

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

In [8]:
#import terrorist dataset
df = pd.read_csv('E:/python/globalterrorism.csv', encoding = 'ISO-8859-1')

In [11]:
df

Out[11]:
   eventid  year  month  day  approxdate  extended  resolution  country  country_txt  region  ...  address  scite1  scite2  scite3  dsource  INT_LOG  INT_IDEO  INT_MISC  INT_ARMY  related
0  1970000001  1970  7  2  NaN  0  NaN  58  Dominican Republic  2  ...  NaN  NaN  NaN  NaN  POIS  0  0  0  0  0  NaN
1  1970000002  1970  0  0  NaN  0  NaN  130  Mexico  1  ...  NaN  NaN  NaN  NaN  POIS  0  1  1  1  1  NaN
2  1970000003  1970  1  0  NaN  0  NaN  160  Philippines  5  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
3  1970000002  1970  1  0  NaN  0  NaN  78  Greece  8  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
4  1970000003  1970  1  0  NaN  0  NaN  101  Japan  4  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
181086  201712210022  2017  12  31  NaN  0  NaN  182  Somalia  11  ...  NaN  "Somalia: Al-Shabaab Military Attack Army Ch...  "Highlights: Somalia Daily Media Highlights 2  "Highlights: Somalia Daily Media Highlights 1  START Primary Collection  0  0  0  0  0  NaN
181087  201712210029  2017  12  31  NaN  0  NaN  200  Syria  10  ...  NaN  "Syria's victory in Syria has ignited IR...  "Two Russian soldiers killed at Hmeymim base L...  "Two Russian servicemen killed in Syria mort...  START Primary Collection  -9  -9  1  1  1  NaN
181088  201712210030  2017  12  31  NaN  0  NaN  160  Philippines  5  ...  NaN  "Maguindanao clashes trap the members' FBI...  NaN  NaN  START Primary Collection  0  0  0  0  0  NaN
181089  201712210031  2017  12  31  NaN  0  NaN  82  India  6  ...  NaN  "Trader escapes grenade attack in Imphal" Ba...  NaN  NaN  START Primary Collection  -9  -9  0  0  -9  NaN
181090  201712210032  2017  12  31  NaN  0  NaN  160  Philippines  5  ...  NaN  "Security tightened in Cotoabato following IED..."  "Security tightened in Cotoabato City."  Manila...  NaN  START Primary Collection  -9  -9  0  0  -9  NaN

181091 rows x 135 columns

In [12]:
print("There are {} rows and {} columns in the dataset".format(df.shape[0],df.shape[1]))

There are 181091 rows and 135 columns in the dataset

In [13]:
df.columns

Out[13]:
Index(['eventid', 'year', 'month', 'day', 'approxdate', 'extended', 'resolution', 'country', 'country_txt', 'region', 'address', 'scite1', 'scite2', 'scite3', 'dsource', 'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_ARMY', 'related'],
      dtype='object', length=135)

In [14]:
no_of_cols=df.columns.nunique()
print("There are total {} columns in the data".format(no_of_cols))

There are total 135 columns in the data

In [15]:
#let's rename some important columns
df=df.rename(columns={'year':'Year','month':'Month','day':'Day','country_txt':'Country','region_txt':'Region','approxdate':'State','city':'City','attacktype_txt':'Attack_Type','targettype_txt':'Target_Type','targetsubtype_txt':'Target_subtype','gname':'Gang_name','weapontype_txt':'Weapon_Type'})

