

« Análisis de mercado a partir de la extracción de datos en la red social Twitter »

by

«Marina Vegas Arenas»

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un
i Universidad
Internacional
de Andalucía
A

November 2018

« Análisis de mercado a partir de la extracción de datos en la red social Twitter »

« Marina Vegas Arenas »

Máster en Economía, Finanzas y Computación

« Juan Diego Borrero Sánchez »

Universidad de Huelva y Universidad Internacional de Andalucía

2018

Resumen

Tras los cambios experimentados estos últimos años por la sociedad, debido al avance de las redes sociales y presencia de las mismas en nuestro día a día, el mundo empresarial ha tenido que tomar medidas al respecto e incorporar estas nuevas herramientas en su estrategia de negocio.

Mediante el análisis de redes sociales, las empresas obtendrán datos sobre el consumidor y conocer su modelo de consumo de una manera rápida y eficiente, ya que esa información es proporcionada directamente por el usuario mediante las interacciones en las redes, tanto con otros usuarios como con las propias empresas. A partir de esto es posible obtener métricas, niveles de audiencia, preferencias, etc.

El fin de este trabajo es aprovechar la herramienta de microblogging Twitter, extrayendo los tweets con RStudio y representándolos con Gephi. El análisis tratará de aportar información a los minoristas del sector de las berries, de la provincia de Huelva, sobre el consumo de su producto en relación con los principales distribuidores de Europa, por lo que las fuentes de información de nuestro estudio serán los supermercados europeos y consumidores del producto que así lo expresen en redes sociales.

Palabras clave: berries, fresas, supermercados, redes sociales, Twitter, análisis de redes sociales, RStudio, Gephi.

Abstract

Society has been experiencing changes due to the new technologies progress and the great development and presence of social networks in our daily live. Consequently, the business world has had to take actions and integrate those new tools in its strategies.

Through social network analysis, companies can get information about consumers and their consumption model in an easy and efficient way, since the most part of that information is directly given by users when interacting in different platforms with other users as well as with different social media channels or companies themselves. From then on, it is possible to get metrics, measure audience rating, preferences, etc. Taking into account all this information, it is easier to take decisions, define the business strategy and the approach of marketing campaigns through social media.

The aim of this work is to take advantage of these analysis tools through Twitter social network, by extracting tweets with RStudio and representing them with RStudio as well as with Gephi. The purpose of the analysis is to provide information about consumption of berries in Europe in relation to biggest business volumes' distributors in the eurozone to retail businesses within this sector in Huelva. For this reason, the main users of our research will be the most relevant european supermarkets and the consumers of the product who show it in social networks.

Keywords: berries, fresas, supermercados, redes sociales, Twitter, análisis de redes sociales, investigación de mercados, estrategia, RStudio, Gephi, strawberries, supermarket, social media, Twitter, social network analysis, market research, strategy.

Tabla de contenidos

1. Introducción	p.7
1.1. Objetivos	p.9
1.2. Justificación de la red social elegida	p.10
1.3. Justificación de los usuarios elegidos	p.12
2. Revisión de la literatura	p.15
3. Metodología	p.17
4. Análisis	p.18
5. Discusión	p.30
6. Conclusiones	p.34
7. Referencias	p.36
8. Anexos	p.38
- Anexo I: Código empleado en R	p.38

Lista de tablas

Tabla 1. Usuarios para el análisis.....	p.14
Tabla 2. Frecuencia de términos junto a las palabras ‘fresas’ y ‘berries’ en Twitter.....	p.22
Tabla 3. Comparativa de los usuarios analizados.....	p.23
Tabla 4. Número de interacciones de los usuarios analizados en torno a los términos estudiados.....	p.28
Tabla 5. Recuento de la muestra tomada de la clasificación de los usuarios clasificados por países.....	p.29

Lista de figuras

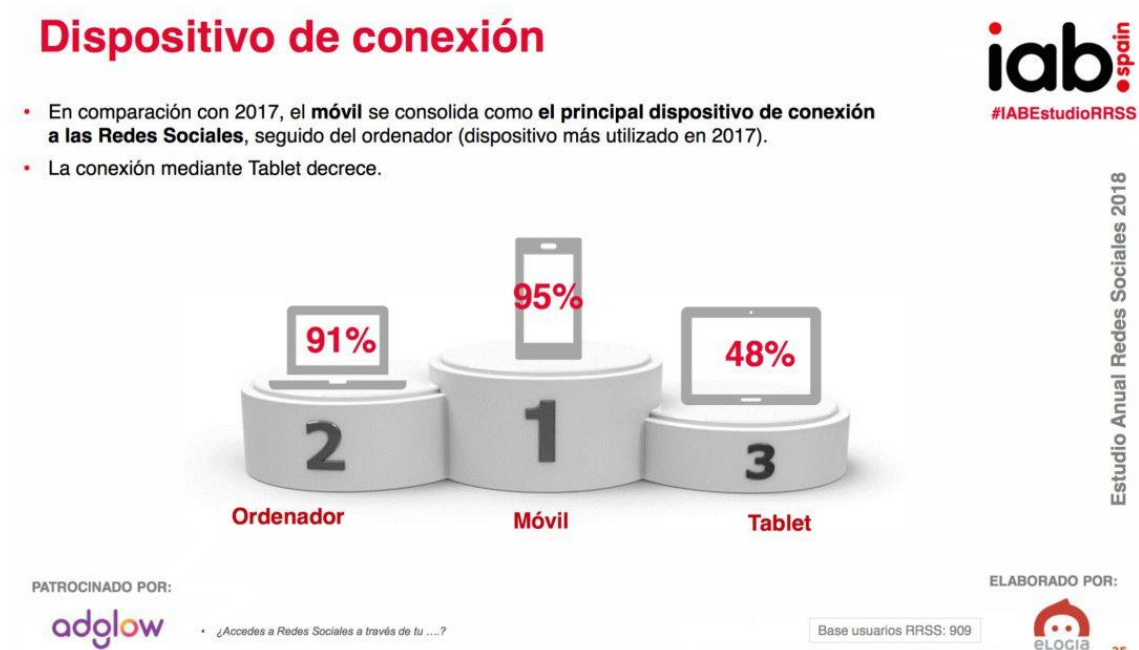
Figura 1. Dispositivos de conexión a las Redes Sociales	p. 7
Figura 2. E-Commerce en Redes Sociales	p. 8
Figura 3. Usuarios activos en las principales Redes Sociales en el mundo.....	p. 10
Figura 4. Logo Twitter Analytics.....	p. 12
Figura 5. Los líderes del Gran Consumo.....	p. 13
Figura 6. Nube de términos (fresas).....	p. 20
Figura 7. Nube de términos (berries).....	p.20
Figura 8. Influencias de las cuentas en la red.....	p.24
Figura 9. Red de hashtags.....	p.26
Figura 10. Red de interacción en Twitter de los principales supermercados europeos.....	p.27
Figura 11. Proporción de los usuarios tomados como muestra de la clasificación de los usuarios por países.....	p.30

1. Introducción

En un entorno tan cambiante como es el de internet y las redes sociales la labor de comunicación de la empresa se alinea con el objetivo empresarial. Con las redes sociales el acceso al público potencial y la posibilidad de enviar un mensaje más personalizado al cliente se ha vuelto mucho más fácil y accesible, haciendo posible satisfacer las necesidades del consumidor de una forma más eficiente (Martínez, 2013). Esta revolución de los canales de comunicación online viene encabezada por una serie de redes sociales, que son las que acaparan las mayores cifras de suscriptores e interacciones en sus plataformas, como son Facebook, Youtube, Whatsapp, Twitter, etc.

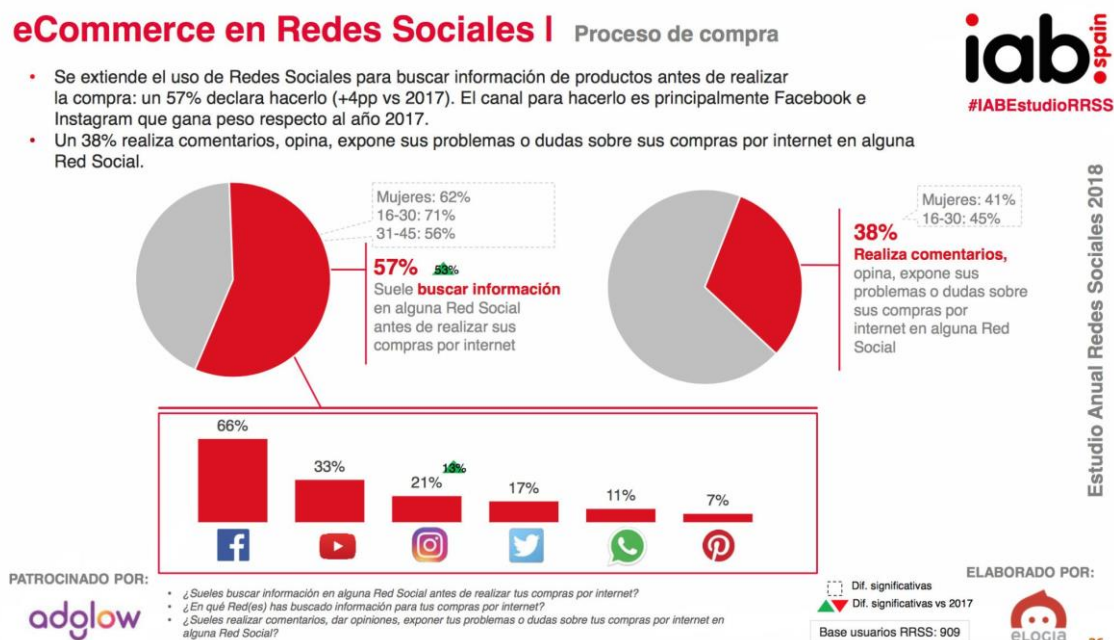
El impulso de las redes sociales se ha visto potenciado por el avance en paralelo de las tecnologías y el desarrollo de los distintos dispositivos que hacen más fácil interactuar con ellas. Es el smartphone la primera opción elegida por los usuarios para interactuar en sus redes sociales, dada facilidad de acceso a través del mismo y cómodo uso; le siguen dispositivos como el ordenador y la tablet (IAB Spain, 2018).

Figura 1: Dispositivos de conexión a las Redes Sociales. Fuente: IAB Spain.



A partir de esto datos es fácil ver la importancia de la publicidad digital para las marcas, con un peso cada vez más relevante en las cifras de ventas e imagen de la marca. Según los datos obtenidos por el estudio de IAB (IAB Spain, 2018), el 81% de los usuarios reconoce seguir a alguna marca y el 27% considera que las marcas con perfiles en redes les inspiran más confianza. A su vez, el 32% acepta positivamente publicidad acorde a sus intereses mientras el 48% admite haber visto publicidad que se ajustaba a su perfil. Al conocer estos datos, observamos como el e-commerce se abre paso gracias a estas nuevas vías de comunicación ya que el 57% busca información sobre el producto o servicio antes de adquirirlo, teniendo en cuenta los comentarios y opiniones de los usuarios respecto a esa marca o producto y uno de cada cuatro usuarios utiliza las redes sociales para comprar el producto. Otros usuarios, aunque no la utilicen como canal de compra, la usan para captar la atención al cliente o para relacionarse directamente con la marca.

