The Sparks Foundation

Auther : Kushal Patel

Data science and Business Analytics intern

Terrorism Data Analysis Project 1

Task : Exploratory Data Analysis

```python
In [ ]:  import pandas as pd
         import numpy as np

         from matplotlib import pyplot as plt
         import seaborn as sns

         import warnings
         warnings.filterwarnings("ignore")
```

Reading csv file

```python
In [2]:  df = pd.read_csv(r"C:\Users\Asus\Downloads\GRIP_data\Grip_data_new.csv")
         pd.options.display.max_columns = None
```

Data Manipulation and Cleaning

```python
In [3]:  df.head()
```

Out[3]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	r
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	
1	1.970000e+11	1970	0	0	NaN	0	NaN	130	Mexico	
2	1.970010e+11	1970	1	0	NaN	0	NaN	160	Philippines	
3	1.970010e+11	1970	1	0	NaN	0	NaN	78	Greece	
4	1.970010e+11	1970	1	0	NaN	0	NaN	101	Japan	

In [4]: `df.tail()`

Out[4]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_
181686	2.017120e+11	2017	12	31	NaN	0	NaN	182	Soma
181687	2.017120e+11	2017	12	31	NaN	0	NaN	200	Sy
181688	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippir
181689	2.017120e+11	2017	12	31	NaN	0	NaN	92	In
181690	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippir
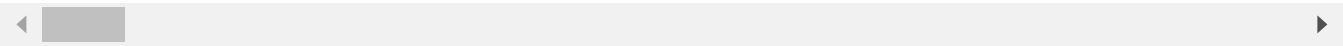
In [5]: `df.columns.values`

```
Out[5]:  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)
```

```python
In [6]:  data = df[['iyear','imonth','iday','country_txt','region_txt','provstate','city','l
                'targtype1_txt','summary','suicide','targsubtype1_txt','natlty1_txt','gn
         data
```

Out[6]:

	iyear	imonth	iday	country_txt	region_txt	provstate	city	latitude	longit
0	1970	7	2	Dominican Republic	Central America & Caribbean	NaN	Santo Domingo	18.456792	-69.951
1	1970	0	0	Mexico	North America	Federal	Mexico city	19.371887	-99.086
2	1970	1	0	Philippines	Southeast Asia	Tarlac	Unknown	15.478598	120.599
3	1970	1	0	Greece	Western Europe	Attica	Athens	37.997490	23.762
4	1970	1	0	Japan	East Asia	Fukouka	Fukouka	33.580412	130.396
...	...	...	...	...	...	...	...	...	...
181686	2017	12	31	Somalia	Sub-Saharan Africa	Middle Shebelle	Ceelka Geelow	2.359673	45.385
181687	2017	12	31	Syria	Middle East & North Africa	Lattakia	Jableh	35.407278	35.942
181688	2017	12	31	Philippines	Southeast Asia	Maguindanao	Kubentog	6.900742	124.437
181689	2017	12	31	India	South Asia	Manipur	Imphal	24.798346	93.940
181690	2017	12	31	Philippines	Southeast Asia	Maguindanao	Cotabato City	7.209594	124.241

181691 rows × 21 columns

◀ ▬▬▬▬▬▬▬▬▬▬▬▬ ▶

In [7]: ```data.shape```

Out[7]: (181691, 21)

In [8]: ```data.isnull().sum()```

Out[8]:
```
iyear                 0
imonth                0
iday                  0
country_txt           0
region_txt            0
provstate           421
city                434
latitude           4556
longitude          4557
attacktype1_txt       0
nkill             10313
nwound            16311
targtype1_txt         0
summary           66129
suicide               0
targsubtype1_txt  10373
natlty1_txt        1559
gname                 0
weaptype1_txt         0
weapsubtype1_txt  20768
dbsource              0
dtype: int64
```

In [7]:
```python
df["region_txt"].nunique()
```

Out[7]:
12

In [10]:
```python
df.country_txt.nunique()
```

Out[10]:
205

Here It should be noted that there are only 195 contries that have been registered in the United Nations. So there might by some error in the country data.

