



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

ADE

Facultad de Administración
y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Análisis de tendencias en contenidos de sitios web de
empresas

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Reyes Vázquez, Ignacio

Tutor/a: Doménech i de Soria, Josep

CURSO ACADÉMICO: 2022/2023

ÍNDICE DE CONTENIDOS

RESUMEN	5
CAPÍTULO 1. INTRODUCCIÓN	6
1.1 Motivación	7
1.2 Objetivos del TFG	7
1.3 Estructura	8
CAPÍTULO 2. MARCO TEÓRICO Y CONTEXTUAL	9
2.1 Introducción	10
2.2 Internet y las empresas	10
2.2.1 Introducción	10
2.2.2 Sitios web	11
2.2.3 Tendencias en los sitios web	12
2.3 Redes Sociales	13
2.3.1 Introducción	13
2.3.2 Evolución de las RRSS	14
2.3.3 Clasificación de las RRSS	15
2.4 COVID-19	17
2.4.1 Introducción	17
2.4.2 Impacto en las empresas	18
2.4.3 La sociedad y la información	20
2.5 Lenguajes empleados para el estudio	21
2.5.1 R y RStudio	21
2.5.2 SQL	22
2.6 JRC	23
CAPÍTULO 3. METODOLOGÍA	25
3.1 Introducción	26
3.2 Construcción de la muestra	26
3.3 Agregación de los estratos de la muestra	27
3.4 Variables de la muestra	28
3.5 Formulación de hipótesis	31
3.6 Métodos estadísticos	32
CAPÍTULO 4. VISUALIZACIÓN	35
4.1 Objetivo de la visualización	36
4.2 Organización de la visualización	36
4.2.1 Gráfico de países	36
4.2.2 Gráfico de sectores económicos	36
4.2.3 Gráfico de tamaños de empresa	37
4.2.4 Mapa de Europa	37
4.3 Implementación de gráficos estáticos	37
4.3.1 Introducción a “ggplot2”	37
4.3.2 Resultado de la implementación	38

4.4 Gráficos interactivos	40
4.4.1 Motivación para la creación de gráficos interactivos	40
4.4.2 Evaluación de alternativas	41
4.4.3 Implementación final	43
4.4.4 Demostración final	46
CAPÍTULO 5. RESULTADOS	47
5.1 Análisis descriptivo	48
5.2 Análisis del uso conjunto de redes sociales	51
5.3 Análisis comparativo	52
5.4 Modelo multivariante	60
CAPÍTULO 6. CONCLUSIONES	64
BIBLIOGRAFÍA	66
ANEXO I. RELACIÓN DEL TRABAJO CON LOS ODS	70

ÍNDICE DE FIGURAS

Figura 1. Evolución del porcentaje de ciudadanos españoles que han utilizado Internet al menos 5 días a la semana. _____	11
Figura 2. Porcentaje de empresas por C.A. en España con conexión a Internet y con página web. _____	12
Figura 3. Porcentaje de empresas españolas por sector que han utilizado en el 2022 las RRSS. _____	14
Figura 4. Millones de usuarios de las RRSS en los meses de Enero de cada año. _____	15
Figura 5. Millones de usuarios contabilizados al comienzo del año 2023 en las cinco RRSS mostradas. _____	17
Figura 6. Evolución del número de contagios (en miles) en España de la enfermedad COVID-19 _____	18
Figura 7. Porcentaje de teletrabajo en España desde el año 2018 al 2022. _____	19
Figura 8. Búsquedas en Internet sobre información relativa al COVID-19 (en relevancia de búsquedas). _____	20
Figura 9. Captura de pantalla de una consulta ejemplo en lenguaje SQL en R. _____	23
Figura 10. Gráfico realizado a través del paquete "ggplot2", desagregado por países de empresa. _____	38
Figura 11. Gráfico realizado a través del paquete "ggplot2", desagregado por tamaños de empresa. _____	39
Figura 12. Gráfico realizado a través del paquete "ggplot2", desagregado por sector de empresa. _____	39
Figura 13. Mapa geográfico realizado a partir del paquete "ggplot2". _____	40
Figura 14. Gráfico de ejemplo realizado con Plotly sobre la evolución del uso de los términos “educación” en los sitios web de los 4 países. _____	41
Figura 15. Gráfico de ejemplo realizado con GoogleVis sobre la evolución del uso de los términos “educación” en los sitios web de los 4 países. _____	42
Figura 16. Gráfico de ejemplo realizado con High Charts sobre la evolución del uso de las RRSS en los sitios web de los 4 países. _____	43
Figura 17. Gráfico realizado con Plotly, desagregado por tamaño de empresa de la evolución del término educación. _____	44
Figura 18. Gráfico realizado con Plotly, desagregado por sector de empresa de la evolución del término educación. _____	45
Figura 19. Mapa interactivo generado con Plotly con los porcentajes de empresas de cada país a lo largo del tiempo. _____	45
Figura 20. Gráfico de líneas de la evolución del uso de la variable RRSS en los sitios web de los cuatro países. _____	56

ÍNDICE DE TABLAS

Tabla 1. Ponderaciones por país de origen de la empresa _____	27
Tabla 2. Ponderaciones por tamaño de empresa _____	27
Tabla 3. Ponderaciones por sector de la empresa _____	28
Tabla 4. Sectores económicos recodificados propios de la muestra _____	29
Tabla 5. Frecuencias relativas y absolutas de los niveles de la variable "Sector" _____	48
Tabla 6. Frecuencias relativas y absolutas de los niveles de la variable "Año" _____	49
Tabla 7. Frecuencias relativas y absolutas de los niveles de la variable "País", "Tamaño", "Covid" y "Redes Sociales" _____	50
Tabla 8. Frecuencias relativas y absolutas de los niveles de las variables propias de las redes sociales _____	51
Tabla 9. Matriz de coeficientes phi de las diferentes redes sociales _____	52
Tabla 10. Tabla de contingencia entre las variables "Covid" y "Tamaño" en los años referentes a la pandemia _____	53
Tabla 11. P. Valores de las pruebas de proporción I _____	53
Tabla 12. Tabla de contingencia entre las variables "Covid" y "País" en los años referentes a la pandemia _____	54
Tabla 13. P. Valores de las pruebas de proporciones II _____	54
Tabla 14. Tabla de contingencia entre las variables "Covid" y "Año" en los años referentes a la pandemia _____	55
Tabla 15. Tabla de contingencia entre las variables "Covid" y "Sector" en los años referentes a la pandemia _____	55
Tabla 16. Tabla de contingencia entre las variables "Redes Sociales" y "País" _____	56
Tabla 17. P. Valores de las pruebas de proporciones III _____	57
Tabla 18. Tabla de contingencia entre las variables "Sector" y "Redes Sociales" _____	58
Tabla 19. P. Valores de las pruebas de proporciones IV _____	58
Tabla 20. Tabla de contingencia entre las variables "Redes Sociales" y "Tamaño" _____	59
Tabla 21. Tabla de contingencia entre las variables "Años" y "Redes Sociales" _____	59
Tabla 22. Estimaciones del modelo de regresión logística _____	60

RESUMEN

El contenido de sitios web de empresas se puede utilizar para ayudar a la monitorización del estado de la economía, de la situación actual del tejido productivo y para evaluar las posibles prioridades y preocupaciones de la sociedad.

El objetivo del proyecto es la puesta en marcha de una herramienta de inteligencia de negocios para observar las tendencias en los sitios web de empresas. Para ello, se deberá definir una muestra de las empresas a analizar, se descargará cada sitio web y se indexará en una base de datos, y se creará un sistema de visualización de los resultados obtenidos.

Finalmente, esta herramienta será empleada para el estudio de dos temas de interés en la sociedad actual: el uso de las redes sociales por parte de las empresas y su evolución en el tiempo, y la capacidad de adaptación de las empresas a la situación de la pandemia provocada por la enfermedad del COVID-19.

ABSTRACT

The content of company websites can be used to help monitor the state of the economy, the current situation of the productive fabric and to assess the possible priorities and concerns of society.

The aim of the project is to implement a business intelligence tool to observe trends in company websites. To do this, a sample of companies to be analysed will be defined, each website will be downloaded and indexed in a database, and a system will be created to visualise the results obtained.

Finally, this tool will be used to study two topics of interest today: the use of social networks by companies and their evolution over time, and the ability of companies to adapt to the pandemic situation caused by the COVID-19 disease.

CAPÍTULO 1. INTRODUCCIÓN

CAPÍTULO 1. INTRODUCCIÓN

1.1 Motivación

La motivación para llevar a cabo este estudio radica en la importancia de comprender los comportamientos de las empresas frente a cualquier situación relevante en la actualidad. Además, conseguir comprender este comportamiento de las empresas, a través de la información que facilitan estas, en sus sitios web, de tal forma que siempre pueda verse actualizada de una forma sistemática.

Para la realización de este proyecto, al abarcar un campo extremadamente amplio, se ha decidido delimitar la observación y el estudio a dos de los ítems más importantes en la sociedad durante estos últimos años: las redes sociales y la enfermedad del COVID-19.

En primer lugar, las redes sociales se han convertido en una revolución para todas las empresas del mundo, la forma de publicitarse a través de ellas, de acercarse más a los clientes y de proporcionar una sensación de prestar más atención a estos, ha hecho que cada día más empresas hagan uso de estas. Día a día la evolución de las redes sociales sigue su curso, tanto con la generación de unas nuevas, como con el abandono del uso de muchas otras. Es por esto, una de las razones que se ha decidido incluir este término para observar su evolución a lo largo del tiempo.

En cuanto a la enfermedad del COVID-19, la importancia que ha tenido en los últimos años en la sociedad mundial hace más que necesario un estudio de como las empresas se han intentado gestionar frente a esta pandemia mundial.

Por último y como es obvio, el estudio de todos los términos en los que existe interés es imposible. Por esto, la gran motivación de la realización de este Trabajo Final de Grado es la creación de una herramienta capaz de realizar las búsquedas deseadas por el usuario y así poder abarcar cualquier término deseado.

1.2 Objetivos del TFG

El objetivo principal de este Trabajo Final de Grado no es otro que analizar las tendencias en el contenido de los sitios web de empresas con el fin de comprender cual es el comportamiento de estas hacia el exterior. En consecuencia y tal como se expone en el resumen del trabajo, se han propuesto los siguientes objetivos, los cuales han sido establecidos como determinantes para medir el éxito de este:

1. Creación de una herramienta de inteligencia de negocios para el estudio de las tendencias en sitios web de empresas a través del lenguaje de programación R y de consultas SQL.
2. Definición, extracción y tratamiento de la muestra de empresas seleccionada.
3. Estudiar la relación entre el uso de redes sociales, la enfermedad del COVID-19 y las páginas web de las empresas.

1.3 Estructura

Para la determinación y el estudio de los diferentes patrones de comportamiento y tendencias en los sitios web de las empresas, se ha dividido el estudio en dos grandes partes el trabajo: una primera sección cuyo objetivo es contextualizar y detallar los conceptos más relevantes en él, y otra, centrada en los análisis y pruebas estadísticas realizadas para estudiar la muestra.

En primer lugar, dentro del capítulo 2, se expondrá en profundidad, el contexto asociado al impacto de Internet en la sociedad y en los sitios web de las empresas, cómo comenzó, y cuál ha sido su evolución a lo largo del tiempo. Dentro de esta contextualización se añadirán a continuación los conceptos propios seleccionados para el estudio, las redes sociales y la enfermedad del COVID-19. Así se ha proporcionado una visión en cuanto a la evolución de la importancia de estos, tanto en la sociedad como en las empresas y por ello en sus sitios web.

Por último, dentro de este capítulo se hace un breve resumen del programa ligado a la realización de este TFG, programa en el que se continúa trabajando, y por ello se ha realizado una breve explicación sobre el proyecto JRC.

Por otro lado, en cuanto a los capítulos 3, 4 y 5 se ha realizado un estudio de carácter numérico y estadístico en relación con los datos para cada una de las cuestiones consideradas. En el capítulo 3 se detallará el muestreo realizado, las variables propias del estudio y los métodos estadísticos utilizados a lo largo del mismo.

En cuanto al capítulo 4, a continuación, se detalla cual ha sido el proceso de depuración de los datos obtenidos de la base de datos, y, cuál ha sido su tratamiento. Por último, se explica el razonamiento del software elegido para la creación de los gráficos del proyecto, además de una serie de ejemplos de estos.

Finalmente, en el capítulo 5 se interpretan los resultados que se han obtenido, elaborando una síntesis del trabajo y detallando en profundidad las conclusiones alcanzadas en el capítulo 6. Adicionalmente, se incluirá al final del documento la bibliografía y referencias utilizadas como documentación para la elaboración del trabajo y un anexo de los ODS.

CAPÍTULO 2. MARCO TEÓRICO Y CONTEXTUAL

CAPÍTULO 2. MARCO TEÓRICO Y CONTEXTUAL

2.1 Introducción

Antes de adentrarnos en el análisis práctico que constituye el objeto de este trabajo, se establecerá el entorno en el que se desarrollará dicho análisis. Dentro del estudio, existen tres conceptos esenciales que son relevantes para comprenderlo: la importancia de Internet, las redes sociales y también por último el COVID-19.

El conocimiento y la integración de estas ideas será de utilidad. Por una parte, como base para determinar cómo las empresas a través de Internet (sitios web), han reaccionado a la necesidad de publicitarse mediante las redes sociales (un punto de vista más a largo plazo y exponencial en el tiempo) y por otra parte como se han adaptado frente a la más que conocida pandemia mundial (un punto de vista mucho más a corto plazo y vertiginoso).

La comprensión y la integración de todos estos conceptos resultan fundamentales, ya que por un lado permiten determinar cuál es el fundamento de la realización de este proyecto, y, por otro lado, facilitan el entendimiento del análisis estadístico posterior correspondiente.

2.2 Internet y las empresas

2.2.1 Introducción

Tal y como explica Apăvăloaie (2014), la aparición de Internet ha tenido un impacto significativo en el mundo en múltiples aspectos, incluyendo la forma en que se vive, trabaja y comunica. En particular, la influencia de Internet en las empresas ha sido significativa y perturbadora. Las empresas han tenido que adaptarse a este nuevo entorno digital para continuar siendo competitivas y aprovechar las oportunidades que ofrece.

Esta nueva tecnología permite a las empresas poder abrirse al mundo de una forma global, no únicamente a través de sus actividades comerciales, sino que también mediante su implicación social en los temas de actualidad. Por primera vez, las empresas a través de Internet son capaces de conectar de una mayor forma con sus clientes a través de una cierta concienciación de sus ideales y políticas.

Desde comienzos de este siglo, la cuantía de usuarios de Internet no ha dejado de aumentar. Hoy en día, es más que necesaria la conexión a Internet en la sociedad, tal es el caso que las empresas tras darse cuenta, comenzaron a hacer uso de esta conexión para la generación de sus propias páginas y sitios web (tal y como se detalla en el apartado siguiente). Las empresas que utilizan Internet en sus actividades diarias, del mismo modo que los usuarios individuales, continúan creciendo día a día, haciendo cada vez más necesarios sus sitios web en una sociedad tan globalizada.

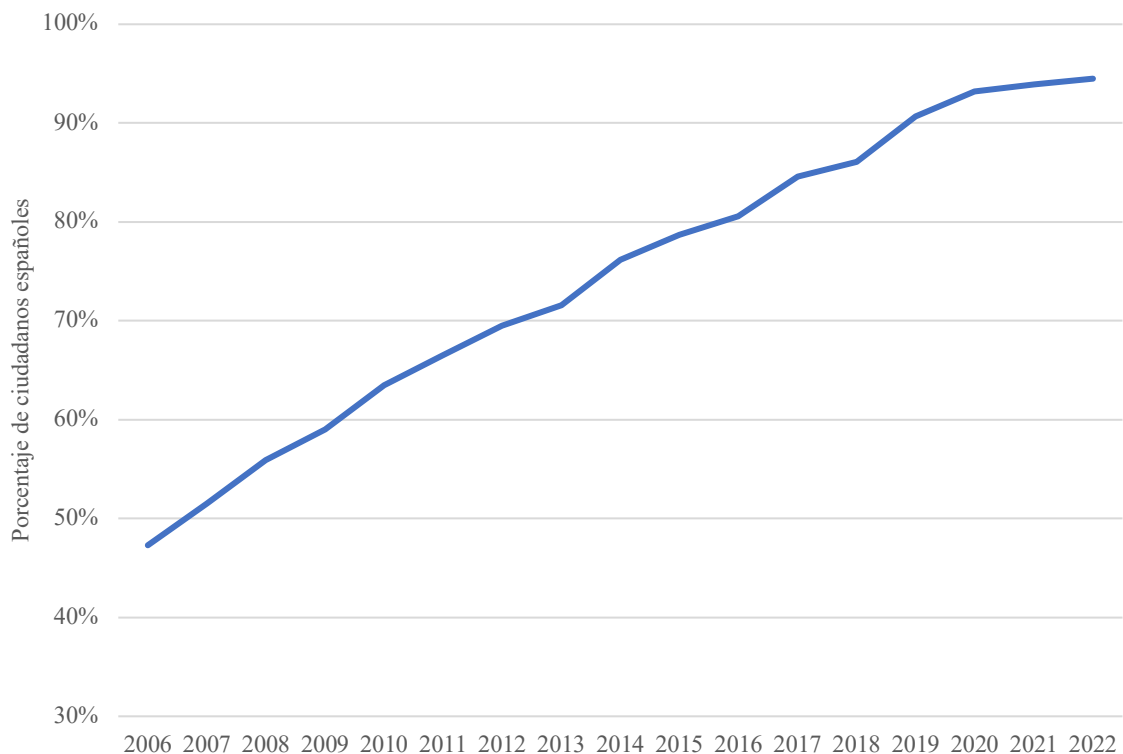


Figura 1. Evolución del porcentaje de ciudadanos españoles que han utilizado Internet al menos 5 días a la semana.

Fuente: Elaboración propia a partir de los datos obtenidos del INE

Tal y como se muestra en la *Figura 1*, el porcentaje de usuarios habituales de Internet en España, al igual que en el resto del mundo continúa creciendo año a año. Se observa como a principios de Siglo XXI, el porcentaje de ciudadanos españoles que utilizaban Internet de una forma asidua no alcanzaba el 50%, mientras que, durante el 2022, el porcentaje ascendió alcanzando una cifra mayor al 90% de la población española.

2.2.2 Sitios web

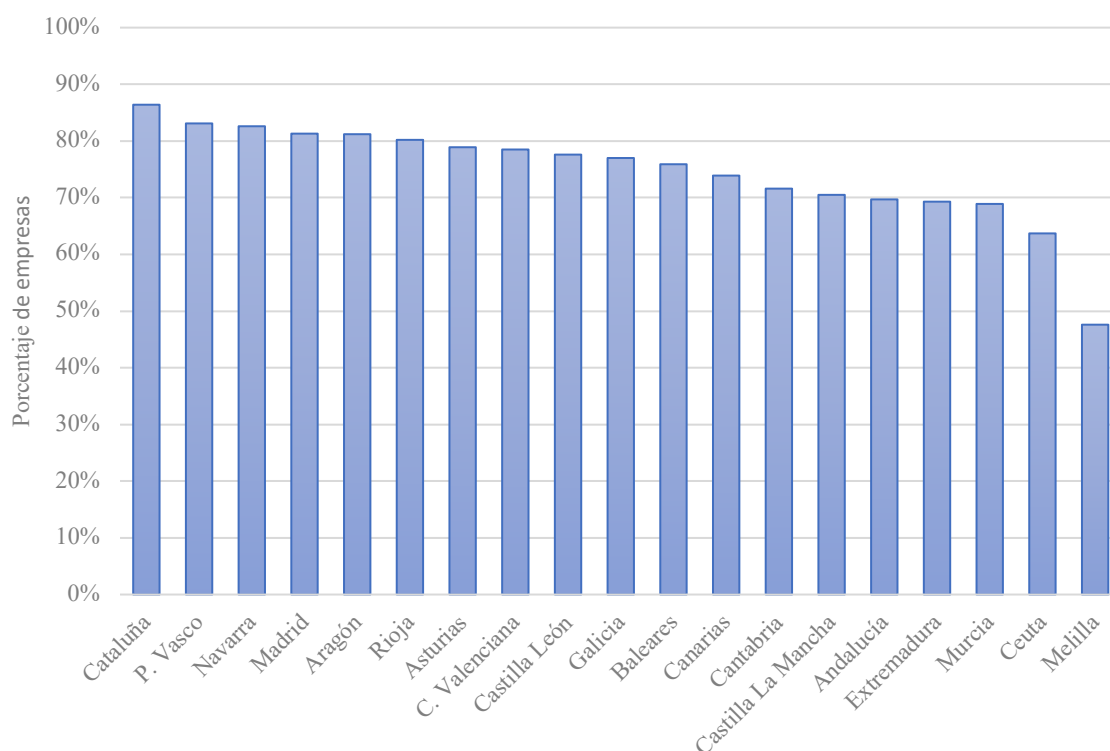
Un sitio web se define como un conjunto de páginas web, vídeos, imágenes y demás recursos relacionados que se almacenan todos ellos en un servidor web, a los que se puede acceder a través de una dirección única, URL o dominio. A través de toda esta cantidad de contenido, las empresas han desarrollado tanto su marca como producto de cara a los clientes y consumidores.

