

Task 4: Exploratory analysis on the dataset "GLOBAL TERRORISM"

The Sparks Foundation-Graduate Rotational Internship Programme(Batch-July(2021))

Data Science and Business Analytics Internship

Submitted By: Chandhini V

Importing all the libraries necessary for analysis

```
In [87]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

Importing and reading the dataset

```
In [88]: data = pd.read_csv("C:\Users\USERA\Desktop\globalterrorismdb_0718dist.csv",encoding='latin1')
print("Data has been successfully imported")
data.head()
```

Data has been successfully imported

```
Out[88]:
```

	eventid	year	month	day	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN
1	1.970000e+11	1970	1	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN
2	1.970000e+11	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN
3	1.970000e+11	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN
4	1.970000e+11	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN

5 rows × 15 columns

Cleaning the dataset

```
In [89]: data.columns.values
```

```
Out[89]: array(['eventid', 'year', 'month', 'day', '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',
       'targettype1', 'targettype1_txt', 'targetsubtype1', 'targetsubtype1_txt',
       'corp1', 'target1', 'natly1', 'natly1_txt', 'targettype2',
       'targettype2_txt', 'targetsubtype2', 'targetsubtype2_txt', 'corp2',
       'target2', 'natly2', 'natly2_txt', 'targettype3', 'targettype3_txt',
       'targetsubtype3', 'targetsubtype3_txt', 'corp3', 'target3', 'natly3',
       'natly3_txt', 'gname', 'gsurname', 'gname2', 'gsurname2',
       'gname3', 'gsurname3', 'motive', 'uncertain1', 'uncertain2',
       'uncertain3', 'individual', 'nperps', 'npercpap', 'claimed',
       'claimmode', 'claimmode_txt', 'claim2', 'claimmode2',
       'claimmode2_txt', 'claim3', 'claimmode3', 'claimmode3_txt',
       'complain', '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', 'nkilliter',
       'nound', 'nwoundus', 'nwoundte', 'property', 'propexten',
       'propexten_txt', 'propvalue', 'propcommen', 'ishostkid',
       'nhostkid', 'nhostkidus', 'nhours', 'ndays', 'divert',
       'kididjour', 'ransom', 'ransomamt', 'ransomamtus',
       'ransompaid', 'ransompaidus', 'ransomote', 'hostkidoutcome',
       'hostkidoutcome_txt', 'freleased', 'addnotes', 'scite1', 'scite2',
       'scite3', 'dsource', 'INF_LOG', 'INF_IDEO', 'INF_MISC', 'INF_ANY',
       'related'], dtype=object)
```

```
In [90]: data.rename(columns={'year':'Year','month':'Month','day':'day','gname':'Group','country_txt':'Country',
       'region_txt':'Region','provstate':'State','city':'City','latitude':'latitude',
       'longitude':'longitude','summary':'summary','attacktype1_txt':'Attacktype','targettype1_txt':'Targettype',
       'weaptype1_txt':'Weapon','nkill':'kill',
       'nwound':'Wound'},inplace=True)
```

```
In [91]: data = data[['Year','Month','day','Country','State','Region','City','latitude','longitude','Attacktype','kill',
       'Wound',
       'target1','summary','Group','Targettype','Weapon','motive']]
```

```
In [92]: data.shape
```

Out[92]: (181691, 18)

```
In [93]: data.isnull().sum()
```

```
Out[93]: 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        536
summary        66129
Group          0
Targettype    0
Weapon         0
motive        131130
dtype: int64
```

```
In [94]: data['Wound'] = data['Wound'].fillna(0)
data['kill'] = data['kill'].fillna(0)
```

```
In [95]: data['Casualties'] = data['kill'] + data['Wound']
```

```
In [96]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 18 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       181257 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     11562 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 [97]: data.describe()
```

