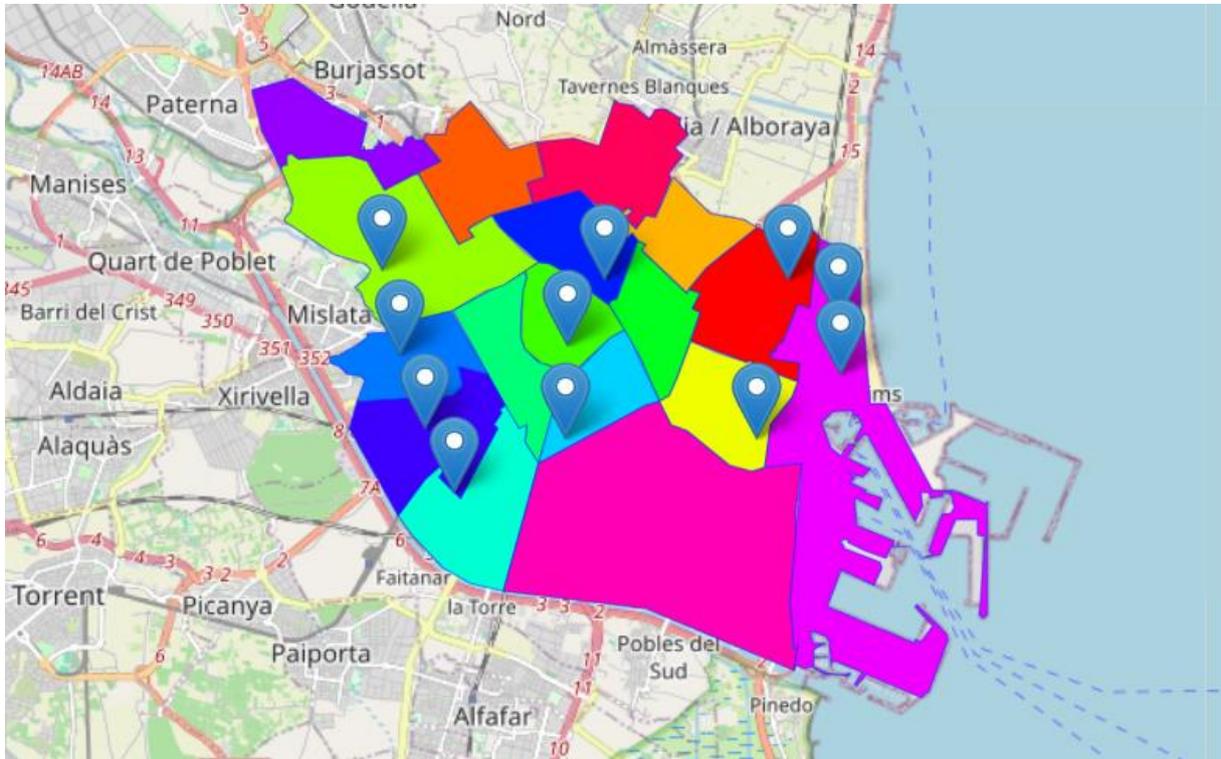


MINIPROYECTO : CONTAMINACIÓN EN VALENCIA

VISUALIZACIÓN DE LOS DATOS



2º Curso Grado de Ciencia de Datos

Autores:

Víctor Piquero Moreno

Víctor Álvarez Palomares

Carlos Heras Pardo

Índice

1. Introducción	3
2. Metodología	3
1. Análisis exploratorio de los datos	3
2. Preprocesado de Datos	4
3. Preprocesado de la información geográfica	5
4. Elección de gráficas para los distintos tipos de datos	6
5. Diseño del mapa e interactividad	7
6. Diseño del cuadro de mandos	8
7. Implementación	8
3. Resultados	9
4. Discusión	13
5. Conclusiones	15
6. Referencias	16

1. Introducción

La contaminación se ha convertido en uno de los mayores problemas de la sociedad. Conforme más avanzamos tecnológicamente, la contaminación se hace notable de más formas diferentes. Con el aumento de la urbanización, la industrialización y el crecimiento poblacional continuo, cada vez se disminuye nuestro nivel de sostenibilidad con el medio ambiente, lo cual tiene consecuencias negativas tanto para la salud, como para la biodiversidad y el clima.

En las ciudades vivimos constantemente rodeados de la contaminación del aire, causada en gran parte por vehículos de transporte, fábricas y plantas de energía. Todo esto produce enfermedades de todo tipo a las que nos sometemos día a día, producidos en parte por diferentes gases que se liberan, como por ejemplo el dióxido de carbono y el metano.

Necesitamos hacer frente a la situación que vivimos y que cada vez la estamos normalizando más, a lo que tenemos que poner fin. Necesitamos ejercer acciones para reducir la contaminación que no solo ayudarán a mejorar la calidad de vida, sino que también garantizarán un futuro sostenible para las futuras generaciones.

Para poder buscar soluciones, primero debemos conocer bien cuáles son las causas de un mayor nivel de contaminación en las ciudades y que medidas se pueden tomar para lidiar con ello. Por eso, en este proyecto buscamos medir en la ciudad de Valencia, los diferentes niveles de contaminación que se han vivido a lo largo de los años, para poder comprobar si con el tiempo hemos tomado medidas efectivas para poder solucionar el problema o si por el contrario cada vez nuestras acciones empeoran el ambiente.

