PROJECT REPORT ON:

"CAUSE OF DEATH"

SUBMITTED BY

Prakash kumar sinha

ACKNOWLEDGMENT

I would like to express my special gratitude to "Flip Robo" team, who has given me this opportunity to deal with a beautiful dataset and it has helped me to improve my analytical skills.

A huge thanks to my academic team "Data trained" who helped me learn and nurtured me through these months. Last but not the least my parents who have been my backbone in every step of my life.

Contents:

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

Analysis :

Importing Libraries

```
In [10]:
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib as mpl
import klib
import warnings
warnings.filterwarnings('ignore')
```

```
In [11]:
df = pd.read_csv("cause_of_deaths dataset.csv")
pd.set_option('display.max_columns', None)
df
```

Imported the necessary libraries and csv data set

Structure of the Dataset

Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinsons Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	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
Somalia	SOM	2008	4154	296	89	8756	1634	756	1817	...	1863	1346	494	8598	2729	2973	2851	4535	507	1110
Cote d'Ivoire	CIV	2018	2367	1137	371	809	27758	609	2323	...	3450	3398	377	758	3404	3638	3958	6661	520	234
Turkmenistan	TKM	1999	146	295	54	23	1	481	311	...	372	477	58	11	446	613	1242	1507	352	100
Solomon Islands	SLB	2005	24	14	14	33	220	101	44	...	277	96	14	33	127	282	101	152	17	13
Lebanon	LBN	1994	50	418	90	13	0	73	279	...	462	650	33	12	370	737	377	586	89	163
Papua New Guinea	PNG	2018	609	387	191	206	945	378	972	...	4505	716	150	202	2185	6954	643	1336	708	91
Kuwait	KWT	2019	12	425	58	1	0	23	42	...	365	322	19	0	583	230	204	338	34	1
Saint Kitts and Nevis	KNA	2015	1	8	3	4	0	3	14	...	26	26	0	3	7	9	10	19	2	0
Italy	ITA	2006	216	34184	5876	469	0	419	666	...	18535	11204	100	325	7154	26286	13429	26019	436	97
Uganda	UGA	2004	6737	1048	261	6462	57092	793	1634	...	3518	2372	548	6326	4804	4639	4181	7092	595	392

6120 rows × 34 columns

Data set having 6120 rows and 34 columns

Columns name of data set

['Country/Territory', 'Code', 'Year', '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'],

There is no null values present in the data set

Out of 34 columns , 32 columns are of int64 type and only two ['Country/Territory', 'Code'] are of object type.

CLEANING OF DATA FRAME—

Shape of cleaned data frame remains same as earlier..

Only memory consumed by data frame is reduced to around 50 percent

Replacing long country name of --

Democratic Republic of Congo', To 'Congo', And

'Saint Vincent and the Grenadines', To 'St. Vincent & the Grenadines'

DATA ANALYSIS

I want to take a look at the total global deaths and which countries lead in total deaths.

This requires aggregating across the DataFrame so we'll need to drop the 'Year' column.

Total Global Deaths from 1990-2019

Let's make a new column called 'Total Deaths'.

Total death

Total global deaths from 1990-2019:--- 1,468,134,716

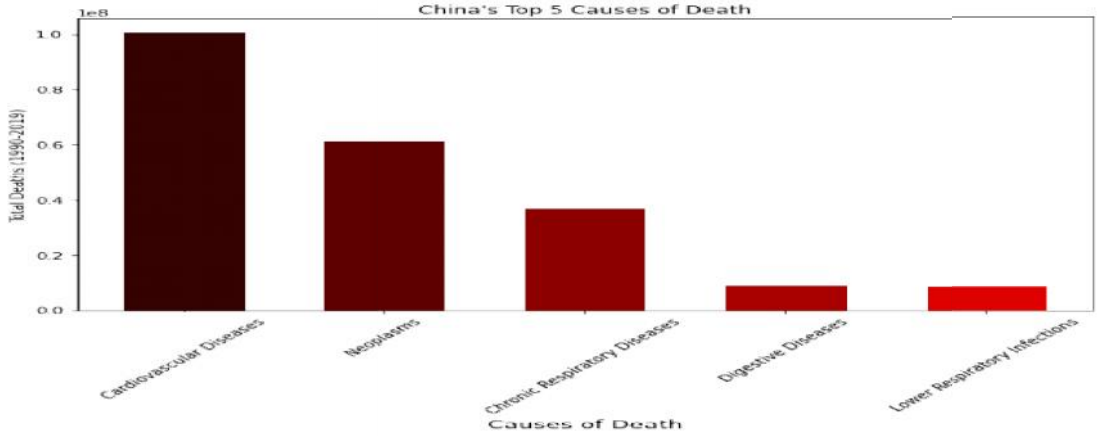
TOP 10 COUNTRIES WITH THE MOST DEATHS FROM 1990-2019

1	CHINA	265408106.0
2	INDIA	238158165.0
3	UNITED STATE	71197802.0
4	RUSSIA	59591155.0
5	INDONESIA	44046941.0
6	NIGERIA	43670014.0
7	PAKISTAN	38151878.0
8	BRAZIL	32674112.0
9	JAPAN	31922807.0
10	GERMANY	25559667.0

As china having world's highest population so the number of death is also highest in the world, followed by India.

