



# DOCUMENTACIÓN SISTEMA DE INTELIGENCIA DE NEGOCIO

Calidad del aire en la Comunidad de Madrid

## Descripción del documento

En este documento se redactan los distintos pasos seguidos en la creación de un sistema de Inteligencia de Negocio para analizar la calidad del aire en la Comunidad de Madrid. Este trabajo corresponde a la segunda Prueba de Evaluación Continua (PEC2) de la asignatura Presentación y Visualización.

**Autor: Alberto Fernández Santos**

Fecha: 21/03/25

Contenido

- 1    Introducción.....2
- 2    Sistema de Inteligencia de Negocio .....3
  - 2.1    Definición de objetivos .....3
  - 2.2    ETL .....3
    - 2.2.1    Extracción ..... 3
    - 2.2.2    Transformación ..... 5
    - 2.2.3    Carga..... 9
  - 2.3    Almacenamiento .....9
  - 2.4    Modelado.....9
  - 2.5    Visualización .....10
- 3    Análisis contaminación Comunidad de Madrid en 2024 .....15

# 1 Introducción

La Comunidad de Madrid dispone de un Sistema Integral de la Calidad del Aire que nos permite conocer en cada momento sus niveles de contaminación atmosférica. Este sistema de control se puso en funcionamiento en el año 2001 y desde ese momento, se han ido guardando todas las mediciones diarias generadas, por lo que actualmente se dispone de un dataset con los datos diarios por anualidades de 2001 a 2024.

En este ejercicio académico, se tomará el dataset proporcionado por la Comunidad de Madrid para desarrollar un sistema de Inteligencia de Negocio con el propósito de analizar la calidad del aire a lo largo de los últimos años.

Por lo que respecta al procedimiento para desarrollar el sistema de Inteligencia de Negocio, se seguirá el proceso mostrado en la Figura 1:

1. Definición de objetivos
2. ETL (Extracción, Transformación y Carga)
3. Almacenamiento
4. Modelado
5. Visualización

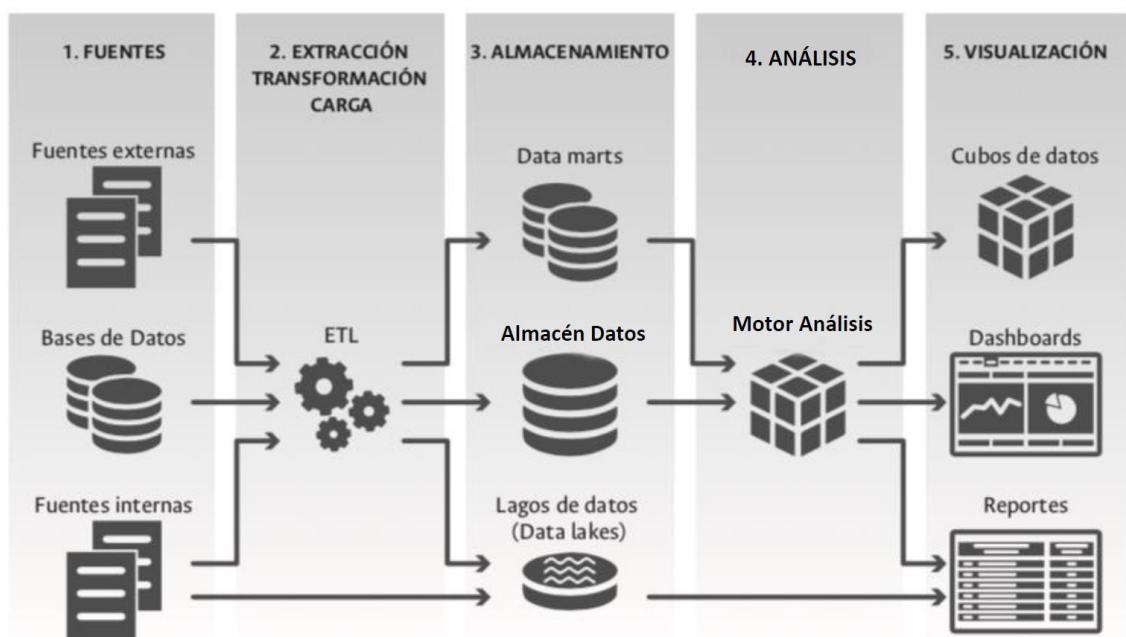


Figura 1. Adaptado de *Arquitectura solución de Inteligencia de Negocio* (p. 23), 2025, Presentación Introducción a la Inteligencia de Negocio de la asignatura Presentación y Visualización del Máster en Ciencia de Datos de UCJC.

Power BI será la herramienta empleada para el desarrollo del sistema de Inteligencia de Negocio en este proyecto.

