CAUSE OF DEATH

Submitted by:

SHRUTI KUMARI

ACKNOWLEDGMENT

I am extremely thankful to google.com that give us indepth knowledge of various topics and diseases that used in this dataset and project.

INTRODUCTION

Health is most important factor of life. It is very important to maintain good health to live life peacefully and happily. Our lifestyle changes cause many diseases from a very young age. Some disease caused by our ignorance on the health. We can focus on Mortality rate that shows the death rate of population, child mortality that shows the death of child under the age of 5 and infant mortality that shows the death of child under the age of 1, all these data 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. We can access health outcomes by both mortality and morbidity (rate of disease in population) provides more encompassing view on health outcomes. We can say sum of mortality and morbidity is referred as 'Burden of Disease' and can be measured by **'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. One DALY represents one lost year of healthy life.

The **Global Burden of Disease Study (GBD)** is a comprehensive regional and global research program of disease burden that assesses mortality and disability from major diseases, injuries, and risk factors. GBD is a collaboration of over 3600 researchers from 145 countries. 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.

We have to understand Global burden of disease in detail. We have to know some concept of these diseases to better understanding of the project. 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, Neoplasm, 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.

Analytical Problem Framing

Dataset contains 6140 rows and 34 columns. Out of 34 columns, 31 columns contain the number of deaths caused by 31 different diseases.

Dataset have detailed number of deaths from year 1990 to 2019 of 204 different countries.

- 1990-2019 have 30 unique years
- 204 unique countries
- 31 unique Diseases

After understanding dataset we have to find out some of the data and trends.

Following are the trends and data that we have to find out:

- Top 5 maximum deaths in which country
- From top 5 we will find out top 2 country with maximum deaths
- We will find out total no. of deaths with respect to year
- Main cause of deaths and and their trends year-wise
- Top 10 main cause of deaths with maximum deaths and their trends
- Year with maximum number of Deaths
- Trends of Disease year wise

Libraries that are used:

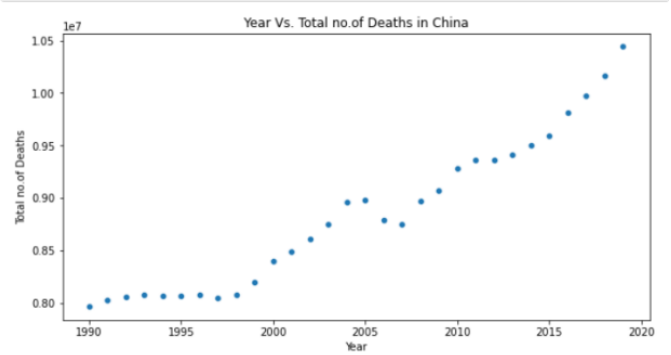
1. Pandas: It is a Python Library and It is used to analyze data.
2. NumPy: It is a Python library used for working with arrays.
3. Matplotlib: It is a comprehensive library for creating static, animated and interactive visualizations in Python.
4. Seaborn: It is a Python data visualization library based on matplotlib.It provides a high level interface for drawing attractive and informative statistical graphics.
5. Warnings: The warning module is actually a subclass of Exception which is a built-in class in Python.

Some of the function used:

- shape: give me the total number of rows and columns
- columns: It shows the name of the columns
- dtypes: It shows the data types of the columns
- Isnull().sum(): It shows the summation of null values in the data
- describe(): shows the statistical summary of the data
- unique(): shows the unique values of the column
- nunique(): shows the number of unique values in the column

