



# **Análisis de la relación entre la distancia recorrida anualmente por los vehículos y la emisión de contaminantes en países europeos**

## **Autores:**

Carlos Pascual Borrue!  
Souhaib El Harrak  
Shaokai Chen  
Alejandro Ortega de Román  
Julio Agüero Encinas

## **Fecha:**

7 de enero de 2025

Universidad de Cantabria  
Universidad Internacional Menéndez Pelayo

# Contents

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Descripción general del proyecto</b>	<b>3</b>
2.1	Objetivos . . . . .	3
2.2	Resultados esperados . . . . .	3
2.3	Requisitos . . . . .	3
2.4	Requerimientos técnicos . . . . .	3
<b>3</b>	<b>Fuentes de datos</b>	<b>4</b>
3.1	EEA . . . . .	4
3.2	Eurostat . . . . .	4
3.3	Our World in Data . . . . .	5
<b>4</b>	<b>Cobertura</b>	<b>6</b>
<b>5</b>	<b>Gestión del Proyecto</b>	<b>6</b>
5.1	3.1 Planificación . . . . .	6
5.1.1	3.1.1 Paquetes de trabajo. . . . .	6
5.2	3.2 Ejecución . . . . .	6
5.2.1	3.2.1 Plan de Gestión de Datos . . . . .	7
5.3	Diagrama de Gantt . . . . .	8
5.4	Curación de datos . . . . .	8
5.4.1	Datasets NOx . . . . .	9
5.4.2	Dataset tráfico . . . . .	10
5.5	Plan de preservación . . . . .	10
<b>6</b>	<b>Análisis</b>	<b>12</b>
6.1	Año 2020 . . . . .	14
6.2	Regresión . . . . .	14
<b>7</b>	<b>Conclusión</b>	<b>14</b>
<b>8</b>	<b>Posibles mejoras</b>	<b>15</b>
<b>9</b>	<b>Referencias</b>	<b>15</b>

# 1 Introducción

La contaminación del aire es uno de los desafíos globales más importantes de hoy en día, y su resolución enfrenta numerosas dificultades, ya que suele requerir de procesos muy costosos en términos de tiempo y de recursos económicos. Una de las estrategias más efectivas para combatirla es reducir las emisiones en su origen, evitando así que los contaminantes lleguen al ambiente, donde su eliminación implicaría del uso de técnicas avanzadas

Este problema está especialmente presente en las zonas urbanas, donde predominan las emisiones de gases contaminantes provenientes de la combustión de los automóviles. Algunos de los contaminantes más nocivos para la salud humana son los siguientes:

- Los óxidos de nitrógeno (NOx): Mezcla de gases que incluye dióxido de nitrógeno (NO<sub>2</sub>) y óxido de nitrógeno (NO), generada cuando el nitrógeno reacciona a altas temperaturas dentro del motor.
- Las partículas finas (PM<sub>2.5</sub> y PM<sub>10</sub>): Son partículas microscópicas que provienen principalmente de la combustión del combustible.
- El ozono troposférico (O<sub>3</sub>): No es emitido directamente por los automóviles. Se forma como contaminante secundario producto de la reacción de los óxidos de nitrógeno con compuestos orgánicos volátiles (hojas secas, cartones, etc) en presencia de la luz solar.
- El dióxido de carbono (CO<sub>2</sub>): Es un gas de efecto invernadero que, en altas concentraciones, contribuye significativamente al calentamiento global y al cambio climático.

En buena parte de las poblaciones europeas, las concentraciones de estos contaminantes superan los límites de calidad del aire fijados por la propia Unión Europea. Por ello, este trabajo se centra en analizar la relación entre la concentración de contaminantes en el aire y su principal fuente de emisión: los vehículos.

Para llevar a cabo el análisis se han escogido 4 países europeos: Francia, España, Suiza y Países Bajos, y, debido a la gran variedad de contaminantes a analizar, se ha considerado únicamente el más abundante y, a su vez, el más nocivo para la salud, que es el NOx.

## 2 Descripción general del proyecto

### 2.1 Objetivos

El objetivo de este proyecto es realizar un análisis sobre la relación entre la emisión de óxidos de nitrógenos (NOx) y la distancia recorrida anualmente por los vehículos en Francia, España, Suiza, Países Bajos. Este objetivo se puede dividir en las siguientes etapas:

- **Búsqueda de fuentes de datos:** Encontrar fuentes de datos que contengan información acerca la emisión de NOx y la distancia recorrida anualmente.
- **Análisis:** Analizar los datos encontrados, como, por ejemplo, observar cómo ha cambiado la cantidad de NOx emitida a lo largo de los años. También se estudiará la correlación entre la distancia recorrida anualmente y la emisión de NOx, y, si es posible, encontrar un parámetro que relacione estas dos variables.
- **Modelo:** Un modelo capaz de hacer estimaciones futuras de las emisiones NOx en función de la distancia recorrida anualmente por los vehículos.

