



Universidad
Zaragoza

Trabajo Fin de Grado

El BIG DATA y nuevas metodologías en el entorno financiero: un estudio del nuevo perfil del cliente a través del uso de tics.

Autor/es

Marta Revuelta

Director/es

Pilar Olave

Facultad de Economía y Empresa
2019

Resumen

Desde principios del siglo XXI, las tecnologías de la comunicación y la información (TICs) se han impuesto como elemento clave en la sociedad y en las empresas como un factor en la digitalización y expansión de la mayoría de negocios. Con las nuevas formas de captar información a través de los medios comunicación, la información disponible para las empresas ha crecido exponencialmente. Por ello, la incorporación de las grandes bases de datos y el análisis de contenidos son dos grandes retos empresariales de este siglo. En las dos últimas décadas, se han desarrollado determinadas técnicas clave dentro de las áreas de la tecnología de la información y comunicación, que nos han proporcionado nuevas herramientas de análisis de contenidos a través del apoyo de técnicas usadas en Big Data.

Este proyecto pretende ahondar en la utilización de análisis de datos para determinar perfiles de clientes usuarios de TICs, y posteriormente determinar estrategias de participación en los nuevos medios de comunicación a través del análisis de grandes bases de datos de información online.

Los resultados muestran que el acceso a internet ha cambiado en los últimos años: aunque se conecta un número similar de personas, el dispositivo más utilizado es el teléfono móvil (94,5%), en comparación al ordenador portátil (55,4%) o el de sobremesa (39,2%). También se ha determinado que existe una serie de factores que influyen en la popularidad de artículos en web, como son la *estructura de contenido*, *las palabras clave en los metadatos* o *el momento de la publicación*.

Así pues, para conseguir notoriedad de imagen y autoridad en la red, las empresas deberán adaptarse a este nuevo tipo de cliente y competir con otras empresas que ya se han acomodado a este nuevo mundo digital.

Abstract

Since the turn of the 21st century, information technologies (ICT) has been imposed as a key element in society and in companies as a factor in the digitalization and expansion of most business. New ways of getting information through media have augmented exponentially the quantity of data available to business, making big data integration and content analysis the new challenges of our century. In the last two decades, certain key techniques have been developed within the areas of information technology to achieve new content analysis tools through the support of techniques used in Big Data.

This project aims to deepen the use of data analysis to determine customer profiles of ICT users, and subsequently determine customer engagement strategies in the new media through the analysis of large online databases

The results show internet access has changed in recent years: Although a similar number of are connected, the most used device is the mobile phone (94,5%) compared to the laptop (55,4%) or the desktop (39,2%). Content structure, keywords in metadata or post timing has been determined as factors that influence the popularity of web articles.

Thus, to achieve brand awareness and domain authority, companies must adapt to this new type of customer and compete with other companies that have already accommodated to the digital world.

Resumen ejecutivo

Las actuales tendencias de comportamiento de nuestra sociedad están creando nuevas demandas de productos y servicios que no están siendo satisfechas por las empresas. Estos nuevos comportamientos están surgiendo debido a la adopción de las tecnologías de la información y la comunicación. Algunas empresas ya han comenzado procesos de digitalización para poder ofrecer sus productos o servicios a través de la web, pero está siendo insuficiente porque no se está considerando el perfil de este nuevo tipo de cliente.

En este proyecto, se pretende utilizar una serie de métodos analíticos cuantitativos para determinar nuevos perfiles de clientes, a través del análisis de las bases de microdatos de la encuesta de uso de TICs del INE de las olas 2012 y 2018. También se pretende analizar los motivos y las causas por las que un determinado contenido se vuelve popular en la red, a través del análisis de la base de artículos de *mashable.com* de 2015. *El objetivo final de este proyecto es ofrecer a una empresa una serie de herramientas que pueda utilizar, tanto a corto plazo como a largo plazo, para desarrollar estrategias de marketing adaptadas a las nuevas necesidades digitales de sus clientes, y conseguir su fidelización.*

En esta nueva era, el 78,3% de la población en España usa internet. Aunque estos datos no han cambiado tanto con respecto al 2012, el método por el cual las personas acceden a internet está cambiando. En el 2018, el 94,5% de las personas han accedido a internet a través del teléfono móvil, un 3,6% más que en el año 2012. Mientras que, en comparación, sólo un 55,4% usó el ordenador portátil y sólo un 39,2% usó el ordenador de sobremesa para acceder a internet.

Por otro lado, el comercio electrónico también está creciendo. En 2018, el 64,1% de los encuestados habían utilizado internet, un 9,7% más que en el año 2012. Hay que destacar que más del 75% de las personas menores de 35 años han comprado a través de internet.

El uso de servicios a través de internet está condicionado por la edad y los estudios terminados de los encuestados. En general, los encuestados que tienen un nivel de estudios elevado son más propensos a utilizar servicios en línea. Por último, los encuestados de entre 25 y 45 años son los que más servicios en línea utilizan.

Un dato importante es que parece existir una relación positiva entre los encuestados que han comprado alguna vez a través de internet, que tienen un nivel educativo alto y que han usado el móvil o tablet para acceder a internet y los que han utilizado los servicios de banca financiera online.

A través del análisis de contenidos de la base de datos *mashable.com* de 2015, se han analizado las causas por las que los artículos publicados de una empresa se hacen populares en la red. La popularidad de un artículo parece depender de cuatro factores: la calidad del contenido del artículo, el número de palabras clave en los metadatos, el día de la semana que se publican y el tipo de artículo.

Los artículos sobre tecnología, estilo de vida y redes sociales se han compartido con más frecuencia que el resto de artículos. Además, los artículos que se han publicado en fin de semana se han compartido con más frecuencia que los que se han compartido durante la semana laboral.

Sin embargo, publicar el fin de semana puede que no sea siempre la mejor opción. A través de un análisis de *diagrama de decisiones*, existen otros factores que pueden tener un papel más importante para que los usuarios los compartan. Estos factores son el número de palabras clave en los metadatos, la estructura del contenido, o las acciones en las palabras clave más importantes.

Aunque el uso de internet haya sido masivo en los últimos años, ha cambiado la forma en la que se usa. El móvil es la forma más común para acceder a la red, mientras que la frecuencia de uso del ordenador portátil y del ordenador de sobremesa para acceder a la web está descendiendo. Por ello es necesario tener en cuenta la planificación de un diseño web adaptable para la mejora de la experiencia de usuario. Se debe tener en cuenta tanto para la creación de nuevos productos o servicios online, y también para el proceso de digitalización en general.

Las personas que tienen un nivel educativo elevado son también las que más compran a través de internet. La edad también ha sido un factor influyente en cualquier tipo de compra online que se ha analizado. Por ello, es importante tener en cuenta estas variables sociodemográficas para determinar nuevos segmentos de mercado y sus perfiles, y adaptar el plan de marketing a sus demandas y necesidades específicas.

En conclusión, la población está cambiando la forma en que usa internet, y la forma de cómo está accediendo a la red. Este proyecto ha probado que el uso del móvil y tablet está aumentando, y por ello las empresas deben adaptar sus estrategias de marketing a un nuevo escenario con actores que tienen perfiles en el 2019 muy diferentes, evidentemente a los del 2012. Además, para conseguir notoriedad de imagen y autoridad en la red, las empresas deberán adaptarse a este nuevo tipo de cliente, que ya se puede considerar un nativo digital.

Contenido

1. Importancia de las tecnologías de la comunicación y del análisis de contenidos: introducción y marco de este proyecto.	1
2. Objetivos del trabajo.....	5
3. Metodología	5
3.1 Encuesta de uso de TICs	5
3.2 Base de datos de popularidad de noticias online 2015	7
4. Resultados	10
4.1 Determinación de nuevos clientes de TICs.....	10
4.1.1 Uso de internet	10
4.1.2 Uso de dispositivos para acceder a Internet	11
4.1.3 Comercio electrónico	12
4.1.4 Volumen de compra en comercio electrónico.....	13
4.1.5 Tipos de producto en comercio electrónico	15
4.1.6 Uso de servicios utilizados a través de internet.....	20
4.1.7 Diagrama de árbol para la banca financiera en el 2018	24
4.2 Análisis de contenidos de mashable.com.....	27
4.2.1 Análisis factorial	27
4.2.2 Análisis de outliers	29
4.2.3 Momento de publicación.....	30
4.2.4 Popularidad.....	32
4.2.5 Diagrama de árbol para la popularidad de un artículo	34
5. Conclusiones.....	36
6. Recomendaciones	37
7. Limitaciones y futuras líneas de trabajo	37
8. Bibliografía.....	38

1. Importancia de las tecnologías de la comunicación y del análisis de contenidos: introducción y marco de este proyecto.

El mundo actual está digitalizado. Cuando las personas realizan actividades diarias a través de internet, como navegar por la web o compartir contenido en redes sociales, dejan un rastro de datos. El incremento del volumen de información que retienen las empresas a través de los medios de comunicación, y el Internet de las Cosas (*Internet of Things*) ha hecho que aumente de forma exponencial el “Data” disponible en el futuro (Bessis & Dobre, 2014). Por ello, la incorporación de las grandes bases de datos y en diferentes soportes, conocidas como *Big Data*, es uno de los retos empresariales de este siglo. La información que ofrecen las bases de datos masivas no pertenecerá a unos gestores limitados, sino que los líderes de todos los sectores tendrán que entender todas las implicaciones del Big Data (Álvarez, 2014).

En este contexto, es de interés señalar que uno de los valores más importantes de una empresa es su llamado capital intelectual. Éste se define como “el conjunto de aportaciones no materiales que en la era de la información se entienden como el principal activo de las empresas del tercer milenio” (Sanchez Medina, Melián Gonzalez, & Hormiga Pérez, 2007). En las dos últimas décadas, se han desarrollado determinadas técnicas clave dentro de las áreas de la tecnología de la información y comunicación, que nos han proporcionado nuevas herramientas con las que hemos construido una economía global. Muchas de estas herramientas aportan beneficios inmateriales que ahora se dan por descontado, pero que antes no existían, hasta el punto de que la organización no puede funcionar sin ellas. La propiedad de tales herramientas proporciona ventajas competitivas y, por consiguiente, constituyen un activo en cualquier empresa y más aún en empresas del sector bancario en las que la fidelización del cliente es un factor clave como estrategia empresarial (Larran Jorge & Muriel de los Reyes, 2007).

La forma de demostrar la importancia del capital intelectual que ha tenido más auge en los últimos 5 años es el análisis de contenidos. El análisis de contenidos es una herramienta de investigación utilizada para determinar la presencia de palabras clave o conceptos dentro de un texto o grupo de textos. Consiste en analizar la presencia, significado y relación de dichas palabras y conceptos, hacer inferencias sobre el mensaje del texto, el autor, la audiencia, la cultura y el momento temporal que éstos ocupan. Un texto puede ser definido como un libro, capítulo, ensayo, entrevista, discusión, conversación, artículo, etc., O cualquier otra forma que tome la comunicación oral o escrita. Este análisis de contenidos sigue un procedimiento sistemático. Primero, el investigador plantea una serie de preguntas o hipótesis. Segundo, se selecciona una

muestra. Tercero, se definen una serie de categorías para codificarlas. Cuarto, se codifica el contenido y se comprueba la fiabilidad. Por último, los datos codificados se analizan y se interpretan. El desarrollo de la estructura moderna del medio digital ha permitido la aparición de nuevas características explotables relacionadas con el mundo de la comunicación, como el *hashtag* de Twitter, y el desarrollo de contextos socio-culturales nuevos que giran en torno a estos nuevos rasgos, que suponen un nuevo terreno para la exploración del análisis de los contenidos (Abela, 2018).

Internet ha introducido nuevos retos al concepto tradicional de análisis de contenidos. Se han identificado problemas para conseguir una muestra representativa debido a la cantidad masivas de internautas activos. Gracias a los móviles, redes, ubicaciones o fotos se producen torrentes de datos difíciles de manejar debido a que estos datos disponibles están desestructurados. Esto es un problema para definir la unidad de análisis y la codificación de forma que la información sea precisa y fiable para cualquier persona que quiera usarla y para anticipar cuándo y dónde va a aparecer nuevos datos relevantes (Olave, 2019). Además, los datos de dominio público suelen tener menos calidad que los de dominio privado, que raramente están disponibles para el personal de investigación (Luque, 2019). Incluso cuando existen datos disponibles no existen una forma de recopilar dichos datos de forma fiable, ya que no se conoce la aleatoriedad de los datos, lo que resta fiabilidad a la representatividad de estos a los usuarios del servicio (Olave, 2019).

Esto ha fomentado la aparición de nuevas formas de realizar este tipo de análisis: han aparecido nuevas ramas de investigación de análisis de contenido con el fin de utilizar la masiva cantidad de datos de *Twitter* para usarlo en periodismo, PYMES y organizaciones sin ánimo de lucro. Estos estudios suelen incluir un análisis de contenidos de cuentas empresariales, el funcionamiento o gestión de las mismas, y un análisis del contexto sociocultural relacionado con ellas (Herrera Damas & Requejo Aleman, 2012).

Es por ello que es necesario realizar este análisis de contenido a través del apoyo de técnicas y herramientas en Big Data para poder desarrollar estudios fiables. El Big Data se define como “conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles” (Powerdata, 2018).

Hay cinco formas establecidas en las que el Big Data pueden crear valor. La primera es que hace que la información se transparente y utilizable de forma rápida y frecuente. La segunda es que,

debido a que las empresas están creando y manteniendo datos de forma digital, se puede conseguir información más detallada y precisa para tomar decisiones que mejoren el desempeño empresarial y de gestión. Tercero, el Big Data permite realizar una segmentación del mercado más reducida y precisa para crear productos y servicios personalizados para el cliente. Cuarto, el análisis de estos datos puede mejorar la toma de decisiones internas, externas y de naturaleza preventiva o proactiva. Y quinto, el Big Data puede usarse para el desarrollo y mejora de productos y servicios futuros. Por ejemplo, a través del análisis de datos del Internet de las Cosas (Opresnik, 2015).

El uso del análisis de Big Data supone una ganancia para todos los sectores. Sin embargo, hay algunos sectores que ganarán más que otros. Según Manyika (2011), que analizó las oportunidades y retos del análisis de Big Data para varios sectores a través de un conjunto de indicadores cuantitativos, previó que los sectores con más potencial de productividad a través del uso del Big Data son los productos electrónicos e informáticos, los productos financieros, los seguros, y el Gobierno.

Por otro lado, las Tecnologías de la Información y la Comunicación (TIC) han revolucionado el mundo en los últimos 20 años. Nos encontramos en la llamada “era digital”, donde estas tecnologías están en todos los ámbitos de nuestra vida y son fundamentales en nuestra sociedad actual, lo que hace la vida más fácil que a la generación anterior (Manyika, 2011).