In [11]:
```python
data['nwound']=data['nwound'].fillna(0)
data['nkill']=data['nkill'].fillna(0)
```

In [12]:
```python
casualities = data.nkill + data.nwound
casualities
```

Out[12]:
```
0          1.0
1          0.0
2          1.0
3          0.0
4          0.0
          ...
181686     3.0
181687     9.0
181688     0.0
181689     0.0
181690     0.0
Length: 181691, dtype: float64
```

In [13]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 21 columns):
 #   Column          Non-Null Count    Dtype
---  ------          --------------    -----
 0   iyear           181691 non-null   int64
 1   imonth          181691 non-null   int64
 2   iday            181691 non-null   int64
 3   country_txt     181691 non-null   object
 4   region_txt      181691 non-null   object
 5   provstate       181270 non-null   object
 6   city            181257 non-null   object
 7   latitude        177135 non-null   float64
 8   longitude       177134 non-null   float64
 9   attacktype1_txt 181691 non-null   object
 10  nkill           181691 non-null   float64
 11  nwound          181691 non-null   float64
 12  targtype1_txt   181691 non-null   object
 13  summary         115562 non-null   object
 14  suicide         181691 non-null   int64
 15  targsubtype1_txt 171318 non-null  object
 16  natlty1_txt     180132 non-null   object
