

# Importación de librerías

```
In [1]: from google.colab import files  
        uploaded = files.upload()
```

Elegir archivos

Ningún archivo seleccionado Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving ViolenceAgainstWomen.json to ViolenceAgainstWomen.json

```
In [2]: import pandas as pd  
  
        # Cargar el archivo JSON línea por línea  
        datos_json = pd.read_json('ViolenceAgainstWomen.json', lines=True)  
  
        # Verificar las primeras filas  
        datos_json.head()
```

Out[2]:

	_id	COUNTRY	SOURCE	DATE	TITLE	BODY
0	{'\$oid': '5ef4c3ca787e0000cc000e40'}	latinoamerica	bbcmundo	3 junio 2019\n	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...
1	{'\$oid': '5ef4c3ca787e0000cc000e43'}	latinoamerica	bbcmundo	21 enero 2019\n	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...
2	{'\$oid': '5ef4c3ca787e0000cc000e47'}	latinoamerica	bbcmundo	8 marzo 2018\n	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...
3	{'\$oid': '5ef4c3ca787e0000cc000e48'}	latinoamerica	bbcmundo	23 noviembre 2018\n	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...
4	{'\$oid': '5ef4c3ca787e0000cc000e49'}	latinoamerica	bbcmundo	24 abril 2019\n	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...

In [3]: *# Dependencies*

```
import numpy as np
import pandas as pd
from pandas import json_normalize
import json

#
import warnings
warnings.filterwarnings('ignore')
```

## 1.- Carga y Exploración de Datos

In [4]: *# <https://zenodo.org/records/6958808>*

```
datos_json = pd.read_json('ViolenceAgainstWomen.json', lines = True)
```

datos\_json

Out[4]:

	_id	COUNTRY	SOURCE	DATE	TITLE	BODY
0	{'\$oid': '5ef4c3ca787e0000cc000e40'}	latinoamerica	bbcmundo	3 junio 2019\n	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...
1	{'\$oid': '5ef4c3ca787e0000cc000e43'}	latinoamerica	bbcmundo	21 enero 2019\n	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...
2	{'\$oid': '5ef4c3ca787e0000cc000e47'}	latinoamerica	bbcmundo	8 marzo 2018\n	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...
3	{'\$oid': '5ef4c3ca787e0000cc000e48'}	latinoamerica	bbcmundo	23 noviembre 2018\n	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...
4	{'\$oid': '5ef4c3ca787e0000cc000e49'}	latinoamerica	bbcmundo	24 abril 2019\n	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...
...	...	...	...	...	...	...
6836	{'\$oid': '60b78fe51b7e26727f6974f8'}	guatemala	aldia	15 apr 2019	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...
6837	{'\$oid': '60b7911b1b7e26727f697533'}	guatemala	aldia	24 jun 2019	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...
6838	{'\$oid': '60b791c91b7e26727f69754b'}	guatemala	aldia	10 jun 2020	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...
6839	{'\$oid': '60b792c51b7e26727f69756c'}	guatemala	aldia	11 apr 2020	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...
6840	{'\$oid': '60b795131b7e26727f6975ae'}	guatemala	aldia	17 dec 2016	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...

6841 rows × 6 columns

In [5]: `!pip install dateparser`

Collecting dateparser

Downloading dateparser-1.2.2-py3-none-any.whl.metadata (29 kB)

Requirement already satisfied: python-dateutil>=2.7.0 in /usr/local/lib/python3.11/dist-packages (from dateparser) (2.9.0.post0)

Requirement already satisfied: pytz>=2024.2 in /usr/local/lib/python3.11/dist-packages (from dateparser) (2025.2)

Requirement already satisfied: regex>=2024.9.11 in /usr/local/lib/python3.11/dist-packages (from dateparser) (2024.11.6)

Requirement already satisfied: tzlocal>=0.2 in /usr/local/lib/python3.11/dist-packages (from dateparser) (5.3.1)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7.0->dateparser) (1.17.0)

Downloading dateparser-1.2.2-py3-none-any.whl (315 kB)

315.5/315.5 kB 9.1 MB/s eta 0:00:00

Installing collected packages: dateparser

Successfully installed dateparser-1.2.2

## Limpieza del dataset

```
In [6]: # Nueva fecha en formato "dd/mm/yyyy"
import dateparser

def parse_fecha(texto):
    if isinstance(texto, str):
        fecha = dateparser.parse(texto, languages=['es', 'en'])
        if fecha:
            return fecha.strftime('%d/%m/%Y')
    return None

datos_json['DATE_formateada'] = datos_json['DATE'].apply(parse_fecha)
datos_json
```

Out[6]:

	_id	COUNTRY	SOURCE	DATE	TITLE	BODY	DATE_formateada
0	{'\$oid': '5ef4c3ca787e0000cc000e40'}	latinoamerica	bbcmundo	3 junio 2019\n	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...	03/06/2019
1	{'\$oid': '5ef4c3ca787e0000cc000e43'}	latinoamerica	bbcmundo	21 enero 2019\n	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...	21/01/2019
2	{'\$oid': '5ef4c3ca787e0000cc000e47'}	latinoamerica	bbcmundo	8 marzo 2018\n	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...	08/03/2018
3	{'\$oid': '5ef4c3ca787e0000cc000e48'}	latinoamerica	bbcmundo	23 noviembre 2018\n	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...	23/11/2018
4	{'\$oid': '5ef4c3ca787e0000cc000e49'}	latinoamerica	bbcmundo	24 abril 2019\n	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...	24/04/2019
...	...	...	...	...	...	...	...
6836	{'\$oid': '60b78fe51b7e26727f6974f8'}	guatemala	aldia	15 apr 2019	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...	15/04/2019
6837	{'\$oid': '60b7911b1b7e26727f697533'}	guatemala	aldia	24 jun 2019	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...	24/06/2019

	<b>_id</b>	<b>COUNTRY</b>	<b>SOURCE</b>	<b>DATE</b>	<b>TITLE</b>	<b>BODY</b>	<b>DATE_formateada</b>
<b>6838</b>	{'\$oid': '60b791c91b7e26727f69754b'}	guatemala	aldia	10 jun 2020	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...	10/06/2020
<b>6839</b>	{'\$oid': '60b792c51b7e26727f69756c'}	guatemala	aldia	11 apr 2020	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...	11/04/2020
<b>6840</b>	{'\$oid': '60b795131b7e26727f6975ae'}	guatemala	aldia	17 dec 2016	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...	17/12/2016

6841 rows × 7 columns

```
In [7]: datos_json.drop(
        ['_id', 'DATE'],
        axis = 1,
        inplace = True
    )

datos_json
```

Out[7]:

	COUNTRY	SOURCE	TITLE	BODY	DATE_formateada
0	latinoamerica	bbcmundo	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...	03/06/2019
1	latinoamerica	bbcmundo	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...	21/01/2019
2	latinoamerica	bbcmundo	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...	08/03/2018
3	latinoamerica	bbcmundo	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...	23/11/2018
4	latinoamerica	bbcmundo	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...	24/04/2019
...	...	...	...	...	...
6836	guatemala	aldia	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...	15/04/2019
6837	guatemala	aldia	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...	24/06/2019
6838	guatemala	aldia	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...	10/06/2020
6839	guatemala	aldia	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...	11/04/2020
6840	guatemala	aldia	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...	17/12/2016

6841 rows × 5 columns

In [8]: *#Renombramos las columnas*

```
In [9]: datos_json.rename( columns = {'COUNTRY':'pais', 'SOURCE':'fuente', 'TITLE':'titulo', 'BODY': 'descripcion', 'DATE_forr
datos_json
```



Out[9]:

	pais	fuelle	titulo	descripcion	fecha
0	latinoamerica	bbcmundo	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...	03/06/2019
1	latinoamerica	bbcmundo	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...	21/01/2019
2	latinoamerica	bbcmundo	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...	08/03/2018
3	latinoamerica	bbcmundo	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...	23/11/2018
4	latinoamerica	bbcmundo	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...	24/04/2019
...	...	...	...	...	...
6836	guatemala	aldia	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...	15/04/2019
6837	guatemala	aldia	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...	24/06/2019
6838	guatemala	aldia	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...	10/06/2020
6839	guatemala	aldia	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...	11/04/2020
6840	guatemala	aldia	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...	17/12/2016

6841 rows × 5 columns

```
In [10]: datos_json = datos_json[['fecha', 'pais', 'fuente', 'titulo', 'descripcion']]

datos_json
```

Out[10]:

	fecha	pais	fuelle	titulo	descripcion
0	03/06/2019	latinoamerica	bbcmundo	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...
1	21/01/2019	latinoamerica	bbcmundo	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...
2	08/03/2018	latinoamerica	bbcmundo	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...
3	23/11/2018	latinoamerica	bbcmundo	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...
4	24/04/2019	latinoamerica	bbcmundo	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...
...	...	...	...	...	...
6836	15/04/2019	guatemala	aldia	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...
6837	24/06/2019	guatemala	aldia	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...
6838	10/06/2020	guatemala	aldia	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...
6839	11/04/2020	guatemala	aldia	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...
6840	17/12/2016	guatemala	aldia	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...

6841 rows × 5 columns

In [11]: `#Eliminamos los espacios en blanco`

In [12]: `datos_json= datos_json.dropna(subset=['fecha','pais','fuente','titulo','descripcion'])`  
`datos_json`

Out[12]:

	fecha	pais	fuelle	titulo	descripcion
0	03/06/2019	latinoamerica	bbcmundo	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...
1	21/01/2019	latinoamerica	bbcmundo	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...
2	08/03/2018	latinoamerica	bbcmundo	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...
3	23/11/2018	latinoamerica	bbcmundo	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...
4	24/04/2019	latinoamerica	bbcmundo	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...
...	...	...	...	...	...
6836	15/04/2019	guatemala	aldia	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...
6837	24/06/2019	guatemala	aldia	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...
6838	10/06/2020	guatemala	aldia	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...
6839	11/04/2020	guatemala	aldia	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...
6840	17/12/2016	guatemala	aldia	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...

4844 rows × 5 columns

In [13]: `#Quitamos los duplicados`

In [14]: `datos_json = datos_json.drop_duplicates(subset = ['titulo', 'descripcion'])`  
`datos_json`

Out[14]:

	fecha	pais	fuelle	titulo	descripcion
0	03/06/2019	latinoamerica	bbcmundo	Qué fue el "genocidio canadiense" por el que m...	Han pasado casi 50 años desde que Helen Betty ...
1	21/01/2019	latinoamerica	bbcmundo	Feminicidio en Ecuador: la advertencia del pre...	"Les hemos abierto las puertas, pero no vamos ...
2	08/03/2018	latinoamerica	bbcmundo	Cómo Los Cabos, uno de los destinos turísticos...	El asesinato fue al mediodía. Cuando salió de ...
3	23/11/2018	latinoamerica	bbcmundo	Caravana de migrantes en Tijuana: "Salí de Hon...	Loli, una mujer transexual de 26 años de El Sa...
4	24/04/2019	latinoamerica	bbcmundo	"Los monstruos de Ecatepec": la pareja acusada...	Juan Carlos N. y Patricia N., la pareja de ase...
...	...	...	...	...	...
6836	15/04/2019	guatemala	aldia	La asesinan y le prenden fuego a su cadaver	Desconocidos desde de un vehículo pasaron aban...
6837	24/06/2019	guatemala	aldia	Capturan a descuartizadores de niña	Luego de cometer el horrendo crimen contra una...
6838	10/06/2020	guatemala	aldia	Violador ofrecía empleo para engañar a sus víc...	Agentes de la Policía Nacional Civil capturaro...
6839	11/04/2020	guatemala	aldia	Es asesinada por su marido enfrente de sus 2 h...	Este martes durante el Toque de Queda, se regi...
6840	17/12/2016	guatemala	aldia	Mujer violada y atravesada muere tras agonizar...	Irma Ferreyra da Rocha era una empleada domést...

4673 rows × 5 columns

## 2.- Preprocesamiento

```
In [15]: # Convertir todo el texto a minúsculas
datos_json = datos_json.applymap(lambda x: x.lower() if isinstance(x, str) else x)
```

datos\_json

Out[15]:

	fecha	pais	fuelle	titulo	descripcion
0	03/06/2019	latinoamerica	bbcmundo	qué fue el "genocidio canadiense" por el que m...	han pasado casi 50 años desde que helen betty ...
1	21/01/2019	latinoamerica	bbcmundo	feminicidio en ecuador: la advertencia del pre...	"les hemos abierto las puertas, pero no vamos ...
2	08/03/2018	latinoamerica	bbcmundo	cómo los cabos, uno de los destinos turísticos...	el asesinato fue al mediodía. cuando salió de ...
3	23/11/2018	latinoamerica	bbcmundo	caravana de migrantes en tijuana: "salí de hon...	loli, una mujer transexual de 26 años de el sa...
4	24/04/2019	latinoamerica	bbcmundo	"los monstruos de ecatepec": la pareja acusada...	juan carlos n. y patricia n., la pareja de ase...
...	...	...	...	...	...
6836	15/04/2019	guatemala	aldia	la asesinan y le prenden fuego a su cadaver	desconocidos desde de un vehículo pasaron aban...
6837	24/06/2019	guatemala	aldia	capturan a descuartizadores de niña	luego de cometer el horrendo crimen contra una...
6838	10/06/2020	guatemala	aldia	violador ofrecía empleo para engañar a sus víc...	agentes de la policía nacional civil capturaro...
6839	11/04/2020	guatemala	aldia	es asesinada por su marido enfrente de sus 2 h...	este martes durante el toque de queda, se regi...
6840	17/12/2016	guatemala	aldia	mujer violada y atravesada muere tras agonizar...	irma ferreyra da rocha era una empleada domést...

4673 rows × 5 columns

```
In [16]: #Cargamos paqueterías
import unicodedata
import string
```

```
In [17]: # Quitar signos de puntuación, incluyendo comas
datos_json = datos_json.applymap(
    lambda x: ''.join(c for c in unicodedata.normalize('NFD', x) if unicodedata.category(c) != 'Mn').translate(str.maketrans('', '')))
datos_json
```

	fecha	pais	fuelle	titulo	descripcion
0	03062019	latinoamerica	bbcmundo	que fue el genocidio canadiense por el que mur...	han pasado casi 50 anos desde que helen betty ...
1	21012019	latinoamerica	bbcmundo	feminicidio en ecuador la advertencia del pres...	les hemos abierto las puertas pero no vamos a ...
2	08032018	latinoamerica	bbcmundo	como los cabos uno de los destinos turisticos ...	el asesinato fue al mediodia cuando salio de C...
3	23112018	latinoamerica	bbcmundo	caravana de migrantes en tijuana sali de hondu...	loli una mujer transexual de 26 anos de el sal...
4	24042019	latinoamerica	bbcmundo	los monstruos de ecatepec la pareja acusada de...	juan carlos n y patricia n la pareja de asesin...
...	...	...	...	...	...
6836	15042019	guatemala	aldia	la asesinan y le prenden fuego a su cadaver	desconocidos desde de un vehiculo pasaron aban...
6837	24062019	guatemala	aldia	capturan a descuartizadores de nina	luego de cometer el horrendo crimen contra una...
6838	10062020	guatemala	aldia	violador ofrecia empleo para enganar a sus vic...	agentes de la policia nacional civil capturaro...
6839	11042020	guatemala	aldia	es asesinada por su marido enfrente de sus 2 h...	este martes durante el toque de queda se regis...
6840	17122016	guatemala	aldia	mujer violada y atravesada muere tras agonizar...	irma ferreyra da rocha era una empleada domest...

4673 rows × 5 columns

```
In [18]: #Tokenizacion
```

```
In [19]: #Instalamos la paquetería
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.2.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.5.1)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
```

```
In [20]: import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
Out[20]: True
```

```
In [21]: #llamamos las paqueterías
from nltk.tokenize import word_tokenize
#Creamos tokens para el título de las noticias y para las descripciones
datos_json['tokens_titulo'] = datos_json['titulo'].apply(lambda x: word_tokenize(x, preserve_line=True))
datos_json['tokens_descripcion'] = datos_json['descripcion'].apply(lambda x: word_tokenize(x, preserve_line=True))
```

```
In [23]: #Resultados de los tokens del título
print(datos_json[['titulo', 'tokens_titulo']].head(10))
```

```

                                titulo \
0  que fue el genocidio canadiense por el que mur...
1  feminicidio en ecuador la advertencia del pres...
2  como los cabos uno de los destinos turisticos ...
3  caravana de migrantes en tijuana sali de hondu...
4  los monstruos de ecatepec la pareja acusada de...
5  quienes eran las cuatro mujeres asesinadas jun...
6  el brutal asesinato de una joven que hizo que ...
7  feminicidios como la muerte de la periodista k...
8  violencia contra la mujer que es el suicidio f...
9  malasia condenan a muerte a dos policias por a...

```

```

                                tokens_titulo
0  [que, fue, el, genocidio, canadiense, por, el,...
1  [feminicidio, en, ecuador, la, advertencia, de...
2  [como, los, cabos, uno, de, los, destinos, tur...
3  [caravana, de, migrantes, en, tijuana, sali, d...
4  [los, monstruos, de, ecatepec, la, pareja, acu...
5  [quienes, eran, las, cuatro, mujeres, asesinad...
6  [el, brutal, asesinato, de, una, joven, que, h...
7  [feminicidios, como, la, muerte, de, la, perio...
8  [violencia, contra, la, mujer, que, es, el, su...
9  [malasia, condenan, a, muerte, a, dos, policia...

```

```

In [24]: #Resultados de Los tokens de descripcion
print(datos_json[['descripcion', 'tokens_descripcion']].head(10))

```



```

                                descripcion \
0 han pasado casi 50 anos desde que helen betty ...
1 les hemos abierto las puertas pero no vamos a ...
2 el asesinato fue al mediodia cuando salio de c...
3 loli una mujer transexual de 26 anos de el sal...
4 juan carlos n y patricia n la pareja de asesin...
5 cinco victimas cinco tiros de gracia con un ar...
6 ocurrio al chasquido de los dedos en un moment...
7 es octubre pero en su escritorio en la redacci...
8 la certeza la agente de policia salvadorena ir...
9 el mas alto tribunal de malasia condeno este m...

```

```

                                tokens_descripcion
0 [han, pasado, casi, 50, anos, desde, que, hele...
1 [les, hemos, abierto, las, puertas, pero, no, ...
2 [el, asesinato, fue, al, mediodia, cuando, sal...
3 [loli, una, mujer, transexual, de, 26, anos, d...
4 [juan, carlos, n, y, patricia, n, la, pareja, ...
5 [cinco, victimas, cinco, tiros, de, gracia, co...
6 [ocurrio, al, chasquido, de, los, dedos, en, u...
7 [es, octubre, pero, en, su, escritorio, en, la...
8 [la, certeza, la, agente, de, policia, salvado...
9 [el, mas, alto, tribunal, de, malasia, condeno...

```

In [25]: `#Stopwords`

In [26]: `#Cargamos la paquetería para stopwords`

```

import nltk
nltk.download('stopwords')

```

[nltk\_data] Downloading package stopwords to /root/nltk\_data...

[nltk\_data] Unzipping corpora/stopwords.zip.

Out[26]: True

In [27]: `#Llamamos la paquetería`

```

from nltk.corpus import stopwords
#Creamos una lista con las palabras más comunes del español
stopwords_es = set(stopwords.words('spanish'))

```

In [28]: `# Stopwords`

```

def filtrar_stopwords(tokens):

```

```

    return [palabra for palabra in tokens if palabra not in stopwords_es]

# Una vez definido el código, se aplica a nuestros tokens realizados anteriormente, es decir a título y descripción
datos_json['tokens_titulo'] = datos_json['tokens_titulo'].apply(filtrar_stopwords)
datos_json['tokens_descripcion'] = datos_json['tokens_descripcion'].apply(filtrar_stopwords)

```

```

In [30]: #Tokens del título sin stopwords
print(datos_json['tokens_titulo'].head(10))

```

```

0    [genocidio, canadiense, murieron, desaparecier...
1    [feminicidio, ecuador, advertencia, presidente...
2    [cabos, destinos, turisticos, mexico, mas, pop...
3    [caravana, migrantes, tijuana, sali, honduras,...
4    [monstruos, ecatepec, pareja, acusada, decenas...
5    [cuatro, mujeres, asesinadas, junto, periodist...
6    [brutal, asesinato, joven, hizo, actriz, anne,...
7    [feminicidios, muerte, periodista, karla, turc...
8    [violencia, mujer, suicidio, feminicida, salva...
9    [malasia, condenan, muerte, dos, policias, ase...
Name: tokens_titulo, dtype: object

```

```

In [31]: #Tokens de la descripción sin stopwords
print(datos_json['tokens_descripcion'].head(10))

```

```

0    [pasado, casi, 50, anos, helen, betty, osborne...
1    [abierto, puertas, vamos, sacrificar, segurida...
2    [asesinato, mediodia, salio, casa, mujer, 26, ...
3    [loli, mujer, transexual, 26, anos, salvador, ...
4    [juan, carlos, n, patricia, n, pareja, asesino...
5    [cinco, victimas, cinco, tiros, gracia, arma, ...
6    [ocurrio, chasquido, dedos, momento, caida, al...
7    [octubre, escritorio, redaccion, prensa, grafi...
8    [certeza, agente, policia, salvadorena, irma, ...
9    [mas, alto, tribunal, malasia, condeno, martes...
Name: tokens_descripcion, dtype: object

```

```

In [32]: #Lematización

```

```

In [33]: !pip install spacy

```

Requirement already satisfied: spacy in /usr/local/lib/python3.11/dist-packages (3.8.7)  
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.0.12)  
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.0.5)  
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.0.13)  
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.11)  
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.0.10)  
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.11/dist-packages (from spacy) (8.3.6)  
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.1.3)  
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.5.1)  
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.10)  
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (0.4.1)  
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (0.16.0)  
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (4.67.1)  
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.2)  
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.32.3)  
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.11.7)  
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.1.6)  
Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-packages (from spacy) (75.2.0)  
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (25.0)  
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.5.0)  
Requirement already satisfied: language-data>=1.2 in /usr/local/lib/python3.11/dist-packages (from langcodes<4.0.0,>=3.2.0->spacy) (1.3.0)  
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)  
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.33.2)  
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (4.14.1)  
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.4.1)  
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4.3)  
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.10)  
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.2.3)

2.13.0->spacy) (2.5.0)  
 Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2025.8.3)  
 Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->spacy) (1.3.0)  
 Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->spacy) (0.1.5)  
 Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (8.2.1)  
 Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (1.5.4)  
 Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (13.9.4)  
 Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from weasel<0.5.0,>=0.1.0->spacy) (0.21.1)  
 Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.11/dist-packages (from weasel<0.5.0,>=0.1.0->spacy) (7.3.0.post1)  
 Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2->spacy) (3.0.2)  
 Requirement already satisfied: marisa-trie>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.1)  
 Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.11/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (4.0.0)  
 Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.19.2)  
 Requirement already satisfied: wrapt in /usr/local/lib/python3.11/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.17.3)  
 Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)

In [34]: !python -m spacy download es\_core\_news\_sm

Collecting es-core-news-sm==3.8.0

Downloading [https://github.com/explosion/spacy-models/releases/download/es\\_core\\_news\\_sm-3.8.0/es\\_core\\_news\\_sm-3.8.0-py3-none-any.whl](https://github.com/explosion/spacy-models/releases/download/es_core_news_sm-3.8.0/es_core_news_sm-3.8.0-py3-none-any.whl) (12.9 MB)

12.9/12.9 MB 80.8 MB/s eta 0:00:00

Installing collected packages: es-core-news-sm

Successfully installed es-core-news-sm-3.8.0

✓ Download and installation successful

You can now load the package via `spacy.load('es_core_news_sm')`

⚠ Restart to reload dependencies

If you are in a Jupyter or Colab notebook, you may need to restart Python in order to load all the package's dependencies. You can do this by selecting the 'Restart kernel' or 'Restart runtime' option.

In [35]: *# Llamamos la paquetería*

```
import re
import spacy
```

In [ ]: *# Cargamos el modelo de spaCy en español*

```
nlp = spacy.load('es_core_news_sm')
# Lemmatización
def lematizar_tokens(tokens):
    doc = nlp(' '.join(tokens))
    lemmas_limpios = []
    for token in doc:
        # obtenemos el lema
        lema = token.lemma_.lower().strip()
        lema = re.sub(r'^a-záéíóúüñ0-9', '', lema) ## como precaución quitamos todo lo que no sea letras/números,
        if lema and len(lema) > 1: # Filtra lemas vacíos o de un solo carácter
            lemmas_limpios.append(lema)
    return lemmas_limpios
```

*# Aplicación sobre los tokens*

```
datos_json['lemmas_descripcion'] = datos_json['tokens_descripcion'].apply(lematizar_tokens)
datos_json['lemmas_titulo'] = datos_json['tokens_titulo'].apply(lematizar_tokens)
```

In [69]: *# Para ver los resultados de la Lemmatización de los títulos en comparación con los tokens*

```
print(datos_json[['tokens_titulo', 'lemmas_titulo']].head(10))
```

```

tokens_titulo \
0 [genocidio, canadiense, murieron, desaparecier...
1 [feminicidio, ecuador, advertencia, presidente...
2 [cabos, destinos, turisticos, mexico, mas, pop...
3 [caravana, migrantes, tijuana, sali, honduras,...
4 [monstruos, ecatepec, pareja, acusada, decenas...
5 [cuatro, mujeres, asesinadas, junto, periodist...
6 [brutal, asesinato, joven, hizo, actriz, anne...
7 [feminicidios, muerte, periodista, karla, turc...
8 [violencia, mujer, suicidio, feminicida, salva...
9 [malasia, condenan, muerte, dos, policias, ase...

```

```

lemmas_titulo
0 [genocidio, canadiense, morir, desaparecer, ci...
1 [feminicidio, ecuador, advertenciar, president...
2 [cabo, destino, turistico, mexico, mas, popula...
3 [caravanar, migrant, tijuán, sali, honduras, m...
4 [monstruo, ecatepec, parejo, acusado, decena, ...
5 [cuatro, mujer, asesinado, junto, periodista, ...
6 [brutal, asesinato, joven, hacer, actriz, anne...
7 [feminicidio, muerte, periodista, karél, turci...
8 [violencia, mujer, suicidio, feminicida, salva...
9 [malasia, condenar, muerte, dos, policia, ases...

```

```

In [70]: # Para ver los resultados de la Lemmatización de la descripción en comparación con los tokens
print(datos_json[['tokens_descripcion', 'lemmas_descripcion']].head(10))

```

```

tokens_descripcion \
0 [pasado, casi, 50, anos, helen, betty, osborne...
1 [abierto, puertas, vamos, sacrificar, segurida...
2 [asesinato, mediodia, salio, casa, mujer, 26, ...
3 [loli, mujer, transexual, 26, anos, salvador, ...
4 [juan, carlos, n, patricia, n, pareja, asesino...
5 [cinco, victimas, cinco, tiros, gracia, arma, ...
6 [ocurrio, chasquido, dedos, momento, caida, al...
7 [octubre, escritorio, redaccion, prensa, grafi...
8 [certeza, agente, policia, salvadorena, irma, ...
9 [mas, alto, tribunal, malasia, condeno, martes...

```

```

lemmas_descripcion
0 [pasado, casi, 50, ano, helar, betty, osborne,...
1 [abierto, puerta, ir, sacrificar, seguridad, n...
2 [asesinato, mediodia, salio, casa, mujer, 26, ...
3 [loli, mujer, transexual, 26, ano, salvador, e...
4 [juan, carlos, patricia, pareja, asesino, seri...
5 [cinco, victima, cinco, tiro, gracia, arma, ca...
6 [ocurrio, chasquido, dedo, momento, caido, alf...
7 [octubre, escritorio, redaccion, prensa, grafi...
8 [certeza, agente, policar, salvadorén, irma, ...
9 [mas, alto, tribunal, malasia, condeno, martes...

```

### 3.- Análisis Temático

## Análisis de Sentimientos

```

In [41]: #Cargamos las paqueterías
import spacy
nlp = spacy.load('es_core_news_sm')

```

```

In [72]: #Debido a que el número de tokens es muy grande para un análisis de sentimientos pre-entrenado (como son el transformador BERT y GPT-3)
#y superan los límites de la memoria de nuestro dispositivo.
#obtenemos por hacer un análisis de sentimientos de manera manual, para ello vamos a entrenar el modelo una lista de palabras
#positivas y negativas que debe identificar el modelo.
palabras_positivas = ['justicia', 'esperanza', 'proteccion', 'rescate', 'seguridad', 'solidaridad', 'derecho', 'igualdad', 'valiente', 'defensa', 'apoyo', 'tribunal', 'ley', 'sentencia', 'condena', 'denuncia', 'captura']

```

```

        'detencion', 'libertad', 'paz', 'superviviente', 'sobrevivio', 'empoderamiento', 'ayuda', 'avance',
        'victoria', 'cambio', 'progreso', 'educacion', 'salvar', 'proteccion', 'refugio', 'esperanzador',
palabras_negativas = ['violencia', 'feminicidio', 'asesinato', 'muerte', 'homicidio', 'desaparicion', 'golpe', 'abuso',
        'violacion', 'agresion', 'crimen', 'asesino', 'victima', 'tortura', 'acoso', 'machista',
        'discriminacion', 'secuestro', 'amenaza', 'impunidad', 'terror', 'horror', 'dolor', 'sufrimiento',
        'cadaver', 'matanza', 'corrupcion', 'tragedia', 'luto', 'descuartizado', 'enfermedad', 'miedo',

```

```

In [73]: #Código de sentimientos
def detectar_sentimiento_manual(fila):
    lemas = fila['lemmas_titulo'] + fila['lemmas_descripcion']
    score = 0
    palabras_detectadas = []

    for lema in lemas:
        lema = lema.lower()
        if lema in palabras_positivas:
            score += 1
            palabras_detectadas.append((lema, 'positivo'))
        elif lema in palabras_negativas:
            score -= 1
            palabras_detectadas.append((lema, 'negativo'))

    if score > 0:
        sentimiento = 'positivo'
    elif score < 0:
        sentimiento = 'negativo'
    else:
        sentimiento = 'neutral'

    return {
        'sentimiento': sentimiento,
        'score': score,
        'palabras_detectadas': palabras_detectadas
    }

```

```

In [74]: datos_json['analisis_sentimiento'] = datos_json.apply(detector_sentimiento_manual, axis=1)

# Columnas de Los sentimientos
datos_json['sentimiento'] = datos_json['analisis_sentimiento'].apply(lambda x: x['sentimiento'])
datos_json['score'] = datos_json['analisis_sentimiento'].apply(lambda x: x['score'])
datos_json['palabras_detectadas'] = datos_json['analisis_sentimiento'].apply(lambda x: x['palabras_detectadas'])

```



```
# Resultados
```

```
print(datos_json[['titulo', 'sentimiento', 'score', 'palabras_detectadas']].head(10))
```

	titulo	sentimiento	score	\
0	que fue el genocidio canadiense por el que mur...	negativo	-19	
1	feminicidio en ecuador la advertencia del pres...	negativo	-6	
2	como los cabos uno de los destinos turisticos ...	negativo	-16	
3	caravana de migrantes en tijuana sali de hondu...	negativo	-11	
4	los monstruos de ecatepec la pareja acusada de...	negativo	-8	
5	quienes eran las cuatro mujeres asesinadas jun...	negativo	-10	
6	el brutal asesinato de una joven que hizo que ...	negativo	-3	
7	feminicidios como la muerte de la periodista k...	negativo	-39	
8	violencia contra la mujer que es el suicidio f...	negativo	-9	
9	malasia condenan a muerte a dos policias por a...	negativo	-2	

	palabras_detectadas
0	[(muerte, negativo), (justicia, positivo), (vi...
1	[(feminicidio, negativo), (asesinato, negativo...
2	[(asesinato, negativo), (crimen, negativo), (s...
3	[(violencia, negativo), (sentencia, positivo),...
4	[(asesinato, negativo), (condena, positivo), (...
5	[(victima, negativo), (asesinato, negativo), (...
6	[(asesinato, negativo), (asesinato, negativo),...
7	[(feminicidio, negativo), (muerte, negativo), ...
8	[(violencia, negativo), (muerte, negativo), (a...
9	[(muerte, negativo), (asesinato, negativo), (t...

In [75]: *#Realizamos un gráfico con la distribución de las palabras entrenadas en el análisis de sentimientos.*

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Estilo del gráfico
```

```
sns.set(style="whitegrid")
```

```
# Conteo de los sentimientos
```

```
conteo = datos_json['sentimiento'].value_counts()
```

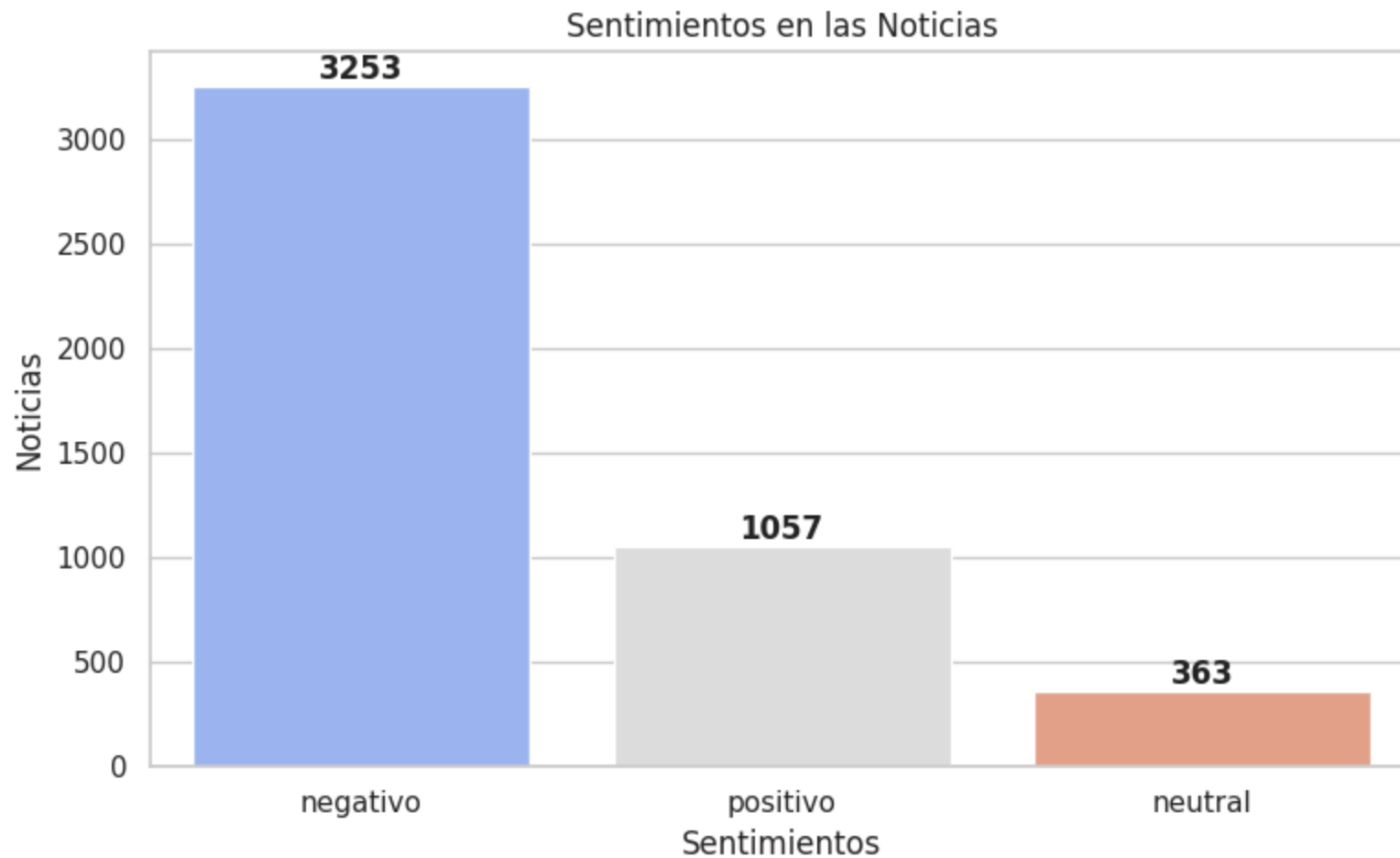
```
# Creación de nuestro gráfico
```

```
plt.figure(figsize=(8, 5))
```

```
ax = sns.barplot(x=conteo.index, y=conteo.values, palette='coolwarm')
```

```
# Agregamos las etiquetas arriba de cada barra
for i, valor in enumerate(conteo.values):
    ax.text(i, valor + 0.5, str(valor), ha='center', va='bottom', fontsize=12, fontweight='bold')

# Títulos y etiquetas de nuestro gráfico
plt.title('Sentimientos en las Noticias')
plt.xlabel('Sentimientos')
plt.ylabel('Noticias')
plt.tight_layout()
plt.show()
```



In [76]: #Las noticias analizadas después de la limpieza de datos es: 4673 de este total:  
 #3253 equivale al 69.6% de las noticias negativas.

*#1057 es igual al 22.6% de las noticias positivas y  
#366 corresponde al 7.7% de las noticias neutrales, puesto que no obtuvieron una clasificación dentro del modelo.*

In [77]: *#Ahora vamos a analizar las palabras más comunes mencionadas en los títulos de las noticias*

```
from collections import Counter
import itertools

# Unir todos los lemmas de descripciones
todas_palabras = list(itertools.chain.from_iterable(datos_json['lemmas_titulo']))

# Contar frecuencia
titulos = Counter(todas_palabras)

# Llamamos las 10 más comunes
print(titulos.most_common(10))
```

```
[('mujer', 923), ('femicidio', 539), ('año', 420), ('violencia', 394), ('hombre', 258), ('feminicidio', 240), ('caso', 235), ('asesinado', 208), ('género', 204), ('pareja', 180)]
```

In [78]: *#Llamamos la librería*

```
import matplotlib.pyplot as plt

# Counter
top_palabras = titulos.most_common(10)
palabras, conteos = zip(*top_palabras)

# Creamos la figura y ejes
fig, ax = plt.subplots(figsize=(10, 6))

# Hacemos los gráficos de barras horizontales
ax.barh(palabras, conteos, color="green")
ax.set_xlabel("Frecuencia")
ax.set_ylabel("Palabras")
ax.set_title("10 palabras más comunes mencionadas en los títulos de las noticias")
ax.invert_yaxis()

# Agregamos las etiquetas a cada barra
for i, valor in enumerate(conteos):
    ax.text(valor + 0.5, i, str(valor), va='center', fontsize=12, fontweight='bold')

plt.tight_layout()
plt.show()
```



```
In [79]: #Como se observa en el gráfico anterior las palabras más comunes en los títulos de noticias son:  
#1.- El sexo  
#2.- El delito de feminicidio.  
#3.- El año en el que se llevó a cabo el descubrimiento o año de la víctima
```

```
In [80]: #Ahora vamos a analizar las palabras más comunes mencionadas en el cuerpo de las noticias  
from collections import Counter  
import itertools  
  
# Unir todos los lemmas de descripciones  
todas_palabras = list(itertools.chain.from_iterable(datos_json['lemmas_descripcion']))
```

```

# Contar frecuencia
descripcion = Counter(todas_palabras)

# Llamamos las 10 más comunes
print(descripcion.most_common(10))

```

```

[('mujer', 17471), ('ano', 13423), ('mas', 8990), ('violencia', 8902), ('ser', 8049), ('decir', 7540), ('hacer', 7144), ('caso', 7103), ('poder', 7092), ('tambien', 4992)]

```

```

In [81]: #Llamamos la librería
import matplotlib.pyplot as plt

# Counter
top_palabras = descripcion.most_common(10)
palabras, conteos = zip(*top_palabras)

# Creamos la figura y ejes
fig, ax = plt.subplots(figsize=(10, 6))

# Hacemos los gráficos de barras horizontales
ax.barh(palabras, conteos, color="orange")
ax.set_xlabel("Frecuencia")
ax.set_ylabel("Palabras")
ax.set_title("10 palabras más comunes mencionadas en los textos de las noticias")
ax.invert_yaxis()

# Agregamos las etiquetas a cada barra
for i, valor in enumerate(conteos):
    ax.text(valor + 0.5, i, str(valor), va='center', fontsize=12, fontweight='bold')

plt.tight_layout()
plt.show()

```



In [83]: *#Podemos concluir que la descripción de las noticias se centra en 3 puntos importante:*  
*#1.- El sexo*  
*#2.- El año de nacimiento o publicación de la noticia, y*  
*#3.- El tipo de violencia que se ejerció contra la víctima*

In [84]: *#Fin del código*