También intentaremos ver cómo los diferentes vehículos y el tráfico en general hacen mella en la contaminación, para intentar ver si uno de nuestros principales objetivos para reducir la contaminación es restringir el nivel de tráfico en las ciudades.

Además, visualizamos si las zonas verdes ayudan a evitar los niveles de contaminación en las diferentes zonas de Valencia, intentando ver si por lo tanto, las zonas con una mayor proporción de zonas verdes son las que tienen un menor nivel de contaminación y si esta es una medida lo suficientemente positiva para tomar.

2. Metodología

1. Análisis exploratorio de los datos

Los datos abiertos del portal de datos de la concejalía de Gobierno Abierto que utilizaremos para visualizar la ciudad de Valencia con Objetivos de Desarrollo Sostenible son los siguientes:

- **Datos diarios calidad aire 2004-2022:** Estos son datos diarios de la calidad del aire de las estaciones de la red de vigilancia de la ciudad de Valencia. Las estaciones que se han utilizado para los datos son las siguientes: Avda. Francia, Bulevar Sur, Molino del Sol, Pista Silla, Politécnico, Viveros, Centro, Consellería Meteo, Nazaret Meteo, Puerto València. Los parámetros utilizados para medir la calidad del aire de estas

zonas han sido: PM1, PM2.5, PM10, NO, NO2, NOx, O3, SO2, CO, NH3, C7H8, C6H6, C8H10, As(ng/m³), Ni(ng/m³), Cd(ng/m³), Pb(ng/m³), B(a)p(ng/m³) ruido, velocidad viento, dirección del viento, temperatura, humedad relativa, presión, radiación Solar, precipitación, velocidad máxima del viento, AS, Ni, Cd, Pb, B(a)p. Estos datos han sido conseguidos de la Red Valenciana de Vigilancia y Control de la Contaminación Atmosférica, y serán muy útiles a la hora de visualizar como ha evolucionado la sostenibilidad de la ciudad de Valencia.

- **Estaciones de contaminación atmosféricas:** Aquí tendremos las estaciones de las que hemos obtenido los parámetros anteriormente mencionados, y las cuáles usaremos para visualizar los datos que tenemos, estando estas repartidas por Valencia.
- **Distritos:** El siguiente fichero contiene las capas de información administrativa proveniente del servicio de Urbanismo e Infraestructuras de “Opendata” referentes a distritos. Estos datos nos serán muy útiles a la hora de situar las estaciones que hemos obtenido anteriormente y a la hora de separa por zonas para visualizar los datos que tenemos.
- **Zonas Verdes (Planificación):** Localización de lugares planificados para la creación de jardines y espacios verdes en la ciudad de Valencia. Estas zonas proporcionan un gran valor positivo para la sostenibilidad de la ciudad, ya sea mejorando la calidad del aire, regulando el clima urbano, dando biodiversidad y hábitat para la fauna o proporcionando mayor bienestar y calidad de vida.
- **Puntos medida tráfico espiras electromagnéticas:** Ubicación de los puntos de medida de tráfico en Valencia (espiras electromagnéticas). Estas zonas también tienen un rol para los objetivos de este proyecto, dando resultados más nocivos como la contaminación del aire, ruido y contaminación acústica, efecto isla de calor e impacto en la biodiversidad entre otros.

2. Preprocesado de Datos

Vamos a listar los métodos que seguimos para limpiar y preparar los datos antes de su análisis.

Tras leer los datos en R, los transformamos de manera que tengan el mismo tipo de dato para poder trabajar con ellos. Al hacer esto nos damos cuenta de que para muchas estaciones no tenemos los datos de todos los parámetros cada vez que se miden estos. Para solucionar este problema, lo que hacemos para poder seguir trabajando con estos parámetros es sustituir los valores faltantes de estos datos que no tenemos por la media que dicho parámetro había

conseguido durante ese mes en esa estación mediante un “group by” por estación y mes, de esta manera logramos no desvirtualizar demasiado los datos que tenemos y conseguimos poder seguir trabajando con ellos para una visualización correcta.

Una vez hemos sustituido estos valores, a los datos que no se les haya podido asignar un valor de ese modo los eliminamos, ya que no harán otra cosa más que estorbar a la hora de visualizar los datos que tenemos más completos y que queramos visualizar con mayor rigor.

Para poder trabajar con los datos de contaminación de una manera más efectiva y para evitar que tengamos que generar gráficas y mapas para cada uno de los parámetros de sostenibilidad que hemos visualizado, lo que hacemos es crear un parámetro único llamado Índice de contaminación, que mediante la unión de diferentes variables que habíamos leído, lo podamos usar de manera más práctica y visual. Este parámetro lo obtenemos mediante los valores de NO, NO₂, NO_x, O₃ y SO₂. Esto nos facilitará mucho a la hora de que nuestros outputs generados sean más fáciles de entender y de representar.

Tras haber hecho estos cambios, volvemos a ordenar los datos que tenemos por año, para evitar que hayamos realizado algún cambio involuntario en el proceso. Guardamos este df como Datos_limpios.csv.

Para poder trabajar con los datos de manera más libre, lo que hacemos es dividir los datos que tenemos a través de los años. De esta forma, creamos un archivo con todos los datos de todas las estaciones que hemos obtenido para cada uno de los años en los que trabajamos, desde 2004 hasta 2022. Así obtenemos 19 archivos, uno de cada año, con los que nos resultará más sencillo trabajar a la hora de representarlos y sacar conclusiones temporales.

