

Videos virales: Factores que contribuyen a la viralidad

Juan Torres, Juliana Rubio y Julián Páez

Noviembre 18 del 2024.

Pontificia Universidad Javeriana

Departamento de Comunicación

Narrativa de Datos

Introducción de la problemática analizada

En un mundo donde el consumo de contenido digital aumenta exponencialmente, YouTube se ha consolidado como una de las principales plataformas de distribución de videos. Sin embargo, entender por qué algunos videos logran captar la atención de millones de usuarios mientras otros pasan desapercibidos sigue siendo un desafío. La viralidad en YouTube no solo depende de la calidad del contenido, sino también de factores estratégicos como las categorías, etiquetas, títulos, descripciones, y el momento de publicación.

Para creadores de contenido, empresas y estrategias de marketing, identificar los elementos clave que contribuyen a la viralidad de un video puede marcar la diferencia entre el éxito y el fracaso. En este análisis, utilizaremos datos recolectados a través de la API de YouTube desde el 15 de octubre hasta el 15 de noviembre de 2024 y herramientas como BERTopic para explorar los factores que hacen que un video se vuelva viral, incluyendo los tópicos más relevantes y las dinámicas de interacción.

Pregunta de investigación

¿Qué factores contribuyen a la viralidad de un video en YouTube, y cómo se relacionan las características del contenido con la interacción de los usuarios y los patrones de visualización?

Esta pregunta se delimita en torno a los siguientes aspectos:

1. **Categoría de Contenido:** Identificar las categorías y temas recurrentes entre los videos virales.
2. **Etiquetas (Tags):** Examinar cómo las etiquetas y la calidad del video influyen en la viralidad.
3. **Impacto del Tipo de Video:** Explorar patrones específicos en videos aptos para niños.
4. **Títulos y Descripciones:** Analizar la longitud, términos frecuentes y estructura de los títulos y descripciones.
5. **Análisis Temporal:** Determinar patrones de publicación efectivos para alcanzar la viralidad.
6. **Tópicos con BERTopic:** Identificar los temas predominantes y su relevancia en los videos virales.

Con este análisis, buscamos no solo identificar qué hace que un video sea viral, sino también proporcionar información procesable para optimizar la creación y publicación de contenido en YouTube.

Metodología: Herramientas

Para llevar a cabo este análisis de los factores que contribuyen a la viralidad de videos en YouTube, se emplearon una serie de herramientas y metodologías específicas que permitieron la recolección, limpieza, análisis y visualización de datos. A continuación, se detalla el proceso y las herramientas utilizadas en cada etapa:

1. Recolección de datos

- **API de YouTube:** Se utilizaron datos recolectados exclusivamente de videos en tendencia (“trending videos”) mediante la API de YouTube. Esto permitió acceder a información detallada de videos relevantes, incluyendo vistas (view_count), likes (like_count), comentarios (comment_count), etiquetas (tags), títulos (title), descripciones (description), duración (duration), idioma detectado (detected_language), y otros atributos.
- **Filtro de idioma:** Los datos recolectados se limitaron a videos en idioma inglés (detected_language == "en"), lo que garantizó coherencia en el análisis textual.
- **Bibliotecas utilizadas:**
 - pandas para almacenar y manipular los datos tabulares.
 - googleapiclient (en Python) para interactuar con la API de YouTube.

2. Limpieza y Preprocesamiento de Datos

- **Limpieza de datos nulos y vacíos:**
 - Se eliminaron columnas irrelevantes o vacías, como dislike_count.
 - Se manejaron valores faltantes en columnas como tags, title y description, asignando valores por defecto o eliminando registros incompletos.
- **Preprocesamiento de texto:**
 - Uso de nltk para eliminar stopwords, aplicar lematización, y limpiar texto.
 - Conversión de texto a minúsculas y eliminación de caracteres especiales mediante expresiones regulares (re).
 - Reemplazo de valores monetarios y numéricos con etiquetas genéricas (Money, NUM) para homogenizar los textos.

3. Análisis Exploratorio de Datos (EDA)

- **Análisis descriptivo:**

- Se utilizaron gráficos como histogramas (matplotlib) para analizar la distribución de vistas, likes y comentarios.
- Cálculo de estadísticas descriptivas (mean, median) para evaluar las métricas clave de los videos.
- **Análisis temporal:**
 - Conversión de published_at a formato datetime y extracción de información como horas (hour) y días de la semana (day_of_week).
 - Gráficos de barras y líneas para identificar patrones temporales en las publicaciones.
- **Categorías de contenido:**
 - Se analizaron las categorías (category) más comunes y con mayor promedio de vistas, likes y comentarios.
 - Gráficos de barras horizontales para visualizar las categorías dominantes.

4. Análisis de Etiquetas y Títulos

- **Etiquetas (Tags):**
 - Extracción y limpieza de etiquetas usando combinaciones únicas (combinations de itertools) para analizar pares, tríos y cuatrigramas frecuentes.
 - Gráficos de barras horizontales para visualizar las etiquetas más comunes.
 - Cálculo del número promedio de etiquetas por video y su distribución.
- **Títulos y Descripciones:**
 - Análisis de longitud de títulos y descripciones mediante histogramas.
 - Identificación de las palabras y frases más comunes utilizando Counter de Python.
 - Generación de nubes de palabras (wordcloud) para visualizar términos frecuentes.
 - Identificación de títulos que terminan en preguntas (?) y su frecuencia relativa.

5. Análisis de Sentimientos

- **TextBlob:**
 - Se empleó la librería TextBlob para analizar el sentimiento de los títulos, categorizando los resultados en positivo, negativo, y neutral.
 - Gráficos de distribución (histogramas y gráficos de pastel) para representar las tendencias de sentimiento.
- **Visualización de la relación con la viralidad:**
 - Comparación entre puntajes de sentimiento y métricas como likes y vistas.

6. Análisis de Duración y Tiempos de Publicación

- **Conversión de duración:**
 - Uso de expresiones regulares (re) para convertir la duración en formato ISO 8601 (PT2H30M15S) a minutos y segundos.
 - Análisis de la relación entre la duración promedio de los videos y métricas de interacción.
- **Frecuencia de publicación:**
 - Conteo de videos publicados por hora y día de la semana, visualizado mediante gráficos de barras.

7. Modelado de Tópicos con BERTopic

- **BERTopic:**
 - Uso de la librería bertopic para realizar análisis de tópicos en los títulos y etiquetas de los videos.
 - Preprocesamiento adicional:
 - Concatenación de títulos (title) y etiquetas (tags) para enriquecer el contexto de los textos analizados.
 - Vectorización de texto mediante embeddings (all-MiniLM-L6-v2).
 - Visualizaciones generadas:
 - Gráfico de barras con los temas más importantes (visualize_barchart).
 - Gráfico de relaciones entre tópicos (visualize_topics).

- **Resultados de tópicos:**

- Identificación de los temas más predominantes en los videos virales.
- Análisis de temas similares mediante comparación de similitud (find_topics).

Herramientas y Bibliotecas

1. Python:

- pandas, numpy para manipulación de datos.
- matplotlib y seaborn para visualización.
- re y nltk para preprocesamiento de texto.
- TextBlob para análisis de sentimientos.
- itertools para análisis de combinaciones de etiquetas.

2. BERTopic:

- Modelado de tópicos basado en embeddings.
- all-MiniLM-L6-v2 como modelo de embeddings semánticos.

3. WordCloud:

- Generación de nubes de palabras para análisis visual de términos frecuentes.

4. YouTube API:

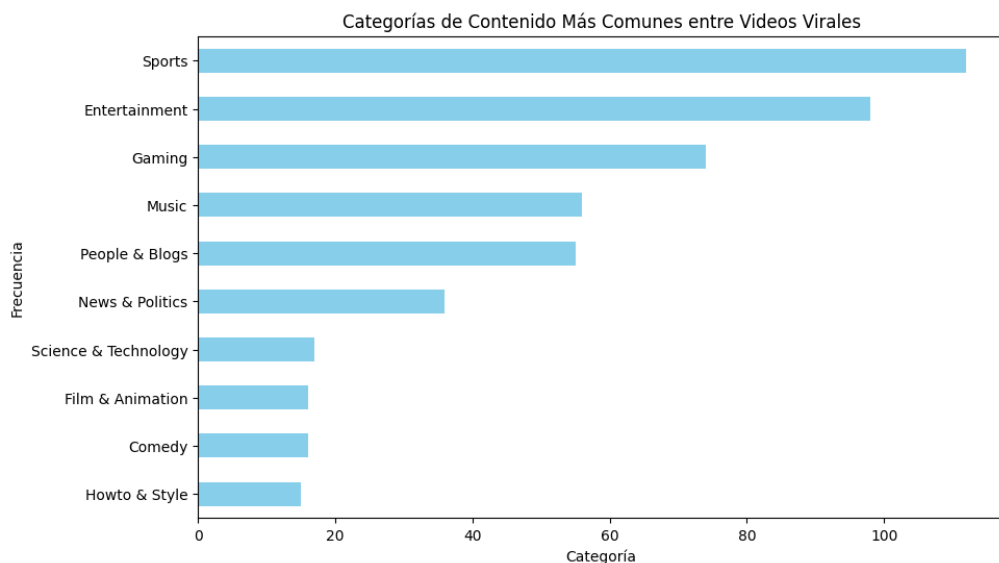
- Extracción de datos de videos en tendencia directamente desde la plataforma.

Resultados

Es necesario dar respuesta a cada una de las preguntas planteadas

1. Categoría de Contenido

¿Cuáles son las categorías de contenido más comunes entre los videos virales?



La gráfica muestra que las categorías más comunes entre los videos virales son:

1. **Sports:** Es la categoría predominante, indicando que los deportes generan un gran interés en términos de cantidad de videos virales.
2. **Entertainment:** El entretenimiento ocupa el segundo lugar, destacándose como una categoría general que atrae a una amplia audiencia.
3. **Gaming:** Los videojuegos son otro tema recurrente en los videos virales, reflejando su popularidad entre las audiencias de YouTube.
4. **Music:** Los videos musicales también representan una parte significativa, dada la amplia base de fanáticos de este tipo de contenido.
5. **People & Blogs:** Los videos centrados en experiencias personales o estilo de vida son muy comunes.

¿Qué categorías suelen tener videos con mayor cantidad de vistas promedio?

