Monthly Intership Program for Professionals@ THE CODERS READY

Data Science and Business Analytics

THE CODERS READY

Name:Ashima Sapehia

Batch:Sept 2022

Task 1:Exploratory Data Analysis Global Terrorism

Problem Statement: 1.Perform "EDA" on dataset "Global Terrorism". 2.As a security/defense analyst try to find out the hot zones of terrorism. 3.What other security issues you can derive through the EDA? 4.You can use Python tools to perform this analysis.

```
!dir/w/s/ent/bytes/dirs
```

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
        warnings.filterwarnings('ignore')
```

```
In [2]: df=pd.read_csv('globalterrorism.csv',encoding='ISO-8859-1')
```

```
In [3]: df.head()
```

	eventid	iyear	imonth	iday	approach	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
0	197000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	197000002	1970	7	5	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
2	197000003	1970	7	6	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
3	197000004	1970	7	8	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
4	197000005	1970	7	9	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
181691	201512270001	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Suspects Offshore Balikatan Exercise p...	"Philippine Marines Kill Two Terrorists ...	"Suspected Militants Kill Two Soldiers, ...	START Primary Collection	0	0	0	0	NaN
181692	201512270002	2015	12	31	NaN	0	NaN	200	Spain	8	...	NaN	"Four Guardia Civil Officers Injured In ...	"Four Guardia Civil Officers Injured In ...	"Four Guardia Civil Officers Injured In ...	START Primary Collection	0	0	1	1	NaN
181693	201512270003	2015	12	31	NaN	0	NaN	200	Spain	8	...	NaN	"Two Women Injured In Spanish Train Bom...	"Two Women Injured In Spanish Train Bom...	"Two Women Injured In Spanish Train Bom...	START Primary Collection	0	0	1	1	NaN
181694	201512270004	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Suspicious Leaches Kill Two Soldiers In...	"Suspicious Leaches Kill Two Soldiers In...	"Suspicious Leaches Kill Two Soldiers In...	START Primary Collection	0	0	1	1	NaN
181695	201512270005	2015	12	31	NaN	0	NaN	92	India	6	...	NaN	"Tribal Insurgent Group Attacks Two Vill...	"Tribal Insurgent Group Attacks Two Vill...	"Tribal Insurgent Group Attacks Two Vill...	START Primary Collection	0	0	0	0	NaN
181696	201512270006	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security Agitated In Cotabato Killing 4(...	"Security Agitated In Cotabato City, Mana...	"Security Agitated In Cotabato City, Mana...	START Primary Collection	0	0	0	0	NaN

181691 rows × 135 columns

```
In [4]: df.columns
```

```
Out[4]: Index(['eventid', 'iyear', 'imonth', 'iday', 'approach',
               'extended',
               'resolution', 'country', 'country_txt', 'region',
               'region_txt',
               'provstate', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG',
               'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'],
              dtype='object', length=135)
```

```
In [5]: No_of_rows=df.columns.nunique()
        print("There are total %s columns in the data",No_of_rows)
```

```
Out[5]: There are total 135 columns in the data 135
```

```
In [6]: df[df.columns.isnull].type('int64').head()
        df[['iyear','imonth','iday','country','country_txt','region','region_txt','target_type',
            'target1','target_type1','target_subtype1','group','group_name','weaptype1_txt','weapon_type','nkill','nkill']]
```