```
Out[97]:
```

	Year	Month	day	latitude	longitude	kill	Wound	Casualties
count	181691.000000	181691.000000	177135.000000	1.771340e+05	1.816910000000	181691.000000	181691.000000	181691.000000
mean	2002.038997	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	11.510446	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	1793667e+02	1570.000000	8191.000000	9574.000000

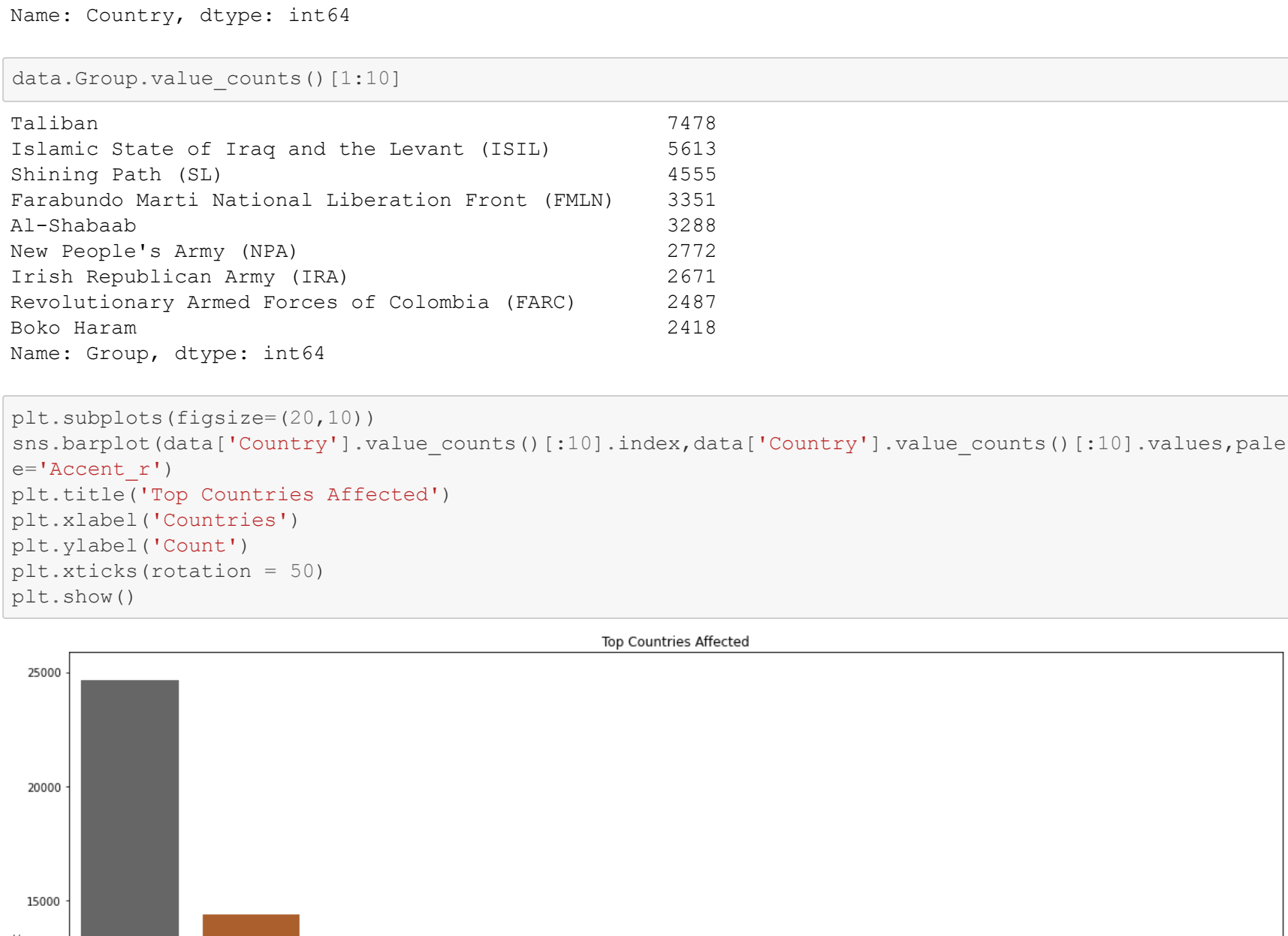
Observations

1. The data consists of terrorist activities ranging from the year : 1970 to 2017
2. Maximum number of people killed in an event : 1570
3. Maximum number of people wounded in an event : 8191
4. Maximum number of total casualties in an event : 9574

Visualizations of the data

1. Yearwise attacks