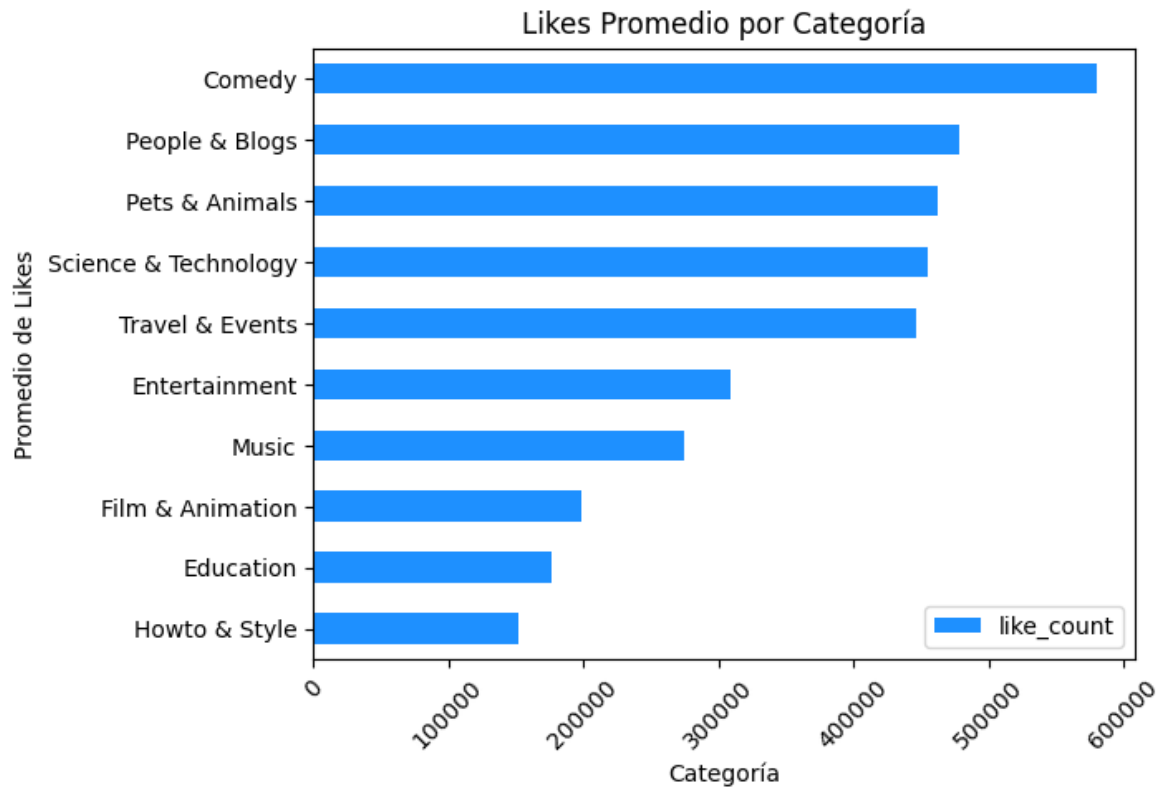


De acuerdo con la gráfica de promedios de vistas por categoría:

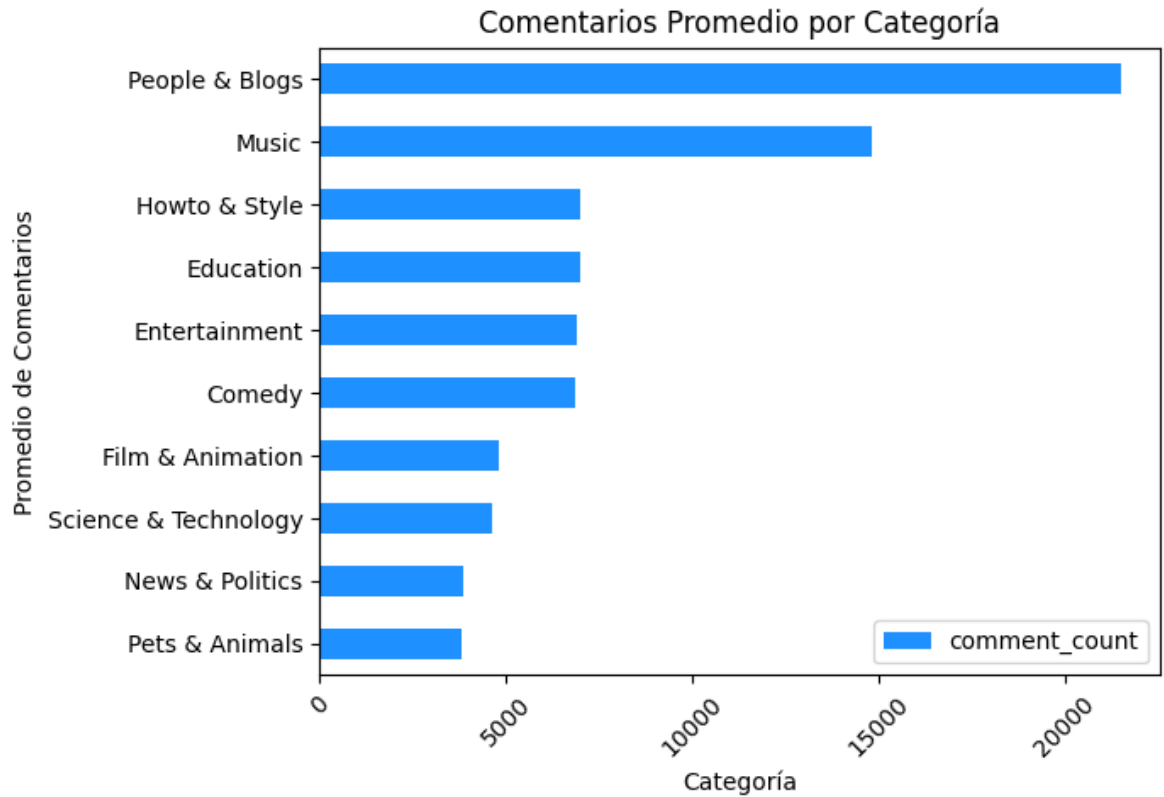
- **Comedy** lidera con la mayor cantidad de vistas promedio, indicando que el humor es altamente efectivo para atraer audiencias masivas.
- **Travel & Events** tiene un promedio alto, posiblemente debido al atractivo visual y la capacidad de este tipo de contenido para inspirar curiosidad.
- **People & Blogs** también obtiene un alto promedio de vistas, lo que refuerza su relevancia entre los videos virales.
- **Pets & Animals** genera gran interés, lo cual no sorprende dada la popularidad de videos relacionados con animales.
- **Science & Technology** aparece en la lista con un promedio considerable, reflejando el creciente interés en contenido educativo y tecnológico.

¿Qué categorías generan mayor interacción por parte de los usuarios (likes y comentarios)?

Los gráficos muestran:



- Comedy genera el mayor número promedio de likes, indicando que el humor resuena fuertemente con la audiencia.
- People & Blogs y Pets & Animals también tienen un alto promedio de likes, sugiriendo que los videos personales y de animales son emocionalmente atractivos.
- Science & Technology y Travel & Events también destacan, mostrando que estos temas generan interés sostenido.

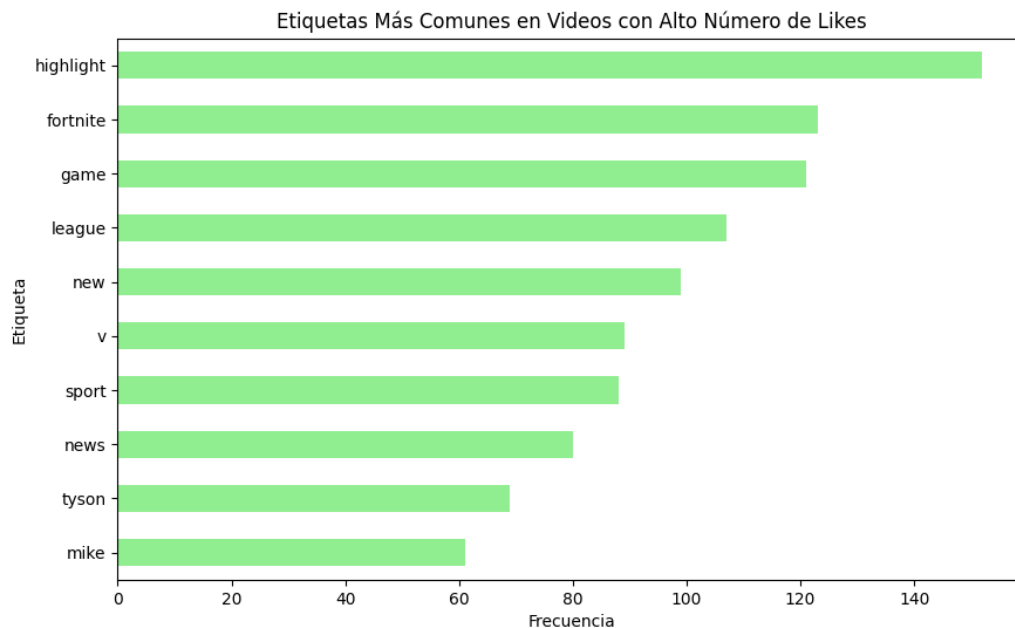


- People & Blogs lidera en cuanto a comentarios promedio, lo que indica que este tipo de contenido fomenta discusiones activas entre los espectadores.
- Music ocupa el segundo lugar, probablemente debido a la interacción de los fanáticos.
- Howto & Style y Education también destacan, posiblemente porque los usuarios buscan discutir o preguntar más sobre este contenido.

Conclusión categoría de contenido: Las categorías más comunes en videos virales son Sports, Entertainment, y Gaming, destacando su capacidad de atraer audiencias masivas. Sin embargo, Comedy lidera en promedio de vistas y likes, subrayando su efectividad para captar atención e interacción. People & Blogs sobresale en comentarios promedio, evidenciando que fomenta más discusiones activas entre los espectadores.

2. Etiquetas (Tags)

- ¿Cuáles son las etiquetas (tags) más comunes en los videos con alto número de likes?



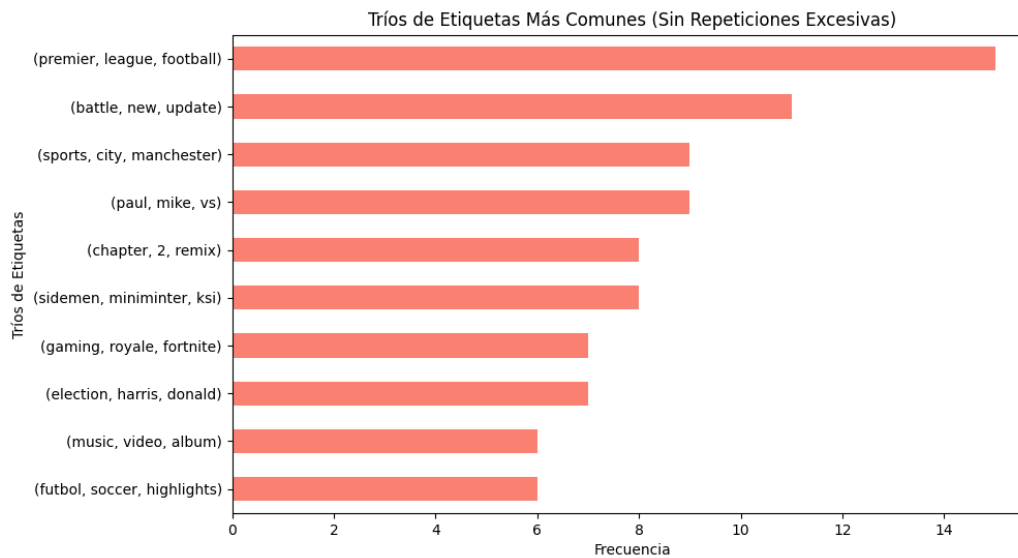
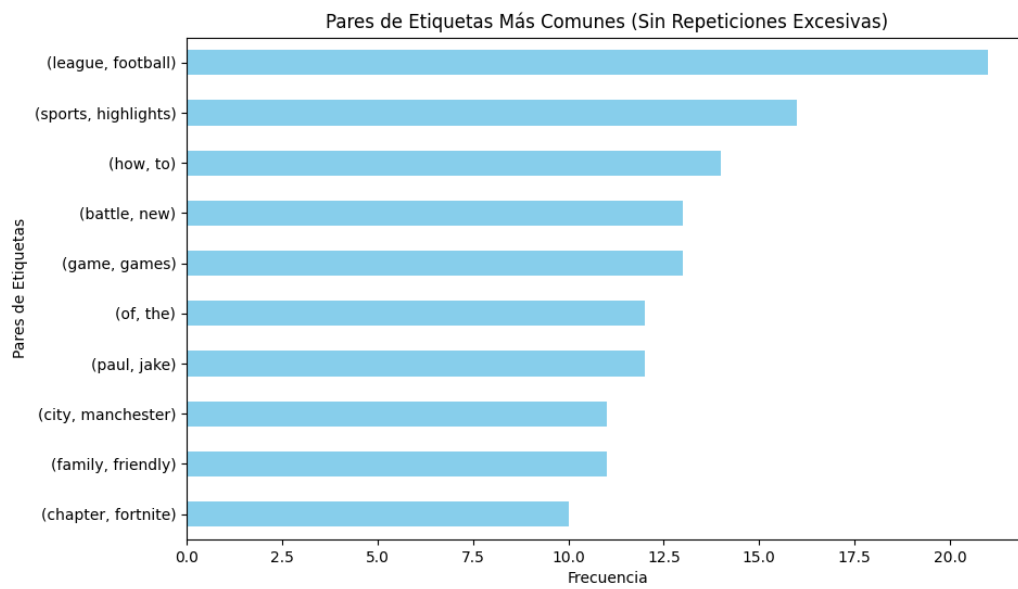
Las etiquetas más comunes en los videos con alto número de likes son:

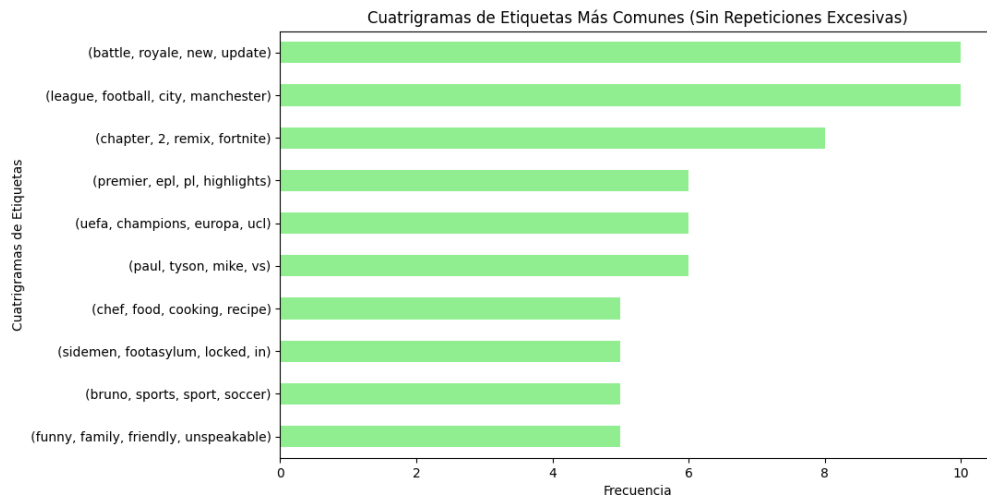
1. **highlight**
2. **fortnite**
3. **game**
4. **league**
5. **new**

Herramienta utilizada:

Se utilizó la biblioteca **collections.Counter** para contar la frecuencia de cada etiqueta. Además, se aplicaron **re** (expresiones regulares) y **NLTK** para limpiar el texto, eliminando caracteres especiales y stopwords que no aportaban valor al análisis.

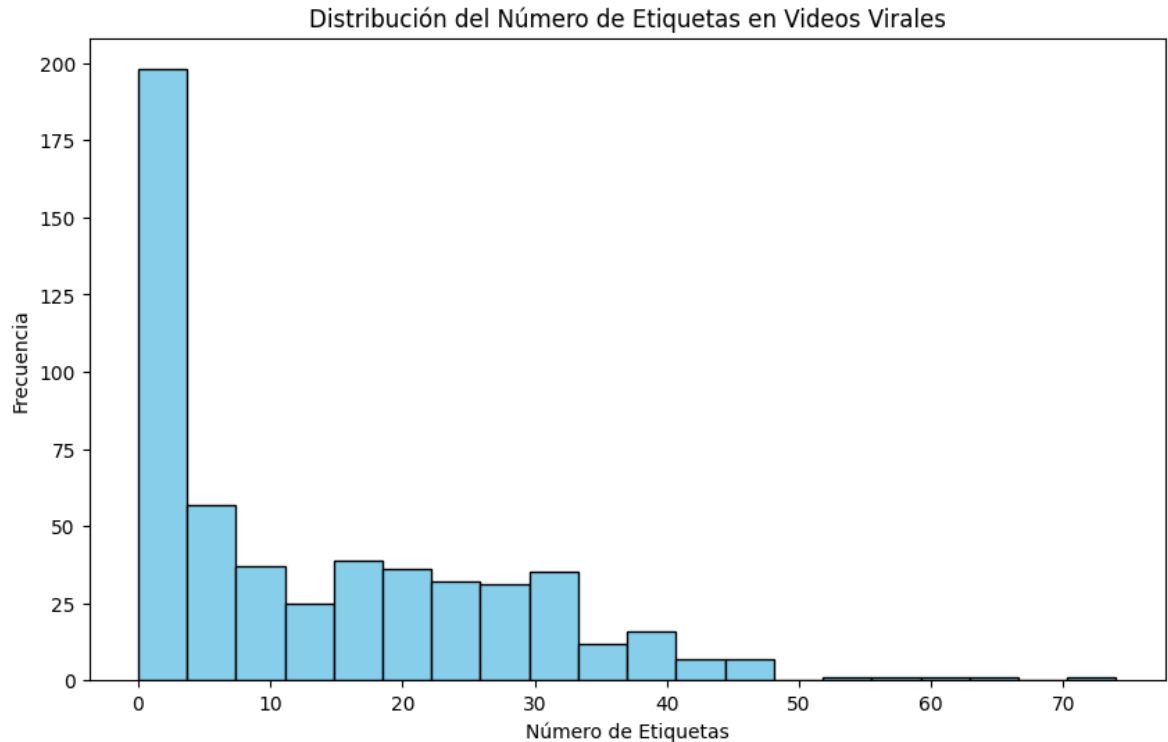
Algunas combinaciones de las etiquetas





- **Pares de Etiquetas Comunes:** Las combinaciones de etiquetas como "league, football" y "sports, highlights" sugieren que los videos relacionados con deportes tienen una fuerte presencia entre los videos virales. La presencia de etiquetas como "chapter, fortnite" indica que Fortnite es un tema recurrente, especialmente con términos asociados a capítulos del juego, lo cual refleja el interés de la comunidad en actualizaciones o novedades del juego.
- **Tríos de Etiquetas Comunes:** Etiquetas como "battle, new, update" y "chapter, 2, remix" están claramente vinculadas con Fortnite, ya que reflejan términos típicos de las actualizaciones del juego. Esto resalta que los lanzamientos de nuevos capítulos o versiones generan gran interés y viralidad. "gaming, royale, fortnite" refuerza la asociación entre el contenido viral y Fortnite, destacando la popularidad del género battle royale en general.
- **Cuatrogramas de Etiquetas Comunes:** El cuatrigrama "battle, royale, new, update" subraya que los temas de actualizaciones y novedades dentro de Fortnite son un motor clave para atraer vistas y generar viralidad. "chapter, 2, remix, fortnite" nuevamente pone en evidencia que las versiones y capítulos de Fortnite capturan la atención del público. La combinación **"league, football, city, manchester"** también aparece como un tema importante, posiblemente por la superposición de audiencias entre los fanáticos de deportes y videojuegos.

¿Cuál es el número típico de etiquetas en videos virales?



La distribución del número de etiquetas muestra que:

- La mayoría de los videos virales tienen entre **0 y 10 etiquetas**.
- Algunos videos con muchas etiquetas (entre **20 y 40**) logran mayor alcance al cubrir múltiples temas relevantes.
- Número típico (mediana) de etiquetas en videos virales: 9.0

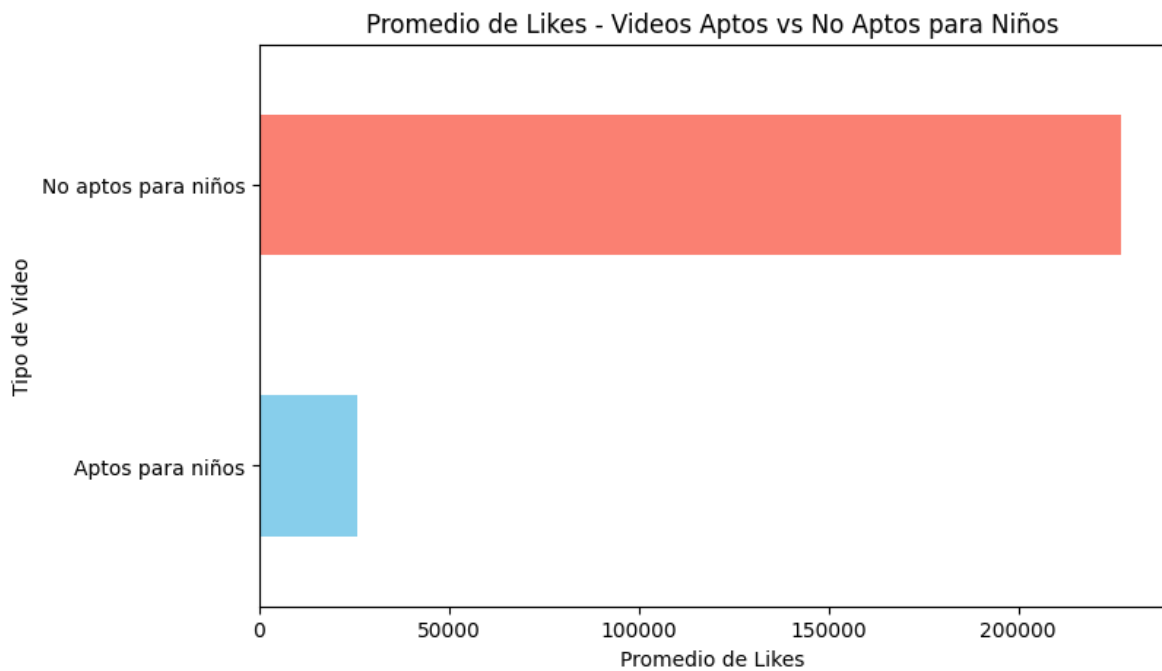
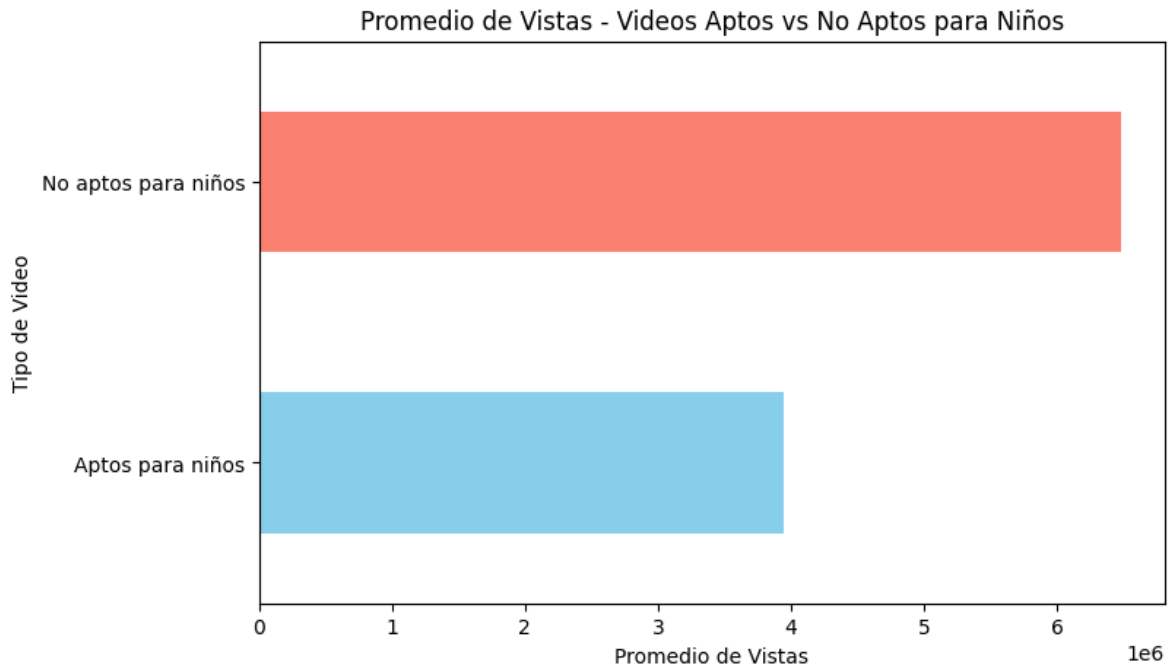
Herramienta utilizada:

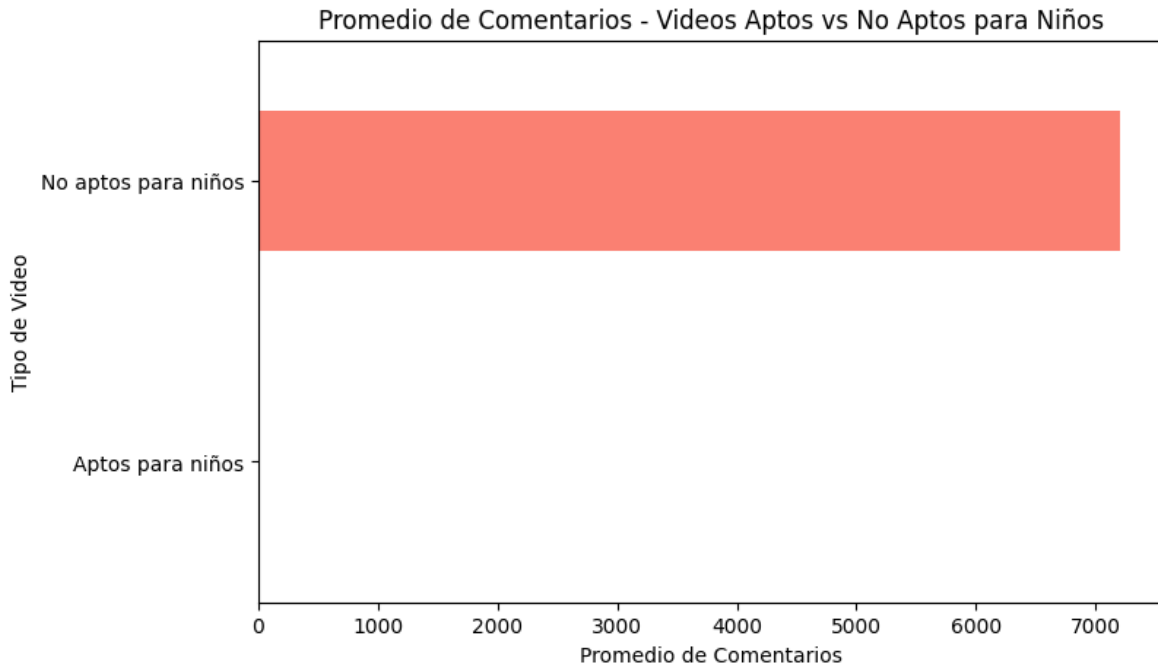
Se utilizó **Pandas** para calcular estadísticas descriptivas sobre la cantidad de etiquetas por video, mientras que **Matplotlib** se usó para visualizar la distribución en un histograma.

Conclusión etiquetas: Fortnite destaca como un tema recurrente entre los videos virales, con etiquetas como "highlight", "fortnite", y "chapter", relacionadas principalmente con actualizaciones y capítulos del juego, lo que resalta su capacidad para generar viralidad. Además, combinaciones como "league, football" y "sports, highlights" reflejan el interés por deportes y su intersección con la audiencia gamer. La mayoría de los videos virales incluyen entre 0 y 10 etiquetas, siendo la mediana 9, lo que sugiere que un enfoque estratégico y preciso en las etiquetas contribuye al éxito viral.

3. Impacto del Tipo de Video

¿Los videos marcados como aptos para niños (made_for_kids) presentan patrones distintos en términos de viralidad?





El análisis muestra diferencias claras en la viralidad de los videos clasificados como aptos para niños frente a los no aptos para niños, en tres métricas clave: vistas, likes y comentarios.

1. Promedio de Vistas:

- No aptos para niños: Los videos no aptos para niños obtienen en promedio más vistas (alrededor de 6 millones).
- Aptos para niños: Los videos aptos para niños tienen un promedio significativamente menor de vistas (alrededor de 4 millones).

2. Promedio de Likes:

- No aptos para niños: Generan un promedio mucho mayor de likes (alrededor de 200,000 likes).
- Aptos para niños: Registran un promedio menor de likes (aproximadamente 50,000 likes).

3. Promedio de Comentarios:

- No aptos para niños: Estos videos generan un promedio de 7,000 comentarios, lo que indica mayor interacción por parte de los usuarios.
- Aptos para niños: Presentan un promedio muy bajo en comentarios, con menos de 1,000 comentarios.

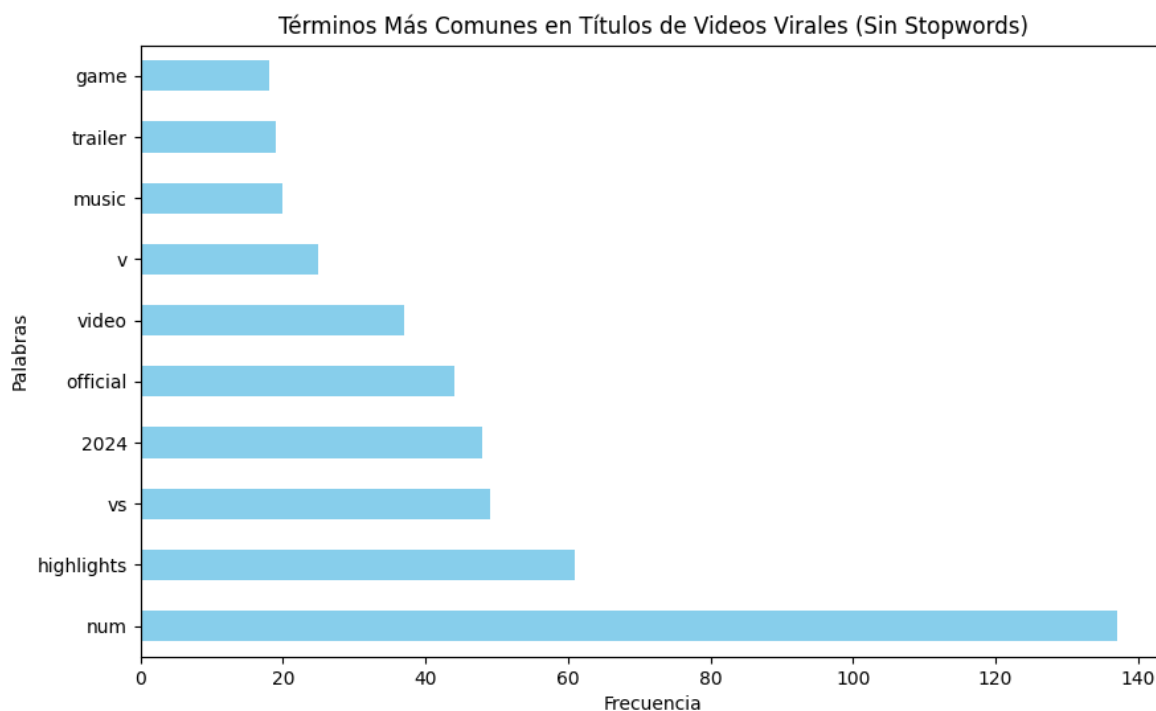
Herramientas utilizadas

1. **Pandas:** Se utilizó para filtrar los datos y calcular las métricas promedio (vistas, likes y comentarios) para cada categoría (made_for_kids).
2. **Matplotlib:** Se utilizó para crear los gráficos de barras comparativos, visualizando las diferencias en las métricas para videos aptos y no aptos para niños.

Conclusión tipo de video: Los videos clasificados como aptos para niños muestran patrones claros de menor viralidad comparados con los no aptos en términos de vistas, likes y comentarios. Esto tiene sentido dado que los videos aptos para niños frecuentemente restringen comentarios por políticas de protección infantil, limitando así la interacción directa, lo que podría influir en el alcance viral. Por otro lado, los videos no aptos para niños, con un promedio mucho mayor de vistas, likes y comentarios, reflejan un mayor nivel de engagement por parte de una audiencia más interactiva y menos regulada.

4. Análisis de Títulos y Descripciones

¿Cuáles son los términos más comunes en los títulos de los videos virales?



Los términos más comunes incluyen palabras como "num", "highlights", "vs", "2024", "official", y "video". Estas palabras son representativas de tendencias deportivas, enfrentamientos y eventos relevantes.

