

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/Ajay rai/content/globalterrorismdb_0718dist.csv")
```

```
C:\Users\Ajay raj\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	197000000001	1970	7	2	NaN	0	NaN
1	197000000002	1970	0	0	NaN	0	NaN
2	197001000001	1970	1	0	NaN	0	NaN
3	197001000002	1970	1	0	NaN	0	NaN
4	197001000003	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	scite1	scite2
0	58	Dominican Republic	2	...	NaN	NaN	NaN
1	130	Mexico	1	...	NaN	NaN	NaN
2	160	Philippines	5	...	NaN	NaN	NaN
3	78	Greece	8	...	NaN	NaN	NaN
4	101	Japan	4	...	NaN	NaN	NaN
..	..	..	..	..	..	..	..
181686	182	Somalia	11	...	NaN	NaN	NaN
181687	200	Syria	10	...	NaN	NaN	NaN
181688	160	Philippines	5	...	NaN	NaN	NaN
181689	92	India	6	...	NaN	NaN	NaN
181690	160	Philippines	5	...	NaN	NaN	NaN
..	..	..	..	..	..	..	..
	scite1	scite2	scite3	scite4	scite5	scite6	scite7
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	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 ...	...	...	...	...	...	...

```

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      NaN
1          PGIS      0        1        1        1      NaN
2          PGIS     -9       -9        1        1      NaN
3          PGIS     -9       -9        1        1      NaN
4          PGIS     -9       -9        1        1      NaN
...
181686    ...      ...      ...      ...      ...
181687    START Primary Collection      0        0        0        0      NaN
181688    START Primary Collection     -9       -9        1        1      NaN
181689    START Primary Collection      0        0        0        0      NaN
181690    START Primary Collection     -9       -9        0       -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  197000000001  1970      7      2      NaN        0        NaN      58  Dominican Republic      2  ...  NaN  NaN  NaN  NaN
1  197000000002  1970      0      0      NaN        0        NaN     130  Mexico      1  ...  NaN  NaN  NaN  NaN
2  197001000001  1970      1      0      NaN        0        NaN     160  Philippines      5  ...  NaN  NaN  NaN  NaN
3  197001000002  1970      1      0      NaN        0        NaN      78  Greece      8  ...  NaN  NaN  NaN  NaN
4  197001000003  1970      1      0      NaN        0        NaN     101  Japan      4  ...  NaN  NaN  NaN  NaN

```

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','dbsource':'dbsource'})`

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

```

Out[9]:   eventid  year  month  day  approxdate  extended  resolution  country  Country  region  ...  addnotes  scite1  scite2  scite3  dbsource
0  197000000001  1970      7      2      NaN        0        NaN      58  Dominican Republic      2  ...  NaN  NaN  NaN  NaN  PG
1  197000000002  1970      0      0      NaN        0        NaN     130  Mexico      1  ...  NaN  NaN  NaN  NaN  PG
2  197001000001  1970      1      0      NaN        0        NaN     160  Philippines      5  ...  NaN  NaN  NaN  NaN  PG
3  197001000002  1970      1      0      NaN        0        NaN      78  Greece      8  ...  NaN  NaN  NaN  NaN  PG
4  197001000003  1970      1      0      NaN        0        NaN     101  Japan      4  ...  NaN  NaN  NaN  NaN  PG

```

5 rows x 135 columns

In [10]: `data.columns`





```

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 ...  "Tvc Russian soldiers killed at Hmeimim 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

```

```
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  197000000001  1970      7      2          0          0          0      58  Dominican Republic  2  ...      0      0      0      PGIS
1  197000000002  1970      0      0          0          0          0     130      Mexico  1  ...      0      0      0      PGIS
2  197001000001  1970      1      0          0          0          0     160  Philippines  5  ...      0      0      0      PGIS
3  197001000002  1970      1      0          0          0          0      78      Greece  8  ...      0      0      0      PGIS
4  197001000003  1970      1      0          0          0          0     101      Japan  4  ...      0      0      0      PGIS
```

