LetsGrowMore

Task:- Exploratory Data Analysis On Dataset - Terrorism

Intermediate Level Task

```
1
```

```
1
```

```
1
```

Imorting Modules

In [58]:
```
1 import numpy as np
2 import pandas as pd
```

Read the Data

In [11]:
```
1 df = pd.read_csv('C:\\Users\\Shivv\\Downloads\\globalterrorismdb.data.csv', encoding = 'ISO-8859-1')
```

In [8]:

```
1 df
```

Out[8]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
181686	201712310022	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab Militants Attack Army Che...
181687	201712310029	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Putin's 'victory' in Syria has turned into a ...
181688	201712310030	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...
181689	201712310031	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...
181690	201712310032	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...

181691 rows × 135 columns

Information and Describe

In [6]:
```
1 df.head()
```

Out[6]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scit
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	Na
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	Na
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	Na
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	Na
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	Na

5 rows × 135 columns

In [7]:
```
1 df.tail()
```

Out[7]:

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

5 rows × 135 columns

In [8]:
```
1 df.describe()
```

Out[8]:

	eventid	iyear	imonth	iday	extended	country	region	latitude	l
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	177135.000000	1.771
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.968501	7.160938	23.498343	-4.586
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.414535	2.933408	18.569242	2.047
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-53.154613	-8.618
25%	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.000000	5.000000	11.510046	4.545
50%	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.000000	6.000000	31.467463	4.324
75%	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.000000	10.000000	34.685087	6.871
max	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.000000	12.000000	74.633553	1.793

8 rows × 77 columns

In [16]:
```
1 df.isna().sum()
```

Out[16]:
```
eventid            0
iyear              0
imonth             0
iday               0
approxdate         0
                ...
INT_LOG            0
INT_IDEO           0
INT_MISC           0
INT_ANY            0
related       156653
Length: 135, dtype: int64
```

In [23]:
```
1 df.isna()
```

Out[23]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scit
0	False	False	False	False	False	False	False	False	False	False	...	False	True	True	Tr
1	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
2	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
3	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
4	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Fa
181687	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Fa
181688	False	False	False	False	False	False	True	False	False	False	...	True	False	True	Tr
181689	False	False	False	False	False	False	True	False	False	False	...	True	False	True	Tr
181690	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Tr

181691 rows × 135 columns

In [32]:
```
1 value = df['related']
2 value = value.fillna(value.mode(), inplace = True)
```

In [28]:
```
1 df.isna()
```

Out[28]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scit
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	Tr
1	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
2	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
3	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
4	False	False	False	False	False	False	True	False	False	False	...	True	True	True	Tr
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Fa
181687	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Fa
181688	False	False	False	False	False	False	True	False	False	False	...	True	False	True	Tr
181689	False	False	False	False	False	False	True	False	False	False	...	True	False	True	Tr
181690	False	False	False	False	False	False	True	False	False	False	...	True	False	False	Tr

181691 rows × 135 columns

In [21]:
```
1 df.columns
2
```

Out[21]:
```
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 [22]:
```
1 df.keys()
```

Out[22]:
```
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)
```

Importing Visualization Modules and Packages

In [64]:
```
1 import seaborn as sns
2 import matplotlib.pyplot as plt
3 import math
4 import warnings
5
```

In [63]:
```
1 warnings.filterwarnings('ignore')
```

Working with Dataset

In [12]:
```
1 df.shape
```

Out[12]:
```
(181691, 135)
```

Rename Columns

In [14]:

```
terr_df=df.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Country',
                           'region_txt':'Region','provstate':'State','city':'City',
                           'attacktype1_txt':'Attack_type','targtype1_txt':'Target_type',
                          'targsubtype1_txt':'Target_subtype','gname':'Gang_name',
                           'weaptype1_txt':'Weapon_type'})
```

In [15]:

```
1 terr_df
```

Out[15]:

	eventid	Year	Month	Day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
181686	201712310022	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab Militants Attack Army Che...
181687	201712310029	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Putin's 'victory' in Syria has turned into a ...
181688	201712310030	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...
181689	201712310031	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...
181690	201712310032	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...

181691 rows × 135 columns

Find Null value and Fill it!

In [18]:
```
1 terr_df.isna().sum()
```

Out[18]:
```
eventid            0
Year               0
Month              0
Day                0
approxdate    172452
               ...
INT_LOG            0
INT_IDEO           0
INT_MISC           0
INT_ANY            0
related       156653
Length: 135, dtype: int64
```

Fill a Null Values

In [24]:
```
1 terr_df['approxdate'] = terr_df['approxdate'].fillna('others')
```

In [26]:
```
1 terr_df['relate']=terr_df['related'].fillna('Unknown')
2
```

In [27]:
```
1 terr_df['addnotes']=terr_df['addnotes'].fillna('Unknown')
```

In [29]:
```
1 terr_df['related']=terr_df['related'].fillna('Unknown')
```

In [30]:
```
1 terr_df.isna().sum()
```

Out[30]:
```
eventid       0
Year          0
Month         0
Day           0
approxdate    0
             ..
INT_IDEO      0
INT_MISC      0
INT_ANY       0
related       0
relate        0
Length: 136, dtype: int64
```

In [32]:
```
1 terr_df.columns.unique()
```

Out[32]:
```
Index(['eventid', 'Year', 'Month', 'Day', 'approxdate', 'extended',
       'resolution', 'country', 'Country', 'region',
       ...
       'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG', 'INT_IDEO',
       'INT_MISC', 'INT_ANY', 'related', 'relate'],
      dtype='object', length=136)
```

In [33]:
```
1 terr_df.nunique()
```

Out[33]:
```
eventid       181691
Year              47
Month             13
Day               32
approxdate      2245
                ...
INT_IDEO           3
INT_MISC           3
INT_ANY            3
related        14307
relate         14307
Length: 136, dtype: int64
```

In [38]:
```
1 print(terr_df['Region'].unique())
```

```
['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']
```

In [40]:
```
1 terr_df['Region'].nunique()
```

Out[40]:
```
12
```

In [41]:
```
1 terr_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 136 entries, eventid to relate
dtypes: float64(55), int64(22), object(59)
memory usage: 188.5+ MB
```

In [37]:
```
1 terr_df['Country'].value_counts().head(10)
```

Out[37]:
```
Iraq              24636
Pakistan          14368
Afghanistan       12731
India             11960
Colombia           8306
Philippines        6908
Peru               6096
El Salvador        5320
United Kingdom     5235
Turkey             4292
Name: Country, dtype: int64
```

In [42]:
```
1 terr_df['Region'].value_counts().head(10)
```

Out[42]:
```
Middle East & North Africa     50474
South Asia                     44974
South America                  18978
Sub-Saharan Africa             17550
Western Europe                 16639
Southeast Asia                 12485
Central America & Caribbean    10344
Eastern Europe                  5144
North America                   3456
East Asia                        802
Name: Region, dtype: int64
```

In [44]:
```
1 terr_df['State'].value_counts().head(10)
2
```

Out[44]:
```
Baghdad               7645
Unknown               4711
Northern Ireland      4498
Balochistan           3710
Saladin               3411
Al Anbar              3299
Nineveh               3241
Sindh                 3206
Khyber Pakhtunkhwa    3084
Diyala                3041
Name: State, dtype: int64
```

In [46]:
```
1 terr_df['Target_subtype'].value_counts().head(10)
```

Out[46]:
```
Unnamed Civilian/Unspecified                                  11596
Police Security Forces/Officers                               11178
others                                                        10373
Military Unit/Patrol/Convoy                                    8277
Military Personnel (soldiers, troops, officers, forces)        7963
Government Personnel (excluding police, military)              6610
Village/City/Town/Suburb                                       6542
Politician or Political Party Movement/Meeting/Rally           6306
Police Building (headquarters, station, school)                5907
Military Barracks/Base/Headquarters/Checkpost                  5211
Name: Target_subtype, dtype: int64
```

In [48]:
```
1 terr_df['City'].value_counts().head(10)
```

Out[48]:
```
Unknown         10209
Baghdad          7589
Karachi          2652
Lima             2359
Mosul            2265
Belfast          2171
Santiago         1621
Mogadishu        1581
San Salvador     1558
Istanbul         1048
Name: City, dtype: int64
```

In [50]:
```
1 terr_df['Attack_type'].value_counts().head(10)
```

Out[50]:
```
Bombing/Explosion                      88255
Armed Assault                          42669
Assassination                          19312
Hostage Taking (Kidnapping)            11158
Facility/Infrastructure Attack         10356
Unknown                                 7276
Unarmed Assault                         1015
Hostage Taking (Barricade Incident)      991
Hijacking                                659
Name: Attack_type, dtype: int64
```

In [51]:
```
terr_df['Gang_name'].value_counts().head(10)

```

Out[51]:
```
Unknown                                                 82782
Taliban                                                  7478
Islamic State of Iraq and the Levant (ISIL)              5613
Shining Path (SL)                                        4555
Farabundo Marti National Liberation Front (FMLN)         3351
Al-Shabaab                                               3288
New People's Army (NPA)                                  2772
Irish Republican Army (IRA)                              2671
Revolutionary Armed Forces of Colombia (FARC)            2487
Boko Haram                                               2418
Name: Gang_name, dtype: int64
```

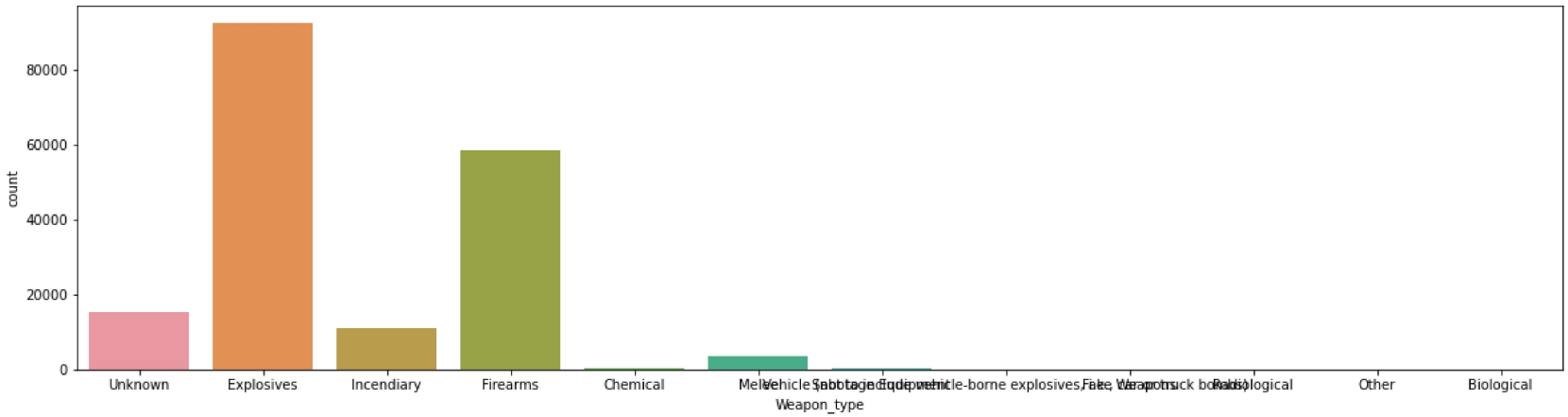
In [52]:
```
terr_df['Weapon_type'].value_counts().head(5)

```

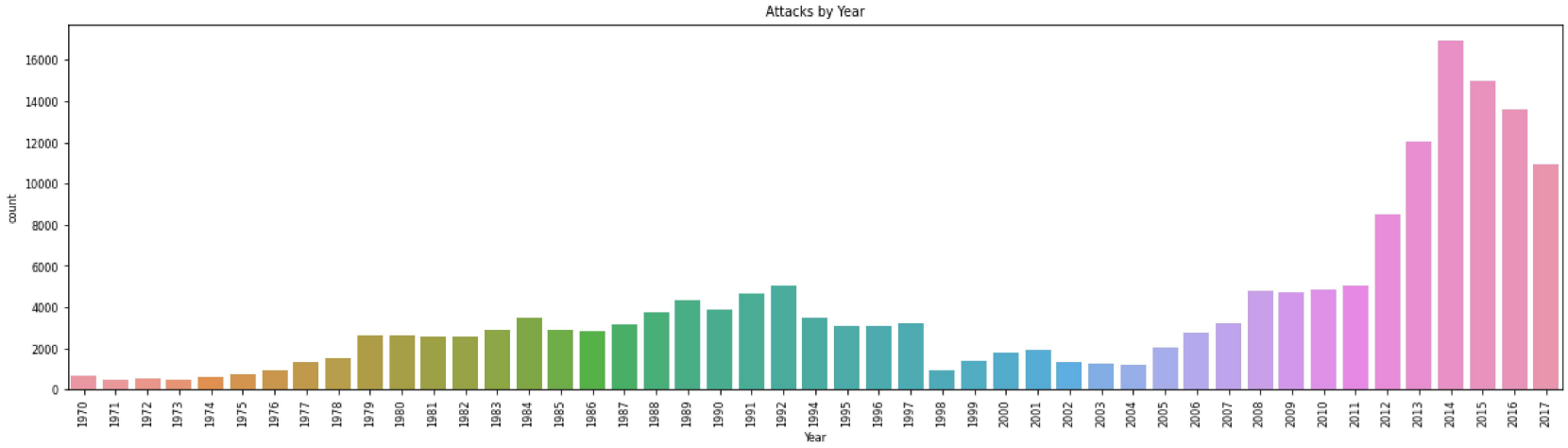
Out[52]:
```
Explosives    92426
Firearms      58524
Unknown       15157
Incendiary    11135
Melee          3655
Name: Weapon_type, dtype: int64
```

Visualization

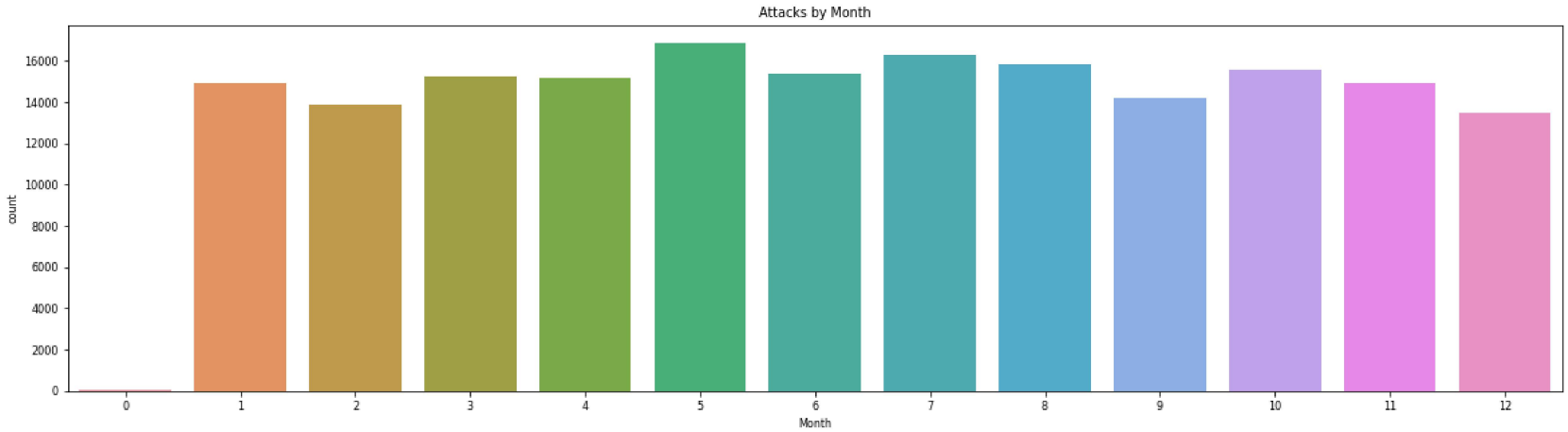
In [53]:
```
plt.figure(figsize=(20,5));
sns.countplot(x=terr_df.Weapon_type);
```



In [59]:
```
plt.figure(figsize=(20,5));
plt.xticks(rotation = 90)
plt.title('Attacks by Year')
sns.countplot(x=terr_df.Year);
```

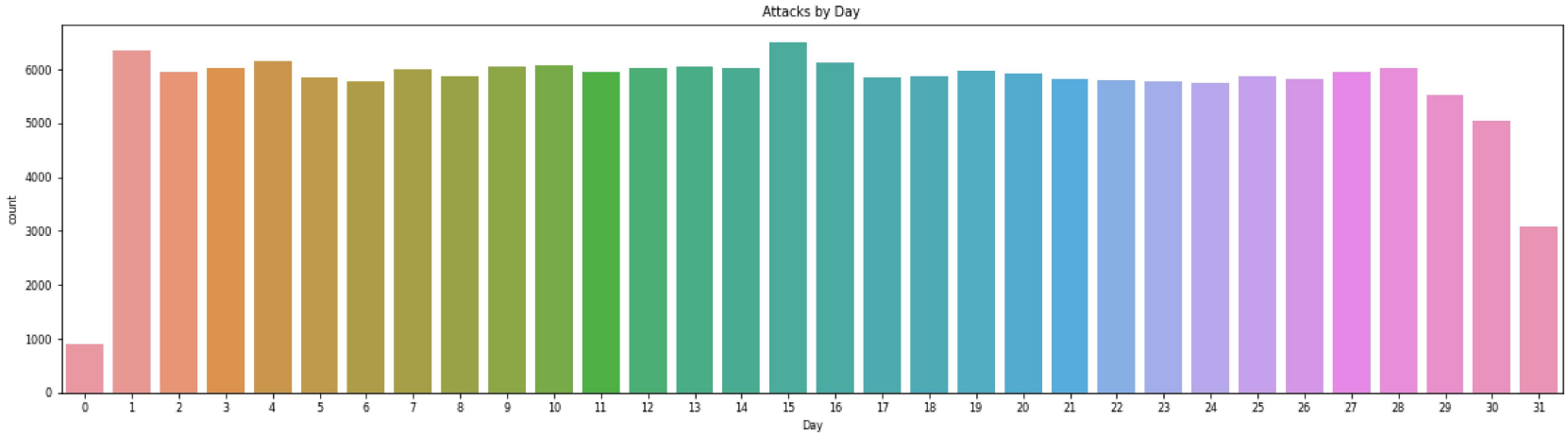


In [60]:
```
1 plt.figure(figsize=(20,5));
2 plt.title('Attacks by Month')
3 sns.countplot(x=terr_df.Month);
```



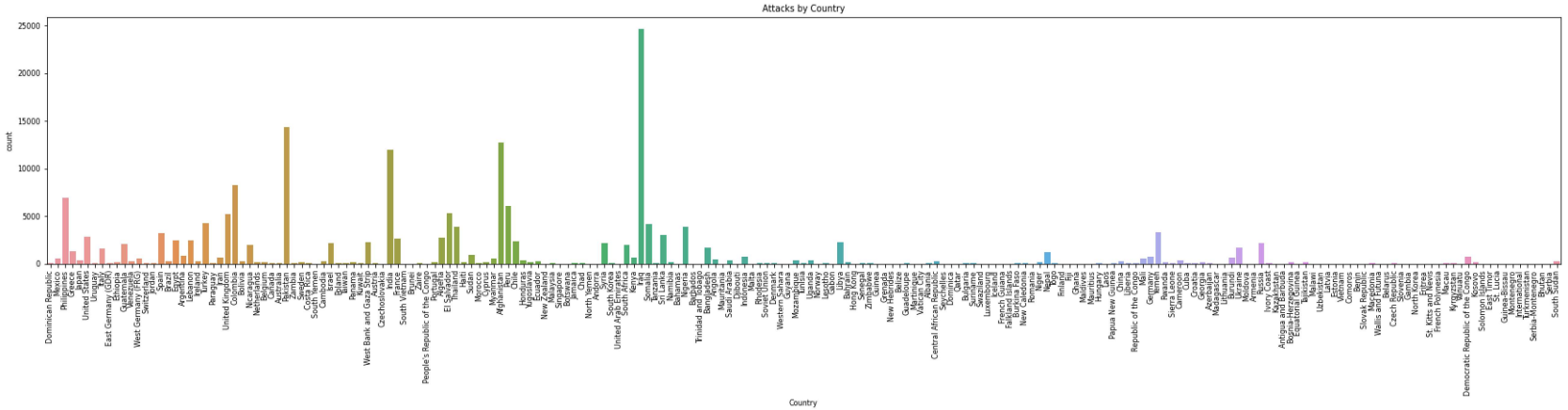
In [61]:
```
1 plt.figure(figsize=(20,5));
2 plt.title('Attacks by Day')
3 sns.countplot(x=terr_df.Day)
```

Out[61]: <AxesSubplot:title={'center':'Attacks by Day'}, xlabel='Day', ylabel='count'>



In [62]:

```
plt.figure(figsize=(30,5));
plt.rcParams.update({'font.size':8});
plt.xticks(rotation = 90)
plt.title('Attacks by Country')
sns.countplot(x=terr_df.Country);
```



In [ ]:

```
1
```

Thanku!!!

In [ ]:

```
1
```