

Casos de uso del Data Science para una Smart City



PIPERLAB

María Medina - PiperLab





Hello my name is



María Medina

Data Scientist en PiperLab

Informática y matemática

@mariamedp
@piperlab_es

“

**Ayudamos a nuestros
clientes a afrontar
problemas de negocio
tomando decisiones
basadas en datos.**



PIPERLAB

#UOCDataDay



@mariamedp
@piperlab_es



#UOCDataDay

ECOMMERCE

@mariamedp
@piperlab_es

www.piperlab.es

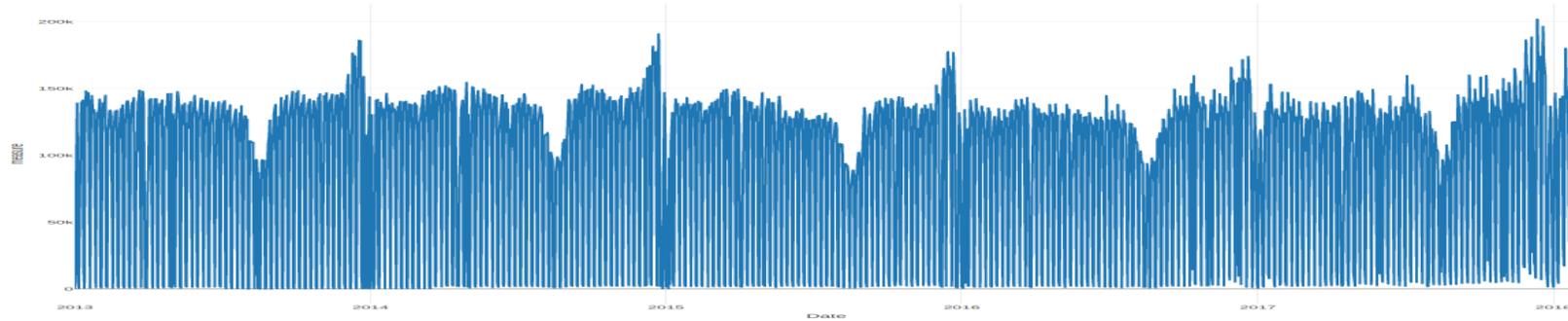
UOC



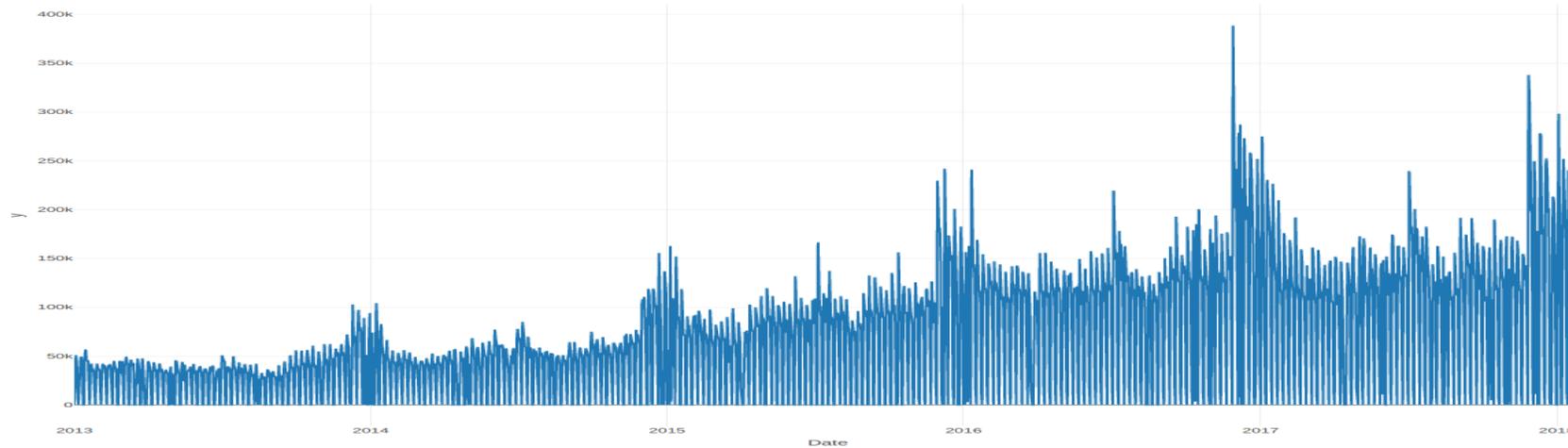
Repartos en SEUR



B2B



B2C

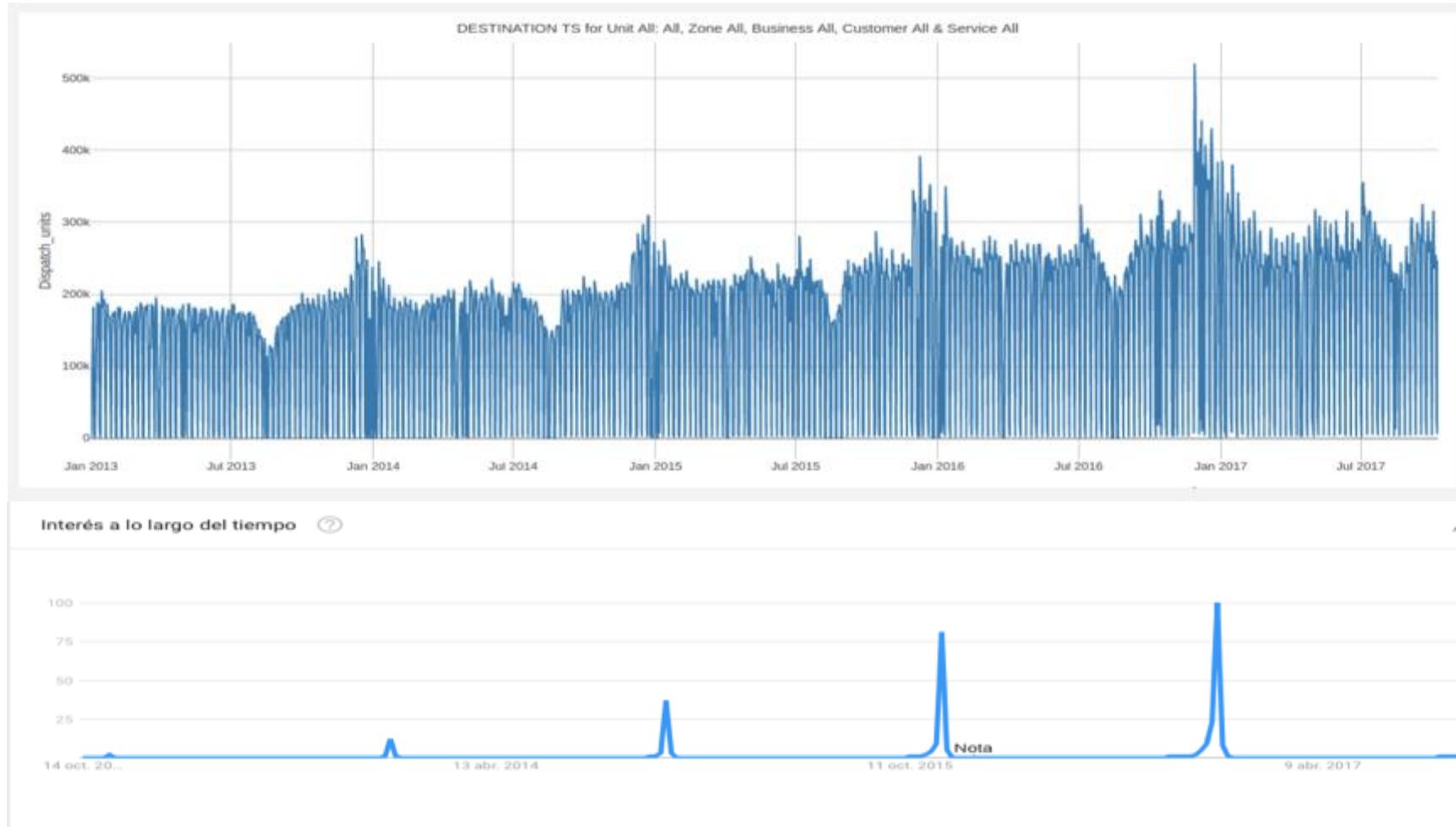


@mariamedp
@piperlab_es



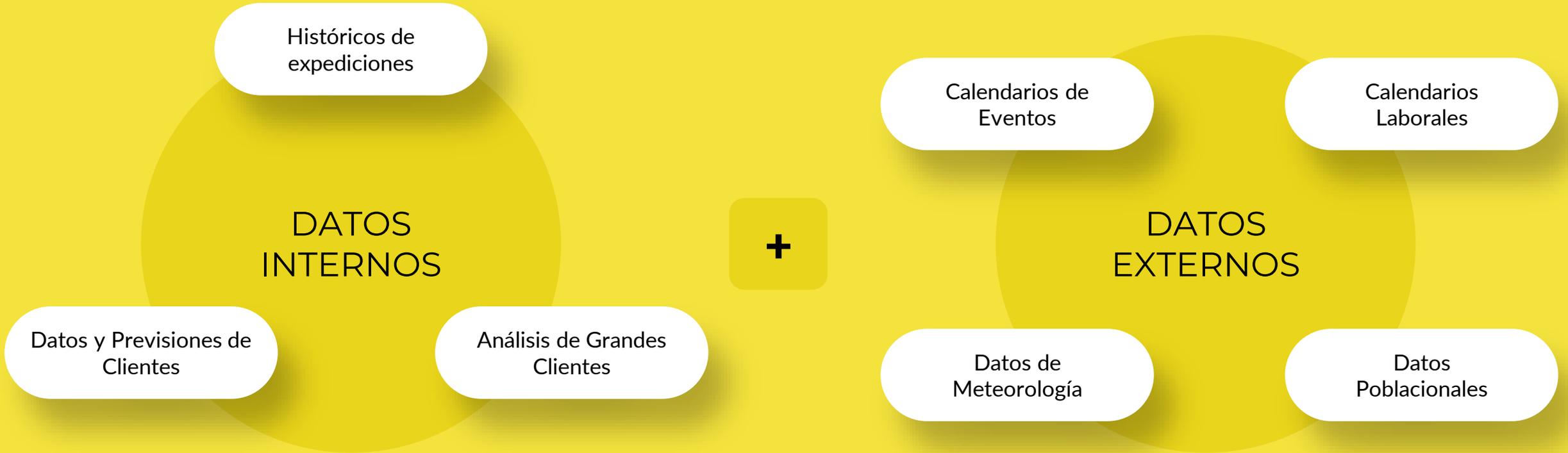


La influencia del Black Friday



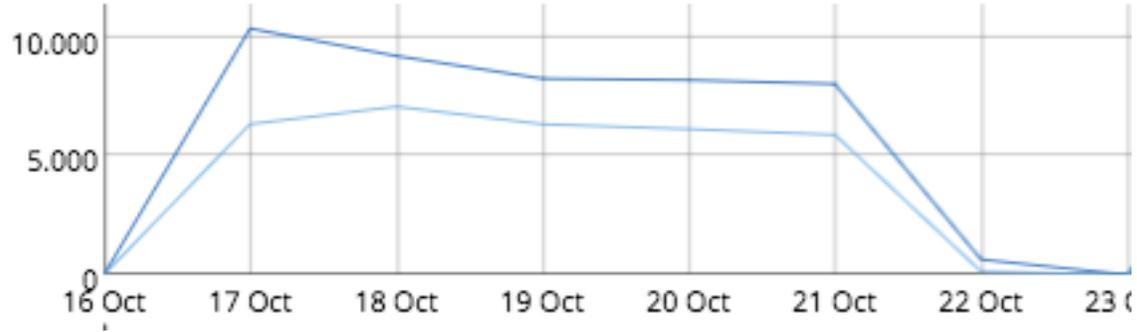


Datos relevantes para el modelo

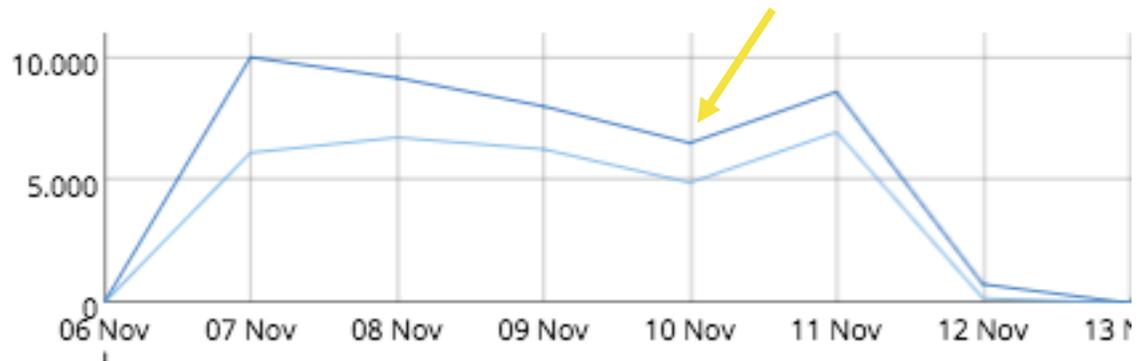




Ejemplo: festivales de otras ciudades



Distribución semanal de las expediciones a repartir en una semana tipo en **Sevilla**

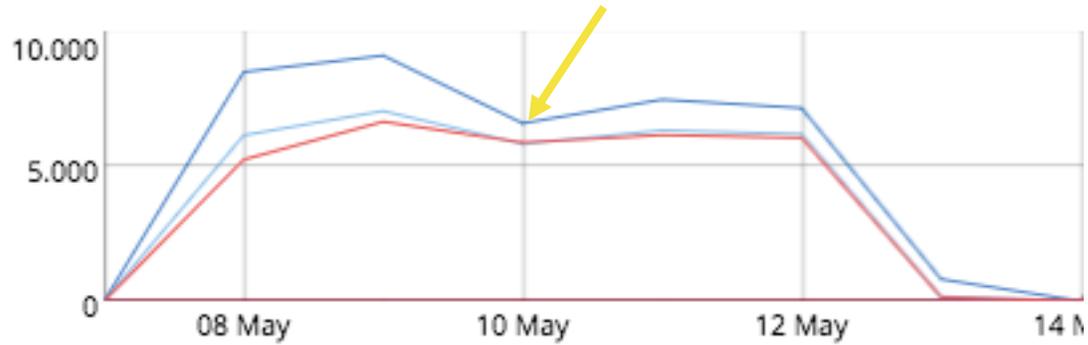


Efecto de un festivo local en Madrid (La Almudena) en la distribución semanal de las expediciones a repartir en **Sevilla**

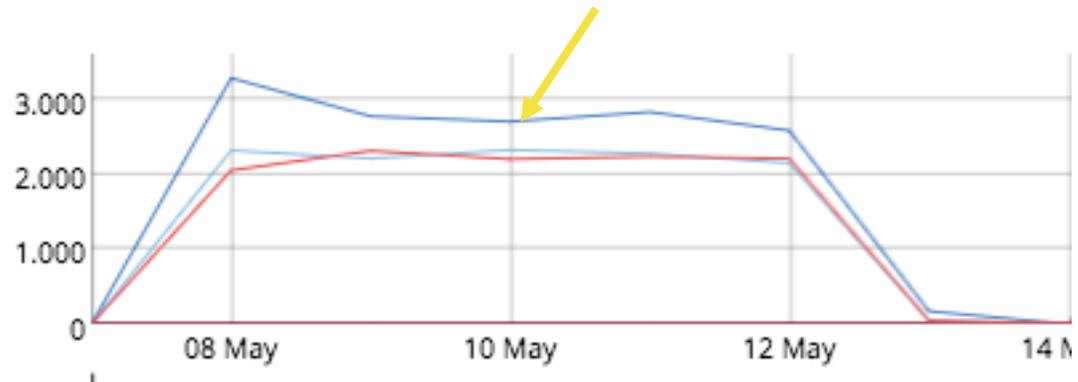




Ejemplo: festivales de otros países



Distribución semanal de las expediciones a repartir en Málaga. Efecto en Málaga del festivo en Francia: 2.000 expediciones menos



Distribución semanal de las expediciones a repartir en Jaén. No existe efecto en Jaén del festivo en Francia





Qué hemos hecho



Más de 9000 modelos

- Para todas las plazas de España, Portugal y Andorra.
- Por origen y destino.
- Por tipo de producto.
- B2B / B2C.
- N° bultos, kilos y N° expediciones.



Herramienta web

- Más de 400 usuarios
- Múltiples departamentos: operaciones, call center, ...



Reentrenamiento automático

- ARIMAs
- Prophet
- SVMs

#UOCDataDay



@mariamedp
@piperlab_es



#UOCDataDay

CONTAMINACIÓN

@mariamedp
@piperlab_es

UOC



Once upon a time...

Ayuntamiento Madrid @MADRID · 11 nov. 2015

Si persiste la alta contaminación, mañana habrá medidas restrictivas de tráfico. Informaremos aquí desde medianoche. madrid.es/portales/munim...

RETWEETS 212 ME GUSTA 68

19:11 - 11 nov. 2015 · Detalles

Ayuntamiento Madrid @MADRID · 12 nov. 2015

Si el nivel de NO2 continúa subiendo, mañana podría aplicarse el escenario 2 del protocolo anticontaminación: prohibido aparcar en zona SER.

RETWEETS 148 ME GUSTA 34

12:00 - 12 nov. 2015 · Detalles

Ayuntamiento Madrid @MADRID · 14 nov. 2015

Este sábado no habrá restricciones al tráfico, se desactiva el protocolo. Gracias a quienes habéis hecho posible que baje el nivel de NO2.

RETWEETS 524 ME GUSTA 264

0:25 - 14 nov. 2015 · Detalles



Ayuntamiento Madrid @MADRID · 11 nov. 2015

Jueves 12 de noviembre, desde las 6:00: máximo 70km/h en la M30 y accesos desde la M40. Se revisará por la noche.

Activado el protocolo por Alta Contaminación (Escenario 1)

- Velocidad máxima de circulación: 70 km/h en M30 y accesos dentro de M40 en ambas direcciones
- Se recomienda usar transporte público
- Limitar ejercicio al aire libre para personas con problemas respiratorios

Protocolo aprobado en marzo de 2015, consta de escenarios 0, 1, 2 y 3.

RETWEETS 564 ME GUSTA 173

22:24 - 11 nov. 2015 · Detalles

Ayuntamiento Madrid @MADRID · 12 nov. 2015

Malas noticias: persiste la alta contaminación, así que a las 6:00 de este viernes 13 entramos en el escenario 2.

Activado el protocolo por Alta Contaminación (Escenario 2)

- Velocidad máxima de circulación: 70 km/h en M30 y accesos dentro de M40 en ambas direcciones
- Está prohibido aparcar vehículos no residentes dentro de la zona SER
- Se recomienda usar transporte público y limitar ejercicio al aire libre para personas con problemas respiratorios

Protocolo de marzo de 2015, consta de escenarios 0, 1, 2 y 3.

RETWEETS 1.124 ME GUSTA 229

23:15 - 12 nov. 2015 · Detalles





Datos de contaminación

Portal de Datos Abiertos del Ayuntamiento de Madrid



Calidad del aire: Datos en tiempo real

Calidad del aire: Datos en tiempo real. El Sistema Integral de la Calidad del Aire del Ayuntamiento de Madrid permite conocer en cada momento los niveles de contaminación atmosférica en el municipio. En este conjunto de datos puede obtener la información actualizada en tiempo real, actualizándose estos datos cada hora, y...

Calidad del aire: Estaciones de control

Calidad del aire: Estaciones de control. El Sistema de Vigilancia está formado por 24 Estaciones Remotas automáticas que recogen la información básica para la vigilancia atmosférica. Poseen los analizadores necesarios para la medida correcta de los niveles de gases y de partículas. Las estaciones remotas son de varios tipos: Urbanas de fondo; representativas de la exposición...

Calidad del aire: Datos horarios años 2003 a 2016

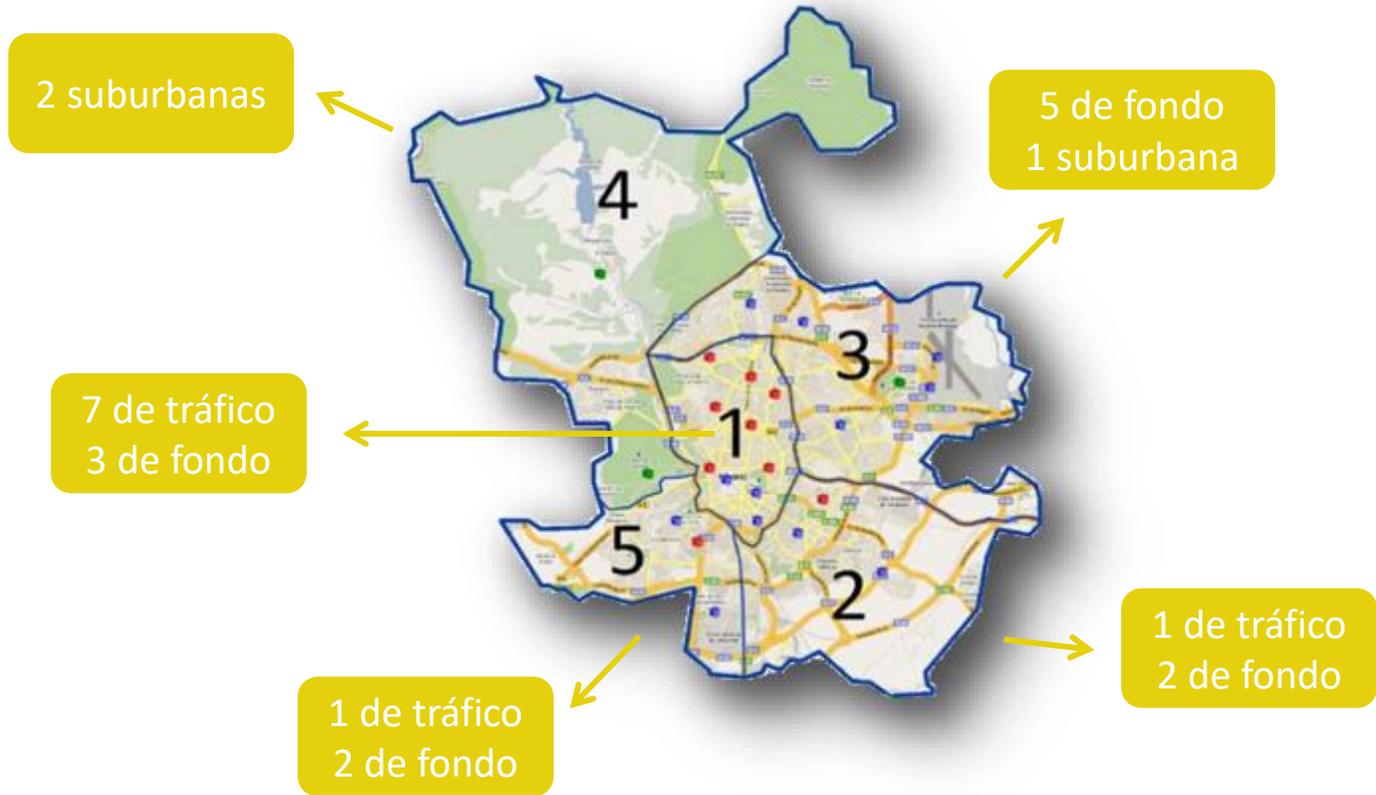
Calidad del aire: Datos horarios años 2003 a 2016. El Sistema Integral de la Calidad del Aire del Ayuntamiento de Madrid permite conocer en cada momento los niveles de contaminación atmosférica en el municipio. En este conjunto de datos puede obtener la información recogida por las estaciones de control de

Calidad del aire: Datos diarios años 2003 a 2016

Calidad del aire: Datos diarios años 2003 a 2016. El Sistema Integral de la Calidad del Aire del Ayuntamiento de Madrid permite conocer en cada momento los niveles de contaminación atmosférica en el municipio. En este conjunto de datos puede obtener la información recogida por las estaciones de control de



Medición contaminación



- Óxidos de nitrógeno
- Ozono
- Partículas en suspensión
- Dióxido de azufre
- Monóxido de carbono
- Niveles sonoros
- ...





Datóxido de Nitrógeno

 @datoxnitro_bot



Datóxido Nitrógeno
@datoxnitro_bot

Bot que publica en tiempo real los niveles de contaminación por dióxido de nitrógeno en la ciudad de Madrid. Datos de datos.madrid.es.

Madrid, Comunidad de Madrid

Tweets 20,2 mil Siguiendo 2 Seguidores 1.676 Siguiendo

Tweets Tweets y respuestas Multimedia

Datóxido Nitrógeno @datoxnitro_bot · 25 ene. 2016
¡Lanzamos el bot coincidiendo con la charla en #databeers #mad!
Hilo explicando el protocolo #contaminaciónmadrid

Datóxido Nitrógeno @datoxnitro_bot
En Madrid hay 24 estaciones que miden la concentración de dióxido de nitrógeno (NO2) en el aire. Mapa ubicaciones...

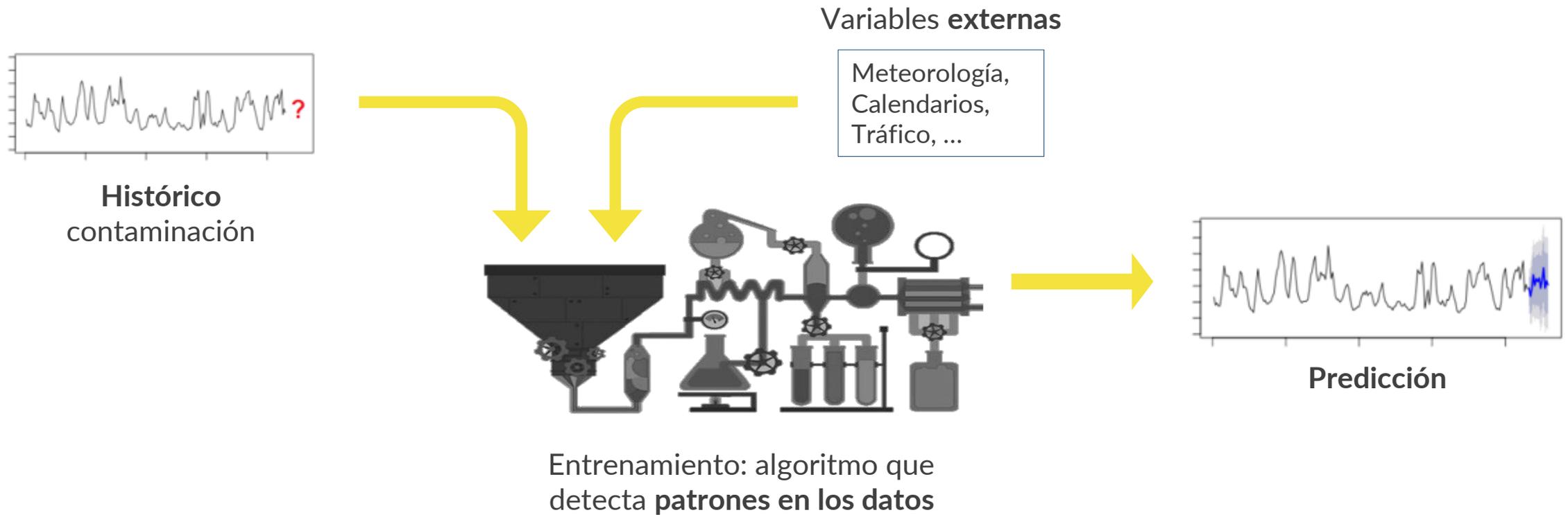




¿Podemos predecir la contaminación?



Proceso de predicción





Proyecto PREDICT

“ *Un enfoque real para un mundo real* ”



Objetivo de Predicción

- Índice de contaminación: activaciones protocolo



Actualización

- Generación de una nueva predicción cada 24 horas



Alcance de la predicción

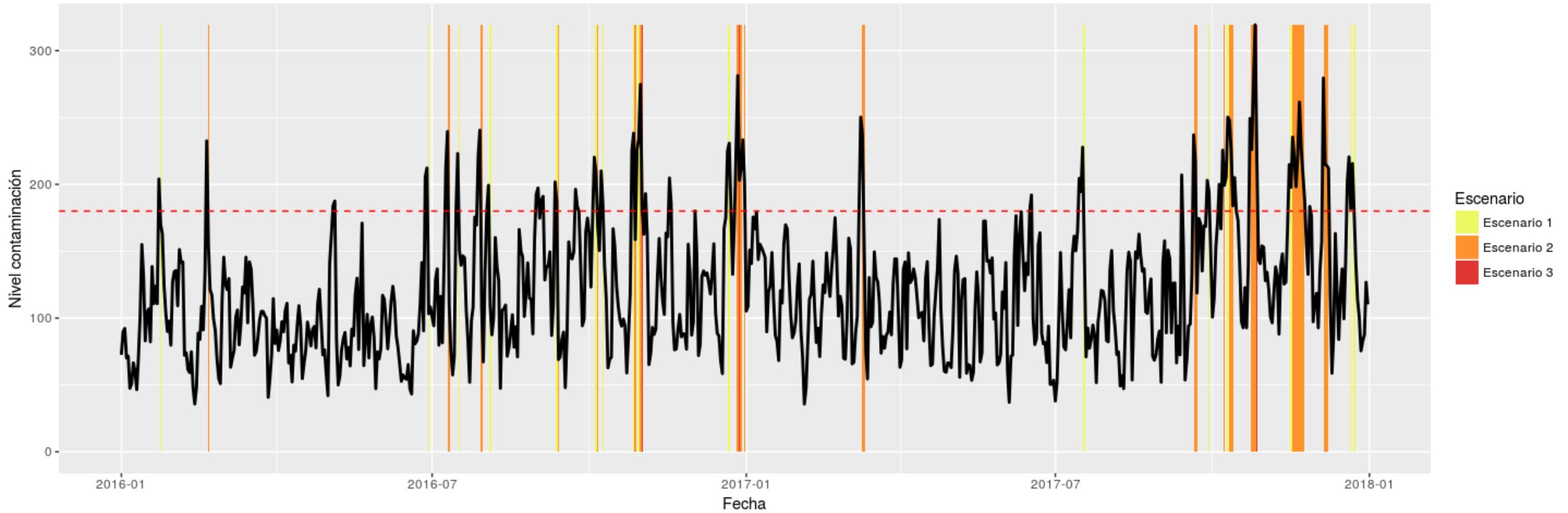
- Corto plazo: 24 horas
- Medio plazo: 48 horas
- Largo plazo: 7 días



Serie contaminación



Objetivo: saber cuándo se va a activar el protocolo





Variables relevantes



METEOROLOGÍA

La contaminación aparece especialmente durante procesos de **inversión térmica** y **poca ventilación**. Esto se detecta a través de: temperatura, presión, humedad relativa, viento, radiación, estabilidad atmosférica.



TRÁFICO

La principal fuente de contaminación por NO₂ es el tráfico rodado. Los **niveles de tráfico** y la **antigüedad del parque móvil** son relevantes a la hora de estimar los niveles de contaminación.



CALENDARIO

Como es difícil conocer el estado del tráfico con antelación, esta información se puede suplir con el **día de la semana**, los **festivos** o puentes locales o nacionales, o el **día del mes**. Además, la contaminación presenta patrones estacionarios y tendencias a más largo plazo que se pueden aproximar usando el **mes** o el **año**.



CONTAMINACIÓN

La tendencia de los niveles previos de contaminación sirve para anticipar futuros valores, incluso de contaminantes diferentes. Por ejemplo: entre NO₂ y O₃ se producen **reacciones químicas** que alteran los niveles.



Variables relevantes



METEOROLOGÍA

La contaminación aparece especialmente durante procesos de **inversión térmica** y **poca ventilación**. Esto se detecta a través de: temperatura, presión, humedad relativa, viento, radiación, estabilidad atmosférica.



TRÁFICO

La principal fuente de contaminación por NO₂ es el tráfico rodado. Los **niveles de tráfico** y la **antigüedad del parque móvil** son relevantes a la hora de estimar los niveles de contaminación.



CALENDARIO

Como es difícil conocer el estado del tráfico con antelación, esta información se puede suplir con el **día de la semana**, los **festivos** o puentes locales o nacionales, o el **día del mes**. Además, la contaminación presenta patrones estacionarios y tendencias a más largo plazo que se pueden aproximar usando el **mes** o el **año**.

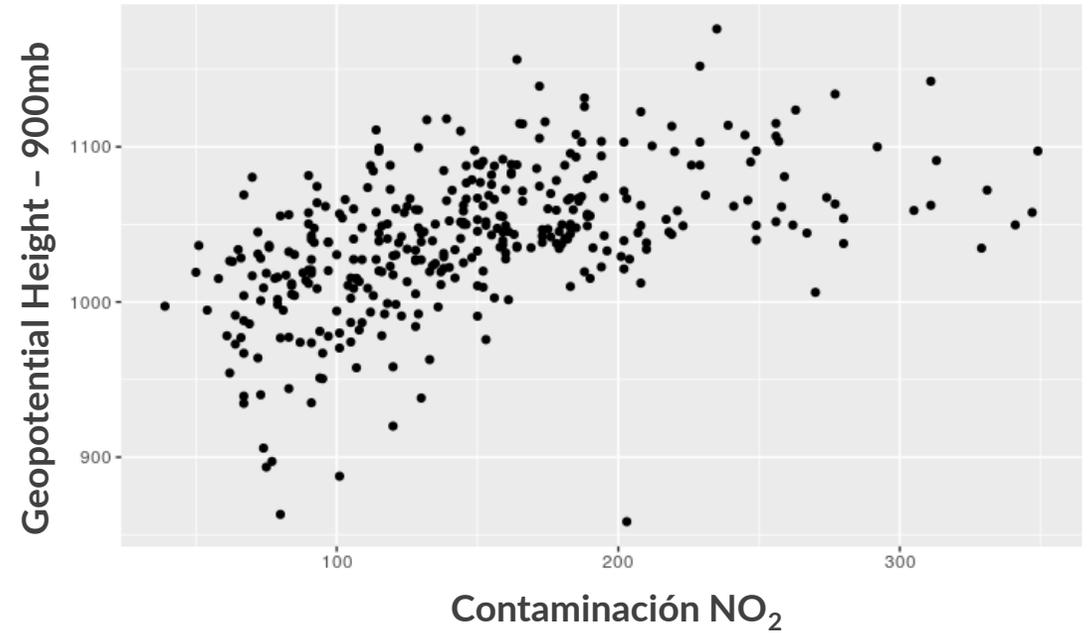
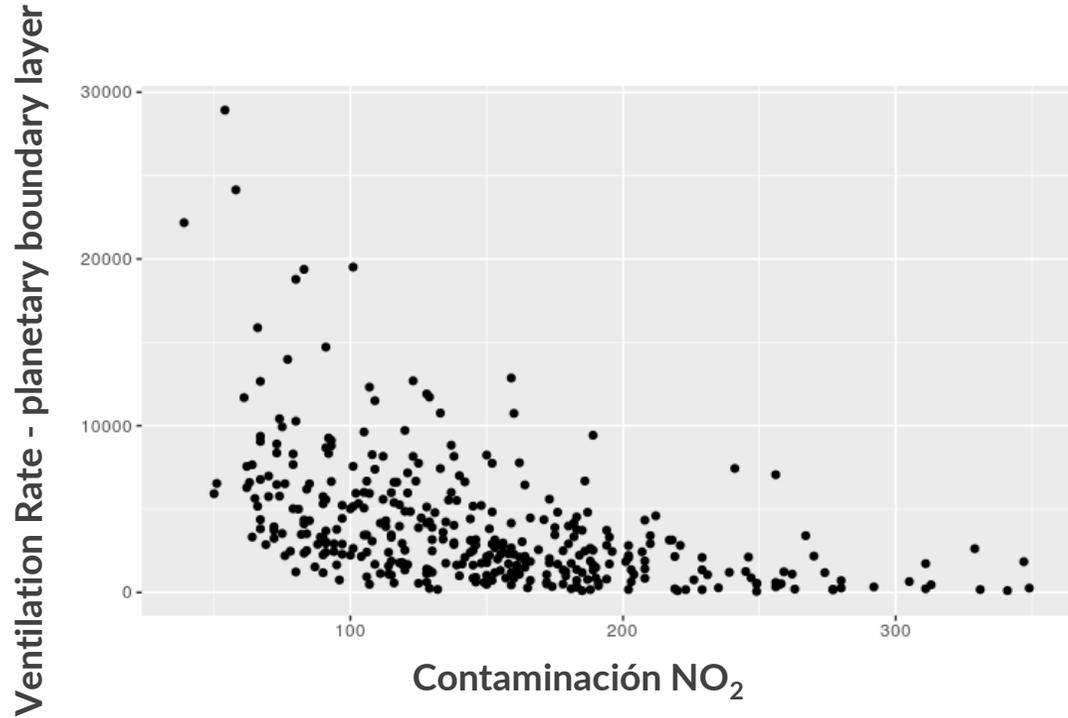


CONTAMINACIÓN

La tendencia de los niveles previos de contaminación sirve para anticipar futuros valores, incluso de contaminantes diferentes. Por ejemplo: entre NO₂ y O₃ se producen **reacciones químicas** que alteran los niveles.



Parámetros meteorología



www.piperlab.es

@mariamedp

@piperlab_es

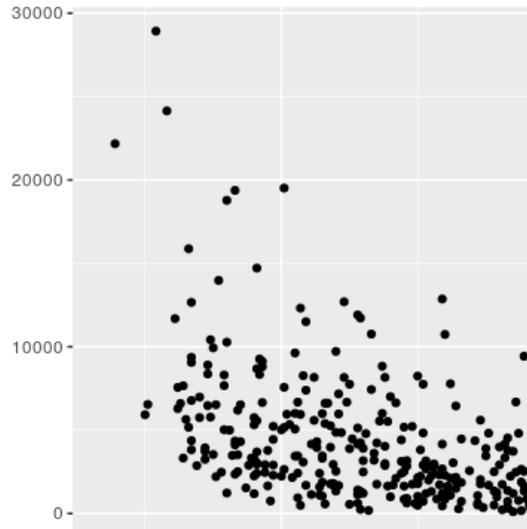




Parámetros meteorología



Ventilation Rate - planetary boundary layer



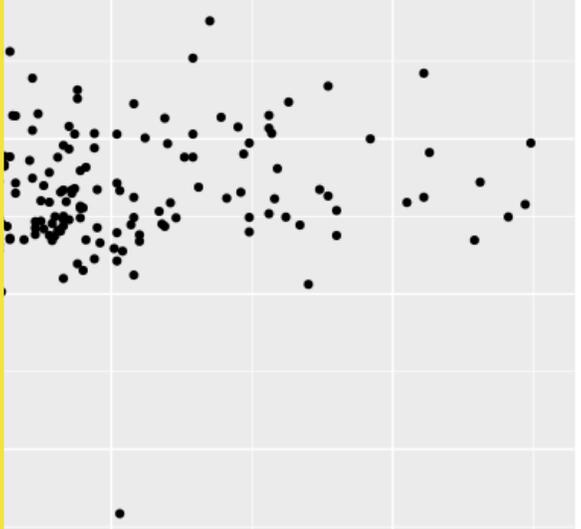
Contaminación NO₂

- Altura geopotencial
- Temperatura
- Humedad relativa
- Altura de la capa límite planetaria
- Velocidad del viento
- Ozono
- Presión
- Ventilación

Geopotential height - 900mb



Contaminación NO₂



www.piperlab.es





Metodología





Proyecto

Resultados



Horizonte	Precisión (accuracy)	Detección picos (recall)
24 horas	87.98%	90.24%
48 horas	87.36%	73.17%
7 días	77.97%	51.22%



#UOCDataDay

¡Gracias!



PIPERLAB

piperlab.es



UOC