### 2.2 Resultados esperados

- **Gráficas para el análisis:** Obtener gráficas que permitan visualizar el cambio en la emisión de NOx en función de los años y de la distancia recorrida, que permitan identificar el patrón de correlación entre los datos.
- **Modelo predictivo:** Se espera obtener un modelo predictivo que permita estimar la emisión de NOx conociendo la distancia recorrida anualmente por los vehículos.

### 2.3 Requisitos

- **Existencia de fuentes de datos abiertos:** Los datos que se utilizarán deben ser accesibles públicamente y, preferiblemente, provenir de fuentes oficiales.
- **Datos de alta calidad:** Los datos deben estar completos y actualizados. Se prefiere que los datasets estén curados previamente y sean descargables en formatos fáciles de trabajar (CSV, TSV, Excel) y que estén correctamente metadados.

### 2.4 Requerimientos técnicos

- **Herramientas de software:** Se utilizará R para leer los datos y generar las gráficas para el análisis. Además, se empleará Python para crear los archivos de metadatos en formato XML. En cuanto a la curación de los datos, se usará OpenRefine, una herramienta especializada en la limpieza de datos.



TIME	2016	2017	2018	2019	2020	2021	2022	2023
GEO								
Belgium	:	:	:	:	:	:	:	:
Bulgaria	:	:	:	:	:	:	:	1 181.680
Czechia	52 775	54 558	56 401	56 401	51 989	53 601	53 601	60 339
Denmark	:	:	:	:	:	:	:	:
Germany	:	:	:	:	:	:	:	:
Estonia	10 387.300 (e)	10 811.300 (e)	11 197.100 (e)	11 659.100 (e)	11 251.300	11 568.200	11 696.100	11 978.900
Ireland	46 836	47 795	47 545	47 072	36 237	41 878	47 604	46 410
Spain	239 353	244 661	250 192	252 055	195 687	239 946	254 459	:
France	613 562	638 134.183	639 405.333	647 331.336	529 898.855	580 436.646	628 217.242	618 266.081
Croatia	21 928	22 750	23 440	24 280	19 850	23 460	24 890	26 250
Latvia	:	:	:	:	:	:	:	:
Lithuania	:	:	:	:	12 160	12 579	:	:
Luxembourg	:	:	:	:	:	:	:	:
Hungary	40 451.330	43 015.720 (e)	44 796.460 (e)	46 415.930 (e)	41 012.560	45 846.970	48 677.260	48 489.500
Malta	:	:	:	:	:	:	:	:
Netherlands	137 836	139 865	140 371	140 709	121 675	127 854	134 861	138 462
Austria	85 014	86 854	:	:	:	:	:	:

Figure 2: Dataset sobre la distancia recorrida anualmente por los vehículos en cada país

Se puede observar que el dataset es bastante incompleto, lo que limita la selección de los países de estudio. Más adelante se discutirá este aspecto en la sección de cobertura.

### 3.3 Our World in Data

Para complementar el análisis, se ha utilizado un tercer conjunto de datos relacionado con el porcentaje de vehículos eléctricos vendidos en cada país en diferentes años. El dataset proviene del sitio web Our world in data, un proyecto del Global Change Data Lab dedicado a crear recursos educativos gratuitos y en abierto para facilitar el acceso y la comprensión de los principales desafíos a nivel global.

El dataset se puede descargar en formato CSV junto con su metadata en formato JSON.

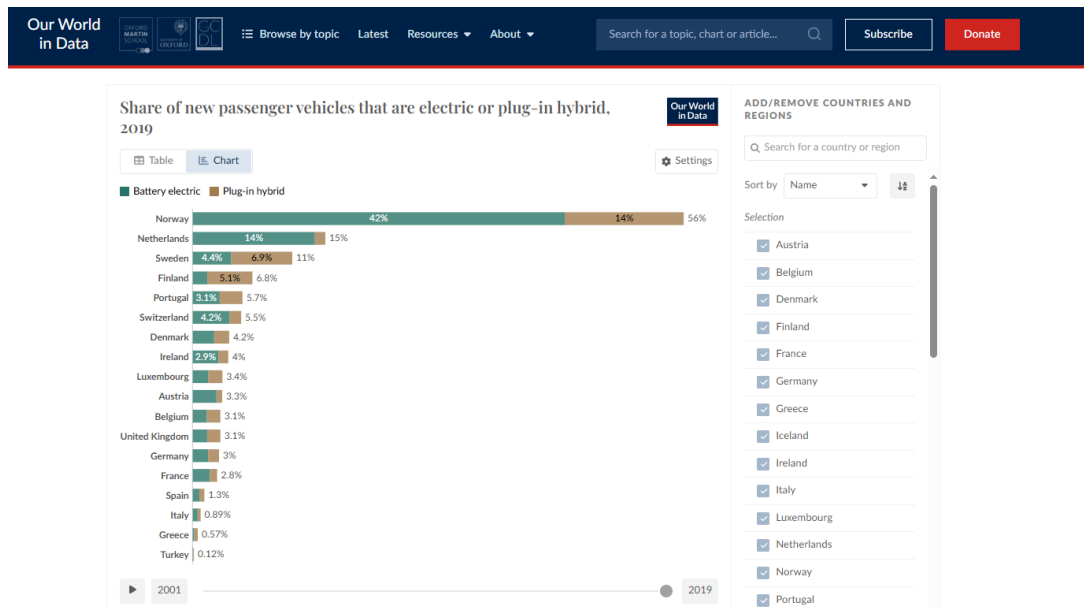


Figure 3: Dataset que contiene el porcentaje de ventas de vehículos eléctricos en diferentes países a lo largo de los años