3. Preprocesado de la información geográfica

Pasamos a trabajar con los datos en QGIS. Para ello, abrimos las capas que hemos obtenido de la base de datos de las diferentes estaciones, distritos y el archivo de cada año que hemos calculado anteriormente en R. Cuando observamos estas capas, notamos que, en distritos, tenemos varios distritos que no van a ser útiles, ya que no tenemos datos de estaciones que se encuentren en esas zonas, y por lo tanto no harían más que entorpecer si queremos visualizar los resultados fielmente. Por tanto, eliminamos los poblados norte y sur para poder trabajar mejor con las capas.

Una vez queremos trabajar con las capas de cada año, primero debemos geolocalizar estos datos, ya que tal como los hemos obtenido, no se encuentran geolocalizados. Así que mediante la capa de estaciones, la unimos por a la capa de cada año por la columna de distrito en cada capa, de manera que conseguimos que todos los datos se encuentren geolocalizados.

Un problema que obtenemos a la hora de juntar estos datos es que el índice de contaminación que habíamos calculado previamente se nos ha transformado a tipo de dato “string”, por lo que simplemente con la calculadora de datos creamos otra columna en la que convertimos el índice de contaminación a tipo de dato real.

Posteriormente, para cada capa de cada año, realizaremos una interpolación IDW sobre la capa de distritos para obtener su raster, el cual tras recortamos lo guardaremos. Repetimos el proceso para cada año desde el 2004 hasta el 2022 y obtenemos un raster para cada uno de estos años.

Una vez realizado esto, para cada uno de los rasters que acabamos de obtener y con la capa de distritos, medimos mediante la función de “estadísticas de zona” la media de contaminación para cada distrito y obtenemos otra capa que guardaremos para cada año. Esta capa nos será muy útil a la hora de visualizar la evolución de nuestros datos.

Como uno de los datos que hemos querido utilizar de la base de datos a la hora de poder indagar un poco más sobre el nivel de sostenibilidad de Valencia son las zonas verdes que hay, queremos averiguar que proporción de zonas verdes por distrito tenemos en los datos que tenemos. Para ello, con la calculadora obtendremos el área de cada uno de los distritos con los que trabajamos. Ahora nos faltaría averiguar el área de zonas verdes en cada uno de los distritos.

Para esto, primero debemos realizar la intersección entre los distritos y las zonas verdes. Haremos esto para asignar a cada una de las diferentes zonas verdes el distrito que le corresponde. Después, mediante la opción de “Estadística por Categorías”, agrupamos por estas zonas por distrito y sumaremos los metros cuadrados de zonas verdes para cada uno de los distritos de los que disponemos. Lo siguiente será unir cada una de las estadísticas con su distrito correspondiente, esto se hará al igual que antes con la herramienta Unir atributos por valor de campo. Una vez que ya tenemos este dato para cada uno de los distritos, simplemente dividimos en cada distrito por el área total que tiene que ya habíamos obtenido previamente, y así habremos logrado calcular la proporción de zonas verdes en cada uno de los distritos con los que trabajamos.

De manera similar a como hemos hecho anteriormente, queremos utilizar los datos de tráfico para averiguar la intensidad del tráfico media en cada uno de los distritos con los que trabajamos.

Lo que haremos será utilizar la variable ih: intensidad hora de vehículos (vehículos/hora) para el fin que queremos. Pero primero debemos arreglar que esta medida se nos muestra como una variable de clase “string”, así que de manera similar a como hemos hecho anteriormente, con la calculadora de campos creamos una nueva columna “ih2” como dato entero. Posteriormente, realizamos la interpolación TIN de las medidas de tráfico con la extensión de distrito, que después con la opción de “Estadísticas por zona” agrupamos por los distritos disponibles, de manera que obtendremos la media de la intensidad del tráfico por distrito. Estos datos nos servirán para explicar de dónde proceden los parámetros de contaminación que tenemos.

4. Elección de gráficas para los distintos tipos de datos

Para las evoluciones temporales, hemos escogido utilizar gráficos de línea. Consideramos que los gráficos lineales son mejores para estas representaciones ya que los datos tienen un orden inherente al estar ordenados por fecha y tener un orden temporal. Este gráfico también nos da una mayor simplicidad y claridad a la hora de visualizar nuestros datos, y no hace más fácil detectar los cambios de tendencias y patrones, ya que nos permite una mayor conexión de puntos sobre los datos.

También utilizamos gráficas de barras cuando tenemos un conjunto de categorías y un valor cuantitativo para cada categoría, con lo que hacemos énfasis principal en estas visualizaciones

en la magnitud de los valores cuantitativos. Usamos barras verticales, que entre sus ventajas tenemos una clara comparación entre categorías, distinción de estas mismas, facilidad para mostrar diferencias e incorporación de múltiples variables.

Por último, hemos decidido combinar en algunas gráficas los dos estilos, representando con barras valores cuantitativos que pudieran explicar los valores de la otra variable, la cual se representará con una línea continua con puntos, a modo de línea de tendencia.

5. Diseño del mapa e interactividad

Para el diseño de los mapas creemos que lo mejor es el uso de rasters y polígonos(Distritos) coloreados. Con los rasters conseguimos una visión continua de los datos, en nuestro caso por ejemplo el índice de contaminación. Mientras que con la otra opción conseguimos un efecto parecido al de un raster, pero de una forma discretizada. Consiguiendo comparar más fácilmente las medidas entre distritos, como saber cuál tiene un mayor índice de contaminación y cuál uno menor.

En los mapas finales tendremos todas las capas calculadas anteriormente en el preprocesado de los datos. Para empezar, tendremos un mapa donde mostraremos el nombre y la ubicación tanto de los distritos escogidos para el estudio, como de las estaciones de contaminación atmosférica que han tomado las medidas necesarias para el proyecto.

Las siguientes dos representaciones serán muy similares, serán mapas que muestran la evolución temporal en todos los distritos, desde 2004 hasta 2022. La única diferencia será la utilización de polígonos para el primero, representando la contaminación en cada uno de los distritos y el uso de los rasters para el segundo, representando la contaminación de una forma continua. La elaboración de estas capas ha sido realizada en Qgis, como ya hemos comentado anteriormente.

En las dos siguientes representaciones la estructura es muy similar. En esta primera parte compararemos la contaminación con la proporción de zonas verdes en cada distrito. Para empezar, tendremos la opción de mostrar dos mapas, en ellos estará la representación de la contaminación en 2022 y en el otro estarán representadas las zonas verdes que hay en Valencia. Elegimos sólo el año 2022 ya que el registro de las zonas verdes es muy actual, por lo que creemos que incluir datos más antiguos podría desvirtuar el resultado. Para estos mapas se utilizará la capa de raster de 2022 ya utilizada anteriormente y el archivo original que contiene las zonas verdes de la ciudad de Valencia. También habrá una segunda opción en la que representaremos con un gráfico de barras para cada distrito la proporción de zona verde que tiene, además representaremos el índice de contaminación con una línea con la que podamos ver si hay algún tipo de relación entre las variables. Para este gráfico se usará la capa de contaminación por distrito que hemos usado anteriormente y la capa de proporción de zonas verdes, la cual hemos calculado y explicado en el preprocesado.

La siguiente representación, como ya hemos comentado, es muy similar a la anterior. En esta ocasión volveremos a representar el índice de contaminación, pero esta vez respecto a la intensidad del tráfico. Tendremos también una primera opción de mostrar dos mapas, en el caso del mapa de la intensidad del tráfico representará esta variable para cada distrito con un mapa de coropletas en función del nivel de intensidad. La segunda opción será igual que en la

representación anterior, pero en este caso el gráfico de barras será representado por la variable de intensidad del tráfico.

Por último, diseñaremos un gráfico de líneas para poder observar la evolución temporal de cada una de las estaciones. El usuario tendrá la opción de elegir el rango de tiempo y la variable a representar. Para este gráfico usaremos el fichero con los datos completos limpios que guardamos al comienzo del preprocesado.

6. Diseño del cuadro de mandos

Para el diseño de la aplicación nos hemos apoyado principalmente en las librerías Shiny y Leaflet de R. Esta primera nos permite crear el menú de nuestra aplicación de una forma fácil e intuitiva, mientras que la segunda nos ayuda a representar de distintas maneras los datos espaciales sobre mapas.

Queríamos un menú en el que pudiéramos elegir entre 6 pestañas, en las que cada una contendrá una de las representaciones que queremos hacer. Dentro de cada pestaña encontraremos los posibles inputs que pueda haber antes de cada representación. Para algunos podremos introducir años específicos, otros nos permitirán introducir rangos de tiempo, en otros podremos seleccionar el tipo de representación que queremos, etc.

Por último, dentro de cada pestaña también podremos encontrar los plots que hemos diseñado. Habrá algunos datos que podemos representar simplemente con la librería de ggplot2, pero también tenemos datos georeferenciados que deberemos de representar con su correspondiente mapa, para lo cual diseñaremos funciones con la ayuda de la librería Leaflet.

7. Implementación

UI (Interfaz de Usuario):

-Se define la apariencia y la estructura de la aplicación utilizando el conjunto de funciones proporcionadas por Shiny y Shinydashboard.

-.Se crea un panel de control con diferentes pestañas (tabItems) que representan las diferentes secciones de la aplicación, como "Distritos y Estaciones", "Contaminación por Distrito", etc.

-Dentro de cada pestaña, se especifican los elementos de la interfaz de usuario, como controles deslizantes o selectores de fecha con funciones como "selectInput" o "sliderInput".

- Por último, se definen los plots de cada pestaña. Según el tipo de representación se elegirá la función de "leafletOutput" (mapas interactivos) o "plotOutput" (gráficos).

Server (Servidor):

-Se definen las funciones de renderizado (renderLeaflet, renderPlot, renderUI) que generan la salida dinámica basada en las entradas del usuario.

-Cada función de renderizado está asociada con un identificador de salida (output\$Id) que se utiliza en la interfaz de usuario para mostrar los resultados.

-Se leen y procesan los datos necesarios para generar los mapas, gráficos y otros elementos visuales.

-Se definen las funciones que representarán los datos utilizando las entradas que introduce el usuario (Variable\$input) con funciones como ggplot() o leaflet().

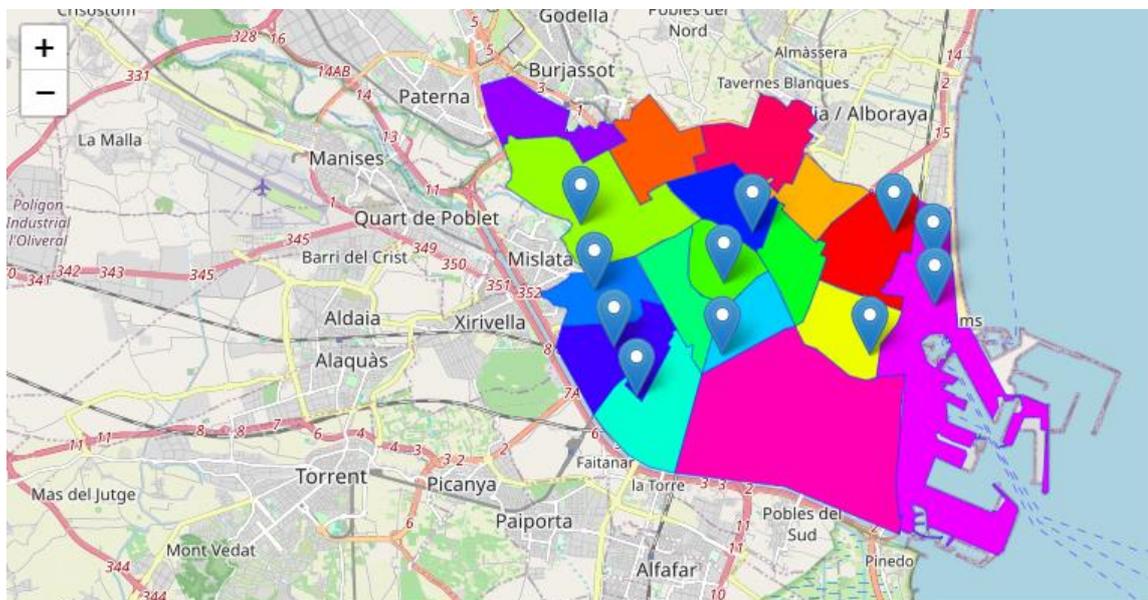
-Se definen también dentro de las funciones principales parámetros estéticos de cada representación con funciones y parámetros como addLegend(), theme(), (opacity = X) o (pal = Y)

En resumen, este código construye una aplicación interactiva que permite a los usuarios explorar y visualizar datos relacionados con la contaminación en la ciudad de Valencia, utilizando mapas interactivos, gráficos y controles para personalizar la visualización de los datos.

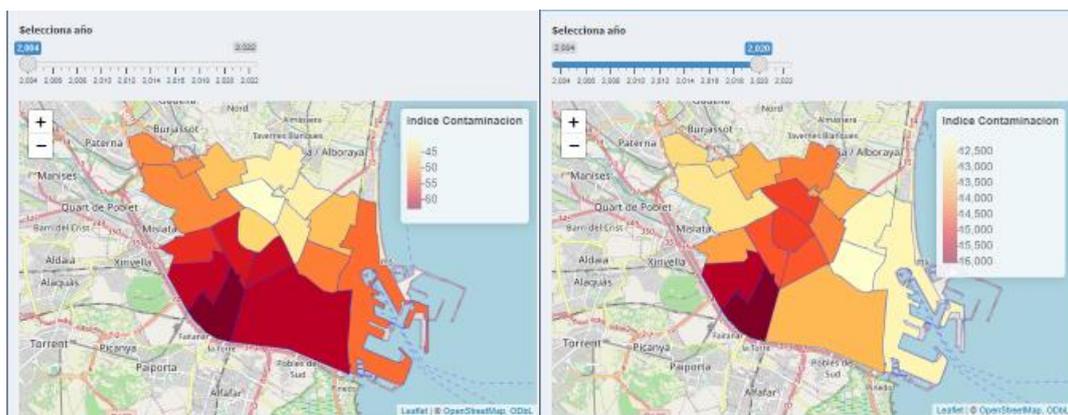
3. Resultados

Vamos a explicar los resultados que hemos obtenido. Para ello, los mostramos en diferentes pestañas de la aplicación de R Shiny, en la que mostraremos los resultados obtenidos y explicaremos nuestras observaciones:

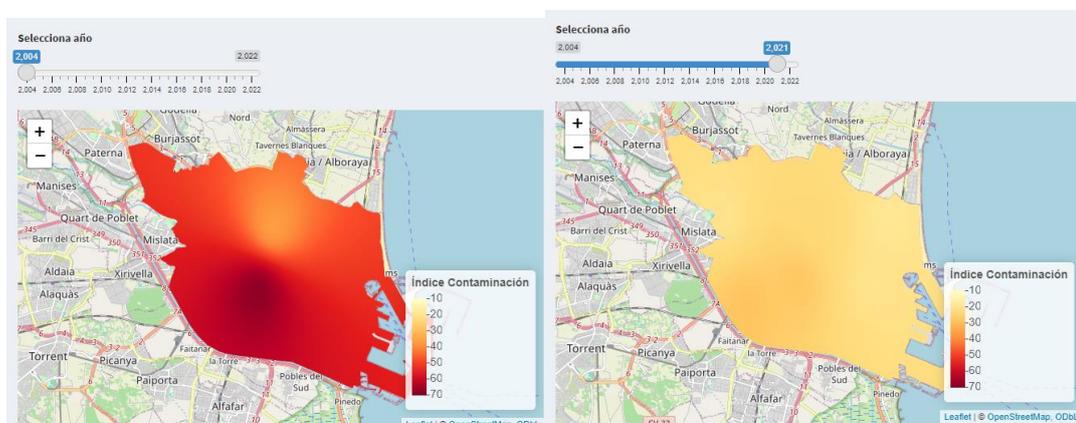
- **Distritos y Estaciones:** Aquí mostramos los diferentes distritos y las estaciones que hemos observado, representadas en diferentes colores sobre un mapa. Nos servirá de apoyo a la hora de mostrar el resto de mapas y gráficas.



