

# EXPLORATORY DATA ANALYSIS ON DATASET TERRORISM

## DONE BY

## S.IRFAAN MUSTAFA

## Importing the necessary libraries

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## importing the dataset

```
In [4]: df=pd.read_csv("C:/Users/Irfaan/Desktop/globalterrorismdb_0718dist.csv")

C:\Users\Irfaan\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: 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_raise = await self.run_ast_nodes(code_ast.body, cell_name,
```

## printing the first five rows of the dataset

```
In [5]: df.head()

Out[5]:
```

	eventid	year	month	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related	
0	197000000001	1970	7	2		NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	197000000002	1970	0	0		NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	1	1	1	NaN
2	197001000001	1970	1	0		NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
3	197001000002	1970	1	0		NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
4	197001000003	1970	1	0		NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN

5 rows × 23 columns

## checking the no of rows and columns in the dataset

```
In [6]: df.shape

Out[6]: (181691, 135)
```

## checking if any null values are present

```
In [7]: df.isnull().sum()

Out[7]:
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
```

## checking if any duplicate values are present

```
In [8]: ndf=df.drop_duplicates()
ndf
# there are no duplicate values present

Out[8]:
```

	eventid	year	month	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related	
0	197000000001	1970	7	2		NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	197000000002	1970	0	0		NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	1	1	1	NaN
2	197001000001	1970	1	0		NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
3	197001000002	1970	1	0		NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
4	197001000003	1970	1	0		NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	...	START Primary Collection	0	0	0	0	NaN
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 Hmeynim base l...	"Two Russian servicemen killed in Syria mortar...	START Primary Collection	-9	-9	1	1	NaN
181688	201712310030	2017	12	31		NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phill...	NaN	NaN	START Primary Collection	0	0	0	0	NaN
181689	201712310031	2017	12	31		NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Inghal" Bus...	NaN	NaN	START Primary Collection	-9	-9	0	-9	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...	NaN	START Primary Collection	-9	-9	0	-9	NaN

181691 rows × 23 columns

## Renaming the columns for better understanding

```
In [49]: ndf.rename(columns={'eventid':'id','iyear':'year','imonth':'month','iday':'day','approxdate':'date','nkill':'killed','nwound':'wounded','weaptype1_txt':'weapon','provstate':'state'})
```

## Getting info about the data

```
In [34]: ndf.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 181691 entries, 0 to 181690
Columns: 23 entries, id to related
dtypes: float64(55), int64(22), object(58)
memory usage: 192.6+ MB
```

## considering only the required columns

```
In [50]: rdf=ndf.reindex(columns=['year','month','day','country_txt','region','killed','wounded','weapon','state','target','attack'])
rdf.head()

Out[50]:
```

	year	month	day	country_txt	region	killed	wounded	weapon	state	target	attack	
0	1970	7	2	Dominican Republic		2	1.0	0.0	Unknown	NaN	Private Citizens & Property	Assassination
1	1970	0	0	Mexico		1	0.0	0.0	Unknown	Federal	Government (Diplomatic)	Hostage Taking (Kidnapping)
2	1970	1	0	Philippines		5	1.0	0.0	Unknown	Tarlac	Journalists & Media	Assassination
3	1970	1	0	Greece		8	NaN	NaN	Explosives	Attica	Government (Diplomatic)	Bombing/Explosion
4	1970	1	0	Japan		4	NaN	NaN	Incendiary	Fukouka	Government (Diplomatic)	Facility/Infrastructure Attack

## Heatmap

```
In [55]: fig=plt.figure(figsize=(15,6))
sns.heatmap(rdf.corr(),annot=True)

Out[55]: <AxesSubplot:~>
```

## Top 10 countries affected

```
In [52]: c_count=ndf['country_txt'].value_counts()
c_count1=c_count[:10]
print(c_count1)

Iraq          24636
Pakistan      14388
Afghanistan   12731
India         11960
Colombia       8306
Philippines   6908
Peru           6096
El Salvador    5320
United Kingdom 5235
Turkey         4292
Name: country_txt, dtype: int64
```

## Terrorist ativities over the year

```
In [41]: fig=plt.figure(figsize=(25,6))
sns.countplot('year',data=rdf,order=rdf['year'].value_counts().index,palette='rainbow')

C:\Users\Irfaan\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(

Out[41]: <AxesSubplot: xlabel='year', ylabel='count'>
```

## printing the most affected states

```
In [42]: rdf['state'].value_counts().head()

Out[42]:
Baghdad      7645
Northern Ireland 4498
Unknown       4290
Balochistan   3710
Saladin       3421
Name: state, dtype: int64
```

## countplot to show the most affected states

```
In [18]: fig=plt.figure(figsize=(15,6))
sns.barplot(rdf['state'].value_counts()[:10].index,rdf['state'].value_counts()[:10].values,palette='rocket')
plt.xlabel('states')
plt.ylabel('counts')
plt.xticks(rotation=90)

C:\Users\Irfaan\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(

Out[18]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
[Text(0, 0, 'Baghdad'),
Text(1, 0, 'Northern Ireland'),
Text(2, 0, 'Unknown'),
Text(3, 0, 'Balochistan'),
Text(4, 0, 'Saladin'),
Text(5, 0, 'Al Anbar'),
Text(6, 0, 'Nineveh'),
Text(7, 0, 'Sindh'),
Text(8, 0, 'Khayr Pakhtunkhwa'),
Text(9, 0, 'Diyala')])
```

## type of the target

```
In [54]: fig=plt.figure(figsize=(15,6))
sns.countplot('target',data=rdf,order=rdf['target'].value_counts().index,palette='coolwarm')
plt.xticks(rotation=90)

C:\Users\Irfaan\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(

Out[54]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21]),
[Text(0, 0, 'Private Citizens & Property'),
Text(1, 0, 'Military'),
Text(2, 0, 'Police'),
Text(3, 0, 'Government (General)'),
Text(4, 0, 'Business'),
Text(5, 0, 'Transportation'),
Text(6, 0, 'Utilities'),
Text(7, 0, 'Unknown'),
Text(8, 0, 'Religious Figures/Institutions'),
Text(9, 0, 'Educational Institution'),
Text(10, 0, 'Government (Diplomatic)'),
Text(11, 0, 'Terrorists/Non-State Militia'),
Text(12, 0, 'Journalists & Media'),
Text(13, 0, 'Violent Political Party'),
Text(14, 0, 'Airports & Aircraft'),
Text(15, 0, 'Telecommunication'),
Text(16, 0, 'NGO'),
Text(17, 0, 'Tourists'),
Text(18, 0, 'Maritime'),
Text(19, 0, 'Food or Water Supply'),
Text(20, 0, 'Abortion Related'),
Text(21, 0, 'Other')])
```

## Most type of attack used

```
In [53]: fig=plt.figure(figsize=(15,6))
sns.countplot('attack',data=rdf,order=rdf['attack'].value_counts().index,palette='coolwarm')
plt.xticks(rotation=90)

C:\Users\Irfaan\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(

Out[53]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8]),
[Text(0, 0, 'Bombing/Explosion'),
Text(1, 0, 'Armed Assault'),
Text(2, 0, 'Assassination'),
Text(3, 0, 'Hostage Taking (Kidnapping)'),
Text(4, 0, 'Facility/Infrastructure Attack'),
Text(5, 0, 'Unknown'),
Text(6, 0, 'Unarmed Assault'),
Text(7, 0, 'Hostage Taking (Barricade Incident)'),
Text(8, 0, 'Hijacking')])
```