

LetsGrowMore Nov2021 internship

Task2 :Exploratory Data Analysis on Dataset - Terrorism

Level :INTERMEDIATE LEVEL Language : Python Environment :jupyter notebook

```
In [1]: #import libraries
```

```
In [2]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

```
In [3]: #load the data
```

```
In [4]: data=pd.read_csv("C:/Users/sri ram/Desktop/letsgrowmore/globalterrorism.csv")
```

C:\Users\sri ram\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: 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.

```
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

```
In [5]: print(data)
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	\
0	1970000000001	1970	7	2	NaN	0	NaN	
1	1970000000002	1970	0	0	NaN	0	NaN	
2	1970010000001	1970	1	0	NaN	0	NaN	
3	1970010000002	1970	1	0	NaN	0	NaN	
4	1970010000003	1970	1	0	NaN	0	NaN	
...	
181686	201712310022	2017	12	31	NaN	0	NaN	
181687	201712310029	2017	12	31	NaN	0	NaN	
181688	201712310030	2017	12	31	NaN	0	NaN	
181689	201712310031	2017	12	31	NaN	0	NaN	
181690	201712310032	2017	12	31	NaN	0	NaN	

	country	country_txt	region	...	addnotes	\
0	58	Dominican Republic	2	...	NaN	
1	130	Mexico	1	...	NaN	
2	160	Philippines	5	...	NaN	
3	78	Greece	8	...	NaN	
4	101	Japan	4	...	NaN	
...	
181686	182	Somalia	11	...	NaN	
181687	200	Syria	10	...	NaN	
181688	160	Philippines	5	...	NaN	
181689	92	India	6	...	NaN	
181690	160	Philippines	5	...	NaN	

	scite1	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
181686	"Somalia: Al-Shabaab Militants Attack Army Che...	
181687	"Putin's 'victory' in Syria has turned into a ...	
181688	"Maguindanao clashes trap tribe members," Phil...	
181689	"Trader escapes grenade attack in Imphal," Bus...	
181690	"Security tightened in Cotabato following IED ...	

	scite2	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
181686	"Highlights: Somalia Daily Media Highlights 2 ...	

```

181687 "Two Russian soldiers killed at Hmeymim base i...
181688                                     NaN
181689                                     NaN
181690 "Security tightened in Cotabato City," Manila ...

```

```

                                scite3 \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
...                              ...
181686 "Highlights: Somalia Daily Media Highlights 1 ...
181687 "Two Russian servicemen killed in Syria mortar...
181688                                     NaN
181689                                     NaN
181690                                     NaN

```

```

                                dbsource INT_LOG INT_IDEO INT_MISC INT_ANY related
0                                PGIS      0      0      0      0      0      NaN
1                                PGIS      0      1      1      1      1      NaN
2                                PGIS     -9     -9      1      1      1      NaN
3                                PGIS     -9     -9      1      1      1      NaN
4                                PGIS     -9     -9      1      1      1      NaN
...                              ...
181686 START Primary Collection      0      0      0      0      0      NaN
181687 START Primary Collection     -9     -9      1      1      1      NaN
181688 START Primary Collection      0      0      0      0      0      NaN
181689 START Primary Collection     -9     -9      0     -9     -9      NaN
181690 START Primary Collection     -9     -9      0     -9     -9      NaN

```

[181691 rows x 135 columns]

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

```

Out[6]:
   eventid  iyear  imonth  iday  approxdate  extended  resolution  country  country_txt  region  ...  addnotes  scite1  scite2  scite3  dbso
0  1970000000001  1970      7      2      NaN         0      NaN      58  Dominican Republic  2  ...      NaN      NaN      NaN      NaN  f
1  1970000000002  1970      0      0      NaN         0      NaN     130    Mexico          1  ...      NaN      NaN      NaN      NaN  f
2  1970010000001  1970      1      0      NaN         0      NaN     160  Philippines      5  ...      NaN      NaN      NaN      NaN  f
3  1970010000002  1970      1      0      NaN         0      NaN      78    Greece          8  ...      NaN      NaN      NaN      NaN  f
4  1970010000003  1970      1      0      NaN         0      NaN     101    Japan           4  ...      NaN      NaN      NaN      NaN  f

```

5 rows x 135 columns

```
In [7]: #rename the columns for the given the data
```

```
In [8]: data.rename(columns={'iyear':'year','imonth':'month','iday':'day','country_txt':'Country','region_txt':'Region','
```