- Contaminación por Distrito:** En este mapa se representa el índice de contaminación (que habíamos calculado previamente basándonos en varios parámetros de contaminación) en cada uno de los distritos a lo largo de los años. Podemos observar que, en contra de la lógica, los primeros años de los que tenemos datos la contaminación era mayor, sobre todo en los distritos del sur oeste. Poco a poco va bajando la contaminación en todos los distritos hasta llegar a su punto más bajo en el año 2020, seguramente debido a la pandemia mundial que vivimos. Después, poco a poco va volviendo a subir la contaminación hasta el 2022, con proyección ascendente. Procedemos a mostrar los mapas de los años 2004 y 2020:



- Ráster de Contaminación:** De forma similar a la gráfica anterior, con este raster observamos la evolución de la contaminación en Valencia por los diferentes años, esta vez sin tener en cuenta los distritos. Al ser los datos los mismos, obtenemos resultados idénticos, con mayor nivel de contaminación en los primeros años y poco a poco disminuye hasta la pandemia que vuelve a subir.



- Contaminación y zonas verdes:** En esta pestaña representaremos dos gráficas diferentes con el fin de ver si las zonas verdes y la contaminación van o no de la mano. Con este propósito en mente, representaremos en un gráfico de barras en verde la la

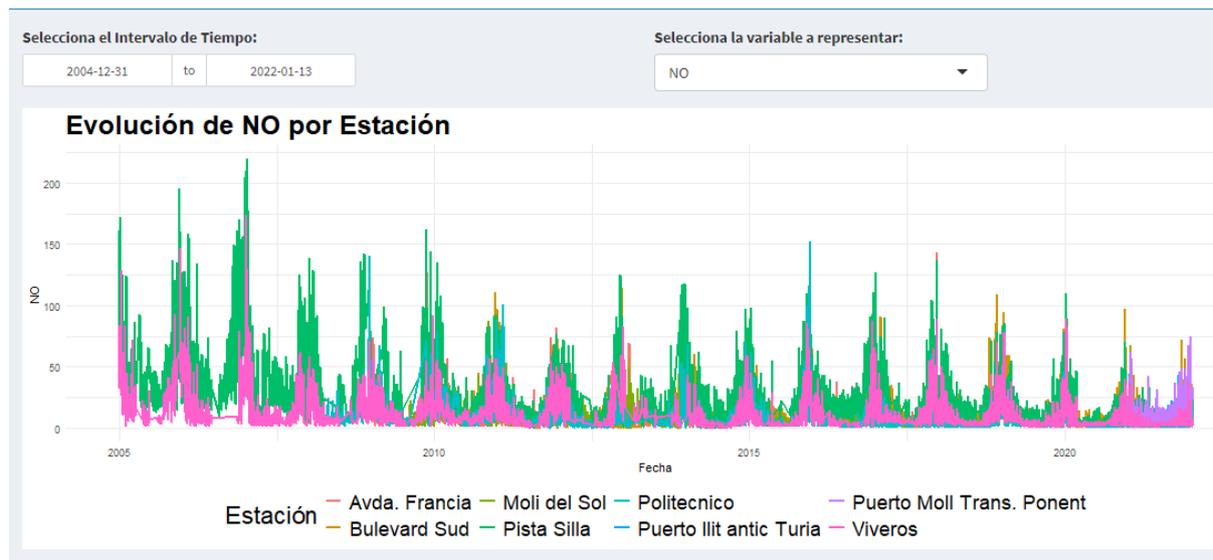
proporción de zonas verdes en las diferentes estaciones de medición de las que tenemos disposición. Al ser estos datos de 2023 tal como nos proporciona la web de descarga, para evitar inconsistencias, los representaremos junto al índice de contaminación del año 2022, siendo este el año más cercano que tenemos disponible. Lo representamos mediante un gráfico de líneas. Podemos comprobar de esta manera que no es totalmente proporcional, pero si que tienen mucha relación. Por ejemplo, observamos que las zonas con mayor índice de contaminación (Jesus y Patraix), son de las zonas con menor proporción de zonas verdes, siendo este último el único distrito sin ninguna zona verde. Por lo general se ve una cierta relación excepto en Poblets Maritims, pero eso tiene una explicación que veremos en nuestra siguiente gráfica.



