

Author : ATANU DAS

Task 4 : Exploratory Data Analysis on dataset Global Terrorism

Objective :

1. Terrorism ia a threat of violence that creates fear in a population. It is important to understand that all violence acts are not Terrorist acts.
 2. As a security / defence analyst , find out the hot zone of terrorism and derive some security issues and insights by EDA.
-

GRIP @ The Spark's Foundation

Role : Data Science And Business Analytics

What this data set is all about ?

- Consists pf information or more than 180,000 Terrorist Attacks.
- Includes information on terrorist attacks around the world from 1970 to 2017.
- Includes systematic data on domestic as well as international terrorist incidents that have occured during specific time period.



THE GLOBAL IMPACT OF TERRORISM

Importing Libraries

In []:

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

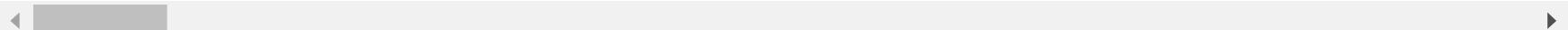
In []:

```
data = pd.read_csv("/content/sample_data/globalterrorismdb_0718dist.csv",encoding='latin1')
data.head()
```

Out[]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	region_txt	provstate	city	latitu
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	Central America & Caribbean	NaN	Santo Domingo	18.4561
1	1.970000e+11	1970	0	0	NaN	0	NaN	130	Mexico	1	North America	Federal	Mexico city	19.3718
2	1.970010e+11	1970	1	0	NaN	0	NaN	160	Philippines	5	Southeast Asia	Tarlac	Unknown	15.4781

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	region_txt	provstate	city	latitude	longitude
3	1.970010e+11	1970	1	0	NaN	0	NaN	78	Greece	8	Western Europe	Attica	Athens	37.9974	23.7272
4	1.970010e+11	1970	1	0	NaN	0	NaN	101	Japan	4	East Asia	Fukouka	Fukouka	33.5802	130.4278



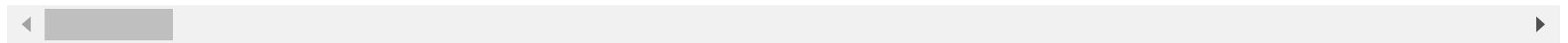
In []:

data.tail()

Out[]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	region_txt	provstate	city	latitude	longitude
181686	2.017120e+11	2017	12	31	NaN	0	NaN	182	Somalia	11	Sub-Saharan Africa	Middle Shebelle	Ceelk Geelo'	5.2000	43.0000
181687	2.017120e+11	2017	12	31	NaN	0	NaN	200	Syria	10	Middle East & North Africa	Lattakia	Jable	35.0000	36.5000
181688	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	Southeast Asia	Maguindanao	Kubento	11.0000	123.0000
181689	2.017120e+11	2017	12	31	NaN	0	NaN	92	India	6	South Asia	Manipur	Imphal	24.8000	91.7000

eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	region_txt	provstate	cit
181690	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	Southeast Asia	Maguindanao Cotabat Cit



In []: `data.columns.values`

```
Out[ ]: 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 [ ]: data.rename(columns={'iyear':'Year','imonth':'Month','iday':"day",'gname':'Group','country_txt':'Country','region_txt':'Region','longitude':'longitude','summary':'summary','attacktype1_txt':'Attacktype','targtype1_txt':'Targettype','weaptype1_txt':'Weapon','nwound':'Wound'},inplace=True)
```

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

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

	Year	Month	day	Country	State	Region	City	latitude	longitude	Attacktype	kill	Wound	target1	summary
0	1970	7	2	Dominican Republic	NaN	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	1.0	0.0	Julio Guzman	NaN
1	1970	0	0	Mexico	Federal	North America	Mexico city	19.371887	-99.086624	Hostage Taking (Kidnapping)	0.0	0.0	Nadine Chaval, daughter	NaN
2	1970	1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741	Assassination	1.0	0.0	Employee	NaN
3	1970	1	0	Greece	Attica	Western Europe	Athens	37.997490	23.762728	Bombing/Explosion	NaN	NaN	U.S. Embassy	NaN
4	1970	1	0	Japan	Fukouka	East Asia	Fukouka	33.580412	130.396361	Facility/Infrastructure Attack	NaN	NaN	U.S. Consulate	NaN

```
In [ ]: data.shape
```

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

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

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

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

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

	Year	Month	day	latitude	longitude	kill	Wound	Casualties
count	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181691.000000	181691.000000	181691.000000
mean	2002.638997	6.467277	15.505644	23.498343	-4.586957e+02	2.266860	2.883296	5.150156
std	13.259430	3.388303	8.814045	18.569242	2.047790e+05	11.227057	34.309747	40.555416
min	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	0.000000	0.000000	0.000000
25%	1991.000000	4.000000	8.000000	11.510046	4.545640e+00	0.000000	0.000000	0.000000
50%	2009.000000	6.000000	15.000000	31.467463	4.324651e+01	0.000000	0.000000	1.000000
75%	2014.000000	9.000000	23.000000	34.685087	6.871033e+01	2.000000	2.000000	4.000000
max	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	1570.000000	8191.000000	9574.000000

