

In [97]:

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

In [71]:

```
encoding_list = ['ascii', 'big5', 'big5hkscs', 'cp037', 'cp273', 'cp424', 'cp437', 'cp50',
                 'cp775', 'cp850', 'cp852', 'cp855', 'cp856', 'cp857', 'cp858', 'cp860',
                 'cp863', 'cp864', 'cp865', 'cp866', 'cp869', 'cp874', 'cp875', 'cp932',
                 'cp1006', 'cp1026', 'cp1125', 'cp1140', 'cp1250', 'cp1251', 'cp1252',
                 'cp1255', 'cp1256', 'cp1257', 'cp1258', 'euc_jp', 'euc_jis_2004', 'eu',
                 'gb2312', 'gbk', 'gb18030', 'hz', 'iso2022_jp', 'iso2022_jp_1', 'iso2',
                 'iso2022_jp_2004', 'iso2022_jp_3', 'iso2022_jp_ext', 'iso2022_kr', 'l',
                 'iso8859_3', 'iso8859_4', 'iso8859_5', 'iso8859_6', 'iso8859_7', 'iso',
                 'iso8859_10', 'iso8859_11', 'iso8859_13', 'iso8859_14', 'iso8859_15',
                 'koi8_r', 'koi8_t', 'koi8_u', 'kz1048', 'mac_cyrillic', 'mac_greek',
                 'mac_roman', 'mac_turkish', 'ptcp154', 'shift_jis', 'shift_jis_2004',
                 'utf_32_be', 'utf_32_le', 'utf_16', 'utf_16_be', 'utf_16_le', 'utf_7']

path= r"C:\\Users\\dtdee\\OneDrive\\Desktop\\Letsupgrade_Python\\Tableau\\data-society-g
for encoding in encoding_list:
    worked = True
    try:
        df = pd.read_csv(path, encoding=encoding)
        print('csvloaded')
    except:
        worked = False
    if worked:
        print(encoding, '\n', df.head())
```

C:\Users\dtdee\AppData\Roaming\Python\Python39\site-packages\IPython\core\interactiveshell.py:3433: DtypeWarning: Columns (4,17,61,62,123) have mixed types.Specify dtype option on import or set low_memory=False.
 exec(code_obj, self.user_global_ns, self.user_ns)

csvloaded

cp437

	eventid	iyear	imonth	iday	approxdate	extended	resolution
country \							
0	197000000001	1970	0	0	NaN	0	NaN
58							
1	197000000002	1970	0	0	NaN	0	NaN
130							
2	197001000001	1970	1	0	NaN	0	NaN
160							
3	197001000002	1970	1	0	NaN	0	NaN
78							
4	197001000003	1970	1	0	NaN	0	NaN
101							

In [72]:

```
df.shape
```

Out[72]:

```
(49566, 137)
```

In [73]:

```
df.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Country',
df=df[['Year','Month','Day','Country','Region','city','latitude','longitude','AttackType
df['casualties']=df['Killed']+df['Wounded']
```

In [74]:

```
df.shape
```

Out[74]:

```
(49566, 18)
```

In [75]:

```
df.duplicated().sum()
```

Out[75]:

```
4650
```

In [76]:

```
df.drop_duplicates(inplace=True)
```

In [77]:

```
df.shape
```

Out[77]:

```
(44916, 18)
```

In [78]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 44916 entries, 0 to 49565
Data columns (total 18 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Year            44916 non-null  int64
1   Month           44916 non-null  int64
2   Day             44916 non-null  int64
3   Country         44916 non-null  object
4   Region          44916 non-null  object
5   city            44916 non-null  object
6   latitude        42486 non-null  float64
7   longitude       42486 non-null  float64
8   AttackType      44916 non-null  object
9   Killed          39826 non-null  float64
10  Wounded         38029 non-null  float64
11  Target          44585 non-null  object
12  Summary         1153 non-null   object
13  Group           44916 non-null  object
14  Target_type     44916 non-null  object
15  Weapon_type     44916 non-null  object
16  Motive          952 non-null    object
17  casualties      37855 non-null  float64
dtypes: float64(5), int64(3), object(10)
memory usage: 6.5+ MB
```

In [79]:

```
df.head(3)
```

Out[79]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	
0	1970	0	0	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	
1	1970	0	0	Mexico	North America	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	
2	1970	1	0	Philippines	Southeast Asia	Unknown	15.478598	120.599741	Assassination	

In [87]:

```
df['latitude'].value_counts()
```

Out[87]:

```
<bound method NDFrame._add_numeric_operations.<locals>.mean of 0      1
8.456792\n1      19.432608\n2      15.478598\n3      37.983773\n4
33.580412\n      ...      \n49561      39.933971\n49562      24.898129\n495
63      32.311973\n49564      23.811388\n49565      54.597269\nName: latitude, L
ength: 44916, dtype: object>      2430
54.597269
1374
-12.046378
1339
13.69288
1198
-33.469115
1112

...
34.022442
1
-13.75
1
13.83759
1
-12.983333
1
24.898129
1
Name: latitude, Length: 9176, dtype: int64
```

In [81]:

```
df.isnull().sum()
```

Out[81]:

```
Year      0
Month     0
Day       0
Country   0
Region    0
city      0
latitude  2430
longitude 2430
AttackType 0
Killed     5090
Wounded    6887
Target     331
Summary    43763
Group      0
Target_type 0
Weapon_type 0
Motive     43964
casualties 7061
dtype: int64
```

In [88]:

```
df['latitude'].fillna(df['latitude'].mean,inplace=True)
df['longitude'].fillna(df['longitude'].mean(),inplace=True)
df['Killed'].fillna(df['Killed'].mean(),inplace=True)
df['Wounded'].fillna(df['Wounded'].mean(),inplace=True)
df['Target'].fillna("Unknown_Tgt",inplace=True)
```

In [89]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 44916 entries, 0 to 49565
Data columns (total 18 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Year            44916 non-null  int64
 1   Month           44916 non-null  int64
 2   Day             44916 non-null  int64
 3   Country         44916 non-null  object
 4   Region          44916 non-null  object
 5   city            44916 non-null  object
 6   latitude        44916 non-null  object
 7   longitude       44916 non-null  float64
 8   AttackType     44916 non-null  object
 9   Killed          44916 non-null  float64
10   Wounded         44916 non-null  float64
11   Target          44916 non-null  object
12   Summary         1153 non-null   object
13   Group           44916 non-null  object
14   Target_type     44916 non-null  object
15   Weapon_type     44916 non-null  object
16   Motive          952 non-null    object
17   casualties      37855 non-null  float64
dtypes: float64(4), int64(3), object(11)
memory usage: 6.5+ MB
```

In [90]:

```
# df.drop(['Summary','Motive','casualties'],axis=1)
```

In [95]:

```
# df['casualties']=df['Killed']+df['Wounded']
```

In [92]:

```
df.head(2)
```

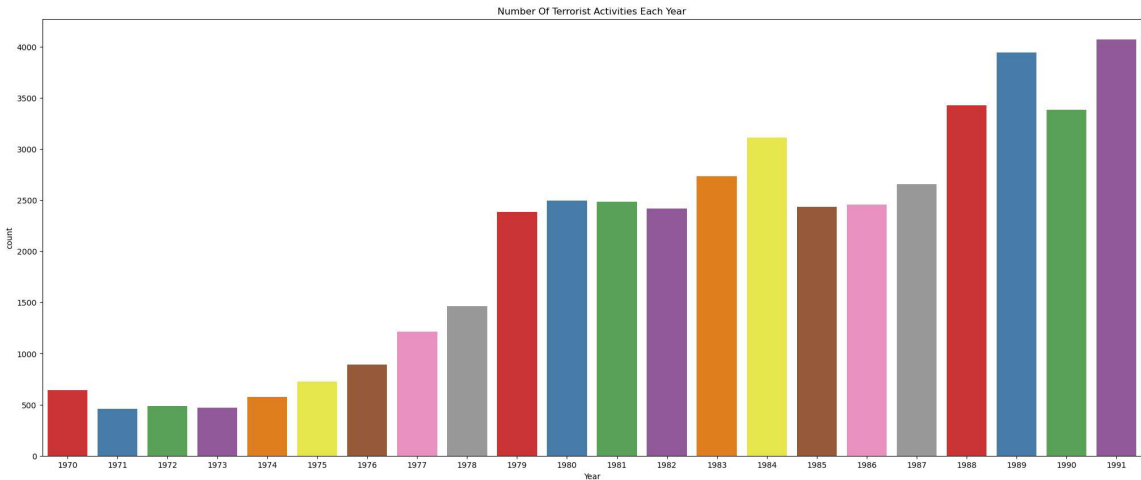
Out[92]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	K
0	1970	0	0	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	
1	1970	0	0	Mexico	North America	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	

