THE SPARKS FOUNDATION

VARSHA REDDY

BATCH-SEP2022

Task4 :- Problem Statement - Exploratory Data Analysis on "Global Terrorism"

Dataset-https://bit.ly/2TK5Xn5

Importing Libraries for Data Manipulation

In [1]:
```
import pandas as pd
import numpy as np
```

Importing Libraries for Data Visualization

In [2]:
```
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import plotly.express as px
import plotly.graph_objects as go
from collections import Counter
import seaborn as sns
%matplotlib inline
```

Importing the Dataset

In [3]:
```
data = pd.read_csv("global terrorism.csv",encoding = "ISO-8859-1")
```

```
C:\Users\om reddy\AppData\Local\Temp\ipykernel_15128\2479220150.py:1: DtypeWarning: Columns (4,6,31,33,61,62,63,76,79,90,92,94,96,114,115,121) have mixed types. Specify dtype option on import or set low_memory=False.
  data = pd.read_csv("global terrorism.csv",encoding = "ISO-8859-1")
```

PreProcessing the Data

In [4]:
```
data.head()
```

Out[4]:

	s	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	1.970000e+11	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	1	1	1	NaN
2	1.970010e+11	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
3	1.970010e+11	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN
4	1.970010e+11	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN

5 rows × 135 columns

In [5]:
```
data.tail()
```

Out[5]:

	s	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
181686	2.017120e+11	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab Militants Attack Army Che...	"Highlights: Somalia Daily Media Highlights 2 ...	"Highlights: Somalia Daily Media Highlights 1 ...	START Primary Collection	0	0	0	0	NaN
181687	2.017120e+11	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Putin's 'victory' in Syria has turned into a ...	"Two Russian soldiers killed at Hmeymim base i...	"Two Russian servicemen killed in Syria mortar...	START Primary Collection	-9	-9	1	1	NaN
181688	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...	NaN	NaN	START Primary Collection	0	0	0	0	NaN
181689	2.017120e+11	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...	NaN	NaN	START Primary Collection	-9	-9	0	-9	NaN
181690	2.017120e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...	"Security tightened in Cotabato City," Manila ...	NaN	START Primary Collection	-9	-9	0	-9	NaN

5 rows × 135 columns

In [6]:
```
data.columns
```

Out[6]:
```
Index(['s', '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)
```

In [7]:
```
print(data.keys())
```

```
Index(['s', '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)
```

In [8]:
```
data.shape
```

Out[8]:
```
(181691, 135)
```

In [9]:
```
data.isnull().sum()
```

Out[9]:
```
s                  0
iyear              0
imonth             0
iday               0
approxdate    172452
               ...  
INT_LOG            0
INT_IDEO           0
INT_MISC           0
INT_ANY            0
related       156653
Length: 135, dtype: int64
```

In [10]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, s to related
dtypes: float64(56), int64(21), object(58)
memory usage: 187.1+ MB
```

In [11]:
```
print(data.duplicated().sum())
```