En particular, destaca la velocidad con la que cada vez una tecnología más moderna sustituye a la anterior. Hace unos años, se creó la manera de enviar correos a través de plataformas de correo electrónico. Esta tecnología quedó obsoleta debido a la aparición de aplicaciones de mensajería instantánea, que ha permitido la comunicación entre dos o más personas como si de una conversación se tratara, y desde cualquier lugar gracias a la portabilidad de los teléfonos móviles (Opresnik, 2015).

El comercio también ha sido revolucionado por las TICs. El rasgo más importante es la aparición del llamado comercio electrónico, referido a todas las actividades de compra venta que se realizan a través de la web. Según la Cámara de Comercio de Zaragoza (2018), las ventas online aragonesas han incrementado su valor en un 30%. Todas las grandes empresas del mundo cuentan con una cadena de distribución física y una plataforma online para ofrecer sus productos. Además, el pequeño comercio está empezando a integrarse en este mercado, ya que también encuentran ventajas competitivas en la distribución online (como la disminución de gastos estructurales) (Bessis & Dobre, 2014).

Uno de los sectores más relevantes a la hora de invertir en nuevas tecnologías ha sido el sector financiero, dado que la confianza del cliente en sus gestores y en la empresa es clave para la

introducción de nuevos productos. Así pues, entendemos que cualquier departamento de marketing en ese sector debe conocer las herramientas de TICs y su puesta en marcha como medio de comunicación con el cliente. Las TICs han ayudado a mejorar la eficiencia en las operaciones, las relaciones e interacciones con los clientes y han aportado nuevas herramientas de seguridad. Jorge Gajardo (2017) dice “Las TIC han sido cruciales para mejorar la interacción y fidelización de los clientes gracias a una atención más rápida y eficiente, y para acelerar los procesos internos de las instituciones financieras” (Kibernum, 2018).

Las TICs también han impactado en los clientes y han cambiado su comportamiento financiero. La evolución de las TICs en el sector financiero ha ido cambiando a lo largo de los años. Desde la aparición de cajeros automáticos al desarrollo de portales en internet donde los bancos pueden dar información propia de la institución y ofertas de sus productos (Larran Jorge & Muriel de los Reyes, 2007).

La tecnología ha permitido también la inclusión y masificación de tarjetas de crédito, impactando, de esta forma la capacidad de compra de las personas y, en los últimos años, la tecnología móvil ha ayudado a que los clientes bancarios accedan a la información oficial de sus productos y realicen transacciones a través de sus dispositivos móviles desde donde se encuentren. “Las personas ya no sólo evitan visitar una sucursal, sino que tampoco requieren de un PC para realizar sus operaciones financieras. (...) En paralelo, otras tecnologías han mejorado los procesos productivos u operacionales de nuestros clientes, reduciendo los costos que después redundan en beneficios para sus clientes finales”, asevera el gerente de Ingeniería de Software de Kibernum (2018).

Es por ello que se va a realizar un análisis de tres bases de datos, dos relacionadas con las TICs y la nueva tipología de cliente de banca, y otra relacionada con el análisis de contenidos digitales, con el fin de proporcionar a la entidad financiera Ibercaja una serie de herramientas que le podrán servir como método de mejora en su relación con clientes y fidelización de cartera, buscando que los departamentos de marketing estén liderando innovación en tecnologías de comunicación.

2. Objetivos del trabajo

El objetivo principal de este trabajo es proporcionar herramientas para la toma de decisiones óptimas en el análisis y estrategias de contenidos a través de análisis de bases de datos masivos.

Como objetivos secundarios, en este proyecto se proponen:

1. Determinar el perfil de uno o varios tipos de clientes del mercado financiero, que es usuarios de las tecnologías de la información y la comunicación (TICs).
2. Analizar las causas por las que los artículos publicados de una empresa se hacen populares en la red.

3. Metodología

Para el primer objetivo secundario de este estudio se ha utilizado la encuesta de equipamiento y uso de tecnologías de la información y comunicación de los hogares (TICs) realizada por el instituto nacional de estadística (INE). Del panel de datos de la base, hemos seleccionado las olas 2012 y 2018. Para el segundo subobjetivo se ha utilizado la base de datos de artículos de *mashable.com*, extraída del repositorio de la universidad de California en Irvine (UCI).

3.1 Encuesta de uso de TICs

La encuesta sobre equipamiento y uso de tecnologías de la información y comunicación en los hogares es realizada por el INE con una periodicidad anual. El ámbito de estudio es la población que reside en viviendas familiares en todo el territorio nacional. Se trata de un muestreo trietápico con estratificación de las unidades de primera etapa. Las unidades de primera etapa son las secciones censales, las de segunda etapa las viviendas familiares principales, y en tercera etapa se selecciona en cada vivienda una persona de 16 o más años. También se obtiene información de todos los niños de 10 a 15 años. El tamaño muestral es de 2500 secciones censales y 25000 viviendas repartidas en las mismas. Las variables investigadas son información de los doce a tres meses anteriores del momento de la entrevista. Para este análisis se han utilizado las encuestas realizadas en las olas 2018 y 2012, utilizando esta última para detectar la existencia de posibles cambios de tendencia en los últimos 6 años.

FICHA TÉCNICA:

Tipo de encuesta: continua de periodicidad anual

Ámbito poblacional: población que reside en viviendas familiares principales

Ámbito geográfico: todo el territorio nacional

Período de referencia de los resultados: el primer trimestre del año

Período de referencia de la información: la mayor parte de las variables investigadas se refieren a los doce y tres meses anteriores al momento de la entrevista

Tamaño muestral: unas 2.500 secciones censales, alrededor de 25.000 viviendas.

Tipo de muestreo: trietápico con estratificación de las unidades de primera etapa. Las unidades de primera etapa son las secciones censales, las de segunda etapa, las viviendas familiares principales. En tercera etapa se selecciona en cada vivienda una persona de 16 o más años. Asimismo, se obtiene información de todos los niños de 10 a 15 años.

Numero de variables en el cuestionario: 241

Se han analizado las siguientes variables de las encuestas: ¿Ha usado internet alguna vez? (USO_INT), Forma de conexión a Internet en la vivienda (INTEFOR1 a 7), Compras a través de internet (COMPRAS), Valor aproximado de compras en los últimos 3 meses (VCOMPRAS) se ha categorizado en cinco grupos (<50, 50-100, 100-500, 500-1000, >1000 euros), Productos/servicios comprados en Internet en los últimos doce meses, 4 productos y 3 servicios diferenciados, que se citan a continuación:

Etiqueta de la variable	Nombre de la variable
bienes hogar	prod2
libros, revistas, periódicos	prod5
alojamiento vacaciones	prod13
entradas para espectáculos	prod15
Participar en redes sociales	serv16_3
Buscar información sobre bienes y servicios	serv16_4
Banca electrónica	serv19_2

Se han utilizado también variables de clasificación. Las variables de clasificación analizadas han sido sexo (SEXO), edad (EDAD) que se ha categorizado en seis estratos (<25 años, 25-34, 35-44, 45-54, 55-64, >64 años), y el nivel de estudios terminados (NIVELEST).

La diferencia entre los niveles de los distintos factores se ha establecido mediante el método de la tabla de contingencia y el Chi-cuadrado de Pearson, utilizando un nivel de significación $p=0,05$.

3.2 Base de datos de popularidad de noticias online 2015

Se ha extraído de la web de recopilación de datos de la Universidad de California en Irvine (UCI), que donó *mashable.com* en el año 2015. El ámbito de estudio son los 39644 artículos de la web publicados en los años 2013 y 2014.

Las variables que se han recogido en la base de datos explican la tipología del artículo, y la representación tanto positiva como negativa del contenido. También recopila información sobre la optimización de la búsqueda web (SEO, por sus siglas en inglés), a través de la recopilación de las palabras clave en los metadatos y en la redacción del artículo. Por último, recoge la información de cuándo se ha publicado el artículo.

FICHA TÉCNICA:

Tipo de encuesta: continua de periodicidad anual

Ámbito poblacional: artículos publicados en los dos últimos años

Período de referencia de los resultados: el primer trimestre del año

Tamaño muestral: 39644 artículos publicados durante el 2013 y 2014.

Tipo de muestreo: Se analizan todos los artículos publicados en los últimos 2 años.

Numero de variables en el cuestionario: 61

Se han analizado 31 de las 61 variables. Las variables analizadas han sido:

Etiqueta de la variable	Nombre de la variable
URL del artículo	URL
Número de palabras en el contenido	n_tokens_content
Número de enlaces en el contenido	num_hrefs
Número de imágenes	num_imgs
Número de vídeos	num_videos
Número de palabras clave en los metadatos	num_kw
¿Es el canal "estilo de vida"?	data_channel_is_lifestyle
¿Es el canal "entretenimiento"?	data_channel_is_entertainment
¿Es el canal "negocios"?	data_channel_is_business
¿Es el canal "social media"?	data_channel_is_socmed
¿Es el canal "tecnología"?	data_channel_is_tech
¿Es el canal "noticias"?	data_channel_is_world
Peor palabra clave (acciones mínimas)	kw_min_min
Mejor palabra clave (acciones máximas)	kw_max_max
Promedio palabra clave (acciones promedio)	kw_avg_avg
Promedio palabra clave (acciones mínimas)	kw_min_avg
Promedio palabra clave (acciones máximas)	kw_max_avg
¿Es el día de la semana "lunes"?	weekday_is_Monday
¿Es el día de la semana "martes"?	weekday_is_Tuesday
¿Es el día de la semana "miércoles"?	weekday_is_Wednesday
¿Es el día de la semana "jueves"?	weekday_is_Thursday
¿Es el día de la semana "viernes"?	weekday_is_Friday
¿Es el día de la semana "sábado"?	weekday_is_Saturday
¿Es el día de la semana "domingo"?	weekday_is_Sunday
¿Es fin de semana?	is_weekend
Máx. Acciones "share" de los artículos referenciados	self_reference_max_shares
Min. Acciones "share" de los artículos referenciados	self_reference_min_shares
Polaridad del artículo	global_sentiment_polarity
Polaridad del título	title_sentiment_polarity
Número de veces que se ha compartido	shares
Popularidad	shares_cat

Se categoriza un artículo como “*popular*” por las veces que se ha compartido un artículo de parte de los usuarios. Se ha considerado que un artículo es popular si se ha compartido más de 1400 veces, que es la mediana del número de veces que se han compartido los artículos (Fabián Abad, 2017).

Las bases de datos como la de *mashable.com* contienen una gran cantidad información, y resultan complicadas tanto de analizar como de interpretar. Por ello existen métodos estadísticos que facilitan el análisis de los datos.

Uno de estos métodos es el denominado “análisis factorial”, que clasifica las variables seleccionadas para definir/identificar factores que reduzcan la complejidad de la información. Cada factor inobservable está definido a través de unas variables que están relacionadas entre sí. La reducción de la dimensión se ha realizado a través de un análisis de los componentes principales con rotación *varimax*, que es una técnica que busca una proyección de las variables donde los datos sean representados mejor a través del cálculo de mínimos cuadrados. Utilizando los nuevos factores extraídos de la reducción de dimensión, se ha identificado y analizado los artículos atípicos (outliers).

Una vez eliminados los outliers, se ha realizado un análisis a través de tablas de contingencia para determinar qué variables pueden influenciar a la popularidad de los artículos en la web. Por último, se ha elaborado un diagrama de árbol con el método de división CHAID. Este método detecta interacciones mediante la prueba Chi-cuadrado y en cada paso elige la variable independiente que presenta la interacción más fuerte con la variable analizada.

La diferencia entre los niveles de los distintos factores se ha establecido mediante el método de la tabla de contingencia y el Chi-cuadrado de Pearson, utilizando un nivel de significación $p=0,05$.

El análisis estadístico se ha realizado con el programa SPSS. De forma análoga, se podría realizar con el software libre R(Rcmdr).

4. Resultados

4.1 Determinación de nuevos clientes de TICs

Para determinar el nuevo perfil de cliente, se ha realizado un análisis de la base de uso de TICs 2018, y se han comparado los resultados con la información de la base de datos de uso de TICs de 2012.

4.1.1 Uso de internet

Un 78,3% de los encuestados han usado internet (gráfico 1).

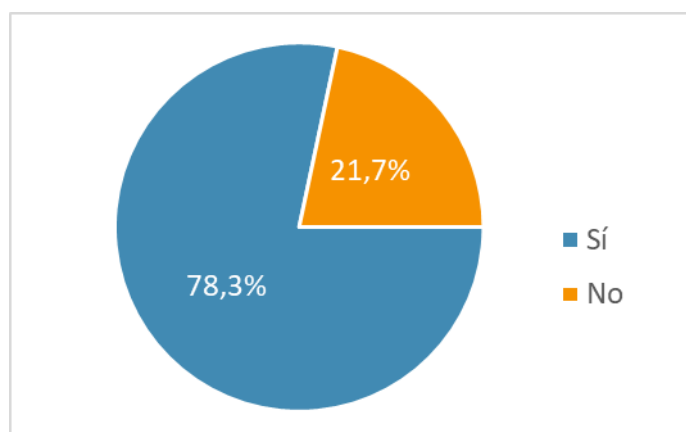


Gráfico 1. Uso de internet en el año 2018

Se ha encontrado que existe una relación entre el uso de internet y la edad de los encuestados. Los mayores a 64 años han usado internet con menos frecuencia que el resto de encuestados: sólo un 33,4% de ellos ha usado internet (gráfico 2).

