Task-2 : INTERMEDIATE LEVEL TASK

Task 01: Exploratory Data Analysis on Dataset - Terrorism

As a security/defense analyst, try to find out the hot zone of terrorism. You can choose any of the tool of your choice

(Python/R/Tableau/PowerBI/Excel/SAP/SAS) Watch Tutorial from here https://youtu.be/CBCfOTePVPo (https://youtu.be/CBCfOTePVPo)

Dataset: https://bit.ly/2TK5Xn5 (https://bit.ly/2TK5Xn5)

Importing Libraries

In [1]:

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

Importing Dataset

In [2]:

```
# Loading iris dataset into the notebook
df = pd.read_csv("globalterrorismdb_0718dist.csv",encoding='latin1')
print(" Dataset loaded successfully")
```

```
 Dataset loaded successfully
```

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(55), int64(22), object(58)
memory usage: 187.1+ MB
```

In [4]:

```
df.head()
```

Out[4]:

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

5 rows × 135 columns

In [5]:

```
df.describe()
```

Out[5]:

	eventid	iyear	imonth	iday	extended	count
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.0000(
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.9685(
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.4145:
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.0000(
25%	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.0000(
50%	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.0000(
75%	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.0000(
max	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.0000(

8 rows × 77 columns

In [6]:

```
df.columns.values
```

Out[6]:

```
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 [7]:

```
df.rename(columns={'iyear':'Year','imonth':'Month','iday':"day",'gname':'Group','country_tx
    'longitude':'longitude','summary':'summary','attacktype1_txt':'Attacktype','targtype1_t
     'nwound':'Wound'},inplace=True)
```

In [8]:

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

In [9]:

```
data.head()
```

Out[9]:

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

In [10]:

```
data.shape
```

Out[10]:

```
(181691, 18)
```

In [11]:

```
data.isnull().sum()
```

Out[11]:

```
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 [12]:

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

In [13]:

```
data['Casualities'] = data['kill'] + data['Wound']
```

In [14]:

