CAUSE OF DEATH PROJECT

Prepared by Suneha Sengar

Data Science Intern at Flip Robo Technologies

Internship Batch 34

— Flip Robo Technologies



SME: Khushboo Garg

Acknowledgement

It is my most profound pleasure and gratification to present this report. Working on this project was an incredible experience that gave me very informative knowledge of data analysis.

Flip Robo Technologies (Bangalore) provided all the required information and dataset, which helped me complete the project.

I want to thank my SME Khushboo Garg for giving me the dataset and instructions to perform the complete case study process.

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. 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.

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.

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

**Data Scientists have to apply their analytical skills to give findings and conclusions in detailed data analysis written in jupyter notebook . Only data analysis is required.**

Benefits of Cause of Death project

1. Helps in analysing trends of disease with passing years.

2. Helps in policy formulation related to health.

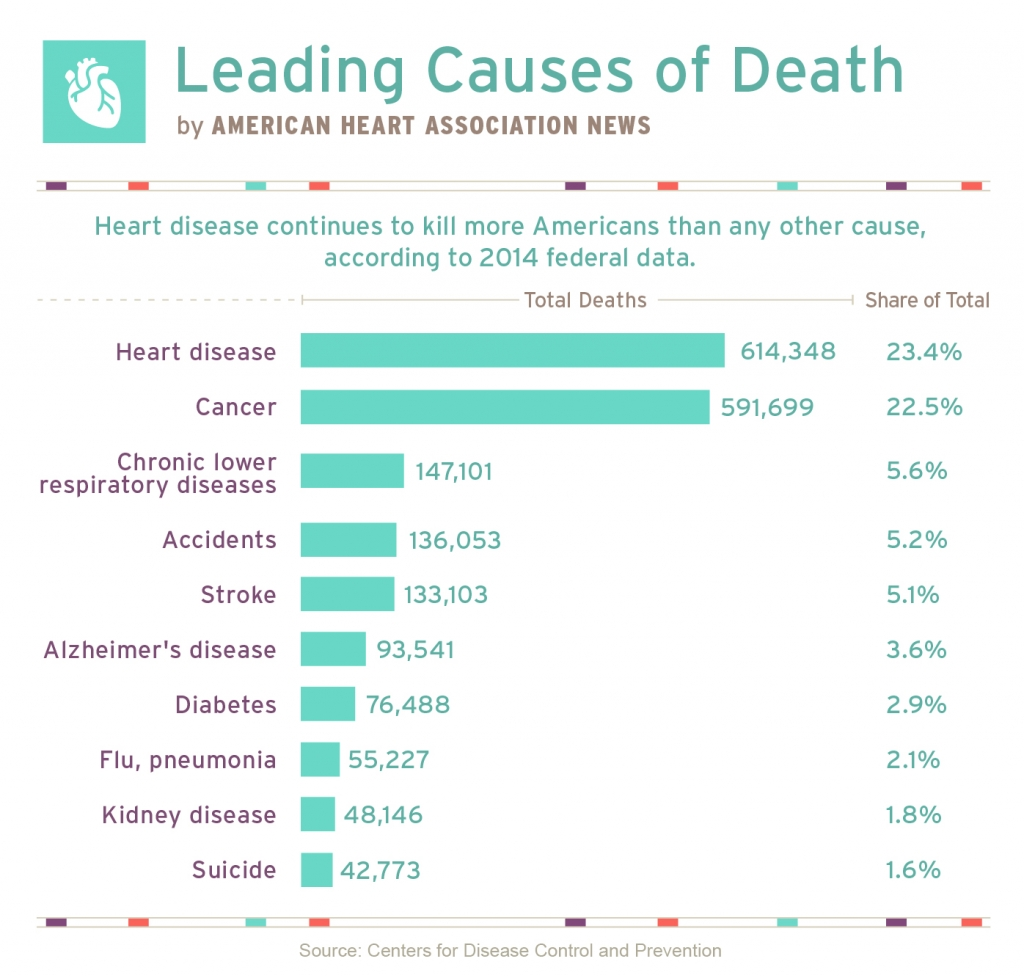
3. Enable evidence based decision making.

