

# DMP

Autores

January 4, 2025

## Descripción del proyecto

En este proyecto se realizará un estudio sobre la relación entre los niveles de concentración de contaminantes (NOx y PM2.5) y la distancia recorrida anualmente por vehículos en 5 países europeos: España, Francia, Países Bajos, Rumanía y Suiza, entre los años 2014 y 2023

## 1 Summary

Los datasets sobre los niveles de contaminantes (NOx y PM2.5) que se analizarán en este proyecto fueron recolectados por la Agencia Europea de Medio Ambiente. Son datos públicos que cualquier persona puede descargar en formato CSV desde su página web oficial: Air quality data. Los datos están separados en diferentes datasets según el tipo de contaminante (NOx o PM2.5) y por país (Para este proyecto se descargarán datasets para Francia, España, Países Bajos, Suiza, Rumanía). Cada dataset contiene datos de NOx o PM2.5 para cada país entre los años 2010 y 2021. Para facilitar el análisis, se combinarán estos datos de diferentes países en un único dataset.

Los datasets sobre el transporte por carretera provienen del Cuestionario Común para Estadísticas de Transporte, desarrollado y analizado por Eurostat. Son datos públicos y pueden ser descargados en formato TSV desde la página web de Eurostat: Traffic data. Se trata de un dataset que contiene la distancia recorrida anualmente por los vehículos (medida en la unidad MIO-KVM, millones de kilómetros por vehículo) para diferentes países europeos entre los años 2014 y 2023.

Se analizará un tercer dataset, sobre la cantidad de coches eléctricos vendidos, este dataset se puede descargar en formato CSV desde la página web: Share of new passenger vehicles. Contiene datos sobre el porcentaje de vehículos eléctricos/híbridos para diferentes países europeos entre los años 2011 y 2019.

### 1.1 What kind of research output are you describing?

Los resultados del análisis incluirán tres datasets: los niveles de contaminantes (NOx y PM2.5), la distancia recorrida anualmente por los vehículos, y el porcentaje de vehículos eléctricos/híbridos vendidos. Además, se generarán visualizaciones gráficas para analizar la relación entre estos indicadores.

### 1.2 Is it physical or digital?

Los resultados obtenidos serán digitales.

### 1.3 Are you generating or re-using it?

Se reutilizará los datos existentes que han sido recolectados por fuentes de datos abiertos. También se generarán nuevos resultados y visualizaciones a partir de estos datos para el análisis.

## **1.4 What is the type of the described dataset?**

Los datasets generados serán de tipo secundario, ya que han sido previamente curados y analizados.

## **1.5 What is its format?**

Se generarán y exportarán los resultados del análisis en formato PNG (para las gráficas obtenidas) y CSV (para los datasets resultantes). Se utiliza PNG porque es un formato sin pérdida de compresión y comúnmente utilizado. Por otro lado, se usa CSV para los datasets debido a que es sencillo y compatible con casi cualquier herramienta y lenguaje de programación.

## **1.6 What is its expected size?**

Se estima que el tamaño final será inferior a 1MB.

## **1.7 To whom might it be useful ('data utility')?**

Los resultados pueden ser de interés para investigadores y académicos en el análisis de la calidad del aire, el cambio climático y el impacto del transporte en la salud pública.

# **2 LINKS BETWEEN OUTPUTS**

Las gráficas generadas mostrarán las relaciones entre los niveles de contaminantes y las distancias recorridas por los vehículos, también se estudiará temas como el impacto de los vehículos eléctricos en la reducción de contaminantes.

# **3 FAIR**

## **3.1 Making data and other outputs findable, including provisions for metadata**

### **3.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?**

Se utilizarán DOIs para los resultados obtenidos, que es un identificador único y persistente que se asigna a un objeto digital, para facilitar su localización y citación a lo largo del tiempo.

### **3.1.2 Will you provide metadata for the described dataset / output?**

Para cada uno de los datasets, se creará un archivo de metadatos en formato JSON utilizando Python y basándose en el esquema Dublin Core.

Se agruparán las gráficas y se creará un único archivo de metadatos para ellas.

## **3.2 Making data and other outputs openly accessible**

### **3.2.1 In which repository will the dataset / output be deposited?**

Los datasets resultantes y las gráficas obtenidas serán almacenados y preservados en repositorios de datos confiables. En este proyecto se utilizará una plataforma de datos abiertos denominado Zenodo. Esta plataforma asegura que los datos se almacenen en formatos accesibles y asigna Identificadores de DOI a los datasets.

## **3.3 DATA**

### **3.3.1 What is the described dataset / output title?**

Study of the relationship between pollutant levels and the annual distance traveled by vehicles in Spain, France, the Netherlands, Sweden, and Romania.

### **3.3.2 Are there any methods or tools required to access the dataset / output?**

A través de DOI correspondiente, cualquier persona podrá acceder y descargar los datasets en formato CSV, de manera gratuita y pública.

### **3.3.3 Please specify how the dataset / output will be accessed during and after the project ends**

Durante el proyecto, los datasets se almacenarán en repositorios locales en los dispositivos de los miembros del grupo. Se utilizarán plataformas en la nube para backup y compartir los datos entre los miembros del equipo.

Una vez finalizado el proyecto, los resultados se subirán a Zenodo para su almacenamiento a largo plazo y su disponibilidad pública.

### **3.3.4 Please specify how long after the project has ended the dataset / output will be made accessible for**

Una vez finalizado el proyecto, el dataset estará disponible de manera permanente en Zenodo.

## **3.4 Metadata**

### **3.4.1 Will you provide metadata even if the described dataset / output can not be openly shared?**

Se proporcionará metadatos para el dataset, incluso si no se pueden compartir abiertamente los datasets.

## **3.5 Making data and other outputs interoperable**

### **3.5.1 Does your (meta)data use a controlled vocabulary?**

Se utilizará Dublin core como controlled vocabulary.

### **3.5.2 Have you applied a standard schema for your (meta)data?**

Se proporcionarán los metadatos en el esquema Dublin Core.

## **3.6 Increasing data and other outputs reuse**

### **3.6.1 What internationally recognised licence will you use for your dataset / output?**

Se publicará los datasets bajo la licencia Creative Commons Attribution 4.0 International (CC BY 4.0), que permite su libre uso, modificación y distribución, siempre que se dé crédito adecuado a los autores originales.

### **3.6.2 What documented procedures for quality assurance do you have in place?**

Para garantizar la calidad de los datos en este proyecto, se seguirán los siguientes procedimientos: 1) Verificar que los datasets descargados provengan de fuentes confiables y oficiales. 2) Limpiar los datos utilizando herramientas como OpenRefine. 3) Revisar los códigos de programación utilizados en el análisis para garantizar su reproducibilidad. 4) Revisión de los resultados finales.

### **3.7 Allocation of resources**

#### **3.7.1 Identify the people who will be responsible and their role(s) in the management of the described output**

### **3.8 SECURITY**

#### **3.8.1 What security measures are followed?**

Los datos se compartirán exclusivamente entre los miembros del grupo a través de la plataforma en la nube, para asegurar el acceso seguro.

### **3.9 ETHICAL ASPECTS**

#### **3.9.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?**

Los datos utilizados en este proyecto han sido obtenidos de fuentes de datos abiertos, como Eurostat y la Agencia Europea de Medio Ambiente, las cuales permiten el acceso y uso público de la información. Por lo tanto, no existen problemas relacionados con la legalidad o la ética en su uso.

#### **3.9.2 Does the described dataset / output contain sensitive information?**

Los datasets que se proporcionarán no contendrán información sensible.

#### **3.9.3 Does the described dataset / output contain personal data?**

Los datasets que se proporcionarán no contendrán información personal.