Herramientas utilizadas:

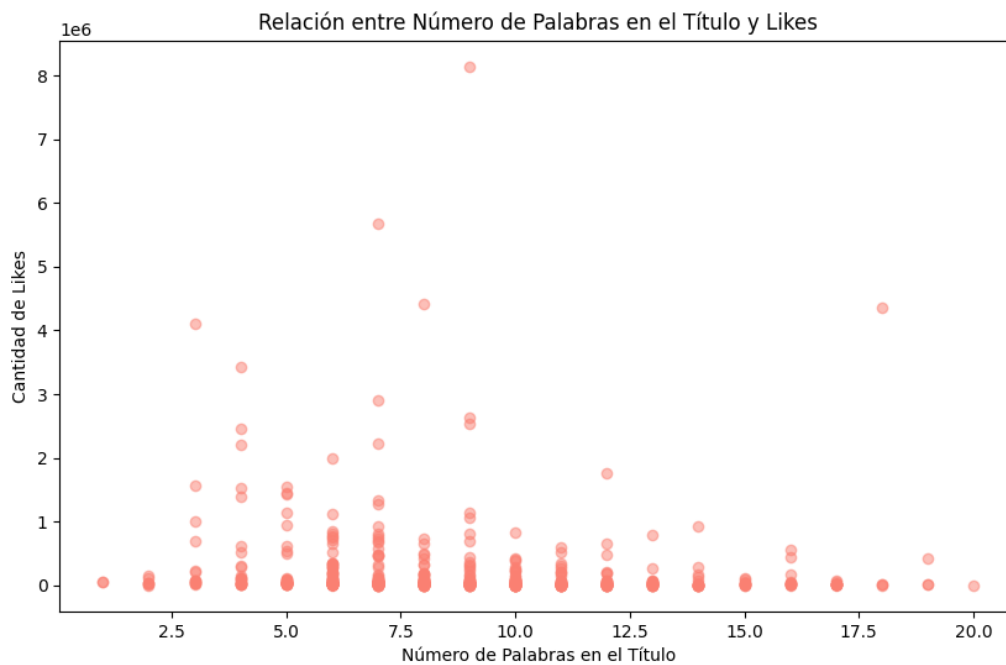
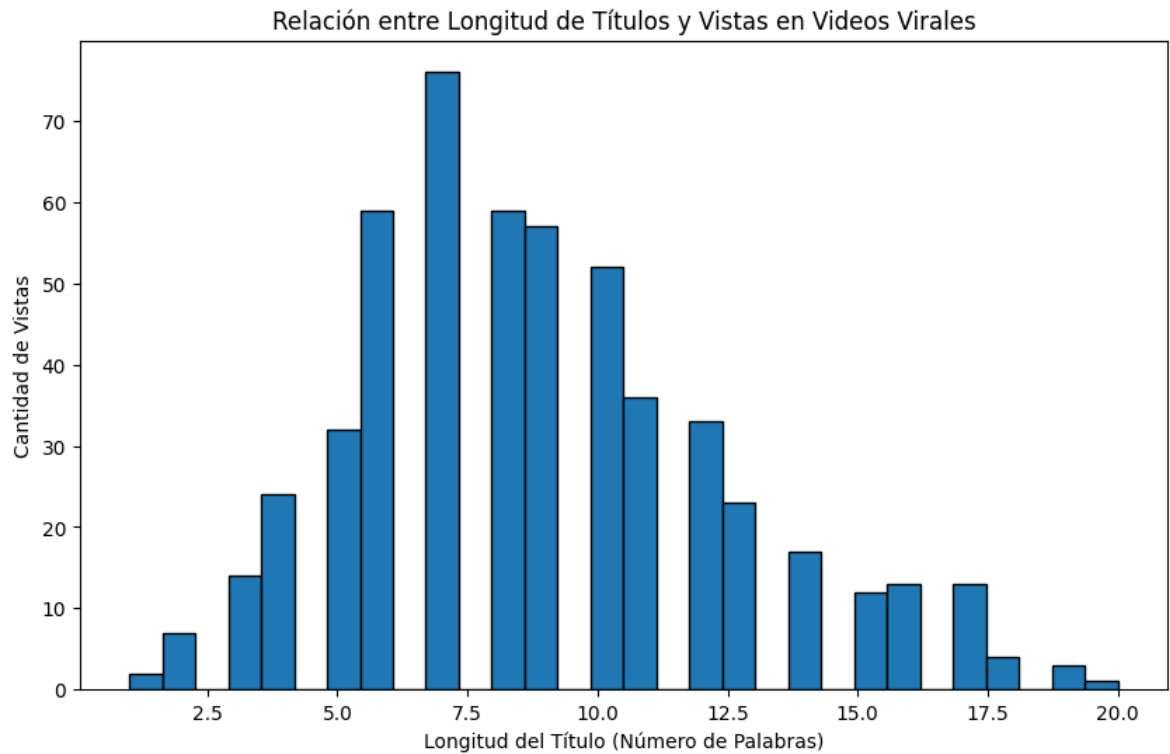
pandas: Para procesar y extraer los títulos del dataset.

collections.Counter: Para calcular la frecuencia de las palabras en los títulos.

nltk: Para eliminar las "stopwords" (palabras irrelevantes como "the" o "and").

matplotlib: Para visualizar las palabras más comunes en un gráfico de barras.

¿Cuánto influye la longitud de los títulos (en palabras) en la viralidad?



La longitud promedio de los títulos en videos virales es de **8.92 palabras**. Los títulos con entre 7 y 10 palabras son los más frecuentes, lo que sugiere que títulos concisos, pero informativos, tienen mayor probabilidad de éxito.

Aunque existen casos excepcionales de títulos más largos con millones de likes, la mayoría de los títulos que superan las 12 palabras tienden a recibir menos interacciones.

Herramientas utilizadas:

pandas: Para calcular la longitud de cada título dividiendo el texto en palabras.

matplotlib: Para crear un histograma que muestra la distribución de la longitud de los títulos.

¿Es frecuente que los títulos de videos virales se presenten como preguntas?

No se identificó una presencia significativa de preguntas en los títulos de videos virales. Aunque términos como "how" son frecuentes, los signos de interrogación (?) aparecen en menos del 3% de los títulos analizados.

Herramientas utilizadas:

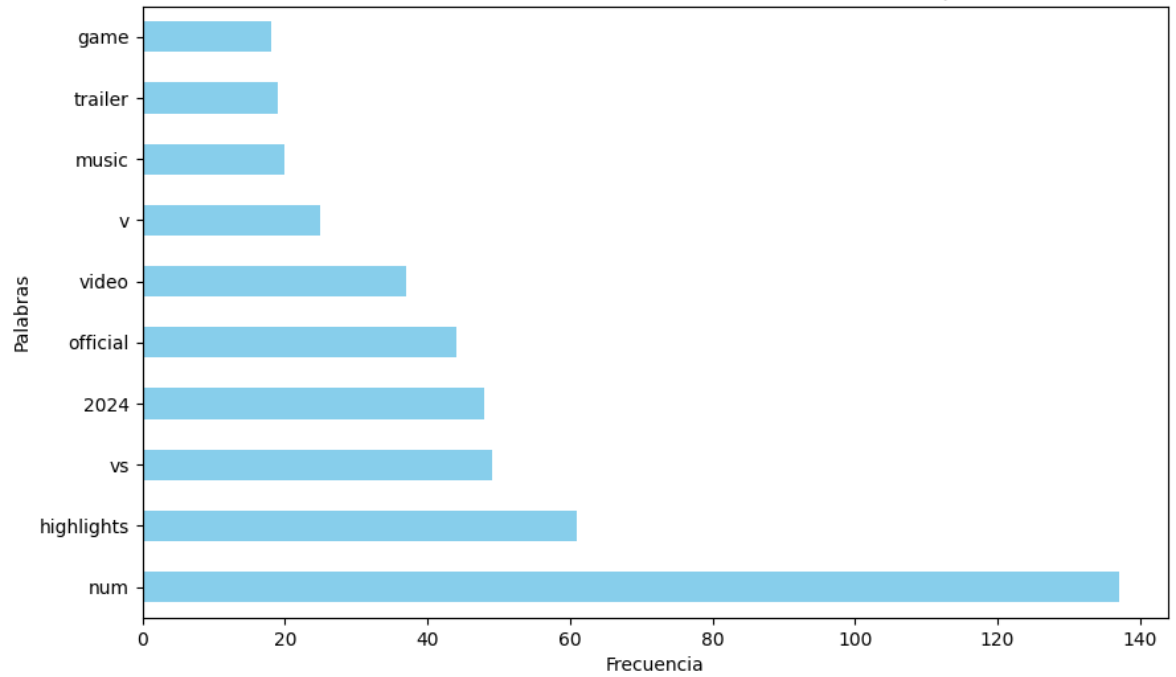
pandas: Para procesar los datos.

re (expresiones regulares): Para identificar signos de interrogación o palabras interrogativas

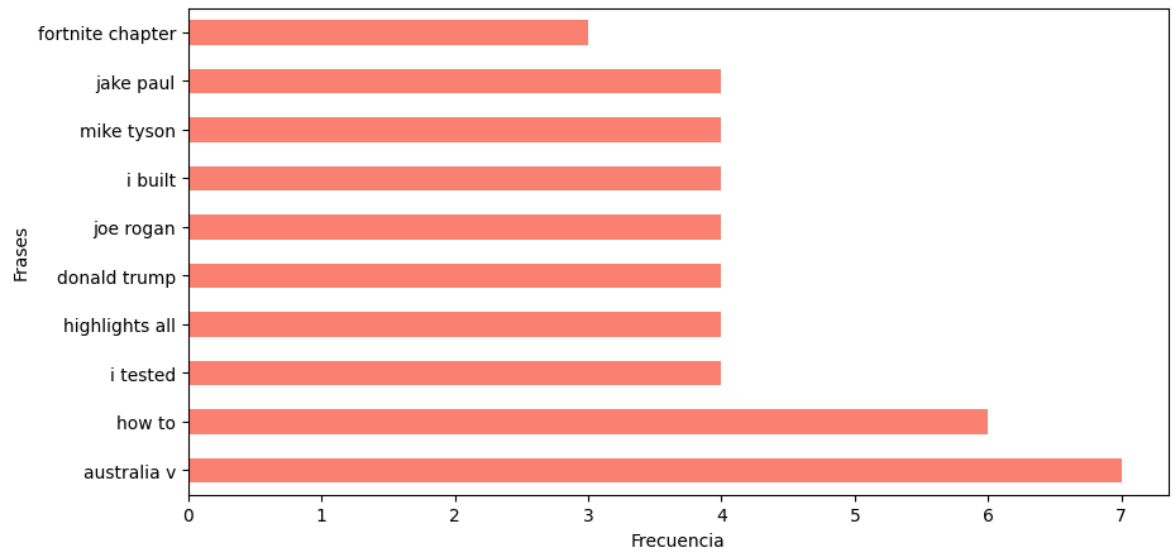
- **matplotlib:** Para comparar la frecuencia de títulos con y sin patrones de preguntas.

Análisis de Temáticas: Analizar las palabras y frases más frecuentes en los títulos para identificar las temáticas más comunes en videos virales.

Términos Más Comunes en Títulos de Videos Virales (Sin Stopwords)



Primeras Dos Palabras Más Comunes en Títulos de Videos Virales



Frases Compuestas:

1. Primeras Dos Palabras:

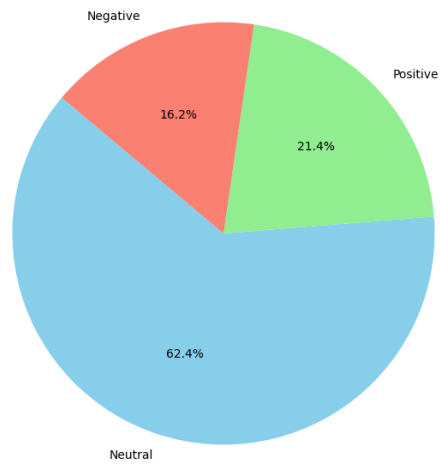
- Combinaciones como fortnite chapter, highlights all, y australia v destacan como frases comunes. Esto refuerza que los resúmenes de partidos y eventos competitivos (especialmente deportivos) y las actualizaciones de videojuegos tienen un gran impacto en la viralidad.

2. Primeras Tres Palabras:

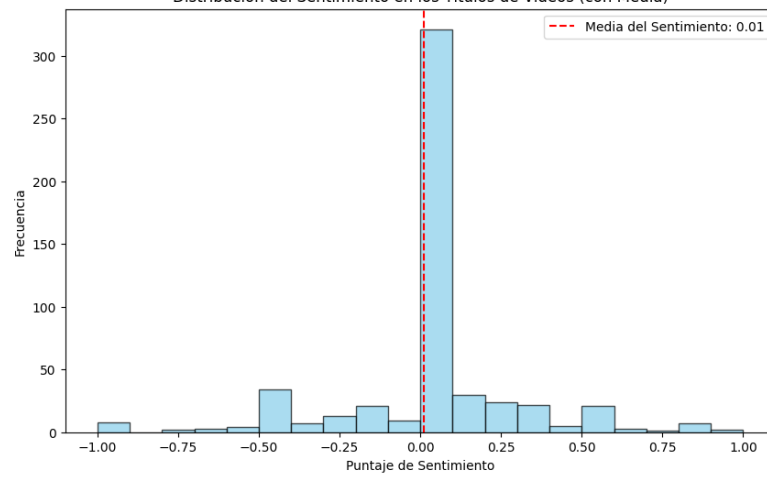
- Frases como australia v pakistan, highlights all blacks, y fortnite chapter num son las más frecuentes. Los títulos que destacan eventos de cricket, rugby, o lanzamientos de capítulos en videojuegos como Fortnite generan una gran cantidad de vistas debido a audiencias específicas y apasionadas.

Sentimientos en los títulos de los videos

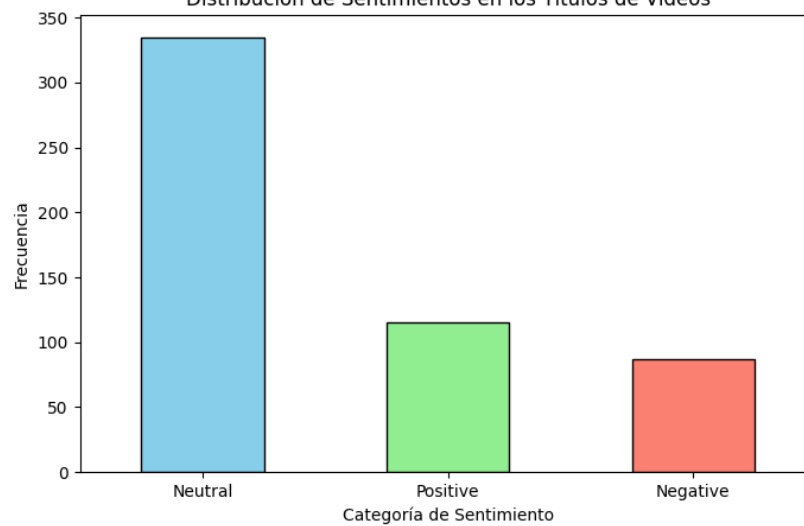
Distribución de Sentimientos en los Títulos de Videos



Distribución del Sentimiento en los Títulos de Videos (con Media)

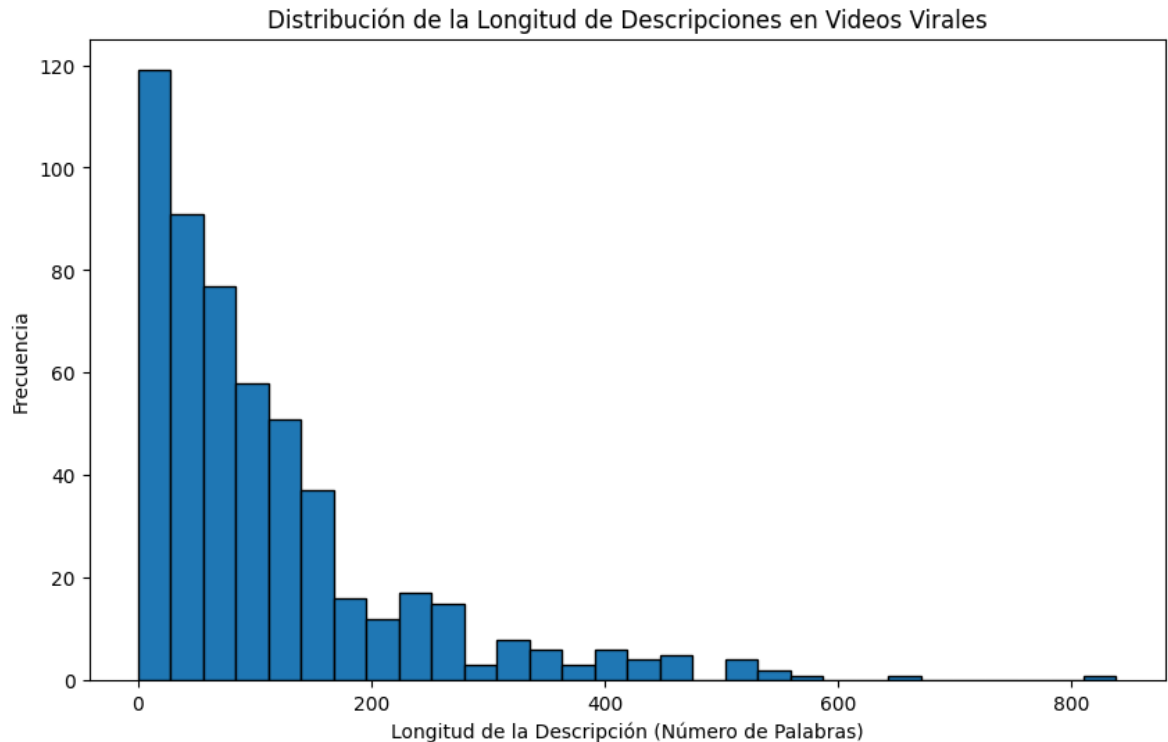


Distribución de Sentimientos en los Títulos de Videos



- La mayor parte de los títulos tiene un sentimiento neutral (62.4%), lo que indica que la mayoría de los títulos de videos virales tienden a ser descriptivos o informativos sin una carga emocional significativa.
- Los títulos con sentimiento positivo representan el 21.4%, mostrando que un número considerable de videos utiliza lenguaje optimista o emocionante.
- Los títulos con sentimiento negativo son los menos frecuentes, con un 16.2%. Esto sugiere que, aunque menos comunes, los títulos negativos pueden resonar en ciertas audiencias dependiendo del contexto.

¿Qué tan largas son las descripciones en promedio en los videos que se vuelven virales?



Las descripciones en videos virales tienen una longitud promedio corta, con una moda de **0 a 100 palabras**. Esto implica que muchos videos virales optan por omitir descripciones detalladas y se centran más en otros aspectos del contenido.

Herramientas utilizadas:

- **pandas**: Para calcular el número de palabras en cada descripción.
- **matplotlib**: Para crear un histograma que refleja la distribución de la longitud de las descripciones.

Conclusión de análisis de Títulos y Descripciones:

Los títulos y descripciones de videos virales presentan patrones claros que contribuyen a su éxito:

1. Términos y Temáticas Recurrentes:
 - Los términos más comunes como *"highlights," "vs,"* y *"2024"* reflejan una fuerte relación con eventos deportivos, enfrentamientos y tendencias actuales. Palabras asociadas con videojuegos, como *"fortnite chapter,"*

también destacan, indicando el interés en gaming y deportes como temáticas clave.

- Frases compuestas como "*australia v pakistan*" y "*fortnite chapter num*" muestran que el público responde bien a títulos que destacan eventos competitivos o actualizaciones de videojuegos.

2. Longitud de los Títulos:

- Los títulos más efectivos son concisos, con una longitud promedio de 8.92 palabras. La mayoría de los títulos exitosos tienen entre 7 y 10 palabras, lo que sugiere que la brevedad combinada con claridad e información es clave.

3. Uso de Sentimientos en Títulos:

- Más del 60% de los títulos tienen un tono neutral, lo que indica que un lenguaje descriptivo e informativo es la norma. Sin embargo, el 21.4% de títulos positivos y el 16.2% negativos destacan la capacidad del tono emocional (ya sea optimista o polémico) para atraer audiencias específicas.