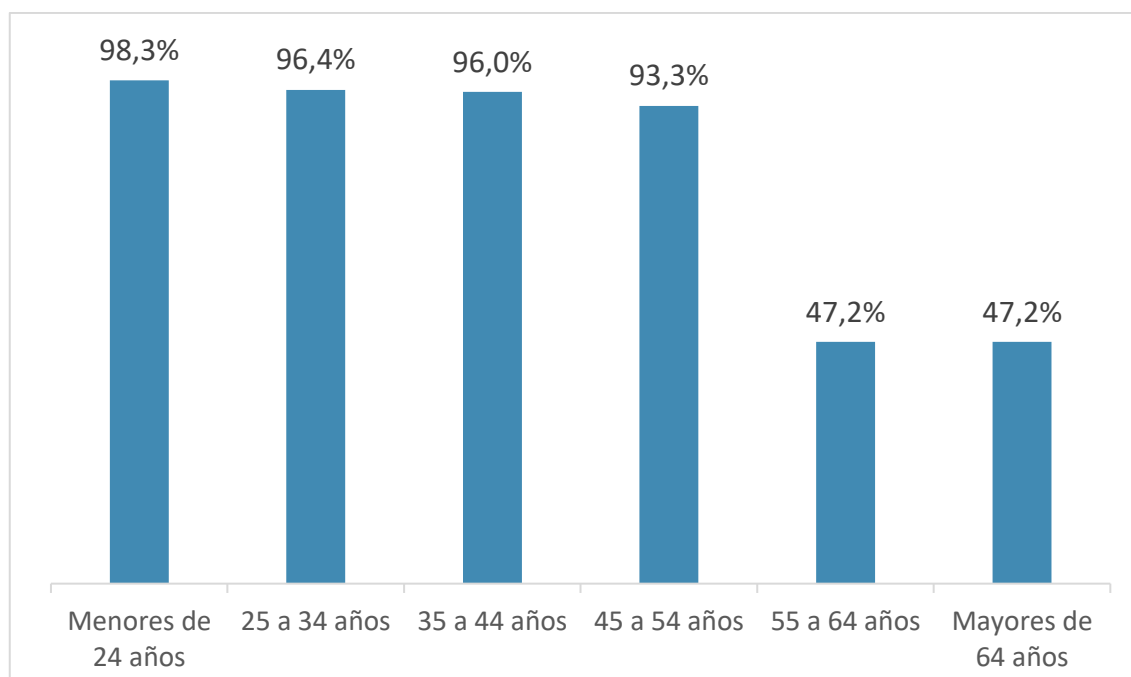


Gráfico 2. Uso de internet según la edad

Las mujeres han usado internet con menos frecuencia que los hombres de forma significativa. Un 76,4% de los hombres encuestados han usado internet, frente al 70,7% de las mujeres (gráfico 3).

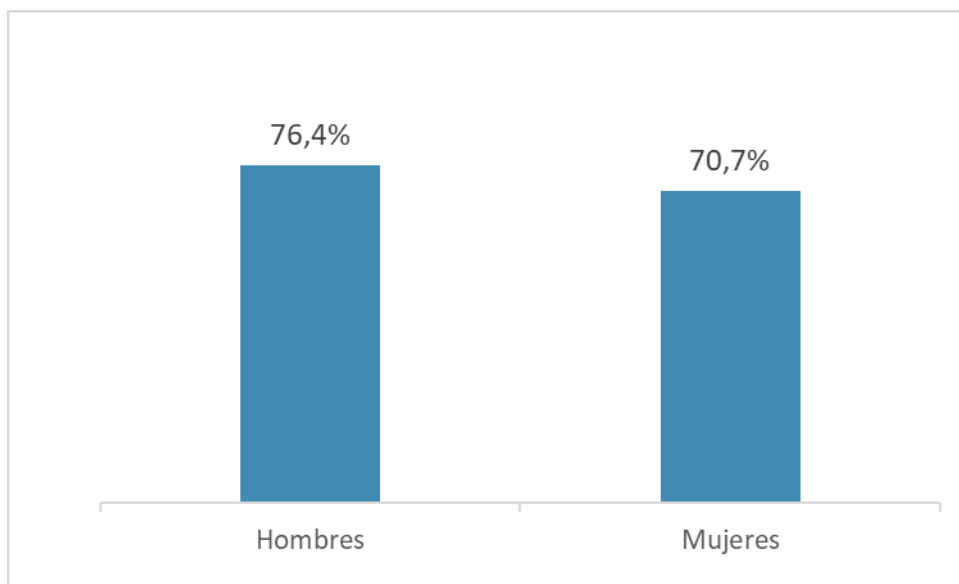


Gráfico 3. Uso de internet según el sexo en el 2018

4.1.2 Uso de dispositivos para acceder a Internet

Los dispositivos para acceder a internet más comunes han sido el móvil, la tablet, el ordenador de sobremesa, y el ordenador portátil.

El uso de estos dispositivos ha ido variando en los últimos 6 años. En el año 2018, el uso del móvil para conectarse a internet ha sido del 94,5%, con un aumento del 3,6% con respecto al año 2012. El 41,4% de los encuestados han utilizado la tablet para acceder a internet en los últimos 3 meses, y se mantuvo parecido al año 2012. Por otro lado, el acceso a internet a través de ordenadores ha descendido con respecto al año 2012. El 39,2% de los encuestados han utilizado el ordenador de sobremesa para acceder a internet, un 15% menos que en el año 2012, y el 55,4% de los encuestados han utilizado el ordenador portátil para acceder a internet, un nivel parecido al del año 2012 (gráfico 4).

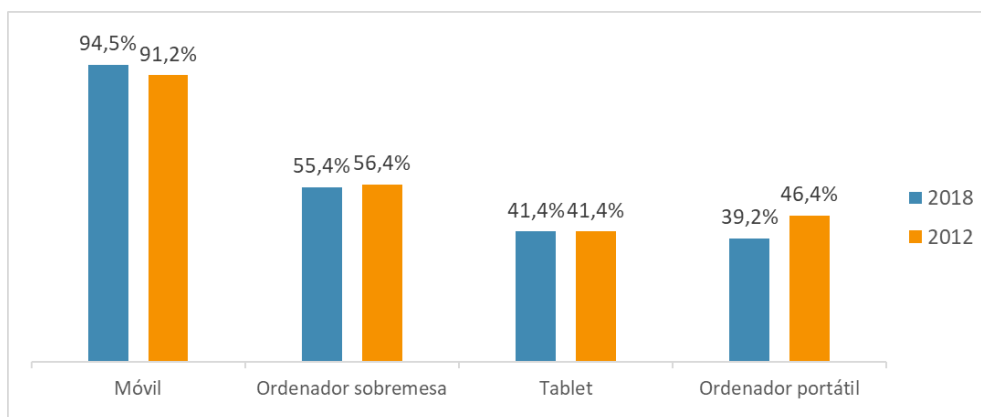


Gráfico 4. Porcentaje del uso de dispositivo para acceder a internet

4.1.3 Comercio electrónico

El comercio electrónico ha aumentado en los últimos 6 años: de las personas que utilizaron internet, el 64,1% han comprado a través de internet en el año 2018, lo que supone un incremento del 9,7% frente al año 2012 (gráfico 5).

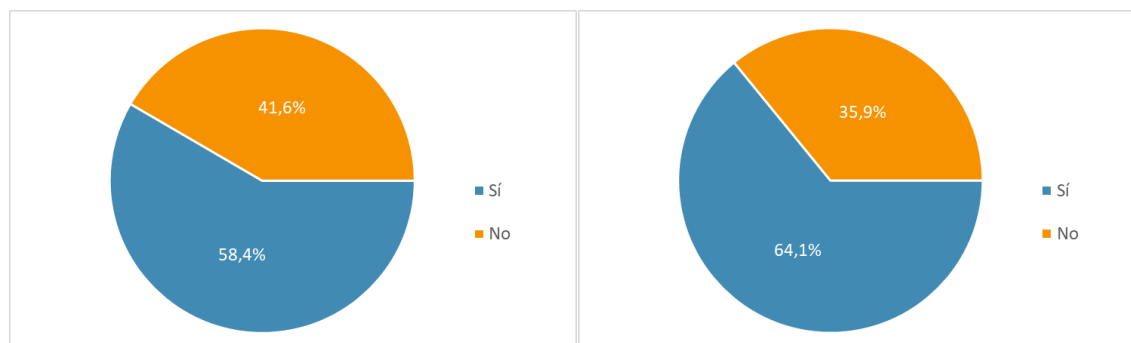


Gráfico 5. Porcentaje de compra a través de internet en los años 2012 (izda.) y 2018 (dcha.)

Las mujeres han comprado con menos frecuencia en internet que los hombres de forma significativa. El 61,9% de las mujeres han comprado alguna vez a través de internet, frente al 66,6% de los hombres (gráfico 6).

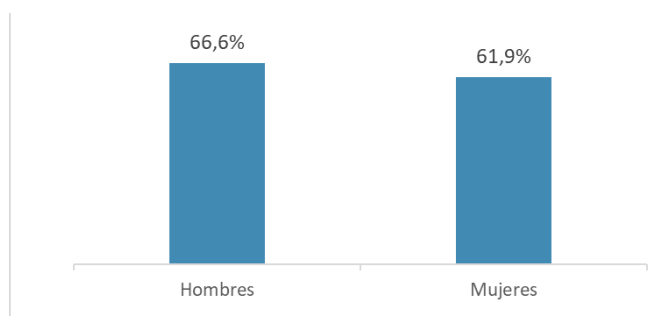


Gráfico 6. Compra a través de internet según sexo

En general, las personas menores de 54 años han comprado por internet con más frecuencia de forma significativa que las que tienen más de 54 años. Así, el 72,3% de los menores de 24 años, el 82,3% de las personas entre 25 y 34 años, el 78,3% de las personas entre 35 y 44 años, y el 67,2% de las personas entre 45 y 54 años han comprado en internet frente al 49,4% de las personas entre 55 y 64 años y el 32,9% de los mayores de 65 años (gráfico 7).

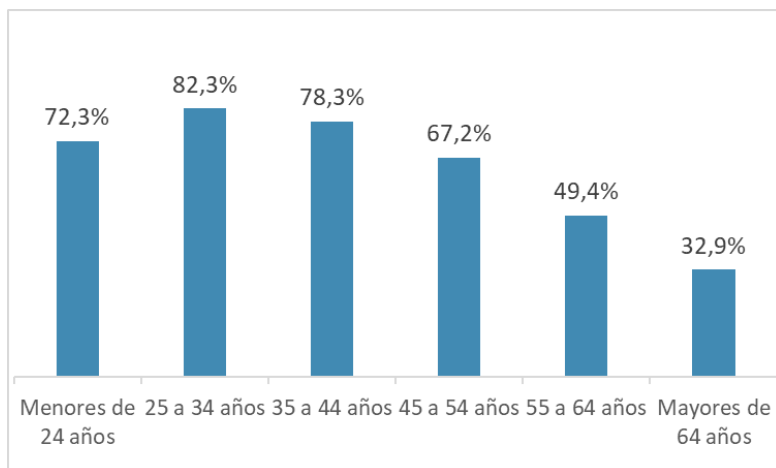


Gráfico 7. Compra a través de internet según edad

4.1.4 Volumen de compra en comercio electrónico

En el año 2018, de los encuestados que han comprado por internet, el 19,6% han gastado menos de 50€, el 28,5% han gastado entre 50 y menos de 100€, el 38,7% han gastado entre 100 y menos de 500€. Por último, el 8,12% han gastado entre 500 y 1000€, y el 5,1% han gastado más de 1000€.

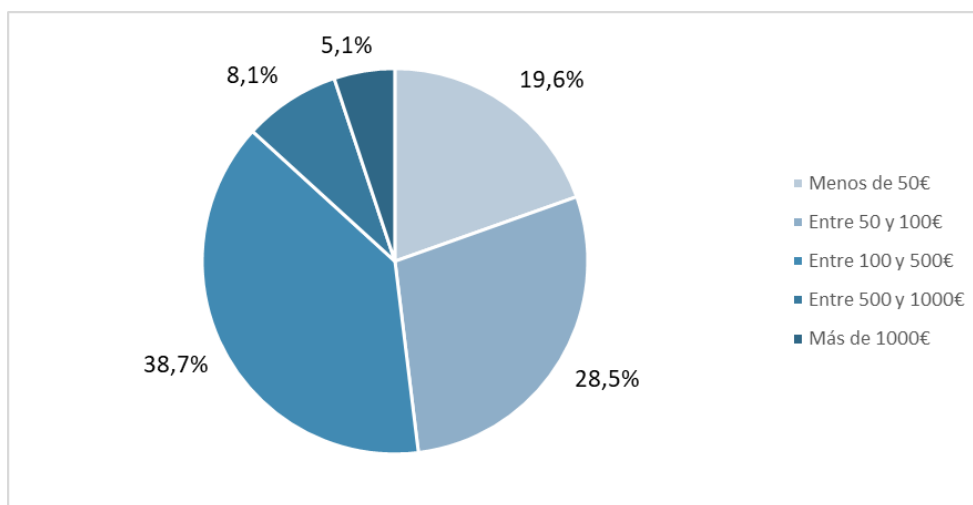


Gráfico 8. Valor aproximado de las compras online en el año 2018

Los hombres han gastado con más frecuencia más de 1000€ en compras por internet que las mujeres. Un 6,5% de los hombres encuestados ha gastado más de 1000€, mientras que sólo un 3,8% de las mujeres han gastado más de 1000€ (gráfico 9).

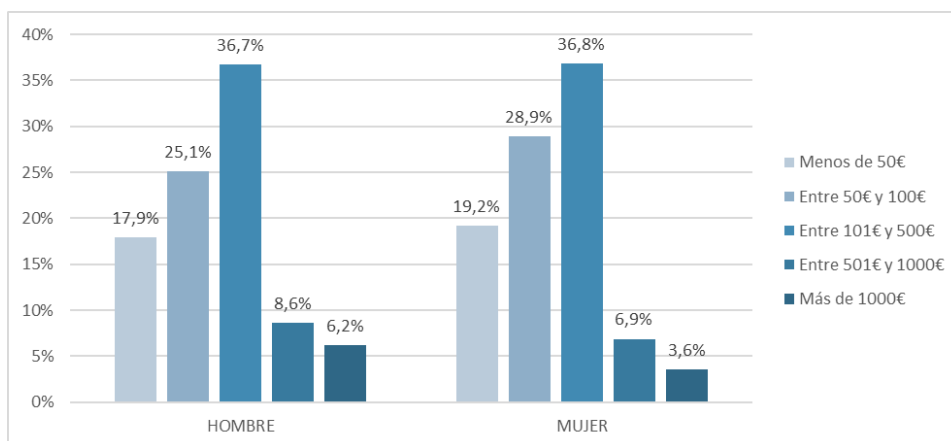


Gráfico 9. Valor aproximado total de las compras por internet en los últimos 3 meses según sexo en el año 2018

Los encuestados menores de 25 años han gastado menos de 50€ y entre 50 y menos de 100€ con más frecuencia que el resto de encuestados: un 33,2% de ellos, y un 33,5% respectivamente. También son los que con menos frecuencia han gastado más de 100€ en compras por internet: el 27,4% han gastado entre 100€ y 500€, el 3,9% han gastado entre 500€ y menos de 1000€, y el 2% ha gastado más de 1000€ (gráfico 10).

