

# Proyectos II, integración, calidad y análisis exploratorio de datos

## Segunda presentación:

### Integración de datos

### Introducción

Antes de cada HITO de presentación, debemos rellenar estas fichas y presentarlas a través de PoliformaT, en la tarea que indiquen los profesores. Cada equipo de trabajo presenta las mismas fichas. Sólo será necesario que las suba uno de los componentes del equipo.

Nombres y apellidos de los autores:

Juan José	Ruiz Esteban
Héctor	Beltrán Pozo
Celia	Villar Arcos
Ivette	Mahmoud-Yousef Puig
Iván	Navarro Martínez

### 2. Interés y alcance del proyecto.

Contesta a las preguntas que plantea este formulario.

#### 2.1. Explica el objetivo principal de tu proyecto ¿qué presenta este estudio?

El objetivo principal del proyecto es analizar la contaminación ambiental de distintos sectores según sus emisiones en las distintas ciudades europeas, así como su contribución en el PM 2.5 (materia particulada: partículas muy pequeñas en el aire que tienen un diámetro de 2.5 micrómetros). Además del estudio de otros factores que pueden afectar a la calidad del aire. Empleando el cruce de datos (Publications Office of the European Union - Urban Atlas Contaminants) y agregaciones nuestras a través de la recogida de datos con web scraping (tráfico aéreo).

#### 2.2. Explica para qué y para quién podría ser de utilidad este estudio

Este estudio puede ser útil para las autoridades y gobiernos locales para identificar las principales fuentes de contaminación y tomar decisiones informadas sobre políticas que mejoren la calidad del aire. También es valioso para organizaciones de salud pública, ya que ayuda a entender cómo la contaminación afecta la salud de las personas y permite crear campañas preventivas. Además, las empresas e industrias pueden utilizar los datos para optimizar sus procesos y reducir sus emisiones. Y por último puede ser empleado con fines de sensibilización social.

#### 2.3. ¿Por qué piensas que es novedoso? ¿has visto estudios similares?

Es novedoso porque combina datos de diferentes fuentes, como las emisiones de sectores específicos, el tráfico aéreo mediante web scraping y la calidad del aire en ciudades europeas, lo que permite un análisis más detallado. Aunque existen estudios sobre la calidad del aire y las fuentes de contaminación, la integración de datos tan específicos, como el tráfico aéreo y las emisiones de sectores urbanos, podría ofrecer una visión más precisa y actualizada. Además del estudio de la contribución de las partículas PM2.5. estimar el índice de contaminación en el sector industrial

### 2.4. Alcance (objetivos definitivos del proyecto)

**Define los objetivos del análisis de datos de tu proyecto.** Se deben presentar 5 objetivos de análisis.

Los objetivos del análisis son:

**OBJETIVO I.** *Cuantificar y caracterizar la contaminación generada por los principales sectores en Europa.*

Analizar los perfiles de emisiones de cada sector en las principales ciudades europeas, analizando los distintos contaminantes emitidos ( $\text{NH}_3$ , COVDM,  $\text{NO}_x$ ,  $\text{PM}_{2,5}$ ,  $\text{SO}_x$ ) y evaluando las diferencias entre regiones. Se busca entender qué tipo de gases son característicos para cada sector, cuál es el sector más contaminante y que sectores contaminan más a la calidad del aire de las ciudades.

**OBJETIVO II.** *Identificar los sectores con huella directa ciudadana y su impacto en la calidad del aire urbana.*

Agrupar sectores en los que la población tiene un impacto directo (transporte, generación eléctrica, gestión de residuos, calefacción residencial, etc.) y analizar cómo afectan a la calidad del aire en diferentes ciudades. Se representarán espacialmente sobre un mapa de Europa los sectores más contaminantes desde el punto de vista de la huella humana, buscando patrones comunes y diferencias regionales.

**OBJETIVO III.** *Evaluar el impacto de los aeropuertos en la contaminación urbana comparando 50 ciudades con aeropuerto frente a 50 sin aeropuerto, cuyas ciudades compartan características demográficas similares.*

Comparar un conjunto de 50 ciudades europeas con aeropuertos de gran tráfico aéreo con otras 50 ciudades de características demográficas y urbanas similares pero sin aeropuerto. El análisis incluirá el tamaño poblacional, nivel de urbanización y ubicación geográfica, con el fin de identificar el peso relativo de los aeropuertos en los perfiles de contaminación urbana.

**OBJETIVO IV.** *Estimar un índice de contaminación por país y analizar los sectores que no tienen relación directa con la huella ciudadana como las industrias o la agricultura.*

Construir una fórmula para estimar un índice de contaminación por país, a partir de los datos de calidad del aire de las ciudades principales ponderadas por su densidad poblacional. Este índice servirá para analizar diferencias entre países y explorar la contaminación atribuible a sectores donde la ciudadanía no tiene acción directa, como la agricultura extensiva o ciertas industrias pesadas, incorporando también indicadores económicos nacionales para analizar de mejor forma las diferencias entre sectores de diversos países.

**OBJETIVO V.** *Comparar patrones de calidad del aire entre regiones europeas con distintas características geográficas y demográficas.*

Comparar grupos de ciudades según su localización (norte-sur, costa-interior, cuenca mediterránea-atlántica), analizando cómo varían los perfiles de contaminación según sus ubicaciones o características. Se evaluarán combinaciones de factores críticos y se pondrán recomendaciones para su mitigación o gestión regional.

### 3. Calidad y Análisis exploratorio.

Describe la calidad de los datos y los resultados del análisis exploratorio efectuado. Explica el trabajo técnico, como por ejemplo, estadísticas aplicadas, visualizaciones o representaciones que has utilizado (puedes poner alguna captura como ejemplo), etc. Valora el esfuerzo del análisis exploratorio de tu proyecto.

No se trata de describir los objetivos resultado del proyecto, sino lo que has tenido que hacer para entender los datos, ver qué nos ofrecen, cómo de “sucia” es la fuente, etc.

Puedes comentar el descarte de datos de las fuentes. Aquellos que no vayas a tener en cuenta por no ser útiles a tus objetivos.

Recomendamos utilizar RMarkdown ya que el análisis exploratorio se espera que lo hayáis hecho en R. Poned un anexo donde veamos las instrucciones, comandos, funciones, etc. además de las gráficas.