5 rows × 136 columns

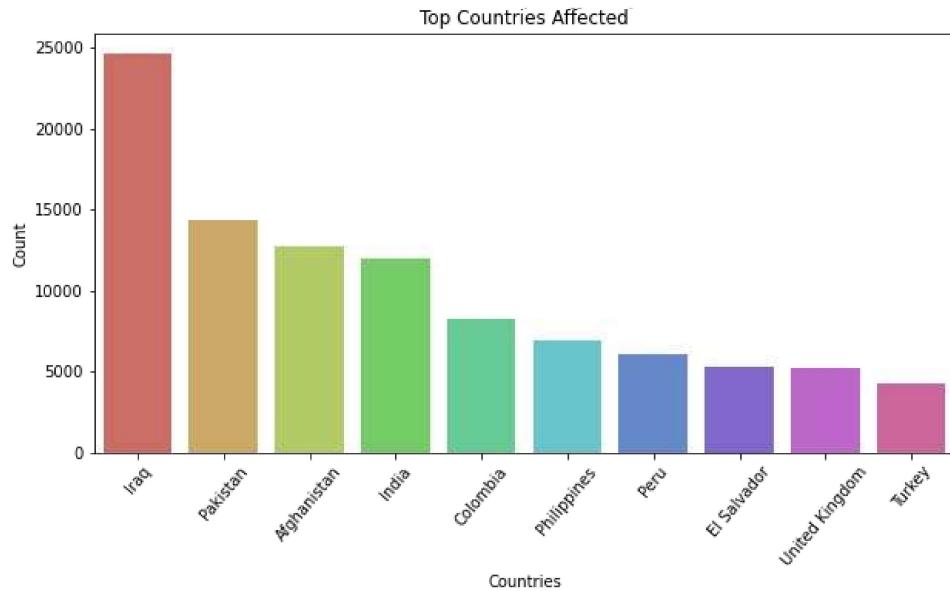
```
In [32]: data['Country'].value_counts()[:10]
```

```
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

```
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()
```

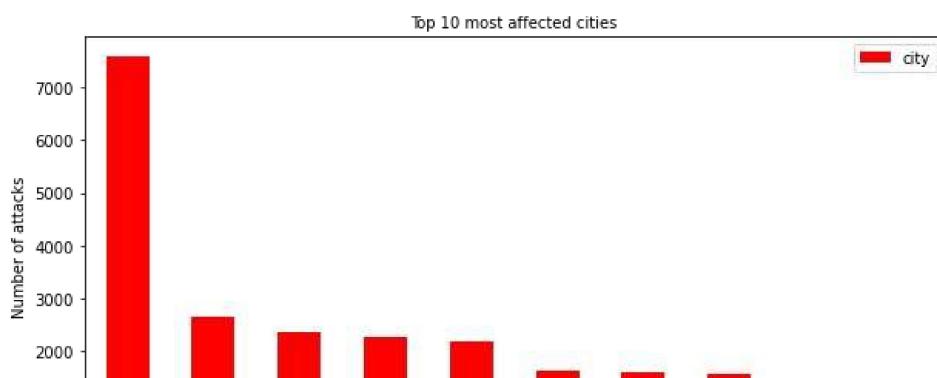
C:\Users\Ajay Raj\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.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



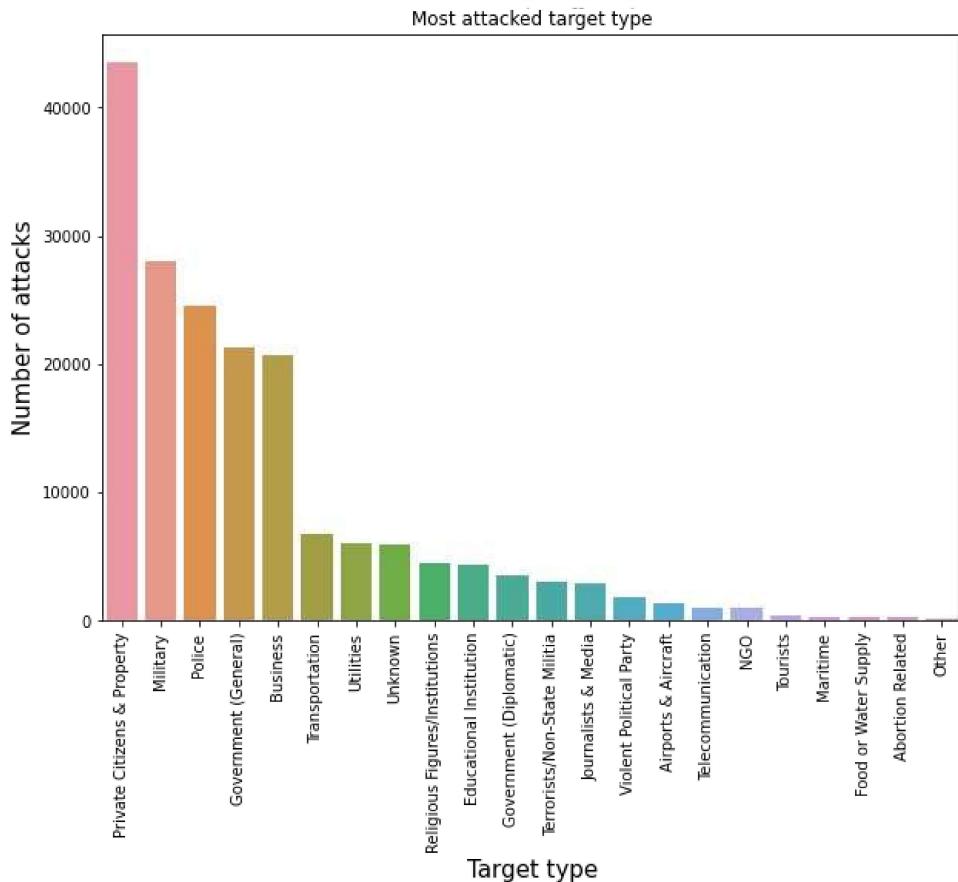
Bag K&S San Mogal San Salw Isgo

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\Ajay Raj\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()
```