## 4 Cobertura

En la sección anterior se mencionó que se utilizarán tres conjuntos de datos: uno sobre la emisión de NOx para los países europeos entre los años 2010 y 2022 (aunque fue publicado en 2024, no contiene datos de 2023 ni de 2024); otro sobre la distancia recorrida anualmente, que abarca desde 2014 hasta 2023; y un tercer dataset sobre la cantidad de coches eléctricos vendidos en los países europeos desde 2001 hasta 2019.

Debido a que estos datos abarcan diferentes períodos de tiempo, se ha seleccionado un período de intersección entre todos ellos, que va de 2014 a 2022 (el conjunto de datos sobre vehículos eléctricos solo servirá como apoyo adicional para comprender los resultados obtenidos con los dos primeros conjuntos, por lo que no influye en la selección del período).

En cuanto a la selección de países, debido a que el conjunto de datos de la distancia recorrida es bastante incompleto (faltan datos de muchos países), se han seleccionado finalmente los siguientes países: Francia, España, Suiza y Países Bajos.

- **Cobertura geográfica:** Francia, España, Suiza y Países Bajos
- **Cobertura temporal:** 2014-2022

## 5 Gestión del Proyecto

### 5.1 3.1 Planificación

#### 5.1.1 3.1.1 Paquetes de trabajo.

Un paso importante en el comienzo de un proyecto son los paquetes de trabajo que representan cómo el proyecto se puede dividir de forma estructurada, lógica y detalladamente en todas las distintas tareas que conlleva realizar un proyecto de investigación. De este modo, al dividir las actividades a realizar se puede asignar distintos paquetes de trabajo a cada persona que trabaja en el proyecto. Por lo tanto, cada tarea queda a realización y responsabilidad de una o varias personas concretas garantizando de este modo su completa ejecución.

Los paquetes de trabajo tienen una cronología, es decir, no se puede empezar la parte de análisis o curación de datos sin primero haber escogido un tema o buscado los recursos necesarios para obtener los datos. Por ello, cabe destacar la importancia de que cada paquete de trabajo tiene su tiempo de entrega o finalización para poder avanzar hacia el siguiente punto.

Por otro lado, un diagrama de Gantt es una herramienta visual para planificar y gestionar proyectos mediante barras horizontales que representan tareas y su duración a lo largo del tiempo. A modo de resumen, nos permite plasmar visualmente los paquetes de trabajo, identificar dependencias entre tareas, asignar recursos y hacer seguimiento del progreso. A su vez, facilita la gestión del tiempo al mostrar qué tareas son más importantes para evitar retrasos. Este diagrama planifica, organiza y controla proyectos de manera eficiente.

### 5.2 3.2 Ejecución

En este proyecto se han utilizado los principios FAIR (Findable, Accessible, Interoperable, Reusable) establecidos por la investigación europea en el programa H2020. Además, se ha diseñado un plan de gestión de datos basado en el Data Management Plan que asegura el manejo ético, eficiente y conforme a estándares internacionales de los datos durante y después de un proyecto.

