

IT Academy

¿ES PREDECIBLE LA GUERRA?

MODELO DE CLASIFICACIÓN DE CONFLICTOS ARMADOS



David Cuxart

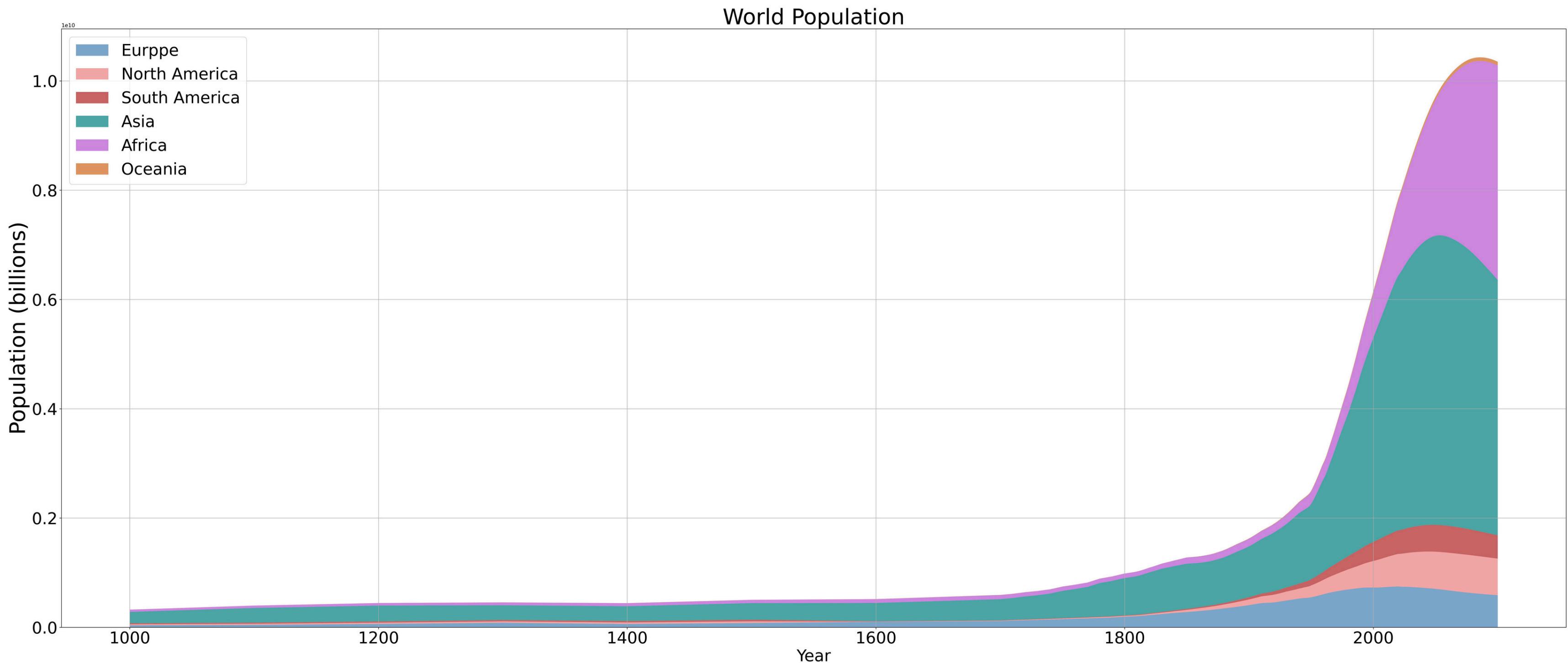
DESCRIPCIÓN GENERAL

- 01 Introducción
- 02 Datos y métodos
- 03 Resultados
- 04 Conclusiones

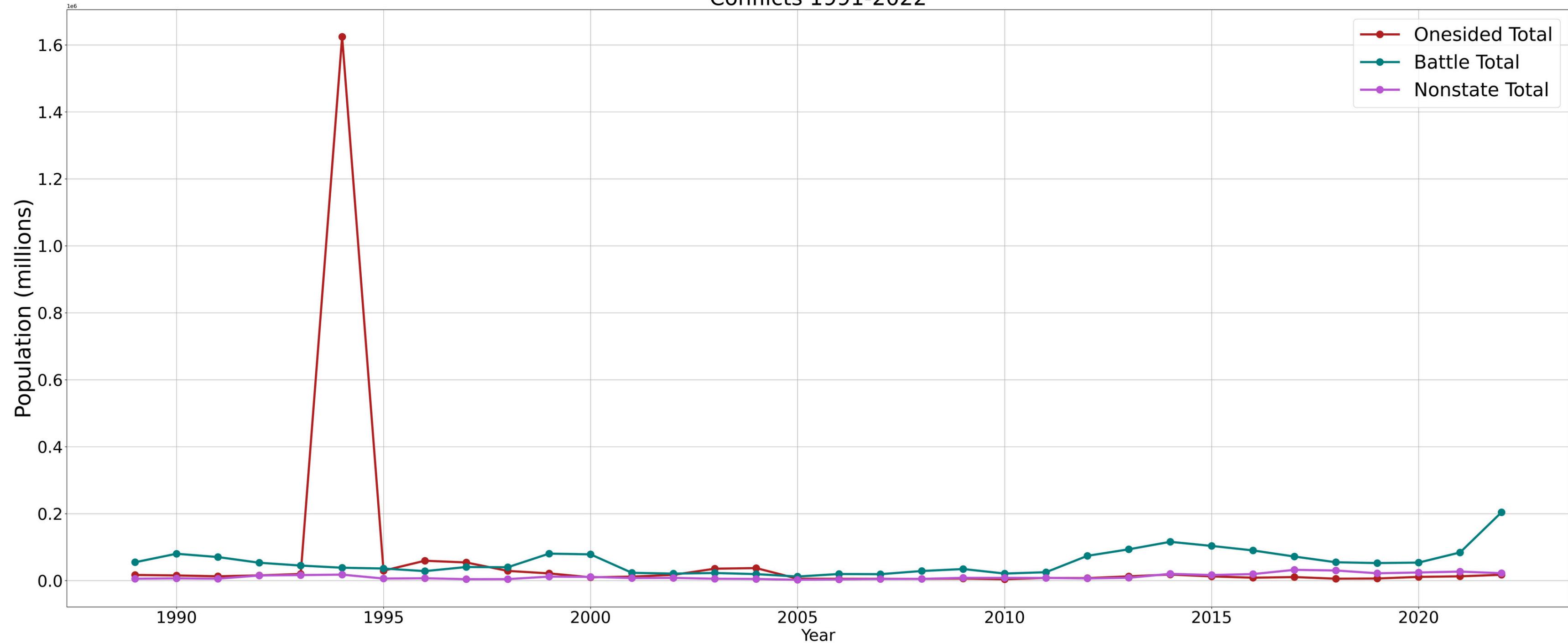


INTRODUCCIÓN

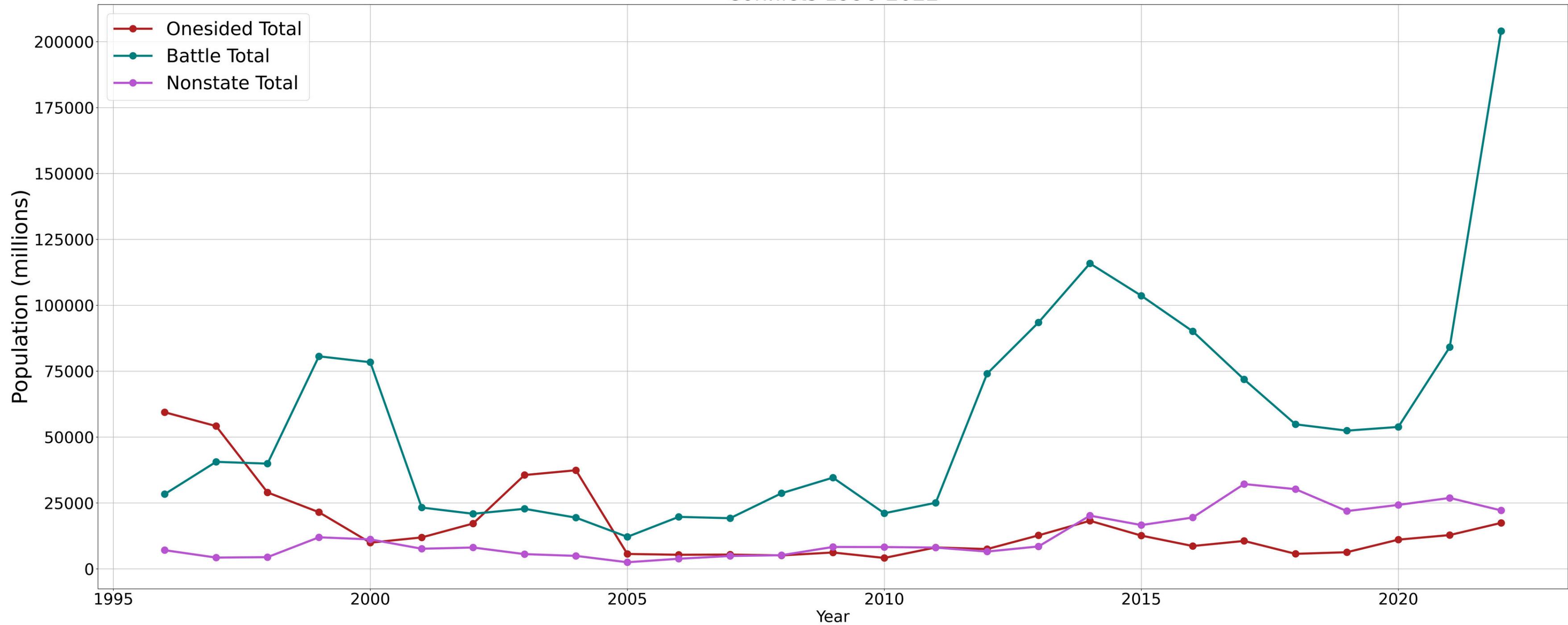
Uganda	8.08	Kenya	5.14	Syria	4.02	Samoa	2.75	Antigua&B.	2.00	Fr. Polynesia	1.50	Aruba	1.00	Japan	0.82
Zambia	7.52	S.-Sudan	5.02	Niger	4.01	Maldives	2.67	Saint Lucia	2.00	Sri Lanka	1.46	Barbados	1.00	Portugal	0.82
Lesotho	7.35	Sao Tome	5.00	Jordan	3.95	Fre. Guiana	2.60	Turkey	2.00	Puerto Rico	1.44	Cuba	1.00	Singapore	0.82
Malawi	7.04	Sierra Leone	4.99	Gabon	3.91	Eq. Guinea	2.53	Algeria	1.98	Uruguay	1.44	Curacao	1.00	Croatia	0.81
Zimbabwe	6.81	Guatemala	4.98	Cabo Verde	3.86	Mexiko	2.52	Qatar	1.90	Tunisia	1.43	Iceland	1.00	Lithuania	0.80
Mali	6.61	Madagascar	4.84	Honduras	3.75	Grenada	2.50	Lebanon	1.88	N. Caledonia	1.43	Seychelles	1.00	Canada	0.79
Burkina F.	6.46	Benin	4.80	Cambodia	3.72	Bhutan	2.47	Israel	1.87	Kazakhstan	1.37	US Virgin	1.00	Slovakia	0.78
Chad	6.40	Namibia	4.75	Nepal	3.65	Libya	2.46	Sau Arabia	1.86	Réunion	1.36	China	0.99	Switzerland	0.77
Afghanistan	5.99	Liberia	4.74	Haiti	3.59	Egypt	2.45	Suriname	1.85	Azerbaijan	1.33	France	0.99	Greece	0.76
Senegal	5.87	Nigeria	4.74	Botswana	3.58	Ecuador	2.41	Colombia	1.84	Chile	1.32	Georgia	0.98	Ukraine	0.76
Angola	5.84	C. African R.	4.72	Micronesia	3.50	Guyana	2.40	Morocco	1.84	Albania	1.31	Thailand	0.98	Austria	0.75
Ethiopia	5.83	Sudan	4.65	Vanuatu	3.50	India	2.38	Brazil	1.81	N. Korea	1.27	United King.	0.96	Bosnia+Herz	0.75
Iraq	5.80	Rep. Congo	4.55	Pakistan	3.39	Peru	2.37	Argentina	1.79	Mauritius	1.20	USA	0.96	Martinique	0.75
Gambia	5.79	Guinea-Bis.	4.45	Bolivia	3.24	Dominican R.	2.35	Fiji	1.77	U. Arab Em.	1.17	Luxembourg	0.95	Poland	0.71
Mozambique	5.77	Guinea	4.42	Djibouti	3.21	Myanmar	2.26	Iran	1.75	Guadeloupe	1.14	Moldova	0.92	Czech R.	0.70
Somalia	5.68	Timor-Leste	4.40	Tajikistan	3.21	Kyrgyzstan	2.25	Oman	1.70	Montenegro	1.10	Taiwan	0.88	Italy	0.70
Swaziland	5.62	C. d'Ivoire	4.37	South Africa	3.19	Turkmenistan	2.25	Kuwait	1.69	New Zealand	1.10	Sweden	0.87	Spain	0.69
Rwanda	5.59	Mayotte	4.33	Belize	3.17	Uzbekistan	2.22	S. Vincent	1.67	Ireland	1.08	Armenia	0.86	Bulgaria	0.68
Palestine	5.46	Comoros	4.30	Nicaragua	3.12	Jamaica	2.19	Brunei	1.64	Serbia	1.08	Malta	0.86	Estonia	0.68
[Gaza]	>6.00]											Romania	0.86	Russia	0.67
Cameroon	5.42	Burundi	4.21	El Salvador	3.02	Kosovo	2.15	Guam	1.60	Cyprus	1.06	Netherlands	0.85	Belarus	0.66
Eritrea	5.41	Ghana	4.19	Kiribati	3.00	Mongolia	2.15	Costa Rica	1.59	Trinid.+Tob.	1.05	S.-Korea	0.85	Germany	0.65
Yemen	5.41	Mauritania	4.17	Tonga	3.00	Venezuela	2.15	W. Sahara	1.57	Australia	1.03	Finland	0.84	Macao	0.65
D.R. Congo	5.29	Papua N.G.	4.13	Bangla Desh	2.98	Panama	2.06	Vietnam	1.53	Norway	1.02	Hungary	0.84	Slovenia	0.63
Tanzania	5.22	Solomon Is.	4.13	Paraguay	2.87	Indonesia	2.05	Bahamas	1.50	Denmark	1.01	Channel Is.	0.83	Latvia	0.62
Togo	5.21	Laos	4.05	Philippines	2.84	Malaysia	2.04	Bahrein	1.50	Macedonia	1.01	Belgium	0.82	Hong Kong	0.59



Conflicts 1991-2022



Conflicts 1996-2022



ÍNDICES DE CONFLICTO



The Hague Centre
for Strategic Studies



ACLED
Bringing clarity to crisis



DATOS Y MÉTODOS

DESCRIPCIÓN DEL DATASET

Variables independientes

16 variables que indican sobre la dimensión económica, política, social, de seguridad o poblacional.

Origen de los datos

Datos de 7 distintas fuentes: universidades, instituciones, organizaciones privadas y públicas.

Variables dependientes

Dispone de 4 variables para predecir distintos tipos de violencia política.

Serie temporal

Disponemos de una serie temporal de 32 años de 1991 hasta 2022.

Bases de datos

El dataset se compone de hasta 13 diferentes archivos de bases de datos

Localizaciones

En total 123 países.

PROCEDENCIA DE LOS DATOS



UPPSALA
UNIVERSITET



VARIABLES INDEPENDIENTES

oil

Exportación de recursos naturales: El porcentaje del total de exportaciones que son recursos naturales, especialmente combustibles fósiles y minerales.

trade

Grado de apertura económica: La suma de las exportaciones e importaciones de bienes y servicios medidas como proporción del producto interno bruto.

gdp

Renta per cápita: Producto interior bruto dividido por el total de la población.

work

Desempleo: Proporción de la población activa laboralmente que está sin trabajo pero disponible y buscándolo.

income

Desigualdad de ingresos: En qué grado hay disparidad significativa en la distribución del ingreso entre individuos, grupos, poblaciones, clases sociales o países.

democracy

Grado de democracia (liberal): Mide la importancia que se le da a la protección de los derechos individuales y de las minorías, así como a la división de poderes y a un alto grado de la protección privada (Rule of Law).

state

Origen del ingreso fiscal del Estado: Esta variable mide si un Estado depende de impuestos sobre recursos naturales, sobre la propiedad o de transacciones económicas.

repression

Nivel de represión: En qué grado hay muertes por cuestiones políticas sin mediación de un proceso judicial.

corruption

Nivel de corrupción: Esta variable va de menos a más corrupción en un Estado, incluyendo tanto a nivel ejecutivo, como legislativo y judicial.

empowerment

Empoderamiento de la mujer: Qué tanto las mujeres están representadas en posiciones de poder y de tomas de decisión, así como presentes organizaciones civiles.

ethnic

Exclusión étnica: En qué medida minorías políticas participan en la toma de decisiones a nivel estatal (no autonómico/regional) o directamente están discriminadas o se autoexcluyen.

pop

Población: Población total de ambos sexos y todas las edades.

youth

Youth Bulge (Población joven): Porcentaje de la población con edad comprendida entre los 15 y 29 años.

war

Índice de guerra de Gunnar Heinsohn: Parte de la población que se encuentra entre 15 y 19 años por la población comprendida entre los 55 y los 59 años.

violence

Tasa de homicidios: Número de defunciones a causa de una muerte violenta ocasionada de manera voluntaria.

VARIABLES DEPENDIENTES

onesided

El uso deliberado de la fuerza armada por parte del gobierno de un estado o de un grupo formalmente organizado **contra civiles** que resulta en al menos 25 muertes en un año.

nonstate

El uso de la fuerza armada entre dos o más grupos armados organizados, **ninguno de los cuales es el gobierno de un estado**, que resulta en al menos 25 muertes relacionadas con batallas en un año.

battle

Un conflicto armado basado en un estado es una incompatibilidad controvertida que afecta al gobierno y/o territorio donde el uso de la fuerza armada entre dos partes, de las cuales al menos una es el gobierno de un estado, resulta en al menos 25 muertes relacionadas con la batalla en un año del calendario.

total

Suma de las anteriores variables dependientes en una sola.

PREPROCESADO DE LOS DATOS

1

En cada una de las fuentes de datos los nombres de los países variaban

4

Fue necesario añadir manualmente Palestina debido a que no hay división territorial.

2

Creación de nuevas columnas a partir de otras columnas. ‘Exclusión étnica’ y variables dependientes.

5

Fue requerido completar algunos años y eliminar países para hacer una lista de países uniforme.

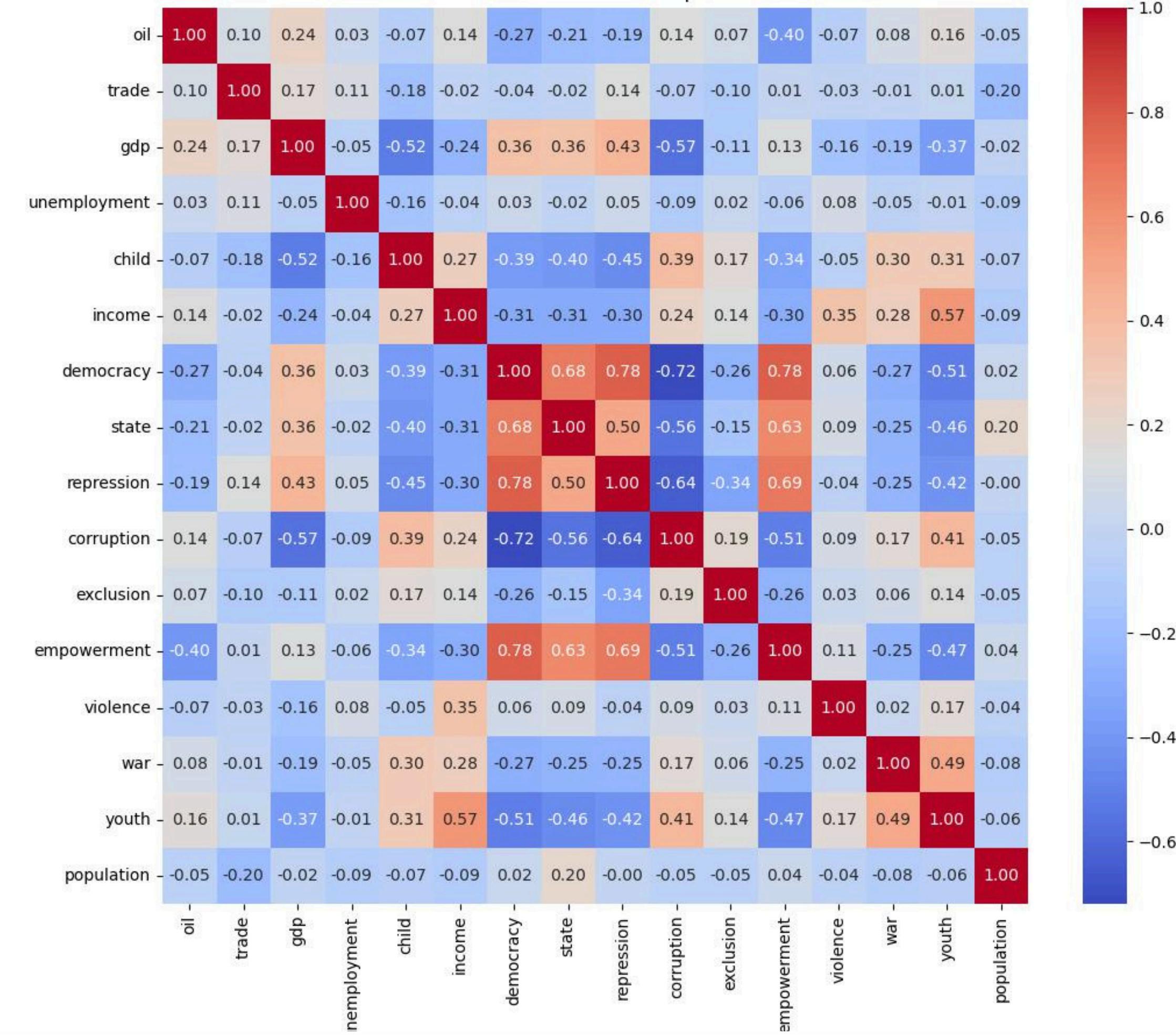
3

Imputación de datos mediante el método LOCF y LOCB.

6

Tan solo fue necesaria la transformación de algunas variables.

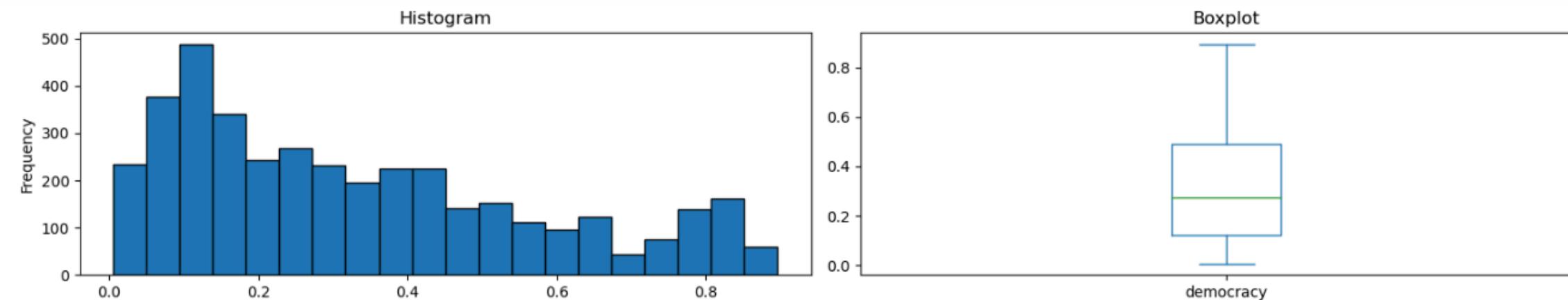
Correlation Heatmap



TRANSFORMACIÓN

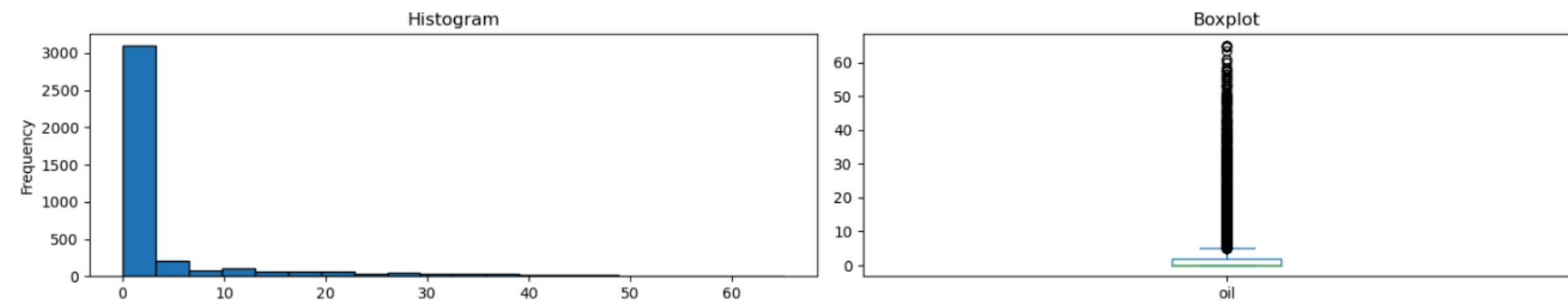
Sin transformar

(V-Dem: democracy, state, repression, corruption)



Robust Scaler

(resto de variables)



SELECCIÓN DEL MODELO

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
ExtraTreesClassifier	0.90	0.89	0.89	0.90	0.85
LGBMClassifier	0.90	0.88	0.88	0.90	0.49
XGBClassifier	0.90	0.88	0.88	0.90	0.46
LabelPropagation	0.89	0.88	0.88	0.89	0.85
LabelSpreading	0.89	0.88	0.88	0.89	1.08
RandomForestClassifier	0.90	0.88	0.88	0.90	2.56
KNeighborsClassifier	0.90	0.88	0.88	0.90	0.19
BaggingClassifier	0.89	0.86	0.86	0.89	0.92
DecisionTreeClassifier	0.87	0.85	0.85	0.87	0.12
ExtraTreeClassifier	0.86	0.84	0.84	0.86	0.04

MODELO XGBOOST

Modelo

Extreme Gradient Boosting es un algoritmo que pertenece a la clase de métodos de aprendizaje conjunto. Esto quiere decir que este modelo combina las predicciones de múltiples árboles de decisión para hacer una predicción final.

Validación

Para evaluar el modelo se aplica una validación cruzada en 10 capas. De este modo, garantiza la evaluación del rendimiento de un modelo en un conjunto de prueba separado.

Hiperparámetros

Mediante GridSearchCV podemos encontrar los mejores hiperparámetros para optimizar el modelo. El proceso implica entrenarlo con diferentes combinaciones de hiperparámetros y seleccionar el que produzca el mejor rendimiento.

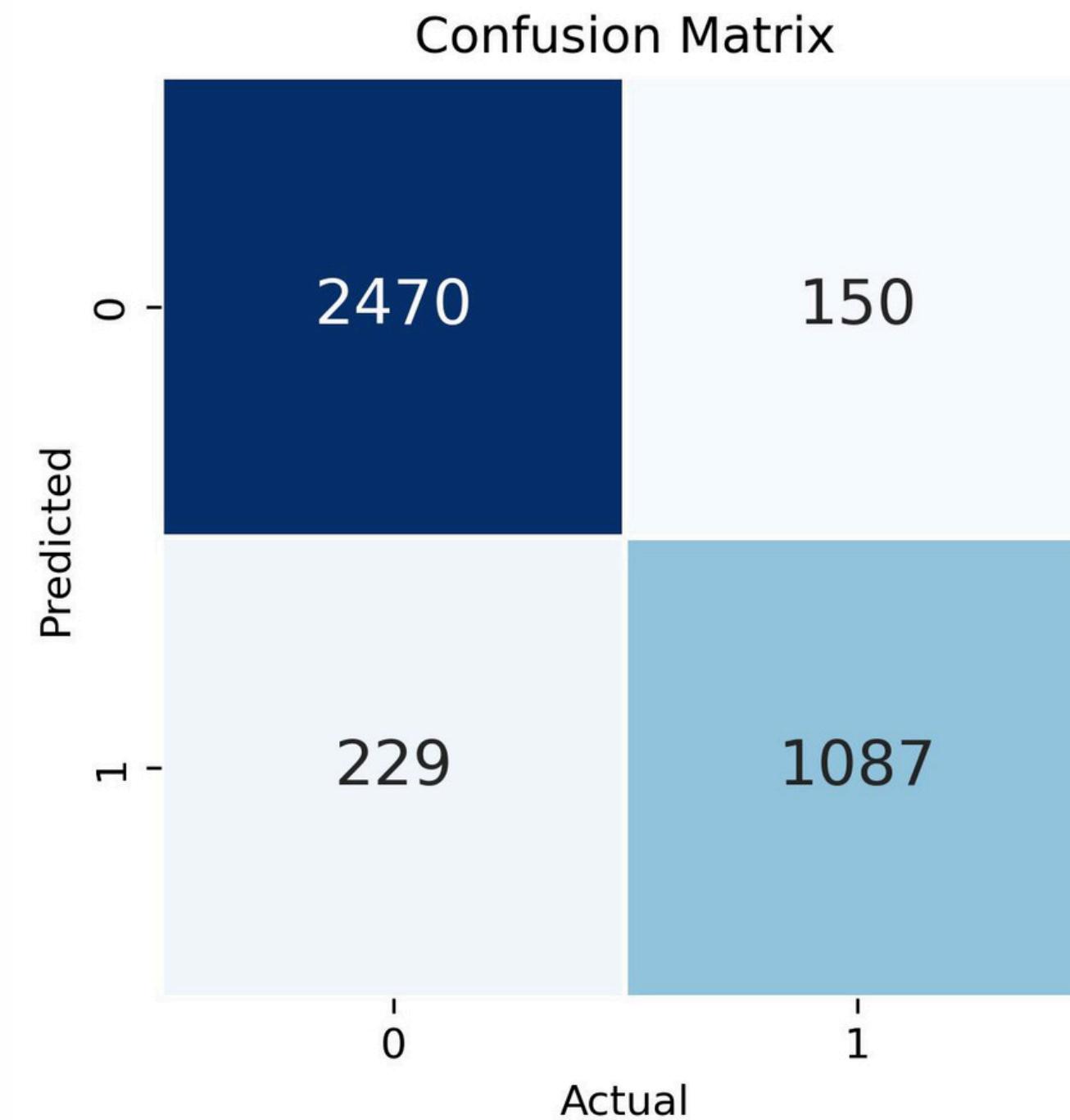


RESULTADOS

INCIDENCIA TOTAL

Classification Report:

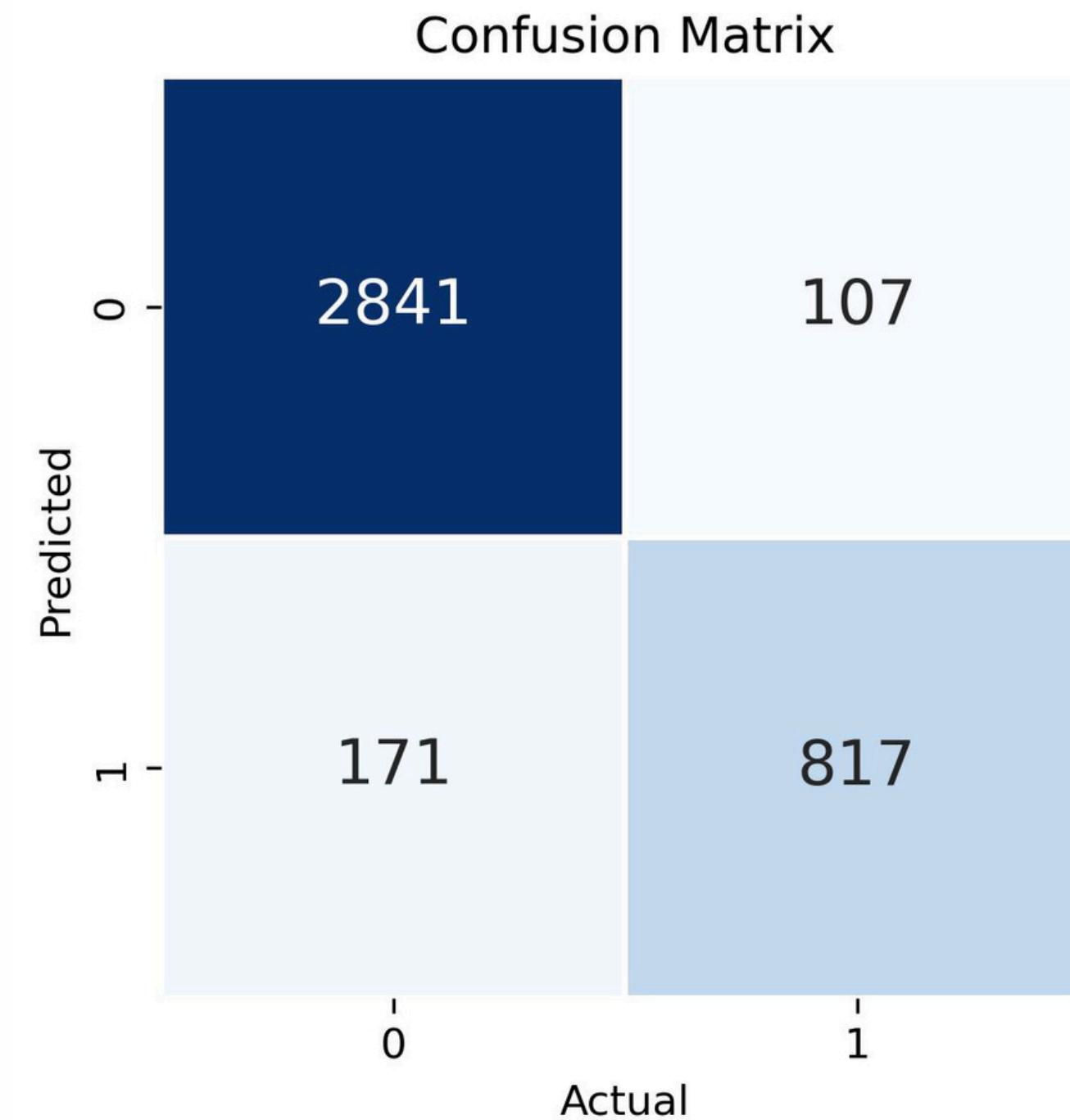
	precision	recall	f1-score	support
0	0.92	0.94	0.93	2620
1	0.88	0.83	0.85	1316
accuracy			0.90	3936
macro avg	0.90	0.88	0.89	3936
weighted avg	0.90	0.90	0.90	3936



INCIDENCIA BATTLE

Classification Report:

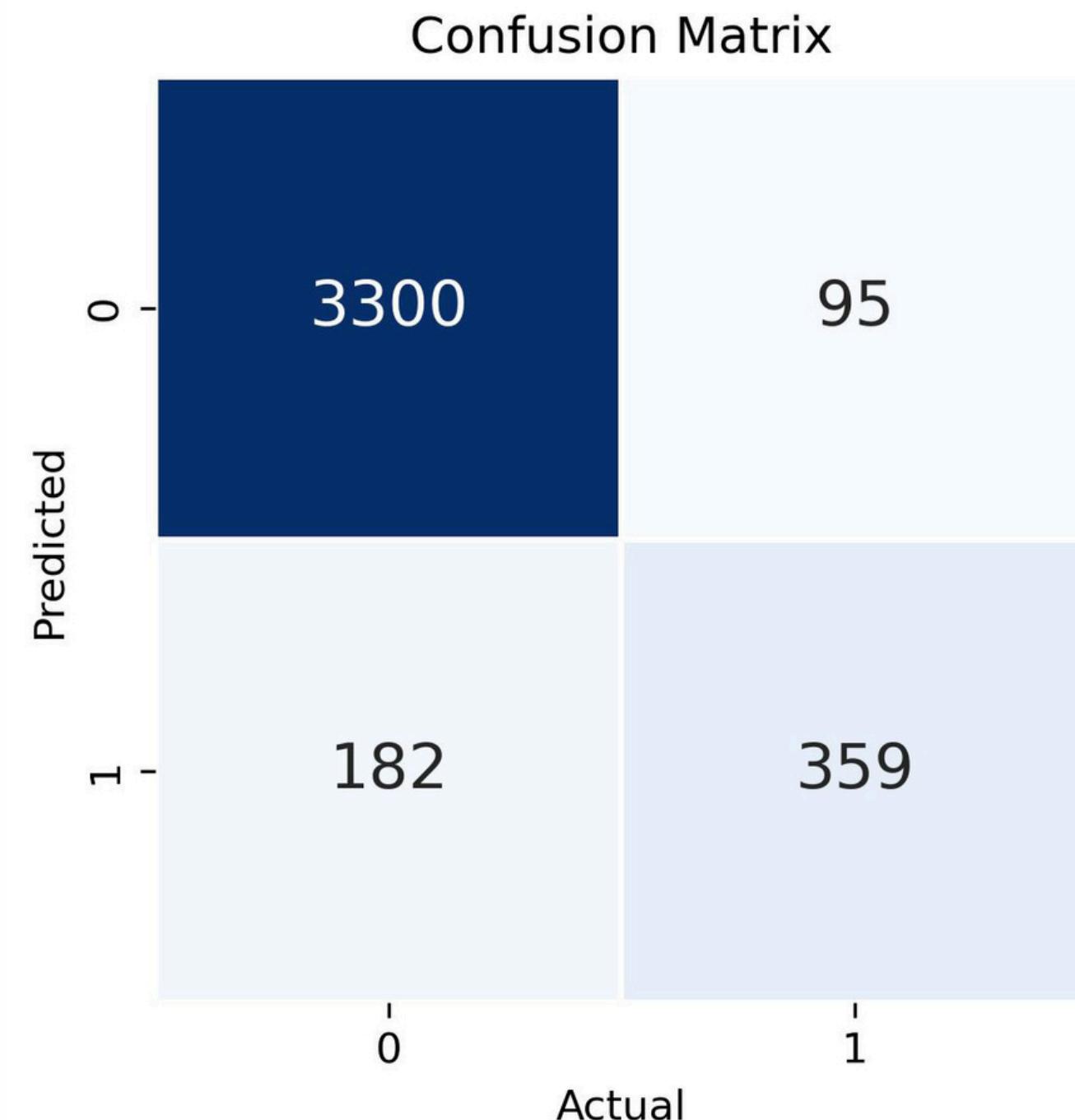
	precision	recall	f1-score	support
0	0.94	0.96	0.95	2948
1	0.88	0.83	0.85	988
accuracy			0.93	3936
macro avg	0.91	0.90	0.90	3936
weighted avg	0.93	0.93	0.93	3936



INCIDENCIA NONSTATE

Classification Report:

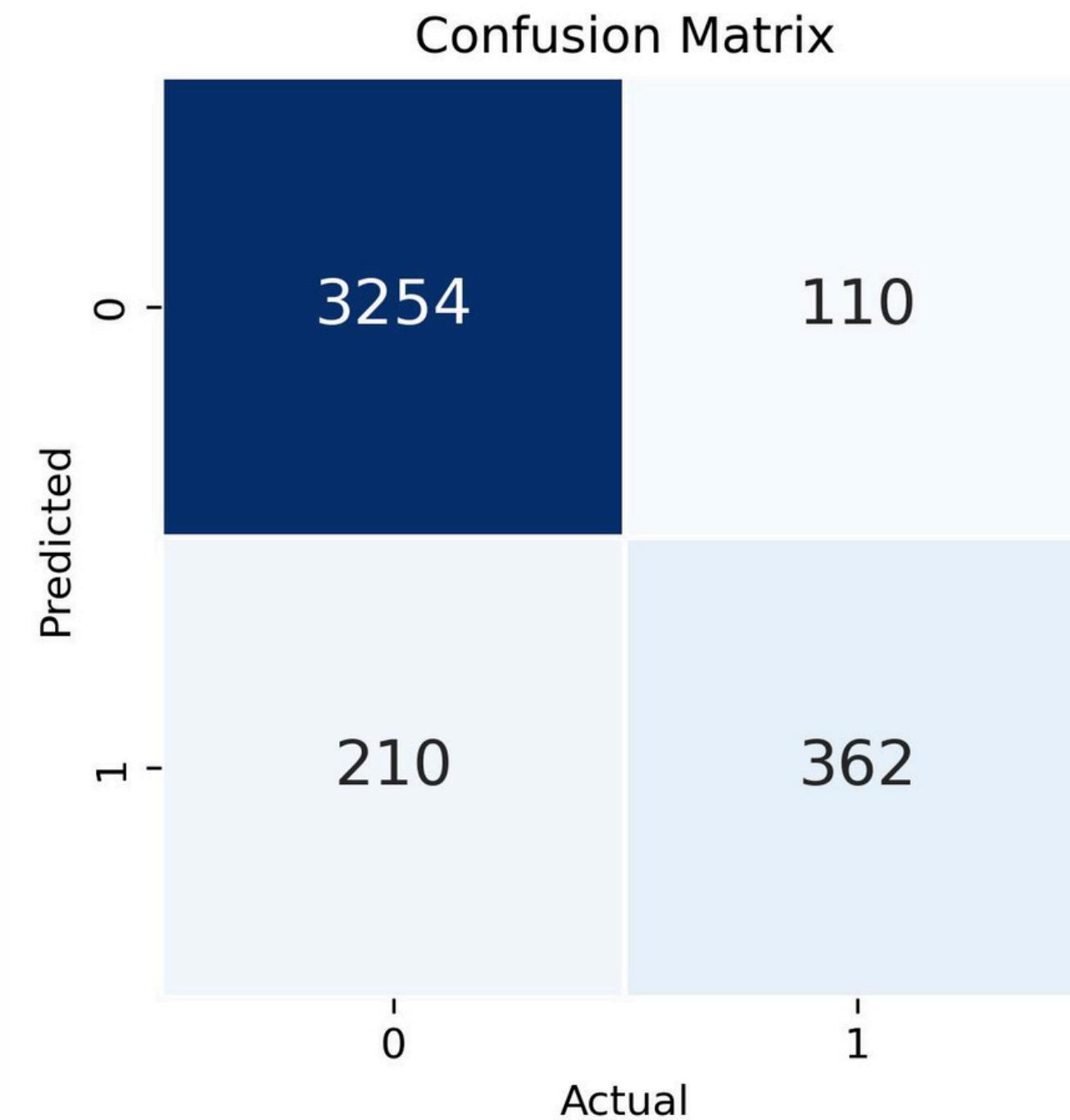
	precision	recall	f1-score	support
0	0.95	0.97	0.96	3395
1	0.79	0.66	0.72	541
accuracy			0.93	3936
macro avg	0.87	0.82	0.84	3936
weighted avg	0.93	0.93	0.93	3936



INCIDENCIA ONESIDED

Classification Report:

	precision	recall	f1-score	support
0	0.94	0.97	0.95	3364
1	0.77	0.63	0.69	572
accuracy			0.92	3936
macro avg	0.85	0.80	0.82	3936
weighted avg	0.91	0.92	0.92	3936





CONCLUSIONES

INVESTIGACIÓN FUTURA

Nuevas variables

Variables referentes a las estados medioambientales como la sequía, o a la nutrición, así como a la expansión de los medios de comunicación, ayudarían a mejorar el rendimiento de este trabajo.

Predicciones por año

Incluir una predicciones para los próximos años mediante un modelo de aprendizaje de clustering, para así poder saber cuál es riesgo de conflicto.

Predicciones por país

Añadir la distinción entre países para poder obtener información de manera separada para cada país.



GRACIAS

usuario github: dcuxart