

# INTRODUCCIÓN A OPEN DATA

**Lidia Contreras**  
*@liconoc*

**Cristina I. Font**  
*@Cristina\_ipunto*



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA

**DSIC**  
DEPARTAMENT DE SISTEMES  
INFORMÀTICS I COMPUTACIÓ



**DCADHA**  
Departamento de Comunicación Audiovisual,  
Documentación e Historia del Arte



# ¿QUÉ ES OPEN DATA?

---



Música, Libros, Películas, Datos científicos, Datos históricos, Datos geográficos, Información Gubernamental, Información de Comunidades Autónomas y Ayuntamientos...

# ¿POR QUÉ, CÓMO Y CUÁNDO?

---



Pese a que el Gobierno Abierto surge a principios de 1900 en Inglaterra, 2009 es una fecha clave para el movimiento.

# ¿QUÉ NO ES OPEN DATA?

Algunos ejemplos:



## IMÁGENES

Usar PDFs con imágenes escaneadas.



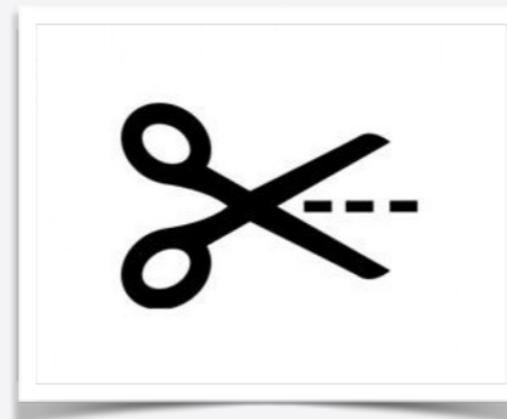
## LEGIBILIDAD

Falta de legibilidad o formatos que dificulten el copiado.



## EXPLICACIÓN

La falta de explicaciones que dificultan la comprensión



## DIVISIÓN

Datasets en pequeños archivos.



## INTEROPERABILIDAD

CUANTOS MÁS DATOS  
PODAMOS UNIR MÁS  
LEJOS LLEGAREMOS

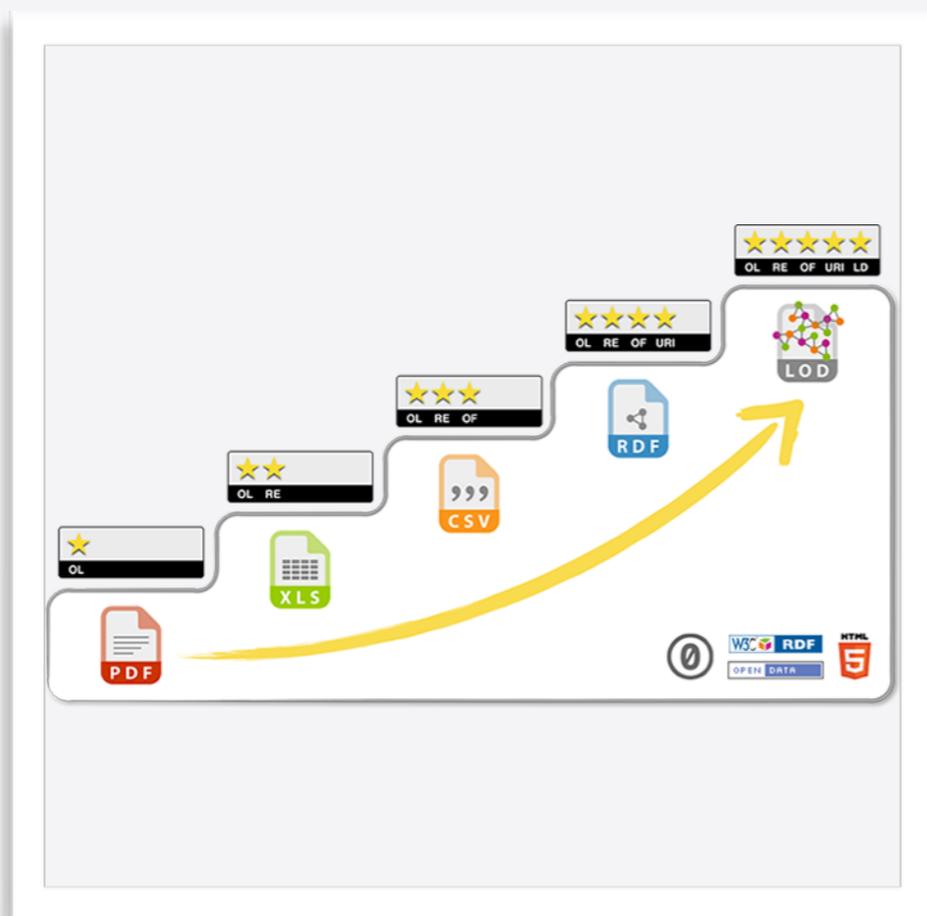


# DATOS ABIERTOS DE

5★

Tim Berners-Lee

# COSTE / BENEFICIO



## GRADO DIFICULTAD / APERTURA

Fácil publicación, difícil reutilización



Valores medios



Reutilización y publicación completa



# SECTOR PRIVADO

---

# INVESTIGACIÓN

ESENCIAL Y CRÍTICO

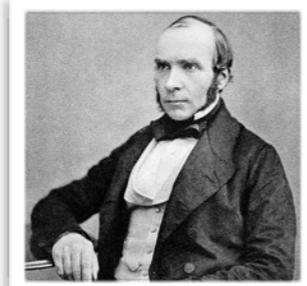
NUEVAS VÍAS Y CAMINOS

VALIDACIÓN Y RECONOCIMIENTO



# EJEMPLOS

---



## JOHN SNOW

Médico inglés s.XIX. Demostró las causas del cólera al cruzar datos de muertes con datos de zonas de agua.



## ADM. PÚBLICAS

Portales con publicación de aplicaciones privadas.



## S. PRIVADO

MWC y su impacto en el turismo.



## DATAUPV

Datos abiertos y sus aplicaciones por parte de estudiantes.



B I G B A N G

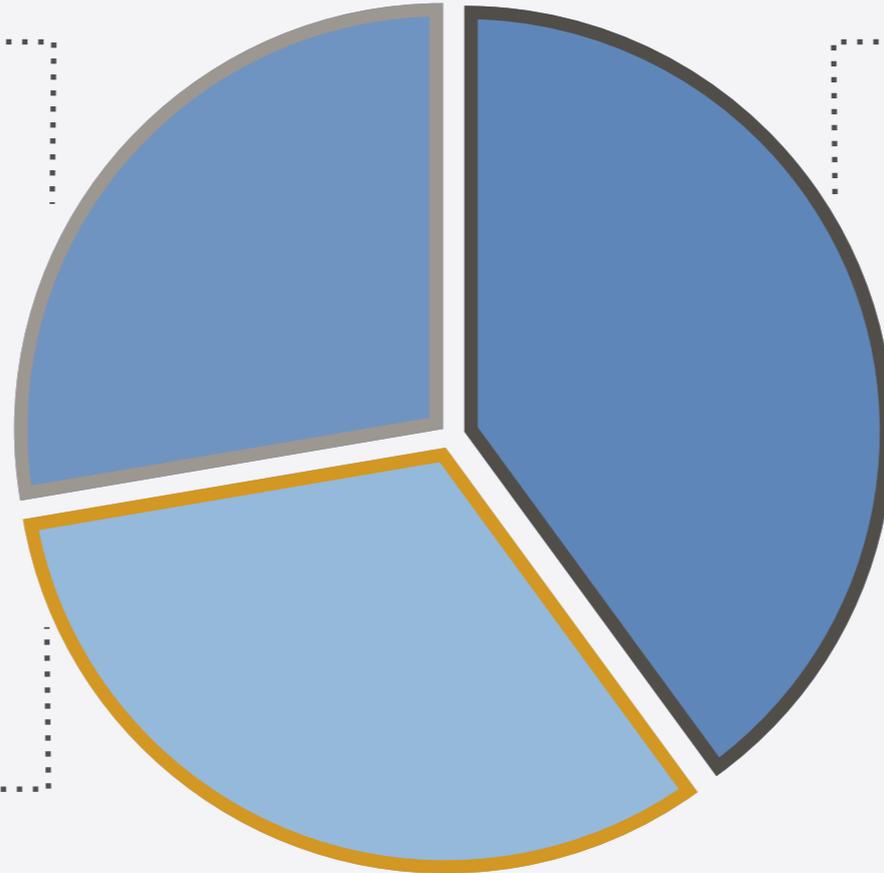


# RECETA

---

MACHINE LEARNING

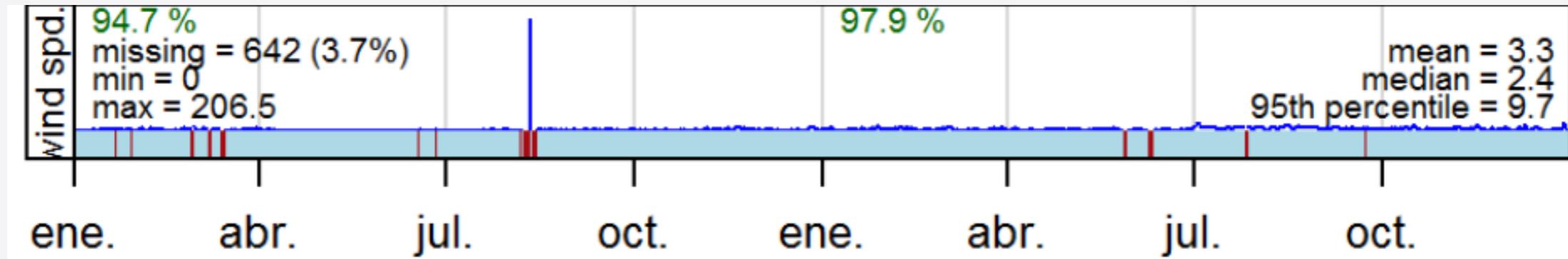
DATOS ABIERTOS



BUENA IDEA



# TRATAMIENTO DE LOS DATOS (I)



1°

## PREPARACIÓN

- ✓ Diferentes formatos de archivo (csv, json, txt...).
- ✓ Diferentes campos (Hora, HORA, H, hora...).
- ✓ Diferentes idiomas (Hora, hour, time...).

2°

## LIMPIEZA

- ✓ Datos anómalos.
- ✓ Datos duplicados.
- ✓ Símbolos (comas, puntos...).
- ✓ Datos irrelevantes.

# TRATAMIENTO DE LOS DATOS (II)

3°

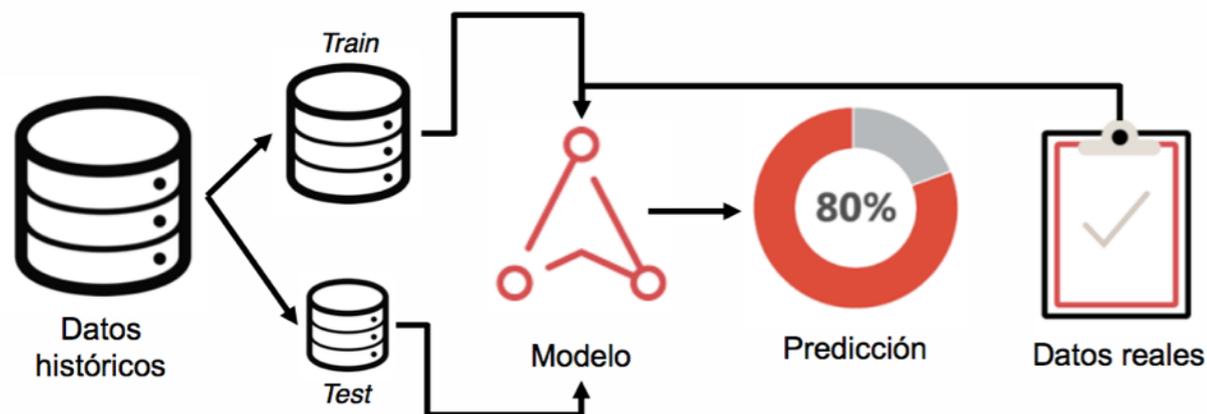
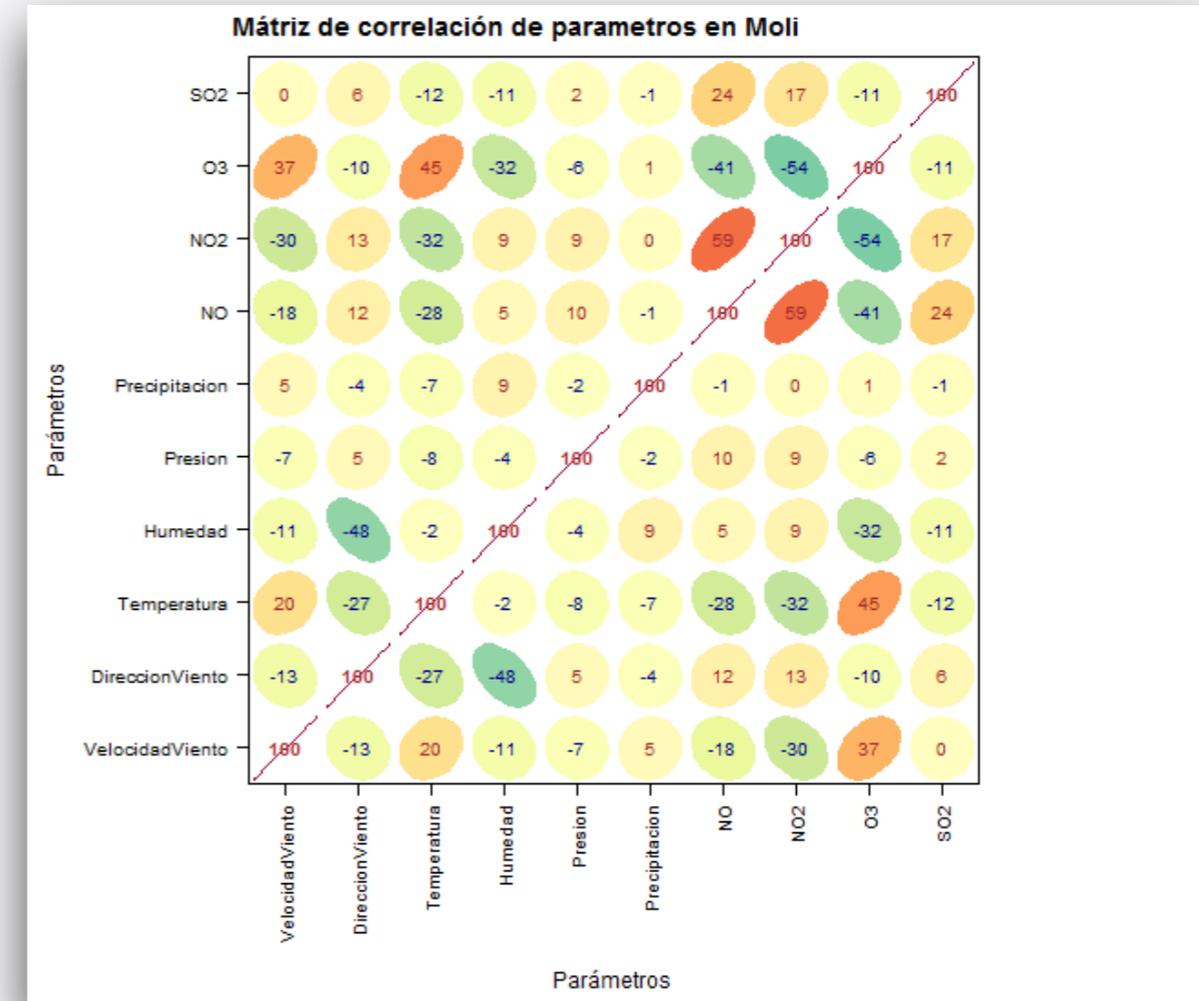
## ESTUDIO DE LOS DATOS

- ✓ Correlaciones entre variables
- ✓ Resúmenes
- ✓ Visualizaciones

4°

## GENERAR NUEVO CONOCIMIENTO

- ✓ Machine Learning



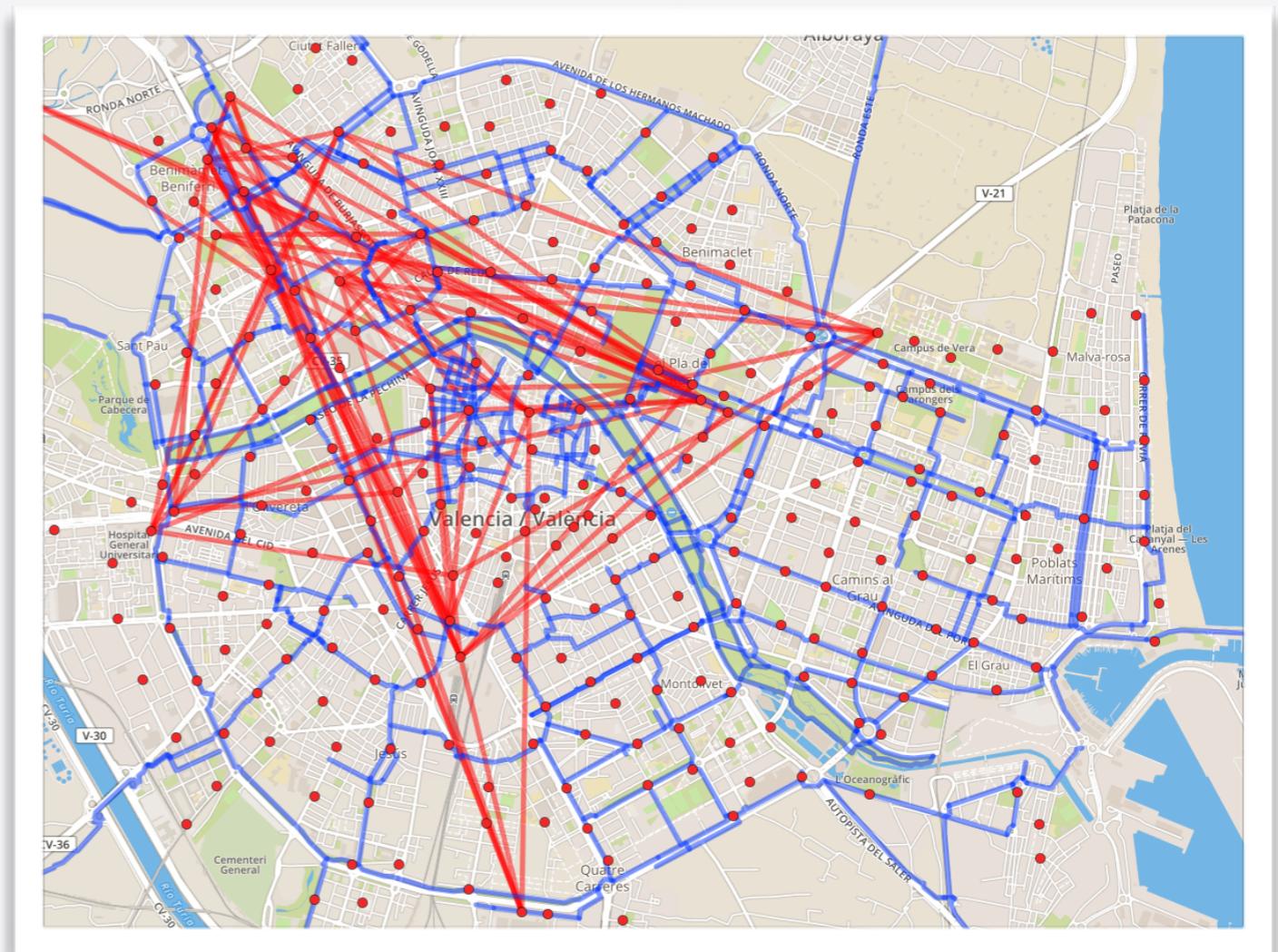
# EJEMPLO 1: MOREBIKELINES

## DATOS Y FUENTES

- RED DE CARRILES BICI DE VALENCIA
  - Fuente: Ayuntamiento de Valencia
  - Parámetros: Coordenadas de los carriles bici.
- USO DE VALENBISI
  - Fuente: Valenbisi
  - Parámetros: n° bornes, bicicletas disponibles, bornes disponibles

## OBJETIVO

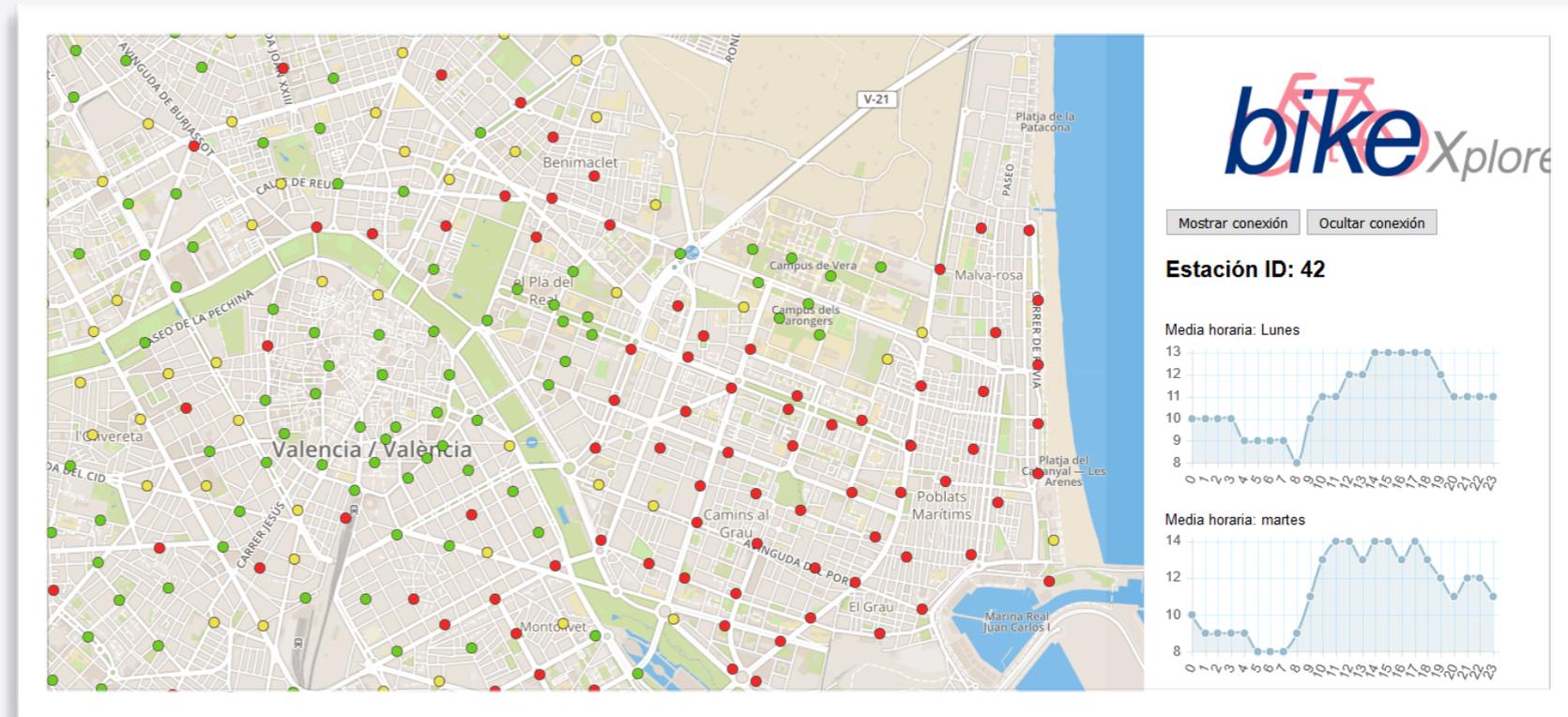
Determinar la necesidad de nuevos carriles bici en la ciudad de Valencia, a partir del uso de las estaciones de Valenbisi.



# EJEMPLO 2: BIKEEXPLORER

## DATOS Y FUENTES

- USO BICICLETAS
  - Fuente: Valenbisi
  - Parámetros: n° bornes, bicicletas disponibles, bornes disponibles
- DATOS DEL CENSO
  - Fuente: Ayuntamiento de Valencia
  - Parámetros: Población por edades
- CLASIFICACIÓN POR ZONAS
  - Fuente: Manual
  - Parámetros: Tipo de zona



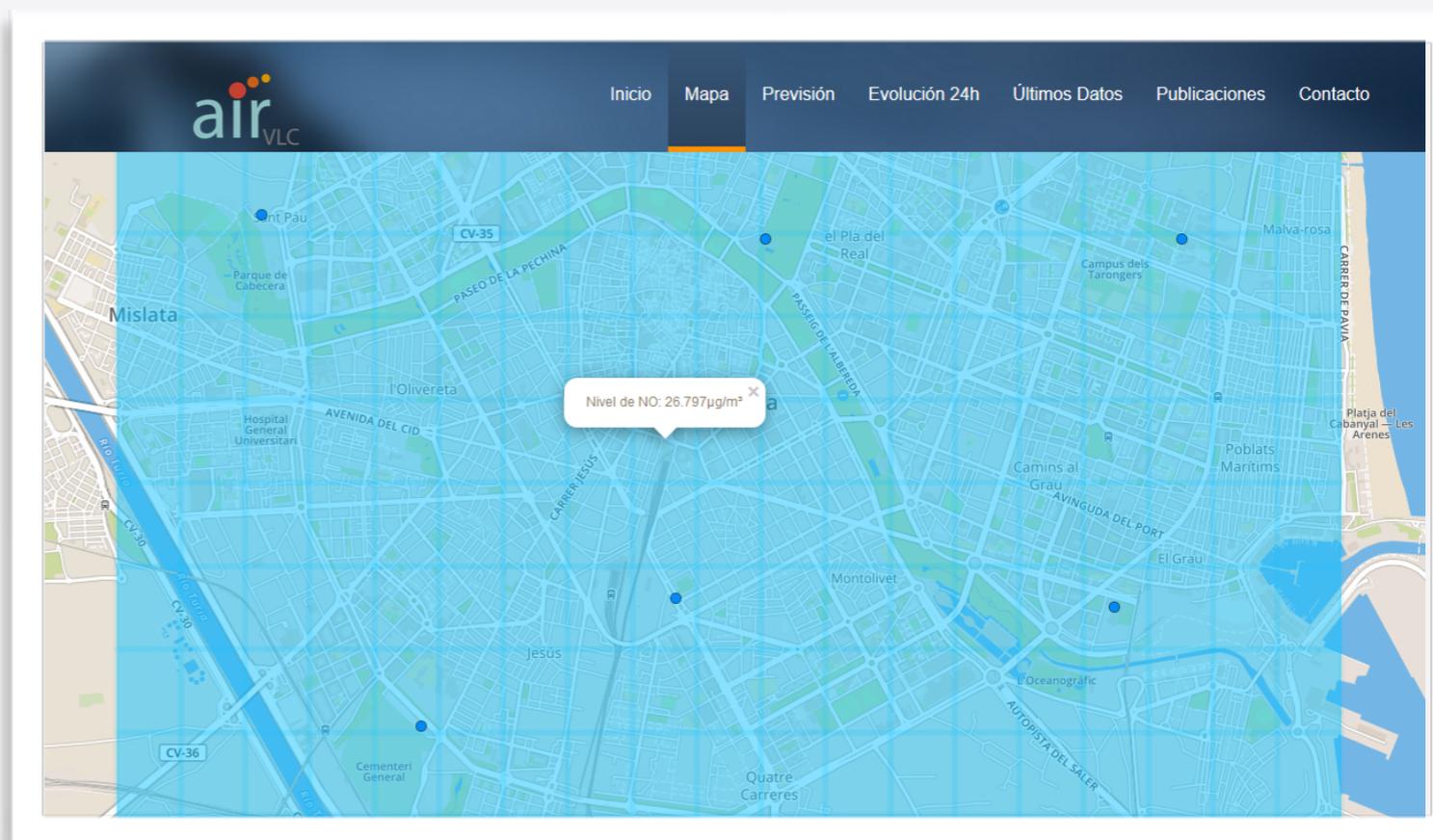
## OBJETIVO

Predicción del uso y disponibilidad de bicicletas en el servicio de Valenbisi.

# EJEMPLO 3: AIRVLC

## DATOS Y FUENTES

- CONTAMINACIÓN
  - Fuente: Generalitat Valenciana
  - Parámetros: NO, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>
- DATOS METEOROLÓGICOS
  - Fuente: AEMET
  - Parámetros: Temperatura, Humedad, Velocidad y dirección del viento, Precipitación y Presión
- TRÁFICO
  - Fuente: Ayuntamiento de Valencia
  - Parámetros: Intensidad de los Puntos de Medida de Tráfico - Espiras electromagnéticas (Vehículos/hora)



## OBJETIVO

Predicción de los niveles de contaminación en Valencia a partir del tráfico y la observación meteorológica.



Thank you!

Lidia Contreras  
@liconoc  
[liconoc@upv.es](mailto:liconoc@upv.es)

Cristina I. Font  
@Cristina\_ipunto  
[crifonju@gmail.com](mailto:crifonju@gmail.com)

