



Dashboard - Datos de Calidad del Aire

Albert Daniel Llica Alvarez

Docente: Mag. Ana Maria Cuadros Valdivia

**UNSA - Universidad Nacional de San Agustín de Arequipa
Junio de 2025**

Índice

1. Introducción	4
2. El Mapa y su Jerarquía de Ubicaciones	4
2.1. Descripción del Mapa	4
2.2. Jerarquía de Ubicaciones	4
2.3. Leyenda Asociada al Mapa	7
3. El Dashboard Interactivo	7
3.1. Serie Temporal y Selección de Contaminantes	7
3.2. Valores Nulos y Atípicos	8
3.3. Mapa de Correlación	9
3.4. Selección de Intervalos y Boxplots	10
3.5. Descripción de Contaminantes: Valores Nulos y Atípicos	11

Índice de figuras

1.	Mapa-ciudades	5
2.	Mapa-Distritos	6
3.	Mapa-Estaciones	7
4.	Serie Temporal	8
5.	Valor Nulo y Atípico	9
6.	Matriz de Correlación	10
7.	Blospot	11
9.	Decripcion Atipico	12
8.	Descripcion Nulos	12

1. Introducción

Presentando una herramienta avanzada, este informe describe el dashboard interactivo desarrollado para analizar la calidad del aire en China. Diseñado con tecnologías como D3.js y SVG, el sistema permite explorar datos geográficos y temporales de manera intuitiva, destacando la distribución de contaminantes en diferentes ubicaciones y su evolución a lo largo del tiempo. El objetivo es proporcionar a los usuarios, como investigadores y responsables de políticas públicas, una plataforma para tomar decisiones informadas basadas en datos.

2. El Mapa y su Jerarquía de Ubicaciones

2.1. Descripción del Mapa

Construyendo una representación geográfica, el mapa interactivo utiliza el archivo `china.json` para dibujar las provincias de China, empleando una proyección Mercator centrada en las coordenadas $[104, 35]$ con una escala inicial de 500. Este mapa sirve como la interfaz principal para explorar las ubicaciones de monitoreo de calidad del aire, permitiendo a los usuarios interactuar mediante zoom y desplazamiento con un rango de escala de $[0, 5, 40]$.

2.2. Jerarquía de Ubicaciones

Organizando los datos en una estructura jerárquica, el mapa presenta tres niveles de ubicaciones:

- **Ciudades:** Representando el nivel superior, las ciudades se visualizan como puntos en el mapa, coloreadas según su clúster de pertenencia (`cluster_id`). Los datos se cargan desde `city.csv`, que incluye identificadores (`city_id`), coordenadas (`latitude`, `longitude`) y clústeres. Al pasar el mouse sobre una ciudad, se muestra un tooltip con información como su nombre en inglés y chino, coordenadas y clúster. Al hacer clic, el mapa realiza un zoom (escala 8) y transiciona al nivel de distritos.

