Name : Akanksha Umak

Data Science and Business Analytics Intern @ The Sparks Foundation

Topic : Exploratory Data Analysis (EDA) - Terrorism

Dataset : globalterrorismdb_0718dist.csv [https://bit.ly/2TK5Xn5]

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
urlfp = "data/globalterrorismdb/globalterrorismdb_0718dist.csv"
df = pd.read_csv(urlfp)
print("The data is imported successfully!")
df.head()
```

```
The data is imported successfully!
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_A	weapsubtype1
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	1
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
2	197000000003	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	1	NaN
3	197000000004	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
4	197000000005	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN

5 rows × 135 columns

In [3]:
```
df.columns values
```

In [4]:
```
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', 'multinat',
       'search', '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', 'nwoundter', 'property', 'ishostkid',
       '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 [5]:
```
df.rename(columns=['iyear', 'imonth', 'iday', 'country_txt', 'region', 'city', 'latitude', 'longitude', 'attacktype1', 'kill',
                   'wound', 'target1', 'summary', 'group', 'targtype1', 'weaptype1', 'weapdetail'], inplace=True)
```

In [6]:
```
df.head()
```

Out[6]:

	eventid	iyear	imonth	iday	country	country_txt	region	region_txt	provstate	city	latitude	longitude	specificity	vicinity	location	summary	crit1	crit2	crit3	doubtterr	alternative	alternative_txt	multinat	search	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	nwoundter	property	ishostkid	ransomamt	ransomamtus	ransompaid	ransompaidus	ransomnote	hostkidoutcome	hostkidoutcome_txt	nreleased	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
0	197	7	2	Dominican Republic	2	Central America & Caribbean	61.454762	-69.961164	Hazomindoro	1.0	0.0	John Duncan	NaN	58443.0	Private Citizens & Property	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	Explosives	5.0	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	197	0	0	Mexico	Federal	North America	Mexico City	19.371887	-99.086624	Foreign Minister and Attaché	0.0	0.0	Nestor Duarte Frayde	NaN	58446.0	Private Citizens & Property	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6	Incendiary	3.0	Government (General)	Diplomatic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
2	197	1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741	Unknown	1.0	0.0	Unknown	NaN	58447.0	Private Citizens & Property	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6	Incendiary	3.0	Government (General)	Diplomatic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	PGIS	0	0	0	1	NaN
3	197	1	0	Greece	Attica	Western Europe	Athens	37.997490	23.762728	Bombing/Explosion	NaN	NaN	U.S. Embassy	NaN	Unknown	Government (General)	Diplomatic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6	Explosives	100.0	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN
4	197	1	0	Japan	Fukushima	East Asia	Fukushima	37.494013	139.928150	Facility/Infrastructure Attack	NaN	NaN	U.S. Consulate	NaN	Unknown	Government (Diplomatic)	Kidnapping	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5	Firearms	22.0	Unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	PGIS	0	0	1	1	NaN

In [7]:
```
df shape
```

Out[7]:
```
(181691, 135)
```

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

Out[8]:
```
eventid            0
iyear              0
imonth             0
iday               0
country            0
country_txt        0
region             0
region_txt         0
provstate       4606
city            434
latitude        4606
longitude       4606
attacktype1        0
kill           9730
wound         16311
target1         434
weaptype1          0
weapdetail     66135
nkill           9682
nkillus        75837
nkillter       76381
nwound         16628
nwoundus       91960
nwoundter      92503
dtype: int64
```

In [9]:
```
df.describe()
```

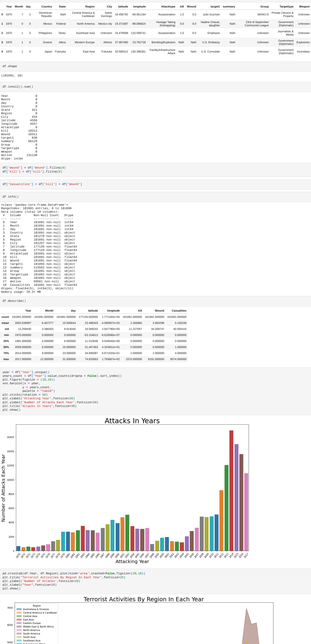
Out[9]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	region	latitude	longitude	specificity	vicinity	crit1	crit2	crit3	doubtterr	alternative	multinat	search	weaptype1	weapsubtype1	weaptype2	weapsubtype2	weaptype3	weapsubtype3	weaptype4	weapsubtype4	nkill	nkillus	nkillter	nwound	nwoundus	nwoundter	property	ishostkid	ransomamt	ransomamtus	ransompaid	ransompaidus	ransomnote	hostkidoutcome	nreleased	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
count	181691.000000	181691.000000	181691.000000	181691.000000	17539.000000	181691.000000	177135.000000	181691.000000	181691.000000	177134.000000	177134.000000	181690.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	3563.000000	181691.000000	181691.000000	181691.000000	168659.000000	10313.000000	2661.000000	437.000000	96.000000	58.000000	171378.000000	166989.000000	170500.000000	167420.000000	162711.000000	159707.000000	181691.000000	181691.000000	50421.000000	50421.000000	50421.000000	50421.000000	50421.000000	50421.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000
mean	201312000297	2013.830319	6.467277	15.505644	23.498469	0.046304	2.546564	131.968501	6.475166	23.498469	-138.580745	2.300962	0.000266	0.571230	0.554753	0.225744	0.035381	0.769474	0.096412	0.071863	4.407457	18.569241	22.351816	36.797823	3.451715	22.789474	2.827586	22.465649	2.325326	0.304615	0.324468	2.139160	0.099308	0.083397	0.117279	0.069622	0.051890	0.051465	0.036544	0.063455	0.043213	0.435480	0.475660	0.866830	0.837493	0.775210	0.814480	0.546269	0.543076	0.563369	0.354213	0.476995
std	1.200986	0.564241	3.463231	8.814314	0.500051	0.210011	0.816689	116.714990	2.188829	0.500051	140.900616	0.907223	0.016293	0.494875	0.496843	0.417930	0.184661	0.421415	0.294981	0.258326	2.549775	9.444963	7.913762	25.549976	1.732780	15.890602	0.378771	9.131314	0.746529	0.460299	0.468053	0.345731	0.298672	0.276781	0.321884	0.254283	0.221829	0.221703	0.187708	0.243875	0.203381	0.495832	0.499375	0.340096	0.369203	0.417566	0.388972	0.497805	0.497993	0.495985	0.478549	0.499475
min	170000000001	1970.000000	0.000000	0.000000	0.000000	0.000000	1.000000	4.000000	1.000000	0.000000	-54.241403	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	201207050284	2013.000000	3.000000	8.000000	23.000000	0.000000	2.000000	78.000000	4.000000	23.000000	-116.772900	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	4.000000	13.000000	15.000000	36.000000	2.000000	8.000000	2.000000	15.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	201312090001	2014.000000	6.000000	15.000000	24.000000	0.000000	2.000000	98.000000	6.000000	24.000000	-76.000000	2.000000	0.000000	1.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	4.000000	22.000000	22.000000	36.000000	3.000000	22.000000	3.000000	22.000000	2.000000	0.000000	0.000000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	201411300001	2014.000000	9.000000	23.000000	24.000000	0.000000	3.000000	160.000000	8.000000	24.000000	23.130000	3.000000	0.000000	1.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	6.000000	23.000000	29.000000	36.000000	5.000000	36.000000	3.000000	36.000000	3.000000	0.000000	0.000000	3.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
max	201612000001	2016.000000	12.000000	31.000000	31.000000	1.000000	4.000000	1004.000000	12.000000	31.000000	179.366667	5.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	1.000000	1.000000	13.000000	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	9.000000	1.000000	1.000000	9.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

1 rows × 135 columns

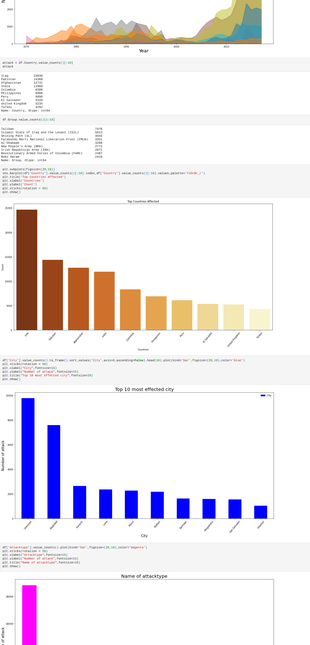
In [10]:
```
iyear = df['iyear'].unique()
attack = df.groupby('iyear').count()['eventid']#count of attacks
plt.figure(figsize = (10,10))
plt.plot(attack,
         iyear,
         c = 'red',
         marker = 'o',
         ms = 5,
         mfc = 'black')
plt.xlabel('Attacking Year', fontsize=20)
plt.ylabel('Number of Attacks', fontsize=20)
plt.title('Attacks In Years', fontsize=30)
plt.show()
```

Out[10]:



In [11]:
```
plt.stackplot(df.iyear, df.Region).plot(kind='area',stackedfalse,figsize=(20,10))
