

Proyecto Final Grupo 5

Social Pulse: Análisis de Interacciones y Tendencias para Marcas en Redes Sociales

Integrantes

Introducción

Objetivos

Objetivo general

Objetivos específicos

Alcance

¿Qué se incluye en el proyecto?

¿Qué no se incluye en el proyecto?

Metas y Resultados

Planificación del Proyecto

Cronograma

Recursos

1. Recursos Humanos

2. Recursos Tecnológicos

3. Infraestructura Adicional

Presupuesto

Estimación de Costos

Asignación de Recursos Financieros

Estructura de Costos

Identificación de Riesgos

Descripción de las fases del proyecto

Fase 1: Investigación y definición de requerimientos

Fase 2: Extracción de datos:

Fase 3: Análisis de Datos Descriptivo

Fase 4: Desarrollo del Componente Predictivo

Fase 5: Creación del Dashboard en Streamlit

Fase 6: Pruebas y Validación

Fase 7: Implementación y Documentación

Fase 8: Actividades de Soporte y Seguimiento

Ejecución del Proyecto

Entregables

Carpeta archivos

Dashboard

Conclusiones

Recomendaciones

Social Pulse: Análisis de Interacciones y Tendencias para Marcas en Redes Sociales



Integrantes

- Andrea Ximena Castaño
 - Angie Arango Zapata
 - Esneider Álvarez Gómez
 - Sebastian Loaiza
 - Alvaro Lozano Rodríguez
-

Diciembre 11 de 2024

Introducción

En un mundo digital en constante evolución, donde las redes sociales son clave para conectar marcas y consumidores, contar con una estrategia de marketing sólida y basada en datos es esencial para cualquier empresa que busque destacar y fidelizar a su audiencia. Sin embargo, muchas marcas enfrentan el reto de interpretar la gran cantidad de datos generados por sus interacciones en redes sociales y traducirlos en acciones estratégicas y efectivas.

Este proyecto propone un análisis descriptivo y predictivo del comportamiento de las audiencias y competidores en plataformas sociales, especialmente orientado a optimizar la presencia de las marcas en canales como Instagram. Se busca transformar datos complejos en insights claros y accionables mediante un dashboard interactivo en Looker. Esta herramienta permitirá a las marcas tomar decisiones informadas y optimizar su estrategia de marketing digital para alcanzar el máximo impacto y engagement en sus audiencias, facilitando un crecimiento orgánico y sostenible en su comunidad online.

Objetivos

Objetivo general

Diseñar e implementar una herramienta digital interactiva que facilite la visualización y análisis de las interacciones y tendencias en el comportamiento de una marca en redes sociales, con énfasis en Instagram.

Objetivos específicos

- Analizar el comportamiento y patrones de interacción de una marca en Instagram.
 - Identificar los tipos de contenido que generan mayor interacción entre los seguidores de una marca.
 - Predecir tendencias de interacción que promuevan el crecimiento orgánico y sostenible de la comunidad digital.
 - Desarrollar un sistema de visualización de datos y métricas clave de marketing digital que facilite la toma de decisiones estratégicas.
-

Alcance

¿Qué se incluye en el proyecto?

El proyecto abarca un conjunto integral de actividades y entregables diseñados para analizar y optimizar la presencia de las marcas en redes sociales, con un enfoque particular en Instagram. Las principales áreas incluidas son las siguientes:

▼ Análisis del comportamiento de la marca en redes sociales

- Evaluación de indicadores clave como el número de seguidores, cantidad de "likes" y comentarios en un periodo de tiempo determinado.
- Monitoreo de tendencias de interacción que permitan identificar oportunidades para fomentar el crecimiento orgánico de la marca.

▼ Análisis descriptivo y predictivo

- Procesamiento y análisis de datos históricos de interacción en redes sociales para comprender patrones actuales.
- Desarrollo de modelos predictivos para anticipar tendencias de comportamiento y proponer estrategias proactivas.

▼ **Identificación de patrones de comportamiento:**

- Estudio de las dinámicas de interacción de la audiencia con el contenido de la marca.

▼ **Desarrollo de un dashboard interactivo**

- Creación de una interfaz visual intuitiva que facilite la exploración y comprensión de datos relevantes.
- Integración de filtros, gráficos y métricas clave para personalizar el análisis según las necesidades específicas de las marcas.

▼ **Generación de recomendaciones estratégicas basadas en datos**

- Elaboración de reportes detallados con insights claros y accionables.
- Propuestas de mejora en las estrategias de contenido, horarios de publicación y segmentación de audiencias.

¿Qué no se incluye en el proyecto?

Análisis en tiempo real de los datos (limitaciones en cuanto los costos de accesos a los datos, y la privacidad de los mismos para un prototipado).

Analizar la audiencia en cuanto a segmentación del mercado. (Limitaciones de acceso a los datos).

- Análisis de la competencia y qué estrategias utilizan.
- Extracción automática de datos desde las plataformas (no se incluye desarrollo de herramientas de web scraping o API, salvo si ya están disponibles).
- Implementación de campañas publicitarias directamente en redes sociales.
- Análisis profundo de plataformas sociales no mencionadas en el alcance (ejemplo: TikTok, LinkedIn).
- Desarrollo de contenido para redes sociales (gráficos, videos, copywriting).
- Integración de sistemas externos adicionales que no sean compatibles con Looker.
- Soporte técnico continuo o mantenimiento del dashboard después de la implementación inicial.

Metas y Resultados

Metas específicas, medibles, alcanzables, relevantes y con un tiempo definido (SMART).

Específicas:

- Diseñar un dashboard que visualice métricas clave como engagement, alcance, y crecimiento de seguidores en Instagram.
- Analizar datos de al menos "6 meses" de interacciones sociales de la marca.
- Generar un informe análisis accionables sobre la audiencia y el mercado.

Medibles:

- Lograr que el dashboard entregue resultados en un tiempo de carga inferior a 5 segundos para las consultas principales.
- Aumentar el porcentaje de decisiones estratégicas basadas en datos en un 30% dentro de los tres meses posteriores a la implementación.
- Obtener una precisión del 85% en las predicciones de tendencias de engagement.

Alcanzables:

- Utilizar fuentes de datos accesibles, como bases de datos existentes, reportes descargables, o APIs de plataformas sociales compatibles.
- Implementar únicamente funcionalidades adaptadas al nivel técnico del cliente para garantizar usabilidad.

Relevantes:

- Potenciar la toma de decisiones estratégicas de la marca en marketing digital para mejorar la conexión con su audiencia objetivo.

- Proveer una herramienta que permita a las marcas posicionarse frente a sus competidores mediante insights innovadores.

Definidas en el tiempo:

- Completar el desarrollo e implementación del dashboard en un plazo de 3 meses desde el inicio del proyecto.
 - Realizar un análisis de resultados post-implementación dentro de los 30 días posteriores a la entrega final.
 - Capacitar al equipo del cliente en el uso del dashboard en un período de 2 semanas tras la entrega.
-

Planificación del Proyecto

Cronograma

Lista de tareas y actividades

▼ **Fase 1: Investigación y Definición de Requerimientos**

- Investigación inicial
- Identificación de métricas clave (seguidores, likes, comentarios, etc.)
- Definición de requisitos para el dashboard
- Análisis de limitaciones de datos y privacidad

▼ **Fase 2: Configuración Técnica**

- Configuración del entorno técnico.
- Recopilación de datos históricos de muestra obtenidos de redes sociales, para marcas en diferentes categorías.
- Extracción de datos.

▼ **Fase 3: Análisis de Datos Descriptivo**

- Análisis de patrones de interacción y crecimiento de seguidores.
- Revisión de datos y tratamiento de outliers.
- Análisis Exploratorio de Datos (EDA).
- Análisis de sentimiento en captions y comentarios.
- Desarrollo de consultas para visualizaciones y métricas clave.
- Creación de visualizaciones iniciales.

▼ **Fase 4: Desarrollo del Componente Predictivo**

- Selección de algoritmos de machine learning
- Entrenamiento y evaluación de modelos predictivos
- Optimización y selección del modelo final

▼ **Fase 5: Creación del Dashboard en Streamlit**

- Implementación de visualizaciones interactivas.
- Inclusión de análisis exploratorios adicionales.
- Representación de los resultados de los modelos predictivos.
- Diseño centrado en la experiencia del usuario.

▼ **Fase 6: Pruebas y validación**

- Validación en dispositivos
- Pruebas con usuarios clave
- Ajustes al modelo predictivo

▼ **Fase 7: Implementación y Documentación**

- Documentación del proceso
- Creación de un manual de usuario
- Capacitación a usuarios clave

▼ Fase 8: Actividades de Soporte y Seguimiento

- Actualización y seguimiento de tareas
- Evaluación de riesgos y contingencias
- Monitoreo y mejora continua

▼ Revisión características adicionales

Fechas de inicio: Noviembre 11

Fechas final: Diciembre 13

Recursos

Para garantizar el éxito del proyecto, es esencial contar con recursos humanos, técnicos y tecnológicos adecuados. A continuación, se detallan los principales recursos necesarios:

1. Recursos Humanos

Identificación del equipo de trabajo: El equipo estará compuesto por profesionales con habilidades multidisciplinarias que cubran tanto el análisis de datos como el desarrollo de herramientas digitales. Los roles y responsabilidades se asignan de la siguiente manera:

Rol	Responsabilidades	Perfil Ideal
Líder del proyecto	<ul style="list-style-type: none"> - Supervisar y coordinar todas las actividades del proyecto. - Establecer objetivos generales y específicos con el equipo. - Tomar decisiones estratégicas y cumplir plazos. 	Profesional con experiencia en gestión de proyectos tecnológicos, capacidad de liderazgo y análisis de datos.
Coordinador	<ul style="list-style-type: none"> - Actuar como enlace entre los equipos. - Asegurar comunicación fluida y cumplimiento de tareas. - Monitorear el progreso y reportar avances al líder. 	Experiencia en gestión de equipos y planificación operativa.
Ingeniero o Analista de Datos	<ul style="list-style-type: none"> - Investigar algoritmos de análisis de datos y tendencias. - Diseñar bases conceptuales y metodológicas para análisis descriptivo y predictivo. 	Experto en ciencias de datos o estadística, con conocimiento en análisis de redes sociales.
Creativo	<ul style="list-style-type: none"> - Proponer ideas innovadoras para la presentación visual. - Diseñar estrategias para mejorar la experiencia del usuario (UX/UI). 	Diseñador gráfico o especialista en diseño de interfaces, con experiencia en dashboards y visualización.
Comunicador	<ul style="list-style-type: none"> - Establecer estrategias de comunicación para resultados. - Preparar reportes, presentaciones y materiales de capacitación para usuarios. 	Especialista en comunicación corporativa, con habilidades para sintetizar información técnica.
Equipo técnico para el dashboard	<ul style="list-style-type: none"> - Desarrollar, implementar y probar el dashboard en Streamlit. - Asegurar funcionalidad y mantenimiento continuo. - Integrar bases de datos y algoritmos de machine learning. 	Desarrolladores y analistas con experiencia en Streamlit

2. Recursos Tecnológicos

1. Bases de datos:

- Necesarias para almacenar y analizar métricas clave de redes sociales, como:
 - Número de seguidores, likes, comentarios, interacciones.
 - Datos históricos de contenido publicado.

- **Ejemplos de tecnologías:** MySQL, PostgreSQL, BigQuery.

2. Algoritmos de machine learning:

- Utilizados para realizar predicciones y análisis de tendencias en las interacciones de las marcas.

• Posibles aplicaciones:

- Modelos de regresión para identificar los tipos de contenido que generan más engagement.
- Análisis de líneas temporales

- **Librerías recomendadas:** Sklearn, Pandas, Seaborn, Matplotlib, Streamlit, Numpy, Plotly

3. Plataforma de desarrollo del dashboard:

- **Herramienta:** Streamlit (o alternativas como Tableau, Power BI).
- Utilizada para diseñar un dashboard visualmente atractivo e interactivo que facilite la toma de decisiones estratégicas.

3. Infraestructura Adicional

1. Servidores y almacenamiento en la nube:

- Espacios seguros y escalables para almacenar datos y ejecutar los análisis.
- **Proveedores recomendados:** AWS, Google Cloud Platform, Microsoft Azure.

2. Herramientas de comunicación y gestión del proyecto:

- **Ejemplos:** Trello, Asana, Slack, Microsoft Teams.
- Facilitan la colaboración y el seguimiento de tareas entre los miembros del equipo.

Presupuesto

Estimación de Costos

La estimación de costos para este proyecto se realizó considerando los recursos necesarios para el desarrollo, implementación, mantenimiento y promoción del dashboard, así como los gastos relacionados con la adquisición de datos, equipo técnico y actividades de marketing.

Asignación de Recursos Financieros

Se estableció un presupuesto total de **\$50,000 USD**, distribuido de la siguiente manera:

1. Desarrollo y Mantenimiento del Dashboard:

- Diseño y programación inicial del dashboard: \$15,000 USD
- Licencias de software (Streamlit, Looker Studio, etc.): \$2,000 USD anuales
- Mantenimiento mensual del dashboard: \$500 USD mensuales.

2. Acceso a Datos y APIs de Redes Sociales:

- Suscripción a APIs de redes sociales (Twitter, Instagram, YouTube): \$5,000 USD anuales.
- Costos de almacenamiento y procesamiento en la nube: \$3,000 USD anuales.

3. Gastos en Marketing y Adquisición de Clientes:

- Campañas publicitarias en redes sociales: \$8,000 USD
- Diseño de contenido promocional: \$1,500 USD
- Eventos y webinars para lanzamiento: \$1,000 USD

4. Sueldos del Equipo Técnico, de Soporte y Ventas:

- **Equipo técnico (desarrolladores y analistas):**
 - **2 desarrolladores** a \$4,000 USD mensuales x 3 meses = **\$24,000 USD**
- **Equipo de soporte técnico:**

- **1 soporte** a \$2,000 USD mensuales x 3 meses = **\$6,000 USD**
- **Equipo de ventas:**
 - **1 vendedor** a \$2,500 USD mensuales x 3 meses = **\$7,500 USD**

Estructura de Costos

La estructura de costos refleja una división clara entre costos fijos y variables, distribuidos según las necesidades del proyecto:

Categoría	Detalle	Costo (USD)
Desarrollo y Mantenimiento del Dashboard	Diseño inicial y programación	\$15,000
	Licencias de software y mantenimiento mensual	\$8,000
Acceso a Datos y APIs de Redes Sociales	APIs y almacenamiento en la nube	\$8,000
Marketing y Adquisición de Clientes	Promoción y contenido	\$10,500
Sueldos del Equipo Técnico, Soporte y Ventas	Salarios totales	\$37,500
Total Estimado del Presupuesto		\$79,000

Supuestos Considerados:

1. Los costos de desarrollo se ajustarán si se requieren funciones adicionales.
2. Las licencias y acceso a APIs están basados en planes estándar de cada plataforma.
3. El equipo de soporte y ventas puede variar dependiendo del alcance real del lanzamiento.
4. Los gastos de marketing consideran una campaña inicial de 3 meses.

Identificación de Riesgos

Tipo de Riesgo	¿Qué podría salir mal?	Plan de contingencia
Técnico: Limitaciones en la recopilación y calidad de datos	<ul style="list-style-type: none"> - Restricciones de las APIs de redes sociales (cambios en permisos o limitaciones de acceso). - Calidad inconsistente o incompleta de los datos extraídos. 	<ul style="list-style-type: none"> - Mantenerse actualizado sobre políticas de plataformas. - Implementar limpieza y validación de datos. - Diseñar un sistema adaptable para incluir fuentes alternativas.
Diseño y funcionalidad: Complejidad del dashboard interactivo	<ul style="list-style-type: none"> - Dificultad para integrar métricas en un dashboard intuitivo. - Baja aceptación por curva de aprendizaje o funcionalidades mal ajustadas. 	<ul style="list-style-type: none"> - Realizar pruebas iterativas con usuarios finales. - Diseñar un sistema modular para ajustes rápidos. - Ofrecer soporte técnico y materiales formativos.
Cronograma: Retrasos en las etapas del proyecto	<ul style="list-style-type: none"> - Retrasos en las etapas por problemas técnicos o cambios en requerimientos. 	<ul style="list-style-type: none"> - Dividir el proyecto en entregables pequeños. - Incorporar tiempo extra en el cronograma. - Priorizar funcionalidades esenciales y planificar mejoras posteriores.
Financiero: Exceder el presupuesto asignado	<ul style="list-style-type: none"> - Aumentos inesperados en costos de software, licencias o personal técnico. 	<ul style="list-style-type: none"> - Realizar análisis de costos inicial detallado y monitorear regularmente. - Buscar herramientas de bajo costo o alternativas open-source. - Contar con margen financiero adicional.
Seguridad:	<ul style="list-style-type: none"> - Vulnerabilidades en el manejo de datos sensibles. - Uso indebido de datos por parte de terceros. 	<ul style="list-style-type: none"> - Implementar protocolos estrictos de seguridad. - Garantizar cumplimiento de normativas como el GDPR. - Realizar auditorías periódicas de seguridad.
Organizativo	<ul style="list-style-type: none"> - Falta de compromiso de marcas o usuarios clave. - Dificultad para recopilar información sobre necesidades específicas de clientes. 	<ul style="list-style-type: none"> - Establecer reuniones regulares con clientes. - Implementar comunicación fluida y destacar beneficios del proyecto. - Garantizar entregables claros en cada etapa para mantener interés y confianza.

Descripción de las fases del proyecto

Fase 1: Investigación y definición de requerimientos

Investigación inicial

Identificación de métricas clave para análisis de predicción

En Instagram, ciertas métricas son clave para evaluar la efectividad de los esfuerzos de marketing. Estas métricas ayudan a medir el alcance, la interacción, el crecimiento de seguidores y el rendimiento del contenido, proporcionando una visión de qué está funcionando y qué áreas necesitan mejoras.

1. Tasa de interacción (engagement rate)

- **Por qué es importante:** La tasa de interacción muestra cómo interactúan activamente los usuarios con el contenido en relación al tamaño de la audiencia, destacando su atractivo.



Fórmula Tasa de interacción por publicación:

$$(Likes + Comments) / Seguidores acumulados (Estimados)$$



Importancia: Indicador directo del éxito de las publicaciones e interés de la audiencia.

- **Tasa de interacción promedio (en varias publicaciones):** Calcula la tasa de interacción para cada publicación en un período y luego saca el promedio.

$$Tasa de Interacción Avg = Mean((Total Interacción / total de seguidores) * 100)$$

2. Interacción total

$$Fórmula : (Total de Likes + Total de Comentarios)$$

3. Tasa de crecimiento de seguidores

Permite ver la rapidez con la que crece la audiencia, ayudando a evaluar la efectividad de tu contenido y estrategias promocionales.



Estimado con el acumulado de comentarios en cada publicación (post).

3. Alcance Promedio (Reach) Estimado

Mide cuántas personas ven las publicaciones y es fundamental para analizar la visibilidad de la marca.



Alcance Total: Sumatoria de VideoPlayCount

Promedio Alcance:

4. Impresiones (Impressions)

Cuenta el número total de veces que el contenido ha sido visto, incluidas las vistas múltiples por los mismos usuarios. Un alto número de impresiones en comparación con el alcance puede indicar interés de la audiencia.

Las impresiones son contadas por Instagram y no requieren cálculo, pero se pueden comparar con el alcance usando:



Estimado impresiones: (total likes + total comentarios + sum(VistasVideos) + Sum(PlayVideos))

- Una proporción alta sugiere múltiples vistas por usuario, lo cual puede indicar un alto interés o valor de repetición.

Si no se tiene el dato de impresiones, este se puede calcular.



Impresiones: Aproximación basada en vistas, plays e interacciones generales.

5. Comentarios por Publicación

Rastrear la cantidad de comentarios por publicación muestra el nivel de conversación en torno a tus publicaciones, indicando una interacción más profunda.

$$\text{ComentariosPromedio por Publicación} = \text{ComentariosTotales en un Período} / \text{Total de Publicaciones en ese Período}$$

6. Viralidad

Basado en hashtags, captions, and media type.



Importancia: Identificar contenido que genera mayor alcance.

7. Periodo óptimo de posteo

- Se calcula con la fecha/hora de posteo y análisis de interacción.



- **Importancia:** Asegura una visibilidad máxima.

- **Frecuencia de posteo.**

8. Éxito por tipo de medio de publicación

Comparación de tasa de interacción por tipo.



Importancia: Determina el mejor formato para futuras campañas.

9. Relevancia de hashtags:

Basado en el uso de hashtags vs. interacción.



Importancia: Guía la estrategia de uso de palabras clave para un mayor descubrimiento de la cuenta.

10. Calculo de nuevos seguidores

$$\text{Seguidores Nuevos} = (\text{núm likes} + \text{Núm Comentarios}) / \text{promedio(tasa de interacción)}$$



Importancia: métrica para proyectar el número de nuevos seguidores basado en las variables.

Descripción de cada métrica para rankear una cuenta.

1. **Seguidores:** Número total de seguidores en la cuenta.
2. **Tasa de interacción:** Mide el nivel de interacción en relación con la base de seguidores.
3. **Promedio de Me gusta:** Promedio de "me gusta" por publicación durante un período.

4. **Promedio de comentarios:** Promedio de comentarios por publicación durante un período.
5. **Tasa de crecimiento basado en nuevos seguidores**
6. **Relevancia: Evaluación basada en la alineación del contenido con la categoría o el nicho.**

Cada métrica proporciona información valiosa y, al monitorear varias de ellas simultáneamente, se puede obtener una visión completa del rendimiento de la cuenta de Instagram. Analizar estas métricas a lo largo del tiempo permite entender tendencias y oportunidades para optimizar la estrategia.

Fase 2. Extracción de datos:

Los datos se han extraído para dos cuentas de Instagram que puedan servir como base para la muestra y la construcción de un prototipo de análisis y visualización de datos.

Cuentas de Instagram seleccionadas por cantidad de seguidores, ubicación en Colombia, y la categoría Educación.

- **AtravelerTeacher:** Cuenta de Instagram de Educación primaria, 35 000 seguidores aprox. a la fecha Nov 20/2024
- **ProfeKyle:** Cuenta de Instagram sobre aprendizaje del idioma inglés. Ubicado en Colombia. 151000 seguidores a Nov20/2024, es considerado influencer en este nicho.

Se investigaron varias formas de obtener los datos. Haciendo web scraping fue la forma más efectiva, dadas las limitaciones de tiempo para hacerlo directamente por la API y el entorno developer de META, era necesario tener permisos para acceder a Instagram insights, de donde se obtienen los datos históricos de interacción, datos a los que solo tiene acceso el creador de la cuenta.

Por lo tanto, se utilizó el aplicativo Apify, que funciona como una extensión del navegador Chrome y permite obtener datos limitados y que son públicos para cualquier cuenta de Instagram Pública. Sin embargo, los datos de insights no son posibles de ser accedidos de esta forma.

Los datos para ambas cuentas se han almacenado en archivos XLSX.

Un archivo de datos descriptivos de la cuenta, un archivo de posts (publicaciones).

1 registro 13 columnas.

id:

- Identificador único de la cuenta de Instagram. Se utiliza como clave para identificar la cuenta en bases de datos o sistemas.

username:

- Nombre de usuario público de la cuenta (por ejemplo, @ejemplo_usuario). Es la forma en que la cuenta es identificada públicamente.

biography:

- La biografía de la cuenta, donde el usuario puede compartir una breve descripción, misión, hashtags, emojis o enlaces.

fullName:

- El nombre completo asociado a la cuenta. Puede ser el nombre real, el nombre de un negocio o un título creativo.

url:

- Probablemente un enlace a la página de perfil de la cuenta o un URL asociado.

followersCount:

- Cantidad total de seguidores de la cuenta. Es una métrica clave para medir el alcance y la popularidad.

postsCount:

- Número total de publicaciones realizadas por la cuenta. Refleja el nivel de actividad de la misma.

isBusinessAccount:

- Indica si la cuenta es una cuenta comercial. Las cuentas comerciales suelen tener acceso a estadísticas adicionales y herramientas promocionales.

verified:

- Valor booleano que indica si la cuenta está verificada por Instagram (con la marca azul). Generalmente, se otorga a figuras públicas, marcas o entidades notables.

followsCount:

- Número de cuentas que esta cuenta sigue. Puede dar una idea de los intereses o hábitos de conexión del usuario.

externalUrl:

- Un enlace externo en la biografía de la cuenta, generalmente utilizado para redirigir a sitios web, tiendas en línea o portafolios.

highlightReelCount:

- Número de historias destacadas (highlight reels) que la cuenta tiene. Estas son colecciones de historias guardadas y mostradas en el perfil.

igtvVideoCount:

- Cantidad de videos en IGTV que la cuenta ha publicado. Esto refleja el uso de la plataforma de videos de formato largo de Instagram.

businessCategoryName:

- Categoría comercial asignada a la cuenta, como "Educación," "Moda" o "Tecnología." Es relevante para clasificar las cuentas comerciales según su contenido y público objetivo.

El archivo de posts tiene los siguientes campos que fueron seleccionados directamente de Apify, considerando los campos que nos puedan servir de utilidad según la investigación preliminar.

La estructura del archivo perfil tiene la siguiente estructura:

El dataset contiene **1130 registros y 47 campos** (Atraverte teacher)

Columnas principales y su interpretación:

1. **id:** Identificador único de cada publicación (ID generado por Instagram).
2. **caption:** Texto de la publicación (descripción o pie de foto).
3. **description:** Parece estar vacío en su mayoría, pero podría ser para descripciones extendidas.
4. **timestamp:** Fecha y hora de publicación en formato ISO (AAAA-MM-DDTHH:MM:SS).
5. **displayUrl:** URL de la imagen principal asociada a la publicación.
6. **videoUrl:** URL del video (si es un video; de lo contrario, puede ser NaN).
7. **ownerUsername:** Nombre de usuario del propietario de la publicación.
8. **locationId:** ID de la ubicación asociada a la publicación (si aplica).
9. **locationName:** Nombre de la ubicación asociada (por ejemplo, ciudad o lugar).
10. **type:** Tipo de publicación (por ejemplo, "Video", "Sidecar").
11. **productType:** Detalles del producto (para publicaciones de tipo video).
12. **firstComment:** Primer comentario en la publicación (si existe).
13. **likesCount:** Cantidad de "me gusta" recibidos.
14. **commentsCount:** Cantidad de comentarios recibidos.
15. **videoPlayCount:** Número de reproducciones del video (si aplica).
16. **videoViewCount:** Número de vistas del video (diferente de reproducciones).
17. **videoDuration:** Duración del video en segundos.
18. **hashtags/0, hashtags/1, ..., hashtags/29:** Hashtags usados en la publicación, distribuidos en 30 columnas.

Fase 3: Análisis de Datos Descriptivo

En esta fase, se realizará un análisis exhaustivo de los datos recopilados para ambas cuentas de Instagram, enfocándose en las métricas clave identificadas en la fase de investigación. Se examinarán patrones de engagement, frecuencias de publicación y la efectividad de diferentes tipos de contenido. El análisis incluirá la identificación de tendencias temporales y la correlación entre variables como el uso de hashtags y las tasas de interacción.

- **Análisis de patrones de interacción y crecimiento de seguidores:**

- Se analizaron los datos históricos para identificar tendencias en el crecimiento del número de seguidores.
- Asimismo, se estudiaron las dinámicas de interacción mediante el análisis de métricas como "likes", comentarios, reproducciones de video y vistas de video.

- **Revisión de datos y tratamiento de outliers:**

- Se realizó un análisis exhaustivo para detectar y tratar valores atípicos (outliers) que pudieran distorsionar los resultados.
- Este proceso incluyó la evaluación del comportamiento de los datos por diferentes métricas, asegurando su consistencia y validez.

- **Análisis exploratorio de datos (EDA):**

- Se exploraron distribuciones de los datos y su comportamiento a lo largo del tiempo, así como en función de variables específicas como tipo de publicación y uso de hashtags.
- Estas exploraciones proporcionaron información clave para determinar posibles relaciones y patrones significativos.

- **Análisis de sentimiento en captions y comentarios:**

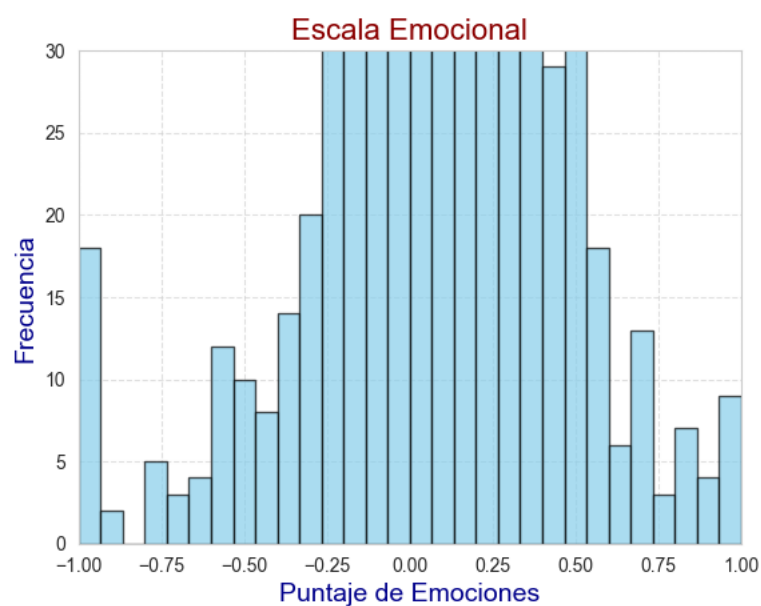
- Se implementó un análisis de sentimiento para evaluar las percepciones y emociones expresadas en los captions de las publicaciones y en los comentarios de los usuarios.
- Este análisis permitió obtener una visión más profunda de la interacción cualitativa de la audiencia.

- **Desarrollo de consultas para visualizaciones y métricas clave:**

- Se diseñaron y ejecutaron consultas específicas para calcular métricas como:
 - Nuevos seguidores.
 - Interacción total (basada en el conteo de "likes", comentarios, reproducciones de video y vistas).
 - Comportamiento por tipo de publicación (imágenes, videos, carruseles).
 - Análisis temporal del rendimiento (línea de tiempo).
- Estas consultas sentaron las bases para las visualizaciones y el análisis predictivo.

- **Creación de visualizaciones iniciales:**

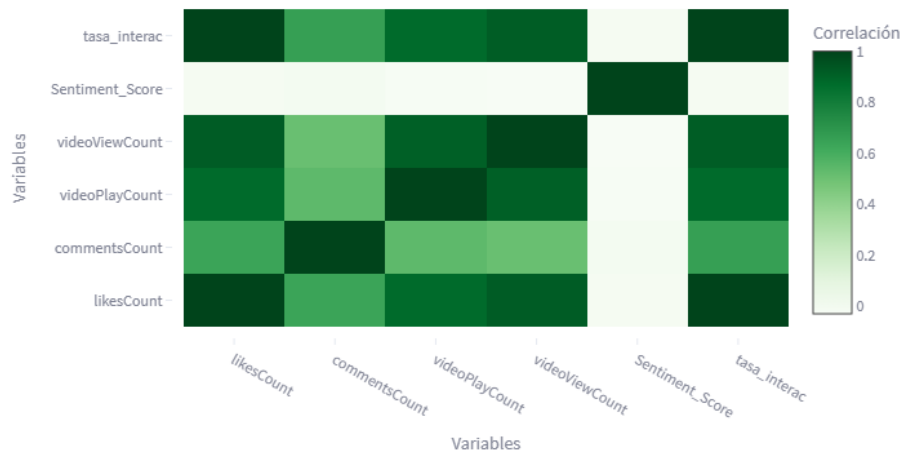
- Se generaron visualizaciones preliminares que incluyeron gráficos de distribución para evaluar la variabilidad de los datos.
- Estas representaciones visuales facilitaron la identificación de tendencias y patrones relevantes para el proyecto.
- Nube de palabras por hashtags, comentarios, mapa de calor de la co-ocurrencia o correlación entre hashtags. visualización de correlación de variables, mapa de ranqueo de palabras segun emociones (positiva, negativa, neutral)



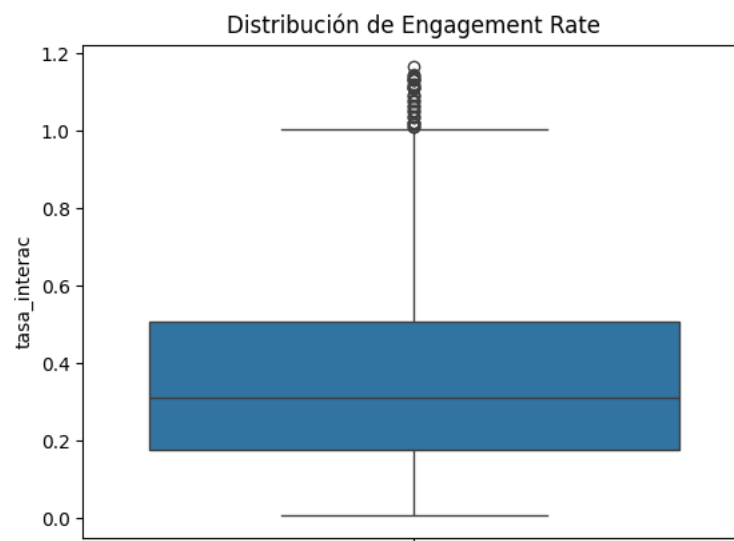
Relación entre Tasa de Interacción y Sentiment Score



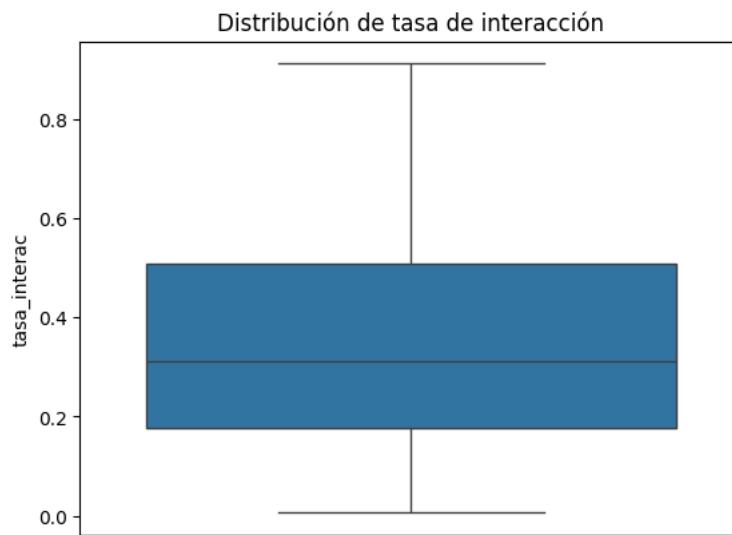
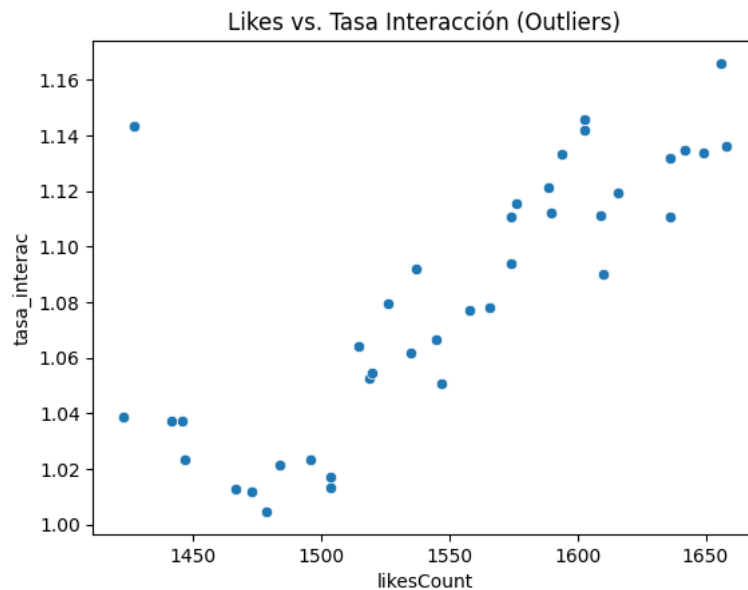
Matriz de Correlación



Matriz de correlación entre las variables



Distribución de la tasa de interacción. Visualización de datos atípicos (outliers)



Fase 4: Desarrollo del Componente Predictivo

Variables utilizadas para la implementación de los modelos

Para la implementación de los modelos predictivos se utilizaron los siguientes datos de entrada:

- **Sentiment_Score:** Para hallar el Sentiment_score se utilizó un modelo predictivo previo de **procesamiento del lenguaje natural (PLN)**, con la librería **Natural Language Toolkit (NLTK)** de Python. En el cual mediante el uso de la columna **"Caption"** se realizó un modelo para catalogar la polaridad en el sentimiento que tenía cada texto en la descripción de una publicación. La escala va de -1 a 1, donde -1 significa una palabra muy negativa, 0 una palabra neutra y 1 una palabra positiva.
- **likesCount:** Corresponde a la cantidad de likes/me gusta que tenga una publicación
- **commentsCount:** Corresponde a la cantidad de comentarios que tenga una publicación
- **videoPlayCount:** Corresponde al número de reproducciones que tiene un video por publicación (aplica solo para los tipos de publicación con video)
- **videoViewCount:** Corresponde al número de vistas que tiene un video por publicación (aplica solo para los tipos de publicación con video)

- **interaction_per_hashtag:** Corresponde a la relación entre la ("tasa de interacción") / (("numero de hashtag" + 1)). Donde la "tasa de interacción" corresponde en primera instancia a "tasa de interacción" = ((likesCount + commentsCount)) / (("Total de seguidores ") * 100). Mientras que el "número de hashtag" corresponde a la sumatoria de hashtag que tiene cada publicación.
- **Interaccion_por_videoduracion:** Corresponde a la relación entre ("tasa de interacción") / (("videoDuration" + 1)). Donde **videoDuration** es la duración que tiene un video en cada publicación

Desarrollo de la fase

En esta fase, se desarrolló el componente predictivo del proyecto mediante la implementación de modelos avanzados de machine learning. A continuación, se detallan las principales acciones realizadas y los resultados obtenidos:

1. Preparación y normalización de los datos:

- Los datos fueron normalizados para garantizar la uniformidad de las variables y mejorar el rendimiento de los modelos predictivos. Este paso fue esencial para evitar sesgos y asegurar la estabilidad de los resultados.

2. Modelos predictivos implementados:

- Se implementó un algoritmo de **procesamiento del lenguaje natural (PLN)**, con la librería **Natural Language Toolkit (NLTK) de Python** . Para hallar el **Sentiment_Score**.
- Posteriormente se unió el resultado del modelo anterior incluida la columna **Sentiment_Score** con las demás variables de entrada a través del método **Merge** de **Pandas** utilizando el **ID** de la publicación como criterio de unificación de ambos datasets.
- Se probaron diversos modelos, incluyendo:
 - Árboles de decisión para regresión.
 - Random Forest.
 - XGBoost.
- Estas técnicas se seleccionaron debido a su capacidad para manejar datos no lineales y su flexibilidad para trabajar con múltiples variables.

3. Selección y prueba de variables:

- Durante el desarrollo, se consideraron diversas variables predictoras, tales como:
 - Número de likes.
 - Número de comentarios.
 - Número de reproducciones y vistas de video.
 - Tasa de interacción por grupo de hashtags y por video.
 - Métrica del sentimiento en los captions o comentarios.
 - Tipo de publicación (imagen, carrusel, video, feed, igtv).
- Sin embargo, no todas las variables se incluyeron en el modelo final debido a la detección de **overfitting** en algunos casos. Este fenómeno se presentó cuando los modelos aprendieron patrones específicos del conjunto de entrenamiento que no generalizaban bien en nuevos datos.

4. Variable objetivo y su cálculo:

- La variable objetivo del modelo fue el número de nuevos seguidores, un indicador clave para evaluar el crecimiento de la audiencia.
- Debido a la falta de datos históricos directos sobre nuevos seguidores, se estimó esta métrica mediante la fórmula:

$$\text{Seguidoresestimado} = \text{tasadeinteracción} / (\text{NumLikes} + \text{NumComentarios})$$

- La tasa de interacción se calculó utilizando la mediana, con el propósito de mitigar el impacto de valores atípicos (outliers) en los resultados.

5. Evaluación y optimización del modelo:

- Los modelos se evaluaron utilizando métricas de desempeño como el error cuadrático medio (MSE) y el coeficiente de determinación (R²).
- Se realizaron iteraciones para ajustar hiperparámetros y mejorar la precisión de las predicciones, seleccionando finalmente el modelo que ofreció el mejor balance entre precisión y capacidad de generalización.

La evaluación y optimización del componente predictivo se llevó a cabo mediante la comparación de tres modelos: **Árboles de Decisión**, **Random Forest** y **XGBoost**. Estos modelos se analizaron en función de sus parámetros y su desempeño, medido principalmente a través del coeficiente de determinación (R²) y métricas complementarias. A continuación, se detalla el proceso:

1. Parámetros iniciales y configuración de los modelos:

Mejores parámetros después de aplicar GridSearch para el árbol de decisión

Mejores Parametros: {'max_depth': 10, 'min_samples_leaf': 5, 'min_samples_split': 10}

- **Árboles de Decisión Regressor:**

- Parámetros ajustados:

```
max_depth=8,
min_samples_leaf=5,
min_samples_split=10,
random_state=42)
```

- Este modelo se utilizó como punto de partida debido a su simplicidad y rapidez en el entrenamiento.
- Cross-Validated MAE: 27.616885440726673
- Training MAE: 14.17, Test MAE: 22.81

Importancia de las columnas

likesCount	0.996909
interaction_per_hashtag	0.001701
commentsCount	0.001321
Interaccion_por_videoduracion	0.000031
videoPlayCount	0.000023
videoViewCount	0.000015
Sentiment_Score	0.000000

- **Random Forest Regressor:**

- Parámetros ajustados: número de árboles (n_estimators), profundidad máxima (max_depth), y el número de características a considerar en cada división (max_features).

```
rf_model = RandomForestRegressor(max_depth=10, n_estimators=100,
                                min_samples_leaf=2, min_samples_split=2, random_state=42)
rf_model.fit(X_train, y_train)
rf_predicciones = rf_model.predict(X_test)
```

Cross-Validated MAE: 17.64421950428951

- Su capacidad para reducir el riesgo de overfitting gracias a la agregación de múltiples árboles fue clave en su elección.

- **XGBoost:**

- Parámetros ajustados: tasa de aprendizaje (learning_rate), número de iteraciones (n_estimators), y el peso de regularización (lambda).

```
xgb_model = XGBRegressor(
    max_depth=8,          # Maximum tree depth
    n_estimators=100,      # Number of trees
```

```

learning_rate=0.05,          # Learning rate
random_state=42,
eval_metric="mae",
early_stopping_rounds=10
)

```

Importancia de las columnas

```

likesCount          0.993991
commentsCount       0.003274
interaction_per_hashtag 0.002550
videoPlayCount      0.000060
videoViewCount      0.000053
Interaccion_por_videoduracion 0.000038
Sentiment_Score     0.000035
dtype: float32

```

Este modelo fue seleccionado por su capacidad de manejo eficiente de datos, flexibilidad y potencia predictiva.

1. Métrica de evaluación principal: Coeficiente de Determinación (R2):

- El R2 fue utilizado para medir la proporción de variabilidad explicada por el modelo en relación con los datos observados.
- Valores más cercanos a 1 indicaron un mejor desempeño en términos de ajuste a los datos.

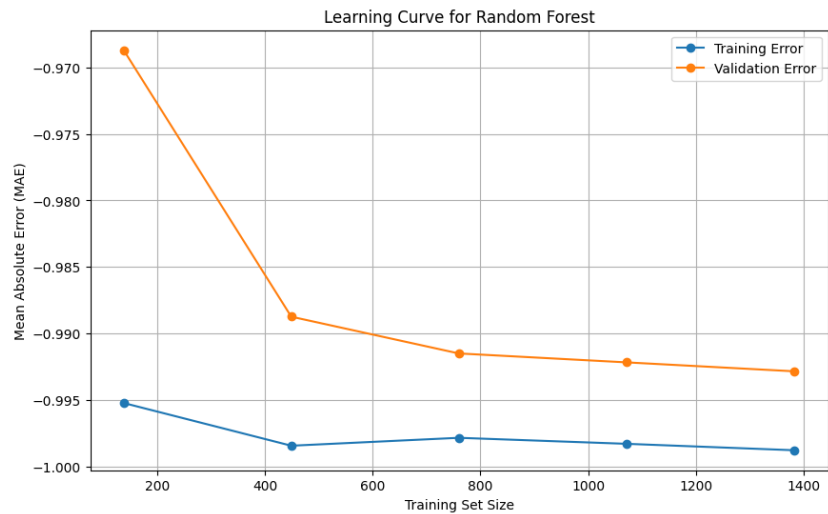
1. Resultados de la comparación:

• Árboles de Decisión:

- R2 promedio: moderado, con una tendencia a sobreajustar en conjuntos pequeños de datos.
- Mejor desempeño en variables altamente categóricas o con relaciones simples.
- Limitación: menor capacidad para capturar relaciones complejas entre las variables.

• Random Forest:

- R²: significativamente más alto que Árboles de Decisión, gracias a su enfoque en la combinación de múltiples árboles.
- Fortalezas: robustez frente a overfitting y mejor manejo de interacciones entre variables.
- Limitación: mayor tiempo de entrenamiento y consumo de recursos computacionales.



Curva de aprendizaje del modelo Bosques Aleatorios

• XGBoost:

- R^2 : el más alto de los tres modelos, demostrando un excelente equilibrio entre ajuste y capacidad de generalización.
- Mejor desempeño en conjuntos de datos grandes y con características correlacionadas.
- Limitación: mayor complejidad en la configuración de hiperparámetros y tiempo de entrenamiento más prolongado.

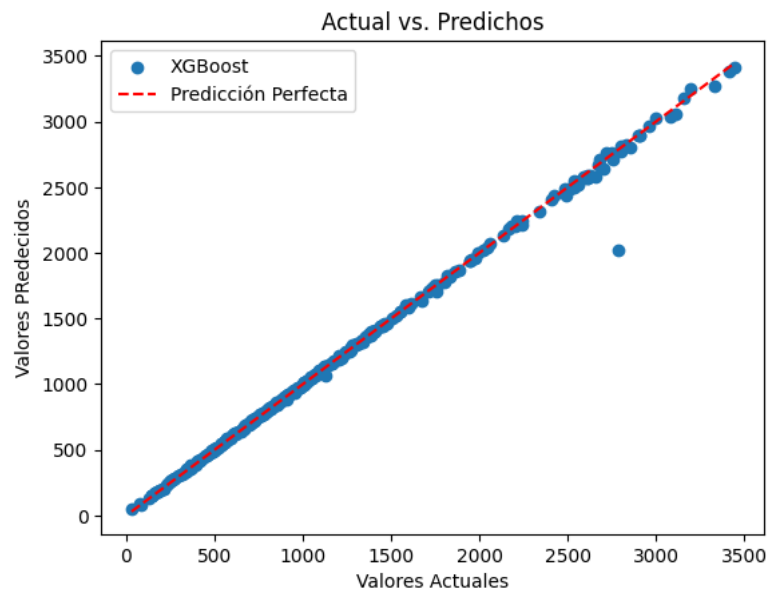
DecisionTree - MAE: 22.81, R^2 : 0.99
 Random Forest - MAE: 12.49, R^2 : 0.99
 XGBoost - MAE: 11.71, R^2 : 1.00

Optimización y selección del modelo final:

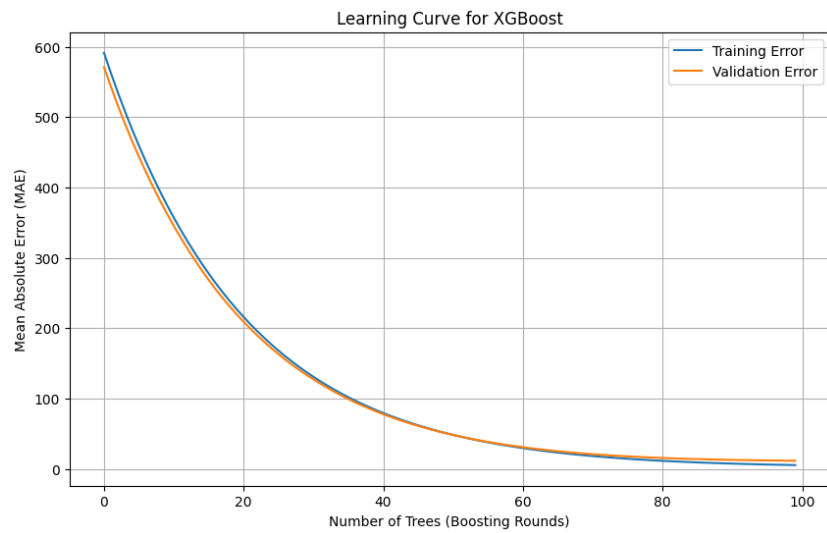
- Los hiperparámetros de cada modelo se ajustaron utilizando una búsqueda en cuadrícula (Grid Search) para identificar las configuraciones que maximizaron el R^2 .
- Se utilizaron validación cruzada (k-fold cross-validation) para garantizar la estabilidad de los resultados y evitar el sobreajuste a un subconjunto específico de datos.
- Tras la comparación, **XGBoost** fue seleccionado como el modelo final debido a su mayor R^2 y su capacidad para generalizar de manera eficiente en nuevos datos. El modelo proporcionó el mejor balance entre precisión predictiva y robustez frente a diferentes configuraciones de variables.

Esta fase permitió no solo predecir el crecimiento en términos de nuevos seguidores, sino también identificar las variables más influyentes en este proceso, proporcionando insights valiosos para la toma de decisiones estratégicas en el ámbito del marketing digital.

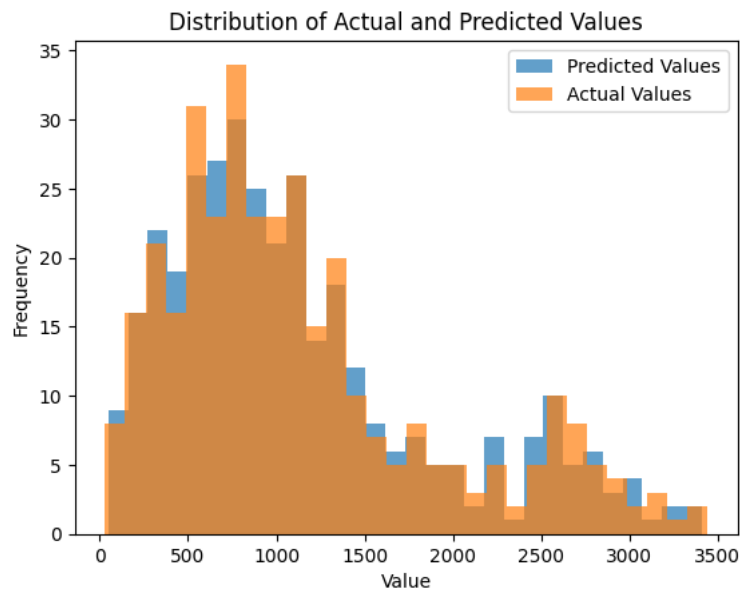
Mejor Modelo por MAE: XGBoost (MAE: 11.71)
 Mejor Modelo por R^2 : XGBoost (R^2 : 1.00)



Resultados finales. Datos predichos



Curva de aprendizaje Modelo XGBoost



Distribución de datos predichos con relación a los datos

Fase 5: Creación del Dashboard en Streamlit

En esta fase se desarrolló un dashboard interactivo utilizando la herramienta **Streamlit**, con el objetivo de facilitar la visualización y exploración de los datos analizados. Este dashboard fue diseñado para presentar resultados relevantes de manera accesible y estructurada, permitiendo a los usuarios finales interactuar con las métricas y obtener insights clave. A continuación, se describen los principales aspectos de esta etapa:



<https://socialpulse2.streamlit.app/>

1. Implementación de visualizaciones interactivas

- Las visualizaciones fueron creadas empleando la biblioteca de visualización de datos de **Streamlit**, la cual permitió una integración fluida con los datasets correspondientes a las cuentas de Instagram: **ProfeKyle** y **atravelerteacher**.
- Se diseñaron gráficos y diagramas interactivos que ofrecieron una representación clara de las métricas clave de las dos cuentas, permitiendo la comparación de datos y el análisis en tiempo real.

