



Universidad Pontificia Comillas

ENGAGEMENT CON EL CONTENIDO DE LAS MARCAS EN SOCIAL MEDIA

Autor: Cecilia Delgado Villanueva
Director: Carmen Valor Martínez

MADRID | Junio 2026

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Cecilia Delgado Villanueva estudiante de E3-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Engagement con el contenido de las marcas en social media", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
11. Generador de datos sintéticos de prueba: Para la creación de conjuntos de datos ficticios.

12. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.

13. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

14. Generador de encuestas: Para diseñar cuestionarios preliminares.

15. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 15 junio 2026

Firma:

A handwritten signature in black ink, consisting of a series of loops and a long horizontal stroke at the end.

ÍNDICE DE CONTENIDO

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado 2

1. INTRODUCCIÓN	6
1.1. Contexto general	6
1.2. Planteamiento del problema teórico	6
1.3. Objetivos y pregunta de investigación	7
1.4. Contribución esperada.....	8
2. MARCO CONCEPTUAL	8
2.1. Antecedentes del engagement: revisión de la literatura	8
2.2. Determinantes del engagement: revisión de literatura	9
2.3. Caso de estudio: Atlético de Madrid	12
3. METODOLOGÍA	13
3.1. Análisis de contenido.....	13
3.2. Selección muestral.....	14
3.3. Variables dependientes e independientes	17
3.4. Codificación y análisis de los datos	22
4. RESULTADOS	24
4.1. Descripción de la muestra	24
4.2. Engagement según formato.....	26
4.3. Engagement según temática	27
4.4. Engagement según presencia de jugadores o afición	28
4.5. Engagement según contexto deportivo	29
4.6. Engagement según finalidad de contenido	32
4.7. Características de textos y variables temporales	32
4.8. Análisis comparativo.....	33
5. CONCLUSIONES	36
5.1. Principales hallazgos	36
5.2. Implicaciones del estudio	37
5.3. Limitaciones del estudio	42
5.4. Futuras líneas de investigación.....	43
6. REFERENCIAS BIBLIOGRÁFICAS	44
Anexo I	48
A.I.1. Partido	48
A.I.2. Celebración	49
A.I.3. Entrenamiento	50
A.I.4. Jugadores y plantilla	51

A.I.5. Fichaje y renovación	52
A.I.6. Institucional.....	53
A.I.7. Merchandising y patrocinio	54
A.I.8. Afición y comunidad	55
A.I.9. Otra	56
Anexo II	57
A.II.1. Descarga de imágenes del corpus	57
A.II.2. Codificación del corpus	58
A.II.3. Generación del corpus final	59
A.II.4. Análisis descriptivo	61
A.II.5. Análisis bivariado	61
A.II.6. Comprobación de multicolinealidad	62
A.II.7. Regresión lineal múltiple	64
A.II.8. Generación de las figuras	66

ÍNDICE DE TABLAS

Tabla 1. Variables analizadas	11
Tabla 2. Resultados de la validación intercodificador mediante el coeficiente alpha de Krippendorff	17
Tabla 3. Variables Independientes.....	18
Tabla 4. Variables de control	22
Tabla 5. Factor de inflación de la varianza (VIF)	23
Tabla 6. Métricas de engagement (N=2.244)	24
Tabla 7. Resultados del análisis bivariado entre las variables independientes y el engagement	26
Tabla 8. Engagement según la temática de la publicación.....	27
Tabla 9. Resultados del análisis bivariado (prueba de Kruskal-Wallis) para las variables de presencia de personas.....	28
Tabla 10. Engagement según la presencia de jugadores y de afición.....	29
Tabla 11. Engagement según el contexto deportivo	29
Tabla 12. Engagement según la finalidad del contenido	32
Tabla 13. Coeficientes del modelo de regresión lineal múltiple sobre el engagement (log)	34
Tabla 14. Decálogo de recomendaciones aplicadas para la gestión del engagement en la cuenta de Instagram del Atlético de Madrid	39

ÍNDICE DE FIGURAS

Figura 1. Distribución del engagement antes y después de la transformación logarítmica (N = 2.244).	25
Figura 2. Engagement mediano según contexto deportivo (N = 2.244). En azul se destacan los contextos que superan la mediana global del corpus y en verde los que se sitúan por debajo.	30
Figura 3. Engagement activo (comentarios) frente a pasivo (likes) por contexto deportivo, expresado como porcentaje de comentarios sobre likes (N = 2.244). El color naranja destaca el caso de las publicaciones post-derrota, claramente diferenciado del resto.	31
Figura 4. Efectos significativos del modelo de regresión múltiple sobre el engagement (N = 2.244). Los porcentajes representan el cambio en el engagement respecto a la categoría de referencia de cada variable. *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$	35

1. INTRODUCCIÓN

1.1. Contexto general

Diariamente, las redes sociales generan contenido que llega a millones de espectadores. Aplicaciones como Instagram, TikTok, X o Facebook se han convertido en el principal canal de comunicación de las marcas para interactuar con su público (Enríquez, 2023). El objetivo de todas estas marcas es poder conseguir un mayor engagement, un concepto que se manifiesta a través de la interacción de los usuarios mediante comentarios o likes, y que actúa como indicador del éxito de sus publicaciones o contenido (Schivinski et al., 2016; Trunfio et al., 2021).

Es por ello que las marcas invierten muchos de sus recursos en el contenido que publican en sus redes sociales para asegurarse de que los usuarios interactúen con él. Sin embargo, no todo el contenido genera altos niveles de engagement, esto ocurre porque las variables que caracterizan cada publicación tienden a ser muy diversas y aún no se ha podido determinar cuáles de ellas se relacionan con mayores niveles de interacción (Barger et al., 2016). La literatura previa señala que variables como el tipo de contenido (informativo o promocional), el tono emocional, la presencia de personas, la frecuencia de publicación y el momento en que se publica se asocian de forma relativamente estable con un mayor engagement (Schreiner et al., 2019). Pero también existen otras variables como el tono emocional del mensaje, el uso de hashtags, la longitud del texto o las llamadas a la acción, que se asocian con resultados mixtos o directamente contradictorios entre estudios. A esto se añade que los estudios que miden el comportamiento del engagement en cuentas deportivas, donde el componente emocional e identitario es central, son menos y por ello están menos explorados. Por ello, esta combinación de variables con evidencia consolidada, evidencia mixta y un sector aún poco estudiado justifica un análisis específico como el que se propone en este trabajo.

1.2. Planteamiento del problema teórico

Por tanto, la relación entre estas variables y el engagement que generan las publicaciones no ha sido probada debido a las diferencias entre las plataformas y culturas, además de la falta de estudios empíricos (Vander Schee et al., 2020). Por ende, es necesario que se siga estudiando

esta relación para poder determinar cuáles son los elementos que producen una mayor interacción de los usuarios.

A partir de esta idea, este trabajo utilizó como objeto de estudio el perfil oficial del Atlético de Madrid en Instagram. La elección de este club obedece a dos razones. En primer lugar, por ser una marca deportiva con una presencia digital consolidada, un alto número de seguidores y gran actividad diaria. En segundo lugar, sus publicaciones varían considerablemente, lo cual permite hacer uso de diferentes variables como formatos o temas.

Durante este trabajo, el análisis se centra en 2.249 publicaciones del perfil oficial del Atlético de Madrid en Instagram, en un periodo de seis meses, entre el 1 de agosto de 2025 y el 4 de mayo de 2026, para poder analizar la comunicación del club durante parte de la temporada en la que coinciden LaLiga y la fase de liga de la Champions. Las publicaciones giran en torno a partidos, resultados, alineaciones, entre otros. Dentro del corpus se incluyeron únicamente las publicaciones visibles del feed del perfil oficial, es decir, imágenes, carruseles y reels. Por el contrario, quedaron fuera del estudio las stories, los directos, es decir, todo el contenido efímero y aquellas publicaciones que hayan sido eliminadas en el momento de la recogida de datos. Esta delimitación permite eliminar todas aquellas publicaciones que impiden una comparación clara; usando únicamente las publicaciones permanentes, siendo así posible recoger de manera más fácil los niveles de interacción de los usuarios, como los me gusta y los comentarios.

Además, las stories y los directos, al funcionar de manera diferente, se eliminan al cabo de un determinado tiempo y no permiten la recogida de datos tan fácilmente. De esta forma, se construyó una base de datos más homogénea y adecuada para analizar las variables del contenido que producen un mayor engagement.

1.3. Objetivos y pregunta de investigación

Pregunta de investigación:

¿Qué variables o características del contenido en las publicaciones del Atlético de Madrid en Instagram influyen en un mayor nivel de engagement por parte de los usuarios?

Objetivo: Con este estudio se propone un análisis cuantitativo del contenido de las publicaciones del Atlético de Madrid. Basado en la literatura previa, se codifican, entre otras variables, aquellas relativas al tema, contexto o formato de la publicación. En particular, este trabajo pretende explorar cómo las emociones específicas asociadas al contenido y el contexto

deportivo inmediato (por ejemplo, victorias, derrotas o anuncios de fichajes) afectan a la interacción de los seguidores. De esta manera, se identifican aquellas variables que generan un mayor engagement, con el fin de reunir evidencias que permitan optimizar las interacciones de las publicaciones de este club de fútbol con sus espectadores.

1.4. Contribución esperada

De esta manera, se pueden identificar aquellas variables que generan un mayor engagement, con el fin de reunir evidencias que permitan optimizar las interacciones de las publicaciones de este club de fútbol con sus espectadores.

2. MARCO CONCEPTUAL

2.1. Antecedentes del engagement: revisión de la literatura

Los antecedentes del engagement se encuentran en la literatura de marketing relacional y de servicios, donde se identificó como una manera valiosa de mejorar el rendimiento de las marcas mediante la interacción con los clientes (Marvi et al., 2024). Bowden (2009) define el engagement como un proceso psicológico que modela los mecanismos subyacentes por los que la lealtad se forma para los nuevos clientes de una marca, mientras que Brodie et al. (2011) lo sitúan dentro de una lógica de interacción y cocreación de valor. A partir de aquí, se puede entender el engagement como fruto de una relación dinámica entre una marca y su público.

A medida que las redes sociales se han ido desarrollando, las marcas han ido evolucionando y mejorando su forma de comunicarse con los consumidores y generar ingresos a través de herramientas que les permiten interactuar en línea (Hollebeek et al., 2014). Así, el engagement ha adquirido mayor relevancia al hacer visible esa relación entre consumidores y marcas a través de interacciones observables (Schreiner et al., 2019). No obstante, la literatura posterior muestra que no existe una única forma de definir o medir el engagement, de hecho, algunos trabajos lo abordan desde dimensiones actitudinales y relacionales, mientras que otros lo observan a través de conductas visibles como la interacción con los contenidos publicados. Por eso, una parte importante de la investigación reciente se centra en cómo analizar el engagement de manera operativa en contextos digitales. Además, varios trabajos de investigación evidencian que los antecedentes del engagement no se comportan igual en todas las plataformas ni en todos los contextos, lo que obliga a no tratarlo como un fenómeno homogéneo.

En el contexto deportivo, para incrementar el consumo deportivo es importante entender las motivaciones de los aficionados (Stavros et al., 2014). En el contexto digital, las redes sociales son el canal que utilizan los aficionados para interactuar con su equipo. Es por eso que el engagement no se usa únicamente con un fin informativo, sino también identitario y emocional, Stavros et al. (2014) observaron que los aficionados interactúan en redes movidos por emociones como la pasión, la esperanza, la autoestima y la camaradería.

En el caso de Instagram, existen diferentes formatos como las Stories: publicaciones que desaparecen después de 24 horas, los Reels o las publicaciones. Es importante hacer esta distinción de formatos de la red social porque ofrece diferentes formas de visibilidad y consumo. Este trabajo se centra en las publicaciones del feed del Atlético de Madrid para analizar cómo determinadas características del contenido se asocian con mayores niveles de engagement.

2.2. Determinantes del engagement: revisión de literatura

El engagement en redes sociales no depende de un único factor, sino de la combinación de distintos elementos del contenido y del contexto en el que este se publica (Schreiner et al., 2019). Por eso, la literatura sobre comunicación digital y redes sociales ha prestado atención a las características de las publicaciones, con el objetivo de identificar qué variables se relacionan con una mayor interacción de los usuarios. Sin embargo, los estudios realizados mantienen que, al no ser todo el contenido igual y existir diferencias culturales y generacionales, no hay una relación clara entre qué variables pueden favorecer una respuesta más alta por parte de la audiencia, ya que depende del tipo de sector, tipo de contenido o incluso plataforma (Annamalai et al., 2021), por lo que se desaconseja extrapolar resultados sin atender al contexto específico de estudio.

Una de las variables más usadas en estudios de redes sociales es el formato de la publicación, por lo que los diferentes tipos de formato presentan diferentes niveles de interacción (Cvijikj y Michahelles, 2013). El algoritmo de Instagram no distribuye de la misma forma una imagen, un carrusel de imágenes o un reel, porque el contenido que se presenta al usuario es diferente en cada uno de ellos. Los reels, en tanto contenido audiovisual breve, tienden a generar mayor alcance según el algoritmo de Instagram y, según Liang y Wolfe (2022), también mayor engagement medio en likes y comentarios. Un carrusel puede alargar el tiempo de atención y ofrecer más información en una sola publicación. Una imagen única, en cambio, puede funcionar bien cuando el mensaje es muy directo. No obstante, en el contexto deportivo,

Annamalai et al. (2021) matizan estos hallazgos: en su estudio de 4.000 publicaciones de seis clubes, las fotografías generaban más likes y comentarios, mientras que los vídeos eran los más compartidos, lo que evidencia que el formato tiene un efecto importante en este estudio.

La temática y finalidad del contenido son dos variables complementarias. En Instagram, hay un sinfín de marcas cuyos contenidos difieren ampliamente. Mientras que la temática responde a "de qué trata" la publicación, en una cuenta como la del Atlético de Madrid, no es lo mismo publicar un resultado, una imagen de entrenamiento o un fichaje. La finalidad describe para qué se publica; las publicaciones pueden tener contenidos relacionales, promocionales, institucionales y de actualidad.

Esto es lo que ocurre con un club de fútbol, ya que se combinan anuncios de partidos, fichajes, mensajes a la afición, alineaciones, merchandising o patrocinadores. Por eso, en este trabajo resulta útil distinguir no solo la temática del contenido, sino también su finalidad, pues no va a generar la misma reacción una publicación puramente informativa de un partido que una con una mayor carga emocional. En la misma línea, Annamalai et al. (2021) demuestran que el contenido social genera el mayor volumen de engagement entre aficionados deportivos, mientras que los contenidos promocionales obtienen las cifras más bajas.

Otro factor es la presencia de personas, y en especial de figuras reconocibles. Bakhshi et al. (2014), en un análisis de un millón de imágenes de Instagram, demuestran que las publicaciones que contienen rostros humanos reciben un 38 % más de likes y un 32 % más de comentarios que las que no los muestran, controlando por número de seguidores y antigüedad de la cuenta. En un club deportivo, este efecto está respaldado por el estudio de Doyle et al. (2022): si aparecen jugadores en las imágenes lo normal es que se capte mejor la atención de la audiencia ya que gran parte de esta es su propia afición. Así, tras analizar 1.727 publicaciones de Instagram de 289 jugadores profesionales de la Major League Soccer, encontraron que las publicaciones cuyo contenido principal es el rendimiento de los futbolistas dan lugar a un engagement más alto. Por ello, dentro del caso analizado tiene más sentido centrar las variables en torno a la presencia o no de jugadores, aficionados o personal del equipo como entrenadores.

Además, en este caso concreto tiene sentido incorporar el contexto deportivo en el que se hace la publicación como variable. Porque una publicación tras una victoria, tras una derrota o durante un anuncio de fichaje no se recibe igual que una publicación institucional sin vínculo directo con la competición. Phua et al. (2018) demuestran experimentalmente que el resultado del partido (victoria/derrota) y la localización (casa/fuera) modifican el engagement que los

aficionados tienen con las publicaciones del club. Esta es una de las dimensiones más útiles en un perfil de club de fútbol, porque la respuesta de los usuarios no depende solo del contenido en sí, sino también del momento en que aparece. Por eso, el contexto deportivo no debe tratarse como un detalle secundario, sino como un elemento que puede ayudar a explicar por qué un mismo tipo de publicación genera respuestas distintas según la situación del equipo.

Por último, también influye el texto que acompaña a la publicación. El uso de hashtags, la longitud del mensaje o los índices de legibilidad pueden modificar la forma en que los usuarios interpretan y responden al contenido. Gkikas et al. (2022) usan estas características en publicaciones de Facebook, demostrando que, a partir de una determinada longitud del texto, el engagement decrece y que el uso moderado de hashtags lo maximiza. No obstante, es importante seleccionar solo aquellas variables textuales que puedan codificarse de manera clara y constante. Por esa razón, es más viable registrar el número de hashtags y la longitud del texto que interpretar la legibilidad del texto de cada publicación.

Por ello, este análisis del engagement en el perfil del Atlético de Madrid se centra en variables que sean fáciles de codificar y directamente identificables en las publicaciones del equipo, además de estar respaldadas en la literatura como predictoras del engagement. Dejando fuera variables como la calidad visual, la legibilidad del texto o el momento exacto de publicación, que encajan mejor como variables de control que como principales variables del análisis. Así, con la selección final de variables se ha buscado mantener un equilibrio entre lo que resulta relevante desde la literatura y lo que realmente puede medirse con rigor dentro del corpus del estudio.

Tabla 1. Variables analizadas

Variable	Categorías	Referencias
Formato de publicación	imagen / carrusel / reel	Liang y Wolfe (2022)
Temática principal	partido / entrenamiento / fichaje y renovación / celebración / merchandising / institucional / afición / otra	Annamalai et al. (2021)

Variable	Categorías	Referencias
Finalidad principal del contenido	informativo / promocional / institucional / entretenimiento / emocional	Schreiner et al. (2019)
Presencia de jugadores	sí / no	Doyle et al. (2022)
Presencia de afición	sí / no	Doyle et al. (2022)
Presencia de otras personas relevantes	entrenador o cuerpo técnico / directivos o representación institucional / otras personas / ninguna	Bakhshi et al. (2014)
Contexto deportivo inmediato	prepartido / post victoria / post empate / post derrota / fichaje o mercado / sin relación deportiva inmediata	Phua et al. (2018)
Uso de hashtags	ninguno / 1 a 2 / 3 o más	Gkikas et al. (2022)
Llamada a la acción en el texto	sí / no	Gkikas et al. (2022)
Longitud del texto	breve / media / larga	Gkikas et al. (2022)

2.3. Caso de estudio: Atlético de Madrid

Como caso de estudio para este trabajo, se ha utilizado el perfil oficial del Atlético de Madrid en Instagram porque cuenta con una cuenta bastante valiosa para analizar el engagement en el contenido de una marca deportiva. En primer lugar, porque se trata de un club con una trayectoria institucional muy amplia, que le proporciona una identidad de marca consolidada y una base de fans que interactúa diariamente con su perfil. El club se fundó en 1903 y, actualmente, recoge 11 títulos de LaLiga, 10 Copas del Rey y 3 UEFA Europa League.

En segundo lugar, el club cuenta con una presencia digital que hace viable y relevante su análisis. La cuenta oficial @atleticodemadrid ronda actualmente los 18 millones de seguidores y supera las 27.000 publicaciones, lo que muestra una actividad sostenida y una comunidad

amplia en la plataforma. Además, el club tiene presencia online en otras redes sociales, donde superan los 90 millones de seguidores.

Además, es relevante tener en cuenta que las publicaciones del Atlético de Madrid son muy dispares; la comunicación del club varía entre información deportiva, identidad de marca, relación con la afición y objetivos comerciales. Precisamente por eso, su perfil de Instagram es útil para comparar publicaciones de distinta naturaleza.

Asimismo, el Atlético de Madrid actualiza su cuenta diariamente, lo cual también es importante para este trabajo, ya que está al día con la actualidad deportiva y puede reflejar distintos momentos del calendario competitivo, lo cual es clave para un análisis de engagement, ya que permite observar cómo cambia la respuesta de los usuarios según el tipo de contenido y el contexto en el que se publica.

Por último, como se va a crear un corpus usando las publicaciones del club durante el calendario oficial del club, incluyendo el arranque de LaLiga y la fase de liga de la UEFA Champions League, va a ser posible que en el corpus se integren publicaciones vinculadas a competición nacional y europea, comunicación de marca y contenidos comerciales.

3. METODOLOGÍA

3.1. Análisis de contenido

Para este estudio, se utilizó el análisis de contenido para poder describir y cuantificar de manera sistemática las variables del contenido publicado, en este caso, por la cuenta del Atlético de Madrid en su perfil oficial de Instagram, con el objetivo de determinar qué variables generan un mayor nivel de engagement. Según Krippendorff (2018), este método es una técnica de investigación que permite realizar inferencias reproducibles y válidas a partir de datos textuales, visuales o simbólicos. Si nos centramos en las redes sociales, esta metodología ayuda a estudiar la relación dinámica entre las marcas y sus usuarios a través de métricas observables como el número de “me gusta”, comentarios o visualizaciones de cada publicación.

El objetivo de este estudio es detectar patrones y asociaciones estadísticas entre las variables de contenido y las métricas de engagement. Para ello, se aplicó un diseño cuantitativo y deductivo, partiendo de conceptos y hallazgos ya consolidados en la literatura sobre comunicación de marca, engagement digital y marketing deportivo (Annamalai et al., 2021;

Gkikas et al., 2022; Schreiner et al., 2019). La operacionalización de las variables, tanto dependientes como independientes, se apoya en investigaciones previas, adaptadas al contexto deportivo y a la red social de Instagram.

El corpus está compuesto por las publicaciones del perfil de Instagram @atleticodemadrid, publicadas entre el 1 de agosto de 2025 y el 4 de mayo de 2026. Se eligió este espacio temporal porque incluye el inicio de LaLiga y la fase de grupos de la UEFA Champions League, que son épocas de alta actividad deportiva. Se usaron las publicaciones de tipo imagen, carrusel y reel, por ser permanentes, excluyendo stories, directos, así como publicaciones fijadas (*pinned posts*), por las razones expuestas en el apartado 3.2 para facilitar la medición del engagement.

Atendiendo a un conjunto de variables previamente definidas (ver apartado 3.3), se codificó cada publicación. Partiendo de una hoja de una tabla de codificación, donde se sistematizaron las variables objeto de análisis, de manera que todas las publicaciones se analizaron de acuerdo con los mismos criterios. Para comprobar que las variables eran claras, se hizo una prueba con un número reducido del corpus, antes de pasar a codificar la muestra completa (Neuendorf, 2017).

A continuación, una vez codificado el corpus, mediante un análisis estadístico descriptivo y exploratorio, se muestra qué variables producen un mayor engagement. Ya que las métricas de likes y comentarios tienen una alta correlación entre sí, se unieron en un único índice: el engagement rate. Este se usó como variable dependiente del análisis (apartado 3.3). Después de haber recogido todos los datos, se aplicó un análisis de regresión lineal múltiple para determinar si las variables están relacionadas, midiendo la relación entre las variables independientes (texto o temática) y las dependientes (número de likes, comentarios o engagement).

En resumen, se usó el análisis de contenido para traducir la comunicación por Instagram del club en datos cuantificables, que permitieron examinar empíricamente la hipótesis de que ciertas variables que se encuentran en las publicaciones del club, por ejemplo, la presencia de jugadores, generan mayores niveles de engagement.

3.2. Selección muestral

Para poder construir el corpus de este estudio con publicaciones que representen la actividad del Atlético de Madrid en Instagram, se han usado todas las publicaciones permanentes de su perfil oficial (@atleticodemadrid) publicadas entre el 1 de agosto de 2025 y el 4 de mayo de

2026, un periodo de nueve meses que coincide con la mayor parte de la temporada futbolística 2025-2026. Haciendo esta delimitación, se ha creado una muestra con diferentes formatos, temáticas, contextos deportivos y fichajes porque coincidía con el inicio de LaLiga y la UEFA Champions League, y así se evitó que el análisis quede condicionado por un único evento.

Como señalan Filo et al. (2015), las redes sociales en el deporte no funcionan únicamente como espacios informativos, sino también como canales de relación con los aficionados y comunicación comercial, por lo que publican de forma continua con contenidos muy diferentes.

La unidad de análisis es la publicación individual del feed, dentro de la cual entran las imágenes, los carruseles y los reels. Se excluyeron del análisis las stories y los directos por ser formatos que no permiten medir el engagement. Además, se eliminaron las publicaciones fijadas (*pinned posts*) porque su exposición prolongada al principio del feed introducía un sesgo en las métricas de engagement que no estaba ligado a las características del contenido. En el caso de los carruseles de fotos, cada publicación se codificó como una unidad única, atendiendo al conjunto de imágenes y vídeos para identificar la temática principal y la presencia de personas, ya que es acorde con cómo funciona Instagram, que registra las métricas de interacción para el conjunto del carrusel y no para cada imagen individual.

En cuanto al procedimiento de selección, ya que cada mes el Atlético de Madrid publica entre 250 y 300 publicaciones y cuenta con más de 27.000, la estrategia muestral fue un análisis poblacional del periodo de estudio. De esta forma, el corpus está formado por publicaciones de todos los meses, lo cual redujo el riesgo de que un evento deportivo condicione los resultados. Finalmente, el corpus final quedó compuesto por 2.249 publicaciones permanentes, distribuidas aproximadamente de manera uniforme a lo largo de los nueve meses analizados (entre 250 y 320 publicaciones por mes).

Todas las publicaciones seleccionadas se registraron con los siguientes datos: fecha, enlace o identificador del post, formato, texto y métricas de engagement disponibles en el momento de la recogida. Los datos se recogieron en un espacio temporal de 2 a 3 días, en este caso del 2 al 4 de mayo de 2026.

En cuanto a la recogida de datos, esta se realizó usando el *Instagram Post Scraper* de Apify, con el que se pudieron extraer los datos públicos de la cuenta en Instagram. No fue necesario desarrollar código propio, ya que el *scraper* se configura a través de un formulario en la plataforma web de Apify. La ejecución se realizó el 2 de mayo de 2026, configurando el *scraper* con los siguientes parámetros: en el campo *Instagram profile URL* se introdujo la

dirección del perfil oficial del club; en *Maximum posts per profile* se estableció un límite de 2.700 publicaciones, valor que excede el volumen estimado para el periodo de estudio, en *Extract posts that are newer than* se indicó la fecha 2025-08-01, lo que restringe la extracción a las publicaciones posteriores al 1 de agosto de 2025. Además, se activó la opción *Skip pinned posts*, para excluir las publicaciones fijadas por el club, y se seleccionó el nivel de detalle *Basic data*, suficiente para los campos que requería el análisis. El resto de parámetros se mantuvieron con sus valores por defecto. El resultado se descargó en formato Excel desde Apify con los siguientes campos: identificador único, URL del post, fecha y hora de publicación (en horario peninsular español), formato (imagen, carrusel o reel), texto del pie (*caption*), hashtags asociados, número de likes, número de comentarios y URL de la imagen principal.

Adicionalmente, se descargaron las imágenes de cada publicación para que pudieran ser correctamente codificadas en el corpus.

Para la creación del corpus, las variables objetivas como el formato, fecha, hora, número de hashtags, longitud del texto, likes, comentarios se extrajeron automáticamente del dataset descargado. Mientras que las variables que requieren interpretación visual temática principal, finalidad, presencia de jugadores, presencia de afición, presencia de otras personas, contexto deportivo y llamada a la acción se codificaron usando la herramienta de IA Claude, que ha procesado la imagen principal de cada publicación y su texto asociado.

En cuanto a la fiabilidad de la codificación, que fue generada por herramientas de IA, para poder validarla correctamente, se codificó manualmente una submuestra aleatoria, que está compuesta por el 5 % del corpus final ($n = 112$ publicaciones). A partir de ahí, se calculó el coeficiente Alpha de Krippendorff (α) entre la codificación generada manualmente y la generada por IA. Este coeficiente se utiliza como medida estándar de fiabilidad intercodificador en análisis de contenido cuantitativo, por varias razones, entre ellas que admite cualquier número de categorías y permite comparar resultados de distintas variables y estudios bajo un mismo criterio (Krippendorff, 2018). Este cálculo se hizo usando la herramienta ReCal2 (Freelon, 2010), versión que se utiliza para dos codificadores y que opera con datos nominales.

Para interpretar los valores generados, se aplicaron los umbrales propuestos por Krippendorff (2018): valores $\alpha \geq 0,800$ indican fiabilidad alta y permiten extraer conclusiones definitivas, mientras que valores entre 0,667 y 0,800 indican fiabilidad aceptable, suficiente para conclusiones, y valores inferiores a 0,667 se consideran insuficientes para análisis

cuantitativos. En el presente estudio, seis de las siete variables analizadas superaron el umbral de fiabilidad alta ($\alpha \geq 0,800$), alcanzando una media global de $\alpha = 0,866$. La única variable que se situó por debajo de ese umbral, Presencia de otras personas ($\alpha = 0,799$), lo hizo por una diferencia mínima y se mantuvo dentro del rango de fiabilidad aceptable. Estos resultados respaldan la solidez de la codificación automatizada y la validez de los análisis estadísticos posteriores. La Tabla 2 recoge los valores obtenidos para cada variable.

Tabla 2. Resultados de la validación intercodificador mediante el coeficiente alpha de Krippendorff

Variable	N	Acuerdos	Discrepancias	% Acuerdo	Alpha de Krippendorff	Interpretación
Temática principal	112	103	9	92,0 %	0,900	Alta
Finalidad principal	112	105	7	93,8 %	0,901	Alta
Presencia de jugadores	112	108	4	96,4 %	0,856	Alta
Presencia de afición	112	104	8	92,9 %	0,842	Alta
Presencia de otras personas	112	100	12	89,3 %	0,799	Aceptable
Contexto deportivo	112	99	13	88,4 %	0,861	Alta
Llamada a la acción (CTA)	112	109	3	97,3 %	0,882	Alta
Media global	—	—	—	92,9 %	0,866	Alta

3.3. Variables dependientes e independientes

Para poder conseguir el objetivo de este trabajo, que es determinar las características del contenido publicado por el Atlético de Madrid en Instagram que generan mayores niveles de engagement, es necesario diferenciar entre variables dependientes e independientes. La elección de las variables independientes y de control del estudio se basa en literatura académica previa.

La variable dependiente es el engagement, que se ha intentado explicar con este trabajo y que mide la respuesta observable de los usuarios ante cada publicación.

Siguiendo la literatura sobre análisis de contenido en Instagram (Schreiner et al., 2019; Annamalai et al., 2021), el engagement se operacionaliza a través de las métricas públicas disponibles en la plataforma, a través del número de “me gusta”, comentarios o visualizaciones. Por ello, la variable dependiente principal es el nivel de engagement que genera cada publicación, usando las métricas explicadas para medirlo.

Las variables de análisis son el engagement total, como la suma de likes y comentarios, y el engagement rate, usando la siguiente fórmula: (engagement total / número de seguidores en la fecha de extracción x 100).

Mientras que las variables independientes son las variables o características de cada publicación que influyen en esa respuesta, como la temática o la presencia de jugadores. Las variables independientes, ya presentadas en el apartado 2.2, son las características de cada publicación que ayudan a explicar diferencias en el nivel de engagement.

Para este trabajo, las variables quedan cerradas así:

Tabla 3. Variables Independientes

Variable	Categorías	Ejemplo del corpus
Formato de publicación	imagen / carrusel / reel	Tipo de publicación dentro del feed de Instagram. Imagen: una sola fotografía. Carrusel: secuencia de varias imágenes o vídeos que el usuario desliza. Reel: vídeo vertical breve.
Temática principal	partido / entrenamiento / fichaje y renovación / celebración / merchandising / institucional / afición / otra	Asunto sobre el que trata la publicación. Partido: publicación de un encuentro (alineación, jugadas, resultado). Entrenamiento: sesiones preparatorias del equipo. Fichaje y renovación: llegada, salida o renovación de un jugador concreto. Celebración: festejos por victorias, títulos o ascensos. Merchandising y patrocinio: productos del club y acciones comerciales con marcas patrocinadoras. Institucional: contenido del club como entidad,

Variable	Categorías	Ejemplo del corpus
		<p>no centrado en jugadores (acuerdos corporativos, mensajes de la directiva, aniversarios institucionales). Afición y comunidad: publicaciones donde los seguidores son protagonistas. Jugadores y plantilla: contenido centrado en los futbolistas fuera del partido (entrevistas, vida diaria, perfiles). Otra: contenido no clasificable en las anteriores. Ver Anexo I para ejemplos visuales.</p>
<p>Finalidad principal del contenido</p>	<p>informativo / promocional / institucional / entretenimiento / emocional</p>	<p>Objetivo comunicativo de la publicación. Informativa: comunica un hecho factual concreto que el seguidor necesita conocer (hora del partido, alineación, resultado, comunicado oficial). Entretenimiento: contenido cuyo fin es divertir o enganchar sin transmitir información factual (vídeos lúdicos, momentos curiosos del entrenamiento, retos entre jugadores). Promocional: busca venta o conversión hacia el aficionado como consumidor (camisetas, abonos, productos). Institucional: refuerza la imagen corporativa del club (mensajes de</p>

Variable	Categorías	Ejemplo del corpus
		la directiva, valores). Emocional: apela al vínculo identitario y afectivo con el club (mensajes de orgullo, agradecimiento a la afición, homenajes).
Presencia de jugadores	sí / no	
Presencia de afición	sí / no	
Presencia de otras personas relevantes	entrenador o cuerpo técnico / directivos o representación institucional / otras personas / ninguna	Aparecen otras personas distintas de jugadores o afición. Entrenador o cuerpo técnico: el entrenador principal, segundos entrenadores, preparadores físicos. Directivos o representación institucional: presidente, miembros de la junta directiva, embajadores oficiales. Otras personas: rivales, árbitros, celebridades, periodistas, patrocinadores en eventos. Si aparecen varias categorías, se asigna la más prominente visualmente o la primera mencionada en el caption.
Contexto deportivo inmediato	prepartido / post victoria / post empate / post derrota / fichaje o mercado / sin relación deportiva inmediata	Momento competitivo en el que se sitúa la publicación. Prepartido: publicación anterior al inicio del encuentro (previas, alineaciones, llegada al estadio). Post-victoria, post-empate, post-derrota: publicación posterior a un partido con ese resultado.

Variable	Categorías	Ejemplo del corpus
		Fichaje o mercado: anuncios de incorporaciones, salidas, renovaciones o rumores oficiales. Sin relación deportiva inmediata: publicaciones no vinculadas a un partido próximo o reciente (contenido institucional, comercial, de afición).
Uso de hashtags Llamada a la acción en el texto	ninguno / 1 a 2 / 3 o más sí / no	El texto de la publicación invita explícitamente al seguidor a realizar una acción concreta: comentar, votar, comprar, visitar un enlace, participar en una campaña, responder a una pregunta o compartir. Ejemplo del corpus: “¿Cuál es tu jugada favorita del partido? Cuéntanoslo en comentarios” (CTA = sí). Una frase descriptiva o emocional sin invitación a actuar se codifica como CTA = no.
Longitud del texto	breve (≤ 3 palabras) / media (4–7 palabras) / larga (≥ 8 palabras)	

La variable de llamada a la acción en el texto, que se codifica como sí o no, dependiendo de si se invita a los seguidores en los comentarios de la publicación a realizar una acción, como comentar, votar, visitar un enlace, participar en una campaña o responder a una pregunta. Por último, se analizó la longitud del texto, que se clasificó como breve, media o larga, siendo breve entre cero y tres palabras, media entre cuatro y siete y larga más de ocho palabras.

VARIABLES DE CONTROL:

Para controlar el efecto de factores temporales que pueden influir en el engagement, independientemente de las variables de contenido de la publicación, se incluyen dos variables de control:

Tabla 4. Variables de control

Variable	Categorías
Momento de publicación	mañana / tarde / noche
Día de publicación	laborable / fin de semana

3.4. Codificación y análisis de los datos

Una vez construido el corpus, habiendo sacado las variables objetivas directamente de los datos extraídos y las variables interpretativas usando la herramienta de IA Claude Haiku 4.5. Partiendo de esa base de datos, se ha decidido dividir en tres partes el análisis de los datos, atendiendo a las diferentes preguntas de investigación.

Antes de empezar la codificación, ha sido necesario transformar la variable del engagement. Las métricas del Atlético de Madrid presentaban una asimetría muy marcada, ya que la mayoría de las publicaciones reciben un nivel de interacción moderado (mediana de 18.800), mientras que solo algunas publicaciones se vuelven virales y llegan al millón, algo común en redes sociales. Lo cual provoca que la media quede distorsionada por estos valores extremos y hace que no se cumpla el supuesto de normalidad. Por ese motivo, además de trabajar con la mediana en los análisis descriptivos (ya que no se ve afectada por valores extremos), se ha aplicado una transformación logarítmica, $\log(\text{engagement} + 1)$, para no usar los datos brutos, reduciendo así la asimetría y permitiendo que se puedan interpretar los coeficientes en términos relativos. También se han borrado del corpus cinco publicaciones en las que se había eliminado la visibilidad de likes en el momento de la recogida de datos, por lo que el corpus final ha quedado en 2.244 publicaciones.

En primer lugar, se empieza por un análisis descriptivo. El objetivo es caracterizar el corpus antes de relacionar las variables, observando cómo están repartidas las características del contenido y qué valores adopta el engagement en términos de likes, comentarios, engagement

total y engagement rate. Con esto, se permite saber cuál es el comportamiento de la cuenta y detectar posibles desequilibrios entre categorías que puedan afectar a los análisis posteriores.

En segundo lugar, se relaciona cada una de las variables independientes con el engagement por separado. Dado que el engagement no sigue una distribución normal, porque los datos están muy sesgados, no se puede usar ANOVA. Por ello, se ha decidido usar la alternativa no paramétrica, Kruskal-Wallis, que compara los niveles de engagement entre las distintas categorías sin asumir normalidad entre los datos. El cálculo se ha realizado en Python mediante la función *kruskal* de la librería SciPy, cuyo script completo puede consultarse en el Anexo II.

Para terminar, se crea un modelo de regresión lineal múltiple sobre el engagement ya transformado. Mientras que en el análisis anterior se ha examinado cada variable de forma aislada, con la regresión se estima el efecto que tiene cada una controlado por todas las demás. Por ejemplo, las publicaciones de celebración suelen coincidir con momentos de victoria, de manera que solo un modelo conjunto puede distinguir si lo que mueve el engagement es la temática o el contexto deportivo. Antes de estimar el modelo, se comprobó que las variables independientes no presentaban problemas de multicolinealidad, mediante el cálculo del factor de inflación de la varianza (VIF) para cada una de ellas. Todos los valores se sitúan por debajo del umbral problemático de 10 (O'Brien, 2007), con un máximo de 6,05 y una media de 2,61, lo que respalda la interpretación independiente de los coeficientes del modelo. La estimación se ha realizado en Python con la librería *statsmodels*, que permite incorporar variables categóricas mediante codificación dummy. No se incluyen las variables de Finalidad principal y Presencia de otras personas, ya que presentan un solapamiento conceptual elevado con Temática principal y con las variables de Presencia de jugadores y afición. La inclusión simultánea generaría multicolinealidad y dificultaría la interpretación de los coeficientes.

Tabla 5. Factor de inflación de la varianza (VIF)

Variable	VIF Máximo
Temática principal	6,00
Presencia de jugadores	1,86
Presencia de afición	1,51
Contexto deportivo	6,05
Llamada a la acción (CTA)	1,15
Longitud del texto	1,49

Variable	VIF Máximo
Momento del día	1,38
Formato	1,32
Hashtags	1,11
Día de publicación	1,04
Media global	2,61

Todo el proceso que se ha utilizado en cuanto al tratamiento de datos y análisis estadístico se ha realizado con Python. Los scripts completos que se han usado en cada fase se encuentran en el Anexo II, para garantizar que sea posible reproducir el estudio.

4. RESULTADOS

4.1. Descripción de la muestra

El corpus completo que se ha analizado lo componen 2.244 publicaciones permanentes del perfil oficial del Atlético de Madrid, que están repartidas de forma homogénea a lo largo de los nueve meses del periodo de estudio. Para cada variable independiente, predomina una categoría que se repite más que las otras a lo largo de las publicaciones objeto de estudio. En cuanto al formato, predominan los carruseles y los reels sobre las imágenes únicas. Por temática, las publicaciones son sobre todo de partidos y jugadores de la plantilla, ambos constituyen más de la mitad del corpus, mientras que se encuentran muchas menos publicaciones comerciales.

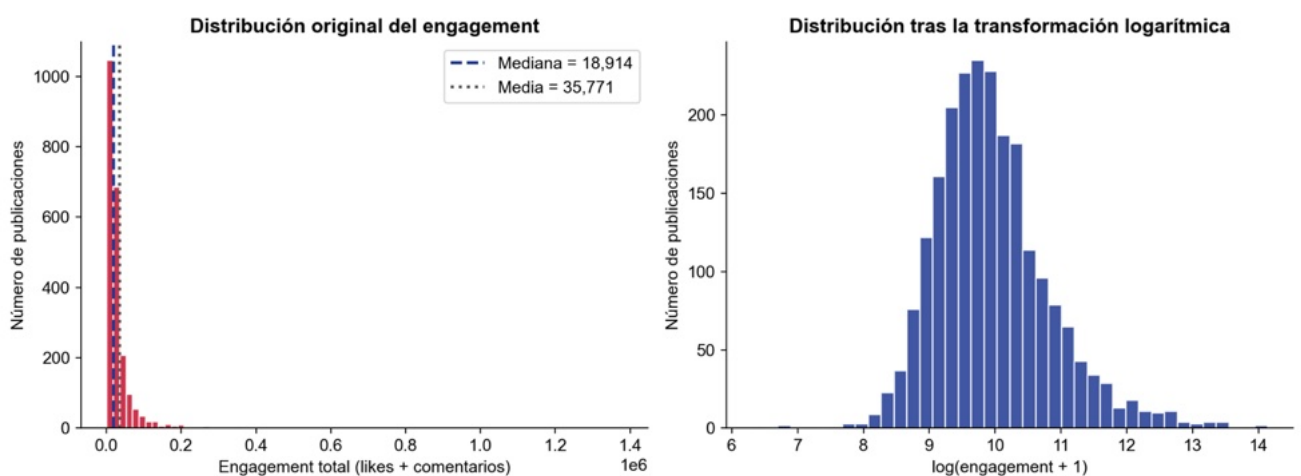
Respecto al engagement, los datos muestran la asimetría que ya se había anticipado. La publicación mediana del corpus tiene 18.914 interacciones contando likes y comentarios, mientras que la media está por encima de las 35.000 por la diferencia entre algunas pocas publicaciones que se han hecho virales y superan el millón de interacciones. Esta diferencia entre media y mediana justifica que se haga uso de la mediana y de la transformación logarítmica en los análisis posteriores. El engagement rate medio es del 0,20 % de la base de seguidores. La Tabla 6 recoge los descriptivos completos de las cuatro métricas analizadas.

Tabla 6. Métricas de engagement (N=2.244)

Métrica	Media	Mediana	Desv. típica	Mínimo	Máximo
Likes	35.242	18.654	71.836	526	1.341.591
Comentarios	528	96	5.557	2	245.060
Engagement total	35.771	18.914	73.249	547	1.378.709
Engagement rate (%)	0,20	0,11	0,41	0	7,66

La Figura 1 muestra gráficamente esta asimetría. A la izquierda, la distribución original del engagement evidencia el sesgo que suele darse en las métricas en redes sociales, con la gran mayoría de publicaciones concentradas en valores bajos-medios y una cola larga formada por unas pocas publicaciones virales que alcanzan cifras muy superiores. A la derecha, tras aplicar la transformación logarítmica $\log(\text{engagement} + 1)$, la distribución se aproxima a una forma simétrica, lo que cumple los supuestos necesarios para aplicar los modelos paramétricos del apartado 4.8.

Figura 1. Distribución del engagement antes y después de la transformación logarítmica ($N = 2.244$).



Antes de presentar los resultados detallados de cada variable, la Tabla 7 ofrece una visión global del análisis bivariado, en la que se comparan los niveles de engagement entre las distintas categorías de las doce variables independientes mediante la prueba de Kruskal-Wallis. Las variables se han ordenado de mayor a menor tamaño del efecto (epsilon cuadrado), lo que permite identificar de un vistazo cuáles muestran una asociación más fuerte con el engagement.

Como se observa en la tabla, el contexto deportivo y la temática principal son las variables con mayor capacidad explicativa, ambas con un tamaño del efecto considerado medio. Les siguen, con efectos pequeños, el momento del día, la finalidad principal del contenido, la presencia de otras personas relevantes y la longitud del texto. Las restantes variables presentan efectos insignificantes (presencia de jugadores, formato, presencia de afición y CTA) o directamente no alcanzan el umbral de significación estadística (tipo de día y hashtags). Los siguientes apartados profundizan en cada variable, especialmente en aquellas que muestran un efecto más relevante.

Tabla 7. Resultados del análisis bivariado entre las variables independientes y el engagement

Variable	H	p-valor	Epsilon ²	Tamaño del efecto
Contexto deportivo	264,52	< 0,001	0,113	Medio
Temática principal	161,24	< 0,001	0,069	Medio
Momento del día	77,65	< 0,001	0,034	Pequeño
Finalidad principal	78,58	< 0,001	0,033	Pequeño
Presencia de otras personas	58,09	< 0,001	0,022	Pequeño
Longitud del texto	39,93	< 0,001	0,017	Pequeño
Presencia de jugadores	22,99	< 0,001	0,009	Insignificante
Formato	14,13	< 0,001	0,005	Insignificante
Presencia de afición	12,46	0,002	0,005	Insignificante
CTA	4,62	0,032	0,002	Insignificante
Tipo de día	2,79	0,095	0,001	No significativo
Hashtags	3,13	0,209	0,001	No significativo

4.2. Engagement según formato

El formato de la publicación muestra una asociación significativa con el engagement en la prueba de Kruskal-Wallis ($p < 0,01$), aunque su efecto sea muy reducido ($\epsilon^2 = 0,005$). Se puede afirmar que el formato explica apenas el 0,5 % por sí mismo, por lo que es insignificante en magnitud. Las tres categorías de formato han obtenido niveles de engagement muy similares; teniendo en cuenta la mediana, las imágenes (19.934) y los carruseles (19.633) se sitúan prácticamente al mismo nivel, mientras que los reels quedan un poco por debajo (17.198). Sin embargo, los reels tienen la media de engagement más alta de los tres formatos (39.789 frente

a unos 33.500 de imágenes y carruseles), lo que se explica por una mayor presencia de publicaciones virales que elevan el promedio sin que el reel mediano destaque. Es importante precisar que esta comparación está basada en likes y comentarios, sin incluir las visualizaciones de los reels, que no se recogieron en la extracción de datos.

En conjunto, al ser las diferencias por formato pequeñas, se puede afirmar que esta variable no es un factor determinante del engagement en el perfil del Atlético de Madrid. Estos resultados, atendiendo a la literatura previa, coinciden con Annamalai et al. (2021), que en su estudio sobre clubes deportivos también encontraron que las fotografías obtenían altos niveles de interacción.

4.3. Engagement según temática

De todas las variables del corpus, la temática es de las que mayor capacidad explicativa tiene ($\epsilon^2 = 0,069$, $p < 0,001$). En la Tabla 5 se puede observar que las publicaciones sobre fichajes/renovaciones y las de celebración concentran un mayor engagement, mientras que las de merchandising/patrocinio y las de entrenamiento son las que menos interacción generan por parte de los usuarios.

Tabla 8. Engagement según la temática de la publicación

Temática	N	Mediana engagement	Media likes	Media comentarios
Fichaje / renovación	63	27.687	44.740	4.586
Celebración	162	25.176	49.104	336
Jugadores / plantilla	632	20.658	39.974	331
Partido	754	18.830	33.043	553
Afición / comunidad	173	18.615	36.099	666
Institucional	94	18.178	31.049	392
Entrenamiento	217	12.711	16.968	103
Merchandising / patrocinio	98	11.432	16.557	120

Con estos resultados se puede afirmar lo que ya se planteó en el marco teórico por Annamalai et al. (2021), que en su estudio observaba que el contenido de carácter social y emocional genera más interacciones entre el público que el contenido promocional. En cuanto a las

publicaciones que se centran en mostrar a jugadores, siguiendo el estudio de Doyle et al. (2022), que observaron que el contenido protagonizado por futbolistas favorece la interacción de la audiencia, los datos del corpus siguen esa misma línea con una mediana de engagement bastante favorable. Además, se puede señalar que las publicaciones que muestran fichajes y renovaciones destacan sobre todo por el volumen de comentarios que generan, muy por encima del resto de temáticas, lo que sugiere que este tipo de contenido no solo gusta, sino que genera conversación y debate entre los seguidores.

Además, mirando a los fichajes y renovaciones, aunque esta categoría representa solo el 2,8 % del corpus (63 publicaciones), concentra una media de 4.586 comentarios por publicación, frente a una media global de 529 comentarios en el resto del corpus. Por lo que, cada publicación de fichaje genera alrededor de nueve veces más conversación que una publicación media. Lo que demuestra que los anuncios de nuevos fichajes son el principal generador de debate entre seguidores.

4.4. Engagement según presencia de jugadores o afición

Tanto la presencia de jugadores como la de afición muestran una asociación significativa con el engagement en el análisis bivariado ($p < 0,001$ y $p = 0,002$ respectivamente), si bien con un tamaño del efecto muy reducido en ambos casos ($\epsilon^2 = 0,009$ y $\epsilon^2 = 0,005$). En el caso de los jugadores, las publicaciones en las que aparecen alcanzan una mediana de engagement de 19.392, frente a 15.443 cuando no aparecen. Una diferencia significativa estadísticamente, pero pequeña en magnitud, donde conviene matizar que los jugadores están presentes en más del 80 % del corpus, por lo que su aparición es casi una constante y discrimina poco entre publicaciones. La presencia de otras personas relevantes, como el cuerpo técnico, rivales o directivos, tiene un efecto algo mayor, aunque todavía pequeño ($\epsilon^2 = 0,022$). La Tabla 9 resume los resultados de la prueba de Kruskal-Wallis para las tres variables.

Tabla 9. Resultados del análisis bivariado (prueba de Kruskal-Wallis) para las variables de presencia de personas

Variable	H	p-valor	Epsilon ²	Tamaño del efecto
Presencia de jugadores	22,99	< 0,001	0,009	Insignificante
Presencia de afición	12,46	0,002	0,005	Insignificante
Presencia de otras personas	58,09	< 0,001	0,022	Pequeño

Tabla 10. Engagement según la presencia de jugadores y de afición

Variable / categoría	N	Mediana eng.	Media likes	Media coment.
Jugadores: sí	1.878	19.392	35.955	411
Jugadores: no	304	15.443	29.867	444
Afición: sí	804	20.570	34.938	873
Afición: no	1.426	18.212	35.413	336

En el caso de la afición, las publicaciones en las que aparecen aficionados alcanzan una mediana de 20.570 frente a 18.212, una diferencia igualmente pequeña. Sin embargo, aquí se puede observar un detalle importante y es que la presencia de afición se asocia con un mayor número de comentarios (873 de media frente a 336 cuando no aparece), por lo que se puede afirmar que las publicaciones con afición fomentan la conversación entre seguidores.

En comparación con el estudio de Doyle et al. (2022), ya mencionado anteriormente, en el que observaron que el contenido que mostraba el rendimiento de los futbolistas generaba un engagement más alto. Los datos obtenidos apuntan en la misma dirección: las publicaciones en las que aparecen jugadores obtienen una mediana de engagement superior (19.392 frente a 15.443).

4.5 Engagement según contexto deportivo

De todo el estudio, esta es la variable que más influye en el engagement ($\epsilon^2 = 0,113$, $p < 0,001$). Siguiendo los datos de la Tabla 11 y el estudio de Phua et al. (2018), donde determinaron que el resultado del partido afecta a cómo los seguidores interactúan con las publicaciones del club, se observa que las publicaciones posteriores a una derrota, a una victoria o relacionadas con el mercado de fichajes son las que más interacción producen, con medianas de engagement próximas a las 25.000 interacciones. La categoría "otro/no identificable" presenta también valores altos, aunque por tratarse de una categoría residual no admite una interpretación temática clara. En el extremo opuesto, las categorías de previa de partido y entrenamiento son las que menos interacción generan, con medianas que no alcanzan las 14.000.

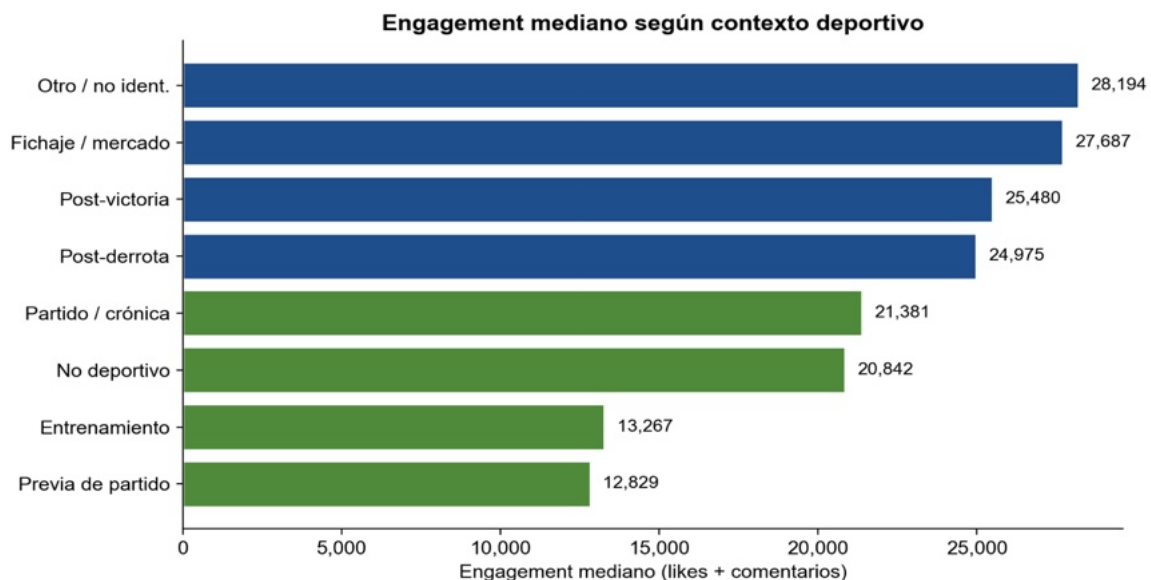
Tabla 11. Engagement según el contexto deportivo

Contexto deportivo	N	Mediana eng.	Media likes	Media coment.
Fichaje / mercado	65	27.687	45.994	668

Contexto deportivo	N	Mediana eng.	Media likes	Media coment.
Post-victoria	375	25.480	44.175	487
Post-derrota	31	24.975	42.467	3.898
Partido en juego / crónica	469	21.381	38.519	408
No deportivo (instit. /comercial)	548	20.842	41.711	939
Entrenamiento / preparación	305	13.267	19.966	105
Previa de partido	389	12.829	21.166	176

La Figura 2 representa visualmente estas diferencias entre contextos deportivos. Se puede observar la separación entre los contextos centrados en resultados o mercado, que se sitúan en la mitad superior con medianas de engagement próximas o superiores a las 20.000 interacciones, y los contextos más rutinarios (previa de partido y entrenamiento), cuyas medianas no alcanzan las 14.000.

Figura 2. Engagement mediano según contexto deportivo (N = 2.244). En azul se destacan los contextos que superan la mediana global del corpus y en verde los que se sitúan por debajo.

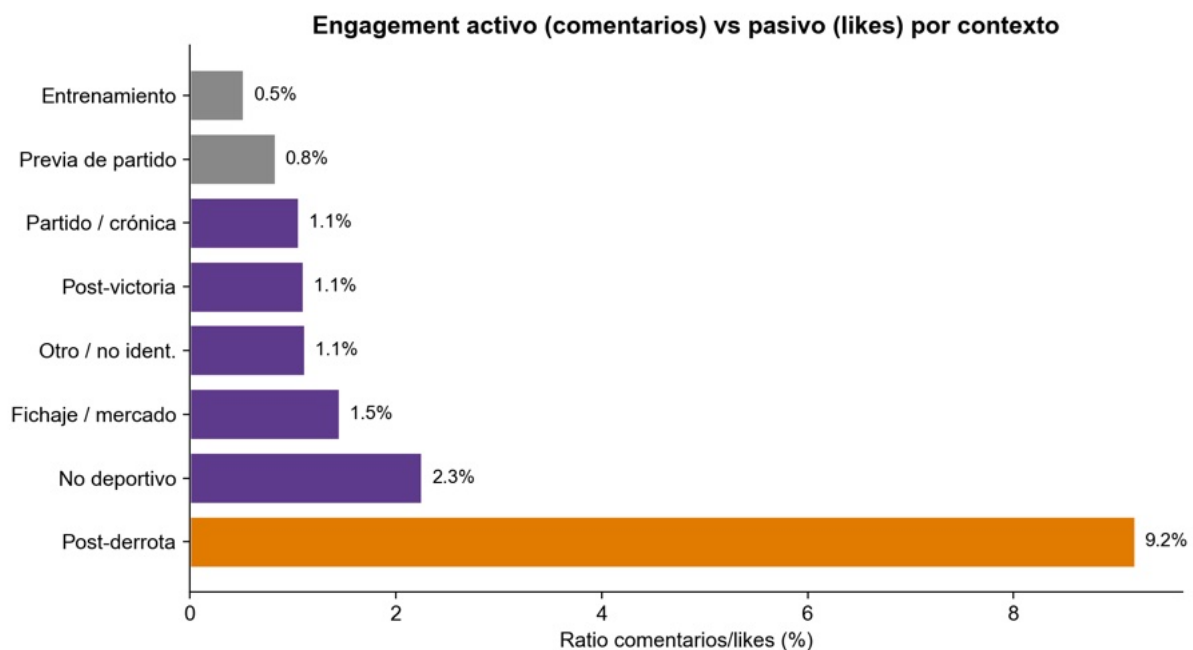


Parece importante destacar que, aunque la mediana de engagement de la derrota es ligeramente inferior a la de la victoria, la métrica de su engagement se compone de manera diferente. Las publicaciones posteriores a una derrota concentran una media de casi 3.900 comentarios, muy por encima de cualquier otra categoría del contexto deportivo, mientras que los likes sí están por debajo de las publicaciones con victoria. Lo cual plantea que la interacción generada por

una derrota no es la misma que genera una victoria, donde el usuario da menos «me gusta» para pasar a comentar, probablemente para expresar crítica o iniciar un debate. Estas diferencias de engagement pasivo y activo aportan más datos sobre cómo se compone el mismo y que el simple recuento conjunto de interacciones tiende a ocultar.

Para visualizar esta diferencia cualitativa entre engagement pasivo y activo, la Figura 3 representa el ratio entre la media de comentarios y la media de likes para cada contexto deportivo. Este indicador, que se expresa en porcentaje, muestra qué proporción de la interacción se presenta en forma de comentario frente a like. Se puede observar que las publicaciones post-derrota presentan un ratio muy superior al del resto de contextos, lo que confirma que la respuesta de la audiencia tras una derrota es cualitativamente distinta de la que se produce tras una victoria u otro contexto positivo.

Figura 3. Engagement activo (comentarios) frente a pasivo (likes) por contexto deportivo, expresado como porcentaje de comentarios sobre likes (N = 2.244). El color naranja destaca el caso de las publicaciones post-derrota, claramente diferenciado del resto.



Conviene matizar este hallazgo señalando que el corpus contiene solo 31 publicaciones posteriores a una derrota, frente a 375 posteriores a una victoria, lo que refleja tanto la trayectoria deportiva del club durante el periodo analizado como una posible estrategia editorial que tiende a comunicar menos los resultados negativos. Aunque el efecto detectado es estadísticamente robusto ($p < 0,001$), su precisión sería mayor con un mayor número de casos. Más allá de la magnitud exacta, lo relevante es la naturaleza cualitativa del fenómeno, que apunta a una activación específica de la conversación tras los resultados adversos.

4.6. Engagement según finalidad de contenido

La finalidad del contenido también se asocia de forma significativa con el engagement ($p < 0,001$). Como recoge la Tabla 12, las publicaciones cuya finalidad principal es el entretenimiento son las que mayor engagement obtienen, seguidas de las de carácter emocional, mientras que las de finalidad informativa y, sobre todo, promocional son las que menos interacción generan.

Tabla 12. Engagement según la finalidad del contenido

Finalidad	N	Mediana eng.	Media likes	Media coment.
Entretenimiento	477	21.381	45.664	438
Emocional	1.260	19.897	35.615	391
Institucional	42	19.082	30.601	319
Informativa	300	14.498	25.781	599
Promocional	165	13.073	20.655	1.771

Estos resultados refuerzan lo observado en el análisis por temática, ya que vuelven a situar el contenido promocional en la posición más baja en cuanto a engagement, lo que confirma de nuevo la idea de Annamalai et al. (2021) sobre el menor rendimiento del contenido comercial. No obstante, es relevante observar los comentarios, puesto que las publicaciones promocionales, pese a recibir pocos likes, concentran la media de comentarios más alta de todas las finalidades (1.771). Esto se explica probablemente por la presencia de sorteos y campañas participativas, que invitan explícitamente a comentar y generan así un tipo de interacción distinto al del resto de contenidos.

4.7. Características de textos y variables temporales

El resto de las variables tienen un impacto menor sobre el engagement. Si observamos la longitud del texto, esta resulta significativa ($\epsilon^2 = 0,017$), ya que la mediana de las publicaciones que tienen texto de longitud media (mediana de 20.667) o larga (20.696) es superior a las que presentan texto breve (16.771). Lo que sugiere que las publicaciones con más desarrollo textual están relacionadas con un mayor engagement.

Sin embargo, el número de hashtags no muestra una asociación significativa con el engagement ($p = 0,209$). En principio, las publicaciones con tres o más hashtags son las que generan más

interacción, sin embargo, solo hay 9 publicaciones en esta categoría. De hecho, el 93 % de las publicaciones del corpus no usan hashtags, lo cual refleja que el uso de hashtags no es habitual en la estrategia comunicativa del club y, en consecuencia, no aparece como un factor relevante para explicar las diferencias de engagement entre publicaciones.

La llamada a la acción presenta una asociación bastante leve y, además, en sentido contrario al esperado. Las publicaciones con llamada a la acción generan una mediana de engagement un poco inferior (17.624) a las que no la incluyen (19.102), aunque las publicaciones que la incluyen generan muchos más comentarios (una media de 1426 frente a 417).

Respecto a las variables temporales, el momento del día sí resulta relevante. Igualmente, en sentido contrario a lo que se esperaba, los posts realizados por la noche generan más engagement (26.885) que los que se publican por la mañana (18.671) y la tarde (17.426); esta diferencia queda confirmada en la regresión, donde publicar de noche aumenta el engagement un 36 % respecto a la mañana. Además, el tipo de día tampoco influye demasiado, aunque publicar en fin de semana genera un engagement ligeramente menor (18.218 frente a 19.273 en días laborables).

4.8. Análisis comparativo

Mediante un modelo de regresión lineal múltiple sobre el engagement, valoramos el efecto conjunto de todas las variables independientes sobre el mismo. En conjunto, el modelo es significativo ($F = 15,37$, $p < 0,001$) y explica un 16,8 % de la varianza del engagement ($R^2 = 0,168$). Aunque este valor sea pequeño, sigue la línea de estudios previos sobre engagement en redes sociales (Lee et al. 2018), en los que gran parte de la variación depende de factores que no se pueden controlar, como son el funcionamiento del algoritmo de la plataforma o contenidos virales puntuales.

La siguiente tabla recoge los coeficientes del modelo, su nivel de significación y su efecto porcentual sobre el engagement. Los resultados del modelo confirman y ordenan lo observado en los análisis previos. El contexto deportivo sigue siendo el principal predictor, ya que, respecto a la previa de partido, las publicaciones posteriores a una derrota incrementan el engagement en torno a un 92 % ($\beta = 0,654$, $p < 0,001$), las posteriores a una victoria un 73 % ($\beta = 0,549$, $p < 0,001$) y las de mercado de fichajes un 66,2 % ($\beta = 0,508$, $p < 0,05$). La temática se mantiene en segundo lugar, donde la categoría que más reduce el engagement es el merchandising, con una caída del 52,5 % respecto a las publicaciones de partido ($\beta = -0,745$, $p < 0,001$). Además, se deben tener en cuenta los efectos secundarios, como son la longitud del

texto, donde las publicaciones de longitud media superan a las breves en torno a un 19,5 % ($\beta = 0,178$, $p < 0,001$), y el del momento del día, ya que los posts nocturnos generan un 35,8 % más de engagement que los publicados por la mañana ($\beta = 0,306$, $p < 0,001$). La Figura 4 representa visualmente los efectos significativos del modelo ordenados por su magnitud.

Una vez controladas el resto de las variables, la presencia de jugadores, la presencia de afición y la llamada a la acción dejan de tener un efecto significativo sobre el engagement, lo que confirma que, al introducir las variables de contexto y temática, estas dejan de ser variables importantes.

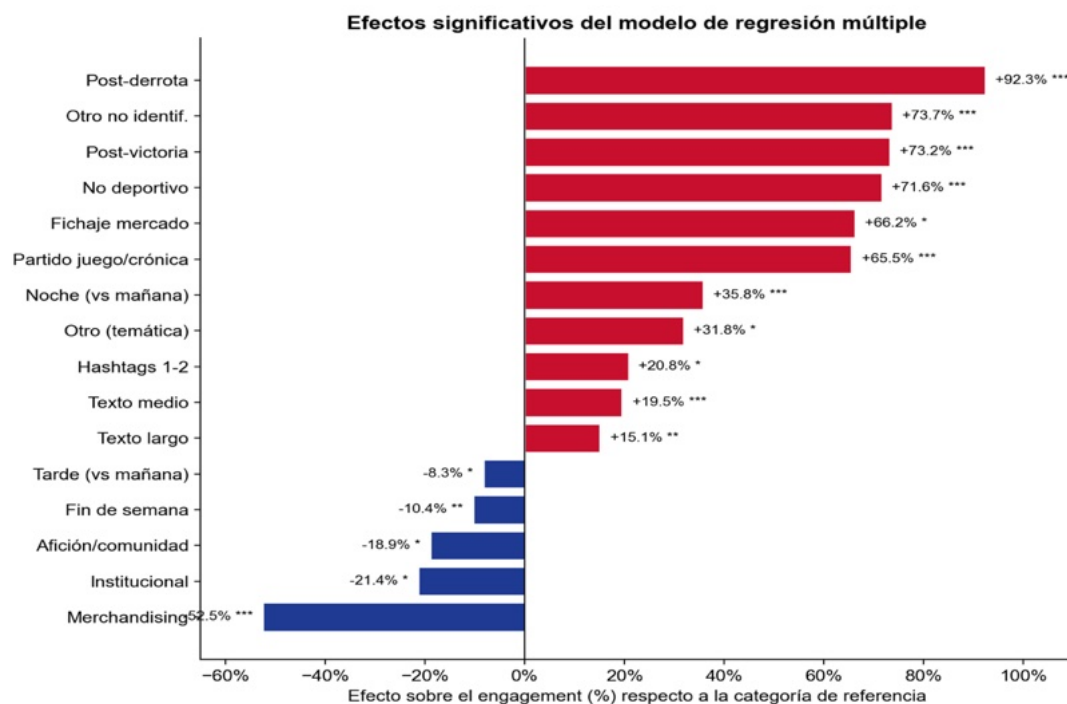
Tabla 13. Coeficientes del modelo de regresión lineal múltiple sobre el engagement (log)

Variable y categoría	β (coef.)	Error típico	Efecto (%)	Significación
FORMATO (ref. imagen)				
Carrusel	0,089	0,046	+9,3	$p = 0,054$
Reel	-0,081	0,048	-7,8	$p = 0,088$
TEMÁTICA (ref. partido)				
Jugadores y plantilla	-0,001	0,056	-0,1	n.s.
Entrenamiento	-0,200	0,105	-18,1	$p = 0,056$
Afición y comunidad	-0,209	0,088	-18,9	$p < 0,05$
Celebración	0,084	0,083	+8,8	n.s.
Merchandising y patrocinio	-0,745	0,111	-52,5	$p < 0,001$
Institucional	-0,241	0,110	-21,4	$p < 0,05$
Fichaje y renovación	0,223	0,249	+25,0	n.s.
Otro	0,276	0,136	+31,8	$p < 0,05$
CONTEXTO DEPORTIVO (ref. previa de partido)				
Partido en juego / crónica	0,504	0,061	+65,5	$p < 0,001$
Post-victoria	0,549	0,070	+73,2	$p < 0,001$
Post-derrota	0,654	0,154	+92,3	$p < 0,001$
Entrenamiento / preparación	0,132	0,095	+14,1	n.s.
Fichaje / mercado	0,508	0,245	+66,2	$p < 0,05$
No deportivo (instit./comercial)	0,540	0,072	+71,6	$p < 0,001$
Otro no identificable	0,552	0,121	+73,7	$p < 0,001$
PRESENCIA DE JUGADORES (ref. no)				
Sí	0,015	0,063	+1,5	n.s.
PRESENCIA DE AFICIÓN (ref. no)				
Sí	-0,023	0,044	-2,3	n.s.
HASHTAGS (ref. 0)				
1 a 2	0,189	0,073	+20,8	$p < 0,05$
3 o más	0,348	0,273	+41,7	n.s.

Variable y categoría	β (coef.)	Error típico	Efecto (%)	Significación
CTA (ref. no)				
Sí	-0,062	0,059	-6,0	n.s.
LONGITUD DEL TEXTO (ref. breve)				
Media	0,178	0,041	+19,5	$p < 0,001$
Larga	0,140	0,047	+15,1	$p < 0,01$
MOMENTO DEL DÍA (ref. mañana)				
Tarde	-0,087	0,040	-8,3	$p < 0,05$
Noche	0,306	0,054	+35,8	$p < 0,001$
TIPO DE DÍA (ref. laborable)				
Fin de semana	-0,110	0,038	-10,4	$p < 0,01$

La Figura 4 muestra visualmente los efectos significativos del modelo de regresión, separando en la parte superior los efectos positivos (en rojo) y en la inferior los efectos negativos (en azul). La asimetría visual del gráfico es informativa por sí misma, ya que los efectos positivos más fuertes provienen del contexto deportivo (post-derrota, post-victoria, fichaje y partido en juego) y el efecto negativo más fuerte corresponde a la temática comercial (merchandising). Junto a ellos, efectos secundarios como el del momento nocturno (+35,8 %) o la longitud media del texto (+19,5 %) se sitúan en la zona intermedia y completan el cuadro de los factores que modulan el engagement en el perfil del Atlético de Madrid.

*Figura 4. Efectos significativos del modelo de regresión múltiple sobre el engagement (N = 2.244). Los porcentajes representan el cambio en el engagement respecto a la categoría de referencia de cada variable. *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.*



5. CONCLUSIONES

Este trabajo se planteó con el objetivo de identificar qué características del contenido publicado por el Atlético de Madrid en Instagram se asocian con un mayor nivel de engagement. Para responderlo, se construyó un corpus de 2.244 publicaciones permanentes del perfil oficial @atleticodemadrid, y dividiendo el análisis en tres partes. Los resultados obtenidos han permitido extraer varias conclusiones que se exponen a continuación.

5.1. Principales hallazgos

La conclusión más clara del estudio es que el engagement en el perfil del Atlético de Madrid está determinado, sobre todo, por el contexto deportivo en el que se publica y la temática del contenido. Son las variables con mayor capacidad explicativa en el análisis bivariado y se mantienen como principales predictores en el modelo de regresión múltiple. El resto de las variables tienen un peso considerablemente menor o dejan de ser relevantes una vez se controla por estas dos.

Dentro del contexto deportivo, los momentos vinculados al resultado y al mercado son los que más movilizan a la afición. Este resultado está ligado al estudio de Phua et al. (2018), en el que se determina que los aficionados tienden a reaccionar más con el club en momentos con mayor significado, como son los resultados de un partido. Lo cual sigue en línea con los resultados

obtenidos de las publicaciones de previas de partido o de entrenamiento, que son las que menos engagement generan.

Un hallazgo que parece relevante mencionar es la diferencia cualitativa entre victorias y derrotas. Aunque la mediana de engagement es similar, su composición es diferente. Según los resultados, se podría afirmar que las derrotas incitan más al debate, ya que tienen una media de comentarios muy superior a las victorias.

En cuanto a la temática, los datos confirman lo planteado por Annamalai et al. (2021) sobre el contenido deportivo en redes sociales, y es que las publicaciones de carácter social y emocional generan más engagement, mientras que las de naturaleza comercial son las que menos interacción generan. La variable de finalidad refuerza esta idea, los resultados posicionan el contenido promocional en el nivel de engagement más bajo. La afición del Atlético responde, por tanto, mejor al contenido que apela a la identidad y a la emoción compartida que al que persigue un objetivo comercial.

Algunos resultados matizan lo planteado en la literatura. La presencia de jugadores, que, según Doyle et al. (2022), debería favorecer la interacción, muestra en este corpus un efecto menor que deja de ser significativo al controlar por el resto de las variables. Una posible explicación es que los jugadores aparecen en más del 80 % del corpus, lo que convierte su presencia en una constante y reduce su capacidad para explicar diferencias. El estudio de Doyle et al. analizaba perfiles individuales de futbolistas, donde su presencia sí es distintiva a la hora de medir el engagement. Sin embargo, en una cuenta de club como la del Atlético, la presencia de jugadores forma parte de la identidad del club. Algo parecido ocurre con el formato, que tiene un efecto casi nulo, aunque los reels tienen la media más alta entre los tres formatos, lo que matiza la idea, extendida en la literatura previa, de que el vídeo corto genera mayor interacción.

5.2. Implicaciones del estudio

Los resultados de este trabajo tienen implicaciones tanto en el plano teórico como en el de la gestión aplicada de la comunicación digital de un club deportivo.

Desde un punto de vista teórico, los resultados confirman que, en una cuenta deportiva, como es un club de fútbol con una gran afición como es la del Atlético de Madrid, las variables que determinan el engagement no son tanto las características formales de la publicación, como el formato, hashtags o llamadas a la acción, sino el momento competitivo en que se publica y el tipo de contenido que se comparte. Por ello, los modelos que analizan el engagement en redes

sociales deben matizarse al aplicarse al ámbito deportivo, en donde la dimensión emocional y la identidad del club tienen más peso que las variables formales, que en otros contextos sí son más influyentes. Adicionalmente, este estudio aporta un matiz cualitativo que no se ha explorado mucho en la literatura previa: como ya se ha mencionado, el engagement no es una métrica uniforme, sino que se descompone en una dimensión pasiva (likes) y una dimensión activa (comentarios) que, dependiendo del contexto, se pueden mover en distintas direcciones. Como ya se ha demostrado, en el caso de las publicaciones posteriores a una derrota, caen los likes, pero se multiplican los comentarios, siendo esto un ejemplo claro de que reducir el engagement a un único agregado oculta diferencias relevantes en el comportamiento de la audiencia.

Desde una perspectiva de comunicación digital, si se quisiera maximizar la interacción con los seguidores, el club debería centrarse en los siguientes puntos:

El primer punto y el más evidente es darle prioridad al contenido que se centra en torno al calendario competitivo, ya que los momentos de resultado y de mercado (post-victoria, post-derrota, fichajes) son los que más interacción concentran. Al mismo tiempo, deberían reducir el número de posts comerciales, dado que el contenido de merchandising y patrocinio es el que sistemáticamente obtiene menos engagement. Como este tipo de publicaciones cumplen objetivos comerciales para el club, no deben eliminarlas por completo, sino reconocer que rinden menos en términos de interacción y plantear si pueden integrarse con elementos emocionales o narrativos que aumenten su atractivo.

Una segunda implicación gira en torno al tratamiento de los resultados deportivos negativos. Lo normal es que los clubes reduzcan su actividad en redes tras una derrota, porque suelen ser publicaciones que no van a gustar a los usuarios. Sin embargo, los datos de este estudio sugieren que tras una derrota, el debate se intensifica notablemente, ya que los comentarios se multiplican por ocho respecto a una victoria. Por lo que, los momentos negativos se convierten en una oportunidad para gestionar la conversación de forma activa, mediante mensajes que canalicen el debate y refuercen la identidad colectiva.

Un tercer punto tiene que ver con la elección del momento en el que se publican los posts. Los resultados muestran que si la publicación se hace de noche se genera un 35,8 % más de engagement que si se publica por la mañana, lo que abre un espacio para revisar las horas habituales de publicación de los clubes, que suelen concentrar su actividad en redes en franjas matutinas o de mediodía. Esta diferencia podría aprovecharse programando los contenidos de

mayor valor (anuncios de fichaje, vídeos de celebración, contenido institucional importante) en franjas nocturnas, especialmente en días laborables, donde el efecto del momento es más pronunciado.

Otro punto a tener en cuenta se relaciona con el contenido textual. Se ha demostrado que hay una asociación significativa entre la longitud media o larga del caption y un mayor engagement, lo cual sugiere que el texto sigue cumpliendo una función importante. Los textos que están un poco más desarrollados, que contextualizan la imagen y aportan información o tono emocional, parecen funcionar mejor que los mensajes más breves. Con estos resultados, se invita a revisar la tendencia que tiene el club hacia los captions mínimos en sus publicaciones, que predominan en el corpus, donde más de un 40 % de las publicaciones se limita a tres palabras o menos.

Por último, los resultados aportan una perspectiva crítica sobre algunas prácticas habituales en redes sociales que en este corpus no muestran efecto: el uso de hashtags y las llamadas a la acción explícitas no se asocian con un mayor engagement una vez controladas las demás variables. Esto sugiere que la efectividad de estas herramientas, ampliamente recomendadas en guías generales de social media, depende del tipo de cuenta. En perfiles institucionales con una comunidad muy identificada como la del Atlético de Madrid, el vínculo emocional con la marca parece suplir la necesidad de estímulos formales para generar interacción.

En conjunto, este estudio aporta un marco empírico que combina implicaciones teóricas (la necesidad de modelos sectoriales para entender el engagement deportivo, y la conveniencia de descomponer el engagement en sus dimensiones pasiva y activa) con implicaciones aplicadas concretas (priorización del calendario competitivo, gestión activa de las derrotas, optimización del momento de publicación, atención al desarrollo del caption y revisión crítica del uso de hashtags y CTA). Esta combinación es la que justifica la utilidad práctica del análisis cuantitativo de contenido más allá del ámbito puramente académico.

Tabla 14. Decálogo de recomendaciones aplicadas para la gestión del engagement en la cuenta de Instagram del Atlético de Madrid

#	Recomendación	Acción concreta	Evidencia del estudio
1	Priorizar el calendario competitivo	Concentrar la actividad en publicaciones vinculadas a resultados,	<i>Contexto deportivo: $\varepsilon^2 = 0,113$ (efecto medio, el mayor del estudio).</i>

#	Recomendación	Acción concreta	Evidencia del estudio
		partidos y mercado de fichajes.	
2	Gestionar activamente las derrotas	Mantener la comunicación tras una derrota con mensajes que canalicen el debate y refuercen la identidad colectiva, en lugar de reducir la actividad.	<i>Post-derrota: media de comentarios casi 8 veces superior a post-victoria.</i>
3	Aprovechar fichajes y celebraciones	Tratar fichajes y celebraciones como contenido estratégico, no como mero anuncio, dado su potencial viral.	<i>Fichaje +66,2 % de engagement; celebración entre las temáticas con mayor mediana.</i>
4	Moderar el contenido comercial	Reducir el volumen de publicaciones puramente promocionales o integrarlas con elementos narrativos y emocionales.	<i>Merchandising: -52,5 % de engagement respecto a publicaciones de partido.</i>
5	Optimizar el momento de publicación	Programar los contenidos de mayor valor en franjas nocturnas, especialmente en días laborables.	<i>Publicaciones nocturnas: +35,8 % de engagement respecto a las matutinas.</i>
6	Desarrollar el caption	Evitar captions ultrabreves y aportar contexto o tono	<i>Texto medio: +19,5 %; texto largo: +15,1 %, respecto a captions breves.</i>

#	Recomendación	Acción concreta	Evidencia del estudio
		emocional con textos de longitud media o larga.	
7	Reconsiderar la apuesta por el vídeo corto	No asumir que los reels rinden sistemáticamente mejor; revisar caso a caso frente a imágenes y carruseles.	<i>Las tres modalidades presentan medianas de engagement equivalentes (17.000–20.000).</i>
8	Revisar el uso de hashtags	Cuestionar el uso sistemático de hashtags en cuentas institucionales con audiencia consolidada.	<i>Hashtags no significativo en bivariado ($p = 0,209$); 93 % del corpus no emplea ninguno.</i>
9	Reconsiderar las llamadas a la acción explícitas	Evaluar críticamente la inclusión de CTA en cuentas con afición muy identificada, donde el vínculo emocional suple la necesidad de estímulos formales.	<i>CTA no significativo en regresión múltiple ($p = 0,290$).</i>
10	Medir el engagement de forma desagregada	Analizar likes y comentarios por separado para captar la dimensión pasiva y activa de la interacción, en lugar de basarse en un único agregado.	<i>Hallazgo distintivo del estudio: la composición del engagement varía según el contexto.</i>

5.3. Limitaciones del estudio

En cuanto a las limitaciones que han surgido durante el trabajo, cabe mencionar las siguientes. En primer lugar, el corpus se centra en un único club y periodo, por lo que los resultados no son necesariamente extrapolables a otros equipos o a otras temporadas. La situación deportiva concreta del Atlético de Madrid durante los meses analizados puede haber influido en algunos de los patrones observados, especialmente en lo relativo al contexto deportivo, que podría no ser igual en otros años o temporadas deportivas.

En segundo lugar, de las 2.564 publicaciones del periodo elegido solo se pudieron descargar las imágenes de 2.249, debido a restricciones técnicas en los servidores de Instagram durante el proceso de extracción.

En tercer lugar, las métricas de engagement utilizadas se basan exclusivamente en likes y comentarios que eran visibles públicamente, lo que deja fuera otras formas de interacción como las visualizaciones, los compartidos o los guardados.

Por último, el modelo de regresión final explica un 16,8 % de la varianza del engagement, lo que indica que una parte importante de la variación sigue siendo atribuible a factores no medidos en este estudio, como el funcionamiento del algoritmo de Instagram o publicaciones que por razones imprevisibles se vuelven virales.

Cabe reflexionar sobre la naturaleza de esos factores, que se pueden agrupar en tres bloques.

En primer lugar, se deben tener en cuenta factores algorítmicos y de distribución. La plataforma de Instagram no muestra cada publicación a todos los seguidores, sino que es el algoritmo quien decide qué contenido aparece en el feed principal de cada usuario, basándose en su historial de interacciones con esa cuenta y el contenido que consume habitualmente. Por lo que, una publicación, en este caso del Atlético de Madrid no siempre llega a la misma audiencia ni en las mismas condiciones, esto introduce una variabilidad importante que el estudio no puede capturar. Además, el algoritmo también decide el orden en el que las publicaciones aparecen en el feed, lo cual afecta a la interacción de esa publicación, si va después de un contenido con el que la gente ha interactuado bastante es más probable que la que va detrás genere mayor interacción que si apareciera tras una secuencia de contenidos irrelevantes para el usuario.

Un segundo bloque tiene que ver con factores propios del comportamiento del usuario que el corpus tampoco registra. El estado del usuario en el momento en que ve una publicación afecta a su interacción con ella. Este puede verse afectado por el tiempo que lleva en la aplicación,

las publicaciones con las que ya ha interactuado, su estado de ánimo o el contexto físico en el que se encuentra son variables que modulan la probabilidad de interacción y que un análisis basado en datos agregados de publicaciones no puede observar. Tampoco se ha controlado la dimensión social de la interacción, es decir, el efecto que produce sobre un usuario ver que otras personas de su entorno (amigos, otros seguidores del club) han interactuado previamente con la publicación, lo que en la literatura se conoce como prueba social (Cialdini, 1984) y se ha demostrado relevante para explicar la viralidad de los contenidos.

Un tercer bloque, más ligado a la coyuntura, lo constituyen los factores contextuales externos al propio club. Una publicación puede beneficiarse o verse perjudicada por eventos ajenos que coinciden con su difusión, como noticias deportivas paralelas (un partido relevante de otro equipo, una declaración polémica, una lesión sonada), tendencias o memes virales del momento, festividades, atención mediática focalizada en otros asuntos o incluso la actividad simultánea de otras cuentas del entorno futbolístico.

Reconocer la existencia de estos tres bloques de variables no medidas no resta valor a los hallazgos obtenidos, sino que delimita con honestidad su alcance. El modelo identifica las variables del contenido y del contexto deportivo que mejor predicen el engagement, pero el engagement real de una publicación concreta también dependerá de factores algorítmicos, individuales y coyunturales que requerirían un diseño de investigación distinto, probablemente con acceso a datos internos de la plataforma o a paneles de usuarios.

5.4. Futuras líneas de investigación

A partir de estas limitaciones se podrían abrir varias líneas de investigación. Una primera consistiría en replicar el análisis sobre otros clubes de fútbol, comparando entidades de tamaño, identidad y resultados deportivos diferentes, para evaluar si los patrones observados son específicos del Atlético de Madrid o se generalizan al sector. Una segunda opción requeriría ampliar el periodo de estudio para agrupar varias temporadas completas para valorar cómo afectan al engagement.

Una tercera línea de interés sería incorporar métricas de engagement adicionales más allá de los likes y los comentarios, como los compartidos, guardados y visualizaciones, que permitirían una mejor caracterización de cómo se comporta la audiencia. Y, por último, partiendo del hallazgo sobre cómo está compuesto el engagement en el caso de las derrotas, consistiría en realizar un análisis cualitativo o de sentimientos sobre el contenido de los comentarios, lo que

permitiría distinguir entre comentarios de apoyo, crítica o debate, y entender con mayor precisión cómo cambia la reacción de la afición según el contexto competitivo.

Por último, una línea de continuación natural de este trabajo dentro del campo del business analytics consistiría en construir un modelo predictivo del engagement esperado para una publicación antes de su difusión, a partir de las variables identificadas como significativas. Este modelo podría incorporarse como herramienta de apoyo a la decisión en el flujo editorial del club, permitiendo estimar el engagement potencial de cada publicación antes de su publicación y orientar la programación del calendario de contenidos. Un desarrollo de este tipo enriquecería el alcance descriptivo del estudio actual con una dimensión prescriptiva, plenamente alineada con la lógica del análisis de datos aplicado al ámbito empresarial.

En resumen, con este trabajo se ofrece evidencia empírica del funcionamiento del engagement en una cuenta deportiva concreta, la del Atlético de Madrid, y afirma que para entender la interacción en redes sociales es importante tener en cuenta tanto el contexto en que se producen como la respuesta de la audiencia.

6. REFERENCIAS BIBLIOGRÁFICAS

Annamalai, B., Yoshida, M., Varshney, S., Pathak, A. A. y Venugopal, P. (2021). Social media content strategy for sport clubs to drive fan engagement. *Journal of Retailing and Consumer Services*, 62, 102648.

<https://www.sciencedirect.com/science/article/abs/pii/S0969698921002149?via%3Dihub>

Bakhshi, S., Shamma, D. A. y Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 965–974). Association for Computing Machinery. https://www.researchgate.net/publication/266655817_Faces_engage_us_photos_with_faces_attract_more_likes_and_comments_on_Instagram

Barger, V., Peltier, J. W. y Schultz, D. E. (2016). Social media and consumer engagement: A review and research agenda. *Journal of Research in Interactive Marketing*, 10(4), 268–287. <https://doi.org/10.1108/JRIM-06-2016-0065>

Bowden, J. L. H. (2009). The process of customer engagement: A conceptual framework. *Journal of Marketing Theory and Practice*, 17(1), 63–74.

Brodie, R. J., Hollebeek, L. D., Jurić, B. y Ilić, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 14(3), 252–271. <https://journals.sagepub.com/doi/10.1177/1094670511411703>

Cialdini, R. B. (1984). *Influence: The psychology of persuasion*. William Morrow and Company.

Cvijikj, I. P. y Michahelles, F. (2013). Online engagement factors on Facebook brand pages. *Social Network Analysis and Mining*, 3(4), 843–861. https://www.researchgate.net/publication/257801193_Online_engagement_factors_on_Facebook_brand_pages

Doyle, J. P., Su, Y. y Kunkel, T. (2022). Athlete branding via social media: Examining the factors influencing consumer engagement on Instagram. *European Sport Management Quarterly*, 22(4), 506–526. https://www.researchgate.net/publication/343839461_Athlete_branding_via_social_media_examining_the_factors_influencing_consumer_engagement_on_Instagram

Enríquez, D. (2023, 26 de octubre). ¿Cómo han cambiado las redes sociales las estrategias de marca? IEBS Business School. <https://www.iebschool.com/hub/como-ha-cambiado-las-redes-sociales-las-estrategias-de-marca/>

Filo, K., Lock, D. y Karg, A. (2015). Sport and social media research: A review. *Sport Management Review*, 18(2), 166–181. https://www.researchgate.net/publication/269730038_Sport_and_social_media_research_A_review

Gkikas, D. C., Tzafilkou, K., Theodoridis, P. K., Garmpis, A. y Gkikas, M. C. (2022). How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. *International Journal of Information Management Data Insights*, 2(1), 100067. https://www.researchgate.net/publication/359245825_How_do_text_characteristics_impact_user_engagement_in_social_media_posts_Modeling_content_readability_length_and_hashtags_number_in_Facebook

Hollebeek, L. D., Glynn, M. S. y Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165. <https://www.sciencedirect.com/science/article/abs/pii/S1094996813000649>

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4.^a ed.). Sage Publications.

Lee, D., Hosanagar, K. y Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, 64(11), 5105–5131. <https://pubsonline.informs.org/doi/10.1287/mnsc.2017.2902>

Liang, S. y Wolfe, J. (2022). Getting a feel of Instagram Reels: The effects of posting format on online engagement. *Journal of Student Research*, 11(4). <https://doi.org/10.47611/jsrhs.v11i4.3600>

Marvi, R., Foroudi, P. y Jerez Jerez, M. (2024). A bibliometric review of customer engagement in the international domain: Reviewing the past and the present. *Corporate Reputation Review*. https://www.researchgate.net/publication/382907923_A_Bibliometric_Review_of_Customer_Engagement_in_the_International_Domain_Reviewing_the_Past_and_the_Present

Neuendorf, K. A. (2017). *The content analysis guidebook* (2.^a ed.). Sage Publications.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690. <https://link.springer.com/article/10.1007/s11135-006-9018-6>

Phua, J., Pan, P. L. y Chen, K. J. (2018). Sport team-endorsed brands on Facebook: Effects of game outcome (win/loss), location (home/away) and team identification on fans' brand evaluations. *Online Information Review*, 42(7), 1064–1083. https://www.researchgate.net/publication/326473990_Sport_team-endorsed_brands_on_Facebook_Effects_of_game_outcome_winloss_location_homeaway_and_team_identification_on_fans'_brand_evaluations

Schivinski, B., Christodoulides, G. y Dabrowski, D. (2016). Measuring consumers' engagement with brand-related social-media content: Development and validation of a scale that identifies levels of social-media engagement with brands. *Journal of Advertising Research*, 56(1), 64–80. https://www.researchgate.net/publication/294424061_Measuring_Consumers'_Engagement_With_Brand-Related_Social-Media_Content_Development_and_Validation_of_a_Scale_that_Identifies_Levels_of_Social-Media_Engagement_with_Brands

Schreiner, M., Fischer, T. y Riedl, R. (2019). Impact of content characteristics and emotion on behavioral engagement in social media: Literature review and research agenda. *International Journal of Research in Marketing*, 36(3), 429–448. https://www.researchgate.net/publication/333008809_Impact_of_content_characteristics_and_emotion_on_behavioral_engagement_in_social_media_literature_review_and_research_agenda

Stavros, C., Meng, M. D., Westberg, K. y Farrelly, F. (2014). Understanding fan motivation for interacting on social media. *Sport Management Review*, 17(4), 455–469. <https://dialnet.unirioja.es/servlet/articulo?codigo=4857076>

Trunfio, M. y Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, 2021(3), 267–292. https://www.researchgate.net/publication/354951823_Conceptualising_and_measuring_social_media_engagement_A_systematic_literature_review

Vander Schee, B. A., Peltier, J. y Dahl, A. J. (2020). Antecedent consumer factors, consequential branding outcomes and measures of online consumer engagement: Current research and future directions. *Journal of Research in Interactive Marketing*, 14(2), 239–268. https://www.researchgate.net/publication/341911896_Antecedent_consumer_factors_consequential_branding_outcomes_and_measures_of_online_consumer_engagement_current_research_and_future_directions

Anexo I

El presente anexo recoge una publicación representativa del corpus para cada una de las nueve categorías de la variable Temática principal. Cada ficha incluye la captura de la publicación, su identificador interno dentro del corpus (ID_post) y una breve justificación de los criterios visuales y textuales que han llevado a su asignación a esa categoría. Esta selección tiene un fin meramente ilustrativo y busca facilitar la comprensión de las definiciones operativas recogidas en la Tabla 3 del apartado 3.3.

A.I.1. Partido

Publicaciones cuyo contenido principal gira en torno a un encuentro deportivo concreto. Incluyen alineaciones, jugadas, momentos del partido, marcadores parciales o crónicas inmediatas. Lo determinante es que la imagen o el texto remiten directamente a un partido específico.



IDENTIFICADOR: ATM_0014

A.I.2. Celebración

Publicaciones que recogen festejos por victorias, títulos, ascensos o efemérides deportivas relevantes. Lo que las distingue del simple post-victoria es la presencia de elementos festivos: vestuarios, abrazos, trofeos, gestos de exaltación. Su carga emocional es alta y su finalidad suele ser estrechar el vínculo con la afición tras un éxito.



IDENTIFICADOR: ATM_0013

A.I.3. Entrenamiento

Publicaciones centradas en sesiones preparatorias del equipo, ya sean ejercicios físicos, tácticos o trabajos específicos por demarcaciones. No corresponden a un partido en juego, sino al trabajo previo. Suelen mostrar la ciudad deportiva, el gimnasio o el campo de entrenamiento.



IDENTIFICADOR: ATM_0050

A.I.4. Jugadores y plantilla

Contenido protagonizado por uno o varios futbolistas de la primera plantilla fuera del contexto de un partido o entrenamiento concreto. Entran aquí entrevistas, perfiles, vida cotidiana, contenido humorístico entre jugadores, retos o presentaciones de protagonismo individual.



IDENTIFICADOR: ATM_0015

A.I.5. Fichaje y renovación

Publicaciones que anuncian la incorporación, salida o renovación de un jugador concreto. Lo que distingue esta categoría de la anterior es que el jugador es noticia por un movimiento contractual, no por su faceta deportiva o personal habitual.



IDENTIFICADOR: ATM_0001

A.I.6. Institucional

Contenido del club como entidad y no centrado en jugadores. Incluye mensajes de la directiva, comunicados corporativos, acuerdos con instituciones, acciones de responsabilidad social corporativa, conmemoraciones del club o eventos de carácter institucional.



IDENTIFICADOR: ATM_0392

A.I.7. Merchandising y patrocinio

Publicaciones de carácter comercial centradas en productos del club (camisetas, equipaciones, merchandising) o en acciones promocionales con marcas patrocinadoras. Su finalidad principal es la venta o la activación de patrocinio.



IDENTIFICADOR: ATM_0103

A.I.8. Afición y comunidad

Publicaciones donde los seguidores son los protagonistas: gradas, peñas, viajes, mosaicos, mensajes dirigidos a la afición, contenido generado por usuarios o iniciativas comunitarias. La diferencia con celebración es que el foco está en los aficionados, no en el equipo.



IDENTIFICADOR: ATM_1201

A.I.9. Otra

Categoría residual para publicaciones que, por su naturaleza o ambigüedad, no encajan claramente en ninguna de las anteriores. Se ha aplicado solo en aquellos casos en los que la asignación a una categoría específica habría forzado la codificación.



IDENTIFICADOR: ATM_0311

Anexo II

El análisis de este trabajo se ha realizado en Python 3.9 sobre macOS. Las librerías utilizadas han sido pandas para el tratamiento de datos, requests para la descarga de imágenes, la librería anthropic para la codificación con el modelo multimodal Claude Haiku 4.5, scipy para la prueba estadística no paramétrica de Kruskal-Wallis y statsmodels para el modelo de regresión lineal múltiple. Estos scripts están presentados en el mismo orden en que se ejecutaron durante la investigación.

A.II.1. Descarga de imágenes del corpus

Este script se ha utilizado para leer el conjunto de datos que se extrajeron mediante Apify, filtrar las publicaciones del periodo de estudio, asignar un identificador único a cada una y descargar la imagen principal. El identificador ha permitido vincular después cada imagen con su codificación y con los datos del análisis.

```
import pandas as pd
import requests
from pathlib import Path
import time

# Leer datos extraídos con Apify y filtrar por periodo
df = pd.read_excel("dataset_apify.xlsx")
df['timestamp'] = pd.to_datetime(df['timestamp']).dt.tz_convert('Europe/Madrid')
df = df[(df['timestamp'] >= '2025-08-01') & (df['timestamp'] <= '2026-05-04 23:59:59')]
df = df[df['isPinned'] != True]
df = df.sort_values('timestamp').reset_index(drop=True)

# Asignar identificador unico a cada publicacion
df['ID_post'] = ['ATM_' + str(i+1).zfill(4) for i in range(len(df))]
df.to_excel("mapeo_corpus.xlsx", index=False)

# Descargar la imagen principal de cada publicacion
Path("imagenes_atletico").mkdir(exist_ok=True)
for _, fila in df.iterrows():
    ruta = Path("imagenes_atletico") / f"{fila['ID_post']}.jpg"
```

```

if ruta.exists():
    continue
try:
    r = requests.get(fila['displayUrl'], timeout=30)
    if r.status_code == 200:
        ruta.write_bytes(r.content)
except Exception:
    pass
time.sleep(5)

```

A.II.2. Codificación del corpus

Este script se ha utilizado para enviar cada imagen con el caption de cada publicación al modelo Claude Haiku 4.5, y guarda la codificación devuelta en un formato estructurado. Cada llamada al modelo devuelve un objeto JSON con la categoría asignada a cada variable interpretativa.

```

import os, json, base64
import pandas as pd
from pathlib import Path
from anthropic import Anthropic

cliente = Anthropic(api_key=os.environ["ANTHROPIC_API_KEY"])

PROMPT = """Analiza esta publicacion del Atletico de Madrid y devuelve un JSON con:
- Tematica_principal: partido, celebracion, entrenamiento, jugadores_plantilla,
  fichaje_renovacion, institucional, merchandising_patrocinio, aficion_comunidad, otro
- Finalidad_principal: emocional, informativa, promocional, entretenimiento, institucional
- Presencia_jugadores: si, no, no_identificable
- Presencia_aficion: si, no, no_identificable
- Presencia_otras_personas: no, cuerpo_tecnico_staff, rivales_arbitros,
  directivos_institucional, celebridades_patrocinadores_otros, no_identificable
- Contexto_deportivo: previa_partido, partido_en_juego_o_cronica, post_victoria,
  post_derrota, entrenamiento_preparacion, fichaje_mercado,
  no_deportivo_institucional_comercial, otro_no_identificable

```

```

- CTA: si, no
- confianza_codificacion: alta, media, baja""""

df = pd.read_excel("mapeo_corpus.xlsx")

with open("codificacion.jsonl", "w") as f:
    for _, fila in df.iterrows():
        ruta = Path("imagenes_atletico") / f"fila['ID_post'].jpg"
        if not ruta.exists():
            continue
        imagen_b64 = base64.b64encode(ruta.read_bytes()).decode("utf-8")
        respuesta = cliente.messages.create(
            model="claude-haiku-4-5",
            max_tokens=500,
            messages=[{"role": "user", "content": [
                {"type": "image", "source": {"type": "base64",
                    "media_type": "image/jpeg", "data": imagen_b64}},
                {"type": "text", "text": PROMPT + f"\n\nCaption: {fila['caption']}"}
            ]}]
        )
        resultado = json.loads(respuesta.content[0].text)
        resultado['ID_post'] = fila['ID_post']
        f.write(json.dumps(resultado, ensure_ascii=False) + "\n")

```

A.II.3. Generación del corpus final

Este script combina los datos objetivos extraídos por Apify con las variables interpretativas codificadas por el modelo, y calcula las variables derivadas que se utilizan después en el análisis: engagement total, engagement rate, longitud del texto, momento del día y tipo de día. El resultado es el archivo único sobre el que se realizan todos los análisis posteriores.

```

import pandas as pd
import json, re

SEGUIDORES = 18000000

```

```

df = pd.read_excel("mapeo_corpus.xlsx")

codificacion = {}
with open("codificacion.jsonl") as f:
    for linea in f:
        obj = json.loads(linea)
        codificacion[obj['ID_post']] = obj

# Variables derivadas
df['Formato'] = df['type'].map({'Image':'imagen', 'Sidecar':'carrusel', 'Video':'reel'})
df['Hashtags_n'] = df['caption'].fillna("").apply(lambda c: len(re.findall(r'#\w+', c)))
df['Hashtags_cat'] = df['Hashtags_n'].apply(
    lambda n: '0' if n == 0 else ('1-2' if n <= 2 else '3+'))
df['Texto_palabras'] = df['caption'].fillna("").apply(
    lambda c: len(re.findall(r"[A-Za-zÑáéíóúüÁÉÍÓÚÿ]", re.sub(r'#\w+|@\w+|http\S+', "", c))))
df['Texto_longitud'] = df['Texto_palabras'].apply(
    lambda n: 'breve' if n <= 3 else ('media' if n <= 7 else 'larga'))
df['Momento_dia'] = df['timestamp'].dt.hour.apply(
    lambda h: 'manana' if 6 <= h <= 13 else ('tarde' if 14 <= h <= 20 else 'noche'))
df['Tipo_dia'] = df['timestamp'].dt.weekday.apply(
    lambda d: 'fin_de_semana' if d >= 5 else 'laborable')

df['Likes'] = df['likesCount']
df['Comentarios'] = df['commentsCount']
df['Engagement_total'] = df['Likes'].fillna(0) + df['Comentarios'].fillna(0)
df['Engagement_rate'] = df['Engagement_total'] / SEGUIDORES * 100

for var in ['Tematica_principal', 'Finalidad_principal', 'Presencia_jugadores',
           'Presencia_aficion', 'Presencia_otras_personas', 'Contexto_deportivo',
           'CTA', 'confianza_codificacion']:
    df[var] = df['ID_post'].apply(lambda x: codificacion.get(x, {}).get(var, 'NA'))

df.to_excel("Corpus_Final_TFG.xlsx", index=False)

```

A.II.4. Análisis descriptivo

Este script calcula los estadísticos descriptivos de las métricas de engagement (media, mediana, desviación típica, mínimo y máximo) y comprueba la reducción de la asimetría tras aplicar la transformación logarítmica. Los resultados son los que se ven en la Tabla 4 y la Figura 1 del apartado 4.1.

```
import pandas as pd
import numpy as np

SEGUIDORES = 18000000

df = pd.read_excel("Corpus_Final_TFG.xlsx")
df = df[df['tiene_imagen'] == True].copy() # publicaciones codificadas visualmente
df = df[df['Likes'] >= 0].copy() # eliminar Likes invalidos
df['Engagement_total'] = df['Likes'] + df['Comentarios']
df['Engagement_rate'] = (df['Engagement_total'] / SEGUIDORES) * 100

# Estadisticos descriptivos
for col in ['Likes', 'Comentarios', 'Engagement_total', 'Engagement_rate']:
    print(f"{col}: media={df[col].mean():.0f}, mediana={df[col].median():.0f}, "
          f"min={df[col].min()}, max={df[col].max()}")

# Comprobacion de la asimetria antes y despues de la transformacion log
print(f"Asimetria original: {df['Engagement_total'].skew():.2f}")
print(f"Asimetria tras log: {np.log1p(df['Engagement_total']).skew():.2f}")
```

A.II.5. Análisis bivariado

Este script aplica la prueba no paramétrica de Kruskal-Wallis a cada variable independiente para comprobar si las diferencias de engagement entre sus categorías son estadísticamente significativas. Además, calcula el tamaño del efecto mediante el coeficiente epsilon cuadrado, que permite comparar la magnitud de la asociación entre variables y construir la Tabla 3 que aparece al inicio del capítulo 4.

```
import pandas as pd
```

```

from scipy import stats

df = pd.read_excel("Corpus_Final_TFG.xlsx")
df = df[df['tiene_imagen'] == True].copy()
df = df[df['Likes'] >= 0].copy()
df['Engagement_total'] = df['Likes'] + df['Comentarios']

variables = ['Formato', 'Tematica_principal', 'Finalidad_principal',
            'Presencia_jugadores', 'Presencia_aficion', 'Presencia_otras_personas',
            'Contexto_deportivo', 'Hashtags_cat', 'CTA', 'Texto_longitud',
            'Momento_dia', 'Tipo_dia']

# Para cada variable: Kruskal-Wallis y tamaño del efecto
for var in variables:
    grupos = [g['Engagement_total'].values for _, g in df.groupby(var)]
    H, p = stats.kruskal(*grupos)
    n = len(df)
    k = len(grupos)
    epsilon2 = (H - k + 1) / (n - k)
    print(f"{var}: H={H:.2f}, p={p:.4f}, epsilon2={epsilon2:.4f}")

# Tabla con engagement mediano por categoría para cada variable
for var in variables:
    tabla = df.groupby(var).agg(N=('Engagement_total', 'size'),
                               Mediana=('Engagement_total', 'median'))
    print(f"\n{var}:")
    print(tabla.sort_values('Mediana', ascending=False))

```

A.II.6. Comprobación de multicolinealidad

Antes de estimar el modelo de regresión múltiple se comprueba que las variables independientes no presentan multicolinealidad severa, mediante el factor de inflación de la varianza (VIF). Valores por debajo de 10 indican ausencia de problemas (O'Brien, 2007).

```

import pandas as pd
import numpy as np
import patsy
from statsmodels.stats.outliers_influence import variance_inflation_factor

df = pd.read_excel("Corpus_Final_TFG.xlsx")
df = df[df['tiene_imagen'] == True].copy()
df = df[df['Likes'] >= 0].copy()
df['Engagement_total'] = df['Likes'] + df['Comentarios']
df['log_Eng'] = np.log1p(df['Engagement_total'])

# Reagrupar categorias residuales mal codificadas por el modelo
df['Contexto_deportivo'] = df['Contexto_deportivo'].replace({
    'post_partido': 'partido_en_juego_o_cronica',
    'post_partido_o_cronica': 'partido_en_juego_o_cronica',
    'post_partido_en_juego_o_cronica': 'partido_en_juego_o_cronica',
    'post_partido_en_juego': 'partido_en_juego_o_cronica',
    'cronica_partido_en_juego': 'partido_en_juego_o_cronica'
})

formula = ("log_Eng ~ C(Formato) + C(Tematica_principal) + C(Contexto_deportivo) "
          "+ C(Presencia_jugadores) + C(Presencia_aficion) + C(Hashtags_cat) "
          "+ C(CTA) + C(Texto_longitud) + C(Momento_dia) + C(Tipo_dia)")

y, X = patsy.dmatrices(formula, df, return_type='dataframe')

# Calcular el VIF para cada variable dummy
for i, col in enumerate(X.columns):
    if col == 'Intercept':
        continue
    vif = variance_inflation_factor(X.values, i)
    print(f" {col}: VIF={vif:.2f}")

```

A.II.7. Regresión lineal múltiple

Este script estima el modelo de regresión lineal múltiple sobre el engagement transformado logarítmicamente, controlando simultáneamente por todas las variables independientes. Los coeficientes del modelo se interpretan como cambios porcentuales en el engagement respecto a la categoría de referencia de cada variable, ya que la variable dependiente está en escala logarítmica. Los resultados de este script son los que se ven en la tabla de coeficientes y la Figura 4 del apartado 4.8.

```
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf

df = pd.read_excel("Corpus_Final_TFG.xlsx")
df = df[df['tiene_imagen'] == True].copy()
df = df[df['Likes'] >= 0].copy()
df['Engagement_total'] = df['Likes'] + df['Comentarios']
df['log_Eng'] = np.log1p(df['Engagement_total'])

# Reagrupar categorías residuales mal codificadas por el modelo
df['Contexto_deportivo'] = df['Contexto_deportivo'].replace({
    'post_partido': 'partido_en_juego_o_cronica',
    'post_partido_o_cronica': 'partido_en_juego_o_cronica',
    'post_partido_en_juego_o_cronica': 'partido_en_juego_o_cronica',
    'post_partido_en_juego': 'partido_en_juego_o_cronica',
    'cronica_partido_en_juego': 'partido_en_juego_o_cronica'
})

# Definir categorías de referencia (la primera de cada lista es la referencia)
df['Formato'] = pd.Categorical(df['Formato'],
    categories=['imagen','carrusel','reel'])
df['Tematica_principal'] = pd.Categorical(df['Tematica_principal'],
    categories=['partido','jugadores_plantilla','entrenamiento','aficion_comunidad',
        'celebracion','merchandising_patrocinio','institucional',
        'fichaje_renovacion','otro'])
```

```

df['Contexto_deportivo'] = pd.Categorical(df['Contexto_deportivo'],
    categories=['previa_partido','partido_en_juego_o_cronica','post_victoria',
        'post_derrota','entrenamiento_preparacion','fichaje_mercado',
        'no_deportivo_institucional_comercial','otro_no_identificable'])
df['Presencia_jugadores'] = pd.Categorical(df['Presencia_jugadores'],
    categories=['no','si','no_identificable'])
df['Presencia_aficion'] = pd.Categorical(df['Presencia_aficion'],
    categories=['no','si','no_identificable'])
df['Hashtags_cat'] = pd.Categorical(df['Hashtags_cat'], categories=['0','1-2','3+'])
df['CTA'] = pd.Categorical(df['CTA'], categories=['no','si'])
df['Texto_longitud'] = pd.Categorical(df['Texto_longitud'],
    categories=['breve','media','larga'])
df['Momento_dia'] = pd.Categorical(df['Momento_dia'],
    categories=['manana','tarde','noche'])
df['Tipo_dia'] = pd.Categorical(df['Tipo_dia'],
    categories=['laborable','fin_de_semana'])

# Estimar el modelo de regresion
formula = ("log_Eng ~ C(Formato) + C(Tematica_principal) + C(Contexto_deportivo) "
    "+ C(Presencia_jugadores) + C(Presencia_aficion) + C(Hashtags_cat) "
    "+ C(CTA) + C(Texto_longitud) + C(Momento_dia) + C(Tipo_dia)")
modelo = smf.ols(formula, data=df).fit()

print(f"R-cuadrado: {modelo.rsquared:.4f}")
print(f"F: {modelo.fvalue:.2f}, p: {modelo.f_pvalue:.4e}")

# Coeficientes con su efecto porcentual sobre el engagement
for nombre, coef, p in zip(modelo.params.index, modelo.params.values,
    modelo.pvalues.values):
    if nombre == 'Intercept':
        continue
    efecto = (np.exp(coef) - 1) * 100
    print(f"{nombre}: coef={coef:.3f}, efecto={efecto:+.1f}%, p={p:.4f}")

```

A.II.8. Generación de las figuras

Este script genera las cuatro figuras incluidas en el capítulo de resultados.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.ticker as mtick

# Configuración visual general
plt.rcParams.update({
    'font.family': 'Arial',
    'font.size': 11,
    'axes.titlesize': 13,
    'axes.spines.top': False,
    'axes.spines.right': False,
})

# Paleta de colores
AZUL = '#1F4E8C'
VERDE = '#4F8A3A'
NARANJA = '#E07B00'
MORADO = '#5E3A8C'
ROJO = '#C8102E'
GRIS = '#888888'

# Carga del dataset y preparación
df = pd.read_excel('Corpus_Final_TFG.xlsx')
df = df[df['tiene_imagen'] == True].copy()
df = df[df['Likes'] >= 0].copy()
df['Engagement_total'] = df['Likes'] + df['Comentarios']
df['log_Eng'] = np.log1p(df['Engagement_total'])

# Reagrupar categorías residuales para las figuras del contexto deportivo
df['Contexto_deportivo'] = df['Contexto_deportivo'].replace({
```

```

'post_partido': 'partido_en_juego_o_cronica',
'post_partido_o_cronica': 'partido_en_juego_o_cronica',
'post_partido_en_juego_o_cronica': 'partido_en_juego_o_cronica',
'post_partido_en_juego': 'partido_en_juego_o_cronica',
'cronica_partido_en_juego': 'partido_en_juego_o_cronica'
})

# Nombres legibles para las categorias del contexto
nombres = {
    'previa_partido': 'Previa de partido',
    'entrenamiento_preparacion': 'Entrenamiento',
    'no_deportivo_institucional_comercial': 'No deportivo',
    'partido_en_juego_o_cronica': 'Partido / cronica',
    'post_derrota': 'Post-derrota',
    'post_victoria': 'Post-victoria',
    'fichaje_mercado': 'Fichaje / mercado',
    'otro_no_identificable': 'Otro / no ident.'
}

# FIGURA 1: histograma engagement original y log
fig, axes = plt.subplots(1, 2, figsize=(12, 4.5))
axes[0].hist(df['Engagement_total'], bins=80, color=ROJO, edgecolor='white')
axes[0].set_title('Distribucion original del engagement', fontweight='bold')
axes[0].set_xlabel('Engagement total (likes + comentarios)')
axes[0].set_ylabel('Numero de publicaciones')
axes[0].axvline(df['Engagement_total'].median(), color=AZUL, ls='--',
                label=f'Mediana = {df["Engagement_total"].median():.0f}')
axes[0].axvline(df['Engagement_total'].mean(), color=GRIS, ls=':',
                label=f'Media = {df["Engagement_total"].mean():.0f}')
axes[0].legend()
axes[1].hist(df['log_Eng'], bins=40, color=AZUL, edgecolor='white')
axes[1].set_title('Distribucion tras la transformacion logaritmica', fontweight='bold')
axes[1].set_xlabel('log(engagement + 1)')
axes[1].set_ylabel('Numero de publicaciones')

```

```

plt.tight_layout()
plt.savefig('Figura_1_engagement_log.png', dpi=200, bbox_inches='tight')
plt.close()

# FIGURA 2: engagement mediano por contexto deportivo
medianas = df.groupby('Contexto_deportivo')['Engagement_total'].median().sort_values()
medianas.index = [nombres.get(i, i) for i in medianas.index]
mediana_global = medianas.median()
colores = [AZUL if v >= mediana_global else VERDE for v in medianas.values]
fig, ax = plt.subplots(figsize=(9, 5))
ax.barh(medianas.index, medianas.values, color=colores, edgecolor='white')
ax.set_xlabel('Engagement mediano (likes + comentarios)')
ax.set_title('Engagement mediano segun contexto deportivo', fontweight='bold')
ax.xaxis.set_major_formatter(mtick.FuncFormatter(lambda x, _: f'{int(x):,}'))
for i, v in enumerate(medianas.values):
    ax.text(v + 400, i, f'{int(v):,}', va='center', fontsize=10)
plt.tight_layout()
plt.savefig('Figura_2_contexto_deportivo.png', dpi=200, bbox_inches='tight')
plt.close()

# FIGURA 3: ratio comentarios sobre likes por contexto
likes_med = df.groupby('Contexto_deportivo')['Likes'].mean()
com_med = df.groupby('Contexto_deportivo')['Comentarios'].mean()
ratio = (com_med / likes_med * 100).sort_values(ascending=False)
ratio.index = [nombres.get(i, i) for i in ratio.index]
colores = [NARANJA if v > 5 else MORADO if v > 1 else GRIS for v in ratio.values]
fig, ax = plt.subplots(figsize=(9, 5))
ax.barh(ratio.index, ratio.values, color=colores, edgecolor='white')
ax.set_xlabel('Ratio comentarios/likes (%)')
ax.set_title('Engagement activo (comentarios) vs pasivo (likes) por contexto',
            fontweight='bold')
for i, v in enumerate(ratio.values):
    ax.text(v + 0.1, i, f'{v:.1f}%', va='center', fontsize=10)
plt.tight_layout()

```

```

plt.savefig('Figura_3_pasivo_activo.png', dpi=200, bbox_inches='tight')
plt.close()

# FIGURA 4: coeficientes significativos del modelo de regresion
coefs = [
    ('Post-derrota', 92.3, '***'),
    ('Otro no identif.', 73.7, '***'),
    ('Post-victoria', 73.2, '***'),
    ('No deportivo', 71.6, '***'),
    ('Fichaje mercado', 66.2, '*'),
    ('Partido juego/cronica', 65.5, '***'),
    ('Noche (vs manana)', 35.8, '***'),
    ('Otro (tematica)', 31.8, '*'),
    ('Hashtags 1-2', 20.8, '*'),
    ('Texto medio', 19.5, '***'),
    ('Texto largo', 15.1, '**'),
    ('Tarde (vs manana)', -8.3, '*'),
    ('Fin de semana', -10.4, '**'),
    ('Aficion/comunidad', -18.9, '*'),
    ('Institucional', -21.4, '*'),
    ('Merchandising', -52.5, '***'),
]
coefs.sort(key=lambda x: x[1])
labels = [c[0] for c in coefs]
values = [c[1] for c in coefs]
fig, ax = plt.subplots(figsize=(9, 7))
colores = [ROJO if v > 0 else AZUL for v in values]
ax.barh(labels, values, color=colores, edgecolor='white')
ax.axvline(0, color='black', lw=0.8)
ax.set_xlabel('Efecto sobre el engagement (%) respecto a la categoria de referencia')
ax.set_title('Efectos significativos del modelo de regresion multiple', fontweight='bold')
ax.xaxis.set_major_formatter(mtick.PercentFormatter(decimals=0))
for i, (v, sig) in enumerate(zip(values, [c[2] for c in coefs])):
    offset = 2 if v > 0 else -2

```

```
ha = 'left' if v > 0 else 'right'  
ax.text(v + offset, i, f'{v:+.1f}% {sig}', va='center', ha=ha, fontsize=9)  
ax.set_xlim(-65, 110)  
plt.tight_layout()  
plt.savefig('Figura_4_coeficientes.png', dpi=200, bbox_inches='tight')  
plt.close()
```