FLIP ROBO

Cause of Death

Submitted By:

Naveen Kumar Ranganathan

ACKNOWLEDGMENT

I would like to thank Flip Robo Technologies who gave me this opportunity to work on **Cause of Death** Project which helped me

to gain in-depth knowledge of EDA Analysis in Data Science to derive insights for organizational goals to meet business needs.

Also, I have utilized a few external resources which helped me to complete this project. All the external resources that were used in creating this project are listed below:

https://ourworldindata.org/causes-of-death-treemap

https://github.com/

https://www.w3schools.com/

https://matplotlib.org/

https://www.geeksforgeeks.org/

https://seaborn.pydata.org/

INTRODUCTION

Business Problem Framing

To make progress towards a healthier world we need to have a good understanding of what health problems we face today. What do people die from?
With the dataset we can identify the major problems which causes death.

Conceptual Background of the Domain Problem

Case Study on Cause of Death:

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

In everyday language we sometimes say that a person died of 'old age' or that they 'died of natural causes'. But there is always an underlying cause that stopped their body from functioning. When the cause of death is not recorded then researchers rely on models to estimate the cause.

Motivation for the Problem Undertaken

What the world dies from is not what is reflected in the media. Some major causes of deaths receive very little attention. Analysing and studying the dataset will give us good understanding ofwhat health problems we face today. We need a good sense of the relative importance of different causes of death and we need to focus our efforts on the biggest causes.

Analytical Problem Framing

Data Set 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.

Getting insights about the dataset:

Imported below required libraries to work on the dataset

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
from sklearn import preprocessing
import warnings
warnings.filterwarnings('ignore')
```

- Using Pandas the csv file of the dataset was imported.

```
# Importing Dataset

pd.set_option('display.max_columns',None)
pd.set_option('display.max_rows',None)
df = pd.read_csv('cause_of_deaths dataset.csv')
df
```

Analysis:

- To identify the dimension of our dataset, used shape() method and found that our dataset contains 6120 columns and 34 rows
- Used dtypes method to identify the data types of the dataset and found all the columns are int type except Country/Territory and Code which are object type
- Used isnull().sum() to identify null values and we don't find any null values in our dataset
- Used nunique() function to identify the uniqueness of the value present in the dataset
- Dataset contains data from 1990 to 2019
- Dropped column code as we have Country/Territory both denotes the same information
- Added Total no of Deaths column to understand total deaths documented by year and country

In [9]:
```
# Dropping the column Code

df2 = df.drop(columns=['Code'],axis=1)
df2.shape
```

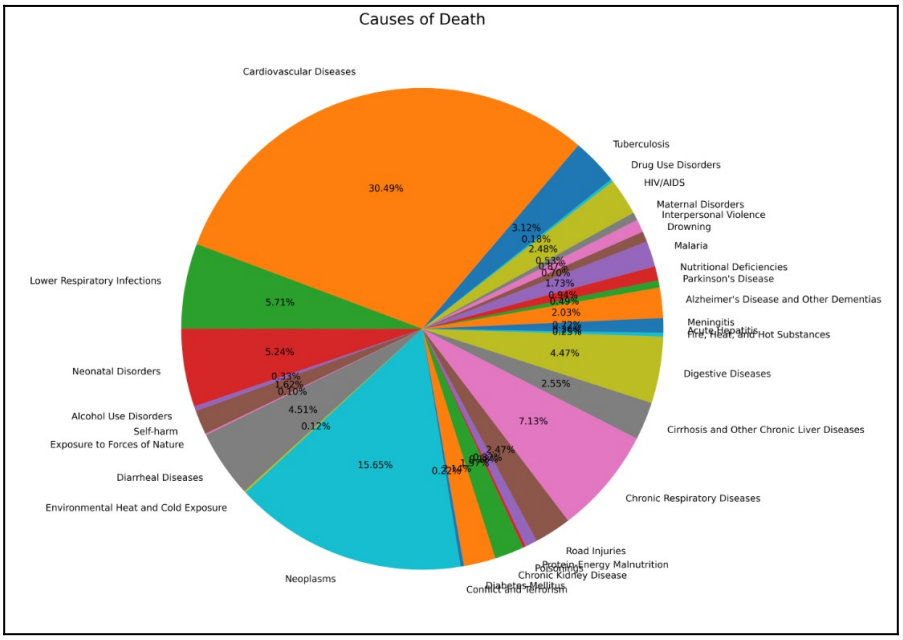
Out[9]: (6120, 33)

In [11]:
```
# Adding new column Total no of deaths

df2['Total no of Deaths'] = df2[list(df2.columns[2:33])].sum(axis=1)
df2.head(10)
```

Data Visualization:

Percentage was calculated to identify which causes maximum deaths