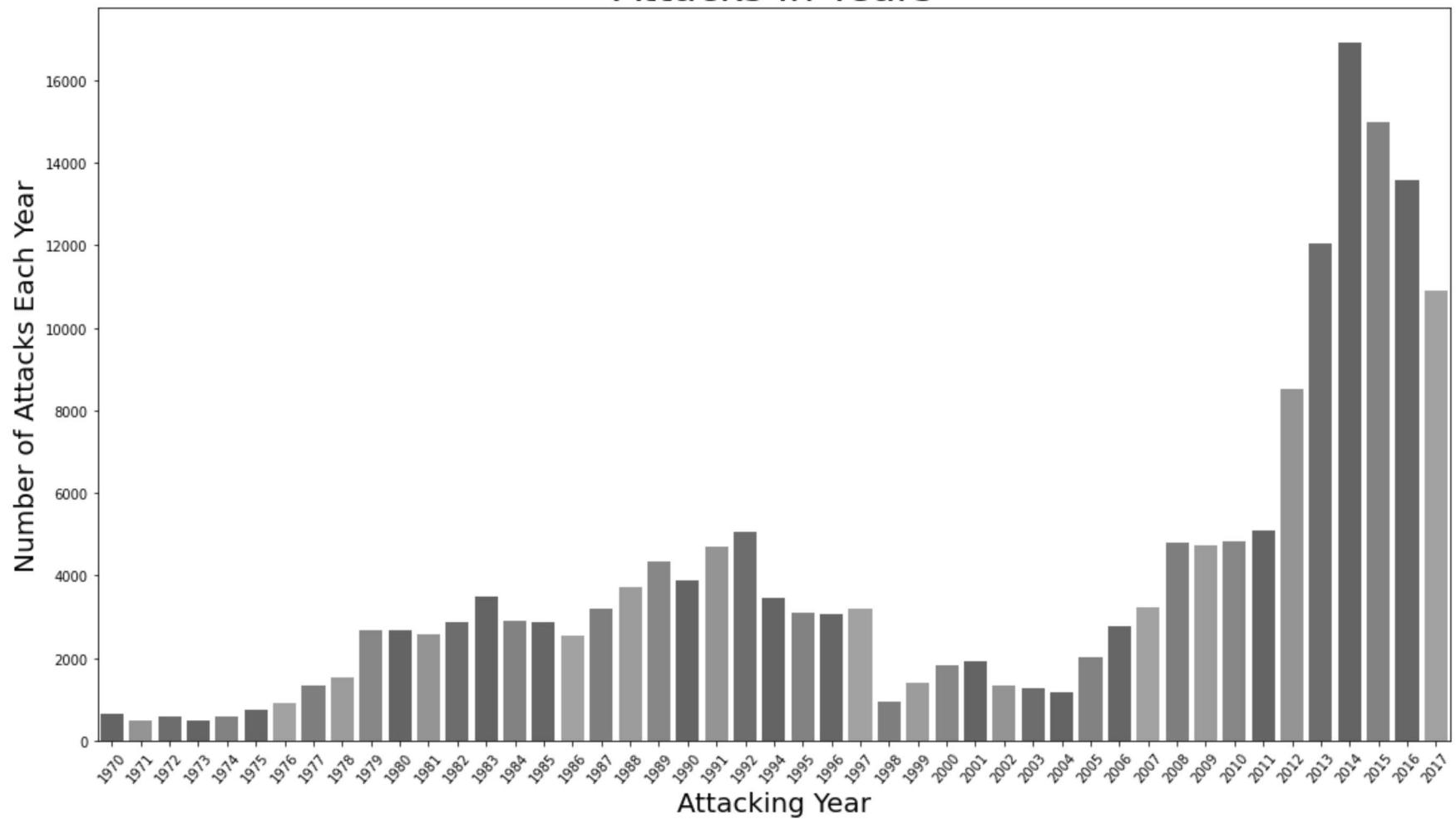
In []:

```

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,
            palette = "tab10")
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

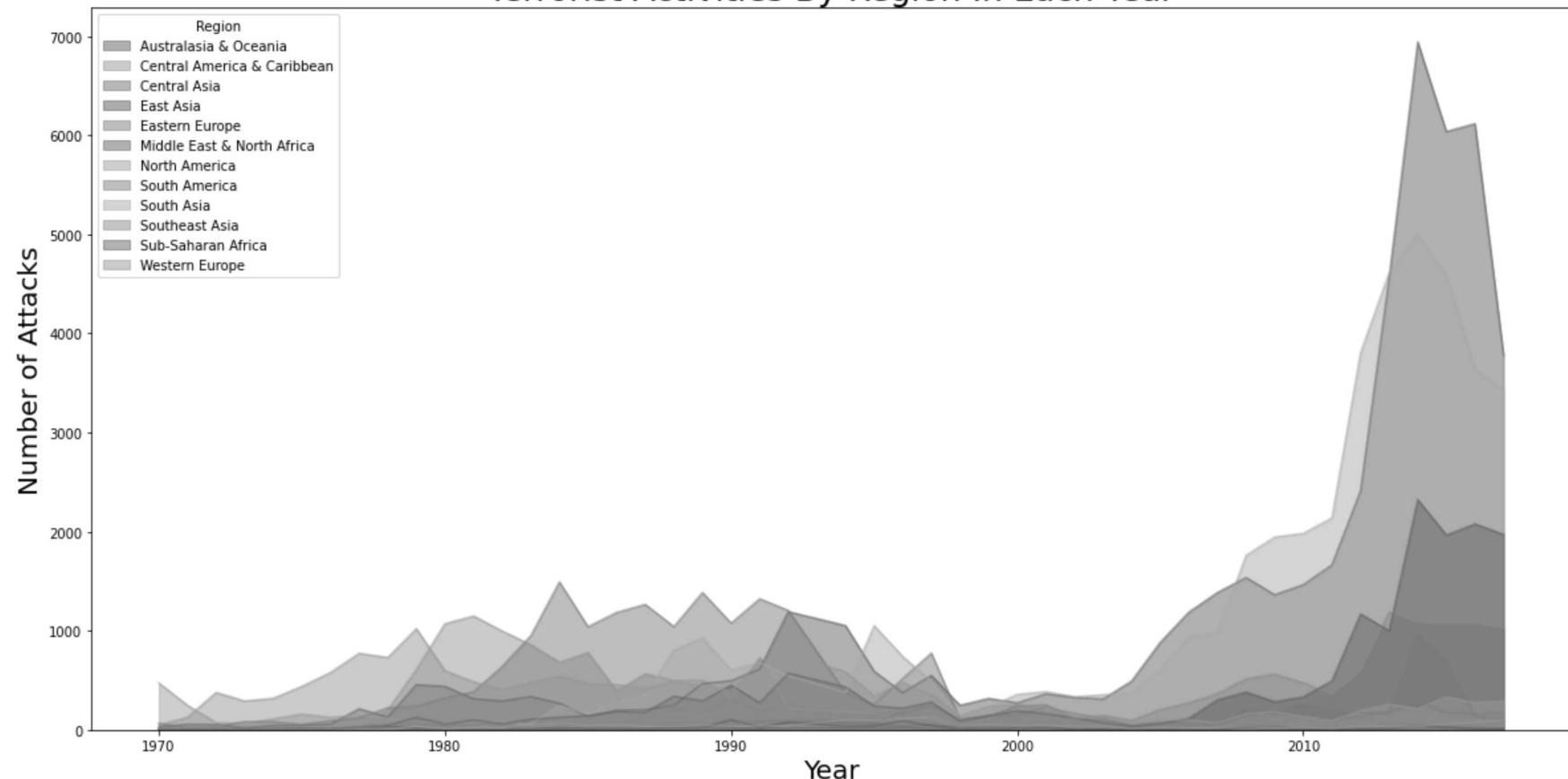


- In the year 2014 the number of attacks are at the highest peak.

In []:

```
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



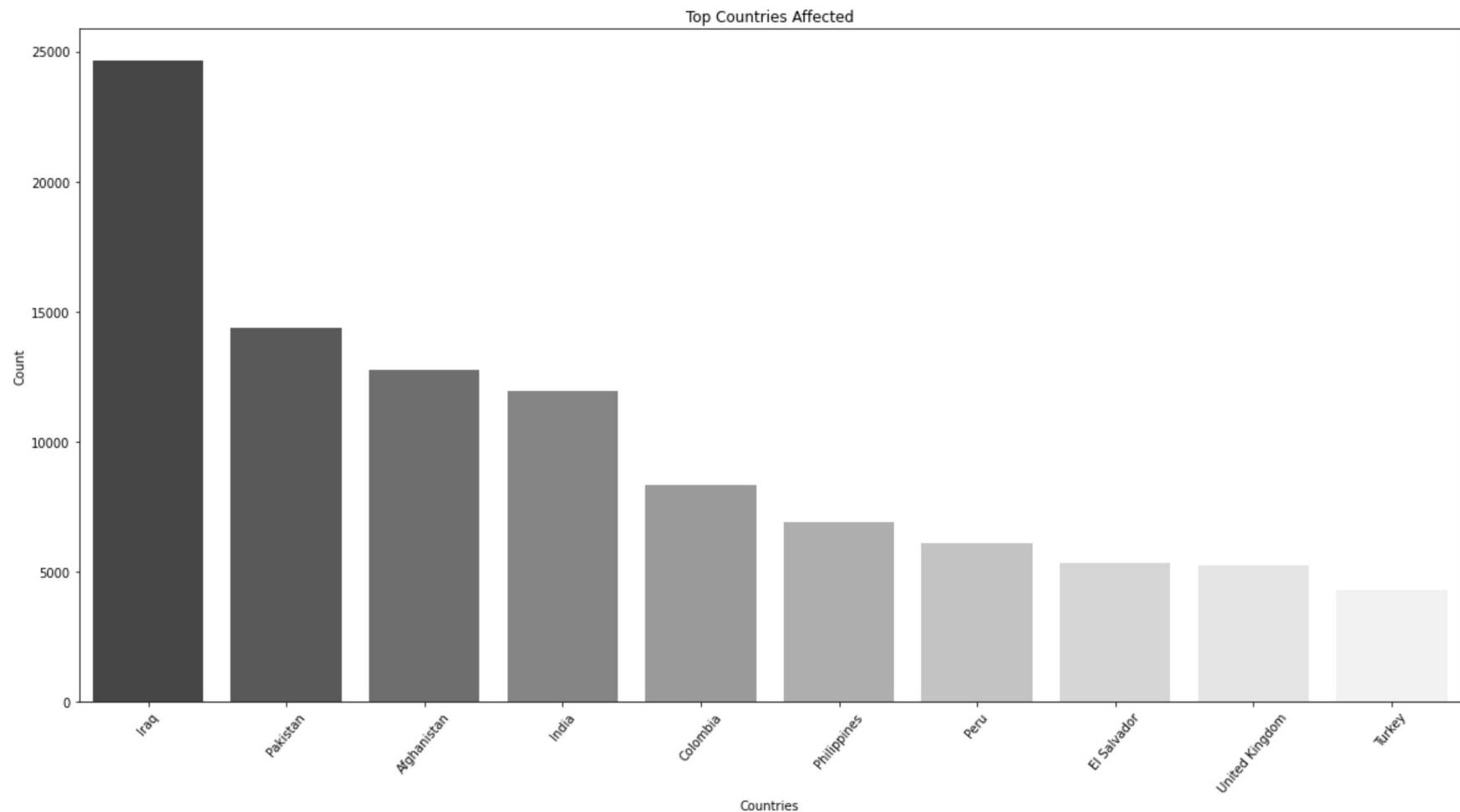
```
In [ ]: attack = data.Country.value_counts()[:10]
attack
```

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

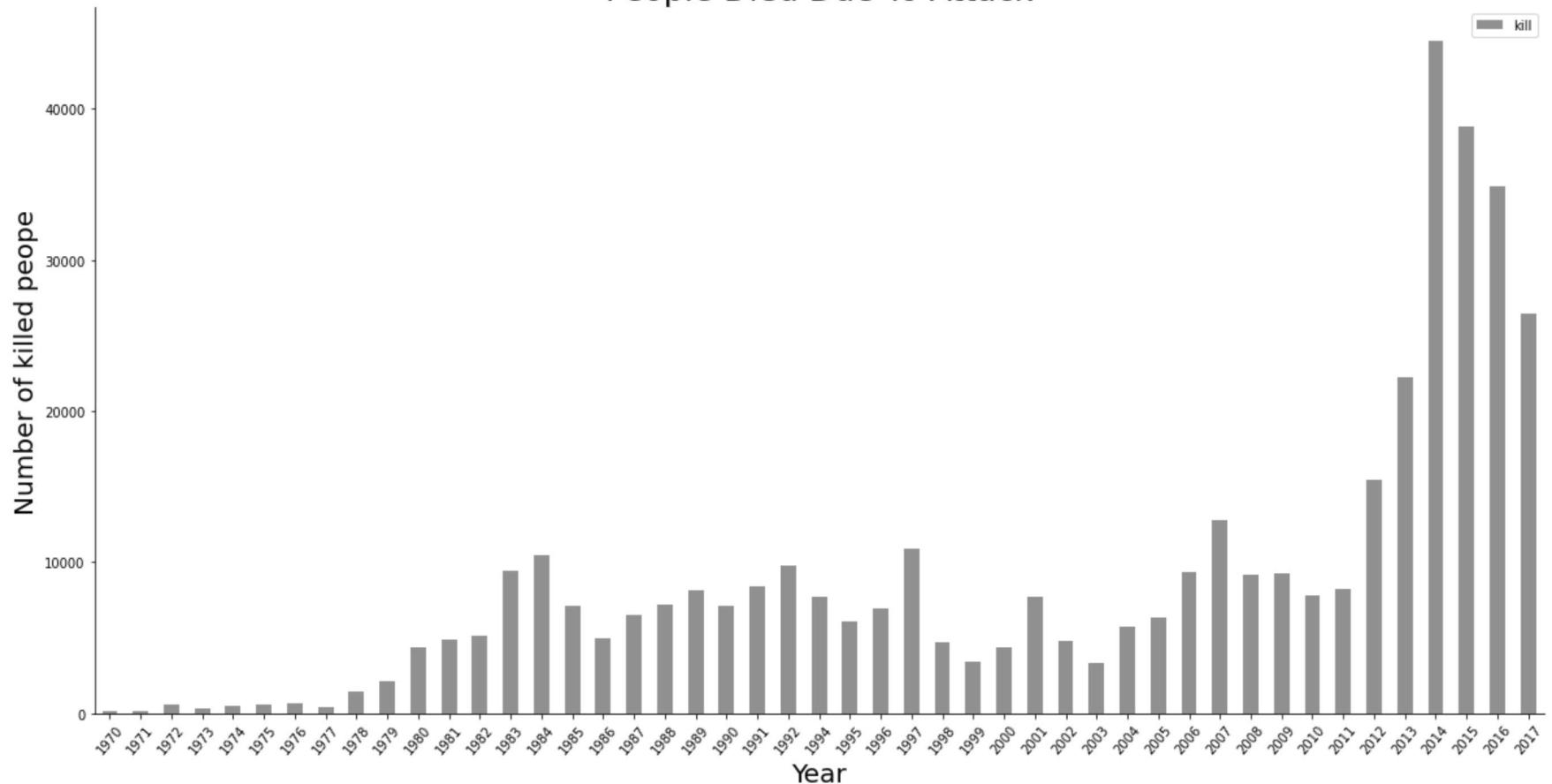
```
Out[ ]: Taliban                                7478
Islamic State of Iraq and the Levant (ISIL)    5613
Shining Path (SL)                             4555
Farabundo Marti National Liberation Front (FMLN) 3351
Al-Shabaab                               3288
New People's Army (NPA)                  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

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