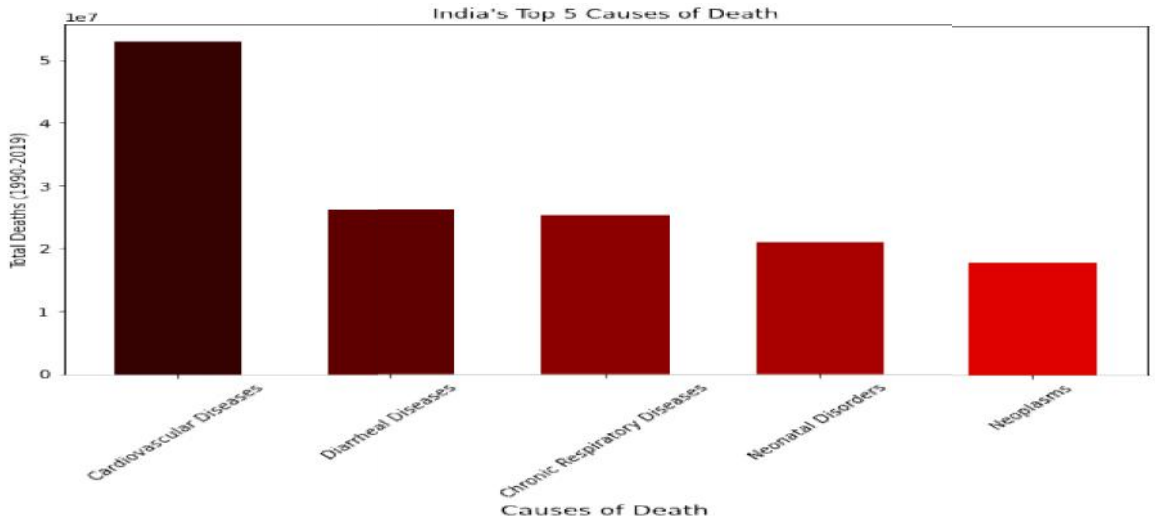
Now we are going to check the Reason of Death in top 10 countries-

1 – China-



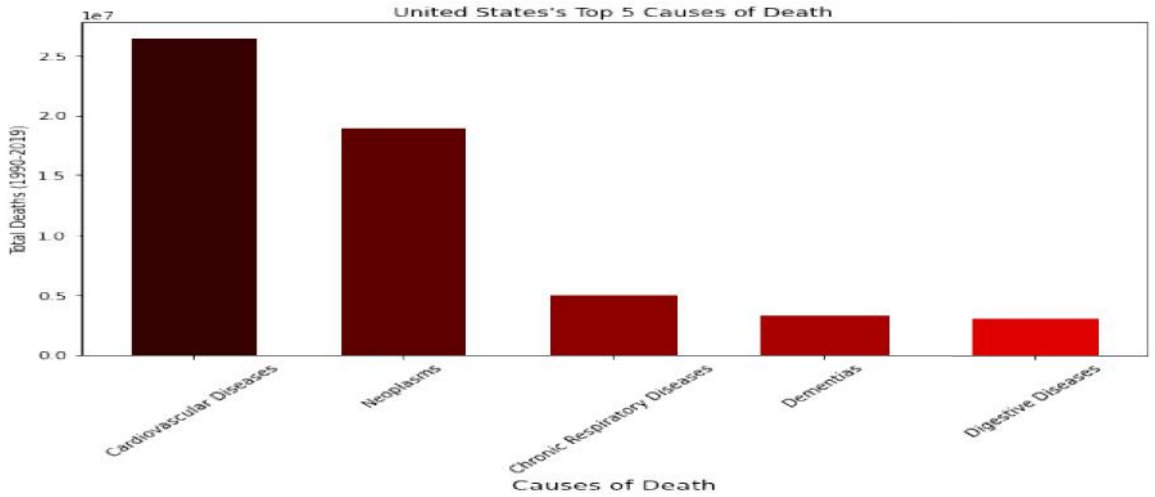
Top 5 cause of death in china – Cardiovascular Disease
Neoplasm
Chronic respiratory disease
Digestive disease
Lower respiratory infection

2- India-



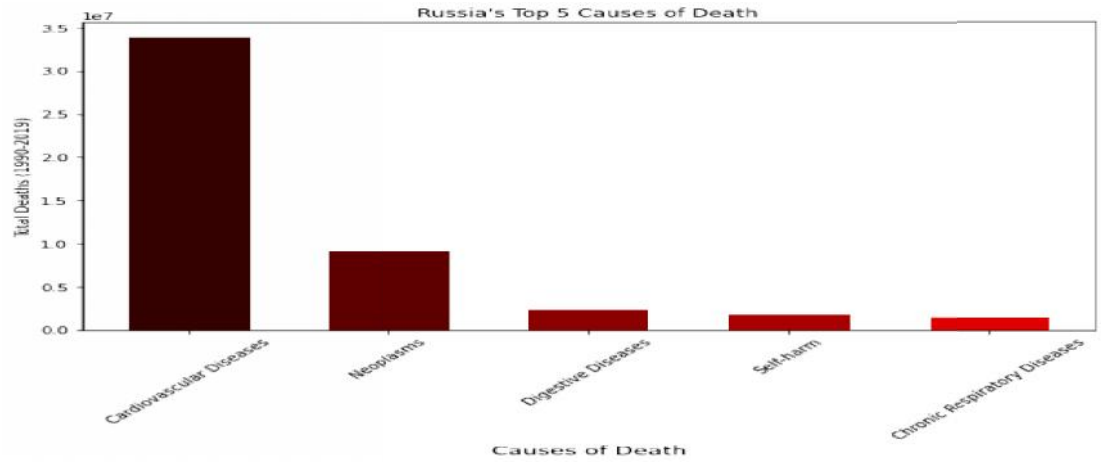
Top 5 cause of death in India – Cardiovascular Disease
Diarrheal disease
Chronic respiratory disease
Neonatal disorders
Neoplasm

