```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest

import warnings
warnings.filterwarnings('ignore')
pd.pandas.set_option('display.max_columns',None)

df=pd.read_csv(r"C:\Users\udayr\cause_of_deaths dataset.csv")
df
```

```
     Country/Territory Code  Year  Meningitis  \
0          Afghanistan  AFG  1990        2159
1          Afghanistan  AFG  1991        2218
2          Afghanistan  AFG  1992        2475
3          Afghanistan  AFG  1993        2812
4          Afghanistan  AFG  1994        3027

...                ...  ...   ...         ...
6115          Zimbabwe  ZWE  2015        1439
6116          Zimbabwe  ZWE  2016        1457
6117          Zimbabwe  ZWE  2017        1460
6118          Zimbabwe  ZWE  2018        1450
6119          Zimbabwe  ZWE  2019        1450


      Alzheimer's Disease and Other Dementias  Parkinson's Disease  \
0                                        1116                  371
1                                        1136                  374
2                                        1162                  378
3                                        1187                  384
4                                        1211                  391

...                                       ...                  ...
6115                                      754                  215
6116                                      767                  219
6117                                      781                  223
6118                                      795                  227
6119                                      812                  232


      Nutritional Deficiencies  Malaria  Drowning  Interpersonal
Violence  \
0                         2087       93      1370
1538
1                         2153      189      1391
2001
2                         2441      239      1514
2299
3                         2837      108      1687
2589
4                         3081      211      1809
2849
```

	...	...	...	
...				
6115	3019	2518	770	1302
6116	3056	2050	801	1342
6117	2990	2116	818	1363
6118	2918	2088	825	1396
6119	2884	2068	827	1434

	Maternal Disorders	HIV/AIDS	Drug Use Disorders	Tuberculosis \
0	2655	34	93	4661
1	2885	41	102	4743
2	3315	48	118	4976
3	3671	56	132	5254
4	3863	63	142	5470
...	...	...	...	...
6115	1355	29162	104	11214
6116	1338	27141	110	10998
6117	1312	24846	115	10762
6118	1294	22106	121	10545
6119	1294	20722	127	10465

	Cardiovascular Diseases	Lower Respiratory Infections \
0	44899	23741
1	45492	24504
2	46557	27404
3	47951	31116
4	49308	33390
...	...	...
6115	16649	12974
6116	16937	13024
6117	17187	12961
6118	17460	12860

```
6119                    17810                        12897

        Neonatal Disorders  Alcohol Use Disorders  Self-harm  \
0                 15612                     72        696
1                 17128                     75        751
2                 20060                     80        855
3                 22335                     85        943
4                 23288                     88        993
...                 ...                    ...        ...
6115               9278                     48       2235
6116               9065                     49       2296
6117               8901                     50       2338
6118               8697                     51       2372
6119               8609                     53       2403

        Exposure to Forces of Nature  Diarrheal Diseases  \
0                                  0                4235
1                               1347                4927
2                                614                6123
3                                225                8174
4                                160                8215
...                              ...                 ...
6115                              16                5102
6116                              31                5002
6117                             251                4948
6118                               0                4745
6119                             660                4635

        Environmental Heat and Cold Exposure  Neoplasms  Conflict and
Terrorism  \
0                                      175      11580
1490
1                                      113      11796
3370
2                                       38      12218
4344
3                                       41      12634
4096
4                                       44      12914
8959
...                                    ...        ...
...
6115                                    37      11161
13
6116                                    37      11465
6
6117                                    37      11744
5
6118                                    37      12038
9
```

```
6119                                                 37        12353
11

        Diabetes Mellitus  Chronic Kidney Disease  Poisonings  \
0                    2108                    3709         338
1                    2120                    3724         351
2                    2153                    3776         386
3                    2195                    3862         425
4                    2231                    3932         451
...                   ...                     ...         ...
6115                 3176                    2108         381
6116                 3259                    2160         393
6117                 3313                    2196         398
6118                 3381                    2240         400
6119                 3460                    2292         405

        Protein-Energy Malnutrition  Road Injuries  \
0                              2054           4154
1                              2119           4472
2                              2404           5106
3                              2797           5681
4                              3038           6001
...                             ...            ...
6115                           2990           2373
6116                           3027           2436
6117                           2962           2473
6118                           2890           2509
6119                           2855           2554

        Chronic Respiratory Diseases  \
0                               5945
1                               6050
2                               6223
3                               6445
4                               6664
...                              ...
6115                            2751
6116                            2788
6117                            2818
6118                            2849
6119                            2891

        Cirrhosis and Other Chronic Liver Diseases  Digestive
Diseases  \
0                                             2673          5005

1                                             2728          5120

2                                             2830          5335
```

```
3                                        2943           5568

4                                        3027           5739

...                                      ...            ...

6115                                     1956           4202

6116                                     1962           4264

6117                                     2007           4342

6118                                     2030           4377

6119                                     2065           4437


      Fire, Heat, and Hot Substances   Acute Hepatitis
0                                 323              2985
1                                 332              3092
2                                 360              3325
3                                 396              3601
4                                 420              3816
...                               ...               ...
6115                              632               146
6116                              648               146
6117                              654               144
6118                              657               139
6119                              662               136

[6120 rows x 34 columns]

df.head()

  Country/Territory Code  Year  Meningitis  \
0        Afghanistan  AFG  1990        2159
1        Afghanistan  AFG  1991        2218
2        Afghanistan  AFG  1992        2475
3        Afghanistan  AFG  1993        2812
4        Afghanistan  AFG  1994        3027

   Alzheimer's Disease and Other Dementias  Parkinson's Disease  \
0                                     1116                  371
1                                     1136                  374
2                                     1162                  378
3                                     1187                  384
4                                     1211                  391
```

	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence
0	2087	93	1370	1538
1	2153	189	1391	2001
2	2441	239	1514	2299
3	2837	108	1687	2589
4	3081	211	1809	2849

	Maternal Disorders	HIV/AIDS	Drug Use Disorders	Tuberculosis \
0	2655	34	93	4661
1	2885	41	102	4743
2	3315	48	118	4976
3	3671	56	132	5254
4	3863	63	142	5470

	Cardiovascular Diseases	Lower Respiratory Infections	Neonatal Disorders \
0	44899	23741	15612
1	45492	24504	17128
2	46557	27404	20060
3	47951	31116	22335
4	49308	33390	23288

	Alcohol Use Disorders	Self-harm	Exposure to Forces of Nature \
0	72	696	0
1	75	751	1347
2	80	855	614
3	85	943	225
4	88	993	160

	Diarrheal Diseases	Environmental Heat and Cold Exposure	Neoplasms \
0	4235	175	11580
1	4927	113	11796
2	6123	38	12218
3	8174	41	12634

			44	12914
4	8215			

	Conflict and Terrorism	Diabetes Mellitus	Chronic Kidney Disease \
0	1490	2108	3709
1	3370	2120	3724
2	4344	2153	3776
3	4096	2195	3862
4	8959	2231	3932

	Poisonings	Protein-Energy Malnutrition	Road Injuries \
0	338	2054	4154
1	351	2119	4472
2	386	2404	5106
3	425	2797	5681
4	451	3038	6001

	Chronic Respiratory Diseases	Cirrhosis and Other Chronic Liver Diseases \
0	5945	2673
1	6050	2728
2	6223	2830
3	6445	2943
4	6664	3027

	Digestive Diseases	Fire, Heat, and Hot Substances	Acute Hepatitis
0	5005	323	2985
1	5120	332	3092
2	5335	360	3325
3	5568	396	3601

| 4 | 5739 | 420 | 3816 |

```python
pd.set_option('display.max_rows',81)

df.dtypes
```

```
Country/Territory                              object
Code                                           object
Year                                            int64
Meningitis                                      int64
Alzheimer's Disease and Other Dementias         int64
Parkinson's Disease                             int64
Nutritional Deficiencies                        int64
Malaria                                         int64
Drowning                                        int64
Interpersonal Violence                          int64
Maternal Disorders                              int64
HIV/AIDS                                        int64
Drug Use Disorders                              int64
Tuberculosis                                    int64
Cardiovascular Diseases                         int64
Lower Respiratory Infections                    int64
Neonatal Disorders                              int64
Alcohol Use Disorders                           int64
Self-harm                                       int64
Exposure to Forces of Nature                    int64
Diarrheal Diseases                              int64
Environmental Heat and Cold Exposure            int64
Neoplasms                                       int64
Conflict and Terrorism                          int64
Diabetes Mellitus                               int64
Chronic Kidney Disease                          int64
Poisonings                                      int64
Protein-Energy Malnutrition                     int64
Road Injuries                                   int64
Chronic Respiratory Diseases                    int64
Cirrhosis and Other Chronic Liver Diseases      int64
Digestive Diseases                              int64
Fire, Heat, and Hot Substances                  int64
Acute Hepatitis                                 int64
dtype: object
```

```python
df['Year']=df.Year.astype('object')

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6120 entries, 0 to 6119
Data columns (total 34 columns):
 #   Column                                      Non-Null Count  Dtype
```

| --- | ------ | ------------- | ----- |
0	Country/Territory	6120 non-null	object
1	Code	6120 non-null	object
2	Year	6120 non-null	object
3	Meningitis	6120 non-null	int64
4	Alzheimer's Disease and Other Dementias	6120 non-null	int64
5	Parkinson's Disease	6120 non-null	int64
6	Nutritional Deficiencies	6120 non-null	int64
7	Malaria	6120 non-null	int64
8	Drowning	6120 non-null	int64
9	Interpersonal Violence	6120 non-null	int64
10	Maternal Disorders	6120 non-null	int64
11	HIV/AIDS	6120 non-null	int64
12	Drug Use Disorders	6120 non-null	int64
13	Tuberculosis	6120 non-null	int64
14	Cardiovascular Diseases	6120 non-null	int64
15	Lower Respiratory Infections	6120 non-null	int64
16	Neonatal Disorders	6120 non-null	int64
17	Alcohol Use Disorders	6120 non-null	int64
18	Self-harm	6120 non-null	int64
19	Exposure to Forces of Nature	6120 non-null	int64
20	Diarrheal Diseases	6120 non-null	int64
21	Environmental Heat and Cold Exposure	6120 non-null	int64
22	Neoplasms	6120 non-null	int64

23	Conflict and Terrorism	6120 non-null	int64
24	Diabetes Mellitus	6120 non-null	int64
25	Chronic Kidney Disease	6120 non-null	int64
26	Poisonings	6120 non-null	int64
27	Protein-Energy Malnutrition	6120 non-null	int64
28	Road Injuries	6120 non-null	int64
29	Chronic Respiratory Diseases	6120 non-null	int64
30	Cirrhosis and Other Chronic Liver Diseases	6120 non-null	int64
31	Digestive Diseases	6120 non-null	int64
32	Fire, Heat, and Hot Substances	6120 non-null	int64
33	Acute Hepatitis	6120 non-null	int64

dtypes: int64(31), object(3)
memory usage: 1.6+ MB

df.isnull().sum()

```
Country/Territory                            0
Code                                         0
Year                                         0
Meningitis                                   0
Alzheimer's Disease and Other Dementias      0
Parkinson's Disease                          0
Nutritional Deficiencies                     0
Malaria                                      0
Drowning                                     0
Interpersonal Violence                       0
Maternal Disorders                           0
HIV/AIDS                                     0
Drug Use Disorders                           0
Tuberculosis                                 0
Cardiovascular Diseases                      0
Lower Respiratory Infections                 0
Neonatal Disorders                           0
Alcohol Use Disorders                        0
Self-harm                                    0
Exposure to Forces of Nature                 0
Diarrheal Diseases                           0
```

```
Environmental Heat and Cold Exposure            0
Neoplasms                                       0
Conflict and Terrorism                          0
Diabetes Mellitus                               0
Chronic Kidney Disease                          0
Poisonings                                      0
Protein-Energy Malnutrition                     0
Road Injuries                                   0
Chronic Respiratory Diseases                    0
Cirrhosis and Other Chronic Liver Diseases      0
Digestive Diseases                              0
Fire, Heat, and Hot Substances                  0
Acute Hepatitis                                 0
dtype: int64
```

```python
# Check Duplicated records
df.duplicated().sum()
```

```
0
```

```python
# split  data to continous and categorical varaiables.
df1_cont=df.select_dtypes(include=['float64','int64'])
df1_cat=df.select_dtypes(include=['object'])
```

```python
# Create Data audit Report for continuous variables
def continuous_var_summary(x):
    return pd.Series([x.count(), x.isnull().sum(), x.sum(), x.mean(),
x.median(),
                      x.std(), x.var(),
x.min(),x.quantile(0.25),x.quantile(0.50),x.quantile(0.75),x.max()],
                  index = ['N', 'NMISS', 'SUM', 'MEAN','MEDIAN',
'STD', 'VAR', 'MIN','P25' ,'P50' ,'P75','MAX'])
```

```python
# Descriptive analysis for continous data
df1_cont.apply(continuous_var_summary)
```

```
        Meningitis  Alzheimer's Disease and Other Dementias  \
N       6.120000e+03                             6.120000e+03
NMISS   0.000000e+00                             0.000000e+00
SUM     1.052457e+07                             2.976884e+07
MEAN    1.719701e+03                             4.864189e+03
MEDIAN  1.090000e+02                             6.665000e+02
STD     6.672007e+03                             1.822066e+04
VAR     4.451568e+07                             3.319924e+08
MIN     0.000000e+00                             0.000000e+00
P25     1.500000e+01                             9.000000e+01
P50     1.090000e+02                             6.665000e+02
P75     8.472500e+02                             2.456250e+03
MAX     9.835800e+04                             3.207150e+05

        Parkinson's Disease  Nutritional Deficiencies        Malaria  \
```

N	6.120000e+03	6.120000e+03	6.120000e+03
NMISS	0.000000e+00	0.000000e+00	0.000000e+00
SUM	7.179795e+06	1.379203e+07	2.534268e+07
MEAN	1.173169e+03	2.253600e+03	4.140960e+03
MEDIAN	1.640000e+02	1.190000e+02	0.000000e+00
STD	4.616156e+03	1.048363e+04	1.842775e+04
VAR	2.130890e+07	1.099066e+08	3.395821e+08
MIN	0.000000e+00	0.000000e+00	0.000000e+00
P25	2.700000e+01	9.000000e+00	0.000000e+00
P50	1.640000e+02	1.190000e+02	0.000000e+00
P75	6.092500e+02	1.167250e+03	3.930000e+02
MAX	7.699000e+04	2.682230e+05	2.806040e+05

	Drowning	Interpersonal Violence	Maternal Disorders	\
N	6.120000e+03	6.120000e+03	6.120000e+03	
NMISS	0.000000e+00	0.000000e+00	0.000000e+00	
SUM	1.030200e+07	1.275284e+07	7.727046e+06	
MEAN	1.683333e+03	2.083797e+03	1.262589e+03	
MEDIAN	1.770000e+02	2.650000e+02	5.400000e+01	
STD	8.877018e+03	6.917006e+03	6.057973e+03	
VAR	7.880146e+07	4.784497e+07	3.669904e+07	
MIN	0.000000e+00	0.000000e+00	0.000000e+00	
P25	3.400000e+01	4.000000e+01	5.000000e+00	
P50	1.770000e+02	2.650000e+02	5.400000e+01	
P75	6.980000e+02	8.770000e+02	7.340000e+02	
MAX	1.537730e+05	6.964000e+04	1.079290e+05	

	HIV/AIDS	Drug Use Disorders	Tuberculosis	\
N	6.120000e+03	6.120000e+03	6.120000e+03	
NMISS	0.000000e+00	0.000000e+00	0.000000e+00	
SUM	3.636442e+07	2.656121e+06	4.585060e+07	
MEAN	5.941899e+03	4.340067e+02	7.491929e+03	
MEDIAN	1.360000e+02	2.000000e+01	4.170000e+02	
STD	2.101196e+04	2.898762e+03	3.954998e+04	
VAR	4.415026e+08	8.402819e+06	1.564201e+09	
MIN	0.000000e+00	0.000000e+00	0.000000e+00	
P25	1.100000e+01	3.000000e+00	3.500000e+01	
P50	1.360000e+02	2.000000e+01	4.170000e+02	
P75	1.879000e+03	1.290000e+02	2.924250e+03	
MAX	3.054910e+05	6.571700e+04	6.575150e+05	

	Cardiovascular Diseases	Lower Respiratory Infections	\
N	6.120000e+03	6.120000e+03	
NMISS	0.000000e+00	0.000000e+00	
SUM	4.477420e+08	8.377004e+07	
MEAN	7.316045e+04	1.368791e+04	
MEDIAN	1.174200e+04	2.126500e+03	
STD	2.915775e+05	4.803172e+04	
VAR	8.501746e+10	2.307046e+09	
MIN	4.000000e+00	0.000000e+00	

```
P25                    2.028000e+03                    3.450000e+02
P50                    1.174200e+04                    2.126500e+03
P75                    4.254650e+04                    1.016125e+04
MAX                    4.584273e+06                    6.909130e+05


          Neonatal Disorders  Alcohol Use Disorders       Self-harm  \
N                6.120000e+03           6.120000e+03  6.120000e+03
NMISS            0.000000e+00           0.000000e+00  0.000000e+00
SUM              7.686073e+07           4.819018e+06  2.371393e+07
MEAN             1.255894e+04           7.874212e+02  3.874825e+03
MEDIAN           9.160000e+02           8.000000e+01  5.330000e+02
STD              5.605837e+04           3.545824e+03  1.842562e+04
VAR              3.142540e+09           1.257287e+07  3.395033e+08
MIN              0.000000e+00           0.000000e+00  0.000000e+00
P25              1.310000e+02           9.000000e+00  9.400000e+01
P50              9.160000e+02           8.000000e+01  5.330000e+02
P75              7.419750e+03           3.160000e+02  1.882250e+03
MAX              8.527610e+05           5.520000e+04  2.203570e+05


          Exposure to Forces of Nature  Diarrheal Diseases  \
N                         6.120000e+03        6.120000e+03
NMISS                     0.000000e+00        0.000000e+00
SUM                       1.490132e+06        6.623551e+07
MEAN                      2.434856e+02        1.082280e+04
MEDIAN                    0.000000e+00        2.965000e+02
STD                       4.717104e+03        6.541617e+04
VAR                       2.225107e+07        4.279276e+09
MIN                       0.000000e+00        0.000000e+00
P25                       0.000000e+00        2.000000e+01
P50                       0.000000e+00        2.965000e+02
P75                       1.200000e+01        3.946750e+03
MAX                       2.226410e+05        1.119477e+06


          Environmental Heat and Cold Exposure       Neoplasms  \
N                                 6.120000e+03  6.120000e+03
NMISS                             0.000000e+00  0.000000e+00
SUM                               1.788851e+06  2.297585e+08
MEAN                              2.922959e+02  3.754224e+04
MEDIAN                            2.100000e+01  5.629500e+03
STD                               1.704466e+03  1.615584e+05
VAR                               2.905206e+06  2.610111e+10
MIN                               0.000000e+00  1.000000e+00
P25                               2.000000e+00  8.097500e+02
P50                               2.100000e+01  5.629500e+03
P75                               1.090000e+02  2.014775e+04
MAX                               2.904800e+04  2.716551e+06


          Conflict and Terrorism  Diabetes Mellitus  Chronic Kidney
Disease  \
N                    6.120000e+03       6.120000e+03
```

```
                                                    6.120000e+03
NMISS            0.000000e+00    0.000000e+00
0.000000e+00
SUM              3.294053e+06    3.144887e+07
2.891169e+07
MEAN             5.382440e+02    5.138705e+03
4.724133e+03
MEDIAN           0.000000e+00    1.087000e+03
8.220000e+02
STD              7.033308e+03    1.677308e+04
1.647043e+04
VAR              4.946742e+07    2.813362e+08
2.712751e+08
MIN              0.000000e+00    1.000000e+00
0.000000e+00
P25              0.000000e+00    2.360000e+02
1.457500e+02
P50              0.000000e+00    1.087000e+03
8.220000e+02
P75              2.300000e+01    2.954000e+03
2.922500e+03
MAX              5.035320e+05    2.730890e+05
2.229220e+05
```

	Poisonings	Protein-Energy Malnutrition	Road Injuries \
N	6.120000e+03	6.120000e+03	6.120000e+03
NMISS	0.000000e+00	0.000000e+00	0.000000e+00
SUM	2.601082e+06	1.203188e+07	3.629647e+07
MEAN	4.250134e+02	1.965994e+03	5.930796e+03
MEDIAN	5.250000e+01	9.200000e+01	9.665000e+02
STD	2.022641e+03	8.255999e+03	2.409778e+04
VAR	4.091075e+06	6.816152e+07	5.807032e+08
MIN	0.000000e+00	0.000000e+00	0.000000e+00
P25	6.000000e+00	5.000000e+00	1.747500e+02
P50	5.250000e+01	9.200000e+01	9.665000e+02
P75	2.540000e+02	1.042500e+03	3.435250e+03
MAX	3.088300e+04	2.022410e+05	3.292370e+05

	Chronic Respiratory Diseases \
N	6.120000e+03
NMISS	0.000000e+00
SUM	1.046053e+08
MEAN	1.709237e+04
MEDIAN	1.689000e+03
STD	1.051572e+05
VAR	1.105803e+10
MIN	1.000000e+00
P25	2.890000e+02
P50	1.689000e+03
P75	5.249750e+03

```
MAX                              1.366039e+06

            Cirrhosis and Other Chronic Liver Diseases  Digestive Diseases
\
N                                          6.120000e+03            6.120000e+03

NMISS                                      0.000000e+00            0.000000e+00

SUM                                        3.747932e+07            6.563864e+07

MEAN                                       6.124072e+03            1.072527e+04

MEDIAN                                     1.210000e+03            2.185000e+03

STD                                        2.068812e+04            3.722805e+04

VAR                                        4.279983e+08            1.385928e+09

MIN                                        0.000000e+00            0.000000e+00

P25                                        1.540000e+02            2.840000e+02

P50                                        1.210000e+03            2.185000e+03

P75                                        3.547250e+03            6.080000e+03

MAX                                        2.700370e+05            4.649140e+05


            Fire, Heat, and Hot Substances   Acute Hepatitis
N                            6.120000e+03      6.120000e+03
NMISS                        0.000000e+00      0.000000e+00
SUM                          3.602914e+06      3.784791e+06
MEAN                         5.887114e+02      6.184299e+02
MEDIAN                       1.260000e+02      1.500000e+01
STD                          2.128595e+03      4.186023e+03
VAR                          4.530917e+06      1.752279e+07
MIN                          0.000000e+00      0.000000e+00
P25                          1.700000e+01      2.000000e+00
P50                          1.260000e+02      1.500000e+01
P75                          4.500000e+02      1.600000e+02
MAX                          2.587600e+04      6.430500e+04

df1_cat.describe(include=['O'])

        Country/Territory  Code  Year
count                6120  6120  6120
unique                204   204    30
```

```
top            Afghanistan   AFG   1990
freq                    30    30   204
```

```
df["Parkinson's Disease"].max()
```

