Business Problem

Import Packages

```
import numpy as np
import pandas as pd
from scipy.stats import kstest
import matplotlib.pyplot as plt
%matplotlib inline
import plotly.express as px
from wordcloud import WordCloud
import seaborn as sns
sns.set(style='whitegrid')
import warnings
warnings.filterwarnings('ignore')
from sklearn.preprocessing import PowerTransformer
```

Create UDFs

```
# 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(0.01),
x.quantile(0.05),

x.quantile(0.10),x.quantile(0.25),x.quantile(0.50),x.quantile(0.75),
                                x.quantile(0.90),x.quantile(0.95),
x.quantile(0.99),x.max()],
                  index = ['N', 'NMISS', 'SUM', 'MEAN','MEDIAN',
'STD', 'VAR', 'MIN', 'P1',
                                 'P5' ,'P10' ,'P25' ,'P50' ,'P75' ,'P90'
,'P95' ,'P99' ,'MAX'])
```

Import Data

```
df=pd.read_csv('E:/DataTrained\Intership_Flip Robo/Projects/Cause of
Death/archive/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

      ...  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]
```

```
#  This will enable us to see truncated columns
pd.set_option('display.max_columns',None)
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	
4	8215	44	12914	

	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
```

Data Inspection

```
# Data information
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   int64
 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(32), object(2)
memory usage: 1.6+ MB
```

Data types is correct for the features but i will change year to categorical for better analysis

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

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

```
0
```

There are no duplicated records

```
# Check the null values
df.isna().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
```

```
df.isin(['NAN','NA','N/A','-',' ','?']).sum().any()
```

```
False
```

There are no null values

```
# Check the number of uniques countries
df['Country/Territory'].nunique()
```

```
204
```

```
# Check the number of uniques code
df['Code'].nunique()
```

```
204
```

```
# Drop Code feature as it same as coutries so will be multicollinarity
df.drop('Code',axis=1,inplace=True)

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

# 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
P1	0.000000e+00	0.000000e+00
P5	0.000000e+00	6.000000e+00
P10	2.000000e+00	1.400000e+01
P25	1.500000e+01	9.000000e+01
P50	1.090000e+02	6.665000e+02
P75	8.472500e+02	2.456250e+03
P90	3.499600e+03	8.715100e+03
P95	6.110100e+03	2.038630e+04
P99	3.501950e+04	9.872486e+04
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	
P1	0.000000e+00	0.000000e+00	0.000000e+00	
P5	2.000000e+00	1.000000e+00	0.000000e+00	
P10	5.000000e+00	2.000000e+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	
P90	2.163400e+03	4.917800e+03	8.596400e+03	
P95	4.707150e+03	9.651100e+03	2.097485e+04	
P99	2.188174e+04	3.240173e+04	8.156846e+04	
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	
P1	0.000000e+00	0.000000e+00	0.000000e+00	
P5	2.000000e+00	2.000000e+00	0.000000e+00	
P10	5.000000e+00	5.000000e+00	1.000000e+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	

P90	2.166000e+03	3.156100e+03	2.206300e+03
P95	4.959400e+03	1.127190e+04	3.636150e+03
P99	3.536042e+04	4.035279e+04	1.778007e+04
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	
P1	0.000000e+00	0.000000e+00	0.000000e+00	
P5	1.000000e+00	0.000000e+00	1.000000e+00	
P10	2.000000e+00	1.000000e+00	3.000000e+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	
P90	1.207730e+04	3.922000e+02	1.124430e+04	
P95	3.368610e+04	9.332000e+02	2.710690e+04	
P99	1.046697e+05	9.163350e+03	1.041743e+05	
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	
P1	2.300000e+01	4.000000e+00	
P5	1.130000e+02	1.400000e+01	
P10	2.420000e+02	3.100000e+01	
P25	2.028000e+03	3.450000e+02	
P50	1.174200e+04	2.126500e+03	
P75	4.254650e+04	1.016125e+04	
P90	1.307558e+05	2.412480e+04	
P95	2.520974e+05	5.346200e+04	
P99	1.273302e+06	1.981039e+05	
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
P1	0.000000e+00	0.000000e+00	1.000000e+00
P5	4.000000e+00	1.000000e+00	3.000000e+00
P10	1.300000e+01	2.000000e+00	8.000000e+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
P90	2.222670e+04	8.880000e+02	5.437500e+03
P95	3.661695e+04	2.891400e+03	1.061205e+04
P99	2.290906e+05	1.686733e+04	7.900556e+04
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	
P1	0.000000e+00	0.000000e+00	
P5	0.000000e+00	1.000000e+00	
P10	0.000000e+00	3.000000e+00	
P25	0.000000e+00	2.000000e+01	
P50	0.000000e+00	2.965000e+02	
P75	1.200000e+01	3.946750e+03	
P90	8.410000e+01	1.768400e+04	
P95	2.320000e+02	2.870115e+04	
P99	2.062340e+03	1.328513e+05	
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	
P1	0.000000e+00	7.000000e+00	
P5	0.000000e+00	7.600000e+01	
P10	0.000000e+00	1.290000e+02	
P25	2.000000e+00	8.097500e+02	
P50	2.100000e+01	5.629500e+03	
P75	1.090000e+02	2.014775e+04	

P90	2.451000e+02	6.457360e+04
P95	6.152000e+02	1.527135e+05
P99	9.935770e+03	6.022461e+05
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
P1	0.000000e+00	3.000000e+00	1.000000e+00
P5	0.000000e+00	1.400000e+01	1.000000e+01
P10	0.000000e+00	4.000000e+01	2.800000e+01
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
P90	4.782000e+02	1.069840e+04	9.375600e+03
P95	1.779000e+03	1.957225e+04	1.680650e+04
P99	1.041000e+04	7.897854e+04	9.882408e+04
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
P1	0.000000e+00	0.000000e+00	2.000000e+00
P5	0.000000e+00	0.000000e+00	6.000000e+00
P10	1.000000e+00	1.000000e+00	1.400000e+01
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
P90	5.951000e+02	4.583000e+03	1.016810e+04
P95	1.208050e+03	8.805100e+03	2.115870e+04
P99	6.599880e+03	2.701049e+04	5.023466e+04
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
P1	3.000000e+00
P5	1.400000e+01
P10	2.800000e+01
P25	2.890000e+02
P50	1.689000e+03
P75	5.249750e+03
P90	2.194620e+04
P95	3.987850e+04
P99	2.171634e+05
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
P1	2.000000e+00	2.000000e+00
P5	8.000000e+00	1.600000e+01
P10	1.800000e+01	3.300000e+01
P25	1.540000e+02	2.840000e+02
P50	1.210000e+03	2.185000e+03
P75	3.547250e+03	6.080000e+03
P90	1.194180e+04	2.274880e+04
P95	2.595360e+04	4.460395e+04
P99	8.720563e+04	1.269085e+05
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
P1	0.000000e+00	0.000000e+00
P5	0.000000e+00	0.000000e+00
P10	2.000000e+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
P90	1.005300e+03	5.902000e+02
P95	2.061300e+03	1.569050e+03
P99	1.136029e+04	9.776920e+03
MAX	2.587600e+04	6.430500e+04

It seems alot of outliers or skewness are in the data

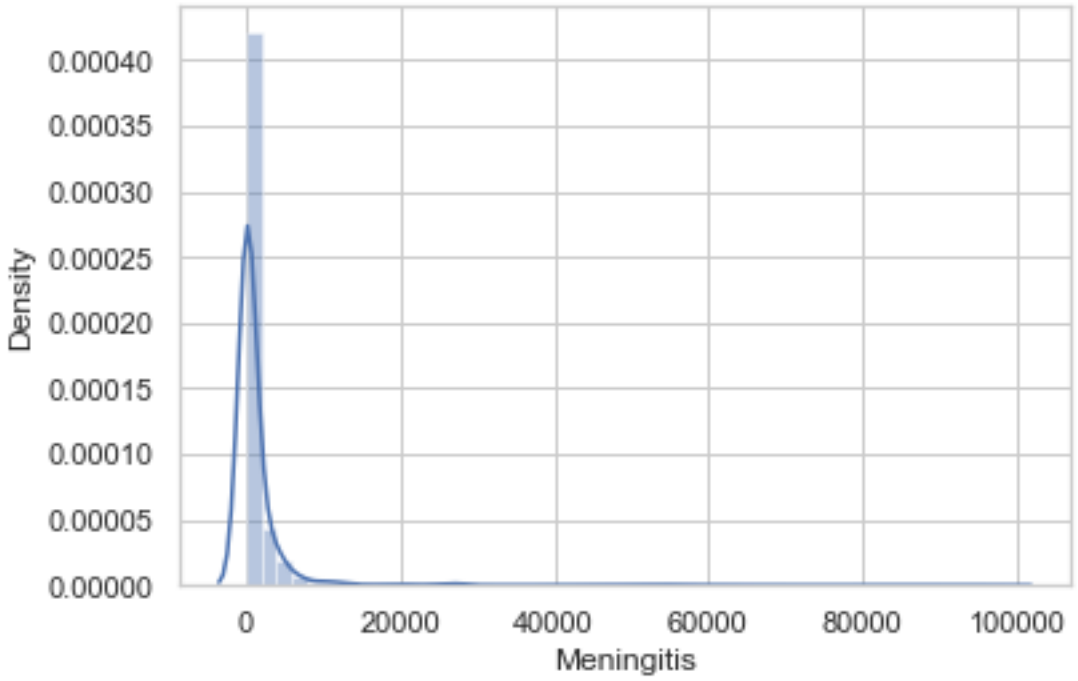
```
df1_cat.describe(include=['O'])
```

	Country/Territory	Year
count	6120	6120

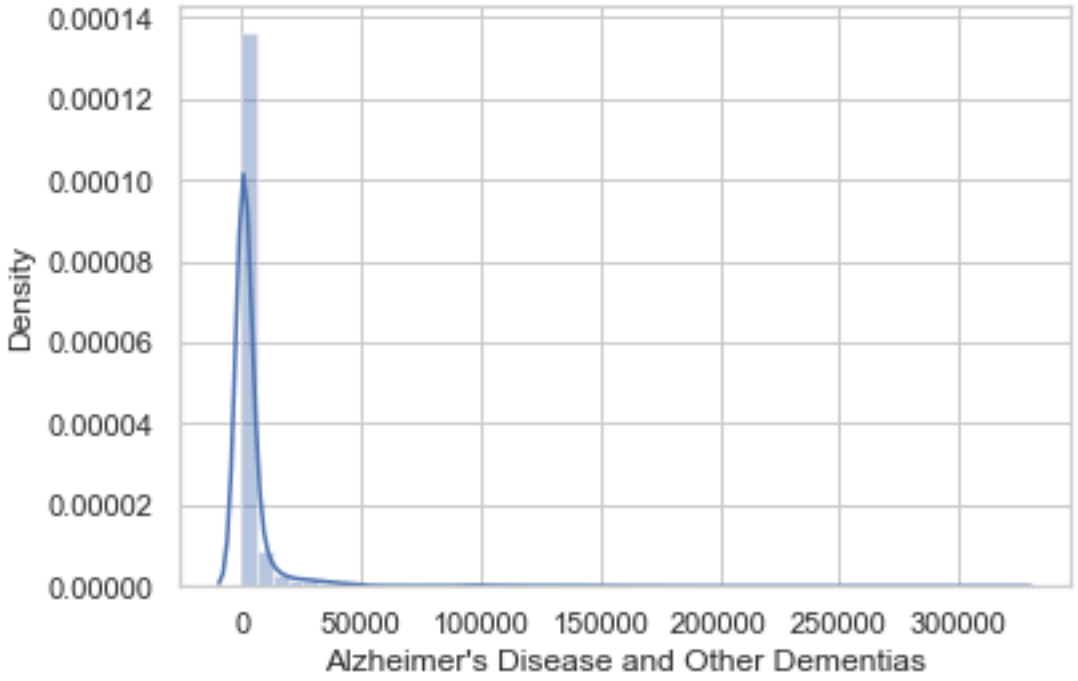
```
unique               204    30
top          Afghanistan  1990
freq                  30   204
```

```
# 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()
```

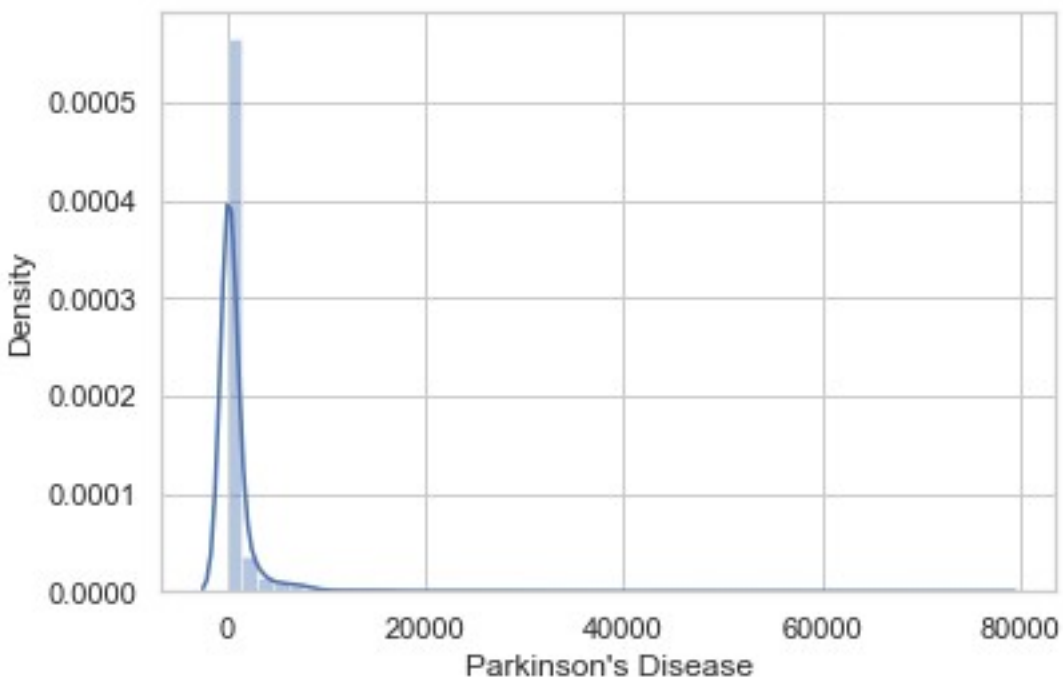
```
KstestResult(statistic=0.8829688223001868, pvalue=0.0)
```



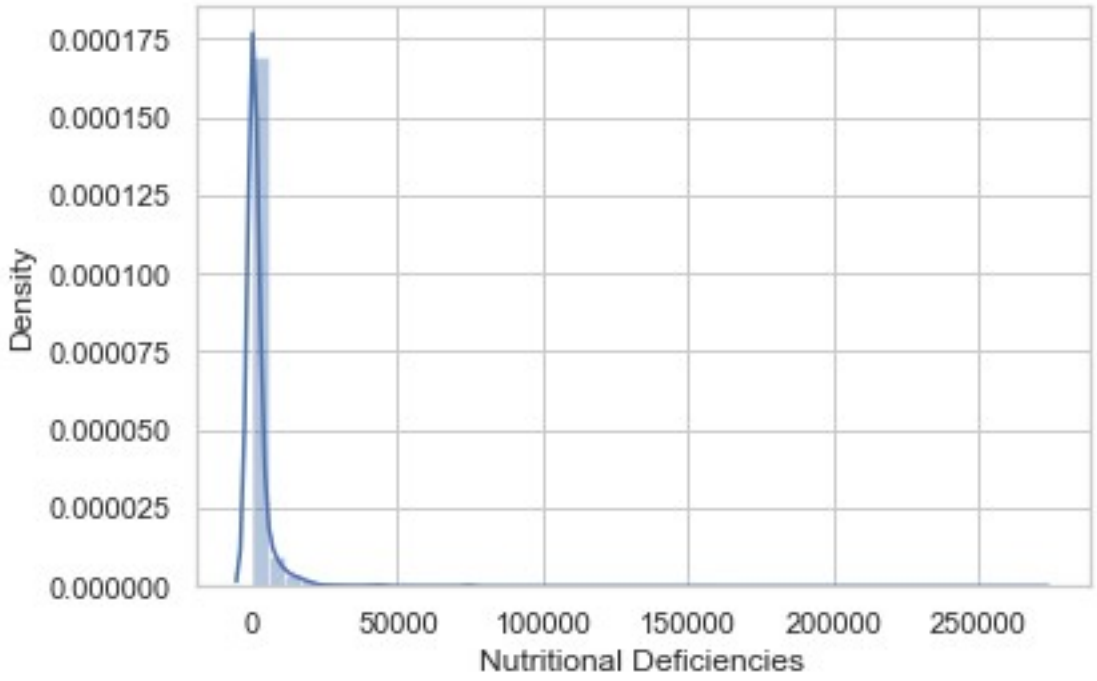

```
KstestResult(statistic=0.9661337620990889, pvalue=0.0)
```



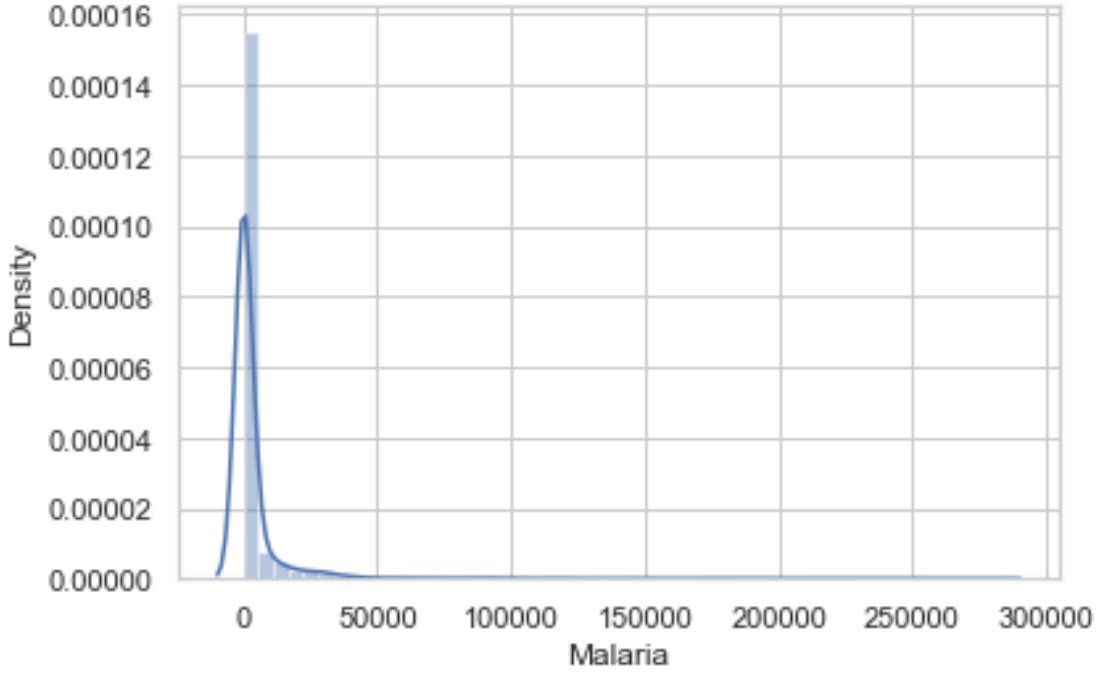

```
KstestResult(statistic=0.9471795137330757, pvalue=0.0)
```



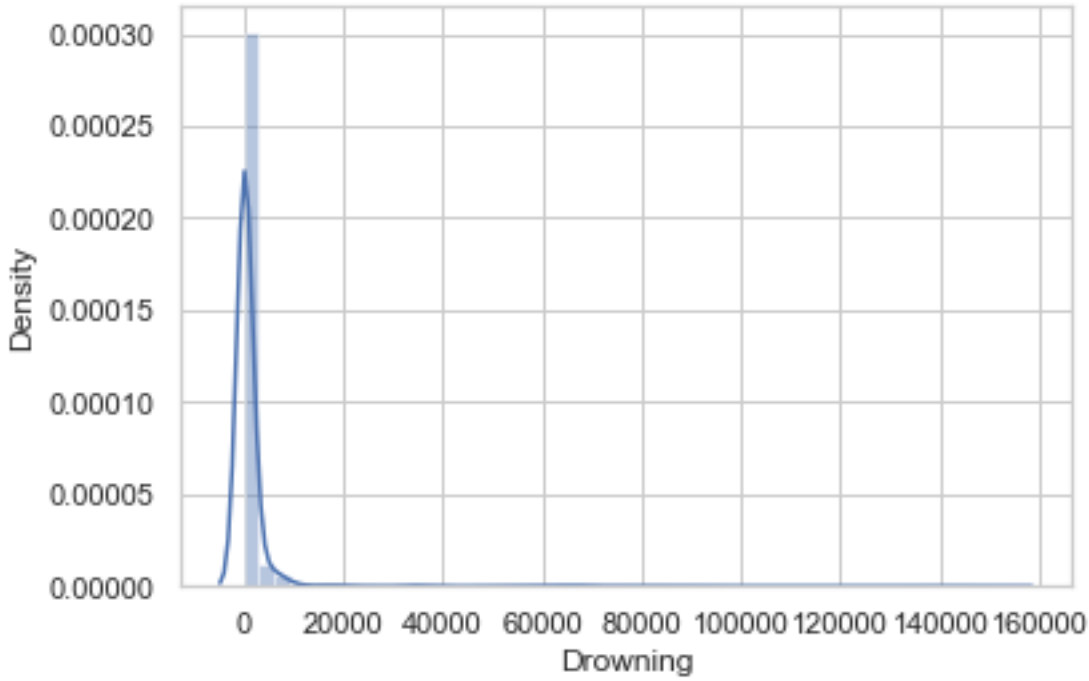

```
KstestResult(statistic=0.8811714366792718, pvalue=0.0)
```



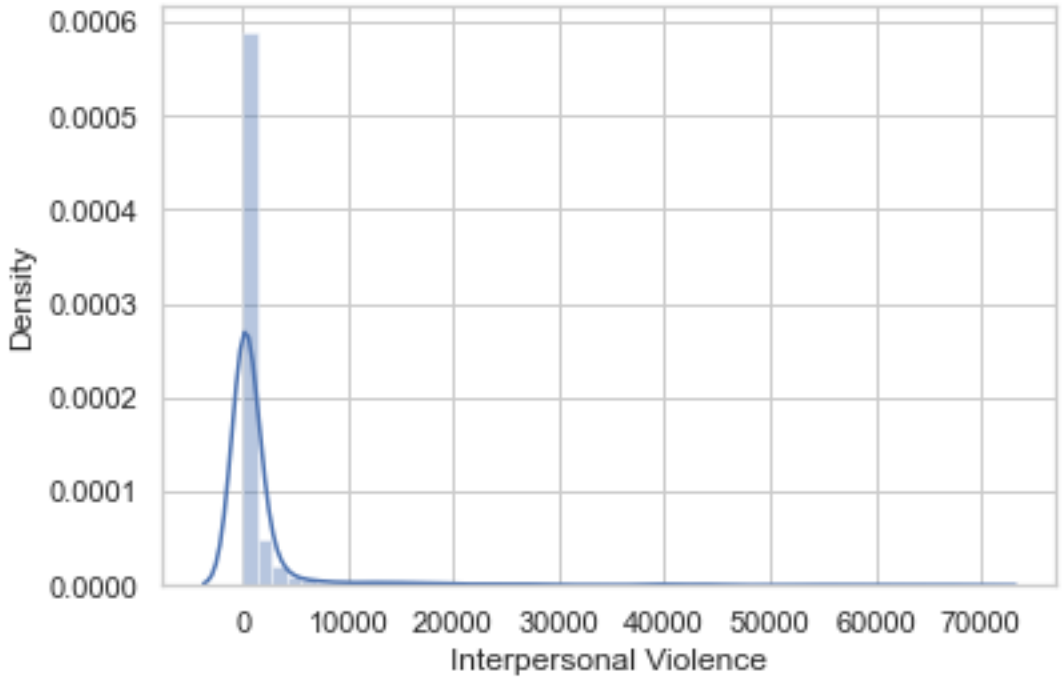