```
In [9]: data.head()
```

```

Out[9]:
   eventid  year  month  day  approxdate  extended  resolution  country  Country  region  ...  addnotes  scite1  scite2  scite3  dbsour
0  1970000000001  1970      7      2      NaN         0      NaN      58  Dominican Republic  2  ...      NaN      NaN      NaN      NaN  PG
1  1970000000002  1970      0      0      NaN         0      NaN     130    Mexico          1  ...      NaN      NaN      NaN      NaN  PG
2  1970010000001  1970      1      0      NaN         0      NaN     160  Philippines      5  ...      NaN      NaN      NaN      NaN  PG
3  1970010000002  1970      1      0      NaN         0      NaN      78    Greece          8  ...      NaN      NaN      NaN      NaN  PG
4  1970010000003  1970      1      0      NaN         0      NaN     101    Japan           4  ...      NaN      NaN      NaN      NaN  PG

```

5 rows x 135 columns

```
In [10]: data.columns
```

```

Out[10]: Index(['eventid', 'year', 'month', 'day', 'approxdate', 'extended',
                'resolution', 'country', 'Country', 'region',
                ...

```

```
'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG',
'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'],
dtype='object', length=135)
```

```
In [11]: data.tail()
```

Out[11]:

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1	scite2
181686	201712310022	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab Militants Attack Army Che...	"Highlights: Somalia Daily Media Highlights 2 ...
181687	201712310029	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...
181688	201712310030	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...	NaN
181689	201712310031	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...	NaN
181690	201712310032	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...	"Security tightened in Cotabato City," Manila ...

5 rows × 135 columns

```
In [12]: data.size
```

Out[12]: 24528285

```
In [13]: data.shape
```

Out[13]: (181691, 135)

```
In [14]: data.describe()
```

Out[14]:

	eventid	year	month	day	extended	country	region	latitude	longitude
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.968501	7.160938	23.498343	-4.586957e+02
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.414535	2.933408	18.569242	2.047790e+05
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-53.154613	-8.618590e+07
25%	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.000000	5.000000	11.510046	4.545640e+00
50%	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.000000	6.000000	31.467463	4.324651e+01
75%	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.000000	10.000000	34.685087	6.871033e+01
max	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.000000	12.000000	74.633553	1.793667e+02

8 rows × 77 columns

data cleaning process(checking the data is NaN & Unknown)

```
In [15]: data.isnull()

Out[15]:
```

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1	scite2	scite3	dbsource
0	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
1	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
2	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
3	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
4	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
...
181686	False	False	False	False	True	False	True	False	False	False	...	True	False	False	False	False
181687	False	False	False	False	True	False	True	False	False	False	...	True	False	False	False	False
181688	False	False	False	False	True	False	True	False	False	False	...	True	False	True	True	False
181689	False	False	False	False	True	False	True	False	False	False	...	True	False	True	True	False
181690	False	False	False	False	True	False	True	False	False	False	...	True	False	False	True	False

181691 rows × 135 columns

```
In [16]: data.isnull().sum

Out[16]:
```

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1	scite2	scite3	dbsource
0	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
1	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
2	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
3	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
4	False	False	False	False	True	False	True	False	False	False	...	True	True	True	True	False
...
181686	False	False	False	False	True	False	True	False	False	False	...	True	False	False	False	False
181687	False	False	False	False	True	False	True	False	False	False	...	True	False	False	False	False
181688	False	False	False	False	True	False	True	False	False	False	...	True	False	True	True	False
181689	False	False	False	False	True	False	True	False	False	False	...	True	False	True	True	False
181690	False	False	False	False	True	False	True	False	False	False	...	True	False	False	True	False

	country	Country	region	...	addnotes	scite1	scite2	scite3	dbsource
0	False	False	False	...	True	True	True	True	False
1	False	False	False	...	True	True	True	True	False
2	False	False	False	...	True	True	True	True	False
3	False	False	False	...	True	True	True	True	False
4	False	False	False	...	True	True	True	True	False
...
181686	False	False	False	...	True	False	False	False	False
181687	False	False	False	...	True	False	False	False	False
181688	False	False	False	...	True	False	True	True	False
181689	False	False	False	...	True	False	True	True	False
181690	False	False	False	...	True	False	False	True	False

	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
0	False	False	False	False	False	True
1	False	False	False	False	False	True
2	False	False	False	False	False	True
3	False	False	False	False	False	True
4	False	False	False	False	False	True
...
181686	False	False	False	False	False	True
181687	False	False	False	False	False	True
181688	False	False	False	False	False	True
181689	False	False	False	False	False	True
181690	False	False	False	False	False	True

[181691 rows x 135 columns]>

```
In [20]: data['year'].value_counts()[:20]

Out[20]:
```

2014	16903
2015	14965
2016	13587
2013	12036
2017	10900
2012	8522
2011	5076

1992 5071
2010 4826
2008 4805
2009 4721
1991 4683
1989 4324
1990 3887
1988 3721
1984 3495
1994 3456
2007 3242
1997 3197
1987 3183
Name: year, dtype: int64

```
In [21]: data.fillna(0)
```

Out[21]:

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1	scite2
0	197000000001	1970	7	2	0	0	0	58	Dominican Republic	2	...	0	0	0
1	197000000002	1970	0	0	0	0	0	130	Mexico	1	...	0	0	0
2	197001000001	1970	1	0	0	0	0	160	Philippines	5	...	0	0	0
3	197001000002	1970	1	0	0	0	0	78	Greece	8	...	0	0	0
4	197001000003	1970	1	0	0	0	0	101	Japan	4	...	0	0	0
...
181686	201712310022	2017	12	31	0	0	0	182	Somalia	11	...	0	"Somalia: Al-Shabaab Militants Attack Army Che...	"Highlights: Somalia Daily Media Highlights 2 ...
181687	201712310029	2017	12	31	0	0	0	200	Syria	10	...	0	"Putin's 'victory' in Syria has turned into a ...	"Two Russian soldiers killed at Hmeymim base i...
181688	201712310030	2017	12	31	0	0	0	160	Philippines	5	...	0	"Maguindanao clashes trap tribe members," Phil...	0
181689	201712310031	2017	12	31	0	0	0	92	India	6	...	0	"Trader escapes grenade attack in Imphal," Bus...	0
181690	201712310032	2017	12	31	0	0	0	160	Philippines	5	...	0	"Security tightened in Cotabato following IED ...	"Security tightened in Cotabato City," Manila ...

181691 rows × 135 columns

```
In [24]: data.isnull().sum()
```

Out[24]:

eventid	0
year	0
month	0
day	0
approxdate	172452
...	...
INT_LOG	0
INT_IDEO	0
INT_MISC	0
INT_ANY	0
related	156653

Length: 135, dtype: int64

```
In [25]: data.isnull().sum(axis=1)
```

```
Out[25]: 0      90
          1      83
          2      88
          3      88
          4      90
          ..
181686    71
181687    68
181688    73
181689    76
181690    78
Length: 181691, dtype: int64
```

```
In [29]: data.isna().sum()
```

```
Out[29]: eventid      0
year      0
month     0
day       0
approxdate 172452
...
INT_IDEO    0
INT_MISC    0
INT_ANY     0
related    156653
Casualties  0
Length: 136, dtype: int64
```

```
In [30]: data.fillna(0,inplace=True)
```

```
In [31]: data['wounded']=data['wounded'].fillna(0)
data['killed']=data['killed'].fillna(0)
data['Casualties']=data['killed']+data['wounded']
data.head()
```

```
Out[31]:
```

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	scite1	scite2	scite3	dbsource	INT_LO
0	1970000000001	1970	7	2	0	0	0	58	Dominican Republic	2	...	0	0	0	PGIS	
1	1970000000002	1970	0	0	0	0	0	130	Mexico	1	...	0	0	0	PGIS	
2	1970010000001	1970	1	0	0	0	0	160	Philippines	5	...	0	0	0	PGIS	
3	1970010000002	1970	1	0	0	0	0	78	Greece	8	...	0	0	0	PGIS	
4	1970010000003	1970	1	0	0	0	0	101	Japan	4	...	0	0	0	PGIS	

5 rows × 136 columns

```
In [32]: data['Country'].value_counts()[:10]
```

```
Out[32]: 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
```

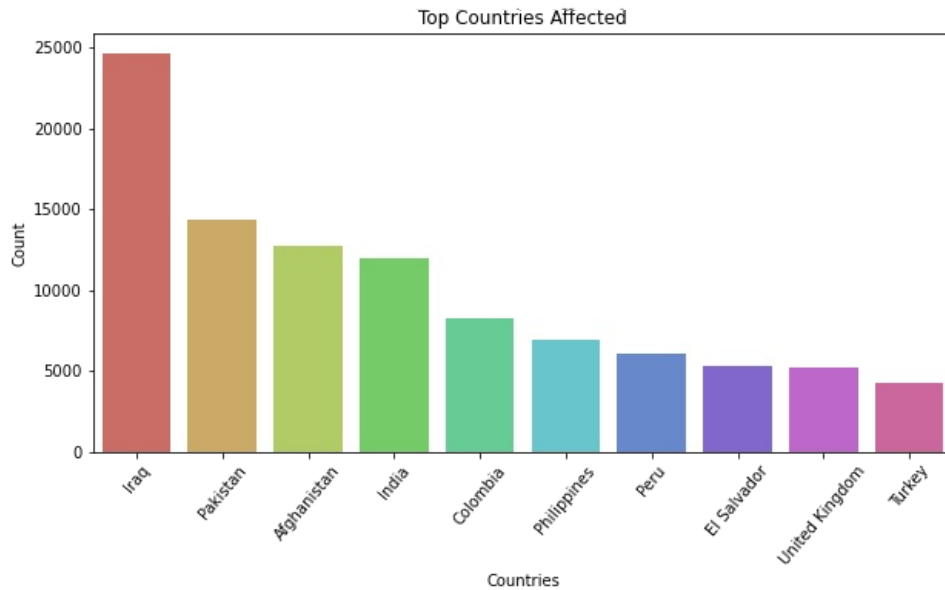
after cleaning data

observing the data how the countries are effected by terrorism

```
In [34]: plt.subplots(figsize=(10,5))
sns.barplot(data['Country'].value_counts()[:10].index,data['Country'].value_counts()[:10].values,palette="hls")
plt.title("Top Countries Affected")
plt.xlabel("Countries")
plt.ylabel("Count")
plt.xticks(rotation=50)
plt.show()
```

C:\Users\sri ram\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

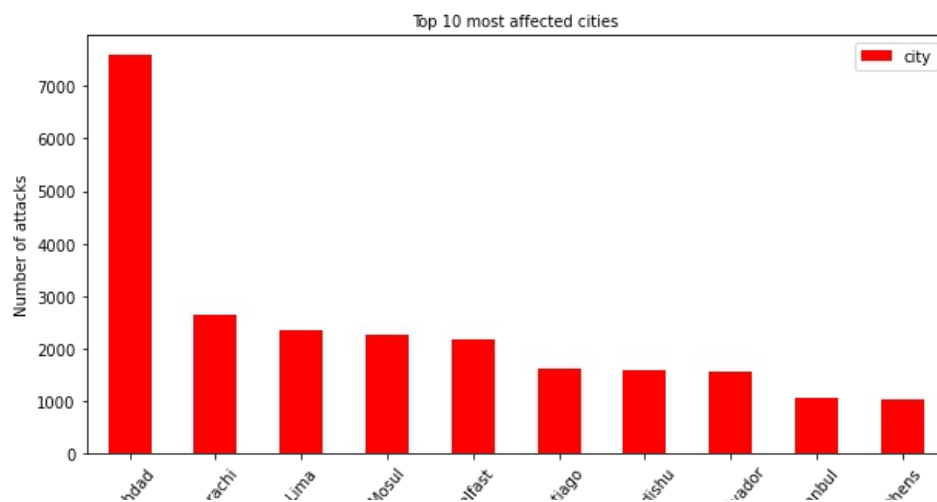
warnings.warn(



```
In [35]: data['Group_name'].value_counts()[:10]
```

```
Out[35]: Unknown                        82782
Taliban                               7478
Islamic State of Iraq and the Levant (ISIL)  5613
Shining Path (SL)                     4555
Farabundo Marti National Liberation Front (FMLN)  3351
Al-Shabaab                           3288
New People's Army (NPA)                2772
Irish Republican Army (IRA)            2671
Revolutionary Armed Forces of Colombia (FARC)  2487
Boko Haram                           2418
Name: Group_name, dtype: int64
```

