

Proyecto Final: Terrorismo Global

Bloque 4: Análisis de Datos con Python

Equipo 1

TEAM MEMBERS



Fernando Guevara



Gil Rodríguez



Ana Rosado



Karla Gómez



José Muñoz

Postwork 01

Obteniendo estimados de locación y variabilidad

1.- Identificar las columnas relevantes para nuestro proyecto.

2.- Obtención de estimados relevantes (promedio, mediana, media truncada, desv. estándar, rango, percentiles, etc.).

iYear, iMonth, iDay (tiempo)



latitude, longitude (ubicación)



```
[6] # El promedio de las columnas  
mean = df_pw_1.mean()  
mean
```

iyear	2002.706499
imonth	6.467478
iday	15.582058
latitude	23.503621
longitude	-460.809929
nkill	2.406442
nwound	3.176587
dtype:	float64

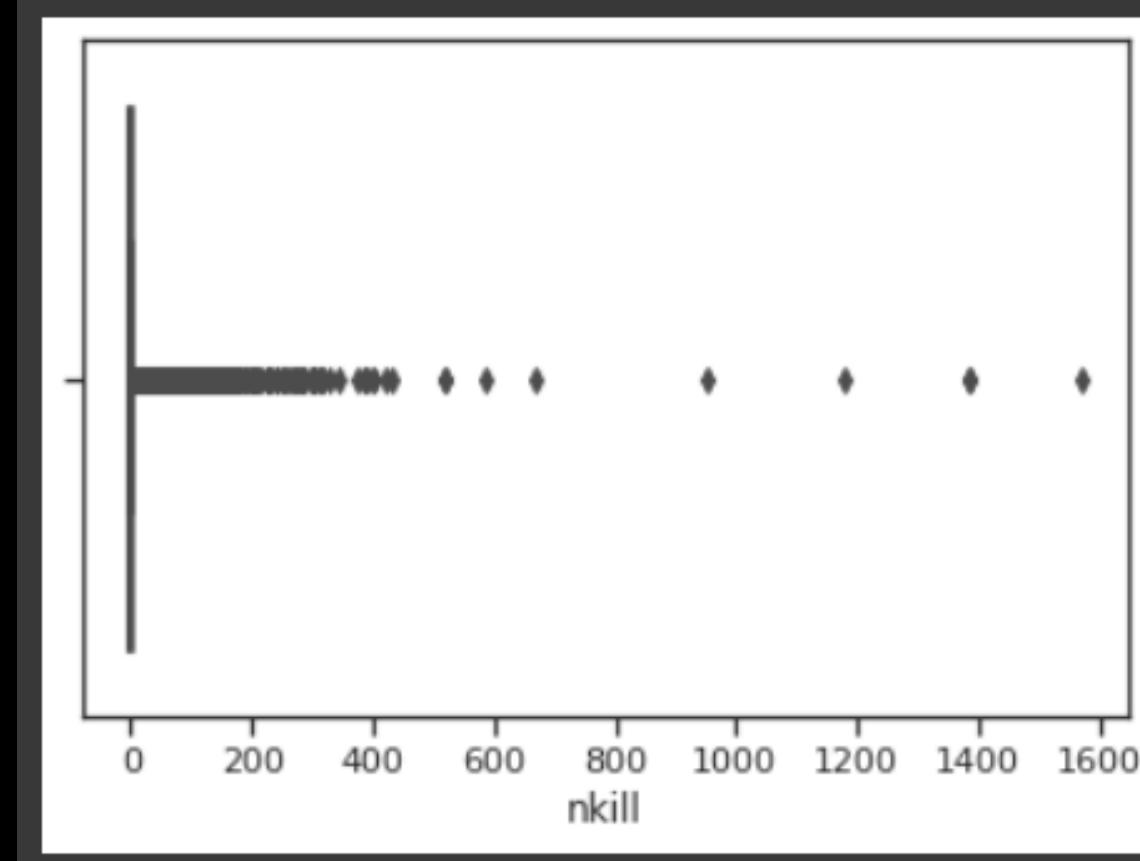
```
[9] # La desviación estándar  
std = df_pw_1.std()  
std
```

iyear	13.229239
imonth	3.386879
iday	8.768099
latitude	18.556458
longitude	205245.316309
nkill	11.566116
nwound	36.025025
dtype:	float64

