Cause-Of-Death

Submitted by:
Sampada Sambrani
Internship batch 34
e-mail:sampada.sambrani@gmail.com

ACKNOWLEDGMENT

I would like to express my deep and sincere gratitude to FLIP ROBO for giving me this opportunity to do this project. As a great bridge between academic and industry, this program educated me how to perform theoretical methodology in real life. I would like to express my sincere thankfulness to our assigned mentors for the continuous support of our queries, for their patience, enthusiasm, motivation and immense knowledge.

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. 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.

The main steps in this project are:

- Exploratory Data Analysis (EDA). By conducting explanatory data analysis, we obtain a better understanding of our data. This yields insights that can be helpful later when building a model, as well as insights that are independently Interesting.

After the EDA for individual countries (year wise) , we are required to get the cause of maximum and minimum deaths caused for each countries and also get the year wise records accordingly.

Analytical Problem Framing

- Mathematical/Analytical Modelling of the Problem

Dataset

	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	Ma
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

- ✓ Our dataset has a total of 6120 rows and 34 columns as seen above.
- ✓ There are a total of 204 countries from first being Afghanistan to last being Zimbabwe.
- ✓ Each country has 30 years from (1990 to 2019) Historical Data of different cause of deaths for all ages around the World.

```
1  df.columns
```

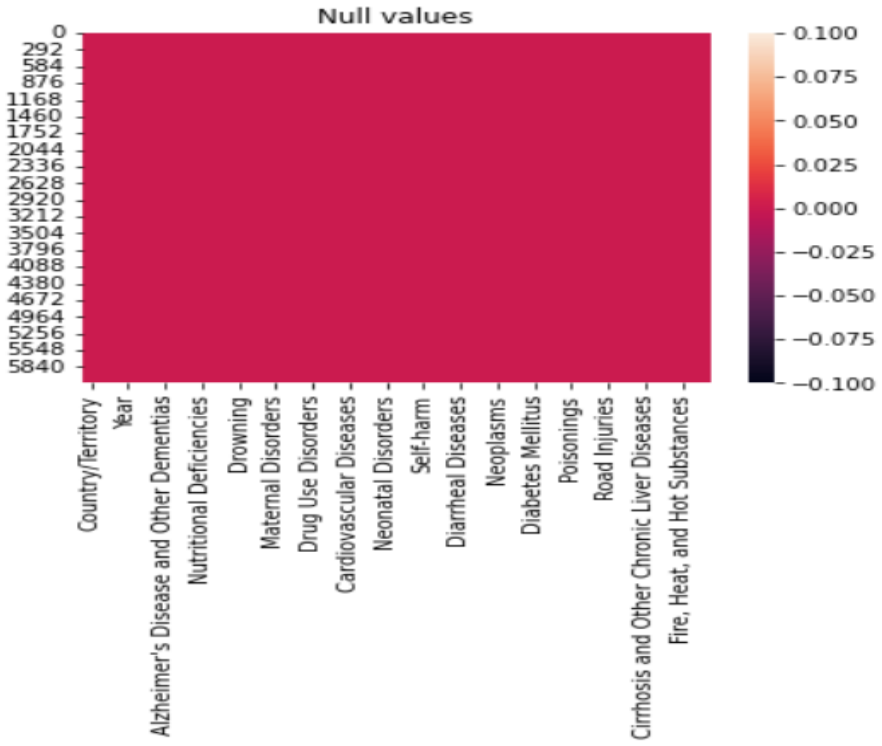
```
Index(['Country/Territory', 'Code', 'Year', 'Meningitis',
       'Alzheimer's Disease and Other Dementias', 'Parkinson's Disease',
       'Nutritional Deficiencies', 'Malaria', 'Drowning',
       'Interpersonal Violence', 'Maternal Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis', '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'],
      dtype='object')
```

```
df.isnull().sum()
```

```
Country/Territory                          0
Code                                       0
Year                                       0
Meningitis                                 0
Alzheimer's Disease and Other Dementias    0
Parkinson's Disease                        0
Nutritional Deficiencies                   0
Malaria                                    0
Drowning                                   0
Interpersonal Violence                     0
Maternal Disorders                         0
HIV/AIDS                                   0
Drug Use Disorders                         0
Tuberculosis                               0
Cardiovascular Diseases                    0
Lower Respiratory Infections               0
Neonatal Disorders                         0
Alcohol Use Disorders                      0
Self-harm                                  0
Exposure to Forces of Nature               0
Diarrheal Diseases                         0
Environmental Heat and Cold Exposure       0
Neoplasms                                  0
Conflict and Terrorism                     0
Diabetes Mellitus                          0
Chronic Kidney Disease                     0
Poisonings                                 0
Protein-Energy Malnutrition                0
Road Injuries                              0
Chronic Respiratory Diseases               0
Cirrhosis and Other Chronic Liver Diseases 0
Digestive Diseases                         0
Fire, Heat, and Hot Substances             0
Acute Hepatitis                            0
dtype: int64
```

- ✓ Our dataset has no null values present nor any duplicate values present in it.
- ✓ The heatmap shows there are no null values in our dataset

```
sns.heatmap(df.isnull())
plt.title("Null values")
plt.show()
```



- Data Sources and their formats

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.

	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	Ma
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

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 Pre-processing Done
  - ✓ As there are no null values present in our dataset as well as no duplicate values present, hence no data pre-processing was done.
  - ✓ Also, there is no any target variable present, the entire data analysis was carried for the dependent variables itself.

- Data Inputs- Logic- Output Relationships
  - ✓ As the data set was large with 6120 rows and 34 columns, where 204 countries were present , where each country had the records for 30 years(1990 to 2019)
  - ✓ The output of each analysis is the maximum and minimum causes of death in each year , where the output depends on various factors for each country.

- Hardware and Software Requirements and Tools Used

Hardware specification

Processor: intel CORE i3 (10th gen) minimum
RAM: 2GB and above
Hard disc capacity: Minimum of 100GB
Display type: Standard VGA

Software specification

Operating system: Windows 10
Front end : Jupyter framework(anaconda)
Programming tool: Anaconda
Internet browser: Google chrome

Libraries

import numpy as np – for numeric algebra
import pandas as pd- for data representation
import matplotlib.pyplot as plt- for data visualization
import seaborn as sns

- Pandas is a popular Python library for data analysis.
- Matplotlib is a very popular Python library for data visualization
- NumPy is a very popular python library for large multi-dimensional

CONCLUSION

- Key Findings and Conclusions of the Study

Our Dataset contained mostly numerical as well as categorical variables, where all the dependent features were used to equally to find the cause of death

.

- Limitations of this work and Scope for Future Work

A machine-learning model for predicting cause of death can be developed using a standardized common data model and an extensible analytical method.

We attempted to use a transparent development process; consequently, the prediction performance of the model was impressive.

The majority of observational data sets lack cause-of-death data and our model can be used to impute this information with high discriminative performance.