```
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  Casualities  181691 non-null  float64
dtypes: float64(5), int64(3), object(11)
memory usage: 26.3+ MB
```

In [15]:

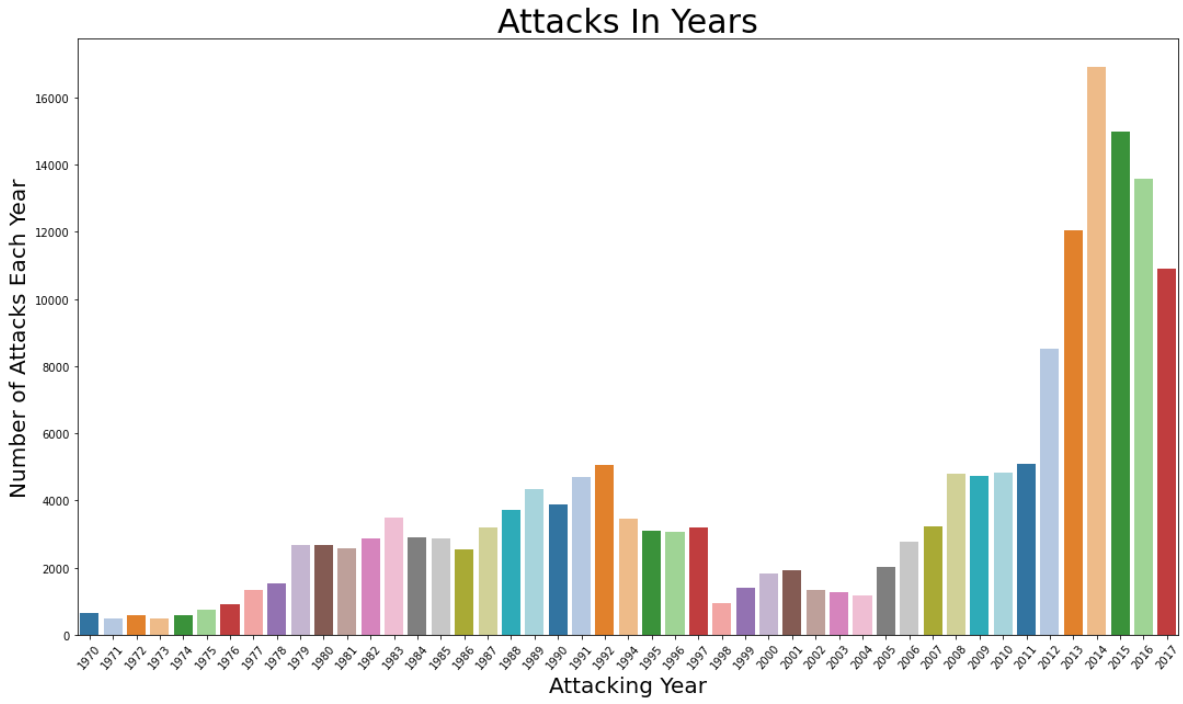
```
data.describe()
```

Out[15]:

	Year	Month	day	latitude	longitude	
count	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181691.0000
mean	2002.638997	6.467277	15.505644	23.498343	-4.586957e+02	2.2668
std	13.259430	3.388303	8.814045	18.569242	2.047790e+05	11.2270
min	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	0.0000
25%	1991.000000	4.000000	8.000000	11.510046	4.545640e+00	0.0000
50%	2009.000000	6.000000	15.000000	31.467463	4.324651e+01	0.0000
75%	2014.000000	9.000000	23.000000	34.685087	6.871033e+01	2.0000
max	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	1570.0000

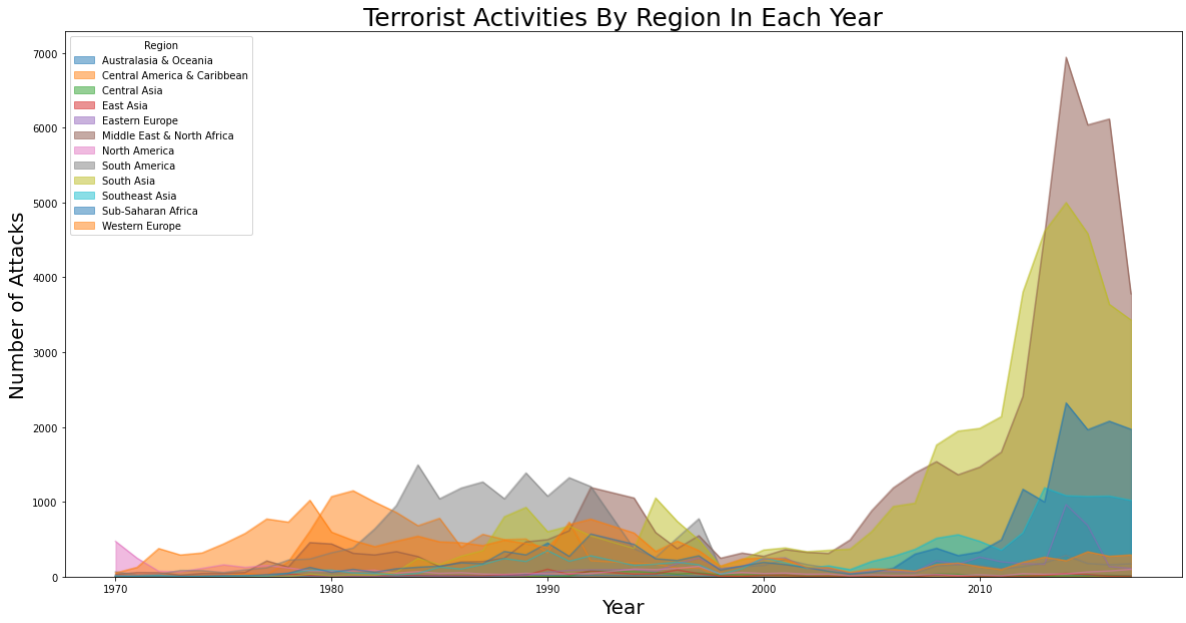
In [16]:

```
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 = "tab20")
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()
```



In [17]:

```
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()
```



In [18]:

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

Out[18]:

```
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 [19]:

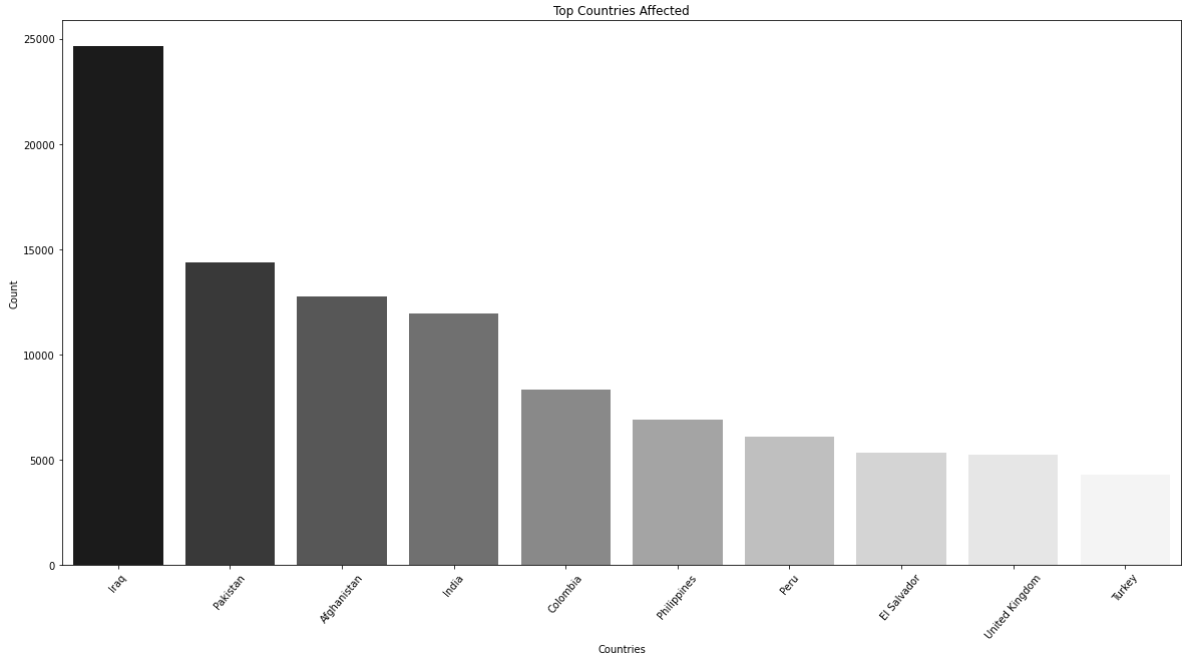
```
data.Group.value_counts()[1:10]
```

Out[19]:

```
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
```

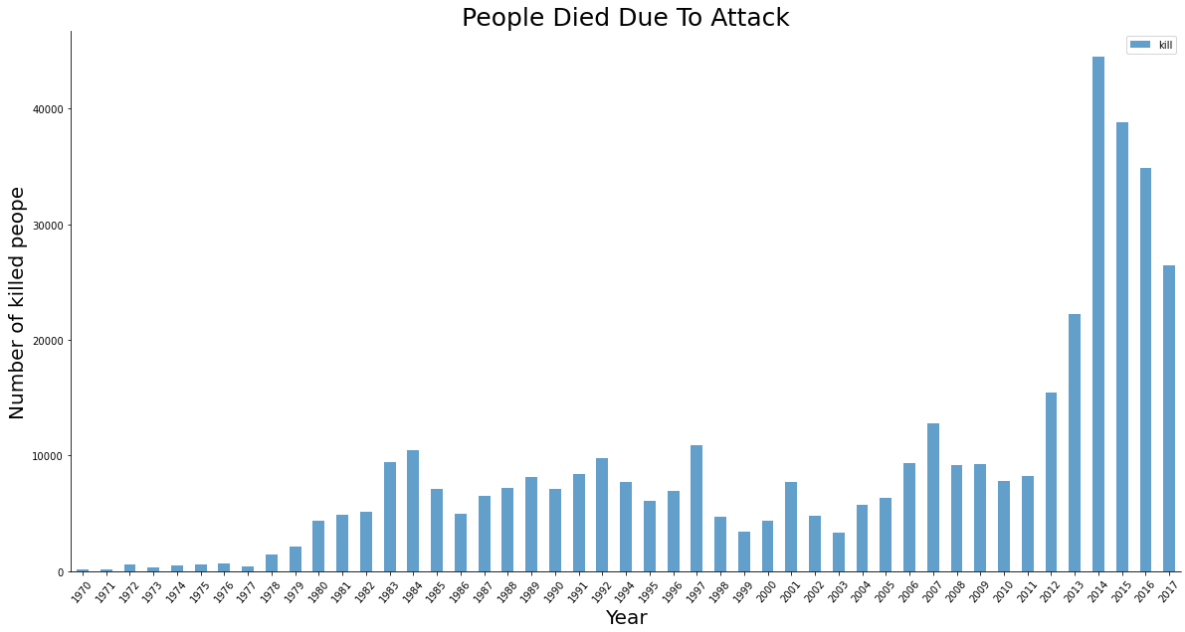
In [20]:

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



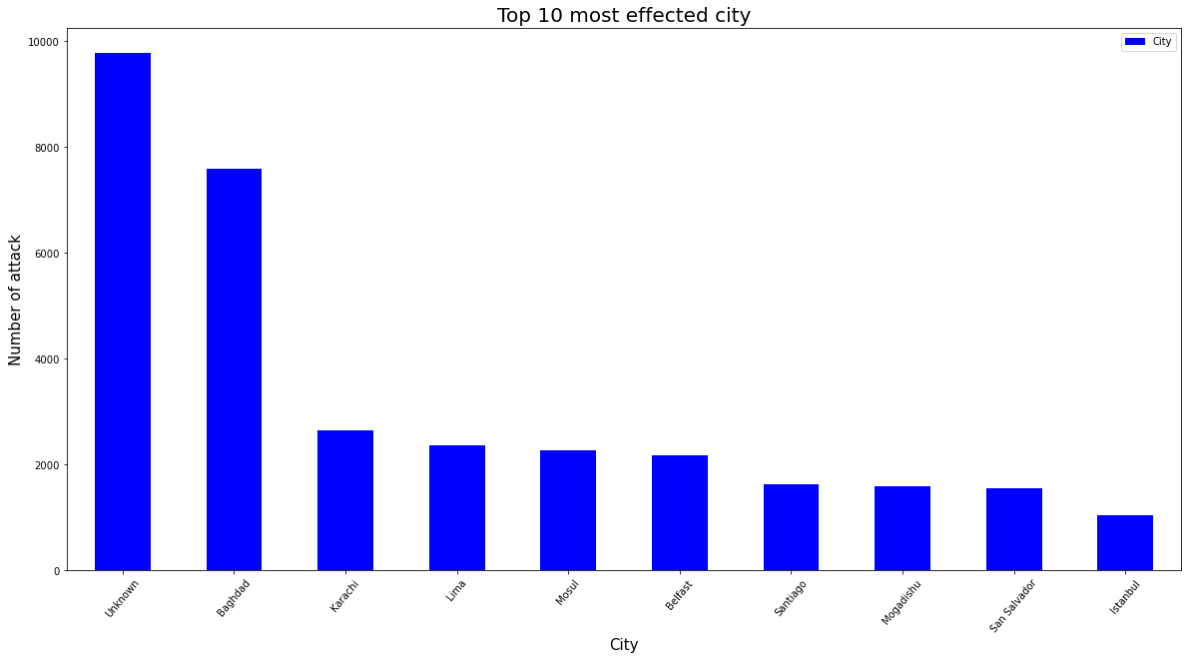
In [21]:

```
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 peope",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)
```



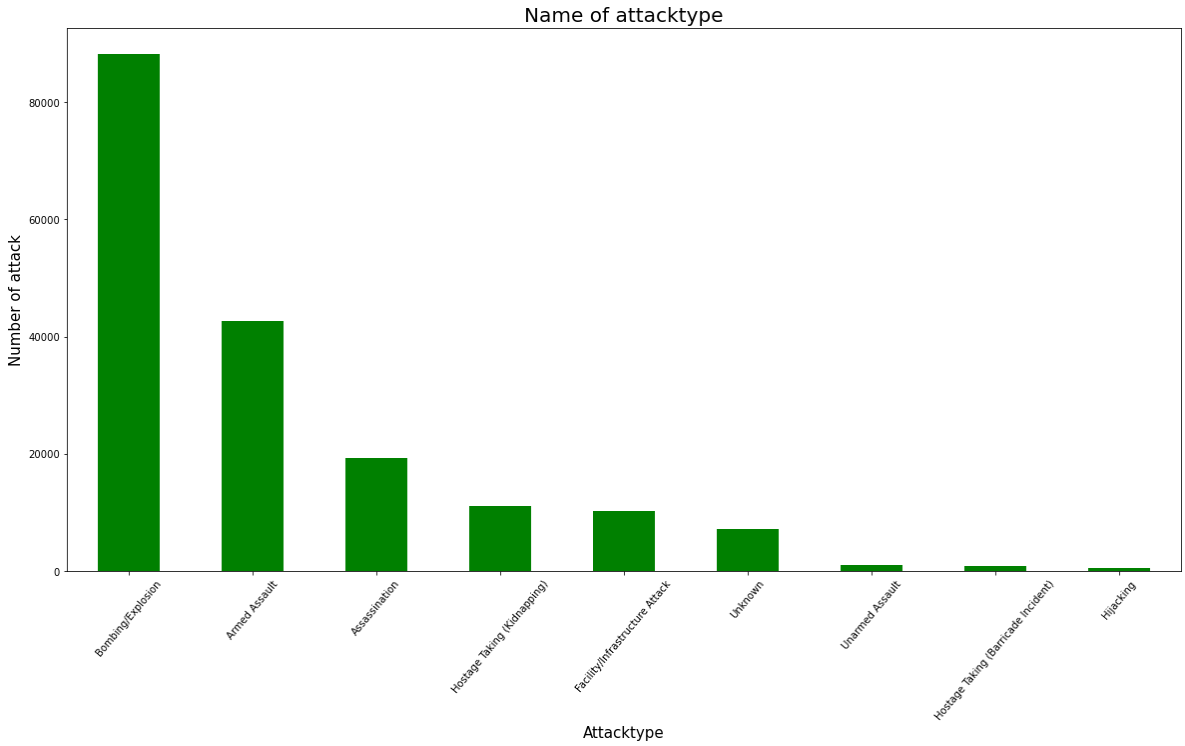
In [22]:

```
data['City'].value_counts().to_frame().sort_values('City',axis=0,ascending=False).head(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()
```