```
KstestResult(statistic=0.5, pvalue=0.0)
```



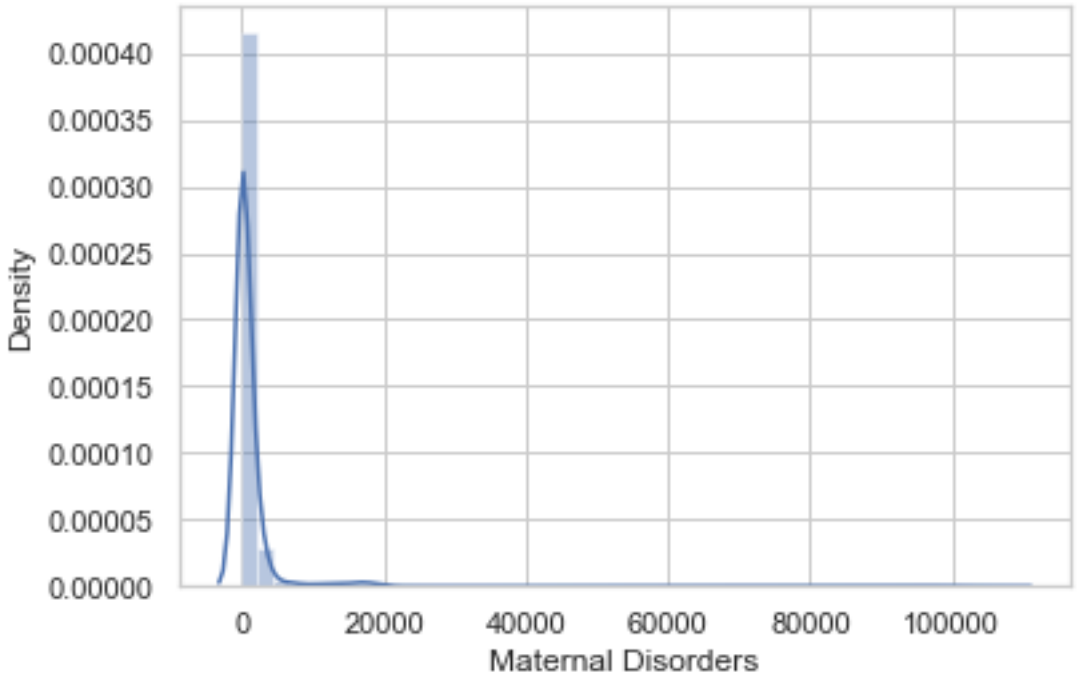

```
KstestResult(statistic=0.9483233045827489, pvalue=0.0)
```



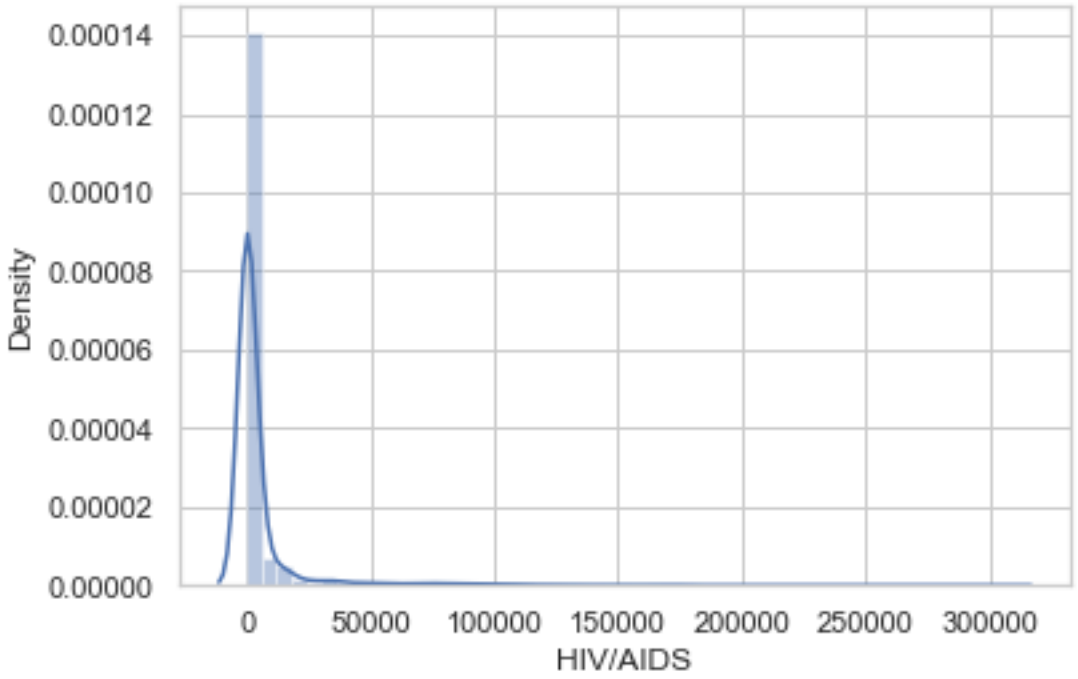

```
KstestResult(statistic=0.9432579451056248, pvalue=0.0)
```



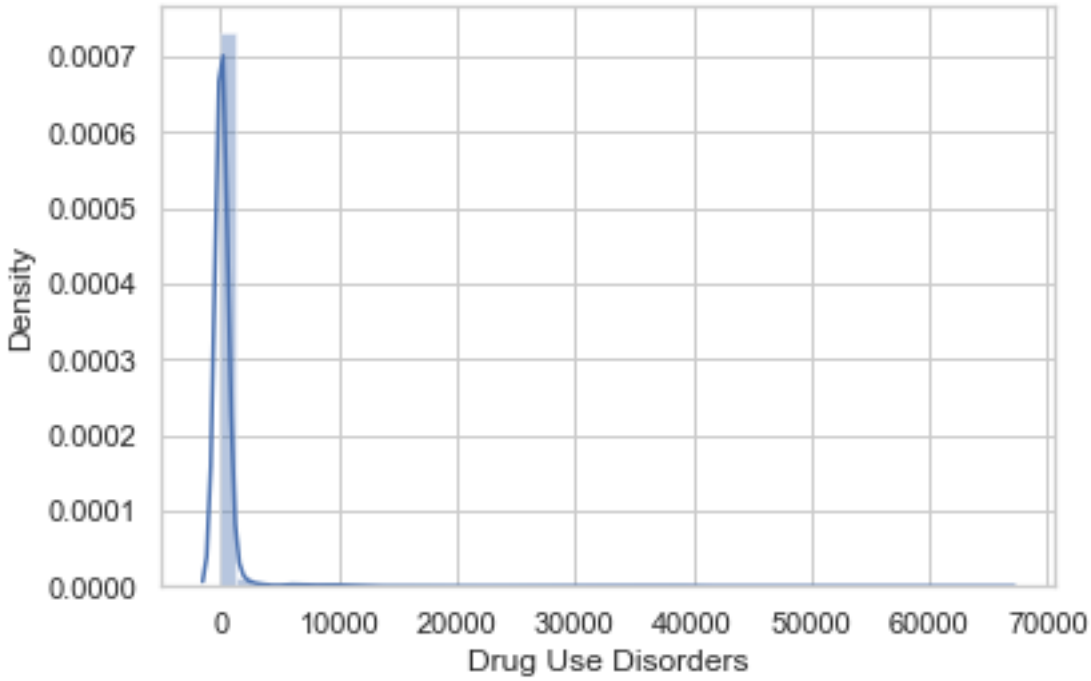

```
KstestResult(statistic=0.8373805870060691, pvalue=0.0)
```



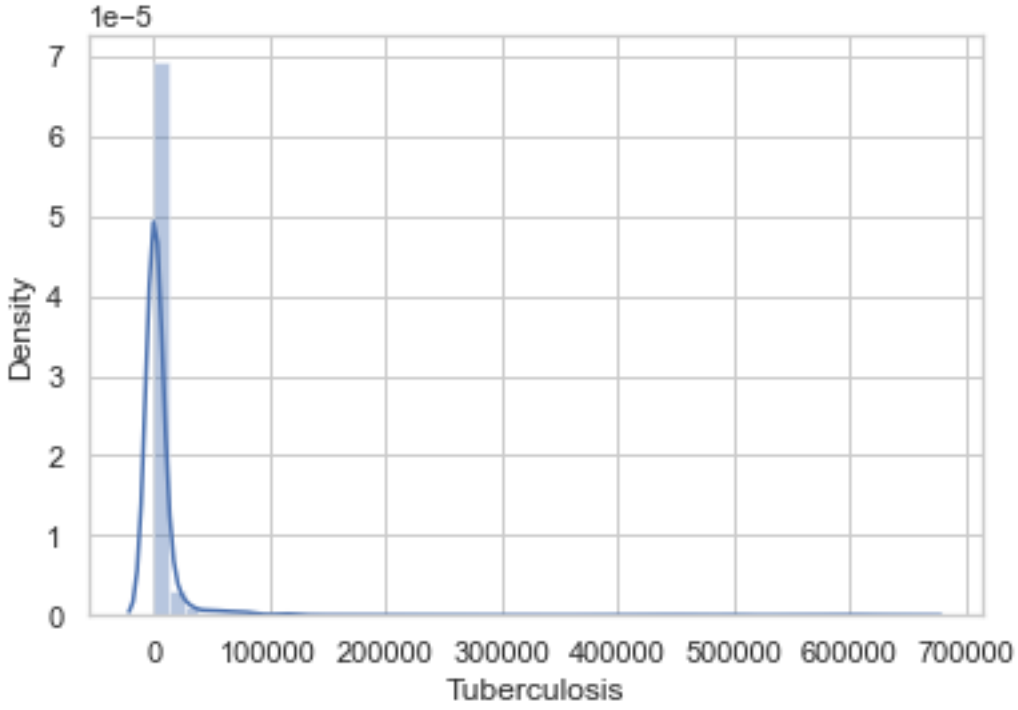

```
KstestResult(statistic=0.8888511752413633, pvalue=0.0)
```



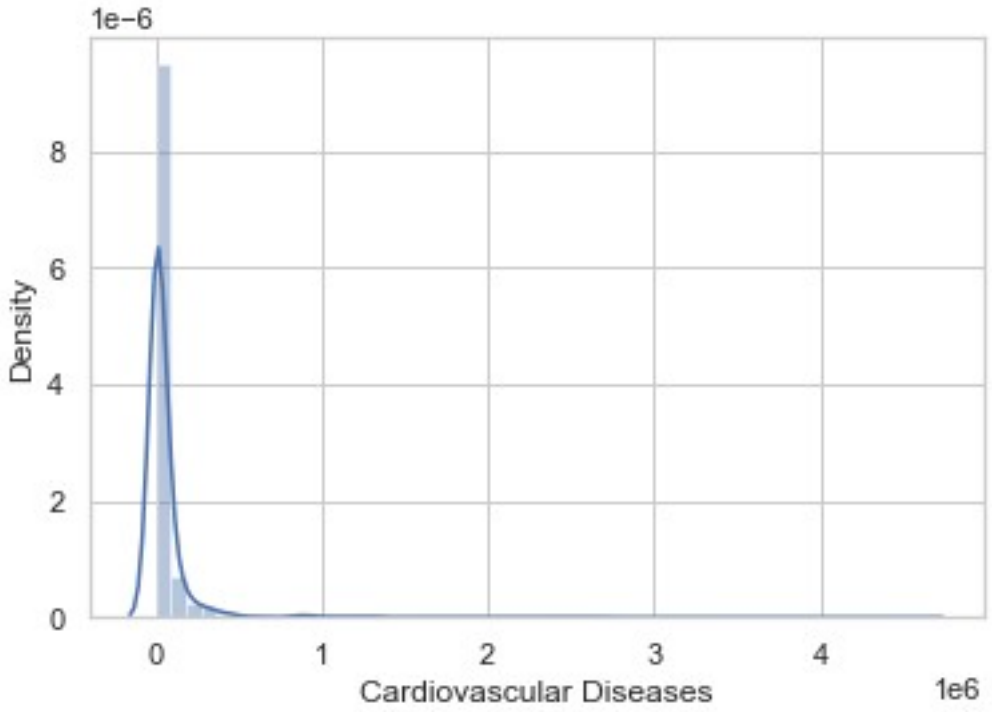

```
KstestResult(statistic=0.7891779726269842, pvalue=0.0)
```



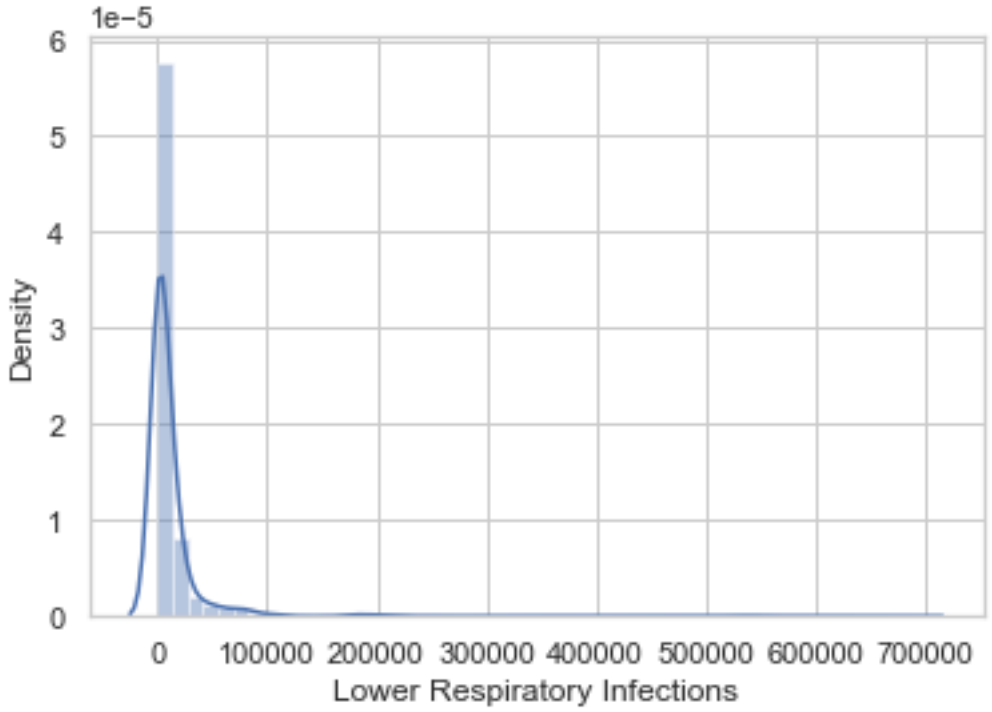

```
KstestResult(statistic=0.9063298405304614, pvalue=0.0)
```



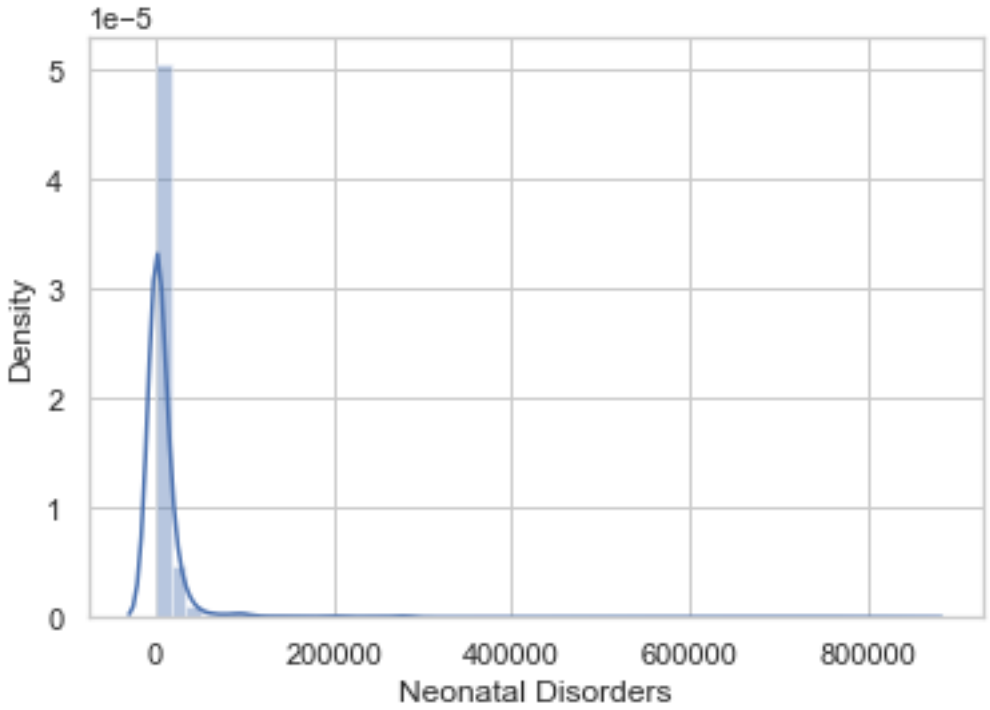

```
KstestResult(statistic=0.9999683287581669, pvalue=0.0)
```



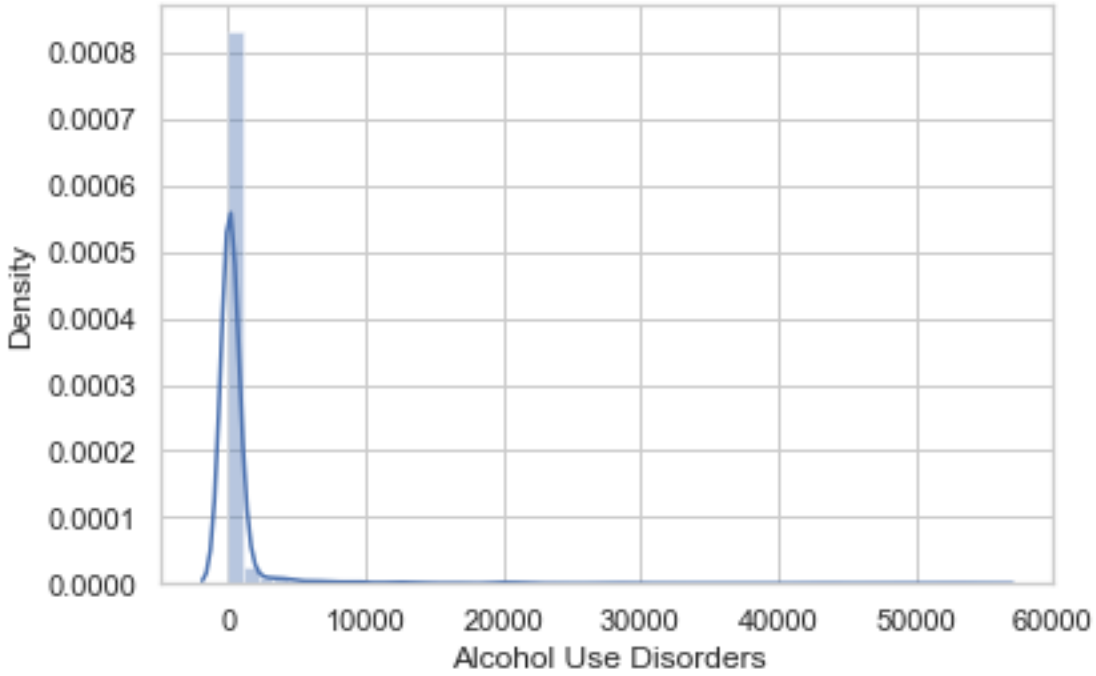

```
KstestResult(statistic=0.9901644071895395, pvalue=0.0)
```



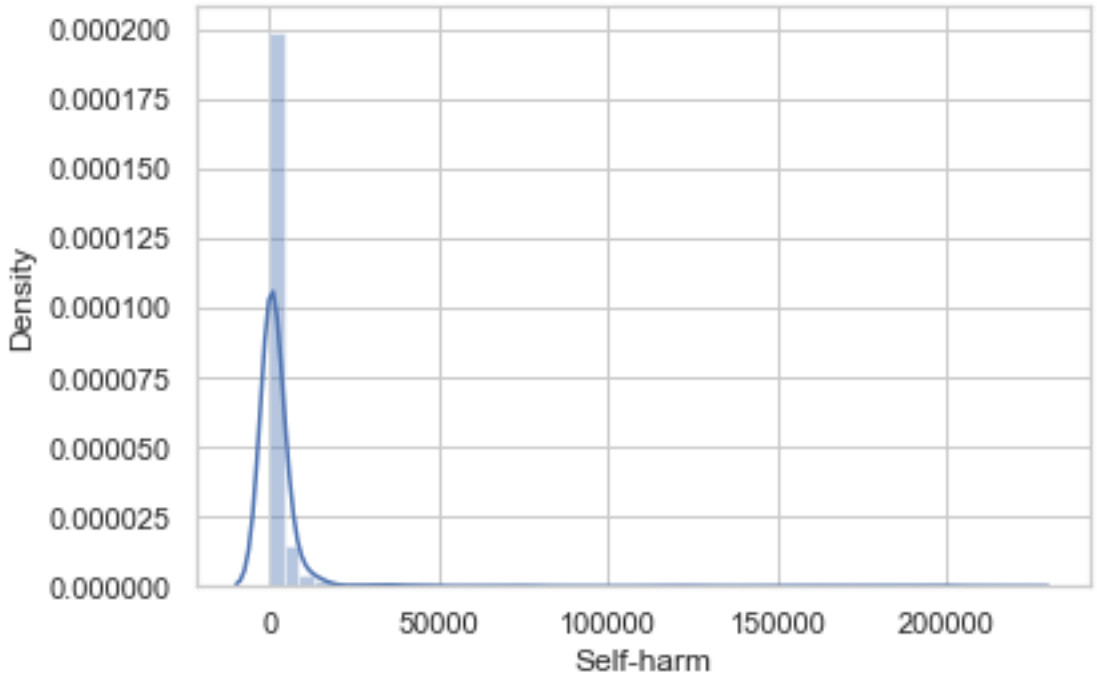

```
KstestResult(statistic=0.9586174222298078, pvalue=0.0)
```



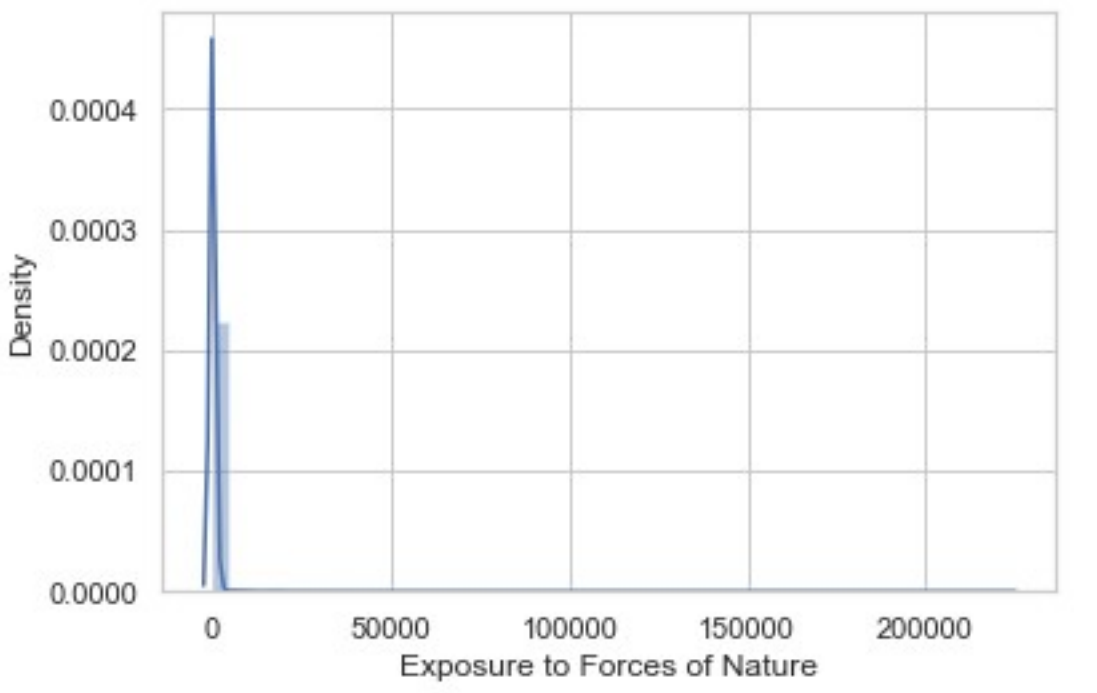

```
KstestResult(statistic=0.8803544432152195, pvalue=0.0)
```



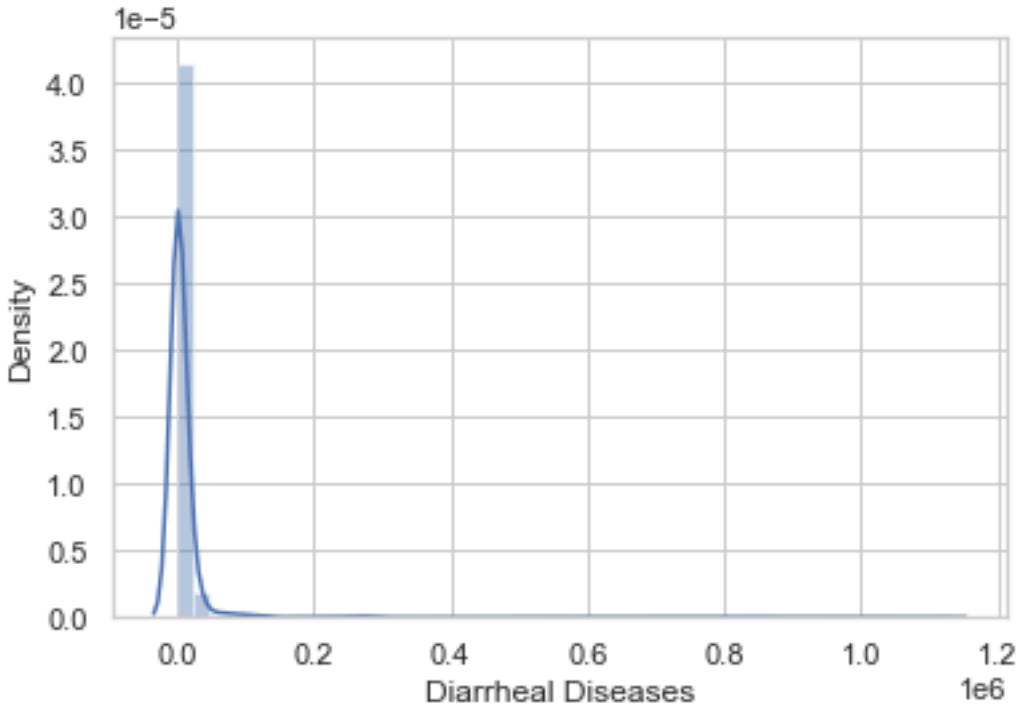

