# PROJECT REPORT

on

Causes of death dataset

# DATA SCIENCE

BY

DATA TRAINED


SUBMITTED BY

POOJA RAJPAL

INTERNSHIP BATCH 33

SUBMITTED TO

MR.SHWETANK MISHRA



***FLIPROBO TECHNOLOGIES***

SESSION 2022-23

# About python

This project is done in a jupyter notebook using python language.

This is widely used high level programming language.

It is a programming language that lets you work quickly and integrate system more efficiently.

Python is designed to be highly readable.

It uses English keywords frequenty.

It is processed at a runtime by the interpreter.

# About

This dataset is about causes of death all over the world for all ages.The causes of death may be because of many diseases .The key features of this dataset are Meningitis, Alzheimer's Disease and Other Dementias, Parkinson's Disease, Nutritional Deficiencies, Malaria, Drowning and many more ..This dataset tells about number of deaths in a particular country in a particular year .

Way to access health status is to focus on mortality.The first GLOBAL BURDEN OF DISEASE (GBD) was GBD 1990.

Around 56 million people die every year .What caused their death?How did the causes of death change over time and differ between countries and regions.

Death rate related to disease ,and other health factors tend to change relatively slowly over time.It is important to understand what is meant by the cause of death and the risk factor associated with a premature death.

# Causes according to this dataset:

1. Meningitis
2. Alzheimer's Disease and Other Dementias
3. Parkinson's Disease
4. . Nutritional Deficiencies
5. Malaria
6. Drowning
7. Interpersonal violence
8. Maternal Disorders
9. Drug Use Disorders
10. Tuberculosis
11. . Cardiovascular Diseases
12. . Lower Respiratory Infections

13. . Neonatal Disorders
14. Alcohol use disorders
15. Self harm
16. Exposure to Forces of Nature
17. . Diarrheal Diseases
18. Environmental Heat and Cold Exposure
19. Neoplasms
20. Conflict and Terrorism
21. Diabetes Mellitus
22. Chronic Kidney Disease
23. Poisioning
24. . Protein-Energy Malnutrition
25. Chronic Respiratory Diseases
26. Cirrhosis and Other Chronic Liver Diseases
27. . Digestive Diseases
28. Fire, Heat, and Hot Substances
29. Acute Hepatitis

# Data analysis steps

In the jupyter notebook

We will first import necessary libraries

Load the dataset

Check the shape

Check for null values in the columns and treat them using imputers

Check for zeros and white spaces and treat them

Data visualization

Data skewness to see whether the data is skewed right or left

And treat the skewness using power transformer to make it normal or to remove the skewness( bell shaped curve )

Splitting the data into into feature and label

Train,test and split the data

Import model library

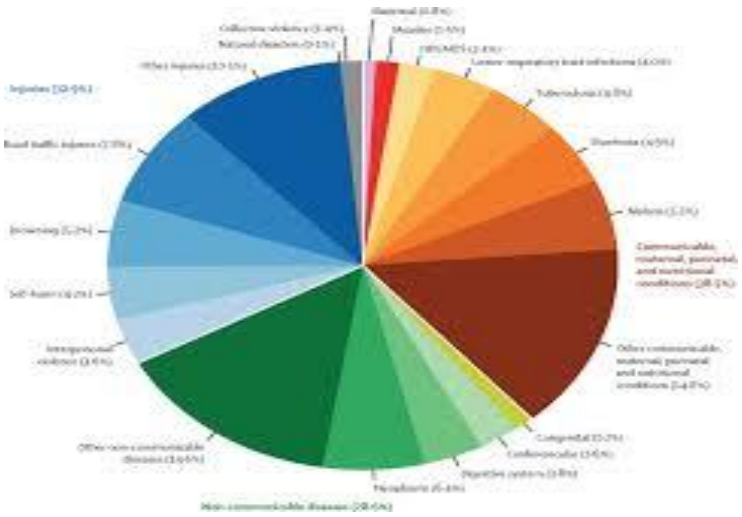
Initialise the model

Train the model

Check the accuracy score for training  and testing

Cross validation score to see if the model is overfitting or not

Hyper parameter tuning using GridSearch CV to increase the accuracy



# Thank  you

30.