

```
In [2]: pip install cufflinks
```

```
Collecting cufflinks
```

```
  Downloading cufflinks-0.17.3.tar.gz (81 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: numpy>=1.9.2 in c:\python\lib\site-packages (from cufflinks) (1.23.5)
Requirement already satisfied: pandas>=0.19.2 in c:\python\lib\site-packages (from cufflinks) (2.1.2)
Requirement already satisfied: plotly>=4.1.1 in c:\python\lib\site-packages (from cufflinks) (5.18.0)
Requirement already satisfied: six>=1.9.0 in c:\python\lib\site-packages (from cufflinks) (1.16.0)
Collecting colorlover>=0.2.1 (from cufflinks)
  Downloading colorlover-0.3.0-py3-none-any.whl.metadata (421 bytes)
Requirement already satisfied: setuptools>=34.4.1 in c:\python\lib\site-packages (from cufflinks) (65.5.0)
Requirement already satisfied: ipython>=5.3.0 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from cufflinks) (8.32.0)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\python\lib\site-packages (from cufflinks) (8.1.6)
Requirement already satisfied: colorama in c:\python\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.4.6)
Requirement already satisfied: decorator in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (0.19.2)
Requirement already satisfied: matplotlib-inline in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (0.1.7)
Requirement already satisfied: prompt_toolkit<3.1.0,>=3.0.41 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (3.0.50)
Requirement already satisfied: pygments>=2.4.0 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (2.19.1)
Requirement already satisfied: stack_data in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (0.6.3)
Requirement already satisfied: traitlets>=5.13.0 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipython>=5.3.0->cufflinks) (5.14.3)
Requirement already satisfied: typing_extensions>=4.6 in c:\python\lib\site-packages (from ipython>=5.3.0->cufflinks) (4.7.0)
Requirement already satisfied: comm>=0.1.3 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from ipywidgets>=7.0.0->cufflinks) (0.2.2)
Requirement already satisfied: widgetsnbextension~=4.0.14 in c:\python\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (4.0.14)
Requirement already satisfied: jupyterlab_widgets~=3.0.14 in c:\python\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (3.0.14)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\python\lib\site-packages (from pandas>=0.19.2->cufflinks) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\python\lib\site-packages (from pandas>=0.19.2->cufflinks) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\python\lib\site-packages (from pandas>=0.19.2->cufflinks) (2023.
```

```
3)
Requirement already satisfied: tenacity>=6.2.0 in c:\python\lib\site-packages (from plotly>=4.1.1->cufflinks) (8.2.3)
Requirement already satisfied: packaging in c:\python\lib\site-packages (from plotly>=4.1.1->cufflinks) (23.1)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from jedi>=0.16->ipython>=5.3.0->cufflinks) (0.8.4)
Requirement already satisfied: wcwidth in c:\python\lib\site-packages (from prompt_toolkit<3.1.0,>=3.0.41->ipython>=5.3.0->cufflinks) (0.2.12)
Requirement already satisfied: executing>=1.2.0 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from stack_data->ipython>=5.3.0->cufflinks) (2.2.0)
Requirement already satisfied: asttokens>=2.1.0 in c:\users\shashank\appdata\roaming\python\python311\site-packages (from stack_data->ipython>=5.3.0->cufflinks) (3.0.0)
Requirement already satisfied: pure-eval in c:\users\shashank\appdata\roaming\python\python311\site-packages (from stack_data->ipython>=5.3.0->cufflinks) (0.2.3)
Downloading colorlover-0.3.0-py3-none-any.whl (8.9 kB)
Building wheels for collected packages: cufflinks
  Building wheel for cufflinks (setup.py): started
  Building wheel for cufflinks (setup.py): finished with status 'done'
  Created wheel for cufflinks: filename=cufflinks-0.17.3-py3-none-any.whl size=68727 sha256=9dc780caa70ee86e200dddc85e4b31f7028c03b2e4296dd2d345a22f92bc32a0
  Stored in directory: c:\users\shashank\appdata\local\pip\cache\wheels\c3\bf\8f\afa0730b29e347e8c1071049f89748c1262e7a14a838820ae6
Successfully built cufflinks
Installing collected packages: colorlover, cufflinks
Successfully installed colorlover-0.3.0 cufflinks-0.17.3
Note: you may need to restart the kernel to use updated packages.
```

```
[notice] A new release of pip is available: 25.0.1 -> 25.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import plotly.express as px
import cufflinks as cf
import plotly.io as pio

import os
```

```
In [18]: print(os.getcwd())
```

```
c:\Users\SHASHANK\OneDrive\Desktop\Evelate labs\indian_army_eda\notebooks
```

```
In [4]: cf.go_offline()
pio.renderers.default = 'notebook'
```

```
In [ ]: data_path = '../Data/cleaned_globalterrorism.csv'
df = pd.read_csv(data_path)

df.head()
```

```
C:\Users\SHASHANK\AppData\Local\Temp\ipykernel_6708\2489049055.py:6: DtypeWarning:
```

```
Columns (3,5,27,29,57,58,59,72,75,86,88,90,92,110,111,117) have mixed types. Specify dtype option on import or set low_memory=False.
```

```
Out[ ]:   iyear  imonth  iday  approxdate  extended  resolution  country  country_txt  region  region_txt  ...  addnotes  scite1  scite2
0  1970        7       2        NaN        0        NaN       58  Dominican Republic      2  Central America & Caribbean  ...        NaN        NaN        NaN
1  1970        0       0        NaN        0        NaN      130      Mexico          1  North America  ...        NaN        NaN        NaN
2  1970        1       0        NaN        0        NaN      160  Philippines          5  Southeast Asia  ...        NaN        NaN        NaN
3  1970        1       0        NaN        0        NaN       78      Greece          8  Western Europe  ...        NaN        NaN        NaN
4  1970        1       0        NaN        0        NaN      101      Japan          4  East Asia  ...        NaN        NaN        NaN
```