UNITED STATE



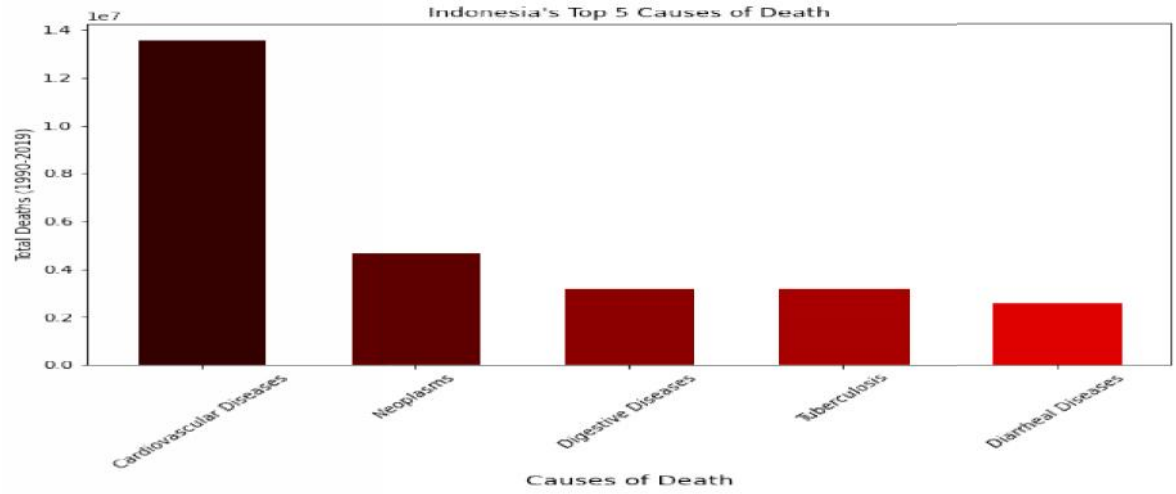
Top 5 cause of death in United States – Cardiovascular Disease
Neoplasm
Chronic respiratory disease
Dementias
Digestive disease

RUSSIA



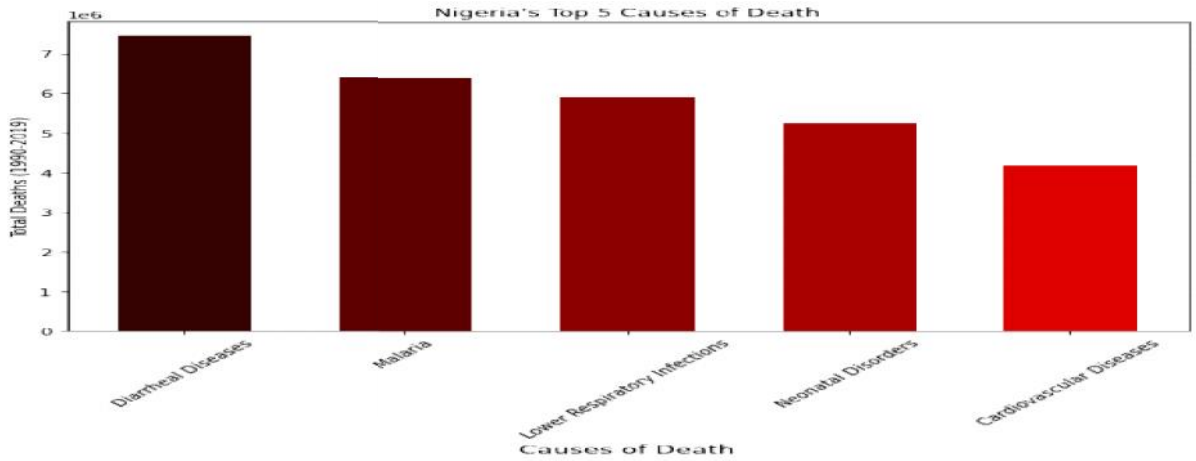
Top 5 cause of death in Russia – Cardiovascular Disease
Neoplasm
Digestive disease
Self harm
Chronic respiratory disease