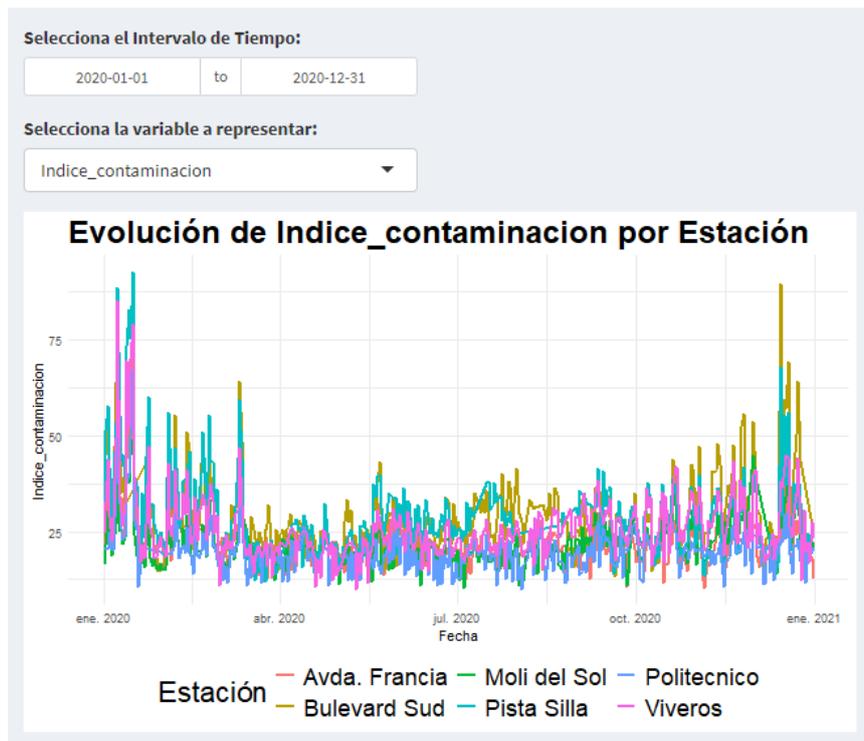
- Contaminación y tráfico:** De manera análoga a como hemos hecho antes, queremos comparar el tráfico en los distintos distritos que tenemos disponibles con su índice de contaminación para ver su relación. Podemos observar que se relacionan bastante, siendo el ejemplo más claro en Poblets Maritims, donde al ser el único distrito sin tráfico es a la vez uno de los que tiene la contaminación más baja. En los distritos con mayor contaminación, también se cumple que tienen un elevado nivel de tráfico, observando que no es un factor imprescindible pero ambos estadísticos tienen bastante que ver el uno con el otro.



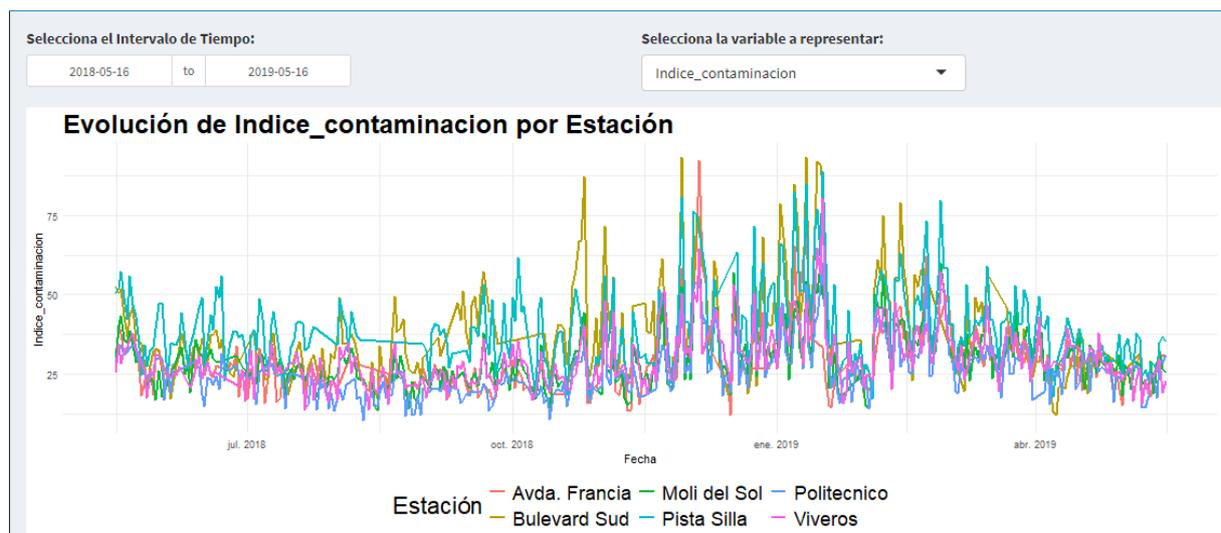
- Evolución temporal de contaminación:** Por último, realizamos una gráfica en la que podemos introducir dos fechas para representar la contaminación en cada una de las estaciones disponibles entre dichas fechas. Las diferentes variables de las que podemos elegir representar son: NO, NO₂, NO_x, O₃, SO₂ y el índice de contaminación. Jugando con las fechas podemos observar varios datos interesantes. Por ejemplo, notamos que, al igual que hemos visto anteriormente, en los primeros años de los datos que tenemos disponibles, la contaminación es mayor.



También, si observamos la evolución de la contaminación en el año de 2020, podemos observar un claro bajón de contaminación a partir de marzo, debido a la pandemia mundial que vivimos.



Por último, si miramos cuál es la época del año en la que tenemos más contaminación, esa sería el invierno, desde noviembre hasta febrero aproximadamente. Esto se podría deber a la Navidad, una época festiva de muchos viajes y mucha producción.



4. Discusión

Viendo las gráficas se observa que los índices de contaminación en 2004 eran muy altos y a partir de 2005 se va reduciendo gradualmente la contaminación con los años, esto podría ser debido a una falta de regulación por parte de la Generalitat, ya que a partir del año 2005 la dirección de calidad ambiental empezó un seguimiento de los niveles de los contaminantes atmosféricos más importantes en las principales áreas urbanas e industriales, la Red Valenciana de Vigilancia y Control de la Contaminación Atmosférica.

