Zarqa Ayub

Achievement 6 Project

Data Source:

The data set retrieved from Kaggle. The web page provides the main source citation. This dataset contains the causes of deaths of people worldwide for all ages. This dataset is collected from the website of Our world in Data.
https://ourworldindata.org/causes-of-death.

The data sourced from an authentic external open platform which is reliable due to its collection and research-oriented purpose which aimed at gaining insights into global causes of mortality.

License:

Why this dataset:

This serves as an illustrative case study demonstrating the impact and significance of data analysis. The dataset comprises numerous continuous variables that may lack meaningful interpretation individually. However, with sufficient curiosity, appropriate tools, and targeted inquiries, a wealth of valuable information can be extracted. In this specific dataset, insights into the health status of a population and the primary causes of mortality can be derived. Such information proves valuable for doctors, researchers, government entities, and the pharmaceutical industry, guiding them in determining where to direct medical production efforts and exploring innovative approaches to healthcare.

Data Collection Method:

The data is collected manually for administrative purpose. Every region was responsible to keep records of its demographics.

Data Contents:

The data set is formed of **6120 rows** and **33 columns** and contains the death causes count by year and country. It contains the causes of deaths of people worldwide for all ages.

Ethical Considerations:

The dataset provides quantitative details about deaths that occurred because of various reasons. It does not include any personal information that could be traced back to individuals or their families.

Nevertheless, it contains sensitive content presenting an ethical concern related to the potential misrepresentation or generalization of certain diseases or lifestyle factors of a

whole country. For example, making inaccurate assumptions about a nation based on the high rates of deaths from alcohol consumption, HIV, or suicide.

Data Limitations:

Potential bias during the collection stage may arise when some deaths have multiple contributing factors. For instance, a person might have died of old age but been recorded as having succumbed to a specific disease. This introduces a source of bias in the data, as the primary cause of death may not accurately reflect the complex circumstances surrounding the individual's death.

Potential human errors during manual data entry; including inputting 20 instead of 2, placing information in the incorrect column or row, or making other mistakes in the process of entering data into the system.

The summary of deaths for a given year and country may vary from the total mortality that occurred because the data does not encompass all potential fatalities.

Data Profile:

- **Variables and Data Types:**

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

```
Acute Hepatitis                                    int64
```

- **Data Integrity**

There are no data integrity issues have been found such as missing values, duplicate values, and NaN values etc.

- **Dimensions**

The dataset consists of **6120** variables and **33** records. ‘Country/Territory’, ‘Code’ are **Qualitative** variables. '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.These are **Quantitative** variables.

Questions (Objectives):

1. What are the most common causes of death? What are the deadliest diseases?
2. Do the causes of death experience changes over the years?
3. Is there a correlation between a country's level of development and the causes of death?
4. What are the chances of dying from an external sudden cause versus an internal one?
5. Does causes of death vary among different countries?