5 rows × 131 columns



```
In [7]: df.info()
df.describe()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 131 entries, iyear to related
dtypes: float64(54), int64(21), object(56)
memory usage: 181.6+ MB

Out[7]: iyear          0
         imonth        0
         iday          0
         approxdate   172452
         extended       0
         ...
         INT_LOG        0
         INT_IDEO        0
         INT_MISC        0
         INT_ANY          0
         related      156653
Length: 131, dtype: int64
```

```
In [8]: fig = px.pie(df, names='attacktype1_txt', title='Distribution of Attack Types')
fig.write_html('../output/attack_type_distribution.html')
fig.show()
```

```
In [ ]: attacks_per_year = df['iyear'].value_counts().sort_index()

fig = px.line(x=attacks_per_year.index,
              y=attacks_per_year.values,
              title='Number of Terrorist Attacks Each Year',
              markers=True,
              labels={'x': 'Year', 'y': 'Number of Attacks'})

fig.write_html('../output/yearly_attacks.html')

fig.show()
```

```
In [11]: top_countries = df['country_txt'].value_counts().nlargest(10)
fig = px.bar(x=top_countries.index, y=top_countries.values, title='Top 10 Countries with Most Attacks', labels={'x':
```

```
fig.write_html('../output/top_countries.html')
fig.show()
```

```
In [12]: heatmap_data = pd.crosstab(df['country_txt'], df['attacktype1_txt'])
fig = px.imshow(heatmap_data, title='Heatmap: Country vs Attack Type')
fig.write_html('../output/country_attack_heatmap.html')
fig.show()
```

```
In [16]: pip install kaleido
```