plt.xlabel('Year', fontsize=20)
plt.ylabel('Number of Attacks', fontsize=20)
plt.title('Terrorist Activities By Region In Each Year', fontsize=30)
plt.show()
```

Out[11]:



In [12]:
```
df.iyear.value_counts[:10].sort
```

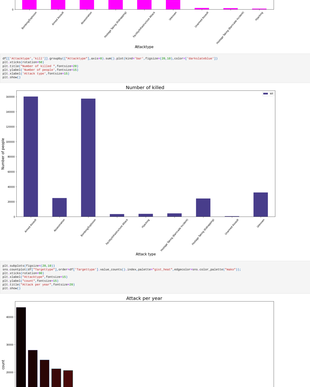
Out[12]:
```
iyear
2014    16639
2013    16996
2012    14606
2016    14308
2015    14965
2011    15996
2010    14036
2009    11994
2008    11663
2007    10900
Name:, dtype: int64
```

In [13]:
```
df.Group.value_counts[:15].sort
```

Out[13]:
```
Taliban                                   7478
Islamic State of Iraq and the Levant (ISIL)  5613
Shining Path (Sendero Luminoso)           4761
New People's Army (NPA)                   3952
Al-Shabaab                                3477
Hamas (Islamic Resistance Movement)       2732
Liberation Tigers of Tamil Eelam (LTTE)   2393
Nakba Day Group                           2024
Boko Haram                                1907
Revolutionary Armed Forces of Colombia (FARC)  1850
Kurds                                     1845
Maoist Group                              1768
Sinai Peninsula Province                  1715
Name: Group, dtype: int64
```

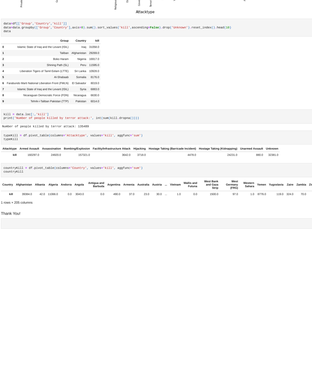
In [14]:
```
plt.subplots(figsize=(15,10))
sns.countplot(df['Group'].value_counts[:10].index,df['Country'].value_counts[:10].values,palette="Oleds_r")
plt.xlabel('Top Countries Affected')
plt.ylabel('Count')
plt.title('Country')
plt.xticks(rotation= 90)
plt.show()
```

Out[14]:



In [15]:
```
df['City'].value_counts[:11].T.plot(kind='bar',figsize=(20,10),color='blue')
plt.xlabel('Country')
plt.ylabel('Number of attack')
plt.title('Top 10 most effected city', fontsize=30)
plt.show()
```

Out[15]:



In [16]:
```
df[ 'Attacktype'].value_counts[:10].plot(kind='bar',figsize=(20,10),color='magenta')
plt.xlabel('Attacktype')
plt.ylabel('Attacktype', fontsize=20)
plt.title('Name of attacktype', fontsize=30)
plt.show()
```

Out[16]:

In [17]:
```
df[ 'Attacktype'].sort 1].groupby('Attacktype').agg(dt.sum).plot(kind='bar',figsize=(20,10),color='darkslateblue')
plt.xlabel('Attacktype')
plt.ylabel('Number of people', fontsize=20)
plt.title('Number of killed', fontsize=30)
plt.show()
plt.title('Attack type', fontsize=30)
```

Out[17]:

In [18]:
```
df.groupby('Group', 'Country', 'kill')
dataset=df.groupby('Country', ).size(0).sort_values(kill').ascending=False).drop('Unknown').reset_index).head(10)
dataset
```

Out[18]:

	Country	Group	kill
0	Islamic State of Iraq and the Levant (ISIL)	Taliban	7478
1	Afghanistan	Islamic State of Iraq and the Levant (ISIL)	5613
2	Peru	Shining Path (Sendero Luminoso)	4761
3	Philippines	New People's Army (NPA)	3952
4	Somalia	Al-Shabaab	3477
5	Israel and the Occupied Territories	Hamas (Islamic Resistance Movement)	2732
6	Sri Lanka	Liberation Tigers of Tamil Eelam (LTTE)	2393
7	Nigeria	Boko Haram	1907
8	Colombia	Revolutionary Armed Forces of Colombia (FARC)	1850
9	Pakistan	Tehrik-i-Taliban Pakistan (TTP)	1810

In [19]:
```
kill = data.loc[:,'kill']
print('Number of people killed by terror attack:', int(sum(kill).dropna()))
```

```
Number of people killed by terror attack: 1254000
```

In [20]:
```
type(kill) # pd.pd_table(columns='Attacktype', values='kill', aggfunc='sum')
type(kill)
```

Out[20]:
```
pandas.core.series.Series
```

In [21]:
```
Attacktype  Armed Assault  Assassination  Bombing/Explosion  Facility/Infrastructure Attack  Hijacking  Hostage Taking (Barricade Incident)  Hostage Taking (Kidnapping)  Unarmed Assault  Unknown
Country
Afghanistan  14187.0  2400.0  15732.0  368.0  1714.0  4479.0  2429.0  80.0  3230.0
```

In [22]:
```
country = df.pivot_table(columns='Country', values='kill', aggfunc='sum')
country[:10]
```

Out[22]:

Country	Afghanistan	Albania	Algeria	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	Australia	Austria	...	Germany	Ghana	Greece	Guatemala	Guinea	Guinea-Bissau	Guyana	Haiti	Honduras	Hungary
kill	30644.0	403.0	11884.0	0.0	3843.0	0.0	4862.0	37.0	153.0	38.0	...	1.0	0.0	2383.0	97.0	0.0	0.0	0.0	0.0	0.0	0.0

1 rows × 205 columns

Thank You