 17  gname           181691 non-null   object
 18  weaptype1_txt   181691 non-null   object
 19  weapsubtype1_txt 160923 non-null  object
 20  dbsource        181691 non-null   object
dtypes: float64(4), int64(4), object(13)
memory usage: 29.1+ MB
```

In [14]: `data.describe()`

Out[14]:

|	iyear	imonth	iday	latitude	longitude	nkill
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
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

Data Visualizations using various charts and Graphs

In [15]:
```python
year=data['iyear'].unique()
years_count=data['iyear'].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('Attack Year',fontsize=20)
plt.ylabel('Numbere of attacks each year',fontsize=20)
```

```python
plt.title('Attacks in years',fontsize=20)
plt.show()
```


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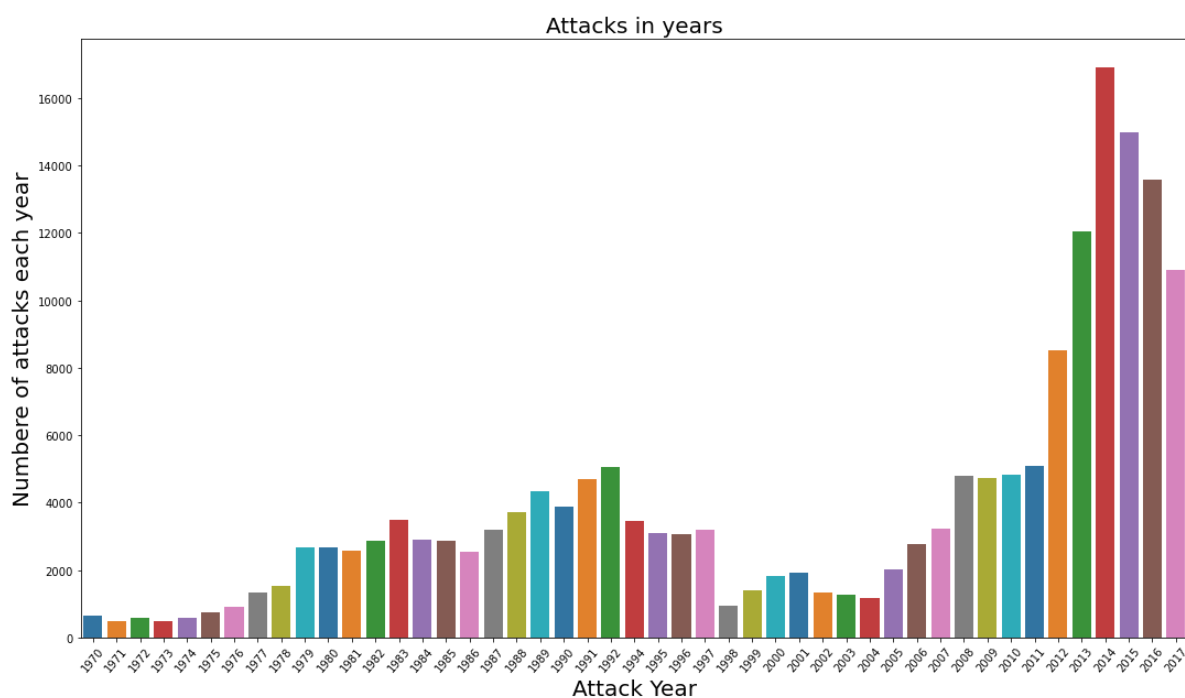
```python
pd.crosstab(data.iyear,data.region_txt).plot(kind='area',stacked=False
                                              ,figsize=(20,10))
plt.title('Terrrist Activities By Region In Each Year',fontsize=25)
plt.xlabel('Year',fontsize=20)
plt.ylabel('Number of Attacks',fontsize=20)
plt.show()
```

In [16]:

Terrrist Activities By Region In Each Year
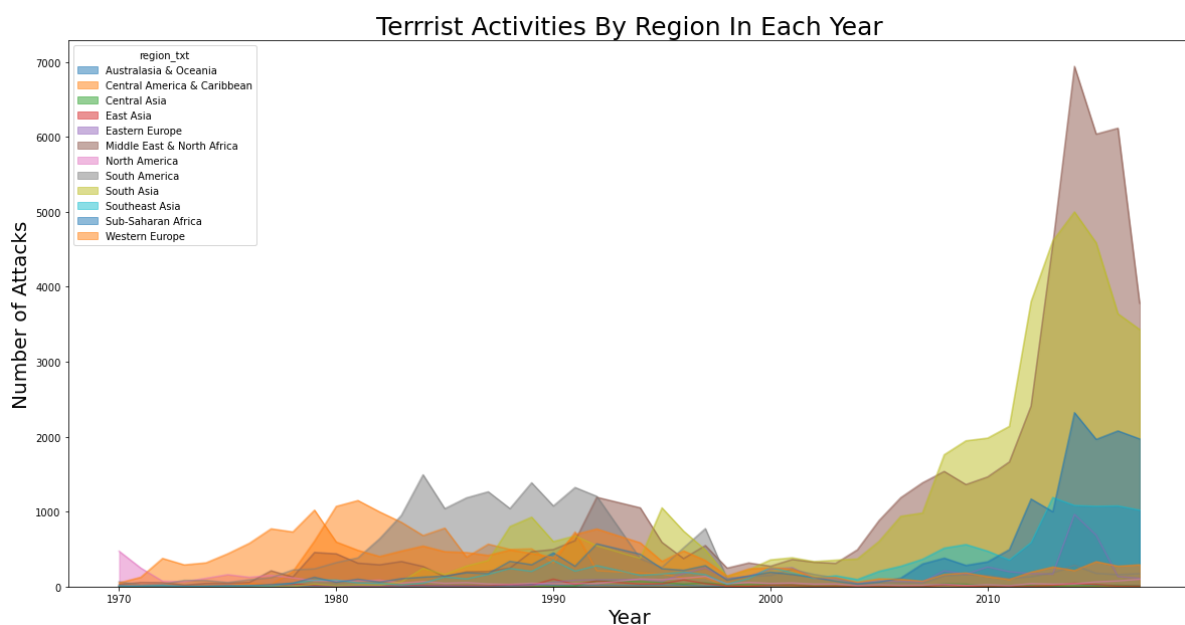
In [17]:
```python
attack= data.country_txt.value_counts()[:10]
attack
```