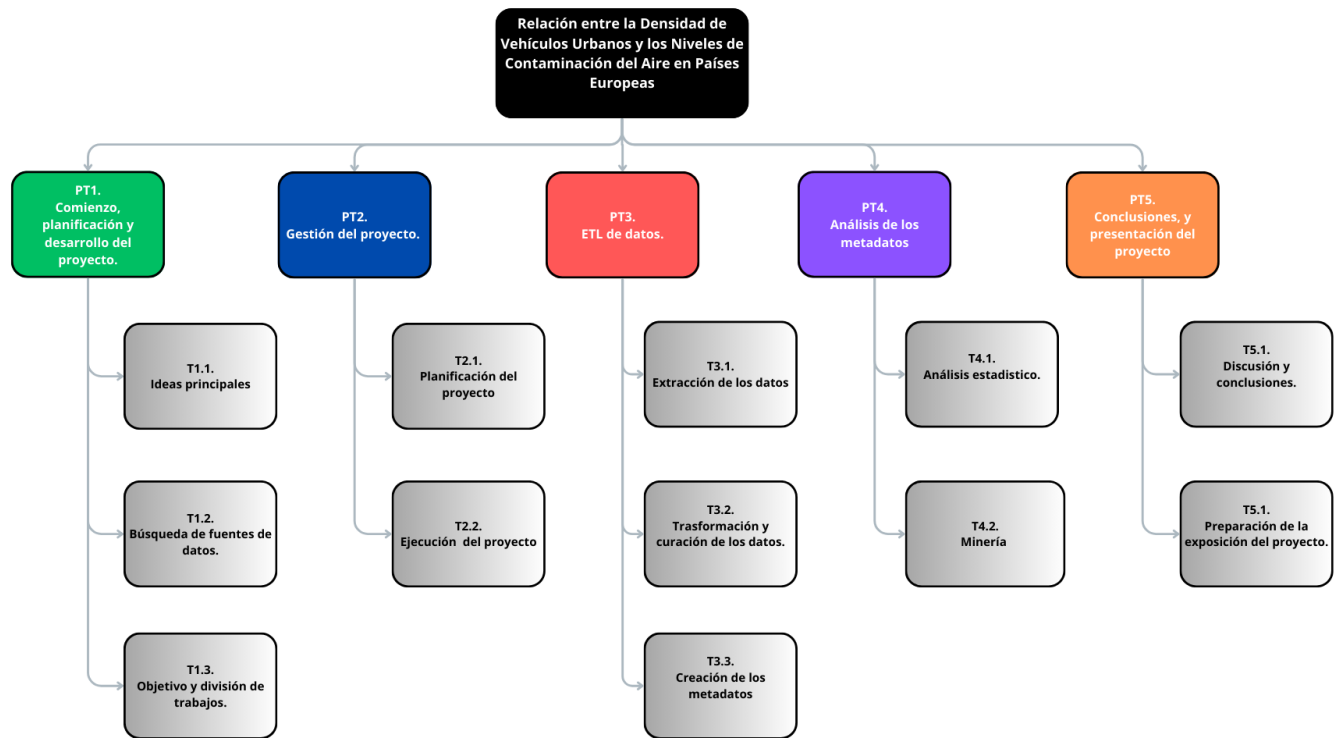


Figure 4: DIAGRAMA DE FLUJO

Los principios FAIR aseguran que los datos sean fácilmente localizables, accesibles para todos los usuarios, compatibles con otros sistemas y reutilizables en otros ámbitos de investigación. Por eso, su uso en este proyecto es indispensable para facilitar colaboraciones científicas y mejorar la transparencia de todo lo utilizado.

El *Checklist for a Data Management Plan* son unas preguntas clave que garantizan que los datos recopilados y procesados sean tratados de forma ordenada, segura y accesible. Este documento abarca aspectos como la recopilación, almacenamiento, descripción, accesibilidad y reutilización de los datos.

Por lo tanto, incluyendo tanto los principios FAIR en investigación europea (H2020) como los requisitos del *Checklist for a Data Management Plan*, se puede usar una gestión efectiva de los datos en el proyecto, lo que puede tener un impacto positivo en los resultados y la calidad de la investigación.

### 5.2.1 3.2.1 Plan de Gestión de Datos

**Resumen:** El objetivo de este estudio es analizar cómo la distancia recorrida anualmente por los vehículos de distintas ciudades europeas afecta los niveles de contaminación del aire. Se recopilaban datos sobre las distancias anuales recorridas, las emisiones contaminantes y del porcentaje de ventas de vehículos eléctricos. La finalidad es determinar patrones de impacto y posibles relaciones de causa-efecto que permitan implementar estrategias de mitigación más efectivas.

#### Metodología:

- **Selección de áreas de estudio:** Se elegirán países representativos de diferentes contextos urbanos en Europa, en concreto Francia, España, Suiza y Países Bajos.
- **Recopilación de datos:**



- **Kilometraje anual:** Datos de los kilómetros recorridos anualmente por los vehículos de distintos países entre los años 2014 y 2022.
- **Contaminación del aire:** Concentraciones de óxidos de nitrógeno NOx provenientes de estaciones de monitoreo ambiental.
- **Herramientas y análisis:** Uso de lenguaje de programación R para procesar la información. Representación de gráficas de dispersión, diagramas de barras y análisis estadísticos de random forest).

**Resultados esperados:** Una relación directa entre la cantidad total de kilómetros recorridos por los vehículos de un país en un año y los niveles de óxidos de nitrógeno registrados. Además, se espera obtener un modelo de predicción capaz de realizar estimaciones futuras.

**Importancia:** La contaminación del aire es uno de los principales problemas ambientales en Europa, siendo el tráfico vehicular una de sus mayores fuentes. Este estudio puede aportar información valiosa para el desarrollo de políticas públicas orientadas a reducir las emisiones contaminantes derivadas del transporte, impulsando alternativas como la mejora de la eficiencia del transporte público y el uso de vehículos eléctricos.

**Documentación y metadatos:** Todos los datos utilizados en este proyecto estarán disponibles en formatos abiertos como CSV, y serán acompañados de metadatos estructurados que describan la fuente de los datos, las variables que contiene y los formatos empleados.

Estos recursos estarán alojados en repositorios públicos para facilitar su consulta y reutilización por parte de otros investigadores.

**Requisitos legales, éticos y códigos de conducta:** Este proyecto utiliza exclusivamente datos públicos y anónimos, por lo que no se requiere una protección adicional para la privacidad o confidencialidad. Todos los datos generados estarán bajo una licencia Creative Commons Attribution (CC-BY 4.0), permitiendo su uso libre con reconocimiento a los autores. Esto fomenta la colaboración y asegura que los hallazgos sean accesibles tanto para la comunidad científica como para el público general. Adicionalmente, se cumplirán las normativas éticas aplicables al uso de datos públicos en el marco europeo, garantizando que el proyecto sea transparente y contribuya a la sostenibilidad ambiental.

