# Cause Of Death EDA

## INTERNSHIP BATCH-33

***Submitted by:***
***Rahul M***

## Context

A straightforward way to assess the health status of a population is to focus on mortality – or concepts like child mortality or life expectancy, which are based on mortality estimates. A focus on mortality, however, does not take into account that the burden of diseases is not only that they kill people, but that they cause suffering to people who live with them. Assessing health outcomes by both mortality and morbidity (the prevalent diseases) provides a more encompassing view on health outcomes. This is the topic of this entry. The sum of mortality and morbidity is referred to as the 'burden of disease' and can be measured by a metric called 'Disability Adjusted Life Years' (DALYs). DALYs are measuring lost health and are a standardized metric that allow for direct comparisons of disease burdens of different diseases across countries, between different populations, and over time. Conceptually, one DALY is the equivalent of losing one year in good health because of either premature death or disease or disability. One DALY represents one lost year of healthy life. The first 'Global Burden of Disease' (GBD) was GBD 1990 and the DALY metric was prominently featured in the World Bank's 1993 World Development Report. Today it is published by both the researchers at the Institute of Health Metrics and Evaluation (IHME) and the 'Disease Burden Unit' at the World Health Organization (WHO), which was created in 1998. The IHME continues the work that was started in the early 1990s and publishes the Global Burden of Disease study.

# Exploratory Data Analysis

1.Checking the missing values

2.Checking for numerical columns

3.Checking for the distribution of numerical variables

4.Checking for Categorical variables

5.Types of categorical variables

6.detecting outliers

7.Checking Skewness

8.Outliers detection

9.correlation on columns in the dataset

**1. Checking the missing value**

Missing value can be checked by the following python code:-

```
missing_value=[feature for feature in df.columns if
df[feature].isnull().sum()>1] missing_value
```

Observation:-

There are no missing values present in the dataset

**2. Checking the numerical columns**

Numerical values can be checked using the following python code:-

```
numerical_feature=[feature for feature in df.columns if df[feature].dtypes!="O"]
df[numerical_feature]
```

Observation:-

There are 32 numerical values present in the dataset

**3. Checking the categorical variable**

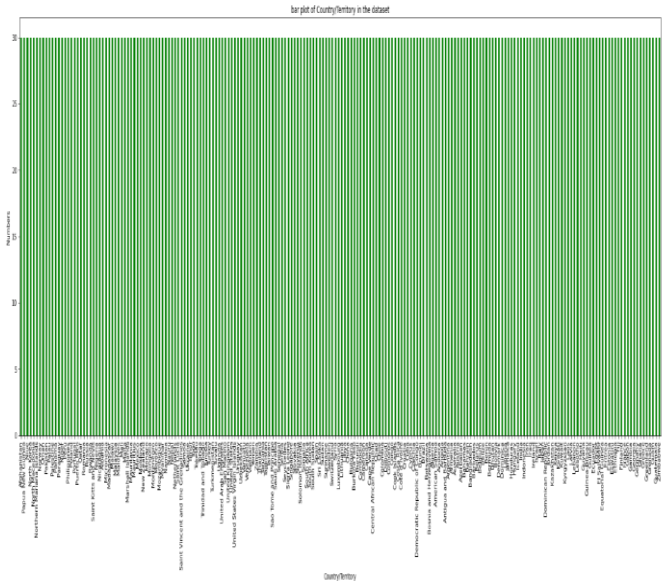
Categorical variable can be checked using the following python code

```
discrete_feature=[feature for feature in df.columns if feature not in numerical_feature]
df[discrete_feature]
```

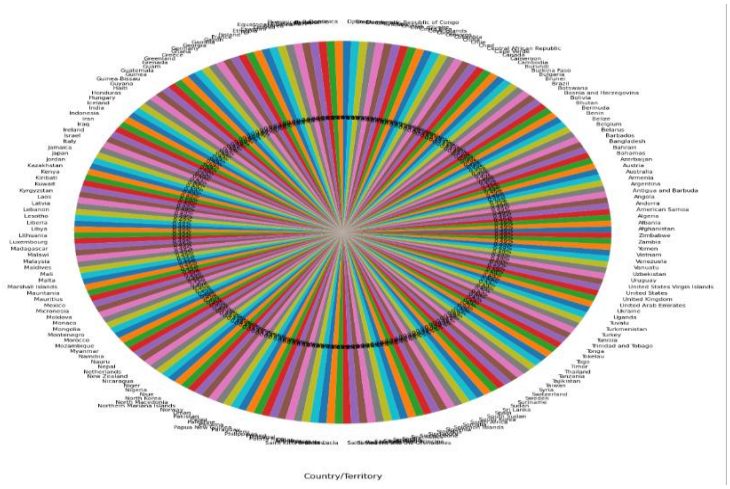
Observation:-

There are 2 categorical values present in dataset

# Data Visualization:-




Country/Territory




observation:-

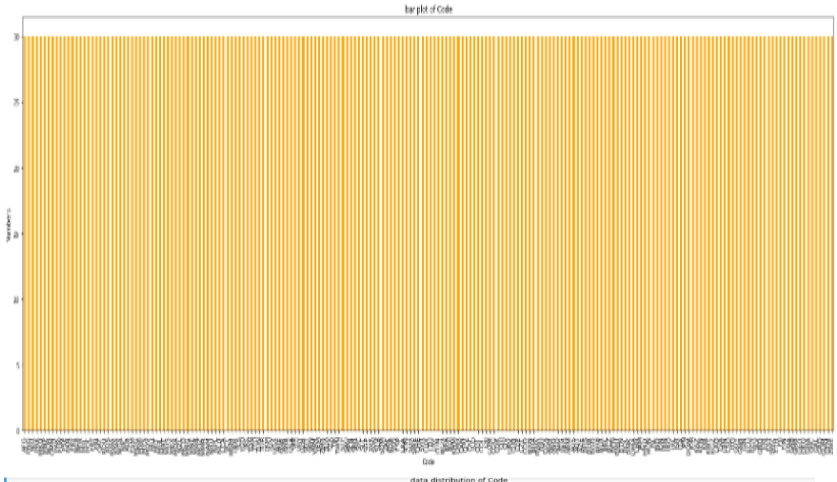
we can observe 204 unique variables in the dataset and

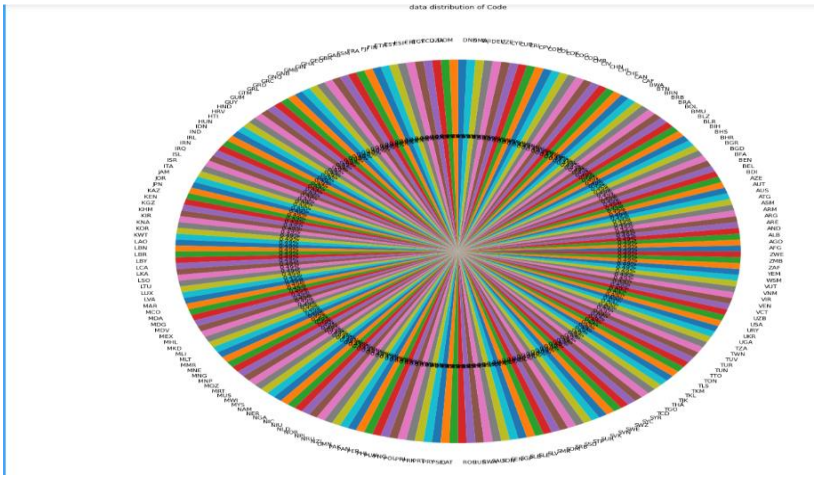
all the variables are having uniformly distributed and

each variables in the column are having count of 30 .

each of the countries in the dataset is 0.49% of the dataset

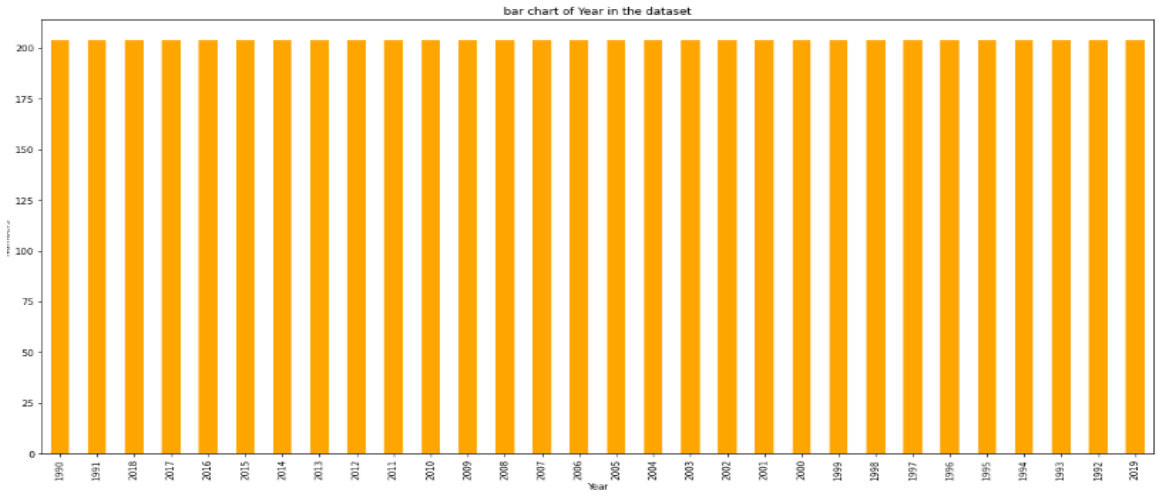
code:

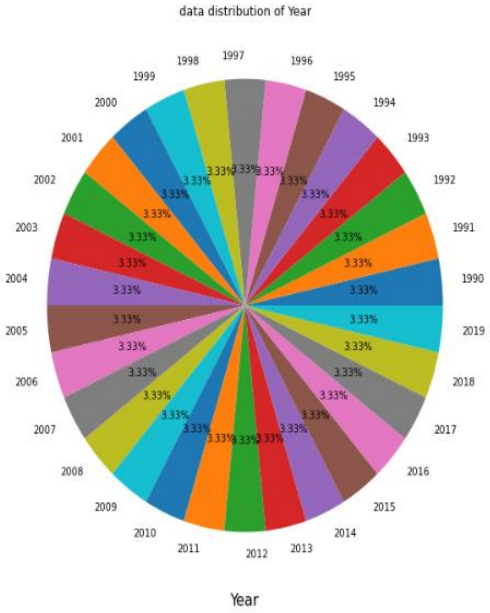


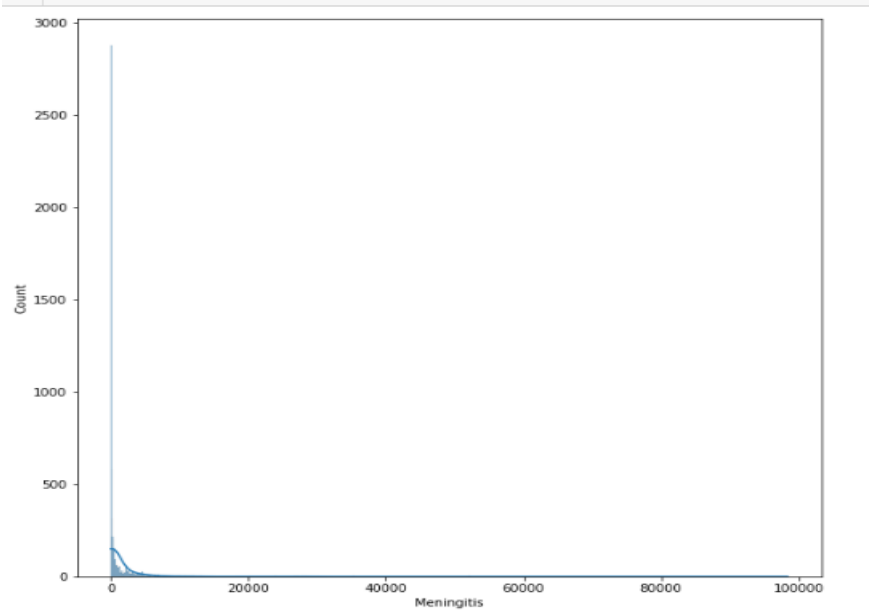

Observation:-

we can observe 204 unique variables in the dataset and
all the variables are having uniformly distributed and
each variables in the column are having count of 30 .
each of the code corresponding to countries in the dataset is 0.49% of the dataset



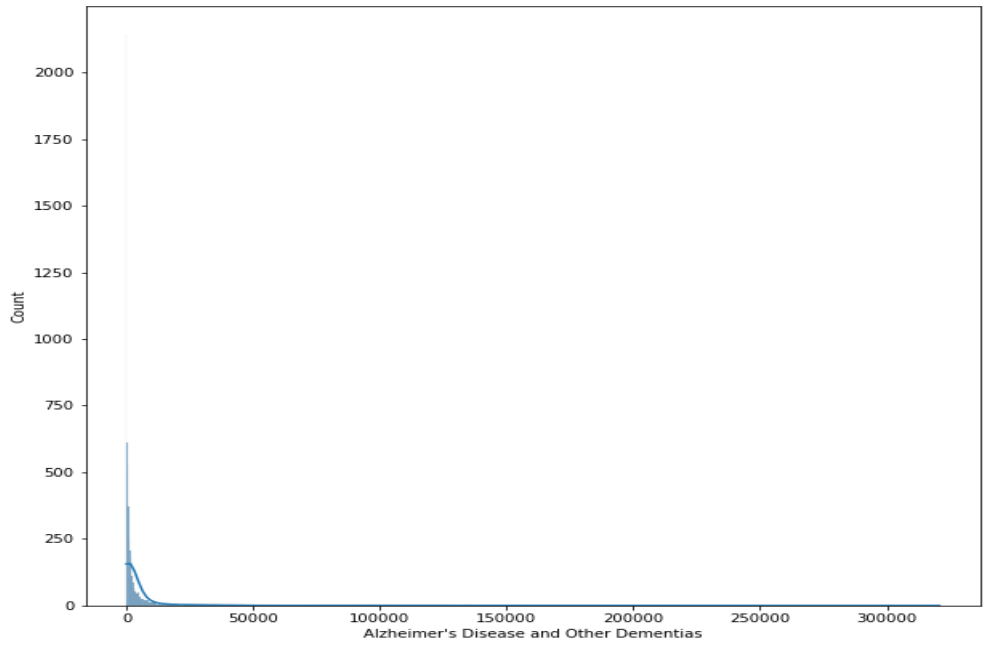




observation:-
We can see that there are 30 unique year in the dataset and
they re uniformly distributed in the data set and their distribution percentage is 3.33% in all the years through out the datsaet.
We can observe 200 observations in each of the year in the datset.
we can see the data is from 1990 to 2019 in the dataset



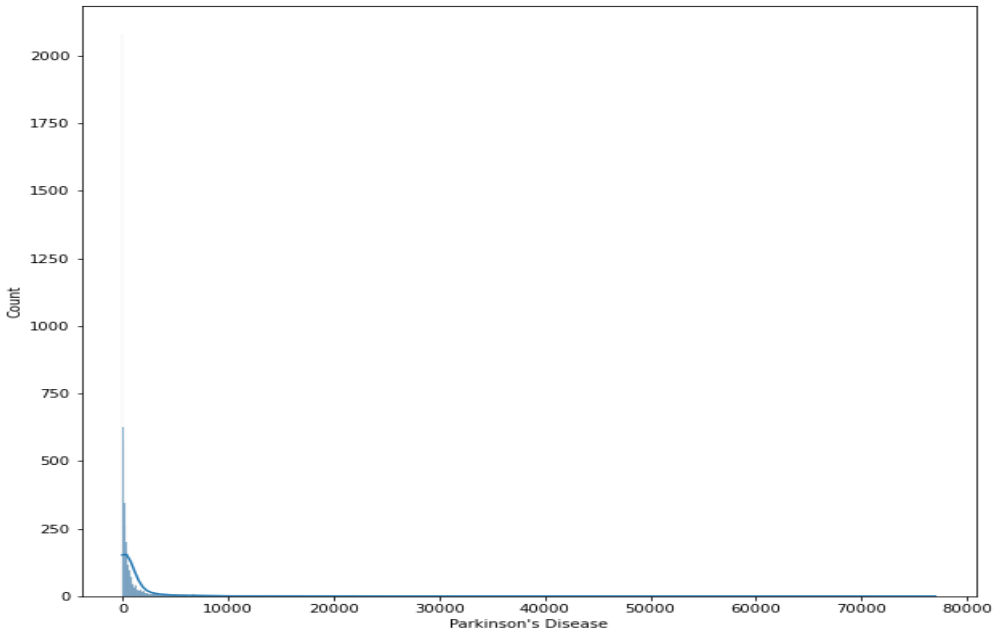

observation:-

we can observe 2020 unique values in the Meningitis column and the min value in this column is zero and the max value in the column is 98358.and we can observe maxumum observations are taking place in zero in the column



