EDA for Morticd9 dataset

You can find this dataset **Morticd9** here.

1. Import libraries

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

2. Import Dataset

In [3]:
```
morticd9 = pd.read_excel("Morticd9_2.xlsx")
```

3. Show Dataset

In [4]:
```
morticd9
```

Out[4]:

	Country	country_name	Year	List	Cause	Sex	Sex_m_f	Frmat	IM_Frmat	Deaths
0	2090	Canada	1991-01-01	09B	B155	2	Female	1	1	6
1	2090	Canada	1991-01-01	09B	B156	1	Male	1	1	
2	2090	Canada	1991-01-01	09B	B156	2	Female	1	1	
3	2090	Canada	1991-01-01	09B	B159	1	Male	1	1	2
4	2090	Canada	1991-01-01	09B	B159	2	Female	1	1	2
...	...	...	...	...	...	...	...	...	...	.
23027	5020	Australia	1997-01-01	09B	B522	1	Male	1	1	
23028	5020	Australia	1997-01-01	09B	B522	2	Female	1	1	
23029	5020	Australia	1997-01-01	09B	B523	1	Male	1	1	7
23030	5020	Australia	1997-01-01	09B	B523	2	Female	1	1	
23031	5020	Australia	1997-01-01	09B	B524	1	Male	1	1	1

23032 rows × 37 columns

4. Basic EDA

- **To get first 5 rows of dataset**

In [6]:
```
morticd9.head()
```

Out[6]:

	Country	country_name	Year	List	Cause	Sex	Sex_m_f	Frmat	IM_Frmat	Deaths1	...

	Country	country_name	Year	List	Cause	Sex	Sex_m_f	Frmat	IM_Frmat	Deaths1	...
0	2090	Canada	1991-01-01	09B	B155	2	Female	1	1	64	...
1	2090	Canada	1991-01-01	09B	B156	1	Male	1	1	0	...
2	2090	Canada	1991-01-01	09B	B156	2	Female	1	1	0	...
3	2090	Canada	1991-01-01	09B	B159	1	Male	1	1	23	...
4	2090	Canada	1991-01-01	09B	B159	2	Female	1	1	21	...

- **To get bottom 5 rows of dataset**

In [7]:
```
morticd9.tail()
```

Out[7]:

	Country	country_name	Year	List	Cause	Sex	Sex_m_f	Frmat	IM_Frmat	Deaths
23027	5020	Australia	1997-01-01	09B	B522	1	Male	1	1	
23028	5020	Australia	1997-01-01	09B	B522	2	Female	1	1	
23029	5020	Australia	1997-01-01	09B	B523	1	Male	1	1	7
23030	5020	Australia	1997-01-01	09B	B523	2	Female	1	1	
23031	5020	Australia	1997-01-01	09B	B524	1	Male	1	1	1

5 rows × 37 columns

In [8]:
```
type(morticd9)
```

Out[8]:
```
pandas.core.frame.DataFrame
```

- **To show all the column names of dataset**

In [9]:
```
morticd9.columns
```

Out[9]:
```
Index(['Country', 'country_name', 'Year', 'List', 'Cause', 'Sex', 'Sex_m_f',
       'Frmat', 'IM_Frmat', 'Deaths1', 'Deaths2', 'Deaths3', 'Deaths4',
       'Deaths5', 'Deaths6', 'Deaths7', 'Deaths8', 'Deaths9', 'Deaths10',
       'Deaths11', 'Deaths12', 'Deaths13', 'Deaths14', 'Deaths15', 'Deaths16',
       'Deaths17', 'Deaths18', 'Deaths19', 'Deaths20', 'Deaths21', 'Deaths22',
       'Deaths23', 'Deaths26', 'IM_Deaths1', 'IM_Deaths2', 'IM_Deaths3',
       'IM_Deaths4'],
      dtype='object')
```

- **To fetch any perticular column from dataset**

In [12]:
```
morticd9['country_name']
```

Out[12]:
```
0           Canada
1           Canada
2           Canada
3           Canada
4           Canada
           ...
23027    Australia
```

```
23028    Australia
23029    Australia
23030    Australia
23031    Australia
```