### 5.3 Diagrama de Gantt

El diagrama de Gantt es una herramienta de planificación y gestión de proyectos que permite visualizar las tareas necesarias, sus duraciones y la relación entre ellas en una línea de tiempo. En el contexto del proyecto sobre la relación entre la densidad de vehículos urbanos y los niveles de contaminación del aire, este diagrama facilita la organización de las fases del estudio, desde la recolección de datos hasta el análisis final. Al distribuir las actividades de manera clara, el equipo puede coordinar esfuerzos, priorizar tareas y asegurarse de que todas las etapas se realicen en el tiempo previsto, minimizando retrasos y conflictos en el cronograma.

La implementación del diagrama de Gantt en este proyecto permite identificar dependencias entre tareas, gestionar recursos de manera eficiente y ajustar el plan en función de imprevistos o cambios en las condiciones del estudio. Además, su naturaleza visual proporciona una comunicación clara y comprensible entre los miembros del equipo y las partes interesadas, favoreciendo una mayor alineación con los objetivos del proyecto y asegurando que todos trabajen hacia un propósito común de manera organizada.

### 5.4 Curación de datos

Para facilitar el trabajo y asegurar que los datos recolectados puedan ser procesados por las herramientas del análisis, se ha realizado la curación de los datos:

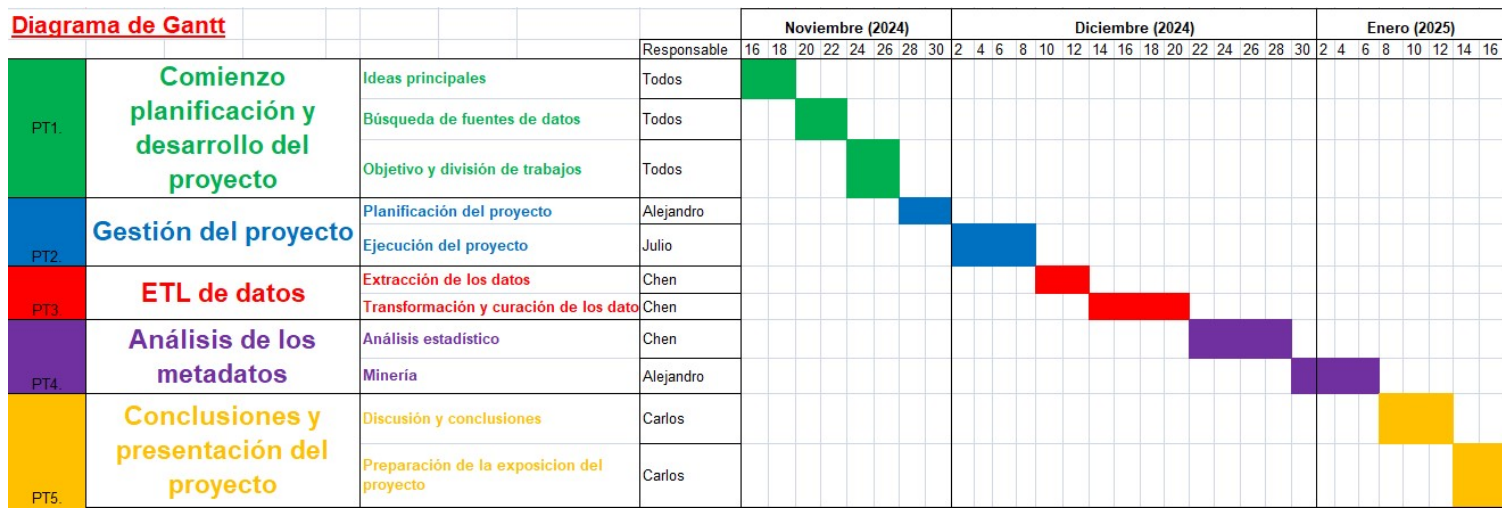


Figure 5: DIAGRAMA DE GANTT

#### 5.4.1 Datasets NOx

Los datos sobre la emisión de los contaminantes pueden ser descargados en formato CSV, y están organizados en diferentes datasets según el tipo de contaminante (en este caso solo se trabaja con la emisión de NOx) y por países (se han descargado datasets para Francia, España, Países Bajos y Suiza). Cada dataset contiene datos de NOx entre los años 2010 y 2022. Para poder llevar a cabo la curación de datos, primero es necesario entender qué información contiene cada columna, por lo que se realiza una descripción breve del dataset:

##### Descripción del dataset:

Tomando como ejemplo la emisión de NOx en Francia, los datos se organizan en las siguientes columnas: Pollutant, Country, Pollutant (columna repetida), Unit, Sector (sector de origen del contaminante), Sub Sector, Country (columna repetida), Column (descripción de las variables) y years.

Pollutant	Country Name	Pollutant_Source	Unit	EEA Sector	EEA Sub-Sector	Country Name 2	Column	2010	2011	2012	2013	2014	2015
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	EEAActivity Percentage	1	1	1	1	1	1
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	EEASector Percentage	23	23	23	23	24	24
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	EEASubSector Percentage	10	11	11	11	12	12
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	Emissions	125	128	127	123	127	129
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	Fixed CountryPollYear	1,236	1,179	1,153	1,134	1,058	1,035
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	Fixed EEAActivity	125	128	127	123	127	129
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	Fixed EEA Sector	279	277	265	257	256	247
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Crops	France	Fixed EEA SubSector	125	128	127	123	127	129
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Energy use	France	EEAActivity Percentage	1	1	1	1	1	1
NOx	France	NOx	Gg (1000 tonnes)	Agriculture	Energy use	France	EEASector Percentage	23	23	23	23	24	24

Figure 6: Dataset sobre la contaminación de NOx en Francia

El propio dataset ya está curado, pero aun así contiene datos irrelevantes para el análisis, se realiza una curación de los datos mediante una herramienta vista en clase, Openrefine, siguiendo los siguientes procesos:

- Eliminar las columnas repetidas.
- Crear un text/numeric facet para las columnas principales, para verificar si existe algún error en los datos (Se ha observado que existen filas con “blanks”, que deben ser eliminadas).
- Filtrar los datos de manera que solo queden los datos de “Emisiones” y de “Nivel promedio anual”.

- Extraer los datos en un nuevo dataset.

Se ha repetido estos procesos para el dataset de cada país.

## 5.4.2 Dataset tráfico

### Descripción del dataset:

En cuanto a los datos la distancia recorrida, el dataset disponible se encuentra en un formato relativamente raro. Por un lado, los datos de la unidad, tipo de vehículo y país están separados por comas, mientras que los datos correspondientes al VKM anual (Vehículos Kilómetros por Millón, que representa la distancia recorrida por los vehículos en un año) están separados por tabulaciones. Además, utiliza el carácter “.” para rellenar las variables que están vacías, e incluye una letra “e” detrás de algunas variables para indicar que son “estimaciones”.

▼ freq,regisveh,unit,vehicle,geoTIME_PERIOD	▼ 2014	▼ 2015	▼ 2016	▼ 2017	▼ 2018	▼ 2019	▼ 2020	▼ 2021	▼ 2022	▼ 2023
A,TERNAT_REG,MIO_VKM,TOTAL,AT <a href="#">edit</a>	80604	82622	85014	86854	:	:	:	:	:	:
A,TERNAT_REG,MIO_VKM,TOTAL,BE	98800	99910	:	:	:	:	:	:	:	:
A,TERNAT_REG,MIO_VKM,TOTAL,BG	:	:	:	:	:	:	:	:	:	1181.600
A,TERNAT_REG,MIO_VKM,TOTAL,CH	64163	65298	66601	67765	68556	69203	57723	60999	64727	66430
A,TERNAT_REG,MIO_VKM,TOTAL,CZ	48335	50934	52775	54558	56401	56401	51989	53601	53601	60339
A,TERNAT_REG,MIO_VKM,TOTAL,EE	9471.600 e	9976 e	10387.300 e	10811.300 e	11197.100 e	11659.100 e	11251.300	11568.200	11696.100	11978.900
A,TERNAT_REG,MIO_VKM,TOTAL,ES	222689	230840	239353	244661	250192	252055	195687	239946	254459	:

Figure 7: Dataset sobre el tráfico

Para limpiar los datos, se han seguido los siguientes pasos:

- Leer el archivo utilizando el tabulador como separador (Al leerlo así, los datos separados por comas formarán una única columna).
- Separar la columna en unit/vehicle/country.
- Usar una expresión en GREL (value.replace(“e”, “”)) para eliminar las letras “e”.
- Extraer los datos de Francia, España, Suiza y Países Bajos en un nuevo dataset.

## 5.5 Plan de preservación

Para garantizar la reutilización de los resultados obtenidos a largo plazo, se utilizarán dos plataformas de repositorio en línea para almacenar los datos:

- Github: es una plataforma de desarrollo colaborativo basada en la nube que permite a los desarrolladores gestionar y compartir proyectos de software. Aunque no se trata de un repositorio común para almacenar los datos que se generarán en este trabajo (principalmente datasets y gráficas), se utilizará para compartirlos entre los miembros del grupo y preservar los resultados finales. A diferencia de los repositorios estándar, Github no asigna los identificadores persistentes a los objetos almacenados, solo proporciona una URL a los archivos digitales, por lo que puede haber problema como la ruptura de enlaces o cambios en la URL. (URL Github)
- Zenodo: es un repositorio digital de acceso abierto utilizado principalmente por investigadores científicos para publicar y preservar datos. Una de las principales diferencias entre Zenodo y Github es que esta primera asigna un DOI a los objetos almacenados, el cual se trata de un identificador persistente que

asegura la identificación de los objetos a largo plazo, y permite que otras personas los localicen.(Al final se decidió no utilizar los recursos online de Zenodo, se ha empleado únicamente Github para la preservación)

## 6 Análisis

Para analizar la correlación entre los indicadores, se han construido una serie de gráficas para cada país de estudio, que incluyen: una gráfica que representa la emisión de NOx en función de los años y otra sobre la distancia recorrida en función de los años. Para no ocupar demasiadas páginas, en este informe solo se mostrarán las gráficas correspondientes a Francia (Las demás se almacenarán en repositorios mencionados).

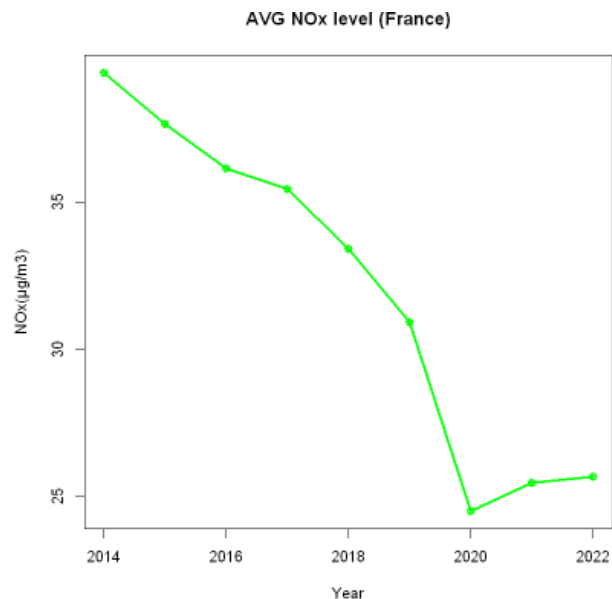


Figure 8: El nivel de NOx (Francia)

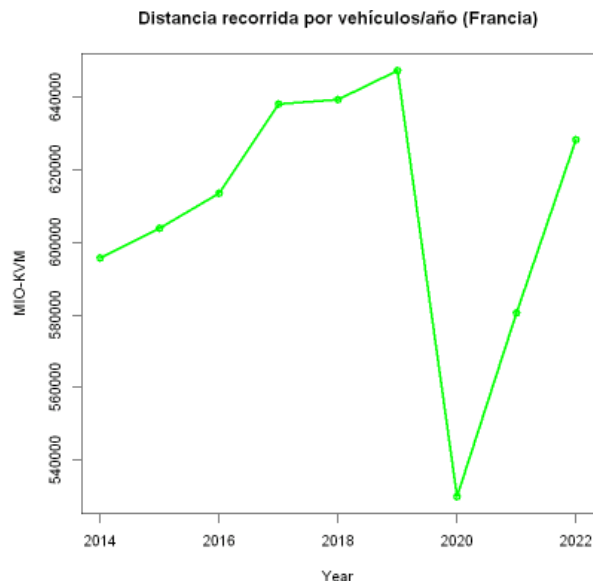


Figure 9: La distancia recorrida anualmente por los vehículos (Francia)

Las gráficas obtenidas presentan resultados algo inesperados. Lo esperado sería que la concentración de NOx aumentara con la distancia recorrida. Sin embargo, las gráficas muestran un comportamiento contrario: la emisión de NOx disminuye a medida que aumenta la distancia recorrida. La causa principal de este resultado podría deberse a los siguientes factores:

- 1) El aumento del porcentaje de coches eléctricos o de coches con motores modernos que reducen la cantidad de NOx emitido. Para verificar esto, se han buscado datos sobre el porcentaje de coches eléctricos vendidos en Francia a lo largo de los años, los cuales se presentan en la siguiente gráfica:

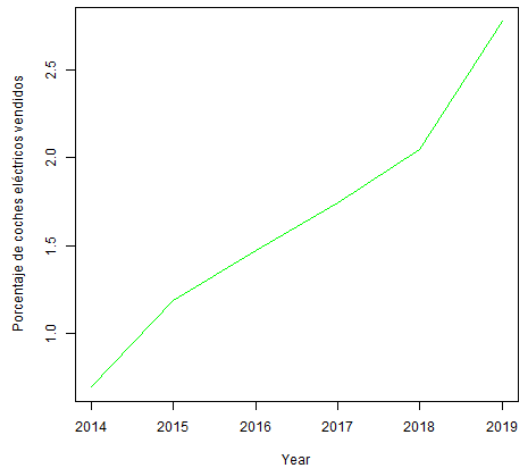


Figure 10: El porcentaje de coches eléctricos vendidos (Francia)

Se observa que el porcentaje de coches eléctricos ha ido aumentando, aunque de forma gradual, lo cual puede ser un factor que explique la reducción de NOx.

- 2) El uso de coches con motores modernos. Los coches hoy en día pueden utilizar motores que están diseñados para reducir la cantidad de contaminantes emitidos.
- 3) El último factor que puede contribuir a la reducción de NOx son los compromisos de reducción de emisiones asumidos por los países, los cuales pueden haber impulsado políticas y medidas más estrictas para controlar la contaminación. En la misma página web de la EEA se pueden encontrar datos sobre la reducción de emisiones que se han logrado, los cuales se muestran en la siguiente gráfica:

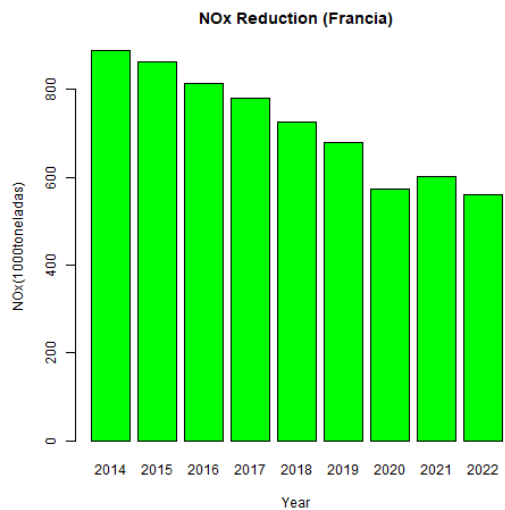


Figure 11: Reducción de emisión de NOx lograda (Francia)

## 6.1 Año 2020

A partir de las figuras (8 y 9) también se ha observado un dato interesante: en el año 2020 se registra un mínimo tanto en la gráfica de la emisión como en la de la distancia recorrida. Este comportamiento puede explicarse por el confinamiento durante ese año, cuando prácticamente nadie salía de casa, lo que resultó en una distancia recorrida mínima y, consecuentemente, en una emisión mínima. Por lo tanto, aunque debido a los compromisos de reducción de emisiones y el aumento de la cantidad de coches eléctricos la relación no parece muy significativa, se observa que si que existe una conexión directa entre estos dos indicadores, que se manifiesta en el mínimo de 2020

## 6.2 Regresión

En la siguiente gráfica se ha representado la emisión de NOx con respecto a la distancia recorrida anualmente por los coches, junto con una línea de regresión (Glm de grado 3):

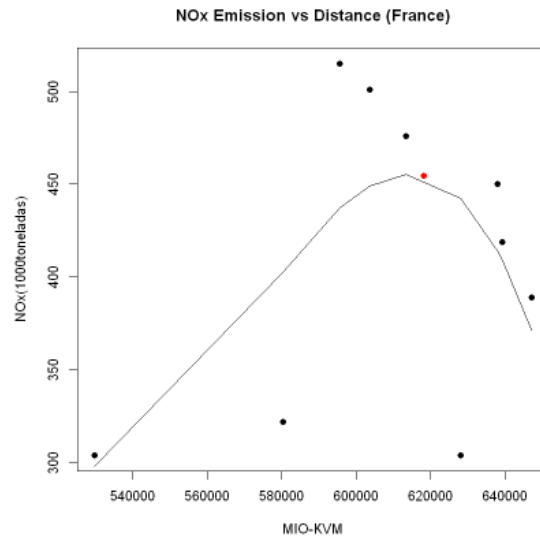


Figure 12: La emisión de NOx frente a la distancia recorrida, junto con la línea de regresión de grado 3.

Esta línea de regresión es capaz de predecir la emisión de NOx conociendo la distancia recorrida en un año concreto. Para ello, se ha realizado una predicción como prueba.

Se ha mencionado anteriormente, el dataset sobre la distancia recorrida tiene datos hasta el año 2023, mientras los datos del dataset sobre la emisión termina en el año 2022, por lo tanto, existe un punto extra de la distancia recorrida con el cual se puede estimar la emisión correspondiente, que es el punto rojo indicado en la figura anterior.

## 7 Conclusión

Se ha realizado un estudio sobre la relación entre la distancia recorrida y las emisiones de NOx, obteniendo como resultado que no se observa una relación relevante entre estas dos variables, lo cual puede deberse a dos posibles factores: El aumento de ventas de coches eléctricos y a la aparición de motores modernos cada vez menos contaminantes. A partir de los datos obtenidos se ha entrenado un modelo de regresión para poder estimar la emisión de NOx en función de la distancia recorrida en un año en concreto.

## 8 Posibles mejoras

Dado que el modelo obtenido no ha sido demasiado exitoso, se ha reflexionado acerca de una serie de mejoras que podrían ser útiles en proyectos futuros similares:

- Considerar y analizar también los datos sobre diferentes motores, ya que dependiendo del fabricante del motor, la emisión podría ser diferente.
- Analizar la vegetación del el país o zona de estudio, ya que es un factor importante a tener cuenta en la construcción de un modelo más sofisticado, puesto que la abundancia de vegetación se relaciona con la rapidez con la que son absorbidos los contaminantes.
- Limitar el estudio a zona más concretas. En lugar de estudiar la emisión a nivel nacional, conviene centrar el estudio a una ciudad concreta, quizás de esta forma las medidas serán más precisas.
- Aplicar un modelo más complejo para realizar la predicción. Utilizar modelos como árboles de regresión, random forest, en lugar de una regresión simple.

## 9 Referencias

### References

- [1] Página web de EEA(datasets sobre la calidad de aire): Calidad de aire
- [2] Página web de Eurostat(dataset sobre el vehículo-kilómetro): Tráfico
- [3] Our world in data(dataset sobre la cantidad de coches eléctricos vendidos): Coches eléctricos
- [4] URL Github URL