# Memoria de Capstone Project

**Predicción de la Calidad del aire en Madrid: Impacto de las ZBE y de la vida urbana**

Máster en Data Science / Convocatoria Junio de 2024



| [Raquel Rodríguez](mailto:raquel.rodriguez.vds@immune.institute) | [**Alejandra lopez sturm**](mailto:alejandra.lopez.vds@immune.institute) | **Maruxa Moreira** |
| --- | --- | --- |

# ÍNDICE

1. **INTRODUCCIÓN**

Predicción y de la calidad del aire

1. **ESTADO DEL ARTE**

Tecnología al servicio de la calidad del aire en las ciudades

1. **DESARROLLO**

Metodología y herramientas para la predicción a partir de los datos

1. **PRUEBAS Y RESULTADOS**
2. **CONCLUSIONES**

Innovación para el futuro de la vida urbana

## 

# 1. INTRODUCCIÓN: Predicción y de la calidad del aire.

La elección de este proyecto surge de la preocupación por el impacto que la calidad del aire tiene en la salud y el bienestar de la población urbana. Nosotras compartimos la convicción de que la protección del medio ambiente es una responsabilidad fundamental y que la tecnología puede y debe ser una aliada en este desafío. La calidad del aire es un factor clave en la sostenibilidad de las ciudades y en la calidad de vida de sus habitantes, especialmente para colectivos vulnerables como las personas con problemas respiratorios.

El objetivo principal de este trabajo es desarrollar un modelo capaz de predecir la calidad del aire en la ciudad de Madrid, utilizando datos meteorológicos y de tráfico procedentes de fuentes públicas. Este modelo pretende servir como herramienta de apoyo para la toma de decisiones proactivas en la gestión de la contaminación urbana. Como objetivos secundarios, se analiza el impacto de las políticas de movilidad, en particular de las Zonas de Bajas Emisiones (ZBE), la influencia de eventos urbanos especiales sobre los niveles de contaminación, y se explora el potencial de la innovación y el análisis de datos para mejorar la calidad de vida en Madrid, aportando evidencia sobre la eficacia de las políticas públicas implementadas.

La metodología seguida se basa en la recopilación, integración y análisis de datos abiertos, empleando herramientas como Python y SQL para la preparación y modelización de los datos. Se han utilizado modelos de aprendizaje automático adecuados para series temporales, como regresión lineal, Random Forest y XGBoost, junto con técnicas de preprocesamiento y validación para asegurar la robustez de los resultados. El análisis incluye tanto variables tradicionales como innovadoras, permitiendo una visión integral del fenómeno estudiado.

El alcance del proyecto abarca el periodo comprendido entre 2021 y 2025, excluyendo el año 2020 del modelado predictivo debido a las circunstancias excepcionales del confinamiento, aunque dicho año se ha considerado en el análisis multi-anual del efecto frontera para comparar situaciones de tráfico mínimo. Además, se ha tenido en cuenta el impacto del teletrabajo y otras medidas excepcionales a la hora de investigar, con el fin de enriquecer las conclusiones y recomendaciones.

La memoria se organiza en cinco secciones principales. Tras esta introducción, el Estado del Arte presenta un análisis de los trabajos previos y el contexto actual de la predicción de la calidad del aire y las políticas urbanas. En Desarrollo se describe detalladamente el proceso de obtención, tratamiento y modelización de los datos. La sección de Pruebas y Resultados documenta los experimentos realizados y los hallazgos obtenidos, con especial atención al efecto frontera y a la influencia de eventos urbanos. Por último, en las conclusiones se sintetizan los resultados, se discuten las implicaciones y se proponen recomendaciones para futuras investigaciones y políticas públicas.

# 2. ESTADO DEL ARTE: Tecnología al servicio de la calidad del aire en las ciudades.

El análisis y la predicción de la calidad del aire en entornos urbanos es un campo que ha experimentado un notable avance en los últimos años, impulsado por la preocupación social, la disponibilidad de datos abiertos y el desarrollo de nuevas tecnologías. A continuación, se revisan los principales enfoques y hallazgos en la literatura científica y técnica, situando nuestro trabajo en el contexto actual.