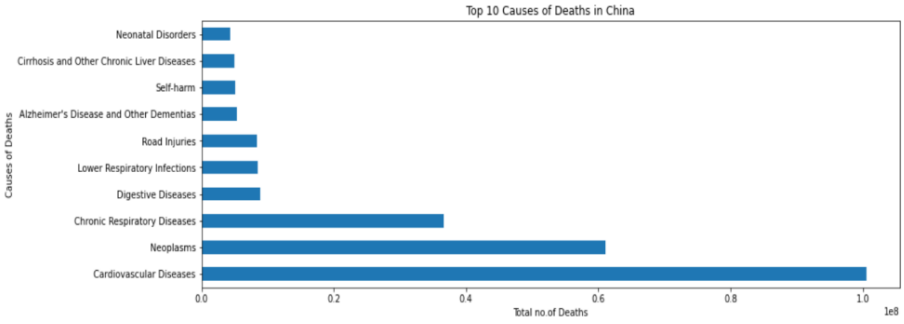
CONCLUSION

- Maximum total Deaths in:
    1. CHINA - 265408106
    2. INDIA – 238158165
    3. UNITED STATES- 71197802
    4. RUSSIA-59591155
    5. INDONESIA - 44046941
- CHINA:
    1. Deaths in china increases from year 1990 to 2019

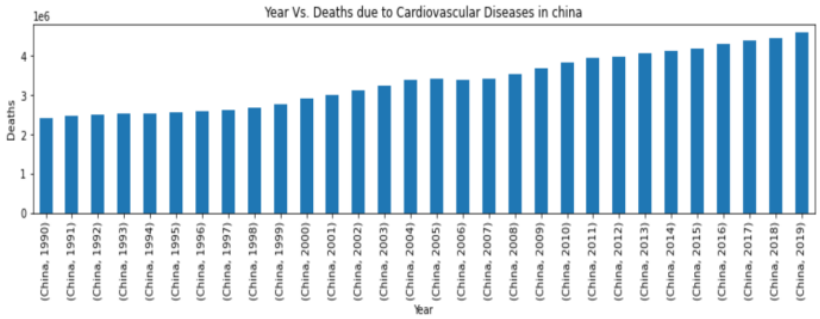


    2. Top 10 main cause of deaths with maximum deaths

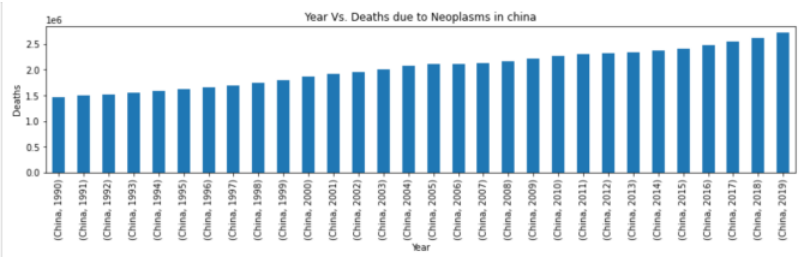
Cardiovascular Diseases	100505973
Neoplasms	61060527
Chronic Respiratory Diseases	36676826
Digestive Diseases	8924906
Lower Respiratory Infections	8525819
Road Injuries	8350399
Alzheimer's Disease and Other Dementias	5381846
Self-harm	5078550
Cirrhosis and Other Chronic Liver Diseases	4918899
Neonatal Disorders	4353666



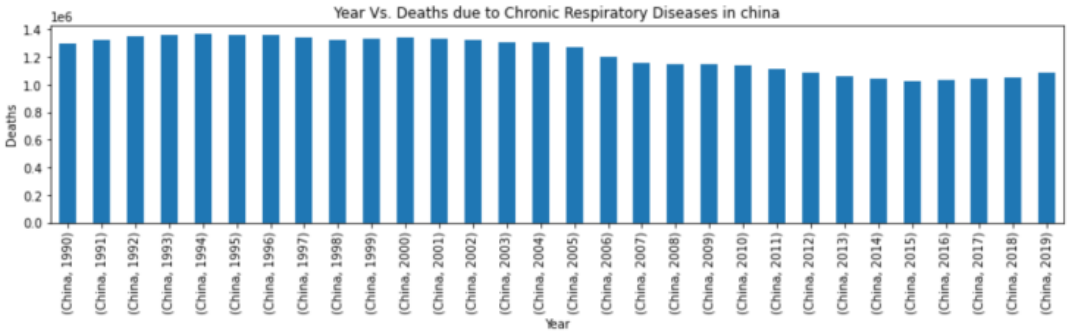
3. Cardiovascular Diseases in China:
   Total deaths due to cardiovascular diseases are 100505973 in China. It increases rapidly from year 1990 to 2019.This disease affects the heart and blood vessels. It is mainly caused by increased smoking, high blood pressure, high cholesterol, unhealthy diet, lack of exercise and obesity.