INDONESIA



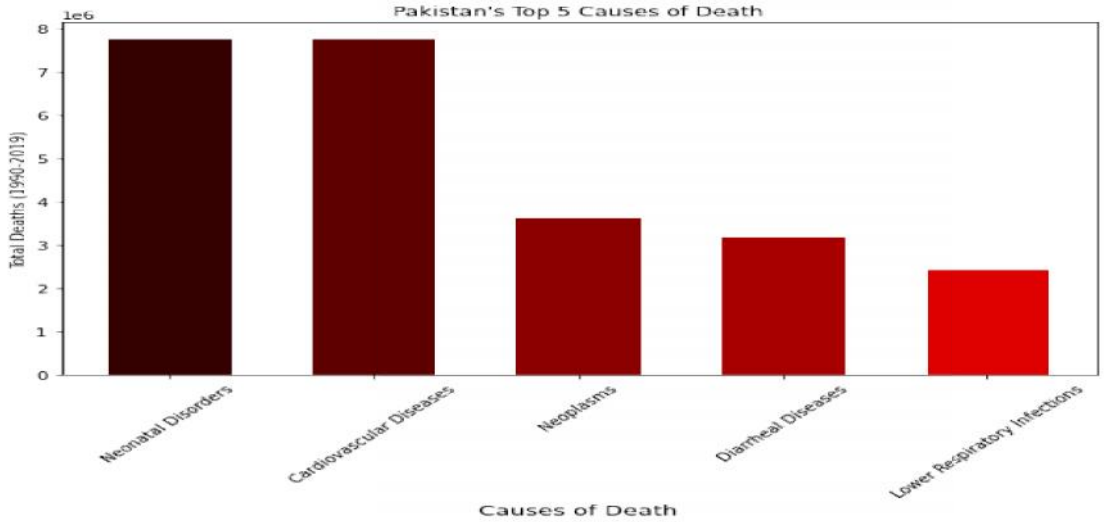
Top 5 cause of death in Indonesia – Cardiovascular Disease
Neoplasm
Digestive disease
Tuberculosis
Diarrheal disease

NIGERIA



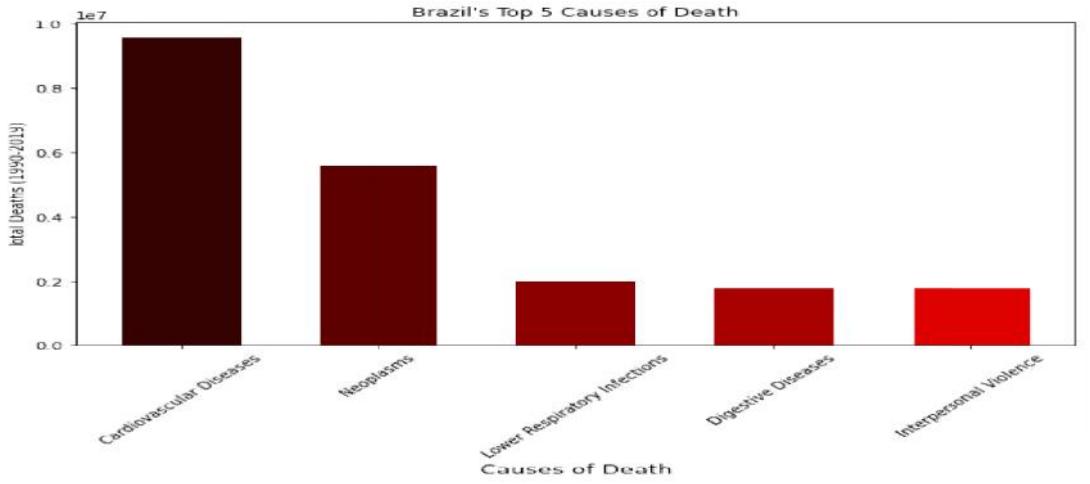
Top 5 cause of death in Nigeria – Diarrheal disease
Malaria
Lower respiratory infection
Neonatal disorder
Cardio vascular disease

PAKISTAN



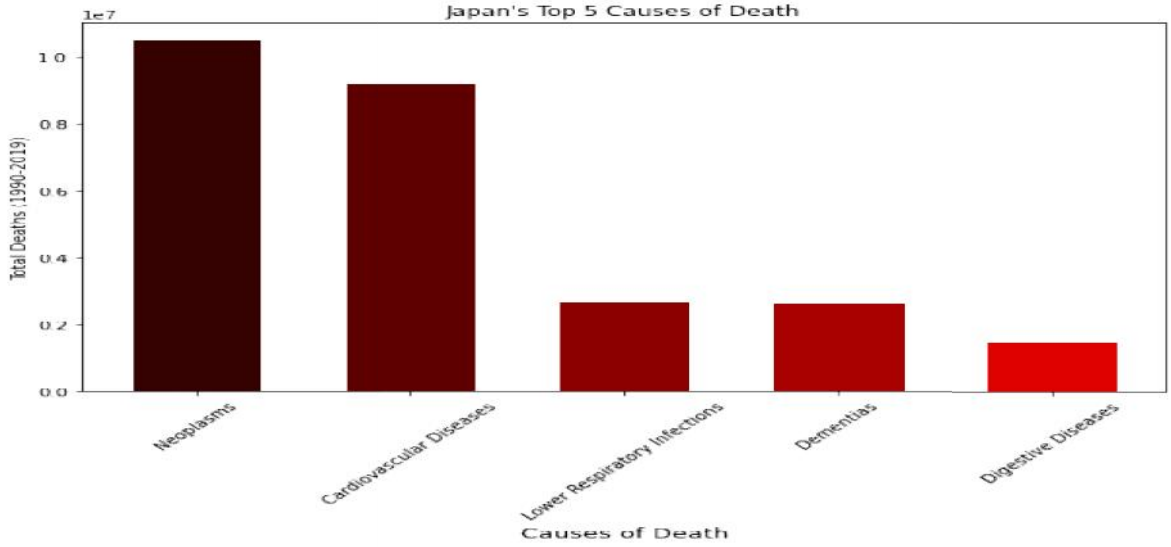
Top 5 cause of death in Pakistan – Neonatal disorder
Cardio vascular disease
Neoplasm
Diarrheal disease
Lower respiratory infection

BRAZIL



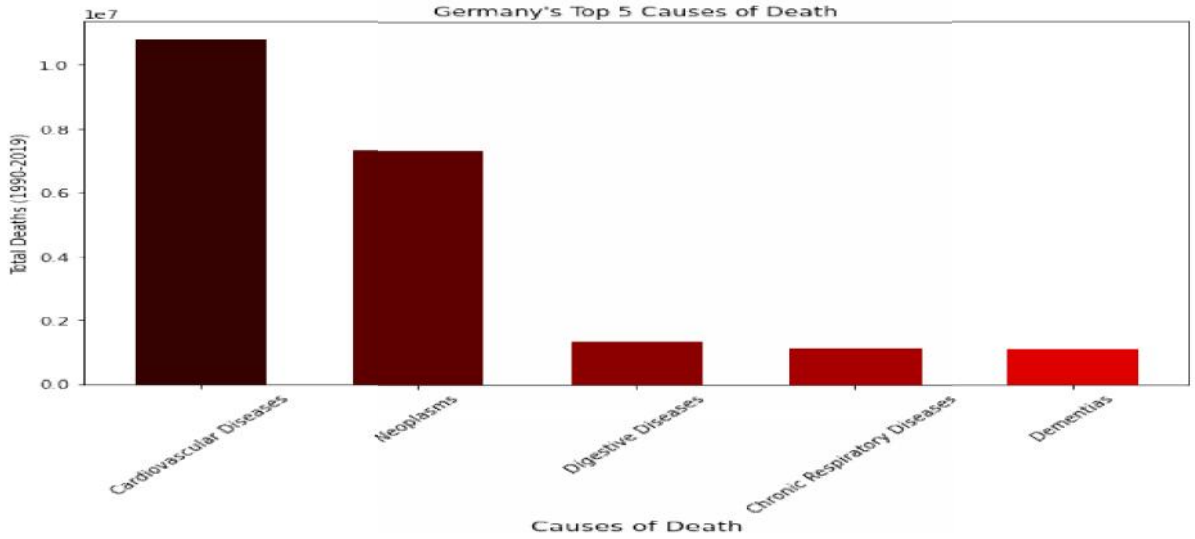
Top 5 cause of death in Brazil– Cardio vascular disease
Neoplasm
Lower respiratory infection
Digestive disease
Interpersonal violence

JAPAN



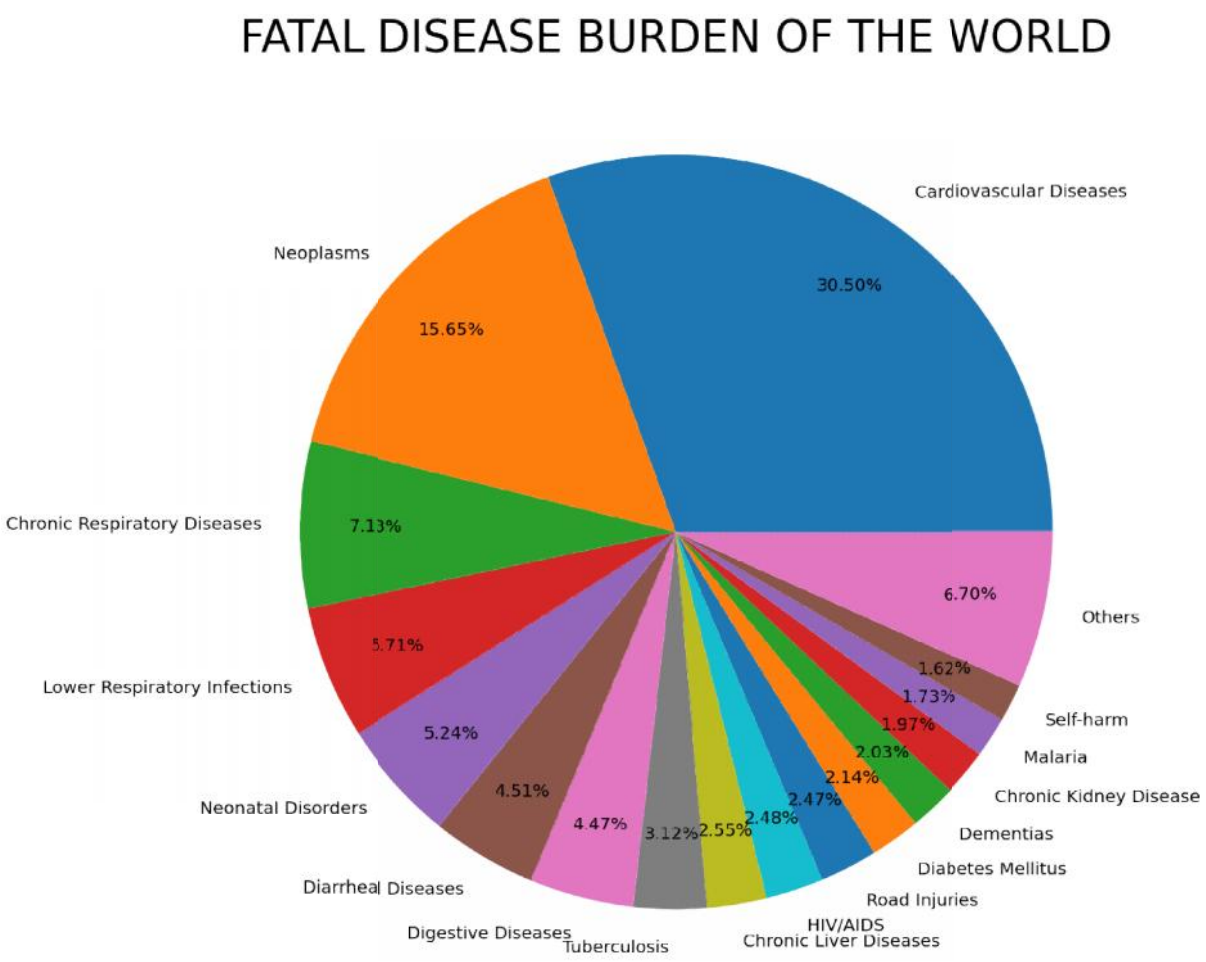
Top 5 cause of death in Japan– Neoplasm
Cardio vascular disease
Lower respiratory infection
Dementias
Digestive disease

