LetsGrowMore

Name : Varsha Reddy

Task - Exploratory Data Analysis on Dataset -Terrorism

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

Level : Intermediate Level

Batch : August 2022

Importing the libraries for Data Manipulation and Data Visualization

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

In [2]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:
```
import plotly.offline as py
import plotly.graph_objs as go
import plotly.express as px
from collections import Counter
import math
import warnings
warnings.filterwarnings('ignore')
```

Importing the Dataset

In [4]:
```
df = pd.read_csv("globalterrorismdb_0718dist.csv",encoding = "ISO-8859-1")
```

Preprocessing the Data

In [5]:
```
df.head()
```

Out[5]:

	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
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 [6]:
```
df.columns
```

Out[6]:
```
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)
```

In [7]:
```
df.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','city':'City','country_txt':'Country','provstate':'State','region_txt':'Region','attacktype1_txt':'AttackType','target1':'Target','nkill':'Killed','nwound':'Wounded',
df=df[['Year','Month','Day','City','Country','State','Region','AttackType','Target','Killed','Wounded','Summary','Group','Target_type','Weapon_type','Motive']]
df.head()
```

Out[7]:

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

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

Out[8]:
```
Year                0
Month               0
Day                 0
City              434
Country             0
State             421
Region              0
AttackType          0
Target            636
Killed          10313
Wounded         16311
Summary         66129
Group               0
Target_type         0
Weapon_type         0
Motive         131130
dtype: int64
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 16 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   City         181257 non-null  object
 4   Country      181691 non-null  object
 5   State        181270 non-null  object
 6   Region       181691 non-null  object
 7   AttackType   181691 non-null  object
 8   Target       181055 non-null  object
 9   Killed       171378 non-null  float64
 10  Wounded      165380 non-null  float64
 11  Summary      115562 non-null  object
 12  Group        181691 non-null  object
 13  Target_type  181691 non-null  object
 14  Weapon_type  181691 non-null  object
 15  Motive       50561 non-null   object
dtypes: float64(2), int64(3), object(11)
memory usage: 22.2+ MB
```

In [10]:
```
print(df.duplicated().sum())
```

```
9579
```

In [11]:
```
df.Country.value_counts()[:10]
```

Out[11]:
```
Iraq              24636
Pakistan          14368
Afghanistan       12731
India             11960
Colombia           8306
Philippines        6908
Peru               6096
El Salvador        5320
United Kingdom     5235
Turkey             4292
Name: Country, dtype: int64
```

In [12]:
```
df.Target.value_counts()
```

Out[12]:
```
Civilians                     6461
Unknown                       5918
Soldiers                      3157
Patrol                        2942
Checkpoint                    2905
                              ...
Manager Sanat Raj                1
Military/ Police patrol          1
Thor 67 Outpost                  1
Employee's Vehicle               1
Hmeymim Air Base                 1
Name: Target, Length: 86006, dtype: int64
```

In [13]:
```
df.Target.value_counts(normalize=True)
```

Out[13]:
```
Civilians                     0.035685
Unknown                       0.032686
Soldiers                      0.017437
Patrol                        0.016249
Checkpoint                    0.016045
                                ...
Manager Sanat Raj             0.000006
Military/ Police patrol       0.000006
Thor 67 Outpost               0.000006
Employee's Vehicle            0.000006
Hmeymim Air Base              0.000006
Name: Target, Length: 86006, dtype: float64
```

In [14]:
```
df['Casualities'] = df.Killed +df.Wounded
```

In [15]:
```
df.describe(include=["object", "bool"])
```

Out[15]:

	City	Country	State	Region	AttackType	Target	Summary	Group	Target_type	Weapon_type	Motive
count	181257	181691	181270	181691	181691	181055	115562	181691	181691	181691	50561
unique	36674	205	2855	12	9	86006	112492	3537	22	12	14490
top	Unknown	Iraq	Baghdad	Middle East & North Africa	Bombing/Explosion	Civilians	09/00/2016: Sometime between September 18, 201...	Unknown	Private Citizens & Property	Explosives	Unknown
freq	9775	24636	7645	50474	88255	6461	100	82782	43511	92426	14889

