

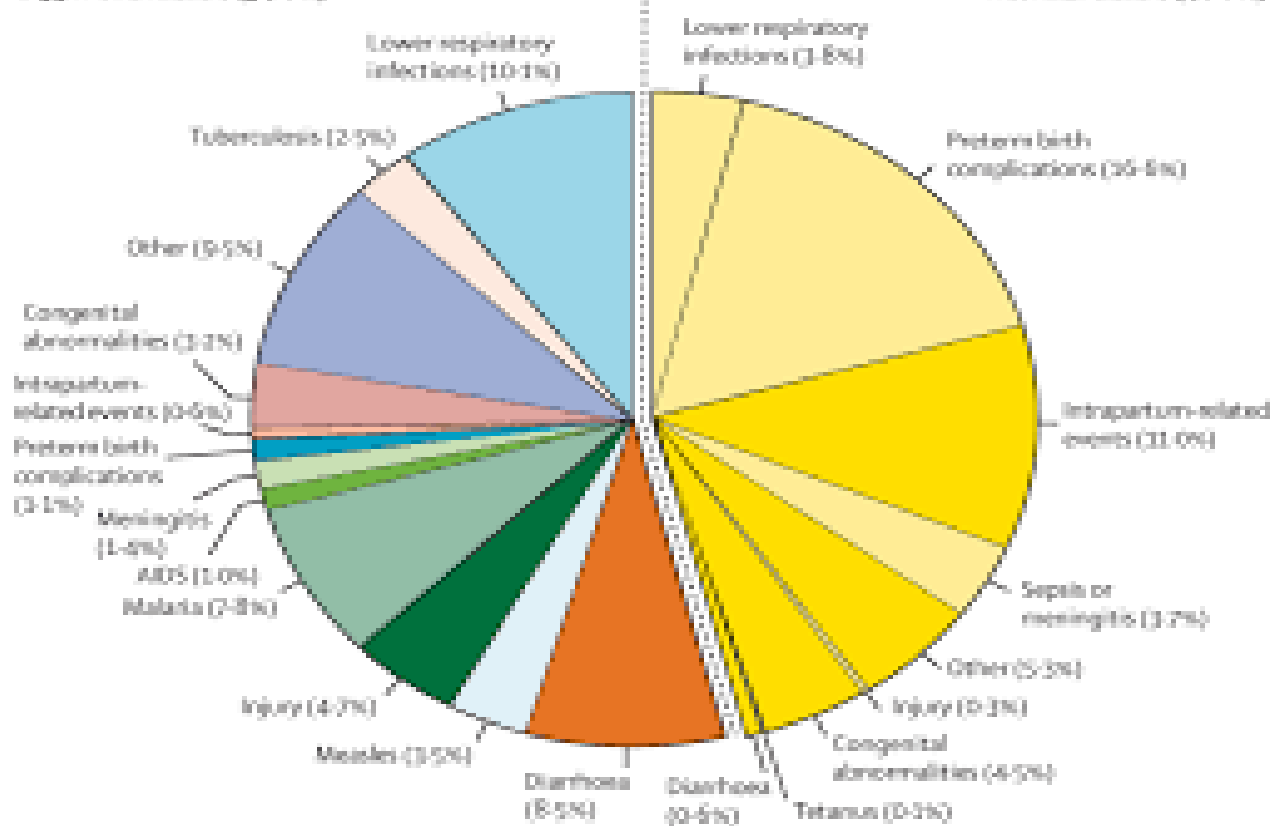
Causes of death dataset

Problem statement

In this dataset we will learn about causes of death in particular country

1-59 month deaths (54.0%)

Neonatal deaths (46.0%)



understanding

- This dataset tells about number of people died in a particular year of a particular country.
- There are number of causes that are responsible for the causes of death .
- Such as 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 etc.

understanding

- It is the way to find the health status of a population .
- The burden of disease is measured by a metric called Disability Adjusted Life Years(DALys).
- The first 'Global Burden of Disease' (GBD) was GBD 1990 .
- The GBD estimates how much people die worldwide from over 300 diseases .

EDA

- EDA is exploratory data analysis
- Which includes the following:-
- Checking the shape (.shape) gives the number of rows and columns.
- Checking the description(.describe()) gives the count,mean,standard deviation,max and min value and quartiles.
- See the datatypes of columns(.dtypes)and if they are object change them into integer and float.

EDA

- Checking for nulls (using `.isnull().sum()`) and treat them using imputers.
- Checking for zeros and treat for it using log, mean and mode.
- Checking for whitespaces and replace them with nulls and then treat them.

Data visualization

- This includes visualizing the data as we can better understand through charts and graphs.
- For that importing necessary libraries.
- Checking for outliers using boxplot.
- There are many techniques to deal with outliers such as z score, inter quartile range.
- After dealing with the outliers lets split the data into features and target.

Model building

- Train, test and split the data.
- After preprocessing model initializing and model building .
- Check for accuracy.
- Cross validation and hyper parameter tuning is necessary.