



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Comunicación Estratégica en el social media marketing:
Propuesta para el sector de running en Instagram

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Primo Aparici, Alba

Tutor/a: Baviera Puig, Tomás

Cotutor/a: Fernández Diego, Marta

CURSO ACADÉMICO: 2021/2022

Agradecimientos

Estar escribiendo estas palabras me emociona, significan el final de un ciclo, que a pesar de su dureza y exigencia voy a echar en falta. Y por supuesto, quiero agradecer a todas aquellas personas que han hecho posible que llegue hasta donde he llegado.

Gracias a mis padres, por darme la maravillosa oportunidad de formarme y apoyarme en todo momento.

Gracias a mis amigos, por levantarme cada vez que yo no me creía capaz de hacerlo. Sin sus ánimos este proyecto hubiera sido mucho menos llevadero.

Gracias a todos los profesionales y docentes que se esfuerzan día a día por la enseñanza y la investigación, sois una pieza clave del puzle en la vida de cualquier estudiante.

Por último, quiero agradecer enormemente a mi tutor, Tomás Baviera, sin él y sin su confianza y apoyo este trabajo no hubiera sido posible. Gracias por darme fuerzas para continuar cada vez que encontrábamos algún que otro problemilla. Pero, sobre todo, gracias por enseñarme a creer en mí y a luchar por superarme cada día. Sin duda este proyecto aporta valor a mi vida.

Índice

AGRADECIMIENTOS	1
ÍNDICE	2
ÍNDICE DE TABLAS	3
ÍNDICE DE FIGURAS	4
1.INTRODUCCIÓN	5
1.1 <i>Resumen</i>	5
1.2 <i>Objetivos</i>	6
1.3 <i>Metodología</i>	6
1.4 <i>Asignaturas relacionadas</i>	7
1.5 <i>Orden documenta</i>	8
2.MARCO TEÓRICO	9
2.1 <i>Comunicación estratégica a través de redes sociales</i>	9
2.1 <i>Engagement en redes sociales</i>	10
2.2 <i>Instagram</i>	12
2.3 <i>Running</i>	16
3.METODOLOGÍA	19
3.1 <i>Descripción del proceso</i>	19
3.2 <i>Identificación de influencers</i>	20
3.3 <i>Extracción de los posts</i>	23
3.4 <i>Limpieza de datos</i>	24
3.5 <i>Variables</i>	25
3.5.1 <i>Variables dependientes</i>	25
3.5.2 <i>Variables de Usuario</i>	25
3.5.3 <i>Variables de Post</i>	27
3.5.4 <i>Variables de Imagen</i>	29
3.6 <i>Modelos Análisis de regresión multivariable</i>	30
3.7 <i>Herramientas utilizadas</i>	31
4.RESULTADOS	34
4.1 <i>Variable dependiente likes</i>	36
4.2 <i>Variables dependiente comentarios</i>	40
4.3 <i>Variable dependiente engagement</i>	44
5.DISCUSIÓN DE RESULTADOS	50
6.PROPOSTA DE ACTUACIÓN.....	54
7.CONCLUSIONES	56
8.BIBLIOGRAFÍA	57
9. ANEXOS	60
<i>Anexo 1. Relación del TFG con los ODS</i>	60
<i>Anexo 2. Lista de los 149 influencers seleccionados</i>	62

Índice de tablas

TABLA 1 PUBLICACIONES CONSULTADAS PARA LA SELECCIÓN DE INFLUENCERS.....	21
TABLA 2 METADATOS DESCARGADOS ASOCIADOS A LOS USUARIOS.	22
TABLA 3 METADATOS DESCARGADOS ASOCIADOS A LOS POSTS.	24
TABLA 4 VARIABLES DEPENDIENTES.....	25
TABLA 5 VARIABLES A NIVEL DE USUARIO.....	26
TABLA 6 VARIABLES A NIVEL DE POST.....	27
TABLA 7 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES PARA EL DATASET COMPLETO.....	34
TABLA 8 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES PARA EL DATASET REDUCIDO.....	35
TABLA 9 MODELOS DE REGRESIÓN MULTIVARIABLE CON DATASET COMPLETO Y VARIABLE DEPENDIENTE LIKES.	37
TABLA 10 MODELOS DE REGRESIÓN MULTIVARIABLE CON DATASET REDUCIDO Y VARIABLE DEPENDIENTE LIKES	39
TABLA 11 MODELOS DE REGRESIÓN MULTIVARIABLE PARA EL DATASET COMPLETO Y VARIABLE DEPENDIENTE COMENTARIOS	41
TABLA 12 MODELOS DE REGRESIÓN MULTIVARIABLE DATASET REDUCIDO Y VARIABLE DEPENDIENTE COMENTARIOS.....	43
TABLA 13 MODELOS DE REGRESIÓN MULTIVARIABLE PARA EL DATASET COMPLETO Y VARIABLE DEPENDIENTE ENGAGEMENT.	45
TABLA 14 MODELO DE REGRESIÓN MULTIVARIABLE DATASET REDUCIDO Y VARIABLE ENGAGEMENT.....	47
TABLA 15 RESUMEN DE LAS VARIABLES SIGNIFICATIVAS EN TODOS LOS MODELOS DE REGRESIÓN MULTIVARIABLE.	51

Índice de figuras

<i>FIGURA 1 ELEMENTOS DE UN PERFIL DE INSTAGRAM.....</i>	<i>14</i>
<i>FIGURA 2 ELEMENTOS DE UN POST DE INSTAGRAM.</i>	<i>15</i>
<i>FIGURA 3 PORCENTAJE DE CORREDORES EN ESPAÑA DE 1996 A 2021.....</i>	<i>16</i>
<i>FIGURA 4 COMPARACIÓN DE MOTIVACIONES PARA CORRER ENTRE CORREDORES PRE-PANDEMIA Y NUEVOS CORREDORES.</i>	<i>17</i>
<i>FIGURA 5 PROCESO METODOLÓGICO DE LA REALIZACIÓN DEL PROYECTO.</i>	<i>19</i>
<i>FIGURA 6 DATOS OBTENIDOS DE LOS INFLUENCERS.....</i>	<i>21</i>
<i>FIGURA 7 PERFILES DE CRISTINA PEDROCHE Y KELLY HOLMES A 26/01/2022.</i>	<i>23</i>
<i>FIGURA 8 ARQUITECTURA CNN ALEXNET.</i>	<i>30</i>
<i>FIGURA 9 ENTORNO DE DESARROLLO VISUAL STUDIO CODE.</i>	<i>31</i>
<i>FIGURA 10 ENTORNO DE TRABAJO RSTUDIO.</i>	<i>33</i>

1.Introducción

1.1 Resumen

En la actualidad, las redes sociales son uno de los mayores canales de comunicación y divulgación existentes. Es una forma muy rentable de llegar a un gran número de personas. Por este motivo, cada vez son más las empresas que suman estas herramientas a sus campañas y estrategias de marketing, consiguiendo así un alcance superior al que pueden conseguir con otros canales de comunicación. Una de las redes sociales favoritas es Instagram. Se trata de una de aplicación móvil que se focaliza en compartir contenido multimedia, como imágenes o videos.

En el marketing de redes sociales, el concepto de *engagement* tiene una gran importancia para evaluar el alcance de una publicación. El *engagement* consiste en la capacidad de interacción con otros usuarios generada por una publicación. En el caso concreto de Instagram, el elemento gráfico juega un papel capital en la generación de *engagement*. Este TFG pretende indagar en el efecto que tienen las imágenes publicadas en Instagram para generar *engagement*. El estudio se circunscribe a la comunidad *runner*.

El objetivo del presente proyecto es elaborar una serie de propuestas para mejorar la comunicación en Instagram de manera orgánica. Las propuestas están dirigidas a las empresas con interés en vender sus productos a la comunidad runner y trabajar con influencers. De esta manera, se pretende que las empresas puedan conseguir un crecimiento sin necesidad de pagos adicionales, es decir, generando interés en los consumidores por el contenido que se comparte.

Para llevar a cabo este proyecto, se han identificado 149 influencers que publican contenido sobre running en Instagram, tanto a nivel nacional como internacional. Mediante técnicas de *big data* se han descargado 24.800 *posts* publicados en el año 2021 por estos usuarios. Tras una limpieza de esta base de datos, el corpus de análisis ha quedado en 23.696 posts. Cada post se ha completado con la información relativa al usuario que lo publicó.

Con el fin de verificar la influencia de las imágenes en el *engagement*, se ha seleccionado una submuestra con aquellos posts que solo tenían una imagen (se han descartado los posts con vídeos y con galerías de imágenes). Esta submuestra asciende a 13.446 posts. A las imágenes de esos posts se han aplicado técnicas de inteligencia artificial basadas en redes neuronales con el fin de extraer información contenida en las fotografías. En concreto, se ha evaluado el número de caras y si la fotografía había sido tomada en el exterior o en el interior.

Con toda esta información, se han elaborado y analizado 21 modelos de regresión multivariable para identificar aquellas variables que más influyen en el *engagement*. Las variables dependientes de los modelos han sido el número de *likes*, el número de

comentarios y la tasa de *engagement*. A partir de los resultados obtenidos se han propuesto 11 pautas de actuación para las empresas relacionadas con el running con el fin de mejorar su marketing de influencers, así como el alcance y crecimiento de su propia cuenta.

Palabras clave: *Engagement; redes sociales; Instagram; comunicación estratégica; running; Redes neuronales.*

1.2 Objetivos

El objetivo general del presente proyecto es conseguir un conjunto de propuestas de actuación en Instagram para que empresas que venden productos a la comunidad runner mejoren su *engagement* de manera orgánica. Se identificarán las posibles estrategias a seguir en base al análisis de 3 variables dependientes y los factores que influyen en ellas, prestando especial atención en los factores provenientes de las propias imágenes.

Este objetivo genérico se descompone en los siguientes objetivos específicos:

- Profundizar en el concepto de *engagement*.
- Identificar influencers runners en la plataforma de Instagram.
- Extraer información de forma directa desde la aplicación de Instagram.
- Extraer características de imágenes recopiladas de Instagram a través de redes neuronales.
- Analizar a través de modelos estadísticos la influencia de las variables predictoras del *engagement*.
- Proponer las pautas a seguir para conseguir un crecimiento orgánico en Instagram, en base a las variables estudiadas.

1.3 Metodología

La metodología básica de este TFG ha sido el análisis de regresión multivariable. Las variables que han intervenido en esta serie de análisis provienen de una base de datos generada específicamente para el presente proyecto. Esta base de datos se compone, en primer lugar, de los datos de Instagram de una selección de influencers en el campo del running y de los posts que publicaron durante un año. Estos datos se obtuvieron directamente de Instagram. En segundo lugar, se obtuvieron dos variables provenientes

de las imágenes publicadas. En este caso, se redujo la muestra inicial a aquellos posts que solo tenían una imagen (un post de Instagram puede tener más contenido multimedia, como se describirá más adelante). La obtención de estas variables requirió el desarrollo de dos herramientas específicas que fueron programadas en Python. Una de estas herramientas consistió en el desarrollo de una red neuronal.

Dado que el proceso seguido es la parte principal del proyecto, se ha incluido un capítulo con toda la descripción detallada de cada fase. El resumen del proceso metodológico seguido se ve gráficamente en la Figura 5.

1.4 Asignaturas relacionadas

Este proyecto se ha podido realizar gracias a haber cursado el Doble Grado de Ingeniería Informática y ADE. La combinación de ambas titulaciones me ha ayudado a aplicar los conocimientos técnicos adquiridos al mundo empresarial.

El inicio de este proyecto fue presentado como TFG de informática el 15 de julio de 2022. En él se trataron las cuestiones más técnicas del presente proyecto, que nos han permitido transformar la información conseguida en conocimiento de interés para las empresas. En el TFG de ADE se analiza todo ese conocimiento, de tal forma que se ha elaborado una propuesta de actuación en Instagram para el aumento del *engagement*.

Realizando este proyecto hemos abordado distintas áreas de estudio. Encontramos una fuerte relación con las siguientes asignaturas:

- Investigación Comercial
- Dirección Comercial
- Introducción a la estadística
- Econometría
- Sistemas de Información estratégicos

Además, ha sido necesaria la formación en distintas áreas no estipuladas en el plan de estudios para llevar a cabo el proyecto. Por ello, se han dedicado más de 2 meses al autoaprendizaje del lenguaje de programación Python, así como a saber aplicarlo en el tratamiento de datos, técnicas de *big data* e incluso inteligencia artificial.

Para abordar el estudio de las 13.446 imágenes que componen uno de nuestros corpus, ha sido necesaria la realización de un curso de Especialización en técnicas de tratamiento de imágenes con inteligencia artificial, proporcionado por la UPV, titulado: “*Deep Learnig* aplicado al tratamiento de señales e imágenes”. Este curso nos ha llevado más de 40 horas de estudio y nos ha permitido plantear la realización del presente proyecto. El curso se realizó en febrero de 2021.

1.5 Orden documenta

A continuación, presentamos el orden documental del proyecto y describimos brevemente de cada uno de los capítulos que componen este documento:

1. **Marco teórico:** En este apartado se da una breve explicación de los elementos más importantes para realizar el trabajo: comunicación estratégica, *engagement* redes sociales, *Instagram* y *running*.
2. **Metodología:** En este apartado se detalla cómo se ha construido el conjunto de datos. Además, se describirá el preprocesado de datos y la elección y creación de variables para los modelos estadísticos.
3. **Resultados:** En este capítulo se presentará toda la información obtenida después de aplicar las metodologías aplicadas en el capítulo anterior.
4. **Discusión de resultados:** Nos centraremos en la discusión y significado de los resultados obtenidos.
5. **Propuesta de actuación:** En este epígrafe detallaremos las estrategias que proponemos para conseguir un mayor *engagement* en base a los datos estudiados.
6. **Conclusiones:** En este apartado se analiza si se han alcanzado los objetivos propuestos y se realiza una valoración del proyecto en su conjunto. Además, se describirán posibles líneas de desarrollo que podrían derivarse de este primer proyecto.
7. **Bibliografía:** Por último, se referencian todas las fuentes bibliográficas utilizadas para la realización del proyecto.

Por último, cabe mencionar que todos los ficheros y bases de datos utilizados a lo largo del trabajo se pueden localizar en el siguiente repositorio creado específicamente para este TFG: https://github.com/albaap007/TFG_ADE.

2.Marco Teórico

2.1 Comunicación estratégica a través de redes sociales

El entorno de la comunicación ha cambiado significativamente durante las dos últimas décadas. Estos cambios han afectado tanto a la teoría como a la práctica de todas las áreas de la comunicación. Entre estos cambios se encuentra un enfoque más integrado de la comunicación de marketing que incluye, entre otras cosas, la difuminación de las líneas de la publicidad y las relaciones públicas. La comunicación estratégica o integrada en marketing ha surgido como resultado de las audiencias cada vez más segmentadas de consumidores dadas las múltiples opciones que existen. Una campaña integrada significa que todos los esfuerzos de comunicación actúan conjuntamente y dan lugar a una sinergia: "el todo es mayor que la suma de sus partes" (Wells, Burnett, & Moriarty, 2003)

La Comunicación Estratégica supone la planificación de los medios de comunicación, tanto los medios de masas como los medios digitales, que utilizan las organizaciones para establecer relaciones y persuadir al público objetivo. La comunicación estratégica estudia cómo una organización puede utilizar herramientas como la publicidad, los medios de comunicación y las relaciones públicas para posicionarse en la mente de sus principales interesados y en la sociedad (Pintado & Sanchez Herrera, 2017).

Schultz, profesor de Comunicaciones Integradas de Marketing (IMC) en la Universidad Northwestern de Chicago y padre fundador de las IMC reconocido en todo el mundo, define las comunicaciones integradas de marketing como "un reajuste de las comunicaciones para verlas como las ve el cliente: como un flujo de información procedente de fuentes indistintas" (Schultz, Tannenbaum, & Lauterborn, 1993). Según Blakeman (2007), los consumidores de hoy en día están tan inundados de mensajes comerciales que ya no es posible confiar en la comunicación unidireccional con el consumidor que ofrecen los medios tradicionales como la televisión, la radio, las revistas y los periódicos. Las comunicaciones de marketing son comunicaciones desarrolladas y transmitidas, normalmente a través de medios de comunicación pagados, a los consumidores con el fin de vender bienes y servicios. Para que las comunicaciones de marketing sean eficaces en el siglo XXI, la publicidad y las relaciones públicas deben ser interactivas mediante una comunicación bidireccional en la que el consumidor pueda realmente relacionarse con la marca.

"El IMC, también conocido como marketing relacional, se esfuerza por construir una relación a largo plazo entre el comprador y el vendedor haciendo que el individuo al que se dirige participe en un intercambio de información interactivo o bidireccional" (Blakeman, 2007, pág. 3).

De ahí nace la importancia de las redes sociales en la comunicación estratégica, ya que son capaces de conseguir esa interacción entre marca-consumidor. Además, este tipo de plataformas permite publicitarse de manera mucho más barata y de una manera

mucho más real, cosa que en los tiempos en los que estamos tiene real importancia en la sociedad.

En muchos sentidos, las redes sociales han facilitado a los consumidores la tarea de exigir responsabilidades a organizaciones, personajes públicos y grandes instituciones (Lloyd, 2016). Los usuarios pueden encontrar y revelar fácilmente información sobre un acontecimiento anterior que implique a una organización, ya sea ventajoso o perjudicial para la marca. Los usuarios también pueden proporcionar información pública instantánea expresando sus opiniones a través de las redes sociales. Además, las redes sociales han hecho que muchas organizaciones tengan que controlar su marca y presentar un mensaje coherente en todas las plataformas. El público puede generar información que puede ser perjudicial para la reputación de una marca.

Esta transparencia en lo que a información circulante en redes sociales se refiere, proporciona una mayor confianza de los consumidores en las marcas, por lo que surge la necesidad para las empresas de implicarse en estos medios de comunicación, preparando buenas estrategias para captar y fidelizar clientes.

Así mismo, las redes sociales se convierten en el canal de comunicación más directo con los consumidores, lo que tiene una gran ventaja para las marcas: cuanta más comunicación exista en estas plataformas más información sobre los consumidores podrá ser recopilada, para posteriormente ser analizada y entender el comportamiento de los consumidores. Además, nos permite también la recopilación de información de competidores.

Comprender el comportamiento de los consumidores y convertir esos datos en información práctica nos ayuda a crear una relación de comprensión y confianza con los consumidores, lo que en última instancia puede traducirse en un aumento de la fidelidad a la marca. Si el consumidor siente que se entienden sus deseos a través de como se comunica la marca será más fácil la captación de clientes.

Sin embargo, no solo podemos entender el comportamiento de los consumidores a través de estas herramientas, sino que además podemos influir en su comportamiento a través de la información compartida en estas plataformas. Por tanto, las redes sociales se convierten en un canal de comunicación estratégica de doble filo.

2.1 *Engagement* en redes sociales

La aparición de las plataformas de medios sociales ha modificado drásticamente el papel de los clientes de las empresas, que han pasado de ser observadores pasivos de los contenidos a participantes activos. Hoy en día, los consumidores son co-productores y co-creadores de contenido a través de sus interacciones y comportamientos en redes sociales. Este comportamiento incluye la creación, la contribución o el consumo de contenido relacionado con la marca por parte de los clientes dentro de una red social. El grado de compromiso difiere de los tipos simples de compromiso (por ejemplo, dar al botón "me gusta" en una publicación en Instagram) a los tipos más altos de compromiso

de los clientes en acciones como la publicación de reseñas (Dolan, Conduit, Frethey-Bentham, Fahy, & Goodman, 2019).

La creciente prevalencia de las redes sociales ha hecho que tanto docentes como profesionales se centren en el concepto de *engagement* (en castellano equivaldría a *compromiso*) en las plataformas de redes sociales. El compromiso de los clientes se ha estudiado en muchos campos, como la psicología, la educación, la gestión, el marketing y los sistemas de información.

En el contexto del marketing digital, Van Doorn (2010) define el *engagement* como "las manifestaciones de comportamiento de un cliente que tienen un enfoque de marca o de empresa, más allá de la compra." Existen varios enfoques del compromiso de los consumidores, entre los que se encuentran las ofertas de productos o servicios, las actividades y los eventos y los medios de comunicación. Algunos reconocieron las propiedades del *engagement*, como la intensidad, y examinaron sus diversos aspectos y antecedentes. Según Dolan et al. (2019), existen seis tipos de comportamientos de compromiso que se pueden encontrar en redes sociales: crear, contribuir, destruir (conocidos como comportamientos de compromiso activo), consumir, dormir y desapegarse (conocidos como formas pasivas y/o más individualizadas de compromiso). El *engagement* es interactivo y dependiente del contexto y se entiende a través del análisis de cada uno de los factores influyentes.

El *engagement* es pues, el conjunto de interacciones y actividad en los contenidos publicados en redes sociales, lo que incluye los "me gusta", los comentarios, las reacciones, etc. Este concepto contribuye a la vinculación de los usuarios con un contenido concreto y tiene un efecto importante en la difusión de información dentro de una red social. Sin embargo, las interacciones exactas que determinan lo que contribuye al compromiso difieren entre las plataformas y tipos de contenido.

Las características interactivas del comportamiento de compromiso pueden conducir a diferentes niveles de intensidad. En un estudio más reciente sobre los niveles de intensidad del *engagement*, Dolan et al. (2019) sugieren dos tipologías que engloban los seis grupos de comportamiento (referidos anteriormente). Según argumentan, puede ser pasivo (bajo) o activo (alto) y también positivo o negativo. El compromiso pasivo se define a través del comportamiento de un miembro que visualiza todo el contenido publicado en redes sociales y aprovecha al máximo los beneficios accesibles, mientras que no participa en ninguna actividad de la plataforma. Por el contrario, el compromiso activo se determina a través del comportamiento de los miembros altamente interesados en comprometerse en una red social participando en actividades, creando mensajes, difundiendo información y proporcionando apoyo emocional a los demás.

Existen diferentes métricas para medir la intensidad de la pasividad o la activación del compromiso con los medios sociales. Por ejemplo, Alhabash et al. (2013) describen los "me gusta" (como "respuesta afectiva") y los comentarios como ("deliberación activa y pública"), como comportamientos activos en los medios sociales, mientras que la lectura de contenidos y los clics son ejemplos de comportamientos de compromiso pasivo.

En concreto, en este proyecto vamos a evaluar el *engagement* desde 3 puntos de vista:

- *Likes*: como un primer nivel de compromiso, en este nivel al usuario le ha llamado la atención el contenido compartido y por tanto se ha producido un primer nivel de interacción.
- Comentarios: como segundo nivel de compromiso, donde el usuario va más allá e invierte su tiempo en comentar el contenido compartido.
- Ratio *Engagement*: La tasa de *engagement* es una métrica utilizada para medir el nivel de compromiso generado a partir de un contenido creado o una campaña de la marca. En otras palabras, la tasa de compromiso se refiere al nivel de interacción con los seguidores que se genera a partir del contenido creado por un usuario en relación con la base de seguidores.

$$Engagement = \frac{Total\ Engagement}{Número\ de\ followers}$$

- Total *engagement*: se refiere al número de interacciones (cuya medición depende de la plataforma, en este caso estará compuesto por los *likes* y los comentarios); y
- El número de *followers* se refiere a la cantidad total de personas que siguen la cuenta.

El índice de compromiso proporciona una representación más precisa del rendimiento de los contenidos que si se limitan a observar medidas absolutas individuales como el número de "me gusta", comentarios, acciones, etc. Se trata de una métrica más completa.

A continuación, vamos a describir la plataforma que va a ser el objeto de este proyecto: la red social Instagram.

2.2 Instagram

Instagram es una aplicación social de fotografía con funciones de toma de fotografías, filtrado para su mejora y compartición. También puede considerarse un nuevo tipo de medio social basado en la interacción con imágenes, ya que sus funciones ayudan a las personas a crear redes sociales basadas en imágenes (Jang, Han, Shith, & Lee, 2015). Esta aplicación móvil se publicó oficialmente en Apple Appstore en octubre de 2010. Su número de usuarios creció rápidamente tras el lanzamiento, con más de 100.000 usuarios registrados en solo una semana. Posteriormente, Instagram se convirtió en el centro de atención del público el 10 de abril de 2012. Ese día, Facebook compró esta empresa, que sólo tenía 551 días y 13 empleados, por mil millones de dólares en efectivo y acciones. En 2018, el número de usuarios de Instagram ya había superado los mil millones (Constine, 2018).

Con varios tipos de mapas y vídeos, no solo es una herramienta para las redes sociales entre amigos, sino también una importante plataforma para que los Influencers trabajen.

Mientras tanto, en el mundo de los negocios, también es un canal para que las empresas globales atraigan el tráfico web e interactúen con los consumidores. Como plataforma social, el éxito de Instagram para destacar en la feroz competencia se debe principalmente a su claro posicionamiento de producto y a sus destacadas ventajas en comparación con otras plataformas de medios sociales.

En comparación con redes sociales tradicionales, que tienen una serie de potentes funciones, como el chat instantáneo, la publicación de estados personales, la publicación de fotos, etc., Instagram abandona las funciones integrales y se centra únicamente en la producción de imágenes (Smith, 2014). Para aprovechar al máximo la comodidad de los *smartphones* y simplificar al máximo la función de su aplicación, Instagram solo conserva algunas funciones básicas, como la captura, la edición de imagen, el intercambio de fotos y los comentarios. Esta estrategia hace que sea fácil de aprender y manejar (Manovich, 2016). Este claro posicionamiento del producto ayuda a que los usuarios a los que les gusta capturar y compartir fotos utilicen esta aplicación, lo que convierte la plataforma en una red social fotográfica única entre las numerosas plataformas de medios sociales. Al mismo tiempo, Instagram también se asocia con otras aplicaciones sociales, como Facebook, Twitter, etc., para ampliar mejor la actividad social de los usuarios al permitirles compartir fotos con sus amigos en estas plataformas con un solo clic. Por lo tanto, la fuerte profesionalidad y el claro posicionamiento del producto permiten a Instagram lograr una diferenciación del producto y una fuerte influencia de la marca en poco tiempo, lo que le permite destacar en la intensa competencia del mercado.

Instagram ha ido actualizándose y adaptándose a las necesidades, por ello han sido habituales la incorporación de nuevas funcionalidades constantemente. A continuación, se detallan las más importantes:

- 2015: fue cuando se introdujo la opción de crear y compartir publicidad.
- 2016: fue incorporada la función de *Stories*, que consiste en compartir contenido que solo será visible durante 24 horas. A este contenido se le pueden incluir un montón de elementos interactivos como encuestas, cuestionarios, preguntas, música, gifs...
- 2018: aparece la función Instagram TV, lo que permite hacer transmisiones en directo además de poder añadir publicaciones con videos de larga duración. Por otro lado, nació la opción de “tiendas” en Instagram lo que permite comprar directamente desde el perfil de una marca.
- 2020: se incorpora la función *Reels*, lo que permite compartir videos de corta duración (menos de un minuto) adaptados a las pantallas de los móviles.

Esta última función es la específica de TikTok, principal competidor de Instagram, lo que nos indica que Instagram es capaz de coger aquellas funcionalidades que se acogen de manera positiva en sus competidores e integrarlas en su aplicación.

La última novedad de Instagram, que se encuentra ahora en fase experimental en Estados Unidos, es la oferta de suscripciones a creadores de contenido de renombre. A través de esta suscripción los usuarios podrán acceder a contenido exclusivo no disponible para los demás usuarios. Por lo tanto, ya no se podrá consumir de manera gratuita todo el contenido de la aplicación.

Para profundizar y tener una visión más completa de la App se van a describir los principales elementos, divididos en dos grupos: elementos de usuario y elementos del post.

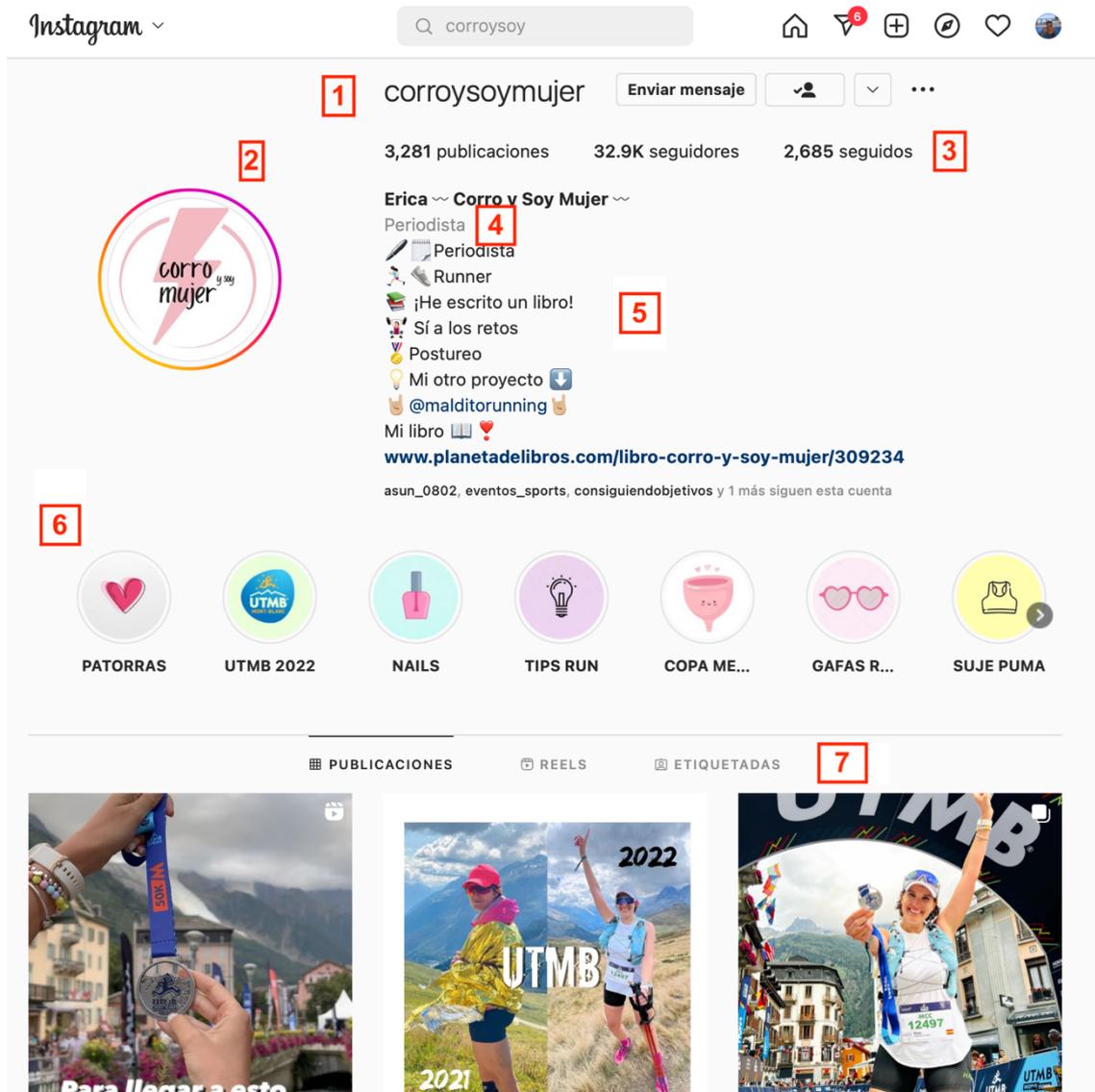


Figura 1 Elementos de un perfil de Instagram

Fuente: Instagram (2022a).

En la Figura 1 podemos visualizar los componentes de un perfil de Instagram, desde la interfaz preparada para ordenador. En primer lugar (1), tenemos el *Nick* o nombre de usuario, que es el nombre con el que se identifica al perfil. Después, encontramos la foto de perfil (2), que normalmente, identifica de manera visual al usuario de la cuenta. Además, también encontramos información sobre el usuario (5), más conocido como biografía, donde se detallan aspectos interesantes sobre la cuenta. Por ejemplo, en este caso nos está indicando a qué se dedica el usuario, sus hobbies y sus principales proyectos. Además, se puede añadir una de las categorías predeterminadas que te ofrece Instagram (4), lo que categoriza el perfil en un sector concreto, en este caso, periodista.

Por otro lado, encontramos los datos propios del perfil (3), compuestos por el número de publicaciones totales de la cuenta, el número de seguidores y el número de cuentas que sigue el usuario. Más abajo encontramos las historias destacadas (6), que recopilan historias subidas por el usuario ordenadas a su gusto. Al destacarlas, se permite la visualización de estas historias pasadas las 24 horas desde su publicación.

En último lugar, encontramos los *posts* publicados por el usuario (7) en formato cuadrado y de cualquier tipo: imágenes, videos, reels...

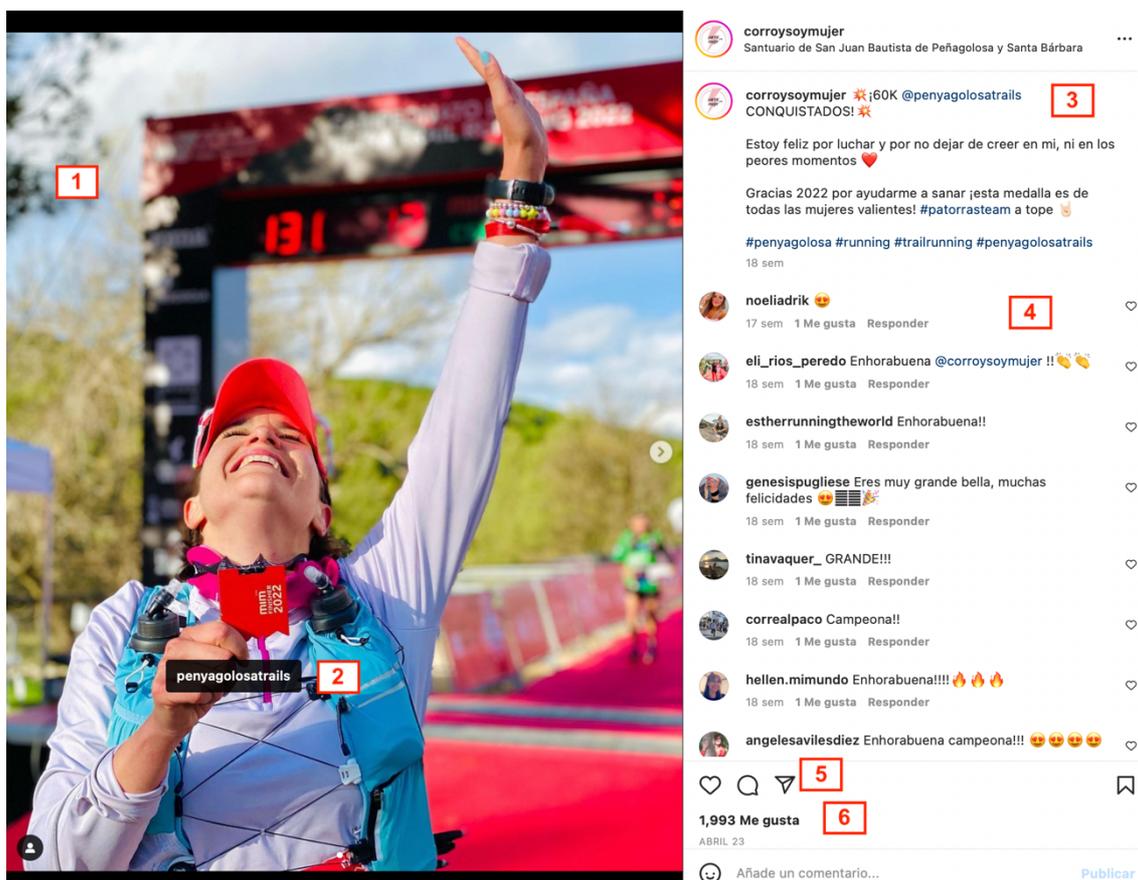


Figura 2 Elementos de un post de Instagram.

Fuente: Instagram (2022b)

En la Figura 2, observamos los elementos que componen una publicación en Instagram. Empezando por el elemento más importante, se trata del contenido multimedia de la publicación (1), que en este caso es una imagen. También podría ser un video o un grupo de varios elementos multimedia. Además, sobre la imagen, encontramos los perfiles etiquetados (2) en la publicación. A la derecha, encontramos el texto o “pie de foto” (3) que complementa el contenido multimedia. En este elemento se incluyen los *hashtags* que utiliza el usuario para posicionar la publicación.

En último lugar encontramos los elementos más relacionados con la interacción entre usuarios. Todas las funciones que se pueden realizar con la publicación (5) son, dar “me gusta” con un *click* en el icono corazón, comentar dando *click* al icono del bocado, compartir por mensaje directo dando *click* sobre el icono del avión de papel y guardar la publicación presionando sobre el icono marcapáginas. Arriba de estas funciones

podemos encontrar los comentarios (4) que ha recibido la publicación y abajo el recuento de *likes* (6) que ha recibido.

Una vez presentada la herramienta de Instagram, vamos a introducir el sector en el que se va a centrar nuestro estudio del *engagement*. Se trata del sector del running.

2.3 Running

El running, es una actividad deportiva que se puede realizar individualmente o de forma colectiva. Se trata de una actividad en la que es muy fácil iniciarse y en la que los beneficios se consiguen rápidamente.

En España la práctica de esta disciplina deportiva cada vez es mayor, como se observa en la Figura 3. Aunque se encuentran fluctuaciones, a lo largo de tiempo la tendencia es creciente. Además, el incremento es mayor desde que a causa de la pandemia COVID-19 se tuvieron que cerrar gimnasios y anular actividades deportivas colectivas.

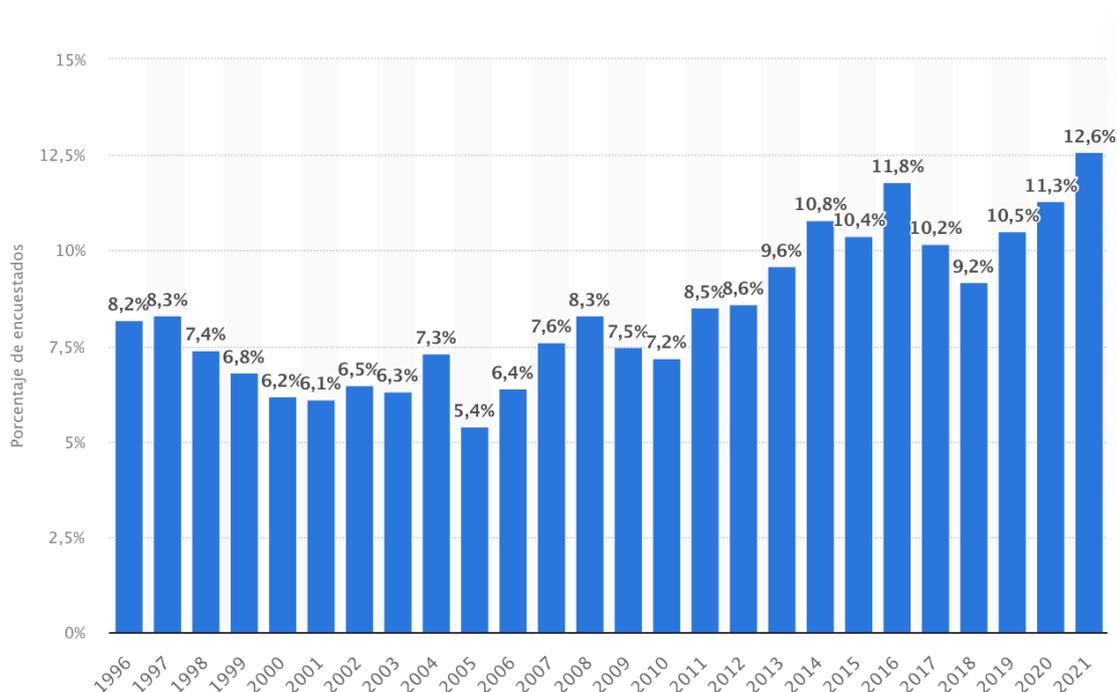


Figura 3 Porcentaje de corredores en España de 1996 a 2021.

Fuente: Statista Research Department (2022)

Este incremento no se refleja solo a nivel nacional. Un informe del organismo deportivo World Athletics (2021), que analizó las tendencias del running recreativo, descubrió que alrededor del 13% de los corredores encuestados se habían iniciado en el último año, y más de una quinta parte de todos los corredores afirmaron haber corrido más durante la pandemia que antes.

En el mismo informe, se les hizo varias preguntas relevantes a los encuestados. Una de ellas, preguntaba sobre la motivación para empezar a correr. En la Figura 4, es evidente

el cambio entre los motivos por los que empezaron a correr los runners antes de la pandemia y los que empezaron a correr después. El 72,78% de los nuevos corredores, se vio motivado a empezar por salud física frente a un 11,59% que dice estar motivado para relacionarse socialmente. Estos datos son relevantes ya que una persona que inicia esta actividad por causas de salud es más probable que se mantenga a largo plazo que una persona que empieza a correr para ampliar su círculo social. Si extrapolamos esto a nivel empresarial, un consumidor constante podrá beneficiar más a la empresa que un consumidor esporádico.

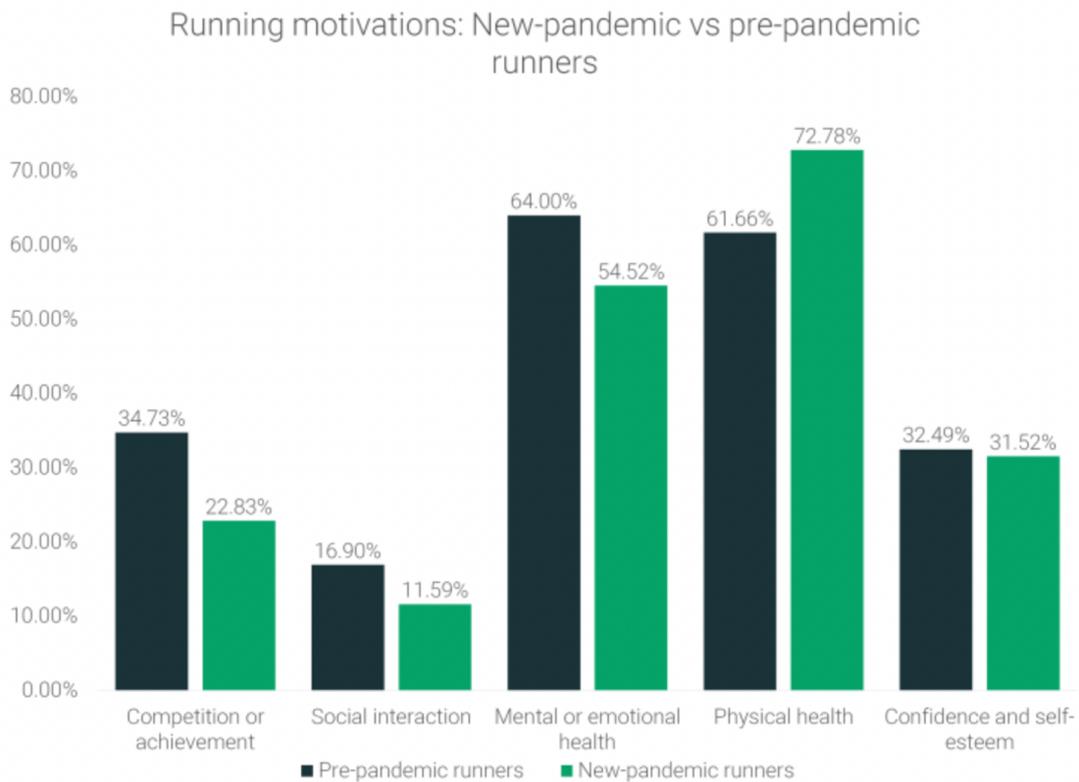


Figura 4 Comparación de motivaciones para correr entre corredores pre-pandemia y nuevos corredores.

Fuente: Rizzo (2021)

Otra pregunta interesante que se le hizo a los encuestados fue su preferencia entre participar en carreras tradicionales, o hacerlo en carreras virtuales. Las carreras virtuales son una nueva modalidad de competición en la que los inscritos corren la distancia pautada y a través de una aplicación móvil comparte tanto el recorrido como el tiempo realizado. Este tipo de carreras evita las aglomeraciones en eventos deportivos, ya que cada participante puede correr la distancia pautada en cualquier momento y de forma individual.

A pesar de que siguen siendo más populares las carreras tradicionales, es evidente el grado de aumento de la preferencia en eventos virtuales. El 31,58% de los corredores postpandemia prefieren correr virtualmente, lo que los hace un 115,37% más propensos a correr una carrera virtual que los corredores que iniciaron la actividad antes de la pandemia.

Estos datos son un claro ejemplo de cómo esta comunidad social se ha adaptado a las circunstancias consecuencia de una pandemia, ya que hace varios años el concepto de carrera virtual ni siquiera existía.

Para las empresas, esta evolución, les supone tener que ajustar sus estrategias de comunicación para poder adaptarse al máximo a toda la comunidad runner, buscando siempre cubrir todas las necesidades de un runner, ya sea tradicional o postpandémico.

En este capítulo se ha introducido el concepto de *engagement*, se ha presentado la red social Instagram y se ha descrito brevemente el sector de running. En el siguiente pasamos a explicar la metodología que se ha seguido en el TFG, con el fin de realizar propuestas de actuación para las empresas a partir de un corpus de datos recopilados de Instagram.

3. Metodología

3.1 Descripción del proceso

El presente trabajo ha seguido el proceso metodológico que se muestra en la, compuesto por 5 fases.

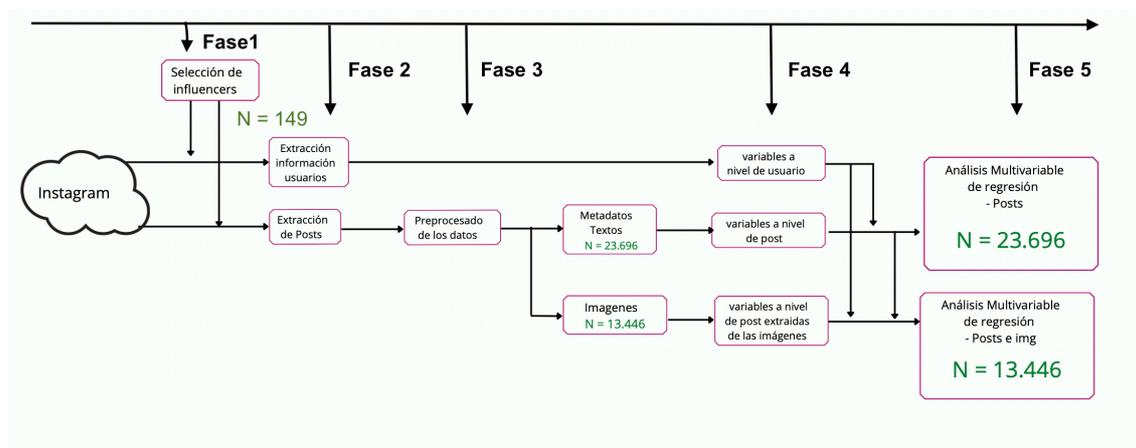


Figura 5 Proceso metodológico de la realización del proyecto.

Fuente: Elaboración propia.

En una primera fase, hicimos la selección de influencers que se ajustaban a nuestro objeto de estudio, recopilando 149 perfiles. Una vez seleccionados, en la segunda fase, procedimos a la extracción de información a través de varios ficheros de código. Por un lado, se hizo una extracción con los datos relacionados con el usuario y, por otro lado, una extracción del contenido publicado por los perfiles y los metadatos asociados (lo que incluía también la obtención tanto de imágenes como de videos). Esta extracción supuso la recopilación de 24.800 publicaciones.

En una tercera fase, nos encargamos de limpiar los datos obtenidos y prepararlos para su uso en modelos de regresión quedando así un corpus de 23.696 *posts*. A partir de este corpus, construimos una submuestra que solo contiene aquellos *posts* de tipo imagen que se ha utilizado para la extracción de características de las imágenes a través de redes neuronales, una técnica de inteligencia artificial muy potente para el tratamiento de imágenes. Esta diferenciación, ha sido necesaria porque algunas de las variables que vamos a generar son específicas a los posts de tipo imagen, por lo que no aplican a posts tipo video o grupo de imágenes.

Una vez los datos estaban preprocesados y los corpus definidos, pasamos a la fase 4, donde generamos nuestras variables, y construimos nuevas a partir de los datos obtenidos. Las variables con las que trabajamos se dividen en 3 grupos:

- 1) Variables a nivel de usuario: son las variables que nos proporcionan información asociada al perfil.

- 2) Variables a nivel de *post*: son las variables específicas de un *post*. Por ejemplo, el número de *likes* o de comentarios que ha tenido un *post*.
- 3) Variables a nivel de *post* extraídas de las imágenes: estas variables también van asociadas y son específicas de un *post* pero, dada la naturaleza de su origen se decide tratarlas en un grupo aparte. Además, el uso de estas variables solo es posible cuando el tipo de publicación es una imagen ya que estas variables han sido generadas a través de la submuestra de los 13.446 *posts* de tipo imagen. Estas variables son, por ejemplo, el número de rostros que aparecen en una imagen o si se trata de una fotografía tomada al aire libre o no.

En la última fase, y con las variables ya definidas se hicieron 21 modelos de análisis jerárquico de regresión multivariantes utilizando ambos corpus. Con el uso del primer corpus (23.696 *posts*) se utilizaron las variables a nivel de usuario y a nivel de *post* y se estudiaron 3 variables dependientes: *likes*, comentarios y *engagement*. Con el segundo corpus (13.446 *post*), se incluyeron, además de todas las variables utilizadas en los primeros modelos, las variables extraídas de las propias imágenes y se volvieron a estudiar las 3 variables dependientes.

A continuación, se detallarán todas las fases mencionadas y las herramientas utilizadas.

3.2 Identificación de influencers

En esta primera fase, se seleccionaron los *influencers* que iban a formar parte del estudio. La primera decisión que se tomó fue no limitar la búsqueda de *influencers* solo al territorio nacional. Esta decisión se respalda por la cantidad de datos que necesitamos obtener para entrenar el modelo de redes neuronales, si bien se han diferenciado para enriquecer el conjunto de datos obtenido y hacer un posterior análisis de esta característica.

Una vez tomada esta decisión, se inició la búsqueda, a través de internet, de publicaciones hechas en redes sociales o en páginas web especializadas en *running*, donde se listarán los influencers de *running* más importantes o reconocidos tanto a nivel mundial como a nivel nacional. También se incluyeron aquellas publicaciones ofrecidas por páginas web especializadas en búsqueda de influencers, como inBeat (2021). Con estos criterios conseguimos un total de 11 publicaciones que se enumeran en la Tabla 1.

Después de revisar las distintas publicaciones, se listaron todos los usuarios de Instagram en un fichero de texto, obteniendo un total de 248 perfiles. El primer paso fue eliminar los perfiles duplicados, ya que en varias publicaciones se mencionaban los mismos perfiles. El número de usuarios repetidos ascendió a 31, dejando así un total de 217 perfiles para la extracción de datos.

	Fuente	Título	Ámbito
1	(Runnea, 2021)	Los 11 bloggers de running más influyentes de España	Nacional
2	(Roldán, 2018)	Las instagramers más influyentes del universo running en España	Nacional
3	(Pastor, s.f.)	8 cuentas de runners en Instagram que deberías estar siguiendo	Nacional
4	(Ungría, 2018)	Los runners españoles más influyentes en Instagram	Nacional
5	(Affde, 2021)	Principales influencers del running en las redes sociales	Internacional
6	(El comercio, 2018)	Los 6 runners más influyentes en Instagram	Internacional
7	(InBeat, 2021)	Top 10 Runners Instagram Influencers In Spain In 2021	Nacional
8	(Godfrey, s.f.)	10 global running influencers you should know	Internacional
9	(Feedspot Media Database Team, 2020)	Top 150 Running Instagram Influencers most followed	Internacional
10	(Libertad Digital, 2016)	Los corredores más influyentes del running en España	Nacional

Tabla 1 Publicaciones consultadas para la selección de influencers

Fuente: Elaboración propia.

Para seguir afinando nuestra selección de perfiles necesitábamos acceder a la información de dichos usuarios, por lo que se procedió a la extracción de información de los usuarios. En la Figura 6 se puede ver una pequeña muestra de la información de usuarios después de su descarga en un fichero Excel.

num	username	full_name	business	verified	bio	category	followers	followers	posts	likes	reels
1	heathercruz	Heather Schultz	0	0	Running through life 🏃‍♀️ Coach 🏃‍♀️ Paych Teacher 🏃‍♀️ Track Shack 🏃‍♀️ 26.2 x 40 Sub 3hrs x 13 Boston x 13.1 🏃‍♀️ 26.2 @ 2:54 🏃‍♀️ 13.1 @ 1:21		92464	521	2266	3	1
2	lucy_bartholomew	Lucy Bartholomew	0	1	Fresh air seeker 🌿 Adventure lover 🌿 Plant nurser 🌿		105003	96	5200	23	1
3	sarah33	Sara Hall	1	1	American record holder- Half marathon 🏃‍♀️ Soccer Mom to 4 from 🏃‍♀️ Wife to @jyrahall	Creators & Celebrities	167841	432	1256	1	1
4	stephrothstein	Stephanie Rothstein Bruce	1	1	Professional runner for HokaOneOne/NAZ Elite, PRO Compression, mom, Celiac. Co-conspirator for Picky Bars, coach of Running w/ Bruce!	Creators & Celebrities	96182	762	2354	8	1
5	trackclubbabe	KIM	0	0	RLN 🏃‍♀️ [26.2-6:08@3:11 🏃‍♀️ 13.1-1:28] here to help ya get FASTER & LOVE running 🏃‍♀️ trackclubbabe@gmail.com track your miles w/ @runkeeper app 🏃‍♀️		128816	283	2064	3	1
6	ian.morgan	Ian Morgan Ultra Runner	0	0	🏃‍♀️ Ultra Runner 🏃‍♀️ M. mujer @hangbar Based in Sitges/Spain 🏃‍♀️ Traveler 🏃‍♀️ Coffee 🏃‍♀️ @mrtajpuffist @ultraico @squeerworld		121772	2330	2904	20	1
7	foodfitnessflora	Flora Beverley	0	1	Fitness Food Sustainability 🌱 Collab: Hannah@themodel.co.uk 🏃‍♀️ @backtothetop @beebow 🏃‍♀️ florabeverley@gmail.com @theforwardlab writer 🏃‍♀️ YOUTUBE/BLOG 🏃‍♀️		114672	1231	4144	6	1
8	desiree.linden	desiree linden	1	1	201= Boston Marathon Champion. 2x US Olympian. Coffee aficionado @lovelysty. Whiskey connoisseur. Music junkie. Book nerd. Travel enthusiast.	Creators & Celebrities	181959	1155	496	7	1
9	runrx	RunRx Learn to run pain free	1	0	RunRxStrong Membership Self-Strength-injury prevention. Learn how to run and stay pain free! 🏃‍♀️ @runrx_runrx	Personal Goods & General Merchandise Stores	211008	100	4736	103	1
10	kangpoucher	Kari Goucher	1	1	Mom, wife to @adamgoucher, obsessed runner. "Be the change." 2 x Olympian, World Championship 🏃‍♀️ Athlete-Advisor @cisselle, @banrunning	Creators & Celebrities	188103	502	1875	4	1
11	jenrifalconer	Jenni Falconer	1	1	@smoothradio Breakfast 🏃‍♀️ Founder podcast "RunPod" 🏃‍♀️ Co-Founder @holthealth Runner/Golfer/Mom Enquiries: tap 🏃‍♀️ https://bit.ly/57yuluf	Creators & Celebrities	247719	6263	5914	53	1
12	tommy_riv	tommy_riv 🏃‍♀️	1	1	keep calm & rage on @reastrix	Creators & Celebrities	227573	1077	1058	17	1
13	courtneyflawalter	Courtney D'awalter	1	1	Ultra runner with a love for sunshine, long inners, and candy. 🏃‍♀️ @balmonrunning 🏃‍♀️ @balwinduction 🏃‍♀️ @bioclicakes	Creators & Celebrities	327598	931	198	3	1
14	shahreflanagan	Shelane Flanagan	1	1	Mom 🏃‍♀️ Nike Coach 🏃‍♀️ BTCLIOlympian/Olympic Silver Medalist NYC Marathon champ/American record holder/NYTimes best selling author X3 Run Fast Eat Slow	Creators & Celebrities	420429	407	1173	10	1
15	compyomujer	Erica -- Como y Soy Mujer --	0	0	🏃‍♀️ Periodista 🏃‍♀️ 🏃‍♀️ Runner 🏃‍♀️ He escrito un libro! 🏃‍♀️ 5 a los retos. 🏃‍♀️ Posture 🏃‍♀️ Mi otro proyecto 🏃‍♀️ @theadrunning 🏃‍♀️ Mi libro 🏃‍♀️		32343	2669	3167	6	1
16	pitufollow	SERGIO TURULL	0	1	Addas Athlete 🏃‍♀️ Foodspring -15%: pitufollowFSG 🏃‍♀️ sergiopitufollow@gmail.com 🏃‍♀️ Autor #Compoquecamineacansa		92209	963	3360	25	1
17	marlagagn	Marta Guerrero	0	0	🏃‍♀️ 2112 x record LOBOS 🏃‍♀️ road or 6 🏃‍♀️ ANCS Fitnesswear 🏃‍♀️ Spais 🏃‍♀️ @otora_baia 🏃‍♀️ @compespain 🏃‍♀️ @baonje.es 🏃‍♀️ marlagagn@gmail.com		38491	3501	1151	3	1
18	delafots	Biel Rábals	1	0	🏃‍♀️ Sports Mix 🏃‍♀️ @salomon.spain 🏃‍♀️ 🏃‍♀️ Orientierung lover 🏃‍♀️ 🏃‍♀️ Camara-runner 🏃‍♀️ 🏃‍♀️ shorturl.at/vfVQ	General Interest	20143	887	1400	68	1

Figura 6 Datos obtenidos de los influencers.

Fuente: Elaboración propia.

En la Tabla 2 se describen los datos recuperados después de la extracción de información de los 217 influencers.

Nombre	Descripción
username	El nombre de usuario.
full_name	El nombre completo del usuario.
nPost	El número de posts total visibles que el usuario ha publicado.
Bio	Descripción de la cuenta.
Followers	El número de perfiles que siguen al usuario.
Followees	El número de perfiles que el usuario sigue.
igtvCount	El número de vídeos subidos a igtv
business	Devuelve '1' si la cuenta es un perfil de negocio y '0' en caso contrario.
category	La categoría de negocio si se trata de un perfil de negocio.
verified	Devuelve '1' si la cuenta es un perfil verificado y '0' en caso contrario.

Tabla 2 Metadatos descargados asociados a los usuarios.

Fuente: Elaboración propia.

Al realizarse la descarga, no se encontraron 6 perfiles. Después de verificar que los perfiles no existían, seguimos adelante con una lista con 211 perfiles.

Para conseguir un dataset lo más consistente posible se estableció que, del listado inicial, nos interesaban aquellos perfiles que tuvieran más de 10.000 seguidores, ya que son los perfiles que consiguen mayor *engagement* (Sicilia, Palazon, Lopez, & Lopez, 2021). Filtrando los datos, seleccionando solo los perfiles con más de 10.000 seguidores, nos deja un total de 165 perfiles al excluir 46 usuarios por no cumplir dicha condición.

Por último, fue necesario inspeccionar los 165 perfiles uno a uno desde la App de Instagram para asegurarnos de que los usuarios compartían contenido sobre running frecuentemente. Este paso se ha llevado a cabo porque en los listados previos había perfiles como el de Cristina Pedroche o Patricia Montero. Estos perfiles no comparten contenido sobre running ya que su actividad principal es otra.

Por ejemplo, en la Figura 7 (izquierda) se puede observar como el perfil de Cristina Pedroche, más conocida por los atuendos utilizados para despedir el año, comparte publicaciones aleatorias sobre su vida. No se trata de un perfil enfocado a compartir publicaciones sobre running o temáticas relacionadas. Sin embargo, en la Figura 7 (derecha) se muestra el perfil de Kelly Holmes en el que, sí se pueden localizar diferentes publicaciones de ella corriendo, lo que nos confirma que se trata de una cuenta donde efectivamente se puede consumir contenido relacionado con el running.

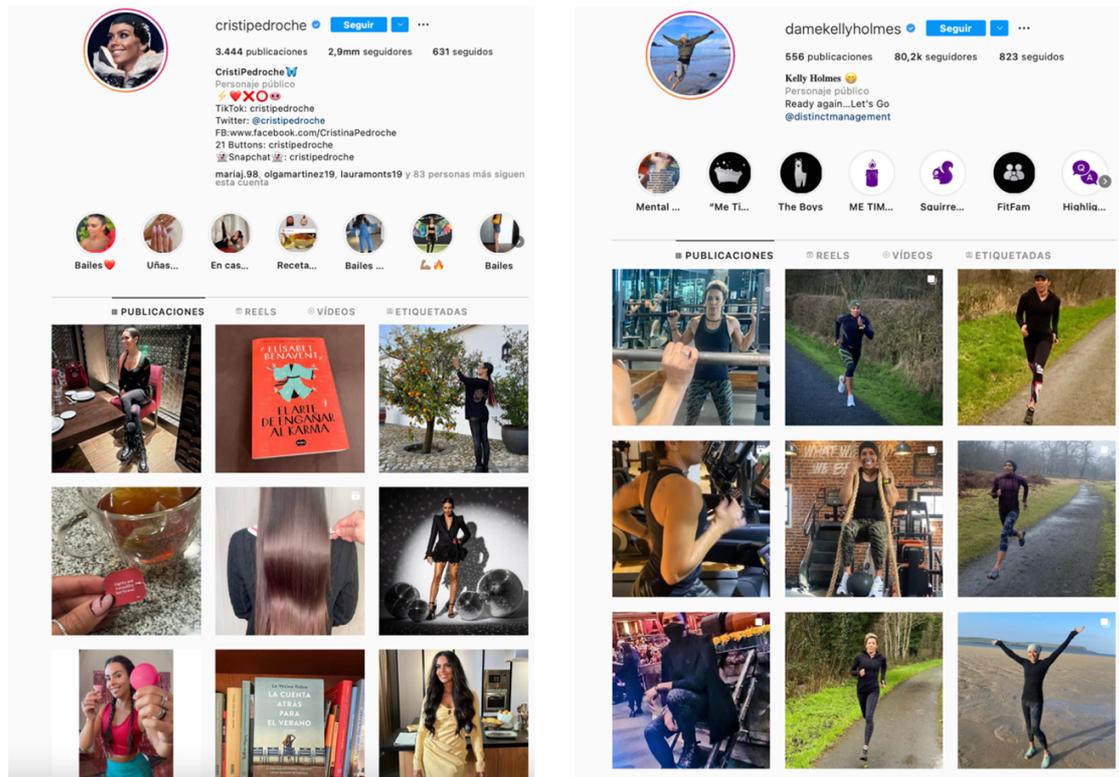


Figura 7 Perfiles de Cristina Pedroche y Kelly Holmes a 26/01/2022.
 Fuente: Elaboración propia.

Al comprobar los 149 influencers se han identificado 17 perfiles en los que no se comparte contenido relacionado con el running en sus 30 posts más recientes, por tanto, no se ha contado con ellos para la creación del dataset. Resumiendo, la lista final de influencers consta de 149 perfiles, que se puede encontrar en el Anexo 9.2: *Anexo 2. Lista de los 149 influencers seleccionados* del presente documento.

3.3 Extracción de los posts

El siguiente paso, con los influencers ya identificados, consistió en el proceso de descarga de los *posts* de cada usuario.

En primer lugar, se tomó la decisión de descargar los posts de un año entero, del 1 de enero de 2021 al 31 de diciembre de 2021. Con el objetivo de optimizar el proceso de descarga y automatizarlo lo máximo posible, construimos un sencillo programa en Python. Conseguimos una descarga de 24.787 *posts*. Los datos obtenidos después de la descarga han sido recopilados y descritos en la Tabla 3.

Metadato	Descripción
Post	El nombre del archivo multimedia que descargamos, en formato UTC.
Shortcode	id del post
User	Alias del usuario que ha compartido la publicación.
Fecha	Día, mes y año en la que se publicó el post.
Hora	Hora exacta con minutos y segundos de la publicación.
Tipo	Especifica si el tipo de contenido multimedia incluido en el post es imagen, vídeo o sidecar.
Mediacount	Número de imágenes en un sidecar.
Likes	Número de “me gusta” que ha recibido el post.
Comentarios	Número de comentarios que ha recibido el post.
Texto	Se trata del texto bajo la foto, escrito por el propio usuario.
Hashtags	una lista de los hashtags que aparecen en el texto de la imagen.
Sponsored	Es 0 si el post no es patrocinado y 1 si se trata de un post patrocinado.
Sponsor	En caso de ser un post patrocinado, quien es el patrocinador o colaborador.
Menciones	Una lista de los usuarios mencionados en el texto de la imagen.
Etiquetas	Una lista de los usuarios etiquetados en la imagen.
videoView	En caso de ser un vídeo, las visitas que ha tenido.
videoTime	En caso de ser un vídeo, la duración que tiene.

Tabla 3 Metadatos descargados asociados a los posts.

Fuente: Elaboración propia.

Además de todos los metadatos descargados, se ha almacenado todo el contenido multimedia (imágenes y videos) clasificado por usuarios. Es decir, por cada uno de los 149 *influencers*, tenemos una carpeta donde almacenamos todo el contenido multimedia que ha publicado en el año 2021.

3.4 Limpieza de datos

Uno de los primeros pasos para construir los dos corpus ha sido unificar todos los metadatos descargados (los asociados a los usuarios y los asociados a los *posts*). Una vez unidas ambas tablas, empezamos la limpieza de los datos.

Una vez agrupados todos los datos, identificamos valores perdidos ('NaN') en el dataset. Este valor nos indica que en la descarga no se estaba encontrando información. Por ejemplo, si en una publicación no se están utilizando hashtags, la variable *hashtags*

tomaba en valor 'NaN'. En nuestro caso, hemos decidido sustituir este valor por un 0. Exceptuando la variable 'category' donde se nos devolvía un NaN si la categoría no estaba especificada. Para solucionarlo, hemos sustituido los 'NaN' por la etiqueta 'Other'.

Además, localizamos valores anómalos en la variable 'likes', en los que especificaba que el número de likes obtenidos por el post son -1, lo que no tiene sentido. Por tanto, se decidió eliminar todos estos posts de nuestro corpus, dejándonos un total de 23.696 posts.

3.5 Variables

3.5.1 Variables dependientes

Las variables dependientes, es decir, las variables objeto del estudio, se recogen en la Tabla 4.

Variables	Valores
Likes	Cualquier valor positivo
Comentarios	Cualquier valor positivo
Engagement	Cualquier valor positivo

Tabla 4 Variables dependientes

Fuente: Elaboración propia.

Tanto los likes como los comentarios, son variables numéricas que hemos conseguido directamente de los metadatos descargados, pueden tomar cualquier valor positivo. En cambio, la variable Engagement la hemos generado nosotros a partir de los datos recuperados. Su cálculo es el siguiente:

$$Engagement = \frac{\text{Número de likes} + \text{Número de comentarios}}{\text{Número de followers}}$$

Este valor multiplicado por 100 nos indica el porcentaje de interacción con cada post respecto a la base de seguidores del perfil que ha publicado el propio post.

3.5.2 Variables de Usuario

Como ya hemos mencionado, este grupo de variables agrupa la información referente al usuario. Esta información es influyente y por lo tanto es necesario tener en cuenta estas variables para estudiar el comportamiento del engagement. En la Tabla 5 se detallan las variables de este grupo, así como los valores que puede tomar.

Variables	Valores
Followers	<ul style="list-style-type: none"> - Nano: seguidores entre 10.000 y 500.000 - Micro: seguidores entre 500.000 y 1.000.000 - Macro: más de 1.000.000 seguidores
Sigue_a_muchos	<ul style="list-style-type: none"> - 0: sigue a menos de 1985 perfiles - 1: sigue a más de 1985 perfiles
Verified	<ul style="list-style-type: none"> - 0: no es un perfil verificado - 1: es un perfil verificado
Business	<ul style="list-style-type: none"> - 0: no es una cuenta de negocio - 1: es una cuenta de negocio
Category	<ul style="list-style-type: none"> - Other: no especifica categoría - Creators and Celebrities - Publishers - General interest - Non-Profits and Religious Organizations - Personal Goods and General Merchandise Stores

Tabla 5 Variables a nivel de usuario

Fuente: Elaboración propia.

La variable *Followers*, que originalmente era numérica, la hemos transformado a una variable categórica. Esta variable nos clasifica el *post* en base a la cantidad de perfiles suscritos. La variable tomará el valor “Nano” cuando el número de seguidores del perfil este entre 10.000 y 500.000 seguidores, tomará el valor “Micro” cuando el número de seguidores esté entre 500.000 y un millón. En último lugar, hablaremos de que los *Followers* son “macro” cuando superan en millón.

La variable *Sigue_a_muchos* también es categórica. Esta variable nos indica la cantidad de perfiles a los que está suscrito el perfil que ha subido un *post* concreto. Esta variable tomará el valor 0 cuando el perfil no siga a más de 1985 perfiles y tomará el valor 1 en caso contrario. Se ha tomado el valor de 1985 porque se trata de la mediana del total de valores de la variable original *Followees*. Se utiliza la mediana para construir esta variable porque se quiere distinguir entre perfiles que siguen a muchos usuarios y perfiles que no siguen a tantos.

Por otro lado, tenemos las variables *Verified* y *Business*, ambas son categóricas y tomarán el valor 0 si los perfiles no son de este tipo. Tomarán el valor 1 en el caso de que sean perfiles verificados y estén configurados como business respectivamente.

Por último, en este grupo encontramos la variable *Category*, en la cual se nos indica la categoría con la que se identifica el perfil que ha publicado un *post* concreto.

3.5.3 Variables de *Post*

En este grupo, encontramos las variables que nos proporcionan información de un *post* concreto. En la *Tabla 6* se muestran las variables finales después de haber transformado los datos originales en variables que nos aporten información sobre el comportamiento de *engagement*.

Variables	Valores
Sponsored	- 0: post no pagado - 1: post pagado
Tipo	- GraphImage - GraphVideo - GraphSidecar: grupo de imágenes o videos.
Día	- Lunes - Martes - Miércoles - Jueves - Viernes - Sábado - Domingo
Estación	- Invierno: Enero, febrero y marzo - Verano: julio, agosto y septiembre - Primavera: abril, mayo y junio - Otoño: octubre, noviembre y diciembre
Momento del día	- Día: desde las 05:00 hasta las 12 :00 horas - Tarde: desde las 13:00 hasta las 21:00 horas - Noche: desde las 22:00 horas hasta las 04:00 horas
Hay_mención	- 0: no se menciona a ningún perfil - 1: se menciona a algún perfil
Hay_hashtags	- 0: no se utilizan hashtags - 1: se utilizan hashtags
Hay_etiquetas	- 0: no se ha etiquetado a ningún perfil - 1: se ha etiquetado a algún perfil
cita_marca	- 0: no se ha citado a ninguna marca - 1: se ha citado a alguna marca

Tabla 6 Variables a nivel de *post*

Fuente: *Elaboración propia.*

En el caso de la variable *Sponsored* no ha hecho falta hacer ninguna conversión a partir de los datos recuperados. Esta variable tomará el valor 0 cuando no se trate de un *post* patrocinado o pagado, y 1 en caso contrario.

La variable *Tipo* nos indica el tipo de *post* que se está publicando según el contenido multimedia. Si el *post* contiene una sola imagen será de tipo *GraphImage*, si contiene solo un video será de tipo *GraphVideo* y si contiene un grupo de 2 a 10 imágenes o videos será de tipo *GraphSidecar*.

Por otro lado, encontramos las variables de *Día*, *Estación* y *Momento del día*. Estas variables han sido creadas a partir del dato Fecha que teníamos en un inicio. En la descarga de los datos, obtuvimos la fecha y la hora de publicación. A partir de estas variables preparamos un pequeño fichero de código que recorriera todos los datos y fuera clasificando los *posts* según las condiciones indicadas.

En el primer caso, a raíz de la fecha completa fuimos capaces de identificar de que día de la semana se trataba. De este modo la variable *Día* toma como valor que día era, de lunes a Domingo, lo que nos deja 7 valores posibles para esta variable.

En el caso de la variable *estación*, lo primero que hicimos fue aislar la parte de la fecha que nos indicaba en que mes se había publicado el *post*. Una vez hecho esto solo fue necesario clasificar el mes:

- Si el número del mes estaba entre 4 y 6 (ambos inclusive), la variable tomaba el valor 'Primavera'.
- Si el número del mes estaba entre 7 y 9 (ambos inclusive), la variable tomaba el valor 'Verano'.
- Si el número del mes estaba entre 10 y 12 (ambos inclusive), la variable tomaba el valor 'Otoño'.
- Si el número del mes estaba entre 1 y 3 (ambos inclusive), la variable tomaba el valor 'Invierno'.

La lógica que se ha utilizado para la variable *Momento del día* es la misma comentada anteriormente. Pero, en este caso trabajando como punto de partida con la hora:

- Si la hora estaba entre 5 y 12 (ambos inclusive), la variable tomaba el valor 'Día'.
- Si la hora estaba entre 13 y 21 (ambos inclusive), la variable tomaba el valor 'Tarde'.
- Si la hora estaba entre 22 y 04 (ambos inclusive), la variable tomaba el valor 'Noche'.

Las variables *Hay_mención*, *Hay_hashtags* y *Hay_etiquetas* se han generado del mismo modo. A través de una pequeña función utilizando la librería de Pandas, en el caso en que se encontraran datos, la variable tomaba el valor 1, en caso contrario 0.

Para obtener la variable *cita_marca* se ha trabajado sobre la variable 'texto', que almacena el texto que acompaña al contenido multimedia en cada publicación. Además, se preparó una lista con las marcas más reconocidas en el mundo del running: Nike, Adidas, Asics, New Balance, Mizuno, Saucony, Hoka, Kalenji, Puma, Reebok, Under Armour.

La variable generada, *cita_marca*, toma el valor 1 si en el texto se encuentra alguna de las marcas de la lista, en caso contrario toma el valor 0.

3.5.4 Variables de Imagen

Además de todas las variables que se han comentado anteriormente asociadas al *post*, hemos querido profundizar y extraer todo el conocimiento posible de los datos recuperados. Para poder entender el comportamiento de los usuarios al interactuar con un *post*, es realmente importante estudiar el contenido multimedia del *post*.

Por ello, seleccionamos un par de posibles características que podrían darnos información interesante sobre el comportamiento de usuarios en el ámbito del running:

- Comprobar el número de personas que aparecen en la imagen. Esto podría ser interesante ya que, aunque generalmente el running es un deporte que se practica de manera individual, puede llevar a ser más atractivo si se realiza en grupo o se participa en eventos deportivos.
- Comprobar si la imagen es al aire libre o está tomada en interior. Esta característica nos ayudaría a verificar si la foto está tomada al aire libre.

Este tipo de características nos supusieron un gran reto ya que no era viable revisar manualmente más de 23.000 *posts*. Pero dada la importancia e influencia que pueden tener en el comportamiento de los usuarios decidimos plantear una manera automática de extraer esta información sin necesidad de revisar manualmente todos los *posts*. Aunque en un principio, se pretendía trabajar con el dataset completo, tuvimos que descartar todos los *posts* de tipo video o sidecar, ya que el tratamiento de este tipo de datos multimedia adquiriría una complejidad fuera de nuestras competencias. Para su eliminación preparamos un pequeño programa escrito en Python que al ejecutarlo redujo el dataset a 13.446 imágenes. Una vez preparado el dataset procedemos a la extracción de características de las imágenes:

Número de caras

Para la generación de esta característica se ha decidido utilizar la librería OpenCV, ya que cuenta con clasificadores para la detección de rostros y es fácil de implementar. Con un sencillo programa en Python y utilizando la librería comentada se ha generado una variable numérica. Esta variable almacena el número de rostros que se detectan en cada imagen.

Espacio cerrado o abierto

En esta ocasión decidimos plantear la creación de una red neuronal que nos clasificara nuestras imágenes en *outdoor* o *indoor*.

Las redes neuronales son un conjunto de algoritmos, modelados a la manera del cerebro humano, que están diseñados para reconocer patrones. Interpretan los datos sensoriales a través de una especie de percepción mecánica, etiquetando o agrupando la información en bruto. Los patrones que reconocen son numéricos, contenidos en vectores, a los que deben traducirse todos los datos del mundo real, ya sean imágenes, sonidos, textos o series temporales (Bosch Rue, Casas Roma, & Lozano Bagen, 2019).

En primer lugar, se investigó si ya existía algún dataset útil para nuestro problema o por lo contrario teníamos que construirlo con un proceso de *web Scraping*. La búsqueda resultó exitosa y conseguimos un dataset de 800 imágenes clasificadas *en outdoor e indoor* (Nadian-Ghomsheh, 2017). La arquitectura elegida para el modelo fue AlexNet, que ganó el reto de reconocimiento visual a gran escala de *Imagenet* en 2012. La Figura 8 muestra las distintas capas de las que se compone este modelo.

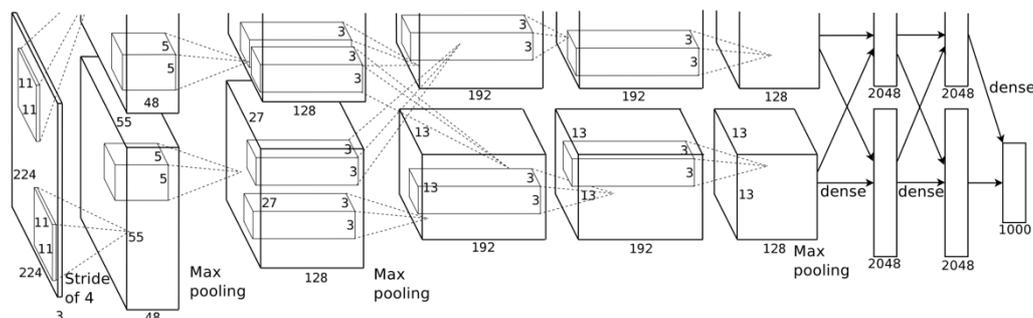


Figura 8 Arquitectura CNN AlexNet.

Fuente: (Krizhevsky, Sutskever, & Hinton, 2012).

Una vez configurada la arquitectura de AlexNet a nuestro dataset procedimos a entrenarla. Los resultados conseguidos fueron muy buenos, consiguiendo una probabilidad de acierto de 83,55 % (Primo Aparici, 2022). Con estos resultados estamos en disposición de utilizar este modelo para reconocer si las imágenes están tomadas al aire libre o no.

3.6 Modelos Análisis de regresión multivariable

Con todas las variables preparadas, pasamos a elegir el tipo de modelo que mejor se pudiera ajustar a nuestro estudio. Dada la amplitud de los corpus, y la cantidad de variables generadas nos decantamos por un análisis jerárquico de regresión multivariable. Este análisis nos permite el estudio de todas las variables al mismo tiempo que nos facilita el trabajo a la hora de interpretar los resultados de los modelos.

Para asegurarnos que podíamos utilizar este modelo, comprobamos que se cumplía la condición de multicolinealidad en ambos corpus. Estas pruebas se encuentran en el repositorio del TFG.

Una vez comprobado, preparamos las variables categóricas múltiples para que el modelo las pudiera entender. Todas las variables de este tipo ha sido necesario codificarlas con *dummies* también conocido como variables ficticias. Este proceso genera $n-1$ *dummies* por variable, siendo n el número total de valores distintos que puede tomar la variable. Para que se entienda el uso de estas variables vamos a ejemplificarlo. En el caso de la variable *Followers*, que toma los valores nano, micro y macro, se generan dos variables ficticias, por ejemplo, nano y micro. Cada una de estas

variables tomará el valor 0 cuando el rango de Followers del perfil que ha subido el post no sea el especificado. En caso de que sí lo sea, tomara el valor 1.

Por último, se planteó la construcción de los modelos de manera jerárquica. De este modo podemos analizar cómo van influyendo las variables de distinta naturaleza conforme se van añadiendo. Para cada corpus se han tratado las 3 variables dependientes. En el caso del corpus completo se han realizado 3 modelos por cada variable dependiente. En el primero incluimos las variables del post, en el segundo añadimos las variables de usuario y en el último el subgrupo de variables de usuario referidas a los seguidores. En el caso del corpus reducido, se han realizado 4 modelos por cada variable dependiente. En este caso, además, incluimos las variables de imagen.

3.7 Herramientas utilizadas

El lenguaje escogido para realizar este proyecto ha sido Python en su última versión disponible (3.11), tanto para la obtención de datos como para el entrenamiento de los modelos propuestos de *Deep Learning*.

El entorno de desarrollo elegido para la realización del proyecto es *VisualStudio Code*, uno de los entornos más conocidos y fáciles de utilizar. Ofrece un montón de extensiones que facilitan el desarrollo y es muy intuitivo. La Figura 9 muestra una captura de pantalla de este entorno.

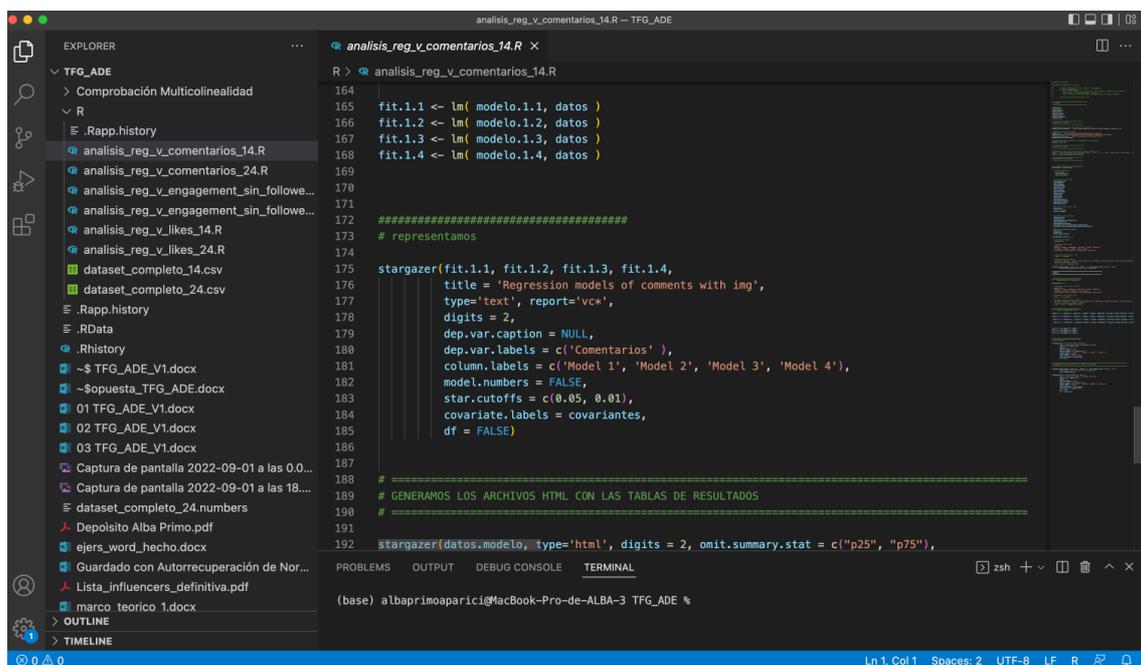


Figura 9 Entorno de desarrollo *Visual Studio Code*.

Fuente: *Elaboración propia*.

A continuación, se detallan las librerías de Python que han sido utilizadas para la realización del proyecto:

- **Instaloader:** se trata de una librería de código libre que te permite descargar imágenes y datos de la aplicación de Instagram de una manera intuitiva y sencilla. Aunque se trata de una librería muy útil presenta ciertas limitaciones que se comentarán posteriormente.
- **NumPy:** implementa vectores multidimensionales y matrices para almacenar grandes cantidades de datos. Cuenta con funciones matemáticas y estructuras de datos de alto nivel.
- **Pandas:** es una biblioteca de Python que se utiliza para trabajar con conjuntos de datos. Tiene funciones para analizar, limpiar, explorar y manipular datos. Pandas nos permite analizar *big data* y sacar conclusiones basadas en teorías estadísticas. Además, puede limpiar conjuntos de datos desordenados y hacerlos legibles y relevantes.
- **Scikit-learn (Sklearn):** es la biblioteca más útil y robusta para el *machine learning* en Python. Proporciona una selección de herramientas eficientes para el aprendizaje automático y el modelado estadístico, incluyendo la clasificación, la regresión, la agrupación y la reducción de la dimensionalidad a través de una interfaz consistente en Python.
- **TensorFlow:** Es un marco de trabajo desarrollado y mantenido por Google que permite la ejecución de operaciones matemáticas, mediante diagramas de flujo de datos, de una forma optimizada. Es flexible, escalable y está constante actualización y mantenimiento.
- **Keras:** El objetivo de esta librería es acelerar la creación de redes neuronales.
- **OpenCV:** es una librería de código abierto para la visión por ordenador y el aprendizaje automático. La biblioteca cuenta con más de 2.500 algoritmos optimizados, que incluyen un amplio conjunto de algoritmos de visión por ordenador y aprendizaje automático, tanto clásicos como de última generación. Estos algoritmos pueden utilizarse para detectar y reconocer caras, identificar objetos y clasificar acciones humanas en vídeos...

Una de las herramientas clave para la realización de este proyecto ha sido RStudio, ya que nos permite usar R. Se trata de un entorno de desarrollo orientado al análisis estadístico y su representación gráfica. Se trata de un software muy fácil de usar. Además, se puede trabajar con multitud de funcionalidades distintas a través de la gran cantidad de librerías que ofrece.

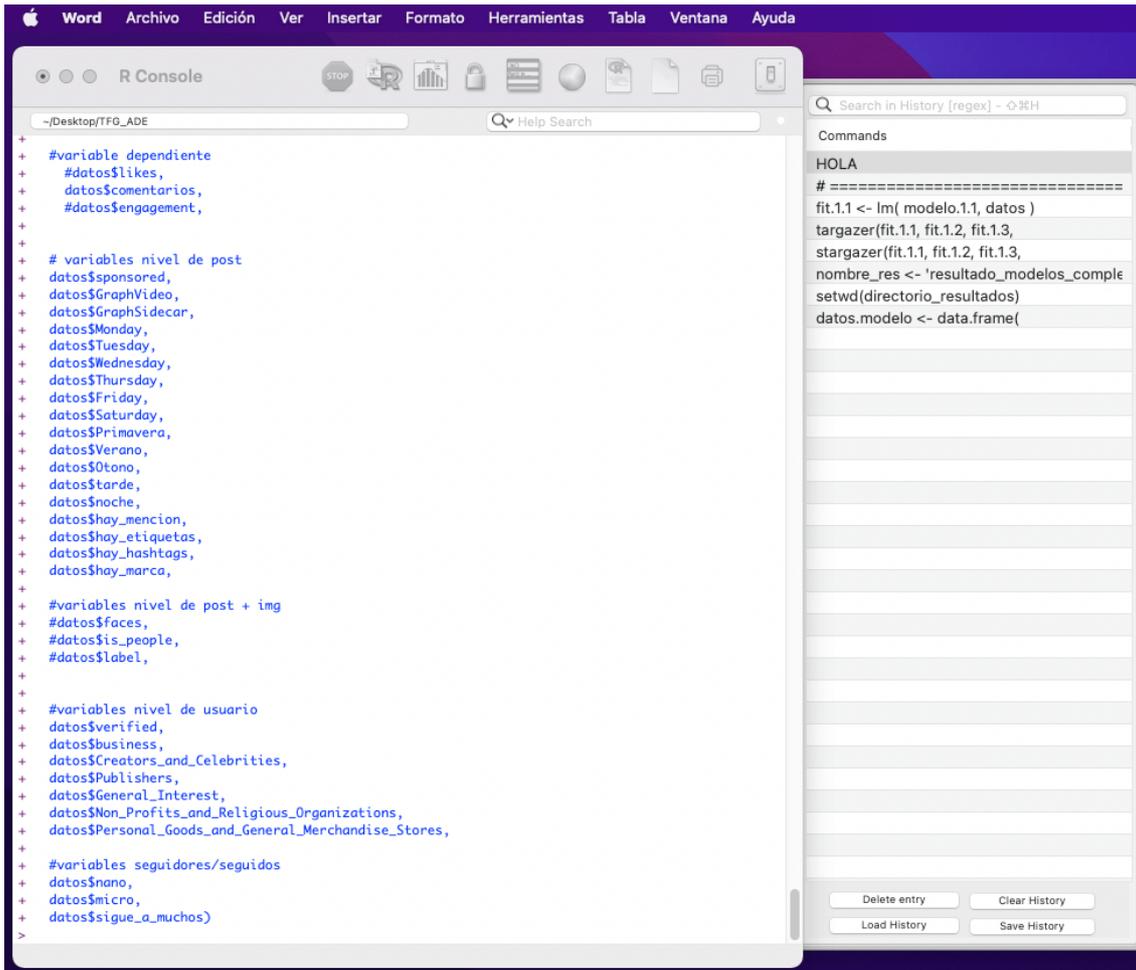


Figura 10 Entorno de trabajo RStudio.
Fuente: Elaboración propia.

4. Resultados

En primer lugar, se presentan los estadísticas descriptivas de las variables para el corpus completo (Tabla 7) y el corpus reducido (Tabla 8).

Descriptive statistics of the variables

Statistic	N	Mean	St. Dev.	Min	Max
Likes	23,696	4,495.00	24,861.00	9	1,529,743
Comentarios	23,696	83.00	503.00	0	52,736
Engagement	23,696	0.05	0.02	0.0005	0.79
sponsored	23,696	0.02	0.15	0	1
GraphVideo	23,696	0.14	0.35	0	1
GraphSidecar	23,696	0.29	0.45	0	1
Lunes	23,696	0.15	0.35	0	1
Martes	23,696	0.14	0.35	0	1
Miércoles	23,696	0.14	0.35	0	1
Jueves	23,696	0.14	0.35	0	1
Viernes	23,696	0.15	0.35	0	1
Sábado	23,696	0.13	0.34	0	1
Primavera	23,696	0.26	0.44	0	1
Verano	23,696	0.25	0.43	0	1
Otoño	23,696	0.24	0.43	0	1
tarde	23,696	0.54	0.50	0	1
noche	23,696	0.19	0.39	0	1
hay_mencion	23,696	0.52	0.50	0	1
hay_etiquetas	23,696	0.60	0.49	0	1
hay_hashtags	23,696	0.73	0.44	0	1
hay_marca	23,696	0.09	0.28	0	1
verified	23,696	0.37	0.48	0	1
business	23,696	0.45	0.50	0	1
Creators_and_Celebrities	23,696	0.34	0.48	0	1
Publishers	23,696	0.01	0.12	0	1
General_interest	23,696	0.02	0.13	0	1
Non_Profits_and_Religious_Organizations	23,696	0.03	0.17	0	1
Personal_Goods_and_General_Merchandise	23,696	0.04	0.19	0	1
nano	23,696	0.96	0.20	0	1
micro	23,696	0.02	0.12	0	1
sigue_a_muchos	23,696	0.31	0.46	0	1

Tabla 7 Estadísticas descriptivas de las variables para el dataset completo.

Fuente: Elaboración propia.

Descriptive statistics of the variables

Statistic	N	Mean	St. Dev.	Min	Max
Likes	13,446	3,582.00	21,457.00	10	690,534.00
Comentarios	13,446	70.00	346.00	0	17,050
Engagement	13,446	0.03	0.03	0.0005	0.70
sponsored	13,446	0.02	0.13	0	1
Lunes	13,446	0.15	0.35	0	1
Martes	13,446	0.14	0.35	0	1
Miércoles	13,446	0.14	0.35	0	1
Jueves	13,446	0.14	0.35	0	1
Viernes	13,446	0.15	0.36	0	1
Sábado	13,446	0.13	0.34	0	1
Primavera	13,446	0.26	0.44	0	1
Verano	13,446	0.25	0.43	0	1
Otoño	13,446	0.24	0.43	0	1
tarde	13,446	0.53	0.50	0	1
noche	13,446	0.20	0.40	0	1
hay_mencion	13,446	0.49	0.50	0	1
hay_etiquetas	13,446	0.61	0.49	0	1
hay_hashtags	13,446	0.72	0.45	0	1
hay_marca	13,446	0.08	0.27	0	1
faces	13,446	0.85	1.30	0	33
is_outdoor	13,446	0.53	0.50	0	1
verified	13,446	0.29	0.45	0	1
business	13,446	0.44	0.50	0	1
Creators_and_Celebrities	13,446	0.34	0.47	0	1
Publishers	13,446	0.01	0.12	0	1
General_interest	13,446	0.02	0.13	0	1
Non_Profits_and_Religious_Organizations	13,446	0.05	0.21	0	1
Personal_Goods_and_General_Merchandise	13,446	0.02	0.13	0	1
sigue_a_muchos	13,446	0.33	0.47	0	1

Tabla 8 Estadísticas descriptivas de las variables para el dataset reducido

Fuente: Elaboración propia.

A la vista de los resultados de los modelos, es importante comentar que en todos los análisis el estadístico F es significativo, lo que nos indica que al menos alguna variable independiente tiene capacidad de explicar la variabilidad de nuestras variables dependientes: *likes*, *comentarios* y *engagement*. De entre todos los modelos, presentan un valor más alto los que estudian en el comportamiento de la variable *likes* por lo que, es la variable de la que mejor podemos entender su comportamiento.

Seguidamente se presentarán los resultados de todos los modelos por variable dependiente.

4.1 Variable dependiente *likes*

En la Tabla 9, tenemos los resultados de los 3 modelos preparados con el dataset completo, es decir, los 23.696 *posts* para el estudio de la variable *likes*. En el primer modelo encontramos las variables a nivel de *post*, en el segundo las variables de *post* incluyendo las de usuario y en el modelo 3 además de todas las variables anteriores se incluyen un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa. Se ha decidido hacer este subgrupo dada la influencia de esta variable en nuestras variables dependientes. En el primer modelo resultan interesantes las siguientes variables ya son significativas al 10 %:

- *GraphVideo*: en este caso, nos está indicando que cuando el tipo de *post* es video frente a un *post* de solo una imagen recibe un mayor número de *likes*.
- *Verano*: las publicaciones compartidas en verano reciben una mejor aceptación frente las publicadas en invierno.
- *Tarde*: según los datos las publicaciones subidas entre las 13h y las 21h consiguen más *likes* frente las subidas entre las 04h y las 12h.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas no se consiguen tantos *likes* frente a los *posts* donde no hay etiquetas.
- *Hay_hashtags*: como en el caso anterior la utilización de hashtags parece disminuir la cantidad de *likes* conseguidos en el *post*.

Además, cabe mencionar que los *posts* subidos entre las 04h y 12h son mejor recibidos frente los publicados por la noche. Aunque el nivel de significación es menor.

Al incluir las variables a nivel de usuario, *hay_menciones* pasa a ser significativa, y parece que se reciben menos *likes* cuando se encuentran menciones en el *post*.

Es destacable en este segundo modelo que todas las variables incluidas son significativas al 10 %. Por un lado, los *posts* subidos por cuentas verificadas son propensas a obtener más *likes* frente a los *posts* subidos por cuentas no verificadas. Pasa lo contrario con la variable *business*, por lo que parece que las publicaciones hechas por cuentas no business reciben un mayor número de *likes*.

Por último, encontramos que, por lo general especificar la categoría del perfil con la que se está subiendo la imagen frente no hacerlo proporciona un mayor número de *likes* en las publicaciones.

Al hacer la última inserción de variables, en el modelo 3, los resultados varían bastante, la variable de verano pierde influencia y la gana la variable Otoño, que nos indica que las publicaciones subidas en esta época tienen menor número de *likes* que las subidas

en invierno. Además, en este modelo observamos que las publicaciones subidas por la noche influyen positivamente en el número de *likes* que obtiene el *post*.

Regression models of likes

	<i>Dependent variable:</i>		
	Likes		
	Model 1	Model 2	Model 3
sponsored	-1,143.00	-864.00	-1,864.00
GraphVideo	5,589.00**	4,174.00**	-694.00
GraphSidecar	389.00	-681.00	-161.00
Lunes	177.00	216.00	226.00
Martes	-834.00	-619.00	-459.00
Miércoles	139.00	408.00	264.00
Jueves	352.00	523.00	351.00
Viernes	-3.80	-2.90	-411.00
Sábado	357.00	592.00	172.00
Primavera	347.00	240.00	-161.00
Verano	2,338.00**	2,186.00**	854.00*
Otoño	397.00	-20.00	-801.00*
tarde	1,555.00**	1,340.00**	2,065.00**
noche	-1,053.00*	-758.00	1,105.00**
hay_mencion	-276.00	-1,359.00**	-514.00
hay_etiquetas	-1,261.00**	-1,650.00**	214.00
hay_hashtags	-2,430.00**	-2,608.00**	-2,088.00**
hay_marca	-844.00	-1,549.00**	458.00
verified		8,300.00**	3,309.00**
business		-3,380.00**	623.00
Creators_and_Celebrities		7,761.00**	911.00
Publishers		5,443.00**	1,605.00
General_interest		7,444.00**	3,897.00**
Non_Profits_and_Religious_Organizations		7,621.00**	2,146.00
Personal_Goods_and_General_Merchandise			
nano			-72,594.00**
micro			-52,186.00**
sigue_a_muchos			-883.00**
Constant	4,929.00**	1,830.00**	73,531.00**
Observations	23,696	23,696	23,696
R ²	0.01	0.05	0.27
Adjusted R ²	0.01	0.05	0.27
Residual Std. Error	24,701.00	24,259.00	21,296.00
F Statistic	18.00**	51.00**	319.00**

Note:

*p<0.05; **p<0.01

Tabla 9 Modelos de regresión multivariable con dataset completo y variable dependiente likes.

Fuente: Elaboración propia.

Centrándonos en las variables a nivel de usuario, *business* deja de ser influyente. Por otro lado, vemos que la categoría *general_interest* sigue siendo influyente en el número de *likes* conseguidos. Sin embargo, las demás categorías dejan de ser influyentes.

Las nuevas variables introducidas en el tercer modelo son todas significativas y sustentan que conforme más seguidores, más *likes* se obtienen, ya que tanto *nano* como *micro* consiguen menos *likes* frente a los perfiles macro (más de 1 millón de seguidores). También observamos que si el perfil que sube la publicación no sigue a muchos perfiles se consigue un mayor número de *likes*.

Por último, se observa que el r^2 mejora de un 0,05 a un 0,27 al incluir el último grupo de variables.

En la Tabla 10, tenemos los resultados de los 4 modelos preparados con el dataset reducido, es decir, los 13.446 *posts*. En el primer modelo encontramos las variables a nivel de post, en el segundo las variables de post incluyendo las extraídas de las propias imágenes, en el modelo 3 además de todas las variables anteriores se incluyen las variables a nivel de usuario y en el cuarto modelo se incluye un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa.

Cabe mencionar que en estos modelos se ha prescindido de las variables *GraphVideo* y *GraphSidecar* ya que este dataset solo está compuesto por imágenes tipo *GraphImage*.

En el primer modelo, resultan interesantes las siguientes variables ya que son significativas al 10 %:

- *Verano*: las publicaciones compartidas en verano reciben una mejor aceptación frente las publicadas en invierno.
- *Tarde*: según los datos las publicaciones subidas entre las 13h y las 21h consiguen más *likes* frente las subidas entre las 04h y las 12h.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas no se consiguen tantos *likes* frente a los posts donde no hay etiquetas.
- *Hay_hashtags*: como en el caso anterior la utilización de hashtags parece disminuir la cantidad de *likes* conseguidos en el *post*.

Regression models of likes with img

	<i>Dependent variable:</i>			
	Likes			
	Model 1	Model 2	Model 3	Model 4
sponsored	-1,270.00	-1,463.00	-569.00	-936.00
Lunes	-662.00	-649.00	-517.00	-202.00
Martes	-1,120.00	-1,151.00	-872.00	-323.00
Miércoles	-369.00	-403.00	-321.00	-139.00
Jueves	-280.00	-300.00	-98.00	81.00
Viernes	723.00	684.00	500.00	123.00
Sábado	-161.00	-193.00	6.50	421.00
Primavera	667.00	651.00	629.00	-75.00
Verano	2,329.00**	2,332.00**	2,459.00**	845.00*
Otoño	30.00	43.00	-108.00	-755.00
tarde	1,481.00**	1,505.00**	1,494.00**	1,585.00**
noche	-672.00	-591.00	-340.00	994.00*
hay_mencion	-299.00	-306.00	-1,326.00**	-403.00
hay_etiquetas	-1,134.00**	-1,041.00**	-1,221.00**	743.00*
hay_hashtags	-3,095.00**	-3,055.00**	-3,067.00**	-2,157.00**
hay_marca	-800.00	-758.00	-1,279.00	25.00
faces		20.00	-50.00	-23.00
is_outdoor		-2,009.00**	-2,089.00**	-1,190.00**
verified			8,892.00**	2,870.00**
business			-2,357.00	535.00
Creators_and_Celebrities			6,064.00**	836.00
Publishers			4,298.00*	806.00
General_interest			4,605.00*	1,126.00
Non_Profits_and_Religious_Organizations			7,235.00**	2,513.00
Personal_Goods_and_General_Merchandise				
nano				-95,560.00**
micro				-84,371.00**
sigue_a_muchos				-115.00
Constant	5,593.00**	6,542.00**	3,070.00**	96,741.00**
Observations	13,446	13,446	13,446	13,446
R ²	0.01	0.01	0.06	0.39
Adjusted R ²	0.01	0.01	0.05	0.39
Residual Std. Error	21,363.00	21,341.00	20,863.00	16,785.00
F Statistic	8.40**	9.10**	33.00**	317.00**

Note:

*p<0.05; **p<0.01

Tabla 10 Modelos de regresión multivariable con dataset reducido y variable dependiente likes

Fuente: Elaboración propia.

Al incluir las variables extraídas de las imágenes, las variables influyentes del modelo anterior se mantienen y, además, observamos que la variable *is_outdoor* sí que es significativa. Esto quiere decir que, si la imagen publicada es tomada en un espacio cerrado, frente a un espacio al aire libre se obtienen más *likes*.

Cuando incluimos las variables a nivel de usuario, *hay_menciones* pasa a ser significativa, y parece que se reciben menos *likes* cuando se encuentran menciones en el *post*.

Es destacable en este tercer modelo que todas las variables incluidas son significativas al 10 %, a excepción de *business*. Por un lado, los *posts* subidos por cuentas verificadas son propensas a obtener más *likes* frente a los *posts* subidos por cuentas no verificadas. Además, encontramos que especificar la categoría del perfil con la que se está subiendo la imagen frente no hacerlo proporciona un mayor número de *likes* en las publicaciones por lo general.

En el último modelo, al incluir el subgrupo de variables de usuario dejan de ser influyentes las variables relacionadas con la categoría del perfil y la variable *hay_mención*. En cambio, la variable *noche* pasa a ser significativa. El subgrupo añadido de variables nos muestra que conforme más seguidores se tiene más *likes* obtienen las publicaciones.

Cabe destacar que el r^2 mejora notablemente en el último modelo. Además, mejora respecto al mejor r^2 conseguido con el dataset completo, de un 0,27 a un 0,39.

4.2 Variables dependiente *comentarios*

En la Tabla 11 tenemos los resultados de los 3 modelos preparados con el dataset completo para la variable dependiente *Comentarios*. En el primer modelo encontramos las variables a nivel de post, en el segundo las variables de post incluyendo las de usuario y en el modelo 3 además de todas las variables anteriores se incluyen un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa. Se ha decidido hacer este subgrupo dada la influencia de esta variable en nuestras variables dependientes.

En el primer modelo, resultan interesantes las siguientes variables ya que son significativas sobre la variable *comentarios*:

- *GraphVideo*: en este caso, nos está indicando que cuando el tipo de *post* es video frente a un *post* de solo una imagen recibe un mayor número de *comentarios*.
- *GraphSidecar*: pasa lo mismo que en el caso anterior cuando el tipo de *post* es *GraphSidecar* frente a *GraphImage*.
- *Verano*: las publicaciones compartidas en verano reciben más *comentarios* frente las publicadas en invierno.

Regression models of Comments

	<i>Dependent variable:</i>		
	Comentarios		
	Model 1	Model 2	Model 3
sponsored	-34.00	-32.00	-36.00
GraphVideo	36.00**	8.90	-22.00*
GraphSidecar	26.00**	9.10	9.40
Lunes	23.00	23.00	23.00
Martes	3.80	6.50	7.00
Miércoles	2.50	4.70	3.60
Jueves	21.00	23.00	22.00
Viernes	4.10	4.30	2.40
Sábado	-4.50	-1.30	-3.20
Primavera	3.10	2.90	0.73
Verano	19.00*	17.00	10.00
Otoño	8.90	4.90	-3.50
tarde	6.50	9.40	12.00
noche	-19.00*	-7.10	2.90
hay_mencion	25.00**	15.00*	17.00*
hay_etiquetas	18.00*	18.00*	28.00**
hay_hashtags	-	-	-45.00**
hay_marca	-1.10	-9.30	0.71
verified		99.00**	77.00**
business		48.00**	68.00**
Creators_and_Celebrities		-25.00	-66.00**
Publishers		-47.00	-56.00
General_interest		12.00	-0.70
Non_Profits_and_Religious_Organizations		-39.00	-67.00**
Personal_Goods_and_General_Merchandise			
nano			-348.00**
micro			-72.00*
sigue_a_muchos			-20.00**
Constant	74.00**	35.00*	381.00**
Observations	23,696	23,696	23,696
R ²	0.005	0.01	0.03
Adjusted R ²	0.004	0.01	0.03
Residual Std. Error	502.00	500.00	496.00
F Statistic	6.30**	14.00**	27.00**

Note:

*p<0.05; **p<0.01

Tabla 11 Modelos de regresión multivariable para el dataset completo y variable dependiente comentarios

Fuente: Elaboración propia

- *Noche*: según los datos las publicaciones subidas entre las 22h y las 04h consiguen menos comentarios frente las subidas entre las 04h y las 12h.
- *Hay_mención*: en los *posts* donde hay menciones se consiguen más comentarios que en los *posts* donde no hay menciones.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas se consiguen más comentarios que los *posts* donde no hay etiquetas.
- *Hay_hashtags*: como en el caso anterior la utilización de hashtags parece disminuir la cantidad de comentarios conseguidos en el *post*.

Al incluir las variables a nivel de usuario, tanto las variables de tipo de *post*, como *verano* y *noche* dejan de ser influyentes. Y pasan a ser significativas *verified* y *business*, por lo que cuando se cumple esta condición se consiguen más comentarios.

Las nuevas variables introducidas en el tercer modelo son todas significativas y sustentan que conforme más seguidores, más comentarios se obtienen, ya que tanto nano como micro consiguen menos comentarios frente a los perfiles macro (más de 1 millón de seguidores). También observamos que si el perfil que sube la publicación no sigue a muchos perfiles se consigue un mayor número de comentarios. Además, es destacable que la variable *Graphvideo* vuelve a ser significativa, pero a la inversa, es decir, según los datos un *post* de tipo *GraphImage* obtendrá más comentarios respecto de un *post* tipo *GraphVideo*.

A pesar de que el r^2 mejora levemente en cada modelo, es muy bajo para poder considerar que estos datos explican el comportamiento de la variable comentarios.

En la Tabla 12 *Modelos de regresión multivariable dataset reducido y variable dependiente comentarios*.

, tenemos los resultados de los 4 modelos preparados con el dataset reducido donde incluimos las variables extraídas de las imágenes. En el primer modelo encontramos las variables a nivel de *post*, en el segundo las variables de *post* incluyendo las extraídas de las propias imágenes, en el modelo 3 además de todas las variables anteriores se incluyen las variables a nivel de usuario y en el cuarto modelo se incluye un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa.

En el primer modelo, resultan interesantes las siguientes variables ya que son significativas para explicar la variabilidad de la variable comentarios:

- *Verano*: las publicaciones compartidas en verano reciben una mejor aceptación frente las publicadas en invierno.
- *Tarde*: según los datos las publicaciones subidas entre las 13h y las 21h consiguen más comentarios frente las subidas entre las 04h y las 12h.

Regression models of comments with img

	<i>Dependent variable:</i>			
	Comentarios			
	Model 1	Model 2	Model 3	Model 4
sponsored	-32.00	-34.00	-27.00	-29.00
Lunes	1.20	1.20	-2.70	-1.70
Martes	-7.10	-7.50	-6.50	-4.40
Miércoles	5.80	5.30	3.20	3.90
Jueves	6.90	6.50	7.80	8.20
Viernes	6.60	6.10	4.50	2.70
Sábado	-6.30	-6.80	-7.40	-5.90
Primavera	5.50	5.30	3.90	1.10
Verano	19.00*	19.00*	17.00*	10.00
Otoño	-0.98	-0.77	-5.30	-8.40
tarde	14.00*	14.00*	20.00**	20.00**
noche	-12.00	-12.00	4.30	8.70
hay_mencion	22.00**	22.00**	14.00*	19.00**
hay_etiquetas	21.00**	22.00**	29.00**	35.00**
hay_hashtags	-49.00**	-48.00**	-42.00**	-36.00**
hay_marca	-6.20	-5.60	-13.00	-7.90
faces		-2.20	-2.40	-2.30
is_outdoor		-24.00**	-21.00**	-18.00**
verified			100.00**	78.00**
business			201.00**	209.00**
Creators_and_Celebrities			-192.00**	-210.00**
Publishers			-192.00**	-197.00**
General_interest			-157.00**	-164.00**
Non_Profits_and_Religious_Organizations			-197.00**	-214.00**
Personal_Goods_and_General_Merchandise				
nano				-371.00**
micro				-317.00**
sigue_a_muchos				-12.00
Constant	70.00**	83.00**	38.00**	404.00**
Observations	13,446	13,446	13,446	13,446
R ²	0.01	0.01	0.03	0.05
Adjusted R ²	0.01	0.01	0.03	0.05
Residual Std. Error	345.00	345.00	341.00	337.00
F Statistic	5.90**	6.10**	18.00**	27.00**

Note:

*p<0.05; **p<0.01

Tabla 12 Modelos de regresión multivariable dataset reducido y variable dependiente comentarios.

Fuente: Elaboración propia

- *Hay_mención*: en los *posts* donde hay menciones se consiguen más comentarios que en los *posts* donde no hay menciones.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas se consiguen más comentarios que los *posts* donde no hay etiquetas.
- *Hay_hashtags*: como en el caso anterior la utilización de hashtags parece disminuir la cantidad de comentarios conseguidos en el *post*.

Al incluir las variables extraídas de las imágenes, las variables influyentes del modelo anterior se mantienen y, además, observamos que la variable *is_outdoor* sí que es significativa. Esto quiere decir que, si la imagen publicada es tomada en un espacio cerrado, frente a un espacio al aire libre se obtienen más comentarios.

Es destacable en el tercer modelo que todas las variables incluidas son significativas al 10 %. Por un lado, los *posts* subidos por cuentas verificadas y son business son propensas a obtener más comentarios frente a los *posts* subidos por cuentas no verificadas o no de tipo business. Además, encontramos que especificar la categoría del perfil con la que se está subiendo la imagen frente no hacerlo proporciona un mayor número de comentarios en las publicaciones generalmente.

En el último modelo, al incluir el subgrupo de variables de usuario dejan de ser influyente la variable *verano*. El subgrupo añadido de variables nos muestra que conforme más seguidores se tiene más comentarios obtienen las publicaciones.

A pesar de que el r^2 mejora levemente en cada modelo, es muy bajo para poder considerar que estos datos explican el comportamiento de la variable.

4.3 Variable dependiente *engagement*

En la Tabla 13, tenemos los resultados de los 3 modelos preparados con el dataset completo, es decir, los 23.696 *posts*. En el primer modelo encontramos las variables a nivel de *post*, en el segundo las variables de *post* incluyendo las de usuario y en el modelo 3 además de todas las variables anteriores se incluyen un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa. Se ha decidido hacer este subgrupo dada la influencia de esta variable en nuestras variables dependientes.

En el primer modelo, resultan interesantes las siguientes variables ya que son significativas explicando la variabilidad de la variable *engagement*:

- *Sponsored*: cuando el *post* es patrocinado obtiene menos *engagement* frente a los *posts* no pagados.

- *GraphVideo*: en este caso, nos está indicando que cuando el tipo de *post* es video frente a un *post* de solo una imagen el *engagement* conseguido es menor.

Regression models of Engagement

	<i>Dependent variable:</i>		
	Engagement		
	Model 1	Model 2	Model 3
sponsored	-0.01**	-0.01**	-0.01**
GraphVideo	-0.01**	-0.003*	-0.003*
GraphSidecar	0.004**	0.005**	0.005**
Lunes	0.0001	0.0004	0.0005
Martes	-0.001	-0.001	-0.001
Miércoles	-0.0005	-0.0001	-0.0001
Jueves	-0.002	-0.002	-0.002
Viernes	-0.002	-0.002	-0.002
Sábado	0.001	0.001	0.001
Primavera	-0.001	-0.0005	-0.0005
Verano	0.003*	0.002	0.002
Otoño	0.002	0.002	0.002
tarde	0.003*	0.003*	0.003*
noche	-0.002	-0.001	-0.001
hay_mencion	0.001	0.0004	0.0003
hay_etiquetas	0.004**	0.004**	0.004**
hay_hashtags	-0.005**	-0.01**	-0.01**
hay_marca	0.01**	0.005**	0.005**
verified		-0.001	-0.002
business		-0.03**	-0.02**
Creators_and_Celebrities		0.02**	0.02**
Publishers		0.03**	0.03**
General_interest		0.07**	0.07**
Non_Profits_and_Religious_Organizations		0.01**	0.01**
Personal_Goods_and_General			
sigue_a_muchos			0.001
Constant	0.03**	0.04**	0.04**
Observations	23,696	23,696	23,696
R ²	0.01	0.02	0.02
Adjusted R ²	0.01	0.02	0.02
Residual Std. Error	0.07	0.07	0.07
F Statistic	8.00**	22.00**	21.00**
Note:		*p<0.05; **p<0.01	

Tabla 13 Modelos de regresión multivariable para el dataset completo y variable dependiente *engagement*.

Fuente: Elaboración propia

- *GraphSidecar*: en este caso, nos está indicando que cuando el tipo de *post* es *sidecar* frente a un *post* de solo una imagen el *engagement* conseguido es mayor.
- *Verano*: las publicaciones compartidas en verano reciben una mejor aceptación frente las publicadas en invierno.
- *Tarde*: según los datos las publicaciones subidas entre las 13h y las 21h consiguen un mayor *engagement* frente las subidas entre las 04h y las 12h.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas se consigue más *engagement* frente a los *posts* donde no hay etiquetas.
- *Hay_hashtags*: la utilización de hashtags parece disminuir el *engagement* conseguido en el *post*.
- *Hay_marca*: los posts donde se cita alguna marca reconocida consiguen un mayor *engagement* frente los *posts* donde no se cita ninguna marca.

Al incluir las variables a nivel de usuario, la variable verano deja de ser significativa.

Es destacable en este segundo modelo que todas las variables incluidas son significativas al 10 %, a excepción de *verified*. Por un lado, los *posts* subidos por cuentas tipo business son propensas a obtener un menor *engagement* frente a los *posts* subidos por cuentas que no son de este tipo.

Por último, encontramos que especificar la categoría del perfil con la que se está subiendo la imagen frente no hacerlo, generalmente proporciona un mayor *engagement* en las publicaciones.

Al hacer la última inserción de variables, en el modelo 3, los resultados se mantienen prácticamente igual que en el modelo 2, ya que la variable introducida en este caso no es significativa.

A pesar de que el r^2 mejora un 0.01 respecto del primer modelo comentado, es muy bajo para poder considerar que estos datos explican el comportamiento de la variable *engagement*.

En la Tabla 14, tenemos los resultados de los 4 modelos preparados con el dataset reducido donde incluimos las variables extraídas de las imágenes. En el primer modelo encontramos las variables a nivel de post, en el segundo las variables de post incluyendo las extraídas de las propias imágenes, en el modelo 3 además de todas las variables anteriores se incluyen las variables a nivel de usuario y en el cuarto modelo se incluye un subgrupo de variables de usuarios específicas que cuantifican la cantidad de seguidores que tiene el perfil que ha subido el *post* y viceversa.

Regression models of engagement with img

	<i>Dependent variable:</i>			
	Engagement			
	Model 1	Model 2	Model 3	Model 4
sponsored	-0.01**	-0.01**	-0.01**	-0.01**
Lunes	-0.003**	-0.003**	-0.003**	-0.002**
Martes	-0.001	-0.001	-0.001	-0.001
Miércoles	-0.002*	-0.002*	-0.002*	-0.002*
Jueves	-0.002*	-0.002*	-0.002**	-0.002**
Viernes	-0.001	-0.001	-0.001	-0.001
Sábado	0.0000	0.0000	-0.0000	0.0000
Primavera	-0.0004	-0.0004	0.0001	0.0002
Verano	0.002**	0.002**	0.001*	0.001*
Otoño	0.0002	0.0002	0.001	0.001
tarde	0.001	0.001	0.0004	0.0004
noche	-0.002**	-0.002**	-0.001	-0.001
hay_mencion	0.0001	0.0000	0.001*	0.001
hay_etiquetas	0.003**	0.003**	0.003**	0.003**
hay_hashtags	-0.002**	-0.002**	-0.003**	-0.003**
hay_marca	0.003**	0.003**	0.001	0.001
faces		0.001**	0.0003	0.0003
is_outdoor		-0.002**	-0.002**	-0.002**
verified			-0.001**	-0.002**
business			-0.03**	-0.03**
Creators_and_Celebrities			0.02**	0.02**
Publishers			0.01*	0.003
General_interest			0.05**	0.05**
Non_Profits_and_Religious_Organizations			0.01**	0.01**
Personal_Goods_and_General_Merchandise				
sigue_a_muchos				0.003**
Constant	0.03**	0.03**	0.04**	0.04**
Observations	13,446	13,446	13,446	13,446
R ²	0.01	0.02	0.09	0.10
Adjusted R ²	0.01	0.01	0.09	0.10
Residual Std. Error	0.03	0.03	0.03	0.03
F Statistic	11.00**	12.00**	58.00**	58.00**

Note:

*p<0.05; **p<0.01;

Tabla 14 Modelo de regresión multivariable dataset reducido y variable engagement.

Fuente: Elaboración propia

En el primer modelo, resultan interesantes las siguientes variables ya que son significativas explicando la variabilidad de la variable *engagement*:

- *Sponsored*: cuando el *post* es patrocinado obtiene menos *engagement* frente a los *posts* no patrocinados.
- *Lunes*: en este caso, nos está indicando que cuando el *post* se comparte un lunes frente a un *post* compartido un domingo el *engagement* conseguido es menor.
- *Miércoles*: en este caso, nos está indicando que cuando el *post* se comparte un miércoles frente a un *post* compartido un domingo el *engagement* conseguido es menor.
- *Jueves*: en este caso, nos está indicando que cuando el *post* se comparte un jueves frente a un *post* compartido un domingo el *engagement* conseguido es menor.
- *Verano*: las publicaciones compartidas en verano reciben una mejor aceptación frente a las publicadas en invierno.
- *Noche*: según los datos las publicaciones subidas entre las 04h y las 12h consiguen un mayor *engagement* frente a las subidas entre las 22h y las 04h.
- *Hay_etiquetas*: en los *posts* donde hay etiquetas se consigue más *engagement* frente a los *posts* donde no hay etiquetas.
- *Hay_hashtags*: la utilización de hashtags parece disminuir el *engagement* conseguido en el *post*.
- *Hay_marca*: los *posts* donde se cita alguna marca reconocida consiguen un mayor *engagement* frente a los *posts* donde no se cita ninguna marca.

Al incluir las variables extraídas de las imágenes, las variables influyentes del modelo anterior se mantienen y, además, observamos que la variable *faces* es significativa, por lo que podría darse un mayor *engagement* cuantas más caras aparezcan en la imagen. Por otro lado, *is_outdoor* también es significativa. Esto quiere decir que, si la imagen publicada es tomada en un espacio cerrado, frente a un espacio al aire libre se obtiene más *engagement*.

Es destacable que en el tercer modelo todas las variables incluidas son significativas. Por un lado, los *posts* subidos por cuentas verificadas o tipo *business* son propensas a obtener un menor *engagement* frente a los *posts* subidos por cuentas que no son de este tipo. Además, encontramos que especificar la categoría del perfil con la que se está subiendo la imagen frente a no hacerlo proporciona un mayor *engagement* en las publicaciones generalmente.

También observamos que en este modelo la variable *noche* pierde su influencia, así como las variables extraídas de las imágenes. Sin embargo, la variable *hay_mencion* influye levemente en este modelo.

Al hacer la última inserción de variables, en el modelo 4, los resultados se mantienen prácticamente igual que en el modelo 3, aunque sí que encontramos que la variable *sigue_a_muchos* es significativa. Por tanto, esto quiere decir que un *post* subido por un usuario que siga a muchos perfiles recibirá más interacción frente a un *post* subido por un usuario que no sigue a muchos perfiles.

5. Discusión de resultados

En este capítulo se discutirán los resultados descritos en el apartado anterior, sacando información de interés para la realización de la propuesta de actuación.

Para una mejor interpretación de los resultados, así como para facilitarnos la extracción de características se ha realizado la Tabla 15, en la que se incluyen todas las variables que han sido significativas en alguno de los modelos.

Para diferenciar el nivel de significación de las variables se han utilizado dos tonos distintos. Por ello, cuando el coeficiente significativo es positivo con más de un 10 % de significatividad, la celda se pinta de verde intenso; en cambio, si el coeficiente es positivo pero con una significatividad entre el 5 y el 10 %, toma un color verde más suave. Por otro lado, cuando el coeficiente es negativo con más de un 10 % de significatividad la celda se pinta de un rojo intenso, y si el nivel de significatividad se sitúa entre el 5 y el 10 %, ese color rojo es más suave.

Pasando a interpretar los resultados, iremos comentando cada una de las variables y su respectivo comportamiento a lo largo de los 21 modelos realizados, destacando aquellos puntos que sean relevantes.

En el caso de la variable *sponsored*, no parece tener relevancia hasta el estudio de la variable dependiente *engagement*. Esto puede deberse a que al ver la totalidad de interacción frente a los seguidores de la cuenta que ha subido el *post* se identifique un empeoramiento de la interacción. Esto quiere decir que un *post* pagado frente a uno que no está pagado teniendo el mismo alcance consigue peores resultados, lo que confirma que los usuarios interaccionan más con posts no esponsorizados.

Siguiendo con las variables del tipo de contenido que se comparte, observamos que en los casos que *Graphsidecar* es significativo al 10 % y además de forma positiva. Esto se explica porque, al subir un grupo de *n* imágenes o videos, se puede mostrar a los usuarios de la plataforma *n* veces el mismo *post* en ocasiones diferentes, por lo que las posibilidades de visualización se multiplican *n* veces. Sin embargo, la variable *GraphVideo* es un poco más contradictoria en este aspecto. A pesar de tener una buena acogida en cuanto a *likes* se refiere, e incluso comentarios. Cuando pasamos a analizar la variable *engagement* este tipo de contenido parece funcionar peor frente a las imágenes. Uno de los factores que podría estar influyendo en esta ocasión es la duración de los videos, ya que el hecho de que sean muy largos podría hacer que decrementara la interacción. Si bien para dar respuesta al comportamiento de esta variable independiente debería hacerse un estudio más a fondo, en el que sería interesante analizar el contenido del propio video.

	Likes							Comentarios							Engagement						
	N= 23696			N= 13446				N= 23696			N= 13446				N= 23696			N= 13446			
Modelos ->	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
esposorizado															-	-	-	-	-	-	-
GraphVideo	+	+						+		-					-	-	-				
Graphsidecar								+							+	+	+				
Lunes																		-	-	-	-
Miércoles																		-	-	-	-
Jueves																		-	-	-	-
Verano	+	+	+	+	+	+	+	+			+	+	+		+			+	+	+	+
Otoño			-																		
tarde	+	+	+	+	+	+	+				+	+	+	+	+	+	+				
noche	-		+				+	-										-	-		
hay_mencion		-				-		+	+	+	+	+	+								+
hay_etiquetas	-	-		-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hay_hashtags	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
hay_marca		-													+	+	+	+	+		
caras																			+		
Es_outdoor					-	-	-								-	-	-			-	-
verificada		+	+			+	+		+	+		+	+							-	-
business		-							+	+		+	+			-	-			-	-
Creators_and_Celebrities		+				+				-						+	+			+	+
Publishers		+				+										+	+			+	
General_interest		+				+										+	+			+	+
Non_profits		+				+				-						+	+			+	+
nano			-				-			-											
micro			-				-			-											
sigue_a_muchos			-							-											+

	Coeficiente positivo con una significatividad entre el 5 y el 10 %
	Coeficiente positivo con una significatividad mayor de un 10 %
	Coeficiente negativo con una significatividad entre el 5 y el 10 %
	Coeficiente negativo con una significatividad mayor de un 10 %

Tabla 15 Resumen de las variables significativas en todos los modelos de regresión multivariable.
Fuente: Elaboración propia.

Analizando las franjas temporales donde se comparte el contenido analizado hemos conseguido extraer conclusiones robustas. En lo que a los días de la semana se refiere, que se trata de la línea temporal que menos parece influir, se refleja que los posts subidos los lunes, miércoles o jueves tienen un peor *engagement* respecto a los posts subidos en domingo. Esto es interesante ya que, confirma que conforme más tiempo libre existe, más interacción hay con el contenido publicado. Lo confirmamos además con el siguiente rango temporal estudiado, ya que, en prácticamente todos los modelos se nos indica que el contenido publicado en verano se posiciona mucho mejor respecto al contenido publicado del invierno, y el verano incluye la época del año donde se concentran normalmente las vacaciones y, en consecuencia, donde más tiempo libre suelen tener las personas. Por último, en lo que a franjas temporales se refiere, el contenido publicado entre las 13h y las 21h es el que mejor se posiciona, siendo esta franja horaria donde se concentran las horas libres de la mayoría de las personas. Sin embargo, en el caso de la franja horaria noche, que incluye de 22h a 04h, los resultados no son tan concluyentes. Podría estar influyendo el ámbito territorial, ya que, por ejemplo, en España nos acostamos más tarde que en el norte de Europa, lo que podría estar sesgando los resultados. También se puede asumir que las primeras horas de esta franja horaria la actividad de la plataforma sea mayor que en altas horas de la madrugada, por lo que se podrían estar dando estas contradicciones. En cualquier caso, según los datos, será mejor publicar por la tarde.

Las siguientes variables de interés son *hay_mencion* y *hay_etiquetas*, las dos se comportan de manera similar y aunque en principio no presentan una buena acogida a nivel de *likes* sí lo hace a nivel de *comentarios*. Al fin y al cabo, cuando mencionamos o etiquetamos a cualquier perfil es más probable que esos perfiles interactúen con la publicación. Si bien en los primeros modelos no parecen funcionar positivamente estas variables, mientras no se abuse de ellas pueden ser beneficiosas para aumentar el *engagement*.

La variable *hay_hashtags* es sin duda la más perjudicial contra todo pronóstico, ya que en los 21 modelos nos indica que es significativa en un 10 % de manera negativa. Esto es relevante ya que normalmente el uso de *hashtags* ayuda a posicionar el contenido para conseguir mayores visualizaciones, si bien, el abuso de ellos puede conseguir el efecto contrario. Aunque sería necesario indagar un poco más para saber porque está influyendo de manera negativa el uso de *hashtags*, por ejemplo, cuantificando los *hashtags* para saber si está abusando de ellos o no, no debería penalizarse el contenido publicado haciendo un buen uso de los *hashtags*. Remitiéndonos a los resultados obtenidos, será mejor utilizar los *hashtags* estrictamente necesarios.

En el caso de la variable, *hay_marca*, a pesar de que en el segundo modelo nos indica que influye de manera negativa, es más relevante cómo se comporta en los modelos que estudian la variable dependiente *engagement*. En este caso su influencia es positiva. Esto se explica por qué al citar a marcas reconocidas en el contenido publicado nuestro alcance se va a incrementar, ya que usuarios que sigan de cerca esas marcas tendrán posibilidad de encontrar nuestro contenido, por lo que la interacción conseguida será mayor que la que podríamos tener cuando el alcance se conforma solamente con los propios seguidores del perfil que publica el *post*.

El siguiente grupo de variables a comentar es uno de los más interesantes del proyecto ya que son variables extraídas del contenido que percibe el usuario visualmente. Y ambas han resultado ser significativas. La más significativa ha sido, *is_outdoor*, ya que en todas los modelos que se ha utilizado ha sido significativa al 10 % de manera negativa. Esto quiere decir que aquellas imágenes no tomadas al aire libre tienen una mejor acogida entre los usuarios, por lo que será beneficioso compartir contenido de estas características. Estos resultados son coherentes con relación a la aparición de carreras virtuales y la predisposición de los nuevos runners a participar en estos eventos (Rizzo, 2021). Si bien, sería interesante profundizar en el tipo de contenido que se comparte en espacios cerrados para afinar las propuestas de actuación. Por otro lado, la variable *faces*, que nos indica el número de caras que se encuentran en la imagen, parece no influir demasiado hasta que llegamos al estudio de la variable dependiente *engagement* donde se ve que se puede conseguir un mayor *engagement* cuantas más caras aparezcan en la imagen.

Pasando a las variables a nivel de usuario, encontramos contradicciones en varias de ellas. Comentando por encima *verified*, ya que la verificación de una cuenta depende de la plataforma de Instagram, vemos que en un principio es significativamente positiva pero cuando estudiamos el *engagement* pasa a ser significativamente negativa. A pesar de este resultado, consideramos que siempre será mejor contar con una cuenta verificada ya que esto aporta fiabilidad y veracidad al perfil, lo que son características positivas. Por otro lado, *business* parece comportarse de manera similar. En este caso sería necesario verificar si los beneficios de tener una cuenta *business* nos afectan, teniendo en cuenta que las propuestas van dirigidas a empresas, es evidente que será beneficioso tener este tipo de cuenta.

A continuación, revisamos las variables que categorizan la cuenta donde se publica el contenido, y en general, será positivo tener la cuenta categorizada a no especificarlo, ya que todas las categorías se comportan igual frente a no especificar este dato.

Por último, el subgrupo de variables específicas para clasificar una cuenta en base al número de seguidores que se tiene confirma que cuantos más seguidores tenga una cuenta más interacción conseguirá por lo que se concluye que conseguir seguidores debe estar siempre presente en las estrategias utilizadas en social media. Por otro lado, la variable *sigue_a_muchos* no tiene un significado muy relevante en el estudio y dado que los resultados obtenidos no son concluyentes no se va a utilizar esta variable para dar recomendaciones.

Siendo conscientes de las limitaciones que hemos asumido en nuestro estudio, los resultados obtenidos aportan información relevante a las empresas. Si bien, de cara a la publicación de un artículo, se pretende hacer un análisis más riguroso estudiando cada una de las variables con un modelo ANOVA. Además, durante el presente proyecto ya se ha trabajado en la generación de nuevas variables para enriquecer el estudio y seguir mejorándolo progresivamente profundizando en el objeto de estudio.

Después analizar la información extraída de los 21 modelos e identificar cual es relevante, vamos a realizar las propuestas de actuación pertinentes en el siguiente capítulo.

6.Propuesta de actuación

Una vez analizados todos los datos procedemos a realizar la propuesta de actuación para que las empresas del sector runner consigan un crecimiento orgánico en Instagram:

- I. **No abusar de publicaciones patrocinadas:** según hemos comprado estos *posts* no parecen funcionar mejor que los *posts* sin patrocinar, por lo que, se recomienda que antes de hacer un *post* pagado, regalen productos a sus colaboradores y ellos libremente hablen sobre la marca en Instagram.
- II. **Priorizar contenido tipo Sidecar:** este contenido va a aumentar el número de visualizaciones por publicación que se hagan, por lo que debería incorporarse de forma habitual a la hora de compartir contenido. A pesar de que los resultados con el tipo de contenido video no han sido tan concluyentes se recomienda su uso **insertando videos de corta duración**, ya que puede resultar beneficioso.
- III. **Priorizar la publicación de contenido entre las 13h y las 21h:** Como hemos visto, las publicaciones subidas en esta franja horaria son las que más interacción consiguen por lo que debería planificarse el contenido de manera que los *posts* se publicaran entre estas horas.
- IV. **Potenciar campañas para captar clientes en verano o periodos vacacionales:** Tal y como hemos comentado, el contenido tiene una mayor acogida en verano ya que, en general, se tiene más tiempo libre. Por eso, debería utilizarse este periodo para lanzar las campañas de publicidad más potentes, ya que la probabilidad de que tengan éxito es mucho mayor. También podrían valer periodos vacacionales como navidad.
- V. **Priorizar la publicación en días diferentes a Lunes, Miércoles y Jueves:** según los datos estudiados será mejor compartir publicaciones en uno de los días que no se mencionan anteriormente, los fines de semana son un buen momento para compartir el contenido más importante.
- VI. **Publicar contenido donde aparezcan rostros frecuentemente:** se ha comprobado que, según los datos estudiados, esta característica favorece a la interacción con el contenido, por lo que publicar contenido donde aparezcan rostros frecuentemente será beneficioso para aumentar el *engagement*.
- VII. **Publicar contenido donde aparezcan espacios cerrados frecuentemente:** se ha comprobado que, según los datos estudiados, esta característica favorece a la interacción con el contenido, por lo que publicar contenido de fotografías tomadas en espacios cerrados frecuentemente será beneficioso para aumentar el *engagement*.

- VIII. No abusar del uso de Hashtags:** Como se ha observado el mal uso de hashtags puede ser realmente perjudicial, por lo que se recomienda hacer uso de los hashtags estrictamente necesarios. Además, sería interesante proponer al equipo de social media y los influencers que trabajen con la marca un curso de especialización o *MasterClass* del uso de hashtags para asegurarse de que todo el equipo sabe utilizarlos correctamente, así como tener un registro con los *hashtags* que se utilizan en cada publicación para poder saber cuáles de ellos funcionan mejor en el sector de interés.
- IX. Potenciar la citación de marcas deportivas:** Como hemos comentado, esta acción puede incrementar el *engagement* de las publicaciones de manera orgánica por lo que se debe potenciar su uso.
- X. Tener el perfil de Instagram categorizado acorde con el público objetivo de interés:** se ha visto que es mejor indicar la categoría a la que pertenece la cuenta donde se publica que no indicarlo, por lo que debemos asegurarnos de que las cuentas con la que se trabaje estén correctamente categorizadas.
- XI. Lanzamiento de sorteos para aumentar el número de seguidores:** Como hemos visto, es realmente importante el número de seguidores que tiene la cuenta donde se vaya a publicar el contenido por lo que este factor debe aumentar. Una manera sencilla es el lanzamiento de concursos con el requisito de que deben seguir a la cuenta. Cuanto más importante sea el sorteo más gente participará, por lo que se deberían planificarse varios sorteos para reclutar seguidores, ya que la consecuencia directa será la mejora de nuestro *engagement*.

7. Conclusiones

En un inicio se propuso abordar el mundo de las redes sociales, en concreto el concepto de *engagement*, durante la trayectoria del proyecto se ha profundizado en este concepto y al mismo tiempo hemos conseguido traducir este concepto para poder trabajar con él en modelos estadísticos. Para ello identificamos a 149 influencers del sector del *running* en Instagram y construimos una base de datos con 23.696 *posts* y un gran número de características relacionadas con este concepto, cumpliendo así con el objetivo de recopilar datos de la propia *app* de Instagram para la investigación. Después de la extracción masiva de datos concluimos que es posible aplicar las técnicas utilizadas para cualquier área o dominio.

Una vez conseguidos los datos trabajamos con ellos en el preprocesado y manipulaciones correspondientes para poder trabajar con ellos, se trata de un proceso muy importante cuando se trabaja con una gran cantidad de datos ya que un mal preprocesado de datos podría llevarnos a conclusiones erróneas. Al final del preprocesado tuvimos dos corpus preparados para su análisis, uno con 23.696 *posts* y una submuestra con 13.446 *posts*.

A su vez, se ha cumplido con el objetivo de crear un modelo de redes neuronales con el propósito de extraer información de las propias imágenes. Por un lado, podemos concluir que el uso de redes neuronales convolucionales es una buena técnica para la extracción de características en imágenes ya que se ha conseguido una precisión elevada (83,55%). Por tanto, concluimos que estas técnicas podrían ser utilizadas para ayudar a la toma de decisiones en estrategias de marketing en redes sociales.

Una vez se ha recopilado toda la información, hemos procedido a analizarla a través de 21 modelos estadísticos jerárquicos de regresión multivariable, lo que nos ha permitido analizar el contenido y el comportamiento de las variables *likes*, comentarios y *engagement*. A partir de dicho análisis, hemos conseguido extraer conclusiones relevantes para realizar una propuesta de actuación con 11 recomendaciones para que las empresas del sector runner consigan un crecimiento orgánico trabajando con influencers, así como un mayor *engagement*.

Por todo ello consideramos este trabajo una buena aproximación inicial tanto a la investigación del *engagement* enfocada al sector del *running* como a la propuesta de recomendaciones a empresas de este sector que trabajan con influencers.

8. Bibliografía

- Affde. (05 de Octubre de 2021). *Principales influencers del running en las redes sociales*. Obtenido de affde: <https://www.affde.com/es/top-running-influencers.html>
- Alhabash, S., McAlister, A., & Hagerstrom, A. (2013). Between Likes and Shares: Effects of Emotional Appeal and Virality on the Persuasiveness of Anticyberbullying Messages on Facebook. *Cyberpsychology, Behavior, and Social Networking*, 16(3), 175-182.
- Armas, E. (21 de Febrero de 2022). *12 joyas 'beauty' que se hicieron virales gracias a la magia de TikTok*. Obtenido de smoda: <https://smoda.elpais.com/belleza/12-joyas-beauty-que-se-hicieron-virales-gracias-a-la-magia-de-tiktok/100474610/image/100474612>
- Beautiful Destinations. (2022). *Beautiful Destinations*. Obtenido de Beautiful Destinations: <https://beautifuldestinations.com>
- Blakeman, R. (2007). *Integrated Marketing Communication: Creative Strategy from Idea to Implementation*. Rowman & Littlefield Publishers.
- Bosch Rue, A., Casas Roma, J., & Lozano Bagen, T. (2019). *Deep learning : principios y fundamentos*. Barcelona: Editorial UOC.
- Constine, J. (2018). *Instagram hits 1 billion monthly users, up from 800M in September*. Obtenido de Tech Crunch: <https://techcrunch.com/2018/06/20/instagram-1-billion-users/?guccounter=1>
- Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior: A framework for engaging customers through social media content. *European Journal of Marketing*, 53(10), 2213-2243.
- El comercio. (14 de Noviembre de 2018). *Los 6 runners más influyentes en Instagram*. Obtenido de el comercio: <https://elcomercio.pe/deporte-total/running/guia-runner/fotos-6-runners-influyentes-instagram-noticia-577183-noticia/>
- Feedspot Media Database Team. (9 de Diciembre de 2020). *Top 150 Running Instagram Influencers most followed*. Obtenido de Feedspot: https://blog.feedspot.com/running_instagram_influencers/
- Godfrey, G. (s.f.). *10 global running influencers you should know*. Obtenido de fashion monitor: <https://www.fashionmonitor.com/blog/wD/10-global-running-influencers-you-should-know>
- InBeat. (2021). *InBeat*. Obtenido de <https://www.inbeat.co>
- Instagram. (31 de Agosto de 2022a). *Perfil de Instagram corroysoymujer*. Obtenido de Instagram: <https://www.instagram.com/corroysoymujer>
- Instagram. (31 de Agosto de 2022b). *Post del perfil de Instagram corroysoymujer*. Obtenido de Instagram: <https://www.instagram.com/corroysoymujer>

- Jang, J. Y., Han, K., Shith, P., & Lee, D. (2015). Generation Like: Comparative Characteristics in Instagram. *Proceedings of the ACM Conference on Human Factors in Computing Systems*. Seúl.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Obtenido de <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- Krueger, A. (21 de Julio de 2016). *This Company Can Predict The Number Of Likes An Instagram Photo Will Get*. Obtenido de forbes: <https://www.forbes.com/sites/alysonkrueger/2016/07/21/this-company-can-predict-the-number-of-likes-an-instagram-photo-will-get/?sh=e8cc55f5a47b>
- Libertad Digital. (31 de Agosto de 2016). *Los corredores más influyentes del running en España*. Obtenido de libertad digital: <https://www.libertaddigital.com/deportes/mas-deporte/2016-08-31/los-corredores-mas-influyentes-del-running-en-espana-1276575752/>
- LikeyAI. (2017). *LikelyAI*. Obtenido de LikelyAI: <https://www.likelyai.com>
- Lloyd, S. (2016). In-Depth Look at a Career in Brand Management. (S. C. Roberts, Entrevistador)
- Manovich, L. (2016). *Designing and Living Instagram Photography: Themes, Feeds, Sequences, Branding, Faces, Bodies*. Obtenido de manovich: <http://manovich.net/index.php/projects/designing-and-living-instagram-photography>
- metricool. (2022). *metricool*. Obtenido de metricool: <https://metricool.com/es/>
- Nadian-Ghomsheh, A. (2017). *Indoor-outdoor image classification using dichromatic reflection model and Haralick features*. Obtenido de figshare: https://figshare.com/articles/dataset/Indoor-Outdoor_dataset/4595323/1
- Pastor, C. M. (s.f.). *8 cuentas de runners en Instagram que deberías estar siguiendo*. Obtenido de carreras populares: <https://www.carreraspopulares.com/noticia/8-cuentas-de-runners-en-instagram-que-deberias-estar-siguiendo>
- Pintado, T., & Sanchez Herrera, J. (2017). *Nuevas tendencias en comunicación estratégica*. Madrid: ESIC.
- Primo Aparici, A. (15 de Julio de 2022). Predicción del engagement en Instagram en el mundo del runnig. Valencia, Valencia, España.
- Redacción PuroMarketing. (22 de Junio de 2021). *En qué redes sociales se están centrando las empresas en su estrategia de Marketing con influencers*. Obtenido de puromarketing: <https://www.puromarketing.com/42/35446/redes-sociales-estan-centrando-empresas-estrategia-marketing-influencers>
- Rizzo, N. (21 de Septiembre de 2021). *Running Boom: 28.76% of runners started during the pandemic*. Obtenido de Run Repeat: <https://runrepeat.com/new-pandemic-runners>
- Roldán, A. (13 de Septiembre de 2018). *Las instagramers más influyentes del universo running en España*. Obtenido de runnea: <https://www.runnea.com/articulos/running-news/2018/09/instagramers-influyentes-universo-running-espana-3805/>

- Runnea. (14 de Octubre de 2021). *Los 11 bloggers de running más influyentes de España*. Obtenido de runnea: <https://www.runnea.com/articulos/running-news/2017/03/bloggers-running-influyentes-espana-2506/>
- Schultz, D., Tannenbaum, S., & Lauterborn, R. (1993). *Integrated marketing Communications*. Lincolnwood: NTC Business Books.
- Sicilia, M., Palazon, M., Lopez, I., & Lopez, M. (2021). *Marketing en redes sociales*. Madrid: Esic Editorial.
- Smith, C. (17 de Agosto de 2014). *Here's Why Instagram's Demographics Are So Attractive To Brands*. Obtenido de Business Insider: <https://www.businessinsider.com/instagram-demographics-2013-12>
- Statista Research Department. (25 de Agosto de 2022). *Porcentaje de individuos que practicó running España 1996-2021*. Obtenido de es.statista: <https://es.statista.com/estadisticas/569559/evolucion-del-porcentaje-de-individuos-que-practicaron-running/>
- Ungria, C. C. (28 de Octubre de 2018). *Los runners españoles más influyentes en Instagram*. Obtenido de trecebits: <https://www.trecebits.com/2018/10/28/los-runners-espanoles-mas-influyentes-en-instagram/>
- Van Doorn, J. L. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, 13(3), 253-266.
- Wells, W., Burnett, J., & Moriarty, S. (2003). *Advertising: Principles & Practice*. Prentice Hall.
- World Athletics. (2 de Junio de 2021). *New research reveals running boom during Covid-19 pandemic*. Obtenido de World Athletics: <https://www.worldathletics.org/be-active/lifestyle/global-running-day-research-nielsen>

9. Anexos

Anexo 1. Relación del TFG con los ODS

ANEXO

OBJETIVOS DE DESARROLLO SOSTENIBLE



Reflexión sobre la relación del TFG con los ODS en general y con el/los ODS más relacionados.

Los Objetivos de Desarrollo Sostenible aprobados por la Organización de las Naciones Unidas se pueden vincular al presente proyecto. Esto se debe a que se ha planteado el proyecto desde un punto de vista innovador.

En primer lugar, este TFG está relacionado con el trabajo decente y crecimiento económico (ODS.8). Con la investigación que hemos llevado a cabo y el resultado conseguido, estos análisis jerárquicos de regresión multivariable podrían ayudar a las empresas a tomar decisiones adecuadas en sus estrategias de marketing en Instagram. De esta manera podrían incrementar su *engagement* y en consecuencia sus ventas, lo que contribuiría al crecimiento económico. Además, las técnicas utilizadas para la extracción de la información podrían extrapolarse a otras áreas, por lo que podría ayudar a empresas de cualquier sector.

Con una relación menos directa, este proyecto se ha desarrollado tras un dominio que se relaciona directamente con la salud y el bienestar (ODS.3), el *running*. Este grupo social realiza una actividad que presenta grandes beneficios para la salud y aunque, este proyecto se ha centrado en el análisis de información recolectada con objetivo de entender el comportamiento del *engagement* de los *posts*, el dataset construido es un material que podría ser válido para estudios relacionados con la salud. Además, dicho dataset se ha almacenado en la nube y puede ser utilizado por cualquier persona sin necesidad de gastar memoria en su disco duro, este hecho relaciona nuestro proyecto con el compromiso de consumo responsable (ODS.12).

Otro de los objetivos que está presente en el proyecto es el de industria, innovación e infraestructura (ODS.9) ya que se han construido dos corpus específicos para nuestro trabajo y no hay evidencia de que se haya investigado el concepto de *engagement* aplicando las técnicas de tratamiento de imágenes con redes neuronales para la extracción de características desde las propias imágenes.

Por último, cabe mencionar una pequeña relación con la reducción de las desigualdades (ODS.10) ya que en el estudio se han utilizado datos de toda la población sin importar la nacionalidad, género o cultura de los perfiles seleccionados.

Anexo 2. Lista de los 149 influencers seleccionados

num	username	followers	followees	business	verified	category
1	heatherunz	92464	521	0	0	
2	lucy_bartholomew	105003	96	0	1	
3	sarahall3	167841	432	1	1	Creators & Celebrities
4	stephrothstein	96182	762	1	1	Creators & Celebrities
5	trackclubbabe	128816	283	0	0	
6	ian.morgan	121772	2330	0	0	
7	foodfitnessflora	114672	1221	0	1	
8	des_linden	181959	1155	1	1	Creators & Celebrities
9	runrx	211008	100	1	0	Personal Goods & General Merchandise Stores
10	karagoucher	188103	502	1	1	Creators & Celebrities
11	jennifalconer	247719	6263	1	1	Creators & Celebrities
12	tommy_rivs	327573	1077	1	1	Creators & Celebrities
13	courtneydauwalter	337598	931	1	1	Creators & Celebrities
14	shalaneflanagan	420429	407	1	1	Creators & Celebrities
15	corroysoymujer	32343	2669	0	0	
16	pitufollow	92309	963	0	1	
17	martaggrun	38491	3501	0	0	
18	bielrafols	20143	987	1	0	General Interest
19	runnea	23442	2617	1	0	Publishers
20	onmytrainingshoes	88081	1020	0	1	
21	maratonfiz	34554	4148	0	1	
22	deparunner	37328	956	0	1	
23	macias1500	23492	1750	0	1	
24	missleggingsrun	91618	633	0	1	
25	thebeautymail	234898	1757	1	1	Personal Goods & General Merchandise Stores
26	trimariapujol	14099	884	0	0	
27	apeleteiob	323584	1236	0	1	
28	mikewardian	40773	1960	1	0	Creators & Celebrities
29	Monitorthebeat	152720	95	1	0	Non-Profits & Religious

						Organizations
30	runsociety	11033	2765	1	0	Publishers
31	runningfervor	814481	461	1	0	Creators & Celebrities
32	ingridpino95	73714	1502	0	1	
33	saleta3	109229	1963	0	1	
34	raulgomez82	164170	912	0	1	
35	valentisanjuan	641492	1390	1	1	Creators & Celebrities
36	theMarathonMarcu s	22171	3466	0	0	
37	TheMirnavator	141974	6923	1	1	Creators & Celebrities
38	UltraMarathon	94351	4173	1	1	Creators & Celebrities
39	HungryRunnerGirl	50675	884	0	0	
40	therunner	10174	7489	1	0	Creators & Celebrities
41	spartanchickie	27387	3243	0	0	
42	mamabear_runs_m arathons	43913	770	0	0	
43	GaryRobbins	86569	2334	0	1	
44	halfcrazymama	22644	2581	0	0	
45	inesmelchor86	13494	108	0	0	
46	milindrunning	1161754	26	0	1	
47	gomofarah	1145930	495	1	1	Creators & Celebrities
48	usainbolt	11112558	317	1	1	Creators & Celebrities
49	ali_mihaela	15621	1936	0	0	
50	marianoont	65685	4306	0	1	
51	noemimj	88140	2017	0	1	
52	samarbau	47371	451	0	0	
53	chemitamartinez	190811	673	1	1	Creators & Celebrities
54	yellowrunner	91366	2118	0	0	
55	Jgomeznoya	312750	309	0	1	
56	carles_Castillejo	35520	201	0	1	
57	mais_runs_trails	11861	724	0	0	
58	omnomnomnomp anda	10213	7122	0	0	
59	nycrunningmama	72613	431	0	1	
60	pittmanfit	34524	1679	0	0	
61	crystalseaver	42706	2734	1	0	Creators & Celebrities
62	Runtothefinish	67158	631	1	1	Creators & Celebrities
63	therunnerbeans	71732	2866	0	1	
64	chrissy.nich.runs	12123	1683	0	0	

65	ncrunnerjacky	33024	4001	0	1	
66	runnylegs	26196	973	0	0	
67	ironwillironjill	28312	1527	0	0	
68	martha_runs	11130	524	1	0	Publishers
69	ultra_damo	22544	372	0	0	
70	tattedrunnergirl	11081	860	0	0	
71	molly_runs_365	11756	1068	0	0	
72	rushbynature	18968	1907	0	0	
73	ultra_sophie	22783	756	0	0	
74	fudgieruns	17588	7444	0	0	
75	run_rhandi_run	15169	665	0	0	
76	gofastmommy	18454	1947	1	0	Creators & Celebrities
77	ruth_runs	20554	3657	0	0	
78	runningmummioof4	20353	3297	1	0	General Interest
79	framefitnesssouthampton	20532	7486	1	0	Creators & Celebrities
80	hayleycarruthers26.2	15559	1383	0	0	
81	heartnsolerunning	16193	1891	0	0	
82	meretherunner14	16068	1362	1	0	Creators & Celebrities
83	little.runner.mel	19976	2083	0	0	
84	gregexploring	16260	689	0	0	
85	tashi_skervinclarke	18052	1294	0	0	
86	themarathonmarcus	22166	3464	0	0	
87	shortgirlrunner	26369	295	0	0	
88	thisbunnyruns	22180	2592	0	0	