```
KstestResult(statistic=0.967604350334383, pvalue=0.0)
```



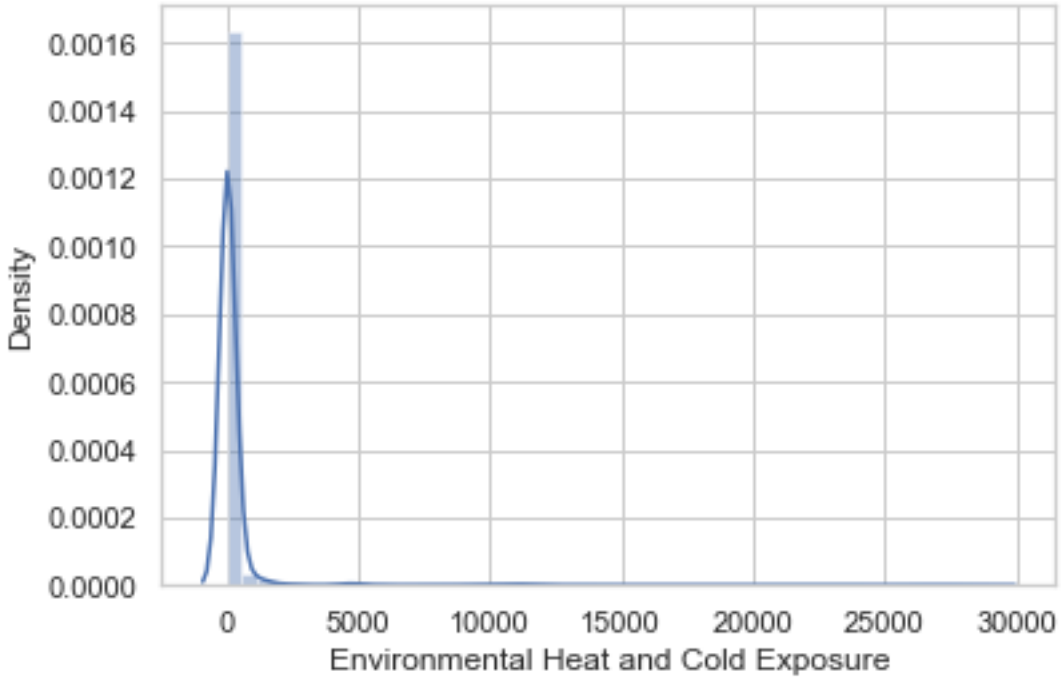

```
KstestResult(statistic=0.5, pvalue=0.0)
```



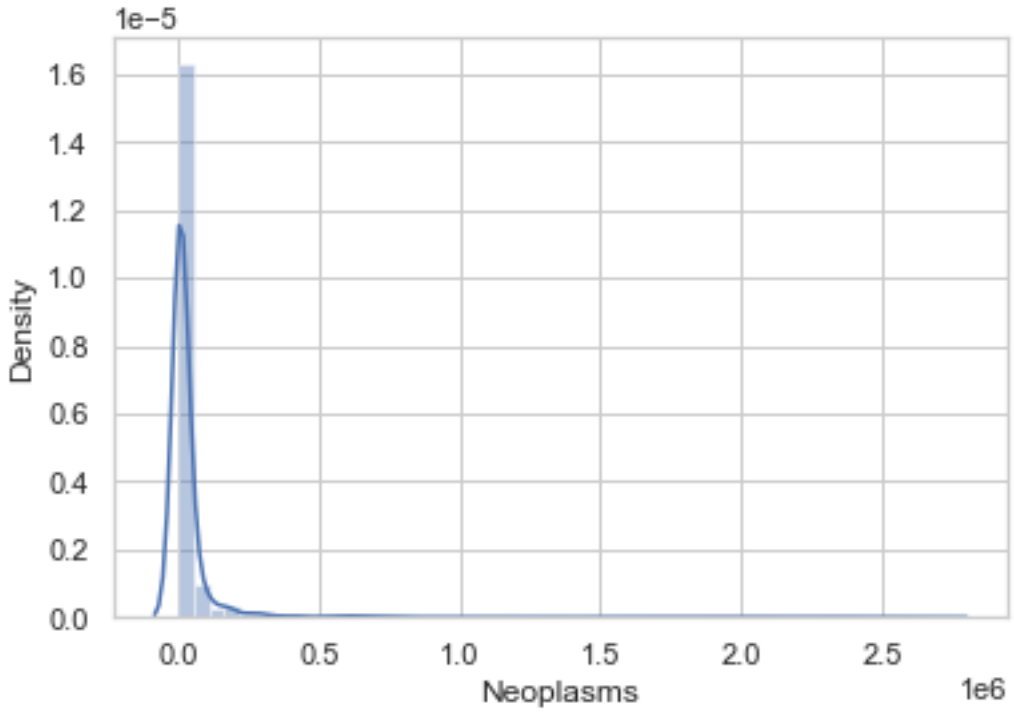

```
KstestResult(statistic=0.9086224170714287, pvalue=0.0)
```



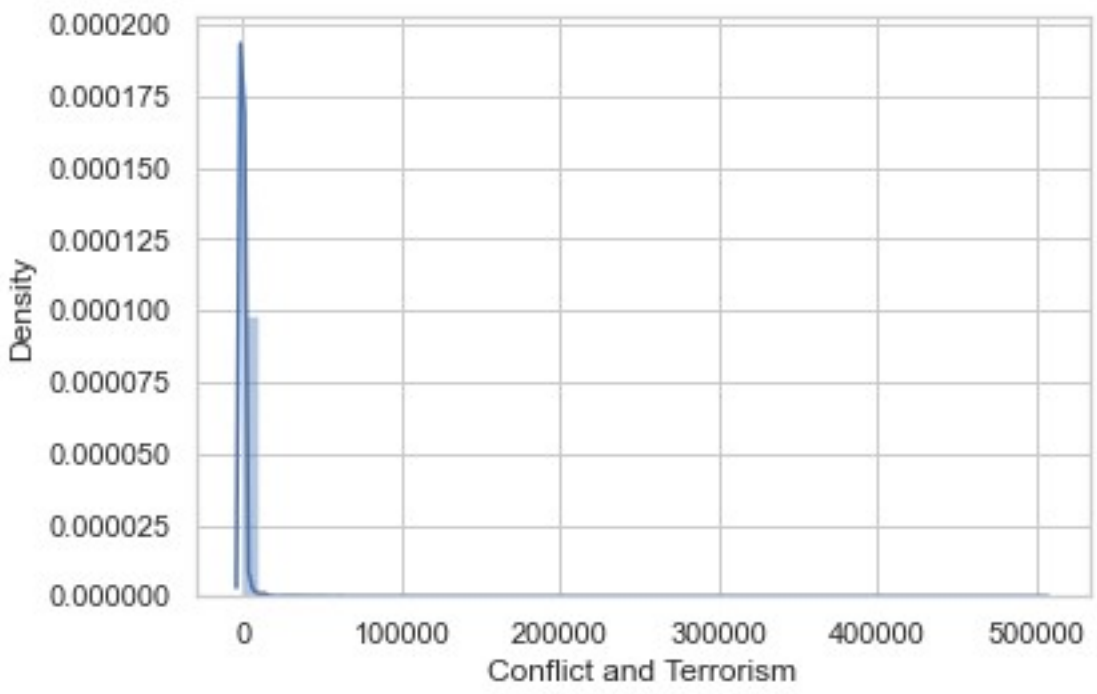

```
KstestResult(statistic=0.7628707830845005, pvalue=0.0)
```



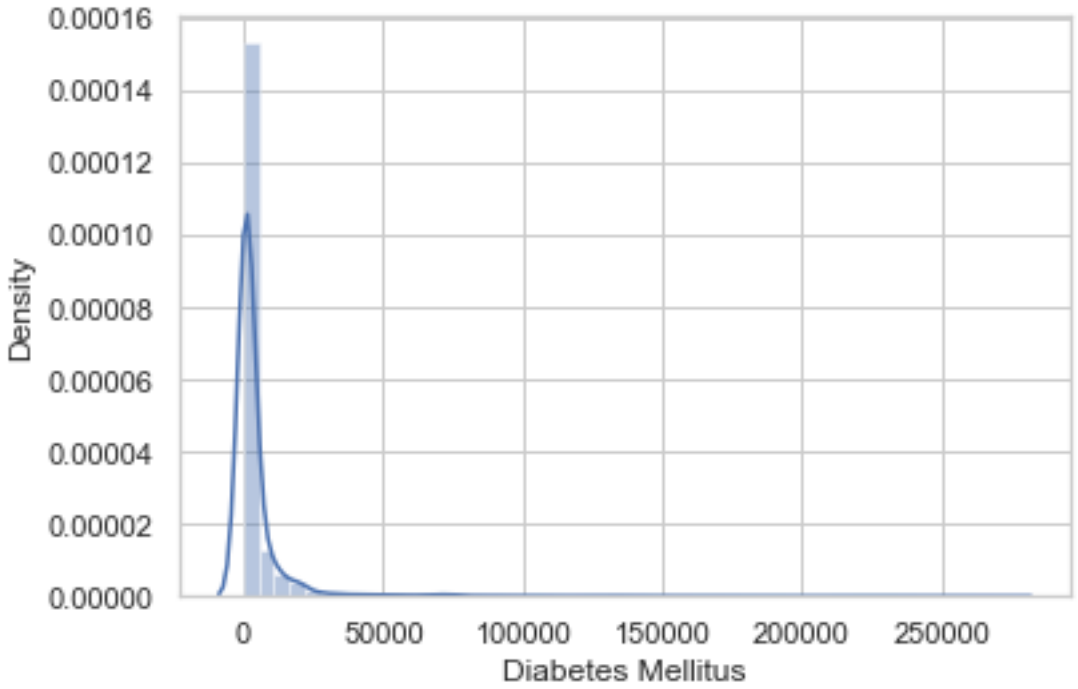

```
KstestResult(statistic=0.9937481411840562, pvalue=0.0)
```



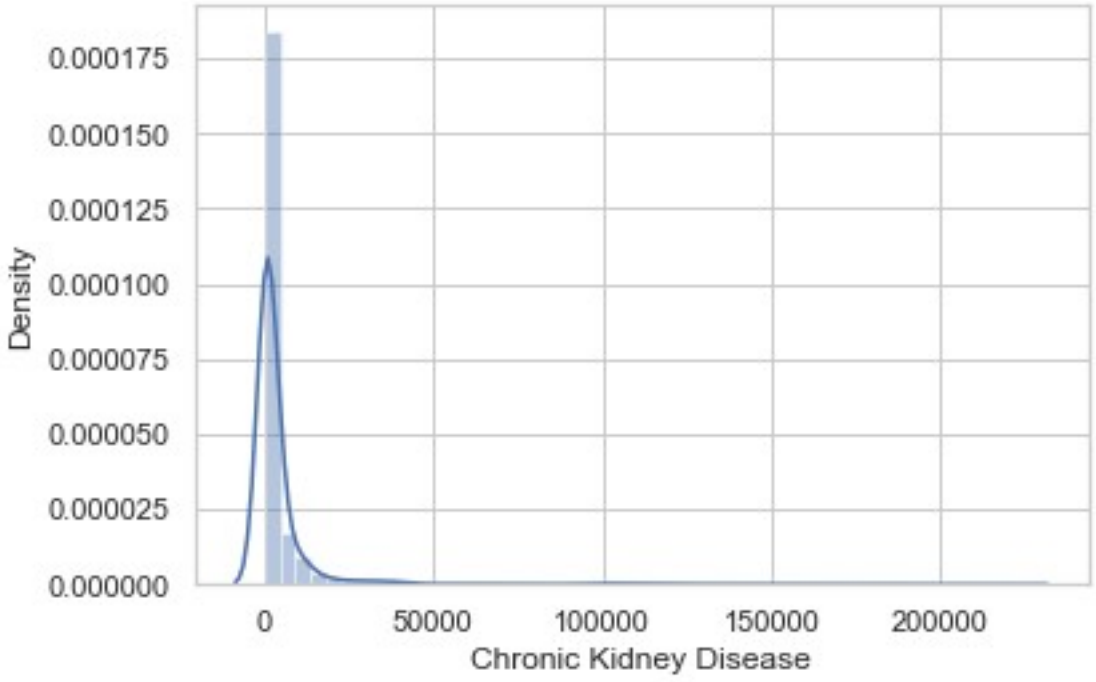

```
KstestResult(statistic=0.5, pvalue=0.0)
```



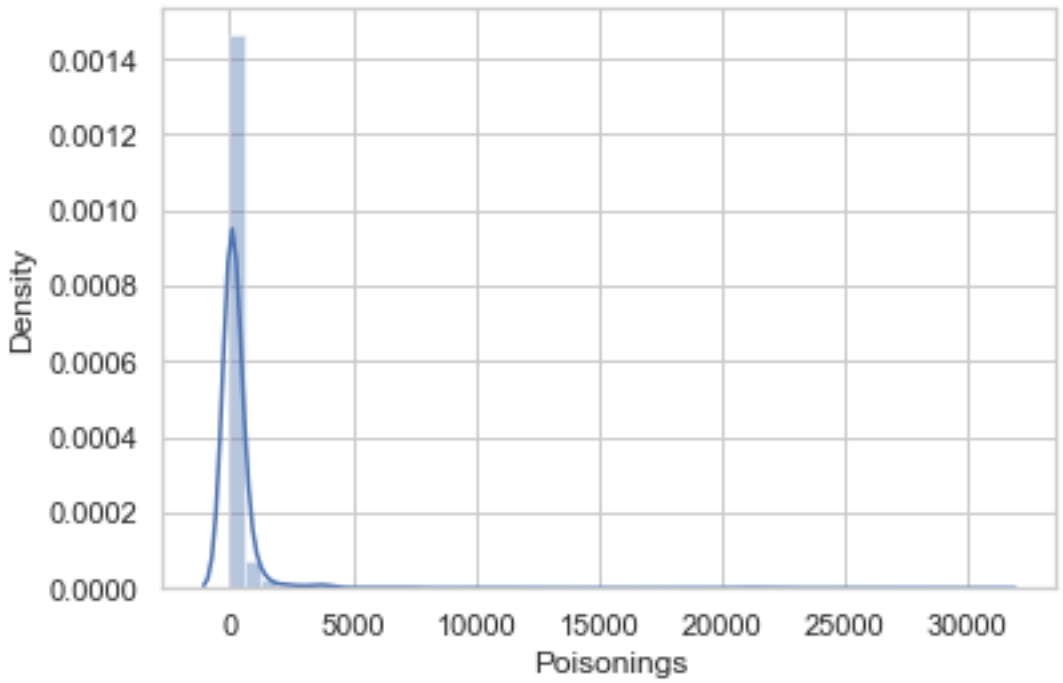

```
KstestResult(statistic=0.9919507555631412, pvalue=0.0)
```



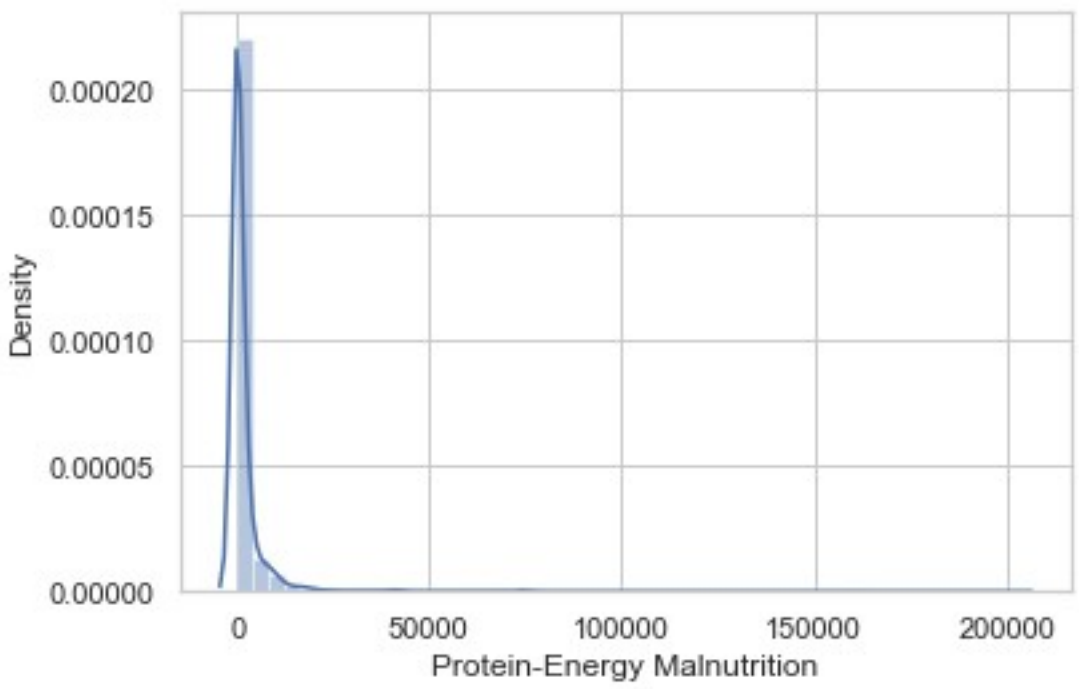

```
KstestResult(statistic=0.9849246117722915, pvalue=0.0)
```



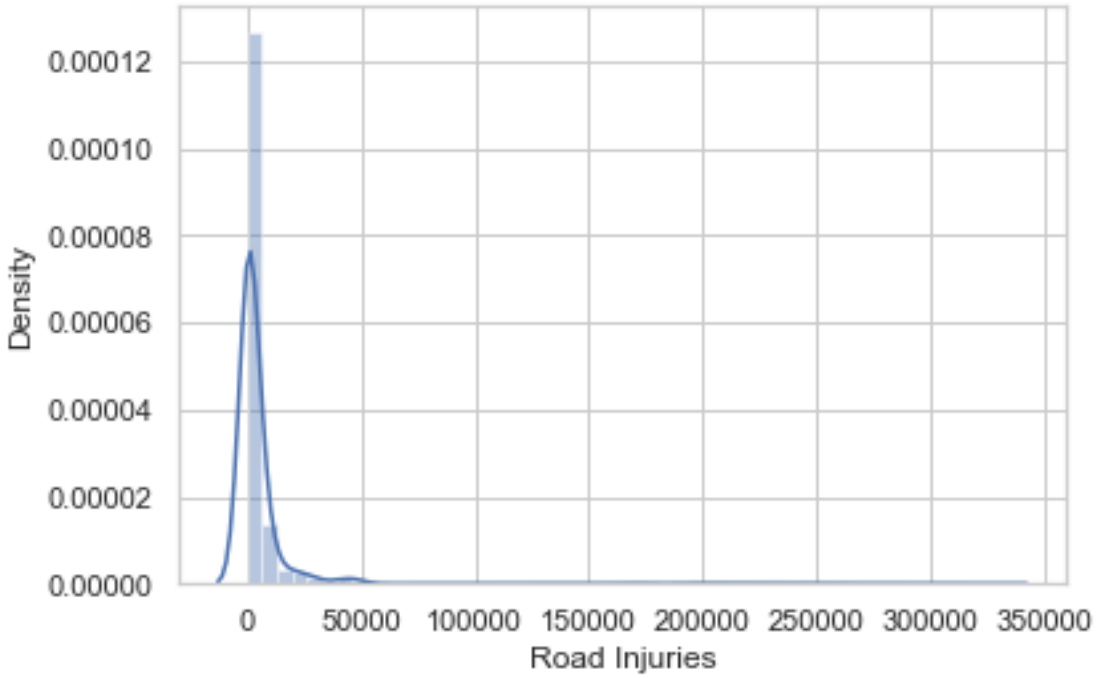

```
KstestResult(statistic=0.8133609791629319, pvalue=0.0)
```



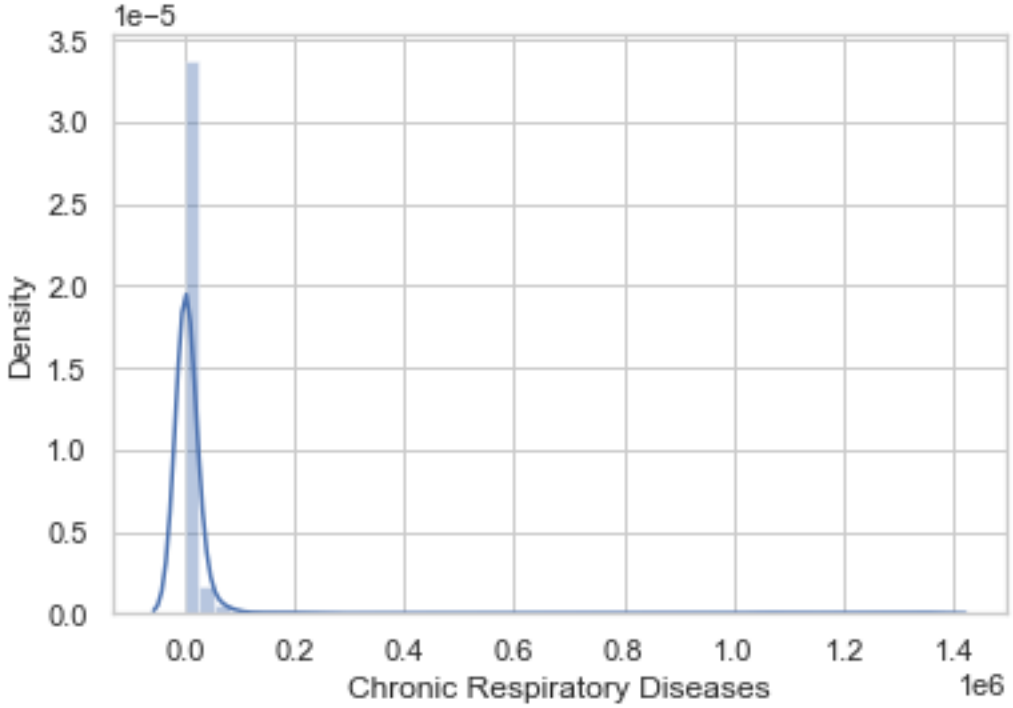

```
KstestResult(statistic=0.8520864693590103, pvalue=0.0)
```



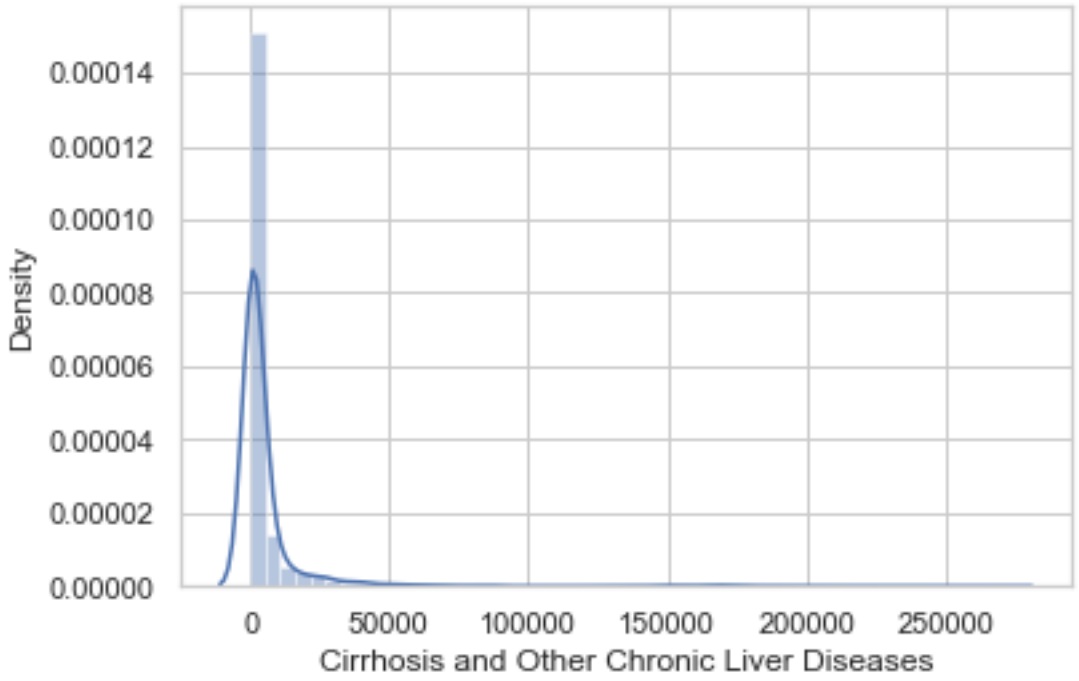