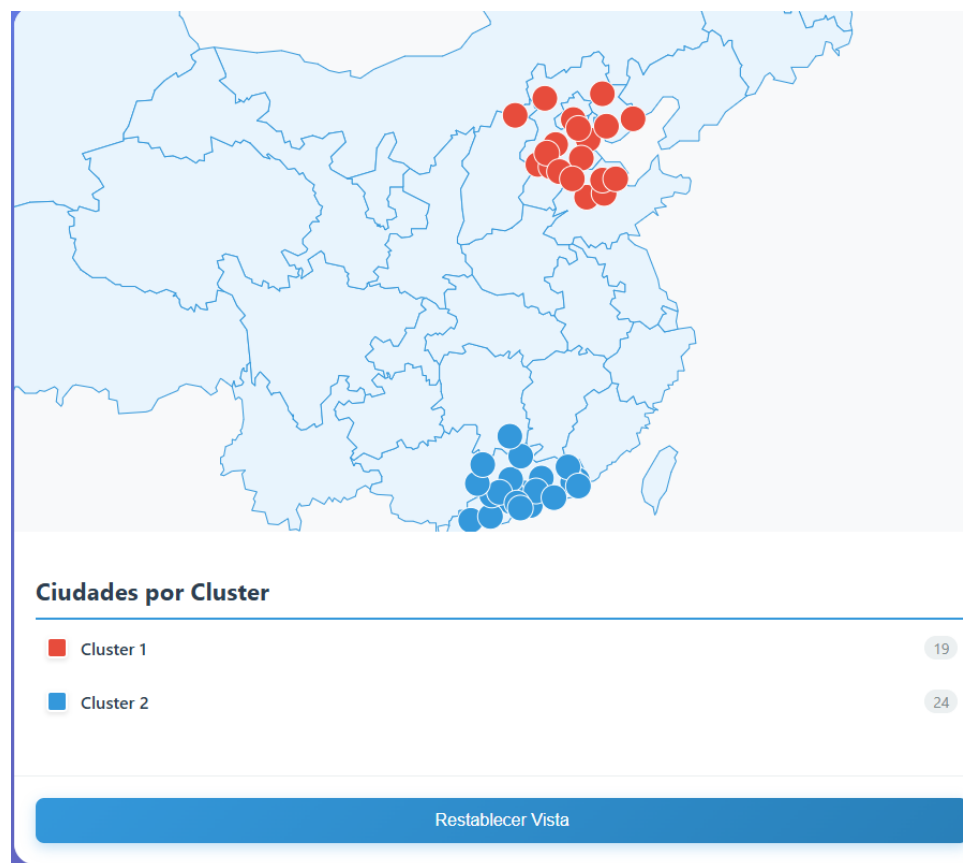


Figura 1: Mapa-ciudades

- **Distritos:** Corresponiendo al nivel intermedio, los distritos se cargan desde `district.csv` y están asociados a una ciudad mediante `city_id`. Se representan como puntos más pequeños, coloreados de manera única dentro de cada ciudad usando una escala de colores (`d3.schemeCategory10`). Al interactuar, los usuarios pueden ver un tooltip con el nombre del distrito e información adicional, y al hacer clic, el mapa realiza un zoom más profundo (escala 16) para mostrar las estaciones de monitoreo.