```
In [98]: 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 = 30)
plt.xlabel('Attacking Year',fontsize=20)
plt.ylabel('Number of Attacks per Year',fontsize=20)
plt.title('Attacks In Years',fontsize=30)
plt.show()
```



```
In [99]: 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 [100]: attack = data.Country.value_counts()[1:10]
```

```
Out[100]: Iraq          24636
Pakistan        14368
Afghanistan     12731
India           11960
Colombia        8306
Philippines     6908
Peru            6096
El Salvador     5325
United Kingdom  5235
Turkey          4292
Name: Country, dtype: int64
```

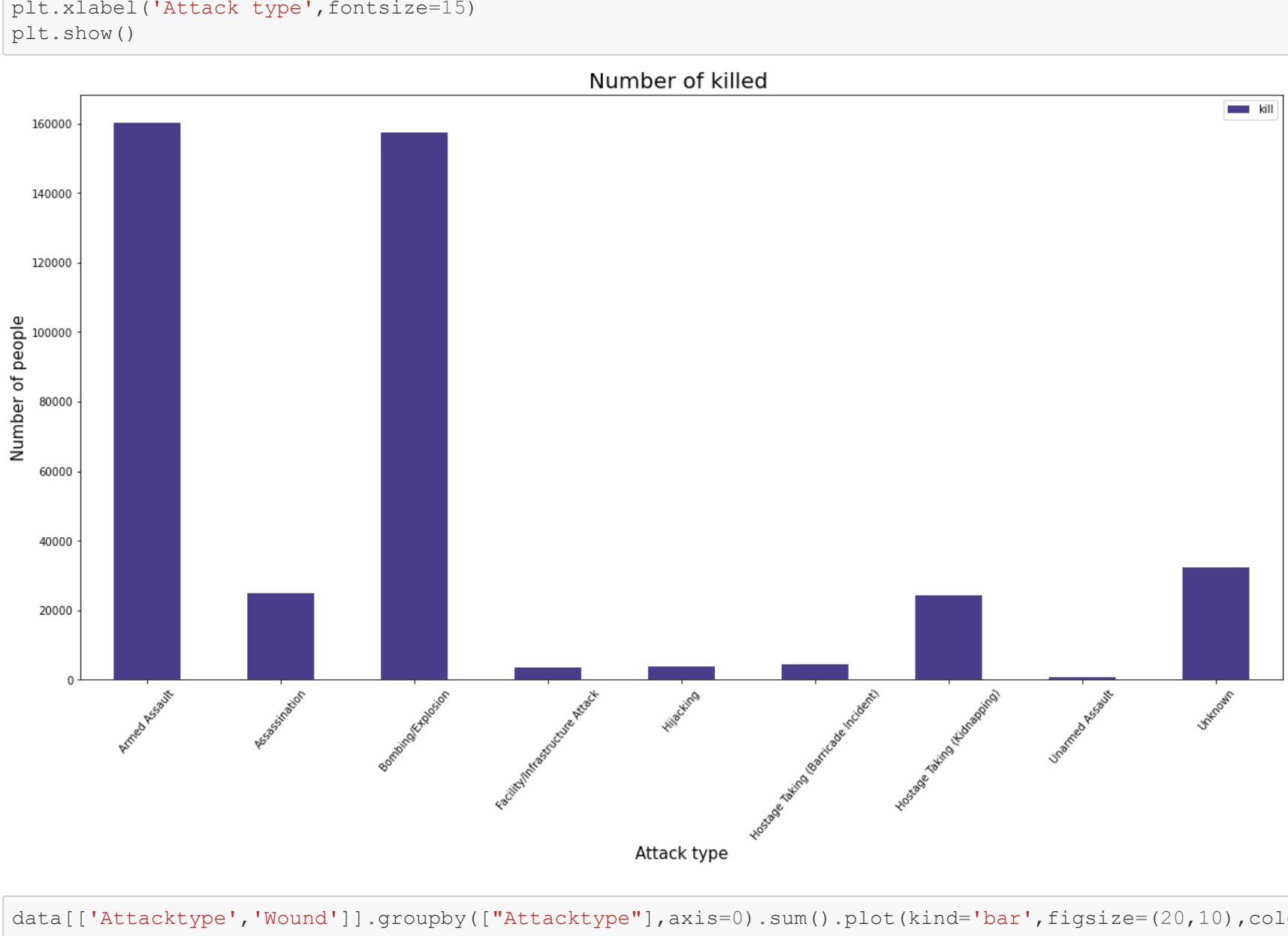
```
In [101]: data.Group.value_counts()[1:10]
```