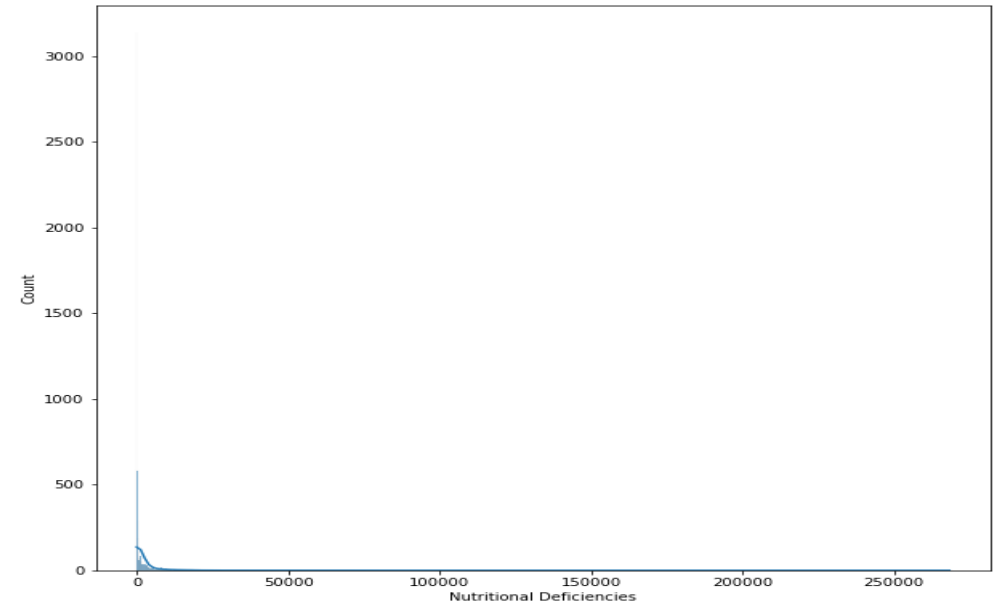

observation:-

we can observe 3037 unique values in the 'Alzheimer s Disease and Other Dementias' column and the min value in this column is zero and the max value in the column is 320715.and we can observe maxumum observations are taking place in zero in the column




Observation:-

we can observe 1817 unique values in the Parkinson's Disease column and
the min value in this column is zero and the max value in the column is 76990.
and we can observe maxumum observations are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000

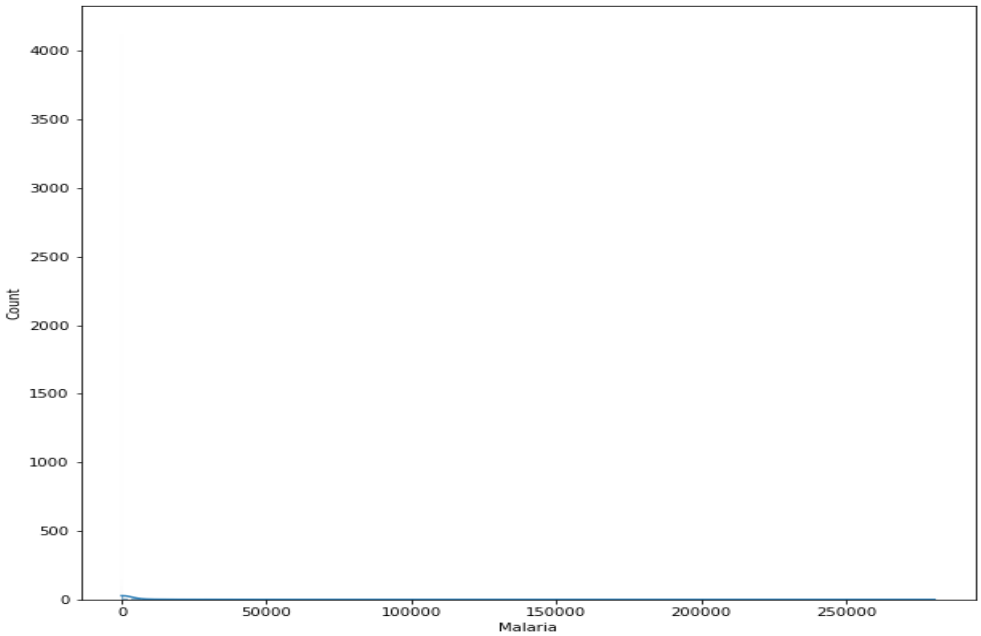



Observation:-

we can observe 2147 unique values in the Nutritional Deficiencies column and the min value in this column is zero and the max value in the column is 268223.

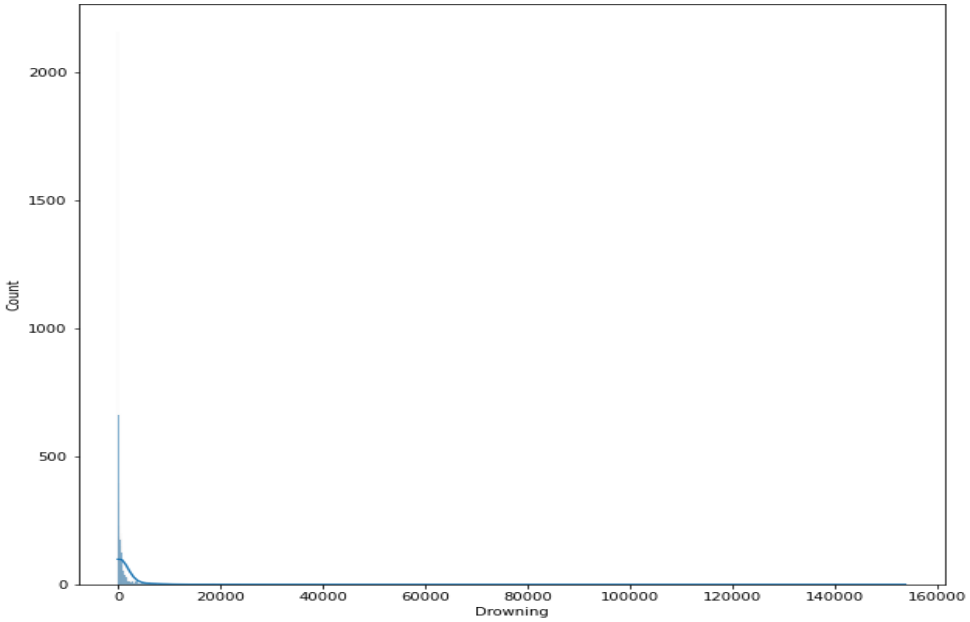
and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 50000



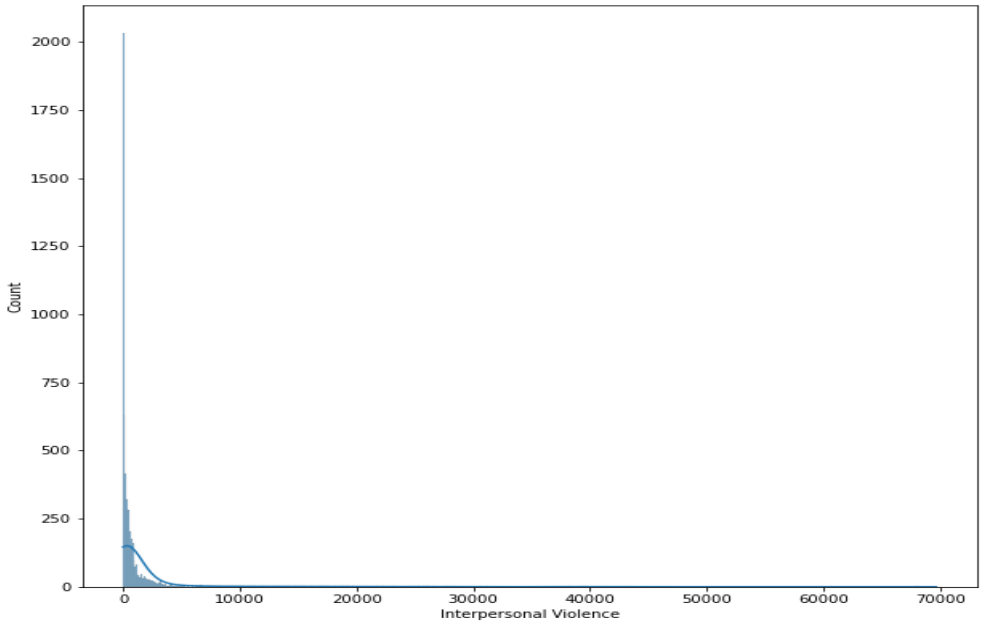

Observation:-

we can observe 1723 unique values in the Malaria column and
the min value in this column is zero and the max value in the column is 280604.
and we can observe maxumum death observed are taking place in zero in the column.



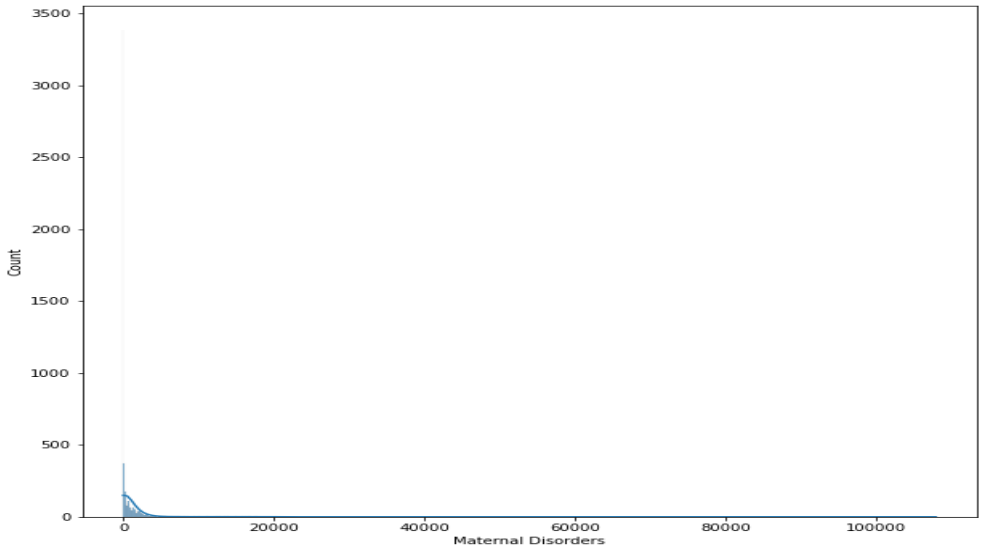

Observation:-

we can observe 1875 unique values in the Drowning column and
the min value in this column is zero and the max value in the column is 153773.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 20000



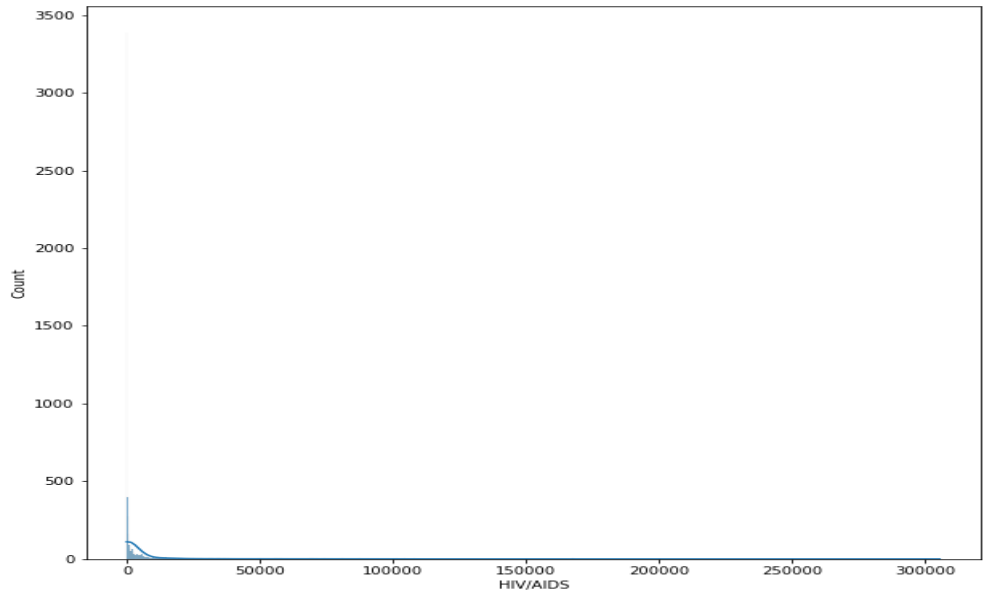

observation:-

we can observe 2142 unique values in the Interpersonal Violence column and
the min value in this column is zero and the max value in the column is 69640.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000



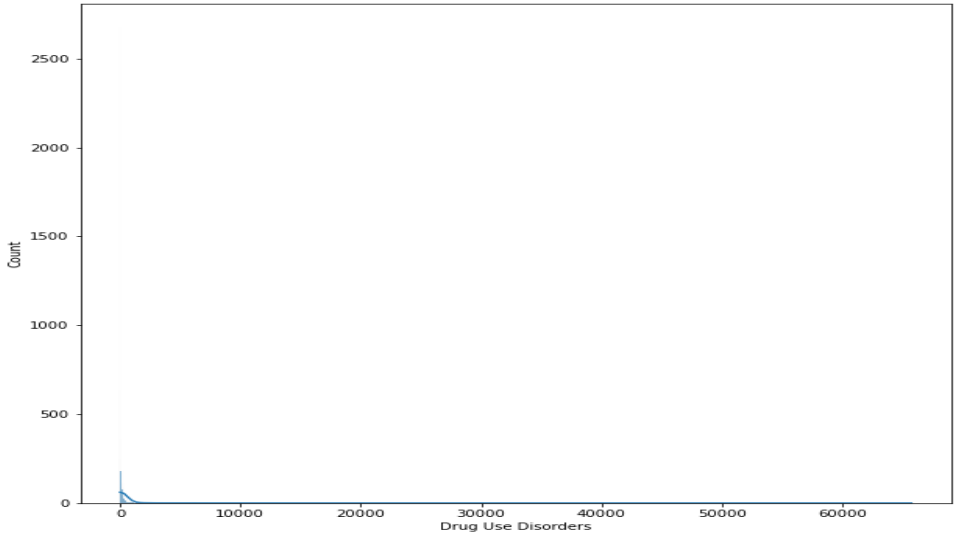

Observation:-

we can observe 1818 unique values in the Maternal Disorders column and
the min value in this column is zero and the max value in the column is 107929.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000