```
Out[17]:  Iraq                24636
          Pakistan            14368
          Afghanistan         12731
          India               11960
          Colombia             8306
          Philippines          6908
          Peru                 6096
          El Salvador          5320
          United Kingdom       5235
          Turkey               4292
          Name: country_txt, dtype: int64
```

```
In [18]: data.gname.value_counts()[1:10]
```

```
Out[18]:  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: gname, dtype: int64
```

```
In [19]: plt.subplots(figsize=(20,10))
         sns.barplot(data['country_txt'].value_counts()[:10].index,data['country_txt'].valu
         plt.title('Terrrist Activities By Region In Each Year',fontsize=25)
         plt.xlabel('Year',fontsize=20)
         plt.ylabel('Number of Attacks',fontsize=20)
         plt.show()
```
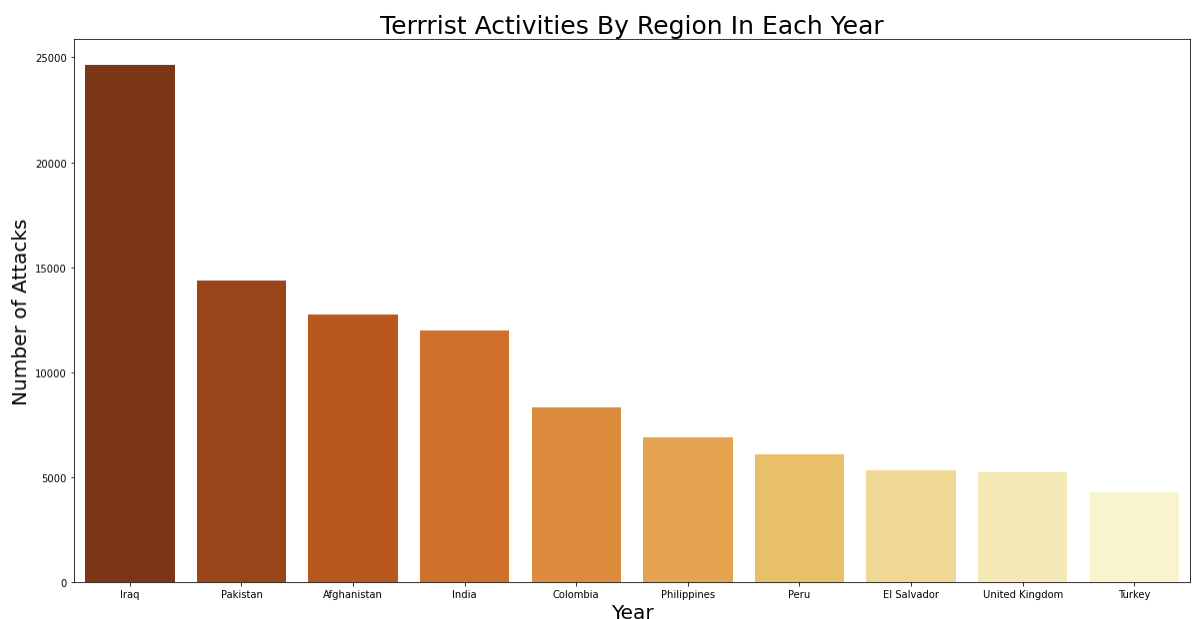