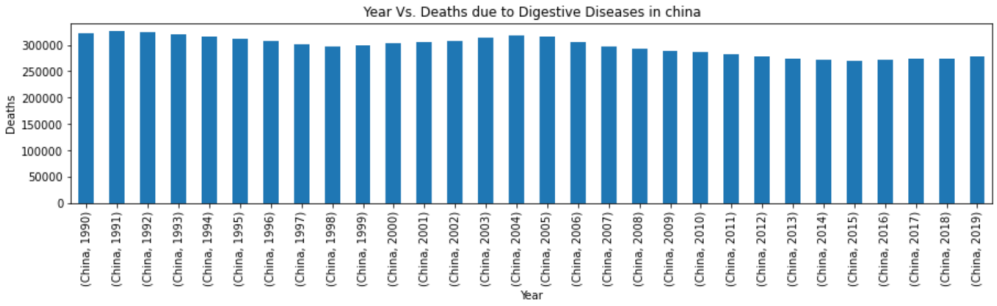
4. Neoplasm Disorder:
   Total numbers of Deaths due to neoplasm are 61060527 in china. It increases from year 1990 to 2019 . Neoplasm is a type of abnormal and excessive growth of tissue.



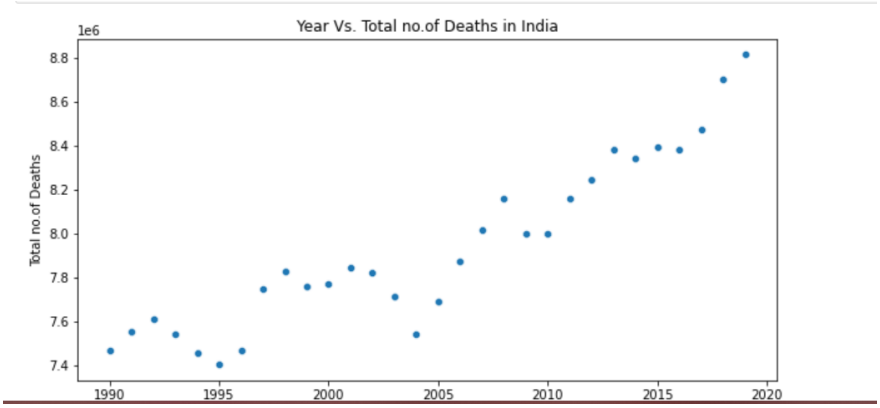
5. Chronic Respiratory Diseases:
   Total number of Deaths due to chronic respiratory diseases are 36676826.It is highly constant throughout the year with very few drop of deaths after 2004.



6. Digestive Diseases:
   Total number of deaths due to Digestive diseases is 8924906.It is almost constant throughout the year from 1990 to 2019.

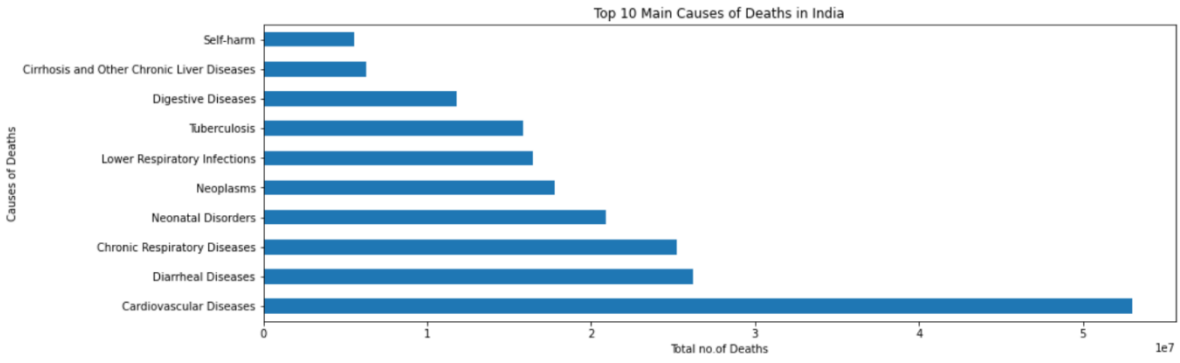


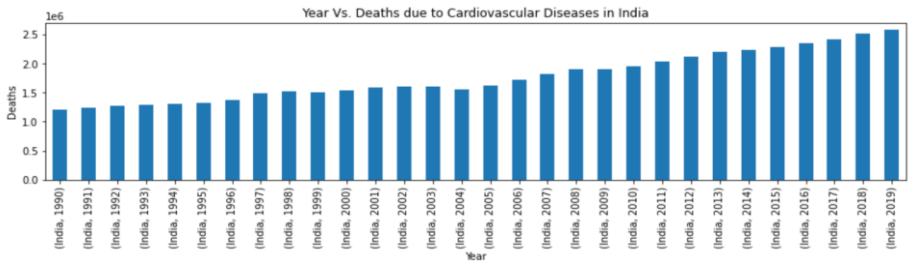
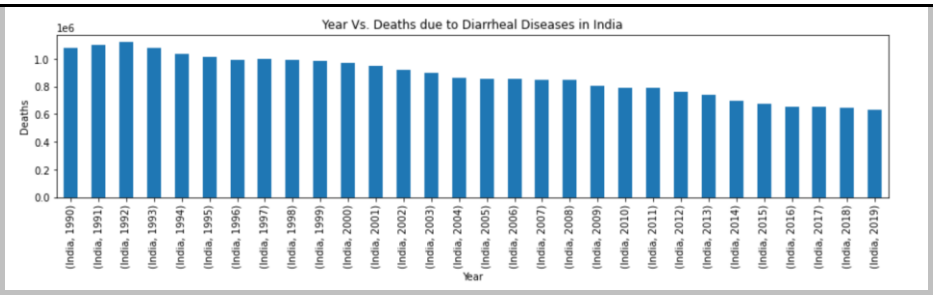
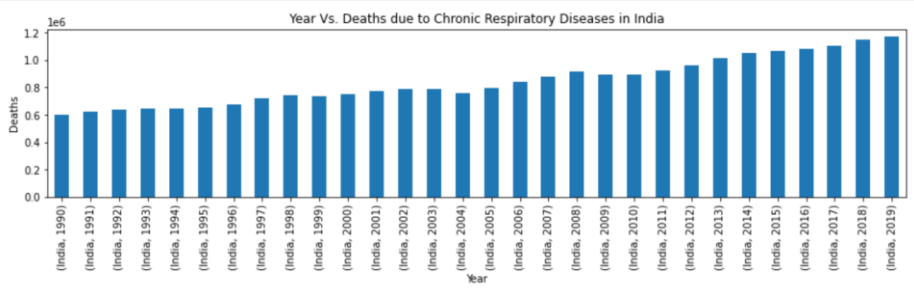
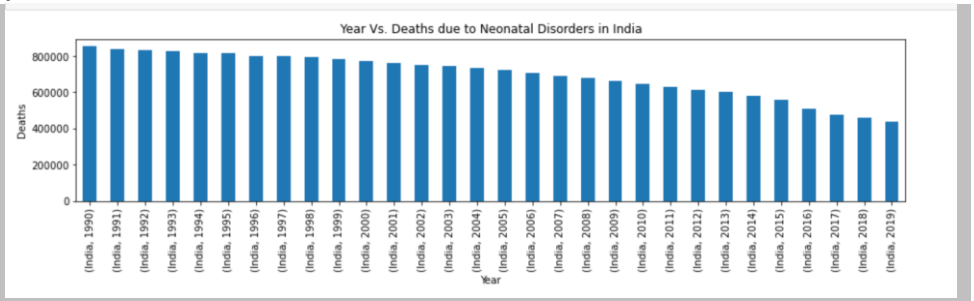
- INDIA:
  - Deaths in India from 1990 to 2019



