



UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

PROYECTO DE INVESTIGACIÓN

Título: Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal.

Trabajo de titulación previo a la obtención del título Magíster en Ciencia de Datos

Autor:

Ing. Henry Saúl Guerrero León

Tutor:

Msc. Segundo Humberto Corrales Beltrán

**LATACUNGA – ECUADOR
2025**

APROBACIÓN DEL TUTOR

En mi calidad de Tutor del Trabajo de Titulación “Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal” presentado por Henry Saúl Guerrero León, para optar por el título Magíster en Ciencia de Datos.

CERTIFICO

Que dicho Trabajo de Titulación ha sido revisado en todas sus partes y se considera que reúne los requisitos y méritos suficientes para ser sometido a la presentación para la valoración por parte del Tribunal de Lectores que se designe y su exposición y defensa pública.

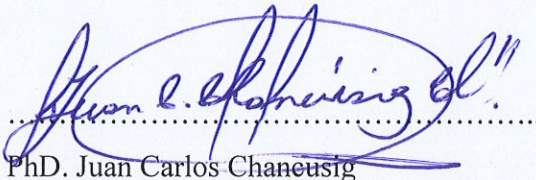
Latacunga, 8 de mayo del 2025


.....
Segundo Humberto Corrales Beltrán, MSc
CC: 0502409287

APROBACIÓN TRIBUNAL

El Trabajo de Titulación: “*Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal*”, ha sido revisado, aprobado y autorizada su impresión y empastado, previo a la obtención del título de Magíster en Ciencia de Datos. El trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

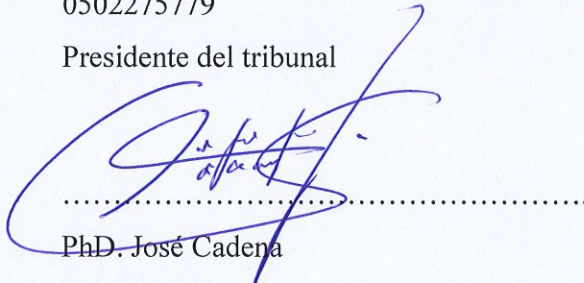
Latacunga, mayo 2025



PhD. Juan Carlos Chaneusig

0502275779

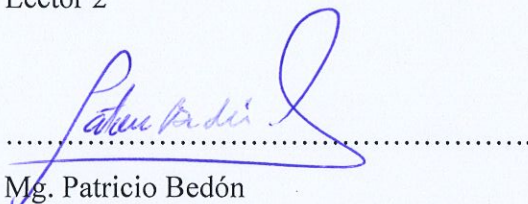
Presidente del tribunal



PhD. José Cadena

0501552798

Lector 2



Mg. Patricio Bedón

0502253271

Lector 3

DEDICATORIA

Dedico este trabajo con todo mi cariño y gratitud a mis padres, por su ejemplo incansable de esfuerzo, dedicación y amor. A ellos, que siempre creyeron en mí incluso cuando yo dudaba.

A mi familia, por su paciencia, comprensión y apoyo en cada paso de este camino.

Y a mis amigos cercanos, que, con una palabra de aliento o un momento de distracción, hicieron más llevadera esta travesía académica.

Este logro es también de ustedes.

- Henry Saúl Guerrero León

AGRADECIMIENTO

Agradezco profundamente a todas las personas que hicieron posible este proyecto.

A mis padres y seres queridos, por acompañarme emocionalmente en este proceso con su apoyo incondicional. A mis amigos, por compartir conmigo tanto los momentos de estrés como los de alegría.

A mis docentes, quienes transmitieron no solo conocimiento, sino también pasión por la ciencia y el análisis.

Y a mis tutores, por su guía, paciencia y orientación crítica durante el desarrollo de esta tesis. Su experiencia y disposición marcaron una gran diferencia en la calidad de este trabajo.

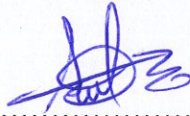
Gracias a todos por ser parte de este esfuerzo y por contribuir, cada uno a su manera, a que hoy esta meta sea una realidad.

- Henry Saúl Guerrero León

RESPONSABILIDAD DE AUTORÍA

Quien suscribe, declara que asume la autoría de los contenidos y los resultados obtenidos en el presente Trabajo de Titulación.

Latacunga, mayo 2025



.....
Ing. Henry Saúl Guerrero León

CC: 1727470229

RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total y/o parcial del presente Trabajo de Titulación a la Universidad Técnica de Cotopaxi.

Latacunga, mayo 2025



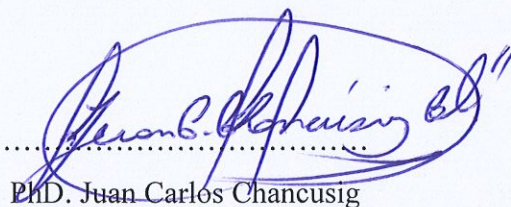
.....
Ing. Henry Saúl Guerrero León

CC: 1727470229

AVAL DEL PRESIDENTE

Quien suscribe, declara que el presente Trabajo de Titulación: “*Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal*”, contiene las correcciones a las observaciones realizadas por los miembros del Tribunal en la predefensa.

Latacunga, mayo, 2025



PhD. Juan Carlos Chancusig

CC: 0502275779

UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

PROYECTO DE DESARROLLO

Título: Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal.

Trabajo de titulación previo a la obtención del título Magíster en Ciencia de Datos

Autor:

Henry Saúl Guerrero León

Tutor:

Segundo Humberto Corrales Beltrán

RESUMEN

El análisis del rendimiento de videos en YouTube es esencial para optimizar estrategias de contenido y mejorar el alcance de los creadores en la plataforma. Sin embargo, la falta de herramientas accesibles para comparar videos limita la toma de decisiones basada en datos. Este trabajo desarrolla una aplicación web que, mediante la YouTube Data API v3, permite extraer y analizar métricas clave para evaluar tendencias y patrones de rendimiento. La metodología utilizada incluye la recopilación y procesamiento de datos, el diseño de una interfaz amigable y la implementación de técnicas de análisis para interpretar los resultados obtenidos. Se espera que la aplicación facilite la identificación de tendencias y fortalezca la toma de decisiones estratégicas en la producción de contenido digital. Con ello, se contribuye a la optimización del rendimiento en YouTube, proporcionando una solución innovadora para el análisis comparativo de videos en la plataforma.

PALABRAS CLAVE: análisis de contenido; YouTube; comparación de videos; métricas; tendencias; estrategia digital; rendimiento.

UNIVERSIDAD TÉCNICA DE COTOPAXI
DIRECCIÓN DE POSGRADO
MAESTRÍA EN CIENCIA DE DATOS
PROYECTO DE DESARROLLO

Topic: "YouTube Video Comparator: Content Analysis, Statistics, and Trends for the Content Strategies' Improvement and Channel Performance".

Titling work prior to obtaining the Master's degree in Data Science

Author:
Guerrero León Henry Saúl

Tutor:
Segundo Humberto Corrales Beltrán

ABSTRACT

The analyzing video performance on YouTube is essential to optimize content strategies and improving creators' reach on the platform. However, the lack accessible tools for comparing videos limits data-driven decision-making. This work develops a web application that, through the YouTube Data API v3, allows users to extract and analyze key metrics to assess trends and performance patterns. The used methodology includes data collection and processing, the design a user-friendly interface, and the implementation analytical techniques to interpret the got results. It is expected, which the application to facilitate the identification trends and strengthen strategic decision-making in the digital content production. With it is contributed to the optimization performance on YouTube, by providing an innovative solution for comparative analysis videos on the platform.

KEYWORDS: Content analysis, YouTube, video comparison, metrics, trends, digital strategy, performance.

Yo, Marco Paúl Beltrán Semblantes con cédula de identidad número: 0502666514 Magíster en Lingüística Aplicada en la enseñanza del Idioma Inglés como Lengua Extranjera con número de registro de la SENESCYT ; 1020-2021-2354162: CERTIFICO haber revisado y aprobado la traducción al idioma Inglés del resumen del trabajo de investigación con el título: "**COMPARADOR DE VIDEOS EN YOUTUBE: ANÁLISIS DE CONTENIDO, ESTADÍSTICAS Y TENDENCIAS PARA LA MEJORA DE ESTRATEGIAS DE CONTENIDO Y RENDIMIENTO DEL CANAL**" de: Guerrero León Henry Saúl, aspirante a Magíster en Ciencia de Datos.

Latacunga, Mayo del 2025.

Atentamente,



Mg. Marco Paúl Beltrán Semblantes

DOCENTE CENTRO DE IDIOMAS-UTC

ÍNDICE DE CONTENIDOS

PROBLEMA DE INVESTIGACIÓN.....	1
1. Introducción	1
1.1. Planteamiento del Problema.....	2
1.1.2. Justificación.....	2
1.1.3. Análisis Crítico.....	3
1.1.4. Prognosis	3
1.1.5. Formulación del Problema.....	3
1.1.6. Interrogante	4
1.1.7. Delimitación del Objetivo de Investigación.....	4
1.2. Objetivos	4
1.2.2. Específicos.....	5
2. Capítulo I. Fundamentación Teórica.....	6
2.1. Área de Investigación.....	7
2.1.1. La Ciencia de Datos aplicada a Plataformas Digitales.....	7
- Análisis de Contenido Mediante Ciencia de Datos.....	8
- Análisis de Métricas de Rendimiento.	9
- La necesidad de Sistemas de Recomendación.	10
2.1.2. Extracción y Análisis de Datos en Redes Sociales	11
- Social Media Analytics (SMA) y Social Network Analysis (SNA)	12
Social Media Analytics (SMA).....	12
Social Network Analysis (SNA).....	12
- Técnicas de Extracción de Datos en YouTube	13
- Metodologías de Análisis de Datos Aplicadas a YouTube	13
2.2. Estado del Arte y Antecedentes	15
- Perspectivas Nativas: YouTube Studio Analytics.....	16
- Plataformas de Análisis de Terceros	18
2.3. Conceptos Clave	21
2.3.1. ¿Qué es YouTube como ecosistema de datos?	21
2.3.2. ¿Qué es el análisis comparativo de contenido?.....	21
2.3.3. ¿Cómo funciona la YouTube Data API v3?	22
2.3.4. ¿Qué técnicas de análisis de datos se pueden usar para comparar videos?....	22
2.4. Tecnologías y Herramientas Utilizadas	23
2.5. Propositiones Teóricas Generales	26
2.5.1. Proposición 1: “La identificación de patrones de interacción entre videos permite inferir preferencias de audiencia.”	27
2.5.2. Proposición 2: “El análisis comparativo de métricas puede optimizar las	

estrategias de contenido en plataformas de video.”	28
3. Capítulo II. Materiales y Métodos	29
3.1 Enfoque General de la Investigación	30
3.1.1 Modalidad o Enfoque de la investigación.....	30
3.1.2 Tipo de Investigación.....	31
3.2 Metodología para el Cumplimiento de los Objetivos Específicos	31
3.2.1 Desarrollo de la Interfaz de Usuario Intuitiva y Funcional.....	32
3.2.2 Implementación del Sistema de Extracción de Datos	34
3.2.3 Diseño y Aplicación de Algoritmos de Comparación y Análisis de Datos.....	37
3.3 Técnicas e Instrumentos.....	42
3.3.1 Instrumento Principal.....	42
3.3.2 Herramienta de Desarrollo	43
3.3.3 Bibliotecas de Python.....	43
3.3.4 Procesamiento y Análisis de Datos	44
3.3.5 Control de Versiones.....	45
3.4 Métodos Específicos de la Especialidad	45
3.4.1 Fundamentos de Gestión y Análisis Estadístico de Datos Digitales	46
3.4.2 Aplicación de Técnicas de Minería de Datos en Contenido Digital	46
4. Capítulo III. Resultados y Discusión	48
4.1 Resultados	48
4.1.1 Resultado del Objetivo Específico 1: Implementación de una Interfaz de Usuario Intuitiva y Funcional	48
4.1.2 Resultado del Objetivo Específico 2: Desarrollo del Sistema de Extracción de Datos	48
4.1.3 Resultado del Objetivo Específico 3: Aplicación de Algoritmos de Comparación y Análisis	51
5. Conclusiones y Recomendaciones	58
6. Referencias Bibliográficas	60
Trabajos citados	60

PROBLEMA DE INVESTIGACIÓN

1. Introducción

En la era digital contemporánea, la creación y el consumo de contenido audiovisual han experimentado un auge sin precedentes, transformando la manera en que interactuamos con la información, el entretenimiento y la cultura. En el epicentro de esta revolución se encuentra YouTube, la plataforma líder mundial para compartir videos en línea, que ha democratizado la producción de contenido y ha creado una comunidad global de creadores y espectadores.

Con más de mil millones de horas de video vistas cada día y millones de usuarios activos mensuales, YouTube se ha convertido en un gigantesco repositorio de conocimiento, entretenimiento y creatividad. Sin embargo, esta abundancia de contenido también plantea desafíos significativos. Para los creadores de contenido, la competencia por la atención del espectador es intensa, y comprender qué tipos de videos resuenan mejor con la audiencia se ha vuelto fundamental para el éxito en la plataforma. Por otro lado, para los usuarios, encontrar contenido relevante y de calidad entre miles de millones de videos puede resultar abrumador. En este contexto, surge la necesidad de desarrollar herramientas y metodologías que permitan comprender y aprovechar al máximo el vasto ecosistema de YouTube. La ciencia de datos emerge como una disciplina clave en este sentido, ofreciendo la capacidad de analizar grandes volúmenes de datos y extraer insights significativos que pueden informar decisiones estratégicas tanto para los creadores como para los usuarios.

El presente trabajo se sitúa en este contexto, proponiendo el desarrollo de un comparador de videos en YouTube que aproveche la YouTube Data API v3 para permitir la extracción, comparación, interpretación y análisis de datos clave de los videos disponibles en la plataforma. Esta herramienta no solo buscará ayudar a los creadores de contenido a entender las tendencias, el rendimiento y el impacto de sus videos, sino también a proporcionar a los usuarios una herramienta útil para descubrir contenido relevante y de calidad en el vasto océano de videos en línea.

1.1. Planteamiento del Problema

1.1.1. Contextualización

YouTube, como la plataforma de video más grande del mundo, alberga una diversidad sin igual de contenido que abarca desde blogs personales hasta tutoriales especializados y transmisiones en vivo. Esta vasta biblioteca de videos no solo refleja la variedad de intereses y pasiones de la audiencia global, sino que también presenta un desafío para los creadores y analistas de datos que buscan comprender las dinámicas subyacentes del contenido y su impacto en la plataforma. En este contexto, la capacidad de analizar y comparar videos de manera efectiva se convierte en un recurso invaluable para identificar tendencias emergentes, comprender el comportamiento del espectador y optimizar estrategias de contenido para mejorar el rendimiento del canal.

1.1.2. Justificación

Ante la vasta cantidad de contenido disponible en YouTube y la competencia cada vez mayor por la atención del espectador, es crucial para los creadores de contenido y analistas de datos contar con herramientas que les permitan comprender y optimizar sus estrategias de contenido. El desarrollo de un comparador de videos en YouTube se presenta como una solución clave para abordar este desafío, ya que proporcionaría una forma eficiente y sistemática de analizar el contenido disponible en la plataforma, identificar tendencias emergentes y evaluar el rendimiento de los videos en relación con diferentes métricas clave. Al ofrecer *insights* significativos sobre el contenido más efectivo y relevante, esta herramienta no solo ayudaría a los creadores a mejorar la calidad y la relevancia de su contenido, sino que también podría mejorar la experiencia del usuario al facilitar la búsqueda y descubrimiento de videos que se adapten a sus intereses y preferencias individuales.

1.1.3. Análisis Crítico

El desarrollo de un comparador de videos en YouTube representa un enfoque innovador y práctico para abordar los desafíos actuales en la gestión y optimización de contenido en la plataforma. Al aprovechar la YouTube Data API v3, esta herramienta ofrece la capacidad de realizar un análisis detallado y sistemático de los videos disponibles en YouTube, proporcionando insights valiosos que pueden informar decisiones estratégicas tanto para los creadores de contenido como para los usuarios.

Sin embargo, es importante tener en cuenta los posibles desafíos y limitaciones asociados con el acceso y la interpretación de los datos, así como la necesidad de garantizar la privacidad y la ética en el manejo de la información de los usuarios.

1.1.4. Prognosis

Se espera que el comparador de videos en YouTube proporcione una herramienta eficaz y accesible para los creadores de contenido y los analistas de datos, facilitando la comprensión y optimización del contenido en la plataforma. Con su capacidad para extraer insights significativos de los datos disponibles en YouTube, se anticipa que esta herramienta contribuirá a mejorar la calidad del contenido, aumentar el compromiso del espectador y optimizar el rendimiento de los canales en la plataforma.

1.1.5. Formulación del Problema

Pregunta Central de Investigación: ¿Cómo podemos desarrollar un comparador de videos en YouTube que permita a los creadores de contenido y analistas de datos extraer, comparar y analizar datos clave de los videos disponibles en la plataforma para mejorar estrategias de contenido y rendimiento del canal de manera efectiva y eficiente?

1.1.6. Interrogante

Preguntas Secundarias:

- ¿Cuáles son las métricas clave que deben considerarse al comparar videos en YouTube para evaluar su rendimiento y relevancia?
- ¿Cómo podemos diseñar y desarrollar una interfaz intuitiva y fácil de usar para el comparador de videos en YouTube que satisfaga las necesidades tanto de los creadores de contenido como de los usuarios?
- ¿Qué técnicas y algoritmos de análisis de datos son más adecuados para extraer insights significativos de los datos disponibles en la YouTube Data API v3 y cómo pueden aplicarse de manera efectiva en el desarrollo del comparador de videos?

1.1.7. Delimitación del Objetivo de Investigación

Este estudio se enfocará en el desarrollo de un comparador de videos en YouTube que permita la comparación de dos vídeos seleccionados, utilizando la YouTube Data API v3. El objetivo principal es extraer, comparar, interpretar y analizar datos clave de los vídeos para mejorar la comprensión de las tendencias de contenido, el rendimiento del canal y las estrategias de contenido. La aplicación se centrará inicialmente en la comparación de metadatos y estadísticas básicas de los vídeos, con potencial para futuras actualizaciones y expansiones para incluir funcionalidades adicionales, como análisis de contenido multimedia y recomendaciones personalizadas.

1.2. Objetivos

1.2.1. General

Desarrollar un comparador de videos de YouTube que permita la extracción, comparación, interpretación y análisis de datos clave de los vídeos disponibles en la plataforma, con el fin de mejorar la comprensión de las tendencias de contenido, el rendimiento del canal y las estrategias de contenido, usando la YouTube Data API v3.

1.2.2. Específicos

- Implementar una interfaz de usuario intuitiva y funcional para el comparador de videos en YouTube, que permita a los usuarios seleccionar y comparar dos o tres vídeos de manera sencilla.
- Desarrollar un sistema de extracción de datos utilizando la YouTube Data API v3 para obtener metadatos y estadísticas relevantes de los vídeos seleccionados, incluyendo vistas, me gusta, comentarios, entre otros.
- Diseñar y aplicar algoritmos de comparación y análisis de datos para identificar patrones, tendencias y diferencias significativas entre los vídeos seleccionados, con el objetivo de proporcionar *insights* útiles para los creadores de contenido.

2. Capítulo I. Fundamentación Teórica

El presente capítulo tiene como propósito establecer el marco teórico que sustenta el desarrollo de la investigación titulada "Comparador de Videos en YouTube: Análisis de Contenido, Estadísticas y Tendencias para la Mejora de Estrategias de Contenido y Rendimiento del Canal." En esta sección se delimitan las áreas de investigación relacionadas, se compendian los avances científicos y tecnológicos pertinentes, y se exponen los conceptos, tecnologías y teorías fundamentales que respaldan el trabajo.

La investigación se sitúa principalmente en dos áreas del conocimiento: la Ciencia de Datos aplicada a Plataformas Digitales, y la Extracción y Análisis de Datos en Redes Sociales. La primera aborda el tratamiento y análisis de grandes volúmenes de datos generados en entornos digitales, aplicando técnicas estadísticas, de minería de datos y aprendizaje automático para descubrir patrones relevantes. La segunda área se enfoca en la obtención, procesamiento y estudio de información proveniente de plataformas sociales, como YouTube, para entender fenómenos de interacción, visualizaciones según el tópico, consumo de contenido, tendencias emergentes y análisis de sentimientos en comentarios. A lo largo de este capítulo se revisarán los antecedentes más relevantes de los últimos años, que incluyen investigaciones previas y proyectos relacionados con el análisis de datos de videos en YouTube y el uso de APIs para la extracción de información. Además, se explicarán los conceptos clave como la caracterización de métricas de desempeño en YouTube, los métodos de comparación de contenido digital y las técnicas de análisis de tendencias. También se abordarán las tecnologías fundamentales que posibilitan el proyecto, entre ellas la YouTube Data API v3, así como los principios teóricos que guían el análisis y visualización de datos.

Con esta fundamentación teórica, se busca proporcionar una base sólida para entender el contexto, alcance y pertinencia del proyecto de investigación, asegurando que las decisiones metodológicas y los desarrollos tecnológicos posteriores estén respaldados por el conocimiento científico existente.

2.1. Área de Investigación

2.1.1. La Ciencia de Datos aplicada a Plataformas Digitales

La Ciencia de Datos se ha convertido en un pilar fundamental para comprender grandes conjuntos de datos en general, también ha incursionado en optimizar el funcionamiento de las plataformas digitales, que es lo que nos compete en este estudio. Esto se define como un campo interdisciplinario que utiliza métodos científicos, procesos, algoritmos y sistemas para extraer conocimiento y perspectivas de datos estructurados y no estructurados. El campo de Big Data ha ganado una popularidad inmensa debido a los grandes avances en tecnología a nivel de hardware y software, ahora existen centros de datos enfocados en el proceso masivo y a gran escala de gigantescas cantidades de información, a la par las herramientas de software vanguardistas son capaces de manejar algoritmos más complejos y aprovecharse de las capacidades del hardware actual. Nace el concepto de BDA (Big Data Analytics), una disciplina emergente que ha revolucionado las herramientas de estadísticas. Acorde a diversos reportes de estudios, existen empresas que han hecho inversiones en inteligencia de negocios y herramientas de análisis gracias a los enormes beneficios que estos brindan, por lo cual el problema que se cree se afrontará en un futuro es la brecha que exista entre las capacidades de estas herramientas y la cantidad de personas con talento capaz de cubrir estas necesidades. (Junling Zhang, 2022)

En el contexto de las plataformas digitales, su aplicación abarca desde la personalización de la experiencia del usuario hasta la optimización de estrategias de contenido y la medición del rendimiento. Su aplicación permite ir más allá de la simple descripción de métricas, posibilitando la identificación de patrones complejos, la predicción de tendencias y la optimización de estrategias basadas en evidencia empírica.

Se sabe que YouTube es el sitio más famoso de videos en el mundo, se tienen cifras increíbles de usuarios viendo 4 mil millones de horas de videos consistentemente. Desde los inicios de YouTube en febrero de 2005, muchas celebridades, empresas y personas en general han compartido una inmensa

cantidad de videos convirtiendo a la plataforma en un repositorio audiovisual de la humanidad. Desde ese entonces hasta el día de hoy, YouTube ha crecido inmensamente en visualizaciones, contenido, usuarios y variedad. Como era de esperarse, una rueda interminable de datos e información se genera y viene en un flujo desestructurado; gigantescas cantidades de información no pueden ser examinadas por ciclos y algoritmos informáticos simples. Es aquí donde entra el análisis paralelo de estas cantidades de información, sistemas de análisis distribuidos, data warehouses de información, y el Big Data y Ciencia de Datos aparecen. Desde que esta data es agregada colosalmente día a día con una velocidad extraordinaria, ha nacido un gran interés en almacenar, limpiar procesar y analizar meticulosamente esta inmensa proporción de información para hacerla usable y entendible. La Ciencia de Datos proporciona el marco metodológico y las herramientas tecnológicas (como Python y sus librerías especializadas) para abordar este desafío. (Eliganti Ramalakshmi, 2022)

- Análisis de Contenido Mediante Ciencia de Datos.

El análisis de contenido en plataformas digitales implica examinar sistemáticamente los materiales comunicativos como texto, imágenes y video para identificar patrones, temas, sentimientos y otras características relevantes. La Ciencia de Datos potencia este análisis mediante la aplicación de técnicas computacionales que permiten procesar grandes cantidades de contenido de manera eficiente y objetiva.

Actualmente, el análisis de contenido digital es potenciado por técnicas computacionales que permiten automatizar la exploración de grandes volúmenes de datos. Herramientas de procesamiento de lenguaje natural (NLP), minería de texto, análisis de sentimiento, reconocimiento de patrones visuales y análisis semántico se integran a flujos de trabajo basados en machine learning para ofrecer una comprensión más profunda, objetiva y escalable de los contenidos. En el caso de YouTube, este enfoque permite analizar desde los títulos y descripciones de los videos hasta los comentarios, etiquetas y transcripciones automáticas, aportando una visión detallada de las estrategias de comunicación utilizadas y de cómo estas inciden en la recepción por parte del público.

Esta capacidad analítica es clave para establecer comparaciones entre videos, evaluar su rendimiento relativo y generar recomendaciones de mejora. Al vincular el contenido con métricas como el engagement, la retención de audiencia o el alcance orgánico, se puede entender no solo qué se comunica, sino cómo se consume e interpreta. De esta manera, la Ciencia de Datos se convierte en una herramienta estratégica para transformar el contenido en información accionable, alineada con los objetivos de optimización y crecimiento de los canales digitales. (T. Aleksandrova, 2019)

- Análisis de Métricas de Rendimiento.

Aplicar técnicas de Big Data a grandes clústeres de datos darán perspectivas muy interesantes a los procesos de negocios que se puedan llevar a cabo en cualquier entidad permitiendo habilitarlos para hacer un marketing más especializado y efectivo gracias al empleo de técnicas analíticas como el minado de datos. El uso de Big Data también permite que los usuarios de plataformas como YouTube tengan facilidades para interactuar con videos que son de su interés, canales recomendados en base a su perfil de preferencias, y también recomendar videos que son de su gusto o están en tendencia y son vistos por muchísimos usuarios. También aquí entra la recomendación de videos, aquel contenido que el algoritmo interpreta que serán de gran interés para el usuario. Es aquí donde YouTube tiene un gran potencial para ser exitoso en consumidores de cualquier tipo. Con los cientos de millones de usuarios dentro de la plataforma, indirectamente se está creando nueva data que es procesada y analizada para nuevos resultados. Para saber qué videos están en tendencia, YouTube usa varios factores incluyendo medición de participación del usuario (fechas de tendencia, videos totales, región y categorías).

Archivar, procesar y analizar eficientemente tal cantidad de datos generada por YouTube es extremadamente difícil. La visualización de data representa la información a través de gráficos, cuadros u otros formatos visuales, esto ayuda a que ciertos patrones y tendencias sean vistos fácilmente. El Machine Learning hace más fácil manejar análisis y métricas de rendimiento y análisis predictivos. En estos tiempos, el análisis eficiente de los datos empresariales y su

almacenamiento en dispositivos adecuados resulta tedioso. Los vídeos de YouTube y, por lo tanto, los datos multimedia que se generan a partir de ellos, suelen estar desestructurados. El análisis perfecto de estos datos multimedia semiestructurados y desestructurados puede suponer un gran reto. (JOHANES FERNANDES ANDRY, 2021)

El análisis de métricas de rendimiento en YouTube es fundamental para el caso de estudio planteado en esta tesis, ya que permite transformar datos complejos y en muchos casos desestructurados en información estratégica y comprensible. Variables como el número de vistas, la retención de audiencia, el engagement, las tasas de clics o la frecuencia de interacción ofrecen una radiografía precisa del comportamiento del usuario frente al contenido. Tal como se menciona en el estudio de Johanes Fernandes Andry (2021), el volumen masivo de datos generados en YouTube requiere del uso de técnicas avanzadas de Big Data, visualización de datos y machine learning para ser procesado adecuadamente. En este contexto, aplicar ciencia de datos para comparar y analizar el rendimiento entre videos no solo optimiza las estrategias de contenido, sino que además permite entender las dinámicas que impulsan el crecimiento o estancamiento de un canal. Así, el análisis de estas métricas no es solo una herramienta técnica, sino una necesidad estratégica para cualquier creador o empresa que busque posicionarse eficazmente dentro de la plataforma.

- La necesidad de Sistemas de Recomendación.

Desde que se notó que el explosivo crecimiento en la abrumadora cantidad de información que podemos encontrar en cualquier medio digital iba a ser constante e incesante, y que la cantidad de visitantes a Internet es equiparable con la población mundial, se ha creado un reto para el ser humano de tener acceso a ciertos recursos de interés específico. Pese a que motores de búsqueda y recuperación de información como Google o DevilFinder han solucionado este problema parcialmente, se sentía la ausencia de un sistema personalizado basado en las preferencias del usuario.

Es aquí donde nacen los Sistemas de Recomendación al Usuario. Son sistemas de filtrado que solucionan el problema de la sobrecarga de información recopilando aquellos

fragmentos de información vital a partir de muchas fuentes. Estos tienen la habilidad de predecir si un usuario preferirá un ítem o no basado en la actividad de su perfil. Los beneficios de esta clase de sistemas de filtrado son enormes. Entre reducir los costos transaccionales de encontrar y seleccionar el mejor contenido, la facilidad de cautivar al usuario con opciones que realmente le interesan y demás. (F.O. Isinkaye, 2015)

Con base en esto, se sabe que los sistemas de recomendación representan un eje crucial dentro de la ciencia de datos aplicada a plataformas digitales porque actúan como una respuesta inteligente al exceso de información, permitiendo identificar patrones de comportamiento del usuario y adaptarse a sus preferencias específicas. En entornos como YouTube, donde la cantidad de contenido nuevo que se publica cada minuto es abrumadora, estos sistemas no solo ayudan a los usuarios a descubrir videos relevantes, sino que también se convierten en herramientas estratégicas para los creadores de contenido. A través del análisis de métricas como tiempo de reproducción, interacciones o frecuencia de clics, es posible alimentar modelos predictivos que optimicen la visibilidad y el rendimiento de los videos.

2.1.2. Extracción y Análisis de Datos en Redes Sociales

El análisis de datos en redes sociales se ha convertido en una disciplina fundamental para comprender cómo se construyen y difunden las dinámicas digitales entre los usuarios y las plataformas. Este tipo de análisis no se limita solo a recolectar grandes volúmenes de datos, sino que implica estudiar las formas en que los usuarios interactúan con el contenido y entre sí, generando rastros digitales que pueden ser procesados para revelar patrones de comportamiento, preferencias e influencias. En plataformas de alta demanda como YouTube, Facebook o X, estos datos incluyen reacciones, comentarios, suscripciones, historial de visualización y otras señales de interacción que, una vez estructuradas y analizadas, permiten extraer información valiosa para múltiples propósitos, desde la personalización de la experiencia del usuario hasta la evaluación del rendimiento del contenido. (KOÇ, 2023)

A diferencia de otras redes, YouTube integra dinámicas sociales en un entorno predominantemente audiovisual, lo que añade capas de complejidad al proceso de

extracción y análisis de datos. Esta plataforma no solo funciona como un medio de consumo pasivo, sino como un ecosistema donde los algoritmos de recomendación, alimentados por la actividad de los usuarios, desempeñan un papel determinante en la visibilidad de los contenidos. Comprender estos mecanismos desde la perspectiva del análisis de datos implica no solo examinar qué se visualiza, sino cómo y por qué ciertos videos alcanzan mayor difusión o engagement. En este sentido, el estudio del comportamiento en redes sociales como YouTube se convierte en un componente esencial para quienes buscan desarrollar herramientas analíticas que mejoren el diseño de estrategias digitales, como lo propone el presente trabajo. (JOHANES FERNANDES ANDRY, 2021)

- Social Media Analytics (SMA) y Social Network Analysis (SNA)

Social Media Analytics (SMA)

Se refiere al proceso de recopilación y análisis de datos de redes sociales para tomar decisiones de negocio o investigación. Implica monitorizar, medir, y analizar conversaciones y tendencias en redes sociales para entender percepciones, rendimiento de campañas, y comportamiento del consumidor. Métricas como el engagement (likes, comentarios, compartidos), alcance, impresiones, y análisis de sentimiento son componentes clave de SMA. (T. Aleksandrova, 2019)

Social Network Analysis (SNA)

Se enfoca en el mapeo y medición de relaciones y flujos entre personas, grupos, organizaciones, u otras entidades conectadas en una red social. Utiliza teoría de grafos para visualizar y analizar estructuras de red, identificar nodos influyentes (centralidad), detectar comunidades o clústeres, y comprender cómo la información o influencia se propaga a través de la red. Aunque menos enfatizado en los snippets para YouTube específicamente, SNA puede aplicarse al análisis de redes de comentarios, colaboraciones entre canales, o patrones de suscripción. (JARMAN, 2022)

- Técnicas de Extracción de Datos en YouTube

La obtención de datos de YouTube es el primer paso crucial. Las principales fuentes y métodos incluyen:

API de YouTube (YouTube Data API & YouTube Analytics API): Es la vía oficial y más robusta para acceder a datos públicos (información de videos, canales, comentarios) y datos privados del propio canal (métricas de rendimiento detalladas, demografía de la audiencia) de forma estructurada y respetando los términos de servicio de la plataforma. Permite recopilar datos como títulos, descripciones, etiquetas, estadísticas de visualización, likes, dislikes, comentarios, información del canal, y métricas de rendimiento como tiempo de visualización, retención de audiencia, fuentes de tráfico, y demás. (upwork, 2025)

Web Scraping: Consiste en extraer datos directamente del código HTML de las páginas web de YouTube. Si bien puede usarse para obtener datos no disponibles fácilmente a través de la API (por ejemplo, datos históricos extensos o de canales específicos sin acceso API), presenta desafíos técnicos (cambios en la estructura del sitio web) y éticos/legales (puede violar los términos de servicio de YouTube). Su uso debe ser cuidadoso y considerado. (Derks, 2020)

Datos Secundarios: Utilización de datasets preexistentes o informes publicados por terceros que contienen datos de YouTube, como Kaggle.

- Metodologías de Análisis de Datos Aplicadas a YouTube

Una vez extraídos los datos, se aplican diversas metodologías de análisis:

- **Procesamiento del Lenguaje Natural (NLP) para Comentarios y Transcripciones:**
 - *Análisis de Sentimiento:* Determinar la polaridad (positiva, negativa, neutra) de los comentarios para medir la recepción de un video o canal. Herramientas de IA y plataformas analíticas a menudo incorporan esta funcionalidad. (Othman, 2022)

- *Modelización de Tópicos (Topic Modeling)*: Identificar los temas principales discutidos en los comentarios o en las transcripciones de los videos utilizando algoritmos como Latent Dirichlet Allocation (LDA). Esto ayuda a comprender los intereses de la audiencia y las temáticas centrales del contenido. (Derks, 2020)
- *Resumen Automático (Summarization)*: Generar resúmenes concisos del contenido de los videos a partir de sus transcripciones, facilitando la búsqueda y el consumo de información.
- *Análisis de Comentarios*: Estudios cualitativos y cuantitativos de comentarios para entender la interacción, el discurso público, y la recepción del contenido.
- **Visión por Computador para Análisis de Video:**
 - Aunque la investigación proporcionada no se centra extensamente en esto, las técnicas de Visión por Computador (CV) pueden aplicarse para analizar el contenido visual de los videos. Esto podría incluir la detección de objetos, reconocimiento facial, análisis de escenas, o extracción de texto incrustado en el video. La integración de CV con NLP permite un análisis multimodal más rico. (Ahmed Al Kuwaiti, 2023)
- **Análisis Estadístico de Métricas de Interacción:**
 - Análisis descriptivo y exploratorio de datos (EDA) para resumir y visualizar métricas clave como vistas, likes, dislikes, comentarios, compartidos, suscriptores, tiempo de visualización, retención de audiencia, etc..
 - Análisis inferencial para probar hipótesis sobre qué factores (del contenido, del canal, de la audiencia) influyen en estas métricas.
 - Análisis de series temporales para identificar tendencias y estacionalidad en el rendimiento del canal o de videos específicos.
 - Comparación de métricas entre diferentes videos, canales o estrategias de contenido. (Maryland, 2025)

- **Sistemas de Recomendación:**
 - Comprender los principios detrás de los algoritmos de recomendación de YouTube es fundamental, ya que influyen enormemente en la visibilidad y el rendimiento del contenido. Estos sistemas utilizan el historial de visualización, las interacciones (likes, dislikes, no interesados), metadatos del video, y características del usuario para sugerir videos relevantes, operando bajo principios de filtrado colaborativo, basado en contenido o híbrido. La robustez de estos sistemas frente a datos ruidosos o ataques maliciosos es un área de investigación activa, ya que datos "sucios" pueden llevar a recomendaciones anómalas. (F.O. Isinkaye, 2015)

2.2. Estado del Arte y Antecedentes

Entender las analíticas de YouTube es crítico para muchas partes dentro del contexto de redes sociales y entretenimiento, tanto para marketers de video, creadores de contenido y científicos de datos. Es por esto que existen un montón de herramientas que pueden ayudar en la mayoría de estos casos, estas herramientas analíticas nos ayudan a entender ciertas métricas clave de YouTube para mejorar el rendimiento del canal y enfoque de videos individuales. Se provee reportería, análisis y comparación de capacidades para monitorear el rendimiento a través del tiempo, comparando tipos de contenido, obtener métricas y perspectivas para videos y shorts. (Newberry, 2025)

Actualmente, el análisis de contenido digital es potenciado por técnicas computacionales que permiten automatizar la exploración de grandes volúmenes de datos. Herramientas de procesamiento de lenguaje natural, minería de texto, análisis de sentimiento, reconocimiento de patrones visuales y análisis semántico se integran a flujos de trabajo basados en machine learning para ofrecer una comprensión más profunda, objetiva y escalable de los contenidos. En el caso de YouTube, este enfoque permite analizar desde los títulos y descripciones de los videos hasta los comentarios, etiquetas y transcripciones automáticas, aportando una visión detallada de las estrategias de comunicación utilizadas y de cómo estas inciden en la recepción por parte del público.

Esta capacidad analítica es clave para establecer comparaciones entre videos, evaluar su rendimiento relativo y generar recomendaciones de mejora. Al vincular el contenido con métricas como el engagement, la retención de audiencia o el alcance orgánico, se puede entender no solo qué se comunica, sino cómo se consume e interpreta. De esta manera, la Ciencia de Datos se convierte en una herramienta estratégica para transformar el contenido en información accionable, alineada con los objetivos de optimización y crecimiento de los canales digitales. (Zote, 2024)

El contexto actual está marcado por la creciente influencia de la inteligencia artificial (IA) en la analítica, prometiendo perspectivas más profundas y automatización, así como una tendencia hacia el análisis multiplataforma para obtener una visión holística del rendimiento del contenido en el panorama digital. Navegar por este complejo ecosistema de herramientas y métricas es esencial para tomar decisiones informadas y optimizar las estrategias de contenido de vídeo. (Samia Idbenssi, 2023)

- Perspectivas Nativas: YouTube Studio Analytics

YouTube Studio se presenta como la fuente principal y gratuita de datos de rendimiento, proporcionada directamente por la plataforma a todos sus creadores. Su accesibilidad e integración directa lo convierten en el punto de partida fundamental para cualquier análisis de canal. (@Iconosquare, 2025)

La forma en que YouTube Studio organiza sus datos analíticos está enfocada en diferentes aspectos de rendimiento nativos propias para el canal.

- Visión General: Ofrece métricas de rendimiento genéricas para cada canal. Se podrán encontrar métricas clave como total de visualizaciones, cantidad de likes y dislikes, tiempo de visualización en horas, gráfica de fluctuación de suscriptores, y ranking de videos más vistos.
- Alcance: Está centrado en cómo los espectadores descubren el contenido del canal. Estas métricas funcionan como un píxel que determina cuántas veces el usuario vio la miniatura del canal y si se le dio clic o no para ver el video, tasa de

clics (CTR), espectadores únicos y fuentes de tráfico. Comprender estas métricas es clave para optimizar la estrategia de descubrimiento.

- Interacción: Esta métrica se encarga de medir el cómo la audiencia del canal interactúa con los videos. Estas claves son el tiempo de visualización en relación a la duración total del video, la retención de audiencia, los puntos más vistos del video (ondas de momento más reproducido), y duración media de reproducciones. Podríamos decir que la retención de audiencia es de los puntos más importantes de este apartado.

- Audiencia: Proporciona información demográfica y de comportamiento sobre los espectadores. Las métricas incluyen espectadores recurrentes vs. nuevos espectadores, espectadores únicos, suscriptores ganados y perdidos, los momentos en que los espectadores están activos en YouTube, otros canales o vídeos que la audiencia ve, edad y género, principales ubicaciones geográficas y uso de subtítulos. Estos datos son cruciales para adaptar el contenido a las preferencias de la audiencia.

- Ingresos: Disponible para los miembros del Programa de Partners de YouTube (YPP). Muestra los ingresos estimados totales, los ingresos por mil visualizaciones (RPM), el coste por mil impresiones basado en reproducciones (CPM), los vídeos con mayores ingresos, los tipos de anuncios que se muestran, los ingresos por transacciones (como membresías del canal o Super Chat/Stickers) y los ingresos de YouTube Premium.

Fortalezas y Limitaciones

La principal fortaleza de YouTube Studio Analytics es que es gratuito, está directamente integrado en la plataforma y proporciona datos fundamentales de primera mano. Ofrece métricas únicas, como el comportamiento de la audiencia en toda la plataforma (qué otros canales ven).

Sin embargo, sus limitaciones son notables. Se centra principalmente en los datos del propio canal, ofreciendo capacidades limitadas para la evaluación comparativa directa y profunda con competidores específicos. Puede carecer de las funciones especializadas que se encuentran en herramientas de terceros, como la

investigación avanzada de palabras clave para SEO, el análisis detallado de sentimientos o el seguimiento del rendimiento multiplataforma. (maekersuite, s.f.)

El énfasis puesto en métricas como el Tiempo de Visualización y la Retención de la Audiencia subraya su papel crítico en el algoritmo de recomendación de YouTube. Un tiempo de visualización elevado y una fuerte retención indican que los espectadores encuentran el contenido atractivo y valioso, lo que satisface el objetivo de YouTube de mantener a los usuarios en la plataforma por más tiempo. En consecuencia, el algoritmo probablemente prioriza los vídeos y canales que demuestran un rendimiento sólido en estas métricas de interacción, sugiriendo que la calidad del contenido y la satisfacción del espectador son primordiales para el crecimiento orgánico, por encima de la simple optimización para obtener clics (CTR). Finalmente, aunque YouTube Studio es completo para el análisis del propio canal, su falta de capacidades profundas de análisis directo de la competencia crea una oportunidad de mercado significativa. Los creadores y especialistas en marketing necesitan comprender su panorama competitivo para identificar oportunidades y amenazas. Esta brecha entre la necesidad de inteligencia competitiva y las limitaciones de la herramienta nativa impulsa directamente la demanda de plataformas externas especializadas en análisis de la competencia. (@Iconosquare, 2025)

- Plataformas de Análisis de Terceros

Dadas las limitaciones de la analítica nativa, especialmente en lo referente al análisis competitivo y funcionalidades avanzadas, los creadores y profesionales del marketing recurren a una amplia gama de herramientas de terceros. Estas soluciones externas ofrecen capacidades extendidas, como información más profunda sobre competidores, investigación avanzada de palabras clave y SEO, análisis multiplataforma, informes especializados, automatización de flujos de trabajo y métricas únicas que no se encuentran en YouTube Studio. (Yadin, 2024)

Categorización de Herramientas de Terceros:

- Suites de Gestión de Redes Sociales Todo en Uno: Plataformas que integran la analítica de YouTube junto con la programación, gestión e informes para múltiples redes sociales. (Ejemplos: Sprout Social, Hootsuite, Metricool)
- Herramientas para crecimiento y SEO de YouTube: Soluciones que se centran en optimizar rendimientos del canal de YouTube, a menudo implementadas como extensiones de navegador (Ejemplos: TubeBuddy, VidIQ)
- Plataformas de Análisis Competitivo y Benchmarking: Herramientas especializadas en el seguimiento y la comparación del rendimiento entre múltiples canales. (Ejemplos: Social Blade, Viewstats, HypeAuditor , Unmetric)
- Plataformas de Inteligencia de Vídeo e Investigación de Mercado: Herramientas de nivel empresarial que proporcionan tendencias amplias del mercado, inteligencia de contenido y análisis de audiencia. (Ejemplos: Tubular Labs, Brandwatch)

Perfiles Detallados de Herramientas Seleccionadas:

- TubeBuddy / VidIQ: Estas dos herramientas son competidores directos y comparten muchas características, operando principalmente como extensiones de navegador. Ofrecen investigación de palabras clave, sugerencias de etiquetas, puntuaciones de optimización SEO, análisis rápidos de competidores (Videolytics en TubeBuddy), pruebas A/B para miniaturas y títulos, y sugerencias sobre el mejor momento para publicar.
- Social Blade: Ampliamente reconocida por rastrear estadísticas públicas de canales (suscriptores, visualizaciones) y proporcionar rankings. Sin embargo, su característica de estimación de ganancias es notoriamente inexacta. Los rangos proporcionados suelen ser extremadamente amplios y a menudo difieren significativamente de los ingresos reales reportados por los creadores. Su valor principal reside en el seguimiento de tendencias de crecimiento a largo plazo y la comparación básica de métricas públicas, pero las estimaciones de ingresos deben tratarse con extrema precaución.
- Viewstats: Una plataforma de análisis más reciente, asociada con el creador MrBeast. Ofrece una extensión de Chrome y un sitio web con

características como búsqueda de vídeos atípicos (Outliers), búsqueda de miniaturas por similitud visual o palabras clave, seguimiento de pruebas A/B realizadas por otros canales y una función de Colecciones para guardar ideas. Su plan Pro tiene un coste de \$49.99 USD al mes, con un descuento por suscripción anual. Se posiciona como una herramienta avanzada para creadores que buscan insights más profundos.

Consideraciones finales:

El ecosistema de herramientas de análisis de YouTube ajenas a YouTube Analytics se caracteriza por una notable segmentación orientada a diversos perfiles de usuario, desde creadores individuales que buscan mejorar su posicionamiento mediante plataformas como TubeBuddy o VidIQ, hasta empresas y agencias que requieren soluciones avanzadas como Tubular Labs o Sprout Social para gestionar múltiples canales y generar reportes integrados. Esta diversidad se refleja también en los modelos de negocio, donde predominan las versiones freemium para usuarios con recursos limitados, mientras que las suites más completas están destinadas a quienes pueden invertir en funcionalidades de análisis profundo. Sin embargo, persiste el desafío de la precisión en estimaciones económicas, como lo evidencia Social Blade, cuyas cifras deben interpretarse con cautela debido a la complejidad de factores que afectan los ingresos reales de un canal. A pesar de estas limitaciones, se observa una evolución en la industria con la aparición de plataformas más sofisticadas como Viewstats, que integran conocimientos del propio ecosistema de creadores y grandes volúmenes de datos, ofreciendo así análisis más precisos y estratégicos para optimizar el rendimiento en la plataforma. (Cena, 2024)

2.3. Conceptos Clave

2.3.1. ¿Qué es YouTube como ecosistema de datos?

YouTube, más que una simple plataforma de distribución de videos, constituye un ecosistema de datos dinámico, complejo y en constante expansión. Cada interacción realizada por los usuarios desde visualizaciones, comentarios y "me gusta", hasta suscripciones, búsquedas y compartidos genera una enorme cantidad de información digital. Esta data se agrupa en formatos estructurados y no estructurados que representan comportamientos, preferencias, hábitos de consumo y tendencias globales. En este contexto, YouTube se convierte en una fuente rica de Big Data que, al ser correctamente extraída, procesada y analizada, permite generar conocimiento útil para la toma de decisiones en áreas como marketing digital, producción de contenido, investigación de audiencias, y optimización de estrategias comunicativas. Su rol como ecosistema de datos lo posiciona como un objeto de estudio valioso para la Ciencia de Datos aplicada a plataformas digitales. (JOHANES FERNANDES ANDRY, 2021)

2.3.2. ¿Qué es el análisis comparativo de contenido?

El análisis comparativo de contenido es una técnica que consiste en examinar y contrastar dos o más unidades de contenido (por ejemplo, videos de YouTube) para identificar similitudes, diferencias y patrones significativos. Este análisis puede realizarse sobre aspectos cualitativos (temática, estilo narrativo, tipo de audiencia) y cuantitativos (visualizaciones, interacciones, tiempo de reproducción). En el caso de YouTube, esta metodología permite evaluar qué factores inciden en el mejor rendimiento de un video sobre otro, cómo responden los usuarios a distintos tipos de mensajes o formatos, y qué contenidos logran mayor posicionamiento. En un entorno competitivo, este enfoque resulta útil para generar insights accionables que permitan optimizar estrategias de contenido, calendarización de publicaciones, y diseño de campañas. (Eliganti Ramalakshmi, 2022)

2.3.3. ¿Cómo funciona la YouTube Data API v3?

La YouTube Data API v3 es una interfaz de programación de aplicaciones proporcionada por Google que permite acceder y manipular datos públicos de YouTube. A través de solicitudes HTTP y respuestas en formato JSON, los desarrolladores pueden extraer información estructurada sobre videos, canales, listas de reproducción, comentarios y estadísticas asociadas. Esta API se convierte en una herramienta poderosa para la Ciencia de Datos, ya que automatiza la recopilación de grandes volúmenes de datos, permite consultas personalizadas con criterios específicos (por ID, palabra clave, fecha, etc.), y facilita el análisis en tiempo real o histórico. Su implementación requiere autenticación mediante claves API y una gestión adecuada de cuotas para no exceder los límites establecidos por Google. Gracias a su versatilidad, es la base tecnológica sobre la que puede construirse una aplicación web orientada al análisis comparativo y extracción de métricas de rendimiento en YouTube.

2.3.4. ¿Qué técnicas de análisis de datos se pueden usar para comparar videos?

Para realizar un análisis comparativo efectivo entre videos de YouTube, pueden emplearse múltiples técnicas dentro del ámbito de la Ciencia de Datos. Entre las más utilizadas están la estadística descriptiva (para observar distribuciones y tendencias centrales de métricas clave), la visualización de datos (para representar gráficamente el rendimiento), el análisis de correlación (para detectar relaciones entre variables como duración y tasa de retención), el clustering (para agrupar videos con comportamientos similares), y el análisis de sentimiento (aplicado a comentarios para evaluar la percepción del público). También pueden aplicarse modelos predictivos y de clasificación mediante algoritmos de machine learning como regresión lineal, árboles de decisión o redes neuronales, con el objetivo de anticipar el rendimiento de futuros contenidos. Estas técnicas permiten enriquecer la interpretación de datos y aportar profundidad al proceso de comparación entre videos.

2.4. Tecnologías y Herramientas Utilizadas

A continuación, se presenta un listado de las tecnologías y herramientas empleadas en el desarrollo del proyecto, junto con una descripción de su propósito y los beneficios que aportan al cumplimiento de los objetivos planteados:

1. Python

Propósito: Lenguaje de programación principal utilizado para implementar la lógica del proyecto.

Beneficios:

Amplia comunidad de usuarios y soporte, tiene una gran cantidad de bibliotecas especializadas para análisis de datos, visualización y desarrollo web. Sintaxis sencilla y legible, lo que facilita el desarrollo y mantenimiento del código.

2. Flask

Propósito: Framework ligero para el desarrollo de aplicaciones web.

Beneficios:

Permite crear una interfaz web interactiva para que los usuarios puedan acceder a las funcionalidades del proyecto. Su flexibilidad facilita la integración con otras tecnologías y bibliotecas. Ideal para proyectos pequeños y medianos debido a su simplicidad y bajo consumo de recursos.

3. Google API Client (YouTube Data API v3)

Propósito: Biblioteca utilizada para interactuar con la API de datos de YouTube y extraer información sobre canales, videos y comentarios.

Beneficios:

Acceso a datos en tiempo real directamente desde la plataforma de YouTube.

Permite realizar consultas específicas y obtener métricas detalladas, como vistas, likes, comentarios y duración de los videos. Facilita la autenticación segura mediante claves API y OAuth.

4. Pandas

Propósito: Biblioteca para manipulación y análisis de datos estructurados.

Beneficios:

Permite transformar los datos obtenidos de la API en estructuras tabulares (DataFrames) para su análisis. Ofrece herramientas avanzadas para limpieza, filtrado y agregación de datos. Su integración con otras bibliotecas facilita la creación de flujos de trabajo eficientes.

5. Matplotlib y Seaborn

Propósito: Bibliotecas para la generación de gráficos y visualizaciones.

Beneficios:

Matplotlib proporciona una amplia gama de opciones para crear gráficos personalizados. Seaborn simplifica la creación de gráficos estadísticos y mejora la estética de las visualizaciones. Ayudan a comunicar los resultados de manera clara y comprensible para los usuarios.

6. WordCloud

Propósito: Biblioteca utilizada para generar nubes de palabras a partir de texto.

Beneficios:

Permite visualizar las palabras más frecuentes en descripciones, etiquetas o comentarios de videos. Facilita la identificación de temas clave y tendencias en los datos textuales.

7. VaderSentiment

Propósito: Herramienta para el análisis de sentimientos en texto.

Beneficios:

Especialmente diseñada para analizar texto corto, como comentarios de redes sociales. Proporciona métricas de polaridad (positiva, negativa, neutral) y un puntaje compuesto que resume el sentimiento general. Fácil de implementar y eficiente en términos de rendimiento.

8. Google Auth OAuthlib

Propósito: Biblioteca para la autenticación segura mediante OAuth 2.0.

Beneficios:

Garantiza que las credenciales de acceso a la API de YouTube sean manejadas de manera segura. Permite a los usuarios autenticarse sin exponer información sensible. Cumple con los estándares de seguridad recomendados por Google.

9. Logging

Propósito: Módulo estándar de Python utilizado para registrar eventos y mensajes en el sistema.

Beneficios:

Facilita la depuración y el monitoreo del comportamiento del sistema.

Permite registrar errores, advertencias e información relevante durante la ejecución del proyecto. Mejora la trazabilidad y el mantenimiento del código.

10. Regular Expressions (re)

Propósito: Módulo estándar de Python para trabajar con patrones en texto.

Beneficios:

Permite extraer, limpiar y transformar datos textuales de manera eficiente.

Es utilizado para procesar descripciones, etiquetas y comentarios de videos.

Facilita la normalización de datos textuales para su análisis posterior.

11. CSV

Propósito: Módulo estándar de Python para leer y escribir archivos en formato CSV.

Beneficios:

Permite importar y exportar datos de manera sencilla. Facilita la interoperabilidad con otras herramientas y sistemas que utilizan este formato. Es útil para almacenar resultados intermedios o finales del análisis.

12. Datetime y Pytz

Propósito: Módulos estándar de Python para trabajar con fechas, horas y zonas horarias.

Beneficios:

Permiten manejar datos temporales de manera precisa, considerando zonas

horarias específicas. Son esenciales para analizar tendencias temporales en los datos, como la frecuencia de publicación de videos o la evolución de comentarios.

13. Logging

Propósito: Módulo estándar de Python para registrar eventos y mensajes en el sistema.

Beneficios:

Facilita la depuración y el monitoreo del comportamiento del sistema.

Permite registrar errores, advertencias e información relevante durante la ejecución del proyecto. Mejora la trazabilidad y el mantenimiento del código.

Conclusión

El uso de estas tecnologías y herramientas no solo garantiza la funcionalidad del proyecto, sino que también asegura su robustez, escalabilidad y facilidad de uso. Cada componente fue seleccionado cuidadosamente para abordar los desafíos específicos del análisis de datos de YouTube, maximizando la eficiencia y la calidad de los resultados obtenidos.

2.5. Proposiciones Teóricas Generales

Las proposiciones teóricas representan afirmaciones orientadoras que sintetizan posibles relaciones entre variables clave del objeto de estudio, y permiten enmarcar conceptualmente el análisis empírico dentro de un marco interpretativo. En este trabajo de titulación, centrado en la comparación de videos en YouTube con base en métricas, contenido y tendencias, dichas proposiciones no solo guían la investigación, sino que permiten trazar una lógica de interpretación entre los hallazgos obtenidos y su aplicabilidad práctica para la optimización de estrategias de contenido.

Desde una perspectiva científica, estas proposiciones facilitan la construcción de conocimiento replicable y aportan a la sistematización de fenómenos observables en el ámbito del análisis comparativo digital. Por ejemplo, al plantear que la

identificación de patrones de interacción entre videos permite inferir preferencias de audiencia, se reconoce la posibilidad de que el comportamiento agregado de los usuarios revele elementos estructurales del contenido que favorecen la fidelización o el interés. Esta idea es fundamental para interpretar por qué ciertos formatos o temáticas tienen mayor tracción, más allá de datos aislados. (F.O. Isinkaye, 2015)

Asimismo, se propone que el análisis comparativo de métricas puede optimizar las estrategias de contenido en plataformas de video, una afirmación que subyace al objetivo general del proyecto. Comparar videos bajo parámetros homogéneos permite obtener inferencias útiles para la toma de decisiones informadas, ya sea en la creación, segmentación o promoción de contenido, contribuyendo así al desarrollo de prácticas editoriales más eficientes y alineadas con el comportamiento de la audiencia.

Una tercera proposición sugiere que la sistematización del análisis mediante herramientas específicas mejora la precisión de las recomendaciones y facilita la escalabilidad del conocimiento generado. Esta premisa se conecta directamente con el desarrollo del aplicativo web como solución tecnológica que operativiza el modelo analítico planteado y permite democratizar el acceso a estos hallazgos para distintos tipos de usuarios.

Estas proposiciones no deben entenderse como verdades absolutas, sino como planteamientos teóricos derivados del análisis riguroso de la problemática, que orientan tanto el enfoque metodológico como la interpretación de resultados en el marco del trabajo de titulación.

2.5.1. Proposición 1: “La identificación de patrones de interacción entre videos permite inferir preferencias de audiencia.”

Referencia Clave: Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273. DOI:

<https://doi.org/10.1016/j.eij.2015.06.005>

• **Aporte:** Este artículo explica cómo los sistemas de recomendación (fundamentales en plataformas como YouTube) utilizan el comportamiento observado del usuario y sus interacciones (patrones de visualización, historial, etc.) para filtrar información y predecir sus preferencias, sugiriendo contenido relevante. Esto sustenta directamente la idea de que los patrones de interacción revelan preferencias.

Referencia Complementaria: Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences Education*, 15(4), es6. DOI: <https://doi.org/10.1187/cbe.16-03-0125>

• **Aporte:** Aunque enfocado en videos educativos, revisa cómo características específicas del video (brevedad, estilo conversacional) impactan el engagement del estudiante (una forma de interacción). Analizar qué características generan mayor engagement permite inferir las preferencias de la audiencia por ciertos estilos o formatos de contenido.

Referencia Adicional: Zhang, K., Cao, Q., Sun, F., Wu, Y., Tao, S., Shen, H., & Cheng, X. (2023). Robust Recommender System: A Survey and Future Directions.

• **Aporte:** Esta encuesta sobre sistemas de recomendación robustos, aunque técnico, se basa en la premisa fundamental de que estos sistemas funcionan analizando patrones de interacción para modelar y predecir preferencias de usuario, reforzando la proposición.

2.5.2. Proposición 2: “El análisis comparativo de métricas puede optimizar las estrategias de contenido en plataformas de video.”

Referencia Clave: Wan Saifudin, W. S. N. (2024). MPM Model: A Cross-Platform Massive Open Online Course (MOOC) Performance Monitoring and Measurement Model Based on MOOC Learning Analytics.

- **Aporte:** Esta tesis doctoral reciente desarrolla y evalúa un modelo (MPM) diseñado específicamente para monitorizar y medir el rendimiento (usando métricas de analítica de aprendizaje) en plataformas digitales (MOOCs). El objetivo explícito es utilizar este análisis comparativo de métricas para obtener insights que permitan justificar el rendimiento y considerar mejoras futuras, es decir, optimizar la estrategia.

Referencia Complementaria: Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. CBE—Life Sciences Education, 15(4), es6. DOI: <https://doi.org/10.1187/cbe.16-03-0125>

- **Aporte:** Al sintetizar principios basados en investigación sobre qué hace efectivo a un video (lo cual se mide a través de métricas de engagement y aprendizaje), implícitamente apoya la idea de que analizar comparativamente qué funciona mejor (métricas) permite optimizar la creación de contenido futuro (estrategia).

Referencia Adicional (Enfoque en Herramientas): Derks, C. C. (2019). *Trends Tool, a tool for abstract textual data analysis used for additive manufacturing trend.*

- **Aporte:** Describe el desarrollo de una herramienta para analizar datos textuales y descubrir tendencias (una forma de métrica derivada). El propósito de la herramienta es facilitar el análisis de datos para obtener insights (comparando tendencias, temas, etc.) que informen la toma de decisiones o estrategias.

3. Capítulo II. Materiales y Métodos

Este capítulo expone la estrategia metodológica adoptada para alcanzar los objetivos específicos de la investigación, detallando las actividades y herramientas empleadas para el desarrollo del sistema web de análisis comparativo de videos en YouTube. La metodología integra enfoques propios de la Ciencia de Datos aplicada a plataformas digitales, considerando la extracción en tiempo real, el

procesamiento y la interpretación de métricas relevantes mediante técnicas computacionales, se anexan los métodos teóricos y empíricos basados en la práctica de desarrollo de software y técnicas de análisis de datos que se han ido aprendiendo y empleando de manera práctica a lo largo del programa de maestría, se enumeran las técnicas e instrumentos, tecnologías y herramientas informáticas que se utilizaron para la construcción de la solución web y finalmente se detalla la tesis de propuesta de investigación final en conjunto con los métodos específicos y el análisis estadístico final como resultado general.

3.1 Enfoque General de la Investigación

3.1.1 Modalidad o Enfoque de la investigación

La investigación adopta un enfoque mixto, ya que integra metodologías cuantitativas y cualitativas para abordar los objetivos del estudio de manera comprensiva.

Componente Cuantitativo: Este componente se centra en la recolección, procesamiento y análisis de datos numéricos extraídos de la plataforma YouTube. Se analizan métricas objetivas como el número de visualizaciones, análisis de "me gusta" y "no me gusta", número de comentarios, duración de los videos, cantidad de videos, entre otras. Estos datos son fundamentales para realizar comparaciones entre canales y videos, identificar patrones estadísticos y evaluar el rendimiento del contenido. La aplicación desarrollada automatiza la extracción de estas métricas y las presenta de forma estructurada para su análisis.

Componente Cualitativo: El aspecto cualitativo se manifiesta en el análisis del contenido textual y semántico de los datos. Esto incluye el análisis de sentimientos de los comentarios de los videos para comprender la recepción de la audiencia, la identificación de temas y palabras clave en las descripciones y etiquetas de los videos, y la interpretación de las tendencias visualizadas. La generación de nubes de palabras a partir de etiquetas y comentarios es un ejemplo de técnica cualitativa que permite explorar la prominencia de ciertos temas. Este

enfoque busca profundizar en el "qué" y el "porqué" detrás de los números, ofreciendo insights sobre las estrategias de contenido y la percepción del público.

3.1.2 Tipo de Investigación

Dadas las características y objetivos del proyecto, la investigación se clasifica principalmente como:

Investigación Descriptiva: La investigación tiene como objetivo detallar de forma sistemática las características, métricas y tendencias pertinentes de los canales y videos de YouTube escogidos por los usuarios. Se describen las características de la herramienta que facilitan la visualización y síntesis de estas características, tales como la comparación de la cantidad de videos, la distribución de su duración, o las palabras más utilizadas en los comentarios. La meta es proporcionar una visión clara del estado y desempeño del contenido estudiado.

Investigación de Desarrollo Tecnológico: Un objetivo principal de este proyecto fue el desarrollo de una herramienta de software práctica: una aplicación web basada en Flask para el estudio de datos de YouTube. Este tipo de estudio se centra en diseñar, desarrollar y validar una solución tecnológica que pueda satisfacer una demanda particular, en este escenario, simplificar a los generadores de contenido, analistas de datos o empresas, la comparación de canales y el estudio de comentarios de forma eficaz e interactiva. La metodología abarca el proceso completo de desarrollo de software, desde la recopilación de requisitos (inherente a los propósitos del proyecto) hasta la ejecución y la propuesta de evaluación.

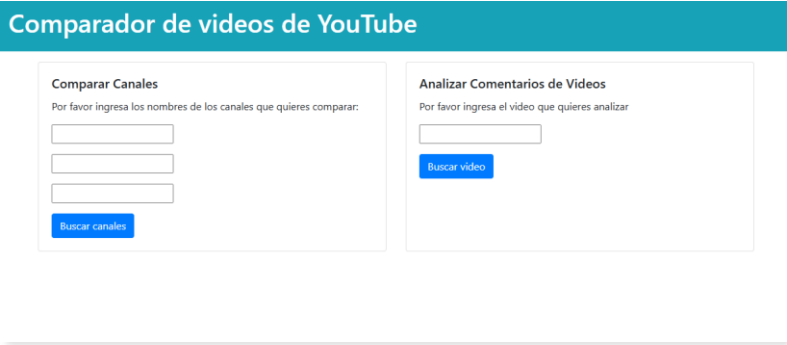
3.2 Metodología para el Cumplimiento de los Objetivos Específicos

La consecución de las metas planteadas en este proyecto de titulación se abordó mediante una serie de fases metodológicas interconectadas, cada una diseñada para satisfacer un objetivo específico. Este enfoque estructurado aseguró que el desarrollo de la herramienta de análisis de YouTube progresara de manera lógica, desde la concepción de su interfaz hasta la implementación de sus capacidades

analíticas y su eventual evaluación. A continuación, se desglosa el camino metodológico seguido para cada objetivo particular.

3.2.1 Desarrollo de la Interfaz de Usuario Intuitiva y Funcional

Es un componente clave el desarrollo de una interfaz de usuario intuitiva y funcional por el hecho que se necesita combinar aspectos clave de diseño como calidad visual, usabilidad y funcionalidad técnica. La interfaz va a constituir el punto de contacto entre el usuario y la solución web para las funcionalidades analíticas que se pretende del sistema. Se espera una experiencia fluida en la que el objetivo central sea la comparativa de canales y el análisis de videos específicos para poder extraer conclusiones a partir de las métricas clave. El uso del framework Bootstrap nos dará ciertos elementos de UI preconstruidos que ayudarán en la simplicidad y escalabilidad de la escritura del código a nivel de Frontend, además de que estandariza los estilos de la web haciéndolos reconocibles y fáciles de usar. A través del framework Flask, se estructuraron vistas web que permiten a los usuarios ingresar los inputs necesarios para realizar su búsqueda ya sea de canales de YouTube, así como el ingreso de enlaces directos a videos para el análisis individual del contenido multimedia.



The image shows a web interface titled "Comparador de videos de YouTube". It is divided into two main sections. The left section, "Comparar Canales", prompts the user to "Por favor ingresa los nombres de los canales que quieres comparar:" and provides three text input fields followed by a blue button labeled "Buscar canales". The right section, "Analizar Comentarios de Videos", prompts the user to "Por favor ingresa el video que quieres analizar" and provides a single text input field followed by a blue button labeled "Buscar video". The interface is clean and uses a light blue header and blue buttons.

Figura 1: Interfaz principal

Para la funcionalidad de Comparativa de Canales, se tienen hasta tres inputs para ingresar los nombres de los canales los cuales se quiere hacer la comparativa, posterior a hacer la búsqueda de estos canales, el usuario podrá seleccionar de una lista informativa los canales que resultaron de la búsqueda tal cual como YouTube lo mostraría, con el logo del canal y el nombre. Una vez seleccionados los canales a ser comparados, se podrá dar clic en el botón de comparación y después de un breve fetch y procesamiento de datos mediante la API de YouTube, se

desplegarán las gráficas analíticas necesarias para una eficiente comparativa de canales.

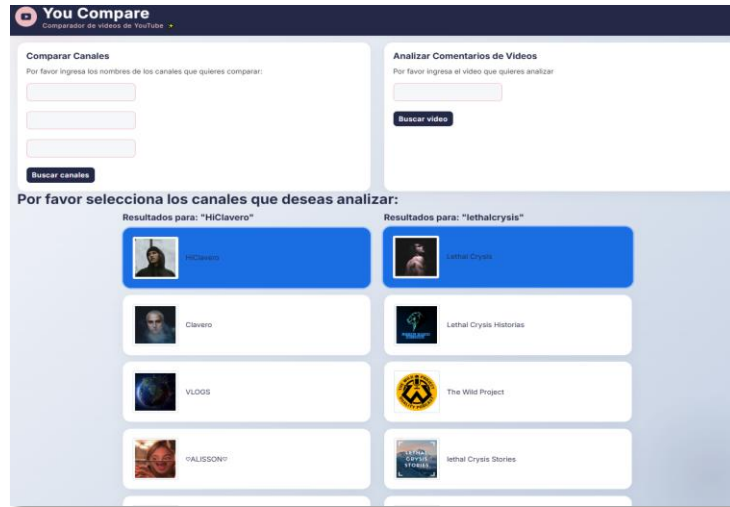


Figura 2: Búsqueda de canales

Mediante el análisis comparativo de métricas se optimizarán varios aspectos clave que un creador de contenido debe considerar a nivel general para crear estrategias determinantes para el éxito en plataformas de video y redes sociales en general. Se consideró importante el comparar videos bajo parámetros homogéneos para así permitir obtener inferencias útiles. Como se mencionó en el listado de herramientas utilizadas, se está utilizando la conocida librería de Python, matplotlib, para la generación de plots o gráficos intuitivos, útiles y descriptivos

Estos parámetros se muestran a continuación tomando el ejemplo de la figura anterior en el que se comparan dos canales dedicados a la creación de contenido de programación e ingeniería en software.

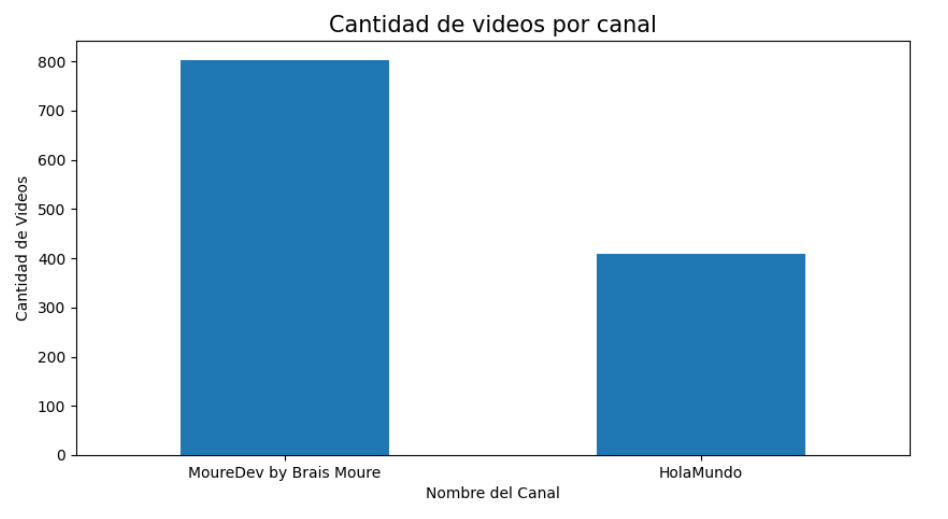


Figura 3: Gráfica de cantidad de videos por canal

Se tienen diferentes gráficas de distintos tipos como nubes de palabras, gráficos de barras, y rankings. La premisa de tener distintas gráficas a nivel de interfaz de usuario que logren explorar diferentes aspectos clave de comparación de canales de YouTube, nos permite conectar con la premisa de que el desarrollo del aplicativo web es una solución tecnológica que pone en marcha el modelo analítico planteado anteriormente y además justifica la finalidad del proyecto como una democratización del acceso a estos hallazgos para todo tipo de usuarios.

3.2.2 Implementación del Sistema de Extracción de Datos

El sistema de extracción de datos de este proyecto está diseñado para interactuar con la API de YouTube Data v3 de diversas maneras que disponibiliza la API, permitiendo obtener información detallada sobre canales, videos y comentarios. A continuación, se describe cómo se implementa esta funcionalidad, incluyendo ejemplos concretos de código. Para interactuar con la API, se utiliza una clave de desarrollador `API_KEY` que se pasa al cliente de la API, la función `youtubeApiKey` es la encargada de configurar el cliente de la API:

```
def youtubeApiKey(DEVELOPER_KEY, OAUTHLIB_INSECURE_TRANSPORT = "1", api_service_name = "youtube", api_version = "v3"):
    os.environ["OAUTHLIB_INSECURE_TRANSPORT"] = OAUTHLIB_INSECURE_TRANSPORT
    youtube = googleapiclient.discovery.build(...)
    return youtube
```

Figura 4: Snippet de código que interactúa con la API de YouTube

La función `youtubeSearchList` permite realizar búsquedas en YouTube, devolviendo una lista de videos o canales según los parámetros proporcionados. Por ejemplo, se puede buscar videos relacionados con una consulta específica tal y como se haría en el buscador propio de YouTube o también se puede obtener la cantidad de videos de un canal específico.

```
def youtubeSearchList(youtube, channel_id=None, q=None, maxResults=50, type=None):
    request = youtube.search().list(
        part="snippet"
        ,channelId=channel_id
        ,maxResults=maxResults
        ,q=q
        ,fields='items(id,snippet),nextPageToken',
        type=type
    )
    responseSearchList = request.execute()
    return responseSearchList
```

Figura 5: Snippet de código que funciona como buscador de videos y canales

Para la obtención de las estadísticas de videos, la función `youtubeSearchListStatistics` amplía la funcionalidad de búsqueda al incluir estadísticas avanzadas, como vistas, likes, dislikes y comentarios. Esto se logra combinando los resultados de búsqueda con las estadísticas de los videos obtenidas mediante la función `video_snippets`.

```
def youtubeSearchListStatistics(youtube, q=None, maxResults=10):
    query_result = youtubeSearchList(youtube, q=q, maxResults=maxResults, type='video')

    # Create a list of video ids
    video_id_list = [x['id']['videoId'] for x in query_result['items']]

    snippets = video_snippets(youtube, video_id_list, maxResults=10, part="statistics")
    assert len(query_result['items']) == len(snippets), 'Query result length does not match statistics length.'

    counter = 0
    for i in query_result['items']:
        i['statistics'] = snippets[counter]['statistics']
        counter += 1

    return query_result
```

Figura 6: Snippet de código que amplía la búsqueda de estadísticas

El siguiente paso es la obtención de Videos de un Canal, para este proceso se usa la función `videoIdList` con la que se permitirá obtener todos los videos públicos de un canal en específico. Este procedimiento se realiza accediendo a la lista de reproducción pública que el canal haya subido. Adicional se utiliza una función de

paginación que provee la API de YouTube para continuar navegando en grandes listas de videos, estos parámetros son configurables y se muestra en el siguiente código.

```
def videoIdList(youtube, channelId):
    videoIdList = []
    requestChannelsList = youtube.channels().list(
        part="contentDetails"
        ,id=channelId
        ,fields='items/contentDetails/relatedPlaylists/uploads'
    )
    responseChannelsList = requestChannelsList.execute()
    # Get upload playlist id from dictionary
    channelUploadPlaylistID = responseChannelsList.get('items')[0].get('contentDetails').get('relatedPlaylists').get('uploads')
    # Get items from upload playlist
    playlistNextPageToken = ''
    while playlistNextPageToken != None:
        requestPlaylistItems = youtube.playlistItems().list(
            part="snippet"
            ,maxResults=50
            ,pageToken=playlistNextPageToken
            ,playlistId=channelUploadPlaylistID
        )
        responsePlaylistItems = requestPlaylistItems.execute()
        for video in responsePlaylistItems['items']:
            videoIdList.append(video['snippet']['resourceId']['videoId'])
            playlistNextPageToken = responsePlaylistItems.get('nextPageToken')
    logger.info(f'The channel {channelId} has {len(videoIdList)} public videos')
    return videoIdList
```

Figura 7: Snippet de código que lista los videos

Se considera también la obtención de comentarios de un video, para este lineamiento se tiene la función `get_all_comments`, la cual extrae todos los comentarios de un video, incluyendo respuestas y metadatos. Esto se realiza mediante la API de `commentThreads`. Esta función devuelve un objeto JSON con todos los comentarios del video y su respectiva metadata, toma como input el objeto de credencial y también el id del video.

```
def get_all_comments(youtube, video_id):
    all_comments = get_comment_threads(youtube, video_id=video_id, part="id,replies,snippet")
    return all_comments
```

Figura 8: Snippet de código que obtiene los comentarios de un video

El sistema de extracción de datos implementado en este proyecto utiliza funciones modulares para interactuar con la API de YouTube. Como se ha visto, estas funciones permiten realizar búsquedas, obtener estadísticas, extraer comentarios y transformar los datos en estructuras fácilmente analizables.

3.2.3 Diseño y Aplicación de Algoritmos de Comparación y Análisis de Datos

Una vez que los datos fueron extraídos y recolectados mediante el uso de la API de YouTube (como se mencionó en el punto 3.2.2), el paso siguiente viene siendo uno de los más importantes. Hacer una inmersión más profunda en los datos, su contenido y estructura es un paso súper importante para extraer conocimiento valioso. Esta etapa no simplemente se trata de tener los datos, si no de prepararlos, interrogarlos y darles un sentido a través de varias técnicas de análisis de Ciencia de Datos. Como objetivo se tiene el transformar la información cruda en insights que puedan ser interpretados y visualizados de una manera mucho más efectiva a través de la aplicación web construida en Flask. Este proceso se ha logrado dividir en varias fases clave, comenzando con el acondicionamiento fundamental de los datos.

- Procesamiento y Transformación de Datos

Antes de poder realizar análisis profundos o hacer comparaciones, los datos que se obtuvieron de la API de YouTube necesitaban pasar por una etapa de preparación. Es como cuando un chef prepara los ingredientes antes de cocinar un plato gourmet: hay que seleccionarlos, limpiarlos y prepararlos bien para que el resultado final sea excelente. Esta primera fase es muy importante, porque la calidad y la forma en que organizamos los datos en ese momento influirán directamente en los análisis que hagamos después.

- *Extracción y Normalización de Datos*

Los datos que ofrece la API de YouTube son útiles, pero no siempre vienen listos para analizar. Fue necesario seleccionar solo lo relevante (como vistas, duración o comentarios) y luego reorganizar todo en un formato más claro y uniforme, es por esto que se hizo una normalización y limpieza de estos datos para usar solo lo que se necesita para el cumplimiento del objetivo del proyecto.

Técnica: Extracción de datos vía API REST y normalización a DataFrame.

Ejemplo de código:

```
# Inserta la data y la normaliza dentro de un dataframe
video_df = pd.DataFrame(video_data_dict)
```

Figura 9: Snippet que convierte los datos crudos de la API en un DataFrame estructurado.

○ Limpieza y conversión de Tipos

Tras extraer los datos, se realizó una limpieza general para corregir valores faltantes y eliminar caracteres innecesarios, asegurando que todo estuviera en el formato adecuado. Luego, se convirtieron los tipos de datos para facilitar el análisis, transformando cifras en números reales, garantizando así precisión y compatibilidad con herramientas estadísticas que se usarán posteriormente.

Técnica: Conversión de tipos y manejo de valores nulos.

Ejemplo de código:

```
# Asegurando que la columna 'like_count' sea numérica
comment_sentiment2['like_count'] = pd.to_numeric(comment_sentiment2['like_count'], errors='coerce')
# Borrando filas con valores NaN en 'like_count' y las columnas de sentimiento
comment_sentiment2 = comment_sentiment2.dropna(subset=['like_count'] + actual_sentiment_columns)
```

Figura 10: Snippet que verifica que los tipos de datos en una columna sean los correctos, hace una conversión y que borra filas con valores nulos.

• Análisis Comparativo de Canales

Una vez que se tengan los datos limpios, normalizados y estructurados de forma correcta en el procesamiento previo, se tiene el escenario listo para el acto principal: enfrentar a los canales de YouTube elegidos por el usuario en una comparativa directa de métricas clave. La idea principal de esta parte es ir más allá de métricas individuales del canal y darle sentido a una comparación realista de métricas de canales que tal vez se relacionen entre sí y ponerlos en perspectiva, esto permitirá al usuario de la aplicación notar las diferencias y similitudes que pueda tener el contenido de diferentes creadores. Para lograr esto se aplicaron varias técnicas que lograron resumir y contrastar la información efectivamente.

○ Agrupación y Agregación

Se organizaron los datos agrupando los videos según el canal al que pertenecen y luego se calcularon estadísticas clave para cada uno, como el total de visualizaciones, número de videos, promedios de duración, vistas, "me gusta" y comentarios. Esta agregación permitió resumir el rendimiento general de cada canal, facilitando una comparación más clara y equitativa entre ellos.

Técnica: GroupBy y agregación para comparar métricas entre canales.

Ejemplo de código:

```
# Agrupar los videos por canal y contar la cantidad de videos
df_all.groupby('channel_title').size().sort_values(ascending=False).plot.bar()
plt.xticks(rotation=ROT)
plt.xlabel("Nombre del Canal")
plt.ylabel("Cantidad de Videos")
plt.title('Cantidad de videos por canal', fontdict = {'fontsize' : 15})
plt.savefig(image_name, dpi=100)
```

Figura 11: Snippet de código que agrupa y muestra por cantidad de videos.

○ *Detección y Eliminación de Outliers*

Para evitar que unos pocos videos atípicos distorsionen los resultados, se consideraron los outliers al calcular métricas como la mediana, que es menos sensible a valores extremos. Aunque no se eliminaron estos datos, se ofrecieron formas de identificar su impacto y mitigar su efecto en las comparaciones entre canales, buscando representar el rendimiento de manera más justa y equilibrada.

Técnica: Cálculo de IQR para filtrar valores atípicos en duración de videos.

Ejemplo de código:

```
# Calcular el outlier (datos atipicos) y limpiar los datos
outlier = (df_all['duration_min'].describe()['75%'] - df_all
['duration_min'].describe()['25%']) * 1.5 + df_all['duration_min'].describe
()['75%']
df_all = df_all[df_all['duration_min'] <= outlier]
```

Figura 12: Snippet que calcula los Outliers y los limpia de la data final

• Análisis de Comentarios

Además de las métricas numéricas, se exploró el contenido textual de los videos, enfocándose en etiquetas y descripciones. Las etiquetas más utilizadas fueron extraídas y cuantificadas para identificar los temas clave de cada canal, lo que permitió visualizar su enfoque mediante nubes de palabras. En cuanto a las descripciones, se contabilizaron los enlaces externos como un indicador de la intención del canal de interactuar fuera de YouTube, lo que añade una dimensión estratégica al análisis comparativo entre creadores. Así como análisis de sentimientos a nivel de comentarios, cálculo de evolución de series de comentarios acumulados a través del tiempo, y correlaciones entre likes y sentimientos de comentarios.

○ *Análisis de Sentimientos*

Para conocer el impacto emocional de los videos, se aplicó un análisis automatizado de sentimientos usando la biblioteca VADER, diseñada para interpretar el lenguaje informal típico de redes sociales. Esta herramienta evaluó los comentarios recolectados, asignando una puntuación de polaridad que permitió identificar la actitud general del público hacia el contenido, proporcionando así una visión rápida sobre la recepción emocional de cada video.

Técnica: Análisis de sentimientos con VADER.

Ejemplo de código:

```

def analyze_comment_sentiments(comment_df):
    '''Analiza los comentarios y devuelve un dataframe con los resultados de la analisis
    analyzer = SentimentIntensityAnalyzer()

    sentiment = {
        'neg': [],
        'neu': [],
        'pos': [],
        'compound': []
    }

    for comment in comment_df['text_original']:
        vs = analyzer.polarity_scores(comment)
        for k, v in vs.items():
            sentiment[k].append(v)

    sentiment_df = pd.DataFrame(data=sentiment)
    comment_sentiment = pd.concat([comment_df.reset_index(), sentiment_df], axis=1)
    return comment_sentiment

```

Figura 13: Función que se encarga de iterar los comentarios de un video, analizarlos mediante métricas de sentimientos y construir un dataframe en base a los resultados usando Vader.

○Series Temporales y Acumulados

El análisis temporal de comentarios permite observar cómo evoluciona la interacción con un video. Agrupando los comentarios según su fecha de publicación y representándolos como series temporales, se identifican patrones como picos de actividad inicial o repuntes posteriores. Además, al visualizar el crecimiento acumulado de comentarios, se evalúa la duración del interés generado, revelando si el video mantiene una conversación sostenida o pierde relevancia rápidamente.

Técnica: Cálculo de suma acumulada para evolución de comentarios.

Ejemplo de código:

```

def split_sentiment_pos_neg(comment_sentiment):
    '''Divide los comentarios en positivos y negativos. Retorna un dataframe con todos
    los comentarios y dos dataframes con los comentarios positivos y negativos.'''

    comment_sentiment.sort_values(by='published_at', inplace=True)
    comment_sentiment['count'] = 1
    comment_sentiment['cumsum'] = comment_sentiment['count'].cumsum()

    # Seleccionar los comentarios positivos y negativos
    neg_sent = comment_sentiment[comment_sentiment['compound'] < -0.5]
    neg_sent['count'] = 1
    neg_sent['cumsum'] = neg_sent['count'].cumsum()
    pos_sent = comment_sentiment[comment_sentiment['compound'] > 0.5]
    pos_sent['count'] = 1
    pos_sent['cumsum'] = pos_sent['count'].cumsum()

    return comment_sentiment, pos_sent, neg_sent

```

Figura 14: Función encargada de dividir los comentarios que se obtuvieron previamente en negativos/positivos, e irlos ordenando en función de su crecimiento a través del tiempo.

○ Visualización de Correlaciones

Finalmente, para comprender la correlación entre la cantidad de likes y los sentimientos de los comentarios, se generó un algoritmo de cálculo para ubicar los patrones de likes versus comentarios positivos o negativos de un video, así podemos saber hacia qué tendencia los usuarios de la plataforma y suscriptores del canal, están enfocándose.

Técnica: Cálculo de correlación entre likes y sentimiento.

Ejemplo de código:

```
# Calcular la correlación entre 'like_count' y las columnas de sentimiento
try:
    like_count_sentiment_corr = round(comment_sentiment2.corr().loc['like_count'][actual_sentiment_columns[0]], 2)
```

Figura 15: Snippet de código que genera el plot en base a los valores obtenidos previamente de conteo de likes y sentimientos en comentarios

• Visualización de Datos

Una vez procesados y analizados los datos, el siguiente paso clave fue presentarlos de forma clara, atractiva y comprensible. La visualización de datos funcionó como un puente entre los resultados técnicos y la interpretación práctica, asegurando que cada gráfico respondiera a una pregunta concreta o transmitiera un mensaje útil. Para esto, se aplicaron diferentes técnicas visuales según el tipo de información: histogramas para mostrar cómo se distribuyen métricas como la duración o las vistas de los videos; gráficos de barras para comparar canales en términos de actividad, reacciones o sentimientos; gráficos de línea para observar cómo evolucionaban los comentarios a lo largo del tiempo; y nubes de palabras, ideales para destacar los términos más repetidos tanto en etiquetas como en los comentarios del público.

Estas visualizaciones se generaron principalmente con bibliotecas de Python como Matplotlib (la base para los gráficos), Seaborn (para comparativas y estética mejorada) y WordCloud (para los mapas de términos). Todo esto se organizó dentro del módulo viz.py, que producía gráficos integrados luego en la interfaz de usuario mediante Flask. Gracias a este enfoque, el usuario final pudo visualizar los resultados directamente en su navegador, con gráficos que hacían más intuitivo y eficiente el proceso de interpretación y comparación entre canales.

Técnicas: Histogramas, barplots, scatterplots, wordclouds. Uso de Matplotlib, Seaborn y WordCloud para visualización.

Ejemplo de código:

```
plt.hist(...)  
plt.bar(...)  
plt.scatter(...)  
plt.imshow(wordcloud, ...)
```

Figura 16: Snippet que lista las técnicas gráficas usadas

3.3 Técnicas e Instrumentos

En esta sección se detalla el proceso de integración del conjunto de técnicas computacionales y herramientas de desarrollo y análisis de datos para llevar a cabo todo el ciclo de trabajo antes mencionado; desde la recolección de los datos, su limpieza y análisis, hasta la presentación de resultados, insights y toma de eventuales decisiones a través de una aplicación web funcional. Se detallarán a continuación los instrumentos principales usados según la fase del proceso en la que fueron empleados:

3.3.1 Instrumento Principal

Como ya se había indicado anteriormente, la obtención de datos se realizó mediante el uso de la YouTube Data API v3, que permitió acceder de forma estructurada a información pública de canales, videos y comentarios. Esta API fue esencial para recuperar métricas como vistas, duración, likes, títulos, descripciones, etiquetas y textos de comentarios.

Fue muy importante el hecho de medir los límites que tenía la API de YouTube antes de considerarla como el instrumento principal para el correcto desarrollo de la tesis central de este proyecto. Como menciona Jason A. Clark en *“Developing a Digital Video Library with the YouTube Data API”*, se necesita tener una clara huella de las capacidades, beneficios y limitaciones de la API y asegurarnos como primera consideración que los requerimientos principales del proyecto puedan ser logrados usando la API. Al notar que la mayoría de características (obtención de videos, búsqueda, estadísticas) podían ser devueltas usando la herramienta, hizo caer en cuenta que sería el instrumento principal adecuado para este estudio. (Jason Clark, 2015)

3.3.2 Herramienta de Desarrollo

Python es el lenguaje de programación predilecto para el manejo, análisis y extracción de datos, lo que lo convierte en una excelente opción para hacer de intermediario con la YouTube Data API v3 (y con casi cualquier proveedor de grandes datos) debido a su gran simplicidad, versatilidad y amplio ecosistema de herramientas, librerías y frameworks. Aunque esto ya se ha expuesto en puntos anteriores, tenemos que se empleó la biblioteca oficial que provee Google para este fin, `google-api-python-client` es, en esencia el intérprete e intermediario construido específicamente para esto, es el mensajero oficial de Google encargado de que una aplicación de Python pueda interactuar de manera nativa con las APIs de Google, incluida obviamente la de YouTube. (Fox, 2021)

En este proyecto se emplearon principalmente técnicas de programación modular y orientada a funciones, aprovechando la flexibilidad y legibilidad que ofrece Python. Se organizaron los distintos procesos en funciones independientes, cada una con una responsabilidad clara, lo que facilita la reutilización y el mantenimiento del código. Por ejemplo, para la extracción de datos desde la API de YouTube, se encapsuló la lógica en funciones como `youtubeAPIkey`, `youtubeSearchList` y `videoIdList`, permitiendo separar la autenticación, la consulta y el procesamiento de resultados. Además, se utilizaron estructuras de datos como diccionarios y DataFrames de pandas para transformar y analizar la información de manera eficiente. Se aplicó manejo de errores mediante bloques `try-except`, así como el uso de logging para el monitoreo y depuración. La integración con Flask se realizó siguiendo el patrón de vistas y rutas, donde cada endpoint ejecuta funciones específicas y retorna los resultados a través de plantillas HTML.

3.3.3 Bibliotecas de Python

En el desarrollo de este proyecto se recurrió a diversas bibliotecas de Python, cada una seleccionada por su utilidad y eficiencia en tareas específicas del análisis de datos y la visualización. Por ejemplo, Flask fue la base para construir la aplicación web, ya que su estructura ligera y su facilidad de integración permitieron crear una interfaz intuitiva donde los usuarios pueden interactuar con los resultados sin complicaciones técnicas. Esta elección facilitó mucho la puesta en marcha del sistema y su despliegue en diferentes entornos.

Por otro lado, pandas fue fundamental para el manejo de los datos. Gracias a esta biblioteca, fue posible transformar la información obtenida de la API de YouTube en tablas ordenadas, lo que simplificó tanto la limpieza como el análisis posterior. La

capacidad de pandas para filtrar, agrupar y resumir datos resultó clave para comparar canales y analizar tendencias en los comentarios de los videos.

La obtención de los datos en sí se realizó mediante `googleapiclient`, que permitió conectarse directamente con la API de YouTube. Esto garantizó que la información utilizada estuviera siempre actualizada y proveniente de la fuente oficial, lo que aporta rigor y confianza a los resultados del análisis. Además, se emplearon bibliotecas como `Matplotlib` y `Seaborn` para la generación de gráficos. Estas herramientas hicieron posible visualizar de manera clara y atractiva los patrones encontrados, desde histogramas de duración de videos hasta gráficos de barras comparando la actividad de distintos canales.

Para el análisis de texto, especialmente en los comentarios, se utilizó `WordCloud`, que ayudó a identificar rápidamente los términos más frecuentes y relevantes. También se incorporó `VaderSentiment`, una herramienta especializada en análisis de sentimientos, que permitió evaluar el tono general de los comentarios y relacionarlo con otras métricas, como la cantidad de “me gusta”.

No menos importante fue el uso de logging, que sirvió para registrar eventos y posibles errores durante la ejecución del sistema, facilitando la depuración y el mantenimiento. Además, se recurrió a módulos estándar como `re` para el manejo de expresiones regulares, `csv` para la lectura y escritura de archivos, y `datetime` junto con `pytz` para gestionar fechas y zonas horarias, lo que resultó esencial al analizar tendencias temporales.

En conjunto, la integración de estas bibliotecas no solo agilizó el desarrollo, sino que también permitió abordar el análisis de datos de YouTube de manera robusta y flexible, asegurando que los resultados fueran tanto precisos como fácilmente interpretables para cualquier usuario interesado.

3.3.4 Procesamiento y Análisis de Datos

El procesamiento y análisis de datos se traduce en la obtención de resultados visuales y cuantitativos que permiten comprender el comportamiento y las características de los canales y videos de YouTube analizados. Se han generado gráficos y tablas que facilitan la interpretación de grandes volúmenes de información, aportando valor a la investigación.

Por ejemplo, mediante la extracción y transformación de datos, se construyen DataFrames que agrupan métricas clave como la cantidad de videos, vistas, duración, etiquetas y enlaces en las descripciones. Estos datos se visualizan en gráficos de barras y de histograma, lo que permite comparar de manera directa la actividad y el alcance de distintos canales. Así, se pueden identificar patrones de publicación, diferencias en la producción de contenido y tendencias de popularidad.

En el análisis de comentarios, se emplean técnicas de procesamiento de lenguaje natural para evaluar el sentimiento de los usuarios. Los resultados se presentan en nubes de palabras, que muestran los términos más frecuentes, y en gráficos de líneas acumuladas, que ilustran la evolución temporal de los comentarios y su polaridad (positiva o negativa). Además, se calcula la correlación entre el sentimiento de los comentarios y la cantidad de “me gusta”, lo que ayuda a entender la relación entre la percepción de la audiencia y la interacción con el contenido.

Estos procesos y visualizaciones no solo permiten comparar canales y videos de manera objetiva, sino que también facilitan la detección de factores que influyen en el éxito o el impacto de un canal. Para la investigación, esto significa contar con herramientas que transforman datos complejos en información clara y accionable, apoyando la toma de decisiones y la generación de conocimiento sobre dinámicas de consumo y producción en plataformas digitales.

3.3.5 Control de Versiones

En este proyecto se utilizó git como sistema de control de versiones para gestionar el desarrollo del código fuente. Git permitió llevar un registro ordenado de los cambios realizados, facilitando la colaboración, la recuperación de versiones anteriores y la resolución de conflictos. Además, se empleó GitHub como plataforma de alojamiento remoto, lo que hizo posible respaldar el proyecto en la nube, compartir avances y mantener un historial accesible y seguro de todo el desarrollo.

3.4 Métodos Específicos de la Especialidad

La realización de este proyecto se nutrió directamente de los conocimientos y competencias desarrollados a lo largo de la Maestría en Ciencia de Datos. A continuación, se detallan las áreas de conocimiento más relevantes y su aplicación específica en el análisis de datos de la plataforma YouTube.

3.4.1 Fundamentos de Gestión y Análisis Estadístico de Datos Digitales

El punto de partida de cualquier iniciativa en ciencia de datos radica en una sólida comprensión de los datos mismos y en cómo gestionarlos y analizarlos estadísticamente para extraer significado. Las materias como Ciencia de Datos, Estadística para la Ciencia de Datos, y Gestión de Bases de Datos, junto con los principios de la Estadística Aplicada, sentaron las bases metodológicas para este proyecto.

- **Conceptualización del Problema y Ciclo de Vida:** La propia definición del proyecto “crear un comparador de videos de YouTube para analizar contenido, estadísticas y tendencias” es un ejercicio de ciencia de datos aplicada. Se siguió un ciclo de vida que abarcó desde la identificación de una necesidad (facilitar el análisis de estrategias de contenido en YouTube), la recolección de datos (vía API), su procesamiento y transformación, el análisis exploratorio y descriptivo, la generación de visualizaciones y, finalmente, el desarrollo de una herramienta (la aplicación web) que entrega estos insights. Este enfoque estructurado es fundamental en la disciplina.

- **Gestión y Estructuración de Datos Digitales:** Aunque el proyecto no implementó una base de datos formal persistente para el almacenamiento a largo plazo de los datos de YouTube (dada su naturaleza dinámica y la extracción bajo demanda), los principios de la Gestión de Bases de Datos fueron cruciales. La información obtenida de la API de YouTube (en formato JSON) tuvo que ser cuidadosamente parseada y estructurada. Aquí es donde la biblioteca pandas y sus DataFrames brillaron, actuando esencialmente como tablas de una base de datos en memoria. Se definieron "esquemas" implícitos para estos DataFrames, asegurando que cada pieza de información (ID del video, título, vistas, duración, texto del comentario, etc.) tuviera su lugar y tipo de dato correcto. Esta organización fue vital para poder realizar consultas, agregaciones y cruces de información de manera eficiente, tal como se haría en un sistema gestor de bases de datos.

3.4.2 Aplicación de Técnicas de Minería de Datos en Contenido Digital

Si bien el proyecto no se centró en la creación de modelos predictivos complejos, sí se aplicaron conceptos y técnicas provenientes de la Minería de Datos y Modelización Predictiva y del Aprendizaje Automático para "minar" el contenido digital y extraer conocimiento valioso, especialmente del texto no estructurado. El objetivo era descubrir patrones y obtener insights a partir de los datos que no son evidentes a primera vista.

- **Análisis de Sentimientos como Clasificación:** El análisis de sentimientos realizado con VaderSentiment puede considerarse una forma de clasificación automática. Aunque Vader utiliza un enfoque basado en léxico y reglas en lugar de un modelo de machine learning entrenado desde cero para este proyecto, el principio es similar: asignar una categoría (positivo, negativo, neutral) a un fragmento de texto. Este proceso es fundamental en el aprendizaje automático y permite convertir datos textuales no estructurados en información estructurada y accionable sobre la percepción de la audiencia.

- **Extracción de Características y Palabras Clave:** La identificación de palabras más usadas en comentarios y etiquetas, después de un preprocesamiento (tokenización, eliminación de stop words), es una técnica básica pero fundamental en la minería de texto. Se trata de una forma de extracción de características del texto, donde las palabras frecuentes se convierten en "características" que describen el contenido o la discusión. Las nubes de palabras son una visualización directa de esta minería.

4. Capítulo III. Resultados y Discusión

4.1 Resultados

4.1.1 Resultado del Objetivo Específico 1: Implementación de una Interfaz de Usuario Intuitiva y Funcional

Como primer resultado tangible de la investigación, se materializó el desarrollo y la implementación de la aplicación web, cuyo punto de entrada es una interfaz de usuario limpia, funcional y diseñada con un enfoque en la simplicidad y la usabilidad. El cumplimiento de este objetivo no fue un mero paso técnico, sino la creación del vehículo a través del cual todos los análisis posteriores son entregados al usuario final.

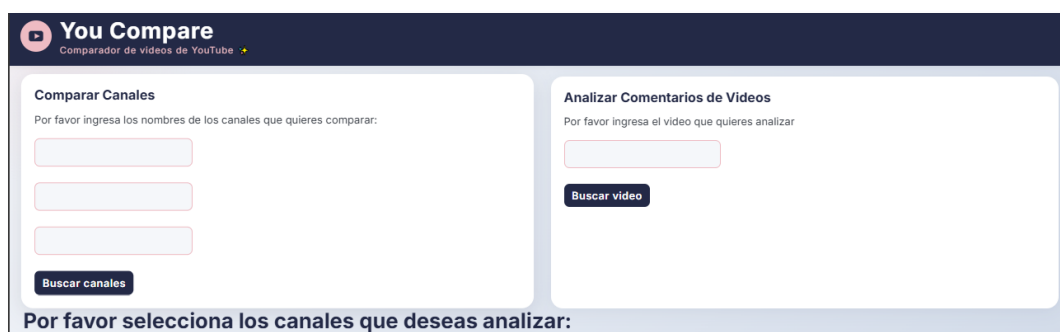


Figura 17: Interfaz principal para elegir entre comparación de Canales o Análisis de Comentarios

El resultado principal es una interfaz web, como se puede apreciar en la **Figura 17**, que presenta al usuario dos rutas de acción claras y diferenciadas desde el primer momento: "Comparar Canales" y "Analizar Comentarios de Videos". Esta separación intencionada evita la sobrecarga de información y guía al usuario de forma natural hacia la funcionalidad que desea utilizar. La idea central fue construir una experiencia fluida, donde la tecnología no fuera una barrera, sino un facilitador. Para esto, se empleó el framework Bootstrap, lo que resultó en el uso de componentes de diseño estandarizados y reconocibles que simplifican la interacción

4.1.2 Resultado del Objetivo Específico 2: Desarrollo del Sistema de Extracción de Datos

Una vez establecida la interfaz, el siguiente resultado fundamental fue la consolidación de un sistema de extracción de datos robusto y funcional. La consecución de este objetivo se demuestra en la capacidad del sistema para actuar como un intermediario eficaz entre las

solicitudes del usuario en la aplicación web y la vasta información disponible a través de la YouTube Data API v3. El sistema no solo recupera datos, sino que los obtiene de manera contextual y estructurada, preparándolos para el análisis. Este proceso se puede observar en dos flujos de trabajo principales que corresponden a las funcionalidades de la herramienta.

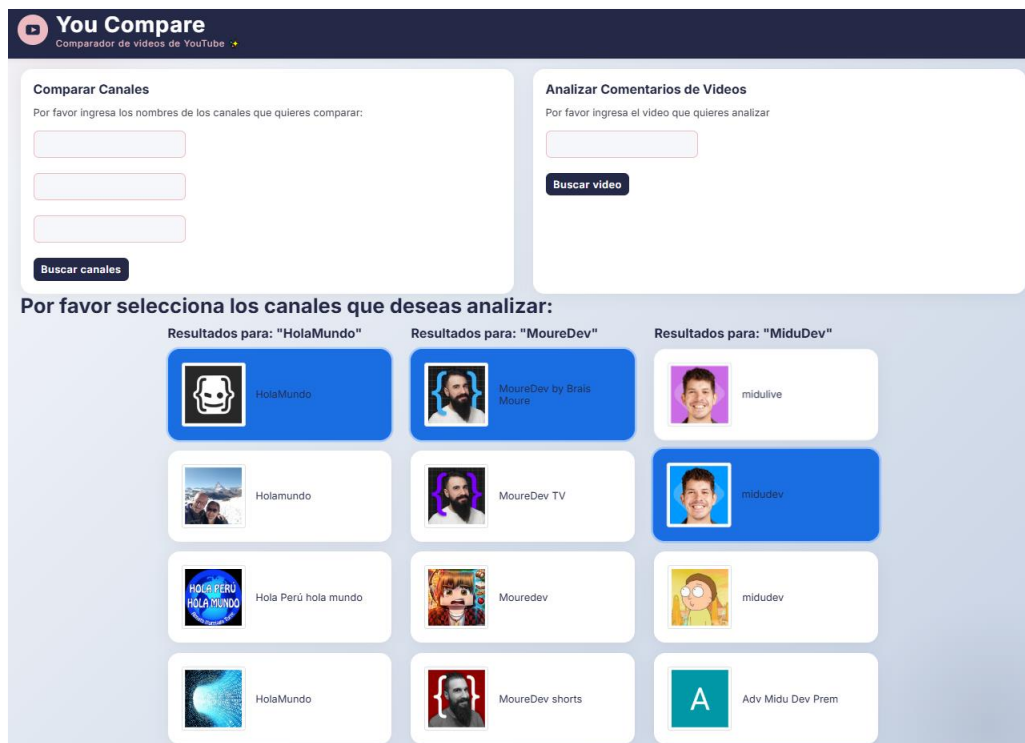


Figura 18: Interfaz de análisis de 3 canales

El primer flujo es el de extracción para el análisis comparativo de canales. Este proceso se inicia cuando un usuario ingresa los nombres de dos o más canales en la interfaz. El sistema interpreta estas entradas y ejecuta una búsqueda a través de la función youtubeSearchList. Como resultado, la aplicación no procede a ciegas, sino que presenta al usuario una lista de los canales encontrados que coinciden con la búsqueda, mostrando su nombre y logo oficial para una correcta identificación, tal como se ilustra en la **Figura 18**. Una vez el usuario selecciona los canales definitivos, el sistema demostró su capacidad para orquestar una secuencia de llamados a la API más profunda. Primero, obtiene un listado completo de todos los videos públicos de cada canal seleccionado, manejando eficientemente la paginación para recopilar catálogos extensos. Posteriormente, para cada video identificado, extrae sus metadatos y estadísticas clave como vistas, "me gusta" y duración. El resultado de esta fase es un conjunto de datos

consolidado y normalizado en un DataFrame de Pandas, listo para ser procesado por los algoritmos de comparación.

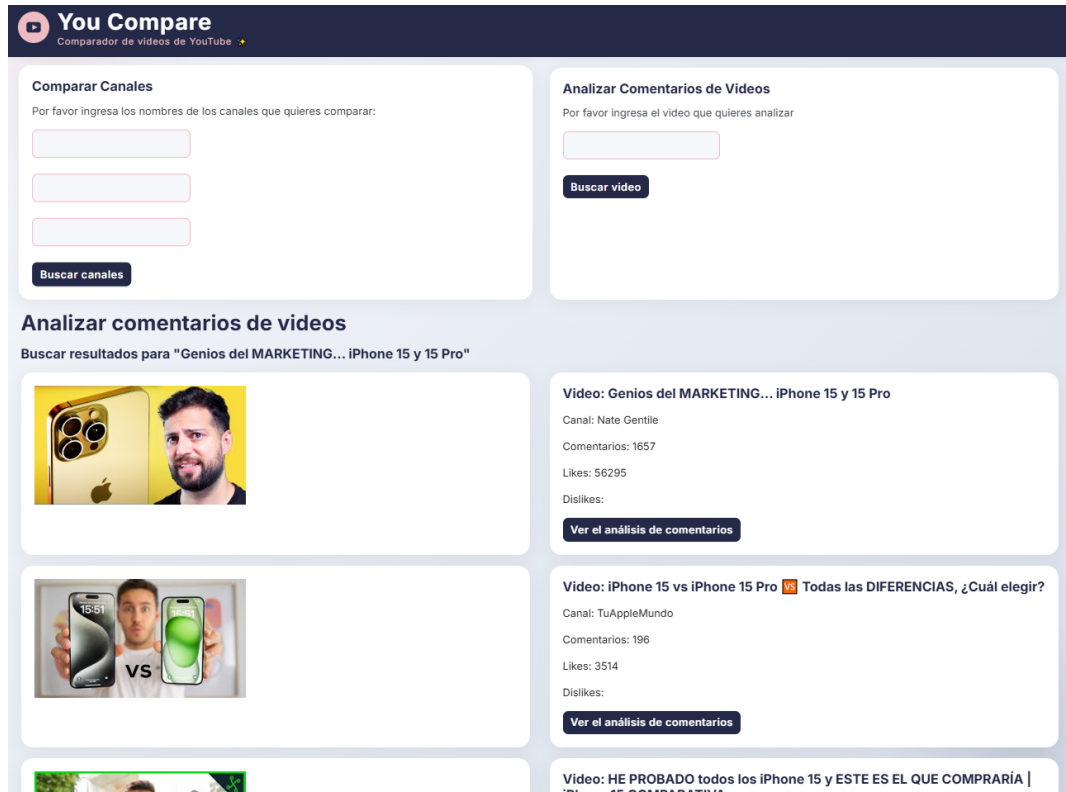


Figura 19: Interfaz de análisis de comentarios

El segundo flujo de trabajo corresponde a la extracción de datos para el análisis de sentimientos, una tarea más focalizada. En este caso, el usuario ingresa la URL de un video específico que desea analizar. El sistema primero procesa esta entrada para aislar el identificador único del video (`video_id`). A continuación, ejecuta la función `get_all_comments`, diseñada específicamente para interactuar con el endpoint `commentThreads` de la API. El resultado es una extracción exhaustiva de todos los comentarios asociados a ese video, incluyendo sus respuestas y metadatos relevantes como la fecha de publicación y la cantidad de "me gusta" por comentario como se muestra en la **Figura 19**. Esta data, que llega en formato JSON, es inmediatamente transformada en una estructura tabular y limpia. El producto final de este flujo es un conjunto de datos ordenado que sirve como la materia prima indispensable para el análisis de texto y la aplicación de los modelos de sentimiento descritos en la metodología.

4.1.3 Resultado del Objetivo Específico 3: Aplicación de Algoritmos de Comparación y Análisis

El cumplimiento del tercer objetivo se materializa en la aplicación exitosa de los algoritmos de procesamiento sobre los datos ya extraídos. El resultado no es un único dato, sino un mosaico de visualizaciones y resúmenes cuantitativos que, en conjunto, ofrecen una narrativa completa del rendimiento y la estrategia de los contenidos analizados. A continuación, se detallan los productos generados por la herramienta, utilizando nuevamente el caso práctico de los canales "MoureDev by Brais Moure" y "HolaMundo" para contextualizar los hallazgos.

Resultados del Análisis Comparativo de Canales

El sistema aplicó un conjunto de algoritmos para contrastar los canales a nivel macro, resultando en los siguientes productos analíticos:

- **Comparativa de Volumen de Producción:** El primer análisis arroja un gráfico de barras que compara la cantidad total de videos por canal (barplot_channel_video_count). Como se observa en la **Figura 20** del documento, este resultado ofrece una medida directa y fundamental del volumen de producción de cada creador.

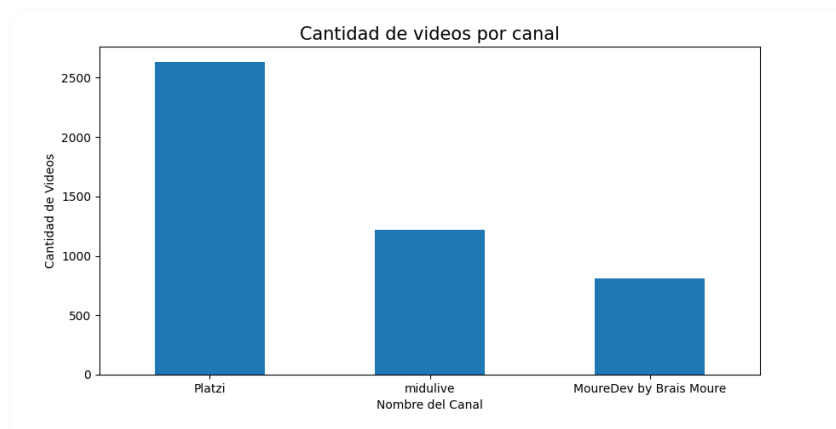


Figura 20: Plot de cantidad de videos por canal

- **Análisis de Formato de Contenido:** Se generó un histograma para cada canal que grafica la distribución de la duración de sus videos (histogram_video_duration_count_single). El resultado, presentado en la **Figura 21**, permite inferir la estrategia de formato de cada uno; si se inclinan por videos cortos y directos o por tutoriales extensos y detallados.