```
Out[6]:
```

	eventid	iyear	imonth	iday	approach	extended	resolution	country	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
0	197000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	NaN
1	197000002	1970	7	5	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	NaN
2	197000003	1970	7	6	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	NaN
3	197000004	1970	7	8	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	NaN
4	197000005	1970	7	9	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	NaN
181691	201512270001	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Suspects Offshore Balikatan Exercise p...	"Philippine Marines Kill Two Terrorists ...	"Suspected Militants Kill Two Soldiers, ...	START Primary Collection	0	0	0	NaN
181692	201512270002	2015	12	31	NaN	0	NaN	200	Spain	8	...	NaN	"Four Guardia Civil Officers Injured In ...	"Four Guardia Civil Officers Injured In ...	"Four Guardia Civil Officers Injured In ...	START Primary Collection	0	0	1	NaN
181693	201512270003	2015	12	31	NaN	0	NaN	200	Spain	8	...	NaN	"Two Women Injured In Spanish Train Bom...	"Two Women Injured In Spanish Train Bom...	"Two Women Injured In Spanish Train Bom...	START Primary Collection	0	0	1	NaN
181694	201512270004	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Suspicious Leaches Kill Two Soldiers In...	"Suspicious Leaches Kill Two Soldiers In...	"Suspicious Leaches Kill Two Soldiers In...	START Primary Collection	0	0	1	NaN
181695	201512270005	2015	12	31	NaN	0	NaN	92	India	6	...	NaN	"Tribal Insurgent Group Attacks Two Vill...	"Tribal Insurgent Group Attacks Two Vill...	"Tribal Insurgent Group Attacks Two Vill...	START Primary Collection	0	0	0	NaN
181696	201512270006	2015	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security Agitated In Cotabato Killing 4(...	"Security Agitated In Cotabato City, Mana...	"Security Agitated In Cotabato City, Mana...	START Primary Collection	0	0	0	NaN

181691 rows × 135 columns

```
In [7]: df[['iyear','imonth','iday','country','region','state','city','attack_type','target_subtype','group_name','weapon_type']]
```

```
Out[7]:
```

```
In [8]: df['target_subtype'] = df['target_subtype'].fillna('unknown')
```

```
In [9]: df['State']=df['State'].fillna('unknown')
```

```
In [10]: df['City']=df['City'].fillna('unknown')
```

```
In [11]: df.replace(['Vehicle (not to include vehicle -car or truck bomb)', 'Vehicle'],inplace = True)
```

```
In [12]: df
```

```
Out[12]:
```

	Year	Month	Day	Country	Region	State	City	Attack_type	Target_type	Target_subtype	Group_name	Weapon_type
0	1970	7	2	Dominican Republic	Central America & Caribbean	NaN	Santo Domingo	Assassination	Private Citizens & Property	NaN	Named Civilian	MANUAL
1	1970	7	5	Mexico	North America	Federal	Mexico City	Hostage Taking (Kidnapping)	Government (Diplomatic)	Diplomatic Personnel (outside of embassy, cons...	23rd of September Communist League	Unknown
2	1970	7	6	Philippines	Southeast Asia	Tarlac	Unknown	Assassination	Journalists & Media	Radio Journalist/Staff/Facility	Unknown	Unknown
3	1970	7	8	Greece	Western Europe	Attica	Athens	Bombing/Explosion	Government (Diplomatic)	Embassy/Consulate	Unknown	Explosives
4	1970	7	9	Japan	East Asia	Fukouka	Fukouka	Facility/Infrastructure Attack	Government (Diplomatic)	Embassy/Consulate	Unknown	Incendiary
181691	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181692	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181693	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181694	2015	12	31	Philippines	Southeast Asia	Maguindanao	Kabuntalang	Facility/Infrastructure Attack	Private Citizens & Property	House/Apartment/Residence	Bangsamoro Islamic Freedom Movement (BIFM)	Incendiary
181695	2015	12	31	India	South Asia	Manipur	Leimkhong	Bombing/Explosion	Government (General)	Government Building/Facility/Office	Unknown	Explosives
181696	2015	12	31	Philippines	Southeast Asia	Maguindanao	Cotabato City	Bombing/Explosion	Unknown	Mail	Unknown	Explosives

181691 rows × 12 columns

Some changes that are required for the dataset.

F4 Nade selects or forget,collapse column as Others

```
In [13]: df['target_subtype'] = df['target_subtype'].fillna('others')
```

```
In [14]: df['State']=df['State'].fillna('unknown')
```

```
In [15]: df['City']=df['City'].fillna('unknown')
```

```
In [16]: df.replace(['Vehicle (not to include vehicle -car or truck bomb)', 'Vehicle'],inplace = True)
```

