

```
# Importing Libraries
import warnings
import numpy as np # linear algebra
import pandas as pd # data processing
import seaborn as sns
import plotly.offline as py
import matplotlib.pyplot as plt
warnings.filterwarnings('ignore')

In [1]: import os
for dirname, _, filenames in os.walk('globalterrorismdb_0718dist.csv'):
    print(os.path.join(dirname, filename))

In [2]: terror = pd.read_csv('globalterrorismdb_0718dist.csv', encoding='ISO-8859-1')

# Basic EDA
terror.head()

Out[4]:
eventid year month day approximate extended resolution country country_iso region ... addnotes scite1 scite2 dbsource NT_LOGO INT_BEOO INT_MISC INT_ANY related
0 1970000001 1970 7 2 NaN 0 NaN 58 Dominican Republic 2 -- NaN NaN NaN NaN POS -0 0 0 0 NaN
1 1970000002 1970 0 0 NaN 0 NaN 130 Mexico 1 -- NaN NaN NaN NaN POS -0 1 1 1 NaN
2 1970000001 1970 1 0 NaN 0 NaN 160 Philippines 5 -- NaN NaN NaN NaN POS -9 -9 1 1 NaN
3 1970000002 1970 1 0 NaN 0 NaN 78 Greece 8 -- NaN NaN NaN NaN POS -0 0 1 1 NaN
4 1970000003 1970 1 0 NaN 0 NaN 101 Japan 4 -- NaN NaN NaN NaN POS -9 -9 1 1 NaN
5 rows x 135 columns

# Terror: columns
Index(['eventid', 'year', 'month', 'day', 'approxdate', 'extended',
       'resolution', 'country', 'country_txt', 'region',
       ...,
       'addnotes', 'scite1', 'scite2', 'dbsource', 'INT_LOGO',
       'INT_BEOO', 'INT_MISC', 'INT_ANY', 'related'],
      dtype='object', length=135)

In [6]: terror.rename(columns={'year':'Year','month':'Month','day':'Day','country_txt':'Country','proxstate':'State',
                             'region_txt':'Region','attacktype1_ext':'AttackType','target1':'Target','kill1':'Killed',
                             'reason':'Wounded','summary':'Summary','gnane':'Group','targettype1':'Target_Type',
                             'weaponstype1':'Weapon_Type','motive':'Motive'}, inplace=True)

In [7]: # Finding only important data
terror.rename({'Year':'Year','Month':'Month','State':'State','Region':'Region','city':'Latitude','Longitude':'Longitude','AttackType':'Killed',
              'Wounded':'Wounded','Target':'Summary','Group':'Target_Type','Weapon_Type':'Motive'})

In [8]: # Checking the null values in data
terror.isnull().sum()

Out[8]:
Month          0
Day            0
Country         0
State        421
Region         0
City           0
Latitude       0
Longitude     457
AttackType     0
Killed       1033
Wounded      1631
Target        638
Summary      66129
Day_Type      0
Weapon_Type   0
Motive       13139
dtype: int64

In [9]: terror.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181692 entries, 0 to 181692
Data columns (total 18 columns):
# Column Non-Null Count Dtype ---
0 Year 181692 non-null int64
1 Month 181692 non-null object
2 Day 181692 non-null int64
3 Country 181692 non-null object
4 State 181270 non-null object
5 Region 181692 non-null object
6 City 181256 non-null object
7 Latitude 177235 non-null float64
8 Longitude 177234 non-null float64
9 AttackType 181692 non-null object
10 Killed 172170 non-null float64
11 Wounded 165380 non-null float64
12 Target 181692 non-null object
13 Summary 115062 non-null object
14 Group 181692 non-null object
15 Target_Type 181692 non-null object
16 Weapon_Type 181692 non-null object
17 Motive 15551 non-null object
dtypes: float64(4), int64(3), object(11)
memory usage: 25.0 MB

In [10]: # Descriptive Feature of data
print("Country with the most attacks:", terror['Country'].value_counts().idxmax())
print("City with the most attacks:", terror['City'].value_counts().index[1]) #as first entry is 'unknown'
print("Region with the most attacks:", terror['Region'].value_counts().idxmax())
print("Year with the most attacks:", terror['Year'].value_counts().idxmax())
print("Month with the most attacks:", terror['Month'].value_counts().idxmax())
print("Group with the most attacks:", terror['Group'].value_counts().index[1])
print("Most Attack Types:", terror['AttackType'].value_counts().idxmax())

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

In [11]: terror['Year'].value_counts(dropna=False).sort_index()

Out[14]:
Year
