Exploratory Data Analysis on Dataset -Terrorism

As a security/defense analyst,try to find out the hot zone of terrorism.reading and analyzing the dataset:

Author Name - Sakshi Itnare

Link for the dataset:https://bit.ly/2TK5Xn5 (https://bit.ly/2TK5Xn5) ¶

Importing the dependencies

In [1]:

```
import numpy as pd
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
from sklearn.preprocessing import LabelEncoder
```

Loading the dataset

In [4]:

```
dataset=pd.read_csv("C:/Users/sakshi itnare/Downloads/Global Terrorism - START data.
```

In [6]:

```
1 dataset
```

Out[6]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	count
0	197000000001	1970	7	2	NaN	0	NaN	58	Dom Re
1	197000000002	1970	0	0	NaN	0	NaN	130	M
2	197001000001	1970	1	0	NaN	0	NaN	160	Phili
3	197001000002	1970	1	0	NaN	0	NaN	78	C
4	197001000003	1970	1	0	NaN	0	NaN	101	
...	...	...	...	...	...	...	...	...	
181686	201712310022	2017	12	31	NaN	0	NaN	182	S
181687	201712310029	2017	12	31	NaN	0	NaN	200	
181688	201712310030	2017	12	31	NaN	0	NaN	160	Phili
181689	201712310031	2017	12	31	NaN	0	NaN	92	
181690	201712310032	2017	12	31	NaN	0	NaN	160	Phili

181691 rows × 135 columns

Exploratory Data Analysis

In [7]:

```
1 dataset.head()
```

Out[7]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan

5 rows × 135 columns

In [8]:

```
1 dataset.tail()
```

Out[8]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	count
181686	201712310022	2017	12	31	NaN	0	NaN	182	S
181687	201712310029	2017	12	31	NaN	0	NaN	200	
181688	201712310030	2017	12	31	NaN	0	NaN	160	Phili
181689	201712310031	2017	12	31	NaN	0	NaN	92	
181690	201712310032	2017	12	31	NaN	0	NaN	160	Phili

5 rows × 135 columns

Printing the shape and features of dataset

In [9]:

```
1 dataset.shape
```

Out[9]:

```
(181691, 135)
```

In [10]:

```
dataset.columns
```

Out[10]:

```
Index(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended',
       'resolution', 'country', 'country_txt', 'region',
       ...
       'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG',
       'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'],
      dtype='object', length=135)
```

In [12]:

```
for columns in dataset.columns:
    print(columns)
```

```
eventid
iyear
imonth
iday
approxdate
extended
resolution
country
country_txt
region
region_txt
provstate
city
latitude
longitude
specificity
vicinity
location
summary
```

Feature Selection

In [17]:

```
data=dataset[["eventid","iyear","imonth","iday","gname","country_txt","provstate","r
              "longitude","latitude","target1","nkill","nwound",
              "summary","gname","motive"]]
```

In [18]:

```
1 data.head()
```

Out[18]:

	eventid	iyear	imonth	iday	gname	country_txt	provstate	region_txt	longitu
0	197000000001	1970	7	2	MANO-D	Dominican Republic	NaN	Central America & Caribbean	-69.951
1	197000000002	1970	0	0	23rd of September Communist League	Mexico	Federal	North America	-99.086
2	197001000001	1970	1	0	Unknown	Philippines	Tarlac	Southeast Asia	120.599
3	197001000002	1970	1	0	Unknown	Greece	Attica	Western Europe	23.762
4	197001000003	1970	1	0	Unknown	Japan	Fukouka	East Asia	130.396

In [19]:

```
1 data.shape
```

Out[19]:

```
(181691, 16)
```

In [20]:

```
1 data.isnull()
```

Out[20]:

	eventid	iyear	imonth	iday	gname	country_txt	provstate	region_txt	longitude
0	False	False	False	False	False	False	True	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
181686	False	False	False	False	False	False	False	False	False
181687	False	False	False	False	False	False	False	False	False
181688	False	False	False	False	False	False	False	False	False
181689	False	False	False	False	False	False	False	False	False
181690	False	False	False	False	False	False	False	False	False

181691 rows × 16 columns

In [21]:

```
1 data.isnull().sum()
```

Out[21]:

```
eventid            0
iyear              0
imonth             0
iday               0
gname              0
country_txt        0
provstate        421
region_txt         0
longitude       4557
latitude        4556
target1          636
nkill          10313
nwound         16311
summary        66129
gname              0
motive        131130
dtype: int64
```

In [25]:

```
1 data.isnull().sum()
```

Out[25]:

```
eventid            0
iyear              0
imonth             0
iday               0
gname              0
country_txt        0
provstate        421
region_txt         0
longitude       4557
latitude        4556
target1          636
nkill          10313
nwound         16311
summary        66129
gname              0
motive        131130
dtype: int64
```

Displaying the shape and info of the data

In [26]:

```
1 data.shape
```

Out[26]:

```
(181691, 16)
```

In [28]:

```
1 data.describe()
```

Out[28]:

	eventid	iyear	imonth	iday	longitude	lati
count	1.816910e+05	181691.000000	181691.000000	181691.000000	1.771340e+05	177135.00
mean	2.002705e+11	2002.638997	6.467277	15.505644	-4.586957e+02	23.49
std	1.325957e+09	13.259430	3.388303	8.814045	2.047790e+05	18.56
min	1.970000e+11	1970.000000	0.000000	0.000000	-8.618590e+07	-53.15
25%	1.991021e+11	1991.000000	4.000000	8.000000	4.545640e+00	11.51
50%	2.009022e+11	2009.000000	6.000000	15.000000	4.324651e+01	31.46
75%	2.014081e+11	2014.000000	9.000000	23.000000	6.871033e+01	34.68
max	2.017123e+11	2017.000000	12.000000	31.000000	1.793667e+02	74.63

In [29]:

```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 16 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   eventid      181691 non-null  int64
 1   iyear        181691 non-null  int64
 2   imonth       181691 non-null  int64
 3   iday         181691 non-null  int64
 4   gname        181691 non-null  object
 5   country_txt  181691 non-null  object
 6   provstate    181270 non-null  object
 7   region_txt   181691 non-null  object
 8   longitude    177134 non-null  float64
 9   latitude     177135 non-null  float64
 10  target1      181055 non-null  object
 11  nkill        171378 non-null  float64
 12  nwound       165380 non-null  float64
 13  summary      115562 non-null  object
 14  gname        181691 non-null  object
 15  motive       50561 non-null   object
dtypes: float64(4), int64(4), object(8)
memory usage: 22.2+ MB
```

Correlation and Heatmap

In [30]:

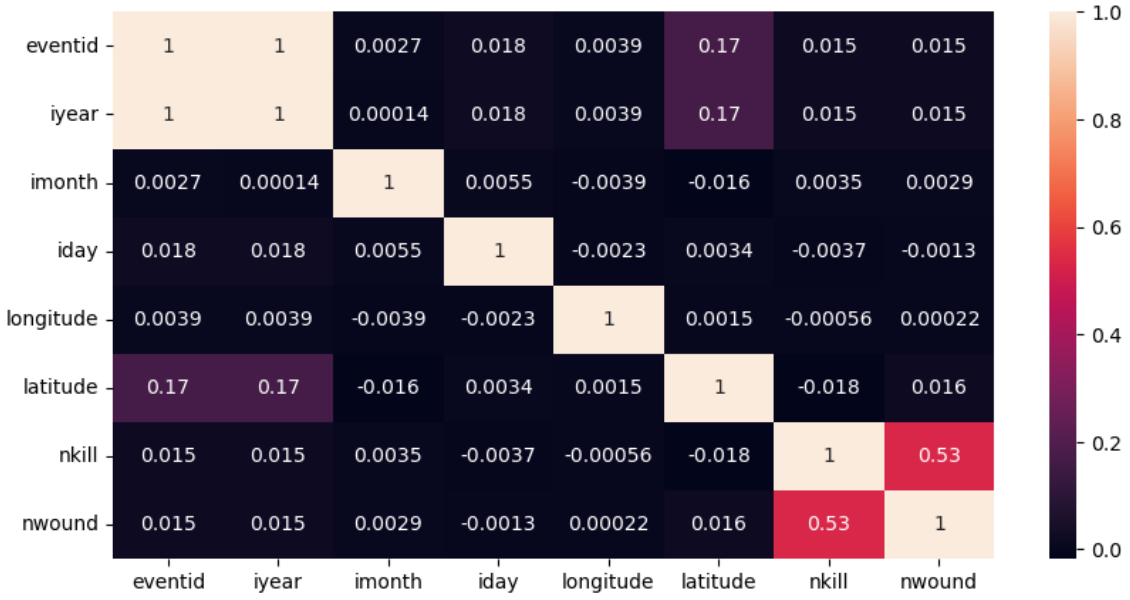
```
1 data.corr()
```

Out[30]:

	eventid	iyear	imonth	iday	longitude	latitude	nkill	nwound
eventid	1.000000	0.999996	0.002706	0.018336	0.003907	0.166886	0.015351	0.015282
iyear	0.999996	1.000000	0.000139	0.018254	0.003917	0.166933	0.015341	0.015273
imonth	0.002706	0.000139	1.000000	0.005497	-0.003880	-0.015978	0.003463	0.002938
iday	0.018336	0.018254	0.005497	1.000000	-0.002285	0.003423	-0.003693	-0.001268
longitude	0.003907	0.003917	-0.003880	-0.002285	1.000000	0.001463	-0.000562	0.000223
latitude	0.166886	0.166933	-0.015978	0.003423	0.001463	1.000000	-0.018124	0.015988
nkill	0.015351	0.015341	0.003463	-0.003693	-0.000562	-0.018124	1.000000	0.534375
nwound	0.015282	0.015273	0.002938	-0.001268	0.000223	0.015988	0.534375	1.000000

In [31]:

```
1 plt.figure(figsize=(10,5))
2 sns.heatmap(data.corr(),annot=True)
3 plt.show()
```



Number of unique values

In [32]:

```
1 data.nunique()
```

Out[32]:

```
eventid        181691
iyear              47
imonth             13
iday               32
gname            3537
country_txt       205
provstate        2855
region_txt         12
longitude       48039
latitude        48322
target1         86006
nkill             205
nwound            238
summary        112492
gname            3537
motive          14490
dtype: int64
```