

In [121...

```
Prajaktasawale
#LGMTask3
#Exploratory Data Analysis on Dataset terrorism
```

In [122...

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [123...

```
df=pd.read_csv(r"C:\Users\Prajakta\Downloads\globalterrorismdb_0718dist.csv",encoding='iso-8859-1')
```

In [124...

```
df.head()
```

Out[124...

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3
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 × 135 columns

In [125...

```
df.tail()
```

Out[125...

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3
181686	201712310022	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somalia: Al-Shabaab	"	"

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1
													Militants Attack Army Che...
181687	201712310029	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Putin's 'victory' in Syria has turned into a ...
181688	201712310030	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguindanao clashes trap tribe members," Phil...
181689	201712310031	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Trader escapes grenade attack in Imphal," Bus...
181690	201712310032	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Security tightened in Cotabato following IED ...

In [126...

```
df.describe()
```

Out[126...

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitud
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+C
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.968501	7.160938	23.498343	-4.586957e+C
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.414535	2.933408	18.569242	2.047790e+C
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-53.154613	-8.618590e+C

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitud
<b>25%</b>	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.000000	5.000000	11.510046	4.545640e+C
<b>50%</b>	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.000000	6.000000	31.467463	4.324651e+C
<b>75%</b>	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.000000	10.000000	34.685087	6.871033e+C
<b>max</b>	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.000000	12.000000	74.633553	1.793667e+C

In [127...

```
df.shape
```

Out[127...

```
(181691, 135)
```

In [128...

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(55), int64(22), object(58)
memory usage: 187.1+ MB
```

In [129...

```
df.isnull().sum() # total Null values
```

Out[129...

```
eventid      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 [130...

```
df.isnull().sum()/df.isnull().count()*100 # % of Null values
```

Out[130...

```
eventid      0.000000
```

```
iyear          0.000000
imonth         0.000000
iday           0.000000
approxdate     94.914993
...
INT_LOG        0.000000
INT_IDEO       0.000000
INT_MISC       0.000000
INT_ANY        0.000000
related        86.219461
```

In [131...

```
df.columns.values
```

Out[131...

```
array(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended',
      'resolution', 'country', 'country_txt', 'region', 'region_txt',
      'provstate', 'city', 'latitude', 'longitude', 'specificity',
      'vicinity', 'location', 'summary', 'crit1', 'crit2', 'crit3',
      'doubtterr', 'alternative', 'alternative_txt', 'multiple',
      'success', 'suicide', 'attacktype1', 'attacktype1_txt',
      'attacktype2', 'attacktype2_txt', 'attacktype3', 'attacktype3_txt',
      'targettype1', 'targettype1_txt', 'targetsubtype1', 'targetsubtype1_txt',
      'corp1', 'target1', 'natlty1', 'natlty1_txt', 'targettype2',
      'targettype2_txt', 'targetsubtype2', 'targetsubtype2_txt', 'corp2',
      'target2', 'natlty2', 'natlty2_txt', 'targettype3', 'targettype3_txt',
      'targetsubtype3', 'targetsubtype3_txt', 'corp3', 'target3', 'natlty3',
      'natlty3_txt', 'gname', 'gsubname', 'gname2', 'gsubname2',
      'gname3', 'gsubname3', 'motive', 'guncertain1', 'guncertain2',
      'guncertain3', 'individual', 'nperps', 'nperpcap', 'claimed',
      'claimmode', 'claimmode_txt', 'claim2', 'claimmode2',
      'claimmode2_txt', 'claim3', 'claimmode3', 'claimmode3_txt',
      'compclaim', 'weaptype1', 'weaptype1_txt', 'weapsubtype1',
      'weapsubtype1_txt', 'weaptype2', 'weaptype2_txt', 'weapsubtype2',
      'weapsubtype2_txt', 'weaptype3', 'weaptype3_txt', 'weapsubtype3',
      'weapsubtype3_txt', 'weaptype4', 'weaptype4_txt', 'weapsubtype4',
      'weapsubtype4_txt', ' weapdetail', 'nkill', 'nkillus', 'nkillter',
      'nwound', 'nwoundus', 'nwoundte', 'property', 'propextent',
      'propextent_txt', 'propvalue', 'propcomment', 'ishostkid',
      'nhostkid', 'nhostkidus', 'nhours', 'ndays', 'divert',
      'kidhijcountry', 'ransom', 'ransomamt', 'ransomamtus',
      'ransompaid', 'ransompaidus', 'ransomnote', 'hostkidoutcome',
      'hostkidoutcome_txt', 'nreleased', 'addnotes', 'scite1', 'scite2',
      'scite3', 'dbsource', 'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_ANY',
      'related'], dtype=object)
```

```
In [132... df.rename(columns={'iyear':'Year', 'imonth':'Month', 'iday':'Day', 'country_txt':'Country', 'region_txt':'Reg', 'latitude':'Latitude', 'longitude':'Longitude', 'attacktype1':'AttackType', 'nkill':'Killed', 'targettype1_txt':'TargetType', 'weaptype1_txt':'WeaponType', 'motive':'Motive'}, inplace=True)
```

```
In [133... df=df[['Year', 'Month', 'Day', 'Country', 'Region', 'State', 'City', 'Latitude', 'Longitude', 'AttackType', 'Killed', 'Wounded', 'Target', 'TargetType', 'Group', 'Summary', 'TargetType', 'WeaponType', 'Motive']]
```

```
In [134... df.describe()
```