```
4550
```

Renaming the Columns

In [12]:
```
data.rename(columns={'iyear':'Year','imonth':'Month','city':'City','iday':'Day','country_txt':'Country','region_txt':'Region','attacktype1_txt':'AttackType','target1':'Target','nkill':'Killed','nwound':'Wounded','summary':'Summary
data['Casualities'] = data.Killed + data.Wounded
data=data[['Year','Month','Day','Country','Region','City','latitude','longitude','AttackType','Killed','Wounded','Casualities','Target','Group','Target_type','Weapon_type']]
data.head(10)
```

Out[12]:

	Year	Month	Day	Country	Region	City	latitude	longitude	AttackType	Killed	Wounded	Casualities	Target	Group	Target_type	Weapon_type
0	1970	7	2	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	1.0	0.0	1.0	Julio Guzman	MANO-D	Private Citizens & Property	Unknown
1	1970	0	0	Mexico	North America	Mexico city	19.371887	-99.086624	Hostage Taking (Kidnapping)	0.0	0.0	0.0	Nadine Chaval, daughter	23rd of September Communist League	Government (Diplomatic)	Unknown
2	1970	1	0	Philippines	Southeast Asia	Unknown	15.478598	120.599741	Assassination	1.0	0.0	1.0	Employee	Unknown	Journalists & Media	Unknown
3	1970	1	0	Greece	Western Europe	Athens	37.997490	23.762728	Bombing/Explosion	NaN	NaN	NaN	U.S. Embassy	Unknown	Government (Diplomatic)	Explosives
4	1970	1	0	Japan	East Asia	Fukouka	33.580412	130.396361	Facility/Infrastructure Attack	NaN	NaN	NaN	U.S. Consulate	Unknown	Government (Diplomatic)	Incendiary
5	1970	1	1	United States	North America	Cairo	37.005105	-89.176269	Armed Assault	0.0	0.0	0.0	Cairo Police Headquarters	Black Nationalists	Police	Firearms
6	1970	1	2	Uruguay	South America	Montevideo	-34.891151	-56.187214	Assassination	0.0	0.0	0.0	Juan Maria de Lucah/Chief of Directorate of in...	Tupamaros (Uruguay)	Police	Firearms
7	1970	1	2	United States	North America	Oakland	37.791927	-122.225906	Bombing/Explosion	0.0	0.0	0.0	Edes Substation	Unknown	Utilities	Explosives
8	1970	1	2	United States	North America	Madison	43.076592	-89.412488	Facility/Infrastructure Attack	0.0	0.0	0.0	R.O.T.C. offices at University of Wisconsin, M...	New Year's Gang	Military	Incendiary
9	1970	1	3	United States	North America	Madison	43.072950	-89.386694	Facility/Infrastructure Attack	0.0	0.0	0.0	Selective Service Headquarters in Madison Wisc...	New Year's Gang	Government (General)	Incendiary

In [13]:
```
data.Country.value_counts()
```

Out[13]:
```
Iraq                   24636
Pakistan               14368
Afghanistan            12731
India                  11960
Colombia                8306
                       ...  
International              1
Wallis and Futuna          1
South Vietnam              1
Andorra                    1
Antigua and Barbuda        1
Name: Country, Length: 205, dtype: int64
```

In [14]:
```
data.describe(include=["object", "bool"])
```

Out[14]:

	Country	Region	City	AttackType	Target	Group	Target_type	Weapon_type
count	181691	181691	181257	181691	181055	181691	181691	181691
unique	205	12	36674	9	86006	3537	22	12
top	Iraq	Middle East & North Africa	Unknown	Bombing/Explosion	Civilians	Unknown	Private Citizens & Property	Explosives
freq	24636	50474	9775	88255	6461	82782	43511	92426

In [15]:
```
data[data["AttackType"] == 1].mean()
```

Out[15]:
```
Year           NaN
Month          NaN
Day            NaN
Country        NaN
Region         NaN
City           NaN
latitude       NaN
longitude      NaN
AttackType     NaN
Killed         NaN
Wounded        NaN
Casualities    NaN
Target         NaN
Group          NaN
Target_type    NaN
Weapon_type    NaN
dtype: object
```

In [16]:
```
data[data["Target"]==1].mean()
```

Out[16]:
```
Year           NaN
Month          NaN
Day            NaN
Country        NaN
Region         NaN
City           NaN
latitude       NaN
longitude      NaN
AttackType     NaN
Killed         NaN
Wounded        NaN
Casualities    NaN
Target         NaN
Group          NaN
Target_type    NaN
Weapon_type    NaN
dtype: object
```

In [17]:
```
data.fillna(0).head()
```

Out[17]:

	Year	Month	Day	Country	Region	City	latitude	longitude	AttackType	Killed	Wounded	Casualities	Target	Group	Target_type	Weapon_type
0	1970	7	2	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	1.0	0.0	1.0	Julio Guzman	MANO-D	Private Citizens & Property	Unknown
1	1970	0	0	Mexico	North America	Mexico city	19.371887	-99.086624	Hostage Taking (Kidnapping)	0.0	0.0	0.0	Nadine Chaval, daughter	23rd of September Communist League	Government (Diplomatic)	Unknown
2	1970	1	0	Philippines	Southeast Asia	Unknown	15.478598	120.599741	Assassination	1.0	0.0	1.0	Employee	Unknown	Journalists & Media	Unknown
3	1970	1	0	Greece	Western Europe	Athens	37.997490	23.762728	Bombing/Explosion	0.0	0.0	0.0	U.S. Embassy	Unknown	Government (Diplomatic)	Explosives
4	1970	1	0	Japan	East Asia	Fukouka	33.580412	130.396361	Facility/Infrastructure Attack	0.0	0.0	0.0	U.S. Consulate	Unknown	Government (Diplomatic)	Incendiary

In [18]:
```
data.dropna(axis=0,inplace=True)
```

Ruinous features of Dataset

In [19]:
```
print("A Year with the highest number of attacks:",data['Year'].value_counts())
```