In [16]:
```
df.sort_values(by="City", ascending=False).head()
```

Out[16]:

	Year	Month	Day	City	Country	State	Region	AttackType	Target	Killed	Wounded	Summary	Group	Target_type	Weapon_type	Motive	Casualities
44393	1990	10	24	zinarag	Philippines	Cagayan	Southeast Asia	Armed Assault	truck	0.0	0.0	NaN	New People's Army (NPA)	Private Citizens & Property	Incendiary	NaN	0.0
42281	1990	4	25	yokohama	Japan	Kanagawa	East Asia	Facility/Infrastructure Attack	residence, japan airport bldg. company *	0.0	0.0	NaN	Chukakuha (Middle Core Faction)	Business	Incendiary	NaN	0.0
19135	1983	6	23	yacan	Peru	Pasco	South America	Assassination	pablo cardenas, justic of peace	2.0	0.0	NaN	Shining Path (SL)	Government (General)	Firearms	NaN	2.0
41289	1990	1	24	wakunai	Papua New Guinea	Bougainville	Australasia & Oceania	Armed Assault	brihan norman islander aircraft	0.0	0.0	NaN	Bougainville Revolutionary Army (BRA)	Airports & Aircraft	Incendiary	NaN	0.0
19055	1983	6	13	vinchos	Peru	Ayacucho	South America	Assassination	constantino supdo, justice of peace	1.0	0.0	NaN	Shining Path (SL)	Government (General)	Firearms	NaN	1.0

In [17]:
```
df[df["AttackType"] == 1].mean()
```

Out[17]:
```
Year          NaN
Month         NaN
Day           NaN
City          NaN
Country       NaN
State         NaN
Region        NaN
AttackType    NaN
Target        NaN
Killed        NaN
Wounded       NaN
Summary       NaN
Group         NaN
Target_type   NaN
Weapon_type   NaN
Motive        NaN
Casualities   NaN
dtype: object
```

In [18]:
```
df[df["Target_type"]==1].mean()
```

Out[18]:
```
Year          NaN
Month         NaN
Day           NaN
City          NaN
Country       NaN
State         NaN
Region        NaN
AttackType    NaN
Target        NaN
Killed        NaN
Wounded       NaN
Summary       NaN
Group         NaN
Target_type   NaN
Weapon_type   NaN
Motive        NaN
Casualities   NaN
dtype: object
```

In [19]:
```
df.fillna(0).head()
```

Out[19]:

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

In [20]:
```
df.dropna(axis=0,inplace=True)
```

Ruinous features of Dataset

In [21]:
```
print("Year with the most attacks:",df['Year'].value_counts().idxmax())
```

```
Year with the most attacks: 2011
```

In [22]:
```
print("Month with the most attacks:",df['Month'].value_counts().idxmax())
```

```
Month with the most attack: 7
```

In [23]:
```
print("Group with the most attacks:",df['Group'].value_counts().index[1])
```

```
Group with the most attack: Taliban
```

In [24]:
```
print("Region with the most attacks:",df['Region'].value_counts().idxmax())
```

```
Region with the most attacks: South Asia
```

In [25]:
```
print("City with the most attacks:",df['City'].value_counts().index[1])
```

```
City with the most attacks: Unknown
```

In [26]:
```
print("Country with the most attacks:",df['Country'].value_counts().idxmax())
```

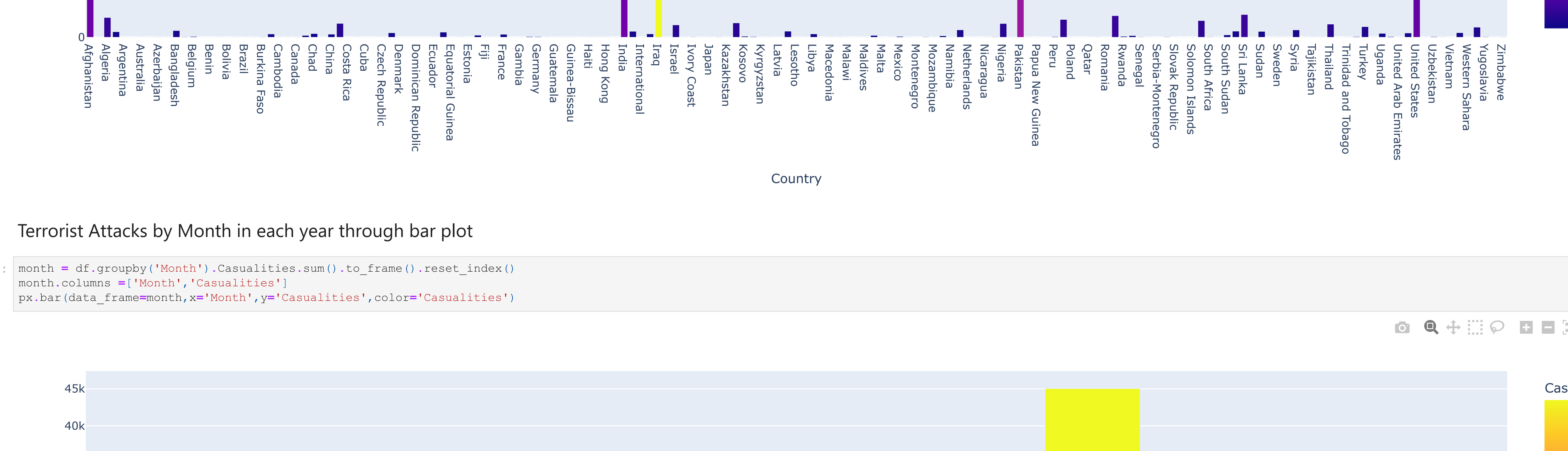
```
Country with the most attacks: Iraq
```

In [27]:
```
print("Most Attack Types:",df['AttackType'].value_counts().idxmax())
```

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

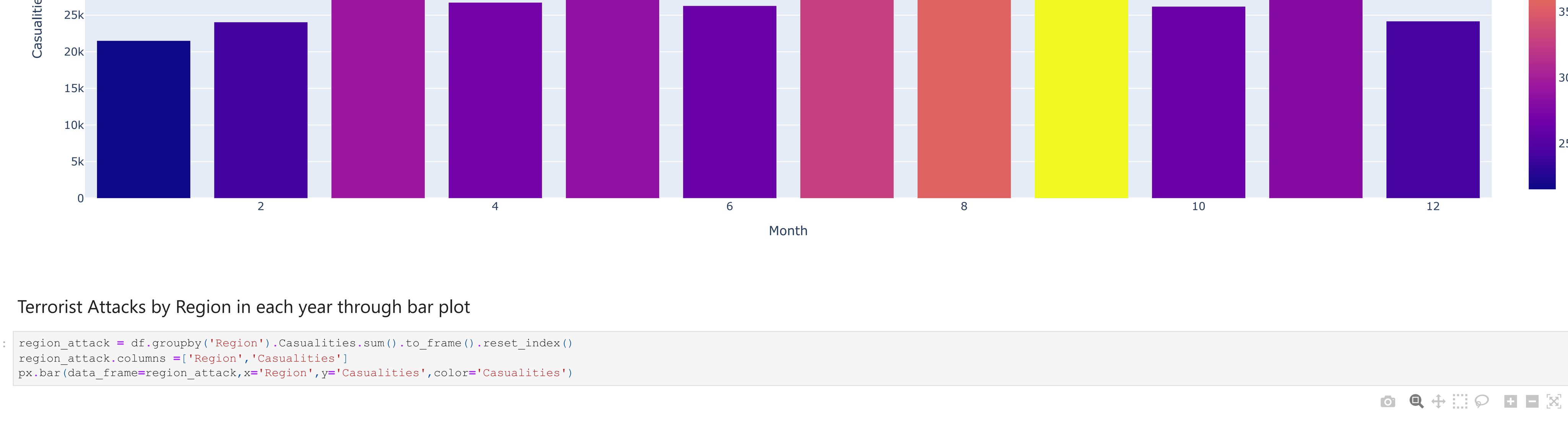
Terrorist Attacks by Country in each year through bar plot

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



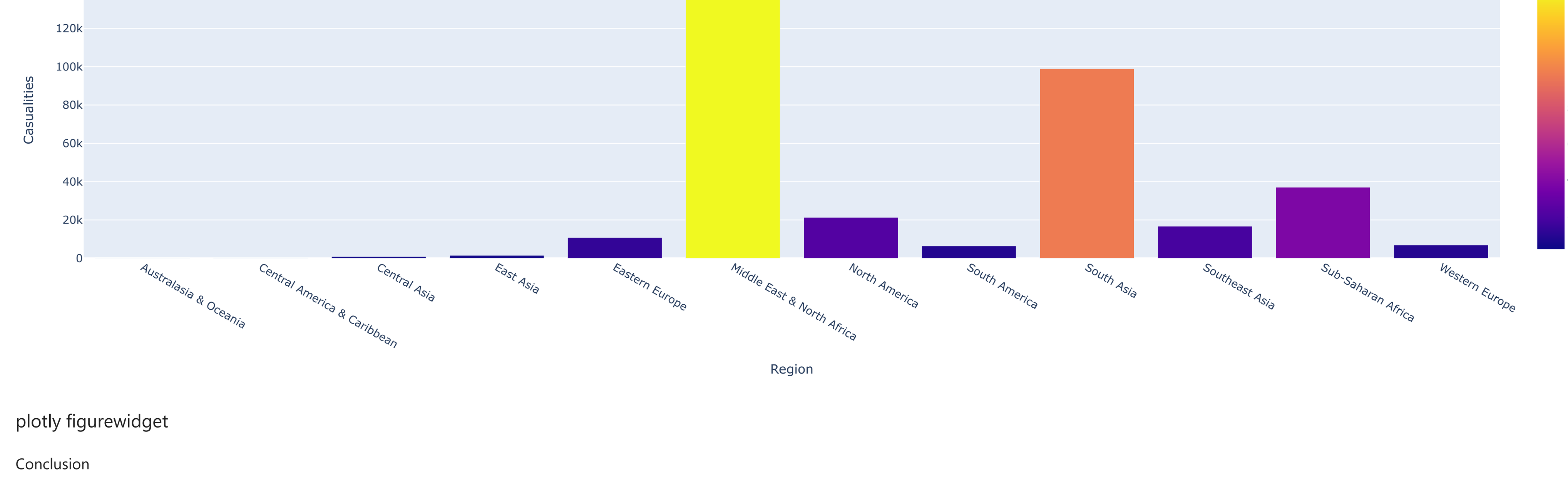
Terrorist Attacks by Month in each year through bar plot

In [29]:
```
month = df.groupby('Month').Casualities.sum().to_frame().reset_index()
month.columns =['Month','Casualities']
px.bar(data_frame=month,x='Month',y='Casualities',color='Casualities')
```



Terrorist Attacks by Region in each year through bar plot

In [30]:
```
region_attack = df.groupby('Region').Casualities.sum().to_frame().reset_index()
region_attack.columns =['Region','Casualities']
px.bar(data_frame=region_attack,x='Region',y='Casualities',color='Casualities')
```



plotly figurewidget

Conclusion

In [31]:
```
f = go.FigureWidget()
f
```

```
FigureWidget({
    'data': [], 'layout': {'template': '...'}
})
```

In [32]:
```
f.add_scatter(y=['Year','Month','Day','Killed','Wounded']);
f.add_bar(y=['Region','State','City','Country','Target'])
f.layout.title = 'Terrorist Attacks '
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: