NAME: KRITHICK BALAJI RAMESH
ROLL_NO: RA2111003011318
SRM EMAIL: kr5623@srmist.edu.in
PERSONAL EMAIL ID: krithickbalaji2@gmail.com
CONTACT NO: +91 6385516155

PROJECT NAME: Exploratory Data analysis on global terrorism

```
#Importing the libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
```

```
#Importing the data set
data=pd.read_csv('/content/globalterrorismdb_0718dist.csv',encoding='latin1')
```

```
/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:3326: DtypeWarning: Columns (4,31,33,62,76,79,94,96,121) have mixed types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

```
#Checking the basic information from the given data set
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18870 entries, 0 to 18869
Columns: 135 entries, eventid to related
dtypes: float64(57), int64(22), object(56)
memory usage: 19.4+ MB
```

```
#Checking the shape of the given data set
data.shape
```

```
(18870, 135)
```

```
#Displaying the features name
data.columns
```

```
Index(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended',
       'resolution', 'country', 'country_txt', 'region',
       ...
       'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG',
       'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'],
      dtype='object', length=135)
```

```
#Checking the first five rows from the given data set
data.head()
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN

5 rows × 135 columns

```
#Checking the last five rows from the given data set
data.tail()
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scit
18865	198305240006	1983	5	24	NaN	0	NaN	45	Colombia	3	...	NaN	NaN	N
18866	198305240007	1983	5	24	NaN	0	NaN	45	Colombia	3	...	NaN	NaN	N
18867	198305240008	1983	5	24	NaN	0	NaN	145	Nicaragua	2	...	NaN	NaN	N
18868	198305240009	1983	5	24	NaN	0	NaN	145	Nicaragua	2	...	NaN	NaN	N
18869	198305240010	1983	5	24	NaN	0	NaN	603	United	8		NaN	NaN	N

```
data.rename(columns={'iyear':'Year','imonth':'Month','iday':"day",'gname':'Group','country_txt':'Country','region_txt':'Region','provstate':'State','city':'City','latitude':'latitude',
    'longitude':'longitude','summary':'summary','attacktype1_txt':'Attacktype','targtype1_txt':'Targettype','weaptype1_txt':'Weapon','nkill':'kill',
     'nwound':'Wound'},inplace=True)
```

```
#Statistical measure of the given data set
data.describe()
```

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude	sp
count	1.887000e+04	18870.000000	18870.000000	18870.000000	18870.000000	18870.000000	18870.000000	18123.000000	1.812300e+04	18
mean	1.978580e+11	1978.517011	6.198781	15.206465	0.019131	177.541176	5.416428	26.370738	-4.786345e+03	
std	3.406406e+08	3.409654	3.448318	8.955007	0.136989	166.162416	3.357730	22.515713	6.402079e+05	
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-45.867889	-8.618590e+07	
25%	1.977042e+11	1977.000000	3.000000	8.000000	0.000000	69.000000	2.000000	13.692880	-8.688120e+01	
50%	1.979101e+11	1979.000000	6.000000	15.000000	0.000000	110.000000	5.000000	33.888523	-6.926293e+00	
75%	1.981082e+11	1981.000000	9.000000	23.000000	0.000000	209.000000	8.000000	43.184180	9.174508e+00	
max	1.983052e+11	1986.000000	12.000000	31.000000	1.000000	605.000000	12.000000	64.837778	1.726362e+02	

8 rows × 79 columns

```
#Checking the sum of the missinf value from the given data set
data.isnull().sum()
```