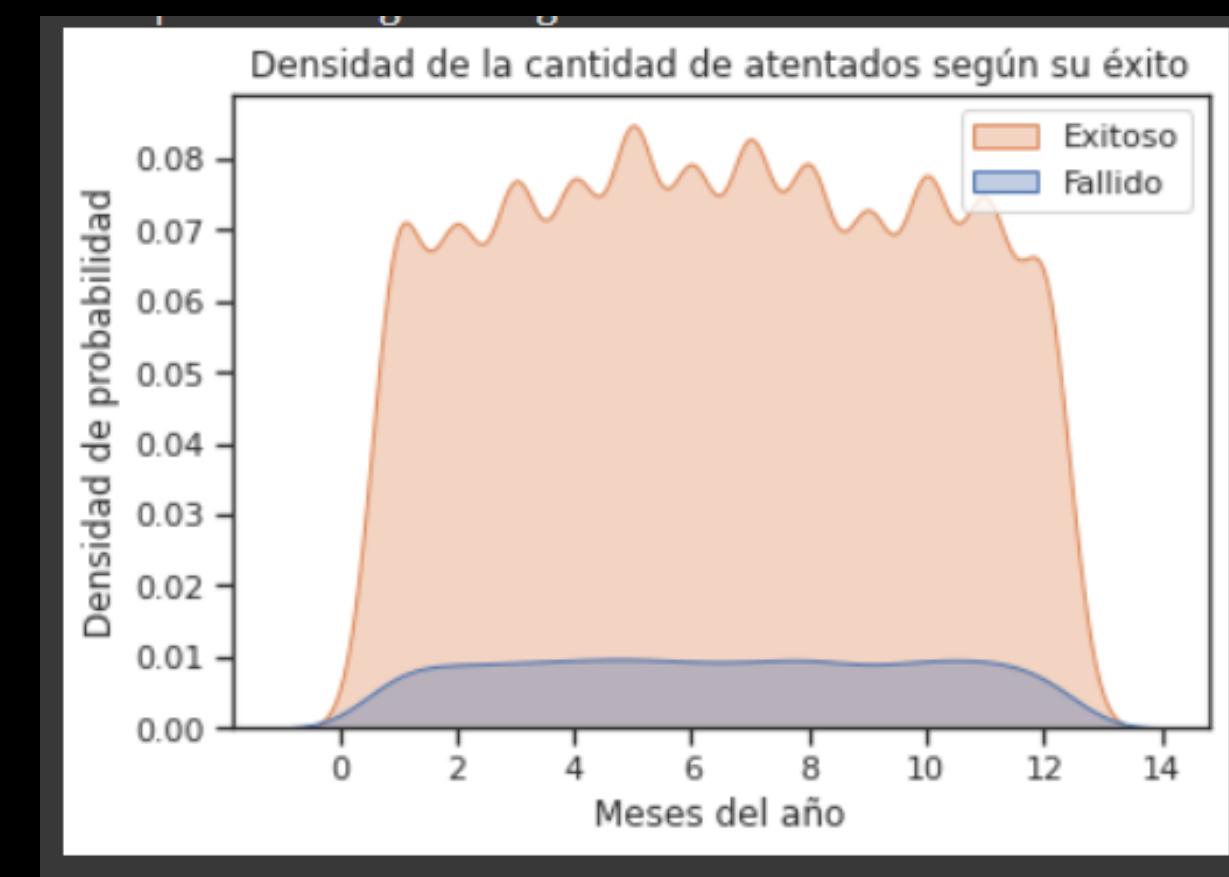
Postwork 02

Introducción a la visualización de datos: Distribuciones

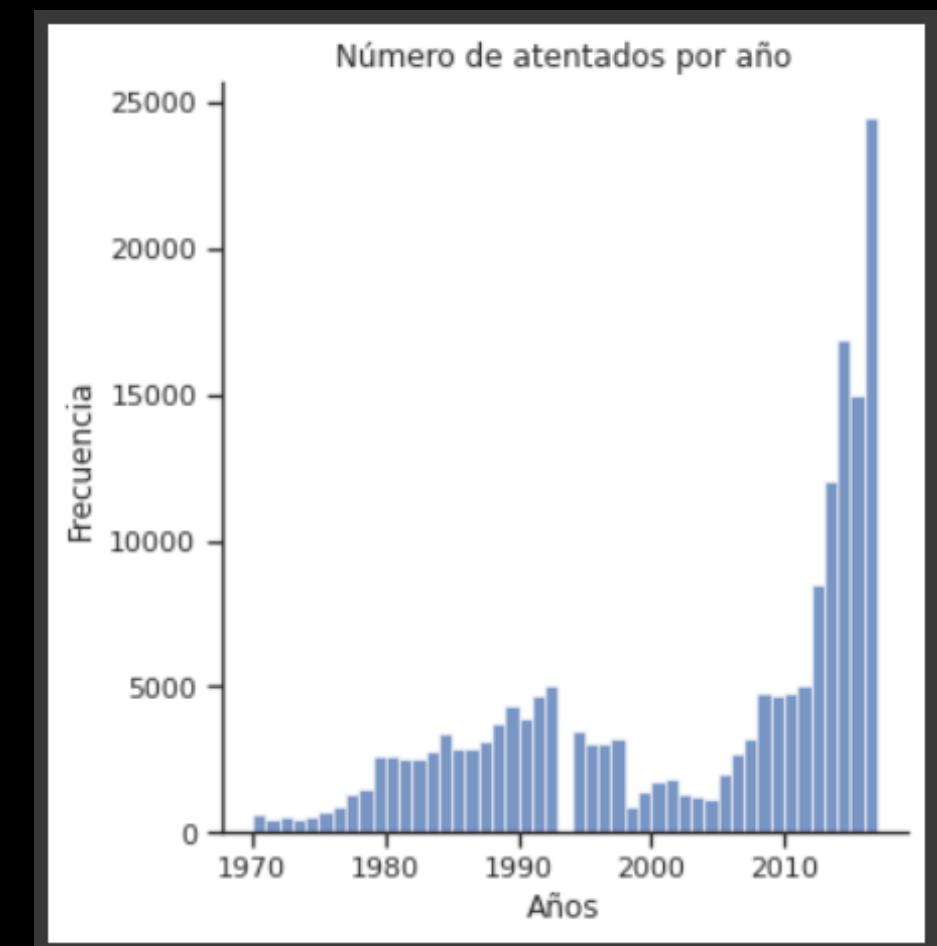
Boxplots



Gráficas de densidad

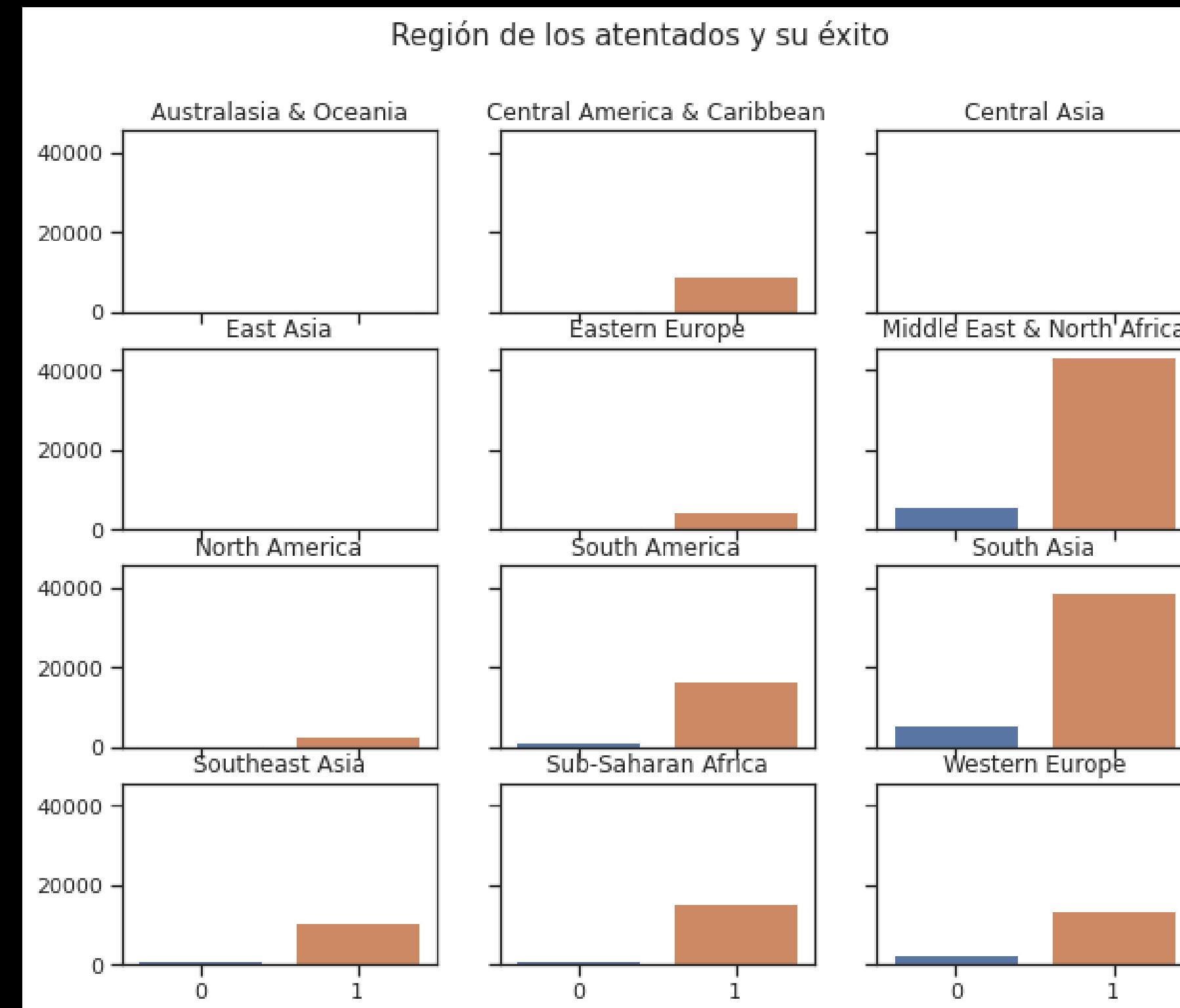


Tablas de frecuencia e histograms

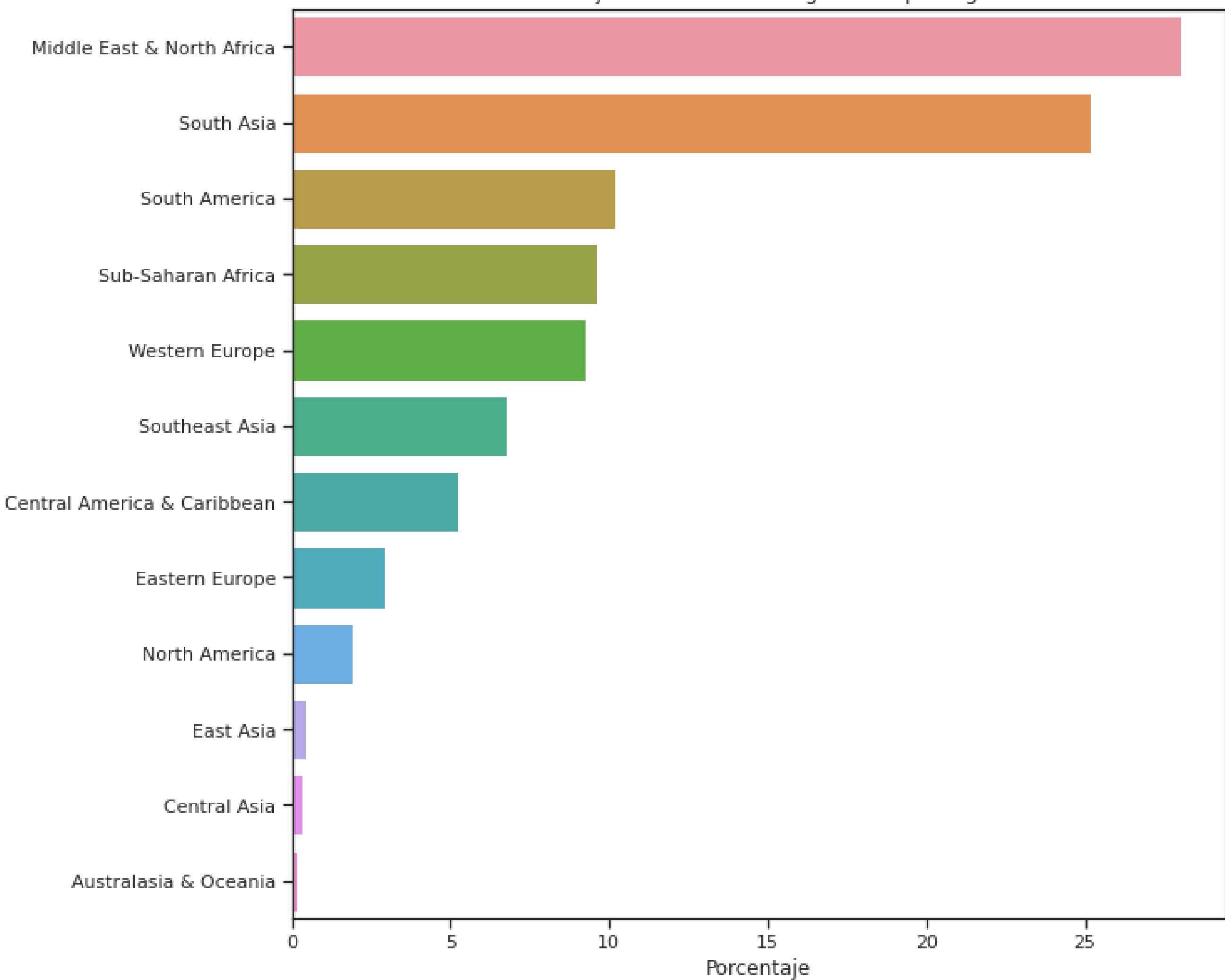