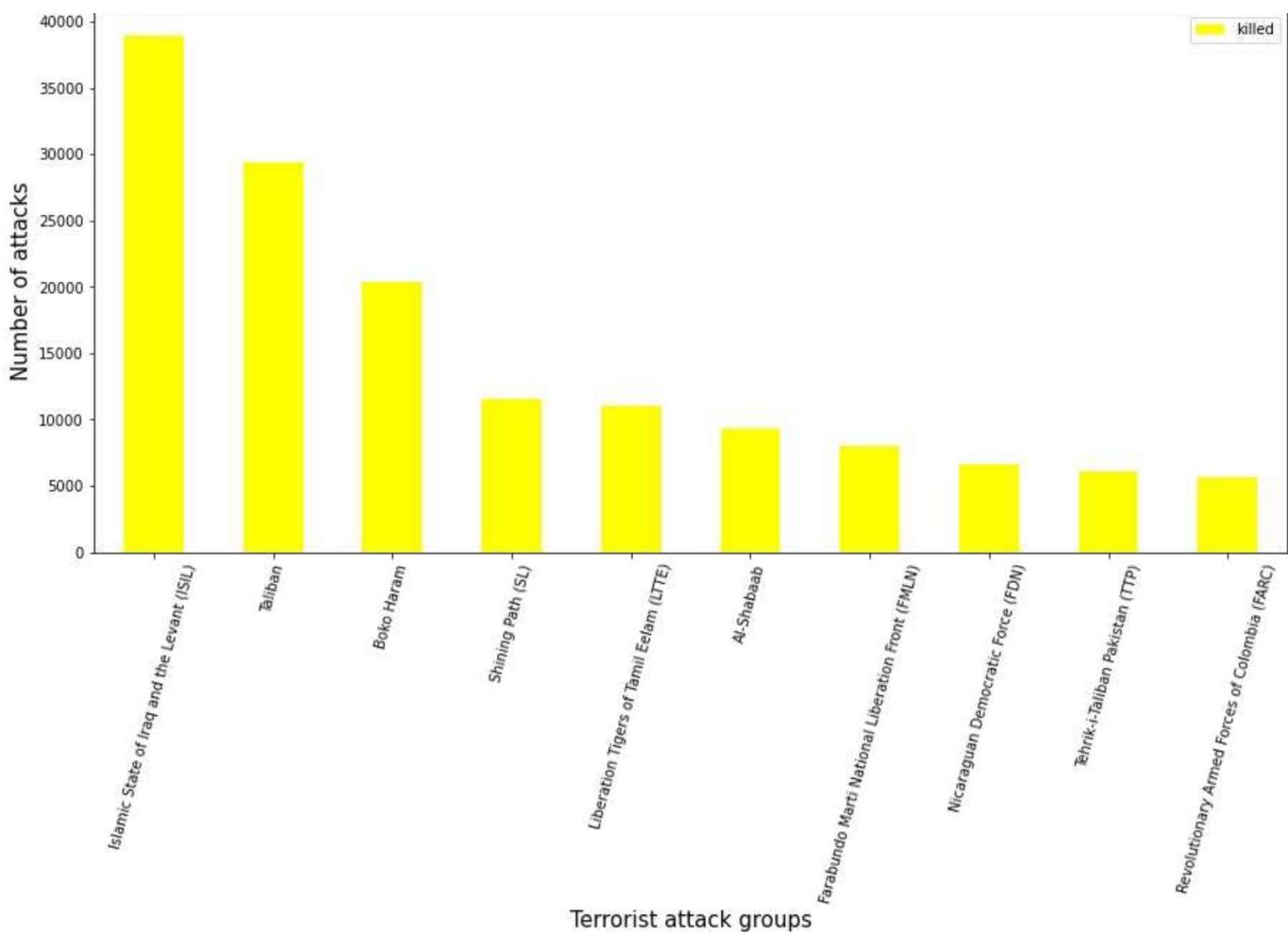
Out[39]:

	eventid	year	month	day	approxdate	extended	resolution	country	Country	region	...	scite1	scite2	scite3	dbsource	INT_LO
0	197000000001	1970	7	2	0	0	0	58	Dominican Republic	2	...	0	0	0	PGIS	.
1	197000000002	1970	0	0	0	0	0	130	Mexico	1	...	0	0	0	PGIS	.
2	197001000001	1970	1	0	0	0	0	160	Philippines	5	...	0	0	0	PGIS	.
3	197001000002	1970	1	0	0	0	0	78	Greece	8	...	0	0	0	PGIS	.
4	197001000003	1970	1	0	0	0	0	101	Japan	4	...	0	0	0	PGIS	.

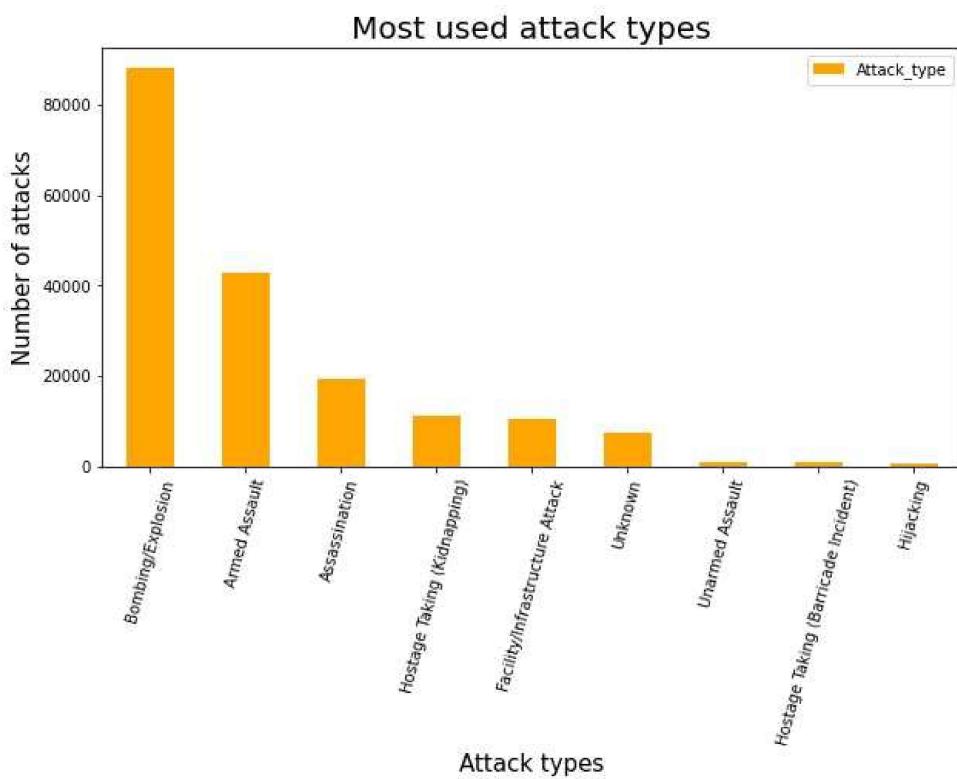
5 rows × 136 columns