```
KstestResult(statistic=0.9829638274585659, pvalue=0.0)
```



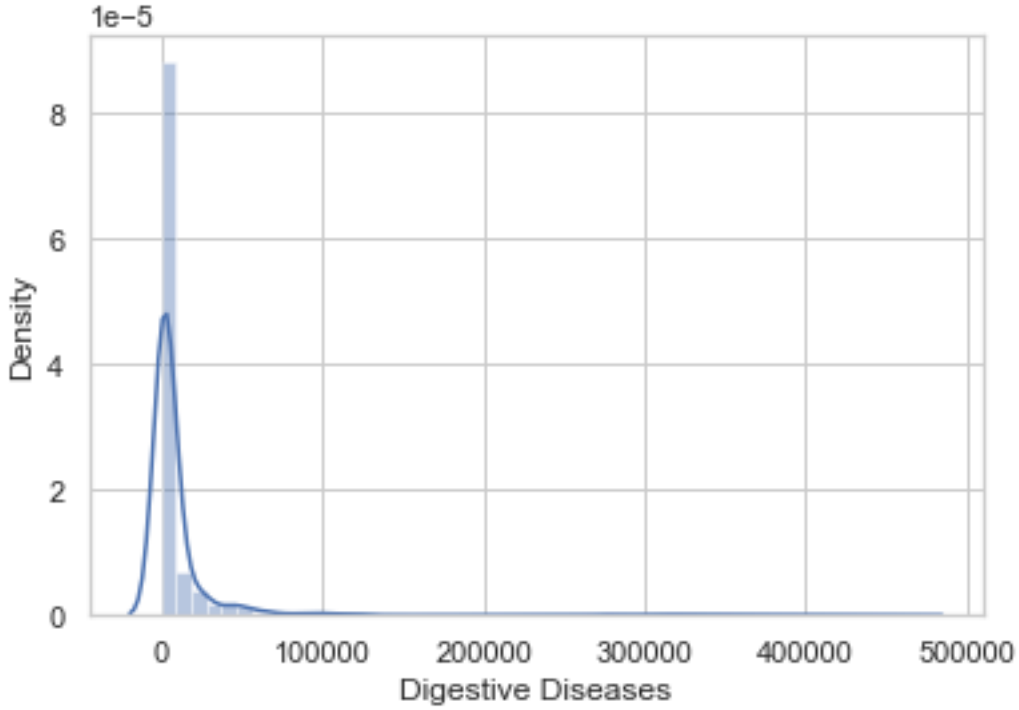

```
KstestResult(statistic=0.9890206163398663, pvalue=0.0)
```



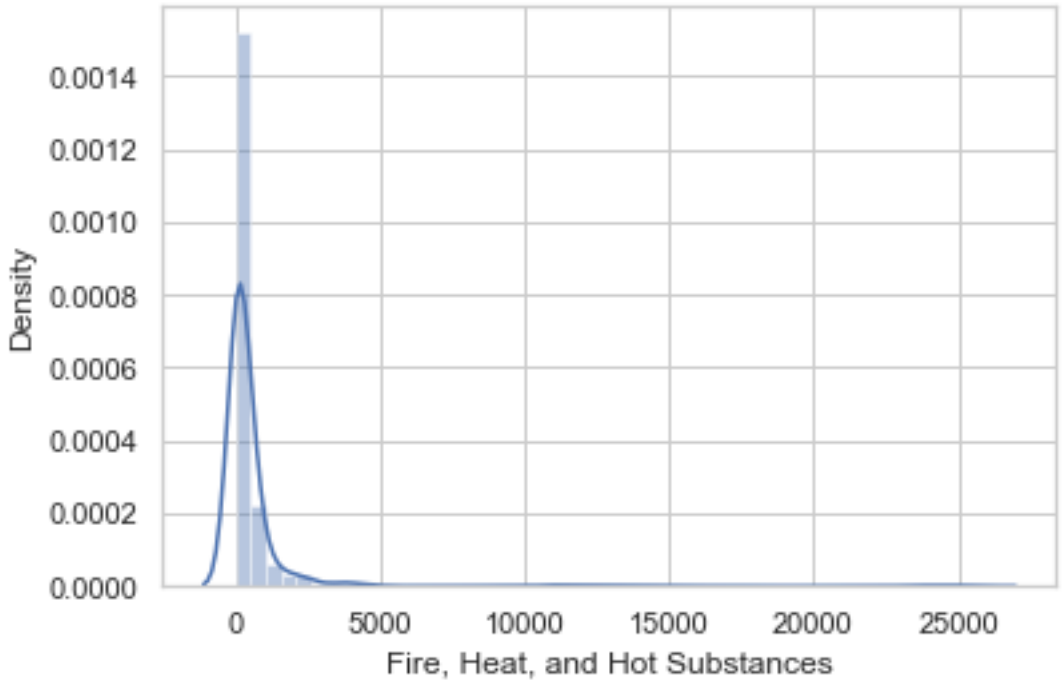

```
KstestResult(statistic=0.9738135006611803, pvalue=0.0)
```



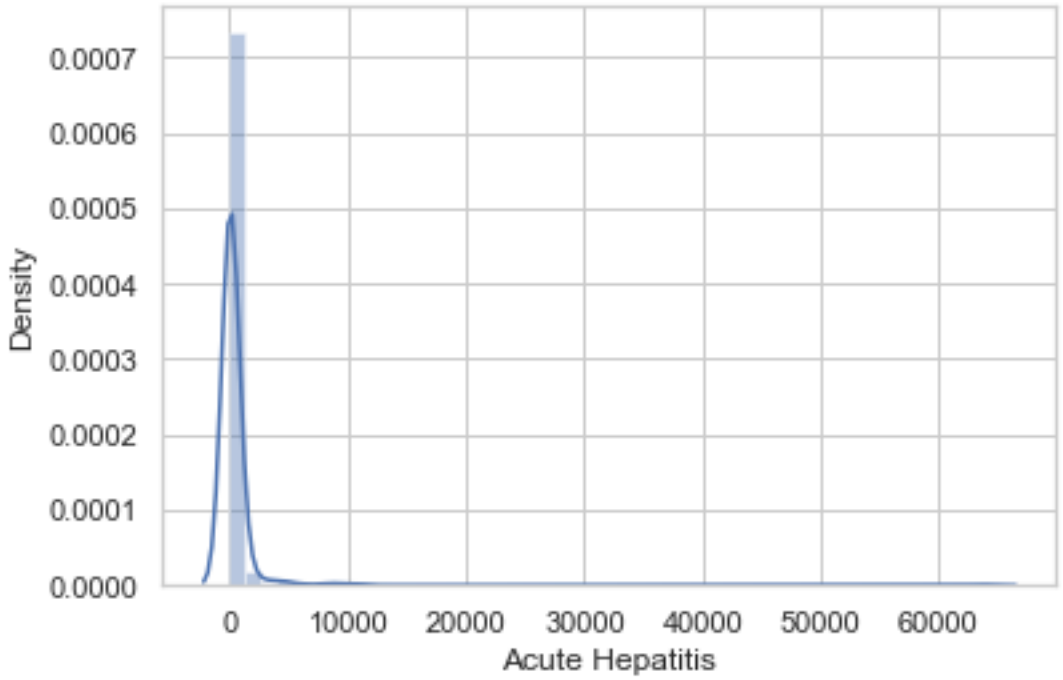

```
KstestResult(statistic=0.9878657882428797, pvalue=0.0)
```




```
KstestResult(statistic=0.895550521646592, pvalue=0.0)
```




```
KstestResult(statistic=0.75551784190803, pvalue=0.0)
```



All of the features are not follow the NDs (Right Skewed) as it shown in both graphs and Statistics

```
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()
```

```
# 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)
```

There are 30 years of statistics in this data set (1990-2019).

```
# 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  Year  Meningitis  \
0          Afghanistan  1990        2159
1          Afghanistan  1991        2218
2          Afghanistan  1992        2475
3          Afghanistan  1993        2812
4          Afghanistan  1994        3027
...                ...   ...         ...
6115          Zimbabwe  2015        1439
6116          Zimbabwe  2016        1457
6117          Zimbabwe  2017        1460
6118          Zimbabwe  2018        1450
6119          Zimbabwe  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
```

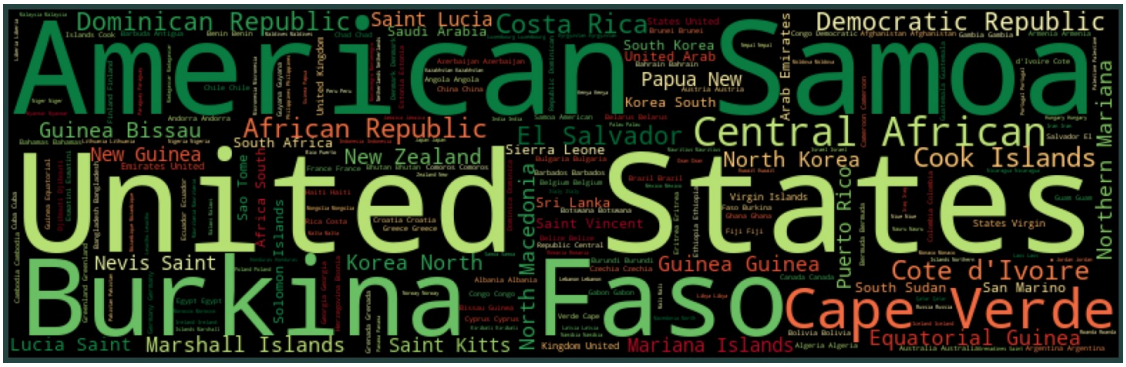
```
6117                                          2007                4342

