



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escuela Técnica Superior de Ingeniería Informática
Universidad Politécnica de Valencia

Estudio de la calidad del aire en Valencia

TRABAJO DE PROYECTO II

Grado en Ciencia de Datos

Autores: Adriá Aguilar Mínguez
Pablo Gandía Miñana
Santiago Font Piña
Fernando Martínez Gómez
Sergio Ortiz Montesinos
Tutores: Sara Blanc Clavero
M^a José Ramírez Quintana

Curso 2024-2025

Índice general

Índice general	1
Índice de figuras	5
1 Introducción	7
1.1 Motivación	7
1.2 Objetivos	7
2 Configuración del proyecto	9
2.1 Fuentes de datos	9
2.2 Integración y transformación de los datos	9
2.3 Herramientas y técnicas utilizadas	10
3 Resultados obtenidos	11
3.1 Objetivo 1: Análisis temporal y espacial de los contaminantes atmosféricos en Valencia	11
3.2 Objetivo 2: Identificación de patrones meteorológicos y su relación con los niveles de contaminación atmosférica	14
3.3 Objetivo 3: Caracterización del tráfico en Valencia y su impacto en la contaminación	17
3.4 Objetivo 4: Estudiar las emisiones de CO ₂ y el comportamiento de los inmuebles de la ciudad de Valencia	19
3.5 Objetivo 5: Patrones entre contaminantes y factores meteorológicos, y relación entre contaminantes y zonas verdes	22
4 Lecciones para futuros proyectos	25
<hr/>	
Apéndices	
A Actualización Hito 1 (CORREGIR DE NUEVO)	27
A.1 Técnicas de obtención de datos y extracción	27
A.2 Análisis de las fuentes: interpretación de los datos, valoración de su utilidad en el proyecto	28
A.2.1 Datos atmosféricos y climatológicos	28
A.2.2 Intensidad del tráfico	29
A.2.3 Certificados Eficiencia Energética	30
A.2.4 Espacios verdes	30
B Actualización Hito 2 (CORREGIR DE NUEVO)	31
B.1 Objetivo principal del proyecto ¿qué presenta este estudio?	31
B.2 ¿Para qué y para quién podría ser de utilidad este estudio?	31
B.3 ¿Por qué es novedoso? ¿Hay estudios similares?	31
B.4 Objetivos definitivos del proyecto	32
C Objetivo 1	33
C.1 Enlaces a los html	33
C.1.1 Estudio descriptivo y temporal de los contaminantes	33
C.1.2 Mapas de calor de los contaminantes	33
C.1.3 Análisis de interrelaciones entre contaminantes	33
C.2 Figuras del Objetivo 1	33

D	Objetivo 2	37
D.1	Enlaces a los códigos	37
D.1.1	PCA sobre las variables meteorológicas	37
D.1.2	Clustering sobre las variables meteorológicas	37
D.1.3	Clustering sobre los contaminantes cruzados con meteorología	37
D.1.4	PLS sobre las variables meteorológicas y los contaminantes	37
D.2	Figuras del Objetivo 2	38
E	Objetivo 3	41
E.1	Enlaces a los html	41
E.1.1	Resumen de los datos y mapa Mapa	41
E.1.2	Pasar datos a horarios	41
E.1.3	Distribuciones variables tráfico	41
E.1.4	Tipos de vías (clustering)	41
E.1.5	Cruce con Contaminación	41
E.1.6	Relaciones con Contaminación	41
E.1.7	Agrupaciones y comportamientos	41
E.2	Figuras del Objetivo 3	42
E.2.1	PCA: Tráfico y contaminación sin relación aparente	42
E.2.2	Agrupación contaminantes altos, tráfico alto, dirección del viento interesante	42
F	Objetivo 4	43
F.1	Enlaces a los html	43
F.1.1	Rmarkdown compilado en HTML	43
F.1.2	Mapas interactivos	43
F.2	Figuras del Objetivo 4	43
F.2.1	Clustering de las referencias catastrales según sus emisiones y antigüedad	43
F.2.2	Clustering de las parcelas catastrales según sus emisiones y antigüedad	44
F.2.3	Correlaciones CO ₂ y contaminates	44
F.2.4	Patrón horario invierno de contaminantes	45
G	Objetivo 5	47
G.1	Enlaces a los html	47
G.1.1	Creación de dataframe para reglas de asociación	47
G.1.2	Reglas de asociación	47
G.1.3	Zonas verdes y contaminación	47
G.2	Figuras del Objetivo 5	47
H	Información extra sobre los objetos de estudio	49
H.1	NO _x (NO + NO ₂)	49
H.1.1	¿Qué es?	49
H.1.2	Valores límite establecidos	49
H.1.3	Causantes	49
H.1.4	Efectos sobre la salud humana y los ecosistemas	50
H.2	O ₃	51
H.2.1	¿Qué es?	51
H.2.2	Valores límites establecidos	51
H.2.3	Causantes	51
H.2.4	Efectos sobre la salud humana y los ecosistemas	52
H.3	PM _{2,5} y PM ₁₀	53
H.3.1	¿Qué es?	53
H.3.2	Valores límite establecidos	53

H.3.3	Causantes	53
H.3.4	Efectos sobre la salud humana y los ecosistemas	54
H.4	SO ₂	55
H.4.1	¿Qué es?	55
H.4.2	Valores límite establecidos	55
H.4.3	Causantes	55
H.4.4	Efectos sobre la salud humana y los ecosistemas	55
H.5	Factores climatológicos y su influencia en la calidad del aire	57
H.6	Contextualización de Valencia	58
H.6.1	Fuentes principales de contaminación	58
H.6.2	Estructura urbana y dispersión atmosférica	58
H.6.3	Movilidad y transporte	58
H.6.4	Turismo y eventos multitudinarios	58

Índice de figuras

3.1	PCA variables meteorológicas	15
3.2	Box Plots de emisiones de CO ₂ por usos del edificio	19
C.1	Ejemplo de maoa de calor	33
C.2	Ejemplo de evolución menusal	34
C.3	Ejemplo de evolución semanal	34
C.4	Ejemplo de evolución horaria	35
C.5	Ejemplo de faltantes	35
C.6	PCA Contaminantes	35
C.7	Clustering de contaminantes	36
D.1	Tabla clustering meteorológico	38
D.2	Tabla clustering contaminantes cruzado con meteorología	39
E.1	PCA	42
E.2	Agrupación	42
F.1	Clustering de las referencias catastrales	43
F.2	Clustering de las referencias catastrales	44
F.3	Correlacion entre CO2 y el testeo de contaminantes	44
F.4	Correlación entre CO2 y el testeo de contaminantes	45
G.1	Correlación entre el área de zona verde y los contaminantes	47

CAPÍTULO 1

Introducción

En este trabajo se hace un estudio de la calidad del aire en la ciudad de Valencia en el periodo de tiempo 2019 - 2024. Para ello, se tiene en cuenta el tráfico, zonas verdes, emisiones de los inmuebles, factores climatológicos y la relación de todos los anteriores con los principales contaminantes atmosféricos.

1.1 Motivación

El proyecto tiene como motivación aportar información y transparencia sobre la calidad del aire en Valencia a instituciones públicas (Ayuntamiento de Valencia, Generalitat Valenciana, Gobierno de España y Parlamento Europeo) y empresas privadas, facilitando la toma de decisiones para adaptarse al Marco legislativo nacional y europeo, así como para mejorar los aspectos de la ciudad que se consideren necesarios.

1.2 Objetivos

1. Explicar la contaminación en la ciudad de Valencia por zonas y contaminantes.
2. Estudiar la influencia de los factores meteorológicos en la contaminación.
3. Caracterizar el tráfico en Valencia y su impacto en la contaminación.
4. Examinar las emisiones de los inmuebles y su relación con la contaminación.
5. Investigar las relaciones entre las zonas verdes y los contaminantes y los patrones entre los factores meteorológicos y los contaminantes.

CAPÍTULO 2

Configuración del proyecto

2.1 Fuentes de datos

Fuentes de contaminación:

- [Emisiones de CO₂ de cada Vivienda](#)

Fuentes de datos atmosféricos:

- [Estaciones Atmosféricas](#)
- [Temperatura en Valencia 2018-2022](#)
- [Temperatura en Valencia 2020-2025](#)

Fuentes de tráfico:

- Datos de circulación de las vías en Valencia, facilitada por ETRA y el Ayuntamiento de Valencia, no podemos difundirla abiertamente.
- [Coordenadas de espiras tráfico](#)

Fuentes de espacios verdes:

- [Espacios verdes en Valencia](#)

2.2 Integración y transformación de los datos

Uno de los retos más importantes de este proceso ha sido transformar y agrupar datos de varios sitios y fuentes distintas para poder extraer información útil.

- Se han juntado bases de datos de contaminantes de años distintos para tener datos más compactos (de una misma estación).
- Se han asociado tanto estaciones como espiras de tráfico a su ubicación real por medio del cruce con bases de datos de coordenadas para poder representar mapas.

- Se ha cambiado la granularidad de los datos de transporte de intervalos de 5 minutos a datos anuales para poder comparar y cruzar con otras bases disponibles, como contaminación y dirección de viento.
- Se han obtenido variables agrupadas (de otras) para ampliar la información y variabilidad en los cambios de granularidad.
- Se han contrastado diversas fuentes como Google Maps para asegurar y corregir a mano ubicaciones poco coherentes.
- Se han integrado API's y Scrappers de Python para obtener, completar y ampliar la información de zonas geográficas urbanas disponibles para así poder realizar análisis más completos (objetivo 3 y 4).

2.3 Herramientas y técnicas utilizadas

Los lenguajes que se han utilizado para la transformación de datos e integración han sido **Python** y **R**.

El primero se ha empleado principalmente para la **obtención de datos mediante técnicas de *web scraping***, utilizando mayoritariamente el entorno interactivo **Jupyter Notebooks**. Las librerías utilizadas en este proceso incluyen: `requests`, `BeautifulSoup`, `pandas`, y `selenium`, entre otras.

Por su parte, el lenguaje **R** se ha utilizado tanto para la **transformación de datos** como para el **análisis completo del proyecto**, incluyendo el estudio de contaminantes, la ejecución del **análisis de componentes principales (PCA)**, el **clustering**, el **PLS**, las **Reglas de Asociación** y la integración de los resultados de los distintos objetivos. Las librerías más utilizadas han sido: `dplyr`, `ggplot2`, `factoextra`, `cluster`, `zoo` y `sf`, entre muchas otras.

También hemos utilizado **GitHub** y el controlador de versiones `Git` para enviarnos los datos y tener un lugar en la nube al que publicar los programas para que fueran accesibles para todos.

CAPÍTULO 3

Resultados obtenidos

3.1 Objetivo 1: Análisis temporal y espacial de los contaminantes atmosféricos en Valencia

Este primer objetivo se ha estructurado en dos fases. En la primera, se estudiaron y caracterizaron estadísticamente los distintos patrones temporales y espaciales de los contaminantes en Valencia. En la segunda, se exploraron las relaciones conjuntas entre dichos contaminantes.

Para ello, se empleó una base de datos con registros horarios de contaminantes por zona, obtenida al unificar las distintas bases de datos de cada área en una única fuente de información.

Para abordar este objetivo, resultó imprescindible contar con una visión general de los contaminantes, entendiendo su naturaleza, sus principales fuentes de emisión y sus efectos sobre la salud humana.

1. Estudio descriptivo y temporal de los contaminantes

En primer lugar, se analizaron las distribuciones de los contaminantes (PM_{10} , $PM_{2.5}$, SO_2 , NO , NO_x , NO_2 , O_3 y CO), comprobando que ninguna seguía una curva normal: todas mostraban sesgo hacia la derecha y, en varios casos, una única moda. Para facilitar su estudio y aproximarse a distribuciones más simétricas, se aplicaron transformaciones logarítmicas a los datos.

En primer lugar, se evaluaron las distribuciones de los contaminantes registrados en las distintas estaciones de medición, observándose patrones globalmente similares entre ellas. Con el fin de profundizar en las posibles diferencias regionales, se representaron gráficamente las medias y desviaciones típicas de cada contaminante por estación, así como mapas de calor que pusieron de manifiesto variaciones aparentemente significativas en las concentraciones según la ubicación geográfica. Vease un ejemplo en la Figura C.1.

Posteriormente, se realizaron pruebas ANOVA para cada contaminante en función del área de medición, lo que permitió confirmar estadísticamente la existencia de diferencias significativas entre zonas.

Cabe destacar que se detectaron valores extremadamente atípicos en algunas estaciones, al parecer relacionados con el episodio meteorológico de la **DANA**. Ante esta situación, se llevó a cabo un análisis más exhaustivo, identificando las estaciones con los registros más elevados: **Massanassa UM**, **Paiporta UM**, **Catarroja UM** y **Benetússer UM**.

Estas estaciones están localizadas en zonas que se vieron especialmente afectadas por la **DANA**. Al analizar la fecha de inicio de sus mediciones, se observó que todas comenzaron a registrar datos pocas semanas después del evento, lo que sugiere que fueron instaladas específicamente para monitorizar las consecuencias de dicho fenómeno.

Seguidamente, se examinó la evolución temporal de las concentraciones:

- **Patrones estacionales:** varios contaminantes mostraron estacionalidad, con ciclos periódicos de subidas y bajadas a lo largo del año.
- **Tendencia anual:** Los valores medios anuales fueron elevados en 2019, cayeron notablemente en 2020 por las restricciones derivadas de la pandemia de **COVID-19** y registraron un repunte en 2023; no obstante, el balance global apunta a un ligero descenso de la media anual a lo largo del periodo, incluso en el caso del O_3 , a pesar de los esfuerzos europeos por mantener concentraciones adecuadas.
- **Variación mensual:** contaminantes como NO , NO_2 , NO_x y CO presentaron concentraciones más altas en invierno que en verano, sugiriendo una influencia de la temperatura.
- **Patrones semanales:** excepto O_3 (que tiende a aumentar los fines de semana), el resto de contaminantes muestra niveles inferiores en sábado y domingo, lo que apunta al impacto del tráfico y la actividad industrial en días laborables.
- **Patrones horarios:** la mayoría registra un primer pico de 6:00 a 12:00, seguido de un descenso hasta las 16:00 y un segundo repunte, coincidiendo con los periodos de entrada y salida al trabajo.

Vease un ejemplo de cada uno en la Figura C.2, Figura C.3 y Figura C.4.

Finalmente, se analizó la distribución de los datos faltantes, observando que la estación de Torrent–El Vedat acumulaba la mayor proporción, aunque sin un patrón claro, lo que sugiere que las ausencias fueron aleatorias. Vease un ejemplo en la Figura C.5.

Ver código de este apartado: 1.1 - Código del análisis descriptivo y temporal y 1.2 Mapas de calor de los contaminantes.

2. Análisis de interrelaciones entre contaminantes

En esta fase se seleccionaron únicamente las estaciones que medían todos los contaminantes salvo CO (excluido por su escasez de registros y alto porcentaje de datos faltantes). Puesto que las ausencias representaban menos del 5 % en cada serie, se imputaron valores mediante la media de las observaciones más cercanas en el mismo intervalo horario y estación, garantizando la conservación de la distribución original.

Con las series completas, se aplicaron dos técnicas multivariantes:

- **Análisis de Componentes Principales (PCA)** utilizando tres componentes, cuyos resultados fueron:
 - Fuerte correlación entre PM_{10} y $PM_{2.5}$.
 - Fuerte correlación entre NO , NO_x y NO_2 .
 - Correlaciones leves entre partículas (PM) y óxidos de nitrógeno (NOs).
 - Relación muy débil de SO_2 con los anteriores.
 - O_3 sin relación o con correlación negativa frente al resto.

Vease en la Figura C.6

- **Clustering Jerárquico:** se identificaron tres grupos de estaciones según su perfil contaminante. Para determinar el número óptimo de clústeres se emplearon dos métodos complementarios: el método del codo y el coeficiente de Silhouette, ambos coincidiendo en que la solución más adecuada era la partición en tres grupos.
 - **Grupo 1:** Massanassa UM.
 - **Grupo 2:** Quart de Poblet, Torrent (El Vedat), València Av. Francia, València Molí del Sol, y la UPV.
 - **Grupo 3:** Sedaví, Pista de Silla, y ambas esatciones del puerto (Port Moll Trans. Ponent y Port llit antic Túrria).

Vease en la Figura C.7

El código de este apartado: 1.2 - Código de PCA y Clustering.

En conclusión, existen diferencias significativas en los niveles de contaminantes según la zona y relaciones claras entre ellos, atribuidas principalmente al tráfico y a factores meteorológicos, cuya influencia se abordará en las siguientes fases del estudio.

Además, en términos globales, excluyendo el periodo de la pandemia, se observa una ligera pero constante disminución de los niveles anuales de contaminantes, lo que apunta a una mejora progresiva de la calidad del aire en Valencia.

3.2 Objetivo 2: Identificación de patrones meteorológicos y su relación con los niveles de contaminación atmosférica

El segundo objetivo del proyecto consistió en analizar de forma conjunta las condiciones meteorológicas y los niveles de contaminantes atmosféricos registrados en Valencia. Se persiguió una doble meta: por un lado, caracterizar estadísticamente los diferentes patrones climáticos que se repiten en la ciudad a lo largo del tiempo; por otro, estudiar cómo estos patrones influyen en la concentración de contaminantes como NO, NO₂, NO_x, O₃, PM10, PM2.5 y SO₂.

Para ello se trabajó con la base de datos de meteorología horaria, descartando la base por zonas de temperaturas diarias ya que su incorporación habría reducido la granularidad temporal del análisis e incrementado significativamente la complejidad del modelo debido al tratamiento espacial de múltiples coordenadas.

Análisis multivariante de condiciones meteorológicas

Primeramente, se estudiaron las distribuciones, mediante histogramas, de las principales variables meteorológicas: temperatura, humedad, viento, nubosidad, visibilidad y precipitaciones. Observando patrones interesantes como la mixtura de poblaciones en la variable viento, conformada por la población de viento de poniente y la población viento de levante. Así mismo se observó también una unión de poblaciones en la variable temperatura, conformada por las temperaturas de invierno y las temperaturas de verano.

Tras estudiar las distribuciones de las variables meteorológicas, la variable ángulo del viento fue sustituida por su seno y coseno (teniendo en cuenta que $\sin Y = 1$ es viento procediente del este y que $\sin X = 1$ es viento procediente del norte). A continuación, se aplicó un PCA con el objetivo de reducir la dimensionalidad y extraer conclusiones sobre dichas variables (véase apéndice D.1.1). A partir del scree plot obtenido seleccionamos las primeras 3 componentes que explicaban un 64 % de la variabilidad total. Los resultados del PCA permitieron observar asociaciones meteorológicas coherentes y significativas. En primer lugar, se detectó una fuerte correlación positiva entre precipitaciones, nubosidad y humedad, variables que, a su vez, mostraban una relación inversa con la temperatura y la visibilidad. Además, el análisis reveló un patrón direccional en los vientos: los vientos del sur se asociaban con temperaturas más elevadas, mientras que los procedentes del norte estaban relacionados con temperaturas más bajas. De forma complementaria, el viento de poniente (proveniente del oeste) se vinculaba con condiciones más frías, en contraste con el viento de levante (desde el este), que se asociaba a temperaturas más altas. Este último fenómeno se explica por una marcada estacionalidad: en invierno predominan los vientos de poniente, mientras que en verano se incrementa notablemente la frecuencia del levante.

A continuación se presenta un gráfico que sintetiza visualmente estas relaciones a través de las cargas de las variables los planos seleccionados del PCA:

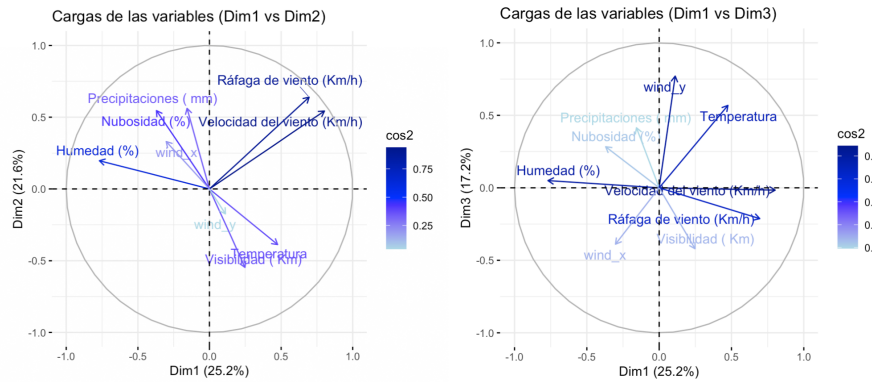


Figura 3.1: PCA variables meteorológicas

Sabiendo ya cómo se estructuraban las variables, se aplicó un clustering jerárquico (véase apéndice D.1.2) estableciendo el número de grupos en 5, mediante los gráficos de silhouette y de WSS. Los clusters seleccionados representaban diferentes situaciones meteorológicas recurrentes en Valencia: días invernales despejados y ventosos, jornadas veraniegas con brisa marina, días fríos y calmados, situaciones de transición primaveral y episodios de lluvia con temperaturas más bajas y fuerte viento (Vease en la Figura D.1).

Integración con datos de calidad del aire

Tras fusionar las bases de datos de meteorología y calidad del aire mediante una unión por fecha y hora, se aplicó un clustering jerárquico (véase Anexo D.1.3) sobre las concentraciones de contaminantes atmosféricos (NO , NO_2 , NO_x , O_3 , PM_{10} , $\text{PM}_{2.5}$ y SO_2). El número óptimo de clusters se determinó mediante el análisis conjunto de los gráficos de silhouette y WSS, estableciéndose en siete grupos.

Cada uno de estos clusters reflejaba un escenario de contaminación atmosférica distinto, mostrándose correspondencias claras con condiciones meteorológicas específicas. En particular, se observó que viento débil, bajas temperaturas y alta humedad favorecían la acumulación de contaminantes primarios (NO , NO_2 , NO_x , PM_{10} y $\text{PM}_{2.5}$), mientras que los niveles elevados de ozono (O_3) se asociaban a días soleados, secos y con brisa de levante, condiciones propicias para su formación por fotólisis (Vease en la Figura D.2)

Modelado de la relación meteorología–contaminación

Para cuantificar las influencias obtenidas de cada contaminante por las variables meteorológicas, se ajustó un modelo de regresión PLS (véase apéndice D.1.4). Este permitió determinar qué factores del clima explicaban mejor la variación en cada tipo de contaminante. Los resultados revelaron patrones consistentes :

- NO , NO_2 y NO_x aumentaban con temperaturas frías, baja humedad, cielos despejados y poco viento, condiciones típicas de inversión térmica.
- O_3 se elevaba en días cálidos, secos y con brisa muy leve del este.
- PM_{10} y $\text{PM}_{2.5}$ se intensificaban con viento débil, alta humedad y temperaturas más bajas, señalando su origen mixto: tanto urbano como de formación secundaria. Además, se observaba como el $\text{PM}_{2.5}$ es más susceptible al viento que el PM_{10} , muy probablemente debido a su tamaño.

- SO₂ mostró poca dependencia del clima, indicando que su variabilidad responde más a fuentes puntuales de emisión que al entorno atmosférico.

3.3 Objetivo 3: Caracterización del tráfico en Valencia y su impacto en la contaminación

Planteamiento

Los datos de los que disponemos son los recogidos por 1129 espiras de tráfico distintas, situadas por toda la ciudad, y con principal presencia en las grandes vías. (*Véase apéndice E.1.1*)

El funcionamiento de estas espiras es simple, son sensores electromagnéticos, colocados a lo largo de un carril de la vía, capaces de registrar si un vehículo pasa por encima por medio de inducción magnética, en nuestro caso registran la cantidad de coches que pasan por la vía aproximado por hora (intensidad), el % de tiempo que la espira tiene un vehículo arriba (ocupación) y la velocidad media.

Puesto que los datos estaban en intervalos de 5 minutos ha sido necesario un cambio de gránulo temporal a horario, generando varios estadísticos nuevos. (*Véase apéndice E.1.2*)

En cuanto a las distribuciones observadas:

(*Véase apéndice E.1.3*)

Tráfico mínimo entre las 22:00 y las 6:00 de la mañana, intensidades y ocupación bajas, al igual que el valor de muchos contaminantes, especialmente los óxidos de nitrógeno. Los picos horarios se alcanzan al inicio y el fin de la jornada laboral 7-8 y 18:00 con máximos medios de 600 vehículos por hora y ocupaciones superiores al 8 %, coincidiendo esta subida con la de algunos contaminantes por la mañana.

También coincide el comportamiento semanal, cuando menor volumen de tráfico hay los Sábados y Domingos, la gente no trabaja y no se desplazan, aunque se dan picos de tráfico los Viernes por posibles salidas de la ciudad a otros destinos. Valores uniformes el resto de días.

Por último vemos un comportamiento mensual muy marcado por festivos, la menor actividad laboral y productiva en Semana Santa (Abril) y Verano (Agosto) lleva a la gente a realizar menos desplazamientos, reduciendo el volumen de tráfico. Paralelamente hay un descenso de los óxidos de nitrógenos y los PM (2.5 y 10), pero sin afirmar que exista relación real directa.

Los patrones de tráfico de cada vía se intuyen con el comportamiento de ocupación, las vías con intensidades y velocidades significativas también tienen valores medios de ocupación, aunque para valores altos presentan intensidades y velocidades bajas, que se da cuando hay coches en la vía pero detenidos (un atasco).

Caracterización de Vías:

(*Véase apéndice E.1.4*)

Esto nos lleva a segmentar el tipo de vía en el que se sitúa la espira para caracterizar mejor posibles influencias distintas en cuanto a contaminación. Utilizando un clustering de partición diferenciamos 3 importantes grupos de tamaño similar:

-Vías con alta intensidad y velocidad, avenidas principales de gran fluidez. -Vías de bajo volumen de tráfico, velocidad ocupación e intensidad bajas, calles de barrio. -Vías de intensidad y velocidad moderada pero con ocupaciones máximas muy altas, principal-

mente situadas en el centro y conectando a grandes avenidas, zonas de congestión que posiblemente contaminen más .

Aunque las distribuciones y patrones que seguían estas vías eran prácticamente idénticas (algunas con mayores o menores valores).

Influencia Contaminante:

Para estudiar su influencia con contaminación hemos seleccionado únicamente 4 zonas que coinciden con la ubicación de nuestras espiras de tráfico y con los contaminantes analizados en el objetivo 1 (en relación a sus nulos):

Molí del Sol, Pista de Silla, El Politécnico (altura a facultad de Arquitectura) y **Avenida de Francia** (*Véase apéndice E.1.5*)

Aunque en Valencia, no hemos seleccionado las dos estaciones del puerto por las pocas espiras de tráfico cercanas y el sesgo que implica la actividad portuaria.

Para el estudio generamos una base de datos horaria con las medias de las espiras de tráfico que se encontraban en un radio dado a la estación (1000 m), diferenciando según el tipo de vía explicado antes, y además se le añadió la dirección del viento categorizada.

Haciendo el estudio de relaciones (*Véase apéndice E.1.6*), visualizando matrices de correlaciones y aplicando modelos como PCA no obtuvimos resultados concluyentes con los que afirmar que existe una relación directa entre el tráfico y la contaminación, al menos no bajo el paradigma de relaciones lineales o con el enfoque simple utilizado. *Véase Figura E.1*

Esto es comprensible ya que las tendencias de tráfico son muy cambiantes, al igual que la contaminación, sumado a que las mediciones generadas por las espiras y centrales de contaminación son bastante imprecisas y se complica mucho encontrar modelos que sí expliquen la varianza de contaminación con tráfico.

Agrupaciones:

No obstante, sí se ha observado una posible influencia aplicando técnicas de agrupación (clustering de partición con tráfico, contaminación y viento en cada una de las estaciones). (*Véase apéndice E.1.7*)

Seleccionando diferente grupos (patrones) de distribuciones en cada zona observamos que los grupos con mayores valores de contaminantes presentan también valores altos en cuanto a volumen de tráfico, y además, la dirección del viento siempre es desde las zonas con mayor tráfico (vías de alta intensidad) hacia los medidores de contaminantes, lo que podría ser interpretado como que el viento arrastra las emisiones de tráfico hacia estas estaciones (como ocurre en el Politécnico con los vientos del Oeste y Sur-Oeste: *Véase Figura E.2*), aumentando los valores de los contaminantes estudiados, menos O_3 y en ocasiones SO_2 que, como habíamos visto, tenían un comportamiento independiente al resto.

Es decir, aunque sí que coincide que en los periodos de mayor tráfico los contaminantes presentan valores más altos, esto puede deberse a otros factores más generales como momentos de mayor actividad humana general (trabajo, productividad...), y sin un enfoque muy complejo es imposible cuantificar estas relaciones.

3.4 Objetivo 4: Estudiar las emisiones de CO₂ y el comportamiento de los inmuebles de la ciudad de Valencia

Planteamiento

Con el fin de profundizar en los factores que podrían influir en la calidad del aire de la ciudad de València, se han incorporado al análisis las estimaciones de emisiones de dióxido de carbono (CO₂) generadas por los inmuebles, a partir de los Certificados de Eficiencia Energética (CEE) presentados ante la Generalitat Valenciana. Estos certificados permiten calcular las emisiones anuales de CO₂ por metro cuadrado construido en función del uso del edificio, su antigüedad y su eficiencia energética. A esta información se han añadido los datos catastrales asociados a cada referencia, lo que ha permitido obtener variables clave como el uso detallado del inmueble, su superficie y el año de construcción. Dicha variable se ha transformado en antigüedad calculando la diferencia respecto al año 2025. Por otro lado, las coordenadas geográficas se han obtenido mediante mapas vectoriales del catastro, facilitando así su integración con los datos espaciales de contaminación. Este objetivo busca conectar los datos energéticos del parque edificatorio con las mediciones reales de contaminación atmosférica, para identificar zonas críticas, patrones horarios y posibles relaciones entre las variables.

Análisis Exploratorio de Datos:

Como primer paso, se llevó a cabo un análisis exploratorio de la distribución de las emisiones de CO₂ en la ciudad, eliminando datos anómalos tras su validación en Google Maps y Catastro para verificar si sus emisiones son correctas. Posteriormente se elaboró un mapa interactivo con información de emisiones y certificados energéticos presentados en cada código postal (Ver 4.1 - Mapa de calor por CCPP) y se identificaron zonas con **valores especialmente elevados, como el centro histórico y el barrio del Cabanyal, ambos caracterizados por su antigüedad y baja eficiencia energética**. También destaca el municipio de Paterna, donde se han detectado usos no residenciales, principalmente oficinas e industrias con emisiones superiores a los 40 kg CO₂/m² anuales.

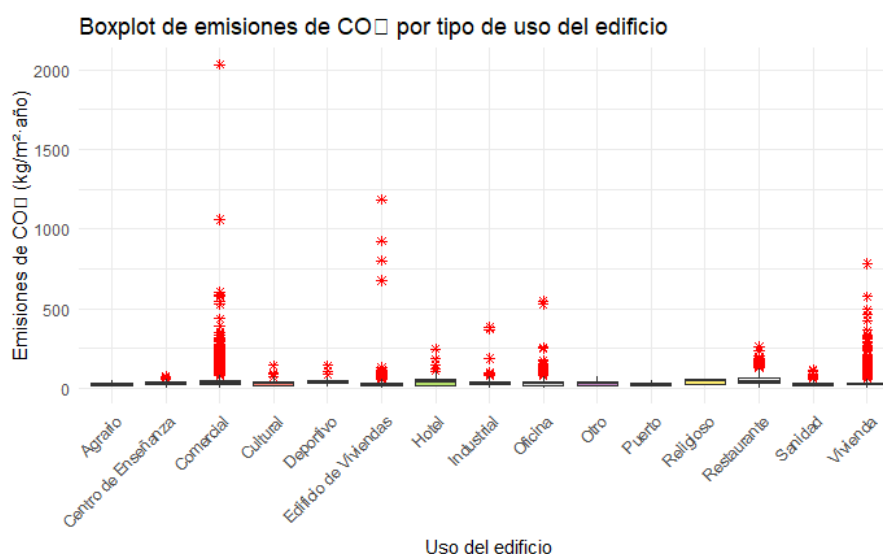


Figura 3.2: Box Plots de emisiones de CO₂ por usos del edificio

Clasificación del comportamiento de emisiones por referencias catastrales:

Los hallazgos anteriores reflejan que no existe un patrón homogéneo en la distribución de emisiones, hecho que se pone de manifiesto en los mapas incluidos en los anexos. Para profundizar en esta heterogeneidad, se aplicaron técnicas de clustering sobre una muestra estratificada por tipo de uso, utilizando como variables las emisiones de CO₂, la superficie y la antigüedad.

Como resultado del agrupamiento, se identificaron ocho clústeres de inmuebles con características energéticas diferenciadas. El primero agrupa viviendas y oficinas con emisiones moderadas, representando el comportamiento medio de la ciudad. El segundo se compone de edificios de reciente construcción, como viviendas nuevas y centros educativos, con buena eficiencia energética. El tercero incluye industrias, centros deportivos, colegios y hoteles antiguos, todos con emisiones superiores a la media. El cuarto clúster está formado por viviendas envejecidas con alta huella de carbono, mientras que el quinto recoge edificaciones antiguas reformadas o sin sistemas de climatización, que presentan bajas emisiones. El sexto agrupa los edificios con las emisiones más elevadas de toda la muestra. El séptimo corresponde a viviendas y hoteles relativamente recientes, pero con un bajo rendimiento energético. Por último, el octavo clúster está dominado por grandes superficies comerciales. El clúster 1 fue el más representativo, agrupando aproximadamente el 40 % de las referencias. (Ver F 4.1 - Clustering de las referencias catastrales)

Clasificación del comportamiento de emisiones por parcelas catastrales:

En una segunda fase, se realizó un nuevo agrupamiento a nivel de parcela, por lo que se pudo usar los datos en su totalidad, tras el agrupamiento se obtuvo unas treinta y mil parcelas, por lo que era viable obtener una visión más general del comportamiento de esta manera. Esta clasificación dio lugar a siete clústeres principales:

1. Grandes edificios recientes con eficiencia media.
2. Edificios antiguos con alta ineficiencia.
3. Edificios nuevos y diversos con alta eficiencia energética.
4. Edificios con poca antigüedad con eficiencia media.
5. Edificios muy antiguos con eficiencia moderada.
6. Edificios reformados con buen rendimiento energético.
7. Casos extremos de ineficiencia energética.

Casi 50 % de las parcelas se encuentran agrupadas en el cluster 4, dándonos a entender que lo normal son 30.59337kg CO₂/m² F 4.2 - Clustering de las parcelas catastrales)

Relaciones con la calidad de aire

Para relacionar estos resultados con la calidad del aire, se asignó cada edificio a su estación de medición más cercana mediante geometrías de Voronoi. Esta asignación permitió representar espacialmente las zonas asociadas a cada estación, identificando **áreas especialmente críticas como la Pista de Silla y Molí del Sol, que destacan los edificios no muy antiguos con eficiencia media, y antiguos con eficiencia moderada**. La zona de la estación de Quart de Poblet tiene más casos de edificios antiguos con alta ineficiencia.

Con el objetivo de explorar posibles vínculos entre las emisiones de los edificios y la concentración de contaminantes, se compararon los niveles estimados de CO₂ con los datos reales de contaminación atmosférica correspondientes al año 2023, seleccionado por ser el último sin anomalías relevantes. **A través del coeficiente de correlación de Pearson, se observó una leve relación positiva entre las emisiones de CO₂ y los niveles de ozono troposférico (O₃) en determinadas zonas** F 4.3 - Gráfico de Correlaciones de CO₂ Y Contaminantes), aunque no se detectaron patrones claros con otros contaminantes, posiblemente debido a la falta de información sobre emisiones industriales. Finalmente, se analizó el comportamiento horario medio de los contaminantes desde el año 2019, centrándose en los meses de verano e invierno, que son los periodos de mayor consumo energético así como de emisiones de CO₂ por parte de las viviendas . En invierno se detectaron picos de concentración entre las 7:00 y las 9:00 horas (uso de agua caliente) y entre las 17:00 y las 22:00 horas (uso de calefacción), mientras que en verano las concentraciones tienden a elevarse entre las 12:00 y las 19:00 horas, coincidiendo con el uso intensivo del aire acondicionado. En particular, se ha constatado que muchos contaminantes aumentan significativamente en esas franjas horarias, especialmente en invierno, lo cual sugiere que las dinámicas de consumo energético podrían estar influyendo en los niveles de contaminación, debido a las emisiones generadas , aunque no de manera exclusiva. F 4.1 - Gráfico de líneas concentración contaminantes)

3.5 Objetivo 5: Patrones entre contaminantes y factores meteorológicos, y relación entre contaminantes y zonas verdes

Para sacar conclusiones de forma abstracta sobre el Objetivo 2: Identificación de patrones meteorológicos y su relación con los niveles de contaminación atmosférica y poder cuantificarlas, se han aplicado reglas de asociación con el fin de ver los patrones más repetitivos que se dan en los factores meteorológicos cuando la concentración de los contaminantes es buena y mala.

Para ello se ha creado un dataframe en el que cada fila representa la hora de un día concreto de un año, y cada columna representa un intervalo de valores para cada variable, tanto meteorológica como contaminante (véase el dataframe en el *apéndice del objetivo 5*).

Las reglas de asociación (véase Anexo G.1.2) se han creado seleccionando como antecedente los factores meteorológicos y como consecuente los contaminantes. Además, se han seleccionado aquellas con un soporte mínimo de 0.05 para evitar encontrar patrones que no se han dado en un número suficiente de horas; una confianza mínima de 0.8 para detectar que los patrones se cumplen en la mayoría de ocasiones; y un lift mayor que 1.1 para evitar patrones que en realidad sean independientes entre el antecesor y el consecuente.

En primer lugar se han filtrado aquellas en las que la concentración de los contaminantes era buena, y se ha observado cada contaminante uno por uno. Los resultados obtenidos son los siguientes:

- $PM_{2.5}$: el 80 % de las horas en las que el viento viene de Ponente y las rachas de viento superaban los 30 km/h, los valores establecidos por la OMS para este contaminante son buenos. Para este contaminante también cabe resaltar que para cuando los valores se consideran de calidad moderada, se observa el 80 % de las veces que el las ráfagas de viento son inferiores a 15 km/h y su velocidad inferior a 10 km/h, así como una temperatura entre 20 y 30 °C.
- NO_x : el 95 % de las horas en las que la humedad está en torno al 30-60 % y hace entre 20 y 30°C, o el viento viene de Levante y su velocidad es superior a 20 km/h, los valores establecidos por la OMS para este contaminante son buenos.
- O_3 : el 85 % de las horas en las que la dirección del viento viene de Ponente, su velocidad es inferior a 10 km/h y la humedad es alta (entre 60 y 90 %), los valores establecidos por la OMS para este contaminante son buenos.

Para el resto de contaminantes no se han detectado patrones para los parámetros establecidos al principio, lo cual indica que no podemos obtener conclusiones robustas sobre ellos.

Después se han filtrado las reglas en las que la concentración de los contaminantes era mala, pero no peligrosa, puesto que en Valencia no se han registrado valores peligrosos durante el 2023. Los resultados obtenidos son los siguientes:

- O_3 : el 80 % de las ocasiones en las que el viento viene de Levante, los valores establecidos por la OMS para este contaminante son muy insalubres.

Para el resto de valores, pese a bajar el valor del soporte a 0.001, no se han encontrado patrones, ya que no hay una muestra suficiente como para establecer patrones que se den con frecuencia. Se ha establecido ese valor de soporte porque es el mínimo para tener una muestra mínimamente significativa (al menos 10 horas).

En cuanto a la relación entre las zonas verdes con los contaminantes, se ha dibujado por un lado un mapa que mide, para cada distrito municipal de Valencia, la suma total de los m^2 de sus zonas verdes. Por otro lado, se ha dibujado un mapa para cada contaminante que mide la concentración media de ese contaminante en cada distrito.

Acompañado de los mapas, se ha estudiado la correlación entre el área de zonas verdes de cada distrito y su concentración de cada contaminante usando la correlación de pearson (véase Figura G.1).

Los resultados obtenidos (véase Anexo G.1.3) indican que hay una correlación positiva con los PMs y los NOs , y una negativa con el SO_2 . Además, no se ha detectado ninguna correlación con el O_3 . No obstante, al tener una muestra para los contaminantes de tan solo 3 distritos, se ha realizado un test t para comprobar si la muestra era significativa y, por tanto, las conclusiones válidas. Los resultados obtenidos en ese test indican que la muestra es pequeña como para afirmar la correlación.

CAPÍTULO 4

Lecciones para futuros proyectos

Aclaración. Reflexiones aplicables a vuestros futuros proyectos, ya bien sea sobre el planteamiento de objetivos, la discriminación de fuentes de datos, sobre el uso de técnicas, sobre el uso de R y/u otros lenguajes, sobre la presentación de resultados o el análisis descriptivo. Haced un planteamiento común entre todos los componentes del equipo.

Deberéis enumerar y valorar también todas las herramientas software utilizadas por vuestro equipo: desde R studio, como repositorios, aplicaciones, chats, etc.

Las principales lecciones que hemos aprendido para futuros proyectos han sido:

1. **Valorar la viabilidad y limitaciones de los datos disponibles para el desarrollo de los objetivos.** En este proyecto hemos sido muy ambiciosos y no hemos tenido en cuenta las limitaciones de los datos, lo cual nos ha afectado en las horas extras que hemos echado en la última parte del proyecto, al tener que ir modificando algunos aspectos.
2. **Organizar el repositorio de trabajo para agilizar el flujo de trabajo, tanto personal como entre los miembros del equipo.** Nos hemos dado cuenta de que usar una herramienta de control de versiones como Git, hace mucho más rápido y cómodo el trabajo.
3. **Entender profundamente el contexto antes de empezar a trabajar con los datos.** Esto es fundamental para manejar los datos correctamente, elegir los modelos adecuados y evitar conclusiones erróneas.
4. **Saber qué se está haciendo y porqué en todo momento a la hora de escribir código.** El uso de la inteligencia artificial para programar puede llegar a causar estragos. Todavía comete muchos errores y la falta de conocimiento sobre la estructura de los archivos por parte de la inteligencia artificial que se esté consultando puede incluso llegar a hacer más lento el trabajo.

En cuanto a las herramientas software utilizadas, las más relevantes han sido:

1. **R studio.** Es un buen editor de código y muy bien integrado para R.
2. **Git.** Fundamental para organizar el control de versiones y la colaboración entre los miembros del equipo.
3. **Discord.** Perfecto para llamadas y comunicación del equipo.

APÉNDICE A

Actualización Hito 1 (CORREGIR DE NUEVO)

A.1 Técnicas de obtención de datos y extracción

- Para la organización de los datos hemos creado un **repositorio de Git** en GitHub, donde hemos depositado los datos que hemos obtenido y las técnicas de obtención utilizadas.
- Principalmente, los datos se han sacado de **archivos .csv** procedentes de las páginas web mostradas en el punto anterior. Pese a que se podrían haber descargado en formato JSON o a través de la API de alguna de las páginas, se ha optado por el formato .csv porque es el más fácil de descargar y es fácil de leer a la hora de tratar con los datos.
- Para la obtención de los datos de la [web de la GVA](#), los obtuvimos por **scraping** con un script en Python3, ya que había que descargar una base de datos por año de muchas zonas diferentes de valencia. El scraper se encuentra incluido en el repositorio de GitHub.
- Análogamente, realizamos un **scraping** sobre la página de la [AVAMET](#), obteniendo todos los datos de todas las fechas disponibles, pero únicamente en las zonas que nos interesaban. El scraper también se encuentra en el repositorio.
- Además, hemos desarrollado otro **scraper** que extrae la información del tiempo histórico de Valencia desde la página [Tiempo3](#), registrando datos cada hora desde 2020 hasta 2025. A diferencia del scraping de AVAMET, este obtiene información de la ciudad de Valencia con una granularidad menor.
- Para la obtención de los datos de intensidad de tráfico hemos contactado con el ayuntamiento de valencia y empresas relacionadas con el tráfico y la movilidad en València, llevando a cabo procesos burocráticos para finalmente poder obtenerlos, ya que no son completamente públicos.
- La obtención de las emisiones de kg CO₂ de los certificados de eficiencia energéticos se han obtenido de la web [web del IVACE GVA](#), posteriormente al no tener datos suficientes para el análisis se han obtenido **víascraper** una a una la información de catastro de cada una de las referencias, pero al tardar mucho por interacción se recurrió al uso de **multiparalelismo y a proxies con 200 IPs**, posteriormente las que fallaron se volvió a realizar de forma automatizada por la web de Catastro .

- Las coordenadas de las referencias catastrales se han obtenido del shapefile del **mapa vectorial de la provincia de Valencia** [web mapas](#)

A.2 Análisis de las fuentes: interpretación de los datos, valoración de su utilidad en el proyecto

Las fuentes de datos utilizadas han sido evaluadas en función de su calidad, disponibilidad y relevancia para el estudio. A continuación, se realiza una valoración de cada una:

A.2.1. Datos atmosféricos y climatológicos

Se han obtenido datos horarios de distintas estaciones de análisis de contaminantes en Valencia. La información es muy útil para analizar la calidad del aire y cómo varía según la localización y el tiempo.

Los datos de contaminación contienen información horaria de diferentes zonas de Valencia, de Valencia ciudad, así como pueblos del exterior, en total 27 estaciones. Cada una tiene un rango de años distintos, algunas van de 2016 a 2025, y otras cuentan con datos desde 2004. Las principales variables con las que contamos son:

- Óxidos de nitrógeno (NO_x) ($\mu g/m^3$)
- Óxido nítrico (NO) ($\mu g/m^3$)
- Dióxido de nitrógeno (NO_2) ($\mu g/m^3$)
- Ozono (O_3) ($\mu g/m^3$)
- Partículas $<2.5 \mu m$ ($PM_{2,5}$) ($\mu g/m^3$)
- Partículas $<10 \mu m$ (PM_{10}) ($\mu g/m^3$)
- Dióxido de azufre (SO_2) ($\mu g/m^3$)

La base de datos extraída de AVAMET proporciona información diaria sobre temperatura, humedad y otras variables meteorológicas. Aunque los datos no son horarios, su cruce con los de contaminación permite evaluar la influencia del clima en la calidad del aire en diferentes zonas de Valencia, y un gran rango de fechas, desde 2004 hasta la actualidad. Las variables son:

- Fecha (día-mes-año)
- Temperatura mínima ($^{\circ}C$)
- Temperatura media ($^{\circ}C$)
- Temperatura máxima ($^{\circ}C$)
- Humedad (HR %)
- Dirección del viento (Categorica)
- Velocidad media de viento (km/h)
- Velocidad máxima de viento (km/h)

- Precipitaciones (mm)

También disponemos de las coordenadas, y municipio de las estaciones de control climático y de contaminación que usamos en el estudio, para poder realizar cálculos en base a la localización.

Además, se ha recopilado información meteorológica horaria desde 2020 hasta 2025 de la página [Tiempo3](#), pero tan solo de la ciudad de Valencia. A diferencia de AVAMET, esta fuente permite analizar la evolución el clima en intervalos más pequeños, aunque perdiendo precisión geográfica, lo que resulta clave para estudiar como factores como la temperatura, la humedad o la velocidad del viento afectan a la calidad del arie en distintos momentos del día. Las variables que se han mantenido son las siguientes:

- Fecha (día-mes-año)
- Hora (0 a 23)
- Temperatura (°C)
- Precipitaciones (mm)
- Probabilidad de nevada (%)
- Humedad (HR %)
- Velocidad del viento (Km/h)
- Ráfagas de viento (Km/h)
- Ángulo del viento (°)
- Dirección del viento (Categórica)
- Nubosidad (%)
- Visibilidad (Km)

Estas variables permiten estudiar la evolución del clima con granularidad horaria y su impacto en las partículas en suspensión.

A.2.2. Intensidad del tráfico

Gracias a la empresa ETRA disponemos de los datos cada 5 minutos durante toto 2023 de la información que miden 1129 espiras de tráfico:

- IDPM (Identificador)
- Nombre
- Fecha (minuto-hora-día-mes-año)
- Intensidad horaria (vehículos / hora)
- Ocupacion (%)
- Velocidad (km/h)
- FiabilidadVelocidad (%)

- FiabilidadIntensidad (%)
- FiabilidadOcupacion (%)
- Latitud (coordenada)
- Longitud (coordenada)
- Agrupaciones horarias de las anteriores

A.2.3. Certificados Eficiencia Energética

Se han eliminado las variables referidas a la energía y solo nos hemos quedado con los identificadores y emisiones de CO₂:

- ANYO_CONSTRUCCION (2025)
- USO_EDIFICIO (HOSPITAL)
- DIRECCIÓN (MANZANA 73 AVENIDA DE FRANCIA, 46023, VALENCIA)
- MUNICIPIO (VALENCIA)
- CÓDIGO_POSTAL (46026)
- REFERENCIA_CATASTRAL (5892401YJ2659D0001UH)
- EMISIONES_CO2_LETRA (A)
- EMISIONES_CO2_VALOR (kg CO₂/m²)

Se han creado:

- SUPERFICIE_M2 (m²)
- lat, long (latitud, longitud)
- PARCELA_CATASTRAL (5892401YJ2659D)

A.2.4. Espacios verdes

Se han eliminado algunas variables que no son de interés para el proyecto que se pretende desarrollar. Estas son: “Número de elementos fitness”, ya que en la fuente de obtención de los datos no se especifica que es y, tras un análisis preliminar, hay muchos datos iguales y más de la mitad son nulos; “Ud. Gestión”, puesto que no aporta información relevante en el contexto del proyecto.

Las variables que se han mantenido son:

- OBJECTID (1234)
- Id Jardín (99123456)
- Nombre (JARDIN CACTUS)
- Barrio (BENIFERRI)
- Área (m²)
- Zona (NORTE, SUR, TURIA)
- Distrito_Municipal (ALGIROS)
- geo_point_2d (latitud, longitud)

APÉNDICE B

Actualización Hito 2 (CORREGIR DE NUEVO)

B.1 Objetivo principal del proyecto ¿qué presenta este estudio?

Entender el comportamiento de la contaminación en la ciudad de Valencia en función del tráfico, meteorología, zonas verdes e inmuebles.

B.2 ¿Para qué y para quién podría ser de utilidad este estudio?

Este estudio puede resultar de gran utilidad para diversas entidades y para la ciudadanía en general.

El Ayuntamiento de Valencia y las administraciones públicas de la Generalitat Valenciana pueden aprovechar los resultados para diseñar estrategias que reduzcan el impacto medioambiental de la ciudad, mejorando la movilidad, la planificación urbana y la gestión de espacios verdes. Además, las empresas encontrarán en este estudio una herramienta valiosa para orientar su transición hacia las cero emisiones netas, adaptándose a las normativas ambientales y fortaleciendo su compromiso con la sostenibilidad.

Por otro lado, el estudio también puede beneficiar a personas con asma y otras afecciones respiratorias, ya que les permitirá comprender mejor las causas de la contaminación y su impacto en la salud, facilitando la toma de decisiones informadas para minimizar su exposición. Finalmente, los ciudadanos de Valencia podrán acceder a información clara sobre la calidad del aire en su entorno, fomentando la concienciación y promoviendo su implicación en iniciativas que contribuyan a mejorar la calidad de vida en la ciudad.

B.3 ¿Por qué es novedoso? ¿Hay estudios similares?

Tras una revisión de estudios relacionados con la contaminación en Valencia, se han encontrado algunos que tratan la contaminación en Valencia:

- [Predicción e interpolación dinámica de los niveles de contaminación atmosférica mediante datos de intensidad de tráfico y dirección del viento.](#) Es un TFM que utiliza datos de intensidad de tráfico y meteorológicos para predecir los niveles de contaminación atmosférica.

- **Análisis de la transición verde de la ciudad de Valencia.** Es un trabajo para una asignatura de una carrera en el que se trabaja con datos de espacios verdes, transporte público y carriles bici, y contaminación atmosférica para hacer un análisis hacia la transición verde de la ciudad.

Ambos estudian la contaminación, pero para el caso del primero, solo tiene en cuenta la intensidad del tráfico y el viento, mientras que el proyecto que planteamos tiene en cuenta todos los factores meteorológicos. Para el segundo, difiere en que sus datos usados para el transporte tiene en cuenta también el tráfico de bicis, mientras que los de nuestro proyecto tiene en cuenta la intensidad del tráfico en general, y además incluye más factores que pueden tener relación con la contaminación como la superficie cultivada, la meteorología y las emisiones CO_2 en $\text{kg CO}_2/\text{m}^2$ año de los inmuebles.

En general, este proyecto puede ser novedoso porque intenta estudiar una posible relación entre diversos campos con la contaminación de Valencia, algo que hasta ahora no hemos podido encontrar. Esto puede ayudar a establecer nuevas relaciones y abrir paso a futuras investigaciones que sigan con esta tendencia.

B.4 Objetivos definitivos del proyecto

1. Explicar la contaminación en la ciudad de Valencia por zonas y contaminantes.
2. Estudiar la influencia de los factores meteorológicos en la contaminación.
3. Caracterizar el tráfico de Valencia y su impacto en la contaminación.
4. Examinar las emisiones de los inmuebles y su relación con la contaminación.
5. Investigar las relaciones y patrones de todos los factores de los objetivos anteriores con respecto a la calidad del aire.

APÉNDICE C

Objetivo 1

C.1 Enlaces a los html

C.1.1. Estudio descriptivo y temporal de los contaminantes

[Enlace al archivo markdown compilado en html](#)

C.1.2. Mapas de calor de los contaminantes

[Enlace al archivo markdown compilado en html](#)

C.1.3. Análisis de interrelaciones entre contaminantes

[Enlace al archivo markdown compilado en html](#)

C.2 Figuras del Objetivo 1

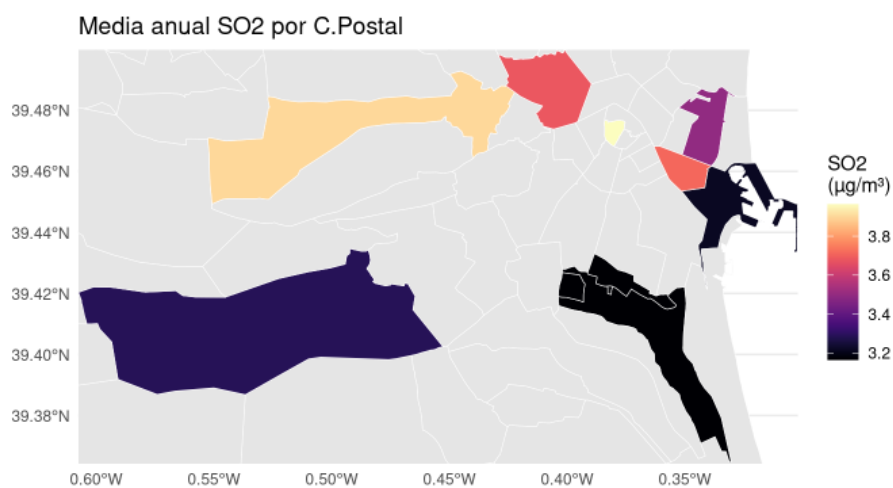


Figura C.1: Ejemplo de maoa de calor

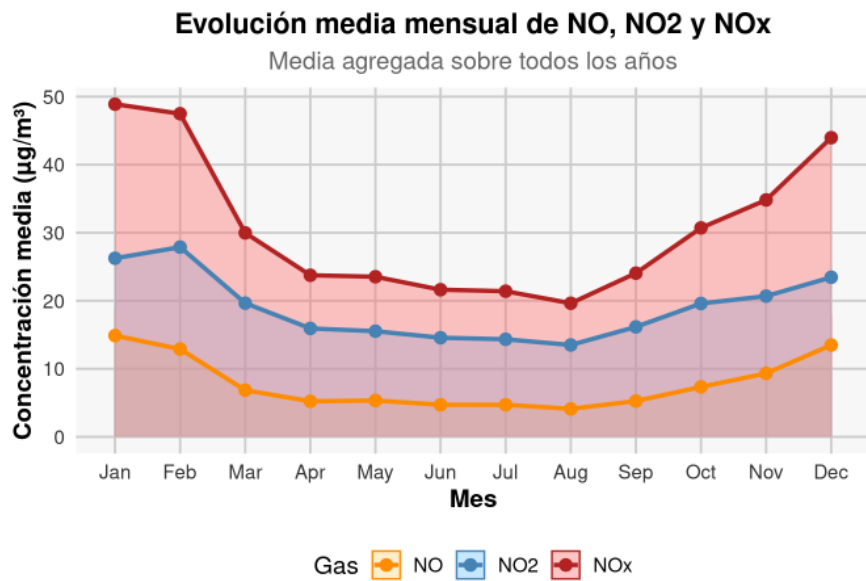


Figura C.2: Ejemplo de evolución mensual

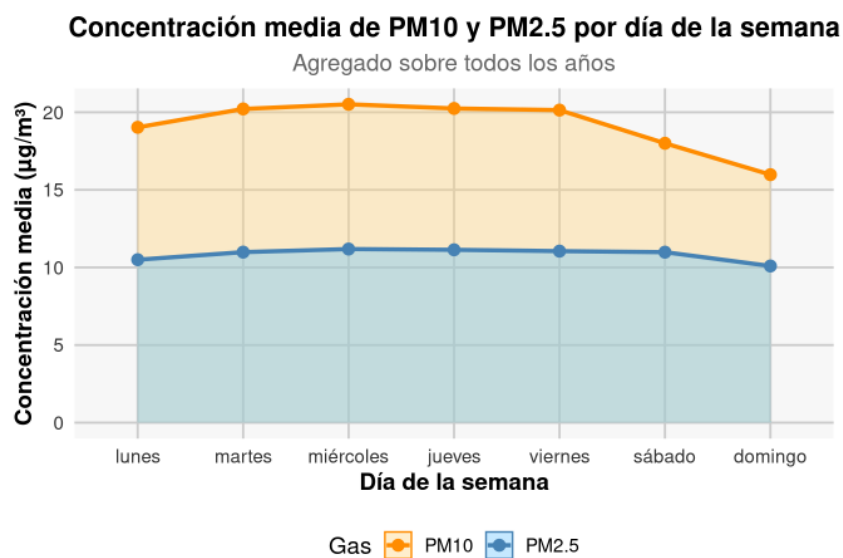


Figura C.3: Ejemplo de evolución semanal

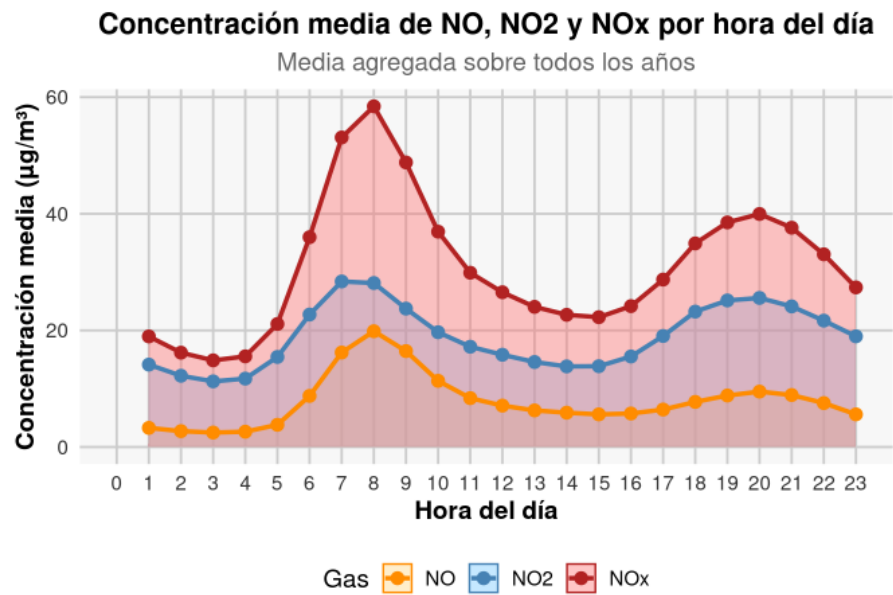


Figura C.4: Ejemplo de evolución horaria

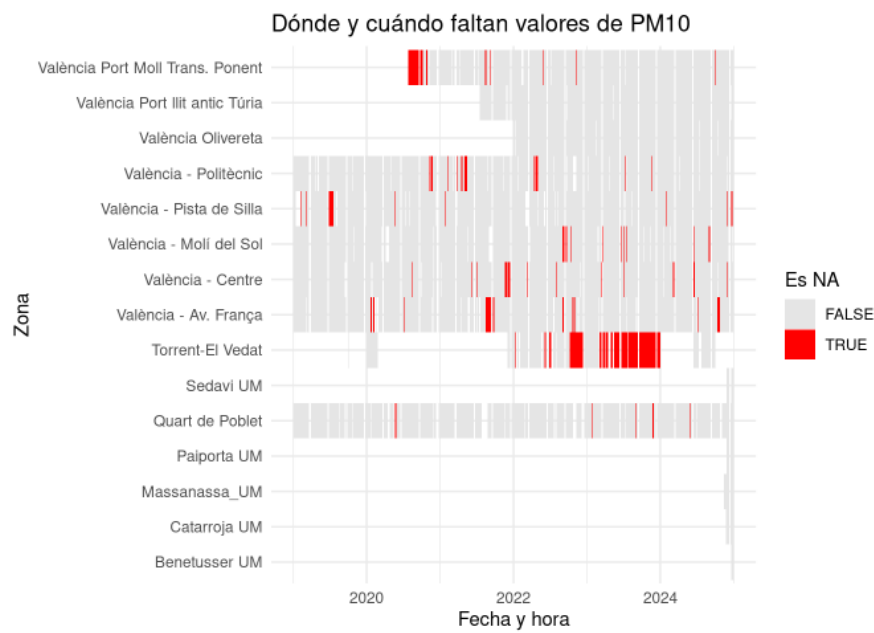


Figura C.5: Ejemplo de faltantes

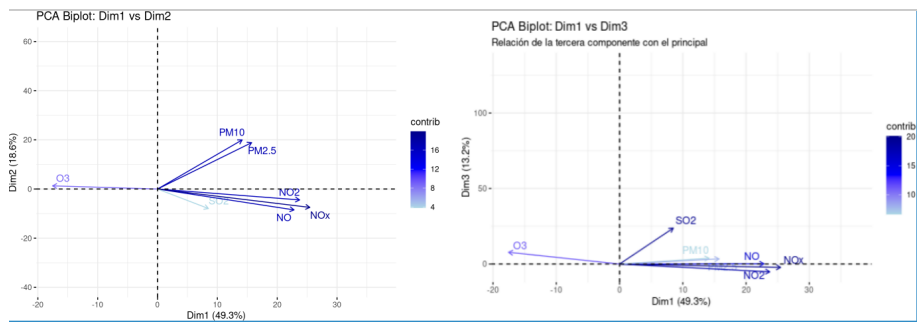


Figura C.6: PCA Contaminantes

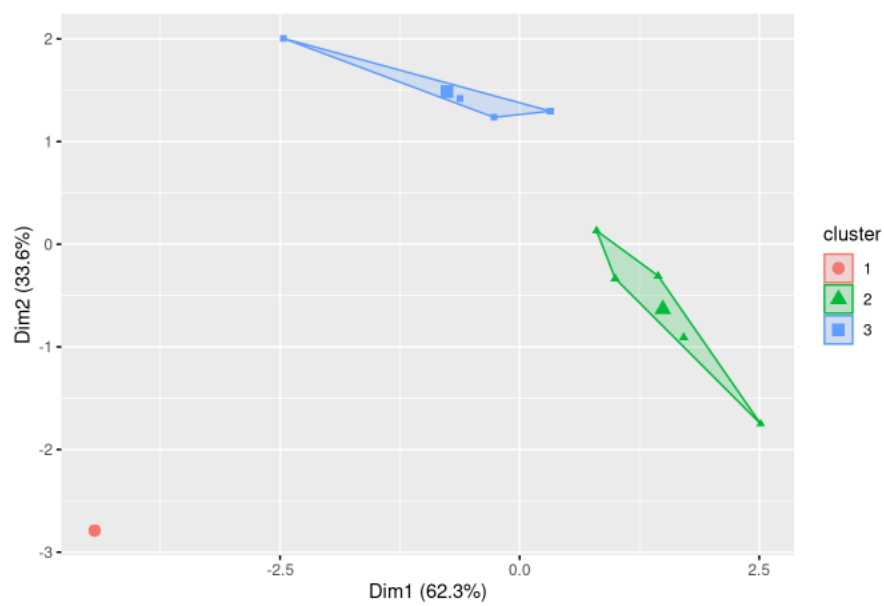


Figura C.7: Clustering de contaminantes

APÉNDICE D

Objetivo 2

D.1 Enlaces a los códigos

D.1.1. PCA sobre las variables meteorológicas

[Enlace al archivo markdown compilado en html](#)

D.1.2. Clustering sobre las variables meteorológicas

[Enlace al archivo markdown compilado en html](#)

D.1.3. Clustering sobre los contaminantes cruzados con meteorología

[Enlace al archivo markdown compilado en html](#)

D.1.4. PLS sobre las variables meteorológicas y los contaminantes

[Enlace al archivo markdown compilado en html](#)

D.2 Figuras del Objetivo 2

Resumen por cluster (bloque 1)

cluster	Temperatura	Precipitaciones (mm)	Humedad (%)	Velocidad del viento (Km/h)	Ráfaga de viento (Km/h)	Ángulo del viento (°)
1	19.24419	0.0026107	42.20280	25.184525	35.32981	278.8131
2	23.93623	0.0018019	55.08636	12.386492	15.50143	108.3932
3	14.52135	0.0017405	64.30357	8.546835	15.82799	283.1877
4	16.54644	0.0925682	74.02156	9.343450	13.82076	159.3866
5	13.25369	1.0167488	82.72906	13.636946	20.52709	160.4901

cluster	Nubosidad (%)	Visibilidad (Km)	wind_x	wind_y	mes_mas_comun	tiempo_mas_comun	tiempo_reducido
1	18.94325	9.988006	0.1885100	-0.8794244	January	Soleado	Despejado
2	12.95075	9.988254	-0.2053213	0.7486150	July	Soleado	Despejado
3	19.46214	9.996225	0.3067778	-0.8126466	January	Despejado	Despejado
4	62.04456	9.852005	-0.0558691	0.1297451	March	Parcialmente nublado	Nublado
5	90.47783	5.340148	0.2434032	0.0456622	March	Llovizna	Lluvia ligera

Figura D.1: Tabla clustering meteorológico

cluster	temperatura	precipitacion	humedad	vel_viento	rafaga	nubosidad
1	20.30938	0.0430244	58.89436	14.292269	19.90719	28.41344
2	17.90135	0.0379291	63.62856	11.253798	17.69162	28.90961
3	17.74781	0.0176305	65.31171	8.877692	13.63552	27.21086
4	17.21716	0.0228906	61.09205	10.030601	15.30743	25.06096
5	15.16874	0.0110421	61.09110	7.829883	12.85459	23.97930
6	14.28194	0.0025806	58.39355	8.420000	14.29484	25.60000

cluster	visibilidad	wind_x	wind_y	angulo_viento	NO	NO2
1	9.892171	0.3873936	0.8705120	176.0096	2.191304	7.885423
2	9.879876	1.0995285	-3.2732826	219.8283	3.072955	12.607286
3	9.923249	0.7109216	-1.1504478	214.5583	4.220028	19.316643
4	9.888131	-0.2757673	-0.4750077	208.4429	12.160678	37.210133
5	9.949137	-0.2047800	-1.9983021	224.5576	36.842650	51.112146
6	9.998064	0.7053997	-2.6906740	233.6903	117.535484	67.793548

cluster	NOx	O3	PM10	PM2.5	SO2	mes_mas_comun
1	11.00241	72.00308	12.00393	6.751631	3.002190	April
2	17.09697	44.58254	10.60619	6.285519	4.512327	October
3	25.60766	40.14469	33.11236	19.593324	3.359897	February
4	55.72568	33.44166	16.65361	9.319742	3.161082	January
5	107.47032	15.91270	31.51932	18.367495	4.701863	January
6	247.89677	12.10323	39.03548	19.006452	6.574193	January

Figura D.2: Tabla clustering contaminantes cruzado con meteorología

APÉNDICE E

Objetivo 3

E.1 Enlaces a los html

E.1.1. Resumen de los datos y mapa Mapa

[Enlace al archivo markdown compilado en html](#)

E.1.2. Pasar datos a horarios

[Enlace al archivo markdown compilado en html](#)

E.1.3. Distribuciones variables tráfico

[Enlace al archivo markdown compilado en html](#)

E.1.4. Tipos de vías (clustering)

[Enlace al archivo markdown compilado en html](#)

E.1.5. Cruce con Contaminación

[Enlace al archivo markdown compilado en html](#)

E.1.6. Relaciones con Contaminación

[Enlace al archivo markdown compilado en html](#)

E.1.7. Agrupaciones y comportamientos

[Enlace al archivo markdown compilado en html](#)

E.2 Figuras del Objetivo 3

E.2.1. PCA: Tráfico y contaminación sin relación aparente

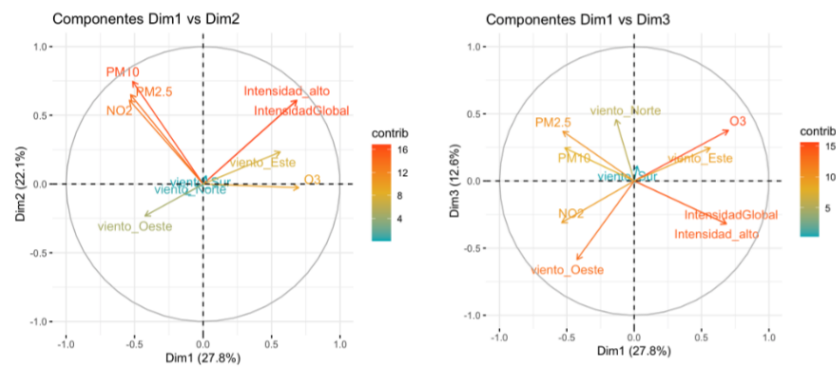


Figura E.1: PCA

E.2.2. Agrupación contaminantes altos, tráfico alto, dirección del viento interesante

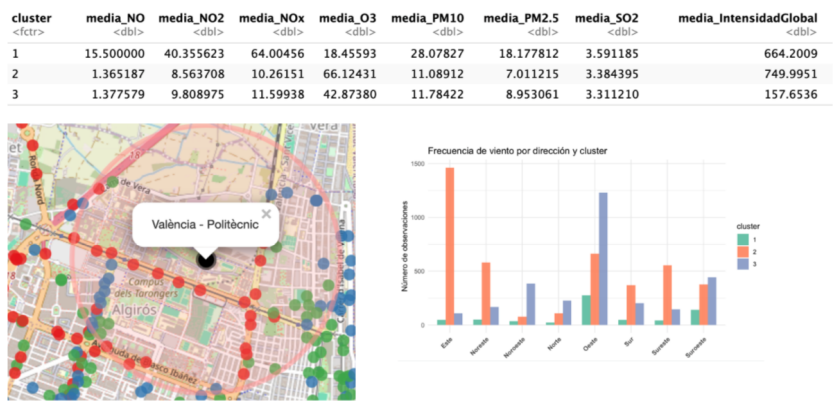


Figura E.2: Agrupación

APÉNDICE F

Objetivo 4

F.1 Enlaces a los html

F.1.1. Rmarkdown compilado en HTML

Puede consultar el código principal en este enlace: [Abrir código compilado](#)

F.1.2. Mapas interactivos

Puedes ejecutar el código desde: [Ejecutar códigos](#)

F.2 Figuras del Objetivo 4

F.2.1. Clustering de las referencias catastrales según sus emisiones y antigüedad

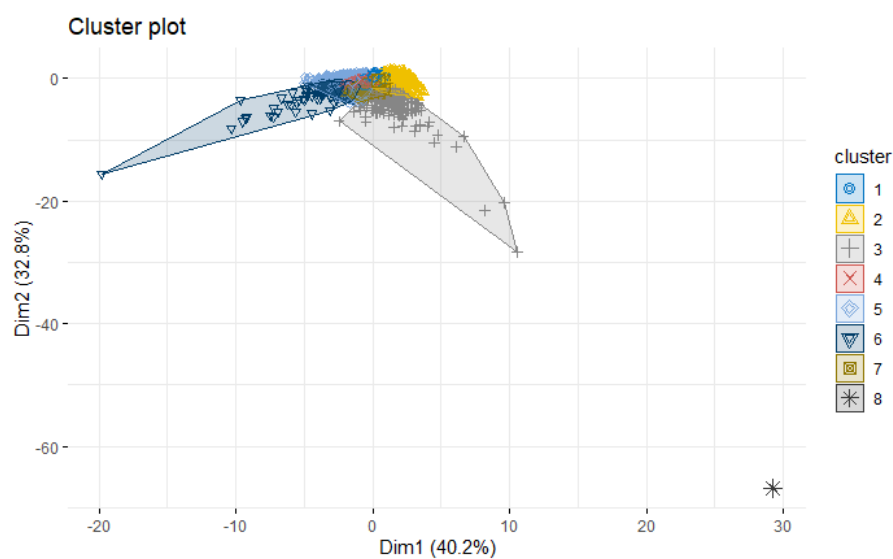


Figura F.1: Clustering de las referencias catastrales

F.2.2. Clustering de las parcelas catastrales según sus emisiones y antigüedad

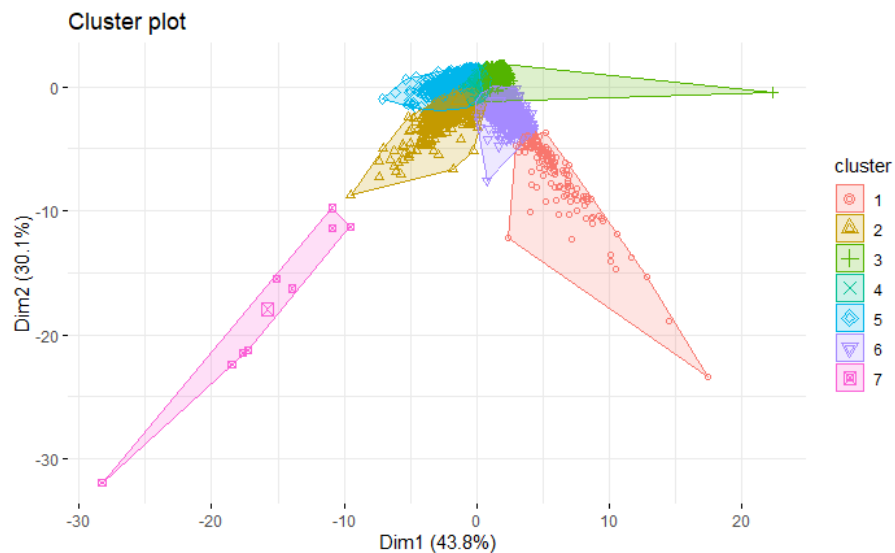


Figura F.2: Clustering de las referencias catastrales

F.2.3. Correlaciones CO₂ y contaminantes

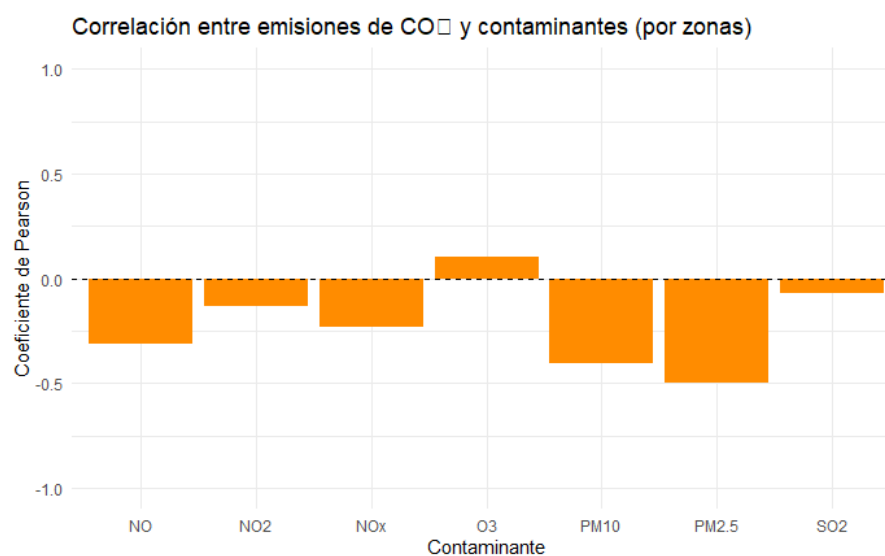
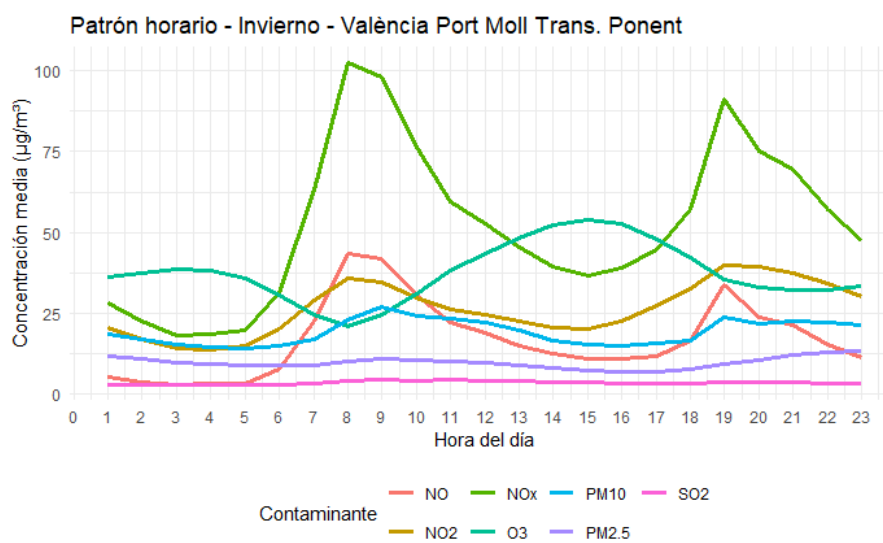


Figura F.3: Correlación entre CO₂ y el resto de contaminantes

F.2.4. Patrón horario invierno de contaminantes**Figura F.4:** Correlación entre CO₂ y el testeo de contaminantes

APÉNDICE G

Objetivo 5

G.1 Enlaces a los html

G.1.1. Creación de dataframe para reglas de asociación

[Enlace al archivo Rmarkdown compilado en html](#)

G.1.2. Reglas de asociación

[Enlace al archivo Rmarkdown compilado en html](#)

G.1.3. Zonas verdes y contaminación

[Enlace al archivo Rmarkdown compilado en html](#)

G.2 Figuras del Objetivo 5

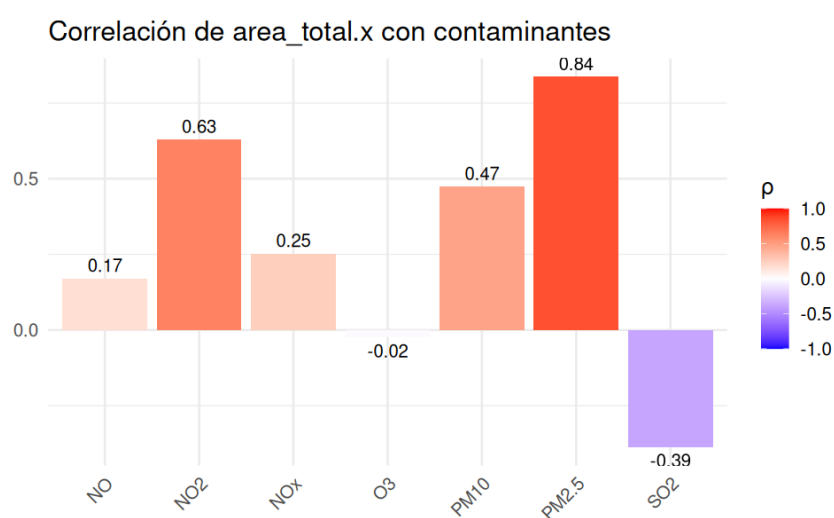


Figura G.1: Correlación entre el área de zona verde y los contaminantes

APÉNDICE H

Información extra sobre los objetos de estudio

H.1 NO_x ($NO + NO_2$)

H.1.1. ¿Qué es?

El NO_x es un término genérico que hace referencia a un grupo de gases muy reactivos (tales como el óxido nítrico (NO) y el dióxido de nitrógeno (NO_2)) que contienen nitrógeno y oxígeno en diversas proporciones.

Es la suma, en partes por mil millones en volumen de monóxido de nitrógeno (NO) y dióxido de nitrógeno (NO_2), expresada como concentración másica de dióxido de nitrógeno en microgramos por metro cúbico, $\mu g/m^3$.

H.1.2. Valores límite establecidos

Los valores de referencia tanto para NO como para NO_2 son los siguientes:

Tiempo promedio	Meta intermedia ($\mu g/m^3$)				
	1	2	3	4	AQG
Anual	40	30	20	–	10
24 horas	120	50	–	–	–

Tabla H.1: Valores recomendados por la OMS

Permitiéndose 3 o 4 días de superación al año.

H.1.3. Causantes

Tienen un origen principalmente antrópico (acción humana).

Como contaminantes, son gases que se emiten en los procesos de combustión que se llevan a cabo en relación con el **tráfico** (sobre todo vehículos automóviles, y en especial de motores diésel) y con el transporte en general, así como en **instalaciones industriales de alta temperatura** y de **generación eléctrica**.

Los focos emisores emiten generalmente NO y NO_2 , a los que denominamos primarios, con el tiempo el NO se oxida y genera NO_2 secundario.

En ambiente urbano, generalmente más del 75 % del NO₂ en aire ambiente es aportado por el tráfico rodado. Esta contribución es mayor que la que aporta al Inventario Nacional de Emisiones debido a que los ciudadanos viven muy próximos al tráfico rodado, y aunque en tonelaje las emisiones son inferiores a las de otras fuentes, su contribución a la exposición humana en ciudades es muy superior.

Los niveles más altos de NO_x se alcanzan en las grandes aglomeraciones urbanas y en sus zonas metropolitanas, así como en el entorno de las vías de comunicación con tráfico más denso.

H.1.4. Efectos sobre la salud humana y los ecosistemas

Los denominados óxidos de nitrógeno engloban tanto al monóxido (NO) como al dióxido de nitrógeno (NO₂). De las dos, es ésta última la principal forma química con efectos adversos sobre la salud; además, el NO se oxida con facilidad, dando lugar a NO₂ rápidamente una vez presente en la atmósfera.

Los principales efectos sobre la **salud** son los siguientes:

- Inflamación de las vías aéreas
- Afecciones de órganos, como hígado o bazo
- Afecciones de sistemas, como el sistema circulatorio o el inmunitario
- Infecciones pulmonares e insuficiencias respiratorias

Los principales efectos sobre los **ecosistemas** son los siguientes:

- Acidificación y eutrofización (exceso de nutrientes en ecosistemas acuáticos)
- Afecciones metabólicas
- Limitación del crecimiento vegetal

Por otra parte, los NO_x actúan como **precursores de la formación de ozono (O₃)** y de otros contaminantes fotoquímicos (por ejemplo, al reaccionar con compuestos orgánicos volátiles, COVs), lo que potencialmente agrava las consecuencias mencionadas sobre la salud y el medio ambiente y conlleva efectos sobre el clima.

H.2 O₃

H.2.1. ¿Qué es?

En la troposfera, el O₃ se forma de manera secundaria a partir de **reacciones químicas** complejas en las que participan principalmente **óxidos de nitrógeno** (NO₂ secundario, junto al emitido como primario) y **compuestos orgánicos volátiles** (COVs, tanto antrópicos como biogénicos procedentes de la vegetación).

H.2.2. Valores límites establecidos

Valores objetivo	Nivel	Período
Valor objetivo para la protección de la salud humana (fecha de cumplimiento: año 20105)	120 µg/m ³	Máxima diaria de las medias móviles octohorarias. No debe superarse en más de 5 ocasiones de promedio en un período de tres años.
Valor objetivo para la protección de la vegetación (fecha de cumplimiento: año 20102)	AOT40 = 18.000 µg/m ³ ·h	Valor acumulado de mayo a julio de promedio en un período de 5 años.

Objetivos a largo plazo	Nivel	Período
Objetivo a largo plazo para la protección de la salud humana	120 µg/m ³	Máxima diaria de las medias móviles octohorarias dentro de un año civil.
Objetivo a largo plazo para la protección de la vegetación	AOT40 = 6.000 µg/m ³ ·h	Valor acumulado de mayo a julio .

Tipos de umbrales	Parámetro	Valor del umbral
Umbral de información	Promedio horario	180 µg/m ³
Umbral de alerta	Promedio horario	240 µg/m ³

H.2.3. Causantes

La velocidad y el grado de formación de O₃ se ven muy incrementados con el **aumento de la radiación solar, las emisiones antropogénicas de precursores y el ciclo biológico de emisiones biogénicas de COVs**. Por ello sus niveles son más elevados en el sur de Europa y en primavera y verano.

Además, sus niveles son superiores en las **periferias de las grandes urbes** y en las **zonas rurales** porque la reacción fotoquímica necesita una cierta distancia para generar O₃ a partir de sus precursores. Una vez formado y en entornos urbanos con altos niveles de NO, el O₃ se consume rápidamente mediante la oxidación de NO a NO₂. Es por ello que en zonas urbanas de tráfico los niveles de O₃ suelen ser muy bajos, mucho más bajos que en entornos poco contaminados, en donde se recibe el O₃ generado durante el transporte de masas de aire desde zonas contaminadas urbanas e industriales, y no existe NO local que lo pueda consumir.

Semanalmente, los niveles de O₃ más elevados se registran los **fin de semana y, sobre todo, el domingo**, en coherencia con la disminución de los niveles de precursores que se registra en tales días (sobre todo NO_x).

H.2.4. Efectos sobre la salud humana y los ecosistemas

El gas ozono (O_3) tiene un **efecto positivo en la estratosfera** (a unos 10-50 km de la superficie terrestre), ya que protege de la radiación ultravioleta. Sin embargo, a cotas inferiores, **en la troposfera** (la capa de la atmósfera en contacto con la tierra), se convierte en un **contaminante** que actúa como un potente y agresivo agente oxidante.

Los principales efectos sobre la **salud** son los siguientes:

- Problemas respiratorios (irritación, inflamación, insuficiencias respiratorias, asma)
- Incremento de la mortalidad prematura

Los principales efectos sobre los **ecosistemas** son los siguientes:

- Problemas de crecimiento de cultivos y bosques
- Reducción de la absorción de CO_2 por las plantas
- Alteración de la estructura de los ecosistemas
- Reducción de la biodiversidad
- Gas de efecto invernadero, que contribuye al calentamiento de la atmósfera

H.3 PM_{2,5} y PM₁₀

H.3.1. ¿Qué es?

Las PM son un contaminante atmosférico formado por partículas sólidas o líquidas en suspensión aerodinámica. Estas partículas incluyen compuestos orgánicos, polvo, hollín y metales.

H.3.2. Valores límite establecidos

Valor legislado–FASE I	Valor límite	Período
Valor límite diario (VLD) de PM ₁₀ para la protección de la salud humana (fecha de cumplimiento: 1 de enero de 2005)	50 µg/m ³	Valor medio en 24 h. No debe superarse en más de 35 ocasiones por año civil
Valor límite anual (VLA) de PM ₁₀ para la protección de la salud humana (fecha de cumplimiento: 1 de enero de 2005)	40 µg/m ³	Año civil

Tabla H.2: Valores legislados para PM₁₀ en aire ambiente

Valor legislado	Valor	Periodo
Valor límite anual para la protección de la salud (fecha de cumplimiento: 1 de enero de 2015).– Fase I	25 µg/m ³	Año natural
Valor límite anual para la protección de la salud (fecha de cumplimiento: 1 de enero de 2020).– Fase II	20 µg/m ³	Año natural

Tabla H.3: Valores legislados para PM_{2,5} en aire ambiente

Contaminante	Tiempo promedio	Meta intermedia				Nivel de las directrices sobre la calidad del aire
		1	2	3	4	
PM _{2,5} µg/m ³	Anual	35	25	15	10	5
PM _{2,5} µg/m ³	24 horas ^a	75	50	37.5	25	15
PM ₁₀ µg/m ³	Anual	70	50	30	20	15
PM ₁₀ µg/m ³	24 horas ^a	150	100	75	50	45

Tabla H.4: Valores recomendados por la OMS

H.3.3. Causantes

Según el Inventario Nacional de Emisiones Contaminantes a la Atmósfera, los focos más influyentes son:

- Combustión en sectores no industriales, tanto para las PM₁₀ como para las PM_{2,5}
- Tráfico rodado (en el caso de las PM₁₀)
- Agricultura (en el de las PM_{2,5})
- Minería
- Industria
- Emisiones domésticas y residenciales

- Polvo mineral de origen africano

Para valorar la situación nacional con respecto a estos contaminantes hay que tener presente que España siempre ha presentado niveles altos de partículas, cuya concentración se incrementa esporádicamente de forma natural por las **intrusiones de polvo africano**.

H.3.4. Efectos sobre la salud humana y los ecosistemas

Son los **contaminantes del aire más importantes** en términos de peligrosidad para la salud humana, ya que pueden ser inhaladas y penetrar así en el sistema respiratorio. Además, las de menor tamaño pueden incluso alcanzar los alveolos pulmonares, lo que les permite de este modo llevar sustancias nocivas a zonas muy sensibles y agravar patologías que pueden conducir incluso a una muerte prematura. De este modo, las partículas (en especial las *PM_{2,5}*) pueden estar implicadas en el **incremento de la mortalidad y de la morbilidad** por causas respiratorias y cardiovasculares.

En líneas generales, tienen los siguientes efectos en el medio ambiente y clima:

- Afección del crecimiento vegetal
- Reducción de la visibilidad
- Influencia en los cambios de temperatura netos
- Alteración de los patrones de precipitación
- Alteración de la relación entre la radiación reflejada y la incidente (albedo superficial)
- Daños en las edificaciones

H.4 SO₂

H.4.1. ¿Qué es?

El dióxido de azufre (SO₂) es un gas incoloro con una olor penetrante y acre. Se genera principalmente en procesos de combustión de combustibles fósiles y en actividades industriales, y es precursor de sulfatos que contribuyen a la formación de material particulado (PM_{2,5} y PM₁₀) y lluvia ácida.

H.4.2. Valores límite establecidos

Categoría	AQI	SO ₂ 1 h (µg/m ³)
Bueno (Good)	0 – 50	0 – 91.7
Moderado (Moderate)	51 – 100	94.3 – 196.5
Insalubre para grupos sensibles (USG)	101 – 150	199.1 – 484.8
Insalubre (Unhealthy)	151 – 200	487.4 – 796.6
Muy insalubre (Very Unhealthy)	201 – 300	799.2 – 1582.7
Peligroso (Hazardous)	301 – 500	1585.3 – 2630.8

Tabla H.5: Índice de calidad del aire (AQI) y rangos de concentración de SO₂ en 1 hora

H.4.3. Causantes

Los causantes antropogénicos son:

- Combustión de combustibles fósiles (centrales eléctricas de carbón y petróleo, refinerías de petróleo y plantas petroquímicas, plantas metalúrgicas, incineradoras de residuos, vehículos y maquinaria pesadas)
- Quemas de biomasa: incendios forestales y agrícolas, estufas y calderas domésticas de leña o carbón.

Los causantes naturales son:

- Erupciones volcánicas es su principal fuente natural
- Emisiones geotermales
- Oxidación de compuestos sulfurados biogénicos

H.4.4. Efectos sobre la salud humana y los ecosistemas

Los principales efectos sobre la **salud humana** son:

- Irritación aguda sobre los ojos, nariz, garganta y vías respiratorias
- Broncoconstricción y agravamiento de asma
- Reducción de la función pulmonar
- Incremento de hospitalizaciones por enfermedades respiratorias y cardiovasculares en poblaciones sensibles

- Mortalidad prematura
- Lesiones oculares y gastrointestinales

Los principales efectos sobre los **ecosistemas** son:

- Lluvia ácida
- Reducción de visibilidad: Dispersa la luz y genera bruma en zonas urbanas
- Corrosión y deterioro de materiales

H.5 Factores climatológicos y su influencia en la calidad del aire

- **Temperatura:** Calor y sol aceleran la fotoquímica y disparan el ozono; noches frías con inversión retienen *NO*, *NO₂*, *CO* y partículas cerca del suelo.
- **Humedad relativa:** Mucha humedad crea nubosidad que frena la fotoquímica pero puede favorecer aerosoles secundarios; aire seco y soleado potencia el ozono.
- **Precipitación:** La lluvia lava rápidamente *PM*, *NO₂* y *SO₂* (el resto son poco solubles). Los periodos secos favorecen la acumulación.
- **Velocidad del viento:** Vientos moderados-fuertes dispersan contaminantes. Calma o brisa débil provoca picos de *NO*, *NO₂*, *CO* y *PM* por estancamiento.
- **Ráfagas:** Aumentan turbulencia y mezcla. Antes de tormentas pueden levantar polvo.
- **Dirección/ángulo del viento:** Determina si llega aire limpio del mar o polvo/contaminación del interior o del Sáhara, y a dónde viajan las emisiones locales.
- **Nubosidad:** Cielo despejado + sol → más ozono. Cielo cubierto reduce la fotoquímica y, con lluvia o viento, limpia el aire.
- **Visibilidad:** Alta visibilidad suele indicar aire limpio. Bruma, calima o smog revelan partículas o humedad con contaminantes.

H.6 Contextualización de Valencia

H.6.1. Fuentes principales de contaminación

- El **tráfico rodado** es la fuente más relevante: las vías de entrada y salida de la ciudad soportan elevadas densidades de vehículos, generando altos niveles de NO_x y partículas.
- El **Puerto de Valencia**, uno de los más activos del Mediterráneo, emite partículas y SO_2 por gases de escape de buques y operaciones de carga/descarga. Está en marcha un plan de “enchufado” de muelles para apagar motores en puerto.
- **Actividad industrial** concentrada en el cinturón metropolitano (Paterna, Sagunto, Almussafes), con emisiones puntuales de NO_x y polvo. En la ciudad quedan básicamente pequeñas industrias y calderas de edificios.

H.6.2. Estructura urbana y dispersión atmosférica

- Valencia ocupa un área **densamente edificada** con escasos espacios verdes más allá del Jardín del Turia, lo que genera isla de calor urbana y favorece la formación de ozono secundario en días soleados.
- El **antiguo cauce del Turia**, convertido en corredor verde, mejora parcialmente la ventilación urbana.
- El **plano costero** y la **ausencia de obstáculos elevados** facilitan la dispersión de contaminantes, a diferencia de otras capitales interiores.

H.6.3. Movilidad y transporte

- Se realizan unos **4,9 millones de desplazamientos diarios** en el área metropolitana, de los cuales el 43 % son a pie. La bicicleta (Valenbisi y 150 km de carriles bici) y el transporte público (Metrovalencia, EMT, MetroBus) han crecido en uso.
- Desde 2021, la **velocidad máxima urbana es de 30 km/h** en la mayor parte de las calles, reduciendo emisiones locales.

H.6.4. Turismo y eventos multitudinarios

- Valencia recibió **2,8 millones de turistas en 2019**. El tráfico de cruceros aporta picos de emisiones en el puerto.
- La quema masiva de monumentos en **Fallas** (marzo) provoca episodios breves de elevada concentración de partículas y gases.
- El **turismo** de sol y playa en verano aumenta temporalmente el parque móvil y la congestión costera.