A través del diseño, contenido y estrategias de marketing digital implementadas en sus sitios web, las empresas presentan su pleno control, en ellas, pueden desde destacar productos a los que necesitan otorgarles más relevancia, hasta postularse a favor de alguna iniciativa con relevancia al cambio climático que mejore así su conciencia de marca. Poseer un sitio web llamativo y acorde a la empresa, es fundamental para el correcto funcionamiento de esta.

Además, las páginas web también son utilizadas para presentar los valores, misión y visión de la empresa, (Fitzsimmons, 2022) lo que permite a los visitantes comprender su postura y compromiso en relación con ciertos aspectos de relevancia en la sociedad. De

hecho, hoy en día, casi todos los sitios web de empresas poseen algún apartado dedicado a ello, otorgando así la posibilidad de que cualquier persona pueda conocerlos.

Las empresas, están otorgando tanta relevancia a este aspecto que, de igual forma que en los usuarios individuales de Internet, la cantidad de empresas que poseen sitios web continúa aumentando día a día. Como muestra la *Figura 2*, prácticamente todas las Comunidades Autónomas españolas, presentan un porcentaje de empresas con conexión a Internet y página web superior al 70%. Es cierto que, Ceuta y sobre todo Melilla, influenciadas por su menor desarrollo tecnológico y recursos económicos, presenta un menor porcentaje. A pesar de esto, este es muy cercano al 50%, ya que, la posesión de un sitio web para la empresa ha comenzado a ser una necesidad.



*Figura 2. Porcentaje de empresas por C.A. en España con conexión a Internet y con página web.
Fuente: Elaboración propia a partir de los datos obtenidos del INE*

2.2.3 Tendencias en la sociedad y en los sitios web

Las preocupaciones relevantes a la sociedad, en un mundo tan globalizado como el de hoy en día, son muy variables y cambiantes, es por esto por lo que, el estudio del comportamiento de las empresas frente a ellos es muy complejo. Por ello, a través del estudio de las tendencias en sus sitios web, capacitados para su actualización prácticamente instantánea, permite acercarse mucho más a la opinión de las empresas hacia estas.

Del mismo modo que los temas de interés en la sociedad se encuentran en continua evolución, las opiniones propias de las empresas hacia ellos se pueden ver de la misma manera modificadas. A razón de esto, a través del estudio de estas tendencias, se puede

llegar a predecir tanto cuál será su posicionamiento futuro frente a una variación de ese mismo problema, como qué actuación realizaría frente a uno semejante.

2.3 Redes Sociales

2.3.1 Introducción

El primero de los casos de estudio de tendencias en los sitios web de las empresas hace referencia al uso que hacen éstas de las redes sociales. Para ello, se ha realizado una diferenciación de que son las redes sociales desde un punto de vista sociológico y desde uno en relación con Internet. Esta es muy relevante, ya que, la percepción de qué es una red social ha variado mucho lo largo del tiempo.

Desde una perspectiva sociológica, una red social se define como un conjunto de relaciones y conexiones entre individuos, grupos u organizaciones. Estas relaciones pueden basarse en diversos tipos de vínculos: la amistad, la familia, o cualquier tipo de interés común. En una red social, los individuos crean interacciones entre ellos, que son capaces de influir en sus comportamientos, actitudes y decisiones.

Por otro lado, desde un punto de vista de Internet, el concepto de red social se entiende como un SNS (Social Networking Sites). Estos, son sitios de Internet, que facilitan que exista esta conexión entre individuos. Ahora, tal y como expresa Zhang (2015), el pensamiento de red social se basa en internet, en la conexión vía web. Estas, se perciben como plataformas basadas en la web, las cuales dan la posibilidad a los usuarios de crear perfiles públicos y generar una lista de otros usuarios con los que compartir una conexión. Así, han sido capaces de transformar la percepción de los adolescentes, e incluso, de la mayoría de los adultos, a la hora de mantener y entablar relaciones entre ellos, ya que, para relacionarse siempre tienen en cuenta el nuevo método en línea (Subrahmanyam, 2008).

Los SNS, son un excelente medio para que las empresas interactúen con sus clientes y consumidores. Estas plataformas les permiten establecer una presencia en línea, promover sus bienes o servicios y establecer relaciones directas con su audiencia. Gracias a estas, el contacto con los clientes es mucho más amplio, permitiendo así la recepción de comentarios, opiniones y preguntas de estos. La cantidad de empresas que hacen uso de las redes sociales es altísima, utilizan a modo de escaparate, sus sitios web para incluir las redes sociales que poseen y así informar sobre ellas a sus clientes. A continuación, en la *Figura 3*, se observa que más del 95% de las empresas españolas han hecho uso de alguna red social durante el año 2022. Esto nos hace ver que las redes sociales son esenciales para la vida diaria de las personas pero que, hoy en día, también las empresas las consideran esenciales para sí mismas.

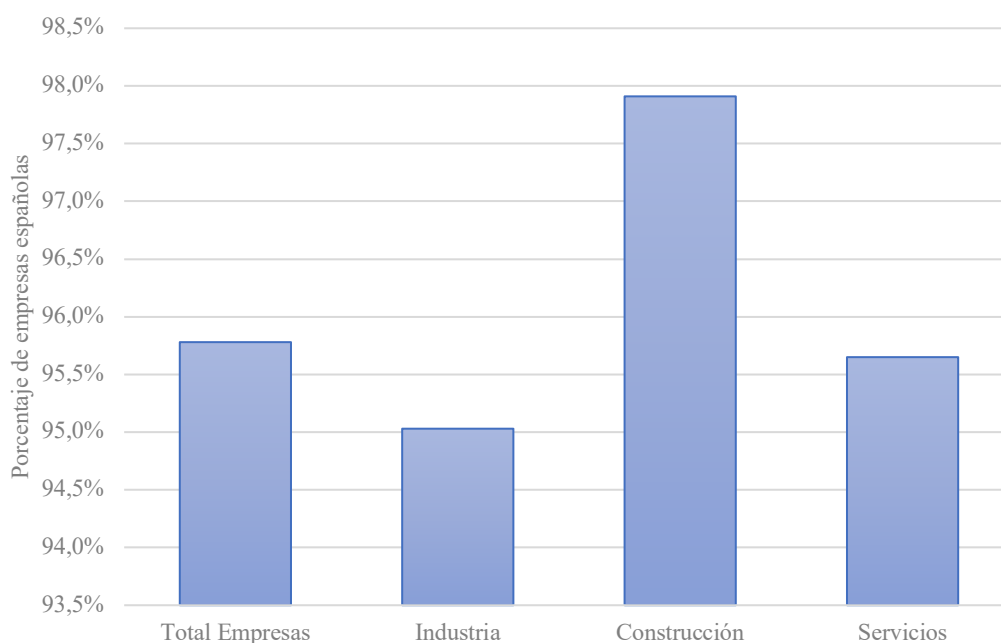


Figura 3. Porcentaje de empresas españolas por sector que han utilizado en el 2022 las RRSS.
Fuente: Elaboración propia a partir de los datos obtenidos del INE

2.3.2 Evolución de las redes sociales

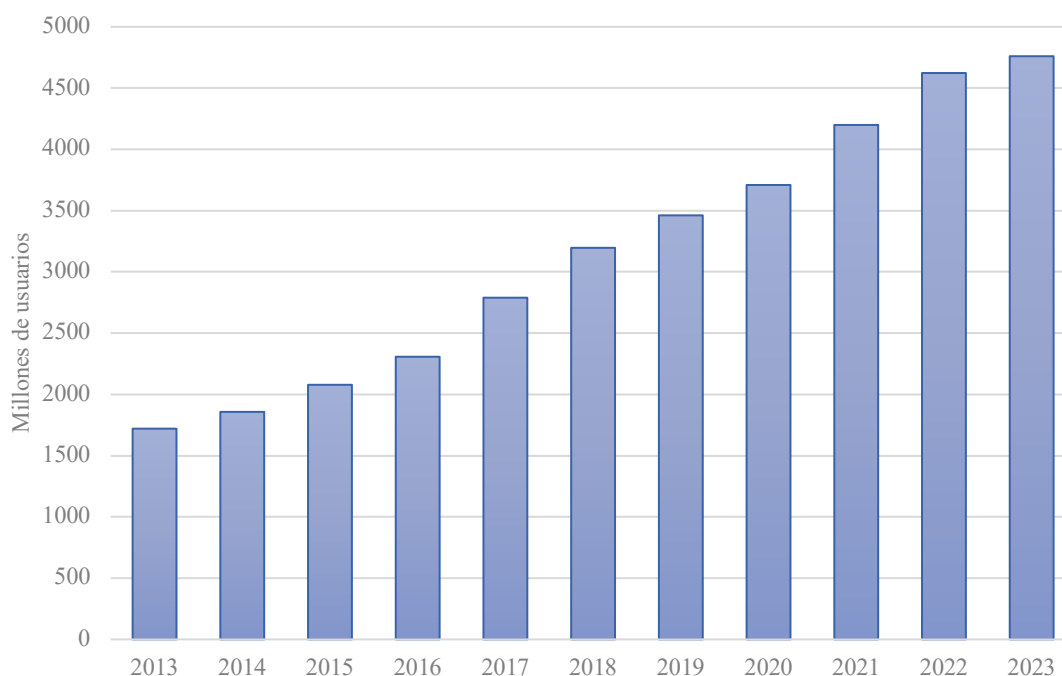
Para tratar de comprender la relevancia que las redes sociales han tenido y tienen en la sociedad, se ha realizado este estudio. Durante los años finales de los años 70 y principios de los 80, se comenzaron a dar los primeros pasos, con sistemas de BBS (Bulletin Board Systems) y salas de chat en línea, como las que se encontraban en sistemas de IRC (Internet Relay Chat). Con los primeros, los usuarios eran capaces de conectarse a un tablón en línea y leer y dejar mensajes en el mismo, en cambio con el IRC, se establecían los primeros chats de mensajería en tiempo real.

Durante los últimos años de la década de los 90 y los principios de los 2000, los sitios web más empleados hacían referencia a la creación de perfiles y listas de amigos. MySpace, Friendster o SixDegrees.com sentaron las bases de lo que serían las futuras redes sociales.

Fue en el año 2004, cuando se lanzó Facebook, una red social inicialmente creada para estudiantes universitarios pero que, en 2006 viendo el éxito que había generado, se abrió a cualquier persona con una dirección de correo electrónico. Se considera a esta la red social más relevante del mundo (ya que se cree que, a partir de ella, muchas otras redes sociales fueron creadas), llegando a ser hoy en día la que más usuarios posee.

Posteriormente, durante los años 2010 – 2020, diversas redes sociales conocidas a nivel mundial como Instagram, TikTok, LinkedIn, Whatsapp o Twitter han sido creadas. Hoy en día, las redes sociales continúan evolucionando para incluir características como las transmisiones en vivo, la realidad aumentada, y han tomado un papel relevante en la economía digital, como canales de atención al cliente, publicidad, y venta de productos y servicios.

La importancia que han tomado dentro de la sociedad se explica a través del número de usuarios que poseen todas ellas. A través de la *Figura 4*, se observa como la evolución de este número de usuarios es únicamente ascendente a lo largo del tiempo, alcanzando a principios del año 2023, unos valores superiores a los 4600 millones de usuarios. Mientras que, a principios del 2013, no se superaban los 1800 millones. Este crecimiento se debe a la tan gran cantidad de redes sociales existentes (como se muestra en el apartado siguiente).



*Figura 4. Millones de usuarios de las RRSS en los meses de Enero de cada año.
Fuente: Elaboración propia a partir de los datos obtenidos de un estudio de Data Reportal*

2.3.3 Clasificación de las redes sociales

Existen una inmensa cantidad de redes sociales, por esto, hoy en día es impensable que una empresa no trate de extender sus alas hacia la búsqueda de más público a través de ellas. Por ello, este es uno de los dos puntos clave en los que se profundizará en el estudio. El mundo se encuentra en constante cambio tal y como ya se ha comentado anteriormente y es por esto por lo que se han desarrollado diferentes redes sociales, de diferentes materias, para diferentes usuarios, creando una clara clasificación entre ellas.

Duong (2020), hace una diferenciación de los tipos de redes sociales que existen en base a su propósito, es decir, a través de cuál es la función que realizan los usuarios en ellas. Esta clasificación expone las siguientes: redes sociales de propósito general, redes sociales de fotografías y contenido multimedia, redes sociales de mensajería instantánea, redes sociales profesionales y redes sociales temáticas o especializadas.

La clasificación que realiza se compone de la siguiente manera:

1. Redes sociales de propósito general (Facebook, Twitter, etc.): Estas son las redes sociales más conocidas y utilizadas a nivel mundial. Se caracterizan por ser de propósito general, es decir, no están enfocadas en una temática o actividad específica, sino que se utilizan para compartir contenido en general, como fotos, videos, noticias, pensamientos, etc. No están realmente definidas en ningún ámbito, cada usuario es libre para otorgarle una utilidad, de la forma que más le agrade.

2. Redes sociales de fotos y videos (Instagram, YouTube, Vimeo, etc.): Este tipo de redes sociales se enfocan en la creación y el consumo de contenido visual. Redes sociales como Instagram se centra en compartir fotos y videos, mientras que YouTube y Vimeo se utilizan principalmente para compartir videos y contenido audiovisual más elaborado.

3. Redes sociales de mensajería instantánea (WhatsApp, Snapchat, etc.): Estas redes sociales se utilizan para la comunicación directa entre usuarios en tiempo real. Se caracterizan por tener una función de chat o mensajería instantánea, que permite el intercambio de mensajes, fotos y videos.

4. Redes sociales profesionales (LinkedIn, Xing, etc.): Estas redes sociales están enfocadas en el ámbito laboral y profesional. Se utilizan para conectar a profesionales de diferentes sectores y compartir información sobre empleos, oportunidades laborales, etc.

5. Redes sociales temáticas o especializadas (Goodreads, Strava, etc.): Este tipo de redes sociales se enfocan en temáticas específicas, como la literatura, el deporte, la música, la cocina, etc. Son utilizadas por personas que tienen un interés específico en una actividad o temática en particular y buscan conectar con otros usuarios con intereses similares.

Como se puede observar en la *Figura 5*, no todas las categorías de redes sociales poseen el mismo número de usuarios. Tal y como se muestra y como se ha comentado al comienzo de este apartado, Facebook, es la red social que más usuarios posee, logrando alcanzar prácticamente la cifra de 3000 millones. Esta, se trata de una red social de propósito general, abierta a todo el mundo y de fácil uso. En cambio, Goodreads, se enfoca en un catálogo de usuarios mucho más reducido como el de los lectores especializados. Pese a ello, el número de usuarios que la integran no es irrisorio, ni mucho menos, con cerca de 150 millones. Por ello, a pesar de pensar que no todas las redes sociales son relevantes para la sociedad, hasta las menos conocidas, presentan unos valores de usuarios increíbles, siendo así fundamentales en el día a día de muchas personas.

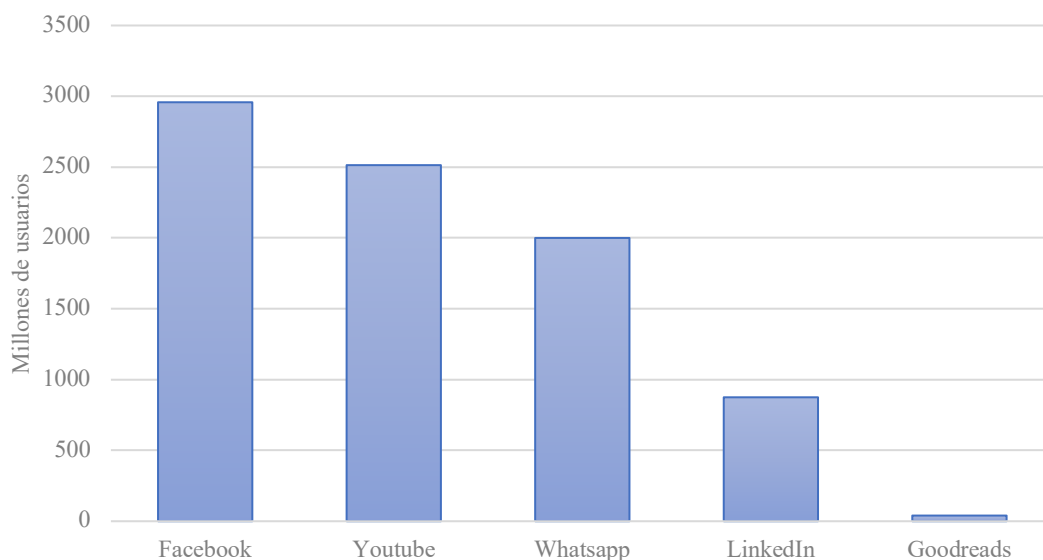


Figura 5. Millones de usuarios contabilizados al comienzo del año 2023 en las cinco RRSS mostradas.
Fuente: Elaboración propia a partir de los datos obtenidos del estudio de Data Reportal

2.4 COVID-19

2.4.1 Introducción

El segundo de los casos de estudio de tendencias en los sitios web de las empresas es sobre la enfermedad COVID-19. Esta, es una enfermedad infecciosa causada por el virus SARS-COV-2 que se detectó por primera vez en Wuhan, China, en diciembre de 2019. Desde entonces, ha afectado a millones de personas en todo el mundo, causando enfermedad y muerte. A raíz de esta, se ha declarado la pandemia de COVID-19 como una emergencia de salud pública de preocupación internacional. (OMS, 2020)

Tal y como expone la OMS (2023), esta enfermedad ha alcanzado los 750 millones de infecciones diagnosticadas y los 6,5 millones de pacientes fallecidos. Es más que evidente que esta enfermedad es conocida por todo el mundo y ya que este no es un proyecto de carácter médico o de ciencias de la salud no se profundizará mucho más en el tema de cómo surge ni como se transmite dicha enfermedad.

En España, la enfermedad ha alcanzado una cifra de contagios superior a los 13 millones, tal y como se muestra en la *Figura 6*. Fue, durante los meses posteriores a Agosto de 2020 (con una cifra de 288.000 casos diagnosticados únicamente), ya con capacidad de realizar pruebas para determinar las personas contagiadas por la enfermedad, cuando el número de contagios comenzó a crecer de manera descontrolada. Por otro lado, gracias a las actuaciones del personal sanitario y a las vacunas creadas para hacer frente a esta pandemia, el número de contagios a finales del año 2022 comenzó a disminuir en gran medida.

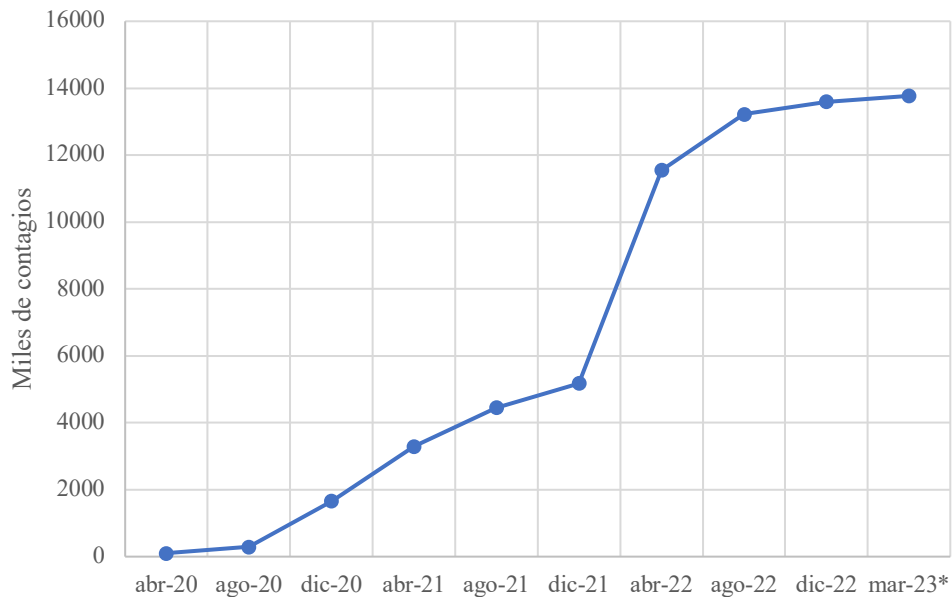


Figura 6. Evolución del número de contagios (en miles) en España de la enfermedad COVID-19.
Fuente: Elaboración propia a partir de los datos obtenidos del Gobierno de España

2.4.2 Impacto en las empresas

A través del increíble impacto de la enfermedad en la sociedad, tanto los gobiernos como los ciudadanos tuvieron que adoptar una serie de medidas para evitar la propagación del virus, esto, como es evidente, fue una medida aceptada y adoptada de igual manera por las empresas de todo el “globo”.

Las empresas se han visto afectadas en distintas áreas, como la cadena de suministro, la demanda de productos y servicios, la disponibilidad de empleados y la implementación de medidas de seguridad (Pinzaru, 2020). Además, para responder a estos desafíos, y su adaptación a la “nueva normalidad”, las empresas han tenido la necesidad de implementar estrategias totalmente innovadoras y flexibles para esta situación.

Es cierto que, existen una gran cantidad de sectores económicos muy diferenciados entre ellos, en los cuales habitan todas estas empresas, y, que no en todos se ha vivido de igual manera la situación causada por la pandemia. Como es evidente, por ejemplificarlo, la hostelería (sector cerrado durante gran parte de la pandemia), ha sufrido en mucha mayor medida que los servicios farmacéuticos (trabajando durante la misma). Este estudio se hará en más profundidad en los siguientes apartados, ya que, en la muestra utilizada para el estudio se diferencia el sector económico en el que se incluyen estas empresas, pero es importante remarcar que no todos los sectores se vieron afectados de igual manera.

Kang (2021), destaca, que las empresas han tenido que adaptarse rápidamente para hacer frente a los desafíos que ha presentado la pandemia. Para ello, las empresas han implementado una serie de respuestas, entre las que se incluyen:

- Adaptación de los procesos de producción: Las empresas han tenido que reorganizar sus procesos de producción para asegurar la salud y seguridad de los

empleados y cumplir con las regulaciones sanitarias, con implementaciones como el teletrabajo o trabajo online.

- Cambios en la cadena de suministro: La pandemia ha interrumpido las cadenas de suministro globales, lo que ha llevado a las empresas a buscar nuevas fuentes de materiales y proveedores.
- Implementación de nuevas tecnologías: Las empresas han adoptado nuevas tecnologías para permitir el ya mencionado teletrabajo y la colaboración en línea con los clientes y proveedores. De esta forma, se limitaba el contacto social entre trabajadores a únicamente los plenamente necesarios.
- Desarrollo de nuevos productos y servicios: Las empresas han desarrollado nuevos productos y servicios para satisfacer las necesidades de los clientes en el contexto de la pandemia. Muchas han sido las empresas textiles que han cambiado su fabricación de productos, para comenzar a realizar mascarillas u otro tipo de material sanitario.
- Cambios en la estrategia de marketing: Las empresas han ajustado su estrategia de marketing para adaptarse a los cambios en la demanda y las necesidades de los clientes.

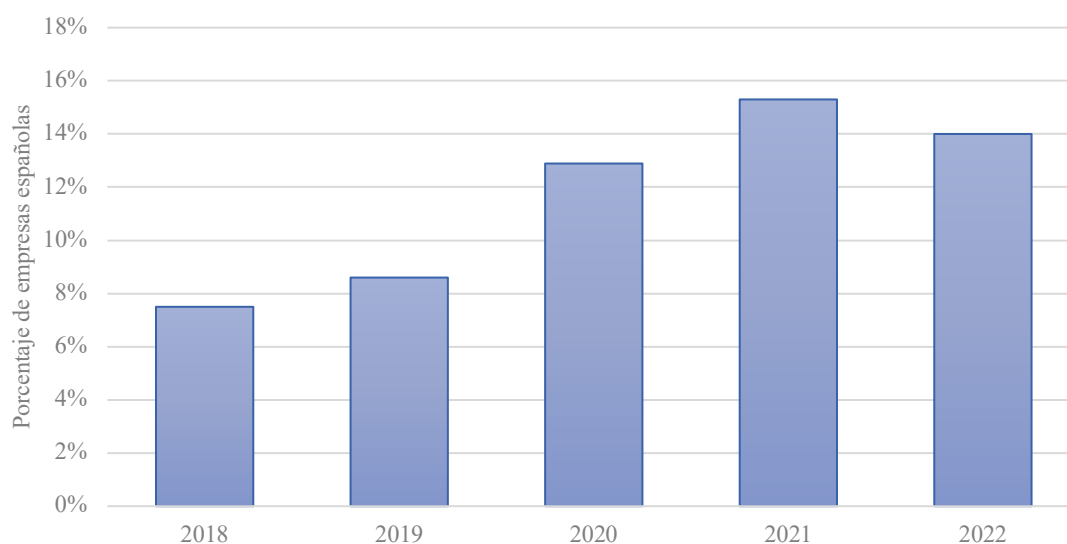


Figura 7. Porcentaje de teletrabajo en España desde el año 2018 al 2022.

Fuente: Elaboración propia a partir de los datos aportados del IVIE (Instituto Valenciano de Investigaciones Económicas)

Como vemos, una de las medidas más adoptada por las empresas españolas para hacer frente a la pandemia fue la implantación del “teletrabajo”, a través de la *Figura 7*. Durante el año 2018, únicamente un 7% de empresas españolas disponían de la opción de trabajar a distancia. A causa de la pandemia, desde el año 2020, hasta el 2022 (en especial el 2021), el porcentaje de empresas que han adoptado esta nueva forma de trabajo ha llegado ha aumentado en una gran cuantía, logrando superar el 15% (más del doble que en el primer año mostrado).

También, como es claro, todos los negocios que han tenido que transformarse de una manera tan clara han debido de redefinir su futuro en un contexto postpandemia (Enaworu, 2021). Para ello, han debido de realizar e implantar medidas de supervivencia y crecimiento a largo plazo. Entre estas, destacan la adopción de tecnologías digitales, la diversificación de los productos y servicios ofrecidos, la implementación de prácticas de trabajo más flexibles y el fortalecimiento de las relaciones con los clientes y los proveedores.

En general, la pandemia no ha pasado desapercibida para ningún tipo de empresa. Estas, pueden haberse visto más o menos afectadas en cuanto a su rendimiento económico, pero a nivel de forma de trabajo, relaciones interpersonales entre sus trabajadores y clientes, pensamiento a largo plazo, etc. todas y cada una de ellas, se han visto transformadas de una forma radical y es por esto por lo que es más que necesario su estudio.

2.4.3 La sociedad y la información

La pandemia del COVID-19 ha sido una de las mayores crisis sanitarias y sociales a nivel mundial en la historia reciente. Desde que la enfermedad fue declarada una pandemia por la Organización Mundial de la Salud (OMS) en marzo de 2020, ha tenido un impacto significativo en la vida de las personas y en la sociedad en su conjunto.

Ante una crisis de tal magnitud, la necesidad de información se ha convertido en una prioridad para la sociedad. La propagación del virus, la evolución de la enfermedad, las medidas de prevención y las últimas novedades en cuanto a tratamientos y vacunas han sido temas cruciales para la población en todo el mundo. La importancia de esta información actualizada y fiable ha sido fundamental, para la toma de decisiones y la gestión de la crisis sanitaria y social provocada por el COVID-19. Esta, ha llevado a un constante seguimiento de las novedades y evolución de la enfermedad por parte de los medios de comunicación, los gobiernos y las organizaciones internacionales de salud e incluso los ciudadanos como vemos en la *Figura 9*.

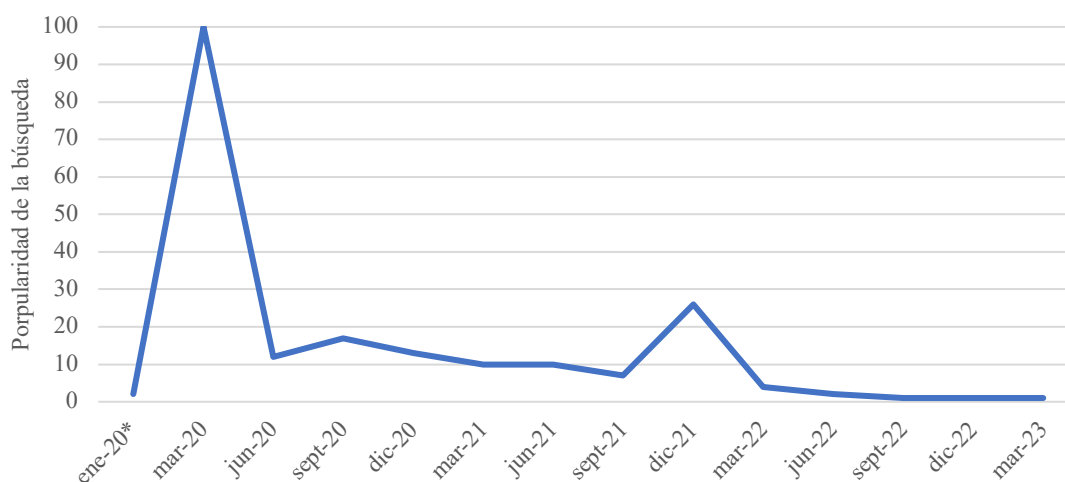


Figura 8. Búsquedas en Internet sobre información relativa al COVID-19 (en relevancia de búsquedas). Fuente: Elaboración propia a partir de los datos obtenidos de Google Trends

Reddy y Gupta (2020), enfatizan en que la comunicación efectiva es crucial para combatir la desinformación respecto a la pandemia y garantizar que la población tenga acceso a información precisa y actualizada sobre la enfermedad. Se destaca que la comunicación debe ser clara, concisa y accesible para el público en general, y que las autoridades sanitarias y los medios de comunicación tienen un papel importante en la transmisión de información precisa y confiable.

Las empresas por ello también han sido un medio de concienciación y comunicación sobre las novedades que había generado la pandemia en su ámbito. A través de sus sitios web, las empresas, daban voz a sus problemas causados por la pandemia, a sus respuestas a las grandes incógnitas de la sociedad, y, de sus clientes sobre si iban a continuar aportando el servicio que realizaban.

La mayoría de las empresas incorporaron información relativa al COVID-19 en sus sitios web por tal de dar tranquilidad y estabilidad a sus clientes (Mahmud, 2021). Estas lo hicieron de formas diferentes, pero siempre tratando de proteger la salud de sus empleados, clientes y otras partes interesadas. De esta forma, las empresas desde un punto de vista responsable y ético comunicaron a sus clientes sus acciones y compromisos frente a la pandemia.

Es por esto, que el análisis de tendencias en los sitios web centrados en este ámbito es muy necesario a la par que interesante. Hoy por hoy, las empresas que han resistido a esta crisis mundial pueden comenzar a “moverse hacia adelante” al igual que la sociedad. En cambio, hay otras muchas que no lograron sobreponerse a una crisis de tal magnitud. De esta forma es muy importante observar y ver cómo ha habido empresas que sí que han logrado continuar realizando sus servicios a través de la buena forma de adaptación a la pandemia y como otras no han sido capaces de hacerlo.

2.5 Lenguajes empleados para el estudio

Para tratar también el contexto tecnológico en el que se involucra el trabajo, se han descrito los dos lenguajes de programación, así como el entorno utilizado para la realización de este. Así, se trata de dar una explicación sobre el razonamiento de su elección y cuál ha sido su uso.

2.5.1 R y RStudio

En una amplia gama de campos, el lenguaje de programación R es una herramienta esencial para el análisis de datos y la estadística (Team, 2022). Es un lenguaje de programación de código abierto (puede ser modificado, estudiado o distribuido por cualquier usuario) con muchas características para el análisis estadístico y la visualización de datos. Se pueden obtener resultados significativos en la investigación científica al explorar sus múltiples funcionalidades y paquetes. Entre sus grandes ventajas, destacan la generación de gráficos y visualizaciones impresionantes y su gran capacidad para manejar grandes conjuntos de datos.

Es por esto, la razón por la que desde la dirección del proyecto se decidió a la utilización de este lenguaje especializado en análisis de datos para todo el trabajo. Además, este

lenguaje ha sido tratado durante el grado de Administración y Dirección de Empresas a través de la Intensificación propia de Inteligencia de Negocios. De esta forma, ambas características eran idóneas para realizar el trabajo mediante este.

Por otro lado, RStudio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, diseñado para facilitar la programación y el análisis de datos. A través de este entorno, se permite realizar la programación en R. (Allaire, 2012) Es una herramienta esencial para este tipo de programación, ya que, proporciona una interfaz de usuario intuitiva y una amplia variedad de funcionalidades que mejoran la eficiencia y la productividad, y también la utilizada mediante el grado.

A través del lenguaje y su entorno, a lo largo del grado, se han realizado estudios estadísticos diversos a través del análisis de datos de empresas ficticias para poder realizar trabajos reales con ellos. De esta forma, se daba una situación idónea mediante la cual, a través de este trabajo, se comenzara a trabajar con empresas reales de la economía. Por todo ello, se consideró que la forma más útil y de más aprovechamiento para nuestro estudio y visualización de los datos se realizara mediante estas dos herramientas mencionadas en este apartado.

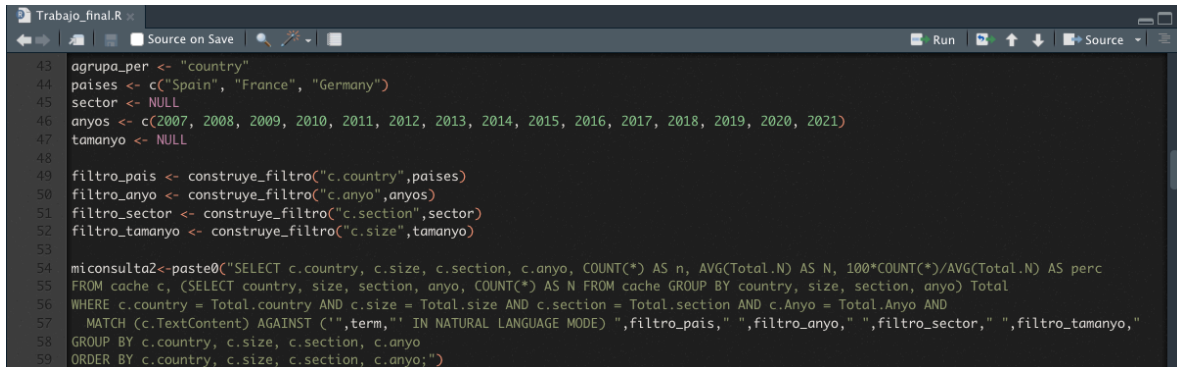
2.5.2 SQL

SQL es otro lenguaje de programación utilizado para administrar y manipular bases de datos relacionales (son un tipo de base de datos que organizan la información en una serie de tablas relacionadas entre sí, en lugar de una estructura jerárquica). Permite realizar diversas operaciones en los datos, como consultar, insertar, actualizar y eliminar información en una base de datos. Este se utiliza en una gran variedad de aplicaciones y sistemas.

MySQL, por otro lado, es un sistema de gestión de bases de datos relacionales de código abierto que utiliza el lenguaje SQL para realizar operaciones en la base de datos. Su facilidad de uso, escalabilidad y velocidad hacen que sea uno de los sistemas más utilizados con este tipo de lenguaje.

RStudio, el entorno seleccionado para la realización del trabajo permite conectarse y trabajar con bases de datos utilizando SQL, mediante el paquete “MariaDB” (O’Leary, 2015). El proyecto con el JRC del que parte este TFG almacenó sus datos en una base de datos relacional compatible con MySQL. Así, a través de las herramientas que proporciona R, se pudo realizar las diferentes consultas y descarga correspondiente necesaria para la obtención de los datos del trabajo.

Para la realización de las consultas, se necesitaron los conocimientos adquiridos durante el grado sobre el lenguaje SQL. Este, al igual que R, posee una determinada sintaxis, muy bien definida y que es muy útil para la consulta en las bases de datos. De esta forma, se han obtenido todos los datos mediante los cuales posteriormente se han realizado los análisis. A continuación, en la *Figura 9*, se observa una de las consultas en lenguaje SQL realizadas.

The image shows a screenshot of an R console window titled 'Trabajo_final.R'. The window contains R code that defines variables for country, sector, and year, and then uses a function 'construye_filtro' to create filters. The main part of the code is an SQL query named 'miconsulta2' which selects data from a 'cache' table, grouped by country, size, section, and year. The query includes a WHERE clause for filtering and a MATCH clause for text analysis. The results are ordered by country, size, section, and year.

```
43 agrupa_per <- "country"
44 paises <- c("Spain", "France", "Germany")
45 sector <- NULL
46 anyos <- c(2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021)
47 tamanyo <- NULL
48
49 filtro_pais <- construye_filtro("c.country",paises)
50 filtro_anyo <- construye_filtro("c.anyo",anyos)
51 filtro_sector <- construye_filtro("c.section",sector)
52 filtro_tamanyo <- construye_filtro("c.size",tamanyo)
53
54 miconsulta2<-paste0("SELECT c.country, c.size, c.section, c.anyo, COUNT(*) AS n, AVG(Total.N) AS N, 100*COUNT(*)/AVG(Total.N) AS perc
55 FROM cache c, (SELECT country, size, section, anyo, COUNT(*) AS N FROM cache GROUP BY country, size, section, anyo) Total
56 WHERE c.country = Total.country AND c.size = Total.size AND c.section = Total.section AND c.Anyo = Total.Anyo AND
57 MATCH (c.TextContent) AGAINST ('",term,"' IN NATURAL LANGUAGE MODE) ",filtro_pais," ",filtro_anyo," ",filtro_sector," ",filtro_tamanyo,"
58 GROUP BY c.country, c.size, c.section, c.anyo
59 ORDER BY c.country, c.size, c.section, c.anyo;")
```

*Figura 9. Captura de pantalla de una consulta ejemplo en lenguaje SQL en R.
Fuente: Elaboración propia mediante los datos obtenidos del proyecto JRC2023*

2.6 JRC

La idea del TFG de análisis de tendencias de sitios web de las empresas surge a raíz de un proyecto llamado JRC2023 de la Universidad Politécnica de Valencia, liderado por el tutor de este TFG, Josep Domènech i de Soria. Este proyecto tiene como objetivo realizar un estudio sobre la sostenibilidad en PYMES de cuatro países de la Unión Europea. El proyecto, debe su nombre a la comunidad de investigación de la CE.