```
eventid           0
iyear             0
imonth            0
iday              0
approxdate    18849
              ...
INT_LOG           1
INT_IDEO          1
INT_MISC          1
INT_ANY           1
related       18143
Length: 135, dtype: int64
```

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

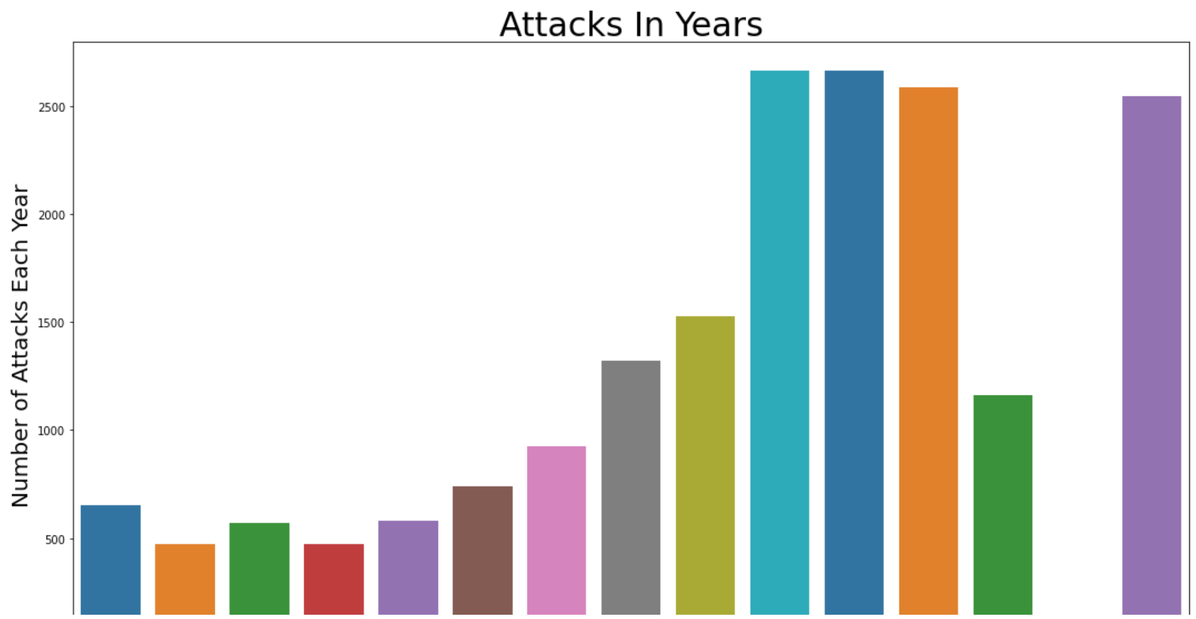
```
#Checking the number of unique values
data.nunique()
```

```
eventid       18870
iyear            15
imonth           13
iday             32
approxdate       19
              ...
INT_LOG           3
INT_IDEO          3
INT_MISC          3
INT_ANY           3
related         154
Length: 135, dtype: int64
```

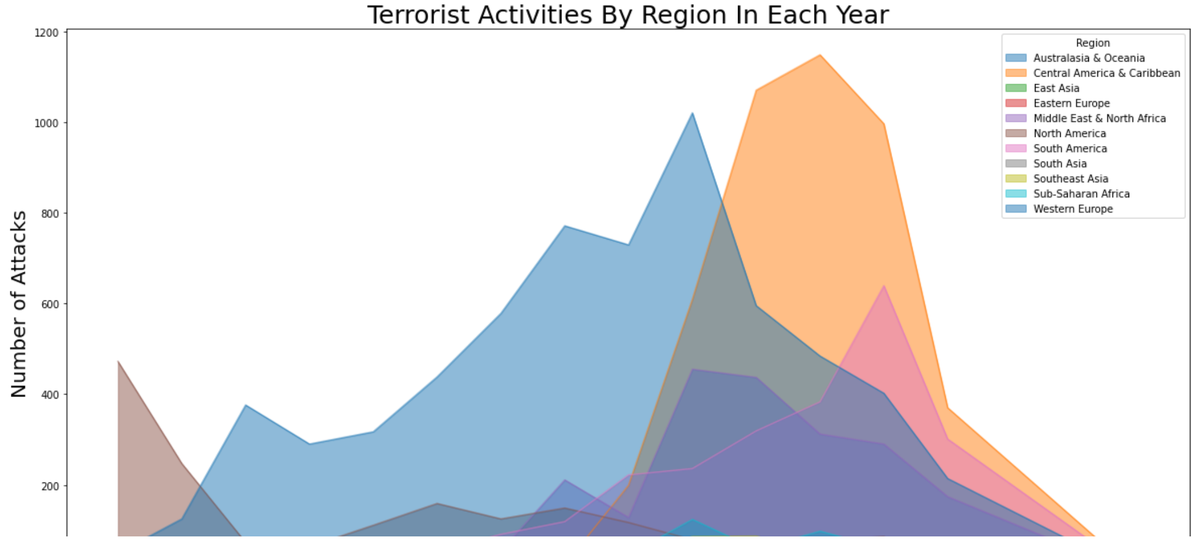
```
data['Casualities'] = data['kill'] + data['Wound']
```