76990

```
df["Parkinson's Disease"].unique().min()
```

0

```
df['Malaria'].unique().max()
```

280604

```
df['Country/Territory'].unique()
```

```
array(['Afghanistan', 'Albania', 'Algeria', 'American Samoa',
'Andorra',
       'Angola', 'Antigua and Barbuda', 'Argentina', 'Armenia',
       'Australia', 'Austria', 'Azerbaijan', 'Bahamas', 'Bahrain',
       'Bangladesh', 'Barbados', 'Belarus', 'Belgium', 'Belize',
'Benin',
       'Bermuda', 'Bhutan', 'Bolivia', 'Bosnia and Herzegovina',
       'Botswana', 'Brazil', 'Brunei', 'Bulgaria', 'Burkina Faso',
       'Burundi', 'Cambodia', 'Cameroon', 'Canada', 'Cape Verde',
       'Central African Republic', 'Chad', 'Chile', 'China',
'Colombia',
       'Comoros', 'Congo', 'Cook Islands', 'Costa Rica', "Cote
d'Ivoire",
       'Croatia', 'Cuba', 'Cyprus', 'Czechia',
       'Democratic Republic of Congo', 'Denmark', 'Djibouti',
'Dominica',
       'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador',
       'Equatorial Guinea', 'Eritrea', 'Estonia', 'Eswatini',
'Ethiopia',
       'Fiji', 'Finland', 'France', 'Gabon', 'Gambia', 'Georgia',
       'Germany', 'Ghana', 'Greece', 'Greenland', 'Grenada', 'Guam',
       'Guatemala', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti',
       'Honduras', 'Hungary', 'Iceland', 'India', 'Indonesia', 'Iran',
       'Iraq', 'Ireland', 'Israel', 'Italy', 'Jamaica', 'Japan',
'Jordan',
       'Kazakhstan', 'Kenya', 'Kiribati', 'Kuwait', 'Kyrgyzstan',
'Laos',
       'Latvia', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
'Lithuania',
       'Luxembourg', 'Madagascar', 'Malawi', 'Malaysia', 'Maldives',
       'Mali', 'Malta', 'Marshall Islands', 'Mauritania', 'Mauritius',
       'Mexico', 'Micronesia', 'Moldova', 'Monaco', 'Mongolia',
       'Montenegro', 'Morocco', 'Mozambique', 'Myanmar', 'Namibia',
       'Nauru', 'Nepal', 'Netherlands', 'New Zealand', 'Nicaragua',
       'Niger', 'Nigeria', 'Niue', 'North Korea', 'North Macedonia',
```

```
       'Northern Mariana Islands', 'Norway', 'Oman', 'Pakistan',
'Palau',
       'Palestine', 'Panama', 'Papua New Guinea', 'Paraguay', 'Peru',
       'Philippines', 'Poland', 'Portugal', 'Puerto Rico', 'Qatar',
       'Romania', 'Russia', 'Rwanda', 'Saint Kitts and Nevis',
       'Saint Lucia', 'Saint Vincent and the Grenadines', 'Samoa',
       'San Marino', 'Sao Tome and Principe', 'Saudi Arabia',
'Senegal',
       'Serbia', 'Seychelles', 'Sierra Leone', 'Singapore',
'Slovakia',
       'Slovenia', 'Solomon Islands', 'Somalia', 'South Africa',
       'South Korea', 'South Sudan', 'Spain', 'Sri Lanka', 'Sudan',
       'Suriname', 'Sweden', 'Switzerland', 'Syria', 'Taiwan',
       'Tajikistan', 'Tanzania', 'Thailand', 'Timor', 'Togo',
'Tokelau',
       'Tonga', 'Trinidad and Tobago', 'Tunisia', 'Turkey',
       'Turkmenistan', 'Tuvalu', 'Uganda', 'Ukraine',
       'United Arab Emirates', 'United Kingdom', 'United States',
       'United States Virgin Islands', 'Uruguay', 'Uzbekistan',
'Vanuatu',
       'Venezuela', 'Vietnam', 'Yemen', 'Zambia', 'Zimbabwe'],
      dtype=object)
```

```python
df['Country/Territory'].value_counts()
```

```
Afghanistan          30
Papua New Guinea     30
Niue                 30
North Korea          30
North Macedonia      30
                     ..
Greenland            30
Grenada              30
Guam                 30
Guatemala            30
Zimbabwe             30
Name: Country/Territory, Length: 204, dtype: int64
```

```python
len(df['Country/Territory'].unique())
```

```
204
```

```python
# Check Normality of continous data
for i in df1_cont.columns:
    sns.distplot(df1_cont[i])
    print(kstest(df1_cont[i].values,'norm'))
    plt.show()
```

```
KstestResult(statistic=0.8829688223001868, pvalue=0.0)
```
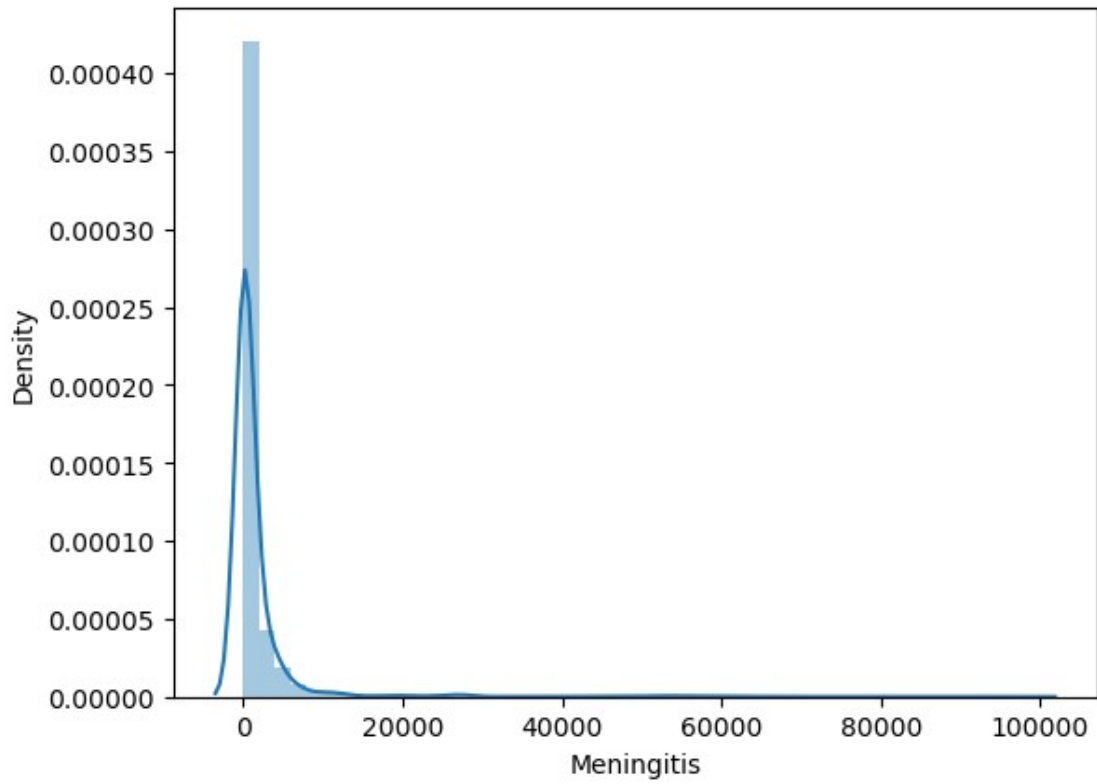
KstestResult(statistic=0.9661337620990889, pvalue=0.0)
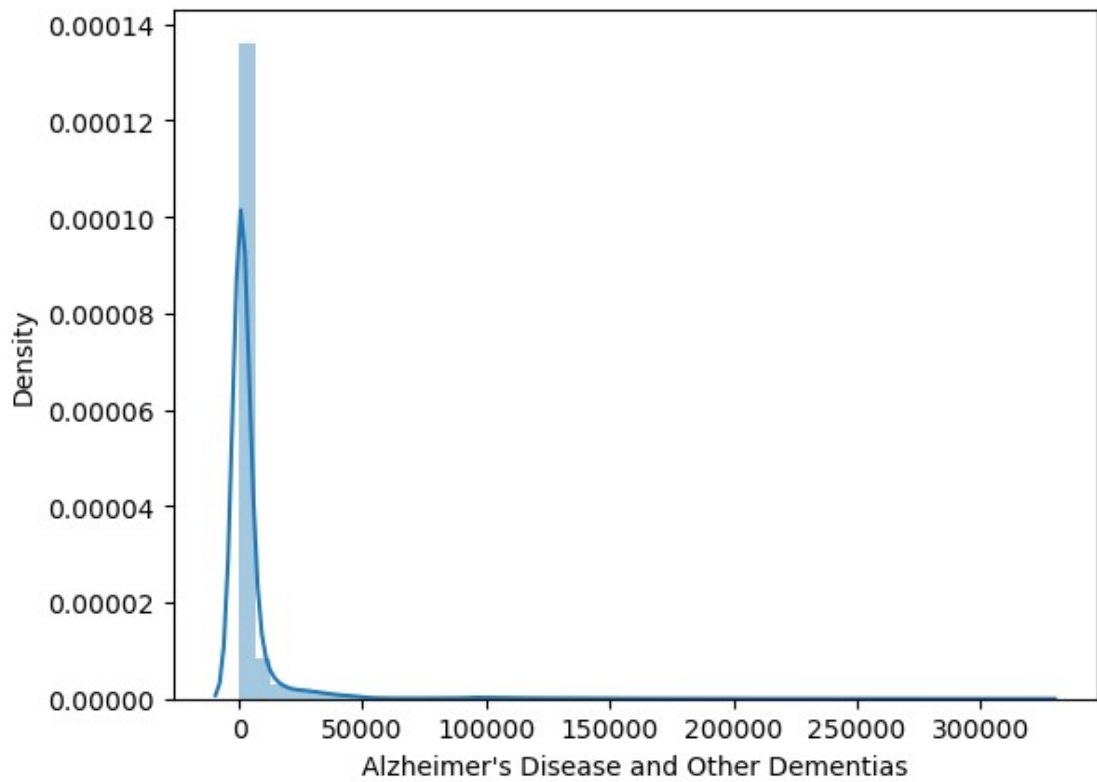
KstestResult(statistic=0.9471795137330757, pvalue=0.0)
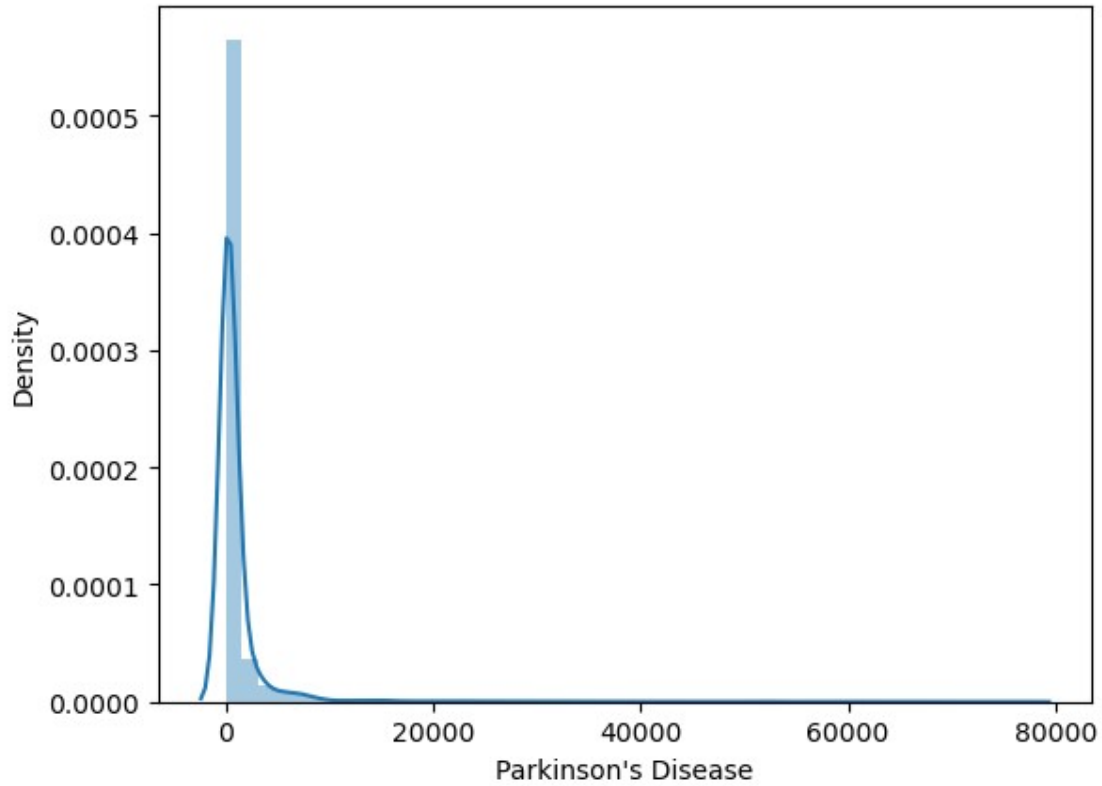
KstestResult(statistic=0.8811714366792718, pvalue=0.0)
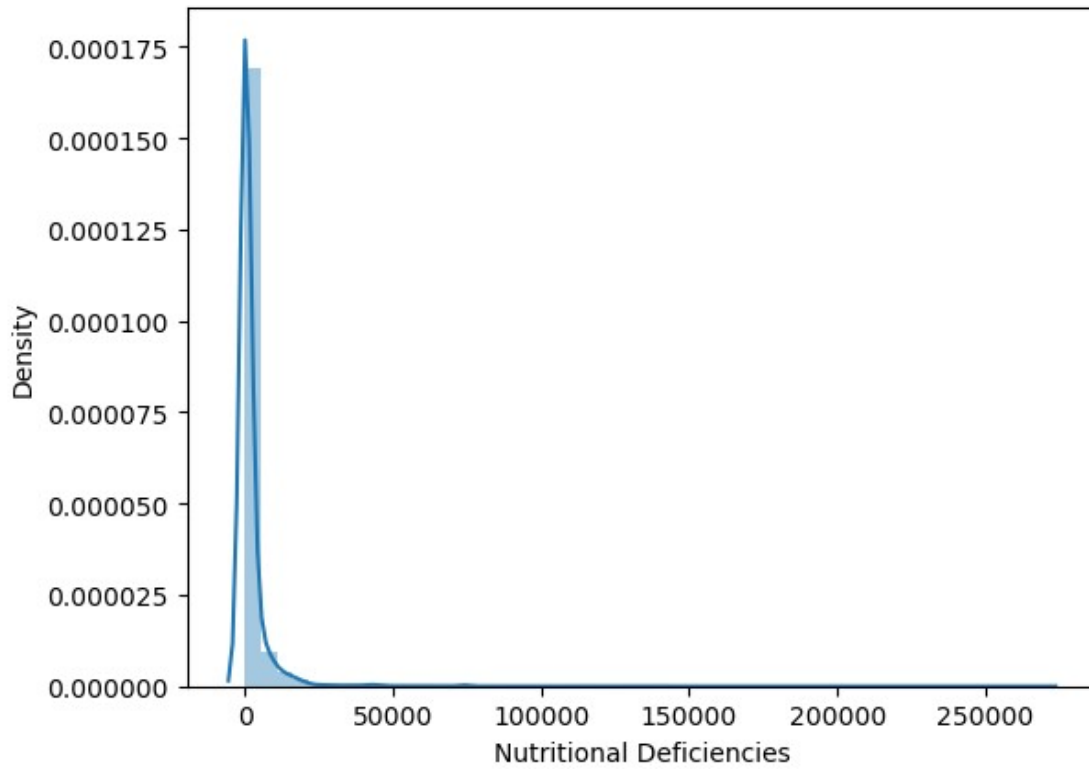
KstestResult(statistic=0.5, pvalue=0.0)
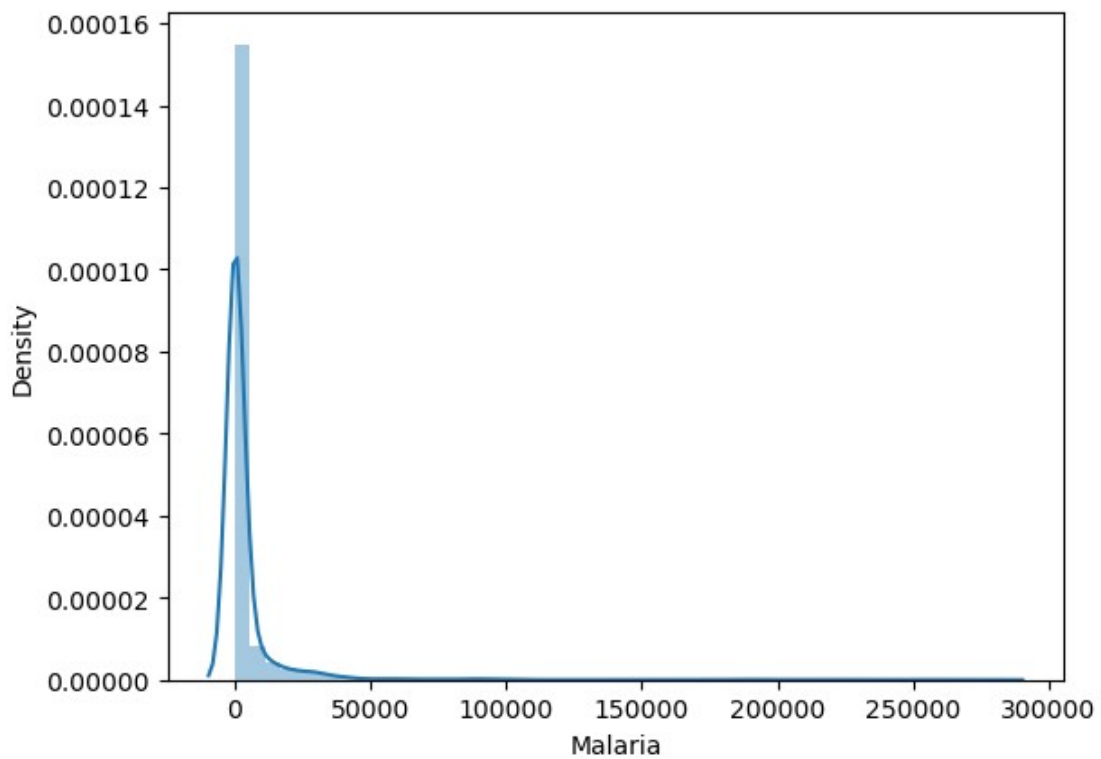
KstestResult(statistic=0.9483233045827489, pvalue=0.0)
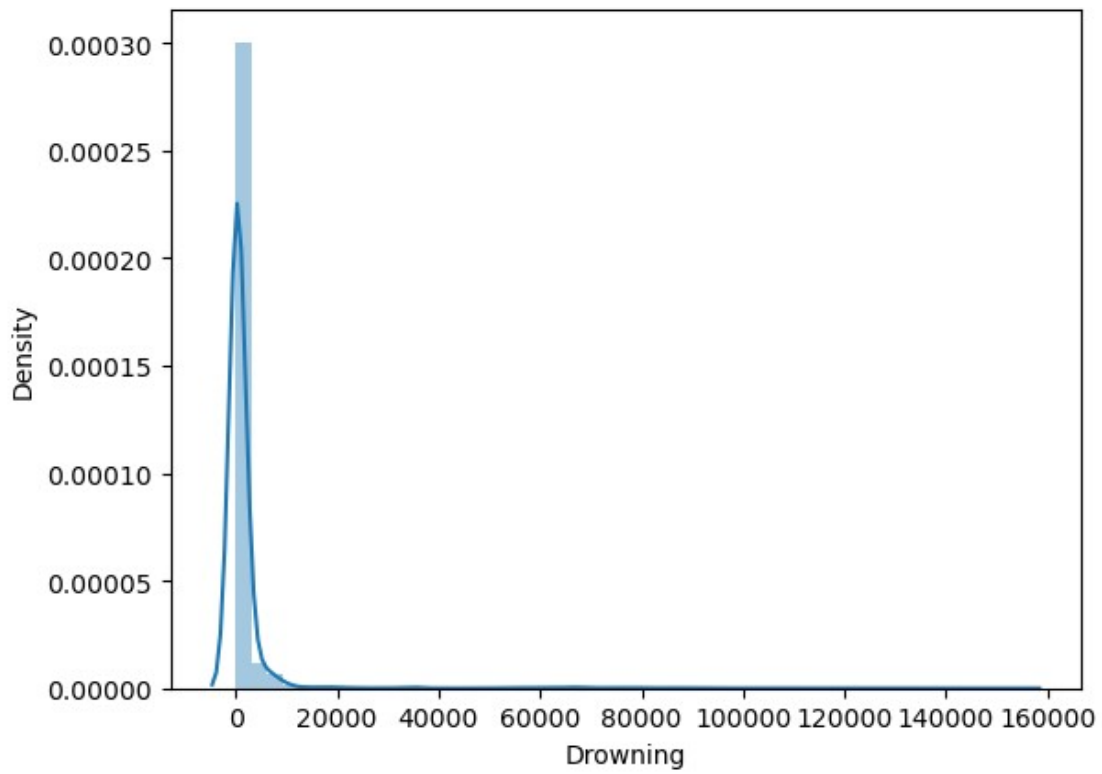
KstestResult(statistic=0.9432579451056248, pvalue=0.0)
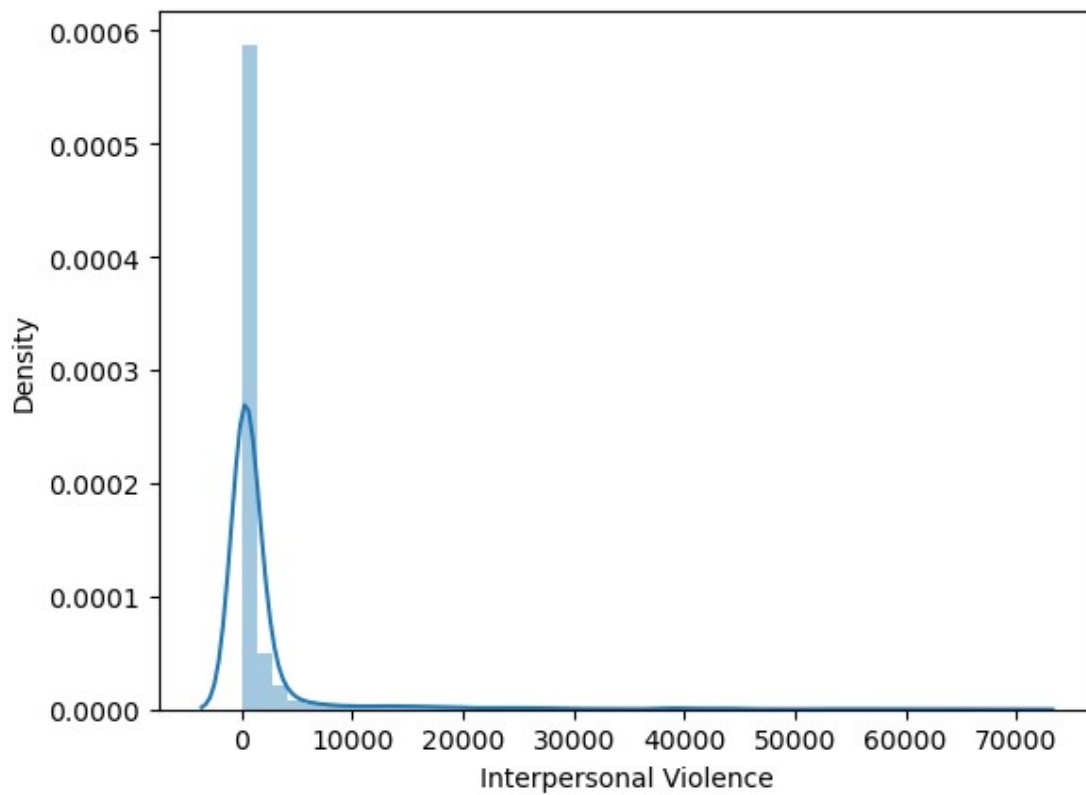
KstestResult(statistic=0.8373805870060691, pvalue=0.0)
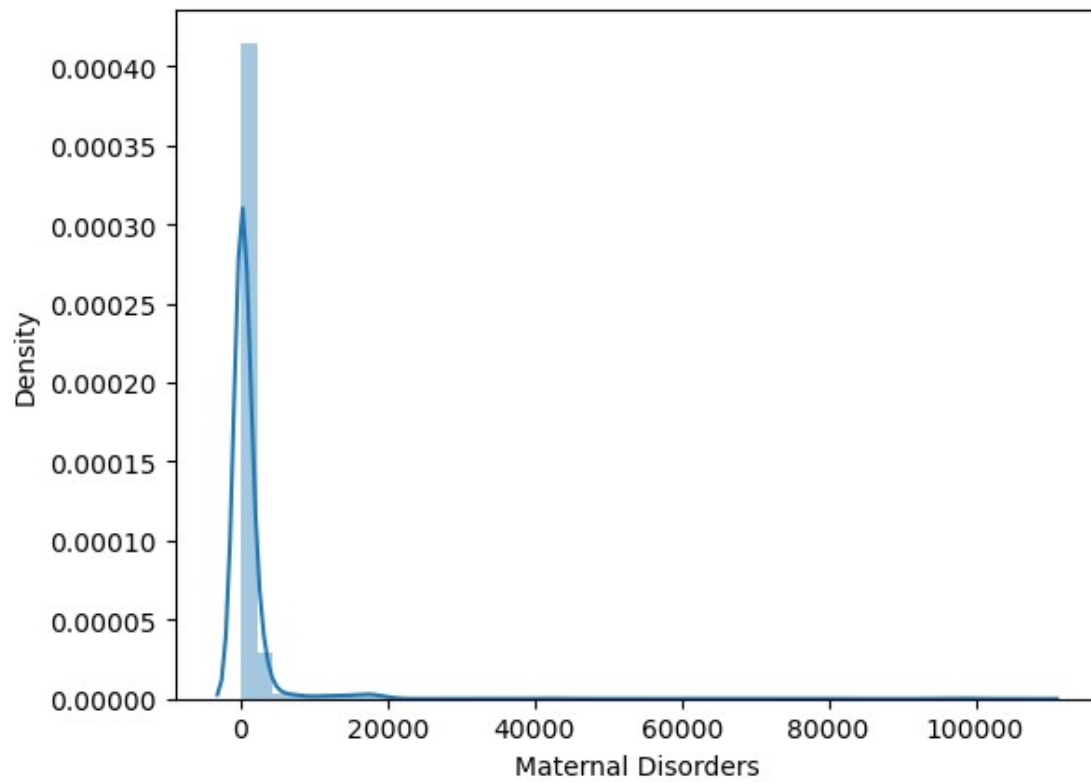
KstestResult(statistic=0.8888511752413633, pvalue=0.0)
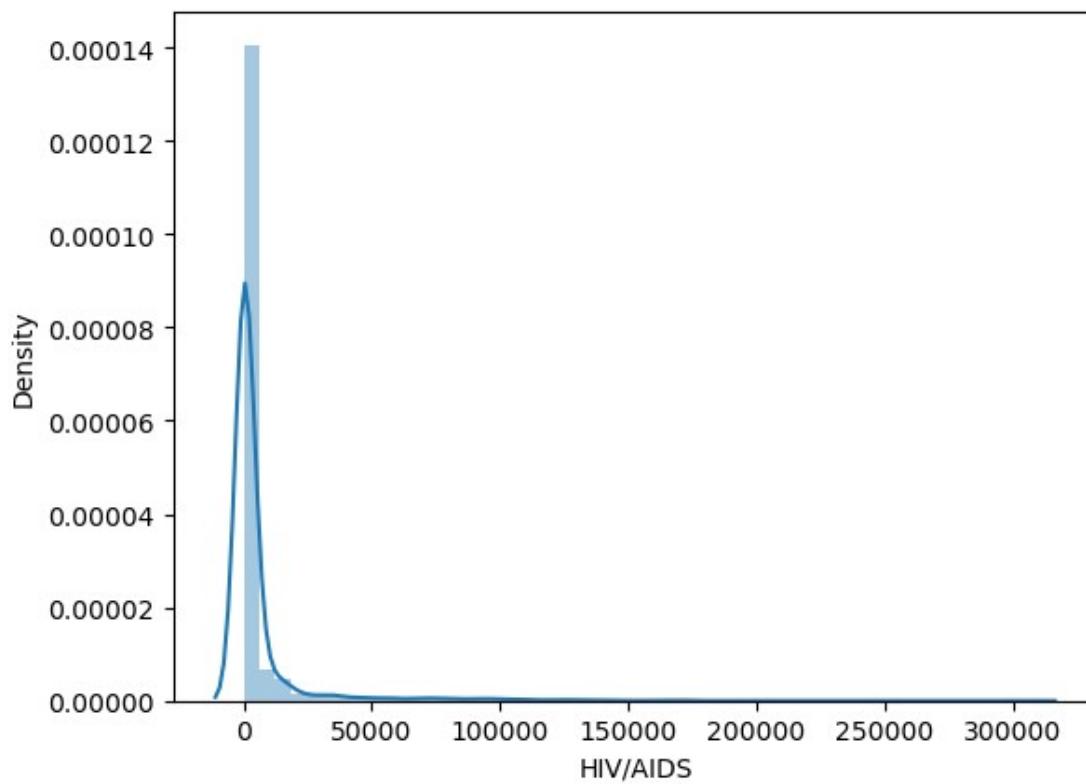
KstestResult(statistic=0.7891779726269842, pvalue=0.0)
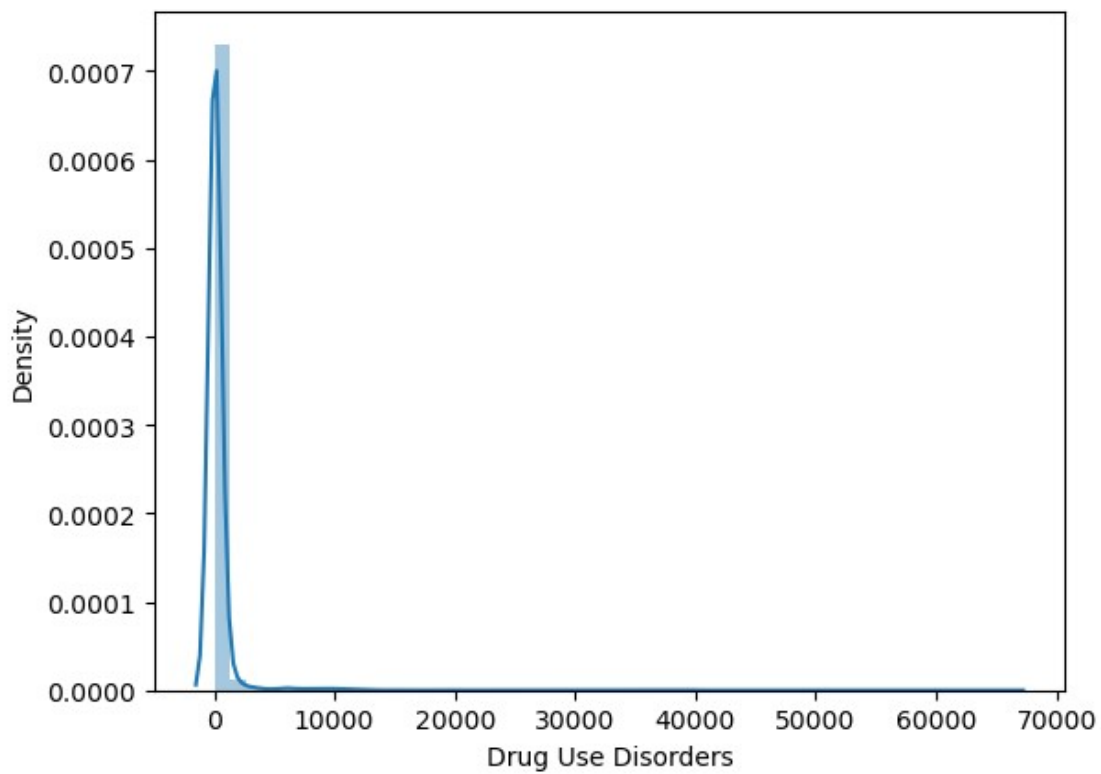
KstestResult(statistic=0.9063298405304614, pvalue=0.0)
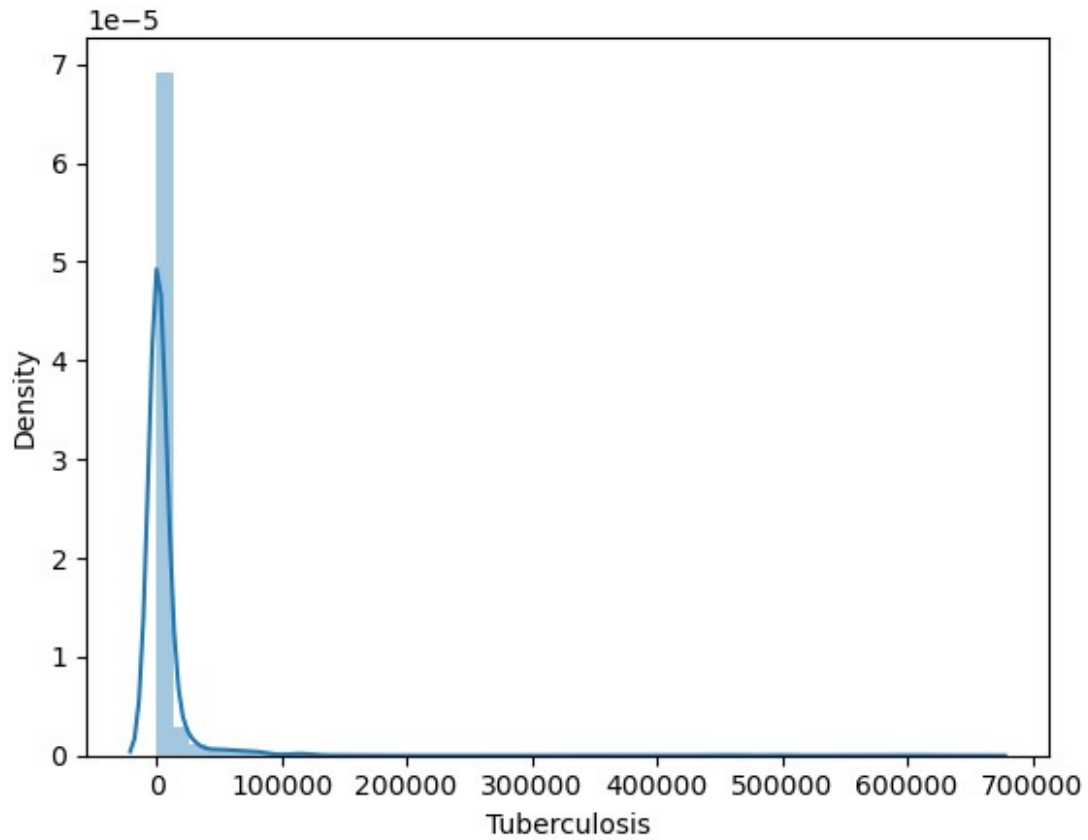
KstestResult(statistic=0.9999683287581669, pvalue=0.0)
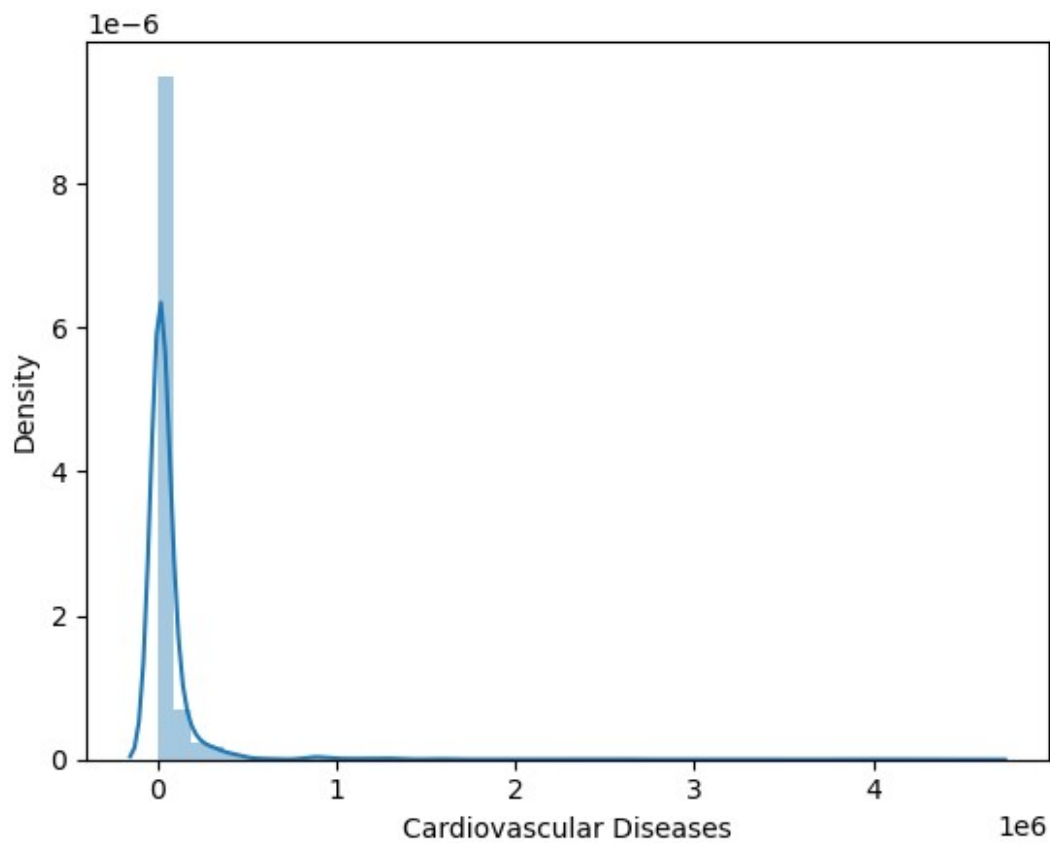
KstestResult(statistic=0.9901644071895395, pvalue=0.0)
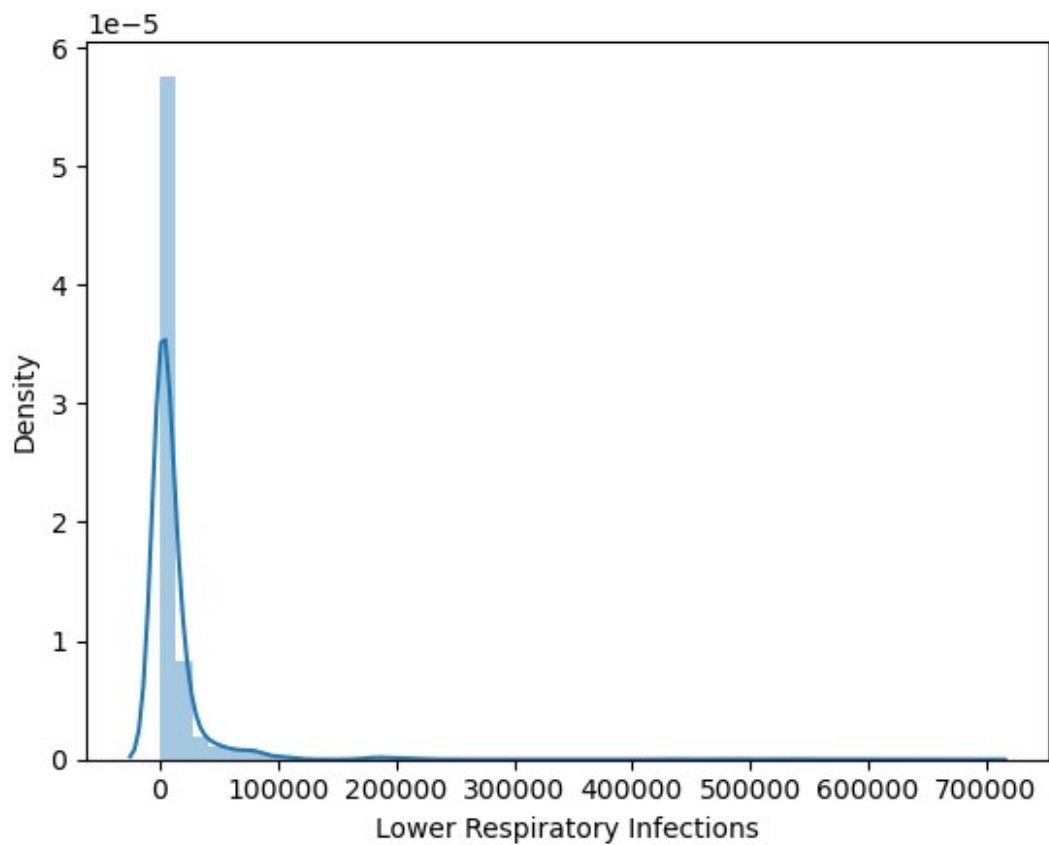
KstestResult(statistic=0.9586174222298078, pvalue=0.0)
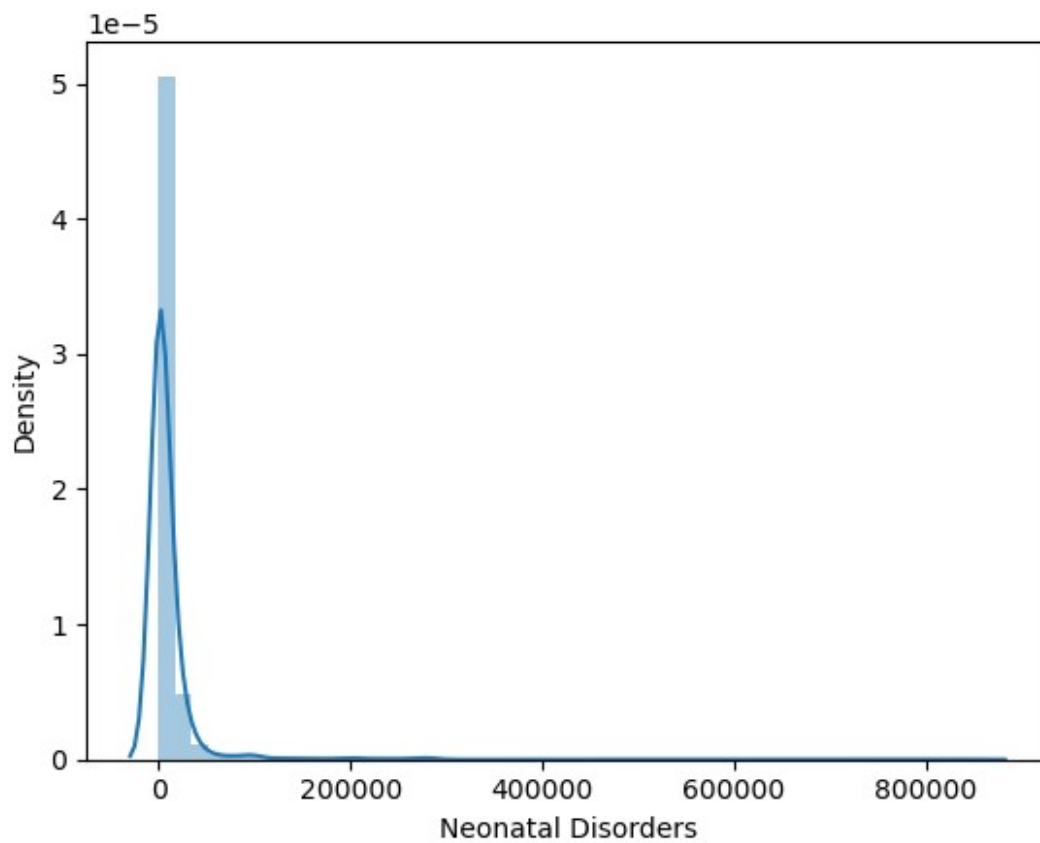
KstestResult(statistic=0.8803544432152195, pvalue=0.0)
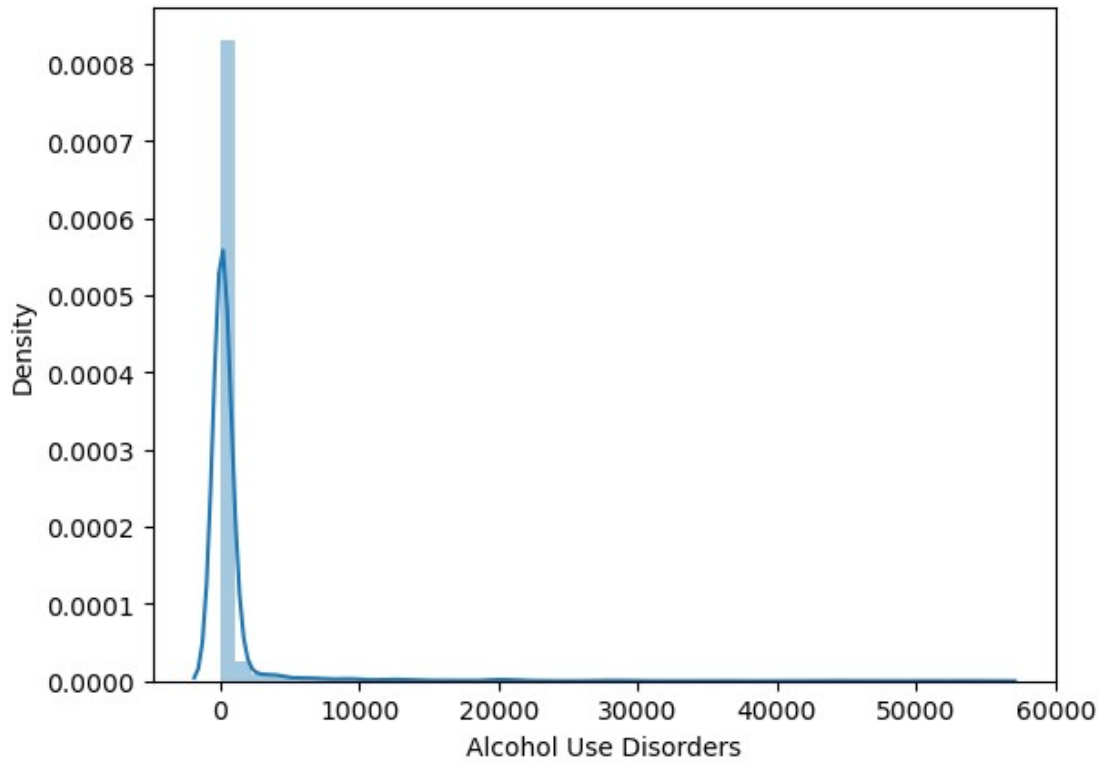
KstestResult(statistic=0.967604350334383, pvalue=0.0)
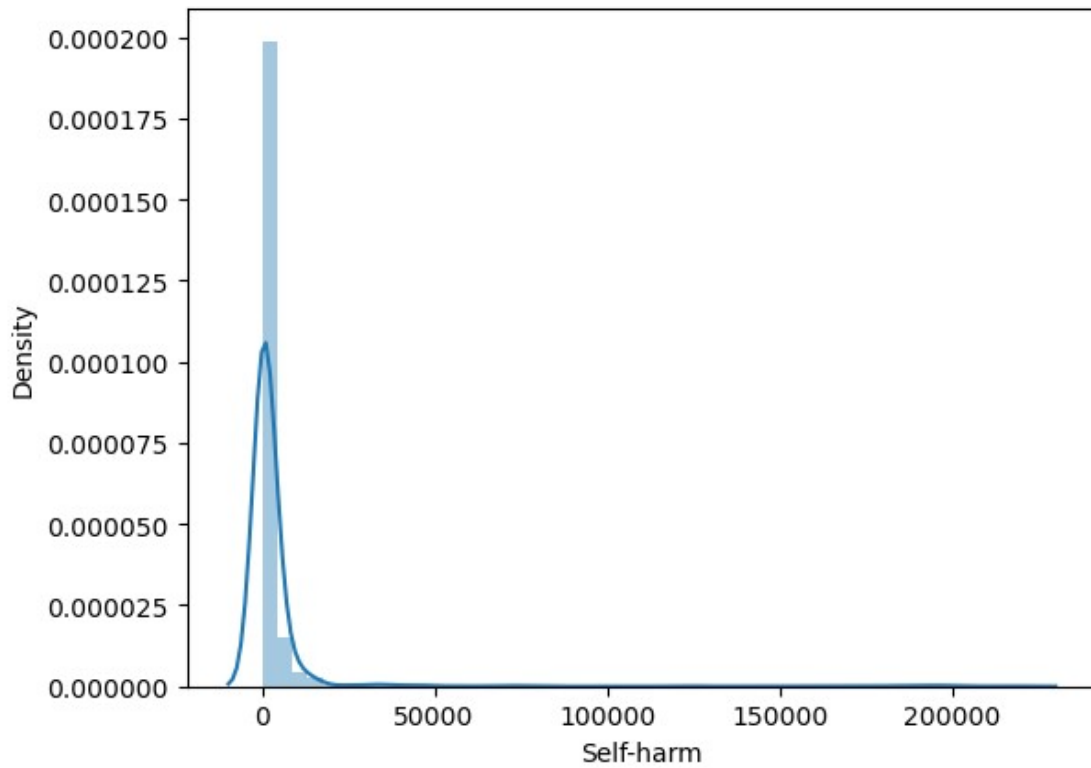
KstestResult(statistic=0.5, pvalue=0.0)
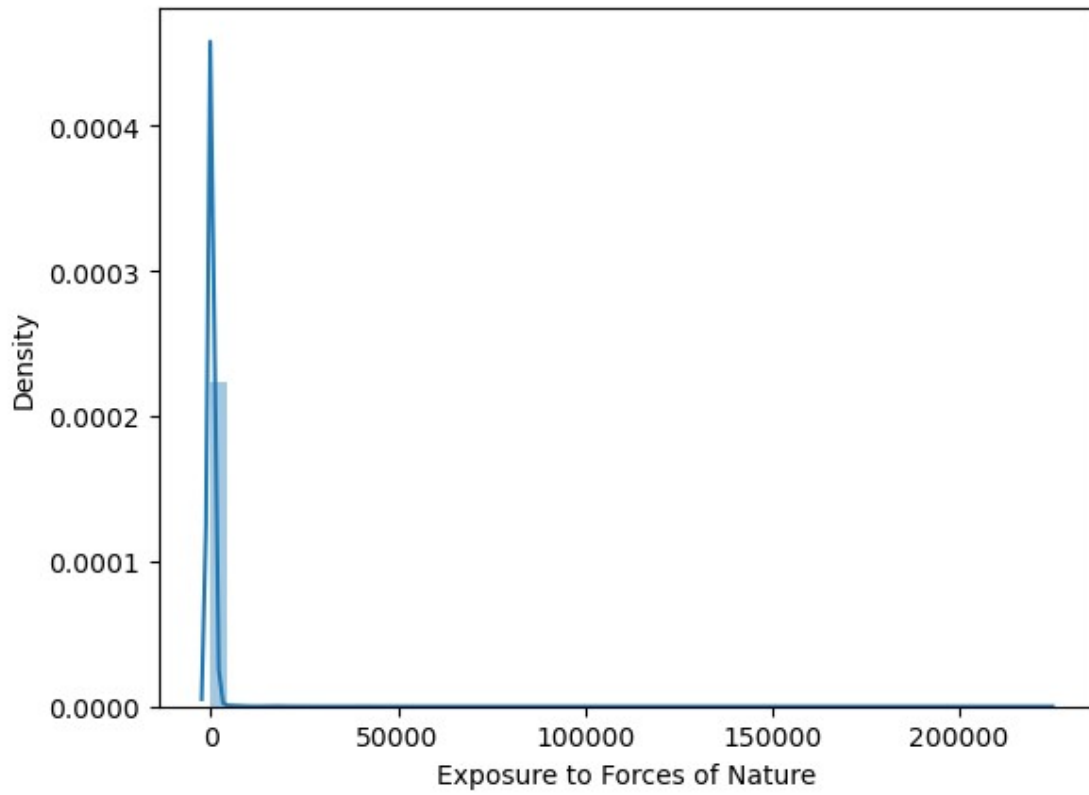
KstestResult(statistic=0.9086224170714287, pvalue=0.0)
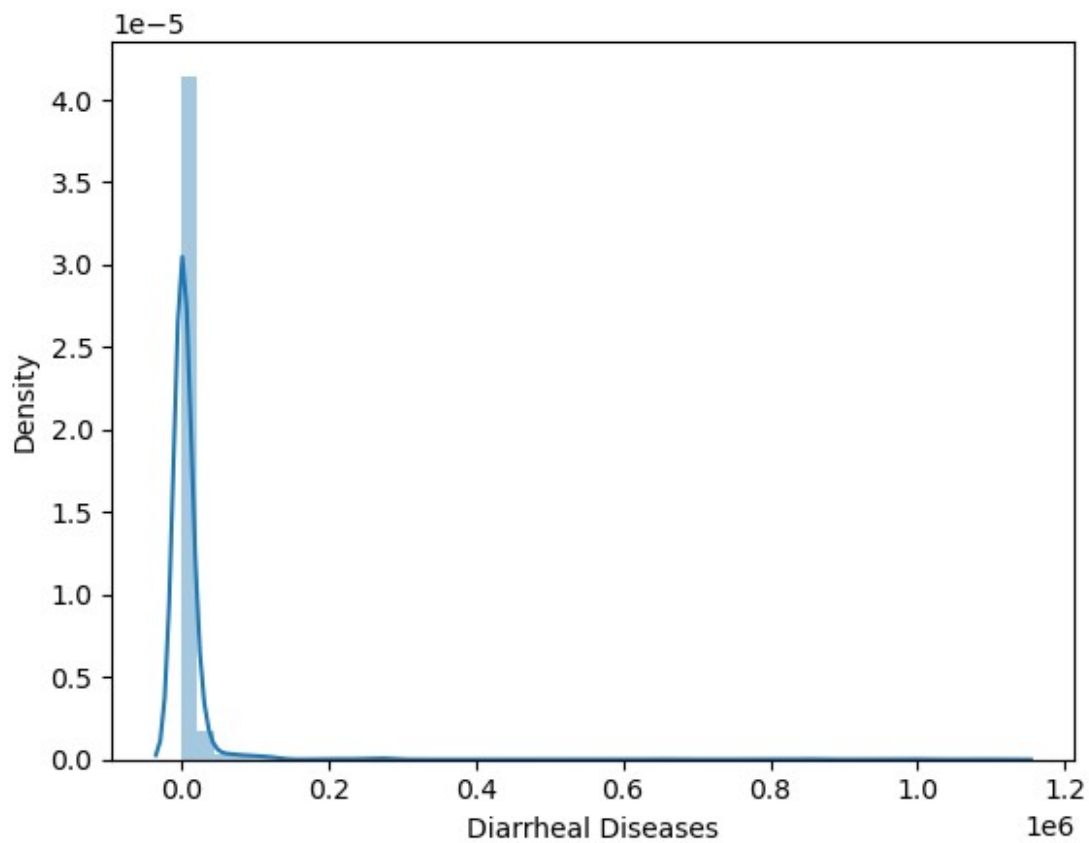
KstestResult(statistic=0.7628707830845005, pvalue=0.0)
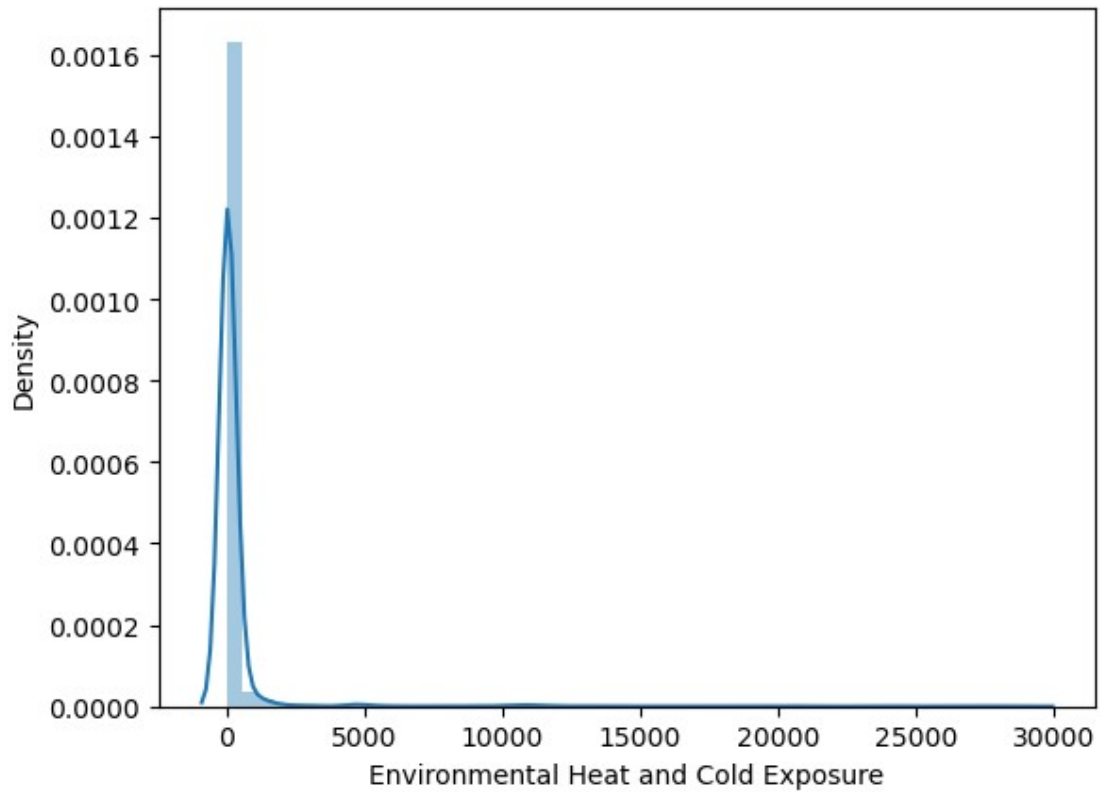
KstestResult(statistic=0.9937481411840562, pvalue=0.0)
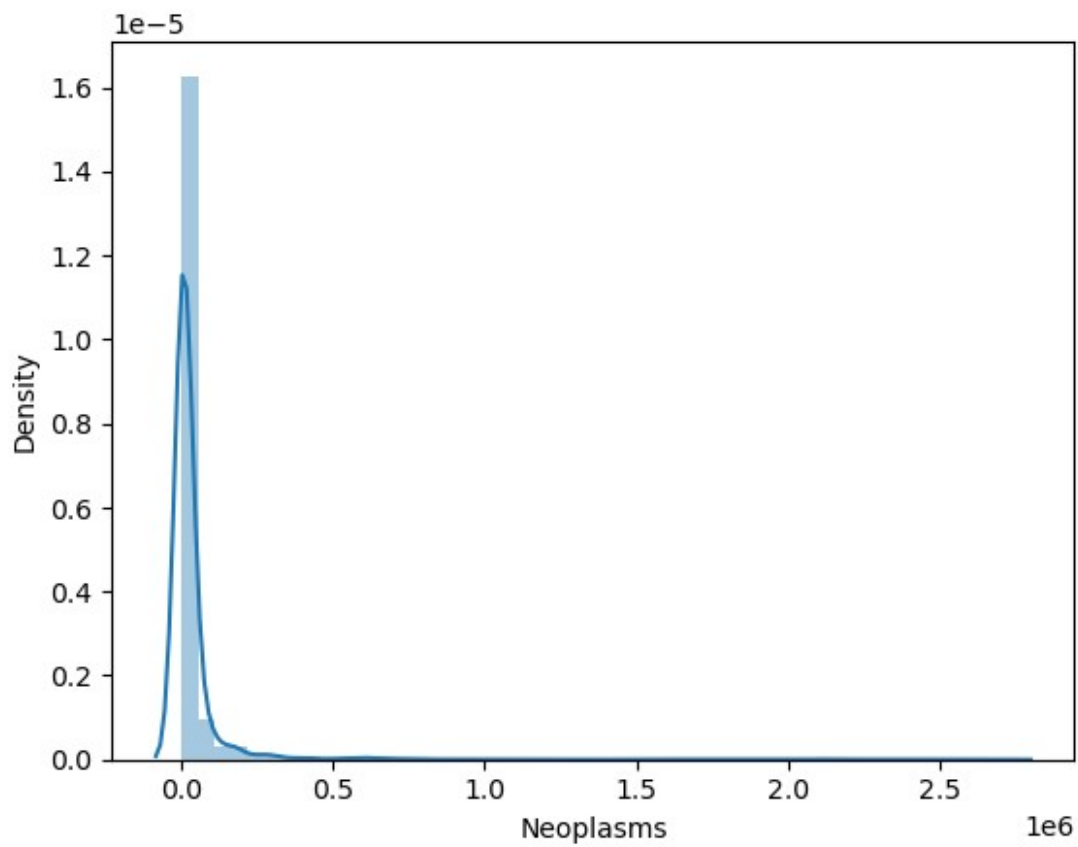
KstestResult(statistic=0.5, pvalue=0.0)
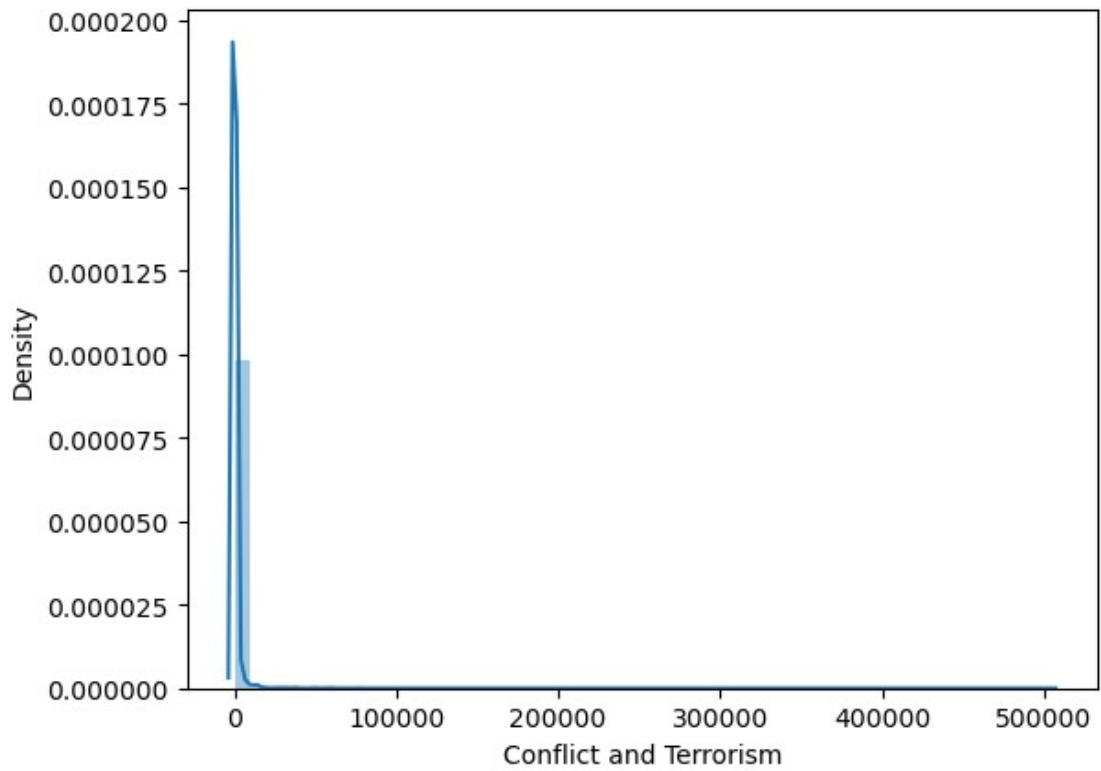
KstestResult(statistic=0.9919507555631412, pvalue=0.0)
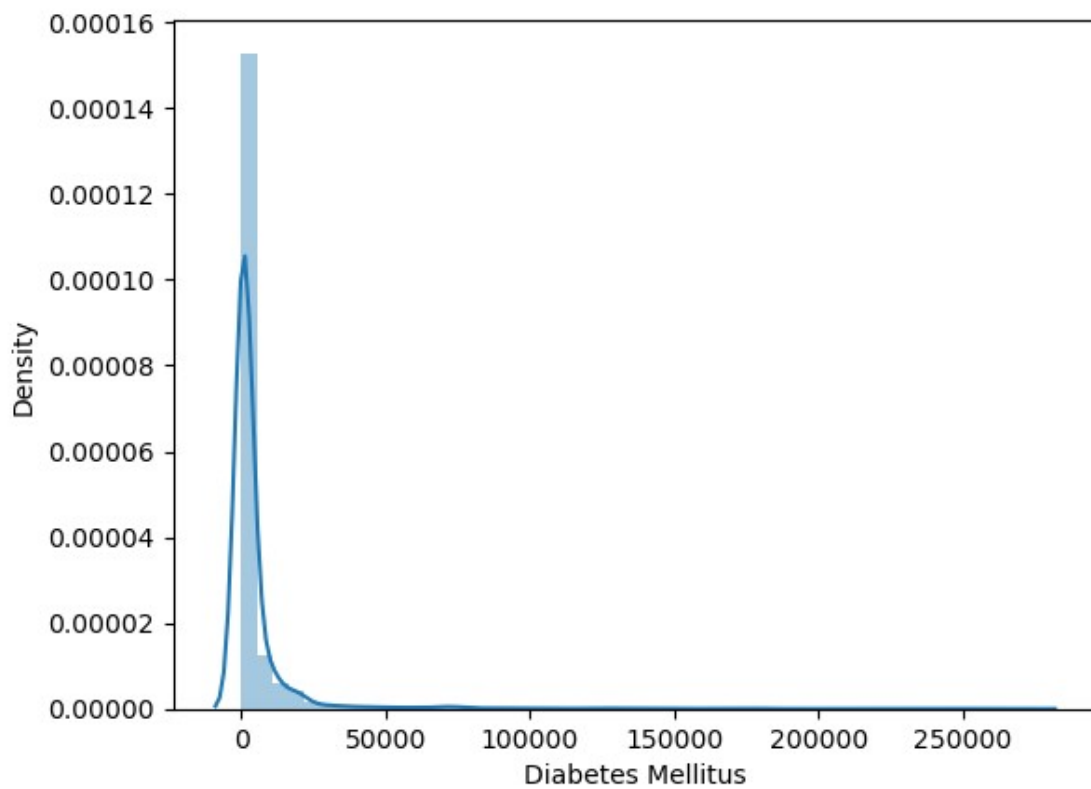
KstestResult(statistic=0.9849246117722915, pvalue=0.0)
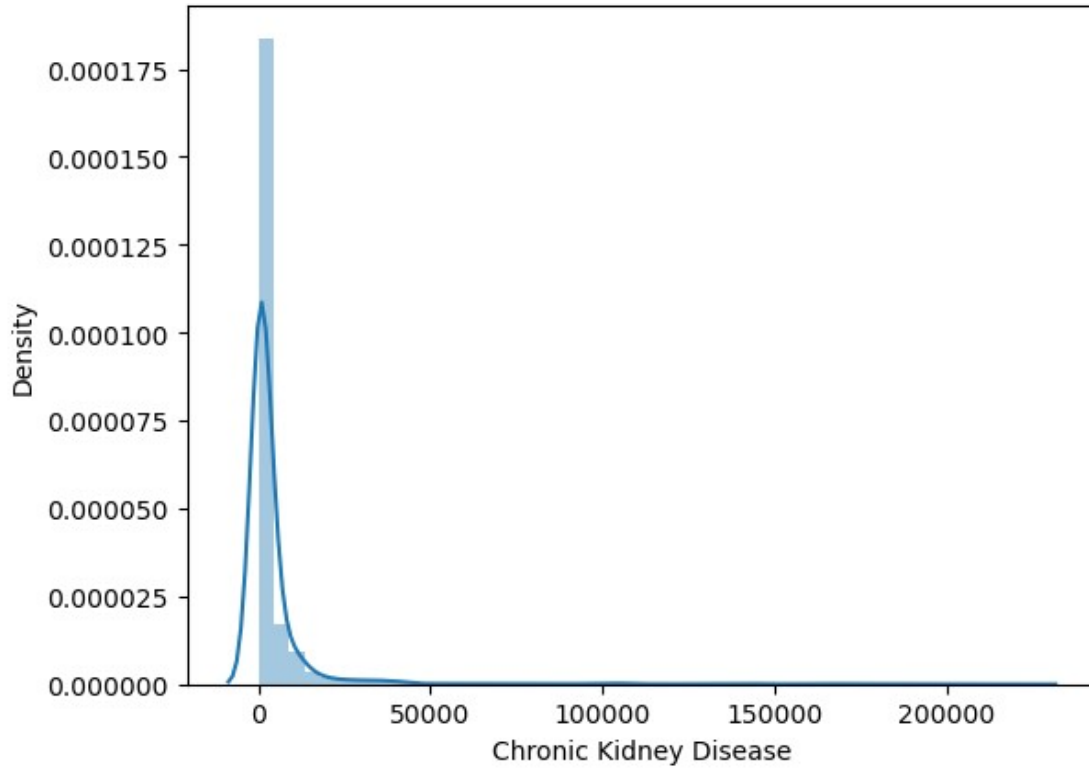
KstestResult(statistic=0.8133609791629319, pvalue=0.0)
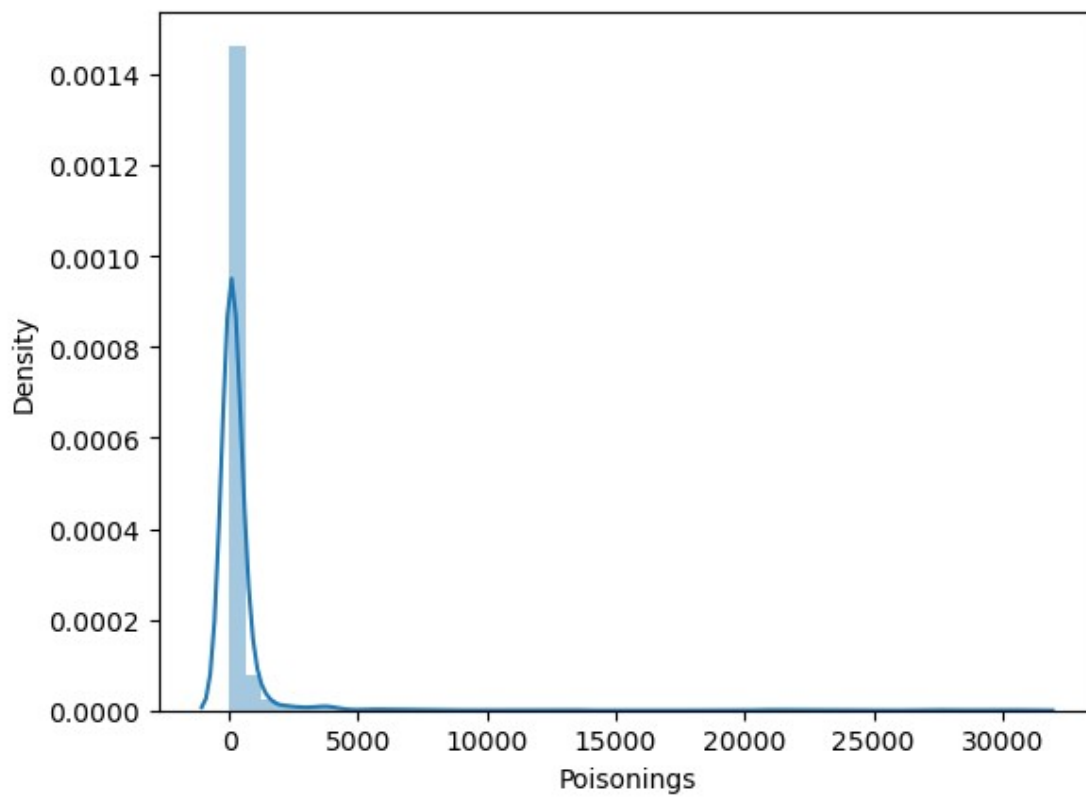
KstestResult(statistic=0.8520864693590103, pvalue=0.0)
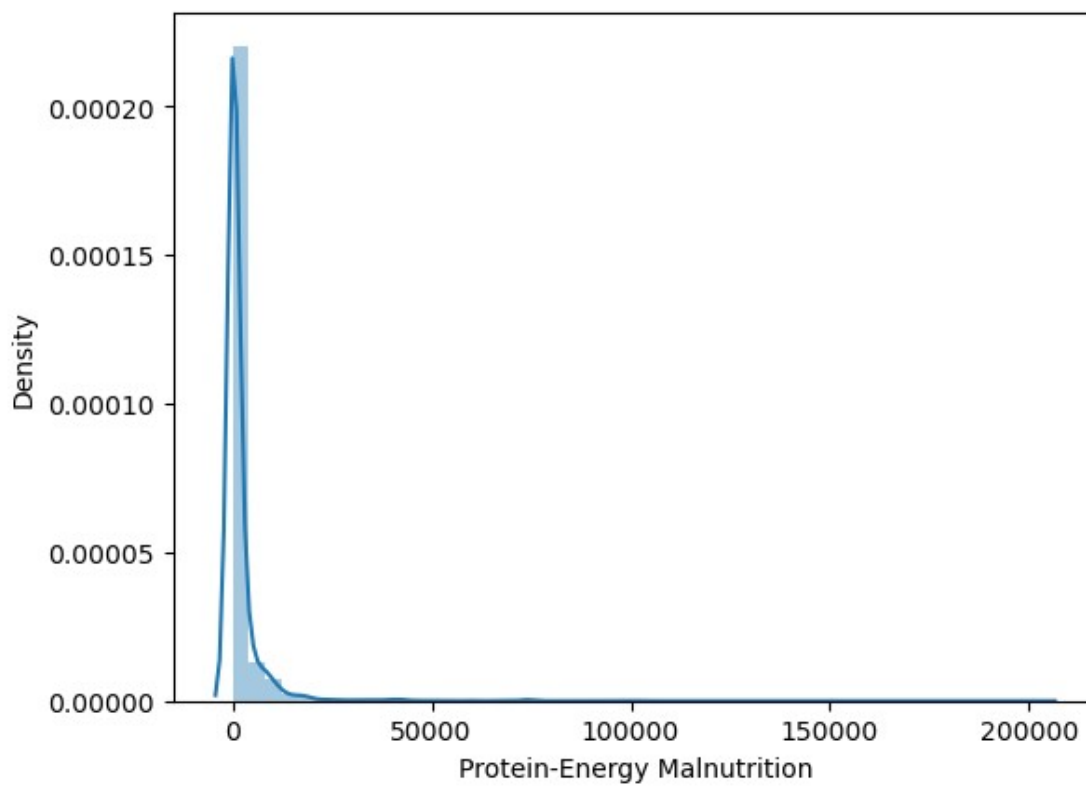
KstestResult(statistic=0.9829638274585659, pvalue=0.0)
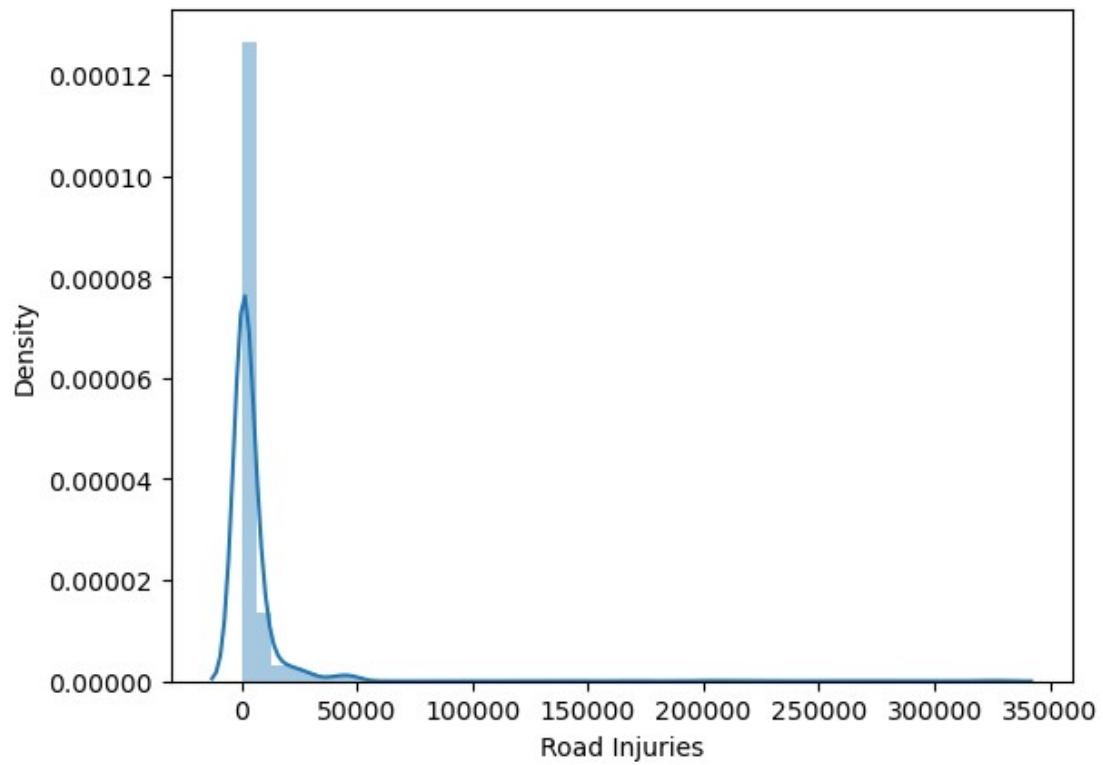
KstestResult(statistic=0.9890206163398663, pvalue=0.0)
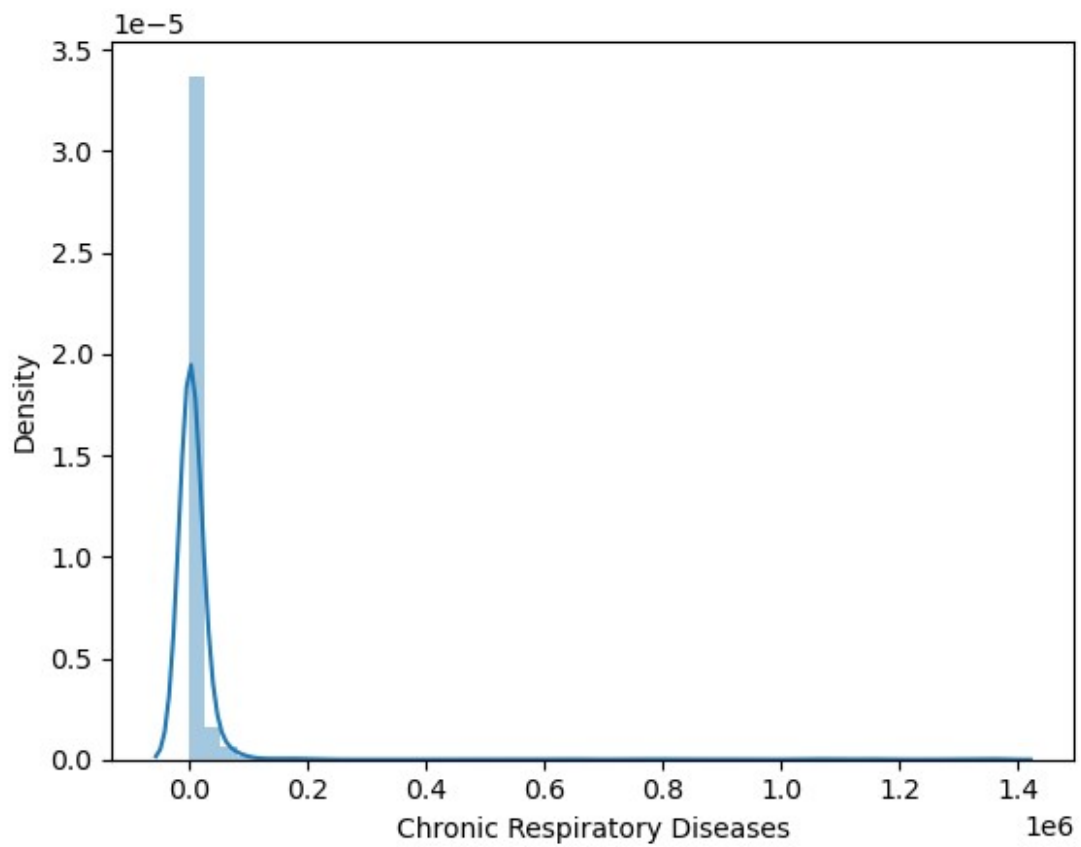
KstestResult(statistic=0.9738135006611803, pvalue=0.0)
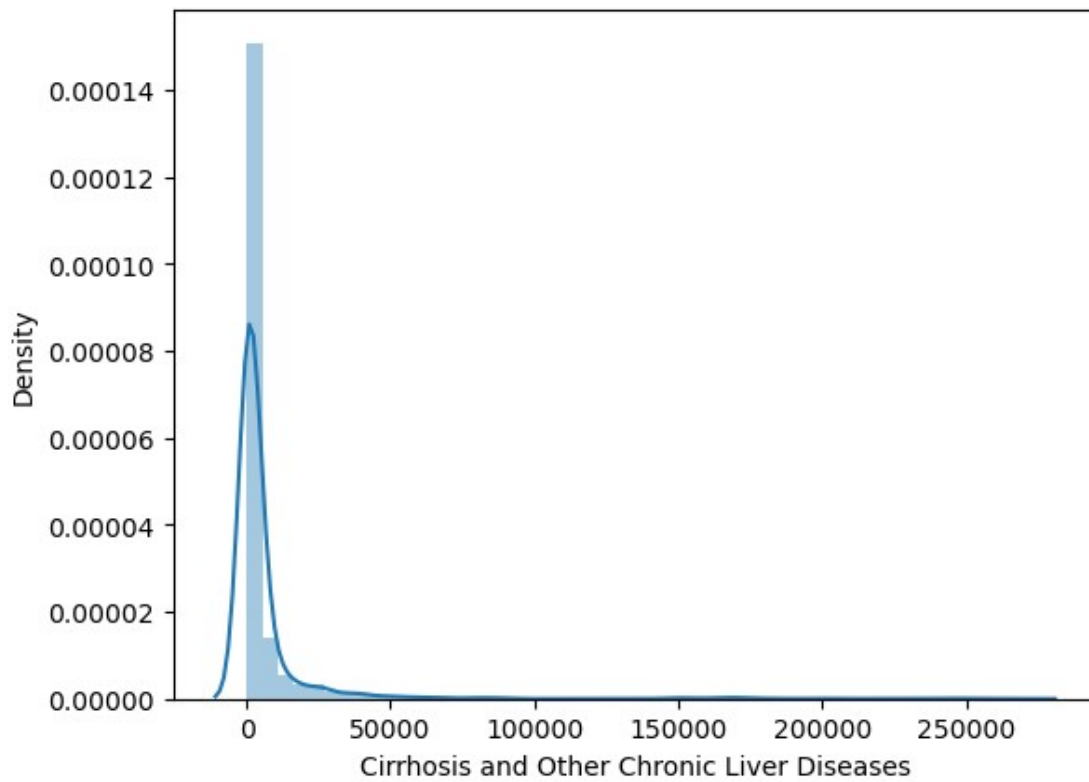
KstestResult(statistic=0.9878657882428797, pvalue=0.0)
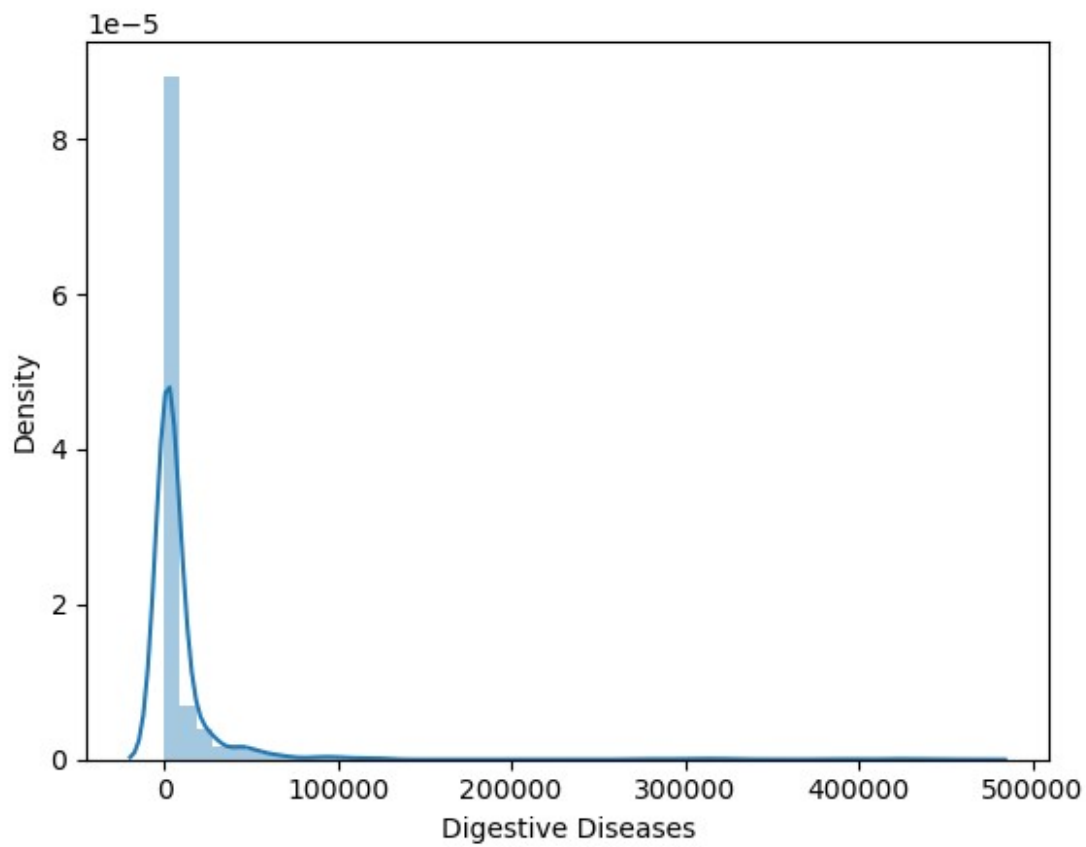
KstestResult(statistic=0.895550521646592, pvalue=0.0)
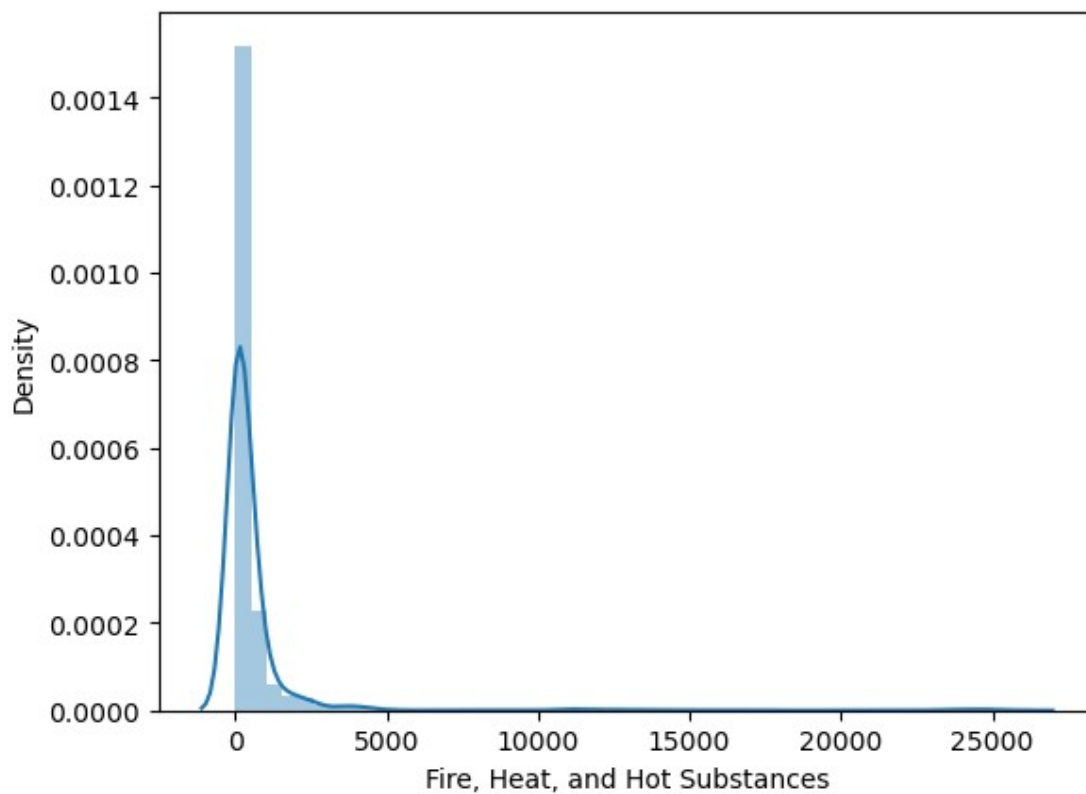
KstestResult(statistic=0.75551784190803, pvalue=0.0)
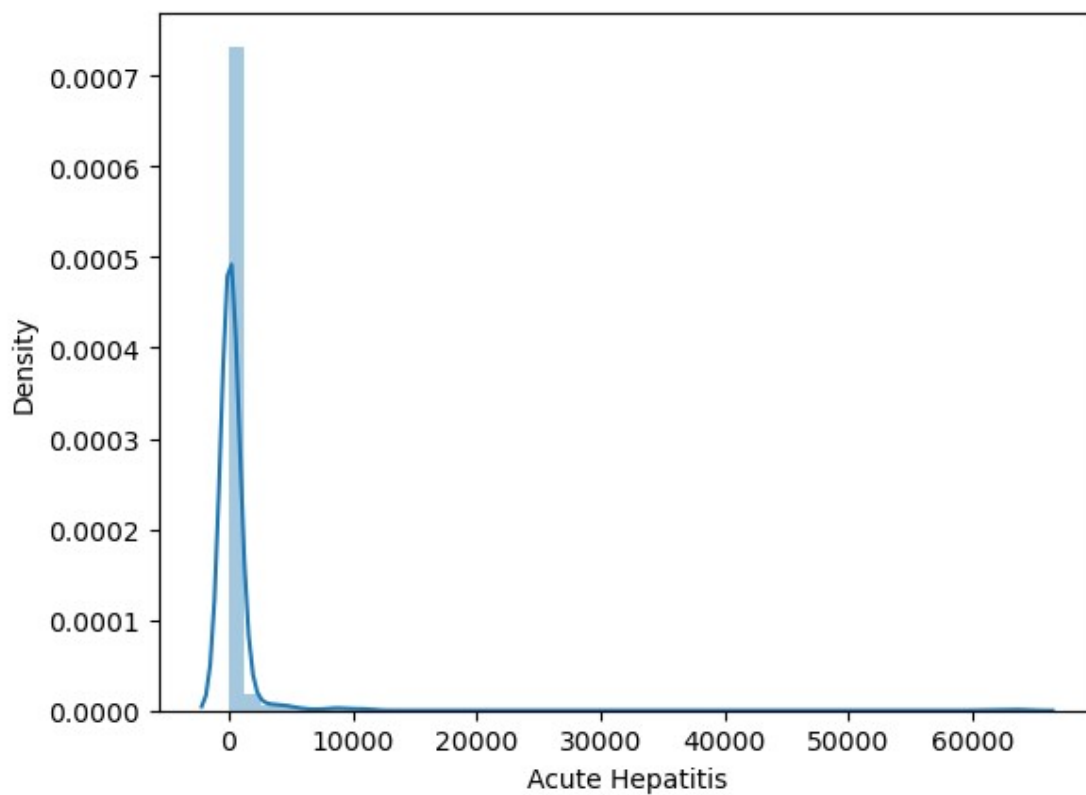
```
df1_cont.skew()
```

```
Meningitis                                    8.248599
Alzheimer's Disease and Other Dementias       8.695288
Parkinson's Disease                           9.321242
Nutritional Deficiencies                     14.863496
Malaria                                       9.159105
Drowning                                     10.553901
Interpersonal Violence                        5.777146
Maternal Disorders                           12.278358
HIV/AIDS                                      6.610169
Drug Use Disorders                           13.367100
Tuberculosis                                 12.022406
Cardiovascular Diseases                       9.411914