```
In [ ]:  
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)
```

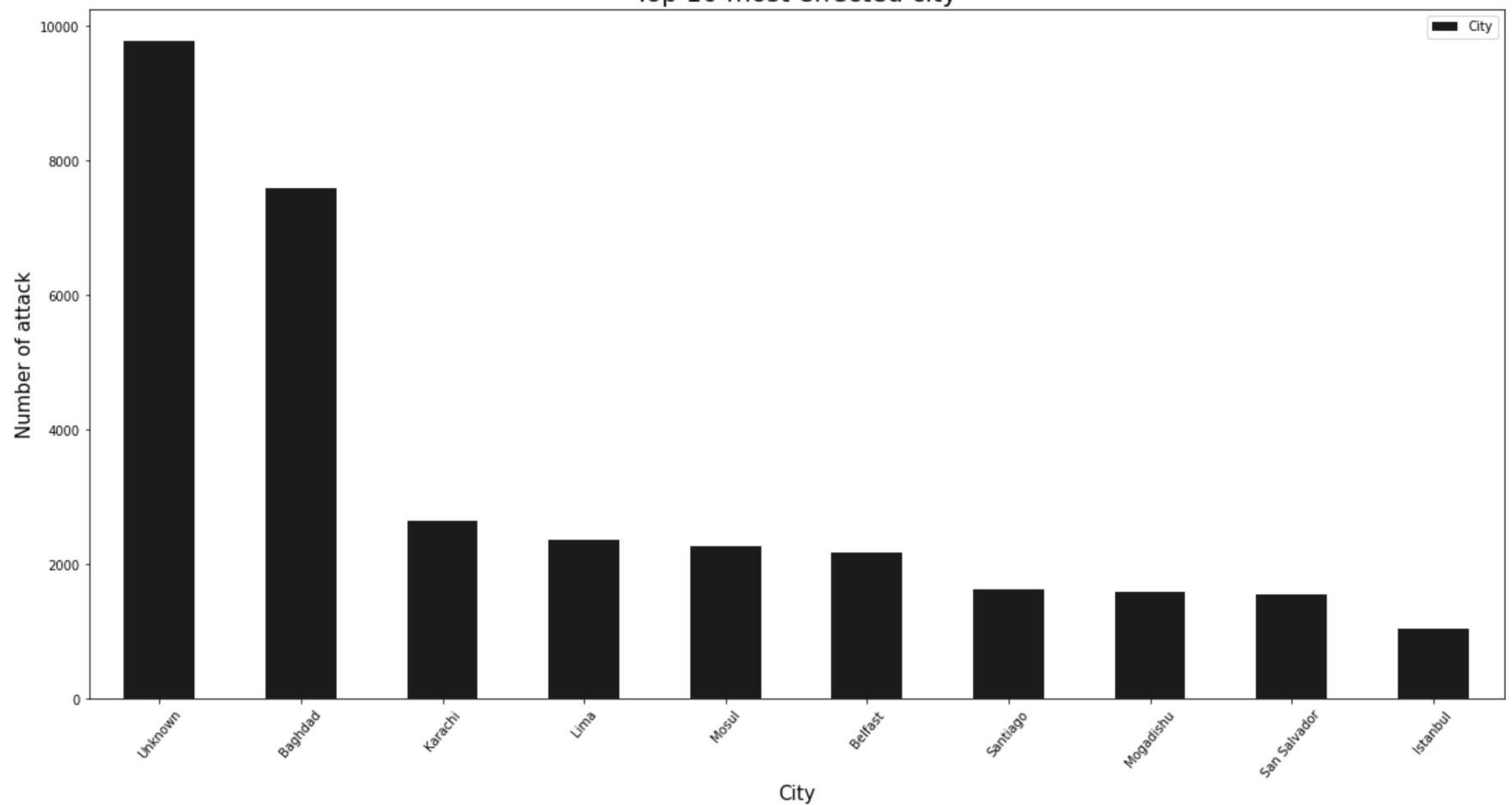
People Died Due To Attack



In []:

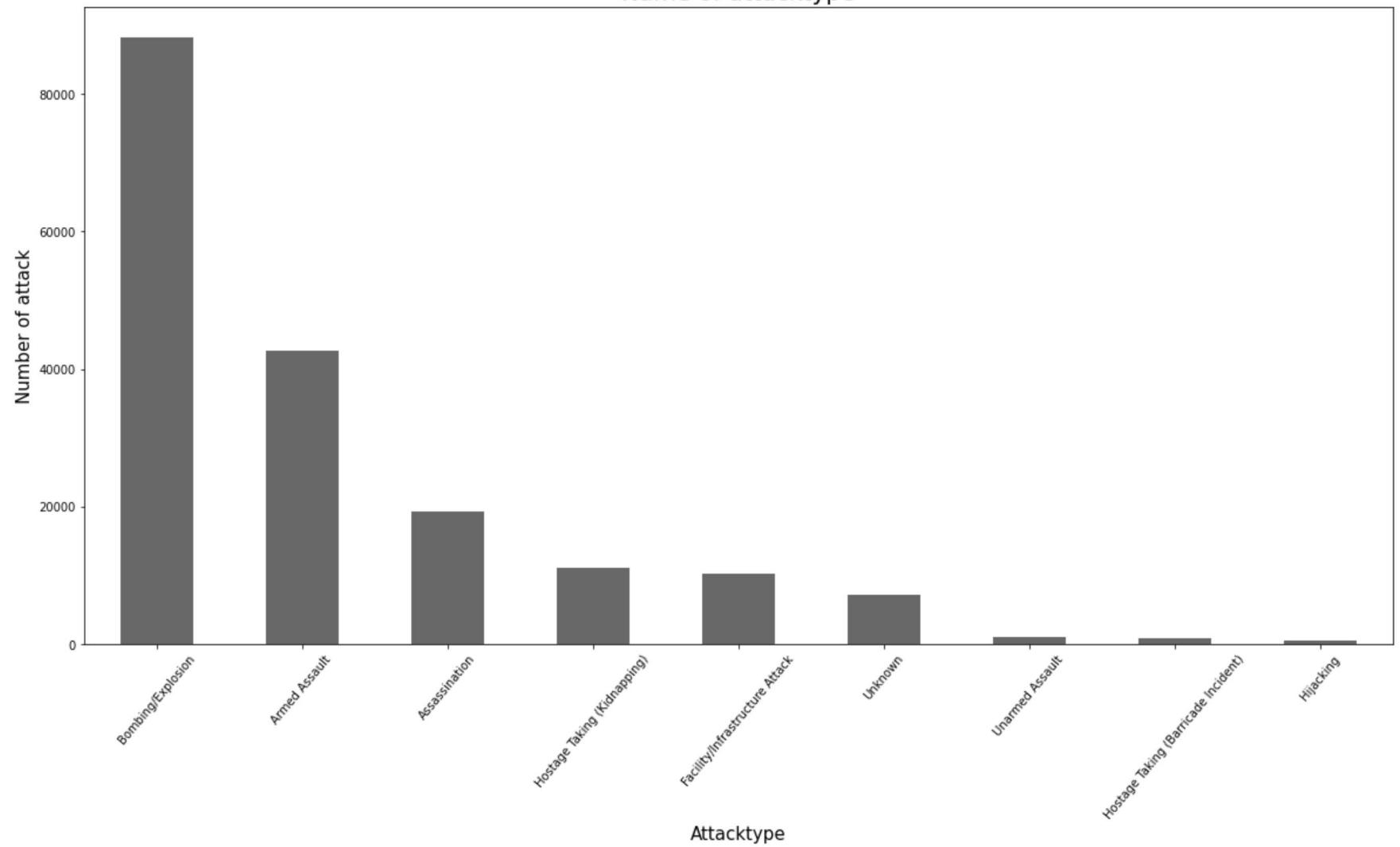
```
data['City'].value_counts().to_frame().sort_values('City',axis=0,ascending=False).head(10).plot(kind='bar',figsize=(20,10))
plt.xticks(rotation = 50)
plt.xlabel("City",fontsize=15)
plt.ylabel("Number of attack",fontsize=15)
plt.title("Top 10 most effected city",fontsize=20)
plt.show()
```

Top 10 most effected city



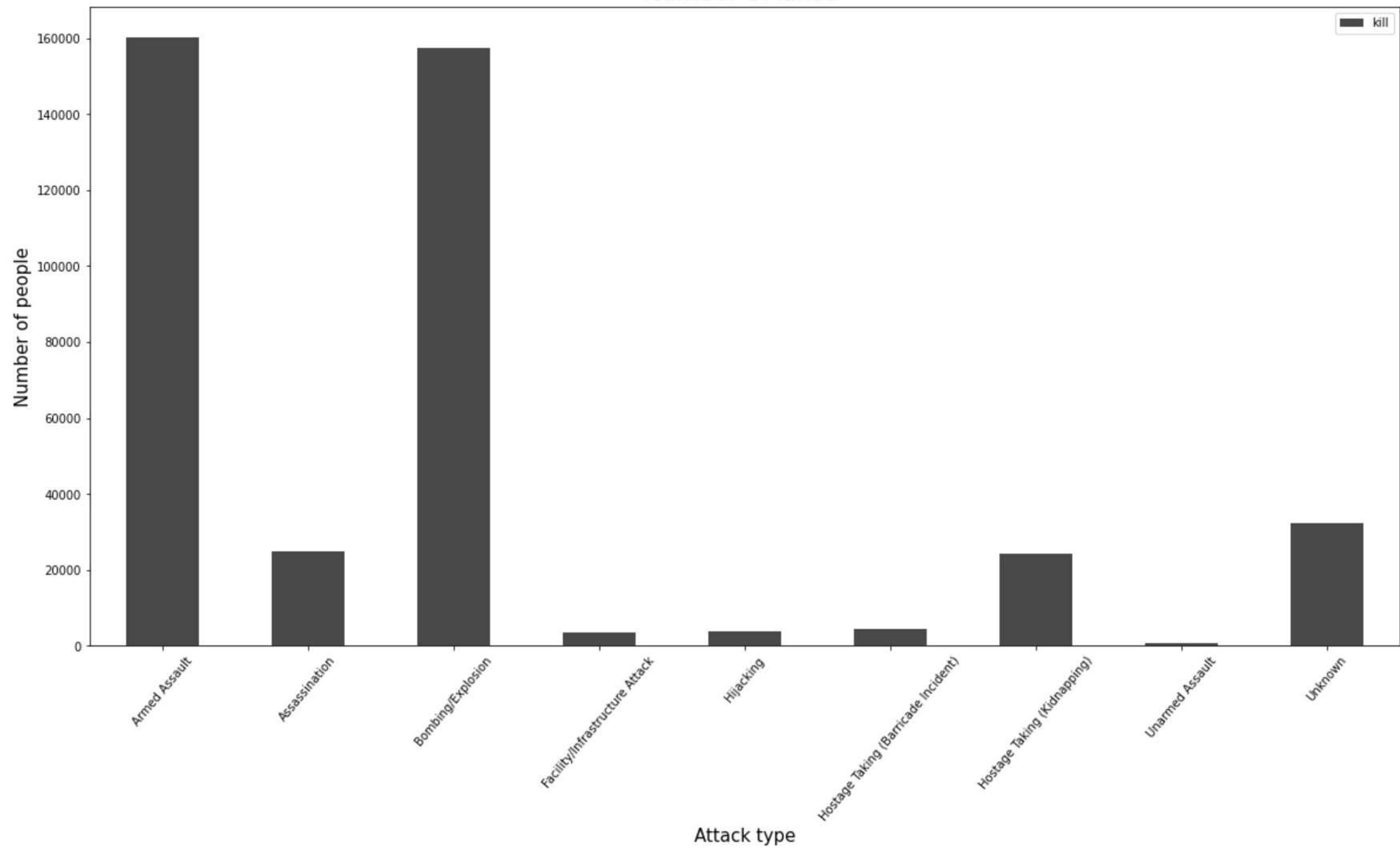
```
In [ ]:  