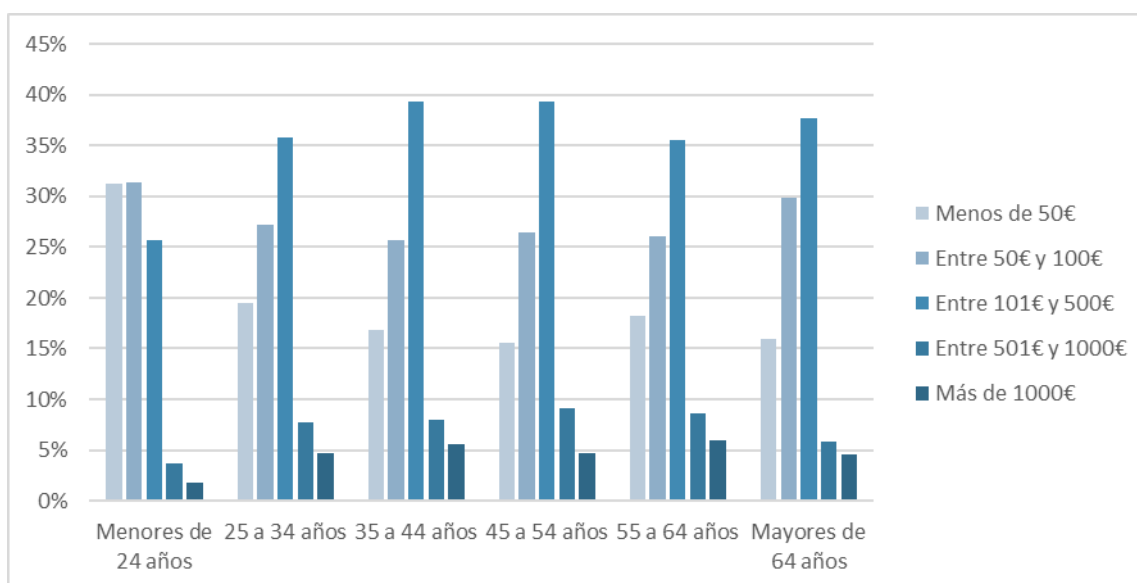


Gráfico 10. Valor de compra por internet según edad

4.1.5 Tipos de producto en comercio electrónico

Se ha analizado la compra de los 4 productos que se han considerados más relevantes. Estos son “bienes del hogar”, “libros y revistas”, “alojamientos vacacionales” y “entradas de espectáculos” (gráfico 11).

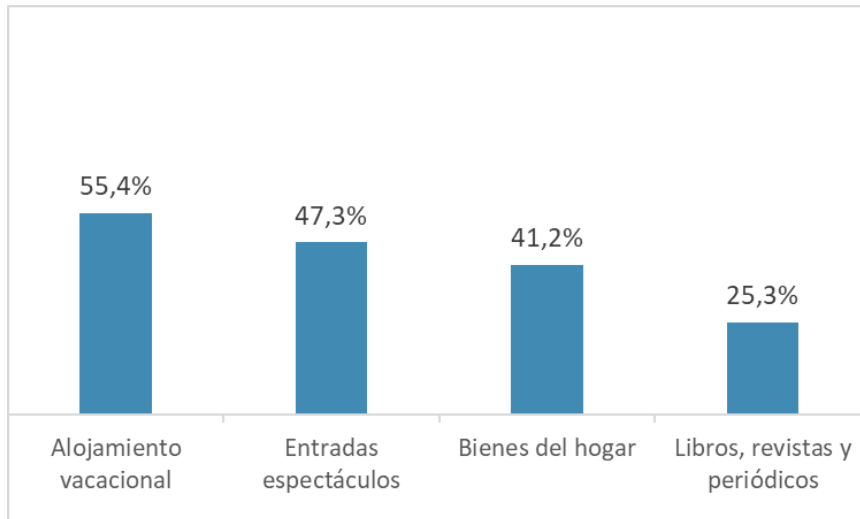


Gráfico 11. Compra online de productos en el año 2018

El 41,2% de las encuestados que han comprado por internet han comprado bienes del hogar (gráfico 12).

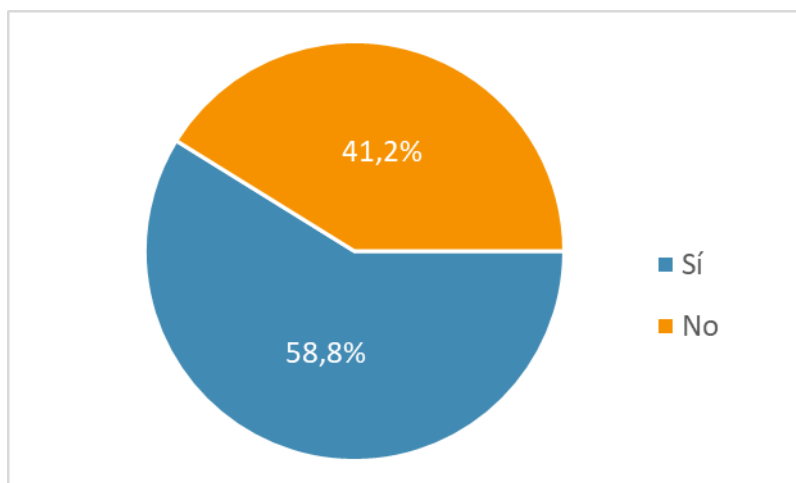


Gráfico 12. Compra online de bienes del hogar en el año 2018

Se han encontrado diferencias significativas entre la compra online de bienes del hogar y la edad. Los encuestados menores de 24 años y los mayores de 55 años han comprado con menos frecuencia bienes del hogar que el resto de los encuestados de forma significativa. Solo el 20,2% de los menores de 24 años, el 33,1% de los encuestados entre 45 y 54 años, y el 33,6% de los mayores de 64 años han comprado por internet, frente a la media del 58.8% (gráfico 13).

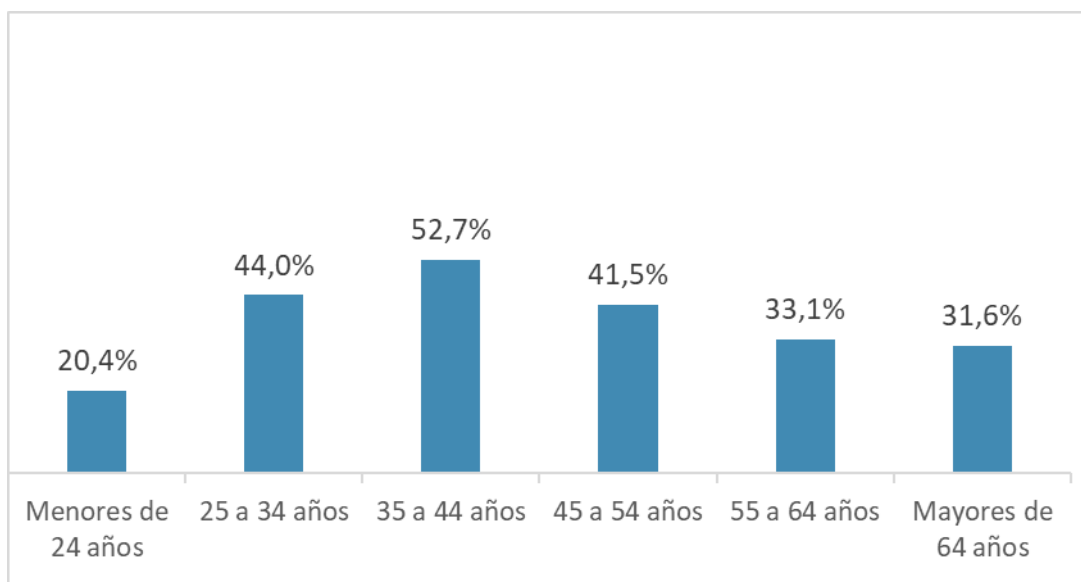


Gráfico 13. Compra online de bienes del hogar según edad

El 25,3% de los encuestados que han comprado a través de internet han comprado libros y revistas (gráfico 14).

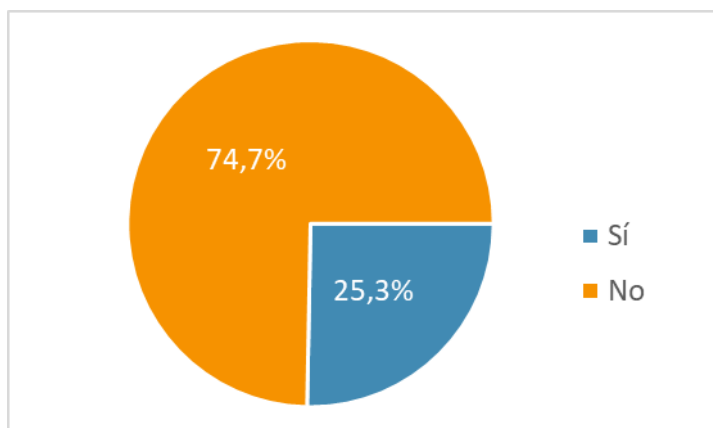


Gráfico 14. Compra online de libros y revistas en el año 2018

Se han encontrado diferencias significativas en la compra online de libros y revistas según la edad. Los menores de 25 años han comprado libros y revistas con menos frecuencia que el resto de encuestados (22,5%). Mientras, los encuestados de entre 45 y 54 años han comprado libros y revistas con más frecuencia que el resto de encuestados (27,2%) (gráfico 15).

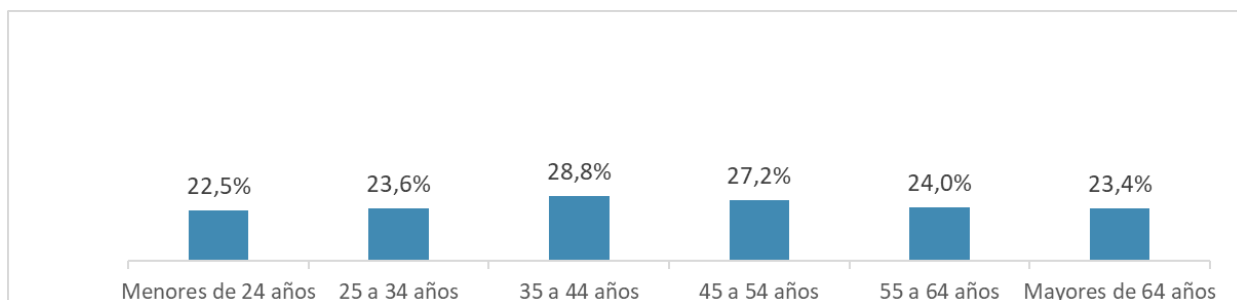


Gráfico 15. Compra online de libros y revistas según la edad

El 55,4% de los encuestados que han comprado a través de internet, han comprado servicios de alojamiento vacacional online (gráfico 16).

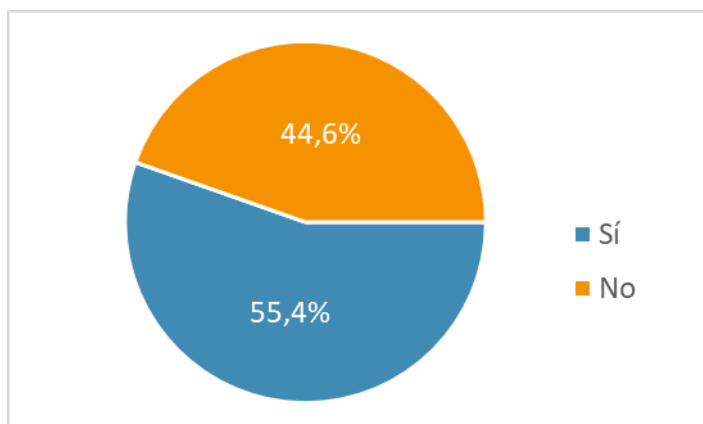


Gráfico 16. Compra online de alojamientos vacacionales en el año 2018

Se han encontrado diferencias significativas en la compra online de alojamiento vacacional según la edad de los encuestados. Los menores de 25 años (37,3%) y los mayores de 64 años (48,4%) han comprado alojamiento vacacional con menos frecuencia que el resto de encuestados. Los encuestados de entre 25 a 34 años (61,0%) y de 35 a 44 años (59,0%) compran alojamiento vacacional con más frecuencia que el resto de encuestados (gráfico 17).

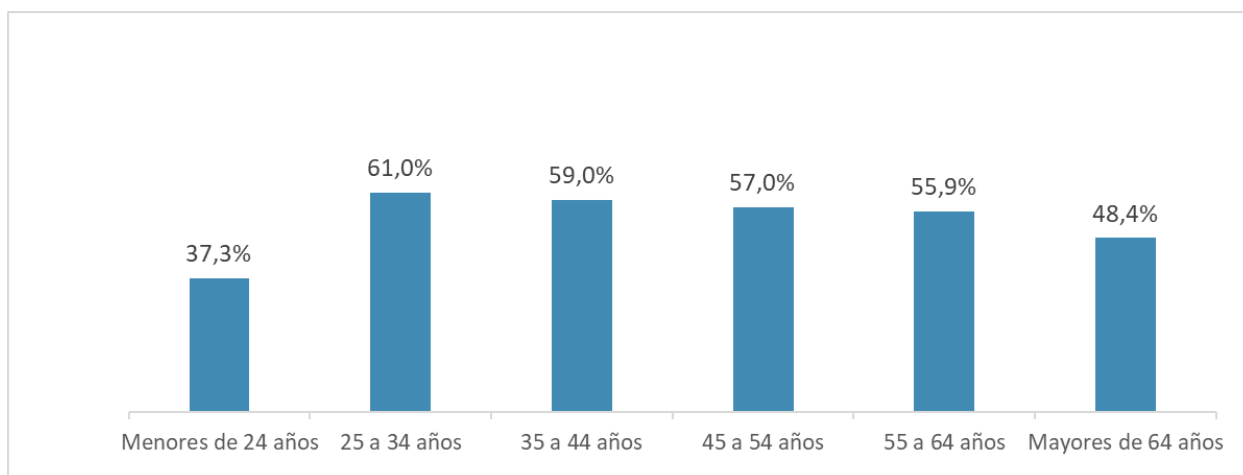


Gráfico 17. Compra online de alojamiento vacacional según edad

El 47,3% de los encuestados que han comprado a través de internet han comprado entradas de espectáculo a través de internet (gráfico 18).

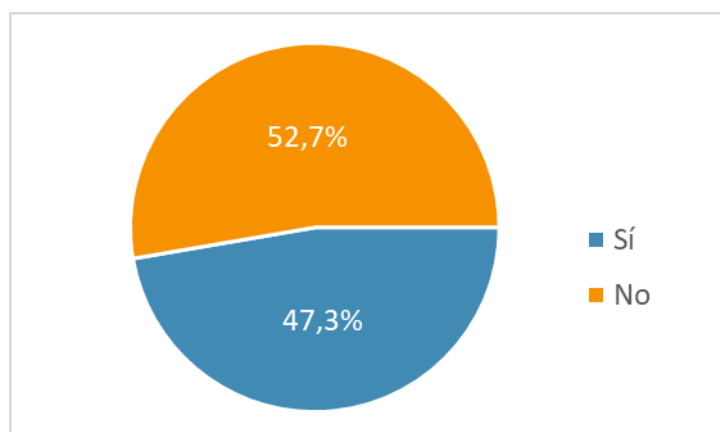


Gráfico 18. Compra online de entradas de espectáculos en el año 2018

Además, los encuestados de entre 55 a 64 años (42,3%) y los mayores de 64 años (38,5%) han comprado entradas de espectáculo con menos frecuencias que el resto de encuestados. Mientras, los encuestados de entre 35 y 44 años (50,6%) han comprado online con más frecuencia entradas de espectáculo que el resto de encuestados (gráfico 19).

