Double-click (or enter) to edit

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

```
data = pd.read_csv('/content/globalterrorism.csv',encoding='ISO-8859-1')
```

```
data.head(10)
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	sc
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	
1	1.970000e+11	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	
2	1.970000e+11	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	
3	1.970000e+11	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	
4	1.970000e+11	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	
5	1.970000e+11	1970	1	1	NaN	0	NaN	217	United States	1	...	The Cairo Chief of Police, William Petersen, r...	"I C Washi Janu
6	1.970000e+11	1970	1	2	NaN	0	NaN	218	Uruguay	3	...	NaN	
7	1.970000e+11	1970	1	2	NaN	0	NaN	217	United States	1	...	Damages were estimated to be between $20,000-$...	Comr Goverr Opera U
8	1.970000e+11	1970	1	2	NaN	0	NaN	217	United States	1	...	The New Years Gang issue a communiqué to a loc...	Tom E "Rads Bomb the A
9	1.970000e+11	1970	1	3	NaN	0	NaN	217	United States	1	...	Karl Armstrong's girlfriend, Lynn Schultz, dro...	Comr Goverr Opera U

10 rows × 135 columns

```
data.shape
```

```
(181691, 135)
```

```
data.info
```

```
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      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  START Primary Collection        0         0        0       0      NaN
181687  START Primary Collection       -9        -9        1       1      NaN
181688  START Primary Collection        0         0        0       0      NaN
181689  START Primary Collection       -9        -9        0      -9      NaN
181690  START Primary Collection       -9        -9        0      -9      NaN

[181691 rows x 135 columns]>
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(56), int64(21), object(58)
memory usage: 187.1+ MB
```

```
data.isnull()
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1
0	False	False	False	False	True	False	True	False	False	False	...	True	True
1	False	False	False	False	True	False	True	False	False	False	...	True	True
2	False	False	False	False	True	False	True	False	False	False	...	True	True
3	False	False	False	False	True	False	True	False	False	False	...	True	True
4	False	False	False	False	True	False	True	False	False	False	...	True	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...
181686	False	False	False	False	True	False	True	False	False	False	...	True	False
181687	False	False	False	False	True	False	True	False	False	False	...	True	False
181688	False	False	False	False	True	False	True	False	False	False	...	True	False
181689	False	False	False	False	True	False	True	False	False	False	...	True	False
181690	False	False	False	False	True	False	True	False	False	False	...	True	False

181691 rows × 135 columns

```
data.isnull().sum()
```

```
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
```

```
data.describe
```

```
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      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  START Primary Collection        0         0        0       0      NaN
181687  START Primary Collection       -9        -9        1       1      NaN
181688  START Primary Collection        0         0        0       0      NaN
181689  START Primary Collection       -9        -9        0      -9      NaN
181690  START Primary Collection       -9        -9        0      -9      NaN

[181691 rows x 135 columns]>
```

```
data.describe()
```

	eventid	iyear	imonth	iday	extended	country	region	latitude	
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	181691.000000	177135.000000	
mean	2.003238e+11	2002.638997	6.467277	15.505644	0.045346	131.968501	7.160938	23.498343	-
std	1.383523e+09	13.259430	3.388303	8.814045	0.208063	112.414535	2.933408	18.569242	
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.000000	1.000000	-53.154613	-
25%	1.990000e+11	1991.000000	4.000000	8.000000	0.000000	78.000000	5.000000	11.510046	
50%	2.010000e+11	2009.000000	6.000000	15.000000	0.000000	98.000000	6.000000	31.467463	
75%	2.010000e+11	2014.000000	9.000000	23.000000	0.000000	160.000000	10.000000	34.685087	
max	2.020000e+11	2017.000000	12.000000	31.000000	1.000000	1004.000000	12.000000	74.633553	

8 rows × 77 columns

```
data.columns
```

```
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)
```

```
data.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Country','provstate':'state',
                     'region_txt':'Region','attacktype1_txt':'AttackType','target1':'Target','nkill':'Killed',
                     'nwound':'Wounded','summary':'Summary','gname':'Group','targtype1_txt':'Target_type',
                     'weaptype1_txt':'Weapon_type','motive':'Motive'},inplace=True)
```

```
#data=data.rename(columns = {'iyear':'year','imonth':'month', 'iday':'day','country':'country','country_txt':'country_name','region':'reg
```

```
data
```

	eventid	Year	Month	Day	approxdate	extended	resolution	country	Country	region	...	addnotes	sc
0	1.970000e+11	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	
1	1.970000e+11	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	
2	1.970000e+11	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	
3	1.970000e+11	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	
4	1.970000e+11	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	2.020000e+11	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN	"Somali Shal Mili Attack C
181687	2.020000e+11	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN	"Pu 'victo Syria turned i
181688	2.020000e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Maguind clashes memb F
181689	2.020000e+11	2017	12	31	NaN	0	NaN	92	India	6	...	NaN	"Tr esc gre atta Imp B
181690	2.020000e+11	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN	"Sec tighten Cota following

181691 rows × 135 columns

```
data['country'].value_counts()
```

```
95     24636
153    14368
4      12731
92     11960
45      8306
       ...
422        1
226        1
428        1
7          1
10         1
Name: country, Length: 205, dtype: int64
```

```
data.drop('eventid',axis=1)
```

	Year	Month	Day	approxdate	extended	resolution	country	Country	region	Region	...	addnotes	scite1
0	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	Central America & Caribbean	...	NaN	NaN
1	1970	0	0	NaN	0	NaN	130	Mexico	1	North America	...	NaN	NaN
2	1970	1	0	NaN	0	NaN	160	Philippines	5	Southeast Asia	...	NaN	NaN
3	1970	1	0	NaN	0	NaN	78	Greece	8	Western Europe	...	NaN	NaN
4	1970	1	0	NaN	0	NaN	101	Japan	4	East Asia	...	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	..
181686	2017	12	31	NaN	0	NaN	182	Somalia	11	Sub-Saharan Africa	...	NaN	"Somalia: Al Shabaab Militants Attack Army Che..
181687	2017	12	31	NaN	0	NaN	200	Syria	10	Middle East & North Africa	...	NaN	"Putin's 'victory' in Syria has turned into a ..
181688	2017	12	31	NaN	0	NaN	160	Philippines	5	Southeast Asia	...	NaN	"Maguindanao clashes trap tribe members, Phil..
181689	2017	12	31	NaN	0	NaN	92	India	6	South Asia	...	NaN	"Trade escapes grenade attack in Imphal, Bus..
181690	2017	12	31	NaN	0	NaN	160	Philippines	5	Southeast Asia	...	NaN	"Security tightened in Cotabato following IED ..

181691 rows × 134 columns

```
data = data.fillna('-')
```

```
data
```

	eventid	Year	Month	Day	approxdate	extended	resolution	country	Country	region	...	addnotes	sc
0	1.970000e+11	1970	7	2	-	0	-	58	Dominican Republic	2	...	-	
1	1.970000e+11	1970	0	0	-	0	-	130	Mexico	1	...	-	
2	1.970000e+11	1970	1	0	-	0	-	160	Philippines	5	...	-	
3	1.970000e+11	1970	1	0	-	0	-	78	Greece	8	...	-	
4	1.970000e+11	1970	1	0	-	0	-	101	Japan	4	...	-	
...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	2.020000e+11	2017	12	31	-	0	-	182	Somalia	11	...	-	"Somali Shal Mili Attack C
													"Pu

```
data.isnull()
```

	eventid	Year	Month	Day	approxdate	extended	resolution	country	Country	region	...	addnotes	scite1	scite
0	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
1	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
2	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
3	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
4	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
181687	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
181688	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
181689	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals
181690	False	False	False	False	False	False	False	False	False	False	...	False	False	Fals

181691 rows × 135 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(1), int64(21), object(113)
memory usage: 187.1+ MB
```

```
data=data[['Year','Month','Day','Country','state','Region','city','latitude','longitude','AttackType','Killed',
           'Wounded','Target','Group','Target_type','Weapon_type','Motive']]
```

```
data
```

	Year	Month	Day	Country	state	Region	city	latitude	longitude	AttackType	Killed	Wound
0	1970	7	2	Dominican Republic	-	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	1.0	0
1	1970	0	0	Mexico	Federal	North America	Mexico city	19.371887	-99.086624	Hostage Taking (Kidnapping)	0.0	0
2	1970	1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741	Assassination	1.0	0
3	1970	1	0	Greece	Attica	Western Europe	Athens	37.99749	23.762728	Bombing/Explosion	-	
4	1970	1	0	Japan	Fukouka	East Asia	Fukouka	33.580412	130.396361	Facility/Infrastructure Attack	-	
...	...	...	...	...	...	...	...	...	...	...	...	
181686	2017	12	31	Somalia	Middle Shebelle	Sub-Saharan Africa	Ceelka Geelow	2.359673	45.385034	Armed Assault	1.0	2
181687	2017	12	31	Syria	Lattakia	Middle East & North	Jableh	35.407278	35.942679	Bombing/Explosion	2.0	7

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 17 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   Country      181691 non-null  object
 4   state        181691 non-null  object
 5   Region       181691 non-null  object
 6   city         181691 non-null  object
 7   latitude     181691 non-null  object
 8   longitude    181691 non-null  object
 9   AttackType   181691 non-null  object
 10  Killed       181691 non-null  object
 11  Wounded      181691 non-null  object
 12  Target       181691 non-null  object
 13  Group        181691 non-null  object
 14  Target_type  181691 non-null  object
 15  Weapon_type  181691 non-null  object
 16  Motive       181691 non-null  object
dtypes: int64(3), object(14)
memory usage: 23.6+ MB
```

```
data.isnull()
```

	Year	Month	Day	Country	state	Region	city	latitude	longitude	AttackType	Killed	Wounded	Target	Group	Ta
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
181686	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
181687	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
181688	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
181689	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
181690	False	False	False	False	False	False	False	False	False	False	False	False	False	False	

181691 rows × 17 columns

```
data.isnull().sum()
```

```
Year           0
Month          0
Day            0
Country        0
```

```
state          0
Region         0
city           0
latitude       0
longitude      0
AttackType     0
Killed         0
Wounded        0
Target         0
Group          0
Target_type    0
Weapon_type    0
Motive         0
dtype: int64
```

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

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

```
print("City with the most attacks:",data['city'].value_counts().index[1]) #as first entry is 'unknown'
```

```
City with the most attacks: Baghdad
```

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

```
Region with the most attacks: Middle East & North Africa
```

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

```
Year with the most attacks: 2014
```

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

```
Month with the most attacks: 5
```

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

```
Group with the most attacks: Taliban
```

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

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

```
print("Country with the less attacks:",data['Country'].value_counts().idxmin())
print("Region with the less attacks:",data['Region'].value_counts().idxmin())
print("Year with the less attacks:",data['Year'].value_counts().idxmin())
print("Month with the less attacks:",data['Month'].value_counts().idxmin())
print("Less Attack Types:",data['AttackType'].value_counts().idxmin())
```

```
Country with the less attacks: Vatican City
Region with the less attacks: Australasia & Oceania
Year with the less attacks: 1971
Month with the less attacks: 0
Less Attack Types: Hijacking
```

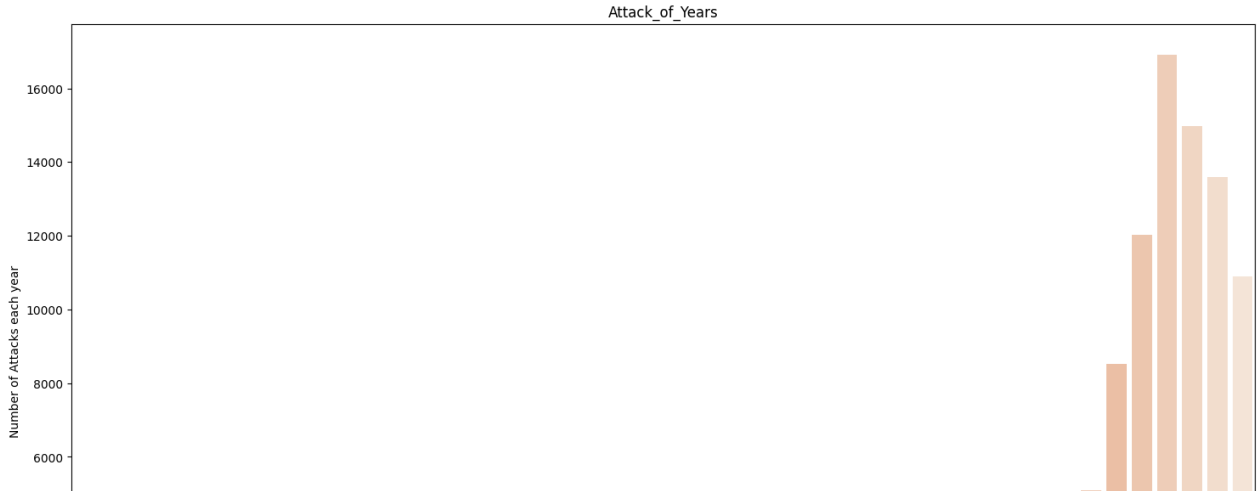
```
x_year = data['Year'].unique()
x_year
```

```
array([1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980,
       1981, 1986, 1982, 1983, 1984, 1985, 1987, 1988, 1989, 1990, 1991,
       1992, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,
       2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014,
       2015, 2016, 2017])
```

```
y_count_years = data['Year'].value_counts(dropna = False).sort_index()
y_count_years
```

```
1970      651
1971      471
1972      568
1973      473
1974      581
1975      740
1976      923
1977     1319
1978     1526
1979     2662
1980     2662
1981     2586
1982     2544
1983     2870
1984     3495
1985     2915
1986     2860
1987     3183
1988     3721
1989     4324
1990     3887
1991     4683
1992     5071
1994     3456
1995     3081
1996     3058
1997     3197
1998      934
1999     1395
2000     1814
2001     1906
2002     1333
2003     1278
2004     1166
2005     2017
2006     2758
2007     3242
2008     4805
2009     4721
2010     4826
2011     5076
2012     8522
2013    12036
2014    16903
2015    14965
2016    13587
2017    10900
Name: Year, dtype: int64
```

```
plt.figure(figsize = (18,10))
sns.barplot(x = x_year,
            y = y_count_years,
            palette = 'rocket')
plt.xticks(rotation = 45)
plt.xlabel('Attack Year')
plt.ylabel('Number of Attacks each year')
plt.title('Attack_of_Years')
plt.show()
```

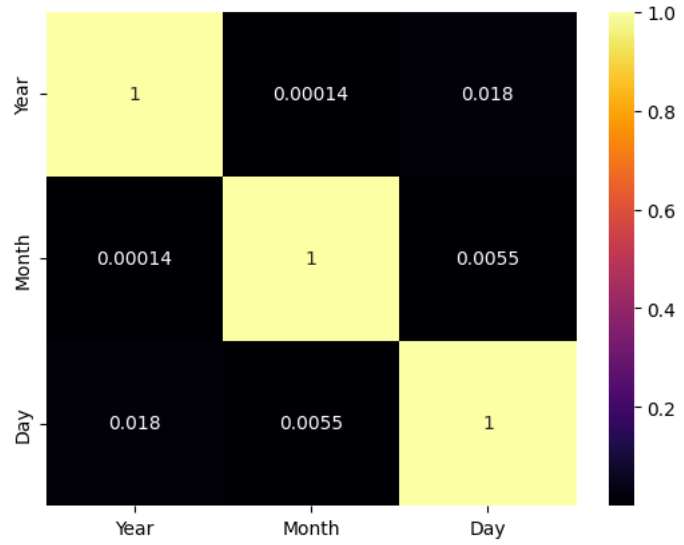



```
correlation=data.corr()
correlation
```

	Year	Month	Day
Year	1.000000	0.000139	0.018254
Month	0.000139	1.000000	0.005497
Day	0.018254	0.005497	1.000000

```
import seaborn as sns
sns.heatmap(data.corr(),annot=True,cmap='inferno')
```

```
<Axes: >
```




```
print(data["city"].unique())
```

```
['Santo Domingo' 'Mexico city' 'Unknown' ... 'Hungrum' 'Ceelka Geelow'
 'Kubentog']
```

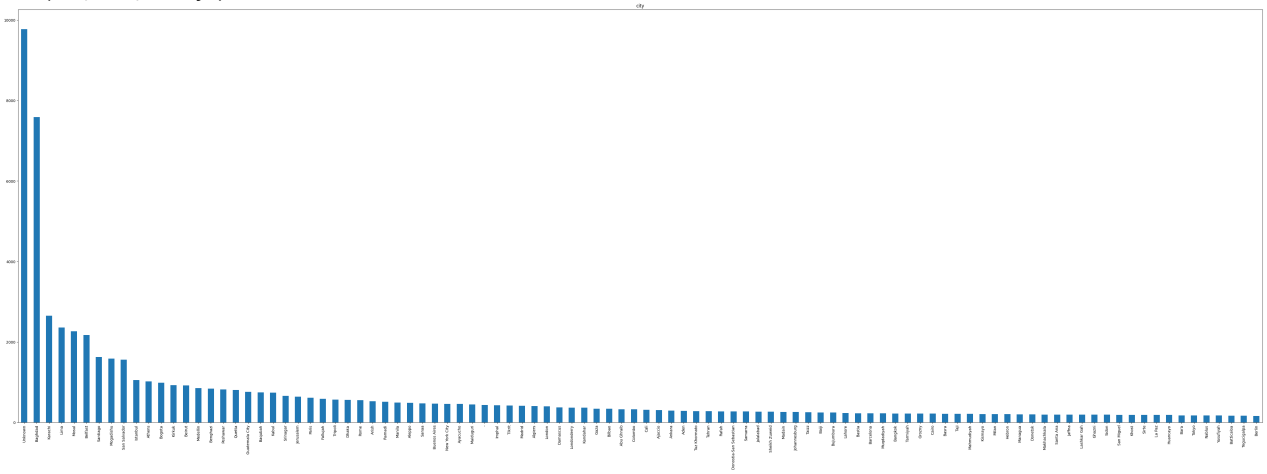
```
print(data["Killed"].unique())
```

```
[1.0 0.0 '-' 7.0 47.0 2.0 36.0 5.0 3.0 4.0 25.0 15.0 26.0 8.0 81.0 6.0 9.0
 16.0 30.0 31.0 12.0 21.0 14.0 88.0 11.0 10.0 27.0 18.0 22.0 19.0 92.0
 13.0 73.0 100.0 42.0 43.0 17.0 98.0 422.0 48.0 34.0 54.0 50.0 35.0 20.0
 41.0 37.0 28.0 40.0 32.0 85.0 23.0 300.0 60.0 24.0 58.0 70.0 87.0 45.0
 38.0 29.0 74.0 83.0 90.0 66.0 80.0 67.0 51.0 39.0 114.0 124.0 76.0 33.0
 75.0 57.0 62.0 46.0 56.0 63.0 120.0 102.0 78.0 79.0 52.0 77.0 200.0 49.0
 111.0 165.0 44.0 241.0 64.0 108.0 65.0 132.0 270.0 228.0 110.0 136.0
 180.0 250.0 93.0 130.0 59.0 94.0 53.0 146.0 329.0 97.0 71.0 240.0 227.0
 126.0 106.0 388.0 68.0 84.0 82.0 171.0 107.0 55.0 112.0 72.0 96.0 140.0
 61.0 105.0 150.0 115.0 89.0 1180.0 170.0 168.0 121.0 375.0 91.0 304.0
 123.0 135.0 256.0 109.0 271.0 206.0 104.0 320.0 275.0 224.0 118.0 129.0
 95.0 259.0 1384.0 1383.0 190.0 119.0 101.0 116.0 518.0 344.0 160.0 188.0
 103.0 205.0 145.0 153.0 127.0 69.0 141.0 134.0 400.0 86.0 184.0 210.0
 142.0 212.0 287.0 315.0 151.0 670.0 1570.0 310.0 298.0 953.0 517.0 201.0
 122.0 158.0 117.0 144.0 208.0 152.0 230.0 280.0 174.0 143.0 383.0 283.0
 154.0 284.0 433.0 266.0 133.0 163.0 128.0 588.0 311.0]
```

```
#sns.barplot(x = 'city',y ='Killed',data = data)
#plt.show()


t1 = data['city'].value_counts()[:100]
t1.plot(kind='bar',figsize=(60,20))
plt.title('city')
```

```
Text(0.5, 1.0, 'city')
```



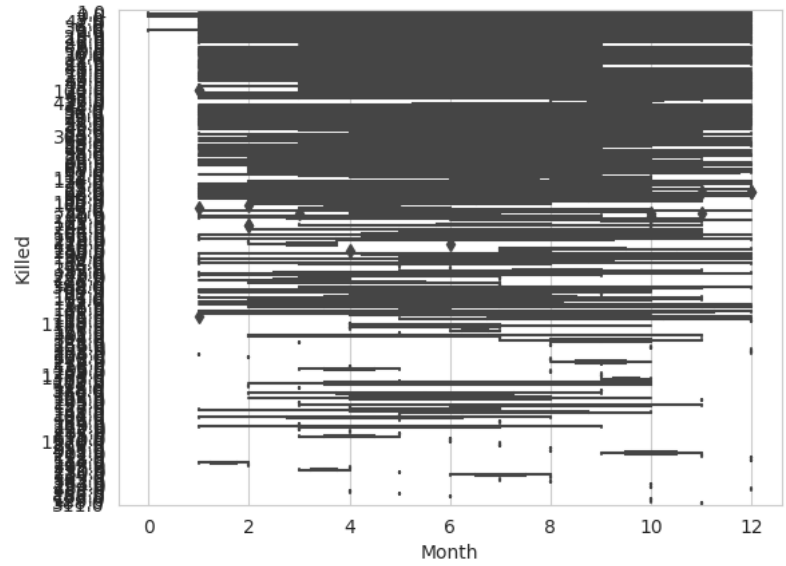
```
data['city'].value_counts()
```

```
Unknown                    9775
Baghdad                    7589
Karachi                    2652
Lima                       2359
Mosul                      2265
                           ...
Sbet                          1
Sukirin                       1
Dehiattakandiya district      1
Oued-Djemaa                   1
Kubentog                      1
Name: city, Length: 36675, dtype: int64
```

```
sns.set_style("whitegrid")

sns.boxplot(x = 'Month', y = 'Killed', data = data)
```

```
<Axes: xlabel='Month', ylabel='Killed'>
```




```
bplot = sns.boxplot(y='Killed', x='Year',
```

```
                data=data,
                width=0.5,
                palette="colorblind")
```