Figura 2: eCommerce en Redes Sociales. Fuente: IAB Spain



Estos nuevos recursos empresariales posibilitan la realización de una investigación del mercado singular, puesto que la información obtenida por estas vías aporta un valor

diferencial al estudio, ya que se genera en un entorno abierto y natural, donde se producen las conversaciones e interacciones entre los participantes (Solo Marketing, 2013).

Este trabajo tratará de aportar un conocimiento más amplio del panorama actual del sector de la distribución de berries en el territorio europeo a aquellos minoristas de la provincia de Huelva interesados en conocer cuáles serían unos clientes potenciales fuera de las fronteras españolas, centrando nuestro estudio en su consumo a través de supermercados, y cómo estos interactúan mediante Twitter con sus seguidores y clientes. Con la minería de texto podremos llegar a discutir los resultados objetivos, ya que junto con la extracción de tweets realizada con RStudio y la representación de los datos en Gephi, conseguiremos mejor perspectiva de los movimientos del mercado y poder extraer unas conclusiones y recomendaciones.

1.1. Objetivos

El objetivo general del trabajo será realizar un estudio del mercado de las berries, trabajando con técnicas de data mining y text mining, a través de la red social Twitter para proporcionar información a aquellos minoristas del sector sobre cómo se percibe su producto en la red de Twitter y cómo se distribuyen en los principales supermercados europeos.

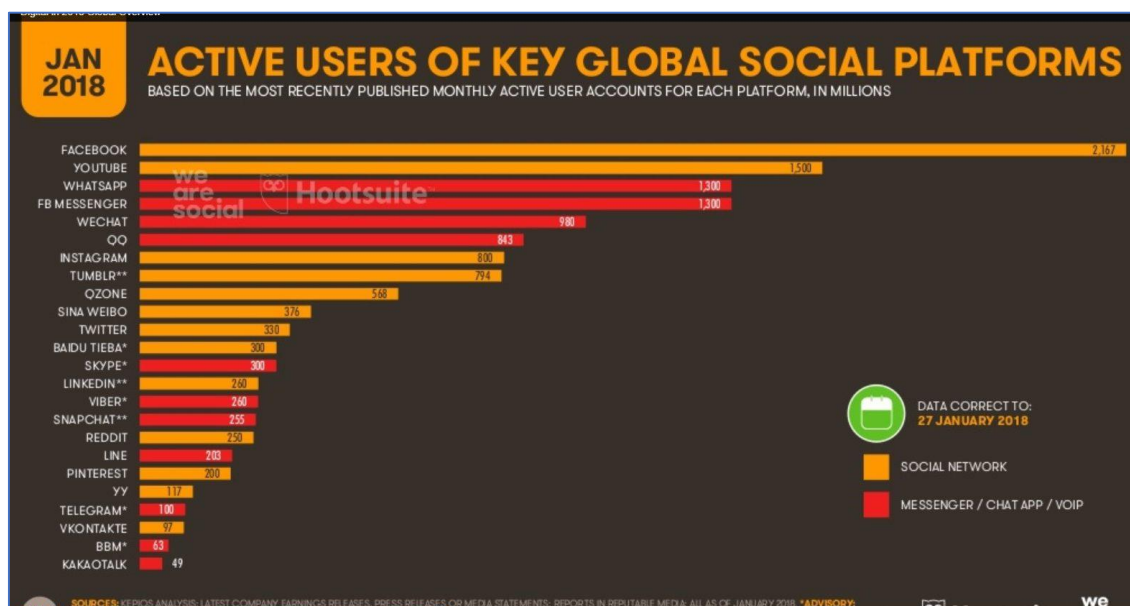
Así, como objetivos específicos, nos planteamos los siguientes:

1. Estudiar de qué manera se expresan los usuarios en Twitter cuando se refieren al producto analizado.
2. Realizar un análisis sobre el posicionamiento del producto en Europa.
3. Analizar las interacciones de los supermercados en Twitter con los usuarios.
4. Clasificar los usuarios analizados en función a su popularidad y seguidores.
5. Determinar en qué países se menciona más a los productos del estudio en la red.
6. Desarrollar una representación gráfica de los datos obtenidos.
7. Proporcionar una guía básica sobre cómo realizar una investigación de mercados en Twitter con RStudio y Gephi.

1.2. Justificación elección red social elegida

Para la investigación que queremos llevar a cabo, la primera opción considerada fue analizar el mercado a través de la red social Facebook, dadas sus características y público. Facebook es la red social por excelencia, con 2.167 millones de usuarios, aproximadamente, activos en el pasado mes de enero de 2018 según el estudio “*Digital in 2018: World’s Internet users pass the 4 billion mark*” llevado a cabo por las empresas *We are Social* y *Hootsuite*. Al ser una plataforma con un público tan amplio y con un gran número de interacciones, las empresas vieron pronto una oportunidad de negocio en ella, incorporándola en sus estrategias empresariales y plan de marketing como herramienta clave para el desarrollo empresarial. Sin embargo, tras la implantación en pasado 25 de mayo del 2018 del nuevo Reglamento General de Protección de Datos (RGPD) la obtención de información a través de los usuarios y sus perfiles se ha visto bastante dificultada, llegando a ser un auténtico reto la aprobación de Facebook para acceder a esos datos. Es por esto por lo que, decidimos finalmente recurrir a Twitter como herramienta de análisis.

Figura 3. Usuarios activos en las principales Redes Sociales en el mundo. Fuente: We Are Social & Hootsuite.



Desde la aparición de Twitter en el año 2006 esta red social ha ido experimentando una evolución y crecimiento muy positivo en su uso y presencia frente a otras plataformas similares. Según las estadísticas arrojadas por Twitter, en enero del 2018 contaban con ya 330 millones de usuarios y una media de 500 millones de tweets al día. Tanto con un enfoque más privado, como con uno más comercial, Twitter arroja información sobre los individuos que interactúan en ella, las preferencias y opiniones de los usuarios, así como las reacciones de las personas frente a la información contenida en cada tweet. Ya sea mediante caracteres en un tuit de elaboración propia, como con un retweet de otro usuario, “me gustas” o menciones, Twitter contiene información muy valiosa acerca del comportamiento del consumidor; además incluso permite conocer la estrategia utilizada por sus competidores para captar la atención de su público y lograr la satisfacción de los clientes. Con el fin de usarla como herramienta empresarial y poder sacar el máximo provecho a esta red como fuente de información sobre el mercado, y en concreto sobre el consumidor, Twitter lanzó Twitter Analytics. Con esta aplicación, Twitter permite llevar un estudio diario de las estadísticas de tu cuenta mediante diversos indicadores, como explican en su web <https://business.twitter.com/es.html>¹:

1. *“El inicio de cuenta es tu card de informes de Twitter, con estadísticas de alto nivel recopiladas mes a mes. También es una galería de tus grandes éxitos: destacaremos tus Tweets con mejor rendimiento y te presentaremos a las personas influyentes de tu red.*
2. *Tu panel de actividad de Tweet es donde encontrarás indicadores de cada uno de tus Tweets. Sabrás exactamente cuántas veces los usuarios de Twitter han visto, retwitteado, indicado que les gusta y respondido cada Tweet.*
3. *Tu panel de información de audiencias contiene información valiosa sobre las personas que te siguen en Twitter. Puedes hacer un seguimiento del crecimiento de tus seguidores a lo largo del tiempo y obtener más información sobre los intereses y los datos demográficos de tus seguidores.”*

¹ <https://business.twitter.com/es.html>. Consultada 18.08.2018

Sin embargo, para este estudio no emplearemos esta herramienta ya que consideramos más adecuada para la investigación a desarrollar el uso combinado de RStudio y Gephi, la primera para la extracción y la segunda para la representación de los datos.

Figura 4. Logo Twitter Analytics. Fuente: Bocetos.com.



1.3. Justificación de los usuarios elegidos

Para poder llevar a cabo un estudio con unos resultados concluyentes que realmente nos aporten la información deseada, debemos desarrollar una investigación minuciosa de los usuarios a analizar y encontrar aquellos que puedan responder a las preguntas planteadas.

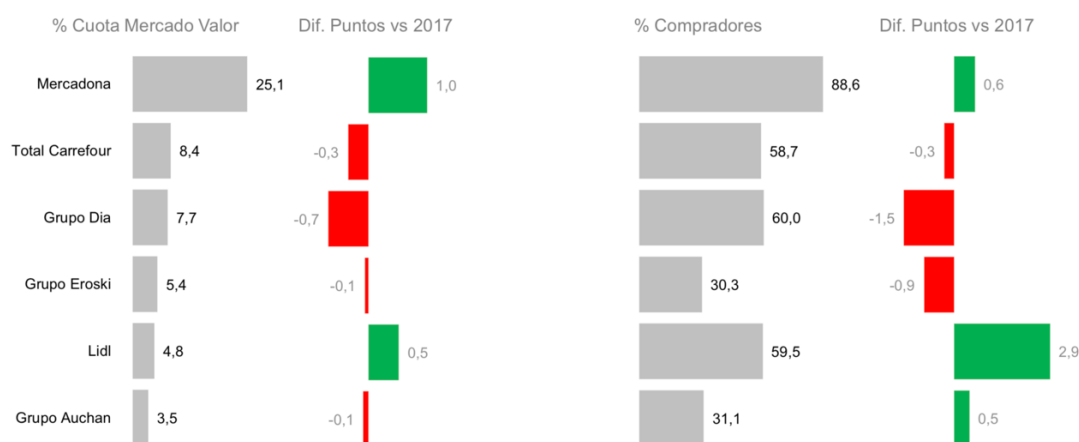
Según el estudio “La situación de la gran distribución en el mercado de consumo” realizado por la EAE Business School (EAE Business School, 2018), la gente acude a realizar sus compras del hogar a los supermercados, autoservicios e hipermercados como principales canales de distribución de productos de alimentación. El resto de opciones, como son los establecimientos tradicionales y canales de venta online quedan por detrás de los anteriores canales. En el sector que a nosotros nos interesa, que es el de los productos frescos, tanto los supermercados, tiendas de autoservicio como las tiendas tradicionales tienen una cuota de mercado muy similar (32-35 %), siendo esta la opción más frecuentada por los consumidores a la hora de elegir dónde adquirir los productos frescos (EAE Business School, 2018). Tras conocer estos datos hemos optado por el

análisis del consumo de los berries en los principales supermercados de España y Europa, puesto que la mayor parte de los berries onubenses se exportan a Europa.