Postwork 03

Exploración de Variables Categóricas y Análisis Multivariable

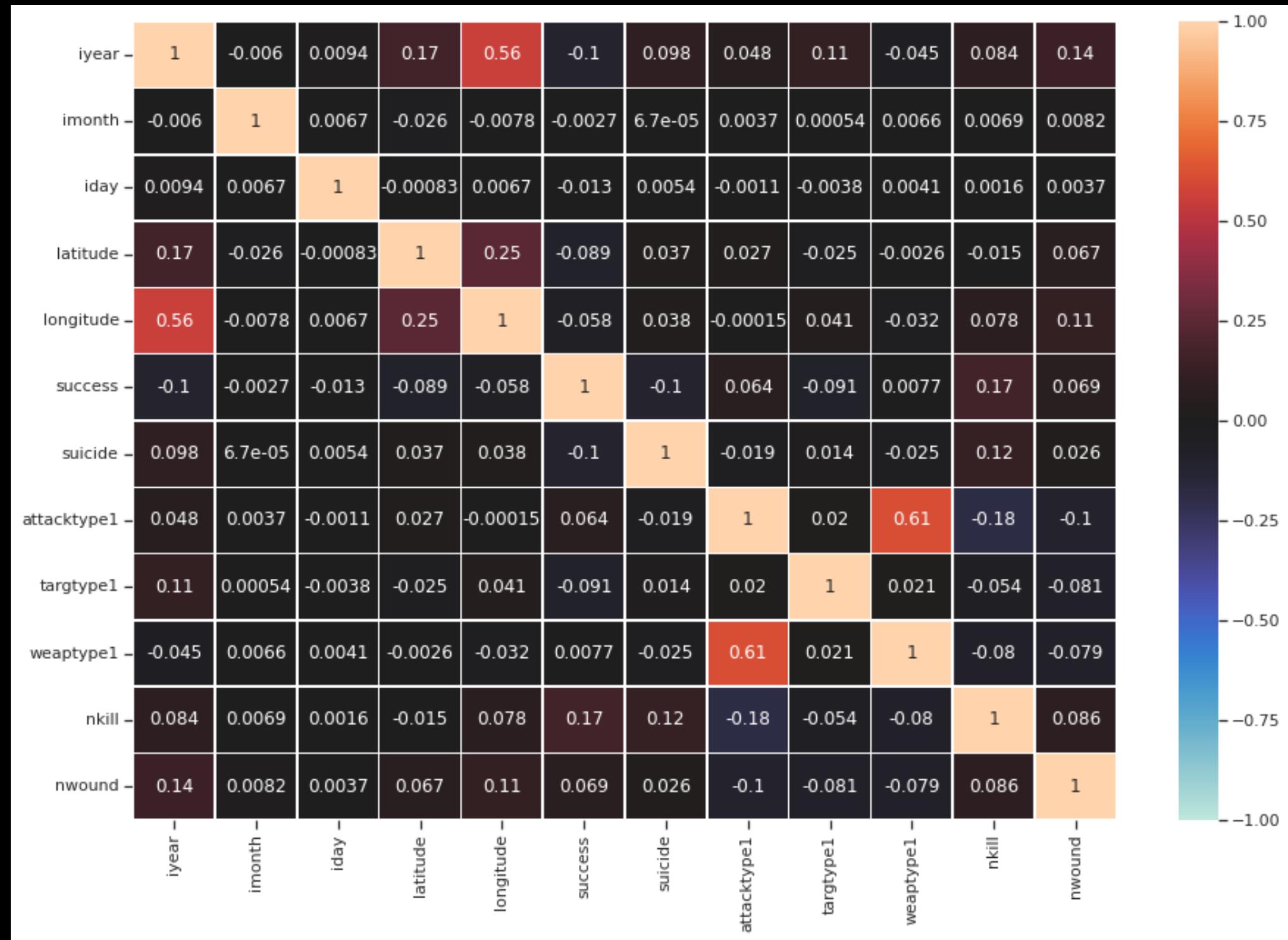


Porcentaje de los atentados registrados por región



Postwork 04

Correlaciones y Regresión Lineal Simple





Notebook



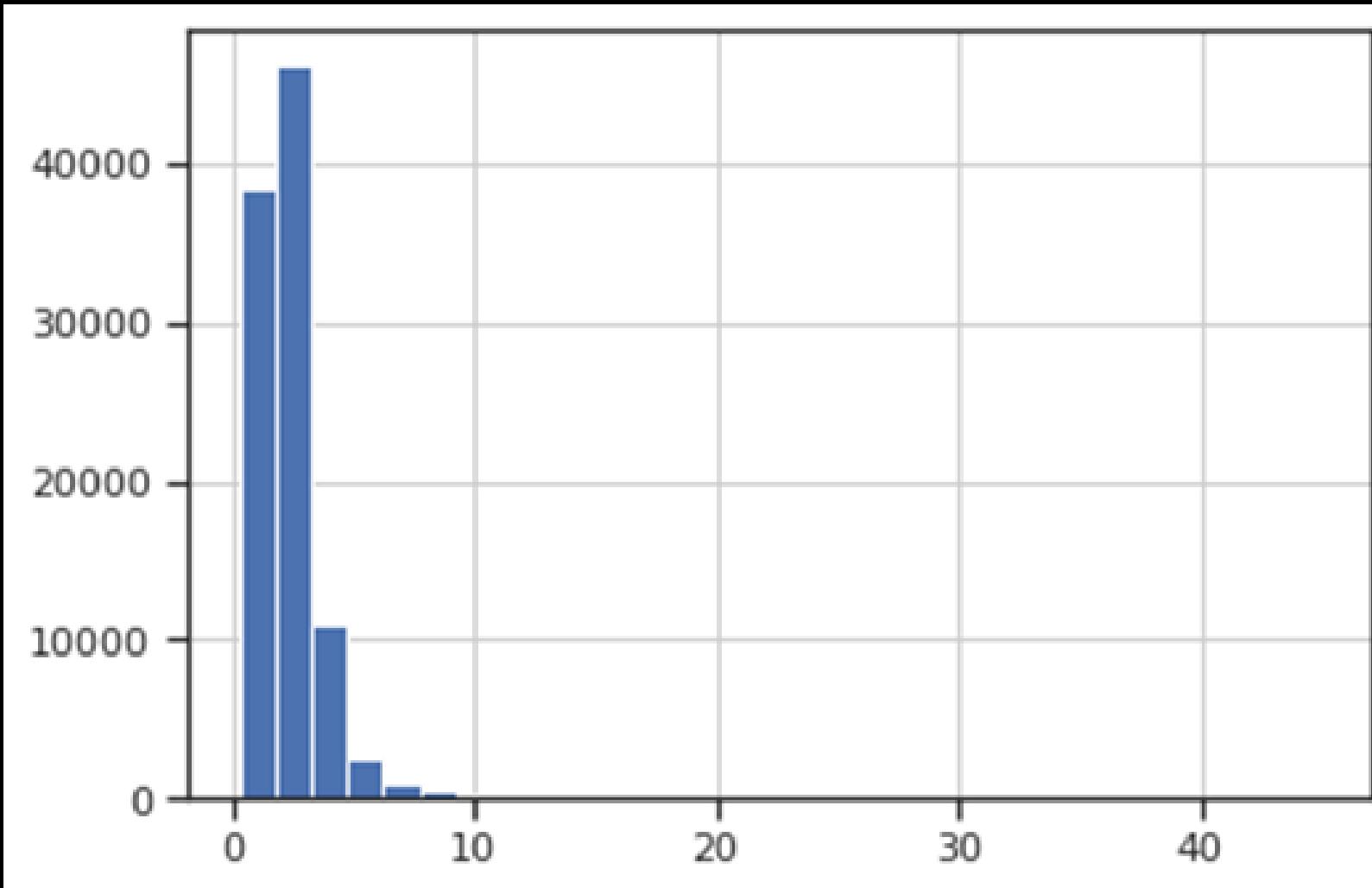
SCAN ME

[https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?](https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?usp=sharing)
usp=sharing

Postwork 05

Distribuciones muestrales

```
# La distribución muestral de las medias para 'nkill'  
means_nkill=bootstrap_mean(50, 100000, df_pw_1['nkill'])
```



