Aplicación interactiva para el análisis y visualización de datos de generación de energía en países de todo el mundo

Multimedia Interactiva 2021-2022

Raúl Colino Singh

# Introducción

En este trabajo se describen las motivaciones de visualización de los datos que se desean analizar y decisiones tomadas durante el desarrollo de la aplicación.

# Caracterización del problema, objetivos de la aplicación y datos

La energía eléctrica es un bien que se considera fundamental para el desarrollo de un país. Por ello se desea ver la relación que existe entre la cantidad de generación de energía que genera un país, su población y su riqueza.

En los últimos años se ha oído hablar frecuentemete sobre la transición energética a las energías renovables y sobre los retos que ello conlleva. Por ello se desea visualizar la evolución que los países han tenido para averiguar si el mundo va progresando hacia un futuro sostenible y respetuoso con el planeta.

Por ello, un objetivo de esta aplicación es permitir observar la evolución de las distintas fuentes de energía a lo largo del tiempo en ciertos países concretos. El gráfico “stacked area” se creó con este fin.

También se desea visualizar la cantidad de energía generada por cada uno de los países de todo el mundo de un solo vistazo. Para ello un mapa coroplético resulta ideal.

Sin embargo, para hacer una comparación más detallada con un ranking y averiguar las diferencias que existen entre varios países, se han incorporado dos diagramas de barras. Además debido a la distribución de los gráficos, estos diagramas mostrados con una altura reducida permite tener al mapa coroplético y el diagrama “stacked area” mas cerca de la sección de eleccion de datos de entrada permitiendo una visualización más rápida y cómoda.

Todos los gráficos, menos el “stacked area chart” están sincronizados con los datos de entrada y se pueden añadir o eliminar los países para centrar la visualización en los países que se deseen analizar.

El dataset ha sido creado a partir de distintas fuentes:

* World in Data
* gapminder.org

El dataset tiene 18 columnas y un indicador en formato de rango de tiempo (en años) distribuido por filas. En total el dataset cuenta con 3588 filas.

Las columnas más relevantes del dataset son:

* Code: Codigo del pais
* Year: Indica el ano al que pertenecen los datos
* GDP: Este indicador proporciona valores per cápita del producto interior bruto (PIB) expresados en dólares internacionales corrientes convertidos por el factor de conversión de la paridad del poder adquisitivo (PPA). El PIB es la suma del valor añadido bruto de todos los productores residentes en el país más los impuestos sobre los productos y menos las subvenciones no incluidas en el valor de los productos. Con este indicador se trata de representar en nuestros gráficos la riqueza de cada país.
* Entity: Nombre del país
* Coal: Electricidad producida por carbón (Teravatios/hora)
* Gas: Electricidad producida por gas (Teravatios/hora)
* Hydro: Electricidad producida por agua (Teravatios/hora)
* Other renewables: Electricidad producida por otras fuentes renovables no explícitas en el dataset (Teravatios/hora)
* Solar: Electricidad producida por energía solar (Teravatios/hora)
* Oil: Electricidad producida por petróleo (Teravatios/hora)
* Wind: Electricidad producida por fuentes eólicas (Teravatios/hora)
* Nuclear: Electricidad producida a partir de energía nuclear (Teravatios/hora)
* Access to electricity (% of population): Porcentaje de población con acceso a electricidad
* Population: Población (estimaciones históricas)
* All sources combined: Total de energía generada (Teravatios/hora)

# Interacción, decisiones de diseño e implementación

Se ha decidido utilizar la librería de desarrollo de aplicaciones de ciencia de datos de Streamlit.io para crear la interfaz web e incorporar los gráficos de la aplicación con Python ya que se tenía experiencia previa con este lenguaje. Se eligió crear la aplicación con Streamlit.io porque es una librería reciente que está teniendo una creciente popularidad y se deseaba probarla.

La librería utilizada para dibujar la mayor parte de los gráficos ha sido Altair debido a la sencillez y minimalismo de sus diseños.

Para el mapa coroplético se decidió utilizar en un principio Altair, pero, debido a la pobre documentación y la poca facilidad de uso de sus mapas, se decidió utilizar en su lugar Folium, una librería mucho más versátil y mejor documentada.

En la fila superior de la aplicación se muestran de izquierda a derecha:

* Los widgets para la elección de los datos de entrada que se desean visualizar. En esta sección se puede elegir la fuente de energía (o la combinación de todas), el año de los datos a mostrar que a través de un slider se puede ir cambiando y visualizar los cambios a través del tiempo en los gráficos y los países que se desean comparar/visualizar en los gráficos.
* El diagrama “scatter plot” para visualizar la relación GDP-Generación de energía-Población de cada país elegido en los widgets anteriormente mencionados.

Se decidió utilizar dos diagramas de barras debajo ya que ocupan poca altura y permiten visualizar el mapa y el diagrama “stacked area chart” lo más cerca posible de la sección de eleccion de datos de entrada y dando a la vez información relevante que permite comparar con precisión la energía eléctrica generada por cada país en forma de ranking y compararlo con el diagrama de % de población con acceso a electricidad.

Abajo del todo se encuentra el mapa coroplético que permite visualizar todos los países de todos los continentes y el gráfico “stacked area” que permite visualizar a través del tiempo los cambios en las fuentes de energía para un país concreto.

# Cómo ejecutar la aplicación

Para ejecutar la aplicación se requiere tener Python instalado. En concreto la aplicación se ha creado y probado en Python 3.8.

Para la instalación de Python se utilizó Conda.

Independientemente de si se utiliza Conda o no para la instalación de Python en el sistema, es recomendable instalar Streamlit a través de pip. Si no se hace así, pueden producirse errores con la librería de numpy.

También es necesario tener instaladas las siguientes librerías: pandas, geopandas, folium, streamlit\_folium y altair.

Con estos comandos se han instalado todas las librerias necesarias para ejecutar la aplicacion:

**pip install streamlit**

**conda install pandas -c conda-forge**

**conda install geopandas -c conda-forge**

**conda install geopandas -c conda-forge**

**conda install altair**

**conda install folium -c conda-forge**

**pip install streamlit-folium**

Una vez todas la librerias estan instaladas, la aplicación se puede ejecutar y visualizar en el navegador con el comando:

**streamlit run dashboard.py**

# Conclusiones

La tarea que más tiempo ha consumido en la práctica ha sido la búsqueda de los datos junto con su procesamiento para formar el dataset final que utiliza la aplicación de visualización de datos. El desarrollo de la aplicación ha requerido menos tiempo. Sin embargo debido a que Streamlit es muy reciente, tiene muchas limitaciones (por ejemplo las opciones de distribución y personalización de la interfaz son limitadas (esto se ve inmediatamente porque todas las aplicaciones de la comunidad se parecen mucho), a veces el texto y/o otros elementos se cubren unos a otros al cambiar el tamaño de la pantalla, para utilizar algunas librerías como Folium es necesario utilizar herramientas de otros desarrolladores (además son limitadas) y no se pueden anidar columnas para hacer distribuciones más complejas). De todas formas, si Streamlit madura puede llegar a ser una herramienta bastante útil y cómoda de utilizar.