```
A Year with the highest number of attacks: 2014    14905
2015    13369
2016    12307
2013    11338
2017     9835
2012     8092
2011     4885
1992     4794
2010     4724
2009     4452
2008     4426
1991     4294
1989     4101
1990     3627
1988     3408
1994     3130
2007     2907
1995     2870
1987     2834
1996     2784
1997     2767
1984     2492
2006     2444
1986     2261
1985     2196
1979     2087
1980     2011
2005     1829
1981     1779
2001     1777
1983     1700
2000     1602
1982     1542
1978     1237
1999     1232
2002     1227
2003     1158
1977     1076
2004     1022
1998      761
1976      598
1970      569
1975      388
1974      344
1971      316
1973      225
1972      224
Name: Year, dtype: int64
```

In [20]:
```
print("A Month with the highest number of attacks:",data['Month'].value_counts())
```

```
A Month with the highest number of attacks: 5     14641
7     14425
8     13947
10    13730
6     13590
4     13327
11    13273
3     13230
1     13001
9     12590
2     12243
12    11942
0         7
Name: Month, dtype: int64
```

In [21]:
```
print("A Region with the highest number of attacks:",data['Region'].value_counts())
```

```
A Region with the highest number of attacks: Middle East & North Africa     45817
South Asia                     42261
South America                  16240
Sub-Saharan Africa             13833
Western Europe                 13816
Southeast Asia                 11226
Central America & Caribbean     7089
Eastern Europe                  4833
North America                   3283
East Asia                        744
Central Asia                     537
Australasia & Oceania            267
Name: Region, dtype: int64
```

In [22]:
```
print("A City with the highest number of attacks:",data['City'].value_counts())
```

