

Mayur Shinde

Data Science and Business Analytics Intern @ The Sparks Foundation

Topic : Exploratory Data Analysis - Terrorism

Dataset : `globalterrorismdb_0718dist.csv` [<https://bit.ly/2TK>]

```
In [27]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

```
In [63]: data = pd.read_csv("D:\Codes\Grip\Global Terrorism - START data\globalterrorismdb_0718d
data.head()
```

Out[63]:

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

5 rows × 135 columns

```
In [29]: data.columns.values
```

```
Out[29]: array(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended',
       'resolution', 'country', 'country_txt', 'region', 'region_txt',
       'provstate', 'city', 'latitude', 'longitude', 'specificity',
       'vicinity', 'location', 'summary', 'crit1', 'crit2', 'crit3',
       'doubtterr', 'alternative', 'alternative_txt', 'multiple',
       'success', 'suicide', 'attacktype1', 'attacktype1_txt',
       'attacktype2', 'attacktype2_txt', 'attacktype3', 'attacktype3_txt',
       'targtype1', 'targtype1_txt', 'targsubtype1', 'targsubtype1_txt',
       'corp1', 'target1', 'natlty1', 'natlty1_txt', 'targtype2',
       'targtype2_txt', 'targsubtype2', 'targsubtype2_txt', 'corp2',
       'target2', 'natlty2', 'natlty2_txt', 'targtype3', 'targtype3_txt',
       'targsubtype3', 'targsubtype3_txt', 'corp3', 'target3', 'natlty3',
```

```
'natlty3_txt', 'gname', 'gsubname', 'gname2', 'gsubname2',
'gname3', 'gsubname3', 'motive', 'guncertain1', 'guncertain2',
'guncertain3', 'individual', 'nperps', 'nperpcap', 'claimed',
'claimmode', 'claimmode_txt', 'claim2', 'claimmode2',
'claimmode2_txt', 'claim3', 'claimmode3', 'claimmode3_txt',
'compclaim', 'weaptype1', 'weaptype1_txt', 'weapsubtype1',
'weapsubtype1_txt', 'weaptype2', 'weaptype2_txt', 'weapsubtype2',
'weapsubtype2_txt', 'weaptype3', 'weaptype3_txt', 'weapsubtype3',
'weapsubtype3_txt', 'weaptype4', 'weaptype4_txt', 'weapsubtype4',
'weapsubtype4_txt', 'weapdetail', 'nkill', 'nkillus', 'nkillter',
'nwound', 'nwoundus', 'nwoundte', 'property', 'propextent',
'propextent_txt', 'propvalue', 'propcomment', 'ishostkid',
'nhostkid', 'nhostkidus', 'nhours', 'ndays', 'divert',
'kidhijcountry', 'ransom', 'ransomamt', 'ransomamtus',
'ransompaid', 'ransompaidus', 'ransomnote', 'hostkidoutcome',
'hostkidoutcome_txt', 'nreleased', 'addnotes', 'scite1', 'scite2',
'scite3', 'dbsource', 'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_ANY',
'related'], dtype=object)
```

In [71]: `data.rename(columns={'iyear':'Year','imonth':'Month','iday':'day','gname':'Group','coun
'longitude':'longitude','summary':'summary','attacktype1_txt':'Attacktype','targtyp
'nwound':'Wound'},inplace=True)`

In [72]: `data = data[['Year', 'Month', 'day', 'Country', 'State', 'Region', 'City', 'latitude', 'longitude',
'Wound', 'target1', 'summary', 'Group', 'Targettype', 'Weapon', 'motive']]`

In [73]: `data.head()`

Out[73]:

	Year	Month	day	Country	State	Region	City	latitude	longitude	Attackt
0	1970	7	2	Dominican Republic	NaN	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassina
1	1970	0	0	Mexico	Federal	North America	Mexico city	19.371887	-99.086624	Hostage Ta (Kidnapp
2	1970	1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741	Assassina
3	1970	1	0	Greece	Attica	Western Europe	Athens	37.997490	23.762728	Bombing/Explo:
4	1970	1	0	Japan	Fukouka	East Asia	Fukouka	33.580412	130.396361	Facility/Infrastruc Atti

In [74]: `data.shape`

Out[74]: (181691, 18)

In [75]: `data.isnull().sum()`

Out[75]:

Year	0
Month	0
day	0
Country	0
State	421

```
Region          0
City           434
latitude       4556
longitude      4557
Attacktype     0
kill           10313
Wound          16311
target1        636
summary         66129
Group          0
Targettype      0
Weapon          0
motive        131130
dtype: int64
```

```
In [76]: data['Wound'] = data['Wound'].fillna(0)
data['kill'] = data['kill'].fillna(0)
```

```
In [77]: data['Casualties'] = data['kill'] + data['Wound']
```

```
In [78]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 19 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          --          --    
 0   Year        181691 non-null   int64  
 1   Month       181691 non-null   int64  
 2   day         181691 non-null   int64  
 3   Country     181691 non-null   object 
 4   State        181270 non-null   object 
 5   Region       181691 non-null   object 
 6   City         181257 non-null   object 
 7   latitude     177135 non-null   float64
 8   longitude    177134 non-null   float64
 9   Attacktype   181691 non-null   object 
 10  kill         181691 non-null   float64
 11  Wound        181691 non-null   float64
 12  target1      181055 non-null   object 
 13  summary       115562 non-null   object 
 14  Group         181691 non-null   object 
 15  Targettype    181691 non-null   object 
 16  Weapon        181691 non-null   object 
 17  motive        50561 non-null   object 
 18  Casualties    181691 non-null   float64
dtypes: float64(5), int64(3), object(11)
memory usage: 26.3+ MB
```

