

# **INFORME DE DATA CURATION**

## **DE LA CIUDAD DE BARCELONA**

### **1. Fuentes de datos**

Los datos que se utilizan en la parte del proyecto orientada a la recolección de datos de la ciudad de Barcelona se han obtenido de diversos sensores colocados en la ciudad.

A continuación, se explican las fuentes utilizadas así como la estructura de los datos elegidos.

#### **1.1. Situación de estaciones de medición de la calidad del aire**

Se trata del conjunto de datos que contiene las localizaciones y datos de las estaciones de medición en Barcelona.

**Fuente:**

<http://opendata-ajuntament.barcelona.cat/data/es/dataset/qualitat-aire-estacions-bcn>

**Estructura:**

Nombre	Descripción
Nom_cabina	Nombre de la estación
Codi_dtes	Identificador de la estación
Zqa	Código de la zona
Codi_eoi	Código europeo de la cabina
Longitud	Longitud geográfica
Latitud	Latitud geográfica
Ubicació	Nombre de la calle
Codi_Districte	Código del distrito
Nom_Districte	Nombre del distrito
Codi_Barri	Código del barrio
Nom_Barri	Nombre del barrio
Ocupacio_sol	Tipo de suelo (Urbana, Suburbana ó Rural)
Emissions_properes	Emisiones dominantes en la zona (Tráfico ó Industrial)
Contaminant_1	Indicador contaminante NO <sub>2</sub>
Contaminant_2	Indicador contaminante O <sub>3</sub>
Contaminant_3	Indicador contaminante PM 10

#### **1.2. Mediciones de la calidad del aire**

Este conjunto de datos contiene las medidas realizadas por las estaciones en tiempo real.

**Fuente:**

<http://opendata-ajuntament.barcelona.cat/data/es/dataset/qualitat-aire-detall-bcn>

#### Estructura:

Nombre	Descripción
Nom_cabina	Nombre de la estación que ha tomado la medida
Qualitat_aire	Calidad del aire (Buena, Regular o Pobre)
Codi_dtes	Identificador de la estación
Zqa	Código de la zona
Codi_eoi	Código europeo de la cabina que ha tomado la medida
Longitud	Longitud geográfica
Latitud	Latitud geográfica
Hora_o3	Hora de la medición de O3 (cada hora)
Qualitat_o3	Calidad del índice O3
Valor_o3	Valor de la medida O3
Hora_no2	Hora de la medición de NO2 (cada hora)
Qualitat_no2	Calidad del índice NO2 (Buena, Regular o Pobre)
Valor_no2	Valor de la medida de NO2
Hora_pm10	Hora de la medición de PM 10 (cada hora)
Qualitat_pm10	Calidad de las partículas en suspensión (Buena, Regular o Pobre)
Valor_pm10	Valor de la medida de PM 10
Generat	Fecha y hora de cuándo se ha generado la medida
DateTime	Timestamp de la hora de creación del fichero

### 1.3. Tramos de carretera

El conjunto de datos que se puede encontrar en el siguiente enlace contiene los detalles de cada tramo de carretera de Barcelona.

#### Fuente:

<http://opendata-ajuntament.barcelona.cat/data/ca/dataset/transit-relacio-trams/resource/036bfde0-b73e-4cb9-93db-5785f032ab68>

#### Estructura:

Nombre	Descripción
Tram	Identificador del tramo de medición
Descripción	Nombre del tramo
Coordenadas	Coordenadas de inicio y fin de cada tramo (lat0, lon0, lat1,lon1)

### 1.4. Mediciones del estado del tráfico

Todas las medidas realizadas por los sensores de tráfico con una frecuencia de 5 minutos son regidas en el conjunto de datos al que dirige el enlace que aparece más abajo.

#### Fuente:

<http://opendata-ajuntament.barcelona.cat/data/es/dataset/trams>

### Estructura:

Nombre	Descripción
IdTram	Identificador del tramo de medición
Data	Fecha de la medición
EstatActual	Estado actual (0=nada, 1=muy fluido, 2=fluido, 3=denso, 4=muy denso, 5=congestionado, 6=atasco)
EstatPrevist	Estado previsto en 15 minutos (0=nada, 1=muy fluido, 2=fluido, 3=denso, 4=muy denso, 5=congestionado, 6=atasco)

## 2. Proceso de curación de los datos

En este apartado, se describen los pasos realizados para la curación de los datos descritos en el primer apartado del informe, que se han realizado con el objetivo de unirlos para obtener, dada una estación de medición, la información del estado del aire y del estado del tráfico cercano.

Los pasos que se muestran a continuación se pueden ver resueltos en el Jupyter Notebook `dataCurationBarcelona.ipynb` recogido en el repositorio del proyecto:

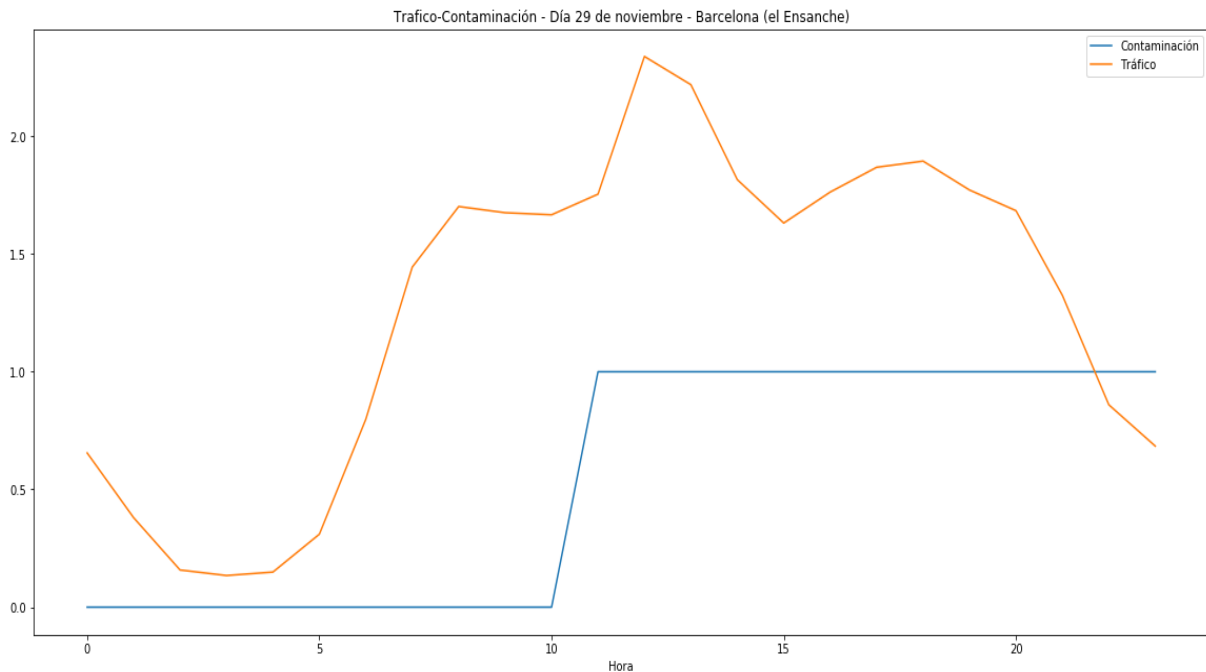
### URL

#### Pasos realizados durante la curación de los datos

1. Por medio de la librería Pandas, en Python, se han cargado los archivos .csv obtenidos del portal de datos en abierto de Barcelona.
2. Se han convertido las fechas en formato String a DateTime con el fin de poder hacer comparaciones y agrupaciones por fecha.
3. Se han mapeado los valores de calidad del aire de Strings a Ints con el objetivo de realizar cálculos numéricos sobre ellos.
4. Para los tramos de carretera, se ha añadido una nueva columna, PuntoMedio, la cual contiene el punto medio de las coordenadas de cada tramo.
5. En base al punto medio y a las coordenadas de las estaciones, se ha asignado a cada tramo una estación cercana. Es decir, se ha añadido una nueva columna llamada NearStation con el identificador de la estación a la que pertenece.
6. Se han agrupado los datos de calidad por estación, día y hora, de tal manera que se puede acceder a la media de calidad de aire horaria por cada estación.
7. Sabiendo el día y el identificador de la estación, se pueden consultar los datos del tráfico con media horaria.

### 3. Análisis de los resultados

En este caso, en el siguiente gráfico, se puede ver claramente una relación entre el tráfico y la calidad del aire. Sobre las seis de la mañana, el tráfico comienza a aumentar y unas horas más tarde, debido a la acumulación de gases y partículas contaminantes, empeora la calidad del aire.



*Ilustración 1. Comparación entre el tráfico y la calidad del aire durante el día 29 de noviembre en Barcelona.*