	Year	Month	Day	Latitude	Longitude	AttackType	Killed	Wounded
count	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181691.000000	171378.000000	165380.000000
mean	2002.638997	6.467277	15.505644	23.498343	-4.586957e+02	3.247547	2.403272	3.167668
std	13.259430	3.388303	8.814045	18.569242	2.047790e+05	1.915772	11.545741	35.949392
min	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	1.000000	0.000000	0.000000
25%	1991.000000	4.000000	8.000000	11.510046	4.545640e+00	2.000000	0.000000	0.000000
50%	2009.000000	6.000000	15.000000	31.467463	4.324651e+01	3.000000	0.000000	0.000000
75%	2014.000000	9.000000	23.000000	34.685087	6.871033e+01	3.000000	2.000000	2.000000
max	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	9.000000	1570.000000	8191.000000

```
In [135... df.shape
```

Out[135... (181691, 19)

```
In [136... df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 19 columns):
#   Column          Non-Null Count  Dtype
---  -
#   Column          Non-Null Count  Dtype
```

```
0   Year      181691 non-null int64
1   Month     181691 non-null int64
2   Day       181691 non-null int64
3   Country   181691 non-null object
4   Region    181691 non-null object
5   State     181270 non-null object
6   City      181257 non-null object
7   Latitude  177135 non-null float64
8   Longitude 177134 non-null float64
9   AttackType 181691 non-null int64
10  Killed    171378 non-null float64
11  Wounded   165380 non-null float64
12  Target    181055 non-null object
13  TargetType 181691 non-null object
14  Group     181691 non-null object
15  Summary   115562 non-null object
16  TargetType 181691 non-null object
17  WeaponType 181691 non-null object
18  Motive    50561 non-null object
dtypes: float64(4), int64(4), object(11)
memory usage: 26.3+ MB
```

```
In [137... df.isnull().sum()/df.isnull().count() * 100#%of null values
```

```
Out[137... Year      0.000000
Month     0.000000
Day       0.000000
Country   0.000000
Region    0.000000
State     0.231712
City      0.238867
Latitude  2.507554
Longitude 2.508104
AttackType 0.000000
Killed    5.676120
Wounded   8.977330
Target    0.350045
TargetType 0.000000
Group     0.000000
Summary   36.396409
TargetType 0.000000
WeaponType 0.000000
Motive    72.171984
```

In [138...

```
df = df.drop('Longitude', axis=1)
```

In [139...

```
df.head()
```

Out[139...

	Year	Month	Day	Country	Region	State	City	Latitude	AttackType	Killed	Wounded	Target	TargetType	Group
<b>0</b>	1970	7	2	Dominican Republic	Central America & Caribbean	NaN	Santo Domingo	18.456792	1	1.0	0.0	Julio Guzman	Private Citizens & Property	MANO-D
<b>1</b>	1970	0	0	Mexico	North America	Federal	Mexico city	19.371887	6	0.0	0.0	Nadine Chaval, daughter	Government (Diplomatic)	23rd of September Communist League
<b>2</b>	1970	1	0	Philippines	Southeast Asia	Tarlac	Unknown	15.478598	1	1.0	0.0	Employee	Journalists & Media	Unknown
<b>3</b>	1970	1	0	Greece	Western Europe	Attica	Athens	37.997490	3	NaN	NaN	U.S. Embassy	Government (Diplomatic)	Unknown
<b>4</b>	1970	1	0	Japan	East Asia	Fukouka	Fukouka	33.580412	7	NaN	NaN	U.S. Consulate	Government (Diplomatic)	Unknown

In [140...

```
df.tail()
```

Out[140...

	Year	Month	Day	Country	Region	State	City	Latitude	AttackType	Killed	Wounded	Target	TargetType	
<b>181686</b>	2017	12	31	Somalia	Sub-Saharan Africa	Middle Shebelle	Ceelka Geelow	2.359673		2	1.0	2.0	Checkpoint	Military
<b>181687</b>	2017	12	31	Syria	Middle East & North Africa	Lattakia	Jableh	35.407278		3	2.0	7.0	Hmeymim Air Base	Military

	Year	Month	Day	Country	Region	State	City	Latitude	AttackType	Killed	Wounded	Target	TargetType	
181688	2017	12	31	Philippines	Southeast Asia	Maguindanao	Kubentog	6.900742		7	0.0	0.0	Houses	Private Citizens & Property
181689	2017	12	31	India	South Asia	Manipur	Imphal	24.798346		3	0.0	0.0	Office	Government (General)
181690	2017	12	31	Philippines	Southeast Asia	Maguindanao	Cotabato City	7.209594		3	0.0	0.0	Unknown	Unknown

In [141...

```
df.shape
```

Out[141...

```
(181691, 18)
```

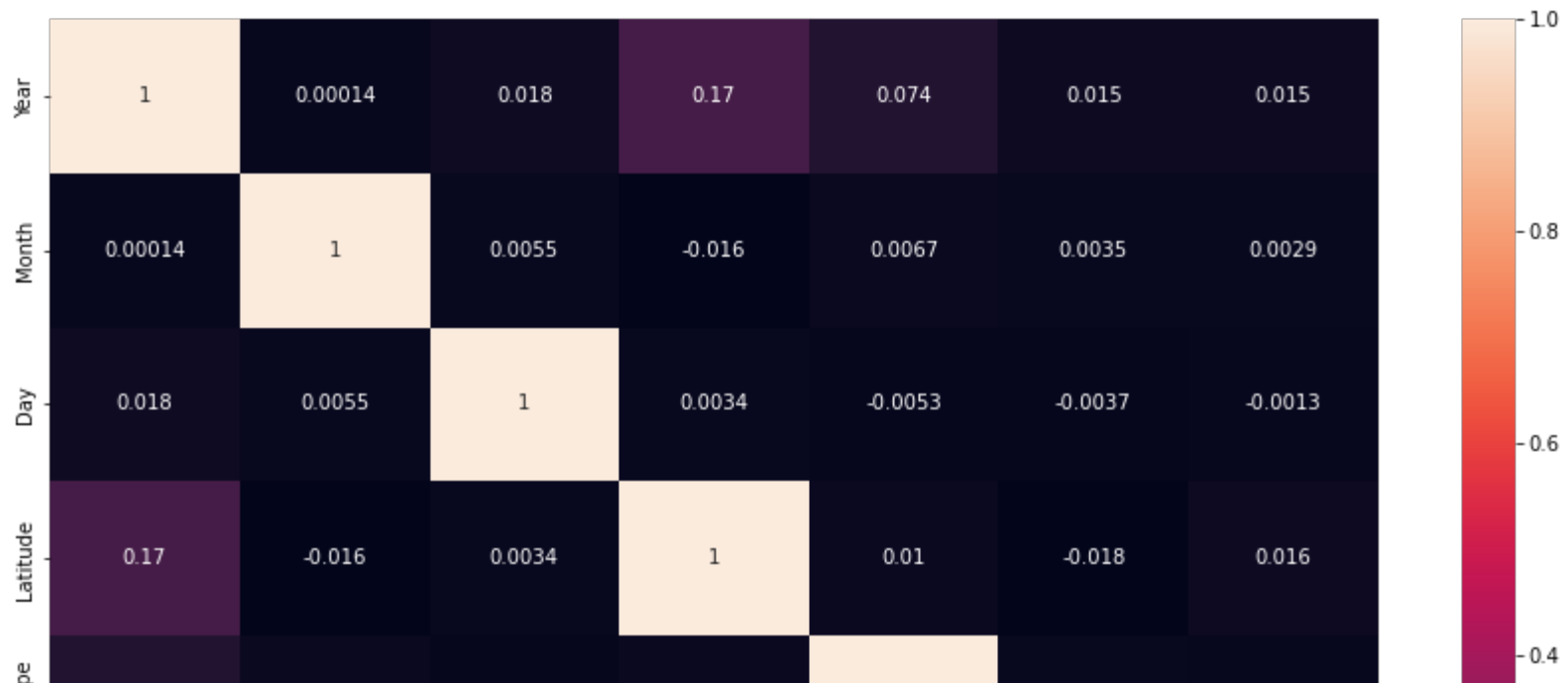
In [142...

```
plt.figure(figsize=(15,10))  
cor = df.corr()  
sns.heatmap(cor, annot=True)
```

Out[142...

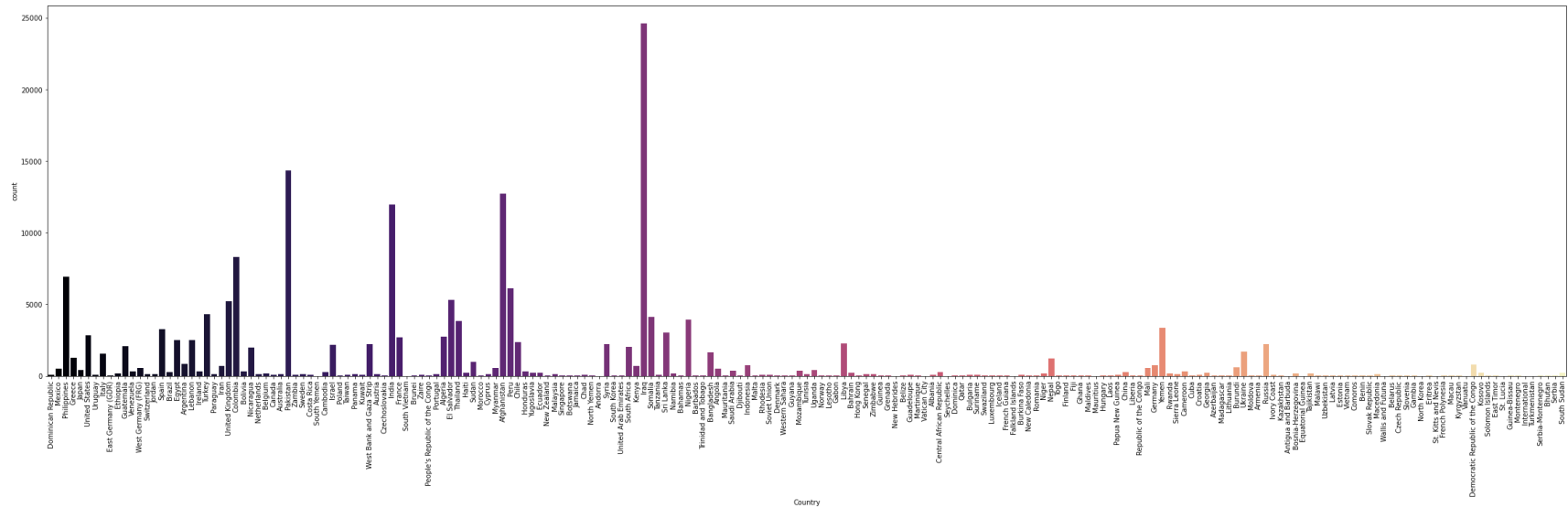
&lt;AxesSubplot:&gt;





In [143...

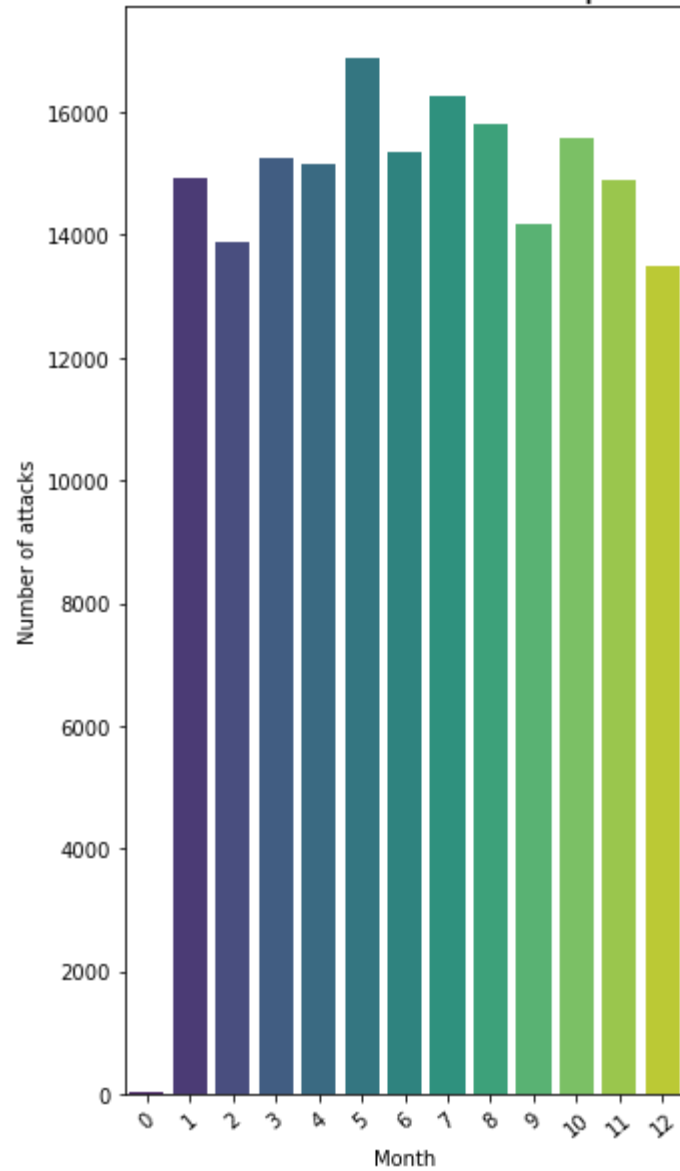
```
plt.figure(figsize=(40,10))
sns.countplot(df['Country'],palette="magma")
plt.xticks(rotation=90)
plt.show()
```



In [155...

```
plt.figure(figsize=(5,10))
sns.countplot(df['Month'],palette='viridis')
plt.title('Number of Terrorist Attacks per Month',fontsize='20')
plt.ylabel("Number of attacks")
plt.xlabel("Month")
plt.xticks(rotation='40')
plt.show()
```

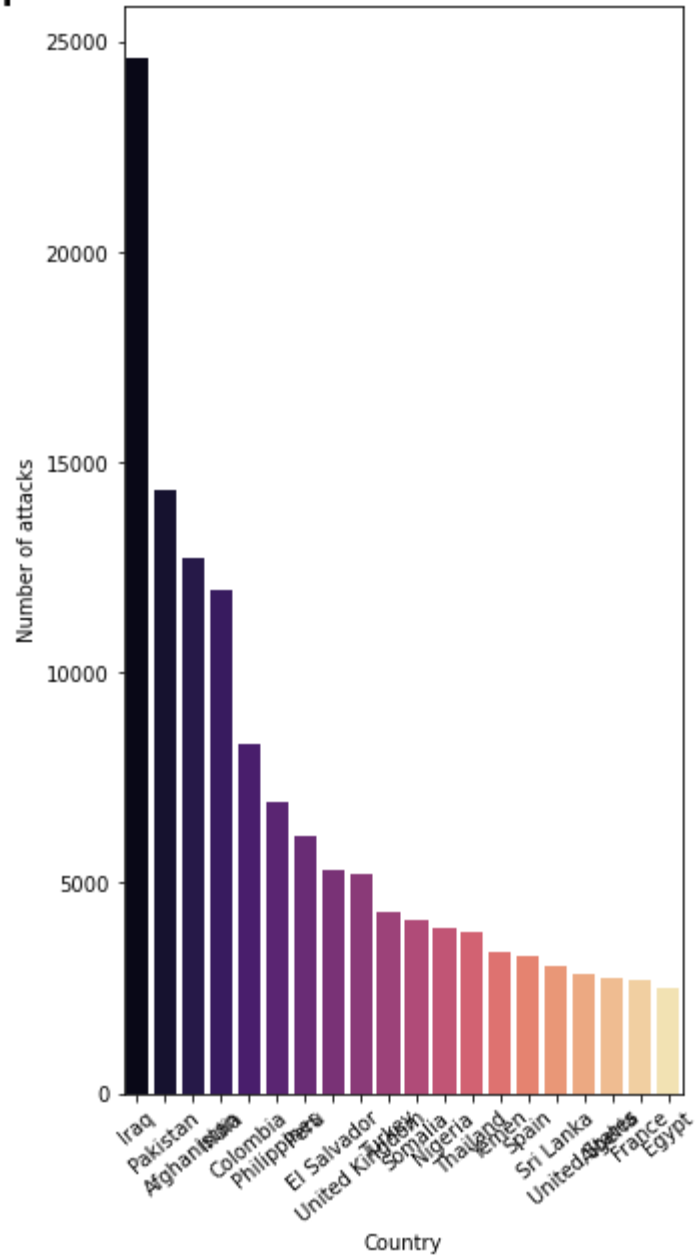
### Number of Terrorist Attacks per Month



In [156...

```
plt.subplots(figsize=(5,10))
sns.barplot(df['Country'].value_counts()[:20].index, df['Country'].value_counts()[:20].values,palette='magma')
plt.title('Top 20 most affected countries', fontsize = 30)
plt.xlabel('Country')
plt.ylabel('Number of attacks')
plt.xticks(rotation = 40)
plt.show()
```

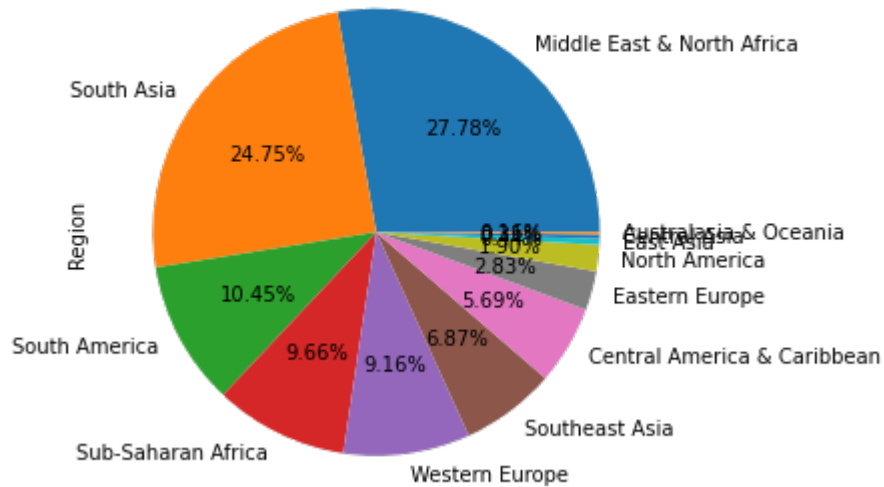
## Top 20 most affected countries



In [157...

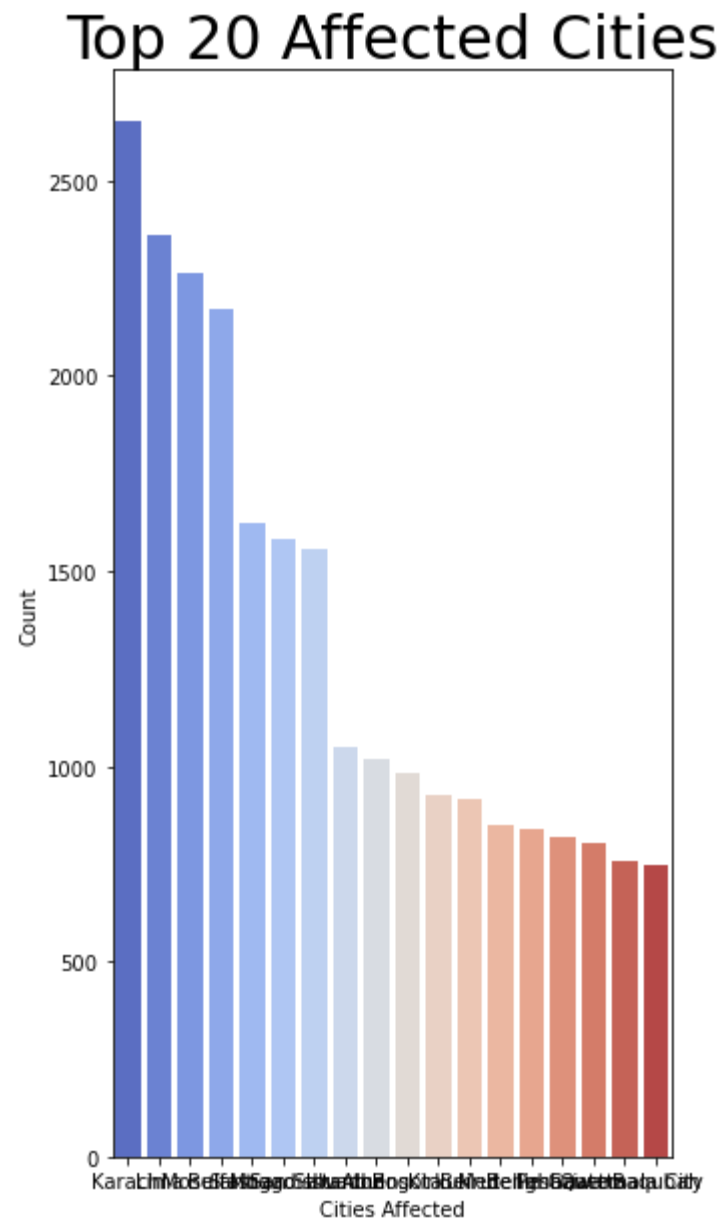
```
plt.figure(figsize=(5,10))
plt.title('Top 10 Affected Regions',fontsize='30')
df['Region'].value_counts().head(30).plot(kind='pie',autopct='%.02f%%')
plt.show()
```

## Top 10 Affected Regions



In [164...

```
plt.figure(figsize=(5,10))
sns.barplot(df['City'].value_counts()[2:20].index,df['City'].value_counts()[2:20].values, palette='coolwarm')
plt.title('Top 20 Affected Cities',fontsize='30')
plt.xlabel('Cities Affected')
plt.ylabel('Count')
plt.show()
```



In [148...

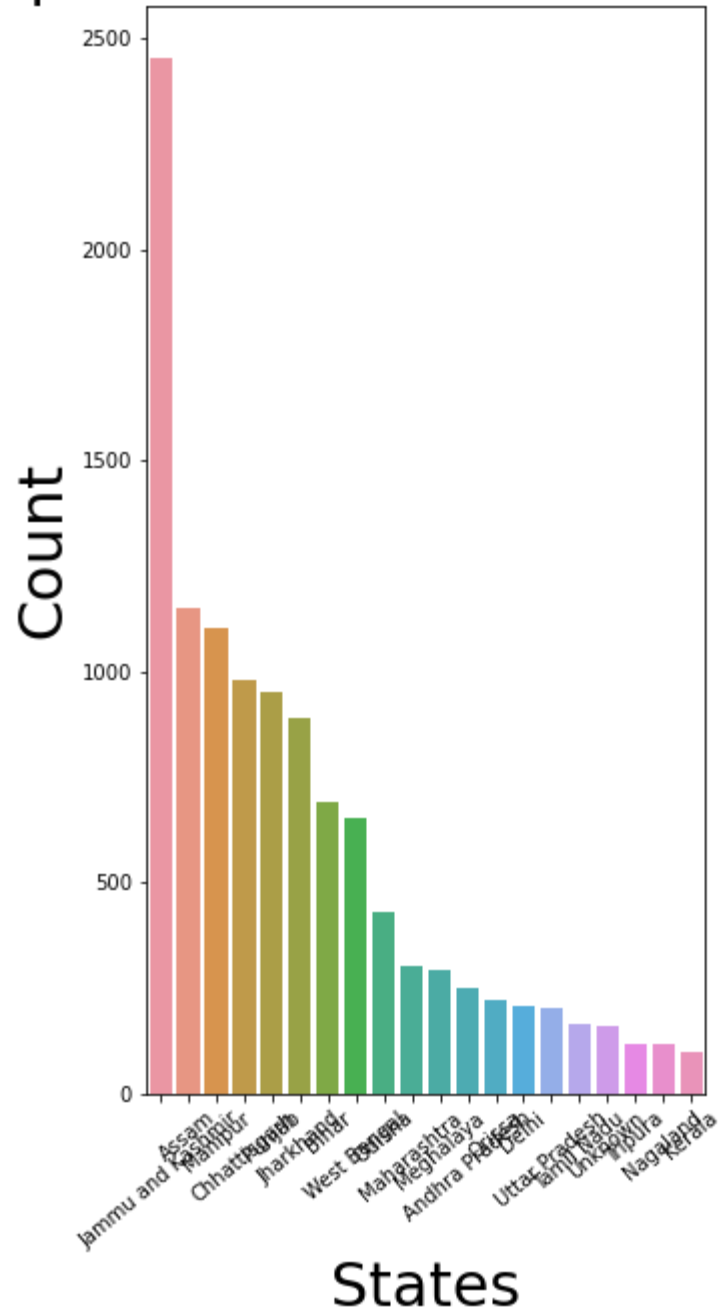
```
df_india = df[df['Country'] == 'India']['State']
```

In [160...

```
plt.figure(figsize=(5,10))
sns.barplot(df_india.value_counts()[ :20].index,df_india.value_counts()[ :20].values)
plt.title('Top 20 Affected States of india',fontsize='30')
plt.xlabel('States',fontsize='30')
plt.ylabel('Count',fontsize='30')
plt.xticks(rotation='40')
plt.show()
```

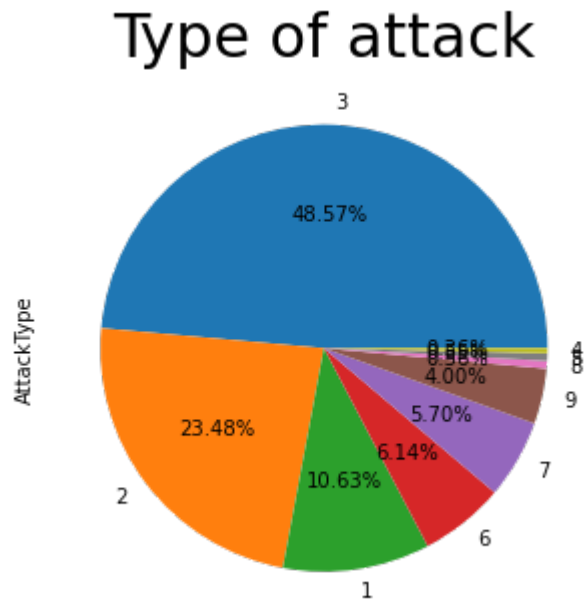


## Top 20 Affected States of india



In [161...