GERMANY



Top 5 cause of death in Germany– Cardio vascular disease
Neoplasm
Digestive disease
Chronic Respiratory disease
Dementias

NOW REASON OF DEATH IN THE WORLD

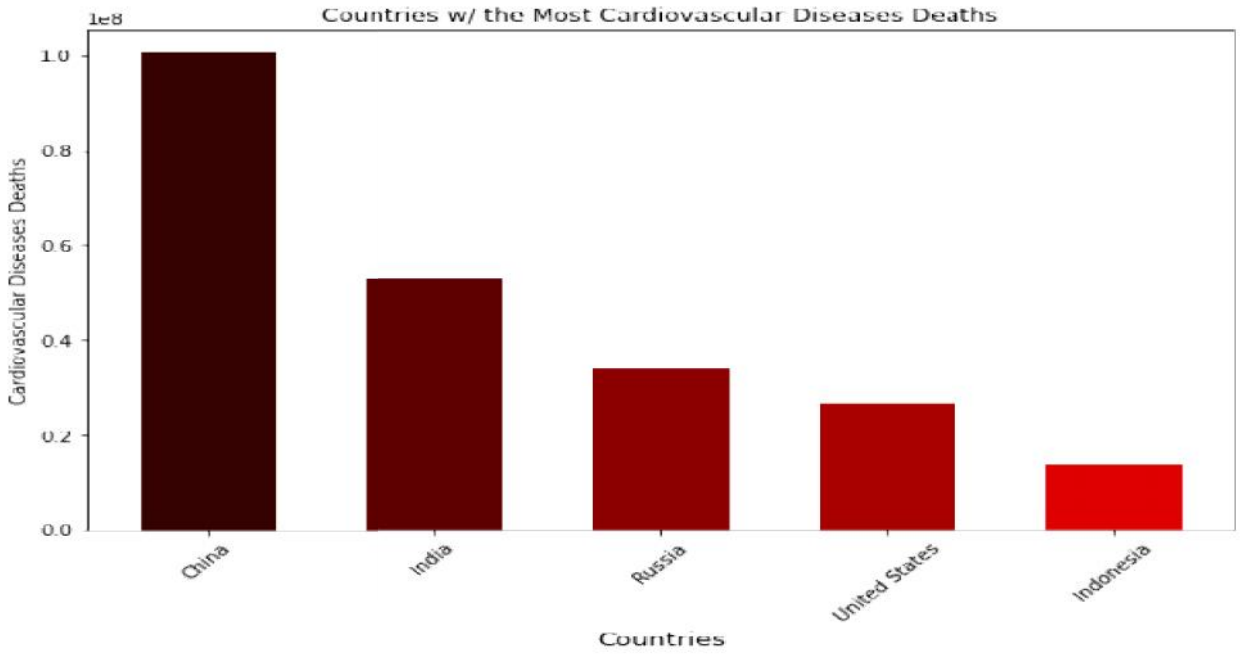


NOW TOP 5 REASONS BEHIND DEATH IN THE WORLD ARE

Cardio vascular disease (30.5%)
Neoplasm (15.65%)
Chronic Respiratory disease (7.13%)
Lower respiratory infection (5.71%)
Neonatal disorders (5.24%)