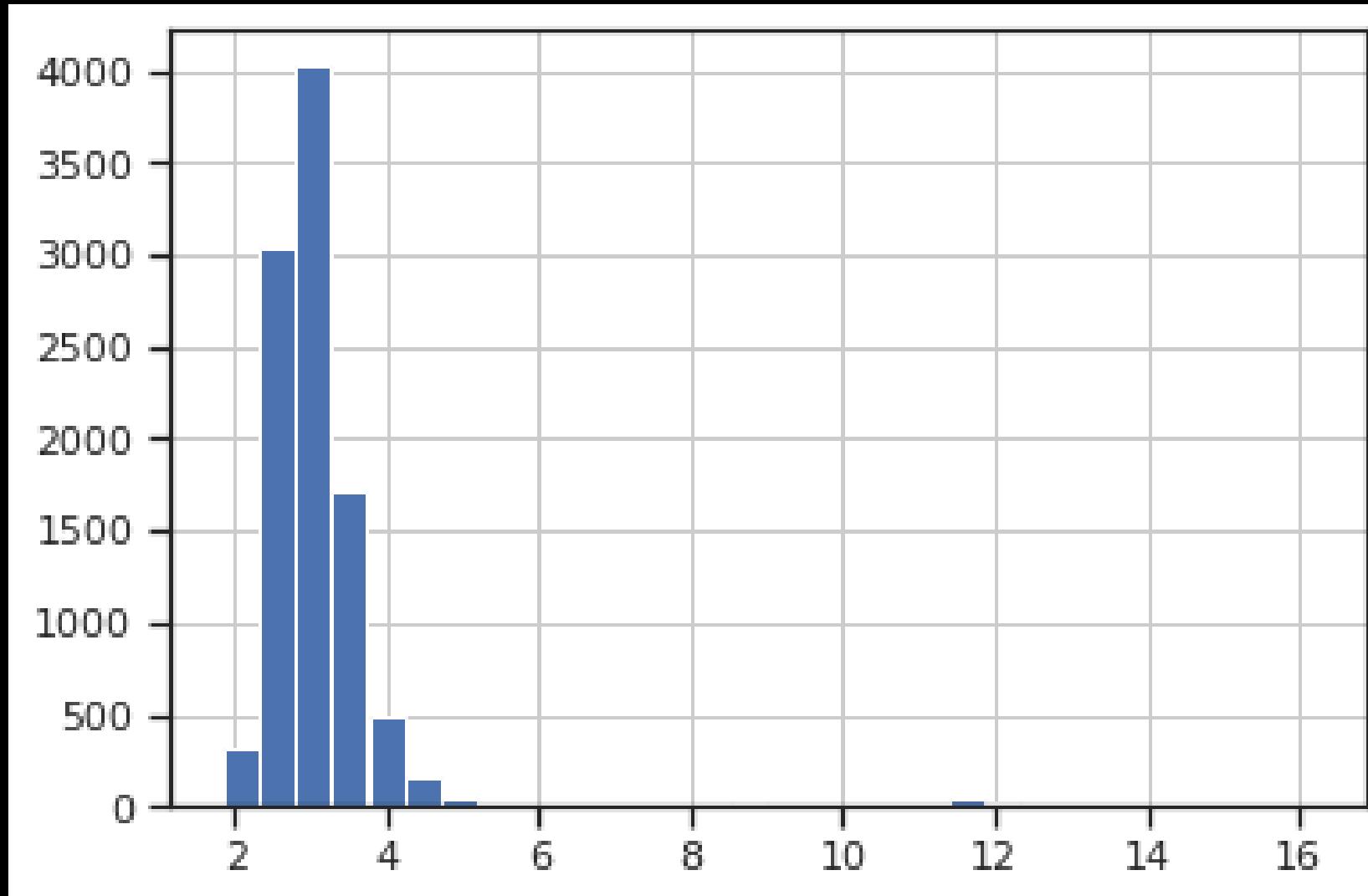
```
# Asimetría  
means_nkill.skew()  
7.91
```

```
# Curtosis  
means_nkill.kurtosis()  
117
```

Postwork 05

Distribuciones muestrales

```
# La distribución muestral de las medias para 'nwound'  
means_nwound=bootstrap_mean(1000, 10000, df_pw_1['nwound'])
```



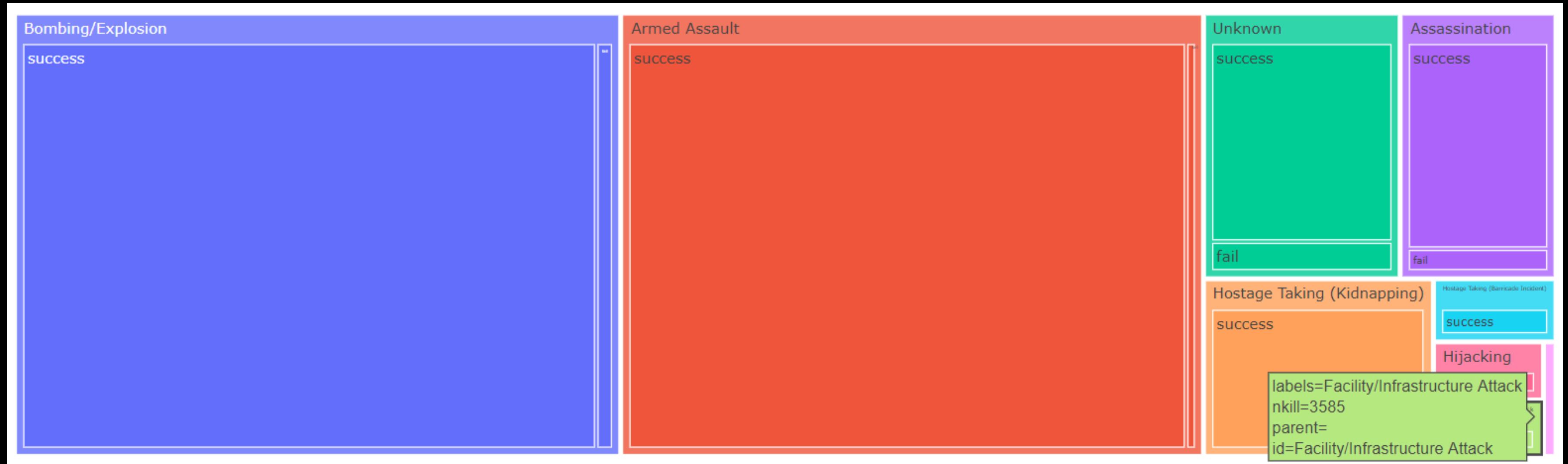
```
# Asimetría  
means_nwound.skew()  
5.80
```

```
# Curtosis  
means_nwound.kurtosis()  
40
```

Postwork 06

Visualización de datos avanzada

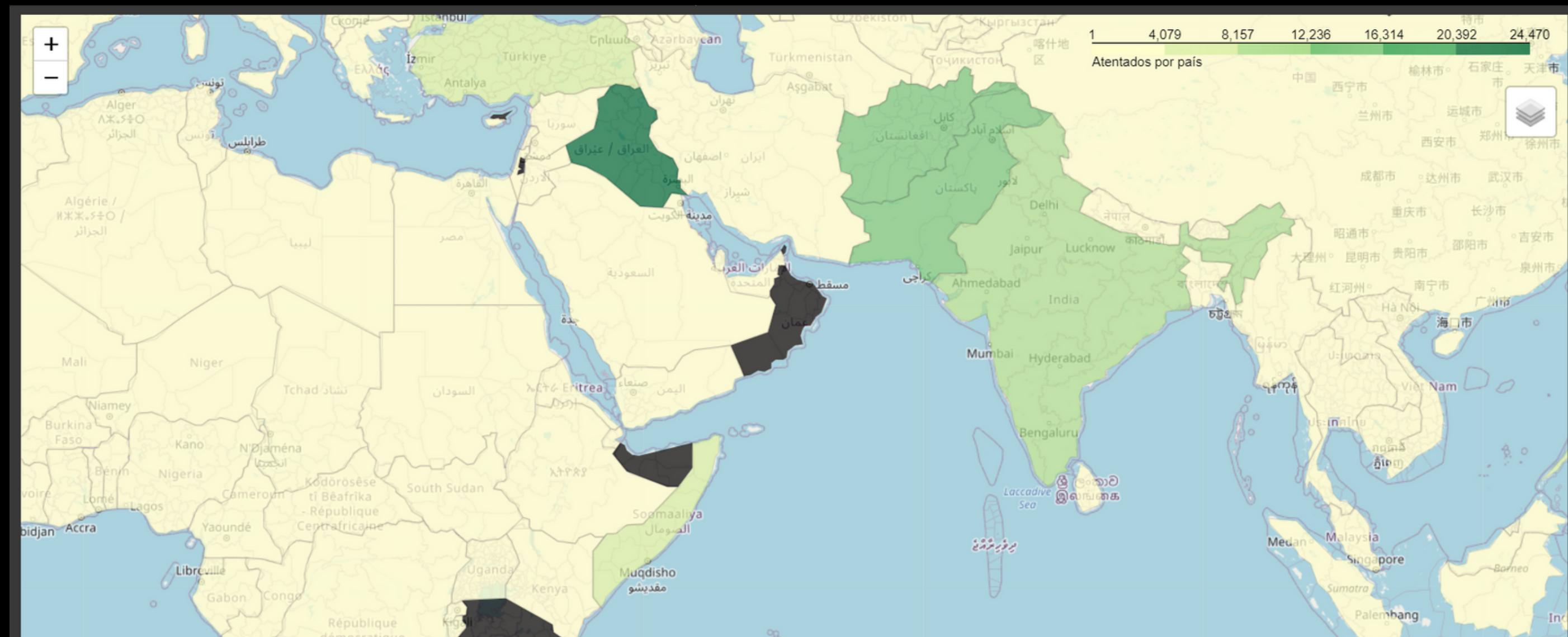
Los **treemaps** se utilizan en relaciones jerárquicas o para variables numéricas agrupadas en múltiples niveles.



Postwork 06

Visualización de datos avanzada

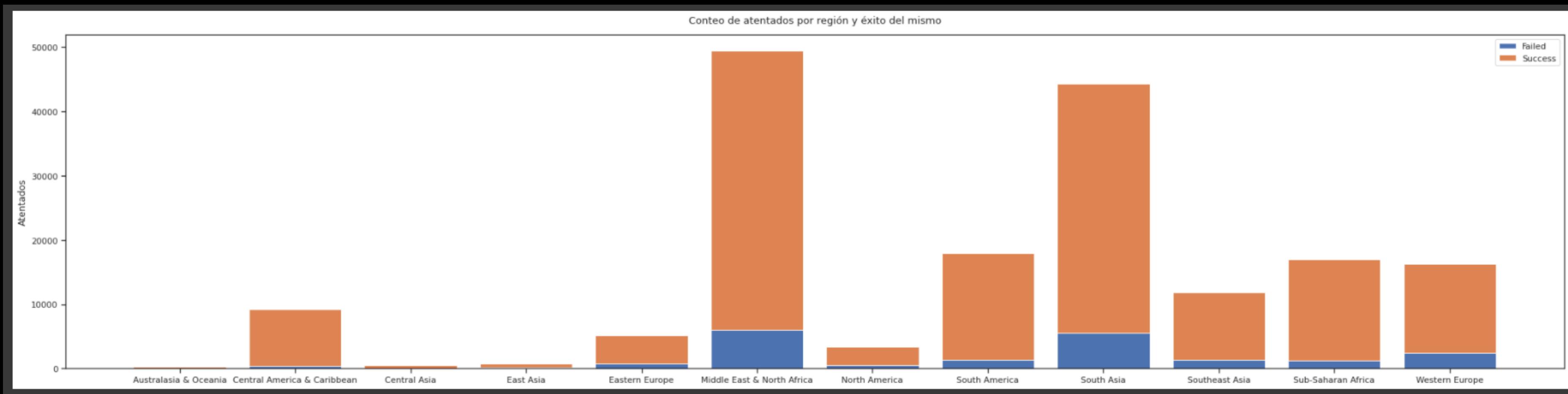
Los **mapas cloropléticos** sirven para explorar datos geográficos o variables numéricas que se pueden agrupar en regiones.



Postwork 06

Visualización de datos avanzada

Las **gráficas de barras apiladas** se utilizan cuando se tiene una variable numérica agrupada en dos niveles por variables categóricas.



