Cause of death Report

This dataset is the way to assess the health status of a population that focuses on the mortality rate. The mortality rate along with morbidity rate provides a stronger view on health outcomes.

In this dataset we have historical data of different cause of death for all ages around the world. The key features 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.

First, we import all the necessary libraries, and the path of the dataset is loaded into the jupyter notebook. There are 6120 rows and 34 columns. The features county/territory, code and year are of no significant for the analysis, hence we drop those columns. Now we have only 31 columns. We analyse the data with these features.

Using info() method, we see that, there are no null values, hence we need to not to handle missing values. Also, we observe that all the features are numerical type. Since there is no categorical data, we need not to encode.

Applying describe() method, we see the descriptive statistics . here we observe that, in some features the standard deviations are greater than the mean value. Hence the distribution is not normal. Also, we see the different quadrant values of the features.

In the visualization using boxplot we observe all the features have some outliers. These outliers would affect the result of the analysis. Hence, we remove all the outliers. After removing outliers, we check for skewness, all are in acceptable range.

Next, check for the correlation among the features using heatmap. We observe that, Meningitis-Nutritional Deficiency, meningitis- maternal disorders, Parkinson’s disease- Alzheimer’s disease, Alzheimer’s disease-cardiovascular disease, Parkinson’s disease- cardiovascular disease, neonatal disease- maternal disorders, neoplasms- cardiovascular disease, cirrhosis and other chronic liver disease- acute hepatitis, cirrhosis and other chronic liver diseases- alcohol uses disorders,- diabetics mellitus , diabetes mellitus- digestive disorders- chronic respiratory disease, neoplasms- chronic respiratory diseases, - cirrhosis and other chronic liver disease-digestive disease re highly correlated.

So, we conclude that, diabetes mellitus, cirrhosis and other chronic liver diseases, chronic respiratory diseases, deficiency in protein and energy, malnutrition, digestive disorders, maternal disorders features are more prone to the cause of death. Also, the natural disasters and alcohol use disorders contribute to the mortality rate. Hence, the precautions and treatment of these diseases and disorders should be emphasized to lead a lower mortality rate and for a healthy life.