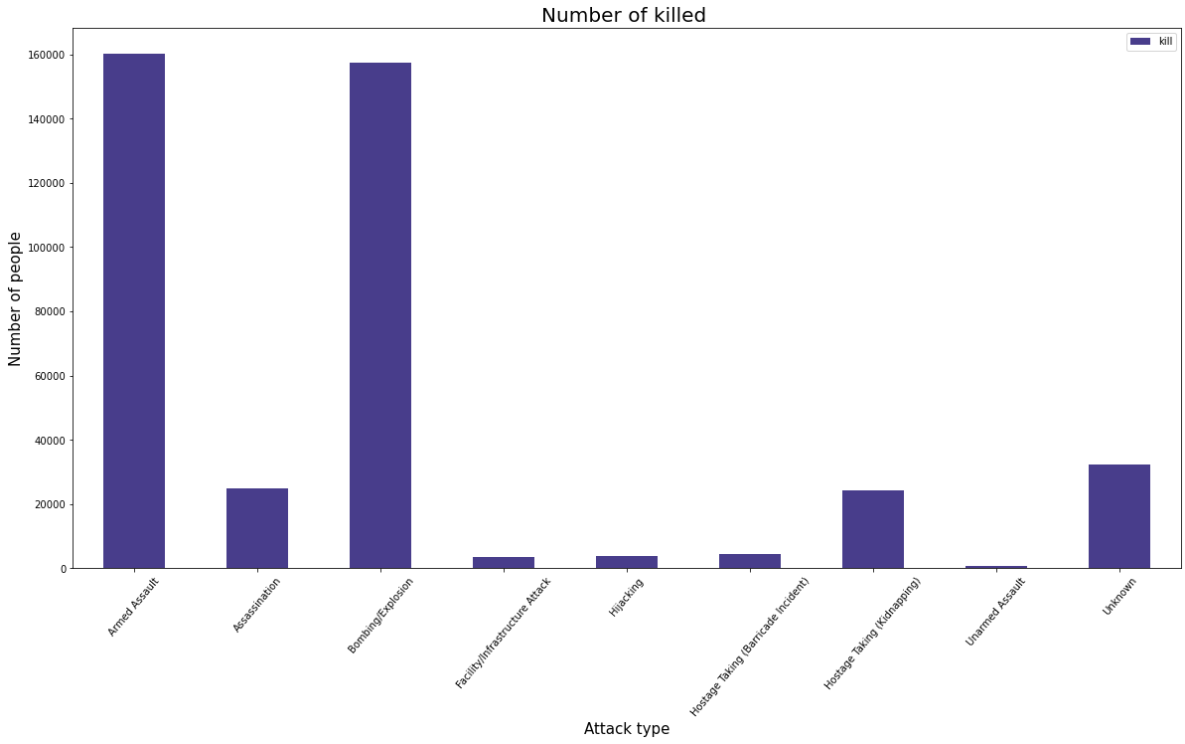
In [23]:

```
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()
```