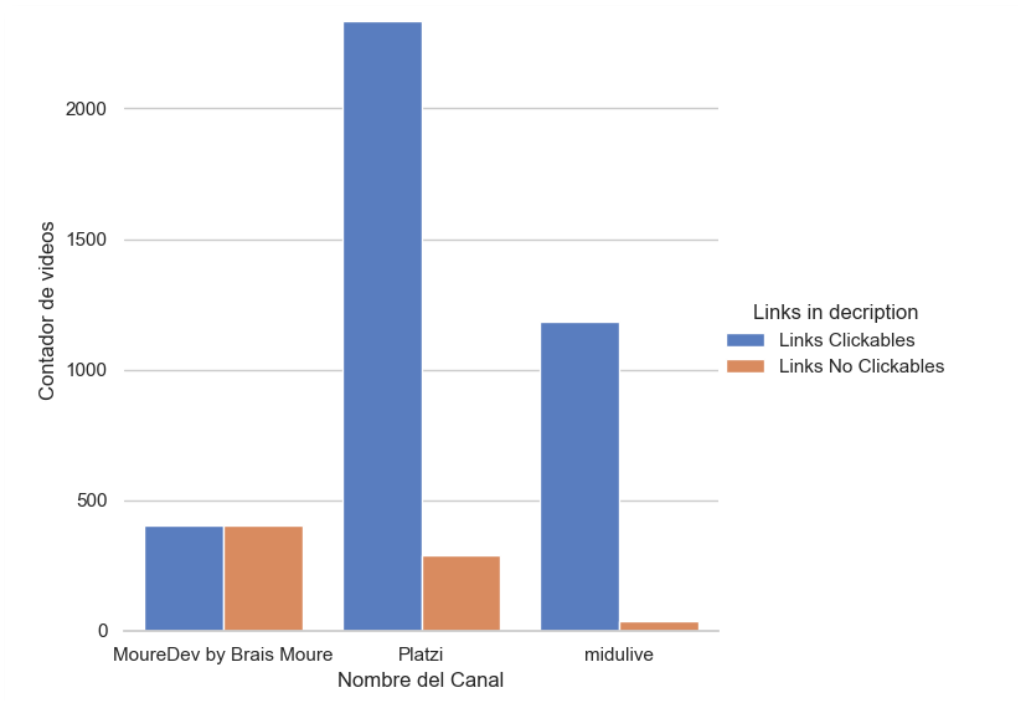


Figura 23:Plot de barras que indica la cantidad de enlaces externos por canal

• **Identificación de Contenido de Mayor Impacto:** Finalmente, el sistema procesa las métricas de todos los videos para generar una tabla (top_videos). Esta **Figura 24**, renderizada en la interfaz, presenta un ranking de los videos con mayor número de vistas de cada canal. Este resultado es de alto valor estratégico, ya que identifica de forma inequívoca las piezas de contenido que han logrado un mayor alcance.

Top 5 Videos más vistos por Canal

Rank	Channel Title	Video Title	View Count
1	Platzi	El mundo está en pausa, tú no Desde casa	5381344
2	Platzi	Hoy estamos a 79 de marzo del 2020... 🤔	5277729
3	Platzi	Técnicas de hackeo de las que serás víctima	4579669
4	Platzi	Cuánto tiempo más seguiremos en cuarentena por el coronavirus	3102469
5	Platzi	Cómo prepararte para una recesión económica Platzi Finanzas	3081422
6	MoureDev by Brais Moure	KOTLIN: Curso ANDROID desde CERO para PRINCIPIANTES	3039854
7	MoureDev by Brais Moure	ANDROID STUDIO: COMO Crear una APP (para Principiantes) 📱 [Tutorial]	2829850
8	MoureDev by Brais Moure	Curso COMPLETO de PYTHON desde CERO para PRINCIPIANTES	2618002
9	midulive	CURSO REACT 2024 - Aprende desde cero	1147886
10	MoureDev by Brais Moure	Curso COMPLETO de SQL y BASES DE DATOS Desde Cero para PRINCIPIANTES	1133708
11	MoureDev by Brais Moure	Curso COMPLETO de GIT y GITHUB desde CERO para PRINCIPIANTES	883443
12	midulive	¡DEACTIVA esto de CHROME YA! 😡	680787
13	midulive	Curso de AWS Desde Cero Amazon Web Services 🖥️	592946
14	midulive	CURSO DE NODE.JS DESDE CERO: Introducción y primeros pasos	564429

Figura 24: Gráfico de posiciones del top canal/video con mayor cantidad de vistas

generó un estallido de comentarios inicial que luego se desvaneció, o si por el contrario ha mantenido un goteo constante de nueva interacción. Para añadir una capa de profundidad a este análisis, la **Figura 27** presenta un segundo gráfico de línea (lineplot_cumsum_video_comments_pos_neg) que desglosa la suma acumulada por sentimiento, trazando por separado la evolución de los comentarios positivos y negativos. El resultado de este gráfico es particularmente revelador, pues permite observar la narrativa emocional de la recepción a lo largo del tiempo.



Figura 26: Gráfico de línea para identificar la evolución de cantidad de comentarios en el tiempo

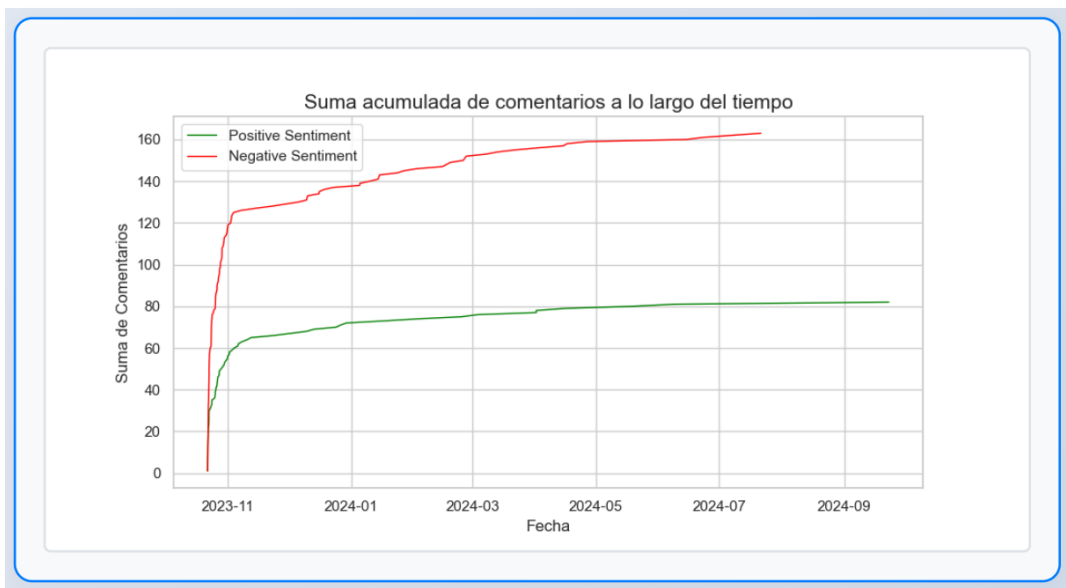


Figura 27: Gráfico de línea que demuestra la cantidad de comentarios negativos y positivos a través del tiempo

• **Estudio de la Relación entre Sentimiento y Validación Social:** Finalmente, el análisis más granular se enfoca en la cultura de la propia comunidad de comentaristas. Para esto, se genera un gráfico de dispersión (scatterplot) que correlaciona el puntaje de sentimiento de cada comentario individual con la cantidad de "me gusta" que ha recibido (scatterplot_sentiment_likecount), tal como se muestra en la **Figura 28**. Este resultado es fundamental para entender qué tipo de aportaciones valora la comunidad. Permite responder a preguntas como: ¿Son los comentarios más positivos los que reciben mayor validación (likes)?, ¿o existe un aprecio por las críticas constructivas (comentarios con sentimiento negativo pero alto número de likes)? La distribución de los puntos en este gráfico ofrece, por tanto, un insight sobre la madurez y las dinámicas internas de la audiencia del canal.

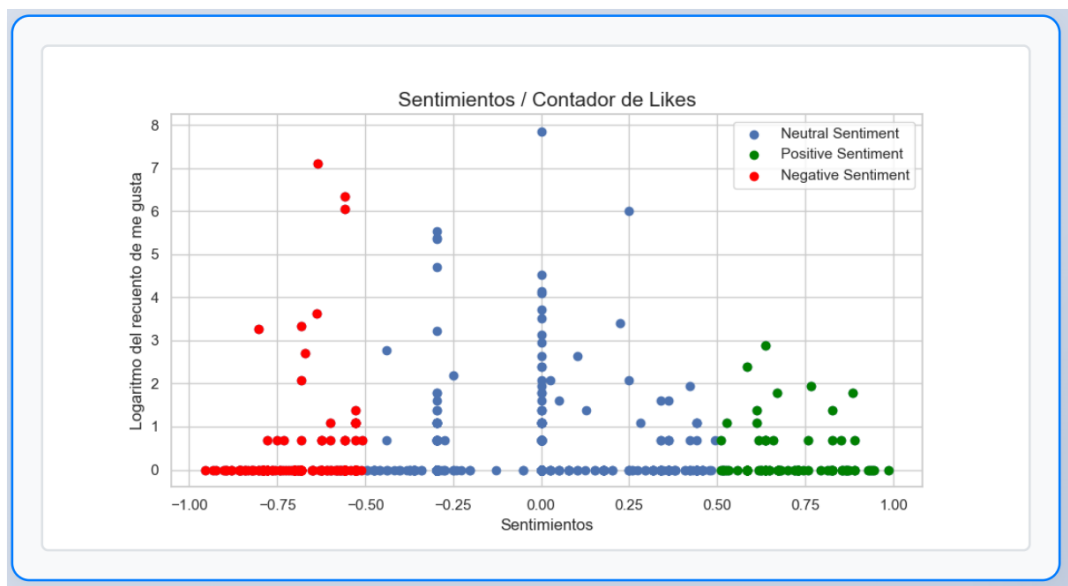


Figura 28: Gráfico de puntos que demuestra la cantidad de comentarios positivos o negativos en función de likes o dislikes

5. Conclusiones y Recomendaciones

A lo largo del desarrollo de este proyecto, se logró construir una herramienta robusta y flexible para el análisis de datos provenientes de YouTube, integrando diversas técnicas de ciencia de datos y visualización. El sistema permite comparar canales, analizar comentarios y extraer información relevante de manera automatizada, lo que representa un avance significativo respecto a los métodos manuales o poco sistematizados que suelen emplearse en el análisis de contenido digital.

Una de las principales conclusiones que se puede extraer es que la integración de diferentes bibliotecas de Python, junto con el uso de la API de YouTube, facilita enormemente la recolección y el procesamiento de grandes volúmenes de datos. Esto no solo agiliza el trabajo del investigador, sino que también reduce la posibilidad de errores humanos y asegura la reproducibilidad de los análisis. El uso de pandas, por ejemplo, permitió transformar datos crudos en estructuras fácilmente manipulables, mientras que herramientas como Matplotlib, Seaborn y WordCloud hicieron posible visualizar patrones y tendencias que de otra manera pasarían desapercibidos.

En cuanto al análisis de comentarios, la aplicación de técnicas de procesamiento de lenguaje natural, como el análisis de sentimientos, aportó una dimensión cualitativa que enriquece el estudio cuantitativo tradicional. Poder identificar el tono general de la audiencia, así como correlacionar este sentimiento con métricas de interacción como los “me gusta”, abre la puerta a investigaciones más profundas sobre el impacto real del contenido y la percepción de los usuarios.

Otro aspecto relevante es la importancia de la visualización de datos. Los gráficos generados no solo sirven para ilustrar resultados, sino que también ayudan a descubrir relaciones y comportamientos que no son evidentes a simple vista. Por ejemplo, los histogramas de duración de videos y los barplots de actividad por canal permiten identificar rápidamente diferencias en las estrategias de publicación, mientras que las nubes de palabras y los gráficos de sentimiento ofrecen una visión clara de los temas y emociones predominantes en la comunidad.

Sin embargo, el desarrollo del proyecto también puso de manifiesto algunos retos y limitaciones. Uno de ellos es la dependencia de la API de YouTube, que impone restricciones en cuanto a la cantidad de datos que se pueden extraer en un periodo

determinado. Esto obliga a diseñar estrategias eficientes de consulta y almacenamiento, así como a considerar posibles alternativas o complementos en caso de que se requiera un análisis a mayor escala. Además, el análisis de sentimientos, aunque útil, presenta limitaciones inherentes a los modelos utilizados, especialmente cuando se trata de comentarios en diferentes idiomas o con expresiones coloquiales y sarcasmo.

Desde el punto de vista técnico, la modularidad del código y la separación de responsabilidades entre los distintos componentes del sistema facilitaron tanto el desarrollo como el mantenimiento. El uso de control de versiones con Git y GitHub resultó fundamental para llevar un registro ordenado de los avances y para garantizar la integridad del trabajo, especialmente en un entorno colaborativo o cuando se requiere volver a versiones anteriores del proyecto.

En términos de impacto, la herramienta desarrollada tiene un gran potencial para ser utilizada no solo en el ámbito académico, sino también por creadores de contenido, agencias de marketing digital y cualquier persona interesada en comprender mejor el funcionamiento y la dinámica de YouTube. La posibilidad de comparar canales, analizar tendencias y explorar la interacción de la audiencia de manera visual y automatizada representa un valor añadido que puede traducirse en mejores estrategias de contenido y una mayor comprensión del entorno digital.

6. Referencias Bibliográficas

Trabajos citados

- @Iconosquare. (Abril de 2025). *iconosquare*. Obtenido de Guía completa de estadísticas de YouTube: <https://www.iconosquare.com/es/blog/guia-estadisticas-youtube>
- Ahmed Al Kuwaiti, K. N.-R. (2023). A Review of the Role of Artificial Intelligence in Healthcare. *National Library of Medicine*.
- Cena, M. D. (26 de Junio de 2024). *Vidpros*. Obtenido de How ViewStats Pro can help you grow your YouTube channel?: <https://vidpros.com/viewstats-pro-review-a-growth-hack-for-your-youtube-channel/>
- Derks, C. (2020). *Trends Tool, a tool for abstract textual data analysis used for additive manufacturing trend*. Enschede, The Netherlands: University of Twente.
- Eliganti Ramalakshmi, A. B. (2022). YouTube Data Analysis and Prediction of Views . *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* , 574.
- F.O. Isinkaye, Y. F. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 273.
- Fox, E. A. (2021). YouTube Video Analysis. *Virginia Tech*, 35.
- JARMAN, D. S. (2022). *Social network analysis and festival relationships: personal, organisational and strategic connections*. Edinburgh.
- Jason Clark. (2015). *code4lib*.
- JOHANES FERNANDES ANDRY, H. T. (2021). BIG DATA ANALYSIS ON YOUTUBE WITH TABLEAU. *Journal of Theoretical and Applied Information Technology*, 10.
- Junling Zhang, H. L. (2022). *The Impact of Big Data Management Capabilities on the Performance of Manufacturing Firms in Asian Economy During COVID-19: The Mediating Role of Organizational Agility and Moderating Role of Information Technology Capability*. Shanghai, China: Faculty of Economics and Management, East China Normal University.
- KOÇ, B. (2023). *The Role of User Interactions in Social Media on Recommendation Algorithms: Evaluation of TikTok's Personalization Practices From User's Perspective*. İSTANBUL: İSTANBUL UNIVERSITY INSTITUTE OF SOCIAL SCIENCES DEPARTMENT OF PUBLIC RELATIONS AND PUBLICITY.
- maekersuite. (s.f.). *Top 10 Tools for YouTube Analytics: A Comparative Review*. Obtenido de <https://maekersuite.com/blog/youtube-analytics-tools>
- Maryland, U. o. (2025). Obtenido de <https://guides.hshsl.umaryland.edu/finding-data/health-data-types>
- Newberry, C. (06 de Febrero de 2025). *Hootsuite*. Obtenido de YouTube analytics: Tools and tips for marketers: <https://blog.hootsuite.com/youtube-analytics/>
- Othman, A. (2022). Advances in Digital Marketing in the Era of AI. *Research Gate*, 314.
- Samia Idbenssi, L. S. (2023). Exploring the Relationship between Social Media and Tourist Experiences: A Bibliometric Overview. *MDPI Open Access Journals*.
- T. Aleksandrova, I. B. (2019). *CORPORATE LEARNING FOR THE DIGITAL WORLD*. Moscow: y Valery Katkalo, Martin Moehrle, Dmitry Volkov.
- upwork. (2025). *upwork*. Obtenido de <https://www.upwork.com/hire/youtube-monetization-experts/>
- Yadin, N. (24 de Julio de 2024). *thoughtleaders*. Obtenido de Best YouTube analytics tools: <https://www.thoughtleaders.io/blog/best-youtube-analytics-tools>
- Zote, J. (19 de Abril de 2024). *SproutSocial*. Obtenido de 11 YouTube analytics tools to improve your marketing in 2025: <https://sproutsocial.com/insights/youtube-analytics-tools/>