

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

MAESTRÍA EN CIENCIA DE DATOS

**ANÁLISIS DE SENTIMIENTOS SOBRE RESEÑAS DE
NETFLIX**

POR:

ACT. DAMIÁN ATILANO MARTÍNEZ ALVARADO

1. Introducción

Para este estudio, se realizó un análisis exhaustivo de las reseñas de usuarios a la plataforma de *streaming Netflix*. El análisis comenzó con la evaluación de la distribución de las reseñas a lo largo de los años, permitiendo identificar tendencias temporales significativas. Posteriormente, se contabilizó la frecuencia de las palabras en general y también de manera anual, lo cual ayudó a discernir los temas más recurrentes en diferentes períodos. Además, se llevó a cabo un estudio detallado de los *n-grams* más comunes para captar combinaciones de palabras que revelan contextos y temas específicos. Finalmente, se implementó un análisis de sentimientos sobre las reseñas, proporcionando una visión clara del estado emocional de los usuarios con respecto a los contenidos de la plataforma. Este enfoque multifacético permite una comprensión profunda de la percepción y reacción de los usuarios ante el servicio ofrecido por *Netflix*.

2. Descripción de datos

El conjunto de datos consiste en reseñas sobre *Netflix* extraídas de *Kaggle*, incluye el identificador único de cada reseña, el contenido textual y la fecha de publicación. Estos datos son fundamentales para realizar un seguimiento de la respuesta del usuario frente a los contenidos específicos a lo largo del tiempo.

3. Metodología

- Visualización de Datos: Antes de iniciar cualquier análisis detallado, realizamos una visualización general de los datos para entender su estructura y distribución. Esto incluye gráficos y diagramas que muestran las características principales de las reseñas recopiladas. Analizamos cómo se distribuyen las reseñas a lo largo del tiempo para identificar patrones temporales, picos de actividad y posibles anomalías. Este paso es fundamental para comprender cómo los eventos externos o cambios en la plataforma de *Netflix* pueden afectar la frecuencia y el tono de las reseñas.

- Se empleó la biblioteca *NLTK* para la limpieza y normalización del texto, incluyendo la eliminación de caracteres especiales, espacios innecesarios y la estandarización del texto. Este paso asegura la consistencia y calidad del análisis posterior.
- Tokenización y Lematización: Para descomponer el texto en componentes manejables y más analizables, utilizamos la tokenización y la lematización. La tokenización divide el texto en palabras o frases individuales, mientras que la lematización reduce las palabras a su forma base o lema, lo que facilita un análisis lingüístico más preciso.
- Conteo de Frecuencias: Contamos la frecuencia de cada palabra o frase en general y por año en el conjunto de datos para identificar los términos más comunes y destacados en las reseñas. Esto ayuda a discernir los temas o aspectos que son más importantes para los usuarios de *Netflix*.
- Vectorización: Se transformó el texto preprocesado en una forma numérica utilizando técnicas de vectorización basadas en *n-grams*, lo cual fue implementado con la ayuda de la biblioteca *scikit-learn*. Este enfoque permite captar combinaciones de palabras que proporcionan más contexto que los términos individuales y facilita el análisis mediante métodos estadísticos y algoritmos de aprendizaje automático.
- Análisis de Sentimiento: Finalmente, se realizó un análisis de sentimientos clasificando cada reseña en categorías emocionales (positivas, neutras o negativas) utilizando la biblioteca *AFINN*. *AFINN* proporciona un léxico de palabras calificadas con valores que indican la polaridad y la intensidad del sentimiento. Este enfoque permite una evaluación cuantitativa directa de los sentimientos expresados en las reseñas. Realizar este análisis es fundamental para comprender la percepción general de los usuarios hacia el contenido de *Netflix* y para informar decisiones estratégicas que mejoren la experiencia del usuario en la plataforma.

4. Resultados

4.1. Tendencia de publicaciones respecto al tiempo

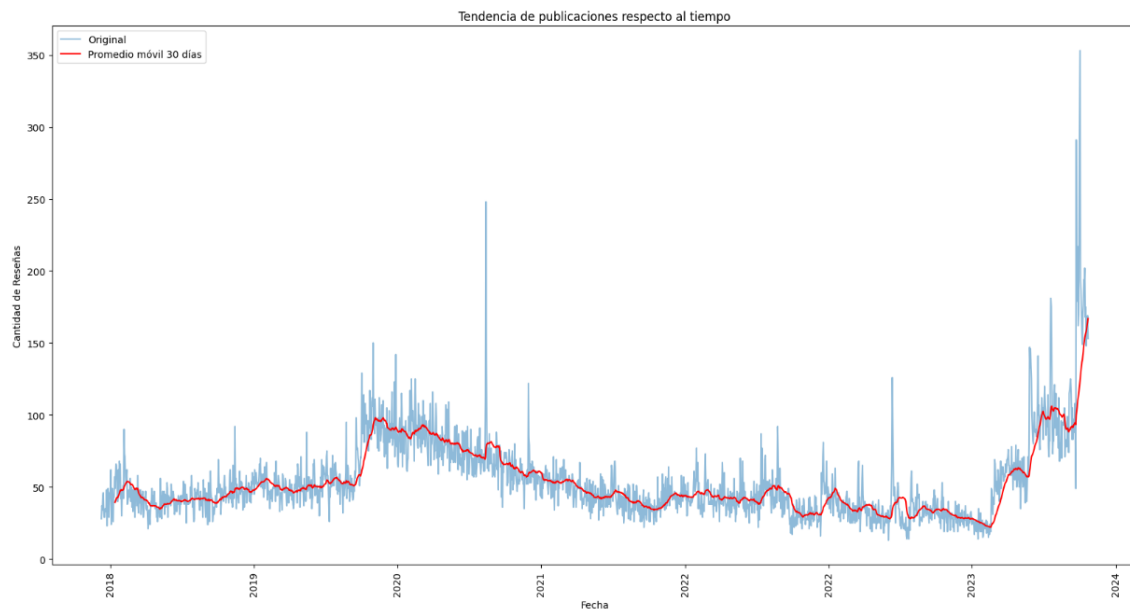


Figura 1: Tendencia de publicaciones respecto al tiempo

Esta gráfica de tiempo muestra picos y valles en la cantidad de reseñas a lo largo del tiempo, con un aumento notable en los últimos períodos, indicando momentos de alta actividad y discusión entre los usuarios.

4.2. Palabras más frecuentes - Reseñas Netflix

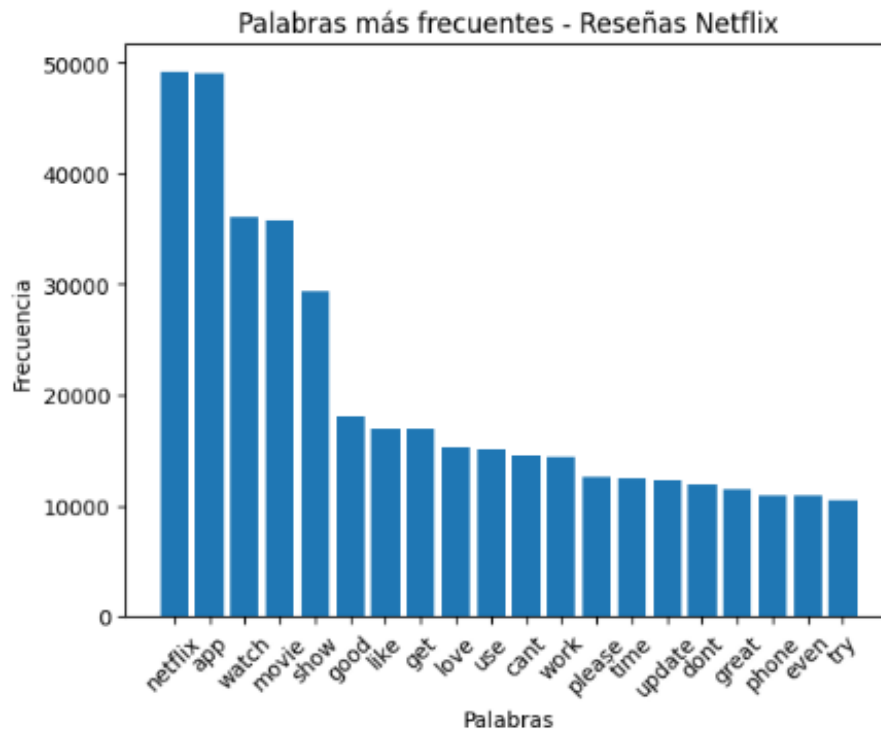


Figura 2: Palabras más frecuentes - Reseñas Netflix

Esta gráfica muestra una lista consolidada de términos comunes en las reseñas, con "Netflix", "watch" y "movie" como los más prominentes, reflejando un enfoque claro en la plataforma y el consumo de contenido.

4.3. Palabras más frecuentes en el año 2018

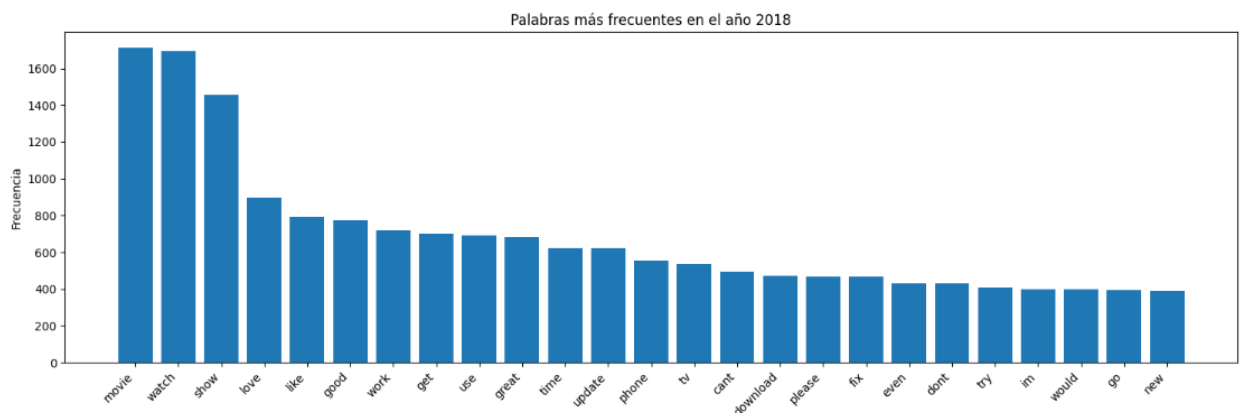


Figura 3: Palabras más frecuentes en 2018

Esta gráfica revela una variedad de palabras frecuentes, con 'movie' y 'watch' a la cabeza, indicando un interés sostenido en el visionado de películas y series. La diversidad de adjetivos positivos sugiere una acogida favorable de las ofertas de la plataforma.

4.4. Palabras más frecuentes en el año 2019

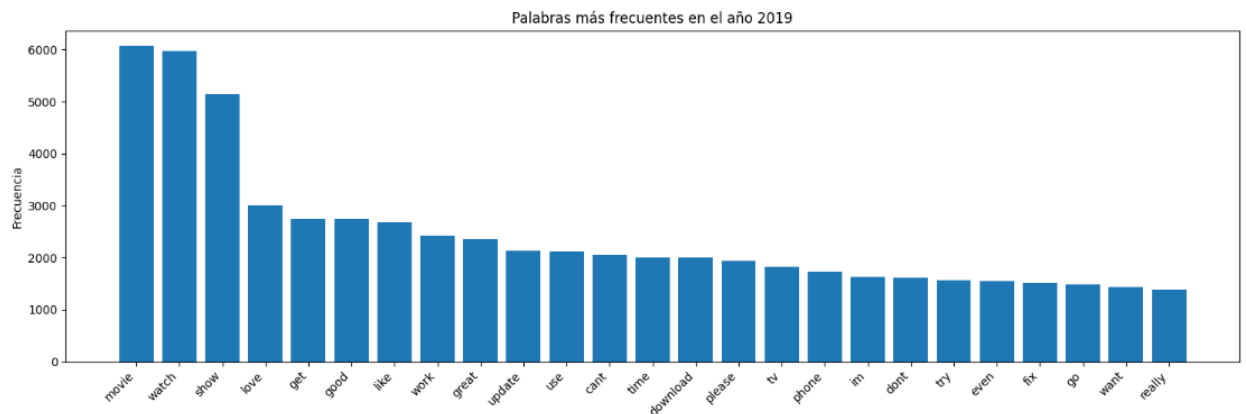


Figura 4: Palabras más frecuentes en 2019

En 2019, 'movie' y 'watch' siguen siendo términos clave, mostrando que el interés y el enfoque principal de las reseñas no ha cambiado significativamente a lo largo de los años. La presencia continua de palabras como 'good' y 'great' subraya una percepción positiva del contenido.

4.5. Palabras más frecuentes en el año 2020

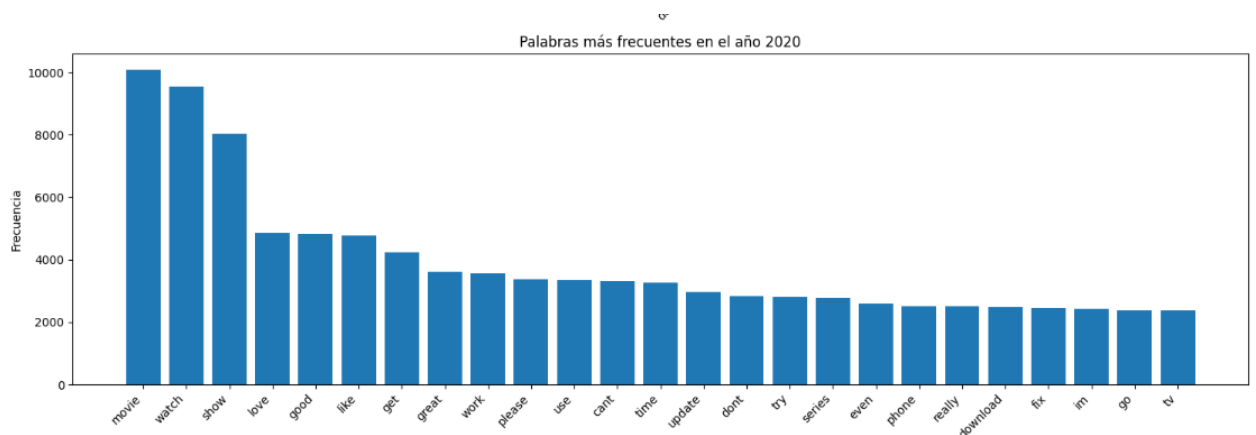


Figura 5: Palabras más frecuentes en 2020

El año 2020 muestra un aumento en la frecuencia de términos relacionados con el contenido visual, como 'movie' y 'watch', lo que puede ser indicativo de un mayor consumo de contenido durante períodos de confinamiento global. El término 'love' aparece nuevamente, reflejando una respuesta emocionalmente positiva hacia el contenido ofrecido durante este período desafiante.

4.6. Palabras más frecuentes en el año 2021

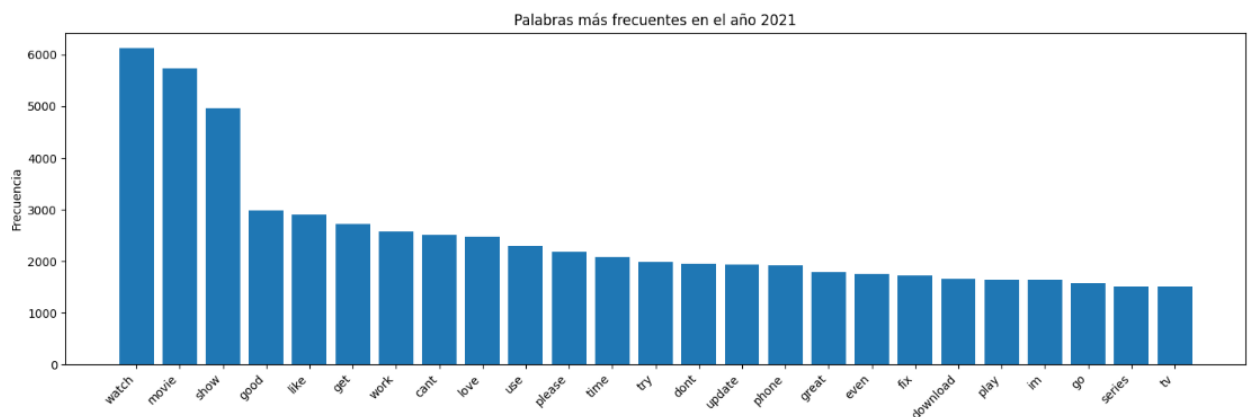


Figura 6: Palabras más frecuentes en 2021

En 2021, 'watch' y 'movie' también son prominentes, acompañados de "good" y "love", lo cual podría sugerir una respuesta particularmente positiva hacia el contenido ofrecido ese año. La presencia de "love" puede indicar un alto nivel de satisfacción y disfrute del contenido visualizado.

4.7. Palabras más frecuentes en el año 2022

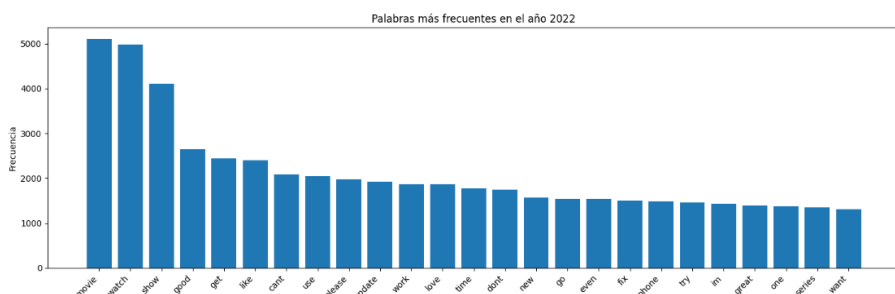


Figura 7: Palabras más frecuentes en 2022

Las palabras más frecuentes en 2022 siguen un patrón similar, con 'movie' y 'watch' liderando la tabla, junto con términos como 'good' y 'like'. Esto indica que

los usuarios continúan enfocándose en la calidad del contenido y la experiencia general de visionado, reiterando la importancia de las funcionalidades que mejoren esta experiencia.

4.8. Palabras más frecuentes en el año 2023

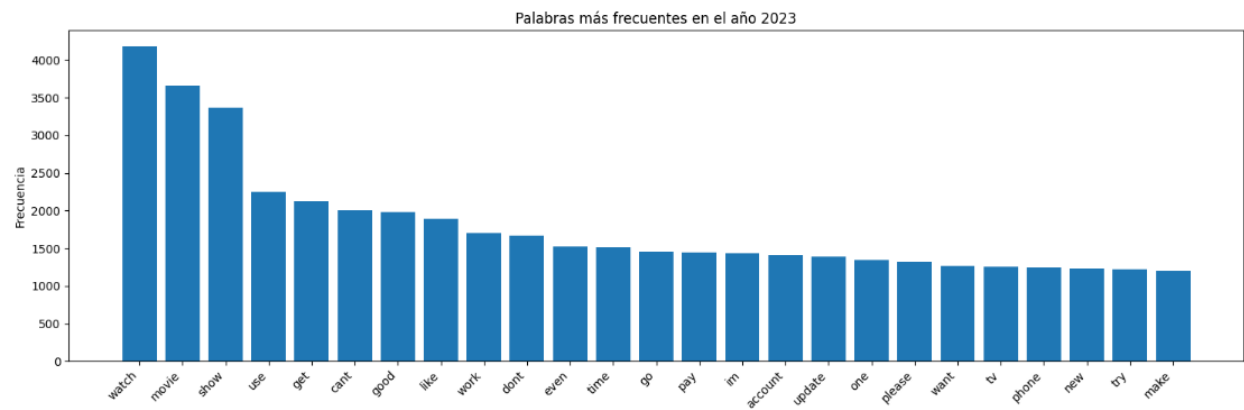


Figura 8: Palabras más frecuentes en 2023

Similar a 2024, las palabras 'watch', 'movie', 'use', y 'good' dominan las reseñas en 2023, lo que subraya una constante valoración del contenido y de la facilidad de uso de la plataforma. La consistencia en las palabras clave a lo largo de los años resalta una experiencia de usuario estable y centrada en el entretenimiento y la funcionalidad.

4.9. Palabras más frecuentes en el año 2024

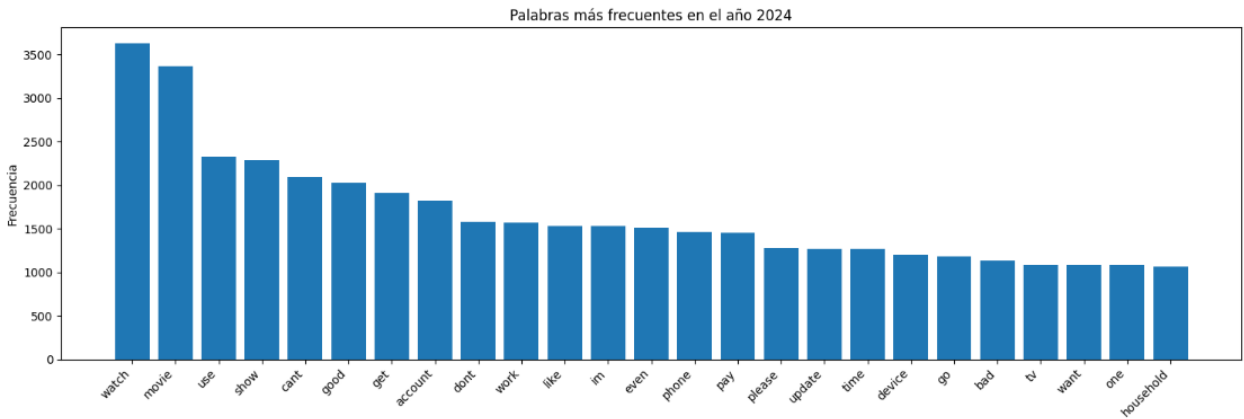


Figura 9: Palabras más frecuentes en 2024

Esta gráfica muestra que las palabras más utilizadas en las reseñas de 2024 incluyen términos clave como 'watch', 'movie', 'use', 'show', y 'good'. La alta frecuencia de estas palabras indica un fuerte enfoque de los usuarios en la visualización de películas y series, así como en la calidad y utilidad de la plataforma. Esto refleja una experiencia de usuario predominantemente orientada hacia el consumo de contenido y la interacción con la plataforma.

4.10. Bigramas más frecuentes

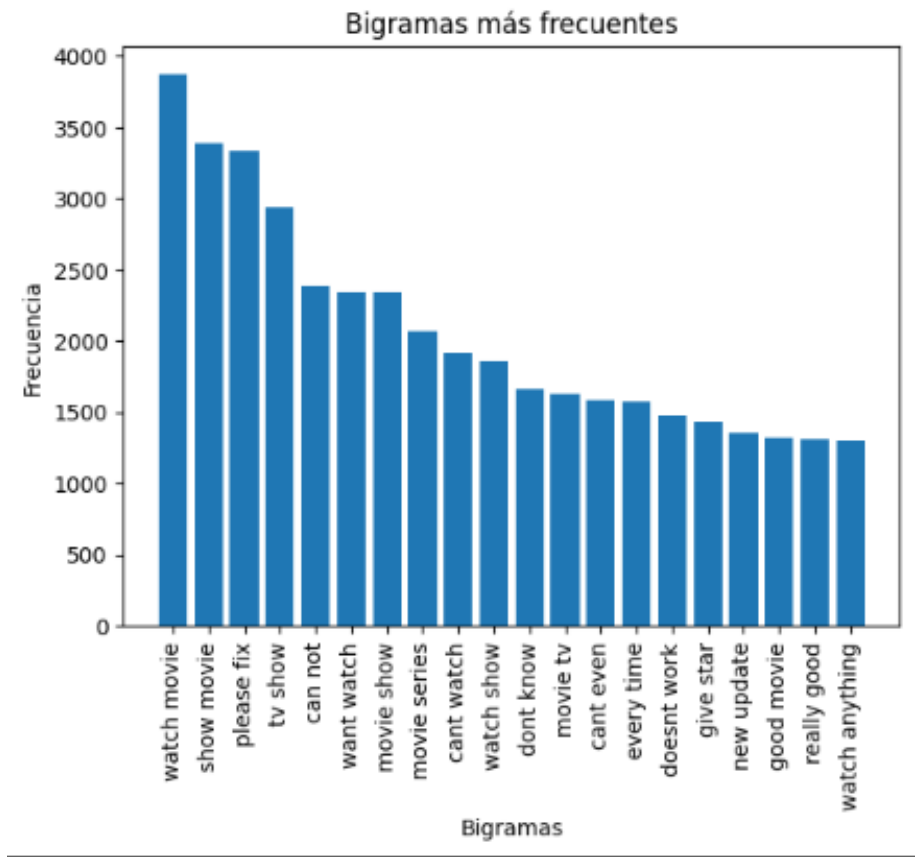


Figura 10: Bigramas más frecuentes

Los bigramas más comunes incluyen 'watch movie', 'please fix', y 'can not', lo que revela preocupaciones comunes y solicitudes de los usuarios. El uso frecuente de 'please fix' puede indicar áreas específicas de mejora solicitadas por los usuarios, mientras que 'watch movie' refuerza la tendencia hacia el contenido cinematográfico.

4.11. Análisis de Sentimiento por Año

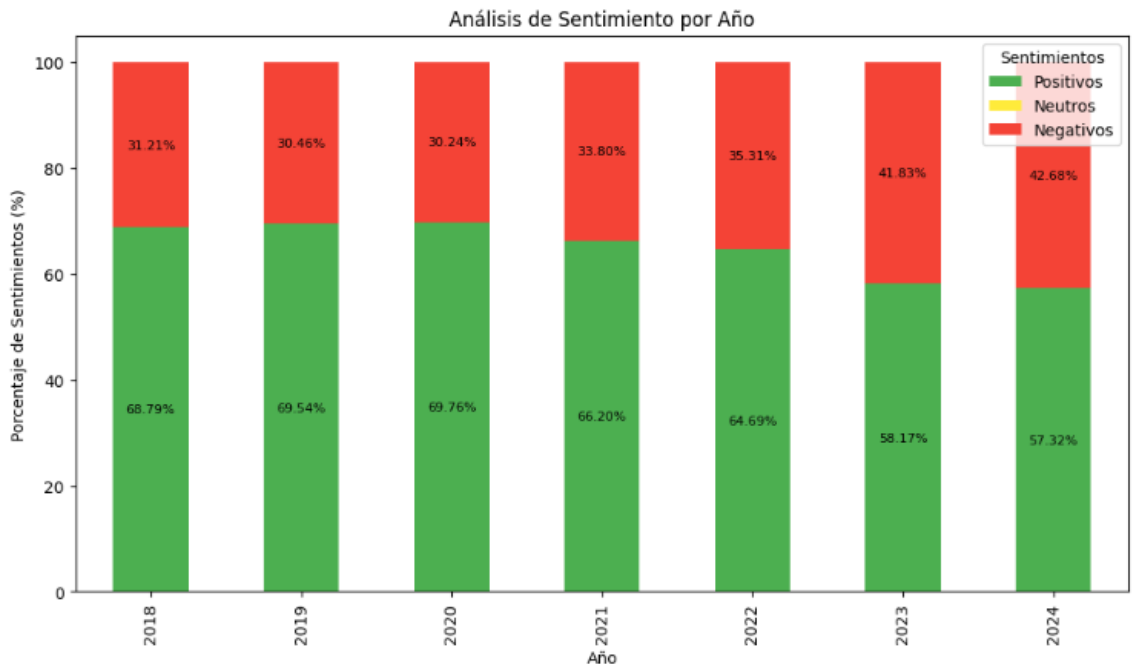


Figura 11: Análisis de Sentimiento por Año

El análisis de sentimientos revela una variación en los porcentajes de sentimientos positivos, neutros y negativos a lo largo de los años. Aunque la mayoría de los años muestran una predominancia de sentimientos positivos, hay un incremento notable en los comentarios negativos en años recientes, posiblemente reflejando cambios en las expectativas de los usuarios o en la calidad del servicio.

5. Conclusiones

La conclusión general del análisis de las reseñas de usuarios en *Netflix* a lo largo de varios años revela patrones consistentes y cambios significativos en las percepciones y expectativas de los usuarios. Términos como 'watch', 'movie', y 'good' han dominado consistentemente las reseñas, destacando un enfoque continuo en la calidad del contenido y la experiencia de visualización, reflejando una fuerte preferencia y expectativa por parte de los usuarios de *Netflix* por encontrar contenido entretenido y de calidad.

A través de los años, aunque ciertos términos han sido constantes, ha habido cambios notables en la frecuencia y el tipo de palabras usadas. El uso de términos como 'love' en años específicos sugiere períodos de alta satisfacción del usuario, mientras que la aparición de solicitudes directas en bigramas y trigramas, como 'please fix', indica áreas de preocupación y necesidad de mejora en la plataforma.

El análisis también muestra un aumento en la actividad de reseñas y cambios en las tendencias de palabras durante eventos globales significativos, como la pandemia en 2020, sugiriendo que *Netflix* juega un papel crucial en la vida de sus usuarios durante períodos de crisis, proporcionando un escape o una forma de entretenimiento esencial. Sin embargo, el análisis de sentimientos a lo largo de los años revela una tendencia hacia un aumento en los comentarios negativos en años recientes, lo que podría reflejar un cambio en las expectativas de los usuarios o posibles desafíos enfrentados por *Netflix* en términos de satisfacer a una base de usuarios diversa y global.

Este estudio resalta la importancia de un análisis continuo y detallado de las reseñas de los usuarios para plataformas de *streaming* como *Netflix*. Entender las tendencias, problemas y cambios en las percepciones de los usuarios puede ayudar a *Netflix* a adaptarse mejor a las necesidades cambiantes de su audiencia, optimizar la experiencia del usuario, y mejorar la calidad del contenido ofrecido.

6. Referencias

Ashish Kumar - Netflix Reviews - <https://www.kaggle.com/datasets/ashishkumarak/netflix-reviews-playstore-daily-updated>