

# PROJECT REPORT- Causes of Death Case Study

Richard Prabhakar

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.

# PROJECT REPORT- Causes of Death Case Study

## ***Problem statement.***

\* 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.

\* We need to find the cause of death for each country each year and each disease against each other

## **Steps :-**

- check the no of rows and columns
- check the names of the columns
- Checking for null values
- Check to see the data type if numeric or not-df.info
- Draw graph to visualize if null values present
- Checking to see the mean and std deviation values of all numeric columns as we as the 25%, 50%, and 75% quintile value
  - - We see that the values in the columns are very varied and lack symmetry ,
  - - We see that in most of the columns the std number is much higher than the mean which is showing that there is no normal distribution in the columns.
  - - We see that the max values in each are really higher than the 75% quintile and we can see that this data shows that over the years the rates have come down for many of the causes due to improvements in medical science
- Printing all the unique values

## *Libraries used*

...

pandas as pd

numpy as np

matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

Import import

plotly\_express as px

from plotly.offline

import

init\_notebook\_mode

---

## DATA SET :

'cause\_of\_deaths  
dataset.csv'

---

## CONTENTS:-

6120 rows collected  
from all the countries  
based on the year and  
type of death cause

## 34 columns:-

- 1 to 3 has demographics specific feature
- 4 to 34 has causes of death

- We see that the majority of the max values in all the columns as seen before are really outliers from the normal data , we see that the highest is cardiovascular diseases
- Visualization of the data as a whole
  - Plotting histplot and pair plot of entire dataset
    - We see that the histogram shows that the death causes all have similar trend and with this univariate analysis we are not able to understand anything which will help us study the cause of death so we will move on to further analysis
- Checking the unique values in year country to know what there are
- Assigning a variable to the causes of death
- Creating a new column for total deaths for that row with total from the causes separately
- Top 10 Total\_no\_of\_Deaths of all countries
- Top 60 Total\_no\_of\_Deaths of all countries
- Visualizing Year and Causes of Disease
- Splitting the year and causes in 2 variable and plotting scatter , line and bar plot
  - We see that the bar plot is thw best in representing the trend , we see for the most part the level are uniform for all years , except for:
    - Diabetes Mellitus,Neoplasm,Chronic Kidney Disease,Alzhelmers and Parkinson - increased
    - Protein Malnutrition ,Tuberculosis ,maternal disorders,nutritional defeciancies & Drowning - decreased
    - Conflict and terrorism ,Exposure to forces of nature are irreglar as they are time or based on events which occurred during that year
- Studying the relationship between the causes and China as its having the most deaths compared to any other Country/Territory
- China - "Total\_no\_of\_Deaths" against "Year"
- Year vs all causes of death in China –we see varied but mainly upward trend
- Second Highest country in total death per year - India against "Year"
- Year vs all causes of death in India –we see varied but mainly upward trend
- Top 3 Countries interms of "Total no.of Deaths" - For All the Years
- China India and USA are top 3
- Top 10 Causes of Deaths in China –both count as well as graph
- Top 10 Causes of Deaths in India – count as well as graph
- Top 10 Causes of Deaths in USA –both count as well as graph
- Disease specific country fatality for span of 30 years
- Again we see India China and USA , in some we see other countries but mainly the 3 countries
- Country specific disease fatality for span of 30 years
- We see cardiovascular diseases at the top, and then neoplasms
- Which are the countries facing overall highest deaths due to diseases-plotted bar graph
- Which are the top killer diseases in the world –plotted pie chart
- Found that Cardiovascular diseases is the highest followed by Neoplasms and Chronic respiratory Diseases.
- Plotting heat map to see the relationship between the features
- Final Observations – China and India have historically had the most deaths and Cardiovascular is the cause of the highest death overall from 1990 to 2019