In [53]:

```
data[['Group_name','killed']].groupby(['Group_name'],axis=0).sum().drop('Unknown').sort_values('killed',ascending=False)
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)
```



```
In [48]: data["Attack_type"].value_counts().to_frame().sort_values('Attack_type', ascending=False).plot(kind="bar", color="orange")
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()
```

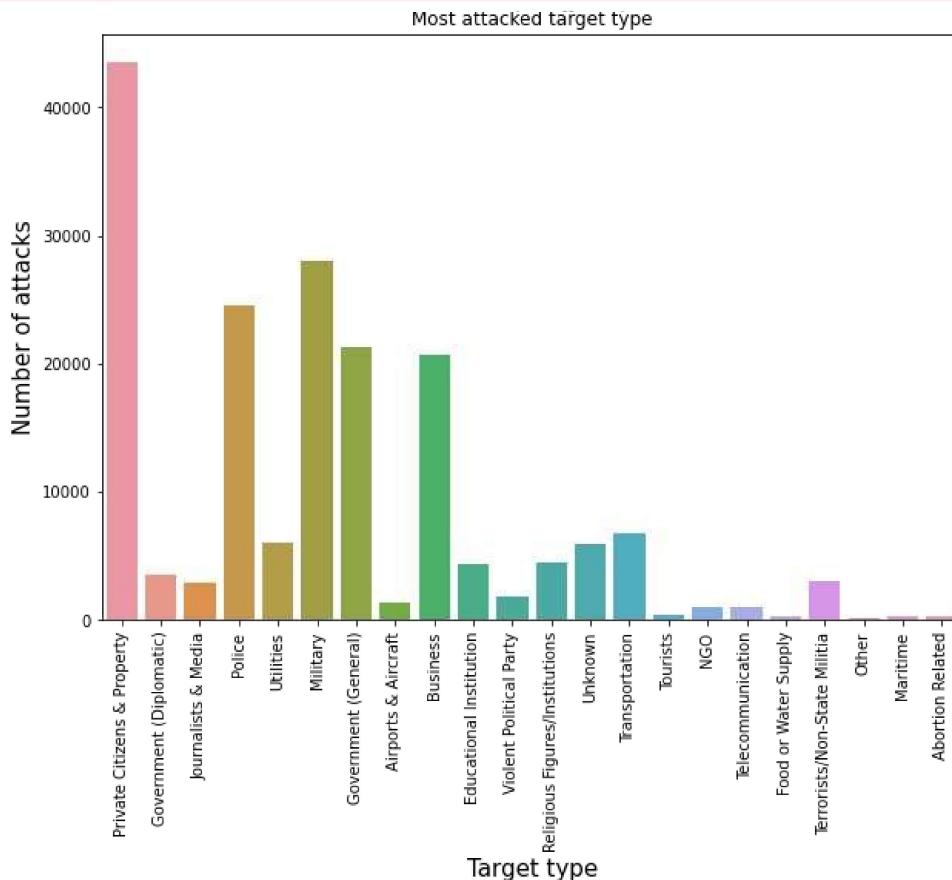


```

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