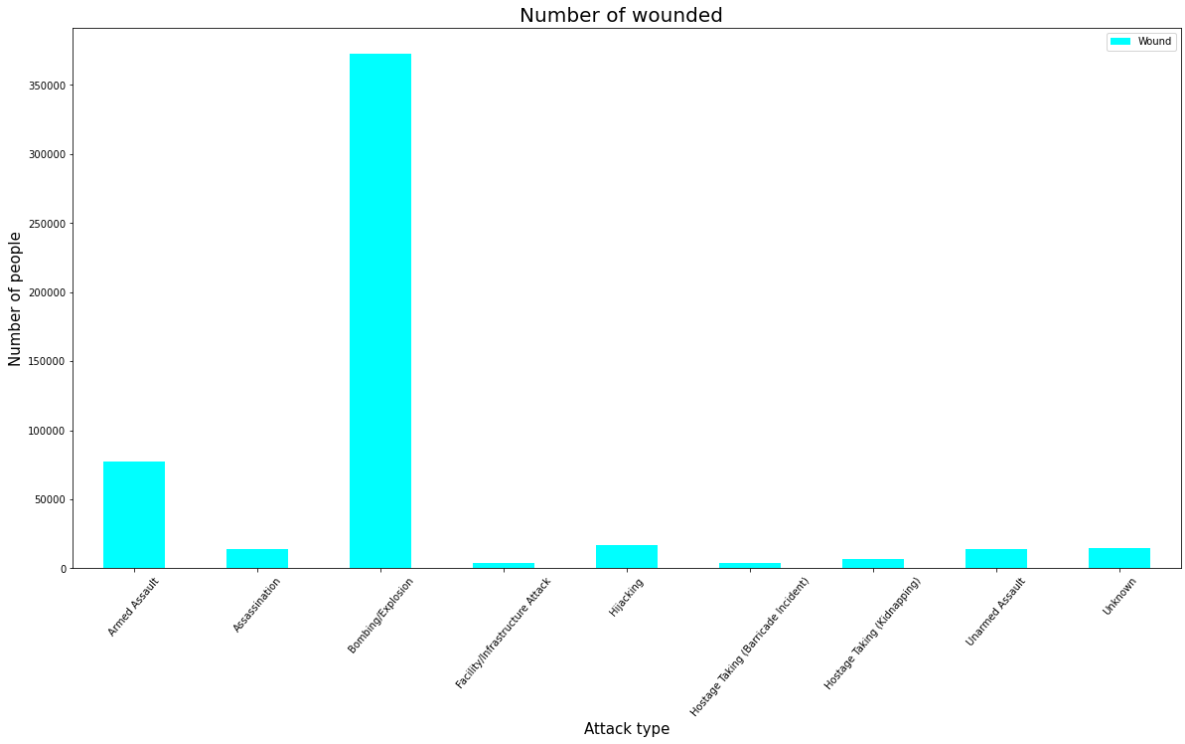
In [24]:

```
data[['Attacktype','kill']].groupby(["Attacktype"],axis=0).sum().plot(kind='bar',figsize=(2
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()
```