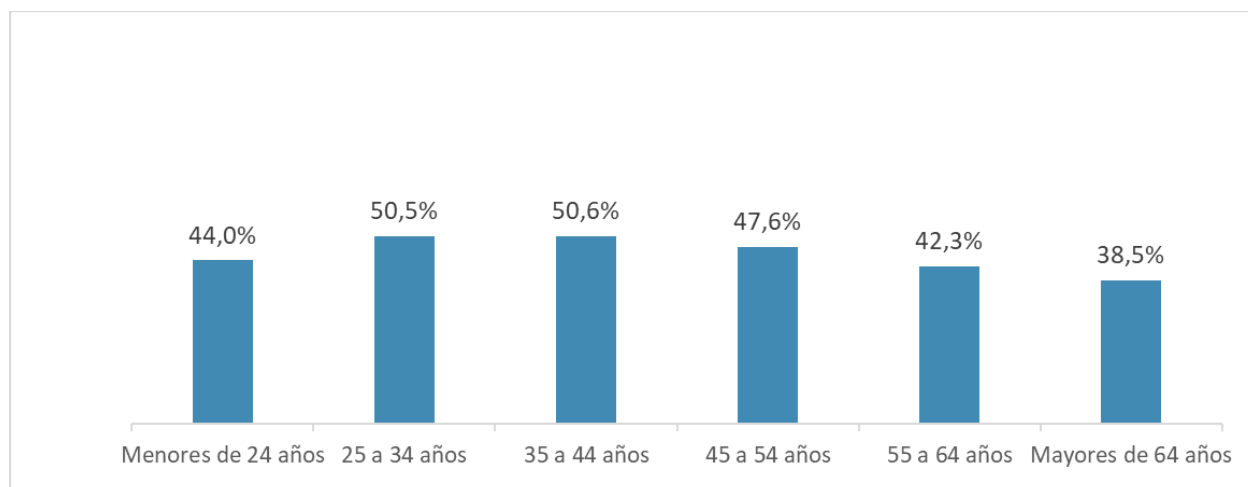


Gráfico 19. Compra online de entradas de espectáculos según edad

4.1.6 Uso de servicios utilizados a través de internet

Se ha analizado los servicios utilizados a través de internet en el año 2018. Los servicios que se han considerado relevantes son la participación en redes sociales, la búsqueda de información de productos y servicios y el uso de la banca electrónica (gráfico 20).

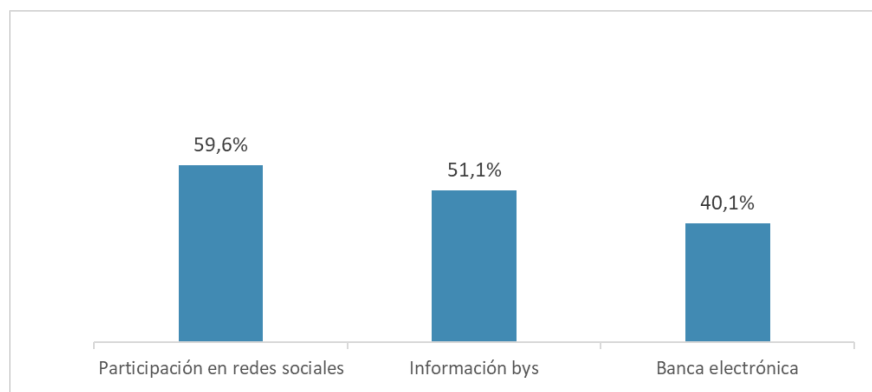


Gráfico 20. Porcentaje de uso de los servicios online en el año 2018

El 61,8% de los encuestados han participado en redes sociales en 2018 (gráfico 21).

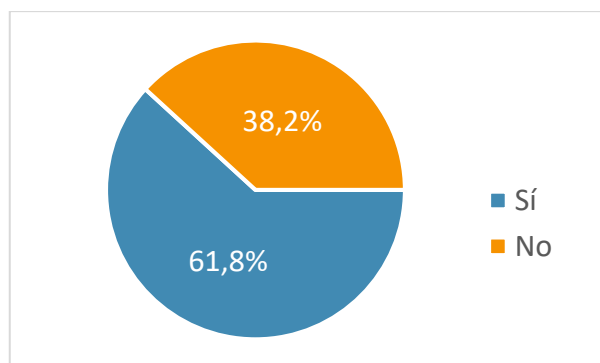


Gráfico 21. Participación en redes sociales en el año 2018

La participación en redes sociales varía según el sexo del encuestado, su edad, y su nivel educativo. Las mujeres utilizan más las redes sociales que los hombres: el 64,6% de las mujeres encuestadas han participado en redes sociales frente al 58,6% de los hombres (gráfico 22).

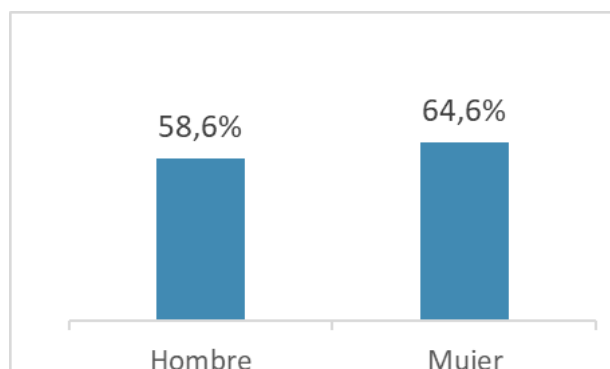


Gráfico 22. Participación en redes sociales según el sexo

La educación también ha influido en el uso de redes sociales. Las personas que no tienen estudios (58,1%) o que tienen un máster o doctorado (53,8%) han participado con menos frecuencia en redes sociales. Las personas que tienen acabados los estudios primarios y los que tienen estudios universitarios participan con más frecuencia en redes sociales, con un 64,9% y un 64,1% respectivamente (gráfico 23).

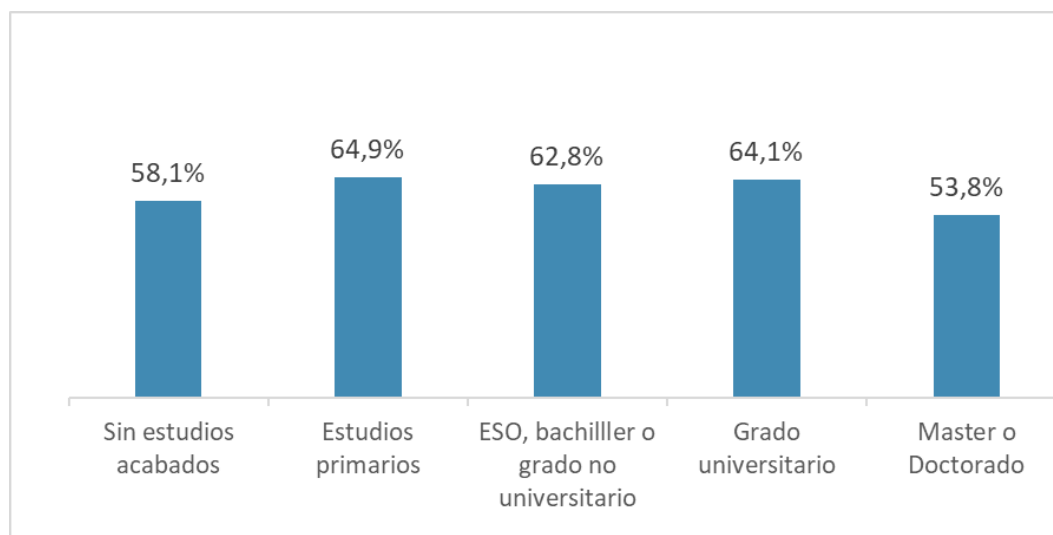


Gráfico 23. Participación en redes sociales según el nivel educativo

La edad ha influido en la participación en redes sociales. Un 90,1% de los menores de 25 años, el 83,6% de las personas entre 25 y 34 años y el 70,9% de los encuestados entre 35 a 44 años han participado en redes sociales. Por otro lado, las personas mayores de 45 años participan en redes sociales con menos frecuencia: sólo el 46,8% de los encuestados de entre 55 a 64 años y el 35,4% de las personas mayores de 64 años han participado en las redes sociales (gráfico 24).

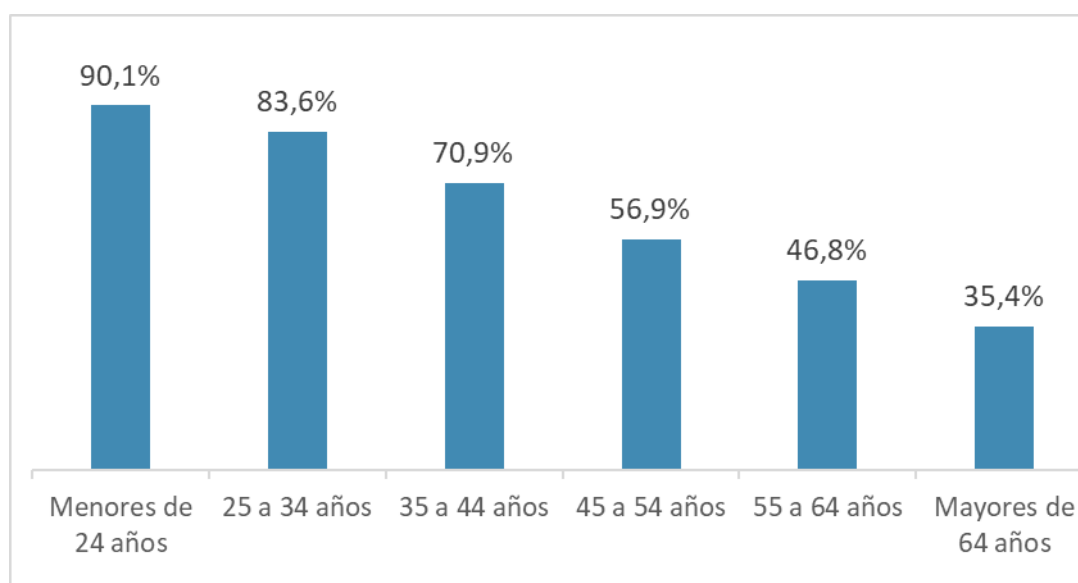


Gráfico 24. Participación en redes sociales según la edad

El 83,3% de los encuestados han buscado información sobre productos y servicios en la red (gráfico 25).

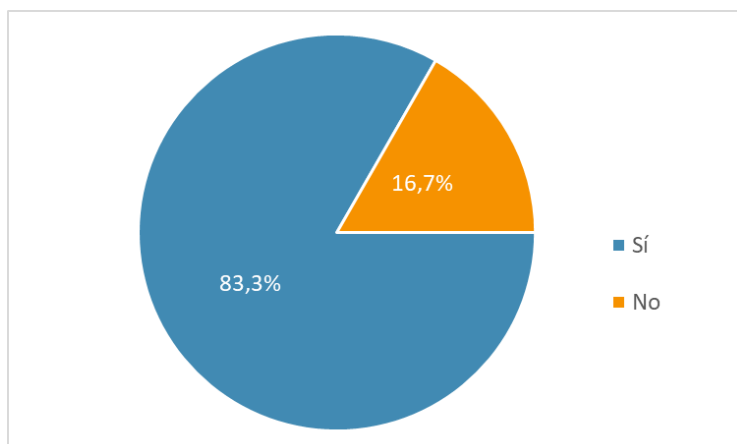


Gráfico 25. Búsqueda de información de bienes y servicios online en el año 2018

El nivel de estudios y la edad ha influido en el número de encuestados que busca información. Las personas sin estudios acabados han buscado información con menos frecuencia (un 71,7%) que los encuestados que tienen estudios primarios o superiores. El 86,0% con nivel educativo de primaria, el 89,8% con nivel educativo de bachiller, grado medio o grado superior, el 94,4% con estudios universitarios, y el 91,9% de las personas que tienen master o doctorado han buscado información sobre bienes y servicios a través de internet con más frecuencia (gráfico 26).

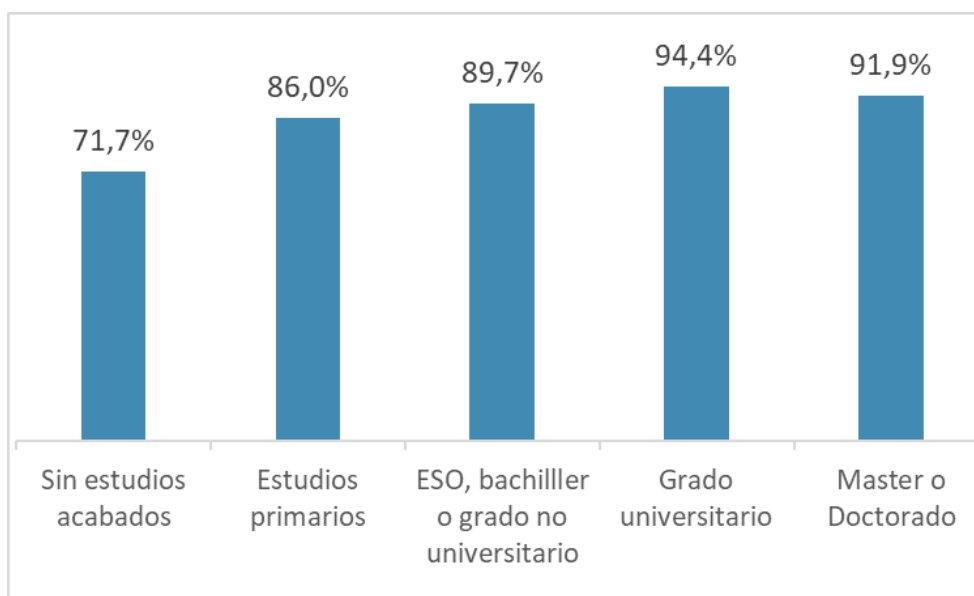


Gráfico 26. Búsqueda de información de productos y servicios online según el nivel de estudios

Por último, un 87,4% de los encuestados entre 25 y 34 años y un 88,1% de los encuestados de entre 35 a 44 años han buscado información de bienes y servicios, con más frecuencia. Mientras, los mayores de 64 años han buscado información de bienes y servicios con menos frecuencia (69,8%) (gráfico 27).