In [13]:
```
type(morticd9['country_name'])
```

Out[13]:
```
pandas.core.series.Series
```

In [14]:
```
morticd9.shape
```

Out[14]:
```
(23032, 37)
```

In [15]:
```
morticd9.describe()
```

Out[15]:

	Country	Sex	Frmat	IM_Frmat	Deaths1	Deaths2	Deaths3
count	23032.000000	23032.000000	23032.0	23032.0	2.303200e+04	23032.000000	23032.000000
mean	3283.518583	1.513677	1.0	1.0	3.022596e+03	37.646883	3.012331
std	1194.366374	0.499824	0.0	0.0	3.052064e+04	468.978579	33.221217
min	2090.000000	1.000000	1.0	1.0	0.000000e+00	0.000000	0.000000
25%	2150.000000	1.000000	1.0	1.0	4.000000e+00	0.000000	0.000000
50%	2450.000000	2.000000	1.0	1.0	7.300000e+01	0.000000	0.000000
75%	4080.000000	2.000000	1.0	1.0	6.630000e+02	1.000000	0.000000
max	5020.000000	2.000000	1.0	1.0	1.172959e+06	21008.000000	1566.000000

8 rows × 32 columns

- **To know how many unique values are there in each columns of dataset**

In [16]:
```
morticd9.nunique()
```

Out[16]:
```
Country            5
country_name       5
Year               7
List               2
Cause            393
Sex                2
Sex_m_f            2
Frmat              1
IM_Frmat           1
Deaths1         4529
Deaths2          595
Deaths3          166
Deaths4          139
Deaths5          114
Deaths6          114
Deaths7          224
Deaths8          236
Deaths9          357
Deaths10         436
Deaths11         488
```

```
Deaths12         612
Deaths13         720
Deaths14         855
Deaths15         940
Deaths16        1044
Deaths17        1168
Deaths18        1389
Deaths19        1661
Deaths20        1860
Deaths21        1953
Deaths22        2073
Deaths23        2465
Deaths26          67
IM_Deaths1       330
IM_Deaths2       285
IM_Deaths3       237
IM_Deaths4       391
```

In [17]:

```
morticd9.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23032 entries, 0 to 23031
Data columns (total 37 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Country       23032 non-null  int64
 1   country_name  23032 non-null  object
 2   Year          23032 non-null  datetime64[ns]
 3   List          23032 non-null  object
 4   Cause         23032 non-null  object
 5   Sex           23032 non-null  int64
 6   Sex_m_f       23032 non-null  object
 7   Frmat         23032 non-null  int64
 8   IM_Frmat      23032 non-null  int64
 9   Deaths1       23032 non-null  int64
 10  Deaths2       23032 non-null  int64
 11  Deaths3       23032 non-null  int64
 12  Deaths4       23032 non-null  int64
 13  Deaths5       23032 non-null  int64
 14  Deaths6       23032 non-null  int64
 15  Deaths7       23032 non-null  int64
 16  Deaths8       23032 non-null  int64
 17  Deaths9       23032 non-null  int64
 18  Deaths10      23032 non-null  int64
 19  Deaths11      23032 non-null  int64
 20  Deaths12      23032 non-null  int64
 21  Deaths13      23032 non-null  int64
 22  Deaths14      23032 non-null  int64
 23  Deaths15      23032 non-null  int64
 24  Deaths16      23032 non-null  int64
 25  Deaths17      23032 non-null  int64
 26  Deaths18      23032 non-null  int64
 27  Deaths19      23032 non-null  int64
 28  Deaths20      23032 non-null  int64
 29  Deaths21      23032 non-null  int64
 30  Deaths22      23032 non-null  int64
 31  Deaths23      23032 non-null  int64
 32  Deaths26      23032 non-null  int64
 33  IM_Deaths1    23032 non-null  int64
 34  IM_Deaths2    23032 non-null  int64
 35  IM_Deaths3    23032 non-null  int64
 36  IM_Deaths4    23032 non-null  int64
dtypes: datetime64[ns](1), int64(32), object(4)
memory usage: 6.5+ MB
```