In [16]:
#columns names were updated in this data frame
df

Out[16]:
   eventid  Year  Month  Day  approxdate  extended  resolution  country  Country  region  ...  address  scite1  scite2  scite3  dsource  INT_LOG  INT_IDEO  INT_MISC  INT_ARMY  related
0  1970000001  1970  7  2  NaN  0  NaN  58  Dominican Republic  2  ...  NaN  NaN  NaN  NaN  POIS  0  0  0  0  0  NaN
1  1970000002  1970  0  0  NaN  0  NaN  130  Mexico  1  ...  NaN  NaN  NaN  NaN  POIS  0  1  1  1  1  NaN
2  1970000003  1970  1  0  NaN  0  NaN  160  Philippines  5  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
3  1970000002  1970  1  0  NaN  0  NaN  78  Greece  8  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
4  1970000003  1970  1  0  NaN  0  NaN  101  Japan  4  ...  NaN  NaN  NaN  NaN  POIS  -9  -9  1  1  1  NaN
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
181086  201712210022  2017  12  31  NaN  0  NaN  182  Somalia  11  ...  NaN  "Somalia: Al-Shabaab Military Attack Army Ch...  "Highlights: Somalia Daily Media Highlights 2  "Highlights: Somalia Daily Media Highlights 1  START Primary Collection  0  0  0  0  0  NaN
181087  201712210029  2017  12  31  NaN  0  NaN  200  Syria  10  ...  NaN  "Syria's victory in Syria has turned into the members' FBI...  "Two Russian soldiers killed at Hmeymim base L...  "Two Russian servicemen killed in Syria mort...  START Primary Collection  -9  -9  1  1  1  NaN
181088  201712210030  2017  12  31  NaN  0  NaN  160  Philippines  5  ...  NaN  "Maguindanao clashes trap the members' FBI...  NaN  NaN  START Primary Collection  0  0  0  0  0  NaN
181089  201712210031  2017  12  31  NaN  0  NaN  82  India  6  ...  NaN  "Trader escapes grenade attack in Imphal" Ba...  NaN  NaN  START Primary Collection  -9  -9  0  0  -9  NaN
181090  201712210032  2017  12  31  NaN  0  NaN  160  Philippines  5  ...  NaN  "Security tightened in Cotoabato following IED..."  "Security tightened in Cotoabato City."  Manila...  NaN  START Primary Collection  -9  -9  0  0  -9  NaN

181091 rows x 135 columns

In [17]:
df=df[['Year','Month','Day','Country','Region','State','City','Attack_Type','Target_Type','Target_subtype','Gang_name','Weapon_Type']]

In [18]:
#our new dataframe
df

Out[18]:
   Year  Month  Day  Country  Region  State  City  Attack_Type  Target_Type  Target_subtype  Gang_name  Weapon_Type
0  1970  7  2  Dominican Republic  Central America & Caribbean  Unknown  Santo Domingo  Assassination  Private Citizens & Property  Named Civlian  MANO-D  Unknown
1  1970  0  0  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  1  0  Philippines  Southeast Asia  Tarlac  Unknown  Assessment  Journalists & Media  Radio Journalists/Staff/Facility  Unknown  Unknown
3  1970  1  0  Greece  Western Europe  Attika  Athens  Bombing/Explosion  Government (Diplomatic)  Embassy/Consulate  Unknown  Explosives
4  1970  1  0  Japan  East Asia  Fukushima  Fukushima  Facility/Infrastructure Attack  Government (Diplomatic)  Embassy/Consulate  Unknown  Incendiary
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
181086  2017  12  31  Somalia  Sub-Saharan Africa  Middle Shebelle  Ceelba Geelow  Armed Assault  Military  Military Checkpoint  Military Checkpoint  Al-Shabaab  Firearms
181087  2017  12  31  Syria  Middle East & North Africa  Latakia  Jdah  Bombing/Explosion  Military  Military Barracks/Basic/Headquarters/Checkpoint  Muslim extremists  Explosives
181088  2017  12  31  Philippines  Southeast Asia  Maguindanao  Kibiting  Facility/Infrastructure Attack  Private Citizens & Property  House/Apartments/Residence  Bangsamoro Islamic Freedom Movement (BIFM)  Incendiary
181089  2017  12  31  India  South Asia  Manipur  Imphal  Bombing/Explosion  Government (General)  Government Building/Facility/Office  Unknown  Explosives
181090  2017  12  31  Philippines  Southeast Asia  Maguindanao  Cotoabato City  Bombing/Explosion  Unknown  NaN  Unknown  Explosives

181091 rows x 12 columns

In [19]:
df['Target_subtype'] = df['Target_subtype'].fillna('others')

In [20]:
df['State'] = df['State'].fillna('Unknown')

In [21]:
df['City'] = df['City'].fillna('Unknown')

In [22]:
df.replace({"Vehicle":not to include vehicle-borne explosives, i.e., car or truck bombs}): "Vehicle",inplace = True)

In [23]:
df

Out[23]:
   Year  Month  Day  Country  Region  State  City  Attack_Type  Target_Type  Target_subtype  Gang_name  Weapon_Type
0  1970  7  2  Dominican Republic  Central America & Caribbean  Unknown  Santo Domingo  Assassination  Private Citizens & Property  Named Civlian  MANO-D  Unknown
1  1970  0  0  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  1  0  Philippines  Southeast Asia  Tarlac  Unknown  Assessment  Journalists & Media  Radio Journalists/Staff/Facility  Unknown  Unknown
3  1970  1  0  Greece  Western Europe  Attika  Athens  Bombing/Explosion  Government (Diplomatic)  Embassy/Consulate  Unknown  Explosives
4  1970  1  0  Japan  East Asia  Fukushima  Fukushima  Facility/Infrastructure Attack  Government (Diplomatic)  Embassy/Consulate  Unknown  Incendiary
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
181086  2017  12  31  Somalia  Sub-Saharan Africa  Middle Shebelle  Ceelba Geelow  Armed Assault  Military  Military Checkpoint  Military Checkpoint  Al-Shabaab  Firearms
181087  2017  12  31  Syria  Middle East & North Africa  Latakia  Jdah  Bombing/Explosion  Military  Military Barracks/Basic/Headquarters/Checkpoint  Muslim extremists  Explosives
181088  2017  12  31  Philippines  Southeast Asia  Maguindanao  Kibiting  Facility/Infrastructure Attack  Private Citizens & Property  House/Apartments/Residence  Bangsamoro Islamic Freedom Movement (BIFM)  Incendiary
181089  2017  12  31  India  South Asia  Manipur  Imphal  Bombing/Explosion  Government (General)  Government Building/Facility/Office  Unknown  Explosives
181090  2017  12  31  Philippines  Southeast Asia  Maguindanao  Cotoabato City  Bombing/Explosion  (General)  others  Unknown  Explosives

181091 rows x 12 columns

In [24]:
df.shape

Out[24]:
(181091, 12)

In [25]:
df.isnull().sum()

Out[25]:
Year      0
Month     0
Day        0
Country    0
Region     0
State      0
City       0
Attack_Type  0
Target_Type  0
Target_subtype  0
Gang_name  0
Weapon_Type  0
dtype: int64

In [26]:
df.head(10)

Out[26]:
   Year  Month  Day  Country  Region  State  City  Attack_Type  Target_Type  Target_subtype  Gang_name  Weapon_Type
0  1970  7  2  Dominican Republic  Central America & Caribbean  Unknown  Santo Domingo  Assassination  Private Citizens & Property  Named Civlian  23rd of September Communist League  Unknown
1  1970  0  0  Mexico  North America  Federal  Mexico city  Hostage Taking (Kidnapping)  Government (Diplomatic)  Diplomatic Personnel (outside of embassy cons...  Unknown
2  1970  1  0  Philippines  Southeast Asia  Tarlac  Unknown  Assessment  Journalists & Media  Radio Journalists/Staff/Facility  Unknown
3  1970  1  0  Greece  Western Europe  Attika  Athens  Bombing/Explosion  Government (Diplomatic)  Embassy/Consulate  Unknown
4  1970  1  0  Japan  East Asia  Fukushima  Fukushima  Facility/Infrastructure Attack  Government (Diplomatic)  Embassy/Consulate  Unknown
5  1970  1  1  United States  North America  Rhode  Cran  Armed Assault  Police  Police Building (headquarters, justice, school)  Black Nationalists  Firearms
6  1970  1  2  Uruguay  South America  Montevideo  Montevideo  Assassination  Police  Police Security Forces/Officers  Tupamaros (Uruguay)  Firearms
7  1970  1  2  United States  North America  California  Oakland  Bombing/Explosion  Utilities  Electricity  Unknown  Explosives
8  1970  1  2  United States  North America  Wisconsin  Madison  Facility/Infrastructure Attack  Military  Military Recruiting Station/Academy  New Year's Gang  Incendiary
9  1970  1  3  United States  North America  Wisconsin  Madison  Facility/Infrastructure Attack  Government (General)  Government Building/Facility/Office  New Year's Gang  Incendiary

In [27]:
df.columns.unique()

Out[27]:
Index(['Year', 'Month', 'Day', 'Country', 'Region', 'State', 'City', 'Attack_Type', 'Target_Type', 'Target_subtype', 'Gang_name', 'Weapon_Type'],
      dtype='object')

In [28]:
df.columns.nunique()

Out[28]:
12

In [29]:
Total_countries=df['Country'].nunique()
print("Total {} countries were noted in the data".format(Total_countries))

Total 285 countries were noted in the data

In [30]:
print(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', 'Australia & Oceania', 'South Asia', 'Central Asia']

In [31]:
print(df['Region'].nunique())

12

In [32]:
print(df['Attack_Type'].unique())

['Assassination', 'Hostage Taking (Kidnapping)', 'Bombing/Explosion', 'Facility/Infrastructure Attack', 'Armed Assault', 'Hijacking', 'Unknown', 'Unarmed Assault', 'Hostage Taking (Barricade Incident)']

In [33]:
print(df['Target_Type'].unique())

['Private Citizens & Property', 'Government (Diplomatic)', 'Journalists & Media', 'Police', 'Utilities', 'Military', 'Government (General)', 'Airports & Airports', 'Business', 'Educational Institution', 'Violent Political Party', 'Religious Figures/Institutions', 'Unknown', 'Transportation', 'Tourists', 'NGO', 'Telecommunication', 'Food or Water Supply', 'Terrorists/Non-State Militia', 'Other', 'Maritime', 'Abortion Related']

In [34]:
print(df['Gang_name'].unique())

['MANO-D', '23rd of September Communist League', 'Unknown', ...
'Farouk Warriors', 'Minuties of Metropolitan Attacks',
'Baluch Republican Party']

In [35]:
print(df['Gang_name'].nunique())

3537

In [36]:
print(df['Weapon_Type'].unique())

['Unknown', 'Explosives', 'Incendiary', 'Firearms', 'Chemical', 'Mele', 'Submachine Equipment', 'Vehicle', 'Fate Weapons', 'Radiological', 'Other', 'Biological']

In [40]:
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181091 entries, 0 to 181090
Data columns (total 12 columns):
#  Column  Non-Null Count  Dtype
---  ---
0  Year  181091 non-null  int64
1  Month  181091 non-null  int64
2  Day  181091 non-null  int64
3  Country  181091 non-null  object
4  Region  181091 non-null  object
5  State  181091 non-null  object
6  City  181091 non-null  object
7  Attack_Type  181091 non-null  object
8  Target_Type  181091 non-null  object
9  Target_subtype  181091 non-null  object
10  Gang_name  181091 non-null  object
11  Weapon_Type  181091 non-null  object
dtypes: int64(3), object(9)
memory usage: 16.6+ MB

In [41]:
df.describe()

Out[41]:
   Year  Month  Day
count  181091.000000  181091.000000  181091.000000
mean    2002.639607    6.487277    15.505444
std     13.264340     3.386303     8.814045
min      1970.000000     0.000000     0.000000
25%     1991.000000     4.000000     8.000000
50%     2009.000000     6.000000    15.000000
75%     2014.000000     9.000000    23.000000
max      2017.000000    12.000000    31.000000

In [42]:
df['Country'].value_counts().head(10)

Out[42]:
Iraq      2436
Pakistan  1458
Afghanistan  1273
India     1186
Colombia   838
Philippines  698
Peru       686
El Salvador  629
United Kingdom  523
Turkey     492
Name: Country, dtype: int64