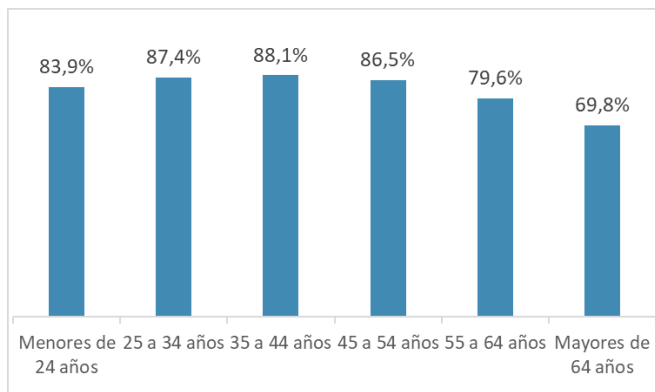


Gráfico 27. Búsqueda de información de productos y servicios online según la edad

El 56,0% de los encuestados han utilizado servicios de banca electrónica (gráfico 28). La edad y los estudios han influido en el uso de este servicio. Los encuestados sin estudios acabados utilizan la banca electrónica con menor frecuencia (32,4%). En cambio, el 71,1% de los encuestados con bachiller, grado medio o grado superior; el 80,1% de los que tienen estudios universitarios; y el 80,6% de los que han terminado los estudios de máster o doctorado utilizan la banca electrónica (gráfico 29).

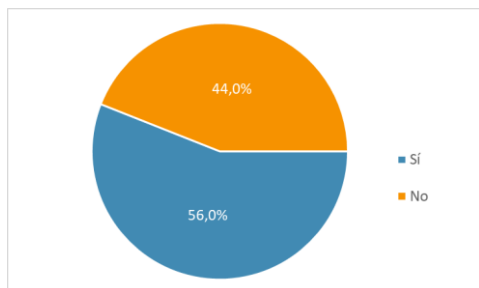


Gráfico 28. Utilización de la banca electrónica en el año 2018

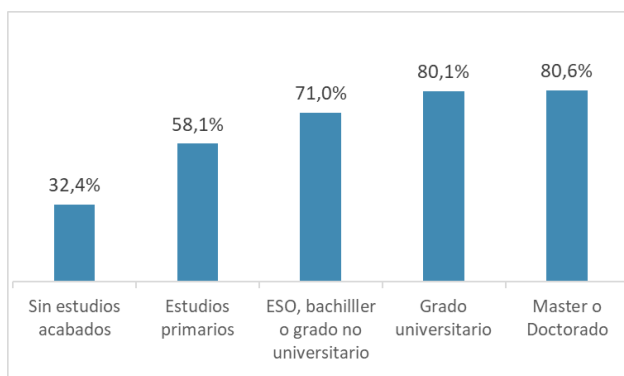


Gráfico 29. Uso de la banca electrónica según el nivel de estudio terminado

Los menores de 25 años y los mayores de 64 años han utilizado la banca electrónica con menos frecuencia de forma significativa: sólo el 34,8% de los menores de 25 años y el 39,4% de los mayores de 65 años han utilizado la banca electrónica. Por otro lado, el 67,9% de los encuestados de 25 a 45 años han utilizado la banca electrónica, siendo los que más usan de forma significativa este servicio (gráfico 30).

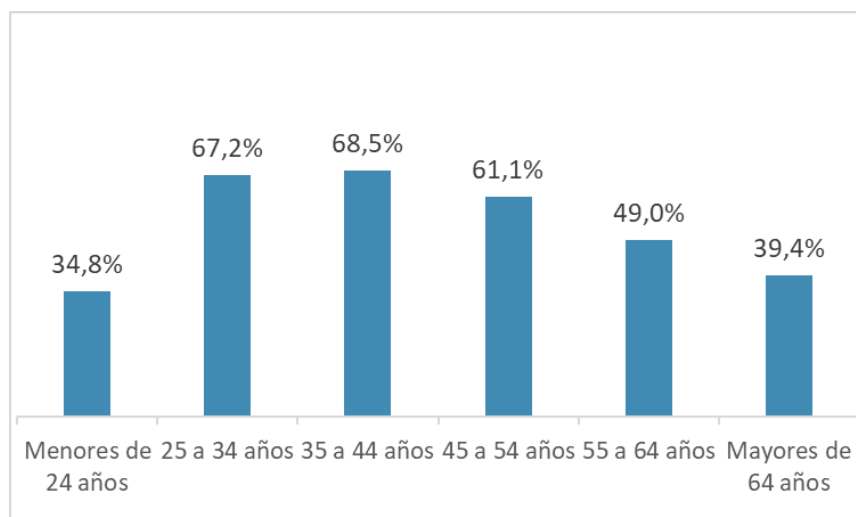


Gráfico 30. Utilización de la banca electrónico según la edad

4.1.7 Diagrama de árbol para la banca financiera en el 2018

La banca online es una de las innovaciones tecnológicas más importantes dentro del sector financiero. Para el desarrollo de este tipo de canales, existe “una creciente demanda de los clientes en estos canales, motivada por la escasez de tiempo, la reducción de costes o la comodidad”. Por otro lado, también existe la necesidad de impulsar este tipo de herramienta, para alcanzar una “una mayor cuota de mercado, dar una imagen de entidad sofisticada y desarrollada tecnológicamente” (Larran Jorge & Muriel de los Reyes, 2007).

Tanto el crecimiento de la banca online como el creciente uso de dispositivos nuevos para acceder a internet (en este caso, móviles y tablets) en los últimos 6 años podría suponer una evolución en el uso que se hace de las distintas herramientas disponibles, y que los dispositivos móviles (teléfonos móviles y tablets) pasen de ser considerado un canal alternativo a uno principal.

Desde el punto de vista del consumidor, se ha analizado qué tipos de persona en el año 2018 utilizan más este tipo de servicios. El 56% de los encuestados en el año 2018 han afirmado que utilizan la banca online. Se han considerado como significativas el nivel educativo, el uso del teléfono móvil para acceder a internet, el uso de la tablet para acceder a internet, y si se ha comprado a través de internet.

Se han analizado 23 nodos. De éstos, se explicarán los nodos 11, 12 y 15, porque son los que presentan un mayor porcentaje de uso de la banca financiera (imagen 1).

- El nodo 11 indica que el 89,6% de los encuestados que han comprado a través de internet, tienen los estudios de nivel de bachiller acabado y utilizan la tablet para acceder a internet han utilizado la banca financiera
- El nodo 12 indica que un 88,2% de los encuestados que han comprado online alguna vez, tienen un grado universitario, un máster o un doctorado y han utilizado el teléfono móvil para acceder a internet también utilizan la banca financiera como servicio online.
- El nodo 15 indica que el 81,2% de los encuestados que han comprado a través de internet, tienen estudios primarios, y acceden a internet a través de la Tablet utilizan el servicio online de banca electrónica.

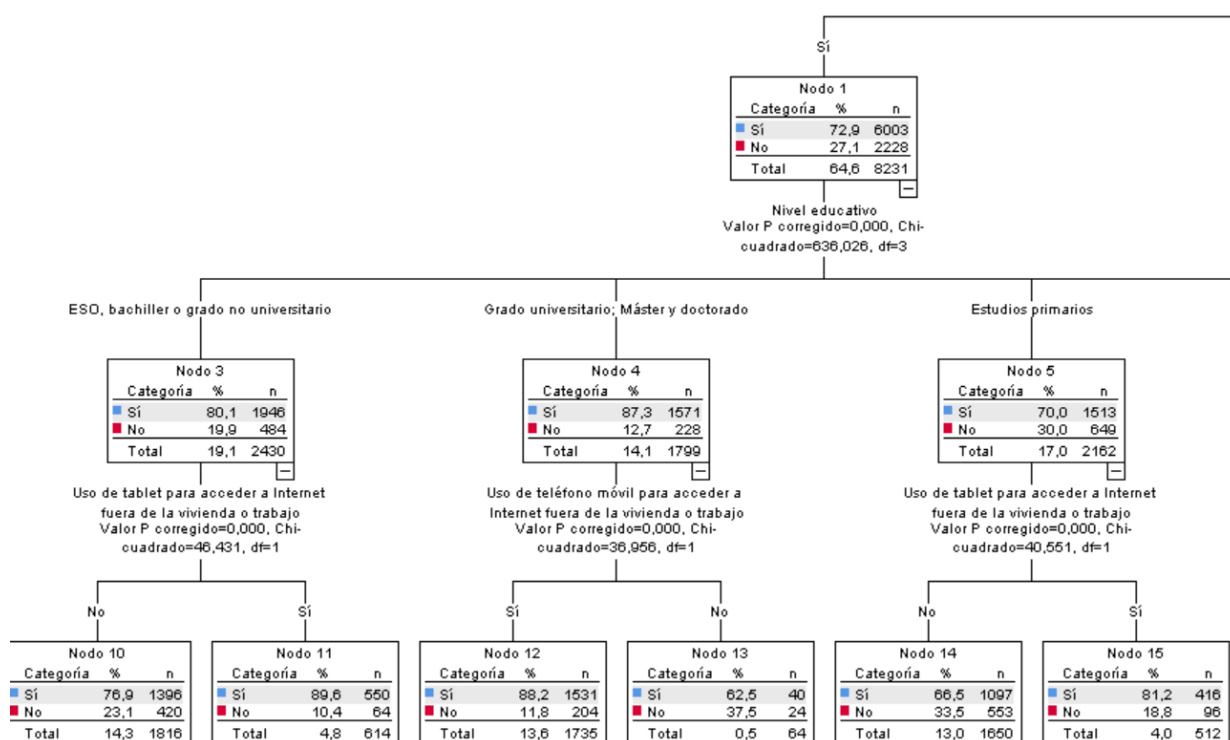


Imagen 1. Nodos 11, 12 y 15 del diagrama de árbol del uso de la banca electrónica (lado izquierdo del árbol, donde el sí indica que sí han comprado por internet).

Que un encuestado haya comprado online alguna vez parece indicar que también utiliza el servicio de la banca financiera. Los encuestados que al menos han finalizado la educación primaria utilizan más la banca electrónica que los que no tienen estudios. Por último, el uso del móvil y de la tablet para acceder a internet, frente al ordenador convencional, es un factor positivo para utilizar la banca electrónica.

Es por ello que quizá se deba tener en cuenta el diseño de aplicaciones de banca financiera para tablets y móviles. Esto incluye un diseño que se adapte a las nuevas formas de interactuar del usuario, y que se tenga en cuenta el nivel de educación y la edad del usuario de estos servicios.

4. 2 Análisis de contenidos de mashable.com

4.2.1 Análisis factorial

Se ha utilizado el análisis factorial en este trabajo para realizar un análisis exploratorio de los datos y para poder realizar a continuación un estudio de outliers (o de datos atípicos).

Los variables seleccionadas para este análisis han sido las variables “Numero de palabras en el título”, “Número de palabras en el contenido”, “Número de enlaces”, “Número de imágenes”, “número de videos”, “Número de palabras clave en los metadatos”, “Peor palabra clave (acciones mínimas)”, “Mejor palabra clave (acciones máximas)”, “Nivel de subjetividad absoluto en el título”, “Max. Acciones ‘share’ de los artículos referenciados en Mashable”, “mejor palabra clave (acciones promedio)”, “Peor palabra clave (acciones promedio)”, “Promedio de palabra clave (acciones mínimas)”, “Promedio de palabra clave (acciones máximas), “promedio de palabra clave (acciones promedio)” y “Min. Acciones ‘share’ de los artículos relacionados de Mashable”

Con el análisis factorial se han seleccionado 7 factores que explican el 73,2% de la varianza de los datos. Estos factores son:

Factor 1. Mejor palabra clave. Se refiere a las palabras clave por las cuales se ha encontrado más veces el artículo a través de buscadores. Incluye “mejor palabra clave (acciones máximas)”, “Mejor palabra clave (acciones promedio)” y “Peor palabra clave (acciones mínimas)” (relacionada de forma negativa).

Factor 2. Peor palabra clave. Se refiere a las palabras clave por las cuales se ha encontrado menos veces el artículo a través de buscadores. Incluye “peor palabra clave (acciones promedio)”, “promedio palabra clave (acciones promedio)” y “Promedio palabra clave (acciones máximas)”

Factor 3. Contenido del artículo. Explica la estructura del artículo, cómo está redactado, y su contenido. Incluye a las variables “número de imágenes”, “número de palabras en el contenido”, y “número de enlaces”.

Factor 4. Número palabras clave. Se refiere a la cantidad de palabras clave que tiene cada artículo y su promedio de visitas. Incluye la variable “numero de palabras clave en los metadatos” y “Promedio de palabras clave (acciones mínimas)”.

Factor 5. Número de veces que los enlaces se han compartido. Este factor relaciona las veces que se ha compartido un artículo y los artículos enlazados a éste. Incluye las variables “Min. Acciones ‘share’ de los artículos relacionados de Mashable” y “Max. Acciones ‘share’ de los artículos relacionados de Mashable”.

Factor 6. Neutralidad del título. Este factor explica la importancia del título y la subjetividad entendida por los lectores. Incluye las variables “nivel de subjetividad absoluta del título” (relacionada de forma negativa), y “número de palabras en el título”

Factor 7. Número de vídeos en el artículo. Incluye la variable “número de vídeos”.

4.2.2 Análisis de outliers

Se ha realizado una serie de gráficos de dispersión 3D con los componentes extraídos del análisis factorial para determinar cuáles de los artículos se tratan de casos outlier, que son los casos 16295, 12301, 17138 y 12682, que se explicarán a continuación.

Comparando los tres primeros factores, se puede observar que existen casos atípicos. El que más destaca es el del artículo 16282 “*Randi Zuckerberg: The Best Thing About Social Media Is Also the Worst*”. Este artículo puntúa bajo en las variables relacionadas con el contenido del artículo y puntúa muy alto en las relacionadas con mejor palabra clave y peor palabra clave. Otros artículos que siguen esta tendencia son el caso 17138 “*12 'Sound of Music Live' Performances You Should Rewatch*” y el caso 16295 “*The 15 greatest videogame sidekicks of all time*”, artículos cuyas mejores palabras clave han conseguido que los usuarios encuentren el artículo más veces, pero el resto de palabras clave han conseguido muy pocos hits. Es decir, el volumen de visitas a estos dos artículos ha dependido en mayor medida que el resto de artículos de sus mejores palabras clave. Estos artículos tienen también menos palabras clave que la media (gráfico 31).