```
A City with the highest number of attacks: Baghdad           7352
Unknown            6724
Karachi            2565
Mosul              2111
Lima               2108
                   ... 
Grasberg              1
Shaheed Anudand       1
Windesheim            1
Galcom                1
Kubentog              1
Name: City, Length: 32441, dtype: int64
```

In [23]:
```
print("A Country with the highest number of attacks:",data['Country'].value_counts())
```

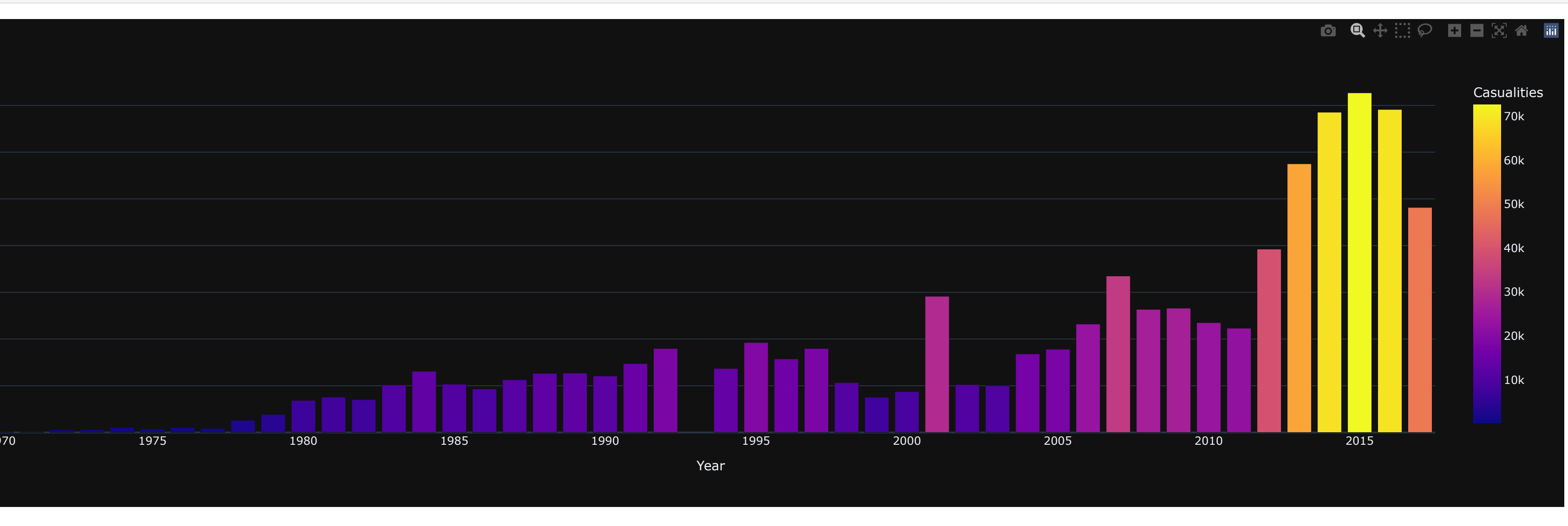
```
A Country with the highest number of attacks: Iraq                 23122
Pakistan               13798
Afghanistan            11716
India                  11293
Colombia                7262
                       ... 
St. Kitts and Nevis        1
St. Lucia                  1
South Vietnam              1
North Korea                1