- **To know about corelation between each column of dataset**

In [18]:
```
morticd9.corr()
```

Out[18]:

	Country	Sex	Frmat	IM_Frmat	Deaths1	Deaths2	Deaths3	Deaths4	D
Country	1.000000	-0.000310	NaN	NaN	-0.040748	-0.040201	-0.042798	-0.043048	-0
Sex	-0.000310	1.000000	NaN	NaN	-0.004678	-0.012977	-0.010930	-0.014858	-0
Frmat	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
IM_Frmat	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Deaths1	-0.040748	-0.004678	NaN	NaN	1.000000	0.803053	0.907804	0.897442	0
Deaths2	-0.040201	-0.012977	NaN	NaN	0.803053	1.000000	0.875312	0.852521	0
Deaths3	-0.042798	-0.010930	NaN	NaN	0.907804	0.875312	1.000000	0.989985	0
Deaths4	-0.043048	-0.014858	NaN	NaN	0.897442	0.852521	0.989985	1.000000	0
Deaths5	-0.042810	-0.017437	NaN	NaN	0.893351	0.839602	0.979104	0.991970	1
Deaths6	-0.043251	-0.015501	NaN	NaN	0.891631	0.827980	0.970387	0.985403	0
Deaths7	-0.044508	-0.017150	NaN	NaN	0.883810	0.813688	0.958198	0.975210	0
Deaths8	-0.041564	-0.022998	NaN	NaN	0.868249	0.800461	0.941320	0.963140	0
Deaths9	-0.033647	-0.031491	NaN	NaN	0.766557	0.719023	0.846652	0.880816	0
Deaths10	-0.030744	-0.036894	NaN	NaN	0.767662	0.723276	0.843066	0.874366	0
Deaths11	-0.034723	-0.036046	NaN	NaN	0.820130	0.763570	0.881170	0.903570	0
Deaths12	-0.036935	-0.034095	NaN	NaN	0.832123	0.762980	0.874427	0.892620	0
Deaths13	-0.039196	-0.031865	NaN	NaN	0.870535	0.779044	0.886791	0.900721	0
Deaths14	-0.040077	-0.031289	NaN	NaN	0.912446	0.794413	0.899167	0.909018	0
Deaths15	-0.040666	-0.029380	NaN	NaN	0.944610	0.796543	0.898018	0.904054	0
Deaths16	-0.042547	-0.027037	NaN	NaN	0.959673	0.796145	0.894818	0.898088	0
Deaths17	-0.042012	-0.026419	NaN	NaN	0.965649	0.796153	0.893466	0.893851	0
Deaths18	-0.041235	-0.025729	NaN	NaN	0.968205	0.795618	0.892641	0.891234	0
Deaths19	-0.041521	-0.022253	NaN	NaN	0.977230	0.796929	0.895080	0.892293	0
Deaths20	-0.042154	-0.016642	NaN	NaN	0.988879	0.796579	0.896718	0.891719	0
Deaths21	-0.043534	-0.008203	NaN	NaN	0.995474	0.793983	0.896671	0.887749	0
Deaths22	-0.038443	0.004127	NaN	NaN	0.991682	0.771114	0.877164	0.860834	0
Deaths23	-0.032804	0.029360	NaN	NaN	0.925016	0.686763	0.791118	0.763143	0
Deaths26	-0.045387	-0.033163	NaN	NaN	0.823359	0.778621	0.879140	0.893368	0
IM_Deaths1	-0.036790	-0.008430	NaN	NaN	0.684300	0.943339	0.739362	0.722142	0
IM_Deaths2	-0.036427	-0.015267	NaN	NaN	0.742553	0.958243	0.814610	0.789409	0
IM_Deaths3	-0.039049	-0.012935	NaN	NaN	0.787456	0.972312	0.860173	0.834738	0
IM_Deaths4	-0.037984	-0.014858	NaN	NaN	0.811292	0.888767	0.889359	0.866430	0

