**TRABAJO INTEGRADOR FINAL**

**ETAPA 1. Planteo de problema**

Integrantes, grupo 13:

Deglise, Tomás

Dodonov, Vladyslav

Rossetti, Natalia

1. **Planteo del Problema**

En la presente investigación se pretende identificar y mensurar posibles cambios en el clima de opinión y en el humor social, por medio de un análisis comparado en redes sociales.

Para este fin se articulará un análisis que toma las interacciones en redes sociales (Youtube) de tres grupos de cuentas: a) cuentas de dos portales de medios digitales[[1]](#footnote-0); b) cuentas de usuarios/as relevantes fuertemente a favor del gobierno; y c) cuentas de usuarios/as relevantes fuertemente en contra del gobierno. De éste modo se pretende rastrear patrones de cambios de opinión y humor social con mayor precisión y menor sesgo: inspirados en la metodología comparativa por diferencia, podremos asumir que, si en los tres grupos se presentan cambios de opinión en dirección similar, ésta no dependerá de las diferencias ideológicas de los grupos, sino que se tratará de un cambio lo suficientemente fuerte como para romper las barreras ideológicas de los respectivos electorados “duros” de cada bando. Al mismo tiempo, los datos pueden permitirnos una aproximación, incluso, a las lógicas de polarización en torno a ciertos issues.

**Objetivos Generales**

Explorar interacciones y relaciones de usuarios/as a la plataforma social *youtube* con las publicaciones de medios de comunicación sobre las políticas del gobierno del presidente Javier Milei, identificando patrones de polarización y cambios de opinión a través del análisis de comentarios en esa plataforma.

**Objetivos Específicos**

* Medir posibles cambios de opinión y de humor social en usuarios/as de canales de *youtube* ante ciertos eventos o publicaciones.
* Medir posibles cambios de opinión con respecto al gobierno entre usuarios/as que interactúan con cuentas a favor del gobierno.
* Medir posibles cambios de opinión con respecto al gobierno entre usuarios/as que interactúan con cuentas en contra del gobierno.
* Identificar si hay algún tipo de “issue” (o tópicos) que genere mayor malestar o rechazo entre los usuarios en las redes sociales
* Detectar posibles lógicas de polarización de opinión en los diferentes grupos estudiados.
* Construir una visualización de palabras principales para proporcionar una comprensión rápida de los términos y conceptos más recurrentes.
* Detectar posibles granjas de usuarios/as de opinión en los diferentes grupos estudiados para observar la influencia sobre estos “issues” (o tópicos).

**Justificación**

Javier Milei es el primer presidente outsider en Argentina que interactúa de manera particular en las redes sociales, a diferencia de otros mandatarios/as y políticos/as. Es una figura política polarizadora, y su gestión está generando una amplia gama de reacciones en la sociedad.

La interacción en el mundo de las redes sociales, un microclima del humor social, nos aproxima a los temas que circulan en la formación de opinión pública y pueden darnos una idea de la movilización de apoyo o rechazo político. Especialmente porque Milei rompe con el esquema de comunicación política tradicional, utilizando las redes y plataformas digitales como forma de comunicación y espacio de expresión no institucional.

Buscaremos comprender cómo las opiniones políticas se agrupan en burbujas informativas y cómo interactúan entre sí en el ámbito digital. Además, nos interesa investigar si las publicaciones pueden influir en los cambios de opinión de los/as usuarios/as o si los comentarios reflejan opiniones sólidas.

Para ampliar el análisis, consideramos necesario observar canales de *youtube de* portales de medios de comunicación masivos y comparar las interacciones de los/as *paradestinatarios/as[[2]](#footnote-1)* en momentos y fechas específicas que serán escogidas para este trabajo, a fin de comparar la conducta en grupos a favor y en contra. Datos que planean usar para responder la pregunta

**Metodología**

La investigación se llevará a cabo por medio del relevamiento de los comentarios a los tres grupos de usuarios/as ya nombrados. A través de éste se buscará poder clasificar y medir las interacciones en al menos tres: a favor del gobierno, en contra del gobierno y neutrales.

Los relevamientos, por su parte, se darán en virtud de ciertas fechas políticamente relevantes a lo largo del tiempo, a fin de poder ver la evolución del humor y la opinión.

Estas fechas de relevancia son retomadas en virtud de construir un corpus compuesto tanto momentos “álgidos” para el gobierno (donde se facilita una opinión en contra de éste), momentos favorables (donde se facilita una opinión favorable) y fechas “neutrales” a fin de controlar la medición evitando sesgos: si los cambios de opinión presentan un patrón que trasciende el tipo de evento podemos afirmar que la tendencia efectivamente es de un cambio de opinión general y no específica en relación a un issue.

En este sentido, se retoman los siguientes eventos:

Como eventos políticos desfavorables: a) el primer paro nacional (24 de enero), b) la marcha federal universitaria (23 de abril), c) la crisis diplomática con España (3ra semana de mayo) y d) la polémica por los alimentos del Ministerio de Capital Humano (30 de mayo) y e) fórmulas de jubilaciones (6 de junio).

Como eventos políticos neutrales se decidió tomar efemérides políticas: a) el 8 de marzo, b) el 24 de marzo, c) el 25 de mayo, y d) el 20 de junio.

Como eventos políticos favorables se tomaron los siguientes: a) la aprobación general de la Ley bases en la cámara de diputados, b) la reunión del Presidente con Elon Musk, c) el anuncio de la reducción de la inflación en abril, d) la presentación del libro del Presidente y e) renuncia a la jubilación al cumplir el mandato.

Éste relevamiento se realizará cotejando los tres grupos de usuarios/as, los cuales se componen del siguiente modo:

| **Condición** | **Plataforma social** | **Usuario** | **Seguidores** |
| --- | --- | --- | --- |
| A favor del gobierno | YouTube | El Peluca Milei | 1,7 mill. |
| Urgente Milei | 573k |
| Neutrales | YouTube | La Nación | 2.86 M |
| Infobae | 1.14 M |
| En contra del gobierno | YouTube | País de boludos | 209 k |
| Rebelión popular | 794 k |

La selección de usuarios/as de los grupos a favor y en contra se da en virtud de tres aspectos. Por un lado, su posicionamiento claro frente a las políticas del gobierno actual y, por lo mismo, una audiencia que comulga con tal posicionamiento. Por otro, la relevancia en términos de influencia: no sólo tienen gran cantidad de seguidores/as, sino también de visualizaciones, interacciones y, en suma, influencia con sus propias comunidades de usuarios/as.

Por otro lado las cuentas “neutrales” se tomaron noticias sobre los acontecimientos a observar en los canales de *youtube* de los portales de noticias de “Infobae” y “La Nación”, considerando que son espacios que intercambian discursos contradictorios de los tres actores que tienen y buscan legitimidad para hablar en público sobre política, que son: los políticos, los periodistas y la opinión pública. En este sentido, las audiencias de estas cuentas no estarán necesariamente sesgadas ideológicamente: muchos seguirán esas cuentas en virtud de la investidura presidencial y otros, incluso, comentaran allí como forma de interacción y expresión, sin necesariamente ser seguidores o no. De éste modo, se entiende que éstas cuentas están más abiertas a un público “general”.

*Modificaciones metodológicas*

Para construir una base fiable que nos permita comparar las distintas cuentas y cumplir nuestros objetivos, tuvimos que realizar cambios en la propuesta metodológica. Ante la dificultad de acceder a la API de Instagram, procedimos a observar la interacción y vínculo de los usuarios en los canales de YouTube. Esto implicó que no observamos desde la cuenta del canal del actual presidente, ya que no tiene un canal propio activo con un caudal significativo de usuarios/as. Así, seleccionamos dos portales de noticias: La Nación e Infobae, que puedan alcanzar los propósitos presentados.

**Hipótesis de investigación**

En la presente investigación se cuentan con ciertas hipótesis al respecto, las cuales se pueden enumerar en las siguientes:

Habrá una evolución en el clima de opinión tendiente al cambio.

Este cambio se dará principalmente en contra del gobierno de manera general, y se podrá ver específicamente en una aparición cada vez mayor de opiniones negativas en las cuentas neutras y pro gobierno.

Los eventos sociales relacionados a issues sociales implicarán una polarización mayor en las opiniones.

**Parte 2. Análisis Exploratorio | Diseño y Documentación**

**Datasets recolectados**

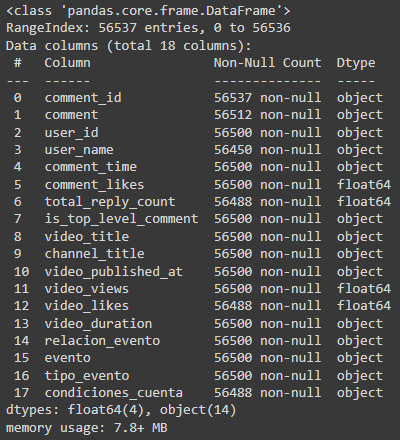
1. **Dataset Youtube**
   * Metadatos de videos de Youtube de las cuentas afines a Milei
   * Muestra de datos y metadatos de los comentarios del video

Los datos contienen información relacionada a los objetos base (videos) y los comentarios relacionados. Hemos tomado datos combinados relacionados a los comentarios, los/as usuarios/as (id y nombre por los límites de privacidad), al video y al canal. También hemos agregado etiquetas específicas relacionadas a los eventos y la posición del canal para poder utilizarlo posteriormente para mayor segmentación.

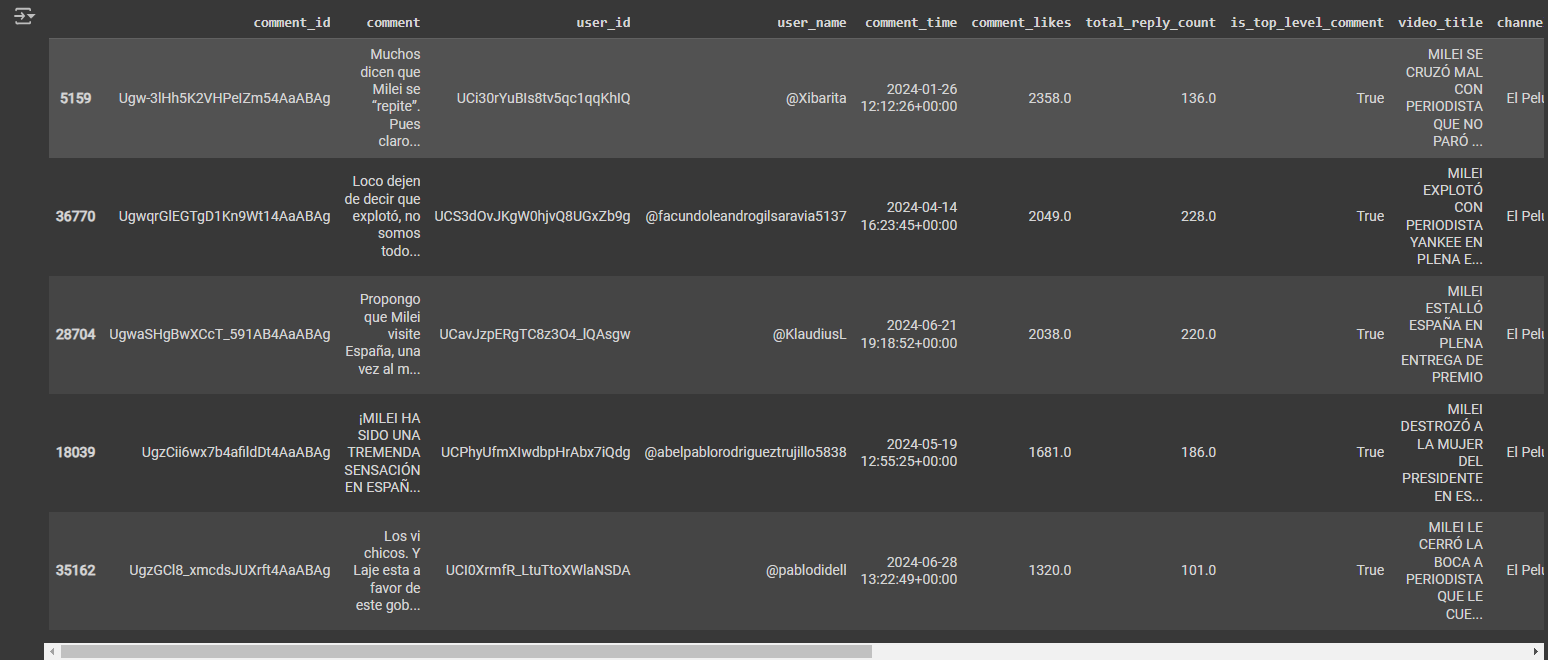
Se realizó una recolección de datos (API) y armado de base dataset. Para esta tarea utilizamos Youtube Data Api oficial.[[3]](#footnote-2)

Los datos que pudimos capturar desde la API de Youtube son los siguientes:

Con *comentarios.info()* podemos observar la base de los resultados

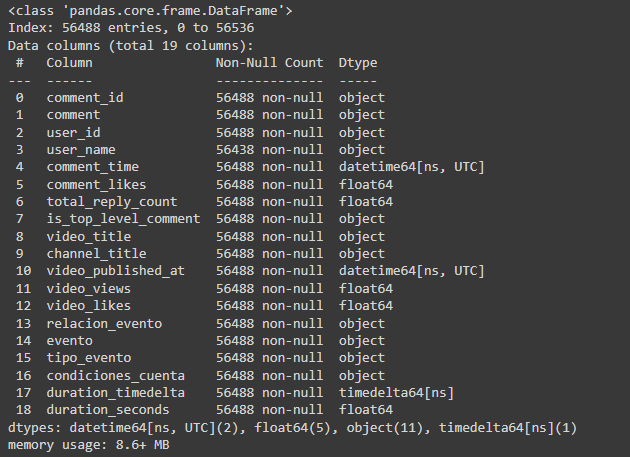


Ejemplos de los datos recolectados (encontrarán un csv adjunto, además del código utilizado)

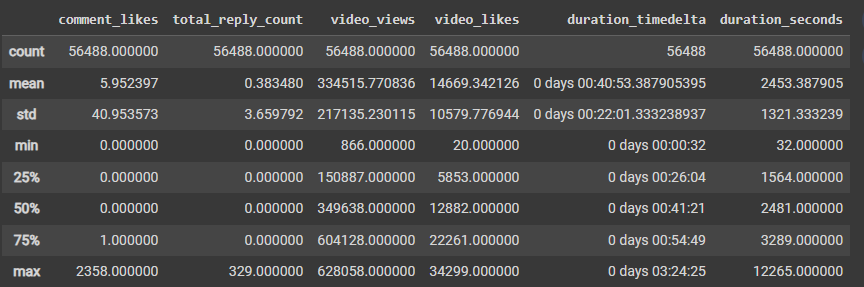


Hemos decidido trabajar juntos los datos de los videos, del canal y de los comentarios en una sóla tabla. Esto quiere decir que la tabla (dataframe) está desnormalizada.

Hecha la limpieza y las transformaciones básicas, hemos llegado al siguiente dataframe:

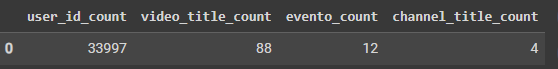


Utilizando el método .describe() hemos obtenido los resultados básicos en cuanto a las medidas de centralización y dispersión.



Podemos observar inicialmente lo siguiente:

* En promedio, son 5,96 los likes por comentario. Sin embargo, tenemos muchos comentarios con 0 likes (llegando al tercer cuartil con 1) y el máximo es de 2358, que explica la desviación estándar tan alta (40,95)
* Las respuestas a los comentarios iniciales constituyen en promedio 0,38. Y tenemos la concentración aún mayor de casos, llegando al tercer cuartil con 0 comentarios y el máximo con 329.
* Las vistas por video en promedio están en 334.515, llegando a un máximo de 628.058.
* Los likes por video empiezan desde 20 y llegan a 34.299 marcando una gran dispersión también. Esto es probable que esté afectado por la presencia de algunos videos denominados “shorts” de corta duración, como también por presencia de canales con menos público.
* La duración de los videos varía considerablemente, yendo desde 32 segundos hasta más de 3 horas. Vemos que, igualmente, prevalecen videos más largos ya que en el primer cuartil ya estamos en 26 minutos.

Realizando en conteo de datos únicos para 4 variables de interés obtenemos los siguientes resultados:  


* Tenemos 33.997 usuarios únicos (teniendo en cuenta que en total son 56.488 registros es significativa la repetición de los mismos usuarios comentando)
* 88 videos en total
* Divididos en 12 eventos
* Tomando 4 canales de youtube

**Etapa 3. Reporte de Avance | Análisis I + Documentación**

*Exploración de Datos*

Luego de descargar el dataset inicial, realizar la limpieza de datos y ver las medidas de centralización y dispersión, se inició la exploración de los datos. Nos centramos en identificar frecuencias básicas y patrones significativos en los mensajes y eventos registrados.

Inicialmente tenemos la distribución de comentarios por canal. “El Peluca Milei” siendo el más significativo, seguido por “LA NACION” y “Pais De Boludos”.



En cuanto a los eventos, el que más conversación generó es la “Crisis diplomática con España” y el “Primer paro nacional” muy cerca de la “Crisis alimentos Capital Humano”.

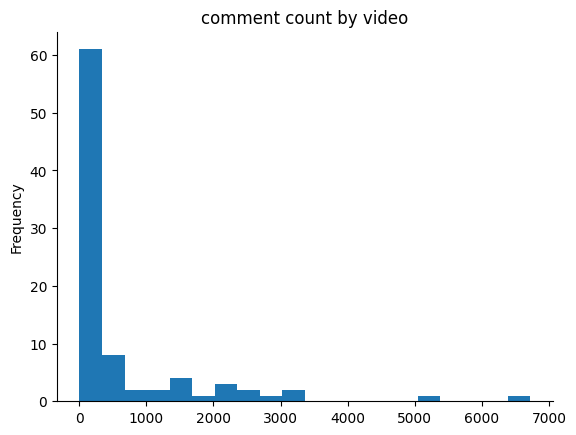


Esto se comprueba si vemos el top 10 de los videos por comentario:



Podemos observar que la mayoría son del canal líder (El Peluca Milei) y que el primer y tercer video están relacionados con la Crisis diplomática con España. Esto se debe a la exposición internacional del evento que generó interés no sólo del público nacional, sino internacional.

Siguiendo con la lógica de ver comentarios por video, observamos esta distribución graficada:



Observamos que la mayoría de los videos están entre 0 y 300 comentarios.

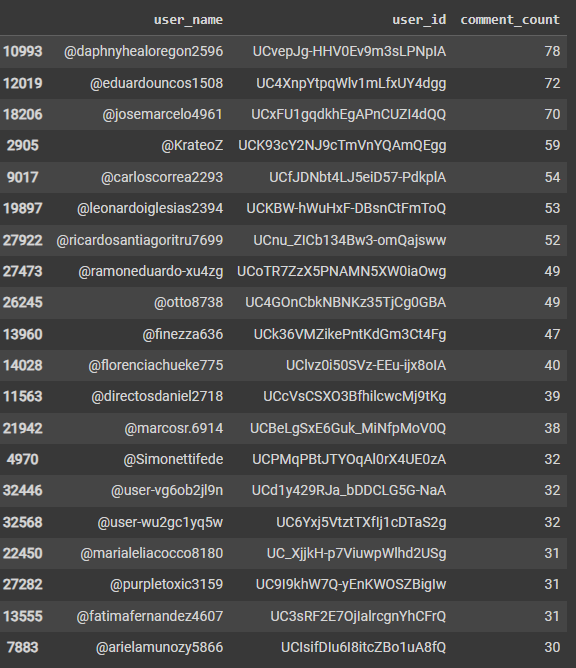
En cuanto a la distribución del tipo de evento (valoración) que se agrupó en 3 categorías, vemos que los eventos desfavorables generaron la mayor cantidad de vistas. Es importante remarcar que el evento que más tracción generó “Crisis diplomática con España” está en la categoría “Desfavorable”.



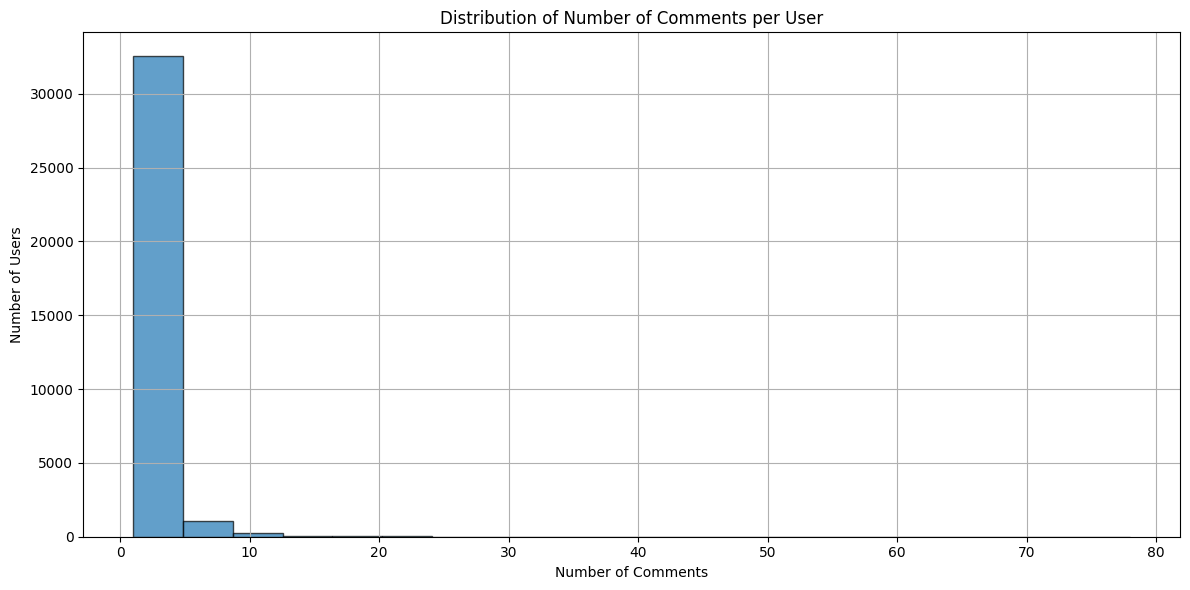
Esta distribución por evento y su tipo (valoración) se puede observar mejor la interacción en la siguiente visualización:



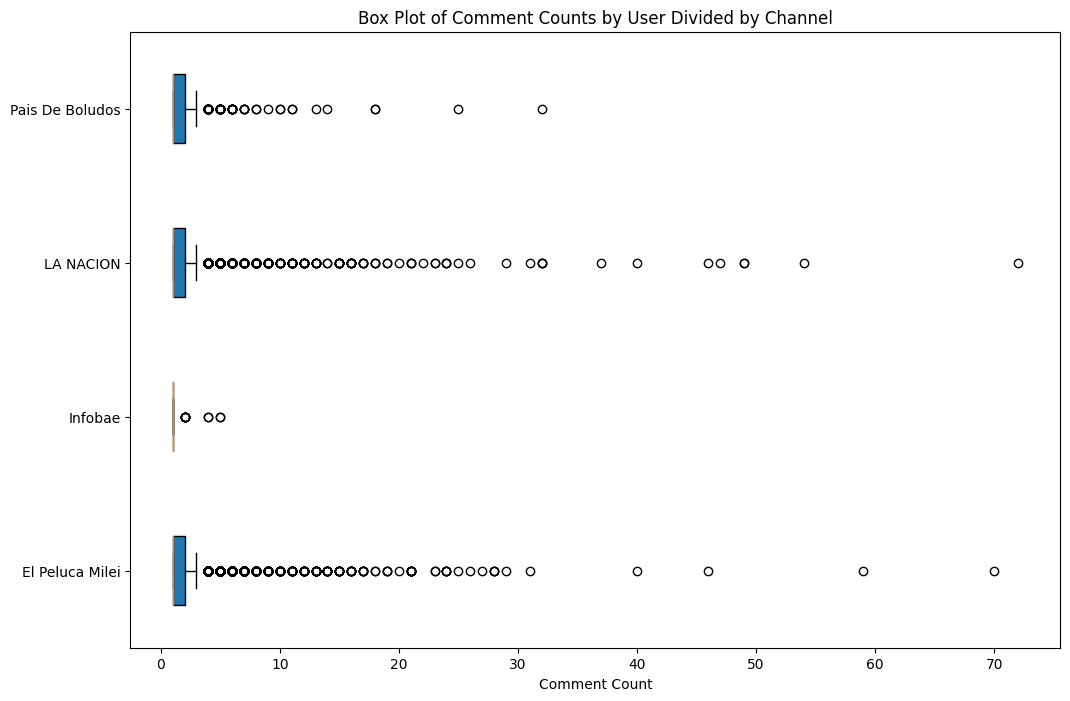
Siguiendo con las distribuciones, podemos observar que el top 20 de los/as usuarios/as por comentario tienen alta concentración de los mismos, aunque también tenemos a los mismos protagonistas de los canales como @Simonettifede en la lista.



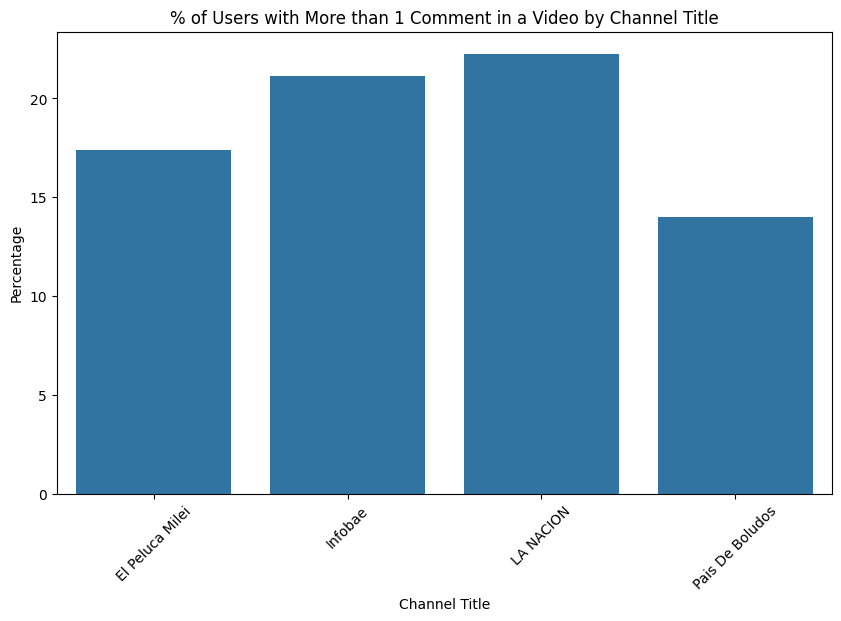
Graficando el conteo total de comentarios por usuarios/as nos encontramos ante el siguiente resultado:



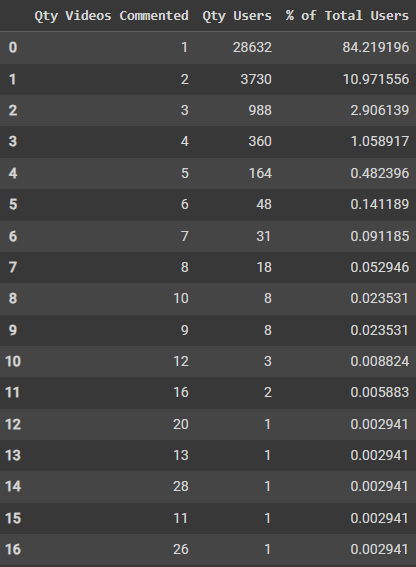
Visualizando estos resultados por canal, vemos que el en cuanto a las frecuencias absolutas el de mayor dispersión es “El Peluca Milei”, seguido por “LA NACIÓN”.



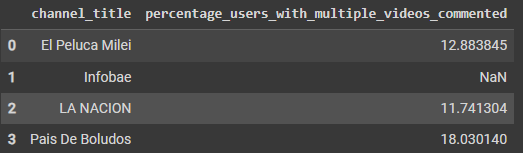
Siguiendo con los/as usuarios/as y los comentarios, podemos observar en el siguiente gráfico la proporción de usuarios/as que dejaron más de un comentario por canal. Vemos que los canales neutros, de noticias, son los que más usuarios con comentarios múltiples tienen en proporción a los/as usuarios /as que comentaron en total.

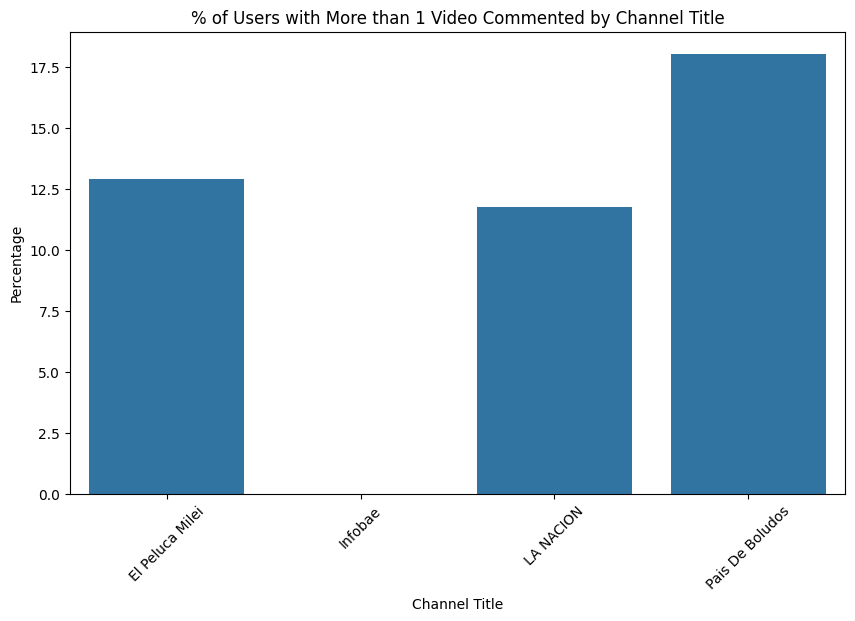


En cuanto a la distribución completa de usuarios/as, por videos comentados podemos observar que el 84,2% sólo comentaron en 1 video y casi un 11% de usuarios/as que comentaron en 2 videos. Sin embargo, aún hay alrededor de 5% que comentaron en 3 videos o más que marca una presencia más estable de estos/as usuarios/as en los canales analizados.

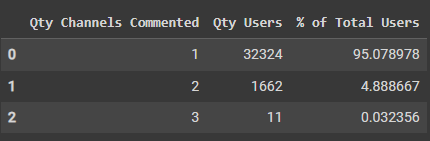


Sin embargo, si analizamos la cantidad de usuarios que dejaron comentarios en más de un video que puede representar mayor “lealtad” a la cuenta, “País de Boludos” está en el primer puesto con el 18%, seguida por “El Peluca Milei” con el 12,9%.





Y por último, en relación a los/as usuarios/as, podemos observar en la siguiente tabla que la mayoría de los/as usuarios/as comentó sólo en un canal, sin embargo, hubo 4,89% de usuarios/as que comentaron en 2 canales y 11 usuarios que comentaron en 3 (de 4 analizados). Indicando que al menos un 5% de los usuarios están mirando videos de cuentas con diferentes afiliaciones en cuanto a Milei.



*Análisis de texto y clasificación*

Se aplicaron técnicas de procesamiento de lenguaje natural (NLP) para el conteo de palabras y generación de nubes de palabras para en la siguiente etapa realizar la clasificación de mensajes en categorías de positivo, negativo y neutro. Esto nos permitió realizar una primera aproximación para comprender las tonalidades y polaridades presentes en los discursos analizados.

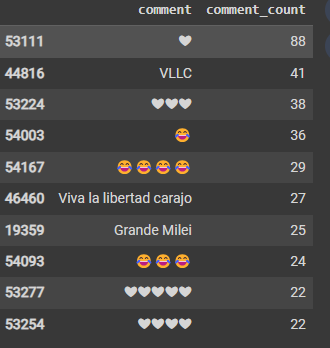
Cuenta de YT: “El Pelucla Milei”



Cuenta YT del portal digital: “La Nación”



En cuanto a la distribución de los comentarios repetidos, inicialmente obtuvimos este resultado para el top 10:

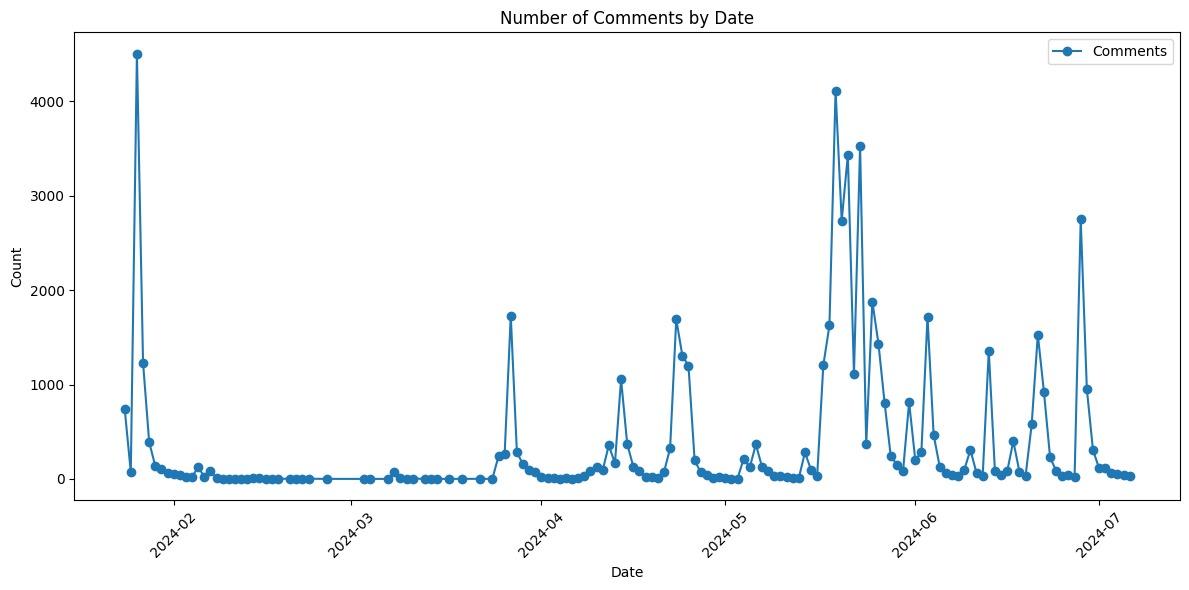
**

Sin embargo, para profundizar aún más en los mensajes, con regex sacamos los valores como emojis y comentarios de 1 sóla palabra y obtuvimos los siguientes resultados. Claramente, podemos observar que “Viva la Libertad, carajo”, en conjunto de sus variaciones es el comentario que más se repite.

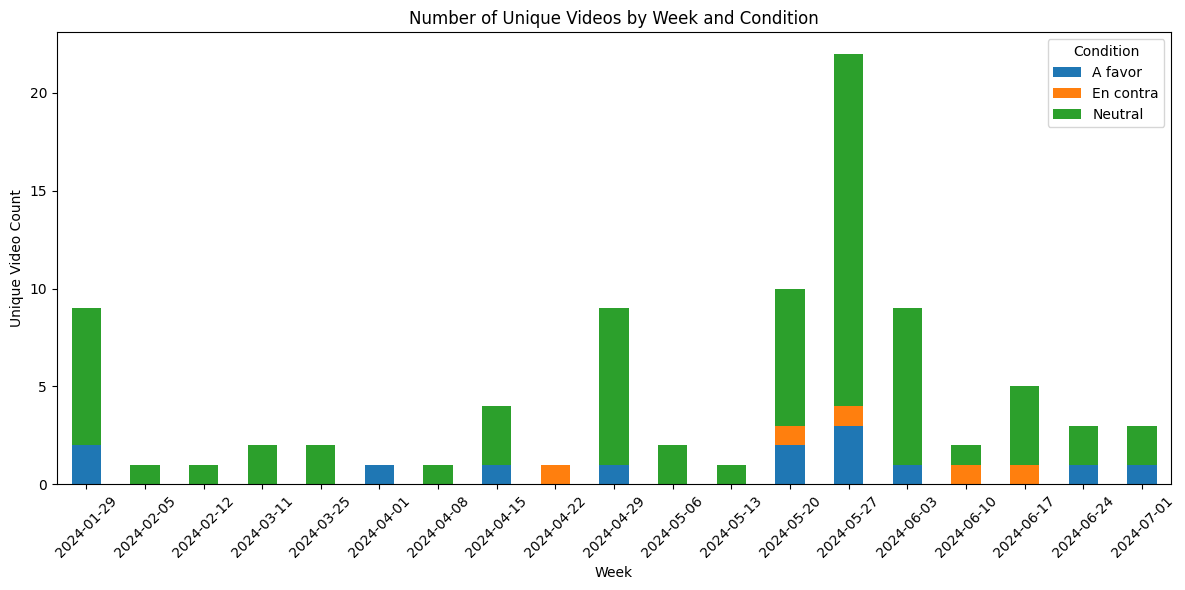


*Distribución y Visualización*

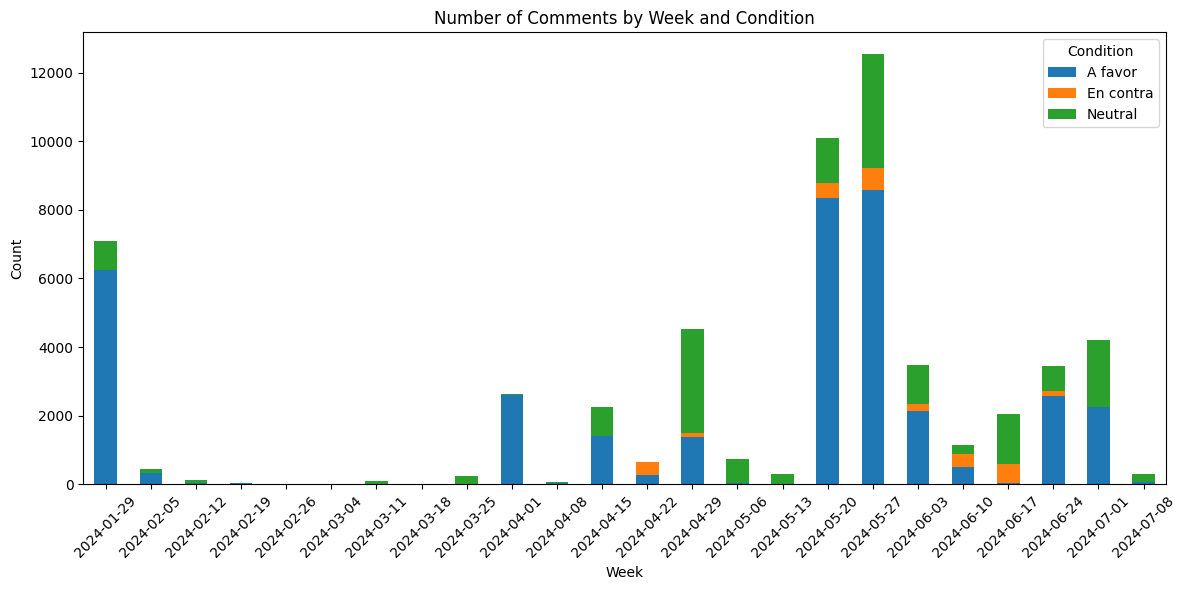
Se realizaron análisis de la distribución temporal de los mensajes, frecuencia de mensajes idénticos, y la identificación de usuarios/as recurrentes. Además, se categorizaron los eventos relevantes por etiquetas temáticas y se clasificaron según su favorabilidad (favorable, desfavorable, neutral).

****

En cuanto al conteo de videos publicados, divididos por la afiliación de la cuenta tenemos esta distribución. Aunque si bien es similar en cuanto al conteo de comentarios, hay varias diferencias notables.

****

Para explayarnos en las diferencias vamos a graficar lo mismo, pero en relación a los comentarios.



Podemos observar que las cuentas a favor, aún sin publicar alta cantidad de videos, reciben la mayor cantidad de comentarios. Por contraste, las cuentas neutrales son las que más vídeos publican, sin embargo no representan el top 1 en los comentarios.

*Visualización y Presentación de Resultados*

Finalmente, para la siguiente etapa, se buscará mejorar la implementación de técnicas de visualización de datos para graficar los resultados obtenidos, destacando las frecuencias, distribuciones y relaciones identificadas durante el análisis. Esto facilitará una aproximación para iniciar la tarea de construir una presentación, con hallazgos preliminares obtenidos.

1. Desde la presentación “etapa 1” hasta la fecha se han realizado modificaciones al corpus de trabajo a fin de facilitar y construir una base confiable y factible que permita comparar los distintos dataset a observar. Esto se detalla en “*modificaciones metodológicas”.*  [↑](#footnote-ref-0)
2. El paradestinatario es precisamente ese destinatario identificado con la “gente común”: aquella que no se interesa por la política, que no lee encuestas ni ve programas políticos, y que decide sus votos a último momento (Montero, 2018, p.46). Disponible en: Montero, A.S. (2018): “Gestionar la duda. La interpelación al paradestinatario en el discurso de Cambiemos”, Revista Mexicana de Opinión Pública. [↑](#footnote-ref-1)
3. Disponible hasta la fecha 11 de julio de 2024 en: <https://developers.google.com/youtube/v3?hl=es-419> [↑](#footnote-ref-2)