```
Out[101]: Taliban          7478
Islamic State of Iraq and the Levant (ISIL)  5613
Shinjitah (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
```

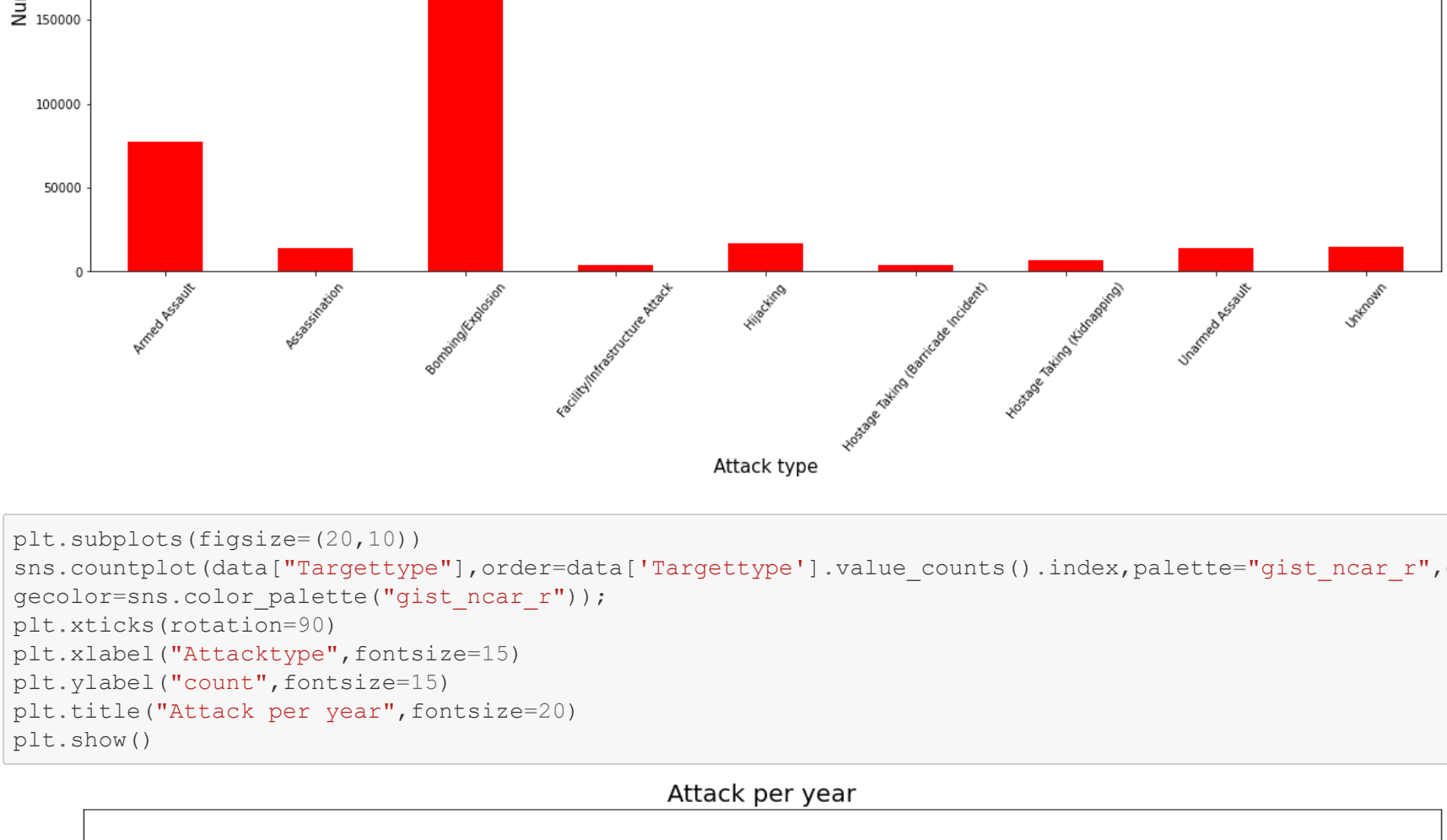
```
In [102]: plt.subplots(figsize=(20,10))
df = df.groupby(['Country']).value_counts()[1:10].index,data['Country'].value_counts()[1:10].values,palette='Accent_r')
plt.title('Top Countries Affected')