Number Of Terrorist Activities Each Year

In [108]:

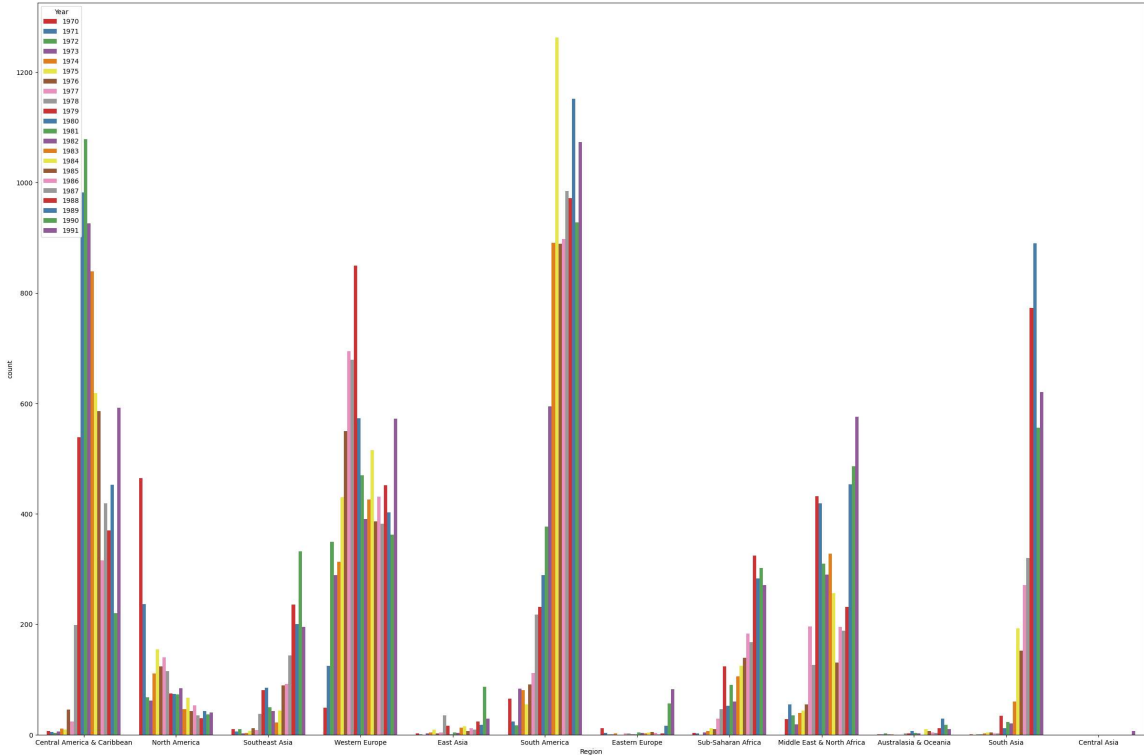
```
plt.subplots(figsize=(25,10))
sns.countplot(x='Year',data=df,palette='Set1')
plt.title('Number Of Terrorist Activities Each Year')
plt.show()
```