***Predicción de la calidad del aire mediante machine learning***

La aplicación de algoritmos de aprendizaje automático para la predicción de la calidad del aire se ha consolidado como una de las líneas de investigación más prometedoras. Se han desarrollado modelos de machine learning, como árboles de decisión y regresión, para anticipar los niveles de contaminantes clave (SO₂, CO, NO₂, O₃ y PM2.5) en entornos urbanos, demostrando una mejora significativa respecto a los métodos tradicionales.

De forma similar, [Tamore (2023)](https://drive.google.com/file/d/1815r7nUlmtgkwqYDBXKQOrGhGK5vhe02/view?usp=drive_link) revisa el uso de técnicas de machine learning e IoT para la monitorización y predicción de la calidad del aire, subrayando la importancia de la selección de variables y la integración de datos meteorológicos y de tráfico para mejorar la precisión de los modelos.

En ciudades como Pekín, Londres o Nueva Delhi, se han implementado sistemas predictivos que combinan datos históricos, meteorológicos y de tráfico, permitiendo anticipar episodios de contaminación y activar alertas tempranas para la población. Estos estudios coinciden en señalar que la calidad de los datos y la correcta selección de variables son factores determinantes para el éxito de los modelos predictivos.

***El Índice Europeo de Calidad del Aire (EAQI) como referencia metodológica***

En Europa, la evaluación y comunicación de la calidad del aire se realiza habitualmente a través del [Índice Europeo de Calidad del Aire (EAQI)](https://ecmwf-projects.github.io/copernicus-training-cams/proc-aq-index.html), gestionado por la Agencia Europea de Medio Ambiente (EEA). El EAQI se calcula a partir de las concentraciones de cinco contaminantes principales (O₃, NO₂, SO₂, PM2.5 y PM10), determinando un subíndice para cada uno y tomando como valor final el subíndice más alto, reflejando así el peor nivel de calidad del aire en cada momento. El índice se actualiza cada hora y se clasifica en seis niveles, desde "Muy bueno" hasta "Extremadamente pobre", facilitando la interpretación para la ciudadanía y las autoridades Agencia Europea de Medio Ambiente, 2024.

En nuestro trabajo, hemos tomado como referencia esta metodología estandarizada a nivel europeo para el cálculo del AQI, complementándola con variables meteorológicas y de tráfico, así como con variables innovadoras como la diferencia de tráfico entre el interior y el exterior de la ZBE y la presencia de eventos especiales. Esta aproximación multidimensional permite un análisis más completo y adaptado a la realidad de Madrid.

***Impacto de las Zonas de Bajas Emisiones (ZBE) y políticas de movilidad***

La implantación de Zonas de Bajas Emisiones (ZBE) es una de las políticas más extendidas en Europa para combatir la contaminación urbana. Diversos estudios han evaluado su impacto en ciudades como Londres, París, Berlín y Madrid, concluyendo que, en general, las ZBE contribuyen a la reducción de contaminantes como el NO₂ y las partículas en suspensión. Sin embargo, la literatura también señala la existencia del llamado “efecto frontera”, por el cual la contaminación puede desplazarse a las zonas limítrofes a la ZBE, reduciendo parcialmente la eficacia global de la medida. En el caso de Madrid, informes de Ecologistas en Acción y otros colectivos han puesto de manifiesto que, aunque se han producido mejoras, los niveles de NO₂ siguen superando los límites recomendados en determinadas áreas, lo que evidencia la necesidad de políticas más integrales y una evaluación continua de su impacto ([Ecologistas en Acción, 2023](https://drive.google.com/file/d/1I72-g4klsm8SdPz65S5ctFOGprYcNa4W/view?usp=drive_link)).

A nivel internacional, la experiencia de ciudades como Milán o Estocolmo muestra que la combinación de ZBE con incentivos al transporte público y la movilidad activa (bicicleta, caminar) puede potenciar los resultados positivos y mejorar la calidad de vida urbana.

***Influencia del teletrabajo y eventos urbanos***

El teletrabajo ha emergido como un factor relevante en la reducción de la contaminación urbana, especialmente durante la [pandemia de 2020, cuando la drástica disminución del tráfico se tradujo en una mejora temporal de la calidad del aire](https://drive.google.com/file/d/1l-jk85EINJ9SfzACMyMhfy_R2IoSJCnM/view?usp=drive_link) en muchas ciudades. Estudios recientes confirman que, aunque [el teletrabajo aumentó en España en 2024](https://www.equiposytalento.com/noticias/2025/04/09/el-teletrabajo-volvio-a-aumentar-en-espana-en-2024-pero-sigue-lejos-de-la-media-europea), su implantación sigue siendo inferior a la media europea, lo que limita su impacto estructural sobre la contaminación (Equipos y Talento, 2025).