```
# Checking percentage of death by its causes

percentage=[]
for i in df.columns[3:34]:
    per = round((df[i].sum()/df.sum()[3:34].sum()*100),2)
    percentage.append(per)

df_tot = pd.DataFrame({"Disease":df.columns[3:34],"Overall Death Rate":percentage})
df_tot["Overall Death Rate"].sort_values(ascending = False)
df_tot
```



- Cardiovascular Diseases deaths is the highest in the world with 30.49% across all the deaths documented followed by Neoplasms with 15.65%
- Exposure to Forces of Nature 0.10%, Environmental Heat and Cold Exposure 0.12% and Poisonings 0.18% are the lowest in deaths

Based on our dataset we can group the causes of deaths into three large categories

Non-communicable diseases

Diseases which cannot be passed from person to person.

Communicable or infectious diseases

Diseases that are caused by a pathogen which can be passed from person to person.

Injuries

This is a very wide category which includes accidents – such as car crashes and falls from stairs or ladders – as well as intentional injuries like homicides, war deaths, and suicides.

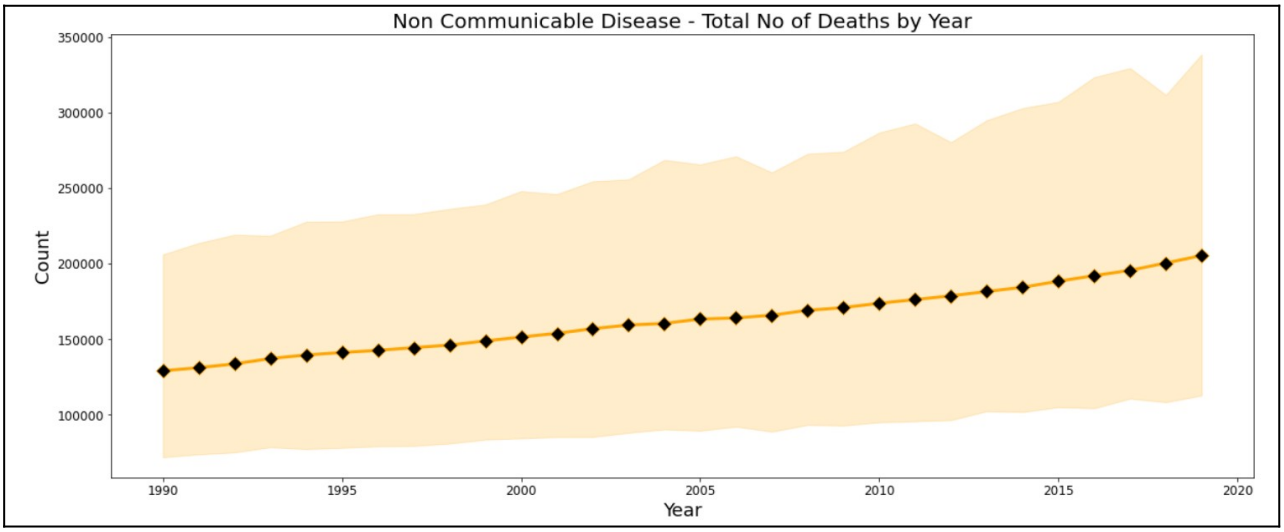
```
# Based on WHO diseases has been categorized - reference--> ourworldindata

# ncd --> Non communicable disease

# ncd = ["Meningitis","Alzheimer's Disease and Other Dementias", "Parkinson's Disease",
#        "Cardiovascular Diseases","Neoplasms","Diabetes Mellitus","Chronic Kidney Disease","Chronic Respiratory Diseases",
#       "Cirrhosis and Other Chronic Liver Diseases","Digestive Diseases","Acute Hepatitis"]

# cd --> Communicable disease

# cd = ["Nutritional Deficiencies","Malaria","Maternal Disorders","HIV/AIDS","Tuberculosis","Lower Respiratory Infections",
#      "Neonatal Disorders","Diarrheal Diseases","Protein-Energy Malnutrition"]

