

# **Práctica 2**

## **Madrid Sostenible**

*Infraestructura de Almacenamiento  
para la Ciudad Inteligente*

El Ayuntamiento de Madrid, a través del Área de Gobierno de Medio Ambiente y Movilidad, ha lanzado la iniciativa "Madrid Sostenible 2030" para transformar la capital en una ciudad inteligente orientada a la sostenibilidad. Como equipo de científicos de datos, se os ha encargado diseñar e implementar una **infraestructura de almacenamiento que integre múltiples fuentes de información sobre la huella ecológica** urbana, patrones de movilidad y calidad ambiental en los diferentes distritos madrileños. El proyecto tiene un objetivo triple: analizar el impacto medioambiental actual, identificar oportunidades de mejora, y crear un entorno donde diferentes actores urbanos (funcionarios municipales, empresas y ciudadanos) puedan acceder a información relevante según sus necesidades, sin requerir conocimientos técnicos avanzados.

## Datos Disponibles

1. **Base de datos municipal** (dump SQL):
  - Registros de consumo energético por los 21 distritos de Madrid y edificios públicos
  - Datos sobre distribución de infraestructuras urbanas en la ciudad
  - Inventario de parques, jardines y arbolado urbano de la ciudad
2. **Archivos de movilidad:**
  - `trafico_horario.csv`: Datos anónimos de los sensores de tráfico de Madrid con conteos de vehículos por hora, ubicación y tipo
  - `bicimad_usos.csv`: Registros de uso del servicio BiciMAD (fechas, estaciones de origen/destino, duración)
  - `parkings_rotacion.csv`: Estadísticas diarias de ocupación de los aparcamientos públicos de rotación (PAR) de Madrid
3. **Conjunto de datos de participación ciudadana:**
  - `avisamadrid.json`: Reportes de ciudadanos sobre problemas urbanos a través de la app Avisa Madrid (categoría, ubicación, descripción, fecha)

## Objetivos del Proyecto

El objetivo principal de este proyecto es desarrollar una **infraestructura de datos integrada** que satisfaga las necesidades de tres perfiles de usuario distintos, cada uno con diferentes habilidades técnicas y requerimientos analíticos.

La infraestructura se considerará **válida** cuando demuestre su capacidad para responder efectivamente a **preguntas específicas** para cada perfil:

### Objetivo 1

Para **científicos de datos** de **empresas** y **organizaciones** locales, la infraestructura debe permitir analizar mediante notebooks de Python: "¿Cuáles son los horarios de mayor congestión de tráfico en Madrid y qué tipos de vehículos predominan en esas franjas?". Solo si estos usuarios pueden acceder, procesar y leer estos datos de forma programática, este componente de la solución se considerará exitoso.

## Objetivo 2

Para **gestores municipales** con conocimientos avanzados de **SQL**, la infraestructura debe proporcionar un modelo de datos que permita responder mediante consultas SQL a: "¿Qué rutas de BiciMAD son más populares entre los usuarios y cómo varían los patrones de uso entre usuarios abonados y ocasionales?" y "¿Cómo se relaciona la densidad de población de los distritos con la presencia de infraestructura de transporte público?". La validez de esta capa de la solución se determinará por la eficiencia y precisión con que estas consultas pueden ejecutarse.

## Objetivo 3

Para **ciudadanos** y **asociaciones** vecinales sin conocimientos técnicos, la infraestructura debe ofrecer **visualizaciones** que respondan claramente a: "¿Qué aparcamientos públicos presentan mayores variaciones de ocupación a lo largo del día y la semana, y cómo se correlacionan con su ubicación en la ciudad?". La solución será válida solo si personas sin formación técnica pueden obtener esta información de manera intuitiva y sin necesidad de escribir código o consultas.

El reto consiste en diseñar una infraestructura que evolucione para satisfacer estos tres niveles de interacción, desde el análisis programático avanzado hasta la visualización simplificada, manteniendo la coherencia y trazabilidad de los datos a lo largo de todo el proceso.

Opcionalmente se puede:

1. **Ampliar el catálogo de preguntas:** Incorporar nuevas preguntas relevantes, especialmente aquellas que aprovechen fuentes adicionales de datos que se hayan integrado para enriquecer el análisis.
2. **Definir KPIs clave:** Proponer y calcular un conjunto de indicadores clave de desempeño (KPIs) relacionados con la sostenibilidad urbana que puedan ser monitorizados regularmente por los diferentes usuarios del sistema.

## Entregables Esperados

1. **Repositorio Git** completo con:
  - a. Código fuente de todos los componentes
    - i. **Archivos Docker** que permitan desplegar toda la infraestructura
    - ii. Scripts y ficheros adicionales necesarios (e.g. notebooks, etc)
  - b. README.md que detalle la **Memoria técnica** con:
    - i. Diagrama de la infraestructura
    - ii. Modelo de datos diseñado
    - iii. Procesos de transformación implementados
    - iv. Guía de puesta en marcha
    - v. Ejemplos de uso y soporte a las consultas
    - vi. \*Detalle de las fuentes adicionales incorporadas y su valor añadido
2. **PDF en Moodle** que contenga:
  - a. URL completa del repositorio Git para poder evaluarlo

- b. Nombres y correos de los integrantes del equipo
- c. Copia del README del proyecto con la memoria técnica

## Desarrollo de la Práctica

Esta es una planificación orientativa para desarrollar la infraestructura de datos en función de las necesidades de los diferentes perfiles de usuario:

### **Sesión 1: Entorno para científicos de datos**

- Creación del repositorio y estructura base del proyecto
- Exploración de las fuentes de datos mediante notebooks de Jupyter
- Desarrollo de análisis que responda a la pregunta

### **Sesión 2: Modelo de datos para analistas SQL**

- Ampliación del entorno con BBDD relacional
- Diseño del modelo de datos
- Diseño de consultas SQL complejas

### **Sesión 3: Visualización para usuarios no técnicos**

- Integración de herramienta de visualización
- Desarrollo de dashboards interactivos para datos de aparcamientos
- Implementación de filtros y controles accesibles para usuarios sin conocimientos técnicos

### **Sesión 4: Documentación**

- Definición de KPIs relevantes y su visualización
- Validación completa de la infraestructura con los tres perfiles de usuario
- Documentación técnica y preparación de la demostración

Se recomienda seguir un enfoque iterativo, implementando primero una solución básica funcional que luego se pueda refinar y extender.

## Evaluación

El trabajo será evaluado según una rúbrica que considerará aspectos como:

- Adecuación de la arquitectura al problema planteado
- Calidad del modelo de datos diseñado
- Efectividad de los procesos de transformación
- Usabilidad para los diferentes perfiles de usuario
- Calidad y relevancia de las visualizaciones
- Valor añadido a través del enriquecimiento con fuentes adicionales
- Documentación y claridad del código
- Capacidad para extraer insights valiosos de los datos