```
In [36]: data.city.value_counts().to_frame().sort_values('city',axis=0,ascending=False).drop("Unknown").head(10).plot(kind='bar')
plt.xticks(rotation=50)
plt.xlabel("City")
plt.ylabel("Number of attacks")
plt.title("Top 10 most affected cities",fontsize=10)
plt.show()
```



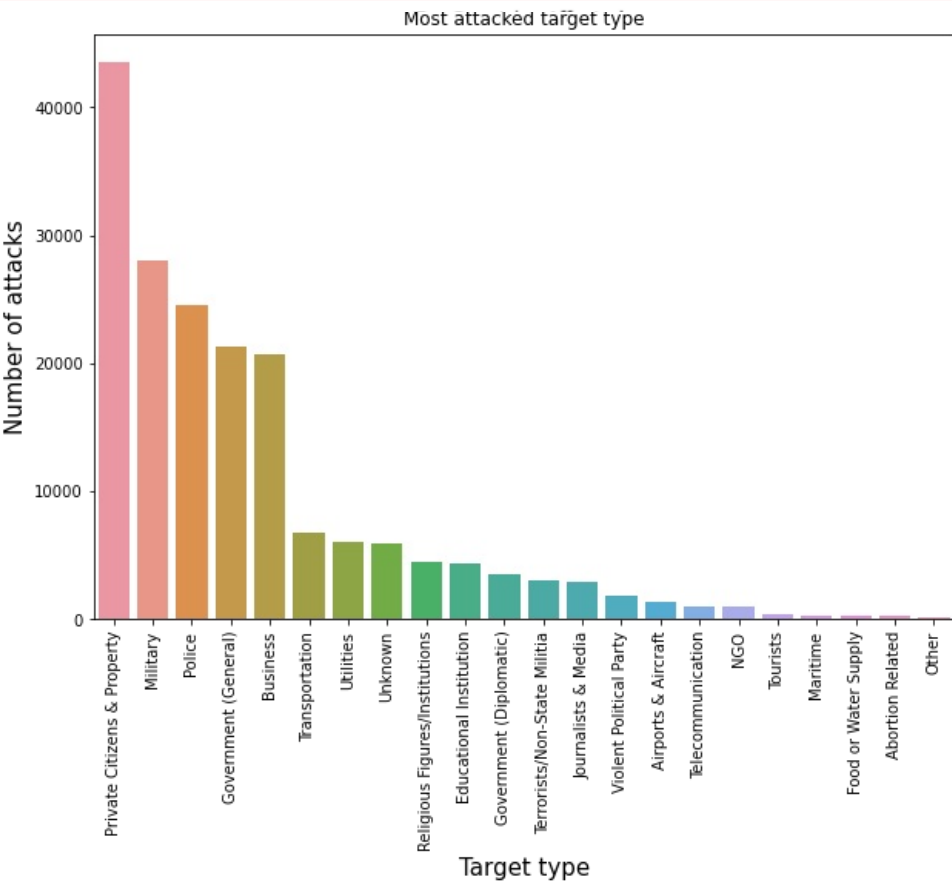
Bagh. Ka. A. Be. San. Mogau. San Salu. Ista. A.

City

```
In [38]: plt.subplots(figsize=(10,7))
sns.countplot(data.Target_type,order=data["Target_type"].value_counts().index);
plt.title("Most attacked target type")
plt.xticks(rotation=90)
plt.xlabel("Target type",fontsize=15)
plt.ylabel("Number of attacks",fontsize=15)
plt.show()
```

C:\Users\sri ram\anaconda3\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(



```
In [39]: data.head()
```

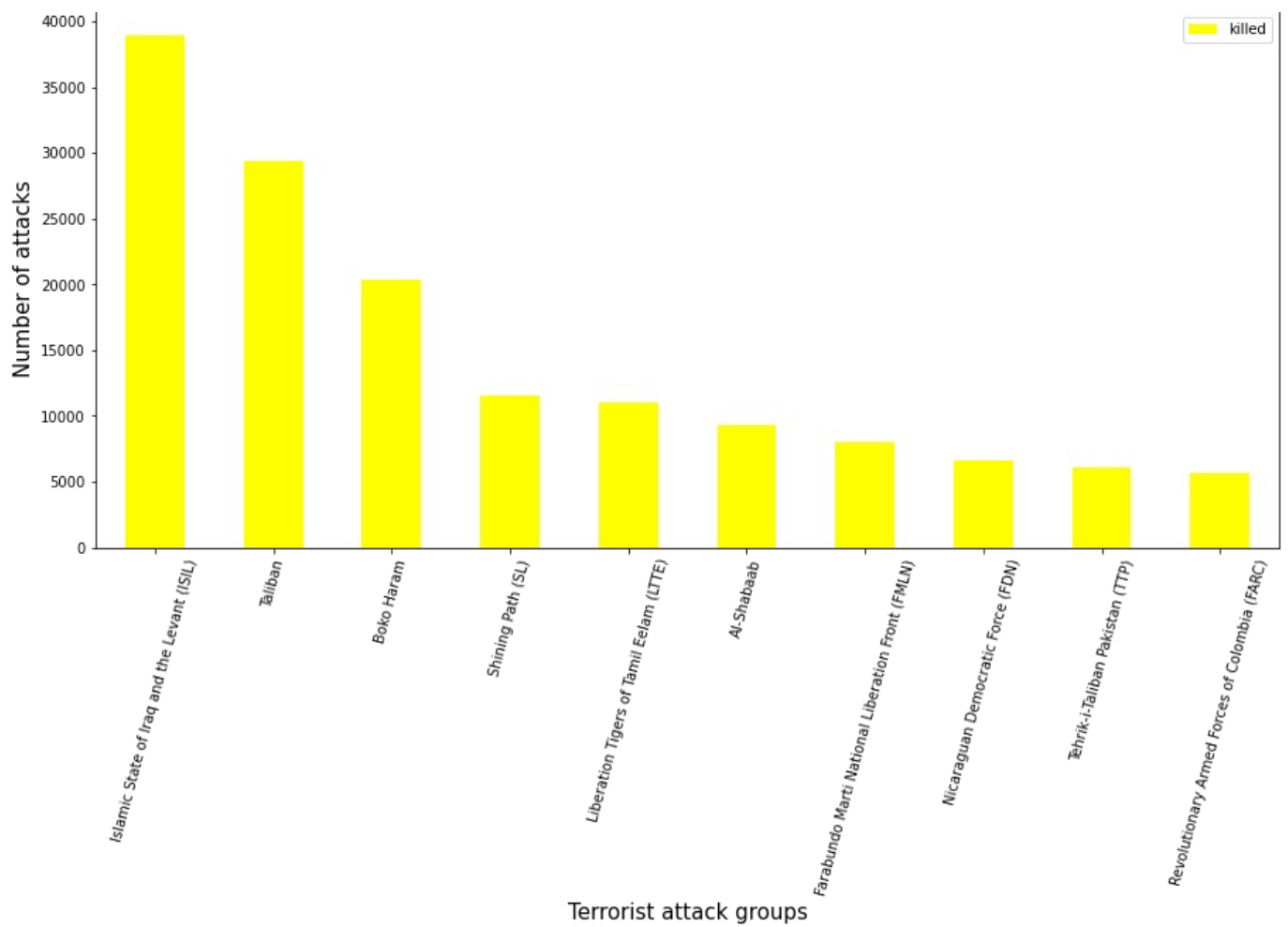
Out[39]:

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	scite1	scite2	scite3	dbsource	INT_LO
0	1970000000001	1970	7	2	0	0	0	58	Dominican Republic	2	...	0	0	0	PGIS	
1	1970000000002	1970	0	0	0	0	0	130	Mexico	1	...	0	0	0	PGIS	
2	1970010000001	1970	1	0	0	0	0	160	Philippines	5	...	0	0	0	PGIS	
3	1970010000002	1970	1	0	0	0	0	78	Greece	8	...	0	0	0	PGIS	
4	1970010000003	1970	1	0	0	0	0	101	Japan	4	...	0	0	0	PGIS	

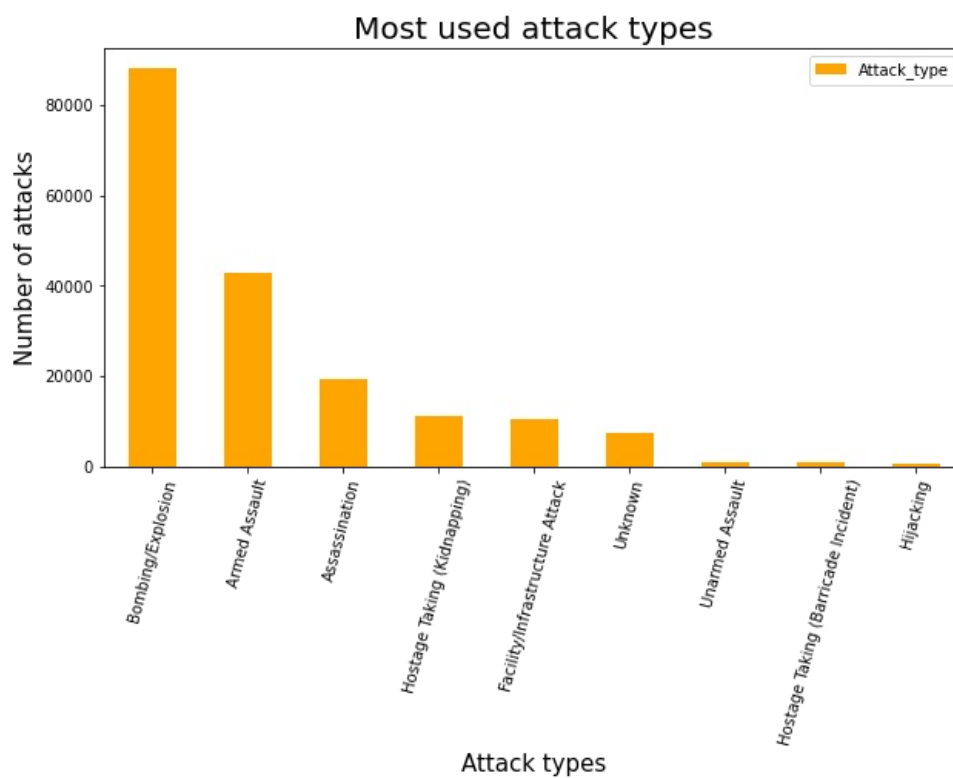
5 rows × 136 columns