```
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()
```




```
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()
```




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

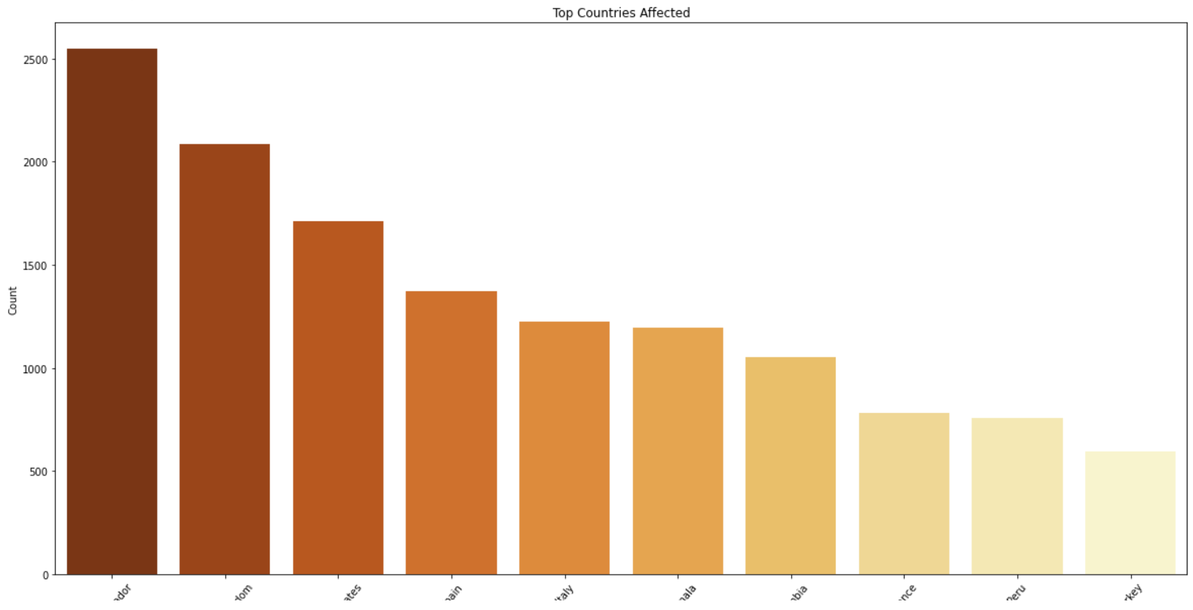
```
El Salvador       2547
United Kingdom    2087
United States     1711
Spain             1374
Italy             1226
Guatemala         1194
Colombia          1051
France             781
Peru               757
Turkey             593
Name: Country, dtype: int64
```

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

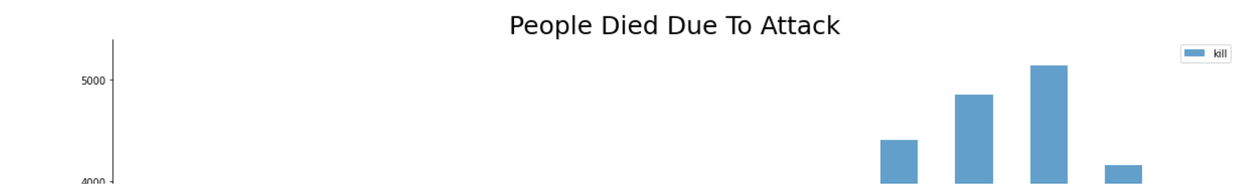
```
Irish Republican Army (IRA)                        1332
Farabundo Marti National Liberation Front (FMLN)    778
Basque Fatherland and Freedom (ETA)                 768
Shining Path (SL)                                   515
Palestinians                                        416
Revolutionary Armed Forces of Colombia (FARC)       310
M-19 (Movement of April 19)                         259
Protestant extremists                               220
Red Brigades                                        209
Name: Group, dtype: int64
```