plt.xlabel('Countries')
plt.ylabel('Count')
plt.show()
```



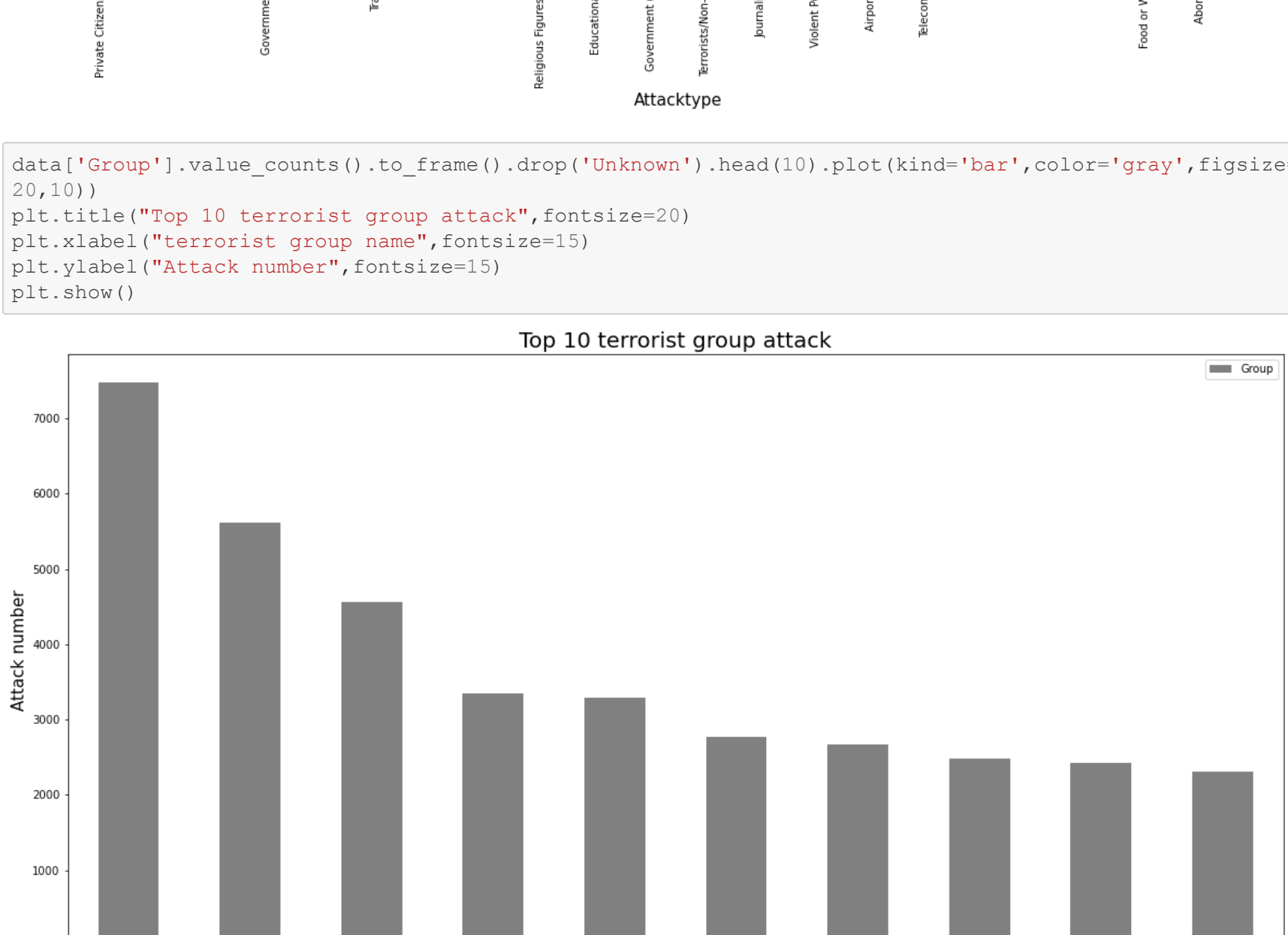
```
In [103]: df = data[['Year','kill']].groupby(['Year']).sum()
fig, ax4 = plt.subplots(figsize=(20,10))
df.plot(kind='bar',alpha=0.7,ax=ax4)
plt.xticks(rotation = 50)
plt.title('People Died Due To Attack',fontsize=25)
plt.ylabel('Number of killed people',fontsize=20)