2. Inclusión de análisis exploratorios adicionales

- Además de los análisis realizados en fases anteriores, el dashboard incorporó métricas descriptivas que no habían sido incluidas previamente, tales como:
 - Varianza, media y mediana** de las principales variables.
 - Estadísticas descriptivas** que resumieron el comportamiento general de los datos.
- Estas métricas proporcionaron una perspectiva más completa y detallada del comportamiento de las cuentas analizadas, ofreciendo un contexto adicional para la interpretación de los resultados.

3. Representación de los resultados de los modelos predictivos

- El dashboard integró los diagramas y resultados derivados de los tres modelos predictivos implementados en la fase anterior (**Árbol de Decisión, Random Forest y XGBoost**), incluyendo:
 - Comparaciones gráficas de las predicciones realizadas por cada modelo.
 - Visualización de métricas de desempeño, como el coeficiente de determinación (R^2) y el error cuadrático medio (MSE).
- Estos elementos permitieron a los usuarios evaluar el desempeño relativo de los modelos y comprender la base estadística detrás de las predicciones realizadas.

4. Diseño centrado en la experiencia del usuario

- Se priorizó un enfoque intuitivo y accesible, asegurando que el dashboard pudiera ser utilizado por personas con conocimientos técnicos limitados.
- La interfaz fue diseñada para ser visualmente atractiva y funcional, con un esquema de navegación sencillo que guió al usuario a través de las diferentes secciones.



Las fases siguientes se describen sobre una hipotética puesta en producción de la presente propuesta.

Fase 6: Pruebas y Validación

La fase de pruebas y validación se diseñará para garantizar que el sistema desarrollado funcione de manera óptima en diferentes entornos y cumpla con las expectativas de los usuarios clave. Las actividades previstas incluyen:

1. Validación en dispositivos:

- Pruebas de compatibilidad en diferentes dispositivos, como computadoras de escritorio, tabletas y teléfonos móviles.
- Verificación del rendimiento del dashboard en diversos navegadores (Chrome, Firefox, Safari, Edge) y sistemas operativos (Windows, macOS, Android, iOS).
- Identificación y solución de problemas relacionados con la resolución de pantalla, tiempos de carga y capacidad de respuesta del dashboard.

2. Pruebas con usuarios clave:

- Selección de un grupo representativo de usuarios finales para realizar pruebas funcionales y de usabilidad.
- Aplicación de escenarios de prueba que simulen el uso real del dashboard para identificar posibles áreas de mejora.
- Recolección de retroalimentación cualitativa y cuantitativa a través de cuestionarios, entrevistas y observaciones directas.

3. Ajustes al modelo predictivo:

- Evaluación del desempeño del modelo predictivo en condiciones reales mediante datos actualizados.
- Identificación de desviaciones en las predicciones y ajuste de hiperparámetros para mejorar la precisión.
- Reentrenamiento del modelo si es necesario, utilizando los datos obtenidos durante las pruebas.

Fase 7: Implementación y Documentación

Esta fase se enfocará en la puesta en producción del proyecto, asegurando que los usuarios finales puedan utilizarlo de manera efectiva y que exista una documentación completa del proceso para futuras referencias. Las actividades principales serán:

1. Documentación del proceso:

- Elaboración de un informe técnico que detalle todas las fases del proyecto, desde la conceptualización hasta la implementación.
- Documentación de las decisiones clave tomadas durante el desarrollo, incluyendo las metodologías utilizadas, los modelos seleccionados y los ajustes realizados.
- Creación de un repositorio de documentación que incluya diagramas, código fuente y análisis de resultados.

2. Creación de un manual de usuario:

- Desarrollo de un manual detallado que explique el uso del dashboard, incluyendo instrucciones paso a paso para la navegación, la interpretación de gráficos y el uso de filtros.
- Inclusión de capturas de pantalla y ejemplos prácticos para facilitar el aprendizaje.
- Traducción del manual, si es necesario, para adaptarlo a diferentes públicos.

3. Capacitación a usuarios clave:

- Diseño e implementación de sesiones de capacitación para los usuarios finales, enfocadas en el uso eficiente del dashboard y la interpretación de los resultados.
- Creación de materiales complementarios como tutoriales en video, presentaciones interactivas y hojas de ayuda.
- Evaluación de la comprensión de los usuarios mediante actividades prácticas y simulaciones.

Fase 8: Actividades de Soporte y Seguimiento

Una vez implementado el sistema, se llevará a cabo un conjunto de actividades orientadas a garantizar su funcionamiento continuo y su mejora progresiva. Estas actividades incluyen:

1. Actualización y seguimiento de tareas:

- Monitoreo regular del rendimiento del dashboard para identificar posibles problemas técnicos.
- Implementación de actualizaciones periódicas para agregar nuevas funcionalidades o mejorar las existentes.
- Priorización y resolución de solicitudes de mejora realizadas por los usuarios.

2. Evaluación de riesgos y contingencias:

- Identificación de riesgos operativos o de seguridad asociados con el uso del sistema en producción.
- Desarrollo de planes de contingencia para abordar posibles fallos técnicos o vulnerabilidades de seguridad.
- Realización de auditorías periódicas para garantizar el cumplimiento de normativas de privacidad y protección de datos.

3. Monitoreo y mejora continua:

- Implementación de métricas de desempeño para evaluar el impacto del dashboard en la toma de decisiones de los usuarios.
- Recolección de retroalimentación continua a través de encuestas y foros de discusión con los usuarios.
- Aplicación de un enfoque iterativo para optimizar la experiencia del usuario y la precisión de los análisis proporcionados por el sistema.

Ejecución del Proyecto

• Comunicación Efectiva

El equipo se reúne de forma sincrónica una o dos veces por semana, en sesiones programadas los jueves o viernes dentro del horario establecido por el programa. Adicionalmente, se trabaja de manera asíncrona a través de un grupo de WhatsApp, utilizado para compartir avances, recursos y materiales relacionados con el proyecto. Este espacio también facilita pequeñas discusiones y conversaciones sobre diversos aspectos del trabajo, promoviendo una colaboración constante y efectiva.

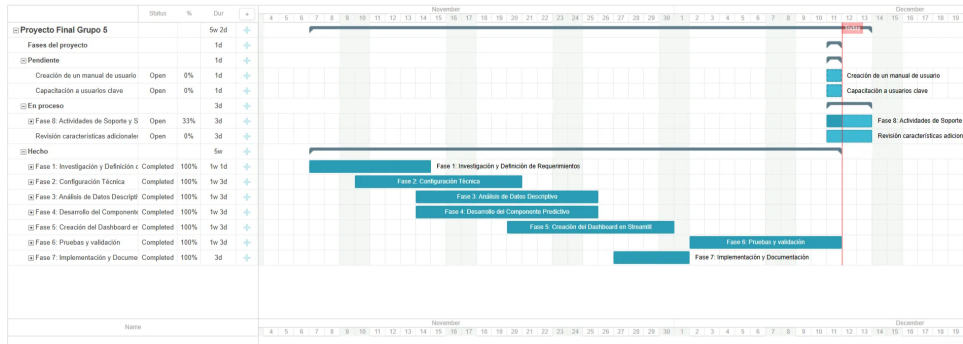
De manera regular, se entregaron informes de progreso semanalmente, conforme a las solicitudes realizadas por el programa.

• Gestión del Equipo

Durante las reuniones, se abordaron temas clave en cada una de las fases del proyecto. Estos encuentros incluyeron discusiones sobre tópicos fundamentales, tales como la selección del tema del proyecto, las fuentes de datos, los métodos de acceso a la información, la metodología del proyecto y aspectos técnicos específicos. Las tareas se asignaron en función de las habilidades individuales de los miembros del equipo. Entre estas asignaciones se incluyeron actividades como la preparación de presentaciones, la redacción de documentos y la implementación de aspectos técnicos. En tareas más complejas, como la codificación de los modelos predictivos, se trabajó de manera conjunta, fomentando la participación activa de todos los integrantes.

• Seguimiento del Progreso

El proyecto ha avanzado de acuerdo con el cronograma establecido en el diagrama de Gantt, garantizando que las actividades se desarrollen dentro de los plazos estipulados.



• Gestión de Cambios

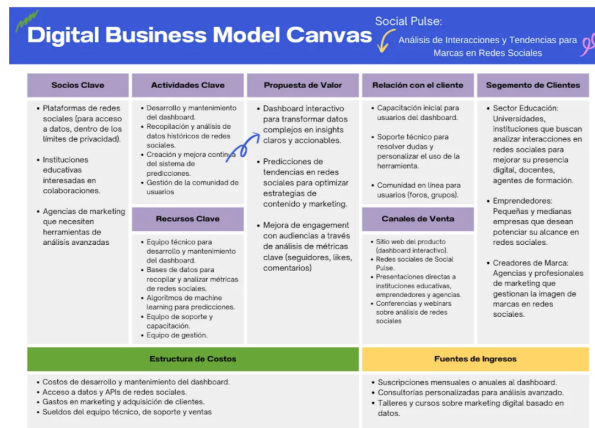
Durante el desarrollo del proyecto, se realizaron ajustes significativos, particularmente en la Fase 1. Fue necesario cambiar el tema del proyecto en dos ocasiones debido a la falta de acceso a los datos requeridos y la deserción de un miembro del equipo. A pesar de estos desafíos, en las Fases 2 y 3 se lograron avances sustanciales. Entre ellos, se destaca la extracción de datos, la investigación relacionada con el nuevo tema, la identificación de métricas relevantes y la realización de pruebas iniciales con datos de muestra. Asimismo, se llevaron a cabo análisis descriptivos y predictivos que establecieron una base sólida para las siguientes fases del proyecto.

• Gestión de la Calidad

Para garantizar la calidad de los resultados, se realizaron pruebas exhaustivas utilizando conjuntos de datos de prueba y datos definitivos. Este enfoque permitió validar los procesos y asegurar la confiabilidad de los análisis y resultados obtenidos.



• Business Model Canvas



Modelo Business Model Canvas. Social Pulse.

Entregables

- Carpeta con los archivos:

La carpeta drive incluye la carpeta proyecto final grupo 5 (streamlit) con los archivos organizados para el despliegue de la visualización.

La carpeta código proyecto final 5, con el código y los archivos xlsx, csv, py, ipynb para su ejecución.

Carpeta archivos



[Carpeta Drive](#)

Dashboard



<https://socialpulse2.streamlit.app/>

Conclusiones

El proyecto "**Social Pulse: Análisis de Interacciones y Tendencias para Marcas en Redes Sociales**" demostró la importancia de integrar análisis descriptivos y predictivos con herramientas de visualización para optimizar estrategias de marketing digital. A continuación, se presentan las principales conclusiones derivadas del desarrollo del proyecto:

1. Impacto del Modelo Predictivo:

- La implementación del modelo predictivo permitió identificar variables clave como la tasa de interacción, los likes, los comentarios, el tipo de publicación, análisis de emociones (sentimiento) de caption y comentarios, las cuales son fundamentales para anticipar el crecimiento en términos de nuevos seguidores.
- El modelo XGBoost mostró ser el más efectivo debido a su capacidad para manejar datos complejos, obteniendo un equilibrio óptimo entre precisión y generalización.

2. Relevancia del Dashboard Interactivo:

- El desarrollo del dashboard en Streamlit permitió consolidar los análisis y resultados en un entorno accesible e intuitivo, facilitando la toma de decisiones informadas por parte de los usuarios finales.
- Las visualizaciones interactivas, junto con la incorporación de métricas adicionales, ofrecieron una perspectiva integral y detallada del comportamiento de las cuentas analizadas.

3. Lecciones Aprendidas:

- El proyecto enfrentó desafíos significativos, como la falta de datos históricos de nuevos seguidores y la necesidad de ajustar modelos para evitar el overfitting. Estos retos resaltaron la importancia de la adaptabilidad del equipo y la planificación estratégica.
- La colaboración efectiva entre los miembros del equipo, junto con una comunicación clara y constante, fue clave para superar los obstáculos y lograr los objetivos planteados.

4. Contribución a la Innovación en Marketing Digital:

- Este proyecto proporciona una herramienta innovadora que combina técnicas avanzadas de machine learning con visualización de datos, permitiendo a las marcas comprender y anticipar mejor el comportamiento de su audiencia en redes sociales.
- La solución presentada sienta las bases para futuros desarrollos que puedan incluir análisis en tiempo real y segmentación más detallada de audiencias.

5. Proyección y Escalabilidad:

- El enfoque utilizado es aplicable a otras plataformas sociales y puede ampliarse para incluir métricas adicionales y funcionalidades avanzadas.
- La versatilidad del modelo y el dashboard lo convierte en una herramienta potencialmente valiosa para diferentes sectores empresariales y académicos.

6. Evaluación de los modelos

- El proceso de evaluación permitió identificar que las variables más influyentes para predecir el número de nuevos seguidores.

Recomendaciones

Se proponen las siguientes recomendaciones para optimizar futuras implementaciones y ampliar el alcance de la herramienta:

Acceso a Datos y Políticas de Privacidad:

- Establecer colaboraciones con plataformas sociales para garantizar el acceso a datos más completos mediante APIs oficiales. Esto reduciría las limitaciones encontradas al utilizar herramientas de terceros como Apify.
- Realizar un análisis exhaustivo de las políticas de privacidad de las plataformas desde la fase inicial del proyecto, asegurando la viabilidad de la recopilación de datos necesarios para cumplir con los objetivos planteados.
- Para proyectos futuros es fundamental evaluar previamente la accesibilidad a los datos y las fuentes de obtención disponibles. Las plataformas digitales y redes sociales suelen implementar estrictas políticas de privacidad que restringen el acceso a ciertos datos sin los permisos correspondientes, lo cual puede limitar significativamente el desarrollo óptimo del proyecto.

1. Optimización del Modelo Predictivo:

- Explorar la implementación de modelos adicionales, como redes neuronales o técnicas híbridas, para mejorar la precisión y robustez de las predicciones.
- Automatizar el proceso de ajuste de hiperparámetros mediante herramientas como búsqueda en cuadrícula o algoritmos genéticos, mejorando la eficiencia en proyectos similares.

2. Ampliación y Mejora del Dashboard:

- Incorporar funcionalidades de análisis en tiempo real para ofrecer insights más dinámicos y relevantes.
- Permitir mayor personalización para los usuarios finales, como la selección de métricas específicas, comparación entre períodos y análisis detallado por tipo de contenido.
- Incluir módulos de exportación de datos y gráficos en formatos accesibles como PDF o Excel, facilitando la generación de reportes.

3. Capacitación y Difusión:

- Diseñar programas de capacitación personalizados para los usuarios finales, asegurando que comprendan y aprovechen al máximo las funcionalidades del dashboard.
- Promover la solución en foros empresariales y académicos para recibir retroalimentación valiosa y explorar oportunidades de colaboración en nuevos contextos.

4. Monitoreo y Mejora Continua:

- Implementar un sistema de monitoreo regular para evaluar el desempeño del modelo predictivo y la herramienta, recopilando datos de uso y retroalimentación de los usuarios.
- Planificar actualizaciones periódicas para adaptar el sistema a cambios en las plataformas sociales y en las necesidades de los usuarios.

5. Proyección a Futuro:

- Explorar la posibilidad de integrar análisis en otras plataformas sociales, como TikTok y LinkedIn, para diversificar las aplicaciones del proyecto.
- Ampliar el análisis hacia la segmentación avanzada de audiencias y su comportamiento, incluyendo métricas demográficas y psicográficas para estrategias más personalizadas.

6. Evaluación de la Viabilidad Financiera y Técnica:

- Asegurar que futuras versiones del proyecto cuenten con un presupuesto realista que contemple recursos adicionales para licencias, infraestructura en la nube y mantenimiento continuo.
- Considerar herramientas open-source o soluciones escalables para optimizar costos sin comprometer la calidad.

**Nota:**

Los datos recopilados de la cuenta de Instagram

@AtravelerTeacher fueron obtenidos con el consentimiento explícito de su creadora. Además, al tratarse de una cuenta pública, los datos son accesibles dentro del marco permitido por las políticas de privacidad de Instagram. De manera similar, los datos de la cuenta **@ProfeKyle**, también pública, se recopilaron respetando las normativas vigentes sobre privacidad y uso de datos. Cabe destacar que toda la información (datos) utilizada en este proyecto tiene un propósito estrictamente académico y no comercial.