Requirement already satisfied: kaleido in c:\python\lib\site-packages (0.2.1)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 25.0.1 -> 25.1
[notice] To update, run: python.exe -m pip install --upgrade pip

```
In [25]: print(df.describe())
print(df.info())
print(df['attacktype1_txt'].value_counts())
```

	iyear	imonth	iday	extended	\	
count	181691.000000	181691.000000	181691.000000	181691.000000		
mean	2002.638997	6.467277	15.505644	0.045346		
std	13.259430	3.388303	8.814045	0.208063		
min	1970.000000	0.000000	0.000000	0.000000		
25%	1991.000000	4.000000	8.000000	0.000000		
50%	2009.000000	6.000000	15.000000	0.000000		
75%	2014.000000	9.000000	23.000000	0.000000		
max	2017.000000	12.000000	31.000000	1.000000		
	country	region	latitude	longitude	\	
count	181691.000000	181691.000000	177135.000000	1.771340e+05		
mean	131.968501	7.160938	23.498343	-4.586957e+02		
std	112.414535	2.933408	18.569242	2.047790e+05		
min	4.000000	1.000000	-53.154613	-8.618590e+07		
25%	78.000000	5.000000	11.510046	4.545640e+00		
50%	98.000000	6.000000	31.467463	4.324651e+01		
75%	160.000000	10.000000	34.685087	6.871033e+01		
max	1004.000000	12.000000	74.633553	1.793667e+02		
	specificity	vicinity	...	ransomamt	ransomamtus	\
count	181685.000000	181691.000000	...	1.350000e+03	5.630000e+02	
mean	1.451452	0.068297	...	3.172530e+06	5.784865e+05	
std	0.995430	0.284553	...	3.021157e+07	7.077924e+06	
min	1.000000	-9.000000	...	-9.900000e+01	-9.900000e+01	
25%	1.000000	0.000000	...	0.000000e+00	0.000000e+00	
50%	1.000000	0.000000	...	1.500000e+04	0.000000e+00	
75%	1.000000	0.000000	...	4.000000e+05	0.000000e+00	
max	5.000000	1.000000	...	1.000000e+09	1.320000e+08	
	ransompaid	ransompaidus	hostkidoutcome	nreleased	\	
count	7.740000e+02	552.000000	10991.000000	10400.000000		
mean	7.179437e+05	240.378623	4.629242	-29.018269		
std	1.014392e+07	2940.967293	2.035360	65.720119		
min	-9.900000e+01	-99.000000	1.000000	-99.000000		
25%	-9.900000e+01	0.000000	2.000000	-99.000000		
50%	0.000000e+00	0.000000	4.000000	0.000000		
75%	1.273412e+03	0.000000	7.000000	1.000000		
max	2.750000e+08	48000.000000	7.000000	2769.000000		
	INT_LOG	INT_IDEO	INT_MISC	INT_ANY		
count	181691.000000	181691.000000	181691.000000	181691.000000		

```
mean      -4.543731      -4.464398      0.090010     -3.945952
std       4.543547      4.637152      0.568457     4.691325
min      -9.000000     -9.000000     -9.000000    -9.000000
25%     -9.000000     -9.000000      0.000000    -9.000000
50%     -9.000000     -9.000000      0.000000     0.000000
75%      0.000000      0.000000      0.000000     0.000000
max      1.000000      1.000000      1.000000     1.000000
```

```
[8 rows x 75 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 131 entries, iyear to related
dtypes: float64(54), int64(21), object(56)
memory usage: 181.6+ MB
None
attacktype1_txt
Bombing/Explosion           88255
Armed Assault                42669
Assassination                 19312
Hostage Taking (Kidnapping)   11158
Facility/Infrastructure Attack 10356
Unknown                        7276
Unarmed Assault                1015
Hostage Taking (Barricade Incident) 991
Hijacking                      659
Name: count, dtype: int64
```

```
In [ ]: fig_nkill = px.histogram(
            df,
            x='nkill',
            nbins=30,
            title='Distribution of Number of Kills',
            labels={'nkill': 'Number of Kills'},
            marginal='box',
            opacity=0.75
        )

fig_nkill.update_layout(
    template='plotly_white',
    title={'x':0.5, 'xanchor':'center'},
    xaxis=dict(tickmode='linear')
```

```
)  
  
fig.write_html('../output/distribution of number of kills.html')  
  
fig_nkill.show()
```

```
In [ ]: data_path = '../Data/cleaned_globalterrorism.csv'  
df = pd.read_csv(data_path)  
  
print(df.isnull().sum())  
  
df = df.dropna()  
  
print(df.info())
```

```
C:\Users\SHASHANK\AppData\Local\Temp\ipykernel_6708\2926496265.py:2: DtypeWarning:  
Columns (3,5,27,29,57,58,59,72,75,86,88,90,92,110,111,117) have mixed types. Specify dtype option on import or set low_memory=False.
```

```
iyear          0
imonth         0
iday           0
approxdate    172452
extended       0
...
INT_LOG        0
INT_IDEO       0
INT_MISC       0
INT_ANY        0
related        156653
Length: 131, dtype: int64
<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Columns: 131 entries, iyear to related
dtypes: float64(54), int64(21), object(56)
memory usage: 0.0+ bytes
None
```

```
In [ ]: print(df.describe())
print(df['attacktype1_txt'].value_counts())
print(df.head())
```

	iyear	imonth	iday	extended	country	region	latitude	longitude	\
count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	specificity	vicinity	...	ransomamt	ransomamtus	ransompaid	\		
count	0.0	0.0	...	0.0	0.0	0.0			
mean	NaN	NaN	...	NaN	NaN	NaN			
std	NaN	NaN	...	NaN	NaN	NaN			
min	NaN	NaN	...	NaN	NaN	NaN			
25%	NaN	NaN	...	NaN	NaN	NaN			
50%	NaN	NaN	...	NaN	NaN	NaN			
75%	NaN	NaN	...	NaN	NaN	NaN			
max	NaN	NaN	...	NaN	NaN	NaN			
	ransompaidus	hostkidoutcome	nreleased	INT_LOG	INT_IDEO	INT_MISC	\		
count	0.0	0.0	0.0	0.0	0.0	0.0			
mean	NaN	NaN	NaN	NaN	NaN	NaN			
std	NaN	NaN	NaN	NaN	NaN	NaN			
min	NaN	NaN	NaN	NaN	NaN	NaN			
25%	NaN	NaN	NaN	NaN	NaN	NaN			
50%	NaN	NaN	NaN	NaN	NaN	NaN			
75%	NaN	NaN	NaN	NaN	NaN	NaN			
max	NaN	NaN	NaN	NaN	NaN	NaN			
	INT_ANY								
count	0.0								
mean	NaN								
std	NaN								
min	NaN								
25%	NaN								
50%	NaN								
75%	NaN								
max	NaN								

[8 rows x 75 columns]
Series([], Name: count, dtype: int64)

Empty DataFrame

Columns: [iyear, imonth, iday, approxdate, extended, resolution, country, country_txt, region, region_txt, provstate, city, latitude, longitude, specificity, vicinity, location, crit1, crit2, crit3, doubtterr, multiple, success, suicide, attacktype1, attacktype1_txt, attacktype2, attacktype2_txt, attacktype3, attacktype3_txt, targtype1, targtype1_tx t, targsubtype1, targsubtype1_txt, corp1, target1, natlty1, natlty1_txt, targtype2, targtype2_txt, targsubtype2, targ subtype2_tx t, corp2, target2, natlty2, natlty2_txt, targtype3, targtype3_tx t, targsubtype3, targsubtype3_tx t, corp3, target3, natlty3, natlty3_tx t, gname, gsubname, gname2, gsubname2, gname3, gsubname3, motive, guncertain1, guncertain 2, guncertain3, individual, nperps, nperpcap, claimed, claimmode, claimmode_tx t, claim2, claimmode2, claimmode2_tx t, claim3, claimmode3, claimmode3_tx t, compclaim, weaptype1, weaptype1_tx t, weapsubtype1, weapsubtype1_tx t, weaptype2, weaptype2_tx t, weapsubtype2, weapsubtype2_tx t, weaptype3, weaptype3_tx t, weapsubtype3, weapsubtype3_tx t, weaptype4, weaptype4_tx t, weapsubtype4, weapsubtype4_tx t, weapdetail, nkill, nkillus, nkillter, nwound, nwoundus, nwoundde, ...]