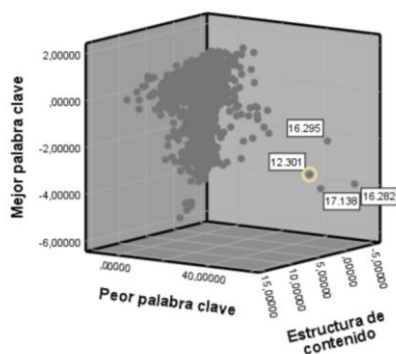


Gráfico 31. Dispersión de casos según su estructura, mejor palabra clave, y peor palabra clave

4.2.3 Momento de publicación

La web de *mashable.com* ha clasificado sus artículos según 5 categorías: entretenimiento, estilo de vida, noticias y temas del mundo, ciencia y tecnología. Los artículos de *mashable.com* se publican según el criterio de sus editores. Según Dobeles (2005), el momento de publicación es un punto crucial a la hora de que un artículo sea popular. Por ello, se ha analizado si existe una relación entre el tipo de publicación y el día de la semana en que el artículo se ha publicado.

Los artículos con referencia a estilo de vida se han publicado con más frecuencia los miércoles y con menos frecuencia los sábados y domingos (gráfico 32).

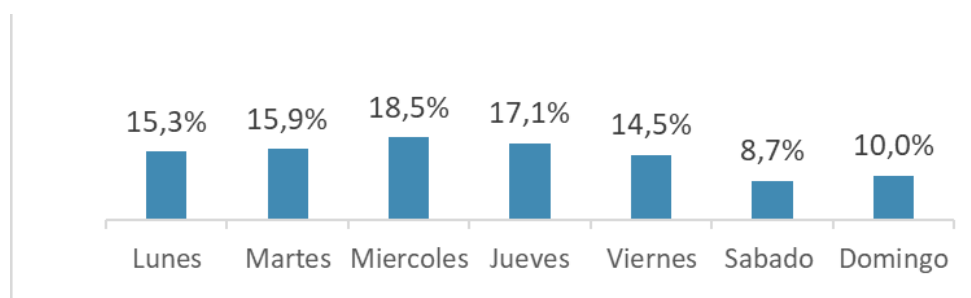


Gráfico 32. Porcentaje de artículos de estilo de vida que se han publicado durante la semana

Por otro lado, los relacionados con el entretenimiento se han publicado con más frecuencia el lunes y el domingo, y con menos frecuencia el jueves y sábado (gráfico 33).

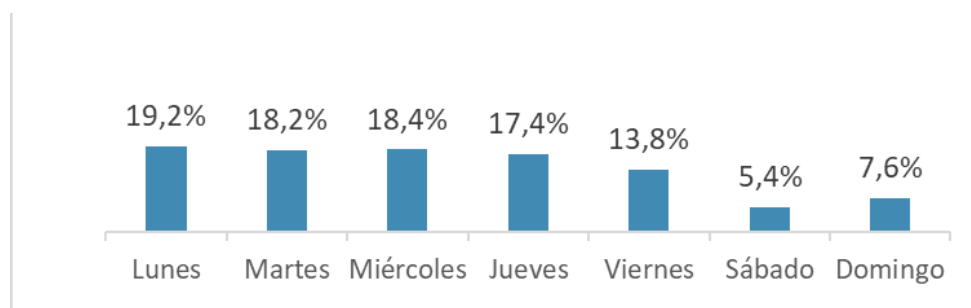


Gráfico 33. Porcentaje de artículos que se han publicado durante la semana

Lunes, miércoles y jueves se han publicado con más frecuencia post con temática de negocios, mientras que se publican con menos frecuencia el viernes, sábado y domingo (gráfico 34).

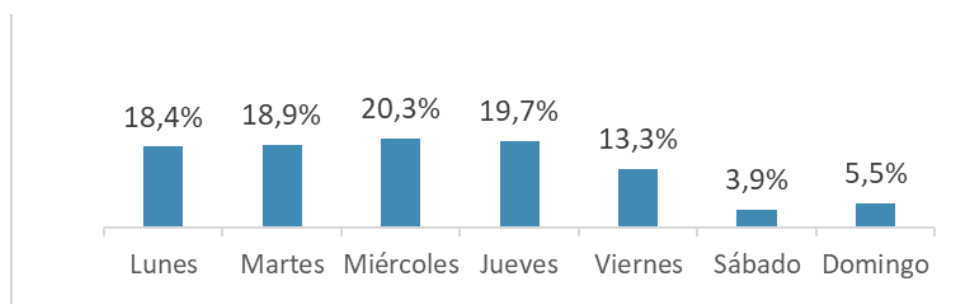


Gráfico 34. Porcentaje de artículos de negocios que se han publicado durante la semana

Las publicaciones de redes sociales se han publicado con más frecuencia los martes, los jueves y los sábados (gráfico 35).

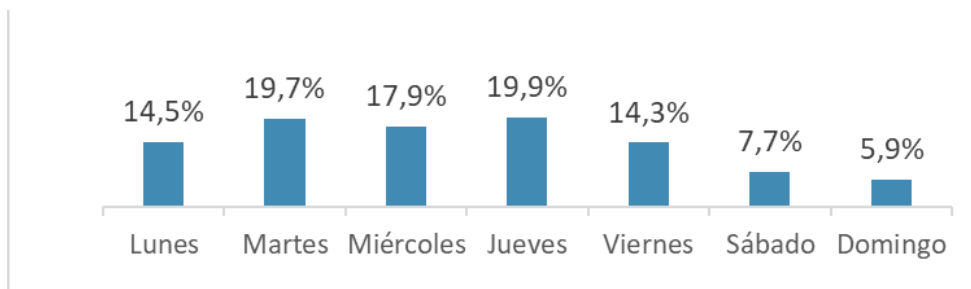


Gráfico 35. Porcentaje de artículos de redes sociales que se han publicado durante la semana

Los martes, miércoles y sábado son los días que más se publica contenido relacionado con la tecnología frente al resto de días (gráfico 36).

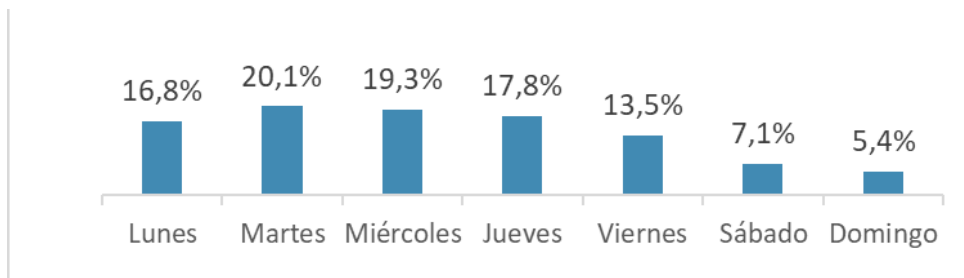


Gráfico 36. Porcentaje de artículos de tecnología que se han publicado durante la semana

Finalmente, las publicaciones sobre noticias globales se publican con más frecuencia el miércoles, jueves, viernes, sábado y domingo (gráfico 37). Aunque hay que tener en cuenta que las noticias se deben publicar cuando sucedan, con lo que el análisis de momento de publicación puede que no es tan importante

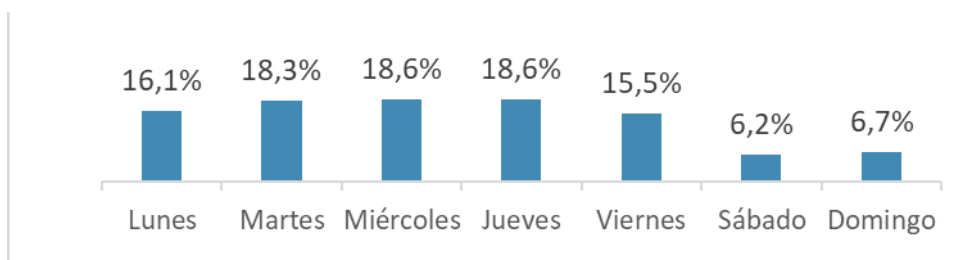


Gráfico 37. Porcentaje de artículos de noticias que se han publicado durante la semana

En general, los artículos se publican con menos frecuencia los fines de semana que durante la semana.

4.2.4 Popularidad

En esta sección se pretende conocer cuáles pueden ser los factores por los cuales un artículo se vuelve popular o no. Por lo tanto, se considera que el 53,4% del total de artículos han sido populares, frente al 46,4% que no (gráfico 38).

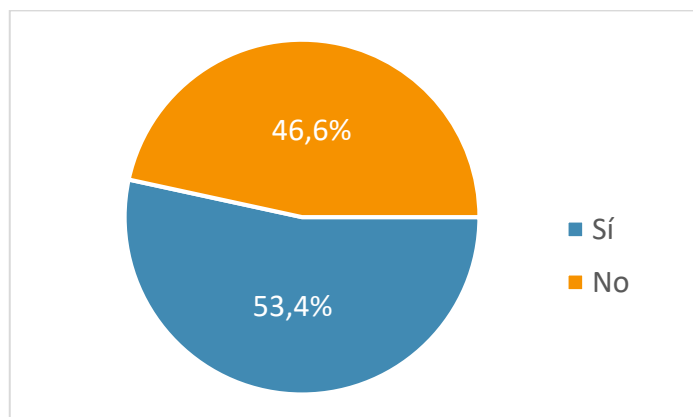


Gráfico 38. Artículos que han sido populares

Existe una relación entre el número de veces que se ha compartido un artículo y el día de publicación del mismo. Los artículos publicados el viernes, sábado y domingo se han compartido con más frecuencia que el resto de días (gráfico 39).

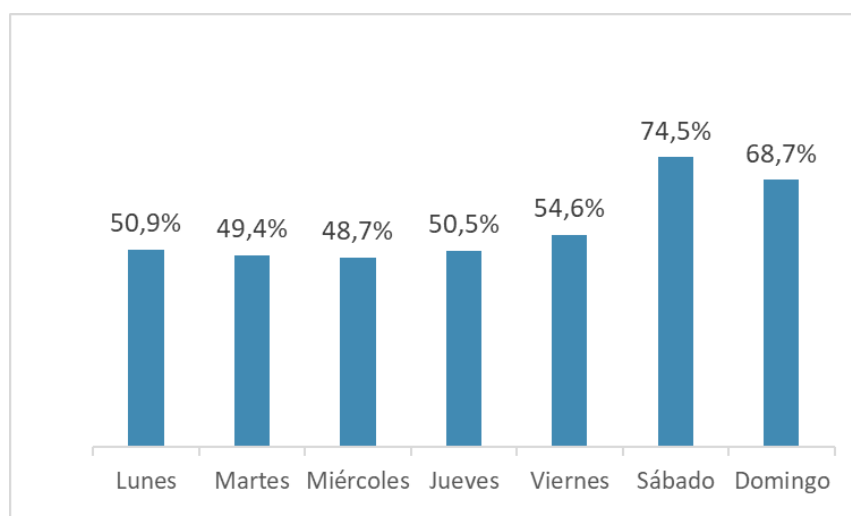


Gráfico 39. Popularidad según el día de publicación

Esto coincide con el artículo de Sall (2014), donde se afirma que los artículos publicados durante el fin de semana tienen mayor número de visitas, mayor número de hits, y se comparten más.

A continuación, se va a analizar si los artículos publicados durante el fin de semana se han compartido más a menudo que el resto de artículos. Se ha realizado una comparación de medias del número de veces que los usuarios han compartido un artículo según éste se haya publicado en fin de semana o no.

La media de veces que se ha compartido un artículo, independientemente de cuándo se publicó, es de 3395 veces. Los artículos publicados el fin de semana se han compartido una media de 3903 veces, y los artículos que se han publicado de lunes a viernes se han compartido una media de 3318 veces. Así, parece ser que los artículos publicados en fin de semana se comparten más a menudo que los artículos publicados entre semana (gráfico 40).

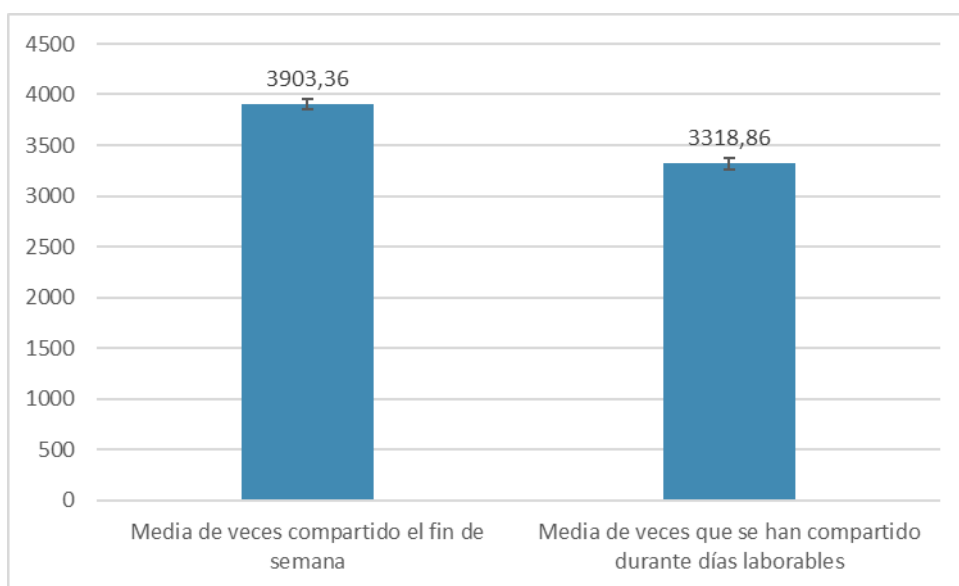


Gráfico 40. Media de veces que se ha compartido un artículo en fin de semana frente a días laborables. La barra vertical es ± 1 error estándar

El número de shares logrados por cada artículo está muy relacionado con el tipo de publicación que es. Los artículos sobre el mundo, del ámbito del entretenimiento, y del ámbito de los negocios se han compartido menos. Mientras tanto, los artículos sobre tecnología, estilo de vida y redes sociales se han compartido más (gráfico 41).