```
In [79]: data.describe()
```

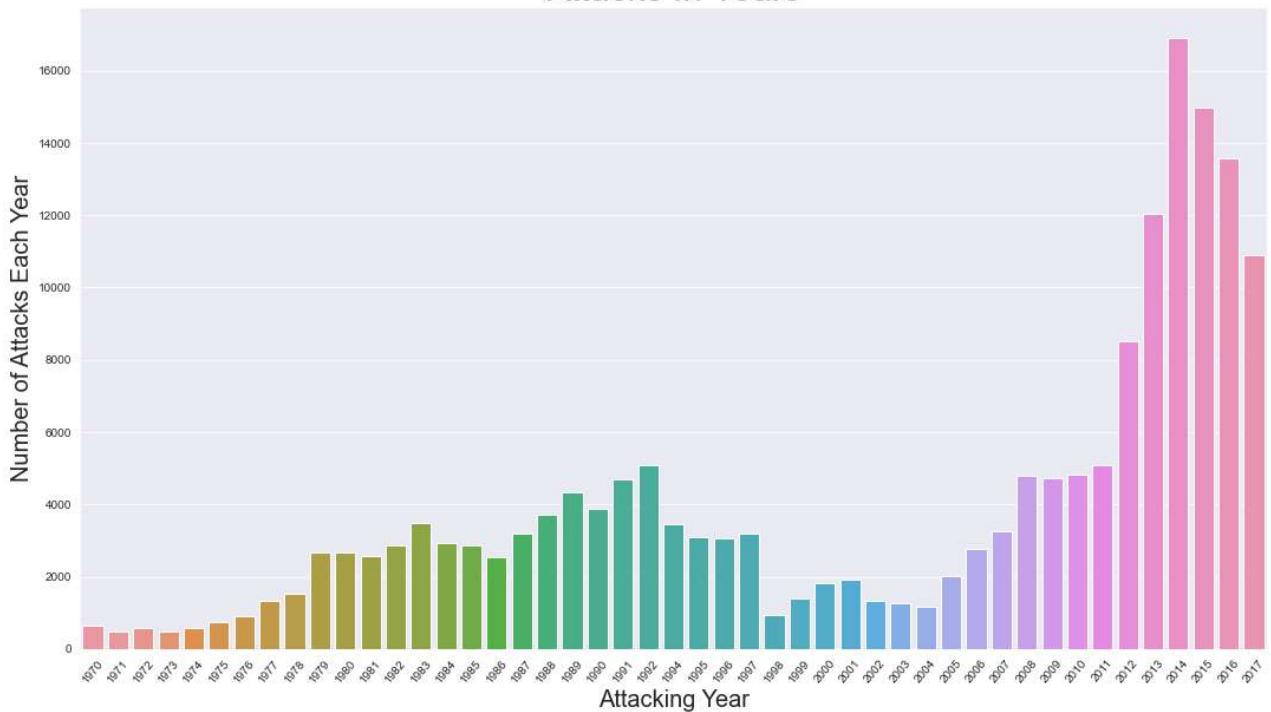
	Year	Month	day	latitude	longitude	kill
count	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181691.000000
mean	2002.638997	6.467277	15.505644	23.498343	-4.586957e+02	2.266860
std	13.259430	3.388303	8.814045	18.569242	2.047790e+05	11.227057
min	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	0.000000
25%	1991.000000	4.000000	8.000000	11.510046	4.545640e+00	0.000000
50%	2009.000000	6.000000	15.000000	31.467463	4.324651e+01	0.000000

	Year	Month	day	latitude	longitude	kill
75%	2014.000000	9.000000	23.000000	34.685087	6.871033e+01	2.000000
max	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	1570.000000

In [80]:

```
year = data['Year'].unique()
years_count = data['Year'].value_counts(dropna = False).sort_index()
plt.figure(figsize = (18,10))
sns.barplot(x = year,
            y = years_count)
sns.color_palette("mako", as_cmap=True)
plt.xticks(rotation = 50)
plt.xlabel('Attacking Year', fontsize=20)
plt.ylabel('Number of Attacks Each Year', fontsize=20)
plt.title('Attacks In Years', fontsize=30)
plt.show()
```

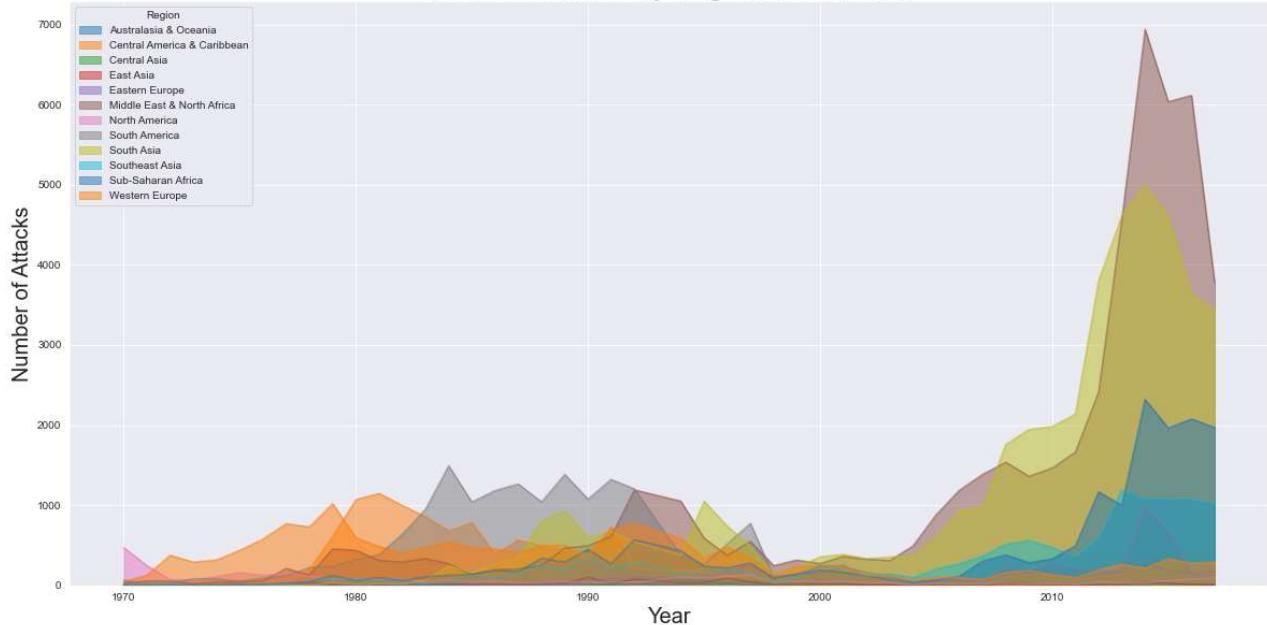
Attacks In Years



In [81]:

```
pd.crosstab(data.Year, data.Region).plot(kind='area', stacked=False, figsize=(20,10))
plt.title('Terrorist Activities By Region In Each Year', fontsize=25)
plt.ylabel('Number of Attacks', fontsize=20)
plt.xlabel("Year", fontsize=20)
plt.show()
```

Terrorist Activities By Region In Each Year



```
In [82]: attack = data.Country.value_counts()[:10]
attack
```

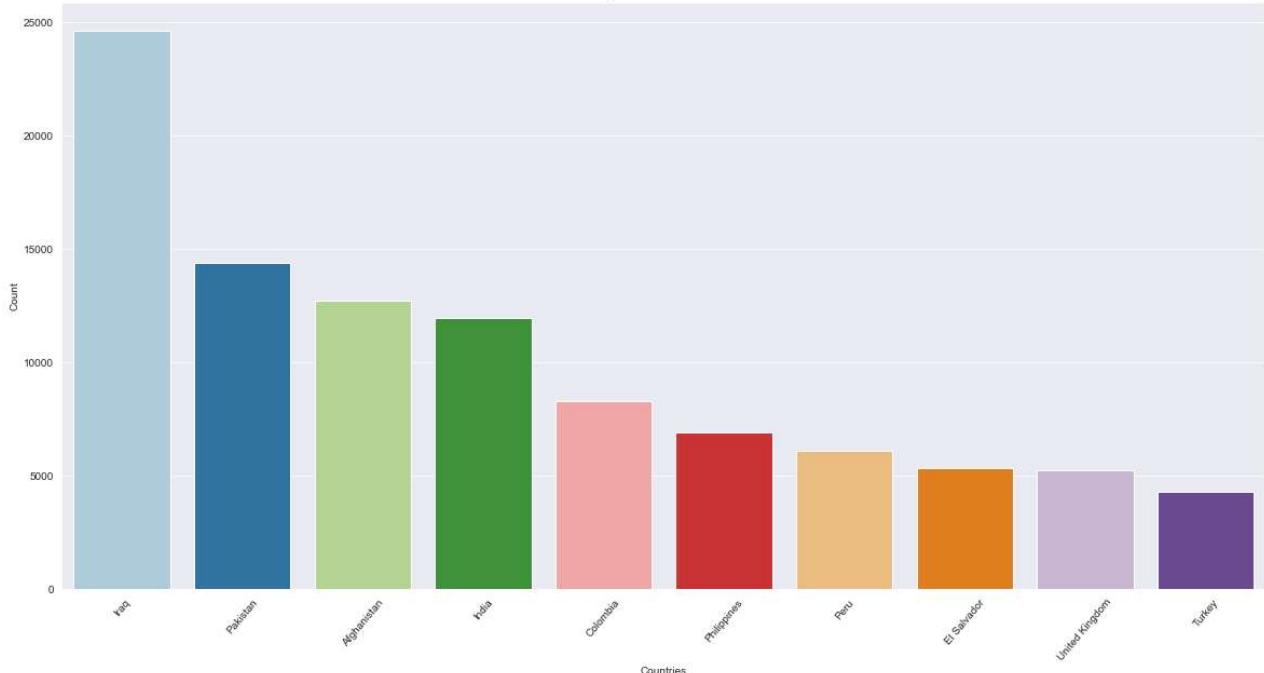
```
Out[82]: Iraq                24636
Pakistan           14368
Afghanistan        12731
India               11960
Colombia            8306
Philippines         6908
Peru                 6096
El Salvador          5320
United Kingdom       5235
Turkey               4292
Name: Country, dtype: int64
```

```
In [83]: data.Group.value_counts()[1:10]
```

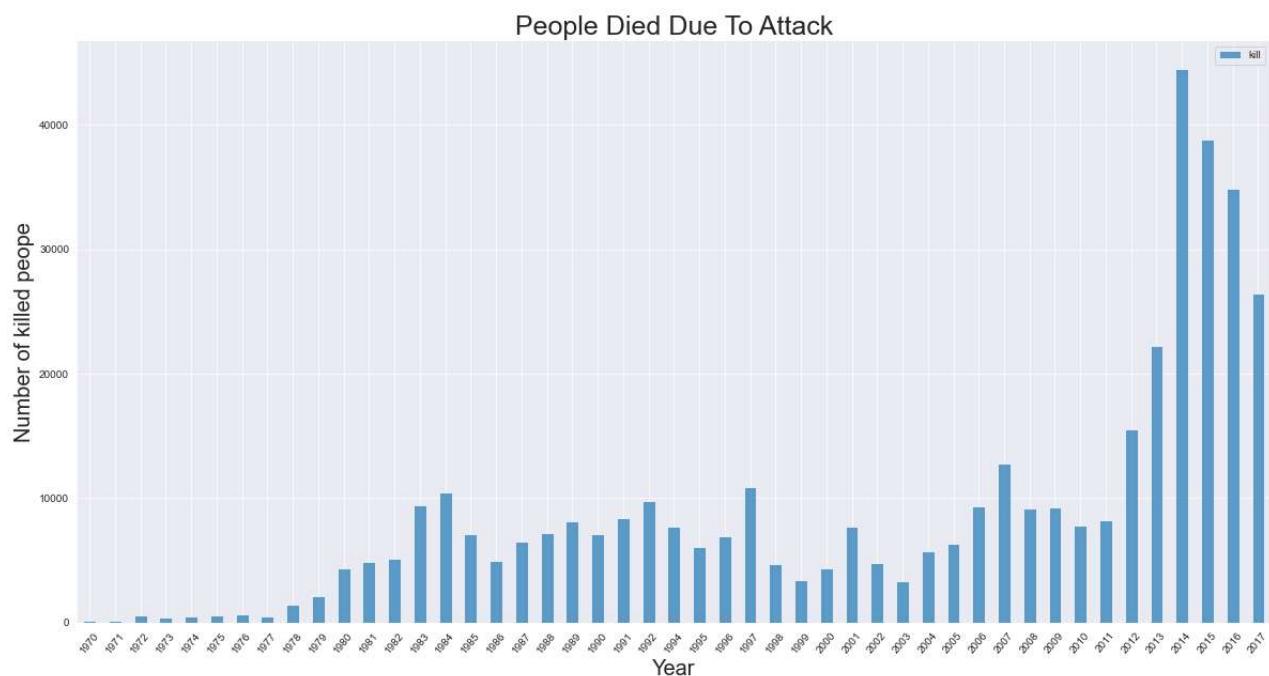
```
Out[83]: Taliban                  7478
Islamic State of Iraq and the Levant (ISIL) 5613
Shining Path (SL)                   4555
Farabundo Marti National Liberation Front (FMLN) 3351
Al-Shabaab                      3288
New People's Army (NPA)            2772
Irish Republican Army (IRA)       2671
Revolutionary Armed Forces of Colombia (FARC) 2487
Boko Haram                        2418
Name: Group, dtype: int64
```

```
In [91]: plt.subplots(figsize=(20,10))
sns.barplot(data['Country'].value_counts()[:10].index,data['Country'].value_counts()[:10].values)
plt.title('Top Countries Affected')
plt.xlabel('Countries')
plt.ylabel('Count')
plt.xticks(rotation = 50)
plt.show()
```

Top Countries Affected

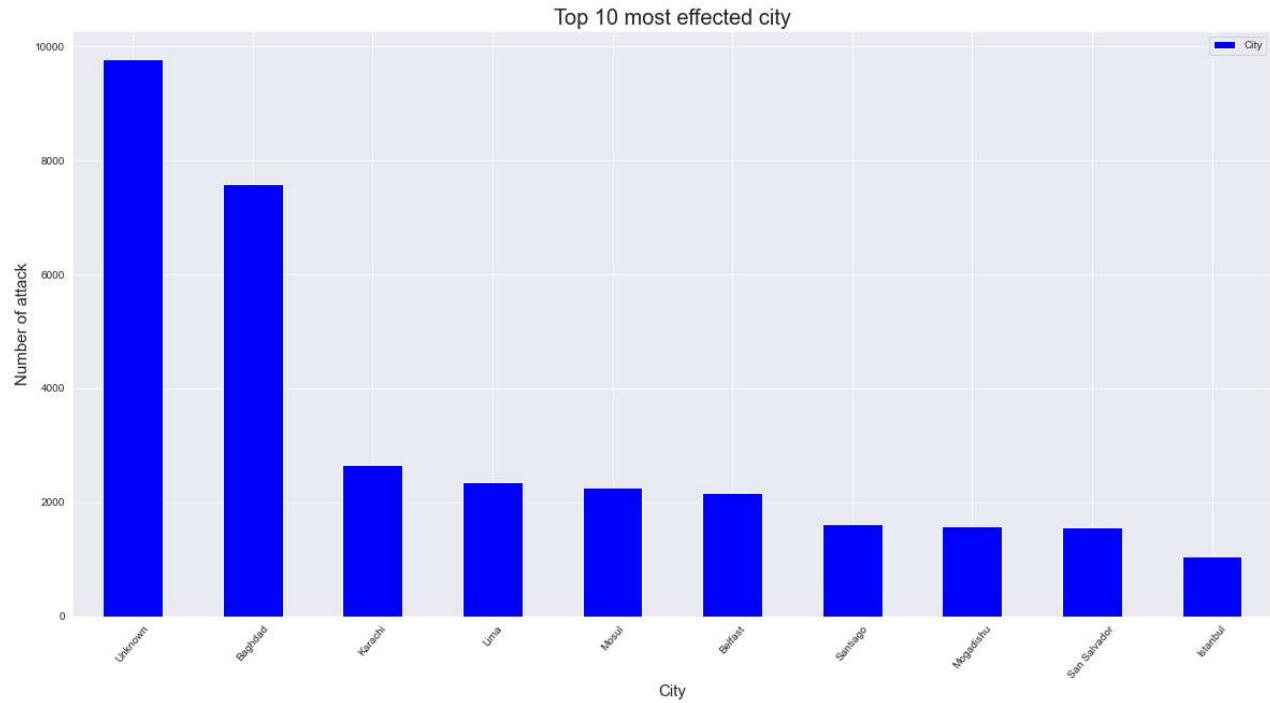