```
In [53]: data[['Group_name', 'killed']].groupby(['Group_name'],axis=0).sum().drop('Unknown').sort_values('killed',ascending=True)
plt.title("Top 10 terrorist groups",fontsize=15)
plt.xlabel("Terrorist attack groups",fontsize=15)
plt.ylabel("Number of attacks",fontsize=15)
plt.xticks(rotation=75)
plt.show()
```

Top 10 terrorist groups



```
In [48]: data["Attack_type"].value_counts().to_frame().sort_values('Attack_type',ascending=False).plot(kind="bar",color="c")
plt.xticks(rotation=75)
plt.xlabel("Attack types",fontsize=15)
plt.ylabel("Number of attacks",fontsize=15)
plt.title("Most used attack types",fontsize=20)
plt.show()
```

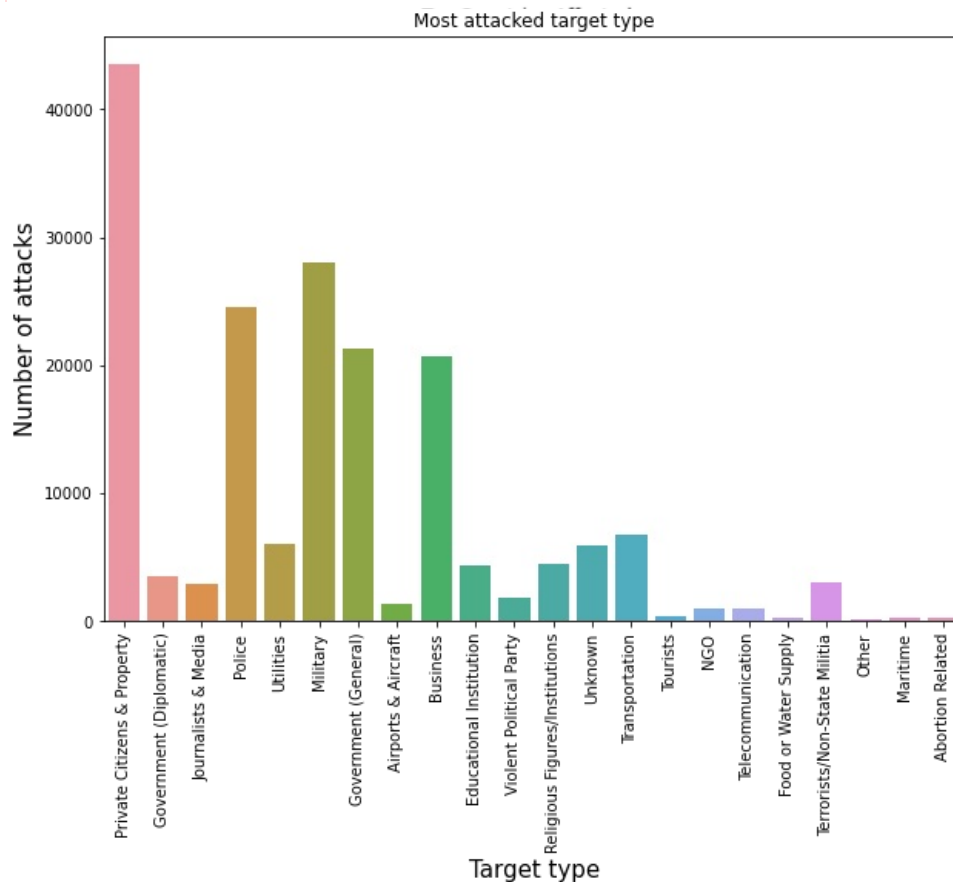


```
In [45]: plt.subplots(figsize=(10,7))
sns.countplot(data["Target_type"]);
plt.title("Most attacked target type")
```