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 34 columns]
```

EDA

Univariate analysis

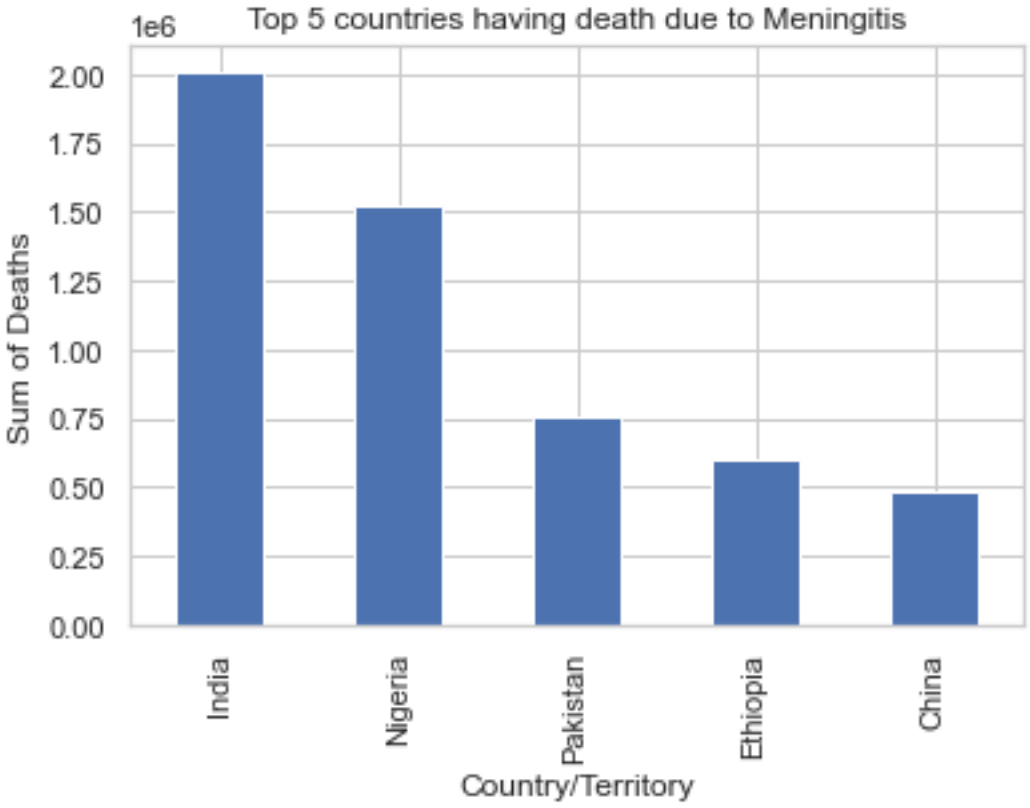
```
text = " ".join(subject_titles for subject_titles in
df1_cat["Country/Territory"])
wordcloud = WordCloud(width = 800, height = 250,
            background_color
="black",colormap="RdYlGn",max_font_size=100, stopwords =None,repeat=
True).generate(text)
plt.figure(figsize = (20, 8),facecolor= "#254441")
plt.imshow(wordcloud)
plt.axis("off")
plt.margins(x=0, y=0)
```

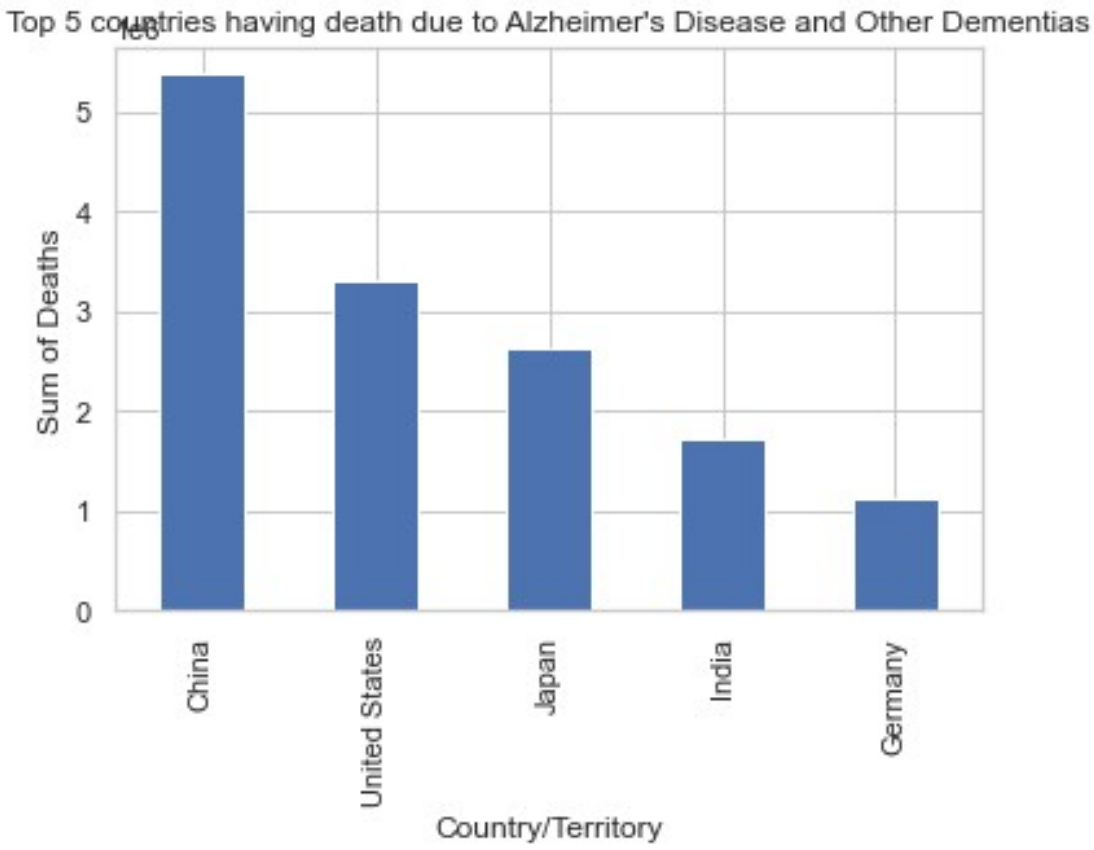
```
plt.tight_layout(pad = 0)
plt.show()
```

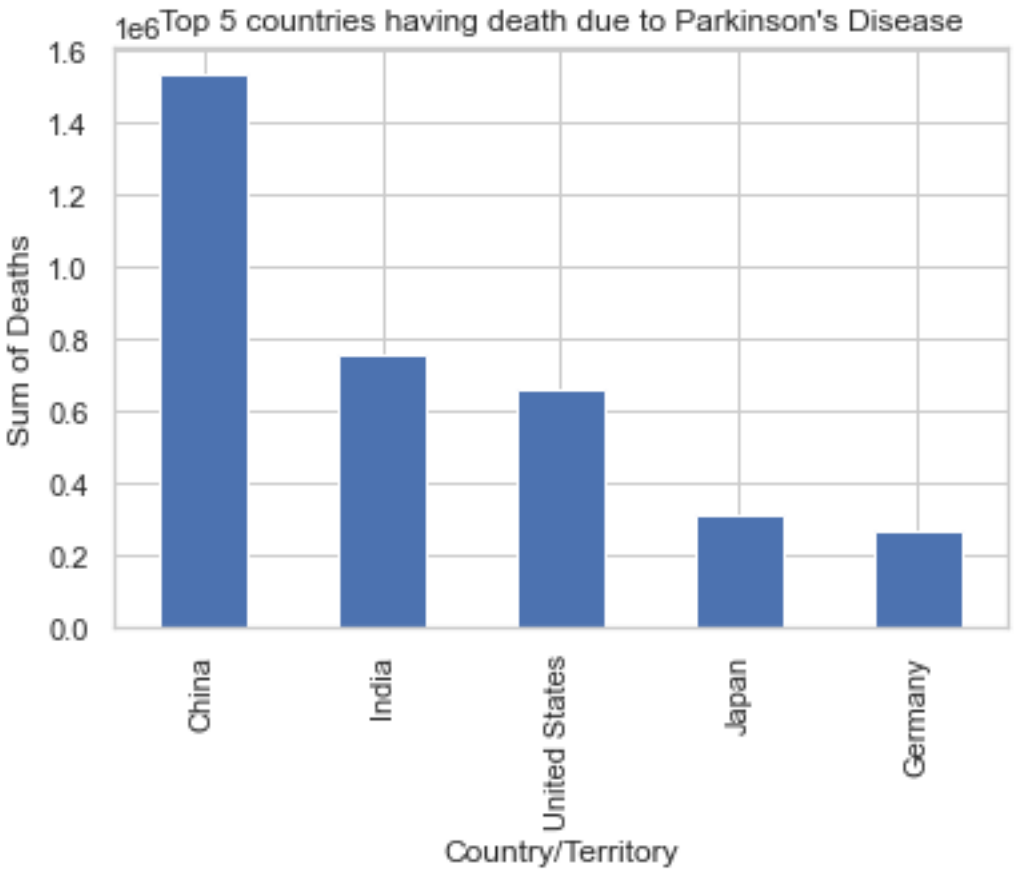


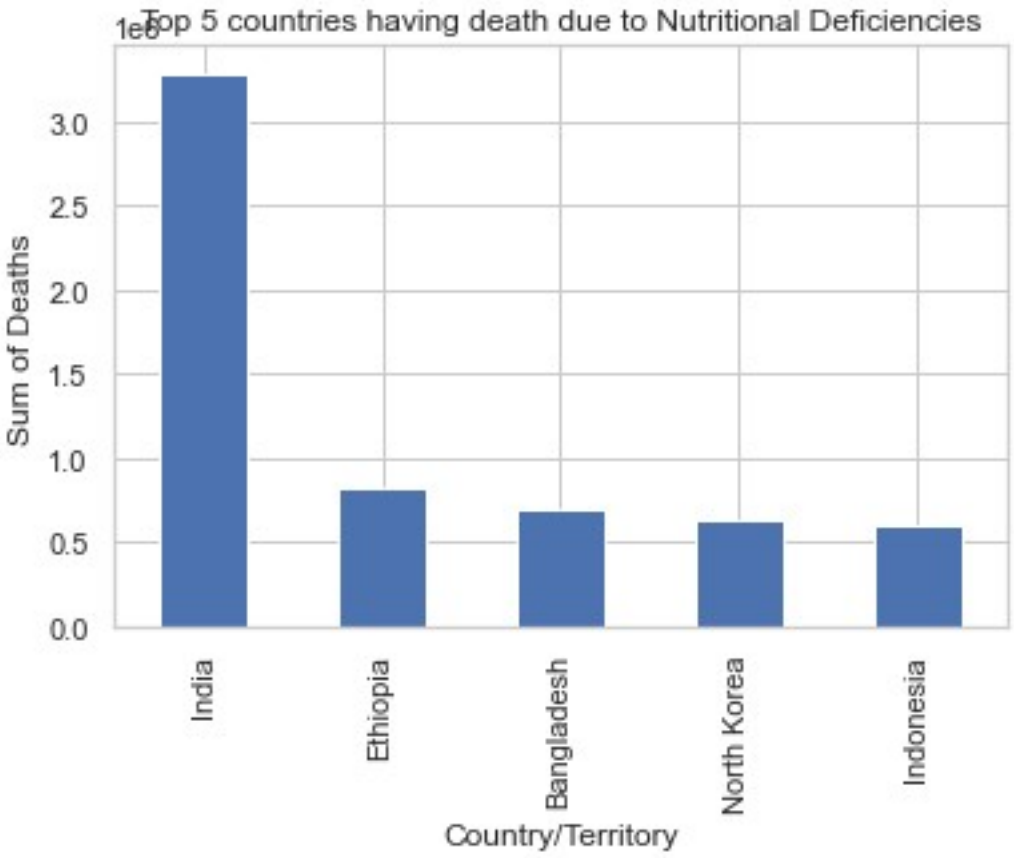
Bivariate analysis

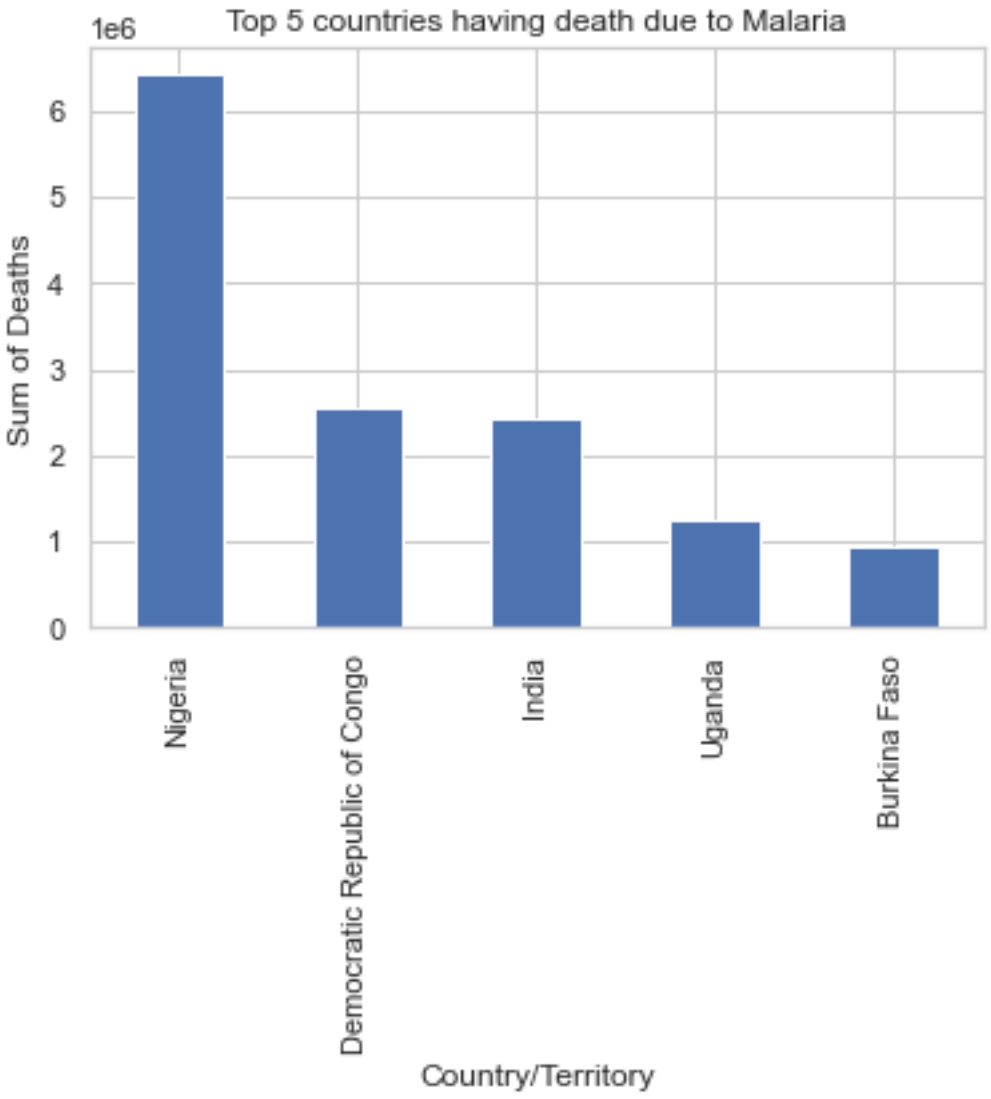
```
for i in df1_cont.columns:
    df2.groupby('Country/Territory')
[i].sum().sort_values(ascending=False).head().plot(kind='bar')
    plt.title('Top 5 countries having death due to '+i)
    plt.ylabel('Sum of Deaths')
    plt.show()
```

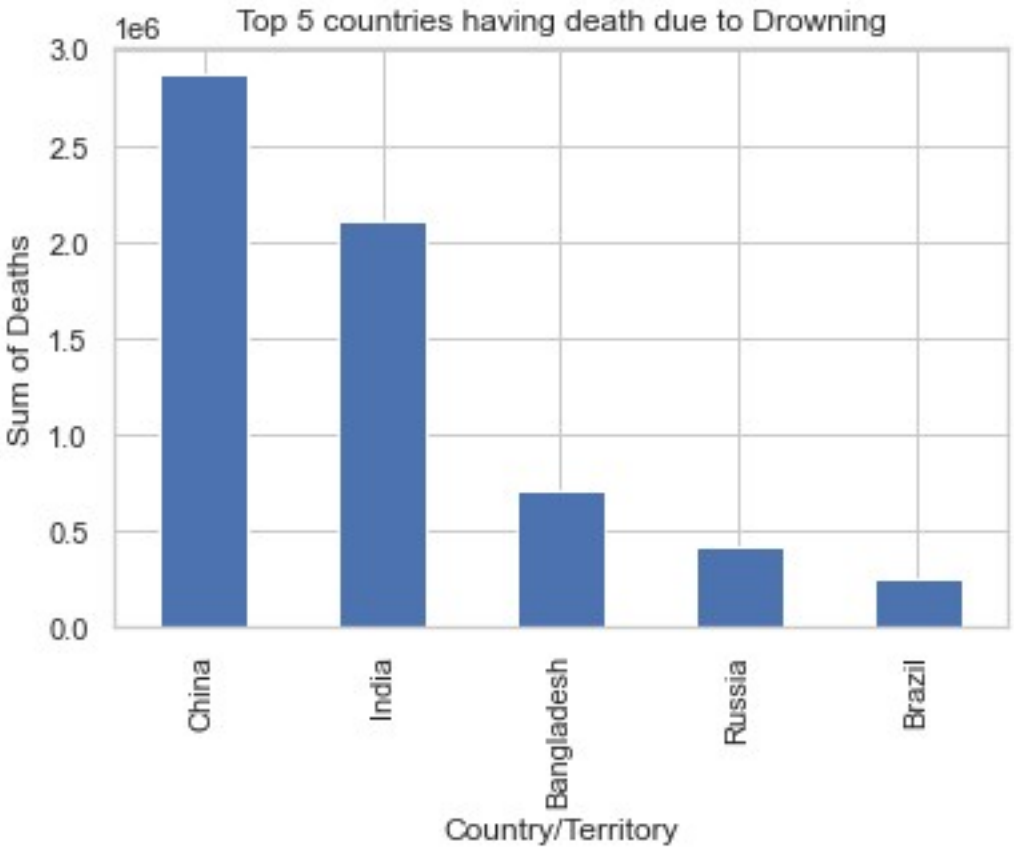



Top 5 countries having death due to Alzheimer's Disease and Other Dementias
1e6
5
4
3
2
1
0
Sum of Deaths
China
United States
Japan
India
Germany
Country/Territory

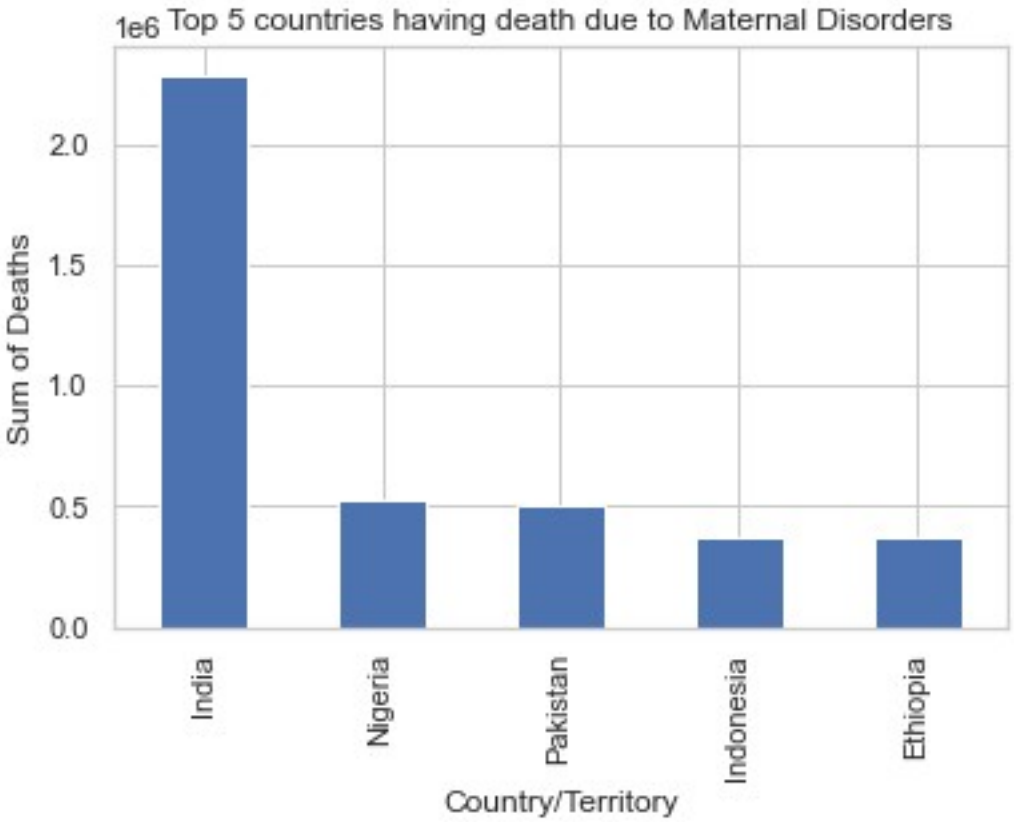

Top 5 countries having death due to Parkinson's Disease
1e6
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
China
India
United States
Japan
Germany
Country/Territory

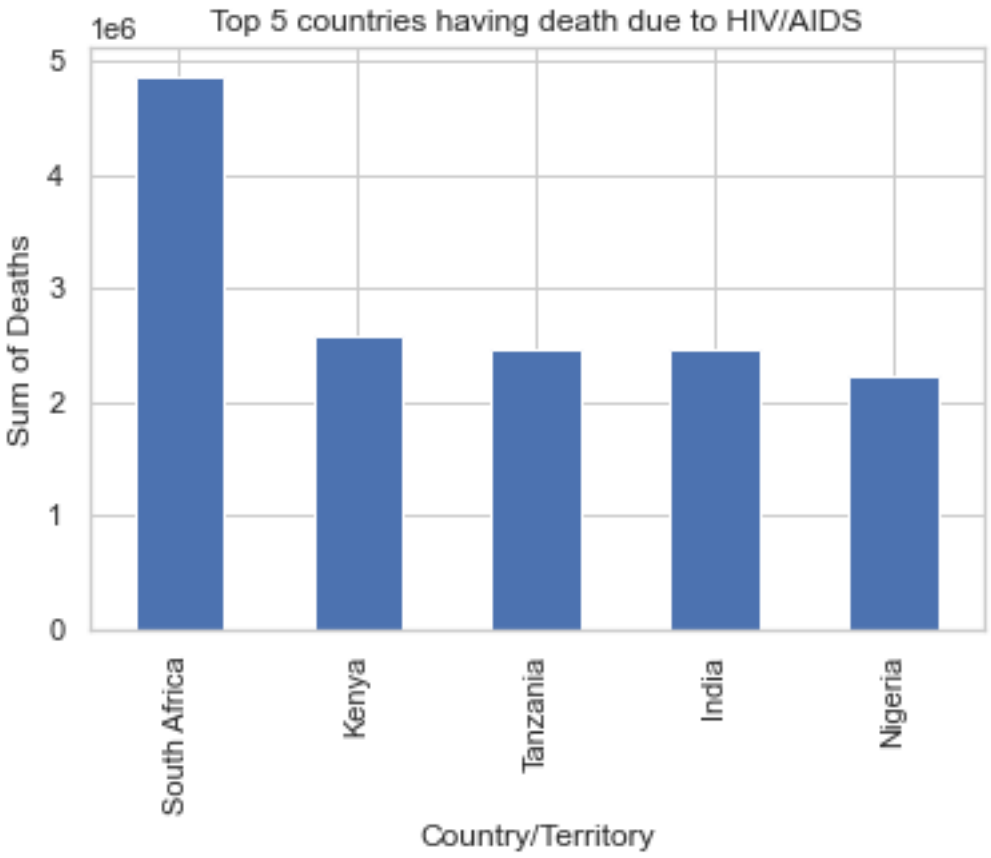

Top 5 countries having death due to Nutritional Deficiencies
1e8
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
India
Ethiopia
Bangladesh
North Korea
Indonesia
Country/Territory

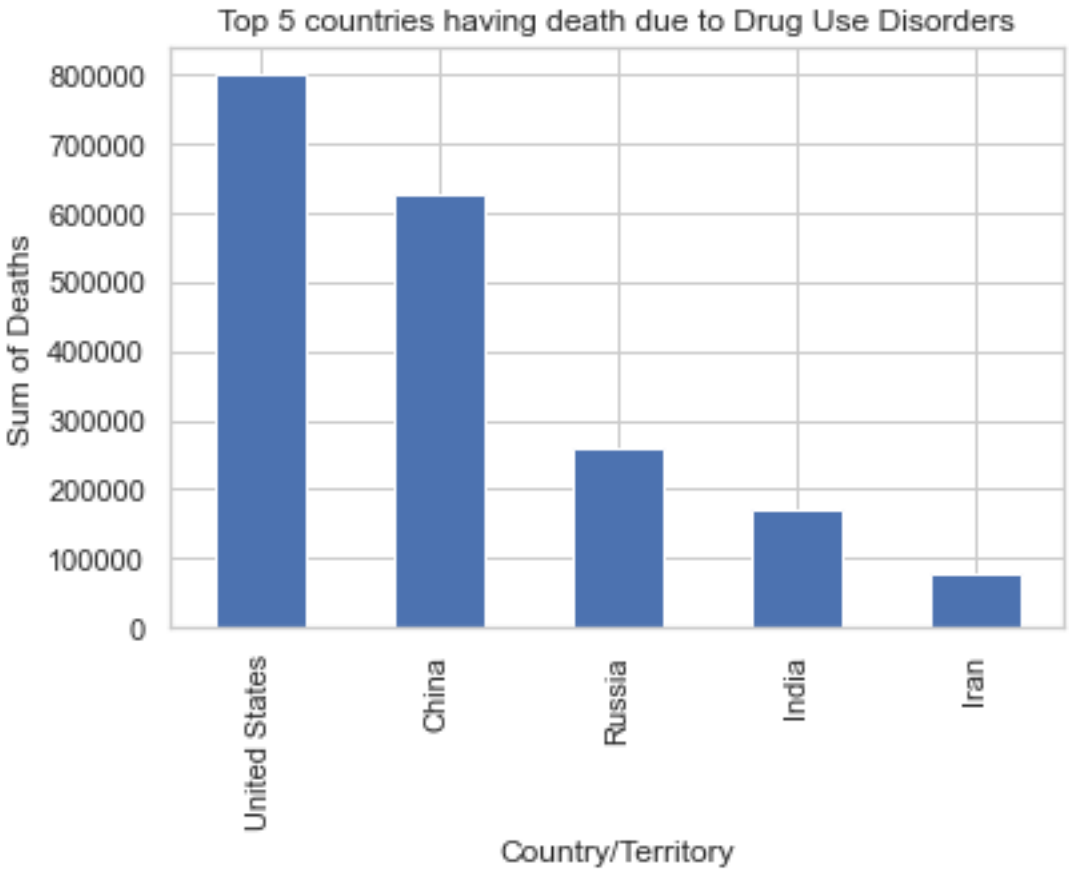

Top 5 countries having death due to Malaria
1e6
Sum of Deaths
0
1
2
3
4
5
6
Nigeria
Democratic Republic of Congo
India
Uganda
Burkina Faso
Country/Territory

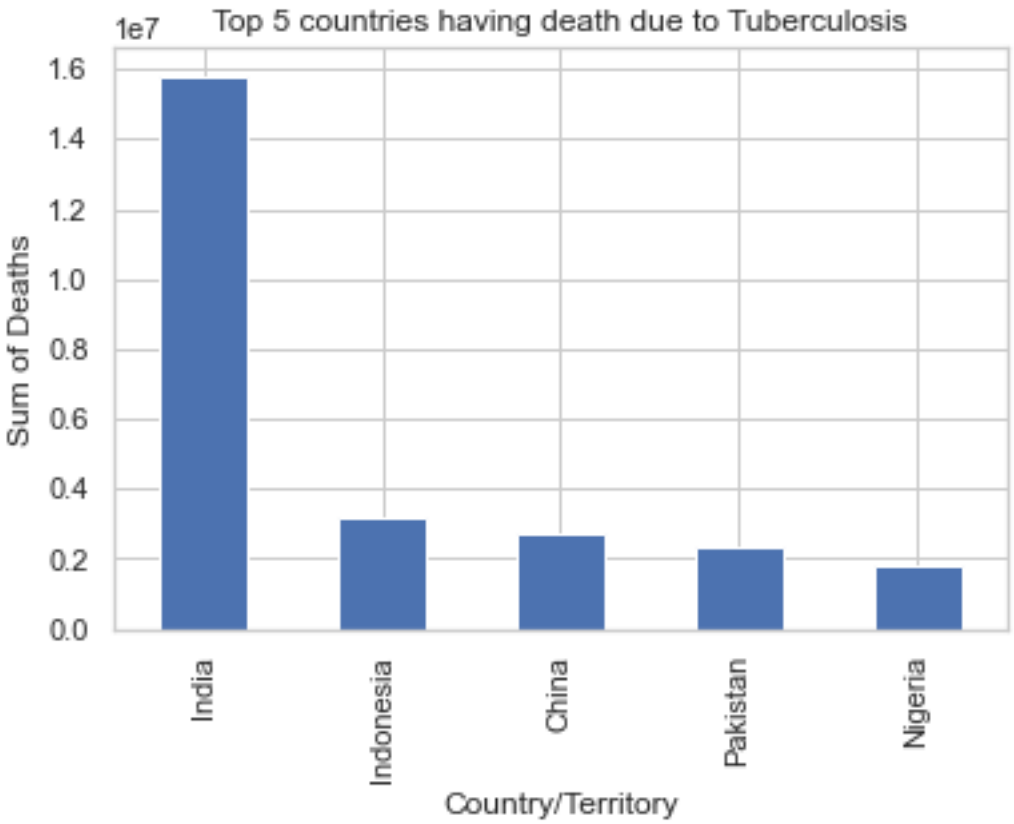

Top 5 countries having death due to Drowning
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
China
India
Bangladesh
Russia
Brazil
Country/Territory

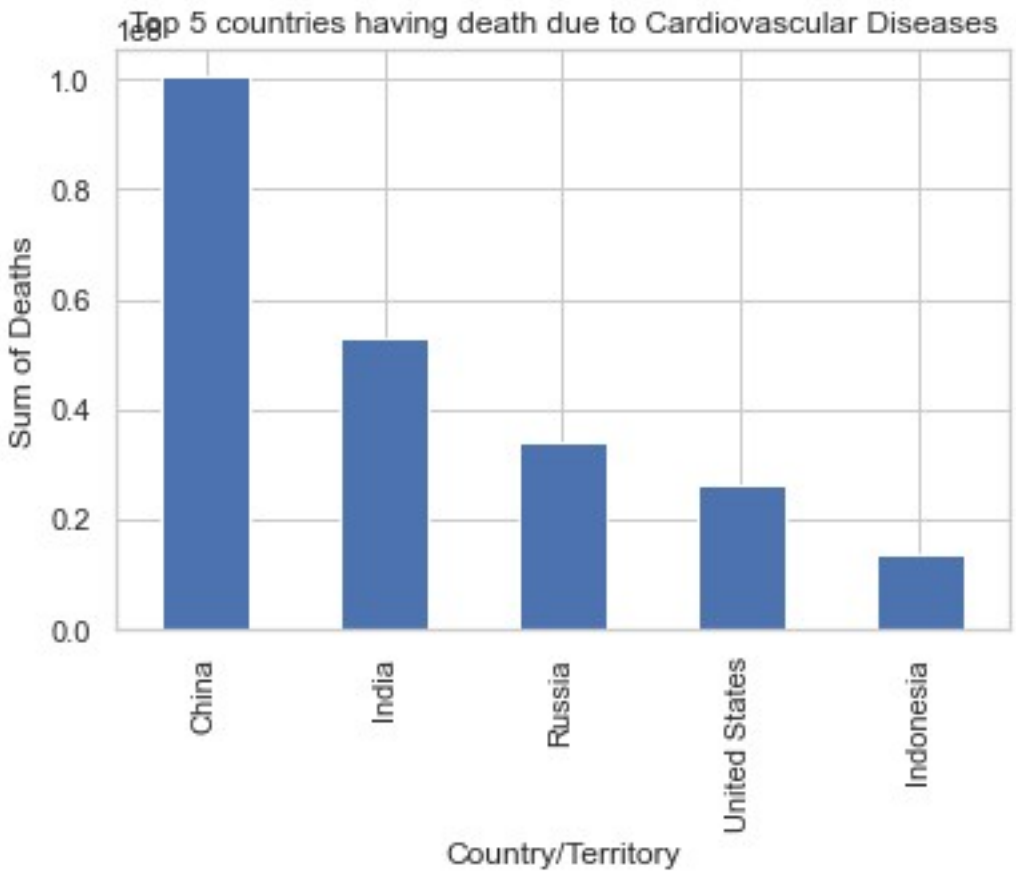

Top 5 countries having death due to Interpersonal Violence
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Sum of Deaths
Brazil
India
Russia
China
Colombia
Country/Territory

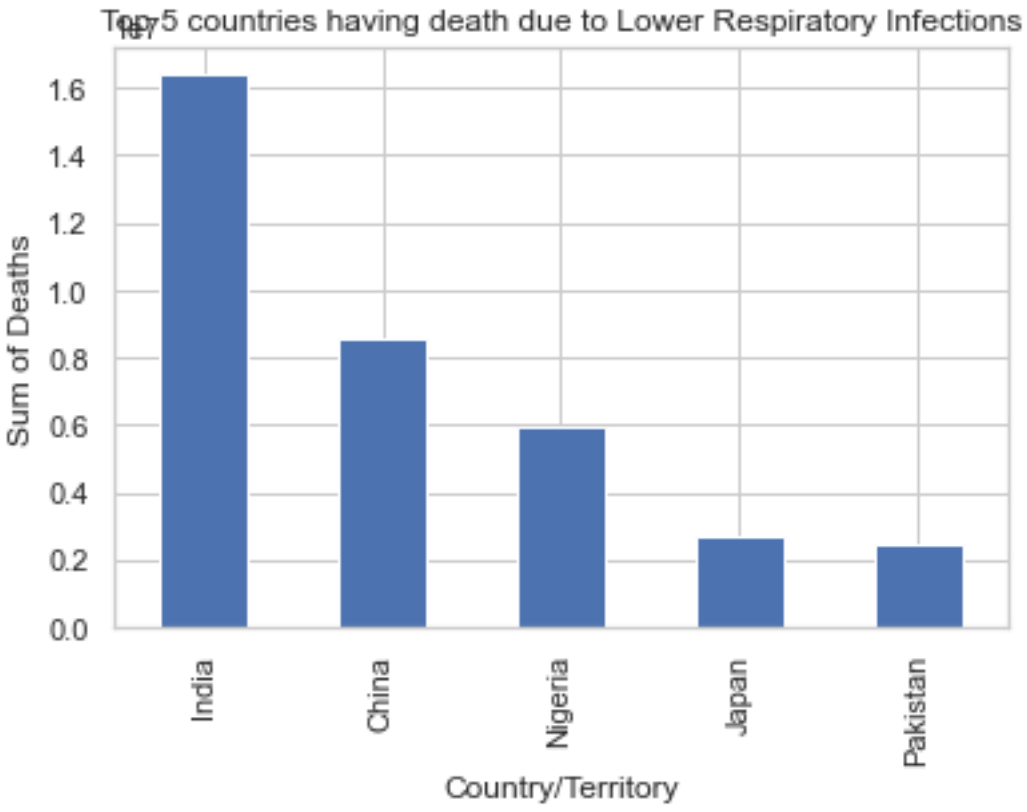

Top 5 countries having death due to Maternal Disorders
1e6
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
India
Nigeria
Pakistan
Indonesia
Ethiopia
Country/Territory

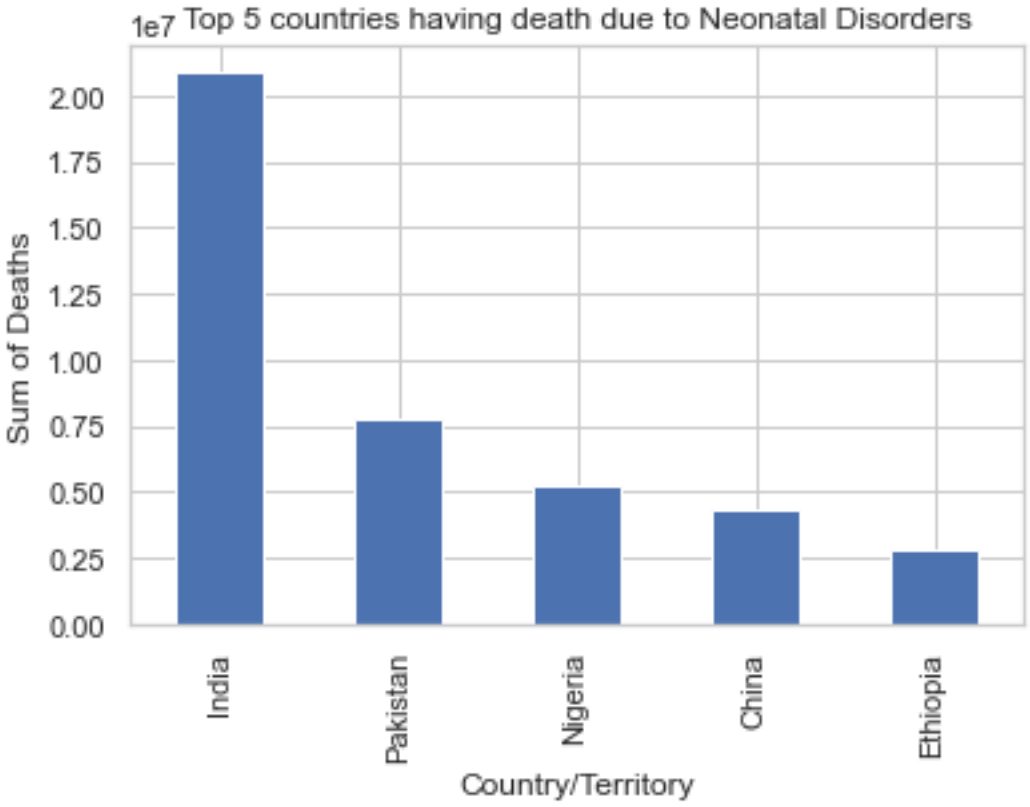

Top 5 countries having death due to HIV/AIDS
1e6
5
4
3
2
1
0
Sum of Deaths
South Africa
Kenya
Tanzania
India
Nigeria
Country/Territory

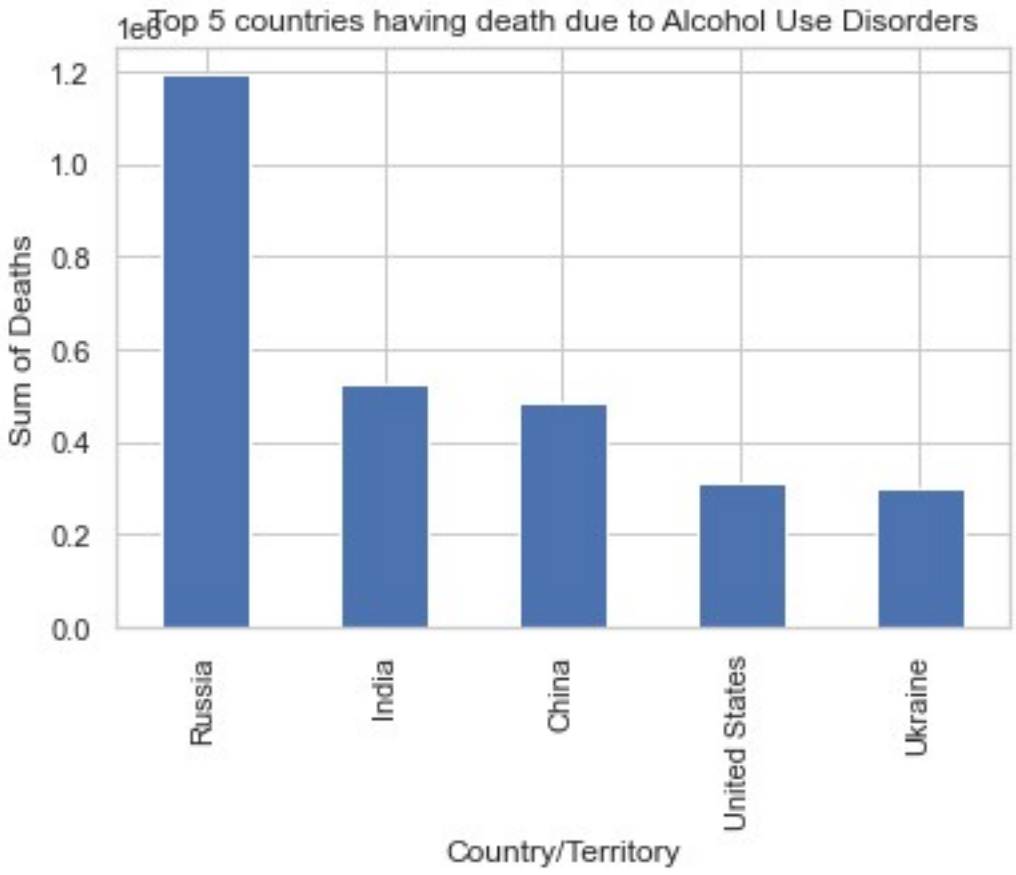

Top 5 countries having death due to Drug Use Disorders
Sum of Deaths
800000
700000
600000
500000
400000
300000
200000
100000
0
United States
China
Russia
India
Iran
Country/Territory

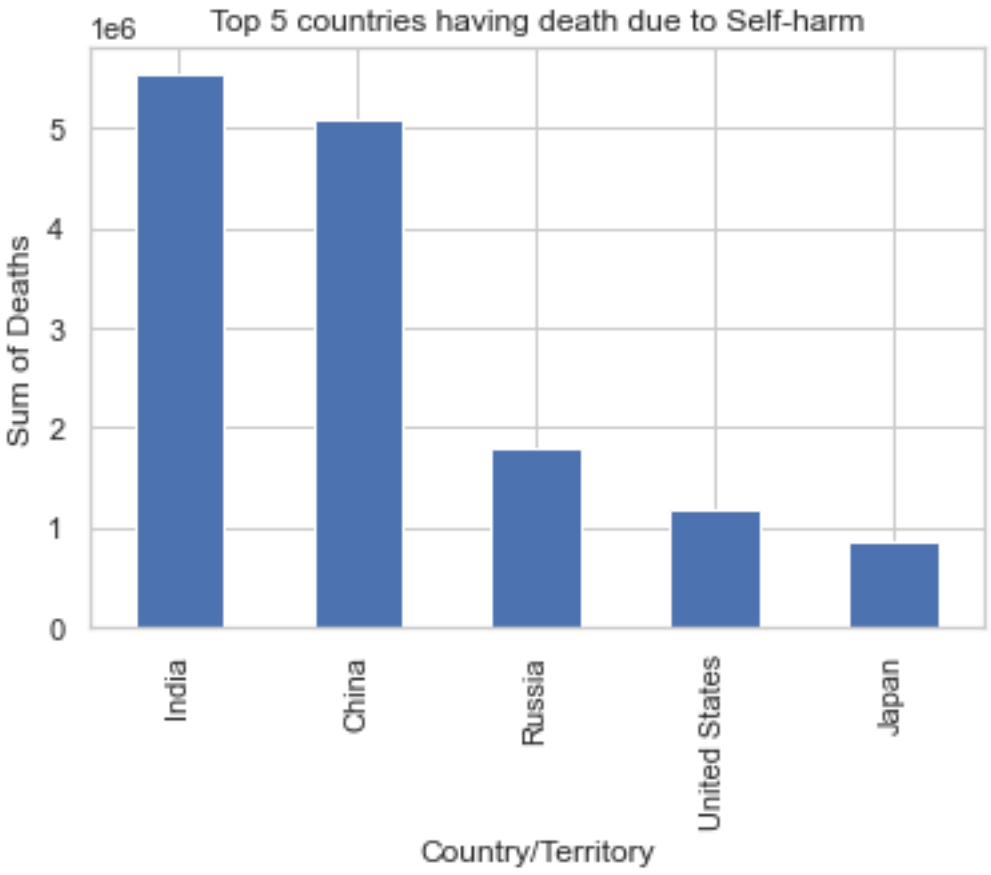

Top 5 countries having death due to Tuberculosis
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
India
Indonesia
China
Pakistan
Nigeria
Country/Territory

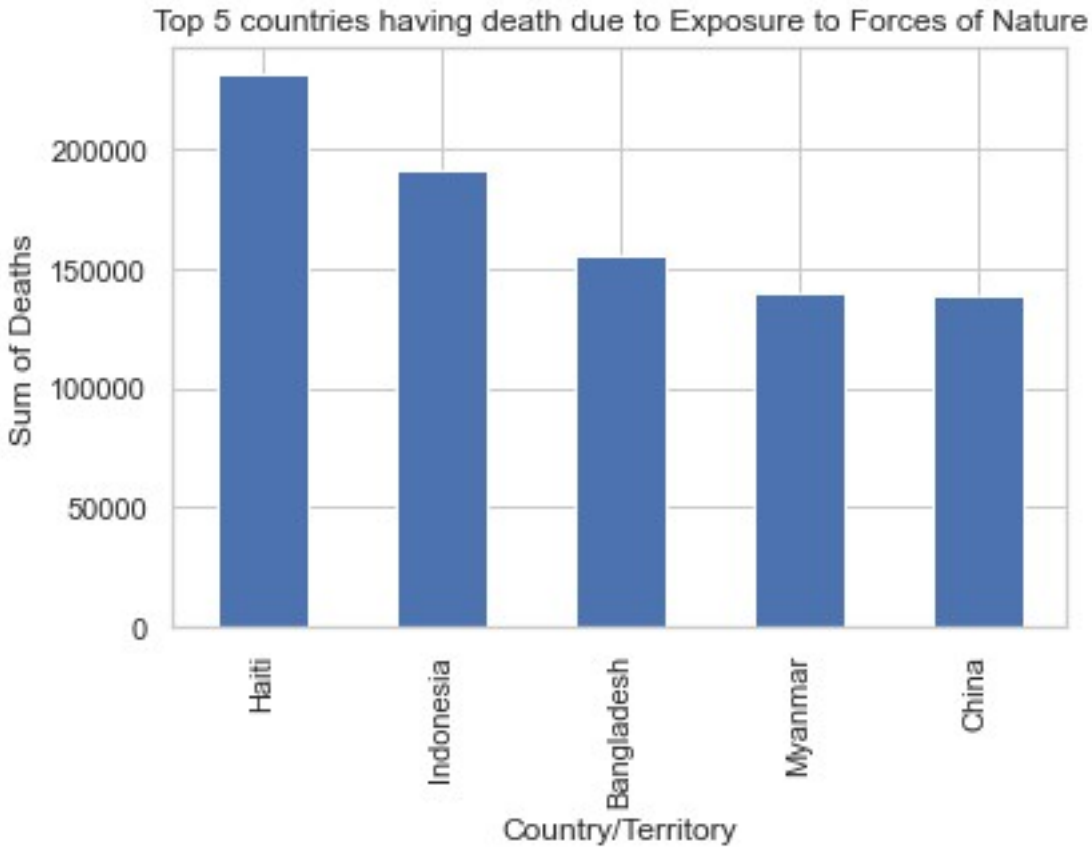

Top 5 countries having death due to Cardiovascular Diseases
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
China
India
Russia
United States
Indonesia
Country/Territory

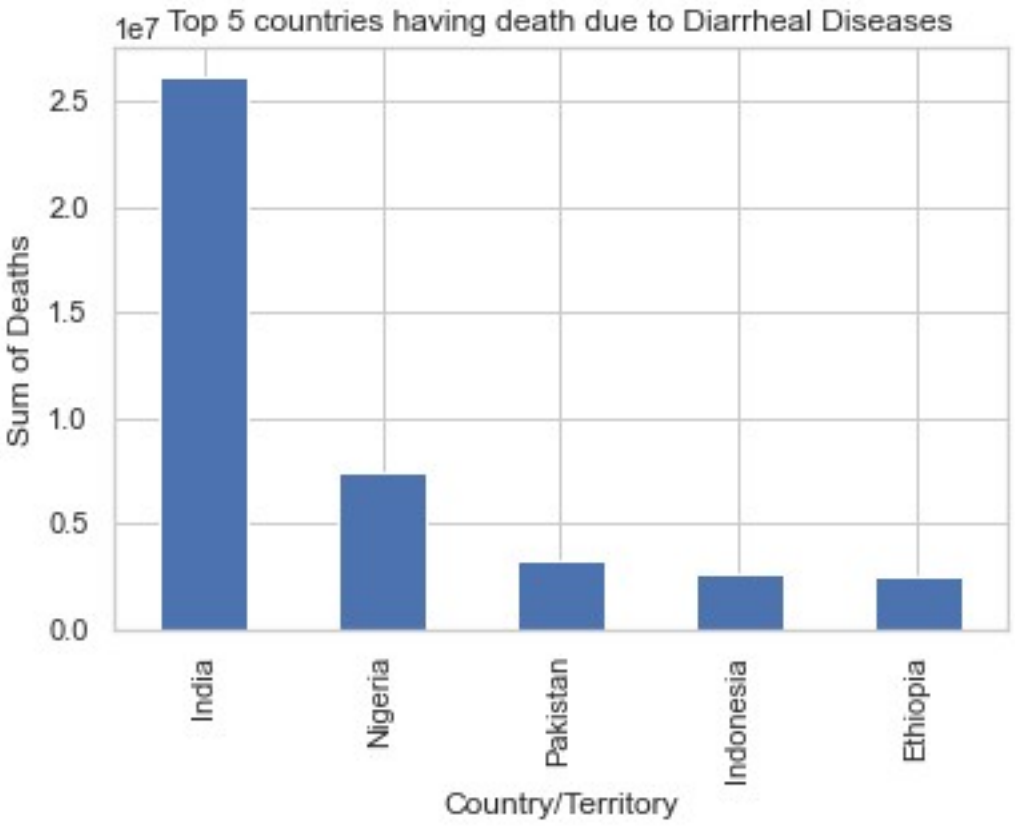

Top 5 countries having death due to Lower Respiratory Infections
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
India
China
Nigeria
Japan
Pakistan
Country/Territory

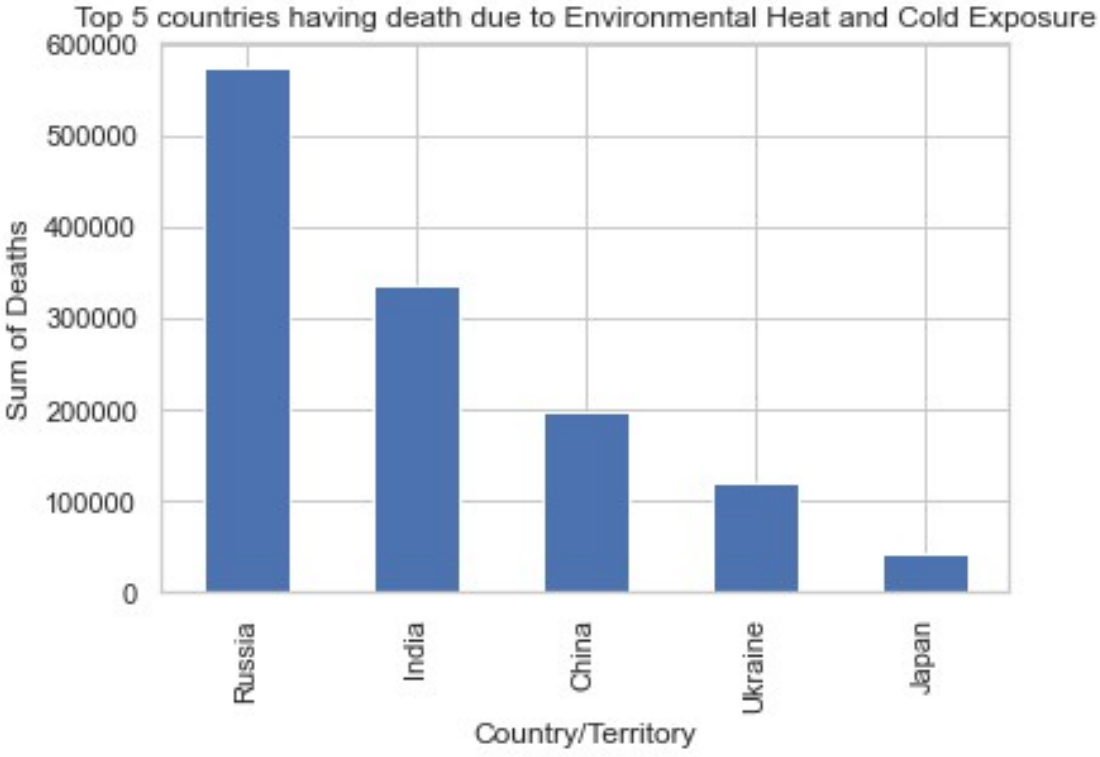

1e7 Top 5 countries having death due to Neonatal Disorders
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Sum of Deaths
India
Pakistan
Nigeria
China
Ethiopia
Country/Territory

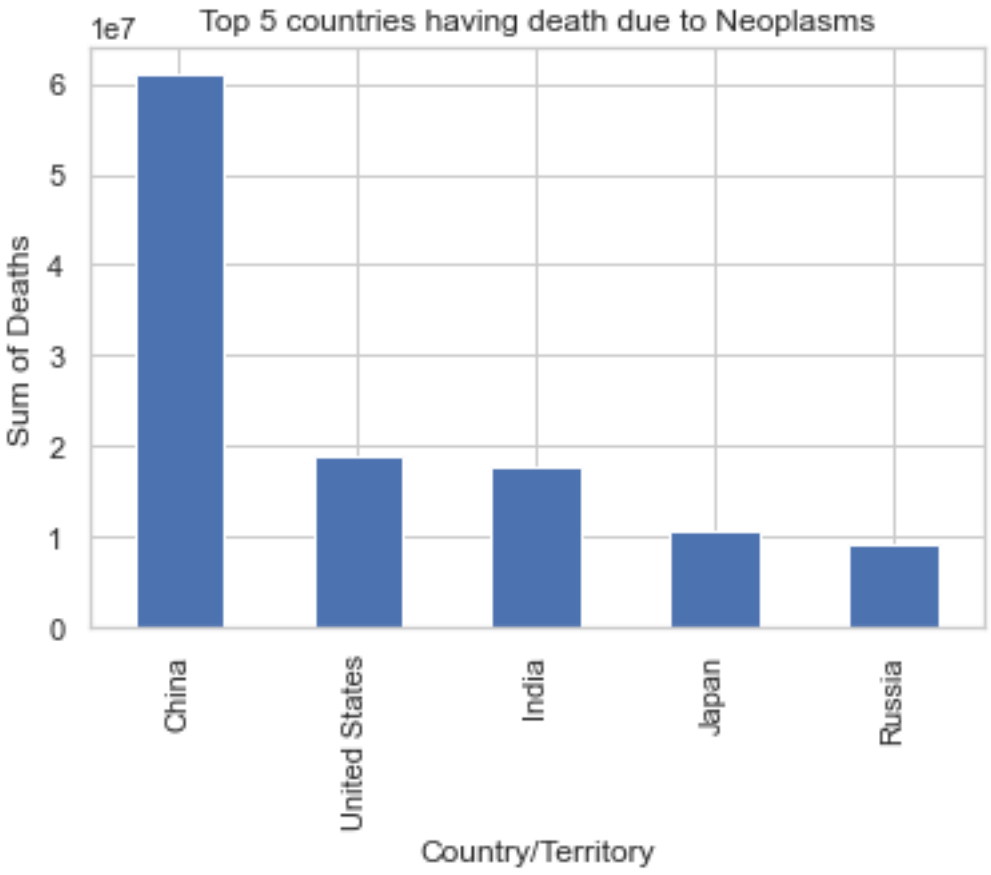

Top 5 countries having death due to Alcohol Use Disorders
1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
Russia
India
China
United States
Ukraine
Country/Territory

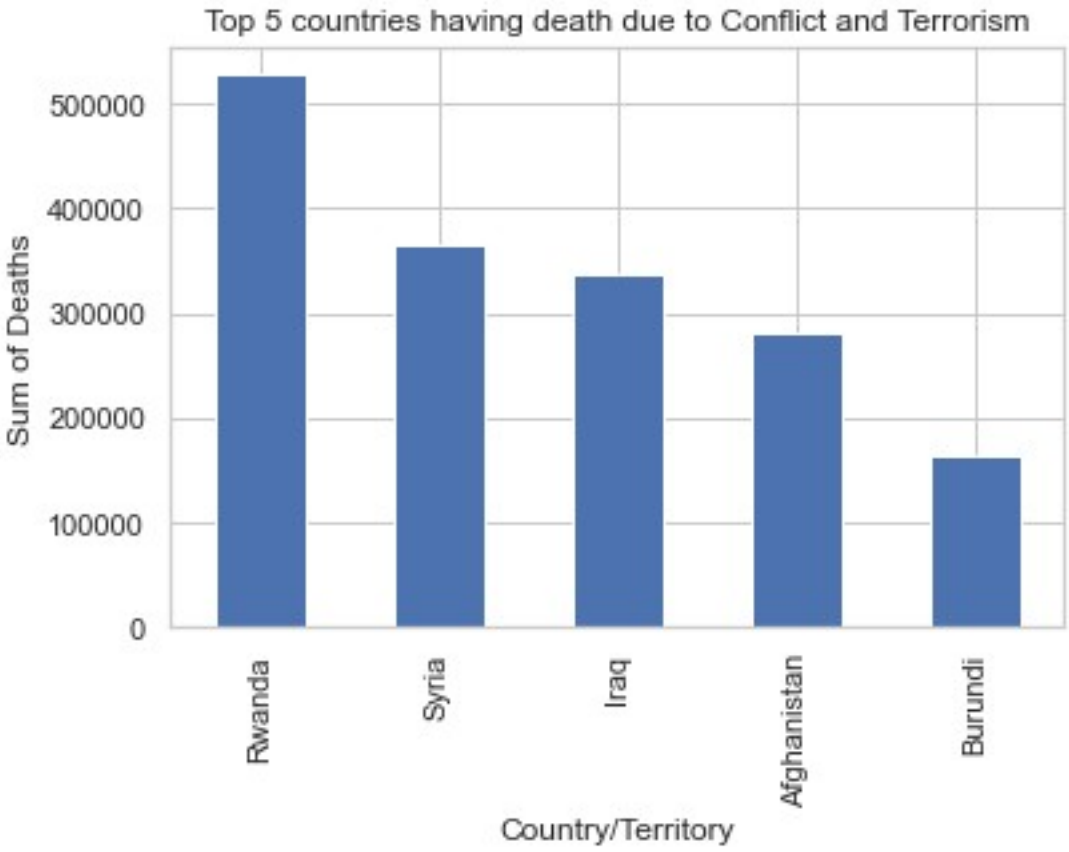

Top 5 countries having death due to Self-harm
1e6
5
4
3
2
1
0
Sum of Deaths
India
China
Russia
United States
Japan
Country/Territory

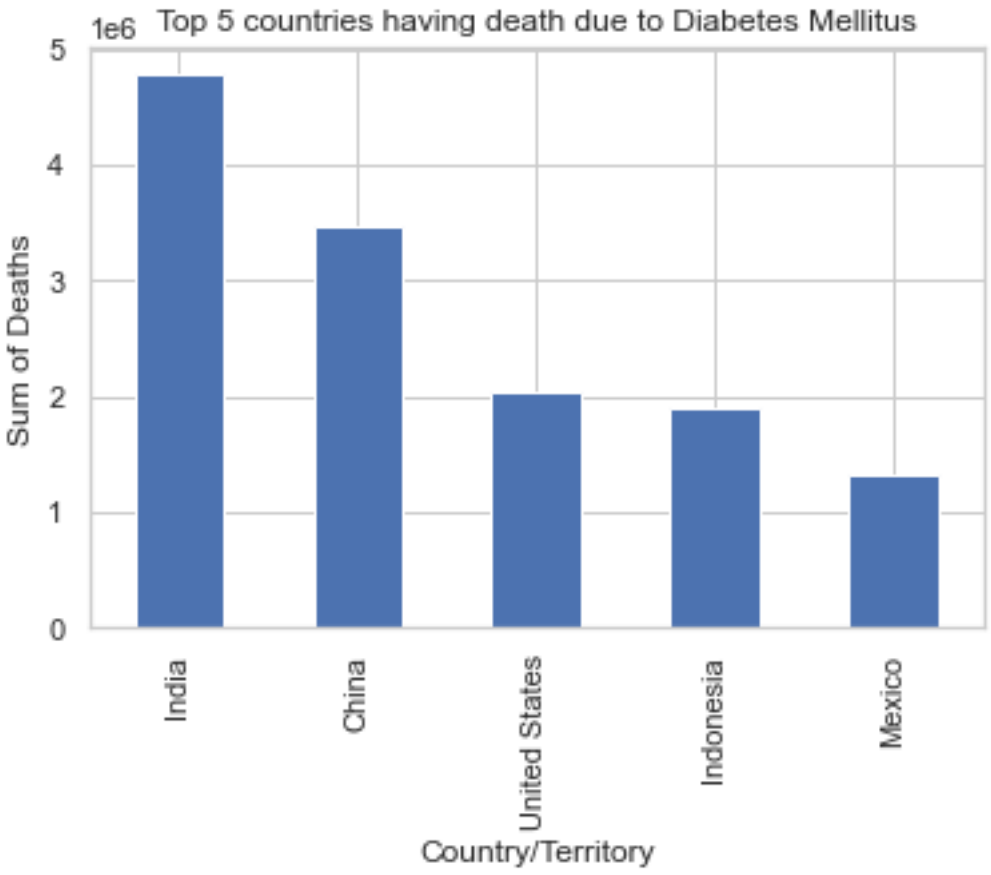

Top 5 countries having death due to Exposure to Forces of Nature
Sum of Deaths
200000
150000
100000
50000
0
Haiti
Indonesia
Bangladesh
Myanmar
China
Country/Territory

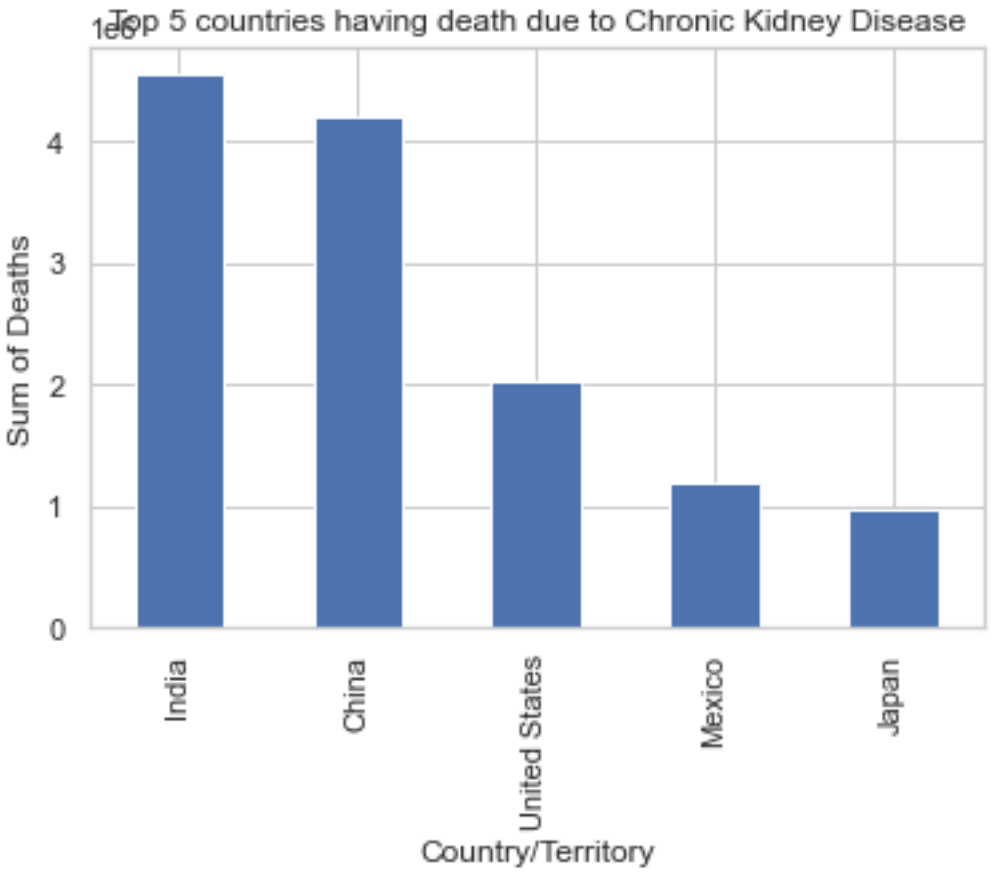

Top 5 countries having death due to Diarrheal Diseases
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
India
Nigeria
Pakistan
Indonesia
Ethiopia
Country/Territory

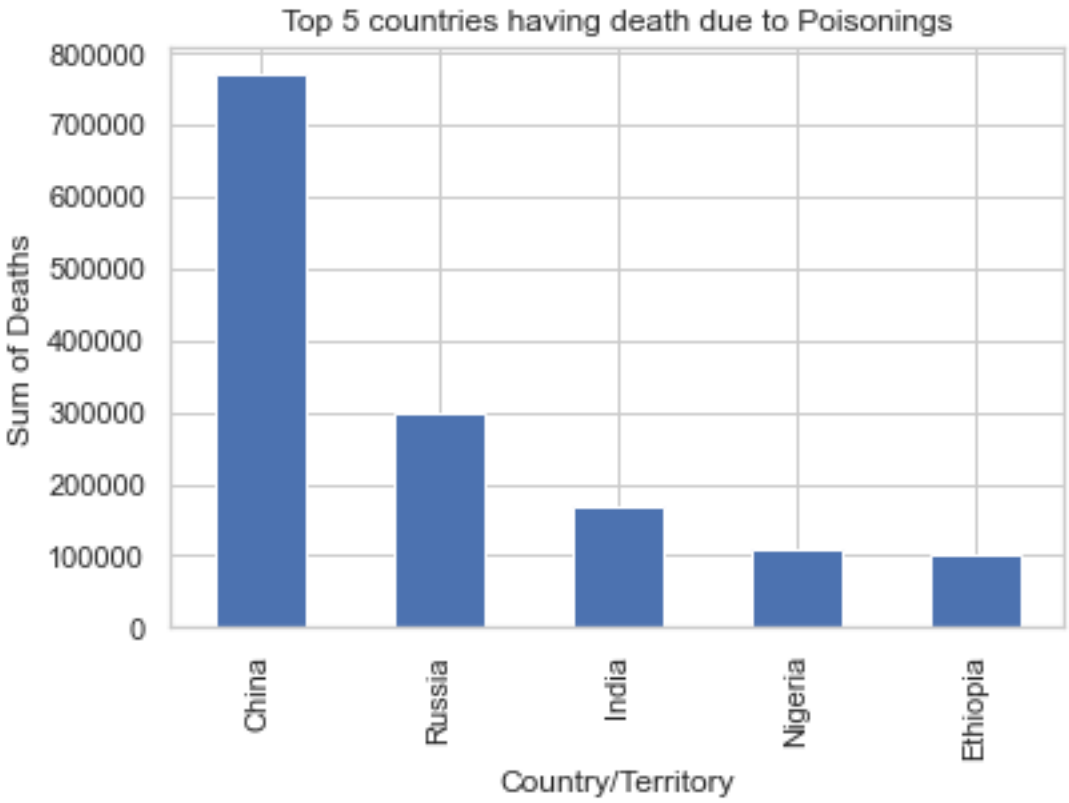

Top 5 countries having death due to Environmental Heat and Cold Exposure
600000
500000
400000
300000
200000
100000
0
Sum of Deaths
Russia
India
China
Ukraine
Japan
Country/Territory

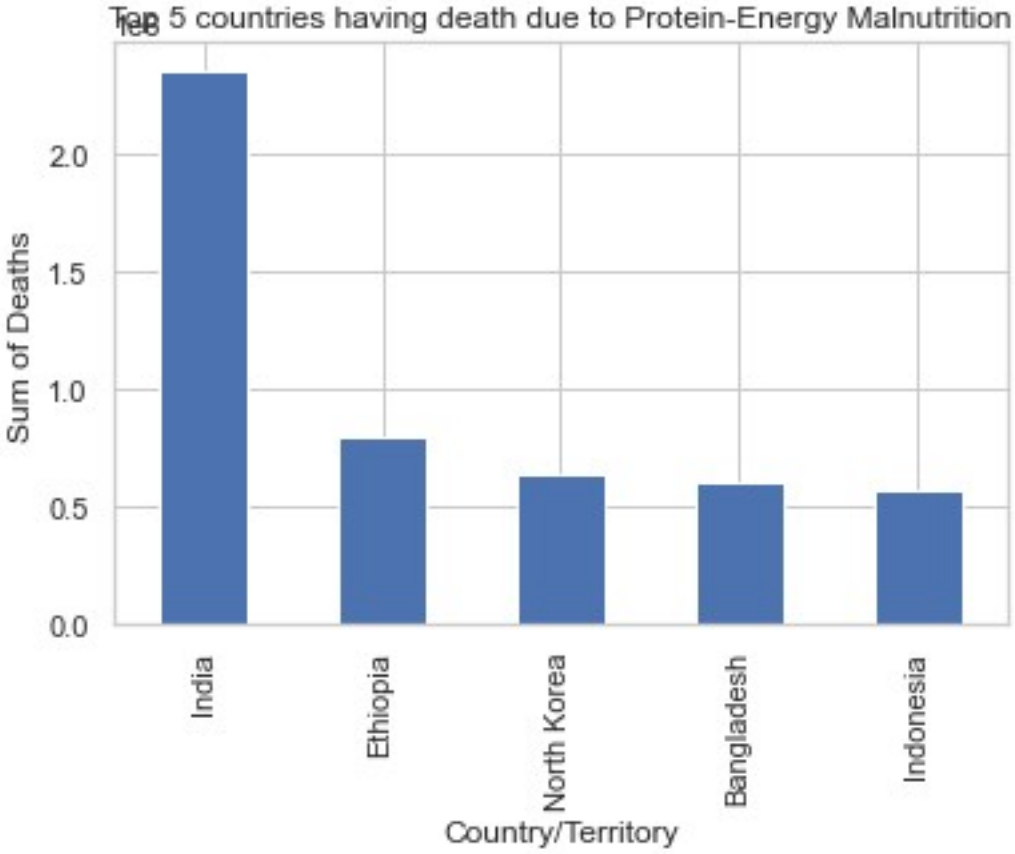

Top 5 countries having death due to Neoplasms
1e7
6
5
4
3
2
1
0
Sum of Deaths
China
United States
India
Japan
Russia
Country/Territory

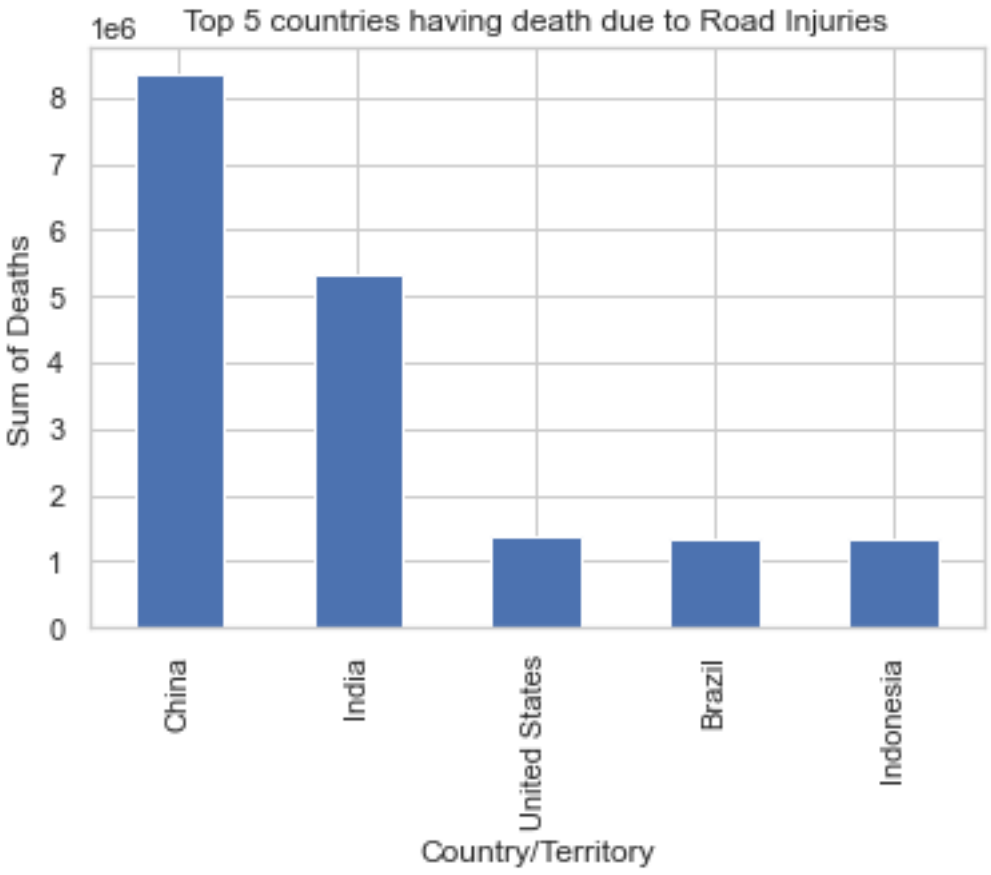

Top 5 countries having death due to Conflict and Terrorism
Sum of Deaths
500000
400000
300000
200000
100000
0
Rwanda
Syria
Iraq
Afghanistan
Burundi
Country/Territory

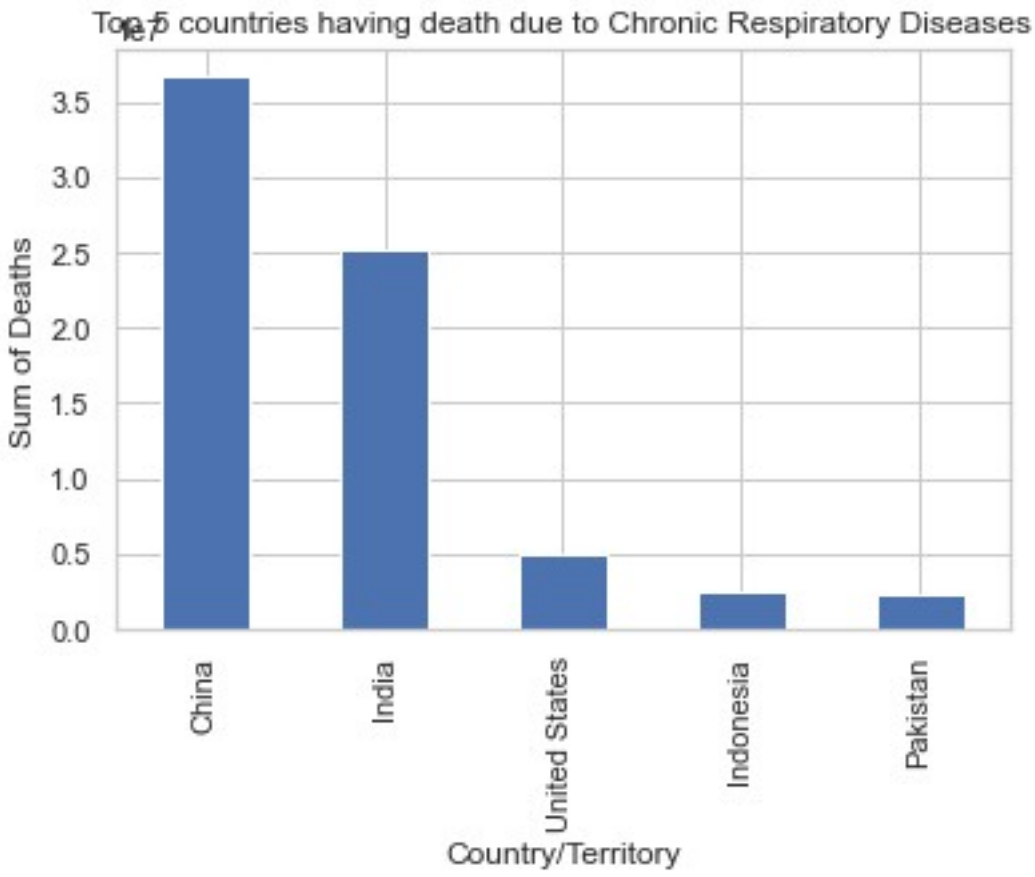

Top 5 countries having death due to Diabetes Mellitus
1e6
5
4
3
2
1
0
Sum of Deaths
India
China
United States
Indonesia
Mexico
Country/Territory

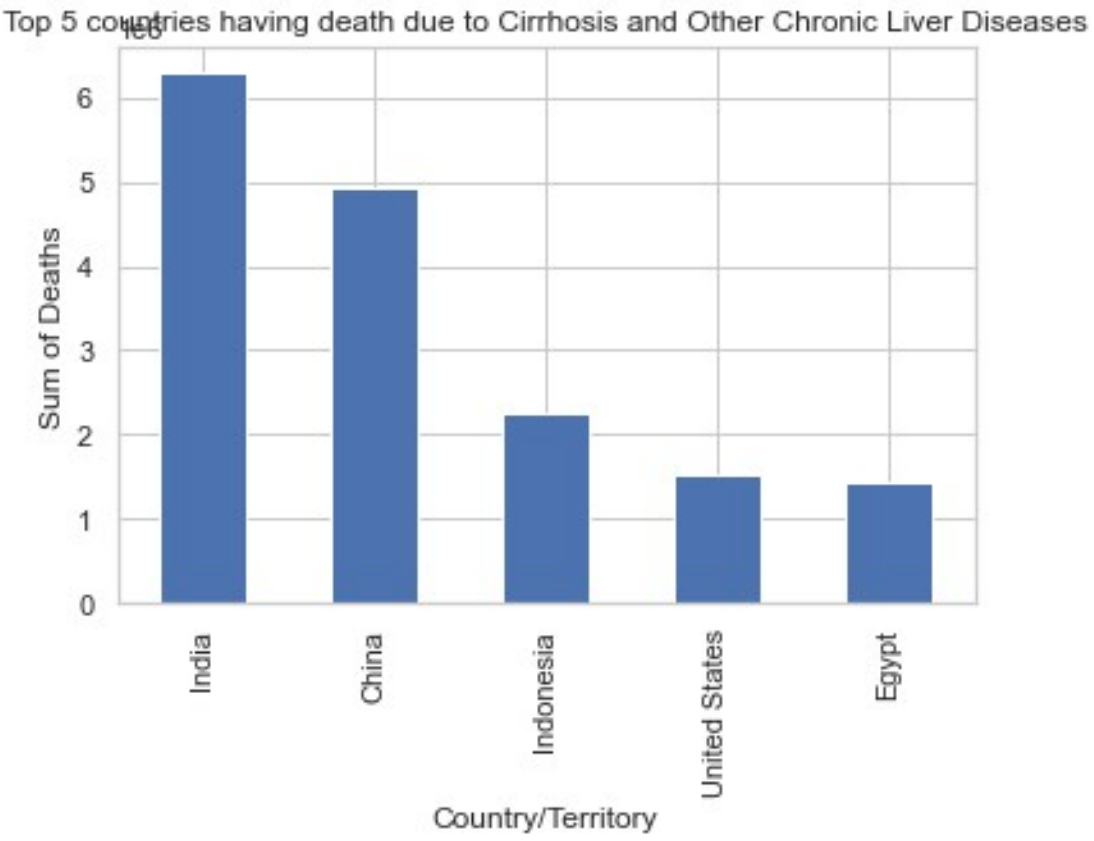

Top 5 countries having death due to Chronic Kidney Disease
1e6
4
3
2
1
0
Sum of Deaths
India
China
United States
Mexico
Japan
Country/Territory

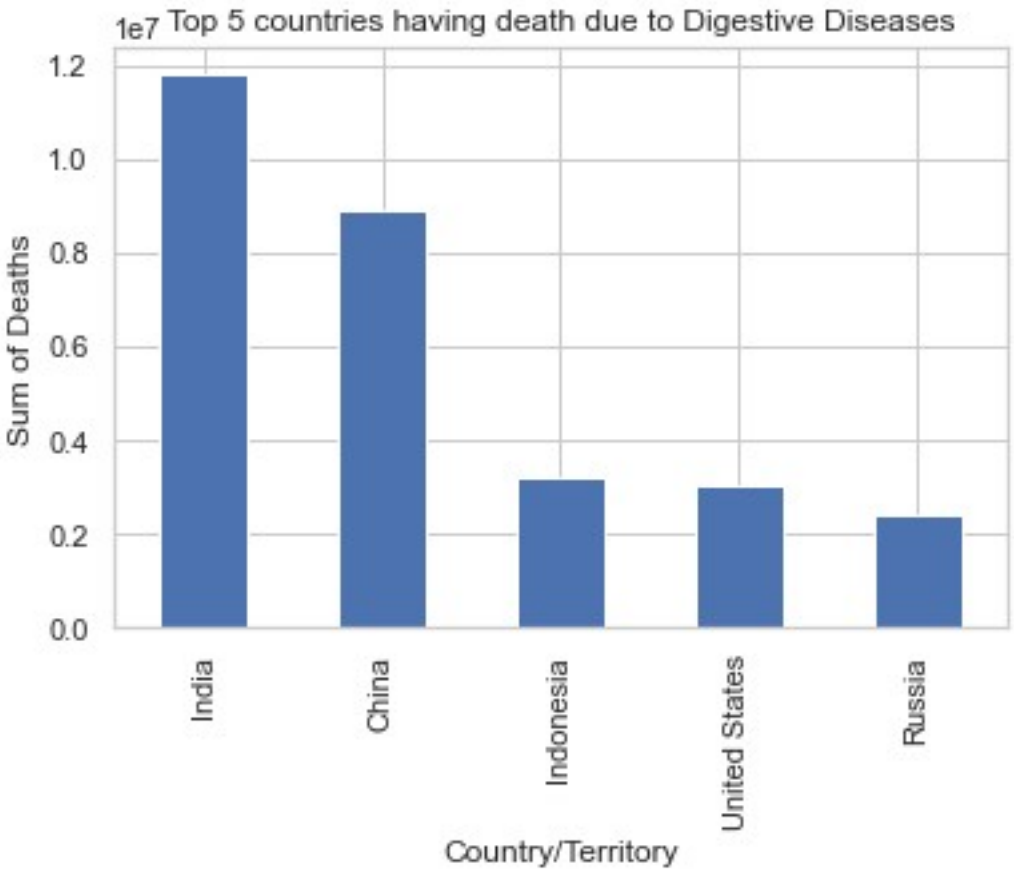

Top 5 countries having death due to Poisonings
800000
700000
600000
500000
400000
300000
200000
100000
0
Sum of Deaths
China
Russia
India
Nigeria
Ethiopia
Country/Territory

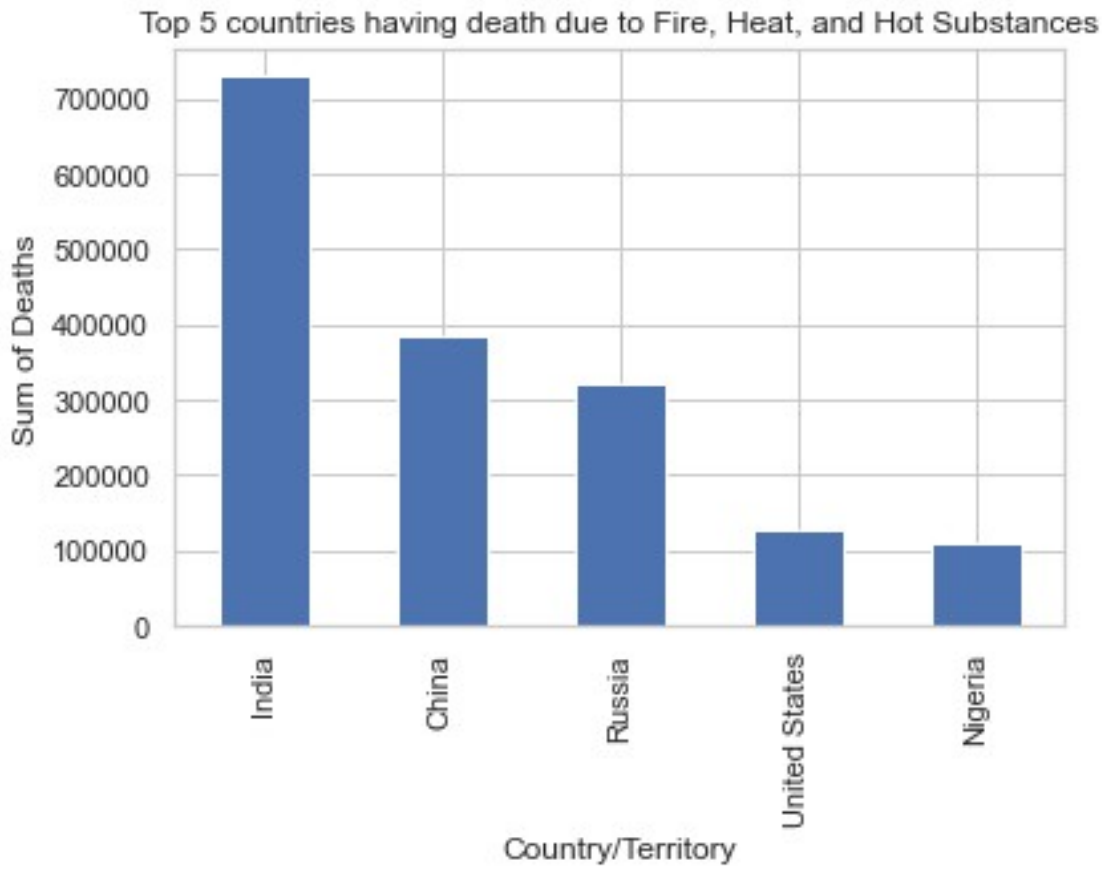

Top 5 countries having death due to Protein-Energy Malnutrition
1e8
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
India
Ethiopia
North Korea
Bangladesh
Indonesia
Country/Territory

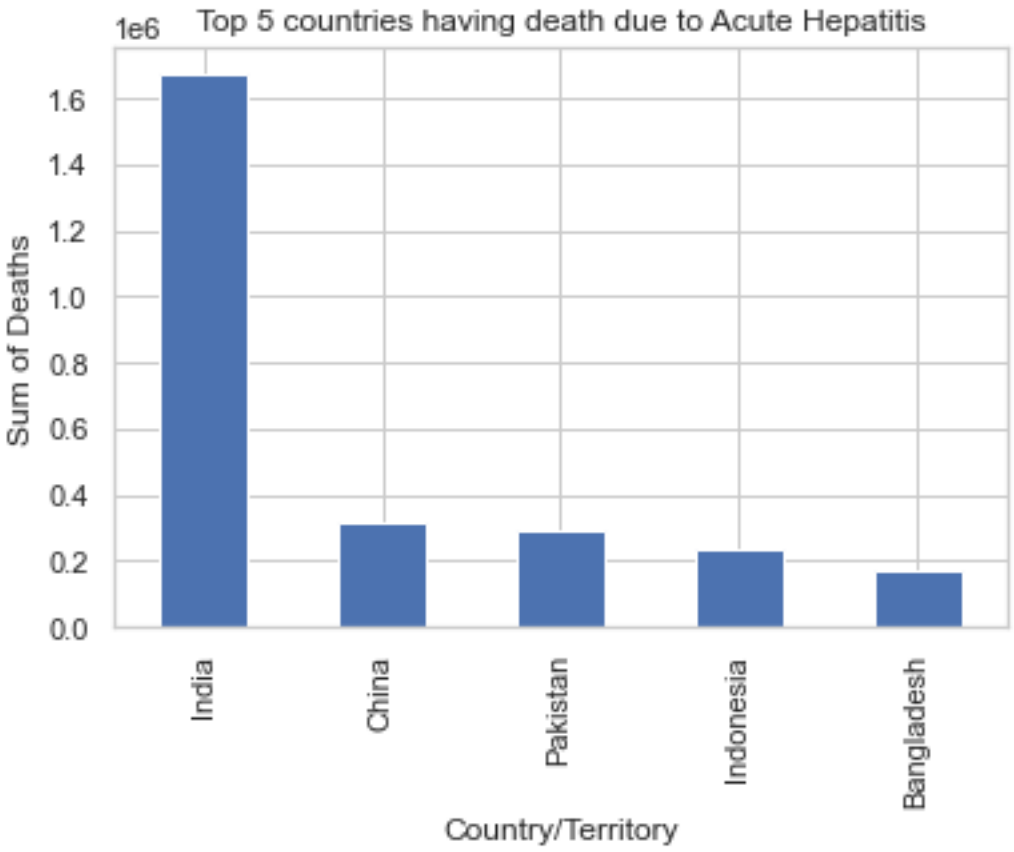

Top 5 countries having death due to Road Injuries
1e6
8
7
6
5
4
3
2
1
0
Sum of Deaths
China
India
United States
Brazil
Indonesia
Country/Territory


Top 5 countries having death due to Chronic Respiratory Diseases
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Sum of Deaths
China
India
United States
Indonesia
Pakistan
Country/Territory


Top 5 countries having death due to Cirrhosis and Other Chronic Liver Diseases
1e6
6
5
4
3
2
1
0
Sum of Deaths
India
China
Indonesia
United States
Egypt
Country/Territory

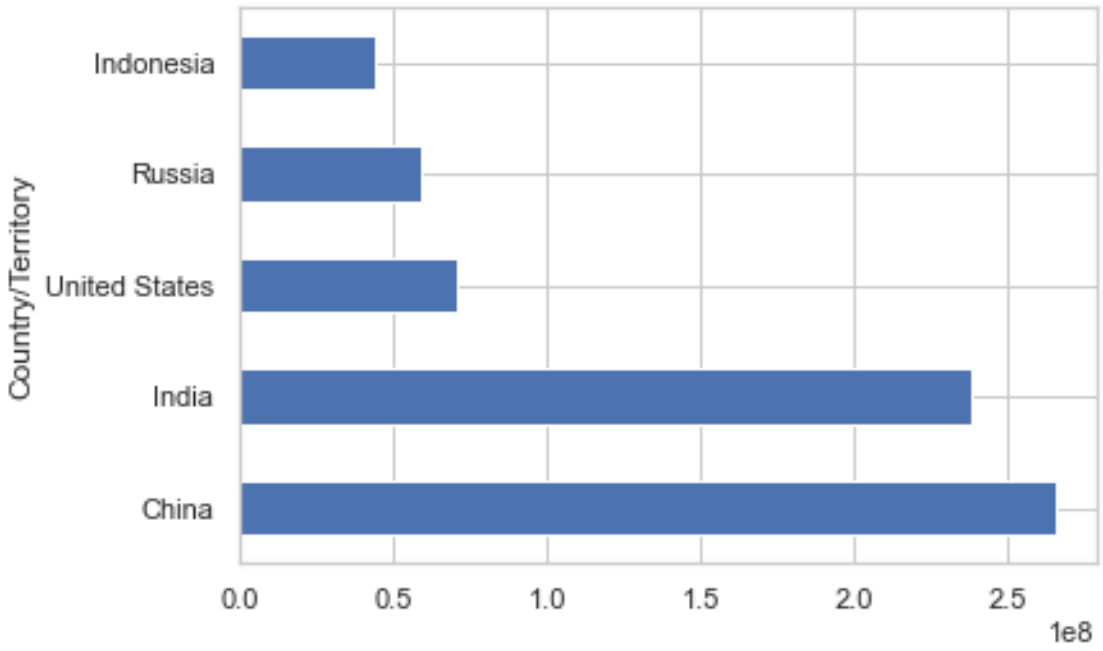

1e7
Top 5 countries having death due to Digestive Diseases
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sum of Deaths
India
China
Indonesia
United States
Russia
Country/Territory


Top 5 countries having death due to Fire, Heat, and Hot Substances
Sum of Deaths
700000
600000
500000
400000
300000
200000
100000
0
India
China
Russia
United States
Nigeria
Country/Territory



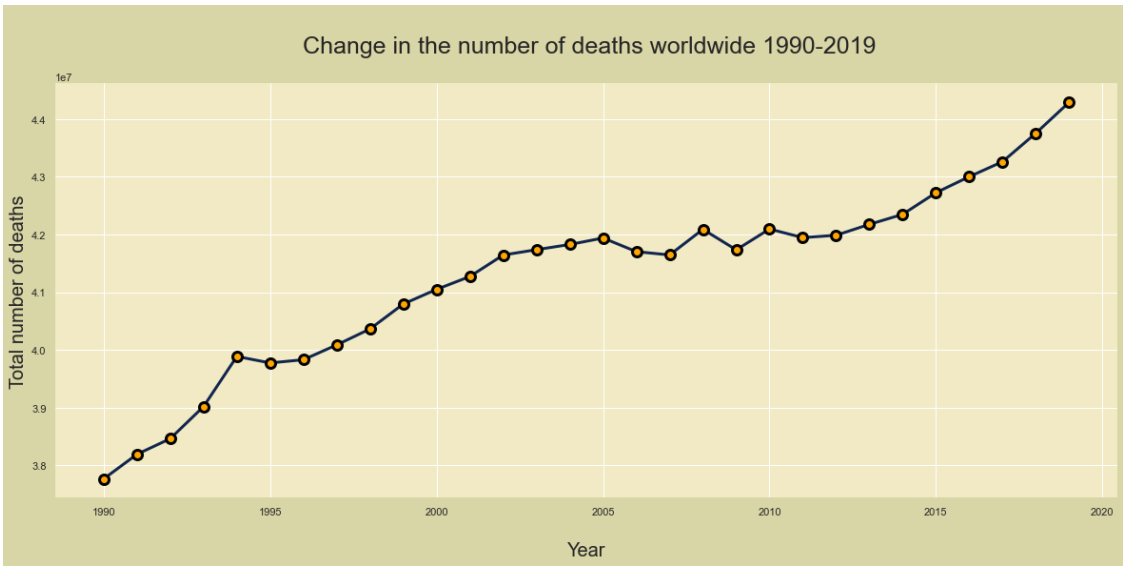