```
In [17]: df
```

```
Out[17]:
```

	Year	Month	Day	Country	Region	State	City	Attack_type	Target_type	Target_subtype	Group_name	Weapon_type
0	1970	7	2	Dominican Republic	Central America & Caribbean	NaN	Santo Domingo	Assassination	Private Citizens & Property	NaN	Named Civilian	MANUAL
1	1970	7	5	Mexico	North America	Federal	Mexico City	Hostage Taking (Kidnapping)	Government (Diplomatic)	Diplomatic Personnel (outside of embassy, cons...	23rd of September Communist League	Unknown
2	1970	7	6	Philippines	Southeast Asia	Tarlac	Unknown	Assassination	Journalists & Media	Radio Journalist/Staff/Facility	Unknown	Unknown
3	1970	7	8	Greece	Western Europe	Attica	Athens	Bombing/Explosion	Government (Diplomatic)	Embassy/Consulate	Unknown	Explosives
4	1970	7	9	Japan	East Asia	Fukouka	Fukouka	Facility/Infrastructure Attack	Government (Diplomatic)	Embassy/Consulate	Unknown	Incendiary
181691	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181692	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181693	2015	12	31	Spain	Western Europe	Balearic Islands	Palma de Mallorca	Assassination	Tourists	Tourist other	Basque Fatherland and Freedom (ETA)	Explosives
181694	2015	12	31	Philippines	Southeast Asia	Maguindanao	Kabuntalang	Facility/Infrastructure Attack	Private Citizens & Property	House/Apartment/Residence	Bangsamoro Islamic Freedom Movement (BIFM)	Incendiary
181695	2015	12	31	India	South Asia	Manipur	Leimkhong	Bombing/Explosion	Government (General)	Government Building/Facility/Office	Unknown	Explosives
181696	2015	12	31	Philippines	Southeast Asia	Maguindanao	Cotabato City	Bombing/Explosion	Unknown	Mail	Unknown	Explosives

181691 rows × 12 columns

```
In [18]: df.shape
```

```
Out[18]: (181691, 12)
```

checking null value

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

```
Out[19]: Year                 0
         Month                0
         Day                  0
         Country              0
         Region               0
         State                0
         City                 0
         Attack_type          0
         Target_type          0
         Target_subtype       0
         Group_name           0
         Weapon_type          0
         dtype: int64
```

unique column names.

```
In [20]: df.columns.unique()
```

```
Out[20]: Index(['Year', 'Month', 'Day', 'Country', 'Region', 'State', 'City',
                'Attack_type', 'Target_type', 'Target_subtype', 'Group_name',
                'Weapon_type'],
               dtype='object')
```

```
In [21]: df.columns.nunique()
```

```
Out[21]: 12
```

```
In [22]: Total_count=df['Country'].nunique()
         print("Total %s countries were noted in the data",Total_count)
```

```
Out[22]: Total 205 countries were noted in the data
```

```
In [23]: df['Region'].unique()
```

```
Out[23]: array(['Central America & Caribbean', 'North America', 'Southeast Asia',
                'Western Europe', East Asia', 'South America', 'Eastern Europe',
                'Sub-Saharan Africa', 'Middle East & North Africa', 'Australasia & Oceania',
                'South Asia', 'Central Asia'],
               dtype='object')
```

```
In [24]: df['Region'].value_counts()
```

```
Out[24]: Region
         Sub-Saharan Africa                 28075
         Southeast Asia                     18565
         South America                      18259
         Middle East & North Africa         18160
         South Asia                         17554
         Western Europe                     16639
         Eastern Europe                     11731
         Central America & Caribbean        10344
         East Asia                           7276
         North America                       3456
         Australasia & Oceania               2971
         Central Asia                        1563
         Name: count, dtype: int64
```

top 10 affected countries