```
In [92]: df = data[['Year', 'kill']].groupby(['Year']).sum()
fig, ax4 = plt.subplots(figsize=(20,10))
df.plot(kind='bar', alpha=0.7, ax=ax4)
plt.xticks(rotation = 50)
plt.title("People Died Due To Attack", fontsize=25)
plt.ylabel("Number of killed people", fontsize=20)
plt.xlabel('Year', fontsize=20)
top_side = ax4.spines["top"]
top_side.set_visible(False)
right_side = ax4.spines["right"]
right_side.set_visible(False)
```

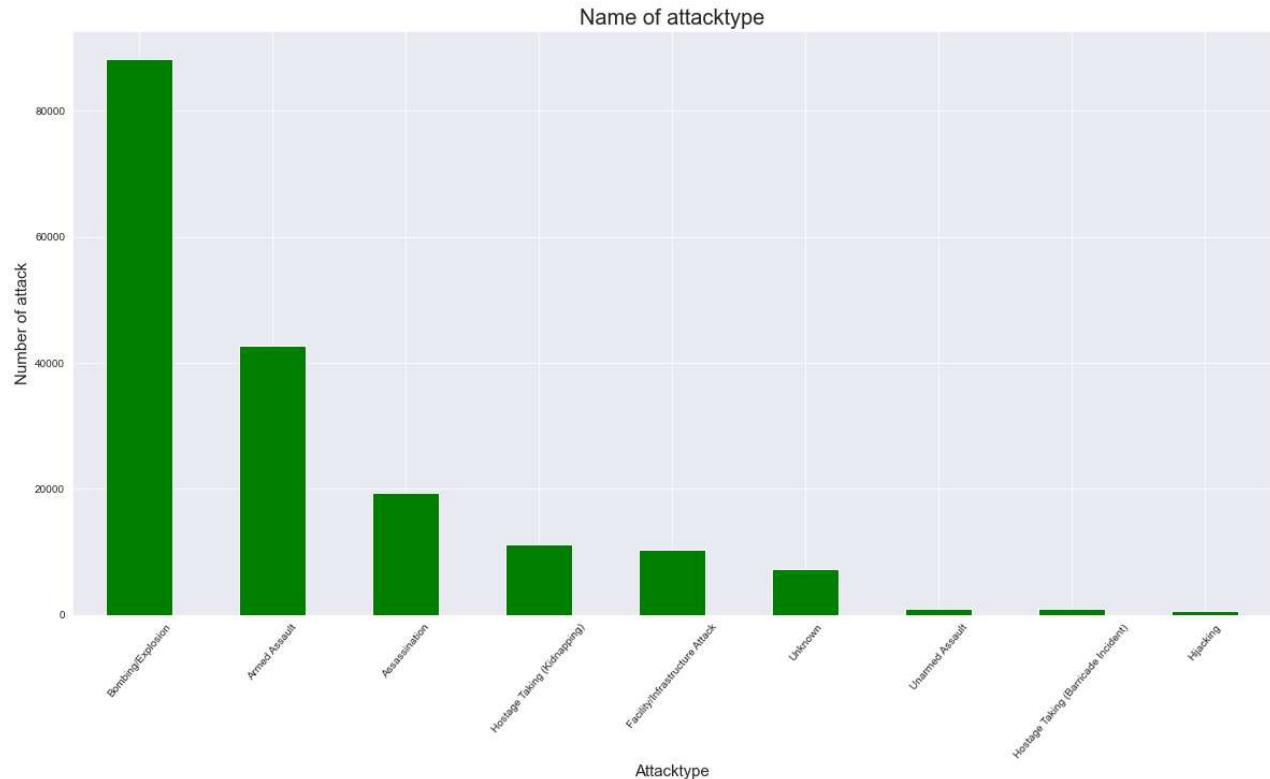


```
In [95]: data['City'].value_counts().to_frame().sort_values('City', axis=0, ascending=False).head()
plt.xticks(rotation = 50)
plt.xlabel("City", fontsize=15)
```

```
plt.ylabel("Number of attack", fontsize=15)
plt.title("Top 10 most effected city", fontsize=20)
plt.show()
```

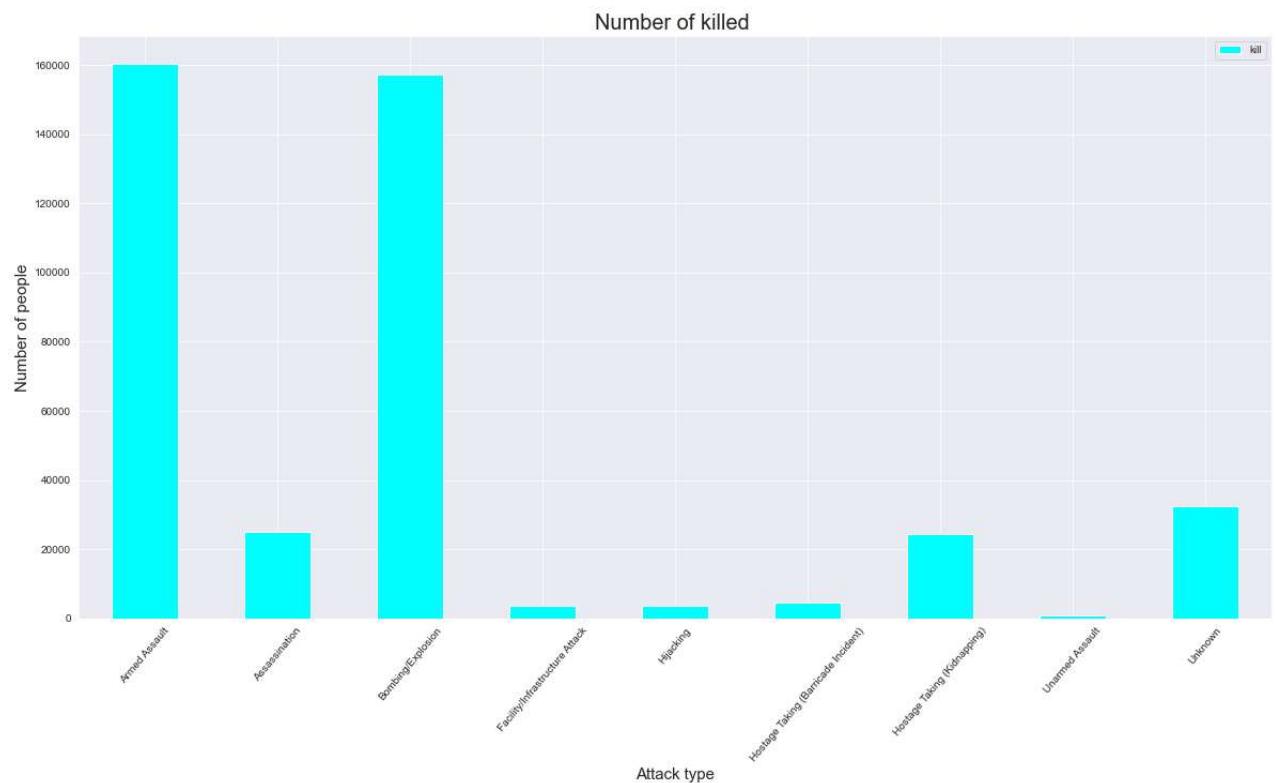


```
In [96]: data['Attacktype'].value_counts().plot(kind='bar', figsize=(20,10), color='Green')
plt.xticks(rotation = 50)
plt.xlabel("Attacktype", fontsize=15)
plt.ylabel("Number of attack", fontsize=15)
plt.title("Name of attacktype", fontsize=20)
plt.show()
```

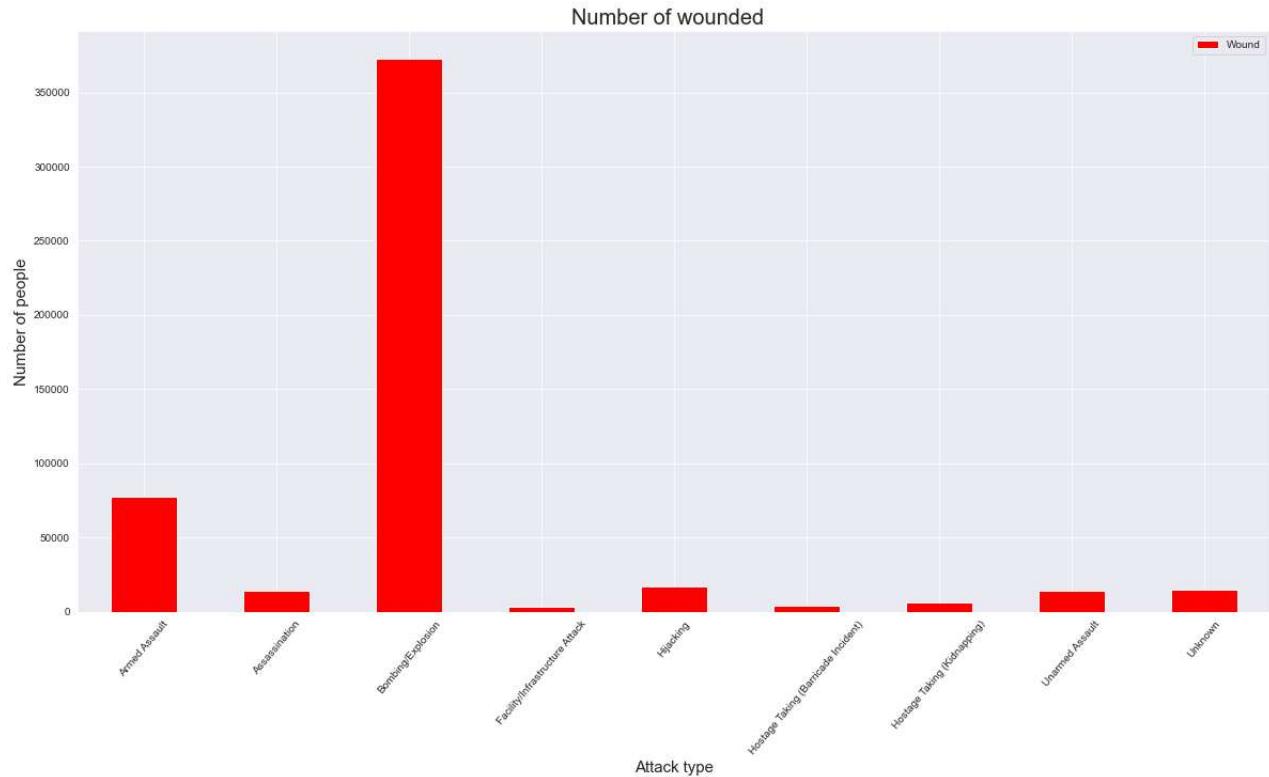


```
In [97]: data[['Attacktype', 'kill']].groupby(["Attacktype"], axis=0).sum().plot(kind='bar', figsize
```

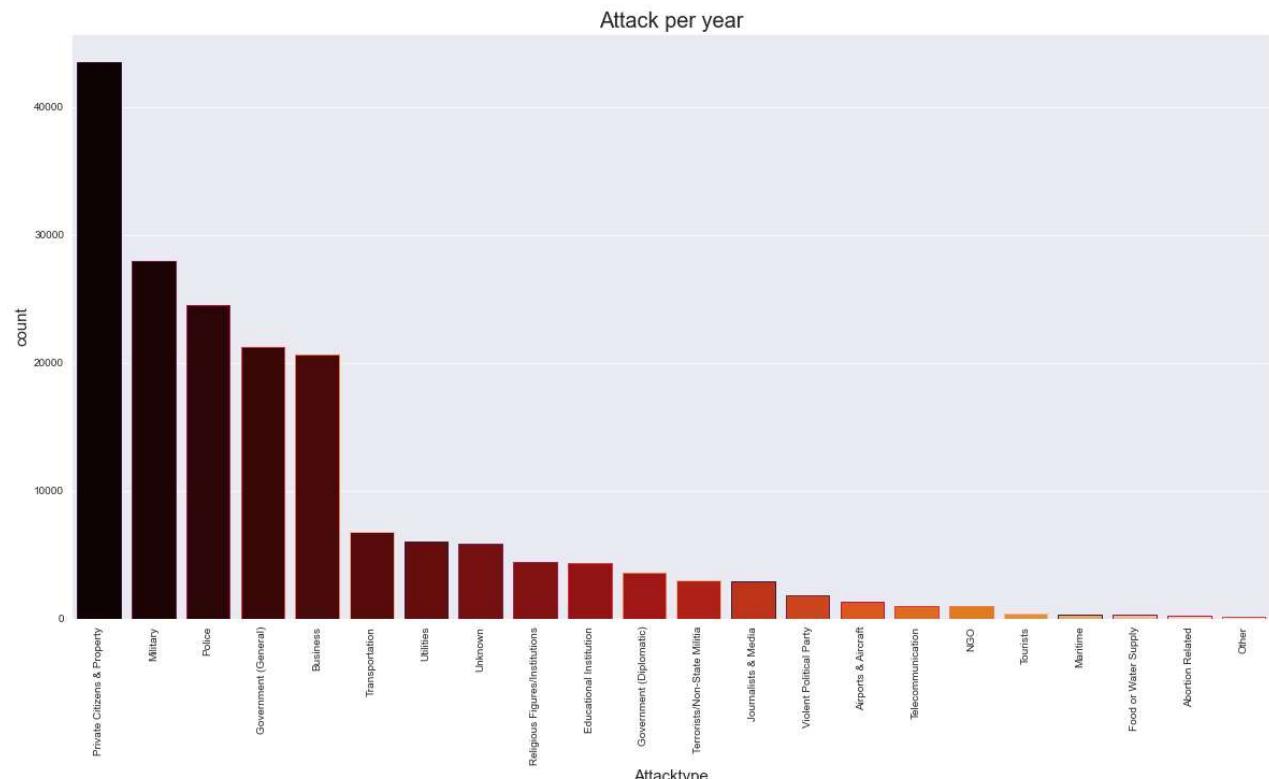
```
plt.xticks(rotation=50)
plt.title("Number of killed ", fontsize=20)
plt.ylabel('Number of people', fontsize=15)
plt.xlabel('Attack type', fontsize=15)
plt.show()
```