```
plt.xticks(rotation=90)
plt.xlabel("Target type",fontsize=15)
plt.ylabel("Number of attacks",fontsize=15)
plt.show()
```

C:\Users\sri ram\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

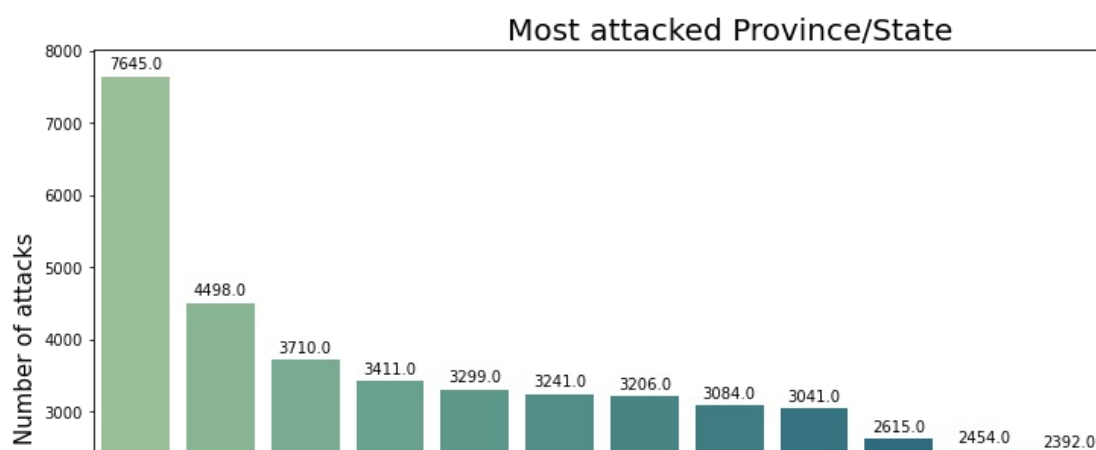


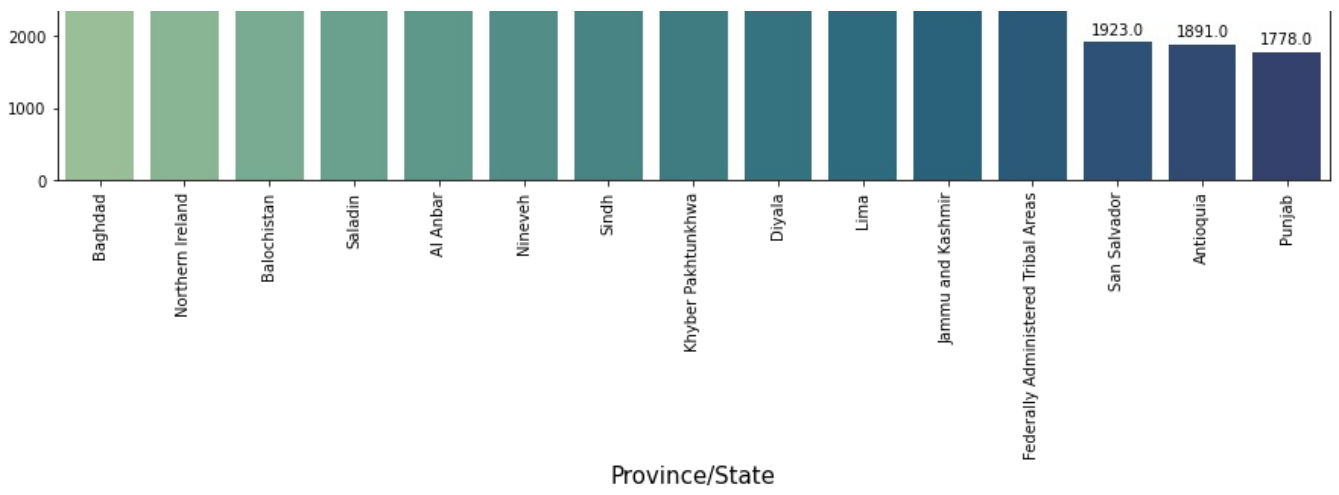
In [54]:

```
dd=data[(data["Province/State"]!='Unknown')]
ds=dd["Province/State"].dropna(axis=0).value_counts()[1:15]
plt.subplots(figsize=(15,7))
g=sns.barplot(ds.index,ds.values,palette="crest")
for val in g.patches:
    g.annotate(format(val.get_height()),(val.get_x() + val.get_width()/2, val.get_height()),
               ha='center',size=10,va='center',textcoords='offset points',xytext=(0, 8))
plt.title("Most attacked Province/State",fontsize=20)
plt.xlabel("Province/State",fontsize=15)
plt.ylabel("Number of attacks",fontsize=15)
plt.xticks(rotation=90,fontsize=10)
plt.show()
```

C:\Users\sri ram\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(





Hot Zones Most attacked country: Iraq Most attacked Province/State: Baghdad Most attacked city: Baghdad Most used attack type: Bombing/Explosion Most attacked target type: Private Citizens & Property Terrorist group with most attacks: Islamic State of Iraq and the Levant

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js