

**Diseño de un tablero interactivo en Power BI para el análisis de datos meteorológicos y de
calidad del aire en Colombia**

Carlos Enrique Sarmiento Aguilera

Asesor

Sixyel Jeyson Castañeda Coronado

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Nota de Aceptación

Sixyel Jeyson Castañeda Coronado

Director de Trabajo de Grado

Mireya García

Jurado

Dedicatoria

Esta tesis de grado está dedicada a las dos personas más importantes de mi vida: mi esposa y nuestro pequeño hijo. Ellos han sido, durante estos años, la principal fuente de fuerza y motivación para no rendirme, para continuar adelante a pesar del cansancio y para superar cada uno de los obstáculos que se presentan en el camino. Cada dificultad superada ha significado un motivo más de orgullo y aprendizaje. Este logro representa un paso más hacia el cumplimiento de nuestros sueños como familia, y me siento profundamente orgulloso de haberlo alcanzado de su mano. Este trabajo es por y para ustedes.

Agradecimientos

Agradezco a todos los tutores de la Especialización en Ciencia de Datos y Analítica por los conocimientos y las herramientas brindadas a lo largo del programa, los cuales fueron fundamentales para la correcta estructuración y desarrollo de este trabajo de grado. De manera especial, expreso mi agradecimiento a mi director de tesis, el Mgtr. Sixyel Jeyson Castañeda Coronado, por su acompañamiento constante, paciencia y valiosas retroalimentaciones, que permitieron orientar adecuadamente los objetivos de la investigación y alcanzar de forma satisfactoria los resultados propuestos. Finalmente, agradezco a la Universidad Nacional Abierta y a Distancia (UNAD) por ofrecer una modalidad educativa flexible, que hace posible la formación académica y el cumplimiento de metas profesionales sin descuidar las demás responsabilidades personales y laborales.

Resumen

Esta investigación explora la influencia de los tableros interactivos en Power BI en el análisis de la relación entre las condiciones meteorológicas y la calidad del aire en ciudades colombianas que cuentan con estaciones de monitoreo. El estudio integra datos ambientales sobre la contaminación del aire (como PM10 y PM2.5) y variables meteorológicas, como la temperatura y la humedad, para desarrollar una plataforma dinámica y fácil de usar para el usuario final. En áreas urbanas, es evidente cómo la contaminación del aire impacta la salud de los habitantes. Por ello, el objetivo de este estudio es mejorar la comprensión de este fenómeno y proporcionar información útil para mitigar los riesgos asociados. Para lograrlo, se emplearán datos abiertos de plataformas como SISAIRE e IDEAM, lo que permitirá a los interesados monitorear la contaminación del aire en tiempo real y reconocer tendencias relacionadas con las condiciones meteorológicas y la contaminación atmosférica.

Palabras claves: Power BI, Calidad del aire, Condiciones meteorológicas, Contaminación del aire, Visualización de datos.

Abstract

This research explores the influence of interactive dashboards in Power BI on the analysis of the relationship between meteorological conditions and air quality in Colombian cities with monitoring stations. The study integrates environmental data on air pollution (such as PM10 and PM2.5) and meteorological variables, including temperature and humidity, to develop a dynamic and user-friendly platform for end users. In urban areas, it is evident how air pollution impacts the health of inhabitants. Therefore, the objective of this study is to enhance the understanding of this phenomenon and provide useful information to mitigate the associated risks. To achieve this, open data from platforms such as SISAIRE and IDEAM will be used, allowing stakeholders to monitor air pollution in real time and identify trends related to weather conditions and atmospheric pollution.

Keywords: Power BI, Air quality, Meteorological conditions, Air pollution, Data visualization.

Tabla de Contenido

Introducción	12
Descripción del Problema	14
Planteamiento del Problema.....	14
Sistematización del Problema	16
Justificación	18
Objetivos	20
Objetivo General	20
Objetivos Específicos.....	20
Marco de Referencia	21
Estado del Arte.....	21
Marco Contextual.....	22
Marco Teórico.....	24
Calidad del Aire y Salud Pública	24
Factores Meteorológicos y su Influencia en la Calidad del Aire	24
Visualización de Datos en el Análisis de la Calidad del Aire	25
Uso de Tableros Interactivos en la Ciencia de Datos.....	25
Estudios Previos sobre Visualización de la Calidad del Aire en Colombia.....	26
Marco Conceptual.....	26
Contaminación del Aire.....	26
Condiciones Meteorológicas	26
Calidad del Aire.....	27
Tablero Interactivo en Power BI	27

Análisis de Datos.....	27
Relación entre Variables	27
Marco Normativo.....	28
Metodología.....	30
Método	30
Tipo de Estudio.....	30
Recolección de Datos.....	31
Resultados.....	32
Dashboard IDEAM.....	32
Dashboard SISAIRE.....	34
Dashboard IDEAM VS SISAIRE.....	42
Dashboard Mapa Coroplético.....	44
Análisis Estadístico de la Relación entre Variables Meteorológicas y Contaminantes.....	45
Análisis Focalizado: Efecto de la Velocidad del Viento sobre PM10 y PM2.5.....	68
Correlación Simple.....	69
Correlación Parcial.....	71
Modelo Lineal Simple.....	72
GAM Reducido.....	74
Regresión Robusta.....	78
Conclusiones.....	81
Recomendaciones.....	83
Referencias Bibliográficas.....	84
Apéndices.....	92

Lista de Tablas

Tabla 1 <i>Primeras Filas del Archivo departamentos_sisaire.csv</i>	35
Tabla 2 <i>Primeras Filas del Archivo municipios_sisaire.csv</i>	35
Tabla 3 <i>Primeras Filas del Archivo estaciones_sisaire.csv</i>	36
Tabla 4 <i>Primeras Filas del Archivo datos_comtaminantes_sisaire.csv</i>	37
Tabla 5 <i>Primeras Filas del Archivo sisaire_merge.csv</i>	39
Tabla 6 <i>Matriz de Correlación</i>	48
Tabla 8 <i>VIF para cada Contaminante y sus Variables Meteorológicas Asociadas</i>	52
Tabla 10 <i>Resultados de los Modelos No Lineales con Interacciones – Año 2021</i>	55
Tabla 11 <i>Correlación Simple PM2.5</i>	69
Tabla 12 <i>Correlación Simple PM10</i>	70
Tabla 13 <i>Correlación Parcial Material Particulado</i>	72
Tabla 14 <i>Modelo de Regresión Lineal Simple para PM10</i>	72
Tabla 15 <i>Modelo de Regresión Lineal Simple para PM2.5</i>	73
Tabla 16 <i>Resumen Estadístico del Modelo GAM para PM2.5</i>	75
Tabla 17 <i>Resumen Estadístico del Modelo GAM para PM10</i>	77
Tabla 18 <i>Coefficientes de Regresión Robusta</i>	79

Lista de Figuras

Figura 1 <i>Dashboard IDEAM</i>	33
Figura 2 <i>Dashboard SISAIRES</i>	41
Figura 3 <i>Dashboard IDEAM VS SISAIRES</i>	43
Figura 4 <i>Dashboard Mapa Coroplético</i>	45
Figura 5 <i>Correlación entre Variables Meteorológicas y Contaminantes</i>	46
Figura 6 <i>Matriz de Valores p</i>	49
Figura 7 <i>Relaciones entre Contaminantes y Meteorología</i>	50
Figura 8 <i>Resultados de los Modelos Lineales</i>	54
Figura 9 <i>Relación No Lineal entre NO₂ y Variables Meteorológicas</i>	61
Figura 10 <i>Relación No lineal entre O₃ y Variables Meteorológicas</i>	62
Figura 11 <i>Relación No Lineal entre PM_{2.5} y Variables Meteorológicas</i>	63
Figura 12 <i>Relación No Lineal entre PM₁₀ y Variables Meteorológicas</i>	64
Figura 13 <i>Comparación del R² entre modelos</i>	65
Figura 14 <i>Correlación Simple PM_{2.5} vs Velocidad del Viento</i>	70
Figura 15 <i>Correlación simple PM₁₀ vs Velocidad del Viento</i>	71
Figura 16 <i>Recta Regresión PM₁₀ vs Velocidad del Viento</i>	73
Figura 17 <i>Recta Regresión PM_{2.5} vs Velocidad del Viento</i>	74
Figura 18 <i>Efecto No Lineal del Viento Sobre PM_{2.5}</i>	76
Figura 19 <i>Efecto No Lineal del Viento Sobre PM₁₀</i>	78

Lista de Apéndices

Apéndice A <i>Scripts Utilizados para la Conversión y Agregación de Datos del IDEAM</i>	92
Apéndice B <i>Scripts Utilizados para Procesamiento y Consolidación de Datos de SISAIRES</i>	96
Apéndice C <i>Script Integración y Filtrado de Datos IDEAM–SISAIRES</i>	110
Apéndice D <i>Script en R para el Análisis Estadístico Avanzado</i>	112

Introducción

En los últimos años, la contaminación del aire se ha convertido en uno de los problemas más importantes a nivel mundial. Diversos estudios han mostrado que respirar aire con altos niveles de contaminantes como el material particulado PM2.5 o el ozono (O₃) puede afectar gravemente la salud de las personas, provocando enfermedades respiratorias y cardiovasculares (Lelieveld et al., 2020). En Colombia, esta situación también se presenta en muchas ciudades, donde factores como el tráfico, la industria y el clima influyen directamente en la calidad del aire (Marín et al., 2025).

Por eso resulta muy necesario contar con herramientas que permitan entender cómo las condiciones del clima —como la temperatura, la humedad o la velocidad del viento— se relacionan con los niveles de contaminación. Aunque instituciones como el IDEAM y el SISAIRE publican información valiosa, muchas veces estos datos no se presentan de una manera que sea fácil de interpretar por el público general o por quienes toman las decisiones ambientales.

En este trabajo se propone diseñar un tablero interactivo en Power BI que permita analizar de forma visual y sencilla la relación entre las variables meteorológicas y los niveles de contaminación del aire en ciudades colombianas. Con esta herramienta se busca facilitar el análisis de la información, reconocer patrones y aportar datos útiles para comprender cómo el clima puede influir en la calidad del aire.

La investigación tiene un enfoque cuantitativo y descriptivo, y un diseño no experimental y longitudinal, ya que se trabaja con datos históricos sin alterar las variables. Para su desarrollo se empleará la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), que

incluye las fases de comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación y despliegue.

El documento se estructura de la siguiente forma: primero se presenta la descripción y planteamiento del problema; luego, la justificación y los objetivos de la investigación. Más adelante se desarrolla el marco de referencia, que abarca el estado del arte, el contexto y las normas que sustentan el estudio. Posteriormente, se explica la metodología aplicada, y finalmente se muestran los resultados, conclusiones y recomendaciones.

Descripción del Problema

Planteamiento del Problema

El diseño de un tablero en Power BI que combine datos sobre el clima y la calidad del aire en una ciudad colombiana aborda un desafío importante en la actualidad, dado que los niveles de contaminación del aire y las condiciones meteorológicas tienen un impacto directo en la salud y bienestar de los ciudadanos. Esta elección se basa en la creciente inquietud por los niveles de polución y sus efectos adversos, especialmente en entornos urbanos. De acuerdo con el Sistema de Información sobre Calidad del Aire (SISAIRE), muchas ciudades en Colombia enfrentan problemas serios relacionados con la calidad del aire, ya que las concentraciones de las partículas en suspensión, tales como elPM10 y PM2.5 exceden los límites indicados por la Organización Mundial de la Salud (OMS) (Aldana & Cárdenas, 2020). Esto subraya la necesidad de tener herramientas que permitan a los ciudadanos y a los responsables de las políticas públicas tomar decisiones informadas basadas en datos que sean accesibles y fáciles de entender.

En cuanto a los antecedentes, varios estudios han documentado cómo la calidad del aire se relaciona con problemas de salud pública. Investigaciones han evidenciado que la exposición a altos niveles de contaminantes atmosféricos está vinculada a un incremento en enfermedades respiratorias y cardiovasculares (Sandoval D. et al., 2019; Galvis-Vizcaíno et al., 2021). Sin embargo, a pesar de esta información, hay una notable falta de integración entre los datos meteorológicos y los índices de la calidad del aire en plataformas que sean accesibles al público general. Los principales aspectos del problema incluyen no solo la carencia de información accesible sobre la calidad del aire, sino también las dificultades para correlacionar estos datos con variables meteorológicas que pueden influir en la calidad del aire. Por ejemplo, fenómenos climáticos como las lluvias o cambios drásticos de temperatura pueden afectar

considerablemente las concentraciones de contaminantes (Acosta & Meléndez, 2024). Esta situación impacta a diversos actores en el ámbito educativo y social, ya que tanto estudiantes como comunidades vulnerables son especialmente sensibles a los efectos negativos de la contaminación (Suárez, 2024).

Dicho lo anterior, se logra plantear la pregunta de investigación: ¿De qué manera el uso de un tablero interactivo en Power BI facilita el análisis de la relación entre las condiciones climáticas y los niveles de calidad del aire en las ciudades colombianas?

Sistematización del Problema

A partir del planteamiento anterior, se identifican los principales elementos que estructuran el problema de investigación. En primer lugar, la calidad del aire en las ciudades colombianas se ha visto afectada por el incremento de contaminantes como el PM10 y PM2.5 (Agudelo-Londoño et al., 2024), lo que genera un impacto directo en la salud de los habitantes y en la sostenibilidad ambiental. Las condiciones meteorológicas, tales como la temperatura, la humedad y las precipitaciones, influyen significativamente en la dispersión o concentración de estos contaminantes, por lo que su análisis conjunto resulta esencial para comprender los patrones de contaminación atmosférica (Martínez Morales, 2024).

Sin embargo, la falta de integración de datos entre las variables meteorológicas y los indicadores de calidad del aire representa una limitación importante. A pesar de la existencia de información pública en plataformas como SISAIRE e IDEAM, la presentación de los datos no siempre permite una interpretación clara ni facilita la toma de decisiones por parte de la ciudadanía o de los responsables de políticas ambientales. Esto evidencia una brecha entre la disponibilidad de los datos y su aprovechamiento efectivo mediante herramientas de análisis visual.

En este contexto, surge la necesidad de desarrollar un tablero interactivo en Power BI que permita integrar, analizar y visualizar de manera dinámica la información proveniente de fuentes oficiales, facilitando la comprensión de la relación entre las condiciones meteorológicas y la calidad del aire. Este tablero busca servir como una herramienta de apoyo tanto para el análisis técnico como para la educación ambiental, promoviendo la accesibilidad y la transparencia de los datos públicos.

De esta manera, la sistematización del problema se puede expresar a través de las siguientes preguntas orientadoras:

¿Cuáles son las variables meteorológicas que presentan mayor influencia sobre los niveles de calidad del aire en las ciudades colombianas?

¿De qué manera la integración de datos abiertos en Power BI puede mejorar el análisis y la comprensión de esta relación?

¿Qué beneficios aporta la visualización interactiva de los datos para la toma de decisiones y la concientización ciudadana respecto a la contaminación del aire?

Justificación

La investigación sobre cómo pueden llegar a relacionarse las condiciones meteorológicas y los niveles de calidad del aire en las ciudades colombianas tiene gran importancia en el contexto actual, donde la contaminación del aire viene siendo un problema crítico para la salud pública. Según un estudio basado en un modelo atmosférico, aproximadamente 8.8 millones de muertes al año a nivel mundial son atribuibles a la contaminación ambiental por partículas PM2.5 y ozono (Lelieveld et al., 2020) (Mora, 2022). Lo que subraya la urgencia de abordar este problema en Colombia, donde el Departamento Nacional de Planeación estima que entre 2015 y 2016 se registraron 8.241 muertes relacionadas con este fenómeno (Parra Sánchez et al., 2020). Además, el material particulado PM2.5 es uno de los contaminantes más perjudiciales, afectando especialmente a las poblaciones vulnerables (Parra Sánchez et al., 2020) (Cabrera Encinoso, 2021).

El diseño de un tablero interactivo usando Power BI permitirá visualizar y analizar estos datos de forma accesible, facilitando la comprensión del impacto que tienen las condiciones meteorológicas sobre la calidad del aire.

Este estudio no solo beneficiará a la comunidad educativa al proporcionar herramientas visuales para el análisis de datos ambientales, sino que también contribuirá al conocimiento académico sobre la interrelación entre variables climáticas y contaminantes atmosféricos. La disponibilidad de datos abiertos a través del SISAIRE y otras plataformas permite a los investigadores acceder a información crucial que puede ser utilizada para estudios epidemiológicos y para informar políticas públicas, además investigaciones previas han demostrado que el uso de datos sobre calidad del aire puede mejorar las decisiones sobre salud pública y políticas ambientales (Cadavid-Giraldo et al., 2021). Al integrar estos datos en un

formato interactivo, se espera fomentar un mayor foco de los ciudadanos sobre los problemas ambientales y sus efectos en su salud.

Finalmente, esta investigación contribuirá a mitigar el problema identificado al proporcionar una herramienta que permita a los tomadores de decisiones evaluar rápidamente la calidad del aire y las condiciones climáticas. Esto es esencial para implementar estrategias efectivas de mitigación y prevención. Por ejemplo, el monitoreo constante de la calidad del aire puede facilitar alertas tempranas para proteger a poblaciones vulnerables durante episodios de alta contaminación. Además, al identificar patrones en los datos meteorológicos y de calidad del aire, se podrán desarrollar intervenciones dirigidas que aborden específicamente las fuentes de contaminación.

Objetivos

Objetivo General

Diseñar un tablero interactivo en Power BI que permita analizar la relación entre las condiciones meteorológicas y los niveles de calidad del aire en ciudades colombianas, evaluando su utilidad como herramienta para interpretar patrones y apoyar la toma de decisiones en gestión ambiental.

Objetivos Específicos

Recopilar datos meteorológicos y de la calidad del aire de fuentes confiables, utilizando los recursos de Datos Abiertos de Colombia y SISAIRE del IDEAM.

Analizar estadísticamente las correlaciones entre variables meteorológicas y los niveles de contaminación del aire.

Diseñar visualizaciones dinámicas e interactivas en Power BI para identificar tendencias y patrones en los datos recopilados, facilitando la comprensión y el análisis de la relación entre las condiciones meteorológicas y la calidad del aire.

Marco de Referencia

Estado del Arte

El análisis de la calidad del aire y su relación con las condiciones meteorológicas constituye un área de estudio cada vez más relevante. Por ejemplo, las investigaciones en China han evidenciado que los factores como la temperatura, la presión atmosférica y la altura de la capa límite pueden explicar hasta cerca de un 30 % de la variabilidad en las concentraciones de PM_{2.5}, mientras que otras contribuciones se atribuyen a emisiones antropogénicas (Li, Daqing, Ashkenazy, Havlin, & etal., 2025). En otro estudio en el que se ha usado un enfoque estadístico en Vietnam, se encontró que la temperatura ($\rho = 0,30$), la radiación ($\rho = 0,30$) y la humedad ($\rho = 0,28$) ejercieron los mayores efectos sobre las concentraciones de PM_{2.5}, especialmente durante episodios de contaminación (Nguyen et al., 2023). Estas evidencias internacionales confirman que las variables meteorológicas tienen un papel significativo en la dinámica de los contaminantes atmosféricos, aunque los efectos varían según la región y la temporada.

En el ámbito de la tecnología, las herramientas de visualización y plataformas de inteligencia de negocio (BI) se han utilizado para facilitar la interpretación de datos ambientales. Por ejemplo, Microsoft Power BI fue empleada en un caso práctico en el municipio de Rustenburg (Sudáfrica) para representar datos de calidad del aire mediante dashboards interactivos, permitiendo a los oficiales de calidad del aire y a profesionales de salud ambiental obtener una visión más clara del cumplimiento de las normas y del comportamiento de las concentraciones (Wright & Wernecke, 2020). De manera similar, se ha documentado un proyecto que integra datos en tiempo real de meteorología y calidad del aire en un dashboard de Power BI, que incluye variables como PM_{2.5}, PM₁₀, visibilidad y humedad, lo que demuestra la viabilidad técnica de tales instrumentos (Tushar117, 2025). Estos avances tecnológicos

evidencian que la combinación de grandes volúmenes de datos ambientales con herramientas de presentación visual puede potenciar la toma de decisiones y la divulgación pública.

En el caso colombiano, aunque existen plataformas institucionales como el SISAIRE del IDEAM que recogen información de estaciones de monitoreo de calidad del aire, la integración de esas variables con factores meteorológicos en herramientas visuales interactivas accesibles al público parece aún limitada. Por tanto, pese a que se dispone de los datos, falta una mayor articulación que permita a diversos actores como lo son los ciudadanos, gestores públicos e investigadores, explorar las relaciones entre clima y contaminación de forma dinámica. En este sentido, el presente estudio pretende contribuir a cerrar esa brecha mediante el diseño de un tablero interactivo en Power BI que integre datos abiertos, analice correlaciones entre variables meteorológicas y contaminantes y presente los resultados de forma intuitiva y útil para la gestión ambiental.

Marco Contextual

En Colombia, la contaminación atmosférica representa uno de los principales retos ambientales y de salud pública. De acuerdo con el Departamento Nacional de Planeación (DNP, 2018), cada año se registran miles de muertes atribuibles a la exposición a contaminantes del aire, especialmente material particulado fino ($PM_{2.5}$) y ozono (O_3). Estos contaminantes son responsables de un aumento considerable en las enfermedades respiratorias y cardiovasculares, afectando de manera particular a las poblaciones urbanas de bajos recursos, niños y adultos mayores. Según el Instituto Nacional de Salud (De Salud, 2022), la exposición prolongada al material particulado se relaciona con incrementos en hospitalizaciones por asma, enfermedad pulmonar obstructiva crónica (EPOC) y complicaciones cardíacas.

El Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), a través del Subsistema de Información sobre Calidad del Aire (SISAIRE), coordina la recolección y publicación de datos provenientes de estaciones de monitoreo distribuidas en todo el país. Estas estaciones registran contaminantes atmosféricos como PM₁₀, PM_{2.5}, NO₂, SO₂, CO y O₃, junto con variables meteorológicas como la temperatura, la humedad relativa, la velocidad del viento y la radiación solar (Vásquez Franco, 2019). Sin embargo, aunque la información es pública, su análisis e interpretación suelen requerir conocimientos técnicos avanzados, lo cual limita el acceso y comprensión por parte del ciudadano común o incluso de actores institucionales que no manejan herramientas de análisis de datos.

En este contexto, el uso de tecnologías de inteligencia de negocio (BI), como Microsoft Power BI, representa una oportunidad significativa para transformar datos crudos en información visual, comprensible y útil para la toma de decisiones ambientales. El desarrollo de tableros interactivos que integren datos meteorológicos y de calidad del aire puede facilitar la identificación de patrones y tendencias, además de permitir un monitoreo más dinámico de los fenómenos contaminantes. Como señalan Wright y Wernecke (2020), las visualizaciones interactivas son herramientas poderosas para democratizar la información ambiental y fomentar la participación ciudadana informada.

En el plano internacional, la Organización Mundial de la Salud (OMS) y el Banco Mundial han advertido que la contaminación atmosférica constituye uno de los principales riesgos ambientales para la salud, con más de 7 millones de muertes prematuras al año asociadas a la mala calidad del aire (OMS, 2021). En Colombia, esta situación se agrava en ciudades como Bogotá, Medellín, Cali, Bucaramanga y Barranquilla, donde las condiciones topográficas, el tránsito vehicular y las emisiones industriales aumentan las concentraciones de material

particulado (Marín et al., 2025). Frente a ello, fortalecer los sistemas de monitoreo y mejorar la comunicación de datos al público general es esencial para promover políticas ambientales más efectivas y sostenibles.

Así que este proyecto se enmarca dentro de la línea de investigación en tecnología aplicada al análisis ambiental, cuyo propósito es desarrollar herramientas que integren la ciencia de datos, visualización y sostenibilidad. El diseño del tablero en Power BI busca además ofrecer una solución práctica, accesible y escalable que permita explorar la relación entre las condiciones meteorológicas y la calidad del aire en distintas ciudades colombianas, apoyando tanto la gestión institucional como la sensibilización ciudadana respecto al impacto ambiental y la salud pública.

Marco Teórico

Calidad del Aire y Salud Pública

La calidad del aire hace referencia a la cantidad de contaminantes presentes en la atmósfera y su efecto en la salud de los ciudadanos y el medio ambiente. Según la Organización Mundial de la Salud (OMS), la exposición prolongada a partículas como PM2.5, PM10, ozono y dióxido de nitrógeno está vinculada a enfermedades respiratorias, problemas cardiovasculares y un aumento en la mortalidad prematura (Cárdaba Arranz, 2012). En Colombia, investigaciones han revelado niveles preocupantes de contaminación en ciudades como Bogotá, Medellín y Cali, afectando especialmente a poblaciones vulnerables, como niños y adultos mayores (Chaparro & Quevedo, 2021; Botello & Arenas, 2023).

Factores Meteorológicos y su Influencia en la Calidad del Aire

Las condiciones meteorológicas tienen un rol importante en cómo se dispersan y concentran los contaminantes en el aire. Hay factores como la temperatura, la humedad relativa, la velocidad del viento y las precipitaciones influyen directamente en cómo se acumular o

dispersan las partículas contaminantes (Ruiz & Ruiz, 2024) (Warthon Olarte et al., 2024). Por ejemplo, temperaturas elevadas pueden acelerar las reacciones químicas en la atmósfera, aumentando los niveles de ozono (Coates et al., 2016), mientras que una alta humedad puede retener partículas finas, empeorando la contaminación en áreas urbanas densamente pobladas (Ji et al., 2020) (Klompaker et al., 2021). A nivel global, diversos estudios han encontrado una relación significativa entre estos factores climáticos y la calidad del aire (González, 2021) (Armesto Arenas & Cañon Barriga, 2023).

Visualización de Datos en el Análisis de la Calidad del Aire

La visualización de datos es una herramienta clave para analizar y comunicar información compleja, especialmente en temas relacionados con la calidad del aire. Herramientas como Power BI permiten presentar datos de manera interactiva, lo que facilita identificar patrones, tendencias y relaciones entre variables. Estos tableros interactivos, al integrar datos en tiempo real, ofrecen a investigadores y tomadores de decisiones una comprensión más clara de cómo las condiciones meteorológicas impactan los niveles de contaminación (Prada Penagos, 2023).

Uso de Tableros Interactivos en la Ciencia de Datos

Los tableros interactivos son herramientas diseñadas para visualizar y analizar grandes cantidades de datos en un único lugar. Estas aplicaciones permiten a los analistas explorar información y descubrir insights relevantes sin necesidad de realizar consultas manuales en bases de datos (Marchetti, 2023). Un ejemplo destacado es Power BI, una plataforma líder en visualización de datos que ofrece funcionalidades como integración, filtros, personalización y gráficos interactivos (Lombana, 2024). Estudios recientes han demostrado su efectividad en el análisis ambiental, ya que facilita la comparación de variables, como las condiciones

meteorológicas y los niveles de contaminación, proporcionando una visión clara y detallada (Prada Penagos, 2023).

Estudios Previos sobre Visualización de la Calidad del Aire en Colombia

En Colombia el monitoreo y la gestión de la calidad del aire han sido prioritarios para varias entidades gubernamentales y académicas. Investigaciones previas realizadas han demostrado que se han utilizado técnicas de modelación para entender cómo los patrones climáticos influyen en los niveles de contaminación en ciudades específicas, pero pocas han explorado el uso de tableros interactivos como Power BI para este propósito. La incorporación de estas herramientas en la investigación permite no solo observar las condiciones actuales de calidad del aire, sino también comprender patrones históricos y predecir tendencias futuras, lo cual es ideal para la toma de decisiones acertadas (Casallas et al., 2022).

Marco Conceptual

Contaminación del Aire

La contaminación del aire se refiere a las partículas suspendidas en el aire que pueden perjudicar la salud de las personas y del medio ambiente. En este estudio, se consideran principalmente partículas en suspensión (PM2.5, PM10) y gases como el ozono (O₃), los cuales nos indican el nivel de contaminación del aire en las ciudades (Sánchez & Bautista, 2019).

Condiciones Meteorológicas

En el contexto de este estudio, las condiciones meteorológicas incluyen temperatura, humedad, precipitación y velocidad del viento, y todas estas son variables que influyen en la dispersión y concentración de los contaminantes atmosféricos. Estos factores se analizarán para entender como llegan a afectar la calidad del aire en las distintas ciudades de Colombia (García, 2019).

Calidad del Aire

La calidad del aire se mide usando varios índices que reflejan la concentración de contaminantes y el riesgo para la salud. En este caso, el Índice de Calidad del Aire (ICA) nos permite interpretar los niveles de contaminación en un contexto accesible para el público general y los tomadores de decisiones (Navas Vázquez, 2024).

Tablero Interactivo en Power BI

Es una herramienta que permite mostrar e integrar datos de forma interactiva y adaptable. Mediante el uso de gráficos, filtros y mapas, los usuarios pueden explorar cómo se relacionan diferentes variables, como la calidad del aire y las condiciones climáticas, lo que hace que el análisis sea más claro y detallado (Bonifaz & Guamán, 2023).

Análisis de Datos

El análisis de datos implica procesar, organizar y examinar la información para encontrar patrones relevantes que ayuden a responder preguntas específicas. En este caso, el objetivo es identificar posibles relaciones entre las variables meteorológicas y la calidad del aire, además de evaluar cómo Power BI puede contribuir a realizar este tipo de análisis de manera más eficiente (Yupa et al., 2023).

Relación entre Variables

Este estudio analiza cómo se relacionan las condiciones meteorológicas, como la temperatura y la humedad, con los niveles de contaminación del aire, representados por las partículas PM2.5 y PM10. Utilizando un tablero interactivo en Power BI, se busca identificar patrones que muestren cómo los cambios en estas variables influyen en la calidad del aire, especialmente durante períodos de alta o baja precipitación en ciudades colombianas.

Marco Normativo

El marco normativo de esta investigación se fundamenta en las disposiciones legales y técnicas que regulan la gestión de la calidad del aire y el monitoreo ambiental en Colombia. Dichas normas establecen las directrices para la medición, control y difusión de los niveles de contaminación atmosférica, así como las responsabilidades de las entidades encargadas de velar por la salud pública y el equilibrio ambiental.

En primer lugar, la Ley 99 de 1993 crea el Ministerio de Ambiente y organiza el Sistema Nacional Ambiental (SINA), sentando las bases institucionales para la gestión ambiental en el país. Esta ley establece que la prevención y el control de la contaminación atmosférica son deberes del Estado y la sociedad, y define la obligación de realizar monitoreos continuos sobre la calidad del aire como parte de la planificación ambiental (Congreso de la República de Colombia, 1993).

Complementariamente, el Decreto 1076 de 2015, la que compila la normatividad ambiental vigente, regula los procedimientos para la vigilancia y control de la contaminación atmosférica. En este decreto se definen también los límites permisibles de emisión para distintas fuentes contaminantes y también los mecanismos de reporte de información ambiental hacia el IDEAM y las autoridades ambientales regionales (Ministerio de Ambiente y Desarrollo Sostenible, 2015).

Por otro lado, la Resolución 2254 de 2017 del Ministerio de Ambiente y Desarrollo Sostenible establece los niveles de calidad del aire y los niveles de prevención, alerta y emergencia por contaminación del aire en Colombia. Esta resolución define los valores máximos permisibles para contaminantes como el material particulado (PM10 y PM2.5), el ozono (O₃), el dióxido de azufre (SO₂) y el dióxido de nitrógeno (NO₂), entre otros, los cuales sirven de

referencia para el análisis en plataformas como el SISAIRES (Ministerio de Ambiente y Desarrollo Sostenible, 2017).

Asimismo, la Resolución 650 de 2010 define los Protocolos para el Monitoreo y Seguimiento de la Calidad del Aire, los cuales establecen los procedimientos técnicos para la instalación, calibración y operación de las redes de monitoreo. Este protocolo garantiza la validez y comparabilidad de los datos utilizados en estudios ambientales y en herramientas de análisis como el tablero interactivo que se propone en esta investigación (Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2010).

Finalmente, la Política Nacional de Calidad del Aire (PNCA), adoptada en 2010, establece los lineamientos estratégicos para mejorar la calidad del aire en el país. Su objetivo es reducir la exposición de la población a contaminantes atmosféricos mediante la implementación de sistemas de información, indicadores y herramientas de apoyo a la toma de decisiones, lo que se alinea directamente con el propósito de este estudio.

En conjunto, estas normas conforman un marco legal robusto que sustenta la importancia del monitoreo y análisis de la calidad del aire en Colombia. La integración de estos lineamientos en herramientas tecnológicas como Power BI no solo fortalece la transparencia y el acceso a la información, sino que también apoya la formulación de políticas públicas basadas en evidencia científica.

Metodología

Método

El enfoque de esta investigación es cuantitativo y descriptivo, dado que busca analizar datos históricos sobre la calidad del aire y las condiciones meteorológicas en diversas ciudades colombianas, con el propósito de identificar patrones y posibles relaciones entre ambas variables. Este tipo de enfoque permite examinar información numérica y objetiva, lo cual es adecuado para establecer correlaciones estadísticas entre contaminantes atmosféricos y factores climáticos.

El método adoptado corresponde al modelo CRISP-DM (Cross Industry Standard Process for Data Mining), ampliamente utilizado en proyectos de análisis de datos. Esta metodología proporciona un marco estructurado que facilita la comprensión del problema, la preparación de la información y el desarrollo de productos analíticos útiles, como en este caso, un tablero interactivo en Power BI.

Tipo de Estudio

El estudio es no experimental y longitudinal, ya que no se manipularán las variables observadas, sino que se analizarán tal como ocurrieron en un período determinado. Esto implica que la investigación se basa en la observación y análisis de registros históricos sobre calidad del aire y variables meteorológicas, con el fin de comprender su comportamiento a lo largo del tiempo.

Dado que el objetivo principal es describir y analizar las relaciones entre condiciones climáticas (como temperatura, humedad relativa o velocidad del viento) y contaminantes (como PM2.5, PM10 u ozono), el estudio no busca establecer causalidad, sino evidenciar tendencias y correlaciones significativas.

Recolección de Datos

La recolección de datos se realizará a partir de fuentes oficiales y abiertas, principalmente del Sistema de Información sobre Calidad del Aire (SISAIRE) y del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Estas plataformas proporcionan información validada sobre contaminantes atmosféricos y variables meteorológicas registradas en estaciones de monitoreo distribuidas en diferentes ciudades del país.

Las variables seleccionadas incluyen los niveles de PM2.5, PM10, ozono (O₃) y dióxido de nitrógeno (NO₂), junto con indicadores climáticos como temperatura promedio, humedad del aire, velocidad del viento y precipitación.

Finalmente, es importante reconocer que las limitaciones del estudio estarán relacionadas con la disponibilidad y calidad de los datos, ya que estos pueden variar según la ciudad, el año o la continuidad del monitoreo. Asimismo, posibles vacíos en las fuentes públicas pueden afectar el alcance temporal o espacial del análisis.

Resultados

Dashboard IDEAM

La primera página del tablero desarrollado en Power BI presenta la integración y visualización inicial de las variables meteorológicas provenientes del IDEAM, procesadas a partir de cinco conjuntos de datos relacionados con precipitación, temperatura ambiente, temperatura máxima del aire, humedad del aire y velocidad del viento. Estos archivos, descargados en formato CSV desde el portal de Datos Abiertos del IDEAM, fueron sometidos a un proceso de optimización con el fin de reducir su tamaño y facilitar su uso dentro del entorno de Power BI.

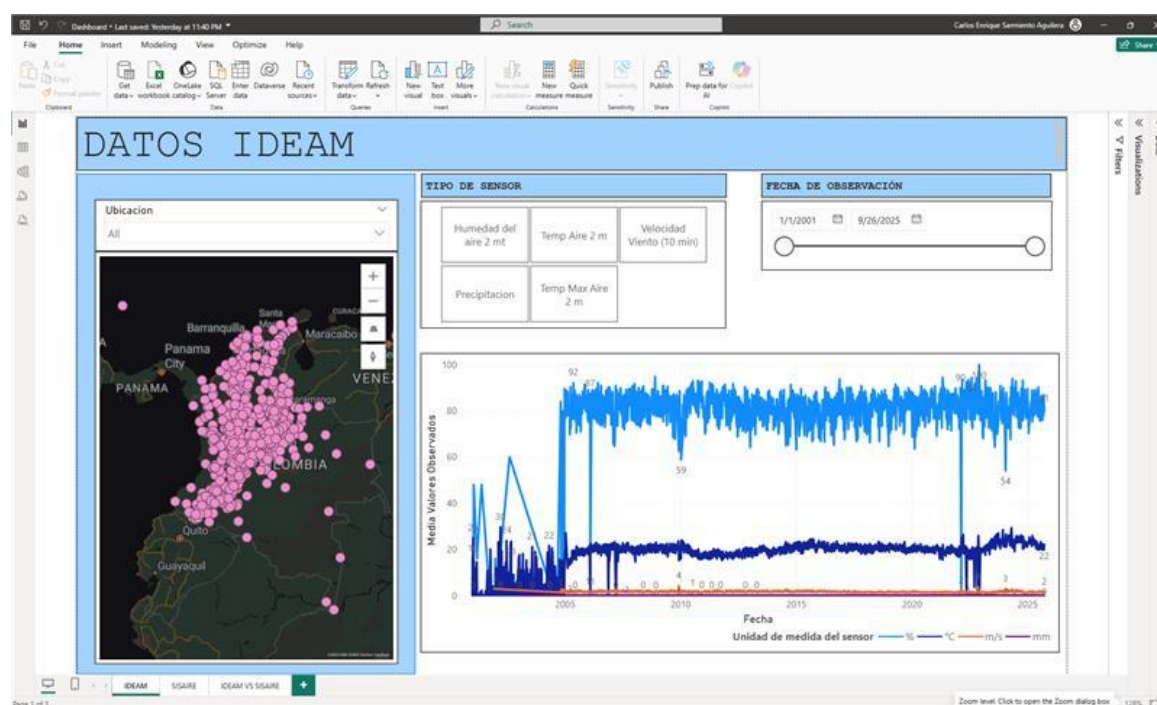
El tratamiento de los datos implicó una conversión inicial a formato Parquet, con el propósito de optimizar el almacenamiento y mejorar la velocidad de lectura en el manejo de grandes volúmenes de información. Este procedimiento se realizó mediante un script desarrollado en DuckDB (véase Apéndice A.1). Posteriormente, cada archivo fue agregado por día, debido a que las estaciones meteorológicas generan múltiples observaciones en una misma fecha, lo cual incrementaba el tamaño y complejidad de los archivos. Al consolidar los valores a través de promedios diarios, sumatorias y conteos, fue posible reducir significativamente el volumen de datos (véase Apéndice A.2).

Una vez completadas estas transformaciones, los cinco archivos resultantes en formato Parquet fueron unificados en un solo archivo maestro mediante un script en Python que utilizó la librería pyarrow. Este paso de unificación (merge) se realizó antes de la importación a Power BI, con el fin de trabajar sobre un único dataset optimizado, evitando operaciones de combinación dentro del entorno del dashboard y mejorando así el rendimiento general del informe. El código empleado para esta consolidación se presenta en el Apéndice A.3.

Este archivo combinado constituyó la base para la construcción de la página inicial del tablero, donde se integraron filtros por estación, rangos de tiempo y visualizaciones descriptivas que permiten explorar el comportamiento de las variables meteorológicas monitoreadas por el IDEAM en las estaciones de interés. En el dashboard se creó una columna adicional denominada “Ubicación”, que combina municipio y departamento para facilitar el mapa interactivo.

Figura 1

Dashboard IDEAM



La interfaz de esta primera página fue diseñada para posibilitar una exploración inicial de los datos meteorológicos registrados por las estaciones del IDEAM. Para ello, se implementó un slicer que permite filtrar las observaciones según la ubicación seleccionada por el usuario, así como un mapa interactivo de Azure Maps en el que se visualizan las estaciones georreferenciadas en el territorio colombiano. Esta representación facilita identificar la distribución espacial de los sensores y su relación con las distintas regiones del país.

Adicionalmente, se incorporó un slicer de tipo tile que permite seleccionar la variable meteorológica de interés (por ejemplo, humedad relativa, temperatura ambiente, precipitación o velocidad del viento), habilitando la visualización dinámica de cada variable según los requerimientos del análisis. También se incluyó un slicer de tipo *between* que permite delimitar el rango temporal en estudio, de modo que el usuario pueda concentrarse en periodos específicos de interés histórico o ambiental.

Finalmente, la visualización principal de esta página corresponde a un gráfico de líneas que muestra la evolución temporal de los valores promedio diarios de la variable seleccionada. Esta gráfica permite identificar tendencias, picos, caídas y patrones de variabilidad en las condiciones meteorológicas del país a lo largo de los años registrados por el IDEAM. En conjunto, la primera página del dashboard funciona como una capa exploratoria que facilita la navegación, comprensión y selección de los datos climáticos antes de continuar con análisis más profundos.

Dashboard SISAIRE

La segunda página del dashboard está enfocada en la visualización de los datos provenientes del Sistema de Información sobre Calidad del Aire – SISAIRE, administrado por el IDEAM. Para construir esta visualización fue necesario realizar un proceso de recolección, limpieza, estandarización y consolidación de datos, debido a que la información disponible en la plataforma no se encuentra unificada y requiere consultas individuales por departamento, municipio, estación y contaminante.

El proceso inició con la consulta de la plataforma SISAIRE (<http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml>), identificando primero cuáles departamentos del país cuentan con estaciones de monitoreo activas. Esto se logró mediante un

script de web scraping (véase Apéndice B.1), el cual generó un archivo CSV con la lista de departamentos disponibles.

Tabla 1

Primeras Filas del Archivo departamentos_sisaire.csv

id	nombre departamento	codigo_departamento
1	ANTIOQUIA	5
2	ARAUCA	81
3	ARCHIPIELAGO DE SAN ANDRES, PROVIDENCIA Y SANTA CATALINA	88
4	ATLÁNTICO	8
5	BOGOTÁ D.C.	11

Nota. La columna codigo_departamento es esencial para realizar la consulta de los municipios.

Tabla 2

Primeras Filas del Archivo municipios_sisaire.csv

id	nombre_municipio	codigo_municipio	departamento_id
1	AMAGÁ	05030	1
2	AMALFI	05031	1
3	ANDES	05034	1
4	ANGELÓPOLIS	05036	1
5	ANGOSTURA	05038	1

Nota. La columna codigo_municipio es esencial para realizar la consulta de las estaciones.

Posteriormente, se elaboró un segundo script (véase Apéndice B.2) que, para cada departamento, consultó los municipios asociados a la red de monitoreo, produciendo así un segundo archivo con el listado de todos los municipios que cuentan con estaciones registradas en SISAIRES (véase tabla 2).

En un tercer paso, se desarrolló un script adicional (véase Anexo B.3) que permitió obtener el conjunto completo de estaciones disponibles para cada municipio. Este procedimiento generó un archivo consolidado con los nombres de las estaciones, sus códigos y su asociación municipal y departamental, información que posteriormente se utilizaría para enriquecer los datos de contaminantes con metadatos geográficos.

Tabla 3

Primeras Filas del Archivo estaciones_sisaires.csv

id	nombre_estacion	codigo_estacion	municipio_id
1	AMA-FER	31849	1
2	AMA-GUA	31880	1
3	CA. AMAGA 2011	30421	1
4	CA. AMAGA 2013	30433	1
5	CA. AMAGA 2016	30715	1

Nota. La columna nombre_estacion es esencial para realizar el merge con los registros descargados de cada estación.

Una vez identificada la estructura territorial del sistema (departamento → municipio → estación), se procedió a la descarga manual de los datos de concentración de los contaminantes O₃, NO₂, PM₁₀ y PM_{2.5} para cada uno de los departamentos identificados. Este proceso generó un

total de 78 archivos CSV independientes, debido a que SISAIRE solo permite descargar un contaminante a la vez para un rango de fechas determinado. Los archivos obtenidos presentaban estructuras similares, con columnas que variaban dependiendo del contaminante reportado.

Para unificar esta información, se desarrolló un script de consolidación (véase Anexo B.4) que integró los 78 archivos en un único dataset. El procedimiento incluyó la identificación automática del contaminante correspondiente en cada archivo, la reorganización de la tabla para normalizar su estructura y la posterior transformación a un formato ancho (wide format) mediante una tabla dinámica. El resultado fue un archivo único con columnas para O₃, NO₂, PM₁₀ y PM_{2.5}, acompañado de su estación de origen y la fecha de medición.

Tabla 4

Primeras Filas del Archivo datos_comtaminantes_sisaire.csv

Estacion	Fecha inicial	NO2	O3	PM10	PM2.5
USME	2020-10-01	23.80273	33.73005	25.80416667	7.68
USME	2020-10-02	31.4967	23.13376364	32.3125	12.55714286
USME	2020-10-03	25.299215	30.811575	27.3375	14.18333333
USME	2020-10-04	20.942955	33.950775	10.05833333	7.63809524
USME	2020-10-05	25.21303	38.872125	24.46666667	11.62916667

Nota. El nombre de las columnas viene tal como se descarga en el portal de SISAIRE.

Una vez consolidado este archivo principal, se procedió a enriquecer la información mediante la asociación de cada estación con su municipio y su departamento correspondiente. Esto se realizó mediante un proceso de normalización y fusión de datos (véase Anexo B.5), empleando coincidencias entre nombres de estaciones, municipios y departamentos previamente

obtenidos durante el proceso de scraping. Este paso me permitió generar un archivo final en el que cada observación de concentración de contaminantes se encuentra vinculada a su ubicación geográfica dentro del país.

El resultado es un dataset unificado con la siguiente estructura: estación, fecha, valores de NO₂, O₃, PM₁₀, PM_{2.5}, departamento y municipio. Este archivo final, denominado SISAIRES_MERGE, fue importado directamente en Power BI para construir la segunda página del dashboard. En dicha página se presentan visualizaciones que permiten analizar la distribución temporal y espacial de los contaminantes atmosféricos en Colombia, lo que facilita la comparación entre regiones y la identificación de tendencias relevantes para el estudio.

Tabla 5

Primeras Filas del Archivo sisaire_merge.csv

Estacion	Fecha inicial	NO2	O3	PM10	PM2.5	nombre_departamento	nombre municipio
USME	2020-10-01	23.80273	33.73005	25.8041666	7.68	BOGOTÁ D.C.	BOGOTÁ D.C.
USME	2020-10-02	31.4967	23.13376364	32.3125	12.55714286	BOGOTÁ D.C.	BOGOTÁ D.C.
USME	2020-10-03	25.299215	30.811575	27.3375	14.18333333	BOGOTÁ D.C.	BOGOTÁ D.C.
USME	2020-10-04	20.942955	33.950775	10.0583333	7.63809524	BOGOTÁ D.C.	BOGOTÁ D.C.
USME	2020-10-05	25.21303	38.872125	24.4666666	11.62916667	BOGOTÁ D.C.	BOGOTÁ D.C.

Nota. En esta tabla ya se encuentra referenciada cada estación con su ubicación geográfica, lo cual es demasiado útil para el dashboard que se va a realizar en PowerBI

Con todo lo anterior, ya se procede a realizar la segunda página del tablero desarrollado en Power BI que corresponde a la visualización del conjunto de datos provenientes del Sistema de Información sobre Calidad del Aire (SISAIRE). Esta sección fue diseñada para facilitar la exploración, comparación y análisis de los niveles de contaminación atmosférica en Colombia, donde se integra múltiples elementos interactivos que permiten ajustar la visualización de acuerdo con las necesidades del usuario.

En primer lugar, se incluyó un slicer que permite filtrar la información por ubicación, la cual combina municipio y departamento para ofrecer una búsqueda más intuitiva de las estaciones.

Adicionalmente, se incorporó un mapa interactivo mediante Azure Maps, donde se muestran las estaciones de monitoreo distribuidas en el país. Cada punto georreferenciado representa un municipio del país que cuenta con estaciones de monitoreo y permite visualizar, de manera espacial, la cobertura del sistema de monitoreo de calidad del aire.

Para complementar la navegación, se añadió un slicer de tipo Dropdown que permite seleccionar una estación específica dentro del conjunto de datos. Esto posibilita realizar análisis focalizados en un punto de monitoreo particular.

Asimismo, se implementó un slicer de tipo Between para filtrar los registros según un rango de fechas. Con esta herramienta el usuario puede restringir la visualización a periodos específicos, lo cual resulta útil para revisar temporadas críticas, episodios de contaminación, o tendencias a lo largo de los años.

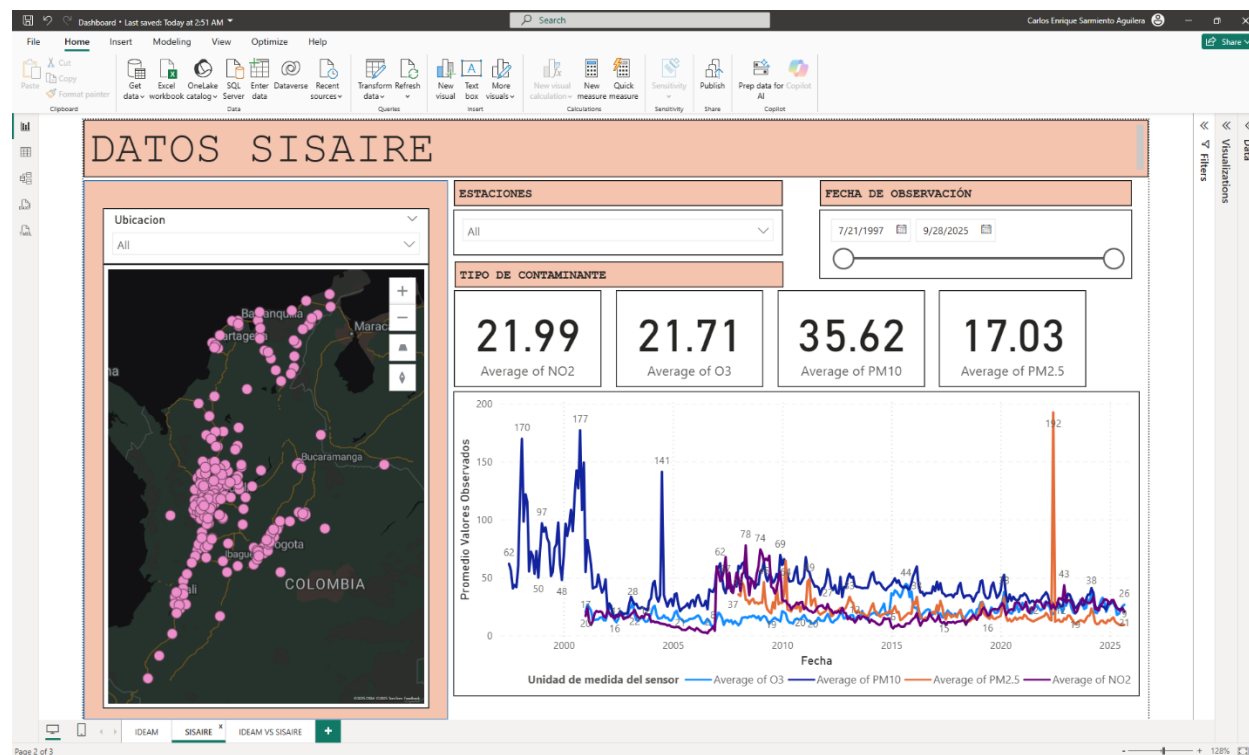
En la parte superior de la página se ubicaron cuatro tarjetas (Cards) diseñadas para mostrar el valor promedio de cada uno de los contaminantes registrados: NO₂, O₃, PM₁₀ y PM_{2.5}. Estas tarjetas proporcionan un resumen inmediato del comportamiento general de los

contaminantes según los filtros seleccionados, permitiendo una lectura rápida del estado de la calidad del aire.

Finalmente, el componente principal de esta página es un gráfico de líneas (Line Chart) que representa los valores promedio diarios de los cuatro contaminantes a lo largo del tiempo. El eje horizontal corresponde a la variable *Fecha Inicial*, mientras que el eje vertical muestra los valores observados. Esta visualización permite detectar comportamientos históricos, patrones estacionales, picos inesperados, variaciones entre contaminantes y posibles tendencias de mejora o deterioro en la calidad del aire.

Figura 2

Dashboard SISAIRE



Dashboard IDEAM VS SISAIRE

La tercera página del tablero corresponde al análisis comparativo entre los valores de calidad del aire provenientes del sistema SISAIRE y las variables meteorológicas registradas por las estaciones del IDEAM. El propósito de esta sección es permitir una visualización integrada que facilite identificar posibles relaciones entre el comportamiento de los contaminantes atmosféricos y la presencia de condiciones meteorológicas específicas.

Para construir este Dashboard ha sido necesario generar previamente un conjunto de datos unificado que contuviera únicamente los municipios y fechas donde existieran registros simultáneos tanto del IDEAM como de SISAIRE. Para ello se desarrolló un script en Python orientado a realizar un proceso de emparejamiento (JOIN) entre ambas fuentes de información (véase Apéndice C). El procedimiento incluyó la carga de los archivos previamente transformados, la normalización de texto para evitar inconsistencias en nombres de municipios y departamentos, la conversión de fechas a un formato uniforme y la aplicación de sustituciones específicas en casos donde existían variaciones reales en la denominación de algunos municipios. Una vez armonizados los campos, se ejecutó un INNER JOIN sobre las columnas de departamento, municipio y fecha, obteniendo así un archivo en formato Parquet que contiene únicamente las intersecciones válidas entre ambas bases.

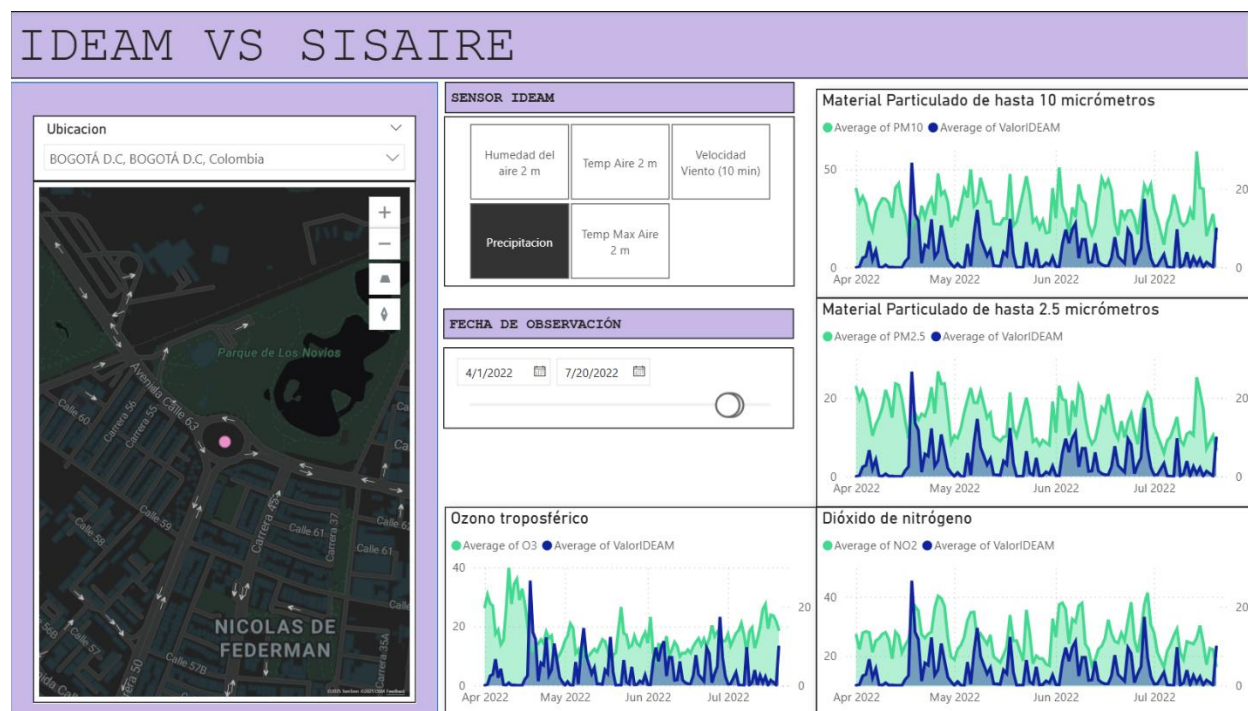
El archivo resultante se cargó posteriormente a Power BI mediante la conexión a SharePoint, lo que permitió construir la página comparativa del dashboard. La interfaz incluye un slicer para filtrar la ubicación, así como un mapa interactivo de Azure Maps que muestra las estaciones georreferenciadas. Adicionalmente, se implementó un slicer de tipo tile que permite seleccionar el sensor meteorológico de interés utilizando la columna `DescripcionSensor`, y un

slicer de tipo between que habilita el filtrado por rangos de fechas, facilitando el análisis temporal de la relación entre ambas fuentes.

El componente principal de este dashboard está conformado por cuatro gráficos de líneas, uno por cada contaminante relevante: NO_2 , PM_{10} , $\text{PM}_{2.5}$ y O_3 . Cada gráfico presenta en el eje X la fecha de observación sin jerarquías temporales, mientras que en el eje Y se muestra el valor promedio diario del contaminante registrado por SISAIRE y de manera complementaria, se incorpora un eje secundario que representa el promedio del parámetro meteorológico proveniente del IDEAM correspondiente al sensor que se seleccione. Esta configuración permite una visualización simultánea y comparativa, facilitando la identificación de patrones, tendencias y posibles correlaciones entre la variación de los contaminantes atmosféricos y el comportamiento de las variables meteorológicas asociadas.

Figura 3

Dashboard IDEAM VS SISAIRE



Dashboard Mapa Coroplético

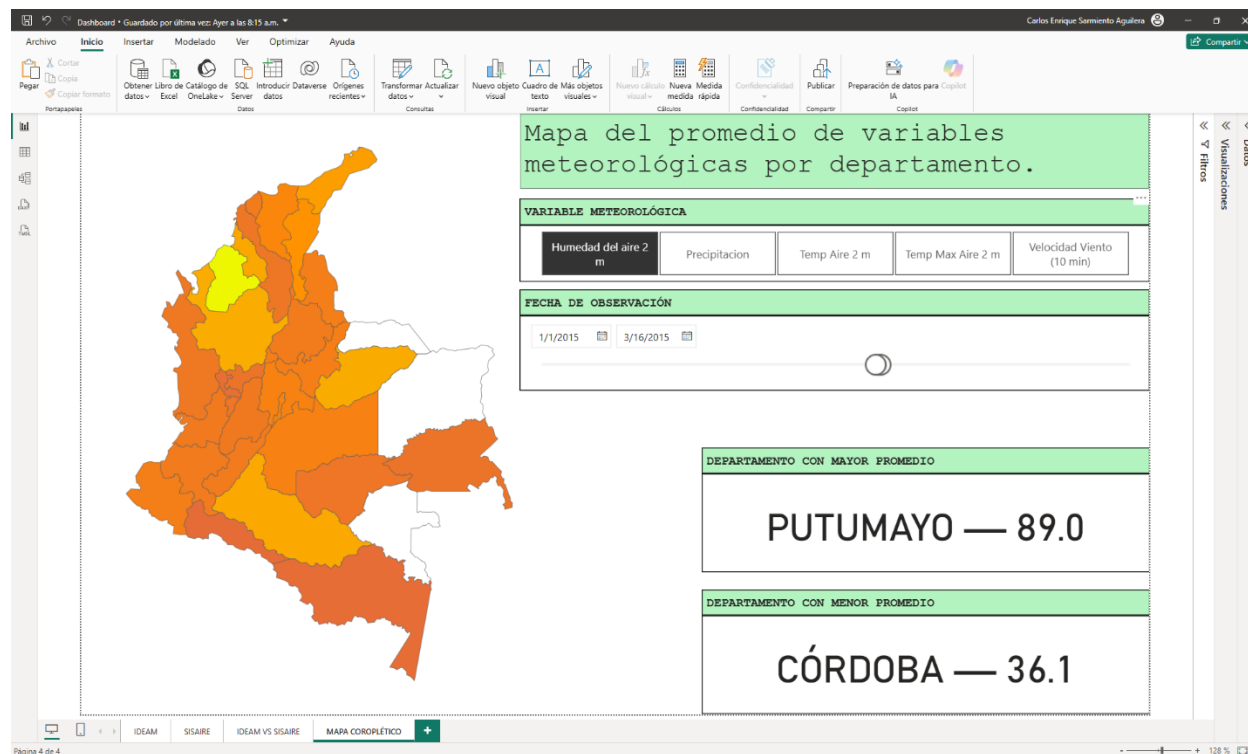
Para complementar el análisis y facilitar la visualización espacial de las variables meteorológicas, se construye un dashboard basado en un mapa de formas. Para poder usar este tipo de visualización se debe habilitar previamente la opción de mapas personalizados en la configuración de Power BI, y luego descargar un archivo GeoJSON con los límites departamentales de Colombia. Ese archivo lo convertí a formato JSON compatible y lo cargué dentro del panel de visualizaciones, para posteriormente asociarlo con la columna Departamento_Final de la tabla del IDEAM. Con esto el mapa pudo reconocer cada departamento y colorearlo según los valores seleccionados.

Después de preparar el mapa, incluí un control de segmentación (slicer) basado en la columna DescripcionSensor, que corresponde a todas las variables meteorológicas disponibles (humedad, precipitación, temperatura del aire, etc.). Este slicer permite cambiar dinámicamente la variable que se quiere analizar, de modo que el mapa se actualiza automáticamente para mostrar el promedio departamental de la variable seleccionada. También agregué un segundo slicer basado en la columna Fecha, con el cual es posible definir un rango de tiempo específico para observar cómo cambian los promedios en el mapa durante un periodo concreto.

Para resaltar los departamentos con valores extremos, construí dos medidas en DAX. La primera identifica el departamento con el mayor valor promedio, y la segunda calcula el departamento con el menor valor promedio, siempre teniendo en cuenta los filtros activos. Estas medidas se muestran en dos tarjetas dentro del dashboard, lo que permite identificar rápidamente qué región presenta las condiciones meteorológicas más altas o bajas según la variable que se esté analizando.

Figura 4

Dashboard Mapa Coroplético



Análisis Estadístico de la Relación entre Variables Meteorológicas y Contaminantes

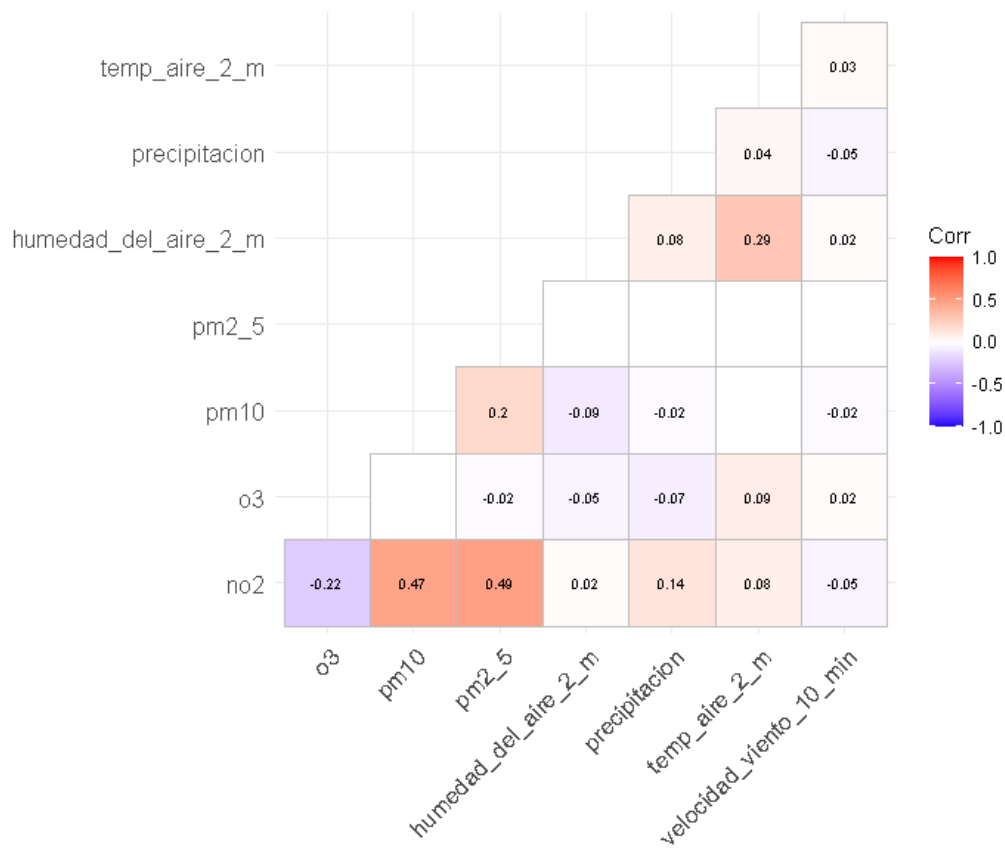
Para esta parte del trabajo lo que hice fue tomar toda la base de datos que tenía integrada entre IDEAM y SISAIRES y luego filtrarla únicamente para el año 2021, porque la idea era trabajar solo con un periodo específico para que los cálculos fueran más manejables. Una vez filtré la información, lo primero que hice en el script de R fue organizar las columnas, estandarizar nombres y convertir todo lo que tenía que ver con fechas al formato correcto (véase Apéndice D).

Cuando tuve todo limpio, pasé a la parte de correlaciones. El script generó dos archivos: uno con los coeficientes de correlación (`correlacion_r.csv`) y otro con los p-values (`correlacion_pvalues.csv`), además de un gráfico (Figura 4). En el gráfico, la relación más visible

es la correlación negativa entre la velocidad del viento y NO_2 , lo cual respalda la idea de que el viento facilita la dispersión de gases contaminantes. Para $\text{PM}_{2.5}$ y PM_{10} este efecto también aparece, pero con menor magnitud. La temperatura presenta asociaciones leves con los gases, reforzando los patrones observados en los modelos GAM: el ozono tiende a aumentar con temperaturas mayores, mientras que NO_2 tiende a reducirse. La precipitación muestra correlaciones muy bajas con todos los contaminantes, lo que sugiere que sus efectos en este conjunto de datos no son lineales o están enmascarados por otros factores, como la frecuencia de lluvias o las características locales de emisión. Estos resultados muestran que las relaciones meteorológicas no son tan directas ni tan lineales como uno pensaría.

Figura 5

Correlación entre Variables Meteorológicas y Contaminantes



En la parte de correlaciones básicamente encontré que la relación entre los contaminantes entre sí era mayor que la relación con las variables meteorológicas. Por ejemplo, NO_2 tenía una correlación moderada con PM_{10} y $\text{PM}_{2.5}$, lo cual tiene sentido porque muchas veces vienen de las mismas fuentes. Pero al comparar los contaminantes con la temperatura, humedad, precipitación o el viento, las correlaciones eran muy bajas, casi siempre menores a 0.1. Aunque los p-values salieron muy pequeños por el tamaño de la muestra, eso no significa que la relación sea fuerte, sino solamente que con tantos datos cualquier correlación mínima termina siendo significativa.

Tabla 6*Matriz de Correlación*

	no2	o3	pm10	pm2_5	humedad_del_aire	precipitación	temp_aire	velocidad_viento
no2	1	-0,2171	0,4695	0,4947	0,022	0,139	0,080	-0,054
o3	-0,217	1	0,0122	-0,02	-0,047	-0,066	0,086	0,022
pm10	0,4695	0,0122	1	0,201	-0,090	-0,020	-0,005	-0,017
pm2_5	0,4947	-0,024	0,2011	1	0,001	-0,003	0,009	-0,004
humedad_del_aire	0,0222	-0,047	-0,0908	0,001	1	0,083	0,289	0,022
precipitación	0,1398	-0,066	-0,020	-0,003	0,083	1	0,039	-0,047
temp_aire	0,0800	0,0866	-0,005	0,009	0,289	0,039	1	0,028
velocidad_viento	-0,054	0,0223	-0,017	-0,004	0,022	-0,047	0,028	1

Nota. En esta tabla se muestra la correlación de Pearson entre las concentraciones de los contaminantes (NO₂, O₃, PM₁₀ y PM_{2.5}) y las variables meteorológicas analizadas (humedad relativa, precipitación, temperatura del aire y velocidad del viento). Los valores cercanos a 1 o -1 indican relaciones fuertes, mientras que los valores próximos a 0 representan correlaciones débiles o casi inexistentes. Esta matriz corresponde exclusivamente a los datos filtrados del año 2021.

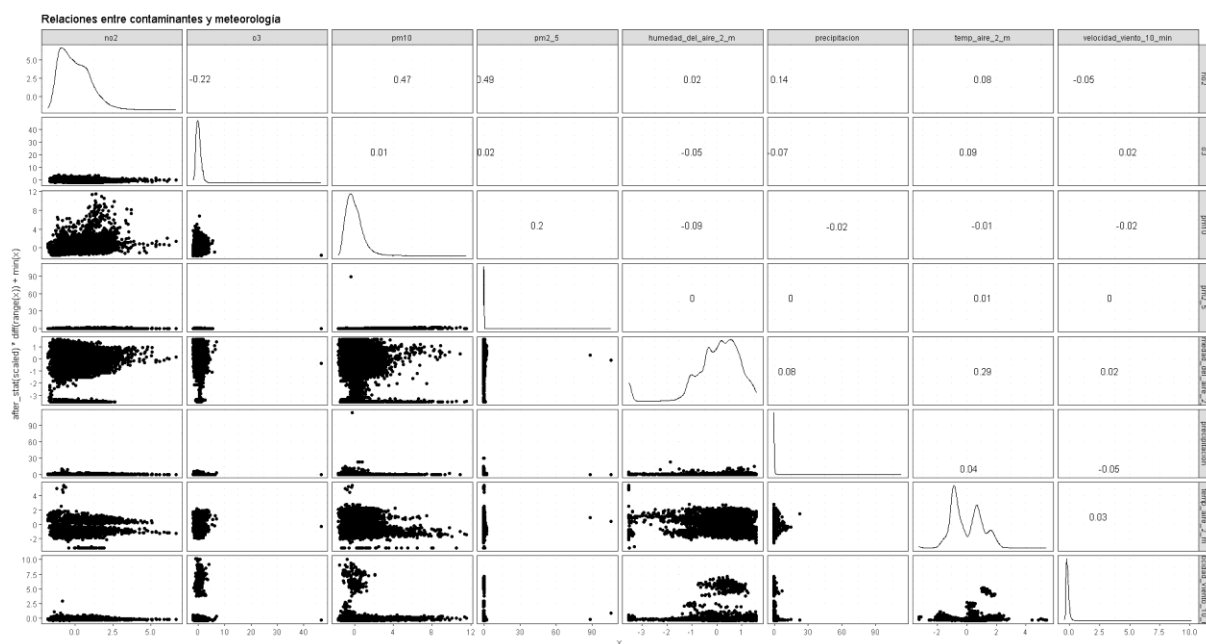
Figura 6*Matriz de Valores p*

	no2	o3	pm10	pm2_5	humedad del aire 2 m	precipitacion	temp aire 2 m	velocidad viento 10 min
no2	1	0	0	0	0,0214	0	4,E-16	2,1,E-08
o3	0	1	0,2492	0,0223	4,E-07	7,E-13	0	0,016
pm10	0	0,2492	1	0	0	0,0036	0,4131	0,0121
pm2_5	0	0,0223	0	1	0,8469	0,6177	0,2055	0,5422
humedad_del_aire_2_m	0,0214	4,34,E-07	0	0,8469	1	0	0	0,0002
precipitacion	0	7,31,E-13	0,0036	0,6177	0	1	4,70,E-10	1,24,E-14
temp_aire_2_m	4,44,E-16	0	0,4131	0,2055	0	4,70,E-10	1	4,09,E-06
velocidad_viento_10_min	2,06,E-08	0,0161	0,0121	0,5422	0,0002	1,24,E-14	4,09,E-06	1

Nota. Matriz de valores p asociados a las correlaciones entre contaminantes y variables meteorológicas. Esta tabla presenta los valores p obtenidos para cada correlación de la matriz anterior. Estos valores permiten evaluar si la correlación observada es estadísticamente significativa. En general, valores p menores a 0,05 indican que la relación entre las variables es significativa. Los valores muy cercanos a cero (por ejemplo, en notación científica como 1,24E-14) reflejan una evidencia estadística muy fuerte. Esta matriz corresponde únicamente a los datos filtrados del año 2021.

Figura 7

Relaciones entre Contaminantes y Meteorología



Del anterior gráfico (Figura 4) podemos decir que corresponde a una matriz de relaciones entre los contaminantes atmosféricos (NO_2 , O_3 , PM_{10} y $\text{PM}_{2.5}$) y las variables meteorológicas seleccionadas (temperatura del aire, precipitación, velocidad del viento y humedad del aire). El diagrama combina histogramas, diagramas de dispersión y coeficientes de correlación, lo que me permite obtener una visión general de cómo interactúan estas variables.

En primer lugar, los histogramas diagonales evidencian que todos los contaminantes presentan distribuciones inclinadas hacia valores bajos, con presencia de algunos picos altos aislados. Esto es típico de contaminantes que suelen tener niveles moderados la mayor parte del tiempo pero que registran episodios puntuales de aumento. Las variables meteorológicas muestran distribuciones más amplias y continuas, especialmente la temperatura y la velocidad del viento.

En cuanto a las correlaciones numéricas podemos observar que la mayoría de las asociaciones entre contaminantes y variables meteorológicas son débiles, lo que ya sugiere la ausencia de relaciones estrictamente lineales. Sin embargo, se identifica por ejemplo que el NO_2 presenta una ligera correlación negativa con la velocidad del viento, lo cual es coherente con el efecto dispersor del viento sobre los contaminantes primarios procedentes del tráfico. También se aprecia una leve relación con la temperatura: a mayor temperatura, los niveles de NO_2 tienden a disminuir, lo cual coincide con el comportamiento esperado por procesos fotoquímicos.

Por otro lado, el ozono (O_3) muestra patrones opuestos a los de NO_2 ya que se nota una tendencia a incrementarse con temperaturas más altas, lo cual es consistente con la naturaleza secundaria del O_3 , que se forma por reacciones fotoquímicas favorecidas por radiación solar y calor. La dispersión de los puntos respalda esta observación, ya que la nube de puntos muestra cierta inclinación ascendente.

Los contaminantes particulados (PM_{10} y $\text{PM}_{2.5}$) exhiben correlaciones muy débiles con las variables meteorológicas, aunque en ambos casos la velocidad del viento parece tener un efecto ligero pero perceptible: cuando el viento aumenta, los niveles de material particulado tienden a reducirse o a dispersarse, lo cual también coincide con lo observado en los modelos GAM.

Después de eso analicé el archivo del VIF. Esto lo hice para verificar si había multicolinealidad entre las variables meteorológicas, porque si las variables estuvieran muy relacionadas entre sí eso dañaría los modelos. Pero en realidad los valores de VIF salieron muy bajos, casi todos entre 1 y 1.3, lo cual es perfecto. Eso quiere decir que las variables que usé no estaban interfiriendo entre ellas y que las podía usar sin problema tanto en los modelos lineales como en los no lineales.

Tabla 7*VIF para cada Contaminante y sus Variables Meteorológicas Asociadas*

Contaminante	Variable	VIF
no2	humedad_del_aire_2_m	1,27855
no2	precipitacion	1,06837
no2	temp_aire_2_m	1,05728
no2	velocidad_viento_10_min	1,38008
o3	humedad_del_aire_2_m	1,19413
o3	precipitacion	1,06300
o3	temp_aire_2_m	1,16669
o3	velocidad_viento_10_min	1,04152
pm10	humedad_del_aire_2_m	1,08927
pm10	precipitacion	1,01619
pm10	temp_aire_2_m	1,05336
pm10	velocidad_viento_10_min	1,05814
pm2_5	humedad_del_aire_2_m	1,12572
pm2_5	precipitacion	1,04383
pm2_5	temp_aire_2_m	1,06785
pm2_5	velocidad_viento_10_min	1,07629

Nota. Esta tabla muestra los valores del VIF (Índice de Inflación de la Varianza) que permite identificar si existe multicolinealidad entre las variables y con estos resultados se confirma que las variables meteorológicas usadas en los modelos no generan redundancia y pueden incluirse sin afectar la estabilidad de los análisis. Año 2021.

Luego seguí con los modelos lineales. El script ajustó un modelo por cada contaminante usando como predictores las variables meteorológicas del año 2021, y todo quedó guardado en el archivo `resumen_lineal_avanzado.csv`. Lo que encontré allí fue que los modelos lineales no explicaban mucho de la variabilidad. Los valores de R^2 estaban muy bajos, en varios casos ni siquiera llegaban al 10%. Eso confirma un poco lo que ya se veía en las correlaciones: las relaciones entre contaminación y meteorología no son tan fuertes como para que un modelo lineal las capture bien. En algunos casos sí había predictores estadísticamente significativos, pero sus efectos eran pequeños.

Como los modelos lineales no funcionaron muy bien, pasé a probar modelos no lineales, específicamente modelos polinomiales. Estos quedaron guardados en el archivo `resumen_no_lineal_avanzado.csv`. Con esto se esperaba ver si había alguna relación curvada o algún comportamiento que no fuera estrictamente lineal. Aunque sí hubo una mejora sobre los modelos lineales, todavía los R^2 seguían siendo bajos. En pocas palabras, los modelos polinomiales capturan un poco mejor la forma de los datos, pero aún así no explican lo suficiente como para considerarlos buenos modelos predictivos en este caso.

Figura 8*Resultados de los Modelos Lineales*

term	estimate	std.error	statistic	p.value	modelo	r2	r2_ajustado	shapiro_p	ad_p	breusch_p	durbin_p	AIC	BIC
(Intercept)	-0,0212	0,0235	-0,9022	0,3670	Lineal	0,0551	0,0547	NA	3.7e-24	8,E-100	2,E-151	26235,8	26278,8
humedad_del_aire_2_m	-0,0656	0,0158	-4,1621	3,2E-05	Lineal	0,0551	0,0547	NA	3.7e-24	8,E-100	2,E-151	26235,8	26278,8
precipitacion	0,4240	0,0300	14,1486	5,5E-45	Lineal	0,0551	0,0547	NA	3.7e-24	8,E-100	2,E-151	26235,8	26278,8
temp_aire_2_m	0,1414	0,0152	9,2807	2,1E-20	Lineal	0,0551	0,0547	NA	3.7e-24	8,E-100	2,E-151	26235,8	26278,8
velocidad_viento_10_min	-1,1566	0,1144	-10,1093	6,6E-24	Lineal	0,0551	0,0547	NA	3.7e-24	8,E-100	2,E-151	26235,8	26278,8
(Intercept)	0,0497	0,0108	4,6023	4,2E-06	Lineal	0,0265	0,0261	NA	3.7e-24	0,8645	3,E-138	29927,8	29971,4
humedad_del_aire_2_m	-0,1053	0,0143	-7,3651	1,9E-13	Lineal	0,0265	0,0261	NA	3.7e-24	0,8645	3,E-138	29927,8	29971,4
precipitacion	-0,1319	0,0231	-5,7134	1,1E-08	Lineal	0,0265	0,0261	NA	3.7e-24	0,8645	3,E-138	29927,8	29971,4
temp_aire_2_m	0,1931	0,0123	15,6766	9,1E-55	Lineal	0,0265	0,0261	NA	3.7e-24	0,8645	3,E-138	29927,8	29971,4
velocidad_viento_10_min	-0,0906	0,0263	-3,4405	0,0006	Lineal	0,0265	0,0261	NA	3.7e-24	0,8645	3,E-138	29927,8	29971,4
(Intercept)	-0,1139	0,0101	-11,2440	3,1E-29	Lineal	0,0216	0,0214	NA	3.7e-24	2,E-11	1,E-205	48275,9	48322,7
humedad_del_aire_2_m	-0,0714	0,0089	-8,0258	1,1E-15	Lineal	0,0216	0,0214	NA	3.7e-24	2,E-11	1,E-205	48275,9	48322,7
precipitacion	-0,0540	0,0115	-4,6821	2,9E-06	Lineal	0,0216	0,0214	NA	3.7e-24	2,E-11	1,E-205	48275,9	48322,7
temp_aire_2_m	0,0479	0,0069	6,9759	3,1E-12	Lineal	0,0216	0,0214	NA	3.7e-24	2,E-11	1,E-205	48275,9	48322,7
velocidad_viento_10_min	-0,8105	0,0451	-17,9910	9,7E-72	Lineal	0,0216	0,0214	NA	3.7e-24	2,E-11	1,E-205	48275,9	48322,7
(Intercept)	0,0285	0,0130	2,1881	0,0287	Lineal	0,0005	0,0002	NA	3.7e-24	6,E-08	0,0947	49636,6	49682,8
humedad_del_aire_2_m	0,0136	0,0120	1,1305	0,2583	Lineal	0,0005	0,0002	NA	3.7e-24	6,E-08	0,0947	49636,6	49682,8
precipitacion	0,0049	0,0171	0,2889	0,7727	Lineal	0,0005	0,0002	NA	3.7e-24	6,E-08	0,0947	49636,6	49682,8
temp_aire_2_m	0,0140	0,0098	1,4287	0,1531	Lineal	0,0005	0,0002	NA	3.7e-24	6,E-08	0,0947	49636,6	49682,8
velocidad_viento_10_min	0,1255	0,0655	1,9150	0,0555	Lineal	0,0005	0,0002	NA	3.7e-24	6,E-08	0,0947	49636,6	49682,8

Nota. En esta tabla se muestran los resultados de los modelos lineales usados para analizar cómo las variables meteorológicas del IDEAM (humedad, precipitación, temperatura y viento) se relacionan con cada contaminante. Se incluyen los coeficientes, los valores p y las métricas del modelo (R^2 , AIC y BIC), que permiten ver qué tan fuerte es la relación y si es estadísticamente significativa. Esta tabla resume el comportamiento general de cada contaminante frente a las condiciones del clima durante el año 2021.

Tabla 8*Resultados de los Modelos No Lineales con Interacciones – Año 2021*

term	estimat e	std.err or	statisti c	p.value	modelo	r2	r2_ajustad o	AIC	BIC
(Intercept)	0,4568	0,0301	15,176	2,0E-51	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
humedad_del_aire_2_m	-0,127	0,0162	-7,838	5,0E-15	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
precipitacion	0,7447	0,0870	8,5567	1,3E-17	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
temp_aire_2_m	0,0515	0,0149	3,4473	0,001	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
velocidad_viento_10_min	-1,058	0,1452	-7,289	3,4E-13	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
I(humedad_del_aire_2_m^2)	-0,352	0,0143	-24,59	1,4E- 129	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
I(precipitacion^2)	-0,197	0,0254	-7,775	8,3E-15	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
I(temp_aire_2_m^2)	-0,263	0,0172	-15,32	2,2E-52	No lineal + Interacción	0,168	0,1676	25035,4	25114,1

term	estimat e	std.err or	statisti c	p.value	modelo	r2	r2_ajustad o	AIC	BIC
I(velocidad_viento_10_min^2)	-2,136	0,4472	-4,778	1,8E-06	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
precipitacion:velocidad_viento_10 _min	0,0843	0,4337	0,1944	0,846	No lineal + Interacción	0,168	0,1676	25035,4	25114,1
(Intercept)	-0,211	0,0263	-8,061	8,4E-16	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
humedad_del_aire_2_m	-0,158	0,0149	-10,62	3,0E-26	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
precipitacion	-0,373	0,0366	-10,22	1,9E-24	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
temp_aire_2_m	0,1812	0,0125	14,461	6,0E-47	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
velocidad_viento_10_min	-1,219	0,0885	-13,77	8,1E-43	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
I(humedad_del_aire_2_m^2)	-0,019	0,0129	-1,480	0,139	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
I(precipitacion^2)	0,1367	0,0183	7,4542	9,7E-14	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
I(temp_aire_2_m^2)	0,0588	0,0163	3,6146	0,000	No lineal + Interacción	0,046	0,0455	29721,3	29801,2

term	estimat e	std.err or	statisti c	p.value	modelo	r2	r2_ajustad o	AIC	BIC
I(velocidad_viento_10_min^2)	0,2852	0,0215	13,241	1,0E-39	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
precipitacion:velocidad_viento_10 _min	-0,089	0,0250	-3,585	0,000	No lineal + Interacción	0,046	0,0455	29721,3	29801,2
(Intercept)	-0,053	0,0182	-2,917	0,004	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
humedad_del_aire_2_m	-0,089	0,0092	-9,628	6,8E-22	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
precipitacion	-0,018	0,0329	-0,563	0,573	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
temp_aire_2_m	0,0649	0,0078	8,2925	1,2E-16	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
velocidad_viento_10_min	-0,854	0,0566	-15,07	4,8E-51	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
I(humedad_del_aire_2_m^2)	-0,065	0,0070	-9,250	2,5E-20	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
I(precipitacion^2)	0,0056	0,0013	4,4561	8,4E-06	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
I(temp_aire_2_m^2)	-0,029	0,0080	-3,630	0,000	No lineal + Interacción	0,030	0,0301	48121,5	48207,2

term	estimat e	std.err or	statisti c	p.value	modelo	r2	r2_ajustad o	AIC	BIC
I(velocidad_viento_10_min^2)	0,1938	0,0544	3,5638	0,000	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
precipitacion:velocidad_viento_10 _min	0,3325	0,1228	2,7076	0,007	No lineal + Interacción	0,030	0,0301	48121,5	48207,2
(Intercept)	0,0975	0,0223	4,3763	1,2E-05	No lineal + Interacción	0,004	0,0033	49590,4	49675,2
humedad_del_aire_2_m	0,0151	0,0124	1,2241	0,221	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
precipitacion	0,0059	0,0388	0,1515	0,880	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
temp_aire_2_m	0,0188	0,0099	1,8962	0,058	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
velocidad_viento_10_min	0,1791	0,0813	2,2034	0,028	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
I(humedad_del_aire_2_m^2)	-0,046	0,0110	-4,237	2,3E-05	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
I(precipitacion^2)	-0,004	0,0043	-1,054	0,292	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
I(temp_aire_2_m^2)	-0,045	0,0115	-3,941	8,1E-05	No lineal + Interacción	0,003	0,0033	49590,4	49675,2

term	estimat e	std.err or	statisti c	p.value	modelo	r2	r2_ajustad o	AIC	BIC
I(velocidad_viento_10_min^2)	0,1558	0,0696	2,2395	0,025	No lineal + Interacción	0,003	0,0033	49590,4	49675,2
precipitacion:velocidad_viento_10 _min	-0,041	0,1530	-0,268	0,788	No lineal + Interacción	0,003 9	0,0033	49590,4 2	49675,2 0

Nota. En esta tabla se presentan los resultados de los modelos no lineales que incluyen términos cuadráticos y de interacción entre las variables meteorológicas del IDEAM. Estos modelos se usaron para evaluar si la relación con los contaminantes muestra comportamientos más complejos que los modelos lineales simples. Aquí se incluyen los coeficientes, valores p y métricas como el R², AIC y BIC, que permiten ver qué tan bien se ajusta cada modelo.

Finalmente analicé los modelos GAM, que fueron los que mejor desempeño tuvieron. Estos modelos permiten ajustar relaciones no lineales más complejas usando splines, y los resultados están en el archivo resumen_gam_avanzado.csv. En estos modelos sí se ve un aumento del R^2 ajustado y también mejora la desviación explicada, pero tampoco es que logren valores muy altos. Los GAM sí muestran que hay cierto efecto de la meteorología sobre los contaminantes, pero ese efecto sigue siendo moderado y no alcanza para explicar la mayor parte del comportamiento.

Igualmente generé los gráficos GAM para cada contaminante. Estos gráficos muestran cómo cambia cada contaminante en función de cada variable meteorológica, permitiendo ver patrones no lineales. En el caso de $PM_{2.5}$, por ejemplo, casi todas las variables meteorológicas muestran curvas prácticamente planas, indicando una relación muy débil. Algo parecido ocurre con O_3 . En contraste, para PM_{10} y sobre todo para NO_2 sí se aprecian ciertas curvaturas: por ejemplo, una ligera respuesta de NO_2 frente a la humedad y la temperatura, aunque sigue siendo una relación modesta. La variable que mostró algo más de estructura en varios contaminantes fue la velocidad del viento, especialmente en $PM_{2.5}$ y PM_{10} , donde se identifica un leve comportamiento curvilíneo.

Figura 9

Relación No Lineal entre NO2 y Variables Meteorológicas

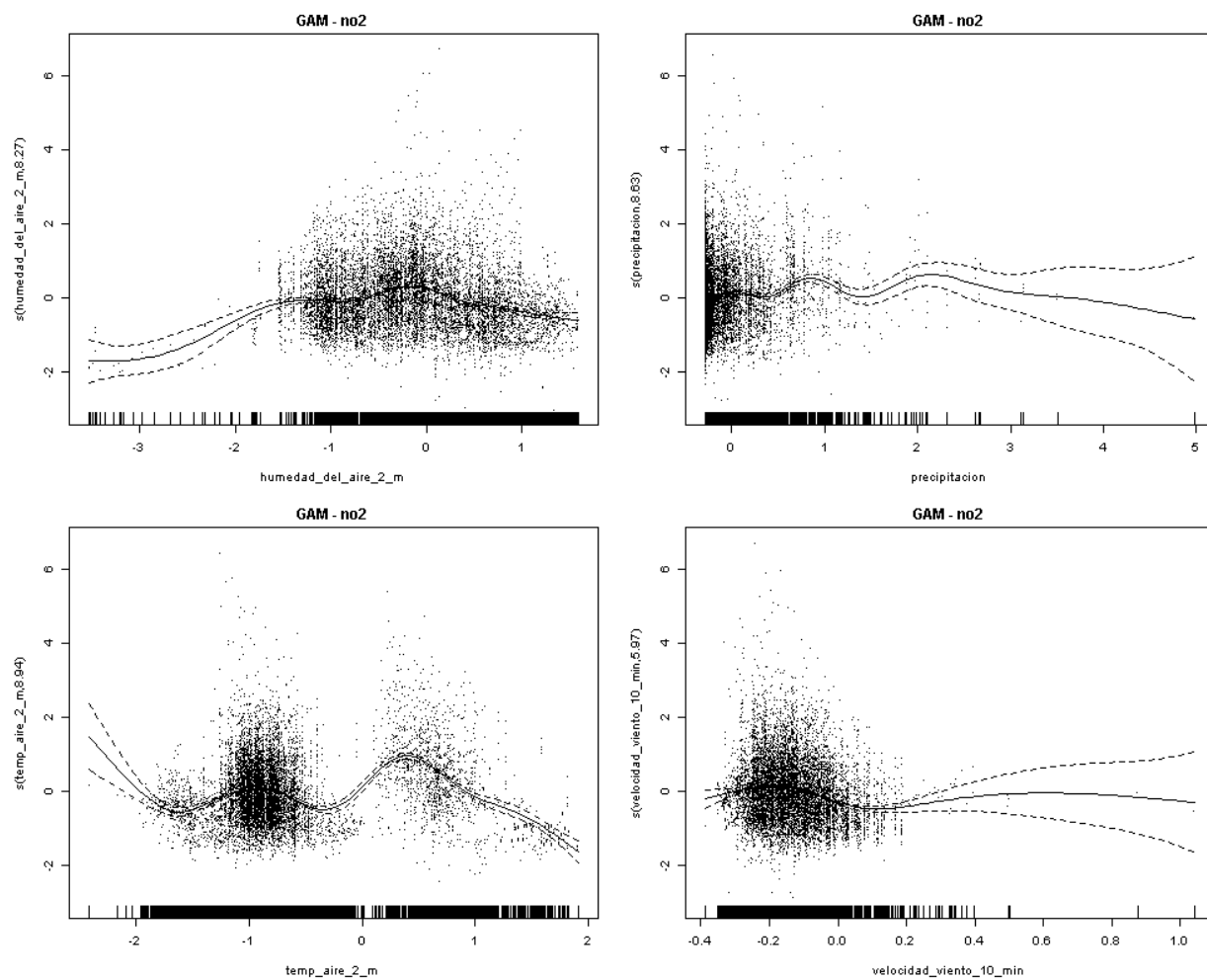


Figura 10

Relación No lineal entre O3 y Variables Meteorológicas

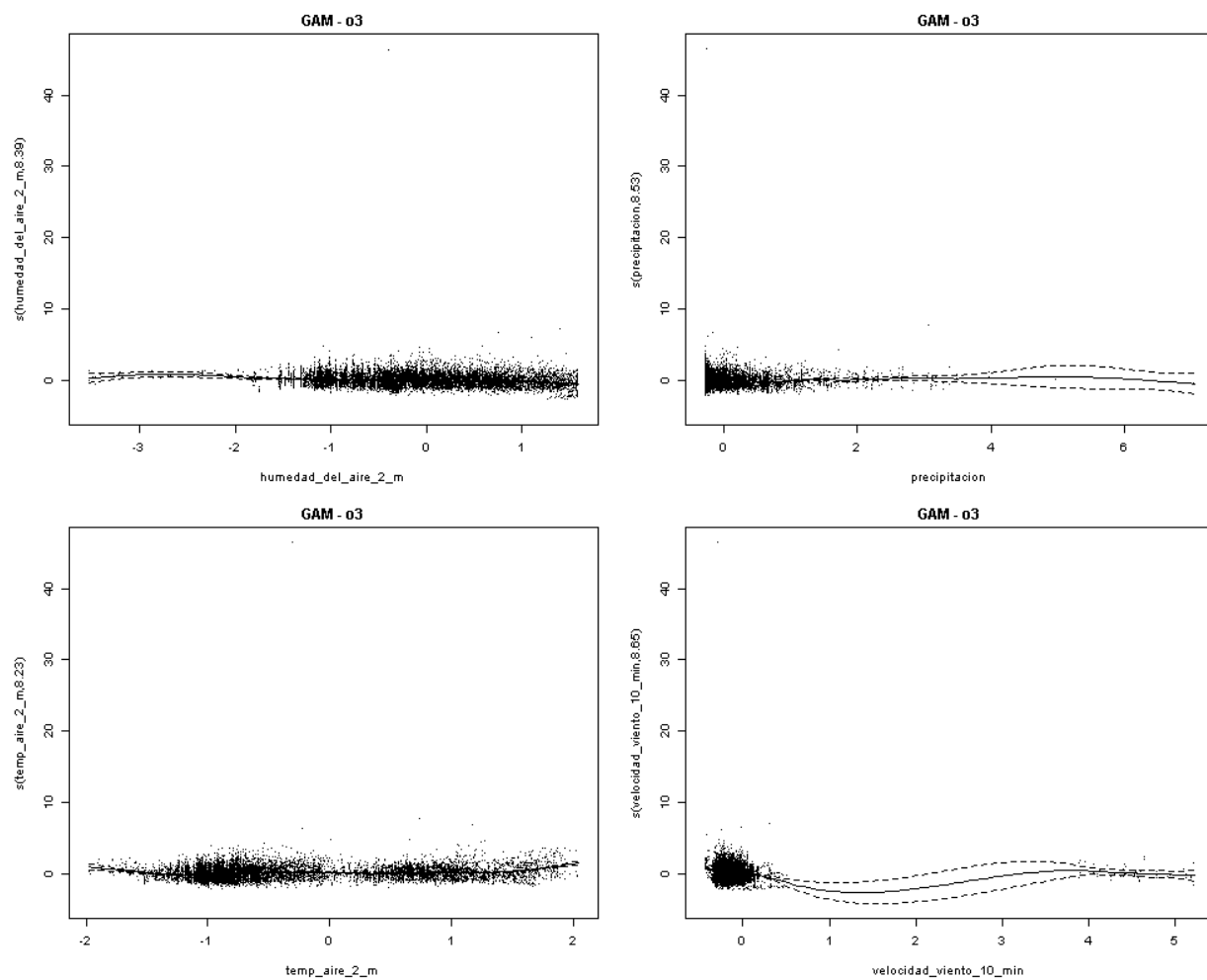


Figura 11

Relación No Lineal entre $PM_{2.5}$ y Variables Meteorológicas

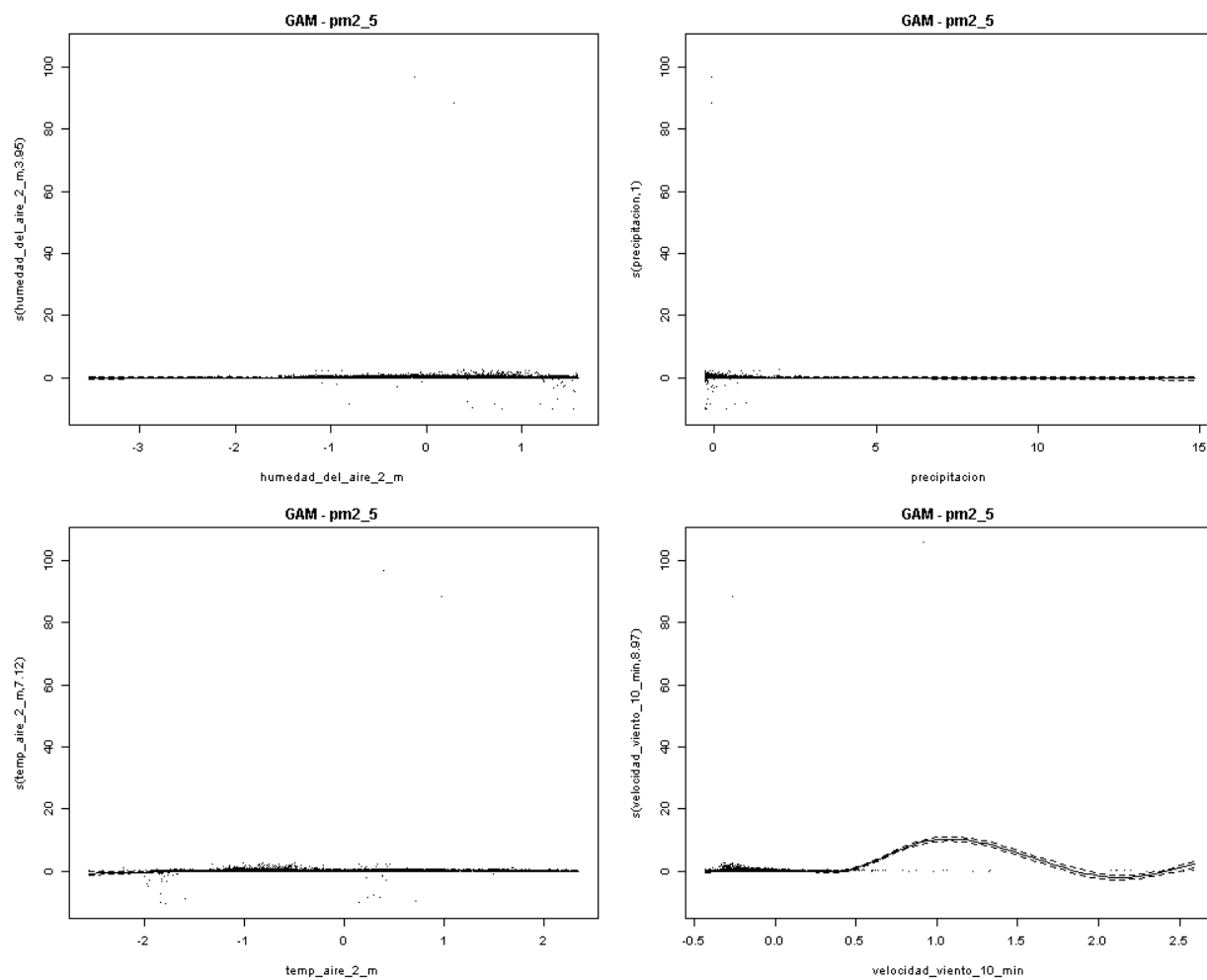
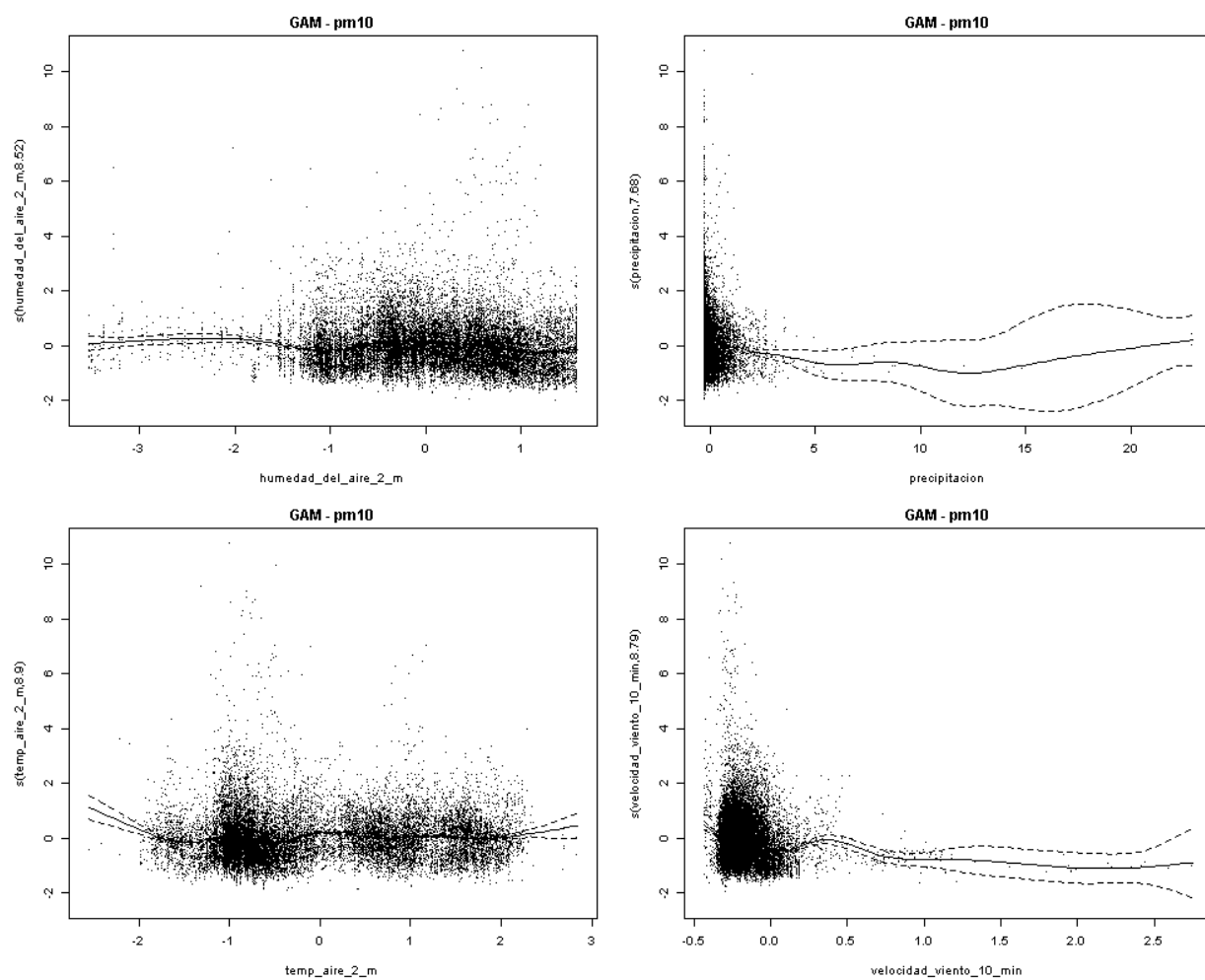


Figura 12

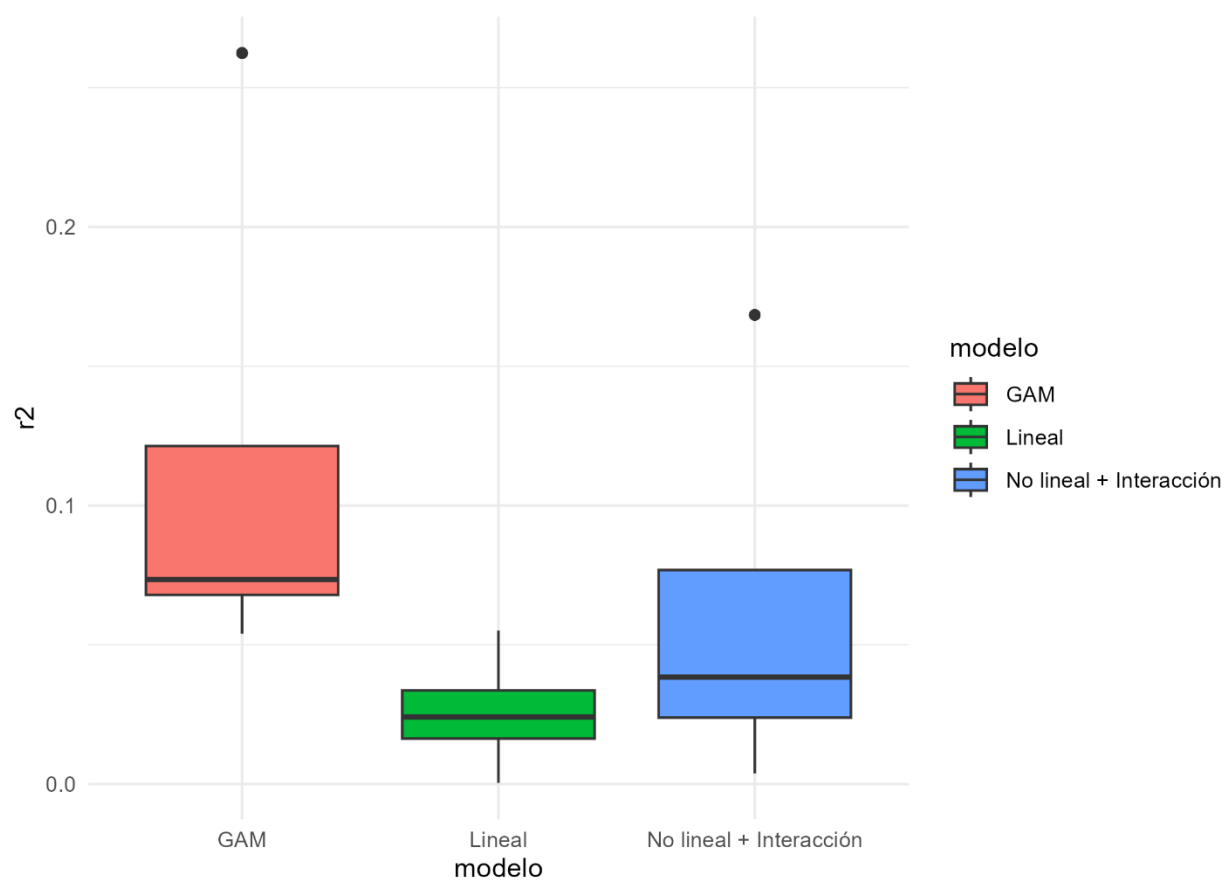
Relación No Lineal entre PM10 y Variables Meteorológicas



Ahora, se procede a realizar una comparación, el objetivo de la comparación del coeficiente de determinación (R^2) entre los modelos lineales, no lineales con interacción y GAM es evaluar cuál de estos enfoques describe mejor la variabilidad de los contaminantes a partir de las variables meteorológicas (Véase Figura 10).

Figura 13

Comparación del R^2 entre modelos



El primer aspecto notable es que los modelos lineales presentan los valores de R^2 más bajos, con una mediana cercana a cero. Esto evidencia que las relaciones entre los contaminantes y las variables meteorológicas no responden adecuadamente a asociaciones estrictamente lineales, lo cual era coherente con lo sugerido por las correlaciones previamente analizadas.

Los modelos no lineales con interacciones muestran una mejora ligera respecto a los lineales, con un aumento pequeño en la mediana y una mayor dispersión en los valores. Esto indica que la introducción de interacciones y relaciones no lineales permite capturar un poco más de variabilidad, aunque todavía de forma limitada.

Los modelos GAM son claramente los de mejor desempeño entre los tres tipos evaluados, ya que su mediana de R^2 es superior a la de los demás modelos, y los valores individuales muestran que, aunque el ajuste sigue siendo moderado, los GAM capturan una mayor proporción de la variabilidad de los contaminantes, así que esto explica la capacidad para modelar relaciones suavizadas y no lineales sin imponer una forma funcional rígida.

Así que este gráfico respalda la conclusión de que las relaciones entre meteorología y contaminación atmosférica son principalmente no lineales y que los modelos GAM representan la aproximación más adecuada para capturar estos patrones en los datos analizados.

En general, después de revisar todas las salidas del script y analizar cada archivo, lo que puedo concluir es que para el año 2021 la meteorología sí tiene un efecto sobre las concentraciones de los contaminantes, pero ese efecto no es tan fuerte ni tan directo como para que los modelos estadísticos los expliquen bien. Tanto los modelos lineales como los no lineales, e incluso los que incluyen interacciones y términos cuadráticos, muestran que sí hay relaciones, pero estas no alcanzan para obtener un buen nivel de predicción. La mayoría de los R^2 son bajos,

lo que indica que las variables del clima solo explican una parte pequeña de lo que pasa con los contaminantes.

Esto también tiene sentido porque la calidad del aire depende de muchos otros factores que no estaban presentes en la base de datos usada. Por ejemplo, la intensidad del tráfico, la actividad industrial, la hora del día, los eventos de quema de combustibles, las condiciones de estabilidad atmosférica y hasta procesos químicos que ocurren en la atmósfera y que no se pueden ver solo con humedad, temperatura, precipitación o viento. Por eso, aunque los modelos sí detectan efectos meteorológicos, ninguno logra capturar completamente el comportamiento real de los contaminantes.

¿Alguna condición meteorológica impacta sobre los contaminantes en 2021?

Sí, algunas variables meteorológicas sí muestran influencia consistente. En varios modelos, la temperatura del aire, la humedad y la velocidad del viento aparecen con coeficientes significativos. En especial, la velocidad del viento suele tener un efecto importante porque ayuda a dispersar los contaminantes y reducir su concentración. La temperatura también influye en procesos químicos que pueden aumentar o disminuir contaminantes como el ozono. La humedad y la precipitación tienden a “limpiar” la atmósfera, aunque este efecto no siempre es lineal.

Sin embargo, aunque estas variables sí tienen impacto, la magnitud de ese impacto no es suficiente para que un modelo explique bien todo el comportamiento observado. Así que, en pocas palabras, las variables meteorológicas estudiadas sí afectan, pero no determinan por completo. La conclusión general es que sí existe influencia, pero se queda corta cuando se busca explicar o predecir la contaminación solo con las variables climáticas disponibles.

Análisis Focalizado: Efecto de la Velocidad del Viento sobre PM10 y PM2.5

Además del análisis general entre clima y contaminación, decidí hacer un análisis aparte únicamente para revisar si la velocidad del viento realmente tiene algún efecto sobre los niveles de material particulado (PM10 y PM2.5). La idea es mirar este tema de forma más detallada porque, en teoría, el viento debería ayudar a dispersar los contaminantes, pero quería comprobarlo con los datos del 2021.

Para eso, lo primero que hice fue calcular la correlación simple entre el viento y cada tipo de material particulado. Con esto básicamente reviso si, cuando el viento aumenta, los valores de PM también suben o bajan. Además, generé gráficos de dispersión con la línea de tendencia para ver visualmente si había alguna relación clara.

Después pasé a una correlación parcial, que sirve para revisar si la relación entre viento y PM se mantiene incluso cuando se toman en cuenta las otras variables meteorológicas (temperatura, humedad y precipitación). Esto es útil porque a veces las variables del clima se “mezclan” entre sí y pueden confundir los resultados, así que esta prueba ayuda a aislar de forma más clara el efecto del viento.

Luego ajusté un modelo lineal simple donde el viento es la única variable que predice PM. Estos modelos me permiten ver, con un valor numérico, cuánto cambiaría el PM si el viento sube una unidad. También revisé si ese cambio es significativo o no, y qué tan bien el modelo logra explicar los datos (que normalmente no es mucho, pero igual aporta información).

Como las relaciones entre contaminación y clima no siempre son lineales, también apliqué un modelo GAM. Este tipo de modelo permite detectar comportamientos curvos, puntos de inflexión o patrones que un modelo lineal no alcanza a mostrar. Esto ayuda a ver si, por

ejemplo, el viento tiene más efecto a velocidades bajas que a velocidades altas, o algo por el estilo.

Y para asegurarme de que los resultados no estuvieran afectados por valores extremos o cosas raras en los datos, también corrí un modelo de regresión robusta, que no se deja afectar tan fácil por outliers.

Correlación Simple

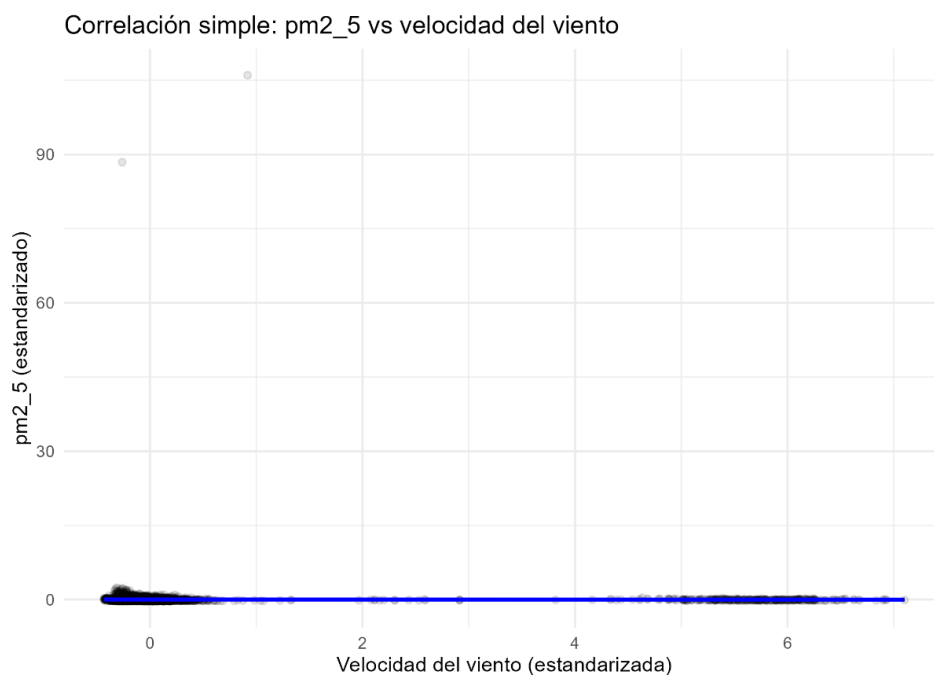
No se encontró evidencia de correlación significativa entre la velocidad del viento y las concentraciones de PM2.5 ($r = -0.004$, $p = 0.542$, IC 95%: -0.019 a 0.010), sugiriendo que otros factores meteorológicos o antropogénicos son más determinantes para este contaminante.

Tabla 9

Correlación Simple PM2.5

variable	cor	p_value	conf_low	conf_high
pm2_5	-0.00447202709111	0.54223363731	-0.0188521926682378	0.00990998829437931

Nota. No hay evidencia de relación lineal entre viento y PM2.5

Figura 14*Correlación Simple PM2.5 vs Velocidad del Viento*

Se encontró una correlación negativa estadísticamente significativa entre la velocidad del viento y las concentraciones de PM10 ($r = -0.017$, $p = 0.012$, IC 95%: -0.031 a -0.004). Sin embargo, la magnitud de esta correlación es muy débil, explicando menos del 0.03% de la variabilidad en PM10 ($R^2 = 0.0003$), lo que sugiere que el viento por sí solo no es un predictor útil de las concentraciones de este contaminante.

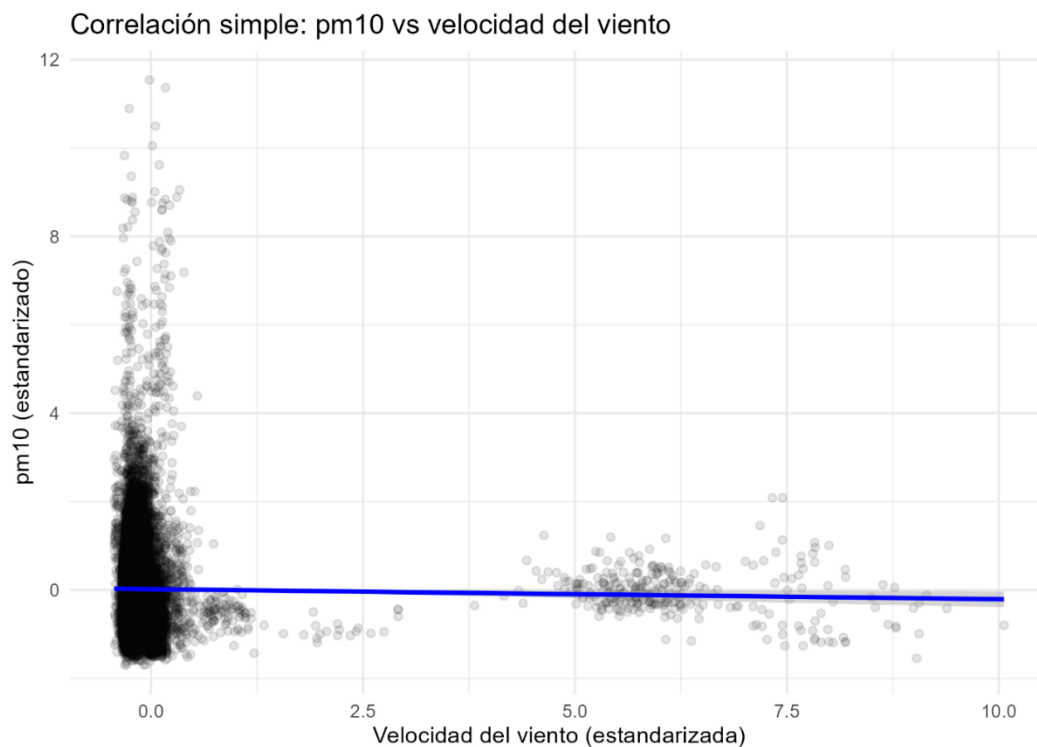
Tabla 10*Correlación Simple PM10*

variable	cor	p_value	conf_low	conf_high
pm10	-0.017484869912167	0.0121184402730	-0.03113826855690	-0.003824947263632

Nota. Aunque es estadísticamente significativa, la correlación es prácticamente nula. El viento explica menos del 0.03% de la variabilidad en PM10 ($R^2 \approx 0.0003$)

Figura 15

Correlación simple PM10 vs Velocidad del Viento



Correlación Parcial

El análisis de correlación parcial reveló que el efecto del viento sobre PM10 es significativamente más fuerte ($r = -0.133$, $p < 0.001$) que lo sugerido por la correlación simple ($r = -0.018$, $p = 0.012$). Este hallazgo indica un efecto de supresión, donde otras variables meteorológicas ocultaban el verdadero impacto del viento. Al controlar temperatura, precipitación, humedad y presión atmosférica, el viento explica 1.77% de la variabilidad en PM10, confirmando su rol en la dispersión de partículas gruesas.

A diferencia de PM10, el viento no mostró un efecto significativo sobre PM2.5 en el análisis de correlación parcial ($r = 0.015$, $p = 0.056$). El cambio de signo respecto a la correlación simple sugiere que el efecto del viento sobre partículas finas está confundido por otras variables

meteorológicas. Esto es consistente con la literatura, que indica que PM2.5, debido a su menor tamaño y mayor tiempo de residencia atmosférica, está más influenciado por procesos químicos y fuentes antropogénicas que por dispersión mecánica.

Tabla 11

Correlación Parcial Material Particulado

contaminante	estimate	p_value	statistic	n
pm10	-0.132904067804166	9.72996307630e-72	-17.9910473714038	18006
pm2_5	0.0149354376441025	0.0555122681012389	1.91498097259771	16441

Nota. El viento tiene un efecto real sobre PM10, aunque débil (1.77% de varianza)

Modelo Lineal Simple

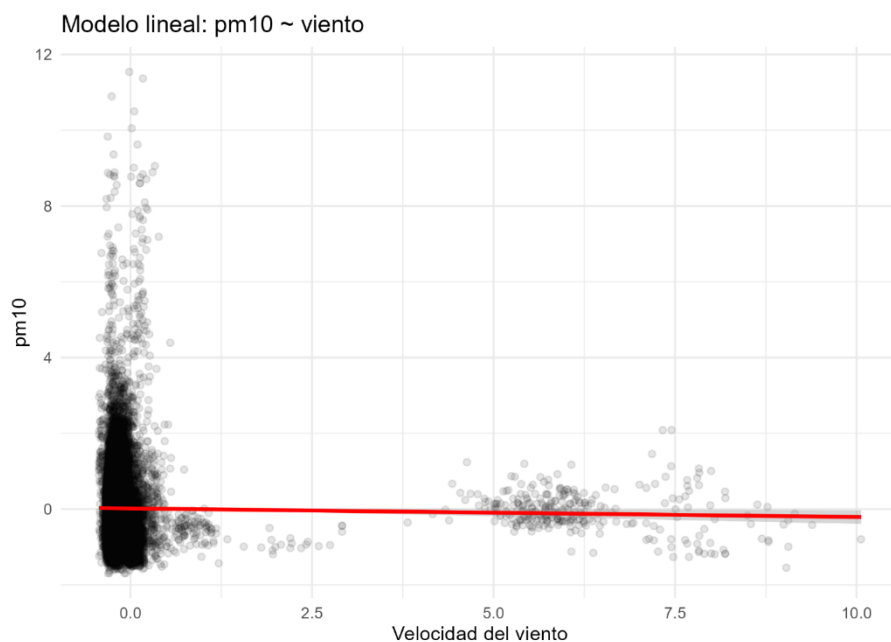
El modelo de regresión lineal simple indica que, por cada aumento de una desviación estándar en la velocidad del viento, las concentraciones de PM10 disminuyen en 0.023 desviaciones estándar ($\beta = -0.023$, $p = 0.012$). Sin embargo, el modelo explica solo 0.03% de la variabilidad en PM10, lo que sugiere que el viento por sí solo no es un predictor útil. El análisis de correlación parcial reveló que, al controlar otras variables meteorológicas, el efecto del viento es 7.6 veces más fuerte ($r = -0.133$, $p < 0.001$), indicando un efecto de supresión donde otras variables ocultaban el verdadero impacto del viento.

Tabla 12

Modelo de Regresión Lineal Simple para PM10

term	estimate	std.error	statistic	p.value	r2
(Intercept)	0.01863138	0.007131238	2.61264387	0.008991015	0.00030572
velocidad_viento_10_min	-0.0226839	0.00904140	-2.5088987	0.01211844	0.00030572

Nota. Correlación simple y regresión lineal dan resultados idénticos. Esto valida la coherencia de los análisis

Figura 16*Recta Regresión PM10 vs Velocidad del Viento*

El modelo de regresión lineal simple no encontró evidencia de un efecto significativo del viento sobre PM2.5 ($\beta = -0.006$, $p = 0.54$, $R^2 = 0.002\%$).

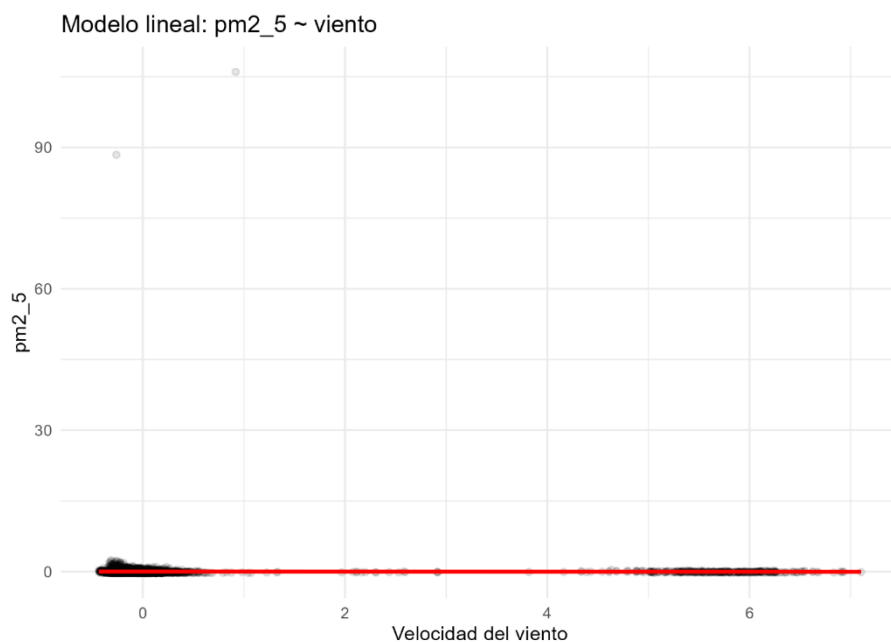
Tabla 13*Modelo de Regresión Lineal Simple para PM2.5*

term	estimate	std.error	statistic	p.value	r2
(Intercept)	0.00397729	0.00758687	0.52423380	0.60012218	1.999907e-05
velocidad_viento_10_min	-0.0060822	0.0099799	-0.60945003	0.5422336	1.999907e-05

Nota. Sin efecto significativo ($p = 0.54$)

Figura 17

Recta Regresión PM2.5 vs Velocidad del Viento



GAM Reducido

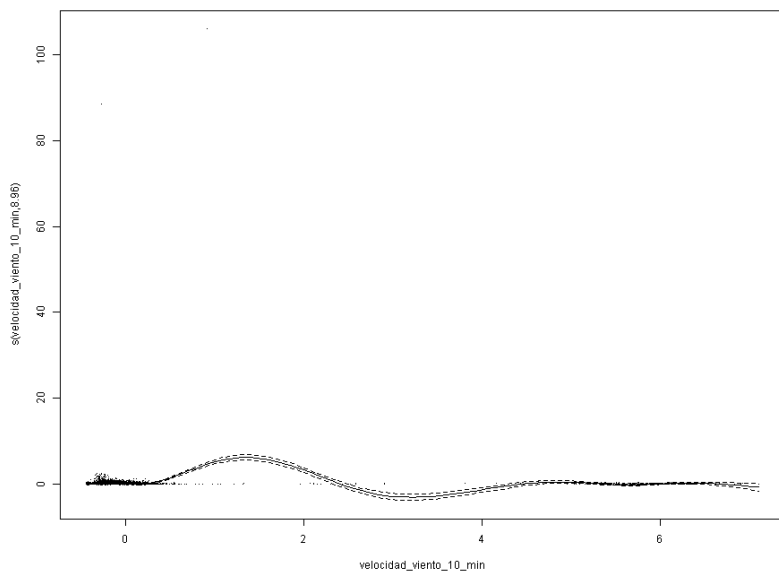
Los resultados presentados en la Tabla 16 son muy acertados, ya que los grados de libertad efectivos (EDF) de 8.965, es un valor considerablemente mayor que 1, y confirma que la relación entre la velocidad del viento y la concentración de PM2.5 es fuertemente no lineal. Esta no linealidad es estadísticamente robusta, como lo demuestra el elevado valor del estadístico F (54.389) y un valor-p prácticamente nulo ($p \approx 0$), lo que indica que el efecto del viento es un predictor significativo de las concentraciones de PM2.5. No obstante, es crucial notar que el coeficiente de determinación ajustado ($R^2 = 2.55\%$) y la varianza explicada (2.60%) son bajos. Esto implica que, si bien la velocidad del viento tiene un efecto discernible y estadísticamente válido, solo explica una pequeña fracción de la variabilidad total de PM2.5, sugiriendo la existencia de otros factores meteorológicos

Tabla 14*Resumen Estadístico del Modelo GAM para PM2.5*

Parámetro	Valor	Interpretación
Grados de Libertad Efectivos (EDF)	8.965	Indica una relación altamente no lineal y compleja.
Estadístico F	54.389	Confirma una fuerte significancia estadística del término de suavizado.
Valor-p	< 0.001	Rechaza la hipótesis nula de no efecto del viento (altamente significativo).
R2 ajustado	0.0255 (2.55%)	Porcentaje de la varianza de PM2.5 explicado por el modelo.
Varianza Explicada	0.0260 (2.60%)	Proporción de la varianza total explicada por el modelo.

Figura 18

Efecto No Lineal del Viento Sobre PM2.5



Nota. La línea continua representa la función de suavizado estimada y el área sombreada corresponde al intervalo de confianza del 95%.

El modelo para PM10 también identifica una relación no lineal significativa (EDF = 7.675, $p < 0.001$). Sin embargo, varios indicadores sugieren que el efecto del viento es considerablemente más débil en comparación con el PM2.5. El valor de EDF, aunque indica no linealidad, es inferior al observado para PM2.5, sugiriendo una curva de respuesta ligeramente menos compleja. La capacidad explicativa del modelo es casi insignificante, con un R2 ajustado de solo 0.20%. Este hallazgo subraya que la velocidad del viento, como variable única, es un predictor muy pobre de las concentraciones de PM10.

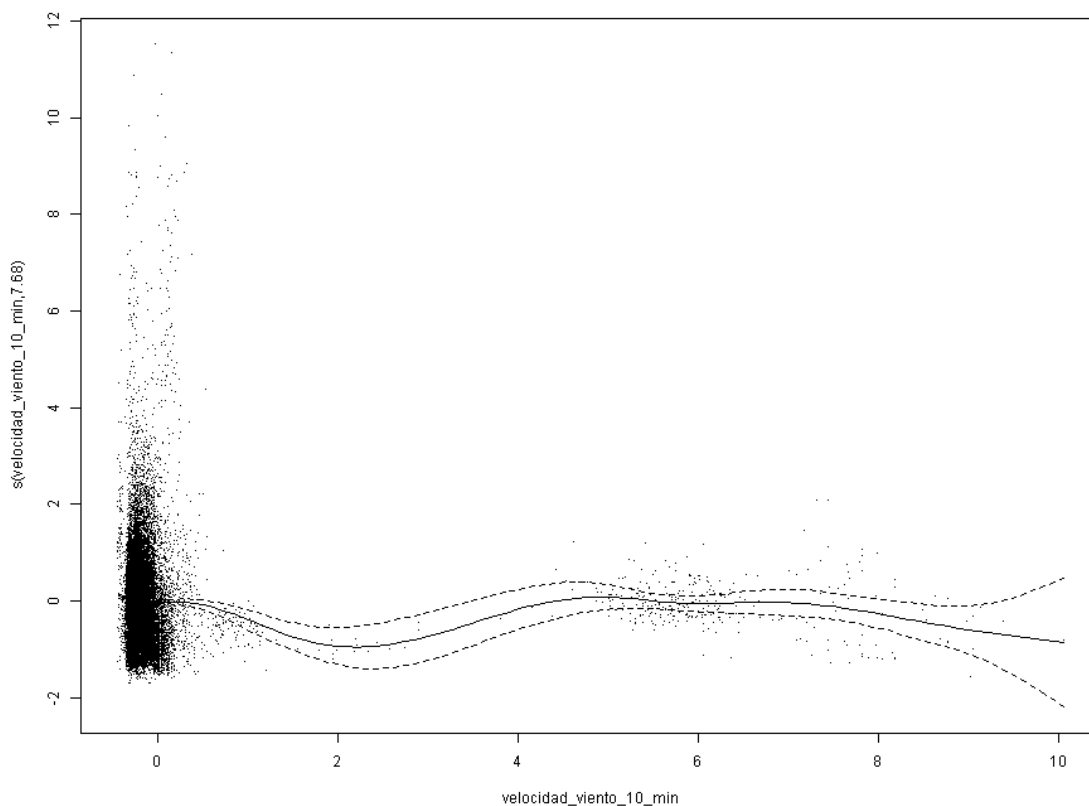
Tabla 15*Resumen Estadístico del Modelo GAM para PM10*

Parámetro	Valor	Interpretación
Grados de Libertad Efectivos (EDF)	7.675	Indica una relación no lineal, pero menos compleja que para PM2.5.
Estadístico F	5.186	Confirma la significancia estadística del término, aunque con menor fuerza.
Valor-p	< 0.001	Rechaza la hipótesis nula de no efecto del viento (significativo).
R2 ajustado	0.0020 (0.20%)	Capacidad explicativa del modelo extremadamente baja.
Varianza Explicada	0.0024 (0.24%)	Proporción de la varianza total explicada por el modelo.

Nota. Resumen estadístico del modelo GAM para el material particulado pm10 en función de la velocidad del viento.

Figura 19

Efecto No Lineal del Viento Sobre PM10



Regresión Robusta

Ahora, siguiendo con el objetivo de evaluar el efecto específico de la velocidad del viento sobre las concentraciones de material particulado, se implementó un modelo de regresión lineal robusta (RLM) utilizando el método de mínimos cuadrados ponderados iterativamente reponderados (IRLS). Este enfoque estadístico fue seleccionado por su capacidad para minimizar la influencia de valores atípicos y observaciones extremas

Tabla 16*Coefficientes de Regresión Robusta*

Contaminante	Intercepto	Coef. Viento	EE (Viento)	Estadístico t	Interpretación
PM10	-0.091	-0.0157	0.0071	-2.22	Significativo ($p < 0.05$)
PM2.5	-0.026	-0.0100	0.0016	-6.41	Altamente significativo ($p < 0.001$)

Nota. Coeficientes de regresión robusta para el efecto de la velocidad del viento sobre material particulado

En esta parte del análisis lo que encontré es que sí existe una relación entre la velocidad del viento y el material particulado, sobre todo en PM10 y PM2.5. Cuando hice la regresión robusta, el coeficiente que salió para PM10 fue negativo (aprox. -0.0157), lo que básicamente significa que cuando el viento aumenta, las concentraciones de PM10 tienden a bajar un poco. El valor t fue de -2.22 y el p-value quedó por debajo de 0.05, así que este resultado no es casualidad: es estadísticamente significativo.

Como todas las variables estaban estandarizadas, eso quiere decir que, si la velocidad del viento sube una desviación estándar, el PM10 disminuye alrededor de 0.0157 unidades estandarizadas. El efecto no es gigante, pero sí es claro, y coincide con lo que normalmente se sabe en temas de calidad del aire: cuando hay más viento, el aire se mueve más y eso “dispersa” las partículas que están suspendidas.

Para PM2.5 pasó algo parecido, pero incluso un poco más fuerte desde el punto de vista estadístico. El coeficiente también fue negativo (-0.0100), pero lo que llamó la atención fue que el valor t fue mucho más alto en magnitud (-6.41) y el p-value quedó muchísimo más pequeño

(menor a 0.001). Esto quiere decir que la relación es más estable y confiable en PM2.5 que en PM10, aunque el tamaño del efecto no sea enorme.

Hay varias razones por las que esto puede pasar: el material particulado PM2.5 se comporta de manera más uniforme cuando cambia el clima, porque es más fina y permanece más tiempo suspendida. Las partículas finas no dependen tanto de fuentes súper locales como sí pasa con PM10 y PM2.5 tiende a moverse más lejos con el viento, por lo que el efecto de dispersión es más claro.

Al final, lo que muestran estos datos es que el viento sí ayuda a bajar la concentración de material particulado, pero no es el único factor que influye. El efecto existe, pero tampoco es tan grande como para explicar por completo lo que pasa con la calidad del aire. Aun así, en PM2.5 la evidencia es más sólida y consistente.

Conclusiones

Al finalizar este proyecto, siento que logré cumplir el objetivo principal que me propuse que era construir un tablero interactivo en Power BI que realmente permita explorar, comparar y entender cómo se relacionan las condiciones meteorológicas con los niveles de calidad del aire en diferentes ciudades del país. Más allá de ser un ejercicio técnico, este trabajo me permitió ver de primera mano lo compleja que es la gestión de datos ambientales en Colombia y lo útil que puede ser una herramienta visual bien construida para apoyar procesos de análisis y toma de decisiones.

El primer gran paso fue recopilar y preparar los datos del IDEAM y de SISAIRE, y aunque al principio pareció sencillo, muy pronto descubrí que cada fuente tenía sus particularidades. Tuve que diseñar varios procesos de limpieza, normalización y consolidación para dejar los datos en un formato manejable, lo cual terminó siendo un aporte importante del proyecto. Tanto el uso de DuckDB para optimizar los archivos del IDEAM como la construcción de los scripts para automatizar el scraping de SISAIRE fueron esenciales para poder trabajar con volúmenes grandes de información sin perder rendimiento.

Una vez organizada toda la información, la construcción del dashboard se volvió una parte más creativa. La primera página, dedicada al IDEAM, me permitió mostrar el comportamiento de las variables meteorológicas de forma intuitiva y flexible. La segunda página, enfocada en los contaminantes del aire, me ayudó a entender mejor las diferencias entre regiones, la distribución temporal y las variaciones típicas de cada contaminante. Finalmente, la tercera página —la comparativa— fue la que más valor añadió al permitir relacionar directamente ambos sistemas y visualizar cómo ciertos patrones meteorológicos parecen acompañar fases específicas de contaminación.

A su vez, el desarrollo de este trabajo permitió cumplir el objetivo de evaluar la relación entre las variables meteorológicas y los principales contaminantes atmosféricos mediante diferentes enfoques de modelamiento estadístico. Por lo que a lo largo del proceso se integraron análisis exploratorios, modelación lineal y no lineal, y finalmente modelos aditivos generalizados (GAM), lo que permitió comprender de manera más profunda la dinámica entre los contaminantes y las condiciones atmosféricas presentes en el área de estudio.

Los resultados obtenidos mostraron que las asociaciones entre meteorología y contaminación no siempre siguen patrones lineales, lo que justifica la incorporación de modelos más flexibles como los GAM. Esta metodología permitió capturar variaciones que los modelos lineales y no lineales tradicionales no lograron describir adecuadamente, destacándose por su mayor capacidad explicativa y mejor desempeño en términos del coeficiente de determinación (R^2).

Usar regresión robusta en vez de una regresión normal fue útil porque este tipo de datos suele tener valores raros o muy altos (por ejemplo, días con incendios o eventos meteorológicos fuertes) y una regresión tradicional puede verse afectada por esos “extremos”. La regresión robusta ayuda a que los resultados sean más estables, sobre todo cuando las distribuciones no son perfectas ni simétricas, como suele pasar con los contaminantes.

Recomendaciones

A partir de los resultados obtenidos en este trabajo, considero que es importante que en futuros estudios sobre la relación entre la contaminación atmosférica y las variables meteorológicas se incluyan más fuentes de información, ya que la meteorología por sí sola no logra explicar suficientemente el comportamiento de los contaminantes. Sería recomendable incorporar datos adicionales como el tráfico vehicular, la actividad industrial, la radiación solar, la estabilidad atmosférica y, en general, factores que influyen directamente en la emisión o transformación de los contaminantes, ya que estos elementos no estuvieron presentes en la base utilizada y podrían mejorar notablemente el desempeño de los modelos.

En cuanto a las líneas de investigación futuras, sería valioso replicar este análisis con más años de datos para identificar si los patrones encontrados en 2021 se mantienen o si corresponden a condiciones particulares de ese periodo.

Referencias Bibliográficas

- Acosta, L., & Meléndez, A. (2024). Implementación de un Sistema de Medición Meteorológica para el Análisis de Datos . RICT Revista de Investigación Científica, Tecnológica e Innovación, 2, 47–55. <https://revista.ccaitec.com/index.php/ridt/article/view/62>
- Agudelo-Londoño, S. M., Blanco-Becerra, L. C., Hernández, M. R., Suárez-Morales, Z. B., Mantilla-León, L. C., & Solís, N. (2024). Injusticia ambiental en la calidad del aire para repartidores de plataformas digitales de Bogotá, Colombia, 2021. *Biomédica*, 44(3), 391-401. <https://doi.org/10.7705/biomedica.7162>
- Aldana, S., & Cárdenas, K. (2020). Análisis Del Estado De Las Capacidades En Gestión De La Calidad Del Aire Urbano De Bogotá Y Barranquilla Como Insumo Para El Fortalecimiento De La Política Pública Ambiental Local Vigente, A Través Del Índice CECA [Universidad Distrital Francisco José De Caldas]. <https://repository.udistrital.edu.co/server/api/core/bitstreams/c2f60ba8-5086-4bc2-9c6f-5d4cc0dde929/content>
- Armesto Arenas, A., & Cañon Barriga, J. E. (2023). Relación entre material particulado, variables meteorológicas e infecciones respiratorias agudas en la zona norte de Ocaña. *Revista de Investigación Agraria y Ambiental*, 15(1), 251–279. <https://doi.org/10.22490/21456453.6537>
- Bonifaz, D., & Guamán, L. (2023). Visualización del comportamiento de la variable temperatura mediante la herramienta “power bi” para un aplicativo web en la provincia de chimborazo [Escuela Superior Politécnica de Chimborazo]. <http://dspace.esPOCH.edu.ec/bitstream/123456789/21721/1/226T0157.pdf>

- Botello, N., & Arenas, J. (2023). Estudio De La Viabilidad De Implementación De Una Estación De Monitoreo De La Calidad Del Aire Utilizando La Tecnología Arduino [Universidad Libre]. <https://core.ac.uk/download/pdf/563909435.pdf>
- Cabrera Encinoso, S. (2021). Análisis de las partículas contaminantes Pm2,5 en Santa Cruz de Tenerife [Universidad de La Laguna]. <https://riull.ull.es/xmlui/handle/915/23035>
- Cadavid-Giraldo, N., Orrego Guarín, A. Z., & Barrera Jiménez, H. (2021). Calidad del aire en el Valle de Aburrá, ¿Qué podríamos esperar de la ciudad inteligente? Ingeniería y Ciencia, 17(33), 185–222. <https://doi.org/10.17230/ingciencia.17.33.8>
- Cárdaba Arranz, M. (2012). Carga de enfermedad por contaminación atmosférica debida a materia particulada (PM10 y PM2,5) y ozono (O3) en la ciudad de Valladolid durante el periodo 1999-2008 [Universidad de Valladolid]. <https://doi.org/10.35376/10324/2043>
- Casallas, A., Córdoba, T., Sánchez-Cárdenas, L., Guevara-Luna, M. A., & Belalcazar, L. C. (2022). Understanding high polluted events in a tropical megacity with air quality observations and ERA5-reanalysis data. Revista Facultad de Ingeniería Universidad de Antioquia. <https://doi.org/10.17533/udea.redin.20220682>
- Chaparro, L., & Quevedo, D. (2021). Evaluacion de Red neuronal recurrente para la prediccion de riesgo por SOx en la ciudad de Bogota. <https://doi.org/10.13140/RG.2.2.35276.74886>
- Coates, J., Mar, K. A., Ojha, N., & Butler, T. M. (2016). The influence of temperature on ozone production under varying NO_x conditions – a modelling study. Atmospheric Chemistry and Physics, 16(18), 11601–11615. <https://doi.org/10.5194/acp-16-11601-2016>

Congreso de la República de Colombia. (1993). *Ley 99 de 1993. Por la cual se crea el Ministerio del Medio Ambiente, se reordena el Sector Público encargado de la gestión y conservación del medio ambiente y los recursos naturales renovables*. Diario Oficial No. 41.146.

Departamento Nacional de Planeación [DNP]. (2018, febrero). Calidad del aire. Una prioridad de política pública en Colombia. [https://colaboracion.dnp.gov.co/CDT/Prensa/PresentaciónCalidad del Aire 15_02_2018.pdf](https://colaboracion.dnp.gov.co/CDT/Prensa/PresentaciónCalidaddelAire15_02_2018.pdf)

De Salud, N. I. N. (2022). Boletín epidemiológico semanal 45 de 2022. *Boletín Epidemiológico Semanal*, 1-28. <https://doi.org/10.33610/23576189.2022.45>

Galvis-Vizcaíno, S. A., León-Barraza, I. R., Padilla-Urueta, K. del S., Pombo-Navas, M. Y., Rada-Carrillo, M. B., Vera-Brand, J., Rebolledo-Cobos, R. C., Galvis-Vizcaíno, S. A., León-Barraza, I. R., Padilla-Urueta, K. del S., Pombo-Navas, M. Y., Rada-Carrillo, M. B., Vera-Brand, J., & Rebolledo-Cobos, R. C. (2021). Material particulado y trastornos de los sistemas cardiovascular y respiratorio en trabajadores de diferentes áreas: una revisión narrativa. *Asociación Española de Especialistas En Medicina Del Trabajo*, 30(3), 362–372. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S3020-11602021000300011&lng=es&nrm=iso&tlng=es

García, C. (2019). Relación entre contaminantes atmosféricos y variables meteorológicas [Universidad Científica del Sur]. <https://repositorio.cientifica.edu.pe/bitstream/handle/20.500.12805/1371/TB-Garcia%20C.pdf?sequence=1>

González, A. (2021). Lineamientos Para La Gestión Del Riesgo Ante La Ocurrencia De Episodios Excepcionales De Contaminación Atmosférica En La Ciudad De Bogotá Y

Municipios Del Área Metropolitana Del Vallé De Aburrá (Colombia) [Universidad Distrital Francisco José De Caldas].

<https://repository.udistrital.edu.co/server/api/core/bitstreams/98fc2b54-adc1-4e8e-b281-56a16d2b5eec/content>

Guerra Gómez, J. A. (2023, 22 marzo). Spike Map of Colombian Cities by population.

Observable. <https://observablehq.com/@john-guerra/spike-map-of-colombian-cities-by-population>

Ji, M., Jiang, Y., Han, X., Liu, L., Xu, X., Qiao, Z., & Sun, W. (2020). Spatiotemporal Relationships between Air Quality and Multiple Meteorological Parameters in 221 Chinese Cities. *Complexity*, 2020, 1–25. <https://doi.org/10.1155/2020/6829142>

Klompaker, J. O., Hart, J. E., James, P., Sabath, M. B., Wu, X., Zanobetti, A., Dominici, F., & Laden, F. (2021). Air pollution and cardiovascular disease hospitalization – Are associations modified by greenness, temperature and humidity? *Environment International*, 156, 106715. <https://doi.org/10.1016/j.envint.2021.106715>

Lelieveld, J., Pozzer, A., Pöschl, U., Fnais, M., Haines, A., & Münzel, T. (2020). Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovascular Research*, 116(11), 1910–1917. <https://doi.org/10.1093/cvr/cvaa025>

Li, Q., Daqing, L., Ashkenazy, Y., Havlin, S., & col. (2025). The origin of long-range links of air pollution in China. *ArXiv preprint*. <https://arxiv.org/abs/2509.05974>

Lombana, W. (2024). Herramienta de Análisis de Desarrollo de Proyectos: Visualización de Tiempos y Eficiencia a través de Power BI. [Universidad Nacional Abierta y a Distancia]. <https://repository.unad.edu.co/handle/10596/63754>

- Marchetti, J. (2023). Análisis de herramientas de ETL y reporting en entornos Big Data [Universidad Nacional de la Plata]. <http://sedici.unlp.edu.ar/handle/10915/165008>
- Marín, D., Herrera, V., Piñeros-Jiménez, J. G., Rojas-Sánchez, O. A., Mangones, S. C., Rojas, Y., Cáceres, J., Agudelo-Castañeda, D. M., Rojas, N. Y., Belalcazar-Ceron, L. C., Ochoa-Villegas, J., Montes-Mejía, M. L., Lopera-Velasquez, V. M., Castillo-Navarro, S. M., Torres-Prieto, A., Baumgartner, J., & Rodríguez-Villamizar, L. A. (2025). Long-term exposure to PM2.5 and cardiorespiratory mortality: an ecological small-area study in five cities in Colombia. *Cadernos de Saúde Pública*, 41(4), e00071024. <https://doi.org/10.1590/0102-311xen071024>
- Martínez Morales, S. (2024, 5 noviembre). Modelación de la dispersión atmosférica de partículas PM2.5, su contenido de metales traza y potencial de riesgo a la salud en el Área Metropolitana de Monterrey. <http://hdl.handle.net/11627/6647>
- Ministerio de Ambiente y Desarrollo Sostenible. (2015). *Decreto 1076 de 2015. Por medio del cual se expide el Decreto Único Reglamentario del Sector Ambiente y Desarrollo Sostenible*. Diario Oficial No. 49.523.
- Ministerio de Ambiente y Desarrollo Sostenible. (2017). *Resolución 2254 de 2017. Por la cual se establecen los niveles de calidad del aire y los niveles de prevención, alerta y emergencia por contaminación del aire*
- Ministerio de Ambiente, Vivienda y Desarrollo Territorial. (2010). *Política Nacional de Calidad del Aire (PNCA)*. Bogotá D.C.
- Ministerio de Ambiente, Vivienda y Desarrollo Territorial. (2010). *Resolución 650 de 2010. Por la cual se establecen los Protocolos para el Monitoreo y Seguimiento de la Calidad del Aire*.

- Mora, R. (2022). Análisis del impacto de las políticas ambientales de la ciudad de Bogotá en los niveles de contaminación del aire en la localidad de Kennedy 2010-2021 [Universidad Colegio Mayor de Cundinamarca].
<https://repositorio.unicolmayor.edu.co/handle/unicolmayor/6782>
- Navas Vázquez, B. (2024). Análisis y estudio para la optimización de la red de puntos de recarga de vehículos eléctricos del Ayuntamiento de Madrid: un enfoque integral para la eficiencia urbana sostenible [Universidad a distancia de Madrid].
<http://hdl.handle.net/20.500.12226/2517>
- Nguyen, G. T. H., La, L. T., Hoang-Cong, H., & Le, A. H. (2023). An exploration of meteorological effects on PM2.5 air quality in several provinces and cities in Vietnam. *Journal Of Environmental Sciences, 145*, 139-151.
<https://doi.org/10.1016/j.jes.2023.07.020>
- Organización Mundial de la Salud (OMS). (2021). *Air pollution*. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- Parra Sánchez, J. S., Oviedo Carrascal, A. I., & Amaya Fernández, F. O. (2020). Analítica de datos: incidencia de la contaminación ambiental en la salud pública en Medellín (Colombia). *Revista de Salud Pública, 22*(6), 1–9.
<https://doi.org/10.15446/rsap.v22n6.78985>
- Prada Penagos, C. R. (2023). Diagnóstico de la incidencia de la calidad del aire en la prevalencia de enfermedades respiratorias, mediante el uso de datos abiertos y el análisis bajo la herramienta Power BI, en la ciudad de Bogotá [Universidad Cooperativa de Colombia].
<https://hdl.handle.net/20.500.12494/55220>

- Ruiz, M., & Ruiz, E. (2024). Estudio de la correlación de la dinámica de material particulado (PM10 y PM2.5) con las condiciones meteorológicas en el municipio de Pasto utilizando estaciones meteorológicas portátiles [Universidad Mariana].
<https://hdl.handle.net/20.500.14112/28264>
- Sánchez, C., & Bautista, M. (2019). Evaluación de la calidad del aire (PM10 y PM2.5) en relación a los parámetros meteorológicos (temperatura, humedad relativa y velocidad de viento) en el sector Cercado- Tarapoto, 2018 [Universidad Peruana Unión].
<https://repositorio.upeu.edu.pe/server/api/core/bitstreams/8bfa4a57-aef2-4474-9d65-7c31b7b7173e/content>
- Sandoval D., B., Reyes R., T., & Oyarzún G., M. (2019). Mecanismos de los efectos nocivos para la salud de la contaminación atmosférica proveniente de incendios forestales. *Revista Chilena de Enfermedades Respiratorias*, 35(1), 49–57.
<https://doi.org/10.4067/S0717-73482019000100049>
- Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.
- Suárez, D. (2024). Efecto del Cambio Climático sobre la Salud [Universidad de Jaén].
<https://hdl.handle.net/10953.1/24341>
- Tushar117. (2025, October 9). Real-Time Weather & Air Quality Insights – Power BI Dashboard. Microsoft Fabric Community.
<https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/Real-Time-Weather-amp-Air-Quality-Insights-Power-BI-Dashboard/m-p/4822559>
- Vásquez Franco, D. F. (2019). *Procesamiento de datos meteorológicos del IDEAM con Python y validación de datos del reanálisis ERA5*. <http://hdl.handle.net/10495/13235>

- Warthon Olarte, B. S., Zamalloa Ponce de Leon, A. I., Warthon Ascarza, J. L., Miranda Hankgo, I. C., Quispe Ccolque, I. R., Ramos Salcedo, V. F., Ponce Amanca, R. E., & Tupayachi Latorre, R. A. (2024). *Evaluación de la Contaminación del Aire por Material Particulado PM2.5 en la ciudad del Cusco Respecto de los Índices de Calidad del Aire entre 2017 y 2018*. *Ambiente, Comportamiento y Sociedad*, 6(1), 1–15.
<https://doi.org/10.51343/racs.v6i1.1112>
- Wright, C. Y., & Wernecke, B. (2020). Using Microsoft© Power BI© to visualise Rustenburg Local Municipality's Air Quality Data. *Clean Air Journal*, 30(1). <https://doi.org/10.17159/caj/2020/30/1.7512>
- Yupa, A., Urgilés, C., Cárdenas, D., & Zenteno, J. (2023). Analítica descriptiva para la comprensión del estado actual del fenómeno delincriminal en el Ecuador. *Pro Sciences: Revista de Producción, Ciencias e Investigación*, 7.
<https://doi.org/https://doi.org/10.29018/issn.2588-1000vol7iss49.2023pp62-87>

Apéndices

Apéndice A

Scripts Utilizados para la Conversión y Agregación de Datos del IDEAM

Este anexo presenta los scripts desarrollados en Python para procesar los datos descargados del IDEAM antes de su análisis en Power BI. Se utilizaron las librerías **DuckDB** y **Python** debido a su eficiencia en el manejo de archivos de gran tamaño y a su compatibilidad con formatos como CSV y Parquet.

Los procesos incluidos corresponden a:

1. Conversión de archivos CSV a formato Parquet.
2. Agregación diaria de las observaciones de presión atmosférica.

Cada script se encuentra comentado y se describe brevemente su propósito para facilitar la replicación del proceso.

A.1 Conversión de archivo CSV a Parquet

Este script convierte el archivo original en formato CSV, descargado del IDEAM, a un archivo Parquet. Este formato permite una lectura optimizada, reduce el tamaño final y facilita su integración con herramientas de análisis como Power BI.

```
import duckdb
```

```
input_csv = r"ruta/del/archivo/IDEAM_Humedad_del_Aire.csv"
```

```
output_parquet = r"ruta/del/parquet/ IDEAM_Humedad_del_Aire.parquet"
```

```
duckdb.sql(f"""
```

```
    COPY (SELECT * FROM read_csv_auto('{input_csv}'))
```

```
    TO '{output_parquet}' (FORMAT PARQUET);
```

```
""")
```

```
print("Conversión completa:", output_parquet)
```

A.2 Agregación diaria de datos

Este script procesa el archivo Parquet generado previamente. Crea un nuevo archivo con los valores agregados por día (promedio, suma, número de registros), manteniendo además los metadatos principales de cada estación.

La función STRPTIME se emplea para transformar el formato original de fecha provisto por el IDEAM antes de agrupar por día.

```
import duckdb
```

```
input_file = r"ruta/del/parquet/IDEAM_Presion_Atmosferica.parquet"
```

```
output_file = r"ruta/del/parquet/IDEAM_Presion_Atmosferica_diaria.parquet"
```

```
con = duckdb.connect()
```

```
print("🕒 Leyendo y agregando datos... esto puede tardar un poco según el tamaño del  
archivo.")
```

```
query = f"""
```

```
COPY (
```

```
SELECT
```

```
    CodigoEstacion,
```

```
    DATE(STRPTIME(FechaObservacion, '%Y %b %d %I:%M:%S %p')) AS Fecha,
```

```
    AVG(ValorObservado) AS ValorPromedio,
```

```
    SUM(ValorObservado) AS ValorTotal,
```

```
    COUNT(*) AS NumRegistros,
```

```
    ANY_VALUE(NombreEstacion) AS NombreEstacion,
```

```
    ANY_VALUE(Departamento) AS Departamento,
```

```
    ANY_VALUE(Municipio) AS Municipio,
```

```
    ANY_VALUE(ZonaHidrografica) AS ZonaHidrografica,
```

```
    ANY_VALUE(Latitud) AS Latitud,
```

```
    ANY_VALUE(Longitud) AS Longitud,
```

```

        ANY_VALUE(DescripcionSensor) AS DescripcionSensor,
        ANY_VALUE(UnidadMedida) AS UnidadMedida
    FROM read_parquet('{input_file}')
    GROUP BY CodigoEstacion, DATE(STRPTIME(FechaObservacion, '%Y %b %d
    %I:%M:%S %p'))
)
TO '{output_file}' (FORMAT PARQUET);
''''

con.execute(query)
con.close()

print(f'Archivo agregado guardado en:\n{output_file}')

```

A.3 Script en Python para la unificación de los archivos Parquet del IDEAM

Este anexo presenta el script desarrollado en Python para realizar la unificación de los cinco archivos en formato Parquet generados tras el proceso de agregación diaria de las variables meteorológicas provenientes del IDEAM. El objetivo de este procedimiento fue consolidar todos los conjuntos de datos optimizados en un único archivo maestro (IDEAM_MERGE.parquet), facilitando así su incorporación al entorno de Power BI y mejorando la eficiencia en el manejo de grandes volúmenes de información.

```

import pyarrow.parquet as pq
import pyarrow as pa
import os

# Carpeta donde están los archivos .parquet previamente agregados
carpeta = r"C:\ruta_archivos_parquet_ideam"

# Listar archivos parquet en la carpeta

```

```
archivos = [f for f in os.listdir(carpeta) if f.endswith(".parquet")]

tablas = []

print(f"Leyendo {len(archivos)} archivos parquet...")

# Lectura de cada archivo parquet y almacenamiento en una lista
for archivo in archivos:
    ruta = os.path.join(carpeta, archivo)
    tabla = pq.read_table(ruta)
    tablas.append(tabla)
    print("Cargado:", archivo)

print("Uniendo tablas...")

# Unificación de todas las tablas en un solo objeto
tabla_unida = pa.concat_tables(tablas, promote=True)

# Ruta para guardar el archivo consolidado
salida = os.path.join(carpeta, "IDEAM_MERGE.parquet")
pq.write_table(tabla_unida, salida)

print("\nArchivo final creado:")
print(salida)
```

Apéndice B

Scripts Utilizados para Procesamiento y Consolidación de Datos de SISAIRES

El presente anexo reúne los scripts utilizados para la obtención, limpieza, estructuración y consolidación de los datos provenientes del Sistema de Información sobre Calidad del Aire – SISAIRES. Estos scripts permitieron automatizar la descarga de información, identificar la estructura territorial de monitoreo (departamentos, municipios y estaciones) y unificar los valores de contaminación atmosférica de los contaminantes O₃, NO₂, PM₁₀ y PM_{2.5}.

Los anexos están organizados según el flujo del proceso realizado.

B.1 Script para obtener los departamentos disponibles en SISAIRES

Este script realiza una consulta automatizada a la plataforma SISAIRES, identifica las opciones disponibles en el menú desplegable de departamentos y genera un archivo CSV con la lista completa de departamentos que poseen estaciones de monitoreo activas.

```
import requests
from bs4 import BeautifulSoup
import csv
import os

# URL de SISAIRES
url_scrap = "http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml"

# Crear sesión y obtener la página
session = requests.Session()
resp = session.get(url_scrap, timeout=10)

if resp.status_code != 200:
    raise Exception(f'No se pudo conectar. Código HTTP: {resp.status_code}')
```

```

soup = BeautifulSoup(resp.text, "html.parser")

# Buscar el select de departamentos
select_departamentos = soup.find("select", {"id": "filtroForm:departamentoSel_input"})

if not select_departamentos:
    raise Exception("No se encontró el select de departamentos")

# Listado para guardar los datos
departamentos = []

# Recorrer opciones
for opt in select_departamentos.find_all("option"):
    codigo = opt.get("value")
    nombre = opt.text.strip()

    if codigo and codigo != "":
        departamentos.append({
            "nombre_departamento": nombre,
            "codigo_departamento": codigo
        })

# Crear CSV
with open("departamentos_sisaire.csv", "w", newline="", encoding="utf-8") as f:
    writer = csv.writer(f)
    writer.writerow(["id", "nombre_departamento", "codigo_departamento"])

    for i, d in enumerate(departamentos, start=1):
        writer.writerow([i, d["nombre_departamento"], d["codigo_departamento"]])

print("CSV generado: departamentos_sisaire.csv")

```

```
print("Archivo guardado en:", os.path.abspath("departamentos_sisaire.csv"))
```

B.2 Script para obtener los municipios por departamento

Para cada departamento identificado en el Anexo B.1, este script consulta la plataforma y extrae los municipios asociados que cuentan con estaciones activas. Los resultados se guardan en un archivo CSV consolidado.

```
import requests
from bs4 import BeautifulSoup
import csv

URL = "http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml"

headers = {
    "Content-Type": "application/x-www-form-urlencoded",
    "Faces-Request": "partial/ajax",
    "User-Agent": "Mozilla/5.0"
}

# Cargamos los departamentos desde un CSV local previamente generado
# formato: id,nombre_departamento,codigo_departamento
DEPARTAMENTOS_CSV = "departamentos_sisaire.csv"

departamentos = []
with open(DEPARTAMENTOS_CSV, newline="", encoding="utf-8") as f:
    reader = csv.DictReader(f)
    for row in reader:
        departamentos.append({
            "id": int(row["id"]),
            "codigo_departamento": row["codigo_departamento"]
        })
```

```

print(f'Departamentos cargados: {len(departamentos)}')

# Crear sesión
session = requests.Session()

#GET inicial para obtener ViewState
resp = session.get(URL, timeout=10)
soup = BeautifulSoup(resp.text, "html.parser")
viewstate = soup.find("input", {"name": "javax.faces.ViewState"})["value"]

# Archivo CSV final
output_file = "municipios_sisaire.csv"
csv_header = ["id", "nombre_municipio", "codigo_municipio", "departamento_id"]

rows = []
autoinc = 1 # contador para id incremental

for dpto in departamentos:
    dep_id = dpto["id"]
    dep_code = dpto["codigo_departamento"]

    print(f"\nBuscando municipios del departamento: {dep_code} ...")

    payload = {
        "javax.faces.partial.ajax": "true",
        "javax.faces.source": "filtroForm:departamentoSel",
        "javax.faces.partial.execute": "filtroForm:departamentoSel",
        "javax.faces.partial.render": "filtroForm msgs-globales",
        "javax.faces.behavior.event": "valueChange",
        "javax.faces.partial.event": "change",
    }

```

```

"filtroForm": "filtroForm",
"filtroForm:departamentoSel_input": dep_code,
"javax.faces.ViewState": viewstate,
}

response = session.post(URL, data=payload, headers=headers)
soup_xml = BeautifulSoup(response.text, "xml")
update_filtro = soup_xml.find("update", {"id": "filtroForm"})

if not update_filtro:
    print(f"No se pudo obtener municipios de {dep_code}")
    continue

soup_html = BeautifulSoup(update_filtro.text, "html.parser")
select_municipios = soup_html.find("select", {"id": "filtroForm:municipioSel_input"})

if not select_municipios:
    print(f"Sin municipios para {dep_code}")
    continue

for opt in select_municipios.find_all("option"):
    value = opt.get("value")
    label = opt.text.strip()

    if value and value != "":
        rows.append({
            "id": autoinc,
            "nombre_municipio": label,
            "codigo_municipio": value,
            "departamento_id": dep_id
        })

```

```

    print(f' {label} ( {value})')
    autoinc += 1

# Guardar CSV final
with open(output_file, "w", newline="", encoding="utf-8") as f:
    writer = csv.DictWriter(f, fieldnames=csv_header)
    writer.writeheader()
    writer.writerows(rows)

print(f"\nArchivo generado: {output_file}")

```

B.3 Script para obtener las estaciones por municipio

Este script consulta SISAIRES municipio por municipio y extrae la información correspondiente a todas las estaciones registradas, incluyendo sus códigos y nombres. El resultado es un archivo con la estructura completa del sistema de monitoreo del país.

```

import requests
from bs4 import BeautifulSoup
import csv

URL = "http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml"

headers = {
    "Content-Type": "application/x-www-form-urlencoded",
    "Faces-Request": "partial/ajax",
    "User-Agent": "Mozilla/5.0"
}

# Cargar departamentos desde CSV
departamentos = {}

```

```

with open("departamentos_sisaire.csv", encoding="utf-8") as f:
    reader = csv.DictReader(f)
    for row in reader:
        departamentos[int(row["id"])] = row["codigo_departamento"]

# Cargar municipios desde CSV
municipios = []
with open("municipios_sisaire.csv", encoding="utf-8") as f:
    reader = csv.DictReader(f)
    for row in reader:
        municipios.append({
            "id": int(row["id"]),
            "codigo_municipio": row["codigo_municipio"],
            "departamento_id": int(row["departamento_id"])
        })

print(f"Municipios cargados: {len(municipios)}")

# Crear sesión
session = requests.Session()

# GET inicial para ViewState
resp = session.get(URL, timeout=10)
soup = BeautifulSoup(resp.text, "html.parser")
viewstate = soup.find("input", {"name": "javax.faces.ViewState"})["value"]

# Archivo final CSV
output_file = "estaciones_sisaire.csv"
csv_header = ["id", "nombre_estacion", "codigo_estacion", "municipio_id"]

rows = []

```

```

autoinc = 1

for mun in municipios:
    mun_id = mun["id"]
    mun_code = mun["codigo_municipio"]
    dep_code = departamentos[mun["departamento_id"]]

    print(f"\nMunicipio {mun_code} (dep {dep_code})...")

# ---- 1. Seleccionar departamento en la página ----
payload_dep = {
    "javax.faces.partial.ajax": "true",
    "javax.faces.source": "filtroForm:departamentoSel",
    "javax.faces.partial.execute": "filtroForm:departamentoSel",
    "javax.faces.partial.render": "filtroForm msgs-globales",
    "javax.faces.behavior.event": "valueChange",
    "javax.faces.partial.event": "change",
    "filtroForm": "filtroForm",
    "filtroForm:departamentoSel_input": dep_code,
    "javax.faces.ViewState": viewstate
}

resp_dep = session.post(URL, data=payload_dep, headers=headers)
soup_dep = BeautifulSoup(resp_dep.text, "xml")

# Actualizar el viewstate si aparece
vs_update = soup_dep.find("update", {"id": "j_id1:javax.faces.ViewState:0"})
if vs_update:
    viewstate = vs_update.text.strip()

# ---- 2. Seleccionar municipio → esto devuelve las estaciones ----

```

```

payload_mun = {
    "javax.faces.partial.ajax": "true",
    "javax.faces.source": "filtroForm:municipioSel",
    "javax.faces.partial.execute": "filtroForm:municipioSel",
    "javax.faces.partial.render": "filtroForm msgs-globales filtroForm:estacionesSel",
    "javax.faces.behavior.event": "valueChange",
    "javax.faces.partial.event": "change",
    "filtroForm": "filtroForm",
    "filtroForm:departamentoSel_input": dep_code,
    "filtroForm:municipioSel_input": mun_code,
    "javax.faces.ViewState": viewstate
}

```

```

resp_mun = session.post(URL, data=payload_mun, headers=headers)

```

```

soup_mun = BeautifulSoup(resp_mun.text, "xml")

```

```

found = False

```

```

# Buscar el bloque que contiene las estaciones

```

```

for upd in soup_mun.find_all("update"):

```

```

    if "estacionesSel" in upd.text:

```

```

        found = True

```

```

        html = BeautifulSoup(upd.text, "html.parser")

```

```

        checkboxes = html.find_all("input", {"type": "checkbox"})

```

```

    for cb in checkboxes:

```

```

        codigo_est = cb.get("value")

```

```

        label_tag = cb.find_next_sibling("label")

```

```

        nombre_est = label_tag.text.strip() if label_tag else "Sin nombre"

```

```

        rows.append({

```

```

        "id": autoinc,
        "nombre_estacion": nombre_est,
        "codigo_estacion": codigo_est,
        "municipio_id": mun_id
    })

    print(f'Estación: {nombre_est} ( {codigo_est} )')
    autoinc += 1

    break

if not found:
    print(f'No se encontraron estaciones para municipio {mun_code} ')

# ---- Guardar el CSV ----
with open(output_file, "w", newline="", encoding="utf-8") as f:
    writer = csv.DictWriter(f, fieldnames=csv_header)
    writer.writeheader()
    writer.writerows(rows)

print(f'\nArchivo generado: {output_file} ')

```

B.4 Script para consolidar y reestructurar los archivos CSV descargados

Este script integra los archivos individuales descargados desde SISAIRES, identifica automáticamente el contaminante contenido en cada archivo, normaliza la estructura de las columnas y genera un archivo unificado en formato ancho (wide format) donde cada contaminante ocupa una columna.

```
import pandas as pd
```

```
import glob

# Carpeta donde están los CSV
ruta = "ruta/*.csv"

# Lista de archivos
archivos = glob.glob(ruta)

dfs = []

for archivo in archivos:
    df = pd.read_csv(archivo)

    # Detectar el contaminante
    contaminante_detectado = None
    for contaminante in ["PM2.5", "PM10", "O3", "NO2"]:
        if contaminante in df.columns:
            contaminante_detectado = contaminante
            df["CONTAMINANTE"] = contaminante
            df["VALOR"] = df[contaminante]
            df = df.drop(columns=[contaminante])
            break

    if contaminante_detectado is None:
        print(f"No se detectó contaminante en: {archivo}")
        continue

    dfs.append(df)

# Unir todo
df_final = pd.concat(dfs, ignore_index=True)
```

```

# PIVOTE: convertir CONTAMINANTE en columnas
df_wide = df_final.pivot_table(
    index=["Estacion", "Fecha inicial"], # ← adapta si tus columnas se llaman distinto
    columns="CONTAMINANTE",
    values="VALOR",
    aggfunc="first" # si no hay duplicados sirve perfecto
).reset_index()

# Quitar el nombre del índice generado por pivot_table
df_wide.columns.name = None

# Guardar
output_path = "unidos_sisaire.csv"
df_wide.to_csv(output_path, index=False)

print("CSV creado:", output_path)

```

B.5 Script para enriquecer el dataset principal con información territorial

Este script toma como entrada los datos consolidados y los cruza con los archivos de departamentos, municipios y estaciones provenientes de los anexos anteriores. El resultado es un archivo final enriquecido con metadatos geográficos para cada observación.

```

import pandas as pd
import unidecode # <--- para eliminar tildes

# 1. Cargar archivos
unidos = pd.read_csv("unidos_sisaire.csv")
estaciones = pd.read_csv("estaciones_sisaire.csv")

```

```

municipios = pd.read_csv("municipios_sisaire.csv")
departamentos = pd.read_csv("departamentos_sisaire.csv")

# === 2. Normalizar nombres de estaciones ===
def normalizar(texto):
    if pd.isna(texto):
        return ""
    return unidecode.unidecode(str(texto)).strip().lower()

unidos["Estacion_norm"] = Unidos["Estacion"].apply(normalizar)
estaciones["nombre_estacion_norm"] = estaciones["nombre_estacion"].apply(normalizar)

# === 3. Merge usando columnas normalizadas ===
df = Unidos.merge(
    estaciones[["nombre_estacion_norm", "municipio_id"]],
    left_on="Estacion_norm",
    right_on="nombre_estacion_norm",
    how="left"
)

# (opcional) puedes borrar las columnas norm después
df = df.drop(columns=["Estacion_norm", "nombre_estacion_norm"])

# === 4. Unir con municipios ===
df = df.merge(
    municipios[["id", "nombre_municipio", "departamento_id"]],
    left_on="municipio_id",
    right_on="id",
    how="left"
)

```

```
# 5. Unir con departamentos
```

```
df = df.merge(  
    departamentos[['id','nombre_departamento']],  
    left_on='departamento_id',  
    right_on='id',  
    how='left'  
)
```

```
# 6. Seleccionar columnas finales
```

```
df_final = df[[  
    'Estacion',  
    'Fecha inicial',  
    'NO2',  
    'O3',  
    'PM10',  
    'PM2.5',  
    'nombre_departamento',  
    'nombre_municipio'  
]]
```

```
# 7. Guardar resultado
```

```
df_final.to_csv("SISAIRE_MERGE.csv", index=False)
```

```
print("Archivo generado: SISAIRE_MERGE.csv")
```

Apéndice C

Script Integración y Filtrado de Datos IDEAM–SISAIRE

El propósito de este anexo es documentar el proceso implementado para obtener un conjunto de datos consolidado entre las fuentes del IDEAM y SISAIRE, utilizado posteriormente en la tercera página del dashboard comparativo. Dado que ambas plataformas reportan información de forma independiente y con estructuras heterogéneas, fue necesario desarrollar un procedimiento de normalización y emparejamiento que permitiera identificar únicamente aquellos casos donde existieran datos coincidentes en términos de ubicación geográfica y fecha de observación.

```
import pandas as pd
```

```
# 1. Cargar archivos
```

```
df_ideam = pd.read_parquet(r"C:ruta\IDEAM_MERGE.parquet")
```

```
df_sisaire = pd.read_csv(r"C:ruta\SISAIRE_MERGE.csv")
```

```
# 2. Normalizar texto
```

```
def normalizar(texto):
```

```
    if pd.isna(texto):
```

```
        return ""
```

```
    return (str(texto).upper()
```

```
        .replace("Á", "A").replace("É", "E")
```

```
        .replace("Í", "I").replace("Ó", "O")
```

```
        .replace("Ú", "U")
```

```
        .replace("Ü", "U")
```

```
        .replace("-", " ")
```

```
        .strip())
```

```
df_ideam["dep_norm"] = df_ideam["Departamento"].apply(normalizar)
```

```
df_ideam["mun_norm"] = df_ideam["Municipio"].apply(normalizar)
```

```

df_ideam["fecha_norm"] = pd.to_datetime(df_ideam["Fecha"]).dt.date

df_sisaire["dep_norm"] = df_sisaire["nombre_departamento"].apply(normalizar)
df_sisaire["mun_norm"] = df_sisaire["nombre_municipio"].apply(normalizar)
df_sisaire["fecha_norm"] = pd.to_datetime(df_sisaire["Fecha inicial"]).dt.date

# 3. Pequeños reemplazos para equivalencias reales

reemplazos_muni = {
    "EL CERRITO": "CERRITO",
}

df_sisaire["mun_norm"] = df_sisaire["mun_norm"].replace(reemplazos_muni)

# -----
# 4. Hacer la unión (INNER JOIN)
# -----

df_merged = df_ideam.merge(
    df_sisaire,
    on=["dep_norm", "mun_norm", "fecha_norm"],
    how="inner",
    suffixes=("_IDEAM", "_SISAIRE")
)

# 5. Guardar resultado

output_path = r"C:ruta\IDEAM_SISAIRE_INTERSECCION.parquet"

df_merged.to_parquet(output_path)
print("Archivo generado:", output_path)
print("Filas finales:", len(df_merged))

```

Apéndice D

Script en R para el Análisis Estadístico Avanzado

```
# --- 1. Reiniciar entorno y cargar librerías ---
rm(list = ls())
gc()

suppressPackageStartupMessages({
  library(tidyverse)
  library(arrow)
  library(car)
  library(lmtest)
  library(MASS)
  library(boot)
  library(broom)
  library(nortest)
  library(ggcorrplot)
  library(mgcv)
  library(Hmisc)
  library(GGally)
  library(ppcor)
})

# --- 2. Cargar dataset ---
data_raw <- arrow::read_parquet(
  "C:\\ruta\\IDEAM_SISAIRE_INTERSECCION.parquet"
) %>%
  as.data.frame()

# --- FILTRAR SOLO 2021 ---
data_2021 <- data_raw %>%
```

```

filter(Fecha >= "2021-01-01", Fecha <= "2021-12-31")

# --- 3. Limpieza inicial ---
data_pivot <- data_2021 %>%
  rename_with(~ str_to_lower(.x)) %>%
  rename_with(~ gsub(" ", "_", .x)) %>%
  rename_with(~ gsub("\\.", "_", .x))

# --- 4. Selección de columnas ---
cols_exist <- intersect(
  c("departamento", "municipio", "fecha",
    "o3", "pm10", "pm2_5", "no2",
    "descripcionsensor", "valorpromedio"),
  names(data_pivot)
)

data_pivot <- data_pivot[, cols_exist, drop = FALSE]

# --- 5. Pivot meteorológico ---
data_pivot <- data_pivot %>%
  tidyr::pivot_wider(
    names_from = descripcionsensor,
    values_from = valorpromedio,
    values_fn = mean,
    values_fill = NA
  )

# --- Normalizar nombres ---
names(data_pivot) <- names(data_pivot) %>%
  str_replace_all("[\\(\\)]", "") %>%
  str_replace_all(" ", "_") %>%

```

```

str_to_lower()

# --- ELIMINAR VARIABLE ALTAMENTE COLINEAL ---
if ("temp_max_aire_2_m" %in% names(data_pivot)) {
  data_pivot$temp_max_aire_2_m <- NULL
}

# --- 5. Definir variables ---
vars_contaminantes <- c("no2", "o3", "pm10", "pm2_5")

# variables meteorológicas = todas excepto contams y columnas administrativas
vars_meteo <- names(data_pivot)[!(names(data_pivot) %in%
  c("departamento", "municipio", "fecha",
    "o3", "no2", "pm10", "pm2_5"))]

# --- 6. Asegurar tipos numéricos ---
for (v in c(vars_contaminantes, vars_meteo)) {
  if (v %in% names(data_pivot)) {
    data_pivot[[v]] <- suppressWarnings(as.numeric(data_pivot[[v]]))
  }
}

# --- 7. Filtrar registros útiles ---
data_pivot <- filter(data_pivot, !if_all(all_of(vars_contaminantes), is.na))

# --- 8. Estandarizar variables ---
data_scaled <- data_pivot %>%
  mutate(across(all_of(c(vars_contaminantes, vars_meteo)), scale))

# BLOQUE A: CORRELACIONES (ROBUSTO)

```

```

vars_exist <- intersect(c(vars_contaminantes, vars_meteo), names(data_scaled))

cor_test <- Hmisc::rcorr(as.matrix(data_scaled[, vars_exist]))
cor_r <- cor_test$r
cor_p <- cor_test$P

cor_r[!is.finite(cor_r)] <- 0
cor_p[!is.finite(cor_p)] <- 1

ggcorrplot(
  cor_r,
  hc.order = FALSE,
  type = "lower",
  lab = TRUE,
  lab_size = 2.5,
  p.mat = cor_p,
  sig.level = 0.05,
  insig = "blank",
  title = "Correlación entre variables meteorológicas y contaminantes"
)

write.csv(cor_r, "correlacion_r.csv")
write.csv(cor_p, "correlacion_pvalues.csv")

# PARES DE VARIABLES (SCATTER MATRIX)

vars_numericas <- vars_exist[sapply(data_scaled[, vars_exist, drop = FALSE], function(x) {
  is.numeric(x) && length(unique(na.omit(x))) > 1
})]

data_pairs <- data_scaled[, vars_numericas, drop = FALSE] %>%

```

```

as.data.frame() %>%
filter(rowSums(is.finite(as.matrix(.))) >= 2)

if (nrow(data_pairs) > 2 && ncol(data_pairs) >= 2) {
  GGally::ggscatmat(
    data = data_pairs,
    color = NULL,
    corMethod = "pearson"
  ) + theme_bw() +
  labs(title = "Relaciones entre contaminantes y meteorología")
}

# CORRELACIONES PARCIALES

cat("\n== CORRELACIONES PARCIALES (CONTROLANDO TODA LA
METEOROLOGÍA) ==\n")

vars_meteo_validas <- vars_meteo[vars_meteo %in% names(data_scaled)]

for (contam in vars_contaminantes) {
  cat("\n\n### Contaminante:", contam, "###\n")

  for (meteo in vars_meteo_validas) {

    # Variables de control = todas las meteorológicas excepto la que analizamos
    controles <- setdiff(vars_meteo_validas, meteo)

    df_valid <- data_scaled %>%
      dplyr::select(all_of(c(contam, meteo, controles))) %>%
      dplyr::filter(if_all(everything(), ~ is.finite(.)))
  }
}

```

```

# Verificar mínimos
if (nrow(df_valid) < 15 ||
    any(sapply(df_valid, function(x) length(unique(x))) < 2)) {
  cat("\n", contam, "-", meteo, ": Datos insuficientes, se omite\n")
  next
}

# Cálculo de correlación parcial
res <- try(
  ppcor::pcor.test(df_valid[[1]], df_valid[[2]], df_valid[, controles]),
  silent = TRUE
)

if (!inherits(res, "try-error")) {
  cat("\n", contam, "-", meteo, "\n")
  print(res)
} else {
  cat("\n", contam, "-", meteo, ": No se pudo calcular (colinealidad o singularidad)\n")
}
}
}

# BLOQUE B: FUNCIONES DE MODELADO

ajustar_modelo_avanzado <- function(var_dep, df) {

  form_lineal <- as.formula(paste(var_dep, "~",
    paste(vars_meteo, collapse = " + ")))

  form_no_lineal <- as.formula(paste(
    var_dep, "~",

```

```

paste(
  c(
    vars_meteo,
    paste0("I(", vars_meteo, "^2)"),
    "velocidad_viento_10_min:precipitacion"
  ),
  collapse = " + "
)
))

df <- filter(df, !if_any(all_of(c(var_dep, vars_meteo)), is.na))
if (nrow(df) < 10) return(list(contaminante = var_dep, error = "Datos insuficientes"))

modelo_lineal <- lm(form_lineal, data = df)
modelo_no_lineal <- lm(form_no_lineal, data = df)

# GAM usando TODAS las meteorológicas
gam_formula <- as.formula(
  paste0(var_dep, " ~ ", paste0("s(", vars_meteo, ")"), collapse = " + ")
)

modelo_gam <- mgcv::gam(gam_formula, data = df)

modelo_robusto <- MASS::rlm(form_lineal, data = df)

n <- nrow(df)
shapiro_p <- if (n >= 3 && n <= 5000) shapiro.test(residuals(modelo_lineal))$p.value else NA
ad_p <- tryCatch(nortest::ad.test(residuals(modelo_lineal))$p.value, error = function(e) NA)
bp <- lmtest::bptest(modelo_lineal)$p.value
dw <- lmtest::dwtest(modelo_lineal)$p.value
vif_values <- car::vif(modelo_lineal)

```

```
boot_fn <- function(df, index) coef(lm(form_lineal, data = df[index, ]))
boot_out <- boot::boot(df, boot_fn, R = 100)
```

```
resumen_lineal <- broom::tidy(modelo_lineal) %>%
  mutate(modelo = "Lineal",
         r2 = summary(modelo_lineal)$r.squared,
         r2_ajustado = summary(modelo_lineal)$adj.r.squared,
         shapiro_p = shapiro_p,
         ad_p = ad_p,
         breusch_p = bp,
         durbin_p = dw,
         AIC = AIC(modelo_lineal),
         BIC = BIC(modelo_lineal))
```

```
resumen_no_lineal <- broom::tidy(modelo_no_lineal) %>%
  mutate(modelo = "No lineal + Interacción",
         r2 = summary(modelo_no_lineal)$r.squared,
         r2_ajustado = summary(modelo_no_lineal)$adj.r.squared,
         AIC = AIC(modelo_no_lineal),
         BIC = BIC(modelo_no_lineal))
```

```
resumen_gam <- data.frame(
  term = names(coef(modelo_gam)),
  estimate = as.numeric(coef(modelo_gam))
) %>%
  mutate(modelo = "GAM",
         r2 = summary(modelo_gam)$r.sq,
         r2_ajustado = summary(modelo_gam)$dev.expl,
         shapiro_p = shapiro_p,
         ad_p = ad_p,
```

```

    breusch_p = bp,
    durbin_p = dw)

list(
  contaminante = var_dep,
  lineal = resumen_lineal,
  no_lineal = resumen_no_lineal,
  gam = resumen_gam,
  vif = vif_values,
  modelo_robusto = modelo_robusto,
  bootstrap = boot_out
)
}

# BLOQUE C: EJECUTAR MODELOS COMPLETOS PARA CADA CONTAMINANTE

resultados <- lapply(vars_contaminantes, function(v) ajustar_modelo_avanzado(v, data_scaled))

resumen_lineal <- bind_rows(lapply(resultados, \(x) x$lineal))
resumen_no_lineal <- bind_rows(lapply(resultados, \(x) x$no_lineal))
resumen_gam <- bind_rows(lapply(resultados, \(x) x$gam))

# BLOQUE D: EXPORTAR RESULTADOS

write.csv(resumen_lineal, "resumen_lineal_avanzado.csv", row.names = FALSE)
write.csv(resumen_no_lineal, "resumen_no_lineal_avanzado.csv", row.names = FALSE)
write.csv(resumen_gam, "resumen_gam_avanzado.csv", row.names = FALSE)

# --- Exportar VIF ---
vif_list <- lapply(resultados, function(x) {
  if (is.null(x$vif) || length(x$vif) == 0 ||

```

```

    is.null(names(x$vif)) || length(names(x$vif)) == 0) {
return(data.frame(
  Contaminante = x$contaminante,
  Variable = NA,
  VIF = NA
))
}

data.frame(
  Contaminante = x$contaminante,
  Variable = names(x$vif),
  VIF = as.numeric(x$vif)
)
})

vif_df <- bind_rows(vif_list)
write.csv(vif_df, "vif_resultados_avanzado.csv", row.names = FALSE)

# COMPARACION R2

comparacion_r2 <- dplyr::bind_rows(
  dplyr::select(resumen_lineal, dplyr::any_of(c("term", "r2", "modelo"))),
  dplyr::select(resumen_no_lineal, dplyr::any_of(c("term", "r2", "modelo"))),
  dplyr::select(resumen_gam, dplyr::any_of(c("term", "r2", "modelo")))
)

ggplot(comparacion_r2, aes(x = modelo, y = r2, fill = modelo)) +
  geom_boxplot() +
  theme_minimal()

ggsave("comparacion_r2_modelos.png", width = 7, height = 5)

```

```
# BLOQUE E: PLOTS GAM
```

```
for (var in vars_contaminantes) {
  gam_formula <- as.formula(
    paste0(var, " ~ ", paste0("s(", vars_meteo, ")"), collapse = " + ")
  )
  modelo_gam <- mgcv::gam(gam_formula, data = data_scaled)

  png(paste0("GAM_", var, ".png"), width = 1000, height = 800)
  plot(modelo_gam, pages = 1, residuals = TRUE, shade = TRUE, main = paste("GAM -", var))
  dev.off()
}
```

```
cat("\n Análisis completado exitosamente \n")
```

```
# ANALISIS FOCALIZADO: efecto de la velocidad del viento sobre PM10 y PM2.5
```

```
cat("\n===== \n")
cat("ANÁLISIS ESPECIAL: Viento → PM \n")
cat("===== \n")
```

```
library(ppcor)
library(mgcv)
library(MASS)
library(ggplot2)
```

```
# Variables objetivo
```

```
variables_pm <- c("pm10", "pm2_5")
var_viento <- "velocidad_viento_10_min"
```

```

# Asegurar que la variable exista
if (!(var_viento %in% names(data_scaled))) {
  stop("No existe la columna velocidad_viento_10_min en el dataset depurado.")
}

# --- 1. Correlaciones simples ---
for (pm in variables_pm) {
  df_plot <- na.omit(data.frame(
    PM = data_scaled[[pm]],
    viento = data_scaled[[var_viento]]
  ))

  # Gráfico correlación simple
  g <- ggplot(df_plot, aes(x = viento, y = PM)) +
    geom_point(alpha = 0.1) +
    geom_smooth(method = "lm", color = "blue", se = TRUE) +
    theme_minimal() +
    labs(
      title = paste("Correlación simple:", pm, "vs velocidad del viento"),
      x = "Velocidad del viento (estandarizada)",
      y = paste(pm, "(estandarizado)")
    )

  ggsave(paste0("01_correlacion_simple_", pm, ".png"), width = 7, height = 5)

  # Tabla con resultados de la correlación
  cor_simple <- cor.test(df_plot$PM, df_plot$viento)
  write.csv(
    data.frame(
      variable = pm,
      cor = cor_simple$estimate,

```

```

    p_value = cor_simple$p.value,
    conf_low = cor_simple$conf.int[1],
    conf_high = cor_simple$conf.int[2]
  ),
  paste0("01_correlacion_simple_tabla_", pm, ".csv"),
  row.names = FALSE
)
}

# --- 2. Correlación parcial controlando meteorología ---
resultados_pc <- list()

for (pm in variables_pm) {
  df_valid <- data_scaled |>
  dplyr::select(all_of(c(pm, var_viento, controles))) |>
  na.omit()

  res_pc <- ppcor::pcor.test(df_valid[[pm]], df_valid[[var_viento]],
    df_valid[, controles])

  resultados_pc[[pm]] <- data.frame(
    contaminante = pm,
    estimate = res_pc$estimate,
    p_value = res_pc$p.value,
    statistic = res_pc$statistic,
    n = res_pc$n
  )
}

tabla_pc <- dplyr::bind_rows(resultados_pc)
write.csv(tabla_pc, "02_correlacion_parcial_pm.csv", row.names = FALSE)

```

```

# --- 3. Modelo lineal simple ---
for (pm in variables_pm) {
  modelo_lm <- lm(as.formula(paste0(pm, " ~ ", var_viento)), data = data_scaled)

  # Guardar tabla
  tabla <- broom::tidy(modelo_lm)
  tabla$r2 <- summary(modelo_lm)$r.squared
  write.csv(tabla, paste0("03_lm_tabla_", pm, ".csv"), row.names = FALSE)

  # Gráfico de la recta de regresión
  df_plot <- na.omit(data.frame(PM = data_scaled[[pm]],
                               viento = data_scaled[[var_viento]]))

  g <- ggplot(df_plot, aes(x = viento, y = PM)) +
    geom_point(alpha = 0.1) +
    geom_smooth(method = "lm", color = "red") +
    theme_minimal() +
    labs(
      title = paste("Modelo lineal:", pm, "~ viento"),
      x = "Velocidad del viento",
      y = pm
    )

  ggsave(paste0("03_lm_grafico_", pm, ".png"), width = 7, height = 5)
}

# --- 4. GAM reducido: PM ~ s(viento) ---
for (pm in variables_pm) {
  modelo_gam_viento <- mgcv::gam(

```

```

as.formula(paste0(pm, " ~ s(", var_viento, ", k = 10)")),
data = data_scaled
)

# Tabla
tabla_gam <- data.frame(
  edf = summary(modelo_gam_viento)$s.table[1, "edf"],
  F = summary(modelo_gam_viento)$s.table[1, "F"],
  p_value = summary(modelo_gam_viento)$s.table[1, "p-value"],
  r2 = summary(modelo_gam_viento)$r.sq,
  deviance_explained = summary(modelo_gam_viento)$dev.expl
)
write.csv(tabla_gam, paste0("04_gam_tabla_", pm, ".csv"), row.names = FALSE)

# Gráfico GAM
png(paste0("04_gam_grafico_", pm, ".png"), width = 900, height = 700)
plot(modelo_gam_viento, residuals = TRUE, shade = TRUE,
     main = paste("Efecto no lineal del viento sobre", pm))
dev.off()
}

# --- 5. Regresión robusta ---
resultados_rlm <- list()

for (pm in variables_pm) {
  modelo_rlm <- MASS::rlm(as.formula(
    paste0(pm, " ~ ", var_viento)), data = data_scaled)

  tabla_rlm <- data.frame(
    contaminante = pm,

```

```
intercept = coef(modelo_rlm)[1],
coef_viento = coef(modelo_rlm)[2],
se_intercept = summary(modelo_rlm)$coefficients[1,2],
se_viento = summary(modelo_rlm)$coefficients[2,2],
t_intercept = summary(modelo_rlm)$coefficients[1,3],
t_viento = summary(modelo_rlm)$coefficients[2,3]
)

resultados_rlm[[pm]] <- tabla_rlm
}

write.csv(
  dplyr::bind_rows(resultados_rlm),
  "05_rlm_resultados.csv",
  row.names = FALSE
)

cat("\n Análisis viento-PM completado \n")
```