plt.xlabel('Year',fontsize=20)
top_side = ax4.spines["top"]
top_side.set_visible("top")
right_side = ax4.spines["right"]
right_side.set_visible(False)
```



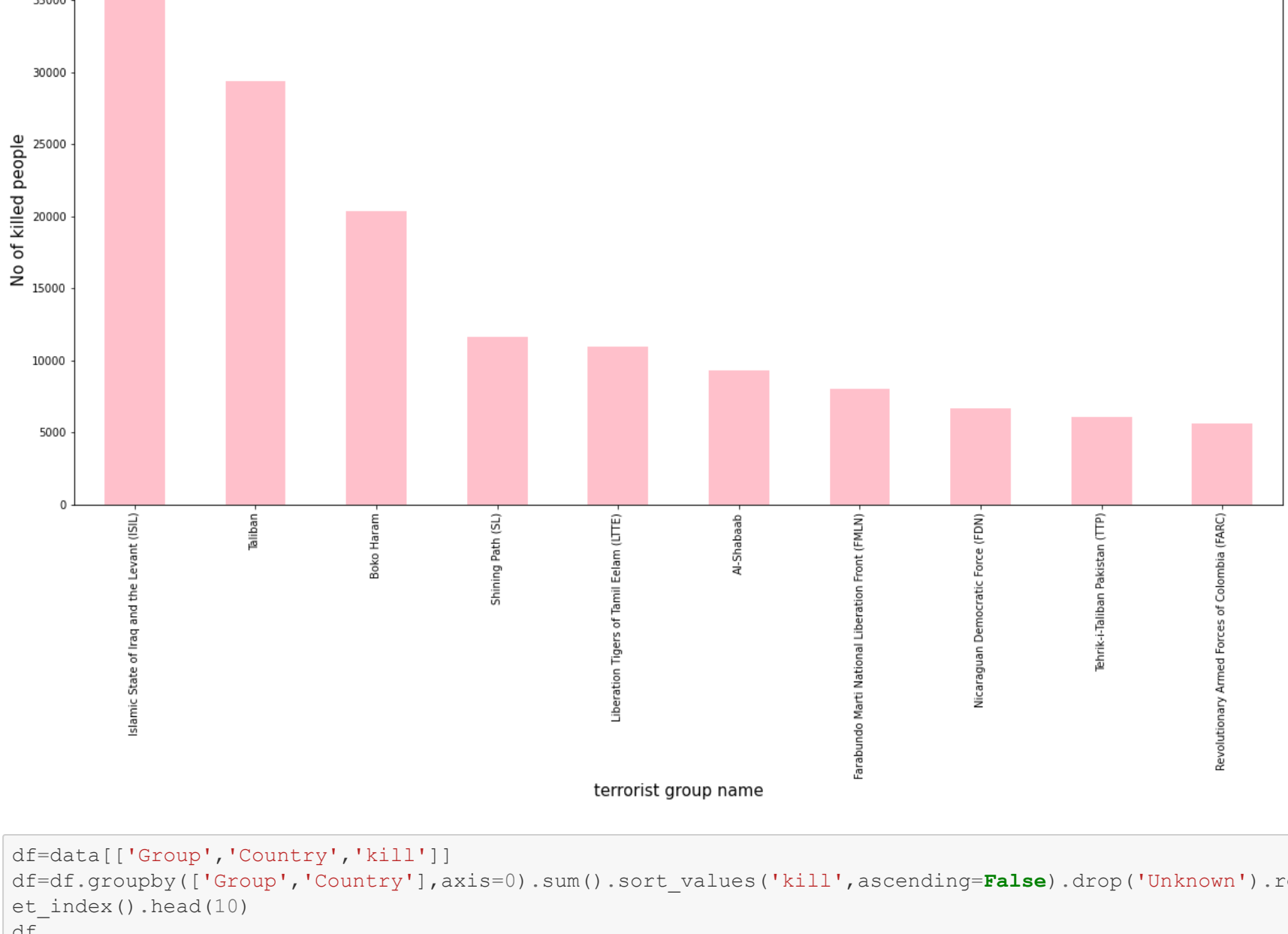
```
In [104]: data['City'].value_counts().to_frame().sort_values('City',axis=0,ascending=False).head(10).plot(kind='bar',figsize=(20,10),color='black')
plt.xlabel('City',fontsize=15)
plt.ylabel('Number of attack',fontsize=15)
plt.title('Top 10 most effected city',fontsize=20)
plt.show()
```



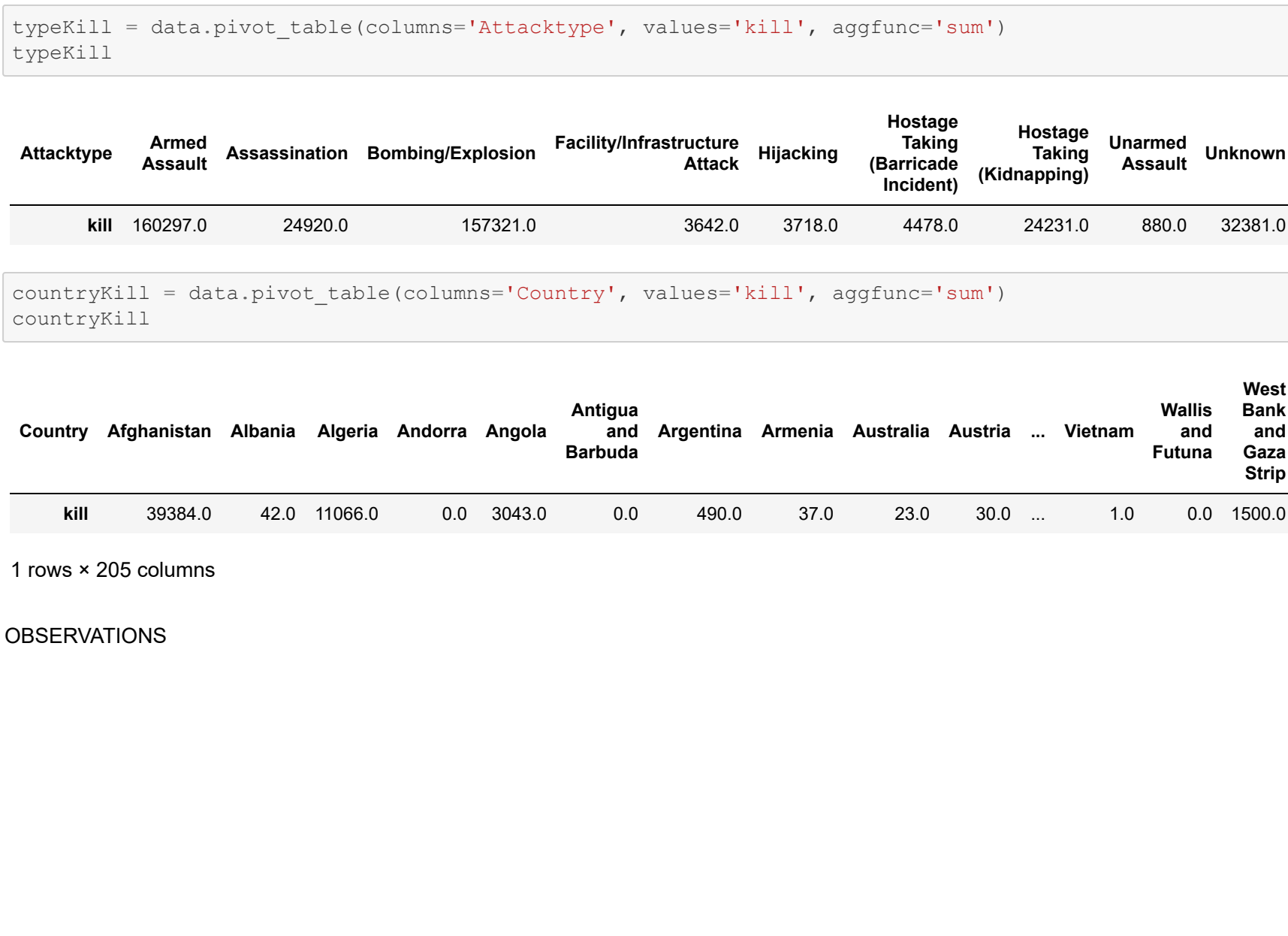
```
In [105]: data['Attacktype'].value_counts().plot(kind='bar',figsize=(20,10),color='violet')
plt.xticks(rotation = 50)
plt.title('Name of attacktype',fontsize=20)
plt.ylabel('Number of attack',fontsize=15)
plt.xlabel('Attack type',fontsize=15)
plt.show()
```



```
In [106]: 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()
```



```
In [107]: data[['Attacktype','Wound']].groupby(['Attacktype'],axis=0).sum().plot(kind='bar',figsize=(20,10),color='red')
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 [108]: plt.subplots(figsize=(20,10))
sns.countplot(data['Targettype'],order=data['Targettype'].value_counts().index,palette='gist_ncar_r',edgecolor=sns.color_palette("gist_ncar_r"))
plt.xticks(rotation=90)
plt.xlabel('Attacktype',fontsize=15)
plt.ylabel('Count',fontsize=15)
plt.title('Attack per year',fontsize=20)
plt.show()
```



```
In [109]: data['Group'].value_counts().to_frame().drop('Unknown').head(10).plot(kind='bar',color='gray',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()
```



```
In [112]: kill = data.loc[:,['kill']]
print('Number of people killed by terror attack:',int(sum(kill.dropna())))
```

Number of people killed by terror attack: 41169

```
In [113]: typeKill = data.pivot_table(columns='Attacktype', values='kill', aggfunc='sum')
typeKill
```

Attacktype	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	Hostage Taking (Baricade Incident)	Hostage Taking (Kidnapping)	Unarmed Assault	Unknown
kill	160297.0	24920.0	157321.0	36420.0	37180.0	44780.0	242310.0	8800.0	323810.0

```
In [114]: countryKill = data.pivot_table(columns='Country', values='kill', aggfunc='sum')
countryKill
```

Country	Afghanistan	Albania	Algeria	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	Australia	Austria	...	Vietnam	Wallis and Futuna	West Bank and Gaza Strip
kill	39384.0	42.0	11066.0	0.0	3043.0	0.0	490.0	37.0	23.0	30.0	...	1.0	0.0	1500.0

1 rows × 205 columns

OBSERVATIONS

1. Regarding the attacks based on yearwise:

A) ATTACKS a) Maximum number of attacks - 16903 in the year 2014 b) Minimum number of attacks - 471 in 1971

B)casualties a) Maximum number of casualties - 85618 in 2014 b) Minimum number of casualties - 255 in 1971

C) Death a) Maximum number of people reported dead - 44490 in 2014 b) Minimum number of people reported dead - 173 in 1971

D) Wounded A) Maximum number of people wounded - 44043 in 2015 B) Minimum number of people wounded - 82 in 1971

1. Regarding region wise attacks:

A)ATTACKS a) Maximum number of attacks - 50474 in "Middle east and North Africa" b) Least number of attacks - 282 in "Australiasia and oceania"

B) casualties a) Maximum number of casualties - 351950 in "Middle east and North Africa" b) Minimum number of casualties - 150 in "Australiasia and oceania"

C) Death a) Maximum number of people reported dead - 137642 in "Middle east and North Africa" b) Minimum number of people reported dead - 150 in "Australiasia and oceania"

D) Wounded a) Maximum number of wounded - 214308 in "Middle east and North Africa" b) Minimum number of wounded - 260 in "Australiasia and oceania"

Conclusion

Terrorist acts in the Middle East and northern Africa have been seen to have fatal consequences. The Middle East and North Africa are seen to be the places of serious terrorist attacks. In addition, even though there is a perception that Muslims are supporters of terrorism, Muslims are the people who are most damaged by terrorist attacks. If you look at the graphics, it appears that Iraq, Afghanistan and Pakistan are the most damaged countries. All of these countries are Muslim countries.