Terrorist Activities by Region in each Year

In [120]:

```
plt.subplots(figsize=(30,20))
sns.countplot(x='Region' , data=df,hue='Year' ,palette='Set1')
plt.show()
```



Number of Total Attacks in Each Country

In [121]:

```
df.head(2)
```

Out[121]:

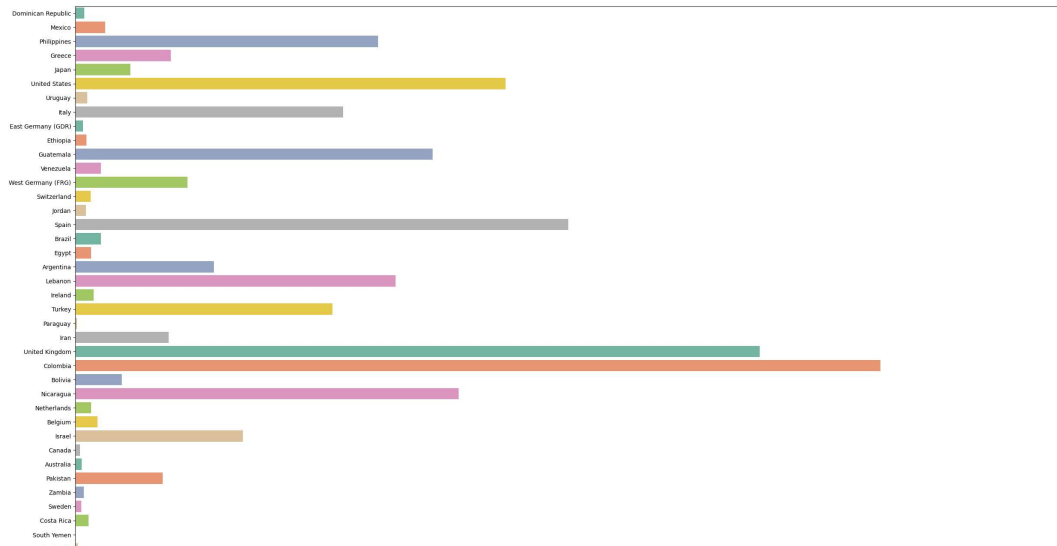
	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	K
0	1970	0	0	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	
1	1970	0	0	Mexico	North America	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	

In [173]:

```
plt.figure(figsize=(30,70))
sns.countplot(y='Country',data=df,palette='Set2')
```

Out[173]:

```
<AxesSubplot:xlabel='count', ylabel='Country'>
```



Top 5 Weapons Used

In [180]:

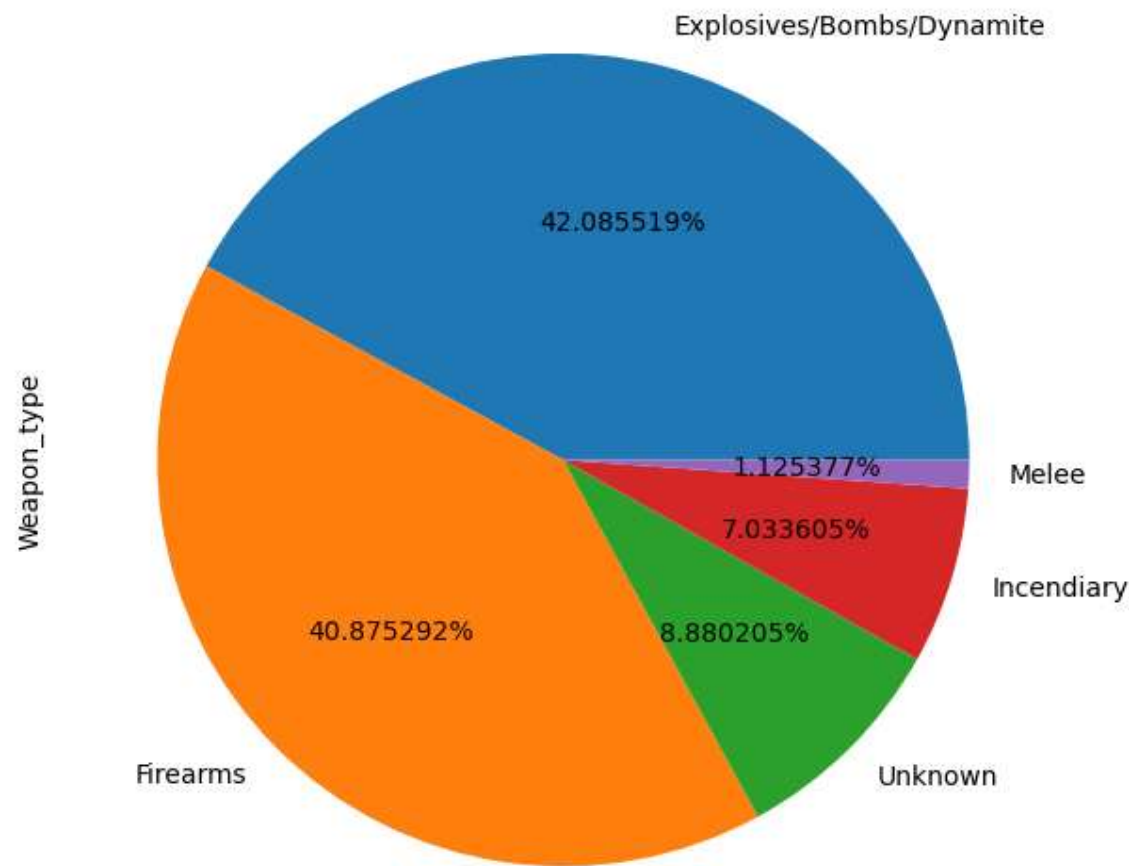
```
weapon=df[ 'Weapon_type' ].value_counts()[0:5]
```


In [184]:

```
weapon.plot(kind='pie',autopct='%1f%%',figsize=(15,7))
```

Out[184]:

<AxesSubplot:ylabel='Weapon_type'>



Number of Total Casualties in Each Country

In [187]:

```
df.head(2)
```

Out[187]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	K
0	1970	0	0	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	
1	1970	0	0	Mexico	North America	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	

In [208]:

```
df.groupby(['Country'])['casualties'].sum().sort_values(ascending=False)
```

Out[208]:

```
Country
El Salvador    21530.698254
Peru           16284.217392
Nicaragua      14427.405383
Sri Lanka      12881.382760
Colombia       12508.228598
...
Armenia        0.000000
Albania        0.000000
New Zealand    0.000000
Bahamas        0.000000
Moldova        0.000000
Name: casualties, Length: 163, dtype: float64
```

In [228]:

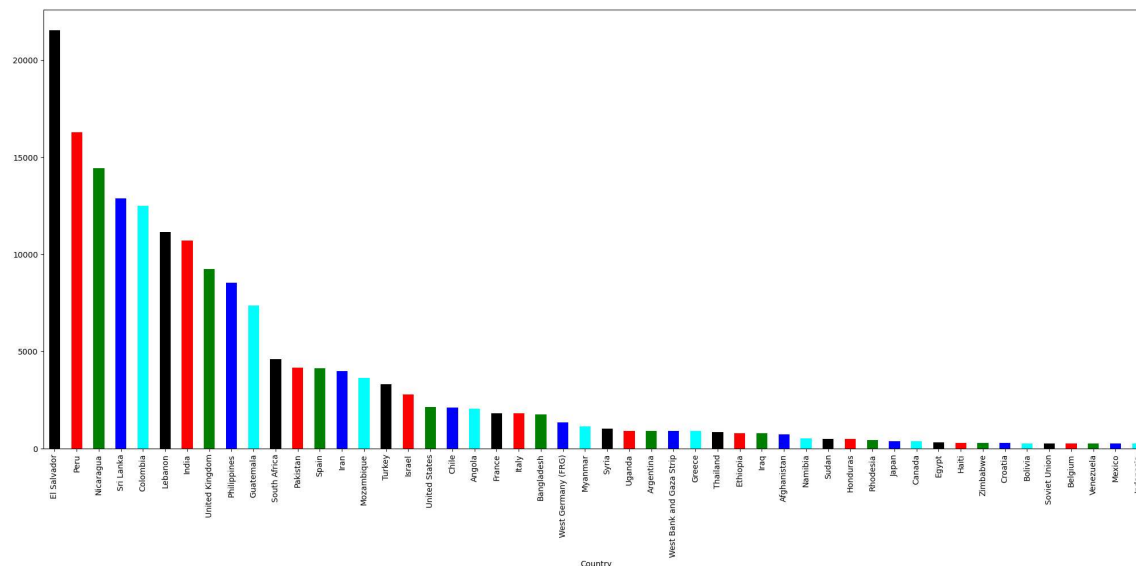
```
casualties=df.groupby(['Country'])['casualties'].sum().sort_values(ascending=False)[0:5
```