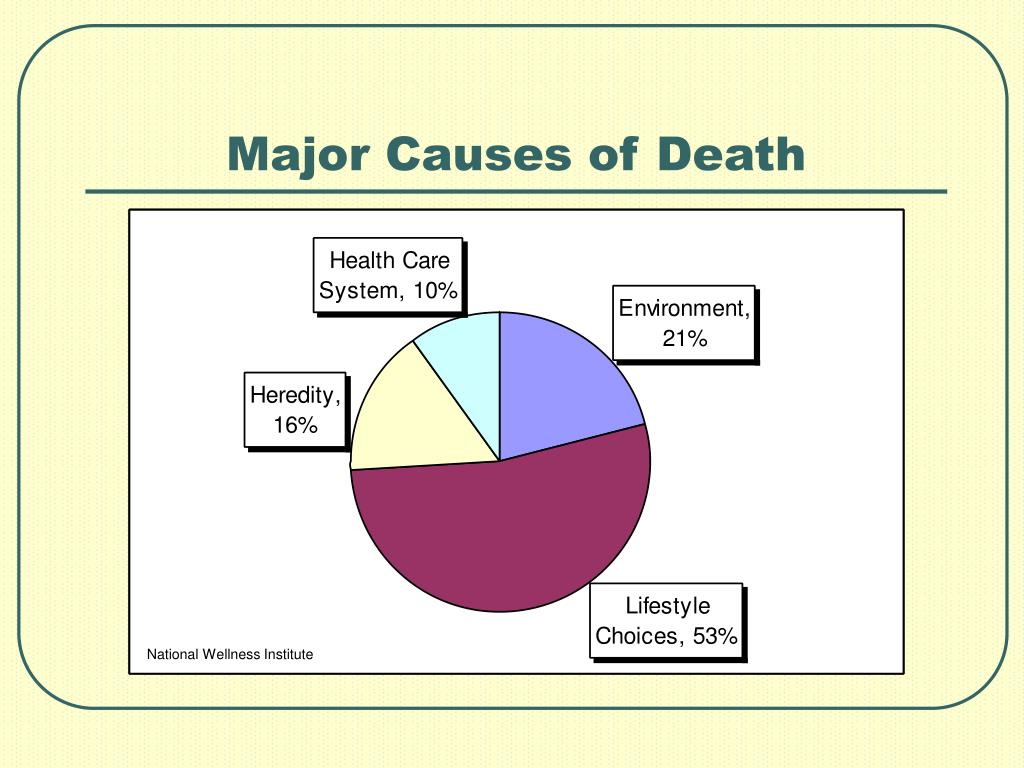
4. Helps in plugging loopholes in heath infrastructure

5.Enable equitable resource distribution

6.Empower citizen regarding their health status.

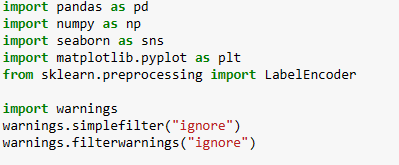
7.To minimise DALY and mortality.





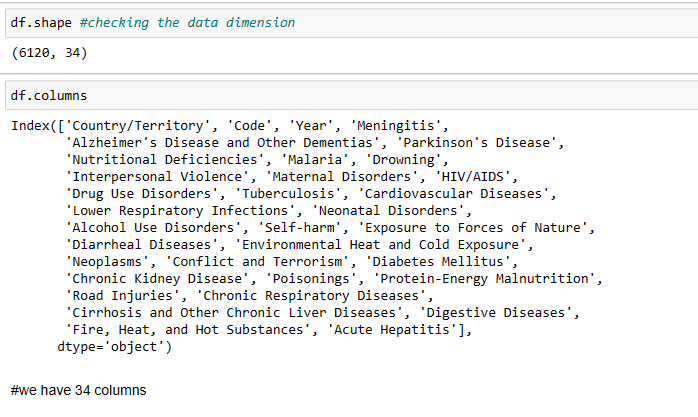
Dataset Details:

First, I imported all the necessary libraries and dependencies to create a detailed data analysis in Python.





**Exploratory Data Analysis (EDA):**

****



There was no missing values.



It described each and every columns , there mean, median , 25, 75 percentile etc.





columns viz conflict and terrorism ,exposure to forces of nature and malaria have many rows with 0 value. It means they are least important columns. they can be dropped.

**Visualization:**

**What is Data Visualization?**

Data visualization is defined as a graphical representation that contains the information and the data.

**Benefits of Good Data Visualization?**

Data visualization is another technique of visual art that grabs our interest and keeps our main focus on the message captured with the help of eyes.