Esta red establece unos límites para ayudar a medir la acumulación de contaminantes en distintas zonas de la comunidad, además pone en práctica la implantación de varias estaciones de muestreo para poder documentar los niveles de contaminación a lo largo de varias zonas de la comunidad valenciana. Además de esta, con el tiempo han ido saliendo otras iniciativas a favor del medioambiente como la Planificación Territorial e Infraestructura Verde y la evaluación ambiental.

A partir de 2005 hasta el día de hoy existen informes detallados de la contaminación a lo largo de toda la comunidad, esta documentación ha facilitado la implementación de diversas medidas que han logrado reducir la contaminación año tras año. Muchas de estas medidas forman parte de la transición ecológica, llevada a cabo también por la Generalitat, se trata de un conjunto de proyectos centrados principalmente en el uso de energías renovables

El decrecimiento del índice de contaminación continua hasta 2020, donde alcanza su punto más bajo, reduciéndose a la mitad debido a la pandemia. Como es lógico en el 2021 aumenta y sería injusto compararlo con el 2020, si se compara con el 2019 se ve una reducción de la contaminación moderada, lo que llevaba ocurriendo los últimos años.

Analizando la contaminación por distritos se observa que las zonas con un índice más elevado se encontraban principalmente al sur y con el paso de los años la contaminación se ha ido desplazando hacia el este, esto podría ser debido a un aumento en la población del interior o un incremento del tráfico en estas zonas entre otras cosas.

Durante este proyecto hemos observado como la contaminación en Navidades aumenta considerablemente. Como es razonable esto ocurre en cualquier lugar donde se celebre la navidad, durante estas fechas se calcula que la persona promedio deja el 5.5% de su huella de carbono anual total, debido a un aumento de los viajes y el consumo en general, sobre todo el de la comida.

Aunque el estudio proporciona información valiosa sobre la evolución sobre la evolución de la contaminación, existen ciertas limitaciones a tener en cuenta. Por ejemplo, las limitaciones geográficas, ya que con este estudio tan solo hemos podido estudiar los distritos de Valencia, por lo que no puede ser fiel al completo con el resto del mundo.

También nos hemos encontrado con datos discontinuos por la pandemia, lo cual ha generado datos atípicos que han afectado a la interpretación de las tendencias generales, aunque no en gran medida.

Este estudio también tiene ciertas limitaciones de los datos, ya que, con más datos como población por distrito, fábricas por distrito u otros datos de interés podríamos haber obtenido unos resultados más exhaustivos.

En cuanto a las posibles áreas para futuros trabajos, se podrían realizar identificaciones de fuentes específicas de contaminación, para poder averiguar por dónde empezar a combatir la contaminación más eficazmente., se podría investigar sobre los efectos de esta contaminación en la salud de los habitantes de los distritos afectados o analizar los cambios demográficos y urbanos.

En relación con las dificultades encontradas una de las principales fue el fichero principal que contiene las mediciones de diferentes variables relacionadas con la contaminación. El problema con este fichero es la cantidad de datos faltantes que hay en él, además de la falta de georreferencia de los datos que obliga a unirlos con otros.

El caso de los valores faltantes es bastante grave, ya que hay estaciones que no empiezan a tener mediciones correctas hasta después de varios años, no pudiendo ni siquiera sustituirlos por la media de la variable en el mismo mes porque no hay ningún dato disponible para hacerla. Por otra parte, hay variables que tienen más valores faltantes que mediciones, lo cual hace imposible su inclusión en el estudio.

Por otro lado, el problema con la georreferencia es doble ya que solo tenemos un conjunto de datos al que poder unirlos, el fichero que contiene las estaciones geolocalizadas, pero en el cual hay estaciones que no coinciden en el nombre con las estaciones que refleja el fichero de medidas y otras estaciones las cuales directamente ni aparecen en el fichero. Esto sigue limitando mucho los datos para un estudio en profundidad.

5. Conclusiones

En conclusión, hemos obtenido información valiosa sobre la contaminación a nivel histórico de los diferentes distritos de Valencia, de cómo afectan las zonas verdes y la intensidad del tráfico a estos mismos, tal y como veíamos en la introducción. Con esto podemos ver qué medidas tomar de manera útil, siendo estas dos una buena forma de empezar para no volver a seguir la corriente del aumento de contaminación que últimamente habíamos tomado.

También hemos visto en que lugares podemos reducir la contaminación de manera más eficaz, siendo en Navidad y en los distritos donde hemos observado un mayor nivel de contaminación.

Empezar por estas reducciones sería muy beneficioso a la práctica, y siendo que el medio ambiente lo necesita, sería muy positivo para el futuro.

6. Referencias

-Generalitat Valenciana. (s.f.). Evaluación Ambiental. Recuperado de:

<https://mediambient.gva.es/es/web/evaluacion-ambiental>

-Generalitat Valenciana. (2020). Informe del Estado del Medio Ambiente en la Comunitat Valenciana. Informe de coyuntura 2017-2020. Recuperado de:

<https://mediambient.gva.es/documents/20552612/90528229/Informe+del+Estado+del+Medio+Ambiente+en+la+Comunitat+Valenciana.+Informe+de+coyuntura+2017-2020.pdf/77e34988-5890-91b6-2562-d275af1a8174?t=1666016722570>

-Cooper, D., & Owen, A. (2013). The Carbon Cost of Christmas. ResearchGate. Recuperado de:

https://www.researchgate.net/publication/263969440_The_Carbon_Cost_of_Christmas