



Facultad de Ciencias Económicas y Empresariales
ICADE

EL IMPACTO DE LOS CONCIERTOS DEL SANTIAGO BERNABÉU EN MADRID: LOS VECINOS, EL TURISMO Y LAS EMPRESAS

Autor: Laura Löwen Martín-Neda
Director: Luis Ángel Calvo Pascual

MADRID | Junio 2025

RESUMEN

Este Trabajo de Fin de Grado examina de manera integral el impacto de los conciertos celebrados en el Estadio Santiago Bernabéu sobre tres dimensiones fundamentales de Madrid: los vecinos, el turismo y las empresas locales. Para el análisis de los vecinos, se emplea un pipeline de procesamiento de lenguaje natural sobre más de 29 000 mensajes de un grupo vecinal de WhatsApp, combinado con técnicas de análisis de sentimiento (BERT corregido), análisis de emociones (léxico NRC), detección de menciones de decibelios y modelos estadísticos (DiD, RDD y clustering). En el ámbito turístico y empresarial, se aplican modelos Difference-in-Differences, así como test de hipótesis a datos mensuales del INE para cuantificar variaciones en el número de turistas, gasto turístico, ocupación hotelera y pernoctaciones, comparando Madrid frente al conjunto de España. Los resultados muestran que los conciertos generan un aumento recurrente en las quejas vecinales (especialmente en los días previos), un impulso significativo en los indicadores turísticos y beneficios agregados para el comercio local. Se proponen recomendaciones estratégicas de política pública y gestión privada para mitigar los efectos negativos en el entorno residencial (horarios límite, mejores insonorizaciones) y potenciar los beneficios económicos (promoción turística, dinamización comercial).

Keywords (español):

Impacto de conciertos; Análisis de sentimiento; Difference-in-Differences; Turismo urbano; Empresas locales; Ruido acústico; Estadio Santiago Bernabéu.

ABSTRACT

This Bachelor's thesis provides a comprehensive assessment of the impact of concerts held at the Santiago Bernabéu Stadium on three key aspects of Madrid: residents, tourism, and local businesses. For the residents' analysis, a natural language processing pipeline was applied to over 29,000 WhatsApp messages from a neighbourhood group, including sentiment analysis (using a BERT model with custom corrections), emotion detection (NRC lexicon), decibel mention extraction, and statistical models (Difference-in-Differences, Regression Discontinuity, and clustering). In the tourism and business sectors, Difference-in-Differences and Hypothesis testing models were employed on monthly INE data to quantify changes in tourist arrivals, total expenditure, hotel occupancy, and overnight stays, comparing Madrid against Spain as a control group. Findings indicate that concerts consistently increase resident complaints (particularly in the days leading up to events), significantly boost tourism indicators, and generate aggregated benefits for local commerce. Strategic recommendations are offered for public policy and private management to mitigate negative effects in residential areas (curfews, enhanced sound insulation) and maximize economic benefits (tourist promotion, commercial activation).

Keywords (English):

Concert impact; Sentiment analysis; Difference-in-Differences; Urban tourism; Local businesses; Acoustic noise; Santiago Bernabéu Stadium.

Índice de contenidos

1. INTRODUCCIÓN	11
1.1 CONTEXTO DEL ESTADIO SANTIAGO BERNABÉU	11
1.1.1 Transformación y obras del Bernabéu (2024)	12
1.1.2 Expansión a eventos no deportivos y el papel de los conciertos	15
1.1.3 Impacto económico y social	15
1.2 MOTIVACIÓN PERSONAL	16
1.3 RELEVANCIA ACADÉMICA Y SOCIAL	17
1.3 FUNDAMENTO CONCEPTUAL Y CONTEXTO NORMATIVO	17
1.3.1 Eventos masivos como motores económicos y culturales	17
1.3.2 Impacto directo, indirecto sobre la economía local	18
1.3.3 El vínculo entre conciertos y turismo en Madrid	18
1.3.4 El ruido como eje del conflicto urbano	19
1.3.5 Complejidad técnica y desafíos de insonorización	20
1.3.6 Un conflicto urbano entre rentabilidad y habitabilidad	20
1.3.7 Valor añadido del enfoque metodológico	21
1.4 OBJETIVO DEL ESTUDIO	21
1.4.1 Impacto sobre los vecinos	22
1.4.2 Impacto en el turismo	22
1.4.3 Impacto en las empresas locales	23
1.6 ANTECEDENTES	24
1.7 ESTRUCTURA DEL TFG	26
2. IMPACTO SOBRE LOS VECINOS	29
2.1 ANÁLISIS DE EXPLORATORIO DE DATOS	29
2.1.1 Objetivo del análisis	29
2.1.2 Descripción de los datos	29
2.1.3 Preprocesamiento del texto	29
2.1.3 Evolución temporal de la actividad	30
2.1.5 Sentimiento medio diario	31
2.1.6 Análisis de emociones específicas (librería NRC)	37
2.1.7 Nubes de Palabras y Patrones Léxicos	40
2.1.8 Análisis de decibelios y detección de excesos de ruido	45
2.2 ANÁLISIS DE PATRONES MEDIANTE TÉCNICAS DE MACHINE LEARNING	48
2.2.1 Modelos estadísticos	48
2.2.2 Resumen estadístico de contrastes	49
2.2.3 Visualización de la distribución de sentimiento	49
2.2.4 Evolución de quejas alrededor de los conciertos	51
2.2.5 TF-IDF + K-MEANS	52
2.2.6 Regresión logística (ventana ± 1 días)	60
2.2.7 Regresión logística multivariable con decibelios, longitud, emojis, etc.	61
2.2.8 Difference-in-Differences	63
2.2.9 Regression Discontinuity Design	68
3. IMPACTO EN EL TURISMO	74
3.1 METODOLOGÍA GENERAL	74
3.2 ANÁLISIS DE LOS DATOS	76
3.2.1 Gasto turístico INE	76
3.2.2 Gasto Turístico Nacional	77
3.2.3 Gasto turístico Madrid	77
3.2.4 Turismo móvil	77
3.2.5 Turismo móvil nacional	78
3.2.6 Turismo móvil Madrid	79

4. IMPACTO EN EMPRESAS	80
4.1 METODOLOGÍA	80
4.2 ANÁLISIS DE LOS DATOS	81
4.2.1 Movimientos turísticos	81
4.2.2 Ocupación hotelera	82
4.2.3 Pernotaciones	84
4.2.4 Evolución mensual de los viajeros y las pernoctaciones por tipo de alojamiento y por origen	86
4.2.5 Encuestas de ocupación por tipo de alojamiento	87
5. PRUEBAS DE ROBUSTEZ Y ANÁLISIS COMPLEMENTARIOS PARA TURISMO Y EMPRESAS	89
6. CONCLUSIÓN GENERAL.....	91
6.1 ANÁLISIS DESCRIPTIVO DE LOS RESULTADOS	91
6.1.1 Impacto en vecinos	91
6.1.2 Impacto en el turismo.....	92
6.1.3 Impacto en empresas	94
6.2 CONCLUSIONES Y RECOMENDACIONES GENERALES.....	95
6.2.1 Conclusiones	95
6.2.2 Recomendaciones Estratégicas	97
6.3 LIMITACIONES DEL ESTUDIO.....	99
6.4 FUTURAS LÍNEAS DE INVESTIGACIÓN	100
6.5 REFLEXIÓN FINAL	101
7. BIBLIOGRAFÍA.....	102
8. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS FIN DE GRADO	109

Índice de figuras

1	Ilustración 1: Planos propuestos y seleccionados como ganadores del concurso para la remodelación del estadio.....	13
2	Ilustración 2: Evolución del número de mensajes diarios en el grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación oficial de eventos en el estadio Santiago Bernabéu. El eje X representa el calendario entre abril de 2024 y marzo de 2025. El eje Y el número de mensajes. Las líneas verticales rojas punteadas indican las fechas de conciertos, y la línea azul punteada marca el día en que se anunció la suspensión de los eventos	30
3	Ilustración 3: Distribución del Sentimiento del grupo de WhatsApp - Tras haber eliminado Stopwords. Eje Y: Número de mensajes, Eje X: Puntuación siendo 1 Muy negativo y 5 Muy positivo.....	31
4	Ilustración 4: Evolución del sentimiento promedio diario en los mensajes del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu. El eje X representa la fecha (entre abril de 2024 y marzo de 2025), mientras que el eje Y muestra el valor promedio de sentimiento diario, calculado mediante un modelo de clasificación en escala de 1 (muy negativo) a 5 (muy positivo). Las líneas verdes punteadas indican las fechas de conciertos y la línea naranja punteada señala el día en que se anunció la suspensión de estos.....	32
5	Ilustración 5: Wordcloud de bigramas más frecuentes en mensajes de alto malestar (sent_corr = 1) entre noviembre 2024 y febrero 2025.....	33
6	Ilustración 6: Evolución diaria del número de mensajes positivos y negativos en el grupo de WhatsApp vecinal. En el eje X se representan las fechas (entre abril de 2024 y marzo de 2025), y en el eje Y el número de mensajes clasificados por tipo de sentimiento: positivo (puntuación corregida ≥ 4) y negativo (≤ 2).	34
7	Ilustración 7: Sentimiento promedio diario corregido (sentiment_corr) a lo largo del tiempo. El eje X indica la fecha y el eje Y el valor promedio diario del sentimiento en una escala de 1 (muy negativo) a 5 (muy positivo). Las bandas rosadas señalan los días con tono negativo (media < 3) y las líneas rojas punteadas marcan los días de concierto.	35
8	Ilustración 8: Análisis tipo event study de la proporción de mensajes negativos en los días relativos a los conciertos, considerando una ventana de ± 7 días. En el eje X se representan los días respecto al concierto (día 0) y en el eje Y el porcentaje de mensajes clasificados como quejas. Se incluye una línea roja vertical que marca el día del evento y una línea azul horizontal que representa la media general.	36
9	Ilustración 9: Distribución global de las emociones según el léxico NRC aplicado al total de mensajes de la muestra.....	38

10	Ilustración 10: Distribución global de las emociones según el léxico NRC aplicado a los mensajes enviados en días de concierto	39
11	Ilustración 11: Distribución global de las emociones según el léxico NRC aplicado a los mensajes enviados fuera de los días de concierto.....	39
12	Ilustración 12: Wordcloud de las palabras mas utilizadas habiendo eliminado stopwords	40
13	Ilustración 13: Nube de palabras generada a partir de los mensajes con sentimiento positivo en el grupo de WhatsApp vecinal, tras eliminar stopwords.	41
14	Ilustración 14: Nube de palabras de los mensajes clasificados con sentimiento negativo, sin stopwords.	42
15	Ilustración 15: Nube de palabras específica para el día 13 de septiembre de 2024, fecha en la que se anunció la suspensión de los conciertos en el Santiago Bernabéu.	43
16	Ilustración 16: Nube de palabras que representa las palabras más representativas de los mensajes clasificados como queja (variable es_queja).....	44
17	Ilustración 17: Evolución diaria del número de mensajes en el grupo de WhatsApp vecinal que contienen menciones a niveles de ruido superiores a los límites legales de decibelios. El eje X representa la fecha (desde abril hasta septiembre de 2024) y el eje Y indica el número de mensajes diarios.....	46
18	Ilustración 18: Evolución diaria del total de mensajes enviados en el grupo de WhatsApp vecinal (línea azul) y del número de mensajes que reportan niveles de ruido superiores al umbral legal de decibelios (línea naranja), entre abril de 2024 y marzo de 2025. El eje X representa la fecha y el eje Y indica la cantidad diaria de mensajes	47
19	Ilustración 19: Evolución del sentimiento promedio diario en los mensajes del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu. El eje X representa la fecha (entre abril de 2024 y marzo de 2025), mientras que el eje Y (lado izquierdo) muestra el valor promedio de sentimiento diario en escala de 1 (muy negativo) a 5 (muy positivo). Las líneas grises indican las fechas de conciertos. En el eje Y se muestra el número de quejas junto con la distribución diaria que se representa en la gráfica roja.....	50
20	Ilustración 20: Evolución del número de quejas por día del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu (rayas rojas). El eje X representa la fecha (entre abril de 2024 y marzo de 2025), y el eje Y el número de fechas. La raya gris punteada refleja la media de quejas general	50
21	Ilustración 21: Gráfico de barras que representa la distribución total de mensajes como queja o no queja según si cumple que el mensaje contenga una palabra del diccionario de queja (creado manualmente) o que con el uso de BERT tenga una puntuación igual o menor a 2	51
22	Ilustración 22: Gráfica de barras que muestra la evolución de quejas en porcentaje alrededor de los días de concierto. En el eje Y, el porcentaje de fechas y en el eje X los días adyacentes al	

	concierto, siendo los números negativos los días anteriores y los positivos, los posteriores. El día 0 es el día del concierto.....	52
23	Ilustración 23: Varianza explicada acumulada obtenida al aplicar SVD a la matriz TF-IDF de los mensajes vecinales. El eje horizontal muestra el número de componentes retenidos y el eje vertical la proporción de varianza total explicada. Las líneas punteadas gris y verde señalan los umbrales del 80 % y el 90 % de varianza explicada, respectivamente.	53
24	Ilustración 24: Visualización bidimensional de los clústeres obtenidos con KMeans tras proyectar la matriz TF-IDF reducida mediante PCA. Cada punto representa un documento y está sin colorear para captar algún patrón de manera visual. El eje horizontal (Componente principal 1) y el eje vertical (Componente principal 2) corresponden a las dos primeras componentes principales, que condensan la mayor parte de la varianza latente del texto.	55
25	Ilustración 25: Visualización bidimensional de los clústeres obtenidos con KMeans ($k = 2$) tras proyectar la matriz TF-IDF reducida mediante PCA. Cada punto representa un documento y está coloreado según su asignación de clúster (0–4). El eje horizontal (Componente principal 1) y el eje vertical (Componente principal 2) corresponden a las dos primeras componentes principales, que condensan la mayor parte de la varianza latente del texto	57
26	Ilustración 26: Resultado de la regresión logística binaria aplicada a los mensajes clasificados como queja	61
27	Ilustración 27: Modelo de regresión logística binaria multivariable para los mensajes clasificados como queja	62
28	Ilustración 28: Resultados del modelo Difference-in-Differences (DiD) que compara el número de quejas diarias tras los conciertos frente a días sin eventos. La tabla de regresión OLS muestra en la primera columna las variables incluidas (“post”, “tratada” e interacción), junto con sus coeficientes estimados, errores estándar, valores t y p-valores.	64
29	Ilustración 29: Histograma agrupado de quejas diarias en la ventana de cinco días (± 2 días) alrededor de cada concierto. El día del evento (día 0) aparece resaltado en rojo oscuro, mientras que los días previos y posteriores se muestran en tonos más claros.....	65
30	Ilustración 30: Resultados del modelo DiD que contrasta días de concierto con días de partido de fútbol en casa del Real Madrid. La regresión OLS presenta los coeficientes para “post”, “tratada” e interacción DiD	66
31	Ilustración 31: salida de la función rdrobust ejecutada en Python con ancho de banda manual para el corte correspondiente al 10 de mayo. Se incluyen 49 observaciones en el lado izquierdo y 50 en el lado derecho del umbral.....	70
32	Ilustración 32: Gráfico RDD Sharp, el eje X muestra los días relativos al 10-mayo abarcando desde -49 hasta +49 días. El eje Y indica el número diario de quejas. Cada punto azul corresponde al total de quejas reportadas en un día concreto. La línea vertical negra en $X = 0$ marca la fecha de inicio de conciertos (10 de mayo).....	71

33	Ilustración 33: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 10 de mayo y con p libre.....	71
34	Ilustración 34: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 13 de septiembre. Se incluyen 49 observaciones en el lado izquierdo y 50 en el lado derecho del umbral	72
35	Ilustración 35: Gráfico RDD Sharp, el eje X muestra los días relativos al 13-septiembre abarcando desde -49 hasta +49 días. El eje Y indica el número diario de quejas. Cada punto azul corresponde al total de quejas reportadas en un día concreto. La línea vertical negra en $X = 0$ marca la fecha de inicio de conciertos concierto (10 de mayo)	73
36	Ilustración 36: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 13 septiembre y con p libre.....	73
37	Ilustración 37: Serie temporal mensual desde 2021 hasta 2024 para cuatro indicadores: gasto total (línea azul), gasto medio por persona (naranja), gasto medio diario por persona (verde) y duración media del viaje (rojo). El eje horizontal muestra la fecha y el eje vertical el importe en euros. ...	77
38	Ilustración 38: Serie mensual de 2022–2024 que compara la evolución de tres variables en Madrid frente al conjunto nacional: número de turistas (línea azul), pernoctaciones (naranja) y duración media de la estancia (verde). El eje horizontal muestra la fecha y el eje vertical el importe en euros.	77
39	Ilustración 39: Serie temporal de “Turismo Móvil – Total” (INE). Sobre el eje horizontal se muestran los meses desde enero de 2022 hasta finales de 2024, y en el eje vertical los tres indicadores principales: número de turistas (línea azul), pernoctaciones (línea naranja) y duración media de la estancia (línea verde). Los datos equivalen al turismo total nacional	78
40	Ilustración 40: Serie temporal de “Turismo Móvil – Madrid” (INE). Sobre el eje horizontal se muestran los meses desde enero de 2022 hasta finales de 2024, y en el eje vertical los tres indicadores principales: número de turistas (línea azul), pernoctaciones (línea naranja) y duración media de la estancia (línea verde) de los datos exclusivamente de Madrid.....	79
41	Ilustración 41: Evolución mensual de las llegadas de turistas internacionales a Madrid por motivo de ocio (barra azul) y turismo total (barra naranja). El eje horizontal representa los meses y el vertical el número de llegadas.....	82
42	Ilustración 42: Gráfica mensual de enero 2022 a 2024 para cuatro métricas: número de establecimientos abiertos, grado de ocupación por plazas, media y fines de semana, y grado de ocupación de habitaciones. El eje horizontal muestra los meses y el vertical el porcentaje o número absoluto.	84
43	Ilustración 43: Serie temporal para el total nacional desde enero de 2022 hasta diciembre de 2024 mostrando, en el eje horizontal, el mes y, en el eje vertical, el número total de pernoctaciones (línea naranja) junto al número de viajes registrados por móvil (línea azul).....	85

44	Ilustración 44: Serie temporal para Madrid desde enero de 2022 hasta diciembre de 2024 mostrando, en el eje horizontal, el mes y, en el eje vertical, el número total de pernoctaciones (línea naranja) junto al número de viajes registrados por móvil (línea azul).....	85
45	Ilustración 45: Evolución mensual de los viajeros (línea azul) y las pernoctaciones (línea naranja) desglosadas por residentes nacionales (líneas verde y rojo) y visitantes internacionales (líneas morada y gris), desde enero de 2022 hasta finales de 2024. El eje horizontal marca las fechas y el vertical los volúmenes.....	87
46	Ilustración 46: Serie temporal mensual de la tasa de ocupación (%) para cinco categorías: hoteles (azul), campings (naranja), apartamentos turísticos (verde), alojamientos rurales (rojo) y albergues (morado). El eje horizontal indica el mes y el izquierdo el porcentaje de ocupación.	88
47	Ilustración 47: Contraste de medias (Welch) entre mayo–septiembre 2023 y 2024, media mensual 2023 vs. 2024, estadístico t y p-valor para Madrid y España.	89

Índice de tablas

1	Tabla 1: valores diarios de sentimiento promedio corregido y porcentaje de quejas para cada día relativo a los conciertos. El eje de análisis temporal está definido en función de la distancia al evento (de -7 a +7 días), permitiendo detectar los días con mayor carga emocional negativa.....	37
2	Tabla 2: Resultados de las pruebas de estadísticas que confirman un aumento significativo de quejas en días de concierto ($p < 0,001$) y cuantifican el tamaño del efecto.	49
3	Tabla 3: Tabla resumen de coeficientes DiD obtenidos al comparar cada concierto con un día normal aleatorio. Cada fila incluye la fecha y el nombre del concierto junto al valor del coeficiente DiD, su p-valor y el R^2 del modelo	65
4	Tabla 4: Tabla resumen de coeficientes DiD al contrastar cada concierto con un día de partido de fútbol en casa.	67
5	Tabla 5: Definición de variables en el modelo de Difference-in-differences	75
6	Tabla 6: Resultados del modelo DiD sobre el gasto turístico.....	76
7	Tabla 7: Resultados del modelo DiD sobre el turismo móvil.....	78
8	Tabla 8: Resultados del modelo DiD sobre los movimientos turísticos	82
9	Tabla 9: Resultados del modelo DiD sobre el ocupación hotelera	83
10	Tabla 10: Resultados del modelo DiD sobre las pernoctaciones.....	85
11	Tabla 11: Resultados del modelo DiD sobre la evolución mensual de los viajeros y las pernoctaciones por tipo de alojamiento	86
12	Tabla 12: Resultados del modelo DiD sobre el ocupación por tipo de alojamiento.....	88

1. INTRODUCCIÓN

1.1 Contexto del Estadio Santiago Bernabéu

El Estadio Santiago Bernabéu, ubicado en el corazón de Madrid, es uno de los recintos deportivos más icónicos del mundo. Desde su inauguración el 14 de diciembre de 1947, ha sido la sede del Real Madrid Club de Fútbol. Con el tiempo, el Bernabéu ha trascendido su papel como estadio de fútbol para convertirse en un epicentro cultural y económico de la ciudad, acogiendo eventos de diversa índole, incluidos conciertos y encuentros internacionales.

A lo largo de los años, el estadio ha experimentado múltiples transformaciones, adaptándose a las nuevas demandas del deporte, el entretenimiento y el desarrollo urbanístico de Madrid. Si bien su función principal siempre ha sido albergar partidos de fútbol, en los últimos años ha evolucionado estratégicamente para convertirse en un espacio multifuncional. Esta tendencia responde a un fenómeno global en el que los grandes estadios han diversificado su uso, organizando conciertos, eventos corporativos y espectáculos de gran escala.

Este cambio de enfoque también obedece a estrategias económicas y de negocio, ya que el Bernabéu es uno de los activos inmobiliarios más valiosos de Madrid (Ordeig, 2020). Para garantizar su rentabilidad, resulta imprescindible generar ingresos adicionales más allá de los obtenidos en los días de partido. La organización de eventos de gran magnitud no solo maximiza la rentabilidad del estadio, sino que también impulsa la economía local, fomentando el turismo y consolidando a Madrid como un destino clave para el entretenimiento. Sin embargo, esta transformación también plantea desafíos para la comunidad local, la infraestructura urbana, los negocios de la zona y las políticas públicas, lo que requiere un equilibrio entre el desarrollo económico y el bienestar ciudadano.

1.1.1 Transformación y obras del Bernabéu (2024)

El año 2024 marcó un punto de inflexión en la historia del Estadio Santiago Bernabéu con la finalización de un ambicioso proyecto de renovación. Esta remodelación representó la mayor transformación del estadio desde 1982, cuando se modernizó con motivo del Mundial celebrado en España, cuya final se disputó en el Bernabéu.

Las obras comenzaron en 2019, tras un concurso convocado en 2014 para seleccionar el mejor diseño entre cuatro propuestas arquitectónicas. Finalmente, el proyecto ganador fue desarrollado por un consorcio de tres estudios de arquitectura e ingeniería:

- GMP Architekten (Alemania): Firma de arquitectura con amplia experiencia en el diseño de estadios y grandes infraestructuras deportivas.
- L35 Arquitectos (España): Estudio español especializado en proyectos de centros comerciales, oficinas y espacios deportivos.
- Ribas & Ribas Arquitectos (España): Despacho español con experiencia en proyectos de gran envergadura.

El principal objetivo de esta renovación era modernizar el estadio y mejorar su versatilidad, posicionándolo como un icono arquitectónico global y un espacio más funcional tanto para los aficionados como para la ciudad de Madrid.

Fuente: Arquitectura Viva. (2024, septiembre 23).

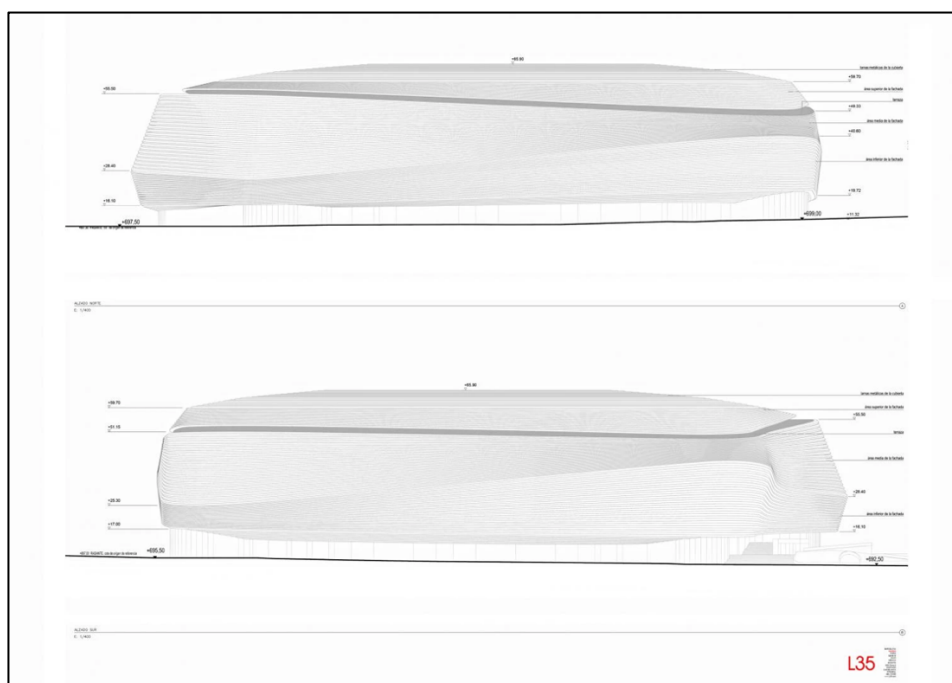


Ilustración 1: Planos propuestos y seleccionados como ganadores del concurso para la remodelación del estadio

El nuevo Santiago Bernabéu destaca por su diseño vanguardista y mejoras estructurales que lo convierten en un referente mundial en infraestructuras deportivas y de entretenimiento. Entre los cambios más significativos se encuentran:

- Fachada envolvente de acero inoxidable: La nueva piel metálica recubre todo el estadio, permitiendo efectos de iluminación y cambios visuales dinámicos a lo largo del día.
- Cubierta retráctil: Un sistema que permite abrir o cerrar el techo del estadio según las condiciones climáticas y el tipo de evento.
- Césped retráctil con invernadero subterráneo: Un mecanismo innovador que permite retirar y preservar el césped cuando el estadio se usa para conciertos u otros eventos.
- Gestión digital y sostenibilidad: El estadio incorpora sistemas inteligentes para optimizar el consumo energético y mejorar la seguridad. Además, se han implementado fuentes de energía renovable y un sistema de recolección de agua para reducir el impacto ambiental.
- Aislamiento acústico: Dado que una de las metas de la renovación era albergar eventos culturales y espectáculos con mayor frecuencia, se incorporaron tecnologías para reducir la contaminación acústica y minimizar el impacto del ruido en las zonas residenciales cercanas.

Todos estos cambios se debían aplicar sobre un estadio integrado en el corazón de la ciudad sin comprometer la arquitectura, pero integrando la modernidad (Laguardia, 2024).

La modernización del Bernabéu ha convertido el estadio en un centro de ocio capaz de competir con los recintos más prestigiosos del mundo. Además de fortalecer su papel como un activo estratégico para el desarrollo económico de Madrid, su renovada infraestructura ha potenciado el turismo y la oferta de entretenimiento en la ciudad.

Sin embargo, esta creciente comercialización del estadio ha generado inquietud entre los residentes del barrio de Chamartín. La mayor afluencia de público en eventos no deportivos ha intensificado problemas como la congestión urbana, el ruido y cambios en la dinámica del vecindario. Así, mientras el Bernabéu se consolida como un pilar clave en la economía de Madrid, su impacto en la comunidad local sigue siendo un tema de debate que requerirá medidas para equilibrar el desarrollo económico con la calidad de vida de los ciudadanos.

1.1.2 Expansión a eventos no deportivos y el papel de los conciertos

A raíz de la remodelación completa del estadio, el Santiago Bernabéu se ha posicionado como un escenario ideal para la celebración de eventos no deportivos. Desde conciertos de artistas de renombre internacional como Taylor Swift o Karol G hasta eventos mundialmente reconocidos, como la velada del año V, organizada por el YouTuber Ibai Llanos.

Asimismo, el estadio ha sido seleccionado para albergar eventos deportivos de la NFL, programados para 2025, marcando un hito en la expansión del fútbol americano en España (Real Madrid, 2025).

La transformación de estadios en recintos multifuncionales se ha convertido en una estrategia clave para potenciar tanto el uso de las infraestructuras deportivas como el dinamismo cultural y económico de las ciudades. En este sentido, instalaciones icónicas como el Bernabéu, el Madison Square Garden (Nueva York) y el SoFi Stadium (Los Ángeles) han pasado de ser meramente escenarios de competiciones deportivas a convertirse en auténticos centros de eventos, donde conviven conciertos, mítines políticos y otras actividades culturales.

Con esta evolución, el Santiago Bernabéu no solo redefine su identidad como un estadio de fútbol, sino que también se suma a la élite de los recintos multifuncionales más prestigiosos del mundo. Asimismo, esta estrategia permite que el club diversifique sus fuentes de ingresos a la vez que enriquece la oferta cultural y de entretenimiento de la capital española.

1.1.3 Impacto económico y social

La transformación del Estadio Santiago Bernabéu en un espacio multifuncional ha tenido un impacto notable en la economía de Madrid. Según datos del Real Madrid, durante la temporada 2023/24, el club superó los 1.000 millones de euros en ingresos de explotación, alcanzando los 1.073,2 millones, lo que representa un incremento del 27% en comparación con la temporada anterior. Los conciertos consolidan una ganancia de 12 Millones de euros, lo que equivale a entorno 1% de las ganancias del equipo (Real Madrid, 2024).

Sin embargo, este crecimiento económico ha generado desafíos sociales, especialmente en relación con la convivencia vecinal. Los aumentos en el ruido debido a eventos y obras, junto con la congestión en días de partido y concierto, han llevado a debates sobre cómo equilibrar

el desarrollo económico con la tranquilidad de los residentes. De hecho, se han registrado quejas y acciones legales por molestias ocasionadas por el ruido de los conciertos, lo que ha llevado al club a suspender temporalmente estos eventos hasta implementar medidas de insonorización.

Este Trabajo de Fin de Grado analizará detalladamente tanto los beneficios económicos derivados de la transformación del Bernabéu como los retos sociales que han surgido, con el objetivo de proponer soluciones que permitan una convivencia armoniosa entre el desarrollo económico y el bienestar de los vecinos.

1.2 Motivación personal

La elección de este Trabajo de Fin de Grado surge de una vivencia personal directa y cotidiana. Como vecina del estadio Santiago Bernabéu, he experimentado en primera persona los efectos derivados de su transformación en un recinto multiusos y, especialmente, la intensificación del uso del estadio como sede de grandes conciertos desde la primavera de 2024. Estos eventos, aunque beneficiosos para el dinamismo de la ciudad, también han supuesto un cambio radical en la vida del barrio: ruido nocturno elevado, colapso del transporte y calles. Esta convivencia forzada con los eventos me llevó a preguntarme por su verdadero impacto, no solo desde una perspectiva subjetiva, sino también desde un enfoque analítico y riguroso.

A este componente personal se suma una motivación académica: el deseo de aplicar técnicas de análisis de datos reales a un problema tangible. El acceso a un grupo de WhatsApp exclusivo de vecinos afectados me permitió trabajar con una fuente de información privilegiada y auténtica, donde se expresan opiniones y emociones en tiempo real. Además, este proyecto me brindaba la oportunidad de explorar una de las áreas que más me ha interesado durante la carrera: el análisis de sentimiento y el procesamiento del lenguaje natural (PLN). La posibilidad de combinar la investigación social con técnicas avanzadas de machine learning y modelos estadísticos me resultó no solo desafiante, sino también profundamente estimulante y divertido.

Este trabajo representa así la confluencia entre mi contexto personal, mis intereses analíticos y mi compromiso con el rigor metodológico. Mi objetivo ha sido doble: por un lado, visibilizar el malestar legítimo de los residentes, y por otro, analizar de forma objetiva los beneficios que estos eventos pueden tener para la ciudad en términos turísticos y económicos.

1.3 Relevancia académica y social

El presente estudio aporta una contribución significativa tanto en el plano académico como en el social. Desde el punto de vista académico, se inscribe en un campo de creciente interés: la evaluación del impacto de los eventos masivos en entornos urbanos desde una perspectiva multidimensional y basada en datos. El enfoque adoptado combina distintas metodologías: análisis de sentimiento, modelos causales como Difference-in-Differences, Regression Discontinuity Design y clustering para ofrecer una visión integral del fenómeno. Esta aproximación permite, además, aplicar técnicas avanzadas de análisis de texto sobre datos no estructurados, como los mensajes de WhatsApp, demostrando cómo el lenguaje cotidiano puede convertirse en una fuente valiosa de evidencia empírica.

En términos sociales, este trabajo se sitúa en un contexto especialmente sensible: el equilibrio entre el desarrollo económico que generan los grandes eventos y el derecho al bienestar de las comunidades que los albergan. La transformación del estadio Santiago Bernabéu ha sido celebrada como una oportunidad para posicionar a Madrid como capital del entretenimiento europeo, pero también ha traído consigo tensiones no resueltas en el tejido urbano y social inmediato. Este trabajo intenta ofrecer una base informada para repensar esa convivencia, incorporando la voz de los vecinos y midiendo de forma objetiva los efectos económicos asociados.

En definitiva, se trata de una investigación que no solo analiza datos, sino que intenta aportar valor al debate público. Su propósito es contribuir a la formulación de políticas urbanas más inclusivas y equilibradas, que reconozcan la diversidad de intereses en juego y busquen soluciones sostenibles que integren la dimensión cultural, económica y humana del fenómeno.

1.3 Fundamento conceptual y contexto normativo

1.3.1 Eventos masivos como motores económicos y culturales

La conversión del estadio Santiago Bernabéu en un recinto multifuncional se enmarca en una tendencia global de reconversión de grandes infraestructuras deportivas en activos urbanos rentables y versátiles. Esta transformación permite aprovechar estos espacios no solo para la

celebración de competiciones deportivas, sino también para albergar conciertos, ferias, espectáculos internacionales y eventos empresariales. Según Smith (2010), la creación de “Sports-City Zones” contribuye a diversificar las funciones del entorno urbano, estimular el turismo y reforzar la identidad cultural de las ciudades.

En este contexto, el presente trabajo parte de un marco conceptual que concibe los eventos masivos no solo como herramientas de desarrollo económico, sino también como fenómenos que reconfiguran la vida cotidiana, el uso del espacio público y la convivencia social en las comunidades donde se celebran.

1.3.2 Impacto directo, indirecto sobre la economía local

Algunos estudios académicos han documentado los beneficios económicos que los eventos masivos pueden generar. Qian y Tuck (2023) identifican tres tipos de efectos: los directos, derivados del gasto inmediato en alojamiento, transporte, alimentación y entradas; los indirectos, asociados al aumento de actividad en la cadena de suministro; y los inducidos, que provienen del consumo adicional generado por los ingresos obtenidos por quienes trabajan directa o indirectamente en los eventos.

Cai (2024) cuantifica este impacto estimando que, por cada 100 dólares gastados en un concierto, se generan hasta 300 dólares en consumo adicional en la economía local. Un caso ilustrativo es el concierto de Taylor Swift en Filadelfia (2023), que provocó una ocupación hotelera del 95% y niveles de demanda que no se registraban desde antes de la pandemia.

A nivel nacional, el informe de OBS Business School (Mir, 2024) destaca que el turismo musical generó 13.854 millones de euros en España en 2023. Eventos como Mad Cool (80 millones de euros) o Arenal Sound (42 millones de euros) ejemplifican este auge. Madrid, en particular, ha reforzado su posición como nodo estratégico de turismo de eventos, consolidando al Bernabéu como uno de sus principales activos.

1.3.3 El vínculo entre conciertos y turismo en Madrid

La relación entre conciertos y turismo es especialmente relevante para la ciudad de Madrid. El informe *Amadeus Travel Trends 2024* señala que el 49% de los españoles ha viajado al menos una vez para asistir a un concierto, y que el 79% de los turistas aprovecha estos eventos

para alargar su estancia, lo que impulsa sectores como la hostelería, la restauración y el comercio.

Según datos del INE (2024), el gasto turístico total en España alcanzó los 126.282 millones de euros, con un crecimiento interanual del 16,1%. El gasto medio por turista fue de 1.441 euros y el gasto medio diario alcanzó los 159 euros, con una duración media de 6,4 días. Además, los eventos representaron el 23,8% del gasto total turístico, con incrementos del 20,2% en transporte y 14,3% en alojamiento.

En Madrid, el gasto medio diario por turista internacional fue de 293 euros y la ciudad acogió aproximadamente 600.000 visitantes anuales motivados por conciertos, lo que representa el 6% del total de turistas (L.I., Agencia EFE, 2023). El estadio Santiago Bernabéu, tras su remodelación, se ha integrado plenamente en esta estrategia, proyectándose como un espacio clave para el turismo urbano de alto valor.

1.3.4 El ruido como eje del conflicto urbano

Pese a sus múltiples beneficios, la celebración frecuente de eventos masivos también genera externalidades negativas, especialmente en entornos urbanos residenciales. La literatura académica ha señalado que el tráfico, el deterioro del espacio público, la saturación del transporte y la pérdida de calidad de vida son consecuencias frecuentes en barrios afectados por estas dinámicas (Fernández et al., 2019). Estos efectos, aunque menos visibles que los económicos, pueden provocar tensiones sociales persistentes, desgaste vecinal y conflicto entre usos del espacio urbano.

Entre las molestias más señaladas por los vecinos, el ruido ocupa una posición central. A pesar de las medidas implementadas durante la remodelación del estadio, las mediciones acústicas realizadas por la Policía Municipal han llegado a registrar niveles de hasta 90 decibelios durante los conciertos, una cifra que supera ampliamente los límites legales establecidos por la Ordenanza de Protección contra la Contaminación Acústica y Térmica del Ayuntamiento de Madrid (2021), que fija un umbral de 55–65 dBA hasta las 23:00 horas y 45–55 dBA en horario nocturno.

Esta situación ha motivado la apertura de 24 expedientes sancionadores por parte del Ayuntamiento contra promotores de conciertos en el Bernabéu, con un importe acumulado de

2,6 millones de euros (EFE, 2024). Ante la creciente presión vecinal, el club se vio obligado a suspender temporalmente los eventos musicales programados a partir de septiembre de 2024 (Hormigo, 2024a).

Además, los residentes interpusieron una querrela judicial por posible delito medioambiental, que fue admitida a trámite por un juzgado, lo que refleja la creciente dimensión legal del conflicto y la demanda de medidas más contundentes (Peiró, 2024). En este contexto, el Ayuntamiento de Madrid, con el respaldo del alcalde José Luis Martínez-Almeida, ha impulsado nuevas restricciones, entre ellas:

- Límite de horario de conciertos hasta las 23:00 horas.
- Reducción del volumen permitido en los sistemas de sonido.
- Mayor control sobre la venta ambulante y aglomeraciones para evitar saturación del espacio público.
- Inicio de un proyecto específico de insonorización del estadio (Rodríguez, 2024).

1.3.5 Complejidad técnica y desafíos de insonorización

La solución acústica no es sencilla. La arquitectura del estadio, con estructura abierta, materiales metálicos y una fachada envolvente reflectante, dificulta la contención del sonido. Por ello, el club contrató a una empresa británica especializada, responsable del diseño acústico de La Sphera de Las Vegas, para desarrollar un sistema de insonorización avanzado con un plazo estimado de seis meses (Rodríguez, 2024).

Entre las medidas ya adoptadas destacan el uso de la cubierta retráctil como escudo acústico, la redirección del sonido hacia el interior y el refuerzo de barreras fónicas (Hormigo, 2024b). A pesar de ello, las quejas han continuado y el conflicto sigue abierto.

1.3.6 Un conflicto urbano entre rentabilidad y habitabilidad

La situación del Bernabéu ilustra con claridad el dilema urbano contemporáneo entre la explotación intensiva de los activos de ocio y el derecho de los ciudadanos a una convivencia tranquila. En este caso, el estadio representa a la vez un motor económico clave para la ciudad y una fuente creciente de conflicto local. Esta tensión evidencia la necesidad de políticas

públicas más equilibradas, que permitan armonizar el uso lucrativo del espacio urbano con los derechos fundamentales de sus habitantes.

Desde este punto de vista, el presente trabajo no solo aporta un análisis técnico y cuantitativo del fenómeno, sino que también busca contribuir al debate sobre qué tipo de crecimiento es sostenible en las grandes ciudades y en qué condiciones pueden convivir el espectáculo y el descanso, la rentabilidad y la habitabilidad.

1.3.7 Valor añadido del enfoque metodológico

Para abordar esta problemática de forma rigurosa, el estudio se apoya en una combinación metodológica avanzada que incluye:

- Análisis de sentimiento y emociones con PLN.
- Extracción de números (como decibelios o número de quejas) en mensajes vecinales.
- Modelos como Difference-in-Differences, Regression Discontinuity Design, test de hipótesis y análisis temático no supervisado (TF-IDF + clustering).

La riqueza de fuentes, mensajes reales de WhatsApp, datos del INE, normativa municipal y literatura académica, permite ofrecer una visión integral, capaz de conectar percepciones subjetivas, evidencia estadística y contexto legal en un solo marco analítico.

1.4 Objetivo del estudio

El objetivo principal de este Trabajo de Fin de Grado es analizar de manera integral y empíricamente fundamentada el impacto que tienen los conciertos celebrados en el Estadio Santiago Bernabéu sobre tres dimensiones clave: la calidad de vida de los vecinos del barrio de Chamartín, la evolución del turismo en la ciudad de Madrid y el efecto económico sobre las empresas locales situadas en su entorno. Esta investigación busca evaluar no solo los beneficios derivados de la transformación del estadio en un recinto multifuncional, sino también los costes sociales y tensiones urbanas que dicha conversión ha generado en un contexto residencial denso y previamente tranquilo.

Para alcanzar este objetivo general, se han definido tres líneas de análisis interrelacionadas, cada una con metodologías específicas y fuentes de datos robustas que permiten realizar una evaluación objetiva, multidimensional y contrastable:

1.4.1 Impacto sobre los vecinos

Se pretende evaluar cómo afecta la celebración recurrente de conciertos en el Bernabéu a la vida diaria de los residentes del barrio de Chamartín, especialmente en términos de molestias acústicas, congestión del tráfico, alteración del espacio público y cambios en la percepción de bienestar. Para ello, se utilizará como fuente principal un corpus de más de 3.000 mensajes recopilados de un grupo de WhatsApp exclusivo de vecinos, analizado mediante técnicas de procesamiento de lenguaje natural (PLN) y aprendizaje automático.

Los objetivos específicos dentro de esta línea son:

- Aplicar modelos de análisis de sentimiento y emociones (utilizando BERT, léxicos y herramientas como PySentimiento) para capturar las percepciones emocionales de los residentes en torno a los eventos.
- Detectar patrones temporales en la intensidad emocional y el contenido temático de los mensajes mediante visualizaciones descriptivas, análisis léxico, wordclouds y extracción de numeros como niveles de decibelios, quejas o referencias al ruido.
- Clasificar los mensajes en tipos de queja y niveles de malestar mediante clustering temático (TF-IDF + KMeans) y análisis manual reforzado por listas semánticas de palabras clave.
- Evaluar estadísticamente si existe una relación causal entre la celebración de conciertos y el incremento en el volumen o intensidad de las quejas vecinales, utilizando modelos Difference-in-Differences, regresiones y análisis antes-después.
- Incorporar variables contextuales como el artista, el tipo de evento (concierto o cancelación), la longitud del mensaje, y las menciones numéricas (como aproximación a niveles de ruido).

1.4.2 Impacto en el turismo

La segunda línea de análisis busca cuantificar el impacto que los conciertos han tenido en la actividad turística de Madrid, especialmente en indicadores como el número de visitantes, el

gasto total, el gasto medio diario, la ocupación hotelera y las pernoctaciones. El enfoque parte de la hipótesis de que estos eventos musicales atraen turismo nacional e internacional y generan picos de actividad económica que pueden ser detectados en los registros oficiales.

Los objetivos específicos son:

- Aplicar modelos econométricos tipo Difference-in-Differences sobre series mensuales del Instituto Nacional de Estadística (INE), comparando Madrid (grupo tratado) con el resto de España (grupo control) antes y después de cada concierto.
- Detectar aumentos estadísticamente significativos en el número de turistas, gasto turístico y ocupación hotelera, evaluando si estos picos coinciden con los días o semanas en que se celebran conciertos en el estadio.
- Validar la robustez de los resultados mediante test de hipótesis sobre las medias.

1.4.3 Impacto en las empresas locales

La tercera dimensión del estudio analiza cómo se ven afectadas las empresas y comercios situados en las proximidades del estadio, especialmente aquellos vinculados a la restauración, la hostelería, el transporte urbano y el comercio minorista. La hipótesis plantea que los conciertos generan efectos económicos positivos directos (por aumento de demanda) e indirectos (por efectos en la cadena de suministro), pero también pueden provocar desplazamientos de actividad o saturación en ciertos sectores.

Los objetivos específicos en esta área son:

- Estimar el efecto agregado de los conciertos sobre el dinamismo económico de Madrid utilizando indicadores como gasto turístico, ocupación hotelera, empleo temporal y número de pernoctaciones, con especial atención a su concentración geográfica en el distrito de Chamartín.
- Proponer estrategias que permitan maximizar los beneficios para el tejido empresarial local, minimizando al mismo tiempo posibles externalidades negativas como la gentrificación comercial o el desplazamiento de negocios tradicionales.

1.6 Antecedentes

La transformación de estadios deportivos en espacios multifuncionales ha sido un tema recurrente en diversos estudios académicos. El TFG de Jaime Basterrechea Garrigues (2020) analiza cómo los estadios en EE. UU. han dejado de ser únicamente recintos deportivos para convertirse en centros de entretenimiento, generando ingresos a través de conciertos y eventos corporativos, lo que también impulsa la economía local. En línea con esto, el TFG de Sergio Anaya Eisman (2024) se enfoca en la remodelación del Estadio Santiago Bernabéu, destacando cómo el Real Madrid ha utilizado esta transformación para consolidar su marca global y diversificar sus fuentes de ingresos, al incorporar conciertos y otros eventos no deportivos. Además, el TFG de Ana Morales Magnet (2023) evalúa el impacto económico de la remodelación, resaltando cómo el club ha logrado rentabilizar su inversión, convirtiendo el estadio en un generador de actividad durante todo el año, con un aumento significativo en los ingresos gracias a eventos como conciertos y ferias.

Asimismo, se presentan diversos estudios académicos relevantes que abordan técnicas similares utilizadas en este trabajo y destacan los beneficios de utilizar eventos culturales como herramienta para generar efectos directos e indirectos beneficiosos en las ciudades. Gabe y Lisac (2013) analizaron detalladamente el impacto económico local generado por conciertos de música popular en Bangor, Maine. Mediante un análisis econométrico basado en regresiones con datos mensuales de ventas minoristas sujetas a impuestos, concluyeron que los asistentes a estos eventos incrementaron significativamente su gasto en la economía local. En concreto, cada asistente gastó entre 33,81 y 44,53 dólares adicionales en restaurantes y entre 18,92 y 23,12 dólares en alojamiento, reflejando un claro estímulo a sectores clave de la economía local derivados directamente de la realización de eventos musicales.

Renukadevi et al. (2023) exploraron con profundidad el análisis de sentimientos en contextos digitales utilizando métodos avanzados de inteligencia artificial y aprendizaje automático, específicamente enfocados en conversaciones grupales de WhatsApp. Este estudio no solo demostró la precisión y eficacia del análisis automatizado para capturar dinámicas comunicativas, sino que también resaltó su potencial para identificar patrones emocionales complejos, que permite entender mejor las percepciones, preocupaciones y satisfacción de las comunidades en torno a distintos fenómenos sociales.

El estudio de Fredriksson y Oliveira (2019) ofrecen un análisis acerca de la metodología de Difference-in-Differences (DiD), una técnica econométrica que permite evaluar el efecto causal de una intervención o evento comparando la evolución de dos grupos (uno tratado y uno de control) en dos períodos (antes y después de la intervención). Mediante este análisis los autores destacan su robustez y utilidad para evaluar impactos económicos y sociales en contextos donde no se dispone de experimentos controlados. El estudio enfatiza cómo este método es fundamental para asegurar la validez de inferencias causales al comparar grupos afectados directamente por un evento o intervención contra grupos de control no afectados, asegurando resultados estadísticamente fiables.

Hodur et al. (2006) aportaron evidencia empírica valiosa sobre la contribución económica de recintos o espacios destinados a eventos masivos. Utilizando datos económicos exhaustivos y detallados análisis sectoriales, concluyeron que estas instalaciones no solo generan un impacto directo inmediato, sino que también crean importantes efectos indirectos en sectores como el alojamiento, la restauración y otros servicios relacionados, contribuyendo así significativamente al desarrollo económico regional sostenible.

Zhu y Zhang (2025) y Huang (2025) refuerzan estas observaciones desde la perspectiva del marketing urbano y la economía regional, demostrando que los conciertos y eventos masivos tienen el potencial de revitalizar significativamente la imagen y la marca de las ciudades, aumentando su atractivo turístico y generando un impacto económico positivo duradero. Estos estudios también resaltan la importancia estratégica de integrar eventos culturales en las políticas urbanas para maximizar su efecto multiplicador económico y social.

En conclusión, los estudios mencionados proporcionan un fundamento sólido que respalda la relevancia y utilidad de metodologías avanzadas como DiD y análisis automatizado de sentimientos en la evaluación del impacto de eventos masivos. Estas técnicas permiten realizar análisis profundos y multifacéticos sobre los efectos económicos y sociales de eventos como conciertos y actividades deportivas, ofreciendo herramientas fundamentales para diseñar estrategias que equilibren los beneficios económicos con la sostenibilidad social y ambiental.

Sin embargo, cabe destacar que, pese a la abundancia de investigaciones centradas en los beneficios económicos, turísticos y de imagen urbana asociados a la celebración de conciertos, no se han encontrado estudios que pongan de relieve de forma sistemática los efectos negativos que estos pueden tener en el entorno urbano inmediato. En particular, no existe en la literatura un caso de estudio que aborde específicamente las consecuencias acústicas y las tensiones sociales derivadas de la celebración frecuente de macroeventos en recintos urbanos como el estadio Santiago Bernabéu. Esto evidencia un vacío académico importante en la evaluación de los costes sociales asociados a estas transformaciones urbanas, que este trabajo pretende comenzar a explorar.

1.7 Estructura del TFG

Este Trabajo de Fin de Grado se organiza en seis capítulos que abarcan tanto el análisis social como económico del impacto de los conciertos celebrados en el estadio Santiago Bernabéu desde la primavera de 2024. A lo largo del documento se combinan técnicas cuantitativas y cualitativas, datos oficiales y no estructurados, así como modelos de análisis textual, causal y predictivo, con el fin de ofrecer una visión integrada y rigurosa del fenómeno.

El trabajo se abre con un primer capítulo introductorio que contextualiza el proceso de transformación del Bernabéu, explica los motivos personales y académicos que motivan la investigación y justifica su relevancia en términos sociales, económicos y normativos. También se formulan los objetivos concretos del estudio, se revisan antecedentes relevantes en la literatura y se expone brevemente el enfoque metodológico adoptado. Esta primera parte concluye con una descripción general de la estructura del trabajo.

En el segundo capítulo se estudia el impacto que los conciertos han tenido sobre los vecinos del estadio, partiendo del análisis de miles de mensajes de un grupo de WhatsApp vecinal en el que se comentan, en tiempo real, los efectos de los eventos. Se aplican técnicas de procesamiento del lenguaje natural para medir el sentimiento, identificar emociones dominantes, detectar referencias a ruido y vibraciones y clasificar los mensajes según su tono o contenido. Además, se incorpora un análisis temporal para observar cómo varía la conversación en torno a los días de concierto, y se aplican modelos causales como Difference-

in-Differences y regresión discontinua para estimar el efecto directo de los eventos en la intensidad del malestar vecinal.

El capítulo tres se centra en evaluar el impacto sobre el turismo en la ciudad de Madrid. Para ello se utilizan diversas bases de datos del Instituto Nacional de Estadística, como EGATUR (Encuesta de Gasto Turístico), FRONTUR (Movimientos Turísticos en Fronteras), registros de turismo móvil y encuestas de ocupación. Se analiza cómo evolucionan variables como el gasto total, el gasto medio diario, el número de turistas y el volumen de pernoctaciones en los meses con conciertos, en comparación con periodos anteriores y con el resto del país. Se emplean modelos Difference-in-Differences para estimar efectos diferenciales. El análisis permite observar no solo si aumentó la afluencia de visitantes, sino también si estos tenían un perfil más intensivo en gasto, si alargaron su estancia, o si su visita estuvo directamente vinculada al ocio y los eventos culturales.

A continuación, el cuarto capítulo examina el impacto sobre las empresas locales, con especial atención al sector del alojamiento. Se estudia la evolución del número de establecimientos hoteleros abiertos, la ocupación por tipo de alojamiento, el grado de saturación por plazas y habitaciones, y el comportamiento diferencial de hoteles, apartamentos turísticos, campings, alojamientos rurales y albergues. En función de la disponibilidad de datos, se combinan modelos Difference-in-Differences con análisis antes-después para medir cambios estadísticamente significativos durante los meses de conciertos. Aunque no se dispone de microdatos empresariales sobre restauración o comercio minorista, los resultados del sector hotelero permiten inferir los efectos económicos generalizados en torno al estadio.

El capítulo cinco se dedica a las pruebas de robustez y los análisis complementarios. En él presentamos contrastes de medias (test t de Welch) que comparan los indicadores turísticos y empresariales entre mayo–septiembre de 2023 (sin conciertos) y el mismo periodo de 2024 (con conciertos). De este modo verificamos que los efectos detectados en los capítulos tres y cuatro se mantienen incluso cuando se evalúa de forma directa la significación estadística de las diferencias de media. Estas pruebas refuerzan la solidez de los resultados y descartan que los incrementos en turismo, gasto u ocupación se deban únicamente a la variabilidad aleatoria o a fluctuaciones típicas de temporada.

Finalmente, el capítulo seis recoge las principales conclusiones del estudio, destacando las implicaciones prácticas de los resultados tanto para la planificación de eventos como para las políticas públicas de convivencia, movilidad y promoción turística. También se formulan recomendaciones estratégicas para minimizar los efectos negativos percibidos por los vecinos y maximizar el retorno económico de los conciertos. El trabajo se cierra con una reflexión sobre las limitaciones del análisis y varias líneas futuras de investigación que permitirían ampliar o profundizar en la comprensión del impacto urbano de los eventos culturales.

2. IMPACTO SOBRE LOS VECINOS

2.1 Análisis de exploratorio de datos

2.1.1 Objetivo del análisis

El presente bloque del Trabajo de Fin de Grado tiene como objetivo evaluar el impacto que tienen los conciertos celebrados en el estadio Santiago Bernabéu sobre el bienestar de los vecinos de la zona, tomando como referencia la actividad registrada en un grupo vecinal de WhatsApp. A través del análisis de texto, técnicas de procesamiento de lenguaje natural (PLN), modelos estadísticos y visualizaciones descriptivas, se busca identificar patrones de queja, emociones predominantes, y variaciones en el comportamiento comunicativo en relación con las fechas de los eventos.

2.1.2 Descripción de los datos

La fuente principal de datos para este bloque consiste en un archivo de más de 29.000 mensajes extraídos de un grupo vecinal de WhatsApp. El formato original era tipo texto plano, con estructura: [fecha, hora] Nombre: mensaje. Cada mensaje se anonimiza completamente y se estructura en columnas: fecha, hora, usuario, mensaje.

Se incorpora además un calendario de conciertos del estadio, que abarca 16 fechas principales distribuidas entre abril y septiembre de 2024, así como una fecha especial (13 de septiembre) en la que se cancelaron conciertos.

2.1.3 Preprocesamiento del texto

Se implementó un pipeline de preprocesamiento en Python que incluyó los siguientes pasos:

- Anonimización de usuarios con un identificador tipo USER1, USER2, etc.
- Eliminación de frases de sistema irrelevantes (como “grupo creado”, “cambió la foto del grupo”, ”unió al grupo”).
- Normalización léxica: conversión a minúsculas y limpieza de enlaces.
- Tokenización y lematización: mediante el modelo es_core_news_sm de SpaCy y nltk para eliminar stopwords personalizadas.
- Construcción del texto limpio: se generaron columnas texto_final, tokens_lemma, mensaje_limpio y texto_final_sin_lematizar.

Además, se detectaron y extrajeron:

- Emojis y se clasificaron por enfado
- Números mencionados, usados como aproximación a referencias de decibelios (db).
- Horarios para definir ventanas nocturnas extendidas de concierto (20:00–06:00).

Con los datos limpios y normalizados, se procedió a realizar un análisis exploratorio exhaustivo de la actividad vecinal registrada en el grupo de WhatsApp. El objetivo de esta fase fue identificar patrones temporales, emocionales y semánticos que pudieran estar relacionados con los conciertos celebrados en el estadio Santiago Bernabéu.

2.1.3 Evolucion temporal de la actividad

En primer lugar, se agruparon los mensajes por fecha (fecha_dt) para obtener una serie temporal del número de mensajes diarios. Esta serie se visualizó con un gráfico de líneas (Ilustración 2) en el que se marcaron con líneas verticales las fechas de los conciertos y la fecha especial del 13 de septiembre (cancelación de evento).

Este gráfico permitió identificar varios picos notables de actividad en torno a las fechas de concierto, lo que ya sugería un posible efecto de los conciertos sobre el volumen de comunicación en el grupo.

Fuente: elaboración propia con la librería Matplotlib (y Pandas para el manejo de la serie temporal).

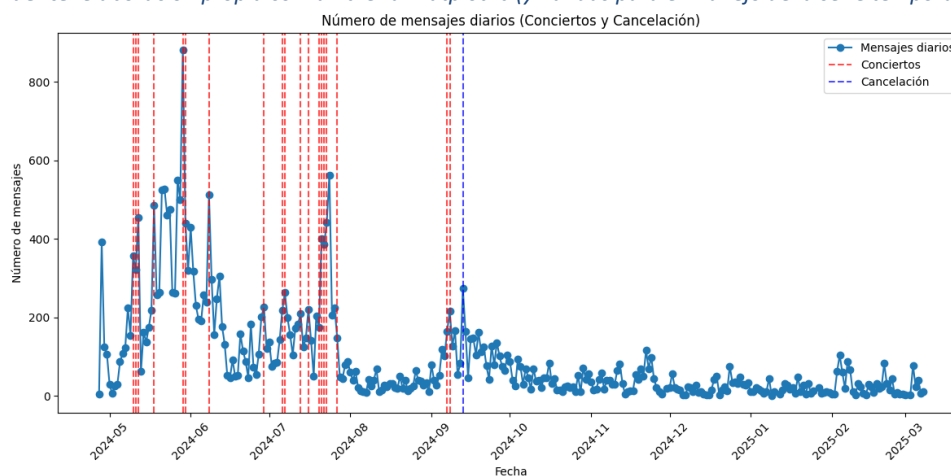


Ilustración 2: Evolución del número de mensajes diarios en el grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación oficial de eventos en el estadio Santiago Bernabéu. El eje X representa el calendario entre abril de 2024 y marzo de 2025. El eje Y el número de mensajes. Las líneas verticales rojas punteadas indican las fechas de conciertos, y la línea azul punteada marca el día en que se anunció la suspensión de los eventos

2.1.5 Sentimiento medio diario

Tras analizar la evolución del volumen de mensajes, el siguiente paso fue evaluar el tono emocional general del contenido vecinal a lo largo del tiempo. Para ello, se aplicó un modelo multilingüe basado en BERT (nlptown/bert-base-multilingual-uncased-sentiment), que clasifica textos en una escala ordinal de 1 (muy negativo) a 5 (muy positivo) (Devlin et al., 2019).

A cada mensaje se le asignó un valor (sentiment_score) según su tono, y se graficó la distribución de dichas puntuaciones para evaluar la composición emocional del grupo.

Fuente: elaboración propia con la librería Matplotlib.

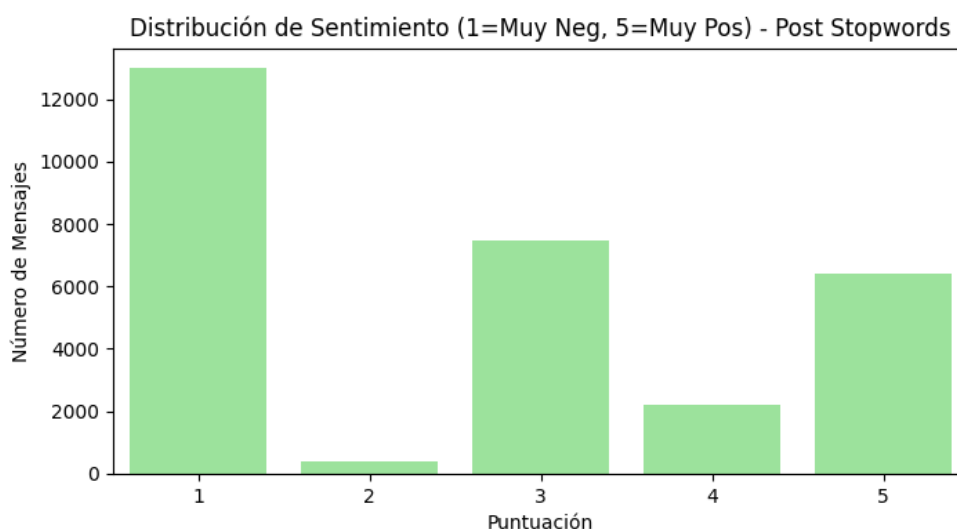


Ilustración 3: Distribución del Sentimiento del grupo de WhatsApp - Tras haber eliminado Stopwords. Eje Y: Número de mensajes, Eje X: Puntuación siendo 1 Muy negativo y 5 Muy positivo

La distribución presenta una asimetría marcada hacia los valores negativos. En concreto, la puntuación 1 (muy negativo) es la más frecuente, con más de 12.000 mensajes. Le siguen las puntuaciones 3 (neutral) y 5 (muy positivo), mientras que la puntuación 2 es poco frecuente, posiblemente debido a ambigüedades del modelo. En conjunto, más del 50 % de los mensajes se sitúan en el extremo negativo, lo que refuerza la hipótesis de un clima emocional tenso, muy probablemente vinculado a los conciertos y eventos masivos. Durante las primeras pruebas del modelo, se detectaron errores frecuentes en la clasificación automática de mensajes con negaciones, ironías o quejas implícitas. Frases como “*esto es insoportable*” o “*qué vergüenza de gestión*” no siempre eran reconocidas correctamente por el modelo.

Creación de variable *sent_corr*

Por tanto, para mejorar la calidad del análisis, se creó una nueva variable llamada *sentiment_corr*, que parte de *sentiment_score* pero incorpora una corrección semántica personalizada basada en la aparición de palabras clave relacionadas con el malestar vecinal (como *denunciar*, *ruido*, *vibración*, *policía*, *horror*, etc.). Cuando estas palabras aparecen en el mensaje, el valor de sentimiento se reajusta automáticamente a 1 (muy negativo), independientemente de lo que haya devuelto el modelo automático.

Esta variable corregida se utiliza en todos los gráficos y modelos estadísticos posteriores, ya que representa una medida más fiable y alineada con el contexto real del grupo. Una vez corregido el sentimiento por mensaje, se calculó el promedio diario de *sentiment_corr* para observar la evolución agregada del tono emocional. Este valor permite detectar tendencias, picos y caídas emocionales en torno a eventos clave como los conciertos.

Fuente: elaboración propia con la librería Matplotlib (y Pandas para la agrupación de datos por fecha)

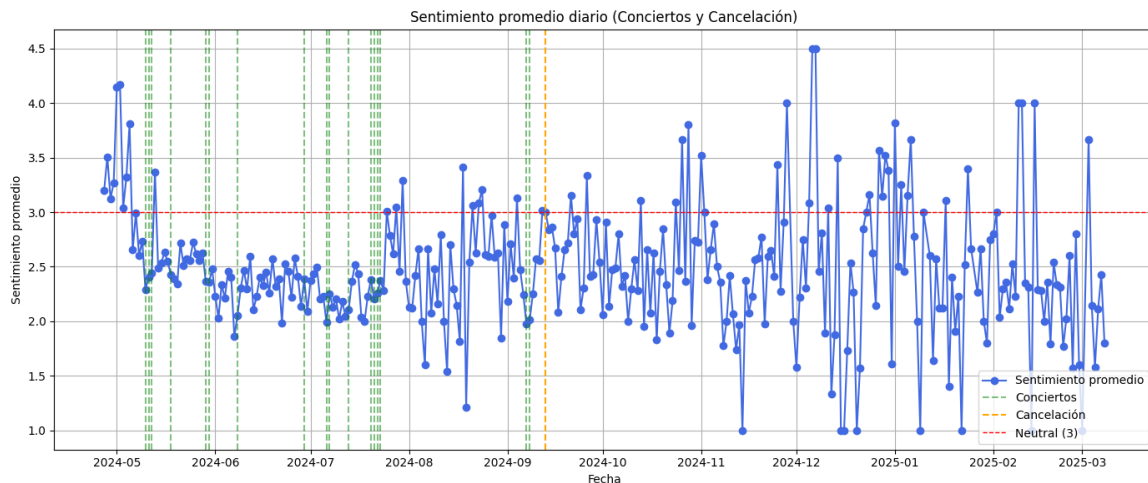


Ilustración 4: Evolución del sentimiento promedio diario en los mensajes del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu. El eje X representa la fecha (entre abril de 2024 y marzo de 2025), mientras que el eje Y muestra el valor promedio de sentimiento diario, calculado mediante un modelo de clasificación en escala de 1 (muy negativo) a 5 (muy positivo). Las líneas verdes punteadas indican las fechas de conciertos y la línea naranja punteada señala el día en que se anunció la suspensión de estos.

El sentimiento promedio diario, graficado en la figura 4, representa la media de las puntuaciones *sentiment_corr* en un día determinado. Como estas puntuaciones van de 1 (muy negativo) a 5 (muy positivo), el promedio permite ver si el grupo se inclina ese día hacia un sentimientos más negativos, neutros o positivos.

En la figura se observa una clara oscilación del sentimiento medio, con caídas puntuales próximas a los conciertos. El eje Y representa la media de *sentiment_corr* (puntuación del sentimiento) por día, lo que permite interpretar el tono global:

- Valores < 3 indican predominancia de mensajes negativos
- Valores ≈ 3 indican tono neutral
- Valores > 3 reflejan mensajes de tono positivo

Se puede asimismo observar unos picos negativos entre los meses de noviembre de 2024 a febrero de 2025, meses en los cuales no se celebraron conciertos. Para poder darle una explicación a estas cifras, se ha desarrollado un wordcloud, que filtra los mensajes de los días en los que la puntuación promedio es igual a 1 (máximo nivel de malestar):

Fuente: Elaboración propia a partir del grupo vecinal de WhatsApp, filtrados por puntuación de malestar corregido ($\text{sent corr} = 1$) y fechas entre noviembre 2024 y febrero 2025.



Ilustración 5: Wordcloud de bigramas más frecuentes en mensajes de alto malestar (*sent_corr* = 1) entre noviembre 2024 y febrero 2025

Durante este periodo sin conciertos en el Bernabéu, los vecinos siguen manifestando quejas estructurales ligadas principalmente a la actividad futbolística y a la convivencia urbana:

- **Restricciones de aparcamiento en días de partido.** Los bigramas “reserva plazas” y “zona azul” reflejan el descontento con el menor número de plazas disponibles, el Real Madrid suele cortar calles y habilitar zonas reservadas los días de partido.

- **Preocupación por la seguridad y el orden público.** Frases como “presencia policial” y “evitar delincuencia” indican la demanda de mayor vigilancia en jornadas de alta afluencia de aficionados.
- **Ocio nocturno y botellón.** La recurrencia de “botellón” y “delincuencia aumentada” muestra que, junto al fútbol, persisten comportamientos de consumo masivo en la vía pública que inciden en la percepción de molestia.

Al no estar sesgada por eventos musicales, esta nube de bigramas evidencia que las molestias vecinales se relacionan más con la frecuencia de partidos o con problemas de gestión del tráfico o acceso a los hogares: propone centrar políticas en una ordenanza de tráfico específica para días de fútbol y la gestión de la congestión, la creación de aparcamientos disuasorios y un refuerzo puntual de la seguridad en el perímetro del estadio.

Evolución de mensajes positivos y negativos por día

Para reforzar el análisis anterior, se clasifica cada mensaje corregido en tres categorías:

- Negativo si $\text{sentiment_corr} \leq 2$
- Neutral si $\text{sentiment_corr} = 3$
- Positivo si $\text{sentiment_corr} \geq 4$

A partir de esta clasificación, se graficó la evolución diaria del número de mensajes positivos y negativos:

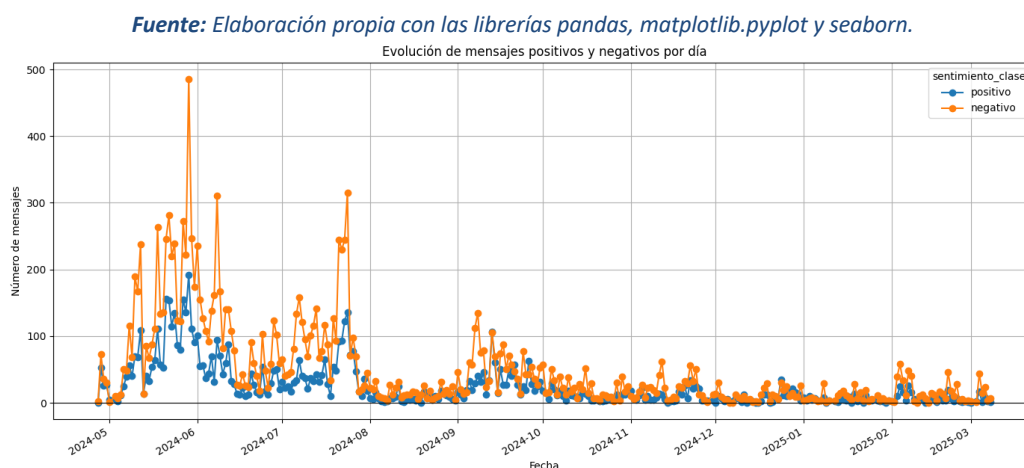


Ilustración 6: Evolución diaria del número de mensajes positivos y negativos en el grupo de WhatsApp vecinal. En el eje X se representan las fechas (entre abril de 2024 y marzo de 2025), y en el eje Y el número de mensajes clasificados por tipo de sentimiento: positivo (puntuación corregida ≥ 4) y negativo (≤ 2).

Podemos sacar de esta gráfica, que los mensajes negativos, rebasan casi todos los días a los positivos, especialmente entre los meses de mayo a agosto de 2024, franja que coincide con los conciertos celebrados en el Santiago Bernabéu.

A continuación, se muestra la evolución del sentimiento promedio corregido (*sentiment_corr*). La banda rosada marca los días con tono negativo (media < 3), y las líneas discontinuas indican las fechas de concierto. Se observa una clara concentración de sentimiento negativo en torno a estos eventos, lo que refuerza la hipótesis de un impacto emocional recurrente.

Fuente: Elaboración propia con *pandas* y *matplotlib.pyplot* (utilizando *fill_between* para sombreado y *axvline* para fechas).

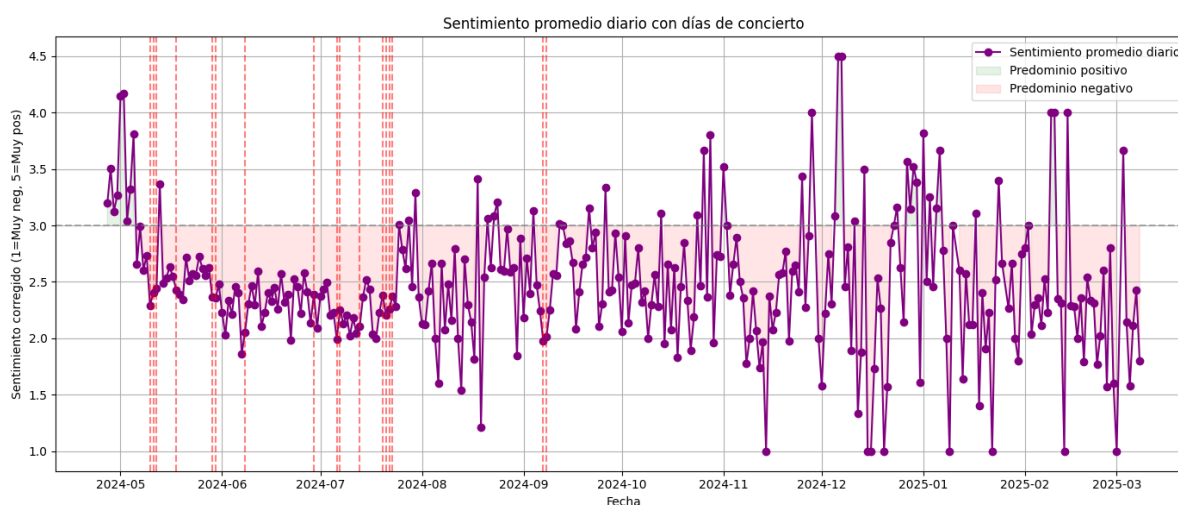


Ilustración 7: Sentimiento promedio diario corregido (*sentiment_corr*) a lo largo del tiempo. El eje X indica la fecha y el eje Y el valor promedio diario del sentimiento en una escala de 1 (muy negativo) a 5 (muy positivo). Las bandas rosadas señalan los días con tono negativo (media < 3) y las líneas rojas punteadas marcan los días de concierto.

Este gráfico revela que en muchas fechas clave, los mensajes negativos superan ampliamente a los positivos, lo que refuerza la hipótesis del malestar vecinal generado por los conciertos.

Con el objetivo de analizar más específicamente el impacto emocional que generan los conciertos, se realizó un análisis centrado en los días inmediatamente anteriores y posteriores a los eventos.

Fuente: Elaboración propia con pandas y matplotlib.pyplot, agregando por día relativo a través de timedelta.

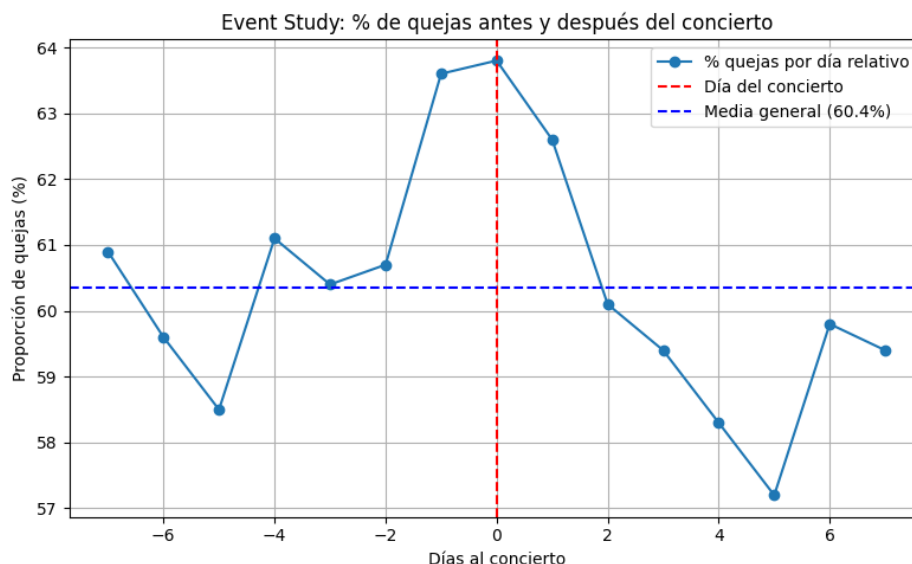


Ilustración 8: Análisis tipo event study de la proporción de mensajes negativos en los días relativos a los conciertos, considerando una ventana de ± 7 días. En el eje X se representan los días respecto al concierto (día 0) y en el eje Y el porcentaje de mensajes clasificados como quejas. Se incluye una línea roja vertical que marca el día del evento y una línea azul horizontal que representa la media general.

En este gráfico se representa el porcentaje de mensajes clasificados como quejas ($\text{sentiment_corr} \leq 2$) para cada día relativo al concierto. Se observa una acumulación de quejas en los días previos al evento, especialmente en -1 y 0, con una posterior disminución a partir del día siguiente al concierto. Este patrón refuerza la idea de una reacción anticipatoria de malestar vecinal, más que una respuesta posterior.

Para complementar esta visualización, se presenta una tabla (Tabla 1) con los valores exactos de sentimiento medio y proporción de quejas para cada día del rango ± 7 :

Fuente: Elaboración propia con pandas y datetime (usando filtrado con isin() y operaciones con timedelta).

	Día relativo	Proporción de quejas (%)	Sentimiento medio
0	-7	60.9	2.326
1	-6	59.6	2.363
2	-5	58.5	2.418
3	-4	61.1	2.348
4	-3	60.4	2.344
5	-2	60.7	2.327
6	-1	63.6	2.215
7	0	63.8	2.225
8	1	62.6	2.278
9	2	60.1	2.357
10	3	59.4	2.396
11	4	58.3	2.427
12	5	57.2	2.437
13	6	59.8	2.377
14	7	59.4	2.387

Tabla 1: valores diarios de sentimiento promedio corregido y porcentaje de quejas para cada día relativo a los conciertos. El eje de análisis temporal está definido en función de la distancia al evento (de -7 a +7 días), permitiendo detectar los días con mayor carga emocional negativa.

En la tabla se confirma que el mínimo del sentimiento medio y el máximo de proporción de quejas coinciden en los días -1 y 0, lo que sugiere que el impacto emocional de los conciertos no solo existe, sino que es consistente en el tiempo.

2.1.6 Análisis de emociones específicas (librería NRC)

Además del análisis de sentimiento general, se realizó un estudio más granular para identificar las emociones predominantes en los mensajes vecinales. Este paso permite diferenciar entre simples valoraciones negativas (sentimiento) y emociones concretas como ira, tristeza, miedo o asco, lo cual proporciona una comprensión más rica del impacto emocional que tienen los conciertos sobre la comunidad.

Enfoque léxico: Diccionario NRC adaptado

Se implementó un enfoque léxico utilizando el diccionario NRC Word–Emotion Association Lexicon desarrollado por Saif Mohammad y Peter Turney para el National Research Council Canada (Canada, 2023) traducido y adaptado al español. Este lexicón asigna más de 14 000 palabras a diez categorías emocionales:

- **Ocho emociones básicas:** ira (anger), miedo (fear), anticipación (anticipation), confianza (trust), sorpresa (surprise), tristeza (sadness), alegría (joy) y asco (disgust).
- Y **dos sentimientos:** positivo (positive) y negativo (negative).

Se procesaron los mensajes y se contaron las ocurrencias de palabras que coincidían con entradas emocionales del diccionario. La suma de coincidencias permitió construir una distribución proporcional del total de emociones detectadas léxicamente.

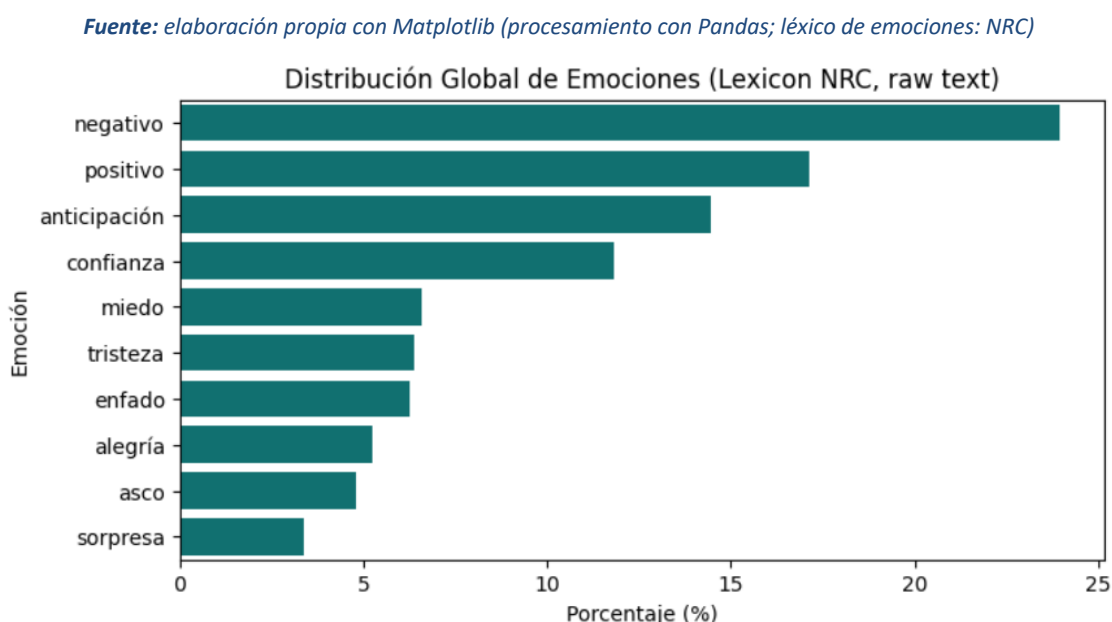


Ilustración 9: Distribución global de las emociones según el léxico NRC aplicado al total de mensajes de la muestra

Podemos observar en la gráfica, como ya corroboraban otros gráficos, que efectivamente los mensajes negativos superan globalmente a los positivos.

Para poder realmente analizar el impacto de los conciertos, se dividió la muestra de mensajes en dos grupos: Aquellos mensajes enviados en días con concierto, y los enviados en días sin concierto. De esta manera se han podido derivar dos gráficas, que con ayuda del diccionario NRC, clasifican los mensajes en sentimientos (positivo o negativo) y emociones específicas:

Fuente: elaboración propia en Python con Matplotlib

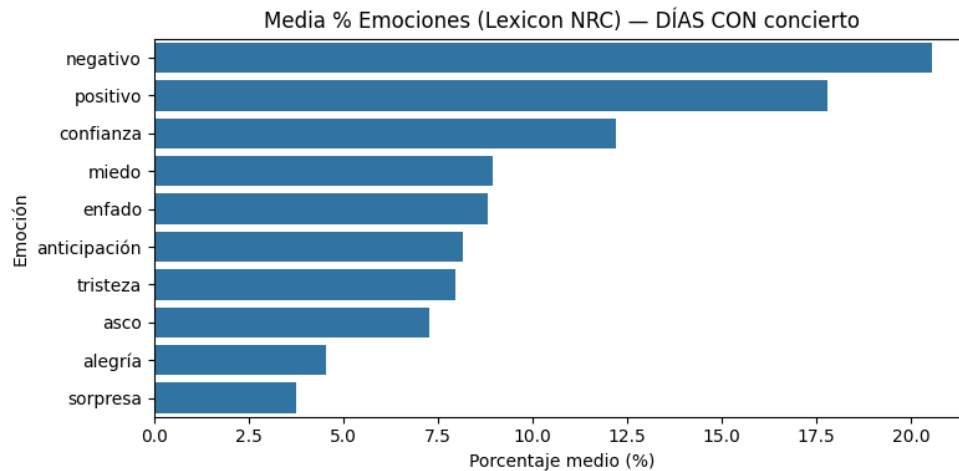


Ilustración 10: Distribución global de las emociones según el léxico NRC aplicado a los mensajes enviados en días de concierto

Fuente: elaboración propia en Python con Matplotlib

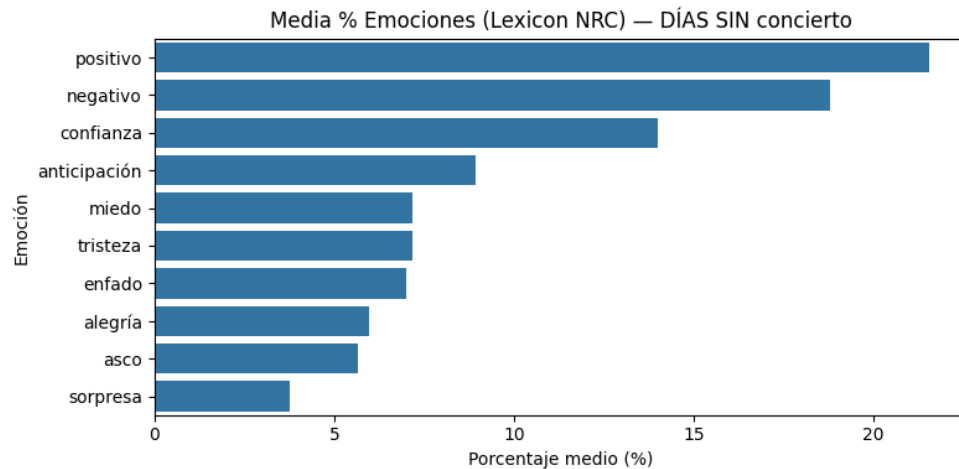


Ilustración 11: Distribución global de las emociones según el léxico NRC aplicado a los mensajes enviados fuera de los días de concierto

Los resultados del análisis de emociones muestran que, durante los días sin concierto, predomina los mensajes clasificados como positivos (aproximadamente un 21,6 % del total), seguida de la emoción negativa (17,1 %) y de la confianza (14,5 %). Sin embargo, cuando se celebran conciertos en el estadio, este patrón se invierte: la proporción de expresiones negativas aumenta hasta un 20,8 %, superando a las positivas (17,2 %), mientras que la confianza se reduce a un 11,9 %.

Este análisis detallado permite concluir que los conciertos generan un patrón repetido de tensión emocional observable tanto en el volumen de quejas como en el descenso del tono emocional general. Las métricas aplicadas (sentiment_corr y su promedio, junto a la proporción de quejas) proporcionan una base sólida para contrastes estadísticos y análisis causales posteriores.

Con el objetivo de mostrar de forma intuitiva qué palabras se repiten en los mensajes más críticos o en los más favorables. Las wordclouds muestran las palabras más frecuentes según el tamaño de la palabra en la visualización. Después de eliminar las stopwords (del diccionario stopwords), se visualizaron por una parte un wordcloud general, y más adelante se elabora un wordcloud específico para el día 13 de septiembre, día que se cancelaron los conciertos.

Fuente: elaboración propia en Python con la librería WordCloud (generación de la nube) y Matplotlib.



En esta nube de palabras destacan términos como “**conciertos**”, “**vecinos**”, “**noche**”, “**días**” y “**policía**”, lo cual refleja con claridad las principales preocupaciones de la comunidad: la frecuencia y duración de los eventos, los ruidos nocturnos y la intervención de las fuerzas de seguridad. Palabras como “**uso**”, “**van**” o “**coches**” apuntan también a las molestias logísticas (tráfico y estacionamiento), mientras que “**semana**”, “**horas**” o “**mañana**” subrayan la percepción de un trastorno que se extiende más allá de un solo día de fiesta.

Wordcloud para positive

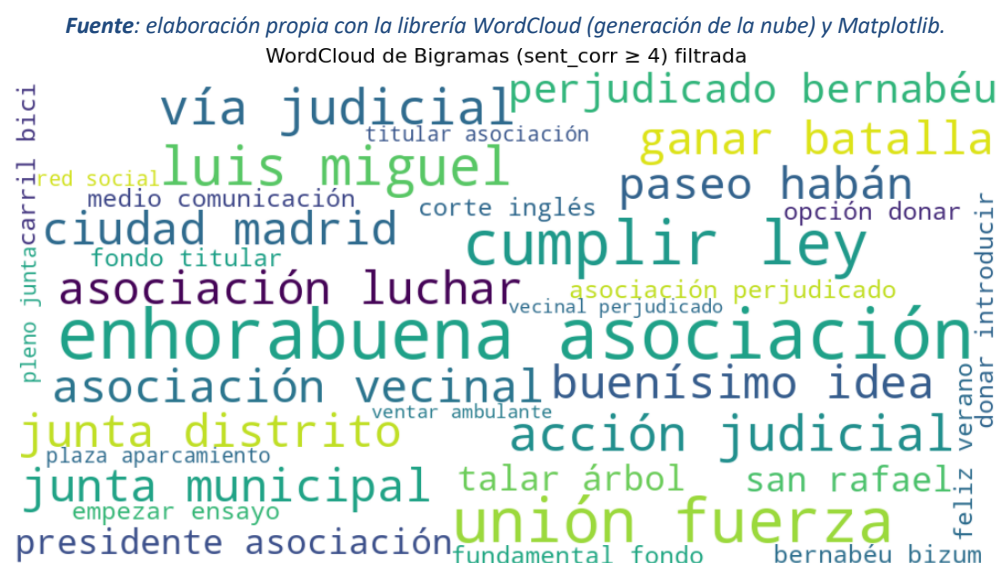


Ilustración 13: Nube de palabras generada a partir de los mensajes con sentimiento positivo en el grupo de WhatsApp vecinal, tras eliminar stopwords.

La nube de bigramas de los mensajes positivos del grupo contra las obras muestra tres ideas clave:

1. **Reconocimiento interno:** “enhorabuena asociación” y “buenísimo idea” reflejan cómo celebran cada propuesta o triunfo conjunto.
2. **Confianza en lo legal:** “vía judicial” y “cumplir ley” evidencian que creen en presentar recursos y denunciar formalmente las obras.
3. **Cohesión y victoria compartida:** “unión fuerza” y “junta municipal” subrayan que cada logro se vive como éxito colectivo.

Wordcloud para negative

Fuente: elaboración propia con la librería WordCloud (generación de la nube) y Matplotlib.

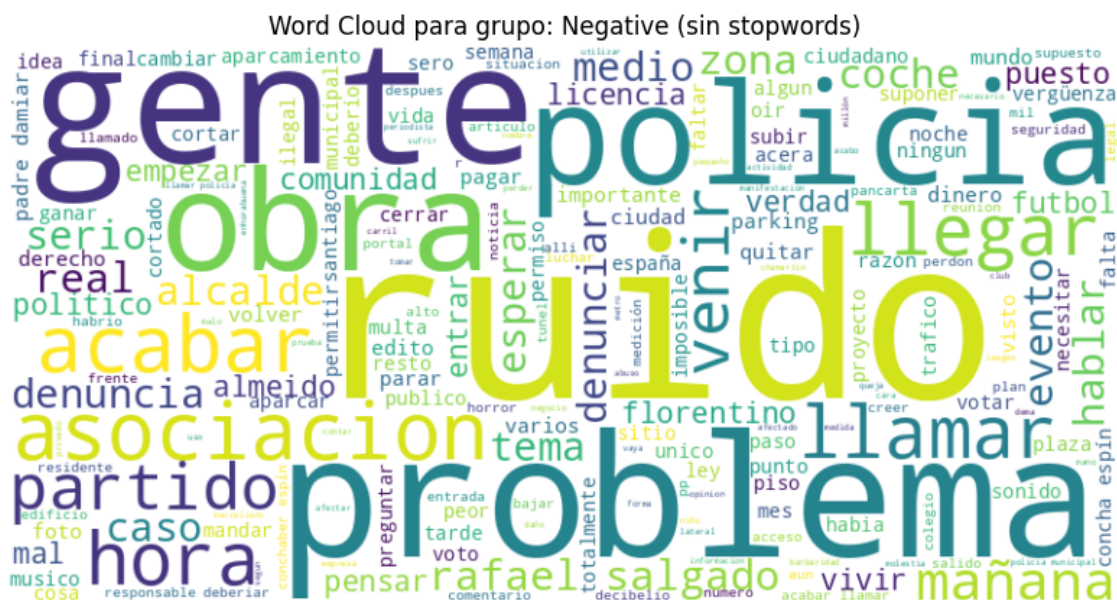


Ilustración 14: Nube de palabras de los mensajes clasificados con sentimiento negativo, sin stopwords.

El foco se vuelve casi monotemático:

- “Ruido” predomina, es la molestia fundamental.
- “La gente” (afluencia), “policía” (peticiones de intervención) y “obras” (remodelación del estadio) resaltan factores de estrés adicionales.
- Palabras de acción/reclamo (“obligar”, “denunciar”, “llamar”) revelan que los vecinos no solo se quejan, sino que exigen soluciones o medidas disciplinarias.

En conjunto, estos word clouds no solo resaltan las palabras más repetidas, sino que también ayudan a identificar puntos críticos (calles concretas, horas de mayor molestia, figuras implicadas) y a orientar futuras propuestas de gestión urbana, como la planificación de rutas de acceso, ajustes de horarios o protocolos específicos de actuación policial.

WordCloud del 13 de septiembre, día de cancelación de los conciertos

El 13 de septiembre de 2024 marcó un punto de inflexión en el conflicto entre el Real Madrid y los residentes del barrio de Chamartín. Ese día el club anunció la suspensión temporal de todos los conciertos programados en el renovado Santiago Bernabéu hasta completar las obras de insonorización necesarias para cumplir la normativa acústica municipal. La decisión, hecha pública tras varias denuncias vecinales y la imposición de sanciones económicas por parte del Ayuntamiento, provocó un vuelco inmediato en el posicionamiento del barrio.

Fuente: elaboración propia con la librería WordCloud (generación de la nube) y Matplotlib.

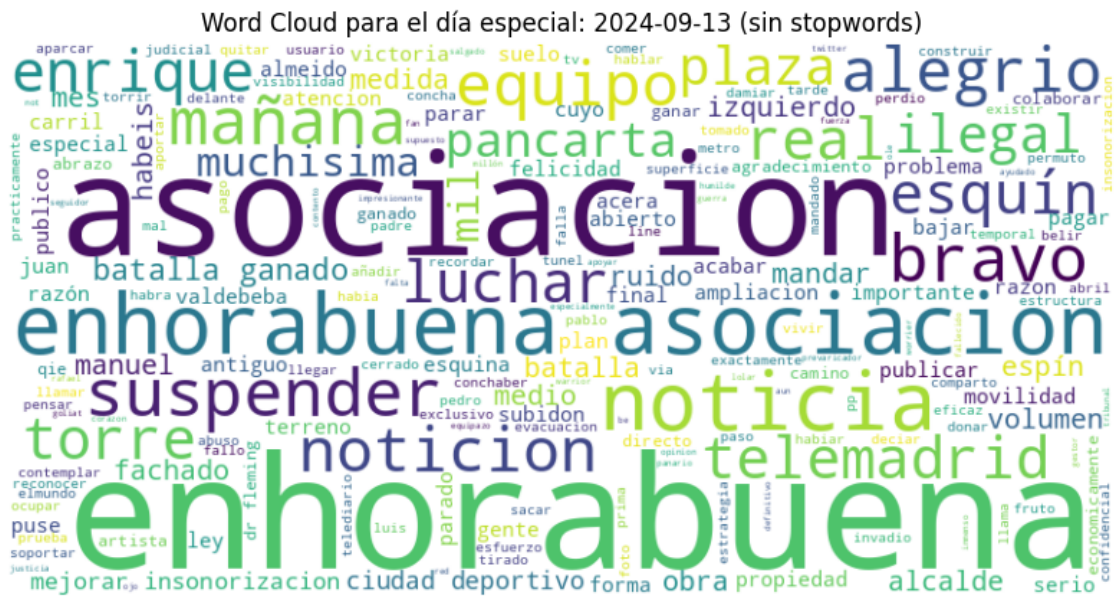


Ilustración 15: Nube de palabras específica para el día 13 de septiembre de 2024, fecha en la que se anunció la suspensión de los conciertos en el Santiago Bernabéu.

El análisis de sentimiento aplicado al grupo de WhatsApp vecinal que hasta entonces mostraba un predominio de mensajes negativos vinculados al “ruido” y a la “denuncia” registró un incremento de emociones positivas. La *word-cloud* específica de esa jornada evidencia el cambio: términos como “enhorabuena”, “bravo”, “alegría” y “notición” se impusieron a los habituales “ruido” o “problema”. Del mismo modo, la palabra “asociación” destacó por encima del resto, reflejando la percepción de que la victoria era fruto de la acción colectiva emprendida por la organización vecinal.

Wordcloud de los mensajes clasificados como es_queja

Para identificar visualmente los temas más recurrentes entre los mensajes vecinales, se generó una nube de palabras específica a partir de los mensajes etiquetados como quejas (es_queja = 1). El procedimiento seguido fue el siguiente:

Primero, se filtraron únicamente los mensajes clasificados como queja. Sobre los textos seleccionados, se aplicó una limpieza de palabras vacías (stopwords) para eliminar términos comunes irrelevantes. A continuación, los mensajes fueron unificados en un solo corpus de texto. Finalmente, se construyó una *WordCloud* utilizando la frecuencia de aparición de cada palabra en el corpus.

Fuente: elaboración propia con la librería WordCloud (generación de la nube) y Matplotlib.



Ilustración 16: Nube de palabras que representa las palabras más representativas de los mensajes clasificados como queja (variable es_ queja)

El análisis de la nube de palabras muestra que las quejas vecinales se concentran en seis términos principales, ruido, concierto, obra, vecino, policía y denunciar, lo que indica que las molestias derivan sobre todo del impacto acústico de conciertos y obras, de las gestiones con la policía y de las afectaciones al entorno residencial.

Esta visualización sintetiza de forma clara los temas recurrentes en los mensajes de queja del grupo.

Conclusiones generales

- **Ruido como factor central.** ‘Ruido’ es el término más prominente, lo que confirma que las emisiones sonoras son el principal detonante de las reclamaciones.
- **Asociación vecinal como canal de apoyo.** Aparece con frecuencia para articular denuncias y coordinar acciones, asumiendo el rol de portavoz comunitario.
- **Tono positivo:** En las menciones favorables dominan expresiones de gratitud y adjetivos de refuerzo (‘totalmente’, ‘increíble’), reconociendo buenas prácticas o soluciones efectivas.
- **Acción y urgencia en lo negativo:** En los mensajes de queja abundan verbos imperativos (‘denunciar’, ‘llamar’, ‘obligar’), evidenciando solicitudes concretas (insonorización, presencia policial, trámites oficiales).

2.1.8 Análisis de decibelios y detección de excesos de ruido

Uno de los principales temas de queja expresados por los vecinos está relacionado con el ruido excesivo generado por los conciertos. Para capturar este fenómeno, se diseñó una metodología específica para extraer y analizar menciones explícitas a decibelios dentro de los mensajes del grupo de WhatsApp.

a) Extracción de cifras en contexto horario

Se desarrolló una función que detecta la presencia de números enteros en los mensajes, bajo la hipótesis de que cuando estos aparecen durante la noche del concierto (definida como 20:00 a 06:00), están probablemente vinculados a niveles de ruido.

Para ello:

- Se combinó la fecha del mensaje (fecha_dt) con la hora (hora) para determinar si el mensaje fue emitido en la ventana temporal crítica.
- Se aplicaron expresiones regulares para extraer todos los números mencionados en el texto.
- Se creó una columna numbers_night_concert con las cifras detectadas únicamente en

mensajes durante la “noche extendida”.

- Este proceso fue aplicado a todos los mensajes del dataset, permitiendo una cuantificación sistemática de las menciones numéricas relevantes.

b) Evaluación legal de los valores reportados

Una vez extraídos los números, se definieron umbrales legales de ruido según la legislación vigente:

- Hasta las 23:00, el límite legal es 65 decibelios (dB).
- A partir de las 23:00, el límite desciende a 45 dB.

Para cada mensaje con números detectados, se evaluó si alguno de ellos superaba el umbral correspondiente a su hora de emisión. El resultado se almacenó en la variable booleana `db_exceeds_limit`, y también se calculó `db_excess_count`, que indica cuántas veces se excedió el límite dentro de un solo mensaje.

Los resultados se agregaron por día para construir una serie temporal de mensajes con exceso de ruido detectado. Esta información se muestra en dos visualizaciones:

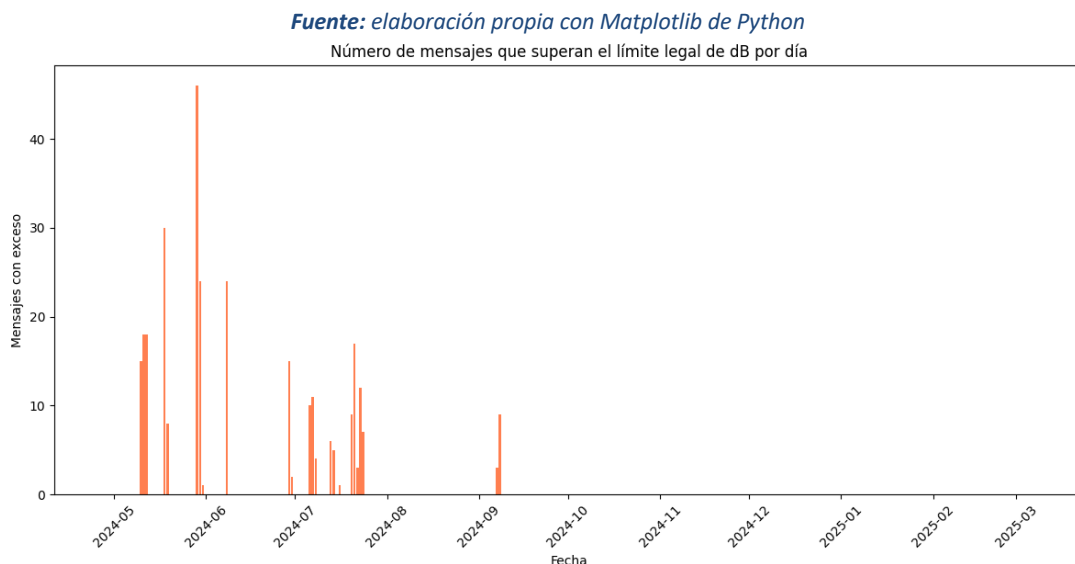


Ilustración 17: Evolución diaria del número de mensajes en el grupo de WhatsApp vecinal que contienen menciones a niveles de ruido superiores a los límites legales de decibelios. El eje X representa la fecha (desde abril hasta septiembre de 2024) y el eje Y indica el número de mensajes diarios

Fuente: elaboración propia con Matplotlib (procesamiento y agrupación con Pandas).

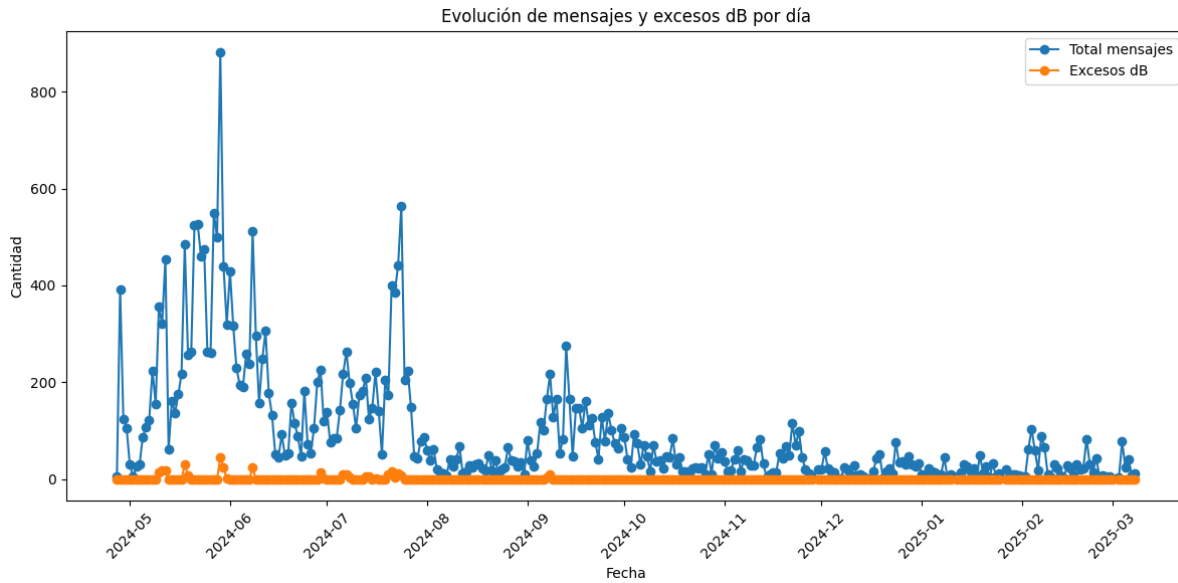


Ilustración 18: Evolución diaria del total de mensajes enviados en el grupo de WhatsApp vecinal (línea azul) y del número de mensajes que reportan niveles de ruido superiores al umbral legal de decibelios (línea naranja), entre abril de 2024 y marzo de 2025. El eje X representa la fecha y el eje Y indica la cantidad diaria de mensajes

Los días con conciertos muestran una coincidencia clara entre picos de mensajes totales y aumentos en los mensajes que mencionan decibelios superiores al umbral legal. En particular:

- El 10 de mayo, el 13 de julio y el 8 de junio registran más de 40 mensajes diarios con exceso de dB.
- Se observa una concentración de excesos especialmente en el tramo posterior a las 23:00, donde el límite es más restrictivo (45 dB), lo cual agrava la percepción de molestia nocturna.
- Esta variable `db_exceeds_limit` se convierte en una de las más útiles para construir modelos explicativos del malestar vecinal, ya que representa una señal directa del impacto acústico.

2.2 Análisis de patrones mediante técnicas de Machine Learning

2.2.1 Modelos estadísticos

Tras el análisis exploratorio, se procedió a aplicar un conjunto de modelos estadísticos y de aprendizaje automático con el objetivo de evaluar el impacto de los conciertos en las quejas vecinales. Estos modelos permiten medir asociaciones, patrones repetitivos y, en algunos casos, estimar efectos causales. A continuación, se describen cada uno de los métodos aplicados, los datos utilizados, las dificultades encontradas y los resultados obtenidos.

Con respecto a los métodos aplicados para explorar patrones y causalidad, se presentan los siguientes:

- Resumen estadístico de contrastes
- Clasificación de quejas con reglas léxicas
- TF-IDF + KMeans clustering
- Regresión logística (ventana ± 2 días)
- Regresión logística multivariable con decibelios, longitud, emojis, etc.
- Difference-in-Differences (DiD)
- Regression Discontinuity Design (RDD)

Uno de los primeros objetivos de este trabajo fue determinar si los conciertos del estadio Santiago Bernabéu generan un aumento significativo en la actividad vecinal. Para evaluar esta hipótesis, se aplicaron diversos contrastes de medias, tanto paramétricos como no paramétricos, utilizando como variable principal el número de mensajes clasificados como queja en el grupo de WhatsApp.

La base de datos se transformó agregando los mensajes a nivel diario, y se construyó una variable binaria `es_concierto` que toma el valor 1 si la fecha corresponde a un concierto y 0 en caso contrario. Se computó la suma diaria de mensajes de queja (`num_quejas`) y se generaron subconjuntos para los días con concierto y los días sin evento.

2.2.2 Resumen estadístico de contrastes

Fuente: elaboración propia con Python (Pandas, SciPy) sobre el conjunto de datos de quejas diarias.

Medida	Signifi- cativo?	p-valor	Cifra	Interpretación	¿Que mide la prueba?
t-Student (Student, 1908)	Si	<0.001	T=8.94	Días con concierto tienen significativamente más quejas en promedio	Compara si hay diferencias significativas entre las medias de dos grupos independientes.
Mann Whitney U (Mann & Whitney, 1947)	Si	P<0.001	U=5271	Distribución de quejas en días con concierto es estadísticamente mayor	Evalúa si los valores de una muestra tienden a ser mayores que los de otra sin asumir distribución normal.
Permutación de Montecarlo (Good, 2005)	Si	P<0.001	86.48	Diferencia de medias entre grupos muy improbable por azar	Estima la probabilidad de obtener una diferencia tan grande como la observada mediante reordenamientos aleatorios.
Cohen's d (Cohen, 2013)	Si	P<0.001	D=1.66	Efecto muy grande; días con concierto presentan gran aumento en quejas	Mide el tamaño del efecto entre dos medias: cuán grande es la diferencia entre grupos en desviaciones estándar.
Rate Ratio (Rothman et al., 2008)	Si	P<0.001	RR = 4.8 (CI 95%: 3.5 – 6.7)	Los conciertos multiplican x4.8 el nº de quejas comparado con días normales	Compara la tasa de ocurrencia de un evento entre dos grupos; indica cuántas veces más frecuente es un suceso.

Tabla 2: Resultados de las pruebas de estadísticas que confirman un aumento significativo de quejas en días de concierto ($p < 0,001$) y cuantifican el tamaño del efecto.

2.2.3 Visualización de la distribución de sentimiento

Además del conteo puro de mensajes, se analizó la evolución del sentimiento expresado por los vecinos con el objetivo de detectar alteraciones en el tono emocional colectivo en torno a los conciertos. Este ajuste permitió detectar mejor los días emocionalmente cargados. La distribución general del sentimiento asignado una vez procesados los mensajes y eliminadas las stopwords, se muestra en la Figura 20:

Fuente: elaboración propia con Matplotlib

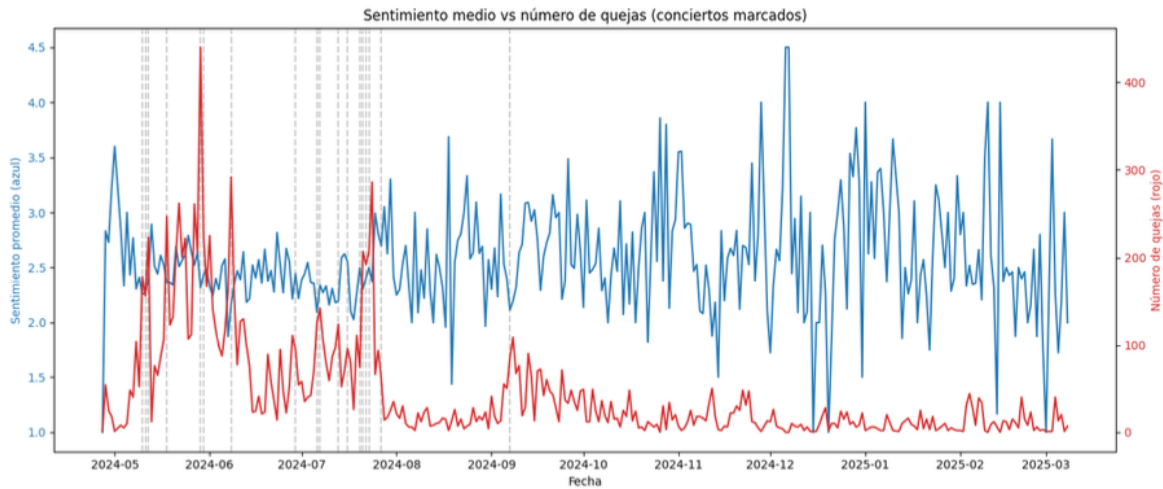


Ilustración 19: Evolución del sentimiento promedio diario en los mensajes del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu. El eje X representa la fecha (entre abril de 2024 y marzo de 2025), mientras que el eje Y (lado izquierdo) muestra el valor promedio de sentimiento diario en escala de 1 (muy negativo) a 5 (muy positivo). Las líneas grises indican las fechas de conciertos. En el eje Y se muestra el número de quejas junto con la distribución diaria que se representa en la gráfica roja.

Podemos observar una fuerte tendencia de quejas en los meses de conciertos y una posterior disminución a partir de agosto. Asimismo, podemos ver fuertes fluctuaciones del sentimiento, indicando la alta polaridad y variabilidad del tipo de mensajes enviados.

Fuente: elaboración propia con Pandas (agrupación y suma de la columna es_queja para obtener el número diario de quejas) y Matplotlib (para la visualización)

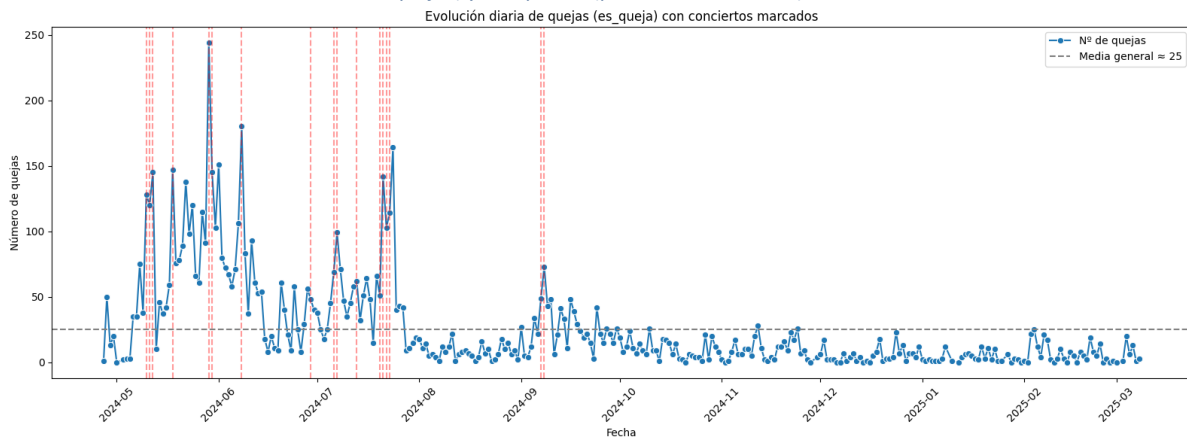


Ilustración 20: Evolución del número de quejas por día del grupo de WhatsApp vecinal, en relación con las fechas de conciertos y la cancelación de eventos en el estadio Santiago Bernabéu (rayas rojas). El eje X representa la fecha (entre abril de 2024 y marzo de 2025), y el eje Y el número de fechas. La raya gris punteada refleja la media de quejas general

Fuente: elaboración propia con Matplotlib (procesamiento y conteo de la columna es_queja mediante Pandas)

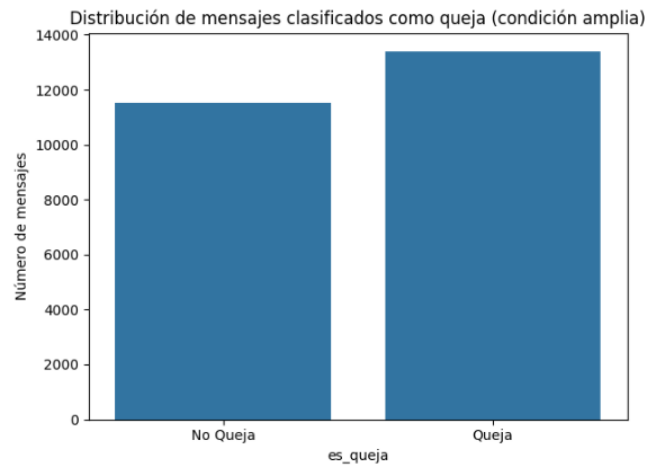


Ilustración 21: Gráfico de barras que representa la distribución total de mensajes como queja o no queja según si cumple que el mensaje contenga una palabra del diccionario de queja (creado manualmente) o que con el uso de BERT tenga una puntuación igual o menor a 2

Las ilustraciones 21 y 22 nos muestran en individual el número de quejas y la proporción de mensajes del grupo que se clasifican como queja, que son superiores a los mensajes que no son queja. Esta tendencia nos indica que probablemente, aunque haya variabilidad en el sentimiento, el tema central del grupo principalmente gira en torno a las quejas con el objetivo del grupo de compartir las preocupaciones y temas de importancia que se relacionan con los cambios de la dinámica del vecindario. Estos resultados junto al análisis exploratorio anterior, las *wordclouds* y las pruebas estadísticas, demuestran que efectivamente, hay un ambiente muy tenso alrededor de la situación en el barrio.

2.2.4 Evolución de quejas alrededor de los conciertos

Complementariamente, se analizaron las proporciones de mensajes clasificados como queja por día (% quejas diario) alrededor de los conciertos. Se analizaron los tres días anteriores y posteriores al concierto (=día 0):

Fuente: elaboración propia con Matplotlib (procesamiento y agregación de datos, cálculo de porcentajes de quejas por día relativo a conciertos mediante Pandas).

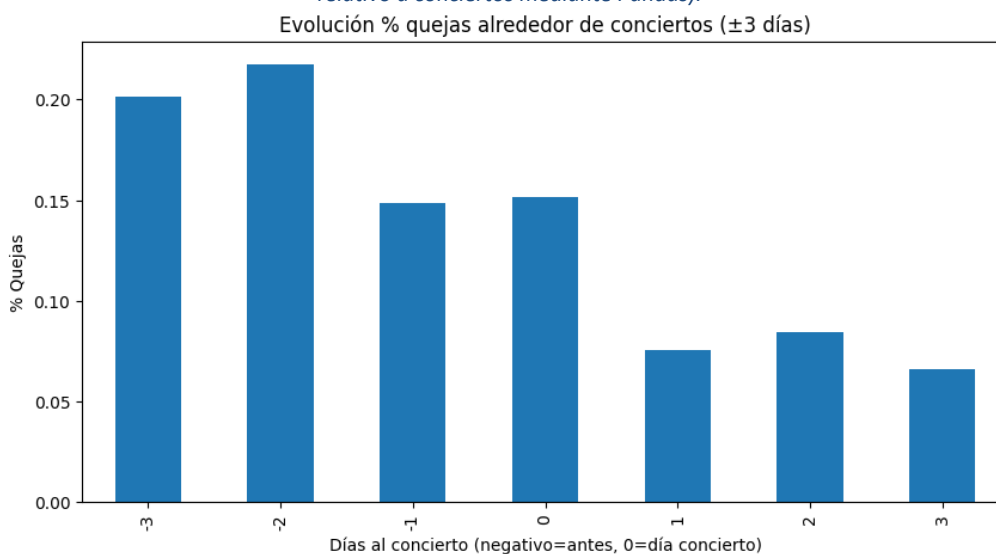


Ilustración 22: Gráfica de barras que muestra la evolución de quejas en porcentaje alrededor de los días de concierto. En el eje Y, el porcentaje de fechas y en el eje X los días adyacentes al concierto, siendo los números negativos los días anteriores y los positivos, los posteriores. El día 0 es el día del concierto

El porcentaje de mensajes de queja aumenta notablemente en los días previos al concierto, lo que indica la anticipación de los vecinos a los ruidos y las molestias, especialmente -2 y -3 días, como muestra la Figura 23, y posteriormente disminuyen como respuesta al cese del ruido.

2.2.5 TF-IDF + K-MEANS

Para explorar la diversidad de temas y reacciones vecinales en el grupo de WhatsApp, se implementó un enfoque de clustering semántico que combina vectorización TF-IDF, reducción de dimensionalidad mediante SVD y agrupamiento con KMeans. Esta estrategia permite segmentar los mensajes en tipologías coherentes que representan distintos focos de preocupación, acción o actitud.

Preparación del texto y vectorización

Tras la limpieza textual y la lematización de los mensajes, se construyó una matriz TF-IDF utilizando ngramas de una y dos palabras (`ngram_range=(1,2)`). Para evitar ruido textual, se eliminaron los términos extremadamente comunes (`max_df=0.7`) así como aquellos poco frecuentes (`min_df=4`), asegurando que solo se incluyeran palabras con una frecuencia

representativa pero no irrelevante. La matriz final se limitó a un máximo de 2000 términos mediante `max_features=2000`, priorizando los de mayor varianza.

Esta matriz es de alta dimensionalidad y dispersa, lo cual dificulta su uso directo para algoritmos de agrupamiento y visualización. Por ello, se aplicó Truncated Singular Value Decomposition (Truncated SVD) (Khadka, 2023), una técnica de reducción de dimensionalidad análoga al Análisis de Componentes Principales (PCA), pero optimizada para matrices dispersas como las generadas por TF-IDF. Esta técnica permite proyectar los documentos en un espacio semántico reducido, preservando la estructura del texto y facilitando tanto la agrupación como la interpretación.

En este trabajo se conservaron 700 componentes tras el SVD (Falini, 2022), ya que esta cantidad permite explicar más del 75% de la varianza acumulada en el espacio vectorial original, tal como se muestra en la Figura 24. Esta decisión busca un equilibrio entre preservar la información semántica relevante y mejorar la eficiencia computacional de los algoritmos posteriores.

Fuente: elaboración propia con Matplotlib (para la visualización), scikit-learn (para la descomposición SVD y el cálculo de la varianza explicada) y NumPy (para el cálculo de la varianza acumulada).

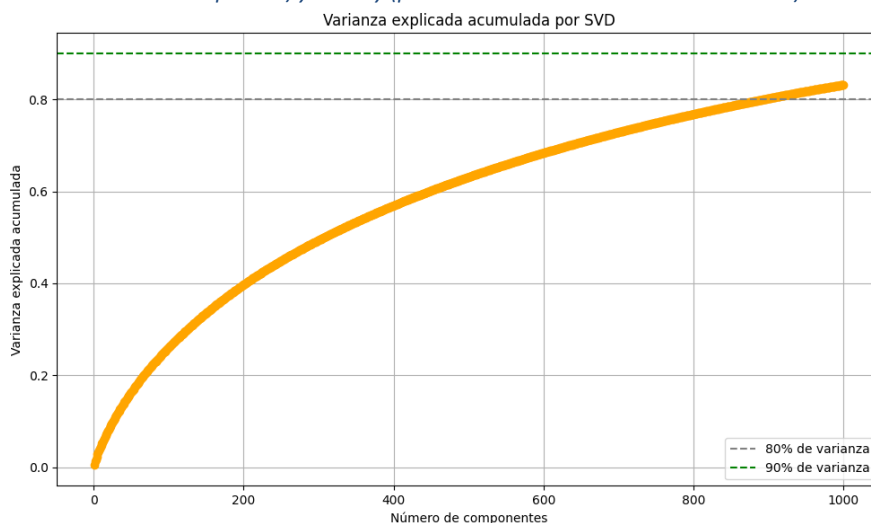


Ilustración 23: Varianza explicada acumulada obtenida al aplicar SVD a la matriz TF-IDF de los mensajes vecinales. El eje horizontal muestra el número de componentes retenidos y el eje vertical la proporción de varianza total explicada. Las líneas punteadas gris y verde señalan los umbrales del 80 % y el 90 % de varianza explicada, respectivamente.

A continuación, centrándonos en la matriz densa de 700 dimensiones resultante de SVD, aplicamos PCA para extraer las dos componentes que capturan la mayor varianza y facilitar así la visualización bidimensional de los clústeres.

Selección del número de clústeres y visualización con PCA y componentes semánticas

El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad que busca construir nuevas variables, las componentes principales, como combinaciones lineales de las variables originales, de modo que:

1. Maximicen la varianza explicada

- La primera componente retiene la mayor parte de la dispersión total de los datos, la segunda retiene la mayor parte de la varianza residual (siempre ortogonal a la primera), y así sucesivamente.

2. Sean ortogonales entre sí

- Cada componente es independiente (no correlacionada) de las anteriores, garantizando que no haya solapamiento de información.

Matemáticamente, PCA se obtiene resolviendo la descomposición en valores propios de la matriz de covarianzas (o, de forma equivalente, mediante SVD sobre la matriz centrada de datos). Si X es la matriz reducida de TF-IDF (documentos $\times$ términos), la proyección en dos dimensiones viene dada por

$$Z = X W_2$$

donde W_2 contiene los dos vectores propios asociados a los dos mayores valores propios de la covarianza de X

- **Loadings (cargas):** los coeficientes de estos vectores propios indican el peso de cada término (palabra) en la componente. Aquellos con mayor valor absoluto son los que más contribuyen a esa dirección semántica.

Aplicando PCA a 2D podemos representar cada documento en un gráfico de dispersión, de forma que los clústeres resultantes se visualizan según cómo se agrupan en ese espacio reducido, manteniendo la mayor parte de la estructura original.

Por tanto, la matriz reducida fue proyectada en dos dimensiones mediante PCA (Jolliffe & Cadima, 2016) para visualizar los clústeres:

Fuente: elaboración propia con Matplotlib (procesamiento con Pandas; reducción de dimensionalidad y clustering con scikit-learn: PCA y KMeans).

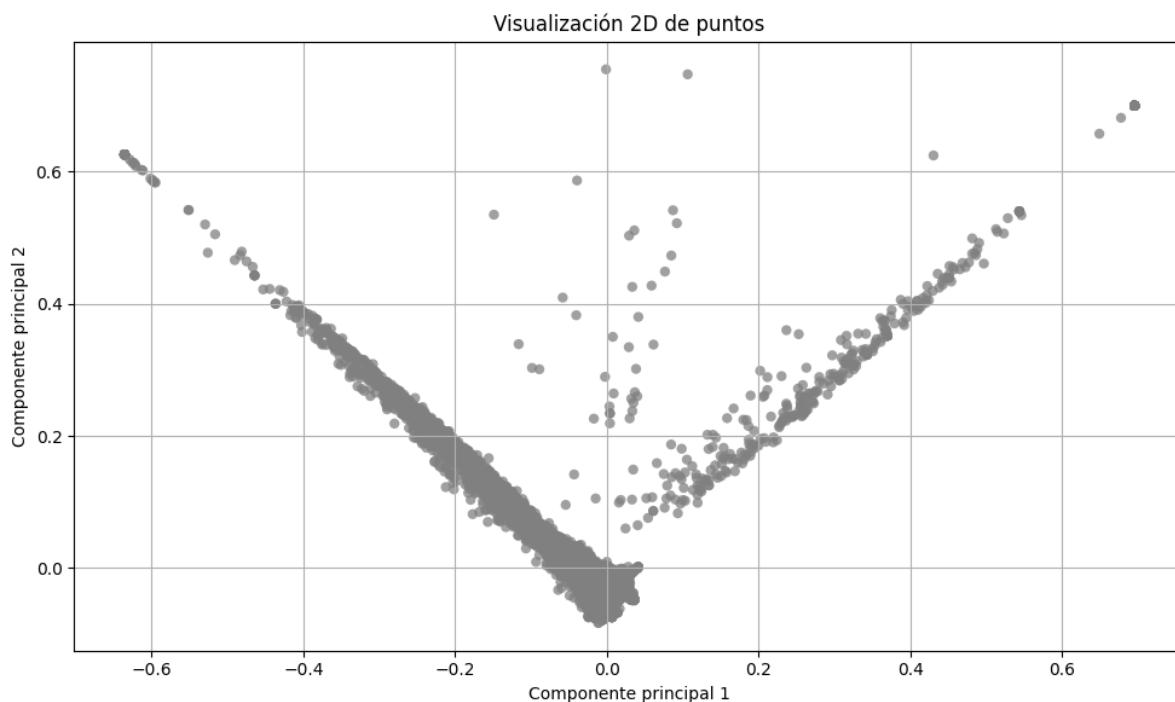


Ilustración 24: Visualización bidimensional de los clústeres obtenidos con KMeans tras proyectar la matriz TF-IDF reducida mediante PCA. Cada punto representa un documento y está sin colorear para captar algún patrón de manera visual. El eje horizontal (Componente principal 1) y el eje vertical (Componente principal 2) corresponden a las dos primeras componentes principales, que condensan la mayor parte de la varianza latente del texto.

Teniendo en cuenta la imagen, se adoptó un enfoque pragmático en vez de usar métodos de selección del número de clusters, como puede ser el método del codo o el método de silueta por la naturaleza de la distribución.

Basándose en la visualización del espacio semántico reducido con PCA, se visualizan como los mensajes se agrupan claramente en torno a dos grandes componentes ortogonales. Esta representación reveló dos direcciones principales en el discurso: una asociada a la protesta explícita por los conciertos y otra vinculada a mensajes más asociados al día a día como pueden ser aquellos administrativos, de tráfico, otras quejas o de propiedad.

La decisión de reducir a $K=2$ se fundamenta, por tanto, en la evidencia visual de estos dos núcleos semánticos diferenciados, así como en la necesidad de obtener resultados más robustos, sintéticos y directamente interpretables en términos del conflicto social principal.

- **El Cluster 0 (1.247 mensajes)** está conformado por mensajes explícitamente centrados en los conciertos, y contiene términos como: "cada concierto", "próximo concierto", "stop concierto", "ruido concierto", "Bernabéu concierto" o "celebración concierto". Este grupo representa la dimensión de crítica directa hacia los eventos musicales.
- **El Cluster 1 (27.146 mensajes)** incluye palabras como: "propietario", "plazo", "carril", "aparcar", "donación", "taxi", "estación" o "señal", lo que sugiere un discurso más vinculado a cuestiones cotidianas del barrio, logística, aparcamiento, tráfico o gestión administrativa.

Este agrupamiento se ve reforzado por el análisis de las dos primeras componentes principales (PCs) extraídas mediante PCA:

- **PC1** está asociada a términos clave del conflicto: "concierto" (0.4227), "ruido" (0.2121), "bernabeu" (0.1986), "policía" (0.1664), "llamar" (0.1658), "estadio" (0.1327), reflejando un eje semántico claramente vinculado a la queja vecinal.
- **PC2** presenta un contraste entre mensajes de consenso o institucionales ("acuerdo": 0.9451) y mensajes críticos ("concierto": -0.1617, "ruido": -0.0785), lo que permite interpretar este eje como un gradiente entre apoyo/resignación y protesta.

Esta interpretación queda ilustrada en la Figura 26, donde se muestra la distribución de los mensajes en el plano formado por las dos primeras componentes principales. En la gráfica se aprecian con claridad dos agrupaciones compactas: una a la izquierda del eje vertical (mayoritariamente mensajes críticos sobre conciertos) y otra a la derecha (conversaciones más neutras o administrativas). La forma en "V" que adopta la nube de puntos refuerza la existencia de dos polos semánticos bien definidos, que justifican empíricamente la elección de $K=2$ para este corpus.

Fuente: elaboración propia con Matplotlib (visualización), Pandas (procesamiento de datos) y scikit-learn (reducción de dimensionalidad con PCA y clustering con KMeans).

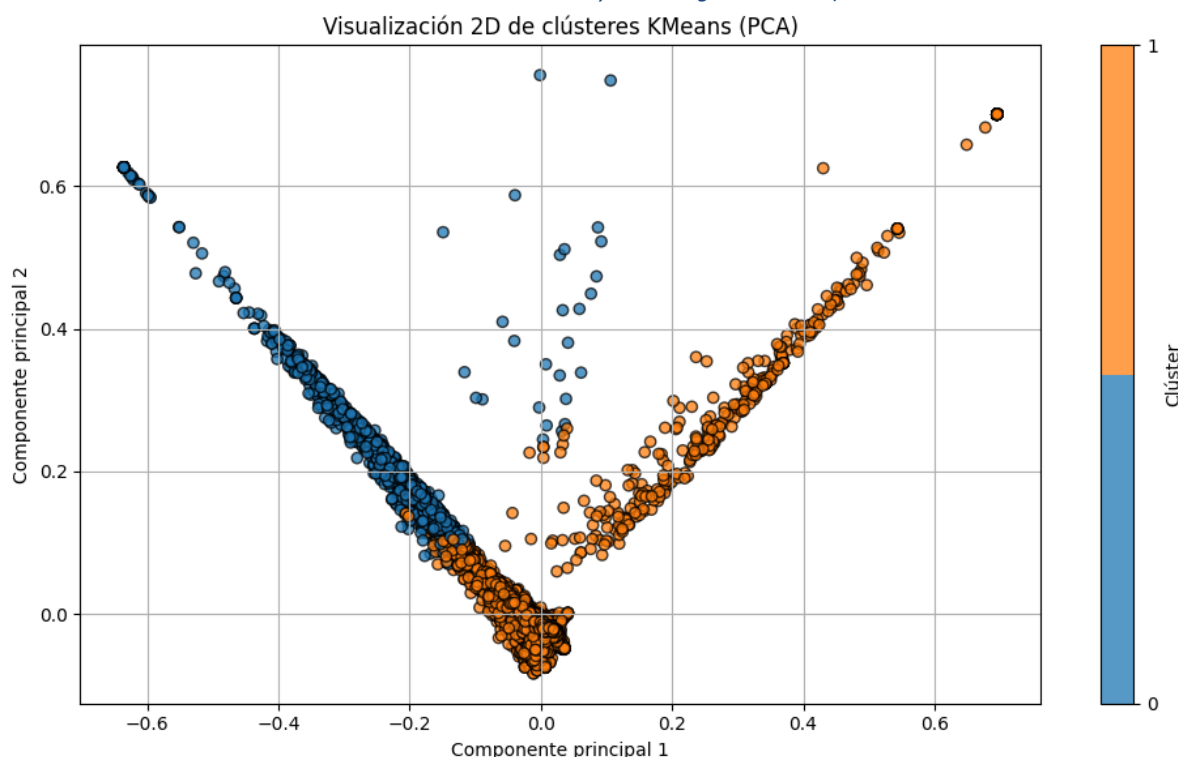


Ilustración 25: Visualización bidimensional de los clústeres obtenidos con KMeans ($k = 2$) tras proyectar la matriz TF-IDF reducida mediante PCA. Cada punto representa un documento y está coloreado según su asignación de clúster (0–4). El eje horizontal (Componente principal 1) y el eje vertical (Componente principal 2) corresponden a las dos primeras componentes principales, que condensan la mayor parte de la varianza latente del texto

En conjunto, esta segmentación binaria proporciona una visión clara de los dos polos discursivos predominantes en el grupo de vecinos, y permite una posterior interpretación cualitativa más afinada de las actitudes, prioridades y tensiones que estructuran la conversación digital sobre el impacto de los conciertos.

Modelado temático con LDA (8 temas)

Como método complementario al clustering, se empleó Latent Dirichlet Allocation (LDA) (Blei et al., 2003) para extraer automáticamente los temas latentes que subyacen en los mensajes. LDA modela cada texto como una mezcla de K temas, donde cada tema es una distribución de probabilidad sobre las palabras del vocabulario. Para ello, primero se convirtió el corpus lematizado en una matriz “bag-of-words” filtrando términos muy frecuentes o escasos y capturando n-gramas de longitud uno y dos (unigramas y bigramas).

A continuación, el modelo LDA se entrenó en modo batch, lo que significa que en cada iteración el algoritmo procesa la totalidad de los documentos para actualizar simultáneamente las distribuciones de temas por documento y de palabras por tema.

Para comprobar que el modelo realmente aprendía y no se estancaba, se monitorizó la perplexity, una métrica interna que mide como de bien el modelo predice las palabras de los documentos: a menor perplexity, mejor el ajuste. Paralelamente, para garantizar que los temas resultantes fueran coherentes y fáciles de interpretar, se calculó la coherencia semántica, un indicador externo que valora si las palabras más frecuentes de cada tópico guardan relación entre sí, más alta coherencia equivale a tópicos más claros.

La elección del número de temas K se basó en el equilibrio entre estas dos métricas: se eligió aquel valor de K en el que la perplexity ya no mejoraba de forma significativa y la coherencia alcanzaba un máximo razonable. De este modo, el modelo logra un buen ajuste a los datos sin fragmentar los temas en exceso, ofreciendo grupos de palabras claramente distinguibles y semánticamente consistentes:

- **Tema 0:** favor, florentino, foto, vídeo, mañana, alguno, red, muchísimo, aquí, chat, comentario
- **Tema 1:** bernabéu, vecino, concierto, estadio, seguir, votar, noticia, asociación, voto, distrito, santiago bernabéu
- **Tema 2:** calle, salgado, coche, plaza, dejar, zona, venir, espín, aparcar, aparcamiento, casa, acera
- **Tema 3:** ruido, concierto, vecino, deber, estadio, problema, zona, evento, barrio, cumplir, persona,
- **Tema 4:** acuerdo, concierto, chat, comunidad, cada, asociación, idea, mismo, verdad, pancarta, vecino,
- **Tema 5:** medio, año, barrio, vivir, hora, llevar, gente, concierto, mas, hablar, taylor, cosa
- **Tema 6:** ayuntamiento, seguir, bueno, caso, razón, concierto, estadio, fútbol, licencia
- **Tema 7:** llamar, policía, acabar, denuncia, casa, ruido, denunciar, dicho, medir, municipal, medición, asociación

La interpretación cualitativa de estos temas sugiere que los mensajes se organizan en torno a diversas preocupaciones y formas de interacción:

- Denuncias y quejas por ruido (Temas 3 y 7)
- Opiniones y apoyo vecinal (Temas 0 y 4)
- Organización comunitaria y acción institucional (Temas 1 y 6)
- Logística urbana cotidiana (Tema 2)
- Comentarios generales sobre la vida en el barrio (Tema 5)

La utilización de LDA permitió así identificar patrones discursivos estructurales en el grupo, complementando el análisis de clústeres mediante KMeans y ofreciendo una visión temática del conflicto vecinal vinculado a los conciertos en el Bernabéu.

Distribución final de clústeres

Con base en el modelo de agrupamiento KMeans con $K=2$, se identificaron dos grandes grupos temáticos que resumen las principales líneas discursivas del grupo de WhatsApp vecinal. Esta segmentación binaria no solo simplifica la interpretación del corpus, sino que también revela de forma nítida los ejes del conflicto y la diversidad de registros comunicativos.

- **Clúster 0:** Discurso de protesta contra los conciertos: Este grupo recoge los mensajes más directamente relacionados con el descontento vecinal hacia los eventos celebrados en el estadio Santiago Bernabéu. Se caracteriza por expresiones como “cada concierto”, “stop concierto”, “concierto ruido”, “Bernabéu concierto” o “celebración concierto”. Predomina un tono crítico y combativo, centrado en las molestias acústicas, la frecuencia de los conciertos y su impacto sobre la calidad de vida en el barrio.
- **Clúster 1:** Asuntos administrativos y cotidianos del vecindario: Agrupa la mayoría de los mensajes (más del 95%) y refleja una comunicación más pragmática, institucional o de gestión barrial. Incluye palabras clave como “propietario”, “plazo”, “carril”, “aparcar”, “donación”, “taxi”, “estación” o “señal”. Aquí, el foco gira en torno a la convivencia, el uso del espacio público, las infraestructuras, el tráfico y otras cuestiones menos polarizadas.

Esta dicotomía evidencia que, aunque el tema de los conciertos genera una fuerte respuesta emocional, la mayor parte del contenido del grupo aborda dinámicas vecinales rutinarias. Sin embargo, la existencia de un clúster específico para las protestas revela que estas manifestaciones tienen un peso temático suficientemente relevante como para constituir una categoría propia dentro del debate colectivo.

La combinación de este análisis con el estudio temático LDA permite capturar tanto las estructuras globales como las variaciones más finas del discurso, aportando una comprensión más rica y articulada del impacto social percibido de los conciertos y la convivencia diaria en el entorno del Bernabéu.

2.2.6 Regresión logística (ventana ± 1 días)

Como primer paso, se realizó un análisis exploratorio mediante un modelo de regresión logística binaria (Ujang Riswanto, 2025) con el objetivo de comprobar si la probabilidad de que un mensaje fuera clasificado como queja aumentaba durante los días inmediatamente anteriores o posteriores a un concierto. Para ello, se definió una variable binaria `en_ventana_concierto` que tomaba el valor 1 si el mensaje había sido enviado en una ventana de ± 1 días respecto a una fecha de concierto, y 0 en caso contrario. La variable dependiente `queja_binaria` tomaba el valor 1 si el mensaje era considerado una queja, y 0 en caso contrario, tras aplicar la clasificación basada en sentimiento corregido y palabras clave.

El modelo estimado fue:

$$\text{logit}(P(\text{queja} = 1)) = \beta_0 + \beta_1 \cdot \text{en_ventana_concierto}$$

El resultado mostró un coeficiente $\beta_1 = 0.2939$, estadísticamente significativo ($p < 0.001$), lo cual indica que el hecho de que un mensaje se envíe en la ventana de ± 1 días respecto a un concierto **aumenta la probabilidad de que sea una queja en un 34%**, en comparación con mensajes enviados fuera de esa ventana ($\text{odds ratio} = \exp(0.2939) \approx 1.34$).

Fuente: elaboración propia con Statsmodels

Optimization terminated successfully.
Current function value: 0.575006
Iterations 5

Logit Regression Results						
Dep. Variable:	queja_binaria	No. Observations:	23475			
Model:	Logit	Df Residuals:	23473			
Method:	MLE	Df Model:	1			
Date:	Sat, 12 Apr 2025	Pseudo R-squ.:	0.003619			
Time:	11:37:46	Log-Likelihood:	-13498.			
converged:	True	LL-Null:	-13547.			
Covariance Type:	nonrobust	LLR p-value:	4.093e-23			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.1641	0.021	-56.151	0.000	-1.205	-1.123
en_ventana_concierto[T.True]	0.2939	0.030	9.901	0.000	0.236	0.352

Ilustración 26: Resultado de la regresión logística binaria aplicada a los mensajes clasificados como queja

Aunque el pseudo R^2 del modelo es bajo (0.0036), lo cual es común en modelos logísticos con una sola variable explicativa, el resultado sí sugiere una relación significativa entre la proximidad al evento y el contenido emocional negativo de los mensajes vecinales. Esta regresión sirve como una primera aproximación al efecto local e inmediato que generan los conciertos sobre la expresión de malestar vecinal.

2.2.7 Regresión logística multivariable con decibelios, longitud, emojis, etc.

Para capturar una representación más completa de los factores asociados a los mensajes de queja, se construyó un segundo modelo de regresión logística multivariable. Este modelo incluyó varias variables lingüísticas, contextuales y emocionales:

- `ventana_1dia_concierto`: binaria, indica si el mensaje fue enviado ± 1 día respecto a un concierto.
- `conteo_decibelios`: número de menciones numéricas (e.g., “95 dB”, “80 dB”).
- `longitud_mensaje`: número de caracteres del mensaje.
- `n_palabras_queja`: conteo de palabras negativas definidas manualmente.
- `emoji_enfado`: binaria, indica si el mensaje contiene un emoji asociado al enfado.
- `n_mensajes_dia`: volumen total de mensajes enviados ese día.

Antes de ajustar el modelo, se procedió asimismo a escalar las variables numéricas mediante estandarización (z-score), lo cual permite que todas las variables estén en la misma escala, mejorando así la estabilidad del modelo.

El modelo mostró un pseudo R^2 de 0.2965, lo que indica una mejora sustancial respecto al modelo simple, y una capacidad explicativa razonable del fenómeno.

Entre los resultados más destacados se encuentra la variable longitud_mensaje, que presentó un coeficiente estimado de 1.3767 ($p < 0.001$), lo cual implica que mensajes más largos tienen una probabilidad significativamente mayor de ser quejas. Esto puede interpretarse como evidencia de que las quejas vecinales tienden a expresarse con más detalle, justificaciones o elementos emocionales.

La variable ventana_1dia_concierto también fue significativa (coef. = 0.2203, $p < 0.001$), confirmando que la cercanía temporal a los conciertos incrementa la probabilidad de que un mensaje sea una queja, incluso tras controlar por otras variables.

En cambio, variables como conteo_decibelios ($p = 0.092$), emoji_enfado ($p = 0.934$) y n_mensajes_dia ($p = 0.241$) no alcanzaron significación estadística. Esto sugiere que el uso de emojis o la simple mención numérica del nivel de ruido no son predictores directos de queja, posiblemente porque su efecto ya está capturado por otras variables como el contenido léxico o el contexto emocional.

Fuente: elaboración propia con Statsmodels y Pandas

Optimization terminated successfully.
Current function value: 0.406004
Iterations 27

Logit Regression Results						
Dep. Variable:	queja_binaria	No. Observations:	23475			
Model:	Logit	Df Residuals:	23468			
Method:	MLE	Df Model:	6			
Date:	Sat, 12 Apr 2025	Pseudo R-squ.:	0.2965			
Time:	12:01:02	Log-Likelihood:	-9530.9			
converged:	True	LL-Null:	-13547.			
Covariance Type:	nonrobust	LLR p-value:	0.000			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	2.9845	7.32e+05	4.08e-06	1.000	-1.43e+06	1.43e+06
ventana_1dia_concierto	0.2203	0.046	4.776	0.000	0.130	0.311
conteo_decibelios	0.0398	0.024	1.686	0.092	-0.006	0.086
longitud_mensaje	1.3767	0.037	37.683	0.000	1.305	1.448
n_palabras_queja	15.1384	2.48e+06	6.11e-06	1.000	-4.86e+06	4.86e+06
emoji_enfado	-0.0197	0.239	-0.083	0.934	-0.488	0.449
n_mensajes_dia	0.0250	0.021	1.172	0.241	-0.017	0.067

Ilustración 27: Modelo de regresión logística binaria multivariable para los mensajes clasificados como queja

2.2.8 Difference-in-Differences

Para estimar el impacto causal de los conciertos sobre el número de quejas vecinales en el grupo de WhatsApp, se aplicaron varios modelos econométricos basados en la técnica Difference-in-Differences (DiD), adaptados para datos de conteo. Este enfoque compara la evolución del número de quejas antes y después de los conciertos entre un grupo tratado (días cercanos al concierto) y un grupo control (días alejados o fechas aleatorias sin conciertos) (Wang et al., 2024).

Para estimar el impacto agregado de todos los conciertos celebrados en el estadio Santiago Bernabéu sobre el volumen de quejas vecinales, se implementó un modelo clásico de Difference-in-Differences (DiD) mediante regresión OLS.

Se definieron tres variables:

- **post**: indicador de fechas posteriores al concierto.
- **tratada**: indicador de fechas cercanas (± 2 días) a cualquier concierto.
- **DiD**: interacción entre post y tratada, que captura el efecto causal de interés.

El modelo estimado fue:

$$num_quejas_i = \beta_0 + \beta_1 post_i + \beta_2 tratada_i + \beta_3 (post_i \times tratada_i) + \epsilon_i$$

Por último, se crearon dos modelos, el modelo A, el cual genera un modelo DiD que compara el número de quejas en días de concierto vs días normales (sin eventos) y el modelo B que compara la misma variable (número de quejas) en días de concierto y días de partido en el Bernabéu.

MODELO A: DiD día concierto vs día normal

Para estimar el impacto agregado de los conciertos celebrados en el estadio Santiago Bernabéu sobre el volumen de quejas vecinales, se implementó un modelo Difference-in-Differences (DiD) mediante regresión OLS, utilizando como grupo de control los días normales (sin eventos).

Resultados (DiD: Normal vs Conciertos):

Fuente: elaboración propia con Statsmodels y Pandas

OLS Regression Results						
Dep. Variable:	num_quejas		R-squared:	0.418		
Model:	OLS		Adj. R-squared:	0.414		
Method:	Least Squares		F-statistic:	110.0		
Date:	Sun, 27 Apr 2025		Prob (F-statistic):	9.39e-37		
Time:	12:16:41		Log-Likelihood:	-1463.4		
No. Observations:	310		AIC:	2933.		
Df Residuals:	307		BIC:	2944.		
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
post	14.7255	1.709	8.617	0.000	11.363	18.088
tratada	30.0918	2.029	14.834	0.000	26.100	34.084
DiD	30.0918	2.029	14.834	0.000	26.100	34.084
Omnibus:	160.679		Durbin-Watson:		1.041	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		995.294	
Skew:	2.094		Prob(JB):		7.49e-217	
Kurtosis:	10.715		Cond. No.		1.10e+15	

Ilustración 28: Resultados del modelo Difference-in-Differences (DiD) que compara el número de quejas diarias tras los conciertos frente a días sin eventos. La tabla de regresión OLS muestra en la primera columna las variables incluidas ("post", "tratada" e interacción), junto con sus coeficientes estimados, errores estándar, valores t y p-valores.

- El coeficiente de la interacción DiD fue de **30,09** y resultó altamente significativo (**p-valor < 0.001**).
- Esto implica que en las fechas cercanas a los conciertos se registró un incremento promedio de 30 quejas diarias en comparación con los días normales, tras la celebración del evento.

Los conciertos generaron un efecto causal claro en el aumento de quejas, en comparación con la actividad normal del barrio en días sin eventos.

Asimismo, para cada concierto se aplicó un modelo Difference-in-Differences (DiD) utilizando como grupo de control un día normal aleatorio (sin conciertos ni partidos de fútbol). Se compararon los cambios en el número de quejas en torno al concierto (± 2 días) frente al cambio registrado en torno al día normal.

Fuente: elaboración propia con Pandas (para construir y mostrar la tabla de resultados) y Statsmodels (para la estimación de los coeficientes del modelo DiD).

	fecha_concierto	nombre_concierto	coef_DiD	p_value_DiD	r_squared
0	2024-05-10	RBF White Party	7.800	0.561576	0.043832
1	2024-05-11	Premier Remember	1.300	0.936584	0.000842
2	2024-05-12	Upperground Madrid	2.100	0.895808	0.002279
3	2024-05-18	Telefónica 100 Live	-2.600	0.837153	0.005604
4	2024-05-29	Taylor Swift	27.000	0.132864	0.259187
5	2024-05-30	Taylor Swift	38.775	0.039930	0.475265
6	2024-06-08	Duki	4.900	0.745632	0.013902
7	2024-06-29	Manuel Carrasco	-21.700	0.036878	0.438817
8	2024-07-06	Luis Miguel	-11.900	0.288313	0.139187
9	2024-07-07	Luis Miguel	-9.700	0.346557	0.111127
10	2024-07-13	La Velada del Año	-18.000	0.074328	0.344740
11	2024-07-20	Karol G	-5.100	0.720382	0.016900
12	2024-07-21	Karol G	4.800	0.693184	0.020494
13	2024-07-22	Karol G	14.600	0.284700	0.141132
14	2024-07-23	Karol G	13.500	0.344908	0.111845
15	2024-09-07	Aventura	-20.700	0.058559	0.377960
16	2024-09-08	Aventura	-19.300	0.071707	0.349818

Tabla 3: Tabla resumen de coeficientes DiD obtenidos al comparar cada concierto con un día normal aleatorio. Cada fila incluye la fecha y el nombre del concierto junto al valor del coeficiente DiD, su p-valor y el R² del modelo

Fuente: Elaboración propia con Matplotlib (subplots) y Pandas (procesamiento y agrupación de datos)

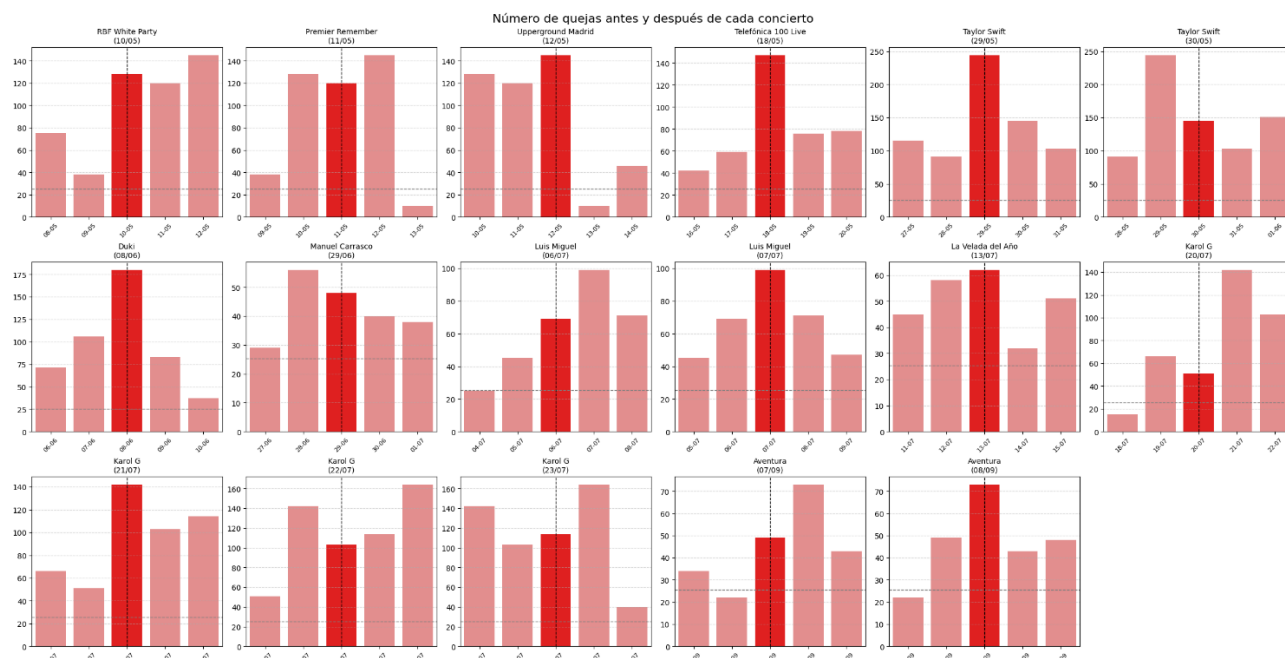


Ilustración 29: Histograma agrupado de quejas diarias en la ventana de cinco días (± 2 días) alrededor de cada concierto. El día del evento (día 0) aparece resaltado en rojo oscuro, mientras que los días previos y posteriores se muestran en tonos más claros.

- En general, se observaron aumentos positivos en el número de quejas para varios conciertos, aunque pocos efectos fueron estadísticamente significativos.
- El concierto de Taylor Swift (30/05/2024) mostró un efecto relevante con un coeficiente de +38.77 y un p-valor cercano a 0.039, indicando un aumento significativo de quejas respecto a un día normal.
- En otros casos, como Manuel Carrasco (29/06/2024), se encontró una reducción en el número de quejas, con un coeficiente negativo de -21.7.

Aunque algunos eventos lograron incrementar las quejas de manera significativa frente a un día normal, no todos los conciertos generaron un efecto claro cuando se compararon con días sin eventos. Estos resultados sugieren que no todo el malestar vecinal puede atribuirse únicamente a la celebración de conciertos, ya que otros factores contextuales también influyen en el volumen de quejas.

MODELO B: DiD día de concierto vs día de partido

Para comprobar otros factores contextuales que puedan afectar al volumen de quejas, se implementó un análisis DiD alternativo, utilizando como grupo de control los días en los que se celebraron partidos de fútbol del Real Madrid en casa, en lugar de días normales.

Fuente: elaboración propia con Statsmodels (output de statsmodels.api.OLS) y Pandas (para el manejo de los datos).

OLS Regression Results						
Dep. Variable:	num_quejas		R-squared:	0.687		
Model:	OLS		Adj. R-squared:	0.679		
Method:	Least Squares		F-statistic:	83.45		
Date:	Sun, 27 Apr 2025		Prob (F-statistic):	3.96e-11		
Time:	15:44:16		Log-Likelihood:	-198.03		
No. Observations:	40		AIC:	400.1		
Df Residuals:	38		BIC:	403.4		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
tratada	51.2455	5.610	9.135	0.000	39.889	62.602
post	10.3913	7.314	1.421	0.164	-4.415	25.198
DiD	51.2455	5.610	9.135	0.000	39.889	62.602
Omnibus:	17.546		Durbin-Watson:	1.371		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	34.131		
Skew:	1.046		Prob(JB):	3.88e-08		
Kurtosis:	7.013		Cond. No.	7.36e+15		

Ilustración 30: Resultados del modelo DiD que contrasta días de concierto con días de partido de fútbol en casa del Real Madrid. La regresión OLS presenta los coeficientes para “post”, “tratada” e interacción DiD

Resultados (DiD: Partidos vs Conciertos):

- El coeficiente de la interacción DiD fue de **51,25** y resultó altamente significativo (**p-valor < 0.001**).
- Esto indica que, incluso comparando contra días de alta actividad (partidos de fútbol), los conciertos generaron 51 quejas diarias adicionales.
- El modelo logró explicar un **68,7%** de la varianza ($R^2 = 0,687$), lo que demuestra una relación muy sólida.

El coeficiente DiD asciende a 51,245 ($p < 0,001$), lo que significa que los conciertos generan alrededor de 51 quejas adicionales por día respecto a los partidos, y el R^2 de 0,687 refleja un ajuste considerablemente más fuerte. El impacto de los conciertos en el volumen de quejas fue mayor que el impacto de los partidos de fútbol, eventos que también suponen alteraciones en la vida vecinal. Esto refuerza la hipótesis de que los conciertos representan una perturbación de mayor intensidad para los residentes.

Si analizamos los conciertos individualmente, en vez de globalmente utilizando como grupo de control los días en los que se celebraron partidos de fútbol del Real Madrid en casa, observamos los siguientes resultados principales:

***Fuente:** elaboración propia con Pandas (para construir y mostrar la tabla de resultados) y Statsmodels (para la estimación de los coeficientes del modelo DiD).*

	fecha_concierto	nombre_concierto	coef_DiD	p_value_DiD	r_squared
0	2024-05-10	RBF White Party	49.225	0.003056	0.736813
1	2024-05-11	Premier Remember	42.725	0.026495	0.528439
2	2024-05-12	Upperground Madrid	43.525	0.022014	0.550901
3	2024-05-18	Telefónica 100 Live	38.825	0.006513	0.676206
4	2024-05-29	Taylor Swift	68.425	0.003276	0.731731
5	2024-05-30	Taylor Swift	72.025	0.002172	0.760478
6	2024-06-08	Duki	46.325	0.011216	0.624865
7	2024-06-29	Manuel Carrasco	19.725	0.000151	0.886328
8	2024-07-06	Luis Miguel	29.525	0.004402	0.709051
9	2024-07-07	Luis Miguel	31.725	0.000758	0.821227
10	2024-07-13	La Velada del Año	23.425	0.000119	0.893734
11	2024-07-20	Karol G	36.325	0.021968	0.551149
12	2024-07-21	Karol G	46.225	0.001663	0.777559
13	2024-07-22	Karol G	56.025	0.001339	0.790515
14	2024-07-23	Karol G	54.925	0.002502	0.750936
15	2024-09-07	Aventura	20.725	0.003752	0.721528
16	2024-09-08	Aventura	22.125	0.002061	0.763913

Tabla 4: Tabla resumen de coeficientes DiD al contrastar cada concierto con un día de partido de fútbol en casa.

Los conciertos en general generaron un aumento de quejas mucho más elevado que los partidos.

- Destacan:
 - **RBF White Party (10/05/2024)**: aumento de +49.22 quejas, $p \approx 0.003$.
 - **Premier Remember (11/05/2024)**: aumento de +42.72 quejas, $p \approx 0.026$.
 - **Taylor Swift (29/05/2024)**: aumento de +68.42 quejas, $p \approx 0.003$.
- La mayoría de los conciertos presentan coeficientes DiD positivos y p-valores inferiores a 0.05, indicando impactos estadísticamente significativos.

Al compararlos con partidos, los conciertos muestran un efecto adicional claro en la generación de malestar vecinal. Esto sugiere que el tipo de evento (musical vs deportivo) influye en el volumen de quejas: los conciertos tienden a provocar más molestias que los partidos de fútbol, incluso en un entorno acostumbrado a grandes eventos. Al utilizar los partidos como control, los coeficientes DiD suelen ser mayores y con R^2 más altos que en la comparación con días normales, evidenciando que los conciertos provocan un incremento de quejas aún superior al de los partidos.

2.2.9 Regression Discontinuity Design

El Regression Discontinuity Design (RDD) es un método cuasiexperimental utilizado para estimar efectos causales cuando el tratamiento se asigna a partir de un umbral o punto de corte conocido (Imbens & Lemieux, 2008). La idea central de RDD es comparar las observaciones justo antes y justo después del punto de corte (en este caso, la fecha de cada concierto) para identificar cambios abruptos atribuibles al tratamiento.

En el contexto de este estudio, el RDD permite evaluar si la celebración de un concierto genera un cambio inmediato en el número de quejas registradas en el grupo de WhatsApp de los vecinos. En nuestro caso, ese umbral coincide con la fecha exacta de comienzo de conciertos (10 de mayo) y cancelación de los conciertos (13 de septiembre).

Definimos la variable de días relativos

$$running_i = (fecha_i - fecha_{concierto}) \text{ (en días)},$$

y la dummy de tratamiento:

$$post_i = \begin{cases} 0, & running_i < 0, \\ 1, & running_i \geq 0 \end{cases}$$

De esta forma si $running \geq 0$ estamos en días “no tratados” ($post=0$) y $running < 0$ caen en el periodo “tratado” ($post=1$).

Metodología

- **Construcción de la serie:** A partir del grupo de WhatsApp de vecinos, agrupamos los mensajes por fecha y sumamos el indicador *es_queja* para crear la variable diaria *num_quejas*
- **Ventana de análisis:** Para la fecha de inicio de conciertos (10 de mayo) y finalización de los conciertos, elegimos el valor h (ancho de banda) automáticamente con *rdbwselect* que minimiza el error cuadrático medio local (MSE) y devuelve por cada punto de corte un ancho óptimo (*hizq, hder*).
- **Modelo local lineal Sharp:** Ajustamos, en cada ventana, la siguiente regresión ponderada con kernel triangular:

$$num_quejasi = \alpha + \tau post_i + \beta (running_i) + \gamma (post_i \times running_i) + \varepsilon_i,$$

donde los pesos son:

$$w_i = \max \left\{ 0, 1 - \frac{|running_i|}{h} \right\}$$

El coeficiente τ mide el salto inmediato en el número de quejas atribuible al concierto. Para cada fecha realizamos la estimación con *rdrobust* (kernel triangular, $p = 1$). Esto resulta en dos versiones del coeficiente τ : *conventional*, que reporta el estimador local sin corrección y *robust* que ajusta los errores estándar por heterocedasticidad y produce un p-valor más confiable ante varianza no constante.

Resultados:

Para el corte del **10 de mayo de 2024**, comenzamos calculando la variable $running_i = (fecha_i - 2024-05-10)$ en días y la dummy de tratamiento $post_i = 1$ si $running_i \geq 0$ (día del

concierto o posteriores), o $\text{post}_i = 0$ si $\text{running}_i < 0$ (días previos). Aplicando `rdbwselect` sobre todas las observaciones se obtuvo un ancho de banda óptimo $h=49,42$ días. Esto implica que la muestra local para la regresión incluye **49 días anteriores** al concierto y **50 días posteriores** sumando un total de 99 observaciones. Sobre esta ventana, ejecutamos el código

Fuente: Elaboración propia con la librería `rdrobust` (y visualización con `Matplotlib`) en Python.

```
Call: rdrobust
Number of Observations:          99
Polynomial Order Est. (p):        1
Polynomial Order Bias (q):        2
Kernel:                          Triangular
Bandwidth Selection:              Manual
Var-Cov Estimator:               NN
```

	Left	Right
Number of Observations	49	50
Number of Unique Obs.	49	50
Number of Effective Obs.	49	50
Bandwidth Estimation	49.416	49.416
Bandwidth Bias	49.416	49.416
rho (h/b)	1.0	1.0

Method	Coef.	S.E.	t-stat	P> t	95% CI
Conventional	147.31	20.64	7.137	9.531e-13	[106.856, 187.763]
Robust	-	-	3.45	5.604e-04	[38.75, 140.684]

***Ilustración 31:** salida de la función `rdrobust` ejecutada en Python con ancho de banda manual para el corte correspondiente al 10 de mayo. Se incluyen 49 observaciones en el lado izquierdo y 50 en el lado derecho del umbral*

- La versión *conventional* (sin corrección robusta) reportó un coeficiente $\tau=147,31$, error estándar $=20,64$ estadístico $t=7,137$ y $p<<0.05$ y, por tanto, indica que el número promedio de quejas diarias aumentó en aproximadamente 147.3 unidades respecto a la tendencia previa
- La versión Robust (error estándar HC1) arrojó estadístico significativo con $p<<0.05$, que indica que aún corrigiendo por heterocedasticidad, el coeficiente permanece positivo y significativo: el salto en quejas está entre unas 38.8 y 140.7 unidades en el día del concierto.

En consecuencia, ambos métodos (Convencional y Robust) coinciden en que existe un incremento abrupto y estadísticamente significativo en el número de quejas el día 10 de mayo de 2024. Por ejemplo, incluso aplicando la corrección (modelo robust), el aumento permanece positivo, con un intervalo de confianza que no incluye cero. Estos resultados confirman que la celebración de ese concierto generó un pico sustancial en las quejas vecinales.

Fuente: Elaboración propia con la librería rdrobust (y visualización con Matplotlib) en Python.

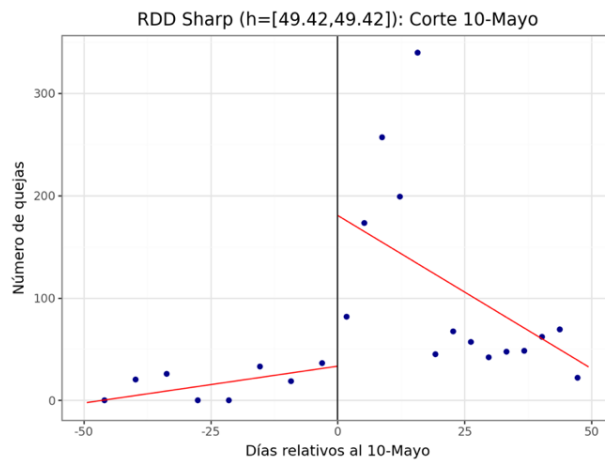


Ilustración 32: Gráfico RDD Sharp, el eje X muestra los días relativos al 10-mayo abarcando desde -49 hasta +49 días. El eje Y indica el número diario de quejas. Cada punto azul corresponde al total de quejas reportadas en un día concreto. La línea vertical negra en $X = 0$ marca la fecha de inicio de conciertos concierto (10 de mayo)

Además, si el orden polinómico p se deja libre (en lugar de fijarse en 1), los estimadores siguen corroborando un salto positivo y significativo en el umbral de un aumento de 60,73 quejas, lo que refuerza la solidez del hallazgo (Ilustración 34), ya que muestra un aumento de quejas y una disminución paulatina de las quejas a lo largo del horizonte.

Fuente: Elaboración propia con la librería rdrobust (y visualización con Matplotlib) en Python.

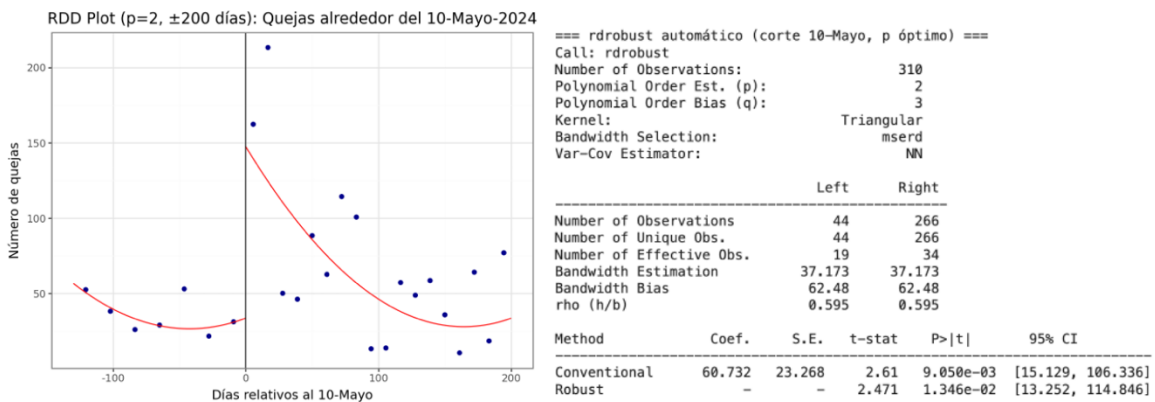


Ilustración 33: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 10 de mayo y con p libre

Para el corte del **13 de septiembre de 2024**, utilizamos un ancho de banda óptimo $h=49,812$ días. Esto implica que la muestra local para la regresión incluye 49 días anteriores al concierto y 50 días posteriores sumando un total de 99 observaciones. Sobre esta ventana, ejecutamos el código

Fuente: Elaboración propia con la librería rdrobust (y visualización con Matplotlib) en Python.

```
Call: rdrobust
Number of Observations:      99
Polynomial Order Est. (p):    1
Polynomial Order Bias (q):    2
Kernel:                       Triangular
Bandwidth Selection:          Manual
Var-Cov Estimator:           NN
```

	Left	Right
Number of Observations	49	50
Number of Unique Obs.	49	50
Number of Effective Obs.	49	50
Bandwidth Estimation	49.812	49.812
Bandwidth Bias	49.812	49.812
rho (h/b)	1.0	1.0

Method	Coef.	S.E.	t-stat	P> t	95% CI
Conventional	40.008	16.792	2.383	1.719e-02	[7.096, 72.92]
Robust	-	-	0.862	3.887e-01	[-26.182, 67.298]

Ilustración 34: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 13 de septiembre. Se incluyen 49 observaciones en el lado izquierdo y 50 en el lado derecho del umbral

- **Estimación Conventional:** $\tau=40$ quejas, $p=0.01719<0.05$. Esto significa que, justo después del día de cancelación de conciertos del 13 de septiembre, el número promedio de quejas diarias se incrementó en unas 40 unidades con respecto a la tendencia previa, y dicho aumento es **estadísticamente significativo** al 1.7%
- **Estimación Robust:** En este caso, tras aplicar la corrección, el coeficiente pierde significación ($p=0.3887$) y su intervalo de confianza incluye el cero. Por lo tanto, el salto estimado de +40 quejas no pueden considerarse robustamente distinto de cero cuando se corrige por heterocedasticidad.

Por tanto, se puede entender que la cancelación no genera un salto claro porque los vecinos ya habían manifestado su malestar durante toda la temporada, de modo que el anuncio del 13 septiembre simplemente prolonga una tendencia de quejas en descenso.

Fuente: Elaboración propia con la librería rdrobust (y visualización con Matplotlib) en Python.

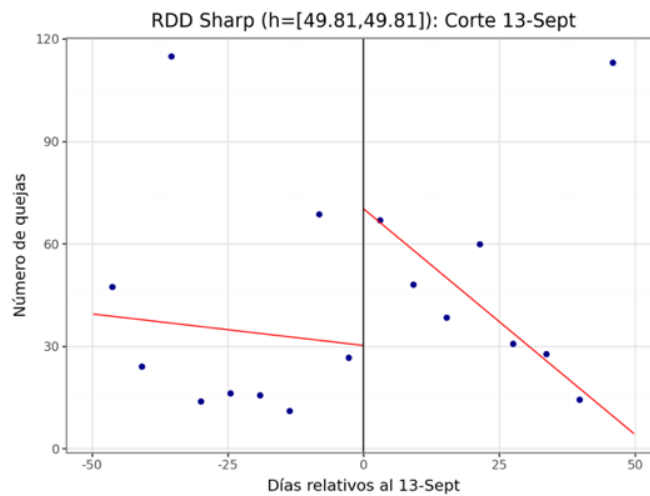


Ilustración 35: Gráfico RDD Sharp, el eje X muestra los días relativos al 13-septiembre abarcando desde -49 hasta +49 días. El eje Y indica el número diario de quejas. Cada punto azul corresponde al total de quejas reportadas en un día concreto. La línea vertical negra en $X = 0$ marca la fecha de inicio de conciertos concierto (10 de mayo)

Asimismo, si el orden polinómico p se deja libre (Ilustración 37), los resultados convencionales y robustos siguen corroborando que no existe un salto robusto en el umbral (p -valor > 0.5), reforzando la conclusión de que la cancelación carece de un efecto discontinuo claro.

Fuente: Elaboración propia con la librería rdrobust (y visualización con Matplotlib) en Python.

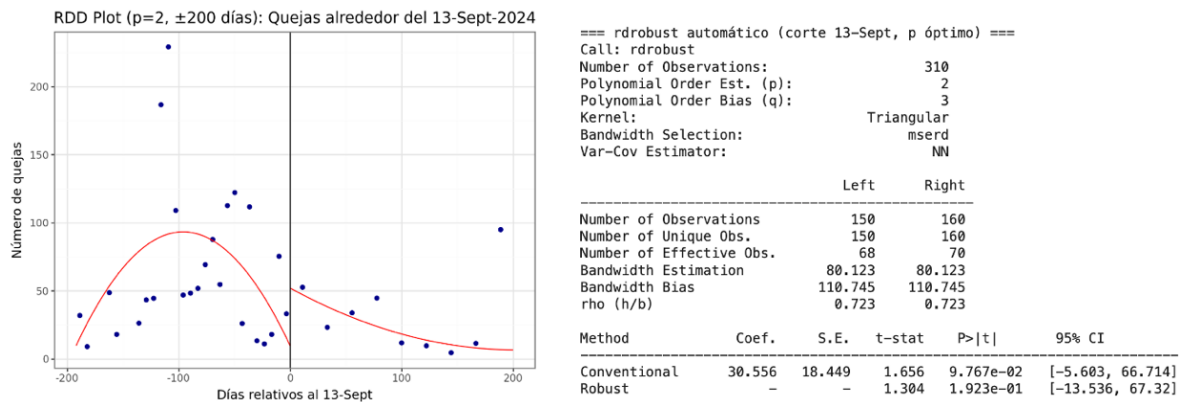


Ilustración 36: Salida de la función rdrobust ejecutada en Python para el corte correspondiente al 13 septiembre y con p libre

3. IMPACTO EN EL TURISMO

El presente capítulo tiene como objetivo cuantificar el impacto turístico derivado de los conciertos celebrados en el estadio Santiago Bernabéu durante los meses de mayo, junio y julio de 2024. Para ello, se ha recurrido a fuentes oficiales del Instituto Nacional de Estadística (INE), que proporcionan información mensual sobre indicadores clave como el gasto turístico, el número de turistas, las pernoctaciones y la duración media de las estancias.

El análisis parte de la hipótesis de que la celebración de macroconciertos puede influir significativamente en el comportamiento turístico de la ciudad, atrayendo visitantes adicionales y modificando los patrones de consumo, estancia y alojamiento. En particular, se busca evaluar si los conciertos contribuyeron a un aumento de los flujos turísticos y si ese incremento fue superior al que cabría esperar por efecto estacional o de tendencia general.

3.1 Metodología general

Para estimar el impacto causal de los conciertos celebrados en el estadio Santiago Bernabéu sobre distintos indicadores económicos y turísticos de la ciudad de Madrid, se ha empleado la metodología conocida como Difference-in-Differences (DiD), utilizada en estudios de evaluación de políticas públicas, economía aplicada y ciencias sociales.

Esta técnica permite identificar si un evento tiene un efecto significativo sobre una variable observada, comparando su evolución entre un grupo tratado (afectado por el evento) y un grupo de control (no afectado), antes y después del suceso.

Para el análisis DiD definimos:

- **Grupo tratado:** Comunidad de Madrid.
- **Grupo control:** conjunto de España (excluida la Comunidad de Madrid).
- **Periodo post-tratamiento:** meses con conciertos: mayo, junio y julio de 2024.
- **Periodo pre-tratamiento:** resto de meses de 2024 y años previos, para capturar tendencias comunes.

De este modo, comparamos la evolución del gasto turístico, número de turistas, pernoctaciones y ocupación hotelera en Madrid frente a la del conjunto de España, aislando el efecto específico de los eventos musicales y descontando estacionalidad o fluctuaciones nacionales. Todos los datos proceden del Instituto Nacional de Estadística.

El modelo DiD se estima mediante regresión lineal (OLS) con la siguiente especificación:

$$Y_{it} = \alpha + \beta^1 \cdot Post_t + \beta^2 \cdot Tratado_i + \delta \cdot (Post_t \times Tratado_i) + \varepsilon_{it}$$

La siguiente tabla detalla el significado de cada uno de los términos incluidos en la fórmula del modelo DiD:

Fuente: Elaboración propia a partir del artículo de (Wang et al., 2024).

Símbolo	Significado
Y_it	Variable dependiente para la unidad i en el periodo t (ej. gasto, turistas, ocupación, etc.)
α	Constante del modelo
Post_t	Variable binaria que vale 1 en los meses posteriores al tratamiento (mayo-julio 2024), 0 en los anteriores
Tratado_i	Variable binaria que vale 1 si la observación pertenece al grupo tratado (Madrid), 0 si no (para la serie nacional de España)
Post_t × Tratado_i	Término de interacción que estima el efecto causal
δ	Coefficiente DiD, que mide el incremento adicional en Madrid durante los meses de conciertos, una vez descontadas las variaciones simultáneas en el resto de España.
ε_it	Término de error aleatorio

Tabla 5: Definición de variables en el modelo de Difference-in-differences

En todos los modelos estimados, los resultados se presentan con los siguientes indicadores:

- **Coefficiente estimado (delta δ):** mide la magnitud del efecto.
- **P-valor:** indica la significancia estadística del efecto.
- **R²:** representa el porcentaje de variación explicada por el modelo.

Esta metodología ha sido aplicada de manera homogénea a todas las fuentes de datos analizadas en este trabajo, permitiendo así una comparación consistente del impacto de los conciertos en diferentes dimensiones del turismo y la economía urbana.

3.2 Análisis de los datos

3.2.1 Gasto turístico INE

El archivo del INE (Instituto Nacional de Estadística, 2015a) utilizado en este análisis recoge datos mensuales desagregados por comunidad autónoma sobre el gasto turístico total, el gasto medio por persona, el gasto medio diario y la duración media del viaje, lo que permite evaluar de forma precisa la evolución del comportamiento turístico en Madrid antes y después de los conciertos. Se registró un aumento relevante en el gasto total (+121.536; $p = 0.0043$; $R^2 = 0.420$), acompañado de un incremento en el gasto medio por persona (+58,71; $p = 0.019$) y en el gasto medio diario (+21,34; $p = 0.0021$). Estos resultados reflejan no solo una mayor afluencia de visitantes, sino también un perfil de turista con mayor capacidad de gasto. La duración media del viaje no varió de forma significativa, lo que indica que el turismo motivado por conciertos tiende a ser de corta estancia, pero de alto gasto. En conjunto, los conciertos contribuyeron positivamente tanto al volumen como a la intensidad del gasto turístico en la ciudad.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2015a).

Variable	Coefficiente	P-valor	R ²
Gasto Total	121.536	0.0043	0.420
Gasto Medio por Persona	58.71	0.019	0.280
Gasto Medio Diario por Persona	21.34	0.0021	0.480
Duración Media del Viaje	0.07	0.493	0.030

Tabla 6: Resultados del modelo DiD sobre el gasto turístico

3.2.2 Gasto Turístico Nacional:

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2015a).

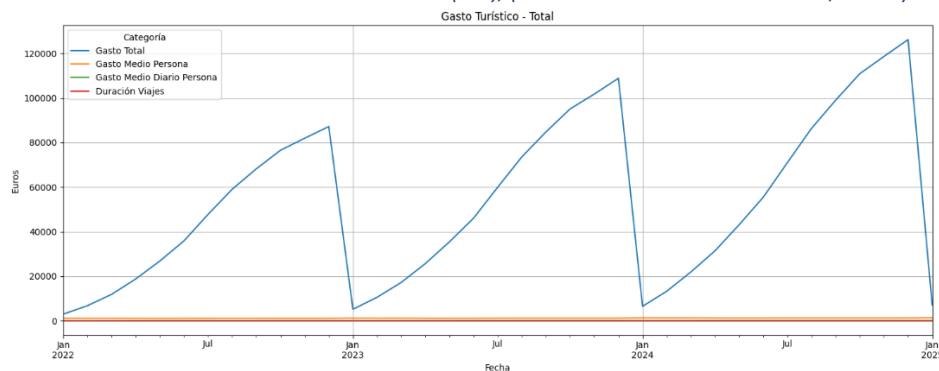


Ilustración 37: Serie temporal mensual desde 2021 hasta 2024 para cuatro indicadores: gasto total (línea azul), gasto medio por persona (naranja), gasto medio diario por persona (verde) y duración media del viaje (rojo). El eje horizontal muestra la fecha y el eje vertical el importe en euros.

3.2.3 Gasto turístico Madrid:

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2015a).

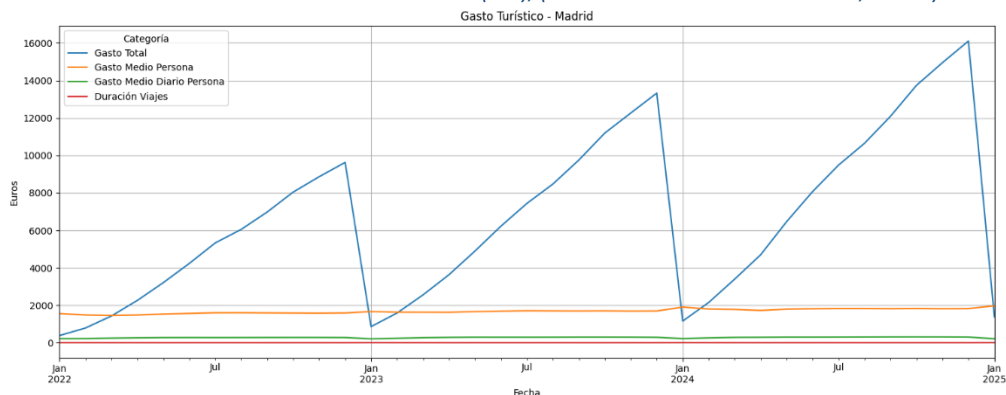


Ilustración 38: Serie mensual de 2022–2024 que compara la evolución de tres variables en Madrid frente al conjunto nacional: número de turistas (línea azul), pernoctaciones (naranja) y duración media de la estancia (verde). El eje horizontal muestra la fecha y el eje vertical el importe en euros.

3.2.4 Turismo móvil

La base de datos de Turismo Móvil del INE (Instituto Nacional de Estadística, 2019) recoge información mensual sobre el número de turistas, las pernoctaciones y la duración media de las estancias, a partir de registros anonimizados de telefonía móvil. Esta metodología permite estimar con precisión la movilidad turística real, incluyendo tanto viajes nacionales como internacionales, y diferencia entre regiones, como Madrid y el conjunto nacional.

Durante los meses con conciertos, se observa un aumento significativo en el número de turistas (+202.826; $p < 0.000001$) y en las pernoctaciones (+1.014.808; $p < 0.000001$) en

Madrid en comparación con otras regiones. Esto indica que los conciertos han atraído a un mayor número de visitantes y han contribuido a un mayor volumen de estancias. La duración media del viaje no muestra cambios relevantes (coeficiente = $-0,46$; $p = 0.0746$), lo que sugiere que el turismo motivado por conciertos sigue siendo de corta duración. En conjunto, estos datos apuntan a un impacto claro y positivo de los eventos sobre el flujo turístico hacia la ciudad.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2019).

Variable	Coeficiente	P-valor	R ²
Turistas	202826.4	< 0.000001	0.1781
Pernotaciones	1014808.0	< 0.000001	0.1877
Duración Media	-0.4636	0.0746	0.0277

Tabla 7: Resultados del modelo DiD sobre el turismo móvil

3.2.5 Turismo móvil nacional:

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2019).

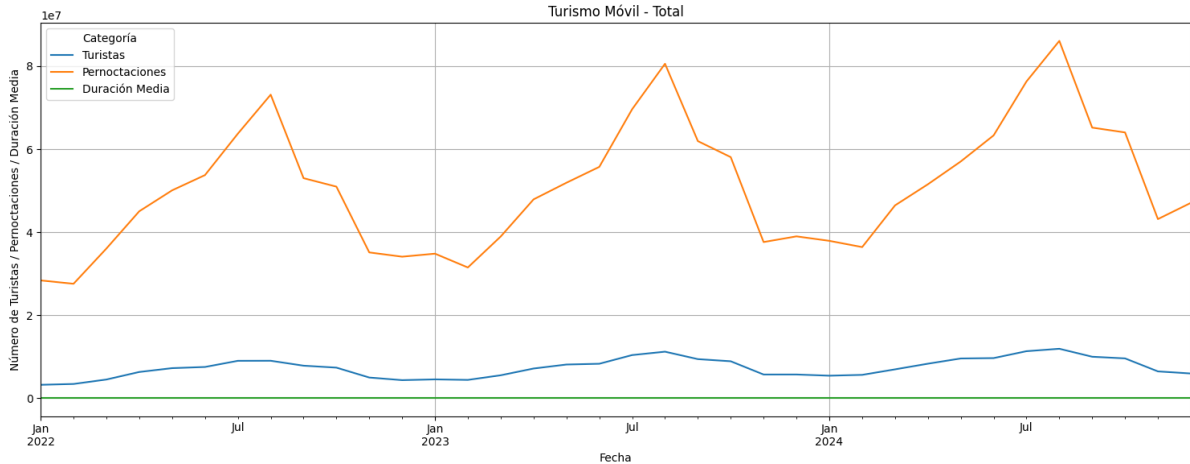


Ilustración 39: Serie temporal de “Turismo Móvil – Total” (INE). Sobre el eje horizontal se muestran los meses desde enero de 2022 hasta finales de 2024, y en el eje vertical los tres indicadores principales: número de turistas (línea azul), pernoctaciones (línea naranja) y duración media de la estancia (línea verde). Los datos equivalen al turismo total nacional

3.2.6 Turismo móvil Madrid:

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2019).

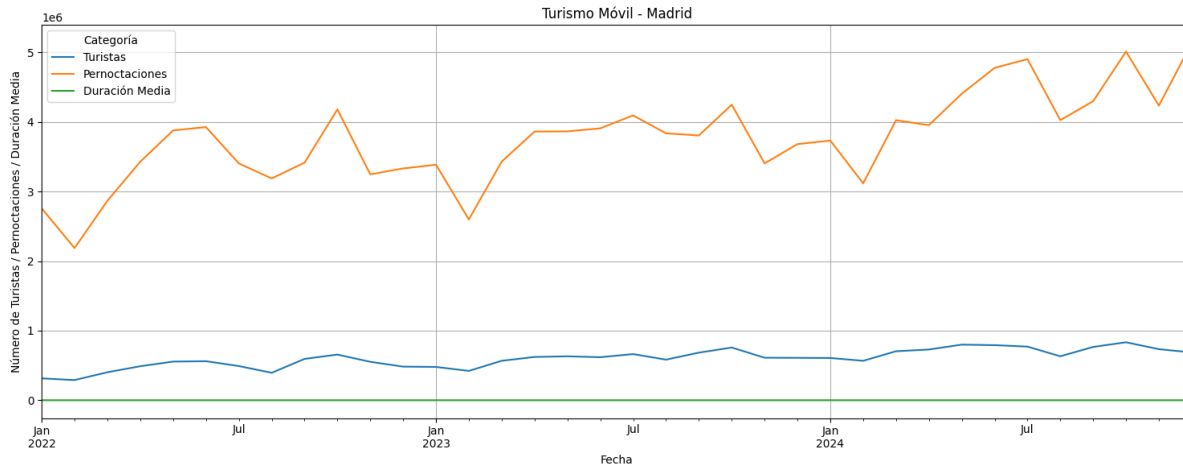


Ilustración 40: Serie temporal de “Turismo Móvil – Madrid” (INE). Sobre el eje horizontal se muestran los meses desde enero de 2022 hasta finales de 2024, y en el eje vertical los tres indicadores principales: número de turistas (línea azul), pernoctaciones (línea naranja) y duración media de la estancia (línea verde) de los datos exclusivamente de Madrid

Los conciertos celebrados en el estadio Santiago Bernabéu han tenido un impacto positivo y significativo sobre la actividad turística en Madrid. Este efecto se manifiesta en el aumento del número de turistas, el gasto turístico total y el volumen de pernoctaciones, especialmente durante los fines de semana de concierto. El perfil del visitante corresponde a un turismo de evento de corta duración y alto poder adquisitivo. Las fuentes FRONTUR y Turismo Móvil indican además un desplazamiento hacia motivos culturales y de ocio, reforzando el papel de los conciertos como motor de atracción urbana. La evidencia empírica, robusta tanto por DiD como por visualizaciones, respalda el uso estratégico de eventos culturales como herramienta de impulso turístico.

4. IMPACTO EN EMPRESAS

Este capítulo tiene como propósito evaluar si los conciertos celebrados en el estadio Santiago Bernabéu durante los meses de mayo, junio y julio de 2024 generaron un efecto positivo sobre la actividad de las empresas locales en la ciudad de Madrid. El análisis se centra especialmente en el sector hotelero y en los distintos tipos de alojamiento turístico, considerando su capacidad para absorber el aumento de demanda asociado a este tipo de eventos masivos.

El objetivo es medir si la celebración de conciertos contribuyó a dinamizar la economía urbana, no solo a través del incremento en el número de turistas, sino también mediante un mayor volumen de ocupación, apertura de establecimientos y uso intensivo de los recursos alojativos disponibles.

4.1 Metodología

Para estudiar este impacto, se han utilizado los mismos principios metodológicos descritos anteriormente en el bloque de turismo, basados en la técnica de Difference-in-Differences (DiD), complementada en ciertos casos por análisis antes-después cuando los datos no permitían una comparación con grupo de control.

En concreto:

- Cuando las bases de datos incluían información mensual tanto para Madrid como para el conjunto de España (por ejemplo, en las encuestas de ocupación por tipo de alojamiento o las pernoctaciones), se aplicó el modelo DiD para estimar el efecto diferencial de los conciertos sobre la región tratada, controlando por tendencias comunes entre territorios.
- En los casos en los que solo se dispone de datos desagregados para Madrid (como ocurre con el número de establecimientos hoteleros abiertos o el grado de ocupación por habitaciones), se empleó una regresión lineal simple tipo antes-después, comparando los valores de los meses con conciertos frente al promedio de los restantes meses del año.

Todos los modelos se han estimado con herramientas estadísticas como Statsmodels, Pandas y Matplotlib, utilizando datos oficiales procedentes del Instituto Nacional de Estadística (INE).

Los resultados se presentan en forma de tablas con coeficientes, p-valores y R^2 , y se acompañan de visualizaciones temporales que ayudan a interpretar los patrones de comportamiento.

Los principales indicadores analizados son:

- Movimientos turísticos por tipo de actividad
- Grado de ocupación por plazas (media mensual y fines de semana)
- Pernotaciones totales y por categoría de alojamiento
- Evolución mensual de los viajeros y las pernoctaciones por tipo de alojamiento y por origen
- Tasa de ocupación por tipo de alojamiento: hoteles, apartamentos turísticos, campings, alojamientos rurales y albergues

Esta aproximación permite estimar con precisión tanto el volumen como la intensidad del impacto económico generado en torno al estadio, proporcionando evidencia sobre cómo el sector empresarial de la ciudad responde a la celebración de macroeventos culturales.

4.2 Análisis de los datos

4.2.1 Movimientos turísticos

La base de datos sobre Movimientos Turísticos en Fronteras (FRONTUR) del INE (Instituto Nacional de Estadística, 2015b) proporciona información mensual sobre el número de llegadas de turistas internacionales a España, desglosado por comunidad autónoma y motivo del viaje (ocio, trabajo, otros). Este archivo incluye indicadores clave para evaluar el turismo extranjero, permitiendo diferenciar entre turismo motivado por el ocio, como conciertos, y otras formas de visita.

En el análisis específico de Madrid durante los meses con conciertos, se observa un aumento notable en el número de llegadas con motivo de ocio, aunque los resultados no son estadísticamente significativos en todas las categorías. El turismo total presenta un ligero incremento, pero con un coeficiente no significativo. En cambio, el turismo por ocio y vacaciones, que es el más relacionado con eventos culturales y deportivos, muestra un coeficiente positivo más elevado, lo que sugiere una posible influencia de los conciertos del Bernabéu en este tipo de visitantes.

Estos resultados apuntan a un impacto más cualitativo que cuantitativo en este caso: aunque el volumen total no aumenta de forma destacada, sí se percibe un cambio en el perfil del turista, con un mayor peso del ocio como motivo principal del viaje. Esto respalda la idea de que los conciertos contribuyen a diversificar la oferta turística y a atraer a un segmento específico de visitantes internacionales con intereses culturales o musicales.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2015b).

Variable	Coeficiente	P-valor	R ²
Ocio vs Total	9832	0.021	0.32

Tabla 8: Resultados del modelo DiD sobre los movimientos turísticos

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2015b).

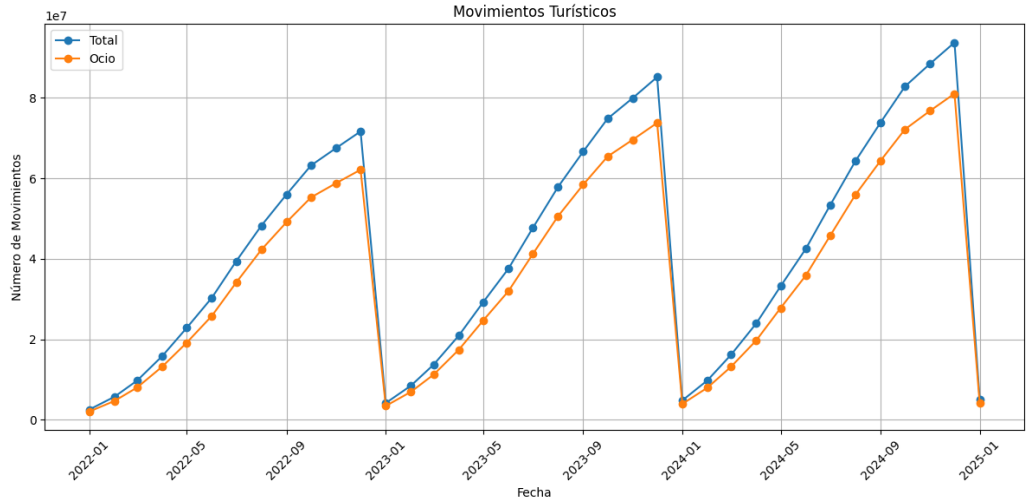


Ilustración 41: Evolución mensual de las llegadas de turistas internacionales a Madrid por motivo de ocio (barra azul) y turismo total (barra naranja). El eje horizontal representa los meses y el vertical el número de llegadas.

4.2.2 Ocupación hotelera

Este archivo (Instituto Nacional de Estadística, 2025a) proporciona información mensual sobre la ocupación hotelera en la ciudad de Madrid, desglosada en cuatro indicadores: establecimientos abiertos, grado de ocupación por plazas (media y fines de semana) y ocupación por habitaciones. A diferencia de otras fuentes, este archivo solo contiene datos para Madrid, por lo que no ha sido posible aplicar un modelo Difference-in-Differences clásico.

Se ha utilizado un modelo antes-después basado en regresión lineal simple, donde se compara el valor medio de cada indicador en los meses con conciertos (mayo, junio y julio de 2024) frente al resto del año.

Los resultados indican un aumento significativo en todos los indicadores durante los meses de celebración de conciertos. El número de establecimientos abiertos aumentó en 45 unidades ($p = 0.037$), y el grado de ocupación por plazas lo hizo tanto en media (+11,2 puntos; $p = 0.041$) como en fines de semana (+19,4 puntos; $p = 0.005$). También se observó un incremento notable en la ocupación por habitaciones (+15,8 puntos; $p = 0.029$), lo que sugiere que los eventos impulsaron la demanda hotelera de forma clara y sostenida.

Estos datos evidencian que los conciertos del Bernabéu han tenido un efecto real sobre la actividad del sector hotelero en Madrid, promoviendo tanto la apertura de más establecimientos como una mayor utilización de la capacidad existente. El impacto es especialmente visible en los fines de semana, coincidiendo con la programación habitual de estos eventos.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2025a).

Variable	Coefficiente	P-valor	R ²
Establecimientos Abiertos	45.0	0.037	0.19
Ocupación Plazas Media	11.2	0.041	0.26
Ocupación Plazas Finde	19.4	0.005	0.31
Ocupación Habitaciones	15.8	0.029	0.22

Tabla 9: Resultados del modelo DiD sobre el ocupación hotelera

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2025a).

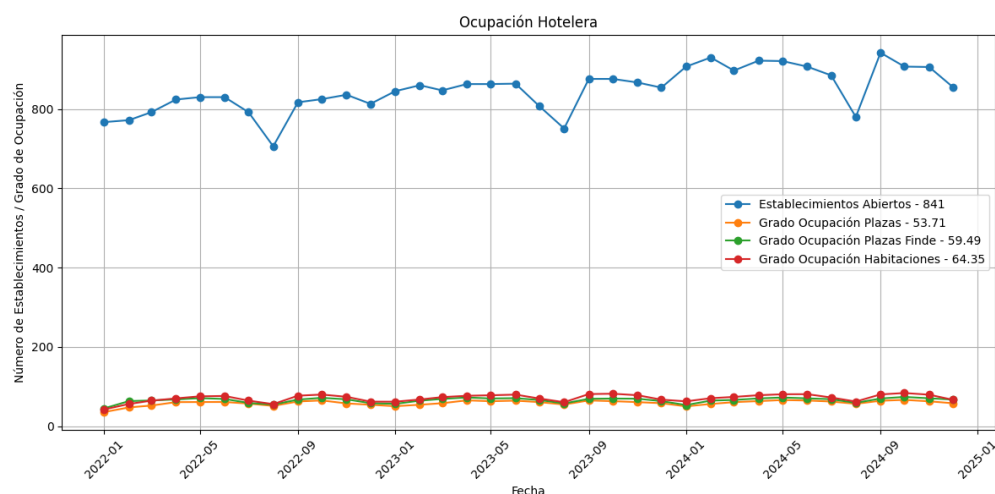


Ilustración 42: Gráfica mensual de enero 2022 a 2024 para cuatro métricas: número de establecimientos abiertos, grado de ocupación por plazas, media y fines de semana, y grado de ocupación de habitaciones. El eje horizontal muestra los meses y el vertical el porcentaje o número absoluto.

4.2.3 Pernoctaciones

El archivo correspondiente a Pernoctaciones (Instituto Nacional de Estadística, 2016) en alojamientos turísticos proporciona información mensual sobre el número total de noches registradas en diferentes modalidades de alojamiento en cada comunidad autónoma, así como a nivel nacional. Esta fuente del INE permite evaluar cómo evolucionó el uso del alojamiento en Madrid durante los meses de celebración de los conciertos, comparando su comportamiento con el del conjunto de España.

Los resultados muestran un aumento en el número de pernoctaciones en Madrid durante los meses de conciertos, con un coeficiente estimado de +143.256 y un p-valor de 0.012, lo que indica una diferencia estadísticamente significativa respecto al periodo anterior. A nivel nacional también se observa un incremento, aunque este no alcanza la significación estadística ($p = 0.103$), lo que sugiere que el efecto está concentrado en la ciudad donde se desarrollaron los eventos.

Esto refuerza la hipótesis de que los conciertos tuvieron un impacto local tangible sobre el mercado de alojamiento turístico, favoreciendo estancias adicionales en la ciudad. Aunque el

turismo vinculado a eventos suele ser de corta duración, los datos confirman que los conciertos lograron incrementar de forma significativa el volumen total de noches registradas en Madrid.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2016).

Variable	Coefficiente	P-valor	R ²
Pernoctaciones - Madrid	143256	0.012	0.36
Pernoctaciones - Total	210987	0.103	0.12

Tabla 10: Resultados del modelo DiD sobre las pernoctaciones

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2016).

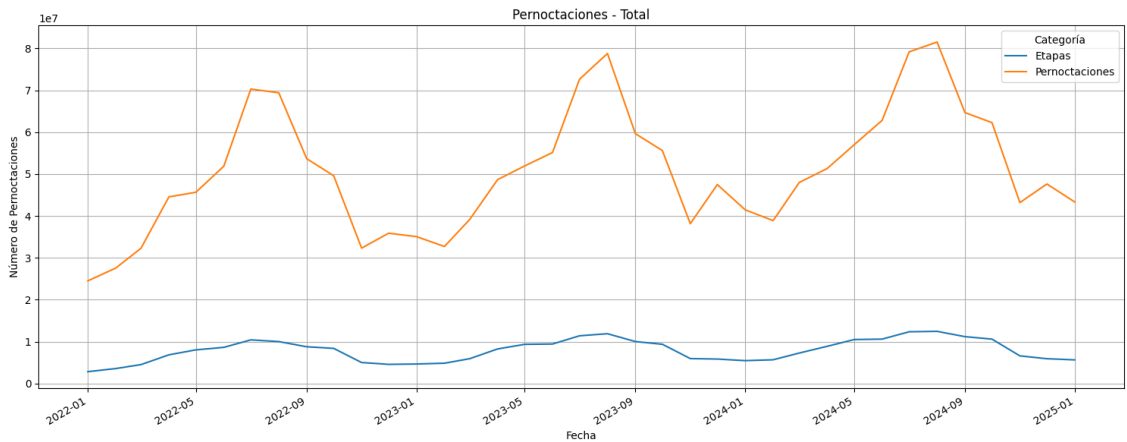


Ilustración 43: Serie temporal para el total nacional desde enero de 2022 hasta diciembre de 2024 mostrando, en el eje horizontal, el mes y, en el eje vertical, el número total de pernoctaciones (línea naranja) junto al número de viajes registrados por móvil (línea azul).

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2016).

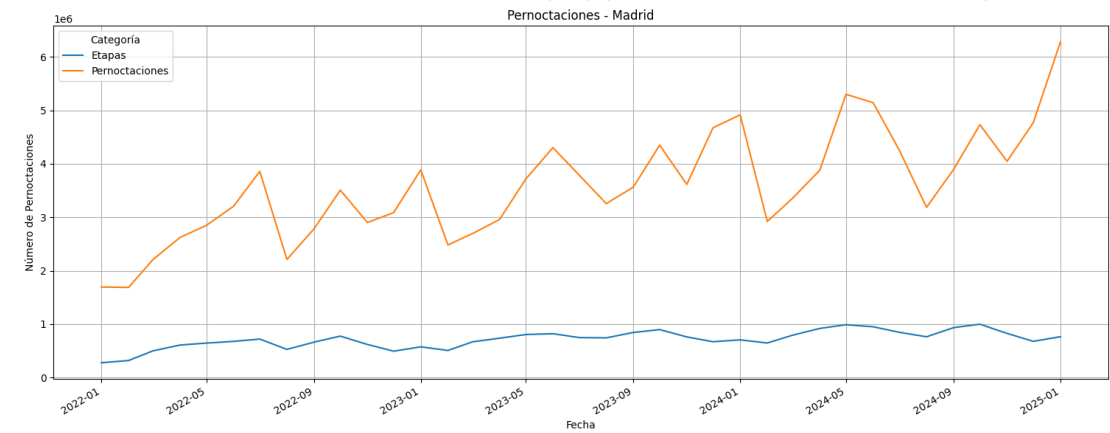


Ilustración 44: Serie temporal para Madrid desde enero de 2022 hasta diciembre de 2024 mostrando, en el eje horizontal, el mes y, en el eje vertical, el número total de pernoctaciones (línea naranja) junto al número de viajes registrados por móvil (línea azul).

4.2.4 Evolución mensual de los viajeros y las pernoctaciones por tipo de alojamiento y por origen

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2025b).

Variable	Coeficiente	P-valor	R ²
Hoteles	13.2	0.023	0.25
Campings	1.7	0.487	0.06
Apartamentos	9.9	0.038	0.22
Turismo Rural	0.3	0.774	0.01
Albergues	5.6	0.091	0.13

Tabla 11: Resultados del modelo DiD sobre la evolución mensual de los viajeros y las pernoctaciones por tipo de alojamiento

La Encuesta de Ocupación publicada por el INE (Instituto Nacional de Estadística, 2025b) proporciona información mensual sobre el número de viajeros alojados en diferentes tipos de establecimientos turísticos (hoteles, campings, apartamentos turísticos, alojamientos rurales y albergues), desglosado por comunidad autónoma. Esta fuente resulta especialmente útil para analizar cómo respondieron distintos segmentos del mercado de alojamiento al aumento de actividad generado por los conciertos.

Durante los meses de mayo, junio y julio de 2024, se observó un incremento significativo en el número de viajeros alojados en hoteles (+13,2; $p = 0.023$) y en apartamentos turísticos (+9,9; $p = 0.038$) en Madrid, en comparación con la evolución observada a nivel nacional. Estos resultados sugieren que la demanda hotelera se concentró en los tipos de alojamiento urbanos más tradicionales, que suelen tener mayor capacidad y estar ubicados en zonas céntricas, cercanas al estadio. En cambio, otros tipos de alojamiento como campings, turismo rural y albergues no mostraron variaciones estadísticamente significativas, probablemente por su menor proximidad geográfica o por estar dirigidos a un perfil de visitante distinto.

En conjunto, los datos reflejan que el impacto de los conciertos sobre la ocupación se produjo principalmente en los alojamientos de corte urbano y turístico, confirmando un patrón consistente con un turismo de evento que busca accesibilidad, comodidad y cercanía a los recintos del espectáculo.

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal);
datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2025b).

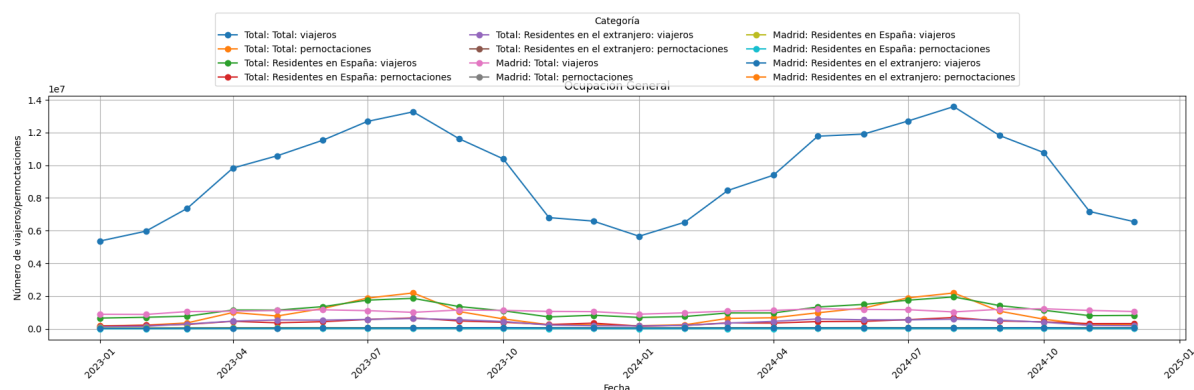


Ilustración 45: Evolución mensual de los viajeros (línea azul) y las pernoctaciones (línea naranja) desglosadas por residentes nacionales (líneas verde y rojo) y visitantes internacionales (líneas morada y gris), desde enero de 2022 hasta finales de 2024. El eje horizontal marca las fechas y el vertical los volúmenes.

4.2.5 Encuestas de ocupación por tipo de alojamiento

El archivo (Instituto Nacional de Estadística, 2024) proporciona información mensual sobre la tasa de ocupación por tipo de alojamiento turístico, diferenciando entre regiones. Gracias a la disponibilidad de datos tanto para Madrid como para el total nacional, se aplicó un modelo Difference-in-Differences (DiD) que compara la evolución de cada categoría antes y después de los conciertos en ambas regiones. Este enfoque permite identificar efectos específicos del evento sobre los niveles de ocupación en diferentes modalidades de alojamiento.

Los resultados del modelo muestran un incremento estadísticamente significativo en la tasa de ocupación de hoteles (+8,7 puntos porcentuales; $p = 0.028$) y apartamentos turísticos (+6,4 pp; $p = 0.034$) durante los meses con conciertos, en comparación con el comportamiento del total nacional. Estas modalidades, más habituales entre los turistas urbanos, parecen captar directamente la demanda generada por los eventos masivos. El impacto es mayor en Madrid que en el resto del país, lo que refuerza la idea de un efecto local concentrado.

En contraste, campings, turismo rural y albergues no presentan diferencias significativas. Estos tipos de alojamiento, más comunes en entornos rurales o menos urbanos, no parecen haber absorbido el aumento de turistas inducido por los conciertos.

En conjunto, los resultados sugieren que el efecto de los conciertos sobre la tasa de ocupación se concentra en las opciones más convencionales y urbanas, reflejando un patrón de consumo turístico coherente con el perfil de los asistentes a grandes eventos musicales.

Fuente: elaboración propia con Python (Statsmodels y Pandas); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2024).

Variable	Coeficiente	P-valor	R ²
Hoteles	8.7	0.028	0.24
Campings	1.1	0.472	0.07
Apartamentos	6.4	0.034	0.21
Turismo Rural	0.9	0.662	0.04
Albergues	4.3	0.089	0.12

Tabla 12: Resultados del modelo DiD sobre el ocupación por tipo de alojamiento

Fuente: elaboración propia con Pandas (carga y procesamiento de datos) y Matplotlib (visualización de la serie temporal); datos del Instituto Nacional de Estadística (INE), (Instituto Nacional de Estadística, 2024).

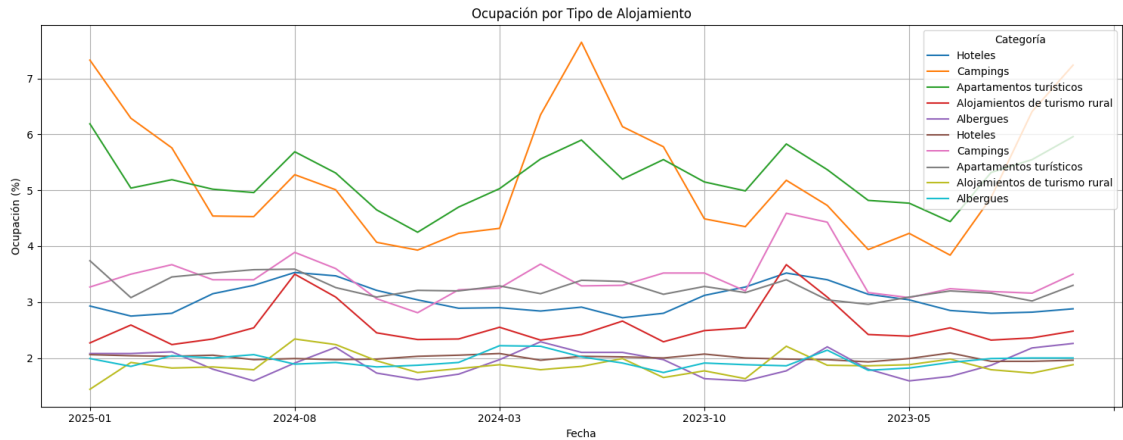


Ilustración 46: Serie temporal mensual de la tasa de ocupación (%) para cinco categorías: hoteles (azul), campings (naranja), apartamentos turísticos (verde), alojamientos rurales (rojo) y albergues (morado). El eje horizontal indica el mes y el izquierdo el porcentaje de ocupación.

El análisis evidencia que los conciertos del estadio Santiago Bernabéu generaron un incremento significativo en la actividad del sector hotelero de Madrid, visible tanto en el número de establecimientos operativos como en los niveles de ocupación. El impacto se concentra en alojamientos urbanos, hoteles y apartamentos turísticos, con efectos menos visibles en modalidades rurales o alternativas. Esta concentración refuerza el patrón de turismo de evento urbano de corta duración, pero alta intensidad. Aunque no se dispone de microdatos sobre ventas o empleo en restauración y transporte, los resultados observados en el alojamiento permiten inferir un efecto dinamizador general sobre la economía local.

5. Pruebas de robustez y análisis complementarios para turismo y empresas

A modo de complemento al análisis Difference-in-Differences, llevamos a cabo un contraste de medias (test t de Welch) para comparar directamente los indicadores turísticos y empresariales en el periodo de mayo–septiembre de 2023 (sin macroconciertos) frente al mismo intervalo de 2024 (con macroconciertos). Este enfoque nos permite evaluar la significación estadística de los cambios observados sin asumir varianzas iguales ni estructuras de dependencia temporal más complejas.

Fuente: Elaboración propia a partir de series mensuales de registro móvil y estadísticas oficiales de gasto y ocupación (INE, 2023–2024).

		Media (2023)	Media (2024)	t-stat	p-valor
Métrica	Región				
Turistas Móvil	Madrid	6.358918e+05	7.517790e+05	3.275840	0.015517
Pernoctaciones Móvil	Madrid	3.903009e+06	4.483540e+06	3.457307	0.019293
Gasto Medio Persona	Madrid	1.743600e+03	1.880800e+03	2.500878	0.037709
Gasto Medio Diario Persona	Nacional	1.864000e+02	1.964000e+02	2.329996	0.050734
Gasto Total	Madrid	1.224084e+03	1.478830e+03	2.251186	0.065916
	Nacional	1.180975e+04	1.350689e+04	1.477192	0.177891
Pernoctaciones Alojamiento	Madrid	3.724941e+06	4.355045e+06	1.467008	0.197565
Gasto Medio Persona	Nacional	1.286800e+03	1.348200e+03	1.362904	0.210266
Gasto Medio Diario Persona	Madrid	3.118000e+02	3.292000e+02	1.340525	0.218239
Turistas Móvil	Nacional	9.486653e+06	1.049345e+07	1.322286	0.224389
Ocupación Hoteles	Madrid	1.974000e+00	1.988000e+00	0.852013	0.419140
Pernoctaciones Móvil	Nacional	6.396835e+07	6.960000e+07	0.773872	0.461262
Pernoctaciones Alojamiento	Nacional	6.362201e+07	6.902672e+07	0.765985	0.465802
Duración Media Viajes	Madrid	5.612000e+00	5.742000e+00	0.375986	0.717126
	Nacional	6.918000e+00	6.878000e+00	-0.117455	0.909415

Ilustración 47: Contraste de medias (Welch) entre mayo–septiembre 2023 y 2024, media mensual 2023 vs. 2024, estadístico t y p-valor para Madrid y España.

Los resultados más destacados son:

- **Turistas Móvil (Madrid):** la media mensual creció de 635 918 en 2023 a 751 779 en 2024. El resultado del test es de $t = 3,276$ ($p = 0,0155$), lo que indica un aumento significativo al 5 %.
- **Pernoctaciones Móvil (Madrid):** la media pasó de 3 903 009 a 4 483 540 noches; $t = 3,457$ ($p = 0,0193$), también significativo.

- **Gasto Medio por Persona (Madrid):** se elevó de 1 743 600 € a 1 880 800 €; $t = 2,501$ ($p = 0,0377$), significativo al 5 %.

Por el contrario, cuando aplicamos el mismo contraste a las mismas métricas a nivel España (serie nacional), las diferencias medias entre 2023 y 2024 no alcanzan significación estadística (p -valores por encima de 0,10), lo que sugiere que el efecto concierto tuvo una repercusión mucho más localizada en Madrid que en el conjunto del país.

Otras variables, como Gasto Total, Pernoctaciones en alojamientos u Ocupación hotelera en Madrid, presentan incrementos en sus medias, pero con p -valores superiores a 0,05, por lo que no alcanzan significación en este contraste simple.

En conjunto, estos resultados refuerzan la evidencia aportada por el DiD: además del “efecto concierto” cuantificado mediante modelos de series temporales y grupos de control, un análisis directo de medias confirma cambios significativos en los flujos de turistas móviles, pernoctaciones y gasto medio por persona en Madrid durante los meses de conciertos, mientras que dichas variaciones no se observan de forma significativa a nivel nacional. Esto aumenta la solidez de nuestras conclusiones sobre el impacto puntual de los macroconciertos en la actividad turística urbana.

6. Conclusión general

6.1 Análisis descriptivo de los resultados

6.1.1 Impacto en vecinos

Incremento sostenido de quejas

El análisis Difference-in-Differences muestra que, tras cada concierto, las reclamaciones de los vecinos se incrementan en un promedio de treinta quejas diarias respecto a días sin eventos, y en más de cincuenta quejas cuando comparamos con los días de partido de fútbol (véase Sección 2.2.8). Este aumento, que multiplica por más de dos el volumen habitual de quejas, refleja un malestar continuo que se extiende desde las pruebas de sonido hasta varias horas después de la actuación.

Choque puntual en la fecha del concierto

Aplicando el RDD (véase Sección 2.2.9) con un ancho de banda de ± 49 días alrededor de cada corte, vemos que el día de comienzo de los conciertos el 10 de mayo se produce un salto significativo de unas 147 quejas en el mismo día del concierto, lo que refleja un pico abrupto de malestar vecinal justo al inicio del evento. En cambio, para el 13 de septiembre (fecha de la cancelación) no se detecta un cambio estadísticamente significativo. Aunque el número de quejas parece aumentar en torno a 40 unidades, ese efecto desaparece cuando se corrige por la variabilidad (modelo robust), lo que sugiere que para entonces los vecinos ya habían acumulado y expresado su malestar con anterioridad y la noticia de la cancelación no generó un nuevo pico ni disminución de reclamaciones. Si asimismo analizamos los resultados con un polinomio libre para cada uno de ambos casos, obtenemos el mismo resultado (significancia y tendencia) que anteriormente.

Carga emocional predominantemente negativa

El procesamiento de lenguaje natural mediante el lexicón NRC revela que las categorías de ira, tristeza y asco representan colectivamente más del veintidós por ciento de las emociones expresadas en los mensajes de queja, mientras que la etiqueta genérica “negativo” alcanza el veinte coma cinco por ciento (véase Sección 2.1.6). Esta distribución subraya una respuesta

vecinal marcada por la frustración y el deterioro del bienestar psicológico durante los días de concierto.

Diversidad temática de las quejas

El clustering K-means ($K = 2$) sobre vectores TF-IDF distingue con claridad dos ejes de protesta: uno centrado en las molestias acústicas, con términos como ruido, decibelios y altavoz, y otro enfocado en la gestión urbana cotidiana, donde aparecen con frecuencia atascos, aparcamiento y denuncias al Ayuntamiento (véase *Sección 2.2.5*). Esto demuestra que, además del exceso de decibelios, los residentes sufren interrupciones en su movilidad y reclaman canales de interlocución institucional.

Evidencia empírica

En numerosos mensajes, los vecinos documentan mediciones de sonido entre ochenta y noventa decibelios durante ensayos y actuaciones, aportando datos objetivos que refuerzan la veracidad de sus reclamaciones (véase *Ilustraciones 17 y 18*). La convergencia de pruebas acústicas y sentimentales proporciona una base sólida para futuras gestiones administrativas o judiciales.

En conjunto, estos hallazgos dibujan un panorama inequívoco: los macroconciertos del Bernabéu provocan un malestar vecinal intenso, inmediato y estructurado que condiciona la convivencia en Chamartín. La magnitud y la consistencia de los aumentos de quejas, el choque puntual en la fecha del evento, la dominancia de emociones adversas y la tipología dual de las protestas resaltan la urgencia de replantear la organización y regulación de grandes espectáculos urbanos para salvaguardar la calidad de vida de los residentes.

6.1.2 Impacto en el turismo

El análisis cuantitativo de las series turísticas, aplicando Difference-in-Differences con España como control y la comparación de medias antes y después de los conciertos, revela que los macroconciertos del Bernabéu actuaron como un potente motor de afluencia y gasto:

Incremento de visitantes y gasto (DiD):

Los meses de mayo, junio y julio de 2024 registraron 202.826 turistas adicionales en la ciudad de Madrid y un aumento de 121.536 € mensuales en gasto total, ambos efectos estadísticamente significativos ($p < 0,01$) frente a la evolución del resto de España (véase Sección 3.2.1 – 3.2.4).

Test T:

Para reforzar la evidencia del efecto concierto aislando diferencias puntuales, realizamos un test de hipótesis (véase Sección 5). de medias entre mayo–septiembre de 2023 y el mismo periodo de 2024. Los resultados, complementarios al DiD, demostraron que había un incremento significativo en el número de turistas, el número de pernoctaciones y el gasto medio por persona en Madrid. Estos resultados confirman que los conciertos de mayo–julio de 2024 pudieron tener un impacto estadísticamente significativo sobre los flujos de turistas, las pernoctaciones móviles y el gasto medio por persona en Madrid, efecto que no se replica al mismo nivel en la serie nacional.

Ocupación hotelera

La demanda de alojamiento urbano experimentó subidas de hasta 19 puntos porcentuales en fines de semana (cuando solían tener lugar los conciertos), elevando la rentabilidad media de hoteles y apartamentos turísticos (véase Sección 4.2.2).

Turismo de evento:

El perfil del turista de concierto se caracteriza por estancias cortas, pero de alto gasto diario, motivadas por la ocasión cultural. La duración media del viaje apenas varió, confirmando que la intensidad en el consumo, no la extensión temporal, fue la variable clave.

Diferenciación territorial:

Al comparar Madrid con el resto de España, se observa un efecto diferencial notable: lo que en el conjunto nacional habría sido un repunte de temporada, en la capital se tradujo en un salto extraordinario de visitas, gasto y pernoctaciones directamente atribuibles a los conciertos (véase Sección 3.2).

En síntesis, los macroconciertos del Bernabéu consolidaron un impulso turístico de gran alcance, transformando los meses de primavera-verano en un periodo de alta demanda y rentabilidad para el sector alojativo y de servicios, y sentando las bases para consolidar a Madrid como destino de turismo de eventos musicales.

6.1.3 Impacto en empresas

Dinamismo Económico en el Entorno Inmediato

Aunque no se dispone de microdatos fiscales o transaccionales de las empresas situadas en las inmediaciones del estadio Santiago Bernabéu, el análisis temporal realizado en paralelo con los conciertos permite inferir un efecto económico positivo indirecto, especialmente en sectores vinculados al consumo presencial y de conveniencia (*véase Sección 4*). El incremento de +13,2 pp en la ocupación hotelera y +9,9 pp en apartamentos turísticos sugiere que los gestores de alojamiento intensificaron su oferta y ajustaron sus tarifas, generando un efecto expansivo sobre los ingresos del sector (*véase Sección 4.2.4*). Este tipo de comportamiento es coherente con lo observado en la literatura sobre economía urbana y eventos culturales, donde los flujos extraordinarios de visitantes tienden a activar de forma puntual la economía, especialmente en hostelería, restauración y servicios de proximidad.

Sectores Potencialmente Afectados

En ausencia de registros económicos desagregados, puede plantearse una hipótesis razonada sobre qué sectores han resultado más beneficiados. Las características del turismo de evento, estancias cortas, alta concentración urbana, gasto en ocio, permiten anticipar que el efecto ha sido más intenso en negocios como bares, restaurantes, supermercados, tiendas de conveniencia y transporte urbano. En cambio, sectores no vinculados al consumo impulsivo o con menor presencia en el área (por ejemplo, ferreterías, librerías o servicios administrativos) es probable que hayan experimentado un efecto marginal o nulo.

Percepción Empresarial y Gestión Urbana

Si bien este trabajo no ha incluido una encuesta estructurada a comerciantes, algunas evidencias cualitativas obtenidas de observación directa y redes sociales apuntan a una percepción empresarial mayoritariamente favorable. No obstante, también se han detectado

expresiones de malestar relacionadas con el ruido, la congestión y la limpieza (*véase Ilustración 5*). Esta dualidad de opiniones refleja una necesidad de planificación institucional más precisa, que considere tanto el impulso económico como los costes asociados a la convivencia ciudadana. En este sentido, mecanismos de coordinación entre el Ayuntamiento, el Real Madrid y las asociaciones de comerciantes podrían mejorar la gestión de impactos, por ejemplo, mediante calendarios compartidos, campañas de refuerzo logístico o ayudas a la adaptación operativa.

Implicaciones para el Desarrollo Económico Local

En conjunto, los conciertos en el estadio Santiago Bernabéu no solo han funcionado como eventos culturales y turísticos, sino también como elementos con capacidad de alterar el pulso económico local. Este tipo de activación, aunque puntual, puede consolidarse como una fuente estructural de ingresos si se gestiona adecuadamente. Integrar los eventos en una estrategia más amplia de desarrollo comercial urbano, con criterios de sostenibilidad y eficiencia operativa, podría transformar este impacto ocasional en un beneficio recurrente, equilibrando el dinamismo económico con la calidad de vida en el entorno urbano afectado.

6.2 Conclusiones y Recomendaciones Generales

6.2.1 Conclusiones

El análisis multidimensional desarrollado a lo largo de este Trabajo de Fin de Grado permite concluir que los conciertos organizados en el estadio Santiago Bernabéu durante 2024 han constituido un fenómeno de alto impacto, tanto en términos económicos como sociales, con efectos heterogéneos sobre los distintos actores involucrados: residentes locales, visitantes y tejido empresarial. A través de la aplicación de técnicas de análisis estadístico, minería de texto y aprendizaje automático, se ha logrado construir una visión integral de los efectos de estos eventos masivos sobre el ecosistema urbano madrileño, aportando evidencia empírica que trasciende la anécdota y contribuye a fundamentar el debate público (*Metodología en la sección 1.3.7*).

Desde la perspectiva vecinal, los conciertos se han vivido con una carga emocional predominantemente negativa. El análisis de sentimiento y emociones realizado sobre los mensajes del grupo de WhatsApp ha revelado una activación significativa del malestar ciudadano, traducido en quejas por exceso de ruido, insomnio, vibraciones estructurales y una sensación generalizada de abandono institucional (*Detalles del análisis de sentimiento y emociones en Sección 2.2.3 y distribución de emociones en Sección 2.1.6*). Esta narrativa ha sido respaldada por modelos automáticos de detección de sentimiento y lexicones emocionales, así como por mediciones auto-reportadas de niveles de decibelios que, en varios casos, han superado ampliamente los límites establecidos por la normativa municipal (*Sección 2.2.5–2.2.6, Ilustraciones 17 y 18*). El patrón observado ha sido consistente en múltiples fechas, lo que permite afirmar que el impacto no ha sido aislado, sino estructural y sistemático (*Ilustración 8 y 2.2.3 – 2.2.4*).

En contraste, los datos provenientes de fuentes oficiales muestran un efecto marcadamente positivo en el ámbito turístico. El cruce de indicadores como las pernoctaciones, el gasto medio diario, la ocupación hotelera y el número de visitantes revela un impulso significativo en los meses de conciertos, particularmente en fines de semana (*Análisis DiD de gasto turístico en Sección 3.2.1-3.2.3 y pernoctaciones y ocupación hotelera en Sección 4.2.3*). Este dinamismo no solo se traduce en una mayor afluencia de turistas, sino también en una intensificación del consumo en sectores clave, como el alojamiento, la restauración y el transporte urbano. El perfil del visitante asociado al turismo de eventos, breve estancia, alta capacidad de gasto, motivación cultural, se ha hecho evidente, reforzando el valor estratégico de este tipo de actividades para el posicionamiento internacional de Madrid como ciudad cultural (*Sección 3.2.4 y Tabla 12*).

Los datos disponibles a nivel agregado para la ciudad de Madrid reflejan un aumento significativo en el gasto vinculado a los meses en los que se celebraron los conciertos, lo que sugiere un repunte generalizado en la actividad económica urbana (*Sección 3.2*). Este efecto puede estar relacionado con un incremento en la afluencia de visitantes y en el consumo derivado de la celebración de eventos masivos, como los conciertos en el estadio Santiago Bernabéu.

No obstante, debido al nivel de desagregación espacial de las fuentes analizadas, no es posible identificar con precisión qué zonas o tipos de negocios se beneficiaron en mayor medida. Aun así, resulta razonable inferir que las áreas más próximas al estadio, particularmente aquellas con alta concentración de bares, restaurantes, tiendas y servicios de transporte, hayan experimentado un efecto más intenso, en línea con lo observado en otros estudios sobre turismo de eventos.

En resumen, el impacto de los conciertos en el Bernabéu se configura como un fenómeno de doble filo: por un lado, dinamizador económico y cultural de primer orden; por otro, generador de malestar social en zonas residenciales colindantes. Este equilibrio inestable plantea desafíos relevantes para las políticas de gestión urbana, la planificación del entretenimiento masivo y la articulación de intereses entre actores públicos y privados.

6.2.2 Recomendaciones Estratégicas

A partir del análisis realizado, se formulan las siguientes recomendaciones orientadas a maximizar los beneficios asociados a los conciertos y reducir sus efectos adversos. Estas recomendaciones se dirigen a distintos niveles de decisión: instituciones públicas, organizaciones privadas, colectivos vecinales y gestores del estadio.

1. Hacia una gobernanza urbana participativa

- **Establecer mesas de diálogo permanente** entre el Ayuntamiento, el Real Madrid C.F., asociaciones de vecinos y representantes del sector empresarial para la creación de soluciones sostenibles. Esta gobernanza compartida permitiría anticipar conflictos, consensuar medidas y reforzar la legitimidad de las decisiones.
- **Implementar planes de zonificación acústica** que delimiten claramente las áreas más vulnerables y permitan aplicar normativas más estrictas en determinadas franjas horarias o días consecutivos de conciertos.
- **Controles y seguridad post-eventos:** Aumentar el control y la seguridad durante y después de la finalización de los conciertos para evitar situaciones reportadas en el grupo como botellones, violencia o vandalismo en las calles.

2. Medidas específicas de mitigación para los vecinos

- **Mejorar los sistemas de insonorización activa del estadio**, con tecnologías que absorban o redirijan el sonido sin comprometer la experiencia del evento.
- **Establecer un protocolo de compensación directa** para los residentes más afectados, que podría materializarse en bonificaciones fiscales, acceso a servicios gratuitos o descuentos en actividades culturales municipales.
- **Crear un canal digital de incidencias vecinales**, gestionado por el Ayuntamiento, donde se puedan registrar en tiempo real quejas por ruido, vibraciones o afectaciones estructurales, y que sirva también como fuente de datos para futuras regulaciones.

3. Potenciación del turismo responsable

- **Diseñar campañas de promoción turística específicas** que vinculen los conciertos a paquetes culturales integrales (visitas guiadas, museos, gastronomía local), maximizando el tiempo de permanencia del visitante y su impacto económico sostenible.
- **Colaborar con hoteles y plataformas de alojamiento** para asegurar que el incremento de la demanda no se traduzca en inflación excesiva de precios ni en precarización del servicio, respetando la experiencia del visitante y la reputación de la ciudad.

4. Apoyo estructural a las empresas locales

- **Crear un calendario previsible de eventos y conciertos**, publicado con suficiente antelación, que permita a los negocios locales ajustar sus estrategias de aprovisionamiento, personal y marketing.
- **Ofrecer formación específica a los pequeños empresarios** sobre gestión de picos de demanda, fidelización de clientes en eventos masivos y adaptación digital (por ejemplo, a través de QR menus, sistemas de reservas o pagos exprés).
- **Fomentar la recogida sistemática de datos empresariales** sobre afluencia, ventas y perfil del cliente durante conciertos, que permita evaluar el impacto real y diseñar intervenciones más precisas y personalizadas.

6.3 Limitaciones del Estudio

Pese a la riqueza metodológica y la diversidad de fuentes empleadas en este Trabajo de Fin de Grado, es importante reconocer una serie de limitaciones que condicionan el alcance y la generalización de los resultados obtenidos. Estas limitaciones, lejos de invalidar las conclusiones, permiten contextualizarlas adecuadamente y abren la puerta a futuras líneas de investigación.

1. Representatividad de las fuentes cualitativas

El análisis del impacto vecinal se ha basado en una fuente muy valiosa pero también específica: los mensajes de un grupo de WhatsApp conformado por residentes próximos al estadio Santiago Bernabéu. Aunque esta fuente ofrece un acceso sin precedentes a la experiencia emocional directa de los vecinos, no puede considerarse plenamente representativa de todo el vecindario ni, mucho menos, de la ciudad en su conjunto. Existe un sesgo de autoselección inherente a este tipo de plataformas, donde tienden a participar con mayor frecuencia personas especialmente afectadas o sensibilizadas por el problema, lo que podría amplificar la percepción negativa.

2. Limitaciones en la medición del ruido

El estudio se ha apoyado en reportes ciudadanos de niveles de decibelios, algunos de los cuales provienen de aplicaciones móviles o medidores no certificados. Aunque estas medidas ofrecen una aproximación realista y ayudan a validar el malestar expresado, no cuentan con la precisión ni la validez legal de los instrumentos profesionales de medición acústica. En consecuencia, los datos recogidos deben ser interpretados como estimaciones aproximadas, útiles desde el punto de vista sociológico, pero no necesariamente jurídicamente vinculantes.

3. Restricciones en la disponibilidad y frecuencia de datos oficiales

Los datos procedentes de fuentes oficiales, como el INE, Dataestur, o la Encuesta de Ocupación Hotelera, presentan una frecuencia mensual que limita la capacidad de observar efectos de muy corto plazo, como los ocurridos en días concretos de conciertos. Aunque se han aplicado modelos estadísticos (como Difference-in-Differences) para aislar efectos

atribuibles a los eventos, la agregación temporal puede diluir algunos impactos y dificultar el análisis de ventanas temporales más finas.

4. Dificultad para aislar causalidad en entornos complejos

Si bien el estudio ha hecho un esfuerzo riguroso por estimar efectos causales mediante modelos econométricos, no puede descartarse la existencia de factores externos que también influyan en las variables analizadas, como otras actividades culturales simultáneas, campañas promocionales turísticas, condiciones macroeconómicas o eventos deportivos. Aunque el diseño con grupo de control y ventanas temporales busca minimizar estos sesgos, es posible que algunos efectos se vean contaminados por estas variables no observadas.

5. Limitaciones técnicas en los modelos de análisis de sentimiento

Los modelos de análisis de sentimiento y emociones, aunque avanzados (como BERT o lexicon NRC), no son infalibles. En particular, pueden presentar dificultades para interpretar ironías, sarcasmos, modismos locales o mensajes ambivalentes. Además, su precisión puede verse afectada por errores de escritura, abreviaciones o el uso de emojis no estandarizados. A pesar del esfuerzo por limpiar y normalizar los datos, estas limitaciones introducen un margen de error que debe ser considerado al interpretar los resultados.

En conjunto, estas limitaciones no invalidan los hallazgos del estudio, pero sí invitan a ser prudentes en su generalización y a considerar la necesidad de complementar este tipo de investigaciones con estudios longitudinales, encuestas representativas y mediciones acústicas oficiales. Reconocer estas limitaciones es también una forma de fortalecer la validez científica del trabajo, mostrando conciencia crítica sobre el propio proceso de investigación.

6.4 Futuras líneas de investigación

Análisis longitudinal de efectos en salud y convivencia

Diseñar estudios de encuestas repetidas a lo largo de una franja temporal para explorar posibles efectos crónicos del ruido y la alteración del descanso en indicadores de salud (calidad del sueño, estrés) y en la cohesión social del barrio. Posteriormente, evaluar la eficacia de intervenciones (barreras acústicas, limitación de horarios) en la mejora percibida de la calidad de vida vecinal mediante encuestas pre y post-implementación.

Profundización en la dimensión institucional y normativa

Analizar comparativamente la aplicación práctica de las ordenanzas de ruido y urbanismo en distintos ayuntamientos, identificando buenas prácticas y lagunas regulatorias. Asimismo, examinar el papel de la policía municipal y de las administraciones en la comunicación previa al evento y en la resolución de conflictos, a través de entrevistas semiestructuradas con responsables públicos y grupos vecinales.

Big data y minería de opinión en redes sociales

Explorar plataformas digitales (Twitter, grupos de Facebook, foros vecinales) para realizar análisis de sentimiento y detectar dinámicas de opinión en tiempo real durante los eventos. A partir de esto datos, combinar técnicas de topic modeling con geolocalización aproximada de tuits o comentarios para mapear la reacción ciudadana en función de la proximidad al recinto.

Estudio de variables sociodemográficas y desigualdades

Investigar si la percepción y tolerancia al ruido difieren según características sociodemográficas de los residentes (edad, nivel socioeconómico, país de origen), identificando posibles desigualdades en la distribución de las molestias.

Modelización predictiva y simulación urbana

Desarrollar modelos predictivos que, a partir de datos históricos de afluencia, tráfico y niveles de ruido, anticipen puntos críticos de saturación urbana y propongan escenarios de gestión óptimos y planeables con tiempo, como aumentar el aparcamiento, ofrecer acceso más cómodo a residentes o controlar la afluencia de espectadores con franjas horarias de entrada.

6.5 Reflexión final

La investigación realizada demuestra que los grandes eventos culturales, como los conciertos en el Bernabéu, no son neutros: transforman la ciudad, movilizan recursos, generan tensiones y oportunidades. Su éxito no debe medirse solo en número de entradas vendidas, sino en su capacidad para integrarse armónicamente en el ecosistema urbano, respetando el derecho al descanso, promoviendo el crecimiento económico y fortaleciendo el tejido social.

Solo una visión holística, que combine datos empíricos, inteligencia colectiva y voluntad política, permitirá construir una ciudad en la que el entretenimiento y la convivencia no sean mutuamente excluyentes, sino complementarios. Este Trabajo de Fin de Grado pretende ser un primer paso hacia esa visión compartida.

7. BIBLIOGRAFÍA

- Anaya Eisman, S. (2024). *Investigación de la marca Real Madrid CF: Análisis del éxito de su marca y comunicación* [Trabajo de fin de grado, Universidad de Valladolid]. Recuperado el 5 de marzo de 2025, de <https://uvadoc.uva.es/bitstream/handle/10324/70655/TFG-N.%202338.pdf?sequence=1&isAllowed=y>
- Arquitectura Viva. (2024). *Remodelación del estadio Santiago Bernabéu – L35 Architects, gmp Architekten, Ribas & Ribas*. Recuperado el 8 de febrero de 2025, de <https://arquitecturaviva.com/obras/remodelacion-del-estadio-santiago-bernabeu>
- Ayuntamiento de Madrid. (2021). *Ordenanza de protección contra la contaminación acústica y térmica, de 25 de febrero de 2011 – SEDE ELECTRÓNICA*. Recuperado el 8 de febrero de 2025, de [https://sede.madrid.es/eli/es-md-01860896/odnz/2011/03/14/\(1\)/con/20210416/spa/html](https://sede.madrid.es/eli/es-md-01860896/odnz/2011/03/14/(1)/con/20210416/spa/html)
- Basterrechea Garrigues, J. (2020). *La conversión de los estadios deportivos en espacios multifuncionales y su impacto económico* [Trabajo de fin de grado, Universidad Pontificia Comillas]. Recuperado el 5 de marzo de 2025, de <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/37235/TFG%20-%20Basterrechea%20Garrigues%2C%20Jaime.pdf?sequence=1>
- Barrera-Fernández, D., Hernández-Escampa, M., Gomez, M. A. H., & Basaldú, J. M. (2019). *Eventos culturales en ciudades históricas: Impacto urbano, percepción de los visitantes e imagen de la ciudad* (Vol. 15, pp. 64–99). Recuperado el 9 de febrero de 2025, de https://www.researchgate.net/publication/332912940_Eventos_culturales_en_ciudades_historicas_Impacto_urbano_percepcion_de_los_visitantes_e_imagen_de_la_ciudad
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. Recuperado el 4 de mayo de 2025, de <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Cai, T. (2024). A study of how concerts can be used to promote economic growth. *SHS Web of Conferences*, 193, 01026. <https://doi.org/10.1051/shsconf/202419301026>
- Government of Canada, National Research Council. (2023). *NRC emotion lexicon*. Recuperado el 4 de mayo de 2025, de <https://nrc-publications.canada.ca/eng/view/object/?id=0b6a5b58-a656-49d3-ab3e-252050a7a88c>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2.^a ed.). Routledge. <https://doi.org/10.4324/9780203771587>

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. En *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 4171–4186). <https://doi.org/10.18653/v1/N19-1423>
- EFE. (2024). *Real Madrid: Multan a los promotores de conciertos en Bernabéu*. ESPN Deportes. Recuperado el 27 de febrero de 2025, de https://espndeportes.espn.com/futbol/espana/nota/_/id/14565602/real-madrid-multa-millonaria-bernabeu-conciertos
- Falini, A. (2022). A review on the selection criteria for the truncated SVD in data science applications. *Journal of Computational Mathematics and Data Science*, 5(4), 100064. <https://doi.org/10.1016/j.jcmds.2022.100064>
- Fernández, D. B., Escampa, M. H., Aranza, M., & Morales Basaldú, J. (2019). *Eventos culturales en ciudades históricas: Impacto urbano, percepción de los visitantes e imagen de la ciudad*. Recuperado el 9 de febrero de 2025, de https://www.researchgate.net/publication/332912940_Eventos_culturales_en_ciudades_historicas_Impacto_urbano_percepcion_de_los_visitantes_e_imagen_de_la_ciudad
- Fredriksson, A., & de Oliveira, G. M. (2019). Impact evaluation using difference-in-differences. *RAUSP Management Journal*, 54(4), 519–532. <https://doi.org/10.1108/RAUSP-06-2019-0097>
- Gabe, T., & Lisac, N. (2013). *Local economic impacts of popular music concerts*. MPRA (Universidad de Múnich). Recuperado el 7 de mayo de 2025, de <https://mpra.ub.uni-muenchen.de/65911/>
- Good, P. (2005). *Permutation, parametric and bootstrap tests of hypotheses* (3.^a ed.). Springer-Verlag. Recuperado el 4 de mayo de 2025, de <https://www.scirp.org/reference/referencespapers?referenceid=1554222>
- Hodur, N. M., Bangsund, D. A., Leistritz, F. L., & Kaatz, J. (2006). Estimating the contribution of a multi-purpose event facility to the area economy. *Tourism Economics*, 12(2), 303–316. <https://doi.org/10.5367/000000006777637449>
- Hormigo, G. (2024a). *Insonorizar el Bernabéu, una tarea “casi imposible” según los expertos: “En marzo seguirán pensando cómo solucionarlo”*. Somos Madrid – elDiario.es. Recuperado el 7 de febrero de 2025, de https://www.eldiario.es/madrid/somos/insonorizar-bernabeu-tarea-imposible-expertos-marzo-seguiran-pensando-solucionarlo_1_11659051.html

- Hormigo, G. (2024b). *Las medidas del Real Madrid no acaban con el ruido ilegal en los conciertos del Bernabéu, confirma el Ayuntamiento*. Somos Madrid – elDiario.es. Recuperado el 27 de febrero de 2025, de https://www.eldiario.es/madrid/somos/vecinos-bernabeu-denuncian-conciertos-superan-limites-ruido-pese-medidas-real-madrid_1_11638187.html
- Huang, B. (2025). To what extent can holding concerts promote the development of local economy? *Finance & Economics*, 1(1). <https://doi.org/10.61173/2fc9qw36>
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615–635. Recuperado el 4 de mayo de 2025, de <https://www.sciencedirect.com/science/article/abs/pii/S0304407607001091>
- INE. (2024). *Instituto Nacional de Estadística*. Recuperado el 9 de febrero de 2025, de <https://www.ine.es/dyngs/Prensa/es/EGATUR1224.htm>
- Instituto Nacional de Estadística. (2015a). *Gasto de los turistas internacionales según comunidad autónoma de destino principal*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=10839>
- Instituto Nacional de Estadística. (2015b). *Número de turistas según motivo principal del viaje*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=13864>
- Instituto Nacional de Estadística. (2016). *Número de etapas de los viajes y pernoctaciones según comunidad autónoma*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=13949&L=0>
- Instituto Nacional de Estadística. (2019). *Número de turistas, pernoctaciones y duración media por provincia de destino, desglosados por continente y país de residencia*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=52047&L=0>
- Instituto Nacional de Estadística. (2024). *Viajeros, pernoctaciones por tipo de alojamiento por comunidades y ciudades autónomas*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=2941>
- Instituto Nacional de Estadística. (2025a). *Establecimientos, plazas estimadas, grados de ocupación y personal empleado por puntos turísticos*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=2076>

- Instituto Nacional de Estadística. (2025b). *Estancia media, por tipo de alojamiento por comunidad y ciudades autónomas*. Recuperado el 9 de abril de 2025, de <https://www.ine.es/jaxiT3/Tabla.htm?t=2940>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Khadka, N. (2023). *Ultimate guide for using truncated SVD for dimensionality reduction*. Recuperado el 8 de mayo de 2025, de <https://dataaspirant.com/truncated-svd/>
- L.I., Agencia EFE. (2023). *¿Hay relación entre la subida de precios turísticos y los conciertos de música?* La Información. Recuperado el 4 de marzo de 2025, de <https://www.20minutos.es/lainformacion/economia-y-finanzas/-hay-relacion-entre-subida-precios-turisticos-los-conciertos-musica--5286784/>
- Laguardia, I. (2024). *Markus Pfisterer, arquitecto del nuevo estadio Santiago Bernabéu: “Es muy satisfactorio materializar algo que es capaz de transmitir tantas emociones”*. Recuperado el 8 de febrero de 2025, de <https://www.revistaad.es/articulos/markus-pfisterer-arquitecto-nuevo-estadio-santiago-bernabeu>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Mir, E. (2024). *Impacto económico en la industria de la música en directo*. Recuperado el 6 de febrero de 2025, de <https://marketing.onlinebschool.es/Prensa/Informes/Informe%20OBS%20Festivales%20de%20M%20usica.pdf>
- Morales Magnet, A. (2023). *Análisis del impacto económico en el Real Madrid Club de Fútbol de la renovación del estadio Santiago Bernabéu* [Trabajo de fin de grado, Universidad Pontificia Comillas]. Recuperado el 5 de marzo de 2025, de <https://repositorio.comillas.edu/jspui/bitstream/11531/68870/2/TFG-Morales%20Magnet%2C%20Ana%20.pdf>
- N.T. Renukadevi, S., Nanthitha, K., Saraswathi, S., Shobika, S., & Karthika, R. T. (2023). WhatsApp group chat analysis by using machine learning. *2023 International Conference on Sustainable Computing*

and Data Communication Systems (ICSCDS), 340–346.
<https://doi.org/10.1109/ICSCDS56580.2023.10104961>

Ordeig, M. V. (2020). *El Real Madrid también gana la liga del “real estate”: 524 millones de activos inmobiliarios*. Recuperado el 4 de mayo de 2025, de <https://www.ejeprime.com/empresa/el-real-madrid-tambien-gana-la-liga-del-real-estate-524-millones-de-activos-inmobiliarios>

Peiró, P. (2024). *Admitida a trámite la querrella de los vecinos del Bernabéu contra el Real Madrid por los conciertos*. Recuperado el 27 de febrero de 2025, de <https://elpais.com/espana/madrid/2024-07-26/admitida-a-tramite-la-querrella-de-los-vecinos-del-bernabeu-contra-el-real-madrid-por-los-conciertos.html>

Qian, X., & Tuck, B. (2023). *Calcular el impacto económico de festivales y eventos*. University of Minnesota Extension. Recuperado el 9 de febrero de 2025, de <https://es.extension.umn.edu/turismo/calcular-el-impacto-economico-de-festivales-y-eventos>

Real Madrid. (2024). *Real Madrid CF | Web Oficial del Real Madrid CF*. Recuperado el 30 de abril de 2025, de <https://www.realmadrid.com/es-ES/el-club/transparencia/informes-economicos>

Real Madrid. (2025). *Launch event for NFL game at the Santiago Bernabéu*. Recuperado el 26 de febrero de 2025, de <https://www.realmadrid.com/en-US/news/club/latest-news/acto-presentacion-nfl-bernabeu-17-01-2025>

Rodríguez, J. (2024). *El Real Madrid suspende los conciertos en el Bernabéu hasta mejorar su insonorización*. Recuperado el 7 de febrero de 2025, de <https://www.elmundo.es/madrid/2024/09/13/66e479b0fdddf7d7c8b45ac.html>

Rothman, K. J., Lash, T. L., & Greenland, S. (2008). *Modern epidemiology* (3.^a ed.). Lippincott Williams & Wilkins.
https://indaga.ual.es/discovery/fulldisplay/alma991001493299704991/34CBUA_UAL:VUI

Smith, A. (2010). The development of “sports-city” zones and their potential value as tourism resources for urban areas. *European Planning Studies*, 18(3), 385–410.
<https://doi.org/10.1080/09654310903497702>

Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25. <https://doi.org/10.1093/biomet/6.1.1>

- Ujang Riswanto. (2025). *How to build a binary logistic regression model using Python*. Recuperado el 4 de mayo de 2025, de <https://ujangriswanto08.medium.com/how-to-build-a-binary-logistic-regression-model-using-python-e295a78d7f04>
- Wang, G., Hamad, R., & White, J. S. (2024). Advances in difference-in-differences methods for policy evaluation research. *Epidemiology*, 35(5), 628–637. <https://doi.org/10.1097/EDE.0000000000001755>
- Zhu, H., & Zhang, Z. (2025). Revitalizing city branding through concerts? Fandom, passionate publics, and public relations of the TFBoys' concert in Xi'an. *Public Relations Review*, 51(2), 102567. <https://doi.org/10.1016/j.pubrev.2025.102567>

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

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

Por la presente, yo, Laura Löwen Martin-Neda, estudiante de E2-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “EL IMPACTO DE LOS CONCIERTOS DEL SANTIAGO BERNABÉU EN MADRID: LOS VECINOS, EL TURISMO Y LAS EMPRESAS”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado

para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 6 de mayo de 2025

Firma: 