In [241]:

```
casualties.plot(kind='bar',figsize=(25,10),color=['black', 'red', 'green', 'blue', 'cyan'
```

Out[241]:

<AxesSubplot:xlabel='Country'>



Number Of Terrorist Activities Middle East & North Africa Each Month

In [242]:

```
df.head(2)
```

Out[242]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	K
0	1970	0	0	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	
1	1970	0	0	Mexico	North America	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	

In [270]:

```
df[df['Region']=='Middle East & North Africa']['Month']
```

Out[270]:

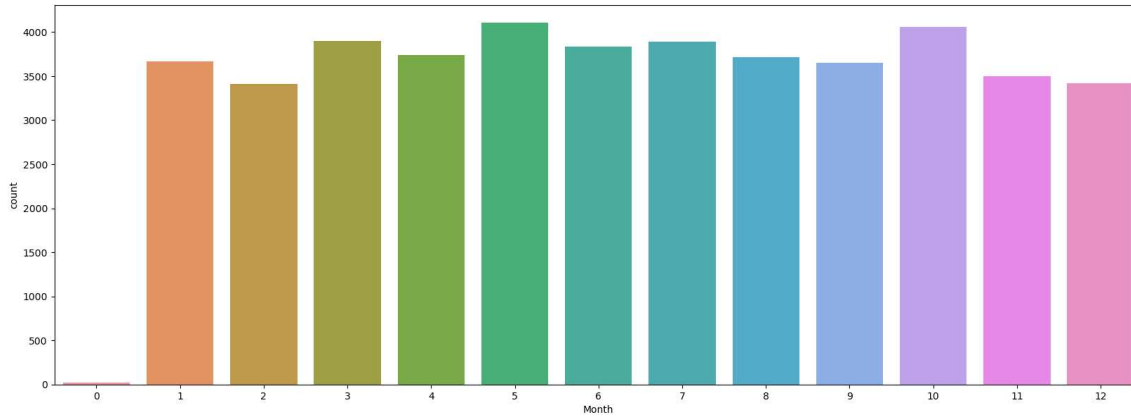
```
0      0
1      0
2      1
3      1
4      1
..
49561  12
49562  12
49563  12
49564  12
49565  12
Name: Month, Length: 44916, dtype: int64
```

In [274]:

```
plt.subplots(figsize=(20,7))
sns.countplot(x=df[df['Region']=='Middle East & North Africa']['Month'],data=df)
```

Out[274]:

```
<AxesSubplot:xlabel='Month', ylabel='count'>
```



Number Of Terrorist Activities Middle East & North Africa Each Year

In [277]:

```
df[df['Region']=='Middle East & North Africa']['Year']
```

Out[277]:

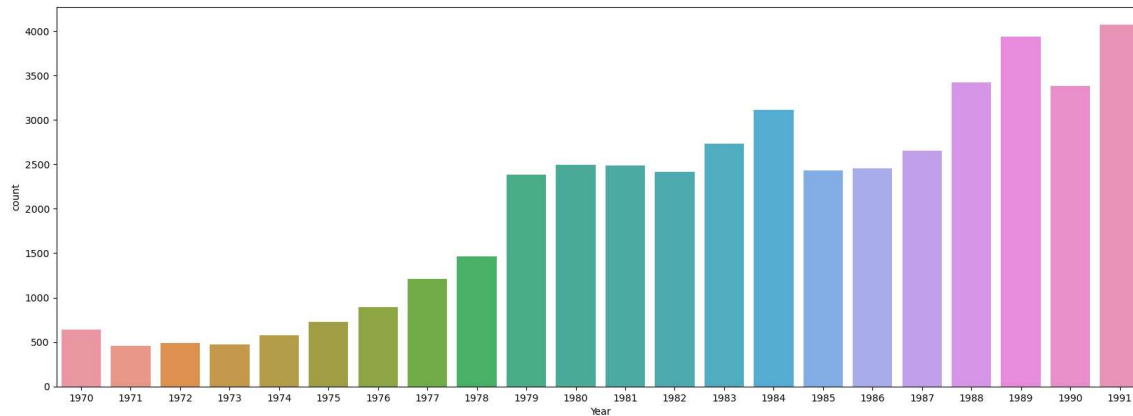
```
0      1970
1      1970
2      1970
3      1970
4      1970
...
49561   1991
49562   1991
49563   1991
49564   1991
49565   1991
Name: Year, Length: 44916, dtype: int64
```

In [278]:

```
plt.subplots(figsize=(20,7))  
sns.countplot(x=df[df['Region']=='Middle East & North Africa']['Year'],data=df)
```

Out[278]:

<AxesSubplot:xlabel='Year', ylabel='count'>



Percentage of casualties due to different ways of attacks

In [286]:

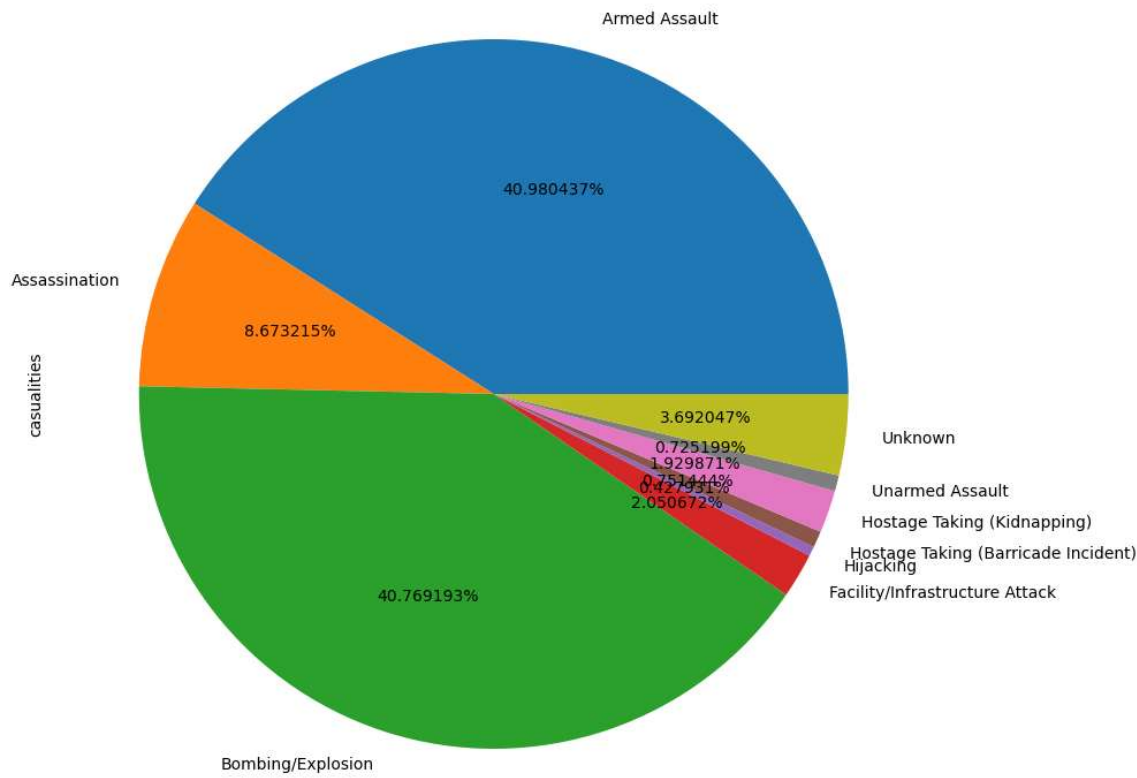
```
attacked=df.groupby(['AttackType'])['casualties'].sum()
```