Postwork 07

Procesamiento de Lenguaje Natural

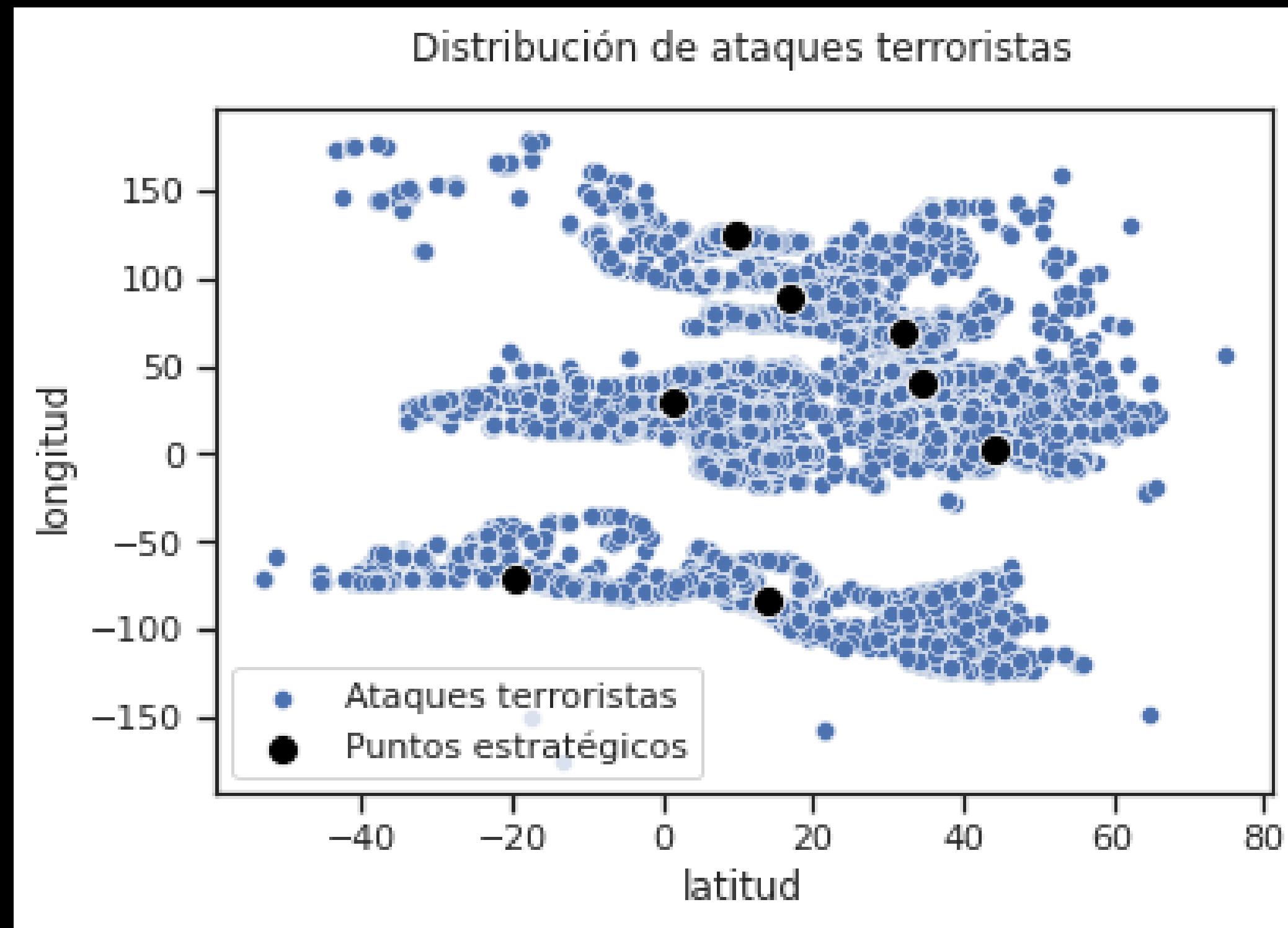


SCAN ME

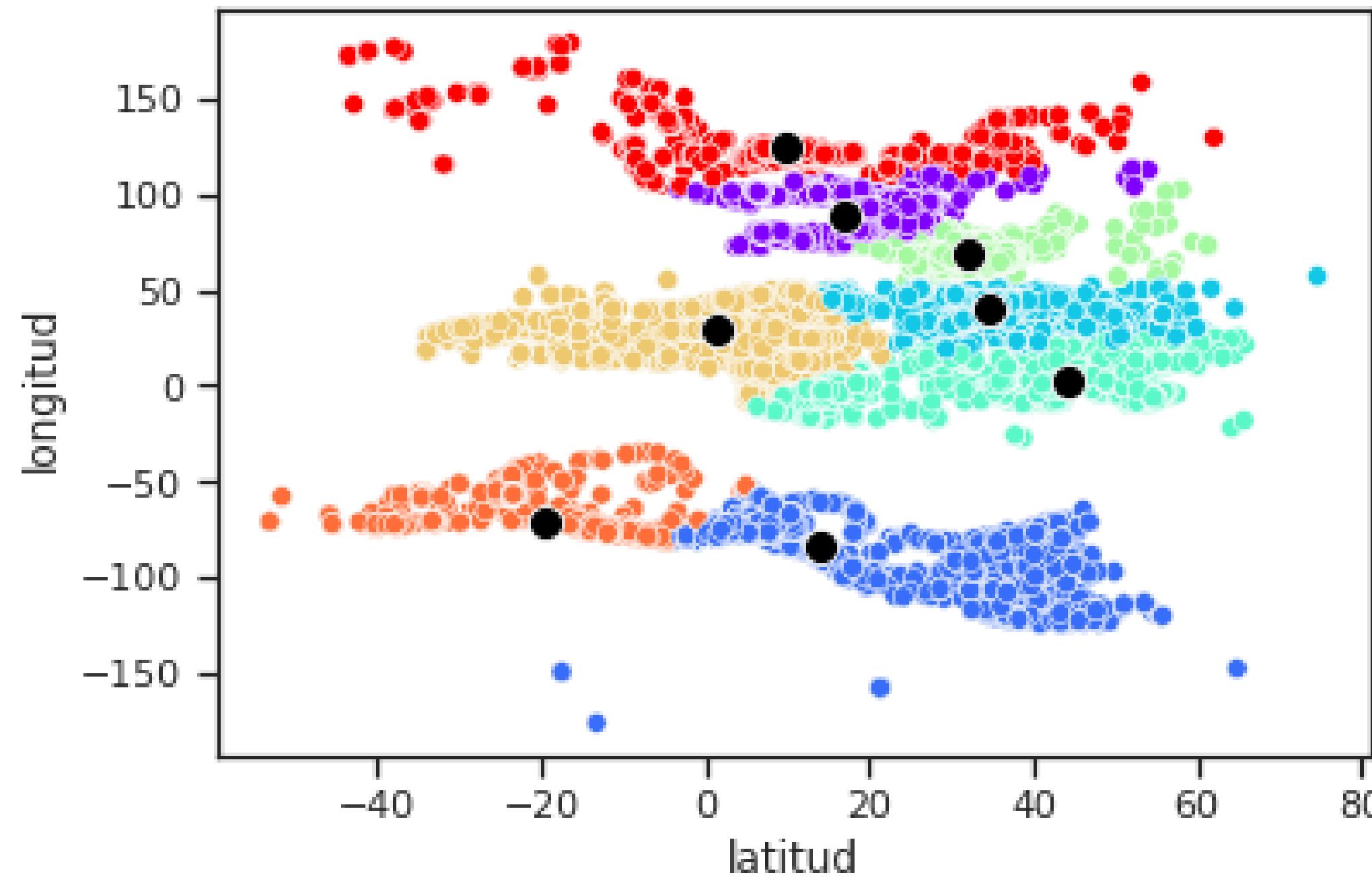
[https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?
usp=sharing](https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?usp=sharing)