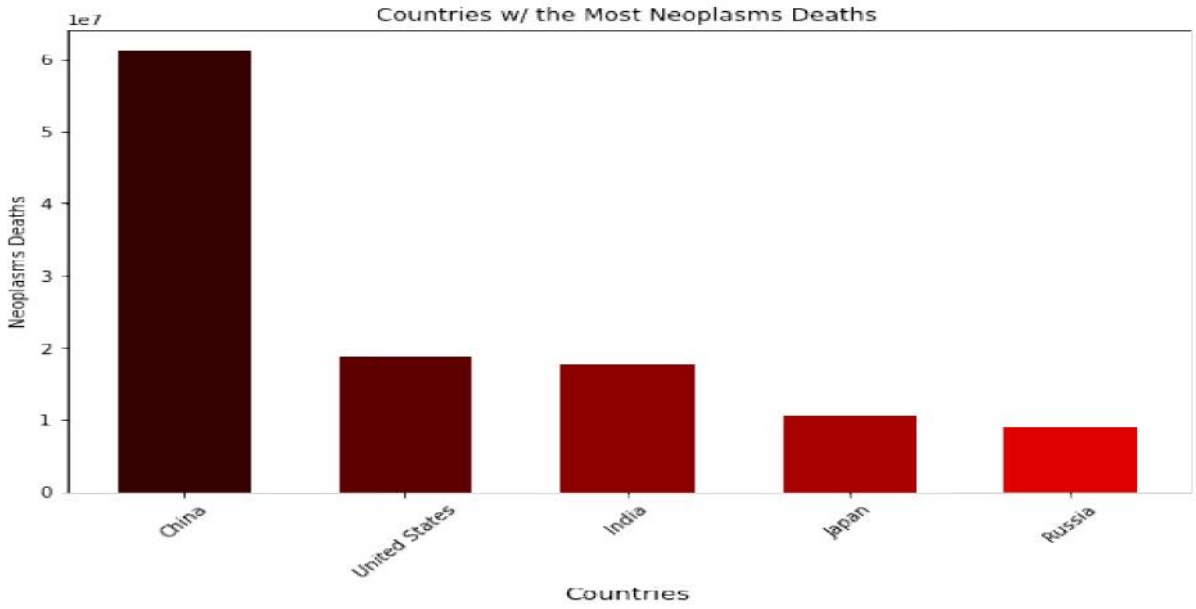
NOW TOP 5 FAVORABLE COUNTRIES FOR THE TOP 5 KILLERS

CARDIO VASCULAR DISEASE (30.5%)



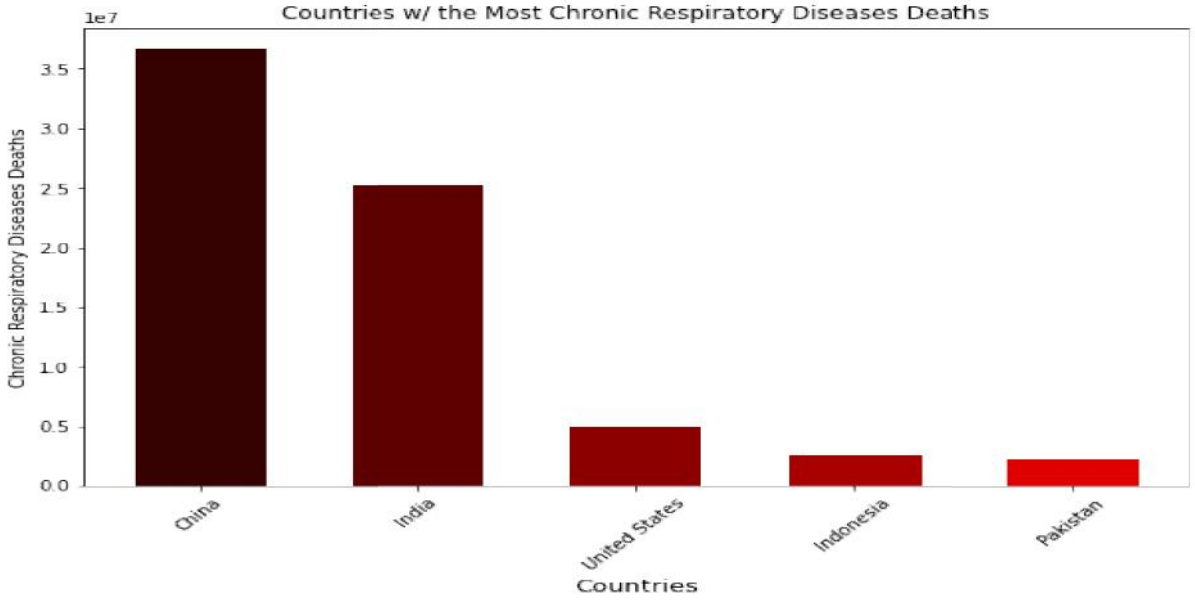
China is the country where Cardiovascular disease occurred highest followed by India

NEOPLASM (15.65%)