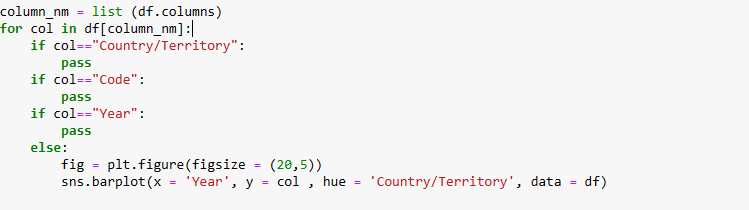
Different Types of Analysis for Data Visualization are:

1. Univariate Analysis: In the univariate analysis, we will be using a single feature to analyze almost all of its properties.

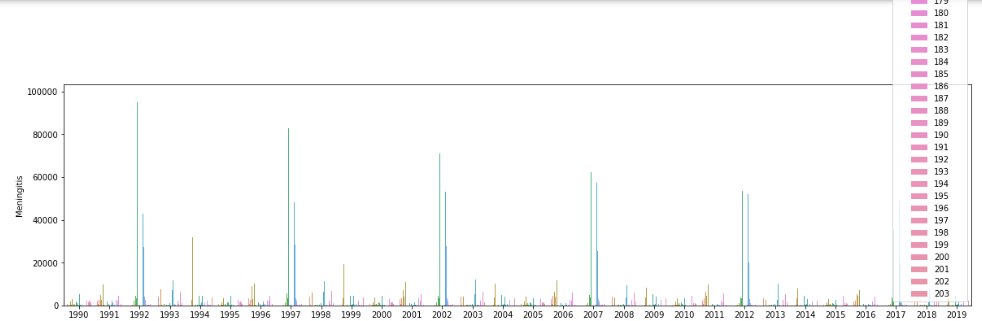
2. Bivariate Analysis: When we compare the data between exactly 2 features then it is known as bivariate analysis.

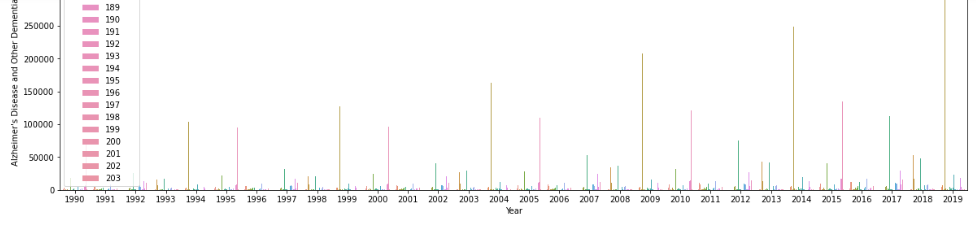
3. Multivariate Analysis: In the multivariate analysis, we will be comparing more than 2 variables.

**Bivariate Analysis**



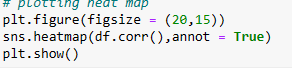
**Output**

****

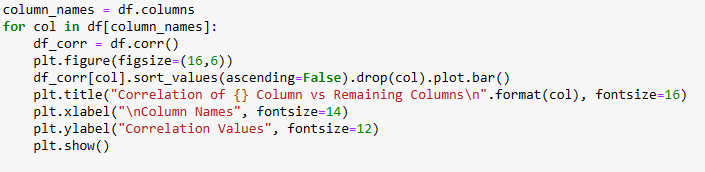
****

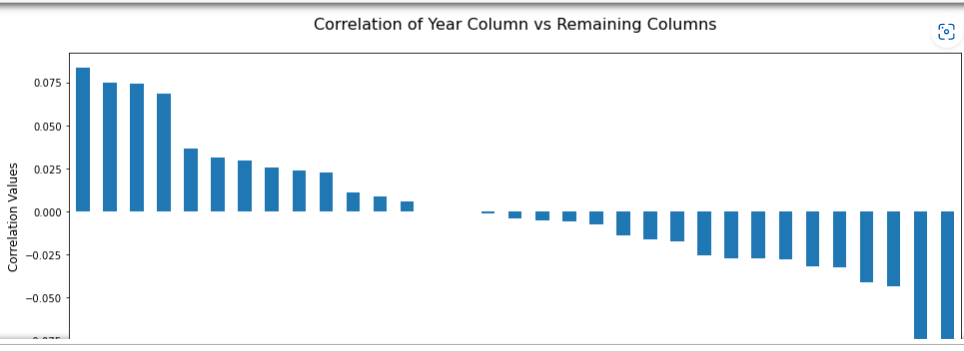
On similar line, graphs are drawn for each “cause of death” column