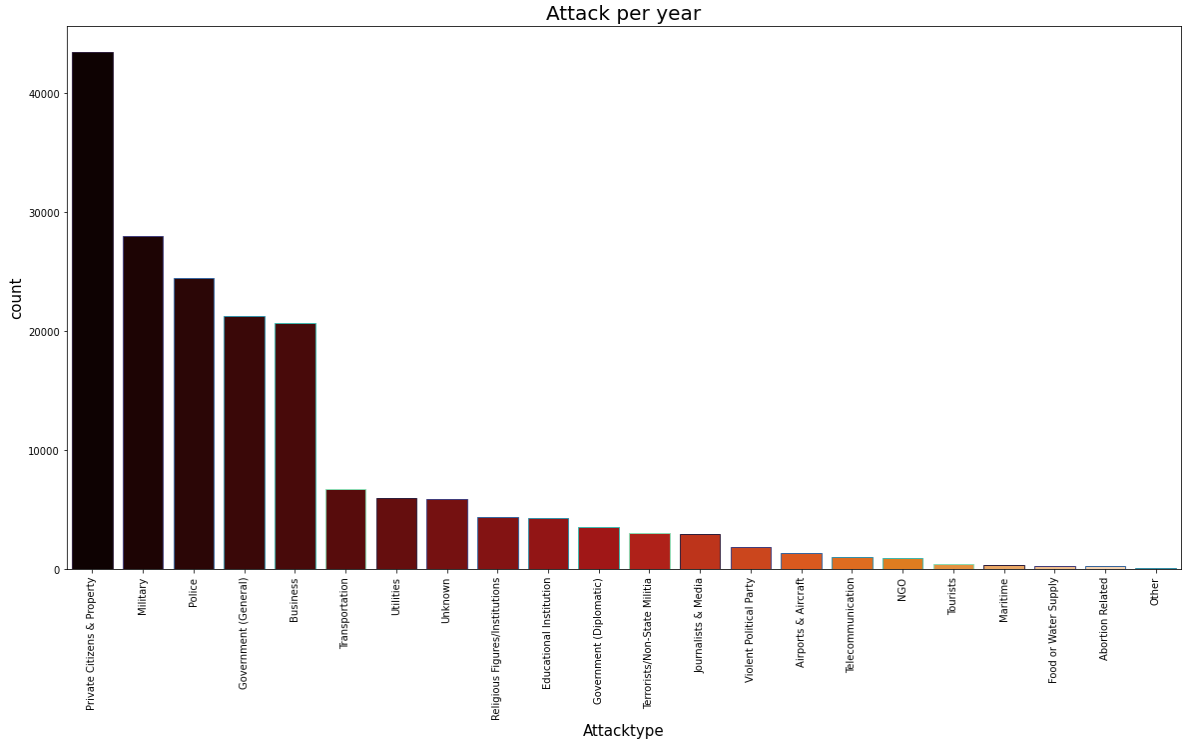
In [25]:

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



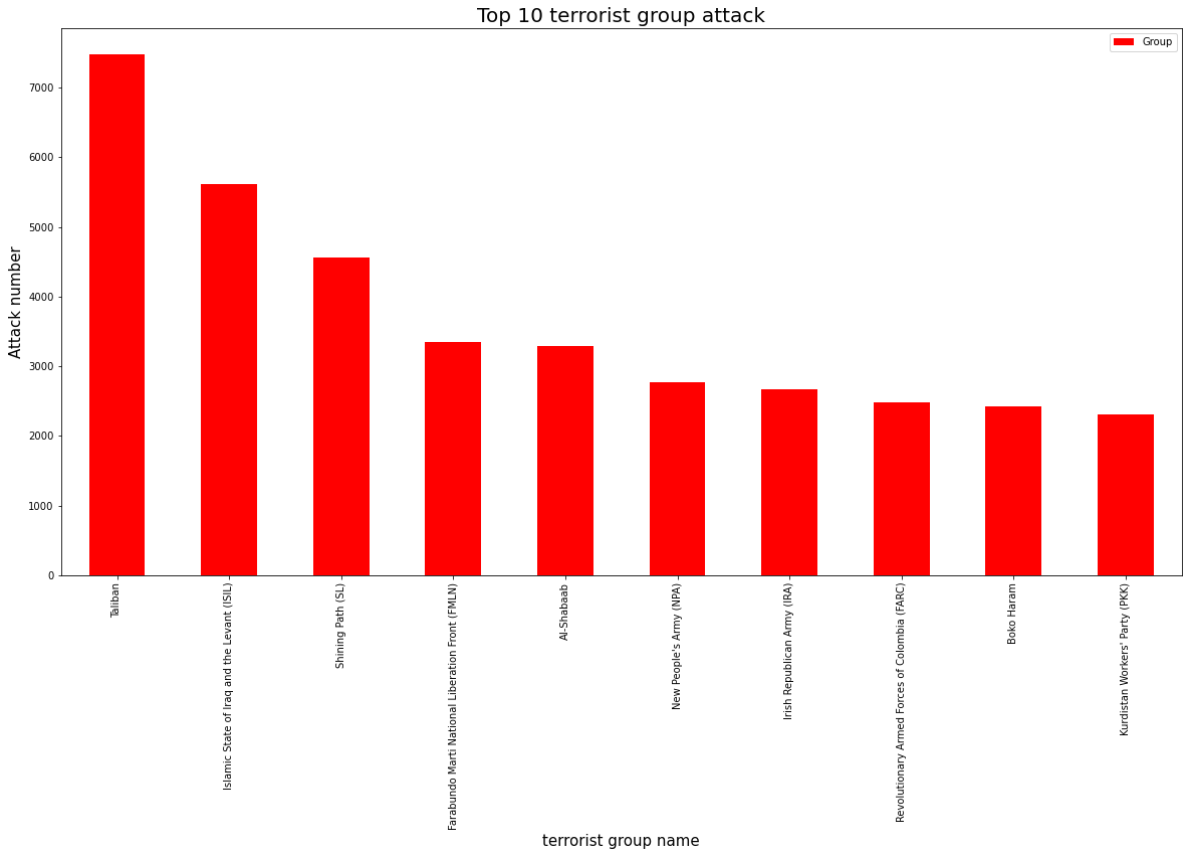
In [26]:

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



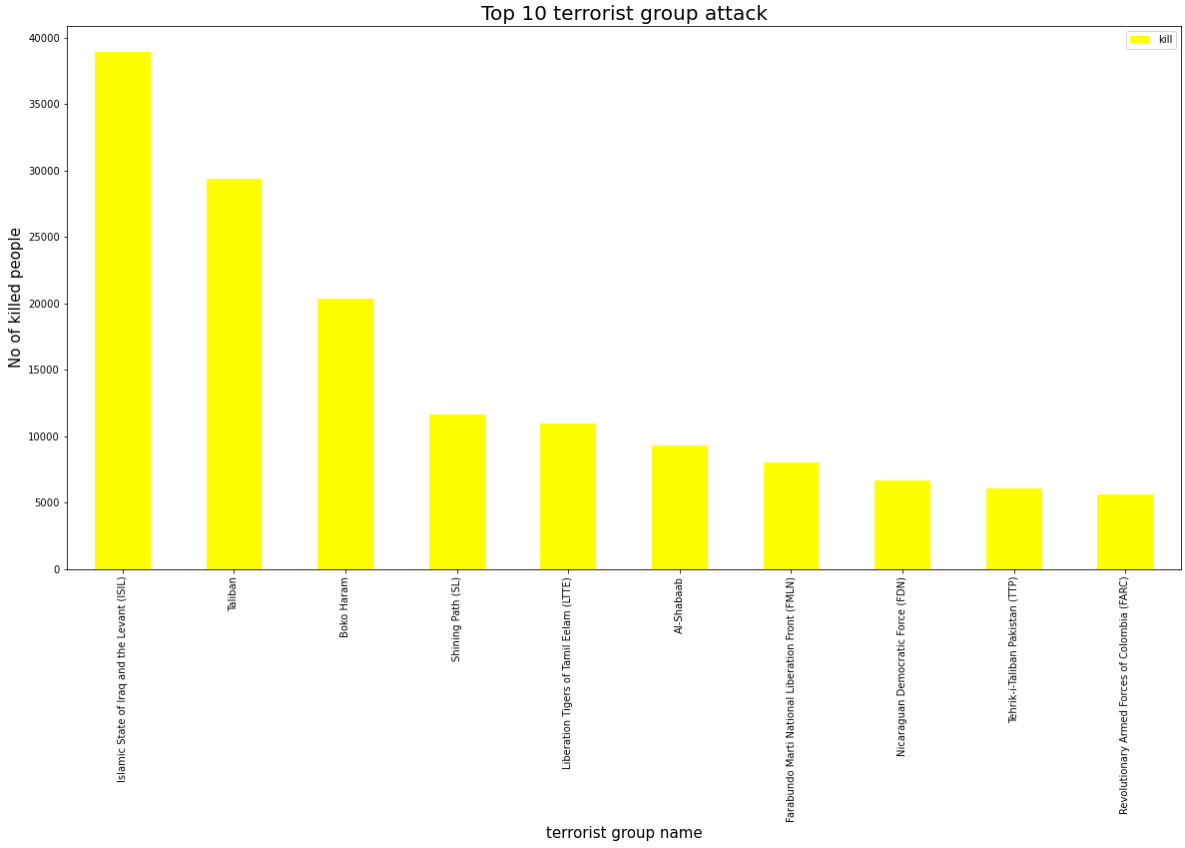
In [27]:

```
data['Group'].value_counts().to_frame().drop('Unknown').head(10).plot(kind='bar',color='red
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 [28]:

```
data[['Group','kill']].groupby(['Group'],axis=0).sum().drop('Unknown').sort_values('kill',a
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 [29]:

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

Out[29]:

	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 [30]:

```
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 [31]:

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

Out[31]:

Attacktype	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	Hos Ta (Barri Incic
kill	160297.0	24920.0	157321.0	3642.0	3718.0	44

In [32]:

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

Out[32]:

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

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