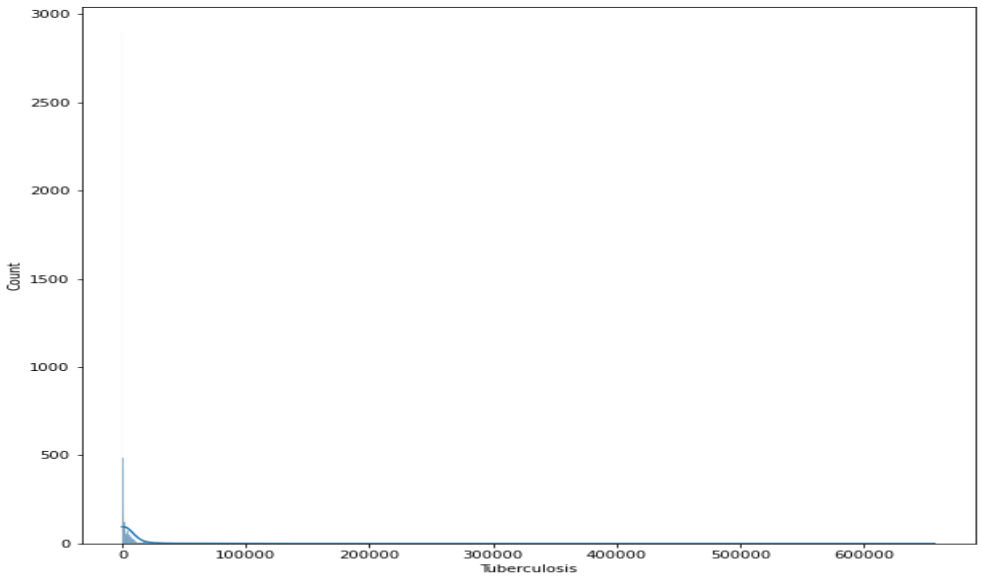

Observation:-

we can observe 2412 unique values in the HIV/AIDS column and
the min value in this column is zero and the max value in the column is 305491.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000



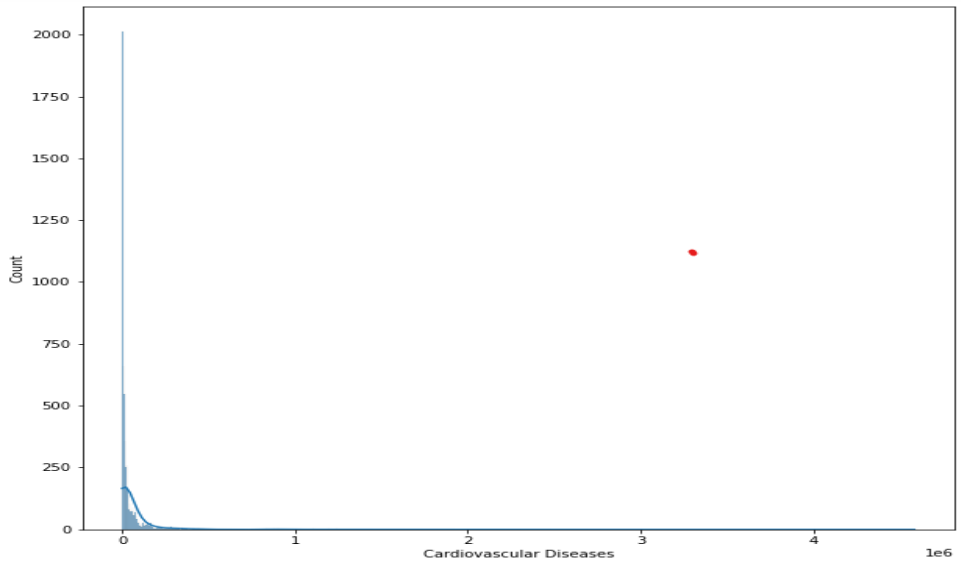

Observation:-

we can observe 876 unique values in the 'Drug Use Disorders' column and
the min value in this column is zero and the max value in the column is 65717.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000




Observation:-

we can observe 2843 unique values in the 'Tuberculosis' column and
the min value in this column is zero and the max value in the column is 657515.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000



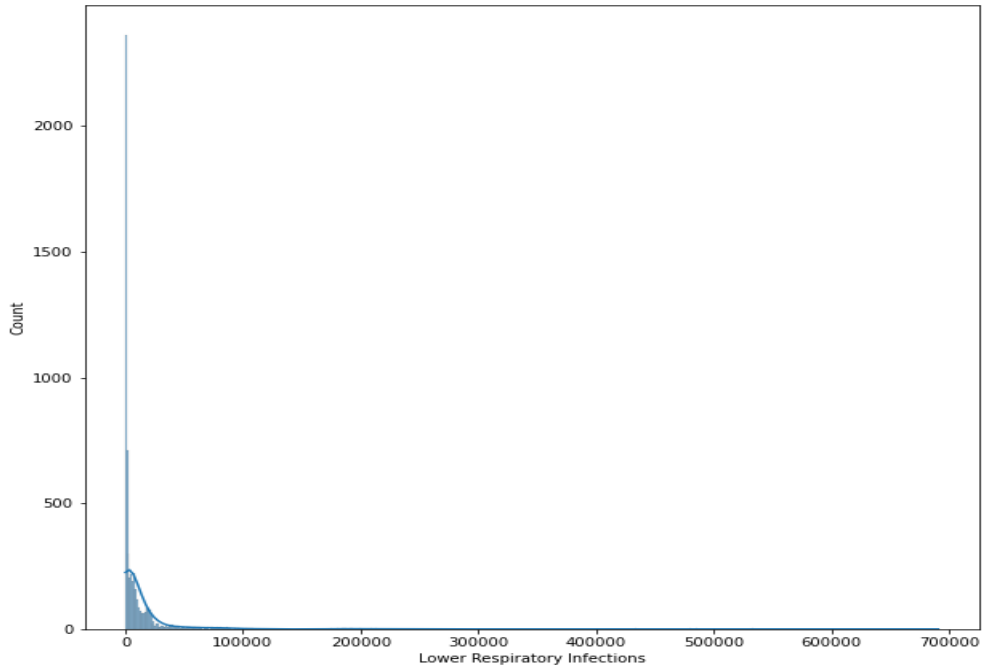

Observation:-

we can observe 5225 unique values in the 'Cardiovascular Diseases' column and

the min value in this column is zero and the max value in the column is 4584273.

and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 1e6



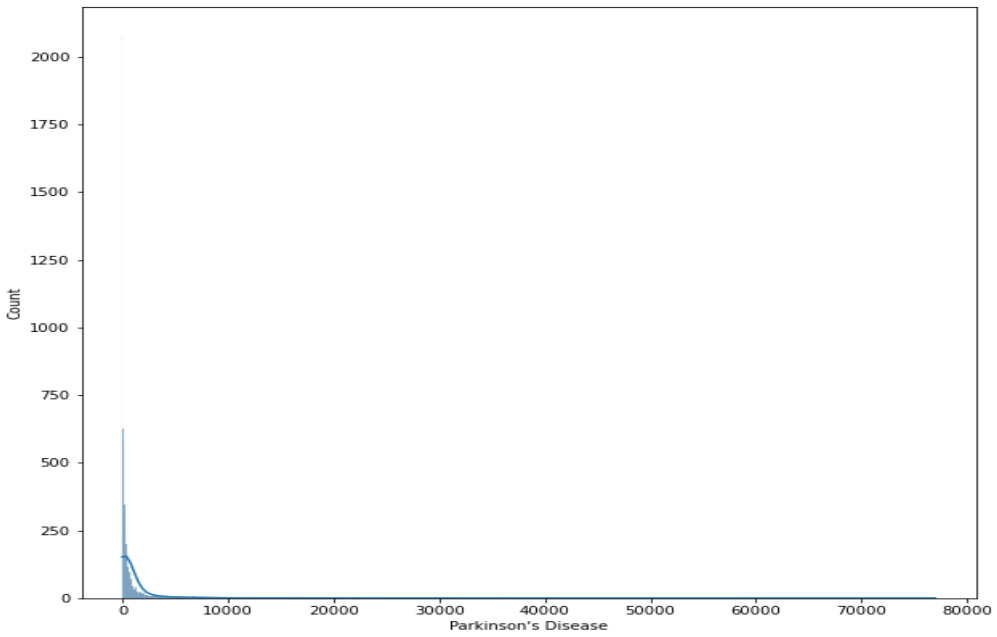

observation:- ]

we can observe 4106 unique values in the 'Lower Respiratory Infections' column and

the min value in this column is zero and the max value in the column is 690913.

and we can observe maxumum death observed are taking place in zero in the column.

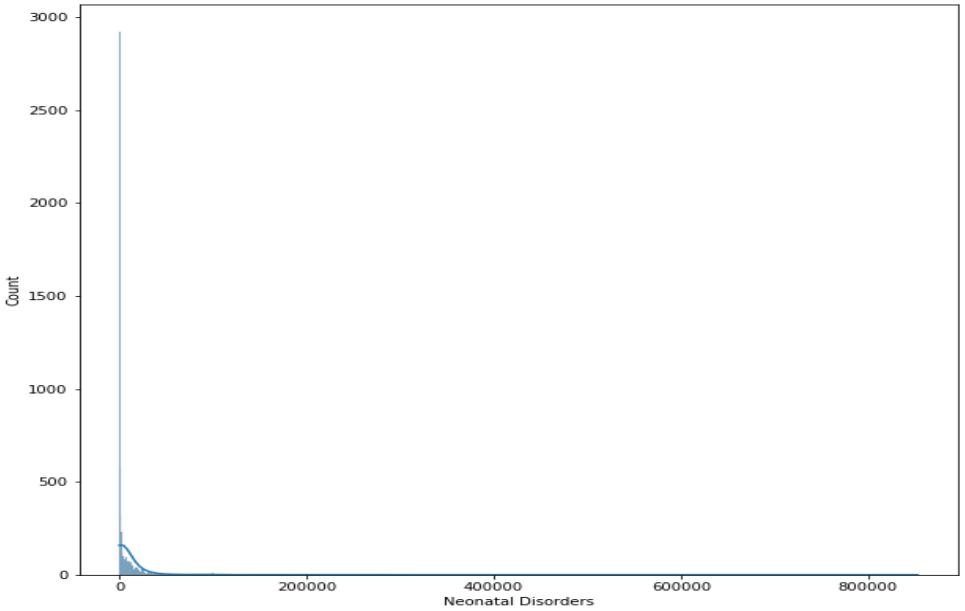
we can observe max count observations in the dataset are below 10000



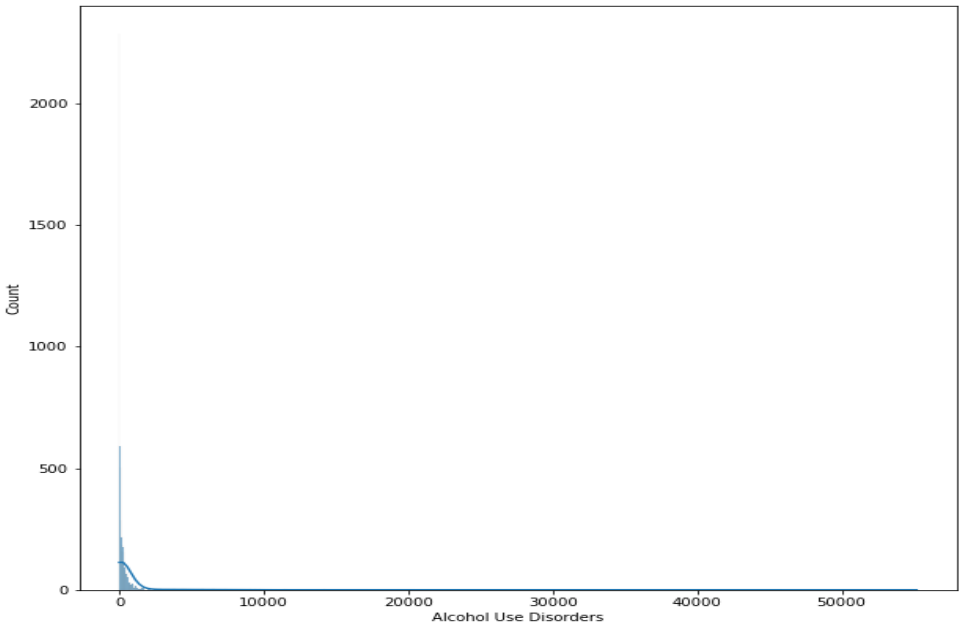

Observation:-

we can observe 1817 unique values in the Parkinson's Disease column and

the min value in this column is zero and the max value in the column is 76990.
and we can observe maxumum observations are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000




Observation:-
we can observe 3553 unique values in the 'Neonatal Disorders' column and
the min value in this column is zero and the max value in the column is 852761.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 20000



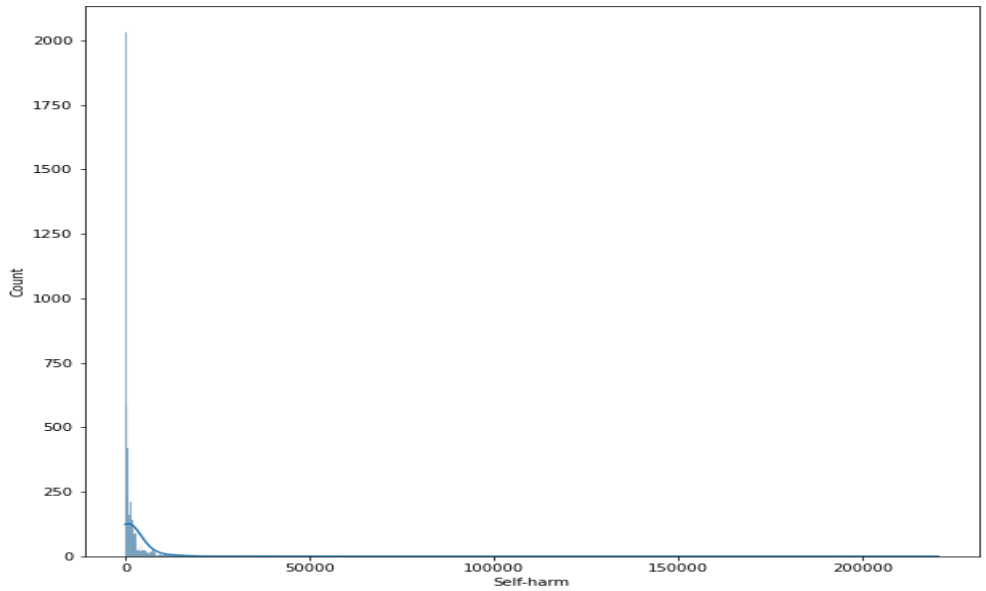

Observation:-

we can observe 1287 unique values in the 'Alcohol Use Disorders' column and

the min value in this column is zero and the max value in the column is 55200.

