



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Facultad de Ciencias Económicas y Empresariales

**LOVEBRAND Y EL ÉXITO DEL
ENGAGEMENT CON EL CLIENTE.
CASO COLA CAO**

Autor: Marta Noguera Martín

Director: Ana Isabel Jiménez Zarco

MADRID | Abril 2023

ÍNDICE

1. INTRODUCCIÓN	4
1.1. CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA	4
1.2. OBJETIVO DEL TRABAJO.....	5
1.3. METODOLOGÍA.....	5
1.4. ESTRUCTURA	6
2. ANALISIS CONCEPTUAL	7
2.1. EL CONCEPTO DE MARKETING	7
2.1.1. <i>Marketing digital</i>	7
2.2. CONCEPTO DE MARCA	8
2.2.1. <i>Tipos de marcas según la legislación española</i>	10
2.2.2. <i>Tipos de marcas según la percepción del consumidor</i>	10
2.3. CONCEPTO DE <i>LOVE BRAND</i>	10
2.3.1. <i>Ventajas de ser una Love Brand</i>	11
2.3.2. <i>Cómo convertirse en una Love Brand. Estrategias</i>	12
2.3.3. <i>Teoría triangular del amor aplicada a las marcas</i>	13
2.4. COMUNICACIÓN DE MARKETING EN LAS REDES SOCIALES.....	15
2.4.1. <i>Digitalización de las marcas</i>	16
2.4.2. <i>¿Qué es una red social?</i>	16
2.4.3. <i>Campañas de publicidad en las redes sociales</i>	17
2.5. APLICACIÓN CASO CONCRETO	18
3. MARCA COLA CAO	20
3.1. HISTORIA DE COLA CAO	20
3.2. ¿ERES MÁS DE COLA CAO O DE NESQUIK?	21
3.3. COLA CAO Y SUS ANUNCIOS. ESTRATEGIAS DE MARKETING.....	21
3.4. COLA CAO ¿ <i>LOVE BRAND</i> ?	23
4. ANÁLISIS EMPÍRICO	26
4.1. <i>TEXT MINING</i>	26
4.1.1. <i>¿Qué es el text mining?</i>	26
4.1.2. <i>Descarga de tweets</i>	27
4.1.3. <i>Limpieza de datos</i>	28
4.1.4. <i>Palabras más repetidas. Visualización</i>	33
4.2. <i>SENTIMENT ANALYSIS</i> O ANÁLISIS DE SENTIMIENTO	34
4.2.1. <i>¿Qué es un análisis de sentimiento?</i>	34
4.2.2. <i>Métodos</i>	35
4.2.3. <i>Método NRC y las ocho emociones de Robert Plutchik</i>	35
4.2.4. <i>Análisis de sentimiento en R Studio</i>	37
5. CONCLUSIONES DEL ESTUDIO	42
6. BIBLIOGRAFÍA	44

TABLA DE ILUSTRACIONES

Ilustración 1: Cambio de logotipo Cola Cao (colacao.es).....	20
Ilustración 2: Preferencias de cacao en España 2022. Fuente Datacentric. (datacentric.es).....	21
Ilustración 3: Instalación de paquetes.....	27
Ilustración 4: Carga de librerías.....	27
Ilustración 5: Descarga de tweets.	28
Ilustración 6: Filtración de variables principales.....	28
Ilustración 7: Fuentes de los tweets.	29
Ilustración 8: Términos de la matriz. Comprobación.	29
Ilustración 9: Creación del corpus.	30
Ilustración 10: Creación de la Document Term Matrix (DTM) y visualización de sus términos.	30
Ilustración 11: Palabras más repetidas en los documentos.....	30
Ilustración 12: Eliminación de stopwords.	31
Ilustración 13: Palabras DTM sin stopwords, signos de puntuación, menciones o páginas web.....	32
Ilustración 14: Palabras más utilizadas sin stopwords, signos de puntuación, menciones o páginas web.	32
Ilustración 15: Stemming.....	33
Ilustración 16: Gráfico de dispersión de las palabras más utilizadas y su frecuencia.	33
Ilustración 17: Código gráfico de dispersión.....	33
Ilustración 18: Nube de palabras.	34
Ilustración 19: Código nube de palabras.	34
Ilustración 20: Rueda de las emociones de Robert Plutchik.	36
Ilustración 21: Análisis de sentimiento con los métodos AFINN, BING y NRC.	37
Ilustración 22: Columnas polaridad en el dataframe.	38
Ilustración 23: Porcentaje de sentimiento de TweetsColaCao.	38
Ilustración 24: Resultado porcentual del sentimiento positivo, negativo y neutro según fuente de emisión del tweet.	38
Ilustración 25: Aplicación del método NRC al dataframe TweetsColaCao.....	39
Ilustración 26: Explicación método NRC.....	39
Ilustración 27: Media de las emociones, máximos y mínimos.....	40
Ilustración 28: Gráfico de barras emociones.	40
Ilustración 29: Tweets con sentimiento positivo sobre Cola Cao.	41
Ilustración 30: Tweets con sentimiento negativo sobre Cola Cao.....	41
Ilustración 31: Código extracción emociones.	41

1. INTRODUCCIÓN

1.1. CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA

Vivimos en un entorno donde el consumo es masivo, en el que la oferta de productos a la que el consumidor tiene acceso sigue incrementando. El nivel de competencia entre las marcas es cada vez mayor, ya que éstas, gracias al entorno digital pueden acceder a un mercado global de una forma sencilla y económica. La época de la venta exclusivamente en el entorno físico ha pasado, y actualmente sin barreras temporales y espaciales, y a un *click de distancia* los consumidores tienen la posibilidad de conocer, comparar, dialogar, comprar y en definitiva relacionarse con una infinidad de marcas.

En esta situación, la posición estratégica de la marca en el mercado no sólo depende de las características funcionales de sus productos; además existen otros factores que la determinan y hacen de ella una marca, reconocida, apreciada e incluso *amada*.

El concepto de *Love Brand* ha ido adquiriendo mucha importancia y cada vez suena más en el mundo del *consumer engagement*. Las marcas están dirigiendo sus esfuerzos a establecer una vinculación con el consumidor, una conexión emocional que no solo cale en su mente, sino también en su corazón. A través de canciones, historias o experiencias, estas marcas buscan situarse en una posición privilegiada con respecto a sus competidores, no por ser la preferida, si no por ser una marca a la que los consumidores “quieren”.

Convertirse en una *Love Brand* requiere mucho esfuerzo y, sobre todo, mucho tiempo, ya que con el paso de los años las marcas van tomando forma en la mente de los consumidores y a crear lazos afectivos que refuerzan la vinculación emocional. Con esto, en esta época de nuevas tecnologías y constante desarrollo, también cobra especial relevancia la posición que la empresa tenga en las redes sociales, ya que es uno de los métodos más utilizados actualmente para poder establecer una conexión con el cliente.

Encontramos multitud de marcas en el mercado, pero solo algunas tendrán la consideración de *Love Brand*, ya que no todas han conseguido establecer esta vinculación emocional con los consumidores. En España, una de las marcas referentes en el *marketing*, especialmente por su capacidad de adaptarse a las circunstancias del mundo que les rodea es Cola Cao. Esta compañía distribuidora de cacao en polvo lleva casi un siglo en el mercado y que ha conseguido situarse en el corazón de muchos españoles y de

muchas generaciones. Con el paso de tiempo ha ido ganando popularidad y, sin ninguna duda, ha conseguido su situación en el mercado español y ha conseguido distinguirse de sus competidores. Sin embargo, ¿es Cola Cao una Love Brand? ¿Cuál es el sentimiento de los consumidores hacia Cola Cao?

1.2. OBJETIVO DEL TRABAJO

El objetivo principal del presente trabajo es analizar cuáles son las emociones y sensaciones que los consumidores tienen hacia la marca Cola Cao. Para ello, realizaré un estudio dual.

Por un lado, busco profundizar en el término *Love Brand*, qué es, cómo se consigue y cuáles son sus características principales. Dentro de esta parte mayormente teórica se analizarán cuestiones como el concepto del amor, las marcas, el marketing y las redes sociales. Para ello, se analizará cómo actúan *Love Brands* en el mercado y sobre cómo adquieren esta posición, delimitando las estrategias utilizadas, las dimensiones en las que actúan y los beneficios que se obtienen de ello.

Por otro lado, una vez comprendidas las características del perfil de las *Love Brands*, se entrará a estudiar una marca concreta, en este caso Cola Cao. Se analizará su historia, su posible vinculación con el consumidor y las cuestiones que hacen que esta marca sea única en el mercado. El objetivo principal es determinar si Cola Cao puede considerarse una *Love Brand* y comprobar si el usuario tiene un sentimiento positivo o negativo hacia la marca y cuáles son las emociones que esta genera en sus consumidores.

1.3. METODOLOGÍA

Para cumplir los objetivos establecidos previamente, se realizará, en primer lugar, un análisis teórico mediante el uso de técnicas cualitativas, mediante la revisión de literatura académica y científica. Con investigación realizada en internet y en documentos relativos a tal efecto se tratará de establecer cuáles son los atributos de una *Love Brand* y de los conceptos relacionados con ella.

En segundo lugar, tras estudiar teóricamente la posición que Cola Cao ostenta en el mercado nacional, se realizará un estudio empírico mediante el uso de *Data Analytics*. A través de las técnicas de *Text Mining* y *Sentiment Analysis*, se analizarán datos de la red social *Twitter* con el lenguaje de programación que proporciona la plataforma R Studio

para poder determinar si el sentimiento generalizado hacia esta marca es positivo o negativo.

1.4. ESTRUCTURA

Este Trabajo de Fin de Grado se dividirá en cinco partes principales. En primer lugar, el presente capítulo, la introducción, el cual explica el contexto y la justificación del estudio, así como los objetivos, la metodología y la estructura que se seguirá durante el trabajo. En segundo lugar, el capítulo relativo a conceptos teóricos, que analizará conceptualmente el marketing, las marcas y el perfil, los atributos y las ventajas de las *Love Brands* en su conjunto. Se tratará también uno de los principales temas que guardan relación con ello en la actualidad, la publicidad en las redes sociales. En tercer lugar, procede realizar una descripción teórica de la marca Cola Cao en su conjunto, su historia, sus competidores y su situación privilegiada en el mercado, así como las estrategias que ha utilizado para lograr situarse en la posición en la que se encuentra. En cuarto lugar, se estudiará empíricamente la marca, a través de las técnicas que proporciona el *Data Analytics*, siguiendo la metodología explicada anteriormente. Por último, se tratará de concluir el trabajo tomando como referencia el análisis empírico realizado y confirmar si Cola Cao puede, efectivamente, considerarse una *Love Brand*. Se añadirá al final del documento el Anexo I, que contendrá el *script* del código de R utilizado para realizar el estudio empírico en el programa R Studio.

2. ANALISIS CONCEPTUAL

2.1. EL CONCEPTO DE MARKETING

Marketing es un término anglosajón que derivado del término “market” hace referencia a las técnicas que las organizaciones desarrollan en el mercado. Para Philip Kotler, el marketing es “es 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”. Para la American Marketing Association (2017) el marketing es 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. Es decir, el objetivo de estas técnicas es, principalmente, llegar a darle valor a la marca y transmitir este valor creado a los consumidores, cuestión que se acerca mucho a la idea de una Love Brand (Fuente, 2022).

El marketing no es una invención por parte de una empresa o de una persona, sino que tiene su origen dentro de un proceso de evolución económica. Su desarrollo ha estado marcado por aspectos políticos y sociales durante el paso de los años y se ha ido adecuando al momento actual en el que vivimos (Ballesteros, 2020). Los métodos habituales de marketing han ido evolucionando en los últimos años, pasando desde la tradicionales 4 Ps de Porter (1960), hasta el actual marketing digital, social media marketing, y en los cuales resulta altamente efectivo el uso de rostros conocidos del momento, hasta canciones adaptadas, covers o gingles del momento. El marketing es una disciplina en constante evolución, pero para innovar es necesario el aprendizaje continuo, y en ocasiones técnicas u elementos que funcionaron, pueden volver a garantizar el éxito si somos capaces de adaptarlos al nuevo contexto (La Brújula del Canto, 2021).

2.1.1. Marketing digital

En el siglo XXI, el siglo de las nuevas tecnologías, considero de especial relevancia hacer una mención al término de marketing digital. Este es un término empresarial muy amplio que hace referencia al marketing que se realiza en el entorno digital, en particular la web 2.0s (Rock Content, 2017).

En un comienzo, el marketing desarrollado en el entorno digital se limitaba a las funcionalidades de la web 1.0, siendo un marketing unidireccional, donde las empresas

decidían qué información se transmitía de sus marcas y de qué manera, tratándose de una comunicación exclusivamente de la empresa al consumidor. Sin embargo, con a comienzos de siglo XXI la aparición de la web 2.0 (las redes sociales, foros o blogs) hace posible dos hechos: (1) la bidireccionalidad en la relación entre la empresa y los consumidores; (2) el empoderamiento del consumidor como creador de contenidos.

El marketing digital está en constante crecimiento, aún sin olvidar las formas tradicionales de marketing. Esto se debe principalmente a las ventajas, que en términos de costes, alcance y eficiencia proporcionan este tipo de acciones.

A modo de ejemplo, podemos tomar en consideración la campaña de publicidad masiva que se ha dado en las redes sociales de las marcas Casio y Twingo. A raíz de una canción viral en la que se mencionaba a estas marcas como “malas”, optaron por aprovecharse del uso de redes sociales y darse mayor publicidad. Sin apenas costes, las marcas publicaron una serie de comentarios en redes sociales, consiguiendo que durante semanas no dejara de hablarse de ellos. Casio, por su parte, aseguró que multiplicaron sus ventas hasta tres veces tras la salida de la canción (Martínez, 2023), mientras que Twingo comentaba que la visitas a la página web de ese modelo llegaron a incrementar en un 3000% en apenas una semana (ReasonWhy, 2023).

2.2. CONCEPTO DE MARCA

Cada vez hay más productos similares en el mercado y lo que antes era una tarea sencilla, como comprar unas zapatillas o una bebida, se ha convertido en algo complicado. Los productos se parecen mucho entre ellos y resulta muy difícil para los consumidores distinguir las cualidades de cada uno para poder elegir. Aquí es donde la marca empieza a cobrar mayor importancia, ya que es, principalmente, un signo distintivo. La marca tiene como objetivo el diferenciar o individualizar en el mercado unos productos o servicios de otros productos o servicios similares, buscando ser un indicador de origen empresarial y un garante de la calidad del producto (OEPM, s.f.).

La definición de marca nos la ofrece la legislación española en la Ley 17/2001, de 7 de diciembre, de Marcas, al establecer que: *“Podrán constituir marcas todos los signos, especialmente las palabras, incluidos los nombres de personas, los dibujos, las letras, las cifras, los colores, la forma del producto o de su embalaje, o los sonidos, a condición de que tales signos sean apropiados para:*

- a) distinguir los productos o los servicios de una empresa de los de otras empresas y*
- b) ser representados en el Registro de Marcas de manera tal que permita a las autoridades competentes y al público en general determinar el objeto claro y preciso de la protección otorgada a su titular” (Ley de Marcas, BOE 2 de diciembre de 2001).*

El concepto de marca, sin embargo, ha sido muy discutido en el ámbito académico y no se ha logrado llegar a un concepto común. Cada experto entiende la marca de una forma distinta, unos tienen en cuenta únicamente el propósito de la marca, otros consideran la función de sus características y hay quien se aproxima a este concepto de una forma más amplia, teniendo en cuenta las ventajas que una marca tiene en un producto. Podríamos mencionar, a título ejemplificativo, varios grupos de expertos que se aproximan a la marca como un concepto distinto. Algunos grupos de autores lo entienden como un instrumento legal, otros lo consideran como un logotipo, y también se ha identificado el concepto de marca con el de empresa. Hay otras corrientes muy vinculadas con el marketing que lo ven como una relación entre la empresa y la marca o como la imagen que se forma en la mente de los consumidores (Gallart-Camahort, Callarisa y Sánchez, 2019).

En línea con lo anterior, y tomando como referencia la conceptualización que ofrece la American Marketing Association (1960), una marca es “un nombre, símbolo o diseño, o una combinación de ellos, dirigida a identificar los productos o servicios de un vendedor o grupo de vendedores y a diferenciarlos del resto de competidores (Gallart-Camahort, Callarisa y Sánchez, 2019).

Por tanto, la marca cumple 3 funciones que se complementan entre sí; identificar, diferenciar y proteger. Esta puede presentar una doble identidad: la escrita y/o verbal definida por el nombre, y la visual, determinada por el símbolo o signo, color o grafía que le la representa. la función visual, comprendida por el logo, el gráfico y el color. Sin embargo, el concepto de marca va más allá. Kotler lo define a la perfección, considerando que independientemente de si la marca se trata de un nombre, una marca comercial, un logotipo u otro símbolo, una marca es la promesa de una parte vendedora de proporcionar, de forma consistente a los compradores, un conjunto específico de características, beneficios y servicios (Amstrong y Kotler, 2013). Es el signo que va a identificar al producto y con el que los consumidores van a establecer una conexión emocional, que podrá ser positiva o negativa (Valera García, 2016).

2.2.1. Tipos de marcas según la legislación española

La Oficina Española de Patentes y Marcas destaca la clasificación de las marcas según dos parámetros concretos, la finalidad y el formato del signo que protege. El primer grupo diferenciará marcas de productos y servicios individualizados, marcas colectivas, que distinguen productos de una asociación de fabricantes, comerciantes o prestadores de servicios, y marcas de garantía, que certifican las características comunes de los productos y servicios elaborados o distribuidos por personas autorizadas para ello. Por su parte, el segundo grupo distingue marcas denominativas, figurativas, de color, de posición, sonoras o de movimiento, entre otros (OEPM, s.f.).

2.2.2. Tipos de marcas según la percepción del consumidor

Una clasificación más adecuada al tema que nos ocupa es la que distingue entre tres tipos de marcas según lo que estas pueden llegar a transmitir a cada consumidor.

- ***Invisible brands***: estas son marcas que pasan desapercibidas para el consumidor, que están a su alrededor pero que no entran en su mundo, no llega a verlas y ni siquiera le interesan.
- ***Everyday brands***: son marcas que llevan tiempo integradas en la vida diaria del consumidor y este confía en ellas, ya que solucionan y aportan algo en su día a día.
- ***Love Brands***: son aquellas marcas que consiguen atrapar al consumidor, ya sea por su imagen, por los beneficios que aportan o por los sentimientos que transmiten. Ya sea de manera consciente o inconsciente estas marcas hacen sentir y el consumidor no solo decide comprar el producto o servicio, si no que se convierte en embajador de la marca por auténtica convicción (Marketing Directo, 2016).

2.3. CONCEPTO DE LOVE BRAND

Las marcas compiten en mercados muy amplios y saturados, donde la confianza hacia las marcas ha ido descendiendo. Sin esa confianza, los clientes pierden interés y desvían su atención hacia la gran variedad de productos similares que pueden encontrar. Por ello, las marcas necesitan individualizar sus esfuerzos y construir relaciones más personales con los consumidores.

Tendemos a pensar que cuando compramos algo lo hacemos siempre con un componente racional, que identificamos las necesidades, las características del producto y las ventajas y los inconvenientes que este puede tener. Sin embargo, esto no suele atender a lo racional, si no que esta influenciado por percepciones y sensaciones. Gerald Zaltman, un profesor de Harvard, explica que el 95% del proceso decisivo de compra ocurre en el subconsciente (Casarotto, 2021).

Las marcas tienen un componente humano. Las marcas nacen, crecen, se desarrollan y, en su caso, muchas mueren. Las personas no las pueden ver, pero, cuando se consigue crear una vinculación con la marca, el cliente la guarda en un rincón de su memoria. Las marcas no son simplemente empresas, sino que son el primer regalo que te hacen de niño, una cena muy especial o el producto que compraste y que te acompañó en aquel viaje.

Una *Love Brand*, como su propio nombre indica en inglés, es una marca que genera una atracción muy fuerte para el consumidor, quien compra y defiende sus productos no porque sea la favorita, sino porque es una marca "amada". Estas son marcas que consiguen crear un vínculo emocional con el cliente, mediante una relación trascendente entre empresa y marca, más allá de lo puramente tradicional. La relación que se consigue es tan fuerte que el consumidor ya no tiene ojos para los competidores y empiezan a defender la marca ante todo lo demás (Jurado González, 2019).

2.3.1. Ventajas de ser una *Love Brand*

Está claro que ser una marca adorada por el consumidor es algo positivo, pero ¿cuáles son las mayores ventajas para las marcas una vez lo consiguen?

En primer lugar, los consumidores tendrán más lealtad hacia la marca. Parece algo evidente, pero lo que se consigue con esto es que las marcas no tengan que preocuparse por sus competidores. El vínculo es tan fuerte que va a ser difícil que el consumidor cambie esa marca por otra y, en muchas ocasiones, incluso se perdonan fallos realizados por la marca debido a esta relación (Casarotto, 2021).

En segundo lugar, los fans de la marca no tendrán ningún problema con el precio de los productos. Es más fuerte la relación con la empresa que el precio de esta, por lo que las *love brands* pueden en muchas ocasiones subir los precios más que sus competidores y no sufrir una disminución en las ventas (Casarotto, 2021).

Por último, la marca será protegida por los clientes. Se convierten en verdaderos defensores de estos productos, ante cualquier tipo de crítica recibida y ayudando a preservar su imagen. Y no solo eso, sino que también se convierten en promotores de la marca, especialmente en redes sociales. Les gusta hablar sobre la compañía, sus beneficios y sus valores. Quieren ver como la marca y crece y contribuir a este crecimiento dentro de lo posible (Casarotto, 2021).

2.3.2. Cómo convertirse en una *Love Brand*. Estrategias.

El objetivo de estas marcas es conectar con los usuarios a través de la imagen, de lo que presentan o de cómo lo expresan. Debe escucharse a los consumidores y descubrir qué le gusta y cuáles son sus opiniones acerca de la marca en cuestión. Además, es importante que el consumidor sienta que se le está prestando atención y que lo que manifiesta sobre su experiencia tiene algo de importancia para la marca.

Las empresas más innovadoras se centran cada vez más en redefinir como sus marcas se perciben, pero, sobre todo, como se viven. El factor fundamental en esta cuestión es la empatía, el pensar desde la perspectiva del cliente y no desde la de la propia empresa. Los clientes priorizarán a las marcas que les comprendan, lo cual reforzará su confianza en ellas. La marca adquiere un rol humano a ojos del cliente, y estos comparten valores (Puig de Falcó, s.f.).

El objetivo es construir una marca de la que todo el mundo se sienta orgulloso de formar parte. Las estrategias que adoptan las marcas para posicionarse como *love brand* en el mercado son numerosas, estas son algunas de ellas (Casarotto, 2021):

- **Construir una personalidad única para la marca.** El objetivo es crear algo diferente que se distinga de las demás marcas del mercado. Se busca construir una serie de valores intrínsecos en la empresa y una visión del mundo que defiendan con todo lo que hacen. Tienen que saber por qué existen y para qué.
- **Responsabilidad social.** Los consumidores de hoy en día quieren sentir una conexión personal con la marca y, para ello, las marcas deben tomar una serie de intereses para construir un mundo mejor. Los fines sociales que toman las marcas en este aspecto son amplios, desde problemas sociales muy frecuentes como el racismo o el feminismo hasta el cambio climático.

- **Tener en cuenta el poder de la nostalgia.** A todo el mundo le gusta recordar tiempos y experiencias pasadas, y las marcas que consiguen vincular al cliente con conexiones positivas del pasado son más propensas a convertirse en love brands.
- **Construir un estilo de vida al que el cliente aspire.** La idea es dar al consumidor algo increíble a lo que aspirar, una forma de ver la vida y de actuar que puede ser suya con una simple compra.
- **Mantenerse al día con las novedades de la sociedad.** Una auténtica love brand nunca deja de moverse ni de adaptarse a las circunstancias. Debe seguir el mismo desarrollo que siguen sus clientes, explorando nuevas modas y conociéndose cada vez más a sí misma y a sus clientes. Esto incluye nuevas tendencias, redes sociales, gustos...
- **Conseguir prensa positiva.** Uno de los elementos más importantes para toda empresa es causar buena impresión al público en general, mejorando la percepción sobre la marca y su reputación. Muchas marcas optan por construir desde dentro un departamento de prensa interno que, casi siempre, se convierte en una muy buena inversión.
- **Construir una marca de la que gente se siente orgulloso por trabajar en ella.** Es importante que los clientes sientan esta conexión, pero es también de especial relevancia que lo sientan los trabajadores. Si la gente que trabaja para la empresa comparte los valores y los defiende y se siente orgullosa de ello, es muy probable que la visión que el público percibe de esta marca sea positiva. Según un estudio de Glassdoor, las compañías que tienen trabajadores felices e inspirados son mucho más propensas a tener clientes satisfechos y fieles a la marca.

2.3.3. Teoría triangular del amor aplicada a las marcas

El amor es una de las emociones más complejas de las personas y, durante muchos años, no despertó ningún tipo de interés en la ciencia, si no que era objeto de literatura, poesía o arte. A partir de la década de 1970, empezaron a surgir teorías científicas en relación con el amor. Entre ellas, Rubin consideraba que el sentimiento esta formado por dos

dimensiones: el amar (atracción física, predisposición por ayudar, exclusividad...) y el gustar (respeto o confianza) y Lee realizaba una analogía entre el amor y los colores, entendiendo que hay tres estilos primarios de amor que dan lugar a otros tipos de amor secundarios (Cassep-Borges, 2009).

Sin embargo, dentro de estas teorías, una de las más relevantes es la Teoría Triangular del Amor de Sternberg, que divide el sentimiento en tres componentes principales: la intimidad, la pasión y la decisión/compromiso. La intimidad se responsabiliza de los sentimientos de proximidad y conexión en la relación, la pasión está marcada por atracción física y sexual y la decisión/compromiso es la certeza de amar y ser amado y las ganas de mantener esa relación a largo plazo (Cassep-Borges, 2009).

En estudios posteriores se ha establecido una conexión de estas distintas formas de amar con la relación del consumidor con la marca, entendiendo que las relaciones humanas son mucho más complejas que las relaciones que la persona puede tener con un producto o empresa. Para ello, se analiza la vinculación que los tres componentes del amor pueden tener con el amor por una marca y, posteriormente se sugieren 8 tipos de sentimientos distintos hacia estas.

La intimidad consiste en gustar, en sentirse de alguna forma cercano y conectado a la marca en cuestión. Ejemplos serían una niña y su inseparable muñeca o un coleccionista con un cuadro, que los tratan como si formaran parte de ellos. La pasión, en cambio, es el componente motivacional, referido a autoestima, cuidado o dominación. En estos casos se relaciona más con los objetos a los que la persona recurre en momentos de tristeza, miedo o ansiedad. La decisión, por su parte, es el reconocimiento a corto plazo de que te gusta algo/alguien, mientras que el compromiso es la seguridad de elegir a esa persona o a ese objeto. A largo plazo los consumidores se convierten en fieles seguidores de las marcas por propio convencimiento, fidelidad o inercia (Madden y Shimp, 1988).

Una vez establecido esto, el análisis se basa en la propuesta de los 8 tipos de amor, según intervengan o no estas tres características, ya sea de forma conjunta o separada.

- ***Nonliking*** (no gustar). No necesariamente se refiere a disgustar, si no que la persona directamente no tiene ningún tipo de sentimiento hacia la marca o el producto.

- **Liking (Gustar)**. Los consumidores sienten algún tipo de afinidad hacia la marca, pero no tienen ningún interés particular en comprar sus productos.
- **Infatuation (Encaprichamiento)**. Se caracteriza por ser un sentimiento muy breve, en el que la persona quiere comprar ese producto, igual por presión social o por querer formar parte de un grupo. Esto suele ocurrir a los niños pequeños, que se empeñan en comprarse una cosa que a los pocos meses van a dejar de utilizar, o, incluso, que ni siquiera les gusta, pero sus amigos tienen.
- **Functionalism (Funcionalismo)**. La compra de ese producto es únicamente porque resulta sencillo. Esto consistiría, principalmente, cambiar un producto de la misma marca por no arriesgarse a probar uno nuevo.
- **Inhibited desire (Deseo inhibido)**. Se da cuando la persona quiere utilizar un producto, pero hay algo que se lo impide. Ejemplos serían una niña a la que su madre no le deja comprarse maquillaje o un padre de familia que no se compra un Ferrari porque no puede llevar a los niños.
- **Utilitarianism (Utilitarismo)**. Un producto resulta útil y lo siguen comprando, pero el consumidor no tiene una vinculación especial con la marca. Es uno de los sentimientos más frecuentes hacia las marcas, el comprar algo porque lo necesitas y no porque efectivamente te guste.
- **Succumbed desire (Deseo sucumbido)**. En ocasiones, los consumidores optan por comprar algo que no necesariamente les gusta en vez de algo que sí por presión que les hace rendirse ante ello.
- **Loyalty (Lealtad)**. Este es el caso de las *love brands*. Es el sentimiento más fuerte que se puede tener hacia una marca y supone la vinculación del cliente con la marca y el convencimiento de querer comprarlo.

2.4. COMUNICACIÓN DE MARKETING EN LAS REDES SOCIALES

Una cuestión que es importante tratar con el auge de las nuevas tecnologías es la digitalización que se está viviendo con las marcas en los últimos años. Como hemos comentado, las marcas están cada vez más presentes en las redes sociales, no solo como

manera de escuchar a los clientes, sino también como medio para realizar campañas de marketing. Este último aspecto queda fuera del objeto de este estudio, pero sí es relevante esta digitalización creciente que estamos viviendo, ya que es lo que permitirá realizar el análisis empírico posterior sobre los comentarios extraídos de Twitter. Por ello, considero importante hacer una introducción previa sobre la presencia de las marcas en internet y el concepto de red social.

2.4.1. Digitalización de las marcas

El incremento del *e-commerce* ha sido masivo en la última década. La digitalización del comercio es el proceso por el que la tecnología digital se implanta en la economía y afecta a la producción, el consumo y a la propia organización en sí misma. Conlleva una mayor automatización y rapidez y puede suponer una oportunidad para el comercio, al permitir conocer en mayor detalle las necesidades de sus clientes (Interempresas, 2022).

Según recientes estudios el 83% de los consumidores necesitan una experiencia online positiva para mantenerse leales a una marca y un 39% lo consideró como un aspecto crítico. La personalización ayuda mucho en este recorrido y para poder lograr una experiencia con el cliente tienes que conocer a tu audiencia. Ofrecer satisfacción es crucial, no con cada transacción, sino con cada contacto que el cliente tenga con la marca. Ofrecer un buen servicio no es difícil, pero construir confianza implica al corazón (Puig de Falcó, s.f.).

En la actualidad, una de las maneras más sencillas de poder escuchar al consumidor y de que este se comunique con la empresa es, como comentábamos antes, las redes sociales. Estas permiten crear comunidades activas en las que preguntar por sabores, colores, embalaje, olores... Si esto se hace bien y se presta atención a los usuarios, puede suponer una fuente de información muy valiosa para la marca y les puede ayudar a adquirir la consideración que esperan de sus clientes (Impact5, s.f.).

2.4.2. ¿Qué es una red social?

Las redes sociales son un espacio digital en el que personas, marcas y entidades pueden crear una red de contactos e interactuar (Florido, 2021). Una red social, al fin y al cabo, no es más que una plataforma online que permite a los usuarios, a través de un perfil, compartir datos e información tanto de sí mismos como de terceros, además de interactuar

y compartir este contenido con otros miembros de la plataforma (Morales Cáceres, s.f.). Con el paso del tiempo estas redes se han ido masificando, convirtiéndose, en muchas ocasiones, en un espacio de anuncio de marcas o para darse a conocer públicamente.

Podemos dividir el uso de redes sociales en dos tipos. Por un lado, encontramos las redes sociales verticales, centradas en un tema común entre todos los usuarios, como LinkedIn o Spotify, y, por otro, las redes sociales horizontales, sin temática concreta, siendo los propios usuarios los que tratan diversos asuntos de su interés, como Instagram, Facebook o Twitter (Florido, 2021).

2.4.3. Campañas de publicidad en las redes sociales

Cabe mencionar brevemente, como complemento a lo expuesto previamente, que las redes sociales no solo dan la posibilidad de establecer una comunicación bidireccional entre empresa y consumidor para conocer sus experiencias, sino que también se utilizan como medio publicitario. A este respecto encontramos tanto el concepto de *influencer* como el de *brand ambassador*, que, aunque similares, tienen algunas diferencias.

Siempre ha sido muy habitual ver anuncios en los que aparecen personajes públicos, debido a su reputación y prestigio, pero en los últimos años, debido al crecimiento de las redes sociales, ha ido apareciendo el concepto de *influencer*. Un *influencer* es, como su propio nombre indica, aquel que busca influenciar a través de las redes sociales, utilizándolas como herramienta para comunicar e impactar en sus seguidores (Translinguo, 2022). Por lo general, estas personas no son en origen famosas, pero, por la actividad que generan en redes sociales, llegan a acumular un gran número de seguidores, siendo por ello objeto de contratación por distintas marcas para publicitar sus productos. Con este tipo de marketing digital se trata de vincular un producto con ese *influencer* y su público, lo cual consigue alcanzar a un mayor número de consumidores *target* de una forma sencilla e integrada, ya que las menciones publicitarias se insertan como parte natural de ciertos contenidos generados diariamente por estos líderes de opinión (su producto de maquillaje favorito, los zapatos que utiliza a diario o el restaurante que recomienda para algún tipo de evento).

Por su parte, los *brand ambassadors* son un elemento clave en el mundo de las *love brands* y surgen como consecuencia de una buena estrategia para convertirse en una marca adorada por sus clientes. Los *brand ambassadors* no son más que clientes que

aman la marca, les gusta hablar sobre ella y se hacen fotos y videos. Son las personas que publicitan tu marca y la recomiendan y defienden en redes sociales.

Los embajadores promocionan y divulgan no solo el producto en sí, sino también los valores y servicios de la marca que representan. Los beneficios son múltiples, ya que refuerzan las dimensiones de la marca y amplifican su visibilidad, la diferencian de sus competidores, impulsan las ventas y fortalecen los valores de la marca, al estar lineados con los suyos propios. Es también muy relevante el hecho de que estas personas consiguen conectar con un público objetivo, que les sigue y percibe su estilo de vida, además de poder sentirse identificado con las situaciones humanas que muestran a diario en las redes sociales (Marketing Spain, 2022).

La diferencia primordial con los *influencers* es que los embajadores de la marca tienen una vinculación especial y mucho más emocional y sincera con la marca que defienden. Cuando un *brand ambassador* comparte contenido sobre una marca en las redes sociales, esto se percibirá de forma mucho menos publicitaria que en el caso de los *influencers*, ya que estos últimos son capaces de anunciar todo tipo de contenido y, en muchos casos, están perdiendo la autenticidad que transmitían cuando empezó a aparecer el concepto.

Según un estudio de *Active Campaign Spain*, el 66% de los profesionales de marketing digital en nuestro país optan por utilizar *influencers* y *brand ambassadors* para sus campañas de publicidad, previendo un crecimiento del 25% anual (Active Campaign, 2022).

2.5. APLICACIÓN CASO CONCRETO

Tomando como referencia lo ya visto sobre las Love Brands, estas tratan, a través de una serie de estrategias, posicionarse en una situación concreta con respecto a sus competidores y a ojos de sus clientes. Dentro de estas estrategias encontramos el construir una personalidad única para la marca, preocuparse por participar activamente en proyectos de responsabilidad social y tener en cuenta el poder de la nostalgia. Una de las marcas españolas que más destaca a este respecto es la marca Cola Cao. Esta es una empresa que ha estado muchos años en el mercado y ha seguido diversas estrategias para conseguir encontrarse en la posición en la que se encuentran hoy en día.

Cola Cao, en este sentido, a lo largo de los años, se ha implicado cada vez más en proyectos de responsabilidad social, especialmente en lo relacionado con los niños y el deporte. Por un lado, al ser patrocinadores del equipo olímpico español, han promovido los deportes y han relacionado su producto a una vida sana. Por otro lado, siempre ha estado muy concienciado de los problemas que afectan a la infancia, entre los cuales, a parte del deporte, encontramos el acoso escolar. La marca sigue promoviendo iniciativas a este respecto, lo cual veremos posteriormente.

Adicionalmente, una de las sensaciones que más utiliza Cola Cao para llegar a sus consumidores, es el poder de la nostalgia, el cual lleva promoviendo desde sus primeros años en el mercado nacional. Es un producto, en origen, destinado a los niños, pero que ha conseguido mantenerse también en la mente de los adultos que empezaron a tomarlo cuando eran pequeños. Es una marca que ha estado presente en la infancia de muchos, tanto en desayunos como en meriendas con amigos y está ligada a su consumo durante estas épocas de niñez. Por ello, ha conseguido posicionarse en el corazón de muchos y continúa haciéndolo, ya que los sentimientos que ahora se dirigen a los niños, también se están destinando a los mayores, y viceversa. Esto se justifica en que a todo el mundo le gusta rememorar experiencias pasadas con cierto cariño, y siempre se va a vincular con un sentimiento positivo que puede llegar a crear esta vinculación emocional de la que veníamos hablando, entre el cliente y la marca.

La red social Twitter es una plataforma en la que más del 50% de los usuarios están entre los 25 y los 44 años y, aunque también tiene usuarios de menos edad, su uso está mayormente consolidado y enfocado en edades posteriores (Santin, 2022). Por ello, este público es en el que se enfoca la marca Cola Cao para despertar el poder de la nostalgia, y despertar emociones relacionadas con su infancia.

Con esto, cabe analizar los comentarios que se han hecho en Twitter sobre Cola Cao, para conocer cuáles son los sentimientos que sus usuarios tienen hacia la marca y cuáles son las emociones principales que se despiertan en ellos. Con carácter previo conviene analizar la historia de Cola Cao, su posición en el mercado y sus estrategias diversas, para comprender qué ha llevado a la marca a despertar estos sentimientos y si esta puede ser considerada como *Love Brand*.

3. MARCA COLA CAO

3.1. HISTORIA DE COLA CAO

Cola Cao surge en un pequeño local de Barcelona, en pleno barrio de Gracia, cuando José María Ventura y José Ignacio Ferrero deciden elaborar artesanalmente un cacao soluble. La marca se registró comercialmente en el año 1945 y se lanzó al mercado español un año más tarde (ColaCao, 2021). Este fue el inicio de uno de los mayores imperios existentes en España, que actualmente está compuesto por marcas muy reconocidas en nuestro país, como Nocilla, Paladín o Cuétara (Invertia, 2019). Está presente en más de 6 millones de hogares y en nuestro país se toman más de 50 vasos de leche con Cola Cao por segundo, casi 5 millones de vasos al día (ColaCao, 2021).

La marca ha experimentado numerosos cambios a lo largo de los años, adaptándose a las circunstancias del momento y a los cambios sociales que han ido ocurriendo en los más de 75 años que llevan en el mercado. Cola Cao ha ido modificando su formato, su logo, su variedad de productos y sus estrategias de marketing. Todo esto ha ayudado a la marca a posicionarse en el mercado de forma privilegiada, siendo una de las marcas más elegidas por los consumidores.

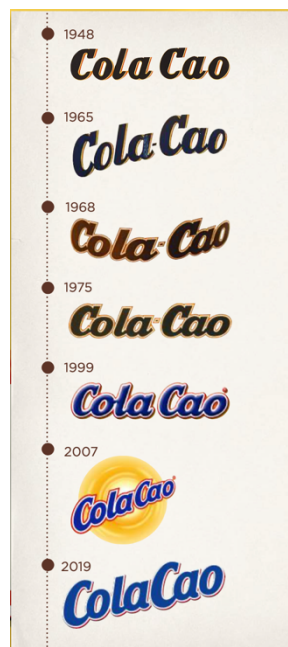


Ilustración 1: Cambio de logotipo Cola Cao (colacao.es)

3.2. ¿ERES MÁS DE COLA CAO O DE NESQUIK?

Durante los más de 75 años que ha estado actuando en el mercado español, Cola Cao se ha resistido a la competencia de distintas marcas de cacao soluble, en especial de Nesquik, que llegó a España casi una década más tarde. La empresa siempre se ha distinguido de sus competidores haciendo referencia a sus “grumitos”, que, como todo español sabe, hacen único este producto. La marca asegura que es un cacao natural sin aditivos (Invertia, 2019).

La rivalidad entre Cola Cao y Nesquik se ha convertido casi en un debate nacional, siendo muy frecuente la pregunta entre amigos o familiares de: “¿tú eres más de Cola Cao o de Nesquik? Por ello, *Datacentric* realizó un estudio para analizar cual era la marca favorita de los españoles y el mapa lo deja claro, Cola Cao es el producto elegido por casi todas las Comunidades Autónomas (OkDiario, 2022).



Ilustración 2: Preferencias de cacao en España 2022. Fuente Datacentric. (datacentric.es)

3.3. COLA CAO Y SUS ANUNCIOS. ESTRATEGIAS DE MARKETING.

La comunicación de Cola Cao, tradicionalmente se ha basado en la publicidad con anuncios como el conocido “*La canción del Cola Cao*” que la gran mayoría de los españoles sigue recordando: “*Yo soy aquel negrito del África tropical, que cultivando cantaba la canción del Cola Cao...*”. Esta canción surgió en el año 1955, siendo uno de los primeros anuncios que se dieron en la radio española, el medio de comunicación por

excelencia en aquella época. Fue la primera marca española en patrocinar una radionovela, concretamente “*Matilde, Perico y Periquín*” en Cadena SER. Tal fue el éxito, que años más tarde se emitiría también este anuncio en cine y televisión. En 1962 aparecieron los primeros *spots* televisivos y en la década de los 80 apostaron por las cadenas televisivas privadas, de las que ni siquiera se obtenían, por aquel entonces, datos de audiencia (Anagrama Comunicación, s.f.).

Desde el inicio, la marca ha vinculado el producto Cola Cao al deporte, lo cual se ha convertido a lo largo de los años en uno de sus signos identificativos. Ya lo decía su primer anuncio: “*lo toma el futbolista para entrar goles, también lo toman los buenos nadadores...*”, pero el desarrollo en este sentido ha sido muy amplio. En 1972 el Cola Cao se convirtió en alimento olímpico y en 1988 la marca se adhirió al plan ADO, convirtiéndose en patrocinador oficial del equipo olímpico español, algo que mantiene a día de hoy, casi 40 años más tarde. Los deportistas del momento se han ido sumando a sus campañas durante todo este tiempo, encontrando spots televisivos con personalidades como Mark Spitz, Emerson Fittipaldi, Rafael Nadal o Dani Pedrosa, estos dos últimos protagonizando el anuncio con sus madres (ColaCao, 2021).

Como comentábamos, las marcas se tienen que adaptar al momento en el que viven, y una de las preocupaciones esenciales de los consumidores es que se tengan en cuenta los valores sociales con los que conviven diariamente. Por ello, Cola Cao, desde el año 2009, ha empezado a dar valor a la superación, la integración y el compañerismo, centrándose más en los niños. La marca hizo múltiples anuncios contando la historia de distintos niños con el deporte, muchos de los cuales a día de hoy han triunfado en ese deporte. Han creado también el programa “*Becas Vamos*” para fomentar el deporte base y cumplir los sueños de muchos niños deportistas en España, junto con el Consejo Superior de Deportes (ColaCao, 2021).

Pero el deporte no es lo único que la marca fomenta, si no que ha incrementado su responsabilidad social corporativa, elemento clave para situar a la marca en una posición estratégica frente a sus consumidores. La Fundación Cola Cao es una entidad sin ánimo de lucro cuya misión es hacer realidad el propósito social de la marca, a través de distintos proyectos de acción social. Cola Cao siempre ha estado especialmente preocupada por el desarrollo de los niños y su protección. Uno de los mayores problemas que afecta la diversidad infantil es el *bullying* o acoso escolar, por lo que, desde la marca, tratan de

poner fin a ello. Esto con lo consiguen con tres elementos principales (ColaCao Acción Social, 2021):

- **Divulgación:** el objetivo es que se hable más de acoso escolar, romper el silencio que hay entorno a esta problemática actual y sensibilizar a la sociedad a través de distintos proyectos y campañas divulgativas.
- **Investigación:** elaborar estudios de investigación que aporten datos sobre el abuso escolar y sus causas en España.
- **Educación:** a través del programa Somos Únicos, Cola Cao educa a los niños sobre el acoso desde la prevención, con la iniciativa principal de “empoderar a los niños a descubrir y potenciar sus singularidades y a luchar contra el acoso escolar” (ColaCao Acción Social, 2021).

3.4. COLA CAO ¿LOVE BRAND?

Cola Cao ha estrenado recientemente *claim* publicitario a través de la frase “eso tan tuyo”, que busca transmitir los valores que han unido a la marca con sus consumidores durante más de 75 años. Un *claim* es una frase pegadiza vinculada a una campaña de publicidad que busca crear emociones y despertar la memoria del consumidor. Cola Cao sabe lo que hace, sabe que tiene una enorme vinculación con sus consumidores y quiere que estos se acuerden de ello.

Si bien es cierto que la marca últimamente está muy centrada en cuestiones relacionadas con los pequeños de la casa, su vinculación va mucho más allá, ya que su relación con el cliente viene de hace muchos años y todos los adultos han sido niños. Según información proporcionada por la empresa, Cola Cao es la marca más consumida de cacao soluble en España desde su creación. Es un producto increíblemente popular, tanto para niños como para mayores, ya que casi el 50% del consumo del producto está en manos de los adultos.

Es una marca que, indudablemente, va creando *engagement* con sus clientes con el paso de los años. Con respecto a los más pequeños de la casa resulta complicado distinguir el nivel de vinculación que se tiene con la marca. Muchos niños toman Cola Cao porque es lo que se compra en su casa y porque les gusta el sabor a chocolate y no han tenido opción de probar otro cacao en polvo. La cuestión más relevante es como un producto que, generalmente, comienza siendo para niños, consigue mantener esta vinculación cuando estos se hacen mayores.

Las generaciones de adultos de hoy en día tienen una conexión y una vinculación muy fuerte con la marca. Esto se debe a que, cuando ellos eran niños, este era el producto estrella en sus desayunos y meriendas, por lo que el Cola Cao y sus grumitos les recuerda a su infancia. Este es el motivo principal de que muchos elijan comprarles a sus hijos lo mismo que tomaron ellos, ya que tienen una fuerte vinculación con lo que el Cola Cao significa para ellos. Incluso actualmente se están elaborando productos para niños de apenas unos meses que contienen Cola Cao, como Blevit Cola Cao, una papilla para que los más pequeños de la casa puedan empezar a probar este producto.

Al respecto de esto cabe mencionar el eterno debate entre Cola Cao o Nesquik y el video que se publicó en 2016 en El País preguntando a los cocineros más importantes de España bajo el título: *“Los chefs responden: ¿Cola Cao O Nesquik? Los cocineros más ilustres de España contestan a una de las preguntas cruciales de la historia de la humanidad”* (Iturriaga, 2016). Es una pregunta que, en realidad, es una tontería, consiste en elegir entre una marca y la otra, pero el propio título de la noticia deja ver que esto es independiente de la elección, sino que son los consumidores los que buscan defender a capa y espada su producto favorito. Es decir, Cola Cao es una auténtica *love brand*, defendida por sus consumidores, quienes guardan una relación especial con ella.

Es indudable que el resultado general fue Cola Cao (15-5), pero lo relevante son las respuestas que dieron y algunas frases que podemos extraer al respecto de la marca. Entre ellas: *“El Cola Cao es de toda la vida”*, *“Cola Cao es mi infancia”*, *“Cola Cao por el negrito”*, *“Continúo tomando Cola Cao”*, *“Yo he tomado Cola Cao desde niño”* o *“Cola Cao es de esas cosas que todavía están ahí en la memoria”*. Son muchos los que también destacan sus grumitos y la posibilidad de tomárselo en cualquier época del año.

La cuestión a analizar no es si prefieren el Cola Cao al Nesquik, si no el porqué de tomar Cola Cao. Como podemos ver, la gran mayoría lo vinculan a su infancia, a “toda su vida” e incluso hay quien dice que es de esas cosas que tiene guardada en la memoria. Porque Cola Cao para mucha gente en España es eso, es un recuerdo de cuando eran niños, y es muy complicado desvincularse de algo a lo que llevas siendo tan fiel tanto tiempo.

Otra cuestión importante es lo que ha conseguido Cola Cao con sus anuncios. Raro es el adulto al que le dices “yo soy aquel negrito” y no continúa con la canción. Incluso niños más jóvenes conocen esta canción porque sus padres se la han cantado desde que son muy

pequeños cada vez que escuchan la palabra Cola Cao. Ya lo vemos reflejado en el video, cuando uno de los cocineros más importantes de España dice “Cola Cao por el negrito”. Por supuesto que no es por eso, pero la vinculación que muchos han creado con la canción, su infancia y el producto, es lo que les hace sentirse conectados emocionalmente a ello.

4. ANÁLISIS EMPÍRICO

Una vez estudiada la posición de Cola Cao en el mercado, se procede a realizar un análisis de la marca en las redes sociales, a través de la plataforma Twitter. Con los comentarios realizados por los usuarios, el objetivo es conocer cuáles son las sensaciones y emociones que estos tienen hacia la marca y, posteriormente, comprobar si estas publicaciones coinciden con lo que Cola Cao trata de transmitir.

Twitter es una plataforma que en la actualidad tiene millones de usuarios realizando comentarios diariamente sobre temas diversos. Es una fuente de contenido inmensa y, por ello, es una de las principales redes utilizadas para llevar a cabo estudios de *Social Media Analytics*. Los beneficios que estos estudios tienen para las marcas son numerosos, ya que permite extraer opiniones de los usuarios, puede ayudar a predecir tendencias, analizar si los anuncios dan resultados o conocer mejor a la audiencia. El objetivo en este caso va a ser el de analizar las opciones y comprobar el sentimiento de los consumidores hacia un producto o una empresa.

4.1. *TEXT MINING*

4.1.1. ¿Qué es el *text mining*?

Es importante tener en cuenta es que el texto es una información no estructurada, lo cual no significa que no tenga estructura, si no que tiene una estructura lingüística entendible para los humanos, pero no para los ordenadores. Por ello, para poder tratar los tweets, estos requieren un preprocesamiento para que los datos puedan ser utilizados por algoritmos de *Machine Learning*. Esto se debe a que el orden de las palabras importa, hay fallos gramaticales (se usan abreviaciones, sinónimos, palabras homógrafas...) y el contexto de la palabra es muy importante, pero es difícil tenerlo en cuenta.

Para poder analizarlo, el *text mining* da una estructura al texto, dividiéndolo en un corpus, que contendrá una serie de documentos, y estos documentos estarán divididos en *tokens*, que son las variables.

Hay diversos métodos para construir estas variables y el enfoque que he seleccionado yo es el llamado *Bag of Words*. Con esto, el documento se convierte en una bolsa de palabras que no tiene en consideración el orden de las mismas. Cada tweet va a tratarse como un

conjunto de palabras individuales, ignorando la gramática y si es signo de puntuación o no. Se va a tratar cada palabra como una variable potencial.

Dentro de esta opción hay dos posibilidades a la hora de determinar el valor que toma cada variable potencial dentro del documento. Hay que darle siempre un valor numérico para que la herramienta de *Machine Learning* pueda interpretarlo. La primera solución es optar por realizar una *Binary Representation*, dándole a la palabra un valor de 1 si aparece en el documento y un 0 si no aparece. La otra opción es hacer una *Term Frequency Representation*, la cual consiste en que esa variable tome como valor la frecuencia con la que la palabra aparece en el documento, es decir, si aparece 3 veces tomará un valor de 3. He optado por realizar la segunda opción, ya que el número de veces que aparece una palabra en un documento puede ser significativo, y se estaría perdiendo información en el caso de optar por la primera.

Para poder realizar este estudio he elegido la herramienta R Studio, un lenguaje de programación que da infinidad de posibilidades para investigar esta cuestión.

4.1.2. Descarga de tweets

Lo primero que hay que hacer es extraer los datos de Twitter, para lo cual hay que instalar una serie de paquetes y abrir las librerías. La única que se empleará en este momento es la librería *Rtweet*, mientras que el resto servirán para el resto del estudio.

```
install.packages("rtweet")
install.packages("syuzhet")
install.packages("rvest")
install.packages("ggplot2")
install.packages("quanteda")
install.packages("tidytext")
install.packages("lubridate")
install.packages("quanteda.textstats")
install.packages("wordcloud")
install.packages("udpipe")
install.packages("twitter")
install.packages("readexcel")
install.packages("readxl")
```

Ilustración 3: Instalación de paquetes.

```
library(rtweet)
library(syuzhet)
library(rvest)
library(xml2)
library(ggplot2)
library(dplyr)
library(stats)
library(base)
library(quanteda)
library(scales)
library(tidytext)
library(lubridate)
library(RColorBrewer)
library(stringr)
library(quanteda.textstats)
library(wordcloud)
library(utf8)
library(udpipe)
library(tidyr)
library(twitter)
library(readxl)
```

Ilustración 4: Carga de librerías.

Una vez están instaladas, se emplea el código para fijar el directorio de trabajo y se descargan los tweets. Previamente hay que solicitar una cuenta de *Twitter Developer*, bastando con introducir usuario y contraseña. Los *tweets* descargados se encuentran en un Excel, primero hay que abrirlo y posteriormente analizar su estructura. Estos se guardarán en una variable a la que he llamado *TweetsColaCao*.

```
setwd("/Users/martanoguera/Desktop/TFG Analytics")
TweetsColaCao<-search_tweets("cola cao", n=2000, include_rts=FALSE, lang="es")

Tweets <-read_excel("Colacao.xlsx")
summary(Tweets) #confirm data is correct
View(Tweets)
```

Ilustración 5: Descarga de tweets.

4.1.3. Limpieza de datos

La estructura de los tweets refleja que el número de variables (columnas) obtenidas es muy amplio y no todas ellas contienen datos relevantes, muchas ni siquiera tienen datos. He obtenido 639 tweets y 43 variables, pero decido quedarme únicamente con cuatro variables principales, que serán las más indicadas para realizar el estudio. Estas son: *id*, *source*, *text* y *created_at*.

- *ID*: número de identificación del usuario de twitter
- *SOURCE*: fuente desde la que se publica (iPhone, Android, Web App)
- *TEXT*: comentario publicado por el usuario de twitter
- *CREATED_AT*: fecha y hora de publicación del contenido

```
TweetsColaCao<-select(Tweets, id, created_at, text, source)
View(TweetsColaCao)
```

Ilustración 6: Filtración de variables principales.

Luego, por comodidad a la hora de tratar los datos, opto por simplificar el texto recogido en la variable *source*, para que quede dividido en los tweets “for Android”, “for iPhone” y “Web App”. El número de tweets baja a 624, una pérdida poco relevante, que se debe a que algunos tweets no venían de ninguna de las tres opciones menciones que son usadas generalmente. Por curiosidad, compruebo cuál es la fuente más utilizada entre los

usuarios al escribir tweets. La más frecuente es Android, con 341, seguida de iPhone con 194 y, por último, la aplicación web, con 89.

```
> TweetsColaCao <- TweetsColaCao %>%
+ select(id, source, text, created_at) %>%
+ extract(source, "source", "Twitter (.*)<") %>%
+ filter(source %in% c("for iPhone", "for Android", "Web App"))
> summary(TweetsColaCao)
      id          source          text          created_at
Min.   :1.629e+18 Length:624      Length:624      Min.   :2023-02-25 10:40:53.00
1st Qu.:1.631e+18 Class :character Class :character 1st Qu.:2023-02-28 20:29:33.00
Median :1.631e+18 Mode  :character Mode  :character Median :2023-03-02 22:16:47.00
Mean   :1.631e+18                               Mean   :2023-03-02 17:42:02.49
3rd Qu.:1.632e+18                               3rd Qu.:2023-03-04 12:07:00.75
Max.   :1.633e+18                               Max.   :2023-03-07 08:54:36.00
> table(TweetsColaCao$source)

for Android  for iPhone  Web App
         341         194         89
```

Ilustración 7: Fuentes de los tweets.

Con estos datos, hay que crear un *corpus* para posteriormente crear una *Document Term Matrix* (en adelante “DTM”), explicado más adelante. Sin embargo, uno de los problemas esenciales que tiene el analizar textos es que hay muchas palabras que no aportan ningún valor al estudio, como pueden ser los signos de puntuación, las menciones de otros usuarios y las páginas web. Esto es necesario eliminarlo para poder percibir de forma más clara cual es el sentimiento real de los consumidores hacia la marca. Debe realizarse antes de todo lo demás porque si no intercede con la creación del corpus y no llegan a eliminarse del todo. De todas formas, lo he creado previamente para comprobar que estos símbolos eran, efectivamente, muy frecuentes, pudiendo ver paréntesis, puntos, comas, menciones y páginas web.

```
> dim(midfm) #He detectado 3372 términos/features
[1] 624 3372
> featnames(midfm) #Observamos cuales son esos términos
 [1] "https://t.co/k8liocufew" "ahorra"      "con"         "cola"       "cao"
 [6] ". "                       "explora"    "todas"      "las"        "ofertas"
[11] "y"                        "promociones" "actuales"   "de"         "para"
[16] "comprar"                 "descuento"  "en"         "su"         "tienda"
[21] "física"                  "o"          "online"     "#saludosyfelizdiavss" "👋"
[26] "👉"                       "https://t.co/s9grxfqm&k" "buenos"     "días"       "!"
[31] "vamos"                   "el"         "martesitos" "cafetito"   "al"
[36] "lío"                      "C"         "también"   "es"         "válido"
[41] ")"                        "👉"        "👉"         "👉"         "👉"
[46] "@massa_43"              "pues"      "matias"    "te"         "va"
```

Ilustración 8: Términos de la matriz. Comprobación.

Una vez realizado este paso previo, pasamos a crear un *corpus* con los comentarios obtenidos, al que denomino *micorpus*, y construyo la DTM. Esta matriz queda organizada

de tal forma que cada fila hace referencia a un documento (tweet) y cada columna a las variables elegidas anteriormente, que son las que caracterizan el registro. Para crear la DTM utilizo la función *dfm*, que, por defecto, convierte todas las palabras a minúsculas, lo cual es muy útil y ayuda a reconocer las palabras como una sola, en vez de como dos palabras distintas, aunque sean iguales y pretendan decir lo mismo.

```
> micorpus<-corpus(TweetsColaCao)
> summary(mi corpus)
Corpus consisting of 624 documents, showing 100 documents:
```

	Text	Types	Tokens	Sentences	id	source	created_at
	text1	22	26	1	1.633028e+18	Web App	2023-03-07 08:54:36
	text2	15	15	1	1.633009e+18	for Android	2023-03-07 07:36:11
	text3	15	16	1	1.633000e+18	Web App	2023-03-07 07:02:20
	text4	9	9	1	1.632996e+18	for iPhone	2023-03-07 06:44:18
	text5	4	4	1	1.632995e+18	for iPhone	2023-03-07 06:42:42

Ilustración 9: Creación del corpus.

```
> midfm<- dfm(mi corpus) #Construccion document-term matrix
> dim(midfm) #He detectado 2549 términos/features
[1] 624 2549
> featnames(midfm) #Observamos cuales son esos términos
 [1] "ahorra"      "con"         "cola"        "cao"         "explora"     "todas"       "las"         "ofertas"
 [9] "y"           "promociones" "actuales"    "de"          "para"        "comprar"     "descuento"  "en"
[17] "su"          "tienda"      "fisica"      "o"           "online"      "saludosyfelizdiavss" "buenos"     "días"
[25] "vamos"      "el"          "martesitos" "cafetito"    "al"          "lío"         "también"    "es"
[33] "válido"     "pues"        "matias"      "te"          "va"          "a"           "poner"      "culo"
[41] "rojío"      "tomate"      "q"           "se"          "enfria"      "un"          "ahora"      "no"
[49] "me"         "vendría"     "nada"        "mal"         "él"          "la"          "riquísima"  "dios"
```

Ilustración 10: Creación de la Document Term Matrix (DTM) y visualización de sus términos.

Detecto que hay 2549 términos (palabras) en estos 624 documentos, dejando una matriz de dimensiones 624x2549. Dentro de estas palabras podemos encontrar muchas que podrían guardar relación con los sentimientos de los consumidores, como “riquísima”, “ofertas”, o “válido”.

Es importante también analizar cuáles son las palabras que más utilizan los usuarios en sus tweets sobre Cola Cao y, para ello, extraigo las 50 palabras más repetidas.

```
> topfeatures(midfm,50) #Primera visualización de las 50 palabras que mas se repiten
 cola    cao    de    el    y    que    con    la    en    a    me    un    no    para    es
676    630    401    338    317    285    239    216    206    196    185    147    144    103    99
 te    le    leche    las    por    o    una    lo    del    café    mi    agua    al    se    droja
77    76    76    71    69    66    66    66    64    59    57    57    56    56    52
 yo    si    los    to    pero    drinks    coma    ya    bebidas    know    sin    coca    conocerme    más    nesquik
49    47    45    44    43    42    40    40    40    40    38    36    35    31    30
 echaron    su    mejor    cerveza    he
30    27    27    27    26
```

Ilustración 11: Palabras más repetidas en los documentos.

Podemos ver que, aparte de “cola” y “cao” y alguna más, el resto son palabras comunes del lenguaje español, conocidas generalmente como *stopwords* (“y”, “el”, “de”, “en”, “que” ...). Estas son aquellas palabras que suelen utilizarse como conectores pero que no aportan ningún valor significativo a la hora de evaluar el mensaje que se quiere transmitir en el texto. El propio programa de R recoge una lista predefinida de *stopwords* para facilitar su eliminación en cada idioma.

Como las *stopwords* no aportan información útil hay que eliminarlas. Esto luego permite crear otra DTM con las palabras reales de los usuarios que buscan aportar información y sentimiento. He optado por eliminar también *stopwords* en inglés, porque he visto que hay alguna palabra en inglés entre las más repetidas. “Drinks” o “know” son palabras muy repetidas, así como “to”, la cual tendría la consideración de *stopword* en inglés. Vemos que se han eliminado bastantes términos, y la DTM ahora tiene una dimensión de 624x2282.

```
> micorpus <- tokens(micorpus, remove_punct = TRUE) %>%
+   tokens_remove(stopwords("es")) %>%
+   tokens_wordstem()
>
> micorpus <- tokens(micorpus, remove_punct = TRUE) %>%
+   tokens_remove(stopwords("en")) %>%
+   tokens_wordstem()
>
> midfm <- dfm(micorpus)
> dim(midfm)
[1] 624 2282
```

Ilustración 12: Eliminación de stopwords.

Una vez hecha esta limpieza de datos previa, vemos que la dimensión de la matriz ha bajado bastante con respecto al inicio. Antes de la eliminación de *stopwords* teníamos alrededor de 2500, pero en la prueba inicial para comprobar los signos de puntuación y menciones el número de palabras alcanzaba 3372. Ya parece adecuado proceder al análisis de sentimiento de los consumidores hacia la marca, con las palabras que verdaderamente son relevantes. Comprobamos otra vez que los datos están completamente limpios, reflejando las palabras de la matriz y las más utilizadas.

```
> topfeatures(midf,30)
cola    cao    lech    café    agua    droja    si    bebida    drink    know    coca    conocerm    nesquik    echaron    cerveza
676    630    77    59    57    52    47    46    42    40    36    35    30    30    28
mejor    zumo    naranja    q    azúcar    toda    bueno    hace    madr    calentito    zero    mañana    chocol    té    tomar
27    27    26    24    22    21    21    21    21    21    19    18    17    17    16
```

Ilustración 13: Palabras DTM sin stopwords, signos de puntuación, menciones o páginas web.

```
> featnames(midf)
[1] "ahorra"      "cola"      "cao"      "explora"   "toda"      "oferta"
[7] "promocion"  "actual"    "comprar"  "descuento" "tienda"    "fisica"
[13] "onlin"      "saludosyfelizdiavss" "bueno"    "dias"      "vamo"      "martesito"
[19] "cafetito"   "lío"      "válido"   "pue"      "matia"     "va"
[25] "poner"      "culo"     "rojoy"    "tomat"    "q"        "enfria"
[31] "ahora"      "vendria"  "mal"      "riquisima" "dio"       "nesquik"
[37] "lado"       "valenciana" "rebujada" "hacer"    "cachimba" "cacahuet"
[43] "si"         "ruibal"   "queja"    "hace"     "prensa"   "mismo"
```

Ilustración 14: Palabras más utilizadas sin stopwords, signos de puntuación, menciones o páginas web.

Otra cuestión importante que se puede realizar en estos casos de análisis de textos es el *stemming* o la lematización, procesos que buscan reducir la palabra a su lexema para que esta no tenga la consideración de dos palabras distintas en el análisis. Es un poco similar a lo que ocurre con las mayúsculas y las minúsculas, si una misma palabra se repite muchas veces, pero en plural y en singular o en femenino y en masculino, debería computar como la misma palabra a la hora de evaluar su frecuencia. Estos procesos, aunque parecidos, tienen algunas diferencias.

El *stemming* es un método que se utiliza para reducir una palabra a su raíz eliminando los sufijos, pero puede cometer dos tipos de errores. Este proceso puede causar *overstemming* o *understemming*. El primero consiste en acortar tanto una palabra que los resultados carezcan de sentido o se acaben reduciendo a una misma raíz palabras que son semánticamente diferentes. Esto puede ocurrir, por ejemplo, al dejar universo y universidad como “univers”. El *understemming* es el proceso contrario, que puede causar que dos palabras semánticamente iguales se reflejen individualmente, algo que en el idioma es español es bastante frecuente que ocurra. Esto ocurriría, por ejemplo con la palabra “absorber” y “absorción”, reconociendo la primera con el lexema “absorb” y la segunda con el lexema “absorc”.

La lematización es un proceso un poco más avanzado y que se da en herramientas más avanzadas. Consistiría en reducir la palabra a su lexema y no a la raíz, lo que supondría un análisis más profundo del texto. Esto podría pasar en el caso del verbo ir y sus conjugaciones, identificando las palabras “ir”, “voy” y “yendo” como un mismo verbo “ir”.

He optado por realizar un proceso de *stemming* en la variable texto de la base de datos *TweetsColaCao*. Sin embargo, para el posterior análisis de las palabras más frecuentes, he eliminado el *stemming* del texto para poder percibir de forma completa cuales son estas palabras más repetidas, que, como veremos, no comparten lexema.

```
> midfm <- dfm_wordstem(midfm, language = "spa")
> dim(midfm)
[1] 624 1886
> featnames(midfm)
 [1] "ahorr"          "col"           "cao"           "explor"        "tod"           "ofert"
 [7] "promocion"     "actual"        "compr"         "descuent"      "tiend"         "fisic"
[13] "onlin"         "saludosyfelizdiavss" "buen"          "dias"          "vam"           "martesit"
[19] "cafetit"       "lio"           "val"           "pue"           "mat"           "va"
[25] "pon"           "cul"           "rojoy"         "tomat"         "q"             "enfri"
[31] "ahor"          "vendr"         "mal"           "riquisim"      "dio"           "nesquik"
[37] "lad"           "valencian"     "rebuju"        "hac"           "cachimb"       "cacahuet"
[43] "si"            "ruibal"        "quej"          "prens"         "mism"          "ment"
[49] "entiend"       "preguntal"    "mam"           "ciert"         "azuc"          "jod"
```

Ilustración 15: Stemming.

4.1.4. Palabras más repetidas. Visualización.

Para visualizar gráficamente estas palabras obtenidas, he optado por realizar dos tipos de gráficos visuales. El primero consiste es un gráfico de dispersión con las palabras y su frecuencia de uso en estos documentos.

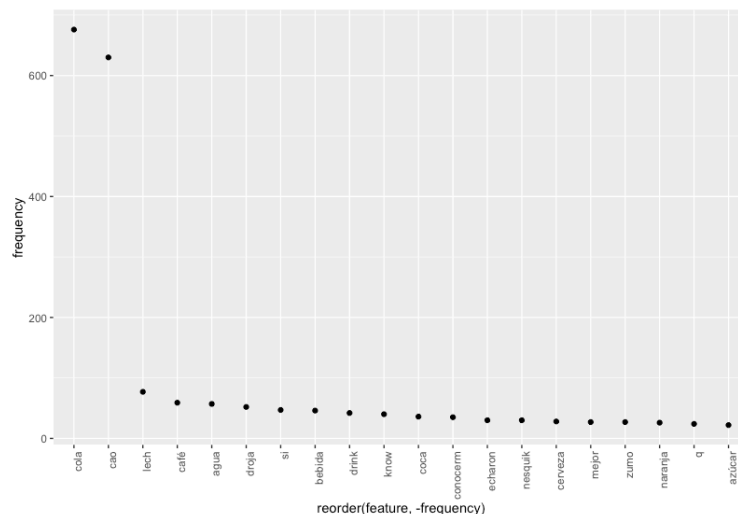


Ilustración 16: Gráfico de dispersión de las palabras más utilizadas y su frecuencia.

```
> #Gráfico de los términos mas frecuentes
> features_freq<-textstat_frequency(midfm, n = 20)
> View(features_freq)
> ggplot(features_freq, aes(x = reorder(feature,-frequency), y = frequency)) +
+   geom_point() +
+   theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Ilustración 17: Código gráfico de dispersión.

Una de las gráficas más visuales que proporciona R al analizar textos es la nube de palabras. Una nube de palabras es una representación gráfica de las palabras clave, que suele utilizarse para representar las etiquetas de un sitio web para facilitar al usuario el lector del documento. Estas palabras aparecerán de distintos colores, en ocasiones formando una forma o imagen. Por lo general, las etiquetas de mayor tamaño y con colores más fuertes son las que aparecen con mayor frecuencia, mientras que aquellas que no se repiten tanto aparecerán más tenues y con un menor tamaño.



Ilustración 18: Nube de palabras.

```
> set.seed(123)
> midfm_t <- t(midfm)
> m <- as.matrix(midfm_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)
```

Ilustración 19: Código nube de palabras.

4.2. SENTIMENT ANALYSIS O ANÁLISIS DE SENTIMIENTO

4.2.1. ¿Qué es un análisis de sentimiento?

Cuando las personas leemos un texto, sabemos captar la intención con la que este ha sido redactado, percibiendo si tiene una carga positiva o negativa. Percibimos sentimientos de los textos, ya sea felicidad, tristeza, euforia, angustia, enfado... Las herramientas de *text mining* dan la posibilidad de intentar capturar automáticamente el contenido emocional de los documentos. Ya sabemos cuáles son las palabras más repetidas en los tweets de Cola Cao y hemos limpiado los datos para que sean las palabras que dan información

relevante, pero ¿cuál es el sentimiento que dan estos tweets? ¿son positivos o negativos hacia la marca?

4.2.2. Métodos

Para poder realizar este análisis de sentimiento encontramos dos enfoques principales. El primero consiste en crear un algoritmo entrenado con varios documentos y sus sensaciones, para poder predecir el sentimiento de un documento nuevo. El segundo es el método que vamos a utilizar, que se apoya en un lexicón de sentimiento. Este es un diccionario que recoge una lista de palabras y el sentimiento que recogen, por lo que, si se encuentra una de esas palabras, va a devolver el sentimiento que tiene guardado para la misma. El lexicón de sentimiento está supeditado a las palabras que recoge el diccionario y, además, no es capaz de captar la negación de un sentimiento. Si encontramos redactado “no estoy triste”, en vez de captarlo como sentimiento positivo, lo captará como negativo al encontrar la palabra “triste”. Sin embargo, el beneficio de este método es que el lenguaje de R tiene integrado este diccionario y resulta más sencillo realizarlo al no exigir un proceso de prueba y error y de entrenamiento del algoritmo.

El programa R contiene tres diccionarios principales para poder realizar este estudio:

- **AFINN**: asignará a cada palabra un valor entre -5 y 5 según el grado que este tenga de negativo o positivo respectivamente.
- **BING**: devuelve un valor de -1, 0 o 1 según el sentimiento positivo o negativo.
- **NRC**: este método no reconoce únicamente sentimientos, si no también emociones. Se basa en las 8 emociones básicas de Plutchik, e incluye cada palabra en alguna de las siguientes categorías, pudiendo ser varias: anger, anticipation, disgust, fear, joy, sadness, surprise y trust. Una vez lo tenga, asignará una polaridad positiva o negativa según las emociones que estén registradas en el diccionario.

4.2.3. Método NRC y las ocho emociones de Robert Plutchik

El diccionario NRC en *Data Analytics* es uno de los que más información puede proporcionar acerca de los sentimientos que se perciben de un documento, y es el que voy a utilizar para analizar la vinculación de los usuarios con la marca Cola Cao. Se diferencia

de los otros dos métodos por su capacidad para tener en consideración las distintas emociones del ser humano, en vez de calificarlo únicamente como positivo o negativo. Parte de la premisa de que no todo es únicamente positivo o negativo, sino que lo frecuente es que las emociones no vengan solas y puedan matizarse dentro de un mismo sentimiento.

Por ello, me parece necesario realizar un breve resumen del concepto de las emociones de Robert Plutchik, ya que es lo que NRC utiliza en su metodología para clasificar como positiva o negativa una palabra.

Plutchik desarrollo su idea entorno a lo que él llamó “Rueda de las emociones”, la cual estaba compuesta por ocho emociones básicas fundamentadas en experiencias relativamente comunes. Estas eran: la alegría, la confianza, el miedo, la sorpresa, la tristeza, la aversión, la ira y la anticipación. El autor argumentaba que estas dimensiones rara vez ocurren solas y que se podían presentar conforme a distintos grados de intensidad. Por ello, una persona suele tener más de una emoción en cada momento, independientemente de que alguna de ellas predomine sobre las otras, ya que todas guardan relación entre sí mismas. Esto es lo que daba la posibilidad de que se produjeran solapamientos entre las emociones, desembocando en un sentimiento más complejo, y Plutchik denominó a estos solapamientos “díadas”. Por ello, en su rueda, encontramos emociones primarias, secundarias o terciarias, según la frecuencia, la intensidad o la combinación que se pudiera producir entre ellas (Mateu-Mollá, 2019).

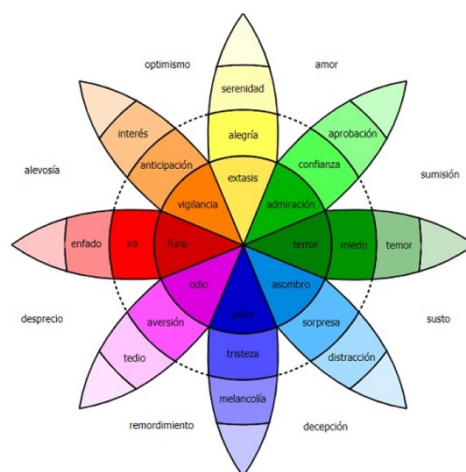


Ilustración 20: Rueda de las emociones de Robert Plutchik.


```
TweetsColaCao<-TweetsColaCao %>% mutate(nrc_polarity= get_sentiment(text, method="nrc"),
bing_polarity=get_sentiment(text, method="bing"))
```

Ilustración 22: Columnas polaridad en el dataframe.

```
> TweetsColaCao %>% summarise(positivos = 100 * sum(nrc_polarity > 0) / n(),
+                               neutros = 100 * sum(nrc_polarity == 0) / n(),
+                               negativos = 100 * sum(nrc_polarity < 0) / n())
# A tibble: 1 × 3
  positivos neutros negativos
  <dbl>    <dbl>    <dbl>
1     7.21    85.3     7.53
```

Ilustración 23: Porcentaje de sentimiento de TweetsColaCao.

He optado por analizar también este sentimiento en función de la procedencia de los *tweets*. Esta procedencia está recogida en la variable *source*, y dividida en “*for iPhone*”, “*for Twitter*” y “*Web App*”. Aquí podemos ver que el resultado cambia un poco según la fuente de emisión del comentario. En los *tweets* emitidos directamente desde un ordenador (*Twitter Web App*), la neutralidad disminuye, y tanto el sentimiento positivo como el negativo se incrementan. En los *tweets* emitidos desde un *smartphone*, la neutralidad sigue bastante parecida al resultado genérico, mientras que el sentimiento positivo en usuarios de *Android* se eleva y en los *iPhone* disminuye.

```
> TweetsColaCao %>% group_by(source) %>%
+   summarise(positivos = 100 * sum(nrc_polarity > 0) / n(),
+             neutros = 100 * sum(nrc_polarity == 0) / n(),
+             negativos = 100 * sum(nrc_polarity < 0) / n())
# A tibble: 3 × 4
  source      positivos neutros negativos
  <chr>         <dbl>    <dbl>    <dbl>
1 Web App      10.1     75.3     14.6
2 for Android   7.62     86.5     5.87
3 for iPhone    5.15     87.6     7.22
```

Ilustración 24: Resultado porcentual del sentimiento positivo, negativo y neutro según fuente de emisión del tweet.

Como ya he mencionado previamente, para poder analizar en profundidad cuál es el sentimiento de los consumidores de Cola Cao hacia la marca, el método que más información puede proporcionar es el método NRC, ya que tiene en consideración las 8 emociones de Plutchik.

```

> TweetsColaCao<-TweetsColaCao %>% mutate (anger=emociones$anger,
+ anticipation=emociones$anger,
+ disgust=emociones$anger,
+ fear=emociones$fear,
+ joy=emociones$joy,
+ sadness=emociones$sadness,
+ surprise=emociones$surprise,
+ trust=emociones$trust,
+ negative=emociones$negative,
+ positive=emociones$positive)

```

Ilustración 25: Aplicación del método NRC al dataframe TweetsColaCao.

El método NRC funciona de tal forma que, según el número de palabras y las sensaciones que estas transmiten en cada documento, da unos valores a las 8 emociones. Estos resultados posteriormente los utiliza en las dos últimas columnas para reflejar si este tiene un sentimiento general positivo o negativo. Como vemos en la siguiente ilustración, un mismo tweet puede dar muchas emociones a la vez y con distintos grados de intensidad. Podemos ver como a través de un mismo tweet se pueden incluso llegar a transmitir las ocho emociones conjuntamente: ira (1), anticipación (1), aversión (1), miedo (1), alegría (4), tristeza (1), sorpresa (1) y confianza (7). En este caso el sentimiento general que percibe el programa es de una negatividad de 3 y una positividad de 7, ya que como vemos tiene mucha alegría y confianza, mientras que refleja valores más pequeños en emociones como la aversión, el miedo o la tristeza.

text	created_at	nrc_polarity	bing_polarity	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
Buenos días Abel Menudo desayuno chin pum te ha...	2023-02-28 11:57:03	0	0	0	0	0	0	5	0	1	3	0	8
Oh que prodigio Que soltura Vivo solo el cola cao m...	2023-03-02 12:17:07	0	0	1	1	1	1	4	1	1	7	3	7
De seguro te dan una ayuda estatal o algo así con mi...	2023-03-05 19:36:33	0	0	0	0	0	0	2	0	0	3	0	6
Bueno igual es la excusa necesaria para tomarte un c...	2023-03-01 16:08:48	0	0	1	1	1	0	3	0	1	4	1	6
sive como desayuno o merienda y esta bastante bue...	2023-02-28 12:30:22	0	0	0	0	0	0	1	0	1	2	0	5
Te deseo suerte en la terapia de seguro te va bien Co...	2023-03-05 19:29:19	0	0	0	0	0	0	3	0	1	3	0	4

Ilustración 26: Explicación método NRC.

Al realizar un sumatorio de los datos de nuestro *dataframe*, podemos ver como, por lo general, la media del sentimiento positivo (0,6474) es mayor que el sentimiento negativo (0,5288). Ocurre también en el caso de las emociones que podemos percibir como “positivas” frente a las “negativas”, peor no únicamente en la media, si no también en los máximos y los mínimos. Estos máximos reflejan el número más alto de esa emoción que se da en el conjunto de los comentarios. Mientras que confianza, tristeza y alegría llegan a 7, 7 y 5 respectivamente, el número máximo de enfado, miedo o aversión es de 3.

```

> summary(TweetsColaCao)
  id                source      text      created_at      nrc_polarity      bing_polarity      anger      anticipation
Min. :1.629e+18   Length:624      Length:624      Min. :2023-02-25 10:40:53.00   Min. :-2.000000   Min. :-1.000000   Min. :0.0000   Min. :0.0000
1st Qu.:1.631e+18   Class :character   Class :character   1st Qu.:2023-02-28 20:29:33.00   1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.:0.0000   1st Qu.:0.0000
Median :1.631e+18   Mode  :character   Mode  :character   Median :2023-03-02 22:16:47.00   Median : 0.000000   Median : 0.000000   Median :0.0000   Median :0.0000
Mean   :1.631e+18                                     Mean   :2023-03-02 17:42:02.49   Mean   : 0.003205   Mean   :-0.06891   Mean   :0.1522   Mean   :0.1522
3rd Qu.:1.632e+18                                     3rd Qu.:2023-03-04 12:07:00.75   3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :1.633e+18                                     Max.   :2023-03-07 08:54:36.00   Max.   : 2.000000   Max.   : 1.000000   Max.   :3.0000   Max.   :3.0000

  disgust      fear      joy      sadness      surprise      trust      negative      positive
Min. :0.0000   Min. :0.0000   Min. :0.0000   Min. :0.0000   Min. :0.0000   Min. :0.0000   Min. :0.0000   Min. :0.0000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
Mean   :0.1522   Mean   :0.1683   Mean   :0.2965   Mean   :0.274   Mean   :0.1026   Mean   :0.3413   Mean   :0.5288   Mean   :0.6474
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
Max.   :3.0000   Max.   :3.0000   Max.   :5.0000   Max.   :7.0000   Max.   :3.0000   Max.   :7.0000   Max.   :8.0000   Max.   :8.0000

```

Ilustración 27: Media de las emociones, máximos y mínimos.

Para comprobar definitivamente, de forma más visual, cual es el sentimiento de los usuarios de Twitter hacia la marca Cola Cao, he optado por realizar un gráfico de barras. Aquí se puede ver claramente que el sentimiento positivo es mayor que el negativo y que las emociones que predominan son la confianza y la alegría, seguidas por la tristeza y el miedo.

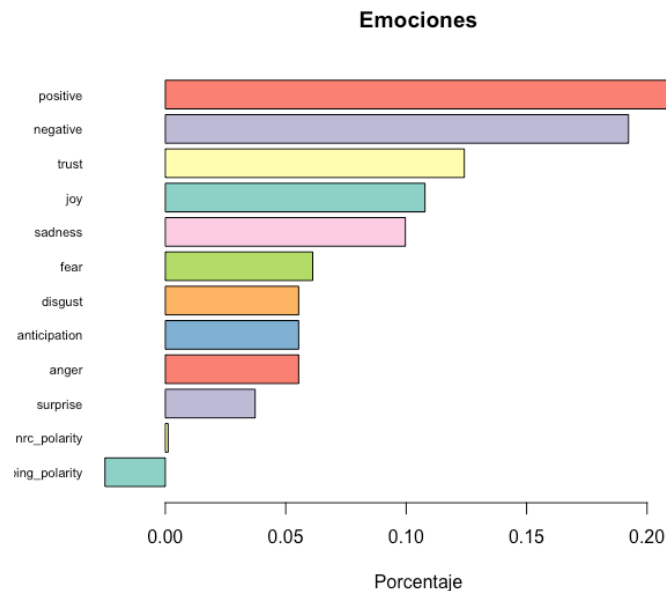


Ilustración 28: Gráfico de barras emociones.

El análisis empírico demuestra claramente que los tweets son, en su mayor parte, positivos hacia la marca. Una de las cuestiones esenciales a este respecto es que, aplicando el método NRC y las emociones de Plutchik, destaca el sentimiento de confianza. Como hemos comentado anteriormente, es de suma importancia que los clientes confíen en la marca y en sus productos, ya que eso será uno de los motivos por lo que elegirán defenderla y promocionarla en sus entornos más cercanos. De este análisis también extraemos un alto grado en la emoción de la alegría, pudiendo ser, en gran medida, por haber establecido una vinculación con la etapa de la niñez.

Como comprobación final, he sacado algunos ejemplos de los tweets con puntuación positiva o negativa más alta, para ver cuáles eran los textos literales. En el caso de los tweets positivos, vemos que, efectivamente, el resultado es positivo hacia ColaCao, ya que habla de lo bueno que está o de que es lo que eligen los consumidores. Sin embargo, algunos de los ejemplos de tweets con sentimiento negativo hablan positivamente de la marca, concretamente tratan sobre la preferencia del Cola Cao porque tiene esos grumitos que tanto les distinguen, o simplemente por lo que les apetece tomarse uno. Aquí se ve reflejado uno de los principales problemas que tiene el lexicón de sentimiento. Si nos fijamos en los dos últimos tweets de la *Ilustración 30*, vemos como el programa no está reconociendo la negación (“no me vendría nada mal” no se está interpretando como “me vendría bien”) o clasifica demasiado rápido una palabra (“lluvia” o “dolor” en este caso ya se están percibiendo como algo negativo, en vez de reconocerlo como algo positivo hacia el producto).

Comiendo pan del bueno con aceite de oliva y un ajo restregao con un cola cao viendo la bola de cristal un sábado por la mañana

Está to bueno el cola cao compa

Bueno pues cola cao

Ilustración 29: Tweets con sentimiento positivo sobre Cola Cao.

Mi hermano a comprado un cola cao falso que parece nesquik y refuerza mi teoría que sabe mejor el Colacao ni grumos tiene eso

Bendito cola Cao para esta lluvia y este dolor de garganta
un cola cao ahora no me vendría nada mal

Ilustración 30: Tweets con sentimiento negativo sobre Cola Cao.

```
tweetsnegativos<-TweetsColaCao[TweetsColaCao$anger > 4, "text"]
print(tweetsnegativos)
View(tweetsnegativos)

tweetspositivos<-TweetsColaCao[TweetsColaCao$positive > 2, "text"]
print(tweetspositivos)
View(tweetspositivos)
```

Ilustración 31: Código extracción emociones.

5. CONCLUSIONES DEL ESTUDIO

Una vez expuesta tanto la parte teórica sobre el marketing y el concepto de Love Brand, como el análisis empírico realizado respecto a la marca Cola Cao, en la que ya se han explicado los resultados obtenidos, cabe destacar:

- 1) El amor por las marcas es algo cada vez más habitual en la sociedad y una cuestión de especial relevancia para el marketing, ya que con la diversidad de opciones que se encuentran en el mercado, lograr un buen posicionamiento en este aspecto es esencial. El objetivo principal es establecer una vinculación con el cliente y que este elija y defienda a la marca frente a sus competidores en el mercado.
- 2) Las nuevas tecnologías, en especial las redes sociales, han permitido que las marcas optimicen sus recursos para lograr sus objetivos y fomentar sus estrategias de marketing, pero también que lo utilicen para conocer la opinión de los clientes.
- 3) Nuevas herramientas como el *text mining* o el *sentiment analysis* dan la posibilidad de extraer comentarios de redes sociales como Twitter y conocer las emociones que una marca despierta en sus consumidores. Respecto al caso que nos ocupa, la marca Cola Cao, cabe extraer las siguientes conclusiones del estudio empírico realizado:
 - a) El sentimiento de los consumidores sobre la marca Cola Cao es un sentimiento positivo. Se ha llegado a esta conclusión al extraer los comentarios y someternos a una serie de herramientas de análisis de datos y obteniendo una polaridad positiva. En concreto, el método NRC ha dado unos valores bastantes altos para emociones consideradas positivas como la alegría o la confianza, mientras que, en el caso de emociones negativas como la ira o la aversión, este resultado da número bastante más pequeños.
 - b) La confianza es la emoción que más destaca en los resultados del estudio. Esta es una de las cuestiones principales que mencionábamos anteriormente para poder situarse como Love Brand en el mercado. Es de especial importancia que el consumidor confíe en la marca y en el producto que obtiene de esta, ya que esto hará que sigan comprando y defendiendo todo lo que hace Cola Cao.

El estudio realizado por Madden y Shimp hablaba de que la mezcla de la intimidad, la pasión y la decisión lleva a los consumidores a sentir auténtica lealtad hacia la marca. La lealtad es una emoción que encuentra íntegramente relacionada con la confianza. Esto dificultará que, con la entrada de otros competidores en este mercado, estos consigan situarse en una posición similar a la de Cola Cao. Sus consumidores más leales confían en ella y en todo lo que esta les puede aportar frente a otros, llegando incluso a perdonarles algunos fallos en el futuro, algo que es muy complicado hacer con marcas en las que no confías y a las que no eres leal.

- c) La alegría y la tristeza son emociones que se despiertan de una forma más o menos similar en la marca Cola Cao. Esto posiblemente esté vinculado a la nostalgia de la infancia, ya que las sensaciones vividas de niños se suelen recordar con mucha alegría, pero también con cierto tono de tristeza por no vivirlas más.

Por todo ello, tomando en consideración las conclusiones extraídas, cabe asegurar que Cola Cao es un claro ejemplo de *Love Brand* en el mercado nacional. El estudio empírico realizado es la prueba de que los consumidores sienten verdadero amor por la marca y que no se trata de un fenómeno aislado, ya que los datos extraídos han sido completamente aleatorios y se han considerado más de 600 comentarios al respecto. Las emociones que la marca despierta son diversas, destacando la confianza y la alegría, algo que resulta muy acorde a los valores que defiende la marca y las estrategias que esta sigue para ganar su posicionamiento en el mercado. Dentro de sus mayores éxitos es que sus consumidores *target* son tanto niños como adultos y una de las emociones con las que más se identifica es la nostalgia, lo cual hace que sus consumidores sientan verdadera pasión por la marca al identificarla con sentimientos que suelen ser, en su mayor parte, de los más felices de su vida.

6. BILIOGRAFÍA

Amstrong, G. y. (2013). *Fundamentos de Marketing*. México: Pearson.

Anagrama Comunicación. (s.f.). *La canción del Cola Cao, referente de la publicidad española*. Obtenido de <https://anagramacomunicacion.com/publicidad/la-cancion-del-cola-cao/>

Active Campaign. (2022). *+100 estadísticas de marketing digital para 2023*. Obtenido de <https://www.activecampaign.com/es/blog/estadisticas-marketing-digital#>

Ballesteros, R. H. (2020). *Marketing más de un siglo de historia: antecedentes, génesis y evolución*. Bogotá.

Casarotto, C. (23 de febrero de 2021). *Love Brand: learn from the most loved brands in the world*. Obtenido de Rockcontent: <https://rockcontent.com/blog/love-brand/>

Cassep-Borges, V. (2009). *Versión reducida de la escala triangular del amor: características del sentimiento en Brasil*. Obtenido de Periódicos Electrónicos en Psicología: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0034-96902009000100004

ColaCao. (2021). *Cola Cao 75 años de grumitos*. Obtenido de <https://colacao.es/bundles/colacaoweb/documents/historia.pdf>

ColaCao Acción Social. (2021). *Descubre nuestro propósito*. Obtenido de <https://colacao.es/accion-social>

Florido, M. (2021). *Redes Sociales: qué son, para qué sirven y cuáles son las más utilizadas por profesionales y empresas*, Escuela Marketing and Web. Obtenido de <https://www.marketingandweb.es/marketing/redes-sociales/redes-sociales-que-son-para-que-sirven/>

Fuente, O. (25 de abril de 2022). *Definición del Marketing ¿Qué es el Marketing Moderno?* Obtenido de IEBS: <https://www.iebschool.com/blog/marketing-marketing/>

Gallart-Camahort, Callarisa y Sánchez. (15 de abril de 2019). *Concepto de marca y sus perspectivas de análisis: una revisión de la literatura*. Revista de Marketing Aplicado, 23 (1), p. 41-56. Obtenido de https://ruc.udc.es/dspace/bitstream/handle/2183/23487/Redmarka_23_1_2019_art_3.pdf

Impact5. (s.f.). *¿Qué es una love brand y por qué todas las marcas quieren serlo?* Obtenido de Impact5: <https://www.impact5.es/que-es-una-love-brand/>

Interempresas. (25 de marzo de 2022). *Guía para la digitalización del comercio minorista*. Obtenido de <https://www.interempresas.net/Juguetes/Articulos/383621-Guia-para-la-digitalizacion-del-comercio-minorista.html>

Invertia. (24 de noviembre de 2019). *ColaCao, el imperio del cacao que crearon dos cuñados en un pequeño local de Barcelona. El Español*.

Iturriaga, M. L. (21 de enero de 2016). Los chefs responden: ¿Cola Cao o Nesquik? *El País*. Obtenido de https://elcomidista.elpais.com/elcomidista/2016/01/11/articulo/1452537563_192653.html.

Jurado González, G. S. (19 de septiembre de 2019). *Love Brand: Relaciones emocionales con marcas*. Obtenido de <https://blog.antropologia2-0.com/es/love-brand/>

La Brújula del Canto. (2021). *La música en la publicidad*. Obtenido de <https://www.labrujuladelcanto.com/2021/08/la-musica-en-la-publicidad.html>

Ley 17/2001, de 7 de diciembre, de Marcas (BOE 2 de diciembre de 2001)

Madden y Shimp, T. y. (1988). *Consumer-Object Relations: a Conceptual Framework Based Analogously on Sternberg's Triangular Theory of Love*. Obtenido de *Advances in Consumer Research* Volume 15: <https://www.acrwebsite.org/volumes/6810/volumes/v15/NA-15>

Marketing Directo. (10 de noviembre de 2016). ¿Cómo convertimos nuestra marca en una "love brand"? Obtenido de <https://www.marketingdirecto.com/especiales/enamorando-al-consumidor/convertimos-nuestra-marca-una-love-brand>

Marketing Spain. (11 de abril de 2022). ¿Qué es un brand ambassador y qué funciones realiza? Obtenido de <https://www.teamlewis.com/es/revista/brand-ambassador/>

Martínez, C. I. (2 de febrero de 2023). CASIO: “Nuestras ventas se incrementaron tres veces en la última semana [tras la canción de Shakira y Bizarrap]” . *El Comercio*.

Mateu-Mollá, J. (16 de octubre de 2019). *La rueda de las emociones de Robert Plutchik: ¿qué es y qué muestra?* Obtenido de *Psicología y Mente*: <https://psicologiyamente.com/psicologia/rueda-emociones-robert-plutchik>

Morales Cáceres, A. ¿Cuál es el impacto jurídico que tiene el uso de redes sociales en lo que respecta al ámbito del derecho a la protección de datos personales?. *Agnitio* Obtenido de <https://agnitio.pe/articulo/cual-es-el-impacto-juridico-que-tiene-el-uso-de-redes-sociales-en-lo-que-respecta-al-ambito-del-derecho-a-la-proteccion-de-datos-personales/>

OEPM. (s.f.). *Oficina Española de Patentes y Marcas*. Obtenido de https://www.oepm.es/es/signos_distintivos/marcas_nacionales/mas_informacion/informacion_marcas_nombres/que_es_marca.html

OEPM. (s.f.). *Oficina Española de Patentes y Marcas*. Obtenido de https://www.oepm.es/es/propiedad_industrial/preguntas_frecuentes/FaqSignos07.html

OkDiario. (2022). ¿ColaCao o Nesquik? El mapa que resuelve la duda y deja claro cuál es el ganador en España. *Ok Diario*.

Puig de Falcó, C. (s.f.). *Las marcas como relaciones*. Obtenido de *Brandward*: <https://branward.com/branderstand/las-marcas-relaciones/>

ReasonWhy. (17 de enero de 2023). Hablamos con Renault: “Esta historia ha permitido que muchísima gente conozca o recuerde un modelo que no tenían en mente”. *ReasonWhy*.

Rock Content. (20 de noviembre de 2017). *Marketing 2.0: ¿de que trata este asunto?* Obtenido de <https://rockcontent.com/es/blog/marketing-2-0/>

Santin, L. (29 de junio de 2022). *¿Cuál es la edad de los usuarios en redes sociales?* Obtenido de Media Click: <https://mediaclick.es/blog/cual-es-la-edad-de-los-usuarios-de-las-redes-sociales/#Redes-sociales-para-adultos>

Translinguo. (2022) *¿Cuál es el origen de la palabra influencers en 2022?* Obtenido de <https://translinguoglobal.com/cual-es-el-origen-de-la-palabra-influencers/>

Valera García, L. (2016). *¿Cómo influye la marca en el consumidor en los productos de consumo no duradero? Estudio empírico entre marcas líderes y marcas blancas.* Obtenido de <https://repositorio.upct.es/xmlui/bitstream/handle/10317/5816/tfg-val-com..pdf?sequence=1&isAllowed=y>

ANEXO I.

```
install.packages("rtweet")
install.packages("syuzhet")
install.packages("rvest")
install.packages("ggplot2")
install.packages("quanteda")
install.packages("tidytext")
install.packages("lubridate")
install.packages("quanteda.textstats")
install.packages("wordcloud")
install.packages("udpipe")
install.packages("twitter")
install.packages("readexcel")
install.packages("readxl")
install.packages("wordcloud2")
```

```
library(rtweet)
library(syuzhet)
library(rvest)
library(xml2)
library(ggplot2)
library(dplyr)
library(stats)
library(base)
library(quanteda)
library(scales)
library(tidytext)
library(lubridate)
library(RColorBrewer)
library(stringr)
library(quanteda.textstats)
library(wordcloud)
library(utf8)
library(udpipe)
library(tidyr)
library(twitter)
library(readxl)
library(wordcloud2)
```

```
setwd("/Users/martanoguera/Desktop/TFG Analytics")
Tweets <- read_excel("Colacao.xlsx")
summary(Tweets) #confirm data is correct
View(Tweets)
```

```
TweetsColaCao<-select(Tweets, id, created_at, text, source)
View(TweetsColaCao)
```

```
TweetsColaCao <- TweetsColaCao %>%
  select(id, source, text, created_at) %>%
  extract(source, "source", "Twitter (.*)<") %>%
  filter(source %in% c("for iPhone", "for Android", "Web App"))
summary(TweetsColaCao)
table(TweetsColaCao$source)
View(TweetsColaCao)
```

```
#Quitamos menciones eliminando el signo “@” y a quien menciona, pero no el tweet
TweetsColaCao$text<-gsub( "@\\w+", "", TweetsColaCao$text)
```

```
#Elimino paginas web que comienzan por http
TweetsColaCao$text<- gsub("http\\S*", "", TweetsColaCao$text)
```

```
#Quito puntuación, exclamaciones... (!"#$%&???)*+,-./:;<=>?@[\\]^_`{|}~)
TweetsColaCao$text<-gsub("[:punct:]", "", TweetsColaCao$text)
```

```
#Elimino dígitos (los números no sirven para capturar emociones)
TweetsColaCao$text<-gsub("[:digit:]", "", TweetsColaCao$text)
```

```
dim(TweetsColaCao)
View(TweetsColaCao)
```

```
##### CONSTRUCCIÓN DEL CORPUS Y DE LA MATRIZ DTM #####
```

```
### Corpus
```

```
micorpus<-corpus(TweetsColaCao)
summary(micorpus)
```

```
### Matriz DFM
```

```
midfm<- dfm(micorpus)
dim(midfm) #He detectado 2549 términos
featnames(midfm) #Observamos cuáles son esos términos
topfeatures(midfm,50) #Primera visualización de las 50 palabras que mas se repiten
```

```
##### LIMPIEZA DE DATOS Y REDUCCIÓN DE LA MATRIZ DTM #####
```

```
#Eliminamos puntuación, números y stopwords, tanto en inglés como en español
```

```
micorpus <- tokens(micorpus, remove_punct = TRUE) %>%
  tokens_remove(stopwords("es")) %>%
  tokens_wordstem()
```

```
micorpus <- tokens(micorpus, remove_punct = TRUE) %>%
```



```
tokens_remove(stopwords("en")) %>%
tokens_wordstem()
```

```
midfm<- dfm(micorpus)
dim(midfm)
featnames(midfm)
topfeatures(midfm,30)
```

#Hacemos stemming (luego se elimina)

```
midfm <- dfm_wordstem(midfm, language = "spa")
dim(midfm)
featnames(midfm)
```

#Gráfico de los términos mas frecuentes

```
features_freq<-textstat_frequency(midfm, n = 20)
View(features_freq)
ggplot(features_freq, aes(x = reorder(feature,-frequency), y = frequency)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

NUBE DE PALABRAS

```
set.seed(123)
midfm_t <- t(midfm)
m <- as.matrix(midfm_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)
```

SENTIMENT ANALYSIS

#Frecuencia de cada término dentro de cada grupo (comentarios positivos y negativos)

```
freq_posneg <- textstat_frequency(midfm, n = 10, groups = "posneg")
ggplot(data = freq_posneg, aes(x = nrow(freq_posneg):1, y = frequency, fill=group)) +
  geom_bar(stat="identity") +
  facet_wrap(~ group, scales = "free") +
  coord_flip() +
  scale_x_continuous(breaks = nrow(freq_posneg):1,
                    labels = freq_posneg$feature) +
  labs(x = NULL, y = "frecuencia de cada termino")
```

USO DE KEY WORDS DEL PAQUETE QUANTEDA

Document Term matrix con valores tf-idf

#Utilizamos la función dfm_tfidf y debemos pasarle la DTM creada con este corpus

```
mitf_idf<-dfm_tfidf(midfm, scheme_tf="prop", base=2)
```

```
##### SENTIMENT ANALYSIS #####
```

```
#Polaridad de los tweets atendiendo a los 3 métodos
```

```
get_sentiment(TweetsColaCao$text, method="afinn")
```

```
get_sentiment(TweetsColaCao$text, method="bing")
```

```
get_sentiment(TweetsColaCao$text, method="nrc")
```

```
#Incorporo dos variables nuevas al dataframe TweetsColaCao, la polaridad de cada tweet con NRC y la polaridad con bing
```

```
TweetsColaCao<-TweetsColaCao %>% mutate(nrc_polarity= get_sentiment(text, method="nrc"), bing_polarity=get_sentiment(text, method="bing"))
```

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

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

```
#Procedencia del mismo porcentaje agrupándolo por source
```

```
TweetsColaCao %>% group_by(source) %>%  
  summarise(positivos = 100 * sum(nrc_polarity > 0) / n(),  
            neutros = 100 * sum(nrc_polarity == 0) / n(),  
            negativos = 100 * sum(nrc_polarity < 0) / n())
```

```
##### Análisis de las emociones de Plutchik #####
```

```
#Capturamos las emociones con NRC y las incorporamos al data frame
```

```
emociones<-get_nrc_sentiment(TweetsColaCao$text, language="spanish")
```

```
TweetsColaCao<-TweetsColaCao %>% mutate (anger=emociones$anger,  
                                           anticipation=emociones$anger,  
                                           disgust=emociones$anger,  
                                           fear=emociones$fear,  
                                           joy=emociones$joy,  
                                           sadness=emociones$sadness,  
                                           surprise=emociones$surprise,  
                                           trust=emociones$trust,  
                                           negative=emociones$negative,  
                                           positive=emociones$positive)
```

```
View(TweetsColaCao)
```

```
summary(TweetsColaCao)
```

```
tweetsnegativos<-TweetsColaCao[TweetsColaCao$anger > 4, "text"]
```

```
print(tweetsnegativos)
```

```
View(tweetsnegativos)
```

```
tweetspositivos<-TweetsColaCao[TweetsColaCao$positive > 2, "text"]
```

```
print(tweetspositivos)
View(tweetspositivos)
```

#Gráfico de barras para ver el peso que tiene cada emoción en el corpus

```
barplot(
  sort(colSums(prop.table(TweetsColaCao[, 5:16]))),
  horiz = TRUE,
  cex.names = 0.7,
  col = brewer.pal(n = 8, name = "Set3"),
  las = 1,
  main = "Emociones", xlab="Porcentaje"
)
```