



# Centro Universitario de la Defensa en la Escuela Naval Militar

## TRABAJO FIN DE GRADO

*Análisis y visualización de datos de redes sociales sobre el  
conflicto de Ucrania*

**Grado en Ingeniería Mecánica**

**ALUMNO:** Jaime López Garay

**DIRECTORES:** Norberto Fernández García

**CURSO ACADÉMICO:** 2023-2024

Universida<sub>de</sub>Vigo





# Centro Universitario de la Defensa en la Escuela Naval Militar

## TRABAJO FIN DE GRADO

*Análisis y visualización de datos de redes sociales sobre el  
conflicto de Ucrania*

**Grado en Ingeniería Mecánica**  
Intensificación en Tecnología Naval  
Cuerpo General

UniversidadeVigo



# **RESUMEN**

La notable evolución de las redes sociales ha supuesto un cambio en el panorama mediático, presentando una transición desde los métodos tradicionales de comunicación, como los periódicos, a la publicación de noticias en redes sociales. La difusión de contenido a través de las redes sociales ha democratizado la transferencia de información, permitiendo una interacción más activa entre usuarios y productores de contenido.

Este trabajo se enmarca en el contexto de la inteligencia de fuentes abiertas en el ámbito de defensa, empleando la red social X/Twitter como fuente de información. La publicación de noticias en X/Twitter por diferentes usuarios registrados permite informarse sobre los eventos actuales en el ámbito de defensa, al igual que conocer la opinión pública sobre un tema de interés global. Esta nueva forma de comunicación facilita la recopilación de información y por consiguiente la realización de un análisis más exhaustivo que permita comprender las causas de los eventos bélicos que se producen en la actualidad.

Por lo tanto, el presente trabajo describe el análisis de una cantidad limitada de datos extraídos de X/Twitter mediante una herramienta de inteligencia de código abierto, la cual permite el procesamiento de estos datos de interés y se postula como un recurso valioso para el examen de una gran cantidad de información de manera eficiente y detallada.

## **PALABRAS CLAVE**

X/Twitter, Redes sociales, Visualización, OSINT



## **AGRADECIMIENTOS**

En primer lugar, querría agradecer a mi tutor, Don Norberto Fernández González, por el tiempo dedicado a este Trabajo Fin de Grado y por la guía que ha desempeñado durante su desarrollo, haciendo posible alcanzar los objetivos de este proyecto de investigación sin ningún contratiempo.

Al LV Grigori Nicolaeff, por haber sido el oficial más didáctico e involucrado durante mis 5 años de escuela. Oficial cuya involucración y contribución fueron determinantes en el alcance del perfil funcional de francés, en la prueba de la Escuela Militar de Idiomas.

A mis compañeros, por enseñarme que la camaradería y el apoyo moral son un pilar fundamental hacia el logro de las metas individuales y saber que siempre podré confiar en ellos.

Por último, a mi hermano Miguel, cuya dedicación y compromiso en la Armada han sido mi guía hacia el desarrollo profesional y por ser una referencia en la que confiar durante numerosas tomas de decisiones.



## CONTENIDO

|                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------|----|
| Contenido .....                                                                                               | 1  |
| Índice de Figuras .....                                                                                       | 4  |
| Índice de Tablas.....                                                                                         | 7  |
| 1 Introducción y objetivos .....                                                                              | 8  |
| 1.1 Contexto y motivación .....                                                                               | 8  |
| 1.2 Objetivos y metodología .....                                                                             | 10 |
| 1.3 Estructura de la memoria .....                                                                            | 10 |
| 2 Estado del arte .....                                                                                       | 12 |
| 2.1 La información en las redes sociales.....                                                                 | 12 |
| 2.1.1 Inicios .....                                                                                           | 12 |
| 2.1.2 Popularidad de las redes sociales.....                                                                  | 12 |
| 2.2 Twitter.....                                                                                              | 13 |
| 2.2.1 Historia de X/Twitter .....                                                                             | 13 |
| 2.2.2 Terminología de X/Twitter .....                                                                         | 16 |
| 2.2.3 Influencia en las noticias .....                                                                        | 17 |
| 2.3 OSINT.....                                                                                                | 17 |
| 2.3.1 Evolución de OSINT .....                                                                                | 17 |
| 2.3.2 Proceso de OSINT .....                                                                                  | 18 |
| 2.4 Conflicto entre Rusia y Ucrania.....                                                                      | 19 |
| 2.4.1 Antecedentes.....                                                                                       | 19 |
| 2.4.2 Intereses geopolíticos.....                                                                             | 21 |
| 2.4.3 Papel de la OTAN.....                                                                                   | 22 |
| 2.4.4 Consecuencias del conflicto.....                                                                        | 22 |
| 2.5 Herramientas utilizadas para el análisis de datos.....                                                    | 23 |
| 2.5.1 Elasticsearch .....                                                                                     | 23 |
| 2.5.2 Kibana.....                                                                                             | 24 |
| 2.6 Alternativas a Kibana.....                                                                                | 24 |
| 2.7 Trabajos académicos relacionados.....                                                                     | 26 |
| 2.7.1 Herramientas de OSINT para la obtención de información en Twitter sobre el conflicto Ucrania-Rusia..... | 26 |
| 2.7.2 Sistema de detección inteligente de usuarios de Twitter interesados en la temática de defensa.....      | 26 |
| 2.7.3 Plataforma OSINT para el almacenamiento y análisis inteligente de datos de la red social Twitter .....  | 26 |

|                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------|----|
| 2.7.4 Análisis de las redes sociales: La guerra de Ucrania .....                                                               | 26 |
| 2.7.5 Análisis y visualización de datos abiertos de carácter informativo para el alumnado de la Universidad de Valladolid..... | 26 |
| 2.7.6 Investigación y extracción de datos con técnicas OSINT.....                                                              | 26 |
| 3 Desarrollo del TFG.....                                                                                                      | 28 |
| 3.1 Instalación de la máquina virtual .....                                                                                    | 28 |
| 3.2 Instalación herramientas a utilizar .....                                                                                  | 29 |
| 3.2.1 Python.....                                                                                                              | 29 |
| 3.2.2 Elasticsearch .....                                                                                                      | 31 |
| 3.2.3 Kibana.....                                                                                                              | 32 |
| 3.3 Tratamiento de los ficheros de datos.....                                                                                  | 33 |
| 3.3.1 Descompresión .....                                                                                                      | 33 |
| 3.3.2 Limpieza y extracción de información de interés.....                                                                     | 33 |
| 3.4 Carga de archivos en Elasticsearch.....                                                                                    | 35 |
| 3.4.1 Creación de un índice .....                                                                                              | 35 |
| 3.4.2 Indexación de datos .....                                                                                                | 37 |
| 3.5 Elección tipo de visualización para cada dato .....                                                                        | 37 |
| 3.5.1 Lens.....                                                                                                                | 38 |
| 3.5.2 TSVB .....                                                                                                               | 40 |
| 3.5.3 Aggregation based .....                                                                                                  | 42 |
| 4 Prueba y validación .....                                                                                                    | 45 |
| 4.1 Creación de las visualizaciones de datos .....                                                                             | 45 |
| 4.1.1 Autor.....                                                                                                               | 46 |
| 4.1.2 Fecha de creación .....                                                                                                  | 47 |
| 4.1.3 Etiquetas .....                                                                                                          | 48 |
| 4.1.4 Menciones.....                                                                                                           | 49 |
| 4.1.5 Anotaciones .....                                                                                                        | 51 |
| 4.1.6 Retuits.....                                                                                                             | 52 |
| 4.1.7 Me gusta (Likes).....                                                                                                    | 54 |
| 5 Conclusiones y líneas futuras .....                                                                                          | 58 |
| 5.1 Conclusiones .....                                                                                                         | 58 |
| 5.2 Líneas futuras .....                                                                                                       | 59 |
| 6 Bibliografía.....                                                                                                            | 60 |
| Anexo I: Implicaciones Sociales, Económicas, y Ambientales .....                                                               | 65 |
| Anexo II: Reflexiones Éticas y Sociales .....                                                                                  | 66 |

|                                                        |    |
|--------------------------------------------------------|----|
| Anexo III: Explicación de los códigos realizados ..... | 67 |
|--------------------------------------------------------|----|

## ÍNDICE DE FIGURAS

|                                                                                                                      |    |
|----------------------------------------------------------------------------------------------------------------------|----|
| Figura 1-1 Gráfico representativo de las aplicaciones más usadas globalmente en 2024. (We are social) .....          | 9  |
| Figura 1-2 Celebrities y medios de comunicación en la red social X/Twitter. (X/Twitter) .....                        | 9  |
| Figura 1-3 Información en tiempo real del conflicto Rusia-Ucrania a través de X/Twitter. (X/Twitter) .....           | 10 |
| Figura 2-1 Adultos que citan el medio como su principal fuente de noticias en porcentaje (Pew Research Center) ..... | 13 |
| Figura 2-2 Medio principal de fuente de noticias por grupo de edad (Pew Research Center) .....                       | 13 |
| Figura 2-3 Primer tuit de la historia. (Marketing4ecommerce) .....                                                   | 14 |
| Figura 2-4 Primer tuit que utiliza la herramienta <i>hashtag</i> . (Marketing4ecommerce) .....                       | 14 |
| Figura 2-5 Objetivos de Elon Musk tras la adquisición de Twitter. (IEBS) .....                                       | 15 |
| Figura 2-6 Cuenta X/Twitter de la Policía Nacional. (Autoría propia) .....                                           | 16 |
| Figura 2-7 Retuit de la cuenta de Armada. (Autoría propia) .....                                                     | 17 |
| Figura 2-8 Documentación gráfico sobre eventos ocurridos en Ucrania. (X/Twitter) .....                               | 18 |
| Figura 2-9 Ciclo de OSINT. (Autoría propia) .....                                                                    | 19 |
| Figura 2-10 Situación geopolítica durante la Guerra Fría. (Britannica) .....                                         | 20 |
| Figura 2-11 Situación geopolítica de la OTAN. (NATO) .....                                                           | 20 |
| Figura 2-12 Porcentaje de la población que indicó ruso como su lengua materna. (Censo Ucrania 2001) .....            | 21 |
| Figura 2-13 Territorio anexionado por Rusia. (El Mundo) .....                                                        | 22 |
| Figura 2-14 <i>Dashboard</i> de Kibana. (Autoría propia) .....                                                       | 24 |
| Figura 3-1 Creación de la máquina virtual. (Autoría propia) .....                                                    | 28 |
| Figura 3-2 Parámetros de la máquina virtual. (Autoría propia) .....                                                  | 29 |
| Figura 3-3 Comando para instalar Python 3.11.6. (Autoría propia) .....                                               | 30 |
| Figura 3-4 Comprobación de la versión de Python instalada. (Autoría propia) .....                                    | 30 |
| Figura 3-5 Comando para instalar la librería <i>simplejson</i> en Python. (Autoría propia) .....                     | 30 |
| Figura 3-6 Instalación de PyCharm. (Autoría propia) .....                                                            | 30 |
| Figura 3-7 Comando para descargar Elasticsearch desde la web. (Autoría propia) .....                                 | 31 |
| Figura 3-8 Descompresión de la carpeta de Elasticsearch. (Autoría propia) .....                                      | 31 |
| Figura 3-9 Servidor de Elasticsearch en localhost. (Autoría propia) .....                                            | 31 |
| Figura 3-10 Librerías instaladas en el editor de texto PyCharm. (Autoría propia) .....                               | 32 |
| Figura 3-11 Comando para instalar Kibana desde la web. (Autoría propia) .....                                        | 32 |
| Figura 3-12 Descompresión de la carpeta de Kibana. (Autoría propia) .....                                            | 32 |
| Figura 3-13 Servidor de Kibana en localhost. (Autoría propia) .....                                                  | 32 |
| Figura 3-14 Descompresión de ficheros de datos con información de interés. (Autoría propia) .....                    | 33 |
| Figura 3-15 Carpeta en la que se almacenan los ficheros descomprimidos. (Autoría propia) .....                       | 33 |

|                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------|----|
| Figura 3-16 Comando para limpieza de los ficheros. (Autoría propia).....                                         | 33 |
| Figura 3-17 Fichero de datos en formato JSON. (Autoría propia).....                                              | 34 |
| Figura 3-18 Comando de instalación librería demoji. (Autoría propia).....                                        | 34 |
| Figura 3-19 Eliminación de información que no es de interés. (Autoría propia).....                               | 34 |
| Figura 3-20 Parámetros almacenados en el diccionario de salida. (Autoría propia).....                            | 34 |
| Figura 3-21 Definición parámetros a eliminar y lectura texto. (Autoría propia) .....                             | 35 |
| Figura 3-22 Comando para la extracción de los datos de interés de los ficheros. (Autoría propia)                 | 35 |
| Figura 3-23 Almacenamiento de los archivos limpiados y los extraídos. (Autoría propia).....                      | 35 |
| Figura 3-24 Extracto del código para la creación de un índice en Elasticsearch. (Autoría propia)                 | 35 |
| Figura 3-25 <i>Mappings</i> definidos en el índice creado en Elasticsearch. (Autoría propia) .....               | 36 |
| Figura 3-26 Creación del índice creado en la plataforma Elasticsearch. (Autoría propia) .....                    | 36 |
| Figura 3-27 Creación de un patrón del índice en Elasticsearch. (Autoría propia) .....                            | 36 |
| Figura 3-28 Comando para la indexación de los ficheros de datos en Elasticsearch. (Autoría propia) .....         | 37 |
| Figura 3-29 Tuits cargados en Elasticsearch. (Autoría propia).....                                               | 37 |
| Figura 3-30 Creación de visualizaciones de datos. (Autoría propia) .....                                         | 38 |
| Figura 3-31 Visualización de la herramienta <i>metric</i> . (Autoría propia) .....                               | 38 |
| Figura 3-32 Visualización de la herramienta <i>table</i> . (Autoría propia) .....                                | 39 |
| Figura 3-33 Visualización de la herramienta <i>bar</i> . (Autoría propia).....                                   | 39 |
| Figura 3-34 Visualización de la herramienta <i>line</i> . (Autoría propia) .....                                 | 39 |
| Figura 3-35 Visualización de la herramienta <i>proportion</i> . (Autoría propia) .....                           | 40 |
| Figura 3-36 Visualización de la herramienta <i>heatmap</i> . (Autoría propia).....                               | 40 |
| Figura 3-37 Visualización de la herramienta <i>Time Series</i> . (Autoría propia).....                           | 41 |
| Figura 3-38 Visualización de la herramienta <i>metric</i> . (Autoría propia) .....                               | 41 |
| Figura 3-39 Visualización de la herramienta <i>Top N</i> . (Autoría propia).....                                 | 41 |
| Figura 3-40 Visualización de la herramienta <i>Gauge</i> . (Autoría propia).....                                 | 41 |
| Figura 3-41 Visualización de la herramienta <i>área</i> . (Autoría propia) .....                                 | 42 |
| Figura 3-42 Visualización de la herramienta <i>Gauge en Aggregation based</i> . (Autoría propia).....            | 42 |
| Figura 3-43 Visualización de la herramienta <i>Goal en Aggregation based</i> . (Autoría propia) .....            | 43 |
| Figura 3-44 Comparativa de las anotaciones más comunes mediante la opción <i>metric</i> . (Autoría propia) ..... | 43 |
| Figura 3-45 Visualización de <i>table en Aggregation based</i> . (Autoría propia).....                           | 44 |
| Figura 3-46 Visualización de la herramienta <i>Tag cloud</i> . (Autoría propia) .....                            | 44 |
| Figura 4-1 Creación de un <i>dashboard</i> en Kibana. (Autoría propia).....                                      | 45 |
| Figura 4-2 Visualizaciones de usuarios más activos mediante <i>Aggregation based</i> . (Autoría propia)          | 46 |

|                                                                                                                     |    |
|---------------------------------------------------------------------------------------------------------------------|----|
| Figura 4-3 Visualizaciones de la cantidad de tuits por fecha. (Autoría propia) .....                                | 47 |
| Figura 4-4 Visualización de las etiquetas más populares. (Autoría propia).....                                      | 48 |
| Figura 4-5 Comparativa en el tiempo de diferentes etiquetas. (Autoría propia).....                                  | 49 |
| Figura 4-6 Volumen de tuits con la etiqueta #Bakmut por fecha. (Autoría propia).....                                | 49 |
| Figura 4-7 Visualización de los usuarios más mencionados. (Autoría propia).....                                     | 50 |
| Figura 4-8 Comparativa de las tres menciones más comunes. (Autoría propia).....                                     | 50 |
| Figura 4-9 Visualizaciones de las anotaciones más comunes. (Autoría propia) .....                                   | 51 |
| Figura 4-10 Cantidad de veces que se menciona a Rusia y a Ucrania. (Autoría propia) .....                           | 52 |
| Figura 4-11 Visualización de los parámetros más retuiteados. (Autoría propia).....                                  | 52 |
| Figura 4-12 Cantidad de retuits por fecha. (Autoría propia) .....                                                   | 53 |
| Figura 4-13 Top 3 tuits más populares según el número de retuits. (Autoría propia).....                             | 54 |
| Figura 4-14 Autores más influyentes según el número total de retuits que reciben. (Autoría propia)<br>.....         | 54 |
| Figura 4-15 Popularidad según el número de <i>likes</i> . (Autoría propia) .....                                    | 55 |
| Figura 4-16 Tabla resumen de los tuits más populares y su autor. (Autoría propia) .....                             | 56 |
| Figura 4-17 Top 5 tuits más populares según el número de “me gusta”. (Autoría propia).....                          | 56 |
| Figura 4-18 Tuit más popular según el número de likes y retuits. [73] .....                                         | 57 |
| Figura 4-19 <i>Dashboard</i> relativo a las gráficas del capítulo de prueba y validación. (Autoría propia)<br>..... | 57 |
|                                                                                                                     |    |
| Figura III - 1 Código para la limpieza de archivos. (Autoría propia) .....                                          | 67 |
| Figura III - 2 Código para la extracción de datos de interés. (Autoría propia) .....                                | 69 |
| Figura III - 3 Código para la creación de un índice en Elasticsearch. (Autoría propia).....                         | 70 |
| Figura III - 4 Código para la indexación de archivos en el índice. (Autoría propia) .....                           | 71 |

## ÍNDICE DE TABLAS

|                                                                                            |    |
|--------------------------------------------------------------------------------------------|----|
| Tabla 1 Relaciones de la OTAN y Rusia con Ucrania. (Autoría propia) .....                  | 22 |
| Tabla 2 Comparativa presupuestos del Ministerio de Defensa. (Autoría propia) .....         | 23 |
| Tabla 3 Comparativa de las diferentes herramientas de visualización. (Autoría propia)..... | 25 |

# 1 INTRODUCCIÓN Y OBJETIVOS

## 1.1 Contexto y motivación

En la actualidad, el uso de redes sociales como X/Twitter utilizadas como medio de información ha experimentado un exponencial crecimiento. Esta plataforma en concreto (X/Twitter) ha revolucionado el consumo de noticias por la sociedad permitiendo llegar a un público más extenso y facilitando la transmisión de información sobre eventos en tiempo real.

X/Twitter [1] es una plataforma de *microblogging* que permite compartir noticias de forma inmediata y posee una considerable diversidad de fuentes para contrastar la información que le llega al usuario. Su funcionamiento consiste en un servicio de comunicación en el que cada usuario es capaz de publicar información de manera rápida y sencilla. Es por ello por lo que esta aplicación se ha convertido en una de las redes sociales más influyentes globalmente, posicionándose en quinta posición en cuanto al uso de redes sociales por parte de los españoles en el año 2024 y siendo la octava más usada en el mundo respecto al porcentaje diario de uso según el último informe digital de la empresa We are social (véase Figura 1-1), con una media de tuits diarios que ronda los 500 millones [2].

No obstante, esta rapidez en el intercambio de información puede dar lugar a la proliferación de las conocidas *fake news* o información no verificada. La falta de contexto en sus mensajes, conocidos como tuits, y el límite de 280 caracteres establecido<sup>1</sup> puede desencadenar un mal entendimiento o interpretación errónea, contribuyendo a la desinformación, lo que ya es una preocupación para las principales democracias en todo el mundo. Por otro lado, este formato de mensaje permite sintetizar la información y procesarla de una forma más rápida. Es por esto por lo que X/Twitter se ha convertido en una red muy popular entre celebridades, ya que les permite interactuar con sus seguidores de una forma única y a una velocidad vertiginosa.

El máximo potencial extraíble de esta aplicación se encuentra en el empleo como fuente abierta de información. La era tecnológica ha traído consigo una línea de transformación de las fuentes abiertas tradicionales, como los periódicos y las revistas, a las fuentes accesibles mediante el uso de internet. Dado el potencial de las redes sociales y su poder de difusión, cabe recalcar la importancia de diferentes cuentas creadas por los medios de comunicación para mantener una información actualizada y accesible, así como también facilitar la interactividad entre usuarios y medio de comunicación, lo que mejora la calidad de transmisión de información. Es por esto por lo que estas cuentas se han

---

<sup>1</sup> Este límite se aplica únicamente a los usuarios no suscritos al servicio de pago Twitter Blue, actualmente conocido como X premium.

convertido en herramientas esenciales para hacer llegar noticias de actualidad a un amplio público en cualquier lugar del mundo.

Un término utilizado para describir esta recopilación de información a través de fuentes abiertas es el de OSINT (*Open-Source Intelligence*). Este término se refiere a la recopilación y análisis de información que es pública. Esta información puede provenir de diversas fuentes como sitios web, redes sociales o noticias, entre otros. El Allied Joint Intelligence de la OTAN [3], en 2001, definía en la AJP-2.0.2001 las técnicas OSINT de la siguiente manera: “La inteligencia derivada de una amplia gama de recursos abiertos, como la radio, la televisión, los periódicos, los libros; a los que el público tiene acceso” [4].

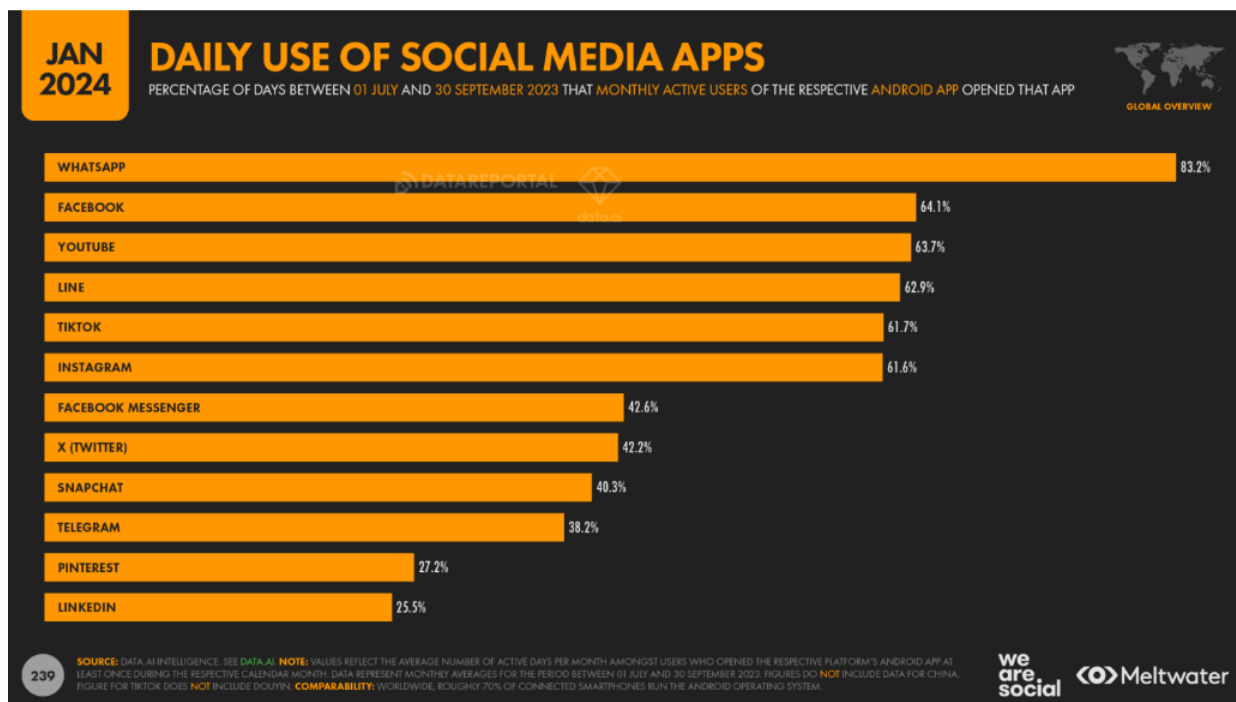


Figura 1-1 Gráfico representativo de las aplicaciones más usadas globalmente en 2024. (We are social)



Figura 1-2 Celebrities y medios de comunicación en la red social X/Twitter. (X/Twitter)

Por consiguiente, las redes sociales se pueden definir como una fuente de información para OSINT debido a la cantidad de información de la que disponen y su accesibilidad al público en general. La principal ventaja de OSINT es esta accesibilidad, así como su coste, ya que únicamente se requiere una conexión a internet y no implica el uso de recursos significativos.

El uso de OSINT se encuentra presente en el ámbito financiero, el tecnológico, el ámbito del periodismo o el de la defensa, entre otros. Este último será el ámbito en el que se empleará esta técnica a lo largo del proyecto, más en concreto en el contexto del reciente conflicto de Ucrania.

## 1.2 Objetivos y metodología

El principal objetivo de este trabajo fin de grado radica en el análisis y visualización de datos recabados de la red social X/Twitter sobre el reciente conflicto de Ucrania (véase Figura 1-3). Para ello, se propone la utilización de herramientas de código libre (como podrían ser Elasticsearch [5] y Kibana [6]) junto con código que ha sido desarrollado en el lenguaje de Python [7] que permitan procesar estos datos para su posterior análisis. Más en concreto, se propone desarrollar un código que permita indexar estos datos en Elasticsearch para su almacenamiento y posterior visualización y análisis a través de Kibana.

Para lograr este objetivo, se pretende ofrecer una interfaz de tipo panel de control (*dashboard*) donde se muestren algunos indicadores obtenidos del análisis de datos y se ofrezca la posibilidad de una consulta rápida sobre temas de actualidad y, más en concreto, del conflicto entre Rusia y Ucrania.

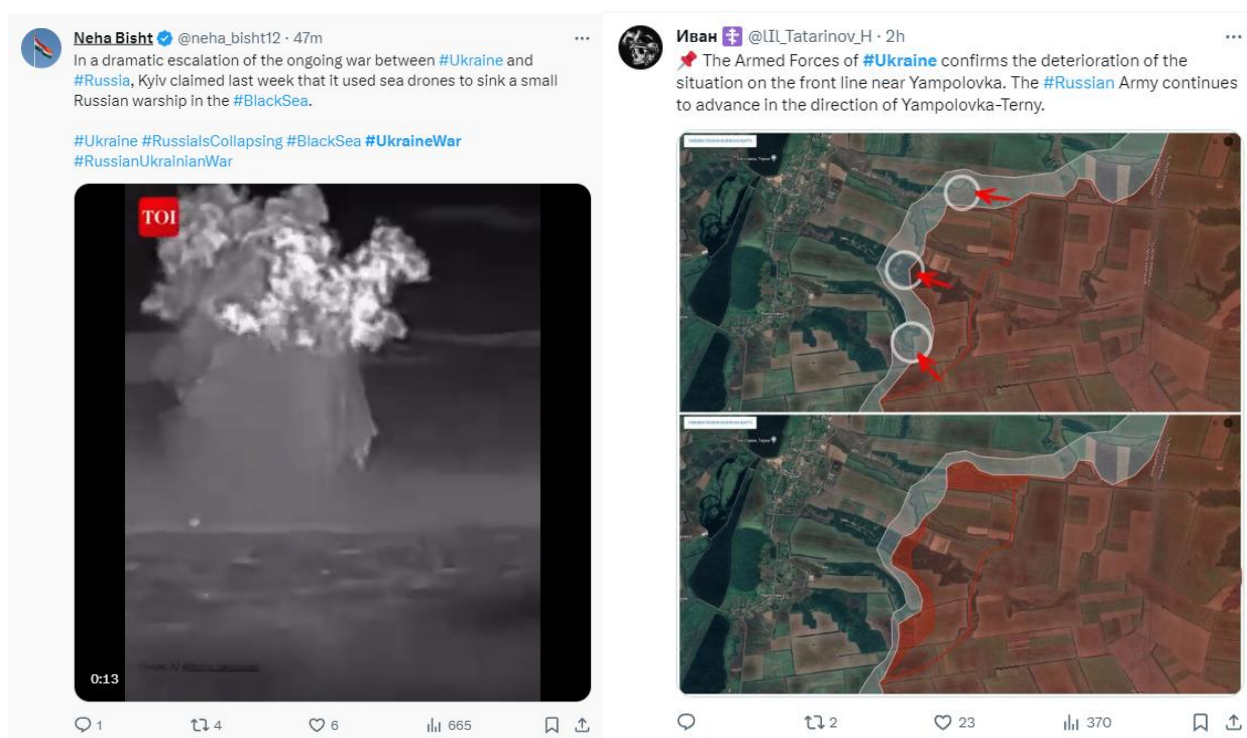


Figura 1-3 Información en tiempo real del conflicto Rusia-Ucrania a través de X/Twitter. (X/Twitter)

## 1.3 Estructura de la memoria

La estructura de este trabajo de fin de grado se divide en 5 capítulos principales, una bibliografía y una serie de anexos con información referente a las diferentes implicaciones éticas, económicas y sociales del trabajo, reflexiones sobre el mismo y por último una explicación de los códigos redactados en el lenguaje de programación Python.

- Capítulo 1: En este capítulo se contextualiza el trabajo y se exponen los objetivos principales del proyecto.
- Capítulo 2: En el segundo capítulo se procede a presentar detalladamente los antecedentes del trabajo. Se comienza comentando la relevancia de la difusión de información a través de las redes sociales, exponiendo así sus inicios y la creciente popularidad que este medio de comunicación ha ganado en los últimos años. A continuación, se detalla la evolución de la red social X/Twitter comentando su historia y su actual influencia en las noticias. También se

habla en este capítulo de OSINT y las ventajas que ofrece la inteligencia de fuentes abiertas. Una vez explicados estos antecedentes, se contextualiza la situación actual entre Rusia y Ucrania para ofrecer un conocimiento que permita comprender determinados detalles del análisis de los datos que se ofrecen en los capítulos posteriores. Finalmente, se explican las herramientas utilizadas para poder aplicar la metodología OSINT a los datos extraídos y se hace una reseña sobre trabajos académicos que se relacionan de alguna manera con el descrito en esta memoria.

- Capítulo 3: Se detallan los pasos previos a la realización del análisis y visualización de los ficheros de datos recogidos, así como el proceso de tratamiento de datos realizado. El capítulo finaliza con un análisis de los diferentes métodos de visualización que la herramienta seleccionada (en este caso, Kibana) proporciona.
- Capítulo 4: En este capítulo se presentan los resultados obtenidos de las diferentes visualizaciones de datos extraídos de la red social X/Twitter. Se explica también el proceso de elección del método de visualización idóneo para cada tipo de dato y se realiza un análisis de los resultados obtenidos en cada caso.
- Capítulo 5: En el último capítulo de la memoria, se presentan las conclusiones derivadas de la realización del trabajo y las reflexiones extraídas de los resultados obtenidos. Se presentan también una serie de posibles líneas futuras de investigación para conseguir una mejora con respecto al alcance de las fuentes de información abiertas con el objetivo de lograr un análisis más exhaustivo y preciso.

En último lugar, se ponen en manifiesto las fuentes de referencia del proyecto, así como los anexos previamente mencionados.

## 2 ESTADO DEL ARTE

### 2.1 La información en las redes sociales

#### 2.1.1 Inicios

Para comprender el inicio de las redes sociales es necesario hablar del impacto de la digitalización. La digitalización se ha llevado a cabo en dos fases sucesivas [8]. La primera fase estaba limitada en cuanto a la interacción del público con las páginas web, ya que estas eran mayoritariamente informativas y no se adaptaban en virtud de la opinión del público. En la segunda fase de esta digitalización, denominada Web 2.0, aparecen los sistemas de gestión de contenidos (*Content Management Systems*, CMS) para gestionar sitios web de medios de comunicación con una mayor sencillez. Posteriormente, aparecen también las redes sociales.

A principios del año 2002 aparecieron los primeros sitios web promocionando las redes de círculos de amigos en línea (Friendster), pero no fue hasta el año siguiente que las redes sociales se hicieron populares con la llegada de sitios tales como MySpace y Xing. La popularidad de estos sitios creció rápidamente incitando a que también muchas compañías abrieran un espacio para dar cabida a las redes sociales de internet [9].

A mediados de la década de los 2000, los medios de comunicación comenzaron a utilizar estas herramientas como medios de difusión de información. Esto permitió un auge en la forma en que las noticias se comparten hoy en día, permitiendo a estos medios llegar a audiencias más amplias. Una de las plataformas más utilizadas con este fin es X/Twitter, creada en 2006 como medio de mensajería breve, ya que facilita la difusión de noticias en tiempo real. Cabe destacar también la red social Facebook; creada 2 años antes que Twitter, esta plataforma tiene la función de generar comunidades virtuales en las que se puede compartir información. Por último, es necesario mencionar la irrupción de redes sociales de mensajería vía internet como WhatsApp, Telegram o Snapchat que permiten lograr una mayor interactividad entre usuarios [8].

Desde la creación de estas plataformas, el uso de redes sociales ha evolucionado hasta desempeñar un papel crucial en la forma en la que las noticias son compartidas en todo el mundo.

#### 2.1.2 Popularidad de las redes sociales

Las redes sociales se encuentran entre una de las herramientas más utilizadas por el público para informarse de noticias de actualidad en tiempo real. El 83,6% [10] de la población española estaba registrado en al menos una red social en enero del año 2024. En cuanto a las cifras mundiales, en 2024 se ha superado la marca de los 5.000 millones de usuarios en las redes sociales, con un crecimiento anual del 5,6% [11].

Según el estudio de 2024 de la agencia We are social [11], la primera motivación en el uso de internet es la búsqueda de información. En algunos casos, sobre todo entre los jóvenes, el usuario recurre más a estas plataformas para informarse de la actualidad que a las plataformas de noticias convencionales (véase Figura 2-1 y Figura 2-2).

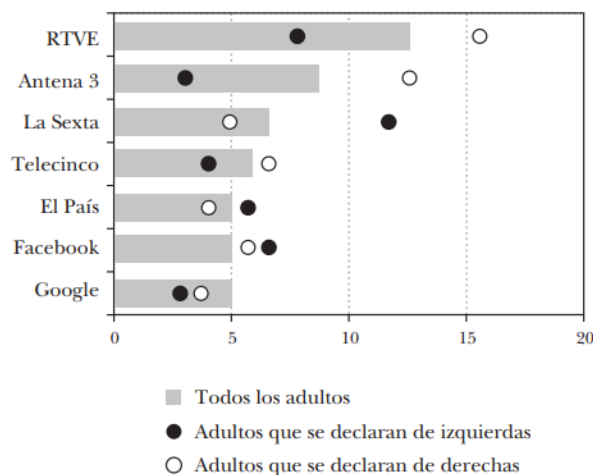


**Figura 2-1 Adultos que citan el medio como su principal fuente de noticias en porcentaje (Pew Research Center)**

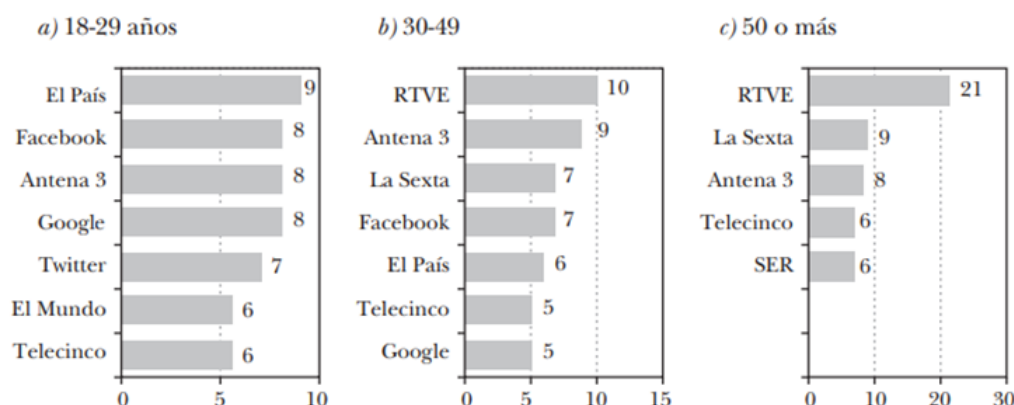


**Figura 2-2 Medio principal de fuente de noticias por grupo de edad (Pew Research Center)**

## 2.2 Twitter

### 2.2.1 Historia de X/Twitter

La historia de X/Twitter se remonta al año 2004, año en el que un programador estadounidense, Noah Glass, desarrolló un proyecto con el nombre de Odeo con el fin de dejar guardados en la nube mensajes grabados en formato MP3 al llamar por teléfono. Odeo fue evolucionando hasta convertirse en una plataforma de *podcast*<sup>2</sup> y se unió el creador de Pyra Labs, el software de creación de Eva Williams. Williams fue adquiriendo un papel más activo hasta convertirse en CEO de la compañía. Concurrentemente, Apple lanzó un reproductor multimedia llamado iTunes junto con un servicio de música y *podcast*; por consiguiente, Odeo se vio obligado a reinventarse para evitar el hundimiento. Es entonces cuando emerge la idea de usar mensajes cortos o SMS (*Short Message System*) para crear conversaciones entre grupos de personas [13].

<sup>2</sup> *Podcast*: anglicismo sin equivalente univocal en español que puede adaptarse al castellano mediante la aplicación de la tilde: pódcast. [12]

La reinención de Odeo dio paso a lo que hoy se conoce como la red social X/Twitter. Sus inicios se remontan al año 2006, con el nombre inicial de “Twtr” que pretendía imitar el trino de un pájaro. El primer tuit de la historia lo escribió Jack Dorsey el 21 de marzo de ese mismo año y decía “*Just setting up my twtr*” (véase Figura 2-3) [13].



**Figura 2-3 Primer tuit de la historia. (Marketing4ecommerce)**

El límite por aquel entonces de los tuits era de 140 caracteres ya que era el límite estándar que los operadores establecían para la mensajería SMS.

En julio de ese año, Jack Dorsey, Biz Stone y Evan Williams compraron los servicios de la plataforma a Odeo y crearon lo que hoy en día conocemos como Twitter Inc., para así explotar este servicio y conseguir capital procedente de inversores. En octubre del año 2006 se lanzó al mercado gracias a la obtención de 35 millones de dólares. Gracias a este salto, la red social adquirió popularidad, pasando de 20.000 textos al día a más de 60.000 [14].

Twitter se convierte en una empresa independiente en el año 2007 y continúa experimentando un notable incremento en el número de tuits diarios. Ese mismo año, un usuario llamado Chris Messina inventaría sin consciencia alguna el uso del *hashtag* al utilizar una almohadilla (#) para identificar un tema en sus tuits (véase Figura 2-4). Twitter vio esto con buenos ojos y lo implementó en la plataforma unos meses más tarde para hablar entre grupos de personas.

Twitter continuó experimentando un auge en 2009, con un aumento de nuevos usuarios del 1.300%, pasando de 5 millones a un total de 71 millones [13]. Este auge fue un éxito para la empresa, aunque trajo consigo dificultades para mantener el servicio activo y evitar colapsos debidos al elevado flujo de usuarios. Twitter por aquel entonces era un caos, ya que, a pesar de tener tal cantidad de usuarios, los beneficios económicos eran escasos.

Estas dificultades para dirigir la empresa provocaron el despido de Evan Williams y su sustitución por Jack Dorsey; Biz Stones abandonaba la compañía en 2011 debido a una notable falta de apoyos. Jack Dorsey se convertiría en CEO en 2015 y traería consigo cambios destacables que revolucionarían el uso de Twitter.

Entre los cambios que introdujo Jack Dorsey destacan el uso de los denominados *trending topic*, que es una forma de hacer más visibles los temas más populares del momento y permiten al usuario un sencillo acceso a información acerca de estos temas gracias al uso de las etiquetas (*hashtags*) que introdujo Chris Messina. También, se añadió la posibilidad de insertar en los tuits imágenes u otros archivos pertenecientes a otras redes sociales.



**Figura 2-4 Primer tuit que utiliza la herramienta *hashtag*. (Marketing4ecommerce)**

Un último cambio radical que introdujo el nuevo CEO fue el incremento hasta los 280 caracteres en los tuits y la eliminación del límite de caracteres en los mensajes directos. Esto supuso un gran avance en la aplicación, ya que el 90% de los usuarios que utilizaban esta plataforma como medio de comunicación redactaban exactamente 140 caracteres, indicativo de que este límite se quedaba escaso. Estos cambios implementados por Jack Dorsey reflejaron un crecimiento del uso diario del 12% durante el año 2017 [15].

A comienzos de abril de 2022, Elon Musk, uno de los mayores influyentes de X/Twitter, adquiere la red social por un precio de 44.000 millones de dólares, con el objetivo de poner fin a las censuras que se producían en la plataforma y fomentar la libertad de expresión, debatiendo, según él, de forma saludable sin recurrir a la violencia (véase Figura 2-5).

El nuevo propietario ha llevado a cabo una estrategia de *rebranding* (proceso de cambio y renovación de la identidad de una marca) con el despido de gran parte del equipo directivo y el cambio de su famoso icono de un pájaro, renombrando la plataforma con el nombre de “X”. Este cambio de nombre, como explica Elon Musk, se debe a que el término Twitter ya no reflejaba la diversidad de funciones que ofrece la plataforma. El nuevo nombre X simboliza la ambición de Musk de crear una aplicación “todo en uno” que integre comunicaciones y servicios financieros [16].

La adquisición de Twitter por parte de Musk ha traído consigo la llegada de Twitter Blue, actualmente denominado X Premium, que se basa en una suscripción para obtener o mantener la verificación de los usuarios en la aplicación y que trae consigo numerosos beneficios entre los que se destacan: un mayor límite de caracteres en cada tuit (4.000), la posibilidad de editar los tuits en un margen de una hora tras haberlos publicado y el más reciente que consiste en un modelo de reparto de ingresos publicitarios a los usuarios premium en función de su participación en la plataforma [9].



Figura 2-5 Objetivos de Elon Musk tras la adquisición de Twitter. (IEBS)

## 2.2.2 Terminología de X/Twitter

Para la comprensión del proyecto y su desarrollo cabe definir los siguientes términos utilizados [17]:

- **Tuitero:** Usuario registrado en la plataforma.
- **Tuit (del inglés *Tweet*):** Publicaciones que se realizan en el perfil social de X/Twitter. Para usuarios que no pagan la subscripción de Twitter Blue, estos tienen un número limitado de caracteres (280) sin incluir el material multimedia.
- **Tuitear:** Acción de publicar un tuit.
- **Retuit:** Acción de compartir un tuit previamente publicado por otro usuario.
- **Citar tuit:** Acción de hacer referencia a un tuit con un texto personalizado.
- **Explorar:** Lugar de la plataforma donde se muestran las tendencias del momento y las noticias de actualidad.
- **Listas:** Las listas permiten personalizar, organizar y priorizar los tuits que se muestran en la cronología.
- **Trending topic:** Tema que por el número de tuits relacionados se pone de moda y es de los más mencionados en la red en ese momento.
- **Impresiones:** Número de veces que se publica un tuit en la cronología o en los resultados de búsqueda.
- **Respuestas:** Número de veces que los usuarios responden a los tuits.
- **Tasa de interacción:** Número de interacciones (todo tipo de acción que se realice sobre un tuit) dividido entre el número de seguidores.
- **Seguidor (del inglés *follower*):** Es el usuario que sigue una cuenta y que, por tanto, puede ver todos los tuits que publica.
- **Siguiendo (del inglés *following*):** La acción de seguir a otro usuario.
- **Etiqueta (Hashtag):** Permite facilitar la búsqueda en un ámbito concreto con ayuda un término precedido por una almohadilla (#).
- **Mención:** Texto de un tuit que contiene el nombre de usuario de otra persona precedido por un @ para hacerle referencia.
- **Cronología (Timeline):** Orden cronológico que sigue la plataforma para organizar los tuits publicados por las cuentas a las que el usuario sigue.
- **Me gusta (Like):** Función que permite a los diferentes usuarios indicar que le ha gustado una publicación. Queda representado por un corazón.

En las Figura 2-6 y Figura 2-7 se muestran de manera práctica algunos ejemplos.



Figura 2-6 Cuenta X/Twitter de la Policía Nacional. (Autoría propia)



Figura 2-7 Retuit de la cuenta de Armada. (Autoría propia)

### 2.2.3 Influencia en las noticias

A la luz de los datos previamente expuestos, resulta evidente que X/Twitter ejerce una influencia considerable en la difusión y el consumo de noticias en el panorama actual. Esta influencia hace que dicha red social posea un gran potencial para moldear la agenda mediática, impactando en diferentes medios de comunicación. Esta influencia se manifiesta en diversas formas:

- En primer lugar, la popularidad en la red social se ha erigido como un criterio relevante para determinar la noticiabilidad de un tema de interés actual. Del mismo modo que se consideran aspectos como la novedad, el conflicto o el interés del protagonista al decidir sobre la inclusión de un tema, se examina si dicho tema está generando debate en X/Twitter. Si el debate generado es *trending topic* (tendencia), aumenta la probabilidad de que los medios le dediquen atención.
- Continuando con su influencia, los debates abiertos en X/Twitter son incorporados a menudo en la difusión de contenidos. Es por ello que la red social influye en la construcción de información, ya que lo que se discute en la plataforma puede llegar a convertirse en noticia de actualidad. De este modo, los usuarios ejercen un impacto significativo en la agenda mediática, recurriendo a X/Twitter como medio de difusión de información en lugar de recurrir a los medios de comunicación tradicionales, que da lugar a que los medios publiquen aquello que sea de interés en las redes.
- La ventaja que posee X/Twitter sobre los medios de comunicación tradicionales recae en el contacto entre periodistas y usuarios que la plataforma permite, ofreciéndoles así a estos periodistas nuevas perspectivas. Esta influencia sobre los periodistas explica la relevancia que adquieren las celebridades o figuras públicas, desatando, en algunos casos, polémicas que tienen repercusión en la agenda mediática. En definitiva, X/Twitter se ha convertido en una forma de noticias que favorece la interactividad entre usuarios y medios de información y que, a su vez, como se observa en la Figura 2-2, va ganando progresivamente más popularidad.

## 2.3 OSINT

### 2.3.1 Evolución de OSINT

El término OSINT se refiere al proceso de recopilar y analizar información de fuentes públicas para obtener inteligencia sobre un tema específico. Tiene su origen en la Segunda Guerra Mundial con el fin de monitorizar las conversaciones radio enemigas para obtener inteligencia; este invento es

considerado un avance imprescindible que permitió recopilar información clave para la victoria de los Aliados [18].

La disciplina de OSINT ha sido utilizada para la recopilación de información en numerosos conflictos posteriores a la Segunda Guerra Mundial como la Guerra Fría, la guerra del Golfo, la de Afganistán y el actual conflicto entre Rusia y Ucrania. En la actualidad, OSINT ha experimentado un crecimiento exponencial gracias a las redes sociales, las cuales se han constituido en el escenario ideal para la proliferación de esta forma de obtención de inteligencia [19].

El auge de OSINT se ha visto reflejado en el actual conflicto ruso-ucraniano, ya que se trata de la primera contienda en un territorio geográfico donde la conexión de red es generalizada y existe una fuerte presencia de documentación gráfica por parte de periodistas y de los propios combatientes que retransmiten imágenes y vídeos que sirven de testimonio de lo que ocurre en el campo de batalla (véase Figura 2-8). Esta documentación, retransmitida a través de las redes sociales, ha permitido a un público extenso seguir el conflicto con información de primera mano prácticamente en tiempo real.



**Figura 2-8 Documentación gráfica sobre eventos ocurridos en Ucrania. (X/Twitter)**

### 2.3.2 Proceso de OSINT

El proceso de recolección de información que sigue OSINT consta de diferentes fases (véase Figura 2-9) debido a la gran cantidad de fuentes de información abiertas de las que se pueden extraer los datos para así realizar una investigación exhaustiva. Entre estas fases se incluyen las siguientes [20]:

1. Establecer necesidades y fijar requisitos.
2. Seleccionar las fuentes de información más fiables.
3. Explotación de las fuentes de información.
4. Procesamiento y análisis de la información recogida.
5. Presentación de la investigación realizada.

El desarrollo de este trabajo fin de grado se relaciona con la quinta etapa de este proceso: la presentación de la información previamente adquirida en otras etapas, lo que permite al usuario acceder a la información de una manera más visual que facilite la comprensión de la misma.

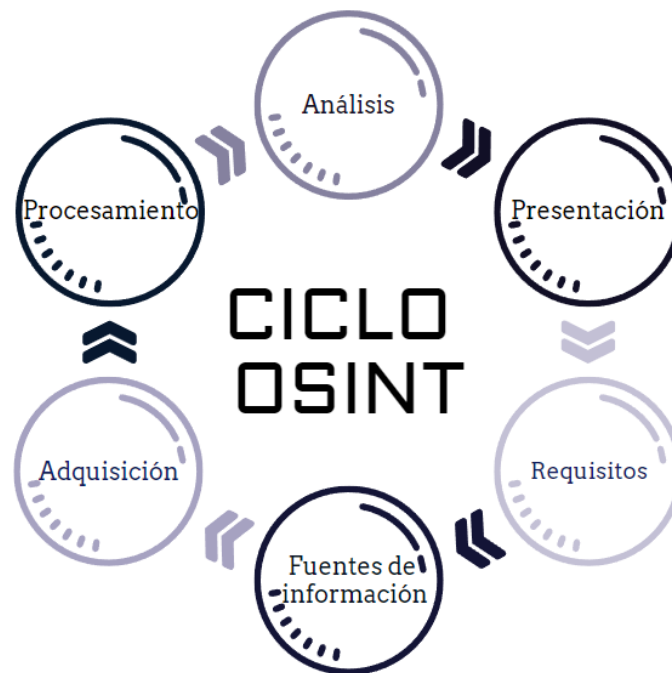


Figura 2-9 Ciclo de OSINT. (Autoría propia)

## 2.4 Conflicto entre Rusia y Ucrania

### 2.4.1 Antecedentes

Los antecedentes del conflicto entre Rusia y Ucrania se remontan al fin de la Guerra Fría y las negociaciones sobre la reunificación de Alemania bajo la Alianza del Atlántico Norte. Durante estas negociaciones, Rusia defiende que el por entonces secretario de Estado de los EE. UU., James Baker, prometió a Mikhail Gorbachev que la OTAN no se expandiría hacia el este de Europa en caso de que Rusia aceptara la unificación de Alemania [18], aunque no existe ningún registro escrito de dicha promesa. Un año más tarde, tras la disolución del Pacto de Varsovia en febrero de 1991, la URSS perdía el apoyo de los países miembros (Albania, Bulgaria, Checoslovaquia, Alemania del Este, Hungría, Polonia y Rumanía) [19] y crecía el miedo de una posible adhesión a la OTAN (véase Figura 2-10). Con la llegada de Clinton a la presidencia de EE. UU., se impulsó la adhesión de los países del este de Europa a la OTAN a lo que el por entonces presidente ruso, Boris Yeltsin, expresó que se trataba de una brecha en el tratado de 1990. Desde ese año, la OTAN se ha ampliado hasta los 32 miembros que componen la Alianza hoy en día (véase Figura 2-11). Es por ello que Putin defiende la invasión de Ucrania debido a la “amenaza” que supone el aumento de los países del este anexionados a la OTAN, muchos de ellos parte del antiguo Pacto de Varsovia.



**Figura 2-10 Situación geopolítica durante la Guerra Fría. (Britannica)**

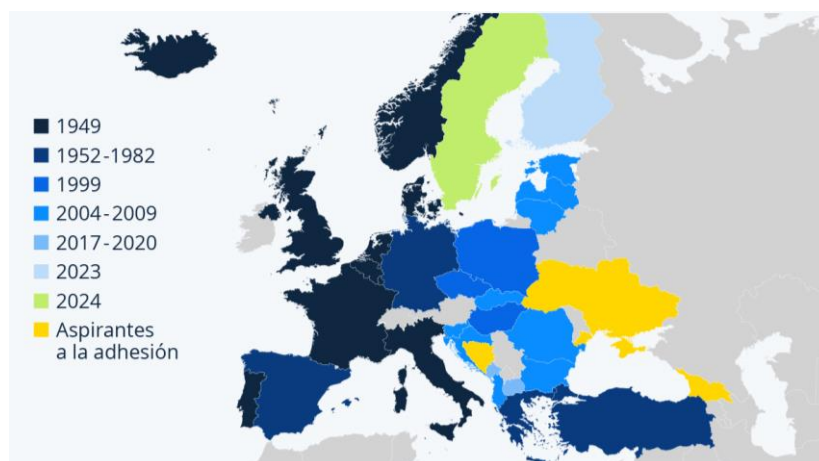


**Figura 2-11 Situación geopolítica de la OTAN. (NATO)**

Es esencialmente necesario comprender la historia entre Ucrania y Rusia, ya que comparten un mismo legado cultural y étnico muy fuerte. Es por ello que el país se encuentra dividido entre los que amparan la independencia ucraniana de la URSS y los que desean mantener esta relación histórica y cultural que une a ambas naciones. Esta división de opiniones se refleja desde que Ucrania se independizó formalmente en 1991, presentando claramente una demografía que reflejaba la opinión de la población prooccidental y la población prorrusa (véase Figura 2-12).

En 2010, este conflicto entre la población ucraniana se incrementó debido a la diferencia de ideología entre los dos candidatos a la presidencia del gobierno: Víktor Yanukovich, candidato prorruso y Yulia Tymoshenko, a favor de la UE y OTAN [21]. El candidato prorruso ganaría por aquel entonces las elecciones, pero sería derrocado a través de un golpe de estado en 2013 [21]; desde esa fecha el gobierno de Ucrania ha estado en manos de la ideología a favor de occidente. El derrocamiento de Yanukovich desencadenó un movimiento de rebelión contra el nuevo gobierno ucraniano, concentrándose mayoritariamente en las regiones del Donbás y Crimea; este movimiento fue apoyado por la Federación Rusa y provocaría la declaración de independencia de Crimea.

Analizando la relevancia de los factores políticos y culturales, se puede apreciar que Ucrania es un país fuertemente dividido entre aquellos a favor de la Europa occidental y quienes desean una relación más estrecha con Rusia.





Figura 2-13 Territorio anexionado por Rusia. (El Mundo)

### 2.4.3 Papel de la OTAN

La OTAN adquiere un beneficio estratégico con respecto a Rusia al posicionar a Ucrania en el bloque occidental, es por ello que la expansión de la Alianza hacia el este constituye el factor más importante de la guerra ruso-ucraniana. La OTAN comenzó a apoyar a Ucrania a raíz de la invasión rusa de la península de Crimea, proporcionando a Ucrania entrenamiento y material militar al igual que aumentando su presencia como medio de disuasión frente a la agresión rusa. Es cierto que Ucrania no es un país miembro de la Alianza, pero su compromiso con la OTAN afianza su poder militar y posiciona a los países occidentales en una manifiesta ventaja estratégica respecto a Rusia. En la Tabla 1 se resume la situación política de Ucrania y se presentan sus relaciones actuales con Rusia y con la OTAN.

|       | Historia y cultura                                            | Política                                                                         | Economía                                                                | Influencia                                                                |
|-------|---------------------------------------------------------------|----------------------------------------------------------------------------------|-------------------------------------------------------------------------|---------------------------------------------------------------------------|
| OTAN  | Actual interés en anexión                                     | Apoyo a la soberanía de Ucrania                                                  | Asistencia económica y apoyo humanitario y militar                      | Aumento de presencia como parte de estrategia de disuasión frente a Rusia |
| RUSIA | Vínculos culturales, lingüísticos y étnicos (URSS hasta 1991) | Conflictos territoriales debidos a sentimientos prorrusos en regiones ucranianas | Dependencia ucraniana en el suministro de energía y disputas económicas | Deseo de mantener su influencia política y económica                      |

Tabla 1 Relaciones de la OTAN y Rusia con Ucrania. (Autoría propia)

### 2.4.4 Consecuencias del conflicto

El conflicto entre Rusia y Ucrania ha tenido una repercusión mundial y ha traído consigo consecuencias significativas en varios niveles, tanto para las dos naciones involucradas como para la sociedad en su conjunto:

- En el aspecto económico, la guerra ha generado un aumento de la inflación de los precios sobre todo en el sector energético debido a que Rusia es el tercer exportador de petróleo a nivel mundial y uno de los suministradores principales de gas natural a Europa, generando el 50% del gas que se consume en el continente [23].
- En cuanto a la alimentación, habiendo señalado la importante producción de grano en el apartado 2.4.2, la inestabilidad política de la zona ha afectado a la producción y exportación de cereales y otras materias primas, provocando una subida de los precios. La situación de incertidumbre ha llevado a los compradores a buscar fuentes alternativas de suministro, lo que ha aumentado la demanda y el precio de estas materias primas [24].
- La última consecuencia por destacar del conflicto es la influencia que ha tenido en el gasto militar de cada país, sufriendo el mayor incremento en Europa desde la Guerra Fría. La guerra ha traído una mayor conciencia de la importancia de las Fuerzas Armadas, favoreciendo enormemente las políticas y los presupuestos destinados a la seguridad y la defensa de cada país. Un claro ejemplo son las cifras del presupuesto del Ministerio de Defensa español de 2023, con el objetivo de potenciar la transformación y modernización de las capacidades de las Fuerzas Armadas; en la Tabla 2 se muestra una comparativa de los presupuestos de defensa de España en los últimos años [25].

|      | Presupuesto defensa<br>(miles de €) | % Respecto PIB | % Variación año anterior |
|------|-------------------------------------|----------------|--------------------------|
| 2020 | 8.994.337,98                        | 0,8%           | 1,5%                     |
| 2021 | 9.411.931,94                        | 0,78%          | 4,6%                     |
| 2022 | 10.155.269,16                       | 0,79%          | 7,9%                     |
| 2023 | 12.827.177,55                       | 0,95%          | 26,3%                    |

**Tabla 2 Comparativa presupuestos del Ministerio de Defensa. (Autoría propia)**

## 2.5 Herramientas utilizadas para el análisis de datos

### 2.5.1 Elasticsearch

Elasticsearch es una herramienta de búsqueda y análisis de código abierto basada en la biblioteca Apache Lucene [25]. Esta plataforma fue creada en 2010 para incorporar la posibilidad de realizar búsquedas de texto completo a documentos con formato JSON<sup>3</sup>.

El funcionamiento de esta herramienta se basa en la recopilación y almacenamiento de datos de diferentes fuentes para su posterior explotación por parte del usuario. Mientras se registra la información, Elasticsearch crea un índice invertido para poder buscar información en tiempo real.

Elasticsearch ofrece una solución de búsqueda de texto con múltiples funciones como:

- Escalabilidad horizontal: posibilidad de distribuir los datos en diferentes nodos, facilitando la gestión de un gran volumen de información y un mayor rendimiento.
- Búsqueda e índices multifuncionales: proporcionan al usuario la capacidad de consultas precisas y eficientes a la vez que se admiten operaciones como la indexación o el análisis.
- Compatibilidad con lenguajes de programación: en el caso de este trabajo fin de grado se utilizará el lenguaje de programación Python 3.

<sup>3</sup> JSON: Del inglés “JavaScript Object Notation”

- Soporte de datos no estructurados: es capaz de manejar datos como documentos JSON o registros de eventos.
- Soporte a la integración y complementos: se integra fácilmente con otros componentes como Kibana para la visualización y análisis de información.

### 2.5.2 Kibana

Kibana es una herramienta que complementa a Elasticsearch, permitiendo realizar un análisis gráfico de los datos que se introducen en la plataforma, así como una consulta de los datos de Elasticsearch de una forma más visual [26].

Entre las posibilidades que nos ofrece Kibana se encuentran [27]:

- Dar forma a nuestros datos: los datos almacenados pueden ser representados de manera gráfica para una comprensión más sencilla por parte del lector.
- Crear objetos de los índices: para poder trabajar sobre ellos y aplicarles la analítica de Kibana, así como diferentes operaciones de aprendizaje automático (*machine learning*).
- Crear visualizaciones de los datos a través de su interfaz gráfica.
- Elaborar *dashboards*, *canvas*, *maps* y *reports* a un nivel profesional con las visualizaciones que previamente se creen (véase Figura 2-14).

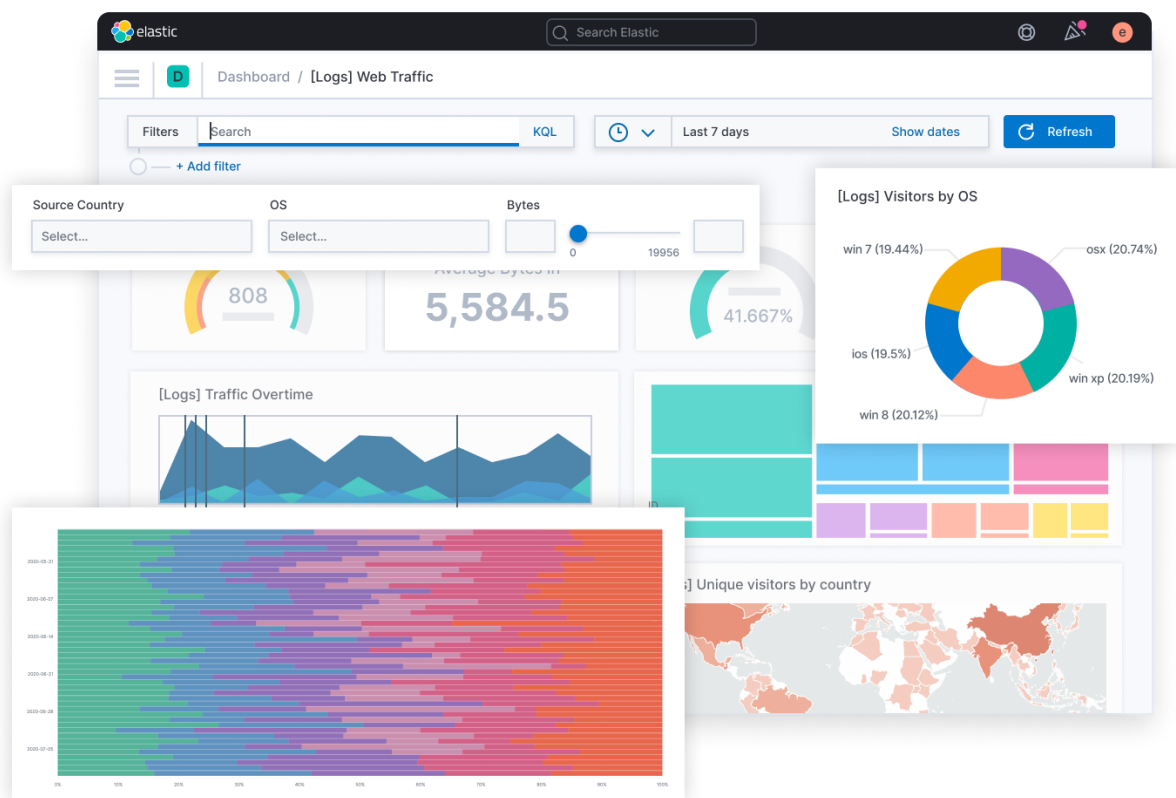


Figura 2-14 Dashboard de Kibana. (Autoría propia)

## 2.6 Alternativas a Kibana

La visualización de datos es una tarea vinculada al análisis de datos cuyo objetivo es la representación gráfica de la información de interés [28]. El análisis que proporciona la visualización juega un papel fundamental en la difusión de una información veraz, realizando un estudio de los temas de actualidad para así explorar, comprender y comunicar la información obtenida de una manera efectiva. Existen diferentes herramientas que, como Kibana, se pueden emplear para la realización de este análisis.

La visualización de datos tiene como objetivo principal la comunicación de patrones, tendencias o relaciones gracias a la representación de datos mediante gráficos, diagramas y otros elementos visuales. Estas visualizaciones pueden variar en formato y complejidad, es por ello por lo que existen diferentes alternativas a Kibana en cuanto a la elección de una herramienta que nos permita el desarrollo de este trabajo fin de grado. Algunos ejemplos de diferentes herramientas son:

- **Grafana:** Grafana [29] es potencialmente el mayor competidor de Kibana ya que ofrece una amplia variedad de fuentes de datos y una interfaz altamente personalizable por el usuario. Por otro lado, su curva de aprendizaje es más pronunciada, lo que ralentizaría el desarrollo del proyecto, y no se centra tanto en el análisis de registros.
- **Tableau Desktop:** Tableau [30] es otra herramienta de visualización de datos que ofrece una interfaz intuitiva y facilita la creación de visualizaciones complejas. Por el contrario, es necesario adquirir una suscripción de pago para obtener la licencia de uso y no es ideal para grandes volúmenes de datos.
- **Power BI:** Esta herramienta [31] es un producto de Microsoft y ofrece capacidades similares a Tableau, aunque es capaz de manejar un mayor volumen de datos. Su principal característica es la posibilidad de integración con otras herramientas de Microsoft como Excel, pero la licencia gratuita es limitada.
- **Google Data Studio:** Esta herramienta [32] es gratuita y permite la integración con otras fuentes de Google, como Google Analytics y Google Sheets, pero tiene limitaciones considerables en términos de personalización en comparación con otras herramientas.

Finalmente, tras un estudio de la herramienta de visualización de datos más indicada, se ha llegado a la conclusión de que Kibana es la más adecuada para el desarrollo de este trabajo, ya que ofrece una integración estrecha con Elasticsearch, un enfoque centrado en el tipo de datos con el que se va a trabajar y una capacidad de personalización para representar los datos obtenidos, al igual que una amplia variedad de opciones de visualización. En la Tabla 3 se expone una comparativa, a modo de resumen, de las características de estas herramientas de visualización.

|                              | <b>Kibana</b>                                    | <b>Grafana</b>                                            | <b>Tableau Desktop</b>                                | <b>Power Bi</b>           | <b>Google Data Studio</b>                                                   |
|------------------------------|--------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------|---------------------------|-----------------------------------------------------------------------------|
| <b>Fuente de datos [33]</b>  | Elasticsearch<br>Search ELK<br>Logstash<br>Beats | Prometheus<br>Graphite<br>InfluxDB<br>Elasticsearch       | Archivos Excel<br>Archivos CSV<br>Datos SQL           | Archivos Excel            | Google Marketing<br>Google AdWords<br>Google Cloud Storage<br>Google Sheets |
| <b>Integración [34] [35]</b> | Integración nativa con Elasticsearch             | Microsoft Dynamics<br>ServiceNow<br>Azure<br>Zendesk Sell | Salesforce,<br>ServiceNow<br>Amazon<br>Redshift Asana | Aplicaciones de Microsoft | Productos de Google                                                         |
| <b>Licencia</b>              | Fuente abierta o comercial                       | Fuente abierta o comercial                                | Comercial                                             | Comercial                 | Gratuita                                                                    |

**Tabla 3 Comparativa de las diferentes herramientas de visualización. (Autoría propia)**

## 2.7 Trabajos académicos relacionados

### 2.7.1 *Herramientas de OSINT para la obtención de información en Twitter sobre el conflicto Ucrania-Rusia*

Trabajo Fin de Grado (TFG) [36] realizado por Álvaro Hernández Ruiz (2022) en el Centro Universitario de la Defensa en Escuela Naval Militar [37]. En este trabajo se realiza un estudio de las diferentes herramientas OSINT y se procede a analizar tuits publicados acerca del conflicto entre Rusia y Ucrania mediante la herramienta T-hoarder. Posteriormente, se realiza un análisis sobre el uso de OSINT para la obtención de información de actualidad, detallando los usuarios que poseen una mayor influencia mediante el número de retuits, menciones, etc.

### 2.7.2 *Sistema de detección inteligente de usuarios de Twitter interesados en la temática de defensa*

Trabajo Fin de Grado [38] realizado por María Fernández Torrejón (2023) en el Centro Universitario de la Defensa en Escuela Naval Militar. En este trabajo se desarrolla un sistema de detección de usuarios de X/Twitter que publiquen contenidos de interés para el ámbito de defensa, para lo que se entrena un clasificador capaz de discernir si la información de entrada es o no de interés. Este clasificador es, a su vez, comparado con la herramienta de modelo de lenguaje ChatGPT [39].

### 2.7.3 *Plataforma OSINT para el almacenamiento y análisis inteligente de datos de la red social Twitter*

Este Trabajo Fin de Grado [40], realizado por Javier Jiménez-Cervantes Pérez (2023) en el Centro Universitario de la Defensa en Escuela Naval Militar desarrolla una plataforma OSINT para la recopilación de información de la red social X/Twitter, creando ficheros JSON que constituyen una base de datos para un futuro procesamiento y presentación de la información recolectada.

### 2.7.4 *Análisis de las redes sociales: La guerra de Ucrania*

Realizado por Sara Pedregosa Gómez (2022) en la Facultad de Comunicación de la Universidad de Sevilla [41]. Este Trabajo Fin de Grado [42] analiza la función de las redes sociales en el conflicto de Ucrania para transmitir información y el impacto que esta transmisión genera en la población durante un conflicto bélico. Este trabajo se cumplimenta con una entrevista a una mujer inmigrante procedente de Ucrania, Svitlana Zarichna, que mantiene el contacto con sus familiares presentes en la zona de guerra.

### 2.7.5 *Análisis y visualización de datos abiertos de carácter informativo para el alumnado de la Universidad de Valladolid*

Realizado por Nadia Pérez Faúndez (2021) en la Facultad de Telecomunicación de la Universidad de Valladolid [43]. Este Trabajo Fin de Grado desarrolla una plataforma OSINT para la recopilación de información de potencial utilidad para estudiantes de dicha universidad. Para ello, se recolectan datos de la base de datos de la Junta de Castilla y León y se procesan con la herramienta de análisis Tableau. Ciertamente es que la herramienta de análisis, la fuente de datos y el escenario en este proyecto son diferentes ya que el trabajo está centrado en un escenario lúdico y la fuente de datos utilizada es la interactividad de los estudiantes dentro del propio campus virtual, pero el proceso de recopilación de datos y visualización se asemeja al de este Trabajo Fin de Grado [44].

### 2.7.6 *Investigación y extracción de datos con técnicas OSINT*

Trabajo Fin de Grado realizado por Irene Perrino Navas (2021) en la Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación [45], el cual está centrado en la extracción de información presente en diferentes redes sociales. En el desarrollo de dicho trabajo [46], se realiza la extracción de información de un usuario ficticio, personificado por la misma autora del trabajo, en diferentes redes

sociales entre las que destaca X/Twitter. El trabajo se basa un escenario de aplicación centrado en la actividad de dicho usuario y su información personal obtenida a través de sus perfiles en diferentes redes sociales. Posteriormente, dicha información se almacena en una base de datos para poder ofrecer una visualización de esta.

## 3 DESARROLLO DEL TFG

### 3.1 Instalación de la máquina virtual

Para la realización de este trabajo será necesaria la instalación de un sistema operativo Linux, con el objetivo de crear un entorno que permita, de manera sencilla, programar los diferentes archivos de código necesarios para el análisis y visualización de datos acerca del conflicto de Ucrania. Para evitar la instalación de otro sistema operativo en el disco físico y utilizar un arranque dual del ordenador, se ha optado por la instalación de una máquina virtual Linux.

La instalación de esta máquina virtual se realiza desde la página oficial de VirtualBox [47]. Se procede a la instalación de la última versión (7.0.14) de Windows hosts, ya que esta debe concordar con el sistema operativo del ordenador en el que se realiza el proyecto. También será necesaria la instalación de un archivo de disco Ubuntu (versión ubuntu-23.10.1) [48] para configurar la máquina virtual.

Una vez se haya instalado todos los ítems necesarios, se procederá a crear una nueva máquina virtual de la siguiente manera:

1. El primer paso para crear el ambiente de trabajo en el que realizaremos el código será seleccionar el icono de “Nueva”. A continuación, se nos abrirá un menú donde se escribirá el nombre de la máquina. Es necesario seleccionar la imagen ISO que hayamos descargado previamente (véase Figura 3-1).

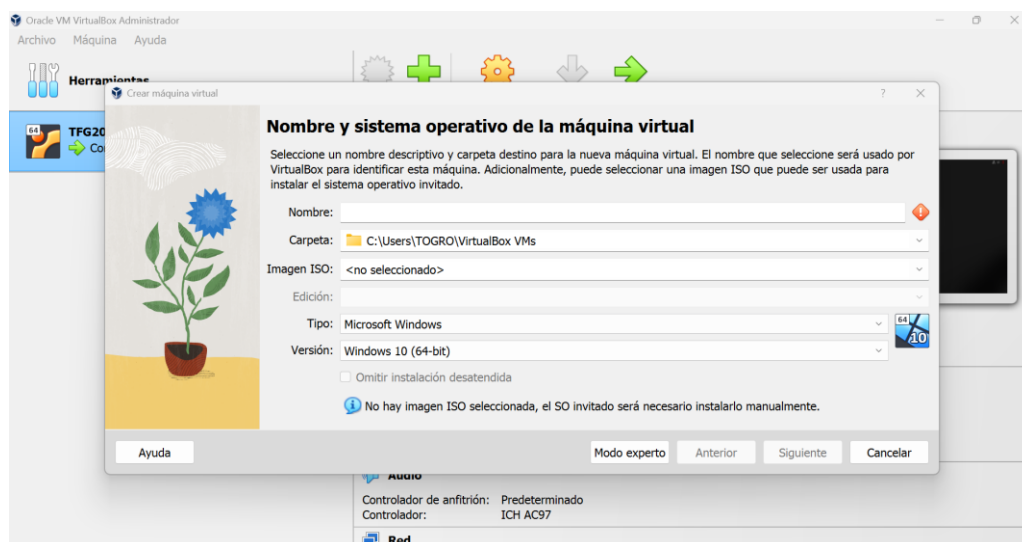


Figura 3-1 Creación de la máquina virtual. (Autoría propia)

- Una vez creada la máquina, será necesario configurar los parámetros de memoria base, número de procesadores y tamaño de disco mediante las ventanas que vayan saltando en pantalla según los requisitos del trabajo. Los parámetros establecidos en la máquina virtual creada para este trabajo se pueden observar en la Figura 3-2.

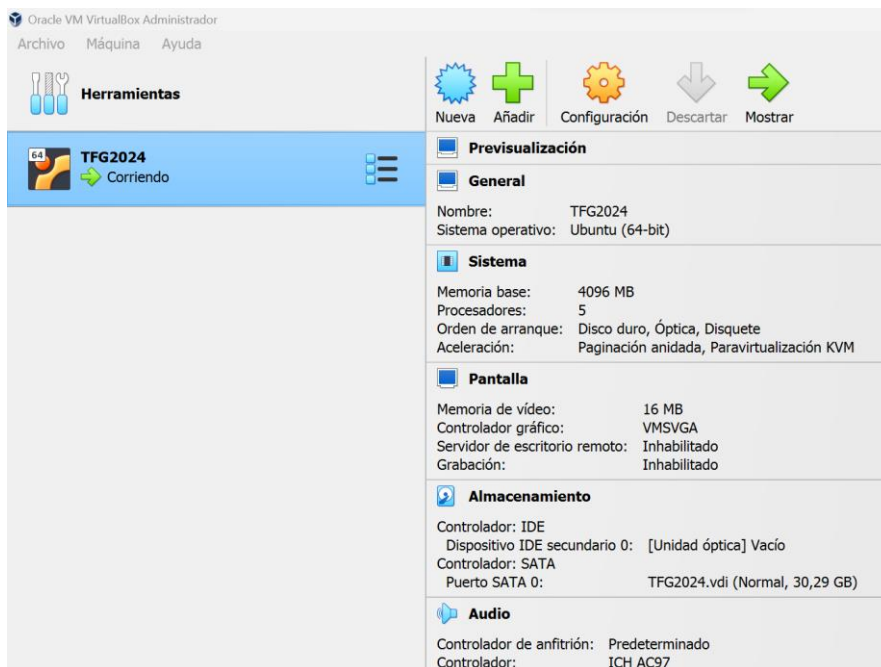


**Figura 3-2 Parámetros de la máquina virtual. (Autoría propia)**

- Finalmente, se ha de seleccionar la opción de “Terminar” en las ventanas emergentes; acto seguido, aparecerá una interfaz donde finaliza la creación y configuración de la máquina virtual creada con la imagen ISO que previamente se haya seleccionado.

## 3.2 Instalación herramientas a utilizar

### 3.2.1 Python

Para la realización de este trabajo fin de grado, así como la posibilidad de trabajar con las diferentes herramientas requeridas para su desarrollo, será necesaria la instalación de Python. Se ha optado por su versión 3.11.6, ya que es una opción estable y, al ser la última versión de la serie 3.11 en el momento de instalación, cuenta con mejoras de rendimiento y correcciones de errores con respecto a las versiones anteriores de la serie. Este lenguaje de programación será empleado para realizar diferentes archivos de código que, al ser ejecutados, permitirán procesar información relativa al conflicto de Ucrania. Esta instalación se llevará a cabo directamente desde el terminal con el comando que se observa en la Figura 3-3. Una vez instalado Python, se comprueba en el terminal que la instalación haya sido correcta (véase Figura 3-4).

Como se menciona en el apartado 2.5, Elasticsearch trabaja con un formato de datos denominado JSON. Ahora bien, Python cuenta con su propia librería JSON para procesar este tipo de datos, la cual viene ya incluida en la instalación, sin embargo, se ha optado por la instalación de una biblioteca externa denominada “simplejson” para el tratamiento de estos datos. “Simplejson” [49] es una biblioteca externa de Python que ofrece una implementación más rápida, sencilla y eficiente de datos JSON en comparación con la librería estándar ya incorporada. Esta librería es externa, por lo que es necesario instalarla por separado (véase Figura 3-5) de la misma manera que se ha instalado Python en el terminal.

```
jaimelpzg@jaimelpzg-1-2:~$ sudo apt install python3
```

Figura 3-3 Comando para instalar Python 3.11.6. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ python3 --version
Python 3.11.6
```

Figura 3-4 Comprobación de la versión de Python instalada. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~$ sudo apt install python3-simplejson
```

Figura 3-5 Comando para instalar la librería simplejson en Python. (Autoría propia)

Juntamente con el lenguaje de programación, será necesaria la instalación de un editor de texto. En este caso se ha seleccionado *PyCharm* [50]. Este editor es un IDE<sup>4</sup> creado por JetBrains en 2010 en respuesta a la creciente demanda de una herramienta eficiente para programar en Python. Este editor ha experimentado una adopción entre un gran número de desarrolladores gracias a su amplia gama de características como: edición de código avanzada, integración con herramientas de control de versiones, depuración interactiva, soporte para *frameworks* populares y análisis de código estático. Es por ello por lo que PyCharm se ha consolidado como un IDE líder en el desarrollo de Python tanto para desarrolladores como principiantes en codificación.

La elección de PyCharm frente a otros editores de texto recae en sus características de edición de código y su fluido flujo de trabajo. Así como la posibilidad de crear y gestionar entornos virtuales y librerías de Python, lo que simplifica la configuración del proyecto. Estas características permiten a los desarrolladores que utilizan PyCharm redactar código de manera más eficiente, y con una reducción de errores en comparación con otros editores de texto básicos.

Este editor de texto no da pie a la posibilidad de instalarlo directamente desde el terminal, es por ello que será necesario acceder a la página web de JetBrains [51] donde se deberá elegir la versión de PyCharm Community Edition para Linux. Una ventaja que proporciona Linux frente a Windows es la opción de instalar el editor de texto PyCharm directamente desde una “*app store*” denominada centro de aplicaciones. Dentro de esta ventana se encuentra disponible el ejecutable del editor de texto PyCharm, siendo posible descargar la versión más reciente (véase Figura 3-6).

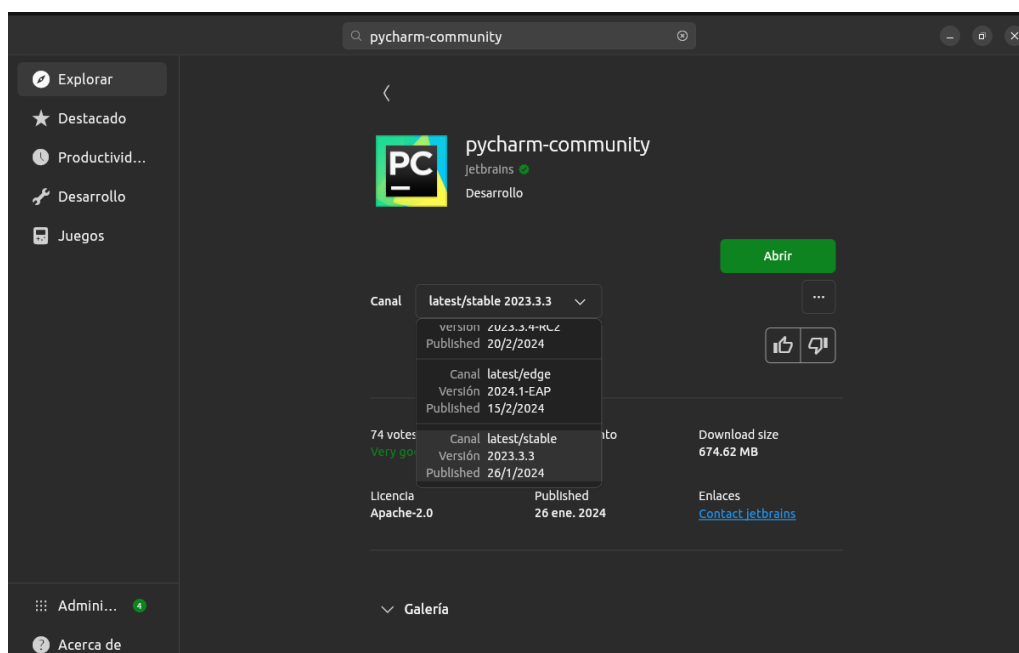


Figura 3-6 Instalación de PyCharm. (Autoría propia)

<sup>4</sup> IDE: Del inglés “*Integrated Development Environment*” o, en español, Entorno de Desarrollo Integrado.

### 3.2.2 Elasticsearch

La versión de esta herramienta ha de ser compatible con la versión de Linux instalada en la máquina virtual. Es por ello por lo que, en lugar de la versión más reciente, se descargará la versión Elasticsearch 7.17.6. Para la instalación de la herramienta, será necesario acceder a la web de Elasticsearch y descargar la versión deseada; esta acción se puede realizar directamente desde el terminal (véase Figura 3-7). La máquina virtual descarga la carpeta de Elasticsearch en versión comprimida, por lo que será necesario descomprimirla con el comando que se observa en la Figura 3-8. Una vez descomprimida, se deberá acceder a la carpeta “bin” dentro de “Elasticsearch-7.17.6” e iniciar el ejecutable de Elasticsearch (./elasticsearch). A continuación, se comprueba la correcta instalación y funcionamiento de dicha herramienta. Para ello, se accederá al servidor escribiendo en el navegador la URL “http://localhost:9200” (véase Figura 3-9).

```
jaimelpzg@jaimelpzg-1-2:~$ wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.17.6-linux-x86_64.tar.gz
```

Figura 3-7 Comando para descargar Elasticsearch desde la web. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~$ tar -zxvf elasticsearch-7.17.6-linux-x86_64.tar.gz
```

Figura 3-8 Descompresión de la carpeta de Elasticsearch. (Autoría propia)

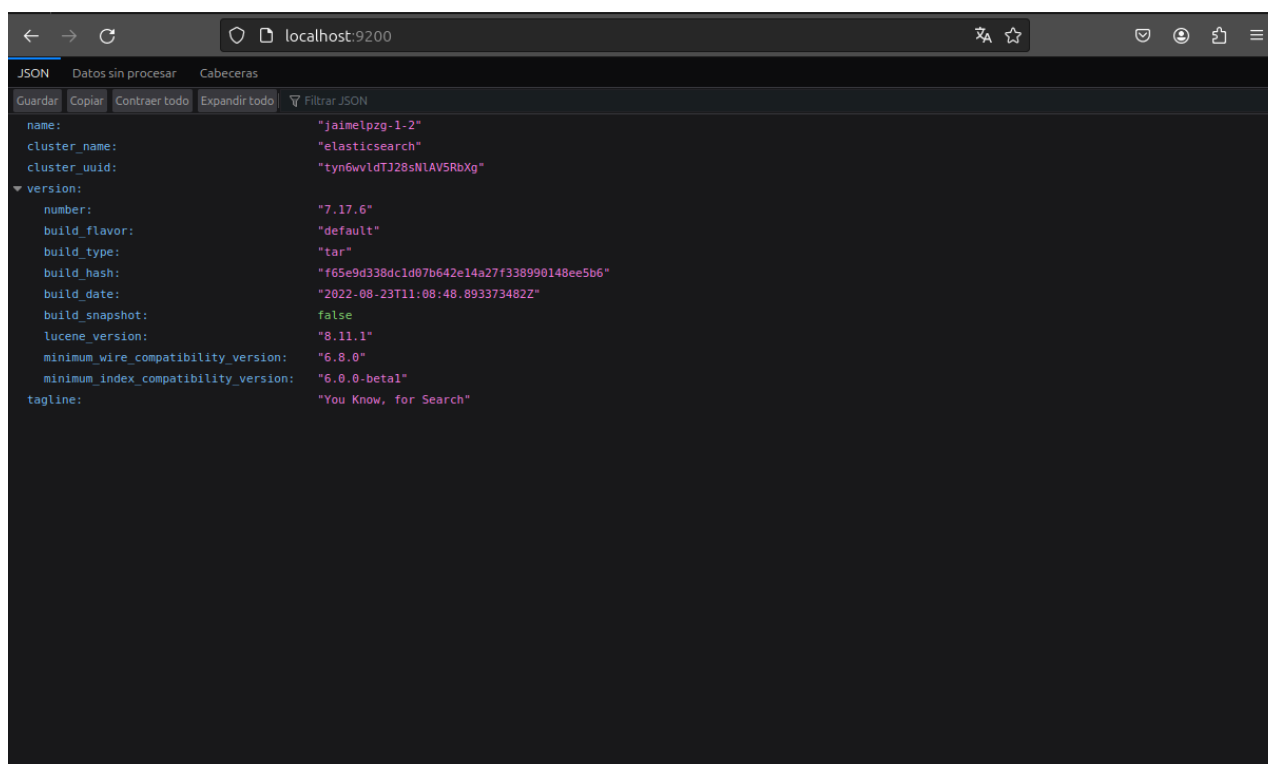


Figura 3-9 Servidor de Elasticsearch en localhost. (Autoría propia)

El cliente de esa misma versión de Elasticsearch ha de ser a su vez instalado en el editor de texto PyCharm como librería de Python. Para ello, en los ajustes del editor de texto, se seleccionará la pestaña “Python Interpreter”. Esta pestaña nos muestra las librerías de Python que el editor ha instalado, así como la versión instalada (véase Figura 3-10). Para la instalación de nuevas librerías se selecciona el símbolo + en la esquina superior izquierda, esto abrirá un desplegable que permite buscar entre un directorio de librerías de código Python, así como seleccionar la versión de la librería que se pretende instalar.

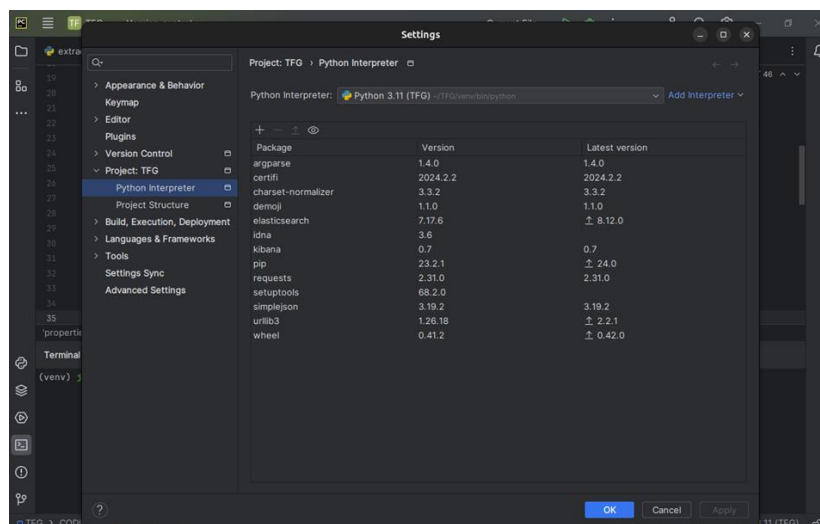


Figura 3-10 Librerías instaladas en el editor de texto PyCharm. (Autoría propia)

### 3.2.3 Kibana

Para descargar Kibana se puede realizar de la misma manera que se ha descargado Elasticsearch, para ello se escribirá el comando que se observa en la Figura 3-11. Cabe resaltar que la versión de Kibana que se descargue ha de coincidir con la versión previamente descargada de Elasticsearch (7.17.6).

```
jaimelpzg@jaimelpzg-1-2:~$ wget https://artifacts.elastic.co/downloads/kibana/kibana-7.17.6-linux-x86_64.tar.gz
```

Figura 3-11 Comando para instalar Kibana desde la web. (Autoría propia)

De la misma manera que en la descarga de Elasticsearch, se descomprime la carpeta cambiando “elasticsearch” por “kibana” (véase Figura 3-12) y se ejecuta con el comando “. /kibana” para inicializar la aplicación dentro de la carpeta “bin” de Kibana.

```
jaimelpzg@jaimelpzg-1-2:~$ tar -zxvf kibana-7.17.6-linux-x86_64.tar.gz
```

Figura 3-12 Descompresión de la carpeta de Kibana. (Autoría propia)

Una vez se ha ejecutado esta herramienta, de nuevo siguiendo el mismo procedimiento que se hizo con Elasticsearch, se comprobará que Kibana se esté ejecutando correctamente ya que, de lo contrario, no será posible acceder a la página principal. Para comprobar que efectivamente Kibana se ha instalado correctamente, se accederá al servidor de Kibana escribiendo la URL “http://localhost:5601” (véase Figura 3-13).

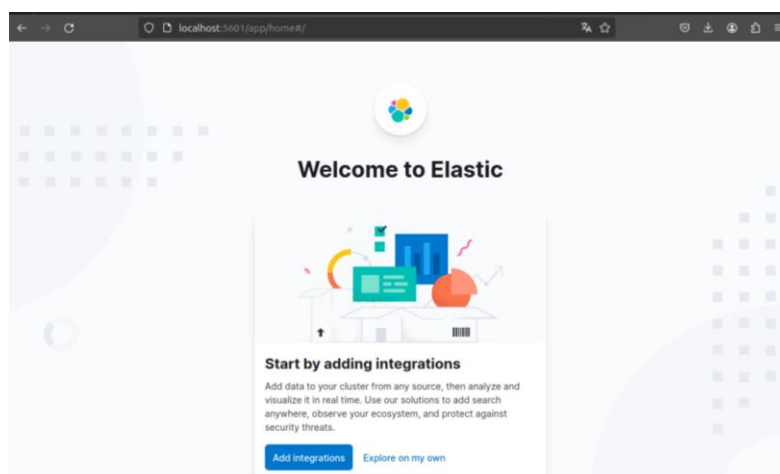


Figura 3-13 Servidor de Kibana en localhost. (Autoría propia)

### 3.3 Tratamiento de los ficheros de datos

Los ficheros de información que se analizan en este trabajo se han generado capturando tuits con la API<sup>5</sup> de *streaming* de X/Twitter entre los meses de abril y junio de 2023. Esta API permite la búsqueda filtrada de tuits en base a palabras clave y etiquetas relacionadas con la guerra de Ucrania. Las capturas de estos tuits se almacenan en formato .txt por lo que ha sido necesario transformarlos a un formato JSON antes de poder trabajar con ellos [34]. De este modo, se empleó un código para la lectura línea por línea del documento .txt y la posterior escritura de su contenido en los documentos .json empleados durante el desarrollo de este trabajo.

#### 3.3.1 Descompresión

Una vez descargados los diferentes ficheros JSON en los que se han almacenado los datos de los tuits acerca de la guerra de Ucrania, se descomprimen para así poder trabajar con ellos (véase Figura 3-14). El producto de este comando será el almacenamiento del fichero descomprimido en la carpeta en la que nos encontremos (~/TFG/CODIGO) (véase Figura 3-15).

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ #Descomprimos fichero
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ gzip -d ucrania-1.log.gz
```

Figura 3-14 Descompresión de ficheros de datos con información de interés. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ ls *.log
ucrania-10.log  ucrania-19.log  ucrania-27.log  ucrania-35.log  ucrania-43.log
ucrania-11.log  ucrania-1.log   ucrania-28.log  ucrania-36.log  ucrania-44.log
ucrania-12.log  ucrania-20.log  ucrania-29.log  ucrania-37.log  ucrania-45.log
ucrania-13.log  ucrania-21.log  ucrania-2.log   ucrania-38.log  ucrania-4.log
ucrania-14.log  ucrania-22.log  ucrania-30.log  ucrania-39.log  ucrania-5.log
ucrania-15.log  ucrania-23.log  ucrania-31.log  ucrania-3.log   ucrania-6.log
ucrania-16.log  ucrania-24.log  ucrania-32.log  ucrania-40.log  ucrania-7.log
ucrania-17.log  ucrania-25.log  ucrania-33.log  ucrania-41.log  ucrania-8.log
ucrania-18.log  ucrania-26.log  ucrania-34.log  ucrania-42.log  ucrania-9.log
```

Figura 3-15 Carpeta en la que se almacenan los ficheros descomprimidos. (Autoría propia)

#### 3.3.2 Limpieza y extracción de información de interés

Los ficheros de datos contienen información sobre tuits en formato JSON, así como información de gestión del proceso de captura, que es de utilidad para la extracción de los tuits, pero no resulta de interés para los propósitos del trabajo, los cuales radican en el análisis y visualización de la información extraída. Es por ello por lo que en el presente apartado se explica el proceso de limpieza de estos ficheros, eliminando toda la información innecesaria. En la Figura 3-17 se muestra un ejemplo de estos caracteres que se desean eliminar. En las primeras cinco líneas se observan los parámetros utilizados para la extracción de los datos de interés, indicando a la API de XTwitter diferentes etiquetas populares de las que obtener la información. A continuación, se encuentra uno de los tuits extraídos, con los diferentes parámetros que la API ha detectado. Finalmente, el tuit acaba con la línea “===”, la cual sirve de separación entre los diferentes tuits dentro de un mismo archivo.

De este modo, el segundo paso para el tratamiento requerido de los datos de interés, una vez se hayan descomprimido los archivos, ha de ser la limpieza de estos. Esta limpieza se realiza mediante la ejecución de un código aplicado al archivo que queramos limpiar (véase Figura 3-17). Para ello, en el terminal se escribirá el comando que se observa en la Figura 3-16, lo que limpiará el archivo JSON y lo almacenará en un nuevo archivo llamado “ucrania-1.log.clean”. En el apartado a) del Anexo III se incluye la explicación del código línea por línea.

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ python3 clean_files.py ucrania-1.log > ucrania-1.log.clean
```

Figura 3-16 Comando para limpieza de los ficheros. (Autoría propia)

<sup>5</sup> API: Del inglés “Application Programing Interface” o, en español, Interfaz de Programación de Aplicaciones.

```
StreamRule(value='#UkraineRussiaWar -is:retweet lang:en', tag='UkraineRussiaWar', id='1625776982368694272')
StreamRule(value='#UkraineWar -is:retweet lang:en', tag='UkraineWar', id='1625776988249096192')
StreamRule(value='#Ukraine -is:retweet lang:en', tag='Ukraine', id='1625776993898795009')
StreamRule(value='((Ukraine OR Russia) war) -is:retweet lang:en', tag='war', id='1625777000051888128')
StreamRule(value='#RussianArmy -is:retweet lang:en', tag='RussianArmy', id='1625777005319839746')
{'author_id': '93793462', 'context_annotations': [{'domain': {'id': '30', 'name': 'Entities [Entity Service]',
'description': 'Entity Service top level domain, every item that is in Entity Service should be in this domain'},
'entity': {'id': '781974596752842752', 'name': 'Services'}}, {'domain': {'id': '47', 'name': 'Brand', 'description':
'Brands and Companies'}, 'entity': {'id': '10045225402', 'name': 'Twitter'}}], 'created_at':
'2023-02-15T08:42:05.000Z', 'edit_history_tweet_ids': ['1625777453854777344'], 'entities': {'annotations': [{'start':
109, 'end': 114, 'probability': 0.9823, 'type': 'Place', 'normalized_text': 'Russia'}, {'start': 141, 'end': 145,
'probability': 0.9722, 'type': 'Place', 'normalized_text': 'India'}, {'start': 187, 'end': 192, 'probability': 0.9761,
'type': 'Place', 'normalized_text': 'Russia'}, {'start': 235, 'end': 239, 'probability': 0.9693, 'type': 'Place',
'normalized_text': 'India'}], 'mentions': [{'start': 0, 'end': 13, 'username': 'SidiqFaizaan', 'id':
'1299737983264776193'}, {'start': 14, 'end': 27, 'username': 'bhogleharsha', 'id': '78941611'}]}, 'id':
'1625777453854777344', 'lang': 'en', 'text': "@SidiqFaizaan @bhogleharsha I don't understand why people are giving so
much time on this tweet of yours 😊😊 \nRussia is tired now due to war. India is fueling somewhat by buying fuel from
Russia. Keep in mind the power of young and new India.", 'local_time_ns': 1676450531201493444}
===
```

Figura 3-17 Fichero de datos en formato JSON. (Autoría propia)

Ahora que el archivo está limpio, será necesario extraer los datos que nos interesan de él. Para ello se ha realizado el código que se explica en detalle en el apartado b) del Anexo III. Para la ejecución de este código es necesaria la instalación de la librería “demoji” (véase Figura 3-18) tanto en el terminal de Linux como en PyCharm desde el “*interpreter*”. La librería “demoji” [52] sirve para manipular emoticonos dentro de cadenas de texto en Python, lo que puede ser útil para el procesamiento de texto que se pretende realizar con estos archivos, eliminando estos emoticonos.

```
jaimelpzg@jaimelpzg-1-2:~$ pip3 install demoji --break-system-packages
```

Figura 3-18 Comando de instalación librería demoji. (Autoría propia)

En los archivos limpios se almacenan tuits en formato JSON, uno por línea. Lo que se hace a continuación es extraer la información de interés para el análisis posterior (autor, fecha de creación, menciones, etiquetas, anotaciones, número de *likes*, etc.) además del texto del tuit, del que se eliminan enlaces, emoticonos, etiquetas y menciones mediante las líneas de código que se observan en la Figura 3-19. Para la extracción de esta información de interés, se crea un diccionario de salida en el que se almacenará la información detectada en cada una de las líneas del archivo. Es necesario definir cada uno de los parámetros que se desean almacenar en este diccionario como se observa en la Figura 3-20.

```
6 def clean_tweet_text(tweet_text):
7     cleaned_tweet = re.sub(pattern=r'https?://\S+\S+', repl='', tweet_text) # Eliminar los enlaces
8     cleaned_tweet = re.sub(pattern=r'@w+\S+', repl='', cleaned_tweet) # Eliminar las menciones
9     cleaned_tweet = re.sub(pattern=r'#(\w+)', repl='', cleaned_tweet) # Eliminar los hashtags
10    cleaned_tweet = demoji.replace(cleaned_tweet, repl='')
11    return(cleaned_tweet)
```

Figura 3-19 Eliminación de información que no es de interés. (Autoría propia)

```
48 if "hashtags" in data_dict["entities"]:
49     etiquetas=data_dict["entities"]["hashtags"]
50     output_dict["hashtags"] = []
51     for etiqueta in etiquetas:
52         output_dict["hashtags"].append(etiqueta["tag"])
53
54 if "public_metrics" in data_dict:
55     if "like_count" in data_dict["public_metrics"]:
56         output_dict["likes"] = data_dict["public_metrics"]["like_count"]
57
58     if "retweet_count" in data_dict["public_metrics"]:
59         output_dict["retuits"] = data_dict["public_metrics"]["retweet_count"]
```

Figura 3-20 Parámetros almacenados en el diccionario de salida. (Autoría propia)

Una vez ejecutado, el código irá leyendo los archivos JSON línea por línea y almacenando en el diccionario de salida los parámetros definidos que detecte (véase Figura 3-21). A continuación, mediante el comando que se observa en la Figura 3-22, se crea un nuevo archivo con la información extraída del anterior. Una vez finalizada la ejecución del comando, el nuevo archivo se creará en la ruta de acceso en la que se haya ejecutado el comando. Se aplica este proceso para todos los archivos de información JSON ya “limpiados” con lo que en la ruta de acceso de trabajo se almacenarán los archivos resultantes del proceso de limpieza, así como los extraídos (véase Figura 3-23).

```
13 file = open(sys.argv[1], encoding='utf-8', errors='ignore')
14 lines = file.readlines()
15 for line in lines:
16     if line.startswith("{"):
17         output_dict = dict()
18         data_dict = json.loads(line)
```

Figura 3-21 Definición parámetros a eliminar y lectura texto. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ python3 extrac_data2.py ucrania-1.log.clean > ucrania-1.extract
```

Figura 3-22 Comando para la extracción de los datos de interés de los ficheros. (Autoría propia)

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ ls *.extract *.log.clean
ucrania-10.extract      ucrania-19.extract      ucrania-27.extract      ucrania-35.extract      ucrania-43.extract
ucrania-10.log.clean    ucrania-19.log.clean    ucrania-27.log.clean    ucrania-35.log.clean    ucrania-43.log.clean
ucrania-11.extract      ucrania-1.extract       ucrania-28.extract      ucrania-36.extract      ucrania-44.extract
ucrania-11.log.clean    ucrania-1.log.clean     ucrania-28.log.clean    ucrania-36.log.clean    ucrania-44.log.clean
ucrania-12.extract      ucrania-20.extract      ucrania-29.extract      ucrania-37.extract      ucrania-45.extract
ucrania-12.log.clean    ucrania-20.log.clean    ucrania-29.log.clean    ucrania-37.log.clean    ucrania-45.log.clean
ucrania-13.extract      ucrania-21.log.clean    ucrania-2.log.clean     ucrania-38.log.clean    ucrania-4.log.clean
ucrania-13.log.clean    ucrania-21.log.clean    ucrania-2.log.clean     ucrania-38.log.clean    ucrania-4.log.clean
ucrania-14.extract      ucrania-22.extract      ucrania-30.extract      ucrania-39.extract      ucrania-5.extract
ucrania-14.log.clean    ucrania-22.log.clean    ucrania-30.log.clean    ucrania-39.log.clean    ucrania-5.log.clean
ucrania-15.extract      ucrania-23.extract      ucrania-31.extract      ucrania-3.extract       ucrania-6.extract
ucrania-15.log.clean    ucrania-23.log.clean    ucrania-31.log.clean    ucrania-3.log.clean     ucrania-6.log.clean
ucrania-16.extract      ucrania-24.extract      ucrania-32.extract      ucrania-40.extract      ucrania-7.extract
ucrania-16.log.clean    ucrania-24.log.clean    ucrania-32.log.clean    ucrania-40.log.clean    ucrania-7.log.clean
ucrania-17.extract      ucrania-25.extract      ucrania-33.extract      ucrania-41.extract      ucrania-8.extract
ucrania-17.log.clean    ucrania-25.log.clean    ucrania-33.log.clean    ucrania-41.log.clean    ucrania-8.log.clean
ucrania-18.extract      ucrania-26.extract      ucrania-34.extract      ucrania-42.extract      ucrania-9.extract
ucrania-18.log.clean    ucrania-26.log.clean    ucrania-34.log.clean    ucrania-42.log.clean    ucrania-9.log.clean
```

Figura 3-23 Almacenamiento de los archivos limpiados y los extraídos. (Autoría propia)

## 3.4 Carga de archivos en Elasticsearch

### 3.4.1 Creación de un índice

Una vez que se tenga los archivos con la información procesada que de verdad interesa para su análisis, será necesario implementar un código Python para la creación de un índice en Elasticsearch en el que se almacenará esta información; en este código será necesario definir la URL del servidor de Elasticsearch y un índice en el que almacenar los datos, para que no exista problema con múltiples índices se elimina cualquier posible índice ya existente en la plataforma (véase Figura 3-24). En el apartado c) del Anexo III se observa la explicación del código mencionado línea a línea.

```
3 index_name="indice_tweets"
4 elastic = Elasticsearch("http://localhost:9200")
5 elastic.indices.delete(index=index_name, ignore=[400, 404])
```

Figura 3-24 Extracto del código para la creación de un índice en Elasticsearch. (Autoría propia)

Los datos que se almacenen en Elasticsearch se han de definir mediante a una estructura de *mappings* para poder ser almacenados y consultados. Los *mappings* especifican los tipos de datos de los campos de un índice de Elasticsearch y cómo se indexan y analizan dichos campings, lo que les hace un elemento crucial para la correcta interpretación, almacenamiento y búsqueda de datos, permitiendo una consulta efectiva de estos (véase Figura 3-25).

```

12 mappings = {
13     'properties': {
14         "url": {"type": "keyword"},
15         "autor": {"type": "keyword"},
16         "profile": {"type": "keyword"},
17         "fecha_creacion": {"type": "date"},
18         "sello": {"type": "long"},
19         "texto": {"type": "text"},
20         "hashtags": {"type": "keyword"},
21         "menciones_nombre": {"type": "keyword"},
22         "menciones_id": {"type": "keyword"},
23         "anotaciones": {"type": "keyword"},
24         "retuits": {"type": "long"},
25         "likes": {"type": "long"}

```

Figura 3-25 *Mappings* definidos en el índice creado en Elasticsearch. (Autoría propia)

En resumen, el código que se define en dicho anexo se implementará para crear un índice en Elasticsearch con sus mencionados ajustes y borrar cualquier índice existente con el mismo nombre que pueda existir en ese instante. Una vez creado el *script*<sup>6</sup> Python denominado “create\_elasticsearch\_index.py”, se implementará el índice en Elasticsearch con el comando que se observa en la Figura 3-26.

```
jainelpzg@jainelpzg-1-2:~/TFG/CODIGO$ python3 create_elasticsearch_index.py
```

Figura 3-26 Creación del índice creado en la plataforma Elasticsearch. (Autoría propia)

Tras creado dicho índice, será necesario configurar un *pattern* asociado a este en el archivo create\_elasticsearch\_index, de modo que el índice sea localizable y accesible desde Kibana. En la ventana “create index pattern” se ha de elegir un nombre para este nuevo *pattern*. Asimismo, es esencial que este *pattern* se asocie con el índice que se haya creado previamente en la plataforma (véase Figura 3-27).

Figura 3-27 Creación de un patrón del índice en Elasticsearch. (Autoría propia)

<sup>6</sup> *Script*: anglicismo que indica un texto en que se expone en detalle el contenido de un programa, en este caso de un código Python. [53]

### 3.4.2 Indexación de datos

Llegados a este punto, Elasticsearch se ha configurado correctamente. El siguiente paso será la redacción de un código que permita indexar los datos de cada archivo en el índice previamente creado. El código redactado leerá línea a línea el archivo que se pretenda indexar e irá cargando los datos de este en Elasticsearch. De esta forma, los parámetros definidos en el diccionario de salida de los ficheros extraídos se añadirán al patrón creado, rellenando cada parámetro del índice definido en Elasticsearch. Consecuentemente, será posible indexar los diferentes archivos JSON en Elasticsearch con el comando que se observa en la Figura 3-28. En el apartado d) del Anexo III se encuentra una explicación de las diferentes líneas de código del *script* “index\_data.py” que permiten la indexación de los archivos de información.

```
jaimelpzg@jaimelpzg-1-2:~/TFG/CODIGO$ python3 index_data.py ucrania-2.extract
```

Figura 3-28 Comando para la indexación de los ficheros de datos en Elasticsearch. (Autoría propia)

Una vez cargados los archivos, es aconsejable realizar una comprobación de que los datos se hayan indexado correctamente. Para ello, se accederá a la pestaña “Discover” que se encuentra en el desplegable de la izquierda de la plataforma Kibana. En esta pestaña se podrá observar cada tuit cargado en Elasticsearch y los campos de interés que se han definido en el índice previamente, los cuales se encuentran resaltados en azul (véase Figura 3-29).

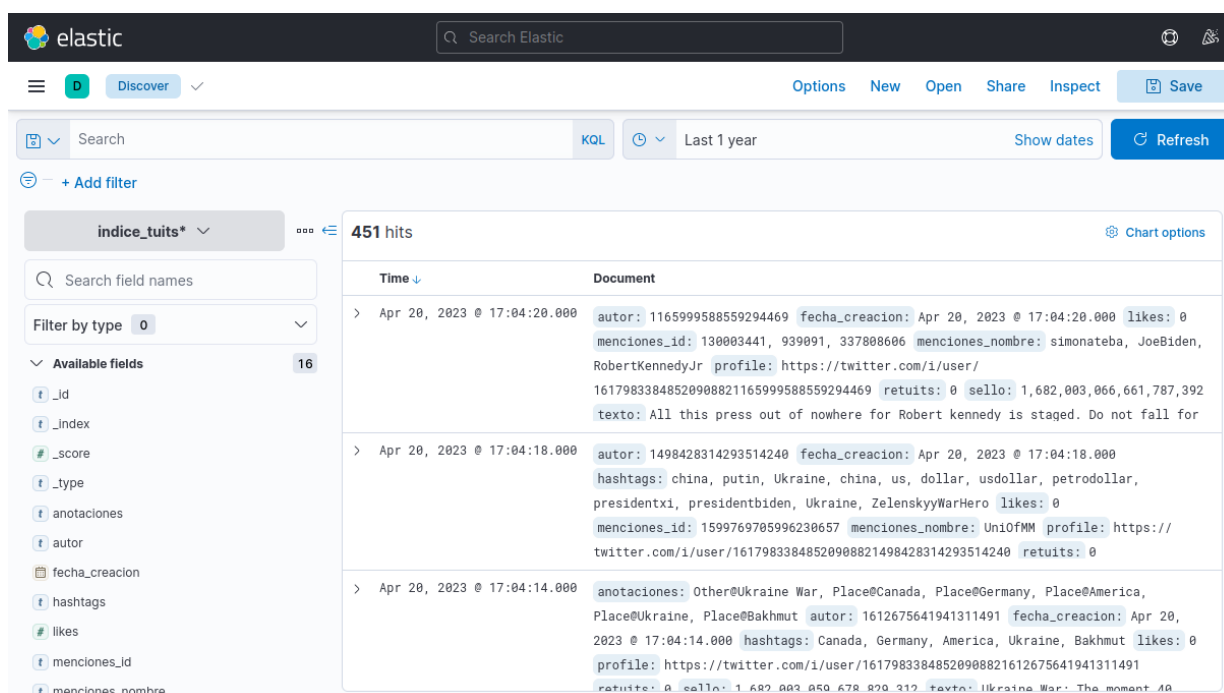


Figura 3-29 Tuits cargados en Elasticsearch. (Autoría propia)

### 3.5 Elección tipo de visualización para cada dato

En esta sección de capítulo 3, se experimentará con la herramienta Kibana para realizar las visualizaciones de los datos que se han extraído y finalmente indexados en Elasticsearch. Para ello, en el desplegable de la izquierda, se seleccionará la pestaña “Dashboard”. A continuación, aparecerá la ventana mostrada en la Figura 3-30, en la que se seleccionará el desplegable “All types”, el cual mostrará todos los tipos de gráficos que la plataforma permite crear para cada dato. Para los datos que se han extraído expresamente en este proyecto, convendrá seleccionar las opciones resaltadas en

amarillo: “*Lens*”, “*TSVB*”<sup>7</sup> y “*Aggregation based*”, las cuales ofrecen una gran variedad de tipos de gráficos a utilizar para la representación de la información obtenida de los ficheros de datos. Asimismo, se podrá ajustar la fecha para visualizar una representación de los datos dentro de un periodo específico de tiempo seleccionando el icono de calendario en la esquina superior derecha en pantalla.

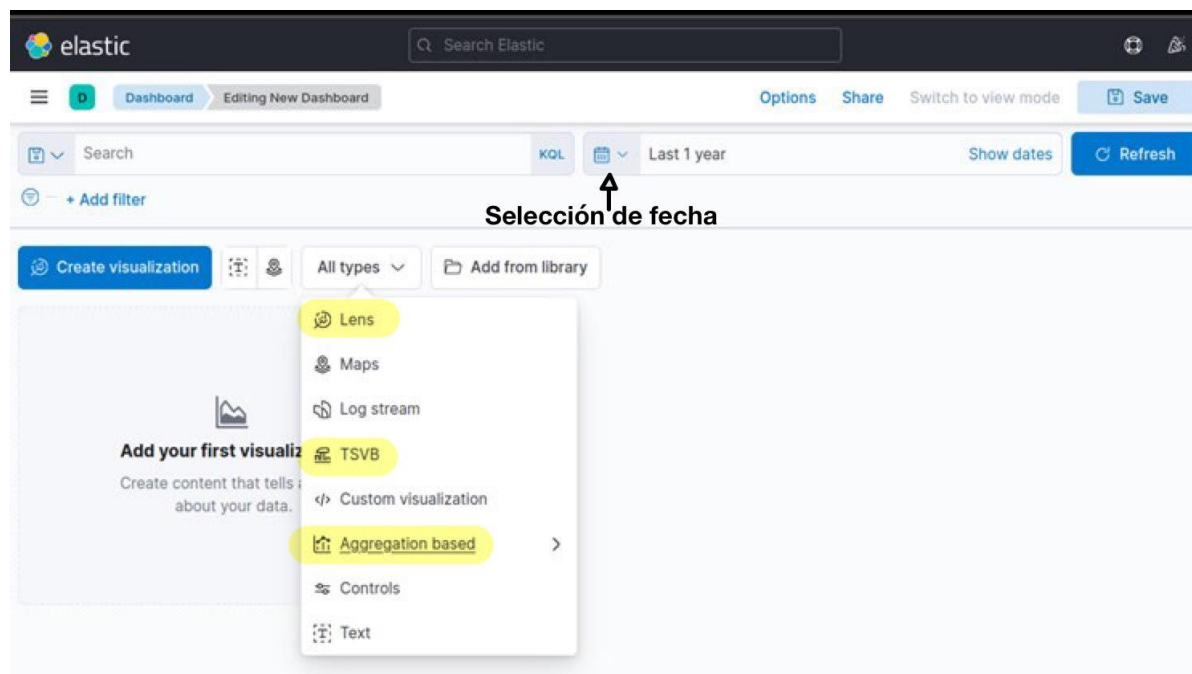


Figura 3-30 Creación de visualizaciones de datos. (Autoría propia)

### 3.5.1 *Lens*

Dentro de la pestaña “*Lens*” Kibana ofrece una experiencia intuitiva y orientada a objetos para crear gráficos de manera fácil y rápida. Esta pestaña destaca por su enfoque en la construcción visual y la interactividad, lo que permite un análisis de datos dinámico. Dentro de dicha pestaña se diferencian 5 tipos de visualización:

- *Tabular and single value*: Dentro de este tipo de visualizaciones se encuentran dos opciones: la primera opción es “*metric*”, la cual ofrece datos numéricos acerca de parámetros específicos que se hayan registrado en el índice de Elasticsearch (véase Figura 3-31). La segunda opción de este tipo de visualización es “*table*”, la cual ofrece tablas que se pueden configurar para que presenten tantas columnas de datos como queramos, con el fin de entrelazar diferentes parámetros entre sí (véase Figura 3-31).

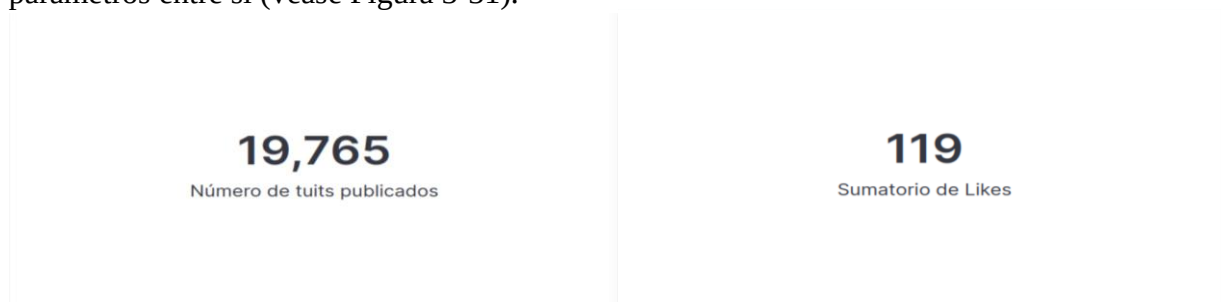


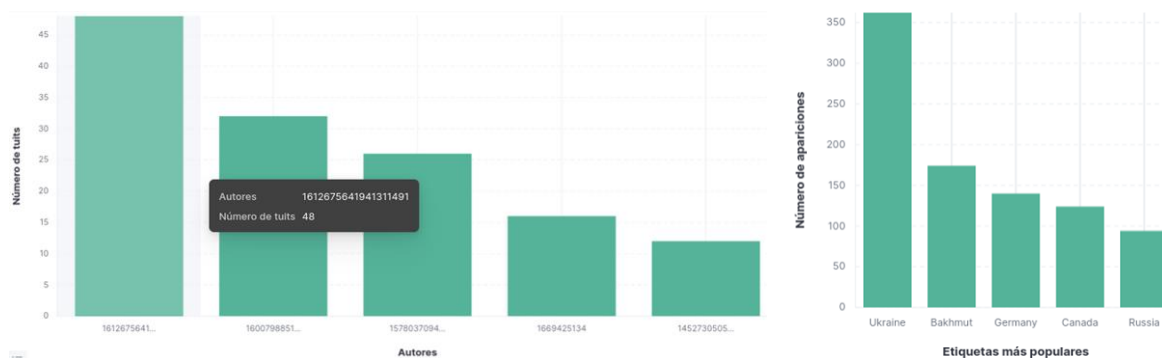
Figura 3-31 Visualización de la herramienta *metric*. (Autoría propia)

<sup>7</sup> TSVB: del inglés “*Time Series Visual Builder*”, es una herramienta utilizada para crear visualizaciones de series temporales.

| Top 10 de usuarios más mencionados | Número de menciones | Entidades más populares | Número de menciones |
|------------------------------------|---------------------|-------------------------|---------------------|
| NATO                               | 245                 | Place@Ukraine           | 210,958             |
| POTUS                              | 239                 | Place@Russia            | 85,163              |
| UN                                 | 211                 | Place@Bakhmut           | 71,078              |
| ZelenskyyUa                        | 180                 | Place@Germany           | 59,509              |
| mfa_russia                         | 118                 | Place@Canada            | 52,726              |
| JoeBiden                           | 110                 | Place@America           | 29,872              |
| SecDef                             | 105                 | Place@Kyiv              | 24,120              |
| antonioguterres                    | 105                 | Person@Putin            | 21,351              |
| CBC                                | 94                  | Place@UkraineWar        | 17,963              |
| JustinTrudeau                      | 94                  | Place@UkraineRussiaWar  | 14,832              |

Figura 3-32 Visualización de la herramienta *table*. (Autoría propia)

- *Bar*: Esta opción ofrece diferentes visualizaciones que muestran los datos en forma de barras, tanto vertical como horizontalmente. Esta modalidad de representación es idónea para comparar diversos valores numéricos, permitiendo una visualización sencilla de la disparidad entre dichos valores (véase Figura 3-33).

Figura 3-33 Visualización de la herramienta *bar*. (Autoría propia)

- *Line and area*: Este tipo de visualización presenta una gráfica lineal en la que se podrá configurar un eje horizontal, como la fecha de publicación, y valores en el eje y que se quieran analizar (véase Figura 3-34). De esta manera, por ejemplo, se pueden mostrar datos visuales acerca de la cantidad de tuits que hablan de un tema específico por fecha. De la misma manera en que los datos se pueden visualizar como líneas, este tipo de visualización ofrece una opción, denominada “*area*”, que sombrea el área bajo cada línea de datos.

Figura 3-34 Visualización de la herramienta *line*. (Autoría propia)

- *Proportion*: Este tipo de visualización ofrece la posibilidad de analizar la proporción relativa de un volumen de datos, creando gráficos en los que los datos se presentan de manera sectorizada, dependiendo del porcentaje de su volumen sobre el total de información. La opción “*proportion*” es útil para discernir la relevancia relativa de categorías de datos como etiquetas más populares o una distribución de los usuarios más activos (véase Figura 3-35).



Figura 3-35 Visualización de la herramienta *proportion*. (Autoría propia)

- *Heatmap*: Este tipo de visualización permite representar datos de manera visual utilizando colores para indicar la densidad o intensidad de los valores en una cuadrícula (véase Figura 3-36).

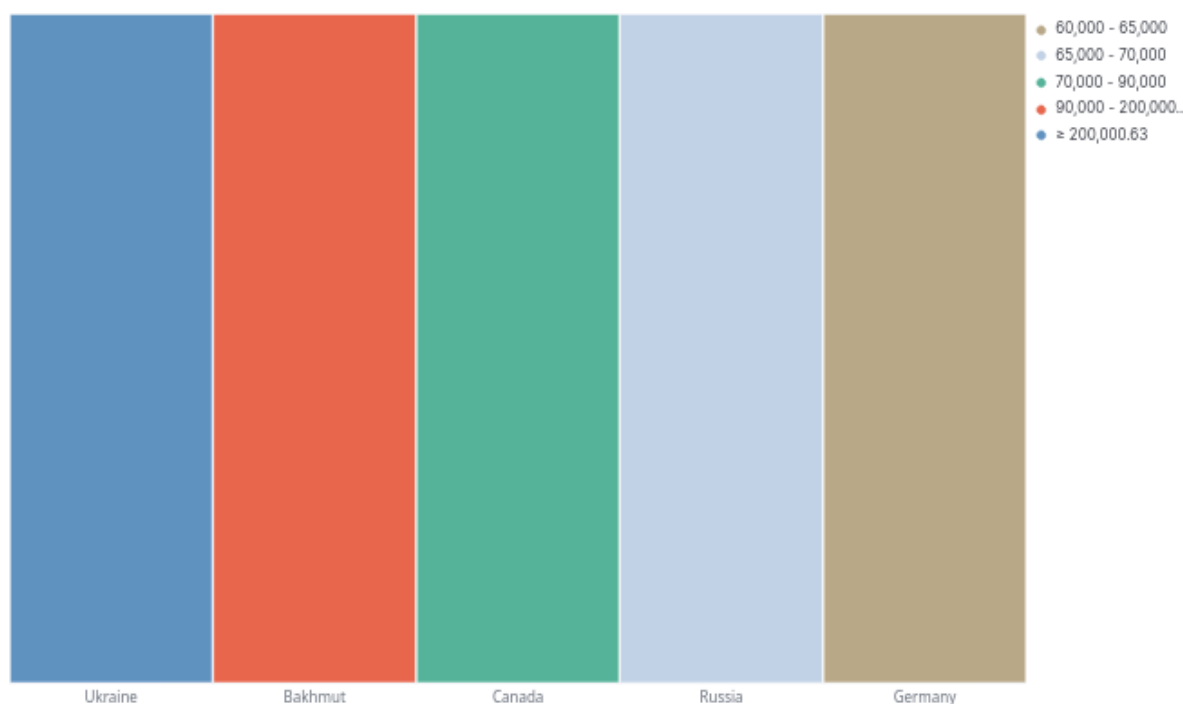


Figura 3-36 Visualización de la herramienta *heatmap*. (Autoría propia)

### 3.5.2 TSVB

Otra forma de obtener las gráficas deseadas es mediante esta opción, TSVB permite construir visualizaciones de series temporales mediante la superposición de múltiples capas. Cada capa puede contener una o varias series temporales con diferentes métricas y agregaciones. Las visualizaciones creadas en TSVB se pueden integrar fácilmente en los *dashboards* de Kibana junto con otras visualizaciones para construir paneles de control completos y personalizados. Dentro de la pestaña TSVB se encuentran 6 opciones para realizar el análisis de datos deseado:

- *Time Series*: Esta opción ofrece una gráfica como la que se observa en la Figura 3-37. En la parte inferior derecha se puede filtrar la serie de datos a representar en la gráfica, del mismo modo se pueden añadir más series de datos.



**Figura 3-37 Visualización de la herramienta *Time Series*. (Autoría propia)**

- *Metric*: Esta opción ofrece un resultado similar al de la opción “*metric*” de la pestaña *Lens*, con la ventaja de comparar dos volúmenes de datos (véase Figura 3-38).



**Figura 3-38 Visualización de la herramienta *metric*. (Autoría propia)**

- *Top N*: Esta opción ofrece una comparativa del volumen de datos de una serie temporal. Esta opción es útil para identificar tendencias e identificar los N valores más altos de un campo. (véase Figura 3-39).



**Figura 3-39 Visualización de la herramienta *Top N*. (Autoría propia)**

- *Gauge*: Esta opción ofrece un gráfico que representa el número de veces que se repite un valor (véase Figura 3-40). *Gauge* es útil para proporcionar una representación visual rápida y clara de un valor y la cantidad de veces que este aparece, lo que facilita la supervisión y comprensión de los datos en los tableros de control y paneles.



**Figura 3-40 Visualización de la herramienta *Gauge*. (Autoría propia)**

- *Table*: Similar a la opción “*table*” de *Lens*, aunque en este caso la tabla se realiza mediante una serie temporal.

### 3.5.3 Aggregation based

Esta pestaña se enfoca en crear visualizaciones basadas en agregaciones de datos, lo que significa que los datos se agrupan y se calcula una métrica sobre ellos para representarla en el gráfico. Adicionalmente, la pestaña permite editar el tipo de agregación para la representación de datos, algunas de las agregaciones comunes incluyen: suma, promedio, mínimo, máximo, conteo, mediana, desviación estándar y percentiles. Dentro de esta pestaña, se encuentran diferentes opciones para la creación de gráficas:

- *Area*: permite representar datos de manera visual en un gráfico de área, lo que facilita la identificación de tendencias, patrones y cambios a lo largo del tiempo o en relación con otras dimensiones (véase Figura 3-41).

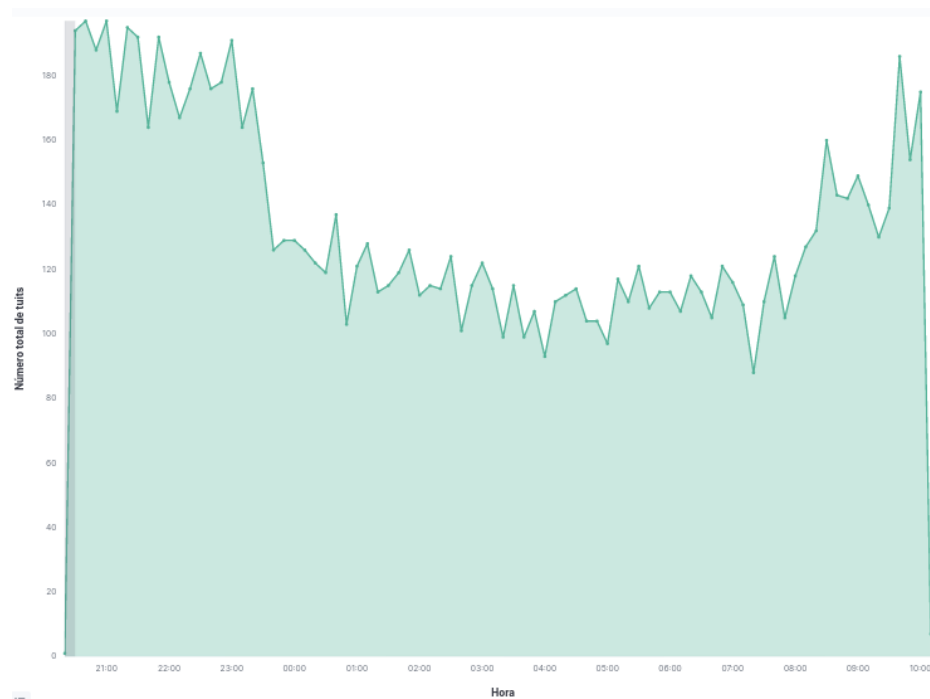


Figura 3-41 Visualización de la herramienta *área*. (Autoría propia)

- *Gauge*: similar a la opción “*gauge*” de TSVB, con la posibilidad de mostrar la cantidad de datos relativos a múltiples parámetros (véase Figura 3-42).

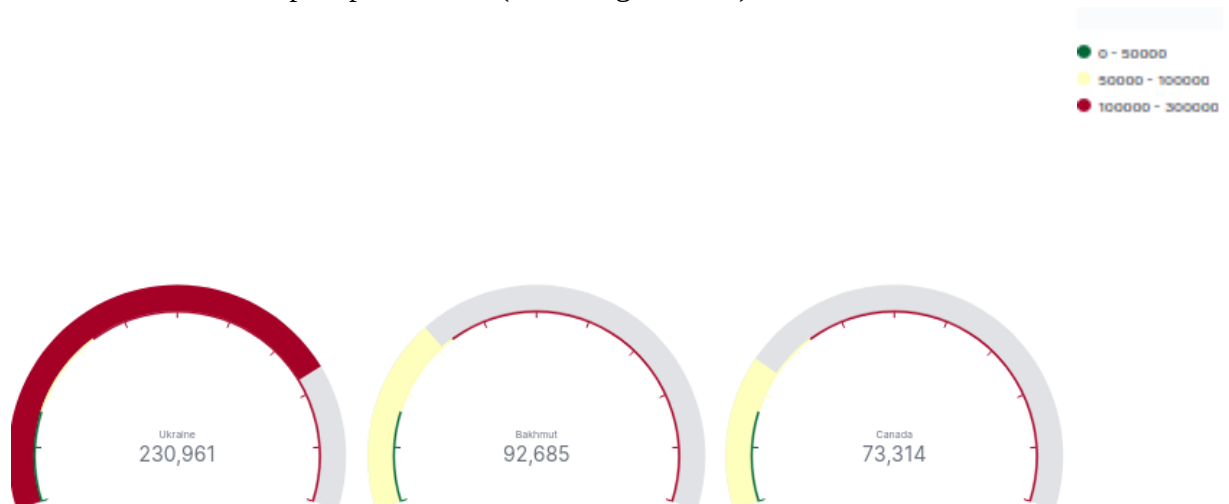


Figura 3-42 Visualización de la herramienta *Gauge* en *Aggregation based*. (Autoría propia)

- *Goal*: Esta opción permite establecer y visualizar objetivos específicos dentro de determinados datos, lo que facilita la comparación y el análisis de métricas en relación con las metas predefinidas (véase Figura 3-43).



**Figura 3-43 Visualización de la herramienta *Goal* en *Aggregation based*. (Autoría propia)**

- *Heatmap*: Similar a la opción “heatmap” de *Lens* aunque la opción de *Aggregation based* proporciona más flexibilidad y control para crear visualizaciones personalizadas utilizando agregaciones de datos de manera más detallada.
- *Vertical/Horizontal bar*: Permite representar datos con barras, adecuado para valores entre diferentes categorías o periodos de tiempo, aunque se precisan un mayor ingreso de datos para generar las visualizaciones que en *Lens*.
- *Line*: Muestra los datos como una serie de puntos conectados por líneas, útil para visualizar tendencias y cambios en los datos a lo largo del tiempo, aunque debido a la agrupación de un alto volumen de datos y sus métricas calculadas, la curva de aprendizaje para saber utilizar esta opción es más elevada.
- *Metric*: Muestra una única métrica como un número o un gráfico de un solo punto, útil para resaltar un valor específico, y del mismo modo que en *Lens* y TSVB. Una ventaja sobre estas es la posibilidad que ofrece de analizar un amplio volumen de datos, aunque únicamente permite comparar datos de un tipo de métrica; es decir, permite comparar la métrica de diferentes valores con respecto a un único parámetro (número de *likes* que reciben, cantidad de retuits, volumen total de veces que se repiten, etc.) (véase Figura 3-44).

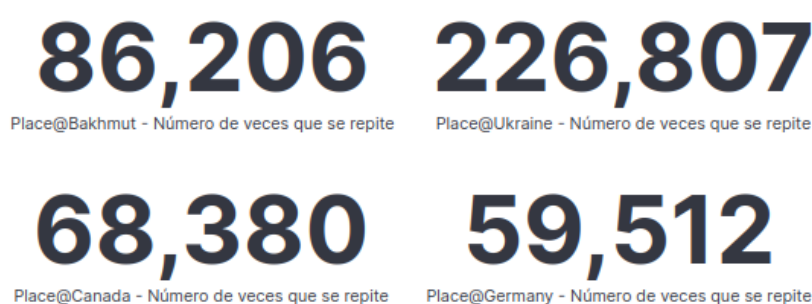


**Figura 3-44 Comparativa de las anotaciones más comunes mediante la opción *metric*. (Autoría propia)**

- *Pie*: De manera similar a *Lens*, representa los datos como segmentos de un círculo.
- *Data table*: Muestra los datos en forma de tabla, útil para ver los datos en detalle y realizar comparaciones específicas entre diferentes entradas. Esta opción ofrece la posibilidad de incluir agregaciones personalizadas para el cálculo de métricas específicas (véase Figura 3-45), así como también ofrece la posibilidad de editar el formato de la tabla. En definitiva, esta opción es más completa que la opción *table* de *Lens* y permite llevar a cabo una visualización eficiente sin la necesidad que requiere TSVB de una mayor familiarización con la herramienta, debida a las diferentes series temporales que se han de incluir en cada capa al emplear esta opción.

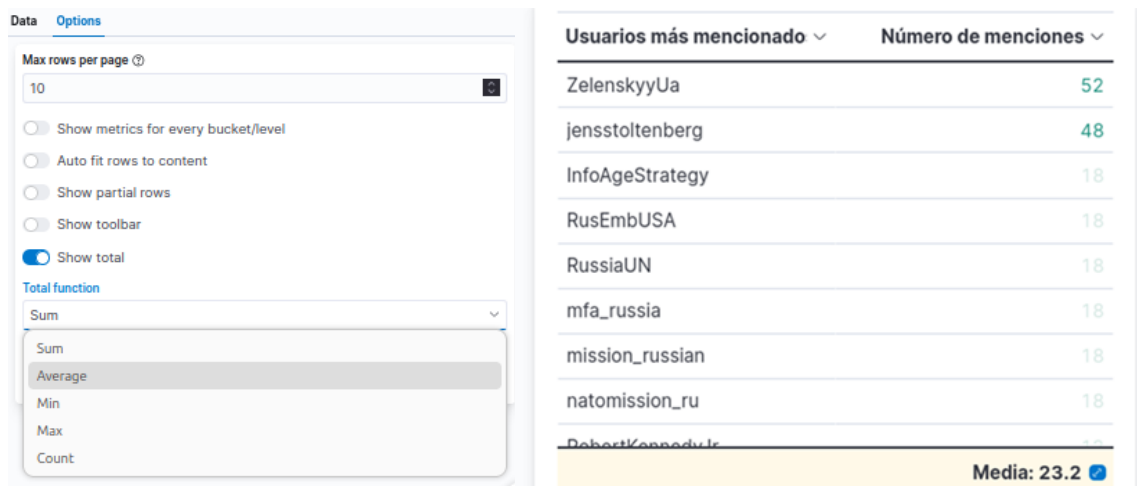


Figura 3-45 Visualización de table en *Aggregation based*. (Autoría propia)

- *Tag cloud*: Esta última opción representa datos de texto, donde cada palabra o término es una etiqueta que aparece en el conjunto de datos. La frecuencia con la que aparece cada término se refleja en el tamaño de la etiqueta en la nube (véase Figura 3-46).



Figura 3-46 Visualización de la herramienta *Tag cloud*. (Autoría propia)

## 4 PRUEBA Y VALIDACIÓN

### 4.1 Creación de las visualizaciones de datos

En este capítulo se dará paso a la creación de visualizaciones utilizando los datos de los diferentes parámetros del índice creado en Elasticsearch. Estas visualizaciones se han llevado a cabo tras la indexación de un total de 45 ficheros de datos, los cuales cuentan con un total de 523.240 tuits relacionados con la guerra de Ucrania que, como previamente se menciona en el apartado 3.3, fueron capturados entre los meses de abril y junio de 2023 utilizando la API de *streaming* de X/Twitter. Previamente al análisis y visualización de los datos, es necesario crear un *dashboard*, seleccionando esta opción dentro de la pestaña “*analytics*” situada en el lateral izquierdo de la pantalla (véase Figura 4-1). En este tablero de control se incrustan los gráficos de la información recogida. Una vez dentro del tablero, el proceso para la creación de los gráficos será semejante al explicado en el apartado 3.5 de la memoria.

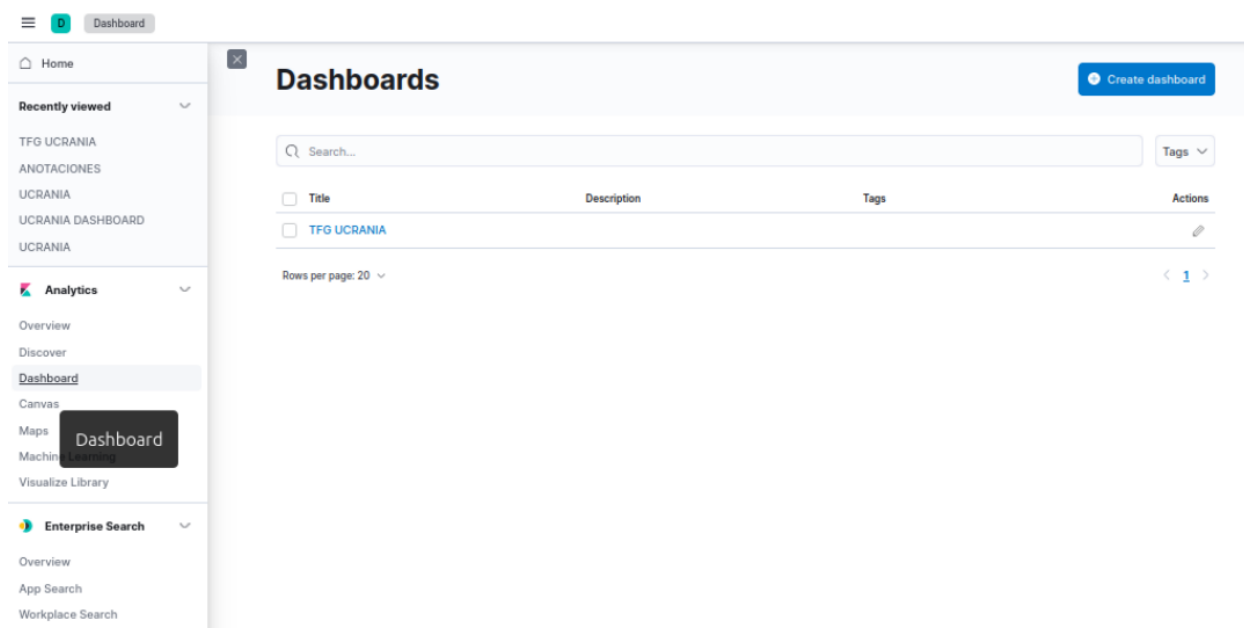


Figura 4-1 Creación de un *dashboard* en Kibana. (Autoría propia)

Previamente a la elección del tipo de gráfico utilizado para cada dato, se ha de diferenciar en función de las necesidades en cuanto a las visualizaciones que más concuerden con cada tipo de dato. Es por ello por lo que en los siguientes apartados se realiza un estudio de las visualizaciones que mejor

encajan con cada parámetro del índice, resultado del análisis completo de todos los ficheros de información.

#### 4.1.1 Autor

Para realizar una representación de los autores más activos sobre el conflicto de Ucrania en la muestra de datos de X/Twitter, es interesante la opción de *Aggregation based* que ofrece Kibana ya que ofrece un mayor control y flexibilidad sobre la estructura y el comportamiento de la visualización. Analizando las otras dos opciones, *Lens* y *TSVB*, estas no se ajustan tanto al objetivo de visualización que se busca; *Lens* ofrece visualizaciones rápidas sin complejidad técnica mientras que *TSVB* ofrece datos temporalmente asociados, lo cual no es el objetivo principal de la representación de los autores más activos.

Habiendo descartado las otras dos opciones, *Aggregation based* permite la representación de este tipo de datos en forma de barras, así como en forma de tabla como se ha explicado en el apartado 3.5.3 de la memoria. A continuación, se procede a representar estos datos mediante las herramientas empleadas (véase Figura 4-2):



Figura 4-2 Visualizaciones de usuarios más activos mediante *Aggregation based*. (Autoría propia)

#### 4.1.2 Fecha de creación

Para el desarrollo de este proyecto, los datos extraídos corresponden a tuits publicados entre los meses de abril y junio de 2023. Es interesante analizar el volumen de datos perteneciente a cada uno de estos meses. Como se explica en el apartado 3.5.1, la opción “*proportion*” dentro de *Lens* permite analizar la proporción relativa de un volumen de datos para poder discernir entre los meses más activos en cuanto a la recopilación de datos de X/Twitter. Dentro de *Lens* también se puede realizar una visualización que concierne el número diario de tuits publicados mediante un gráfico de barras (véase Figura 4-3).

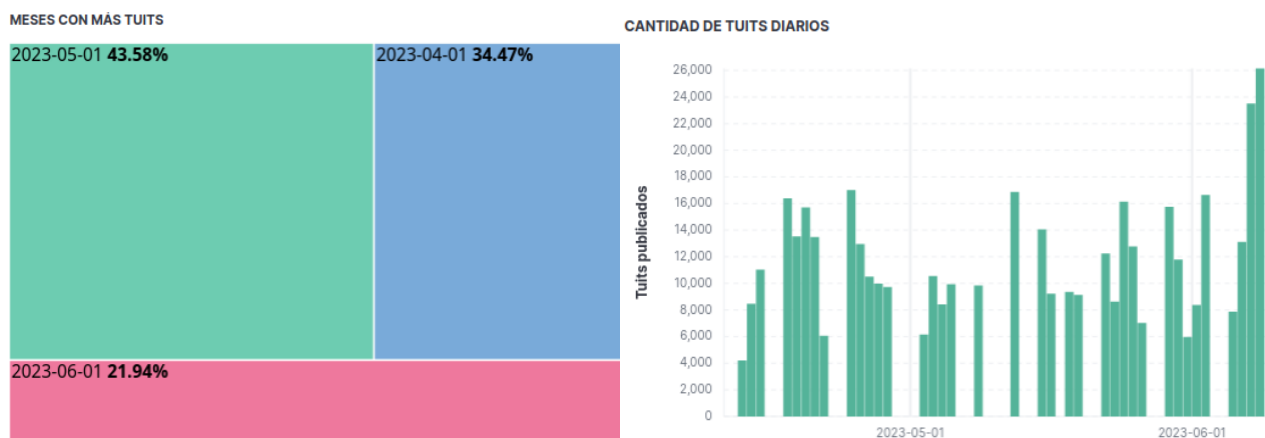


Figura 4-3 Visualizaciones de la cantidad de tuits por fecha. (Autoría propia)

Realizando un análisis más detenido sobre el mes de mayo, el cual es el mes más activo por parte de los usuarios de la red social. Se puede observar que en un total de 12 días no se extrajo ningún dato de tuits relativos al conflicto entre Rusia y Ucrania, esto se debe a días donde se produjeron caídas en el programa o apagados programados del servidor de captura. A partir de junio el acceso pasó a ser de pago, lo que puso fin a la captura de datos. Los días donde más se tuiteó sobre este conflicto durante el mes de mayo fueron el 12, 24 y 29, por lo que conviene analizar más profundamente la causa de este holgado volumen de datos:

- En el 12 de mayo el presidente ucraniano Zelensky afirmaba que los ocupantes rusos estaban listos para la derrota [54]. Por otro lado, el líder del grupo mercenario Wagner retaba al ministro de defensa ruso a visitar la zona disputada de Bakhmut debido a la difícil situación operativa, mientras amenazaba con retirar a su grupo de mercenarios de esta zona debido a la falta de munición. En este mismo día se confirmaba el envío de misiles de largo alcance a Ucrania por parte del Reino Unido igual que 14 aviones de combate MIG-29 por parte de Polonia.
- La madrugada del 24 de mayo se produce un ataque ucraniano con drones tanto en la ciudad rusa de Belgorod como contra el buque de reconocimiento Ivan Hurs en aguas de Turquía [55]. También esa mañana, el jefe del grupo Wagner admitía el fracaso de la campaña militar rusa en Ucrania, ya que, debido al apoyo recibido por múltiples naciones, Ucrania se ha militarizado aún más y esto habría imposibilitado el cumplimiento de los objetivos rusos [56].
- Por último, durante la mañana del 29 de mayo se produce una ofensiva rusa con misiles y drones sobre Kiev. El ejército ruso aprovechó esta ofensiva para atacar objetivos militares e infraestructuras estratégicas de Ucrania, en el que resultaba ser el mayor ataque con drones sobre Ucrania desde el comienzo de la guerra.

Dejando de lado los sucesos del mes de mayo, cabe destacar igualmente el 16 y 24 de abril al igual que el 6 y 8 de junio, los cuales registran el mayor volumen de datos extraídos sobre el conflicto, 23,510 y 26,148 tuits respectivamente:

- El día 16 de abril se produce un intercambio de prisioneros entre ambas naciones, evento que coincide con la Pascua ortodoxa. Por otro lado, el Papa expresaba su deseo de ver a los rusos y ucranianos encontrar la paz “Pienso también en nuestros hermanos y hermanas de Rusia y Ucrania que hoy celebran la Pascua. Que el Señor esté cerca de ellos y les ayude a hacer la paz”, dijo el pontífice ante los cerca de 20.000 fieles reunidos para escucharle en la plaza de San Pedro del Vaticano [57].
- El 24 de abril, según desveló The Washington Post, estaba previsto un ataque en Moscú por parte del ejército ucraniano, el cual finalmente no se llevó a cabo a petición de Estados Unidos [58]. Paralelamente, la inteligencia británica revelaba una estrategia de Putin para la integración de las áreas ocupadas de Ucrania en la burocracia de la Federación Rusa [59].
- El 6 de junio de 2023 soldados ucranianos ocupan posiciones rusas a las afueras de Bakhmut mientras Rusia afirmaba haber frustrado un ataque en Donetsk. Adicionalmente, una detonación producía la voladura de una presa en Kiev, provocando la inundación de la región de Jerson y la evacuación de más de 6.000 ucranianos.
- Por último, el día en el que se extrajeron más datos, durante el 8 de junio de 2024 crecen las tensiones entre el grupo Wagner y Rusia mientras que el secretario general de la OTAN instaba a los países miembros a acelerar la ayuda humanitaria. Este mismo día, el secretario de estado de EEU, Antony Blinken, calificaba de “fracaso estratégico” la ofensiva rusa ya que, según señalaba, la invasión había unido a Occidente, dificultando el avance ruso.

#### 4.1.3 Etiquetas

El uso de las etiquetas facilita la búsqueda de un tema específico por parte de los usuarios ya que convierten las palabras en enlaces en los que, al hacer clic, se muestran todos los tuits que comparten esa misma etiqueta. Esto permite la creación de tendencias y foros de discusión sobre un tema específico.

A lo largo de este apartado, se realiza un estudio de las etiquetas más populares y por ende de los temas que han causado más tendencia. Para la representación de este estudio, las opciones más interesantes a la hora de crear las visualizaciones son las que presenta la opción *Lens* mediante el diagrama circular y *Aggregation based* con la nube de datos (véase Figura 4-4). La gráfica que se muestra a la derecha permite representar una comparativa de las etiquetas que más se repiten en los tuits capturados, en concreto representa las 6 etiquetas más populares por porcentaje de uso. Ahora bien, cabe destacar que el porcentaje que se representa no equivale al porcentaje total de los tuits, sino al porcentaje relativo a los tuits donde se encuentra cualquiera de estas etiquetas, el cual variará dependiendo del número de etiquetas que se comparen.



Figura 4-4 Visualización de las etiquetas más populares. (Autoría propia)

De la misma forma que se pueden visualizar las etiquetas más populares, también se pueden hacer comparaciones extrayendo algunas de ellas, como se observa en la Figura 4-5. En esta figura se

visualiza la cantidad de veces que se utilizan las etiquetas #Rusia, #Ucrania y #UkraineWar. Para la realización de estas gráficas lineales interesa la opción *Lens* ya que permite la obtención de datos visuales respecto al tiempo de una forma intuitiva, lo cual facilita la interpretación del gráfico. Por el contrario, aunque esta tabla represente una serie temporal, la opción TSVB únicamente ofrece gráficas de áreas, lo cual perjudica a la visualización de las líneas cuando se comparan más de dos volúmenes de datos. Es por ello que la opción elegida para representar esta comparativa entre las etiquetas nombradas ha sido la gráfica lineal que ofrece *Lens*.

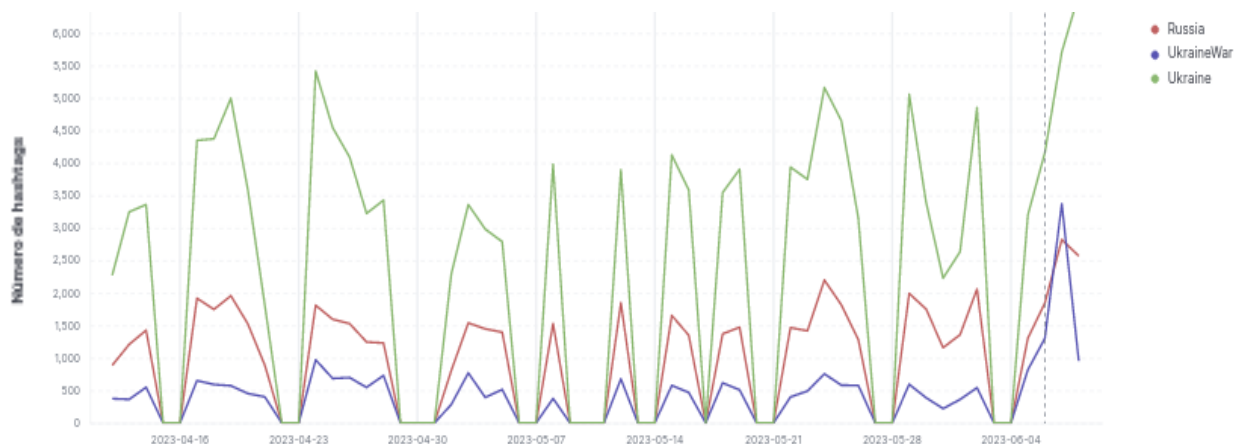


**Figura 4-5 Comparativa en el tiempo de diferentes etiquetas. (Autoría propia)**

Hay que tener en cuenta que las diferentes etiquetas cobran importancia según los eventos que vayan sucediendo. Es por ello por lo que, observando lo explicado en el apartado 4.1.2 de la memoria, la mayor cantidad de etiquetas con respecto a Bakhmut corresponden con el mes de mayo (véase Figura 4-6). El volumen de etiquetas sobre un tema específico, como por ejemplo la etiqueta #Bakhmut en el mes de mayo, es un indicativo de la popularidad del tema y del alcance global que tiene ya que, a mayor utilización de la etiqueta, mayor es el alcance y popularidad sobre el tema a nivel mundial.

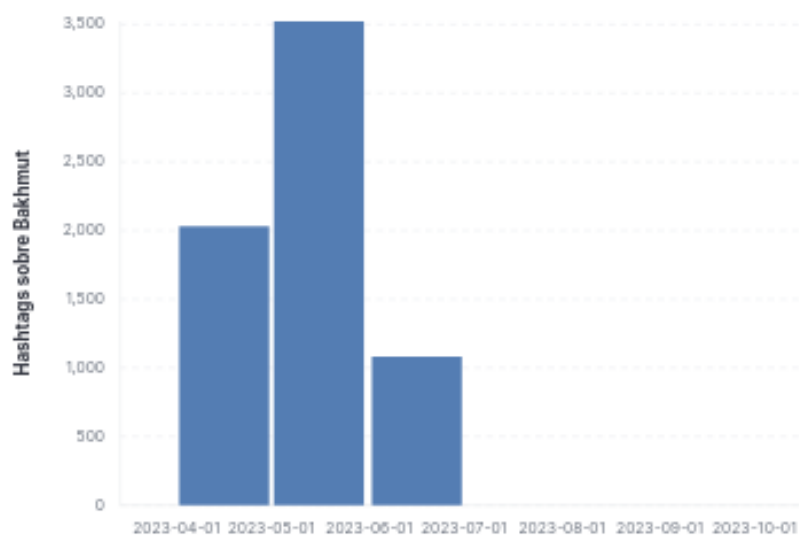


**Figura 4-6 Volumen de tuits con la etiqueta #Bakhmut por fecha. (Autoría propia)**

#### 4.1.4 Menciones

Cabe señalar la presencia de menciones a cuentas vinculadas a personajes públicos o medios de comunicación que aparecen en los tuits relacionados con el conflicto de Ucrania. Para ello, gracias a la herramienta de *Aggregation based*, se pueden representar tablas y nubes de datos que presenten este

volumen de información (véase Figura 4-7). En la tabla de datos se observan tres cuentas destacables con respecto al volumen de menciones que han obtenido entre los meses de abril y junio de 2023: @mfa\_russia pertenece a la cuenta oficial del ministerio de asuntos exteriores ruso [60], @nexta\_tv es un medio de comunicación bielorruso que opera principalmente en las redes sociales [61] y @ZelenskyyUa se trata de la cuenta oficial del presidente de Ucrania, Volodymyr Zelensky [62].

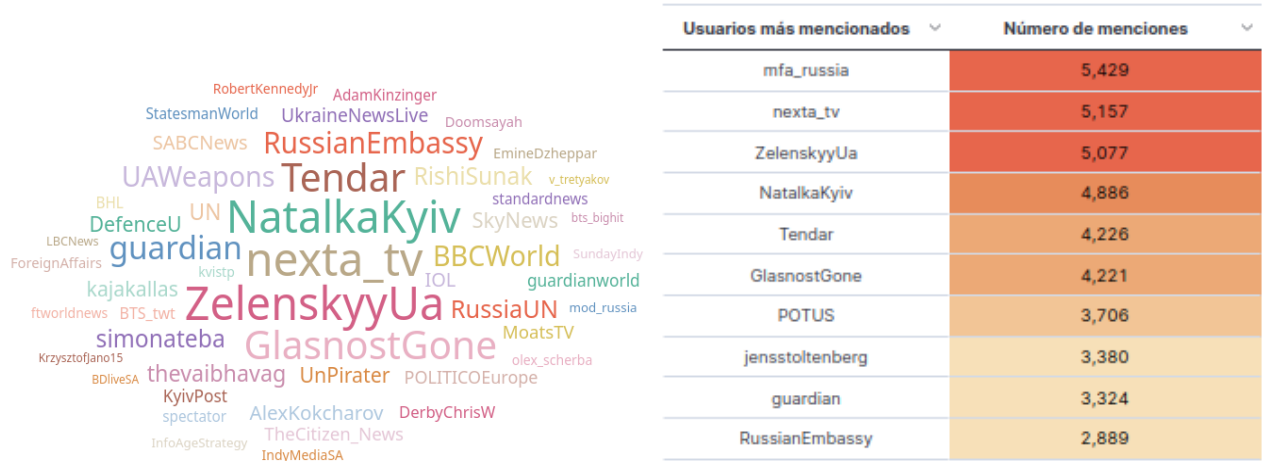


Figura 4-7 Visualización de los usuarios más mencionados. (Autoría propia)

Del mismo modo que los eventos sucedidos tienen repercusión en las etiquetas que se publican, las menciones a estas cuentas también tienen una explicación. Como se observa en la Figura 4-8, el presidente ucraniano ganó popularidad en el mes de abril, mientras que en mayo hubo más menciones hacia el medio de comunicación Nexta TV y en junio la cuenta más popular fue la del Ministerio de Asuntos Exteriores ruso.

Con respecto a la popularidad de Zelenksy en el mes de abril, destaca el día 24 de abril con un total de 629 menciones. En este día el presidente ucraniano firmaba una declaración en la que se pedía el ingreso de Ucrania en la OTAN y en la Unión Europea [63]. El día más activo de Nexta TV, con un total de 464 menciones, fue el 12 de mayo; en este día el medio de comunicaciones confirmaba los avances ucranianos en la región de Bakhmut, confirmación que coincide con el enfrentamiento del líder de Wagner y el ministro de defensa ruso. Destacable con respecto al Ministerio de Asuntos Exteriores ruso, el 1 de junio se menciona a su cuenta en un total de 526 ocasiones, este volumen de menciones coincide con la acusación de Moscú a Kiev de preparar ataques contra Rusia desde EEUU aunque Ucrania se comprometiese a no utilizar contra territorio ruso los misiles que esperaba recibir de los estadounidenses [64].

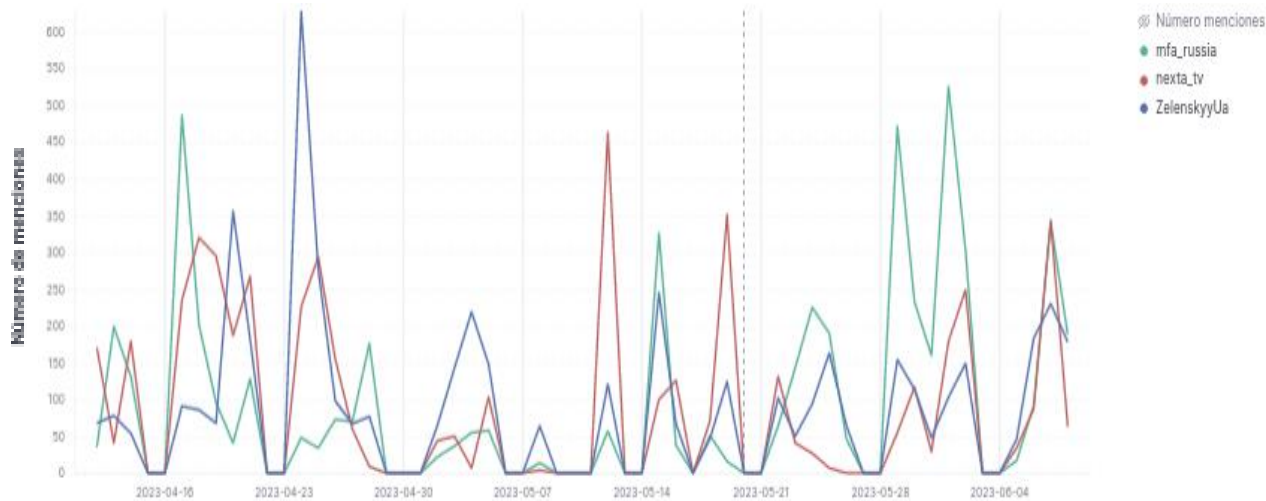


Figura 4-8 Comparativa de las tres menciones más comunes. (Autoría propia)

### 4.1.5 Anotaciones

Las anotaciones son referencias a entidades con nombre (personas, organizaciones, localizaciones, etc.) extraídas del texto del tuit. Mediante la herramienta *Aggregation based* se puede hacer un análisis de las anotaciones más detectadas mediante un mapa de calor, el cual divide por colores la densidad de datos de cada anotación. Esta visualización se complementa con la opción “*metric*” de TSVB, ya que permite comparar múltiples anotaciones, ofreciendo una visualización más clara de la cantidad exacta del número de veces que una palabra es detectada en el período de tiempo que se extrajeron los datos (véase Figura 4-9).

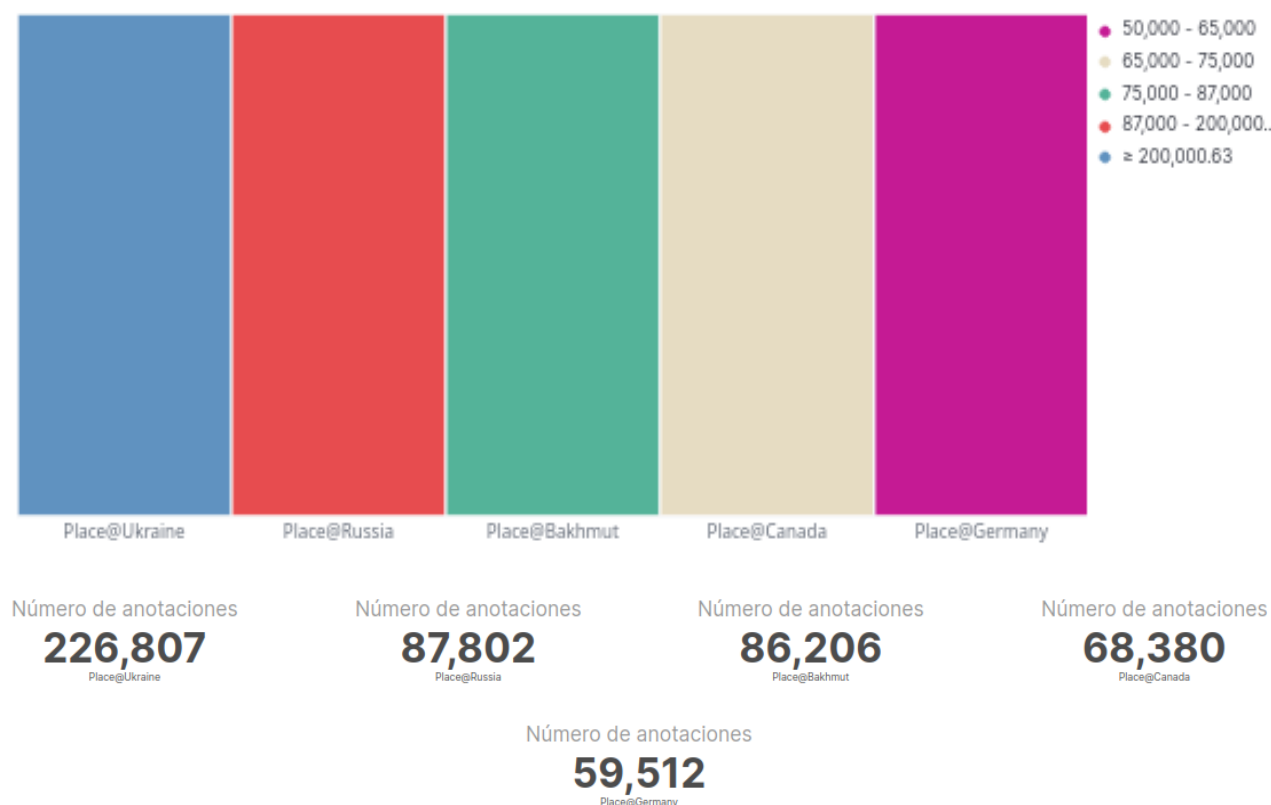


**Figura 4-9 Visualizaciones de las anotaciones más comunes. (Autoría propia)**

Se observa que la mayoría de estas anotaciones pertenecen a lugares, esto se debe a la gran repercusión que el conflicto de Ucrania y Rusia tiene sobre el mundo. Es por ello que, según cálculos realizados por el autor sobre el volumen de datos indexado en Elasticsearch, la cantidad de tuits que contienen anotaciones de lugares corresponde al 72% de los tuits totales extraídos.

Otra utilidad interesante es la de comparar dos anotaciones, por ejemplo, la comparación entre las localizaciones de Rusia y Ucrania. En la Figura 4-10 se presenta una gráfica, mediante TSVB, de la cantidad de veces que aparecen estos dos lugares en función del tiempo. Esta vez, al ser únicamente la comparativa entre a dos volúmenes de datos, la opción elegida permite crear una serie temporal y ofrece una clara comparativa entre la cantidad de veces que aparecen los lugares “Ucrania” y “Rusia” durante los meses en los que se extrajeron los tuits.

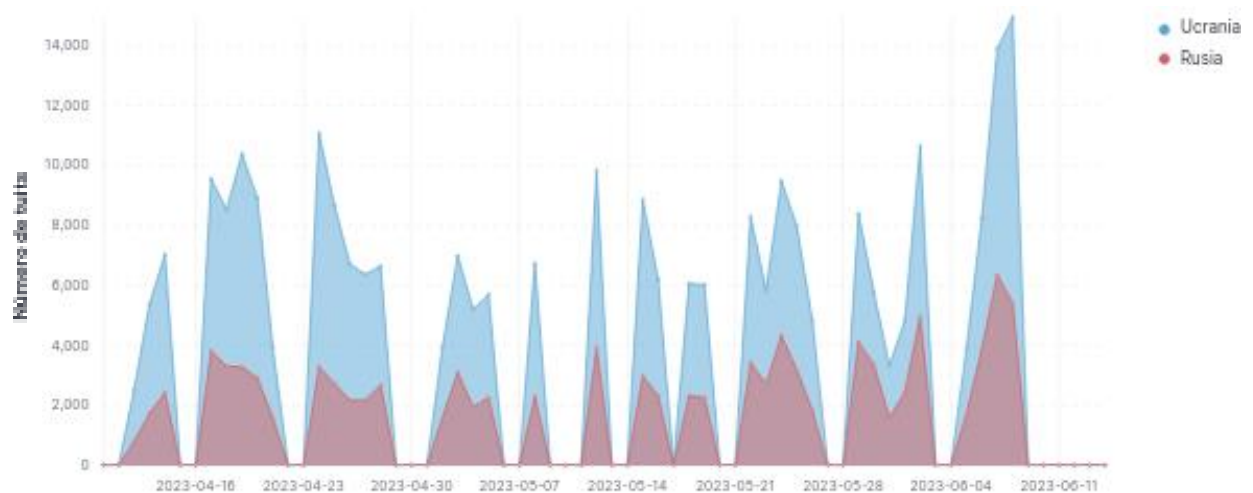


**Figura 4-10** Cantidad de veces que se menciona a Rusia y a Ucrania. (Autoría propia)

#### 4.1.6 Retuits

Los retuits son un indicativo del nivel de interés que suscita un tuit en la plataforma X/Twitter. Asimismo, la cantidad de retuits puede ser considerada como una métrica que refleja la difusión y popularidad de tuit en la comunidad de usuarios de la plataforma. Los datos en formato JSON extraídos indican un total de 1,555 retuits. Como se muestra en la Figura 4-11, el volumen de menciones en estos retuits se reparte equitativamente entre organizaciones, como la Corte Penal Internacional (@IntlCrimCourt) o la OTAN (@NATO), y figuras públicas como el presidente de los EE. UU. (@POTUS) y de Ucrania (@ZelenskyyUa). En cuanto a las etiquetas incluidas en los tuits más retuiteados, estas fueron mayoritariamente las que mencionaban a la nación ucraniana, con más del 50% del total de retuits. Del mismo modo que en la distribución previamente explicada, el porcentaje mostrado en las figuras representa el porcentaje relativo a la cantidad de tuits en los que se incluía alguna de las etiquetas que se muestran.

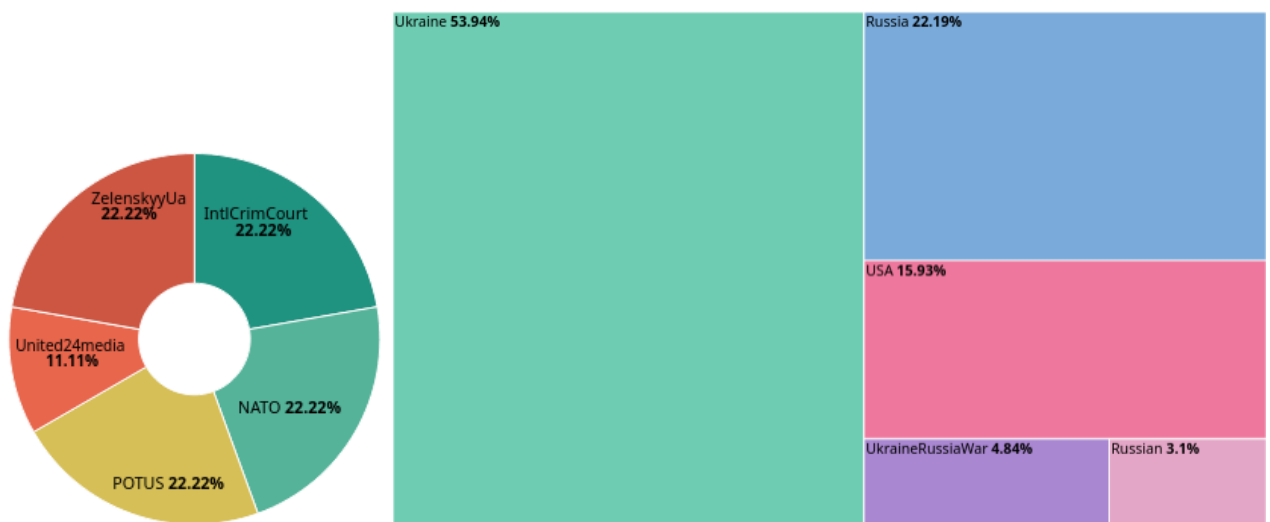


**Figura 4-11** Visualización de los parámetros más retuiteados. (Autoría propia)

Se observa en la Figura 4-12, la cual representa las entidades más populares según el número de retuits por fecha, que durante el mes de junio hay una visible dominancia de los retuits relacionados con la OTAN, en especial el día 2 de junio. Esta fecha coincide con la cumbre de la OTAN celebrada en Oslo [65], que consistió en una reunión informal entre ministros de Asuntos Exteriores de la Alianza, en dicha reunión se discutieron principalmente las claves de la continuidad del apoyo que los países miembros prestan a Ucrania. Otro tema relevante de esa reunión fue la solicitud de ingreso en la

Alianza por parte de Suecia y la por entonces reciente anexión de Finlandia. Estos países han sido los dos últimos en unirse a la OTAN desde la invasión de Rusia, el 4 de abril de 2023 [66] y el 7 de marzo de 2024 [67] respectivamente. La solicitud de Suecia fue rechazada por dos países miembros en 2022, Turquía y Hungría, pero finalmente ambos han ratificado la adhesión de Suecia a la OTAN este 2024.

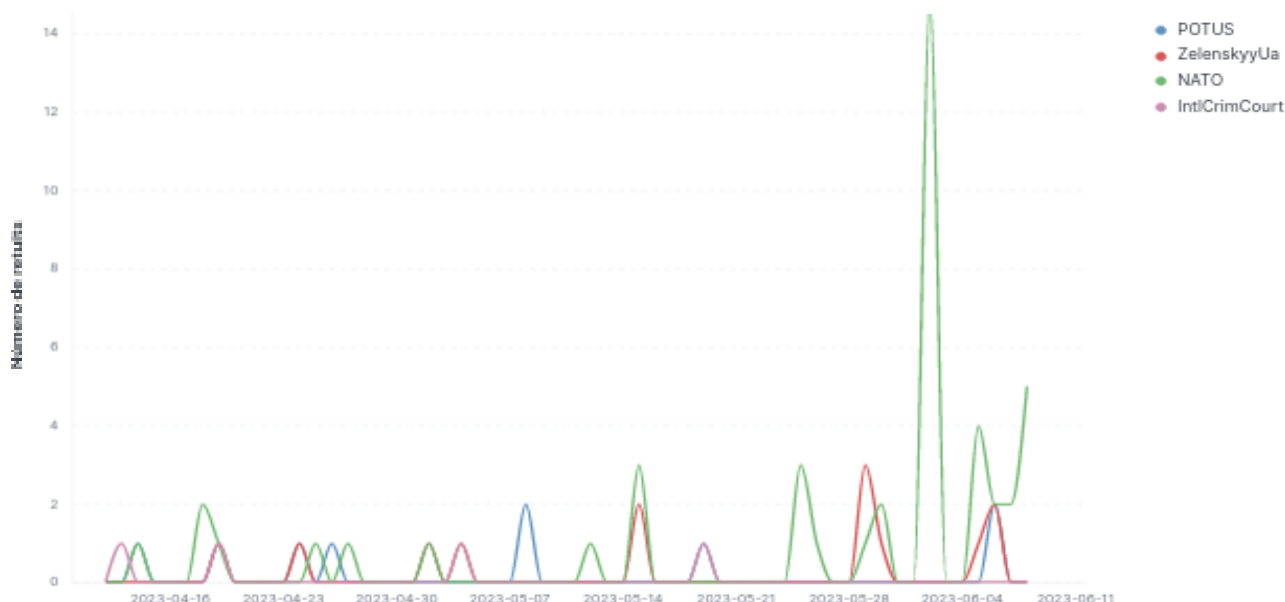


Figura 4-12 Cantidad de retuits por fecha. (Autoría propia)

Para alcanzar un análisis más detallado con respecto a aquellos tuits que suscitaron un mayor interés entre el público, en la Figura 4-13 se presenta un extracto de los tres tuits que más veces fueron retuiteados. En el contenido de estos tuits destacan figuras públicas como el secretario de Estado Antony Blinken, el cual habla sobre las Fuerzas Armadas rusas como una potencia mundial, o el embajador de Rusia en Reino Unido Adrei Kelin, el cual critica a los extranjeros que se alistaron para luchar en favor del ejército ucraniano. Desde que el conflicto estalló en 2022, Ucrania comenzó a reclutar personal extranjero para su frente, creando con este propósito una web para que aquellos voluntarios que quisiesen defender Ucrania de la invasión rusa pudieran alistarse en la legión extranjera ucraniana. Actualmente, un total de 20.000 extranjeros de 55 nacionalidades diferentes sirven en el ejército de Ucrania [68] y lo que denuncia Kelin en el mencionado tuit es la supuesta implicación de diferentes países del oeste de Europa para facilitar este reclutamiento.

Para la cumplimentación del análisis, en la Figura 4-14 se muestra la influencia de diferentes usuarios dependiendo de la cantidad de retuits que obtienen. El usuario más popular, con un total de 261 retuits y una media de 130,5 retuits, corresponde con el perfil de @\_bambkb. Este perfil corresponde a un activista que publica tuits sobre temas de actualidad [69]. Observando el autor de aquellos tuits que más retuits obtuvieron, se observa que el más popular (resaltado en amarillo) fue de este mismo autor, con un total de 259 retuits, el cual menciona a las naciones de Rusia, Ucrania y E.E.U.U. (véase Figura 4-13). Los perfiles restantes corresponden a una cuenta eliminada, @RussianEmbassy, @dahboo7 y @BillyJnr17, respectivamente. Estos dos últimos usuarios corresponden a un *streamer*<sup>8</sup> de noticias actuales (@dahboo7) y a un usuario que retuitea publicaciones sobre la actualidad global (@BillyJnr17).

<sup>8</sup> *Streamer*: anglicismo sin equivalente univocal en español, puede adaptarse usando expresiones pluriverbales, como, por ejemplo “realizador de transmisiones en directo” [70].

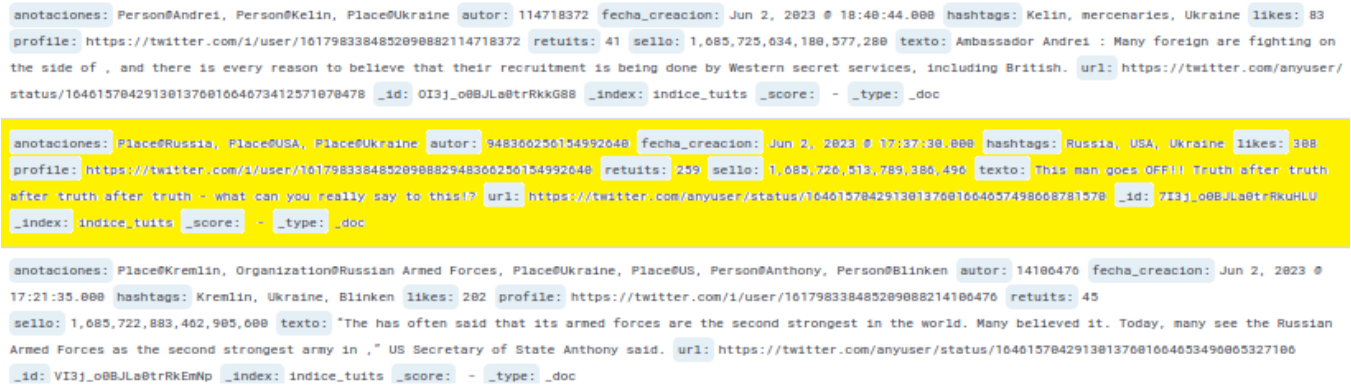


Figura 4-13 Top 3 tuits más populares según el número de retuits. (Autoría propia)

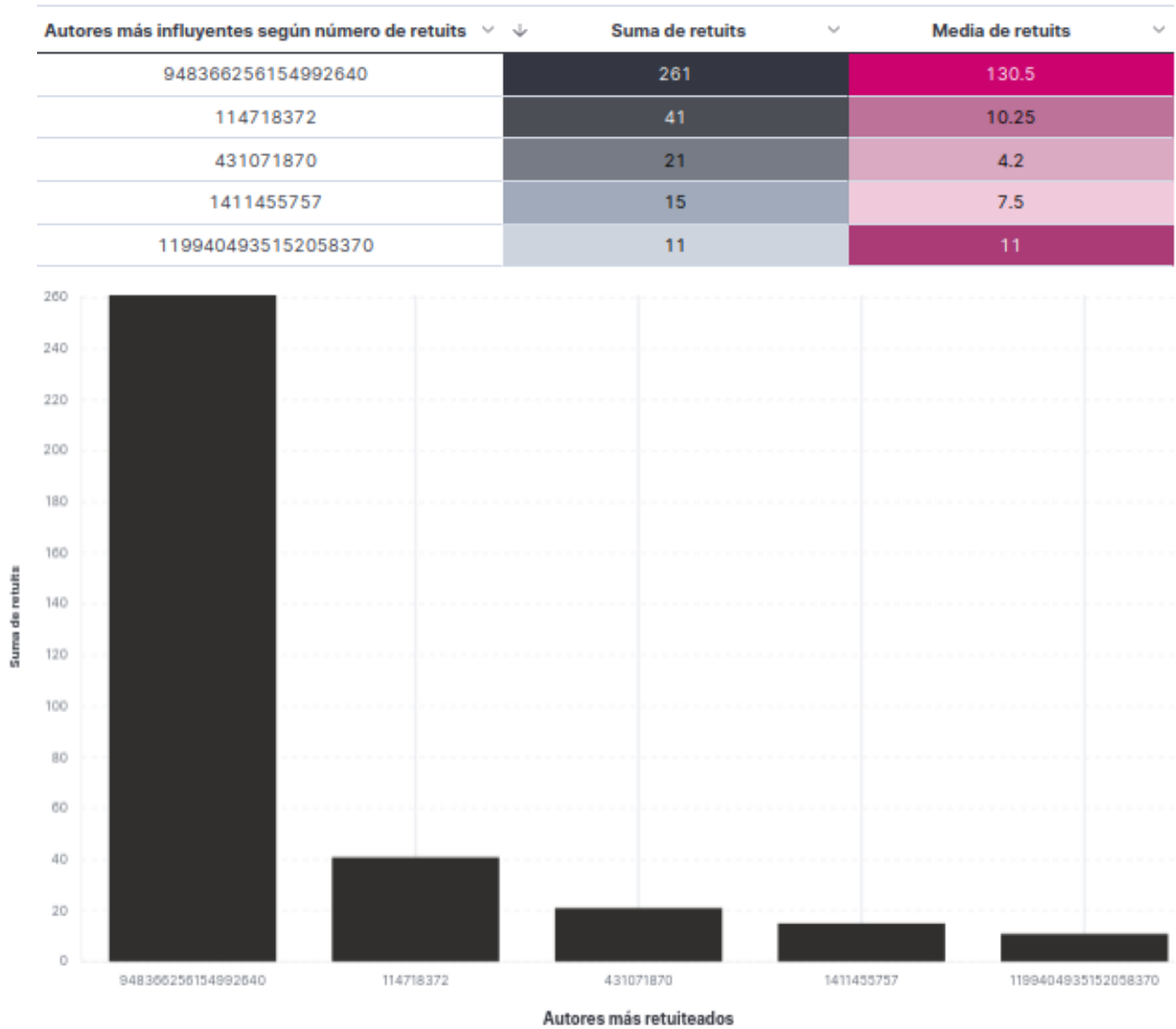


Figura 4-14 Autores más influyentes según el número total de retuits que reciben. (Autoría propia)

### 4.1.7 Me gusta (Likes)

El término *likes* permite reconocer una materia que recibe el apoyo o interés del público. Los “me gusta” son también indicativo de la popularidad de un tuit, de la misma manera que los retuits, aunque su forma de interacción es diferente a la de estos: los *likes* reflejan la apreciación personal del

contenido mientras que los retuits indican una acción más activa de compartir y difundir el contenido. Es interesante analizar la popularidad de diferentes índoles entre los usuarios activos de la red social; para ello se ha realizado un estudio de los parámetros previamente mencionados que más *likes* han recibido, con el objetivo de comprender la influencia que estos parámetros tienen.

En la Figura 4-15 se pueden observar 3 gráficas que presentan los parámetros más populares según el número de *likes*. La gráfica de la izquierda muestra las etiquetas que contenían los tuits con mayor popularidad, la siguiente consiste en una gráfica de los usuarios mencionados en los tuits que más “me gusta” obtuvieron, lo cual es un dato que no es necesariamente positivo para dichos usuarios, ya que estos tuits pueden llevar connotaciones negativas o críticas hacia ellos, finalmente la gráfica de la derecha expone las anotaciones encontradas en los tuits con mayor número de *likes*. En las presentes gráficas destaca la mención de personajes públicos como Zelensky, Putin o Blinken, al igual que la presencia de diferentes medios de comunicación como BBC World, Sky News o The Guardian.

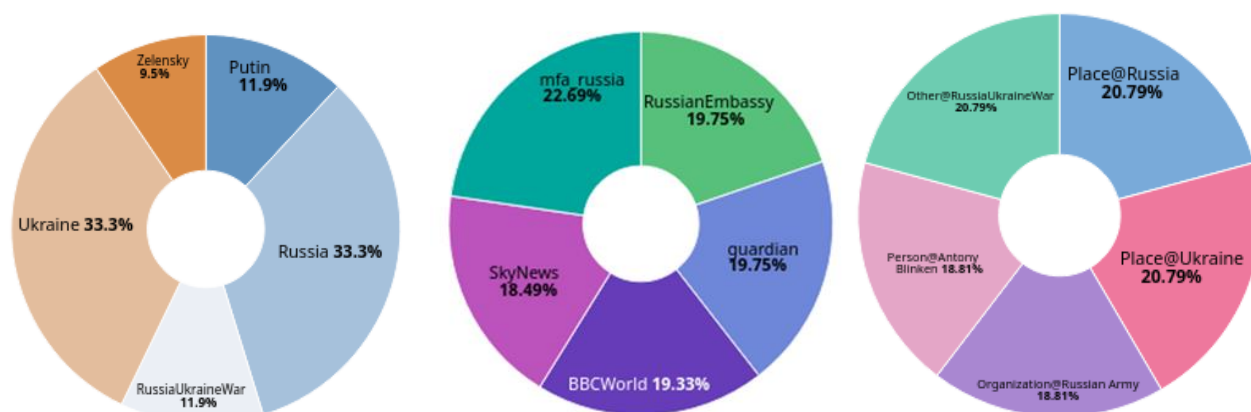


Figura 4-15 Popularidad según el número de *likes*. (Autoría propia)

Realizando un estudio más intensivo de la interactividad según el número de *likes*, se ha observado una mayor interactividad durante el día 2 de junio, fecha que coincidía con la cumbre de la OTAN previamente mencionada, en la que se discutieron temas relevantes del conflicto ruso-ucraniano.

Se puede observar en la Figura 4-17, la cual representa los 5 tuits que han obtenido más *likes*, que en ellos se menciona la cumbre que tuvo lugar en Noruega y el apoyo recibido por Ucrania, destacando sobre todo algunos países miembros como por ejemplo en el tuit, redactado por @GlasnostGone, que dice: “*President Zelensky said he’d received ‘powerful support’ from allies at Summit. Several countries, including UK, Denmark, Poland, Netherlands & Belgium said they’d help procure F-16s for & US endorsed F-16 training programme for pilots*”. Esta cuenta de X/Twitter pertenece a un escritor inglés llamado Alex King que desde la invasión rusa de Crimea en 2014 ha estado activo en la difusión de noticias sobre el conflicto, desde entonces muchas etiquetas de X/Twitter respecto a la guerra han ganado popularidad gracias a él [71]. También, entre los tuits más populares, se menciona al secretario de Estado Antony Blinken, el cual destacó en varios de ellos el fracaso de Rusia en la invasión de Ucrania. Según Blinken, la fallida estrategia de Rusia por ocupar Ucrania ha mostrado la debilidad del país, el cual se creía tener una de las Fuerzas Armadas más poderosas a nivel mundial, y esta debilidad ha tenido repercusiones desfavorables tanto en la influencia como en los intereses de Rusia durante los últimos años. Con la amenaza rusa, Putin ha conseguido que los países de occidente se unan más, lo cual ha mostrado la situación ventajosa que la OTAN posee frente a Rusia. Tal y como han mencionado figuras como el presidente estadounidense [72] o el Secretario General de la OTAN [73], la invasión de Ucrania ha sido un claro error ya que el efecto que ha acarreado es totalmente contrario al que Vladimir Putin buscaba, y ha provocado que todavía más países de Europa del Este quieran formar parte de la esfera que conforma la Alianza.

El tuit con más *likes*, con un total de 308 en el momento de la extracción, se presenta en la Figura 4-17 resaltado en amarillo, este tuit corresponde con un tuit publicado por el mencionado @\_bambkb. Este tuit no solo es el más popular en cuanto a número de “me gusta”, sino que también es el tuit extraído que más retuits obtuvo, con un total de 259 retuits, el cual se analizó en el previo apartado. Es por ello por lo que se puede concluir que ha sido el tuit más influyente y el que ha causado más interés de entre los tuits extraídos para la realización de este análisis. Para analizar el contexto de este tuit, se ha extraído el tuit directamente de X/Twitter (véase Figura 4-18). En este tuit se hace referencia a un discurso de Bantu Holomisa, miembro de la Asamblea Nacional de Sudáfrica, el cual critica a Estados Unidos por utilizar el conflicto de Ucrania para aprovecharse de la situación y debilitar a Rusia. Según dice este político, la verdadera causa del caos y la destrucción a nivel mundial es a causa de los Estados Unidos y de los países occidentales, los cuales, según pronuncia en su discurso, son la verdadera amenaza [74]. Desde el momento de su captura, el tuit ha ganado todavía más popularidad hasta los 3.400 retuits y 5.100 *likes* en el momento de escribir este texto.

En la Figura 4-16 se presenta una tabla resumen con los identificadores de los tuits con más *likes*, indicando a su vez el autor que ha publicado cada uno de ellos. Los siguientes tuits más populares corresponden a tuits publicados por @KiyvPost, @NatalkaKyiv, @anno1540 y al mencionado @GasnostGone, respectivamente. El primero de estos corresponde a un diario de noticias ucraniano [75], y los demás autores son *influencers*<sup>9</sup> que escriben sobre la guerra de Ucrania.

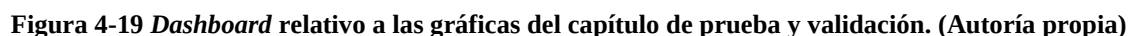
| ID Tuit                   | Autor del tuit      | Número de likes recibidos |
|---------------------------|---------------------|---------------------------|
| 1,685,726,500,000,000,000 | 948366256154992640  | 308                       |
| 1,685,722,800,000,000,000 | 14106476            | 202                       |
| 1,685,723,900,000,000,000 | 1507037811278553099 | 144                       |
| 1,685,726,100,000,000,000 | 2604194205          | 119                       |
| 1,685,722,900,000,000,000 | 2758100418          | 115                       |

Figura 4-16 Tabla resumen de los tuits más populares y su autor. (Autoría propia)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>anotaciones: Person@Zelensky, Place@Moldova, Place@UK, Place@Denmark, Place@Poland, Place@Netherlands, Place@Belgium, Place@Ukraine, Place@US autor: 2758100418<br/>         fecha_creacion: Jun 2, 2023 @ 17:45:29.000 hashtags: Moldova, Ukraine likes: 115 profile: https://twitter.com/1/user/16179833848520908822758100418 retuits: 19<br/>         sello: 1,685,722,916,363,471,488 texto: President Zelensky said he'd received 'powerful support' from allies at summit. Several countries, including UK, Denmark, Poland, Netherlands &amp; Belgium said they'd help procure F-16s for &amp; US endorsed F-16 training programme for pilots url: https://twitter.com/anyuser/status/164615784291381376016646558711231488 _id: jY3j_o8BJLa8trRkFWMP _index: indice_tuits _score: - _type: _doc</p> | <p>anotaciones: Place@Russia, Place@USA, Place@Ukraine autor: 948366256154992640 fecha_creacion: Jun 2, 2023 @ 17:37:30.000 hashtags: Russia, USA, Ukraine likes: 308<br/>         profile: https://twitter.com/1/user/16179833848520908822758100418 retuits: 259 sello: 1,685,726,513,789,380,490 texto: This man goes OFF!! Truth after truth after truth after truth - what can you really say to this? url: https://twitter.com/anyuser/status/16461578429138137601664657498668781578 _id: 713j_o8BJLa8trRkUHLU _index: indice_tuits _score: - _type: _doc</p>                                                                                                                                                                                |
| <p>anotaciones: Person@Blinken, Place@Ukraine, Place@Russia, Person@putin, Place@Russia autor: 2604194205 fecha_creacion: Jun 2, 2023 @ 17:29:22.000 hashtags: Russian, Ukraine likes: 119 profile: https://twitter.com/1/user/16179833848520908822604194205 retuits: 20 sello: 1,685,726,150,733,230,330 texto: Blinken said that the invasion of was its "strategic failure" that weakened Russia's influence and interests "for many years." putin tried to show strength but made Russia's weakness obvious to the world, what the Russian dictator sought to prevent, he... url: https://twitter.com/anyuser/status/16461578429138137601664655451210891265 _id: bo3j_o8BJLa8trRkqXGt _index: indice_tuits _score: - _type: _doc</p>                                                                      | <p>anotaciones: Place@Kremlin, Organization@Russian Armed Forces, Place@Ukraine, Place@US, Person@Anthony, Person@Blinken autor: 14106476 fecha_creacion: Jun 2, 2023 @ 17:21:35.000 hashtags: Kremlin, Ukraine, Blinken likes: 202 profile: https://twitter.com/1/user/1617983384852090882214106476 retuits: 45<br/>         sello: 1,685,722,883,462,905,000 texto: "The has often said that its armed forces are the second strongest in the world. Many believed it. Today, many see the Russian Armed Forces as the second strongest army in , " US Secretary of State Anthony said. url: https://twitter.com/anyuser/status/16461578429138137601664653496065327106 _id: VI3j_o8BJLa8trRkEmMp _index: indice_tuits _score: - _type: _doc</p> |
| <p>anotaciones: Person@Antony Blinken, Organization@Russian Army, Place@Russia, Other@RussianArmy, Place@Ukraine, Other@RussiaUkraineWar autor: 1507037811278553099<br/>         fecha_creacion: Jun 2, 2023 @ 17:18:46.000 hashtags: Russia, RussianArmy, Ukraine, RussiaUkraineWar likes: 144 profile: https://twitter.com/1/user/161798338485209088221507037811278553099 retuits: 28 sello: 1,685,723,939,073,341,440 texto: Antony Blinken talks about the Russian Army. url: https://twitter.com/anyuser/status/16461578429138137601664652787551827280 _id: Qo3j_o8BJLa8trRkRmgd _index: indice_tuits _score: - _type: _doc</p>                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

Figura 4-17 Top 5 tuits más populares según el número de “me gusta”. (Autoría propia)

<sup>9</sup> *Influencer*: anglicismo usado en referencia a una persona con capacidad para influir sobre otras, principalmente a través de las redes sociales [76].



## 5 CONCLUSIONES Y LÍNEAS FUTURAS

### 5.1 Conclusiones

La guerra de Ucrania ha dejado múltiples lecciones aprendidas, tanto en la sociedad como en las relaciones internacionales. Como consecuencia directa de este conflicto, los países miembros de la OTAN han reforzado su presupuesto militar y se ha incrementado la percepción de la importancia de las Fuerzas Armadas en la sociedad. Esta conciencia ha traído consigo una mayor implicación de las personas en documentar lo que está pasando en Ucrania, proliferando el uso de la red social X/Twitter como medio de transmisión de información.

Tras el desarrollo de este Trabajo Fin de Grado, se ha conseguido visualizar las características principales del presente conflicto en Ucrania y destacar numerosos eventos vinculados con el conflicto que han suscitado un destacado interés en las redes sociales a nivel mundial, gracias a la recolección de datos pertinentes, procedentes de la red social X/Twitter, entre los meses de abril y junio de 2023. Tras el análisis realizado, se demuestra nuevamente que las redes sociales se han convertido en un elemento valioso y una herramienta poderosa para el intercambio de información ya que poseen un alcance mundial. Los eventos que ocurren en el mundo tienen una repercusión social gracias a la interconexión que las redes sociales permiten a los ciudadanos, la cual posibilita el intercambio de información de manera inmediata. Y, lo que es más importante, la guerra de Ucrania ha demostrado un nuevo poder de las redes sociales, que ahora son capaces de reportar información verídica y en tiempo real, desde el frente de batalla, lo cual supone una novedad en cuanto a la transmisión de la información de carácter bélico.

Una vez presentados los resultados recogidos en el Capítulo 4, se concluye que el código ha sido exitoso y ha cumplido con todos los objetivos del epígrafe 1.2, llevando a cabo un exhaustivo análisis de los datos obtenidos a través de X/Twitter. Asimismo, tal y como indicaban los objetivos expuestos, se ha culminado la creación de la interfaz de tipo panel de control, en la que se permite la consulta de los datos recabados de una forma rápida y sencilla, posibilitando de igual manera la configuración del periodo de tiempo por parte del usuario con el fin de profundizar en el análisis deseado.

Además, el estudio de datos ha podido ser realizado gracias a la publicación de contenidos en la red social no sólo por parte de las cuentas oficiales de los medios de comunicación, sino también por *influencers*, los cuales, como se ha observado, pueden llegar a tener una mayor influencia sobre el público que los propios medios de información tradicionales. Esto es un claro indicio de la transformación significativa en el panorama informativo a nivel mundial, el cual ha sobrepasado la barrera geográfica y ha permitido una mayor diversidad de perspectivas acerca de un conflicto.

Gracias a esta diversidad de perspectivas, se han podido representar diferentes posturas por parte de los usuarios de los que se ha extraído la información, creando así un análisis completo y facilitando

la propia interpretación de las noticias, lo cual es muchas veces difícil debido a la postura con la que los diferentes medios de comunicación comparten esta información. Es por ello que se puede concluir que las redes sociales han facilitado el proceso de transferencia de información y el análisis de esta. No obstante, la difusión masiva de noticias falsas se ha convertido en una estrategia de propaganda que puede llegar a generar una desinformación a nivel mundial, es por ello la necesidad de corroborar esta información con múltiples perfiles que regularmente publiquen sobre temas de interés.

## 5.2 Líneas futuras

En lo que concierne a posibles mejoras para el almacenamiento y análisis de datos extraídos, se podrían explorar diferentes plataformas, como por Instagram, TikTok o Telegram, las cuales se han convertido en plataformas muy populares a nivel mundial [11], con el objetivo de captar más información y permitir el análisis de un mayor volumen de datos. Centrándose en el análisis, independientemente de la plataforma utilizada para la extracción de datos, estas mejoras se podrían resumir en los siguientes puntos:

- Monitorizar los datos en tiempo real, mediante diferentes técnicas de recolección de datos que permitan una actualización constante de la información, para así conseguir un análisis actualizado en lugar de un análisis centrado en un periodo de tiempo.
- Incluir, además de la información extraída en redes sociales, datos de otras fuentes, como encuestas o datos demográficos, para obtener una imagen más completa y precisa.
- Analizar no solo en contenido de los tuits, sino también las imágenes y enlaces adjuntos, los cuales pueden contener información de interés.
- Incorporar técnicas de NLP<sup>10</sup> para identificar el sentimiento de un texto y profundizar el análisis de opiniones y comentarios.

En resumen, futuros trabajos relacionados con el presente se podrían beneficiar de una captación de información a través de múltiples plataformas y una mejora tanto en las técnicas de análisis como en la monitorización de los datos de interés.

---

<sup>10</sup> NLP: del inglés “*Natural Language Processing*” o, en español, Procesamiento de Lenguaje Natural, es una forma de inteligencia artificial enfocada en las formas en que pueden interactuar las computadoras y las personas usando lenguaje humano [77].

## 6 BIBLIOGRAFÍA

- [1] «Home / X», X (formerly Twitter). Accedido: 21 de febrero de 2024. [En línea]. Disponible en: <https://twitter.com/home>
- [2] Jack Sheperd, «23 Essential Twitter (X) Statistics You Need to Know in 2024», The Social Shepherd. Accedido: 12 de enero de 2024. [En línea]. Disponible en: <https://thesocialshepherd.com/blog/twitter-statistics>
- [3] «Homepage», NATO. Accedido: 10 de marzo de 2024. [En línea]. Disponible en: <https://www.nato.int/cps/en/natohq/index.htm>
- [4] Y. Rodríguez, «Inteligencia de fuentes abiertas (OSINT): características, debilidades y engaño | GESI». Accedido: 16 de enero de 2024. [En línea]. Disponible en: <https://www.seguridadinternacional.es/?q=es/content/inteligencia-de-fuentes-abiertas-osint-caracter%C3%ADsticas-debilidades-y-enga%C3%B1o>
- [5] «Elasticsearch: Motor de búsqueda y analítica distribuido oficial», Elastic. Accedido: 21 de febrero de 2024. [En línea]. Disponible en: <https://www.elastic.co/es/elasticsearch>
- [6] «Kibana: Explora, visualiza y descubre datos», Elastic. Accedido: 21 de febrero de 2024. [En línea]. Disponible en: <https://www.elastic.co/es/kibana>
- [7] «Welcome to Python», Python. Accedido: 26 de febrero de 2024. [En línea]. Disponible en: <https://www.python.org/>
- [8] B. Bruseta Dupré, A. Escribá Esteve, G. López García, J. Maudos Villarroja y F. Pascual Lavilla, «Los medios de comunicación en la era digital». Accedido: 22 de enero de 2024. [En línea]. Disponible en: [https://www.fbbva.es/wp-content/uploads/2023/07/DE\\_2023\\_MediosComunicacion\\_Ivie.pdf](https://www.fbbva.es/wp-content/uploads/2023/07/DE_2023_MediosComunicacion_Ivie.pdf)
- [9] «Los orígenes de las redes sociales y los medios de comunicación», Deloitte. Accedido: 21 de febrero de 2024. [En línea]. Disponible en: [https://www2.deloitte.com/content/dam/Deloitte/mx/Documents/risk/Gobierno-Corporativo/GC\\_Boletin\\_invierno2014\\_tema2.pdf](https://www2.deloitte.com/content/dam/Deloitte/mx/Documents/risk/Gobierno-Corporativo/GC_Boletin_invierno2014_tema2.pdf)
- [10] «Digital 2024: Spain», DataReportal – Global Digital Insights. Accedido: 14 de marzo de 2024. [En línea]. Disponible en: <https://datareportal.com/reports/digital-2024-spain>
- [11] A. Thompson, «Digital 2024: 5 billion social media users», We Are Social. Accedido: 14 de marzo de 2024. [En línea]. Disponible en: <https://wearesocial.com/uk/blog/2024/01/digital-2024-5-billion-social-media-users/>
- [12] «“podcast”», Real Academia Española. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://www.rae.es/observatorio-de-palabras/podcast>
- [13] M. Bastero, «Historia de Twitter X: nacimiento y evolución de una red social», Marketing 4 Ecommerce - Tu revista de marketing online para e-commerce. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://marketing4ecommerce.net/historia-de-twitter/>

- 
- [14] J. A. Carballar Falcón, Twitter. Marketing Personal y Profesional. Editorial RC Libros, 2011, ISBN 9788493831226.
- [15] R. Jiménez Cano, «Twitter aumenta el límite a 280 caracteres para todos los usuarios», El País, Madrid, 8 de noviembre de 2017. Accedido: 30 de enero de 2024. [En línea]. Disponible en: [https://elpais.com/tecnologia/2017/11/07/actualidad/1510037858\\_189762.html](https://elpais.com/tecnologia/2017/11/07/actualidad/1510037858_189762.html)
- [16] R. Escuredo, «De Twitter a X: ¿Qué se esconde detrás de este rebranding?», Thinking for Innovation, ago. 2023, Accedido: 30 de enero de 2024. [En línea]. Disponible en: <https://www.iebschool.com/blog/rebranding-twitter-x-marketing-estrategico/>
- [17] «¿Cómo usar Twitter?: Términos de Twitter», GCFGGlobal. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://edu.gcfglobal.org/es/como-usar-twitter/terminos-de-twitter/1/>
- [18] «Introducción e Historia del OSINT», IUCPOL. Accedido: 14 de marzo de 2024. [En línea]. Disponible en: <https://iucpol.com/evolucion-osint-inteligencia-iucpol/>
- [19] S. García Rosado, «Conflictos | La inteligencia de fuentes abiertas (OSINT) en la guerra de Ucrania», Revista Ejércitos. Accedido: 5 de febrero de 2024. [En línea]. Disponible en: <https://www.revistaejercitos.com/conflictos/la-inteligencia-de-fuentes-abiertas-osint-en-la-guerra-de-ucrania/>
- [20] D. L. Chimbo Chillogalli, «Estudio de herramientas OSINT para ciberseguridad». Accedido: 5 de febrero de 2024. [En línea]. Disponible en: <https://dspace.ups.edu.ec/bitstream/123456789/26658/1/UPS-CT011062.pdf>
- [21] F. Tudela, «Crimea, Donbass y la guerra ruso-ucraniana. El efecto de los Protocolos de Minsk en la finlandización de Ucrania», Agenda Internacional, vol. 29, n.º 40, pp. 7-29, 2022.
- [22] F. Oropeza Fabián, «La guerra Rusia-Ucrania, considerada a partir de la teoría geopolítica del Heartland de Mackinder».
- [23] A. Bonet, «La Guerra Rusia-Ucrania y la crisis energética mundial | Club de Exportadores e Inversores Españoles». Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://clubexportadores.org/la-guerra-rusia-ucrania-y-la-crisis-energetica-mundial/>
- [24] L. Darder de la Poza y F. J. Calahorra Fernández, «Efectos de la Guerra de Ucrania en el sector de la alimentación para el ganado | Unidad de Divulgación Científica de la Facultad de Veterinaria». Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://www.ucm.es/udcvet/efectos-guerra-ucrania-ganado>
- [25] «Presupuesto del Ministerio de Defensa», Ministerio de Defensa. Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://www.defensa.gob.es/Galerias/presupuestos/presupuesto-MINISDEF-2023.pdf>
- [26] «Dashboard de Kibana», Elastic. Accedido: 12 de febrero de 2024. [En línea]. Disponible en: <https://www.elastic.co/es/kibana/kibana-dashboard>
- [27] D. Gutiérrez Blanco, «Qué es Kibana», OpenWebinars.net. Accedido: 31 de enero de 2024. [En línea]. Disponible en: <https://openwebinars.net/blog/que-es-kibana/>
- [28] «Las herramientas de visualización de datos más populares», Ministerio de Transformación digital. Accedido: 12 de febrero de 2024. [En línea]. Disponible en: <https://datos.gob.es/es/blog/las-herramientas-de-visualizacion-de-datos-mas-populares>
- [29] «Grafana: The open observability platform», Grafana Labs. Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <https://grafana.com/>
- [30] «Software de análisis e inteligencia de negocios», Tableau. Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <https://www.tableau.com/es-es>
- [31] «Power BI: visualización de datos», Microsoft Power Platform. Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <https://www.microsoft.com/es-es/power-platform/products/power-bi>
- [32] «Data Studio: creación de visualizaciones interactivas de datos», Google News Initiative. Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <https://newsinitiative.withgoogle.com/es-es/resources/trainings/data-studio-make-interactive-data-visualizations/>

- [33] A. Yigal, «Grafana vs. Kibana: The Key Differences to Know», Logz.io. Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://logz.io/blog/grafana-vs-kibana/>
- [34] R. Ruiz, «Diferentes herramientas de Business Intelligence», HobbieCode. Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://www.hobbiecode.com/diferentes-herramientas-de-business-intelligence/>
- [35] «Tableau Desktop and Grafana Integrations | Tableau Desktop and Grafana Connectors and Automation», Tray.io. Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://tray.io/connectors/tableau-desktop-grafana-integrations>
- [36] A. Hernández Ruiz, «Herramientas de OSINT para la obtención de información en Twitter sobre el conflicto Ucrania - Rusia». Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <http://calderon.cud.uvigo.es:8080/api/core/bitstreams/e699ad77-d094-48b0-9894-14b5a1264cb3/content>
- [37] «Centro Universitario de la Defensa – ENM – Sitio web del Centro Universitario de la Defensa, en la Escuela Naval Militar». Accedido: 13 de febrero de 2024. [En línea]. Disponible en: <https://cud.uvigo.es/>
- [38] M. Fernández Torrejón, «Sistema de detección inteligente de usuarios de Twitter interesados en la temática de defensa». Accedido: 22 de enero de 2024. [En línea]. Disponible en: <http://calderon.cud.uvigo.es:8080/api/core/bitstreams/3f6e7990-1a72-41fb-b7e7-3d003180c75b/content>
- [39] «ChatGPT». Accedido: 22 de febrero de 2024. [En línea]. Disponible en: <https://chat.openai.com>
- [40] J. Jiménez-Cervantes Pérez, «Plataforma OSINT para el almacenamiento y análisis inteligente de datos de la red social Twitter». Accedido: 29 de febrero de 2024. [En línea]. Disponible en: <http://calderon.cud.uvigo.es:8080/api/core/bitstreams/861a814a-a1c3-47c7-8331-ab784e274097/content>
- [41] «Inicio | Facultad de Comunicación». Accedido: 22 de febrero de 2024. [En línea]. Disponible en: <https://fcom.us.es/>
- [42] S. Pedregosa Gómez, «Análisis de las redes sociales: la guerra de Ucrania». Accedido: 29 de febrero de 2024. [En línea]. Disponible en: [https://idus.us.es/bitstream/handle/11441/136871/PER\\_PEDREGOSAGOMEZ\\_TFG.pdf](https://idus.us.es/bitstream/handle/11441/136871/PER_PEDREGOSAGOMEZ_TFG.pdf)
- [43] «Universidad de Valladolid». Accedido: 28 de febrero de 2024. [En línea]. Disponible en: <https://www.uva.es/export/sites/uva/>
- [44] N. Pérez Faúndez, «Análisis y visualización de datos abiertos de carácter informativo para el alumnado de la Universidad de Valladolid». Accedido: 28 de febrero de 2024. [En línea]. Disponible en: <https://uvadoc.uva.es/bitstream/handle/10324/50024/TFG-G5203.pdf>
- [45] «Grado en Ingeniería de Telecomunicaciones en Madrid - Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación - UPM», Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación. Accedido: 1 de marzo de 2024. [En línea]. Disponible en: <https://www.etsist.upm.es/>
- [46] I. Perrino Navas, «Investigación y extracción de datos con técnicas OSINT». Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://oa.upm.es/72302/>
- [47] «Oracle VM VirtualBox», VirtualBox. Accedido: 13 de febrero de 2024. [En línea]. Disponible en: <https://www.virtualbox.org/>
- [48] «Download Ubuntu Desktop | Download», Ubuntu. Accedido: 13 de febrero de 2024. [En línea]. Disponible en: <https://ubuntu.com/download/desktop>
- [49] «simplejson: Simple, fast, extensible JSON encoder/decoder for Python», PyPI. Accedido: 26 de febrero de 2024. [En línea]. Disponible en: <https://pypi.org/project/simplejson/>
- [50] «PyCharm: el IDE de Python para desarrolladores profesionales, por JetBrains», JetBrains. Accedido: 19 de febrero de 2024. [En línea]. Disponible en: <https://www.jetbrains.com/es-es/pycharm/>

- [51] «Otras versiones: PyCharm», JetBrains. Accedido: 17 de febrero de 2024. [En línea]. Disponible en: <https://www.jetbrains.com/es-es/pycharm/download/other.html>
- [52] «demoji: Accurately remove and replace emojis in text strings», PyPI. Accedido: 26 de febrero de 2024. [En línea]. Disponible en: <https://pypi.org/project/demoji/>
- [53] «“script”», Real Academia Española. Accedido: 14 de marzo de 2024. [En línea]. Disponible en: <https://www.rae.es/dpd/script>
- [54] «Guerra Ucrania - Rusia: resumen 12/05/2023», El País. Accedido: 7 de marzo de 2024. [En línea]. Disponible en: <https://elpais.com/internacional/2023-05-12/guerra-ucrania-rusia-ultimas-noticias-en-directo.html>
- [55] «Los rusos informan de un ataque con drones contra el buque de reconocimiento “Ivan Hurs”», Military. Accedido: 14 de marzo de 2024. [En línea]. Disponible en: <https://mil.in.ua/es/news/los-rusos-informan-de-un-ataque-con-drones-contra-el-buque-de-reconocimiento-ivan-hurs/>
- [56] «Guerra Ucrania Rusia: últimas noticias del 24 de mayo de 2023», The Huffington Post. Accedido: 7 de marzo de 2024. [En línea]. Disponible en: <https://www.huffingtonpost.es/global/guerra-ucrania-rusia-ultima-hora-directo-combates-bajmut.html>
- [57] G. Colchen, «Intercambio de prisioneros entre Ucrania y Rusia en el día de la Pascua ortodoxa», France 24. Accedido: 7 de marzo de 2024. [En línea]. Disponible en: <https://www.france24.com/es/europa/20230416-intercambio-de-prisioneros-entre-ucrania-y-rusia-en-el-d%C3%ADa-de-la-pascua-ortodoxa>
- [58] «Guerra Ucrania - Rusia: resumen 24/04/2023», El País. Accedido: 7 de marzo de 2024. [En línea]. Disponible en: <https://elpais.com/internacional/2023-04-24/guerra-ucrania-rusia-ultimas-noticias-en-directo.html>
- [59] R. González, «La inteligencia británica revela una nueva estrategia de Putin», Diario AS. Accedido: 7 de marzo de 2024. [En línea]. Disponible en: <https://as.com/actualidad/politica/la-inteligencia-britanica-revela-la-estrategia-de-putin-de-cara-a-las-elecciones-rusas-de-2024-n/>
- [60] «MFA Russia (@mfa\_russia) / X», X (formerly Twitter). Accedido: 8 de marzo de 2024. [En línea]. Disponible en: [https://twitter.com/mfa\\_russia](https://twitter.com/mfa_russia)
- [61] «NEXTA (@nexta\_tv) / X», X (formerly Twitter). Accedido: 8 de marzo de 2024. [En línea]. Disponible en: [https://twitter.com/nexta\\_tv](https://twitter.com/nexta_tv)
- [62] «Volodymyr Zelenskyy / Володимир Зеленський (@ZelenskyyUa) / X», X (formerly Twitter). Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://twitter.com/ZelenskyyUa>
- [63] «Ucrania y Estonia firman declaración pidiendo ingreso de Kiev en la OTAN», SWI (swissinfo.ch). Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://www.swissinfo.ch/spa/ucrania-y-estonia-firman-declaración-pidiendo-ingreso-de-kiev-en-la-otan/48459044>
- [64] M. Rey, D. Álvarez, C. Pérez, y N. G. Santiago, «Resumen de la Guerra entre Ucrania y Rusia el 1 de junio | EE.UU. enviará a Ucrania sistemas de lanzamientos de misiles», RTVE.es. Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://www.rtve.es/noticias/20220601/guerra-ucrania-directo-ultima-hora-noticias/2357281.shtml>
- [65] «Informal meeting of NATO Ministers of Foreign Affairs in Oslo, Norway», NATO. Accedido: 9 de marzo de 2024. [En línea]. Disponible en: [https://www.nato.int/cps/en/natohq/events\\_214929.htm](https://www.nato.int/cps/en/natohq/events_214929.htm)
- [66] «Finlandia ya es miembro de la OTAN», Euronews. Accedido: 9 de marzo de 2024. [En línea]. Disponible en: <https://es.euronews.com/my-europe/2023/04/04/finlandia-ya-es-miembro-de-la-otan-tras-un-proceso-accelerado>
- [67] «OTAN: Suecia entra oficialmente en la Alianza atlántica y acaba con décadas de neutralidad internacional», BBC News Mundo. Accedido: 9 de marzo de 2024. [En línea]. Disponible en: <https://www.bbc.com/mundo/articles/cekengd87llo>

- [68] V. Melkozerova, «Zelenskyy allows foreigners to serve in Ukraine's National Guard», Politico. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://www.politico.eu/article/volodymyr-zelenskyy-allow-foreigners-serve-national-guard-ukraine-war/>
- [69] «Posts with replies by (@bambkb) / X», X (formerly Twitter). Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://twitter.com/bambkb>
- [70] «streamer», Real Academia Española [@RAEinforma]. Accedido: 31 de marzo de 2024. [En línea]. Disponible en: <https://twitter.com/RAEinforma/status/1369954633133744131>
- [71] Y. Zoria, «Twitter storm makes #CrimeaIsUkraine trend again», Euromaidan Press. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://euromaidanpress.com/2020/08/01/twitter-storm-makes-crimeaisukraine-trend-again/>
- [72] M. Reynolds, «Biden advierte a Putin que la OTAN está “más unida” tras invasión de Ucrania», Los Angeles Times. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://www.latimes.com/espanol/eeuu/articulo/2022-02-26/biden-advierte-a-putin-que-la-otan-esta-mas-unida-tras-invasion-de-ucrania>
- [73] A. Taylor, «Analysis | Did Putin inadvertently create a stronger NATO?», Washington Post, 30 de junio de 2022. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://www.washingtonpost.com/world/2022/06/30/putin-nato-sweden-finland-stronger/>
- [74] «@bambkb on X: “This man goes OFF!! Truth after truth after truth after truth - what can you really say to this!? #Russia #USA #Ukraine <https://t.co/ED6WcWItb0>” / X», X (formerly Twitter). Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://twitter.com/bambkb/status/1664657498668781570>
- [75] «Get the Latest Ukraine News Today - KyivPost», Kyiv Post. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://www.kyivpost.com>
- [76] «“influencer”», Real Academia Española. Accedido: 18 de marzo de 2024. [En línea]. Disponible en: <https://www.rae.es/observatorio-de-palabras/influencer>
- [77] «¿Qué es el procesamiento de lenguaje natural (NLP)? | Una guía completa del NLP», Elastic. Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://www.elastic.co/es/what-is/natural-language-processing>

## ANEXO I: IMPLICACIONES SOCIALES Y AMBIENTALES

El empleo de las redes sociales como medio de comunicación implica una serie de consecuencias, las cuales es crucial analizar debido a su impacto en la sociedad. Las redes sociales han alterado radicalmente las relaciones entre las personas, tanto a nivel profesional como social. Este anexo pretende explorar las implicaciones mencionadas, examinando su alcance y consecuencias para así comprender el papel de las redes sociales en la sociedad contemporánea.

Socialmente, tales plataformas han transformado la manera en que los individuos interactúan, facilitando una conexión instantánea y global entre personas de diversas geografías. No obstante, este fenómeno también ha generado problemas como la propagación de desinformación, la polarización de opiniones y un incremento en el uso de las redes sociales, lo cual puede acarrear posibles repercusiones negativas con respecto a la privacidad del usuario ya que el beneficio que generan estas corporaciones se basa en la gestión de *cookies* y recolección de información concerniente a la actividad de los usuarios en línea, lo cual suscita inquietudes fundamentales sobre la privacidad de datos.

Centrándose más en las implicaciones derivadas del presente trabajo, las herramientas utilizadas tienen el potencial de integrar información de distintas fuentes como los medios de comunicación o la información publicada por *influencers* a través de las redes sociales. Esta integración facilita al público interesado el acceso a información sobre el conflicto de Ucrania y la propagación de noticias sobre dicho ámbito bélico, noticias que no necesariamente han de ser verídicas, sino que son objeto de análisis por parte del usuario que tiene acceso a ellas.

Finalmente, en el ámbito ambiental, la monitorización de las redes sociales supone grandes costos computacionales, ya que requiere servidores para sostener dichas plataformas además de un significativo consumo de energía. El consumo energético asociado a la infraestructura digital no solo posee una implicación sobre el cambio climático, sino que también plantea interrogantes éticos sobre la priorización de recursos ante el impacto ambiental asociado al funcionamiento continuo de estos servidores. Es aquí donde surge la disyuntiva entre la información de la actualidad acerca del conflicto de Ucrania y las repercusiones que un excesivo uso de las herramientas de captación puede tener sobre el medio ambiente.

En resumen, si bien las redes sociales han revolucionado la comunicación y la interacción humana, la guerra de Ucrania ha tenido varias implicaciones significativas en su uso, aumentando el intercambio de información a través de estas plataformas y, por consiguiente, facilitando el acceso a un amplio volumen de datos con el objetivo de ser analizados. Las implicaciones de las redes sociales tienen un papel trascendental en la sociedad, tanto en términos de difusión de información como en la desinformación y la propagación del activismo en línea. Es asimismo esencial reflexionar sobre implicaciones de las redes sociales, con miras a garantizar un uso responsable y sostenible. Es imperativo también que las políticas de privacidad sean transparentes y que los usuarios estén debidamente informados acerca del uso y protección de sus datos para evitar posibles usurpaciones de identidad.

## ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

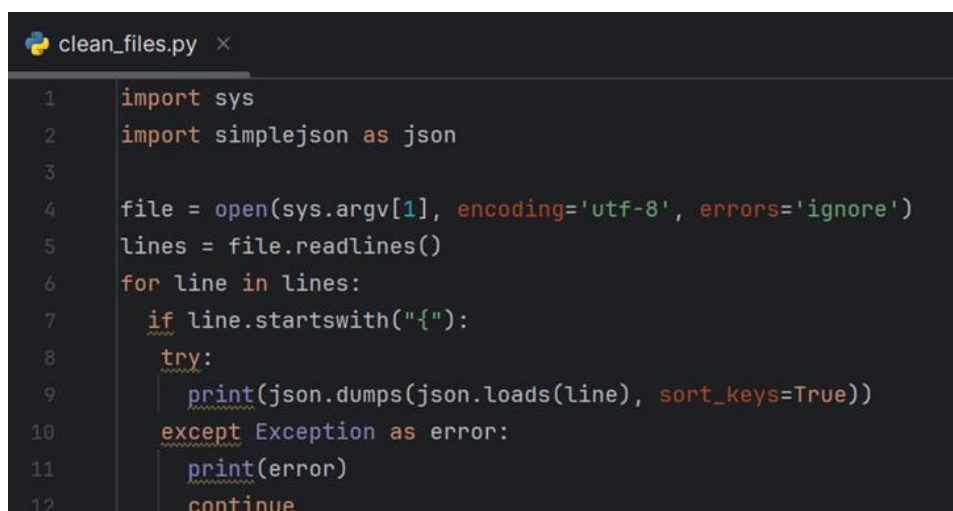
El uso de *Open Source Intelligence* (OSINT) para el análisis y visualización de datos en el contexto de un conflicto como el de Ucrania y Rusia plantea reflexiones que trascienden más allá de la tecnología y se adentran en la responsabilidad humana. Por un lado, la capacidad de recopilar y analizar información proveniente de fuentes abiertas ofrece un potencial sin precedentes para comprender y abordar conflictos de manera más informada, rápida y efectiva. Sin embargo, este poder también conlleva una serie de consideraciones éticas, como la protección de la privacidad acerca de información del usuario y la seguridad de los individuos involucrados en el conflicto. Otro factor de importancia sobre el que reflexionar durante el uso de OSINT es la veracidad y la objetividad en la interpretación de los datos recopilados, ya que cualquier individuo puede publicar información falsa acerca de un conflicto. De este modo, el uso ético de OSINT requiere un compromiso con los principios de transparencia, responsabilidad y respeto a los derechos humanos, así como una reflexión profunda sobre el papel de la tecnología en la construcción de un mundo más justo y pacífico. A pesar de estos interrogantes, la información obtenida gracias a OSINT puede ser de gran ayuda para la información de la sociedad y conocer día a día los principales asuntos de interés a nivel mundial.

## ANEXO III: EXPLICACIÓN DE LOS CÓDIGOS REALIZADOS

El desarrollo de un análisis exhaustivo de datos extraídos de fuentes de información abiertas como son las redes sociales requiere la elaboración de diversos archivos de código que permitan extraer la información de interés además de almacenar esta en una plataforma para permitir su visualización y análisis detallado, como el que se busca en este trabajo de fin de grado. En concreto para la elaboración de este proyecto, ha sido necesaria la elaboración de cuatro archivos de código que, con sus diferentes utilidades, han permitido realizar un tratamiento sobre los datos extraídos para posibilitar el deseado análisis. De este modo, es oportuno la definición de cuatro subapartados para explicar cada código por separado.

- a) En primer lugar, se destaca el código empleado para limpiar los archivos y eliminar la información que no es de interés (véase Figura III - 1). A continuación, se explicará dicho código línea por línea:
  - Líneas 1-2: Se importan los módulos de las librerías necesarias para escribir el código.
  - Línea 4: Abre un archivo cuyo nombre se ha de pasar como argumento en el terminal y lo asigna a la variable “file”. Se especifica una codificación ‘UTF-8’ y se ignoran los errores.
  - Línea 5: Lee las líneas del archivo y las asigna a la variable “lines”.
  - Línea 6: Itera sobre cada línea en la variable previamente creada.
  - Línea 7: Este condicional dará paso a la implementación del código para extraer los datos cuando se detecte el símbolo “{“, lo que sugiere que podría contener datos en formato JSON.
  - Línea 8: Inicia un bloque de código que puede generar una excepción.
  - Línea 9: Intenta cargar el contenido de la línea como un objeto JSON y luego lo vuelve a serializar en formato JSON con json.dumps() y con las claves ordenadas mediante sort\_keys=True. Esto es útil para imprimir el contenido JSON de forma legible y ordenada. Si la línea no contiene datos JSON válidos, se generará una excepción.
  - Línea 10: Captura cualquier excepción que ocurra durante el intento de carga y serialización de JSON.
  - Línea 11: En caso de que se produzca un error, se imprimirá dicho mensaje.
  - Línea 12: La última línea del código permite continuar con la siguiente iteración del bucle ‘for’ sin ejecutar el resto del código en el bucle actual.

A modo de resumen, el código analizado lee un archivo de texto e intenta interpretar cada línea como un objeto JSON y en caso de éxito, imprime el JSON formateado, ignorando las líneas que no contienen datos con formato JSON.



```

1  import sys
2  import simplejson as json
3
4  file = open(sys.argv[1], encoding='utf-8', errors='ignore')
5  lines = file.readlines()
6  for line in lines:
7      if line.startswith("{"):
8          try:
9              print(json.dumps(json.loads(line), sort_keys=True))
10         except Exception as error:
11             print(error)
12         continue
  
```

Figura III - 1 Código para la limpieza de archivos. (Autoría propia)

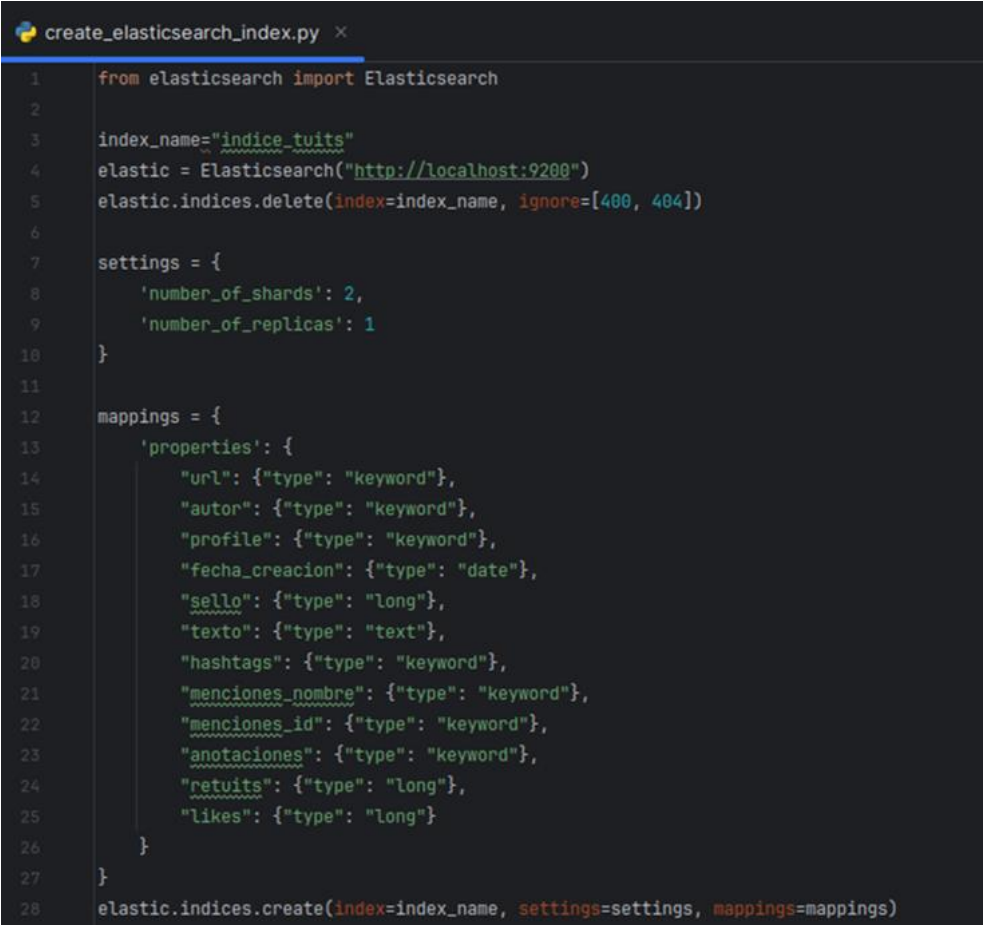
- b) En segundo lugar, se explicará el código empleado para la extracción de los datos de interés (véase Figura III - 2):
- Líneas 1-4: Se importan las librerías necesarias para el funcionamiento del código.
  - Línea 6: Define una función llamada “clean\_tweet\_text”.
  - Línea 7: Se eliminan los enlaces de la cadena de texto y los reemplaza con una cadena vacía. La expresión “https?://\S+\s\*” coincide con cualquier subcadena que comience con “http://” o “https://” seguido de uno o más caracteres.
  - Línea 8: Elimina las menciones de la cadena de texto “cleaned\_tweet” y los reemplaza con una cadena vacía. La expresión “@\w+\s\*” coincide con “@” seguido de uno o más alfanuméricos.
  - Línea 9: Elimina las etiquetas de la cadena de texto “cleaned\_tweet” y los reemplaza con una cadena vacía. La expresión “@\w” coincide con “@” seguido de uno o más alfanuméricos.
  - Línea 10: Elimina los emojis en la cadena de texto “cleaned\_tweet”. De esta manera, gracias a la librería “demoji”, se limpia el texto de emoticonos que puedan afectar a su lectura.
  - Línea 11: Devuelve la cadena de texto tras terminar su limpieza.
  - Línea 13: Abre un archivo con el nombre especificado en el primer argumento de la línea de comandos y lo asigna a la variable “file”. La codificación se establece en UTF-8 y los errores se ignoran.
  - Línea 14: Lee las líneas del archivo y las asigna a la variable “lines”.
  - Línea 15: Itera sobre cada línea en la variable previamente creada.
  - Línea 16: Este condicional dará paso a la implementación del código para extraer los datos cuando se detecte el símbolo ‘{’, el cual es el símbolo por el que empieza cada tuit representado en formato JSON.
  - Línea 17: Crea un diccionario vacío en el que se irán almacenando los valores que posteriormente se indiquen.
  - Línea 18: Convierte la línea en un diccionario de Python llamado “data\_dict”, utilizando la función “loads” de la biblioteca JSON.
  - Líneas 20-62: A lo largo de estas líneas, se extraen los valores que se quieran imprimir en el diccionario de salida. El mecanismo de extracción se realiza rellenando el diccionario de salida cuando detecta los parámetros de interés dentro de “data\_dict”.
  - Línea 64: Convertimos a JSON el diccionario Python de salida y lo imprimimos en pantalla.
- c) El siguiente código a explicar es el que crea un índice en la plataforma Elasticsearch (véase Figura III - 3). A continuación, se describe el código:
- Línea 1: El comando importa el módulo de Elasticsearch, proporcionando su interfaz.
  - Línea 3: Se crea una variable indicando el nombre del índice “índice\_tuis”.
  - Línea 4: Se crea un objeto denominado “elastic” que conecta con el puerto local del servidor de Elasticsearch (9200).
  - Línea 5: La siguiente línea de código elimina un posible índice ya cargado en Elasticsearch con el mismo nombre.
  - Línea 7: Se definen los ajustes para el índice, estableciendo el número de particiones (*shards*) a 2 y el de réplicas a 1.
  - Línea 12: Se define el *mapping* del índice de Elasticsearch, especificando los campos de los datos que se suban a este índice, así como su tipo de datos.
  - Línea 28: La última línea del código crea en el índice mencionado la estructura indicada por los *mappings*.

```

extract_data2.py ×
1  import simplejson as json
2  import re
3  import demoji
4  import sys
5
6  1 usage
7  def clean_tweet_text(tweet_text):
8      cleaned_tweet = re.sub(pattern=r'https?://\S+\s*', repl: '', tweet_text) # Eliminar los enlaces
9      cleaned_tweet = re.sub(pattern=r'@\w+\s*', repl: '', cleaned_tweet) # Eliminar las menciones
10     cleaned_tweet = re.sub(pattern=r'#(\w+)', repl: '', cleaned_tweet) # Eliminar los hashtags
11     cleaned_tweet = demoji.replace(cleaned_tweet, repl: '')
12     return(cleaned_tweet)
13
14     file = open(sys.argv[1], encoding='utf-8', errors='ignore')
15     lines = file.readlines()
16     for line in lines:
17         if line.startswith("{"):
18             output_dict = dict()
19             data_dict = json.loads(line)
20
21             if "author_id" in data_dict:
22                 output_dict["autor"] = data_dict["author_id"]
23                 output_dict["profile"] = "https://twitter.com/i/user/1617983384852898882" + data_dict["author_id"]
24
25             if "id" in data_dict:
26                 output_dict["url"] = "https://twitter.com/anyuser/status/1646157842913813768" + data_dict["id"]
27
28             if "text" in data_dict:
29                 output_dict["texto"] = clean_tweet_text(data_dict["text"])
30
31             if "created_at" in data_dict:
32                 output_dict["fecha_creacion"] = data_dict["created_at"]
33
34             if "entities" in data_dict:
35                 if "mentions" in data_dict["entities"]:
36                     menciones=data_dict["entities"]["mentions"]
37                     output_dict["menciones_nombre"] = []
38                     output_dict["menciones_id"] = []
39                     for mencion in menciones:
40                         output_dict["menciones_nombre"].append(mencion["username"])
41                         output_dict["menciones_id"].append(mencion["id"])
42
43                 if "annotations" in data_dict["entities"]:
44                     anotaciones=data_dict["entities"]["annotations"]
45                     output_dict["anotaciones"] = []
46                     for anotacion in anotaciones:
47                         output_dict["anotaciones"].append(anotacion["type"] + "@" + anotacion["normalized_text"])
48
49                 if "hashtags" in data_dict["entities"]:
50                     etiquetas=data_dict["entities"]["hashtags"]
51                     output_dict["hashtags"] = []
52                     for etiqueta in etiquetas:
53                         output_dict["hashtags"].append(etiqueta["tag"])
54
55             if "public_metrics" in data_dict:
56                 if "like_count" in data_dict["public_metrics"]:
57                     output_dict["likes"] = data_dict["public_metrics"]["like_count"]
58
59                 if "retweet_count" in data_dict["public_metrics"]:
60                     output_dict["retuits"] = data_dict["public_metrics"]["retweet_count"]
61
62             if "local_time_ns" in data_dict:
63                 output_dict["sello"] = data_dict["local_time_ns"]
64
65             print(json.dumps(output_dict))

```

Figura III - 2 Código para la extracción de datos de interés. (Autoría propia)



```

1  from elasticsearch import Elasticsearch
2
3  index_name="indice_tuits"
4  elastic = Elasticsearch("http://localhost:9200")
5  elastic.indices.delete(index=index_name, ignore=[400, 404])
6
7  settings = {
8      'number_of_shards': 2,
9      'number_of_replicas': 1
10 }
11
12 mappings = {
13     'properties': {
14         "url": {"type": "keyword"},
15         "autor": {"type": "keyword"},
16         "profile": {"type": "keyword"},
17         "fecha_creacion": {"type": "date"},
18         "sello": {"type": "long"},
19         "texto": {"type": "text"},
20         "hashtags": {"type": "keyword"},
21         "menciones_nombre": {"type": "keyword"},
22         "menciones_id": {"type": "keyword"},
23         "anotaciones": {"type": "keyword"},
24         "retuits": {"type": "long"},
25         "likes": {"type": "long"}
26     }
27 }
28 elastic.indices.create(index=index_name, settings=settings, mappings=mappings)

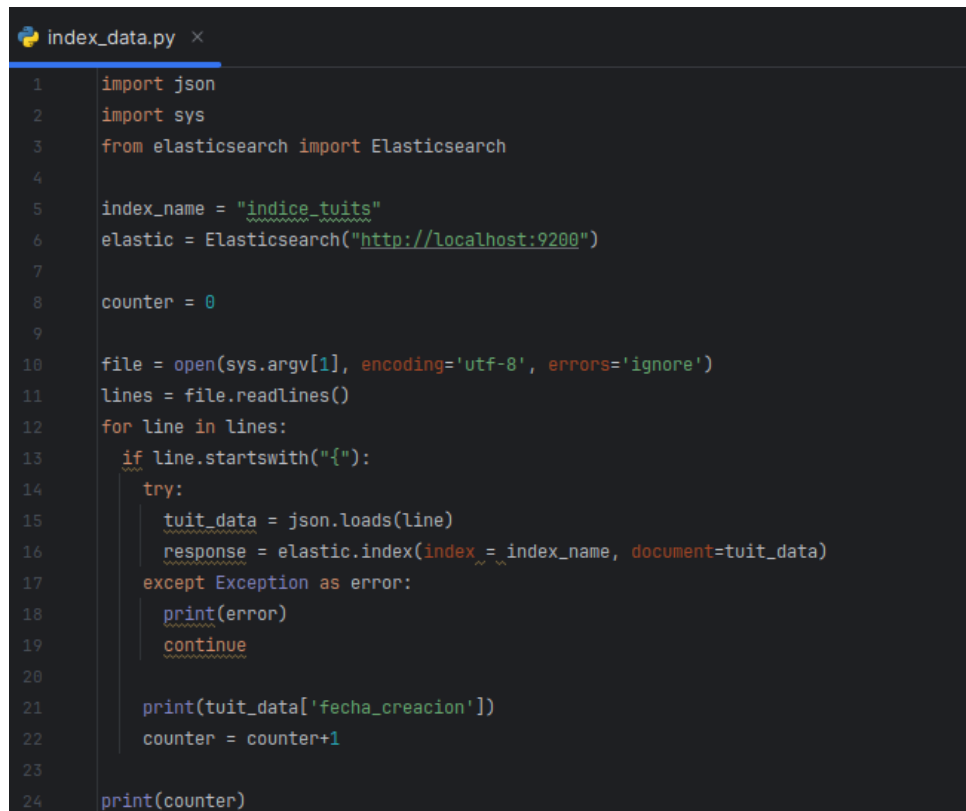
```

Figura III - 3 Código para la creación de un índice en Elasticsearch. (Autoría propia)

- d) El último código que se ha realizado para este trabajo de fin de grado carga los archivos de datos JSON en Elasticsearch y los indexa en el índice previamente creado en la plataforma (véase Figura III - 4); a continuación, se describe dicho código:
- Líneas 1-3: se importan las librerías de Python necesarias para el funcionamiento del código.
  - Línea 5: Se indica el nombre del índice que se quiere utilizar.
  - Línea 6: Se inicializa la conexión con Elasticsearch, indicando el *host* y puerto del servidor.
  - Línea 8: Se crea un contador y se inicializa a 0, el cual que contará los tuits que se hayan indexado en Elasticsearch.
  - Línea 10: Abre un archivo con el nombre especificado en el primer argumento de la línea de comandos y lo asigna a la variable "file". La codificación se establece en UTF-8 y los errores se ignoran. En esta línea de texto procesará el fichero de datos.
  - Línea 11: Lee las líneas del archivo y las asigna a la variable "lines".
  - Línea 12: Itera sobre cada línea en la variable previamente creada.
  - Línea 13: Este condicional dará paso a la implementación del código para extraer los datos cuando se detecte el símbolo "{", el cual es el símbolo por el que empieza cada tuit representado en formato JSON.
  - Línea 14: el comando 'try' inicia un bloque de código que puede generar una excepción.
  - Línea 15: Carga el contenido de la línea como un objeto JSON y lo guarda en la variable "tuit\_data".
  - Línea 16: Indexa los datos del tuit en Elasticsearch, dentro del índice especificado.
  - Línea 17: Captura cualquier excepción que ocurra durante el proceso de carga o indexación.
  - Línea 18: Imprime el mensaje de error en caso de que se genere una excepción durante el proceso.

- Línea 19: Permite continuar con la siguiente iteración del bucle for sin ejecutar el resto del código en el bucle actual.
- Línea 21: Imprime la fecha de creación del tuit, asumiendo que hay un campo llamado “fecha\_creacion” en los datos del tuit.
- Línea 22: Incrementa el contador de tuits indexados.
- Línea 24: Imprime el total de tuits indexados al final del proceso.

En resumen, este código Python lee un archivo de texto e indexa esos datos en Elasticsearch al mismo tiempo que cuenta cuántos tuits fueron indexados correctamente.



```
1 import json
2 import sys
3 from elasticsearch import Elasticsearch
4
5 index_name = "indice_tuits"
6 elastic = Elasticsearch("http://localhost:9200")
7
8 counter = 0
9
10 file = open(sys.argv[1], encoding='utf-8', errors='ignore')
11 lines = file.readlines()
12 for line in lines:
13     if line.startswith("{"):
14         try:
15             tuit_data = json.loads(line)
16             response = elastic.index(index=index_name, document=tuit_data)
17         except Exception as error:
18             print(error)
19             continue
20
21     print(tuit_data['fecha_creacion'])
22     counter = counter+1
23
24 print(counter)
```

**Figura III - 4** Código para la indexación de archivos en el índice. (Autoría propia)