In [43]:
df['State'].value_counts().head(10)

Out[43]:
Baghdad      7645
Unknown      4711
Northern Ireland  108
Balkhistan   3719
Baluchistan  3198
Al-Anbar     3239
Kinshasa     3241
Sindh        3076
Khyber Pakhtunkhwa  3044
Siyakha       3041
Name: State, dtype: int64

In [44]:
df['Target_subtype'].value_counts().head(10)

Out[44]:
Unarmed Civilian/Unspecified      11596
Police Security Forces/Officers    11178
others                             18373
Military Unit/Retal/Convoy         8277
Military Personnel (soldiers, troops, officers, forces)  8277
Government Personnel (excluding police, military)       8618
Village/City/Town/Settlement      8542
Politician or Political Party Movement/Meeting/Rally    6385
Police Building (headquarters, station, school)          5987
Military Barracks/Base/Headquarters/Checkpoint          5211
Name: Target_subtype, dtype: int64

In [45]:
df['City'].value_counts().head(10)

Out[45]:
Unknown      18289
Baghdad      7689
Karachi       2652
Lima          2359
Mosul         2205
Belfast       2171
Santiago      1621
Mogadishu     1581
San Salvador  1558
Tehran        1468
Name: City, dtype: int64

In [46]:
df['Attack_Type'].value_counts().head(10)

Out[46]:
Bombing/Explosion      88255
Armed Assault          42669
Assassination          19322
Hostage Taking (Kidnapping)  12158
Facility/Infrastructure Attack  10396
Unknown                9791
Unarmed Assault        2818
Hostage Taking (Barricade Incident)  2278
Hijacking              659
Name: Attack_Type, dtype: int64

In [47]:
df['Target_Type'].value_counts().head(10)

Out[47]:
Private Citizens & Property  45511
Military                    27984
Police                     24566
Government (General)        22283
Business                    20669
Transportation               6789
Utilities                   6642
Unknown                     5988
Religious Figures/Institutions  4448
Educational Institution      4362
Name: Target_Type, dtype: int64

In [48]:
df['Gang_name'].value_counts().head(10)

Out[48]:
Unknown      82782
Taliban      7479
Islamic State of Iraq and the Levant (ISIL)  5613
Shining Path (SL)  4555
Farabundo Marti National Liberation Front (FMLN)  3551
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 [49]:
df['Weapon_Type'].value_counts().head(5)

Out[49]:
Explosives  92426
Firearms    58524
Unknown     35157
Incendiary  31336
Mele        3655
Name: Weapon_Type, dtype: int64

visualizations

In [50]:
plt.figure(figsize=(20,5));
sns.countplot(x=df.Weapon_Type);

In [51]:
plt.figure(figsize=(20,5));
sns.countplot(x=df.Weapon_Type);

In [52]:
plt.figure(figsize=(20,5));
plt.xticks(rotation = 90);
plt.title('Attacks by Year');
sns.countplot(x=df.Year);

In [53]:
plt.figure(figsize=(20,5));
<Figure size 1440x368 with 0 Axes>

In [54]:
plt.title('Attacks by Month');
sns.countplot(x=df.Month);

In [55]:
plt.figure(figsize=(20,5));
<Figure size 1440x368 with 0 Axes>

In [56]:
plt.title('Attacks by Day');
sns.countplot(x=df.Day);

In [57]:
plt.figure(figsize=(30,5));
plt.xticks(rotation = 90);
plt.title('Attacks by Country');
sns.countplot(x=df.Country);

In [58]:
plt.figure(figsize=(20,5));
plt.xticks(rotation = 90);
plt.title('Attacks by Region');
sns.countplot(x=df.Region);

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

In [60]:
plt.figure(figsize=(20,5));
plt.xticks(rotation = 90);
plt.title('Attacks by Attack_Type');
sns.countplot(x=df.Attack_Type);

In [61]:
plt.figure(figsize=(20,5));
plt.xticks(rotation = 90);
plt.title('Attacks by Attack_Type');
sns.countplot(x=df.Attack_Type);
```