

**TIF GRUPO 13**

Tomás Deglise	DNI 41104846
Vladyslav Dodonov	DNI 94087539
Natalia Rossetti	DNI 34320233

# **El impacto de las redes sociales en el humor social: un estudio comparado de cuentas “pro” y “anti” gobierno en *YouTube*.**

**Resumen**

En la presente investigación se desea identificar y mensurar posibles cambios en el clima de opinión y en el humor social a través de un análisis comparado en redes sociales, específicamente en YouTube.

**Planteo de problema**

Se buscará rastrear patrones de cambios de opinión y humor social con mayor precisión y menor sesgo, asumiendo que cambios similares en los tres grupos indican una tendencia general más allá de diferencias ideológicas. Esta investigación se enfoca en tres grupos de cuentas:

- Cuentas de medios de comunicación digitales: La Nación e Infobae.
- Cuentas pro-gobierno: El Peluca Milei y Urgente Milei.
- Cuentas contra el gobierno: País de Boludos y Rebelión Popular.

**Objetivos**

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).

### **Hipótesis de investigación**

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

Habrà 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.

### **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* en momentos y fechas específicas que serán escogidas para este trabajo, a fin de comparar la conducta en grupos a favor y en contra. A estos canales los definiremos como “otros”, ya que es un espacio de interacción donde usuarios/as pueden tener una mirada a favor, indistinta o en contra sobre la agenda pública del actual gobernante. Cabe aclarar, que consideramos que pueden existir sesgos por la misma dinámica que los medios masivos ya que son actores de influencia y referencia en la arena política.

## Metodología

Se recopilaron comentarios de usuarios/as de YouTube de cada canal que agrupamos en tres posiciones en relación al gobierno actual. La selección de canales “otros” fueron los portales de noticias (La Nación e Infobae) debido a la ausencia de un canal oficial del presidente con suficiente actividad.

Cuentas de YouTube de acuerdo a su influencia y relevancia en la temática política:

Condición	Plataforma social	Usuario	Seguidores
A favor del gobierno	YouTube	El Peluca Milei	1,7 mill.
		Urgente Milei	573k
Otros	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 “otros” 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 necesariamente tienen un posicionamiento ideológico de pertenencia o sí, pero se encuentran en interacción con posibles usuarios/as que tiene una postura opuesta o similar. Por lo tanto, muchos/as 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 este modo, se entiende que éstas cuentas están más abiertas a un público “general”.

Los eventos se clasificaron como desfavorables, neutrales y favorables para medir cambios de opinión en diferentes contextos.

*Eventos desfavorables:* considerados como hechos públicos que la comunicación oficial del gobierno clasificó como acciones negativas para la agenda de gobierno. Asimismo, son manifestaciones en la vía pública o por la opinión pública que pueden repudiar medidas políticas-económicas y la agenda de gobierno del actual presidente.

- Primer paro nacional
- Marcha federal universitaria
- Crisis diplomática con España
- Polémica por alimentos del Ministerio de Capital Humano
- Fórmulas de jubilaciones

*Eventos neutrales:* considerados como sucesos públicos que la opinión pública puede tener una evaluación diversa sobre estos.

- Efemérides políticas y sociales (8 de marzo, 24 de marzo, 25 de mayo, 20 de junio)

*Eventos favorables:* considerados como hechos públicos que la comunicación oficial del gobierno clasificó como acciones positivas para la agenda de gobierno. Asimismo, son manifestaciones en la vía pública o por la opinión pública que pueden aprobar las medidas políticas-económicas y la agenda de gobierno del actual presidente.

- Aprobación de la Ley bases
- Reunión del Presidente con Elon Musk
- Anuncio de reducción de la inflación
- Presentación del libro del Presidente
- Renuncia a la jubilación al cumplir el mandato

Es relevante destacar que el recorte temporal entre enero de 2024 e inicios de agosto de 2024 se debe a la distribución de los eventos de interés. Este recorte temporal también está dictado por las dinámicas mismas de los canales elegidos, que normalmente publican contenido audiovisual cercano a los eventos de interés público para capturar la mayor cantidad de interacciones, y por lo tanto, relevancia en el espacio social e informativo.

## **Descripción y definición de variables**

### **Variables**

- *Likes por comentario*: número de "me gusta" obtenidos por cada comentario en los videos. Este mide el nivel de aprobación o apoyo que recibe cada comentario de otros usuarios.

```
comment_likes      56488 non-null float64
```

- *Respuestas por comentario*: promedio de respuestas que recibe cada comentario. Mide el grado de interacción y discusión generada por cada comentario.

```
total_reply_count  56488 non-null float64
```

- *Vistas por video*: número total de visualizaciones que tiene cada video. Señala el alcance y la popularidad del contenido del video.

```
video_views
```

- *Comentarios por video*: número total de comentarios que tiene cada video. Subraya la relevancia del video, pero también el espacio para la expresión que presenta.

```
comment_count
```

- *Likes por video*: número total de "me gusta" recibidos por cada video. Indica el nivel de aprobación general de los espectadores sobre el contenido del video.

```
video_likes
```

- *Duración de los videos*: tiempo total de cada video, medido en segundos. Mide y permite analizar cómo la longitud del video puede influir en las interacciones y el compromiso de los usuarios.

```
duration_timedelta 56488 non-null timedelta64[ns]
```

- *Usuarios únicos*: número total de usuarios diferentes que han comentado en los videos. Indica la diversidad y la cantidad de personas participando en las discusiones.

```
user_id
```

```
user_name
```

- *Número de videos*: cantidad total de videos analizados en el estudio. Indica la base de datos utilizada para el análisis y la representatividad del estudio.

```
video_title_count
```

- *Tipo de eventos y número de eventos*: cantidad total de eventos políticos relevantes considerados en el análisis. Permite clasificar y segmentar el análisis y evaluar la reacción de los usuarios ante diferentes situaciones políticas.

```
evento
```

```
tipo_evento
```

- *Número de canales:* cantidad total de canales de YouTube incluidos en el estudio. Permite observar y analizar el alcance y la variedad de fuentes de información analizadas.

`channel_title_count`

### Indicadores para la etapa de análisis

#### *Dimensión: Humor social*

Indicadores:

- Análisis de sentimiento
- Detección de sarcasmo
- Utilización de emojis

#### *Dimensión polarización*

Indicadores:

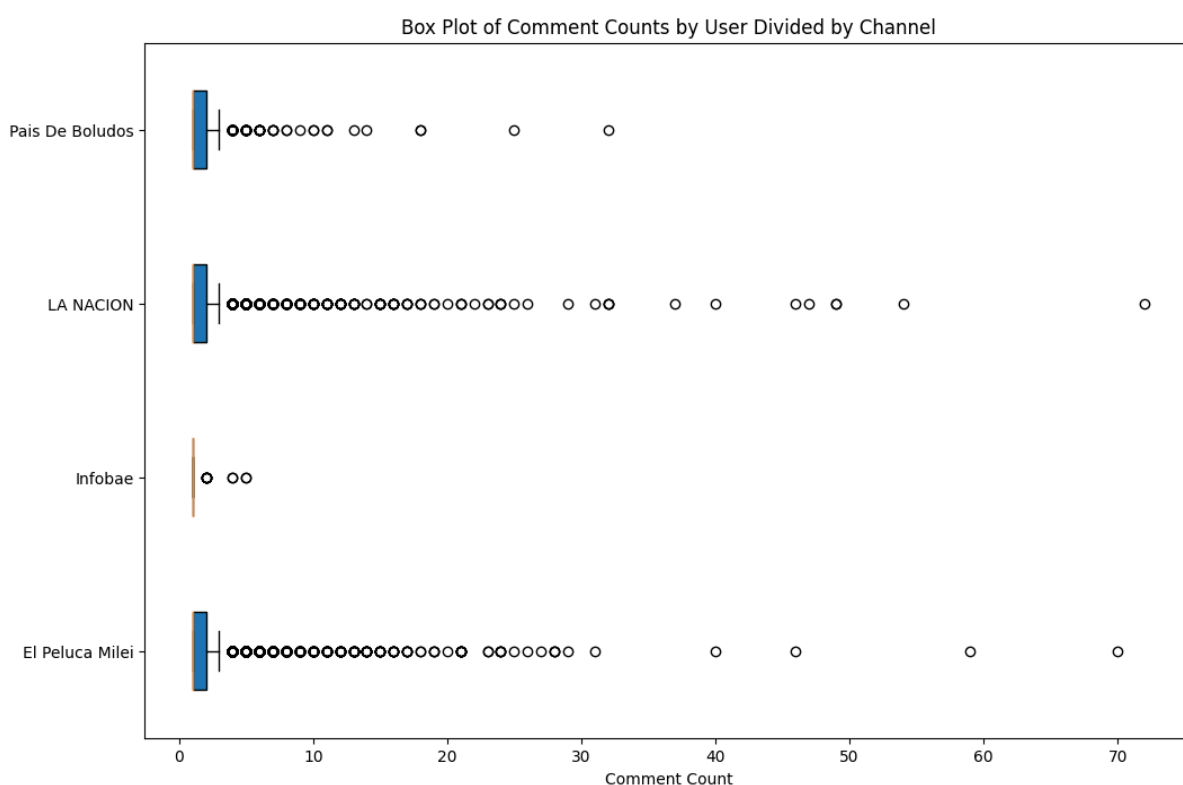
- Unidades lingüísticas que identifican a un grupo (por ejemplo: “viva la libertad”, “VVLLC”, etc.)
- Unidades lingüística que representen ideas (por ejemplo: “ñoquis”, “presiduende”, etc.)
- Análisis de sentimiento
- Detección de sarcasmo
- Utilización de emojis

### Identificación de outliers

Para el trabajo que se desea presentar hemos avanzado en el análisis de medidas de centralización y dispersión, así como de la visualización de distribuciones y gráficos relacionados con los datos recolectados. A continuación, se detallan los outliers identificados en diferentes variables:

- *Likes por comentario:* comentarios con 2358 "me gusta" frente a un promedio de 5.96 y una desviación estándar de 40.95. Esto nos indica la presencia de outliers en la cola derecha de la distribución.
- *Respuestas por comentario:* comentarios con hasta 329 respuestas, muy por encima del promedio de 0.38. Esto nos indica outliers extremos en la cantidad de respuestas.
- *Vistas por video:* videos con hasta 628,058 visualizaciones en comparación con un promedio de 334,515. Esta dispersión significativa apunta a outliers en las visualizaciones.

- Likes por video: videos con hasta 34,299 "me gusta", muy superiores al mínimo de 20 y al promedio calculado. La amplia dispersión sugiere la presencia de outliers.
- *Duración de los videos*: videos que duran más de 3 horas frente a un primer cuartil de 26 minutos, indicando outliers en la duración de los videos.
- Comentarios promedio por usuario y el canal: tenemos una distribución concentrada alrededor de 1,66 comentarios por usuario. Vemos, en el siguiente gráfico, la distribución y sus outliers.



## Revisión del Dataset

### Distribución de Comentarios por Canal:

Se observó que el canal "El Peluca Milei" es el más significativo en términos de cantidad de comentarios, seguido por "LA NACION" y "País de Boludos".

### Conteo de Datos Únicos:

- Usuarios únicos: 33,997
- Total de registros: 56,488
- Total de videos: 88
- Total de eventos: 12

- Total de canales: 4

## **Medidas de Centralización y Dispersión:**

### *Likes por comentario:*

Promedio: 5.96

Mediana: 1

Máximo: 2358

Desviación estándar: 40.95

### *Respuestas a comentarios:*

Promedio: 0.38

Mediana: 0

Máximo: 329

### *Vistas por video:*

Promedio: 334,515

Máximo: 628,058

### *Likes por video:*

Rango: 20 a 34,299

### *Duración de videos:*

Rango: 32 segundos a más de 3 horas

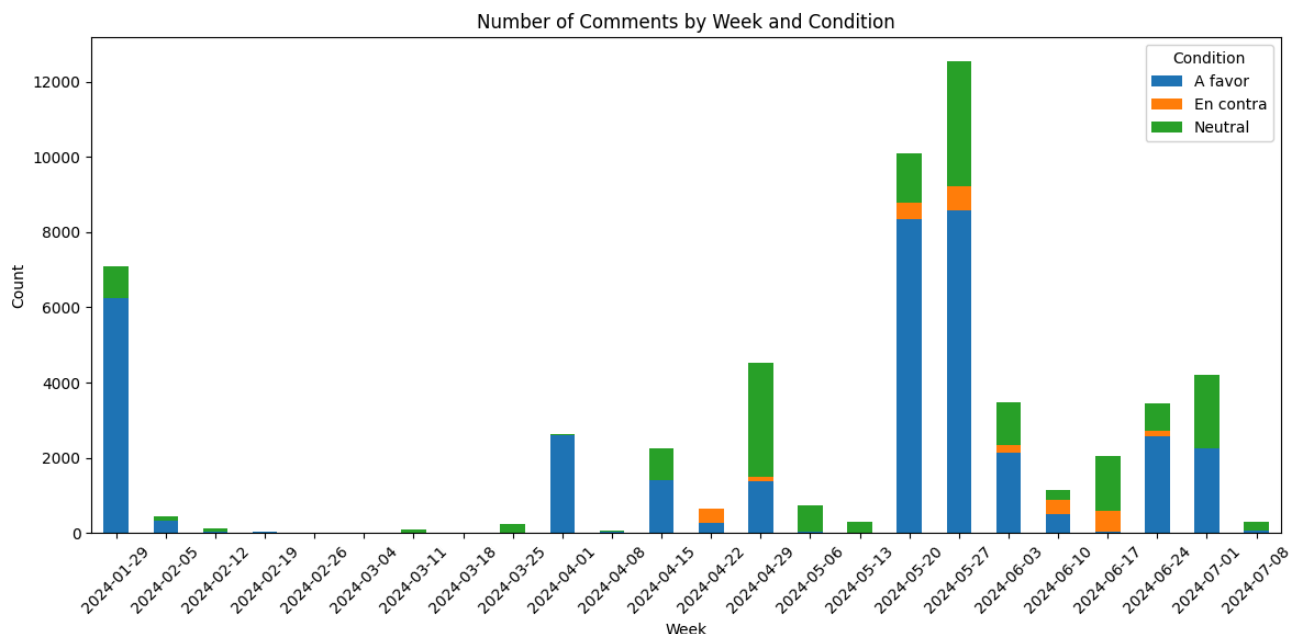
Primer cuartil: 26 minutos



## Visualizaciones

### 1. Distribución Temporal de los Mensajes:

Se realizaron gráficos de la distribución temporal de los mensajes para identificar patrones de actividad.



### 2. Frecuencia de Mensajes Idénticos:

Se identificaron frecuencias y patrones en los comentarios repetidos (raw y limpio de emojis + 1 palabra).

Raw. Top 10:

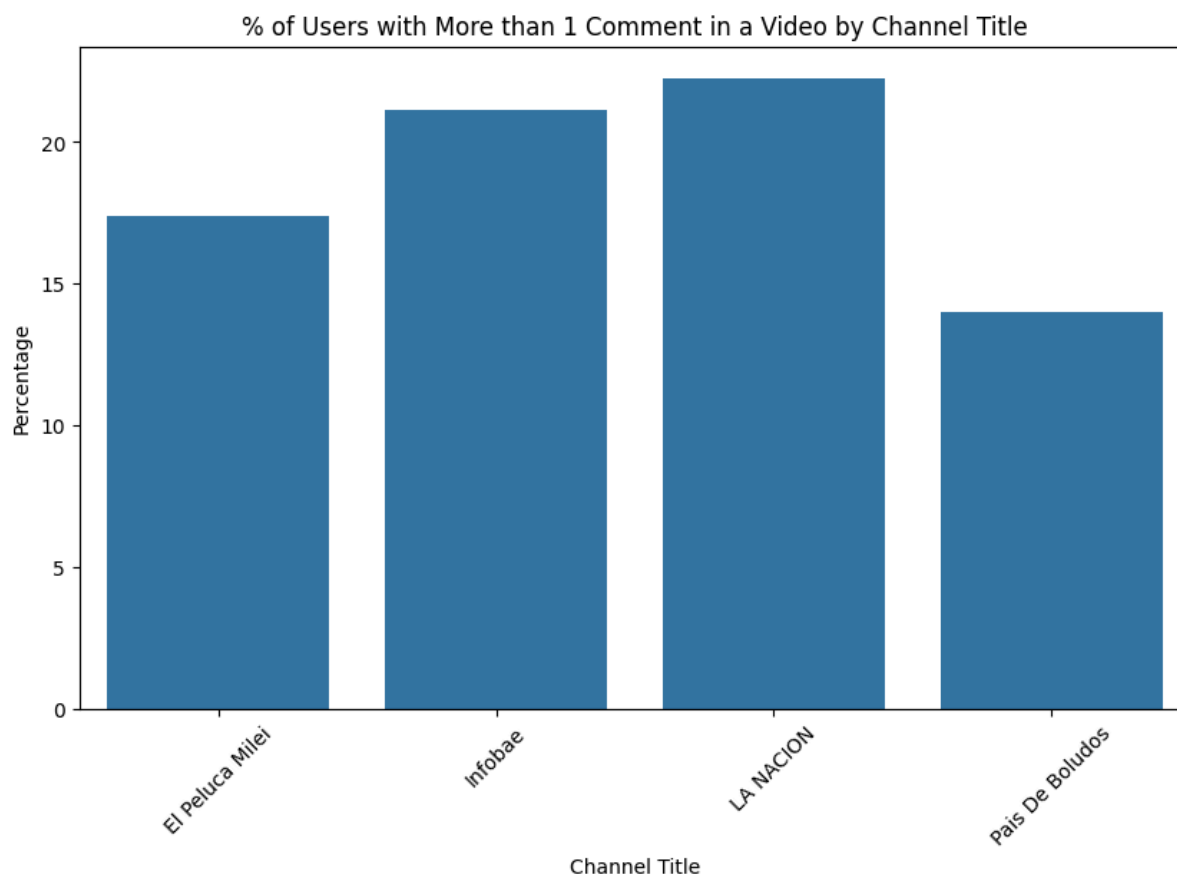
comment	comment_count
♥	88
VLLC	41
♥♥♥♥	38
😂	36
😂😂😂😂	29
Viva la libertad carajo	27
Grande Milei	25
😂😂😂	24
♥♥♥♥♥♥	22
♥♥♥♥♥	22

Limpio. Top 20:

comment	comment_count
Viva la libertad carajo	27
Grande Milei	25
RESOLVE LA INSEGURIDAD PELOUDO NO TE VALLAS A ...	17
para mi es un payaso hablan tantp de los rojos...	15
Jhon Perkins confesiones de un Ganster Econòmi...	14
VIVA LA LIBERTAD CARAJO	13
Viva la libertad	12
Fuera kircho!!!!!!!!!!!! vaya con su &quot;cond...	11
Grande milei	11
Viva Milei	11
EL CONO SUR ES UN PAÍS  AMSTRALIA  AMERI...	10
Nuestro &quot;PINOCHO&quot; (milei) se olvido ...	10
Viva la libertad carajo!	10
Y encima la piba después dice que no hay adoct...	9
Gracias Milei	8
<a href="http://www.youtube.com/results?search...	8
Viva la libertad carajo!!!	8
VAMOS MILEI	8
Vamos Milei	8
VIVA LA PATRIA CARAJO 🇦🇷 🦍 VIVA LA PATRIA CARA...	7

### 3. % de Usuarios con comentarios Recurrentes:

Se analizó la recurrencia de usuarios y usuarias que comentan frecuentemente.



#### 4. Distribución de Comentarios por Afiliación de Cuenta:

Se graficó la cantidad de comentarios en relación con la cantidad de videos publicados por cuentas a favor, neutrales y en contra.

channel_title	comment_count
El Peluca Milei	36925
Infobae	73
LA NACION	16639
Pais De Boludos	2851

#### 5. Análisis de Texto y Clasificación:

NLP para Conteo de Palabras y Nubes de Palabras:

Se generaron nubes de palabras para visualizar las palabras más frecuentes.

Clasificación de Mensajes:

Los comentarios se clasificaron en categorías de positivo, negativo y neutro para analizar las tonalidades y polaridades de los discursos.

## Discusión de la selección del modelo y su implementación (experimentos / pruebas) y métricas

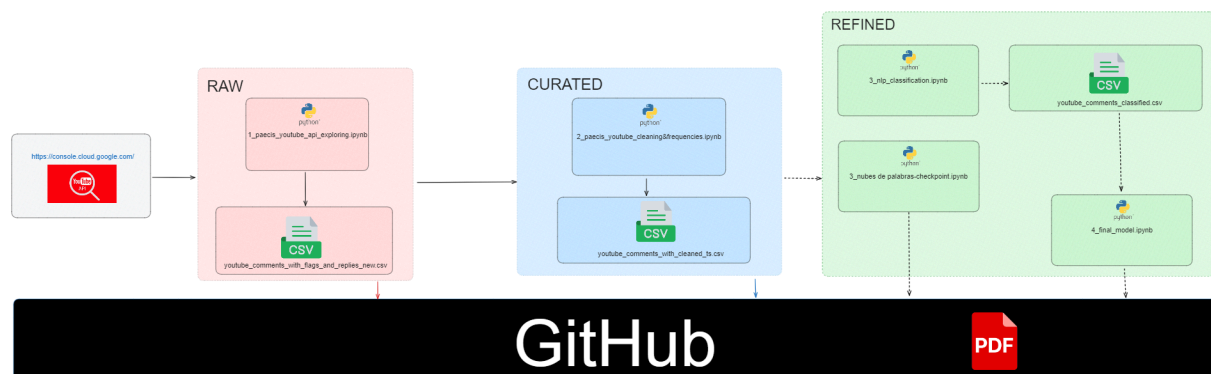
(Esta etapa está en proceso de elaboración)

### Explicación de pipeline de datos

Nuestra primera versión del pipeline va a ser polietápica, dónde vamos trabajando los outputs en formato de .csv de la etapa anterior. Finalmente, estaríamos llegando a la zona “Refined” dónde estará el output final de la clasificación NLP y un modelo descriptivo en cuanto a la relación de las variables. Actualmente estamos en la etapa 3 de dicho pipeline.

Todos estos archivos ya se encuentran en un repositorio disponible al público: [https://github.com/vladyslavbdhm/paecis\\_grupo\\_13](https://github.com/vladyslavbdhm/paecis_grupo_13)

Una vez finalizada esta primera versión, se harán los ajustes para reorganizar el código siguiendo las indicaciones vistas en clase.



### Descripción del código fuente utilizado para conducir el análisis

Actualmente tenemos 3 ipynb que trabajan distintas etapas del pipeline descrito en el punto anterior.