```
In [ ]: df1= data[['iyear','nkill']].groupby(['iyear']).sum()
        fig,ax4 = plt.subplots(figsize=(20,10))
        df.plot(kind='bar',alpha=0.7,ax=ax4)
        plt.xticks(rotation=50)
        plt.title('People Died during attack',fontsize=25)
        plt.xlabel('Number of killed people',fontsize=20)
        plt.ylabel('Year',fontsize=20)
        top_side = ax4.spines["top"]
        top_side.set_visible(False)
        right_side = ax4.spines['right']
        right_side.set_visible(False)
```

In [8]:
```python
data.city.value_counts().to_frame().sort_values('city',axis = 0,ascending=False).he
plt.xlabel('City',fontsize=15)
plt.xticks(rotation=50)
plt.ylabel('Number of Attack',fontsize=15)
plt.title('Top 10 most Effected Cities',fontsize=20)
plt.show()
```

In [8]:
```python
data.city.value_counts().to_frame().sort_values('city',axis = 0,ascending=False).he
plt.xlabel('City',fontsize=15)
plt.xticks(rotation=50)
plt.ylabel('Number of Attack',fontsize=15)
plt.title('Top 10 most Effected Cities',fontsize=20)
plt.show()
```
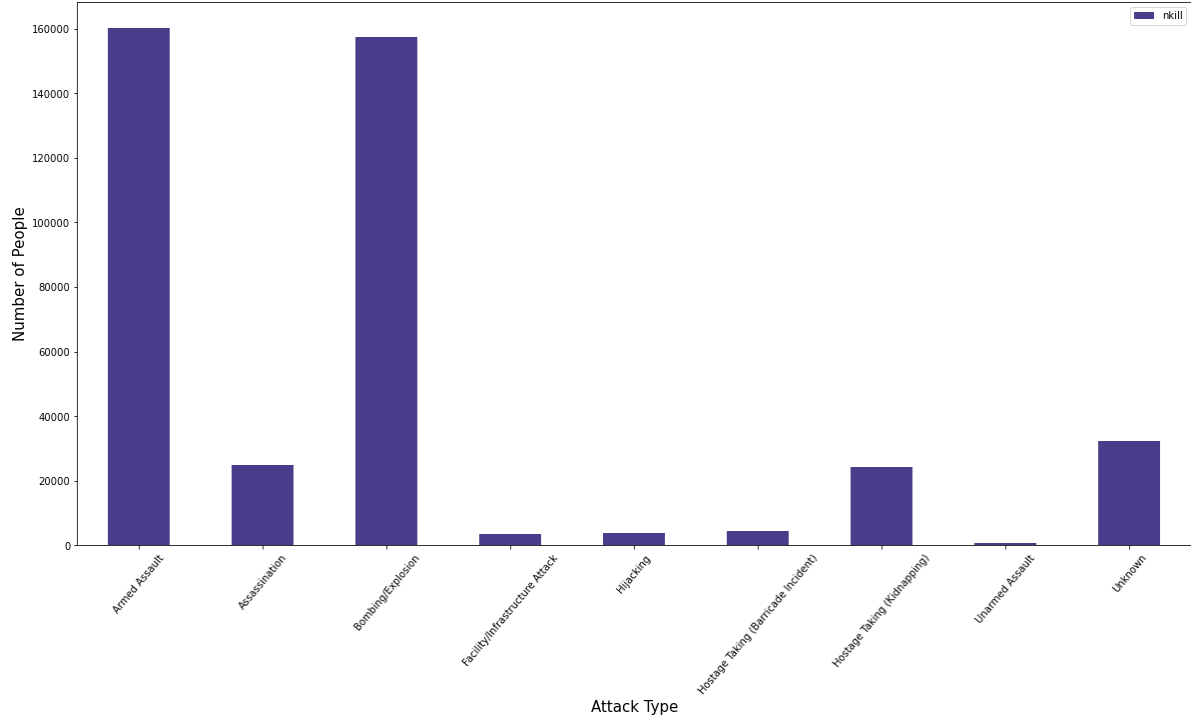
Top 10 most Effected Cities

City
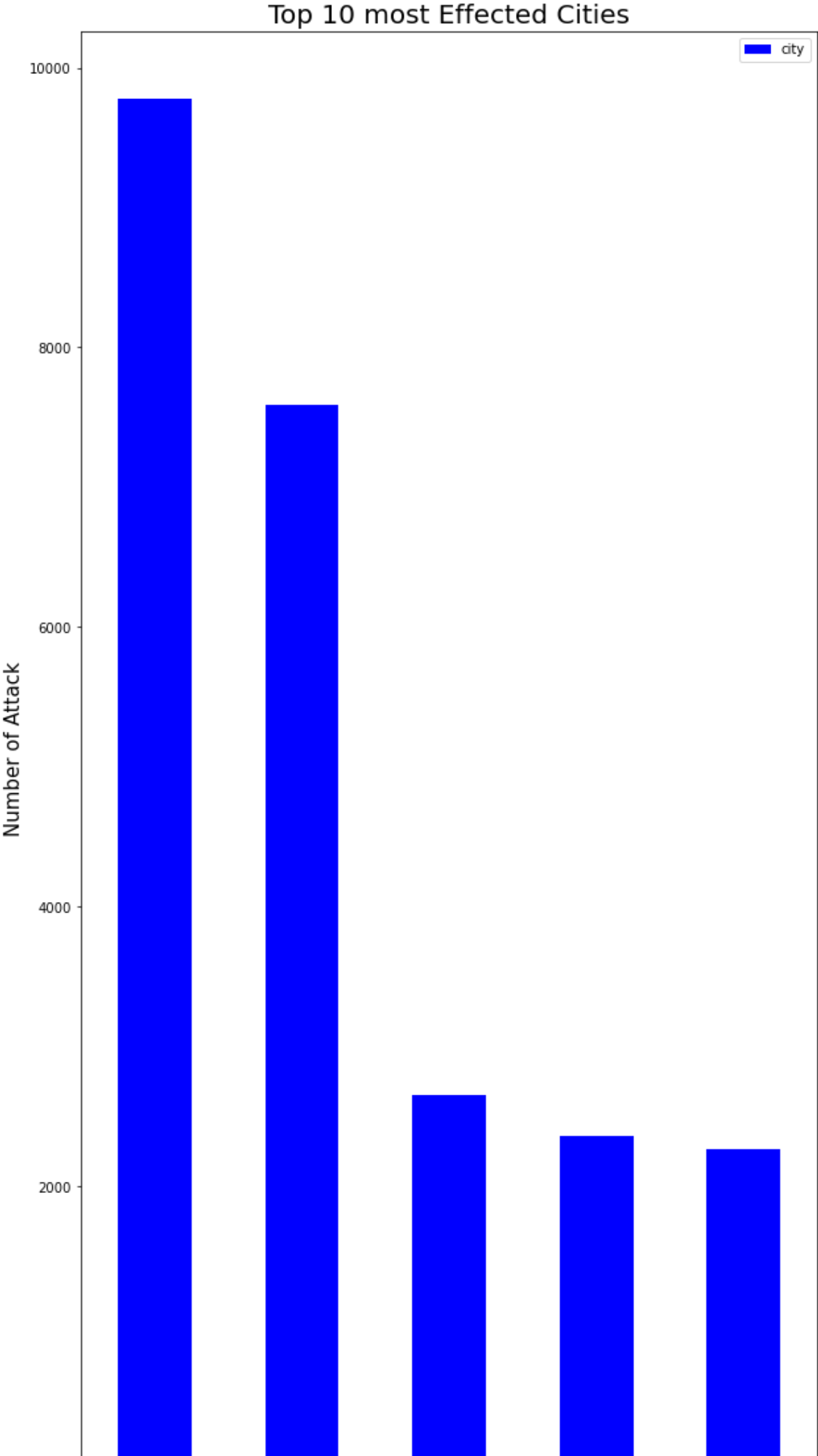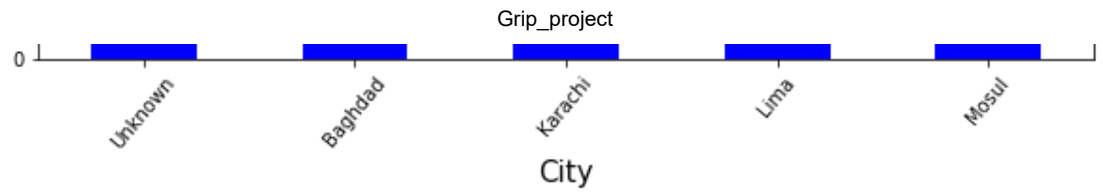
```
In [9]:  data.attacktype1_txt.value_counts().plot(kind='bar',figsize=(20,10),color='yellow'
         plt.xlabel('Attack type',fontsize=15)
         plt.xticks(rotation=50)
         plt.ylabel('Number of atttack',fontsize=15)
         plt.title('Type of attack',fontsize=15)
         plt.show()
```
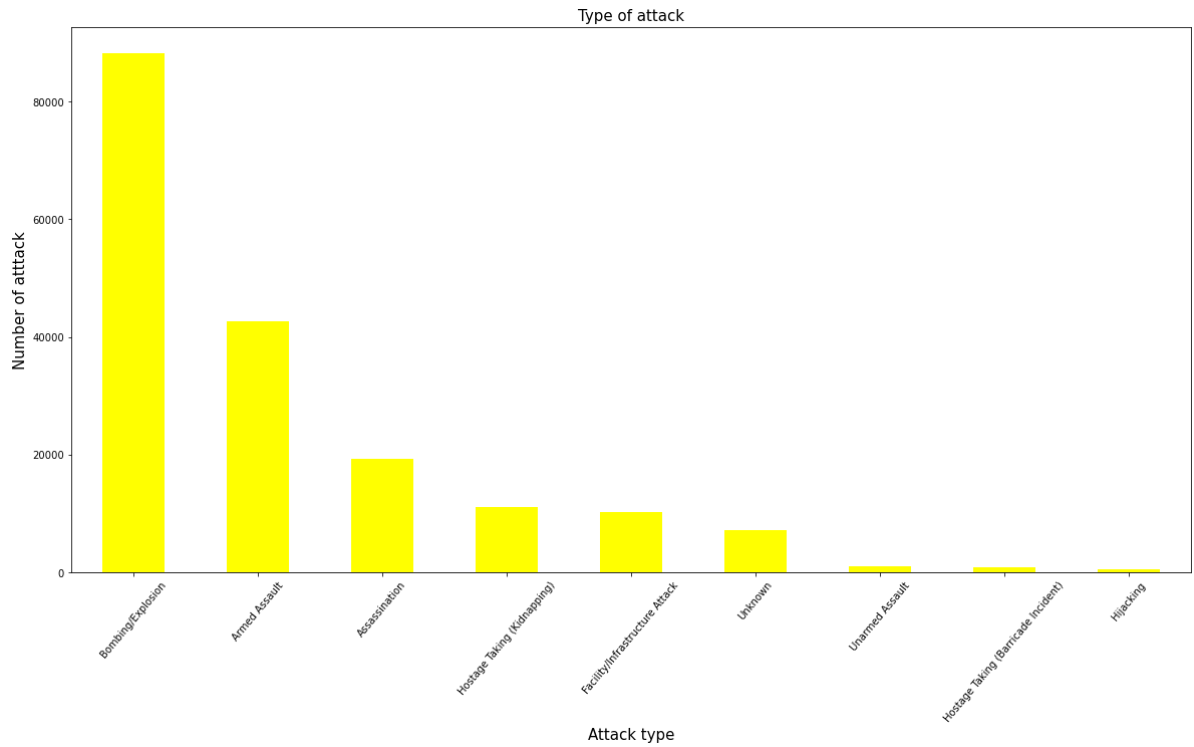