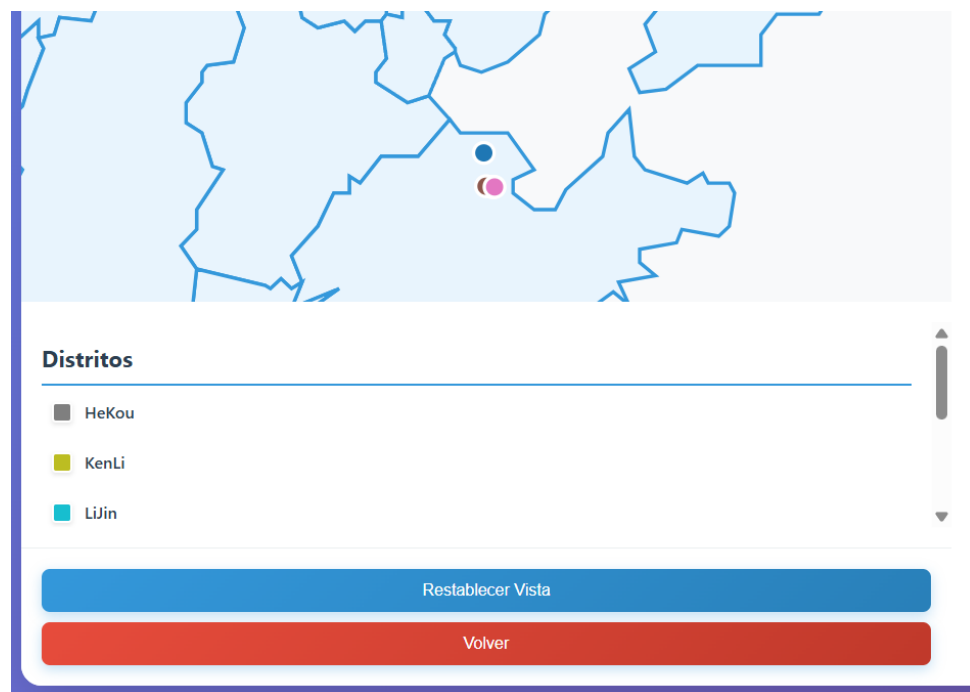


Figura 2: Mapa-Distritos

- **Estaciones de Monitoreo:** Constituyendo el nivel más bajo, las estaciones se cargan desde `station.csv`, con datos como `station_id`, `latitude`, `longitudo` y `district_id`. Se dibujan como círculos pequeños con colores únicos (`d3.schemeTableau10`) dentro de cada distrito. Al pasar el mouse, un tooltip muestra detalles como el nombre de la estación y sus coordenadas, y al hacer clic, se genera una serie temporal para esa estación con un zoom adicional (escala 25).

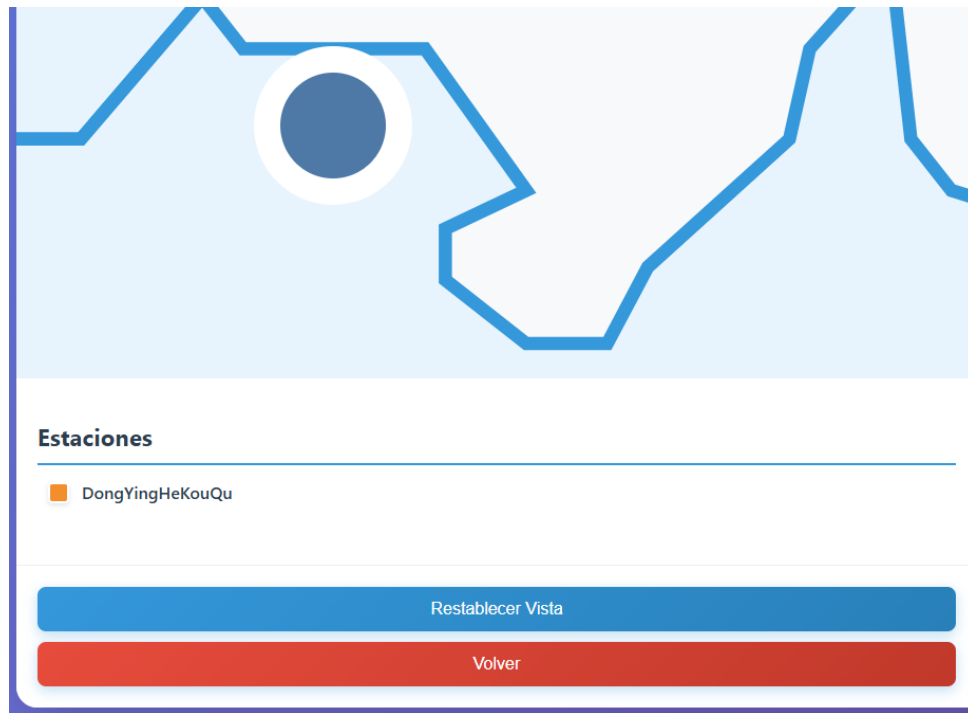


Figura 3: Mapa-Estaciones

2.3. Leyenda Asociada al Mapa

Complementando la visualización, la leyenda se actualiza dinámicamente según el nivel de vista:

- En el nivel de ciudades, muestra los clústeres con colores y la cantidad de ciudades por clúster.
- En el nivel de distritos, lista los distritos de la ciudad seleccionada con sus colores correspondientes.
- En el nivel de estaciones, enumera las estaciones de monitoreo con sus identificadores y colores.

3. El Dashboard Interactivo

3.1. Serie Temporal y Selección de Contaminantes

Generando visualizaciones temporales, la sección de series temporales se activa al seleccionar una estación de monitoreo. Los datos de calidad del aire, cargados desde `air-quality.csv`, incluyen mediciones de seis contaminantes principales: PM2.5, PM10, NO2,

CO, O₃ y SO₂, junto con un timestamp (`time`). Cada contaminante se representa con un color distintivo:

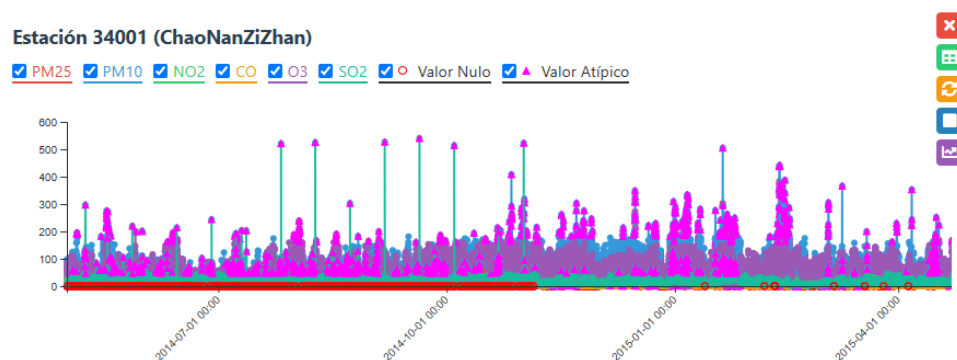


Figura 4: Serie Temporal

Mostrando las series temporales, los gráficos de líneas permiten a los usuarios:

- Visualizar la evolución de cada contaminante a lo largo del tiempo.
- Seleccionar contaminantes específicos mediante casillas de verificación en la leyenda, permitiendo ocultar o mostrar líneas individuales.

3.2. Valores Nulos y Atípicos

Detectando anomalías en los datos, el dashboard identifica valores nulos y atípicos:

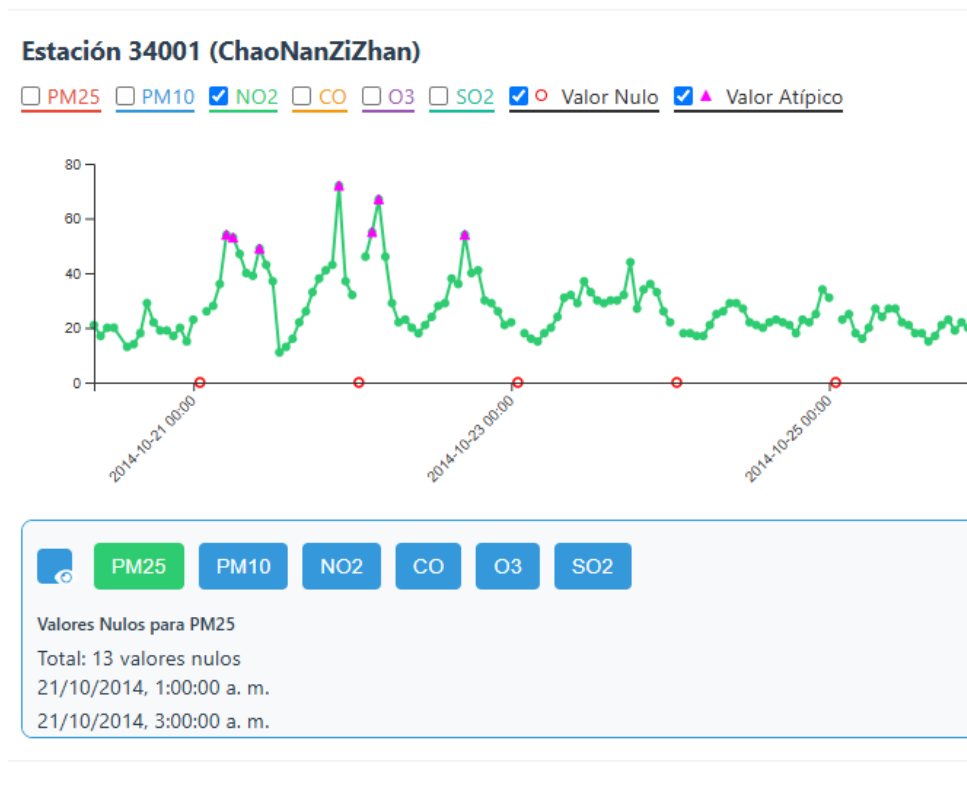


Figura 5: Valor Nulo y Atípico

- **Valores Nulos:** Representados con círculos rojos en la base del gráfico ($y=0$), los valores nulos se pueden ocultar o mostrar mediante una casilla en la leyenda. Al pasar el mouse, el tooltip indica "Valor Nulo" junto con la fecha.
- **Valores Atípicos:** Calculados usando el método del rango intercuartílico (IQR), se identifican como valores fuera del rango $[Q1 - 1,5 \times IQR, Q3 + 1,5 \times IQR]$. Se marcan con triángulos magenta y, al pasar el mouse, el tooltip muestra el valor y la fecha, con una nota "(Atípico)." en color magenta.

3.3. Mapa de Correlación

Creando un análisis estadístico, el dashboard genera una matriz de correlación al hacer clic en el botón correspondiente. Utilizando el coeficiente de correlación de Pearson, se calculan las relaciones entre los contaminantes:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y}$$

donde x_i y y_i son los valores de dos contaminantes, \bar{x} y \bar{y} sus medias, σ_x y σ_y sus desviaciones estándar, y n el número de observaciones.

Visualizando la matriz, el mapa de calor usa una escala de colores (`interpolateRdBu`) que va de rojo ($r = -1$) a azul ($r = 1$), con blanco en $r = 0$. Al pasar el mouse sobre

cada celda, un tooltip muestra el valor de correlación entre los dos contaminantes correspondientes.

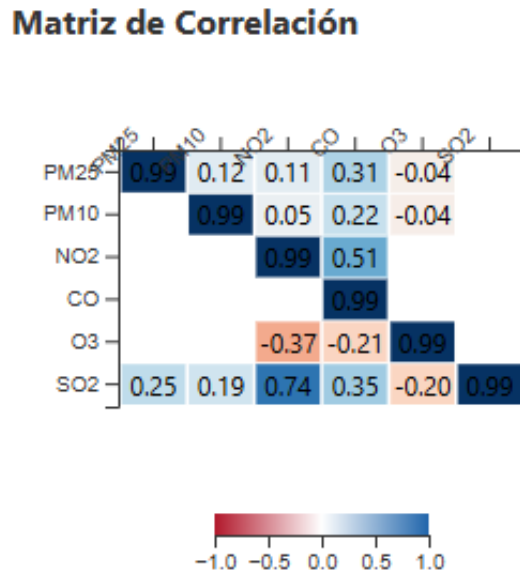


Figura 6: Matriz de Correlación

3.4. Selección de Intervalos y Boxplots

Proporcionando herramientas de filtrado, el dashboard permite:

- **Selección de Intervalos:** Activando el modo de brushing (*class="fas fa-square"*), los usuarios pueden seleccionar un rango de tiempo en la serie temporal. Los datos se filtran para mostrar solo los valores dentro del intervalo seleccionado, actualizando también el mapa de correlación y el boxplot.
- **Boxplots:** Alternando la vista con el botón (*class="fas fa-box"*), se genera un boxplot para cada contaminante seleccionado. Los boxplots muestran:
 - Mediana, cuartiles (Q1, Q3), y bigotes (mínimo y máximo dentro del rango IQR).
 - Valores atípicos como puntos individuales, con tooltips que indican su valor.