El gran porcentaje del gasto de los consumidores en el sector de la distribución en España ha ido evolucionando a lo largo de los años, pero actualmente recae sobre un número reducido de cadenas de supermercados. Con un gasto medio de 4.190 euros al año por hogar, la mayor parte de este gasto se realiza en la cadena valenciana Mercadona (25,1%) y en Carrefour (8,4%); seguido por los supermercados del Grupo Día (7,7%) y Grupo Eroski (5,4%), Lidl (4,8%) y el Grupo Auchan/Alcampo (3,5%). Con un total del 54,9% de la cuota de mercado, estos serían las seis cadenas de distribución elegidas por los usuarios para sus compras del hogar en el año 2018 según un estudio realizado por *Kantar Worldpanel* (2018).

Figura 5. Los líderes del Gran Consumo. Fuente: Kantar Worldpanel.

Total FMCG
Acumulado a 12 de Agosto de 2018 (YTD)



Gran Consumo: Alimentación (con frescos perecederos) + Droguería + Perfumería familiar + Baby + Pet Food
Súpers Regionales no incluye Supermercados Eroski ni El Arbol

KANTAR WORLD PANEL

YTD a 12 de Agosto de 2018
Fuente: Kantar Worldpanel

No obstante, en este estudio vamos a ampliar el alcance a los supermercados europeos para así darle a los proveedores españoles una visión más amplia sobre la internacionalización y posibilidades de negocio en el extranjero. El estudio se realizará por tanto en los principales supermercados europeos basándonos en estudio “*Global Powers of Retailing 2016. Navigating the new digital divide*” publicado por Deloitte

(Deloitte, 2016), descartando de la lista aquellos que no cuentan con una página oficial de Twitter ni cuenta con el número de seguidores o tweets necesarios para ser relevantes en este análisis (a partir de 3000 seguidores y 1000 tweets publicados). Tras aplicar este filtro nuestro estudio se centrará en 24 cuentas de usuarios (cuentas de supermercados). En el caso que un mismo supermercado cuente con varias cuentas por los diferentes países en los que tiene presencia, seleccionaremos aquella orientada al público hispanohablante para facilitar el análisis por palabra o término. Para que esto no provoque una segmentación de la información también incluiremos hashtags en inglés, ya que todas las cuentas, independientemente de su idioma, publican ciertos términos en inglés. En la tabla 1 se muestran las cuentas utilizadas y sus cifras a fecha 02.09.2018:

Tabla 1. Usuarios para el análisis. Fuente: Elaboración propia.

NOMBRE	USUARIO	SEGUIDORES	Nº TWEETS	ORIGEN
Aldi	@Aldi_es	10,4 mil	5.882	Alemania
Alcampo/Auchan	@alcampo	65,2 mil	20,9 mil	Francia
Carrefour	@CarrefourEs	218 mil	45,7 mil	Francia
Casino	@casino_super	15,7 mil	3.120	Francia
Co-op	@coopukfood	129 mil	104 mil	Reino Unido
Conad	@conad	21,4 mil	3.015	Italia
Coop Group	@coop_ch	19,8 mil	5.924	Suiza
Coop Italia	@Coopitalia	27,2 mil	8.007	Italia
Día Supermercado	@DIAEsp	57,9 mil	13,8 mil	España
Eroski	@EROSKI	53,5 mil	24 mil	España
EuroSpin	@EurospinItalia	5.221	1.572	Italia
Iceland	@Icelandfoods	150 mil	73,8 mil	Reino Unido
Intermarché	@intermarche	13,8 mil	1.442	Francia
Jumbo Supermarkten	@JumboSupermarkt	12 mil	48,7 mil	Países Bajos
Leclerc	@LeclercBonPlan	20,2 mil	13,3 mil	Francia
Lidl	@lidlspana	120 mil	17,5 mil	Alemania
Mercadona	@Mercadona	153 mil	72,3 mil	España
Migros	@migros	115 mil	19 mil	Suiza
Morrison	@morrison	269 mil	300 mil	Reino Unido
Rewe	@REWE_Supermark	24 mil	23,6 mil	Alemania
Sainsbury	@sainsburys	511 mil	723 mil	Reino Unido
Spar	@SPARInt	3.514	2.022	Países Bajos
Tesco	@Tesco	537 mil	2,06 M	Reino Unido
Waitrose&Partners	@waitrose	318 mil	163 mil	Reino Unido

2. Revisión de la literatura

Debido a los cambios en los modelos empresariales y en las vías de comunicación, los usuarios han ido adoptando cambios en sus hábitos de consumo y compra. Todo esto debe ser considerado a la hora de diseñar la estrategia empresarial y de marketing. En Andrews & Currim (2004) se compara el comportamiento de los consumidores asiduos a realizar la compra por la vía tradicional, yendo a la tienda física, frente a aquellos que se decantan por las compras online y el porqué de estas decisiones. La información proporcionada por este artículo junto a la expuesta por Degeratu, Rangaswamy & Wu (2000), donde de manera práctica estudian cómo este aumento de información a disposición de los usuarios vía internet afecta a la imagen del supermercado en cuestión, y a su manera de consumir, podemos crearnos una idea inicial de la evolución de todo el entorno empresarial en función a las nuevas tecnologías e incremento de popularidad de las redes sociales, y de las consecuencias que suponen para los distintos supermercados.

Para tener una visión más general de la forma de interactuar de los usuarios en diversas redes sociales hemos acudido al trabajo de Smith, Fischer & Yongjian (2012), que analiza el uso que le dan los usuarios a las redes de Youtube, Facebook y Twitter y qué supone esto para las marcas, remarcando cómo Twitter y Youtube son opuestos en muchos sentidos. Twitter se destina más a la exposición de los usuarios de su vida diaria, y a través de esta red los clientes se ponen en contacto con las marcas para formularles preguntas y resolver dudas. A su vez, destacan que en Twitter es menos probable que una marca cuente con la autopromoción de sus consumidores. Lo contrario sucede en Youtube, donde es más frecuente la mención de multimarcas por los usuarios, promocionando así el producto o servicio, resultando una forma de marketing bastante relevante. En lo respectivo a Facebook, esta plataforma podríamos situarla entre Twitter y Youtube, tanto en funcionalidad como en publicidad dentro de la misma, ya que los usuarios suelen acudir a ella tanto para contactar con la marca como para conocer más sobre su oferta. Son estos los motivos, entre otros, por lo que en este estudio se recomienda a los mercadólogos invertir en estas redes sociales y su administración y organización para maximizar sus beneficios. Cuidar al cliente y cubrir sus necesidades es un punto clave en todo plan de marketing, y estas plataformas hacen que esto sea posible de una forma más personalizada y eficiente (Smith, Fischer & Yongjian, 2012).

Como trabajos de investigación basados en el servicio de micro-blogging Twitter, encontramos el realizado por Pablo Barberá y Gonzalo Rivero “*¿Un tweet, un voto? Desigualdad en la discusión política en Twitter*” (Barberá & Rivero, 2011) donde a partir de la obtención de datos gracias a API Twitter se estudia el comportamiento del usuario en la red a partir del seguimiento de hashtags, entre otros métodos. Recurrir a estos métodos más actuales para estudiar el comportamiento del consumidor es cada vez más frecuente; ya que, además de aumentar la eficiencia, nos aproxima más al pensamiento del consumidor que las técnicas más tradicionales, como pueden ser las encuestas marketing (Chamlertwat, Bhattarakosol, Rungkasiri & Haruechaiyasak, 2012).

En relación con la importancia de la presencia de las marcas en redes sociales para establecer lazos con los consumidores Kwon, Kim, Sung & Yoo (2015) exponen en su trabajo que los usuarios tienen cuatro motivaciones principales para seguir a las marcas en Twitter, y estas son: búsqueda de incentivos, búsqueda de interacción social, uso / simpatía de la marca y búsqueda de información. Esto, sumado a la afirmación por ellos explicada en este artículo de que “*las personas que se comunican con frecuencia con sus pares sobre marcas y, en general, tienen actitudes más favorables hacia la publicidad, tienen más probabilidades de tener actitudes positivas hacia las comunicaciones de marca a través de Twitter*” tiene especial interés para nuestro análisis posterior.

Estudios relacionados con cómo interactúa el consumidor en Twitter cuando habla sobre la alimentación solo hemos encontrado el artículo “Using Twitter data for food-related consumer research: A case study on “what people say when tweeting about different eating situations” publicado en el periódico “Food Quality and Preference” (Vidal, Ares, Machín & Jaeger, 2015). En dicho estudio realizan una extracción de tweets en los que aparecen ciertas palabras seleccionadas relacionadas con las actividades relacionadas con la alimentación. Con esto se consigue un gran volumen de información sobre cómo, dónde, qué y con quién consumen los usuarios determinados alimentos, con lo que se obtiene a su vez las opiniones de los usuarios sobre temas como la salud, el bienestar. Para la obtención de datos emplearon el paquete *twitteR* del software R, al igual que nosotros en nuestro estudio. Este paquete permite conectar el software R con la Interfaz de Programación de Aplicaciones de Twitter (Vidal, Ares, Machín & Jaeger, 2015).

3. Metodología

Para alcanzar los objetivos propuestos usaremos las herramientas **R/RStudio** y **Gephi**, además de abrimos una cuenta en Twitter.

R² es un entorno y lenguaje de programación multiparadigma con un enfoque al análisis estadístico y a la representación gráfica de los resultados obtenidos. Es una implementación de software libre y se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística y en la minería de datos. R es un proyecto GNU, por lo tanto, los usuarios son libres de modificarlo y extenderlo mediante diferentes bibliotecas o paquetes disponibles para que el usuario los instale y amplíe su funcionalidad. Esto lo hace muy práctico y eficiente para nuestro análisis.

Para un uso más cómodo y ágil de R existen una gran variedad de entornos de desarrollo de éste donde desarrollar los scripts, como son RCommander, RKWard y RStudio. En nuestra investigación trabajaremos con **RStudio³** que, como hemos dicho, es un entorno de desarrollo integrado para el lenguaje de programación R, creado con el fin de poder desarrollar el análisis en R de una forma más eficiente y visual. Incluye una consola, un editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo. A su vez, también cuenta con una serie de paquetes y librerías que puedes instalar, según tus necesidades, para ampliar la funcionalidad de R y desarrollar todo tipo de análisis estadísticos y representaciones gráficas.

Tanto R como RStudio están disponibles para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.

A la hora de realizar un buen análisis de datos no solo debemos preocuparnos por la propia extracción y análisis de información, ya que también es muy importante saber representarlos de una forma visual y comprensible para que el usuario que acceda a él

² <https://www.r-project.org>. Consultada 12.11.18

³ <https://www.rstudio.com>. Consultada 12.11.18

pueda comprenderlo y obtener unas conclusiones del mismo. **Gephi**⁴ es una herramienta de código abierto y libre que facilita a sus usuarios la representación y comprensión de sus datos mediante gráficos, permitiendo interactuar con la representación, manipular las estructuras, formas y colores para revelar patrones ocultos. Su objetivo es ayudar a los analistas de datos para hacer hipótesis, intuitivamente descubrir patrones, aislar las singularidades de la estructura o fallos durante el aprovisionamiento de datos.