```
In [13]: data[['attacktype1_txt','nkill']].groupby(['attacktype1_txt'],axis=0).sum().plot(k:
         plt.xlabel('Attack Type',fontsize=15)
         plt.xticks(rotation=50)
         plt.ylabel('Number of People',fontsize=15)
         plt.title('Number of Killed',fontsize=20)
         plt.show()
```

Number of Killed


```
In [ ]:  data[['attacktype1_txt','nwound']].groupby(['attacktype1_txt'],axis=0).plot(kind='l
         plt.xlabel('Attack Type',fontsize=15)
         plt.xticks(rotation=50)
         plt.ylabel('Number of People Wounded',fontsize=15)
         plt.title('Number of Wounded',fontsize=20)
         plt.show()
```

```
In [7]:  plt.subplots(figsize=(20,10))
         sns.countplot(data['targtype1_txt'].value_counts().index,palette='gist_heat')
         plt.title('Attacks per year',fontsize=25)
         plt.xlabel('Attacktype',fontsize=20)
         plt.ylabel('Count',fontsize=20)
         plt.show()
```
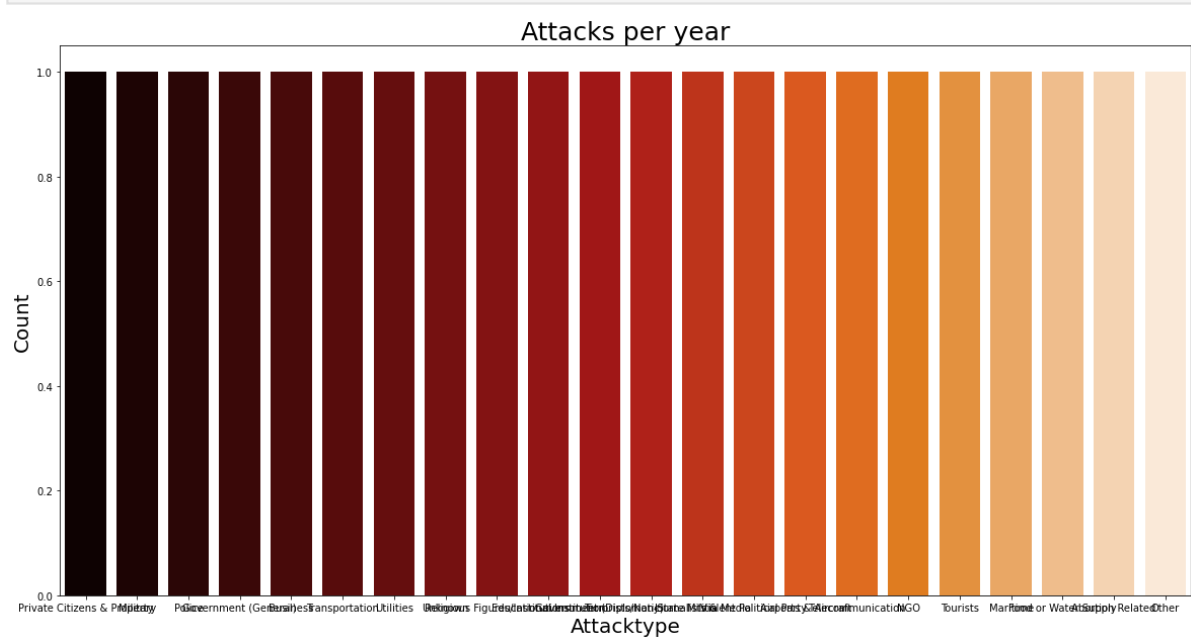

Country with the most Attacks:Iraq

```
In [8]:  data.gname.value_counts().to_frame().drop('Unknown').head(10).plot(kind='bar',colo
         plt.xlabel('terrorist group name',fontsize=15)
```

```
plt.ylabel('Number of attacks',fontsize=15)
plt.title('Top 10 Terror Attacks',fontsize=20)
plt.show()
```
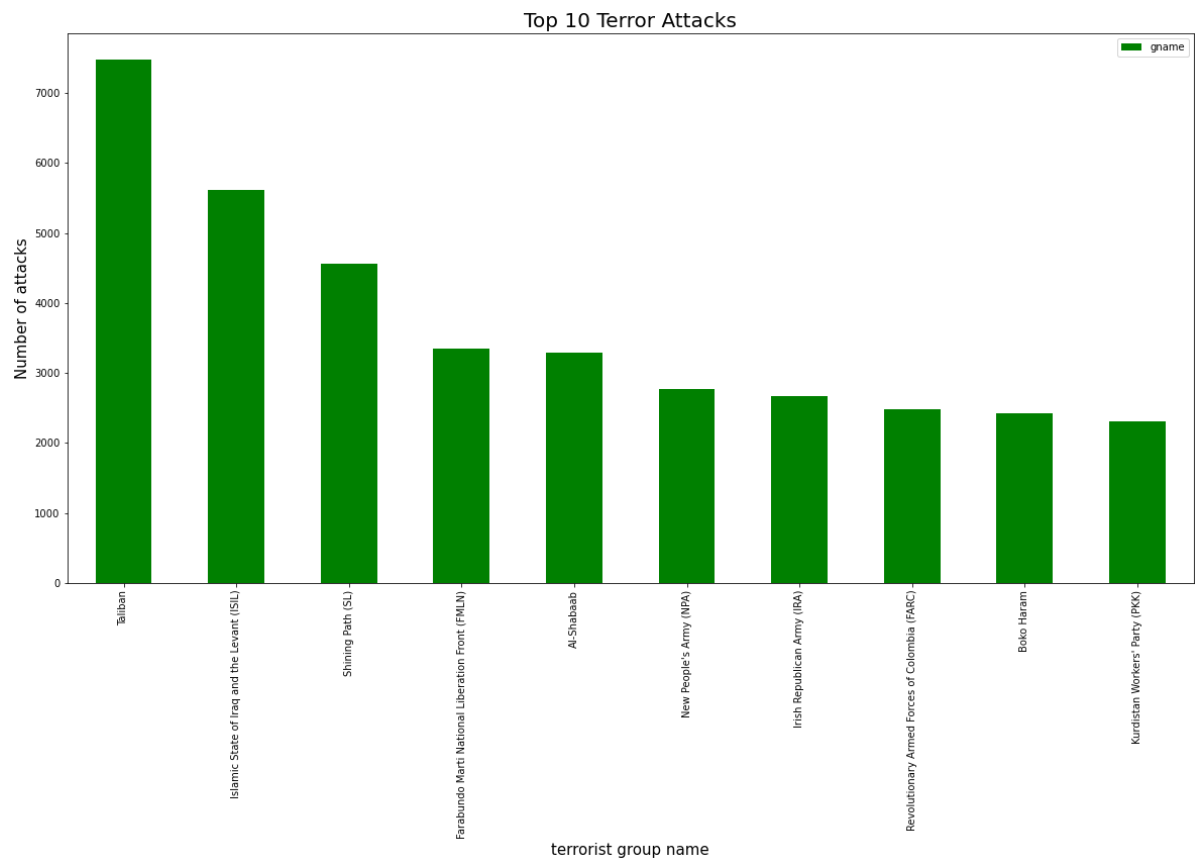