Lower Respiratory Infections                  9.036604
Neonatal Disorders                           10.447536
Alcohol Use Disorders                         9.253511
Self-harm                                     8.973653
Exposure to Forces of Nature                 34.507640
Diarrheal Diseases                           12.563758
Environmental Heat and Cold Exposure         10.631493
Neoplasms                                    10.748333
Conflict and Terrorism                       60.707004
Diabetes Mellitus                             8.022653
Chronic Kidney Disease                        7.919364
Poisonings                                   10.929760
Protein-Energy Malnutrition                  13.115196
Road Injuries                                 9.466209
Chronic Respiratory Diseases                 10.298131
Cirrhosis and Other Chronic Liver Diseases    7.764715
Digestive Diseases                            8.026536
Fire, Heat, and Hot Substances                8.578848
Acute Hepatitis                              12.756098
dtype: float64
```

```
df2=df.copy()
```

```python
# The unique Year data in the Dataframe
```

```
df1_cat['Year'].unique()
```

```
array([1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999,
2000,
       2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
2011,
       2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019], dtype=object)
```

```python
# Creating a new column for 'Total_no_of_Deaths' for individual
Country and Year
```

```
df2['Total_no_of_Deaths'] = df2.sum(axis=1)
```

```
df2

     Country/Territory Code  Year  Meningitis  \
0          Afghanistan  AFG  1990        2159
1          Afghanistan  AFG  1991        2218
2          Afghanistan  AFG  1992        2475
3          Afghanistan  AFG  1993        2812
4          Afghanistan  AFG  1994        3027
...                ...  ...   ...         ...
6115          Zimbabwe  ZWE  2015        1439
6116          Zimbabwe  ZWE  2016        1457
6117          Zimbabwe  ZWE  2017        1460
6118          Zimbabwe  ZWE  2018        1450
6119          Zimbabwe  ZWE  2019        1450

      Alzheimer's Disease and Other Dementias  Parkinson's Disease  \
0                                        1116                  371
1                                        1136                  374
2                                        1162                  378
3                                        1187                  384
4                                        1211                  391
...                                       ...                  ...
6115                                      754                  215
6116                                      767                  219
6117                                      781                  223
6118                                      795                  227
6119                                      812                  232

      Nutritional Deficiencies  Malaria  Drowning  Interpersonal
Violence  \
0                         2087       93      1370
1538
1                         2153      189      1391
2001
2                         2441      239      1514
2299
3                         2837      108      1687
2589
4                         3081      211      1809
2849
...                        ...      ...       ...
...
6115                      3019     2518       770
1302
6116                      3056     2050       801
1342
6117                      2990     2116       818
1363
6118                      2918     2088       825
1396
```

```
6119                        2884      2068        827
1434
```