# inj --> Injuries

# inj = ["Drowning","Interpersonal Violence","Drug Use Disorders","Alcohol Use Disorders","Self-harm",
#        "Exposure to Forces of Nature","Environmental Heat and Cold Exposure","Conflict and Terrorism","Poisonings",
#        "Road Injuries","Fire, Heat, and Hot Substances",]
```

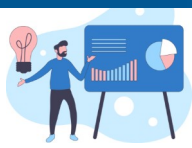
Non-communicable Diseases: "Meningitis","Alzheimer's Disease and Other Dementias", "Parkinson's Disease", "Cardiovascular Diseases","Neoplasms","Diabetes Mellitus","Chronic Kidney Disease","Chronic Respiratory Diseases","Cirrhosis and Other Chronic Liver Diseases","Digestive Diseases","Acute Hepatitis"

- Meningitis death count is reported high in India, Nigeria and low in Uganda
- Alzheimer's Disease and Other Dementias death count is reported high in China followed by United States and reported low in United Kingdom
- Parkinson's Disease death count is reported high in China, India and low in Brazil
- Cardiovascular Diseases death count is reported very high in China and India followed by Russia
- Neoplasms death count is reported very high in China and very low in France
- Diabetes Mellitus death count is reported high in India and China
- Chronic Kidney Disease death count is very high and almost equal in India and China
- Chronic Respiratory Diseases death rate is very high in China and India
- Cirrhosis and Other Chronic Liver Diseases and Digestive Diseases death counts are very high both in India and China
- Death by Acute Hepatitis in India is the highest among all the other countries

As we see the most of death by Non Communicable Disease is very high in India and China, this is due to high density of population in both the country

Below line plot chart shows us that death by Non Communicable disease is increasing yearly based on the data from 1990 to 2019

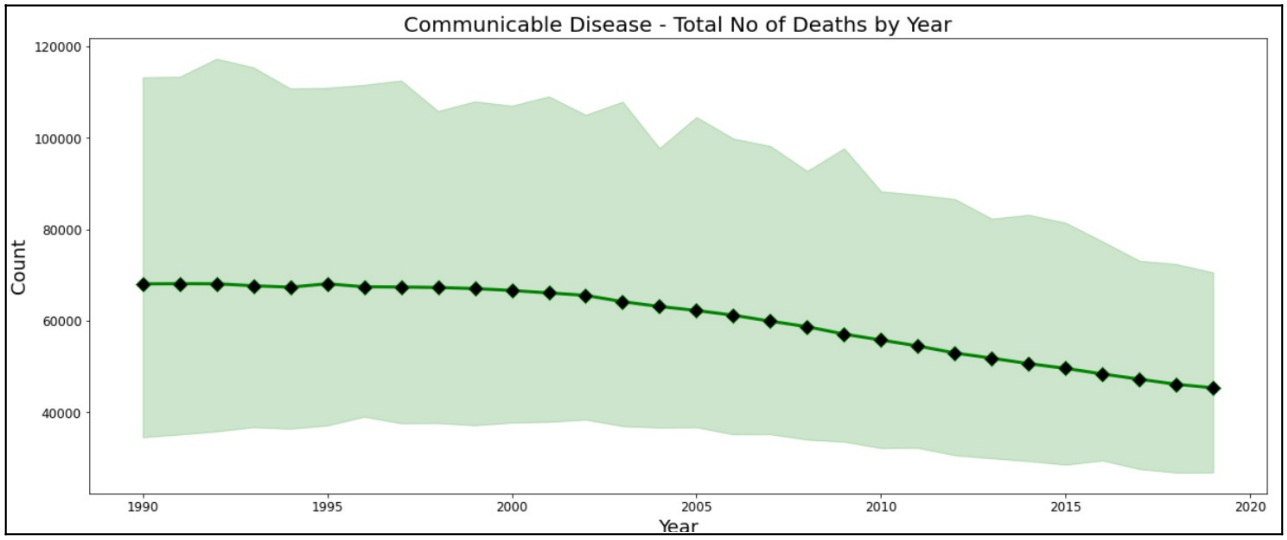




Communicable Diseases: Nutritional Deficiencies","Malaria","Maternal Disorders","HIV/AIDS","Tuberculosis","Lower Respiratory Infections","Neonatal Disorders","Diarrheal Diseases","Protein-Energy Malnutrition”

- Death by Nutritional Deficiencies count is very high in India and low in Tanzania
- Nigeria has the most death count by Malaria
- Maternal Disorders death is highest in India compared to other countries
- Death by HIV/AIDS is high in South Africa followed by Kenya