## 2 Sistema de Inteligencia de Negocio

### 2.1 Definición de objetivos

#### Objetivos

##### Objetivo general

El proyecto consiste en desarrollar un sistema de Inteligencia de Negocio para analizar la calidad del aire en Madrid desde 2001 hasta 2024, utilizando Power BI.

##### Objetivos específicos

- Realizar todas las fases de desarrollo de un sistema de Inteligencia de Negocio mostradas en la Figura 1, culminando en la creación de un dashboard con la información clave de la calidad del aire.
- Analizar la contaminación en Madrid en el año 2024 creabdo un Data-Storytelling para comunicar los resultados mediante una presentación.
- Entregables:
  - Archivo comprimido compuesto por:
    - Cuaderno de bitácora con los pasos seguidos para crear el sistema (este documento).
    - Fichero del dashboard con formato .pbix creado en Power BI.
    - Vídeo con la presentación del análisis y la historia de los resultados.
- Cronograma:
  - Fecha límite de entrega: 23/03/2025

### 2.2 ETL

#### 2.2.1 Extracción

Nos apoyamos de la herramienta *Obtener datos* de Power BI para extraer las bases de datos directamente de las páginas webs de la Comunidad de Madrid.

Bases de datos:

- Datos de calidad del aire (una tabla por año desde 2001 a 2024):  
<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=aecb88a7e2b73410VgnVCM2000000c205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>
- Listado de Contaminantes:  
[https://datos.madrid.es/FWProjects/egob/Catalogo/MedioAmbiente/Aire/Ficheros/Interprete\\_ficheros\\_%20calidad\\_%20del\\_%20aire\\_global.pdf](https://datos.madrid.es/FWProjects/egob/Catalogo/MedioAmbiente/Aire/Ficheros/Interprete_ficheros_%20calidad_%20del_%20aire_global.pdf)
- Estaciones de control:  
<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=9e42c176313eb410VgnVCM1000000b205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>

Gracias a los conectores proporcionados por Power BI, no es necesario descargar en local las bases de datos para posteriormente importarlas en Power BI. Además, al importarlas

directamente desde la dirección URL, podemos actualizar las bases de datos de forma automática y periódicamente.

**Nota.** El separador decimal empleado en las BBDD es el “.”. Para evitar errores en la extracción de los datos, se cambia la configuración regional del archivo .pbix de Power BI sobre el que vamos a desarrollar todo el sistema de Inteligencia de Negocio.

Para cambiar la configuración regional del archivo, basta con ir a *Archivo/Opciones y configuración/Opciones/ARCHIVO ACTUAL/Configuración regional* y seleccionar *Inglés (Estados Unidos)*:

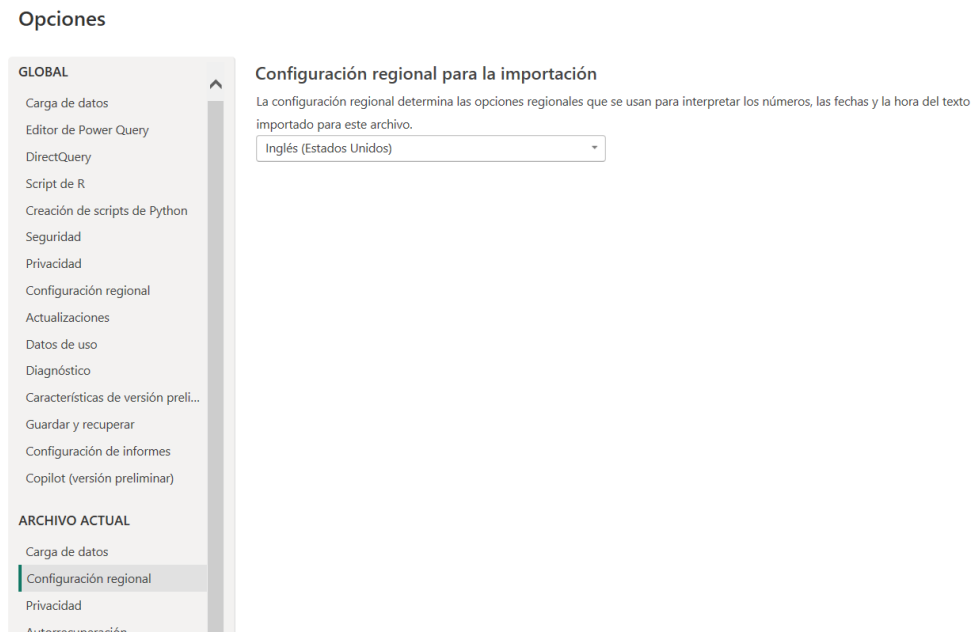


Figura 2. Modificación de Configuración regional del archivo actual para tomar "." como separador decimal.

Una vez extraídas todas las bases de datos mediante el conector web, ya tenemos disponibles dichas bases de datos en Power BI:

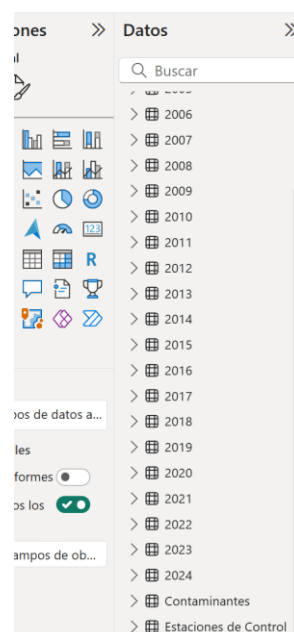


Figura 3. BBDD de la Comunidad de Madrid importadas en Power BI.

### 2.2.2 Transformación

Para la fase de transformación de los datos, nos apoyaremos de la herramienta de *Transformar datos* (Power Query) proveída por Power BI.

A continuación, se detallan las transformaciones realizadas sobre las bases de datos:

#### Tabla Calidad del aire

- Anexar tablas de calidad del aire separadas en los distintos años en una única tabla mediante la herramienta *Combinar/Anexar consultas para crear una nueva*:



Figura 4. Proceso de anexar tablas de calidad del aire en una única tabla nueva

Una vez realizado este procedimiento, obtenemos una nueva tabla compuesta por las 69 columnas correspondientes a las tablas de calidad del aire en los distintos años y un total de 34,769 filas. Cada fila corresponde a las muestras tomadas de un contaminante durante todos los días de un mes y año concreto.

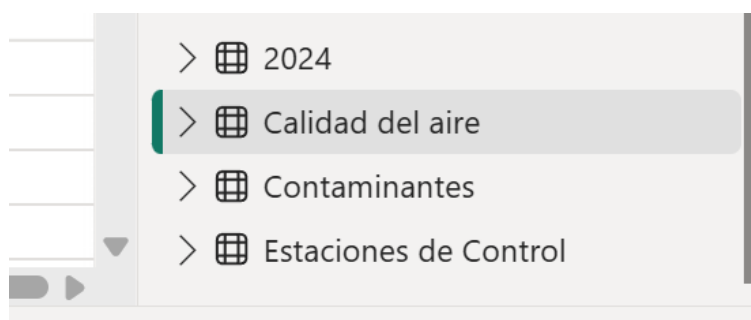


Figura 5. Nueva tabla creada con todas las muestras de calidad del aire desde 2001 a 2024.

- Sustituir por null todos los valores de las columnas Ds cuyo valor de la derecha (columnas Vs) sea "N". El valor "N" en las columnas Vs indica que la medición no se ha realizado o no es válida.

A continuación, se muestra el código empleado en M para el remplazo de valores:

let

```
Origen = Table.Combine({#"2024", #"2023", #"2022", #"2021", #"2020", #"2019",  
#"2018", #"2017", #"2016", #"2015", #"2014", #"2013", #"2012", #"2011", #"2010",  
#"2009", #"2008", #"2007", #"2006", #"2005", #"2004", #"2003", #"2002", #"2001"}),
```

```
// Definir las columnas D01, D02, ..., D31
ColumnasD = List.Transform({1..31}, each "D" & Text.PadStart(Text.From(_), 2, "0")),

// Reemplazar los valores de las columnas D01, D02, ..., D31 en función de las
columnas V01, V02, ..., V31
#"Valores reemplazados" = List.Accumulate(ColumnasD, Origen, (estado, columnaD)
=>
    let
        // Obtener la columna correspondiente en el grupo V (por ejemplo, V01 para
        D01, V02 para D02, etc.)
        columnaV = "V" & Text.End(columnaD, 2),

        // Reemplazar valores en la columna D según el valor de la columna V
        reemplazo = Table.ReplaceValue(
            estado,
            each Record.Field(_, columnaD),
            each if Record.Field(_, columnaV) = "N" then null else Record.Field(_,
columnaD),
            Replacer.ReplaceValue,
            {columnaD}
        )
    in
        reemplazo
)
in
    #"Valores reemplazados"
```

- Pivotar columnas para que cada fila corresponda a un único día. Podemos pivotar la tabla fácilmente seleccionando todas las columnas Ds y Vs y haciendo click en la herramienta *Transformar/Anular dinamización de columnas*.

	PUNTO_MUESTREO	AÑO	MES	DIA	Valor
1	28079004_6_48	2024	7	D01	0,5
2	28079004_6_48	2024	7	V01	V
3	28079004_6_48	2024	7	D02	0,6
4	28079004_6_48	2024	7	V02	V
5	28079004_6_48	2024	7	D03	0,4
6	28079004_6_48	2024	7	V03	V
7	28079004_6_48	2024	7	D04	0,4
8	28079004_6_48	2024	7	V04	V
9	28079004_6_48	2024	7	D05	0,3
10	28079004_6_48	2024	7	V05	V
11	28079004_6_48	2024	7	D06	0,4
12	28079004_6_48	2024	7	V06	V
13	28079004_6_48	2024	7	D07	0,5

**Configuración de la consulta**

**PROPIEDADES**

Nombre: Calidad del aire

Todas las propiedades

**PASOS APLICADOS**

- Origen
- ColumnasD
- Valores reemplazados
- Columna de anulación de din...
- Columnas con nombre cambi...

Figura 6. Tabla Calidad del aire pivotada.

- Filtrar columna DIA por valores que empiecen por “D”. De este modo, ocultamos los valores correspondientes a las columnas Vs en las visualizaciones.

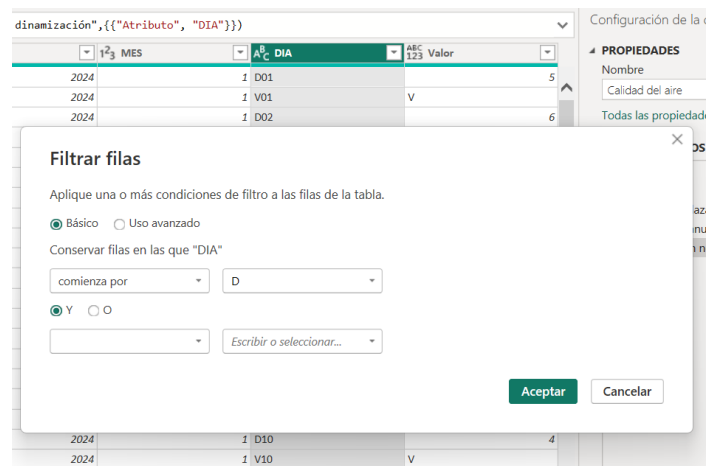


Figura 7. Filtrado de columna DIA

- Reemplazar valores y cambiar tipo a los valores de la columna DIA:
  - Texto que empiezan por "D0" a ""
  - Texto que empiezan por "D" a ""
  - Cambiar tipo de la columna de texto a valor numérico entero

The screenshot shows a data table with columns: PUNTO\_MUESTREO, ANO, MES, DIA, and Valor. The data includes rows for the year 2024 with various DIA values (D01, V01, D02, etc.). The 'DIA' column now contains empty strings. The 'Valor' column contains numerical values. The table is filtered to show only rows where 'DIA' is empty. The 'Configuración de la consulta' (Query configuration) panel on the right shows the 'PASOS APLICADOS' (Applied steps) section, which includes 'Origen' (Origin), 'ColumnasD' (ColumnsD), 'Valores reemplazados' (Replaced values), 'Columna de anulación de din...' (Cancellation column of din...), 'Columnas con nombre cambi...' (Columns with changed name...), 'Filas filtradas' (Filtered rows), 'Valor reemplazado' (Replaced value), 'Valor reemplazado1' (Replaced value1), and 'Tipo cambiado' (Type changed).

Figura 8. Reemplazo de valores y cambio de tipo a número entero en la columna DIA.

Una vez realizado este paso, ya tenemos preparada la columna día en formato numérico de tipo entero.

- Cambiar a orden ascendente en las columnas ANO, MES y DIA.  
De este modo, ya tenemos la tabla *Calidad del aire* ordenada en orden cronológico.

The screenshot shows a data table with columns: PUNTO\_MUESTREO, ANO, MES, DIA, and Valor. The data is sorted by ANO, MES, and DIA in ascending order. The 'DIA' column now contains numerical values. The 'Valor' column contains numerical values. The table is filtered to show only rows where 'DIA' is empty. The 'Configuración de la consulta' (Query configuration) panel on the right shows the 'PASOS APLICADOS' (Applied steps) section, which includes 'Origen' (Origin), 'ColumnasD' (ColumnsD), 'Valores reemplazados' (Replaced values), 'Columna de anulación de din...' (Cancellation column of din...), 'Columnas con nombre cambi...' (Columns with changed name...), 'Filas filtradas' (Filtered rows), 'Valor reemplazado' (Replaced value), 'Valor reemplazado1' (Replaced value1), 'Tipo cambiado' (Type changed), and 'Filas ordenadas' (Sorted rows).

Figura 9. Orden ascendente en columnas ANO, MES y DIA



- Cambiar tipo a valor numérico decimal en la columna Valor (valor de medida realizada de un contaminante en un día concreto):

	PUNTO_MUESTREO	ANO	MES	DIA	1.2 Valor
1	28079040_12_8	2001	1	29	67
2	28079008_1_38	2001	1	5	10
3	28079008_12_8	2001	1	16	211
4	28079040_12_8	2001	1	28	33
5	28079008_7_8	2001	1	4	99
6	28079008_1_38	2001	1	18	11
7	28079004_6_48	2001	1	18	0,6
8	28079011_7_8	2001	1	4	20
9	28079011_7_8	2001	1	3	22
10	28079018_6_48	2001	1	9	2,4
11	28079040_12_8	2001	1	27	77
12	28079008_20_59	2001	1	19	19,7
13	28079008_7_8	2001	1	11	95
14	28079008_1_38	2001	1	2	10
15	28079008_12_8	2001	1	14	354
16	28079004_8_8	2001	1	16	59

Figura 10. Cambio de tipo a valor numérico decimal en la columna Valor

- Se añade una nueva columna con la fecha de cada medición en formato dd/mm/yyyy.

	ANO	MES	DIA	FECHA	1.2 Valor
1	2001	1	29	29/01/2001	
2	2001	1	5	05/01/2001	
3	2001	1	16	16/01/2001	
4	2001	1	28	28/01/2001	
5	2001	1	4	04/01/2001	
6	2001	1	18	18/01/2001	
7	2001	1	18	18/01/2001	
8	2001	1	4	04/01/2001	
9	2001	1	3	03/01/2001	
10	2001	1	9	09/01/2001	
11	2001	1	27	27/01/2001	
12	2001	1	19	19/01/2001	
13	2001	1	11	11/01/2001	
14	2001	1	2	02/01/2001	
15	2001	1	14	14/01/2001	
16	2001	1	16	16/01/2001	
17	2001	1	20	20/01/2001	
18	2001	1	2	02/01/2001	
19	2001	1	22	22/01/2001	

Figura 11. Nueva columna Fecha en la Tabla Calidad del aire.

Una vez realizadas todas las transformaciones, la tabla de *Calidad del aire* está formada por 10 columnas y 1,038,236 filas:

PROVINCIA	MUNICIPIO	Estación	MAGNITUD	PUNTO_MUESTREO	Año	Mes	Día	Valor	Fecha
28	79	38	8	28079038_8_8	2001	4	15	47	domingo, 15 de abril de 2001
28	79	16	8	28079016_8_8	2001	5	22	22	martes, 22 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	23	25	miércoles, 23 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	24	22	jueves, 24 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	25	24	viernes, 25 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	26	20	sábado, 26 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	27	14	domingo, 27 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	28	32	lunes, 28 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	29	26	martes, 29 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	30	32	miércoles, 30 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	31	37	jueves, 31 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	4	23	viernes, 4 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	5	15	sábado, 5 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	6	9	domingo, 6 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	7	23	lunes, 7 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	8	24	martes, 8 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	9	20	miércoles, 9 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	10	26	jueves, 10 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	11	38	viernes, 11 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	12	25	sábado, 12 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	13	10	domingo, 13 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	14	21	lunes, 14 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	15	12	martes, 15 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	16	19	miércoles, 16 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	17	17	jueves, 17 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	18	9	viernes, 18 de mayo de 2001
28	79	16	8	28079016_8_8	2001	5	19	17	sábado, 19 de mayo de 2001

Tabla: Calidad del aire (1.038.236 filas)

Figura 12. Estructura Tabla Calidad del aire tras realizar transformación de datos.

### 2.2.3 Carga

Debido a como se ha diseñado el proceso ETL en Power BI, la aplicación podrá extraer de forma directa los datos de las direcciones URL de la Comunidad de Madrid, a continuación realizará automáticamente el proceso de transformación de datos y finalmente, el propio archivo con extensión .pbix actuará como almacén de datos.

## 2.3 Almacenamiento

Tal y como se ha indicado al final del anterior punto, para este ejercicio académico, el propio archivo con extensión .pbix creado por Power BI se encargará de almacenar los datos una vez extraídos y transformados.

## 2.4 Modelado

El siguiente paso es ir a la *Vista de modelo* para realizar la fase de modelado. Es decir, establecer las relaciones entre las distintas tablas con el propósito de, posteriormente, visualizar y analizar los datos correctamente.

La Figura 13 muestra el modelo que relaciona las 3 tablas:

- Tabla *Calidad del aire*: actúa como tabla de hechos donde cada fila de la tabla se corresponde a la medición de un tipo de contaminante en un día concreto.
- Tabla *Contaminantes*: dimensión del modelo donde se registran todos los tipos de contaminantes y sus características.
  - Relaciones establecidas: columna *Cod.* de la tabla *Contaminantes* con la columna *MAGNITUD* de la tabla *Calidad del aire*.
- Tabla *Estaciones de control*: dimensión del modelo donde se registran las ubicaciones de las distintas estaciones donde se realizan las mediciones.
  - Relaciones establecidas: columna *CODIGO\_CORTO* de la tabla *Estaciones de control* con la columna *ESTACION* de la tabla *Calidad del aire*.

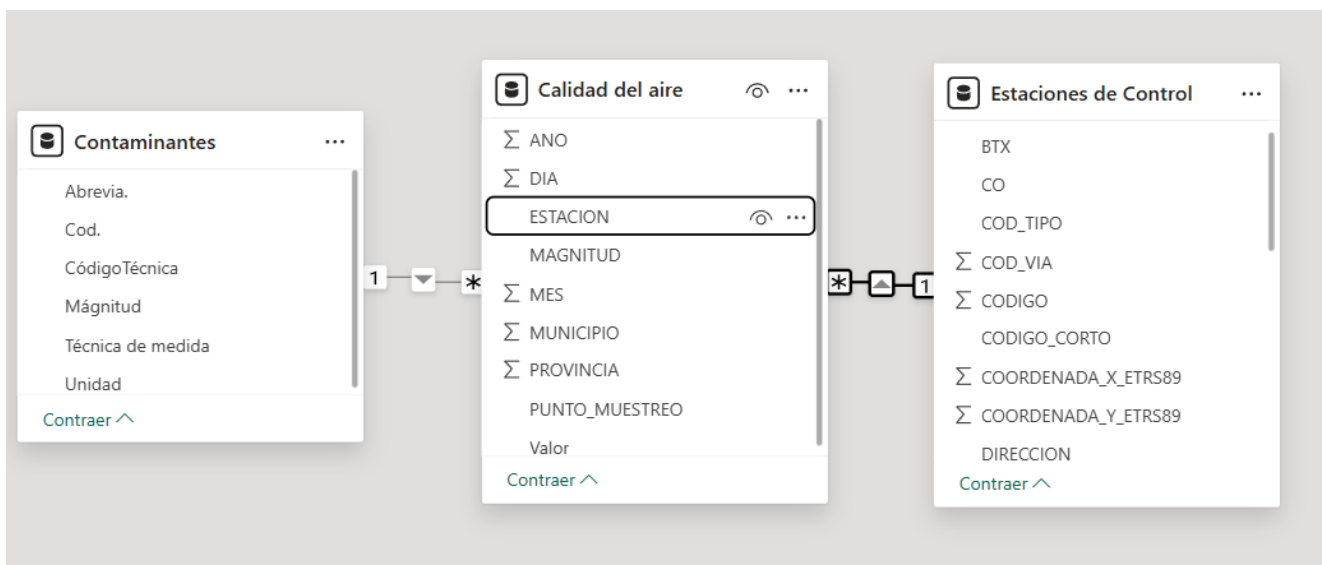


Figura 13. Relaciones establecidas entre las tres tablas.

## 2.5 Visualización

Una vez extraídos los datos, completado el proceso ETL y construido el modelo, llega el momento de visualizar la información con el objetivo de monitorear, analizar y facilitar la toma de decisiones de manera rápida y efectiva. Para esta etapa de visualización, utilizaremos el dashboard mostrado en la Figura 14.



Figura 14. Dashboard creado con Power BI para analizar la calidad del aire en la Comunidad de Madrid desde 2001 a 2024.

A continuación, se detalla las distintas visualizaciones y funcionalidades presentes en el Dashboard creado:

- **Segmentador de datos (filtros):** con el propósito de facilitar la interacción del usuario, el dashboard presenta una sección con los siguientes filtros:



Figura 15. Segmentadores de datos presentes en el dashboard.

- **Segmentador para selección de Estación de control:** desplegable con todas las Estaciones de control de la Comunidad de Madrid.

Figura 16. Configuración segmentador de Estaciones de control.

- Segmentador para el filtrado del Contaminante: desplegable con el listado de distintos compuestos químicos medidos.

Figura 17. Configuración segmentador de Contaminantes.

- Segmentador de fecha.

Figura 18. Configuración segmentador de fechas.

Al aplicar uno o varios de los filtros vistos, afectará a las visualizaciones que veremos a continuación.

- **Mapa localizador de las Estaciones de control**: en esta visualización se muestra las ubicaciones de las 24 estaciones de control actualmente activas en la Comunidad de Madrid.



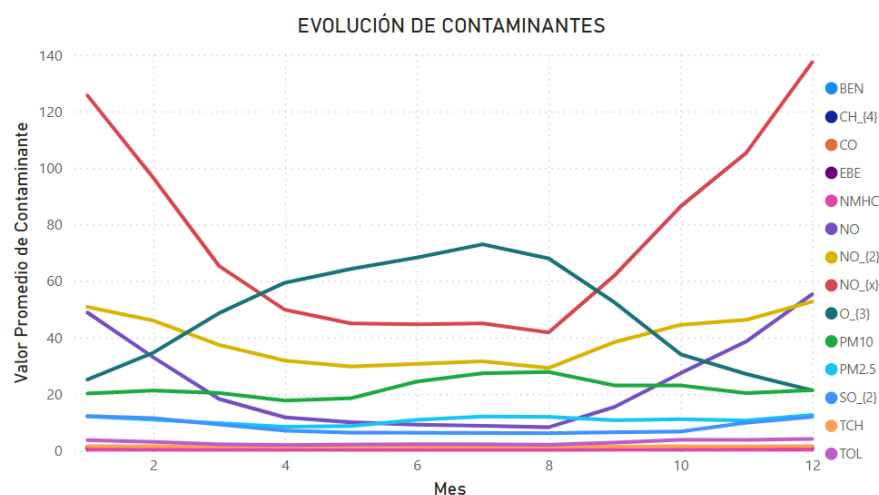


Figura 22. Evolución temporal de contaminantes con granularidad mensual.

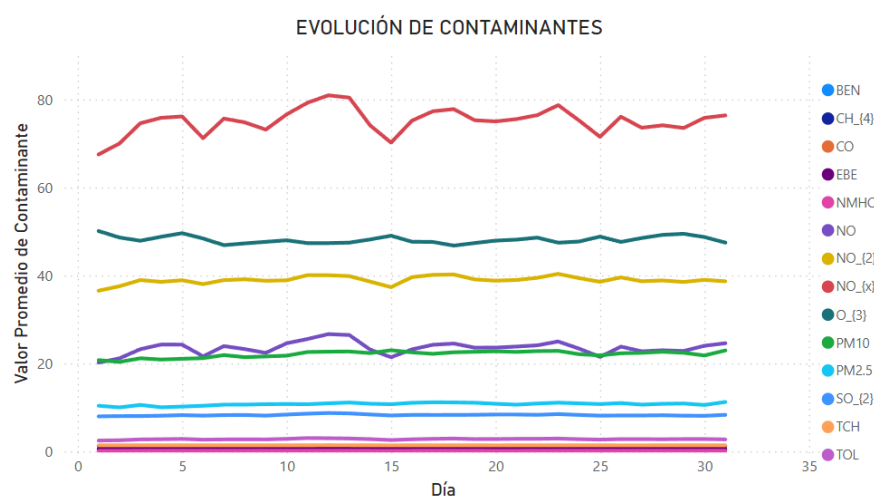


Figura 23. Evolución temporal de contaminantes con granularidad diaria.

Eje X	
Fecha Jerarquía	✓ ✕
Año	✕
Mes	✕
Día	✕
Eje Y	
Promedio de Valor	✓ ✕
Eje Y secundario	
Agregar campos de datos a...	
Leyenda	
Símbolo	✓ ✕

Figura 24. Configuración de la visualización del gráfico de líneas.

**Nota.** Los valores visualizados corresponden al promedio obtenido de las distintas muestras medidas por las estaciones de control y en la escala de tiempo especificada por el usuario.

- Gráfico de barras con la evolución de la cantidad de medidas tomadas:** en este gráfico de muestra la evolución temporal del número de muestras tomadas y al igual que en el anterior gráfico de líneas, el gráfico de barras está diseñado para que el usuario pueda cambiar la granularidad temporal según se muestra en las figuras Figura 25, Figura 26 y Figura 27.



Figura 25. Evolución temporal del número de medidas tomadas con granularidad anual.

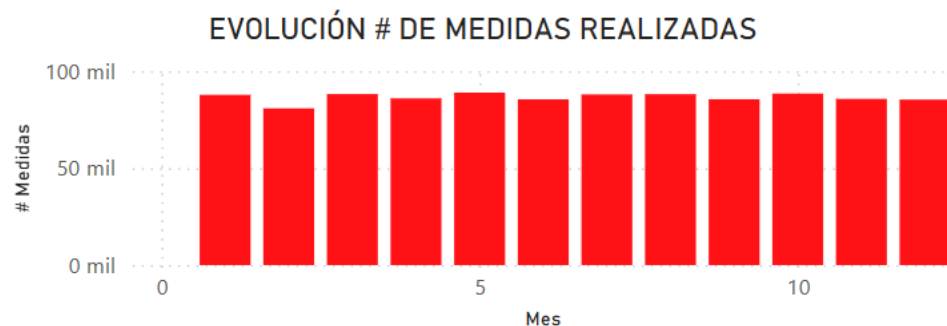


Figura 26. Evolución temporal del número de medidas tomadas con granularidad mensual.

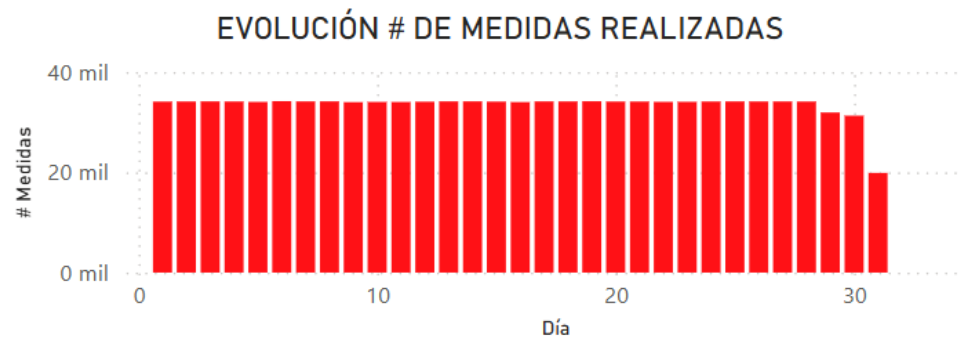


Figura 27. Evolución temporal del número de medidas tomadas con granularidad diaria.

Eje X	
Fecha Jerarquía	✓ ✕
Año	✕
Mes	✕
Día	✕
Eje Y	
Recuento de Valor	✓ ✕
Leyenda	
Agregar campos de datos a...	

Figura 28. Configuración del gráfico de barras.

- **Ventana flotante con información de las estaciones de control y contaminantes:** al situarnos sobre el gráfico de líneas, aparece una ventana flotante con información sobre las estaciones de control y los contaminantes visualizados, como su símbolo químico, el valor visualizado y la unidad de medida.

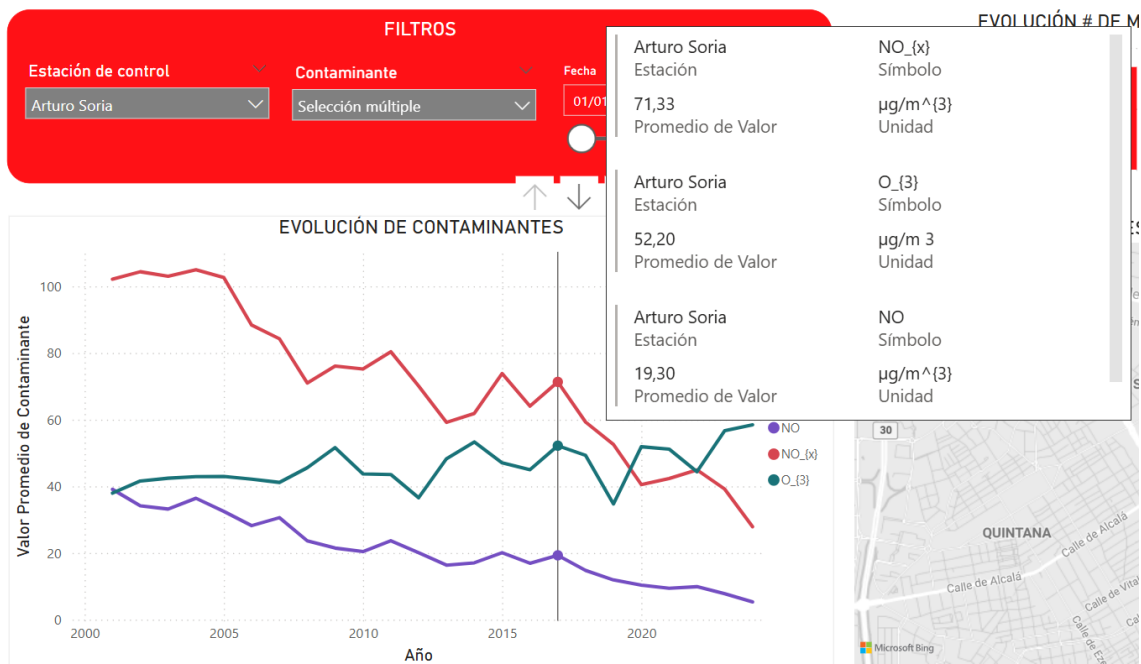


Figura 29. Ventana flotante en el gráfico de líneas.

### 3 Análisis contaminación Comunidad de Madrid en 2024

Para finalizar este proyecto, se realizará una presentación con el análisis detallado de la contaminación en la Comunidad de Madrid en el año 2024 (ver documento de presentación).

Con el propósito de profundizar en el análisis, nos apoyaremos de la base de datos meteorológicos de 2024 proporcionado por la Comunidad de Madrid en el siguiente enlace web: <https://datos.madrid.es/egob/catalogo/300351-15-meteorologicos-diarios.csv>

En esta BBDD adicional se pueden encontrar mediciones diarias de distintas magnitudes como:

- Velocidad y dirección del viento
- Temperatura
- Humedad relativa
- Presión barométrica
- Radiación solar
- Precipitación