

PRACTICA OPEN DATA & DYNAMIC VIZ

Introducción

Esto es una breve explicación de lo que hace la práctica que he realizado. Primero mencionar que he utilizado datos del INE, del Ministerio de Interior y de GitHub para sacar los geojson que representan las comunidades y provincias españolas.

Explicación parte open data

En este programa se realiza la obtención y procesamiento de datos oficiales de España sobre criminalidad y renta media por municipio. Para que funcione correctamente, es necesario que exista una carpeta data en el directorio de ejecución, ya que ahí se van a guardar todos los archivos generados. Además, se dispone de un fichero requirements.txt, que permite instalar todas las librerías necesarias para ejecutar el programa.

Los datos de criminalidad se obtienen del Ministerio del Interior utilizando Selenium. Se automatiza la navegación por la web, haciendo clic en botones para aceptar cookies, abrir menús y acceder a los datos trimestrales de cada municipio. En este proceso, es necesario interactuar con un iframe, ya que los enlaces a los datos están dentro de un HTML incrustado dentro de la página principal. Una vez dentro del iframe, se seleccionan todas las tipologías de delitos y todos los periodos disponibles para cada trimestre, y se descarga la tabla de datos.

Los datos de renta media y mediana provienen del INE, también extraídos con Selenium. El programa selecciona automáticamente todos los municipios de cada provincia y los indicadores de renta relevantes para 2022 y 2023. Luego se guardan en archivos Excel, uno por provincia, dentro de la carpeta data.

Una vez obtenidos los datos, se realiza un procesamiento adicional: se agrupan los trimestres para obtener los totales anuales, se calculan las variaciones porcentuales entre 2022 y 2023, y se corrigen los nombres de comunidades y provincias para que sean consistentes. Esto permite que los datos estén listos para su análisis y visualización en la app de Streamlit.

Elegí estos datos principalmente porque ya estaba familiarizado con ellos, tanto por mi TFG previo como por discusiones con amigos, y ya había trabajado antes con indicadores de renta media. En mi TFG no utilicé web scraping: en su lugar, partía de ficheros Excel oficiales y escribí un parser en Java para extraer los datos y almacenarlos en una base de datos. Ahora, este programa me permite automatizar la descarga de los mismos datos y obtenerlos en un formato directamente utilizable para análisis y visualización.

Explicación parte visualización dinámica

Lo que se puede ver en el Streamlit es un dashboard interactivo que permite explorar tanto la criminalidad como la renta de los municipios y comunidades autónomas de España. La aplicación combina datos oficiales de criminalidad y de renta, permitiendo analizar, comparar y visualizar la información de manera clara y flexible.

Sección de Delitos

1. Tabla interactiva de delitos

En esta pestaña, llamada “Tabla interactiva”, se muestra un resumen completo de los datos de criminalidad en España. Aquí se pueden explorar todas las columnas del dataset, incluyendo delitos por tipo, región, año y trimestre, con una vista tabular que permite filtrar y examinar los datos fácilmente.

2. Histograma por tipo de delito

En esta sección, se puede seleccionar un tipo de delito específico mediante un menú desplegable. La aplicación genera un histograma comparativo entre 2022 y 2023, mostrando la cantidad total de delitos registrados en cada año para el tipo seleccionado. Esto permite identificar tendencias o cambios en la criminalidad por tipo de delito.

3. Gráfico por región

Esta pestaña permite analizar los delitos desglosados por región, ya sea comunidad autónoma, provincia o municipio.

- Primero se selecciona el nivel de agregación deseado y luego la región específica.
- La visualización se genera como un gráfico de barras, mostrando el total de delitos por tipo en el año 2023, con información adicional al pasar el cursor sobre cada barra (total 2022 y variación respecto al año anterior). Esto permite comparar la criminalidad dentro de cada nivel geográfico de manera intuitiva.

4. Mapa interactivo de España

Esta sección permite visualizar la distribución geográfica de los delitos mediante un mapa interactivo basado en Folium:

- Se puede elegir el nivel de agregación (comunidad o provincia), el tipo de delito y el trimestre o el total anual.
- Cada región se colorea según la cantidad de delitos, usando una escala de colores que facilita identificar las zonas con mayor incidencia.
- Al pasar el cursor sobre una región, se muestra un tooltip con información detallada: nombre de la región, tipo de delito, número de delitos en 2022 y 2023, y la variación porcentual.
- Debajo del mapa se incluye un gráfico de barras complementario para comparar visualmente los valores entre las diferentes regiones. Esta sección permite tener una visión geoespacial inmediata de la criminalidad en España.

Sección de Rentas

5. Exploración interactiva de la renta

En la pestaña “Rentas”, los datos de renta se presentan mediante dos tabs principales.

○ Primer tab: “Exploración interactiva renta”

Aquí se pueden seleccionar una o varias comunidades autónomas, elegir la métrica de renta a analizar (por persona, por hogar, por unidad de consumo, media o mediana) y definir cuántos municipios visualizar en los gráficos.

La sección muestra:

- Métricas principales: media y mediana de la renta, así como el número de municipios incluidos según los filtros aplicados.
- Gráfico de barras: ranking de los municipios con mayor renta según la métrica seleccionada, con barras coloreadas por comunidad autónoma.
- Tabla de datos: vista tabular de los mismos datos filtrados, para explorar valores individuales con detalle.

○ Segundo tab: “Agrupación por CP y Comunidad”

Este tab muestra un resumen agregado por comunidad autónoma:

- Calcula las medias de renta por persona, por hogar y por unidad de consumo, así como la mediana y el número de municipios que componen cada comunidad.
- Incluye un gráfico de barras que compara la renta media por persona entre comunidades.
- Un gráfico de dispersión muestra la evolución de la renta de 2022 a 2023 por municipio, permitiendo identificar qué comunidades han experimentado mayor crecimiento o descenso.
- Una tabla agregada permite explorar la información de manera comparativa y detallada.