```
In [98]: data[['Attacktype', 'Wound']].groupby(['Attacktype'], axis=0).sum().plot(kind='bar', figsize=(10, 6))
plt.xticks(rotation=50)
plt.title("Number of wounded ", fontsize=20)
plt.ylabel('Number of people', fontsize=15)
plt.xlabel('Attack type', fontsize=15)
plt.show()
```

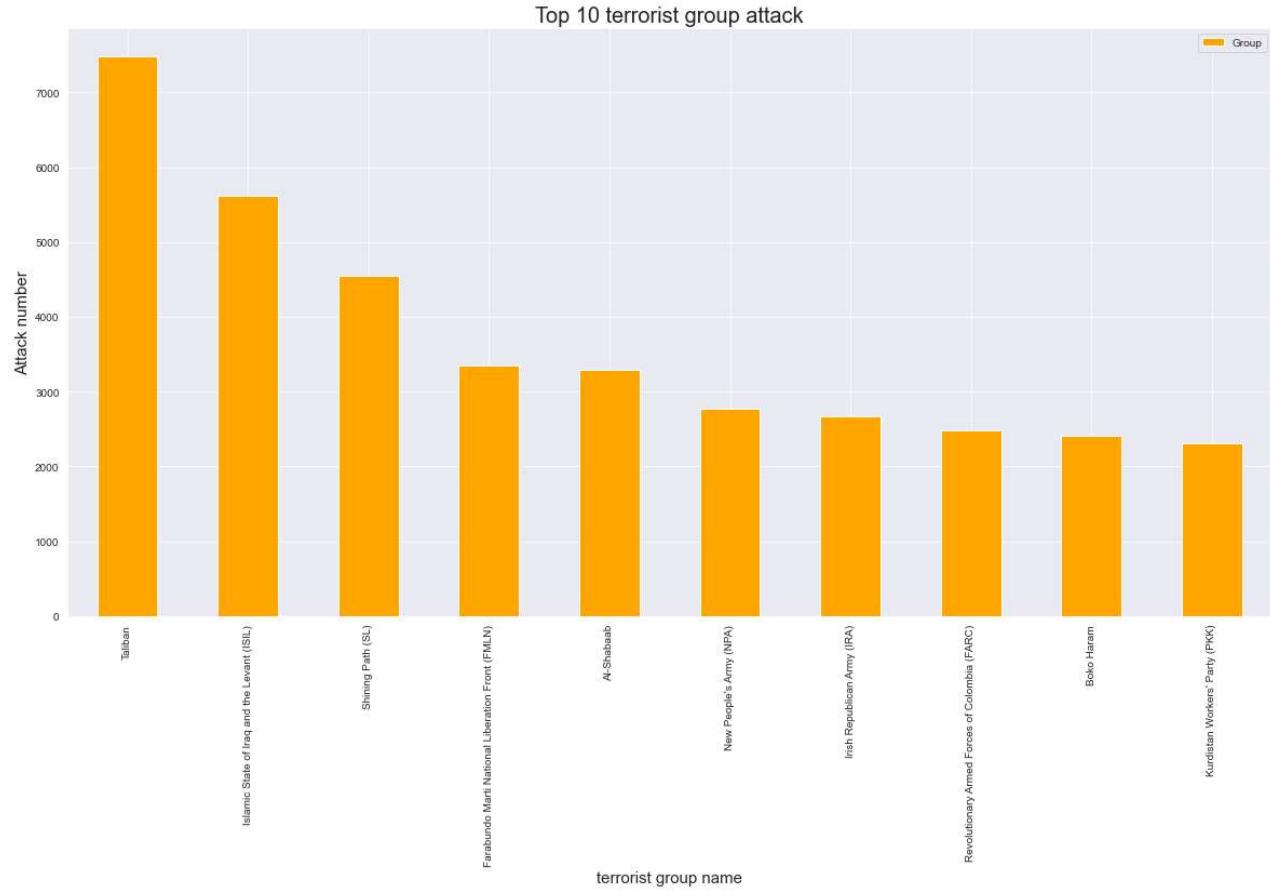