```
#Printing the top countries affected
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()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y.
  FutureWarning
```

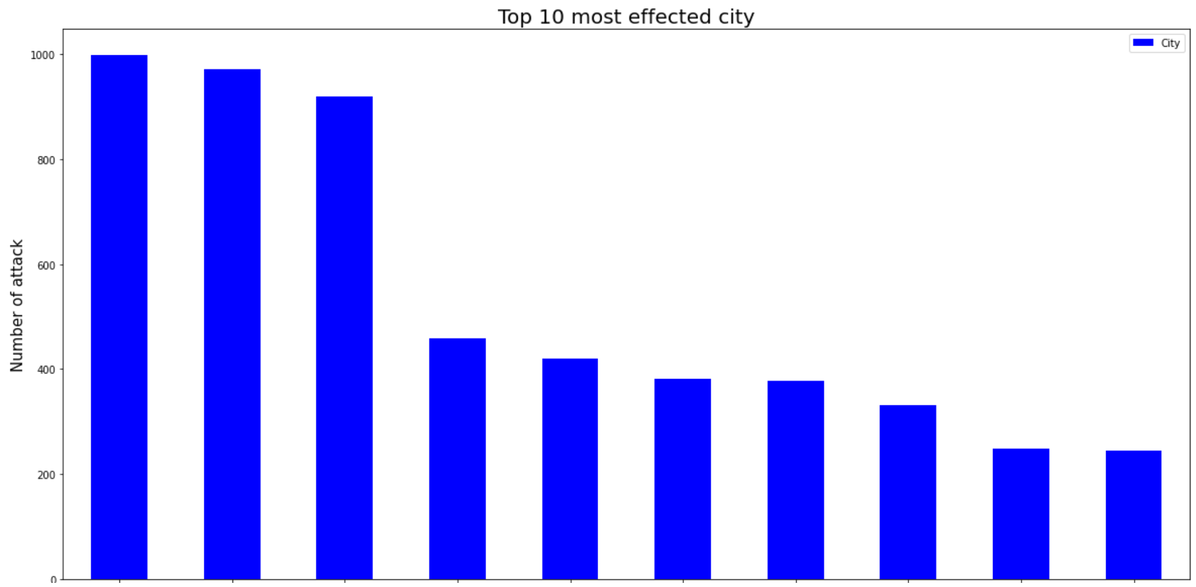



```
#Printing people died due to attack
df = data[['Year','kill']].groupby(['Year']).sum()
fig, ax4 = plt.subplots(figsize=(20,10))
df.plot(kind='bar',alpha=0.7,ax=ax4)
plt.xticks(rotation = 50)
plt.title("People Died Due To Attack",fontsize=25)
plt.ylabel("Number of killed peope",fontsize=20)
plt.xlabel('Year',fontsize=20)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



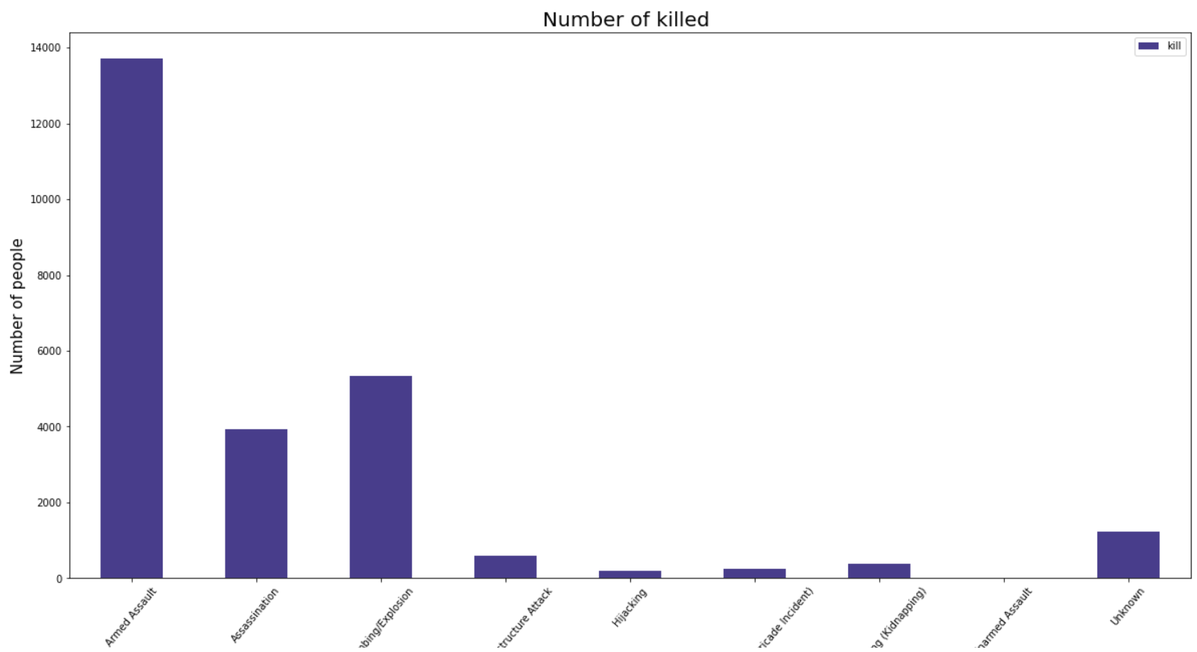

```
#Printing the top most effected city

data['City'].value_counts().to_frame().sort_values('City',axis=0,ascending=False).head(10).plot(kind='bar',figsize=(20,10),color='blue')
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()
```