Wallis and Futuna          1
Name: Country, Length: 202, dtype: int64
```

Visualizing the Data

An analysis of terrorists attacks by country through bar plots for each year

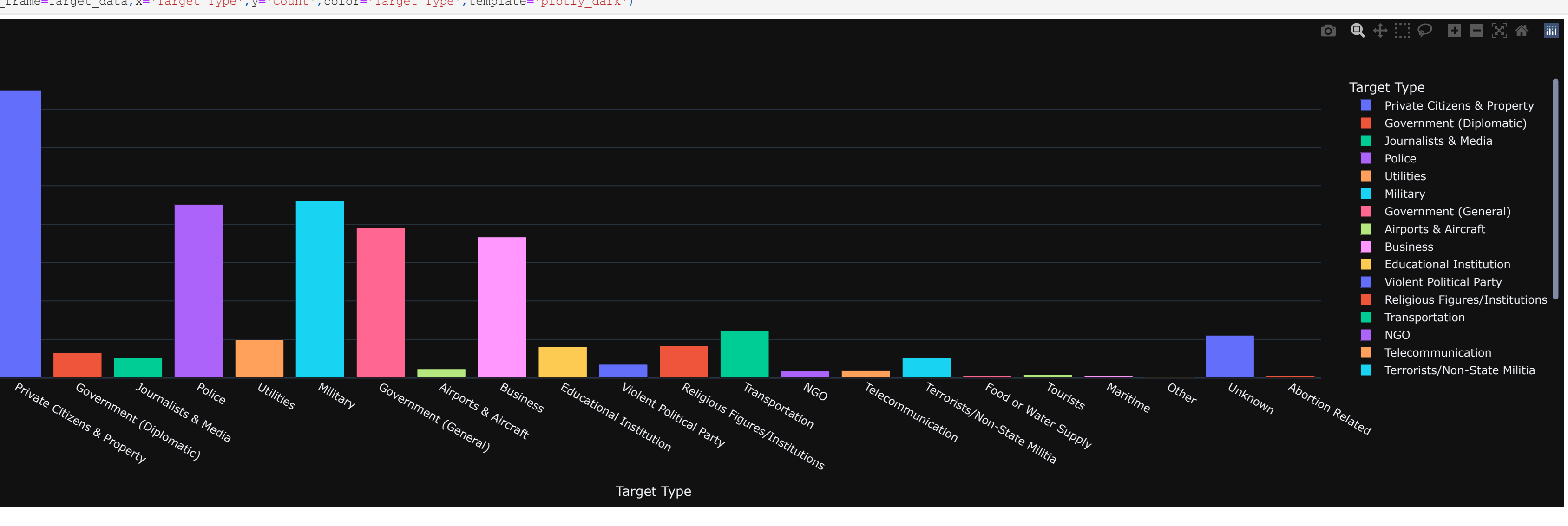
In [24]:
```
year = data.groupby('Year').Casualities.sum().to_frame().reset_index()
year.columns = ['Year','Casualities']
```

In [25]:
```
px.bar(data_frame=year,x="Year",y="Casualities",color="Casualities",template='plotly_dark')
```



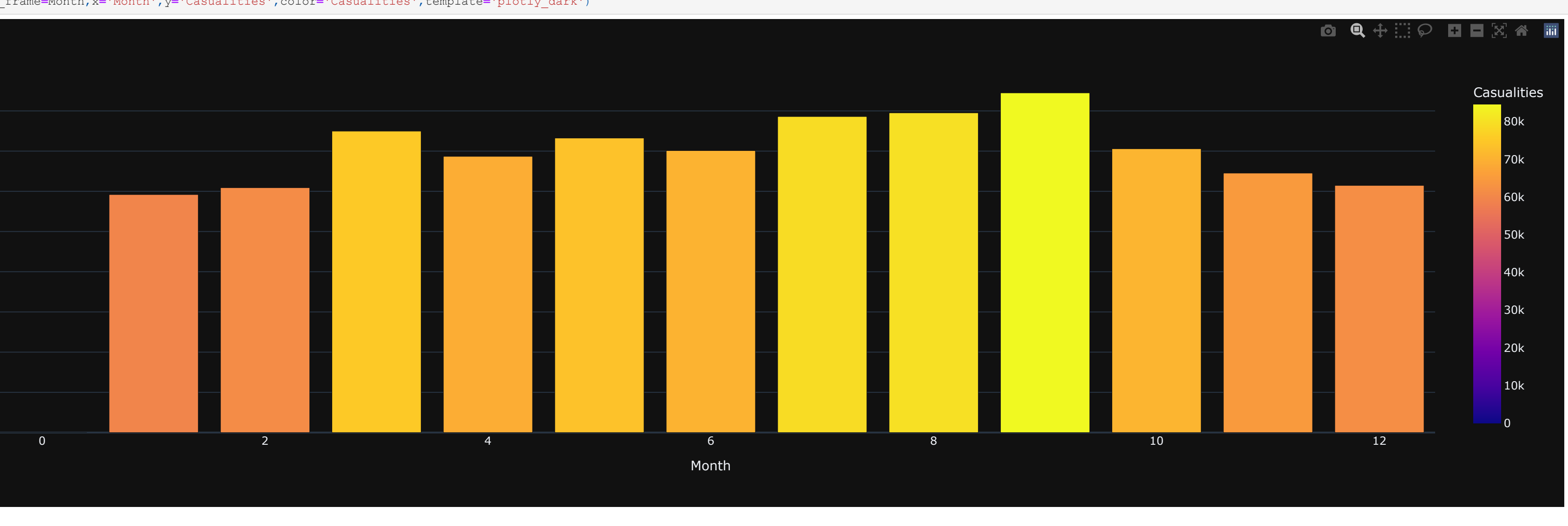
Targeted Attacks

In [26]:
```
Target = list(data['Target_type'])
Target_map = dict(Counter(Target))
Target_data = pd.DataFrame(Target_map.items())
Target_data.columns = ['Target Type','Count']
px.bar(data_frame=Target_data,x='Target Type',y='Count',color='Target Type',template='plotly_dark')
```



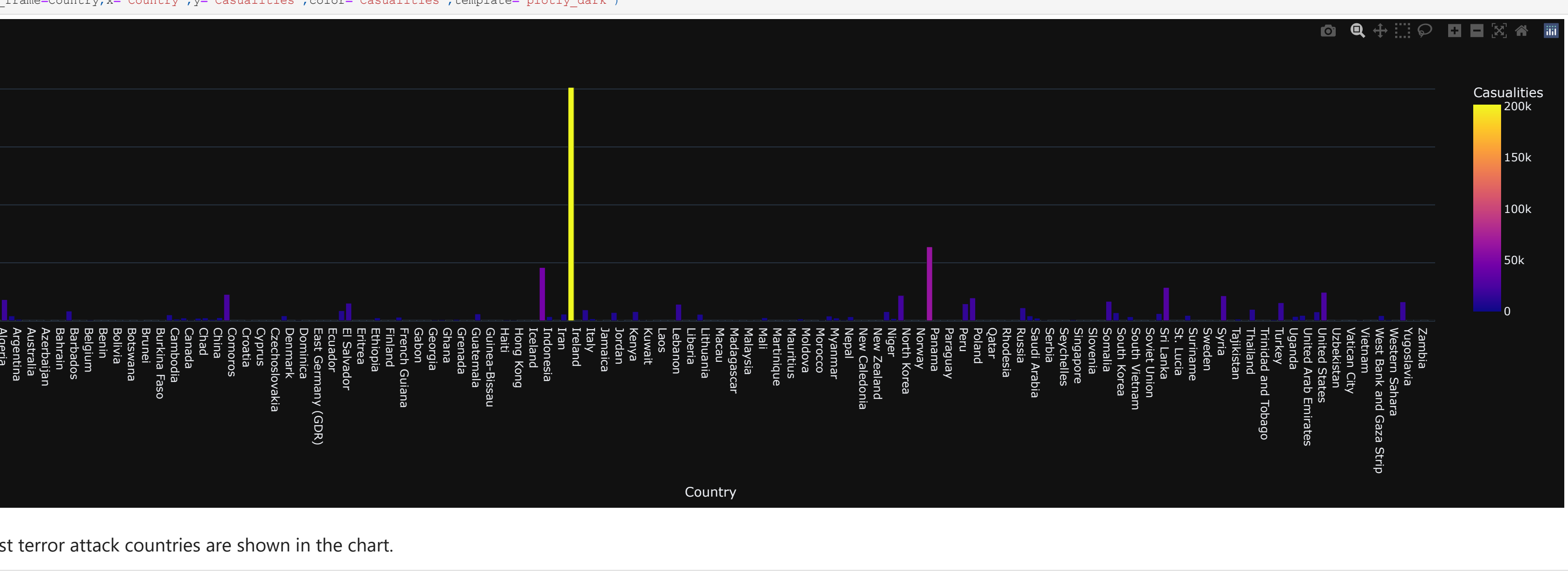
The number of terrorist attacks by month in each year is shown through a bar graph

In [27]:
```
Month= data.groupby('Month').Casualities.sum().to_frame().reset_index()
Month.columns=['Month','Casualities']
px.bar(data_frame=Month,x='Month',y='Casualities',color='Casualities',template='plotly_dark')
```



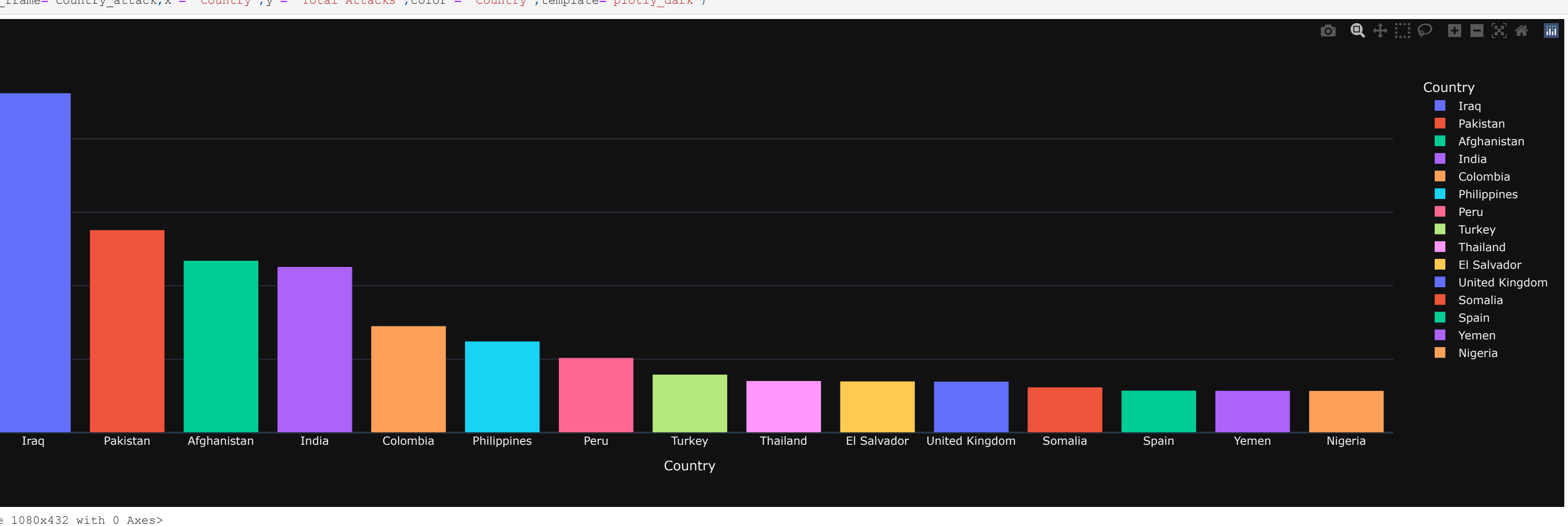
An analysis of terrorism attacks by country through bar plots for each year

In [28]:
```
Country = data.groupby('Country').Casualities.sum().to_frame().reset_index()
Country.columns=['Country','Casualities']
px.bar(data_frame=Country,x='Country',y='Casualities',color='Casualities',template='plotly_dark')
```



The highest terror attack countries are shown in the chart.

In [29]:
```
plt.figure(figsize=(15,6))
country_attack=data.Country.value_counts()[:15].reset_index()
country_attack.columns= ["Country", "Total Attacks"]
px.bar(data_frame= country_attack,x = 'Country',y = 'Total Attacks',color = 'Country',template='plotly_dark')
```



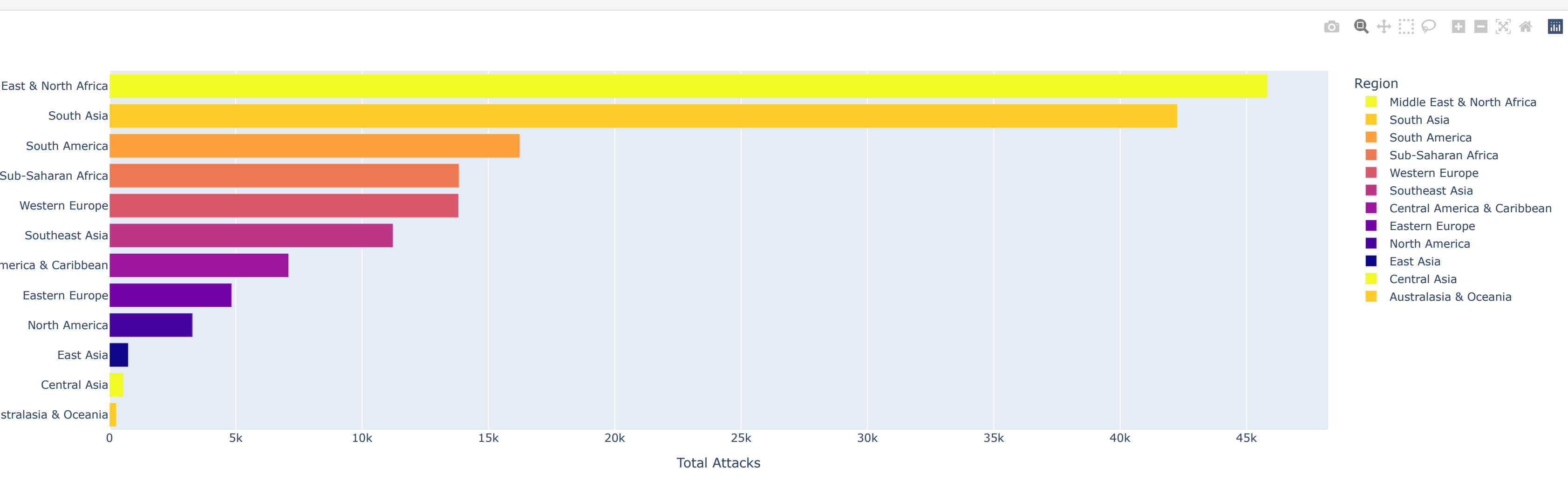

```
<Figure size 1080x432 with 0 Axes>
```

scrutiny:

Iraq, once again the highest, but this time followed by Afghanistan and Pakistan, had the highest death rate.

Region-wise Count of Terror Attacks

In [30]:
```
regional_attacks = data.Region.value_counts().to_frame().reset_index()
regional_attacks.columns = ['Region', 'Total Attacks']
fig = px.bar(data_frame=regional_attacks,x = 'Total Attacks',y='Region',color = 'Region',
             color_discrete_sequence= px.colors.sequential.Plasma_r)
fig.show()
```



Scrutiny:

Middle East and North Africa shows the highest followed by South Asia.