```
# Top countries in different death disease/enviroment  from above
graphs
# India---- Meningits, Nutritional Deficiencies, Maternal Disorders,
Tuberculosis, Lower Respiratory Infections, Self-harm
# Neonatal Disorders,Diarrheal Diseases, Diabetes Mellitus,Choronic
kidney,Protein-energy malnutrition,Digestive,Acute 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()
```

```
<AxesSubplot:ylabel='Country/Territory'>
```

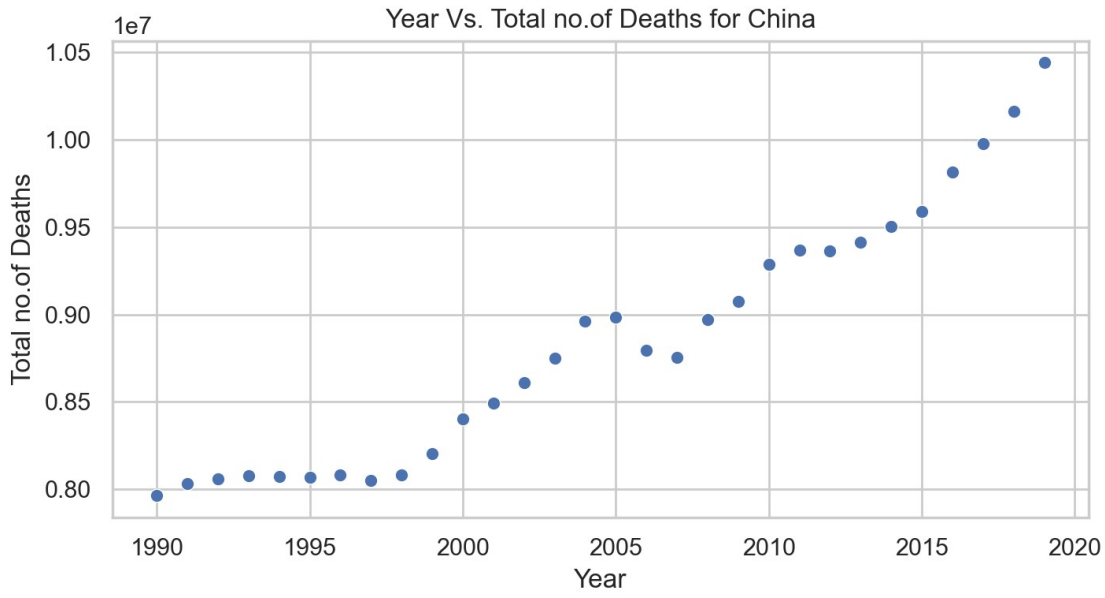