El JRC (Joint Research Centre) es un centro de investigación conjunta de la Comisión Europea, que tiene como objetivo proporcionar asesoramiento científico y técnico a los responsables políticos de la Unión Europea. JRC realiza investigaciones y análisis en una amplia variedad de áreas, desde la energía y el medio ambiente hasta la salud y la seguridad alimentaria.

El proyecto se centraba en la importancia que posee la sostenibilidad en las PYMES, ya que, pese a ser empresas que no poseen los recursos necesarios para provocar un cambio a gran escala, son responsables de la gran mayoría de gases que provocan el efecto invernadero. Así, se enfocaron en el estudio de los diferentes sitios web de estas empresas para realizar el análisis de su nivel de concienciación sobre esta problemática. Su idea, era basarse en si las empresas informaban acerca de este problema en sus sitios web, y de ese modo concienciaban también a la sociedad.

Como ya se ha comentado, este trabajo se encuentra muy ligado a dicho proyecto, y es por esto, que, se decidió extrapolar la idea de únicamente centrarse en el estudio de la sostenibilidad, al estudio de cualquier término que tuviera relevancia en la sociedad. La obtención de los datos de este estudio se realizó en el año 2022. Estos, son mediante los cuales se ha realizado el trabajo (explicado más detalladamente en el punto siguiente).

Durante el año 2023, momento en el que se comienza este trabajo, se ha continuado con la recopilación de nuevos datos sobre estos cuatro países, así como, del resto de países de la Unión Europea. Todavía, al tratarse de una tan gran cantidad de datos, se encuentran en proceso de limpieza y procesamiento y es por esto por lo que este estudio se ha centrado en los datos recopilados y tratados durante el año 2022.

Por último, cabe destacar que el proyecto se encuentra en su fase final y se espera que sea a lo largo de este 2023 cuando se finalice el estudio y se pueda realizar el informe pertinente al cliente, en este caso, a la Comisión Europea.

CAPÍTULO 3. METODOLOGÍA

CAPÍTULO 3. METODOLOGÍA

3.1 Introducción

Esta sección es fundamental para entender cómo se llevó a cabo el análisis de los datos tratados, y en más profundidad, en los dos conceptos explicados con anterioridad (redes sociales y COVID -19).

Se describen detalladamente los métodos y técnicas utilizadas para recopilar y analizar los datos de los sitios web de las empresas. Además, se explican los criterios de selección de las empresas y los sitios web a analizar, así como la planificación y ejecución del análisis en sí.

3.2 Construcción de la muestra

Cabe destacar, como ya se ha mencionado en el apartado anterior, que, la selección de las empresas propias de la muestra del estudio se realizó el año anterior del proyecto JRC. Para realizar la selección de las empresas se llevó a cabo una búsqueda exhaustiva en dos bases de datos y en línea, SABI y ORBIS. De esta forma, SABI fue utilizada para la obtención de los datos propios de las empresas españolas, mientras que las demás fueron obtenidas de ORBIS, ya que, esta se trata de una base de datos mucho más amplia.

La población que ha sido seleccionada en este estudio son las empresas que poseen sitio web de una selección de sectores económicos de los principales cuatro países de la Unión Europea: España, Francia, Italia y Alemania. Se han excluido, por motivo evidente, las empresas sin sitio web, pues el trabajo se centra en estudiar las tendencias en los sitios web de las empresas. Los sectores de actividad económica, según la clasificación NACE, que se estudiarán son: (B-J y L-N). Los códigos NACE (Clasificación Estadística de Actividades Económicas en la Comunidad Europea) son un sistema de clasificación que es utilizado para la categorización y organización de las actividades económicas en diferentes sectores.

La muestra ha sido construida mediante un muestreo estratificado sobre esta población, donde cada estrato se compone de una combinación de país, tamaño y sector económico. Cada grupo tiene una muestra de aproximadamente 200 empresas, con un nivel de confianza del 95%, un margen de error del 7% y un P=Q del 50%. Se ha realizado un muestreo exhaustivo en aquellos sectores donde la población total de empresas no supera este número (como en las actividades mineras). Finalmente, alrededor de 33.000 empresas forman parte de la muestra total.

Las dos bases de datos proporcionan información sobre el código NACE de 4 dígitos de las empresas; con el cual posteriormente se ha recodificado con la letra correspondiente al sector económico de la empresa. A continuación, para clasificar las empresas de la muestra por tamaño, se ha considerado el número de empleados del 2021, ya que, en el momento de la obtención de los datos de la muestra, era el último año disponible. Además, los datos obtenidos de los sitios web para cada empresa recorren desde el 2007 hasta el 2021. Es por esto por lo que, el conjunto final de datos consta de un total de 495.000 observaciones (las 33.000 empresas con una observación por cada año, durante 15 años).

3.3 Agregación de los estratos de la muestra

La muestra obtenida no tiene la misma distribución de sectores económicos, tamaños de empresa y países que la población subyacente, dado que todos los estratos tienen el mismo número de observaciones. Por ello, cuando se quieren obtener medidas de la población, se deben aplicar pesos a cada estrato de forma que se recupere la distribución de empresas original.

Para la realización de todas las agregaciones mostradas en el trabajo (gráficos, obtención de las variables propias del estudio de la COVID-19 y las Redes Sociales...), ha sido necesaria la ponderación de pesos. Fue necesario disponer de una clara muestra de cómo de relevantes eran tanto los tamaños de las empresas y los sectores económicos para la economía del propio país, cómo cuán importante eran las economías de estos países para la economía mundial.

De esta forma, se han obtenido las ponderaciones relativas al país de la empresa estudiada (*Tabla 2*), la referente al tamaño de la empresa, mostrada en la *Tabla 3* y, por último, la ponderación por sectores económicos, las cuales se explican mediante la *Tabla 4*.

Tabla 1. Ponderaciones por país de origen de la empresa

País	ES	IT	FR	DE
Valor	13,27%	19,60%	27,51%	39,62%

Fuente: Informe técnico del proyecto con JRC en 2022

Tabla 2. Ponderaciones por tamaño de empresa

Tamaño	ES	IT	FR	DE
Micro	95,38%	94,28%	93,01%	89,38%
Pequeña	3,89%	5,03%	5,79%	8,44%
Mediana	0,60%	0,59%	0,99%	1,80%
Grande	0,14%	0,10%	0,21%	0,38%

Fuente: Informe técnico del proyecto con JRC en 2022

Tabla 3. Ponderaciones por sector de la empresa

Sector	ES	IT	FR	DE
B	0,07%	0,05%	0,07%	0,08%
C	6,44%	10,17%	8,39%	8,54%
D	0,43%	0,34%	1,12%	2,61%
E	0,26%	0,26%	0,31%	0,42%
F	15,46%	13,32%	14,32%	14,19%
G	26,73%	29,21%	26,23%	22,75%
H	7,54%	3,27%	5,00%	4,18%
I	10,40%	9,16%	9,54%	9,07%
J	2,55%	2,97%	3,45%	4,85%
L	6,84%	6,46%	6,50%	6,21%
M	15,76%	20,50%	18,45%	19,10%
N	7,53%	4,29%	6,61%	8,01%

Fuente: Informe técnico del proyecto con JRC en 2022

3.4 Variables de la muestra

Variables comunes

Las variables comunes del estudio son cuatro: la primera hace referencia al país de cada una de las empresas. Es una variable categórica nominal discreta, la cual puede tomar cuatro valores distintos, tal y como se ha expresado en la definición de la muestra: España, Francia, Italia y Alemania.

La segunda variable, es el sector económico, la cual puede tomar 13 valores distintos. Del mismo modo que la anterior, se trata de una variable de tipo categórico, discreta, nominal. Para ello, se ha hecho uso del ya mencionado código NACE que se obtiene de las bases de datos y su posterior recodificación con las siguientes letras (dependiendo del sector al que pertenecen).

Tabla 4. Sectores económicos recodificados propios de la muestra

SECCIÓN	SECTOR ECONÓMICO
B	Industrias extractivas
C	Industria manufacturera
D	Suministro de energía eléctrica, gas, vapor y a. a
E	Suministro de agua, actividades de saneamiento
F	Construcción
G	Comercio al por mayor y al por menor vehículos de motor
H	Transporte y almacenamiento
I	Hostelería
J	Información y comunicaciones
K	Actividades financieras y de seguros
L	Actividades inmobiliarias
M	Actividades profesionales, científicas y técnicas
N	Actividades administrativas y servicios auxiliares

Fuente: Elaboración propia a través de la clasificación NACE del INE

En tercer lugar, la variable referente al tamaño. Esta, se trata de una variable categórica discreta, la cual, hace referencia al número de empleados. Esta, se mide por la clase de tamaño a la que pertenecen las empresas. Haciendo uso de la clasificación del Instituto Nacional de Estadística, se han definido 4 tamaños diferentes de empresa: micro, pequeña, mediana y grande. Para la determinación de que una empresa perteneciera al grupo micro, ésta, debe poseer entre 1 y 9 trabajadores. Se considera pequeña a toda empresa de 10 a 49 trabajadores. Por su parte, una empresa de entre 50 y 249 empleados se clasifica como mediana, y una mayor a estos números (250 trabajadores o más), como grande. Por último, se encuentra la variable referente al año del estudio. Esta, es una variable categórica discreta nominal, la cual, puede tomar 15 valores distintos (2007 – 2021).

Estas cuatro variables son propias de las bases de datos ORBIS y SABI, pero a través de la herramienta de búsqueda creada, se permite generar multitud de variables dicotómicas (binarias), “1” si el término deseado aparece en el sitio web, y “0” si no lo hace. Se pueden crear tantas variables como el usuario quiera, ya que, la búsqueda de términos es ilimitada. Pese a ello, para la realización de este trabajo final de grado, y así a modo de ejemplificación de cómo funciona la herramienta creada, las variables que se utilizarán son las que se mencionan a continuación.

Variables propias del estudio del Redes Sociales

En este apartado, se tienen en consideración, las variables propias que se han utilizado para la obtención de los resultados en el capítulo V de este trabajo, en el ámbito del estudio de la evolución y la situación actual de la información de las redes sociales en los sitios web de las empresas. Para ello, se han tenido en consideración las siguientes variables:

La primera variable es Redes sociales: Es una variable binaria, es decir, se trata de una variable categórica con únicamente dos valores: “1” si aparece en el sitio web de la empresa y “0” cuando no aparece. Para este estudio, se han tenido en cuenta los siguientes términos: “redes sociales”, “red social”, “social network”, ‘social media’, “soziale Medien”, “soziale Netzwerke”, “rot sozial”, “rosso sociale”, “sociale rosso”, “media sociali”, “médias sociaux”, “réseau social”, “réseau social”, “instagram”, ‘facebook’, ‘linkedIn’, ‘youtube’, ‘godreads’, ‘whatsapp’, ‘twitter’, ‘vimeo’, ‘snapchat’, ‘strava’, ‘xing’ teniendo así en cuenta tanto los idiomas de los cuatro países, como también todas las redes sociales que se han utilizado en el proyecto.

La segunda variable es el término de la redes sociales propia de la red social Instagram, para la cual se ha realizado una búsqueda con este término: “instagram”. Se trata de una variable categórica con únicamente dos valores y por tanto binaria. Tomará el valor de “0” cuando no aparezca en el sitio web consultado, mientras que tomará el valor de “1” cuando sí que se encuentre en este.

La tercera variable es la variable propia de la red social Twitter y para la cual se ha realizado una búsqueda con este término: “twitter”. Es una variable categórica y binaria con únicamente dos valores. Tomará el valor de “0” cuando no aparezca en el sitio web consultado, mientras que tomará el valor de “1” cuando sí que se encuentre en este.

La cuarta variable que se ha tenido en cuenta en el estudio hace referencia a la red social Facebook y para la cual se ha realizado una búsqueda con este término: “facebook”. Del mismo modo que las anteriores y todas las pertenecientes a este grupo, se trata de una variable categórica con únicamente dos valores posibles (binaria). De esta forma, tomará el valor “1” si se encuentra en el sitio web consultado y el valor “0” si no lo hace.

La quinta variable es Youtube y de la misma forma que las anteriores se ha realizado una búsqueda con este término: “youtube”. Como las demás, es una variable categórica nominal y binaria con únicamente dos valores. Tomará el valor de “0” cuando no aparezca en el sitio web consultado, mientras que tomará el valor de “1” cuando sí que se encuentre en este.

La sexta variable es propia de la red social LinkedIn y su término buscado ha sido “linkedIn”. Es una variable categórica y binaria con únicamente dos valores. Tomará el valor de “0” cuando no aparezca en el sitio web consultado, mientras que tomará el valor de “1” cuando sí que se encuentre en este.

Por último, esta variable es Whatsapp con el término utilizado “whatsapp”. Es una variable categórica y binaria con únicamente dos valores. Tomará el valor de “0” cuando no aparezca en el sitio web consultado, mientras que tomará el valor de “1” cuando sí que se encuentre en este.

Variables propias del estudio de menciones a la enfermedad COVID-19

Del mismo modo que se ha realizado en el apartado anterior, en este se han tenido en consideración las variables relativas a la enfermedad del COVID-19. En este caso, únicamente se ha generado una variable y es la siguiente:

La variable Covid: Del mismo modo que las variables propias de las Redes Sociales, se trata de una variable categórica con únicamente dos valores posibles “1” y “0”, y por tanto binaria. En cuanto a este estudio, se han tenido en cuenta las siguientes términos: “COVID-19”, “coronavirus”, “SARS-CoV-2”, “pandemia”, “pandémie”, “pandemie”, “covid”, teniendo así en cuenta los idiomas de los cuatro países, y pudiendo ampliar todavía más la búsqueda. De esta forma, si la variable toma el valor “1” quiere decir que se encuentra en el sitio web donde se ha realizado la búsqueda, y el valor “0” cuando no aparezca en este.

3.5 Formulación de hipótesis

Para la elaboración de los dos casos de estudio sobre los contenidos de los sitios web de las empresas, se van a formular cinco hipótesis, que posteriormente se estudiarán en el capítulo 5. De esta forma, se podrá corroborar que el estudio de dicho análisis ha tenido éxito y utilidad. Las cinco hipótesis son las siguientes:

H1. Las empresas en países con una mayor proporción de contagios de la enfermedad COVID-19 en su población, poseen una mayor inclinación por publicar información sobre esta enfermedad en sus sitios web.

Basándose en la información proporcionada por la OMS (Organización Mundial de la Salud) (2023), los cuatro países muestreados presentan los siguientes números de contagios: España (13.868.227), Francia (39.010.097), Alemania (38.423.300) e Italia (25.842.595). Por otro lado, la población de cada uno de los cuatro países (Unión Europea, 2023) es la siguiente: España (47.615.034), Francia (67.871.925), Alemania (83.794.000) e Italia (59.030.133). De esta forma, los países que mayor proporción de contagios presentan y por ende los correspondientes a la H1, son Alemania (45,85%) y Francia (57,48%), mientras que España (29,13%) e Italia (43,78%), presentan una proporción menor. Por tanto, cabe esperar que más empresas alemanas y francesas traten asuntos relacionados con la enfermedad COVID-19 comparadas con las españolas e italianas.

H2. Las empresas en países más desarrollados tecnológicamente tienden a hacer un mayor uso de redes sociales.

Tal y como expone la Comisión Europea (2019), Alemania es considerado el país de la Unión Europea junto a Suecia, Dinamarca y Finlandia más desarrollado a nivel tecnológico y en el sector de la innovación. Al tratarse de un país altamente desarrollado, debe existir mayor acceso generalizado a internet y por ello mayor utilización de las redes sociales. Por ello, en esta muestra, Alemania será el país que se espera que posea una mayor tendencia a la utilización de redes sociales en sus sitios web, frente a los otros tres países muestreados.

H3. El uso de redes sociales es mayor en los sectores económicos que se encuentran más cerca del consumidor final.

Los sectores más próximos al consumidor final son los siguientes: Comercio al por mayor y al por menor vehículos de motor (G) y Hostelería (I) (INE, 2009). Por ello, estos, serán los que posean una mayor tendencia a la utilización de las redes sociales en los sitios web

de sus empresas. Esto es debido a que las redes sociales se centran como ya se ha comentado en el apartado correspondiente en la búsqueda de cercanía y confianza con los clientes, más que con terceros u otras empresas.

H4. El uso de las diferentes redes sociales en los sitios web implica el uso de otras redes sociales en estos.

El uso de las redes sociales por los usuarios (Wilson, 2010) permite que se cree la necesidad de utilización de otras redes sociales para diferentes acciones. Cuando se utiliza una red social y se descubre su utilidad, surge el deseo de utilizar otras para cumplir con otras acciones y necesidades, tanto en individuos como en empresas. De esta forma, se va a estudiar una posible correlación entre el uso de las seis de las redes sociales más utilizadas por los usuarios (Auxier, 2021): ‘Instagram’, ‘Twitter’, ‘Facebook’, ‘LinkedIn’, ‘Whatsapp’ y ‘Youtube’.

H5. La tendencia a la información relativa a la enfermedad COVID-19 es mayor en las empresas más concienciadas en la imagen de marca y en generar confianza en sus clientes.

Tal y como expone Spence (2010), las empresas de menor tamaño (micro y pequeñas), deben cuidar mucho más su imagen de marca y dar confianza a sus clientes, que las de gran tamaño, ya que, estas segundas poseen una fidelización muy superior de sus clientes. Cuando una empresa es de un tamaño elevado, gracias a sus recursos económicos alcanza a más clientes, generando una mayor reputación y reconocimiento de marca.

De esta forma, en el capítulo 5, dedicado al análisis de los resultados obtenidos a través de los métodos estadísticos utilizados, se aceptarán o rechazarán estas cinco hipótesis. Así, se verá de una forma muy representativa, lo interesante y la inmensa capacidad de análisis que posee esta herramienta.

3.6 Métodos estadísticos

En este apartado, se explicarán todos y cada uno de los métodos estadísticos utilizados para aceptar o rechazar las cinco hipótesis planteadas. De esta forma, se fundamenta el razonamiento que se ha seguido para llegar a las conclusiones del trabajo.

Coefficiente Phi

La prueba estadística utilizada para medir la correlación entre variables binarias se llama prueba del coeficiente phi (Ekström, 2011). El coeficiente phi es una medida de asociación que se utiliza específicamente para evaluar la relación entre dos variables categóricas binarias. Esta prueba se basa en una tabla de contingencia que muestra la frecuencia conjunta de las categorías de las variables.

La prueba del coeficiente phi, evalúa si hay una asociación significativa entre las variables binarias. La hipótesis nula es que no hay asociación entre las variables, mientras que la hipótesis alternativa plantea que existe una asociación significativa. De esta forma, se ha

utilizado este método estadístico para evaluar la posible asociación entre el uso de las diferentes redes sociales en los sitios web de las empresas con estas hipótesis:

$$H_0 = \text{No existe ningún tipo de asociación entre las variables}$$
$$H_1 = \text{Existe asociación entre las variables}$$

Por último, este coeficiente fluctúa entre valores del rango $[-1 - 1]$. De esta forma, se considerará que posee una relación de asociación positiva las dos variables, siempre que, el coeficiente sea positivo y viceversa. En cuanto a la “fuerza” de esa asociación, se considera que existen tres grados. El primero, es el correspondiente al coeficiente phi menor que $|0.1|$, considerando que existe una asociación débil. El segundo, el coeficiente phi fluctúa entre $|0.1|$ y $|0.2|$, considerando esta como una asociación moderada. Por último, el grado del coeficiente phi mayor que $|0.3|$, el cual hace referencia a una asociación fuerte.

Modelo Regresión Logística

La regresión logit, también conocida como regresión logística, es una técnica de modelado estadístico utilizada para predecir y analizar variables binarias o categóricas que tienen dos categorías mutuamente excluyentes, como sí/no (Nick, 2007).

La fórmula propia de la regresión logit es la siguiente: $p = \frac{1}{1+e^{-(\beta_0+\beta_1X)}}$, en ella, la probabilidad (p), que posee un rango de entre 0 y 1, se calcula a través de estos valores: X, son los valores de las variables predictoras utilizadas en el modelo. β_0 es el término constante del modelo y β_1 es el coeficiente del predictor. En resumen, esta expresión nos permite predecir la probabilidad de ocurrencia del evento binario en función de la variable predictora y los coeficientes del modelo.

De esta forma, la función logit se utiliza para convertir cualquier valor que se desee en un valor con un rango entre 0 y 1. Así, será utilizada posteriormente para concluir si alguna variable es influyente en la aparición de otra en los sitios web de las empresas.

Pruebas de proporción de dos muestras

Las pruebas de proporción de dos muestras son un tipo de prueba estadística utilizada para comparar las proporciones de dos grupos o muestras independientes (Agresti, 2012). Estas pruebas, son aplicables cuando se tienen dos variables categóricas binarias y se desea determinar si hay evidencia suficiente para concluir que las proporciones de éxito en ambos grupos son diferentes.

Esta prueba de hipótesis se ha utilizado para comprobar las diferencias entre grupos dentro de las variables estudiadas en este trabajo, y, observar, como han afectado las variables creadas en ellos. De esta forma, se han establecido las siguientes hipótesis nula y alternativa:

$H_0 =$ La proporción del primer grupo es mayor/menor que la del segundo grupo

$H_1 =$ La proporción del primer grupo es menor/mayor que la del segundo grupo

En esta prueba de hipótesis, al igual que en todas las demás, se establece un nivel de significación del 5% (0.05). Así, cuando los p. valores calculados sean inferiores a este nivel de significación, H_0 , es decir, la hipótesis nula se rechazará. Por otro lado, cuando el p. valor sea superior al nivel de significación, se aceptará H_0 .

El estadístico de esta prueba sigue una distribución chi cuadrado, y se calcula de la siguiente forma: $\chi^2 = \frac{(O_1 - E_1)^2}{\frac{E_1 + (O_2 - E_2)^2}{E_2}}$

Para la explicación de la fórmula del estadístico, se explica que O_1 y O_2 son las frecuencias observadas de las dos muestras (número de éxitos de cada muestra), mientras que E_1 y E_2 , son las frecuencias esperadas, calculadas bajo la hipótesis nula, siendo: $E_x = \frac{n_x * e_T}{n_x + n_y}$. En esta segunda fórmula, n_x y n_y son los tamaños de cada muestra, y e_t , la suma de éxitos de ambas muestras.

CAPÍTULO 4. VISUALIZACIÓN

CAPÍTULO 4. VISUALIZACIÓN

4.1 Objetivo de la visualización

El objetivo de esta se basa en, a través de un término (seleccionado por el usuario), proporcionar una evolución temporal de este en los sitios web de las empresas, por país, sector y tamaño de empresa. Además de esta forma, esta visualización es útil tanto para el equipo de trabajo del JRC, junto al cual se ha colaborado en este trabajo, como para el usuario externo final.

Para cada término seleccionado, se permite hacer una selección de los niveles de agregación propios de este trabajo: el sector, el tamaño y el país de la empresa. Con esta funcionalidad, se facilita el análisis de todos los datos relevantes relacionados con el término seleccionado, abarcando desde el año 2007 hasta el 2021, y brindando la capacidad de interactuar con ellos en diferentes niveles.

En definitiva, la creación de esta herramienta de visualización busca permitir a los usuarios comprender y acceder de una manera sencilla a la información de los sitios web de empresas relativa al término seleccionado. De esta forma, se permite el estudio de las tendencias de los sitios web de los países muestreados y las posibles diferencias entre los sectores de estos.

4.2 Organización de la visualización

Para el correcto entendimiento de la visualización generada, se ha decidido mostrar la misma en cuatro paneles diferenciados, además de un cuadro de texto donde incluir el término que se desea analizar. Cada uno de los paneles muestra un gráfico de la evolución del término seleccionado, siendo el primero para el gráfico mostrado por países, el segundo por los tamaños de empresa, el tercero el que muestra los diferentes sectores de la economía y por último el perteneciente a un mapa de la UE.

4.2.1 Gráfico de países

En primer lugar, el primer gráfico que se implementará es el que muestra la evolución del término seleccionado en cada uno de los cuatro países pertenecientes a la muestra. Este, es un gráfico de líneas, que muestra la información desde el año 2007 hasta el 2021. Al ser un gráfico diferenciado por países, no se incluirá información referente a los sectores económicos, ni tamaños de empresa. En el eje horizontal, al igual que se hará en los otros gráficos, se muestran los años propios del estudio, y en el eje vertical, se muestran los porcentajes de empresas por país, que utilizan el término seleccionado en su sitio web.

4.2.2 Gráfico de sectores económicos

El segundo de los gráficos que se va a realizar en esta visualización es el que muestra la información desagregada por los distintos sectores que se han estudiado en este trabajo. De esta forma, se permite la diferenciación de cada uno de ellos. Del mismo modo que en el apartado anterior, la información que se muestra corresponde a los años desde el

2007 al 2021. En este gráfico, los porcentajes de las empresas que utilizan el término escogido en sus sitios web se muestran en el eje vertical, y en el horizontal, se muestra el año de cada observación.

4.2.3 Gráfico de tamaños de empresa

En cuanto al tercer gráfico generado, muestra la información desagregada por el tamaño de empresa. Así, se permite en este, diferenciar entre las empresas micro, pequeñas, medianas y grandes, a través de un gráfico de líneas. En el eje horizontal, se muestran los diferentes años estudiados, mientras que en el vertical se muestran los porcentajes de empresas que utilizan el término seleccionado en su sitio web.

4.2.4 Mapa de Europa

Por último, el gráfico generado en este caso es un mapa de Europa. Se muestra en él, el porcentaje de empresas pertenecientes a cada país que, en un año determinado, muestran en sus sitios web el término seleccionado. En él, aparecen coloreados los países muestreados: España, Francia, Italia y Alemania, en función del ya mencionado porcentaje.

4.3 Implementación de gráficos estáticos

El programa elegido para esta implementación fue el ya mencionado anteriormente R. Con él, y más concretamente a través del paquete propio de dicho lenguaje, “ggplot2” (del que se hablará en profundidad en el siguiente apartado), se ha realizado esta visualización de datos.

4.3.1 Introducción a ggplot2

Se ha llevado a cabo un primer análisis de datos utilizando el paquete “ggplot2” de RStudio. El paquete “ggplot2” (Wickham, 2011) es una herramienta poderosa que permite la creación de visualizaciones de datos con alta calidad y consistentes en R. Entre sus características más relevantes, destaca la capacidad de generación de gráficos a partir de diferentes tipos de datos, una gramática coherente y una sintaxis muy intuitiva para el usuario.

Wickham expone que entre los gráficos más empleados y de mayor utilidad, destacan los gráficos de dispersión (donde los puntos en el gráfico representan la relación entre dos variables continuas), los gráficos de barras (comparaciones entre categorías), de líneas (cómo evolucionan las variables continuas a lo largo del tiempo) y los de áreas (representando la distribución de los datos con áreas sombreadas).

4.3.2 Resultado de la implementación

Tal y como se ha expuesto anteriormente en este apartado, los cuatro gráficos han sido realizados a través de la herramienta de “ggplot2” en lenguaje de programación R. Estos, son los correspondientes a los gráficos diferenciados por país, tamaño y sector de empresa, así como el mapa de los cuatro países.

Para la muestra de los gráficos generados, se ha realizado una consulta del término “Educación” a modo de ejemplo para observar cuál es el funcionamiento de la herramienta creada. En este término, se han incluido estos cinco conceptos: educación, education, bildung, educazione y éducation. Así, se ha tenido en cuenta el idioma propio de cada país y el inglés por si algún sitio web de dichos países estuviera en este idioma. De esta forma, se han obtenido estos cuatro gráficos:

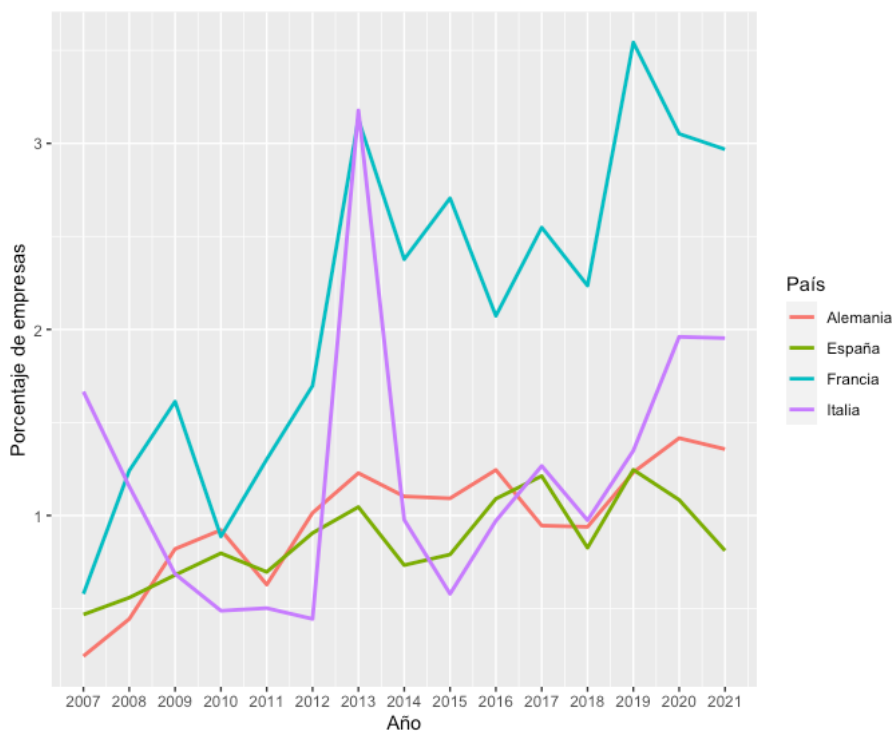


Figura 10. Gráfico realizado a través del paquete “ggplot2”, desagregado por países de empresa.
Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

Como se observa en la Figura 10, se ha generado tal y como se explicaba al comienzo del capítulo, un gráfico de líneas desagregando la información en los cuatro países muestreados. Así se observa la evolución de la utilización del término “educación” en los sitios web de cada uno de los cuatro países.

A continuación, a través de la Figura 11, se muestra la información también, con un gráfico de líneas, pero esta vez, diferenciando según el tamaño de las empresas estudiadas. Del mismo modo y como se expuso anteriormente, en el eje vertical se incluyen el porcentaje de empresas que han utilizado alguno de los términos mencionados en sus sitios web. Por otro lado, en el eje horizontal se muestran los años propios del estudio (2007 – 2021).

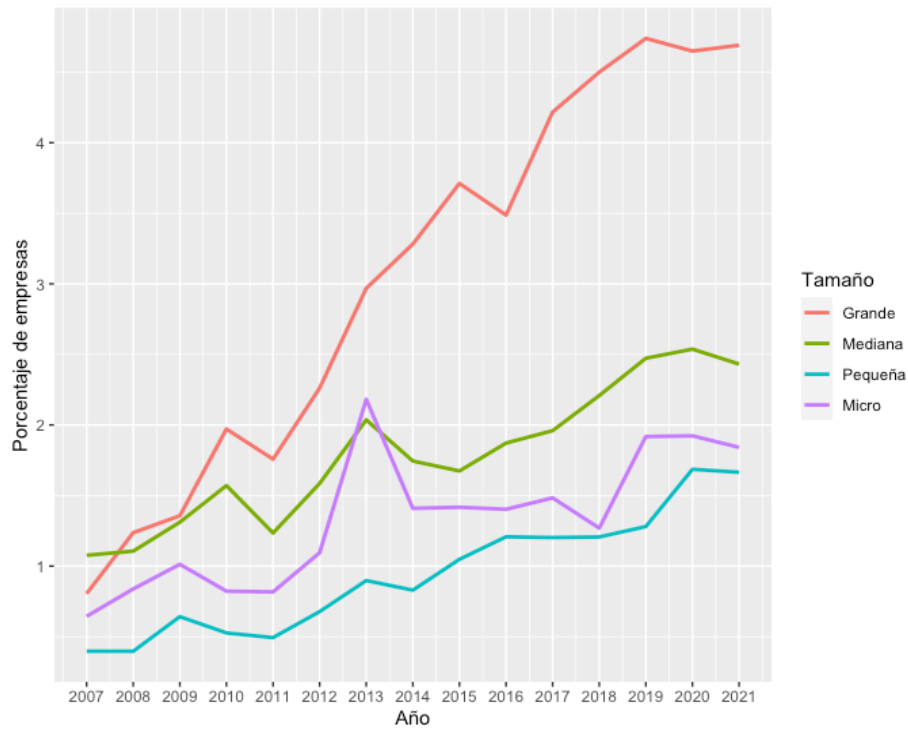


Figura 11. Gráfico realizado a través del paquete "ggplot2", desagregado por tamaños de empresa.
Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

En cuanto al último gráfico de líneas generado (Figura 12), la información se muestra desagregada en función del sector económico al que pertenece la empresa estudiada. Por ello, se muestran los sectores pertenecientes al rango (B-J y L-N).

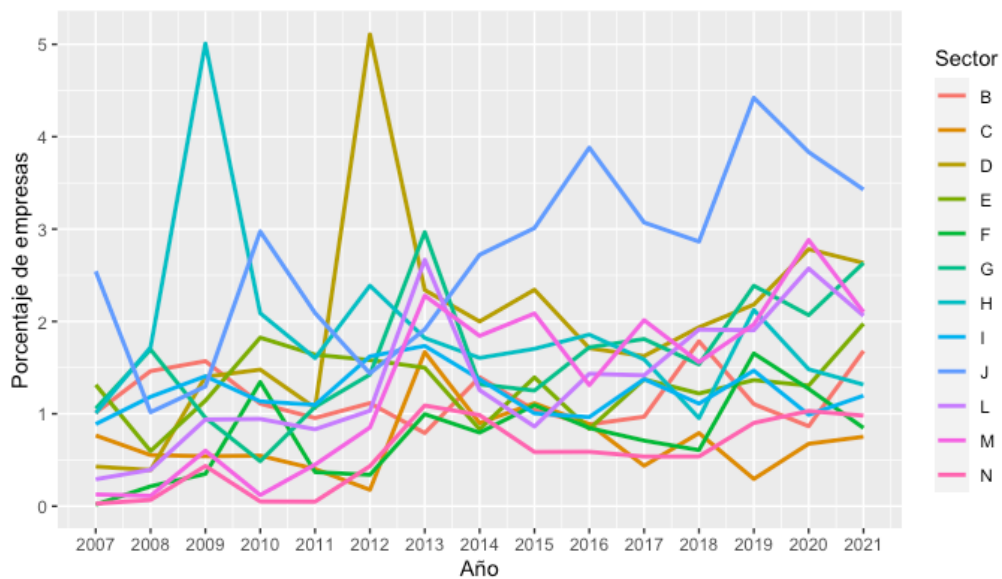
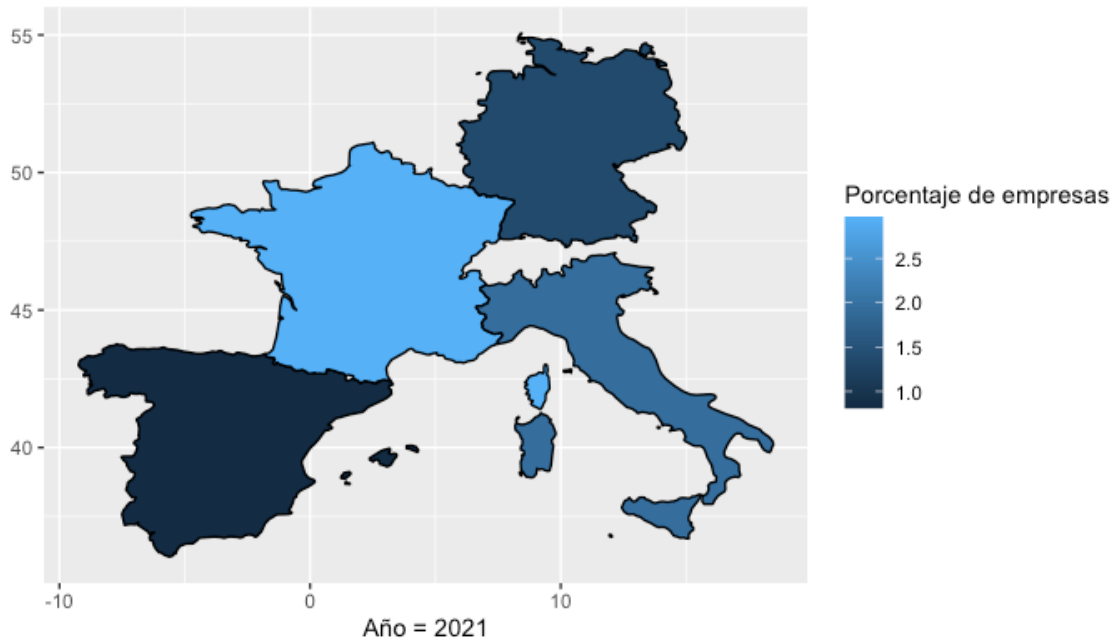


Figura 12. Gráfico realizado a través del paquete "ggplot2", desagregado por sector de empresa.
Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

Por último, el último gráfico realizado a través del paquete de visualización “ggplot2”, es el que se muestra mediante la *Figura 13*. En ella, se observa el mapa geográfico de los cuatro países estudiados, coloreados en función del porcentaje de empresas que presentan los términos relevantes a la educación en sus sitios web. Como vemos, y al tratarse de un gráfico estático, se ha realizado el perteneciente a los porcentajes de los países en el año 2021.



*Figura 13. Mapa geográfico realizado a partir del paquete "ggplot2".
Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"*

4.4 Gráficos interactivos

Tras lograr generar todos los gráficos requeridos para el proyecto en estático y gracias a la herramienta de “ggplot2”, el equipo de trabajo se ha planteado la posibilidad de generar estos mismos gráficos, pero con un contenido interactivo. De esta forma, permitir la interacción con ellos y aumentar la capacidad de información.

4.4.1 Motivación de la creación de los gráficos interactivos

La principal motivación de la realización de este tipo de gráficos es mejorar la calidad del trabajo y aumentar su impacto a través de un más sencillo entendimiento por parte del lector. Estos, permiten al usuario explorar los datos de una manera más dinámica y atractiva, lo que puede hacer que la información sea más comprensible y memorable. Entre sus características propias destacan la facilitación de exploración de los gráficos, la capacidad de realizar zoom para observar un tramo específico de este, la utilización de filtros en ellos y el resaltado de líneas, entre otros. Hoy en día, es necesario hacer más atractivo cualquier tipo de visualización de datos, ya que, es la manera más sencilla de captar la atención del usuario y lograr su completo entendimiento.

4.4.2 Evaluación de alternativas

Para lograr el objetivo de la generación de los gráficos interactivos, el equipo decidió realizar un análisis de las herramientas a utilizar para esta generación. Estas herramientas seleccionadas en R incluyeron: Plotly, GoogleVis y Highcharts.

1. Plotly es un paquete de visualización de datos en RStudio que permite la creación de gráficos interactivos y animados de alta calidad (Sievert, 2021). Su sintaxis simple y su interfaz de usuario intuitiva lo hacen fácil de usar. Su principal ventaja es su posibilidad de trabajar mediante los gráficos ya creados con “ggplot2”. Este paquete permite generar gráficos interactivos a través de gráficos planos generados anteriormente. Así, se podía reutilizar todo el trabajo previo.

