Proyecto Python

* Objetivo: Identificar un patrón de opinión con respecto a la campaña publicitaria navideña de Balenciaga del 2022

<https://www.elastic.co/es/what-is/sentiment-analysis>

TRABAJO INTEGRADOR PCC 2024-1 - GRUPOS 103 Y 104

A partir de un tema de interés específico proponer una solución aplicada a un producto de comunicación en cualquiera de sus especialidades haciendo uso de código en Python. Pueden ser análisis o visualizaciones de datos, buscadores de información, aplicaciones interactivas, sistema de reservas, software específico para automatizar algún proceso, entre otros.

Se entrega un documento (Code Book) y el cuaderno exportado de Google Colab. Éste último debe estar #comentado paso a paso. El documento inicial, con citación en formato APA consiste en:

Título

Integrantes

Introducción al problema (Justificar el problema a resolver, >= 200 palabras)

Propuesta de solución (Explicar cómo va a funcionar el código, incluir un diagrama de flujo, >= 300 palabras)

Reflexión final (Explicar las limitaciones del proyecto y qué otros pasos podrían aplicar para mejorar la solución en una ocasión posterior; >=200 palabras).

Algunas bases de datos gratuitas son:

Plataforma nacional de datos abiertos: https://www.datosabiertos.gob.pe/

Tendencias de Google (hacen la búsqueda sobre un tema de su preferencia y descargan el CSV): https://trends.google.com/

Google Public Data Explorer: https://www.google.com/publicdata/directory

Singapore Data: https://beta.data.gov.sg/

Registry of Open Data in AWS (Amazon Web Services): https://registry.opendata.aws/

Algunos ejemplos de código interactivos [no pueden repetirlos] son:

Juego de Piedra, Papel o Tijera: https://colab.research.google.com/drive/1fnwVRy2iEqGD9gcRSt87CoOetjNZI2d1?usp=sharing

Juego de Adivina el Número: <https://colab.research.google.com/drive/1-J-1hR4NIfzrcoLZSYNq5RUrkVdX0BRU?usp=sharing>

<https://www.kaggle.com/datasets/shivamb/netflix-shows/data>

<https://www.kaggle.com/datasets/micheldc55/publicity-influence-on-sales>

<https://www.kaggle.com/>

<https://developer.x.com/en/docs/twitter-api>

<https://www.octoparse.es/blog/scraper-gratis-de-twitter>

**Herramienta de Análisis de Sentimiento en Redes Sociales:**

* **Descripción:** Un programa que analice los comentarios y menciones en redes sociales para determinar el sentimiento hacia una marca o producto.
* **Tecnologías:** Python, APIs de redes sociales (Twitter API, Facebook Graph API), procesamiento de lenguaje natural (NLP).
* **Características:** Recolección de datos de redes sociales, análisis de sentimiento, visualización de datos.

### **Herramienta de Análisis de Sentimiento en Redes Sociales**

**Descripción del Proyecto:** El objetivo de esta herramienta es analizar los comentarios y menciones en redes sociales sobre una marca o producto específico, y determinar el sentimiento (positivo, negativo o neutral) de estos comentarios. Esta información puede ser muy valiosa para las empresas, ya que les permite entender la percepción pública y ajustar sus estrategias de marketing en consecuencia.

**Componentes del Proyecto:**

1. **Recolección de Datos:**
   * **API de Redes Sociales:** Utilizar APIs como la de Twitter (Twitter API) y Facebook (Facebook Graph API) para recolectar menciones, comentarios y publicaciones sobre la marca o producto.
   * **Frecuencia de Recolección:** Configurar el sistema para que recolecte datos de manera periódica (por ejemplo, cada hora o cada día).
2. **Procesamiento de Lenguaje Natural (NLP):**
   * **Limpieza de Datos:** Eliminar enlaces, menciones de usuarios, hashtags y otros elementos que no aporten valor al análisis de sentimiento.
   * **Tokenización:** Dividir el texto en palabras o tokens para un análisis más detallado.
   * **Eliminación de Palabras Vacías:** Quitar palabras comunes que no aportan significado (stopwords) como "el", "de", "y".
3. **Análisis de Sentimiento:**
   * **Modelos Preentrenados:** Utilizar modelos de análisis de sentimiento preentrenados como VADER (Valence Aware Dictionary and sEntiment Reasoner) o TextBlob para obtener una puntuación de sentimiento.
   * **Modelos Personalizados:** Entrenar un modelo propio usando técnicas de machine learning (como Random Forest o SVM) o deep learning (con TensorFlow o PyTorch) en caso de necesitar un análisis más preciso y específico para el dominio.
4. **Almacenamiento de Datos:**
   * **Base de Datos:** Utilizar una base de datos (como PostgreSQL o MongoDB) para almacenar los comentarios y sus correspondientes análisis de sentimiento.
   * **Estructura de Datos:** Almacenar datos con campos como ID, fecha, texto del comentario, autor, puntuación de sentimiento, etc.
5. **Visualización de Resultados:**
   * **Dashboard:** Crear un dashboard interactivo con herramientas como Tableau, Power BI, o bibliotecas de visualización en Python (como Dash, Plotly, o matplotlib).
   * **Gráficos y Tablas:** Incluir gráficos de tendencias de sentimiento a lo largo del tiempo, distribución de sentimientos, palabras más frecuentes en comentarios positivos/negativos, etc.
6. **Alertas y Reportes:**
   * **Alertas:** Configurar alertas automáticas cuando haya un cambio significativo en el sentimiento (por ejemplo, un aumento repentino de comentarios negativos).
   * **Reportes Periódicos:** Generar reportes semanales o mensuales que resuman el análisis de sentimiento y los principales hallazgos.

<https://colab.research.google.com/drive/1lOaP8JUFiQllNG9y41sRom5f3-yba8UL?usp=sharing>

Las marcas necesitan estar presentes y activas en al menos tres redes sociales relevantes para su target, ya que son esenciales para la comunicación y construcción de relaciones con los usuarios, convirtiéndose en una herramienta indispensable en el mundo digital. Un diálogo fluido en estas plataformas aumenta el compromiso de los usuarios, generando confianza al mostrar la historia de la empresa y testimonios de clientes, humanizando la marca a través del storytelling y facilitando el contacto como un canal adicional de atención al cliente que permite corregir problemas en tiempo real. Esta interacción cercana no solo fortalece los vínculos sino que también facilita la concreción de negocios y el aumento de las ventas mediante contenido de valor y call to actions efectivos.

<https://www.woowup.com/blog/la-presencia-en-redes-sociales#:~:text=En%20la%20actualidad%2C%20las%20redes,usuario%20tenga%20con%20nuestra%20marca>.

Tener presencia en redes sociales no es solo una opción para las grandes empresas, es una necesidad imperiosa en el mundo actual. Las redes sociales permiten una conexión directa y efectiva con los clientes, ofreciendo múltiples beneficios que van más allá de la simple publicidad.

<https://mktmarketingdigital.com/empresas-tener-presencia-en-redes-sociales/>

<https://www.nextibs.com/campanas-publicitarias-mas-polemicas/>

<https://www.puromarketing.com/30/213774/redes-sociales-buscadores-grandes-canales-reparten-pastel-marketing-publicidad-digital#:~:text=Las%20redes%20sociales%20se%20han,3%25%20respecto%20al%20a%C3%B1o%20anterior>.

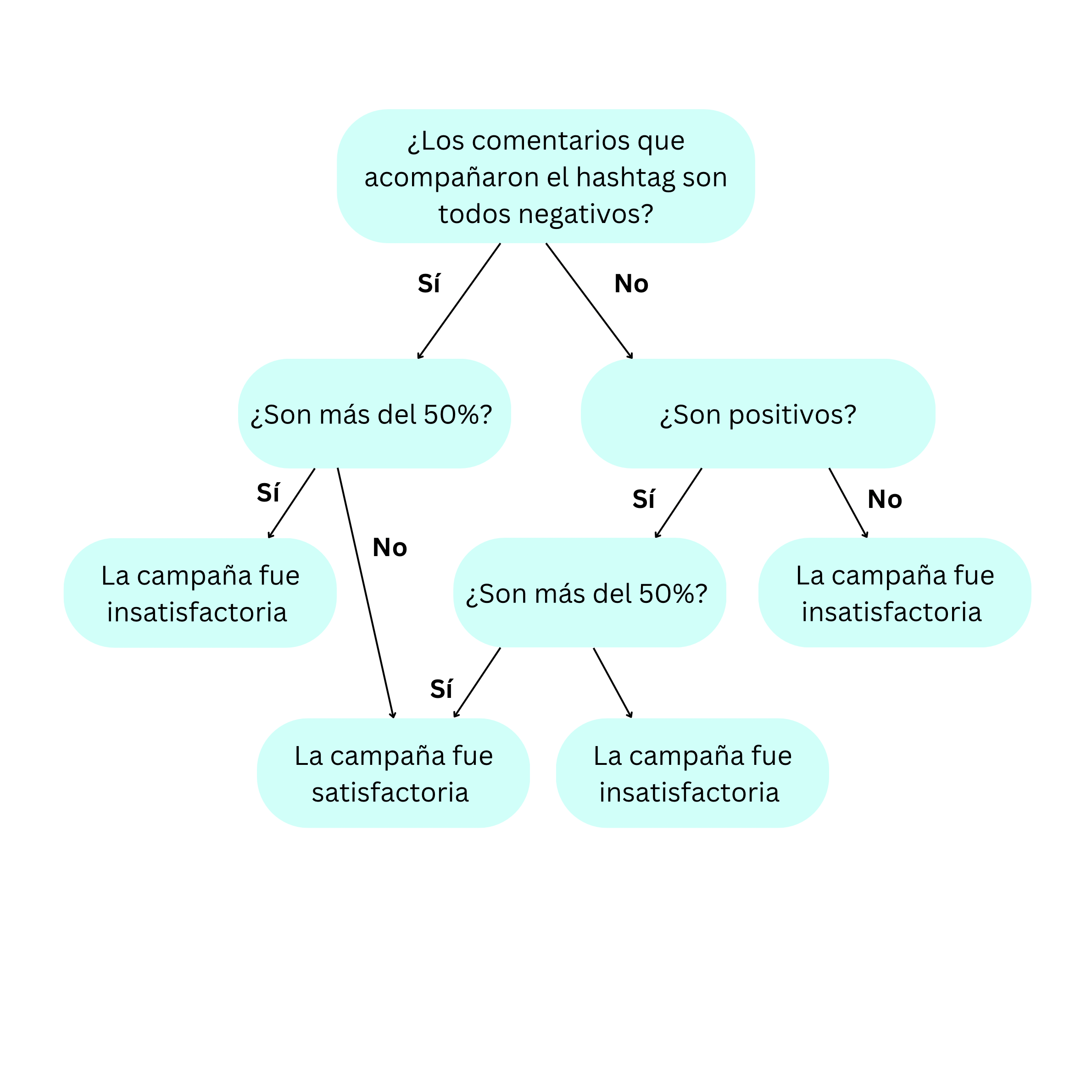
<https://tweetdelete.net/es/recursos/twitter-sentiment-analysis-tool-track-your-audiences-views/>.

**Introducción**

Las redes sociales son plataformas digitales que permiten el intercambio de ideas, intereses u opiniones entre diversos usuarios. Su naturaleza interactiva las han convertido en una herramienta indispensable para las grandes empresas, puesto que resultan efectivas al momento de crear conexión entre una marca y su público objetivo. Es así que estas plataformas se han posicionado como el principal canal publicitario a nivel mundial. Asimismo, son fundamentales para conocer y comprender las opiniones o percepciones del público sobre una campaña publicitaria. Las empresas pueden rastrear, en tiempo real, menciones, comentarios y conversaciones relacionadas, y así saber cómo el público está respondiendo a la campaña. Este tipo de análisis les permite mejorar sus estrategias de marketing para abordar las preocupaciones del cliente y mejorar la percepción de su marca.

En 2022, Balenciaga lanzó una campaña publicitaria muy polémica, pues presentaba a niños y niñas junto a disfraces bondage. Este evento se viralizó rápidamente en redes sociales, siendo X una de las plataformas más notables: el hashtag “#cancelBalenciaga” fue utilizado por los usuarios para criticar a la marca. El presente trabajo tiene como objetivo elaborar un código que nos permita analizar el uso de este hashtag y los comentarios asociados, con el fin de comprender los sentimientos que los usuarios de X manifestaron respecto a esta controversial publicidad.

**Diagrama de flujo**



API

API Key:

wbkfbuuHmHDhXzj5uBqVB76OD

API Key Secret:

xo4D8FRrd9szxo75g0UlBGsx5m2d4bh7uwEYtx9jhKGEfwIYxa

Access Token:

1501410058835578883-edh1QYxHXjdkXGfBUN0dD1vp1By8sy

Access Token Secret:

MSPiGH7MNCePVvmyDAlmrnUZkQZ5lAdwtPjYA8JKoxyqO

<https://geekytheory.com/analisis-sentimientos-twitter-python-apache-spark-streaming-tiempo-real/>

<https://cienciadedatos.net/documentos/py25-text-mining-python>

Para poder realizar este análisis se evaluarán las emociones mediante las opiniones que se dieron a conocer a través de la plataforma de X. De este modo, cabe resaltar que un análisis de sentimientos tiene el objetivo de identificar, extraer y estudiar información subjetiva. Por lo tanto, clasifica el texto como positivo, negativo y neutro, como se podrá visualizar en nuestro gráfico final. Asimismo, resulta importante mencionar que tanto el número de vistas, me gustas, reposteos y comentarios, son las métricas más utilizadas para realizar un análisis de los resultados de una campaña. Sin embargo, el análisis de sentimientos va más allá de estos números, por lo que importa más la calidad de las interacciones, es decir, las palabras más utilizadas y los patrones que se pueden identificar mediante lo escrito en cada tweet.

Metodología:

En primer lugar, mediante el uso de la herramienta de Octoparse, pudimos obtener una base de datos con información como el nombre de usuario, el texto que acompaña el hashtag, el número de likes, vistas, comentarios y compartidos. De esta manera, la base fue descargada e importada en el colab utilizado para ser analizada.

En segundo lugar, realizamos 7 códigos, de los cuales el primero ordena en un cuadro la base de datos que estaremos utilizando. El segundo código nos brinda un gráfico de barras con respecto a la frecuencia de tweets en cada mes desde el inicio de la campaña, es decir, en diciembre del 2022. El tercer código nos brinda un gráfico de barras con respecto a la frecuencia de publicación de tweets por parte de cada usuario, es decir, la cantidad de veces que un usuario utilizó el hashtag en sus tweets. El cuarto código nos brinda un gráfico de barras con la cantidad de likes que obtuvieron los 10 tweets más likeados. El quinto código nos brinda un gráfico de barras con la cantidad de retweets que obtuvieron los 10 tweets más compartidos. El sexto código nos brinda un wordcloud con palabras que fueron repetidas en más de un tweet. El séptimo código nos brinda un gráfico circular con la distribución de sentimientos que contienen los tweets.

<https://www.octoparse.es/>

# **Herramienta de Web Scraping Gratis**

Como primer paso del proyecto, se procedió a realizar el scraping de la plataforma Octoparse para recopilar tweets asociados al hashtag #cancelbalenciaga. Posteriormente, se construyó una base de datos en Excel utilizando los datos obtenidos, con el objetivo de realizar un análisis de sentimientos detallado.

De esta manera, al contar con la información necesaria dentro del Colabnote, procedimos a ordenar los datos obtenidos de la base de datos según tres criterios principales: el hashtag asociado, el nombre de usuario del autor del tweet y el contenido textual del propio tweet. Esta organización nos permitió estructurar y preparar los datos de manera efectiva para análisis posteriores y visualización de patrones significativos.

Mediante el análisis de los textos de los tweets recopilados, obtuvimos los siguientes resultados:

Primero, identificamos una asociación negativa entre la modelo Kim Kardashian y la marca Balenciaga. Esta percepción se puede identificar como parte de una intención negativa de los usuarios, la cual tuvo como consecuencia una tendencia emergente de cancelación también dirigida hacia esta figura pública.