```
# Divide the causes of death into 3 main categories:

communicable_diseases_df = df2[["Year", "Nutritional Deficiencies",
"Malaria", "Maternal Disorders", "HIV/AIDS","Drug Use
Disorders","Tuberculosis","Neonatal Disorders","Alcohol Use
Disorders","Diarrheal Diseases"]]

non_communicable_diseases_df = df2[["Year", "Meningitis","Alzheimer's
Disease and Other Dementias", "Parkinson's Disease",
"Cardiovascular Diseases","Lower Respiratory Infections", "Acute
Hepatitis", "Digestive Diseases", "Cirrhosis and Other Chronic Liver
Diseases",
"Chronic Respiratory Diseases", "Diabetes Mellitus","Chronic Kidney
Disease"]]

injures_df = df2[["Year","Drowning", "Interpersonal Violence", "Fire,
Heat, and Hot Substances", "Road Injuries", "Poisonings" ,
"Protein-Energy Malnutrition", "Conflict and Terrorism", "Self-harm",
"Exposure to Forces of Nature",
"Environmental Heat and Cold Exposure"]]

communicable_diseases_df = df.assign(sumRow =
communicable_diseases_df.sum(axis=1) -
communicable_diseases_df['Year'])
sum_by_year_communicable_diseases_df =
communicable_diseases_df[['Year','sumRow']].groupby('Year').sum().rese
t_index(drop=False)
```

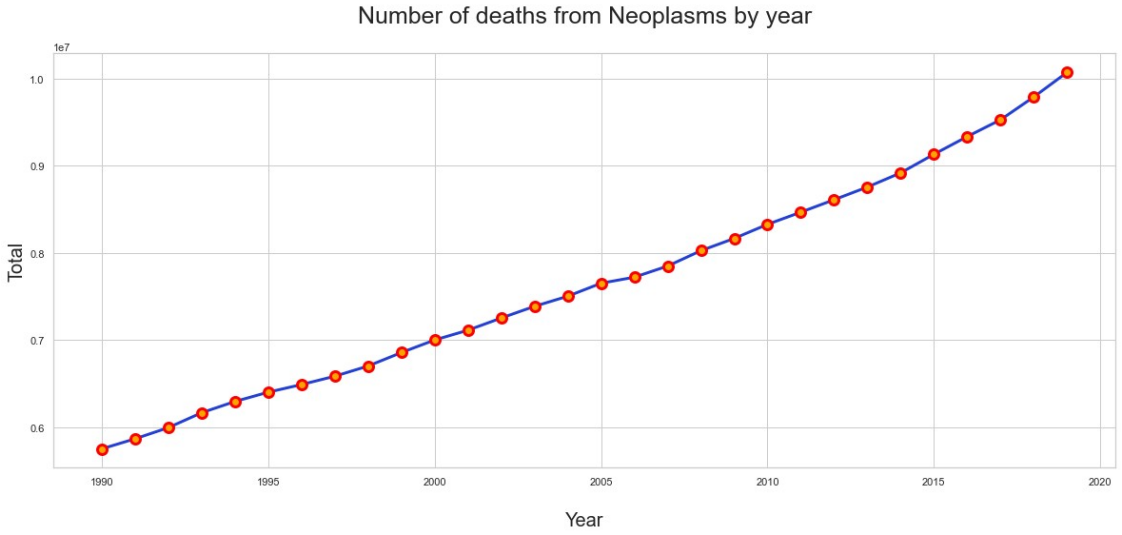
```
non_communicable_diseases_df =
non_communicable_diseases_df.assign(sumRow =
non_communicable_diseases_df.sum(axis=1) -
non_communicable_diseases_df['Year'])
sum_by_year_non_communicable_diseases_df =
non_communicable_diseases_df[['Year','sumRow']].groupby('Year').sum().
reset_index(drop=False)

injures_df = injures_df.assign(sumRow = injures_df.sum(axis=1) -
injures_df['Year'])
sum_by_year_injures_df =
injures_df[['Year','sumRow']].groupby('Year').sum().reset_index(drop=F
alse)

sum_by_year_df =
sum_by_year_communicable_diseases_df.merge(sum_by_year_non_communicabl
e_diseases_df, on='Year').merge(sum_by_year_injures_df,on='Year')
sum_by_year_df.rename(columns={'sumRow_x': 'communicable_diseases',
'sumRow_y': 'non_communicable_diseases', 'sumRow': 'injures'},
inplace=True)

#  "Total_no_of_Deaths" against "Year"

sum_by_year_df["Total"]=sum_by_year_df["communicable_diseases"]
+sum_by_year_df["non_communicable_diseases"]+sum_by_year_df["injures"]
plt.subplots(figsize=(20,8))

p=sns.lineplot(x=sum_by_year_df["Year"] ,y=sum_by_year_df["Total"],dat
a=sum_by_year_df,color="#11264e",marker="o",linewidth=3,markersize=10,
markerfacecolor="orange",markeredgecolor="black",markeredgewidth=3)
p.axes.set_title("\n Change in the number of deaths worldwide 1990-
2019\n",fontsize=25)
p.axes.set_xlabel("\nYear",fontsize=20)
p.axes.set_ylabel("Total number of deaths",fontsize=20)

sns.despine(left=True, bottom=True)
plt.show()
```



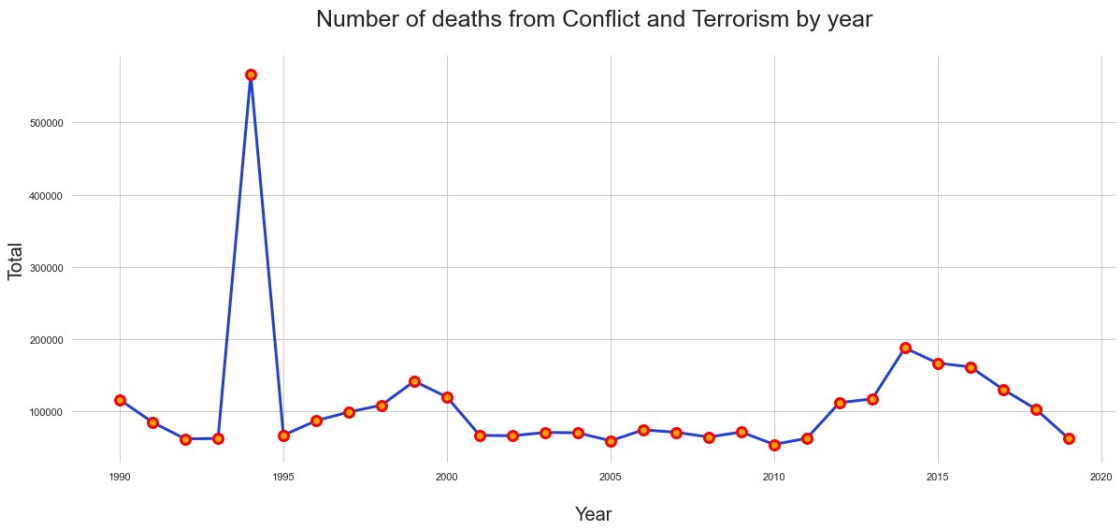
```
# China - "Total_no_of_Deaths" against "Year"

China_Total_no_of_Deaths =
df2[df2['Country/Territory']=='China'].sort_values(by='Total_no_of_Deaths',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();
```



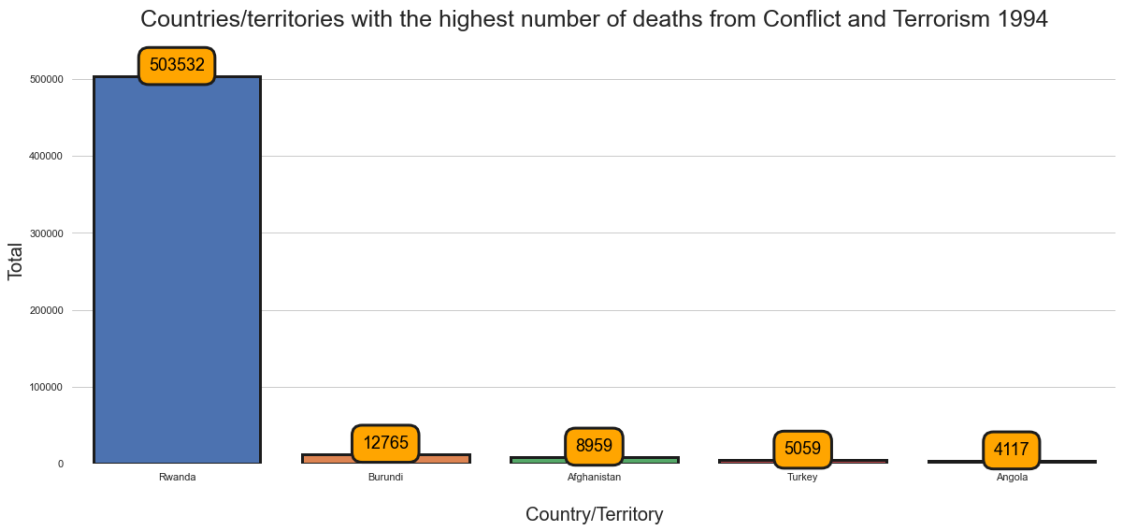

```
# Cancer
Neoplasms=pd.DataFrame(df2.groupby("Year")
["Neoplasms"].sum()).reset_index()
plt.subplots(figsize=(20,8))
sns.set(rc={"axes.facecolor":"#F2EAC5","figure.facecolor":"#D8D5A6"})

p=sns.lineplot(x=Neoplasms["Year"],
               y=Neoplasms["Neoplasms"],
               data=Neoplasms,
               color="#2540D5",
               marker="o",linewidth=3,markersize=10,

markerfacecolor="orange",markeredgecolor="red",markeredgewidth=3)
p.axes.set_title("\nNumber of deaths from Neoplasms by year\
n",fontsize=25)
p.axes.set_xlabel("\nYear",fontsize=20)
p.axes.set_ylabel("Total",fontsize=20);
```



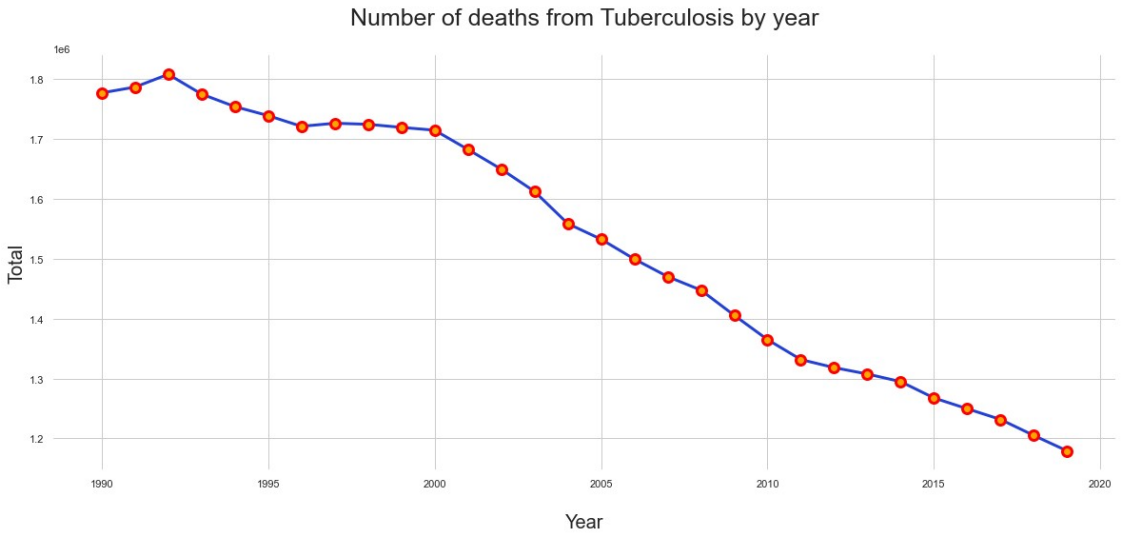

```
# Conflict/Terrorism
Conflict_Terrorism=pd.DataFrame(df2.groupby("Year")["Conflict and
Terrorism"].sum()).reset_index()
plt.subplots(figsize=(20,8))

p=sns.lineplot(x=Conflict_Terrorism["Year"],
               y=Conflict_Terrorism["Conflict and Terrorism"],
               data=Conflict_Terrorism,
               color="#2540D5",
               marker="o",linewidth=3,markersize=10,

markerfacecolor="orange",markeredgecolor="red",markeredgewidth=3)
p.axes.set_title("\nNumber of deaths from Conflict and Terrorism by
year\n",fontsize=25)
p.axes.set_xlabel("\nYear",fontsize=20)
p.axes.set_ylabel("Total",fontsize=20)

sns.despine(left=True, bottom=True)
plt.show()
```



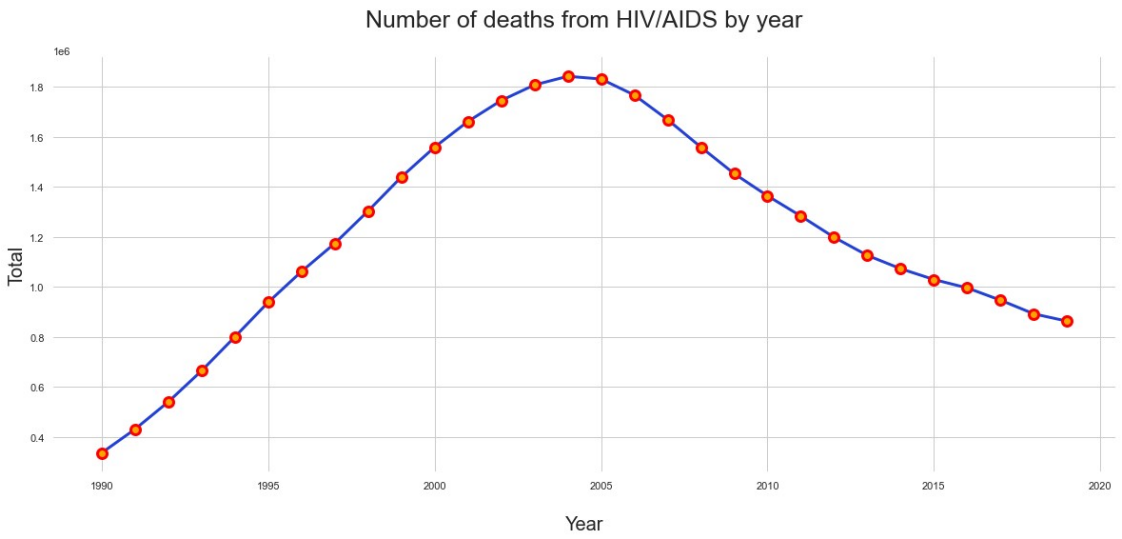

```
temp=pd.DataFrame(df2.groupby("Year")["Conflict and
Terrorism"]).reset_index()
temp=df2[df2["Year"]==1994].sort_values(['Conflict and
Terrorism'],ascending=False)

plt.subplots(figsize=(20,8))
p = sns.barplot(x=temp["Country/Territory"][:5],y=temp["Conflict and
Terrorism"], saturation=1,edgecolor = "#1c1c1c", linewidth = 3)
p.axes.set_title("\nCountries/territories with the highest number of
deaths from Conflict and Terrorism 1994\n",fontsize=25)
plt.xlabel("\nCountry/Territory",fontsize=20)
plt.ylabel("Total",fontsize=20)
plt.xticks(rotation=0)
for container in p.containers:

p.bar_label(container,label_type="edge",padding=2,size=18,color="black
",rotation=0,
    bbox={"boxstyle": "round", "pad": 0.6, "facecolor": "orange",
"edgecolor": "#1c1c1c", "linewidth" : 3, "alpha": 1})

sns.despine(left=True, bottom=True)
plt.show()
```




```
# Number of deaths from Tuberculosis by year
Tuberculosis=pd.DataFrame(df2.groupby("Year")
["Tuberculosis"].sum()).reset_index()
plt.subplots(figsize=(20,8))

p=sns.lineplot(x=Tuberculosis["Year"],
               y=Tuberculosis["Tuberculosis"],
               data=Tuberculosis,
               color="#2540D5",
               marker="o",linewidth=3,markersize=10,

markerfacecolor="orange",markeredgecolor="red",markeredgewidth=3)
p.axes.set_title("\nNumber of deaths from Tuberculosis by year\
n",fontsize=25)
p.axes.set_xlabel("\nYear",fontsize=20)
p.axes.set_ylabel("Total",fontsize=20)

sns.despine(left=True, bottom=True)
plt.show()
```




```
# HIV/AIDS
HIV_AIDS=pd.DataFrame(df2.groupby("Year")["HIV/AIDS"].sum()).reset_ind
ex()
plt.subplots(figsize=(20,8))

p=sns.lineplot(x=HIV_AIDS["Year"],
               y=HIV_AIDS["HIV/AIDS"],
               data=HIV_AIDS,
               color="#2540D5",
               marker="o",linewidth=3,markersize=10,

markerfacecolor="orange",markeredgecolor="red",markeredgewidth=3)
p.axes.set_title("\nNumber of deaths from HIV/AIDS by year\
n",fontsize=25)
p.axes.set_xlabel("\nYear",fontsize=20)
p.axes.set_ylabel("Total",fontsize=20)

sns.despine(left=True, bottom=True)
plt.show()
```




```
# Check correlation

plt.figure(figsize = (200,70))
sns.heatmap(df2.corr(),
annot=True,cmap='summer',mask=np.triu(np.ones_like(df2.corr())))
plt.show()
```