

CAUSE OF DEATH

Submitted by:

RAVI PRAKASH BAJPAI

**ACKNOWLEDGMENT**

In this project I have refered to the detail provided by flip robo and I have also refered to google for many terms and about data, I have also refered to wikipedia and other websites

**INTRODUCTION**

* Business Problem Framing

This data set is related to deaths cause in different parts of world and the disease through which death happened and other reason thorugh which death was caused. This data set also describe about death of people of different age group and the reason of the death

* Conceptual Background of the Domain Problem

In this project there are two concepts that are morality rate and morbidity rate which should be measured. ‘Burden of disease’ is also major feature that is sum of morality and morbidity that should be considered

* Review of Literature

In this project data set has been provided through which we can analyse total numbers of death cause by different disease or other reasons that has cause death.

First four most causing death diseases are shown

Cardiovascular Diseases 447741982

Neoplasms 229758538

Chronic Respiratory Diseases 104605334

Lower Respiratory Infections 83770038

We can analyse that death cause by cardiovasular Diseases is highest and death by Exposure to force of nature is least as shown in jupyter notebook

We can also check which area of the world is affected more by any particular disease so more medical treatment can be provided to that particular part

We can also check the data according to the year that is in which year particular disease has affected more or less in a given area

**Analytical Problem Framing**

Data source provided is a csv file provided by flip robo. In the given data set there are 6120 rows and 34 features as seen through dot shape function.

Nan values are nil in all the features that is no data is missing.

Country and country code both are giving the same information so droping country name and using only code value.

Code is given in string format as seen in data so applying label encoder to convert string data into numeric format.

Summation of number of death has also been done to find which disease is causing more no of death.

Through describe function it can be seen that data is not in normalised form there are outliers in data and skewness is present in the data.taking the example of disease ‘Malaria’ in this feature we can see mean value is 4140 but its standard deviation is showing nearly 18000 which is not possible, In min value, first quantile, second quantile it is showing 0 value but in third quantile it is 393 and max is showing 280604. Through this we can analyse that data is not properly distributed and there is skewness present in the data.

Similarly we can check for other features also which shows that data is not properly distributed.

Data Preprocessing Done

For more analyses subplot has been drawn, here in subplot it can be clearly seen that data is not normalized and there is no proper distribution of data,this can be seen nearly in all the features are have been ploted in subplot. ignoring the years and country code column.

For clarifying more about data .skew function has been used,here from skewness it can be seen all the values are nearly greater than 5 (considering 5 as normal value),this shows that skewness is present in the data.

Zscore method

Now applying Zscore method to remove the skewness from the data.checking the data again through ploting subplot and skew function.there is clear difference showing in the subplot and skew values that skewness is remove to large extend.

HEAT MAP

Ploting the heat map for the given data set,through heat map it can be visualize that there is multicolinearity present in the data. We can see that

1.nutritional deficiencies and protein energy malnutrition are closely related to eath other as value is showing 1

2. neoplasms is closely related to alzheimer disease and parkinsons disease.

Simarily we can see many features are corelated to other features in the given data set .

Multicolinearity can be removed from VIF function.

CONCLUSION

In this project we have drawn up with many conclusions

* Death rate in the particular country.
* Death rate in the particular year.
* Disease that is mostly affecting to larger part of the world.
* Any particular country which is affected more from any particular disease.
* Disease which are growing year by year.

From above conclusion death rate can be reduced.focusing more on particular area affected by particular disease so more life can be saved by providing medical treatment to the people.