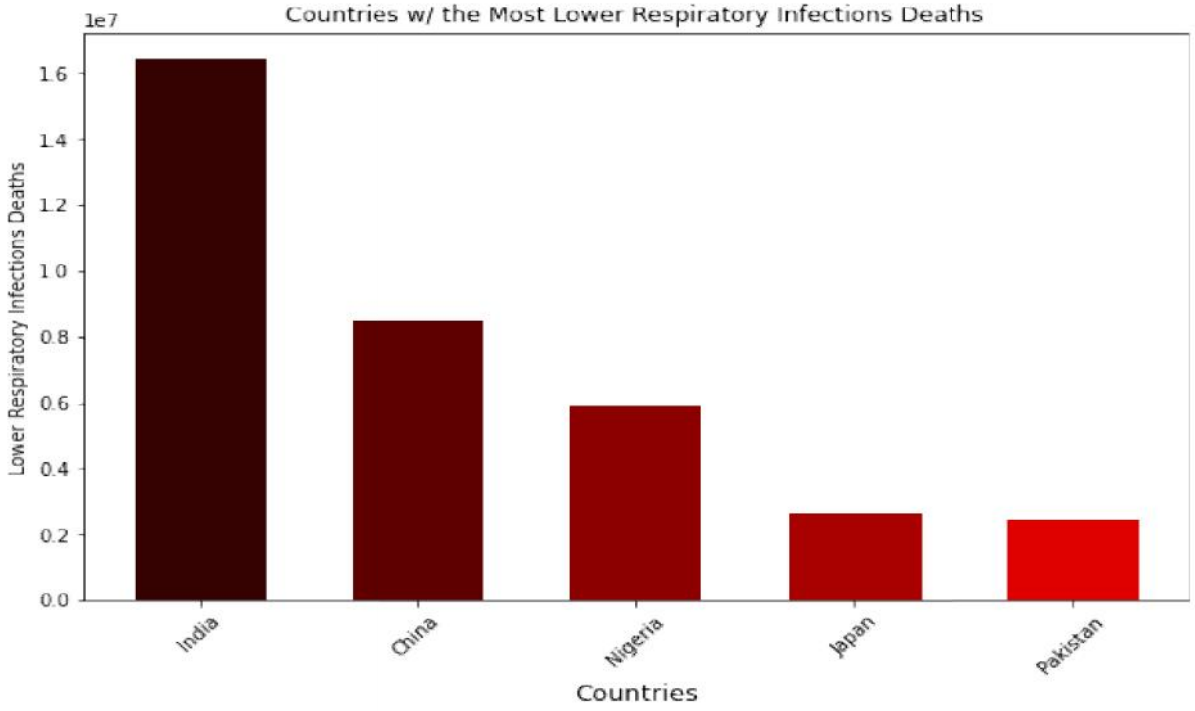
China is the country where Neoplasm occurred highest followed by United states and India

CHRONIC RESPIRATORY DISEASE (7.13%)



China is the country where Chronic Respiratory disease occurred highest followed by India

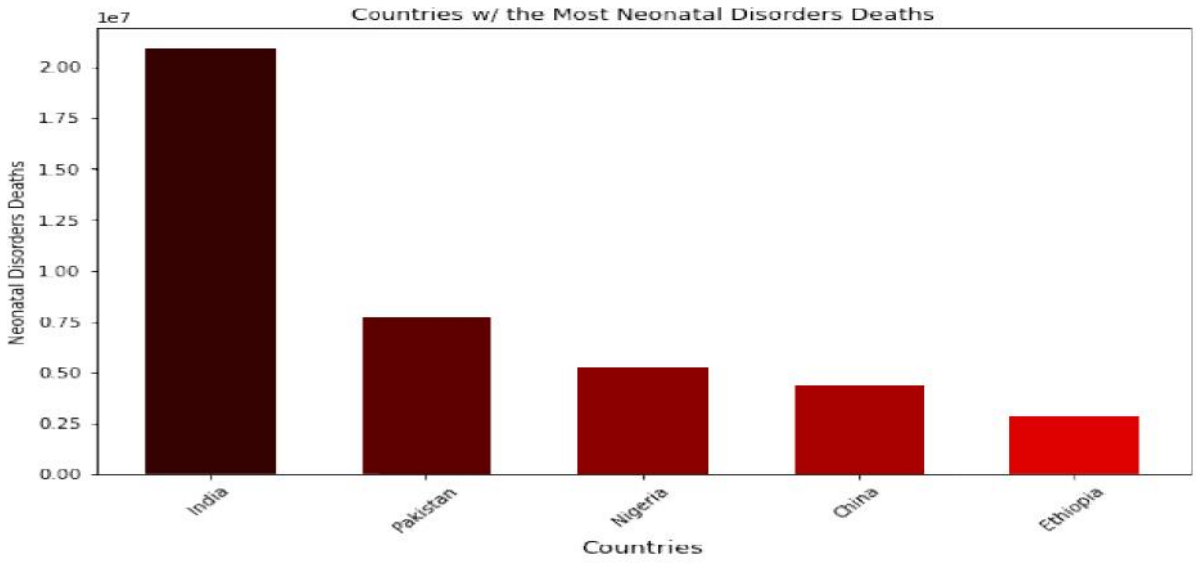
LOWER RESPIRATORY INFECTION (5.71%)



.

India is the country where Lower Respiratory infections occurred highest followed by China

NEONATAL DISORDERS (5.24%)



India is the country where Neonatal disorder occurred highest followed by Pakistan.

Conclusion-

China, India and USA face the largest brunt of deaths due to disease in the world Cardiovascular disease, Neoplasm, and lower respiratory Tract infection . As these are the top killers in the world