32 rows × 32 columns

5. To chceck missing, incomplete, or NULL value in each column

In [19]:
```
morticd9.isnull().sum()
```

Out[19]:
```
Country         0
country_name    0
Year            0
List            0
Cause           0
Sex             0
Sex_m_f         0
Frmat           0
IM_Frmat        0
Deaths1         0
Deaths2         0
Deaths3         0
Deaths4         0
Deaths5         0
Deaths6         0
Deaths7         0
Deaths8         0
Deaths9         0
Deaths10        0
Deaths11        0
Deaths12        0
Deaths13        0
Deaths14        0
Deaths15        0
Deaths16        0
Deaths17        0
Deaths18        0
Deaths19        0
Deaths20        0
Deaths21        0
Deaths22        0
Deaths23        0
Deaths26        0
IM_Deaths1      0
IM_Deaths2      0
IM_Deaths3      0
IM_Deaths4      0
dtype: int64
```

6. To check outliers in the dataset

In [66]:
```
morticd9.plot(kind="box", subplots=True, layout=(8,4), figsize=(30,30))
```

Out[66]:
```
Country          AxesSubplot(0.125,0.799681;0.168478x0.0803191)
Sex            AxesSubplot(0.327174,0.799681;0.168478x0.0803191)
Frmat          AxesSubplot(0.529348,0.799681;0.168478x0.0803191)
IM_Frmat       AxesSubplot(0.731522,0.799681;0.168478x0.0803191)
Deaths1          AxesSubplot(0.125,0.703298;0.168478x0.0803191)
Deaths2        AxesSubplot(0.327174,0.703298;0.168478x0.0803191)
Deaths3        AxesSubplot(0.529348,0.703298;0.168478x0.0803191)
Deaths4        AxesSubplot(0.731522,0.703298;0.168478x0.0803191)
Deaths5          AxesSubplot(0.125,0.606915;0.168478x0.0803191)
```

```
Deaths6       AxesSubplot(0.327174,0.606915;0.168478x0.0803191)
Deaths7       AxesSubplot(0.529348,0.606915;0.168478x0.0803191)
Deaths8       AxesSubplot(0.731522,0.606915;0.168478x0.0803191)
Deaths9          AxesSubplot(0.125,0.510532;0.168478x0.0803191)
Deaths10      AxesSubplot(0.327174,0.510532;0.168478x0.0803191)
Deaths11      AxesSubplot(0.529348,0.510532;0.168478x0.0803191)
Deaths12      AxesSubplot(0.731522,0.510532;0.168478x0.0803191)
Deaths13         AxesSubplot(0.125,0.414149;0.168478x0.0803191)
Deaths14      AxesSubplot(0.327174,0.414149;0.168478x0.0803191)
Deaths15      AxesSubplot(0.529348,0.414149;0.168478x0.0803191)
Deaths16      AxesSubplot(0.731522,0.414149;0.168478x0.0803191)
Deaths17         AxesSubplot(0.125,0.317766;0.168478x0.0803191)
Deaths18      AxesSubplot(0.327174,0.317766;0.168478x0.0803191)
Deaths19      AxesSubplot(0.529348,0.317766;0.168478x0.0803191)
Deaths20      AxesSubplot(0.731522,0.317766;0.168478x0.0803191)
Deaths21         AxesSubplot(0.125,0.221383;0.168478x0.0803191)
Deaths22      AxesSubplot(0.327174,0.221383;0.168478x0.0803191)
Deaths23      AxesSubplot(0.529348,0.221383;0.168478x0.0803191)
Deaths26      AxesSubplot(0.731522,0.221383;0.168478x0.0803191)
IM_Deaths1          AxesSubplot(0.125,0.125;0.168478x0.0803191)
IM_Deaths2       AxesSubplot(0.327174,0.125;0.168478x0.0803191)
IM_Deaths3       AxesSubplot(0.529348,0.125;0.168478x0.0803191)
IM_Deaths4       AxesSubplot(0.731522,0.125;0.168478x0.0803191)
```