Segundo, pudimos identificar la relevancia del tipo de lenguaje utilizado por los usuarios a través de las palabras más utilizadas, los tweets más reposteados y los más likeados. De este modo, identificamos que las palabras más mencionadas tienen una intención negativa. Asimismo, los tweets más reposteados y likeados contienen opiniones en contra de la marca Balenciaga, lo cual representa esta misma intención negativa identificada en las palabras que más se mencionaron. Las palabras más usadas por los usuarios de X suelen estar asociadas a vocablos fuertes como pedofilia o relacionado, lo cual pone en manifiesto la percepción negativa que se ha creado de la marca a raíz de su campaña controversial. También se pueden encontrar algunas relacionadas con satanismo o explotación infantil, que van de la mano con lo primero y son un reflejo de como la imagen de la marca ha sido fuertemente moldeada con base en pensamientos negativos.

Tercero, identificamos que existen usuarios que twittearon más de una vez con el hashtag analizado, lo cual representa también una frecuencia relevante para este proyecto. Esto se puede identificar como una intención de manifestar su incomodidad o descontento con la campaña.

Cuarto, identificamos que la mayoría de los tweets se publicaron en diciembre de 2022; se ahí, ha disminuido significativamente. Esto parece coherente considerando que este fue el periodo en el que la campaña se dio a conocer al público.

Por último, con respecto al objetivo principal del proyecto, el análisis de sentimientos que realizamos en última instancia, proporcionó información bastante relevante. Esto gracias a que pudimos identificar que casi el 50% de los comentarios se encuentran dentro del rango “neutro”, mientras que una minoría del 1% se encuentra dentro del rango “positivo”. Asimismo, el rango “negativo” obtuvo más del 50% de los comentarios, lo cual supera por un pequeño porcentaje al rango “neutro”. De este modo, podemos concluir que la mayoría de usuarios que utilizaron el hashtag #cancelbalenciaga se encuentra en un rango “neutro” y “negativo” con respecto a los sentimientos que generó la campaña navideña de Balenciaga en ellos. Esto se puede inferir gracias al código generado a partir de las palabras que más se repitieron en los comentarios, siendo estas clasificadas como positivas, negativas y neutras. De esta manera, podemos manifestar que una gran mayoría de los usuarios tuvo una intención de expresar sus sentimientos negativos y neutrales con respecto a la campaña.

Flores, S. (2022). ¿Por qué es importante la presencia de mi marca en redes sociales? Retail CRM Marketing. <https://www.woowup.com/blog/la-presencia-en-redes-sociales#:~:text=En%20la%20actualidad%2C%20las%20redes,usuario%20tenga%20con%20nuestra%20marca>.

Cecilia, M. (2020). ¿Por qué las grandes empresas deben tener presencia en redes sociales? Marketing Digital. <https://mktmarketingdigital.com/empresas-tener-presencia-en-redes-sociales/>

Sánchez, M. (2023). Campañas publicitarias más polémicas. Marketing Digital. <https://www.nextibs.com/campanas-publicitarias-mas-polemicas/>

Puro Marketing (2024). Redes sociales y buscadores: los grandes canales que se reparten el pastel del marketing y la publicidad digital. <https://www.puromarketing.com/30/213774/redes-sociales-buscadores-grandes-canales-reparten-pastel-marketing-publicidad-digital#:~:text=Las%20redes%20sociales%20se%20han,3%25%20respecto%20al%20a%C3%B1o%20anterior>.

[Bagyaratnam](https://tweetdelete.net/es/recursos/author/arnel-bagyaratnam/), A. (2024). Herramienta de análisis de sentimiento de twitter: siga las opiniones de su audiencia. Tweet delete. <https://tweetdelete.net/es/recursos/twitter-sentiment-analysis-tool-track-your-audiences-views/>

Martinez, J. (2022). Balenciaga, censurada por su campaña de Navidad con niños y ositos “bondage”. Fashion United. <https://fashionunited.es/noticias/moda/balenciaga-censurada-por-su-campana-de-navidad-con-ninos-y-ositos-bondage/2022112239849>