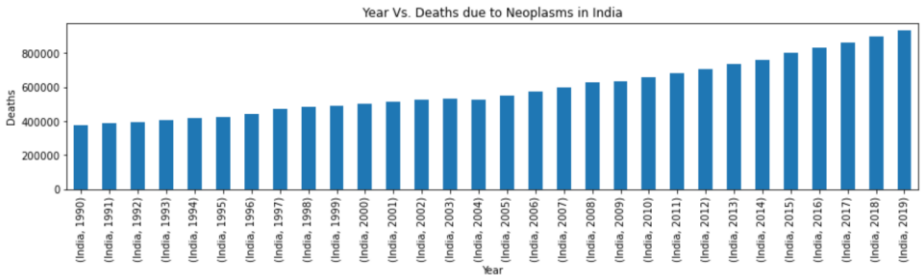
  - Top 10 main cause of Death in India:

```
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
```



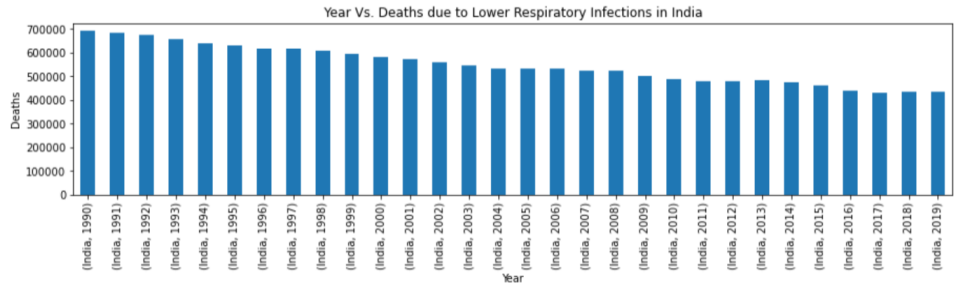
Cardiovascular Diseases	 It is a heart disease that rapidly increases from year 1990 to 2019.Total Cardiovascular Diseases cause death in India are 52994710.
Diarrheal Diseases	 Total causes of deaths due to diarrheal diseases are 26243547.This diseases is 2nd most cause of deaths in India with extremely high during 1990 then decreases slowly till 2019.
Chronic Respiratory Diseases	 Chronic respiratory diseases are chronic diseases of the airways and other parts of lungs. Deaths increases from year 1990 to 2019.
Neonatal Diseases	 Neonatal Diseases is related to newborn children. It is decreases rapidly from 1990 to 2019

Neoplasm



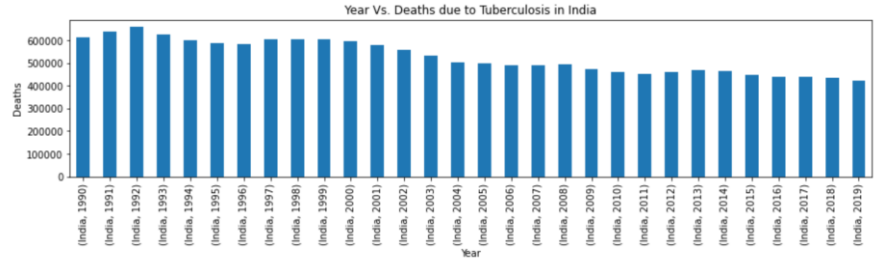
An abnormal mass of tissue that forms when cells grow and more than they should or do not die when they should. Cause of deaths due to neoplasm diseases are rapidly increases.

Lower Respiratory Infection



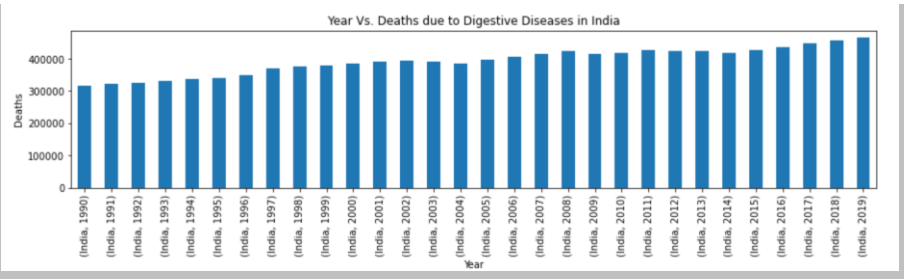
This Infection cause deaths that decreases from year 1990 to 2019

Tuberculosis

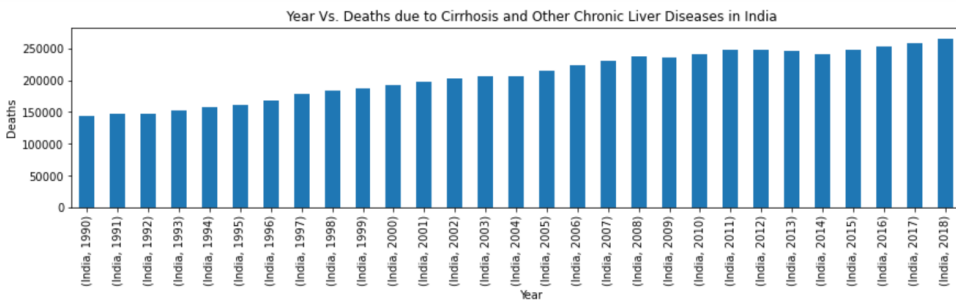
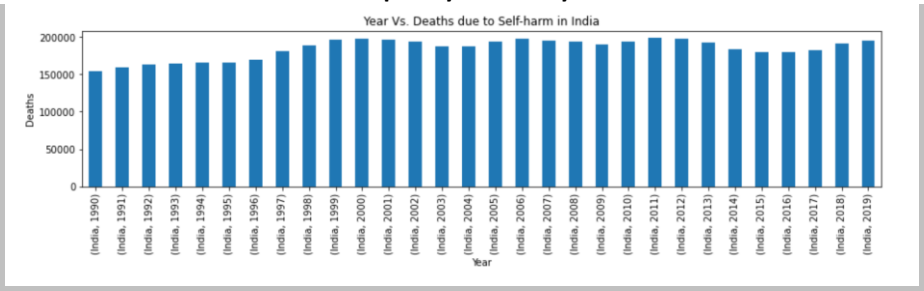


Cause of deaths due to tuberculosis is decreases from year 1990 to 2019 but it is still high number of deaths.

Digestive Diseases

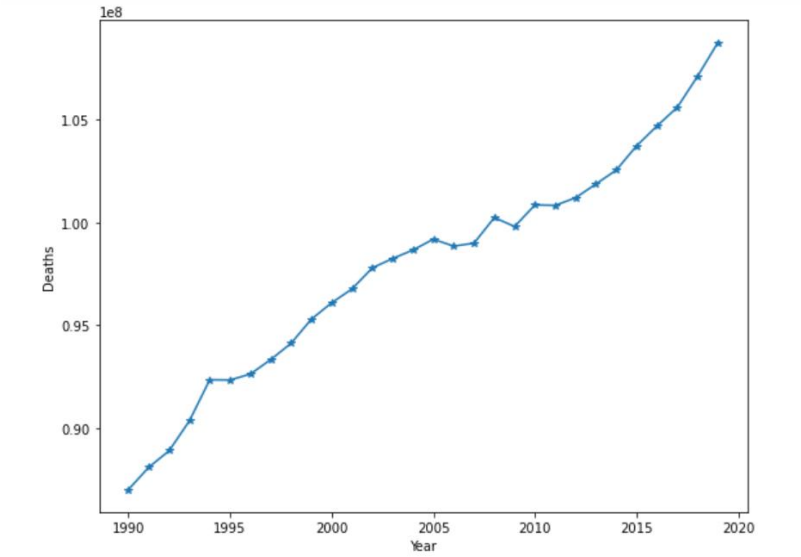


Causes of deaths due to digestive diseases are increases from year 1990 to 2019.

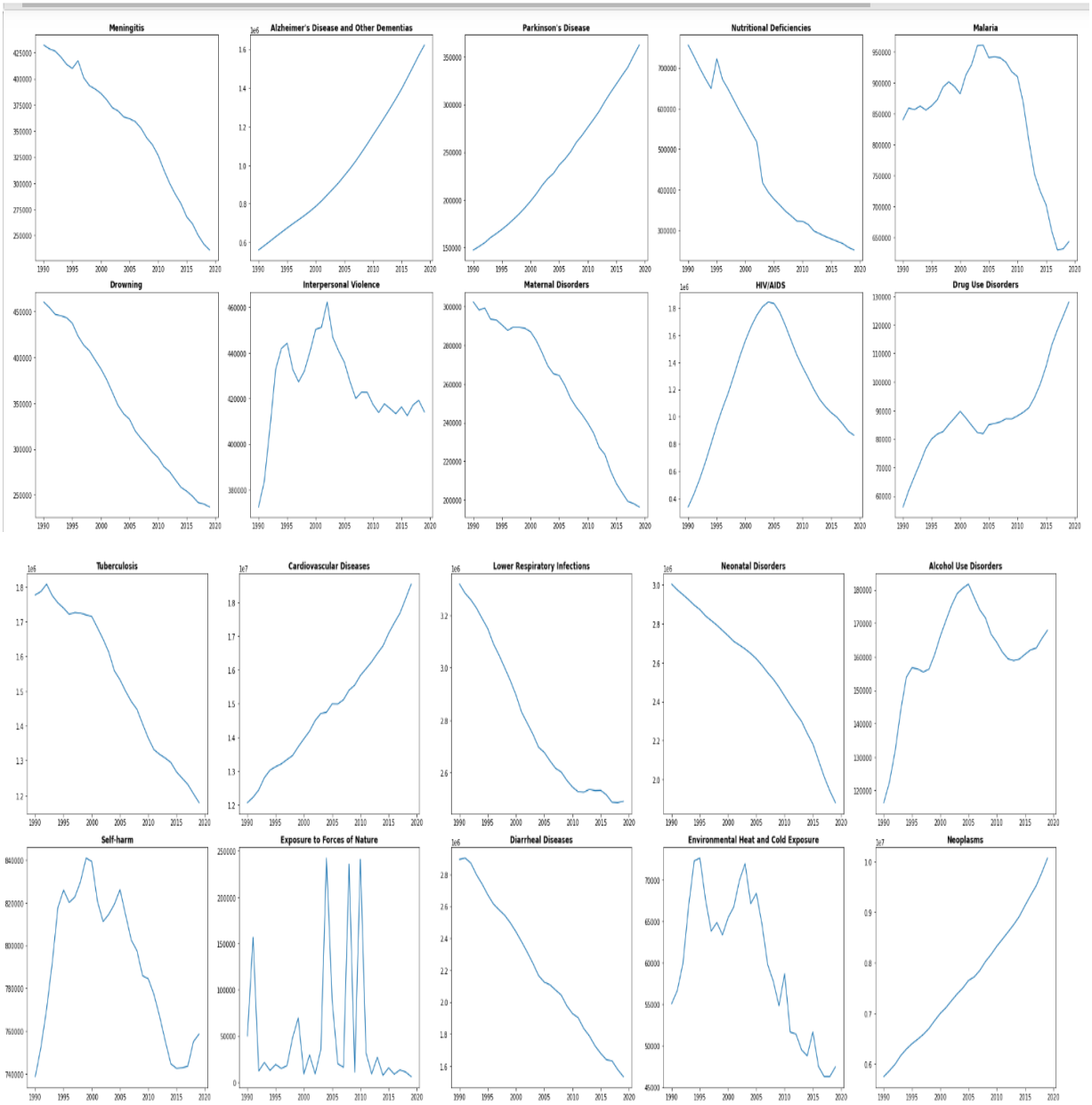
Cirrhosis and other chronic liver infection	 Cause of deaths due to cirrhosis and other chronic liver infection is increases rapidly from year 1990 to 2019.
Self-harm	 Cause of deaths due to self harm is extremely high throughout the year from 1990 to 2019.