Index: []

[0 rows x 131 columns]

```
In [ ]: print("Data Shape:", df.shape)
print("Column Names:", df.columns)

print("Distribution of Attack Types:\n", df['attacktype1_txt'].value_counts())

print("Correlation Matrix:\n", df.corr())
```

```
Data Shape: (0, 131)
Column Names: Index(['iyear', 'imonth', 'iday', 'approxdate', 'extended', 'resolution',
       'country', 'country_txt', 'region', 'region_txt',
       ...
       'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG',
       'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'],
      dtype='object', length=131)
```

Distribution of Attack Types:

```
Series([], Name: count, dtype: int64)
```

Correlation Matrix:

	iyear	imonth	iday	approxdate	extended	resolution	country	\
iyear	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
imonth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
iday	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
approxdate	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
extended	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	
INT_LOG	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
INT_IDEO	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
INT_MISC	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
INT_ANY	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
related	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	country_txt	region	region_txt	...	addnotes	scite1	scite2	\
iyear	NaN	NaN	NaN	...	NaN	NaN	NaN	
imonth	NaN	NaN	NaN	...	NaN	NaN	NaN	
iday	NaN	NaN	NaN	...	NaN	NaN	NaN	
approxdate	NaN	NaN	NaN	...	NaN	NaN	NaN	
extended	NaN	NaN	NaN	...	NaN	NaN	NaN	
...	
INT_LOG	NaN	NaN	NaN	...	NaN	NaN	NaN	
INT_IDEO	NaN	NaN	NaN	...	NaN	NaN	NaN	
INT_MISC	NaN	NaN	NaN	...	NaN	NaN	NaN	
INT_ANY	NaN	NaN	NaN	...	NaN	NaN	NaN	
related	NaN	NaN	NaN	...	NaN	NaN	NaN	

	scite3	dbsource	INT_LOG	INT_IDEO	INT_MISC	INT_ANY	related
iyear	NaN	NaN	NaN	NaN	NaN	NaN	NaN
imonth	NaN	NaN	NaN	NaN	NaN	NaN	NaN
iday	NaN	NaN	NaN	NaN	NaN	NaN	NaN
approxdate	NaN	NaN	NaN	NaN	NaN	NaN	NaN
extended	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
...     ...     ...     ...     ...     ...     ...     ...
INT_LOG      NaN      NaN      NaN      NaN      NaN      NaN      NaN
INT_IDEO      NaN      NaN      NaN      NaN      NaN      NaN      NaN
INT_MISC      NaN      NaN      NaN      NaN      NaN      NaN      NaN
INT_ANY       NaN      NaN      NaN      NaN      NaN      NaN      NaN
related      NaN      NaN      NaN      NaN      NaN      NaN      NaN
```

[131 rows x 131 columns]

