MÁSTER EN BIG DATA

CASOS DE ANALÍTICA

ANÁLISIS SENTIMENTAL TWITTER

REALIZADO POR:

David Mauricio López Sandoval

LA SALLE BARCELONA - UNIVERSIDAD RAMON LLULL

ABRIL, 2024

INDICE

0. ADQUISICION DE DATOS

Este segundo caso gira entorno al análisis sentimental de los tweets hechos por usuarios en la plataforma de Twitter, los datos se han extraído de una base de datos en un formato CSV de con la siguiente dirección

0.1 EXAMEN PRELIMINAR DEL DATASET

Formato CSV

El uso de archivos CSV para el examen preliminar del dataset presenta varias ventajas que facilitan el análisis inicial de los datos:

Carga Rápida de Datos:

Los archivos CSV pueden ser cargados rápidamente en memoria utilizando bibliotecas eficientes como pandas en Python. Esto permite realizar un análisis exploratorio de datos (EDA) sin demoras significativas.

Visualización Sencilla:

Dado que los datos en formato CSV están organizados en filas y columnas, es fácil visualizarlos y realizar inspecciones preliminares utilizando herramientas como pandas o Excel.

Identificación de Inconsistencias:

La estructura tabular de los archivos CSV facilita la detección de inconsistencias y anomalías en los datos, como valores faltantes, duplicados o atípicos, que pueden ser abordados durante la fase de limpieza.

Filtrado y Agrupamiento Eficientes:

Con las herramientas adecuadas, los datos en CSV pueden ser filtrados y agrupados rápidamente para identificar patrones iniciales y tendencias en el dataset. Esto es útil para comprender la distribución de los datos y planificar análisis más detallados.

En resumen, el formato CSV ofrece una combinación de simplicidad, flexibilidad y eficiencia que lo convierte en una opción ideal para la limpieza de datos y el análisis preliminar en proyectos de análisis de sentimientos en Twitter.

1 LIMPIEZA DE DATOS

El proceso de limpieza de datos es crucial para preparar el dataset de Twitter para análisis posteriores. A continuación, se describe detalladamente cada paso del script clean_data, que realiza varias transformaciones y limpiezas en el dataset original:

- Definición de Nombres de Columnas: Se asignan nombres significativos a las columnas del DataFrame original. Esto facilita el acceso y manipulación de los datos.

- Eliminación de Columnas Innecesarias:
- Conversión de Fechas: La columna tweet_date se convierte al formato datetime para facilitar la extracción de información temporal.
- Extracción de Información Temporal: Se extraen la hora del día y el día de la semana a partir de la columna de fecha y se añaden como nuevas columnas.
- Extracción de la Fecha: Se extrae solo la fecha (sin la hora) y se mantiene en la columna tweet_date.

Se aplica una función personalizada clean_text al texto de los tweets para limpiarlos y tokenizarlos. Esto incluye la eliminación de enlaces, menciones de usuarios, números y caracteres especiales, así como la conversión a minúsculas, eliminación de stopwords y lematización, se

escogió la lematización sobre el stemming en este proyecto de análisis de sentimientos en Twitter debido a su mayor precisión, capacidad de conservación del significado y consideración del contexto. Esto ayuda a obtener un análisis más confiable y útil de los datos textuales.

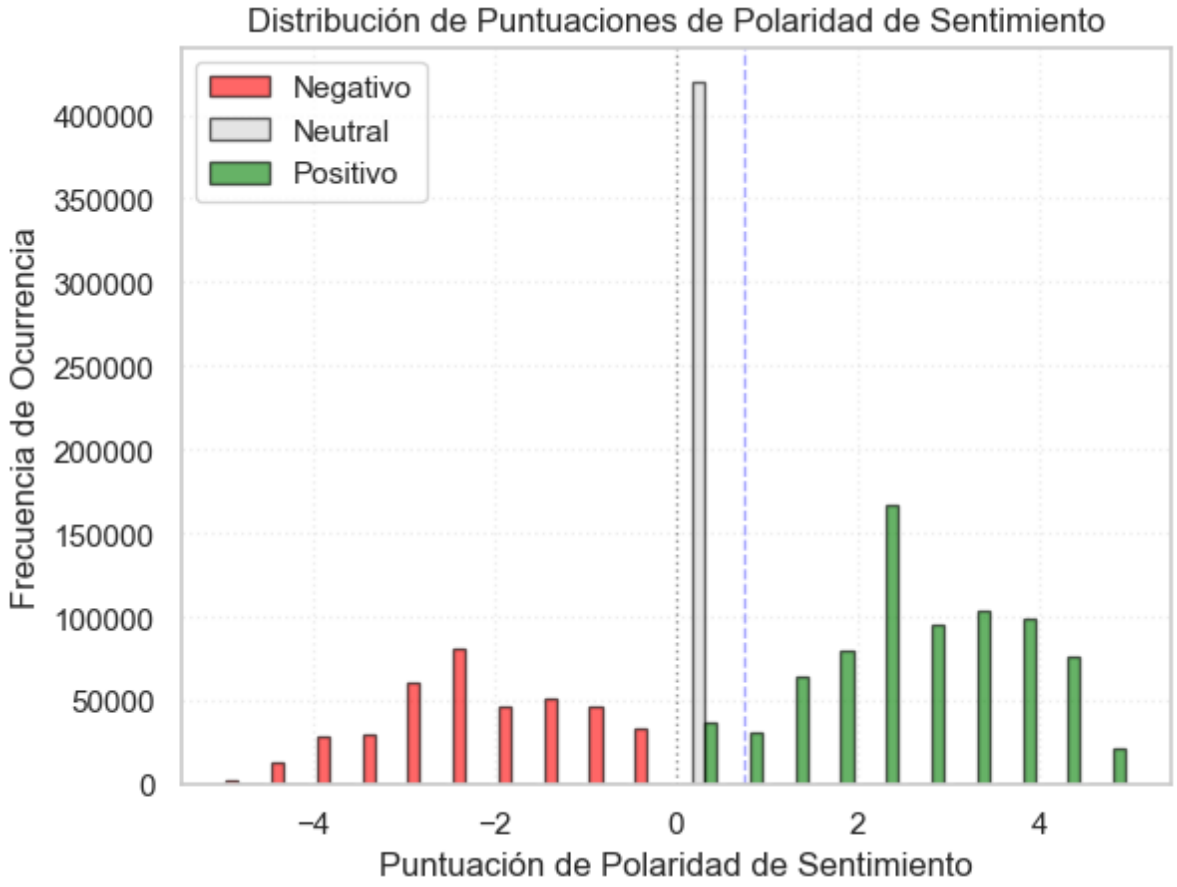
Se eliminan las columnas originales de texto (tweet_text) ya que ahora tenemos una versión limpia del texto.

- Se eliminan las filas donde clean_tweet es una cadena vacía.
- Eliminación de Filas con Valores Nulos:
- Se eliminan las filas con valores nulos en cualquier columna restante.

Este proceso de limpieza asegura que los datos están en un estado óptimo para el análisis de sentimientos y otras tareas de procesamiento de lenguaje natural

2 ANALISIS

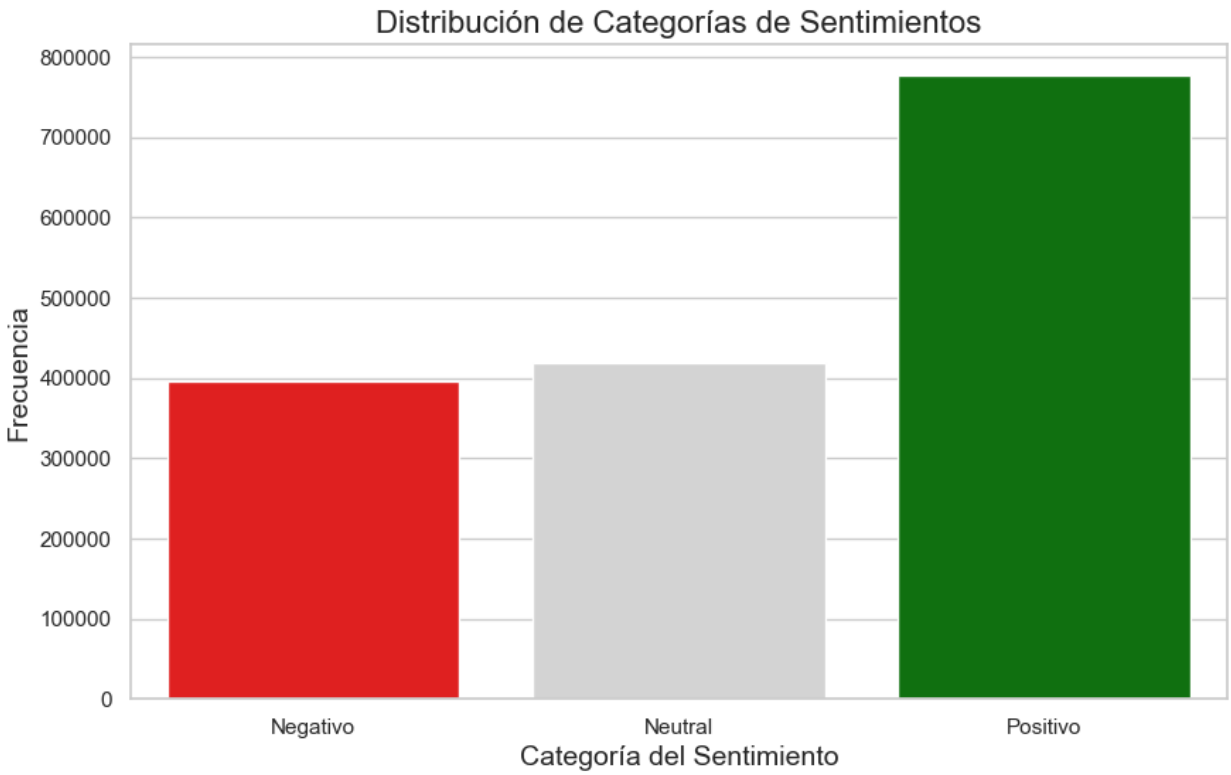
2.1.1 ¿Cuál es la distribución de las polaridades y complejidad de lectura/escritura de los tweets en el dataset?



Fuente: Elaborado por los autores

Se puede observar que la mayor parte de los datos se distribuyen en una polaridad entre 0 y 1, lo que indica una tendencia hacia lo neutro y positivo. Además, se percibe que los datos con mayor distribución después de los mencionados anteriormente son aquellos que mantienen una polaridad positiva, oscilando entre 2.5 y 3.

2.1.2 ¿Hay una mayor cantidad de tweets positivos, negativos o neutrales?



Fuente: Elaborado por los autores

Por otro lado, también podemos visualizar que la mayor cantidad de tweet son positivos seguido de los tweets neutros y la menor cantidad son de tweet negativos

2.1.3 ¿Cómo se relacionan las distintas polaridades según la complejidad de lectura/escritura de los tweets?

Para poder verificar la complejidad de un texto en ingles podemos utilizar una librería que se llama textstat,

que proporciona un buen manejo de índices para el idioma inglés, utilizaremos la métrica flesch reading ease que nos permite obtener

La escala del índice de puntuación es la siguiente:

90-100: Muy fácil de leer. Entendido por un estudiante promedio de 11 años.

80-90: Fácil de leer. Conversacional en inglés para consumidores.

70-80: Bastante fácil de leer.

60-70: Texto estándar. Entendido por estudiantes de 13-15 años.

50-60: Moderadamente difícil de leer.