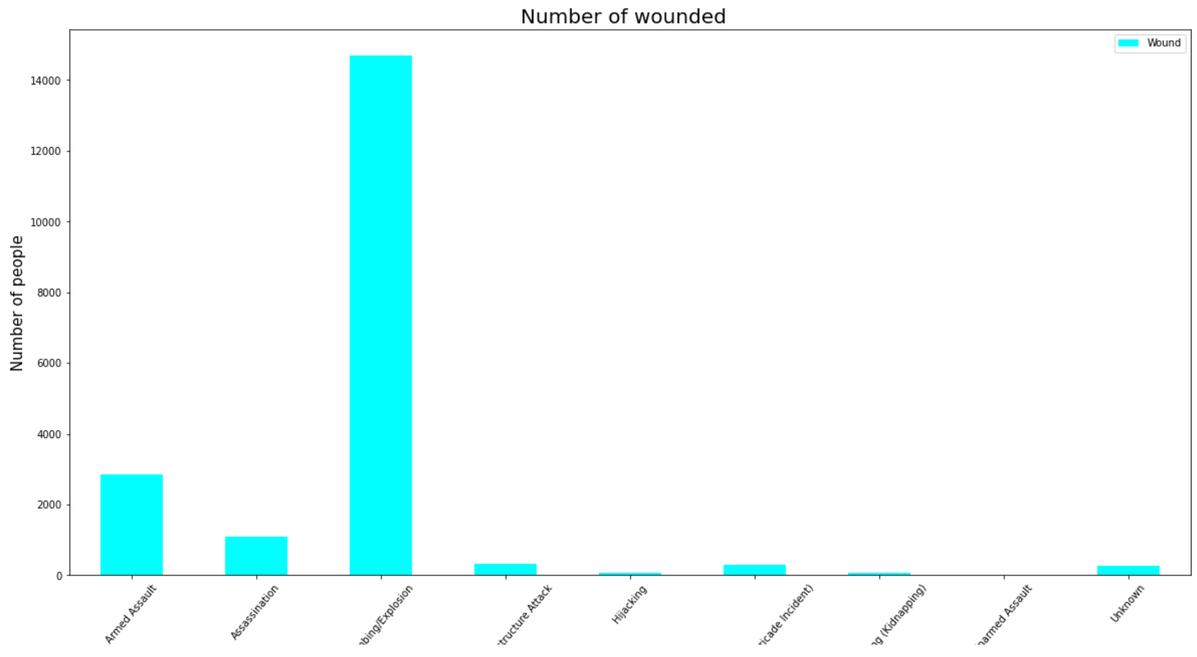



```
#Printing the name of the attack type
data['Attacktype'].value_counts().plot(kind='bar',figsize=(20,10),color='magenta')
plt.xticks(rotation = 50)
plt.xlabel("Attacktype",fontsize=15)
plt.ylabel("Number of attack",fontsize=15)
plt.title("Name of attacktype",fontsize=20)
plt.show()
```

```
#Printing the Number of killed
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()
```



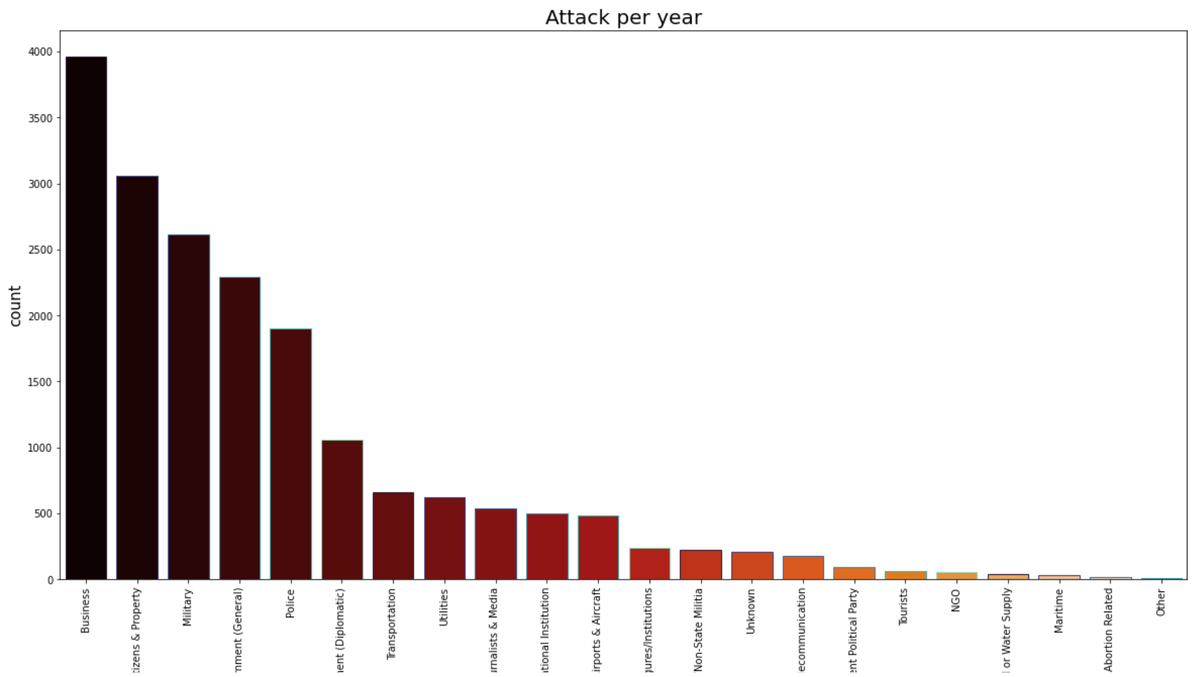

```
#Printing the number of wounded
data[['Attacktype','Wound']].groupby(["Attacktype"],axis=0).sum().plot(kind='bar',figsize=(20,10),color=['cyan'])
plt.xticks(rotation=50)
plt.title("Number of wounded  ",fontsize=20)
plt.ylabel('Number of people',fontsize=15)
plt.xlabel('Attack type',fontsize=15)
plt.show()
```