Por otro lado, la celebración de eventos urbanos masivos (deportivos, culturales, etc.) puede provocar picos de contaminación, tanto por el aumento del tráfico como por la concentración de personas y actividades. Esta es la razón por la que recomendamos incorporar estas variables en los modelos predictivos para obtener una visión más realista y útil para la gestión municipal.

***Tecnología, apps y ciencia de datos para la gestión ambiental***

El uso de aplicaciones móviles y plataformas digitales, como Waze o Google Maps, ha revolucionado la forma en que se monitoriza y gestiona la movilidad urbana. Estas apps recopilan datos en tiempo real sobre el tráfico, rutas, velocidad de los vehículos y patrones de movilidad, lo que permite a investigadores y autoridades analizar el impacto de las políticas de movilidad, como las Zonas de Bajas Emisiones (ZBE), de manera mucho más precisa y dinámica.

Además de facilitar la recogida de datos de alta resolución espacial y temporal, estas aplicaciones ofrecen funcionalidades que afectan directamente a la efectividad de las ZBE y, por extensión, a la calidad del aire: monitorización del tráfico en las zonas afectadas, análisis de cambios en los flujos de vehículos antes y después de la implantación de la ZBE, detección de rutas alternativas utilizadas por los conductores para evitar la ZBE (lo que puede desplazar la contaminación a barrios colindantes), alertas y avisos sobre restricciones de acceso, y datos agregados para evaluar la efectividad de las políticas públicas.

Entre las ventajas de utilizar estos datos destacan la inmediatez, la cobertura amplia y la posibilidad de analizar patrones de movilidad a gran escala. Sin embargo, también existen limitaciones, como posibles problemas de privacidad, sesgos en la muestra de usuarios, desplazamiento del tráfico y la contaminación a zonas limítrofes, y el hecho de que estas apps no miden directamente los niveles de contaminantes atmosféricos. En el contexto de Madrid, la integración de datos de apps de navegación permite optimizar rutas, reducir tiempos de congestión y, potencialmente, disminuir las emisiones. Además, facilita la identificación de puntos críticos de congestión y la evaluación del impacto real de las ZBE en la movilidad y la calidad del aire.

***Inteligencia artificial y sostenibilidad urbana***

La inteligencia artificial se perfila como una herramienta clave para el diseño de ciudades más sostenibles y resilientes. Según Retina Tendencias, la IA puede ayudar a mitigar entre el 5% y el 10% de las emisiones globales de gases de efecto invernadero para 2030, gracias a aplicaciones como la optimización del tráfico, la monitorización ambiental y la gestión eficiente de recursos ([Retina Tendencias, 2023](https://retinatendencias.com/negocios/ciudades-del-futuro-tejiendo-sostenibilidad-con-hilos-de-inteligencia-artificial/)).

Proyectos como GreenLight, que optimiza semáforos para reducir emisiones, o el uso de IA en la reforestación y el análisis de datos ambientales, demuestran el potencial transformador de estas tecnologías en la lucha contra la contaminación urbana. No obstante, la literatura también advierte sobre la necesidad de garantizar la transparencia, la ética y la equidad en el uso de la IA, para evitar sesgos y asegurar que los beneficios lleguen a toda la ciudadanía.

***Aportación diferenciadora de este trabajo***

A pesar de los avances descritos, la mayoría de estudios se centran en la predicción general de la calidad del aire o en la evaluación global de políticas. Nuestro trabajo aporta un enfoque innovador al analizar el efecto frontera de las ZBE en Madrid, incorporar variables diferenciadoras como eventos urbanos y teletrabajo, y utilizar técnicas de machine learning para evaluar el impacto real de las políticas de movilidad. De este modo, situamos nuestro proyecto en la vanguardia de la investigación aplicada, contribuyendo a la construcción de ciudades más saludables y sostenibles.

# 3. DESARROLLO: Metodología y herramientas para la predicción a partir de los datos.

**Organización, trabajo en equipo y comunicación**

El desarrollo de este proyecto ha sido posible gracias a una organización interna muy cuidada y a una comunicación constante y eficaz entre las integrantes del equipo. Desde el inicio, establecimos reuniones semanales a través de Google Meet para planificar tareas, revisar avances y resolver dudas de forma conjunta. La comunicación diaria se mantuvo mediante WhatsApp y Discord, lo que nos permitió tomar decisiones rápidas, compartir recursos y coordinarnos eficazmente ante cualquier imprevisto.

La gestión de archivos y documentación se centralizó en Google Drive, donde creamos una estructura de carpetas clara y accesible para todas. Para evitar confusiones y pérdidas de información, elaboramos una guía de archivos que describe la función y el contenido de cada documento, asegurando que cualquier persona ajena al equipo pueda navegar fácilmente por el repositorio.

La distribución de tareas se realizó de manera estratégica, buscando la máxima eficiencia y fomentando siempre un entorno de aprendizaje compartido y apoyo mutuo. Todas participamos activamente en la toma de decisiones y en la comprensión global del proceso, asumiendo cada una la responsabilidad principal de determinadas fases para optimizar el flujo de trabajo y garantizar la calidad de los resultados.

**Procedimientos y metodología**

El proceso comenzó con la recopilación de más de 600 archivos de datos abiertos del Ayuntamiento de Madrid, abarcando registros horarios y diarios de contaminantes atmosféricos (NO₂, O₃, PM10, PM2.5, SO₂), variables meteorológicas (temperatura, viento, humedad) y datos de tráfico de diferentes estaciones repartidas por la ciudad, tanto dentro como fuera de la ZBE.

La limpieza y el preprocesamiento de estos datos fue una tarea fundamental y laboriosa. Utilizamos Python y MySQL para unificar archivos mensuales, eliminar duplicados, gestionar valores nulos y homogeneizar los nombres de columnas y los formatos de fecha. En Power Query, se aplicaron transformaciones adicionales como la creación de columnas auxiliares (año, mes, estación), la desanidación de columnas y la unificación de formatos para facilitar el análisis conjunto.

Se prestó especial atención a la coherencia temporal y espacial de los datos, validando que todas las fuentes estuvieran correctamente alineadas antes de proceder a la integración. La integración de los datos supuso fusionar todas las fuentes en un único dataset consolidado, que sirvió de base para el análisis exploratorio y el modelado. Este proceso incluyó la validación cruzada de la coherencia y consistencia de los datos, asegurando que la información estuviera alineada temporal y espacialmente para su posterior análisis y modelado.

**Creación e integración de variables exógenas innovadoras**

Uno de los principales valores añadidos del proyecto ha sido la creación de variables exógenas innovadoras, que enriquecen el análisis y la predicción de la calidad del aire en Madrid, diferenciando nuestro AQI personalizado de los índices tradicionales empleados en la Unión Europea.

Se diseñó una variable diaria que mide la diferencia de intensidad de tráfico entre estaciones situadas dentro y fuera de la ZBE. Para ello, primero se clasificaron todas las estaciones según su localización respecto a la ZBE, y posteriormente se calculó la diferencia diaria de tráfico entre ambos grupos. Esta variable permite analizar el “efecto frontera” y el posible desplazamiento de la contaminación hacia zonas limítrofes, aportando una visión más realista del impacto de las políticas de movilidad.

Además, también creamos una variable binaria (0/1) que identifica los días en los que se celebraron eventos urbanos relevantes (conciertos, partidos, manifestaciones, ferias, etc.), integrando información de portales oficiales y eventos de la ciudad. Para ello, se generó un DataFrame con todas las fechas del periodo de estudio (2021-2025) y se marcó con un 1 los días con evento y con 0 el resto. Esta variable permite capturar el impacto real de la vida urbana sobre la calidad del aire y analizar su influencia en los patrones de tráfico y contaminación.