```
In [25]: df['Country'].value_counts().head(10)
```

```
Out[25]: Country
         Iraq                  24636
         Pakistan              14368
         Afghanistan           12731
         India                 11960
         Colombia              8491
         Philippines           6908
         Peru                  6096
         El Salvador           5320
         United Kingdom        5235
         Turkey                4292
         Name: count, dtype: int64
```

top 10 affected regions

```
In [26]: df['Region'].value_counts().head(10)
```

```
Out[26]: Region
         Middle East & North Africa    50474
         South Asia                    40671
         South America                 28818
         Sub-Saharan Africa            28072
         Southeast Asia                18647
         Southeast Asia                15385
         Central America & Caribbean   14040
         Eastern Europe                 9341
         North America                  3456
         East Asia                      3351
         Name: Region, dtype: int64
```

top 10 affected regions.

```
In [27]: df['State'].value_counts().head(10)
```

```
Out[27]: State
         Baghdad                    7640
         unknown                    5603
         Northern Ireland           4240
         Mindanao                   3719
         Punjab                     3526
         Al anbar                   3256
         Sindh                      2848
         Sinaloa                    2394
         Khyber Pakhtunkhwa         2382
         Bogota                     2351
         Antioquia, Bajo Cauca      2300
         Name: count, dtype: int64
```

Affected states

```
In [28]: df['Target_subtype'].value_counts().head(5)
```

```
Out[28]: Target_subtype
         Private Citizens&Property                   12198
         Military Checkpoint                          8337
         others                                       6973
         Military Barracks/Base/Headquarter/Checkpost 6872
         Military Personnel (soldiers, troops, officers, forces)  7962
         Name: count, dtype: int64
```

chosen attacking modes

```
In [29]: df['Target_type'].value_counts().head(10)
```

```
Out[29]: Target_type
         Private Citizens & Property       45811
         Military                          25543
         Police                            24998
         Government (General)              16965
         Business                          15954
         Transportation                    10157
         Utilities                          9832
         Religious Figures/Institutions     7575
         Educational Institution            6922
         Terrorist/Non-State Militia        7422
         Name: count, dtype: int64
```

targets in attack

```
In [30]: df['Group_name'].value_counts().head(5)
```

```
Out[30]: Group_name
         Unknown                            82782
         Taliban                             7478
         Islamic State of Iraq and the Levant (ISIL)  5613
         Shining Path (SL)                   4761
         Farabundo Marti National Liberation Front (FMLN)  3351
         Name: count, dtype: int64
```

Gangsters and leading top 10 groups

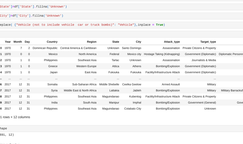
```
In [31]: df['Weapon_type'].value_counts().head(5)
```

```
Out[31]: Weapon_type
         Explosives    153390
         Firearms      115836
         Unknown        16847
         Incendiary     11265
         Melee           3662
         Name: Weapon_type, dtype: int64
```

Visualizations

Some but some using visualizations. For this, we have to use matplotlib and seaborn which we have installed at the beginning.

```
In [32]: df.Weapon_type.value_counts().plot.bar(figsize=(20,8))
         sns.countplot(x=df.Weapon_type);
```


Explosives
Firearms
Unknown
Incendiary
Melee
Sharp Object
Other
Chemical
Fire
Vehicle
Biological
Radiological
Other Explosive Device
Mechanical Device
Other Firearm
Other Incendiary
Other Melee
Other Sharp Object
Other Chemical
Other Fire
Other Vehicle
Other Biological
Other Radiological
Other Mechanical Device
Unknown Explosive Type
Unknown Firearm
Unknown Incendiary
Unknown Melee
Unknown Sharp Object
Unknown Chemical
Unknown Fire
Unknown Vehicle
Unknown Biological
Unknown Radiological
Unknown Mechanical Device
Unknown Other Explosive Type
Unknown Other Firearm
Unknown Other Incendiary
Unknown Other Melee
Unknown Other Sharp Object
Unknown Other Chemical
Unknown Other Fire
Unknown Other Vehicle
Unknown Other Biological
Unknown Other Radiological
Unknown Other