Para la extracción de datos y tweets desarrollaremos una API a partir de nuestra cuenta en Twitter. Una API (Application Programming Interface) es una interfaz de programación de aplicaciones, con la que podremos acceder, en nuestro caso, a ciertos datos que nos proporciona Twitter para analizar el contenido, comportamiento y actividad de los usuarios de la plataforma. A través de la página Application Management de Twitter se obtienen las claves necesarias para acceder a API Twitter desde RStudio y Gephi:

- Api key
- Api secret
- Access token
- Access token secret

Emplearemos estas claves en el código de R para poder acceder a la extracción de datos, así como en Gephi para continuar con la extracción y representación de la información a analizar (ver Anexo I).

4. Análisis

Iniciamos el análisis con una extracción en Twitter de aquellos tweets con los hashtags #fresas o #berries con RStudio, recopilando un total de 1000 tweets por hashtag (últimos tweets publicados a fecha del 03.09.2018). Esto lo hemos representado en una nube de términos basándonos en las frecuencias de los caracteres que aparecen junto a dichos hashtags.

⁴ <https://gephi.org>. Consultada 12.11.2018

En el momento de tener que decidir qué términos utilizar para cada apartado del análisis nos hemos basado en un análisis previo realizado en la red a nivel de usuario, para observar cuáles eran más populares y así poder extraer la mayor cantidad de información posible, así como en las diferencias tendencias de consumo de las berries en Europa. Aunque los términos generales que engloban a los productos del análisis sean ‘frutos rojos’ para las publicaciones en español y ‘berries’ en inglés, hemos decidido en el caso la búsqueda en español emplear la palabra ‘fresas’ como término representativo de todos los frutos rojos ya que el consumo de estas predomina claramente frente a los otros frutos de la misma categoría en nuestro país, siendo muy reducido el número de publicaciones encontradas en Twitter con el término ‘frutos rojos’ frente a las encontradas con el término ‘fresa’. Al contrario que en España, en el resto de países europeos hay mayor diversidad en el consumo de frutos rojos, por lo que el término ‘berries’ representa de manera exacta el producto de nuestro análisis. (Eurofresh Distribution, 2016)

Para este estudio previo hemos realizado un seguimiento de los términos del análisis durante cinco meses para ver cómo eran empleados por los usuarios en Twitter y qué información nos aportaba cada uno a la investigación. A partir de la búsqueda de dichos hashtags y términos en Twitter durante este periodo hemos seleccionado los términos que estudiaremos a continuación. En lo que referente a las cuentas de los supermercados analizadas para este estudio, hemos estado observando su actividad durante cinco meses, siguiendo a estos usuarios desde una cuenta de Twitter particular para así poder ofrecer una mejor interpretación de los datos obtenidos a continuación, y poniendo especial atención a los términos que empleaban relacionados con las fresas y frutos rojos para completar nuestra elección de términos para el estudio.

En las figuras 6 y 7 aparecen en el centro de la nube de términos la palabra en torno a la que se ha estudiado la frecuencia de los términos que la acompañan. Ambas se representan en color gris, y en un tamaño destacablemente superior al resto de términos representados. En la Figura 6, las palabras aparecen en un tamaño creciente en proporción a su frecuencia en tweets, y a su vez diferenciadas por colores. El color gris está destinado al término analizado, fresas, y los colores rosa, lila y verde a las palabras que van de mayor a menor número de apariciones por tweets.

En la Tabla 2, se muestra la lista de las 25 palabras más comunes en los tweets en los que los usuarios emplean los hashtags #fresas y #berries. Como observamos, coinciden en las posiciones más altas de la tabla aquellas palabras que aparecían en las Figuras 8 y 9 con un tamaño superior. En la columna ‘Frecuencia’ aparece el número de veces que el usuario ha empleado el término en cuestión, tomando como muestra la extracción de los últimos 1000 tweets publicados en los que se menciona ‘#fresas’ o ‘#berries’ a 4.09.2018. De esos 1000 tweets hemos eliminado aquellos que no habían sido publicados en alguna lengua europea. Coincide que en ambos términos, la mayoría de los términos que suelen acompañarlos son palabras relacionadas con la alimentación.

La Tabla 2 ha sido elaborada con RStudio a partir de los datos extraídos por la API Twitter y perfeccionada en Excel.

Con el fin de conocer algunas cifras relevantes de las cuentas de supermercados analizadas en nuestro estudio para saber su actuación en Twitter, hemos elaborado la Tabla 3. Hemos extraído y creado la tabla en RStudio a partir del código especificado en el Anexo I y la hemos exportado a Excel para adecuarla a nuestro estudio y hacerla más visual. En ella encontramos la columna ‘Cuenta’ con el usuario de los distintos supermercados y las columnas ‘Tweets’ (número de tweets publicados), ‘Seguidores’ (número de usuarios que siguen a esa cuenta), ‘Favoritos’ (número de tweets que la cuenta del supermercado ha marcado como favorito), ‘Seguidos’ (número de usuarios a los que sigue la cuenta del supermercado), ‘Nombre’ (nombre del supermercado que aparece en la descripción de su perfil), ‘Idioma’ (idioma en el que publica el usuario), ‘Listas’ (número de listas creadas por la cuenta del supermercado). En las cuatro primeras columnas hemos aplicado un filtro para representar en gráficas de barras horizontales la cifra correspondiente de cada celda, y así tener una visión más clara de estas cuatro columnas que reflejan algunos de las características más importantes de las cuentas analizadas, como son: los tweets publicados, el número de seguidores, número de seguidos y favoritos. Así, a simple vista podemos destacar cómo los supermercados de Reino Unido son los que más interactúan en la red, seguidos por aquellas cuentas en español de los supermercados analizados.

Tabla 2. Frecuencia de términos junto a las palabras ‘fresas’ y ‘berries’ en Twitter.
Fuente: Elaboración propio a partir de los datos obtenidos en Rstudio.

Nº	Palabra	Frecuencia	Nº	Palabra	Frecuencia
1	fresas	885	1	berries	750
2	chocolate	418	2	autumn	143
3	jugo	334	3	berry	115
4	desayuno	333	4	breakfast	69
5	bañados	332	5	jam	58
6	empleada	332	6	beautiful	54
7	huevitos	332	7	fruit	53
8	pancakes	332	8	garden	45
9	pedía	332	9	red	38
10	revueltos	332	10	healthy	36
11	naranja	330	11	strawberries	36
12	tecreesfresay	77	12	strawberry	36
13	fresa	61	13	summer	36
14	quiero	57	14	birds	36
15	crema	56	15	botanic	34
16	comer	33	16	botanicgarden	34
17	frutas	32	17	dawyck	34
18	más	32	18	scottishborders	34
19	yogur	29	19	food	34
20	pimientos	25	20	las	34
21	alimentos	24	21	hawthorn	33
22	verde	24	22	good	33
23	café	23	23	smoothie	33
24	huevos	23	24	rowan	31
25	sandía	23	25	goji	30

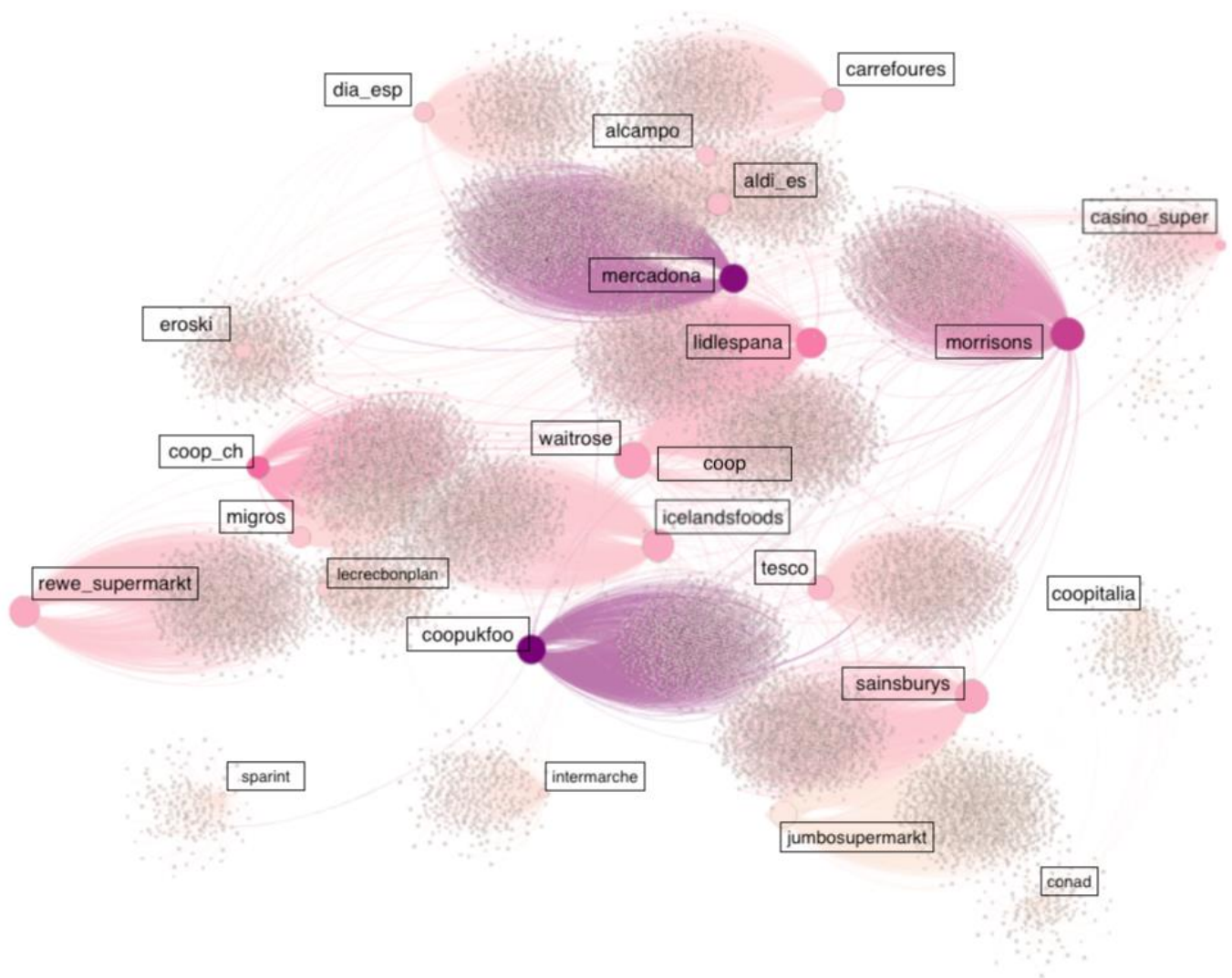
Tabla 3. Comparativa de los usuarios analizados. Fuente: Elaboración propia a partir de los datos obtenidos en RStudio.

Cuenta	Tweets	Seguidores	Favoritos	Seguidos	Nombre	Idioma	Listas
Aldi_es	5900	10509	3037	133	Aldi Supermercados	Español	76
Alcampo	20923	65349	6276	518	Alcampo	Español	387
CarrefourEs	45731	218481	42340	1657	Carrefour España	Español	906
casino_super	3129	15693	80	190	Casino Supermarchés	Francés	12
coopukfood	103889	129301	27081	13428	Co-op Food	Inglés	640
conad	3024	21598	3550	1802	Conad	Italiano	69
coop_ch	5963	19800	1137	436	Coop	Alemán	106
Coopitalia	8024	27252	3354	751	Coop Italia	Italiano	249
DIA_Esp	13835	57995	2852	2287	DIA España	Español	304
EROSKI	23999	53653	807	2639	EROSKI	Español	503
EurospinItalia	1572	5233	43	179	Eurospin Italia	Italiano	35
Icelandfoods	74140	150447	22236	2437	Iceland Foods	Inglés	427
intermarche	1442	13894	3754	137	Intermarché	Francés	153
JumboSuperm	48872	12035	2709	1253	Jumbo Supermarkten	Holandés	83
LeclercBonPlan	13283	20244	11106	465	Bons Plans E.Leclerc	Francés	139
lidlespana	17592	120561	8895	916	Lidl España	Español	453
Mercadona	72619	153856	850	43	Mercadona	Español	733
migros	19127	115253	9667	1179	Migros	Alemán	272
morrisons	300733	269614	41440	19464	Morrisons	Inglés	777
REWE_Superm	23701	24037	634	217	REWE	Alemán	154
sainsburys	724363	512938	4326	57125	Sainsbury's	Inglés	1936
SPARInt	2045	3549	1855	375	SPAR International	Inglés	48
Tesco	2062548	538441	4132	172021	Tesco	Inglés	1642
waitrose	163418	319244	17518	6721	Waitrose & Partners	Inglés	1479

A partir de la Figura 8 podemos observar la influencia de cada cuenta analizada en la red. Esto lo hemos representado a través de Gephi, combinando varias ‘Layouts’ para una mejor representación, y aplicando ‘Longitud media de camino’ en la pestaña de ‘Estadísticas’. La gráfica sigue una distribución ‘OpenOrd’, para tratar grandes grupos de datos y facilitar la creación de comunidades. Para una mejor visualización hemos ajustado la gráfica a partir de ‘Noverlap’ y ‘Ajuste de etiquetas’ para evitar tanto el solapamiento de los nodos como el de las etiquetas. Los nodos están representados por colores en función de sus valores de intermediación, y el tamaño en función de su grado. Para ello seleccionamos en ‘Apariencia’ el valor ‘Betweenness Centrality’, y se representarán los nodos en un grado de color de más claro a más oscuro dependiendo de su influencia en la red (basaremos la influencia en el número de veces que hace de puente entre un usuario y otro). Las aristas siguen la misma clasificación de color en función de

su peso (número de actores que intervienen en esa interacción). El tamaño del nodo viene reflejado por el ‘grado’ (número de relaciones del nodo), por lo que va creciendo su tamaño conforme aumentan sus interacciones.

Figura 8. Influencias de las cuentas en la red. Fuente: Elaboración propia a partir de Gephi.



En la Figura 9 hemos extraído los datos con ‘Twitter Streaming Importer’ de Gephi y representado la red de hashtags más frecuentes que aparecen justo a los hashtags #fresas #strawberries #berries y #frutosrojos. Para una visualización óptima los datos siguen una distribución ‘Force Atlas 2’ y hemos marcado ‘Evitar el solapamiento’ para que se vieran

claramente los nodos. En la pestaña de ‘Estadísticas’ hemos activado ‘Modularidad’ y como complemento, en ‘Apariencias’ hemos seleccionado tanto para los nodos como para las aristas la representación en base a las agrupaciones creadas en ‘Modularidad’ para así diferenciar las distintas comunidades por colores. Con ello conseguimos distinguir seis comunidades, basadas en qué hashtags suelen ir juntos cuando se publican los tweets que incluyen los hashtags analizados.

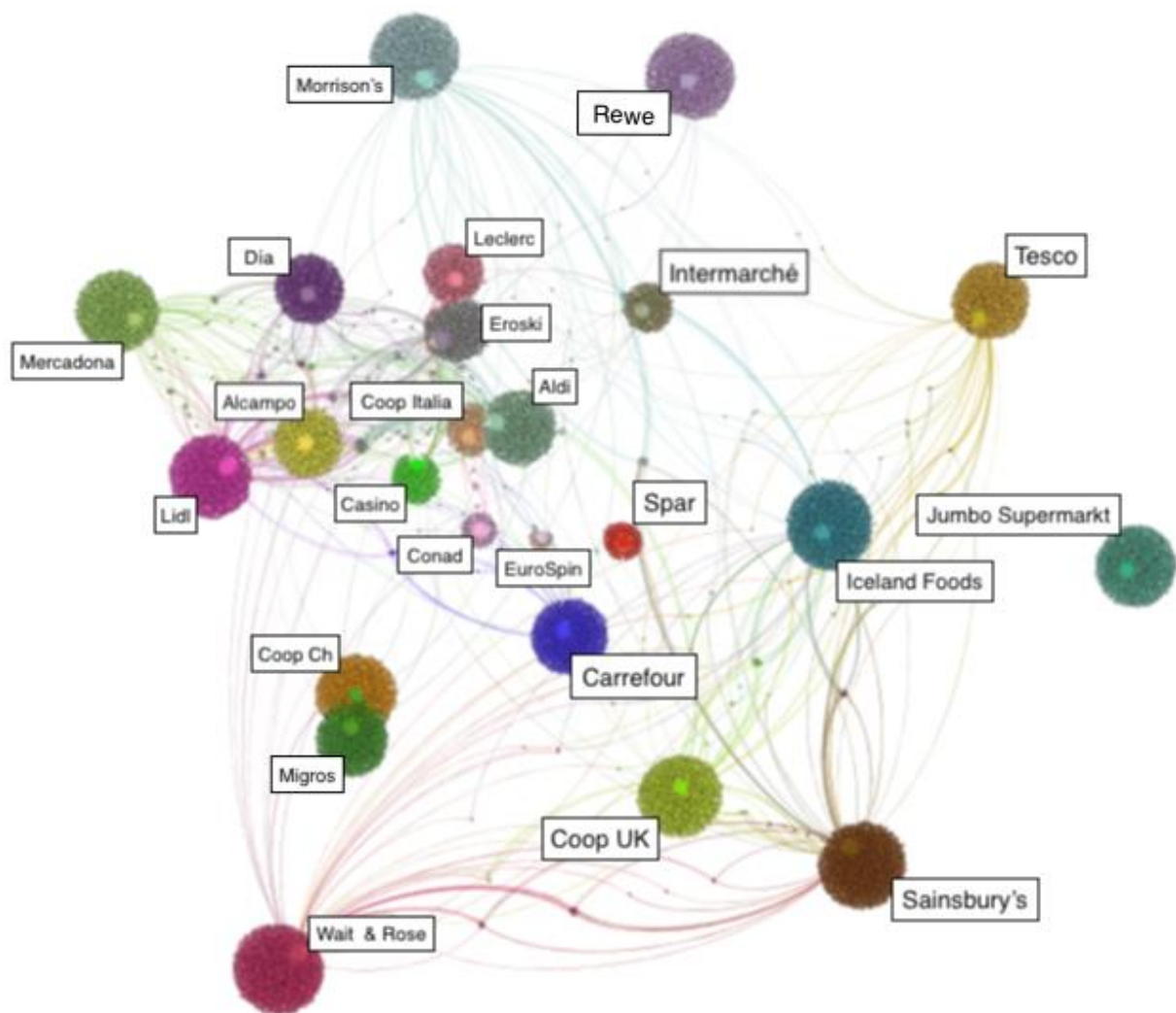
La comunidad predominante es la verde, aquella que guarda hashtags como #naturals, #foodismedicine, #herbal, #nochemicals... podríamos con ello identificar este grupo como una corriente de hashtags enfocados a la vida sin químicos, consumo de productos naturales y no procesado y una vida sana. El segundo grupo que más destaca en el gráfico es el azul, que enfoca los hashtags en un tema concreto que es la vida fitness y aparecen hashtags referidos a la comida en concreto. A continuación, los grupos naranja y rosa representan una tendencia en la artesanía y joyería que es la representación de frutas, en este caso fresas y frutos rojos, en cristal, y la dieta vegana, respectivamente. Por último, la comunidad gris engloba términos como #medioambiente, #sostenible y #natural.

[illegible]

El objetivo de la Figura 10 es representar las relaciones de los supermercados con sus clientes en Twitter. Hemos extraído los datos en RStudio e importado este archivo en formato .cvs a Gephi, obteniendo el nivel de respuesta de los supermercados a sus clientes. Cada nodo y arista está representado de un color, en función al nodo principal con el que más interactúe. Como nodos principales, aquellos de mayor tamaño, tenemos los 24 supermercados del estudio. Lo primero que destaca al observar la gráfica es que aquellos supermercados con más nodos a su alrededor son los supermercados ingleses,

por lo tanto, son los que más interactúan con sus clientes mediante respuestas en forma de tweet. A su vez, los supermercados con cuentas españolas Mercadona, Lidl, Día y Carrefour, y el alemán Rewe, también tienen un buen nivel de respuesta a sus usuarios. Si observamos las aristas, nos muestra que las cuentas de supermercados ingleses son las que más usuarios comparten entre sí, al igual que sucede con el círculo de cuentas españolas.

Figura 10. Red de interacción en Twitter de los principales supermercados europeos.
Fuente: Elaboración propia a partir de Gephi.



Para llegar a esta representación de la Figura 10. Hemos aplicado una distribución 'Force Atlas 2' y hemos marcado 'Evitar el solapamiento' para que se vieran claramente los nodos. En la pestaña de 'Estadísticas' hemos activado 'Modularidad' y como complemento, en 'Apariencias' hemos seleccionado tanto para los nodos como para las

aristas la representación en base a las agrupaciones creadas en ‘Modularidad’. Para remarcar aquellos nodos con más peso y relevancia en este estudio hemos hecho una partición de los mismos en la pestaña ‘Apariencia’, seleccionando una clasificación por frecuencia y estableciendo un intervalo de tamaño para ello entre 1 y 100. Con esto conseguimos destacar en la gráfica a los nodos correspondientes a los supermercados, ya que son los que tienen mayor frecuencia dado su alto número de interacciones.

Tabla 4. Número de interacciones de los usuarios analizados en torno a los términos estudiados. Fuente: Elaboración propia a partir de los datos obtenidos en Gephi.