In [288]:

```
attacked.plot(kind='pie', autopct='%1f%%', figsize=(25,10))
```

Out[288]:

<AxesSubplot:ylabel='casualties'>



Number of people died in each year

In [289]:

```
df.head(2)
```

Out[289]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	Killed
0	1970	0	0	Dominican Republic	Middle East & North Africa	Santo Domingo	18.456792	-69.951164	Assassination	1
1	1970	0	0	Mexico	Middle East & North Africa	Mexico city	19.432608	-99.133207	Hostage Taking (Kidnapping)	0

In [314]:

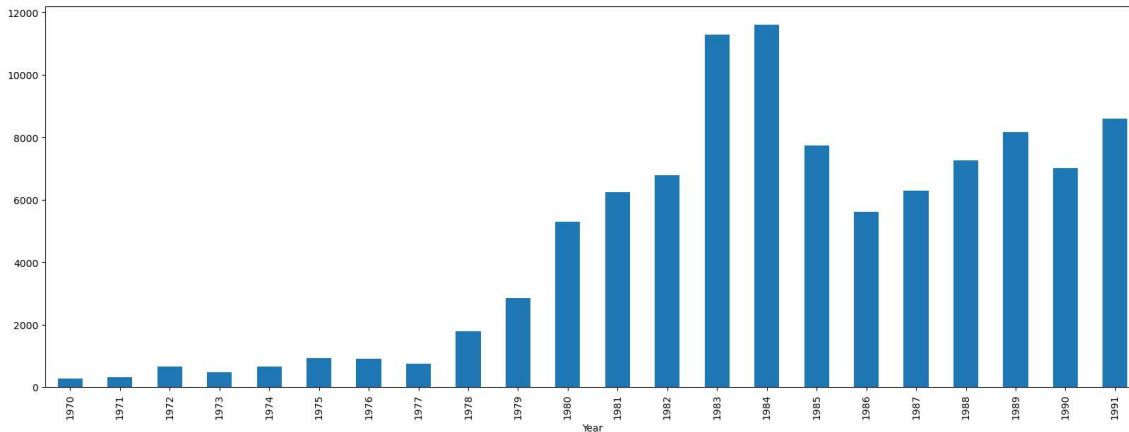
```
X=df.groupby(['Year'])['Killed'].sum()
```

In [316]:

```
X.plot(kind='bar',figsize=(20,7))
```

Out[316]:

<AxesSubplot:xlabel='Year'>



Conclusion

1. Most No of people died year 1983-85.
2. So most Terrorist activities started in 1984 year and onwards peak in 1989-1991
3. Most Terrorist activites were in the Soth America and Central America& Caribbean
4. Most cities which were attacked were Peru,Columbia,El Salvador
5. Most weapons used for attacking was Explosives,Bombs,Dynamites and Firearms
6. No of casualties found were more in El Salvador,peru,Salvador,Srilanka
7. Number Of Terrorist Activities Middle East & North Africa were most in 1989 and 1991 and most in months of May and October
8. Percentage of casualties were most due to attacks like Armed Assaukt and Bombing
9. Most no of people died in the year 1983 and 1984