



Escola  
Técnica  
Superior  
de Enxeñaría



# Tableau, Power Bi, Apache Superset

**Pablo Martínez González,  
Hugo Vázquez Docampo**

Almacenes y Minería de Datos

[Práctica 5 – 13/11/2022]

# Contenido

1. Introducción .....	3
2. Instalación .....	3
3. Datos iniciales y limpieza .....	4
4. Decisiones iniciales generales .....	5
5. Tableau .....	6
6. Power Bi .....	13
7. Apache Superset .....	21
8. Dificultades encontradas .....	24
9. Conclusiones .....	25
10. Horas de Trabajo .....	25

## 1. Introducción

En el siguiente informe, se realiza la explicación de las distintas acciones realizadas para implementar los ejercicios propuestos en las tres aplicaciones indicadas, siendo estas Tableau, Power Bi y Apache Superset. En ellas, se utilizarán distintas herramientas de representación de información de manera que se puedan mostrar diagramas, tablas y mapas sobre el consumo eléctrico y el crecimiento económico mundial de cada país en un cierto tiempo hasta la actualidad.

De esta manera, se crearán paneles de control interactivos para poder ver de manera sencilla los distintos resultados asociados a estos datos, iterando a través de los diferentes años o países.

Además, también se realizarán diagramas para representar información partiendo de la base de datos “Adventure Works”, con la que se ha trabajado en prácticas anteriores. Consecuentemente, se realizarán diagramas y gráficas para mostrar datos referentes a los clientes por países, y a las ventas totales por divisas.

## 2. Instalación

Para la realización de esta práctica, será necesario instalar las tres aplicaciones en las que se realizarán los distintos paneles de control para representar la información referente al consumo eléctrico y el crecimiento económico.

En primer lugar, para la instalación de la aplicación Tableau, simplemente será necesario descargarla y crear una cuenta. Posteriormente, se introducirá un código para poder acceder de manera gratuita a las herramientas de la aplicación. Una vez instalado, simplemente se creará un nuevo libro en el que estará el proyecto, y se cargará el fichero Excel con los datos sobre los que se desea hacer los diagramas.

De igual manera, la instalación de Power Bi es relativamente sencilla. Al tratarse de una aplicación derivada de Microsoft, esta se puede instalar en Windows de manera directa a partir de la tienda de aplicaciones. Una vez instalada, simplemente importando el fichero de fuente de datos, que en este caso será un fichero Excel, este se cargará y ya se podrá trabajar sobre él.

Para la instalación de Apache Superset se hace uso de la instalación docker disponible en la [web oficial de superset](#).

Una vez realizada la instalación es necesario aplicar una serie de configuraciones para poder realizar las diferentes tareas requeridas para este software:

- En la sección de **database connections**, se debe habilitar, sobre la base de datos de ejemplos, la opción de editar > advanced > security > Allow file uploads to database. Esto permitirá cargar el fichero excel para llevar a cabo las diferentes gráficas.
- Por otra parte, a la hora de conectar la base de datos a Apache Superset nos encontramos que todos los puertos se encuentran cerrados y, por tanto, la conexión se rechaza. Para solucionarlo, es necesario acceder al fichero docker-compose-non-dev.yml y añadir, sobre el contenedor de la base de datos la opción: “ports: -5432:5432”.
- Además, se opta por utilizar el contenedor con Postgres disponible ya en el programa. De este modo, se crea una nueva BBDD con el nombre de adventureworks y se hace uso del comando `psql -d adventureworks < install.sql` para cargar la base de datos. Una vez realizado este proceso, bastará con acceder a Superset y escoger la opción de nueva base de datos. En ella se indicará como IP: 172.18.0.1, como puerto 5432 y como base de datos la anteriormente creada.

### 3. Datos iniciales y limpieza

Para la realización de esta práctica se han utilizado una serie de datos extraídos del World Bank Databank. Con el objetivo de poder realizar gráficos y medidas relacionadas con el consumo eléctrico de los países y el crecimiento económico mundial, se ha buscado dentro de las distintas series proporcionadas por la base de datos, aquellas que coincidan con estas medidas. Consecuentemente, se han extraído dos variables, “Electric power consumption (KWh per capita), que mide el consumo eléctrico de cada país medido en kilo watios hora, y “GDP per capita (anual %) que mide el porcentaje de crecimiento o decrecimiento del producto interior bruto de cada país con respecto al año anterior.

Por otro lado, se han seleccionado únicamente aquellos datos de años comprendidos entre 1992 y 2021, ya que se ha considerado que, para la realización de un estudio útil y significativo de los datos, un periodo de tiempo de 30 años muestra información representativa de cada país.

Finalmente, seleccionando todos los países incluidos en la base de datos, se ha obtenido un fichero Excel en el que se refleja toda esta información. Cabe mencionar que el formato del fichero se trata de dos filas por cada país, una para cada variable de medida, y donde las columnas muestran todos los años desde 1992 hasta 2021 con los valores de las medidas. Además, también se incluye otra información como el nombre de la serie medida, el código, y el nombre de cada país.

	A	B	C	D	E	F	G	H	I
1	Country Name	Country Code	Series Name	Series Code	1992 [YR1992]	1993 [YR1993]	1994 [YR1994]	1995 [YR1995]	1996 [YR1996]
2	Afghanistan	AFG	Electric power	EG.USE.ELEC.K	..	..	..	..	..
3	Afghanistan	AFG	GDP per capita	NY.GDP.PCAP	..	..	..	..	..
4	Albania	ALB	Electric power	EG.USE.ELEC.K	453.6440739	536.3638251	596.4079593	663.7839954	9
5	Albania	ALB	GDP per capita	NY.GDP.PCAP	-6.622551212	10.22994935	8.969761659	14.02449593	9
6	Algeria	DZA	Electric power	EG.USE.ELEC.K	559.7830129	547.9904718	550.9719596	559.9178908	5
7	Algeria	DZA	GDP per capita	NY.GDP.PCAP	-0.564788753	-4.250986637	-2.931116597	1.836410561	2
8	American Sam	ASM	Electric power	EG.USE.ELEC.K	..	..	..	..	..
9	American Sam	ASM	GDP per capita	NY.GDP.PCAP	..	..	..	..	..
10	Andorra	AND	Electric power	EG.USE.ELEC.K	..	..	..	..	..
11	Andorra	AND	GDP per capita	NY.GDP.PCAP	-2.869215672	-4.427065042	-0.397087332	0.852320249	..
12	Angola	AGO	Electric power	EG.USE.ELEC.K	53.48666282	52.00747317	50.65258525	49.26424531	5
13	Angola	AGO	GDP per capita	NY.GDP.PCAP	-8.876931195	-26.41177135	-1.87789751	11.35953867	9
14	Antigua and B	ATG	Electric power	EG.USE.ELEC.K	..	..	..	..	..
15	Antigua and B	ATG	GDP per capita	NY.GDP.PCAP	-0.561528048	3.170343763	4.41621052	-6.408268751	4
16	Argentina	ARG	Electric power	EG.USE.ELEC.K	1410.377544	1496.904499	1576.767198	1648.895228	1
17	Argentina	ARG	GDP per capita	NY.GDP.PCAP	6.487631818	6.802926902	4.505456801	-4.032158811	4
18	Armenia	ARM	Electric power	EG.USE.ELEC.K	1962.635282	1195.024488	1045.478466	1050.243539	1
19	Armenia	ARM	GDP per capita	NY.GDP.PCAP	-40.74465636	-6.638470109	7.950115708	9.103388041	7

A pesar de que se han conseguido todos los datos necesarios para realizar las medidas y estudios correspondientes, se ha realizado una limpieza y filtrado de datos con el objetivo de facilitar el trabajo con estos y ajustarlos a los formatos aceptados por las herramientas a utilizar.

En primer lugar, en las medidas de los valores, para aquellos datos que no tenían información, se representaban los nulos mediante “..”, lo cual se ha remplazado por “null” para que las distintas aplicaciones detecten esto como un valor vacío.

Además, para facilitar la comprensión, se ha cambiado el formato de los nombres de las columnas, ya que estos venían con uno distinto al convencional.

Por último, se ha comprobado que existen países dentro del fichero que no presentan ningún valor de medida en ningún año, por ello, se han eliminado estas filas que presenten todos los valores nulos.

Consecuentemente, se ha obtenido una tabla que almacena toda la información por país sobre el consumo eléctrico y el crecimiento económico anual en cada año desde 1992 hasta 2021.

	A	B	C	D	E	F	G
1	Country Name	Country Code	Series Name	Series Code	1992	1993	1994
2	Afghanistan	AFG	Electric power EG.USE.ELEC.*	null	null	null	null
3	Afghanistan	AFG	GDP per capita NY.GDP.PCAP	null	null	null	null
4	Albania	ALB	Electric power EG.USE.ELEC.*	453.6440739	536.3638251	596.407	
5	Albania	ALB	GDP per capita NY.GDP.PCAP	-6.622551212	10.22994935	8.96976	
6	Algeria	DZA	Electric power EG.USE.ELEC.*	559.7830129	547.9904718	550.971	
7	Algeria	DZA	GDP per capita NY.GDP.PCAP	-0.564788753	-4.250986637	-2.93111	
8	American Samoa	ASM	Electric power EG.USE.ELEC.*	null	null	null	
9	American Samoa	ASM	GDP per capita NY.GDP.PCAP	null	null	null	
10	Andorra	AND	Electric power EG.USE.ELEC.*	null	null	null	
11	Andorra	AND	GDP per capita NY.GDP.PCAP	-2.869215672	-4.427065042	-0.39708	
12	Angola	AGO	Electric power EG.USE.ELEC.*	53.48666282	52.00747317	50.6525	
13	Angola	AGO	GDP per capita NY.GDP.PCAP	-8.876931195	-26.41177135	-1.8778	
14	Antigua and Barb	ATG	Electric power EG.USE.ELEC.*	null	null	null	
15	Antigua and Barb	ATG	GDP per capita NY.GDP.PCAP	-0.561528048	3.170343763	4.4162	
16	Argentina	ARG	Electric power EG.USE.ELEC.*	1410.377544	1496.904499	1576.76	
17	Argentina	ARG	GDP per capita NY.GDP.PCAP	6.487631818	6.802926902	4.50545	
18	Armenia	ARM	Electric power EG.USE.ELEC.*	1962.635282	1195.024488	1045.47	
19	Armenia	ARM	GDP per capita NY.GDP.PCAP	-40.74465636	-6.638470109	7.95011	
20	Aruba	ABW	Electric power EG.USE.ELEC.*	null	null	null	
21	Aruba	ABW	GDP per capita NY.GDP.PCAP	0.270153782	1.009406484	2.26506	
22	Australia	AUS	Electric power EG.USE.ELEC.*	8550.328665	8701.760344	8812.48	

## 4. Decisiones iniciales generales

El propósito de la utilización de estas tres aplicaciones es representar la información mediante distintas gráficas y diagramas en paneles de control para poder ver dicha información en su totalidad de forma sencilla y pudiendo filtrar los datos a visualizar directamente. Consecuentemente, se ha optado por realizar los siguientes diagramas para representar la información adecuadamente.

En primer lugar, se mostrará un panel de control para ver la información de manera general, pudiendo ver dos gráficas, una de cada medida, en las cuales se representa el promedio de los datos a lo largo de los años. Además, se incluirán filtros interactivos en los que se podrá filtrar por los países sobre los que se quiere representar las gráficas y cálculos, así como un rango de años específico.

Para el crecimiento económico, se ha decidido utilizar un diagrama lineal para representar de forma adecuada los crecimientos y decrecimientos de cada año con respecto al anterior, al ser este un valor porcentual. Por otro lado, para el consumo energético se ha establecido un diagrama de barras para mostrar el consumo total promedio de los países en cada año.

Por otro lado, dentro de esta primera página, se podrá ver la media total de cada medida sobre los años y países seleccionados en los filtros, gracias a la incorporación de dos gráficas de tarjeta.

En un segundo dashboard, se podrán ver dos rankings de países según los valores medios de cada serie, de forma que se podrán apreciar aquellos con mayor crecimiento económico y mayor consumo energético. Además, se incluye una opción para cambiar el orden del ranking y poder ver aquellos países con menor GDP o consumo. De igual manera que en la página anterior, los resultados podrán ser filtrados por un rango específico de años y por países, para poder ver el orden de un conjunto concreto de ellos.

Finalmente, se incluirá otra página en el panel de control en la que se representa un mapa mundial interactivo donde se apreciará el valor medio de los valores en cada país según tonalidades distintas de colores. Para poder ver los resultados de las dos variables, se incluirá un filtro que cambiará los resultados del mapa entre el crecimiento económico y el consumo eléctrico medio. Además, se podrá incluir un filtro de años para ver solo resultados de un periodo de tiempo concreto en el mapa mundial.

De esta manera, se podrá navegar en las diferentes páginas del panel de control para poder ver toda la información de manera clara y aplicar así filtros de manera sencilla para ver información precisa de forma interactiva.

## 5. Tableau

A continuación, se empleará la aplicación Tableau para realizar las representaciones comentadas anteriormente sobre el consumo energético y el crecimiento económico mundial de cada año. De esta forma, se comenzará importando el fichero Excel explicado antes tras haber realizado su limpieza.

Un aspecto importante es que, de la forma en la que los datos están representados en el fichero, donde los años están uno en cada columna y hay dos entradas de cada país, una para cada variable, se complica la utilización de las herramientas y la manera para trabajar con la información, ya que no se podrían extraer los datos de los años de manera sencilla.

Data	Data	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Country Name	Country Code	Series Name	Series Code	1992	1993	1994	1995	1996	1997	1998
Afghanistan	AFG	Electric power consumption ...	EG.USE.ELEC.KH.PC	null	null	null	null	null	null	null
Afghanistan	AFG	GDP per capita growth (annu...	NY.GDP.PCAP.KD.ZG	null	null	null	null	null	null	null
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	453.644	536.364	596.408	663.784	904.347	694.665	734.85
Albania	ALB	GDP per capita growth (annu...	NY.GDP.PCAP.KD.ZG	-6.62255	10.2299	8.96976	14.0245	9.78018	-10.3611	9.51648
Algeria	DZA	Electric power consumption ...	EG.USE.ELEC.KH.PC	559.783	547.99	550.972	559.918	570.517	575.699	618.493
Algeria	DZA	GDP per capita growth (annu...	NY.GDP.PCAP.KD.ZG	-0.564789	-4.25099	-2.93112	1.83641	2.29082	-0.519902	3.53437
American Samoa	ASM	Electric power consumption ...	EG.USE.ELEC.KH.PC	null	null	null	null	null	null	null
American Samoa	ASM	GDP per capita growth (annu...	NY.GDP.PCAP.KD.ZG	null	null	null	null	null	null	null
Andorra	AND	Electric power consumption ...	EG.USE.ELEC.KH.PC	null	null	null	null	null	null	null
Andorra	AND	GDP per capita growth (annu...	NY.GDP.PCAP.KD.ZG	-2.86922	-4.42707	-0.397087	0.85232	3.8319	9.14398	3.48118

De esta forma, será necesario hacer una modificación sobre la tabla importada. Tableau permite realizar el cambio necesario, transformando las columnas de los años en una sola, donde para cada país haya tantas filas como años. Consecuentemente, para realizar esto, se seleccionará dentro de la aplicación aquellas columnas que se desean modificar, y se escoge la opción pivotar. De esta forma, la tabla con la información quedará con el formato adecuado y fácil para trabajar.