- 1\_paecis\_youtube\_api\_exploring.ipynb
- 2\_paecis\_youtube\_cleaning&frequencies.ipynb
- 3\_nubes de palabras-checkpoint.ipynb

En la etapa uno (*1\_paecis\_youtube\_api\_exploring.ipynb*), estamos utilizando la API Youtube v3.0 para traer la información de interés desde youtube. Hacemos algunos ajustes simples en cuanto a la data y la desnormalización de las tablas (videos, comentarios) para poder trabajarla en un sólo dataset.

En la segunda etapa (*2\_paecis\_youtube\_cleaning&frequencies.ipynb*) realizamos la limpieza del dataset, creamos las columnas calculadas y realizamos el primer análisis exploratorio de los datos. Eso nos permite entender la información y realizar ajustes y consideraciones necesarias en cuanto a su calidad. Aplicamos los primeros análisis estadísticos, de dispersión y también las medidas centrales.

Finalmente, en la etapa tres (*3\_nubes de palabras-checkpoint.ipynb*) realizamos la primera aproximación al NLP donde creamos las nubes de palabras que representan, de forma exploratoria, las palabras más usadas por canal.

Estaremos creando los últimos dos códigos dónde se profundizará la aplicación de las técnicas en el contexto de NLP para clasificar los comentarios en cuanto a su intención (positivo, neutro, negativo) y su dirección (hacia Milei, hacía los otros usuarios, hacía el creador del vídeo/canal). Por último, estaremos cerrando la investigación con un modelo de clasificación para poder responder las preguntas planteadas al inicio de esta investigación.

## Interpretación de resultados y su relación con los objetivos iniciales

Compartimos algunas interpretaciones de resultados que hemos observado hasta la fecha pero aún están en proceso de revisión y deben ser puestos nuevamente a prueba para evitar sesgos de interpretación.

### 1. Cambios de Opinión y Humor Social:

- *Likes y Respuestas*: la dispersión en likes por comentario y respuestas sugiere que algunos comentarios generan reacciones muy fuertes, lo cual podría reflejar cambios de opinión significativos en ciertos segmentos de la audiencia.
- *Vistas y Likes por Video*: los videos relacionados con eventos desfavorables tienen más vistas y likes, lo que podría indicar un mayor interés y posible cambio de humor social en respuesta a estos eventos..
- *Comentarios por video y Vistas por video*: la distribución muestra que ciertos videos tienen una alta cantidad de visualizaciones y comentarios, reflejando mayor interés en temas específicos.
- *Duración de los videos*: La longitud de los videos no parece tener una relación directa con la cantidad de interacciones, ya que hay videos largos y cortos con alta participación.

### 2. Polarización:

- *Unidades lingüísticas y Análisis de Sentimiento*: se identificaron frases y términos recurrentes que evidencian una clara polarización, como "viva la libertad carajo" y "emojis" para apoyar al gobierno y términos despectivos para criticarlo.

- La clasificación de comentarios en positivo, negativo y neutro, junto con las nubes de palabras, permitió identificar términos recurrentes y tópicos que generan reacciones polarizadas. Términos asociados con crisis diplomáticas, paros nacionales y políticas alimentarias fueron particularmente destacados.
- 3. Granjas de Usuarios de Opinión:
  - Usuarios/as recurrentes: la posible identificación de usuarios/as recurrentes en los comentarios sugiere la posible existencia de granjas de usuarios o grupos organizados que buscan influir en la opinión pública. Este es más evidente en canales con fuerte inclinación política, como "El Peluca Milei".

### **Aportes, posibles aplicaciones, futuros pasos**

Finalmente, para la entrega de este trabajo final integrador, 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.

Se buscará profundizar el trabajo en para la obtención de resultados de cambios de opinión en sentido analítico, a fin de explorar si hay focos de polarización y así poder corroborar o contrastar nuestra hipótesis de trabajo.