



Facultad de Ciencias Empresariales

**PLATAFORMAS DE VÍDEO BAJO DEMANDA: EL VALOR DE SU
PRESENCIA Y DE LA INTERACCIÓN CON SUS USUARIOS EN
REDES SOCIALES**

Autora: Fátima Güell Aldama

Directora: Lucía Barcos Redín

MADRID | Junio 2023

RESUMEN:

Nuestro objetivo con el presente trabajo es identificar posibles patrones o correlaciones en las interacciones de las plataformas de Servicios de Video a la Carta por Suscripción con sus usuarios en Redes Sociales. En particular, queremos analizar de qué manera los antecedentes relativos al mensaje de la Participación del Cliente en Redes Sociales -sobre todo el tono y la carga emocional de los mensajes- afectan a la manifestación de la Participación del Cliente en Redes Sociales. Con tal propósito, presentamos la siguiente estructura de trabajo. En primer lugar, hablaremos del Marketing y las Redes Sociales, centrándonos en el concepto de Participación del Cliente y su protagonismo en el estado actual del campo del marketing. para a continuación profundizar sobre la industria de los Servicios de Video a la Carta por Suscripción -de nuevo prestando atención a su relación con las Redes Sociales. En tercer lugar, trataremos de manera teórica algunas técnicas de *Text Mining*, y de *Sentiment Analysis* en particular, así como de su aplicación en el entorno RStudio. Todo ello encaminado a presentar en último lugar un estudio de campo sobre el *Twitter Sentiment Analysis* aplicado al caso de las plataformas Servicios de Video a la Carta por Suscripción con RStudio.

PALABRAS CLAVE: Servicios de Video a la Carta por Suscripción, Twitter, Text Mining, Sentiment Analysis, RStudio.

ABSTRACT:

Our goal with the present work is to identify possible patterns or correlations in the interactions of Video on Demand Subscription Services platforms with their users on Social Networks. In particular, we want to analyze how the antecedents related to the message of Customer Engagement in Social Networks -especially the tone and emotional charge of the messages- affect the manifestation of Customer Engagement on Social Networks. For this purpose, we present the following structure for the paper. First, we will discuss Marketing and Social Networks, focusing on the concept of Customer Engagement and its prominence in the current state of the marketing field. We will then delve into the Video on Demand Subscription Services industry - again paying attention to its relationship with Social Networks. Thirdly, we will discuss some Text Mining techniques, from a theoretical approach, and Sentiment Analysis in particular, as well as their application in the RStudio environment. Finally, we will present a field study on Twitter Sentiment Analysis applied to the case of Video on Demand Subscription Services platforms with RStudio.

KEYWORDS: Subscription Video On Demand Services, Twitter, Text Mining, Sentiment Analysis, RStudio

ÍNDICE GENERAL

| | |
|--|-----------|
| ÍNDICE DE FIGURAS | 5 |
| GLOSARIO | 6 |
| INTRODUCCIÓN | 7 |
| 1. OBJETIVO DEL TRABAJO | 7 |
| 2. METODOLOGÍA | 7 |
| 3. ESTADO DE LA CUESTIÓN: CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA | 7 |
| 4. ESTRUCTURA DEL TRABAJO | 8 |
| CAPÍTULO I. MARKETING Y RRSS | 9 |
| 1. EL MARKETING Y SU EVOLUCIÓN | 9 |
| 2. LAS RRSS Y SU PROTAGONISMO EN MARKETING DURANTE LAS ÚLTIMAS DÉCADAS | 10 |
| 3. LA PARTICIPACIÓN O COMPROMISO DEL CLIENTE (PDC) | 12 |
| 3.1. Conceptualización de la PDC | 12 |
| 3.2. Antecedentes, Decisiones y Resultados de la PDC en RRSS | 13 |
| a. Antecedentes de la PDC en RRSS | 13 |
| i. Las manifestaciones de la PDC en RRSS | 15 |
| ii. Resultados de la PDC en RRSS | 15 |
| CAPÍTULO 2. LAS PLATAFORMAS DE SVOD Y LAS RRSS | 16 |
| 1. UN CAMBIO RADICAL EN LA MANERA DE CONSUMIR ENTRETENIMIENTO: LAS PLATAFORMAS DE SVOD | 16 |
| 2. EL “BOOM” DE LA INDUSTRIA DE LAS PLATAFORMAS SVOD | 17 |
| 3. LA IMPORTANCIA DE LA PRESENCIA DE LAS PLATAFORMAS DE SVOD EN RRSS | 18 |
| CAPÍTULO 3. APROXIMACIÓN TEÓRICA AL SENTIMENT ANALYSIS Y A SU APLICACIÓN A DATOS TEXTUALES OBTENIDOS DE TWITTER CON RSTUDIO | 20 |
| 1. INTRODUCCIÓN AL TEXT MINING | 20 |
| 1.1. ¿Qué es el <i>Text Mining</i> ? | 20 |
| 1.2. El <i>Text Mining</i> paso a paso | 21 |
| i. Formas de preprocesamiento de los datos textuales | 23 |
| ii. Posibles operaciones de limpieza de los datos | 24 |
| 1.3. Tipos o formas de <i>Text Mining</i> según el análisis | 24 |
| a. <i>Word Frequency Analysis</i> | 24 |
| b. <i>Sentiment Analysis</i> | 24 |
| i. El <i>Sentiment Analysis</i> y sus técnicas | 24 |
| ii. El lexicón de emociones NRC | 26 |
| 2. SENTIMENT ANALYSIS APLICADO A TWITTER CON RSTUDIO | 27 |
| 2.1. RStudio como herramienta para llevar a cabo un <i>Sentiment Analysis</i> | 27 |
| 2.2. Twitter como fuente de datos para <i>Sentiment Analysis</i> | 28 |
| CAPÍTULO 4. ESTUDIO DE CAMPO: TWITTER SENTIMENT ANALYSIS APLICADO AL CASO DE LAS PLATAFORMAS SVOD CON RSTUDIO | 30 |
| 1. DEFINICIÓN DEL MARCO DE NUESTRO ANÁLISIS | 30 |
| 2. LOS DATOS EMPLEADOS EN NUESTRO ANÁLISIS | 31 |
| 2.1. Primera selección de las plataformas de SVOD e identificación de sus perfiles en Twitter | 31 |

| | | |
|-----------|--|-----------|
| 2.2. | Extracción de los Tuits y Menciones con la API de Twitter desde RStudio | 32 |
| 2.1. | Caracterización de los datos para nuestro análisis | 34 |
| a. | Identificación de las variables más relevantes | 34 |
| b. | Definición del marco temporal de los tuits que serán objeto de análisis | 35 |
| c. | Mayor acotación del objeto de análisis de las Menciones en función de la fuente de publicación del tuit | 37 |
| d. | Características de los tuits a tener en cuenta para el análisis | 38 |
| i. | Consideraciones generales relativas a los caracteres | 38 |
| ii. | El bigrama 'RT' y los re-tuits | 39 |
| 3. | SENTIMENT & EMOTION ANALYSIS CON EL LEXICÓN NRC | 43 |
| 3.1. | Cuestiones previas | 43 |
| 3.2. | Análisis de la polaridad o sentimiento y su visualización | 45 |
| a. | Para los Tuits publicados por las plataformas SVOD | 45 |
| b. | Para las Menciones a las plataformas SVOD | 46 |
| 3.3. | Análisis de las emociones y su visualización | 46 |
| a. | Para los Tuits publicados por las plataformas SVOD | 46 |
| b. | Para las Menciones a las plataformas SVOD | 48 |
| | CONCLUSIONES | 52 |
| | BIBLIOGRAFÍA | 54 |
| | ANEXOS | 59 |
| | Anexo #1: Evolución del uso de las RRSS de 2012-23: número de usuarios en RRSS (en millones) y variación anual (de Enero a Enero) | 59 |
| | Anexo #2: Las RRSS más populares del mundo clasificadas por número mensual de usuarios activos (en millones), Enero 2023 | 60 |
| | Anexo #3: Antecedentes , Decisiones y Resultados de la PDC en RRSS según Lim y Rasul (2022) | 61 |
| | Anexo #4: Código para la descarga de los datos en RStudio y comentarios explicativos | 62 |
| | Anexo #5: Código empleado para la caracterización de nuestros datos | 67 |
| | Anexo #6: Código para el WORD FREQUENCY ANALYSIS | 77 |
| | Anexo #7: Código para el SENTIMENT ANALYSIS | 83 |

ÍNDICE DE FIGURAS

| | |
|--|-----------|
| Ilustración 1: El proceso de Text Mining paso a paso..... | 21 |
| Ilustración 2: Rueda de las emociones de Robert Plutchik | 27 |
| Ilustración 3: Ejemplos de librerías disponibles en RStudio..... | 28 |
| Ilustración 4: Resumen de las etapas seguidas en nuestro estudio de campo | 31 |
| Ilustración 5: Cuota de mercado mundial de las principales plataformas SVOD en 2022 | 32 |
| Ilustración 6: Resumen de los perfiles de Twitter de las plataformas SVOD (Abril de 2023)..... | 32 |
| Ilustración 7: Funciones definidas en RStudio para la descarga de los tuits..... | 33 |
| Ilustración 8: Ejemplo de ejecución de las funciones para la descarga de los datos | 34 |
| Ilustración 9: Dimensiones de los Datasets de Menciones y Tuits para cada plataforma SVOD | 35 |
| Ilustración 10: Franja temporal de los Tuits de cada plataforma SVOD | 35 |
| Ilustración 11: Nueva franja temporal y volumen de Tuits para cada plataforma SVOD | 36 |
| Ilustración 12: Franja temporal y volumen de Menciones para cada plataforma SVOD | 37 |
| Ilustración 13: Nuevo volumen y franja temporal de las Menciones para cada plataforma SVOD | 38 |
| Ilustración 14: Visualización del cuerpo de los 10 primeros Tuits de Netflix..... | 39 |
| Ilustración 15: Etapas del Word Frequency Anlysis realizado para los Tuits y Menciones | 41 |
| Ilustración 16: Las 20 palabras más repetidas en los Tuits de Netflix | 42 |
| Ilustración 17: Nube de palabras con las palabras más repetidas en las Menciones de Amazon Prime Video..... | 42 |
| Ilustración 18: Proporción de re-tuits en los Tuits y Menciones de cada plataforma SVOD | 43 |
| Ilustración 19: Esquema del proceso observado para el análisis de los Tuits..... | 44 |
| Ilustración 20: Esquema del proceso observado para el análisis de las Menciones..... | 45 |
| Ilustración 21: Proporción de las polaridades asociadas a los Tuits de cada plataforma | 45 |
| Ilustración 22: Proporción de las polaridades asociadas a las Menciones de cada plataforma | 46 |
| Ilustración 23: Peso de las emociones en los Tuits de las distintas plataformas..... | 47 |
| Ilustración 24: Peso de las emociones en las Menciones a las distintas plataformas | 49 |

GLOSARIO

PDC: Participación (o Compromiso) del Cliente. En inglés *Customer Engagement*.

RRSS: Redes Sociales. En inglés *social media*.

SVOD: Servicios de Video a la Carta por Suscripción. En inglés *Subscription Video On Demand Services*.

tuit (en minúscula) y **Tuit** (con mayúscula): 1. “tuit” hace referencia, de manera genérica, a todo mensaje de texto publicado por un usuario en Twitter; 2. “Tuit”, en nuestro estudio de campo, hace referencia a los tuits publicados por las plataformas SVOD en sus *timelines* de Twitter

mención (en minúscula) y **Mención** (con mayúscula): 1. “mención” se refiere, de manera genérica, a todo tuit en el que un usuario apela directamente a otro por su nombre; 2. “Mención”, hace referencia de en nuestro estudio de campo, a las menciones que los usuarios de Twitter hacen a las plataformas SVOD

Streaming = Transmisión

Sentiment Analysis = Análisis de Sentimiento

Word Frequency Analysis = Análisis de la Frecuencia de las Palabras

Text Mining = Text Mining

INTRODUCCIÓN

1. OBJETIVO DEL TRABAJO

Nuestro objetivo con el presente trabajo es identificar posibles patrones o correlaciones en las interacciones de las plataformas de SVOD con sus usuarios en RRSS. En particular, queremos analizar de qué manera los antecedentes de la PDC en RRSS relativos al mensaje -sobre todo el tono y la carga emocional de los mensajes- afectan a la manifestación de la PDC en RRSS.

2. METODOLOGÍA

Como muestra de los antecedentes de la PDC en RRSS relativos al mensaje analizaremos los tuits publicados por determinadas plataformas de SVOD en Twitter (en adelante Tuits, no confundir con tuits, ver **Glosario**), en un periodo temporal determinado que definiremos más adelante. Puesto que nos interesa sobre todo analizar el tono y la carga emocional empleada por dichas plataformas en sus tuits, al margen de realizar un *Word Frequency Analysis* sobre los mismos, realizaremos igualmente un *Sentiment Analysis*.

Como manifestación de la PDC en RRSS analizaremos los tuits publicados por usuarios individuales de Twitter -en un periodo de tiempo determinado- en los que mencionan de manera directa, es decir, que apelan directamente a las mismas plataformas de SVOD (en adelante Menciones). Nos centraremos sobre todo en el volumen de los mismos y de nuevo, en el tono y carga emocional que los usuarios ponen en sus tuits (*Sentiment Analysis*). En este sentido, partimos de la presunción que los usuarios que interactúan de manera tan directa con las plataformas de SVOD en Twitter, si no son consumidores efectivos de los servicios ofrecidos por éstas, al menos son consumidores potenciales

3. ESTADO DE LA CUESTIÓN: CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA

Dentro del predominante Marketing Relacional, resulta de particular interés el concepto de Participación – o Compromiso- del Cliente (“PDC”, del inglés *Customer Engagement*), - que la marcas deben tener muy presente de cara a construir relaciones a largo plazo con sus consumidores y solidificar la lealtad de los mismos hacia la marca.

La PDC también ha entrado en el incipiente universo digital y de las RRSS, donde presenta antecedentes, manifestaciones y resultados particulares muy interesantes para las marcas. En efecto, las RRSS -entre ellas, y en particular, Twitter- tienen un increíble potencial marketiniano, pues se presentan cómo un espacio idóneo para favorecer la comunicación bidireccional entre marca y consumidor que caracteriza el Marketing Relacional.

Por todo ello, resulta evidente que las marcas deben comprender cómo interactuar con sus clientes en línea y qué factores tienen un impacto positivo en la PDC para así poder rentabilizar al máximo su presencia y actividad en RRSS.

Así las cosas, a aquellas empresas con presencia en RRSS les puede resultar de gran utilidad el analizar no solo el tono empleado por ellas en RRSS, sino la manera en que los usuarios reaccionan a dicho tono, por ejemplo, a través del *Text Mining* -y en particular el *Sentiment Analysis*.

Por otra parte, debemos mencionar la competitividad de la industria de los servicios SVOD, dónde no cesa de crecer tanto la oferta de plataformas de *streaming* como la demanda de sus servicios. Una industria de carácter predominantemente tecnológico, que tiene una presencia más que notoria en RRSS, y que parece estar beneficiándose del concepto de PDC, aunque sea de manera inconsciente.

4. ESTRUCTURA DEL TRABAJO

En primer lugar, hablaremos del Marketing y las RRSS, para a continuación profundizar sobre la industria de los servicios SVOD -de nuevo prestando atención a su relación con las RRSS. En tercer lugar trataremos de manera teórica algunas técnicas de *Text Mining*, y de *Sentiment Analysis* en particular, así como de su aplicación en el entorno RStudio. Todo ello encaminado a presentar en último lugar un estudio de campo sobre el *Twitter Sentiment Analysis* aplicado al caso de las plataformas SVOD con RStudio.

CAPÍTULO I. MARKETING Y RRSS

1. EL MARKETING Y SU EVOLUCIÓN

Philip Kotler, conocido como el “padre del Marketing moderno” define el marketing como “un proceso social y administrativo mediante el cual grupos e individuos obtienen lo que necesitan y desean a través de generar, ofrecer e intercambiar productos de valor con sus semejantes” (cómo se citó en Fuente, 2022). A su vez, la *American Marketing Association* define el marketing como “una función de la organización y un conjunto de procesos para crear, comunicar y entregar valor a los clientes, y para manejar las relaciones con estos últimos, de manera que beneficien a toda la organización” (cómo se citó en Fuente, 2022).

Asimismo, el Marketing, se inscribe dentro un proceso económico evolutivo más amplio, quedando su desarrollo marcado por avatares políticos y sociales, que van adaptando el Marketing a las circunstancias del momento (Ballesteros, 2020). De hecho, hoy en día los consumidores tienen altas expectativas en las campañas de marketing de las distintas empresas y sus marcas, pues esperan que las marcas hagan algo más que simplemente venderles un producto. Esperan que las marcas conecten con ellos.

Esto sería indicativo de un cambio en la manera de hacer marketing. Concretamente, el paso de un Marketing Transaccional a un Marketing Relacional, que se viene produciendo en las últimas décadas (Islam et al., 2019) y que está obligando a los profesionales del Marketing a hallar formas innovadoras de conectar con los clientes. ¿Pero en qué consiste este cambio?

Mientras que el Marketing Transaccional busca aumentar constantemente las ventas a través de una captación indiferenciada de consumidores potenciales, el Marketing Relacional busca mejorar los resultados a través de la fidelización de consumidores confirmados del producto, es decir, construyendo una relación satisfactoria con las personas que ya son consumidores (Díaz, 2019).

Así pues, en el Marketing Transaccional la comunicación entre la marca y el consumidor es unidireccional: el vendedor es quién se comunica con sus potenciales clientes, y lo hace en masa, de manera indiscriminada y sin obtener información acerca de ellos. En cambio, en el Marketing Relacional, la comunicación marca-consumidor es bidireccional:

la marca obtiene información valiosa de sus consumidores, lo que le permitirá a su vez crear acciones más personalizadas hacia sus consumidores, favoreciendo así la fidelización de los mismos.

Pues bien, dentro de este contexto de cambio, las RRSS se presentan como el espacio idóneo para que tenga lugar esa comunicación bidireccional entre marca y consumidor que caracteriza al marketing Relacional. En ellas, los consumidores pueden expresar fácil y abiertamente sus actitudes y opiniones hacia diversas marcas, ya sea a través de comentarios, o de acciones como "me gusta" o "compartir" (Lim y Rasul, 2022).

2. LAS RRSS Y SU PROTAGONISMO EN MARKETING DURANTE LAS ÚLTIMAS DÉCADAS

Las RRSS pueden definirse como "herramientas para la interacción social, que utilizan técnicas de comunicación altamente accesibles y escalables -como las tecnologías móviles basadas en la web- para convertir la comunicación en un diálogo interactivo" (Coulson, 2013).

Las RRSS, son muchas y muy variadas, pudiendo poner a disposición de los usuarios técnicas de comunicación diversas: desde imágenes y vídeos, pasando por comentarios y publicaciones de texto, hasta mensajería instantánea.

A lo largo de la última década, las RRSS se han convertido en una auténtica fuerza, contando con cerca de 4 760 millones de usuarios en todo el mundo en enero de 2023, ósea algo menos del 60% del total de la población mundial (Kemp, 2023).

Si bien parece que el crecimiento del número de usuarios en RRSS se ha frenado en los últimos meses, habiéndose contabilizado en 2023 un crecimiento anual de sólo el 3% - frente al crecimiento anual de más del 10% observado en 2022 (**Anexo 1**) -, es indiscutible las RRSS siguen creciendo, y a pasos agigantados.

Además, el aumento del uso de *smartphones* ("teléfonos inteligentes") y teléfonos móviles también ha espoleado el uso de las RRSS (Hooda y Ankur., 2018). A principios de 2023, el número de personas que utilizan el teléfono móvil alcanzó los 5 440 millones a nivel mundial. Este número -que ha crecido un 3% en 2022, y casi un 2% en 2021 (Kemp, 2023; Kemp, 2022), equivale al 68% de la población mundial.

Según Dixon (2023), entre las RRSS más populares en enero de 2023 (**Anexo 2**) - popularidad medida en el número de usuarios mensuales-, cabe destacar Facebook (fundada en 2003, cuenta con poco menos de 3 000 millones de usuarios mensuales), YouTube (fundada en 2005, cuenta con algo más de 2 500 millones) e Instagram (fundada en 2014, suma 2 000 millones). En cambio, Twitter -fundada en 2006- ocupa un lugar menos importante, con cerca de 550 millones de usuarios mensuales.

Históricamente, el principal objetivo de estas RRSS era permitir a personas con intereses comunes compartir información y expresar opiniones (Barreto, 2013). Sin embargo, el uso de las RRSS ya no se limita sólo a esto. En la práctica, las RRSS -entre las que cabe destacar Facebook, Instagram, y Twitter, entre otras- se han convertido en poderosas y populares herramientas de Marketing Digital¹, empleadas por las organizaciones tanto para promocionar sus productos, como construir relaciones entre los consumidores y sus marcas (Lim y Rasul, 2022) -concretamente para para construir y desarrollar una imagen de marca, así como la identidad corporativa de la empresa (Ibrahim et Al., 2017).

Prueba del potencial marketiniano que presentan las RRSS, es el hecho de que, en la última década, éstas han sido el objeto de múltiples adquisiciones multimillonarias (Escriche, 2022), ya sea a manos de otras RRSS -como la compra de WhatsApp y de Instagram por parte de Facebook- o de alguna de las grandes empresas tecnológicas -por ejemplo, la adquisición de LinkedIn por parte de Microsoft; o la de YouTube por parte de Google-. La más reciente de estas ha sido la adquisición de Twitter por parte de Elon Musk, en Octubre de 2022, valorada en \$44 000 millones (Southern, 2022).

Y mayor prueba aún de dicho potencial es el hecho de que marcas en todo el mundo han multiplicado sus inversiones en el espacio digital social (compuesto por una gran variedad de RRSS, como ya hemos podido comprobar): éstas han pasado de cerca de \$380 000 millones en 2020 a más de \$491 000 millones en 2021, previéndose que las mismas alcancen los \$785 000 millones en 2025 (Lim y Rasul, 2022).

¹ El Marketing Digital se define como la promoción de productos o marcas mediante herramientas y técnicas en línea.

3. LA PARTICIPACIÓN O COMPROMISO DEL CLIENTE (PDC)

Dentro del Marketing Relacional, en el que como ya hemos explicado, las RRSS tienen un papel protagonista, resulta de particular interés el concepto de Participación – o Compromiso- del Cliente (“PDC”, del inglés *Customer Engagement*), - que definiremos en mayor profundidad más adelante- pues su presencia ha probado ser fundamental para construir interacciones y relaciones con los clientes a largo plazo y solidificar la lealtad del cliente hacia la marca (Lim y Rasul, 2022).

De hecho, líneas de investigación anteriores han puesto en evidencia que la PDC afecta positivamente a numerosos comportamientos del cliente como pueden ser las evaluaciones de marca, la lealtad, la confianza, la intención de compra y la satisfacción del cliente (Lim y Rasul, 2022). En consonancia con esto, el Marketing Science Institute -organismo mundial de Marketing profesional-, ha incluido en dos ocasiones consecutivas (primero 2018, y luego 2020) la PDC en su lista de prioridades de cara a construir un compromiso sostenible con los consumidores.

3.1. Conceptualización de la PDC

Según Lim y Rasul (2022), la PDC se presenta como un campo de estudio emergente en el mundo de la investigación, comenzando alrededor del año 2010 tanto su consideración a nivel académico como su conceptualización. A su vez, la PDC en RRSS es un línea de investigación que en los últimos años ha ido ganando cada vez más atención académica.

En particular, van Doorn et al. (2010, p. 253), quienes enfocan la PDC como un construcción teórica de carácter conductual, definen la misma como "la manifestación conductual de los clientes hacia una marca o empresa, más allá de la compra, resultante de impulsores motivacionales".

El proceso tradicional de PDC *offline* -esto es, al margen de las RRSS- se basa en una comunicación unidireccional que, permite a las marcas relacionarse con los clientes, pero limita la capacidad de los clientes para relacionarse con las marcas (Lim y Rasul, 2022). A su vez, la estrategia contemporánea de PDC *online* -es decir, en línea, basada en RRSS- permite a marcas y clientes relacionarse recíprocamente (Lim y Rasul, 2022), y resulta, por ello mismo, mucho más potente. En efecto, las RRSS permiten a las marcas construir rápidamente la PDC a través de una comunicación directa y flexible, proporcionándoles

funciones analíticas que les permiten evaluar el impacto y la eficacia de su comunicación en sus relaciones con sus clientes objetivo.

No obstante, las RRSS también han hecho que la comunicación social se vuelva más compleja y dinámica, pues no solo permite intercambios directos entre clientes y marcas, sino también entre los propios clientes (Lim y Rasul, 2022). Así, podemos decir que las RRSS han contribuido a la transformación de los clientes en participantes activos y cocreadores de las marcas.

3.2. Antecedentes, Decisiones y Resultados de la PDC en RRSS

Lim y Rasul (2022), han realizado un análisis sistemático de los estudios conceptuales y empíricos, sobre la PDC en las RRSS, publicados en revistas académicas entre los años 2000 y 2020. El objetivo de dicho análisis -basado en el marco conceptual ADO²- era identificar los antecedentes, decisiones y resultados de la PDC en RRSS.

En el marco ADO (Paul y Benito, 2018), los antecedentes son precursores directos de las decisiones -explican directamente por qué se toma o no una decisión- y los precursores indirectos de los resultados -explican indirectamente por qué se obtiene o no un resultado.

¿Pero qué significan estos conceptos, aplicados a la PDC en RRSS? Debemos comenzar por señalar que las “decisiones” hacen referencia a las diversas manifestaciones de la PDC que podrían darse en RRSS. A su vez, los "antecedentes" hacen referencia a los factores que influyen en que se produzcan dichas decisiones, es decir, en que se produzca o no la PDC en RRSS. Por último, los "resultados" hacen referencia a las principales consecuencias de dicha PDC en RRSS.

Pues bien, en su revisión de la PDC en RRSS, Lim y Rasul (2022) descubrieron ocho categorías de antecedentes, tres categorías de decisiones y cuatro categorías de resultados (**Anexo 3**).

a. Antecedentes de la PDC en RRSS

² ADO es el acrónimo inglés de “Antecedents, Decisions, Outcomes” (Antecedentes, Decisiones y Resultados), un marco conceptual, elaborado por Paul y Benito (2018), para la revisión sistemática de estudios y artículos académicos.

De entre las ocho categorías de antecedentes identificadas por Lim y Rasul (2022), cabe destacar las siguientes:

- **1: Antecedentes relativos al Mensaje:** se trata de aquellos atributos de los mensajes diseñados y difundidos por los profesionales del marketing y dirigidos a los consumidores, que podrían influir en la participación de éstos en RRSS. A saber: el estilo de comunicación, el mensaje, la fuente del contenido, la personalización, el boca a boca electrónico, el entretenimiento, la riqueza informativa, la interacción, el enfoque normativo y la tendencia.

Por ejemplo, se ha probado que los mensajes con un estilo de comunicación amable aumentan tanto el número de comentarios, como el tono positivo de los mismos. Así mismo, si bien el contenido informativo por sí solo no es muy eficaz para suscitar la PDC, puede llegar a serlo al combinarse con contenido emocional o expresivo de la personalidad, como pueden ser la emoción o el humor, especialmente cuando la intención es construir comunidades de marca.

Por ello, las empresas deben cuidar el tono empleado en su estrategia de comunicación en RRSS, no solo para que se ajuste a la imagen de marca que desean proyectar online, sino, sobre todo, para favorecer la PDC online de sus consumidores, por todos los beneficios -o resultados- que ésta conlleva (y que expondremos más adelante). En esta línea, las empresas con presencia en RRSS pueden encontrar interesante y provechoso analizar tanto el tono empleado por ellas en RRSS, como la manera en que los usuarios reaccionan a dicho tono. Una forma de realizar dicho análisis sería a través del *Sentiment Analysis* [Análisis de sentimiento], que proponemos en nuestro trabajo y sobre el que profundizaremos más adelante.

- **2: Antecedentes relativos a la Industria:** se trata de aquellos factores influidos o moldeados por la industria que podrían explicar la PDC en las RRSS. Por una parte, tenemos la implicación en la industria de la moda, y por otra, la orientación tecnológica. Mientras que la motivación para implicarse en la moda en sí misma no parece tener un impacto significativo en la PDC en RRSS, el uso de tecnologías que requieren la implicación tanto de clientes como de vendedores sí que lo tiene.

En este sentido, las plataformas de SVOD -parte del objeto de estudio de este trabajo- podrían entrar en la categoría de industria con orientación tecnológica, pues los

usuarios consumen estos servicios, normalmente online -salvo los supuestos en los que puedan descargar el contenido para reproducir offline- y en todo caso a través de un dispositivo tecnológico.

- **3: Antecedentes relativos a la Plataforma:** se trata de las características de las plataformas de RRSS que podrían influir en la PDC en estas mismas RRSS. Dichas características son: la expectativa de esfuerzo, la expectativa de rendimiento, el disfrute, riesgo y confianza percibidos. Concretamente, las plataformas de RRSS que exigen un mayor esfuerzo y suponen un mayor riesgo para los usuarios (por ejemplo, problemas de privacidad) tienden a disuadir la PDC en dichas plataformas. Por el contrario, las plataformas de RRSS que permiten a sus usuarios disfrutar de ellas de manera fiable y sin incidencias en su funcionamiento, motivarán la PDC en ellas.

En este sentido, Twitter no es la plataforma más segura a ojos de los usuarios (*¿Cuál es la red social más segura?*, 2022), si bien en los últimos años, la red social ha hecho esfuerzos en reforzar la seguridad de sus usuarios (Martín Barrbero, 2022).

i. Las manifestaciones de la PDC en RRSS

Resumiendo, los hallazgos de Lim y Rasul (2022) en este ámbito, podemos decir que la PDC se manifiesta en RRSS principalmente de tres maneras: cognitiva, afectiva y conductual. La más sencilla de comprender es la manifestación conductual, que se traduce, por ejemplo, en el número de comentarios y publicaciones que un usuario realiza en un periodo de tiempo determinado.

Las manifestaciones cognitiva y afectiva, suelen ir ligadas a la conductual, y están asociados a conceptos marketinianos más amplios. En resumen, algunas de las formas concretas de este tipo de manifestaciones se dan cuando los usuarios dan “like” (“me gusta”), comparten o comentan en las publicaciones de una marca. En estos supuestos, ellos mismos están generando un vínculo con la marca, e incluso llegan a actuar de manera inconsciente como embajadores de la marca.

ii. Resultados de la PDC en RRSS

De entre las cuatro categorías de resultados identificadas por Lim y Rasul (2022), cabe destacar las siguientes:

- **1: Resultados relativos a los Negocios:** hace referencia a aquellas actividades empresariales que terminan influyendo -generalmente de manera positiva- en el rendimiento de la empresa, pudiendo medirse éste de distintas maneras, entre ellas, el rendimiento financiero, el rendimiento de mercado y los beneficios/rentabilidad netos de la empresa. Cabe señalar, que, en la mayoría de los casos, estos resultados trascienden al rendimiento de la marca, ya que una empresa puede poseer simultáneamente varias marcas. En todo caso, lo que trasciende del análisis efectuado por Lim y Rasul (2022) es que, la PDC en RSS, puede accionar actividades empresariales que mejoren el rendimiento de la empresa.
- **#2: Resultados relativos a la Marca:** hace referencia a las consecuencias que se derivan de la PDC en RRSS directamente para las marcas, que, con carácter general, además de ser positivas, son las más destacables. Concretamente, la PDC en RSS permite a las marcas conectar mejor con los clientes al permitirles construir una mejor imagen de marca y dotarles de capacidad de respuesta a ojos de los usuarios. Igualmente, la PDC en RRSS suscita, entre los usuarios, el apego y lealtad a la marca, la intimidad con la marca, la confianza en la marca, y el uso de la misma. A través de la PDC en RSS, las marcas pueden adquirir y desarrollar una sólida base de clientes que no solo sean conscientes y conocedores de sus marcas, sino que también puedan convertirse en buenos embajadores de sus marcas.

A la luz de todo lo expuesto, resulta evidente que las marcas deben comprender cómo interactuar con sus clientes en línea y qué factores tienen un impacto positivo en la PDC (Tavares y Nogueira, 2021), para así rentabilizar al máximo su presencia y actividad en RRSS (Santini et al., 2020).

CAPÍTULO 2. LAS PLATAFORMAS DE SVOD Y LAS RRSS

1. UN CAMBIO RADICAL EN LA MANERA DE CONSUMIR ENTRETENIMIENTO: LAS PLATAFORMAS DE SVOD

El consumidor contemporáneo, que vive rodeado por lo digital y lo tecnológico, ha cambiado radicalmente su manera de relacionarse con los medios de difusión y la industria del entretenimiento en los últimos años (Pérez, 2020). Concretamente, la aparición del servicio de *streaming* (“transmisión”) de Netflix en 2007, detonó la revolución en la industria del entretenimiento (McDonald y Smith-Rowsey., 2016).

A partir de Netflix, el consumidor, cansado de los paquetes y ofertas de los tradicionales servicios de televisión por cable, va a “migrar” hacia los servicios en línea de vídeo a la carta por suscripción (SVOD, del inglés “*Suscription Video On Demand*”), que le ofrecen contenidos originales y a precios más bajos (Pasirayi et al. 2021; Rataul et al. 2018).

Si bien los servicios de cable tradicionales, con sus emisiones en directo o de consumo compartido siguen estando disponibles, el consumidor contemporáneo se ha acostumbrado a tener acceso inmediato a una gran cantidad y variedad de contenidos multimedia y de *streaming*, que puede consumir, individualmente, cuando y donde quiera, gracias a los a los servicios SVOD (Pérez, 2020).

Strangelove (2015³) define el modelo de los servicios SVOD como un "buffet libre", y no es coincidencia que la industria se denomine “servicios de vídeo a la carta”. En efecto, los servicios SVOD ofrecen a sus suscriptores un amplio catálogo de productos de entretenimiento en formato vídeo (películas, programas de televisión, etc.), disponibles las 24 horas al día, 7 días a la semana, y accesibles tanto en pantalla de televisión como a través de diversos dispositivos móviles (*smartphones*, tabletas, ordenadores portátiles, etc.), lo que ha favorecido un consumo más independiente y móvil (Pérez, 2020).

2. EL “BOOM” DE LA INDUSTRIA DE LAS PLATAFORMAS SVOD

Amazon Prime Video y Netflix abrieron las puertas de la industria de los servicios SVOD, arrancando sus servicios de video *streaming* en 2006 y 2007 respectivamente. Tras el éxito del modelo de negocio de los citados precursores de la industria SVOD, un gran número de empresas dentro de la industria del entretenimiento lanzaron sus propios servicios de SVOD. Así vieron la luz, entre otras muchas plataformas de *streaming*, las de Hulu (en 2008) y HBO (en 2010, cuya plataforma SVOD ha pasado a llamarse Max en 2023); más recientemente las de Disney+ y Apple TV+ (ambas en 2019), y por último la de Peacock (en 2020) (Bond, 2019; Kim, 2022).

Las últimas incorporaciones a la industria SVOD han venido motivadas en gran parte por el cambio en el comportamiento digital global que trajo consigo la pandemia de la

³ Strangelove, M. (2015). *Post-TV: Piracy, cord-cutting, and the future of television*. [Post-TV: Piratería, corte de cables y el futuro de la televisión]. University of Toronto Press.

COVID-19. Y es que, las plataformas de SVOD se encuentran entre las principales beneficiarias de la pandemia.

Durante la pandemia, con pocas opciones para el entretenimiento, los consumidores acudieron en masa a las plataformas de SVOD. Y, si bien, la pandemia ya ha terminado, parece que muchos de los comportamientos digitales que comenzaron entonces, perduran aún hoy en día, pues el *streaming* sigue representando una parte cada vez mayor de la audiencia televisiva mundial (Kemp, 2023).

En efecto, según los últimos informes de GWI -empresa británica de investigación de audiencias- los servicios de SVOD (entre ellos los ofrecidos por Netflix y Disney+) representan ahora más del 45% del tiempo que los internautas en edad laboral dedican a ver la televisión, habiendo aumentado dicha cuota un 10% desde 2019. Por último, cabe mencionar que el usuario típico de Internet pasa ahora más de 1 hora y media al día consumiendo servicios de *streaming* y televisión online.

Todo parece indicar que el crecimiento que hasta ahora han conocido las plataformas de SVOD, va a continuar en los próximos años. Netflix -plataforma de SVOD más popular en el mundo- contaba con 231 millones de suscriptores mundiales a finales de 2022, y se espera que alcance 251 millones en 2027. Igualmente se proyecta que Amazon Prime Video y Disney+ (que ocupan la segunda y tercera posición en el ranking mundial de plataformas de SVOD más populares) pasen de sus respectivos 200 millones y 162 millones de suscriptores en 2022, a 250 millones y 210 millones en 2027 (Mena Roa, 2023; Orús 2022).

En resumen, la industria de los servicios SVOD es, sin duda, un mercado competitivo, dónde no cesa de crecer el número tanto de plataformas de *streaming* como de usuarios de las mismas.

3. LA IMPORTANCIA DE LA PRESENCIA DE LAS PLATAFORMAS DE SVOD EN RRSS

En un mercado tan saturado, las plataformas de *streaming* se ven obligadas a competir por la atención y la lealtad de unos usuarios que, normalmente, no pueden costearse la suscripción de todos y cada uno de los diferentes servicios SVOD disponibles, y que

parecen sentirse cada vez más “frustrados” y “fatigados” por la gran variedad de plataformas disponibles (Spangler, 2019).

Según, Rahe et al., (2021), las armas de que disponen las plataformas de SVOD para atraer suscriptores son principalmente dos: la calidad del contenido que ofrecen y la percepción de su marca. ¿Pero cómo usarlas?

Conviva (2021) -empresa de *Big Data*, especializada en la medición y análisis de audiencias de empresas en el sector de las telecomunicaciones, cuyos servicios son utilizados por las principales plataformas de *streaming*- informa que son cuatro los principales canales a través de los cuales los usuarios descubren nuevo contenido de entretenimiento: el boca a boca, la publicidad, las RRRSS y las recomendaciones que les hacen las propias plataformas de *streaming*.

Y es que, tal y como explica Pérez (2020), el cambio en la manera de consumir entretenimiento -hoy en día mucho más individual y dominada por las plataformas de SVOD- ha ido acompañado de un cambio en la manera de compartir dicho entretenimiento con otros consumidores.

En efecto, los espectadores hoy en día son proclives a expandir su experiencia de consumo de entretenimiento hasta las RRSS, donde no solo comparten sus hábitos de visionado y sus impresiones sobre el contenido, sino que interactúan directamente con los creadores de dicho contenido, adquiriendo más información, aportando contenido humorístico (por ejemplo, a través de “memes”) y comprometiéndose con la producción (Wohn y Na, 2011).

Pero los espectadores no son los únicos con una presencia significativa en RRSS. Y es que las cadenas de medios y los propios servicios SVOD favorecen cada vez más el marketing en RRSS -frente a canales de marketing tradicionales (Benes, 2020, Okazaki et al. 2015)- dónde tratan siempre de ser a la vez el mediador y el objeto de la discusión social en curso (Delgado et al., 2018).

Así pues, podemos decir que las RRSS, entre ellas Twitter, son un espacio de encuentro para las plataformas de *streaming* y sus suscriptores, donde ambas partes pueden interactuar, ya sea compartiendo impresiones acerca del contenido o formulando preguntas y respuestas (Pérez, 2020).

Fernández-Gómez y Martín-Quevedo (2018), han puesto de manifiesto con su estudio, que con las estrategias de marketing en RRSS, cuando consisten en favorecer las interacciones sociales entre las plataformas de *streaming* y sus suscriptores, pueden tener efectos positivos considerables para las marcas, sobre todo en lo que respecta a su posicionamiento, notoriedad y a la fidelización de sus suscriptores.

Concretamente, la interacción entre la marca y el consumidor en RRSS influye positivamente en la percepción, el valor y el conocimiento de la marca por el consumidor, favorece el boca a boca, y llega incluso a influir en posteriores decisiones de compra (Liu et al. 2016).

Todo ello, según Fernández-Gómez y Martín-Quevedo (2018), incrementaría la fidelidad de los suscriptores a las plataformas de SVOD, lo que debería traducirse en una mayor retención de los mismos y una disminución de las tasas de cancelación de servicios. De ahí que las RRSS se hayan convertido en una pieza clave en las estrategias de marketing de las plataformas de *streaming* (Pérez, 2020).

CAPÍTULO 3. APROXIMACIÓN TEÓRICA AL *SENTIMENT ANALYSIS* Y A SU APLICACIÓN A DATOS TEXTUALES OBTENIDOS DE TWITTER CON RSTUDIO

1. INTRODUCCIÓN AL *TEXT MINING*

1.1. ¿Qué es el *Text Mining*?

Como exponíamos anteriormente, es esencial que las empresas cuiden el tono empleado en su estrategia de comunicación en RRSS, sobre todo de cara a favorecer la PDC online de sus consumidores, por todos los beneficios que ésta conlleva, en particular los relativos a la marca y a los negocios.

Por ello, a aquellas empresas con presencia en RRSS les puede resultar de gran utilidad el analizar no solo el tono empleado por ellas en RRSS, sino la manera en que los usuarios reaccionan a dicho tono, por ejemplo, a través del *Sentiment Analysis*, técnica que utilizaremos en nuestro estudio de campo y sobre la que profundizaremos en el presente capítulo.

El *Sentiment Analysis* es una rama o aplicación del *Text Mining*, que, a su vez, es una forma de *Data Mining*. Según el centro de estudios tecnológicos Immune Technology Institute (*¿Qué son y para qué sirven el Text mining y Data mining?*, 2022), el *Text Mining* se podría definir como un proceso de análisis de datos que “abarca el análisis de datos textuales de toda índole (pertenecientes a diferentes medios, idiomas, etc.)” con el objetivo de identificar posibles patrones o correlaciones entre los términos del texto, y así lograr comprender nueva información que no estuviera necesariamente explícita en el texto.

1.2. El *Text Mining* paso a paso

Antes de entrar en materia, nos parece conveniente introducir de manera esquemática las etapas o pasos que se suelen seguir a la hora de llevar a cabo cualquier *Text Mining*, con independencia de su forma o del campo de aplicación del mismo, puesto que este esquema es el mismo que hemos seguido en nuestro estudio de campo.

Ilustración 1: El proceso de Text Mining paso a paso



Nota: Diagrama elaborado en base al blog *¿Qué son y para qué sirven el Text mining y Data mining?* (28 de abril de 2022)

a. Especial mención a la etapa de “Tokenización”

Así pues, el elemento distintivo del *Text Mining* es el tipo de datos que conforman su objeto: los datos textuales. ¿Pero por qué los textos, para su análisis merecen un tratamiento diferenciado dentro del *Data Mining*? Pues bien, simplemente por la dificultad de su procesamiento por parte de los algoritmos de *Machine Learning* debido al carácter no-estructurado de los datos textuales (Provost y Fawcett, 2013).

Esto no significa que un texto carezca de estructura, de hecho, el lenguaje verbal reposa sobre estructuras gramaticales y sintácticas: el orden de las palabras importa, y mucho. El carácter no-estructurado de un texto implica más bien que la estructura lingüística que resulta entendible para el hombre, no lo es para una máquina o algoritmo.

Así pues, todo proceso de *Text Mining* ha de pasar por una etapa de “Tokenización” para darle una estructura al texto, convertirlos en variables reconocibles y entendibles para el algoritmo -como por ejemplo una matriz- para el posterior procesamiento y análisis de los datos textuales (Provost y Fawcett, 2013).

Hay diversos métodos de Tokenización, pero nosotros hemos optado por el denominado *Bag of Words* (Bolsa de Palabras) por su simplicidad. Con este método, el documento se convierte en una “bolsa de palabras”, sin importar el orden de las mismas, ni la gramática, ni los signos de puntuación. Cada palabra se considerará individualmente como una variable potencial, a la que se asignará un valor numérico para que el algoritmo pueda interpretarla.

Dentro del método *Bag of Words*, hay a su vez dos posibles formas de asignar un valor a cada variable potencial. Estas serían, por una parte, la *Binary Representation* (Representación Binaria), según la cual, a la aparición de una palabra concreta en el texto se le asignará el valor 1, y a la “no aparición” (o ausencia) de esa misma palabra se le asignará el valor 0 (Provost y Fawcett, 2013).

Por otra parte, estaría la *Term Frequency Representation* (Representación de la Frecuencia del Término), según la cual, a cada palabra -o variable potencial- se le asigna el valor numérico de su frecuencia de aparición o repetición en el texto. Es decir, si la palabra aparece 5 veces, se le asignará un valor de 5 (Provost y Fawcett, 2013).

Nosotros optaremos por esta última en nuestra aplicación práctica, para evitar posibles pérdidas de información, ya que la frecuencia de una palabra en un texto determinado puede resultar relevante, pudiendo afectar a la carga emocional del texto.

b. Especial mención a la etapa de preprocesamiento y limpieza de los datos

Además, debemos tener en cuenta que, a la complejidad gramatical y sintáctica de cada lengua (conjugaciones, palabras sinónimas, homógrafas, etc.), debemos añadir otros factores como la mezcla de idiomas, las faltas de ortografía, la semántica, el contexto o subtexto de cada palabra, el uso de abreviaciones...

De ahí que todo proceso *Text Mining* requiera pasar, además de por la tokenización, por una etapa de preprocesamiento -para extraer los elementos más importantes o representativos del texto- y/o de limpieza de los datos -para eliminar cualquier elemento

que pueda conducir a error o que no aporte valor al análisis. Dependiendo del análisis y del enfoque convendrá aplicar unas técnicas u otras.

Estas fases de preprocesamiento y limpieza suelen confundirse como etapas de “reducción del texto”, pudiendo darse por separado o simultáneamente, antes de la etapa de tokenización o incluso a la vez que ésta, dependiendo de las herramientas de *Machine Learning* empleadas.

i. Formas de preprocesamiento de los datos textuales

El preprocesamiento de los datos textuales está enfocado a reducir el texto conservando únicamente elementos esenciales y representativos, que aporten verdadero valor al análisis que se desea hacer de los mismos. Entre las posibles técnicas que tienen cabida en esta etapa, podemos destacar las siguientes:

- El *stemming*: consiste en reducir una palabra a su raíz o lexema mediante la eliminación de sufijos y prefijos (Provost y Fawcett, 2013). Esta técnica tiene dos posibles limitaciones, conocidas como *overstemming* o *understemming*, que se darían cuando, respectivamente, las palabras se acortan demasiado o no lo suficiente. En el *overstemming*, palabras que pertenecen a distintas familias léxicas quedan reducidas bajo un mismo término (p. ej. reducir “bandana” y “bandido” a “band”). Por el contrario, en el *understemming*, palabras de una misma familia léxica no llegan a reducirse a la raíz común, y se procesan como términos diferenciados (p. ej. reducir “reducción” a “reduc” y “reductor” a “reduc”).
- La lematización: consiste en reducir cada palabra a su forma primitiva -que no a su raíz- por lo que implica un procesamiento más sofisticado que tenga en cuenta no solo prefijos y sufijos, sino elementos gramaticales, como pueden ser las conjugaciones verbales (Provost y Fawcett, 2013).
- La eliminación de *stop-words*: consiste en eliminar las palabras más comunes o frecuentes del idioma del texto, pues se trata de palabras “vacías”, es decir, que aportan poco valor para el análisis. Además, pueden darse *stop-words* propias a la industria, temática o contexto en el que se realice el análisis (p. ej. la abreviación

“rt” en Twitter, o las palabras “película” o “serie” en la industria de las plataformas SVOD).

ii. Posibles operaciones de limpieza de los datos

Caben mencionar las siguientes técnicas: transformar todas las palabras a mayúsculas o minúsculas, por ejemplo, para no distinguir entre nombres propios o comunes; eliminar números y dígitos si no aportan valor; eliminar todos los signos de puntuación.

En el caso de *Text Mining* aplicado a Twitter, otras posibles técnicas que puede ser de utilidad serían, entre otras: eliminar direcciones webs (URLS); eliminar emoticonos; eliminar tweets duplicados; eliminar signos frecuentes en Twitter como puede ser la almohadilla “#”, la arroba “@” o el *ampersand* “&”; filtrar los tweets por idiomas.

1.3. Tipos o formas de *Text Mining* según el análisis

a. *Word Frequency Analysis*

El *Word frequency Analysis* es una de las técnicas de *Text Mining* más utilizadas, sobre todo para el análisis de grandes volúmenes de datos textuales. Su principal función es identificar las palabras más repetidas en el texto, así como establecer posibles patrones y relaciones entre ellas.

Aunque no resulta el objeto principal de nuestro estudio, también hemos empleado esta técnica, como punto de partida, para entender mejor la composición de nuestros datos textuales, de cara a realizar una limpieza y reducción de los mismos que nos permita obtener unos datos lo más afinados posibles para el posterior análisis de sentimiento.

b. *Sentiment Analysis*

i. El *Sentiment Analysis* y sus técnicas

Como explicábamos antes, hay determinados factores o elementos del lenguaje que, si bien nosotros entendemos de manera natural, no resultan entendibles a primera vista para los algoritmos de *Machine Learning*. La intencionalidad o emoción subyacente en un texto forma parte de estos elementos lingüísticos más complejos y “ajenos” al lenguaje computacional. No obstante, hay técnicas como el *Sentiment Analysis* que permiten

precisamente captar la carga emocional de un texto con ayuda de algoritmos (Provost y Fawcett, 2013).

El *Sentiment Analysis* puede abordarse principalmente de dos maneras distintas. Se puede hacer desde un enfoque predictivo: se construye un corpus de textos, etiquetando manualmente cada texto en base a su carga sentimental (positiva, negativa o neutra) o emocional (un nivel de detalle mayor en el sentimiento, hace referencia a emociones concretas como la felicidad, la ira, la tristeza, etc.), y con éste se entrena un algoritmo de clasificación que sea capaz de “predecir” la carga emocional de un nuevo texto en base a la similitud que guarda con la estructura lingüística registrada en textos anteriores.

Otra forma de análisis sería partir de un lexicón de sentimiento -ya sea de elaboración propia o pre-existente, como el lexicón de emociones NRC- para determinar la carga sentimental y/o emocional de un texto. Las principales ventajas de este método frente al anterior son, primero, que no se debe pasar por un proceso de prueba y error y de entrenamiento del algoritmo, y segundo, en caso de utilizar un lexicón preexistente, nos ahorramos el trabajo de etiquetación manual.

¿Pero en que consiste este método? Un lexicón de sentimiento es una suerte de diccionario o glosario de emociones, que recoge un listado de palabras y el sentimiento y/o emociones asociadas a cada una de ellas en el lenguaje natural. El análisis de sentimiento según este método consistiría en proporcionarle al algoritmo, por una parte, el texto a analizar, y por otra el lexicón deseado. Si el algoritmo detecta en el texto a analizar alguna de las palabras listadas en el lexicón proporcionado, entonces nos devolverá la carga emocional que dicha palabra tiene asociada en el lexicón.

Como podemos observar es una forma particular de *Term Frequency Representation*, en la que a cada palabra se le va a asociar un valor numérico, no solo en base a su aparición en el texto, sino también en base a la carga emocional de la palabra.

Ahora bien, esta técnica tiene sus limitaciones, entre ellas que el análisis va a estar enteramente supeditado a las palabras recogidas en el lexicón (si la palabra no aparece en nuestro diccionario, no podremos asignarle una carga sentimental y/o emocional), y que, al tomar como variable potencial cada palabra, no seremos capaces de captar la negación de un sentimiento (p. ej. “No estoy nada contenta” puede llevar asociada una carga

sentimental positiva al detectarse la palabra “contenta” a pesar de que el verdadero sentimiento subyacente es lo contrario).

Aun así, por las ventajas que presenta el segundo método, que ya hemos expuesto, éste será el que utilicemos en el análisis que expondremos más adelante. En concreto, usaremos el lexicón de emociones NRC: un lexicón muy utilizado, sobre todo en el campo de la investigación académica y la docencia, debido a su carácter público y a su incorporación en la mayoría de los entornos y lenguajes de programación (entre ellos RStudio).

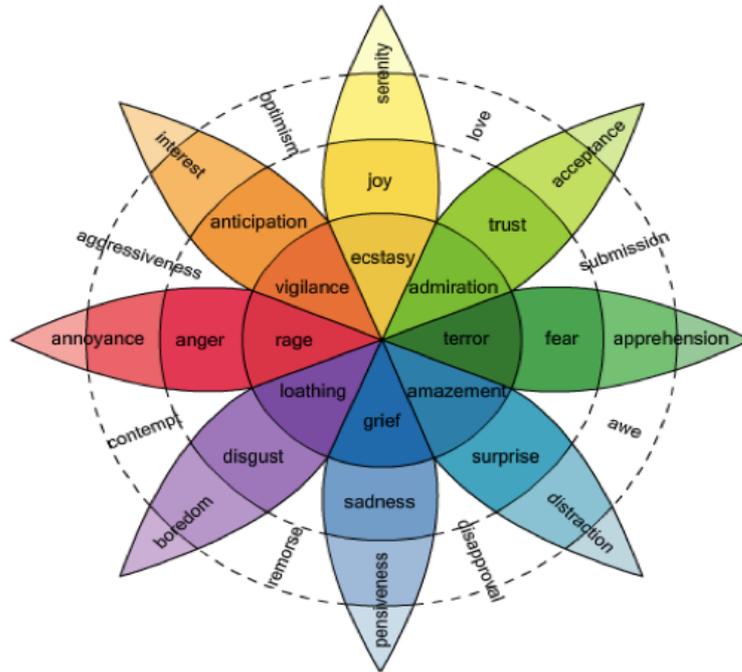
ii. El lexicón de emociones NRC

El lexicón de emociones NRC, publicado en julio de 2011 y originalmente en inglés, fue elaborado por los doctores Saif M. Mohammad y Peter Turney, quienes se apoyaron en el *crowdsourcing* para llevar a cabo la fase de anotación inicial o etiquetación de las palabras en base a su sentimiento y emoción (Mohammad y Turney, 2013). En Agosto de 2022 se publicó la última actualización del lexicón, quedando disponible en un total de 108 idiomas diferentes (Mohammad, *s.f.*).

El método de asociación de la carga sentimental o polaridad (positiva/negativa) según este lexicón es el siguiente. Si la palabra aparece en el texto, y ésta tiene asociada un sentimiento positivo, se le asigna un valor 1 para la categoría “positivo” (ahora bien, si la palabra aparece 5 veces en el texto, a cada repetición le habrá sido asociado un 1, por lo que, en total, habrá un valor positivo total de 5). Lo mismo aplica para el sentimiento negativo. Finalmente, la carga sentimental o polaridad del texto se obtiene sustrayendo el total de valores “negativos” al total de valores “positivos” (si hay 5 palabras positivas y 8 negativas, el texto tendrá una polaridad de -3).

Pero lo más interesante de este lexicón es que las 14.182 palabras en él listadas -que suman un total de 25.000 acepciones o sentidos- llevan asociadas no solo un sentimiento (positivo o negativo), sino también en muchos casos una de las 8 emociones básicas de Robert Plutchik (a saber, *anger, fear, anticipation, trust, surprise, sadness, joy, and disgust* [ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y aversión]). Esto permite añadir un matiz extra a la hora de valorar la carga sentimental de un texto.

Ilustración 2: Rueda de las emociones de Robert Plutchik



Nota: Ilustración obtenida de Mohammad y Turney (2013).

Llegados a este punto, y dado que se trata de uno de los pilares del lexicón de emociones NRC, conviene detenerse brevemente a resumir la teoría de las emociones del psicólogo Robert Plutchik. Esta teoría gira entorno a lo que Plutchik denominó “Rueda de las emociones” (**Ilustración 2**), compuesta por ocho emociones básicas, a saber: la ira, el miedo, la anticipación, la confianza, la sorpresa, la tristeza, la alegría y la aversión.

El psicólogo defendía que, en la vida real, las emociones están relacionadas entre sí, de ahí que las personas suelen experimentar varias emociones de manera simultánea, aunque no se den todas en la misma intensidad, pudiendo una predominar sobre las demás. Estos solapamientos entre emociones básicas abocaban a lo que Plutchik llamaba “díadas” que a su vez podían ser primarias, secundarias o terciarias (Alabau, 2021).

2. SENTIMENT ANALYSIS APLICADO A TWITTER CON RSTUDIO

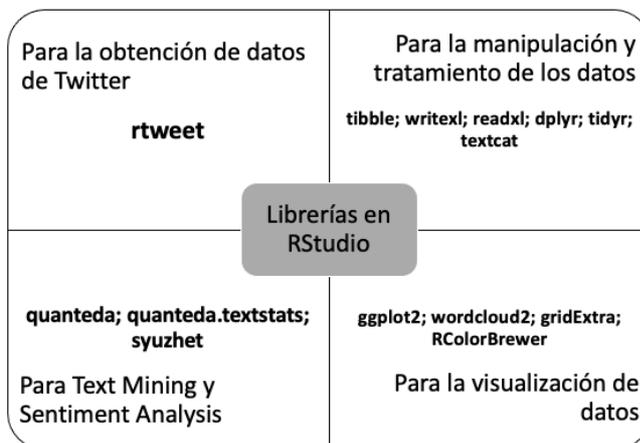
2.1. RStudio como herramienta para llevar a cabo un *Sentiment Analysis*

A nuestro parecer, el entorno y lenguaje de programación de RStudio resulta idóneo para llevar a cabo un proceso de *Text Mining*, y más en particular de *Sentiment Analysis*, sobre todo aplicado a datos extraídos de Twitter por varios motivos.

Primero por la facilidad de dicho lenguaje y su amplia utilización en el mundo de la investigación académica y la docencia. Segundo, por la posibilidad de emplear el entorno para extraer datos de Twitter a través de una API. Tercero, por la cantidad y variedad de herramientas que ofrece el entorno la manipulación y procesamiento de datos, y para la aplicación de *Text Mining* y *Sentiment Analysis*.

En relación con este último ámbito, cabe señalar que RStudio, incorpora hasta tres lexicones de sentimiento diferentes,: además del NRC que ya hemos analizado, están AFINN (que asigna a cada palabra un valor entre -5 y 5 en función de la intensidad de su polaridad positiva/megativa) y BING (asigna a cada palabra, o bien el valor -1, si su polaridad es negativa, 0 si es neutra, o bien 1 si es positiva).

Ilustración 3: Ejemplos de librerías disponibles en RStudio



2.2. Twitter como fuente de datos para *Sentiment Analysis*

La red social Twitter es sin duda una fuente de datos muy interesante de cara a aplicar técnicas de *Sentiment Analysis*, fundamentalmente por dos motivos: el potencial marketiniano de la red social y la idoneidad de los datos de Twitter para el tipo concreto de análisis.

En cuanto al potencial marketiniano de Twitter cabe señalar que, desde su aparición, la plataforma tuvo una acogida muy rápida entre la población mundial: en 2016, ya contaba con más de 313 millones de usuarios activos al mes en todo el mundo (Zimbra et al., 2018), y en 2022 ese número ha ascendido a 556 millones (Kemp, 2023). Ciertamente es que Twitter ya no es la red social más utilizada del mundo, pero su uso sigue siendo

considerable, sobre todo entre la población adulta. Y mayor prueba de su interés económico es el hecho de que la plataforma fuera adquirida por Elon Musk el pasado octubre de 2022, por un valor de \$44 000 millones (Southern, 2022).

En cuanto a la idoneidad de los datos de Twitter para ser objeto de un Análisis de Sentimiento, debemos comenzar por explicar cómo se genera y comparte contenido en esta plataforma. Según Zimbra et al. (2018) podemos definir Twitter como una plataforma de microblogging en la que los usuarios son autores de “tuits”, en los que expresan sus opiniones acerca de un tema de interés. Estos tuits quedan publicados en el *timeline* (blog personal) de cada usuario, al cual tendrán acceso sólo sus seguidores (si el autor del tuit tiene una cuenta privada), o todos los demás usuarios de Twitter (si la cuenta del autor es pública).

Debemos resaltar que, en Twitter, a diferencia de RRSS más populares -como Instagram o TikTok- las palabras siguen teniendo tanto peso como las imágenes, ya que la comunicación entre los usuarios sigue siendo principalmente verbal a través de los “tuits”, que en el fondo son breves textos. Ciertamente, estos textos pueden ir acompañados de imágenes, vídeos o enlaces, pero el elemento verbal sigue siendo clave dentro del tuit.

Además, estos textos son cortos (actualmente se permite una longitud máxima de 280 caracteres), por lo que su extracción, almacenamiento y manipulación resultan muy cómodos. De ahí que Twitter ya haya sido objeto de muchas investigaciones recientes sobre análisis de sentimiento (Zimbra et al., 2018).

Por otra parte, hasta hace poco, el acceso a los datos -que se hace a través de las API⁴ de la plataforma- era prácticamente ilimitado y gratuito, al menos para cuestiones de investigación académica. Bastaba con solicitar una cuenta de desarrollador de Twitter, introducir un nombre de usuario y contraseña, crear una aplicación -pudiendo crear un número ilimitado de aplicaciones- y conectarse a las API de Twitter. Con ello se tenía acceso a toda la información pública en Twitter.

⁴ Una API es una pieza de código que permite a diferentes aplicaciones comunicarse entre sí y compartir información y funcionalidades.

No obstante, el acceso a los datos públicos” -y más desde la reciente adquisición de la plataforma- se ha ido restringiendo y monetizando. Actualmente, y desde el 18 de mayo de 2023, el plan gratuito que ofrece Twitter a los desarrolladores ya no permite descargar tuits. Con el plan básico -suscripción de pago mensual, que cuesta 100 dólares al mes- sólo permite crear una única aplicación, y no permite descargar o “leer” más de 10.000 tuits al mes. Además, para cada solicitud o interacción con la plataforma -por ejemplo, una descarga o búsqueda de tuits-, obliga a los desarrolladores a autenticarse con sus credenciales propias (la “API key” y el “bearer token” asociados a una aplicación de desarrollador) que se obtienen al crear la aplicación con la cuenta de desarrollador.

Igualmente, al buscar tuits por palabras clave específicas, si bien es cierto que no hay límite de volumen, sí, que hay un límite temporal: pudiendo únicamente descargar los tuits publicados entre 6 y 9 días antes de la fecha de extracción de los datos. En RStudio, esto se hace a través de la función `search_tweets` (del paquete `rtweet`), que, por defecto, no devuelve más de 18.000 tuits por solicitud.

Y, del mismo modo, si se pretenden extraer los tuits publicados por cuentas específicas sólo se obtendrá una muestra de volumen muy limitado, que no podrá exceder los 3.200 tuits. En RStudio, esto se hace con la función `get_timelines` (también del paquete `rtweet`), que por defecto extraerá los últimos publicados.

CAPÍTULO 4. ESTUDIO DE CAMPO: *TWITTER SENTIMENT ANALYSIS* APLICADO AL CASO DE LAS PLATAFORMAS SVOD CON RSTUDIO

1. DEFINICIÓN DEL MARCO DE NUESTRO ANÁLISIS

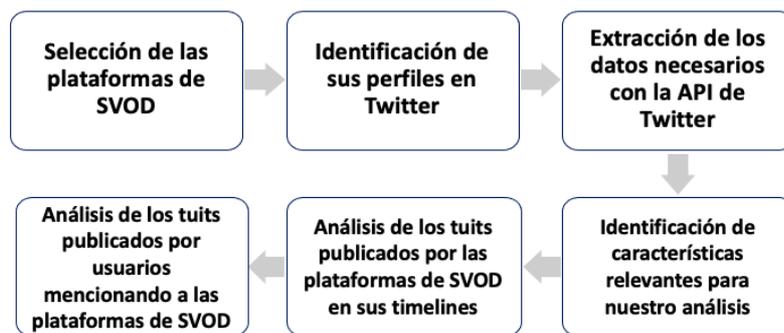
Teniendo en consideración todo lo expuesto en los anteriores capítulos, a continuación, pasamos a reproducir el estudio de campo que hemos efectuado con el objetivo de identificar posibles patrones o correlaciones en las interacciones de las plataformas de SVOD con sus usuarios en RRSS. En particular, queremos analizar de qué manera los antecedentes de la PDC en RRSS relativos al mensaje -en particular el tono y la carga emocional de los mensajes- afectan a la manifestación de la PDC en RRSS.

Como muestra de los antecedentes de la PDC en RRSS relativos al mensaje analizaremos los tuits publicados por determinadas plataformas de SVOD en Twitter (en adelante Tuits, no confundir con tuits, ver **Glosario**), en un periodo temporal determinado que

definiremos más adelante. Puesto que nos interesa sobre todo analizar el tono y la carga emocional empleada por dichas plataformas en sus tuits, al margen de realizar un *Word Frequency Analysis* sobre los mismos, realizaremos igualmente un *Sentiment Analysis*.

Como manifestación de la PDC en RRSS analizaremos los tuits publicados por usuarios individuales de Twitter -en un periodo de tiempo determinado- en los que mencionan de manera directa, es decir, que apelan directamente a las mismas plataformas de SVOD (en adelante Menciones). Nos centraremos sobre todo en el volumen de los mismos y de nuevo, en el tono y carga emocional que los usuarios ponen en sus tuits (*Sentiment Analysis*). En este sentido, partimos de la presunción que los usuarios que interactúan de manera tan directa con las plataformas de SVOD en Twitter, si no son consumidores efectivos de los servicios ofrecidos por éstas, al menos son consumidores potenciales.

Ilustración 4: Resumen de las etapas seguidas en nuestro estudio de campo

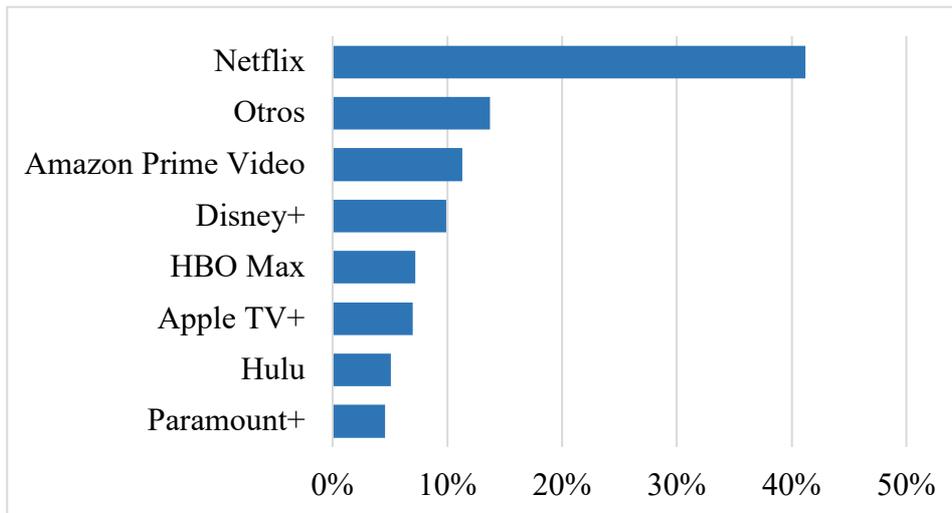


2. LOS DATOS EMPLEADOS EN NUESTRO ANÁLISIS

2.1. Primera selección de las plataformas de SVOD e identificación de sus perfiles en Twitter

Teniendo en cuenta que la industria de las plataformas SVOD es muy competitiva, el primer paso es decidir cuáles de todas las plataformas queremos seleccionar para nuestro estudio. Nosotros hemos optado por usar la cuota de mercado mundial (medida en base a la demanda mundial de contenidos originales) como criterio de selección (**Ilustración 5**). Así, en un primer momento vamos a extraer los datos de las 7 plataformas con mayor demanda global de contenidos originales, a saber: Netflix, Amazon Prime Video, Disney+, HBO Max (), Apple TV+, Hulu, Paramount+. Luego podremos ir acotando más la selección de plataformas en función de la calidad de los datos extraídos.

Ilustración 5: Cuota de mercado mundial de las principales plataformas SVOD en 2022



Nota: Cuota de mercado mundial medida en base a la demanda mundial de contenidos originales.

Fuente: Orús (2023) para Statista.

El siguiente paso es identificar los perfiles de estas plataformas en Twitter, puesto que debemos saber cuáles son sus nombres de usuario exactos (**Ilustración 6**) para poder extraer sus respectivos Tuits y Menciones. Nosotros vamos a centrarnos en los perfiles principales -no los regionales- de estas plataformas, que suelen encargarse de la estrategia de comunicación global y, por tanto, es probable que generen un mayor tráfico de tuits.

Ilustración 6: Resumen de los perfiles de Twitter de las plataformas SVOD (Abril de 2023)

| Plataforma | Nombre de usuario | Creación | Seguidores | Tuits |
|--------------------|-------------------|-----------|---------------|-----------|
| Netflix | @netflix | Oct. 2008 | 21,9 millones | 49,4 mil |
| Amazon Prime Video | @PrimeVideo | Nov. 2008 | 3,2 millones | 73,9 mil |
| Disney+ | @DisneyPlus | Abr. 2018 | 5,5 millones | 43,3 mil |
| HBO Max | @hbomax | Ago. 2010 | 1,3 millones | 24,3 mil |
| Hulu | @hulu | Jun. 2008 | 1,1 millones | 151,2 mil |
| Apple TV+ | @AppleTVPlus | Feb. 2019 | 186,4 mil | 4,6 mil |
| Paramount + | @paramountplus | Oct. 2017 | 372,4 mil | 88 mil |

Nota: A partir de Mayo de 2023 se hacen efectivos los siguientes cambios de marca: HBO Max pasa a llamarse Max, y migra su perfil de Twitter a @StreamOnMax; Apple TV+ migra su perfil de Twitter a @AppleFilms. Nosotros descargamos los tuits de los perfiles antes del cambio de marca.

2.2.Extracción de los Tuits y Menciones con la API de Twitter desde RStudio

Una vez identificados los nombres de usuario en Twitter de las plataformas SVOD que queremos estudiar, procedemos a descargar desde RStudio sus respectivos Tuits y Menciones. Se incluye, como **Anexo #4** al trabajo, el código empleado para la descarga, con comentarios explicativos.

Llegados a este punto, y antes de entrar más en detalle, conviene señalar que nuestra cuenta de desarrollador de Twitter está adscrita al plan gratuito, por lo que actualmente no podemos descargar datos. Afortunadamente, antes de que entraran en vigor los términos y condiciones descritos en el capítulo anterior, fuimos capaces de hacer una descarga de datos suficiente para realizar nuestro trabajo de campo.

La descarga la hemos hecho gracias al paquete **rtweet**, concretamente con las funciones **get_timelines** y **search_tweets**, cuyos funcionamiento y limitaciones hemos expuesto en el capítulo anterior, y a los cuales nos referimos. Además, utilizamos el paquete **tibble** para el procesamiento de los datos, y el paquete **writexl** para volcar los datos en un archivo Excel y así almacenarlos fuera de la interfaz de RStudio (para su posterior difusión e incluso procesamiento en otros entornos).

Puesto que tuvimos que hacer 14 descargas de datos en total -las dos clases de tuits, para cada una de las siete plataformas SVOD- y a su vez volcar todos los datos en ficheros Excel, automatizamos el proceso de descarga y escritura de los datos en Excel definiendo las siguiente funciones propias en RStudio. El código para estas funciones podrá ser utilizadas en el futuro para otros posibles objetos de estudio.

Ilustración 7: Funciones definidas en RStudio para la descarga de los tuits

```
# Para los tuits publicados por las plataformas SVOD en sus timelines

crear_excel_timeline <- function(plataforma, fecha) {
  df_timeline <- get_timeline(c(plataforma), n=3200, token=auth)
  path <- '/Users/fatima/Desktop/TFG Analytics/DDBB/get_timelines/'
  namefile <- paste0(path, plataforma, '_', fecha, '.xlsx')
  write_xlsx(df_timeline, namefile) }

# Para los tuits publicados por otros usuarios mencionando a las plataformas SVOD

crear_excel_menciones <- function(plataforma, fecha){
  id_plataforma <- paste0('@', plataforma)
  df_menciones <- search_tweets(id_plataforma, n = Inf, token = auth)
  path<-' /Users/fatima/Desktop/TFG Analytics/DDBB/search_tweets/'
  namefile<-paste0(path,'menciones_', plataforma,'_',fecha, '.xlsx')
  write_xlsx(df_menciones,namefile)}
```

Luego, para descargar los datos, bastará con ejecutar las funciones (**Ilustración 8**), dándole los inputs deseados, a saber: en lugar de 'plataforma', introducimos el nombre de usuario en Twitter (sin el signo @) de la plataforma SVOD en cuestión; y en lugar de 'fecha' podremos introducir lo que queramos, por ejemplo, la fecha de descarga de los datos, de cara a identificarlos en el futuro. Muy importante que escribamos estos inputs entre comillas simples, para que R lo interprete como elementos de tipo 'character'.

Ilustración 8: Ejemplo de ejecución de las funciones para la descarga de los datos

```
crear_excel_timeline('netflix','19Apr')
crear_excel_timeline('PrimeVideo','19Apr')
crear_excel_timeline('DisneyPlus','19Apr')
crear_excel_timeline('hbomax','19Apr')
crear_excel_timeline('hulu','19Apr')
crear_excel_timeline('AppleTVPlus','19Apr')
crear_excel_timeline('paramountplus','19Apr')

crear_excel_menciones('netflix','19Apr')
crear_excel_menciones('PrimeVideo','19Apr')
crear_excel_menciones('DisneyPlus','19Apr')
crear_excel_menciones('hbomax','19Apr')
crear_excel_menciones('hulu','19Apr')
crear_excel_menciones('AppleTVPlus','19Apr')
crear_excel_menciones('paramountplus','19Apr')
```

Nota: Cabe mencionar que las funciones tardan un tiempo en ejecutarse.

2.1. Caracterización de los datos para nuestro análisis

Una vez descargados los datos, y antes de proceder a su análisis, vamos a realizar un primer visionado y manipulación de los mismos en RStudio para hacernos una idea del volumen y la calidad los tuits que tenemos a nuestra disposición. Esta etapa es fundamental para identificar posibles necesidades de limpieza y reducción de los datos, o incluso para acotar aún más el marco u objeto del análisis, de cara a mejorar la calidad y congruencia del mismo. Los códigos de R empleados para esta sección se incluyen, como **Anexo #5** y **#6** al trabajo

a. Identificación de las variables más relevantes

Tras cargar los datos en RStudio, lo primero que observamos es que, tanto para los Tuits como para las Menciones, cada línea u observación corresponde a un tuit, y que *a priori*,

tenemos más o menos el mismo volumen de Tuits y de Menciones para cada plataforma SVOD.

Ilustración 9: Dimensiones de los Datasets de Menciones y Tuits para cada plataforma SVOD

| | |
|---------------------------|---------------------------|
| 🔵 Menciones_AppleTVPlus | 9810 obs. of 43 variables |
| 🔵 Menciones_DisneyPlus | 9909 obs. of 43 variables |
| 🔵 Menciones_hbomax | 9906 obs. of 43 variables |
| 🔵 Menciones_hulu | 9915 obs. of 43 variables |
| 🔵 Menciones_netflix | 9915 obs. of 43 variables |
| 🔵 Menciones_paramountplus | 9915 obs. of 43 variables |
| 🔵 Menciones_PrimeVideo | 9915 obs. of 43 variables |
| 🔵 Tweets_AppleTVPlus | 3187 obs. of 43 variables |
| 🔵 Tweets_DisneyPlus | 3200 obs. of 43 variables |
| 🔵 Tweets_hbomax | 3200 obs. of 43 variables |
| 🔵 Tweets_hulu | 3195 obs. of 43 variables |
| 🔵 Tweets_netflix | 3169 obs. of 43 variables |
| 🔵 Tweets_paramountplus | 3199 obs. of 43 variables |
| 🔵 Tweets_PrimeVideo | 3191 obs. of 43 variables |

Además, con la descarga de datos de Twitter, hemos obtenido, a parte del “cuerpo” o texto de los tuits, un montón de datos en torno a los tuits -en total 43 variables (las columnas)- de los cuales no todos aportan información relevante. Nosotros vamos a conservar únicamente 4 de las 43 variables, a saber:

- **id**: identificador numérico anonimizado del usuario de Twitter que publicó el tuit
- **source**: fuente desde la que publicó el tuit (entre otras: iPhone, Android, Web app, Media Studio, TweetDeck)
- **text**: el “cuerpo” o texto del tuit
- **created_at**: fecha y hora de publicación del tuit

b. Definición del marco temporal de los tuits que serán objeto de análisis

Ahora debemos analizar las fechas de publicación de los Tuits y Menciones para definir una franja temporal para el análisis.

Ilustración 10: Franja temporal de los Tuits de cada plataforma SVOD

| | Plataforma_SVOD | Fecha_Minima | Fecha_Maxima |
|---|-----------------|---------------------|---------------------|
| 1 | netflix | 2022-09-09 15:07:57 | 2023-04-19 18:15:32 |
| 2 | AppleTVPlus | 2022-05-24 02:06:26 | 2023-04-04 23:29:16 |
| 3 | DisneyPlus | 2022-06-11 16:16:16 | 2023-04-19 18:00:56 |
| 4 | hbomax | 2022-04-05 18:00:00 | 2023-04-19 19:03:00 |
| 5 | hulu | 2023-02-19 17:49:58 | 2023-04-19 18:35:48 |
| 6 | paramountplus | 2022-12-20 21:43:22 | 2023-04-19 19:05:58 |
| 7 | PrimeVideo | 2022-09-11 14:30:51 | 2023-04-19 18:03:47 |

Como podemos observar en la **Ilustración 10** los tuits de Paramount+ y Hulu comienzan en Diciembre de 2022 y Febrero de 2023, mientras que todas las demás plataformas tienen publicaciones desde Septiembre de 2022 y hasta de abril de 2023. Por ello las descartamos.

De entre las restantes plataformas vamos a filtrar los datos para conservar únicamente los tuits publicados entre el 9 de noviembre de 2022 y el 4 de abril de 2023 (**Ilustración 11**). Observamos que se reduce considerablemente el número de tweets, sobre todo para HBO Max que se queda en pocos Tuits por encima de 1000. Por ello vamos a descartar también esta última plataforma. Por tanto, nuestro análisis se centrará en Netflix, Amazon Prime Video, Apple TV+ y Disney +.

Ilustración 11: Nueva franja temporal y volumen de Tuits para cada plataforma SVOD

| | Plataforma_SVOD | Fecha_Minima2 | Fecha_Maxima2 | Volumen_Tuits |
|---|-----------------|---------------------|---------------------|---------------|
| 1 | netflix | 2022-11-09 08:01:00 | 2023-04-04 23:03:23 | 2026 |
| 2 | AppleTVPlus | 2022-11-09 16:01:26 | 2023-04-04 23:29:16 | 1421 |
| 3 | DisneyPlus | 2022-11-09 16:20:05 | 2023-04-04 20:01:44 | 1286 |
| 4 | hbomax | 2022-11-09 01:16:40 | 2023-04-04 22:00:01 | 1050 |
| 5 | PrimeVideo | 2022-11-09 01:44:43 | 2023-04-04 23:07:30 | 2022 |

Nótese además que hay dos niveles en el volumen de Tuits. En el primer nivel, con un mayor volumen (por encima de 1.500 Tuits), estarían Netflix y Amazon Prime Video - que son, además, las dos plataformas de SVOD con mayor cuota de mercado. En un segundo lugar, estarían AppleTV+ y Disney+, con un volumen más reducido de Tuits (entre 1.000 y 1.500), teniendo además todas ellas cuotas de mercado similares, en torno al 6% (ver **Ilustración 5**).

Por otra parte, en lo que respecta a las Menciones, tal y como se observa en la tabla *infra* (**Ilustración 12**) nos encontramos con una franja temporal mucho más reducida, de apenas unos días. Pero también más homogénea -con una amplitud mínima de 2 días y máxima de 10 para cada plataforma- por lo que en este caso no necesitaremos filtrar por fechas. Así la franja temporal para las Menciones queda se fija entre el 9 y el 19 de abril de 2023.

El hecho de que las franjas temporales de los Tuits y las Menciones no sean idénticos ciertamente afecta a nuestro análisis, pues se reduce el nivel de correlación que pudiera haber entre el tono y la carga sentimental empleados por las plataformas SVOD y los empleados por sus consumidores en Twitter. No obstante, teniendo en cuenta que ambas franjas sí guardan cierta relación, en tanto la franja temporal de las Menciones se inscribe dentro de la franja de los Tuits, coincidiendo con el final de ésta. Y teniendo igualmente en cuenta que, debido a las restricciones que Twitter ha puesto en el acceso a los datos públicos, no podremos obtener una franja temporal más precisa, procederemos en nuestro estudio con los datos de los que disponemos.

Ilustración 12: Franja temporal y volumen de Menciones para cada plataforma SVOD

| | Plataforma_SVOD | Fecha_Minima3 | Fecha_Maxima3 | Volumen_Menciones |
|---|-----------------|---------------------|---------------------|-------------------|
| 1 | netflix | 2023-04-17 23:09:16 | 2023-04-19 19:15:55 | 9915 |
| 2 | AppleTVPlus | 2023-04-09 23:30:57 | 2023-04-19 19:40:46 | 9810 |
| 3 | DisneyPlus | 2023-04-16 20:25:34 | 2023-04-19 19:27:37 | 9909 |
| 4 | hbomax | 2023-04-15 20:04:41 | 2023-04-19 19:31:10 | 9906 |
| 5 | PrimeVideo | 2023-04-14 18:42:25 | 2023-04-19 19:21:28 | 9915 |

- c. Mayor acotación del objeto de análisis de las Menciones en función de la fuente de publicación del tuit

Recordemos que, con el análisis de las Menciones, queremos analizar tuits publicados por consumidores potenciales de los servicios ofrecidos por SVOD. Por ello, nos parece conveniente filtrar las Menciones de cada plataforma SVOD, en base a la fuente desde dónde publicase el tuit (variable **source** en nuestros datos).

Concretamente conservaremos solamente aquellas Menciones que se publicaran desde un dispositivo electrónico inteligente (iPhone, iPad o Android), o desde la aplicación Web de Twitter, asumiendo que los usuarios que publican desde tales interfaces son, con

mucha seguridad, individuos y por tanto consumidores potenciales de los servicios SVOD. El número de observaciones se ha reducido muy poco tras aplicar el filtro, y queda reproducido en la tabla a continuación (**Ilustración 13**).

Ilustración 13: Nuevo volumen y franja temporal de las Menciones para cada plataforma SVOD

| | Plataforma_SVOD | Volumen_Menciones | Fecha_Minima | Fecha_Maxima |
|---|-----------------|-------------------|---------------------|---------------------|
| 1 | netflix | 9833 | 2023-04-17 23:09:16 | 2023-04-19 19:15:55 |
| 2 | AppleTVPlus | 9653 | 2023-04-09 23:30:57 | 2023-04-19 19:40:46 |
| 3 | DisneyPlus | 9815 | 2023-04-16 20:25:34 | 2023-04-19 19:27:37 |
| 4 | PrimeVideo | 9755 | 2023-04-14 18:42:25 | 2023-04-19 19:21:28 |

- d. Características de los tuits a tener en cuenta para el análisis
 - i. Consideraciones generales relativas a los caracteres

En esta sección vamos a realizar un primer análisis del cuerpo de los Tuits y Menciones, y sobre todo visualizar la literalidad de los textos, de cara a identificar posibles características de los mismos que habremos de tener en cuenta a la hora de preprocesar los datos para el *Sentiment Analysis*. Con una visualización rápida de una pequeñísima muestra de Tuits publicados por Netflix (**Ilustración 14**) ya podemos hacer las siguientes observaciones:

- Conviene unificar todo el texto en minúsculas o mayúsculas para evitar errores en caso de que los algoritmos y diccionarios utilizados en el análisis fueran *case sensitive*, es decir que al algoritmo procesa como palabras diferentes aquellas que contienen mayúsculas y minúsculas.
- Habrá que eliminar del cuerpo de los tuits: signos de puntuación; los dígitos; los caracteres especiales (entre ellos emoticonos, y otros muy frecuentes en Twitter, por

ejemplo, las '@' de las menciones⁵ y '#' de los *hashtags*⁶); y las URLs o direcciones webs, que normalmente comienzan por la frecuencia de caracteres 'https://'. Esto se debe a que pueden inducir a error en el procesamiento del texto.

- Haciendo esto es posible que algunos tuits queden “vacíos” de contenido, no obstante, no vamos a eliminarlos porque queremos que sigan siendo parte del conjunto de tuits observados, de cara a calcular frecuencias de repetición de palabras y otras proporciones. Incluso se podría argumentar que estos tuits vacíos, si bien no tienen carga polaridad positiva o negativa, si tienen una carga neutra.

Ilustración 14: Visualización del cuerpo de los 10 primeros Tuits de Netflix

```
> print(Tweets_netflix[4])
# A tibble: 2,026 × 1
  text
  <chr>
1 "@SouthWalton ok but what about Watching Selling Sunset?????"
2 "The fourth (and formerly final) Shrek film is now on Netflix! http
s://t.co/E5ZeTOYldZ"
3 "RT @fxckoklahoma: NETFLIX I LOVE YOU 🥹🥹 https://t.co/7I18wb3tVA"
4 "This is one of the realest things anyone has ever said on Love is B
lind. Chelsea deserves the world. https://t.co/hJM9B4uzlK"
5 "RT @Teta_2023: Not Tiffany admitting she has sleeping problem durin
g her casting 🥹🥹#LoveIsBlind #LoveIsBlindS4 #LoveisBlind4 htt...
6 "@Need4LSpeed our forever #CoupleGoals"
7 "RT @contodonetflix: To sign the record deal of their dreams, these
artists have to impress @YandelOficial, @RauwAlejandro, @Tainy,...
8 "Sandra Bullock gives one of her best performances ever in 28 Days.
https://t.co/Juow9zemvD"
9 "RT @strongblacklead: My Name is Mo'Nique is now streaming on Netfli
x! https://t.co/q2BgBPgB3C"
10 "RT @moworldwide: WHAT A MOMENT STANDING ON THIS STAGE!!! WE MADE I
T!!! \r\nI LOVE US 4REAL & 4LIFE❤️❤️❤️ https://t.co/BP60CiUFMV"
```

ii. El bigrama 'RT' y los re-tuits

En la ilustración anterior también parece que el bigrama 'RT', muy característico de Twitter, aparece con mucha frecuencia, por lo que conviene profundizar un poco más en ello.

⁵ En Twitter se puede apelar públicamente a otro usuario (“mencionarlo” en jerga tuitera) incluyendo en el cuerpo del tuit su nombre de usuario precedido por el signo @.

⁶ Conjunto de caracteres precedidos por una almohadilla (#) que se incluyen en el cuerpo de un mensaje, sobre todo en las webs de microblogging como Twitter, y sirven de identificador o etiqueta del mensaje.

El bigrama 'RT' en Twitter, que figura al inicio de muchos tuits, es indicativo de que el mensaje es una forma particular de publicación, denominada re-tuit en español, que consiste en compartir nuevamente un tuit ya publicado, por otro usuario o por uno mismo.

Twitter ofrece a los usuarios la función 'Retweet' para generar este tipo de mensaje de manera automática, quedando el bigrama 'RT' incorporado en el cuerpo del tuit, aislado, a modo de apertura, y seguido de una mención (ver nota a pie de página 6) al usuario que publicó el tuit original. Una vez explicado texto, caben hacer dos puntualizaciones.

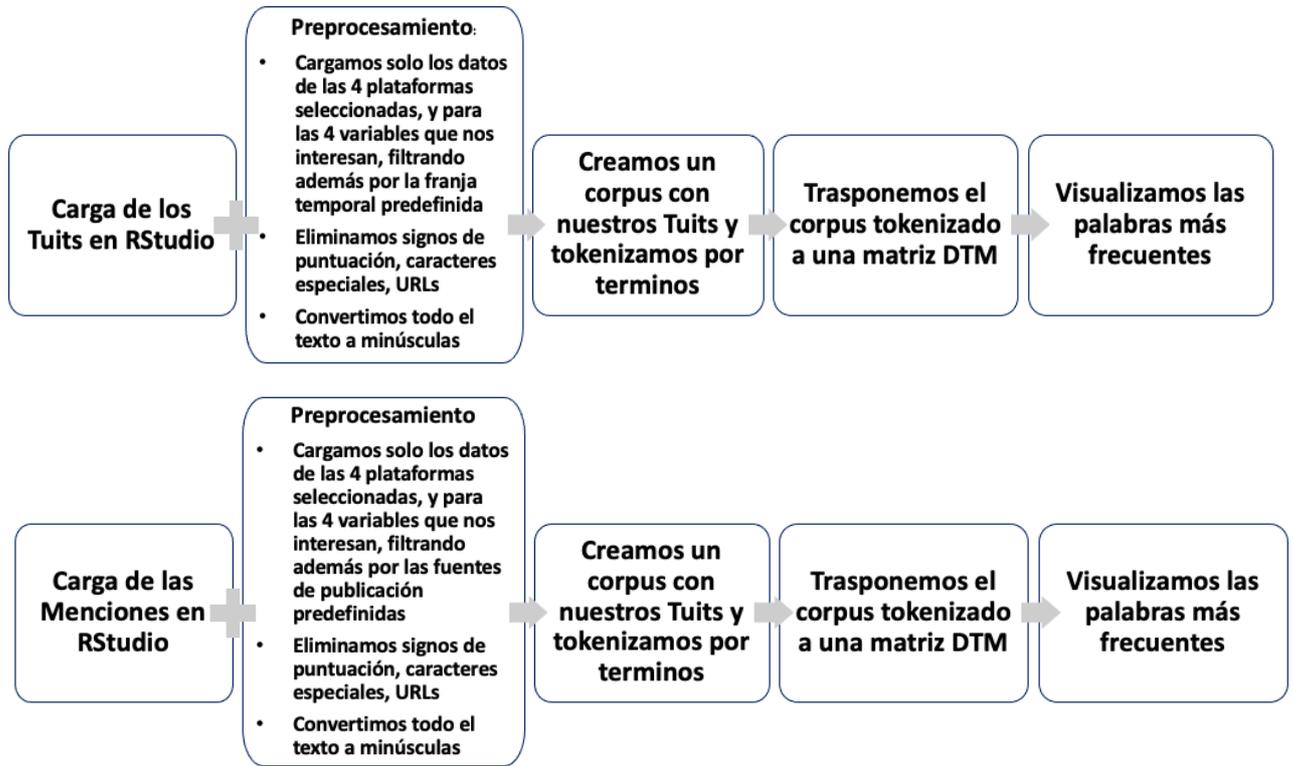
Primero, que debido al concepto de re-tuit es muy posible que el texto de muchos tuits esté duplicado, no obstante, no vamos a eliminar duplicados porque aunque el texto siga siendo el mismo, lo ha compartido otra persona, porque le ha parecido importante o se ha sentido identificada con la carga emocional del tuit en cuestión. Por tanto, su re-difusión de un tuit previo añadiría a la carga sentimental universal de los conjuntos de tuits observados.

En segundo lugar, podemos observemos que el bigrama 'rt', queda individualizado, como si se tratase de una palabra propia, a pesar de no se usa fuera de Twitter, pues carece de significado en los idiomas que utilizan el alfabeto latino. Además, por este mismo motivo, podemos asumir que el bigrama 'rt' solo aparecerá una vez por tuit, por lo que su frecuencia de repetición global en el corpus de tuits es una aproximación bastante fiable del volumen de re-tuits contenido en el corpus.

Así las cosas, debemos señalar que el bigrama 'RT' no aporta contenido emocional al texto, por lo que no aporta valor de cara a un Sentiment Analysis. No obstante, puede ser interesante analizar su frecuencia -como complemento a un análisis de sentimiento-, puesto que, el volumen de re-tuits -presente tanto en nuestra colección de Tuits como de Menciones-, puede servir de métrica del nivel de interacción entre las plataformas SVOD y sus consumidores.

Por eso hemos realizado un primer proceso de *Text Mining* en Rstudio, cuyo código de R se detalla en el **Anexo #6**. Concretamente un *Word Frequency Analysis*, cuyas etapas se detallan en la **Ilustración 15**.

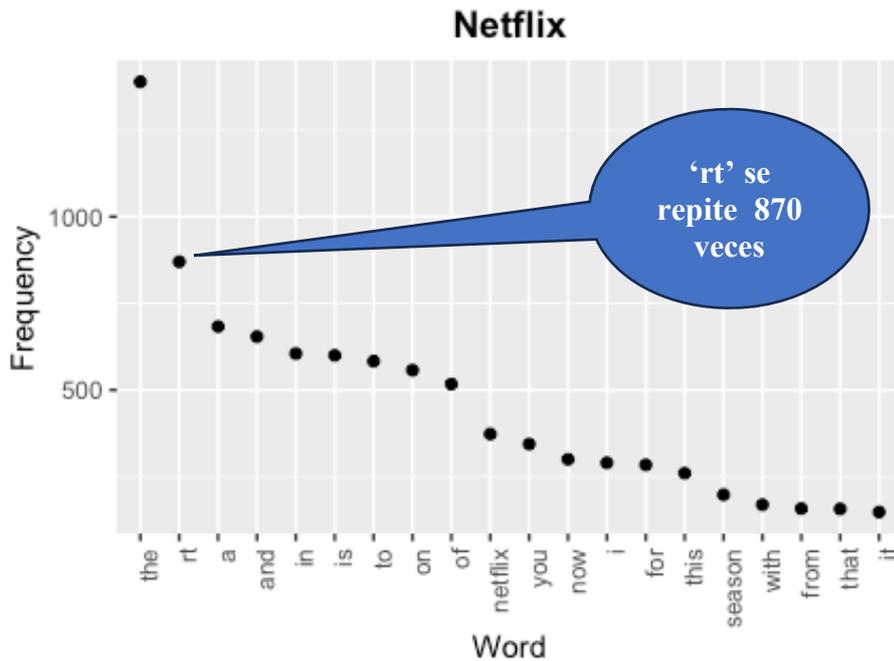
Ilustración 15: Etapas del *Word Frequency Analysis* realizado para los Tuits y Menciones



En este caso hemos nos hemos abstenido de relizar *stemming* o lematización, y también de eliminar las *stop-words* de los tuits, puesto que, como queremos simplemente analizar la frecuencia de repetición del bigrama ‘RT’, estas herramientas no eran necesarias.

Gracias a las herramientas de visualización observamos que, efectivamente, el bigrama RT se encuentra entre las palabras más frecuentes, tanto en los Tuits como en las Menciones, y para las cuatro plataformas observadas. A modo de ejemplo, se incluyen las **ilustraciones 16 y 17**.

Ilustración 16: Las 20 palabras más repetidas en los Tuits de Netflix



Nota: el universo de Tuits analizado comprende los 2.206 tuits publicados por la plataforma Netflix entre el 9 de noviembre de 2022 y el 4 de abril de 2023.

Ilustración 17: Nube de palabras con las palabras más repetidas en las Menciones de Amazon Prime Video

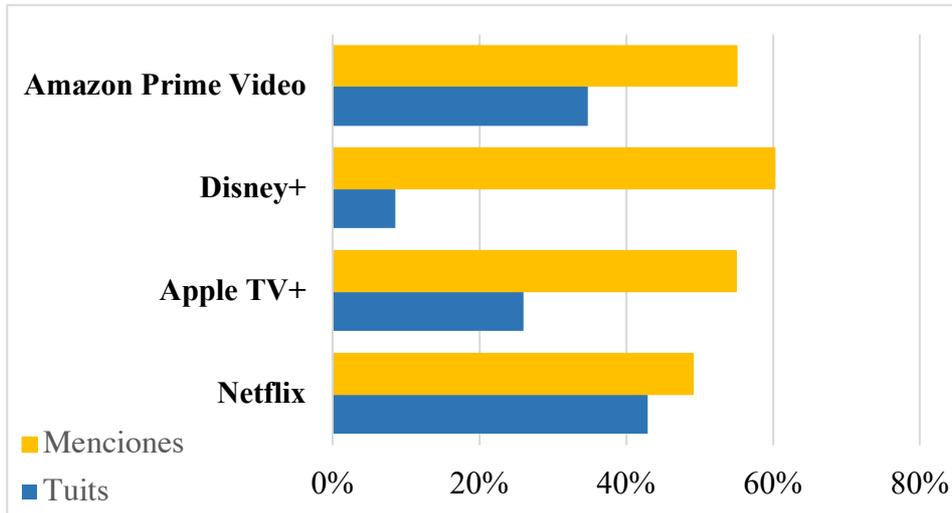


Nota: el universo de Menciones analizado comprende los 9.915 tuits, publicados entre el 14 y el 19 abril de 2023, en los que se menciona a Amazon Prime Video. Vemos que el bigrama 'rt' es sin duda el término más frecuente, con 6.495 repeticiones.

Ahora, recordemos la premisa expuesta anteriormente de que la frecuencia de repetición del binomio 'RT' es una aproximación bastante fiable del volumen de re-tuits que hay en cada uno de los 8 conjuntos de tuits (uno de Tuits y otro de Menciones, para cada una de las cuatro plataformas SVOD observadas). Podemos entonces obtener una aproximación

de la proporción de re-tuits presente en cada uno de los corpus de tuits (**Ilustración 18**), calculada como el cociente entre la frecuencia de repetición del binomio ‘RT’ y el número total de observaciones en cada conjunto de tuits.

Ilustración 18: Proporción de re-tuits en los Tuits y Menciones de cada plataforma SVOD



Nota: Proporción de retuits calculada como la frecuencia de repetición del binomio ‘RT’ entre el número total de observaciones para cada conjunto de tuits (Tuits y Menciones, para cada plataforma). El número de total de observaciones empleado son los que figuran en las Ilustraciones 10 y 12 para los Tuits y Menciones respectivamente.

Así, observamos las plataformas SVOD han adoptado con intensidad diferente la práctica de retuitear contenido de otros usuarios. La que más contenido “no-original” comparte en Twitter es Netflix, con una proporción de re-tuits de algo más del 40%, y la que menos es Disney+, con menos una proporción inferior al 10%.

Igualmente, podemos observar que las plataformas SVOD, consideradas como usuarios individuales hacen mucho menos uso de la función “Retweet”, que sus consumidores, puesto que la proporción de re-tuits es mucho mayor en los conjuntos de Menciones que en los de Tuits, para todas las plataformas observadas, siendo en todos los casos de al menos el 50%.

3. *SENTIMENT & EMOTION ANALYSIS* CON EL LEXICÓN NRC

3.1. Cuestiones previas

Antes de entrar en detalle, no está de más recordar cual será el objetivo de nuestro análisis: identificar posibles patrones, relativos al tono y a la carga emocional de los tuits, en la interacción en Twitter de las plataformas de SVOD con sus consumidores potenciales..

Para ello vamos a realizar un Sentiment Analysis de dos fuentes de datos: por una parte, los Tuits publicados por dichas plataformas entre el 9 de noviembre de 2022 y el 4 de abril de 2023; por otra parte, las Menciones a las mismas plataformas, publicadas entre el 9 y el 19 de abril de 2023, desde un iPhone, un iPad, un Android o desde la Web App de Twitter.

Las plataformas sobre la que se centra el estudio son Netflix, Amazon Prime Video, Apple TV+ y Disney+, por ser las que cuentan una mayor cuota de mercado y por ser las que generaron un mayor tráfico de mensajes en Twitter para el periodo observado

Conviene igualmente recordar la diferencia entre polaridad o sentimiento y emociones. Mientras que la polaridad o sentimiento se centra en atribuir al texto una carga meramente positiva, negativa o neutra, el análisis de emociones tiene más profundidad y permite analizar el peso que tiene un determinado conjunto de emociones en el texto.

Para ambos casos nosotros utilizaremos, en el entorno RStudio, el método de Lexicón NRC, que nos permitirá atribuir a cada conjunto de tuits para cada plataforma, no solo una polaridad, sino el peso que en ellas tiene cada una de las 8 emociones de Plutchik (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y aversión). El código de R empleado en el análisis se adjunta como **Anexo #7**.

Ilustración 19: Esquema del proceso observado para el análisis de los Tuits

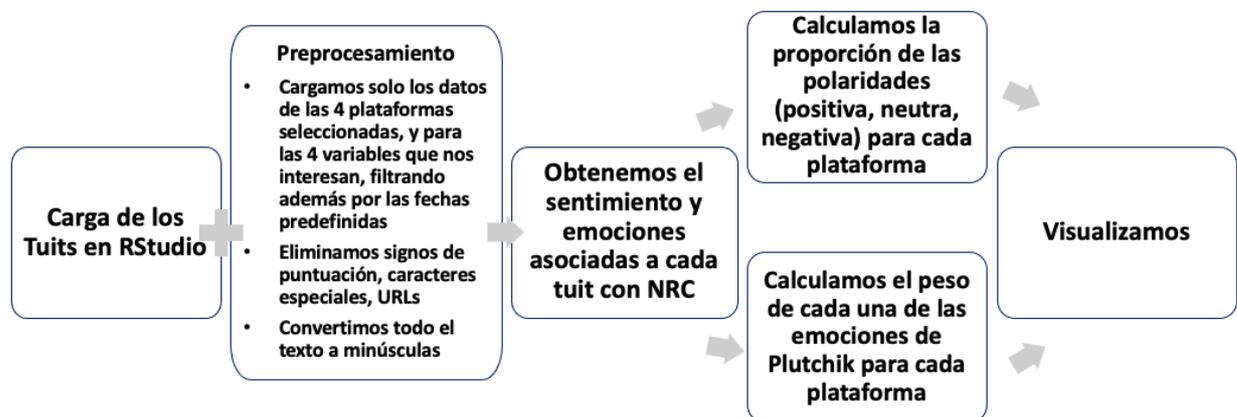
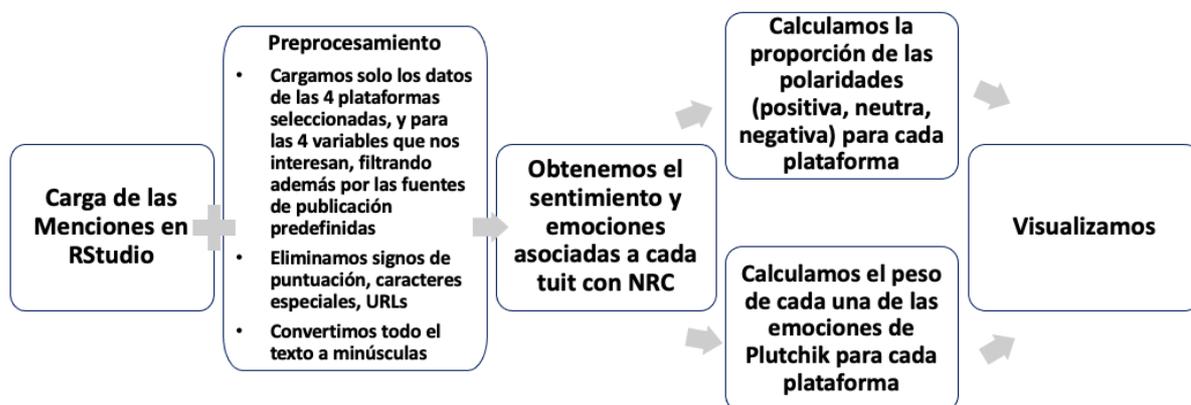


Ilustración 20: Esquema del proceso observado para el análisis de las Menciones



3.2. Análisis de la polaridad o sentimiento y su visualización

a. Para los Tuits publicados por las plataformas SVOD

Con carácter general, parece que las plataformas de SVOD tienden a publicar tuits con un tono sobre todo neutro, típico en comunicaciones meramente informativas (p.ej. informar del lanzamiento de un nuevo producto en la plataforma). Con una proporción muy cercana, tendríamos los mensajes de carga positiva: mensajes alabando las virtudes de un nuevo producto (p.ej. retuitear una crítica positiva que un seguidor ha publicado acerca de contenido disponible en la plataforma), o generando expectación ante un nuevo lanzamiento, lo cual es típico de las campañas de marketing.

Amazon Prime Video destaca por ser la plataforma más neutral, siendo también la que ofrece el catálogo de contenido más adulto, y la mayor oferta de documentales. Por el contrario, la plataforma que tiene a la vez la mayor carga positiva y la menor carga negativa es Disney+, probablemente por su posicionamiento en el mercado como una marca “family-friendly”.

Ilustración 21: Proporción de las polaridades asociadas a los Tuits de cada plataforma

| Plataforma | Volumen Tuits | Positiva | Neutra | Negativa |
|--------------------|---------------|--------------|--------------|--------------|
| Netflix | 2.026 | 38,2% | 48,6% | 13,2% |
| Amazon Prime Video | 2.022 | 31,4% | 59,6% | 9,0% |
| Apple TV+ | 1.421 | 45,5% | 40,2% | 14,4% |
| Disney+ | 1.286 | 47,4% | 46,0% | 6,5% |
| MEDIA | 1.689 | 40,6% | 48,6% | 10,8% |

b. Para las Menciones a las plataformas SVOD

En cuanto a las Menciones, merece la pena destacar que la muestra de tuits disponible tiene un volumen casi 6 veces mayor. Y observamos que, a mayor tamaño de la muestra, más parece aumentar el peso de la polaridad neutra, probablemente, porque se amplifica el efecto neutralizador de las *stop-words*.

Aquí, destaca la “negatividad” de las menciones directas a Netflix, lo que podría indicar que la plataforma y su contenido son objeto de mayores críticas en las redes, o simplemente, que hacen publicaciones acerca de temáticas o géneros que tienen connotaciones negativas (p.ej. contenido provocador, películas de terror o de violencia, etc.).

Igualmente cabe destacar la quasi-idéntica positividad de Amazon Prime Video y de Apple TV+, ambas en el primer puesto en cuanto a carga positiva de sus Menciones. Y por último, observamos que a contrario de lo que sucedía con los Tuits, en este caso, los las menciones a Disney+ son las más neutras de la muestra.

Ilustración 22: Proporción de las polaridades asociadas a las Menciones de cada plataforma

| Plataforma | Volumen Menciones | Positiva | Neutra | Negativa |
|--------------------|-------------------|----------|--------|----------|
| Netflix | 9.833 | 37,2% | 47,4% | 15,4% |
| Amazon Prime Video | 9.755 | 40,1% | 52,8% | 7,1% |
| Apple TV+ | 9.653 | 40,2% | 51,4% | 8,3% |
| Disney+ | 9.815 | 26,4% | 63,9% | 9,7% |
| MEDIA | 9.764 | 36,0% | 53,9% | 10,1% |

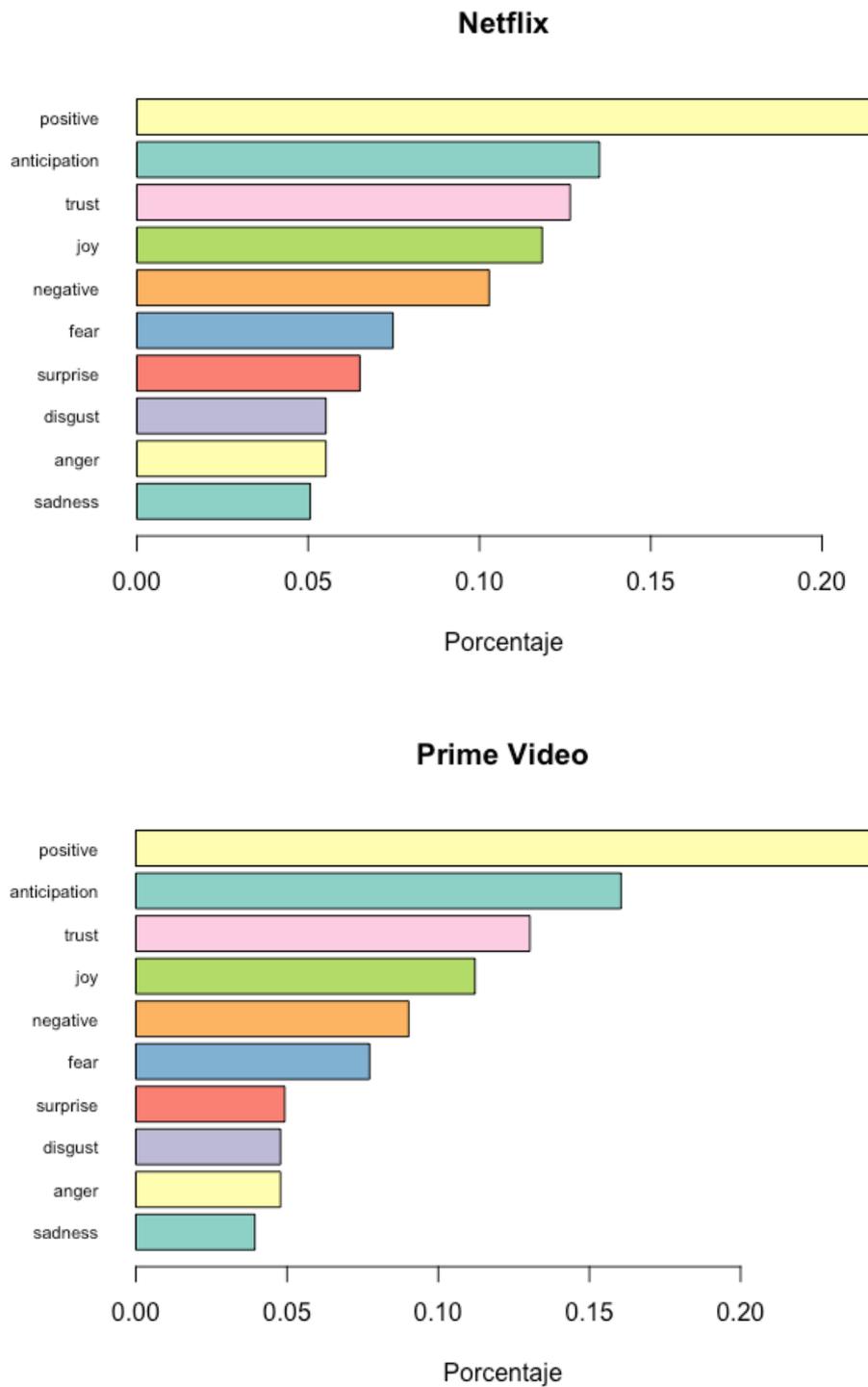
3.3.Análisis de las emociones y su visualización

a. Para los Tuits publicados por las plataformas SVOD

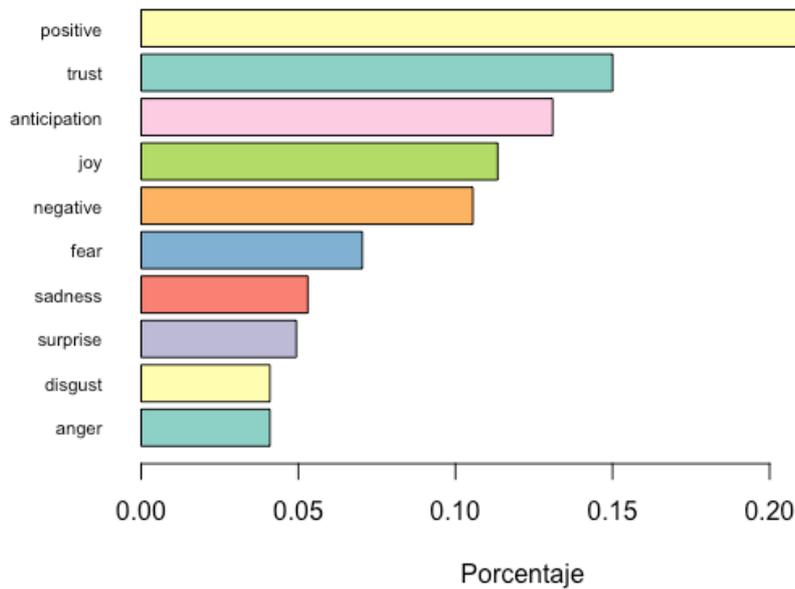
Las emociones con mayor peso para Netflix y Amazon Prime Video son la anticipación la confianza y la alegría, por ese orden, todas ellas de connotación positiva (que es la emoción más presente). Para Apple TV+ y Disney+ son las mismas tres emociones

positivas pero en distinta posición. A su vez, las emociones negativas o más “neutras” como puede ser la sorpresa, tienen un peso mucho menor en los tuits de todas las plataformas. Todo ello estaría en consonancia con el análisis de polaridad realizado, en el cual todas las plataformas han tenido mayor polaridad positiva que negativa.

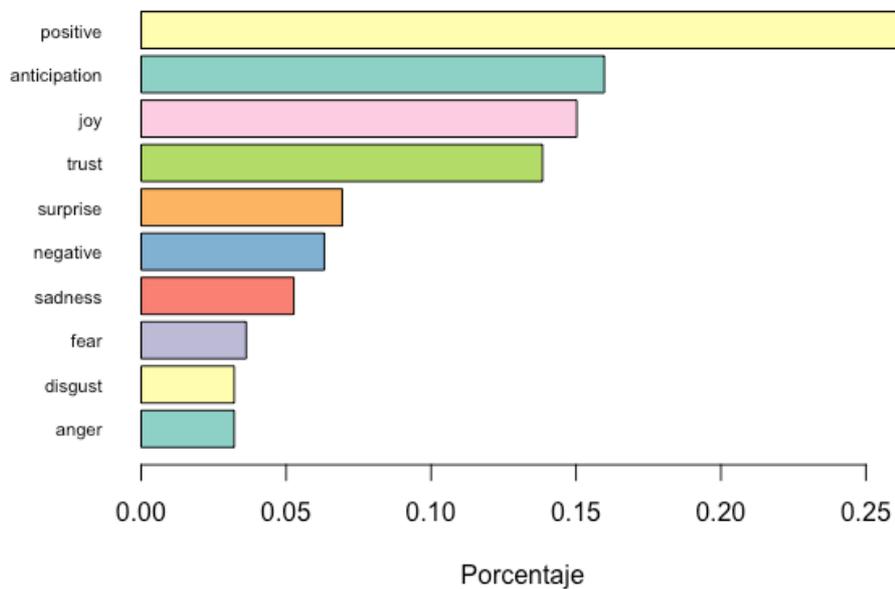
Ilustración 23: Peso de las emociones en los Tuits de las distintas plataformas



AppleTVPlus



DisneyPlus

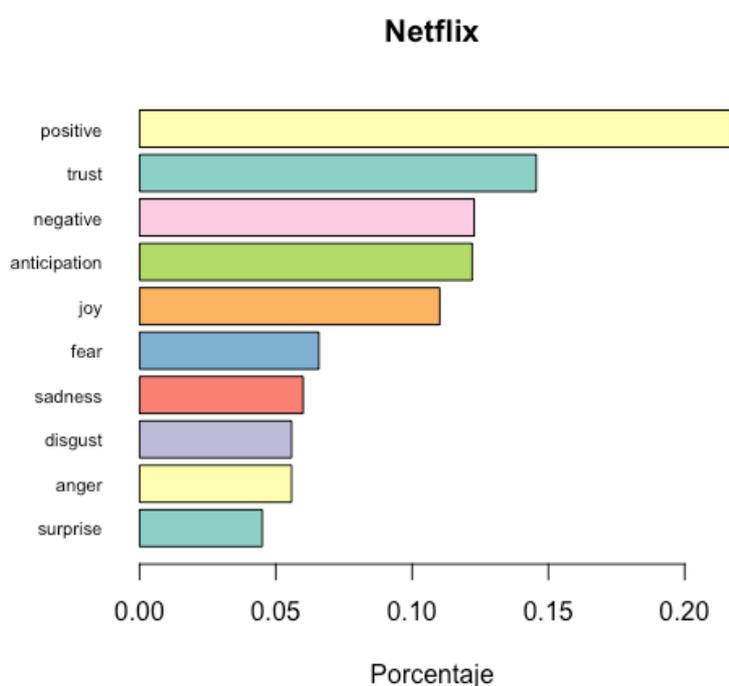


b. Para las Menciones a las plataformas SVOD

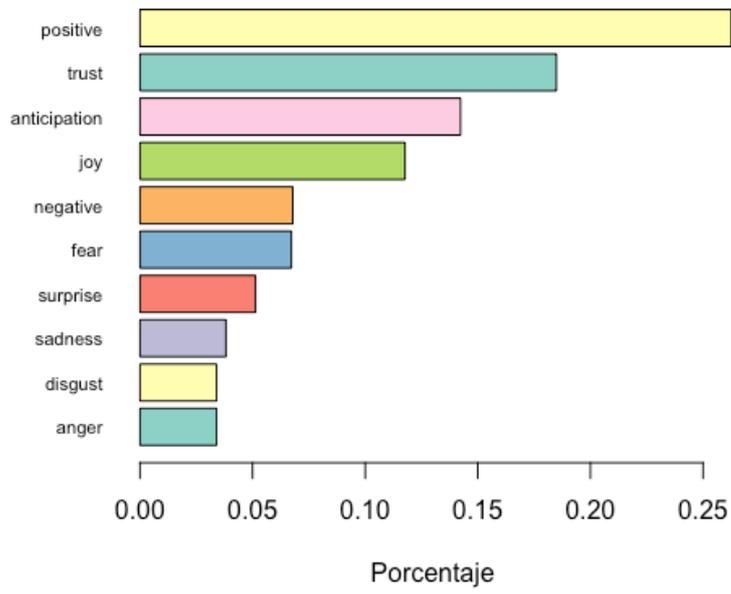
En el caso de las Menciones, las tres emociones preponderantes para todas las plataformas SVOD, siguen siendo la confianza, la anticipación y la alegría -si bien no necesariamente por ese orden. Para Amazon Prime Video, Apple TV+ y Disney+, sería aplicable lo

mismo expuesto en el epígrafe anterior, en cuanto a que existe una gran diferencia entre las emociones de connotación positiva y las de connotación negativa. No obstante, en el caso de Netflix no podemos decir lo mismo, aquí, siendo el escalón que separa las emociones de connotación positiva y negativa mucho menor para el líder del sector, algo que de nuevo está en consonancia con el análisis de polaridad, en el que se apreciaba que las Menciones a Netflix presentaban una mayor polaridad negativa que las demás. Por último, cabe destacar el peso que tiene la anticipación en las menciones a Disney+, lo que podría indicar que el público en Twitter se muestra entusiasta hacia los estrenos de Disney+.

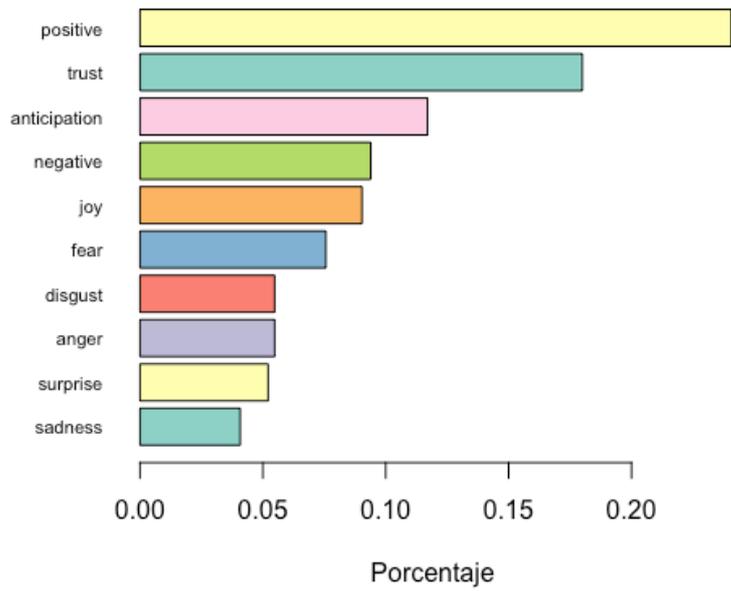
Ilustración 24: Peso de las emociones en las Menciones a las distintas plataformas



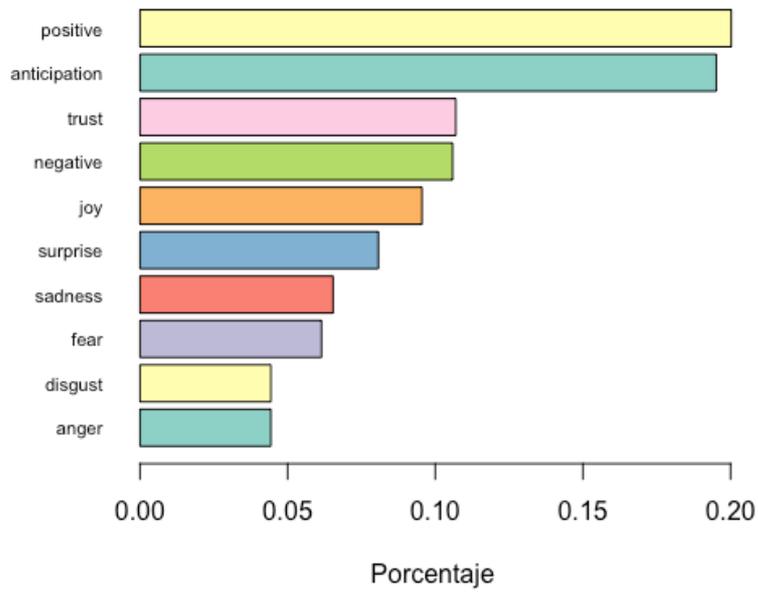
Prime Video



AppleTVPlus



DisneyPlus



CONCLUSIONES

- Las plataformas SVOD han adoptado con intensidad diferente la práctica de retuitear contenido de otros usuarios. La que más contenido “no-original” comparte en Twitter es Netflix, con una proporción de re-tuits de algo más del 40%, y la que menos es Disney+, con menos una proporción inferior al 10%.
- Las plataformas SVOD, consideradas como usuarios individuales hacen mucho menos uso de la función “Retweet”, que sus consumidores.
- Con carácter general, parece que las plataformas de SVOD tienden a publicar tuits con un tono sobre todo neutro, típico en comunicaciones meramente informativas (p.ej. informar del lanzamiento de un nuevo producto en la plataforma). Con una proporción muy cercana, tendríamos los mensajes de carga positiva: mensajes alabando las virtudes de un nuevo producto (p.ej. retuitear una crítica positiva que un seguidor ha publicado acerca de contenido disponible en la plataforma), o generando expectación ante un nuevo lanzamiento, lo cual es típico de las campañas de marketing.
- Amazon Prime Video destaca por ser la plataforma más neutral, siendo también la que ofrece el catálogo de contenido más adulto, y la mayor oferta de documentales. Por el contrario, la plataforma que tiene a la vez la mayor carga positiva y la menor carga negativa es Disney+, probablemente por su posicionamiento en el mercado como una marca “family-friendly”.
- Las emociones con mayor peso para Netflix y Amazon Prime Video son la anticipación la confianza y la alegría, por ese orden, todas ellas de connotación positiva (que es la emoción más presente). Para Apple TV+ y Disney+ son las mismas tres emociones positivas pero en distinta posición. A su vez, las emociones negativas o más “neutras” como puede ser la sorpresa, tienen un peso mucho menor en los tuits de todas las plataformas. Todo ello estaría en consonancia con el análisis de polaridad realizado, en el cual todas las plataformas han tenido mayor polaridad positiva que negativa.
- En cuanto a las Menciones, merece la pena destacar que la muestra de tuits disponible tiene un volumen casi 6 veces mayor. Y observamos que, a mayor tamaño de la

muestra, más parece aumentar el peso de la polaridad neutra, probablemente, porque se amplifica el efecto neutralizador de las *stop-words*.

- Aquí, destaca la “negatividad” de las menciones directas a Netflix, lo que podría indicar que la plataforma y su contenido son objeto de mayores críticas en las redes, o simplemente, que hacen publicaciones acerca de temáticas o géneros que tienen connotaciones negativas (p.ej. contenido provocador, películas de terror o de violencia, etc.).
- Igualmente cabe destacar la quasi-idéntica positividad de Amazon Prime Video y de Apple TV+, ambas en el primer puesto en cuanto a carga positiva de sus Menciones. Y por último, observamos que a contrario de lo que sucedía con los Tuits, en este caso, las menciones a Disney+ son las más neutras de la muestra.
- En el caso de las Menciones, las tres emociones preponderantes para todas las plataformas SVOD, siguen siendo la confianza, la anticipación y la alegría -si bien no necesariamente por ese orden. Para Amazon Prime Video, Apple TV+ y Disney+, sería aplicable lo mismo expuesto en el epígrafe anterior, en cuanto a que existe una gran diferencia entre las emociones de connotación positiva y las de connotación negativa. No obstante, en el caso de Netflix no podemos decir lo mismo, aquí, siendo el escalón que separa las emociones de connotación positiva y negativa mucho menor para el líder del sector, algo que de nuevo está en consonancia con el análisis de polaridad, en el que se apreciaba que las Menciones a Netflix presentaban una mayor polaridad negativa que las demás. Por último, cabe destacar el peso que tiene la anticipación en las menciones a Disney+, lo que podría indicar que el público en Twitter se muestra entusiasta hacia los estrenos de Disney+.

BIBLIOGRAFÍA

- Alabau, I. (8 de noviembre de 2021). La rueda de las emociones de Robert Plutchik. *Psicología online*. <https://www.psicologia-online.com/la-rueda-de-las-emociones-de-robert-plutchik-4707.html>
- Ballesteros, R. H. (2020). *Marketing más de un siglo de historia: antecedentes, génesis y evolución*. Bogotá.
- Barreto, A. M. (2013). Do users look at banner ads on Facebook? [¿Los usuarios miran los “banners” publicitarios en Facebook?]. *Journal of Research in Interactive Marketing*, 7(2), 119–139. <http://dx.doi.org/10.1108/JRIM-Mar-2012-0013>
- Benes, R. (21 de octubre de 2020). Streaming Services Spend Heavily on Marketing amid Pandemic [Los servicios de streaming invierten mucho en marketing en plena pandemia]. *Insider Intelligence*. <https://www.insiderintelligence.com/content/streaming-services-spend-heavily-on-marketing-amid-pandemic#:~:text=Most%20advertisers%20have%20pulled%20back,ad%20measurement%20company%20iSpot.tv>.
- Bond, E. (14 de Agosto de 2019). New Streaming Services From Disney and Warner Bros. Spell Good News for Localization [Los nuevos servicios de streaming de Disney y Warner Bros. auguran buenas noticias para la localización]. *Slator: Language Industry Intelligence*. <https://slator.com/new-streaming-services-from-disney-and-warner-bros-spell-good-news-for-localization/>
- Conviva. (2021). State of Streaming: Content Discovery 2021 [Estado del streaming: Descubrimiento de contenidos 2021]. https://pages.conviva.com/rs/138-XJA-134/images/RPT_Conviva_SoS_Content_Discovery_2021.pdf
- Coulson, D. C. (2013). Dynamics of social media, politics and public policy in the Arab World [Dinámicas de las redes sociales, la política y las políticas públicas]. *Global Media Journal*, 12(22), 1–20. <https://www.globalmediajournal.com/open-access/dynamics-of-social-media-politics-and-public-policy-in-the-arab-world.pdf>
- ¿Cuál es la red social más segura? (3 de junio de 2022) *El Confidencial Digital*. <https://www.elconfidencialdigital.com/articulo/tendencias/cual-es-red-social-mas-segura/20220602112918405712.html>
- Delgado, M., Navarro, C., Garcia-Muñoz, N., LLuís, P. y Paz, E. (2018) TV news and social audience in Europe (EU5): On-screen and Twitter Strategies [Informativos de televisión y audiencia social en Europa (EU5): Estrategias en pantalla y en Twitter]. *Observatorio (OBS*)*, 12(4). <https://doi.org/10.15847/obsOBS12420181317>
- Díaz, P. (30 de Octubre de 2019). Marketing transaccional o marketing relacional: ¿cuál es el mejor? *Blog Acumbamail*. <https://acumbamail.com/blog/marketing-transaccional-o-marketing-relacional-cual-es-el-mejor/#:~:text=El%20marketing%20transaccional%20se%20enfoca,permite%20obtener%20información%20sobre%20ellos>.

- Dixon, S. (14 de Febrero de 2023). Global social networks ranked by number of users 2023 [Clasificación de las redes sociales mundiales por número de usuarios 2023]. *Statista*. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Escrache, E. (26 de Abril 2022). ¿Por cuánto se compraron las otras redes sociales? *Diario ARA*. https://es.ara.cat/economia/redes-sociales-precio-valor-compra-linkedin-twitter-whatsapp-instagram_1_4351369.html#:~:text=En%202014%20era%20la%20compañ%C3%A9a,meses%20antes%20de%20la%20compra
- Fernández-Gómez, E. & Martín-Quevedo, J. (2018). La estrategia de engagement de Netflix España en Twitter. *El Profesional de La Información*, 27(6), 1292–1302. <https://doi.org/10.3145/epi.2018.nov.12>
- Fuente, O. (25 de abril de 2022). Definición de Marketing ¿Qué es el Marketing Moderno? *IEBS. Marketing. Marketing Estratégico*. <https://www.iebschool.com/blog/marketing-marketing/>
- Hooda, A. y Ankur, A. (2018). Acceptance of Social Media as a Marketing Tool : A Quantitative Study [Aceptación de las Redes Sociales como Herramienta de Marketing: Estudio Cuantitativo]. *East Asian Journal of Business Management*, 8(3), pp. 5-12. <https://doi.org/10.13106/eajbm.2018.vol8.no3.5>
- Ibrahim, N.F.; Wang, X. y Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from Twitter [Explorando el efecto de la participación del cliente en comunidades de marca en línea: Evidencia en Twitter]. *Computers in Human Behavior*, (72), pp. 321-338. <https://dx.doi.org/10.1016/j.chb.2017.03.005>
- Islam, J., & Rahman, Z. (2017). The impact of online brand community characteristics on customer engagement: An application of Stimulus-Organism-Response paradigm [El impacto de las características de la comunidad de marca online en la participación del cliente: Una aplicación del paradigma Estímulo-Organismo-Respuesta]. *Telematics and Informatics*, 34(4), pp. 96–109. <https://doi.org/10.1016/j.tele.2017.01.004>
- Kemp, S. (26 de Enero 2022). Digital 2022: Global Overview Report [Digital 2022: Informe Resumen Global]. *DataReportal, We Are Social, Hootsuite*. <https://datareportal.com/reports/digital-2022-global-overview-report>
- Kemp, S. (26 de Enero 2023). Digital 2023: Global Overview Report [Digital 2023: Informe Resumen Global]. *DataReportal, We Are Social, MeltWater*. <https://datareportal.com/reports/digital-2023-global-overview-report>
- Kim, T. (2022). Critical Interpretations of Global-Local Co-Productions in Subscription Video-on-Demand Platforms: A Case Study of Netflix’s YG Future Strategy Office [Interpretaciones críticas de las coproducciones globales-locales en las plataformas de suscripción de vídeo a la carta: Un estudio de caso de la Oficina de Estrategia de Futuro de YG de Netflix]. *Television & New Media*, 23(4), 405–421. <https://doi.org/10.1177/1527476421999437>

- Lim, W. M. y Rasul, T. (2022). Customer engagement and social media: Revisiting the past to inform the future [Compromiso con el cliente y redes sociales: Revisitando el pasado para informar el futuro]. *Journal of Business Research* 148, pp. 325–342. <http://dx.doi.org/10.1016/j.jbusres.2022.04.068>
- Liu, Y., y Lopez, R. A. (2016). The impact of social media conversations on consumer brand choices [El impacto de las conversaciones en las redes sociales en la elección de marca de los consumidores]. *Marketing Letters*, 27(1), pp. 1-13. <https://doi.org/10.1007/s11002-014-9321-2>
- Martín Barbero, I. (9 de marzo de 2022). Twitter quiere que mandes mensajes con más seguridad, ¿cómo lo hará? *El País (versión online)*. https://cincodias.elpais.com/cincodias/2022/03/09/lifestyle/1646814121_095844.html
- McDonald, K. y Smith-Rowsey, D. (Ed.). (2016). *The Netflix effect: Technology and entertainment in the 21st century* [El efecto Netflix: Tecnología y entretenimiento en el siglo 21]. Bloomsbury.
- Mena Roa, M. (29 de marzo de 2023). La plataforma de streaming con más suscriptores en cada país del mundo. *Statista*. <https://es.statista.com/grafico/29600/servicio-de-streaming-con-el-mayor-numero-de-suscriptores-por-pais/#:~:text=Netflix%2C%20la%20plataforma%20con%20m%C3%A1s,el%20mayor%20n%C3%BAmero%20de%20pa%C3%ADses>.
- Mohammad, S.M (s.f.) *NRC Word-Emotion Association Lexicon* [Léxico NRC de asociación palabra-emoción]. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- Mohammad, S.M. y Turney, P. (2013) Crowdsourcing a Word-Emotion Association Lexicon [Crowdsourcing de un léxico de asociación palabra-emoción]. *Computational Intelligence*, 29(3), pp. 436-465. <https://arxiv.org/pdf/1308.6297.pdf>
- Okazaki, S., Díaz Martín, A.M., Rozano Suplet, M. y de Menéndez Benito, H. (2015). Using Twitter to engage: a data mining approach [Utilizar Twitter para captar la atención: un enfoque de minería de datos]. *Internet research: Electronic networking applications and policy*, 25(3), pp. 416-434. DOI:[10.1108/IntR-11-2013-0249](https://doi.org/10.1108/IntR-11-2013-0249)
- Orús, A. (9 de diciembre de 2022). Número de suscriptores de vídeo bajo demanda en todo el mundo por servicio en 2027. *Statista*. <https://es.statista.com/estadisticas/1289305/numero-de-suscriptores-de-svod-en-todo-el-mundo-por-servicio/>
- Orús, A. (10 de febrero de 2023). Ranking de los servicios de streaming de vídeo más demandados del mundo en 2022. *Statista*. <https://es.statista.com/estadisticas/879769/ranking-de-los-servicios-de-video-en-streaming-mas-demandados-del-mundo/>

- Pasirayi, S., y Fennell, P. B. (2021). The effect of subscription-based direct-to-consumer channel additions on firm value [El efecto de los canales de suscripción directos al consumidor en el valor de las empresas]. *Journal of Business Research*, (123), pp.355–366. <https://doi.org/10.1016/j.jbusres.2020.09.067>
- Paul, J., & Benito, G. R. (2018). A review of research on outward foreign direct investment from emerging countries, including China: What do we know, how do we know and where should we be heading? [Una revisión de la investigación sobre la salida de inversión extranjera directa de los países emergentes, incluida China: ¿Qué sabemos, cómo lo sabemos y hacia dónde deberíamos dirigirnos?] *Asia Pacific Business Review*, 24(1), pp. 90–115. <http://justinpaul.uprrp.edu/wp-content/uploads/2019/11/APBRPaginatednov817.pdf>
- Pérez, E. P. (Mayo 2020). Next episode, next market: social media marketing for SVOD: the case of netflix [Siguiendo episodio, siguiente mercado: marketing en redes sociales para SVOD: el caso de Netflix]. *On Research: Journal of EU Business School* (4), pp.8-13. https://www.researchgate.net/publication/348549827_Next_episode_next_market_Social_media_marketing_for_SVoD_the_case_of_Netflix
- Provost, F., & Fawcett, T. (2013). Representing and mining text [Representación y extracción de textos]. *Data science for business: what you need to know about data mining and data-analytic thinking* [Ciencia de datos para la empresa: lo que necesita saber sobre minería de datos y pensamiento analítico de datos] (pp.263-290). O'Reilly.
- ¿Qué son y para qué sirven el Text mining y Data mining? (28 de abril de 2022) Immune Technology Institute. <https://immune.institute/blog/text-mining-data-mining-que-son/#:~:text=El%20Text%20mining%20o%20miner%C3%ADa,relaci%C3%B3n%20entre%20los%20distintos%20contenidos.>
- Rahe, V., Buschow, C., y Schlütz, D. (2021). How users approach novel media products: brand perception of Netflix and Amazon Prime video as signposts within the German subscription-based video-on-demand market [Cómo se acercan los usuarios a los nuevos productos mediáticos: percepción de marca de Netflix y Amazon Prime Video como indicadores dentro del mercado alemán de vídeo a la carta por suscripción]. *Journal of Media Business Studies*, 18(1), pp. 45–58. <https://doi.org/10.1080/16522354.2020.1780067>
- Rataul, P., Tisch, D.G., y Záborský, P. (2018). Netflix: Dynamic capabilities for global success [Netflix: Capacidades dinámicas para el éxito mundial]. *SAGE Publications: SAGE Business Cases Originals*.
- Santini, F. de O., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: a framework and meta- análisis [Compromiso del cliente en los medios sociales: marco y meta-análisis]. *Journal of the Academy of Marketing Science*, 48(6), pp. 1211–1228. <https://doi.org/10.1007/s11747-020-00731-5>
- Southern, M. G. (16 de Noviembre de 2022). Elon xMusk's Twitter Takeover: A Timeline Of Events [Elon Musk se hace con Twitter: Cronología de los

acontecimientos]. *Search Engine Journal*.
<https://www.searchenginejournal.com/elon-musks-twitter-takeover-a-timeline-of-events/470927/#:~:text=Elon%20Musk%20became%20the%20owner,changes%20implemented%20under%20his%20leadership>

Spangler, T. (18 de marzo de 2019). 'Subscription Fatigue': Nearly Half of U.S. Consumers Frustrated by Streaming Explosion, Study Finds ["Fatiga de suscripción": Casi la mitad de los consumidores estadounidenses frustrados por la explosión del streaming, según un estudio]. *Variety*.
<https://variety.com/2019/digital/news/streaming-subscription-fatigue-us-consumers-deloitte-study-1203166046/>

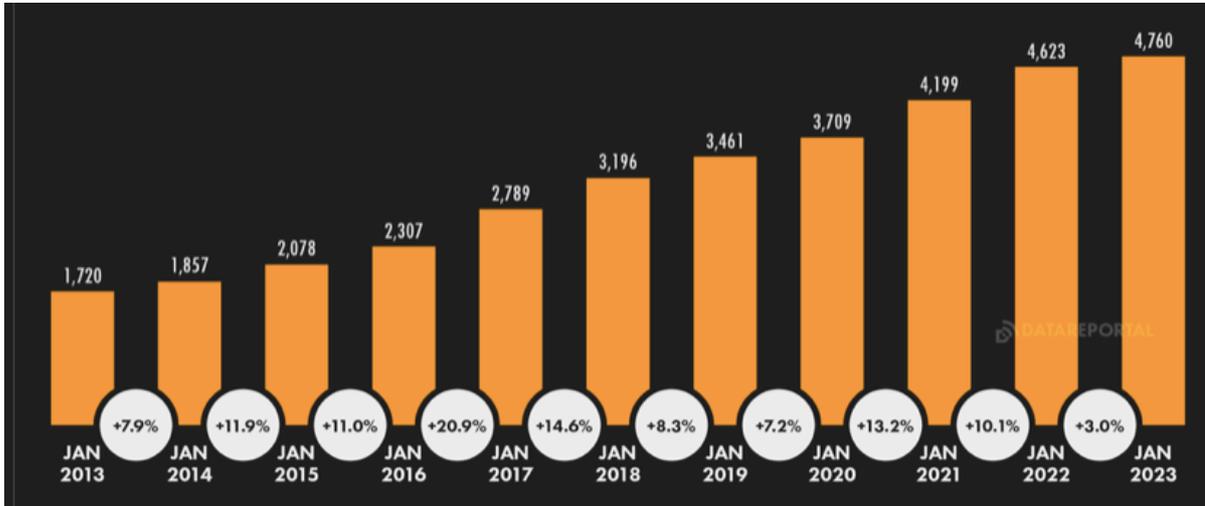
Tavares, C., & Nogueira, M. (2021). Identifying and Exploring Key Drivers of Customer Engagement on Social Media: A Netnography Approach [Identificar y explorar los factores clave del compromiso de los clientes en las redes sociales: Un enfoque netnográfico]. En Á. Rocha, J. L. Reis, M. K. Peter, R. Cayolla, S. Loureiro, & Z. Bogdanovic (Eds.), *Marketing and Smart Technologies* [Marketing y tecnologías inteligentes]. (vol.205, pp. 507–519). Springer. https://doi.org/10.1007/978-981-33-4183-8_40

Wohn, D. Y. y Na, E.-K. (2011) Tweeting about TV: Sharing television viewing experiences via social media message streams [Tuitear sobre la tele: compartir experiencias televisivas a través de las redes sociales]. *First Monday*, 16(3).
<https://doi.org/10.5210/fm.v16i3.3368>

Zimbra, D., Abbasi, A., Zeng, D. y Chen, H. (2018). The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation [El estado del arte en el análisis del sentimiento de Twitter: Revisión y evaluación comparativa]. *ACM Transactions on Management Information Systems*, 9(2) pp. 1–29
<https://doi.org/10.1145/3185045>

ANEXOS

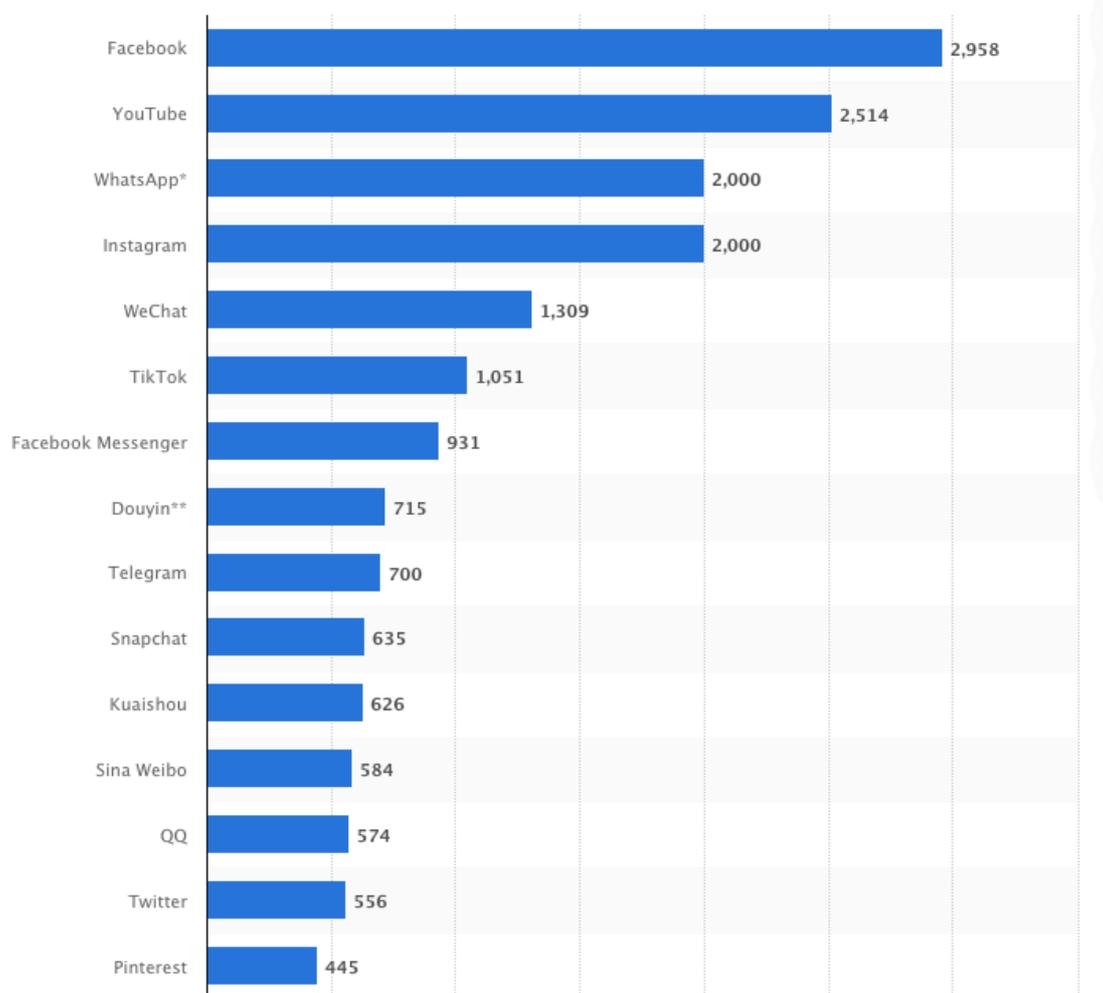
ANEXO #1: EVOLUCIÓN DEL USO DE LAS RRSS DE 2012-23: NÚMERO DE USUARIOS EN RRSS (EN MILLONES) Y VARIACIÓN ANUAL (DE ENERO A ENERO)



Nota: Los usuarios pueden no representar individuos únicos

Fuente: Kemp, S. (26 de Enero de 2023). Digital 2023: Global Overview Report [Digital 2023: Informe Resumen Global]. *DataReportal, We Are Social, MeltWater*.
<https://datareportal.com/reports/digital-2023-global-overview-report>

ANEXO #2: LAS RRSS MÁS POPULARES DEL MUNDO CLASIFICADAS POR NÚMERO MENSUAL DE USUARIOS ACTIVOS (EN MILLONES), ENERO



2023

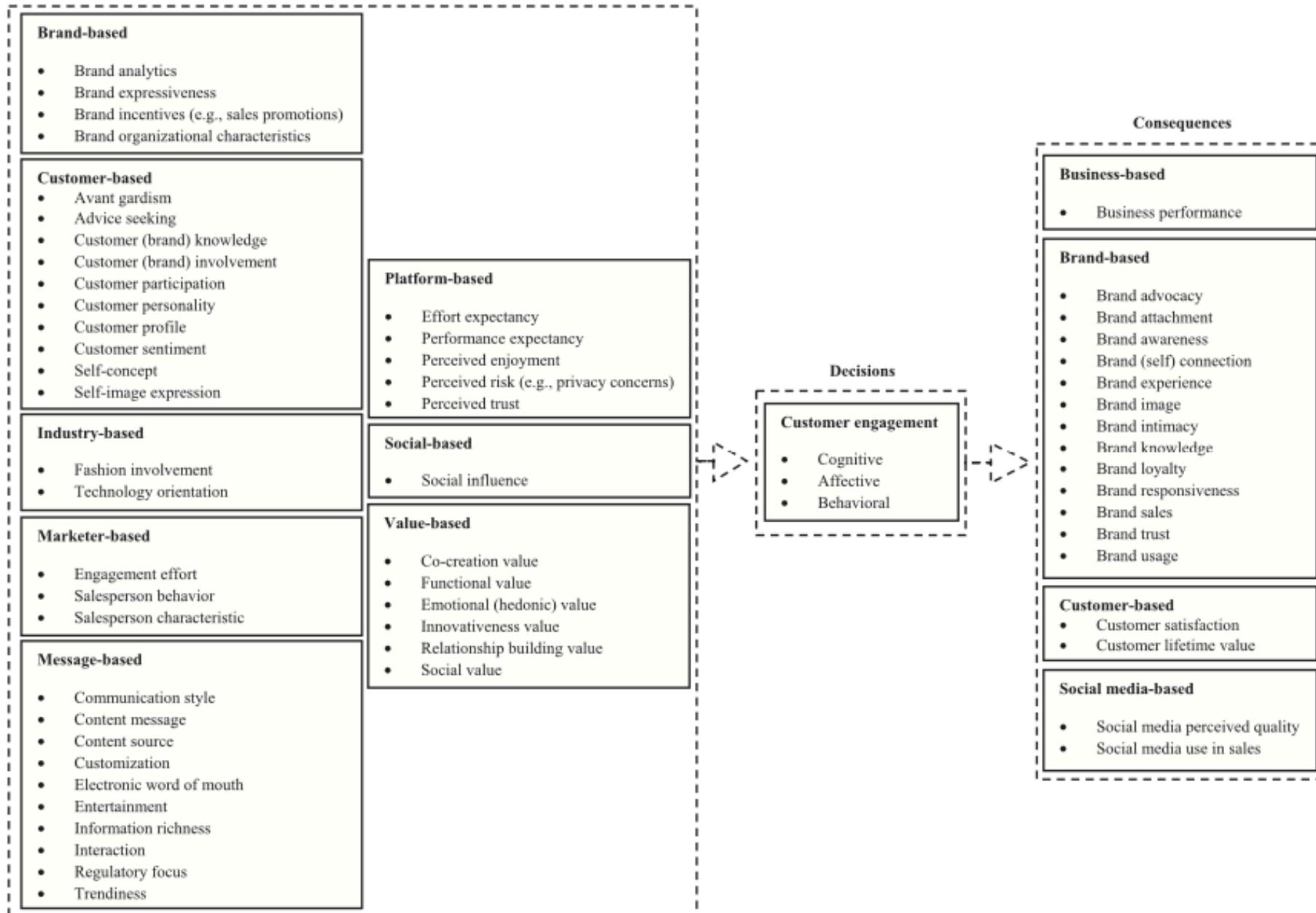
Nota:

**Las plataformas no han publicado cifras actualizadas de usuarios en los últimos 12 meses, las cifras pueden estar desfasadas y ser menos fiables.*

*** La cifra contabiliza usuarios activos diarios, por lo que es probable que el número de usuarios activos mensuales sea mayor.*

Fuente: Dixon, S. (14 de Febrero de 2023). Global social networks ranked by number of users 2023 [Clasificación de las redes sociales mundiales por número de usuarios 2023]. Statista. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

ANEXO #3: ANTECEDENTES , DECISIONES Y RESULTADOS DE LA PDC EN RRSS SEGÚN LIM Y RASUL (2022)



ANEXO #4: CÓDIGO PARA LA DESCARGA DE LOS DATOS EN RSTUDIO Y COMENTARIOS EXPLICATIVOS

NOTA: Los comentarios al código están precedidos del signo ‘#’ para que el lector pueda copiar y pegar el código en el entorno RStudio para su ejecución.

```
# 0. Pasos previos opcionales
# 0.1 Limpiar entorno de trabajo

rm(list = ls())

options(encoding = "utf8")

# 0.2. Instalar paquetes necesarios para la descarga de los datos,
si no se habían descargado anteriormente

install.packages("rtweet")
# Para acceder a los datos de Twitter

install.packages("tibble")
# Para manipular y estructurar datos

install.packages("writexl")
# Para crear ficheros excel

# 1. Pasos previos necesarios para la descarga

# 1.1 Cambiar el directorio de trabajo al directorio donde se
encuentra el archivo de R en el que se está trabajando

setwd(dirname(rstudioapi::getActiveDocumentContext())$path)

# 1.2. Llamar a las librerías contenidas en los paquetes
descargados previamente, si no se había hecho antes

library(rtweet)
library(tibble)
library(writexl)
```

```
# 2. Autenticar sesión en Twitter
```

```
auth <- rtweet_app()
```

```
# Al ejecutar nos va a pedir el 'bearer token' de la aplicación  
que hemos creado con nuestra cuenta de desarrollador en Twitter.  
Lo introducimos y ya se queda guardado en el objeto para toda la  
sesión. Si limpiamos de nuevo el entorno de trabajo o reiniciamos  
de sesión de RStudio, habrá que volver a ejecutar el código y  
reintroducir el token. Este objeto lo vamos a introducir  
posteriormente como parámetro 'token = auth' en las funciones que  
vamos a definir para la descarga de los tuits, para que cada vez  
que ejecutemos la función nuestras credenciales se verifiquen de  
manera automática
```

```
# 3. Definir la función para automatizar la descarga de los datos
```

```
# 3.1. Para los tuits publicados por las plataformas SVOD en sus timelines
```

```
crear_excel_timeline <- function(plataforma, fecha) {  
# 'plataforma' y 'fecha' son los inputs de la función, por ahora  
no tienen ningún valor asignado, habrá que dárselo en el momento  
de ejecutar la función. En lugar de 'plataforma' introduciremos  
el nombre de usuario en Twitter (sin el signo @) de la plataforma  
SVOD. En lugar de 'fecha' podremos poner lo que queramos, p.ej.  
la fecha de descarga de los datos, de cara a identificarlos en el  
futuro. Muy importante que escribamos estos inputs entre comillas  
simples.
```

```
df_timeline <- get_timeline(c(plataforma), n=3200, token=auth)  
# Este objeto contendrá los datos extraídos de Twitter gracias a  
la función 'get_timeline'. Le damos el valor 'auth' (objeto que  
contiene nuestro bearer token de desarrollador) al parámetro  
'token' para automatizar el proceso de autenticación de la API de  
Twitter.
```

```
path <- '/Users/fatima/Desktop/TFG  
Analytics/DDBB/get_timelines/'  
# Este objeto contendrá la ruta donde se archivará el fichero  
Excel donde volcaremos los datos descargados de Twitter
```

```
namefile <- paste0(path, plataforma , '_', fecha, '.xlsx')  
# Este objeto contendrá el nombre del fichero Excel donde  
volcaremos los datos descargados de Twitter. Se compone de: la  
ruta; los inputs de la función que estamos definiendo(plataforma  
y fecha) por lo que tendremos que introducir su valor en el momento  
de ejecutar la función; y la extensión '.xlsx'
```

```
write_xlsx(df_timeline , namefile)  
# Creamos el fichero Excel (con el nombre predefenido)y volcamos  
los tuits descargados  
}
```

```
# 3.2. Para los tuits publicados por otros usuarios mencionando a las plataformas SVOD
```

```
crear_excel_menciones <- function(plataforma, fecha){  
# De nuevo 'plataforma' y 'fecha' son los inputs de la función, y es aplicable lo explicado previamente al respecto.
```

```
  id_plataforma <- paste0('@', plataforma)  
# En Twitter se puede apelar públicamente a otro usuario ("mencionarlo" en jerga tuitera) escribiendo un tuit dónde figure su nombre de usuario precedido por el signo @. Nosotros queremos filtrar "menciones" directas a las plataformas SVOD, así que comenzamos definiendo un objeto en el que se añada el signo '@' al nombre de la plataforma introducido como input. Este objeto luego lo añadiremos como parámetro para filtrar la búsqueda de tuits.
```

```
  df_menciones <- search_tweets(id_plataforma, n = Inf, token = auth)
```

```
  # Con 'search_tweets' obtenemos un dataframe con los tuits en los que figure el texto contenido en 'id_plataforma'. No tenemos límite de volumen, pero si temporal: solo nos devolverá tuits publicados en los últimos 6-9 días.
```

```
  path<-'/Users/fatima/Desktop/TFG  
Analytics/DDBB/search_tweets/'  
# De nuevo, la ruta donde queremos que se archive el fichero Excel donde vamos a volcar los datos descargados de Twitter.
```

```
  namefile<-paste0(path,'menciones_',  
plataforma,'_',fecha,'.xlsx')  
  # De nuevo, el nombre que tendrá el fichero Excel donde vamos a volcar los tuits descargados
```

```
  write_xlsx(df_menciones,namefile)  
  # Creamos el fichero Excel y volcamos los tuits descargados  
}
```

```
# 4.Ejecutar las funciones predefinidas para proceder a la
descarga de las dos clases de tuits para cada una de las 7
plataformas SVOD
```

```
crear_excel_timeline('netflix','19Apr')
```

```
crear_excel_timeline('PrimeVideo','19Apr')
```

```
crear_excel_timeline('DisneyPlus','19Apr')
```

```
crear_excel_timeline('hbomax','19Apr')
```

```
crear_excel_timeline('hulu','19Apr')
```

```
crear_excel_timeline('AppleTVPlus','19Apr')
```

```
crear_excel_timeline('paramountplus','19Apr')
```

```
crear_excel_menciones('netflix','19Apr')
```

```
crear_excel_menciones('PrimeVideo','19Apr')
```

```
crear_excel_menciones('DisneyPlus','19Apr')
```

```
crear_excel_menciones('hbomax','19Apr')
```

```
crear_excel_menciones('hulu','19Apr')
```

```
crear_excel_menciones('AppleTVPlus','19Apr')
```

```
crear_excel_menciones('paramountplus','19Apr')
```

ANEXO #5: CÓDIGO EMPLEADO PARA LA CARACTERIZACIÓN DE NUESTROS DATOS

NOTA: Los comentarios al código están precedidos del signo ‘#’ para que el lector pueda copiar y pegar el código en el entorno RStudio para su ejecución.

```
# 0. Pasos previos
```

```
# 0.1 Limpiar entorno de trabajo
```

```
rm(list = ls())
```

```
options(encoding = "utf8")
```

```
# 0.2. Instalar los siguientes paquetes si no se habían descargado anteriormente
```

```
install.packages("readxl")
```

```
# Para leer ficheros Excel
```

```
install.packages("dplyr")
```

```
install.packages("tidyr")
```

```
# Para crear y manipular datasets
```

```
# 0.3 Cambiar el directorio de trabajo al directorio donde se encuentra el archivo de R en el que se está trabajando
```

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# 0.4. Llamar a las librerías contenidas en los paquetes
descargados previamente, si no se había hecho antes

library(readxl)

library(dplyr)

library(tidyr)

# 1. Cargamos los datos a RStudio

#1.1 Definimos una función para cargar los datos desde nuestros
ficheros Excel

# Para los tuits de las plataformas SVOD

carga_datos_Tweets <- function(nombre_plataforma){

  # 'nombre_plataforma' es el input de la función, por ahora no
  tiene ningún valor, habrá que dárselo en el momento de ejecutar
  la función. En nuestro caso introduciremos el nombre de usuario
  en Twitter (sin el signo @) de la plataforma SVOD. Muy importante
  que lo escribamos entre comillas simples.

  directorio_analytics<-getwd()
```

```
# Objeto que contiene la ruta hasta la carpeta donde se encuentra
el archivo de R en el que estamos trabajando
```

```
ruta_fichero<-
paste0(directorio_analytics,'/DDBB/get_timelines/',nombre_plataf
orma,'_19Apr.xlsx')
```

```
# Ruta hasta los ficheros Excels que contienen los datos
```

```
Tweets <- read_excel(ruta_fichero)
```

```
# Leemos los datos del fichero Excel
```

```
}
```

```
# Para los tuits que mencionan a las plataformas SVOD
```

```
carga_datos_Menciones<- function(nombre_plataforma){
  directorio_analytics<-getwd()
  ruta_fichero<-
  paste0(directorio_analytics,'/DDBB/search_tweets/','menciones_',
nombre_plataforma,'_19Apr.xlsx')
  Menciones<-read_excel(ruta_fichero)}
```

```
#1.2. Cargamos los datos de nuestros ficheros Excel con estas
funciones
```

```
Tweets_netflix<-carga_datos_Tweets('netflix')
```

```
Tweets_AppleTVPlus<-carga_datos_Tweets('AppleTVPlus')
```

```

Tweets_DisneyPlus<-carga_datos_Tweets('DisneyPlus')

Tweets_hbomax<-carga_datos_Tweets('hbomax')

Tweets_hulu<-carga_datos_Tweets('hulu')

Tweets_paramountplus<-carga_datos_Tweets('paramountplus')

Tweets_PrimeVideo<-carga_datos_Tweets('PrimeVideo')

Menciones_netflix<-carga_datos_Menciones('netflix')

Menciones_AppleTVPlus<-carga_datos_Menciones('AppleTVPlus')

Menciones_DisneyPlus<-carga_datos_Menciones('DisneyPlus')

Menciones_hbomax<-carga_datos_Menciones('hbomax')

Menciones_hulu<-carga_datos_Menciones('hulu')

Menciones_paramountplus<-carga_datos_Menciones('paramountplus')

Menciones_PrimeVideo<-carga_datos_Menciones('PrimeVideo')

# 2. Analizamos la franja temporal de los Tuits de cada Plataforma
SVOD:

# 2.1. Obtenemos una tabla con la fecha mínima y máxima de los
Tuits de cada plataforma

Plataforma_SVOD<-
c('netflix','AppleTVPlus','DisneyPlus','hbomax','hulu','paramoun
tplus','PrimeVideo')

# Nombre de usuario de la plataforma

```

```

Fecha_Minima<-
c(min(Tweets_netflix$created_at),min(Tweets_AppleTVPlus$created_
at),min(Tweets_DisneyPlus$created_at),min(Tweets_hbomax$created_
at),min(Tweets_hulu$created_at),min(Tweets_paramountplus$created
_at),min(Tweets_PrimeVideo$created_at))

# Fecha de publicación minima

Fecha_Maxima<-
c(max(Tweets_netflix$created_at),max(Tweets_AppleTVPlus$created_
at),max(Tweets_DisneyPlus$created_at),max(Tweets_hbomax$created_
at),max(Tweets_hulu$created_at),max(Tweets_paramountplus$created
_at),max(Tweets_PrimeVideo$created_at))

# Fecha de publicación máxima

df <- data.frame(Plataformas, Fecha_Minima, Fecha_Maxima)

# Ponemos los datos en formato tabla

View(df)

# Visualizamos

# Descartamos hulu y paramount por sus franjas temporales

# 2.3 Filtramos para quedarnos con los Tuits publicados entre el
09/11/22 y el 04/04/2023

filtrado_por_fechas<- function(datos){

```

```

datos %>% filter(created_at >= as.POSIXct("2022-11-09
00:00:00", tz="UTC")

          & datos$created_at <= as.POSIXct("2023-04-04
24:00:00", tz="UTC"))}

# Funcion para filtrar por las fechas deseadas

# Ejecutamos la función

Tweets_netflix<-filtrado_por_fechas(Tweets_netflix)

Tweets_PrimeVideo<-filtrado_por_fechas(Tweets_PrimeVideo)

Tweets_AppleTVPlus<-filtrado_por_fechas(Tweets_AppleTVPlus)

Tweets_DisneyPlus<-filtrado_por_fechas(Tweets_DisneyPlus)

Tweets_hbomax<-filtrado_por_fechas(Tweets_hbomax)

# Sacamos las nuevas fechas mínimas, máximas y el volumen de Tuits
tras el filtrado

Plataforma_SVOD<-
c('netflix','AppleTVPlus','DisneyPlus','hbomax','PrimeVideo')

Fecha_Minima2<-
c(min(Tweets_netflix$created_at),min(Tweets_AppleTVPlus$created_
at),min(Tweets_DisneyPlus$created_at),min(Tweets_hbomax$created_
at),min(Tweets_PrimeVideo$created_at))

Fecha_Maxima2<-
c(max(Tweets_netflix$created_at),max(Tweets_AppleTVPlus$created_

```

```
at),max(Tweets_DisneyPlus$created_at),max(Tweets_hbomax$created_at),max(Tweets_PrimeVideo$created_at))
```

```
Volumen_Tuits<-
```

```
c(nrow(Tweets_netflix),nrow(Tweets_AppleTVPlus),nrow(Tweets_DisneyPlus),nrow(Tweets_hbomax),nrow(Tweets_PrimeVideo))
```

```
# Ponemos los datos en formato tabla
```

```
df2 <- data.frame(Plataforma_SVOD,Fecha_Minima2,Fecha_Maxima2,Volumen_Tuits)
```

```
# Visualizamos
```

```
# Volumen de Tuits mas pequeño para todos, sobre todo hbomax asi que tambien descartamos esta plataforma
```

```
View(df2)
```

```
# 3. Analizamos la franja temporal de las Menciones de cada Plataforma SVOD:
```

```
Plataforma_SVOD<-
```

```
c('netflix','AppleTVPlus','DisneyPlus','hbomax','PrimeVideo')
```

```
Fecha_Minima3<-
```

```
c(min(Menciones_netflix$created_at),min(Menciones_AppleTVPlus$cr
```

```
eated_at),min(Menciones_DisneyPlus$created_at),min(Menciones_hbo
max$created_at),min(Menciones_PrimeVideo$created_at))
```

```
Fecha_Maxima3<-
c(max(Menciones_netflix$created_at),max(Menciones_AppleTVPlus$cr
eated_at),max(Menciones_DisneyPlus$created_at),max(Menciones_hbo
max$created_at),max(Menciones_PrimeVideo$created_at))
```

```
Volumen_Menciones<-
c(nrow(Menciones_netflix),nrow(Menciones_AppleTVPlus),nrow(Menci
ones_DisneyPlus),nrow(Menciones_hbomax),nrow(Menciones_PrimeVide
o))
```

```
# Ponemos los datos en formato tabla
```

```
df3 <- data.frame(Plataforma_SVOD,Fecha_Minima3,Fecha_Maxima3,
Volumen_Menciones)
```

```
# Visualizamos
```

```
View(df3)
```

```
# 4. Filtramos las Menciones en base a su fuente de publicación:
```

```
# Definimos una funcion para filtrar por las fuentes deseadas
```

```
filtrado_por_fuentes<- function(datos){
```

```
  datos<-datos %>% extract(source, "source", "Twitter (.*)<")
  %>%
```

```
filter(source %in% c("for iPhone", "for Android", "for iPad",  
"Web App"))}
```

```
# Ejecutamos la función
```

```
Menciones_netflix<-filtrado_por_fuentes(Menciones_netflix)
```

```
Menciones_PrimeVideo<-filtrado_por_fuentes(Menciones_PrimeVideo)
```

```
Menciones_AppleTVPlus<-  
filtrado_por_fuentes(Menciones_AppleTVPlus)
```

```
Menciones_DisneyPlus<-filtrado_por_fuentes(Menciones_DisneyPlus)
```

```
# Visualizamos nuevo volumen de menciones tras filtrado
```

```
Plataforma_SVOD<-  
c('netflix','AppleTVPlus','DisneyPlus','PrimeVideo')
```

```
Volumen_Menciones<-  
c(nrow(Menciones_netflix),nrow(Menciones_AppleTVPlus),nrow(Menci  
ones_DisneyPlus),nrow(Menciones_PrimeVideo))
```

```
Fecha_Minima<-  
c(min(Menciones_netflix$created_at),min(Menciones_AppleTVPlus$cr  
eated_at),min(Menciones_DisneyPlus$created_at),min(Menciones_Pri  
meVideo$created_at))
```

```
Fecha_Maxima<-  
c(max(Menciones_netflix$created_at),max(Menciones_AppleTVPlus$cr  
eated_at),max(Menciones_DisneyPlus$created_at),max(Menciones_Pri  
meVideo$created_at))
```

```
df4      <-      data.frame(Plataforma_SVOD,Volumen_Menciones,  
Fecha_Minima,Fecha_Maxima)
```

```
View(df4)
```

ANEXO #6: CÓDIGO PARA EL WORD FREQUENCY ANALYSIS

NOTA: Los comentarios al código están precedidos del signo '#' para que el lector pueda copiar y pegar el código en el entorno RStudio para su ejecución.

```
# PRIMER ANÁLISIS DEL TEXTO: WORD FREQUENCY ANALYSIS

# 1. PREPROCESAMIENTO DE LOS DATOS

#1.1 CARGA DE LOS DATOS A R Y LIMPIEZA DE LOS MISMOS

# Limpiamos entorno de trabajo, definimos directorio, descargamos
paquetes y cargamos librerías necesarias (si no lo habíamos hecho
ya)

rm(list = ls())

options(encoding = "utf8")

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

install.packages("readxl")
install.packages("dplyr")
install.packages("tidyr")

library(readxl)
library(dplyr)
library(tidyr)

# Definimos una función para cargar los Tuits de Excel a RStudio
y preprocesarlos al mismo tiempo

carga_datos_Tweets <- function(nombre_plataforma) {

  # Reutilizamos el código para cargar los datos de Excel a RStudio
del Anexo anterior
  directorio_analytics<-getwd()
```

```

ruta_fichero<-
paste0(directorio_analytics,'/DDBB/get_timelines/',nombre_plataf
orma,'_19Apr.xlsx')
Tweets<-read_excel(ruta_fichero)

# Adapataamos también el código para filtrar por fechas
Tweets <- filter(Tweets, created_at >= as.POSIXct("2022-11-09
00:00:00", tz="UTC") & Tweets$created_at <= as.POSIXct("2023-04-
04 24:00:00", tz="UTC"))

# Y añadimos nuevo código para el preprocesamiento del texto:

Tweets<-select(Tweets, id, created_at, text, source) #
Conservamos solo las variables relevantes para nuestro analisis

Tweets$text<- gsub("&", "", Tweets$text) # quitamos los &
Tweets$text<-gsub( "@\\w+", "", Tweets$text) #quitamos las
menciones
Tweets$text<- gsub("http\\S*", "", Tweets$text) #quitamos las
URLS
Tweets$text<-gsub("[:punct:]", "", Tweets$text) #quitamos
signos de puntuación
Tweets$text<-gsub("[:digit:]", "", Tweets$text) #quitamos
dígitos
Tweets$text<-gsub("[^\\x01-\\x7F]", "", Tweets$text) #quitamos
todos los caracteres no ASCII p.ej. emoticonos
Tweets$text<-tolower(Tweets$text) #pasamos todas las palabras a
minúsculas
Tweets
}
# Hacemos lo mismo para las Menciones (sin filtrado por fechas,
pero filtrando por fuentes):

carga_datos_Menciones <- function(nombre_plataforma){
  directorio_analytics<-getwd()
  ruta_fichero<-
paste0(directorio_analytics,'/DDBB/search_tweets/','menciones_',
nombre_plataforma,'_19Apr.xlsx')

```

```

Menciones<-read_excel(ruta_fichero)
Menciones<-Menciones %>% extract(source, "source", "Twitter
(.*?)<") %>%
  filter(source %in% c("for iPhone", "for Android", "for iPad",
"Web App"))
Menciones<-select(Menciones, id, created_at, text, source)
Menciones$text<-gsub("&", "", Menciones$text)
Menciones$text<-gsub("@\\w+", "", Menciones$text)
Menciones$text<-gsub("http\\S*", "", Menciones$text)
Menciones$text<-gsub("[[:punct:]]", "", Menciones$text)
Menciones$text<-gsub("[[:digit:]]", "", Menciones$text)
Menciones$text <- gsub("^\\s+|\\s+$", "", Menciones$text)
Menciones$text<-tolower(Menciones$text)
Menciones}

# Ejecutamos las funciones para nuestras plataformas SVOD
(netflix, PrimeVideo, AppleTVPlus, DisneyPlus)

Tweets_netflix<-carga_datos_Tweets('netflix')
Tweets_PrimeVideo<-carga_datos_Tweets('PrimeVideo')
Tweets_AppleTVPlus<-carga_datos_Tweets('AppleTVPlus')
Tweets_DisneyPlus<-carga_datos_Tweets('DisneyPlus')

Menciones_netflix<-carga_datos_Menciones('netflix')
Menciones_PrimeVideo<-carga_datos_Menciones('PrimeVideo')
Menciones_AppleTVPlus<-carga_datos_Menciones('AppleTVPlus')
Menciones_DisneyPlus<-carga_datos_Menciones('DisneyPlus')

#2.1 CONSTRUCCION DEL CORPUS Y DE LA MATRIZ DTM

#install.packages("quanteda")
library(quanteda)

# Definimos una funcion para automatizar la creacion del corpus y
matriz dtm para cada una de las plataformas SVOD

#A. CORPUS

```

```

creacion_corpus <- function(datos) {
  micorpus<-corpus(datos) #construccion del corpus: cada tweet
sera una linea y cada palabra un token
  corpus_tokenizado<-tokens(micorpus) #tokenizacion corpus
}

```

```

corpus_Tweets_netflix<-creacion_corpus(Tweets_netflix)
corpus_Tweets_PrimeVideo<-creacion_corpus(Tweets_PrimeVideo)
corpus_Tweets_AppleTVPlus<-creacion_corpus(Tweets_AppleTVPlus)
corpus_Tweets_DisneyPlus<-creacion_corpus(Tweets_DisneyPlus)

```

```

corpus_Menciones_netflix<-creacion_corpus(Menciones_netflix)
corpus_Menciones_PrimeVideo<-
creacion_corpus(Menciones_PrimeVideo)
corpus_Menciones_AppleTVPlus<-
creacion_corpus(Menciones_AppleTVPlus)
corpus_Menciones_DisneyPlus<-
creacion_corpus(Menciones_DisneyPlus)

```

#B. MATRIZ DTM

```

dtm <- function(corpus_tokenizado) {
  midtm<- dfm(corpus_tokenizado) #la funcion dfm por defecto
convierte el texto a minúsculas porque su parametro 'tolower'
tiene 'TRUE' como valor predeterminado
  midtm}

```

```

dtm_Tweets_netflix<-dtm(corpus_Tweets_netflix)
dtm_Tweets_PrimeVideo<-dtm(corpus_Tweets_PrimeVideo)
dtm_Tweets_AppleTVPlus<-dtm(corpus_Tweets_AppleTVPlus)
dtm_Tweets_DisneyPlus<-dtm(corpus_Tweets_DisneyPlus)

```

```

dtm_Menciones_netflix<-dtm(corpus_Menciones_netflix)
dtm_Menciones_PrimeVideo<-dtm(corpus_Menciones_PrimeVideo)
dtm_Menciones_AppleTVPlus<-dtm(corpus_Menciones_AppleTVPlus)
dtm_Menciones_DisneyPlus<-dtm(corpus_Menciones_DisneyPlus)

```

```
# 2.2 VISUALIZACION DE LAS PALABRAS MÁS REPETIDAS
```

```
#install.packages("quanteda.textstats")  
#install.packages("ggplot2")  
#install.packages("wordcloud2")  
#install.packages('gridExtra')
```

```
library(quanteda.textstats)  
library(ggplot2)  
library(wordcloud2)
```

```
#A. Primera visualización de las 50 palabras mas frecuentes
```

```
topfeatures(dtm_Tweets_netflix,50)  
topfeatures(dtm_Tweets_PrimeVideo,50)  
topfeatures(dtm_Tweets_AppleTVPlus,50)  
topfeatures(dtm_Tweets_DisneyPlus,50)
```

```
topfeatures(dtm_Menciones_netflix,50)  
topfeatures(dtm_Menciones_PrimeVideo,50)  
topfeatures(dtm_Menciones_AppleTVPlus,50)  
topfeatures(dtm_Menciones_DisneyPlus,50)
```

```
#B. Scatter plot de las 20 palabras más frecuentes para  
colecciones de Tuits y Menciones seleccionadas
```

```
scatter_plot_WFA <- function(matriz_dtm,plataforma){  
  features_freq<-textstat_frequency(matriz_dtm, n = 20)  
  ggplot(features_freq, aes(x = reorder(feature,-frequency), y =  
frequency)) +  
    geom_point() +  
    theme(axis.text.x = element_text(angle = 90, hjust = 1)) +  
    xlab("Word")+  
    ylab("Frequency") +  
    ggtitle(plataforma)+  
    theme(plot.title = element_text(hjust = 0.5))+  
    theme(plot.title=element_text(face="bold"))
```

```
}
```

```
scatter_plot_WFA(dtm_Tweets_netflix, 'Netflix')
```

```
#C. Nube de Palabras
```

```
word_cloud_WFA <- function(matriz_dtm,plataforma){  
  set.seed(123)  
  midtm_t <- t(matriz_dtm)  
  m <- as.matrix(midtm_t)  
  v <- sort(rowSums(m),decreasing=TRUE)  
  d <- data.frame(word = names(v),freq=v)  
  wordcloud2(data = d, size = 2, shape = "cloud", color="random-  
dark", ellipticity = 0.5)  
}
```

```
word_cloud_WFA(dtm_Menciones_PrimeVideo)
```

ANEXO #7: CÓDIGO PARA EL SENTIMENT ANALYSIS

NOTA: Los comentarios al código están precedidos del signo '#' para que el lector pueda copiar y pegar el código en el entorno RStudio para su ejecución.

```
# SEGUNDO ANÁLISIS DEL TEXTO: SENTIMENT ANALYSIS

# 1. PREPROCESAMIENTO DE LOS DATOS

#1.1 CARGA DE LOS DATOS A R Y LIMPIEZA DE LOS MISMOS

# Limpiamos entorno de trabajo, definimos directorio, descargamos
paquetes y cargamos librerías necesarias (si no lo habíamos hecho
ya)

rm(list = ls())

options(encoding = "utf8")

setwd(dirname(rstudioapi::getActiveDocumentContext())$path)

install.packages("readxl")
install.packages("dplyr")
install.packages("tidyr")
install.packages("syuzhet")
install.packages("RColorBrewer")

library(readxl)
library(dplyr)
library(tidyr)

# Definimos una función para cargar los Tuits de Excel a RStudio
y preprocesarlos al mismo tiempo (añadimos filtro por fechas)

carga_datos_Tweets <- function(nombre_plataforma){
  directorio_analytics<-getwd()
```

```

ruta_fichero<-
paste0(directorio_analytics,'/DDBB/get_timelines/',nombre_plataforma, '_19Apr.xlsx')
Tweets<-read_excel(ruta_fichero)
Tweets <- filter(Tweets, created_at >= as.POSIXct("2022-11-09 00:00:00", tz="UTC") & Tweets$created_at <= as.POSIXct("2023-04-04 24:00:00", tz="UTC"))
Tweets<-select(Tweets, id, created_at, text, source) #
Conservamos solo las variables relevantes para nuestro analisis
Tweets$text<- gsub("&"," ", Tweets$text) # quitamos los &
Tweets$text<-gsub( "@\\w+", "", Tweets$text) #quitamos las menciones
Tweets$text<- gsub("http\\S*", "", Tweets$text) #quitamos las URLs
Tweets$text<-gsub("[[:punct:]]", "", Tweets$text) #quitamos signos de puntuación
Tweets$text<-gsub("[[:digit:]]", "", Tweets$text) #quitamos dígitos
Tweets$text<-gsub("[^\\x01-\\x7F]", "", Tweets$text) #quitamos todos los caracteres no ASCII p.ej. emoticonos
Tweets$text<-tolower(Tweets$text) #pasamos todas las palabras a minúsculas
Tweets
}

# Hacemos lo mismo para las Menciones (sin filtrado por fechas, pero filtrando por fuentes):

carga_datos_Menciones <- function(nombre_plataforma){
  directorio_analytics<-getwd()
  ruta_fichero<-
paste0(directorio_analytics,'/DDBB/search_tweets/', 'menciones_', nombre_plataforma, '_19Apr.xlsx')
  Menciones<-read_excel(ruta_fichero)
  Menciones<-Menciones %>% extract(source, "source", "Twitter (.*)<") %>%
  filter(source %in% c("for iPhone", "for Android", "for iPad", "Web App"))
}

```

```

Menciones<-select(Menciones, id, created_at, text, source)
Menciones$text<- gsub("&", "", Menciones$text)
Menciones$text<-gsub( "@\\w+", "", Menciones$text)
Menciones$text<- gsub("http\\S*", "", Menciones$text)
Menciones$text<-gsub("[:punct:]", "", Menciones$text)
Menciones$text<-gsub("[:digit:]", "", Menciones$text)
Menciones$text <- gsub("^\\s+|\\s+$", "", Menciones$text)
Menciones$text<-tolower(Menciones$text)
Menciones}

# Ejecutamos las funciones para nuestras plataformas SVOD
(netflix, PrimeVideo, AppleTVPlus, DisneyPlus)

Tweets_netflix<-carga_datos_Tweets('netflix')
Tweets_PrimeVideo<-carga_datos_Tweets('PrimeVideo')
Tweets_AppleTVPlus<-carga_datos_Tweets('AppleTVPlus')
Tweets_DisneyPlus<-carga_datos_Tweets('DisneyPlus')

Menciones_netflix<-carga_datos_Menciones('netflix')
Menciones_PrimeVideo<-carga_datos_Menciones('PrimeVideo')
Menciones_AppleTVPlus<-carga_datos_Menciones('AppleTVPlus')
Menciones_DisneyPlus<-carga_datos_Menciones('DisneyPlus')

# 2. SENTIMENT ANALYSIS PARA LOS TUIITS

library(syuzhet)

#Vamos a trabajar con el lexicon NRC por que permite analizar
sentimientos (polaridad: positiva/negativa)
# y además emociones (anger, anticipation, disgust, fear, joy,
sadness, surprise y trust)

#4.1. Definición de una funcion que nos agregue la polaridad y
emociones de cada tweet a los datasets

# Definimos una funcion para agregar la polaridad usando bing y
nrc (num palabras positivas - num palabras negativas)
# y emociones usando nrc de cada tweet al dataset

```

```

agregar_polaridad_emocion <-function(dataset) {
  dataset <- dataset %>% mutate(nrc_polarity= get_sentiment(text,
method="nrc")) #polaridad nrc
  dataset <- dataset %>% mutate(bing_polarity=
get_sentiment(text, method="bing")) #polaridad bing
  emociones<-get_nrc_sentiment(dataset$text, language="english")
#emociones
  dataset<-dataset %>% mutate (anger=emociones$anger,
anticipation=emociones$anticipation,
                                disgust=emociones$anger,
                                fear=emociones$fear,
                                joy=emociones$joy,
                                sadness=emociones$sadness,
                                surprise=emociones$surprise,
                                trust=emociones$trust,
                                negative=emociones$negative,
                                positive=emociones$positive)
}

```

```

SA_Tweets_netflix<-agregar_polaridad_emocion(Tweets_netflix)
SA_Tweets_PrimeVideo<-
agregar_polaridad_emocion(Tweets_PrimeVideo)
SA_Tweets_AppleTVPlus<-
agregar_polaridad_emocion(Tweets_AppleTVPlus)
SA_Tweets_DisneyPlus<-
agregar_polaridad_emocion(Tweets_DisneyPlus)

```

```
# 2.1 Análisis de la polaridad usando Usando nrc
```

```
# Porcentaje de tweets positivos, negativos y neutros
```

```
polaridad_nrc<-function(dataset) {
```

```

dataset %>% summarise(positivos = 100 * sum(nrc_polarity > 0) /
n(),
                    neutros = 100 * sum(nrc_polarity == 0) /
n(),
                    negativos = 100 * sum(nrc_polarity < 0) /
n())
}

```

```

polaridad_nrc(SA_Tweets_netflix)
# positivos neutros negativos
#      38.2    48.6     13.2

```

```

polaridad_nrc(SA_Tweets_PrimeVideo)
# positivos neutros negativos
#      31.4    59.6     9.00

```

```

polaridad_nrc(SA_Tweets_AppleTVPlus)
# positivos neutros negativos
#      45.5    40.2     14.4

```

```

polaridad_nrc(SA_Tweets_DisneyPlus)
# positivos neutros negativos
#      47.4    46.0     6.53

```

```
# 2.2. Análisis de emociones
```

```
library(RColorBrewer) #paquete con colores para barplot
```

```
# Definimos funcion para que nos plotee un Gráfico de barras para
ver el peso que tiene cada emoción en el corpus
```

```

peso_emociones<- function(dataset,nombre_plataforma){
  barplot(
    sort(colSums(prop.table(dataset[, 7:16]))),
    horiz = TRUE,
    cex.names = 0.7,

```

```

col = brewer.pal(n = 8, name = "Set3"),
las = 1,
main =nombre_plataforma, xlab="Porcentaje"
)
}

peso_emociones(SA_Tweets_netflix, 'Netflix')
peso_emociones(SA_Tweets_PrimeVideo, 'Prime Video')
peso_emociones(SA_Tweets_AppleTVPlus, 'AppleTVPlus')
peso_emociones(SA_Tweets_DisneyPlus, 'DisneyPlus')

#####
#####

# Repetimos el proceso para las Menciones

SA_Menciones_netflix<-
agregar_polaridad_emocion(Menciones_netflix)
SA_Menciones_PrimeVideo<-
agregar_polaridad_emocion(Menciones_PrimeVideo)
SA_Menciones_AppleTVPlus<-
agregar_polaridad_emocion(Menciones_AppleTVPlus)
SA_Menciones_DisneyPlus<-
agregar_polaridad_emocion(Menciones_DisneyPlus)

# POLARIDAD

# A. Usando nrc

polaridad_nrc(SA_Menciones_netflix)
# positivos neutros negativos
#      37,2      47.4      15.4

polaridad_nrc(SA_Menciones_PrimeVideo)

```

```
# positivos neutros negativos
#      40,1      52,8      7.12
```

```
polaridad_nrc(SA_Menciones_AppleTVPlus)
```

```
# positivos neutros negativos
#      40.2      51.4      8.32
```

```
polaridad_nrc(SA_Menciones_DisneyPlus)
```

```
# positivos neutros negativos
#      26.4      63.9      9.69
```

```
# 4.3. Análisis de emociones
```

```
peso_emociones(SA_Menciones_netflix, 'Netflix')
```

```
peso_emociones(SA_Menciones_PrimeVideo, 'Prime Video')
```

```
peso_emociones(SA_Menciones_AppleTVPlus, 'AppleTVPlus')
```

```
peso_emociones(SA_Menciones_DisneyPlus, 'DisneyPlus')
```