- Tuberculosis, Lower Respiratory Infections, Neonatal Disorders, Diarrheal Diseases and Protein-Energy Malnutrition death count is high in India compared to other countries

Based on the dataset we can see Communicable disease is in decreasing trend

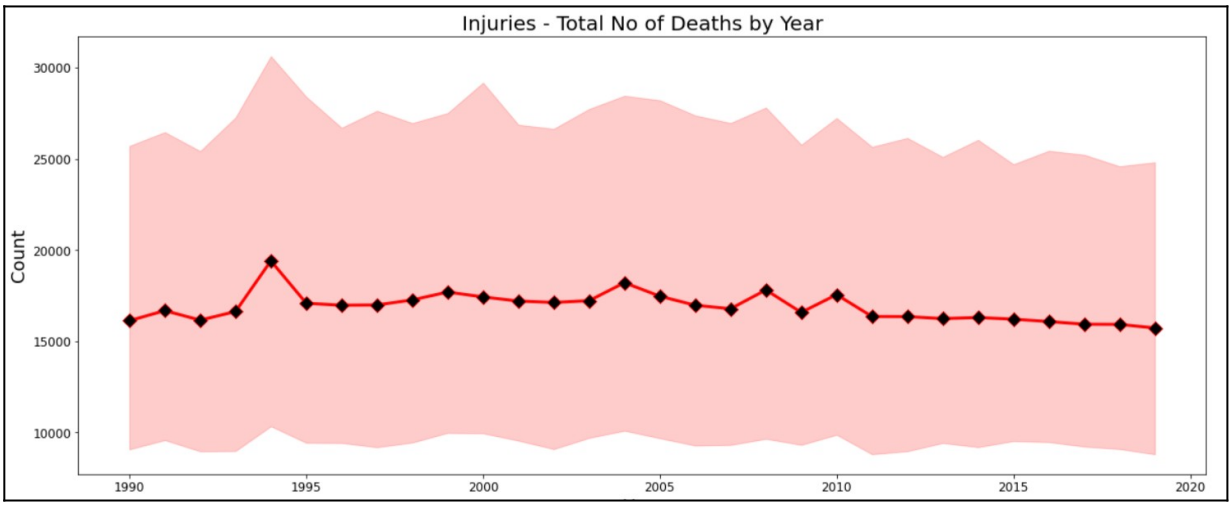


Injuries: "Country/Territory","Year","Drowning","Interpersonal Violence","Drug Use Disorders","Alcohol Use Disorders","Self-harm","Exposure to Forces of Nature","Environmental Heat and Cold Exposure","Conflict and Terrorism","Poisonings",”Road Injuries","Fire, Heat, and Hot Substances"

- Drowning, Self Harm, Road Injuries, and Fire, Heat, and Hot Substances death count are high in China and India
- Brazil got the highest death count by Interpersonal Violence
- United States and China got the highest death count by Drug Use Disorders
- Alcohol Use Disorder is high in Russia followed by India
- Haiti, Indonesia and Bangladesh death count is high by Exposure to Forces of Nature
- Environmental Heat and Cold Exposure death count is high in Russia
- Conflict and Terrorism death counts are higher in Rwanda, Syria and Iraq

- Death by Poisonings is very high in China across the globe

Death by Injuries is less when compared to Non communicable and Communicable disease, based on the dataset we can see the death counts are almost same from 1990 to 2019



CONCLUSION:

Non Communicable disease is in increasing trend and are responsible for one-in-three deaths in the world

Communicable Diseases that are caused by a pathogen which can be passed from person to person. This is where we have made most progress and still we are rapidly developing in this direction. Very few people die from rich countries compared to poorer countries. Many of these deaths can be prevented with modern technologies, such as vaccines, antibiotics, and public health infrastructure like sanitation and clean water.
Maternal deaths, the deaths of newborns, and deaths from nutritional deficiencies are often closely linked to infectious diseases. Most of these are preventable.

Some major causes of deaths receive very little attention. Violence is, fortunately, a relatively rare cause of death. While it receives a lot of media attention, more people die from diarrheal disease than from all forms of violence put together and preventing deaths from diarrhoea would save more lives than bringing an end to all violence. We should be focusing on these areas as this should be less challenging to achieve.