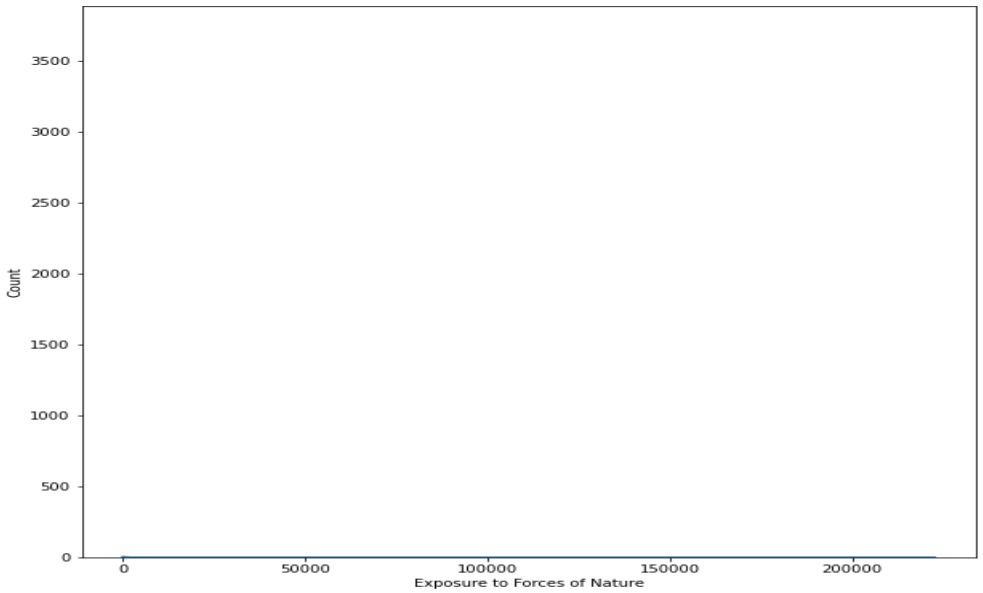
and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 10000




Observation:-

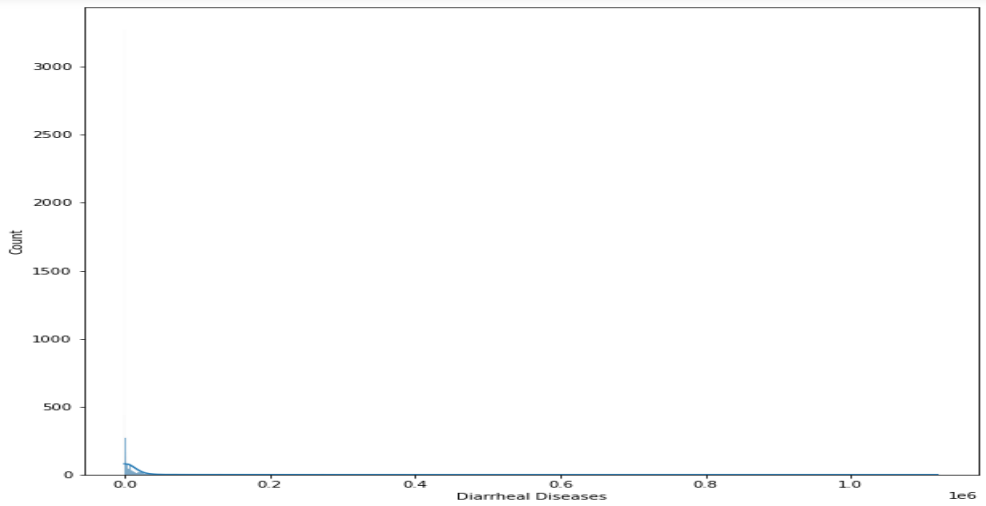
we can observe 2758 unique values in the 'Self-harm' column and
the min value in this column is zero and the max value in the column is 220357.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000



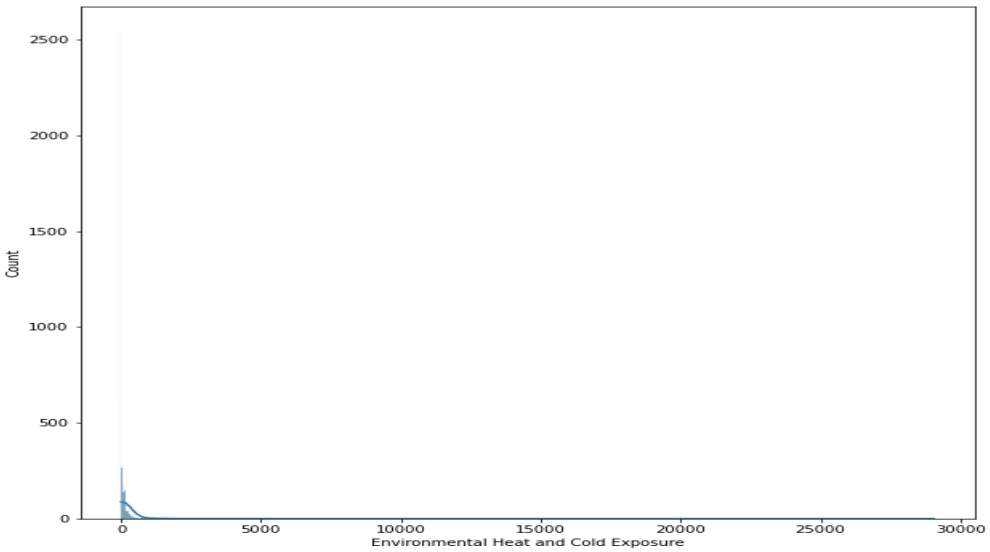

Observation:-

we can observe 478 unique values in the 'Exposure to Forces of Nature' column and

the min value in this column is zero and the max value in the column is 222641.and we can observe maxumum death observed are taking place in zero in the column

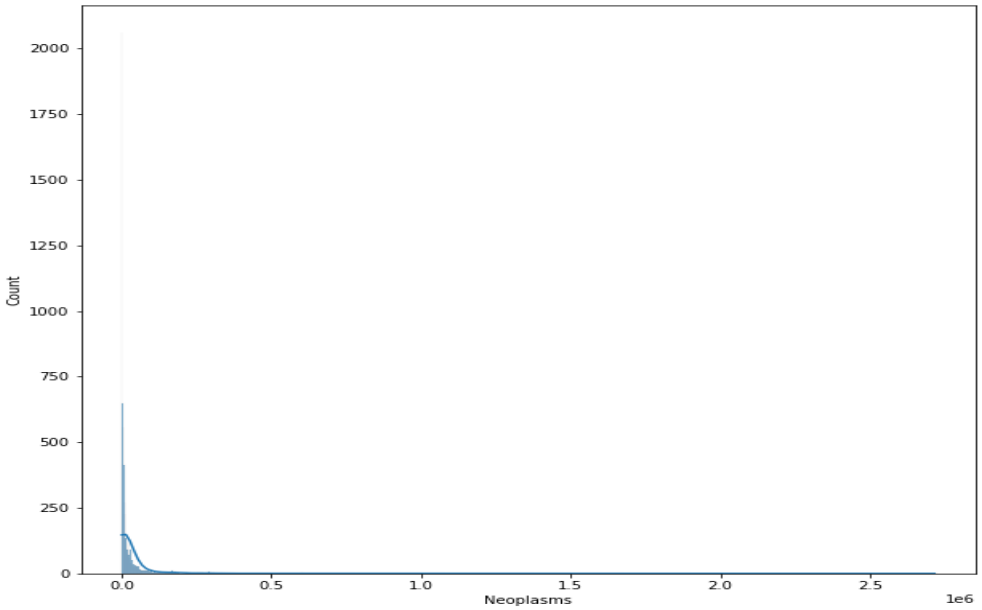



Observation:- we can observe 2874 unique values in the 'Diarrheal Diseases' column and the min value in this column is zero and the max value in the column is 1119477. and we can observe maxumum death observed are taking place in zero in the column.we can observe max count observations in the dataset are below 0.2 unit




Observation:-

we can observe 714 unique values in the 'Environmental Heat and Cold Exposure' column and the min value in this column is zero and the max value in the column is 29048.and we can observe maxumum death observed are taking place in zero in the column.we can observe max count observations in the dataset are below 5000



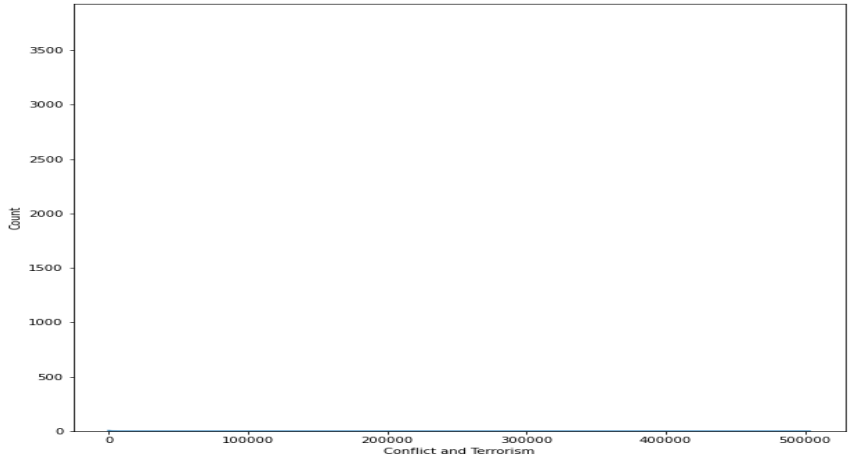

observation:-

we can observe 4814 unique values in the 'Neoplasms' column and

the min value in this column is zero and the max value in the column is 2716551.

and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 0.5 unit

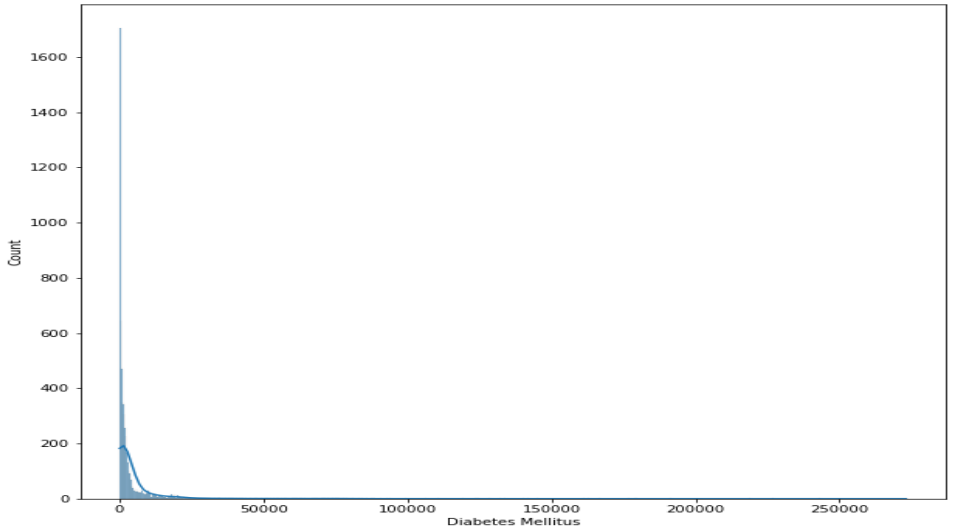



Observation:-

we can observe 918 unique values in the 'Conflict and Terrorism' column and

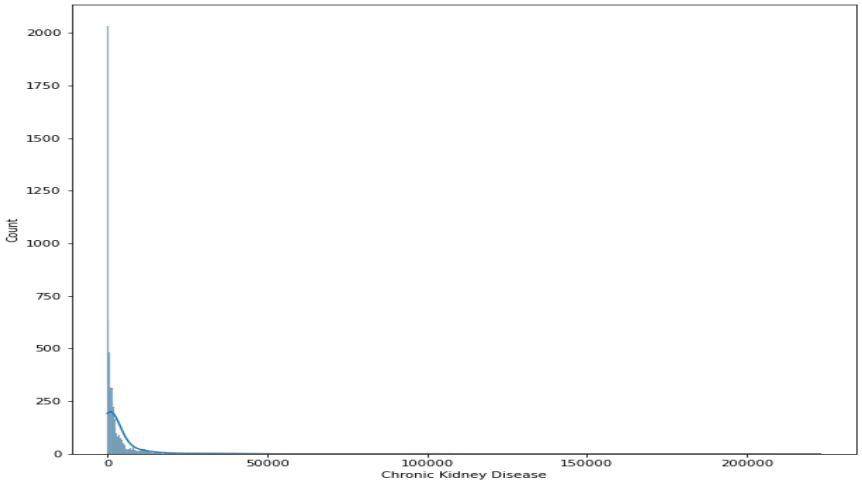
the min value in this column is zero and the max value in the column is 503532.

and we can observe maxumum death observed are taking place in zero in the column



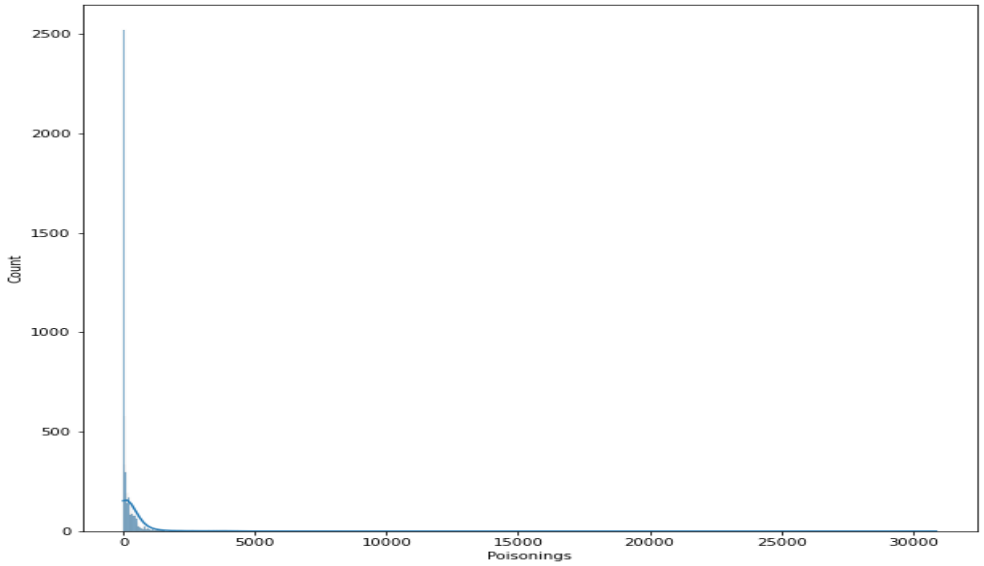

Observation:-

we can observe 3366 unique values in the Diabetes Mellitus column and
the min value in this column is zero and the max value in the column is 273089.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000




Observation:-

we can observe 3246 unique values in the 'Chronic Kidney Disease' column and
the min value in this column is zero and the max value in the column is 222922.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000



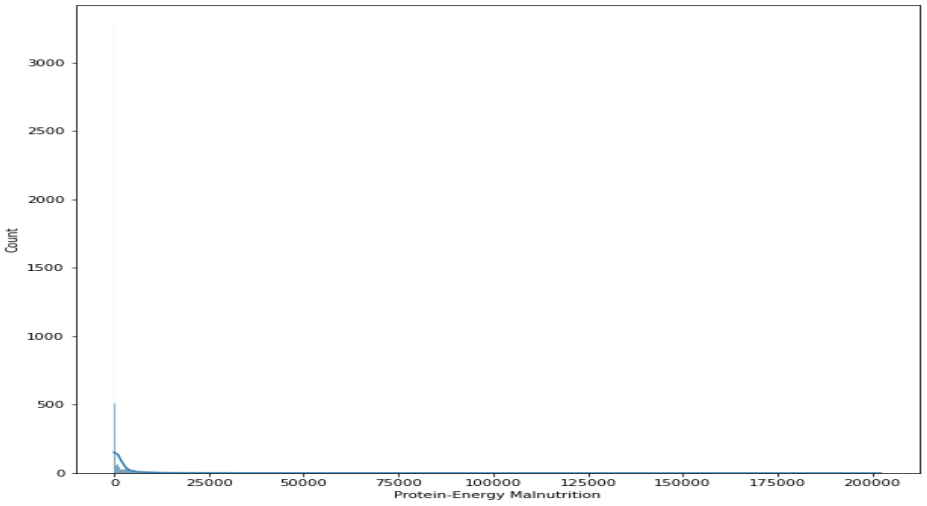

Observation :-

we can observe 1087 unique values in the 'Poisonings' column and

the min value in this column is zero and the max value in the column is 30883.

and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 5000



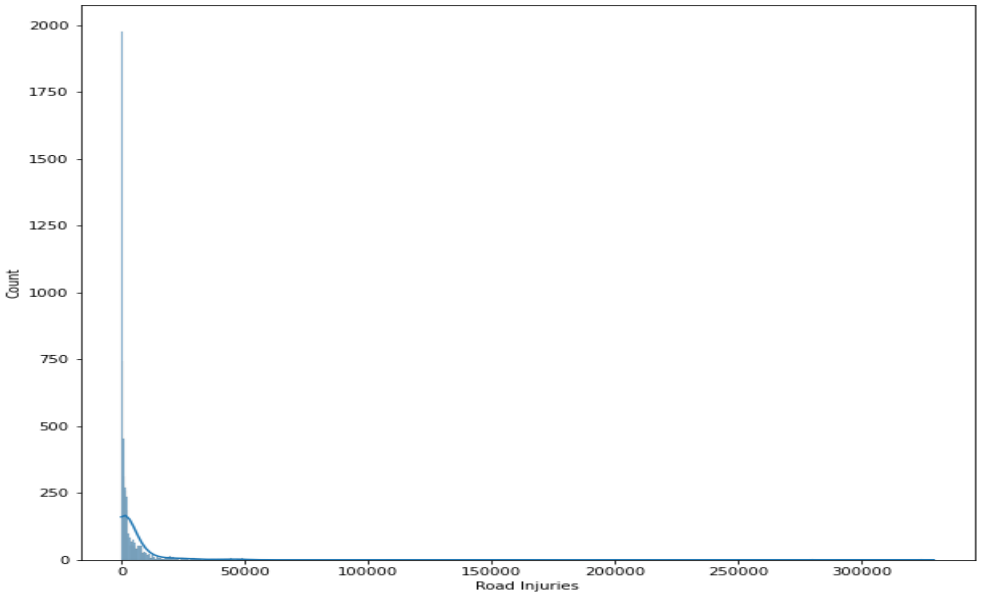

Observation:-

we can observe 2091 unique values in the 'Protein-Energy Malnutrition' column and

the min value in this column is zero and the max value in the column is 202241.

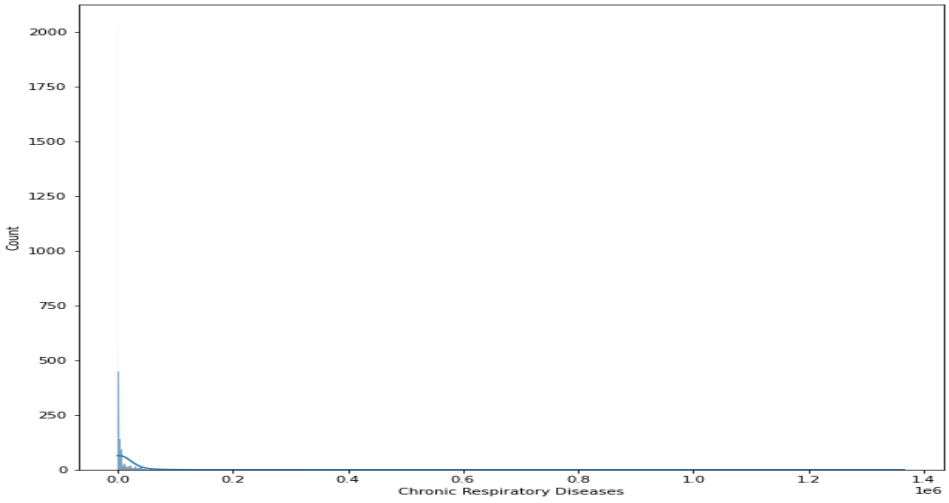
and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 25000



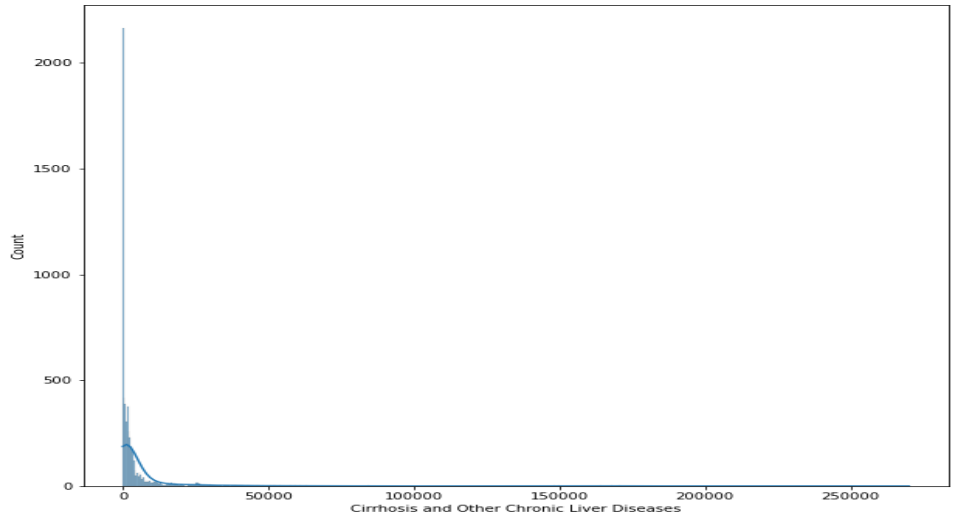

Observation:-

we can observe 3393 unique values in the 'Road Injuries' column and
the min value in this column is zero and the max value in the column is 329237.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000




Observation:-

we can observe 3393 unique values in the 'Chronic Respiratory Diseases' column and
the min value in this column is zero and the max value in the column is 1366039.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 0.2 units



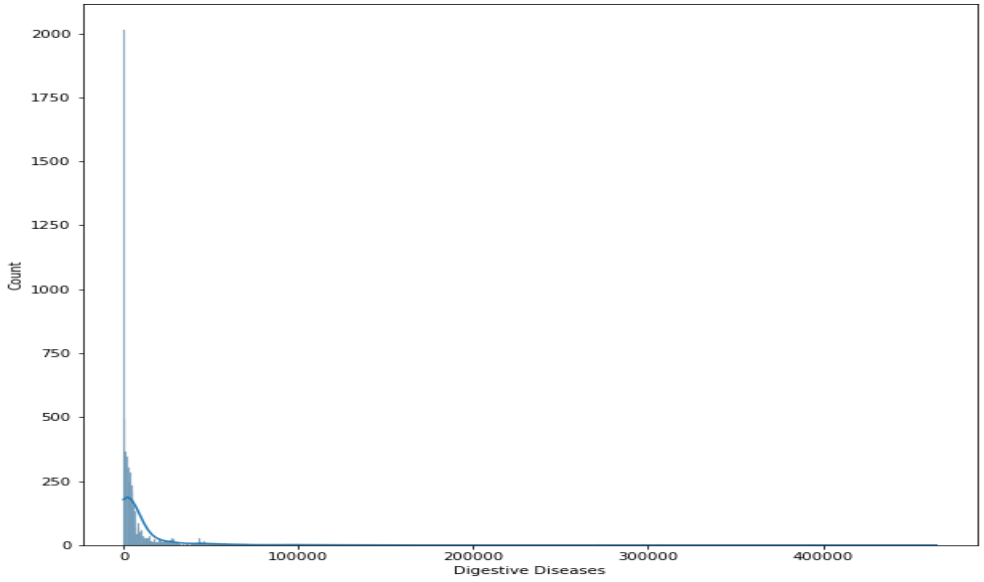

Observation:-

we can observe 3443 unique values in the 'Cirrhosis and Other Chronic Liver Diseases' column and the min value in this column is zero and the max value in the column is 270037.

and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 50000

.




Being able to guarentee the privacy of the customer

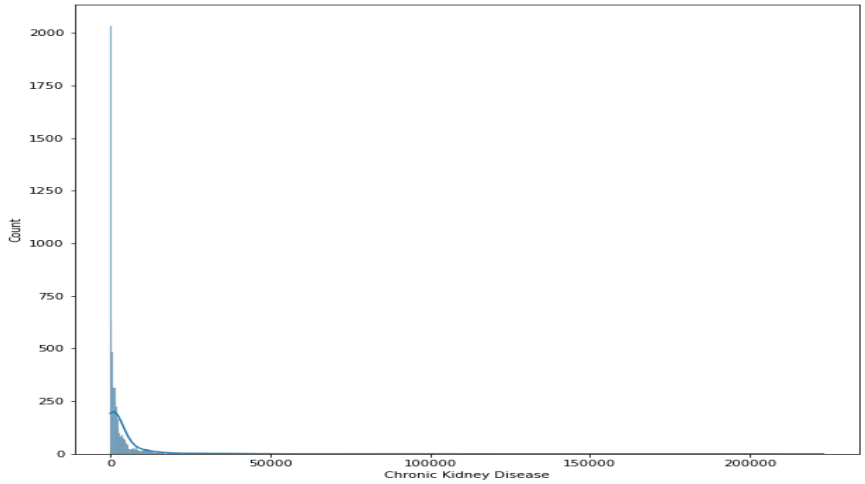
Observation:-

we can observe 4023 unique values in the 'Digestive Diseases' column and

the min value in this column is zero and the max value in the column is 464914.

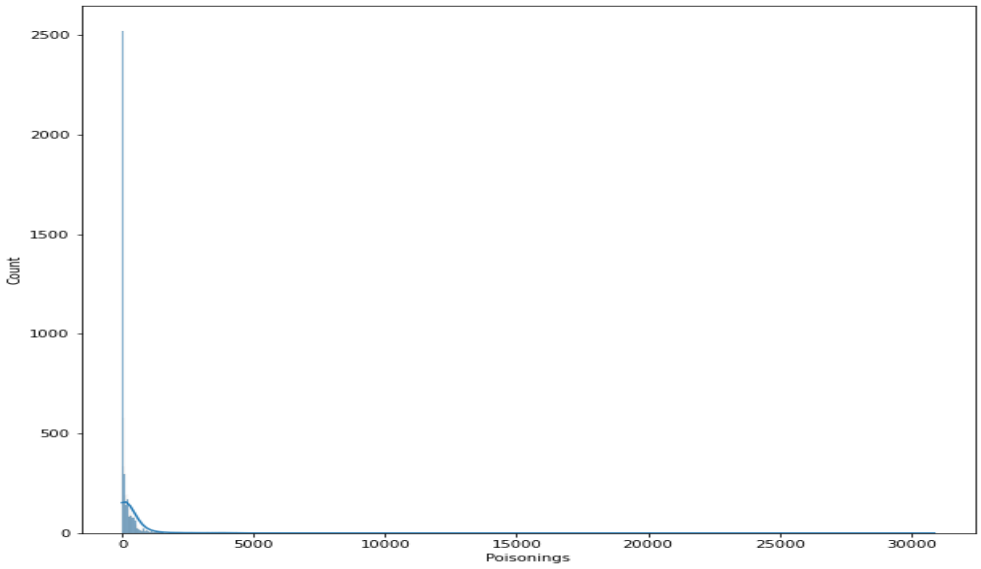
and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 1000000



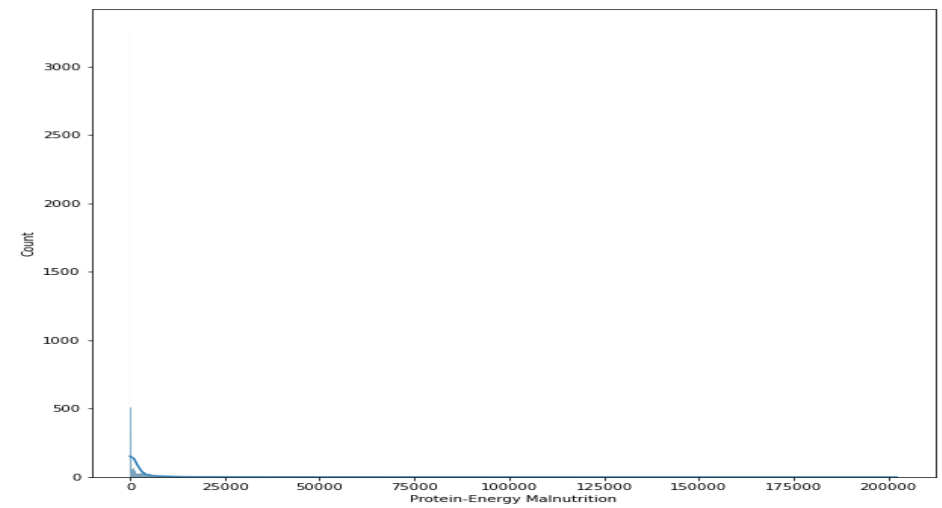

Observation:-

we can observe 3246 unique values in the 'Chronic Kidney Disease' column and
the min value in this column is zero and the max value in the column is 222922.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000




Observation:-

we can observe 1087 unique values in the 'Poisonings' column and
the min value in this column is zero and the max value in the column is 30883.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 5000



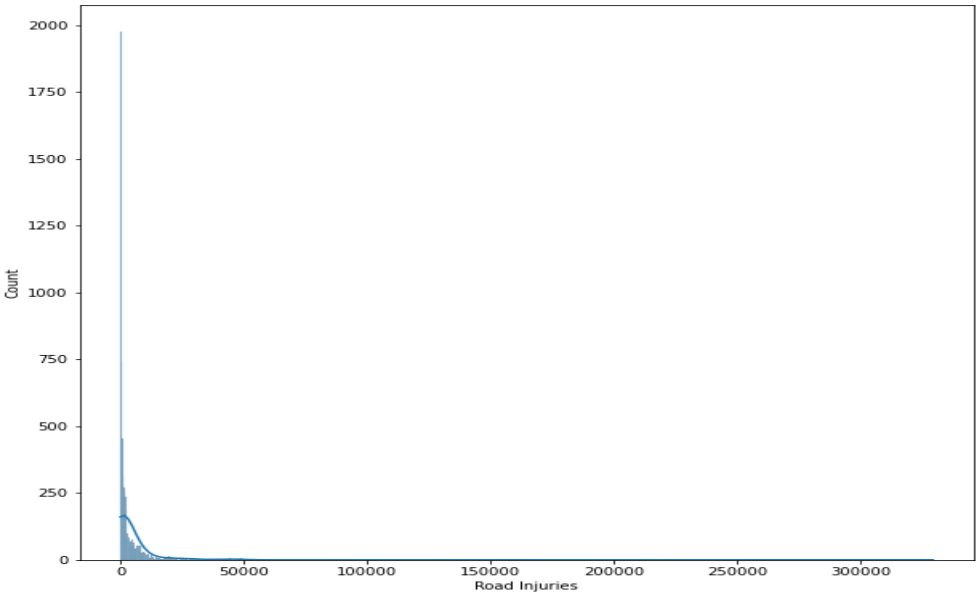

Observation:-

we can observe 2091 unique values in the 'Protein-Energy Malnutrition' column and

the min value in this column is zero and the max value in the column is 202241.

and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 25000



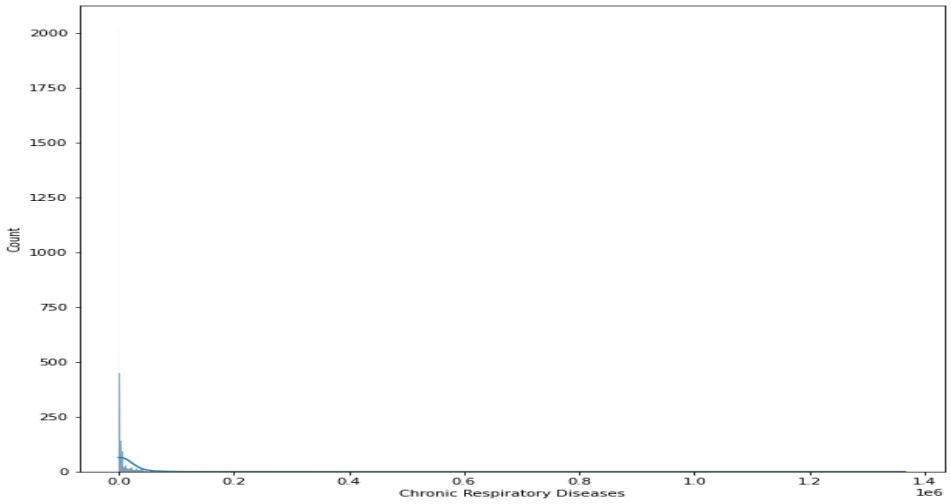

Observation:-

we can observe 3393 unique values in the 'Road Injuries' column and

the min value in this column is zero and the max value in the column is 329237.

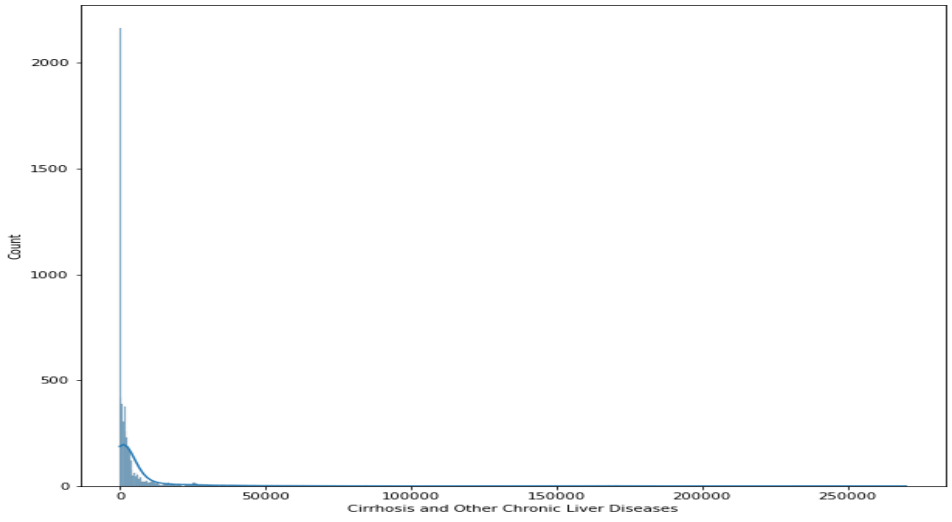
and we can observe maxumum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 50000



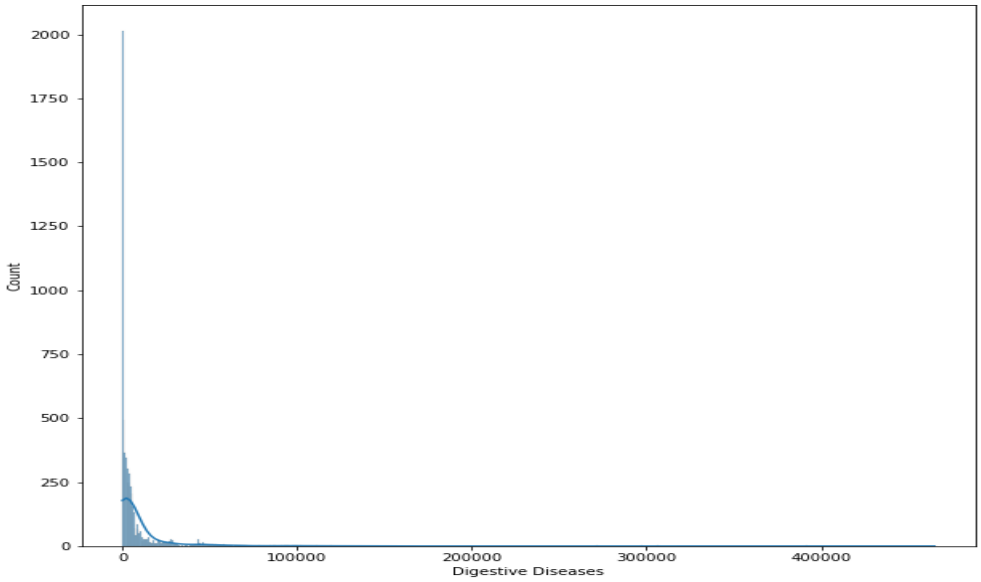

Observation:-

we can observe 3393 unique values in the 'Chronic Respiratory Diseases' column and the min value in this column is zero and the max value in the column is 1366039.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 0.2 units




Observation:-

we can observe 3443 unique values in the 'Cirrhosis and Other Chronic Liver Diseases' column and the min value in this column is zero and the max value in the column is 270037.
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 50000



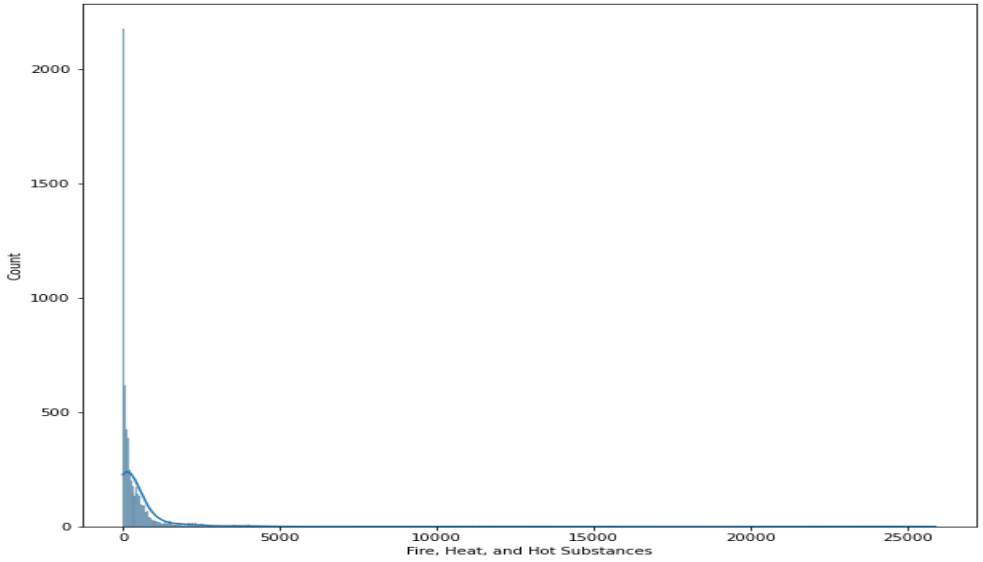

Observation:-

we can observe 4023 unique values in the 'Digestive Diseases' column and

the min value in this column is zero and the max value in the column is 464914.

and we can observe maximum death observed are taking place in zero in the column.

we can observe max count observations in the dataset are below 1000000



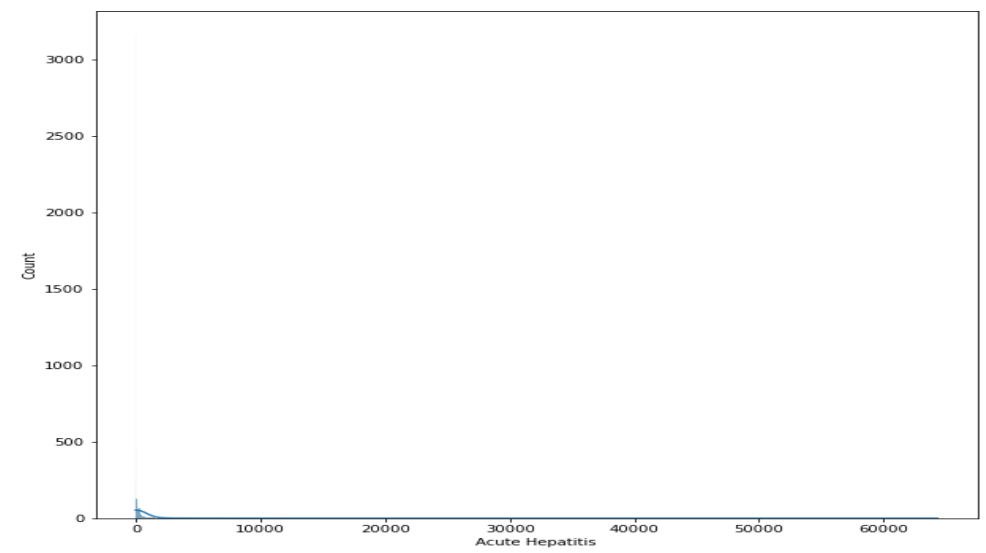

Observation:-

we can observe 1406 unique values in the 'Fire, Heat, and Hot Substances' column and

the min value in this column is zero and the max value in the column is 25876.

and we can observe maxumum death observed are taking place in zero in the column.

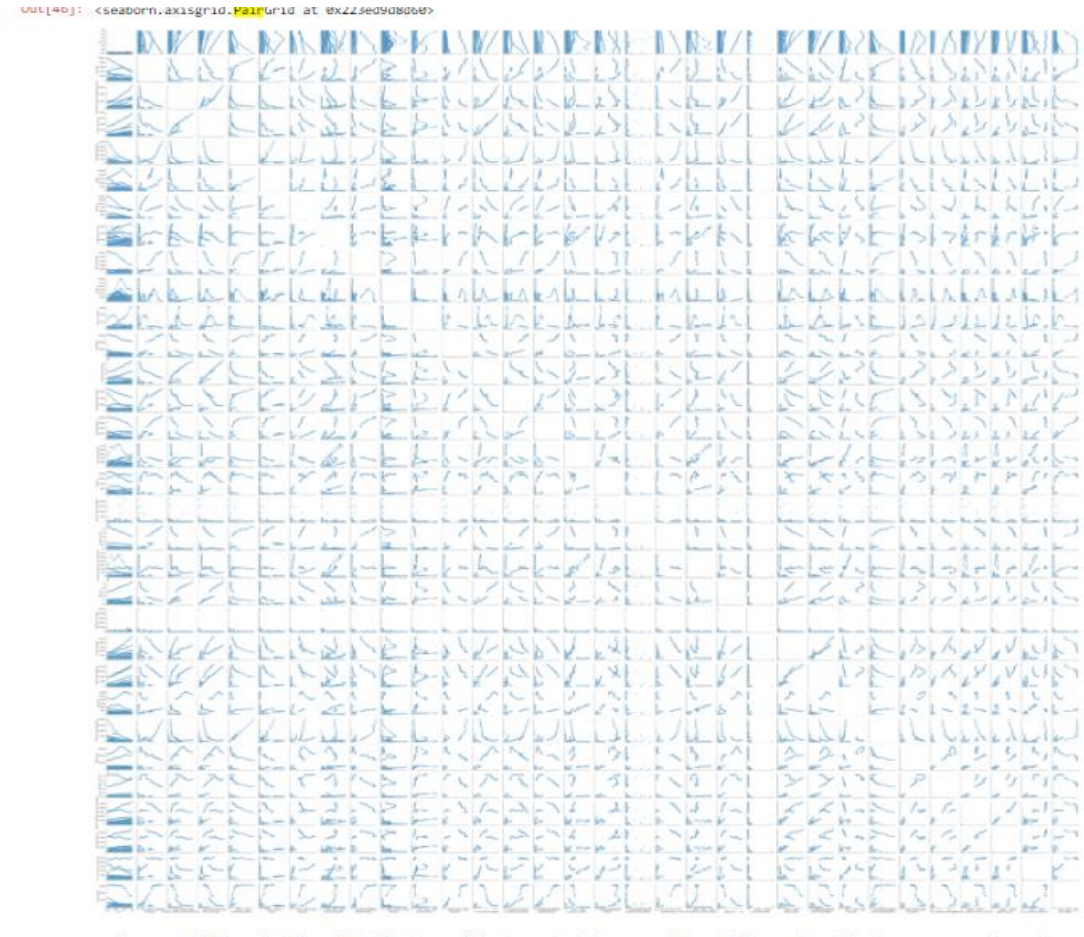
we can observe max count observations in the dataset are below 5000



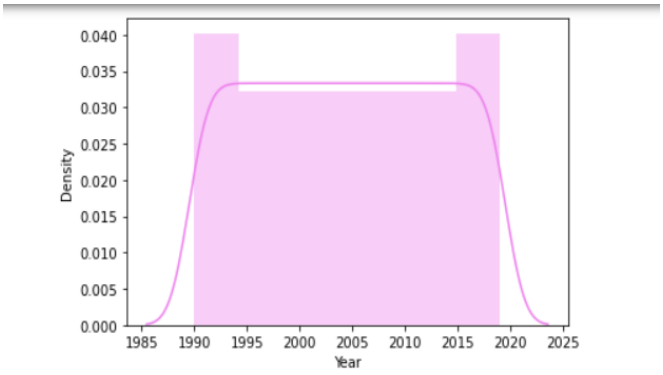

Observation:-

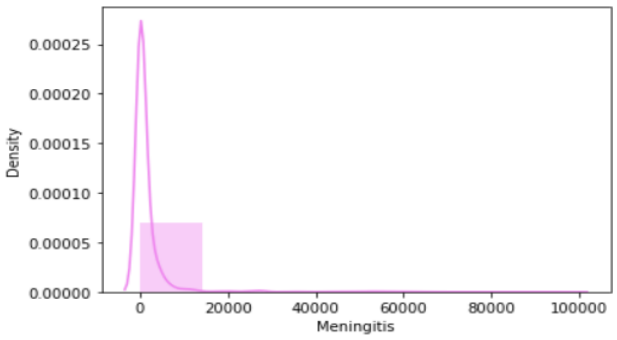
we can observe 1059 unique values in the 'Acute Hepatitis' column and
the min value in this column is zero and the max value in the column is 64305
and we can observe maxumum death observed are taking place in zero in the column.
we can observe max count observations in the dataset are below 10000

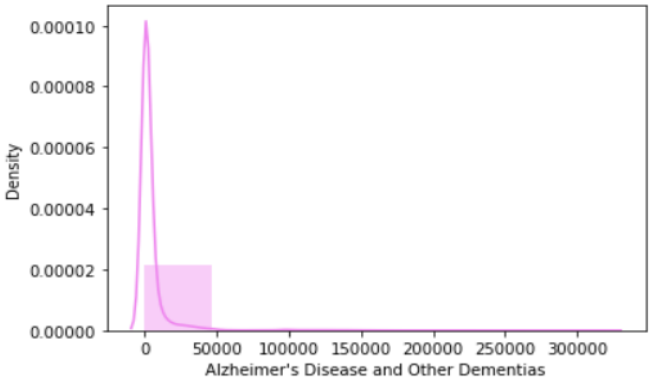
Pair plot

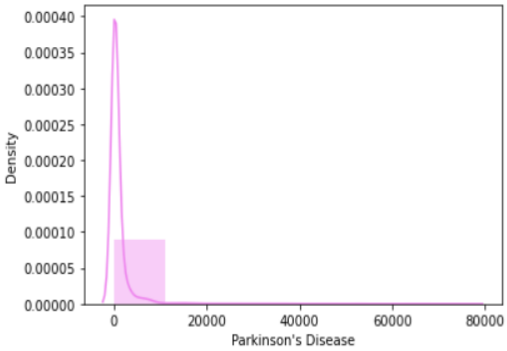


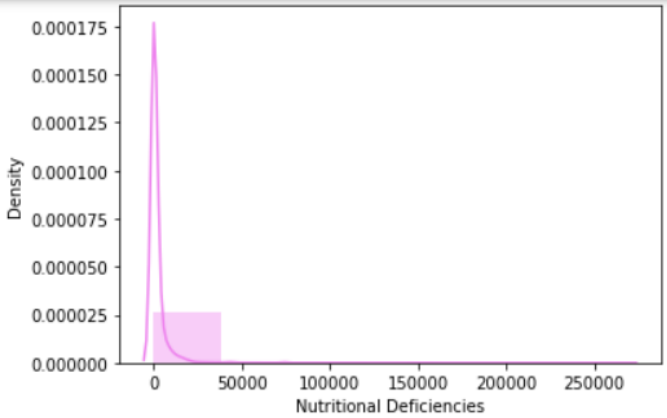

## Visualizing the Skewness

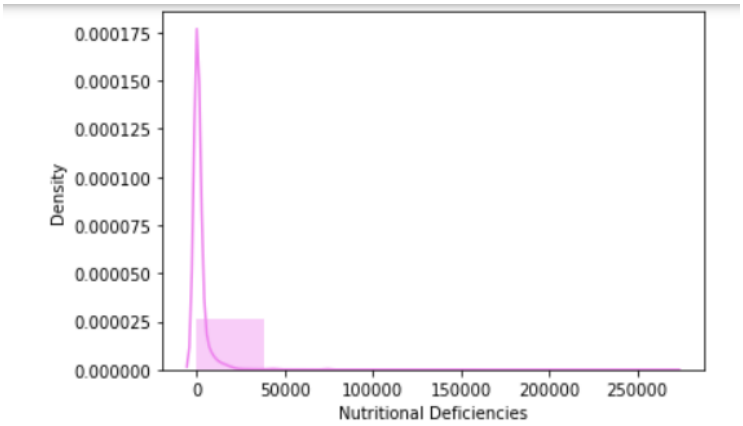

Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
1985
1990
1995
2000
2005
2010
2015
2020
2025
Year

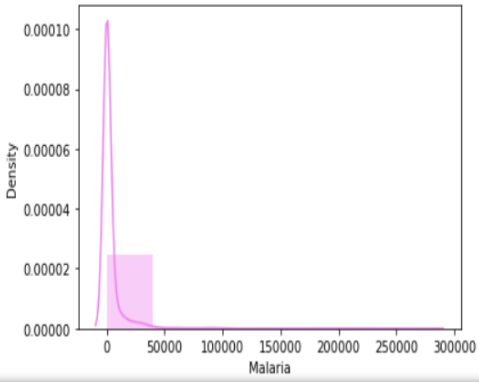

Density
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0
20000
40000
60000
80000
100000
Meningitis

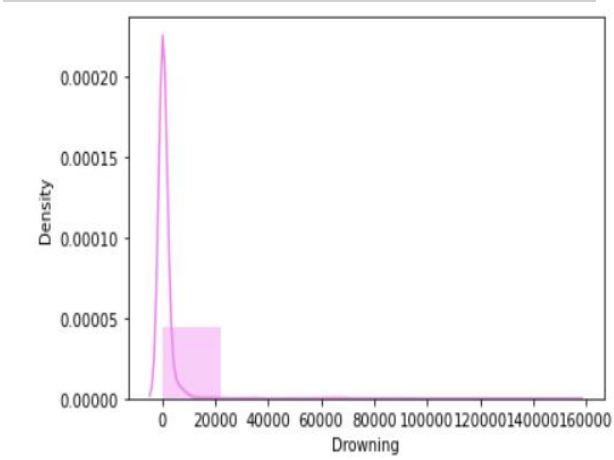

Density
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
50000
100000
150000
200000
250000
300000
Alzheimer's Disease and Other Dementias

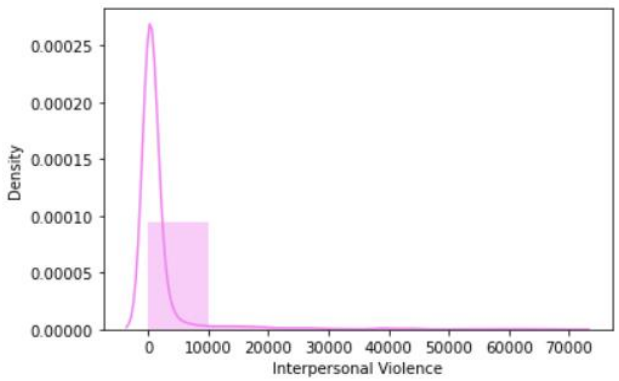

Density
0.00040
0.00035
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0
20000
40000
60000
80000
Parkinson's Disease

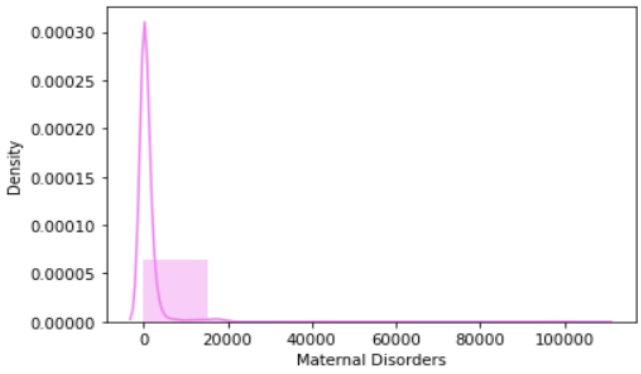

Density
0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
0
50000
100000
150000
200000
250000
Nutritional Deficiencies

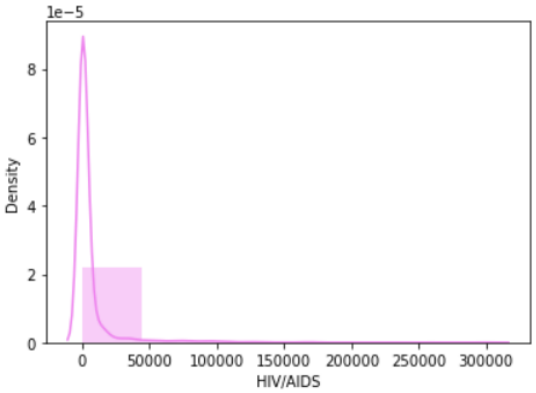

0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
Density
0
50000
100000
150000
200000
250000
Nutritional Deficiencies

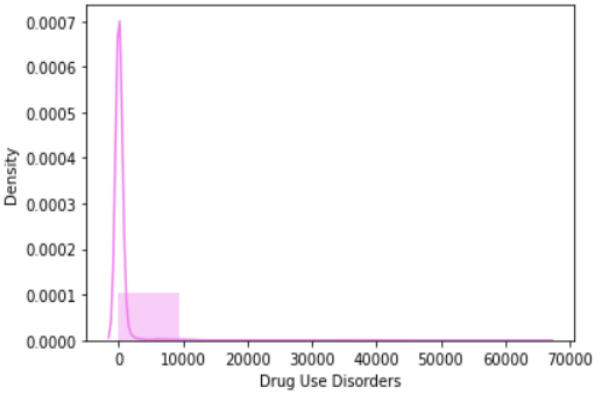

0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Density
0
50000
100000
150000
200000
250000
300000
Malaria

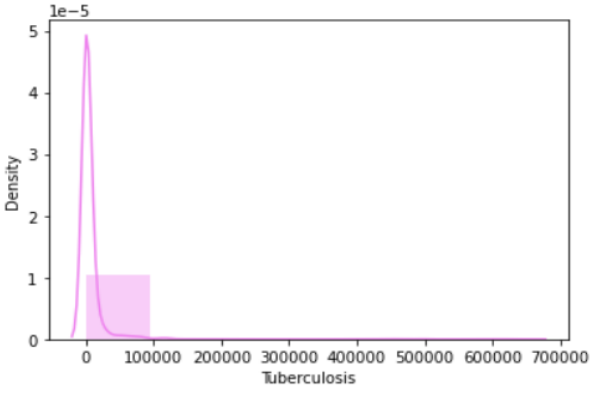

0.00020
0.00015
0.00010
0.00005
0.00000
Density
0
20000
40000
60000
80000
100000
120000
140000
160000
Drowning

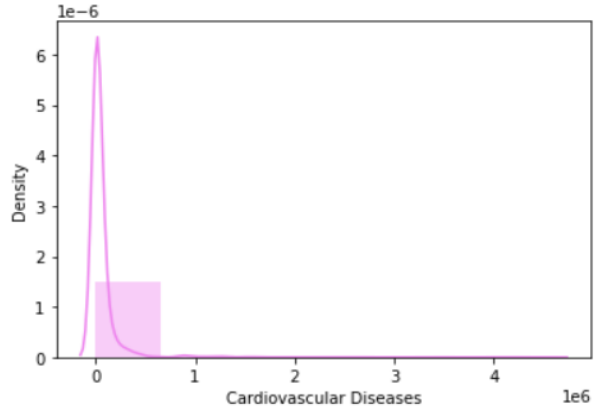

0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
Density
0
10000
20000
30000
40000
50000
60000
70000
Interpersonal Violence

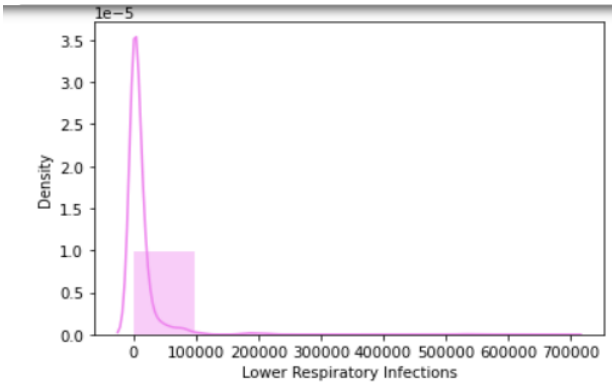

0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
Density
0
20000
40000
60000
80000
100000
Maternal Disorders

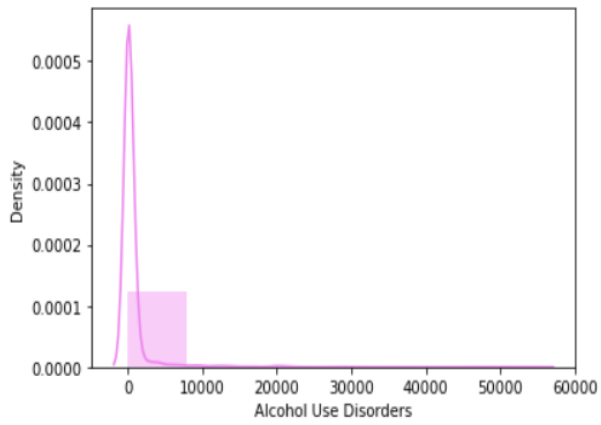

1e−5
8
6
4
2
0
Density
0
50000
100000
150000
200000
250000
300000
HIV/AIDS

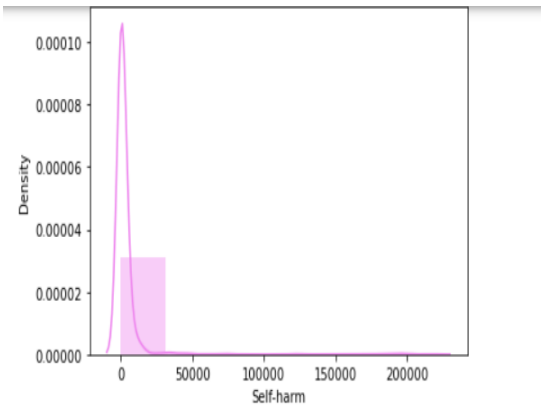

0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
Density
0
10000
20000
30000
40000
50000
60000
70000
Drug Use Disorders

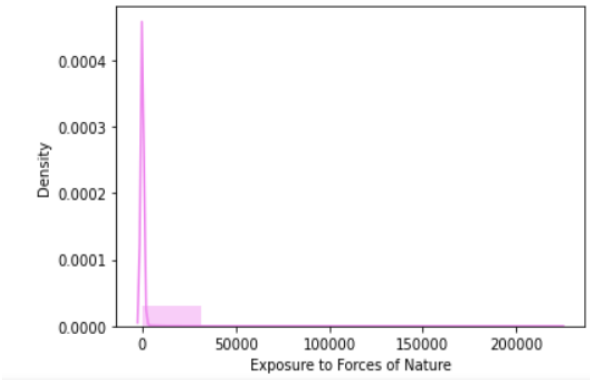

1e−5
5
4
3
2
1
0
Density
0
100000
200000
300000
400000
500000
600000
700000
Tuberculosis

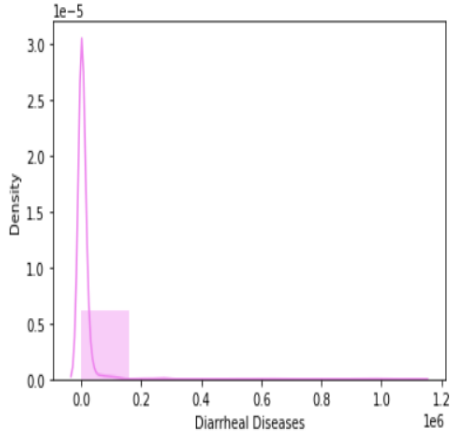

1e−6
Density
Cardiovascular Diseases
1e6

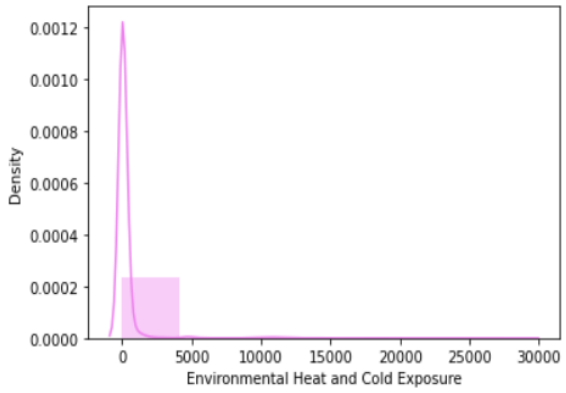

1e−5
Density
Lower Respiratory Infections

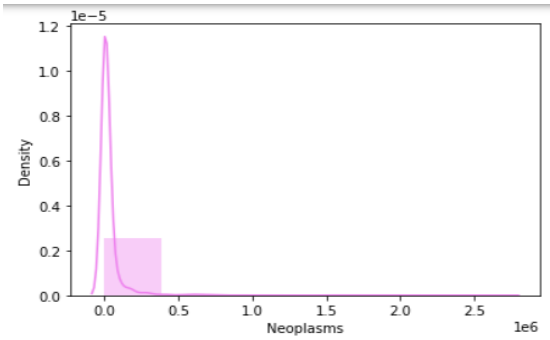

Density
Alcohol Use Disorders

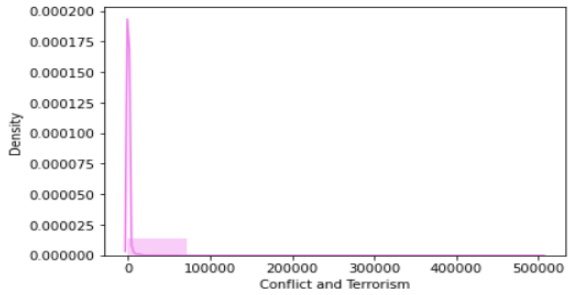

Density
Self-harm

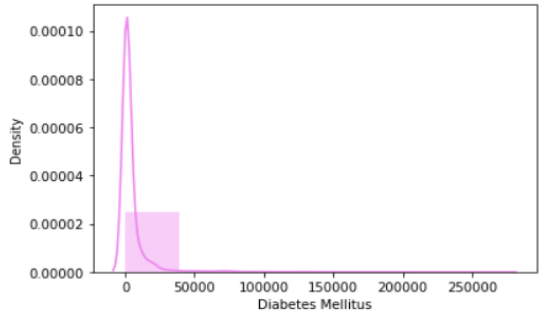

0.0004
0.0003
0.0002
0.0001
0.0000
Density
0
50000
100000
150000
200000
Exposure to Forces of Nature

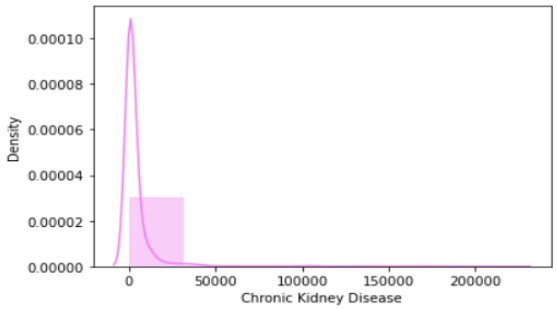

1e−5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Diarrheal Diseases
1e6

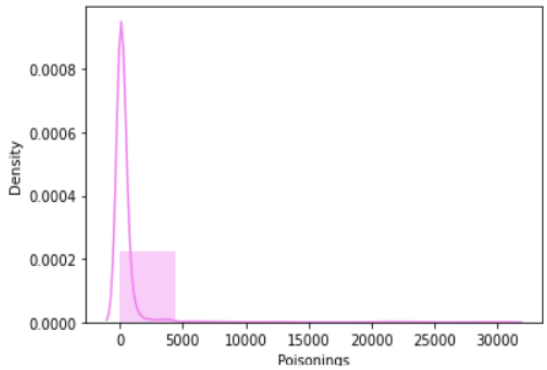

0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
Density
0
5000
10000
15000
20000
25000
30000
Environmental Heat and Cold Exposure

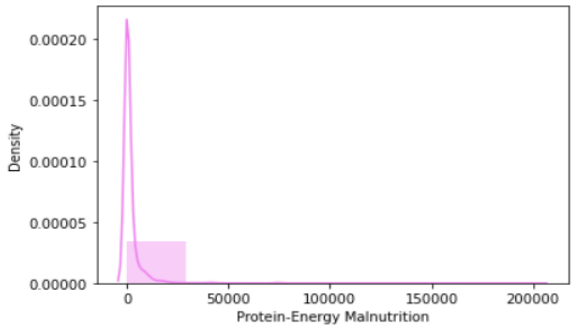

1e−5
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Density
0.0
0.5
1.0
1.5
2.0
2.5
Neoplasms
1e6

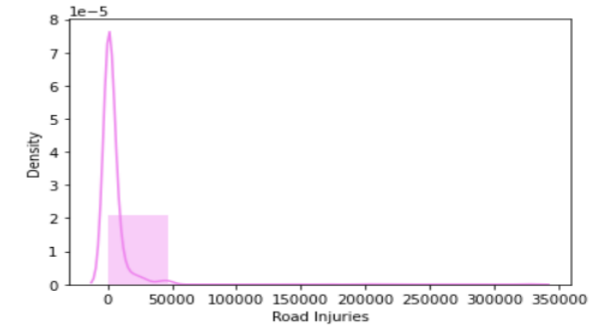

0.000200
0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
Density
0
100000
200000
300000
400000
500000
Conflict and Terrorism

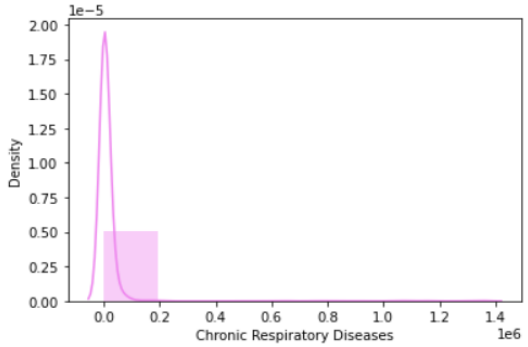

0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Density
0
50000
100000
150000
200000
250000
Diabetes Mellitus

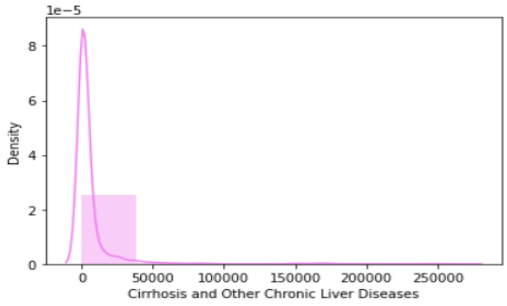

0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Density
0
50000
100000
150000
200000
Chronic Kidney Disease

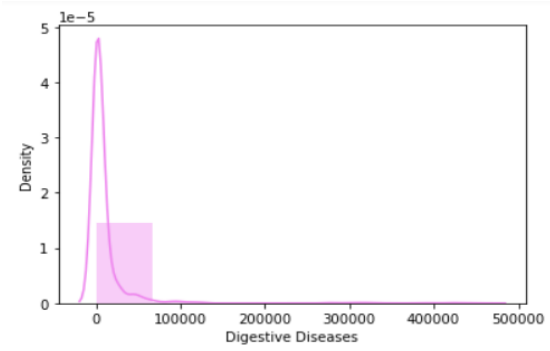

0.0008
0.0006
0.0004
0.0002
0.0000
Density
0
5000
10000
15000
20000
25000
30000
Poisonings

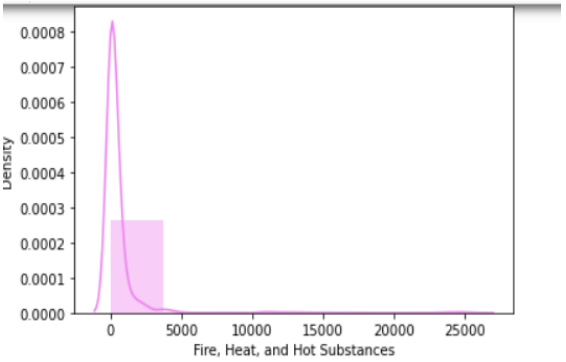

0.00020
0.00015
0.00010
0.00005
0.00000
Density
0
50000
100000
150000
200000
Protein-Energy Malnutrition

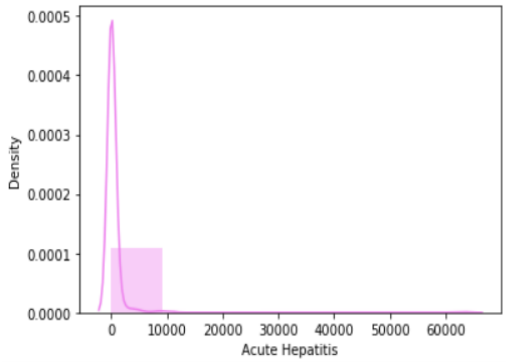

1e−5
Density
Road Injuries


1e−5
Density
Chronic Respiratory Diseases
1e6


1e−5
Density
Cirrhosis and Other Chronic Liver Diseases

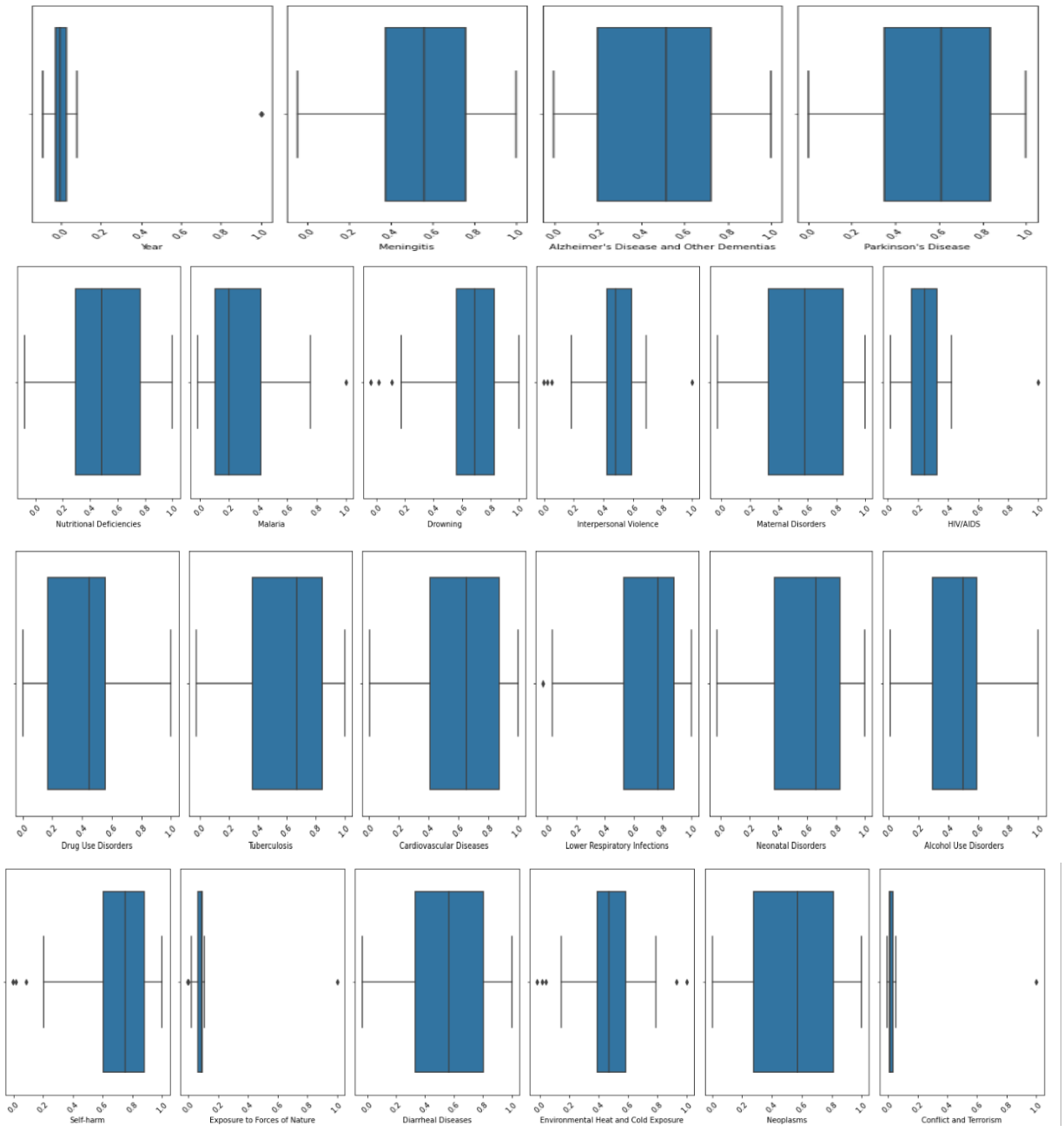

1e−5
Density
Digestive Diseases

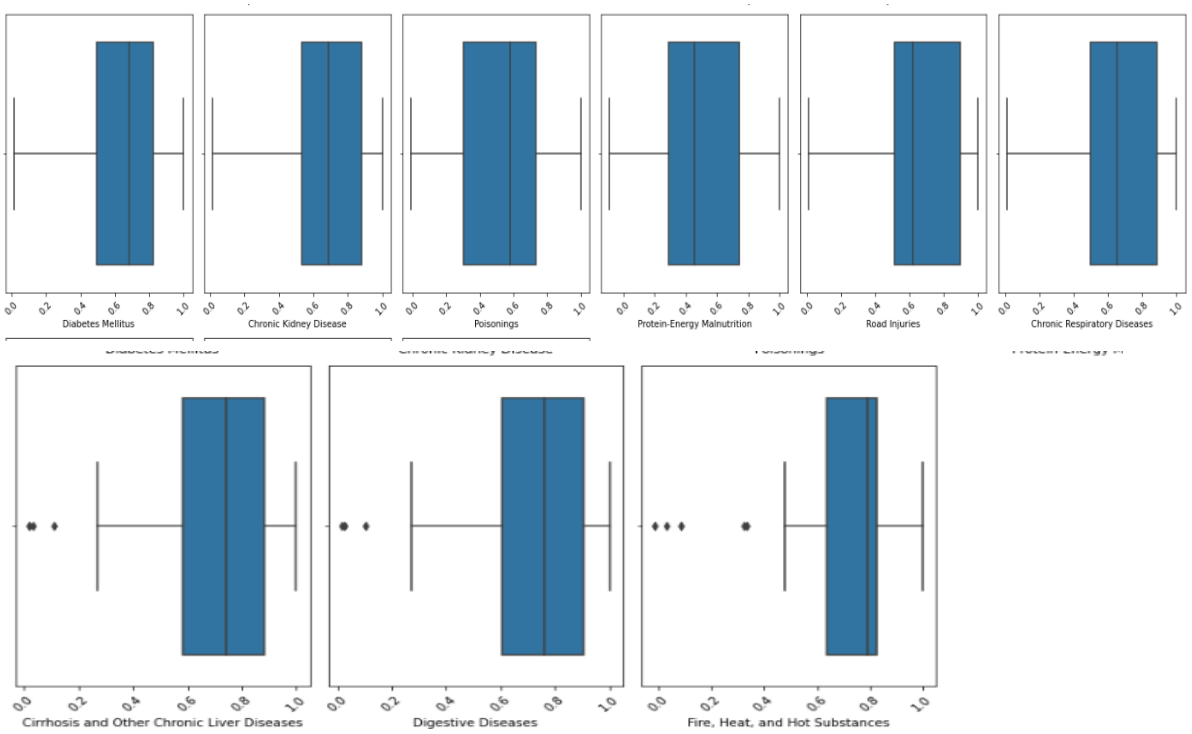

Density
Fire, Heat, and Hot Substances




Able to observe skewness in all columns except year in the continuous variables so we need to treat it

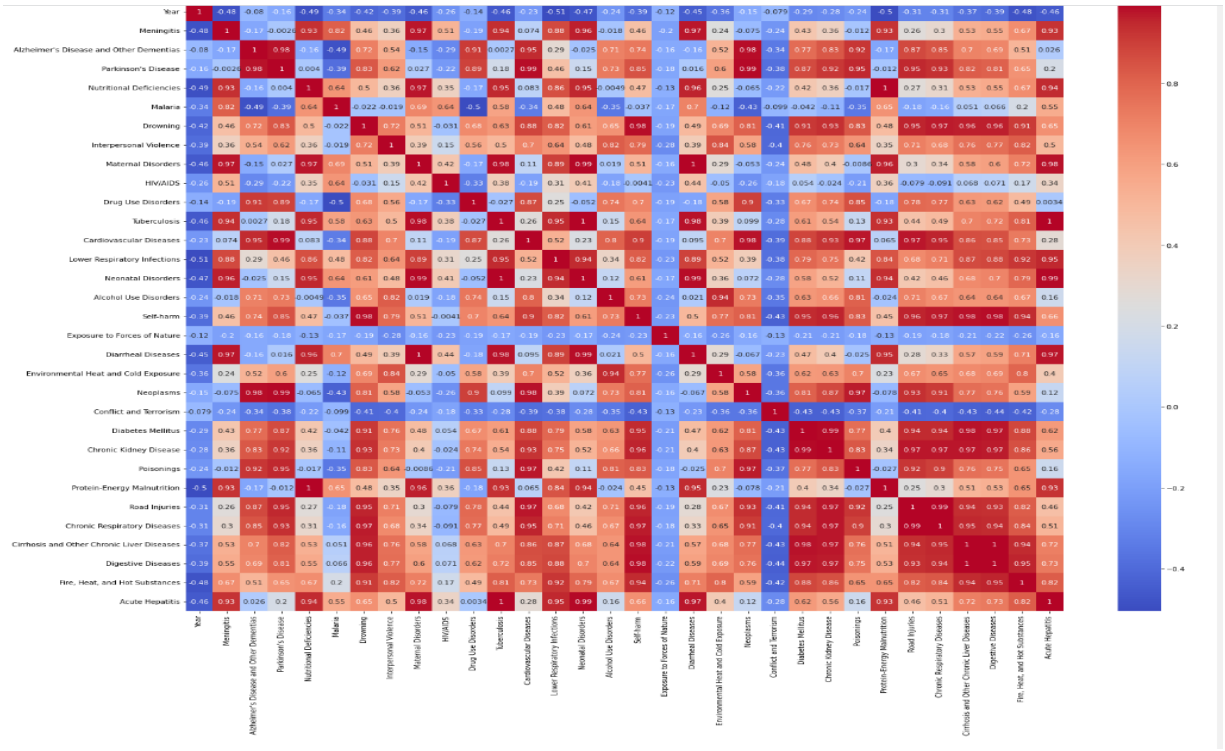
# Outliers visulaization






The outliers in the dataset is acceptable as they are unique data as it can occur due to difference in climate and environmental differences

## Heatmap



# CONCLUSION

This study helped in demonstrating proficiency with statistical analysis of data. This study helped in developing the ability to build and assess data-based models. It also gave a deep understanding of EDA process, handling and cleaning data. This study also helped in demonstrating skill in data management.