Id	Label	Interval	frequency ▼	type
casino_super	casino_super		3184	source; target
eroski	eroski		2208	source; target
sparint	sparint		2053	source; target
dia_esp	dia_esp		2030	source; target
alcampo	alcampo		2025	source; target
lidlspana	lidlspana		2020	source; target
migros	migros		2020	source; target
leclercbonplan	leclercbonplan		2018	source; target
coopukfood	coopukfood		2012	source; target
aldi_es	aldi_es		2012	source; target
coop_ch	coop_ch		2008	source; target
morrisons	morrisons		2004	source; target
rewe_supermarkt	rewe_supermarkt		2004	source; target
icelandfoods	icelandfoods		2002	source; target
conad	conad		2002	source; target
jumbosupermarkt	jumbosupermarkt		2001	source; target
carrefour.es	carrefour.es		2001	source; target
waitrose	waitrose		2001	source; target
tesco	tesco		2000	source; target
mercadona	mercadona		2000	source; target
sainsburys	sainsburys		2000	source; target
coopitalia	coopitalia		1999	source; target
eurospitalia	eurospitalia		1567	source; target
intermarche	intermarche		1494	source; target

Para enfocar más las interacciones de los supermercados analizados en torno a la difusión en Twitter de los productos de interés (fresas, strawberries, berries y frutosrojos) hemos elaborado la Tabla 4, donde aparecen representadas las interacciones de las distintas cuentas de supermercados en las que emplean estos términos. Para ello hemos empleado la herramienta ‘Twitter Streaming Importer’ de Gephi, acotando la búsqueda a esos términos y esos usuarios, a la vez que hemos delimitado con coordenadas geográficas la

extracción de tweets al continente europeo (hemos seleccionado todos los tweets con estas características que Gephi ha extraído en un periodo de 30 minutos).

Con esta misma extracción que hemos realizado para la Tabla 4 hemos obtenido los datos representados en la Tabla 5. A partir de la nacionalidad de los usuarios reflejada en la columna ‘Location’ del laboratorio de datos, hemos tomado una muestra aleatoria de 344 usuarios que interactúan empleando los términos #fresas, #strawberries, #berries y #frutosrojos en sus publicaciones en Twitter. Con estos datos hemos creado en Excel la tabla que se muestra a continuación.

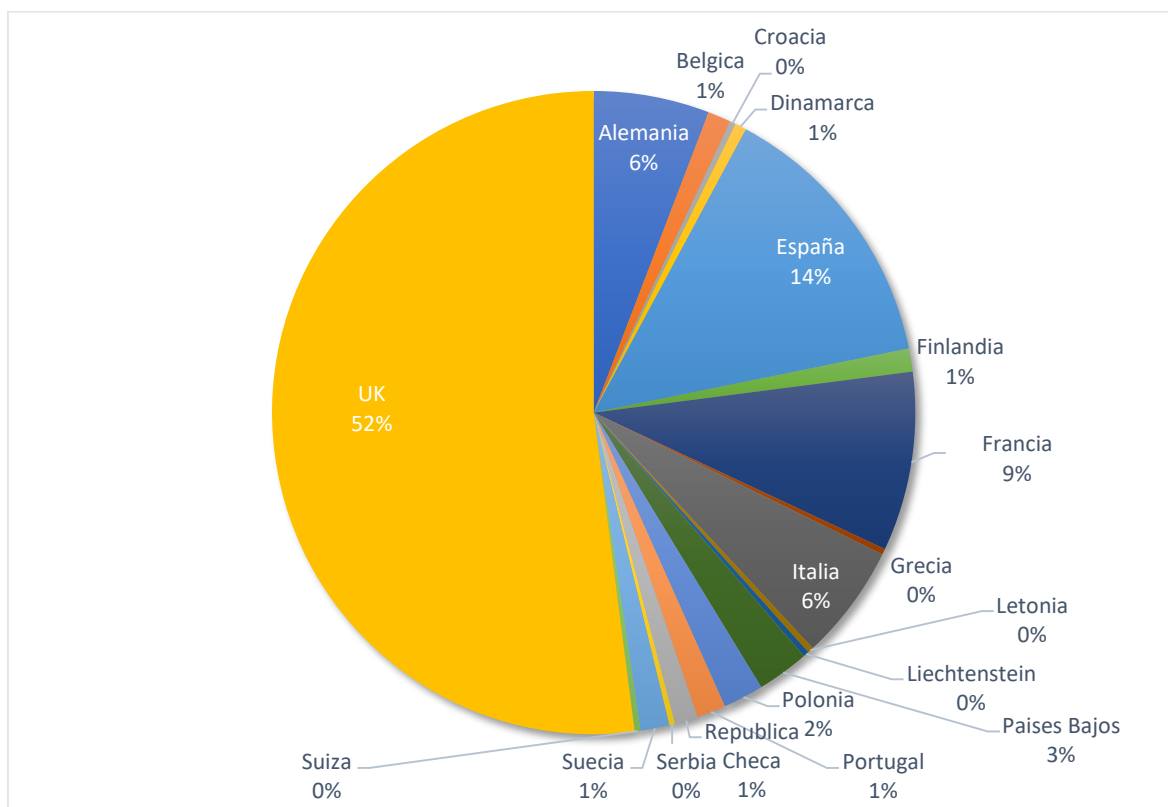
Tabla 5. Recuento de la muestra tomada de la clasificación de los usuarios clasificados por países. Fuente: Elaboración propia a partir de los datos obtenidos en Gephi.

Alemania	20
Belgica	4
Croacia	1
Dinamarca	2
España	48
Finlandia	4
Francia	31
Grecia	1
Italia	20
Letonia	1
Liechtenstein	1
Países Bajos	9
Polonia	7
Portugal	5
Republica Checa	4
Serbia	1
Suecia	5
Suiza	1
UK	179
TOTAL	344

Para conseguir una visualización más representativa de la Tabla 5 desarrollamos la Figura 11. En ella se aprecia, en porcentajes, las distintas nacionalidades de los usuarios que interactúan en Twitter con los términos analizados en este estudio, expresando cada país de las seleccionadas nacionalidades con un color, tamaño y porcentaje determinado en la gráfica en proporción a los datos expresados en la Tabla 5. Cabe destacar, una vez más,

la destacable presencia de la población inglesa en cuanto a su frecuente actividad en Twitter con los productos de estudio

Figura 11. Proporción de los usuarios tomados como muestra de la clasificación de los usuarios por países. Fuente: Elaboración propia a partir de los datos obtenidos en Gephi.



5. Discusión

En una sociedad tan inmersa en las redes sociales como es la sociedad actual, resulta cada vez más fácil acceder a información de cualquier lugar del mundo y contactar con cualquier persona independientemente de la distancia física entre ambos puntos. Esto, junto con los efectos de la globalización, ha permitido la realización de todo tipo de actividades a través de las distintas plataformas online, como puede ser el comercio. Ante esta variación, tanto en el patrón de consumo como en las vías de comercio, los empresarios se han visto obligados a evolucionar junto a este entorno tecnológico para no quedar desfasados ni obsoletos. Todo esto conlleva una larga lista de consecuencias para el empresario, como puede ser el aumento del número de competidores.

En nuestro estudio hemos realizado una investigación del mercado de las frutas del bosque, haciendo hincapié en las fresas, aplicando nuevas técnicas propiciadas por esta revolución de la red, como son la Minería de Texto y la Minería de Datos. A través de Twitter, gracias a las propias interacciones de los usuarios con los supermercados analizados y las de las propias cadenas en sus redes, hemos podido observar algunas características del mercado en el que estamos interesados.

Examinando los términos que aparecen con más frecuencia en aquellos tweets en los que se emplean los hashtags #berries o #fresas (Tabla 2), observamos que estas palabras suelen ir acompañadas de otras del mismo campo semántico, como es el gastronómico (desayuno, comer, alimentos, sandía...) y por palabras que denotan la presencia de una tendencia cada vez más presente en el estilo de vida de las personas, como son aquellas palabras relacionadas con llevar una vida sana, una dieta equilibrada y una vida fitness (healthy, smoothie, fruits...), siendo más frecuente emplear términos en inglés para aquellas publicaciones referentes a una dieta equilibrada.

Reforzando la teoría de que los términos tanto fresas, como frutos rojos, y sus homólogos en inglés (berries y strawberries) cuando más presencia tienen en Twitter es cuando lo asociamos al estilo de vida fitness y saludable, aumentando el consumo de productos naturales y sin químicos (Figura 9). En esta imagen se representan aquellos hashtags más comunes con los que se acompañan los términos #fresas #strawberries #berries y #frutosrojos en los tweets. Han sido agrupados en función a los hashtags que suelen aparecer juntos en un mismo tweet. De todos los grupos creados es evidente que aquel que más suele usarse unido a los términos estudiados son aquellos de color verde, que contiene palabras como #healthyeating, #healthylifestyle, #nochemical, #natural, #digestion... Todos ellos relacionados a su vez con el concepto de una alimentación saludable, basada en productos fresco, naturales y sin químicos. La red azul, la segunda más empleada por los usuarios, está más relacionada con el concepto fitness, que engloba tanto a la alimentación como a un estilo de vida activo y ligado al deporte. En este gráfico también destaca la escasa presencia de términos en español (solo aparecen los hashtags #mediambiente, #natural y #sostenible en un subgrupo poco relevante), y esto es debido al predominio del inglés frente al español en las Redes Sociales; con esto no damos a entender que hay una minoría de hispano hablantes en Twitter, sino que solemos adoptar el inglés a la hora de expresarnos en las redes independientemente de la lengua materna,

sobre todo en el momento de emplear hashtags, ya que el inglés es la lengua más universal y mediante hashtag en inglés llegamos a un público más amplio.

Conocer esto es importante para tenerlo en cuenta en el momento de elegir qué canal de distribución, en este caso supermercados, refleja mejor estas tendencias que están en continuo crecimiento. En la Tabla 3 podemos ver las principales cifras de las cuentas de Twitter de los 24 principales supermercados europeos. Tesco, Sainsbury's y Morrison's son los supermercados más seguidos en la red, y con más publicaciones realizadas. Estas tres cadenas tienen su sede social en Reino Unido y llevan sus cuentas en Twitter en inglés. Tesco, además de ser la multinacional de venta al por menor con más presencia en Twitter y una mayor actividad en su cuenta, es la tercera tienda minorista más grande del mundo, y está presente en 12 países de Asia y Europa, por lo que a través de su cuenta se proyecta a un público muy amplio y variado. Cabe destacar que, dentro de los 5 principales, en lo que respecta a actividad en Twitter basándonos en los tweets publicados y seguidores, aparecen la cadena española Mercadona y la cuenta de Carrefour España como preferida por los usuarios. A su vez, es importante destacar la información representada en la Figura 8 para conocer cómo actúan los clientes en Twitter. Con la Figura 8 demostramos como los supermercados Mercadona, Coop UK y Morrison's son los más influyente, de todas las cuentas analizadas, en Twitter. Con esto conseguimos reflejar que estas son las cuentas que más actúan como puente entre los seguidores de los distintos supermercados. Si centramos nuestra atención en las aristas y en la distribución de los distintos nodos en la gráfica, percibimos un claro grupo de nodos con una mayor cercanía entre sí. Este grupo lo conforman todas las cuentas españolas analizadas, resultando Mercadona, en su mayoría, la cuenta que hace de conexión entre los seguidores de los distintos seguidores de este grupo de supermercados.

En la Figura 10 se representa el nivel de respuesta que ofrecen las cuentas de Twitter de los supermercados a aquellos usuarios que se ponen en contacto con ellos vía tweets. Vemos que los supermercados que tienen más usuarios puente (usuarios en común, que interactúan con ambas cuentas) son: Mercadona, Lidl, Día, Eroski, Alcampo y Aldi por un lado (cuentas españolas) y Morrison's, Sainsbury's, Tesco, Icelanfoods y Waitrose por otro lado (cadenas de Reino Unido). El tamaño de los distintos nodos del gráfico depende del número de interacciones de respuesta registradas por cada cuenta de supermercado. Todas ellas tienen un tamaño similar, a excepción de las cuentas de Spar, Conad,

Eurospin, Coop Italia, Casino e Intermarché que presentan un tamaño más reducido al tener un nivel menor de respuesta. Si observamos el nodo de Jumbo Supermarkt apreciamos su distanciamiento en el gráfico del resto de supermercados, y la ausencia de usuarios puente. Esto es debido a que es la única cuenta que interactúa en holandés, quedando sus usuarios aislados del resto.

Si analizamos de manera más detallada el número de veces que emplean estos términos las cuentas de supermercado analizadas, comprobamos en la Tabla 4 que los supermercados Casino (3184), Eroski (2208), Spar (2053), Día (2030) y Alcampo (2025) son los que más emplean estos vocablos.

Por último, a través de la opción ‘Twitter Streaming Importer’ de Gephi hemos realizado la extracción de datos reflejada en la Tabla 5. Acotando los usuarios a observar a las cuentas de supermercados analizadas en este estudio y sus seguidores y la zona geográfica a Europa hemos recabado la lista de países desde donde interactúan los usuarios de Twitter con las cuentas de los supermercados empleando los términos: fresa, fresas, strawberry, strawberries, berry y berries. Con un total de 20081 nodos(usuarios) y 20647 aristas (conexiones entre usuarios) se confirma que la mayoría de usuarios que interactúan con estos supermercados con relación a estos términos se encuentran ubicados en Reino Unido (52%), España (14%), Francia (9%), Italia (6%) y Alemania (6%). Estas proporciones aparecen representadas en un gráfico circular de porcentajes en la Figura 11. Para alcanzar estas cifras hemos tomando una muestra aleatoria de 344 usuarios.

6. Conclusiones

1. Para hacerse visible en internet, en especial en las Redes Sociales, es recomendable expresarse en inglés, ya que con ello llegarás a más gente al ser este el idioma que predomina entre los usuarios. A su vez, seguir hashtags e interactuar con ellos potenciará la difusión de la información que los acompañe, y tendrá una proyección e impacto mayor en la red.
2. La tendencia que lidera en Twitter que potencia el consumo de las berries y frutos rojos es el estilo de vida saludable y llevar una vida fitness. Para que las publicaciones que contengan promociones o información de estos frutos aumenten el número de impresiones por tweet y llegue a más usuarios se recomienda incluir un contenido que promueva este estilo de vida saludable, natural y sin químicos. Los hashtags #healthyeating, #healthylifestyle, #nochemical, #natural, #fitness... Son algunos de los más populares entre las publicaciones de este movimiento, que a su vez son algunos de los términos más frecuentes utilizados junto a los hashtags #berries y #fresas.
3. El buen uso de las redes sociales para el negocio es esencial, siendo una de las vías de comunicación con el cliente más presente en la actualidad. Es por ello, por lo que el buen uso y mantenimiento de las cuentas de los supermercados en Twitter es algo presente en su estrategia empresarial. Aquellas cuentas analizadas con un mayor número de tweets publicados son, a su vez, las que más seguidores tienen (Tesco, Sainsbury's, Morrison's y Waitrose). Por lo tanto, cuentan con cierta ventaja para comunicarse y escuchar a sus clientes al tener esa cercanía en la red. No obstante, en referencia al sector observado, no coinciden los supermercados con más publicaciones que incluyen las palabras analizadas en su cuenta de Twitter (Casino, Eroski, Spar y Alcampo), con las cuentas de los supermercados que más seguidores tienen que interactúen con esos mismos términos (Tesco, Sainsbury's, Morrison's y Waitrose).
4. Como hemos expresado en el punto anterior, aquellos supermercados con más peso en Twitter, en función de su número de tweets y seguidores, son los supermercados Tesco, Sainsbury's, Morrison's y Waitrose. A continuación, las

cuentas españolas de Carrefour España, Mercadona y Lidl España cuentan con un gran número de seguidores, encontrándose entre las diez primeras cuentas analizadas con más followers. Por el contrario, las cuentas de los supermercados SPAR, Eurospin y Aldi España cuenta con el menor número de seguidores. En todos los casos mencionados coincide el tener un número superior de seguidores con un número elevado de tweets publicados, y viceversa.

5. Reino Unido cuenta con un gran volumen de usuarios en Twitter que interactúa con los supermercados más populares de Europa, siendo el país donde más usuarios promueven en redes los términos berry y strawberry y sus derivados, tanto mediante interacciones con los supermercados, como de forma independiente. Esto, unido a que las cadenas Tesco, Sainsbury's y Morrison's (con sede social en Reino Unido) cuentan con un gran volumen de negocio y presencia en redes, hace muy atractivo los lazos empresariales entre los minoristas españoles y estas franquicias.
6. Gracias a RStudio y Gephi podemos realizar un estudio de mercado a partir de información de las redes sociales.

7. Referencias

1. Andrews, R. L., & Currim, I. S. (2004). Behavioural differences between consumers attracted to shopping online versus traditional supermarkets: implications for enterprise design and marketing strategy. *International Journal of Internet Marketing and Advertising*, 1(1), 38-61.
2. Barberá, P., & Rivero, G. (2012). ¿Un tweet, un voto? Desigualdad en la discusión política en Twitter. En I Congreso Internacional en Comunicación Política y Estrategias de Campaña.
3. Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks.
4. Brust, A. V. (2018). *RPubs*. Obtenido de Análisis de social media: captura y mapeo de datos geo-referenciados.
5. Chamlerwat, W., Bhattarakosol, P., Rungkasiri, T., & Haruechaiyasak, C. (2012). Discovering Consumer Insight from Twitter via Sentiment Analysis. *J. UCS*, 18(8), 973-992.
6. Degeratu, A. M., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing*, 17(1), 55-78.
7. EAE Business School. (2018). La situación de la gran distribución en el mercado de consumo.
8. Eurofresh Distribution. (2016). eurofresh-distribution. Obtenido de Record sales for European berry market: https://www.eurofresh-distribution.com/news/record-sales-european-berry-market#.V_ufTR8s0G0.twitter
9. Global, D. (2016). Global Powers of Retailing 2016. Navigating the new digital divide. Deloitte Touche Tohmatsu Limited.
10. Greene, D. & Cunningham, P. (2013). Producing a Unified Graph Representation from Multiple Social Network Views. *Proc. ACM Web Science'13*.

11. How to use Twitter analytics. (2018). Retrieved from <https://business.twitter.com/en/analytics.html>
12. IAB Spain. (2018). *Estudio anual de Redes Sociales 2018*.
13. Kantar WorldPanel. (2018). *Tendencias en la distribución 2018*. Retrieved from <https://www.kantarworldpanel.com/es/Noticias/el-gran-consumo-crece-un-1porciento-en-los-8-primeros-meses>
14. Martínez, A. C. (2013). Estrategias empresariales en la Web 2.0. Las redes sociales online. Editorial Club Universitario.
15. Måster, T. I. S. A. (2013). The R project for statistical computing.
16. Reynolds, D. (2018). El Mercado de Berries en UK. Retrieved from http://congresofrutosrojos.com/wp-content/uploads/2017/07/Drew-presentationv2-Huelva-Congress-June-2017_ES.pdf
17. RStudio. (2018). Retrieved from <https://es.wikipedia.org/wiki/RStudio>
18. Sánchez López, A. (2017). Diseño de una herramienta para la caracterización financiera de usuarios de redes sociales.
19. Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter. *Journal of interactive marketing*, 26(2), 102-113
20. Solo Marketing. (2013). SoloMarketing. Obtenido de Uso de las redes sociales en la investigación de mercados: <https://solomarketing.es/uso-de-redes-sociales-en-la-investigacion-de-mercados/>
21. Sook Kwon, E., Kim, E., Sung, Y., & Yun Yoo, C. (2014). Brand followers: Consumer motivation and attitude towards brand communications on Twitter. *International Journal of Advertising*, 33(4), 657-680.
22. Vidal, L., Ares, G., Machín, L., & Jaeger, S. R. (2015). Using Twitter data for food-related consumer research: A case study on “what people say when tweeting about different eating situations”. *Food Quality and Preference*, 45, 58-69.

8. Anexos

➤ Anexo I: Código empleado en R.

```
#1. EXTRACCIÓN DE DATOS EN TWITTER

#Instalamos y cargamos las librerías necesarias para la extracción en
Twitter

install.packages("twitterR")
install.packages("dplyr")
install.packages("tidyr")
library(twitterR)

rm(list = ls())
setwd("~/Desktop/TFM")

#Cargamos nuestras credenciales obtenidos de API Twitter y
autenticamos.

api_key <- "XXXXXXX"
api_secret <- "XXXXXXXXXXXX"
access_token <- "XXXXXXXXXXXX"
access_token_secret <- "XXXXXXX"

setup_twitter_oauth(api_key, api_secret, access_token,
access_token_secret)

#Extraemos los ultimos 2000 tweets de las cuentas seleccionadas y
creamos un dataframe

listado_supermercados<-
list('Aldi_es','alcampo','CarrefourEs','casino_super','coopukfood','co
nad','coop_ch','Coopitalia','DIA_Esp','EROSKI','EurospinItalia','Icela
ndfoods','intermarche','JumboSupermarkt','LeclercBonPlan','lidlspana'
,'Mercadona','migros','morrison','REWE_Supermarkt','sainsburys','SPAR
Int','Tesco','waitrose')
listado_supermercados_espanol<-
list('Aldi_es','alcampo','CarrefourEs','DIA_Esp','EROSKI','lidlspana'
,'Mercadona')

agrupadosTFM_df<-data.frame()
for (i in listado_supermercados){
  tweets<-twListToDF(userTimeline(i,n=2000,includeRts = T))
  agrupadosTFM_df<-rbind(agrupadosTFM_df,tweets)
```

```

    print(paste('Se extrajeron satisfactoriamente los tweets de la
cuenta ',i))
}

agrupadosTFM_espanol_df<-data.frame()
for (i in listado_supermercados_espanol){
    tweetse<-twListToDF(userTimeline(i,n=2000,includeRts = T))
    agrupadosTFM_espanol_df<-rbind(agrupadosTFM_espanol_df,tweetse)
    print(paste('Se extrajeron satisfactoriamente los tweets de la
cuenta ',i))
}

#Guardamos los archivos como csv.
save(agrupadosTFM_df, file="agrupadosTFM_df.RData")
save(agrupadosTFM_espanol_df, file="agrupadosTFM_espanol_df.RData")

repliesTFM_df <-
agrupadosTFM_df[complete.cases(agrupadosTFM_df[, 'replyToSN']),]
repliesTFM_df2 <-
data.frame(source=tolower(agrupadosTFM_df$screenName),
target=tolower(agrupadosTFM_df$replyToSN), type="Directed",weight=1)
tweetsTFM_df <-
agrupadosTFM_df[!complete.cases(agrupadosTFM_df[, 'replyToSN']) &
agrupadosTFM_df$isRetweet == 'FALSE',]
tweetsTFM_espanol_df <-
agrupadosTFM_espanol_df[!complete.cases(agrupadosTFM_df[, 'replyToSN'])
& agrupadosTFM_df$isRetweet == 'FALSE',]

write.csv(repliesTFM_df2, "repliesTFM_df.csv")
write.csv(tweetsTFM_df, "tweetsTFM_df.csv")
write.csv(tweetsTFM_espanol_df, "tweetsTFM_espanol_df.csv")

#2. COMPARACIÓN CUENTAS TWITTER

#Instalamos y cargamos las librerías que vamos a utilizar.
library(stringr) #tratar las string (cadenas de palabras)
library(wordcloud) #representar nube de palabras

#Extraemos la información de las cuentas a analizar y creamos una
tabla

comparacion_supermercados<-
lookupUsers(c("Aldi_es", "Alcampo", "CarrefourEs", "casino_super", "coopuk

```

```

food", "conad", "coop_ch", "Coopitalia", "DIA_Esp", "EROSKI", "EurospinItali
a", "Icelandfoods", "intermarche", "JumboSupermarkt", "LeclercBonPlan", "li
dlespana", "Mercadona", "migros", "morrison", "REWE_Supermarkt", "sainsbur
ys", "SPARInt", "Tesco", "waitrose"))
tabla_comparacionsupermercados<-twListToDF(comparacion_supermercados)

#Exportamos la tabla en formato csv y excel.
write.csv(tabla_comparacionsupermercados,"tabla_supermercados.csv")
write.table(tabla_comparacionsupermercados, file =
"tabla_comparacionsupermercados.txt", sep = "\t", eol = "\n", dec =
".", row.names = TRUE, col.names = TRUE)

#3. ANÁLISIS POR PALABRAS - NUBE DE TÉRMINOS

#Instalamos y cargamos las librerías que vamos a utilizar.

require(vosonSML)
require(magrittr)
library(NLP) #procesamiento del lenguaje natural
library(RColorBrewer) #colores gráficos
library(tm) #funciones de text mining
library(wordcloud) #representar nube de palabras

#ANÁLISIS ESPAÑOL- TÉRMINO FRESAS

#Creamos un archivo que recoga la información extraída en relación al
término fresa.
myTwitterData_fresas <- Authenticate("twitter",
                                     apiKey=api_key,
                                     apiSecret=api_secret,
                                     accessToken=access_token ,
                                     accessTokenSecret=access_token_secret)

%>%
  Collect(searchTerm="fresas", numTweets=1000, writeToFile=FALSE,
verbose=TRUE) #hashtag y n°tuits.

#Extraemos el texto de los tweets y comenzamos a filtrar y depurar el
texto

myTwitterData_fresas$text <- iconv(myTwitterData_fresas$text, to =
'utf-8-mac') #errores codificacion text.
txt_fresas = myTwitterData_fresas$text

#Eliminamos retweets
txtclean_fresas = gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", txt_fresas)

```

```

#Eliminamos menciones(@)
txtclean_fresas= gsub("@\\w+", "", txtclean_fresas)

#Eliminamos símbolos de puntuación
txtclean_fresas= gsub("[[:punct:]]", "", txtclean_fresas)

#Eliminamos números
txtclean_fresas= gsub("[[:digit:]]", "", txtclean_fresas)

#Eliminamos los links
txtclean_fresas = gsub("http\\w+", "", txtclean_fresas)

#Creamos un corpus (colección de documentos de texto)
corpus_fresas = Corpus(VectorSource(txtclean_fresas))

#Convertimos a minúsculas
corpus_fresas = tm_map(corpus_fresas, tolower)

#Eliminamos palabras vacías (en español)
corpus_fresas = tm_map(corpus_fresas, removeWords,
c(stopwords("spanish"))))

#Cargamos archivo de palabras vacías personalizada y convertimos a
ASCII
sw_fresas <- readLines("~/Desktop/stopwords.es.txt",encoding="UTF-8")
sw_fresas = iconv(sw_fresas, to="ASCII//TRANSLIT")

#Eliminamos palabras vacías personalizadas
corpus_fresas = tm_map(corpus_fresas, removeWords, sw_fresas)

#Eliminamos espacios en blanco extras
corpus_fresas = tm_map(corpus_fresas, stripWhitespace)

#Creamos una matriz de términos
tdm_fresas <- TermDocumentMatrix(corpus_fresas)

#Convertimos a matriz
m_fresas = as.matrix(tdm_fresas)

#Conteo de palabras en orden decreciente
wf_fresas <- sort(rowSums(m_fresas),decreasing=TRUE)

#Creamos un data frame con las palabras y sus frecuencias
dm_fresas <- data.frame(word = names(wf_fresas), freq=wf_fresas)

```

```

#Creamos una tabla con las frecuencias y la exportamos
write.table(dm_fresas, file = "tabla_frecuencias_fresas.txt", sep =
"\t", eol = "\n", dec = ".", row.names = TRUE, col.names = TRUE)

#Representación de nube de términos(fresa)
wordcloud(dm_fresas$word, dm_fresas$freq, random.order=FALSE,
colors=brewer.pal(8, "Dark2"))

# ANALISIS INGLÉS-TÉRMINO BERRIES

#Creamos un archivo que recoga la información extraída en relación al
término fresa.
myTwitterData_berries <- Authenticate("twitter",
                                     apiKey=api_key,
                                     apiSecret=api_secret,
                                     accessToken=access_token ,

accessTokenSecret=access_token_secret) %>%
  Collect(searchTerm="#berries", numTweets=1000, writeToFile=FALSE,
verbose=TRUE) #hashtag y n°tuits.

#Extraemos el texto de los tweets y comenzamos a filtrar y depurar el
texto
myTwitterData_berries$text <- iconv(myTwitterData_berries$text, to =
'utf-8-mac') #errores codificación text.
txt_berries = myTwitterData_berries$text

#Eliminamos retweets
txtclean_berries = gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "",
txt_berries)

#Eliminamos menciones(@)
txtclean_berries = gsub("@\\w+", "", txtclean_berries)

#Eliminamos símbolos de puntuación
txtclean_berries = gsub("[[:punct:]]", "", txtclean_berries)

#Eliminamos números
txtclean_berries = gsub("[[:digit:]]", "", txtclean_berries)

#Eliminamos los links
txtclean_berries = gsub("http\\w+", "", txtclean_berries)

#Creamos un corpus (colección de documentos de texto)
corpus_berries = Corpus(VectorSource(txtclean_berries))

```

```

#Convertimos a minúsculas
corpus_berries = tm_map(corpus_berries, tolower)

#Eliminamos palabras vacías (en inglés)
corpus_berries = tm_map(corpus_berries, removeWords,
c(stopwords("English")))

#Cargamos archivo de palabras vacías personalizada y convertimos a
ASCII
sw_berries <- readLines("~/Desktop/stopwords_en.txt",encoding="UTF-8")
sw_berries = iconv(sw_berries, to="ASCII//TRANSLIT")

#Eliminamos palabras vacías personalizada
corpus_berries = tm_map(corpus_berries, removeWords, sw_berries)

#Eliminamos espacios en blanco extras
corpus_berries = tm_map(corpus_berries, stripWhitespace)

#Creamos una matriz de términos
tdm_berries <- TermDocumentMatrix(corpus_berries)

#Convertimos a matriz
m_berries = as.matrix(tdm_berries)

#Conteo de palabras en orden decreciente
wf_berries <- sort(rowSums(m_berries),decreasing=TRUE)

#Creamos un data frame con las palabras y sus frecuencias
dm_berries <- data.frame(word = names(wf_berries), freq=wf_berries)

#Creamos una tabla con las frecuencias y la exportamos
write.table(dm_berries, file = "tabla_frecuencias_berries.txt", sep =
"\t", eol = "\n", dec = ".", row.names = TRUE, col.names = TRUE)

#Representación de nube de términos(berries)
wordcloud(dm_berries$word, dm_berries$freq, random.order=FALSE,
colors=brewer.pal(8, "Dark2"))

#ANALISIS SUPERMERCADOS ESPAÑOL

#Extraemos el texto de los tweets y comenzamos a filtrar y depurar el
texto
txt_espanol = tweetsTFM_espanol_df$text

```

```

#Eliminamos retweets
txtclean_espanol = gsub("(RT|via) ((?:\\b\\W*@\\w+)+)", "",
txt_espanol)

#Eliminamos menciones(@)
txtclean_espanol = gsub("@\\w+", "", txtclean_espanol)

#Eliminamos símbolos de puntuación
txtclean_espanol = gsub("[[:punct:]]", "", txtclean_espanol)

#Eliminamos números
txtclean_espanol = gsub("[[:digit:]]", "", txtclean_espanol)

#Eliminamos links
txtclean_espanol = gsub("http\\w+", "", txtclean_espanol)

#Creamos un corpus (colección de documentos de texto)
corpus_espanol = Corpus(VectorSource(txtclean_espanol))

#Convertimos a minúsculas
corpus_espanol = tm_map(corpus_espanol, tolower)

#Eliminamos palabras vacías (en espanyol)
corpus_espanol = tm_map(corpus_espanol, removeWords,
c(stopwords("spanish")))

#Cargamos archivo de palabras vacías personalizada y convertimos a
ASCII
sw_espanol <- readLines("~/Desktop/stopwords.es.txt",encoding="UTF-8")
sw_espanol = iconv(sw, to="ASCII//TRANSLIT")

#Eliminamos palabras vacías personalizada
corpus_espanol = tm_map(corpus_espanol, removeWords, sw_espanol)

#Eliminamos espacios en blanco extras
corpus_espanol = tm_map(corpus_espanol, stripWhitespace)

#Creamos una matriz de términos
tdm_espanol <- TermDocumentMatrix(corpus_espanol)

#Convertimos a matriz
m_espanol = as.matrix(tdm_espanol)

#Conteo de palabras en orden decreciente
wf_espanol <- sort(rowSums(m_espanol),decreasing=TRUE)

#Creamos un data frame con las palabras y sus frecuencias
dm_espanol <- data.frame(word = names(wf_espanol), freq=wf_espanol)

#Creamos una tabla con las frecuencias y la exportamos
write.table(dm_espanol, file =
"tabla_frecuencias_supermercados_espanol.txt", sep = "\t", eol = "\n",
dec = ".", row.names = TRUE, col.names = TRUE)

```

```
#Representación de nube de términos  
wordcloud(dm_espanol$word, dm_espanol$freq, random.order=FALSE,  
colors=brewer.pal(8, "Dark2"))
```