Postwork 08

Clasificación No Supervisada y Supervisada

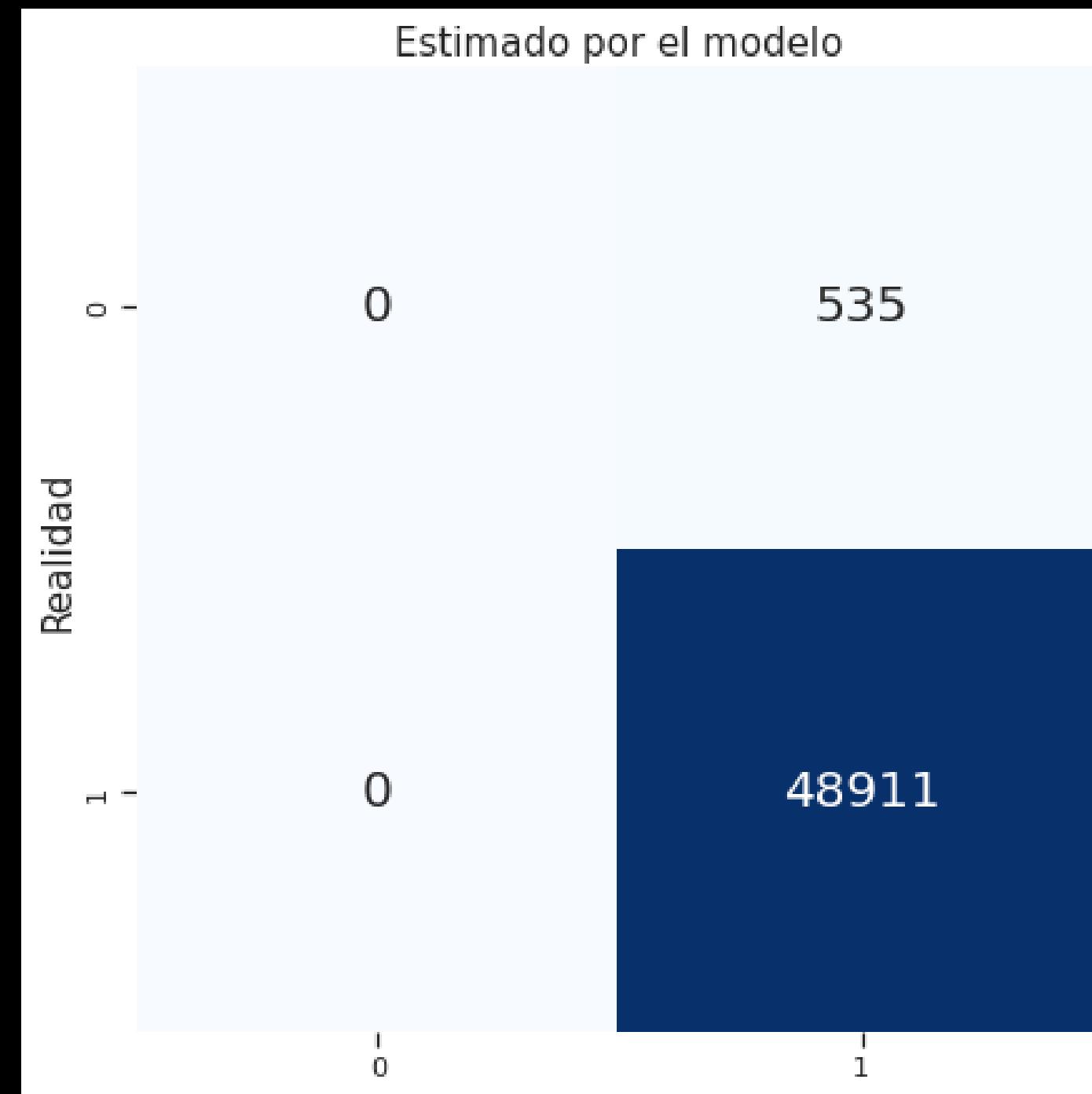


Distribución de ataques terroristas



LR.score(X_test, c1_test)

98.84%



Notebook



SCAN ME

[https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?](https://colab.research.google.com/drive/17ZT9D-cVKDLk4TTtMrNY_9fLwqpkJg?usp=sharing)
usp=sharing

Retos a
los que
enfrento
el equipo.



Repositorio

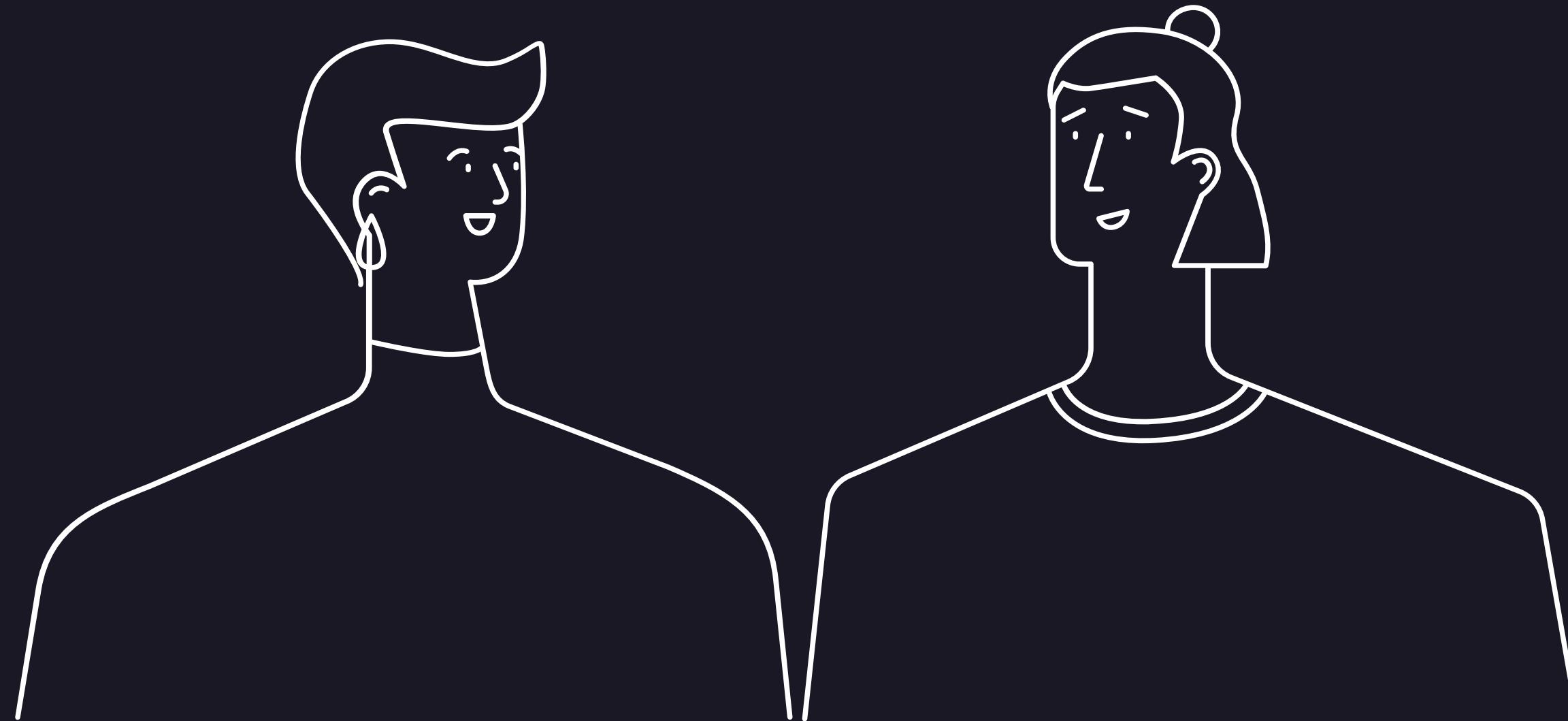


SCAN ME

https://github.com/gilesitorr/DataScience3_Bloque4

GRACIAS POR SU ATENCIÓN

B*EDU



Equipo 2