data['Attacktype'].value_counts().plot(kind='bar', figsize=(20,10), color='magenta')  
plt.xticks(rotation = 50)  
plt.xlabel("Attacktype", fontsize=15)  
plt.ylabel("Number of attack", fontsize=15)  
plt.title("Name of attacktype", fontsize=20)  
plt.show()
```

Name of attacktype



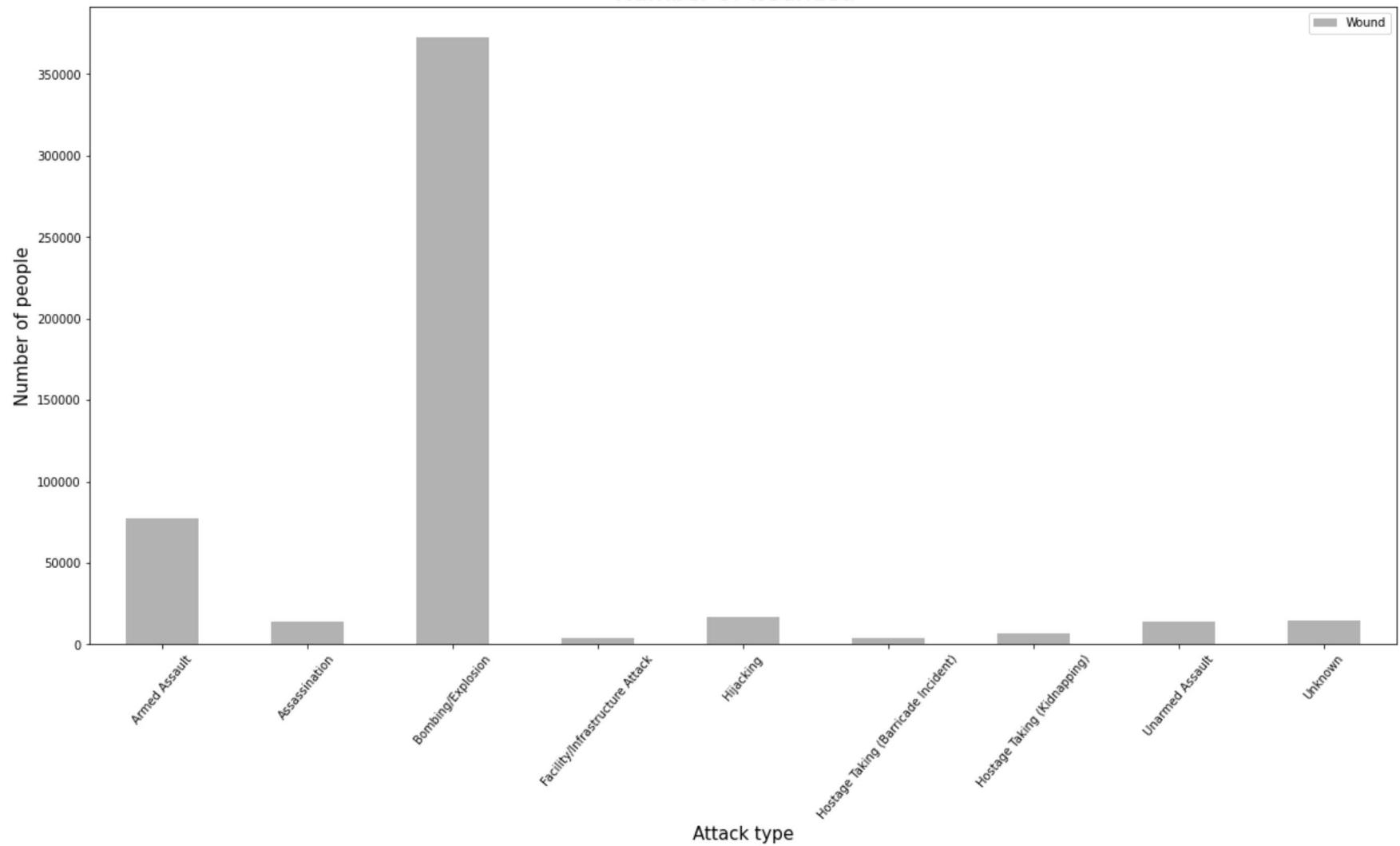
```
In [ ]: data[['Attacktype','kill']].groupby(["Attacktype"],axis=0).sum().plot(kind='bar',figsize=(20,10),color=['darkslateblue'])
plt.xticks(rotation=50)
plt.title("Number of killed ",fontsize=20)
plt.ylabel('Number of people',fontsize=15)
plt.xlabel('Attack type',fontsize=15)
plt.show()
```

Number of killed