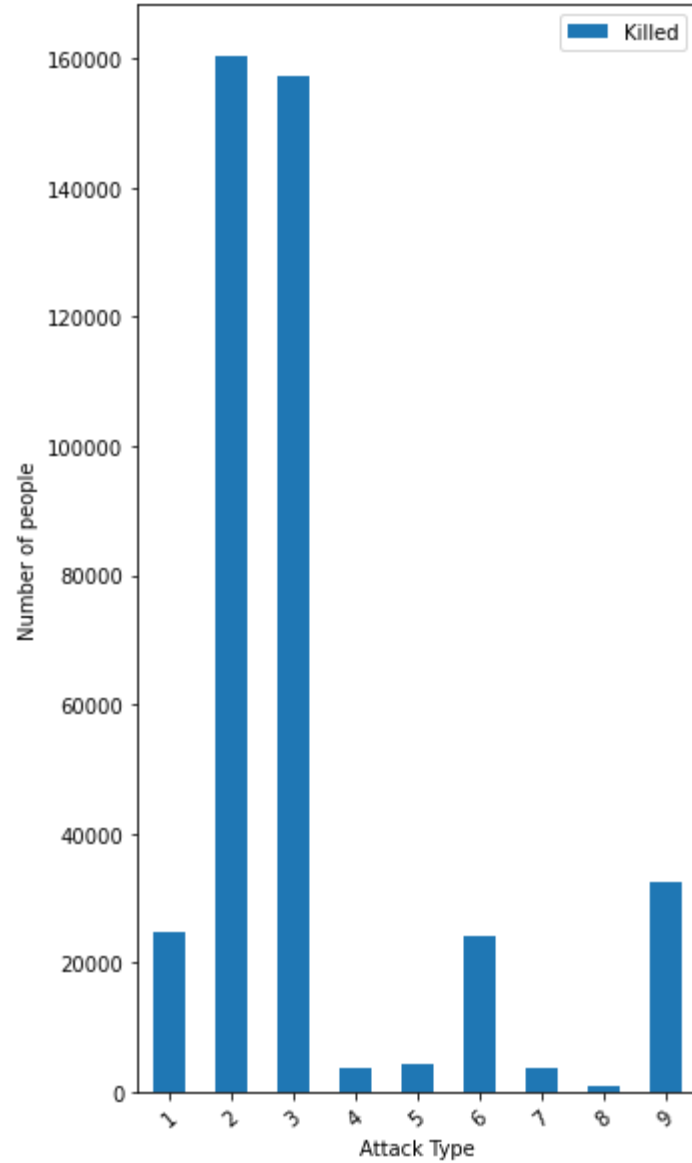
```
plt.figure(figsize=(5,10))
df['AttackType'].value_counts().plot(kind='pie', autopct='%0.02f%%')
plt.title('Type of attack', fontsize='30')
plt.show()
```



In [162...

```
df[['AttackType', 'Killed']].groupby(['AttackType']).sum().plot(kind='bar', figsize=(5,10))
plt.title('Number of Persons Killed', fontsize='30')
plt.xlabel('Attack Type')
plt.ylabel('Number of people')
plt.xticks(rotation='40')
plt.show()
```