	Maternal Disorders	HIV/AIDS	Drug Use Disorders	Tuberculosis
0	2655	34	93	4661
1	2885	41	102	4743
2	3315	48	118	4976
3	3671	56	132	5254
4	3863	63	142	5470
...	...	...	...	...
6115	1355	29162	104	11214
6116	1338	27141	110	10998
6117	1312	24846	115	10762
6118	1294	22106	121	10545
6119	1294	20722	127	10465

	Cardiovascular Diseases	Lower Respiratory Infections
0	44899	23741
1	45492	24504
2	46557	27404
3	47951	31116
4	49308	33390
...	...	...
6115	16649	12974
6116	16937	13024
6117	17187	12961
6118	17460	12860
6119	17810	12897

	Neonatal Disorders	Alcohol Use Disorders	Self-harm
0	15612	72	696
1	17128	75	751
2	20060	80	855
3	22335	85	943
4	23288	88	993
...	...	...	...
6115	9278	48	2235

```
6116                9065                     49      2296
6117                8901                     50      2338
6118                8697                     51      2372
6119                8609                     53      2403

      Exposure to Forces of Nature  Diarrheal Diseases  \
0                              0                4235
1                           1347                4927
2                            614                6123
3                            225                8174
4                            160                8215
...                          ...                 ...
6115                          16                5102
6116                          31                5002
6117                         251                4948
6118                           0                4745
6119                         660                4635

      Environmental Heat and Cold Exposure  Neoplasms  Conflict and
Terrorism  \
0                                      175      11580
1490
1                                      113      11796
3370
2                                       38      12218
4344
3                                       41      12634
4096
4                                       44      12914
8959
...                                    ...        ...
...
6115                                    37      11161
13
6116                                    37      11465
6
6117                                    37      11744
5
6118                                    37      12038
9
6119                                    37      12353
11

      Diabetes Mellitus  Chronic Kidney Disease  Poisonings  \
0                  2108                    3709         338
1                  2120                    3724         351
2                  2153                    3776         386
3                  2195                    3862         425
4                  2231                    3932         451
...                 ...                     ...         ...
```

```
6115                3176              2108        381
6116                3259              2160        393
6117                3313              2196        398
6118                3381              2240        400
6119                3460              2292        405

        Protein-Energy Malnutrition  Road Injuries  \
0                              2054           4154
1                              2119           4472
2                              2404           5106
3                              2797           5681
4                              3038           6001
...                             ...            ...
6115                           2990           2373
6116                           3027           2436
6117                           2962           2473
6118                           2890           2509
6119                           2855           2554

        Chronic Respiratory Diseases  \
0                               5945
1                               6050
2                               6223
3                               6445
4                               6664
...                              ...
6115                            2751
6116                            2788
6117                            2818
6118                            2849
6119                            2891

      Cirrhosis and Other Chronic Liver Diseases  Digestive
Diseases  \
0                                           2673            5005

1                                           2728            5120

2                                           2830            5335

3                                           2943            5568

4                                           3027            5739

...                                          ...             ...

6115                                        1956            4202

6116                                        1962            4264
```

```
6117                                    2007                    4342

6118                                    2030                    4377

6119                                    2065                    4437


        Fire, Heat, and Hot Substances   Acute Hepatitis
Total_no_of_Deaths
0                                 323                2985
147971
1                                 332                3092
156844
2                                 360                3325
169156
3                                 396                3601
182230
4                                 420                3816
194795
...                               ...                 ...
...
6115                              632                 146
130080
6116                              648                 146
128274
6117                              654                 144
126515
6118                              657                 139
123506
6119                              662                 136
123540

[6120 rows x 35 columns]
```

Bivariate analysis

```python
for i in df1_cont.columns:
    df2.groupby('Country/Territory')
[i].sum().sort_values(ascending=False).head().plot(kind='bar')
    plt.title('Top 5 countries having death due to '+i)
    plt.ylabel('Sum of Deaths')
    plt.show()
```

Top 5 countries having death due to Meningitis
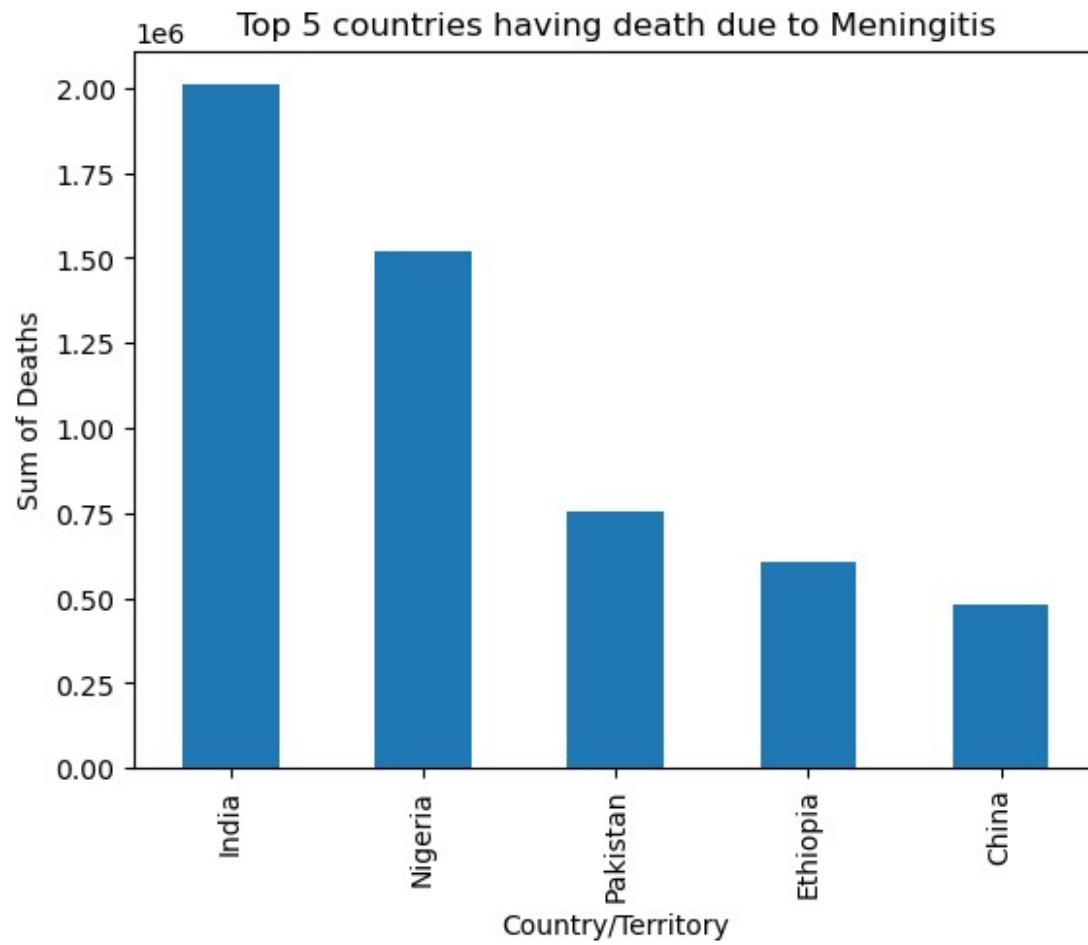
Top 5 countries having death due to Alzheimer's Disease and Other Dementias

Top 5 countries having death due to Parkinson's Disease
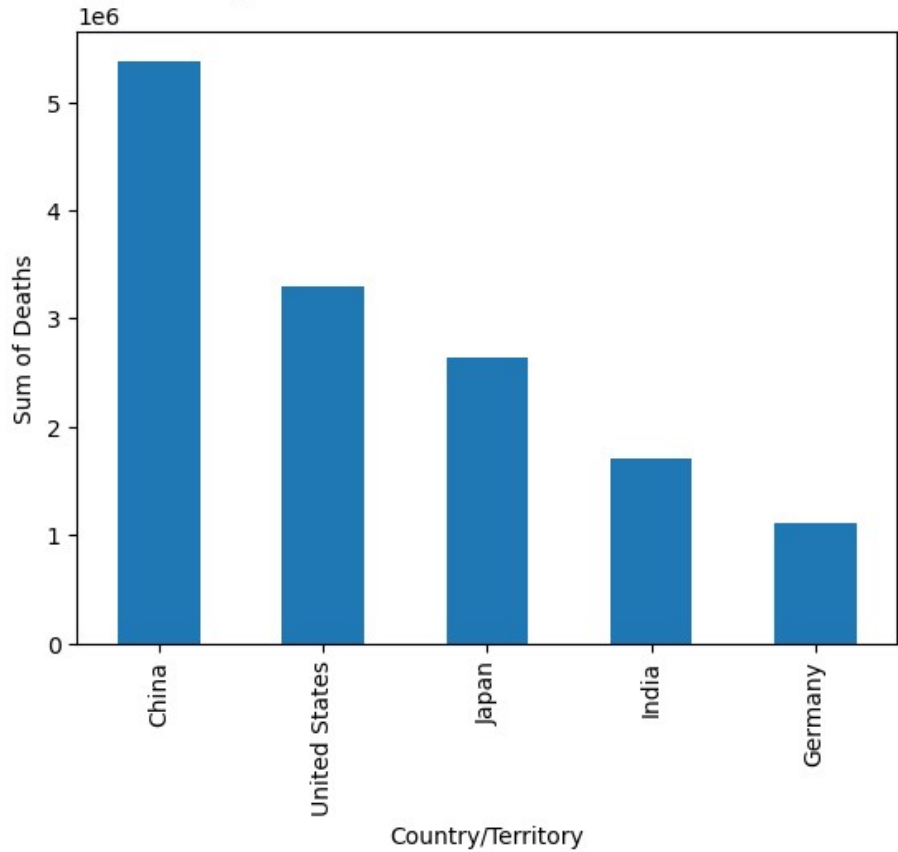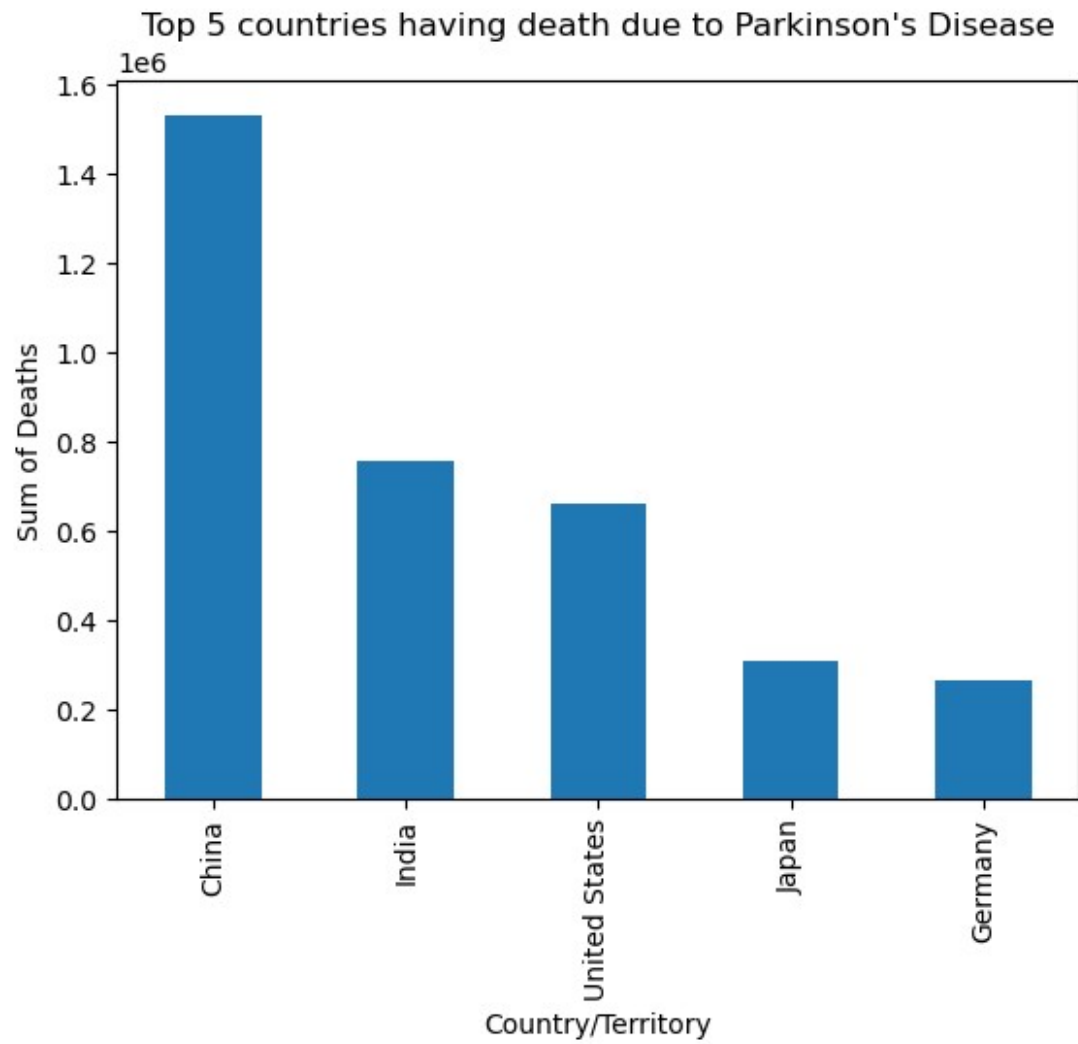
Top 5 countries having death due to Nutritional Deficiencies

Top 5 countries having death due to Malaria

Top 5 countries having death due to Drowning

Top 5 countries having death due to Interpersonal Violence
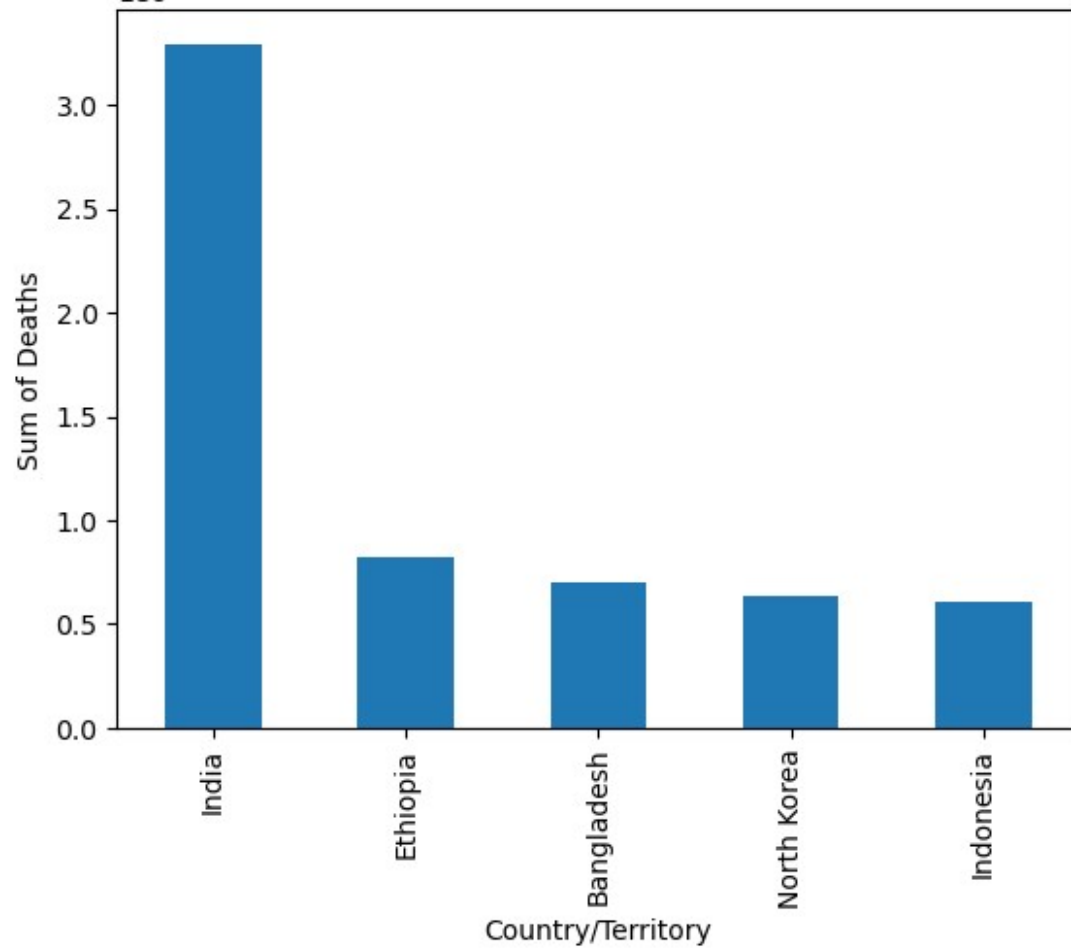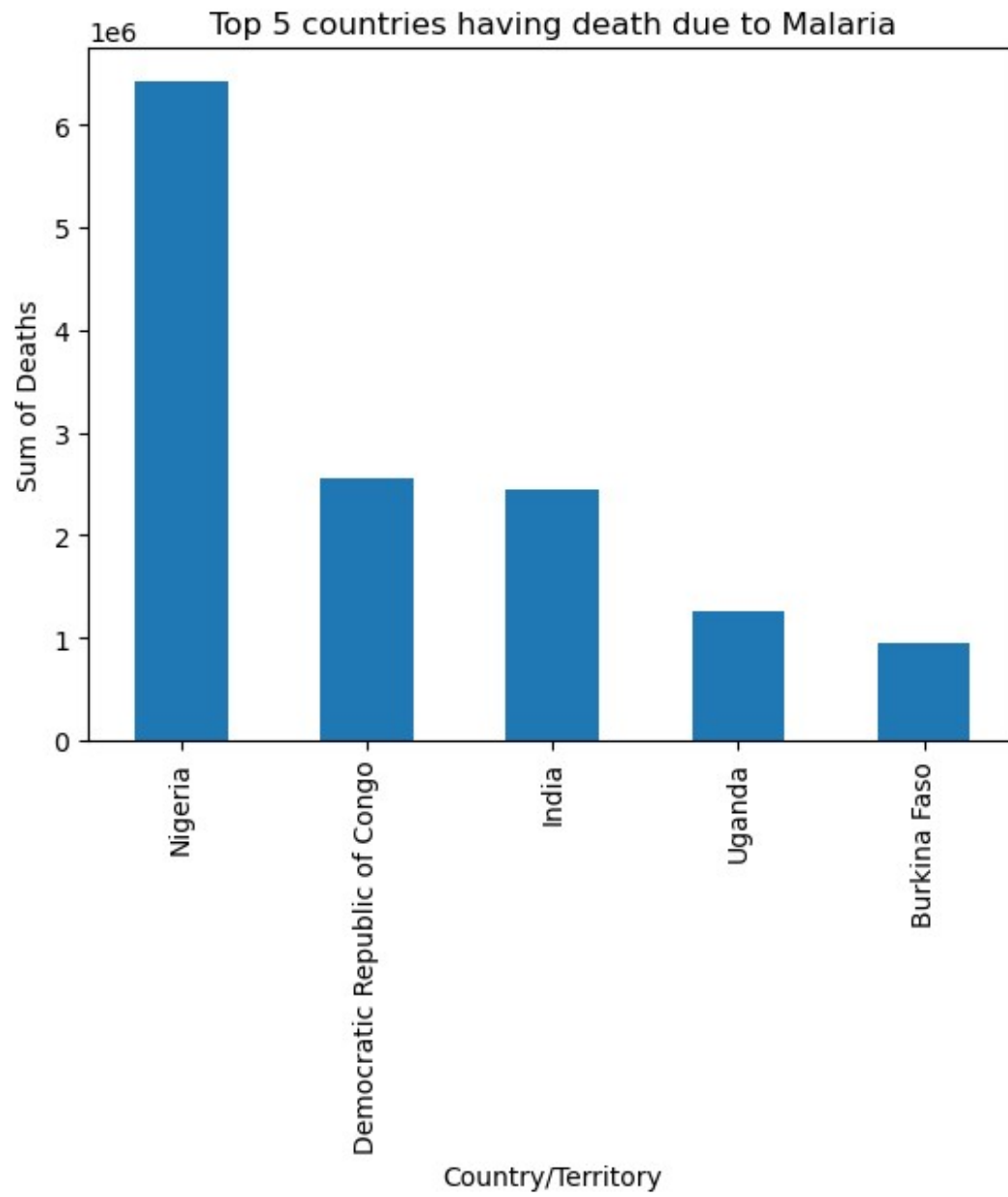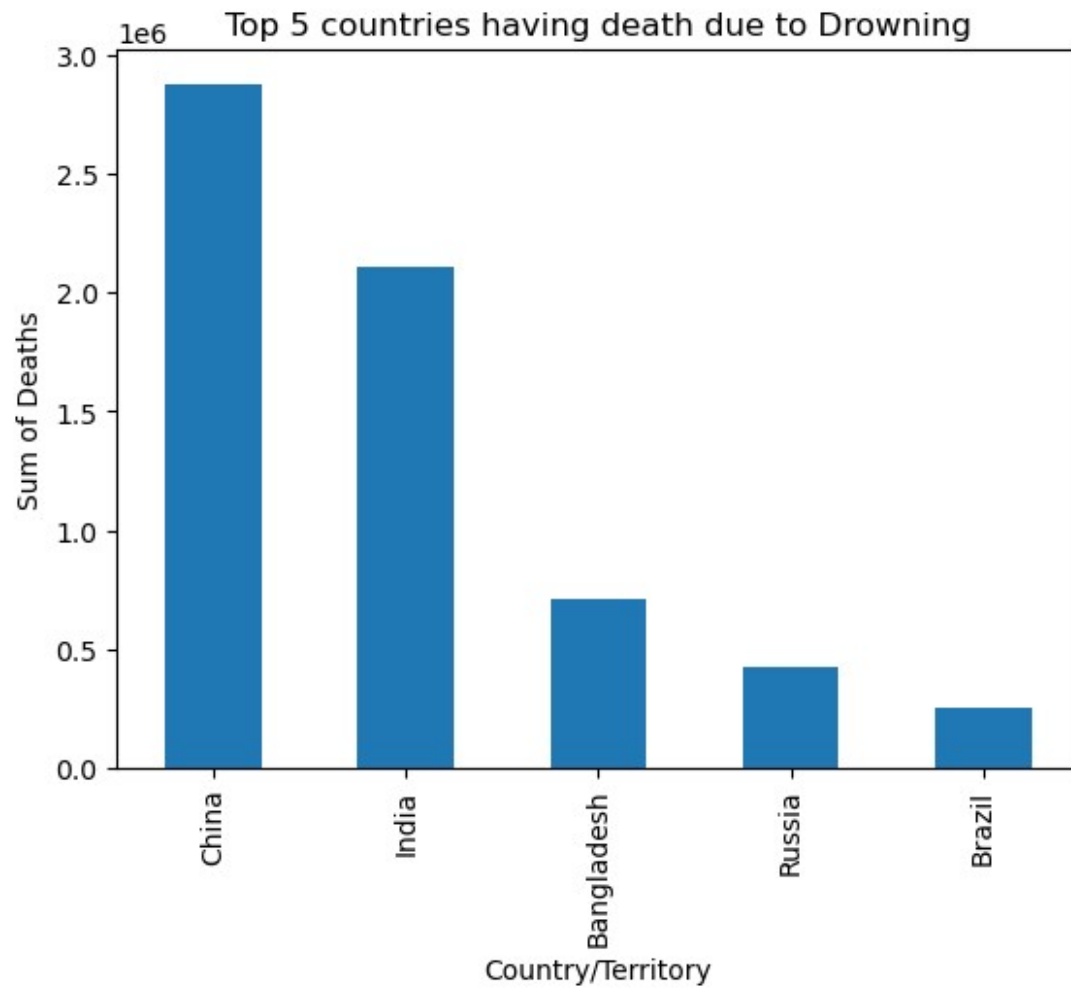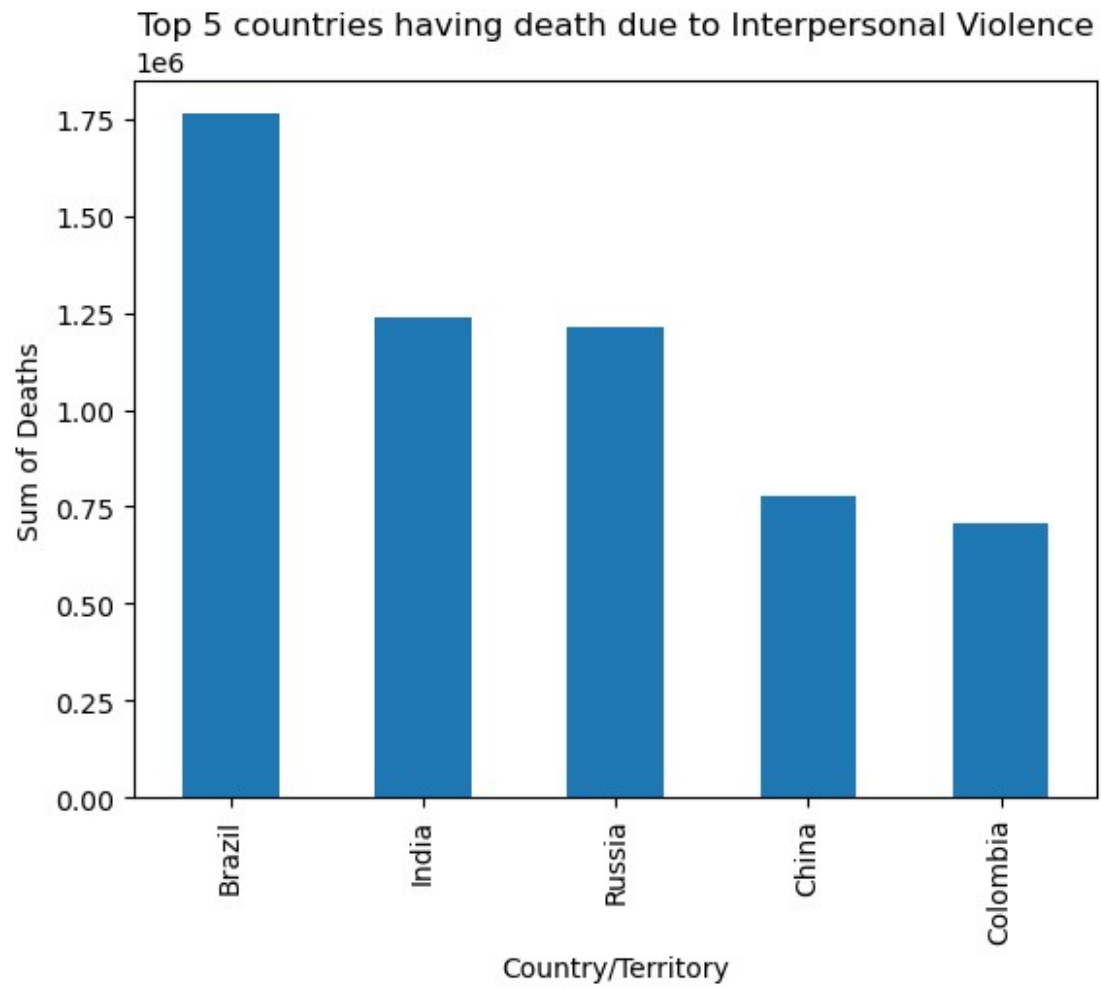
Top 5 countries having death due to Maternal Disorders

Top 5 countries having death due to HIV/AIDS
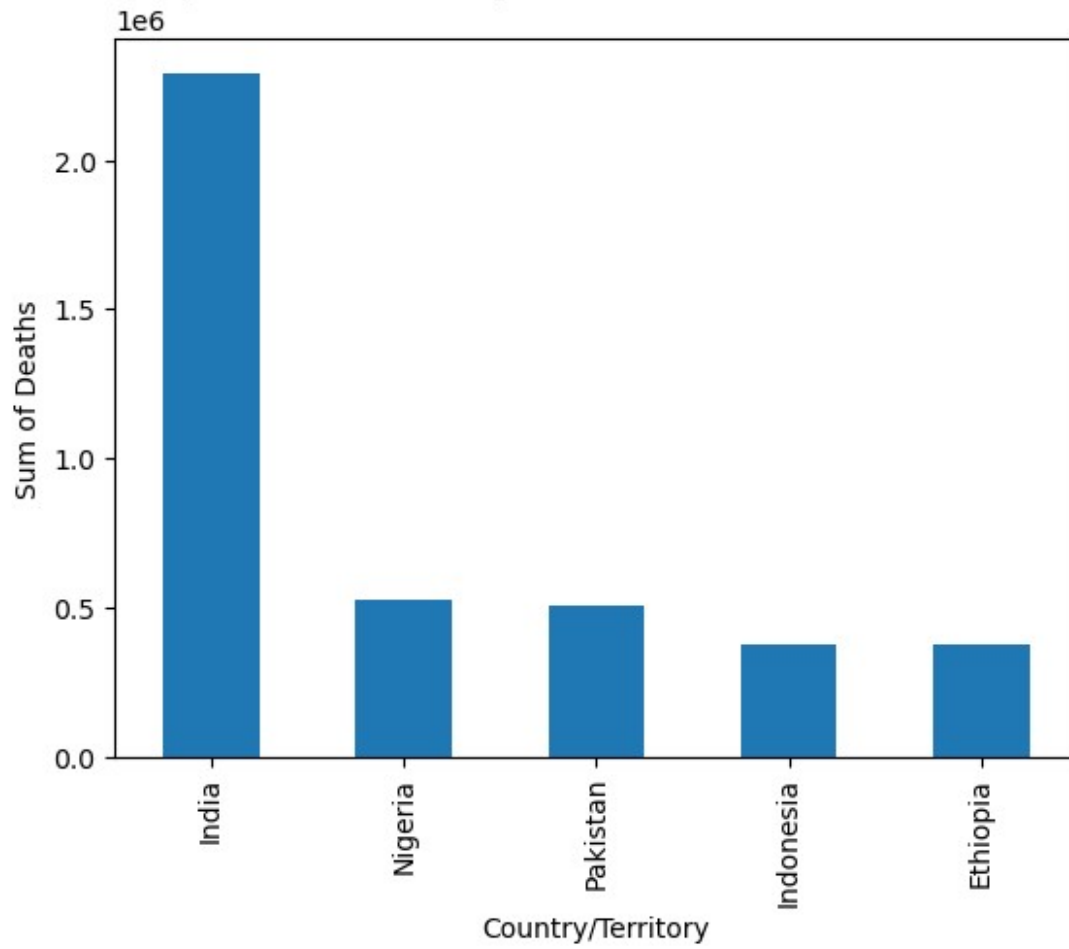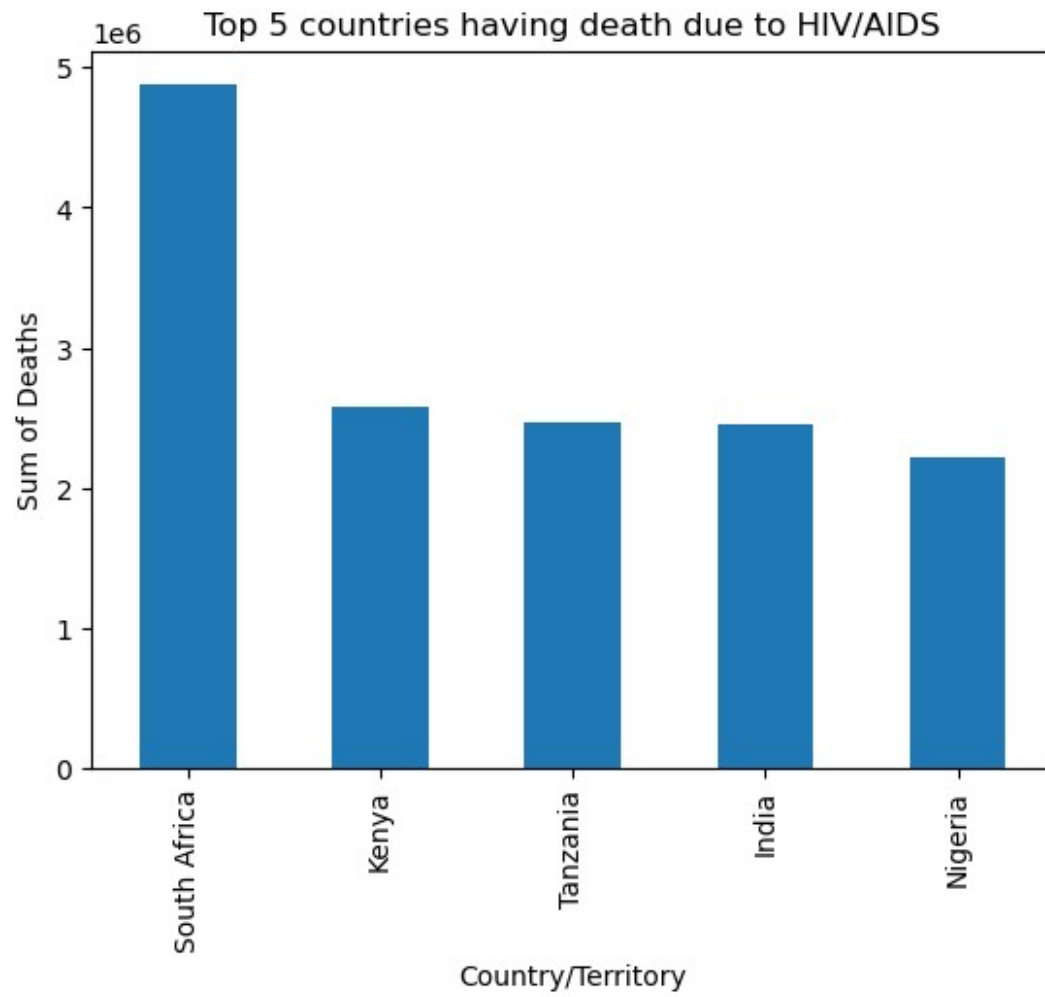
Top 5 countries having death due to Drug Use Disorders

Top 5 countries having death due to Tuberculosis
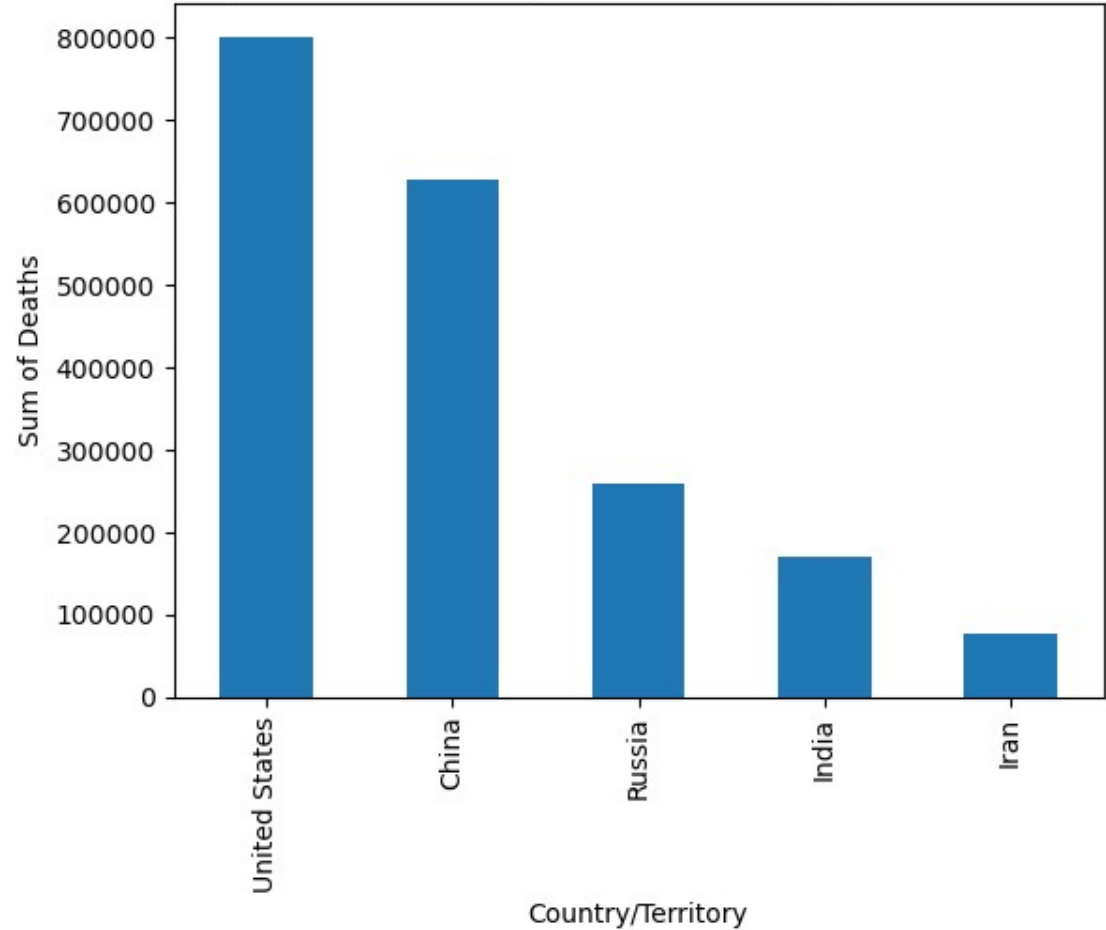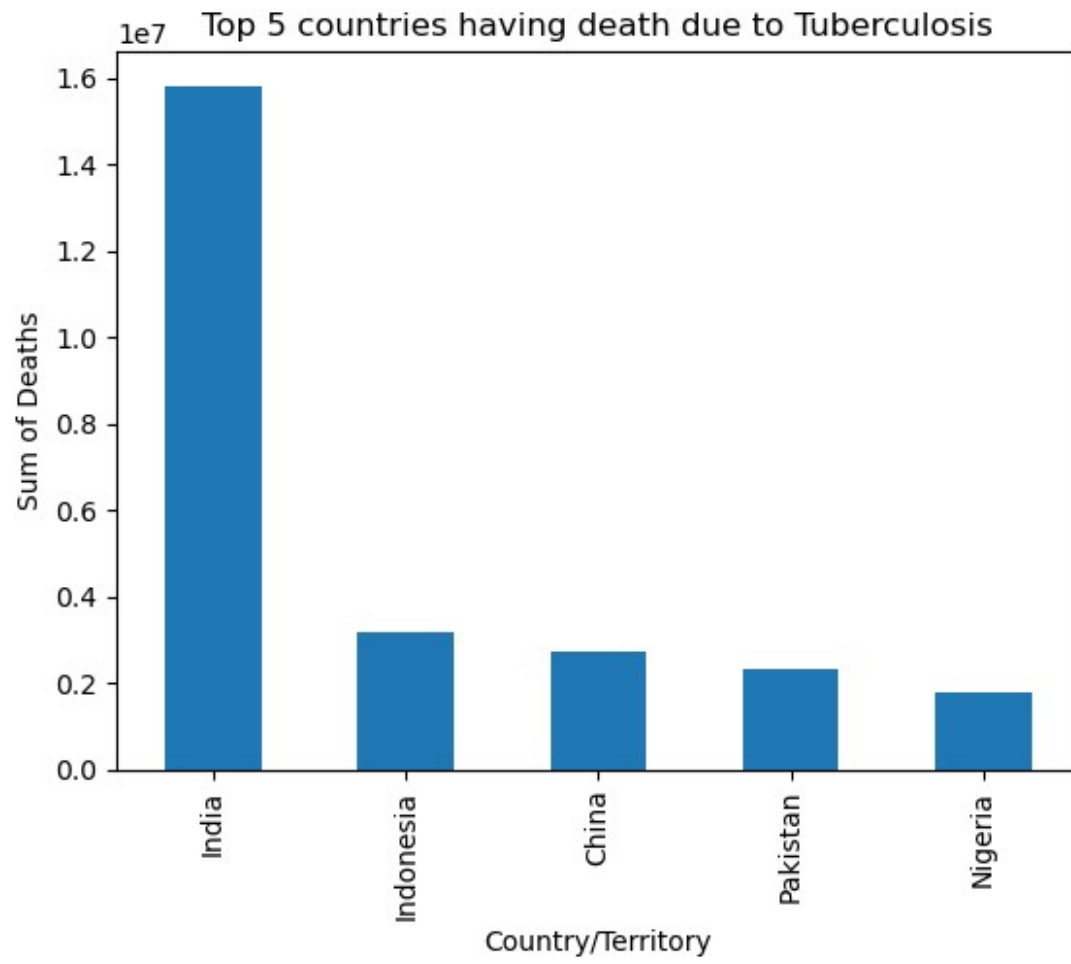
Top 5 countries having death due to Cardiovascular Diseases
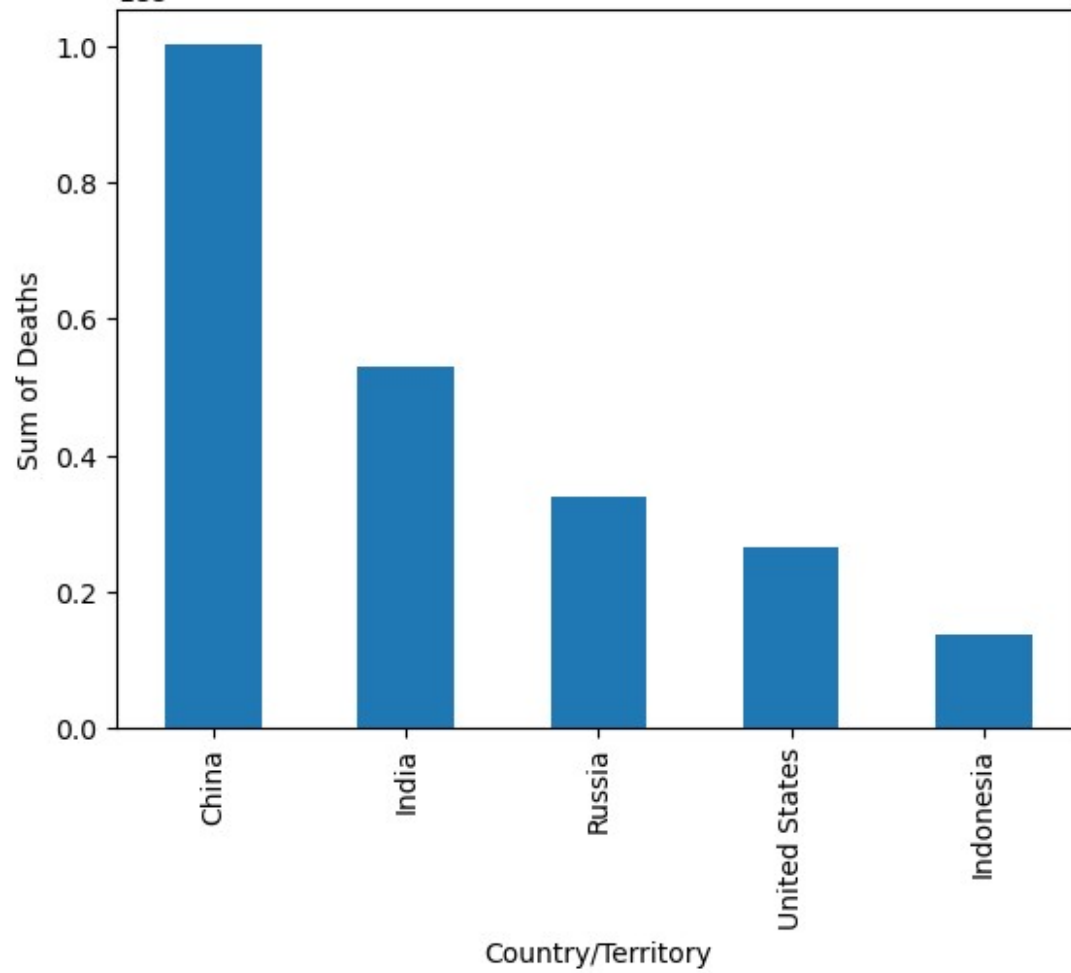
Top 5 countries having death due to Lower Respiratory Infections

Top 5 countries having death due to Neonatal Disorders

Top 5 countries having death due to Alcohol Use Disorders

Top 5 countries having death due to Self-harm

Top 5 countries having death due to Exposure to Forces of Nature
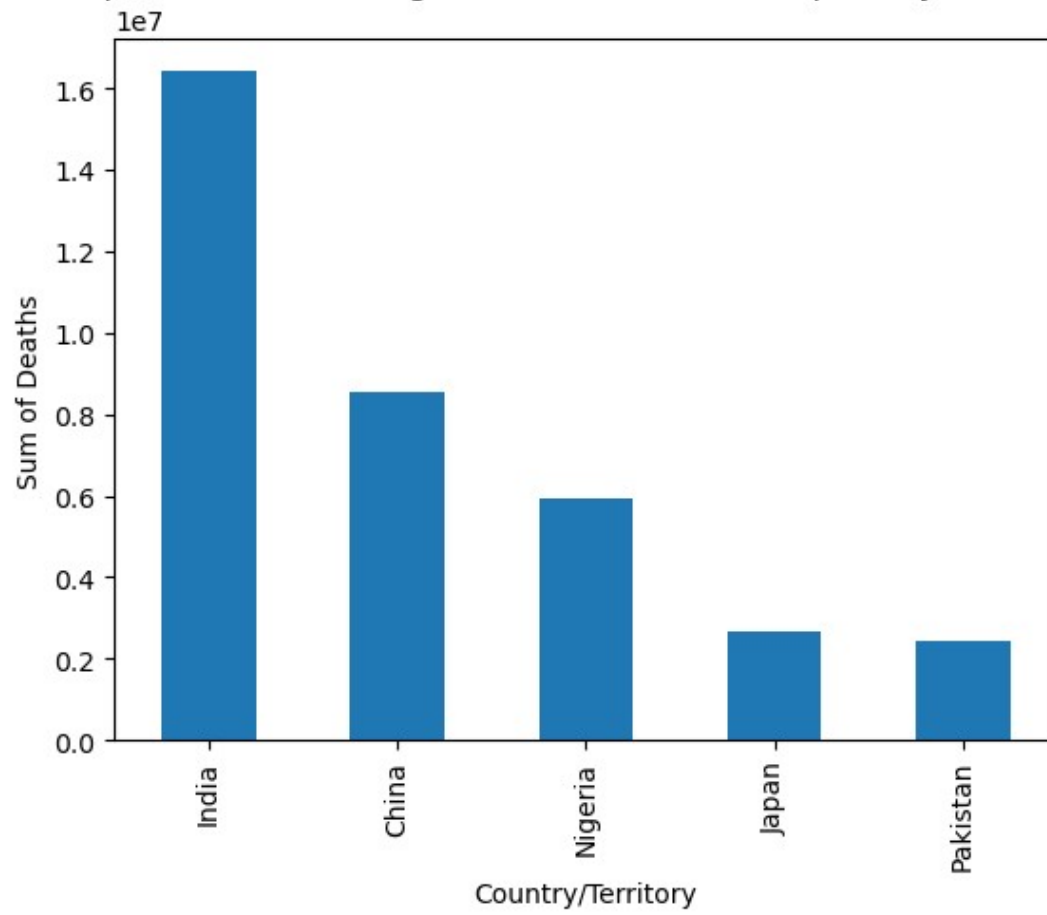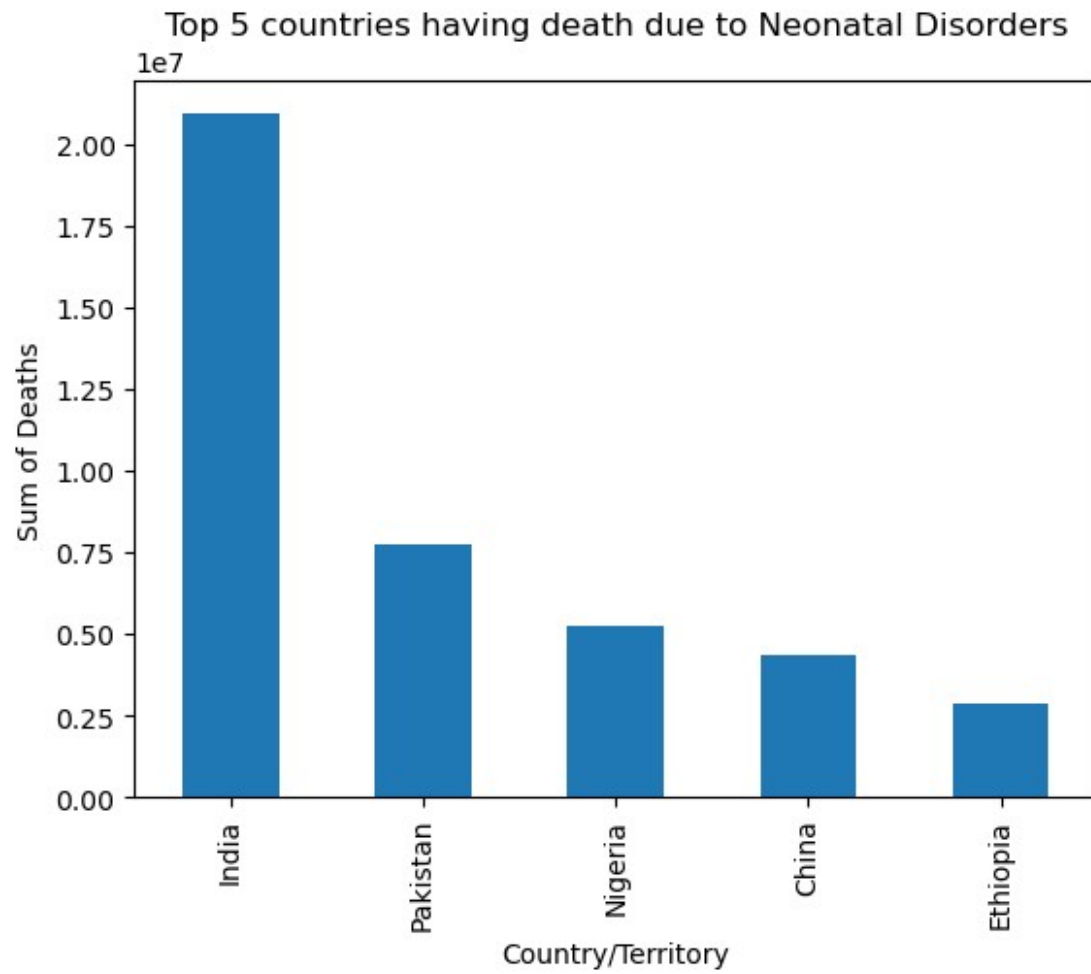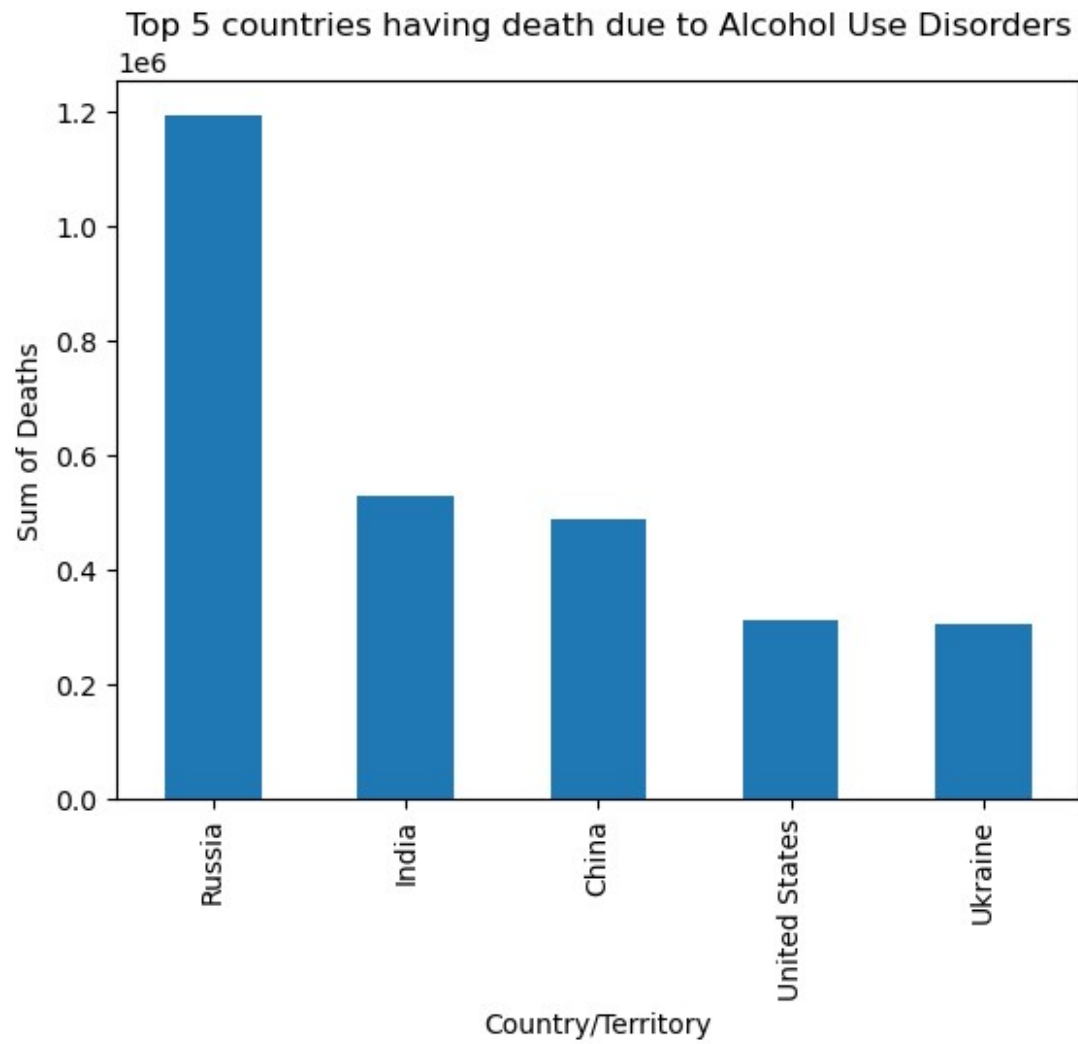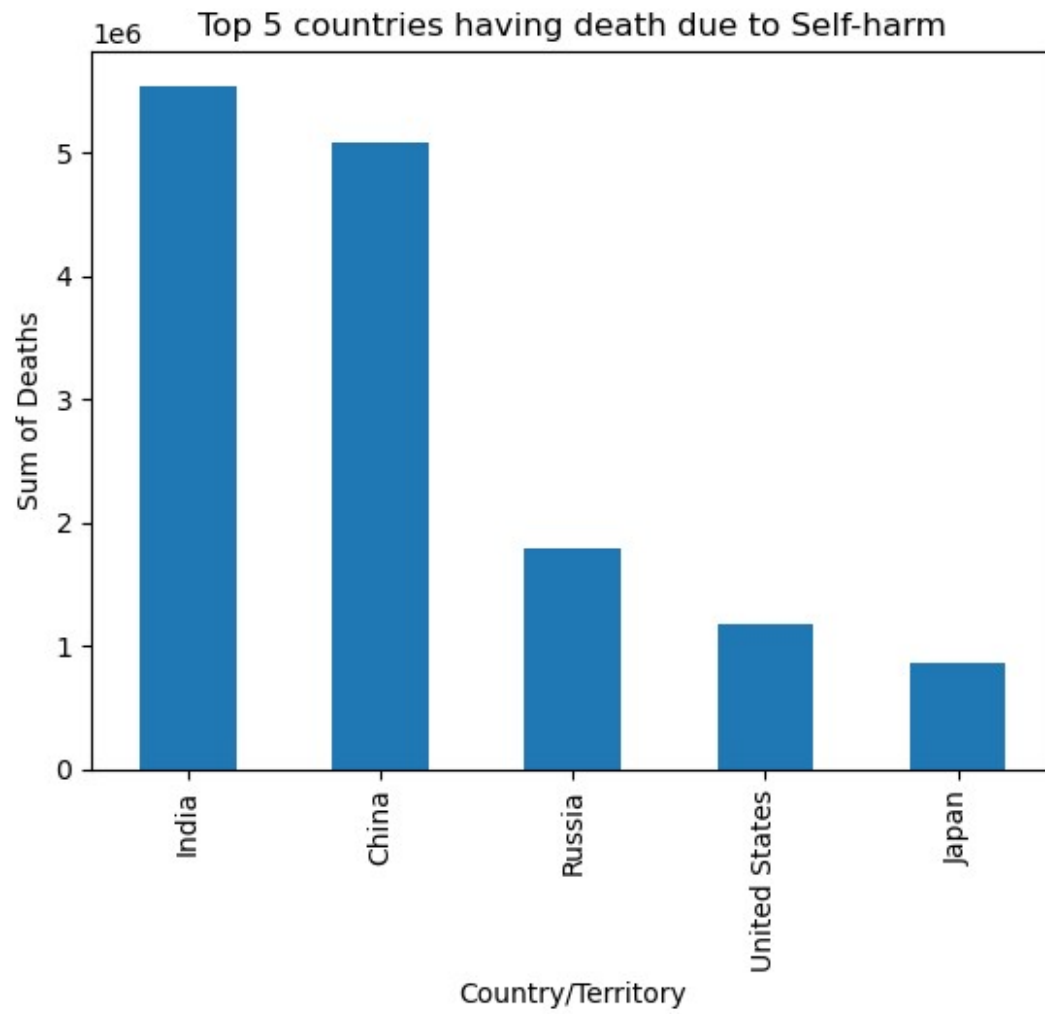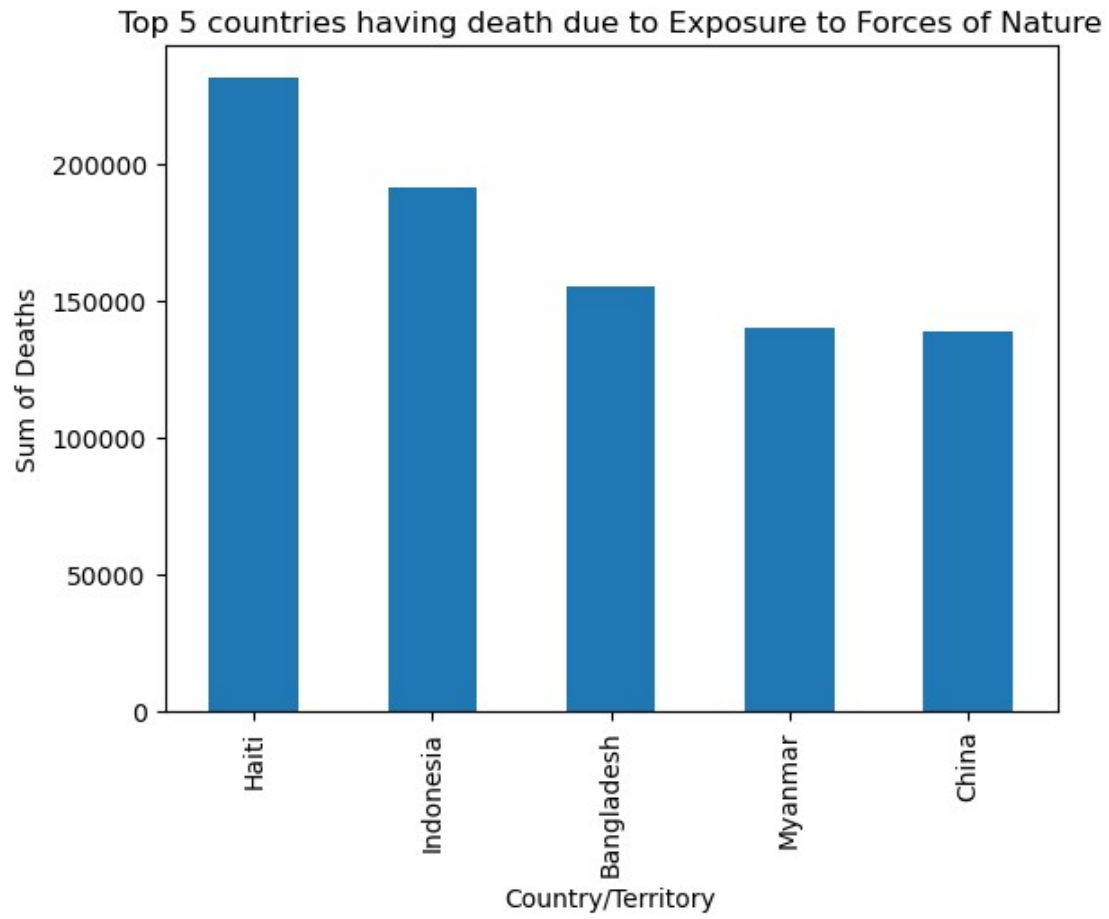
Top 5 countries having death due to Diarrheal Diseases

