Project Report

on

"Cause of Death"

Submitted by

Sumit Pandey

Submission Date: 18th December 2022

Data Description

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 of 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 allows 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.

Objectives

The main objective of EDA on this project is – "To know FATAL DISEASE BURDEN OF THE WORLD"

Datasets

In this Dataset, we have Historical Data on different causes of death 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.

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings
0	Afghanistan	AFG	1990	2159	1116	371	2087	93	1370	1538	...	2108	3709	338
1	Afghanistan	AFG	1991	2218	1136	374	2153	189	1391	2001	...	2120	3724	351
2	Afghanistan	AFG	1992	2475	1162	378	2441	239	1514	2299	...	2153	3776	386
3	Afghanistan	AFG	1993	2812	1187	384	2837	108	1687	2589	...	2195	3862	425
4	Afghanistan	AFG	1994	3027	1211	391	3081	211	1809	2849	...	2231	3932	451
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6115	Zimbabwe	ZWE	2015	1439	754	215	3019	2518	770	1302	...	3176	2108	381
6116	Zimbabwe	ZWE	2016	1457	767	219	3056	2050	801	1342	...	3259	2160	393
6117	Zimbabwe	ZWE	2017	1460	781	223	2990	2116	818	1363	...	3313	2196	398
6118	Zimbabwe	ZWE	2018	1450	795	227	2918	2088	825	1396	...	3381	2240	400
6119	Zimbabwe	ZWE	2019	1450	812	232	2884	2068	827	1434	...	3460	2292	405

6120 rows × 34 columns

It provides information that includes over 6120 records and 34 attributes. The data set is in CSV (Comma Separated Value) format which is further prepared to data frame as supported by the pandas' library in python.

EDA methods

1. Univariate visualization—provides summary statistics for each field in the raw data set
2. Bivariate visualization—is performed to find the relationship between each variable in the dataset and the target variable of interest
3. Multivariate visualization—is performed to understand interactions between different fields in the dataset

Conclusion

	Disease	total_death
0	Cardiovascular Diseases	447741982
1	Neoplasms	229758538
2	Chronic Respiratory Diseases	104605334
3	Lower Respiratory Infections	83770038
4	Neonatal Disorders	76860729
5	Diarrheal Diseases	66235508
6	Digestive Diseases	65638635
7	Tuberculosis	45850603
8	Cirrhosis and Other Chronic Liver Diseases	37479321
9	HIV/AIDS	36364419
10	Road Injuries	36296469

Top 10 fatal diseases burden of the world.