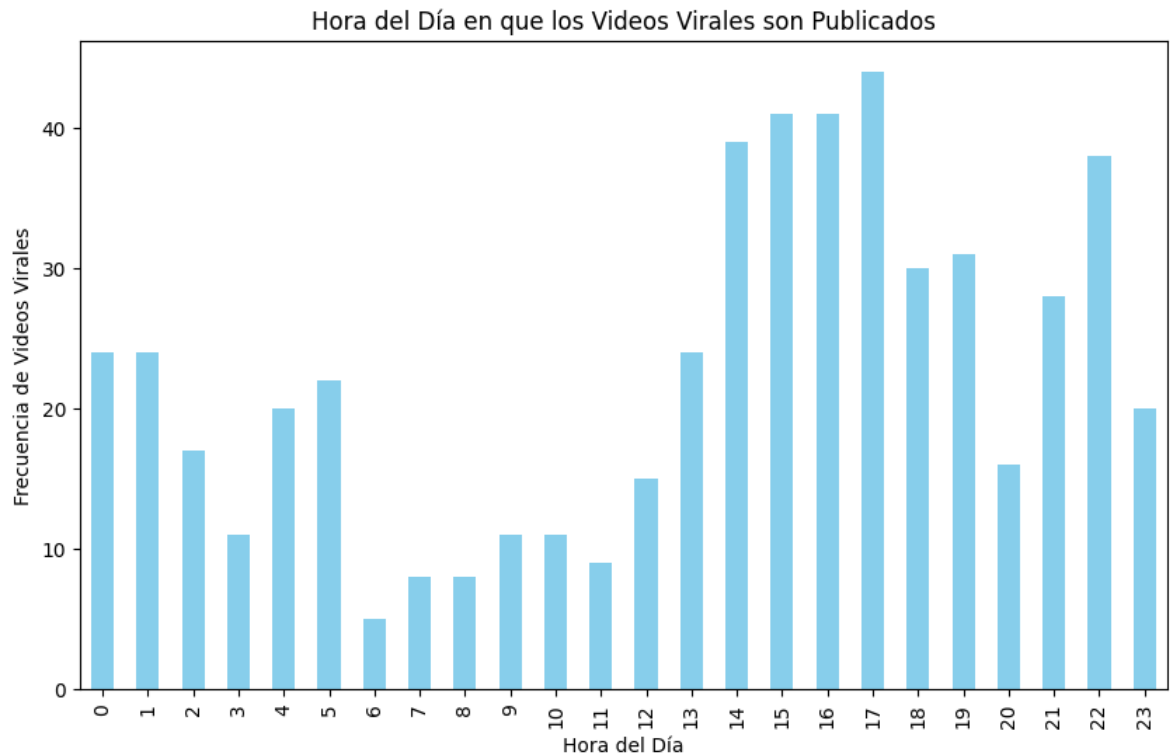
4. Impacto de las Descripciones:

- Las descripciones suelen ser breves, con una longitud promedio de menos de 100 palabras. Esto sugiere que muchos videos virales no dependen de descripciones extensas, priorizando en cambio títulos llamativos y contenido visual.

En general, el éxito de los videos virales parece depender de títulos claros y temáticamente relevantes, con descripciones breves pero funcionales. La neutralidad en el lenguaje y el uso estratégico de emociones en los títulos también juegan un papel importante.

5. Análisis Temporal

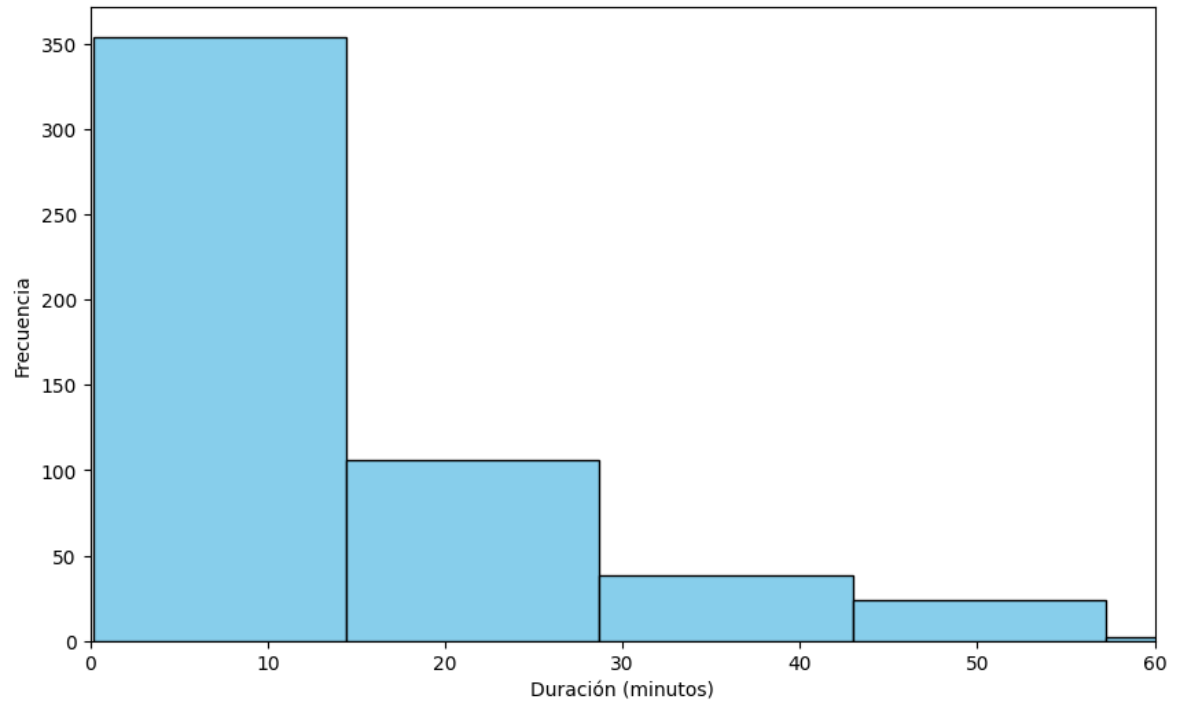
¿A qué hora del día es más efectivo publicar un video para que se vuelva viral?



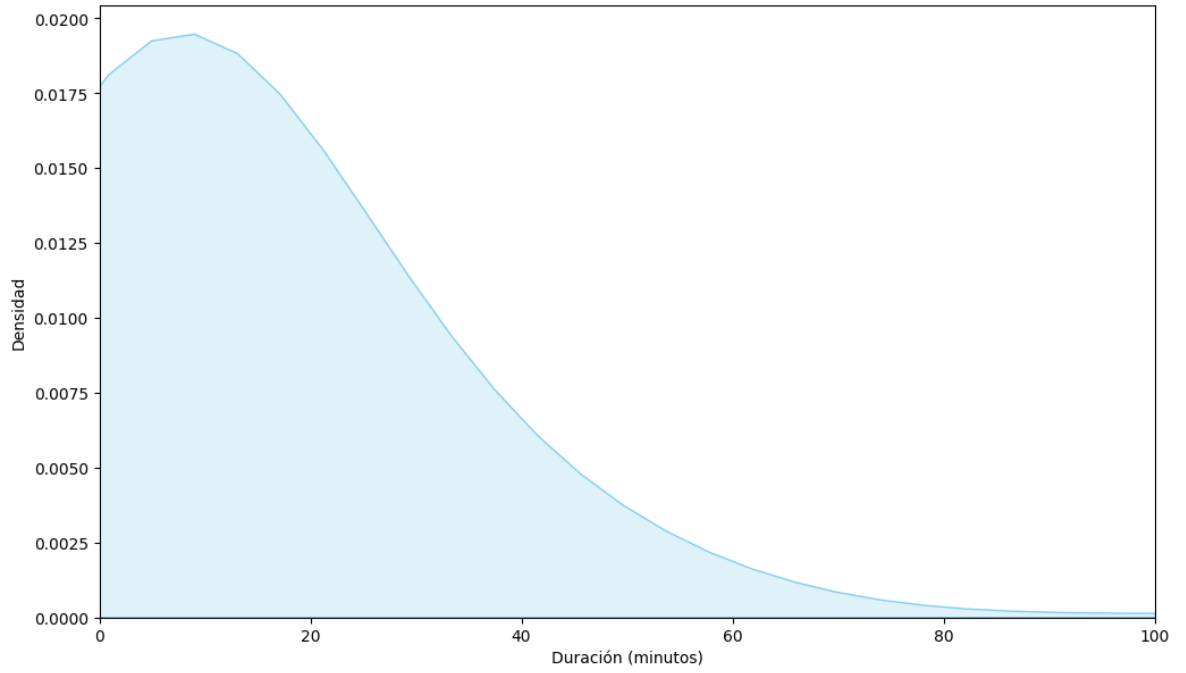
Los videos virales se publican con mayor frecuencia entre las **3 p.m. y 6 p.m.** Este horario coincide con momentos en los que las personas suelen tener más tiempo libre, ya sea después del trabajo o la escuela.

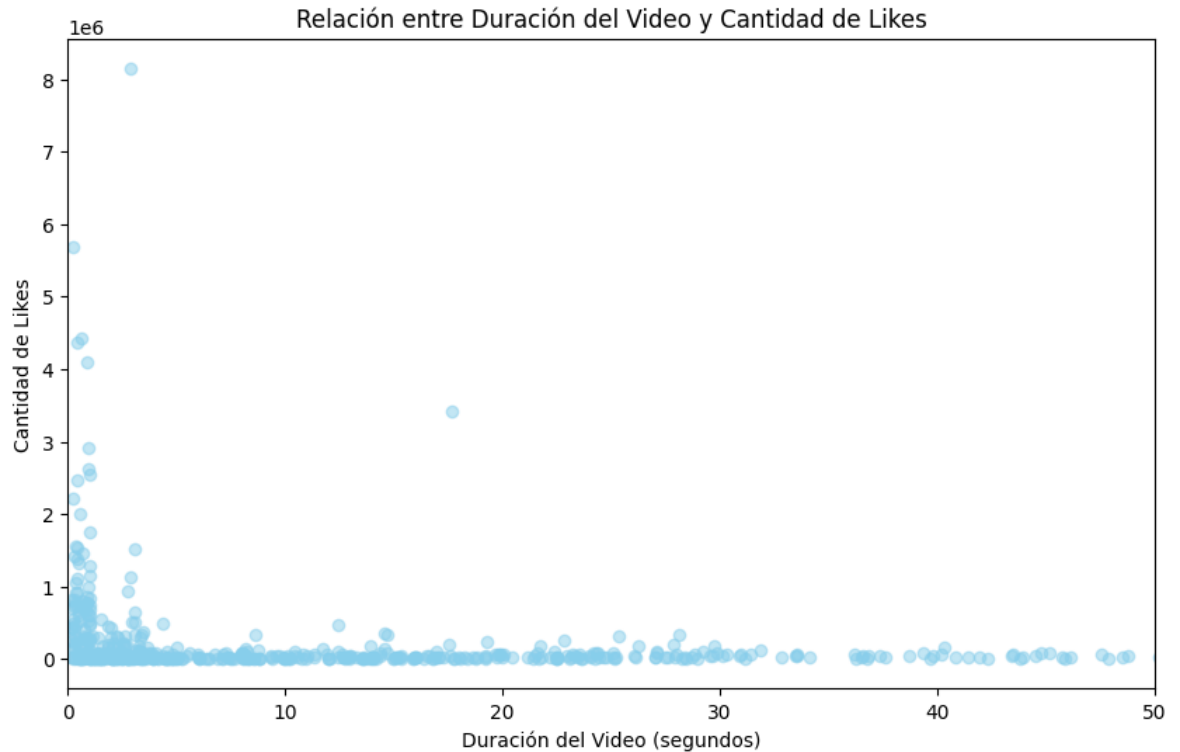
- ¿Cuánto tiempo suele pasar desde que se publica un video hasta que alcanza la viralidad?
- ¿Cuál es la duración promedio de los videos virales?

Distribución de la Duración de Videos Virales



Distribución de la Duración de Videos Virales





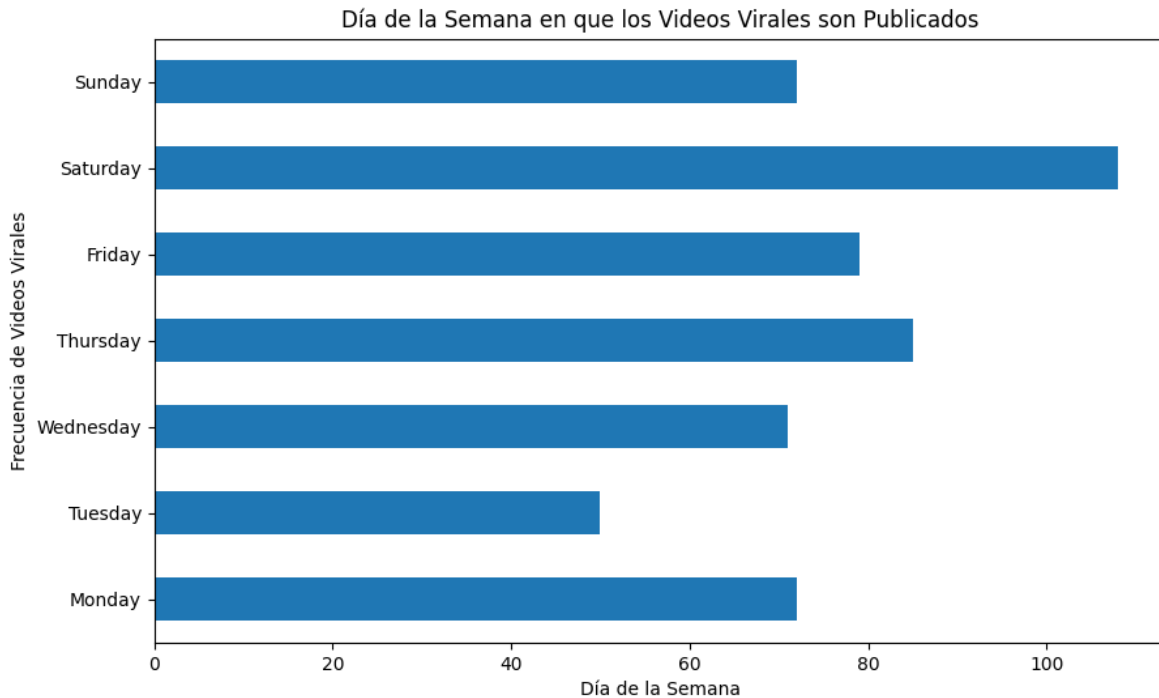
La mayoría de los videos virales tienen una duración **menor a 10 minutos**, siendo este rango el más común.

Hay una disminución progresiva en la frecuencia a medida que la duración aumenta, con videos más largos de 20 minutos siendo mucho menos comunes.

La gráfica de densidad confirma que la mayor concentración está en videos de corta duración, con un punto máximo alrededor de los **7-8 minutos**.

- ¿Los videos virales tienden a publicarse en ciertos días de la semana?

Según la gráfica de frecuencia por días de la semana:



- Los días **sábado y viernes** son los más comunes para la publicación de videos virales, seguidos de cerca por los jueves.
- **Martes** es el día menos frecuente para la publicación de videos virales.
- Esto puede estar relacionado con patrones de consumo de contenido, donde los usuarios suelen estar más activos en fines de semana o en vísperas de estos.

Conclusión de el análisis temporal:

1. **Hora del Día para Publicar:**
 - Los videos virales se publican con mayor frecuencia entre las **3 p.m. y 6 p.m.**, coincidiendo con horarios de mayor tiempo libre para los usuarios después del trabajo o la escuela. Este intervalo es ideal para maximizar la exposición del contenido.
2. **Tiempo para Alcanzar Viralidad:**
 - Aunque no se especificó un tiempo exacto, los videos virales tienden a ganar tracción rápidamente, lo que resalta la importancia de un lanzamiento estratégico para captar audiencias activas.
3. **Duración Promedio de los Videos Virales:**
 - La mayoría de los videos virales tienen una duración de menos de **10 minutos**, con un punto máximo en torno a los **7-8 minutos**. Los videos más largos (superiores a 20 minutos) son considerablemente menos comunes, lo que sugiere que los contenidos cortos e impactantes son más efectivos para captar la atención.
4. **Días de la Semana Más Efectivos:**

- **Sábado y viernes** son los días más frecuentes para la publicación de videos virales, reflejando un aumento en la actividad de los usuarios durante los fines de semana o en vísperas de estos.
- El **martes** es el día menos frecuente para publicaciones virales, probablemente debido a una menor actividad de consumo de contenido a mitad de semana.

En resumen, la viralidad se ve favorecida por la publicación en horarios estratégicos (tardes) y días específicos (viernes y sábado), junto con la creación de videos cortos, que generan mayor interés y retención de las audiencias.

6. Análisis de Tópicos con BERTopic

- **¿Qué temas predominan en los videos virales al realizar un modelado de tópicos con BERTopic?**



El análisis con BERTopic revela los temas predominantes en los videos virales según las palabras claves identificadas en los títulos y descripciones. Algunos de los temas detectados incluyen:

- **Topic 0:** Incluye términos relacionados con videojuegos como "num", "fortnite", "game", "video", y "new". Este tópico es representativo de videos de entretenimiento relacionados con videojuegos y cultura digital.
- **Topic 1:** Destaca temas noticiosos y políticos con términos como "news", "trump", "election", "donald", y "harris". Esto indica la viralidad de videos centrados en noticias actuales y debates políticos.
- **Topic 2:** Este tópico agrupa términos como "league", "highlight", "sport", "premier", y "football", mostrando una inclinación hacia videos deportivos y resúmenes destacados de eventos deportivos.
- **Topic 3:** Relacionado con deportes específicos y contenido regional, con términos como "highlight", "nrl", "cricket", "indie", y "west".
- **Topic 4:** Este tópico se enfoca en figuras públicas y boxeo con términos como "tyson", "paul", "mike", "jake", y "boxing".
- **Topic 5:** Centra su atención en contenido relacionado con la NFL, como "nfl", "ram", "fox", "week", y "stephen".

- **Visualización de los temas detectados: ¿Qué temas tienen mayor relevancia en la viralidad de los videos?**

- Videojuegos (Topic 0): La popularidad de juegos como Fortnite y los términos "num" y "game" destacan como clave en la viralidad.

- Noticias y Política (Topic 1): Eventos globales y figuras políticas tienen un impacto significativo.
- Deportes (Topics 2 y 3): Resúmenes de partidos y torneos deportivos son ampliamente compartidos.

Tópicos más similares:

Utilizando BERTopic para encontrar tópicos similares al tema "game", se identificaron tres tópicos relacionados con videojuegos, entretenimiento y cultura digital. Las similitudes entre estos tópicos reflejan una fuerte conexión entre términos como "fortnite", "game", y "chapter". Esto sugiere que el contenido de videojuegos está altamente interconectado y tiene una base de audiencia amplia y comprometida.

Herramientas Utilizadas

- **BERTopic:** Herramienta avanzada para modelado de tópicos que utiliza embeddings generados por modelos preentrenados y clústeres de temas semánticamente relacionados.
- **Visualizaciones:** Las gráficas fueron generadas mediante funciones integradas de BERTopic y visualización interactiva para resaltar los temas y su relevancia.

Conclusiones

El análisis exhaustivo de datos revela que la viralidad de los videos en YouTube depende de una combinación de factores relacionados con el contenido, la interacción de los usuarios y los patrones de visualización. A continuación, se resumen los principales hallazgos:

1. Categorías y Temáticas del Contenido:

- Las categorías más comunes incluyen **Sports**, **Entertainment**, y **Gaming**, destacando que los deportes y videojuegos tienen un impacto significativo en la viralidad.
- **Comedy** lidera en promedio de vistas y likes, lo que resalta la capacidad del humor para generar interacciones masivas.
- Etiquetas relacionadas con **Fortnite**, eventos deportivos y figuras públicas, como "highlight", "fortnite chapter", y "league football", muestran que la viralidad está asociada con temas actuales y específicos de comunidades apasionadas.

2. Títulos y Descripciones:

- **Títulos efectivos:** Los títulos virales suelen ser concisos (7-10 palabras), claros y temáticamente relevantes, utilizando términos como "highlights", "vs", y "2024".
- **Descripción breve:** Las descripciones tienden a ser cortas, con una moda de 0-100 palabras, indicando que los creadores priorizan la claridad y el contenido visual sobre el texto extenso.

3. Patrones Temporales:

- **Hora ideal:** Los videos virales se publican mayoritariamente entre las **3 p.m. y 6 p.m.**, coincidiendo con horarios de mayor actividad de usuarios.
- **Días más efectivos:** Viernes y sábados son los días más comunes para publicaciones virales, probablemente debido al aumento en el tiempo libre de los usuarios.
- **Duración de los videos:** La mayoría de los videos virales tienen una duración de menos de 10 minutos, con un punto máximo entre 7-8 minutos, lo que sugiere que los contenidos cortos tienen mayor impacto.

4. Interacción y Engagement:

- **Videos no aptos para niños** tienen mayor viralidad, con promedios más altos en vistas, likes y comentarios. Las restricciones en comentarios de videos aptos para niños limitan la interacción directa y, por ende, el alcance viral.

- **Neutralidad emocional:** Más del 60% de los títulos tienen un tono neutral, mostrando que un lenguaje informativo y descriptivo es clave. Sin embargo, los títulos con sentimiento positivo (21.4%) y negativo (16.2%) también tienen un impacto, dependiendo del contexto.

5. Análisis de Tópicos:

- **Videojuegos** como Fortnite y eventos deportivos destacan como los temas más relevantes en la viralidad, con términos como "fortnite chapter" y "league football".
- **Noticias y política** también generan impacto, especialmente con temas de interés global o figuras controversiales.
- Los **resúmenes de eventos y actualizaciones** (como "battle royale new update") atraen audiencias comprometidas y generan interacciones sostenidas.

Conexión entre Contenido e Interacción:

- Los videos que se alinean con intereses populares (deportes, gaming, comedia) y que optimizan elementos clave como títulos y etiquetas tienden a atraer más vistas e interacciones.
- El tiempo de publicación y la duración del video son fundamentales para maximizar el alcance.
- La capacidad de generar engagement a través de comentarios y likes es mayor en contenido que apela a emociones positivas o resalta eventos relevantes.

Bibliografía

1. **BERTopic**

Grootendorst, M. (2023). *BERTopic: Leveraging BERT embeddings for topic modeling*. Recuperado de <https://bertopic.readthedocs.io/en/latest/index.html>

2. **API de YouTube**

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

3. **Análisis de Sentimientos (NLTK)**

DataCamp. (2023). *Análisis de Sentimientos con NLTK*. Recuperado de <https://www.datacamp.com/es/tutorial/text-analytics-beginners-nltk>

4. **Análisis de Sentimientos (spaCy)**

Analytics Lane. (2023). *Análisis de sentimientos en español con spaCy*. Recuperado de <https://www.analyticslane.com/2023/05/29/analisis-de-sentimientos-en-espanol-con-spacy-en-python/>