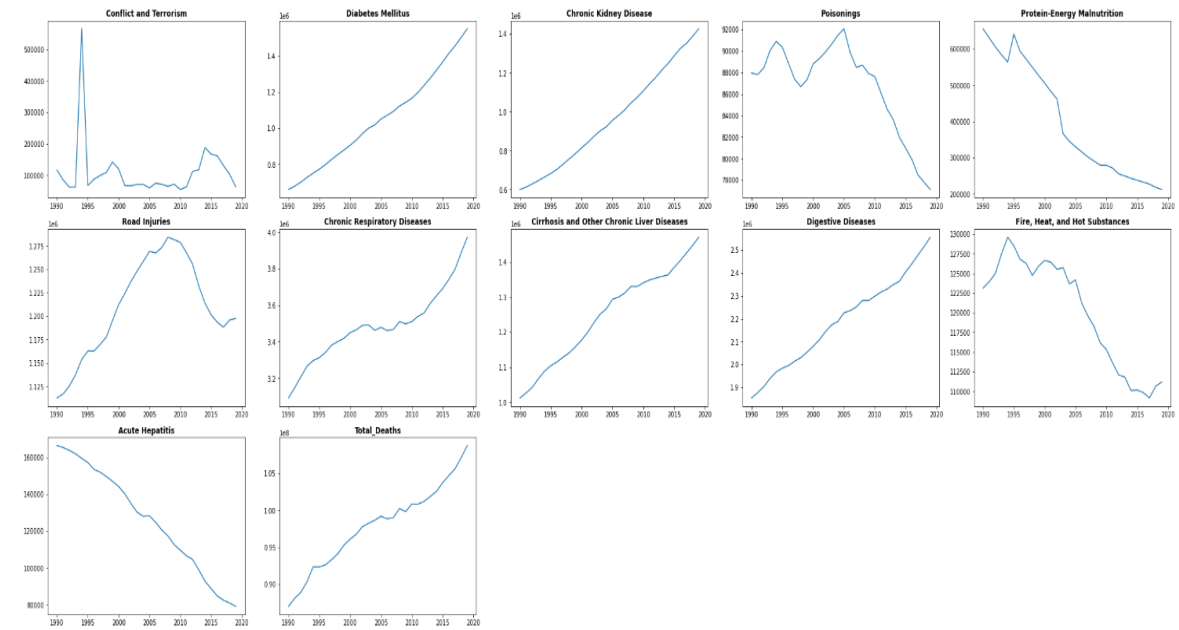
Year with maximum Number of Deaths:

This graph shows maximum number of deaths in 2020.



Year-wise Trend of Diseases:





These Graphs shows the trends of diseases from year 1990 to 2019.

Cause of Deaths that **decreases** are:

-Meningitis

-Nutritional Deficiencies

-Malaria

-Drowning

-Maternal Disorders

-Tuberculosis

-Lower Respiratory Infections

-Neonatal Disorders

-Diarrheal Diseases

-Poisonings

-Protein Energy Malnutrition

-Fire, heat and hot substances

-Acute Hepatitis

Cause of Deaths that **increases** are:

-Alzheimer's Disease and other dementias

-Parkinson's Disease

-Drug use disorders

-Cardiovascular Diseases

-Neoplasm

-Diabetes Mellitus

-Chronic Kidney Diseases

-Chronic Respiratory Diseases

-Cirrhosis and other Chronic liver Diseases

-Digestive Diseases