```
In [ ]:  
data[['Attacktype','Wound']].groupby(["Attacktype"],axis=0).sum().plot(kind='bar',figsize=(20,10),color=['cyan'])  
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()
```

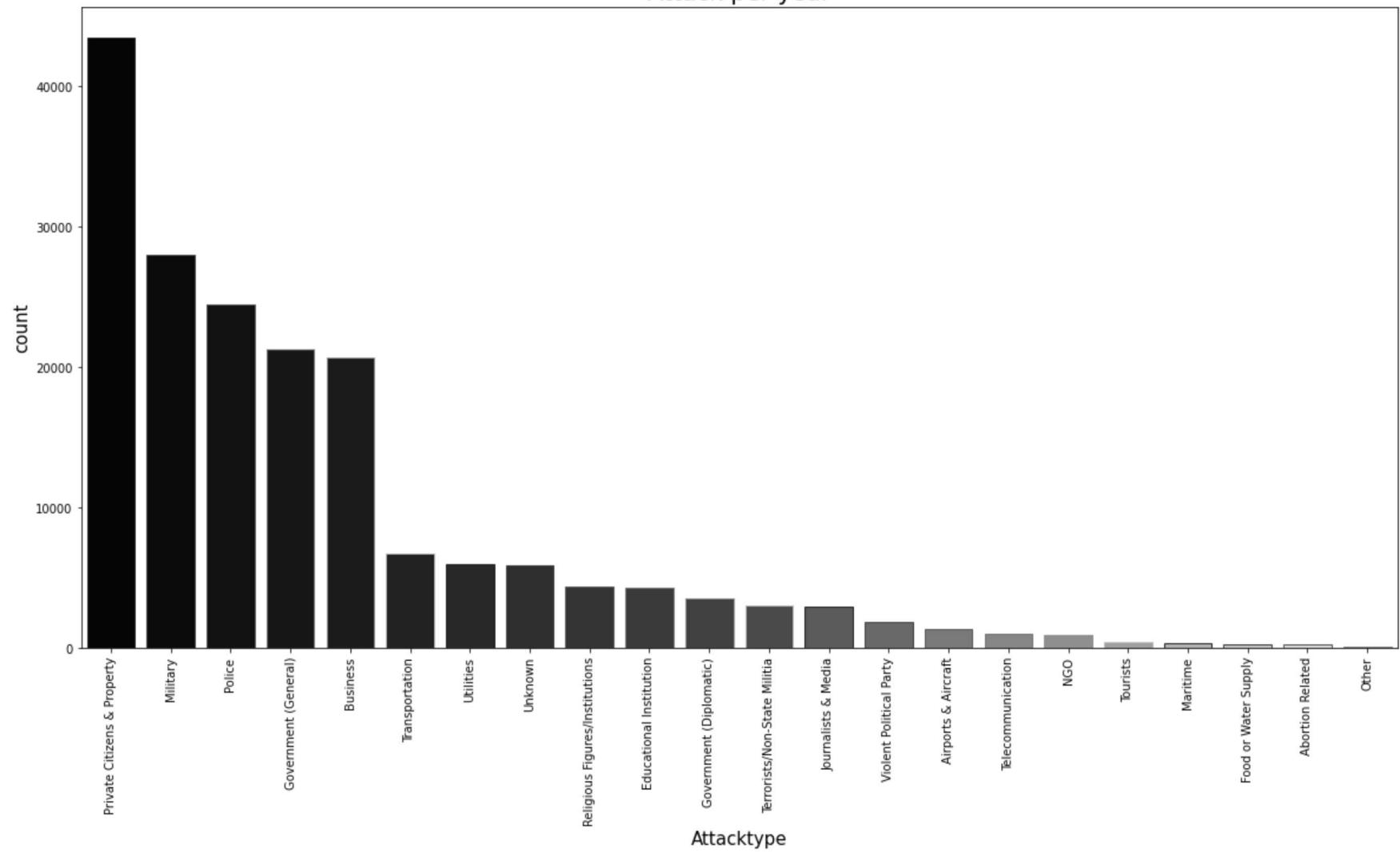
Number of wounded



In []:

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

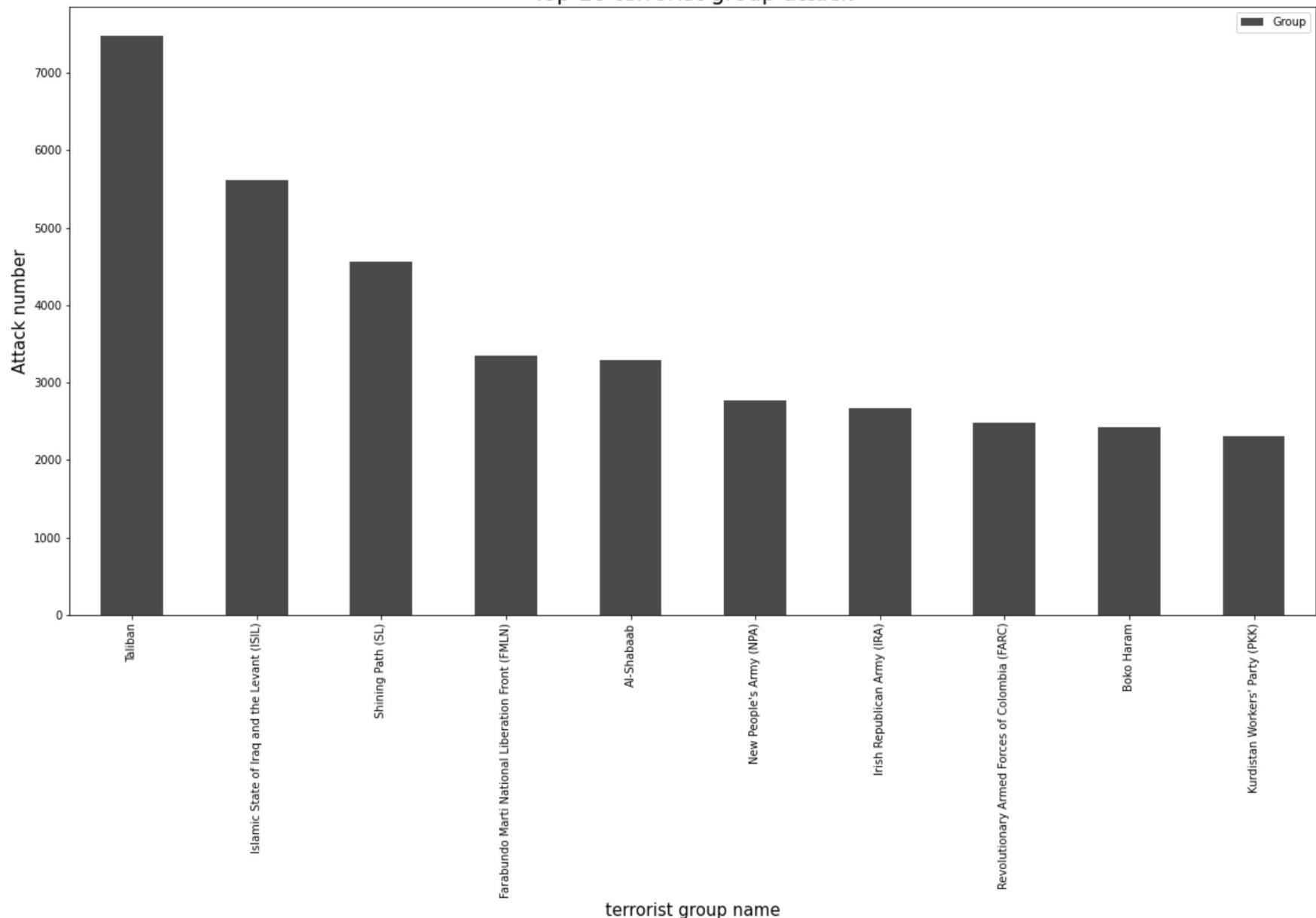
Attack per year