1979    651
1980    534
1981    568
1982    412
1983    478
1984    748
1985    822
1986    1538
1987    2682
1988    2586
1989    2544
1990    2978
1991    3495
1992    3097
1993    2968
1994    3058
1995    3197
1996    334
1997    1395
1998    1014
1999    1906
2000    1533
2001    1278
2002    1166
2003    2017
2004    2758
2005    3242
2006    4805
2007    4555
2008    4628
2009    5078
2010    8522
2011    12636
2012    15903
2013    14905
2014    15903
2015    15903
2016    15903
2017    15903
Name: count, dtype: int64

In [12]: #Number of terrorist activities per year
x_year = terror['Year'].unique()
y_count_years = terror['Year'].value_counts(dropna=False).sort_index()
plt.figure(figsize=(18,10))
sns.barplot(x=x_year, y=y_count_years, palette='rocket')
plt.xticks(rotation=45)
plt.xlabel('Attack Year')
plt.ylabel('Number of Attacks each year')
plt.title('Attack of Years')
plt.show()

In [13]: pd.crosstab([terror.Year, terror.Region], kind='area', figsize=(15,6))
plt.title('Terrorist Activities by Region in each Year')
plt.xlabel('Number of Attacks')
fig, ax = plt.subplots(1, 1)

In [14]: # Terrorist Activities by Region in each Year
fig, ax = plt.subplots(1, 1)

In [15]: terror['Year'] = terror['Year'].astype(int)
terror['Killed'] = terror['Killed'].fillna(0).astype(int)
terror['casualties'] = terror['casualties'].fillna(0).astype(int)

In [16]: terror = terror.sort_values(by='casualties', ascending=False)[ :40] #Values are sorted by the top 40 worst terror attacks as to keep the heatmap simple and fast

In [17]: heat = terror.pivot_table(index='Country', columns='Year', values='casualties')
heat.style.background_gradient(cmap='inferno')

In [18]: heat.head()

Year 1982 1984 1992 1994 1996 1998 1999 2000 2004 2005 2006 2007 2008 2009 2014 2015 2016 2017
Country
Alghansian 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
Iraq 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
Afghanistan 0.0 0.0 500.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
France 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
India 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

In [19]: import plotly.offline as py
py.init_notebook_mode(connected=True)
import plotly.graph_objs as go
colorscale = ['#f08080', '#c0c0c0', '#e0e0e0', '#d0d0d0', '#a0a0a0', '#808080', '#606060', '#404040', '#202020', '#000000']
heatmap = go.Heatmap(z=heat.values, x=heat.columns, y=heat.index, colorscale=colorscale)
data = [heatmap]
fig = Figure(data=[heatmap])
fig.layout.update(title='Top 40 Worst Terror Attacks in History from 1982 to 2016',
                  xaxis=dict(ticks='', nticks=20),
                  yaxis=dict(ticks='', nticks=20))
fig.show()

In [20]: Top 40 Worst Terror Attacks in History from 1982 to 2016

United States 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Syria 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Sri Lanka 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
South Sudan 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Rwanda 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Somalia 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Nigeria 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Yemen 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Spain 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2017
Name: count, dtype: int64

In [21]: #Top Countries affected by Terror Attacks
x_year = terror['Year'].unique()
y_count_years = terror['Year'].value_counts(dropna=False).sort_index()
plt.figure(figsize=(18,10))
sns.barplot(x=x_year, y=y_count_years, palette='rocket')
plt.xticks(rotation=45)
plt.xlabel('Countries')
plt.ylabel('Count')
plt.title('Top Countries Affected')
plt.show()

In [22]: Top Countries Affected

Country
Iraq 24636
Pakistan 14368
Afghanistan 12731
India 11968
Colombia 8306
Philippines 6908
Peru 6098
El Salvador 5329
New People's Army 5225
Turkey 4292
Somalia 4162
Nigeria 3907
Thailand 3849
Yemen 3341
Spain 3249
Name: count, dtype: int64

In [23]: test = terror(terror.groupby(['Shining Path (SL)', 'Taliban', 'Islamic State of Iraq and the
```