Otra de sus ventajas es el tamaño del archivo creado, ya que, este es bastante reducido. Este, permite realizar toda la gama de gráficos necesaria y el tamaño de su HTML generado es muy reducido. Para realizar la prueba de cómo funcionaba esta herramienta, se realizó el gráfico mostrado en la *Figura 14*, siguiente:

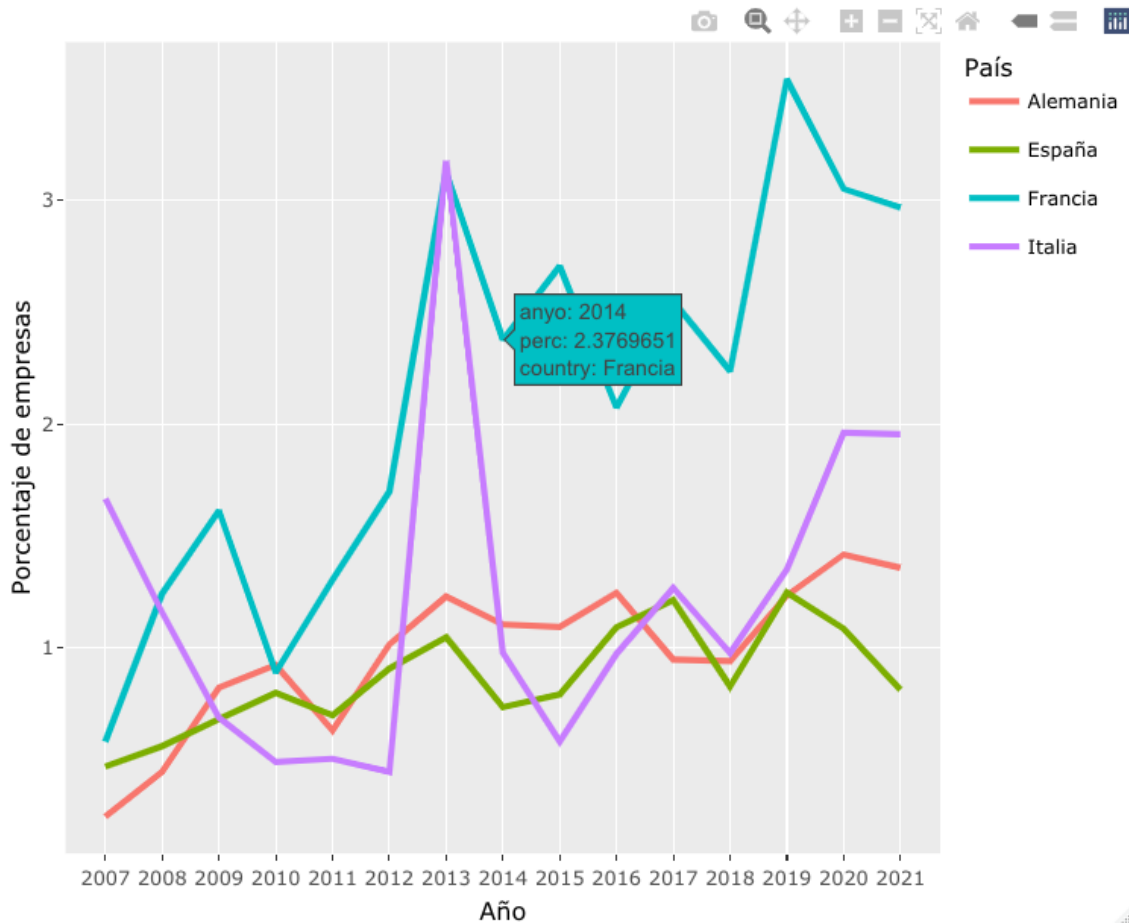


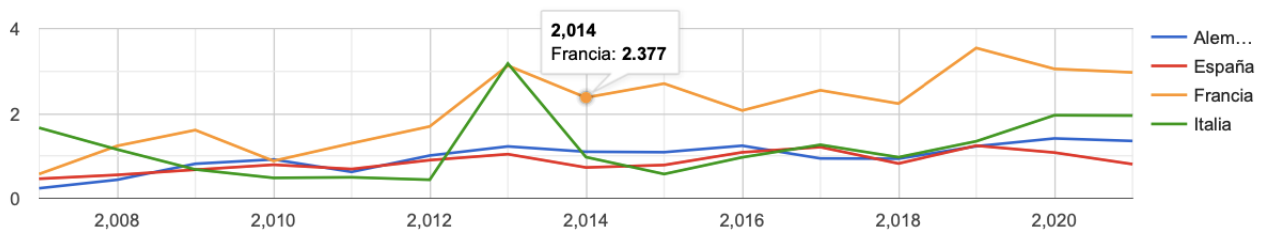
Figura 14. Gráfico de ejemplo realizado con Plotly sobre la evolución del uso de los términos “educación” en los sitios web de los 4 países.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

2. El paquete “GoogleVis” (Gesmann, 2013) proporciona una interfaz entre R y las herramientas de gráficos de Google. Este, permite a los usuarios crear páginas web con gráficos interactivos basados en marcos de datos de R, utilizando estas herramientas, y mostrarlos a partir del servidor de ayuda HTTP de R o en sus propios sitios web, sin tener que cargar los datos en Google.

Esta, es su gran ventaja, ya que, al utilizar esta librería tan amplia, las posibilidades de generación de cualquier tipo de gráfico son inmensas. Este paquete, es capaz de generar todos los gráficos necesarios para el proyecto.

En cuanto a la parte negativa y razón final para no seleccionar dicho paquete para la generación de los gráficos interactivos, es el archivo HTML que genera. Este, es necesario para la visualización y su tamaño es demasiado grande, es decir, ocupa mucho espacio. Es un gran inconveniente, ya que, puede limitar en exceso esa generación de gráficos requerida. Finalmente, a través de la *Figura 15*, se observa el gráfico generado y mediante el que se evaluó este paquete.



Data: data • Chart ID: LineChartID4b668dd14aa • googleVis-0.7.0
R version 4.2.1 (2022-06-23) • Google Terms of Use • Documentation and Data Policy

Figura 15. Gráfico de ejemplo realizado con GoogleVis sobre la evolución del uso de los términos “educación” en los sitios web de los 4 países.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

3. Por último, Highcharts a su vez, es un paquete de visualización de datos en RStudio que permite crear gráficos con un estilo semejante (Kargl, 2018), como gráficos de líneas, de dispersión, etc. Destaca por comprimir todos ellos en una única función “hchart(x)” y es conocido por su estética atractiva y sus opciones de personalización.

Otra característica destacada es su capacidad para compartir visualizaciones con otros. Las visualizaciones creadas con Highcharts se pueden compartir fácilmente en línea, ya sea como una imagen estática o como una visualización interactiva en la nube. Además, este paquete permite la generación de archivos HTML de mucho menor tamaño, que favorece mucho su uso.

Pese a ello, el paquete finalmente tampoco ha sido seleccionado. Esta vez, se debe en gran parte a su falta de gama de gráficos. Esta no era tan avanzada como la necesaria para llevar a cabo el proyecto y finalmente se descartó. El gráfico que se realizó a modo de prueba para seleccionar o no este paquete de herramientas, es el que se muestra a continuación en la *Figura 16*.

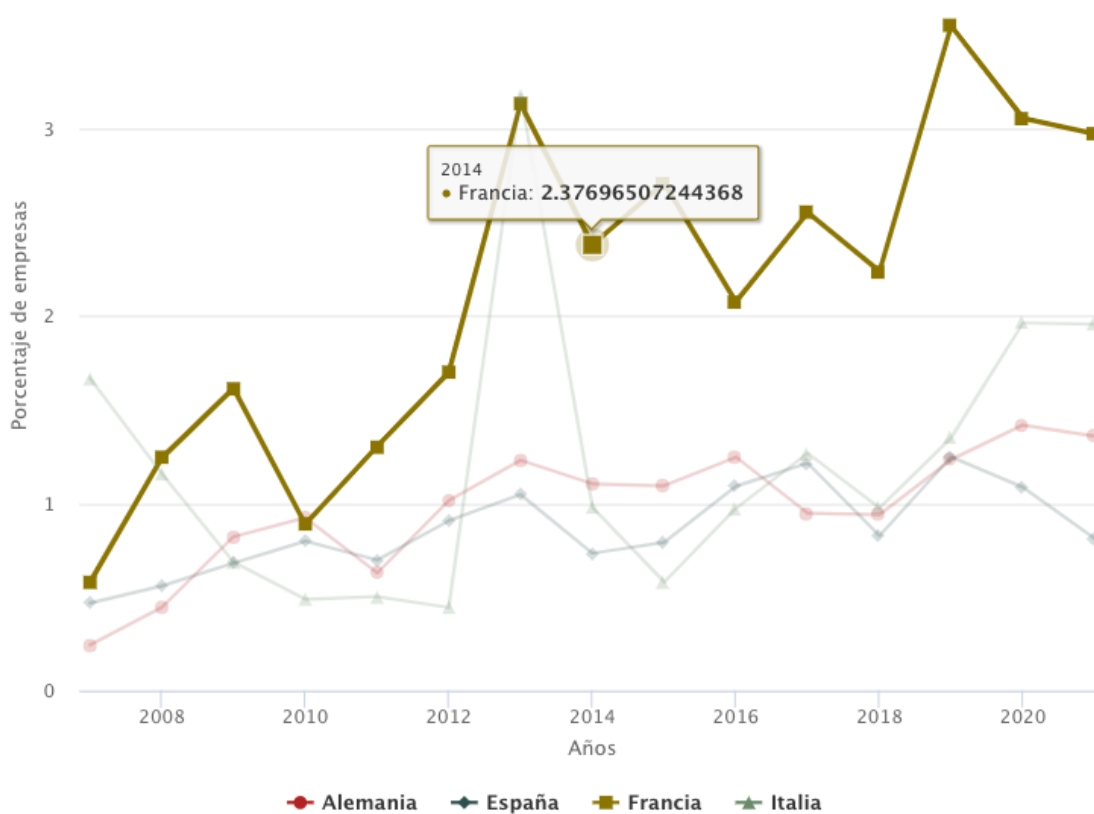


Figura 16. Gráfico de ejemplo realizado con High Charts sobre la evolución del uso de las RRSS en los sitios web de los 4 países.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

Cabe destacar que los tres gráficos mostrados en este apartado del trabajo son realmente interactivos, es decir, se pueden visualizar los valores que están marcados, pero esto no es posible incluirlo en un documento estático. Para que estos sean más representativos, se ha seleccionado la observación del país francés en el año 2014. Así, en cada uno de ellos como se muestran las interacciones propias de cada gráfico y paquete.

Finalmente, a pesar de haber tenido en cuenta tres paquetes de gráficos diferentes, se decidió utilizar Plotly, ya que, la manera más eficiente y eficaz de realizar gráficos en el lenguaje de programación R es a través de la librería “ggplot2”, y este, se encuentra muy ligada a ella, pudiendo así aprovechar todo el trabajo previo.

4.4.3 Implementación final

Finalmente, para mostrar la selección final del paquete utilizado para la herramienta de visualización, se muestra a continuación toda la implementación realizada con ella. De esta manera, se muestran los cuatro tipos de gráficos realizados: por países, por tamaños, por sectores y el mapa interactivo final a través del paquete “Plotly”.

Por ello, el primero de los gráficos realizado, es el desagregado por países y ya realizado a través de este paquete. De esta forma, se permite así, mostrar su interacción. Este, es el

anteriormente mostrado en la evaluación de las alternativas para la realización de los gráficos, en el apartado anterior. Corresponde con el gráfico mostrado en la *Figura 14*.

En segundo lugar, el gráfico generado para la demostración de la información desagregada por tamaño de empresa se muestra en la *Figura 17*. De esta forma, al haber sido realizado mediante Plotly, permite la interacción mostrada en el gráfico siguiente:

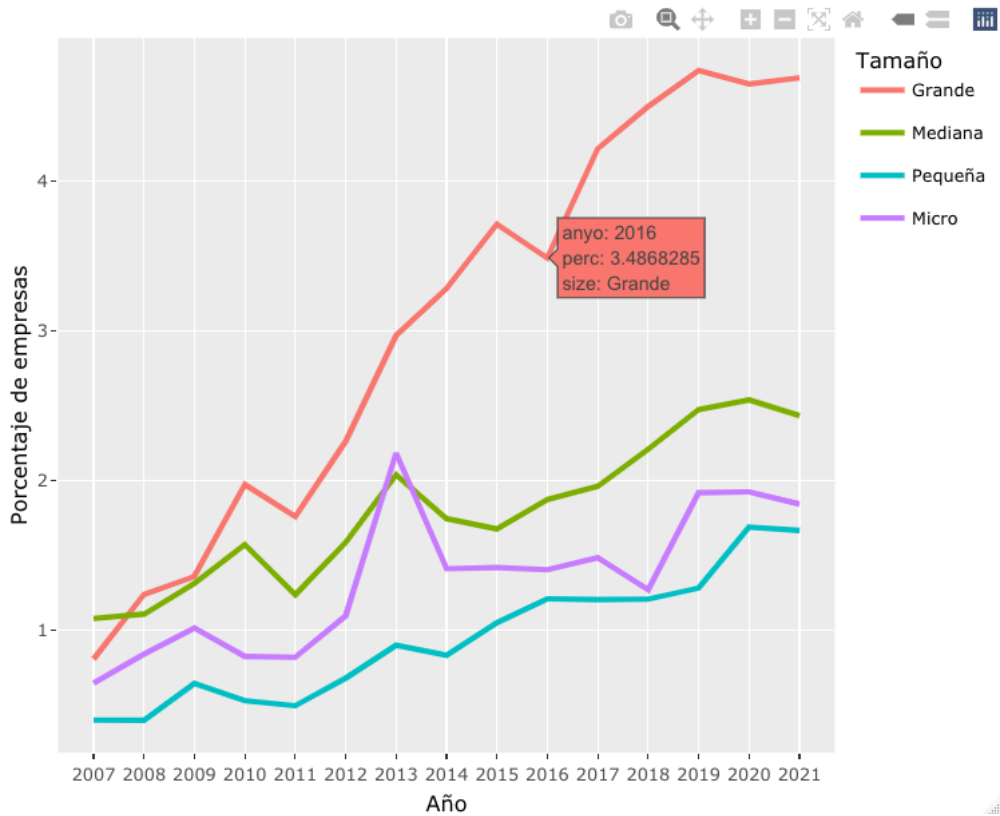


Figura 17. Gráfico realizado con Plotly, desagregado por tamaño de empresa de la evolución del término educación.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

Por otro lado, y del mismo modo, se muestra a través de la siguiente figura (*Figura 18*), la implementación del paquete Plotly en la realización del gráfico correspondiente a la desagregación por sector económico.

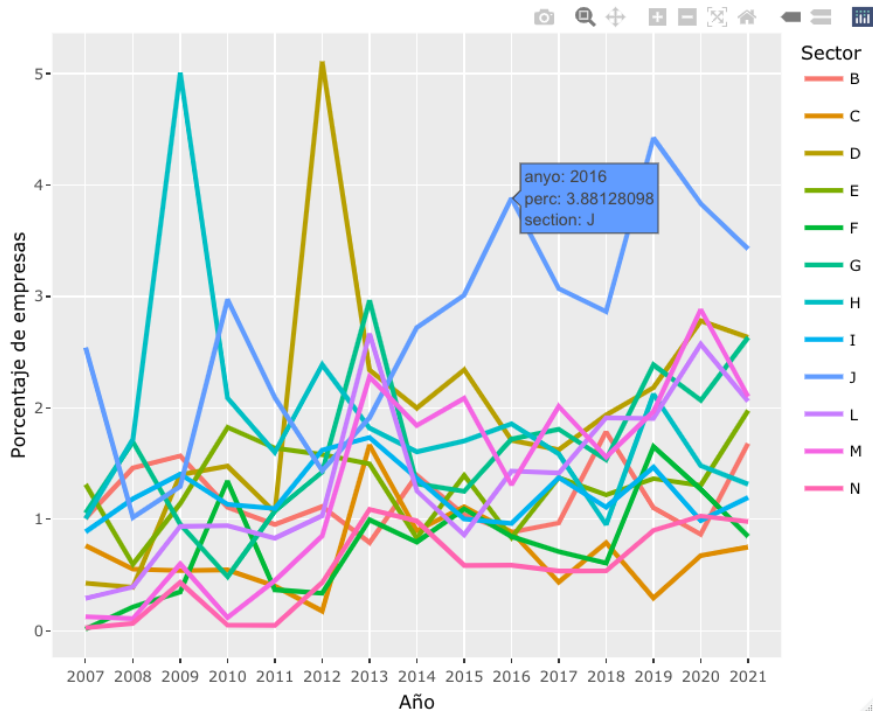


Figura 18. Gráfico realizado con Plotly, desagregado por sector de empresa de la evolución del término educación.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

Por último, se ha generado el mapa interactivo de los cuatro países europeos. En él, se permite observar la evolución de los porcentaje de empresas a lo largo del tiempo a través de una barra inferior inteligente. Este, se muestra en la Figura 19 siguiente:



Figura 19. Mapa interactivo generado con Plotly con los porcentajes de empresas de cada país a lo largo del tiempo.

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos “btrends”

4.4.4 Demostración final

En este apartado, se ha adjuntado un [enlace](#) de una demostración de la utilidad de la herramienta de visualización creada. Cabe destacar que, para el correcto funcionamiento de esta, el usuario debe encontrarse conectado al wifi de la Universidad Politécnica de Valencia (local o VPN). Con él, a través de la elección del término deseado, se generarán los cuatro gráficos explicados en este apartado.

CAPÍTULO 5. RESULTADOS

CAPÍTULO 5. RESULTADOS

En este capítulo se mostrarán los resultados obtenidos de la ejecución de los análisis, ya mencionados en los apartados anteriores, a través de la programación en lenguaje R. De esta manera, se podrá hacer una comprobación sobre qué hipótesis se aceptan y cuáles se rechazan finalmente. A través de este análisis, se busca realizar un claro y completo estudio de las tendencias en los sitios web de las empresas.

5.1 Análisis descriptivo

El estudio de los datos comenzará con un análisis descriptivo de las variables seleccionadas. Estas, al tratarse todas ellas de variables cualitativas, se mostrará en la siguiente tabla, la frecuencia absoluta y relativa de cada nivel de estas. Se han obtenido así los siguientes resultados, comenzando con la *Tabla 5*, la cual expresa las frecuencias de los niveles de la variable sector. En ella, se observa que, el sector que menos observaciones posee es el “Sector B”, mientras que el “Sector M”, con una frecuencia de 9,81%.

Tabla 5. Frecuencias relativas y absolutas de los niveles de la variable "Sector"

	Frecuencia Absoluta	Frecuencia Relativa
Sector = B	16865	5,23%
Sector = C	27177	8,42%
Sector = D	19133	5,93%
Sector = E	26350	8,16%
Sector = F	29299	9,08%
Sector = G	28267	8,76%
Sector = H	28948	8,97%
Sector = I	29114	9,02%
Sector = J	31381	9,72%
Sector = L	25449	7,89%
Sector = M	31672	9,81%
Sector = N	29096	9,01%
Sector = Total	322751	100%

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

En cuanto al análisis descriptivo de la variable “Año”, se muestran a través de la *Tabla 6*, del mismo modo que con la variable “Sector”, se muestra a través de la *Tabla 6*, sus frecuencias absolutas y relativas, referentes a cada año del estudio. En este, destaca el año 2018 con una frecuencia relativa cercana al 9%, mientras que el año 2007 es el que menor frecuencia presenta con un porcentaje cercano al 4%.

