

**UNIVERSIDAD EUROPEA MIGUEL
DE CERVANTES**

ESCUELA POLITECNICA SUPERIOR

**TITULACIÓN:
MÁSTER EN GESTIÓN Y ANÁLISIS DE
GRANDES VOLÚMENES DE DATOS: BIG
DATA**



TRABAJO FIN DE MÁSTER

**ANÁLISIS DE LOS
COMENTARIOS EN
YOUTUBE CON RELACIÓN A
LA GUERRA EN UCRANIA
EN EL AÑO 2024**

AUTOR

VLADYSLAV DODONOV

TUTOR

Fernando Alonso Pastor

VALLADOLID, 19 DE MAYO DE 2025

Contenido

1.	Objetivos del trabajo	3
2.	Análisis de la situación	4
2.1	Descripción del conflicto armado 2022-2025	4
2.2	Historia general del conflicto (2022-2025)	4
2.3	Eventos clave de 2024: batallas, acuerdos y cambios territoriales.....	7
2.4	Cobertura mediática, desinformación y opinión pública.....	9
2.5	Identificación de los Canales relevantes en España.	10
3.	Obtención, procesado y almacenamiento de los datos	11
3.1	Diseño de Proyecto	11
3.2	Obtención de información desde la API de YouTube	14
3.3	Limitaciones y observaciones	14
3.4	Almacenamiento de los datos	15
3.5	GitHub - versionamiento y documentación	15
4.	Análisis exploratorio	15
5.	Diseño e implementación de los modelos o técnicas necesarios	17
6.	Análisis de los resultados obtenidos	18
7.	Conclusiones y planes de mejora	18
8.	Bibliografía	19
9.	Anexo con el código fuente desarrollado.	21

1. Objetivos del trabajo

El objetivo de este trabajo consiste en analizar, aplicando métodos propios de ciencia de datos, los comentarios extraídos de los videos de Youtube y que tienen relación con la Guerra de Ucrania - Rusia. En particular se toman los videos publicados en el año 2024 y en español. En particular se hace énfasis en youtubers españoles y que cuentan con una base de seguidores en el país. Lo que se quiere analizar es una muestra de los comentarios que aparecen en los canales preseleccionados, aplicando metodologías NLP (Análisis de Lenguaje Natural). El objetivo es poder clasificar su postura, su polaridad, describir los patrones discursivos que allí aparecen y argumentos que se dejan entrever. Para complementar el análisis, también se suman las cuentas de los noticieros, que, aunque no representan la neutralidad, sí son un campo de lucha discursiva en común. Parte del objetivo es entender que posturas y discursos predominan en dichos espacios, y abrir una línea de investigación futura para compararla con otros países. También, se intentarán identificar la concentración de comentarios y cuentas que asimilan comportamientos de bots o en “núcleo duro”. Se intentará dar respuesta a las preguntas tales como: ¿Qué actores (personas, agrupaciones políticas y civiles, ONGs, etc.) son los que más aparecen nombradas en los discursos enfrentados? ¿En el año 2024, que noticia fue la que disparó el mayor interés y debate relacionados a la guerra? ¿Qué tan radicalizados se encuentran los usuarios al comentar? Finalmente, al adentrarnos en los discursos que se encuentran frente a una realidad empírica de la guerra, nos preguntaremos por sus motivaciones que dejan entrever en sus argumentos. ¿Qué peso ocupan al considerar su posición la correcta?

2. Análisis de la situación

2.1 Descripción del conflicto armado 2022-2025

La guerra en Ucrania (2022-2025) representa el conflicto armado más importante en Europa desde la Segunda Guerra Mundial, con profundas implicaciones geopolíticas, estratégicas y sociales. Iniciada con la invasión rusa a gran escala en febrero de 2022, la contienda ha dejado decenas de miles de víctimas y millones de desplazados, al tiempo que ha transformado el panorama de la seguridad europea y polarizado la opinión pública mundial. En el ámbito de la comunicación, esta guerra ha estado marcada por una intensa cobertura mediática internacional, una avalancha de información (y desinformación) en redes sociales, y una lucha narrativa entre los actores involucrados. Este trabajo ofrece un análisis de la comunicación a través de la red social YouTube con énfasis en los hechos clave de 2024, los principales actores en juego, la cobertura mediática y el ecosistema informativo en torno a la guerra. Asimismo, se justifica por qué este conflicto se erige como un campo fértil para el análisis comunicacional, dado que los medios de comunicación y las redes sociales se han convertido en escenarios estratégicos de disputa discursiva. A continuación, se presenta una descripción general de la guerra de Ucrania, seguida de secciones dedicadas a los eventos de 2024, al rol de los actores principales, al tratamiento mediático del conflicto (incluyendo fenómenos como la desinformación y las tendencias de búsqueda en el mundo hispanohablante), y finalmente a la relevancia comunicacional del caso ucraniano.

2.2 Historia general del conflicto (2022-2025)

El conflicto armado entre Rusia y Ucrania escaló dramáticamente el 24 de febrero de 2022, cuando Rusia lanzó una invasión militar a gran escala del territorio ucraniano. En los primeros días, las fuerzas rusas atacaron múltiples ciudades con misiles y avanzaron desde el norte (Bielorrusia), el este (Donbás) y el sur (Crimea), con el aparente objetivo de tomar Kiev rápidamente, derrocar al gobierno prooccidental de Volodímir Zelenski y devolver a Ucrania a la esfera de influencia rusa (Kirby, 2025). La ofensiva inicial causó graves daños a infraestructura civil como los hospitales, escuelas y viviendas fueron alcanzados por bombardeos indiscriminados y numerosas bajas entre la población civil. Según Human Rights Watch (2022), en la primera semana de hostilidades más de un millón de personas huyeron de sus hogares, muchas refugiándose en países vecinos, mientras en Rusia se intensificaba la censura

y se reprimían protestas internas contra la guerra. La invasión violó abiertamente el derecho internacional y sentó las bases de lo que sería un conflicto prolongado y devastador. A pesar de los avances iniciales, el plan ruso de una victoria rápida fracasó. Para marzo de 2022, la resistencia ucraniana -respaldada por suministros de armas de Occidente- logró detener el asalto sobre Kiev, obligando a las tropas rusas a retirarse del norte de Ucrania. Rusia entonces concentró su esfuerzo en el Donbás (este de Ucrania) y el sur donde sus avances fueron significativos. Durante 2022, las fuerzas rusas tomaron la ciudad portuaria de Mariúpol tras un sitio de tres meses, y llegaron a ocupar alrededor del 22% del territorio ucraniano, incluyendo la península de Crimea (anexada en 2014) y partes de las regiones de Donetsk, Lugansk, Zaporíyia y Jersón. Sin embargo, el ejército ucraniano llevó a cabo contraofensivas efectivas en la segunda mitad de 2022: en septiembre reconquistó amplias zonas de la provincia de Járkov, y en noviembre recuperó la ciudad de Jersón, la única capital regional que Rusia había logrado capturar desde la invasión de 2022. Estas victorias ucranianas demostraron las debilidades logísticas y de moral de los rusos, a la vez que dieron impulso a Ucrania en el frente diplomático, consolidando el apoyo occidental. En 2023, la guerra entró en una fase de desgaste. Rusia nombró nuevos comandantes y lanzó ofensivas concentradas en el Donbás, logrando avances muy costosos en localidades como Soledar y Bajmut tras meses de combates encarnizados. La ciudad de Bajmut cayó en manos rusas en mayo de 2023, en lo que Moscú proclamó como una victoria significativa, aunque estratégica y simbólicamente limitada. Ucrania, por su parte, recibió sistemas de armamento más avanzados (artillería de largo alcance, defensas antiaéreas, tanques occidentales) y en junio de 2023 inició una contraofensiva en los frentes sur y este. No obstante, esta contraofensiva ucraniana avanzó lentamente frente a las densas líneas defensivas rusas (campos minados, trincheras y “dientes de dragón”) y la superioridad aérea rusa. Hacia finales de 2023, el frente se había estabilizado con solo cambios territoriales marginales: ninguna de las dos partes consiguió una ruptura decisiva, y el conflicto se estancó en una guerra de atrición, con combates intensos, pero líneas relativamente fijas. Para 2024, el equilibrio militar seguía siendo frágil pero relativamente estable en términos territoriales. Rusia aún ocupaba aproximadamente una quinta parte del territorio ucraniano, incluyendo la mayor parte de las regiones orientales de Donetsk y Lugansk y amplias zonas del sur (Zaporíyia y Jersón al este del río Dniéper), además de Crimea. Ucrania conservaba el control de las ciudades principales (Kiev, Járkov, Odesa, Leópolis, etc.) y había logrado proteger con éxito su capital y la mayor parte del centro-occidente del país. Las hostilidades continuaban

en extensos frentes de batalla de más de 1.000 km de longitud, especialmente en Donbás. Como resume un informe de BBC News, en el este “la maquinaria de guerra de Moscú avanza lentamente, apenas unos metros a costa de grandes pérdidas, mientras la línea del frente ha cambiado muy poco en dos años”. Ambas partes parecían desgastadas por la contienda prolongada, pero ninguna daba señales de ceder en sus objetivos fundamentales.



Figura 1. Mapa del este de Ucrania que muestra en rojo las zonas bajo ocupación militar rusa (aprox. 20% del país) hacia mayo de 2025. Las regiones sombreadas indican los límites del avance ucraniano antes del estancamiento del frente. A esa fecha, la guerra había dejado al menos 12.654 civiles muertos y 29.392 heridos confirmados por la ONU desde 2022 (incluyendo 673 niños) (Oficina de la ONU para los Derechos Humanos, 2025). Además, cerca de 7 millones de ucranianos se convirtieron en refugiados en Europa y otras regiones, y unos 3,7 millones eran desplazados internos, configurando la mayor crisis humanitaria en Europa en décadas. Grandes extensiones de Ucrania (alrededor de 139.000 km²) han quedado contaminadas con minas y municiones sin detonar. La dimensión destructiva y humana del conflicto ilustra su escala: Ucrania estima pérdidas militares propias en decenas de miles de soldados, y atribuye a Rusia pérdidas aún mayores, si bien las cifras exactas son objeto de propaganda y difícil verificación (NYT, 2023; Le Grand Continent, 2025). En cuanto a los esfuerzos diplomáticos, hasta 2025 no se había logrado una negociación de paz efectiva. La ONU condenó la invasión en resoluciones de la Asamblea General con amplio respaldo, pero Rusia -miembro permanente del Consejo de Seguridad- bloqueó cualquier acción en ese órgano. Se sucedieron

intentos de mediación o planes de paz propuestos por diversos actores (ONU, Turquía, Francia, Santa Sede, China, entre otros), sin consenso. A mediados de 2023, se alcanzó y luego colapsó un acuerdo para la exportación de granos ucranianos vía el Mar Negro bajo mediación de Turquía y la ONU, reflejando la fragilidad de pactos limitados en medio de la desconfianza. Hacia finales de 2024 e inicios de 2025, surgieron informes sobre contactos exploratorios -incluyendo la posibilidad de negociaciones indirectas entre Estados Unidos y Rusia-, pero Ucrania insistió en su exigencia de retirada rusa como condición previa (CFR, 2025). En suma, al cumplirse tres años de guerra, el conflicto seguía abierto y sin un final claro, con las partes atrincheradas en sus posiciones militares y políticas.

2.3 Eventos clave de 2024: batallas, acuerdos y cambios territoriales

El año 2024 estuvo marcado por algunos acontecimientos militares y políticos significativos en la guerra de Ucrania, aunque en términos generales consolidó las tendencias de 2023. Un hecho destacado a inicios de 2024 fue la **batalla de Avdiivka**, una localidad fortificada en Donetsk que durante años había sido línea de frente. En febrero de 2024, tras intensos asaltos, las tropas ucranianas realizaron una retirada de Avdiivka para evitar su cerco, lo que permitió a Rusia proclamar la captura de la localidad. La caída de Avdiivka supuso “la mayor victoria rusa desde la toma de Bajmut en mayo de 2023” (Reuters, 2024), otorgando a Moscú un trofeo propagandístico y una mejor posición táctica al norte de la ciudad de Donetsk. Aunque Avdiivka quedó devastada casi por completo después de meses de bombardeos, su captura evidenció que Rusia todavía podía avanzar localmente a costa de grandes pérdidas. Otro foco de actividad en 2024 fue la frontera nororiental entre Ucrania y Rusia. A mediados de año, se reportaron incursiones y escaramuzas transfronterizas, incluyendo una ofensiva limitada ucraniana en la región rusa de Bélgorod/Kursk, donde grupos armados opuestos al Kremlin (supuestamente compuestos de voluntarios rusos anti-Putin) realizaron breves incursiones. Estos incidentes, aunque de escala menor, señalaron que Ucrania estaba llevando el conflicto más allá de sus fronteras, explotando la vulnerabilidad de algunas zonas rusas limítrofes. Asimismo, Ucrania intensificó en 2024 el uso de **drones de largo alcance para atacar objetivos en territorio ruso**, como bases aéreas, depósitos de combustible e incluso la ciudad de Moscú. En particular, en **noviembre de 2024** se produjo el mayor ataque con drones contra la capital rusa hasta la fecha, logrado penetrar las defensas aéreas y demostrando la creciente capacidad ucraniana de

golpear en la retaguardia de Rusia (Council on Foreign Relations, 2025). Aunque estos ataques tuvieron más efecto psicológico y simbólico que impacto estratégico, obligaron al Kremlin a reconocer la extensión del conflicto al propio suelo ruso.

En el sur de Ucrania, a lo largo de 2024 continuaron los combates en la provincia de Zaporíyia, especialmente en torno al eje Orihiv-Tokmak, donde las fuerzas ucranianas habían logrado liberar algunas localidades (como Robotine) a fines de 2023. Sin embargo, no se registraron **cambios territoriales de gran escala**: la esperada ruptura ucraniana hacia Melitópol o Berdiansk no se materializó, en parte debido a la férrea defensa rusa y a la superioridad aérea de Moscú. Del lado ruso, más allá de Avdiivka, sus fuerzas hicieron avances marginales en el frente oriental, expandiendo gradualmente el área bajo su control en Donetsk y Lugansk. Un análisis del Institute for the Study of War estimó que Rusia ganó unos 4.168 km² de territorio en 2024, mayormente campos y pequeñas poblaciones en el este de Ucrania. No obstante, este incremento es modesto comparado con lo obtenido en la invasión inicial de 2022, y vino al alto costo de bajas rusas en el año 2024.

En el plano político y diplomático, 2024 trajo desarrollos relevantes. Por un lado, la guerra propició la expansión de la OTAN: Finlandia se había adherido en 2023 y en marzo de 2024 Suecia se convirtió oficialmente en el miembro número 32 de la Alianza Atlántica, ampliando el flanco nor-europeo de la OTAN y contradiciendo uno de los objetivos declarados de Putin (frenar la expansión de la Alianza). Además, Ucrania continuó acercándose a Occidente institucionalmente: la Unión Europea mantuvo el estatus de país candidato de Ucrania y, tras reformas internas, anunció a finales de 2024 su disposición a iniciar el proceso formal de negociaciones de adhesión con Kiev. Estas acciones reforzaron el mensaje político de que, lejos de alejar a Ucrania de Occidente, la agresión rusa había solidificado esos lazos.

Por otro lado, 2024 evidenció ciertas fisuras en la unidad occidental respecto al apoyo indefinido a Ucrania. En Estados Unidos, la proximidad de las elecciones y el relevo en el liderazgo de la Cámara de Representantes generaron debates sobre la continuidad de los paquetes de ayuda militar multimillonarios. No obstante, a pesar de algunas voces disidentes, hacia 2025 Estados Unidos seguía siendo el principal sostén de Ucrania. Desde 2022 hasta inicios de 2025, Ucrania recibió aproximadamente 270.000 millones de euros en asistencia internacional, de los cuales cerca del 43% provino de EE. UU. y una cifra casi equivalente de los países europeos en conjunto. La Unión Europea y sus estados miembros aportaron asistencia financiera, humanitaria y militar valuada en más de **132.000 millones de euros** en ese periodo, destacando iniciativas como el Fondo Europeo de Apoyo a la Paz,

utilizado por primera vez para suministrar armamento a un país en guerra. Estos datos dan cuenta de un compromiso sostenido de Occidente, aunque en 2024 crecieron los llamados -desde sectores políticos en EE. UU. y Europa- a buscar una solución negociada al conflicto si la situación militar seguía estancada. En suma, 2024 consolidó la guerra en Ucrania como un conflicto prolongado, de intensidad media-alta, donde ningún bando logró ventajas definitivas, aunque sí marcó avances territoriales para Rusia.

2.4 Cobertura mediática, desinformación y opinión pública

Desde el inicio de la guerra en 2022, el conflicto en Ucrania fue ampliamente cubierto por medios internacionales, dominando titulares y generando gran interés público. Durante los primeros meses, cadenas como BBC, CNN, *El País* o *The New York Times* ofrecían actualizaciones constantes, reflejando tanto la gravedad del conflicto como la demanda informativa. Sin embargo, con el paso del tiempo, la cobertura disminuyó y se centró solo en eventos impactantes. Un estudio de Díez-Gracia (2024) mostró que la atención mediática cayó notablemente después del primer semestre de 2022, y muchas noticias sobre la guerra tuvieron bajo nivel de lectura y difusión, evidenciando fatiga informativa.

En España, la guerra fue intensamente cubierta al inicio, con gran presencia en medios y alto interés ciudadano. No obstante, el conflicto fue perdiendo protagonismo frente a otras crisis, como la guerra entre Israel y Hamás en 2023, que desplazó el foco informativo en países como España e Italia (Le Grand Continent, 2023). Aun así, la cobertura repuntó en ciertos momentos políticos clave, como la presidencia española del Consejo de la UE.

El conflicto también ha estado marcado por una intensa guerra informativa. Rusia desplegó una estrategia sistemática de desinformación, utilizando medios estatales como RT y Sputnik, así como redes sociales, para justificar la invasión y dividir la opinión pública. Pese a estar bloqueados en Europa, estos medios han prosperado en español, especialmente en América Latina, aprovechando el déficit de medios alternativos y el escepticismo hacia EE. UU. (Reuters Institute, 2023). En España, aunque más restringidos, estos contenidos han circulado por canales de Telegram o medios marginales.

Ucrania, por su parte, ha usado la comunicación estratégica principalmente para ganar apoyo internacional, promoviendo contenidos que destacan la resistencia civil y militar, con iniciativas de verificación como StopFake. Si bien también ha empleado

elementos propagandísticos, su enfoque ha sido más defensivo y centrado en la visibilidad global.

Las redes sociales han jugado un rol central: Twitter, Facebook, Telegram y TikTok han sido espacios de difusión de información veraz y falsa. Plataformas como TikTok dieron lugar a lo que algunos denominaron la “primera guerra TikTok”, por la abundancia de vídeos desde el frente, muchos de ellos virales pero difíciles de verificar (Hasan, 2024). En YouTube, los comentarios en vídeos sobre la guerra reflejan una clara polarización ideológica: desde apoyos fervientes a Ucrania hasta discursos prorrusos con fuerte carga propagandística. El interés público también se reflejó en las búsquedas en Google. En 2022, “Ucrania” fue uno de los términos más buscados en España. Sin embargo, en 2023 desapareció del Top 10, cediendo espacio a otros temas como elecciones o conflictos más recientes. Aun así, encuestas han mostrado que la mayoría de los ciudadanos españoles siguen apoyando la ayuda a Ucrania (Real Instituto Elcano, 2023), mientras en América Latina las posturas están más divididas.

En resumen, la guerra de Ucrania ha sido una batalla no solo militar, sino informativa. La evolución de su cobertura, el impacto de la desinformación, y la disputa narrativa en redes la convierten en un caso ejemplar para el análisis comunicacional contemporáneo.

2.5 Identificación de los Canales relevantes en España.

Se han identificado las cuentas de YouTube en España (contenido en español) más relevantes que han publicado videos sobre la guerra en Ucrania durante 2024 y 2025. Se han clasificado en tres grupos ideológicos según su postura predominante: pro-ucranianas, noticieros y pro-rusas.

VisualPolitik es un canal español centrado en geopolítica con un enfoque abiertamente liberal y prooccidental. Entre 2024 y 2025 ha publicado numerosos videos sobre la guerra, destacando las debilidades rusas y el apoyo occidental a Ucrania. Su narrativa defiende las sanciones a Rusia y la alianza transatlántica.

Memorias de Pez ofrece contenidos educativos semanales sobre historia y actualidad. Durante 2024-25 ha cubierto la guerra en su serie “La Pecera de Memorias”, mostrando simpatía hacia Ucrania y criticando las acciones rusas, especialmente en el trato a soldados y civiles.

RTVE Noticias ofrece cobertura constante y objetiva desde el servicio público. La guerra de Ucrania ha sido uno de sus temas más vistos. Aunque mantiene neutralidad

institucional, adopta términos como “invasión rusa” y refleja el marco europeo de condena.

El País y **El Mundo** son diarios de referencia en España. Ambos han producido videos y coberturas especiales sobre la guerra. El País refuerza la narrativa europea contra la invasión, mientras El Mundo se enfoca en análisis legales y diplomáticos sin sesgo aparente.

Miguel Ruiz Calvo es un abogado y streamer con emisiones diarias sobre la guerra. Su discurso es claramente prorruso, amplificando supuestos logros militares de Rusia y criticando duramente a Ucrania y la OTAN.

Rubén Gisbert ha viajado al Donbás y publicado contenidos que cuestionan la versión occidental del conflicto. Ha aparecido en RT y promovido teorías prorrusas, como negar crímenes atribuidos a Rusia.

3. Obtención, procesado y almacenamiento de los datos

3.1 Diseño de Proyecto

El proyecto se diseñó con el siguiente árbol de carpetas que se explicita en el archivo README.md.

```
proyecto_guerra_ukrania_sentimientos/
|
|—— data/          # Datos descargados y procesados
|   |—— raw/        # Datos originales desde la API de YouTube
|   |—— processed/   # Datos limpios y enriquecidos
|
|—— notebooks/      # Notebooks exploratorios
|   |—— 01_scraping_youtube.ipynb
|
|—— src/            # Código Python reutilizable
|   |—— youtube_api.py  # Funciones de extracción de videos y comentarios
|   |—— preprocessing.py # Limpieza de texto y etiquetas
|   |—— analysis.py     # Análisis de sentimiento y gráficos
|
|—— models/         # Modelos entrenados o embeddings
```

```

├── reports/          # Figuras y resultados para informe
|   └── figures/
|
├── requirements.txt  # Paquetes necesarios para reproducir
├── README.md        # Este archivo
├── .gitignore       # Archivos a excluir en el repositorio
└── .env             # Archivo de configuración de las variables de ambiente. No es
visible en github

```

Por el momento esta estructura es la que marca el diseño principal, sin embargo, se están trabajando las actividades en notebooks .ipynb hasta tener la versión final productiva para entregar. Eso permite crear visualizaciones intermedias y realizar *debugging* al momento de desarrollo.

A dicha estructura se suma la carpeta venv por fuera del proyecto para ejecutar el proceso en un ambiente virtual. De esa forma se siguen las buenas prácticas de la industria. Por otro lado, como se puede observar por la estructura propuesta el trabajo con los datos se dividió en varias etapas para dar claridad, y acotar riesgos al separar los procesos y archivos en el proceso. Tenemos checkpoints intermedios entre la obtención de los datos desde la API de Youtube que queda en la siguiente ubicación relativa: \data\raw\0_comments_raw.csv

Esta etapa se caracteriza por la extracción, por batches de los datos principales de la API de Youtube y los metadatos iniciales:

Campo	Descripción breve
comment_id	ID único del comentario
comment	Texto completo del comentario
comment_text_length	Longitud en caracteres del comentario
user_id	ID del canal/usuario que hizo el comentario
user_name	Nombre de pantalla del autor
comment_time	Fecha y hora en que se publicó el comentario
comment_likes	Número de “me gusta” recibido por el comentario
total_reply_count	Número de respuestas (replies) al comentario

Campo	Descripción breve
video_title	Título del vídeo al que pertenece el comentario
channel_title	Nombre del canal del vídeo
video_published_at	Fecha y hora de publicación del vídeo
video_views	Cantidad de visualizaciones del vídeo
video_likes	Cantidad de “me gusta” del vídeo
video_duration	Duración ISO 8601 del vídeo (p.ej. PT5M30S)
video_tags	Lista de etiquetas (tags) asociadas al vídeo
video_category_id	ID numérico de la categoría del vídeo
relacion_evento	Tu campo: relación con el evento de interés
evento	Tu campo: nombre del evento (p.ej. “Ucrania”)
tipo_evento	Tu campo: tipo de evento (p.ej. “Favorece a Rusia”, “Favorece a Ucrania”, “Neutro”)
condiciones_cuenta	Tu campo: clasificación del canal (p.ej. “pro-ruso”, “noticiero”, etc.)

El siguiente paso es el enriquecimiento de ese dataset inicial con los datos de usuarios. Para eso, también se realiza una primera limpieza que es sacar los comentarios con Id duplicado. Realmente no presentan un número significativo y probablemente se deba al trabajo en *batches* que se realizó al recuperar esos datos. Con estos datos ya procesados se crea un nuevo dataset en la ubicación relativa `\data\processed\1_comments_youtube_clean.csv` y se suman las siguientes columnas:

Campo	Descripción breve
account_created_at	Fecha y hora de creación de la cuenta del usuario que comenta
channel_id	ID único del canal donde se publicó el vídeo al que pertenece el comentario
subscriber_count	Número actual de suscriptores del canal

En la tercera etapa se realiza el trabajo a través del análisis EDA (Análisis Explorativo de Datos) que también guarda una nueva versión del dataset. Este se ubica en la siguiente ruta: `\data\processed\2_comments_youtube_refined.csv`. En este punto se ajustan los campos que representan tiempo a `datetime`, se eliminan todas filas con

campos nulos y también se crea un nuevo campo para calcular hace cuanto se creó la cuenta que deja el comentario en días entre su creación hasta la fecha del comentario. De esa forma se trata de entender que tan reciente fue la cuenta al escribir dicho comentario.

3.2 Obtención de información desde la API de YouTube

La información utilizada en este trabajo se recolectó mediante la API oficial de YouTube Data v3, que permite acceder a datos públicos como comentarios, detalles de videos, estadísticas y metadatos relacionados. La implementación técnica se realizó en Python, usando la biblioteca `googleapiclient` para manejar solicitudes autenticadas a través de una clave API almacenada en un archivo `.env` para mayor seguridad y confidencialidad (Google Developers, 2024). Por otro lado, para enriquecer los datos, además de la información de los comentarios, hemos sumado la información adicional de los canales y los usuarios. La información que hemos recabado por cuenta y canal la hemos unificado en el dataset principal. Se decidió trabajarlo de forma desnormalizada para la facilidad del análisis inicial con Pandas. También para la facilidad de la interpretación posterior, la información se extrajo categorizando los canales analizados según su postura editorial hacia la guerra en Ucrania: pro-ucraniano, pro-ruso o noticiero/neutro. Esta categorización inicial facilita análisis posteriores de polarización política en los comentarios obtenidos.

3.3 Limitaciones y observaciones

El proceso de extracción presentó ciertas limitaciones técnicas inherentes a la API de YouTube, como la cuota diaria de peticiones y el límite máximo de comentarios recuperables por video, lo que impacta directamente en la completitud de los datos obtenidos. Otra limitación relevante fue la restricción o bloqueo de comentarios por parte de ciertos canales en videos específicos, algo común en temas controvertidos (Google Developers Policy, 2024). En ese contexto por la API de la llamada pública no pude acceder a los videos de los canales VisualPolitik, sin embargo, se trabajará para lograrlo en la iteración final de este trabajo. También hubo problemas para ordenar y filtrar efectivamente los videos de interés ya que el nombre por el que se buscaban dichos videos no era altamente restrictivo y trajo videos que además de la temática principal contenía otras no relacionadas. Es el caso que se da fuertemente en el canal Miguel Ruiz Calvo que agrupa sus videos por varias temáticas a la vez.

3.4 Almacenamiento de los datos

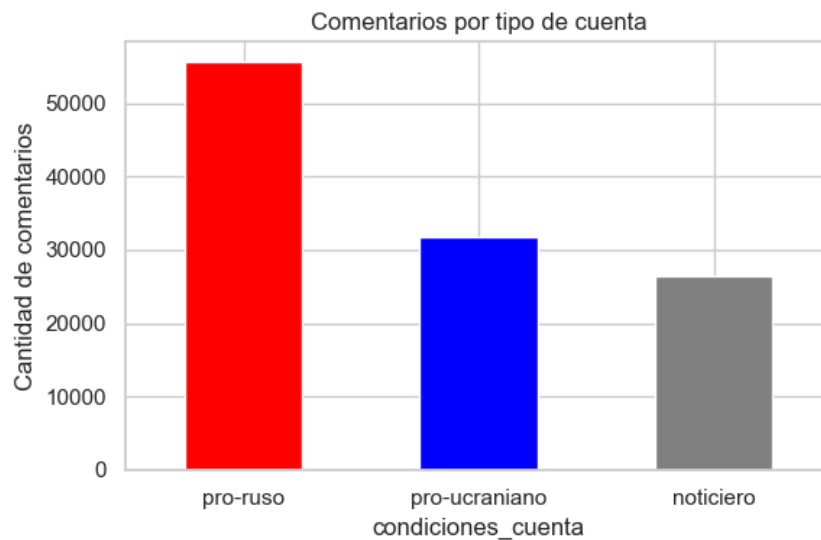
Los datos inicialmente fueron almacenados en archivos CSV, facilitando la manipulación inicial y análisis exploratorios rápidos. Considerando la escalabilidad del proyecto y futuros análisis más avanzados, se estará migrando estos datos a un almacén estructurado más robusto como DuckDB que ofrece una opción económica y eficiente para análisis locales con limitados recursos computacionales (Raasveldt & Mühleisen, 2022). Esta solución va a permitir escalar, con cero costos extra, el *pipeline* de datos y acelerar el procesamiento de estos. Se va a trabajar con DuckDB con la tabla principal denormalizada y desde el mismo código principal de Python. Las tablas se podrán importar y exportar y se realizarán las consultas en SQL para la exploración de datos. Finalmente, estas tablas se estarán explorando desde PowerBI para lograr la presentación final en un dashboard, además de la documentación de la investigación complementaria que va a estar disponible en formato PDF.

3.5 GitHub - versionamiento y documentación

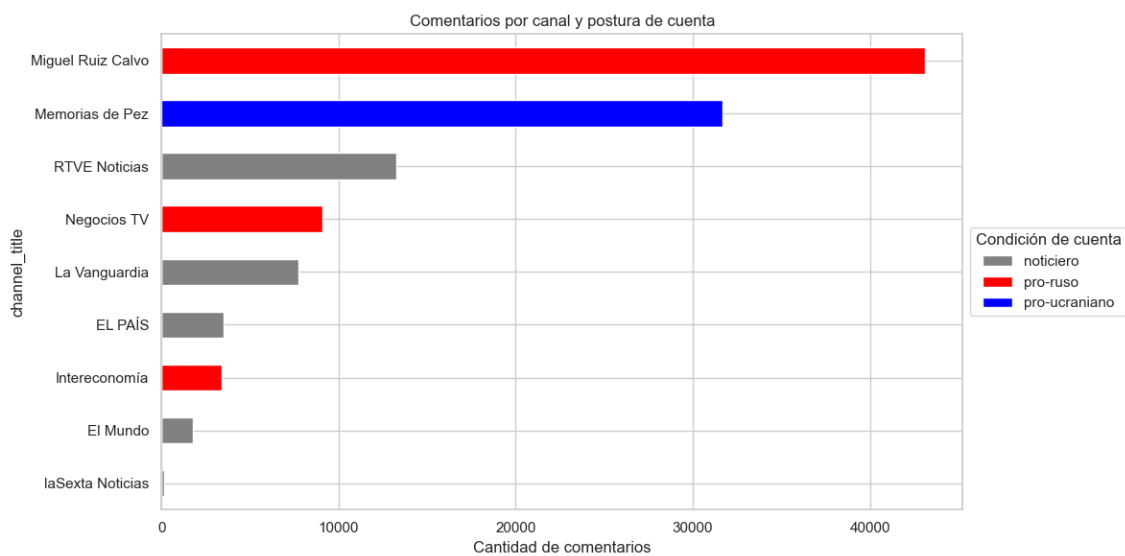
Se utiliza GitHub como sistema de control de versiones para el código fuente del proyecto, garantizando reproducibilidad y trazabilidad de los cambios realizados. El código desarrollado es documentado mediante comentarios claros y detallados que explican las funciones, variables y metodologías aplicadas, siguiendo prácticas recomendadas para proyectos de ciencia de datos. Consecuentemente, además de la presentación formal en los anexos de este trabajo y el Tablero final en PowerBI los resultados se podrán consultar también en ese enlace de GitHub que será compartido en los anexos.

4. Análisis exploratorio

El análisis exploratorio se inició con la utilización de la biblioteca Pandas (McKinney, 2022). Inicialmente se hizo la limpieza de los nulos que ya se mencionó para calcular el total de comentarios recuperados por canal y otras distribuciones claves. Esto permitió obtener una clara visualización inicial del volumen relativo de participación según la orientación editorial del canal.



Observamos que, así como están los datos, tenemos mayor presencia de comentarios en los canales pro-rusos. Al representar dichos resultados a mayor detalle vemos que el canal pro-ruso de Miguel Ruiz Calvo está arriba del ranking. Es sorprendente ya que le gana a Memorias de Pez que tiene hoy cinco veces más de suscriptores.



En cuanto a la distribución temporal de los comentarios, justamente podemos observar una incidencia alta en noviembre-diciembre de 2024 marcando la relevancia del canal pro-ruso mencionado en ese periodo en particular.

- c. Bots, trolls o activistas (cuentas nuevas con muchos comentarios y me gusta entre los videos)
- d. Ejes argumentativos
- e. Clustering
- f. Palabras claves x tipo (político, militar, autores del video).

6. Análisis de los resultados obtenidos

- a. Dashboard con los resultados
- b. Resultados de como se encuentra la polaridad
- c. Resultados de quien tiene más comentarios (pro-rusos, pro-ucranianos, neutros) y sus características principales.

7. Conclusiones y planes de mejora

- a. Poder incluir otras redes
- b. Videos que no se pudieron consultar

8. Bibliografía

Agencia EFE. (2025, 17 de febrero). ONU hace balance de 3 años de crímenes en Ucrania: 12.600 civiles muertos, 29.300 heridos. **SWI swissinfo.ch**. Recuperado de [swissinfo.chswissinfo.ch](https://www.swissinfo.ch/spa/ONU-hace-balance-de-3-años-de-crímenes-en-Ucrania-12600-civiles-muertos-29300-heridos-2025-02-17)

Díez-Gracia, A. (2024). Agenda y demanda informativa sobre la guerra de Ucrania en la prensa internacional. **Revista de Comunicación**, 23(2), 1-21. Recuperado de revistadecomunicacion.comrevistadecomunicacion.com

Hasan, M. (2024). Russia-Ukraine Propaganda on Social Media: A Bibliometric Analysis. **Journalism and Media**, 5(3), 980-992.

<https://doi.org/10.3390/journalmedia5030062> [mdpi.com](https://www.mdpi.com)

Human Rights Watch. (2022). Guerra entre Rusia y Ucrania (nota descriptiva, 24 de febrero de 2022). Human Rights Watch en español. Recuperado de [hrw.org](https://www.hrw.org/hrw.org)[hrw.org](https://www.hrw.org)

Kahn, G. (2023, 31 de marzo). Bloqueada en Occidente, la propaganda rusa prospera en español en TV y redes sociales. **Radio Televisión Martí / Reuters Institute**.

Recuperado de [martinoticias.com](https://www.martinoticias.com)[martinoticias.com](https://www.martinoticias.com)

Kirby, P. (2025, 15 de mayo). Why did Putin's Russia invade Ukraine? **BBC News**.

Recuperado de [bbc.co.uk](https://www.bbc.co.uk/bbc.co.uk)[bbc.co.uk](https://www.bbc.co.uk)

Le Grand Continent. (2023, 30 de diciembre). Las tendencias de búsqueda en Google en 2023. Le Grand Continent. Recuperado de

[legrandcontinent.eu](https://www.legrandcontinent.eu)[legrandcontinent.eu](https://www.legrandcontinent.eu)

Marinero, I. (2022, 7 de diciembre). Lo más buscado en Google en 2022 en España: de Ucrania y el Wordle a Tamara Falcó y Will Smith. **El Español - Omicrono**. Recuperado de [elespanol.com](https://www.elespanol.com)[elespanol.com](https://www.elespanol.com)

Reuters (Guy Faulconbridge). (2024, 17 de febrero). Russia says its forces move forward after Ukraine withdraws from Avdiivka. **Reuters**. Recuperado de [reuters.com](https://www.reuters.com)

Reuters (Max Hunder & Guy Faulconbridge). (2023, 31 de marzo). Russia vs Ukraine: the biggest war of the fake news era. **Reuters** (traducido en Forbes México).

Recuperado de [forbes.com.mx](https://www.forbes.com.mx)

Statista. (2024). Global Google search interest in "Ukraine war" (2022-2024). Statista Research. Recuperado de [legrandcontinent.eu](https://www.legrandcontinent.eu)

Stoltenberg, J. (2023). Speech by NATO Secretary General at the Munich Security Conference. **NATO.int** (18 de febrero de 2023).

The Economist. (2023, 20 de mayo). How Ukraine is fighting the propaganda war. **The Economist**.

The New York Times. (2023, 22 de agosto). Troop Deaths and Injuries in Ukraine War

Near 500,000, U.S. Officials Say. **NYTimes**.

Zelenski, V. (2022). Discurso ante el Congreso de los EE. UU. (16 de marzo de 2022). Transcript via **El País**.

Google Developers. (2024). YouTube Data API v3. Recuperado de <https://developers.google.com/youtube/v3>

Google Developers Policy. (2024). API Usage Limits. Recuperado de <https://developers.google.com/youtube/terms>

Raasveldt, M., & Mühleisen, H. (2022). DuckDB: An Embeddable Analytical Database. VLDB Endowment.

McKinney, W. (2022). Python for Data Analysis (3rd ed.). O'Reilly Media.

9. Anexo con el código fuente desarrollado.



06_classification.ipynb



05_EDA_comentarios.ipynb



04_batch_get_comments_final.ipynb