```
In [100]: plt.subplots(figsize=(20,10))
sns.countplot(data["Targettype"],order=data['Targettype'].value_counts().index,palette=
plt.xticks(rotation=90)
plt.xlabel("Attacktype",fontsize=15)
plt.ylabel("count",fontsize=15)
plt.title("Attack per year",fontsize=20)
plt.show()
```



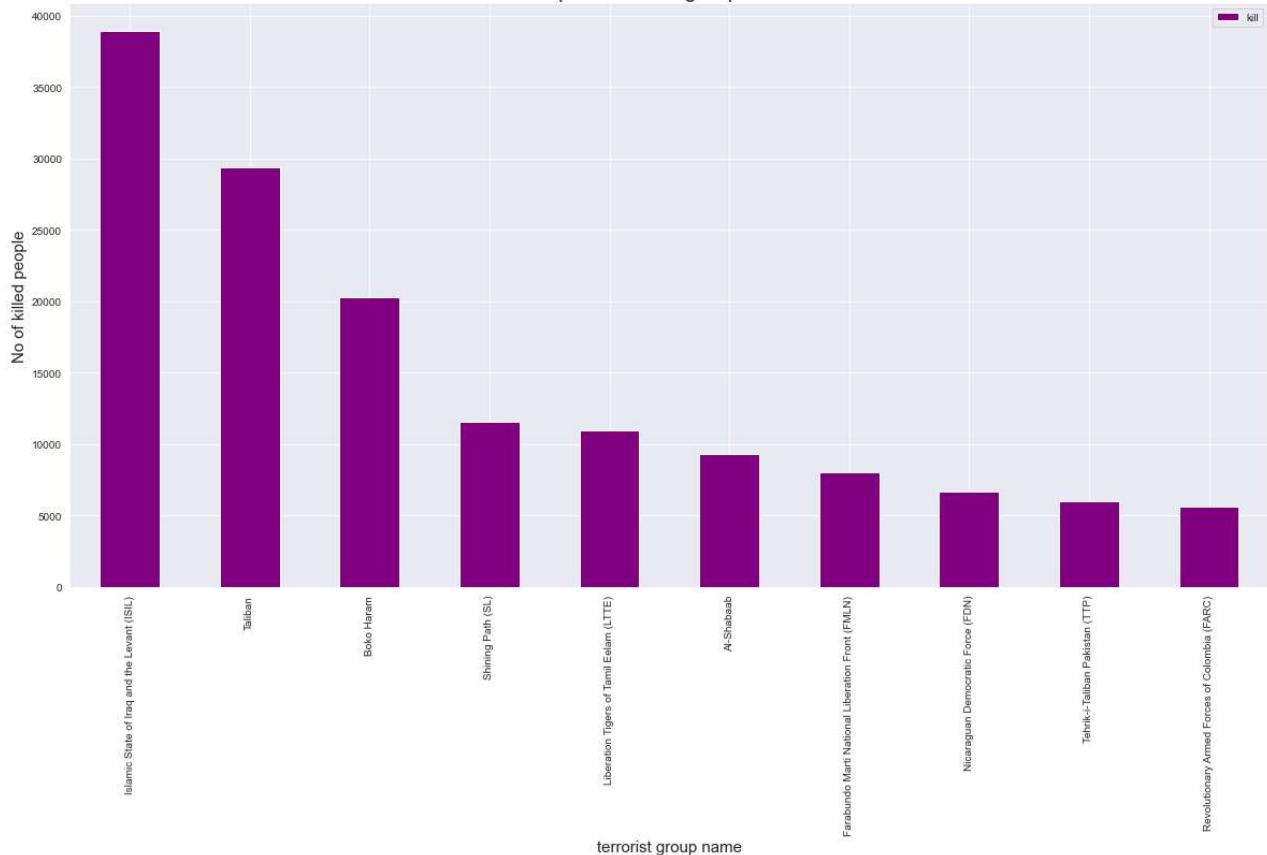
```
In [101]: data['Group'].value_counts().to_frame().drop('Unknown').head(10).plot(kind='bar',color=
plt.title("Top 10 terrorist group attack",fontsize=20)
```

```
plt.xlabel("terrorist group name", fontsize=15)
plt.ylabel("Attack number", fontsize=15)
plt.show()
```



```
In [102]: data[['Group', 'kill']].groupby(['Group'], axis=0).sum().drop('Unknown').sort_values('kill', ascending=False)
plt.title("Top 10 terrorist group attack", fontsize=20)
plt.xlabel("terrorist group name", fontsize=15)
plt.ylabel("No of killed people", fontsize=15)
plt.show()
```

Top 10 terrorist group attack



In [103...]

```
df=data[['Group','Country','kill']]
df=df.groupby(['Group','Country'],axis=0).sum().sort_values('kill',ascending=False).dropna()
df
```

Out[103...]

	Group	Country	kill
0	Islamic State of Iraq and the Levant (ISIL)	Iraq	31058.0
1	Taliban	Afghanistan	29269.0
2	Boko Haram	Nigeria	16917.0
3	Shining Path (SL)	Peru	11595.0
4	Liberation Tigers of Tamil Eelam (LTTE)	Sri Lanka	10928.0
5	Al-Shabaab	Somalia	8176.0
6	Farabundo Martí National Liberation Front (FMLN)	El Salvador	8019.0
7	Islamic State of Iraq and the Levant (ISIL)	Syria	6883.0
8	Nicaraguan Democratic Force (FDN)	Nicaragua	6630.0
9	Tehrik-i-Taliban Pakistan (TTP)	Pakistan	6014.0

In [104...]

```
kill = data.loc[:, 'kill']
print('Number of people killed by terror attack:', int(sum(kill.dropna())))
```

Number of people killed by terror attack: 411868

In [105...]

```
typeKill = data.pivot_table(columns='Attacktype', values='kill', aggfunc='sum')
```

Out[105...]

Attacktype	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	Hostage Taking (Barricade Incident) (K)
kill	160297.0	24920.0	157321.0	3642.0	3718.0	4478.0



In [106...]

```
countryKill = data.pivot_table(columns='Country', values='kill', aggfunc='sum')
countryKill
```

Out[106...]

Country	Afghanistan	Albania	Algeria	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	Australia	Au
kill	39384.0	42.0	11066.0	0.0	3043.0	0.0	490.0	37.0	23.0	

1 rows × 205 columns



Conclusion and Results :

Country with the most attacks: Iraq

City with the most attacks: Baghdad

Region with the most attacks: Middle East & North Africa

Year with the most attacks: 2014

Month with the most attacks: 5

Group with the most attacks: Taliban

Most Attack Types: Bombing/Explosion