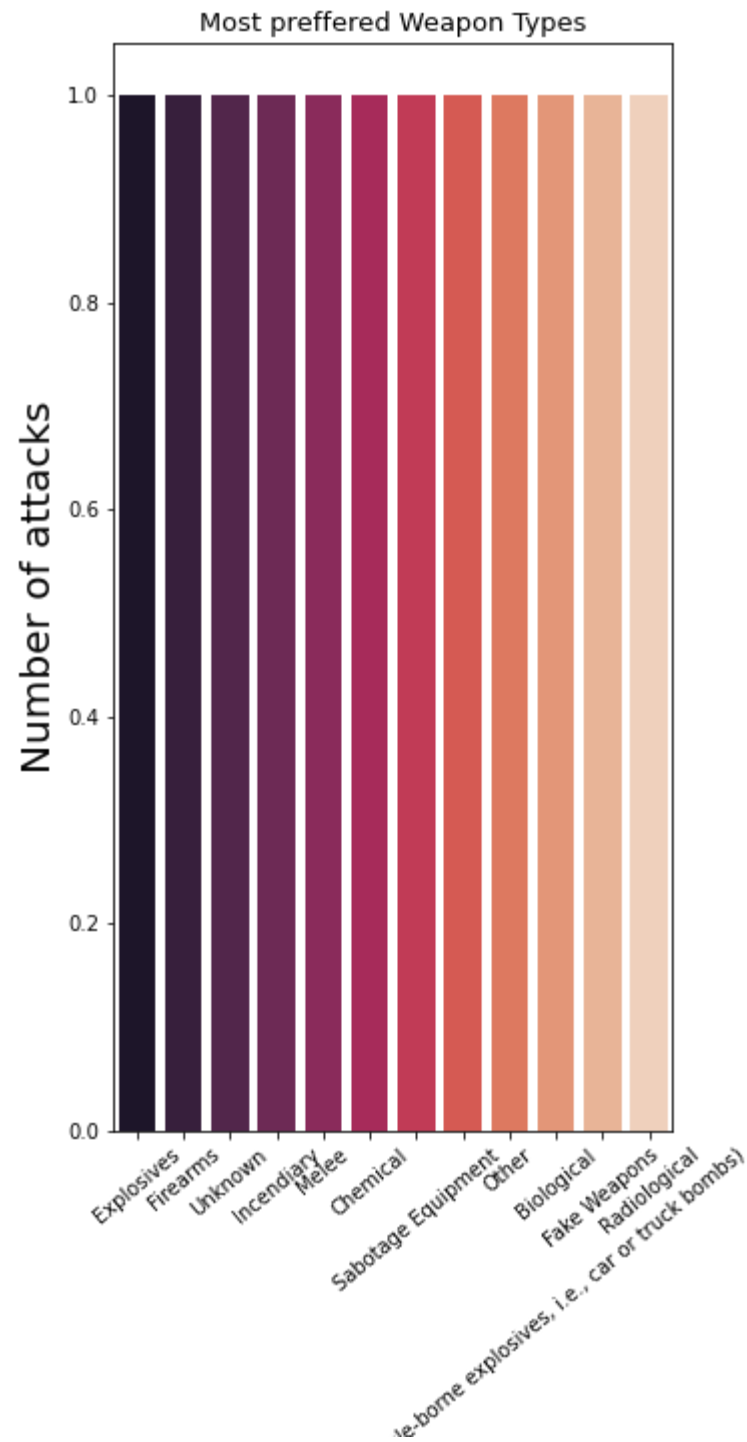
## Number of Persons Killed



In [152...

```
plt.subplots(figsize=(5,10))
sns.countplot(df['WeaponType'].value_counts().index,palette="rocket")
plt.title("Most preferred Weapon Types",fontsize='13')
plt.xlabel("WeaponType",fontsize=20)
plt.ylabel("Number of attacks",fontsize=20)
plt.xticks(rotation='40')

plt.show()
```

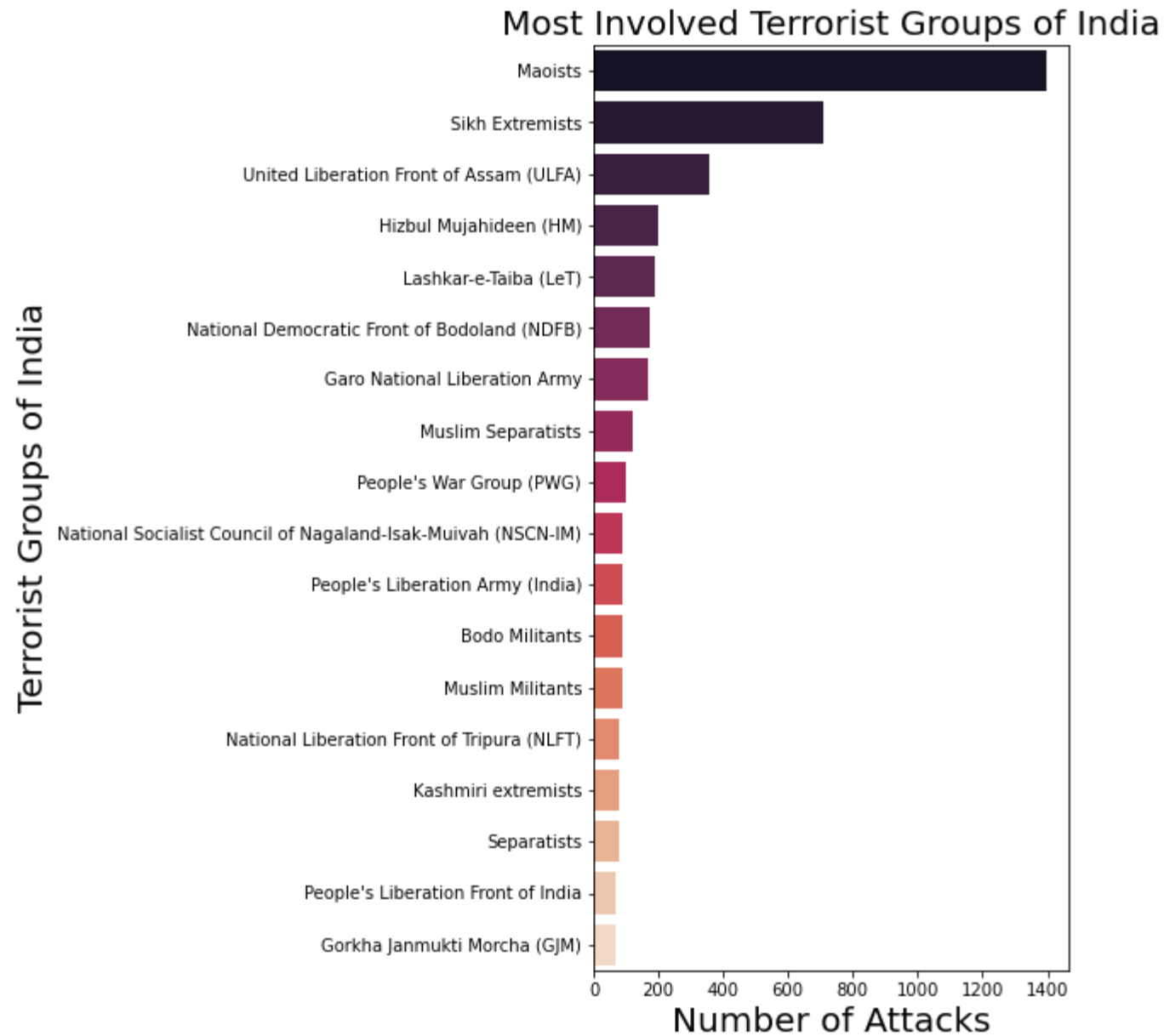


Vehicle (not to include vehicle)

## WeaponType

```
In [153... df_india_grp = df[df['Country'] == 'India']['Group']
```

```
In [163... plt.figure(figsize=(5,10))
sns.barplot(df_india_grp.value_counts()[2:20].values, df_india_grp.value_counts()[2:20].index,palette="rocke
plt.title('Most Involved Terrorist Groups of India',fontsize='20')
plt.xlabel('Number of Attacks',fontsize='20')
plt.ylabel('Terrorist Groups of India', fontsize='20')
plt.show()
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: