



Facultad de Ciencias Económicas y Empresariales

ANÁLISIS DE LA RELACIÓN ENTRE ACTIVIDAD FUTBOLÍSTICA Y LLAMADAS AL SERVICIO 016 EN EL PAÍS VASCO

Autor: María Guadalupe Gut Lladó

Tutor: Riccardo Ciacci

MADRID | Junio de 2026

RESUMEN

El presente trabajo analiza la posible relación entre la actividad futbolística y las llamadas pertinentes registradas en el Servicio 016 en el País Vasco durante el periodo 2021-2026. Para ello, se construyó una base de datos diaria integrando información oficial del Servicio 016 con una base futbolística elaborada específicamente para el estudio a partir de datos obtenidos mediante la API deportiva API-Football.

La base final incorpora información diaria para las provincias de Álava, Bizkaia y Gipuzkoa, incluyendo variables relacionadas con la existencia de actividad futbolística, resultados deportivos, goles anotados y balance de goles. A partir de esta información se realizaron análisis descriptivos, comparaciones de medias y estimaciones econométricas mediante modelos de regresión lineal OLS.

Los resultados muestran que, aunque existen pequeñas diferencias descriptivas entre determinados contextos futbolísticos, las principales variables deportivas analizadas no presentan una relación estadísticamente significativa con el número diario de llamadas pertinentes registradas en el Servicio 016 una vez incorporados los controles temporales y territoriales. En cambio, sí aparecen diferencias asociadas a determinadas variables estructurales, especialmente entre provincias y entre fines de semana y días laborables.

El trabajo aporta además una base de datos original y una metodología de integración de información institucional y deportiva que pueden servir como punto de partida para futuras investigaciones relacionadas con acontecimientos deportivos e indicadores sociales.

PALABRAS CLAVE

Violencia de género, Servicio 016, fútbol, análisis econométrico, País Vasco, datos diarios.

ÍNDICE

CAPÍTULO I. INTRODUCCIÓN	3
I.1. VIOLENCIA DE GÉNERO, FÚTBOL Y CONTEXTO SOCIAL	3
I.2. OBJETIVOS E HIPÓTESIS DEL TRABAJO	4
I.3. ESTRUCTURA DEL TRABAJO	4
CAPÍTULO II. DATOS Y CONSTRUCCIÓN DE LA BASE	6
II.1. DATOS DEL SERVICIO 016	6
II.2. DATOS FUTBOLÍSTICOS	7
II.2.1. Construcción de la base futbolística	7
II.2.2. Limpieza y preparación de la base futbolística	8
II.3. CONSTRUCCIÓN DE LA BASE FINAL	10
II.4. CONSTRUCCIÓN DE VARIABLES	11
CAPÍTULO III. METODOLOGÍA	13
III.1. ESTRATEGIA EMPÍRICA	13
III.2. ANÁLISIS DESCRIPTIVO	14
III.3. COMPARACIÓN DE MEDIDAS Y CONTRASTES ESTADÍSTICOS	14
III.4. MODELO ECONOMETRICO	15
III.5. LIMITACIONES METODOLÓGICAS	16
CAPÍTULO IV. ANÁLISIS EMPÍRICO Y RESULTADOS	18
IV.1. DESCRIPCIÓN GENERAL DE LA MUESTRA	18
IV.2. EVOLUCIÓN TEMPORAL DE LAS LLAMADAS	19
IV.3. ACTIVIDAD FUTBOLÍSTICA Y LLAMADAS AL 016	21
IV.4. FINES DE SEMANA Y DERBIS VASCOS	23
IV.5. CORRELACIONES Y ANÁLISIS PROVINCIAL	24
IV.6. RESULTADOS DE LA REGRESIÓN	27
CAPÍTULO V. DISCUSIÓN	29
V.1. INTERPRETACIÓN GENERAL DE LOS RESULTADOS	29
V.2. INTERPRETACIÓN Y DISCUSIÓN DE LOS RESULTADOS	30
V.3. LIMITACIONES DEL TRABAJO	31
CAPÍTULO VI. CONCLUSIONES	33
VI.1. CONCLUSIONES PRINCIPALES	33
VI.2. APORTACIONES DEL TRABAJO	34
ANEXO	35
A. TABLAS	35
B. OUTPUTS DE LA REGRESIÓN OLS	38
C. GRÁFICOS	38
D. SCRIPTS	42

CAPÍTULO I. INTRODUCCIÓN

I.1. VIOLENCIA DE GÉNERO, FÚTBOL Y CONTEXTO SOCIAL

La violencia de género representa actualmente uno de los principales problemas sociales en España, tanto por las consecuencias que tiene sobre las víctimas como por el impacto social e institucional que genera. Durante las últimas décadas, este fenómeno ha adquirido una presencia cada vez mayor dentro del debate público, impulsando el desarrollo de distintas políticas orientadas a la prevención, la atención y la protección de las víctimas. En este contexto, las instituciones públicas han ido incorporando nuevos mecanismos de apoyo y asistencia con el objetivo de mejorar la capacidad de respuesta frente a este tipo de situaciones.

Entre estos recursos destaca el Servicio 016, un servicio oficial de información, atención y asesoramiento especializado en violencia de género. Su actividad se desarrolla de forma continuada en todo el territorio nacional y registra diariamente un elevado número de llamadas relacionadas con situaciones de violencia de género, solicitudes de ayuda o demandas de información. La disponibilidad de datos diarios desagregados por provincias permite, además, analizar cómo evoluciona esta actividad a lo largo del tiempo y estudiar posibles diferencias territoriales o temporales en el comportamiento de las llamadas registradas.

Al mismo tiempo, el fútbol ocupa una posición especialmente relevante dentro de la sociedad española y, de forma más concreta, dentro del contexto social y cultural del País Vasco. Más allá del ámbito estrictamente deportivo, el fútbol genera una importante movilización social y emocional entre aficionados y seguidores. Determinados encuentros pueden concentrar niveles elevados de tensión, expectación o rivalidad, especialmente cuando se trata de partidos con una fuerte carga competitiva o simbólica. Los resultados deportivos también pueden provocar reacciones emocionales intensas, tanto positivas como negativas, que trascienden el propio acontecimiento deportivo.

En los últimos años, algunos estudios internacionales y distintos debates sociales han planteado la posibilidad de que acontecimientos deportivos de elevada carga emocional puedan estar relacionados con cambios en determinados comportamientos sociales. Aunque esta posible relación ha generado interés en distintos ámbitos sociales y académicos, analizar este tipo de fenómenos resulta complejo debido a la dificultad de estudiar comportamientos sociales mediante información observable y datos cuantitativos diarios.

Partiendo de este contexto, el presente trabajo analiza la posible relación entre la actividad futbolística y las llamadas pertinentes registradas en el Servicio 016 en el País Vasco durante el periodo 2021-2026. Para ello, se construye una base de datos diaria que integra información institucional y deportiva procedente de distintas fuentes. El objetivo principal es estudiar si determinados contextos relacionados con la actividad futbolística presentan algún

tipo de asociación observable con la evolución diaria de las llamadas registradas en el Servicio 016.

I.2. OBJETIVOS E HIPÓTESIS DEL TRABAJO

El objetivo principal del presente trabajo consiste en analizar si existe algún tipo de relación entre la actividad futbolística y las llamadas pertinentes registradas en el Servicio 016 en el País Vasco durante el periodo 2021-2026.

Para ello, se construye una base de datos diaria integrando información institucional y deportiva procedente de distintas fuentes, permitiendo estudiar conjuntamente la evolución temporal de las llamadas y distintos indicadores relacionados con partidos de fútbol, resultados deportivos y variables asociadas a la actividad futbolística.

Junto al objetivo principal, el trabajo plantea también varios objetivos específicos. En primer lugar, se pretende analizar si existen diferencias entre días con y sin actividad futbolística. En segundo lugar, se estudia si determinados resultados deportivos, como victorias, empates o derrotas, presentan comportamientos diferenciados respecto al volumen diario de llamadas registradas. Finalmente, el análisis incorpora también comparaciones territoriales entre las provincias del País Vasco y diferencias asociadas a fines de semana y días laborables.

A partir de estos objetivos, el trabajo plantea varias hipótesis iniciales. La primera hipótesis considera la posibilidad de que los días con actividad futbolística presenten diferencias respecto a los días sin fútbol en términos de llamadas pertinentes registradas en el Servicio 016. La segunda hipótesis plantea que determinados contextos deportivos asociados a resultados negativos podrían relacionarse con mayores niveles de llamadas. Por último, también se considera la posibilidad de que variables territoriales y temporales influyan sobre la evolución diaria observada en los datos.

I.3. ESTRUCTURA DEL TRABAJO

El trabajo se organiza en varios capítulos. La introducción presenta el contexto general del estudio, así como los principales objetivos e hipótesis planteados inicialmente.

El segundo capítulo describe las bases de datos utilizadas y el proceso de construcción de la base final empleada en el análisis. En este apartado se explica tanto la obtención y limpieza de los datos del Servicio 016 como la recopilación, preparación e integración de la información futbolística mediante scripts desarrollados en Python.

El tercer capítulo expone la metodología utilizada en el trabajo, incluyendo el análisis descriptivo, las comparaciones de medias, los contrastes estadísticos y el modelo

econométrico estimado. El cuarto capítulo recoge el análisis empírico y los principales resultados obtenidos a partir de la información analizada, incorporando tablas, gráficos y estimaciones econométricas.

Posteriormente, el quinto capítulo desarrolla la discusión e interpretación de los resultados, así como las principales limitaciones del trabajo. Y por último, el sexto capítulo presenta las conclusiones principales y las aportaciones realizadas por el estudio.

CAPÍTULO II. DATOS Y CONSTRUCCIÓN DE LA BASE

II.1. DATOS DEL SERVICIO 016

Los datos utilizados en este trabajo corresponden a registros diarios del Servicio 016, obtenidos a partir de las bases publicadas por el Ministerio de Igualdad de España. El Servicio 016 es un recurso oficial de atención, información y asesoramiento en materia de violencia de género, dirigido principalmente a víctimas, familiares y personas del entorno cercano. Entre los distintos indicadores disponibles, este trabajo utiliza específicamente el número de llamadas pertinentes registradas diariamente en cada provincia.

Las llamadas pertinentes hacen referencia a aquellas comunicaciones consideradas relevantes dentro de la actividad de atención relacionada con violencia de género. Esta variable permite aproximar el volumen diario de demanda de atención y asistencia asociado a posibles situaciones de violencia de género.

La base de datos original contiene información diaria desagregada por provincias españolas, incluyendo variables temporales y territoriales como año, mes, día y provincia. Posteriormente, se añadió la variable de comunidad autónoma a partir de una asignación territorial de provincias. En este trabajo, el análisis se limita al País Vasco, por lo que únicamente se seleccionaron las observaciones correspondientes a Álava, Bizkaia y Gipuzkoa.

El periodo analizado abarca desde 2021 hasta 2026, lo que permite trabajar con una muestra suficientemente amplia para estudiar patrones temporales y comparar días con y sin actividad futbolística. Además, la frecuencia diaria de los datos resulta especialmente útil para relacionar la fecha de los partidos con el número de llamadas registradas en cada provincia.

Antes de realizar el análisis, los datos fueron limpiados y preparados para su tratamiento posterior. En primer lugar, se revisó la estructura de la base y los tipos de variables con el objetivo de detectar posibles problemas de formato. Se observó que algunas variables temporales aparecían almacenadas como texto, por ejemplo, el año, que figuraba en formatos como “Año 2021”. Por ello, fue necesario eliminar la parte textual y convertir esa variable a formato numérico.

También se creó un diccionario de equivalencias para transformar los meses escritos en texto a formato numérico. Este paso era necesario para construir correctamente una variable de fecha completa y trabajar posteriormente con series temporales diarias.

Una vez transformadas las variables temporales, se creó una nueva variable denominada *fecha*, combinando el año, el mes y el día de cada observación. Esta variable resultaba fundamental para ordenar cronológicamente la base y facilitar posteriormente el cruce con la base futbolística mediante una referencia temporal común.

Después, los datos se ordenaron por provincia y fecha para facilitar el análisis temporal y la integración posterior con los datos deportivos. Además, se seleccionaron únicamente las

variables necesarias para el trabajo, eliminando aquellas columnas que no resultaban relevantes para el análisis empírico.

Por otro lado, se incorporó una variable correspondiente a la comunidad autónoma mediante un proceso de asignación territorial de provincias. Esto permitió filtrar posteriormente únicamente las observaciones pertenecientes al País Vasco, que constituye el ámbito geográfico del estudio.

Finalmente, la base limpia fue exportada en formato CSV para utilizarla en las siguientes fases del análisis y en la construcción de la base final integrada.

II.2. DATOS FUTBOLÍSTICOS

II.2.1. Construcción de la base futbolística

La base de datos futbolística utilizada en el presente trabajo fue construida específicamente para el análisis a partir de información obtenida mediante la API deportiva *API-Football*. Esta plataforma proporciona acceso a datos históricos de partidos, resultados y eventos futbolísticos a través de consultas automatizadas, aunque el acceso completo a determinadas funcionalidades requiere una suscripción de pago.

La recopilación de datos se realizó mediante scripts desarrollados en Python. El objetivo era construir una base homogénea que reuniese información detallada de los principales equipos españoles a lo largo de distintas temporadas. Para ello, se desarrolló un script base y, en cada ejecución, se modificaban manualmente distintos parámetros en función del equipo y la temporada que se quisiesen descargar. En concreto, se cambiaban variables como el nombre del equipo (*EQUIPO_REFERENCIA*), el identificador oficial utilizado por la API (*TEAM_ID*) y la temporada analizada (*SEASON*). Además, también se definían otras variables auxiliares como la división, la comunidad autónoma y el formato de la temporada.

Una vez definidos esos parámetros, el script realizaba consultas automáticas a la API mediante solicitudes HTTP. Esto se conseguía usando dos endpoints principales: uno destinado a obtener información general de los partidos (*fixtures*) y otro destinado a recuperar los eventos específicos de cada encuentro (*fixtures/events*). Esta separación permite obtener tanto información agregada del partido como información detallada relacionada con los goles.

Posteriormente, al ejecutar el script, se recorría automáticamente distintas competiciones almacenadas en un diccionario de ligas, incluyendo La Liga, Copa del Rey, Supercopa, Champions League y Europa League. Para cada competición, se descargaban todos los partidos correspondientes al equipo y temporada seleccionados.

A partir de la información descargada, se extraían distintas variables relacionadas con cada partido. Entre ellas se encontraban la temporada (*Temporada*), utilizada para identificar el periodo competitivo analizado; el equipo de referencia (*Equipo referencia*), correspondiente al equipo sobre el que se construía cada base; la división (*División*); el mes y número de mes;

el año; la fecha del partido (*Fecha*); la jornada (*Jornada*); el equipo local (*Equipo_Local*) y visitante (*Equipo_Visitante*); el resultado final (*Resultado*); la hora de inicio del partido (*Hora_Inicio*); la competición disputada (*Competición*); y distintas variables relacionadas con los goles, como el minuto exacto del gol (*Minuto_Gol*), el equipo que marcaba (*Equipo_Marcador*) y si el gol era a favor o en contra del equipo analizado (*Gol A FAVOR / EN CONTRA*). También se incorporaron variables temporales adicionales, como las franjas horarias del partido y de los goles, utilizadas posteriormente para agrupar temporalmente los eventos deportivos.

Las fechas originales obtenidas a través de la API no tenían un formato compatible con el análisis que se iba a realizar, por lo que fueron transformadas al formato *datetime* de Python. Este formato permite almacenar las fechas como variables temporales estructuradas, facilitando operaciones cronológicas como ordenar observaciones, comparar fechas o trabajar con series temporales. Esta transformación permitiría posteriormente ordenar cronológicamente los partidos, trabajar correctamente con fechas y facilitar el cruce posterior con la base diaria del Servicio 016.

Después, era necesario recuperar los eventos específicos asociados a cada partido, especialmente los goles, porque la información general de los encuentros no incluía suficiente detalle para construir determinadas variables utilizadas posteriormente en el análisis. Para ello, se realizaba una segunda consulta utilizando el identificador único de cada partido (*fixture_id*), el cual permitía acceder a todos los eventos registrados durante ese partido concreto. Una vez descargados, se filtraban únicamente aquellos relacionados con goles, extrayendo información como el equipo que marcaba, el minuto exacto del gol y si el tanto era a favor o en contra del equipo analizado.

A continuación, se desarrollaron funciones específicas para transformar las horas exactas de los partidos y los goles en franjas horarias homogéneas de treinta minutos. El objetivo de esta clasificación era agrupar temporalmente los eventos deportivos y construir variables temporales más fáciles de analizar posteriormente. De esta manera, tanto el inicio de los partidos como el momento de los goles quedaban asociados a intervalos horarios comunes.

Toda la información obtenida de cada competición y temporada se iba incorporando progresivamente a una misma base de datos en Excel hasta construir un dataset completo para cada equipo analizado. Finalmente, cada una de estas bases fue exportada individualmente en formato Excel y posteriormente integrada en un archivo maestro utilizado como base futbolística principal del trabajo.

II.2.2. Limpieza y preparación de la base futbolística

Una vez construida la base futbolística general y consolidados todos los archivos individuales en un único Excel maestro, fue necesario realizar un proceso de limpieza y preparación de los datos antes de comenzar el análisis empírico. Para ello, se trabajó únicamente con la hoja

denominada “Registro de Goles”, que contenía la información completa de partidos, goles y variables temporales relevantes para el estudio.

En primer lugar, se importó la base de datos a Python mediante la librería *pandas* y se revisó la estructura general del dataset. Esta comprobación inicial permitió identificar el número de observaciones, los tipos de variables disponibles y algunos posibles problemas de formato. La base incluía información correspondiente a distintas temporadas, equipos, competiciones y eventos de gol, además de variables temporales y territoriales.

Durante esta revisión se observó que la variable *Fecha* estaba almacenada como texto y no como una variable temporal utilizable para el análisis cronológico. Por este motivo, fue necesario convertirla al formato *datetime* utilizando el parámetro *dayfirst=True*, ya que las fechas originales estaban registradas en formato día/mes/año. Esta transformación permitió trabajar correctamente con fechas, ordenar cronológicamente las observaciones y facilitar posteriormente el cruce con la base diaria del Servicio 016.

Después, se reorganizaron las columnas de la base para mantener una estructura más clara y facilitar el tratamiento posterior de los datos. Se seleccionaron únicamente las variables relevantes para el análisis, eliminando información innecesaria y manteniendo aquellas relacionadas con la fecha, la temporada, los equipos participantes, la competición, los goles y las variables temporales.

Más adelante, la base fue ordenada cronológicamente por equipo y fecha. Este paso permitió mantener una estructura temporal consistente y facilitar la construcción de variables diarias relacionadas con la actividad futbolística.

También se realizó un proceso de homogeneización territorial de las comunidades autónomas. Durante la revisión de la base se observó que una misma comunidad autónoma podía aparecer escrita de dos formas distintas, como ocurría con “Comunitat Valenciana” y “Comunidad Valenciana”. Aunque ambas denominaciones hacen referencia al mismo territorio, mantenerlas separadas podía generar duplicidades o errores en los análisis agrupados por comunidad autónoma. Para corregirlo, dentro del propio script de limpieza se indicó que todos los casos registrados como “Comunitat Valenciana” pasaran a figurar como “Comunidad Valenciana”. De esta forma, la base mantenía una única denominación para cada comunidad autónoma. Después de realizar esta corrección, se revisó el listado final de comunidades registradas para comprobar que no quedaran nombres duplicados o inconsistentes.

Finalmente, una vez completado el proceso de limpieza y validación, la base fue exportada en formato CSV bajo el nombre *futbol_limpio.csv*. Este archivo constituyó la versión definitiva de la base futbolística utilizada posteriormente en la construcción de la base integrada final del trabajo.

II.3. CONSTRUCCIÓN DE LA BASE FINAL

Una vez preparadas por separado la base de datos del Servicio 016 y la base futbolística, el siguiente paso consistió en integrar ambas fuentes en un único dataset. El objetivo era construir una base diaria común que permitiese relacionar el número de llamadas pertinentes registradas con la actividad futbolística correspondiente a la misma fecha y territorio.

Para ello, se importaron en Python los archivos previamente limpiados *016_diario_limpio.csv* y *futbol_limpio.csv*. Tras cargar ambas bases, se revisó la estructura de las variables y se comprobó que las fechas presentasen un formato homogéneo antes de realizar el cruce. Aunque las variables temporales ya habían sido tratadas previamente durante la fase de limpieza, fue necesario volver a convertir la variable *Fecha* al formato *datetime*, ya que al importar archivos CSV Python interpreta inicialmente las fechas como texto. Esta conversión permitió asegurar la compatibilidad temporal entre ambas bases y facilitar el merge.

La base del Servicio 016 incorporaba inicialmente información provincial, por lo que también fue necesario añadir una variable correspondiente a la comunidad autónoma mediante un diccionario de equivalencias entre provincias y territorios. Esto era importante porque el cruce final no se hacía únicamente utilizando la fecha como referencia, sino también la comunidad autónoma. Así, cada observación diaria podía vincularse con la actividad futbolística registrada en el mismo territorio.

La información futbolística tuvo que transformarse previamente antes de integrarse con la base del Servicio 016. La base original contenía múltiples observaciones para una misma fecha, ya que cada fila representaba un evento de gol concreto. Esto obligó a agregar la información a nivel diario, agrupando las observaciones por fecha y CCAA, construyendo así una estructura temporal comparable con la utilizada en la base institucional.

A partir de este proceso se generaron distintas variables relacionadas con la actividad futbolística. En primer lugar, se construyeron las variables *gol_favor* y *gol_contra*, utilizadas para identificar si cada gol correspondía a un tanto marcado o recibido por el equipo analizado. Después, estas variables se agregaron para calcular el número total de goles a favor (*goles_favor*) y goles en contra (*goles_contra*) registrados diariamente en cada comunidad autónoma. También se calculó la variable *goles_totales*, que indica el total de goles registrados en cada fecha.

Además, también se conservaron algunos elementos descriptivos relacionados con la actividad futbolística presente en cada día. Se agruparon y combinaron los nombres de los equipos participantes (*equipos*) y de las competiciones disputadas (*competiciones*). Esto permitía identificar qué actividad futbolística aparecía registrada en cada observación diaria.

Una vez construida la base futbolística diaria, se realizó el cruce definitivo entre ambas fuentes mediante un *merge* utilizando como claves comunes la fecha y la comunidad autónoma. Se utilizó una unión de tipo *left*, tomando como referencia principal la base del Servicio 016. Esto fue para conservar todas las observaciones diarias de llamadas, incluso en aquellas fechas donde no existía actividad futbolística registrada.

Tras realizar el *merge*, las variables futbolísticas que presentaban valores faltantes fueron reemplazadas por cero. Estos valores aparecían principalmente en días sin actividad futbolística dentro de la base. Después, se creó la variable *hay_futbol*, que toma valor 1 cuando existe actividad futbolística registrada y valor 0 cuando no.

También se construyó la variable *balance_goles*, calculada como la diferencia entre goles a favor y goles en contra. A partir de esta variable se generó *resultado_partido*, utilizada para clasificar cada observación diaria según el resultado deportivo registrado. Las observaciones quedaron clasificadas como *Victoria*, *Derrota*, *Empate* o *Sin fútbol* dependiendo del balance de goles y de la existencia de actividad futbolística en esa fecha.

Por último, la base integrada fue exportada bajo el nombre *master_016_futbol_cruzado.csv*. Este archivo constituye la base final utilizada en el trabajo y contiene información diaria sobre llamadas pertinentes al Servicio 016, variables territoriales y variables relacionadas con la actividad futbolística. A partir de esta base se desarrollaron posteriormente los análisis descriptivos y econométricos del estudio.

II.4. CONSTRUCCIÓN DE VARIABLES

Una vez integrada la base final, se procedió a la construcción de las variables utilizadas después en el análisis descriptivo y econométrico. Estas variables pueden agruparse en variables dependientes, variables relacionadas con la actividad futbolística y variables temporales y territoriales.

La variable dependiente principal utilizada en el trabajo es *Llamadas pertinentes*, correspondiente al número diario de llamadas registradas por el Servicio 016 en cada provincia. Esta variable se utiliza como aproximación cuantitativa a la demanda diaria de atención relacionada con situaciones de violencia de género. Dado que la base se encuentra desagregada diariamente, la variable permite analizar posibles variaciones temporales asociadas a la actividad futbolística.

Entre las variables futbolísticas, una de las principales es *hay_futbol*, construida como una variable dicotómica que toma valor 1 cuando existe actividad futbolística registrada en la base para una fecha y territorio determinados, y valor 0 en caso contrario. Esta variable permite comparar el comportamiento medio de las llamadas en días con y sin actividad futbolística registrada.

También se construyeron variables relacionadas con los goles anotados durante los partidos. La variable *goles_favor* recoge el número total de goles marcados por los equipos analizados en cada fecha, mientras que *goles_contra* representa los goles recibidos. A partir de ambas variables se construyó *goles_totales*, correspondiente al total de goles registrados diariamente.

Posteriormente, se generó la variable *balance_goles*, calculada como la diferencia entre goles a favor y goles en contra. Esta variable permite aproximar el resultado deportivo diario de los equipos analizados y captar posibles diferencias entre resultados positivos y negativos.

Además, se construyó la variable categórica *resultado_partido*, utilizada para clasificar las observaciones diarias en función del balance deportivo registrado. Las observaciones podían clasificarse como *Victoria*, *Derrota*, *Empate* o *Sin fútbol*, dependiendo de la diferencia de goles y de la existencia de actividad futbolística registrada en la base.

También se incorporaron variables temporales adicionales. Entre ellas destaca *fin_semana*, construida a partir de la fecha mediante la identificación del día de la semana. Esta variable toma valor 1 cuando la observación corresponde a sábado o domingo y valor 0 para el resto de días. Su incorporación resulta relevante porque tanto la actividad futbolística como el volumen de llamadas pueden presentar comportamientos distintos durante los fines de semana. Por otro lado, también se utilizaron variables temporales ya presentes en la base, como *Año*, *Mes_num* y *Día*, que permiten analizar posibles patrones estacionales y controlar diferencias temporales entre periodos. Asimismo, se mantuvieron variables territoriales como *Provincia* y *CCAA*, utilizadas para distinguir geográficamente las observaciones y realizar comparaciones entre territorios.

Finalmente, se construyó la variable *derbi_vasco*, utilizada para identificar aquellos días en los que se registraban partidos entre equipos vascos relevantes, concretamente entre Athletic Club y Real Sociedad. Esta variable toma valor 1 cuando en la información deportiva diaria aparece registrado dicho enfrentamiento y valor 0 en caso contrario.

CAPÍTULO III. METODOLOGÍA

III.1. ESTRATEGIA EMPÍRICA

El presente trabajo utiliza una metodología cuantitativa basada en el análisis de datos diarios con el objetivo de estudiar la posible relación entre la actividad futbolística y el número de llamadas pertinentes registradas en el Servicio 016 en el País Vasco. Para ello, se construyó una base de datos integrada que combina información territorial y temporal procedente del Servicio 016 con distintas variables relacionadas con partidos y eventos futbolísticos.

La base utilizada en el análisis está compuesta por observaciones diarias desagregadas por provincia. Trabajar con este nivel de detalle permite observar cómo evolucionan las llamadas registradas en distintos momentos del tiempo y comparar días con características futbolísticas diferentes. Además, el uso de datos diarios resulta especialmente útil en este tipo de análisis porque permite relacionar cada jornada futbolística con las llamadas registradas exactamente en esa misma fecha, evitando trabajar con información demasiado agregada.

A partir de esta base integrada, el análisis empírico se desarrolló en varias fases. En una primera etapa se realizó un análisis descriptivo general con el objetivo de conocer mejor la estructura de los datos y observar el comportamiento de las principales variables incluidas en el estudio. Para ello, se calcularon estadísticas descriptivas y se elaboraron distintos gráficos temporales y comparaciones entre grupos de observaciones.

También se analizaron las diferencias existentes entre días con y sin actividad futbolística registrada. Del mismo modo, se compararon las llamadas pertinentes según el resultado deportivo observado en cada fecha, diferenciando entre victorias, derrotas, empates y días sin fútbol. Estas comparaciones permitieron obtener una primera aproximación a posibles diferencias en el comportamiento de las llamadas dentro de distintos contextos deportivos.

Junto con el análisis descriptivo, se incorporaron contrastes estadísticos mediante pruebas de diferencias de medias (*t-test*). El objetivo de estas pruebas era comprobar si las diferencias observadas entre distintos grupos podían considerarse estadísticamente relevantes o si, por el contrario, podían deberse simplemente a variaciones aleatorias dentro de la muestra analizada.

Por último, se estimó un modelo de regresión lineal mediante Mínimos Cuadrados Ordinarios (OLS). Este modelo permitió analizar conjuntamente varias variables futbolísticas y observar su relación con el número diario de llamadas pertinentes registradas en el Servicio 016. Además, la inclusión de controles territoriales y temporales permitió tener en cuenta parcialmente algunas diferencias existentes entre provincias y entre distintos tipos de días dentro de la muestra.

III.2. ANÁLISIS DESCRIPTIVO

Una vez construida la base final, se realizó un análisis descriptivo inicial con el objetivo de comprender la estructura general de los datos y observar posibles patrones relacionados con las llamadas al Servicio 016. Esta primera fase permitió obtener una visión general del comportamiento de las variables antes de desarrollar el análisis estadístico y econométrico posterior.

El análisis se centró únicamente en las observaciones correspondientes al País Vasco, seleccionando las provincias de Álava, Bizkaia y Gipuzkoa. A partir de esta muestra se calcularon estadísticas descriptivas básicas, como el número total de observaciones, la media diaria de llamadas pertinentes y la distribución de las principales variables futbolísticas.

También se estudiaron las diferencias entre días con y sin actividad futbolística registrada. Para ello, se compararon las medias diarias de llamadas pertinentes entre ambos grupos de observaciones. Del mismo modo, se analizaron las diferencias asociadas al resultado deportivo, diferenciando entre victorias, derrotas, empates y días sin actividad futbolística registrada.

La evolución temporal de las llamadas se analizó mediante agregaciones diarias, mensuales y anuales. Estas representaciones permitieron observar posibles patrones de comportamiento a lo largo del tiempo y detectar variaciones entre distintos periodos del año. Además, se compararon las medias de llamadas entre fines de semana y días laborables, incorporando posteriormente esta variable al análisis econométrico.

El análisis también incluyó algunas situaciones específicas relacionadas con el contexto deportivo, como los derbis vascos entre Athletic Club y Real Sociedad. Para ello, se construyó una variable indicadora que permitía identificar este tipo de enfrentamientos y comparar posteriormente el comportamiento medio de las llamadas en esos días concretos.

Por último, se calcularon matrices de correlación entre las principales variables cuantitativas de la base, incluyendo llamadas pertinentes, goles y balance de goles. Este análisis permitió observar el grado de relación lineal existente entre las distintas variables antes de estimar los modelos de regresión.

III.3. COMPARACIÓN DE MEDIDAS Y CONTRASTES ESTADÍSTICOS

Tras el análisis descriptivo inicial, se realizaron distintas comparaciones de medias con el objetivo de estudiar si existían diferencias apreciables en el número de llamadas pertinentes según el contexto futbolístico y temporal de cada observación.

En primer lugar, se comparó la media diaria de llamadas entre días con actividad futbolística registrada y días sin ella. Esta comparación permitió observar si el comportamiento medio de

las llamadas variaba entre ambos grupos antes de introducir controles adicionales en el análisis econométrico.

Posteriormente, se analizaron las diferencias existentes según el resultado deportivo registrado. Para ello, las observaciones fueron clasificadas en cuatro categorías: *Victoria*, *Derrota*, *Empate* y *Sin fútbol*. A partir de esta clasificación se calcularon las medias correspondientes a cada grupo, permitiendo comparar el comportamiento de las llamadas bajo distintos contextos deportivos.

Además, las comparaciones también se realizaron diferenciando entre provincias y entre fines de semana y días laborables. Este análisis permitía comprobar si determinadas diferencias observadas podían estar relacionadas con factores territoriales o temporales y no únicamente con la actividad futbolística.

Para complementar el análisis descriptivo, se aplicó un contraste estadístico de diferencias de medias mediante un *t-test* para muestras independientes. En concreto, se comparó el número de llamadas pertinentes entre los grupos de observaciones con y sin actividad futbolística registrada. Para ello, la muestra fue dividida en dos subconjuntos y posteriormente se estimó el estadístico correspondiente utilizando la versión del test que no asume igualdad de varianzas entre grupos (*equal_var=False*).

La utilización de este contraste permitía evaluar si las diferencias observadas entre ambos grupos podían considerarse estadísticamente significativas dentro de la muestra analizada. De esta forma, el análisis descriptivo no se limitaba únicamente a comparar medias observadas, sino que incorporaba también una primera aproximación inferencial antes de la estimación econométrica posterior.

III.4. MODELO ECONOMETRICO

Con el objetivo de analizar de forma conjunta la relación entre las variables futbolísticas y el número de llamadas pertinentes registradas en el Servicio 016, se estimó un modelo de regresión lineal mediante el método de Mínimos Cuadrados Ordinarios (OLS). Este tipo de modelo permite estudiar cómo varía la variable dependiente ante cambios en distintas variables explicativas, incorporando además controles territoriales y temporales.

La variable dependiente utilizada en el modelo es el número diario de *Llamadas pertinentes*. Como variables explicativas principales se incorporaron *hay_futbol*, *goles_totales* y *balance_goles*, con el objetivo de captar distintas dimensiones de la actividad futbolística registrada en la base. Además, se añadió la variable *fin_semana*, ya que los patrones de llamadas pueden variar entre días laborables y fines de semana independientemente de la existencia de actividad deportiva.

El modelo también incorpora variables categóricas correspondientes a la provincia y al año mediante efectos fijos simples. La inclusión de estas variables permite controlar parcialmente

diferencias estructurales entre territorios y posibles cambios temporales a lo largo del periodo analizado.

La especificación general estimada puede representarse de la siguiente forma:

$$Llamadas_{it} = \beta_0 + \beta_1 HayFutbol_{it} + \beta_2 GolesTotales_{it} + \beta_3 BalanceGoles_{it} + \beta_4 FinSemana_{it} + \gamma Provincia_i + \delta Año_t + \varepsilon_{it}$$

donde i representa la provincia y t la fecha de la observación.

La estimación se realizó utilizando la librería *statsmodels* de Python mediante la función *ols()*. Posteriormente, se obtuvieron los coeficientes estimados, errores estándar y niveles de significatividad estadística asociados a cada variable.

El objetivo principal de esta regresión no es establecer una relación causal estricta entre fútbol y violencia de género, sino analizar si existe una asociación estadística observable entre determinadas variables futbolísticas y la evolución diaria de las llamadas pertinentes registradas en el Servicio 016.

III.5. LIMITACIONES METODOLÓGICAS

Aunque la base de datos construida permite realizar un análisis detallado de la relación entre actividad futbolística y llamadas al Servicio 016, el trabajo presenta algunas limitaciones que conviene tener en cuenta al interpretar los resultados obtenidos.

Una de las principales limitaciones está relacionada con la propia variable utilizada como aproximación a situaciones de violencia de género. Las llamadas pertinentes al 016 permiten observar variaciones en la demanda de atención y asistencia, pero no representan de forma exacta el número real de casos existentes en cada momento. Además, el comportamiento de las llamadas puede verse influido por otros factores sociales, institucionales o mediáticos que no han sido incorporados en el análisis.

La información futbolística utilizada también presenta ciertas limitaciones. La base construida a partir de la API deportiva recoge actividad relacionada con partidos y goles, pero existen dimensiones del contexto deportivo que resultan difíciles de medir directamente mediante las variables disponibles. Aspectos como la importancia subjetiva de algunos partidos, el nivel de rivalidad percibido por los aficionados o el seguimiento real de cada encuentro no quedan completamente reflejados en la base utilizada.

El análisis se realiza utilizando datos agregados diarios y provinciales. Aunque este nivel de desagregación permite trabajar con una muestra amplia y mantener una estructura temporal homogénea, también dificulta identificar dinámicas individuales o efectos más inmediatos que podrían producirse pocas horas después de determinados eventos deportivos.

Otra limitación importante es que el modelo estimado permite identificar asociaciones estadísticas entre variables, pero no establecer relaciones causales estrictas. Que existan diferencias entre días con y sin actividad futbolística registrada no implica necesariamente que dichas diferencias estén provocadas exclusivamente por el fútbol, ya que pueden existir otros factores no observados que afecten simultáneamente al comportamiento de las llamadas.

También debe tenerse en cuenta que algunas variables utilizadas en el análisis fueron construidas específicamente para este trabajo mediante procesos de agregación y transformación de datos. Aunque se realizaron comprobaciones y tareas de limpieza para asegurar la consistencia de la base final, este tipo de construcción empírica puede incorporar determinadas limitaciones derivadas del propio proceso de recopilación y tratamiento de la información.

CAPÍTULO IV. ANÁLISIS EMPÍRICO Y RESULTADOS

IV.1. DESCRIPCIÓN GENERAL DE LA MUESTRA

La base de datos final utilizada en el análisis contiene un total de 4.644 observaciones diarias correspondientes al País Vasco para el periodo comprendido entre 2021 y 2026. Cada observación representa una combinación diaria entre fecha y provincia, integrando simultáneamente información del Servicio 016 y variables relacionadas con la actividad futbolística.

La muestra incorpora información correspondiente a las tres provincias del País Vasco: Álava, Bizkaia y Gipuzkoa. Del total de observaciones, Bizkaia concentra el mayor número de registros, seguida de Gipuzkoa y Álava. Esta diferencia se debe principalmente al peso poblacional de cada territorio y al distinto volumen de llamadas registradas por provincia.

En conjunto, la base recoge 12.837 llamadas pertinentes registradas durante el periodo analizado, con una media diaria aproximada de 2,76 llamadas por observación. No obstante, la distribución de las llamadas presenta una elevada dispersión, observándose días con valores muy reducidos y otros con niveles considerablemente superiores. El valor máximo registrado en la muestra alcanza las 20 llamadas en una única observación diaria.

	Llamadas pertinentes	goles_favor	goles_contra	goles_totales	hay_futbol	balance_goles
count	4644.0	4644.0	4644.0	4644.0	4644.0	4644.0
mean	2.764	0.565	0.526	1.092	0.276	0.038
std	2.163	1.290	1.215	2.263	0.447	1.078
min	0.0	0.0	0.0	0.0	0.0	-6.0
25 %	1.0	0.0	0.0	0.0	0.0	0.0
50 %	2.0	0.0	0.0	0.0	0.0	0.0
75 %	4.0	0.0	0.0	1.0	1.0	0.0
max	20.0	10.0	9.0	17.0	1.0	10.0

TABLA 1: Estadísticas descriptivas de las principales variables

Respecto a la actividad futbolística, la mayoría de observaciones corresponden a días sin fútbol registrado en la base. En concreto, 3.361 observaciones pertenecen a días sin actividad futbolística, frente a 1.283 observaciones en las que sí se registró al menos un evento deportivo asociado a los equipos analizados.

Además, las observaciones con actividad futbolística fueron clasificadas según el resultado deportivo registrado. Dentro de los días con fútbol, las victorias representan la categoría más frecuente, seguidas por las derrotas y los empates. Paralelamente, también se mantuvieron las observaciones correspondientes a días sin actividad futbolística, utilizadas posteriormente como grupo de comparación en el análisis descriptivo y econométrico.

En relación con las variables deportivas, la base incorpora información diaria agregada sobre goles a favor, goles en contra y goles totales. Estas variables fueron construidas a partir de los eventos registrados en los partidos descargados mediante la API deportiva y posteriormente agregadas por fecha y comunidad autónoma antes de realizar el cruce con la base del Servicio 016.

Finalmente, antes de desarrollar el análisis empírico, se realizaron comprobaciones generales sobre la consistencia de la base de datos, incluyendo revisión de tipos de variables, valores nulos y estructura temporal de las observaciones. Estas verificaciones permitieron confirmar la correcta integración de ambas fuentes de información y garantizar la coherencia interna del dataset final utilizado en el estudio.

IV.2. EVOLUCIÓN TEMPORAL DE LAS LLAMADAS

Con el objetivo de observar el comportamiento general de las llamadas pertinentes a lo largo del tiempo, se realizó un análisis temporal utilizando la información diaria agregada para el País Vasco. Este análisis permite identificar posibles patrones de evolución, variaciones estacionales y diferencias entre distintos periodos del intervalo analizado.

En primer lugar, se construyó una serie temporal diaria agregando el número total de llamadas pertinentes registradas en cada fecha. La evolución observada muestra un comportamiento relativamente estable durante gran parte del periodo analizado, aunque con fluctuaciones diarias importantes y determinados picos puntuales de llamadas. Estas variaciones reflejan la existencia de una elevada dispersión diaria en el comportamiento de la variable.

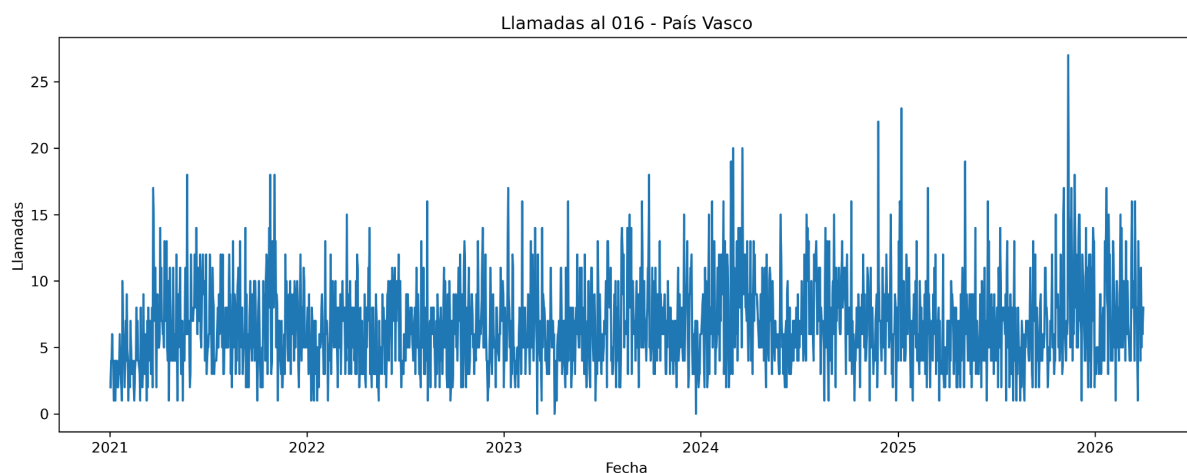


GRÁFICO 1: Evolución temporal de las llamadas al 016 en el País Vasco

Luego, se analizó la evolución mensual de las llamadas mediante el cálculo de medias agrupadas por mes. Este análisis permite observar posibles patrones estacionales en el comportamiento de las llamadas registradas. Aunque las diferencias entre meses no son

extremadamente pronunciadas, sí se aprecian ligeras variaciones a lo largo del año, lo que sugiere la existencia de ciertos componentes temporales en la serie.

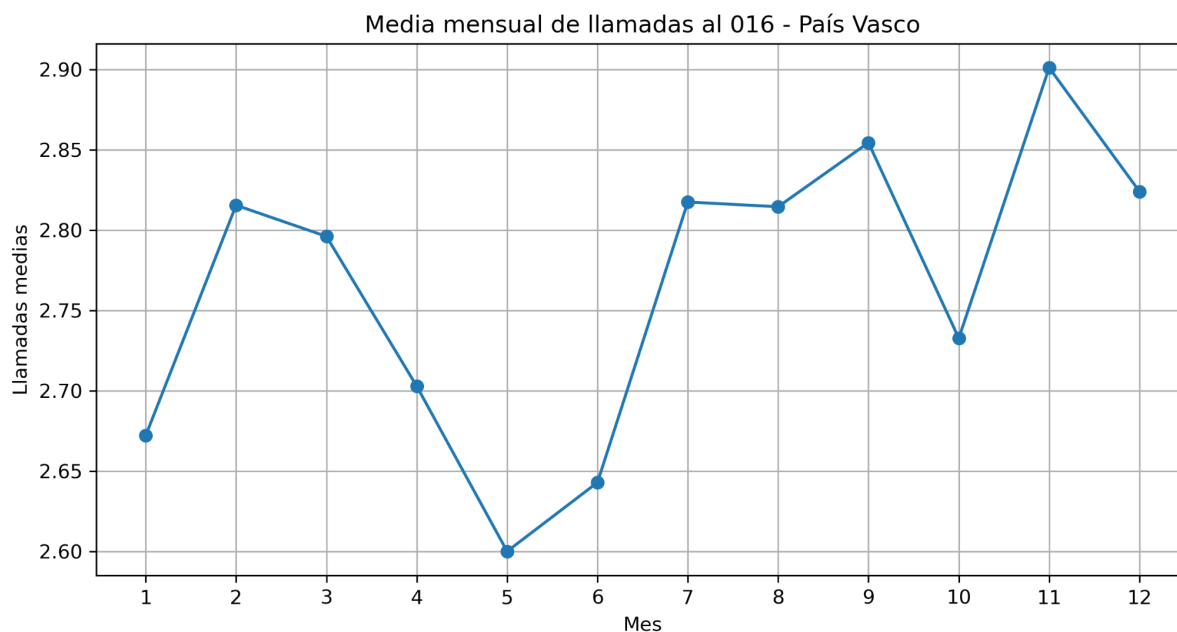


GRÁFICO 2: Evolución mensual media de las llamadas al 016

Después, se miró la evolución anual de las llamadas pertinentes durante el periodo comprendido entre 2021 y 2026. Los resultados muestran que las medias anuales presentan pequeñas variaciones entre años, aunque sin cambios estructurales especialmente bruscos. No obstante, sí se observa un ligero incremento en algunos de los años más recientes de la muestra.

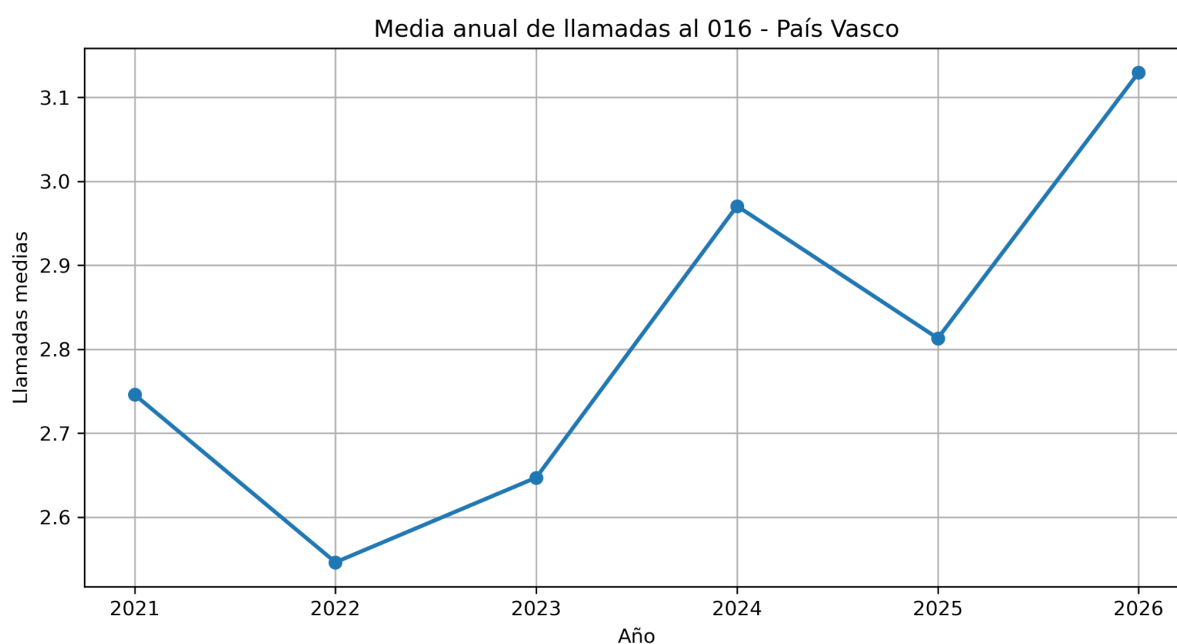


GRÁFICO 3: Evolución anual media de las llamadas al 016

En conjunto, el análisis temporal permite comprobar que las llamadas al Servicio 016 presentan cierta estabilidad agregada a lo largo del tiempo, aunque coexistiendo con fluctuaciones diarias relevantes. Estas diferencias justifican posteriormente la incorporación de controles temporales dentro de los modelos econométricos utilizados en el trabajo.

IV.3. ACTIVIDAD FUTBOLÍSTICA Y LLAMADAS AL 016

Una vez analizada la evolución temporal general de las llamadas pertinentes, el siguiente paso consistió en estudiar si existían diferencias en el comportamiento de las llamadas entre días con y sin actividad futbolística registrada en la base.

Para ello, en primer lugar se compararon las medias diarias de llamadas entre ambos grupos de observaciones. Los resultados muestran diferencias reducidas entre días con actividad futbolística y días sin ella, aunque sí se observan ligeras variaciones en el número medio de llamadas registradas.

hay_futbol	Llamadas pertinentes
0	2.819
1	2.620

TABLA 2: Comparación de llamadas entre días con y sin actividad futbolística

Después, el análisis se amplió incorporando el resultado deportivo registrado en cada fecha. Para ello, las observaciones fueron clasificadas en cuatro categorías: *Victoria*, *Empate*, *Derrota* y *Sin fútbol*. Esta clasificación permite comparar si el comportamiento medio de las llamadas varía dependiendo del contexto deportivo asociado a cada día.

Los resultados descriptivos muestran diferencias moderadas entre categorías, aunque sin cambios extremadamente pronunciados entre unas situaciones y otras. En términos generales, las observaciones correspondientes a derrotas y empates presentan medias ligeramente superiores respecto a otros contextos deportivos, mientras que los días sin actividad futbolística muestran un comportamiento relativamente estable dentro de la muestra.

resultado_partido	Llamadas pertinentes
Derrota	2.564
Empate	2.573
Sin fútbol	2.819
Victoria	2.690

TABLA 3: Comparación de llamadas por resultado del partido

Con el objetivo de representar visualmente estas diferencias, se elaboró un gráfico de barras utilizando la media de llamadas correspondiente a cada tipo de resultado deportivo. Esta representación permite observar de forma más clara las diferencias existentes entre categorías y facilita la comparación entre contextos futbolísticos distintos.

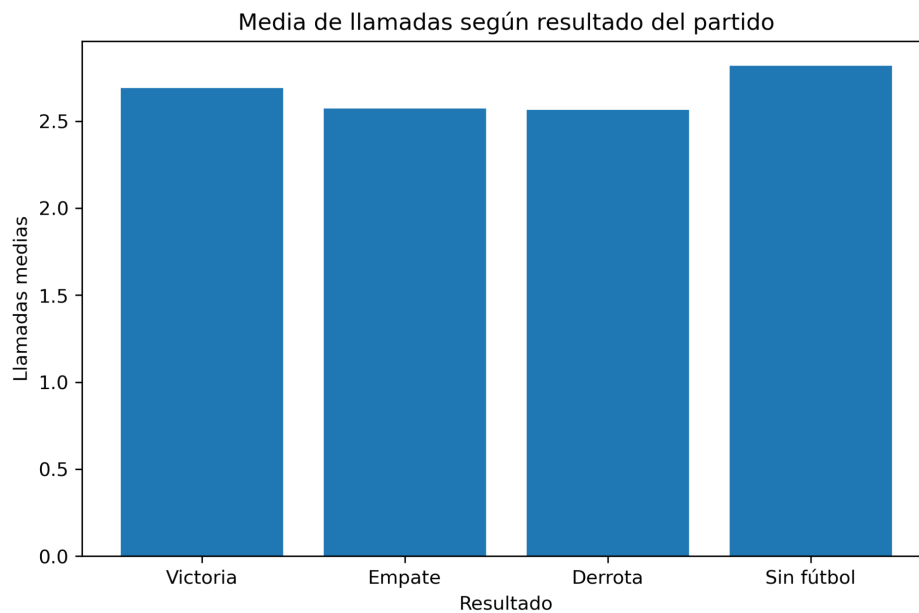


GRÁFICO 4: Media de llamadas al 016 según resultado del partido

Además de las medias, también se analizó la distribución de las llamadas mediante diagramas de caja (*boxplots*). Este tipo de representación permite observar no solo las diferencias centrales entre grupos, sino también la dispersión, los rangos intercuartílicos y la presencia de posibles valores extremos dentro de cada categoría.

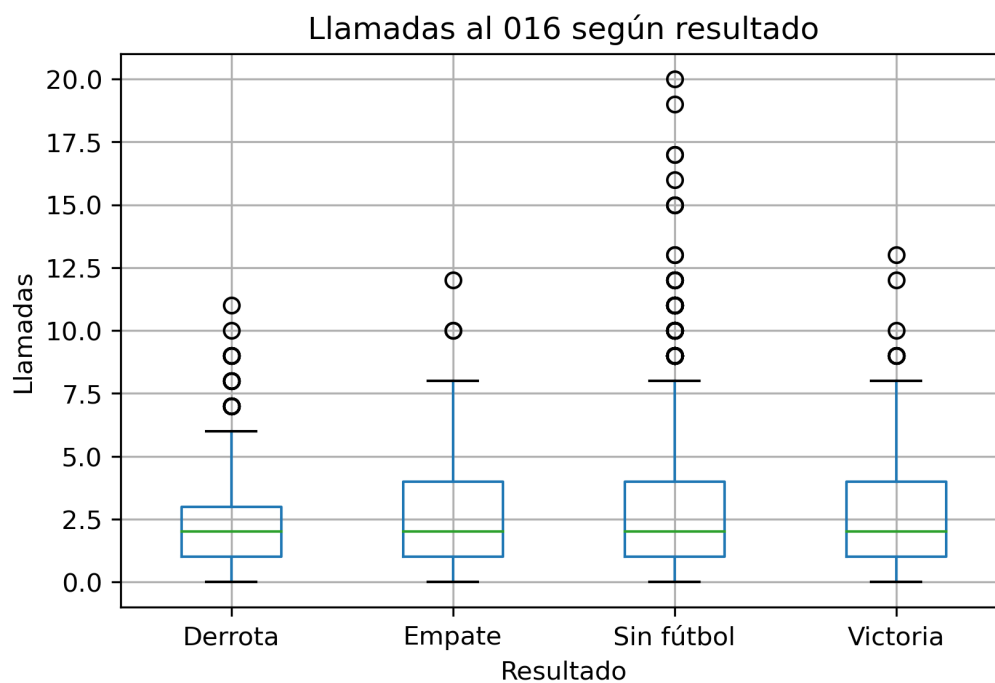


GRÁFICO 5: Distribución de llamadas al 016 según resultado del partido

En conjunto, el análisis descriptivo sugiere que pueden existir ciertas diferencias en el comportamiento de las llamadas dependiendo del contexto futbolístico registrado en cada fecha. No obstante, estas diferencias deben interpretarse con cautela y serán analizadas posteriormente con mayor detalle mediante contrastes estadísticos y modelos econométricos.

IV.4. FINES DE SEMANA Y DERBIS VASCOS

Además de las variables directamente relacionadas con los resultados deportivos, el análisis también incorporó algunas situaciones específicas que podían estar asociadas al comportamiento de las llamadas pertinentes registradas en el Servicio 016.

Una de ellas fue la posible existencia de diferencias entre días laborables y fines de semana. Para estudiar esta cuestión, se construyó la variable *fin_semana*, utilizada para identificar las observaciones correspondientes a sábados y domingos. A partir de esta clasificación, se compararon las medias diarias de llamadas entre ambos grupos de observaciones.

Los resultados muestran que las llamadas pertinentes presentan un comportamiento ligeramente distinto según el tipo de día analizado. En concreto, las medias registradas durante los fines de semana son algo inferiores a las observadas en días laborables. Estas diferencias podrían estar relacionadas con cambios en los patrones de actividad diaria, la movilidad o la utilización de servicios de atención en distintos momentos de la semana.

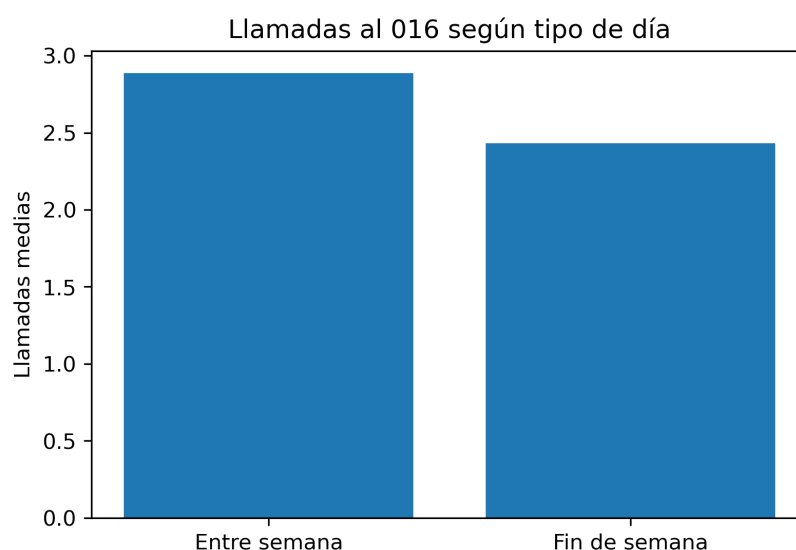


GRÁFICO 6: Comparación de llamadas al 016 entre semana y fin de semana

El análisis también incluyó los denominados derbis vascos, es decir, los enfrentamientos entre Athletic Club y Real Sociedad. Para ello, se creó una variable específica (*derbi_vasco*)

que identifica aquellas fechas en las que ambos equipos aparecen simultáneamente registrados dentro de la actividad futbolística diaria.

A partir de esta variable, se comparó la media de llamadas pertinentes entre días con derbi y días sin este tipo de enfrentamientos. Los resultados muestran ciertas diferencias entre ambos grupos, aunque el número de observaciones correspondientes a derbis es considerablemente menor en comparación con el resto de la muestra. Por este motivo, los resultados asociados a esta variable deben interpretarse con cierta prudencia.

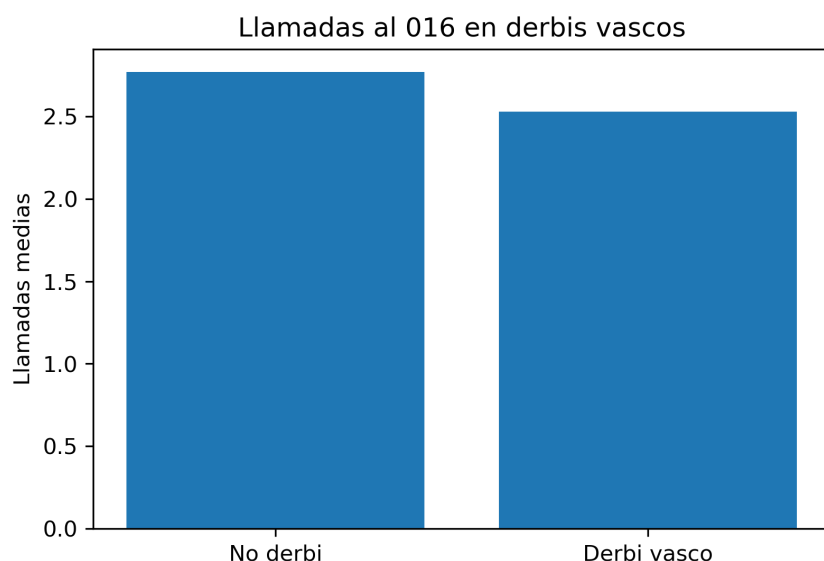


GRÁFICO 7: Llamadas al 016 en días de derbi vasco

Estos análisis complementarios permitieron incorporar algunas dimensiones temporales y contextuales adicionales al estudio. Tanto los fines de semana como determinados partidos considerados especialmente relevantes pueden estar asociados a dinámicas diferentes dentro de la evolución diaria de las llamadas. En cualquier caso, el posible impacto de estas variables debe valorarse conjuntamente con el resto de factores incluidos posteriormente en el análisis econométrico.

IV.5. CORRELACIONES Y ANÁLISIS PROVINCIAL

Con el objetivo de complementar el análisis descriptivo anterior, se estudiaron las relaciones existentes entre las principales variables cuantitativas de la base de datos. Para ello, se calcularon coeficientes de correlación lineal entre las llamadas pertinentes y distintas variables relacionadas con la actividad futbolística.

En términos generales, las correlaciones observadas entre el número de llamadas y las variables futbolísticas son relativamente reducidas. Esto sugiere que, al menos de forma lineal y agregada, la relación directa entre ambas dimensiones no resulta especialmente intensa dentro de la muestra analizada. Sin embargo, sí se observan correlaciones más

elevadas entre algunas de las propias variables deportivas, especialmente entre goles a favor, goles en contra y goles totales, debido a la relación mecánica existente entre ellas.

	Llamadas pertinentes	hay_futbol	goles_favor	goles_contra	goles_totales	balance_goles
Llamadas pertinentes	1.0	-0.041	-0.029	-0.039	-0.037	0.010
hay_futbol	-0.041	1.0	0.709	0.702	0.781	0.058
goles_favor	-0.029	0.709	1.0	0.631	0.909	0.486
goles_contra	-0.039	0.702	0.631	1.0	0.897	-0.371
goles_totales	-0.037	0.781	0.909	0.897	1.0	0.078
balance_goles	0.010	0.058	0.486	-0.371	0.078	1.0

TABLA 4: Matriz de correlación entre las principales variables

Para facilitar la interpretación visual de estas relaciones, también se elaboró una matriz gráfica de correlación. Esta representación permite identificar de manera más intuitiva la intensidad y dirección de las relaciones entre variables, diferenciando correlaciones positivas y negativas.

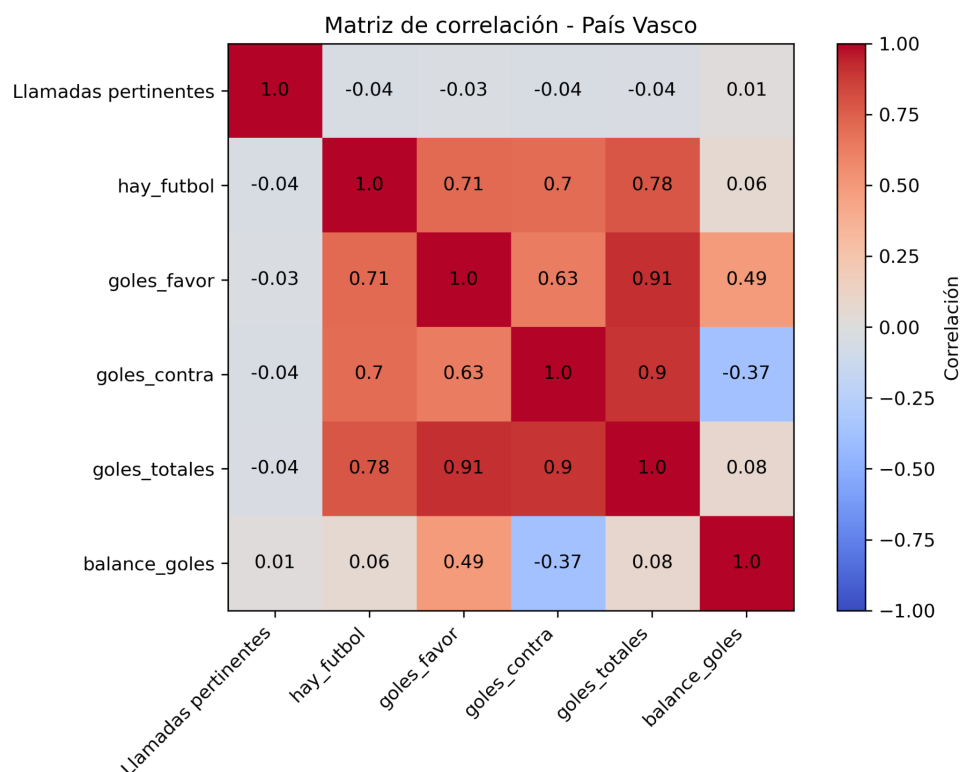


GRÁFICO 8: Matriz de correlación entre variables del estudio

Además del análisis de correlaciones, se realizó una comparación provincial de las llamadas pertinentes dentro del País Vasco. Para ello, se calcularon las medias correspondientes a Álava, Bizkaia y Gipuzkoa, observándose diferencias relevantes entre territorios. Bizkaia presenta el mayor volumen medio de llamadas de la muestra, seguida de Gipuzkoa y Álava. Estas diferencias pueden estar relacionadas principalmente con el distinto tamaño poblacional de cada provincia.

Provincia	Llamadas pertinentes
Araba/Álava	1.665
Bizkaia	4.064
Gipuzkoa	2.079

TABLA 5: Distribución provincial de las llamadas pertinentes

Posteriormente, también se compararon las medias provinciales según el resultado deportivo registrado. Este análisis permite observar si las posibles diferencias asociadas al contexto futbolístico presentan comportamientos similares o distintos entre provincias. Aunque existen ciertas variaciones territoriales, los patrones observados mantienen una estructura relativamente homogénea dentro de la muestra.

Provincia	resultado_partido	Llamadas pertinentes
Araba/Álava	Derrota	1.572
Araba/Álava	Empate	1.586
Araba/Álava	Sin fútbol	1.669
Araba/Álava	Victoria	1.769
Bizkaia	Derrota	3.654
Bizkaia	Empate	3.701
Bizkaia	Sin fútbol	4.197
Bizkaia	Victoria	3.813
Gipuzkoa	Derrota	2.019
Gipuzkoa	Empate	1.896
Gipuzkoa	Sin fútbol	2.120
Gipuzkoa	Victoria	1.958

TABLA 6: Comparación provincial de llamadas según resultado del partido

En conjunto, el análisis provincial y las matrices de correlación permiten contextualizar mejor los resultados descriptivos obtenidos anteriormente y ofrecen una primera

aproximación a las posibles relaciones existentes entre las variables antes de realizar la estimación econométrica final.

IV.6. RESULTADOS DE LA REGRESIÓN

Con el objetivo de analizar conjuntamente la relación entre las variables futbolísticas y el número de llamadas pertinentes registradas en el Servicio 016, se estimó un modelo de regresión lineal mediante Mínimos Cuadrados Ordinarios (OLS). El modelo incorpora variables relacionadas con la actividad futbolística, controles temporales y efectos territoriales para las provincias del País Vasco.

La estimación obtenida presenta un coeficiente de determinación (R^2) de aproximadamente 0,263, lo que indica que el modelo consigue explicar alrededor del 26 % de la variabilidad observada en las llamadas pertinentes dentro de la muestra analizada. Aunque este nivel explicativo no es especialmente elevado, resulta razonable teniendo en cuenta la complejidad del fenómeno estudiado y la existencia de múltiples factores no observables que pueden influir sobre el comportamiento diario de las llamadas.

En relación con las variables futbolísticas principales, los resultados muestran que ni *hay_futbol*, ni *goles_totales*, ni *balance_goles* presentan coeficientes estadísticamente significativos dentro del modelo estimado. Esto sugiere que, una vez incorporados los controles temporales y territoriales, no puede identificarse una relación estadísticamente robusta entre estas variables y el volumen diario de llamadas pertinentes registradas.

Por el contrario, la variable *fin_semana* sí presenta un coeficiente negativo y estadísticamente significativo. Los resultados indican que, manteniendo constantes el resto de variables del modelo, los fines de semana se asocian con un menor número medio de llamadas respecto a los días laborables. Este resultado coincide con los patrones observados previamente en el análisis descriptivo.

Asimismo, los efectos territoriales muestran diferencias relevantes entre provincias. En particular, Bizkaia presenta un volumen medio de llamadas significativamente superior respecto a la categoría de referencia utilizada en la estimación. Gipuzkoa también muestra diferencias positivas, aunque de menor magnitud. Estos resultados reflejan la existencia de diferencias estructurales entre territorios dentro del País Vasco.

En cuanto a los controles temporales, algunos años presentan coeficientes estadísticamente significativos respecto al año base utilizado en el modelo. No obstante, las diferencias observadas entre periodos son relativamente moderadas y no muestran una tendencia completamente uniforme a lo largo del intervalo analizado.

OLS Regression Results						
Dep. Variable:	Q("Llamadas pertinentes")		R-squared:	0.263		
Model:	OLS		Adj. R-squared:	0.261		
Method:	Least Squares		F-statistic:	150.2		
Date:	Tue, 26 May 2026		Prob (F-statistic):	1.05e-296		
Time:	18:23:00		Log-Likelihood:	-9463.9		
No. Observations:	4644		AIC:	1.895e+04		
Df Residuals:	4632		BIC:	1.903e+04		
Df Model:	11					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.7652	0.082	21.430	0.000	1.604	1.927
C(Provincia) [T.Bizkaia]	2.4132	0.069	35.200	0.000	2.279	2.548
C(Provincia) [T.Gipuzkoa]	0.4233	0.071	5.959	0.000	0.284	0.563
C(Año) [T.2022]	-0.1864	0.090	-2.069	0.039	-0.363	-0.010
C(Año) [T.2023]	-0.0742	0.090	-0.828	0.407	-0.250	0.101
C(Año) [T.2024]	0.2570	0.089	2.886	0.004	0.082	0.432
C(Año) [T.2025]	0.0725	0.091	0.796	0.426	-0.106	0.251
C(Año) [T.2026]	0.3743	0.144	2.596	0.009	0.092	0.657
hay_futbol	-0.0603	0.100	-0.604	0.546	-0.256	0.136
goles_totales	0.0104	0.020	0.528	0.597	-0.028	0.049
balance_goles	0.0054	0.026	0.210	0.834	-0.045	0.056
fin_semana	-0.5090	0.069	-7.410	0.000	-0.644	-0.374
Omnibus:	1413.214	Durbin-Watson:	1.822			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6537.881			
Skew:	1.403	Prob(JB):	0.00			
Kurtosis:	8.091	Cond. No.	17.5			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

TABLA: Resultados de la regresión OLS

Los resultados obtenidos sugieren que las variables futbolísticas incorporadas en el análisis no presentan una relación estadísticamente significativa con las llamadas pertinentes al Servicio 016 dentro de la muestra utilizada. Sin embargo, sí se observan diferencias temporales y territoriales relevantes, especialmente asociadas a los fines de semana y a determinadas provincias del País Vasco.

CAPÍTULO V. DISCUSIÓN

V.1. INTERPRETACIÓN GENERAL DE LOS RESULTADOS

El objetivo principal de este trabajo consistía en analizar si existía algún tipo de relación entre la actividad futbolística y el comportamiento de las llamadas pertinentes registradas en el Servicio 016 en el País Vasco. Para ello, se construyó una base de datos diaria integrando información procedente de dos fuentes distintas: por un lado, los datos oficiales publicados por el Ministerio de Igualdad y, por otro, una base futbolística elaborada específicamente para el estudio a partir de información obtenida mediante la API deportiva *API-Football*.

La construcción de esta base permitió relacionar diariamente la evolución de las llamadas pertinentes con distintos elementos vinculados a la actividad futbolística registrada en cada territorio. En concreto, se incorporaron variables relacionadas con la existencia o no de partidos, el número de goles anotados, el balance entre goles a favor y goles en contra y los distintos resultados deportivos registrados en cada fecha. Esto permitió trabajar con una estructura diaria homogénea y desarrollar posteriormente tanto el análisis descriptivo como la estimación econométrica del estudio.

A nivel descriptivo, los resultados muestran ciertas diferencias entre días con y sin fútbol, aunque estas diferencias no resultan especialmente elevadas dentro de la muestra analizada. También aparecen pequeñas variaciones dependiendo del resultado deportivo registrado en cada jornada. En términos generales, las derrotas y los empates presentan medias ligeramente superiores respecto a las victorias, mientras que los días sin actividad futbolística mantienen un comportamiento relativamente estable durante la mayor parte del periodo estudiado.

El análisis temporal permitió observar que las llamadas pertinentes mantienen una evolución relativamente estable a lo largo del intervalo analizado, aunque existen oscilaciones diarias importantes y algunos picos puntuales en determinadas fechas. Estas fluctuaciones muestran que el comportamiento de las llamadas no sigue una evolución completamente uniforme y que existen variaciones temporales relevantes dentro de la serie diaria. También se observaron diferencias entre provincias, siendo Bizkaia el territorio que registra un mayor volumen medio de llamadas durante prácticamente todo el periodo analizado. Esto probablemente esté relacionado con diferencias de población y tamaño entre los distintos territorios incluidos en la muestra.

Los resultados obtenidos mediante la regresión lineal muestran que las principales variables futbolísticas incorporadas en el modelo no presentan coeficientes estadísticamente significativos una vez introducidos los controles territoriales y temporales. Por tanto, dentro de la muestra utilizada no puede identificarse una relación sólida entre la actividad futbolística y el número diario de llamadas pertinentes registradas en el Servicio 016.

Aun así, sí aparecen diferencias relevantes asociadas a otras variables incluidas en el análisis. En particular, los fines de semana presentan un comportamiento distinto respecto a los días laborables y determinadas provincias muestran niveles medios de llamadas significativamente superiores respecto al resto. Estos resultados sugieren que factores

territoriales y temporales parecen tener una mayor capacidad explicativa dentro del modelo que las propias variables futbolísticas incorporadas en el estudio.

V.2. DISCUSIÓN DE LOS RESULTADOS

Los resultados obtenidos a lo largo del análisis no permiten confirmar de forma clara la existencia de una relación estadísticamente significativa entre la actividad futbolística y las llamadas pertinentes registradas en el Servicio 016. Aunque en algunos análisis descriptivos sí aparecen pequeñas diferencias entre determinados contextos deportivos, estas variaciones pierden fuerza cuando se incorporan controles territoriales y temporales dentro del modelo econométrico.

Las variables relacionadas con los partidos de fútbol —como la existencia de actividad futbolística, el número de goles o el balance entre goles a favor y en contra— muestran una capacidad explicativa limitada dentro de la muestra analizada. Esto sugiere que la evolución diaria de las llamadas parece estar condicionada por factores mucho más amplios y complejos que los acontecimientos deportivos incluidos en el estudio.

La ausencia de resultados estadísticamente significativos tampoco implica necesariamente que el contexto futbolístico no pueda influir sobre determinadas dinámicas sociales. La violencia de género constituye un fenómeno multidimensional en el que intervienen factores estructurales, económicos, culturales, psicológicos y sociales difíciles de representar mediante variables agregadas construidas a partir de resultados deportivos.

También resulta posible que algunos efectos asociados a determinados partidos no se reflejen de forma inmediata en los datos utilizados. Parte de esos posibles impactos podrían aparecer con cierto retraso temporal, quedar diluidos dentro de observaciones diarias agregadas o incluso manifestarse a través de otros indicadores distintos a los analizados en este trabajo.

Los resultados sí muestran diferencias relevantes asociadas a determinadas variables territoriales y temporales. Los fines de semana presentan comportamientos distintos respecto a los días laborables y algunas provincias registran niveles medios de llamadas superiores al resto de territorios analizados. Estos patrones parecen indicar que existen componentes estructurales y sociales con una mayor capacidad explicativa dentro del modelo estimado.

El trabajo también pone de manifiesto la dificultad de estudiar empíricamente fenómenos sociales complejos utilizando información agregada. Aunque existen debates sociales y algunos estudios previos que sugieren posibles relaciones entre acontecimientos deportivos y determinados comportamientos sociales, identificar esas dinámicas de manera estadísticamente robusta resulta considerablemente más complicado en la práctica.

Los resultados obtenidos deben interpretarse con prudencia. Más allá de confirmar o rechazar completamente la hipótesis inicial, el análisis realizado permite observar la complejidad del

fenómeno estudiado y las limitaciones existentes a la hora de medir este tipo de relaciones mediante datos agregados diarios.

Uno de los principales aportes del trabajo reside también en el propio proceso de construcción de la base de datos y en la metodología utilizada para integrar información institucional y deportiva dentro de una misma estructura diaria de análisis. La base construida puede servir como punto de partida para futuras investigaciones relacionadas con acontecimientos deportivos y distintos indicadores sociales.

V.3. LIMITACIONES DEL TRABAJO

El presente trabajo presenta diversas limitaciones que conviene tener en cuenta al interpretar los resultados obtenidos. Una de las principales está relacionada con la propia naturaleza de la variable dependiente utilizada. Las llamadas pertinentes registradas por el Servicio 016 constituyen una aproximación indirecta a situaciones relacionadas con violencia de género, pero no representan de forma exacta el número real de casos existentes en cada momento. Por tanto, las variaciones observadas en las llamadas no deben interpretarse automáticamente como cambios directos en la incidencia real de este tipo de situaciones.

El comportamiento de las llamadas también puede verse afectado por numerosos factores externos que no han sido incorporados dentro del análisis. Entre ellos pueden encontrarse campañas institucionales, cobertura mediática, cambios legislativos, acontecimientos sociales concretos o modificaciones relacionadas con el acceso a servicios de atención y asistencia. Todos estos elementos pueden influir en la utilización del Servicio 016 y resultar difíciles de controlar mediante un modelo cuantitativo como el utilizado en este trabajo.

La base futbolística utilizada fue construida específicamente para este estudio a partir de datos obtenidos mediante una API deportiva. Aunque este proceso permitió elaborar una base amplia, homogénea y adaptada a los objetivos de la investigación, también incorpora algunas limitaciones relacionadas con la disponibilidad y estructura de la información proporcionada por la propia plataforma. Determinados datos o características de los partidos no siempre aparecen recogidos con el mismo nivel de detalle.

Además, muchas de las variables futbolísticas utilizadas constituyen aproximaciones relativamente simples al contexto deportivo real. Factores como la importancia emocional de determinados partidos, el nivel de rivalidad percibido por los aficionados, el seguimiento real de los encuentros o el impacto mediático concreto de cada resultado son dimensiones difíciles de medir cuantitativamente. Este tipo de elementos no quedan completamente reflejados mediante variables agregadas de goles, resultados o existencia de actividad futbolística.

Otra limitación importante está relacionada con el nivel de agregación temporal utilizado en el análisis. El trabajo emplea datos diarios agregados por provincia, lo que permite mantener una estructura homogénea y trabajar con una muestra suficientemente amplia. Sin embargo, este nivel de agregación dificulta analizar dinámicas más inmediatas o posibles efectos de

muy corto plazo que podrían producirse pocas horas después de determinados acontecimientos deportivos. Es posible que algunos efectos concretos se diluyan al trabajar únicamente con información diaria agregada.

Por último, el modelo econométrico utilizado permite identificar asociaciones estadísticas entre variables, pero no establecer relaciones causales estrictas. Por este motivo, los resultados obtenidos deben interpretarse principalmente como una aproximación exploratoria dentro de un fenómeno social considerablemente más amplio y complejo. En consecuencia, las relaciones observadas en el análisis no implican necesariamente la existencia de un efecto causal directo entre actividad futbolística y llamadas registradas en el Servicio 016.

CAPÍTULO VI. CONCLUSIONES

VI.1. CONCLUSIONES PRINCIPALES

El presente trabajo tenía como objetivo analizar la posible relación entre la actividad futbolística y las llamadas pertinentes registradas en el Servicio 016 en el País Vasco durante el periodo 2021-2026. Para ello, se construyó una base de datos diaria integrando información institucional y deportiva procedente de distintas fuentes. La combinación de ambas bases permitió relacionar la evolución diaria de las llamadas con distintas variables asociadas a la actividad futbolística registrada en cada fecha y territorio.

A partir del análisis realizado, los resultados muestran que existen algunas diferencias descriptivas entre determinados contextos futbolísticos, aunque estas variaciones no resultan especialmente elevadas dentro de la muestra analizada. Las variables relacionadas con la existencia de partidos, el número de goles o el balance de resultados no presentan coeficientes estadísticamente significativos dentro del modelo econométrico una vez incorporados los controles territoriales y temporales. Esto implica que, utilizando la metodología y los datos empleados en este estudio, no puede identificarse una relación sólida entre la actividad futbolística y el volumen diario de llamadas pertinentes registradas en el Servicio 016.

El análisis también muestra que otras variables incluidas en el modelo parecen tener una mayor capacidad explicativa. En particular, aparecen diferencias asociadas a determinadas provincias y entre días laborables y fines de semana. Estas diferencias sugieren que el comportamiento de las llamadas puede estar más relacionado con factores territoriales, sociales o temporales que con las propias variables futbolísticas incorporadas en el análisis.

Asimismo, los resultados obtenidos muestran que la incorporación de variables temporales, especialmente el efecto fin de semana, modifica parcialmente la relación observada en los análisis descriptivos iniciales. Este resultado refuerza la importancia de considerar factores temporales en la interpretación de la relación entre actividad futbolística y llamadas al Servicio 016.

En cualquier caso, los resultados obtenidos permiten extraer algunas conclusiones adicionales respecto al objetivo planteado inicialmente. El hecho de no encontrar una relación estadísticamente significativa entre las variables futbolísticas y las llamadas pertinentes al Servicio 016 permite descartar que la actividad futbolística registrada constituya, al menos en los términos analizados en este trabajo, un factor especialmente relevante para explicar las variaciones observadas en las llamadas. Además, los resultados ponen de manifiesto la importancia de considerar otros elementos presentes en los datos, especialmente aquellos relacionados con el territorio o con la dimensión temporal. En este sentido, el análisis realizado ayuda a delimitar mejor qué factores parecen estar asociados a las diferencias observadas y cuáles no muestran una influencia apreciable dentro del periodo estudiado.

Aunque determinados acontecimientos deportivos pueden generar contextos emocionales intensos y situaciones de elevada tensión entre aficionados y seguidores, los resultados

obtenidos no permiten observar un efecto estadísticamente claro sobre la evolución diaria de las llamadas registradas. Esto no significa necesariamente que este tipo de acontecimientos no puedan influir en determinados comportamientos sociales, sino que dicho efecto resulta difícil de identificar mediante información agregada diaria y variables cuantitativas relativamente limitadas.

El trabajo también pone de manifiesto la complejidad de analizar fenómenos sociales mediante modelos cuantitativos agregados. La violencia de género constituye un fenómeno condicionado por múltiples factores sociales, económicos, institucionales y personales que no siempre pueden medirse de forma directa mediante variables observables. Por este motivo, las relaciones identificadas en el estudio deben interpretarse como asociaciones estadísticas dentro de un contexto social mucho más amplio y complejo.

En conjunto, los resultados obtenidos no permiten confirmar las hipótesis que planteaban una relación directa entre la actividad futbolística y el volumen diario de llamadas pertinentes registradas en el Servicio 016. Por el contrario, el análisis sugiere que las diferencias observadas en las llamadas parecen estar más asociadas a factores territoriales y temporales que a las variables deportivas incorporadas en el estudio.

VI.2. APORTACIONES DEL TRABAJO

Una de las principales aportaciones del trabajo reside en el propio proceso de construcción de la base de datos utilizada para el análisis. La investigación integra información diaria procedente del Servicio 016 con una base futbolística elaborada específicamente para el estudio mediante consultas automatizadas a una API deportiva.

La construcción y limpieza de ambas bases permitió generar una estructura homogénea de datos diarios a nivel territorial, incorporando variables relacionadas con partidos, goles, resultados deportivos y otros elementos temporales relevantes para el análisis empírico posterior.

Además, el trabajo aporta una aproximación aplicada al estudio de posibles relaciones entre acontecimientos deportivos y determinados indicadores sociales utilizando datos diarios desagregados para el País Vasco. Este enfoque permite trabajar con un mayor nivel de detalle temporal respecto a otros estudios basados únicamente en datos agregados anuales o nacionales.

Por otro lado, el desarrollo de scripts automatizados para la recopilación, limpieza y agregación de la información deportiva constituye también una aportación metodológica relevante dentro del proceso de elaboración del trabajo.

ANEXO

A. TABLAS

TABLA 1: Estadísticas descriptivas de las principales variables
(descriptivos_generales)

	Año	Mes_num	Día	Llamadas pertinentes	goles_favor	goles_contra	goles_totales	hay_futbol	balance_goles
count	4644.0	4644.0	4644.0	4644.0	4644.0	4644.0	4644.0	4644.0	4644.0
mean	2023.171	6.365	1.574	2.764	0.565	0.527	1.092	0.276	0.038
std	1.505	3.492	8.783	2.163	1.290	1.215	2.262	0.447	1.078
min	2021.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	-6.0
25 %	2022.0	3.0	8.0	1.0	0.0	0.0	0.0	0.0	0.0
50 %	2023.0	6.0	16.0	2.0	0.0	0.0	0.0	0.0	0.0
75 %	2024.0	9.0	23.0	4.0	0.0	0.0	1.0	1.0	0.0
max	2026.0	12.0	31.0	20.0	10.0	9.0	17.0	1.0	10.0

TABLA 2: Comparación de llamadas entre días con y sin actividad futbolística
(*futbol_no_futbol*)

hay_futbol	Llamadas pertinentes
0	2.819
1	2.620

TABLA 3: Comparación de llamadas por resultado del partido
(*resultados_partido*)

resultado_partido	Llamadas pertinentes
Derrota	2.564
Empate	2.573
Sin fútbol	2.819
Victoria	2.690

TABLA 4: Matriz de correlación entre las principales variables
(*matriz_correlacion*)

	Llamadas pertinentes	hay_futbol	goles_favor	goles_contra	goles_totales	balance_goles
Llamadas pertinentes	1.0	-0.041	-0.029	-0.039	-0.037	0.010
hay_futbol	-0.041	1.0	0.709	0.702	0.781	0.058
goles_favor	-0.029	0.709	1.0	0.631	0.909	0.486
goles_contra	-0.039	0.702	0.631	1.0	0.897	-0.371
goles_totales	-0.037	0.781	0.909	0.897	1.0	0.078
balance_goles	0.010	0.058	0.486	-0.371	0.078	1.0

TABLA 5: Distribución provincial de las llamadas pertinentes
(*provincias*)

Provincia	Llamadas pertinentes
Araba/Álava	1.665
Bizkaia	4.064
Gipuzkoa	2.079

TABLA 6: Comparación provincial de llamadas según resultado del partido
(provincias_resultado)

Provincia	resultado_partido	Llamadas pertinentes
Araba/Álava	Derrota	1.572
Araba/Álava	Empate	1.586
Araba/Álava	Sin fútbol	1.669
Araba/Álava	Victoria	1.769
Bizkaia	Derrota	3.654
Bizkaia	Empate	3.701
Bizkaia	Sin fútbol	4.197
Bizkaia	Victoria	3.813
Gipuzkoa	Derrota	2.019
Gipuzkoa	Empate	1.896
Gipuzkoa	Sin fútbol	2.120
Gipuzkoa	Victoria	1.958

B. OUTPUTS DE LA REGRESIÓN OLS

OLS Regression Results						
Dep. Variable:	Q("Llamadas pertinentes")	R-squared:	0.263			
Model:	OLS	Adj. R-squared:	0.261			
Method:	Least Squares	F-statistic:	150.2			
Date:	Tue, 26 May 2026	Prob (F-statistic):	1.05e-296			
Time:	18:23:00	Log-Likelihood:	-9463.9			
No. Observations:	4644	AIC:	1.895e+04			
Df Residuals:	4632	BIC:	1.903e+04			
Df Model:	11					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.7652	0.082	21.430	0.000	1.604	1.927
C(Provincia)[T.Bizkaia]	2.4132	0.069	35.200	0.000	2.279	2.548
C(Provincia)[T.Gipuzkoa]	0.4233	0.071	5.959	0.000	0.284	0.563
C(Año)[T.2022]	-0.1864	0.090	-2.069	0.039	-0.363	-0.010
C(Año)[T.2023]	-0.0742	0.090	-0.828	0.407	-0.250	0.101
C(Año)[T.2024]	0.2570	0.089	2.886	0.004	0.082	0.432
C(Año)[T.2025]	0.0725	0.091	0.796	0.426	-0.106	0.251
C(Año)[T.2026]	0.3743	0.144	2.596	0.009	0.092	0.657
hay_futbol	-0.0603	0.100	-0.604	0.546	-0.256	0.136
goles_totales	0.0104	0.020	0.528	0.597	-0.028	0.049
balance_goles	0.0054	0.026	0.210	0.834	-0.045	0.056
fin_semana	-0.5090	0.069	-7.410	0.000	-0.644	-0.374
Omnibus:	1413.214	Durbin-Watson:	1.822			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6537.881			
Skew:	1.403	Prob(JB):	0.00			
Kurtosis:	8.091	Cond. No.	17.5			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

C. GRÁFICOS

GRÁFICO 1: Evolución temporal de las llamadas al 016 en el País Vasco
(*evolucion_temporal*)

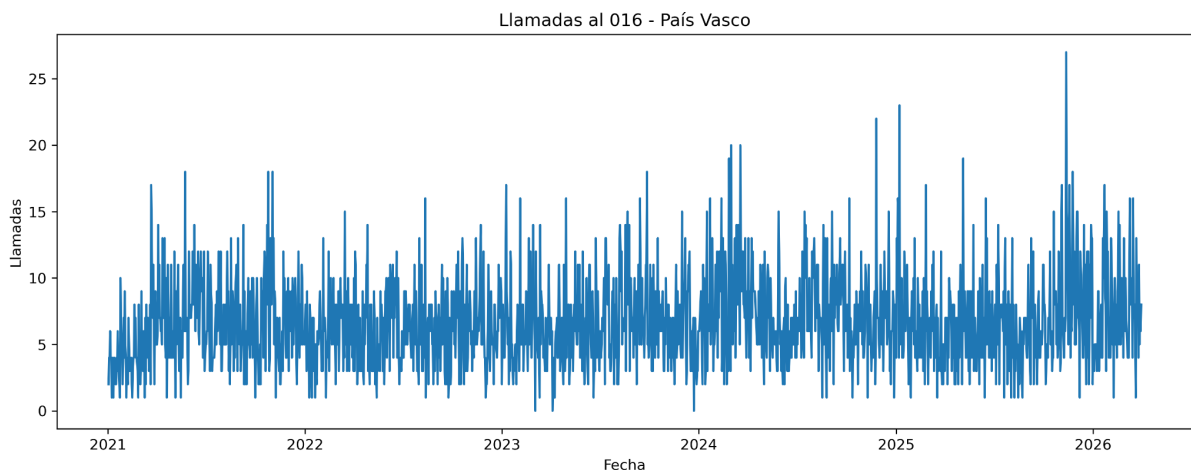


GRÁFICO 2: Evolución mensual media de las llamadas al 016
(evolucion_mensual)

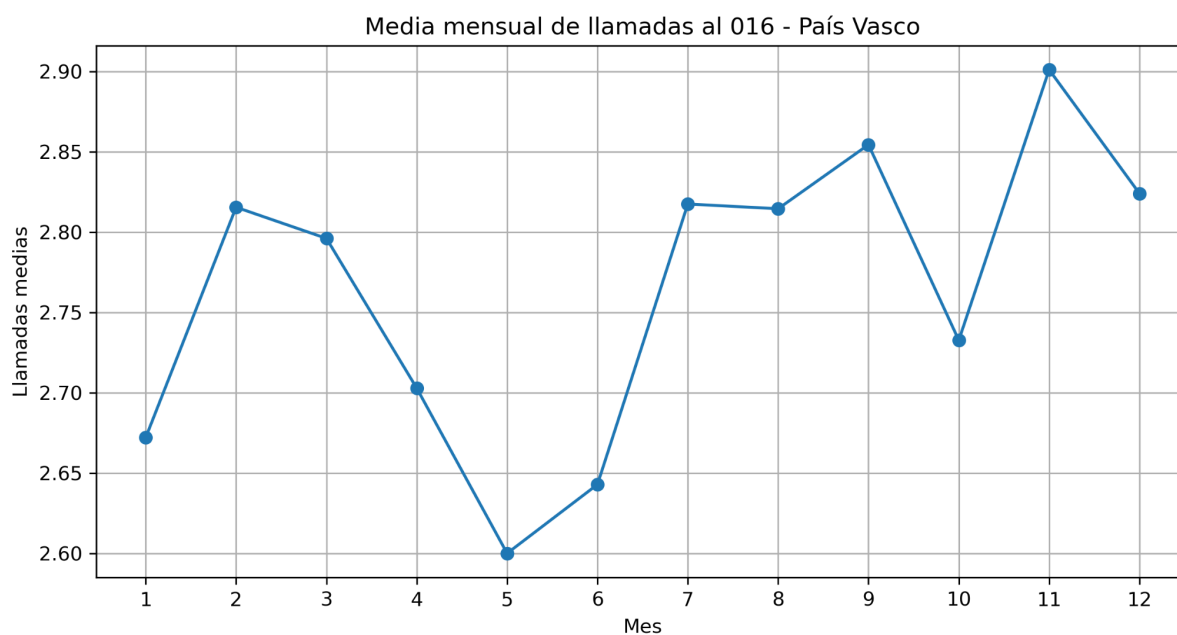


GRÁFICO 3: Evolución anual media de las llamadas al 016
(evolucion_anual)

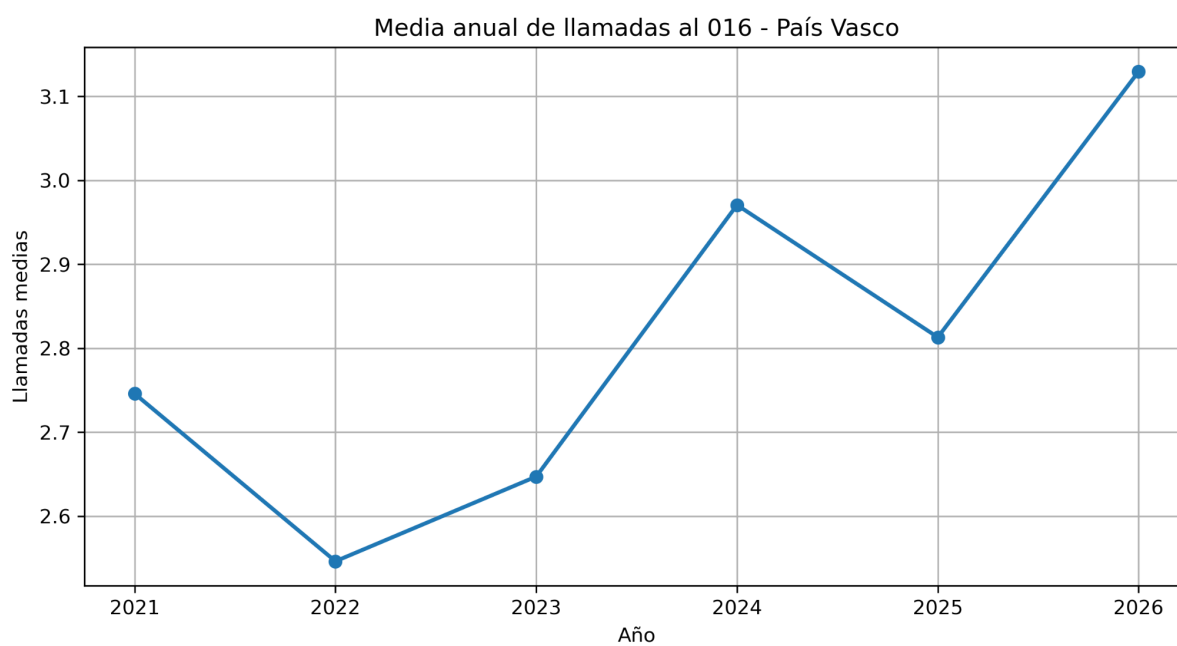


GRÁFICO 4: Media de llamadas al 016 según resultado del partido
(*resultados_partidos*)

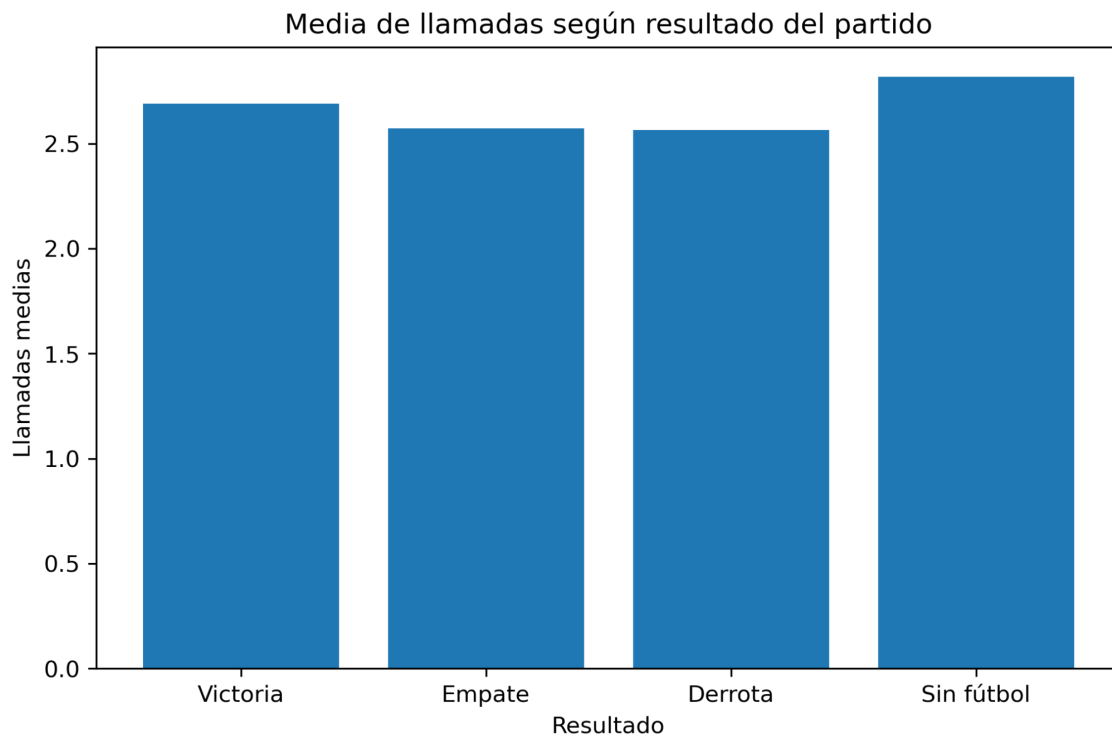


GRÁFICO 5: Distribución de llamadas al 016 según resultado del partido
(*boxplot_resultado*)

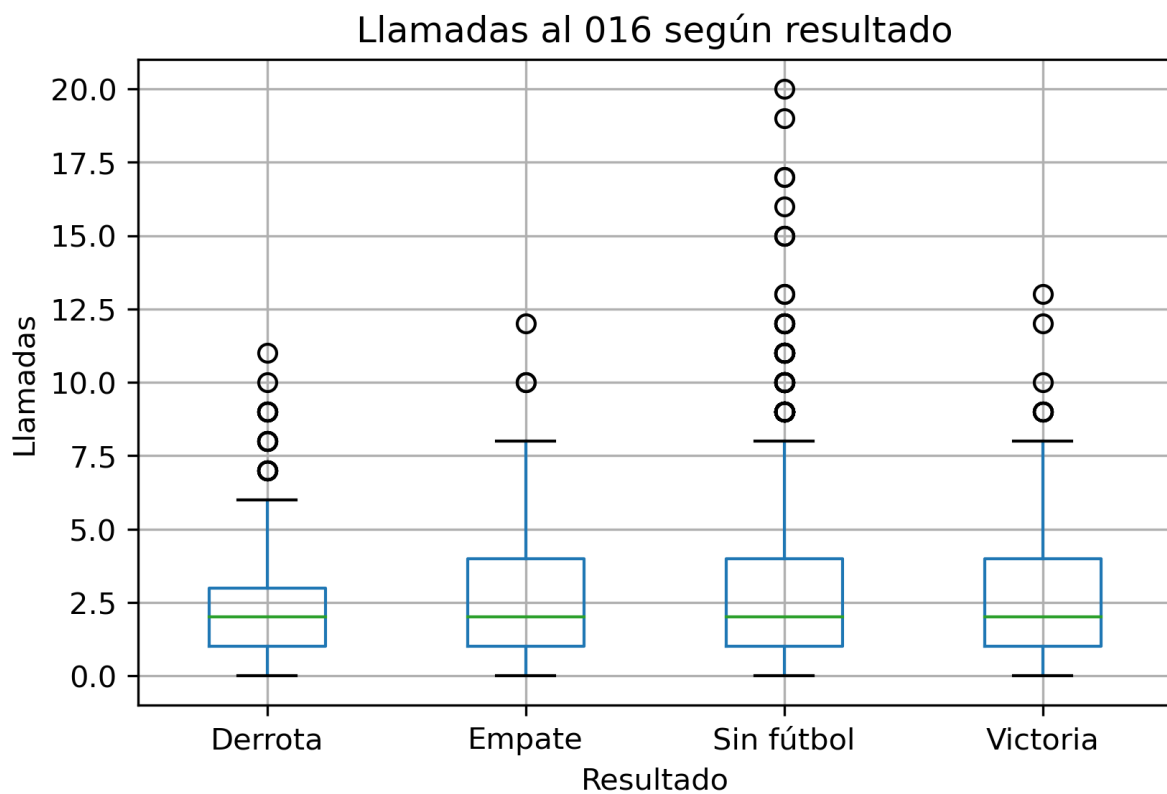


GRÁFICO 6: Comparación de llamadas al 016 entre semana y fin de semana
(fines_semana)

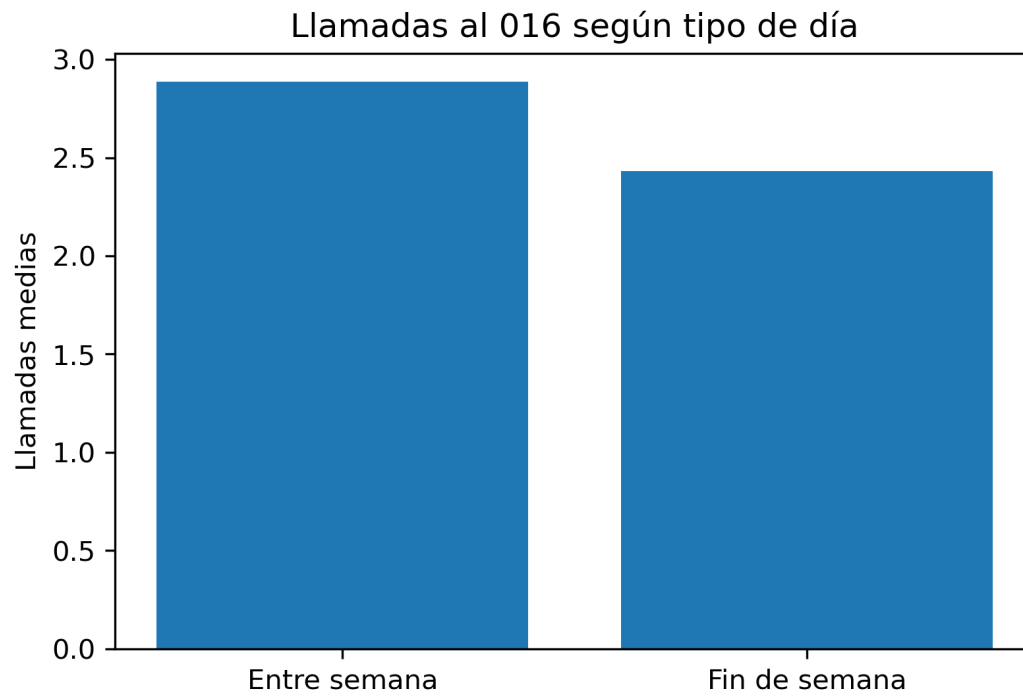


GRÁFICO 7: Llamadas al 016 en días de derbi vasco
(*derbis_vascos*)

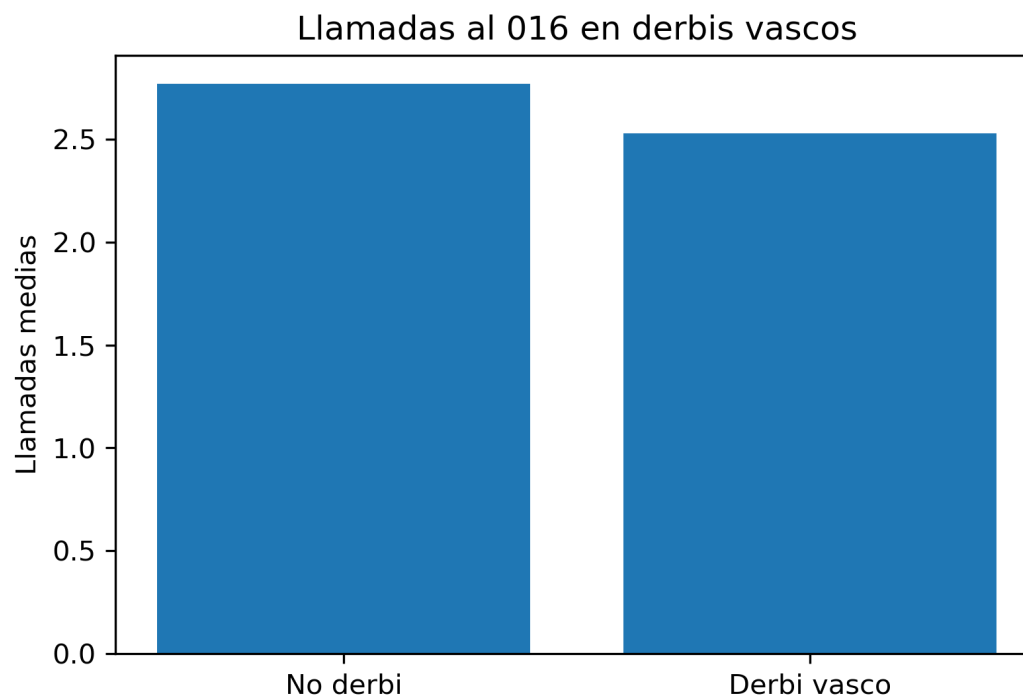
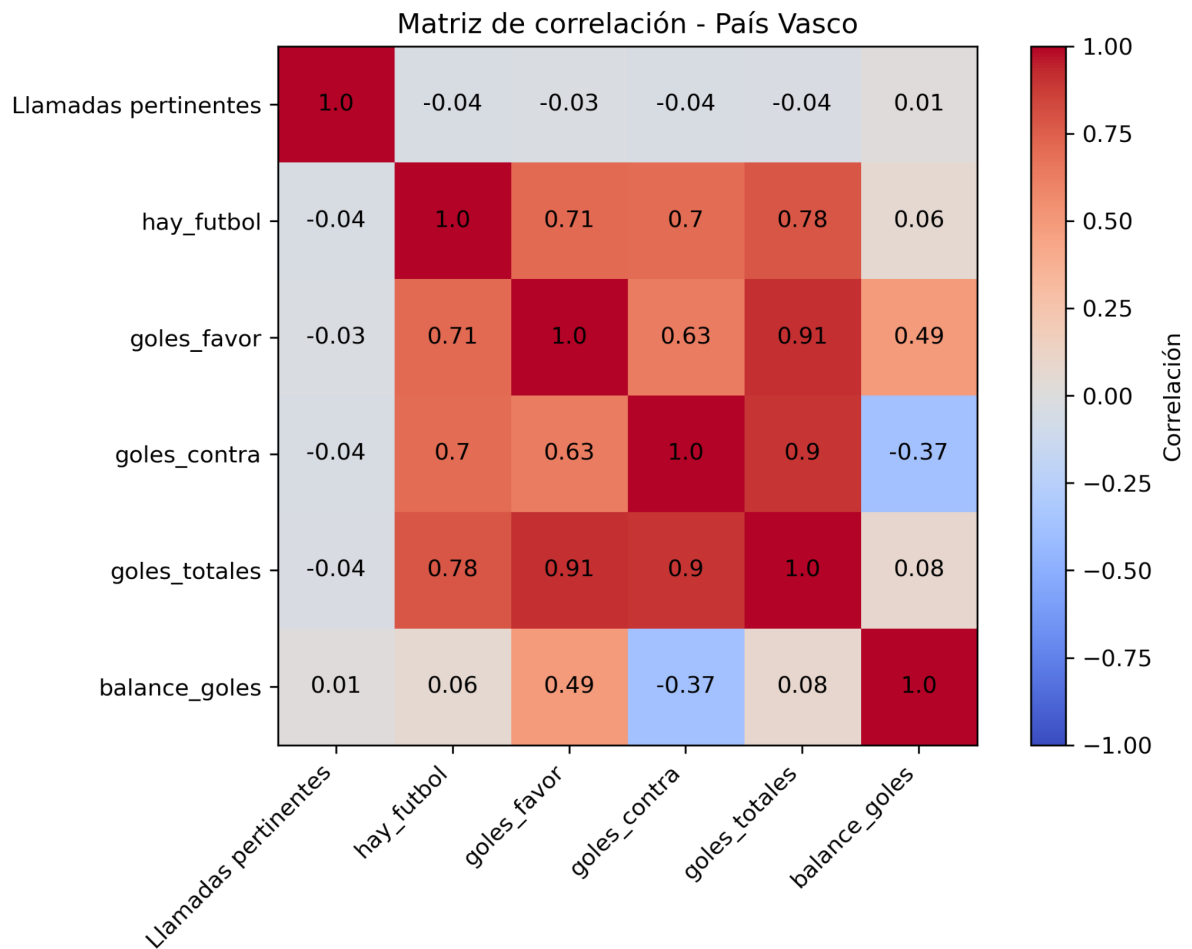


GRÁFICO 8: Matriz de correlación entre variables del estudio
(*matriz_correlacion*)



D. SCRIPTS

SCRIPT 1: 00_create_data_futbol_Real Sociedad_19

```
import requests
import pandas as pd
import time

API_KEY = "aabf02928fb9766eebdc2cc0f11ab56a"

headers = {
    "x-apisports-key": API_KEY
}

EQUIPO_REFERENCIA = "Real Sociedad" #se cambia
TEAM_ID = 548 #se cambia

SEASON = 2019 #se cambia
```

```

TEMPORADA = "2019-20" #se cambia

DIVISION = "1ª" #se cambia
CCAA = "País Vasco" #se cambia

url_fixtures = "https://v3.football.api-sports.io/fixtures"
url_events = "https://v3.football.api-sports.io/fixtures/events"

LEAGUES = {
    "La Liga": 140,
    "Copa del Rey": 143,
    "Supercopa": 556,
    "Champions League": 2,
    "Europa League": 3
}

def franja_30_minutos(fecha):
    hora = fecha.hour
    minuto = fecha.minute

    inicio_min = (minuto // 30) * 30
    fin_min = inicio_min + 30

    inicio = f"{hora:02d}:{inicio_min:02d}"

    if fin_min == 60:
        fin = f"{hora+1:02d}:00"
    else:
        fin = f"{hora:02d}:{fin_min:02d}"

    return f"{inicio}-{fin}"

def franja_gol(fecha, minuto_gol):
    total_min = fecha.hour * 60 + fecha.minute + minuto_gol

    hora = total_min // 60
    minuto = total_min % 60

    inicio_min = (minuto // 30) * 30
    fin_min = inicio_min + 30

    inicio = f"{hora:02d}:{inicio_min:02d}"

    if fin_min == 60:
        fin = f"{hora+1:02d}:00"
    else:

```

```

        fin = f"{hora:02d}:{fin_min:02d}"

    return f"{inicio}-{fin}"

meses_es = {
    1: "Enero",
    2: "Febrero",
    3: "Marzo",
    4: "Abril",
    5: "Mayo",
    6: "Junio",
    7: "Julio",
    8: "Agosto",
    9: "Septiembre",
    10: "Octubre",
    11: "Noviembre",
    12: "Diciembre"
}

LEAGUES = {
    "La Liga": 140,
    "Copa del Rey": 143,
    "Supercopa": 556,
    "Champions League": 2,
    "Europa League": 3
}

df_total = pd.DataFrame()

for nombre_liga, league_id in LEAGUES.items():

    print(f"Descargando {nombre_liga}...")

    params_fixtures = {
        "team": TEAM_ID,
        "league": league_id,
        "season": SEASON
    }

    response = requests.get(url_fixtures, headers=headers,
params=params_fixtures)
    fixtures_data = response.json()

    if fixtures_data["results"] == 0:
        continue

    filas = []

```

```

for partido in fixtures_data["response"]:

    fixture_id = partido["fixture"]["id"]
    fecha_dt =
pd.to_datetime(partido["fixture"]["date"]).tz_convert(None)

    equipo_local = partido["teams"]["home"]["name"]
    equipo_visitante = partido["teams"]["away"]["name"]
    competicion = partido["league"]["name"]
    jornada_raw = partido["league"]["round"]

    if "Regular Season" in jornada_raw:
        jornada = "J" + jornada_raw.split("-")[-1].strip()
    else:
        jornada = jornada_raw

    goles_home = partido["goals"]["home"]
    goles_away = partido["goals"]["away"]
    resultado = f"{goles_home}-{goles_away}"

    hora_inicio = fecha_dt.strftime("%H:%M")
    franja_partido = franja_30_minutos(fecha_dt)

    params_events = {"fixture": fixture_id}
    response_events = requests.get(url_events, headers=headers,
params=params_events)
    events_data = response_events.json()

    time.sleep(0.3)

    for e in events_data["response"]:
        if e["type"] == "Goal":

            equipo_marcador = e["team"]["name"]
            minuto = e["time"]["elapsed"]
            extra = e["time"]["extra"]

            minuto_gol = f"{minuto}+{extra}" if extra else minuto

            local_visit_marcador = "Local" if equipo_marcador ==
equipo_local else "Visitante"
            favor_contra = "A FAVOR" if equipo_marcador ==
EQUIPO_REFERENCIA else "EN CONTRA"

            filas.append({
                "Temporada": TEMPORADA,

```

```

        "Equipo referencia": EQUIPO_REFERENCIA,
        "División": DIVISION,
        "Mes": meses_es[fecha_dt.month],
        "Nº Mes": fecha_dt.month,
        "Año": fecha_dt.year,
        "Fecha": fecha_dt.strftime("%d/%m/%Y"),
        "Jornada": jornada,
        "Equipo Local": equipo_local,
        "Equipo Visitante": equipo_visitante,
        "Equipo Marcador": equipo_marcador,
        "Resultado": resultado,
        "Minuto Gol": minuto_gol,
        "Franja Partido": franja_partido,
        "Hora Inicio": hora_inicio,
        "Franja Gol": franja_gol(fecha_dt, minuto),
        "Local/Visit.": local_visit_marcador,
        "Competición": competicion,
        "Notas": "",
        "CCAA": CCAA,
        "Gol A FAVOR / EN CONTRA": favor_contra
    })

```

```

df_liga = pd.DataFrame(filas)
df_total = pd.concat([df_total, df_liga], ignore_index=True)

```

```
df_total.head()
```

```

nombre_archivo = f"{EQUIPO_REFERENCIA.replace(' ',
'_' )}_{TEMPORADA}_COMPLETO.xlsx" #se cambia
df_total.to_excel(nombre_archivo, index=False)

```

```

df_total["Competición"].value_counts()
df_total.shape

```

SCRIPT 2: 01_limpieza_016

```
import pandas as pd
```

```
# LIMPIEZA DATOS 016
```

```
# Datos del Servicio 016 franja horaria (Provincia).xlsx
```

```

df = pd.read_excel(
    "Datos del Servicio 016 franja horaria (Provincia).xlsx",
    header=3
)

```

```

df.head()
df.columns

df.info()
df.head()

# Convertir el año a numérico
df["Año"] = df["Año"].str.replace("Año ", "").astype(int)
df.head()
df.info()

# transformar los meses de texto (Abril, Agosto, etc.) en números.
meses = {
    "Enero": 1,
    "Febrero": 2,
    "Marzo": 3,
    "Abril": 4,
    "Mayo": 5,
    "Junio": 6,
    "Julio": 7,
    "Agosto": 8,
    "Septiembre": 9,
    "Octubre": 10,
    "Noviembre": 11,
    "Diciembre": 12
}

df["Mes_num"] = df["Mes"].map(meses)
df.head()

df.sample(20)
df["Mes"].unique()

# Datos del Servicio 016 (Provincia diarios).xlsx
df_dia = pd.read_excel(
    "Datos del Servicio 016 (Provincia diarios).xlsx",
    header=3
)

df_dia.head()
df_dia.columns

# Convertir año
df_dia["Año"] = df_dia["Año"].str.replace("Año ", "").astype(int)

# Convertir meses a número

```

```

df_dia["Mes_num"] = df_dia["Mes"].map(meses)

df_dia.head()

# Crear una fecha real (datetime)
df_dia["Fecha"] = pd.to_datetime({
    "year": df_dia["Año"],
    "month": df_dia["Mes_num"],
    "day": df_dia["Día"]
})

df_dia.head()

df_dia.info()

df_dia = df_dia.sort_values(["Provincia", "Fecha"])

df_dia = df_dia[
    [
        "Fecha",
        "Año",
        "Mes",
        "Mes_num",
        "Día",
        "Provincia",
        "Llamadas pertinentes"
    ]
]

df_dia.head()

df_dia.to_csv("016_diario_limpio.csv", index=False)

# Diccionario Provincia → CCAA
provincia_ccaa = {

    "A Coruña": "Galicia",
    "Álava": "País Vasco",
    "Araba/Álava": "País Vasco",

    "Albacete": "Castilla-La Mancha",

    "Alicante": "Comunidad Valenciana",
    "Alicante/Alacant": "Comunidad Valenciana",

    "Almería": "Andalucía",
    "Asturias": "Asturias",

```

"Ávila": "Castilla y León",

"Badajoz": "Extremadura",
"Barcelona": "Cataluña",
"Bizkaia": "País Vasco",
"Burgos": "Castilla y León",

"Cáceres": "Extremadura",
"Cádiz": "Andalucía",
"Cantabria": "Cantabria",

"Castellón": "Comunidad Valenciana",
"Castellón/Castelló": "Comunidad Valenciana",

"Ciudad Real": "Castilla-La Mancha",
"Córdoba": "Andalucía",
"Cuenca": "Castilla-La Mancha",

"Gipuzkoa": "País Vasco",
"Girona": "Cataluña",
"Granada": "Andalucía",
"Guadalajara": "Castilla-La Mancha",

"Huelva": "Andalucía",
"Huesca": "Aragón",

"Illes Balears": "Balears",

"Jaén": "Andalucía",

"La Rioja": "La Rioja",
"Las Palmas": "Canarias",
"León": "Castilla y León",
"Lleida": "Cataluña",
"Lugo": "Galicia",

"Madrid": "Comunidad de Madrid",
"Málaga": "Andalucía",
"Murcia": "Región de Murcia",

"Navarra": "Navarra",

"Ourense": "Galicia",

"Palencia": "Castilla y León",
"Pontevedra": "Galicia",

```

    "Salamanca": "Castilla y León",
    "Santa Cruz de Tenerife": "Canarias",
    "Segovia": "Castilla y León",
    "Sevilla": "Andalucía",
    "Soria": "Castilla y León",

    "Tarragona": "Cataluña",
    "Teruel": "Aragón",
    "Toledo": "Castilla-La Mancha",

    "Valencia": "Comunidad Valenciana",
    "Valencia/València": "Comunidad Valenciana",

    "Valladolid": "Castilla y León",

    "Zamora": "Castilla y León",
    "Zaragoza": "Aragón",

    "Ceuta": "Ceuta",
    "Melilla": "Melilla",
    "No consta": "No consta"
}

# Crear columna CCAA
df_dia["CCAA"] = df_dia["Provincia"].map(provincia_ccaa)

# Comprobar si queda alguna provincia sin asignar
df_dia[df_dia["CCAA"].isna()]["Provincia"].unique()

```

SCRIPT 3: 02_limpieza_futbol

```

import pandas as pd

futbol = pd.read_excel(
    "Goles versión definitiva.xlsx",
    sheet_name="Registro de Goles",
    header=2
)

futbol.head()
futbol.columns

futbol.info()
futbol.head()

```

```

# Convertir fecha
futbol["Fecha"] = pd.to_datetime(
    futbol["Fecha"],
    dayfirst=True
)

futbol.info()

# Ordenar columnas
futbol = futbol[
    [
        "Fecha",
        "Temporada",
        "Año",
        "Mes",
        "Nº Mes",
        "Equipo referencia",
        "Equipo Local",
        "Equipo Visitante",
        "Resultado",
        "Minuto Gol",
        "Hora Inicio",
        "Franja Partido",
        "Franja Gol",
        "Competición",
        "Local/Visit.",
        "Gol A FAVOR / EN CONTRA",
        "CCAA",
        "División",
        "Jornada"
    ]
]

# Ordenar temporalmente
futbol = futbol.sort_values(["Equipo referencia", "Fecha"])

# Unificar nombres de CCAA
futbol["CCAA"] = futbol["CCAA"].replace({
    "Comunitat Valenciana": "Comunidad Valenciana"
})

# Comprobar CCAA únicas
sorted(futbol["CCAA"].unique())

futbol.to_csv("futbol_limpio.csv", index=False)

```

SCRIPT 4: 03_merge_datasets

```
import pandas as pd

# Cargar datasets limpios
df_dia = pd.read_csv("016_diario_limpio.csv")
futbol = pd.read_csv("futbol_limpio.csv")

df_dia.info()
futbol.info()

df_dia.head()
futbol.head()

# Convertir fechas otra vez a datetime
df_dia["Fecha"] = pd.to_datetime(df_dia["Fecha"])
futbol["Fecha"] = pd.to_datetime(futbol["Fecha"])

# Comprobar tipos
df_dia["Fecha"].dtype
futbol["Fecha"].dtype

# Añadir CCAA al 016
provincia_ccaa = {
    "A Coruña": "Galicia",
    "Álava": "País Vasco",
    "Araba/Álava": "País Vasco",
    "Albacete": "Castilla-La Mancha",
    "Alicante": "Comunidad Valenciana",
    "Alicante/Alacant": "Comunidad Valenciana",
    "Almería": "Andalucía",
    "Asturias": "Asturias",
    "Ávila": "Castilla y León",
    "Badajoz": "Extremadura",
    "Barcelona": "Cataluña",
    "Bizkaia": "País Vasco",
    "Burgos": "Castilla y León",
    "Cáceres": "Extremadura",
    "Cádiz": "Andalucía",
    "Cantabria": "Cantabria",
    "Castellón": "Comunidad Valenciana",
    "Castellón/Castelló": "Comunidad Valenciana",
    "Ciudad Real": "Castilla-La Mancha",
    "Córdoba": "Andalucía",
    "Cuenca": "Castilla-La Mancha",
    "Gipuzkoa": "País Vasco",
    "Girona": "Cataluña",
```

```

"Granada": "Andalucía",
"Guadalajara": "Castilla-La Mancha",
"Huelva": "Andalucía",
"Huesca": "Aragón",
"Illes Balears": "Balears",
"Jaén": "Andalucía",
"La Rioja": "La Rioja",
"Las Palmas": "Canarias",
"León": "Castilla y León",
"Lleida": "Cataluña",
"Lugo": "Galicia",
"Madrid": "Comunidad de Madrid",
"Málaga": "Andalucía",
"Murcia": "Región de Murcia",
"Navarra": "Navarra",
"Ourense": "Galicia",
"Palencia": "Castilla y León",
"Pontevedra": "Galicia",
"Salamanca": "Castilla y León",
"Santa Cruz de Tenerife": "Canarias",
"Segovia": "Castilla y León",
"Sevilla": "Andalucía",
"Soria": "Castilla y León",
"Tarragona": "Cataluña",
"Teruel": "Aragón",
"Toledo": "Castilla-La Mancha",
"Valencia": "Comunidad Valenciana",
"Valencia/València": "Comunidad Valenciana",
"Valladolid": "Castilla y León",
"Zamora": "Castilla y León",
"Zaragoza": "Aragón",
"Ceuta": "Ceuta",
"Melilla": "Melilla",
"No consta": "No consta"
}

df_dia["CCAA"] = df_dia["Provincia"].map(provincia_ccaa)

df_dia[df_dia["CCAA"].isna()]["Provincia"].unique()

# Crear variables de goles
futbol["gol_favor"] = (futbol["Gol A FAVOR / EN CONTRA"] == "A
FAVOR").astype(int)
futbol["gol_contra"] = (futbol["Gol A FAVOR / EN CONTRA"] == "EN
CONTRA").astype(int)

# Agregar fútbol por Fecha y CCAA

```

```

futbol_diario = futbol.groupby(["Fecha", "CCAA"]).agg(
    goles_favor=("gol_favor", "sum"),
    goles_contra=("gol_contra", "sum"),
    goles_totales=("Minuto Gol", "count"),
    equipos=("Equipo referencia", lambda x: ",
".join(sorted(x.unique()))),
    competiciones=("Competición", lambda x: ", ".join(sorted(x.unique()))
).reset_index()

futbol_diario.head()

# MERGE
master = df_dia.merge(
    futbol_diario,
    on=["Fecha", "CCAA"],
    how="left"
)

master.head()

master["goles_favor"] = master["goles_favor"].fillna(0).astype(int)
master["goles_contra"] = master["goles_contra"].fillna(0).astype(int)
master["goles_totales"] = master["goles_totales"].fillna(0).astype(int)

master["hay_futbol"] = (master["goles_totales"] > 0).astype(int)

# Crear balance de goles
master["balance_goles"] = master["goles_favor"] - master["goles_contra"]

# Crear resultado correctamente
master["resultado_partido"] = "Sin fútbol"

master.loc[
    (master["hay_futbol"] == 1) & (master["balance_goles"] > 0),
    "resultado_partido"
] = "Victoria"

master.loc[
    (master["hay_futbol"] == 1) & (master["balance_goles"] < 0),
    "resultado_partido"
] = "Derrota"

master.loc[
    (master["hay_futbol"] == 1) & (master["balance_goles"] == 0),
    "resultado_partido"
] = "Empate"

```

```

# Comprobar
master["resultado_partido"].value_counts()

master.head()

master.to_csv("master_016_futbol_cruzado.csv", index=False)

master.info()
master.head()

```

SCRIPT 5: 04_analisis_pais_vasco

```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.formula.api as smf

import os

os.makedirs("graficos", exist_ok=True)
os.makedirs("tablas", exist_ok=True)
os.makedirs("modelos", exist_ok=True)

# IMPORTAR Y CARGAR BASE
master = pd.read_csv(
    "master_016_futbol_cruzado.csv"
) # cargar base final

master["Fecha"] = pd.to_datetime(
    master["Fecha"]
) # cambiar fecha datetime

pv = master[
    master["CCAA"] == "País Vasco"
].copy()
#filtrar por mi ccaa (pais vasco)

pv.head()

pv.info()

pv.describe()

```

```

pv.describe().to_csv(
    "tablas/descriptivos_generales.csv"
)
#comprobar base

# DESCRIPCIÓN GENERAL
print("Número de observaciones:")
print(len(pv))

print("\nTotal llamadas pertinentes:")
print(
    pv["Llamadas pertinentes"].sum()
)

print("\nMedia diaria de llamadas:")
print(
    round(
        pv["Llamadas pertinentes"].mean(),
        2
    )
)

print("\nDías con fútbol:")
print(
    pv["hay_futbol"].value_counts()
)

print("\nResultados:")
print(
    pv["resultado_partido"].value_counts()
)

print("\nProvincias:")
print(
    pv["Provincia"].value_counts()
)

# EVOLUCION TEMPORAL
pv_diario = pv.groupby(
    "Fecha"
)[ "Llamadas pertinentes"].sum()

plt.figure(figsize=(14,5))

plt.plot(pv_diario)

plt.title(

```

```

    "Llamadas al 016 - País Vasco"
)

plt.xlabel("Fecha")

plt.ylabel("Llamadas")

plt.savefig(
    "graficos/evolucion_temporal.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

pv.groupby(
    "hay_futbol"
)["Llamadas pertinentes"].mean()
#ver si hay mas llamadas cuando hay futbol

pv.groupby(
    "resultado_partido"
)["Llamadas pertinentes"].mean()
#resultado del partido

# solo dias que hay futbol
pv_futbol = pv[
    pv["hay_futbol"] == 1
]

pv_futbol.groupby(
    "resultado_partido"
)["Llamadas pertinentes"].mean()

# boxplot para la distribucion (visual)
plt.figure(figsize=(8,5))

pv.boxplot(
    column="Llamadas pertinentes",
    by="resultado_partido"
)

plt.title(
    "Llamadas al 016 según resultado"
)

plt.suptitle("")

```

```

plt.xlabel("Resultado")

plt.ylabel("Llamadas")

plt.savefig(
    "graficos/boxplot_resultado.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

#los dias con mas llamadas
pv.sort_values(
    "Llamadas pertinentes",
    ascending=False
).head(20)

# DERBIS VASCOS
pv["derbi_vasco"] = pv[
    "equipos"
].str.contains(
    "Athletic Club Bilbao, Real Sociedad",
    na=False
).astype(int)

pv.groupby(
    "derbi_vasco"
)["Llamadas pertinentes"].mean()

# CORRELACION
pv[
    [
        "Llamadas pertinentes",
        "goles_favor",
        "goles_contra",
        "goles_totales"
    ]
].corr()

# MATRIZ DE CORRELACION
variables_corr = [
    "Llamadas pertinentes",
    "hay_futbol",
    "goles_favor",
    "goles_contra",

```

```

    "goles_totales",
    "balance_goles"
]

corr = pv[variables_corr].corr()

corr

# visualizar matriz
plt.figure(figsize=(8,6))

plt.imshow(corr, cmap="coolwarm", vmin=-1, vmax=1)

plt.colorbar(label="Correlación")

plt.xticks(
    range(len(corr.columns)),
    corr.columns,
    rotation=45,
    ha="right"
)

plt.yticks(
    range(len(corr.columns)),
    corr.columns
)

# meter valores dentro de cada celda
for i in range(len(corr.columns)):
    for j in range(len(corr.columns)):
        plt.text(
            j,
            i,
            round(corr.iloc[i, j], 2),
            ha="center",
            va="center"
        )

plt.title("Matriz de correlación - País Vasco")
plt.tight_layout()
plt.savefig(
    "graficos/matriz_correlacion.png",
    dpi=300,
    bbox_inches="tight"
)
plt.show()

```

```

# COMPARACIÓN POR PROVINCIAS
# media de llamadas por prov
pv.groupby("Provincia")["Llamadas pertinentes"].mean()

# llamadas con fútbol por prov
pv.groupby(
    ["Provincia", "hay_futbol"]
)["Llamadas pertinentes"].mean()

# resultado del partido por prov
pv.groupby(
    ["Provincia", "resultado_partido"]
)["Llamadas pertinentes"].mean()

# EVOLUCIÓN MES
# media mensual de llamadas
pv.groupby("Mes_num")["Llamadas pertinentes"].mean()

mensual = pv.groupby("Mes_num")["Llamadas pertinentes"].mean()

plt.figure(figsize=(10,5))

plt.plot(
    mensual.index,
    mensual.values,
    marker="o"
) # visualizar

plt.xticks(range(1,13))

plt.title("Media mensual de llamadas al 016 - País Vasco")
plt.xlabel("Mes")
plt.ylabel("Llamadas medias")

plt.grid(True)

plt.savefig(
    "graficos/evolucion_mensual.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

# comparar fútbol por mes
pv.groupby(

```

```

    ["Mes_num", "hay_futbol"]
)["Llamadas pertinentes"].mean()

# FINES DE SEMANA
# CREAR VARIABLE FIN DE SEMANA
pv["dia_semana"] = pv["Fecha"].dt.dayofweek
pv["fin_semana"] = (
    pv["dia_semana"] >= 5
).astype(int)

pv[[
    "Fecha",
    "dia_semana",
    "fin_semana"
]].head()
# comparar llamadas (finde)
pv.groupby(
    "fin_semana"
)["Llamadas pertinentes"].mean()

# futbol + finde
pv.groupby(
    ["fin_semana", "hay_futbol"]
)["Llamadas pertinentes"].mean()

from scipy.stats import ttest_ind

# COMPARAR FÚTBOL VS NO FÚTBOL
con_futbol = pv[
    pv["hay_futbol"] == 1
]["Llamadas pertinentes"]

sin_futbol = pv[
    pv["hay_futbol"] == 0
]["Llamadas pertinentes"] # separar por grupos

# t-test
ttest_ind(
    con_futbol,
    sin_futbol,
    equal_var=False
)

# evol anual + grafico
pv.groupby("Año")["Llamadas pertinentes"].mean()

```

```

anual = pv.groupby("Año")["Llamadas pertinentes"].mean()

plt.figure(figsize=(10,5))

plt.plot(
    anual.index,
    anual.values,
    marker="o",
    linewidth=2
)

plt.title("Media anual de llamadas al 016 - País Vasco")
plt.xlabel("Año")
plt.ylabel("Llamadas medias")

plt.grid(True)

plt.savefig(
    "graficos/evolucion_anual.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

pv.groupby(
    ["Año", "hay_futbol"]
)["Llamadas pertinentes"].mean() #comp. futbol x año

# REGRESION
import statsmodels.formula.api as smf

pv["fin_semana"] = (
    pv["Fecha"].dt.dayofweek >= 5
).astype(int) # var fin de semana

modelo2 = smf.ols(
    formula="""
    Q("Llamadas pertinentes")
    ~ hay_futbol
    + goles_totales
    + balance_goles
    + fin_semana
    + C(Provincia)
    + C(Año)
    """,

```

```

        data=pv
    ).fit()

print(modelo2.summary())

# media
media_resultado = pv.groupby(
    "resultado_partido"
)["Llamadas pertinentes"].mean()

media_resultado = media_resultado[
    ["Victoria", "Empate", "Derrota", "Sin fútbol"]
] #orden

plt.figure(figsize=(8,5))

plt.bar(
    media_resultado.index,
    media_resultado.values
)

plt.title(
    "Media de llamadas según resultado del partido"
)

plt.xlabel("Resultado")
plt.ylabel("Llamadas medias")

plt.savefig(
    "graficos/resultados_partido.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

pv["fin_semana"] = (
    pv["Fecha"].dt.dayofweek >= 5
).astype(int) #var finde

pv.groupby(
    "fin_semana"
)["Llamadas pertinentes"].mean() #media

# grafico

```

```

media_fs = pv.groupby(
    "fin_semana"
)["Llamadas pertinentes"].mean()

plt.figure(figsize=(6,4))
plt.bar(
    ["Entre semana", "Fin de semana"],
    media_fs.values
)
plt.title(
    "Llamadas al 016 según tipo de día"
)
plt.ylabel("Llamadas medias")

plt.savefig(
    "graficos/fines_semana.png",
    dpi=300,
    bbox_inches="tight"
)
plt.show()

# DERBIS
pv["derbi_vasco"] = (
    pv["equipos"]
    .str.contains(
        "Athletic Club Bilbao, Real Sociedad",
        na=False
    )
).astype(int) #var derbi

pv.groupby(
    "derbi_vasco"
)["Llamadas pertinentes"].mean() #media llamadas

# grafico
media_derbi = pv.groupby(
    "derbi_vasco"
)["Llamadas pertinentes"].mean()

plt.figure(figsize=(6,4))

plt.bar(
    ["No derbi", "Derbi vasco"],
    media_derbi.values
)

```

```

plt.title(
    "Llamadas al 016 en derbis vascos"
)

plt.ylabel("Llamadas medias")

plt.savefig(
    "graficos/derbis_vascos.png",
    dpi=300,
    bbox_inches="tight"
)

plt.show()

print("=====")
print("RESUMEN DEL ANÁLISIS EMPÍRICO")
print("=====")

# futbol vs no futbol
media_futbol = pv.groupby(
    "hay_futbol"
)["Llamadas pertinentes"].mean()

print("\n1. MEDIA DE LLAMADAS")

print(
    "\nSin fútbol:",
    round(media_futbol[0], 2)
)

print(
    "Con fútbol:",
    round(media_futbol[1], 2)
)

# resultado partido
resultado_media = pv.groupby(
    "resultado_partido"
)["Llamadas pertinentes"].mean()

print("\n2. RESULTADO DEL PARTIDO")

print(resultado_media)

# finde
fs_media = pv.groupby(

```

```

        "fin_semana"
    )["Llamadas pertinentes"].mean()

print("\n3. FIN DE SEMANA")

print(
    "\nEntre semana:",
    round(fs_media[0], 2)
)

print(
    "Fin de semana:",
    round(fs_media[1], 2)
)

# derbis vascos
derbi_media = pv.groupby(
    "derbi_vasco"
)["Llamadas pertinentes"].mean()

print("\n4. DERBIS VASCOS")

print(
    "\nNo derbi:",
    round(derbi_media[0], 2)
)

print(
    "Derbi:",
    round(derbi_media[1], 2)
)

# provs
prov_media = pv.groupby(
    "Provincia"
)["Llamadas pertinentes"].mean()

print("\n5. PROVINCIAS")

print(prov_media)

print("=====")

# EXPORTAR TABLAS Y MODELO
pv.to_csv(
    "base_final_pais_vasco.csv",

```

```

        index=False
    )

    pv.groupby(
        "hay_futbol"
    )["Llamadas pertinentes"].mean().to_csv(
        "tablas/futbol_no_futbol.csv"
    )

    pv.groupby(
        "resultado_partido"
    )["Llamadas pertinentes"].mean().to_csv(
        "tablas/resultados_partido.csv"
    )

    pv.groupby(
        "Provincia"
    )["Llamadas pertinentes"].mean().to_csv(
        "tablas/provincias.csv"
    )

    pv.groupby(
        ["Provincia", "resultado_partido"]
    )["Llamadas pertinentes"].mean().to_csv(
        "tablas/provincias_resultado.csv"
    )

    corr.to_csv(
        "tablas/matriz_correlacion.csv"
    )

    with open(
        "modelos/regresion_ols.txt",
        "w"
    ) as f:
        f.write(
            modelo2.summary().as_text()
        )

```