Junto a estas variables, se integraron también las variables meteorológicas (temperatura, viento, humedad, etc.), fundamentales para capturar la influencia del clima en la dispersión y concentración de contaminantes. Todas estas variables, junto con las tradicionales, se integraron en un DataFrame final que sirvió de base para el modelado y el análisis exploratorio. Este dataset permitió analizar el impacto de la movilidad urbana, los eventos especiales y las condiciones meteorológicas sobre el tráfico y la contaminación, desarrollar un AQI personalizado adaptado a la realidad de Madrid y mejorar la capacidad predictiva y la utilidad práctica de los modelos para la gestión municipal.

**Herramientas y tecnologías utilizadas**

El desarrollo del proyecto ha requerido el uso de una amplia variedad de herramientas y tecnologías, seleccionadas por su capacidad para facilitar la obtención, procesamiento, análisis y visualización de los datos. Se emplearon lenguajes de programación como Python y SQL para la manipulación y consulta de datos, así como DAX y M (Power Query) para el modelado y transformación en Power BI. Las principales librerías utilizadas fueron pandas, numpy, matplotlib, seaborn, scikit-learn y XGBoost, junto con utilidades como requests, zipfile, glob, os y tempfile para la gestión automatizada de archivos.

Los entornos de trabajo incluyeron Jupyter Notebook, Google Colab, Visual Studio Code, Cursor para el desarrollo y documentación de los análisis, MySQL para la limpieza y transformación estructural de los datos, y Power BI Desktop para el modelado, análisis y visualización interactiva. La colaboración y el control de versiones se gestionaron a través de Google Drive y GitHub, mientras que la comunicación y la coordinación se apoyaron en Google Docs, WhatsApp, Discord y Google Meet.

Esta combinación de recursos permitió automatizar la descarga y el procesamiento de datos, realizar análisis exploratorios avanzados, crear variables innovadoras y preparar datasets robustos y listos para el modelado, asegurando la reproducibilidad y la eficiencia en todas las fases del proyecto.

**Evolución del modelado: de la exploración a la personalización**

El proceso de modelado de este proyecto se ha caracterizado por una evolución continua, en la que cada etapa ha supuesto un avance metodológico y conceptual hacia un índice de calidad del aire más ajustado a la realidad urbana de Madrid. El punto de partida fue el análisis de la Zona de Bajas Emisiones (ZBE) y el denominado “efecto frontera”.

Se estudió el comportamiento de los principales contaminantes tanto dentro como fuera de la ZBE, evaluando si la implantación de esta política provocaba un desplazamiento de la contaminación hacia las zonas limítrofes. Para ello, en primer lugar, se identificaron y clasificaron todas las estaciones de medición de tráfico y calidad del aire según su localización geográfica respecto a la ZBE, utilizando coordenadas y mapas oficiales del Ayuntamiento de Madrid. Posteriormente, se agruparon los datos de tráfico de cada estación y se calcularon medias diarias de intensidad de tráfico para los dos grupos: estaciones dentro de la ZBE y estaciones fuera de la ZBE.

A continuación, se construyó una nueva variable diaria que representa la diferencia de tráfico entre ambos grupos, restando la media de tráfico fuera de la ZBE a la media dentro de la ZBE para cada día del periodo de estudio. Este procedimiento permitió analizar de forma cuantitativa si existía un “efecto frontera”, es decir, si la reducción del tráfico dentro de la ZBE se correspondía con un aumento en las zonas limítrofes.

Finalmente, esta variable “diferencia de tráfico” se integró como predictor en los modelos de calidad del aire, permitiendo captar de forma más precisa el impacto real de las políticas de movilidad sobre la contaminación y facilitando la identificación de posibles desplazamientos de la polución debidos a la implantación de la ZBE.

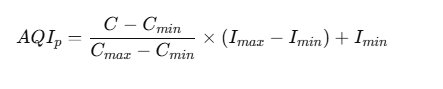
A continuación, se incorporó el análisis del impacto de los eventos especiales en la ciudad, como conciertos, partidos o manifestaciones, que alteran los patrones de tráfico y, en consecuencia, la concentración de contaminantes. Para ello, se creó una variable binaria que identifica los días con eventos relevantes, enriqueciendo el modelo y adaptándolo a la dinámica real de la vida urbana madrileña. La integración de esta variable permitió analizar de manera diferenciada el efecto de la vida social y cultural sobre la calidad del aire, algo ignorado en los modelos tradicionales.

