/Exploratory_data_analysis#:~:text=In%20statistics%2C%20exploratory%20data%20analysis,and%20other%20data%20visualization%20methods