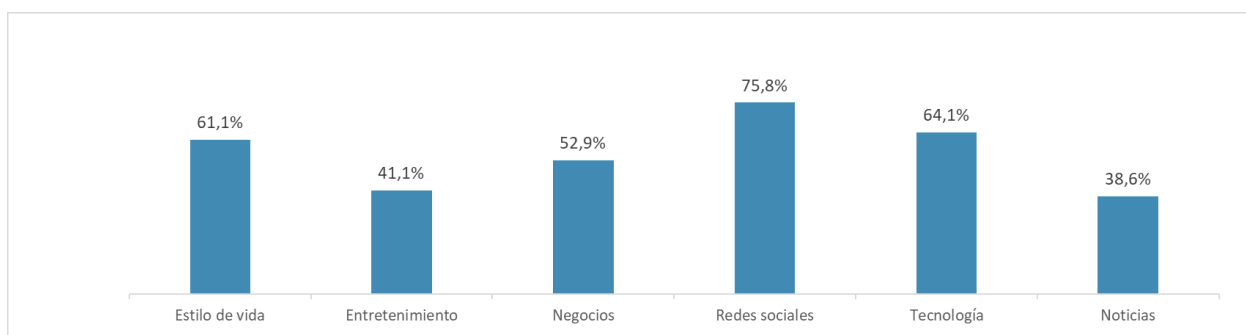


Gráfico 41. Popularidad según el tipo de artículo

La polaridad del artículo, referido a si el artículo se considera que es positivo, neutro o negativo, es otro factor que afecta a la popularidad del artículo: un 56,4% de los artículos positivos se han considerado populares, con lo que se han vuelto populares con mayor frecuencia. Los artículos negativos se han compartido con menor frecuencia: sólo el 48,8% de los artículos negativos han sido populares (gráfico 42).

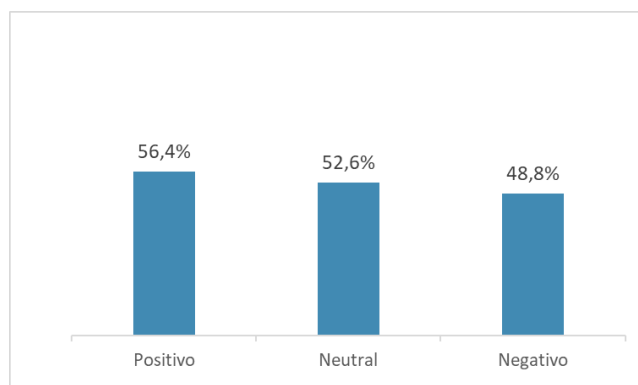


Gráfico 42. Popularidad según la polaridad del artículo

4.2.5 Diagrama de árbol para la popularidad de un artículo

En el siguiente análisis, se pretende buscar una posible estrategia de creación, redacción y gestión de artículos para conseguir que éstos se vuelvan populares o virales en la red de forma eficiente. Es una de las formas más novedosas en la metodología de análisis de contenidos desde el punto de vista de estrategias de marketing. Para definir esta estrategia, se usará un diagrama de árbol para encontrar relaciones entre eventos y para clasificar, segmentar los datos y predecir los posibles sucesos en cada toma de decisiones. Se han considerado como variables relevantes si se publicó el artículo el fin de semana, la estructura del contenido, mejor palabra clave y el número de palabras en los metadatos.

Se han analizado 18 nodos, también denominados situaciones de decisión. De éstos, los nodos con mayor frecuencia de artículos populares son el 15, el 18, y el 16 (imagen 2).

- El nodo 15 incluye los artículos que no se publicaron en fin de semana, con escasas palabras clave en los metadatos, y con pocas acciones en la mejor palabra clave. De estos artículos, el 83,0% de ellos fueron populares.
- El nodo 18 incluye los artículos que no se publicaron el fin de semana, con pocas palabras claves en los metadatos, y con pocas acciones en las mejores palabras claves. De estos, el 79,1% fueron populares.

- El nodo 17 incluye los artículos que no se publicaron en fin de semana, con muchas palabras clave en los metadatos, con muchas acciones en su mejor palabra clave, y con una calidad de la estructura de mala a buena. De estos, el 71,8% fueron populares.

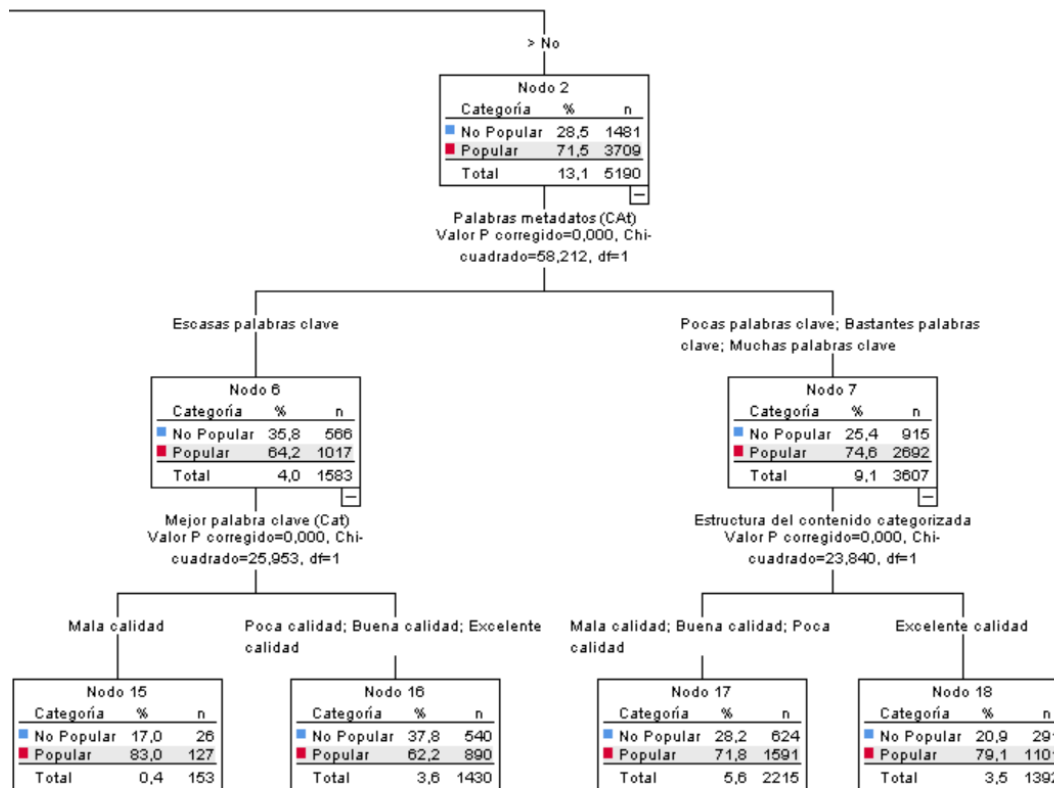


Imagen 2. Nodos 15, 18 y 17 del diagrama de árbol de la popularidad de un artículo (lado derecho del árbol, donde el no indica que el artículo no se ha publicado el fin de semana).

Hay que destacar que, aunque en general los artículos publicados el fin de semana son los que se han compartido más entre los usuarios, existen otros factores que podrían determinar si un artículo es popular o no: en particular el número de palabras claves, la estructura del contenido, o las acciones realizadas en la mejor palabra clave. Sall M. (2014) mencionaba que, aunque el momento de publicación es importante, hay otros factores que influyen más, como la calidad del artículo, el número de seguidores o cómo se interactúa con estos últimos, lo que concuerda con nuestro resultado.

Para mejorar en el uso y gestión de las palabras clave, es necesario aprender a utilizar herramientas que mejoren el SEO (optimización en buscadores web, por sus siglas en inglés). Este tipo de estrategias son importantes porque dan visibilidad a una página web, y puede implicar que llegue más tráfico y más oportunidades para encontrar y fidelizar clientes. Además, tiene su valor como herramienta para crear notoriedad de marca y posicionarse como un sitio web autoritario y de confianza (Herrera Damas & Requejo Aleman, 2012).

5. Conclusiones

Desde hace 10 años, el uso de internet en las viviendas es prácticamente universal. Sin embargo, la forma en cómo lo usamos ha ido cambiando en los últimos 6 años.

Para empezar, mientras que el uso del móvil y tablets está creciendo, el uso de ordenadores de sobremesa y portátil para acceder a internet está bajando.

El perfil del comprador estándar según los resultados es un hombre mayor de 35 años y con grado universitario. Son el grupo que más productos compran en todas las categorías estudiadas, y los que más dinero se gastan en ello. Curiosamente, los hombres compran más a través de internet que las mujeres, incluso en bienes del hogar y consumo, donde se esperaría que las mujeres consumieran más.

El nivel de estudios y la edad influye en el número de encuestados que utiliza banca electrónica. Mas del 71.1% de los encuestados con titulación superior al bachiller utilizan la banca electrónica y la mayor frecuencia de uso se da en el rango de edad de 25 a 45 años.

Las personas que tienen un grado universitario buscan información con más frecuencia en la red que el resto de personas. Además, el valor medio de sus compras online también es más alto. En general, el valor de las compras online realizadas por una persona aumenta cuanto mayor sea el nivel de estudios que tenga la misma.

Parece ser que la popularidad de un artículo en la red depende, principalmente, de cuatro factores. El primero de ellos es la calidad del contenido del artículo, evaluada a través de la longitud del artículo, el número de enlaces a otros artículos, y el número de imágenes. El segundo factor son las palabras clave en los metadatos, que deben de ser escogidas cuidadosamente según el tipo de noticia. En tercer lugar, hay que cuidar de la subjetividad del artículo. En general, las personas reaccionan más ante artículos negativos. Pero según los análisis realizados en los artículos de *mashable.com*, al final los usuarios prefieren compartir contenido que es positivo o, al menos, neutro. Por último, hay que tener en cuenta el día de publicación del artículo. En general, parece que los usuarios ven y comparten artículos el fin de semana. Pero durante la semana laboral, el día de publicación parece depender del tipo de artículo que se publique.

6. Recomendaciones

1. Parece que cada vez más usuarios están cambiando los ordenadores por los móviles, y por tanto es necesario para el mantenimiento de una buena relación entre empresa y cliente crear sistemas menos frustrantes de utilizar. Por ello, es importante adaptar la página web empresarial y los servicios o productos que se ofrezcan de forma online a la interfaz de los móviles modernos.
2. El comercio electrónico y el volumen de compra aún sigue creciendo, con lo cual puede ser una buena oportunidad para encontrar nuevos clientes o nuevas formas de fidelización de clientes.
3. Aunque el grupo que parece interaccionar más online sean los varones de entre 35 y 45 años con ingresos altos, puede resultar un grupo de clientes que no interese en particular a una empresa. Hay que tener en cuenta la segmentación de clientes realizada por la empresa y a cuáles de esos segmentos interesa atraer, fidelizar o mantener.
4. El crecimiento del comercio electrónico también puede implicar la aparición de nuevas formas de pago online, o el aumento de clientes que vayan a registrarse en esos servicios. Es importante para las entidades financieras tener un servicio bueno y fiable para este tipo de clientes.
5. El denominado marketing de contenidos se considera actualmente una de las formas de ganar autoridad y calidad de imagen para una empresa. Por ello, es recomendable tener una estrategia de marketing de contenidos anual.

7. Limitaciones y futuras líneas de trabajo

1. Aunque se pueden extraer las estadísticas de cuánto se usa internet o qué servicios y productos se compran, es posible que, según algunas variables sociodemográficas, no se esté teniendo en cuenta el cómo lo usa cada uno de estos grupos.
2. Algunas variables sociodemográficas pueden ser interesantes para este estudio. Por ejemplo, el tipo de ocupación o estudios de cada caso, para poder hacer una serie de perfiles más precisos.
3. La base de datos de *mashable.com* sólo da información sobre sus artículos. Una línea de trabajo futura interesante sería averiguar información sobre los propios usuarios, e intentar indagar cuándo, cómo y por donde se están viendo o compartiendo los artículos. También sería posible evaluar qué factores sociodemográficos están influyendo en su navegación por el portal.

8. Bibliografía

Abela, J. A. (febrero de 2018). *Las técnicas de análisis de contenido: una revisión actualizada*. Obtenido de Mastor: <http://mastor.cl/blog/wp-content/uploads/2018/02/Andreu.-analisis-de-contenido.-34-pags-pdf.pdf>

Álvarez, D. (2014). *Big Data. La nueva dimensión de los datos*. Recuperado el 22 de Julio de 2019, de Lewis & Carroll: <https://lewisandcarroll.com/big-data/>

Bessis, N., & Dobre, C. (2014). *Big Data and Internet of Things: a roadmap for smart environments*. Springer.

Camara de Comercio de Zaragoza. (2018). *Cámara Zaragoza*. Obtenido de <https://www.camarazaragoza.com/actualidad/los-comercios-electronicos-de-aragon-venden-un-30-mas-en-el-ultimo-ano/>

Fabián Abad, A. (2017). *Repositorio institucional de la universidad de Zaragoza*. Recuperado el 2019, de <https://zaguan.unizar.es/record/65133?ln=es>

Herrera Damas, S., & Requejo Aleman, J. (2012). 10 Good Practices for News Organizations Using Twitter. *Journal of Applied Journalism & Media Studies* , 79-95.

Kibernum. (2018). *TICs en la banca: ¿Cómo han impactado al sector?* Obtenido de <https://www.kibernum.com/noticias/tics-en-la-banca-como-han-impactado-al-sector/>

Larran Jorge, M., & Muriel de los Reyes, M. (2007). La banca por internet como innovación tecnológica en el sector bancario. *Investigaciones Europeas de Dirección y economía de la empresa*, Vol 13, 2. Obtenido de <file:///C:/Users/marta/Downloads/Dialnet-LaBancaPorInternetComoInnovacionTecnologicaEnElSec-2356670.pdf>

Luque, C. (2019). *Plataforma de gobierno abierto*. Obtenido de <https://www.ogoov.com/es/blog/open-data-y-fair-data-diferencias-y-similitudes/>

Manyika, J. (Mayo de 2011). *Big Data: the next frontier for innovation, competition and productivity*. Obtenido de https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx

Nguyen, M. (2019). *Content Marketing Strategy For a Highly Specialized B2B Niche Company*. Obtenido de Hame university of applied science: <https://www.theseus.fi/bitstream/handle/10024/227394/Phuong.nguyen.BBIBP16A7.final.thesis.pdf?sequence=2&isAllowed=y>

Olave, P. (Noviembre de 2019). Big Data: La fiebre del siglo XXI. *conCiencias*. Revista de la Facultad de Ciencias. Universidad de Zaragoza.

Opresnik, D. (Julio de 2015). The value of big data in servitization. *International Journal of Production Economics*, 174-184.

Powerdata. (2018). *Centro de recursos de Big Data*. Obtenido de <https://www.powerdata.es/big-data>

Sall, M. (9 de diciembre de 2014). *Medium*. Obtenido de <https://medium.com/data-lab/when-is-the-best-time-to-publish-wrong-question-8f0b15be89c2>

Sanchez Medina, A. J., Melián Gonzalez, A., & Hormiga Pérez, E. (2007). El concepto del capital intelectual y sus dimensiones. *Investigaciones europeas de economía y empresa*, 97-111.