En paralelo, se avanzó hacia la construcción de un modelo supervisado general, en el que se probaron distintos algoritmos de regresión (LinearRegression, RandomForestRegressor, KNeighborsRegressor y XGBRegressor). Tras un proceso riguroso de validación cruzada y ajuste de hiperparámetros, el modelo XGBRegressor demostró ser el más robusto y preciso, gracias a su capacidad para manejar relaciones complejas y evitar el sobreajuste. Este modelo permitió calcular un AQI personalizado para toda la ciudad, integrando no solo los contaminantes atmosféricos, sino también variables meteorológicas (temperatura, viento, humedad) y las variables exógenas desarrolladas específicamente para este proyecto. Así, se consiguió una visión mucho más realista y contextualizada de la calidad del aire, diferenciando claramente entre el interior y el exterior de la ZBE y detectando posibles barreras al flujo de aire.

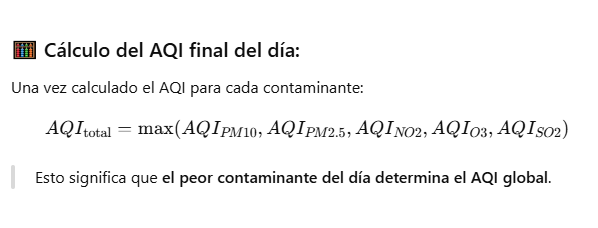
Con el objetivo de profundizar aún más en la comprensión de los patrones de calidad del aire, se aplicaron técnicas de modelado no supervisado, concretamente clustering mediante K-Means. Este enfoque permitió identificar grupos de días con características ambientales similares, revelando tendencias ocultas y validando la importancia de las variables innovadoras propuestas. La selección óptima del número de clusters se realizó mediante el método del codo, y los resultados obtenidos confirmaron que tanto el clima como el tráfico, junto a los contaminantes, son factores determinantes en la calidad del aire de Madrid. Este análisis no sólo reforzó la validez de las conclusiones obtenidas con el modelo supervisado, sino que aportó una perspectiva adicional sobre la diversidad de situaciones que se dan en la ciudad a lo largo del año.

El siguiente paso en la evolución del modelado fue el desarrollo de un modelo supervisado por zonas, alineado con la división oficial del Ayuntamiento de Madrid en cinco áreas geográficas. Para ello, se trabajó en la separación y geolocalización de los datos, permitiendo calcular un AQI personalizado para cada zona, siempre integrando tanto los contaminantes como las variables meteorológicas, de tráfico y eventos especiales. Este enfoque permitió visualizar en un mapa los diferentes niveles de calidad del aire ajustados a la realidad de cada área, proporcionando una herramienta innovadora y de gran utilidad para la gestión urbana y la toma de decisiones a nivel local.

Un aspecto fundamental de nuestro trabajo ha sido el cálculo riguroso del AQI, siguiendo el estándar europeo. Para cada contaminante, se aplicó la fórmula oficial de interpolación, que permite situar la concentración observada dentro del rango correspondiente y obtener así el subíndice AQI:



donde C es la concentración observada del contaminante, Cmin y Cmax los límites del rango, e Imin e Imax los valores mínimo y máximo del subíndice para ese rango. Una vez calculado el AQI para cada contaminante (PM10, PM2.5, NO₂, O₃ y SO₂), el AQI global del día se determina tomando el valor máximo entre todos ellos:



Esto significa que la calidad del aire diaria queda definida por el contaminante más desfavorable registrado, siguiendo la metodología oficial de la Unión Europea.

Sin embargo, la gran aportación de este trabajo ha sido la propuesta de un AQI personalizado, que no sólo replica el cálculo estándar, sino que lo enriquece con variables meteorológicas y exógenas, permitiendo un análisis mucho más completo y adaptado a la realidad de Madrid. La integración de datos meteorológicos ha sido clave para capturar la influencia del clima en la dispersión y concentración de contaminantes, mientras que las variables exógenas (diferencia de tráfico y eventos especiales) han permitido reflejar la complejidad de la vida urbana y el impacto real de las políticas y eventos sobre la contaminación atmosférica.

**Análisis y visualización interactiva**

En paralelo, se desarrolló un modelo de datos relacional en Power BI, integrando todas las fuentes y permitiendo el análisis desde diferentes perspectivas (temporal, espacial, por contaminante, etc.). Se crearon medidas DAX para calcular medias, máximos, mínimos, superaciones de umbrales y clasificaciones de calidad del aire, y se diseñaron dashboards interactivos que facilitan la interpretación de los resultados tanto para expertos como para usuarios no especializados. La validación de los resultados se realizó comprobando la coherencia de los valores calculados con los umbrales legales y revisando la consistencia de los datos tras cada transformación.

**Innovaciones, dificultades y soluciones**

El proyecto destaca por la incorporación de variables exógenas innovadoras (diferencia de tráfico ZBE, eventos especiales) y el desarrollo de un AQI personalizado adaptado a la realidad urbana de Madrid. La combinación de análisis supervisado y no supervisado ha permitido obtener una visión más completa y robusta del fenómeno, reforzando la validez de las conclusiones.

Entre las principales dificultades superadas se encuentran la identificación y delimitación de las ZBE (que requirió la consulta de mapas oficiales y documentación técnica), la obtención y organización de grandes volúmenes de datos heterogéneos, la integración de fuentes diversas y la automatización de procesos para garantizar la reproducibilidad. Estos retos se resolvieron mediante un trabajo minucioso, la aplicación de reglas de limpieza y validación, y el desarrollo de scripts y notebooks específicos para cada fase. La gestión de valores nulos y atípicos se abordó mediante reglas de limpieza y, en algunos casos, imputación o eliminación, asegurando la calidad y fiabilidad de los datos analizados.

El desarrollo de este proyecto ha sido un proceso multidisciplinar y colaborativo, en el que la organización, la comunicación y la innovación han sido claves para superar los retos y alcanzar los objetivos propuestos. El resultado es un trabajo sólido, reproducible y de alto valor aplicado, que sienta las bases para futuras investigaciones y para la mejora de la gestión ambiental en entornos urbanos complejos como Madrid.

# 

# 4. PRUEBAS Y RESULTADOS.

Todos los archivos, notebooks, datos y dashboards del proyecto están disponibles en el siguiente enlace de Google Drive:

[**Calidad del Aire\_Raquel\_Alejandra\_Maruxa**](https://drive.google.com/drive/folders/1ZjoDm0-VQhzamNQURBtNYVvctD0Sdv6i?usp=drive_link)

Incluimos una guía de navegación por las carpetas:

[**Guía de Entregables del Proyecto**](https://docs.google.com/document/d/1KrUi7Lzqky6BMThZkymNbjwAxFBaIrWui5CaslcILzY/edit?usp=drive_link)

# 5. CONCLUSIONES: Innovación para el futuro de la vida urbana.

El desarrollo de este proyecto ha supuesto un salto cualitativo en la forma de analizar y comprender la calidad del aire en Madrid, gracias a la integración de variables exógenas, la creación de un índice AQI personalizado y el uso de herramientas de visualización avanzadas.

El proceso se inició con la recopilación y limpieza exhaustiva de datos de contaminantes, meteorología y tráfico, superando importantes retos de integración y homogeneización. La construcción de una tabla maestra y la aplicación de técnicas de análisis exploratorio permitieron identificar patrones, relaciones y anomalías que no serían visibles con enfoques tradicionales.

La innovación principal reside en la creación de un AQI personalizado, que no solo incorpora los contaminantes regulados por la normativa europea, sino que añade variables meteorológicas y contextuales como la intensidad de tráfico y la celebración de eventos especiales. Este enfoque ha permitido captar la complejidad real de la vida urbana madrileña y anticipar episodios de contaminación asociados a situaciones excepcionales, como grandes eventos o condiciones meteorológicas adversas.

El análisis de los dashboards desarrollados en Power BI ayudan a visualizar la evolución temporal y espacial de la calidad del aire, identificar diferencias entre zonas y periodos, y evaluar el impacto de las políticas públicas.

Se ha constatado, por ejemplo, que la implantación de las Zonas de Bajas Emisiones (ZBE) ha contribuido a una reducción progresiva del NO₂, aunque el efecto frontera es gradual y requiere medidas complementarias a mayor escala.

La integración de la variable “evento especial” ha permitido detectar alteraciones significativas en los patrones de tráfico y contaminación, subrayando la necesidad de adaptar las políticas de movilidad y transporte público durante estos días para minimizar el impacto ambiental.

La validación del modelo, basada en la coherencia con los umbrales legales y la consistencia interna de los datos, garantiza la fiabilidad de las conclusiones y la utilidad práctica de las recomendaciones.

**Recomendaciones y líneas de mejora**

A partir de los hallazgos obtenidos, se recomienda reforzar y adaptar las políticas de movilidad, especialmente en torno a las ZBE, para maximizar su impacto y evitar el desplazamiento de la contaminación a zonas limítrofes.

Es fundamental mejorar la gestión del transporte público y la comunicación con la ciudadanía durante eventos especiales, anticipando restricciones y promoviendo alternativas sostenibles. Además, se aconseja mantener y ampliar la recogida de datos abiertos, integrar más sensores y explorar la aplicación de modelos predictivos avanzados, como un modelo de LLM, para capturar relaciones aún más complejas.

El AQI personalizado desarrollado en este proyecto constituye una herramienta innovadora y flexible, capaz de adaptarse a la realidad cambiante de la ciudad y de servir de referencia para otras urbes interesadas en mejorar la calidad del aire y la salud de sus habitantes. La metodología y los resultados obtenidos sientan las bases para el desarrollo de sistemas de alerta temprana, la evaluación continua de políticas públicas y la promoción de una movilidad más sostenible y equitativa.

En definitiva, este trabajo demuestra que la integración de datos, la innovación metodológica y la visualización avanzada son claves para afrontar los retos ambientales de las ciudades del siglo XXI, contribuyendo de manera tangible a la mejora de la calidad de vida urbana y a la protección de la salud pública.

# 

# 6. REFERENCIAS BIBLIOGRÁFICAS

* **Artículos científicos y revisiones citados**

[Tamore, M](https://drive.google.com/file/d/1815r7nUlmtgkwqYDBXKQOrGhGK5vhe02/view?usp=drive_link). (2023). Machine learning and IoT for air quality monitoring: A comprehensive review. Journal of Urban Technology.**Informes y organismos oficiales**

[Agencia Europea de Medio Ambiente. (2024)](https://ecmwf-projects.github.io/copernicus-training-cams/proc-aq-index.html). Índice Europeo de Calidad del Aire (EAQI).

[Ecologistas en Acción. (2023).](https://drive.google.com/file/d/1I72-g4klsm8SdPz65S5ctFOGprYcNa4W/view?usp=drive_link) Informe sobre la calidad del aire en Madrid.

[Ministerio para la Transición Ecológica y el Reto Demográfico. (2019)](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/glosario-de-terminos/glosario-contaminantes/oxidos-nitrogeno.html). Glosario de contaminantes: Óxidos de nitrógeno.

* **Artículos de prensa y medios digitales**

[El Mundo. (2023, 24 de abril)](https://www.elmundo.es/madrid/2023/04/24/64450216e4d4d886258b458b.html). La implantación del teletrabajo reduce los desplazamientos al trabajo en Madrid.

[Equipos y Talento. (2025, 9 de abril)](https://www.equiposytalento.com/noticias/2025/04/09/el-teletrabajo-volvio-a-aumentar-en-espana-en-2024-pero-sigue-lejos-de-la-media-europea). El teletrabajo volvió a aumentar en España en 2024, pero sigue lejos de la media europea.

[Retina Tendencias. (2023)](https://retinatendencias.com/negocios/ciudades-del-futuro-tejiendo-sostenibilidad-con-hilos-de-inteligencia-artificial/). La inteligencia artificial puede ayudar a mitigar entre el 5% y el 10% de las emisiones globales de gases de efecto invernadero para 2030.

* **Recursos y otros informes**

[Ayuntamiento de Madrid](https://datos.madrid.es/portal/site/egob) Nuestra ciudad en datos abiertos.

[Ecologistas en Acción. (2023)](https://www.ecologistasenaccion.org/308909/ecologistas-en-accion-denuncian-la-contaminacion-de-dioxido-de-nitrogeno-no2-del-trafico-en-la-ciudad-de-murcia/). Ecologistas en Acción denuncian la contaminación de dióxido de nitrógeno (NO₂) del tráfico en la ciudad de Murcia.

[Ver normas APA](https://biblioguias.uam.es/citar/estilo_apa_7th_ed)