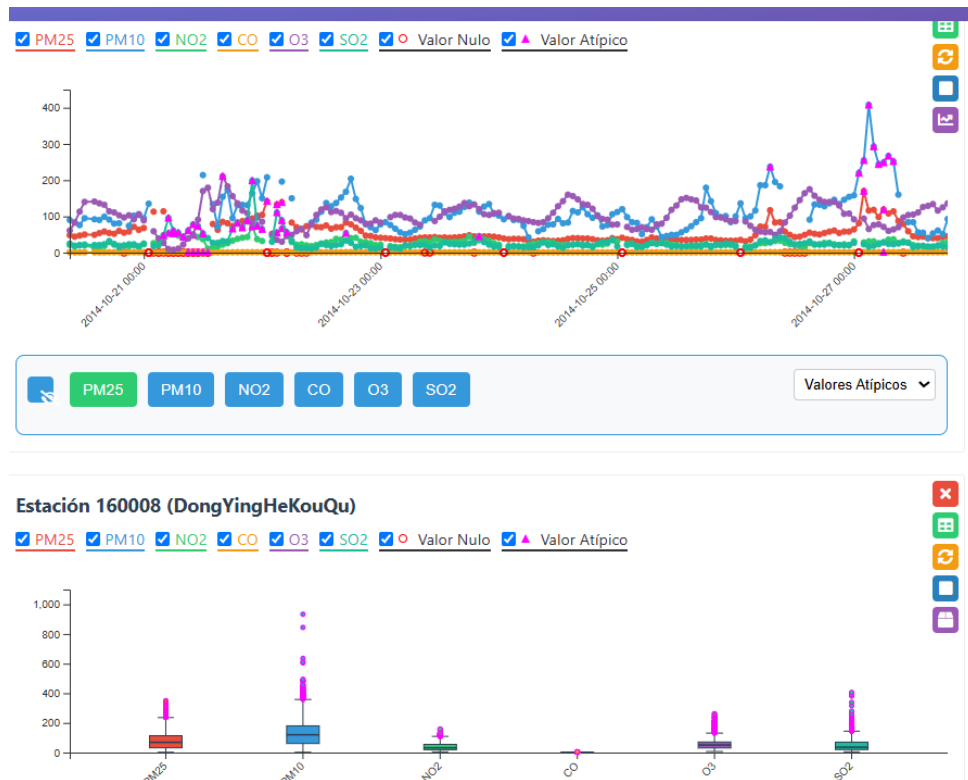


Figura 7: Bloxpot

3.5. Descripción de Contaminantes: Valores Nulos y Atípicos

Analizando los datos de cada estación, el dashboard incluye un cuadro de información que detalla:

- **Valores Nulos:** Para cada contaminante, se lista el total de valores nulos y las fechas específicas en las que ocurren.
- **Valores Atípicos:** Se enumera el total de valores atípicos por contaminante, junto con las fechas y valores específicos, permitiendo identificar patrones de anomalías.

Ejemplo de resultado para una estación:

- **PM2.5:** 5 valores nulos (fechas: 2025-01-01 00:00, ...), 3 valores atípicos (2025-02-01 12:00: 150.23 $\mu\text{g}/\text{m}^3$, ...).
- **NO2:** 2 valores nulos (fechas: 2025-01-02 06:00, ...), 1 valor atípico (2025-03-01 18:00: 80.45 $\mu\text{g}/\text{m}^3$).

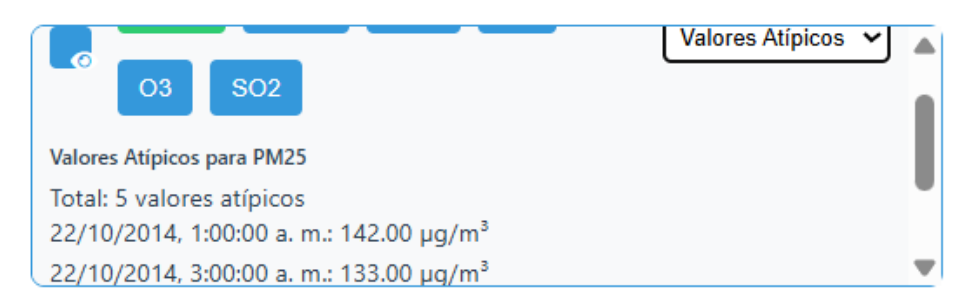


Figura 9: Descripción Atípico



Figura 8: Descripción Nulos

Anexo

- Código fuente: <https://github.com/AlbertLlica/Dashboard-TCD-AirQuality>