1e7

Sum of Deaths

Country/Territory

India
Nigeria
Pakistan
Indonesia
Ethiopia

Top 5 countries having death due to Environmental Heat and Cold Exposure

Top 5 countries having death due to Neoplasms
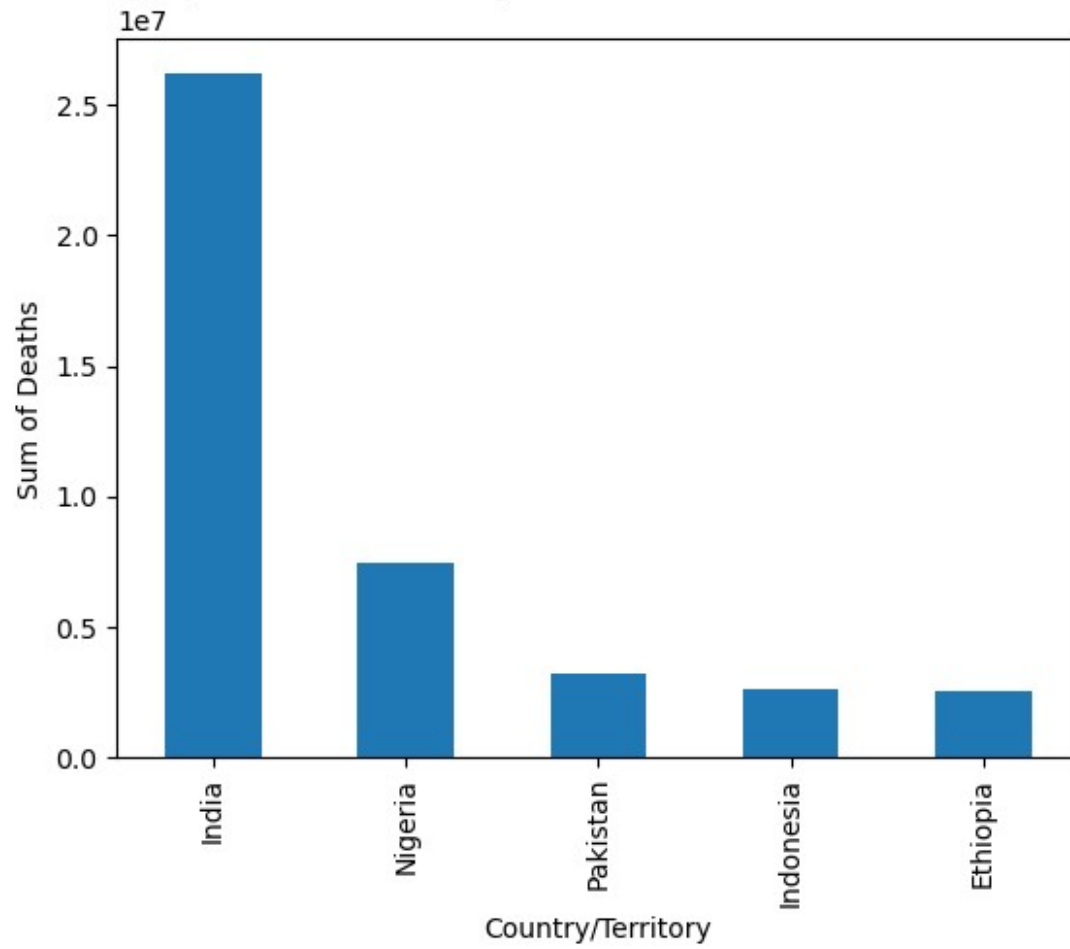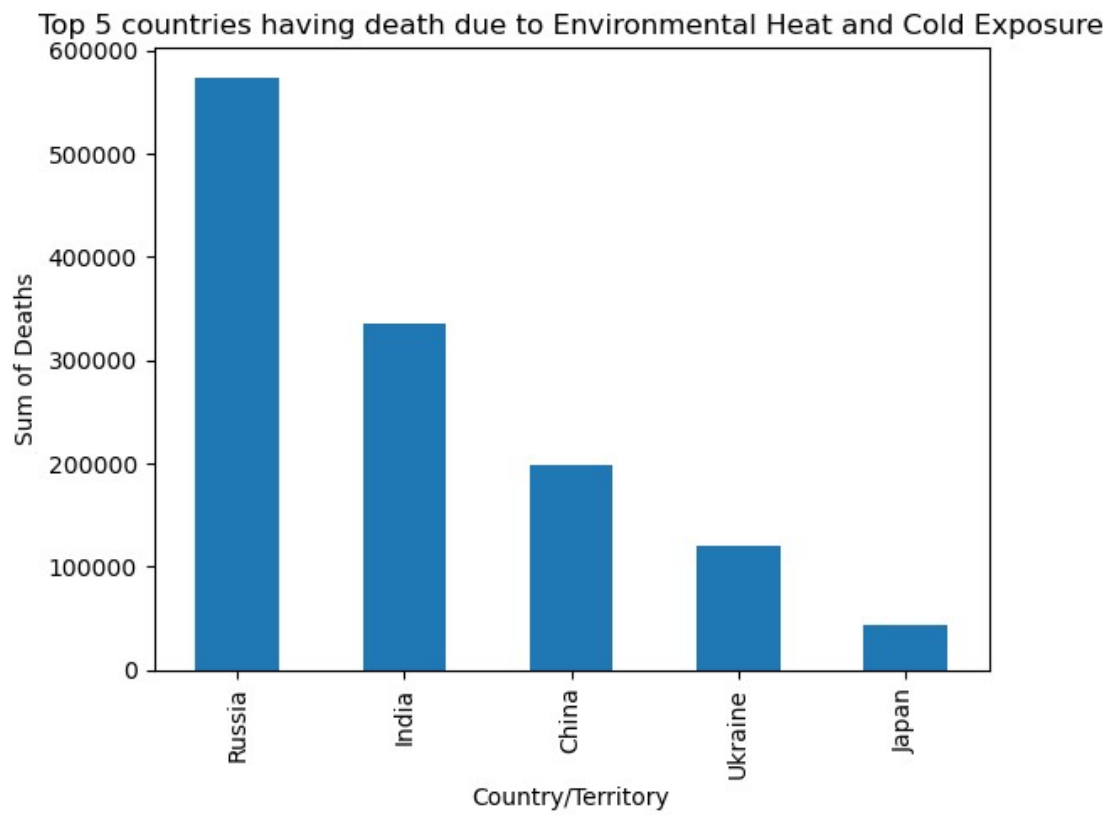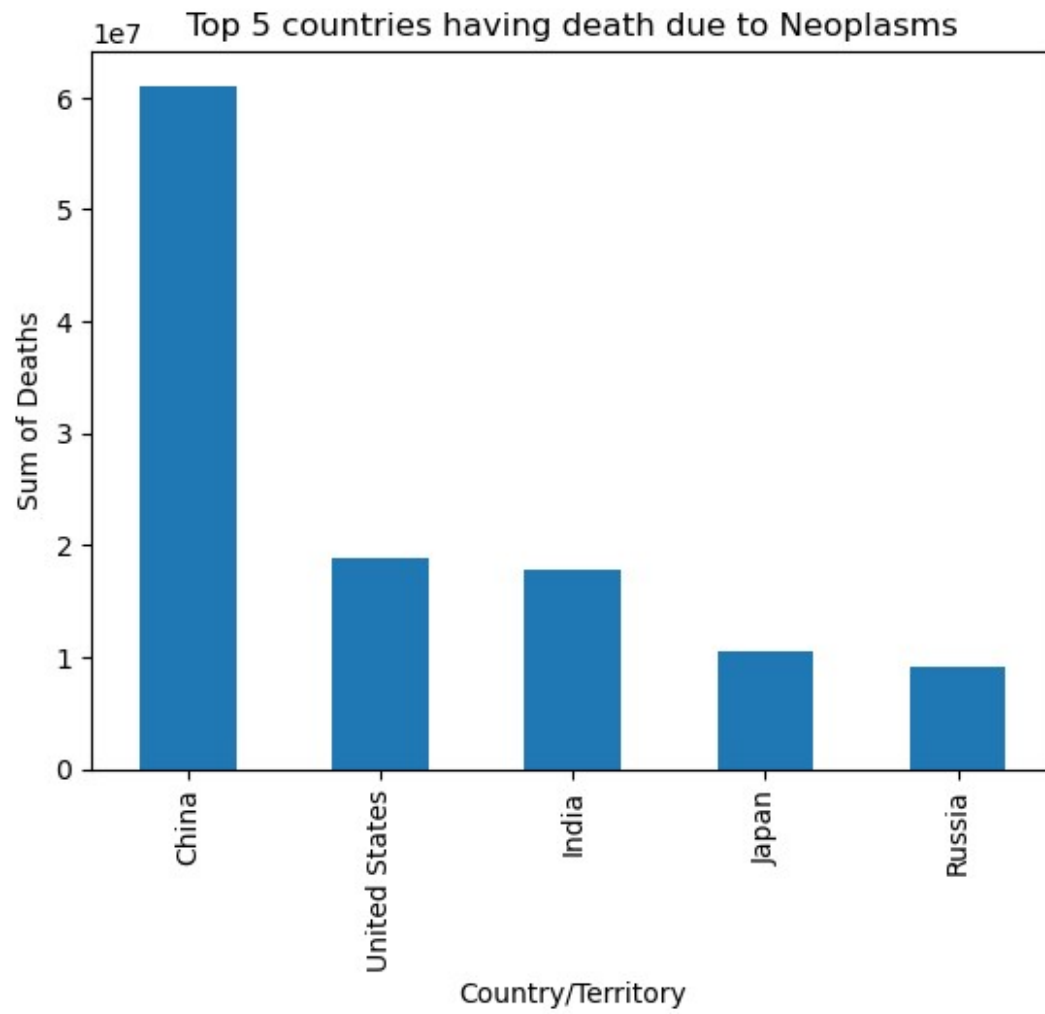
Top 5 countries having death due to Conflict and Terrorism

Top 5 countries having death due to Diabetes Mellitus

Top 5 countries having death due to Chronic Kidney Disease

Top 5 countries having death due to Poisonings
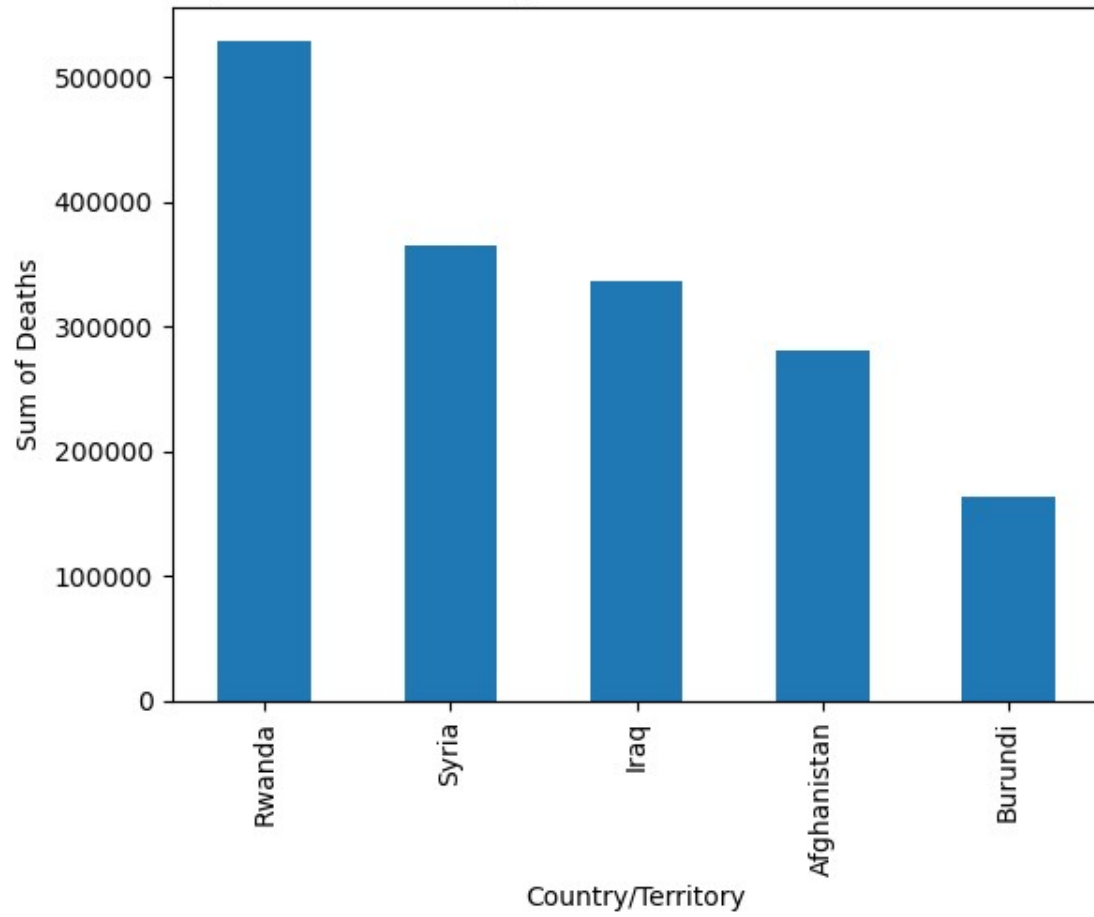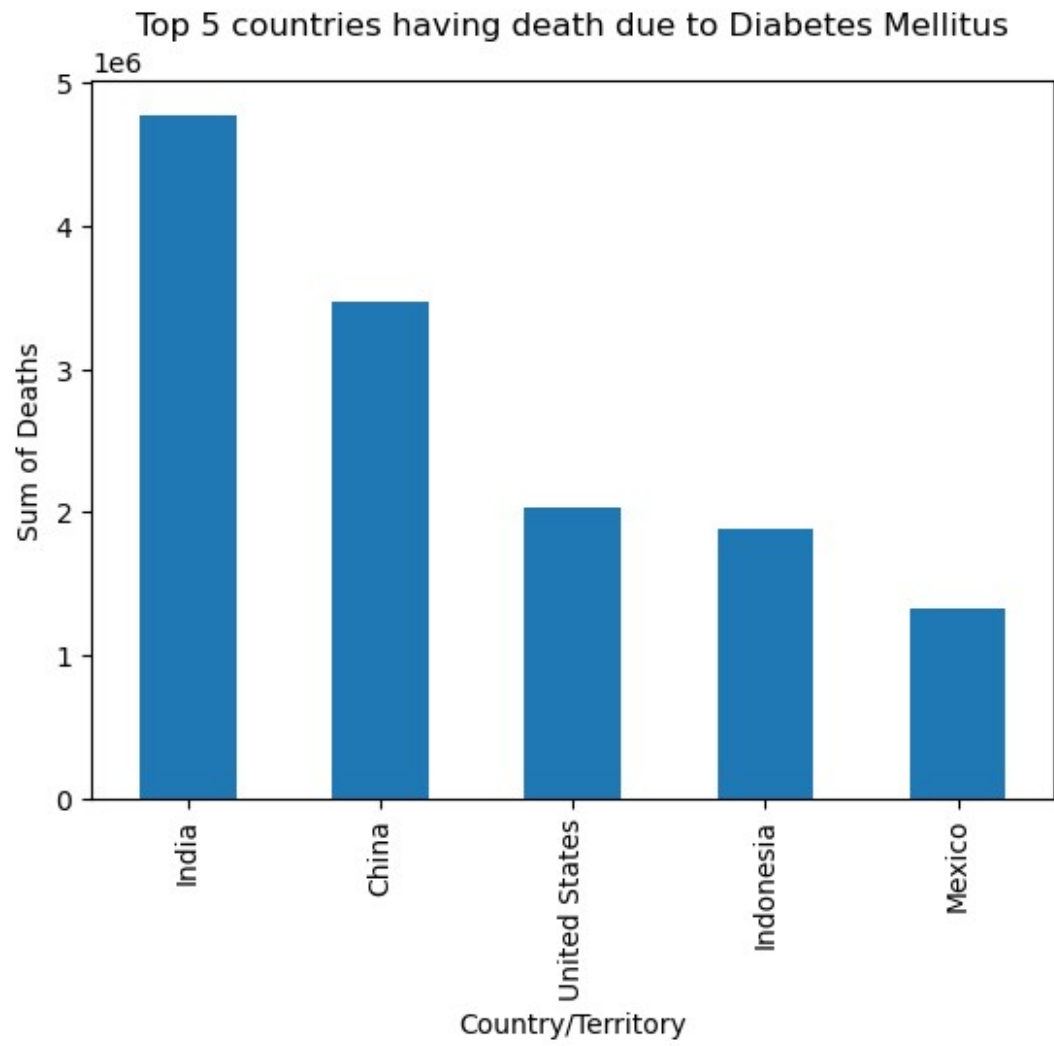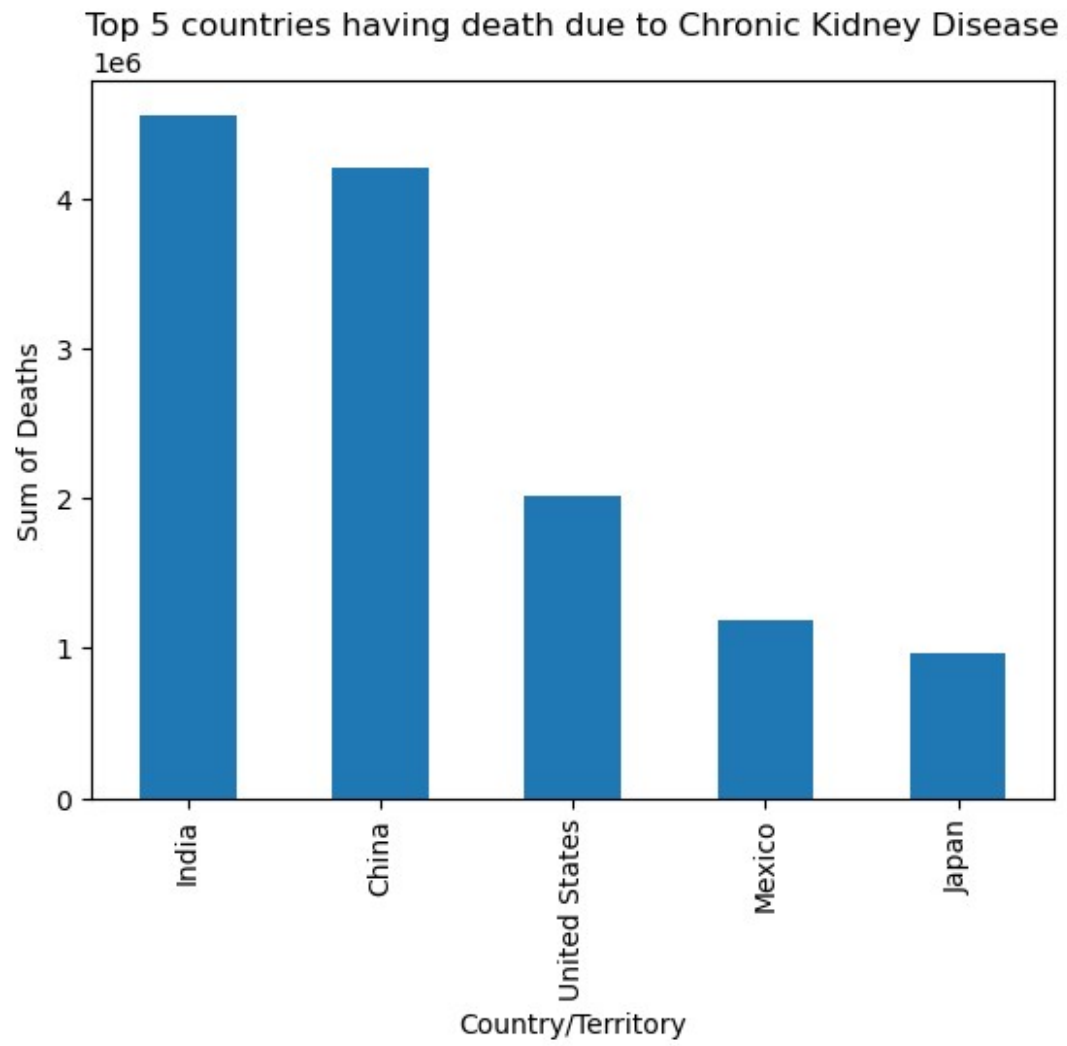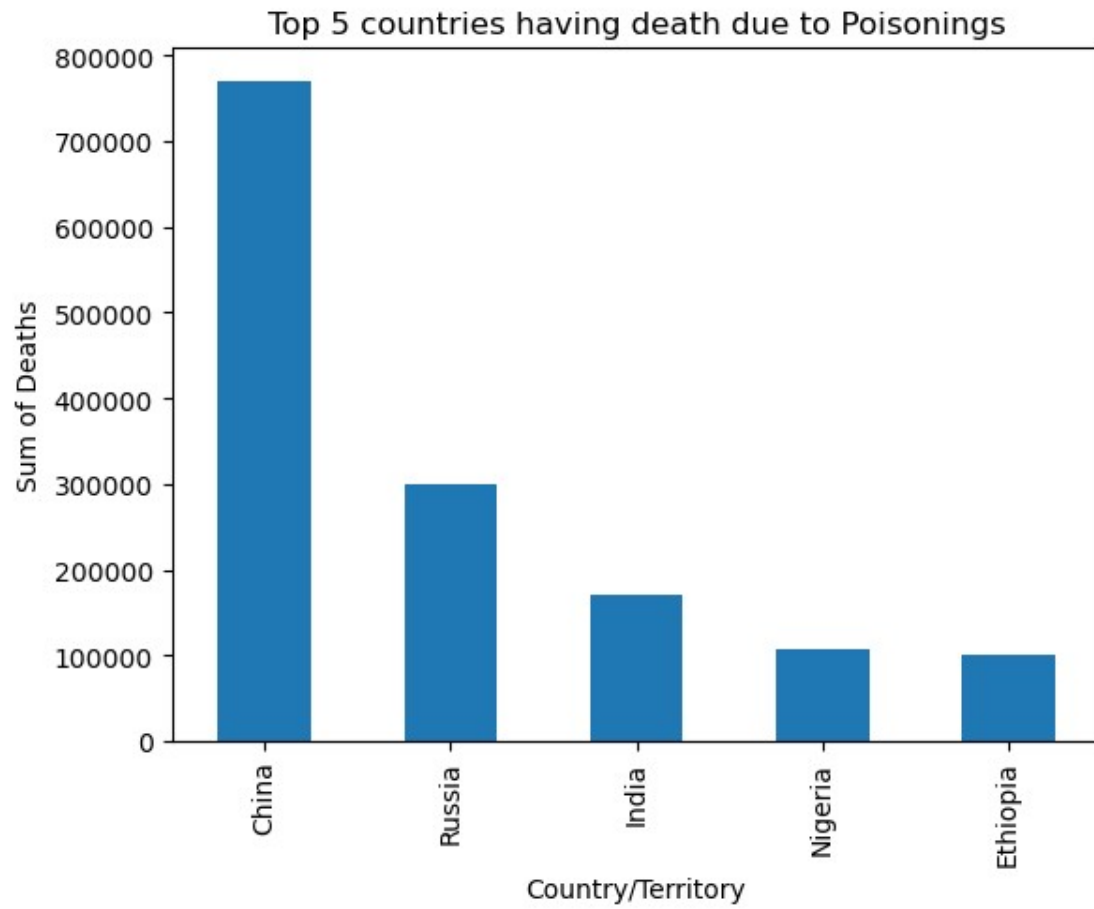
Top 5 countries having death due to Protein-Energy Malnutrition