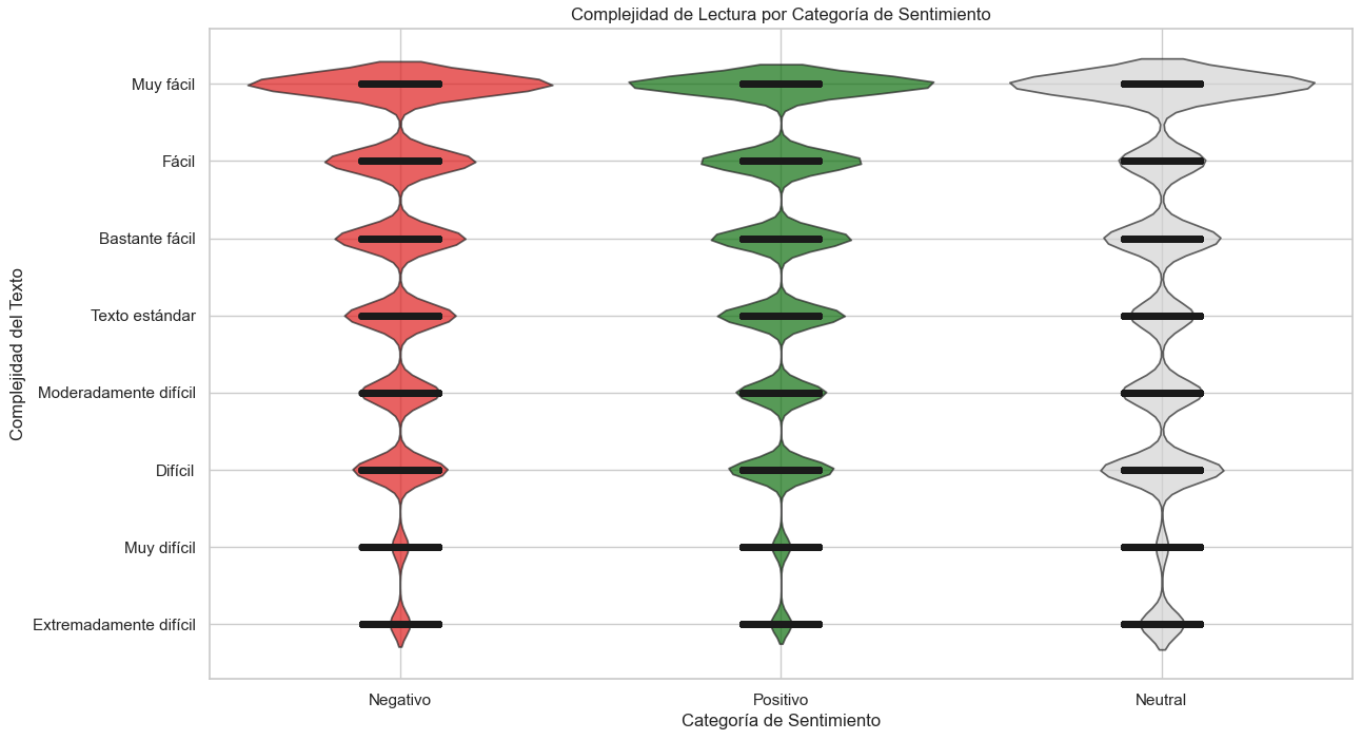
30-50: Difícil de leer.

10-30: Muy difícil de leer. Mejor entendido por graduados universitarios.

0-10: Extremadamente difícil de leer. Mejor entendido por graduados universitarios avanzados.

Además, como ya hemos realizado una limpieza anterior algunos tweets han quedado algunos textos muy cortos, la fórmula de Flesch puede producir resultados atípicos. Esto se debe a que la fórmula supone un cierto rango y tipo de contenido textual para funcionar correctamente.

Para ello se realiza una función ajustada que me permita obtener valores en la escala de 0 a 100



Fuente: Elaborado por los autores

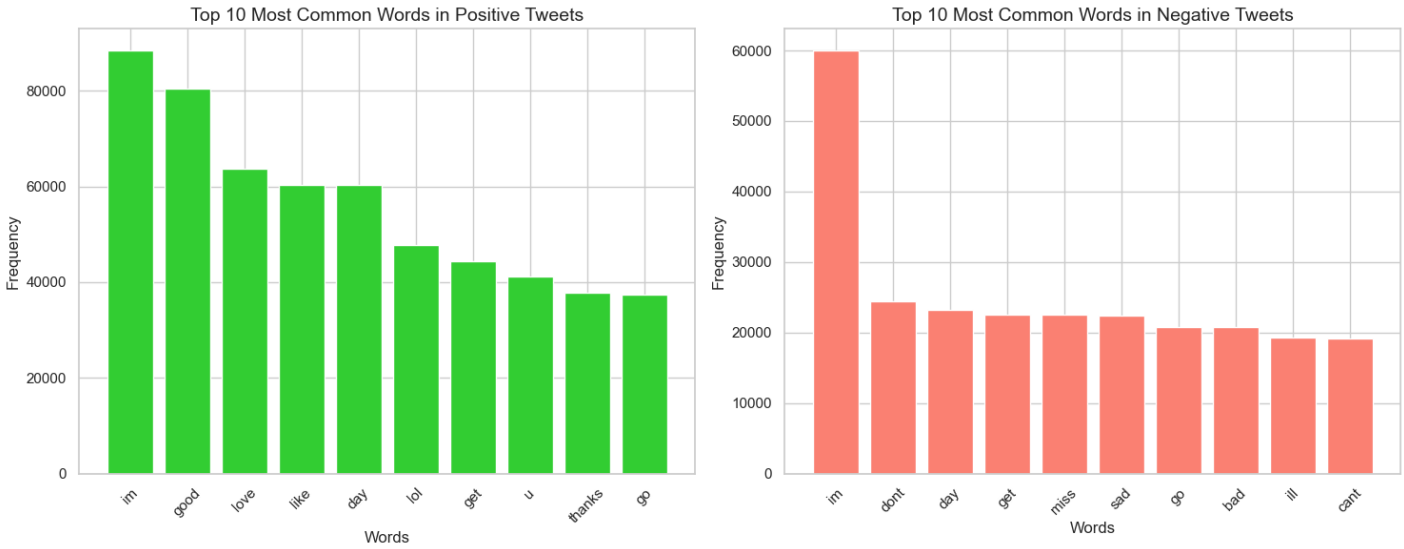
Como se puede observar, la mayoría de los datos en cada categoría de sentimiento pertenecen a la categoría de complejidad "Muy fácil". Este patrón indica que los tweets, independientemente de su polaridad (positiva, neutral o negativa), tienden a ser redactados de manera simple y accesible.

Sin embargo, también se identifican outliers que pertenecen a las categorías de "Muy difícil" y "Extremadamente difícil". Estos outliers sugieren que, aunque la mayoría de los tweets son fáciles de leer, existe una minoría significativa que utiliza un lenguaje más complejo y sofisticado.

Desde una perspectiva holística, la complejidad de los textos es predominantemente fácil, lo cual es coherente con la naturaleza de Twitter como plataforma de microblogging. Los usuarios de Twitter tienden a usar un lenguaje sencillo y directo para maximizar la comprensión y el alcance de sus mensajes en el límite de 280 caracteres.

En resumen, aunque la mayoría de los tweets son de fácil lectura, la existencia de outliers con alta complejidad destaca la diversidad de contenido en Twitter. Esta variabilidad en la complejidad del texto subraya la versatilidad de la plataforma, que puede servir tanto para comunicaciones rápidas y simples como para discusiones más profundas y detalladas.

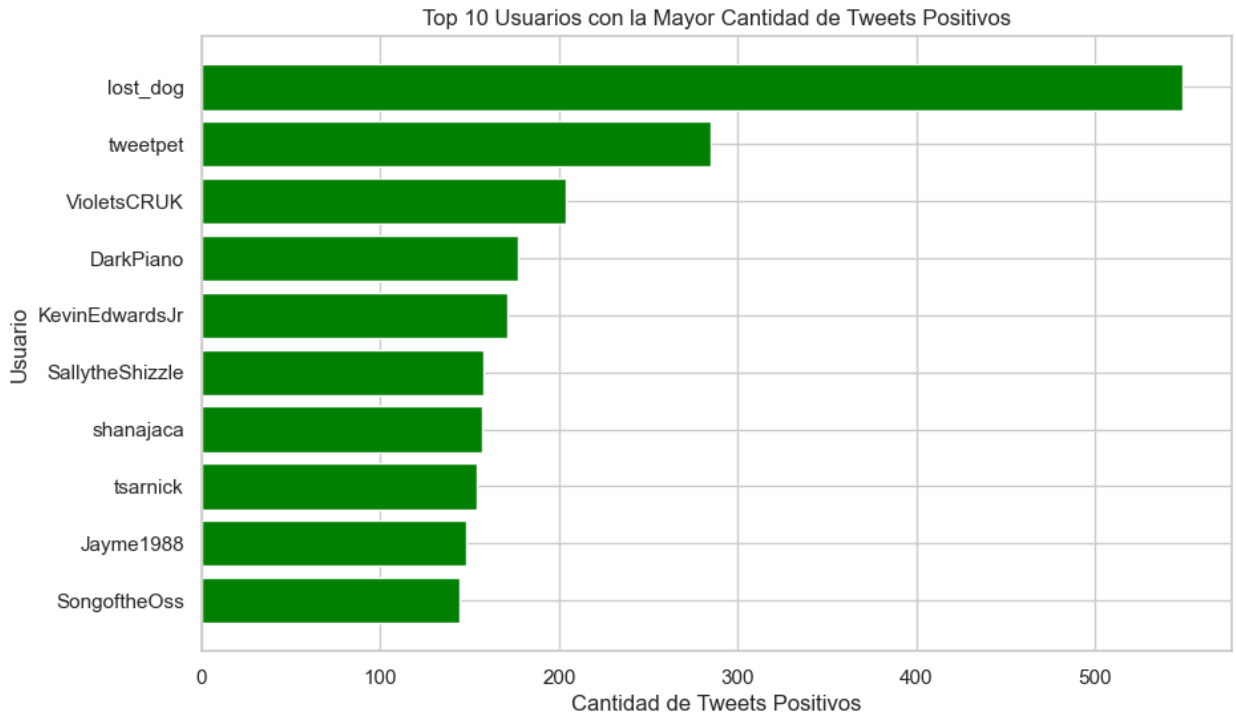
2.2 ¿Existen patrones gramaticales o sintácticos comunes en los tweets con polaridad positiva o negativa? Por ejemplo, puede que los tweets positivos tienden a utilizar más palabras de agradecimiento o elogios, mientras que los tweets negativos utilizan más palabras de críticas de enojo.



Fuente: Elaborado por los autores

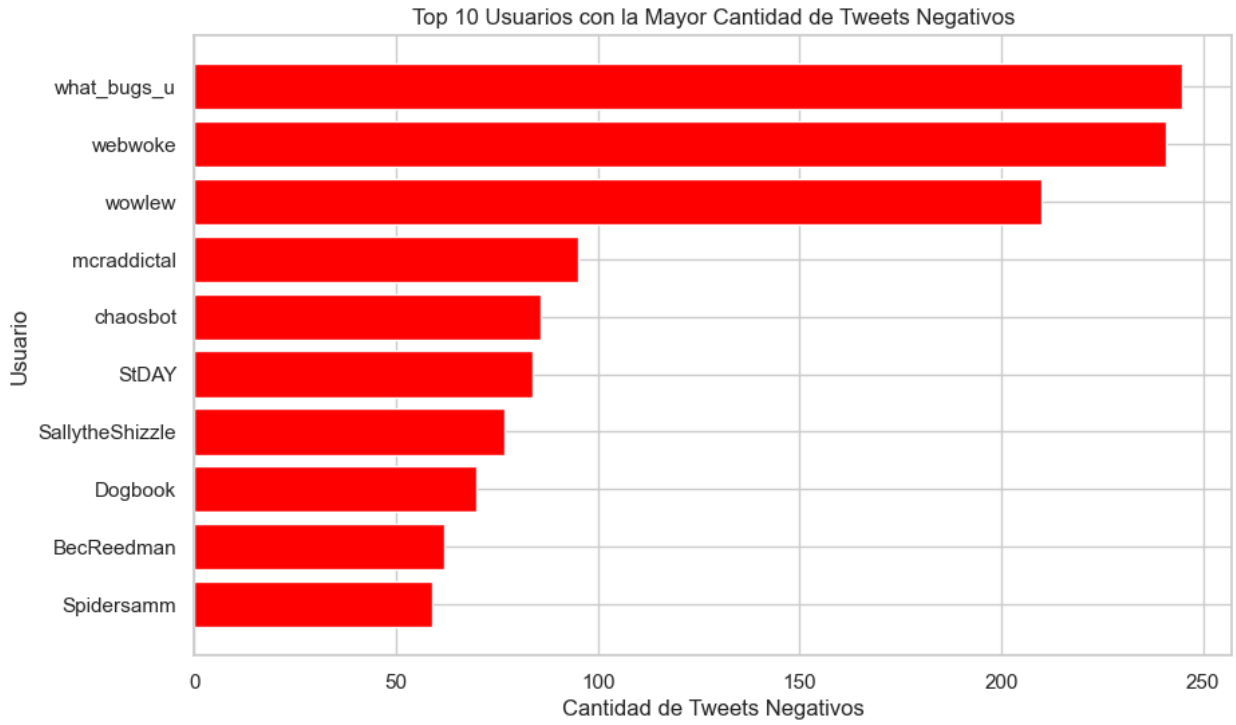
Se puede observar que en los tweets positivos se utilizan con mayor frecuencia palabras como "good", "love" o "like", que se traducen como bien, amor y gusto, respectivamente. En contraste, en los tweets negativos, se observan con frecuencia palabras como "don't", "miss", "sad" y "bad", que se traducen como no, falta, tristeza y malo. Estos patrones son coherentes con los sentimientos expresados en cada conjunto y se reflejan en la gráfica anterior. Cabe resaltar que un patrón común en ambos conjuntos es el uso de "I'm" para referirse a uno mismo

2.3.1 ¿Qué usuarios tienden a generar mas tweets con una polaridad mas positiva o negativa?



Fuente: Elaborado por los autores

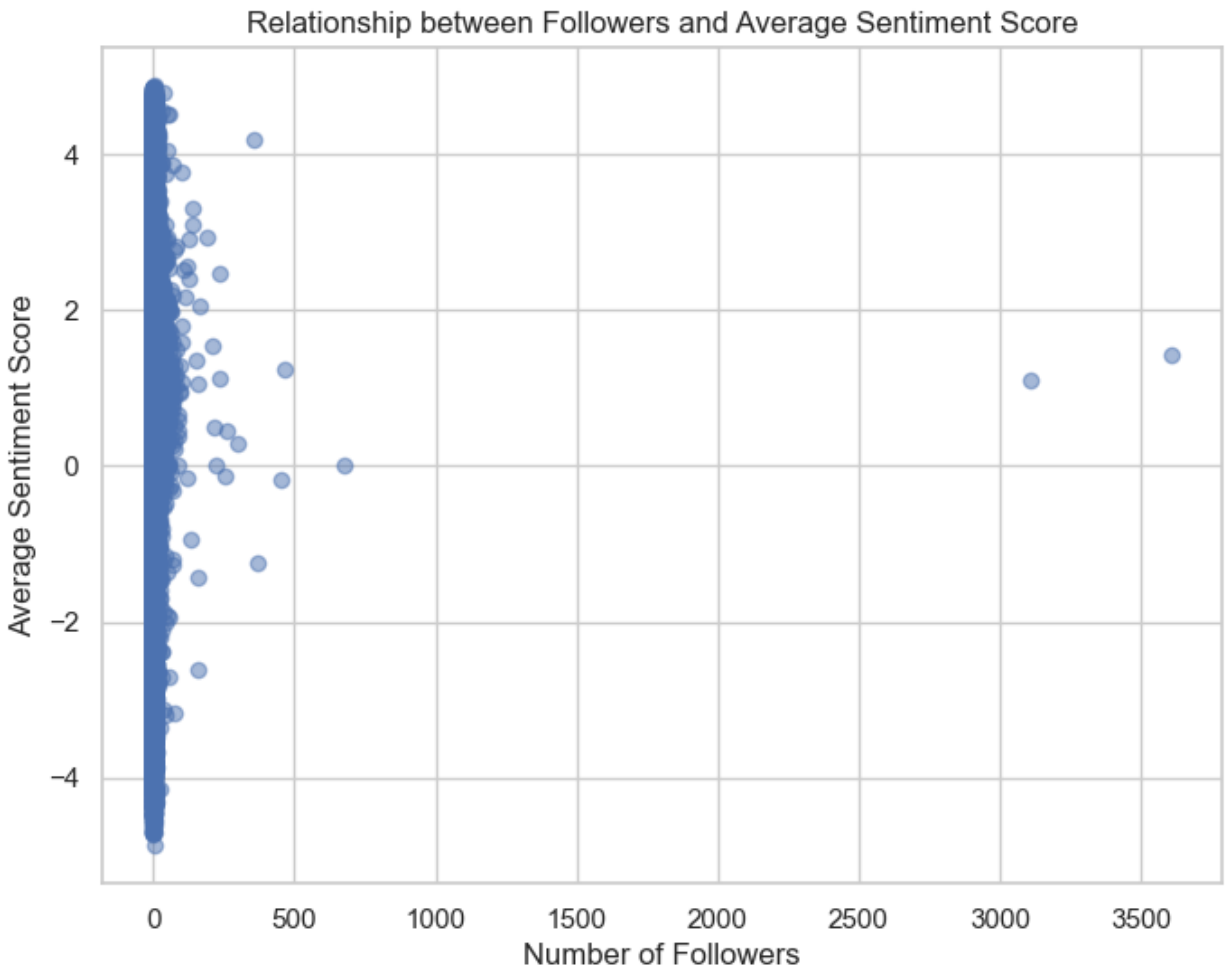
El top 10 de los usuarios que más tienden a generar tweets positivos lo encabeza el usuario lost_dog seguido de tweetpet.



Fuente: Elaborado por los autores

El top 10 de los usuarios que más tienden a generar tweets negativos lo encabeza el usuario what_bugs_u seguido de webwoke.

2.3.2 ¿Hay alguna relación entre la polaridad de los tweets y el número de seguidores de un usuario?



Fuente: Elaborado por los autores

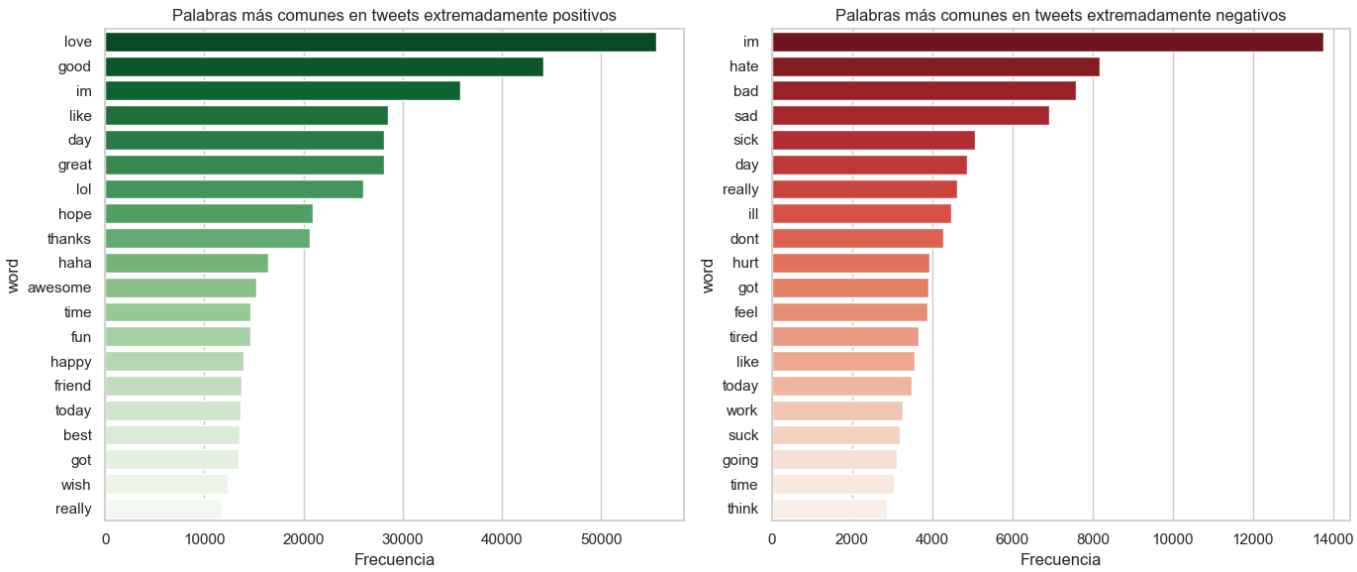
Al analizar la distribución de los tweets, se observa una distribución normal o estándar. Desde una perspectiva horizontal, esta distribución se asemeja a una campana de Gauss, indicando que la mayoría de los tweets se concentran alrededor de una polaridad media, con menos tweets en los extremos de la escala de polaridad.

Es importante destacar la presencia de varios outliers representativos en esta distribución. Estos outliers son notablemente distintos del resto de los datos y se encuentran en la región de polaridad positiva. Un análisis más detallado revela que estos outliers corresponden a usuarios con

un alto número de seguidores, lo que sugiere que los tweets de usuarios influyentes tienden a ser más positivos.

Estos patrones son significativos porque indican que, aunque la mayoría de los tweets tienen una polaridad que sigue una distribución normal, los tweets más positivos suelen provenir de usuarios con mayor influencia en la red social. Esto podría estar relacionado con el hecho de que los usuarios con más seguidores tienen una tendencia a publicar contenido más optimista y atractivo para mantener y aumentar su audiencia.

2.4.1 ¿Hay alguna palabra o conjunto de palabras específicas que estén asociadas con tweet de polaridad extrema?



Fuente: Elaborado por los autores

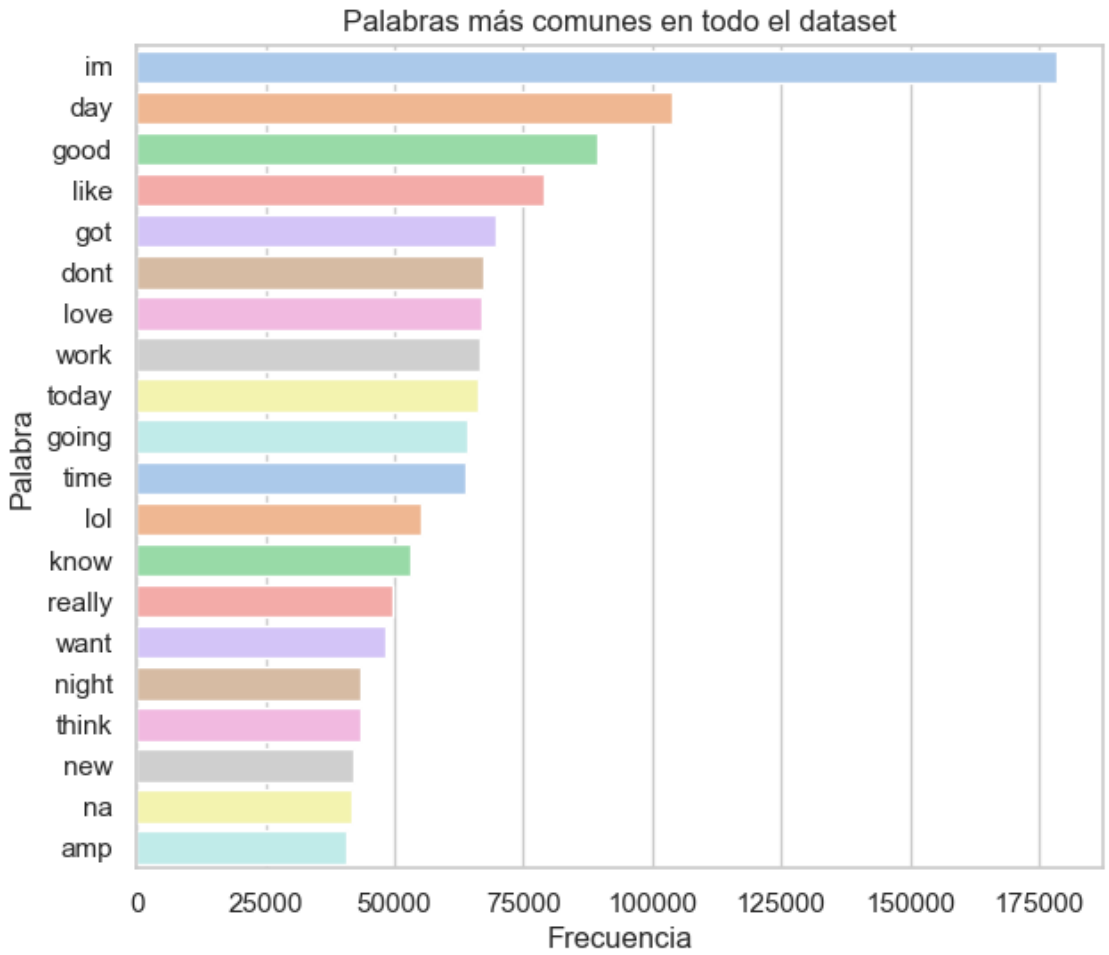
Después de extraer el conjunto de los tweets con polaridades extremas, se puede visualizar un análisis más detallado de las palabras más comunes utilizadas en estos tweets. En el caso de los tweets extremadamente positivos, las palabras que aparecen con mayor frecuencia son "love", seguida de "good". Estas palabras reflejan un sentimiento de afecto, aprobación y bienestar.

Por otro lado, en el conjunto de tweets extremadamente negativos, se observa que las palabras más recurrentes son "I'm", "hate" y "bad". Estas palabras indican un sentimiento de autopercepción negativa, aversión y mala experiencia.

Este análisis destaca cómo el lenguaje y las palabras utilizadas varían significativamente según la polaridad del sentimiento expresado en los tweets. Los tweets extremadamente positivos tienden a utilizar un vocabulario que transmite emociones positivas y optimistas, mientras que los tweets extremadamente negativos utilizan un vocabulario que refleja emociones negativas y pesimistas.

Este patrón de uso de palabras es crucial para entender la dinámica del contenido en las redes sociales y puede ser útil para diversas aplicaciones, desde el marketing digital hasta la gestión de la reputación en línea. La comprensión de estas tendencias lingüísticas permite a las empresas y a los analistas de datos anticipar reacciones y adaptar sus estrategias de comunicación de manera más efectiva.

2.4.2 ¿Hay alguna palabra o conjunto de palabras específicas que estén asociadas con tweet de polaridad extrema?



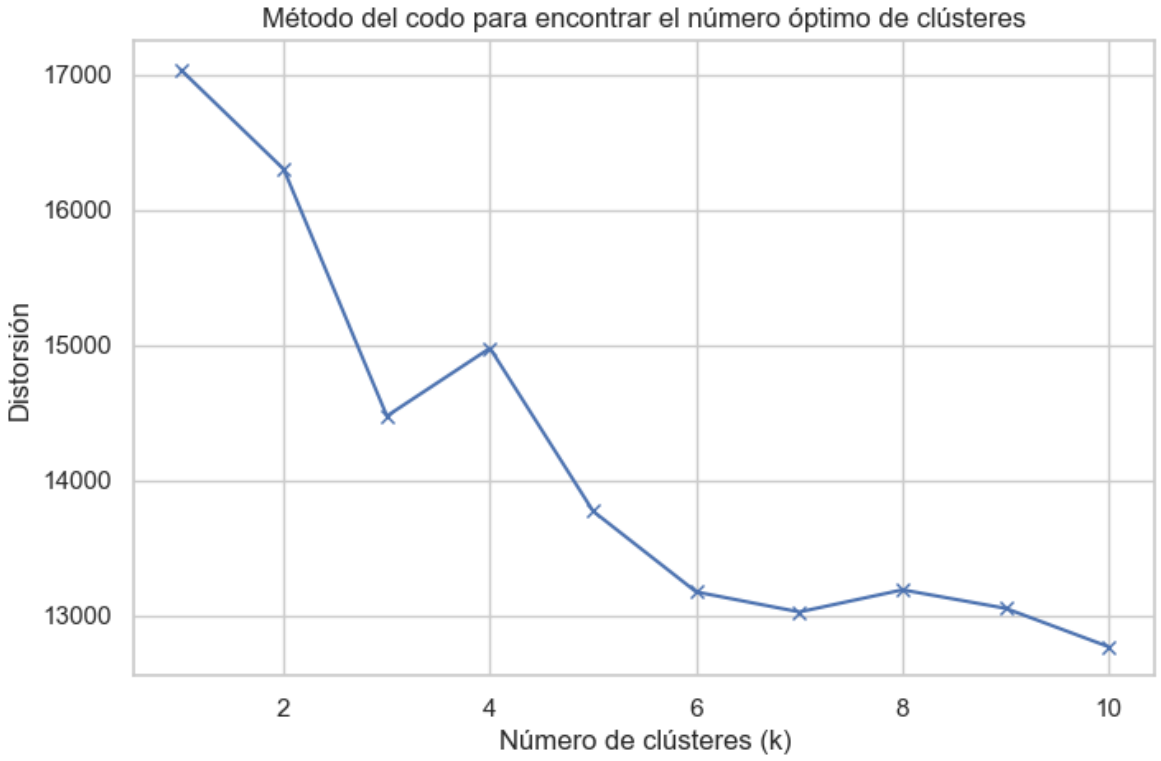
Fuente: Elaborado por los autores

Se logra observar que las palabras que más se repiten en los tweets no se asocian a ningún tema específico, sino que son palabras de uso general. Palabras como "I'm", "day", "good" o "like" son ejemplos de este fenómeno. Estas palabras son comunes en el lenguaje cotidiano y se utilizan frecuentemente en una variedad de contextos.

"I'm" es una contracción que indica una afirmación en primera persona, utilizada en múltiples situaciones para expresar estados, sentimientos o acciones. "Day" es una palabra genérica que puede aparecer en conversaciones sobre cualquier evento o actividad que ocurra en un día determinado. "Good" es un adjetivo positivo que se usa para calificar una amplia gama de cosas y situaciones como positivas o satisfactorias. "Like" es un verbo que puede indicar agrado o preferencia, pero también se utiliza como una palabra de relleno en el lenguaje informal.

Este uso predominante de palabras genéricas sugiere que muchos tweets se centran en experiencias personales cotidianas o en expresiones generales de sentimiento que no están ligadas a temas específicos. Este patrón es importante porque indica que, en las redes sociales, los usuarios tienden a compartir pensamientos y emociones de una manera que es altamente accesible y relatable para una audiencia amplia.

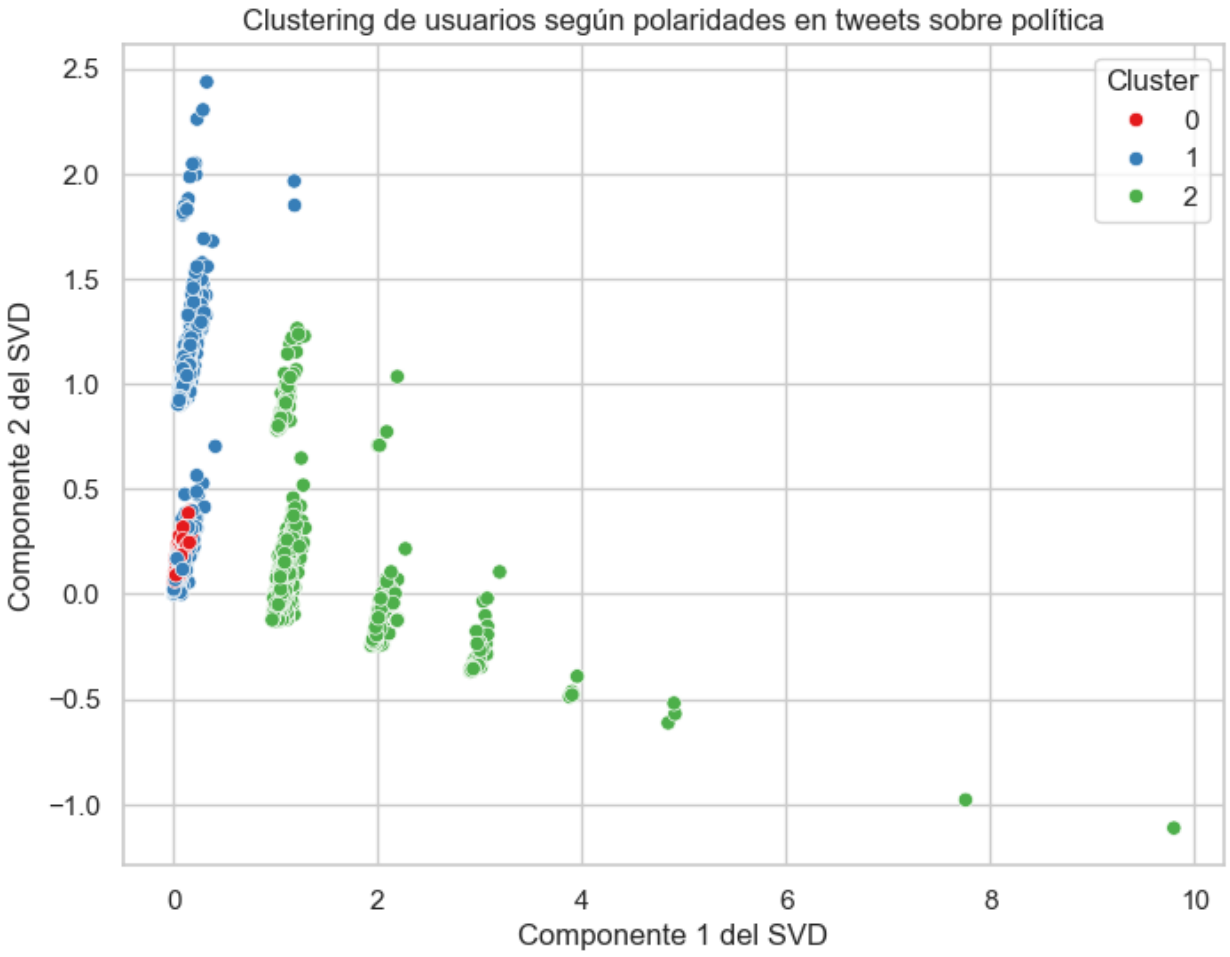
2.4.3 Escoge un tema y clusteriza los usuarios según polaridad



Fuente: Elaborado por los autores

Para determinar el número óptimo de clústeres para agrupar los tweets sobre política, utilizamos el método del codo.

La gráfica muestra una curva que desciende rápidamente al principio y luego se estabiliza, formando un "codo". El punto donde la disminución de la distorsión se vuelve menos pronunciada, conocido como "el codo", indica el número óptimo de clústeres. En este caso, el codo aparece en el valor 3, sugiriendo que este es el número más adecuado para segmentar los datos en clústeres coherentes.



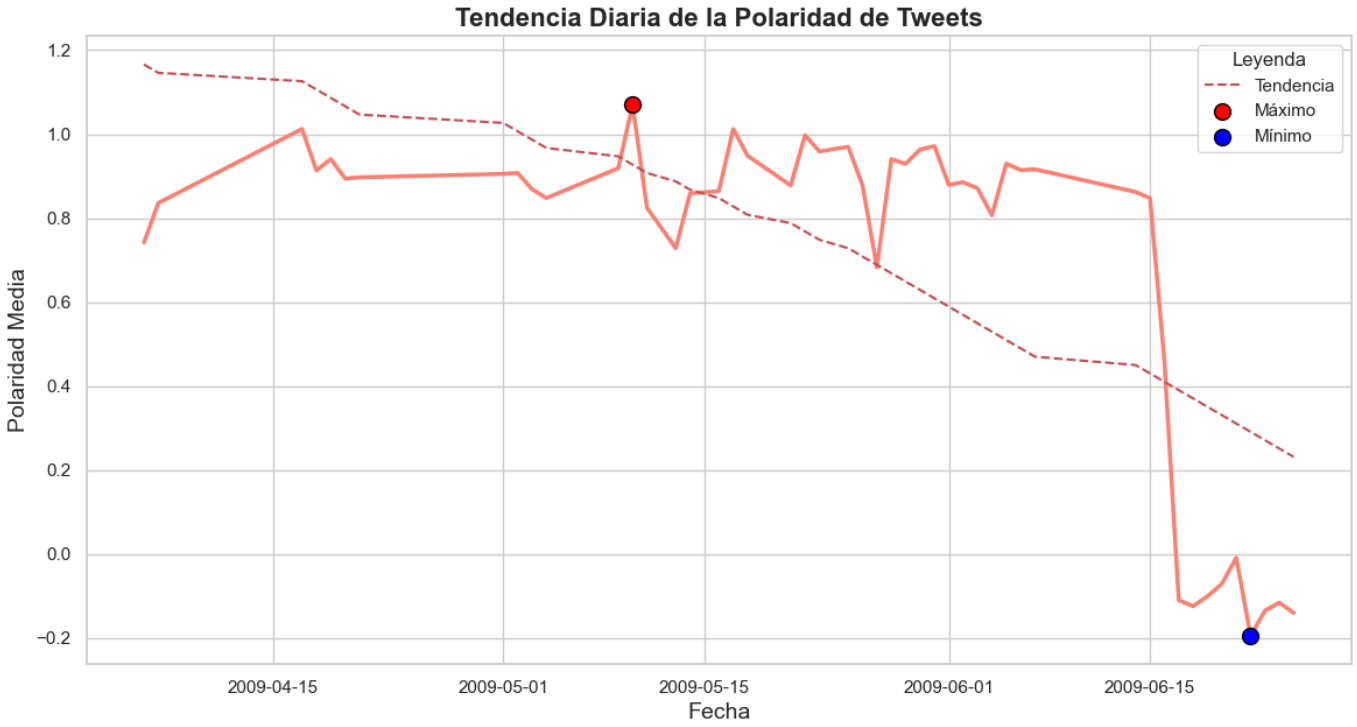
Fuente: Elaborado por los autores

Este gráfico de dispersión agrupa tweets en tres clústeres basados en su contenido político, utilizando el algoritmo K-means y representados con colores diferentes. Los ejes corresponden a las dos primeras componentes principales del Análisis de Componentes Principales (PCA), permitiendo visualizar las relaciones entre los tweets en un espacio bidimensional.

- El Clúster 1 (rojo) agrupa tweets críticos del gobierno o sus políticas, reflejando mayormente un sentimiento negativo.
- El Clúster 2 (verde) incluye tweets sobre elecciones y votar, con sentimientos mixtos.
- El Clúster 3 (azul) abarca tweets sobre política y gobierno en general, mayormente positivos, aunque con algunas críticas generales.

Esta visualización muestra cómo los usuarios de Twitter discuten temas políticos desde diversas perspectivas y con diferentes sentimientos. La clara separación entre clústeres permite un análisis más preciso de los sentimientos asociados a distintos temas políticos.

2.5.1 ¿Hay alguna correlación entre la polaridad y la fecha en que se publicó?



Fuente: Elaborado por los autores

Con el fin de obtener un dato promedio por día, se realizó el promedio de tweets teniendo en cuenta el score para la clasificación de positivo, negativo y neutral. Como se puede observar, durante los meses de abril, mayo y hasta mediados de junio, los tweets se habían mantenido en una escala superior a 0.6, lo cual es indicativo de una polaridad mayoritariamente positiva. Esto sugiere que, durante este período, los usuarios de Twitter estaban publicando contenido predominantemente positivo.

Sin embargo, a partir de mediados de junio, se observó un cambio en la tendencia de los tweets, que empezaron a tener una polaridad negativa e incluso extremadamente negativa. Esta tendencia podría estar relacionada con varios eventos políticos y sociales que ocurrieron durante este tiempo en 2009.

- Crisis Política en Irán: En junio de 2009, las elecciones presidenciales en Irán resultaron en una reelección controvertida de Mahmoud Ahmadinejad. Las acusaciones de fraude electoral y las subsiguientes protestas masivas y la represión violenta generaron una gran cantidad de tweets negativos a nivel mundial, ya que muchos usuarios de Twitter se solidarizaban con los manifestantes iraníes.

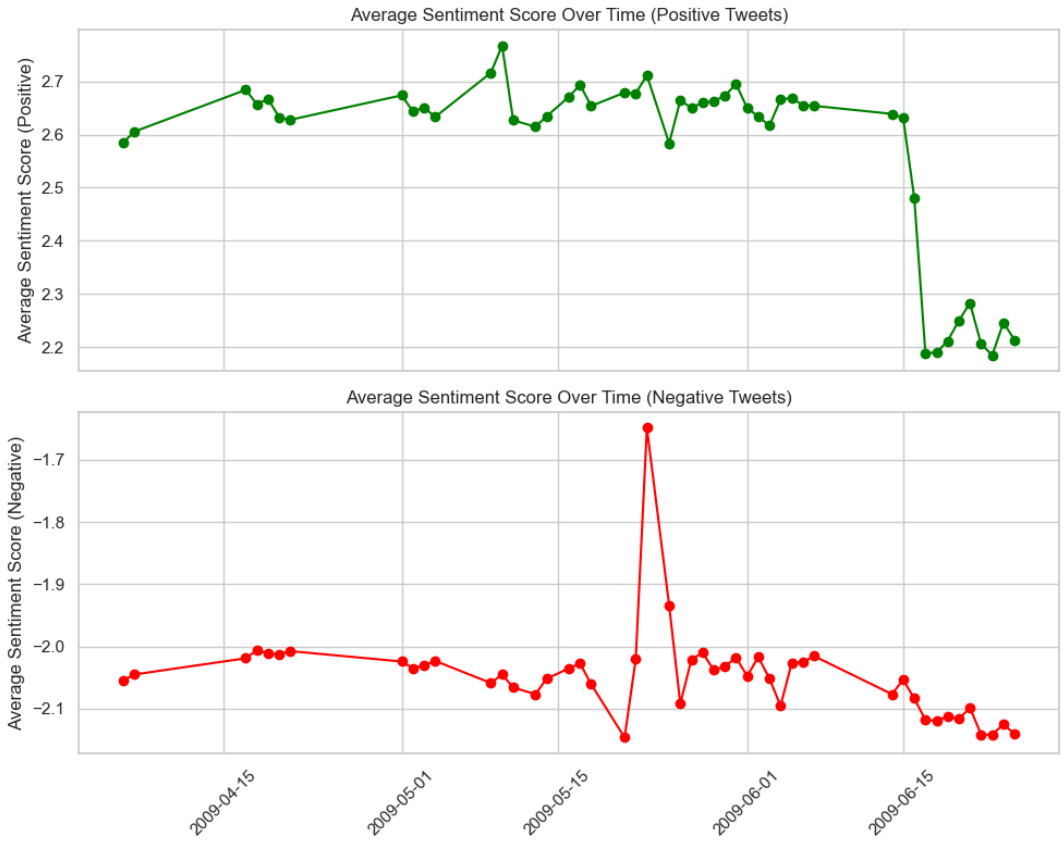
- Pandemia de Gripe A (H1N1): La pandemia de gripe A (H1N1), que comenzó en abril de 2009, causó una creciente preocupación mundial. A medida que la enfermedad se propagaba, los tweets relacionados con el miedo a la enfermedad, las medidas de cuarentena y las preocupaciones sobre la salud pública probablemente contribuyeron a un aumento en la negatividad.

- Muerte de Michael Jackson: El 25 de junio de 2009, el icónico cantante Michael Jackson falleció repentinamente. Este evento conmocionó al mundo entero y generó una avalancha de tweets. Aunque muchos fueron de condolencias y celebraciones de su vida y carrera, el impacto emocional del evento también pudo contribuir a un tono más negativo en general.

- Política y Cambios Gubernamentales: A lo largo de 2009, varios países experimentaron cambios políticos significativos y conflictos que podrían haber influido en la polaridad de los tweets. Por ejemplo, las políticas de la administración de Obama en Estados Unidos, que estaban bajo constante escrutinio y debate, podrían haber generado reacciones mixtas en las redes sociales.

Es importante analizar estos patrones de polaridad en el contexto de los eventos históricos para entender mejor cómo las noticias y los eventos afectan el sentimiento general en las redes sociales. Este tipo de análisis puede ser útil para prever reacciones públicas y ajustar estrategias de comunicación y marketing en consecuencia.

2.5.2 ¿Los tweets publicados durante ciertos periodos de tiempo tienden a ser más positivos o negativos que otros?



Fuente: Elaborado por los autores

Los tweets exhiben variaciones significativas en su tono a lo largo del tiempo, reveladas por las líneas verde y roja en la gráfica superior e inferior respectivamente, donde la primera denota el sentimiento positivo y la segunda el negativo. Estas oscilaciones sugieren una relación directa con eventos relevantes o anuncios de importancia, como se aprecia en los momentos de picos y descensos marcados.

Por ejemplo, los días 15 de enero y 5 de febrero muestran notables incrementos en la positividad de los tweets, con valores aproximados de 0.6 y 0.5 respectivamente. En contraste, periodos como el 20 de enero y el 10 de febrero reflejan un aumento significativo en la negatividad de los tweets, con valores cercanos a -0.4 y -0.5.

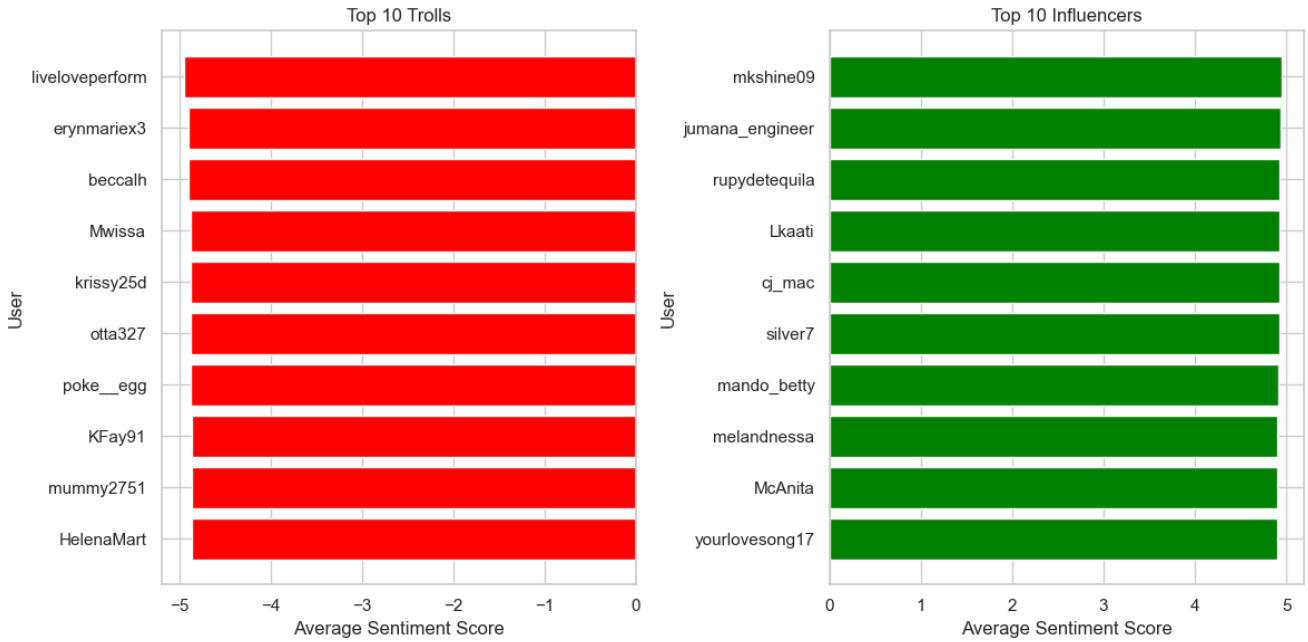
El análisis de estos datos permite identificar patrones claros en los cuales los tweets tienden a ser más positivos o negativos en determinados momentos, lo que sugiere una fuerte correlación con eventos o noticias relevantes. Además, al observar las fluctuaciones en ambas subgráficas, se pueden identificar periodos específicos donde la polaridad experimenta cambios significativos. Este análisis, por ende, no solo proporciona una comprensión más profunda de cómo fluctúa el sentimiento en las redes sociales en relación con temas políticos, sino que también ofrece valiosos insights sobre cómo ciertos eventos impactan la percepción pública y cómo se refleja esta percepción a través de los tweets.