In []:

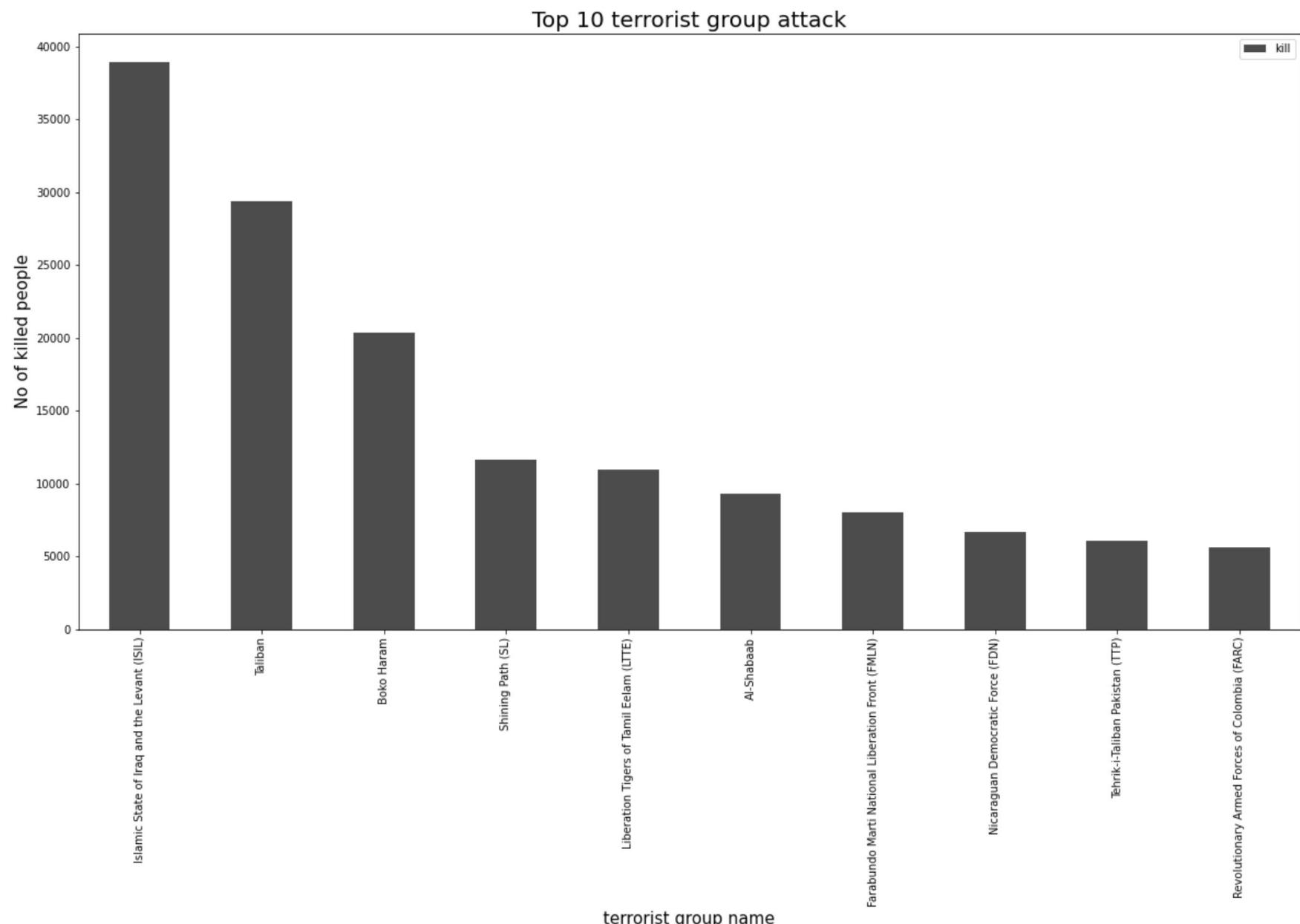
```
data['Group'].value_counts().to_frame().drop('Unknown').head(10).plot(kind='bar', color='green', figsize=(20,10))
plt.title("Top 10 terrorist group attack", fontsize=20)
plt.xlabel("terrorist group name", fontsize=15)
plt.ylabel("Attack number", fontsize=15)
plt.show()
```

Top 10 terrorist group attack



In []:

```
data[['Group','kill']].groupby(['Group'],axis=0).sum().drop('Unknown').sort_values('kill',ascending=False).head(10).plot()  
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()
```



In []:

```
df=data[['Group','Country','kill']]
df=df.groupby(['Group','Country'],axis=0).sum().sort_values('kill',ascending=False).drop('Unknown').reset_index().head(10)
df
```

Out[]:

	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 Marti 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 []:

```
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 []:

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

Out[]:

Attacktype	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	Hostage Taking (Barricade Incident)	Hostage Taking (Kidnapping)	Unarmed Assault	Unknown
kill	160297.0	24920.0	157321.0	3642.0	3718.0	4478.0	24231.0	880.0	32381.0

In []:

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

Out[]:

Country	Afghanistan	Albania	Algeria	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	Australia	Austria	Azerbaijan	Bahamas	Bahrain	Ba
kill	39384.0	42.0	11066.0	0.0	3043.0	0.0	490.0	37.0	23.0	30.0	258.0	1.0	44.0	

Conclusion :

1. Country with the most attacks: **Iraq**
2. City with the most attacks: **Baghdad**
3. Region with the most attacks: **Middle East & North Africa**
4. Year with the most attacks: **2014**
5. Month with the most attacks: **5**
6. Group with the most attacks: **Taliban**
7. Most Attack Types: **Bombing/Explosion**