```
In [ ]: corr_matrix = df.corr()
```

```
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
plt.title('Correlation Heatmap of Numeric Features')
fig.write_html('../output/correlation_heatmap.html')
plt.savefig('../output/correlation_heatmap.png', dpi=300, bbox_inches='tight')
plt.show()
```

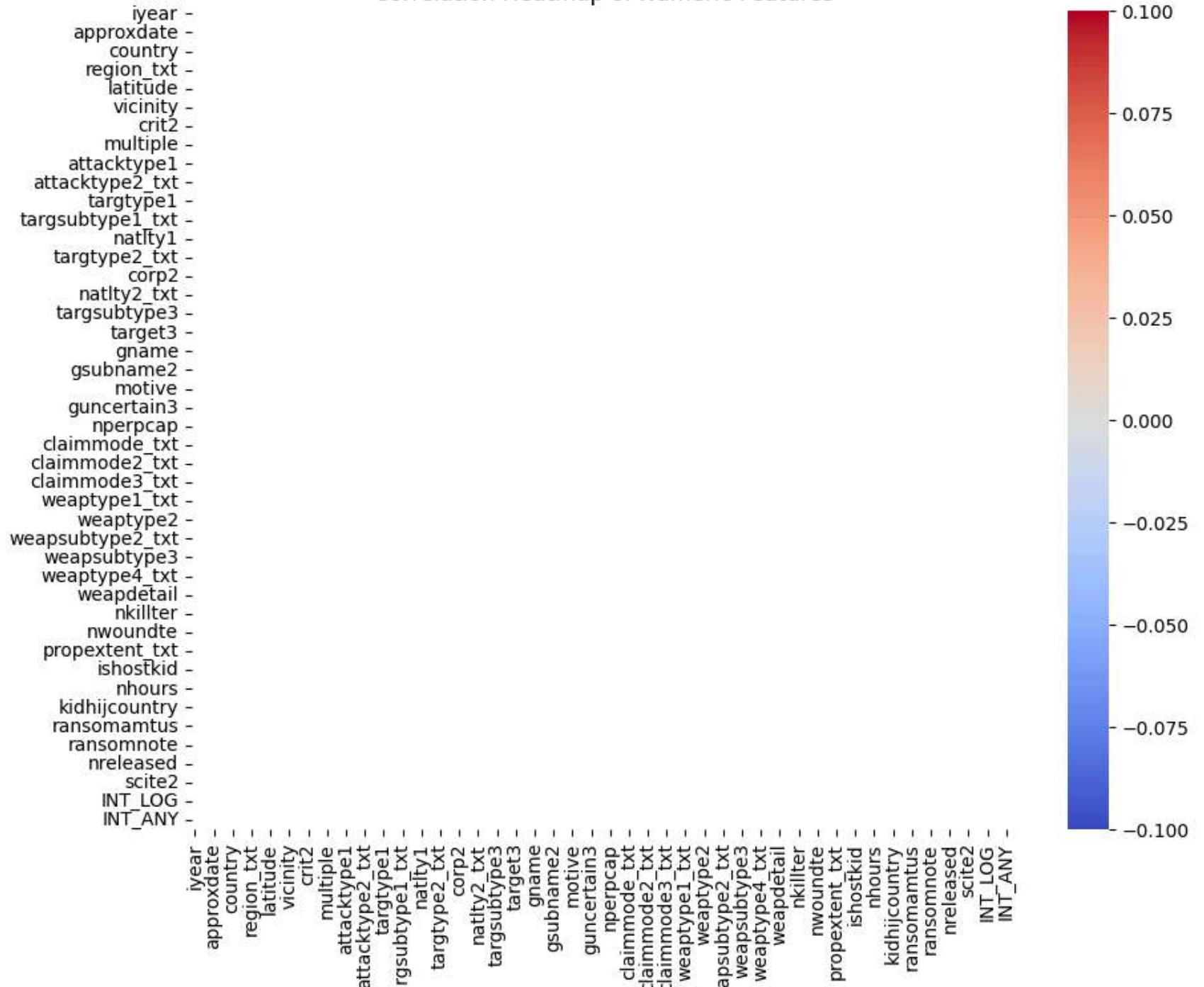
```
c:\python\Lib\site-packages\seaborn\matrix.py:202: RuntimeWarning:
```

All-NaN slice encountered

```
c:\python\Lib\site-packages\seaborn\matrix.py:207: RuntimeWarning:
```

All-NaN slice encountered

Correlation Heatmap of Numeric Features



te

-
-

we