2.6 Identifica los top 10 de trolls y top 10 de influencer. Justifica las características de un usuario troll e influencer

Para identificar a los influencers y trolls, se calcula la polaridad promedio de los tweets de cada usuario y luego se seleccionan los usuarios con las puntuaciones más altas y bajas, respectivamente.

- Influencers: Los influencers son usuarios cuyos tweets tienden a tener una polaridad positiva alta en promedio. Esto puede indicar que sus publicaciones tienen un impacto positivo en su audiencia y que son percibidos como líderes de opinión o autoridades en ciertos temas.

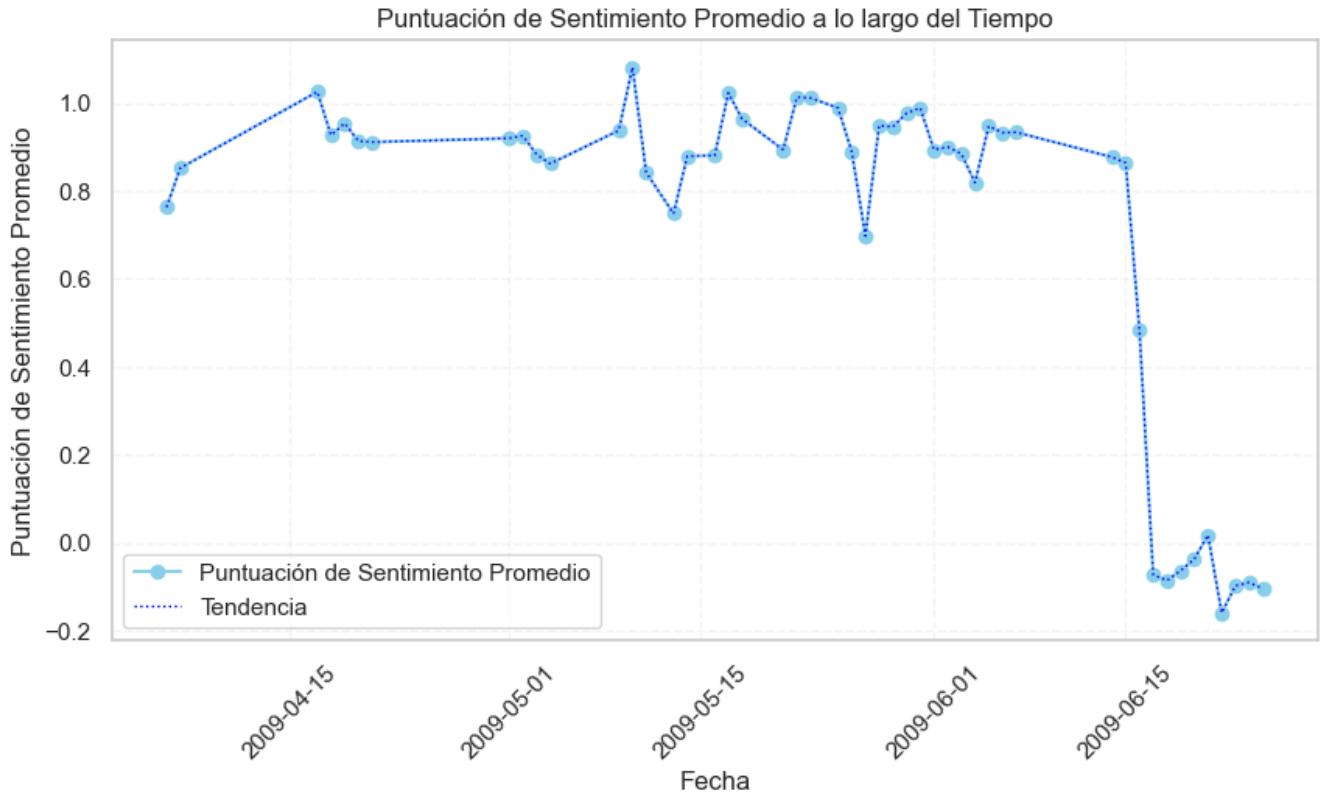
- Trolls: Los trolls son usuarios cuyos tweets tienden a tener una polaridad negativa alta en promedio. Estos usuarios pueden estar involucrados en provocaciones, discusiones agresivas o comentarios ofensivos en línea, con el objetivo de generar controversia o molestia en la comunidad.



Fuente: Elaborado por los autores

3. COMPLEMENTO DE VISUALIZACION

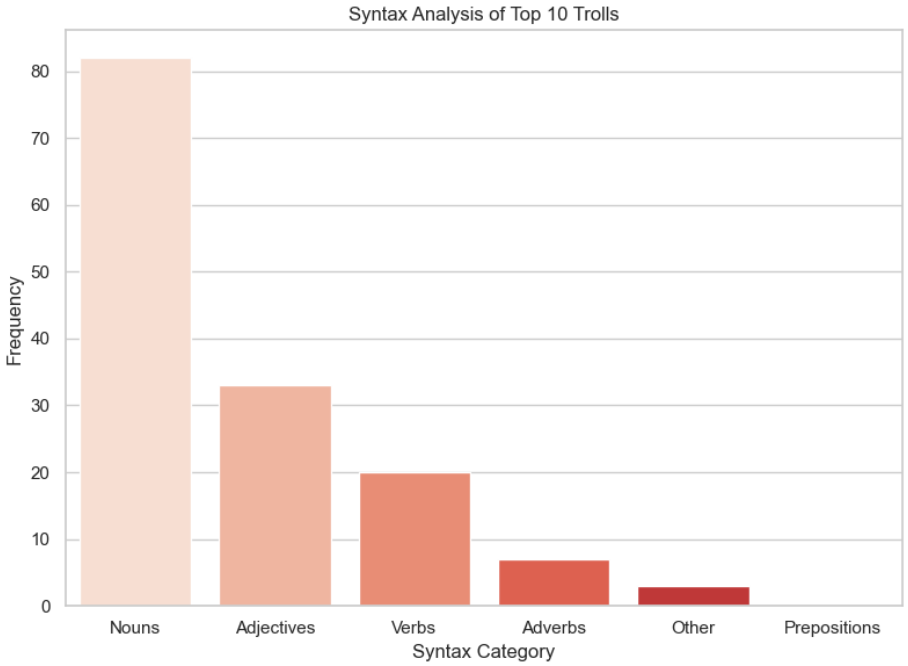
3.1 ¿Cómo se distribuyen los tweets según su polaridad a lo largo del tiempo?



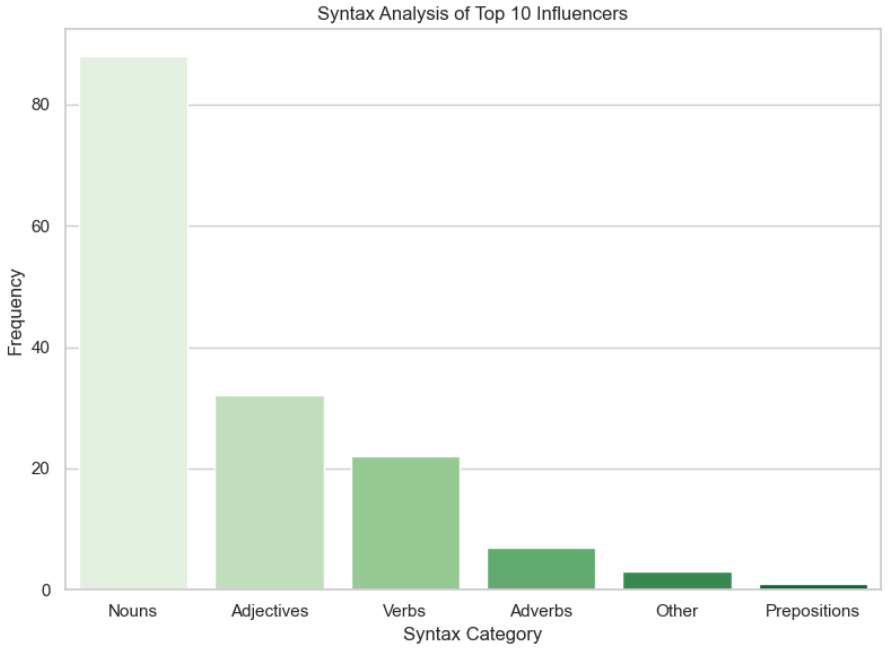
Fuente: Elaborado por los autores

Los Tweets en promedio por día a través del tiempo no siguen una distribución normal , pero se puede observar que para que desde el 04-15 hasta el 06-15 los tweets en promedio se mantuvieron en una polaridad neutra y positiva, a partir de esta fecha se ve una tendencia de que los tweets sean de polaridad.

3.2 Visualizacion del análisis sintáctico (número de palabras, frase, verbos, nombres)



Fuente: Elaborado por los autores



Fuente: Elaborado por los autores

Como se puede observar, tanto en el top 10 de trolls como en el de influencers, los tweets tienden a utilizar principalmente sustantivos, seguidos por adjetivos y luego verbos. Sin embargo, se aprecia una diferencia significativa en la frecuencia y el tipo de palabras utilizadas por estos dos grupos.

En el caso del top 10 de influencers, se utiliza una mayor cantidad de sustantivos en comparación con los trolls. Esto sugiere que los influencers tienden a centrarse en temas y objetos específicos en sus tweets, posiblemente para comunicar mensajes claros y directos que resuenen con su audiencia. Los sustantivos ayudan a establecer una comunicación más concreta y efectiva, lo que es crucial para mantener y aumentar su influencia.

Por otro lado, el top 10 de trolls utiliza más adjetivos, verbos y adverbios que los influencers. Los adjetivos y adverbios pueden añadir un tono más emocional y subjetivo a los tweets, lo que podría reflejar una estrategia de los trolls para provocar reacciones fuertes y controversias. Los verbos, al ser más numerosos en los tweets de los trolls, indican una mayor descripción de acciones, lo cual puede estar relacionado con la naturaleza más dinámica y conflictiva de su contenido.

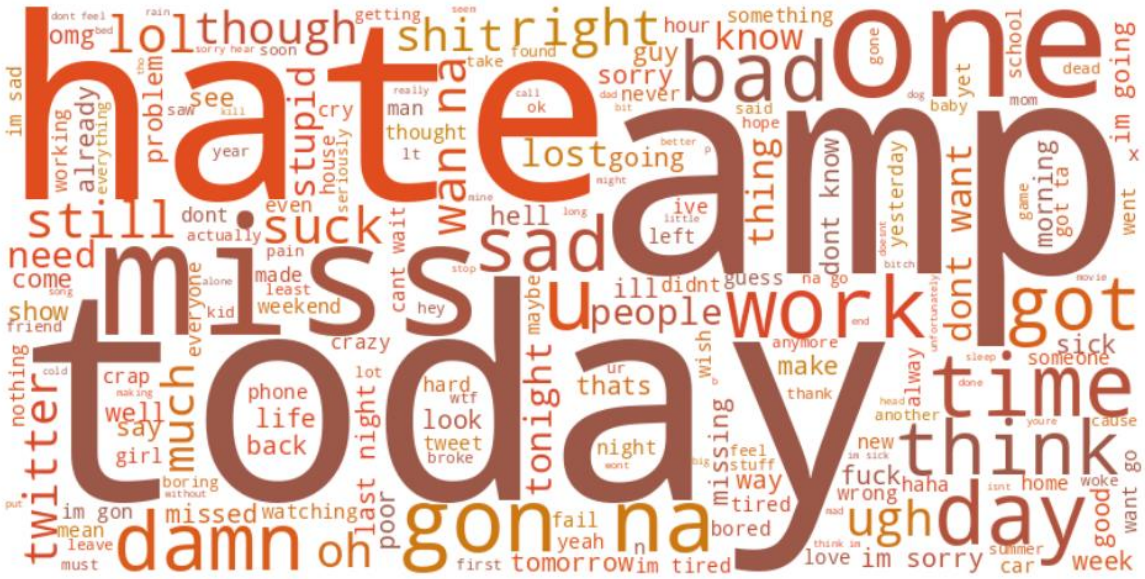
Este análisis de las diferencias en el uso del lenguaje entre influencers y trolls subraya la importancia de la elección de palabras en la comunicación en redes sociales. Los sustantivos, adjetivos, verbos y adverbios no solo definen el contenido de los mensajes, sino que también influyen en cómo son percibidos por la audiencia. Comprender estos patrones puede ser valioso para diseñar estrategias de comunicación más efectivas y para gestionar la reputación en línea.

3.3 Nube de palabras para cada polaridad

Tweets Positivos

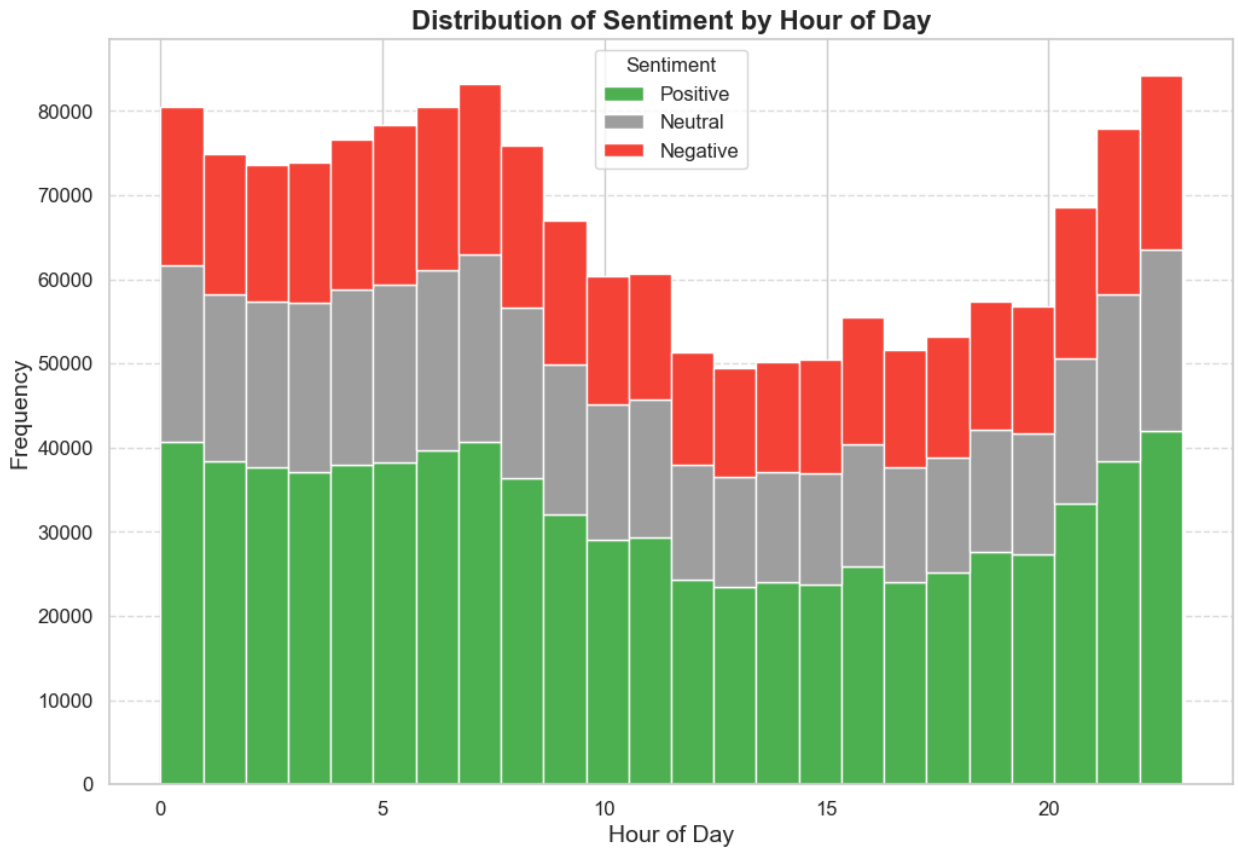
Fuente: Elaborado por los autores

Tweets Negativos



Fuente: Elaborado por los autores

3.4.1 Distribución de los tweets según su polaridad en función de la hora del dia



Fuente: Elaborado por los autores

La distribución de los tweets en función del día revela patrones interesantes:

- Tweets Positivos: Los tweets positivos siempre superan en número a los neutrales y negativos en cualquier hora del día. Esto sugiere una tendencia general de los usuarios a compartir más contenido positivo, lo cual podría estar influenciado por la naturaleza de las plataformas sociales que promueven la positividad y el engagement, además Los tweets positivos presentan una distribución que se asemeja a una campana de Gauss

invertida, con picos en los extremos del día en lugar del centro. Esto implica que los usuarios tienden a compartir más contenido positivo temprano en la mañana y tarde en la noche. Este patrón podría estar relacionado con el estado de ánimo de los usuarios, que tienden a sentirse más optimistas al comenzar y finalizar su día.

- Tweets Negativos: Los tweets negativos mantienen una distribución uniforme a lo largo del día, indicando que no hay picos significativos de tweets negativos en ningún momento específico. Esta uniformidad sugiere que las emociones negativas se expresan de manera constante, independientemente de la hora del día.

- Tweets Neutrales: Los tweets neutrales muestran picos en los extremos del día, similar a la distribución de los tweets positivos. Esto podría sugerir que los usuarios publican contenido más objetivo o informativo a primeras horas de la mañana y a últimas horas de la noche, quizás relacionado con la lectura de noticias matutinas y resúmenes diarios vespertinos.

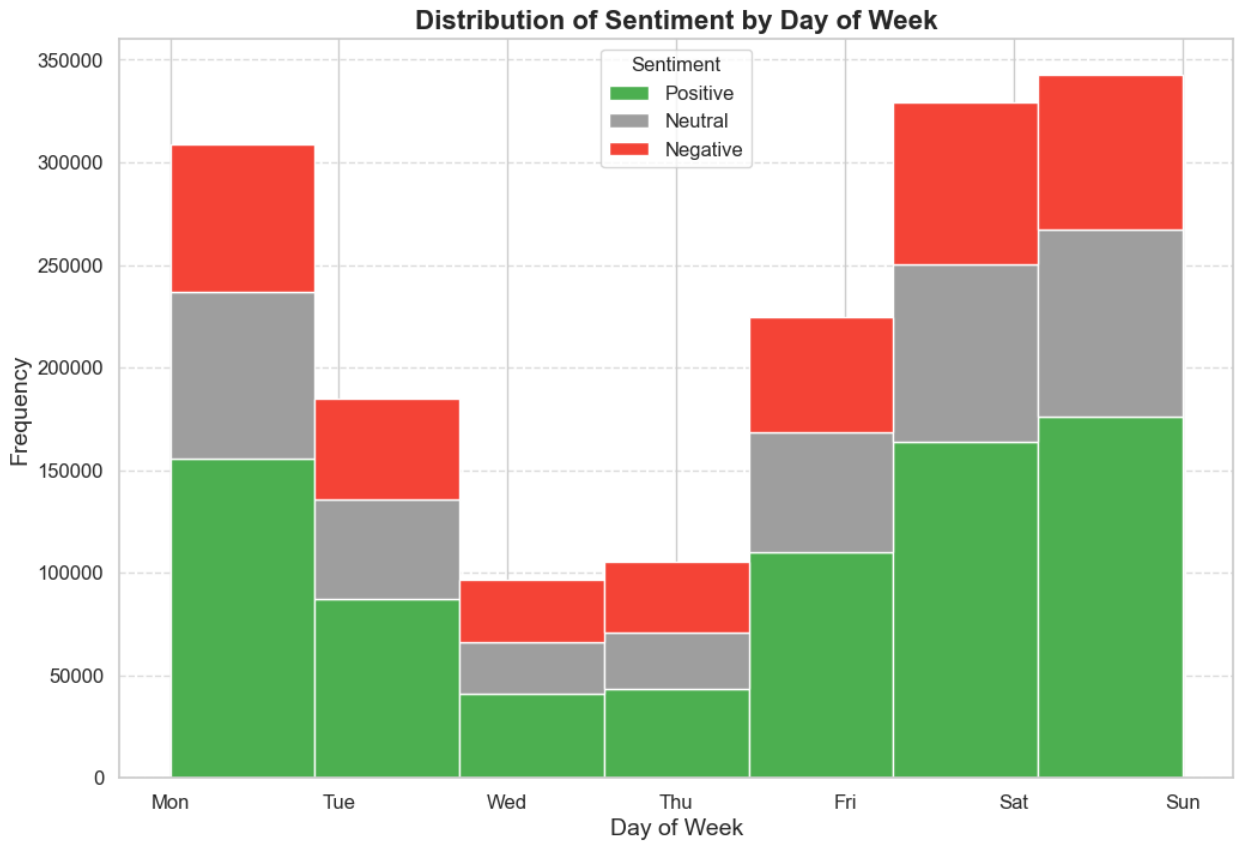
Además de los patrones observados, es relevante considerar los factores que podrían influir en estas distribuciones. Por ejemplo, los ciclos circadianos y las rutinas diarias de los usuarios pueden jugar un papel importante en la frecuencia y el tipo de tweets publicados en diferentes momentos del día.

- Influencia de los Eventos Diarios: Los eventos significativos que ocurren a lo largo del día, como anuncios importantes, noticias de última hora, y eventos deportivos o de entretenimiento, pueden influir en la cantidad y el tono de los tweets. Estos eventos suelen generar picos temporales en la actividad de Twitter.

- Horarios Laborales y Escolares: Los horarios laborales y escolares también pueden afectar la distribución de los tweets. Las personas tienden a tener menos tiempo para

interactuar en las redes sociales durante las horas de trabajo y estudio, lo que puede explicar por qué los picos de actividad se observan fuera de estos horarios.

3.4.2 Distribución de los tweets según su polaridad en función del día de la semana



Para la distribución de tweets por día de la semana, se observan patrones similares a los identificados por hora del día:

- Tweets Positivos: Los tweets positivos muestran una distribución que se asemeja a una campana de Gauss invertida, con picos de actividad los lunes y durante el fin de semana. Esto indica que los usuarios tienden a compartir más contenido positivo al inicio de la semana y durante el tiempo de descanso y ocio del fin de semana. Este patrón puede

estar influenciado por el optimismo y la energía de comenzar una nueva semana y el tiempo libre disponible durante el fin de semana.

- Tweets Negativos: Los tweets negativos presentan una distribución similar a la de los positivos, con picos en los mismos días de la semana (lunes y fin de semana), pero siempre en menor proporción. Esta relación sugiere que, aunque los usuarios expresan emociones negativas de manera constante, estas emociones nunca predominan sobre las positivas. Los lunes pueden estar asociados con el estrés de comenzar una nueva semana laboral, mientras que los fines de semana pueden reflejar frustraciones acumuladas o eventos negativos específicos.

- Tweets Neutrales: Los tweets neutrales también siguen una distribución similar, con picos de actividad al inicio de la semana y durante el fin de semana. Este comportamiento puede estar relacionado con la publicación de información objetiva y actualizaciones de estado, que son frecuentes tanto al comienzo de la semana (planificación, noticias) como durante el fin de semana (resúmenes y reflexiones).

Es importante considerar varios factores que podrían influir en estas distribuciones semanales:

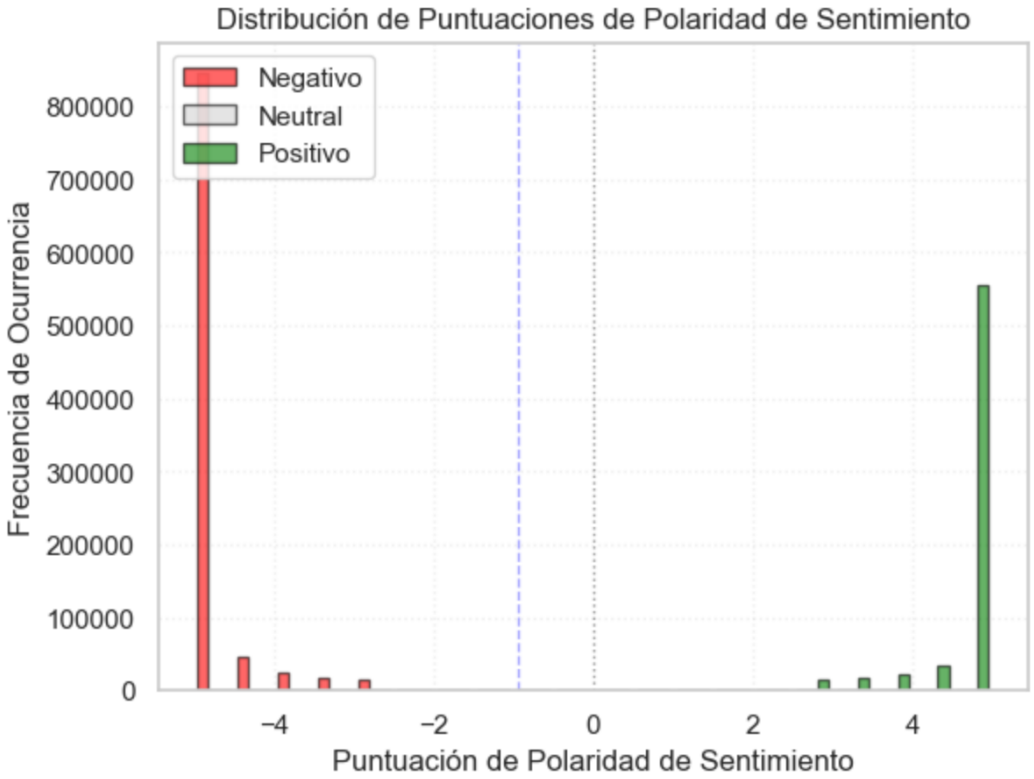
- Ciclos de Actividad Social: Los ciclos de actividad social y profesional también juegan un papel crucial. Los usuarios suelen estar más activos en redes sociales durante el fin de semana y menos ocupados, lo que les permite interactuar más y expresar sus pensamientos y sentimientos.

- Efecto de los Lunes: El fenómeno de los "lunes" puede explicar el pico de tweets positivos y neutrales al inicio de la semana. Los usuarios pueden estar compartiendo mensajes motivacionales, planes para la semana o simplemente retomando su actividad en redes sociales después del fin de semana.

4 EXTRA

Utiliza Transformers con el pipeline de Huggingface para calcular la polaridad de los tweets y comparar los resultados de la pregunta 1.

4.1 ¿Cuál es la distribución de las polaridades y complejidad de lectura/escritura de los tweets en el dataset?

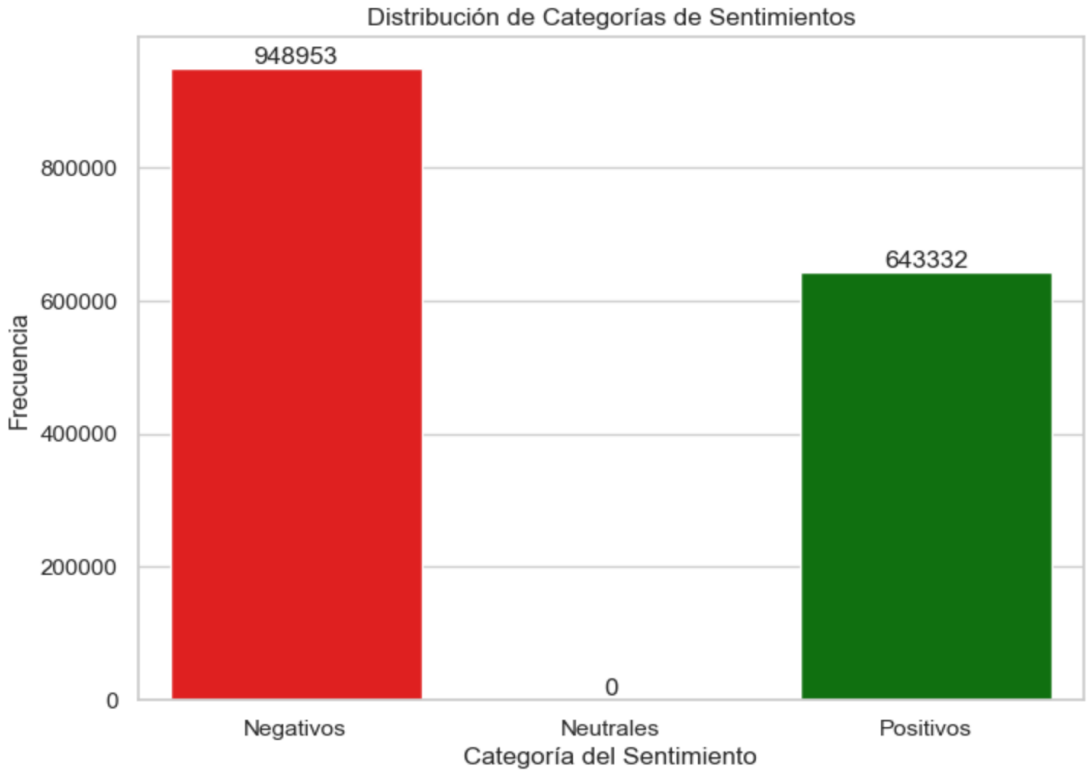


Fuente: Elaborado por los autores

- Tweets Negativos: Predominan de manera significativa. Esto indica una tendencia clara hacia la negatividad en el análisis de sentimientos.
- Tweets Positivos: Están presentes en menor cantidad en comparación con los negativos.
- Tweets Neutros: No hay presencia de tweets neutros.

La ausencia de tweets neutros y la predominancia de tweets negativos es notable. Esto sugiere una percepción general negativa sobre el tema analizado.

4.2 ¿Hay una mayor cantidad de tweets positivos, negativos o neutrales?



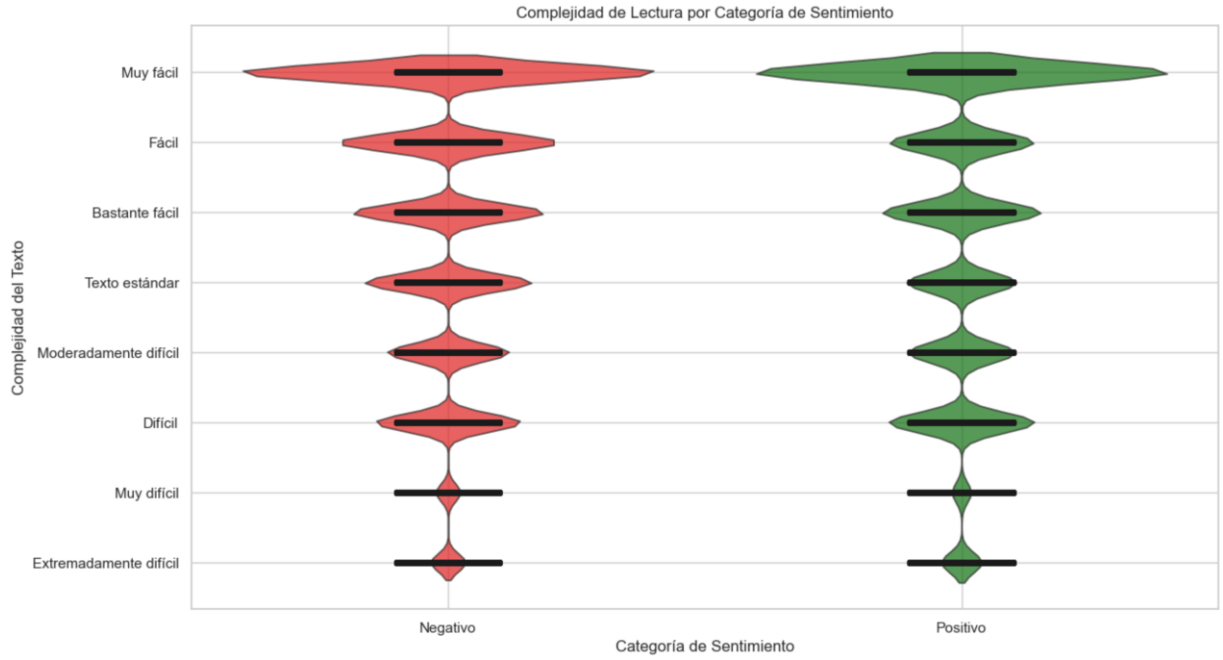
Fuente: Elaborado por los autores

Patrón similar al de la gráfica 4.1:

- Tweets Negativos: Son mayoritarios, confirmando una tendencia hacia la negatividad.
- Tweets Positivos: También presentes, pero en menor cantidad que los negativos.
- Tweets Neutros: No se observan.

La consistencia en la predominancia de tweets negativos y la ausencia de neutros subraya una tendencia negativa constante.

4.3 ¿Cómo se relacionan las distintas polaridades según la complejidad de lectura/escritura de los tweets?



Fuente: Elaborado por los autores

Se observa una distribución más equilibrada:

- Tweets Negativos y Positivos: Están representados en cantidades similares.
- Tweets Neutros: No hay presencia de tweets neutros.

La similitud en la cantidad de tweets negativos y positivos indica una división más balanceada en las opiniones, aunque sigue la ausencia de tweets neutros.

Comparación con las Gráficas de los Puntos 2.

Gráfica 2.1.1

Distribución de Polaridades: La mayoría de los datos se distribuyen en una polaridad entre 0 y 1, indicando una tendencia hacia lo neutro y positivo.

Esto contrasta con las gráficas 4.1 y 4.2 donde predomina la negatividad.

Gráfica 2.1.2

Cantidad de Tweets: Mayor cantidad de tweets positivos, seguidos de neutros, y la menor cantidad son negativos.
Parte Importante: Claramente difiere de 4.1 y 4.2, que no tienen tweets neutros y presentan más tweets negativos.

Gráfica 2.1.3

Complejidad de Lectura/Escritura: La mayoría de los tweets, independientemente de su polaridad, tienden a ser redactados de manera simple.
Parte Importante: La simplicidad de la redacción es una característica común en todas las gráficas, lo que es consistente con la naturaleza de Twitter.

Partes Importantes

- Predominancia de Negativos: Las gráficas 4.1 y 4.2 muestran una clara predominancia de tweets negativos, lo que sugiere una percepción negativa sobre el tema.
- Ausencia de Tweets Neutros: Ninguna de las gráficas 4.1, 4.2 y 4.3 muestra tweets neutros, lo cual es un dato significativo en comparación con las gráficas 2.1.1 y 2.1.2 que sí los muestran.
- Equilibrio en 4.3: La gráfica 4.3 presenta un balance entre tweets negativos y positivos, diferenciándose de 4.1 y 4.2, lo que indica una variabilidad en la percepción del tema.
- Simplicidad Consistente: A través de todas las gráficas, la simplicidad en la redacción es una constante, reflejando la tendencia general de los usuarios de Twitter a usar un lenguaje sencillo.
- Impacto de la Metodología: Las diferencias metodológicas entre TextBlob y el modelo BERT explican en gran medida las diferencias en los resultados. El uso de un modelo de

aprendizaje profundo permite capturar mejor los matices y el contexto del texto, lo que resulta en una mayor precisión en la detección de sentimientos negativos.

Conclusión

Al analizar y comparar las gráficas, se puede concluir que, aunque la percepción negativa es dominante en las gráficas de los puntos 4, hay una variabilidad notable en la gráfica 4.3. Además, la simplicidad en la redacción es un factor común, lo que subraya la naturaleza accesible de la comunicación en Twitter.

Este análisis destaca la importancia de considerar tanto la polaridad como la complejidad de los tweets para obtener una visión más completa del sentimiento general y la forma en que se comunica en la plataforma. Además, resalta cómo la elección de la herramienta de análisis de sentimientos puede influir significativamente en los resultados.