```

C:\Users\Ajay Raj\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(



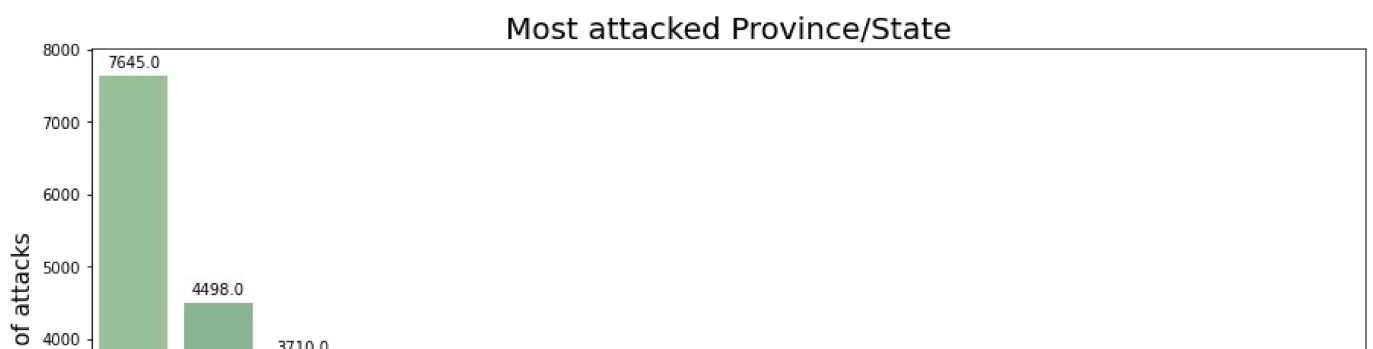
In [54]:

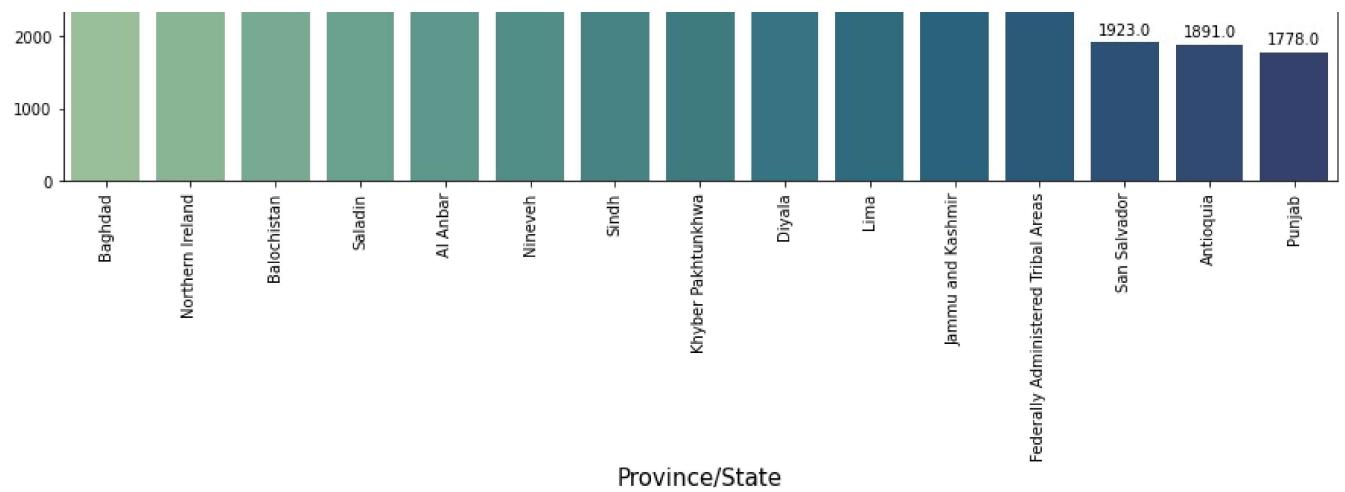
```

dd=data[(data["Province/State"]!="Unknown")]
ds=dd[["Province/State"]].dropna(axis=0).value_counts()[: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\Ajay Raj\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