Tabla 6. Frecuencias relativas y absolutas de los niveles de la variable "Año"

	Frecuencia Absoluta	Frecuencia Relativa
Año = 2007	13533	4,07%
Año = 2008	25572	7,69%
Año = 2009	15354	4,61%
Año = 2010	15180	4,56%
Año = 2011	19338	5,81%
Año = 2012	18151	5,45%
Año = 2013	24590	7,39%
Año = 2014	24032	7,22%
Año = 2015	22278	6,70%
Año = 2016	27723	8,33%
Año = 2017	25677	7,72%
Año = 2018	29695	8,92%
Año = 2019	22975	6,90%
Año = 2020	23009	6,91%
Año = 2021	25644	7,71%
Año = Total	332751	100%

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

A través de la siguiente tabla (Tabla 7), se muestran las frecuencias propias de los diferentes niveles de la variable "País", "Tamaño", "Redes Sociales" y "Covid". Como se observa, el país que más peso posee en el estudio es Alemania con un 28.14%, mientras que, por su parte, Francia, es el que menos frecuencia presenta (21,75%). Del mismo modo, las empresas medianas, presentan una frecuencia superior al resto con un 27.86%, y las de tamaño micro, uno inferior con un 22.49%. Del mismo modo, se observa que la variable "Redes Sociales" toma el valor de "1", es decir, aparece en los sitios web con una frecuencia del 31.16%, mientras que la variable "Covid", lo hace con un 2.25%.

Tabla 7. Frecuencias relativas y absolutas de los niveles de la variable "País", "Tamaño", "Covid" y "Redes Sociales"

	Frecuencia Absoluta	Frecuencia Relativa
País = España	89449	27,71%
País = Italia	72261	22,39%
País = Alemania	90833	28,14%
País = Francia	70208	21,75%
País = Total	322751	100%
Tamaño = Micro	72577	22,49%
Tamaño = Pequeña	80857	25,05%
Tamaño = Mediana	89912	27,86%
Tamaño = Grande	79405	24,60%
Tamaño = Total	322751	100%
Covid = 1	7266	2,25%
Covid = 0	315485	97,75%
Covid = Total	322751	100%
Redes Sociales = 1	100570	31,16%
Redes Sociales = 0	222181	68,84%
Redes Sociales = Total	322751	100%

Fuente: Elaboración propia a partir de los datos obtenidos de la base de datos "btrends"

Por último, a través de la *Tabla 8*, se muestran las frecuencias relativas y absolutas de las diferentes redes sociales utilizadas para el estudio: "Instagram", "Twitter", "Facebook", "Youtube", "Linkedin" y "Whatsapp". La red social y por ello la variable que menor frecuencia relativa presenta es WhatsApp, con un valor del 0.48%. En cambio, Facebook, es la que mayor frecuencia presenta con un 10,42%.

Tabla 8. Frecuencias relativas y absolutas de los niveles de las variables propias de las redes sociales

	Frecuencia Absoluta	Frecuencia Relativa
Instagram = 1	8942	2,77%
Instagram = 0	313809	97,23%
Instagram = Total	322751	100%
Twitter = 1	24355	7,55%
Twitter = 0	298396	92,45%
Twitter = Total	322751	100%
Facebook = 1	33621	10,42%
Facebook = 0	289130	89,58%
Facebook = Total	322751	100%
Youtube = 1	15370	4,76%
Youtube = 0	307381	95,24%
Youtube = Total	322751	100%
Linkedin = 1	12314	3,82%
Linkedin = 0	310437	96,18%
Linkedin = Total	322751	100%
Whatsapp = 1	1561	0,48%
Whatsapp = 0	321190	99,52%
Whatsapp = Total	322751	100%

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

5.2 Análisis del uso conjunto de redes sociales

Para continuar con el estudio de los resultados de los datos, se ha realizado un estudio de posibles correlaciones. Este, hace referencia a la posible relación del uso de las redes sociales: "Youtube", "Facebook", "Whatsapp", "LinkedIn", "Twitter" e "Instagram". Se ha decidido realizar este estudio a través de la comprobación de dichos términos en los sitios web de las empresas durante el año 2021. Para realizar este estudio de posibles correlaciones, al tratarse de variables cualitativas discretas todas ellas, (binarias en este caso), se ha utilizado el método estadístico del coeficiente "phi", a través de su función en el lenguaje de programación R. Y se han obtenido a través del paquete "vcd" el cual, proporciona herramientas y funciones para visualizar y analizar datos categóricos (Meyer, 2020). Se ha obtenido los siguientes valores:

Tabla 9. Matriz de coeficientes phi de las diferentes redes sociales

	Instagram	Youtube	Facebook	Whatsapp	Linkedin	Twitter
Instagram	X	0.43947	0.5716	0.0804	0.3102	0.4320
Youtube	0.43947	X	0.4673	0.0309	0.4446	0.4858
Facebook	0.5716	0.4673	X	0.0830	0.4308	0.5861
Whatsapp	0.0804	0.0309	0.0830	X	0.0346	0.0535
LinkedIn	0.3102	0.4446	0.4308	0.0346	X	0.5141
Twitter	0.4320	0.4858	0.5861	0.0535	0.5141	X

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

De esta forma, y mediante la *Tabla 9*, se sacan las siguientes conclusiones: Al tratarse de valores ampliamente superiores al valor de $[-0.3 - 0.3]$, se considera que todos los pares de variables a excepción de los que poseen a la variable Whatsapp, poseen un nivel de asociación fuerte. Cabe destacar que a pesar de todos los pares poseer esa asociación “fuerte”, la variable cuyos pares presentan mayor nivel de asociación es Facebook.

Del mismo modo, todos los valores presentados son positivos, por lo que, todos los pares de variables poseen una asociación positiva, es decir, las redes sociales son complementarias entre ellas, no sustitutivas.

La variable Whatsapp por su parte, es la que menos asociación presenta a través de sus pares de variables, cabía esperar ese resultado, ya que, además era la que menor frecuencia absoluta presentaba desde un inicio. Al ser todos sus valores menores a $[-0.1 - 0.1]$ se considera que el nivel de asociación de los pares de esta variable es nulo o muy débil, por no llegar a ser exactamente 0.

De esta forma, se evalúa a la cuarta hipótesis planteada (H4): *El uso de las diferentes redes sociales en los sitios web implica el uso de otras redes sociales en estos*. Como vemos con los resultados obtenidos, la hipótesis se acepta. De las seis redes sociales escogidas, se ha observado que cinco de ellas poseen una clara relación en cuanto al uso dentro de los sitios web se refiere. Es cierto que, la red social “Whatsapp” no posee esa cierta asociación, pero, debido a que presenta un bajo número de observaciones, se acepta la hipótesis planteada.

5.3 Análisis comparativo

En este apartado se ha realizado un estudio comparativo sobre cómo se han implementado en los sitios web de las empresas de los cuatro países muestreados, los términos referentes a la variable Covid y los referentes a la variable Redes Sociales. De esta forma, se han evaluado las dos hipótesis iniciales referentes a los países procedentes de la muestra. (H1 y H2).

Análisis comparativo variable Covid

Por su parte el estudio referente a la enfermedad COVID-19, a través de la variable “Covid”, ha sido realizado a través de los datos disponibles desde el año 2020 al 2021. De esta forma, se abarcan los dos años donde la población mundial y por ello sus empresas, se vio afectada por esta pandemia. Los resultados obtenidos han sido por tanto han sido los siguientes:

Tabla 10. Tabla de contingencia entre las variables "Covid" y "Tamaño" en los años referentes a la pandemia

	Grande	Mediana	Pequeña	Micro
Covid = 1	2806	2260	819	1341
Covid = 0	8506	10669	11050	11202

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos “btrends”

A través de la *Tabla 10*, se ha realizado el análisis comparativo referente a los tamaños de empresas y su tendencia a la información del COVID.19 en sus sitios web. Como se puede observar, en cuanto a la frecuencia absoluta, las empresas de tamaño grande son las que más veces (2806) han implementado en sus sitios web la variable “Covid”.

Tras este análisis, se ha realizado la comprobación estadística de la quinta hipótesis (H5): “*La tendencia a la información relativa a la enfermedad COVID-19 es mayor en las empresas más concienciadas en la imagen de marca y en generar confianza en sus clientes.*”

Para ello se ha realizado la prueba de proporciones de dos muestras, a través de la función “prop.test” en lenguaje de programación R. Al tener cuatro niveles dentro de la variable tamaño, se han realizado los seis pares posibles de comprobaciones según esta prueba, por tal de determinar cuál es la proporción mayor.

Se ha establecido el parámetro correspondiente a que la hipótesis alternativa, sea igual a que la proporción del primer nivel sea superior a la del segundo. Así se han establecido las siguientes hipótesis nulas y alternativas para cada una de las pruebas realizadas:

$H_0 =$ La proporción del primer grupo es menor que la del segundo grupo

$H_1 =$ La proporción del primer grupo es mayor que la del segundo grupo

Tabla 11. P. Valores de las pruebas de proporción I

Pruebas de Proporción			
Niveles	Estadístico	Grados de Libertad	P. Valor
Grande - Pequeña	519.15	1	< 2.2e-16
Grande - Micro	834.43	1	< 2.2e-16
Mediana - Pequeña	162.43	1	< 2.2e-16
Mediana - Micro	384.39	1	< 2.2e-16

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos “btrends”

Como se puede observar en la *Tabla 11*, todas las pruebas estadísticas realizadas son significativas y por ello se rechaza la hipótesis nula. De esta forma, se acepta que la proporción mayor de empresas que hablan sobre la enfermedad del COVID-19 en sus sitios web, corresponde al grupo de empresas de tamaños Grande y Mediana. En cambio, las empresas Pequeñas y Micro son las que menor proporción presentan. Por todo ello, se rechaza la quinta hipótesis (H5), ya que, las empresas de mayor tamaño son las que realmente informan en mayor proporción de la enfermedad en sus sitios web.

A continuación, y a través de la siguiente tabla de contingencia (*Tabla 12*), se ha realizado el análisis comparativo entre los diferentes países muestreados y su tendencia a la información sobre la enfermedad en sus sitios web.

Tabla 12. Tabla de contingencia entre las variables "Covid" y "País" en los años referentes a la pandemia

	España	Francia	Alemania	Italia
Covid = 1	2060	1873	1708	1585
Covid = 0	10684	9005	11813	9925

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

La información, en términos absolutos, a través de sus sitios web sobre la enfermedad COVID-19 es mayor en España que en el resto de los países muestreados. Por su parte, Italia, presenta la menor frecuencia absoluta (1585). Posteriormente, a este análisis sobre sitios web informantes de la enfermedad COVID-19 de cada uno de los países muestreados, y del mismo modo que en el análisis anterior, se ha realizado la comprobación estadística de la primera hipótesis (H1). *Los países que presentan una mayor proporción de contagios de la enfermedad COVID-19, poseen una mayor tendencia a la información sobre esta enfermedad en sus sitios web.*

Para ello se ha realizado el siguiente análisis de la misma naturaleza que los anteriores (prueba de proporciones de dos muestras). Se ha establecido el parámetro correspondiente a que la hipótesis alternativa, sea igual a que la proporción del primer nivel sea superior a la del segundo. Así se han establecido las siguientes hipótesis nulas y alternativas para cada una de las pruebas realizadas:

H_0 = La proporción del primer grupo es menor que la del segundo grupo

H_1 = La proporción del primer grupo es mayor que la del segundo grupo

Tabla 13. P. Valores de las pruebas de proporciones II

Pruebas de Proporción			
Países	Estadístico	Grados de Libertad	P. Valor
España - Francia	7.11	1	0.99
España - Alemania	50.03	1	7.54e-13
Francia - Italia	40.89	1	8.02e-11
Francia - Alemania	92.05	1	< 2.2e-16
Italia - Alemania	8.62	1	0.001

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

A través de los p. valores, de la *Tabla 13*, mayores al nivel de significación 0.05, se acepta la hipótesis nula (H0), y se rechaza esta cuando el p. valor de la prueba es menor que el nivel de significación. Por ello, el país que mayor proporción de sitios web en los que informa sobre la enfermedad COVID-19 es Francia, mientras que el menor es Alemania.

De esta forma se rechaza la hipótesis inicial (H1) la cual exponía que los países que mayor proporción de contagios presentaban (Francia y Alemania), eran las empresas que más debían de haber informado sobre la enfermedad. Esto es porque Alemania es el país que menor porcentaje relativo a los términos del COVID-19 presenta.

Por otro lado, se ha realizado el análisis comparativo de los dos años durante los cuales la población mundial sufrió las consecuencias de la pandemia. Esta, se muestra en la siguiente tabla (*Tabla 14*), donde se han comparado los sitios web pertenecientes al año 2020 y a los del 2021. Como se observa, el año que mayor frecuencia relativa y absoluta presenta es el año 2020, año en el que la información de la situación y gravedad de la enfermedad fue más relevante. Fue durante este año cuando explotó la pandemia mundial, y todos los sitios web del mundo se hicieron eco de la noticia.

Tabla 14. Tabla de contingencia entre las variables "Covid" y "Año" en los años referentes a la pandemia

	2020	2021
Covid = 1	3951	3275
Covid = 0	19058	22369

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

Por último, y por tal de finalizar este primer análisis comparativo, se ha realizado un estudio de los diferentes sectores económicos y como ellos, a través de los sitios web de sus empresas informaron sobre la pandemia. Los resultados obtenidos se muestran en la *Tabla 15* siguiente:

Tabla 15. Tabla de contingencia entre las variables "Covid" y "Sector" en los años referentes a la pandemia

	B	C	D	E	F	G	H	I	J	L	M	N
Covid = 1	165	451	470	691	278	526	719	709	945	598	905	769
Covid = 0	2370	3429	2610	3383	4161	3630	3630	3762	3841	3119	3864	3628

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

Los sectores J (Información y comunicaciones), M (Actividades profesionales, científicas y técnicas) y N (Actividades administrativas y servicios auxiliares) son los que más información relativa al COVID-19 presentan en sus sitios web. Por contrario, sectores como el B (Industrias extractivas) y F (Construcción) son los que menor frecuencia a la información presentan.

Análisis comparativo variable Redes Sociales

En cuanto al estudio de la variable Redes Sociales, se han realizado con los datos completos del estudio, es decir, desde el año 2007 al 2021. De esta forma, se observa el crecimiento que ha tenido estas, a lo largo del tiempo y la importancia relativa que poseen en los cuatro países muestreados y en relación con las diferentes variables del estudio.

Tabla 16. Tabla de contingencia entre las variables "Redes Sociales" y "País"

	España	Francia	Alemania	Italia
Redes Sociales = 1	6251	5931	5333	5783
Redes Sociales = 0	6493	4947	8188	5727

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

A través de esta primera tabla (Tabla 16), se observa el uso de la variable "Redes Sociales" en cada uno de los cuatro países muestreados. Como se muestra en ella, el país que mayor frecuencia absoluta presenta es España (6251), mientras que el que presenta una menor frecuencia es Alemania con 5333 observaciones.

Por otro lado, se observa a través de la *Figura 20*, la comparativa de las evoluciones del uso de las redes sociales en los sitios web de los cuatro países. Como vemos, la evolución de Alemania es la menor, ya que, pese de haber comenzado todos los países desde un número de observaciones parecido, este, ha sido el que menos ha aumentado. Por otro lado, destaca las evoluciones de Francia y de España por ser muy semejantes.

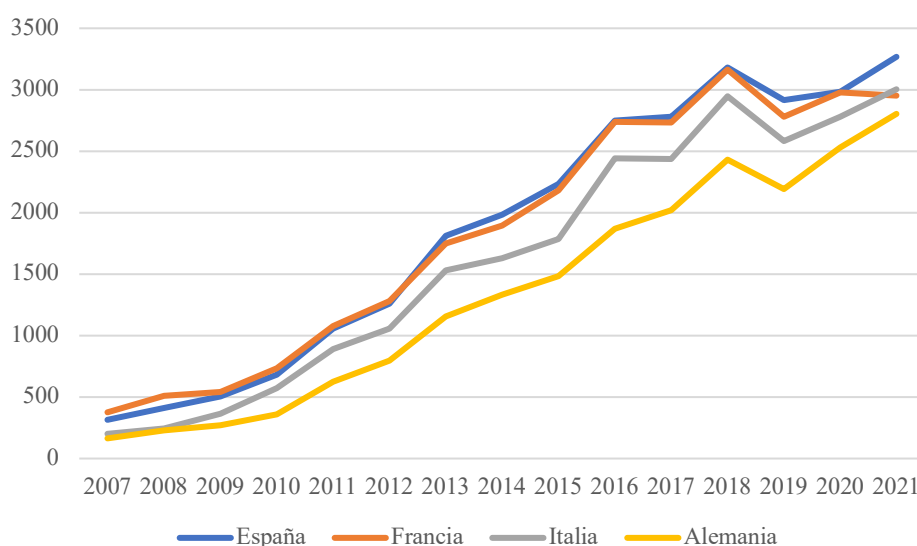


Figura 20. Gráfico de líneas de la evolución del uso de la variable RRSS en los sitios web de los cuatro países.

Fuente: Elaboración propia mediante los datos obtenidos de la base de datos "btrends"

De nuevo, tras realizar este análisis comparativo de los sitios web que utilizan la variable Redes Sociales, de cada uno de los países muestreados, y del mismo modo que en el análisis de la enfermedad COVID-19, se ha realizado la comprobación estadística de la primera hipótesis (H2) *El país que posee una mayor potencia tecnológica posee a su vez*

una mayor tendencia a la utilización de las redes sociales en los sitios web de sus empresas.

De igual manera, se ha procedido a la realización de la prueba estadística de proporciones de dos muestras para la aceptación o rechazo de la misma. Así se han establecido las hipótesis nulas y alternativas de cada una de las pruebas, siendo las mismas que en los otros apartados, obteniendo los resultados siguientes:

$$H_0 = \text{La proporción del primer grupo es menor que la del segundo grupo}$$

$$H_1 = \text{La proporción del primer grupo es mayor que la del segundo grupo}$$

Tabla 17. P. Valores de las pruebas de proporciones III

Pruebas de Proporción			
Países	Estadístico	Grados de Libertad	P. Valor
España - Alemania	245.29	1	< 2.2e-16
Francia - Alemania	551.01	1	< 2.2e-16
Italia - Alemania	293.36	1	< 2.2e-16

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

De esta forma, a través de los resultados mostrados en la *Tabla 17*, se observa que, a través del análisis estadístico, se rechaza la hipótesis nula en las tres pruebas realizadas. Así, se rechaza también la hipótesis inicial (H2), ya que, Alemania, uno de los países europeos más punteros en cuanto a tecnología, de la UE (y el más avanzado a nivel tecnológico de los cuatro muestreados) es el país que menos utiliza las redes sociales en sus sitios web.

En segundo lugar, se ha realizado un análisis comparativo de los diferentes sectores económicos. Se ha realizado la comparación de como los sitios web de sus empresas han implementado en ellos la variable "Redes Sociales". De esta forma, a través de la *Tabla 18*, se muestra la información obtenida:

Tabla 18. Tabla de contingencia entre las variables "Sector" y "Redes Sociales"

	Redes Sociales = 1	Redes Sociales = 0
B	3534	16865
C	7380	27177
D	5910	19133
E	7343	26350
F	7578	29299
G	9039	28267
H	8240	28948
I	9117	29114
J	14234	31381
L	7638	25449
M	11548	31672
N	9009	29096

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

Como se muestra, el sector que mayor frecuencia absoluta presenta, sobre la utilización de la variable "Redes Sociales", es el sector M (Actividades profesionales, científicas y técnicas), mientras el que menor frecuencia presenta es el sector B (Industrias extractivas). Tras este primer análisis comparativo de la variable "Redes Sociales" y de los sectores económicos, se ha realizado la comprobación estadística, a través de la comparación de proporciones, de la tercera hipótesis (H3). *El uso de redes sociales es mayor en los sectores económicos que se encuentran más cerca del consumidor final.*

Para ello se ha realizado el siguiente análisis de la misma naturaleza que previamente realizado. El primer paso para la realización de este ha sido agrupar los sectores por mayor o menor cercanía al consumidor final. De esta forma, basándose en la hipótesis inicial planteada, el grupo de sectores agrupado por ser los más cercanos al consumidor final son: G (Comercio al por mayor y al por menor vehículos de motor) e I (Hostelería), mientras que el resto de los sectores se ha agrupado como más lejanos al consumidor final.