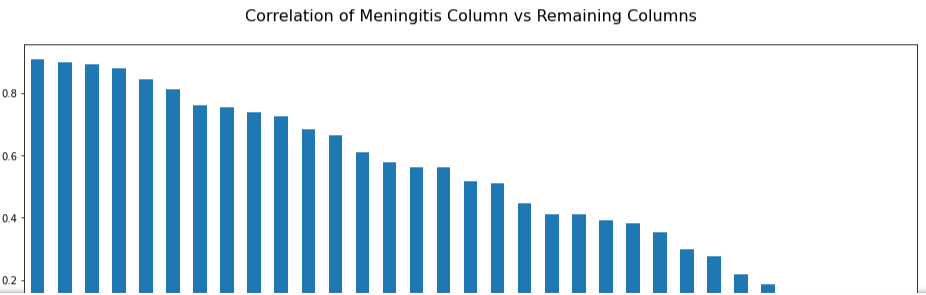
**Multivariate Analysis**

****

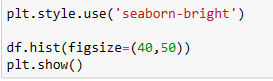
****

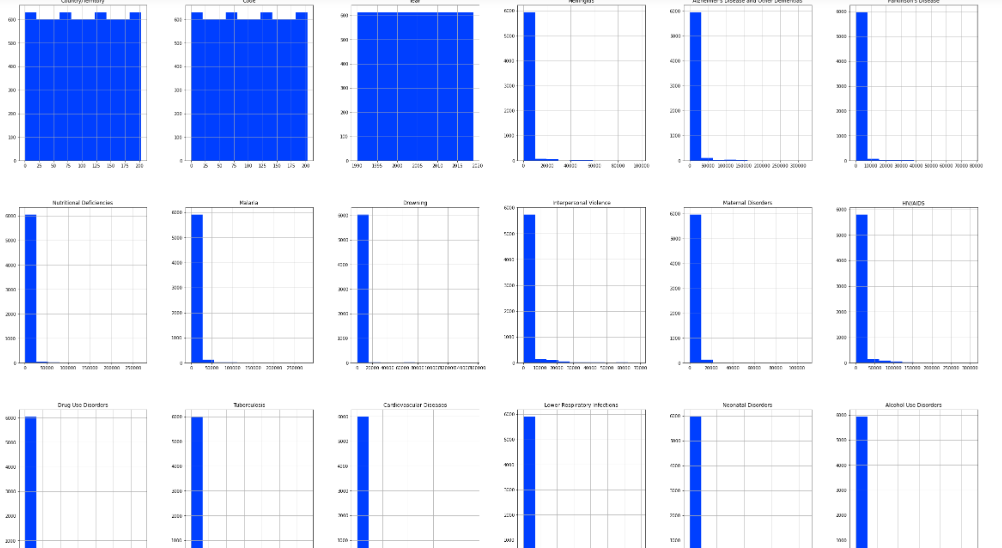
****

****

****

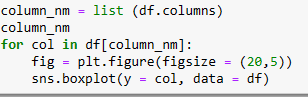
**On similar lines other correlations are made.**

****

****pairplot is also one way of bivariate analysis



It help us to find skewness in dataset. In this dataset almost every column is having skewness. Which need to be cured before giving data to model.



#outlier is present in almost every column except country, code and year column.

# 

# **Inference**

**following inference can be drawn from EDA on Cause of Death Project:**

1.Every other disease is related to other eg protein and eneregy malnutrition is related to alzheimer and dementia disease or chronic liver disease. etc

2.Almost all disease are preventable with good social, cultural and medical determinant of heath. eg equitable distribution of helath resource, safe water and sanitation, good environment etc

3.There are 203 different countries covered in survey. Poor countries suffer from “disease of underdevelopment” whereas Rich countries suffer from “poor lifestyle induced disease”

4.Need for early maternal and child care to minimise the DALY.

5.There are some social determinant of health viz interpersonal relationship, war etc that need to be taken care of.

6.There are some environmental determinant of health viz clean air, water, etc that need to be taken care of to handle disease viz diarrhea, malaria etc.

**FUTURE WORK:**

COUNTRY SPECIFIC FOCUS IS REQUIRED .

EG

U.S, BRUNIE, ANDORRA REQUIRE INTERVENTION IN DISEASES VIZ CARDIOVASCULAR DISORDER ETC

AFGHANISTAN , GUINEA ETC REQUIRE INTERVENTION FOR MALARIA, DIARRHEA, RESPIRATORY DISORDER ETC