Top 5 countries having death due to Road Injuries

Top 5 countries having death due to Chronic Respiratory Diseases
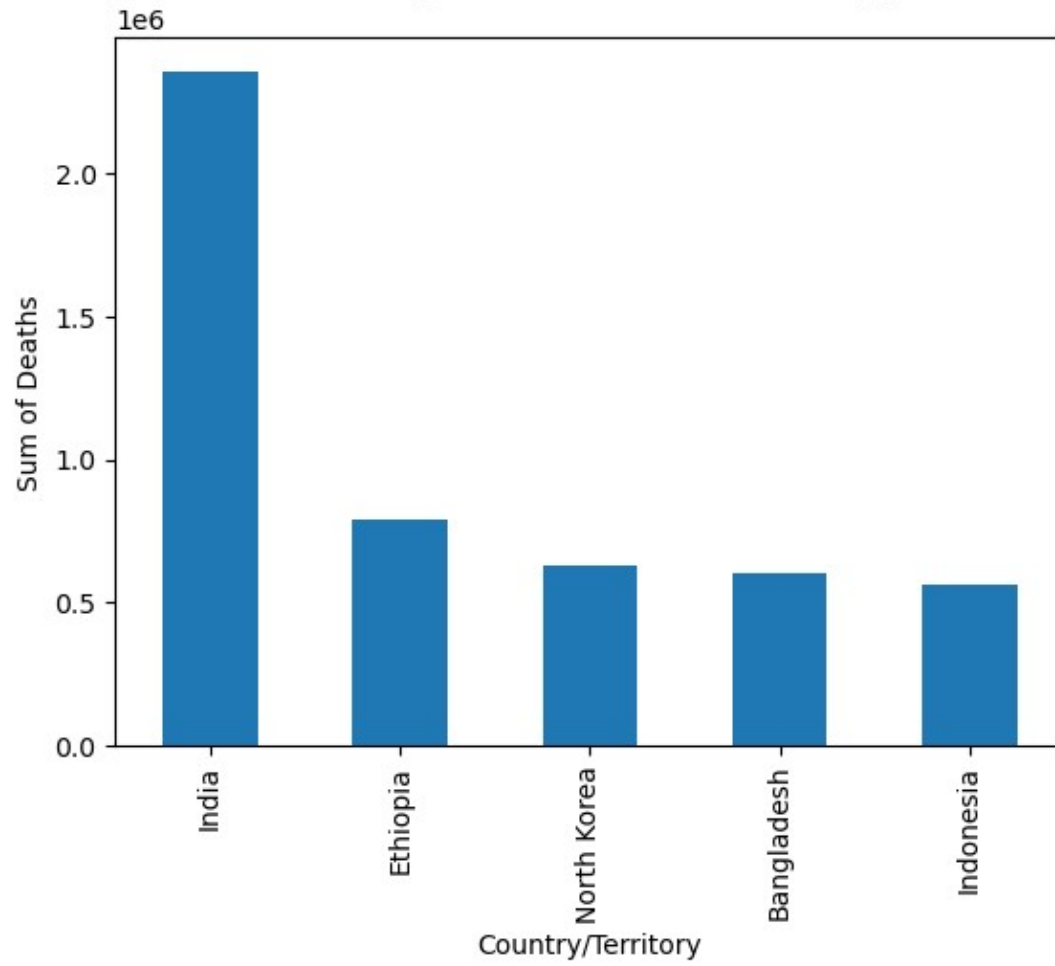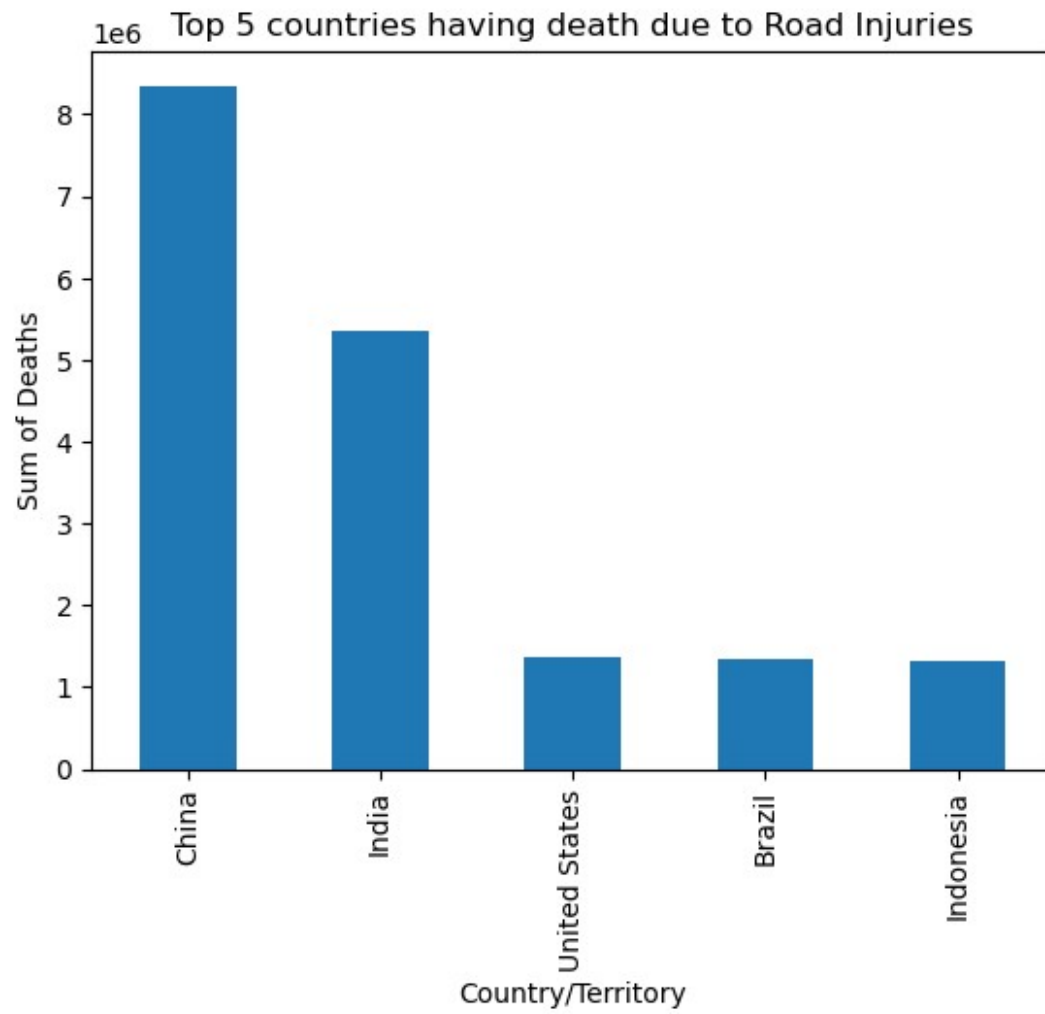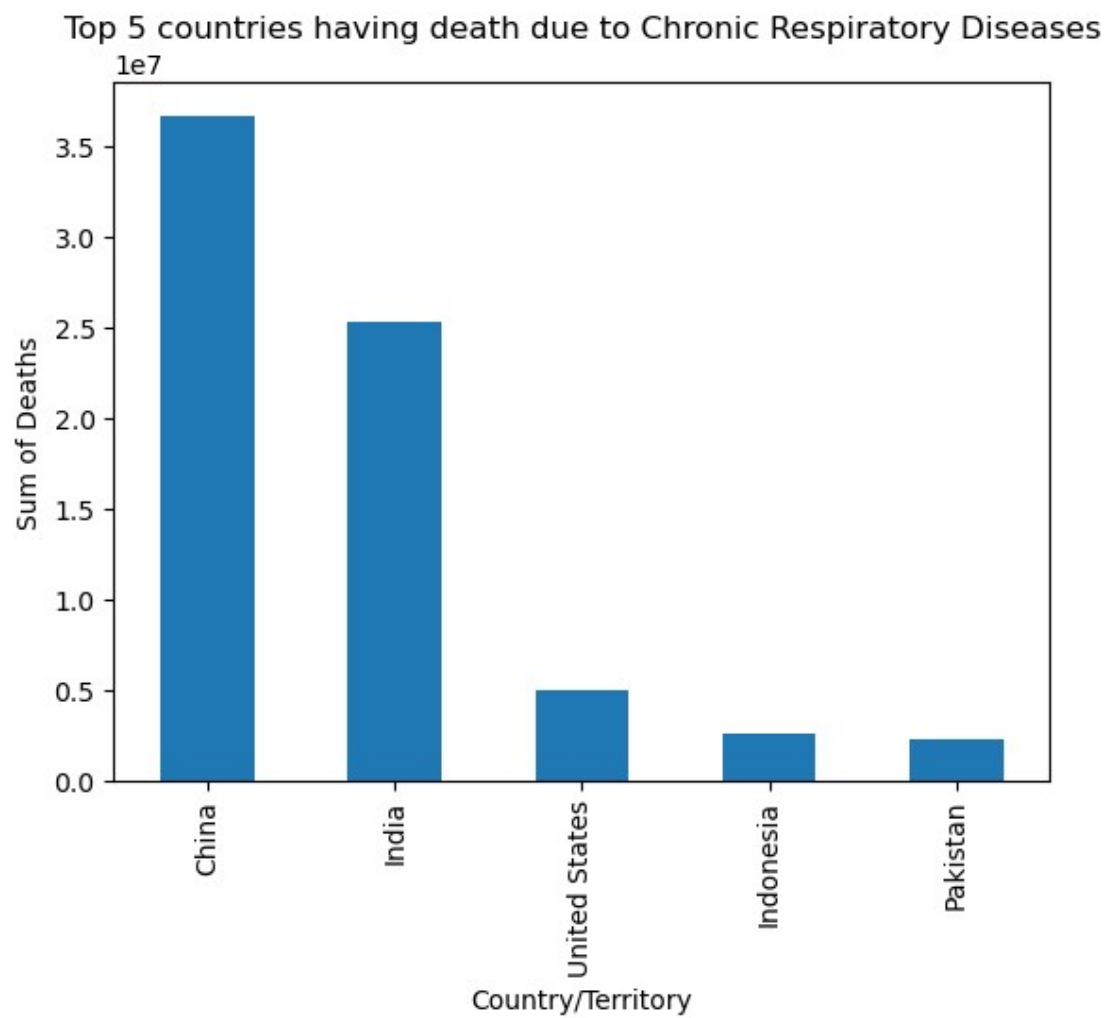
Top 5 countries having death due to Cirrhosis and Other Chronic Liver Diseases

Top 5 countries having death due to Digestive Diseases
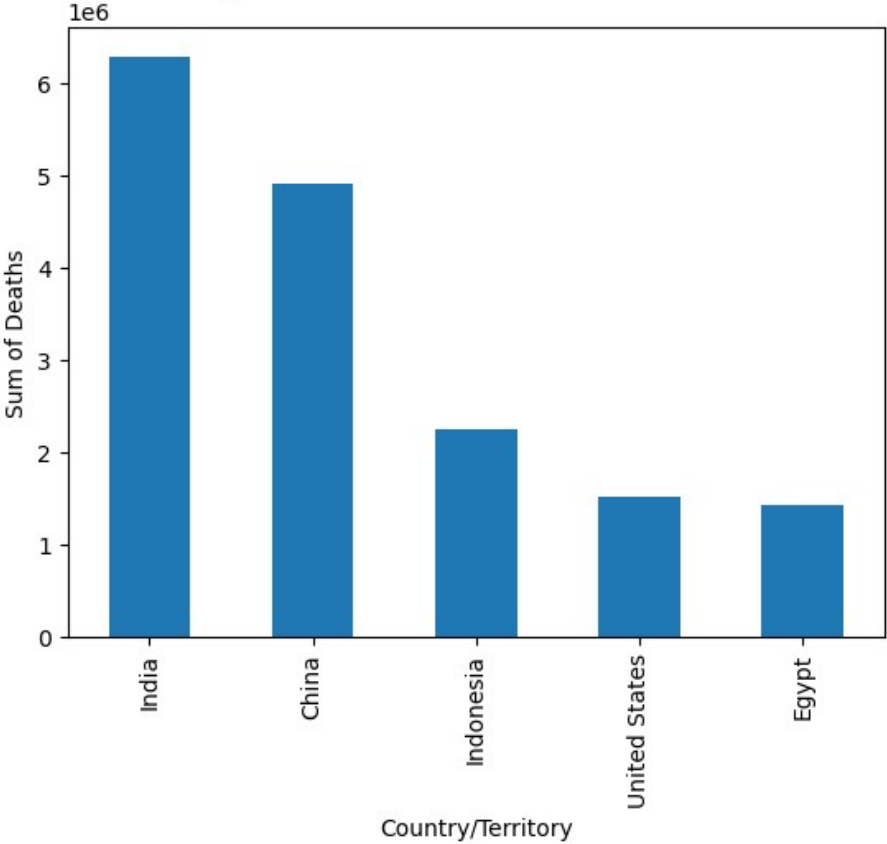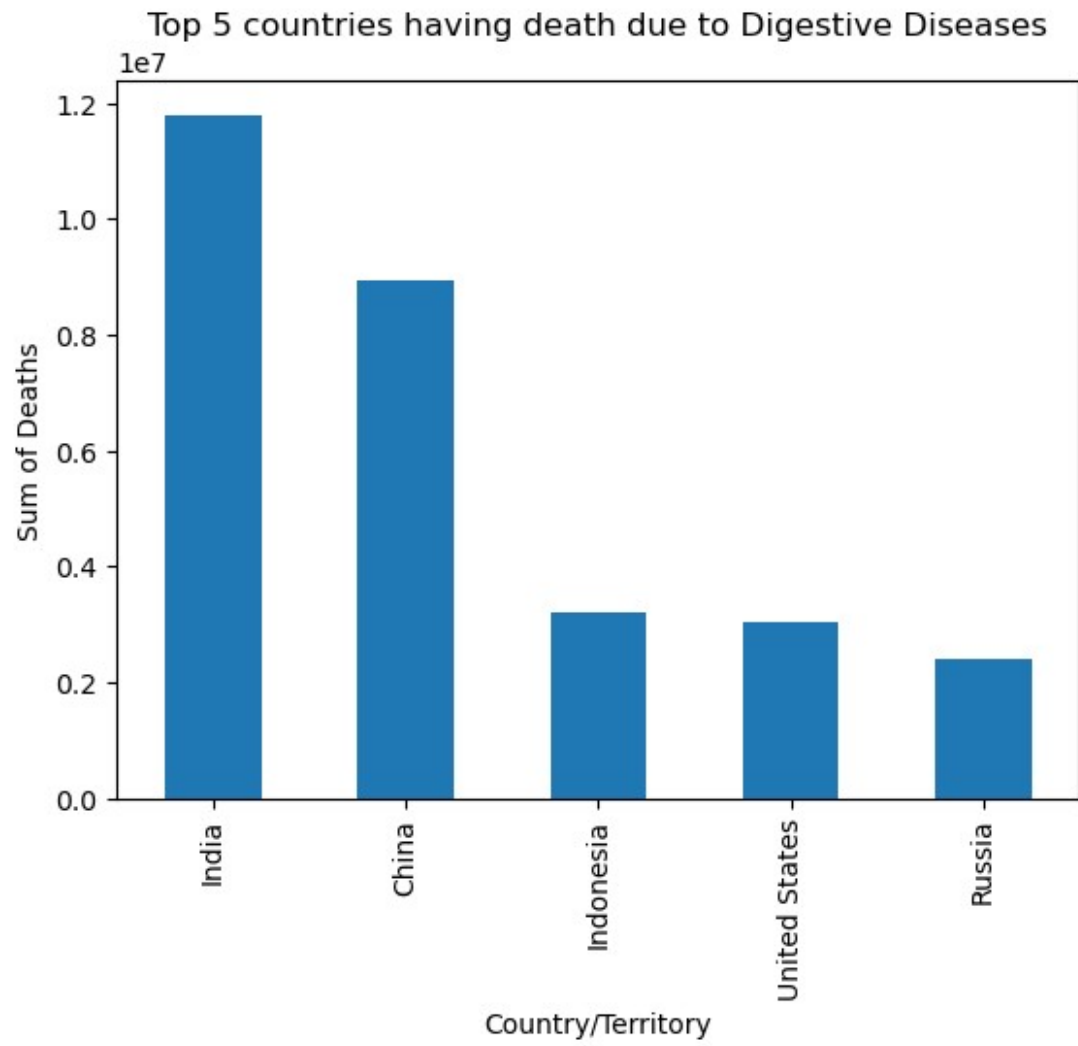
Top 5 countries having death due to Fire, Heat, and Hot Substances

Top 5 countries having death due to Acute Hepatitis
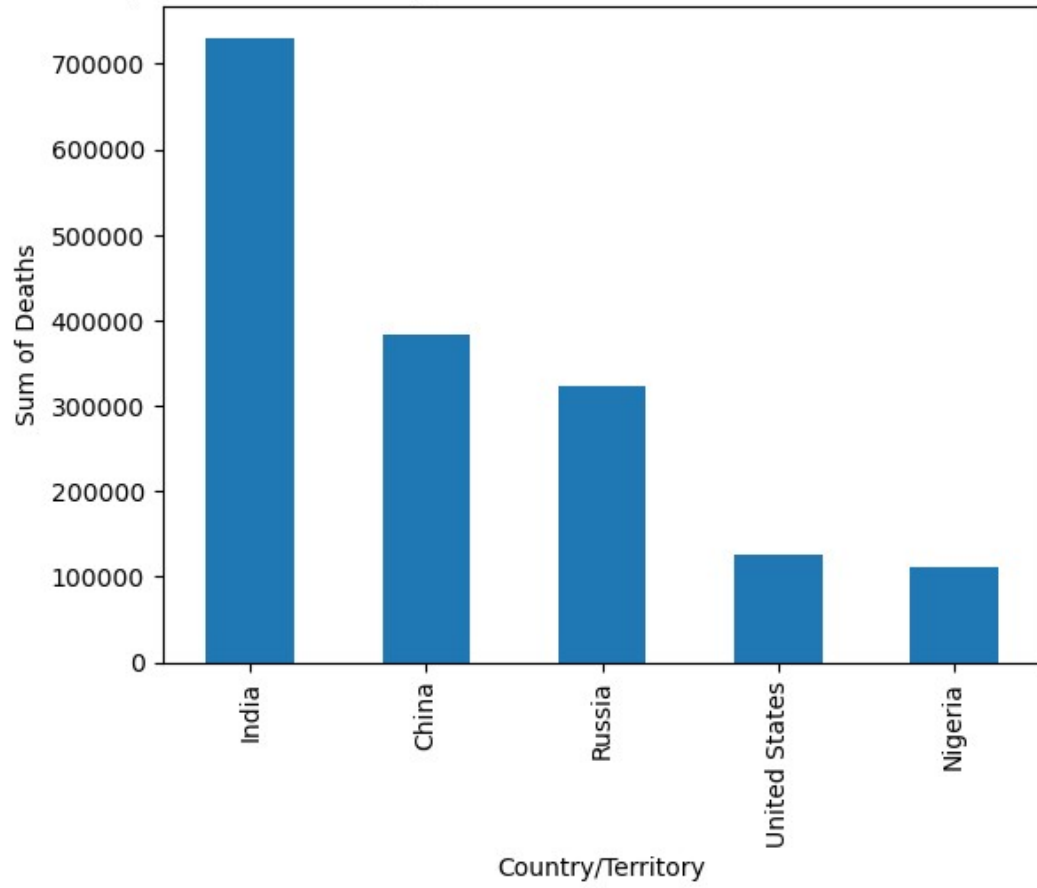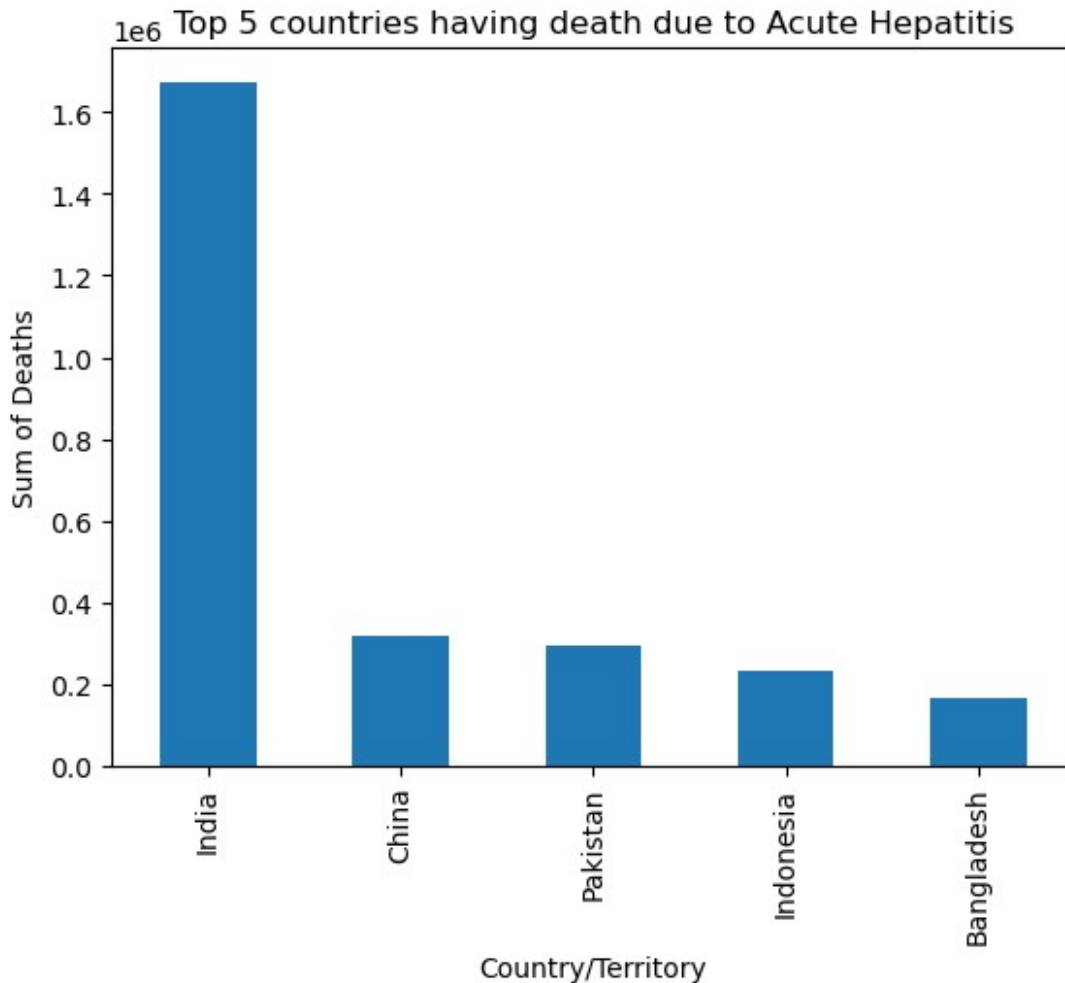

```python
# Top countries in different death disease/enviroment  from above
graphs
# India---- Meningits, Nutritional Deficiencies, Maternal Disorders,
Tuberculosis, Lower Respiratory Infections, Self-harm
# Neonatal Disorders,Diarrheal Diseases, Diabetes Mellitus,Choronic
kidney,Protein-energy malnutrition,Digestive,Acute Hepatitis
# Cirrhosis & other chronic liver, Fire/heat & hot substances

# China---- Alzheimer, Parkinson, Drowning, Cardiovascular Diseases,
Neoplasms, Poisonings, Road injuries, Chronic respiratory
# Nigeria---- Malaria
# Brazil---- Interpersonal Violence
# South Africa---- HIV/AIDS
# USA---- Drug use Disorders
# Russia---- Alcohol use Disorders, Environmental Heat&cold exposure
# Haiti---- Exposure to forces of nature
# Rwanda---- Conflict & Terrorism

df2.groupby('Country/Territory')
['Total_no_of_Deaths'].sum().sort_values(ascending=False).head().plot(
```
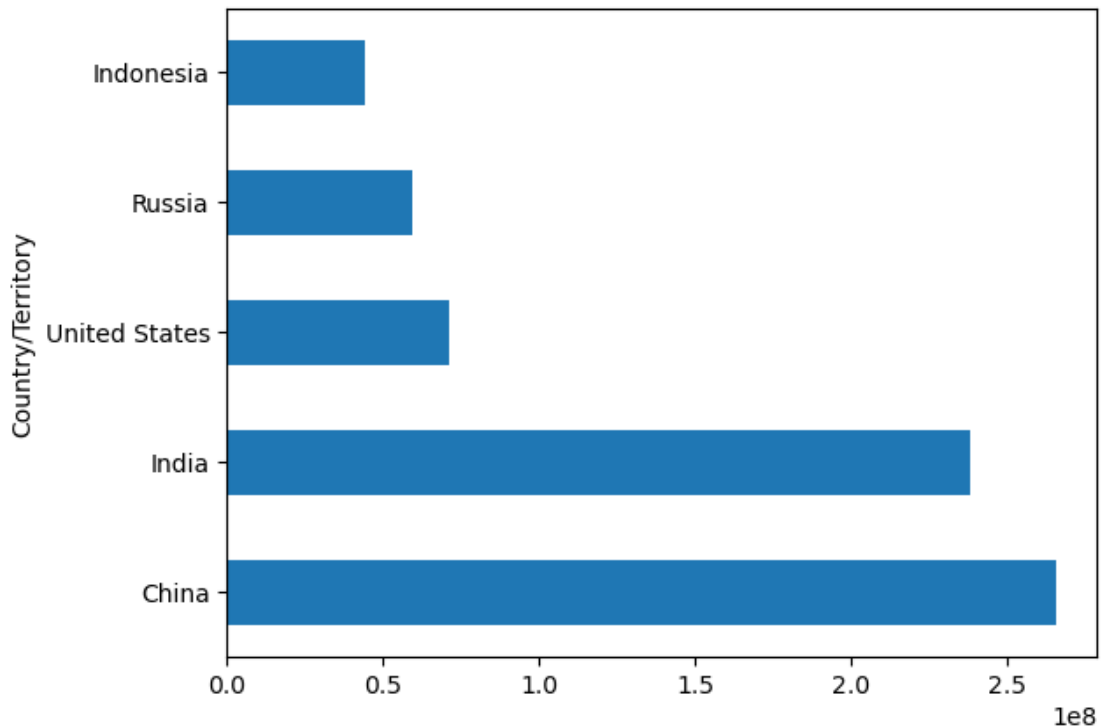
```
kind='barh')
plt.show()
```



```python
# China - "Total_no_of_Deaths" against "Year"

China_Total_no_of_Deaths =
df2[df2['Country/Territory']=='China'].sort_values(by='Total_no_of_Dea
ths',ascending=False)

plt.figure(figsize=(8,4),dpi=200)
sns.scatterplot(data=China_Total_no_of_Deaths, x='Year',
y='Total_no_of_Deaths')
plt.xlabel("Year")
plt.ylabel("Total no.of Deaths")
plt.title("Year Vs. Total no.of Deaths for China")
plt.show();
```

Year Vs. Total no.of Deaths for China