Se ha establecido el parámetro correspondiente a que la hipótesis alternativa, sea igual a que la proporción del primer nivel sea superior a la del segundo. Así se han establecido las siguientes hipótesis nulas y alternativas para cada una de las pruebas realizadas:

H_0 = La proporción del primer grupo es menor que la del segundo grupo

H_1 = La proporción del primer grupo es mayor que la del segundo grupo

Tabla 19. P. Valores de las pruebas de proporciones IV

Pruebas de Proporción			
Grupos de Sectores	Estadístico	Grados de Libertad	P. Valor
Más cercanos (G-I) - Resto	7.49	1	0.003

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos "btrends"

Como se observa a través de la *Tabla 19*, el p. valor de la prueba es inferior al nivel de significación del 5%, y por ello, se rechaza la hipótesis nula. Por ello, la proporción de utilización de la variable “Redes Sociales” es mayor en los sectores más próximos al consumidor final, G (Comercio al por mayor y al por menor vehículos de motor) e I (Hostelería). Por ello se acepta la hipótesis inicial planteada (H3).

Por otro lado, se ha realizado también y de igual forma que con la variable “Covid”, un análisis comparativo del uso de la variable “Redes Sociales” en función del tamaño de empresa. De esta forma y a través de la *Tabla 20*, se ha diferenciado entre el uso propio de las empresas micro, pequeñas, medianas y grandes. Así, se observa como son las empresas de mayor tamaño (Grandes y Medianas), las que más frecuencia absoluta y relativa presentan. Por ello, son las empresas de menor tamaño, las Micro y Pequeñas, las que presentan una frecuencia menor.

Tabla 20. Tabla de contingencia entre las variables "Redes Sociales" y "Tamaño"

	Grande	Mediana	Pequeña	Micro
RRSS = 1	31990	28900	21924	17756
RRSS = 0	47415	61012	58933	54821

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos “btrends”

Del mismo modo, se ha realizado un análisis comparativo de las observaciones por cada uno de los años estudiados. En esta tabla (*Tabla 21*), se muestra como los años que más frecuencia absoluta presentan son los últimos 4 años (2018-2021). Del mismo modo, son estos últimos cuatro años también los que presentan una mayor frecuencia relativa. Esto, como es lógico se debe al grandísimo incremento de usuarios de las redes sociales, año tras año.

Tabla 21. Tabla de contingencia entre las variables "Años" y "Redes Sociales"

	Redes Sociales = 1	Redes Sociales = 0
2007	1056	12477
2008	1393	14179
2009	1681	13673
2010	2344	12836
2011	3652	15686
2012	4396	13755
2013	6250	18430
2014	6846	17186
2015	7682	14596
2016	9793	17930
2017	9981	15696
2018	11724	17971
2019	10474	12501
2020	11268	11741
2021	12030	13614

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos “btrends”

5.4 Análisis multivariante

En este estudio se ha decidido realizar un modelo de regresión logística en el que se incluyan los países, los sectores y los tamaños de las empresas pertenecientes al año 2021. Se ha decidido realizar una regresión de este tipo (logit), ya que, la variable dependiente seleccionada ha sido la variable “Covid”, la cual, es binaria, y, por ello, puede tomar dos valores: 0 y 1. Del mismo modo también se ha incluido en el estudio como variable independiente, la variable: “Redes Sociales”.

Modelo de regresión logística

De esta forma, a través del programa utilizado (R), se realizó el modelo siguiente de regresión logit en el cual se incluyeron todas las variables restantes como independientes y la variable “Covid” como dependiente, y de esta forma, se han obtenido los siguientes resultados, los cuales se muestran a través de la *Tabla 22*.

$$Covid = f(\text{Sector}, \text{País}, \text{Tamaño}, \text{Redes Sociales})$$

Tabla 22. Estimaciones del modelo de regresión logística

	Estimate	Odds Ratio	Error	Pr(> t)
(Intercepto)	0,11020	1,11650	0,01108	2,63E-23
País = Alemania	-0,01433	0,98577	0,00587	0,01458
País = Italia	-0,00981	0,99024	0,00605	0,10457
País = España	0,01155	1,01162	0,00591	0,05042
Tamaño = Mediana	-0,04881	0,95236	0,00583	6,25E-17
Tamaño = Micro	-0,12269	0,88454	0,00602	1,48E-91
Tamaño = Pequeña	-0,09843	0,90626	0,00598	1,22E-60
Sector = C	0,00001	1,00001	0,01144	0,99896
Sector = D	0,05068	1,05199	0,01188	1,99E-05
Sector = E	0,06146	1,06339	0,01117	3,82E-08
Sector = F	-0,03219	0,96832	0,01105	0,00357
Sector = G	0,01032	1,01037	0,01129	0,36086
Sector = H	0,05573	1,05732	0,01110	5,13E-07
Sector = I	0,07802	1,08115	0,01105	1,67E-12
Sector = J	0,05984	1,06166	0,01100	5,44E-08
Sector = L	0,06584	1,06806	0,01145	8,94E-09
Sector = M	0,06481	1,06696	0,01097	3,48E-09
Sector = N	0,06609	1,06832	0,01110	2,63E-09
Redes Sociales	0,10343	1,10897	0,00417	5,57E-134

Fuente: Elaboración propia a través de los datos obtenidos de la base de datos “btrends”

Ahora, se analizarán las variables que según el modelo se hayan considerado significativas. Para ello, se debe tener en cuenta que se está modelando la probabilidad de que aparezca el término covid en la página web de una empresa. Como hay variables cualitativas codificadas como independientes, se han establecido algunos niveles base, ya que los coeficientes mostrados se interpretan como cambios con respecto al nivel base. Así, se han establecido: Francia como el nivel base para la variable “País”, grande para la variable “Tamaño”, y B para la variable “Sector”. De esta forma, los coeficientes se interpretan de la siguiente forma:

El coeficiente de País = Alemania es -0.01433, por lo que el odds-ratio es $\exp(-0.01433) = 0.985772186$. La interpretación de este odds-ratio es: Si la empresa está ubicada en Alemania, la probabilidad de que incluya el término covid se multiplica por 0.986. Es decir, el hecho de ser alemana hace que la probabilidad de hablar de covid disminuya en 1.4 puntos porcentuales con respecto a la que tendría una empresa equivalente en Francia.

El coeficiente de Tamaño = Mediana es -0,04881, y por tanto su odds-ratio es 0,95236. La interpretación de este odds-ratio es: Si la empresa es mediana, la probabilidad de que incluya el término covid en su sitio web se multiplica por 0.952. Es decir, el hecho de ser una empresa mediana hace que la probabilidad de hablar de covid disminuya en 4.8 puntos porcentuales con respecto a la que tendría una empresa equivalente de tamaño grande.

El coeficiente de Tamaño = Micro es -0,12269, y por tanto su odds-ratio es 0,88454. La interpretación de este odds-ratio es: Si la empresa es micro, la probabilidad de que incluya el término covid en su sitio web se multiplica por 0.884. Es decir, el hecho de ser una empresa micro hace que la probabilidad de hablar de covid disminuya en 12.2 puntos porcentuales con respecto a la que tendría una empresa equivalente de tamaño grande.

El coeficiente de Tamaño = Pequeña es -0,09843, y por tanto su odds-ratio es 0,90626. La interpretación de este odds-ratio es: Si la empresa es pequeña, la probabilidad de que incluya el término covid en su sitio web se multiplica por 0.906. Es decir, el hecho de ser una empresa pequeña hace que la probabilidad de hablar de covid disminuya en 9.8 puntos porcentuales con respecto a la que tendría una empresa equivalente de tamaño grande.

El coeficiente de Sector = D es 0,05068, y por tanto su odds-ratio es 1,05199. La interpretación de este odds-ratio es: Si la empresa pertenece al sector D (Suministro de energía eléctrica, gas, vapor y a. a), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.052. Es decir, el hecho de ser una empresa del sector D hace que la probabilidad de hablar de covid aumente en 5.1 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = E es 0,06146, y por tanto su odds-ratio es 1,06339. La interpretación de este odds-ratio es: Si la empresa pertenece al sector E (Suministro de agua, actividades de saneamiento), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1,063. Es decir, el hecho de ser una empresa del sector E hace que la probabilidad de hablar de covid aumente en 6.1 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = F es -0,03219, y por tanto su odds-ratio es 0,96832. La interpretación de este odds-ratio es: Si la empresa pertenece al sector F (Construcción), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.052. Es decir, el hecho de ser una empresa del sector F hace que la probabilidad de hablar de covid disminuya en 3.2 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = H es 0, 0,05573, y por tanto su odds-ratio es 1,05732. La interpretación de este odds-ratio es: Si la empresa pertenece al sector H (Transporte y almacenamiento), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.057. Es decir, el hecho de ser una empresa del sector H hace que la probabilidad de hablar de covid aumente en 5.5 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = I es 0,07802, y por tanto su odds-ratio es 1,08115. La interpretación de este odds-ratio es: Si la empresa pertenece al sector I (Hostelería), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.081. Es decir, el hecho de ser una empresa del sector I hace que la probabilidad de hablar de covid aumente en 7.8 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = J es 0,05984, y por tanto su odds-ratio es 1,06166. La interpretación de este odds-ratio es: Si la empresa pertenece al sector J (Información y comunicaciones), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.062. Es decir, el hecho de ser una empresa del sector J hace que la probabilidad de hablar de covid aumente en 5.9 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = L es 0,06584, y por tanto su odds-ratio es 1,06806. La interpretación de este odds-ratio es: Si la empresa pertenece al sector L (Actividades inmobiliarias), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.068. Es decir, el hecho de ser una empresa del sector L hace que la probabilidad de hablar de covid aumente en 6.5 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = M es 0,06481, y por tanto su odds-ratio es 1,06696. La interpretación de este odds-ratio es: Si la empresa pertenece al sector M (Actividades profesionales, científicas y técnicas), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.067. Es decir, el hecho de ser una empresa del sector M hace que la probabilidad de hablar de covid aumente en 6.5 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

El coeficiente de Sector = N es 0,06609, y por tanto su odds-ratio es 1,06832. La interpretación de este odds-ratio es: Si la empresa pertenece al sector N (Actividades administrativas y servicios auxiliares), la probabilidad de que incluya el término covid en su sitio web se multiplica por 1.068. Es decir, el hecho de ser una empresa del sector N hace que la probabilidad de hablar de covid aumente en 6.6 puntos porcentuales con respecto a la que tendría una empresa equivalente del sector B (Industrias extractivas).

Por último, coeficiente de la variable Redes Sociales es 0,10343, y por tanto su odds-ratio es 1,10897. La interpretación de este odds-ratio es: Si la empresa en el sitio web de la empresa aparece la variable “Redes Sociales”, la probabilidad de que incluya el término covid en su sitio web se multiplica por 1,109. Es decir, el hecho de hablar de las redes sociales hace que la probabilidad de hablar de covid aumente en 10.3 puntos porcentuales con respecto a la que tendría una empresa equivalente que no hablara de las redes sociales.

CAPÍTULO 6. CONCLUSIONES

CAPÍTULO 6. CONCLUSIONES

Este Trabajo Final de Grado se ha centrado en la generación de una herramienta automatizada de visualización, que permitiera el estudio del análisis de tendencias de las empresas a través del análisis de sus sitios web y posteriormente su validación a través del análisis de los resultados.

Para ello se formularon cinco hipótesis iniciales, a través de las cuales, se trataba de analizar las tendencias de las empresas clasificadas por país, tamaño y sector. Se formuló también un modelo de regresión logística explicando la posible relación entre los países y sectores que informan sobre la enfermedad del COVID-19 y el uso de las redes sociales (términos seleccionados para la demostración de la potencia de la herramienta).

De esta forma, se han cumplido los objetivos propuestos en un inicio, que han hecho que la generación de este trabajo haya sido un éxito, de la siguiente manera: Se han logrado generar la herramienta de visualización propia del estudio. Esto, ha sido a través de la programación en lenguaje R y las consultas generadas en SQL. Con estas, se ha permitido generar todas las líneas de código necesarias para la puesta en marcha de esta. Finalmente, a través de la evaluación de los diferentes paquetes disponibles para llevarla a cabo, se ha generado la ya mencionada herramienta, permitiendo así la generación de una demostración mostrada en el trabajo.

En relación con el segundo objetivo, se han establecido los términos mediante los cuales se ha realizado la extracción de datos, a través de la generación de las variables propias del estudio, tanto las referentes a la enfermedad COVID-19, como las referentes a las Redes Sociales. De nuevo, gracias al lenguaje R y SQL, se ha permitido lograr este objetivo, a través de la programación de los códigos oportunos.

Finalmente, a través de los diferentes análisis realizados (descriptivo, comparativo, correlaciones y modelo de regresión logística), y de la comprobación de las cinco hipótesis iniciales formuladas, se ha logrado estudiar la relación existente entre el uso de las redes sociales y de la información relativa a la enfermedad del COVID-19 en los sitios web de las empresas.

Cabe destacar que el principal objetivo del trabajo era la generación de la herramienta automatizada de estudio de los sitios web de las empresas. Es cierto que, para la comprobación de su utilidad se ha realizado el estudio de las ya mencionadas hipótesis iniciales, pero este no era el objetivo primordial. Por ello se concluye que se han logrado todos los objetivos propuestos para la realización del trabajo.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Agresti, A. (2012). *Categorical data analysis* (Vol. 792). John Wiley & Sons.
- Allaire, J. (2012). RStudio: integrated development environment for R. *Boston, MA*, 770(394), 165-171.
- Apvăloaie, E. I. (2014). The impact of the internet on the business environment. *Procedia Economics and finance*, 15, 951-958.
- Auxier, B., & Anderson, M. (2021). Social media use in 2021. Pew Research Center, 1, 1-4
- Duong, C. T. P. (2020). Social media. A literature reviews. *Journal of Media Research- Revista de Studii Media*, 13(38), 112-126.
- Ekström, J. (2011). The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate. *UCLA: Department of Statistics, UCLA*. Retrieved from <https://escholarship.org/uc/item/7qp4604r> (30 de abril de 2023)
- Enaworu, N. M., & Yusuf, N. (2021). COVID-19 and Beyond: Redefining the Future of Business. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(3), 75.
- European Commission. (2019). The 2019 EU Industrial R&D Investment Scoreboard. Luxembourg: Publications Office of the European Union, 2020. Recuperado desde: https://iri.jrc.ec.europa.eu/scoreboard/2019-eu-industrial-rd-investment-scoreboard#field_infographicscoreboard
- Fitzsimmons, A. B., Qin, Y. S., & Heffron, E. R. (2022). Purpose vs mission vs vision: persuasive appeals and components in corporate statements. *Journal of Communication Management*, 26(2), 207-219.
- Gesmann, M., de Castillo, D., & Gesmann, M. M. (2011). Using the Google Visualisation API with R. *The R Journal*, 3(2):40-44, December 2011.
- INE (Instituto Nacional de Estadística). (s.f.). Clasificación Nacional de Actividades Económicas (CNAE-2009). Real Decreto 475/2007 de 13 de abril. Recuperado desde: <https://www.ine.es/daco/daco42/clasificaciones/cnae09/cnae2009.pdf>
- Kang, J., Diao, Z. and Zanini, M.T. (2021), "Business-to-business marketing responses to COVID-19 crisis: a business process perspective", *Marketing Intelligence & Planning*, Vol. 39 No. 3, pp. 454-468. <https://doi.org/10.1108/MIP-05-2020-0217>
- Kargl, M., Mehic, A., Prodanovic, Z., & Seywald, D. (2018). Interactive Data Visualization in R, Graz University of Technology. Recuperado desde: <https://courses.isds.tugraz.at/ivis/surveys/ss2018/ivis-ss2018-g5-r-datavis-survey.pdf>

- Mahmud, A., Ding, D., & Hasan, M. M. (2021). Corporate social responsibility: Business responses to coronavirus (COVID-19) pandemic. *SAGE open*, 11(1), 2158244020988710.
- Meyer D, Zeileis A, Hornik K (2023). *vcd: Visualizing Categorical Data*. R package version 1.4-11, <<https://CRAN.R-project.org/package=vcd>>.
- Nick, T. G., & Campbell, K. M. (2007). Logistic regression. *Topics in biostatistics*, 273-301.
- O'Leary, M., & O'Leary, M. (2015). MySQL and MariaDB. *Cyber Operations: Building, Defending, and Attacking Modern Computer Networks*, 565-604.
- Pinzaru, F., Zbucea, A., & Anghel, L. (2020). The Impact of the COVID-19 Pandemic on Business. A preliminary overview. *Strategica. Preparing for Tomorrow, Today*, 721-730.
- Reddy, B. V., & Gupta, A. (2020). Importance of effective communication during COVID-19 infodemic. *Journal of family medicine and primary care*, 9(8), 3793.
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., ... & Sievert, M. C. (2021). Package 'plotly'.
- Spence, M., & Hamzaoui Essoussi, L. (2010). SME brand building and management: an exploratory study. *European Journal of Marketing*, 44(7/8), 1037-1054.
- Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of applied developmental psychology*, 29(6), 420-433.
- Team, R. C. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Wickham, H. (2011). *ggplot2*. *Wiley interdisciplinary reviews: computational statistics*, 3(2), 180-185.
- Wilson, K., Fornasier, S., & White, K. M. (2010). Psychological predictors of young adults' use of social networking sites. *Cyberpsychology, behavior, and social networking*, 13(2), 173-177.
- World Health Organization. (2020). Coronavirus disease (COVID-19): situation report, 198. World Health Organization. WHO from national authorities by 10:00 CEST 5 August 2020. Recuperado desde: <https://apps.who.int/iris/handle/10665/333735>
- World Health Organization. (n.d.). WHO Coronavirus (COVID-19) Dashboard. Retrieved March 25, 2023, from <https://covid19.who.int/table>

Zhang, Y., & Leung, L. (2015). A review of social networking service (SNS) research in communication journals from 2006 to 2011. *New media & society*, 17(7), 1007-1024.

ANEXO I. RELACIÓN DEL TRABAJO CON LOS ODS

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.				X
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.			X	
ODS 9. Industria, innovación e infraestructuras.	X			
ODS 10. Reducción de las desigualdades.				X
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

Este Trabajo Final de Grado puede verse relacionado con dos de los Objetivos de Desarrollo Sostenible (ODS). El principal objetivo que se puede relacionar con el trabajo no es otro que el Objetivo 9. Industria, innovación e infraestructuras. El motivo de este razonamiento se debe a la gran innovación planteada con la creación de la herramienta de visualización. Esta, se encuentra muy ligada a la meta 9.5 “*Aumentar la investigación científica y mejorar la capacidad tecnológica de los sectores industriales de todos los países...*” De esta forma, se permite a cualquier usuario de la herramienta creada, realizar análisis de ámbito científico de cualquier término deseado.

Por otro lado, y con un poco de menos relación, el TFG se puede enmarcar en el objetivo 8. Trabajo decente y crecimiento económico. La realización de este proyecto se encuentra ligado a la meta 8.2 “*Lograr niveles más elevados de productividad económica mediante la diversificación, la modernización tecnológica y la innovación...*” Su relación se basa en la capacidad mediante la herramienta de visualización creada, el análisis de carácter económico de las empresas mediante sus sitios web.

Para finalizar, este TFG tiene posee una fuerte relación con los ODS relacionados con el ámbito industrial y empresarial, y más concretamente con la innovación en estos, ya que el objetivo principal es evaluar a las empresas mediante sus sitios web.

Aunque es posible que haya algunas conexiones menores con otros de los Objetivos de Desarrollo Sostenible (ODS), estas serían insignificantes en comparación con las propuestas mencionadas anteriormente, que están directamente relacionadas con los objetivos de este proyecto. Por esta razón, en la tabla se ha marcado el resto de los objetivos con la opción "No procede", ya que no hay una relación tan directa con el trabajo, aunque pueda haber algunos puntos en común.