

```
In [6]: # import libraries

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

```
In [7]: df = pd.read_csv(r"C:\Users\sobie\Downloads\countries-by-intentional-homicid
df
```

```
Out[7]:
```

	Location	Region	Subregion	Rate	Count	Year
0	Afghanistan	Asia	Southern Asia	6.7	2474	2018
1	Albania	Europe	Southern Europe	2.1	61	2020
2	Algeria	Africa	Northern Africa	1.3	580	2020
3	Andorra	Europe	Southern Europe	2.6	2	2020
4	Angola	Africa	Middle Africa	4.8	1217	2012
...
190	Venezuela	Americas	South America	36.7	10598	2018
191	Vietnam	Asia	South-Eastern Asia	1.5	1358	2011
192	Yemen	Asia	Western Asia	6.8	1703	2013
193	Zambia	Africa	Eastern Africa	5.4	853	2015
194	Zimbabwe	Africa	Eastern Africa	7.5	981	2012

195 rows × 6 columns

```
In [8]: df.shape
```

```
Out[8]: (195, 6)
```

```
In [9]: df.dtypes
```

```
Out[9]: Location      object
Region      object
Subregion     object
Rate         float64
Count        int64
Year         int64
dtype: object
```

```
In [10]: df.isnull().sum()
```

```
Out[10]: Location      0
Region      0
Subregion    0
Rate        0
Count       0
Year        0
dtype: int64
```

```
In [11]: df.nunique()
```

```
Out[11]: Location      195  
Region          5  
Subregion       19  
Rate           103  
Count          151  
Year            16  
dtype: int64
```

```
In [14]: df1 = df.sort_values('Count', ascending = False).head(5)  
df1
```

```
Out[14]:
```

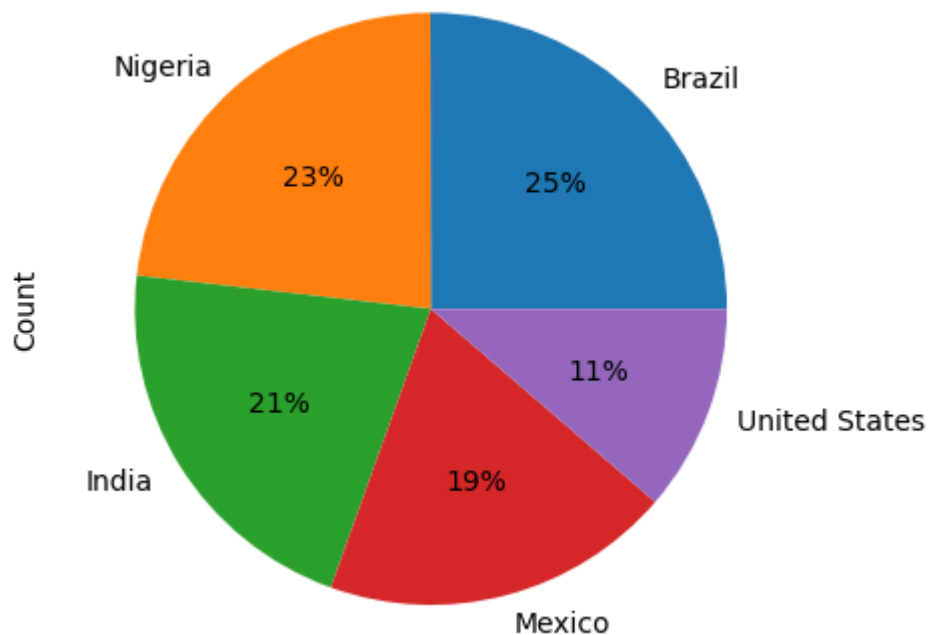
	Location	Region	Subregion	Rate	Count	Year
26	Brazil	Americas	South America	22.5	47722	2020
125	Nigeria	Africa	Western Africa	22.0	44200	2019
78	India	Asia	Southern Asia	3.0	40651	2020
111	Mexico	Americas	Central America	28.4	36579	2020
186	United States	Americas	Northern America	6.5	21570	2020

```
In [15]: df1 = df[['Location', 'Count']].sort_values(by = 'Count', ascending = False).  
df1
```

```
Out[15]:
```

	Location	Count
26	Brazil	47722
125	Nigeria	44200
78	India	40651
111	Mexico	36579
186	United States	21570

```
In [16]: df1.plot(x = 'Location', y = 'Count', kind = 'pie', autopct = '%1.0f%%', label  
plt.legend().set_visible(False)
```

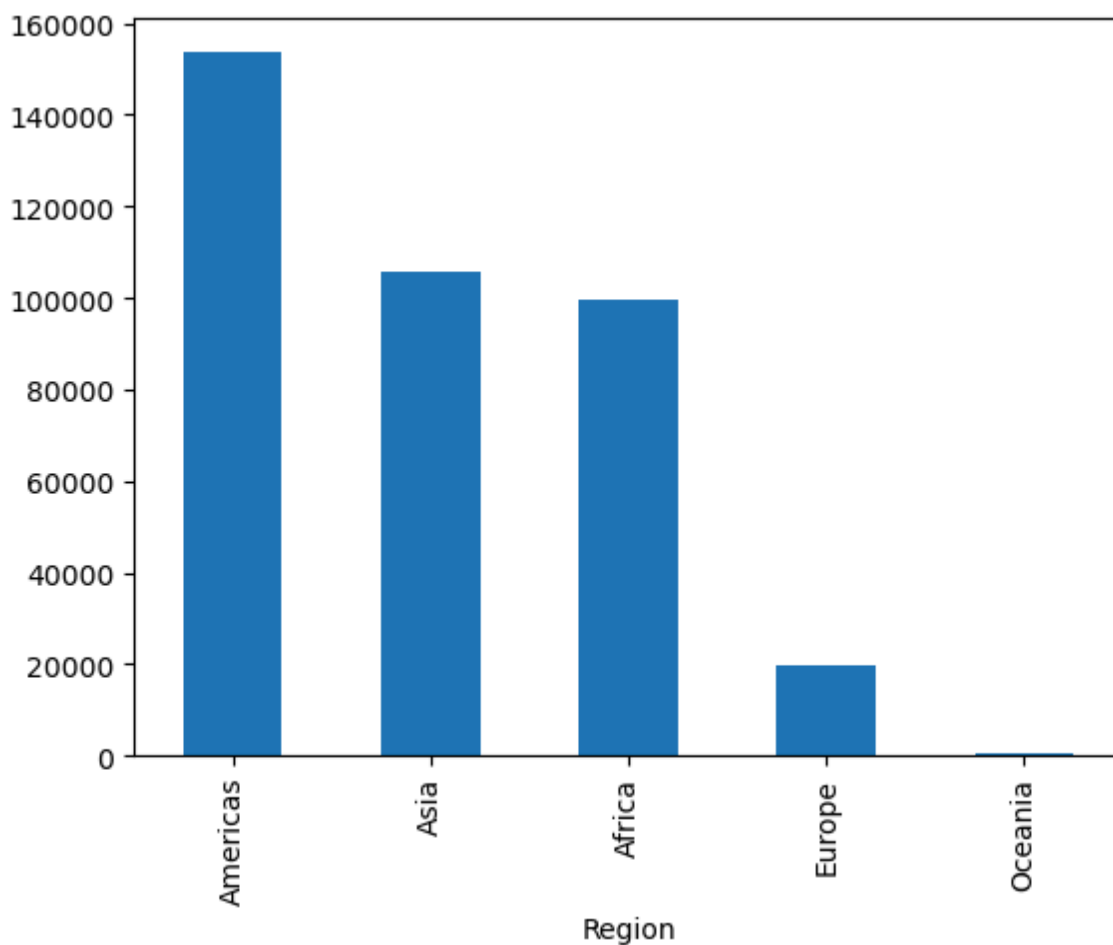


```
In [17]: df2 = df.groupby('Region')['Count'].sum().sort_values(ascending = False)  
df2
```

```
Out[17]: Region  
Americas    153597  
Asia        105552  
Africa       99481  
Europe       19869  
Oceania         347  
Name: Count, dtype: int64
```

```
In [18]: df2.plot(kind = 'bar')
```

```
Out[18]: <AxesSubplot:xlabel='Region'>
```



```
In [19]: df3 = df.groupby('Subregion')['Count'].sum().sort_values(ascending = False)
df3
```

```
Out[19]: Subregion
South America      78872
Southern Asia      58631
Central America    47371
Western Africa     46318
South-Eastern Asia 25314
Eastern Africa     23669
Northern America   22317
Southern Africa    21479
Eastern Europe     14604
Western Asia       11638
Eastern Asia       8563
Northern Africa    5538
Caribbean         5037
Middle Africa      2477
Northern Europe    2097
Western Europe     2075
Central Asia       1406
Southern Europe    1093
Australia, New Zealand 347
Name: Count, dtype: int64
```

```
In [20]: df3.index
```

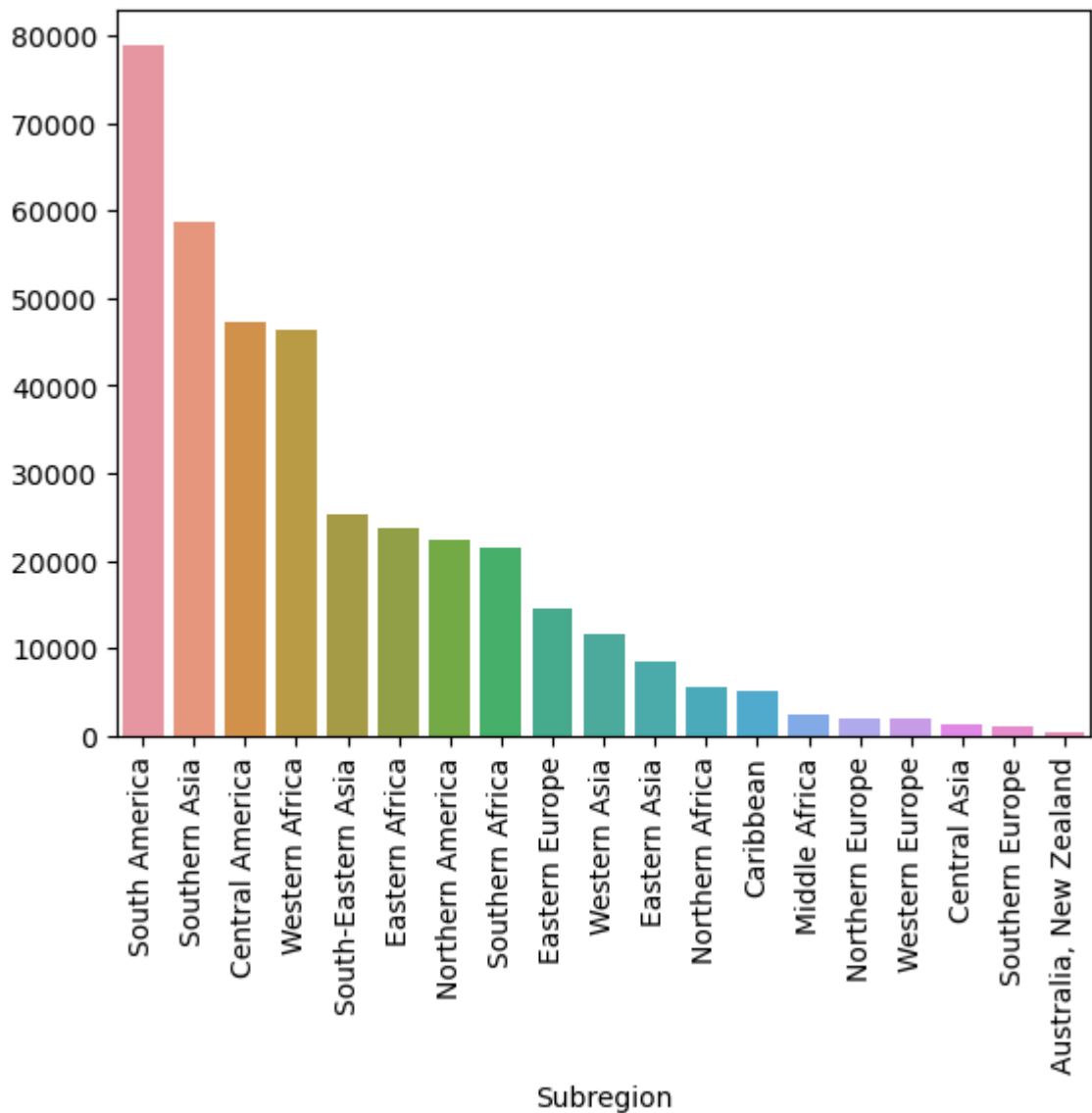
```
Out[20]: Index(['South America', 'Southern Asia', 'Central America', 'Western Africa',  
              'South-Eastern Asia', 'Eastern Africa', 'Northern America',  
              'Southern Africa', 'Eastern Europe', 'Western Asia', 'Eastern Asia',  
              'Northern Africa', 'Caribbean', 'Middle Africa', 'Northern Europe',  
              'Western Europe', 'Central Asia', 'Southern Europe',  
              'Australia, New Zealand'],  
              dtype='object', name='Subregion')
```

```
In [21]: df3.values
```

```
Out[21]: array([78872, 58631, 47371, 46318, 25314, 23669, 22317, 21479, 14604,  
              11638, 8563, 5538, 5037, 2477, 2097, 2075, 1406, 1093,  
              347], dtype=int64)
```

```
In [22]: sns.barplot(x = df3.index, y = df3.values)
plt.xticks(rotation = 90)
```

```
Out[22]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18])),
[Text(0, 0, 'South America'),
 Text(1, 0, 'Southern Asia'),
 Text(2, 0, 'Central America'),
 Text(3, 0, 'Western Africa'),
 Text(4, 0, 'South-Eastern Asia'),
 Text(5, 0, 'Eastern Africa'),
 Text(6, 0, 'Northern America'),
 Text(7, 0, 'Southern Africa'),
 Text(8, 0, 'Eastern Europe'),
 Text(9, 0, 'Western Asia'),
 Text(10, 0, 'Eastern Asia'),
 Text(11, 0, 'Northern Africa'),
 Text(12, 0, 'Caribbean'),
 Text(13, 0, 'Middle Africa'),
 Text(14, 0, 'Northern Europe'),
 Text(15, 0, 'Western Europe'),
 Text(16, 0, 'Central Asia'),
 Text(17, 0, 'Southern Europe'),
 Text(18, 0, 'Australia, New Zealand')]]
```



```
In [23]: df.Year.value_counts()
```

```
Out[23]: 2020    94
          2019    20
          2018    13
          2012    11
          2017    10
          2016     9
          2015     9
          2009     6
          2014     5
          2013     5
          2011     5
          2010     3
          2006     2
          2007     1
          2021     1
          2008     1
          Name: Year, dtype: int64
```

```
In [24]: df4 = df[df['Region'].isin(['Africa', 'Europe'])]
          df4
```

```
Out[24]:
```

	Location	Region	Subregion	Rate	Count	Year
1	Albania	Europe	Southern Europe	2.1	61	2020
2	Algeria	Africa	Northern Africa	1.3	580	2020
3	Andorra	Europe	Southern Europe	2.6	2	2020
4	Angola	Africa	Middle Africa	4.8	1217	2012
11	Austria	Europe	Western Europe	0.7	65	2020
...
182	Ukraine	Europe	Eastern Europe	6.2	2751	2017
184	United Kingdom	Europe	Northern Europe	1.1	755	2018
185	Tanzania	Africa	Eastern Africa	6.5	3439	2016
193	Zambia	Africa	Eastern Africa	5.4	853	2015
194	Zimbabwe	Africa	Eastern Africa	7.5	981	2012

91 rows × 6 columns

```
In [25]: df4 = df4[df4['Year'] > 2016][['Region', 'Year', 'Count']]
df4
```

```
Out[25]:
```

	Region	Year	Count
1	Europe	2020	61
2	Africa	2020	580
3	Europe	2020	2
11	Europe	2020	65
17	Europe	2019	225
...
170	Europe	2020	47
177	Africa	2020	562
181	Africa	2020	4460
182	Europe	2017	2751
184	Europe	2018	755

62 rows × 3 columns

```
In [26]: df4 = df4.groupby(['Region', 'Year']).sum()['Count']
df4
```

```
Out[26]:
```

Region	Year	Count
Africa	2017	1467
	2019	44210
	2020	28460
Europe	2017	3670
	2018	883
	2019	249
	2020	15066

Name: Count, dtype: int64

```
In [27]: df_unstacked = df4.unstack(level = 0)
df_unstacked
```

```
Out[27]:
```

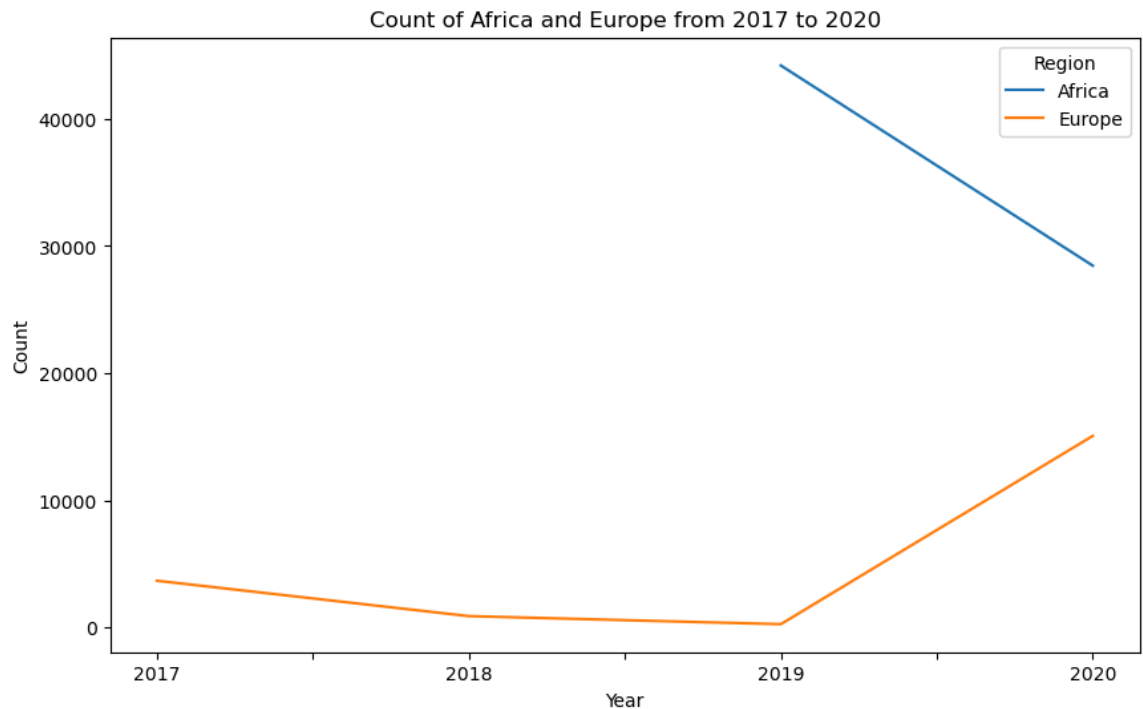
Region	Africa	Europe
Year		
2017	1467.0	3670.0
2018	NaN	883.0
2019	44210.0	249.0
2020	28460.0	15066.0

```
In [28]: df_unstacked.index = df_unstacked.index.astype(int).astype(str)
```



```
In [29]: df_unstacked.plot(kind = 'line', figsize =(10,6))
plt.xlabel('Year')
plt.ylabel('Count')
plt.title('Count of Africa and Europe from 2017 to 2020')
```

Out[29]: Text(0.5, 1.0, 'Count of Africa and Europe from 2017 to 2020')



```
In [30]: df5 = df.groupby(['Year'])['Rate'].sum().sort_values(ascending = False)
df5
```

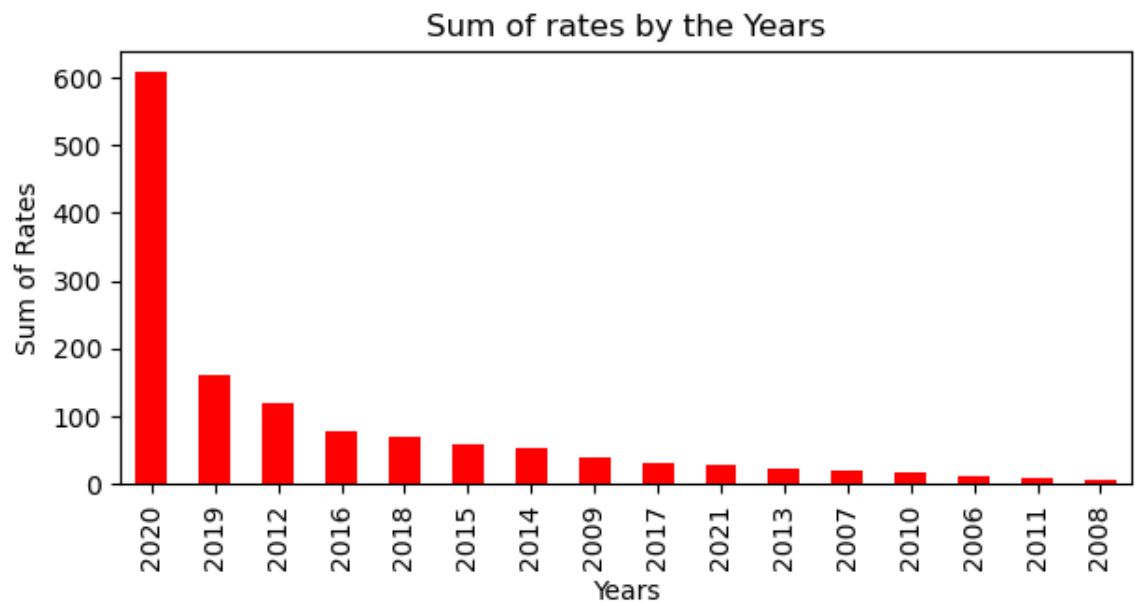
Out[30]:

Year	Rate
2020	608.0
2019	161.0
2012	119.5
2016	78.8
2018	70.0
2015	58.5
2014	54.2
2009	39.5
2017	30.3
2021	28.4
2013	21.7
2007	19.0
2010	18.2
2006	12.5
2011	10.1
2008	5.1


Name: Rate, dtype: float64

```
In [31]: df5.plot(kind = 'bar', figsize = (7,3), color = 'red')
plt.xlabel('Years')
plt.ylabel('Sum of Rates')
plt.title('Sum of rates by the Years')
```

Out[31]: Text(0.5, 1.0, 'Sum of rates by the Years')



```
In [32]: df6 = df[['Year', 'Region', 'Count']]  
df6 = df6.groupby(['Year', 'Region']).sum().sort_values(by = 'Year', ascending=True)  
df6
```

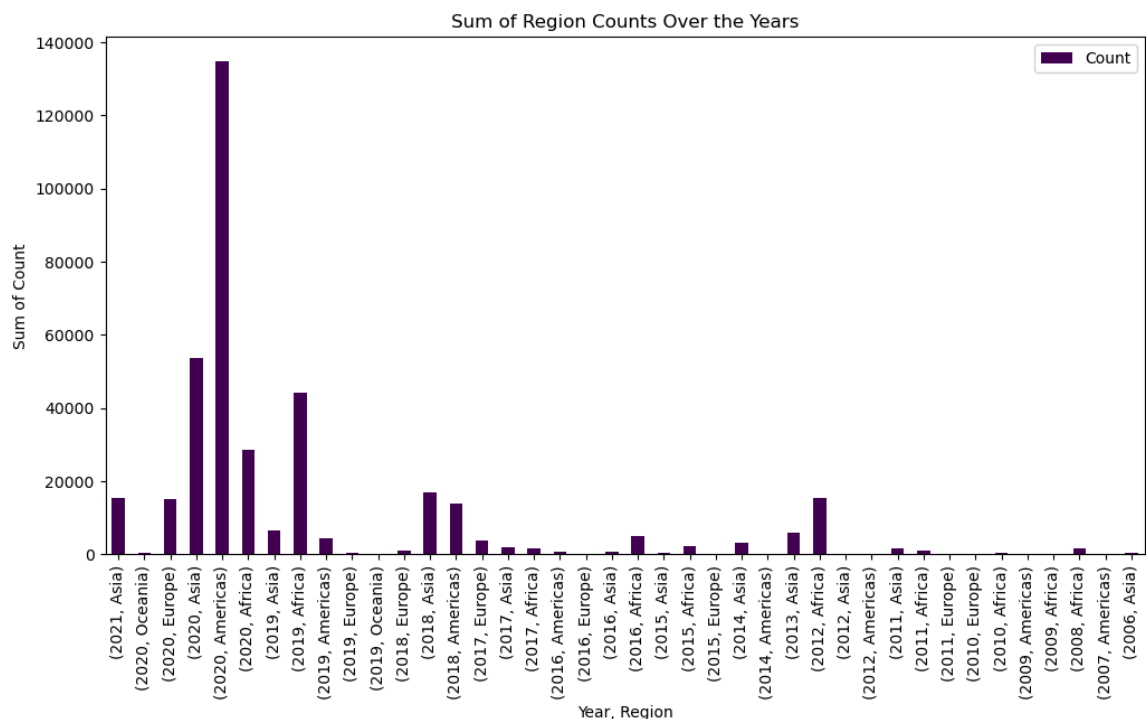


Out[32]:

		Count
Year	Region	
2021	Asia	15299
	Oceania	221
	Europe	15066
2020	Asia	53516
	Americas	134784
	Africa	28460
	Asia	6458
	Africa	44210
2019	Americas	4276
	Europe	249
	Oceania	126
	Europe	883
2018	Asia	16923
	Americas	13793
	Europe	3670
2017	Asia	1787
	Africa	1467
	Americas	608
	Europe	0
2016	Asia	627
	Africa	4987
	Asia	241
2015	Africa	2206
	Europe	0
	Asia	3029
2014	Americas	13
	Asia	5782
2013	Africa	15264
	Asia	61
2012	Americas	53
	Asia	1626
	Africa	855
2011	Europe	0
	Europe	1
	Africa	303
2010	Americas	42
	Africa	27
2009	Africa	1702
2008	Africa	1702

Count		
Year	Region	
2007	Americas	26
2006	Asia	203

```
In [33]: df6.plot(kind = 'bar', figsize = (12,6), colormap = 'viridis')
plt.xlabel('Year, Region')
plt.ylabel('Sum of Count')
plt.title('Sum of Region Counts Over the Years')
plt.show()
```



```
In [34]: df7 = df.groupby(['Subregion'])['Count'].mean().sort_values(ascending = False)
df7
```

```
Out[34]: Subregion
Southern Asia          6514.56
South America         6067.08
Central America       5921.38
Northern America      4463.40
Southern Africa       4295.80
Western Africa        4210.73
South-Eastern Asia    2531.40
Eastern Africa        1577.93
Eastern Europe        1460.40
Eastern Asia          1223.29
Northern Africa       1107.60
Middle Africa         619.25
Western Asia          581.90
Central Asia          281.20
Western Europe        230.56
Caribbean            201.48
Australia, New Zealand 173.50
Northern Europe       139.80
Southern Europe        64.29
Name: Count, dtype: float64
```

In [35]: df7.index

Out[35]: Index(['Southern Asia', 'South America', 'Central America', 'Northern America',
 'Southern Africa', 'Western Africa', 'South-Eastern Asia',
 'Eastern Africa', 'Eastern Europe', 'Eastern Asia', 'Northern Africa',
 'Middle Africa', 'Western Asia', 'Central Asia', 'Western Europe',
 'Caribbean', 'Australia, New Zealand', 'Northern Europe',
 'Southern Europe'],
 dtype='object', name='Subregion')

In [36]: df7.values

Out[36]: array([6514.56, 6067.08, 5921.38, 4463.4 , 4295.8 , 4210.73, 2531.4 ,
 1577.93, 1460.4 , 1223.29, 1107.6 , 619.25, 581.9 , 281.2 ,
 230.56, 201.48, 173.5 , 139.8 , 64.29])


In [37]: data = {'Category': df7.index, 'Value': df7.values, 'Info': df7.values}
 df = pd.DataFrame(data)
 df

Out[37]:

	Category	Value	Info
0	Southern Asia	6514.56	6514.56
1	South America	6067.08	6067.08
2	Central America	5921.38	5921.38
3	Northern America	4463.40	4463.40
4	Southern Africa	4295.80	4295.80
5	Western Africa	4210.73	4210.73
6	South-Eastern Asia	2531.40	2531.40
7	Eastern Africa	1577.93	1577.93
8	Eastern Europe	1460.40	1460.40
9	Eastern Asia	1223.29	1223.29
10	Northern Africa	1107.60	1107.60
11	Middle Africa	619.25	619.25
12	Western Asia	581.90	581.90
13	Central Asia	281.20	281.20
14	Western Europe	230.56	230.56
15	Caribbean	201.48	201.48
16	Australia, New Zealand	173.50	173.50
17	Northern Europe	139.80	139.80
18	Southern Europe	64.29	64.29

In []: pip install notebook --upgrade

```
In [ ]: fig = px.treemap(df, path = ['Category'], values = 'Value', title = 'Treemap')
fig.update_traces(hovertemplate = 'Category:%{label}<br>Value:%{value}')
fig.show()
```



```
In [ ]:
```