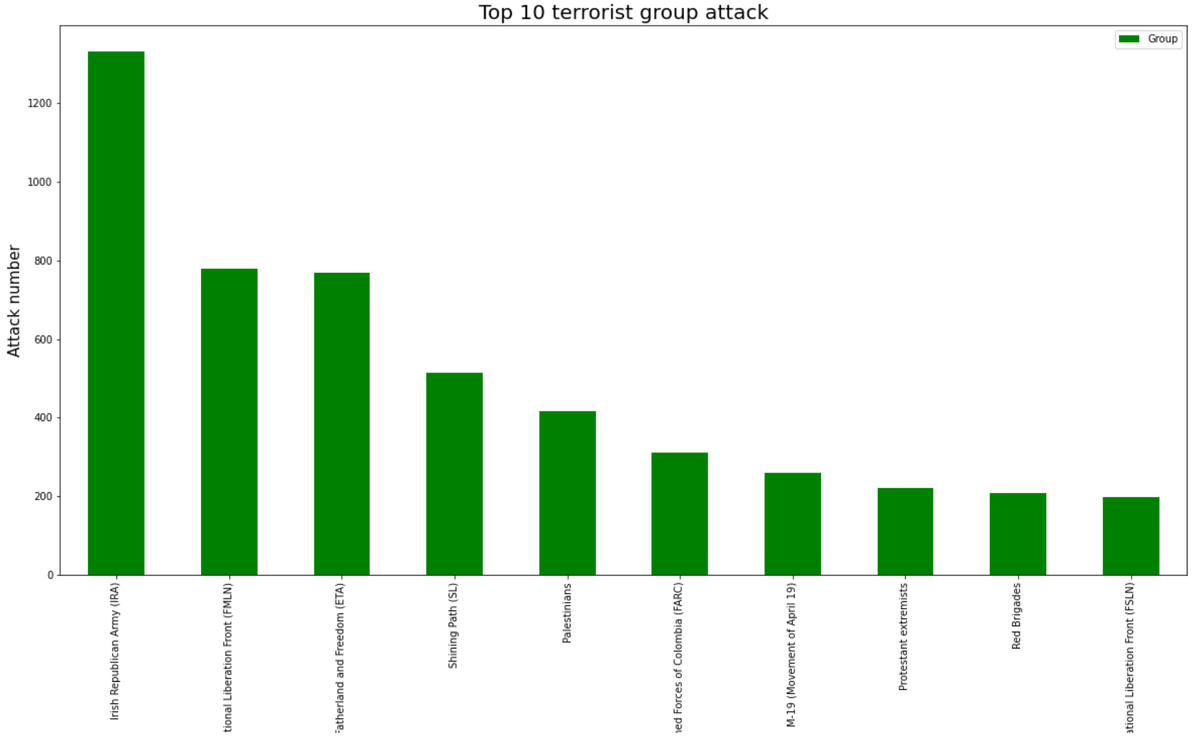

```
#Printing the attack per year
plt.subplots(figsize=(20,10))
sns.countplot(data["Targettype"],order=data['Targettype'].value_counts().index,palette="gist_heat",edgecolor=sns.color_palette("mako"));
plt.xticks(rotation=90)
```

```
plt.xlabel("Attacktype",fontsize=15)
plt.ylabel("count",fontsize=15)
plt.title("Attack per year",fontsize=20)
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variable as a keyword arg: x. Fro
  FutureWarning
```




```
#Printing the top 10 terrorist group attack
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()
```




```
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
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/generic.py:4150: PerformanceWarning: dropping on a non-lexsorted multi-index witho
  obj = obj._drop_axis(labels, axis, level=level, errors=errors)
```

	Group	Country	kill
0	Farabundo Marti National Liberation Front (FMLN)	El Salvador	3022.0
1	Irish Republican Army (IRA)	United Kingdom	1176.0
2	Shining Path (SL)	Peru	1165.0
3	Nicaraguan Democratic Force (FDN)	Nicaragua	951.0
4	Revolutionary Armed Forces of Colombia (FARC)	Colombia	700.0
5	Guerrilla Army of the Poor (EGP)	Guatemala	533.0
6	Mujahedin-e Khalq (MEK)	Iran	489.0
7	Contras	Nicaragua	420.0
8	Basque Fatherland and Freedom (ETA)	Spain	371.0
9	Moro National Liberation Front (MNLF)	Philippines	369.0

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

```
Number of people killed by terror attack: 25667
```

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

Attacktype	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking	Hostage Taking (Barricade Incident)	Hostage Taking (Kidnapping)	Unarmed Assault	Unknowr
kill	13718.0	3941.0	5348.0	591.0	186.0	240.0	389.0	19.0	1235.(

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

Country	Afghanistan	Albania	Algeria	Andorra	Angola	Argentina	Australia	Austria	Bahamas	Bahrain	...	Uruguay	Vatican City	Vene
kill	53.0	0.0	0.0	0.0	134.0	329.0	6.0	12.0	0.0	0.0	...	4.0	0.0	

1 rows × 129 columns

```
Conclusion
1. Country with the most attacks: Iraq
2. City with the most attacks: Baghdad
Regsion with the most attacks: Middle East & North
Africa.
3. Year with the most attacks: 2014

4. Month with the most attacks: 5

5. Group with the most attacks: Taliban

6. Most Attack Types: Bombing/Explosion
```

Conclusion

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

Colab paid products - Cancel contracts here

✓ 0s completed at 10:46 AM