```
In [9]:  df2=data[['gname','country_txt','nkill']]
         df=df.groupby(['gname','country_txt'],axis=0).sum().sort_values('nkill',ascending=
         df
```

gname	country_txt	eventid	iyear	imonth	iday	extended	country	region	latitude
Islamic State of Iraq and the Levant (ISIL)	Iraq	9.668001e+14	9667687	31408	76074	325	455715	47970	165025.991200
Taliban	Afghanistan	1.494438e+15	14943888	49482	117122	683	29692	44538	250969.882708
Boko Haram	Nigeria	4.203335e+14	4203203	13202	33374	143	306789	22957	23771.586275
Shining Path (SL)	Peru	9.025177e+14	9024885	29154	70421	32	722019	13623	-51556.667335
Liberation Tigers of Tamil Eelam (LTTE)	Sri Lanka	3.187303e+14	3187199	10154	24656	16	296856	9576	12841.009096
Al-Shabaab	Somalia	5.775416e+14	5775226	18970	44999	191	521794	31537	7720.968002
Farabundo Marti National Liberation Front (FMLN)	El Salvador	6.613985e+14	6613775	20983	49414	14	203130	6660	40761.351577
Islamic State of Iraq and the Levant (ISIL)	Syria	1.197167e+14	1197127	3918	9938	105	118800	5940	20802.549564
Nicaraguan Democratic Force (FDN)	Nicaragua	1.758443e+14	1758386	5730	12427	2	128470	1772	8697.532230
Tehrik-i-Taliban Pakistan (TTP)	Pakistan	2.686405e+14	2686321	8368	20704	126	204255	8010	43572.555983
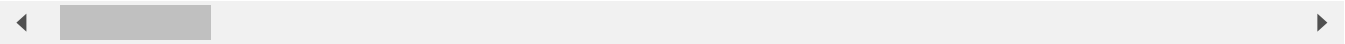
```
In [10]:  kill=data.loc[:,'nkill']
          print('Number of people killed by terror attack: ',int(sum(kill.dropna())))
```

Number of people killed by terror attack: 411868

```
In [11]:  typekill=data.pivot_table(columns='attacktype1_txt',values='nkill',aggfunc='sum')
          typekill
```

Out[11]:

attacktype1_txt	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	H (Ba In
nkill	160297.0	24920.0	157321.0	3642.0	3718.0	

In [12]:
```python
countrykill=data.pivot_table(columns='country_txt',values='nkill',aggfunc='sum')
countrykill
```

Out[12]:

country_txt	Afghanistan	Albania	Algeria	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	Aus
nkill	39384.0	42.0	11066.0	0.0	3043.0	0.0	490.0	37.0	

Conclusion and Results:

City woth the most attacks: Baghdad

Region with the most attacks: Middle East & North Africa

Year with the most attacks: Middle East & North Africa

Month with the most attacks: 5

Most Dangerous Attacking Group : Taliban

Most Frequent Used Attack Types : Bombing/Explosions

Thank You!