



Cause of Death

Submitted by:

Suraj Kumar

ACKNOWLEDGMENT

References which are taken for some coding and steps help for completing the project:

<https://www.kaggle.com/code/spscientist/a-simple-tutorial-on-exploratory-data-analysis>

<https://www.freecodecamp.org/news/sort-dictionary-by-value-in-python/>

<https://sparkbyexamples.com/pandas/pandas-sum-rows/#:~:text=To%20sum%20Pandas%20DataFrame%20rows,of%20columns%20use%20axis%3D1%20.>

<https://seaborn.pydata.org/generated/seaborn.barplot.html>

INTRODUCTION

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. 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). 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.

In this Dataset, we have Historical Data of different cause of deaths for all ages around the World.

We need to give findings and conclusions in detailed data analysis.

Analytical Problem Framing

- Data Sources and their formats

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.

The data is in CSV format and it contains most of the data in numerical format which contain number of deaths due to different causes two or three columns contains object data which contain country name and code.

- Hardware and Software Requirements and Tools Used

We have used following tools for analysis of our data:

Pandas: For loading, handling, analysis and visualisation of Dataframe.

NumPy: For array works.

Matplotlib: For visualisation and analysis of data.

Seaborn: For visualisation and analysis of data.

CONCLUSION

- Key Findings and Conclusions of the Study

After analysis and visualisation, we have found following things from the data:

- There is certain disease which are highly correlated to each other it means that if a person has one disease, then it is quite possible that he can suffer from other disease also.
- There are certain countries in which cause of death are high due to same type of disease.
- Most number of deaths occurs due to certain 10 to 15 types of disease these are more than half of total death.
- Every year showing similar behaviour in case of death due to certain disease i.e., every year the greatest number of people dies due to certain same type of disease.
- There is certain disease due to which there is no any death occurs in some countries.
- Some countries have highest death numbers.