

Name - ARYAN BAJAJ

Task 4 - Exploratory Data Analysis - Terrorism (Level - Intermediate)

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [2]: # Reading data from the link
url = "https://bit.ly/2TK5Xn5"
data = pd.read_csv('C:/Users/HP/Downloads/globalterrorismdb_0718dist.csv', encoding='latin1')
```

E:\Anaconda\lib\site-packages\IPython\core\interactiveshell.py:3146: DtypeWarning: Columns (4,6,31,33,61,62,63,76,79,90,92,94,96,114,115,121) have mixed types.Specify dtype option on import or set low_memory=False.
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,

Understanding the Data

```
In [3]: data.head(5)
```

```
Out[3]:
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource
0	1970000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS
1	1970000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS
2	1970010000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS
3	1970010000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS
4	1970010000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS

5 rows × 135 columns



In [4]: data.tail(5)

Out[4]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2
181686	201712310022	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab Militants Attack Army Che...	"Highlights: Somalia Daily Media Highlights 2 ...
181687	201712310029	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Putin's 'victory' in Syria has turned into a ...	"Two Russian soldiers killed at Hmeymim base i...
181688	201712310030	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...	NaN
181689	201712310031	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...	NaN
181690	201712310032	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...	"Security tightened in Cotabato City," Manila ...

5 rows × 135 columns

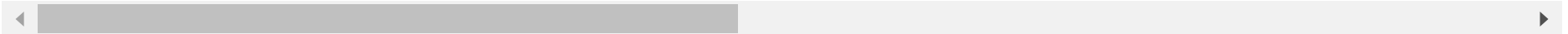


In [5]: `data.describe()`

Out[5]:

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude	
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.968501	7.160938	23.498343	-4.586957e+02	
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.414535	2.933408	18.569242	2.047790e+05	
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-53.154613	-8.618590e+07	
25%	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.000000	5.000000	11.510046	4.545640e+00	
50%	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.000000	6.000000	31.467463	4.324651e+01	
75%	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.000000	10.000000	34.685087	6.871033e+01	
max	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.000000	12.000000	74.633553	1.793667e+02	

8 rows × 77 columns



In [6]: `data.shape`

Out[6]: (181691, 135)

In [7]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(55), int64(22), object(58)
memory usage: 187.1+ MB
```

```
In [8]: data.dtypes
```

```
Out[8]: eventid      int64
        iyear       int64
        imonth      int64
        iday        int64
        approxdate   object
        ...
        INT_LOG      int64
        INT_IDEO     int64
        INT_MISC     int64
        INT_ANY      int64
        related      object
        Length: 135, dtype: object
```

```
In [9]: data.nunique()
```

```
Out[9]: eventid      181691
        iyear        47
        imonth       13
        iday         32
        approxdate   2244
        ...
        INT_LOG       3
        INT_IDEO      3
        INT_MISC      3
        INT_ANY       3
        related      14306
        Length: 135, dtype: int64
```

```
In [10]: data.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)
```

Cleaning the Data

In [11]: *# Finding all the NULL Values*

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

```
Out[11]: eventid          0
         iyear           0
         imonth          0
         iday            0
         approxdate    172452
         ...
         INT_LOG         0
         INT_IDEO        0
         INT_MISC        0
         INT_ANY         0
         related        156653
         Length: 135, dtype: int64
```

In [12]: `data.drop(['approxdate', 'related'], axis=1,inplace = True)`

In [13]: `data.isnull().sum()`

```
Out[13]: eventid          0
         iyear           0
         imonth          0
         iday            0
         extended        0
         ..
         dbsource        0
         INT_LOG         0
         INT_IDEO        0
         INT_MISC        0
         INT_ANY         0
         Length: 133, dtype: int64
```

Relationship Analysis

```
In [14]: correlation = data.corr()
correlation
```

Out[14]:

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude	specificity	...	ransomamt	ransomamtus
eventid	1.000000	0.999996	0.002706	0.018336	0.091761	-0.135039	0.401371	0.166886	0.003907	0.030641	...	-0.009990	-0.018001
iyear	0.999996	1.000000	0.000139	0.018254	0.091754	-0.135023	0.401384	0.166933	0.003917	0.030626	...	-0.009984	-0.018216
imonth	0.002706	0.000139	1.000000	0.005497	-0.000468	-0.006305	-0.002999	-0.015978	-0.003880	0.003621	...	-0.000710	0.046989
iday	0.018336	0.018254	0.005497	1.000000	-0.004700	0.003468	0.009710	0.003423	-0.002285	-0.006991	...	0.012755	-0.010502
extended	0.091761	0.091754	-0.000468	-0.004700	1.000000	-0.020466	0.038389	-0.024749	0.000523	0.057897	...	-0.008114	0.028177
...
nreleased	-0.181612	-0.181556	-0.011535	0.001765	-0.192155	-0.044331	-0.149511	0.002790	-0.017745	-0.030631	...	0.054571	0.034843
INT_LOG	-0.143600	-0.143601	-0.002302	-0.001540	0.071768	0.069904	-0.082584	-0.099827	0.002272	0.073022	...	0.035821	0.031079
INT_IDEO	-0.133252	-0.133253	-0.002034	-0.001621	0.075147	0.067564	-0.071917	-0.094470	0.002268	0.071333	...	0.039053	0.041983
INT_MISC	-0.077852	-0.077847	-0.002554	-0.002027	0.027335	0.207281	0.043139	0.097652	0.000371	-0.019197	...	0.023815	0.125162
INT_ANY	-0.175605	-0.175596	-0.006336	-0.001199	0.080767	0.153118	-0.047900	-0.041530	0.002497	0.061389	...	0.028054	0.053484

77 rows × 77 columns



```
In [15]: import matplotlib.pyplot as plt
plt.figure(figsize=(8,4))
sns.heatmap(correlation,cmap='Blues',annot=False)
```

Out[15]: <AxesSubplot:>