Data	Data	Abc	Abc	#	#
Country Name	Country Code	Series Name	Series Code	Pivotar	Pivotar
				Año	Valores
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1992	453.64
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1993	536.36
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1994	596.41
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1995	663.78
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1996	904.35
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1997	694.66
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1998	734.85
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	1999	1414.70
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	2000	1449.65
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	2001	1351.23
Albania	ALB	Electric power consumption ...	EG.USE.ELEC.KH.PC	2002	1578.17

Tras esto, se procederá a la creación de cada una de las gráficas que se incluirán en los Dashboards que se pretenden realizar (comentados anteriormente).

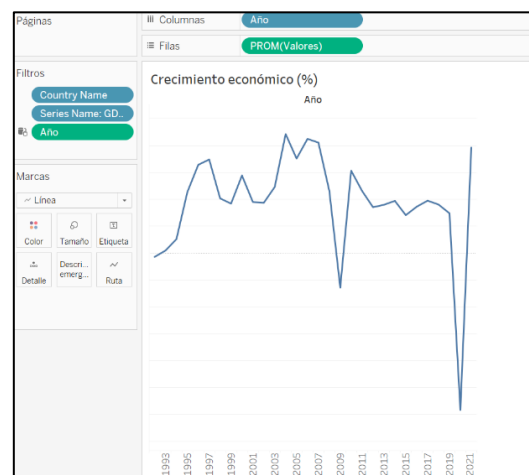
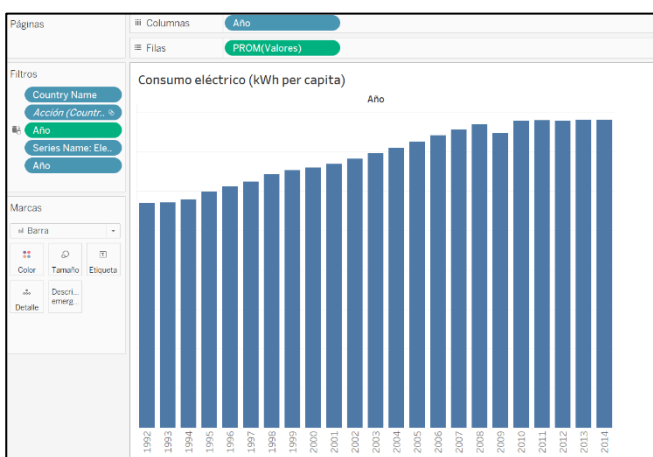
Tableau muestra en la parte izquierda de la aplicación las distintas columnas de la tabla, de forma que, para la creación de tablas y la aplicación de filtros, simplemente será necesario arrastrar aquellas columnas sobre las que se quiere hacer la gráfica para poder crearla.

- **DASHBOARD 1: ANÁLISIS GENERAL. CRECIMIENTO DE LA ECONOMÍA Y CONSUMO ELÉCTRICO**

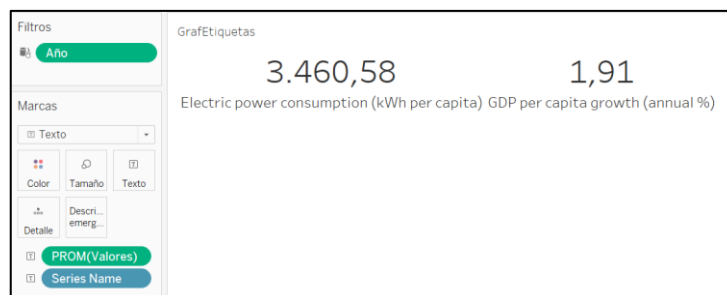
En primer lugar, es importante mencionar que todas las gráficas, tanto del promedio de consumo eléctrico como de crecimiento económico, estarán en función del valor de cada campo y del año correspondiente, por lo que se incluirá esta última variable en la sección de columnas, y el valor en la fila.

Además, como se pretende calcular el promedio general para todos los países, se especificará en el valor, que se use el promedio. Finalmente, se crearán una serie de segmentaciones de datos para que posteriormente en el Dashboard se puedan filtrar los resultados según un rango específico de años, el país, y además un filtro para que cada gráfica sea de la variable correspondiente.

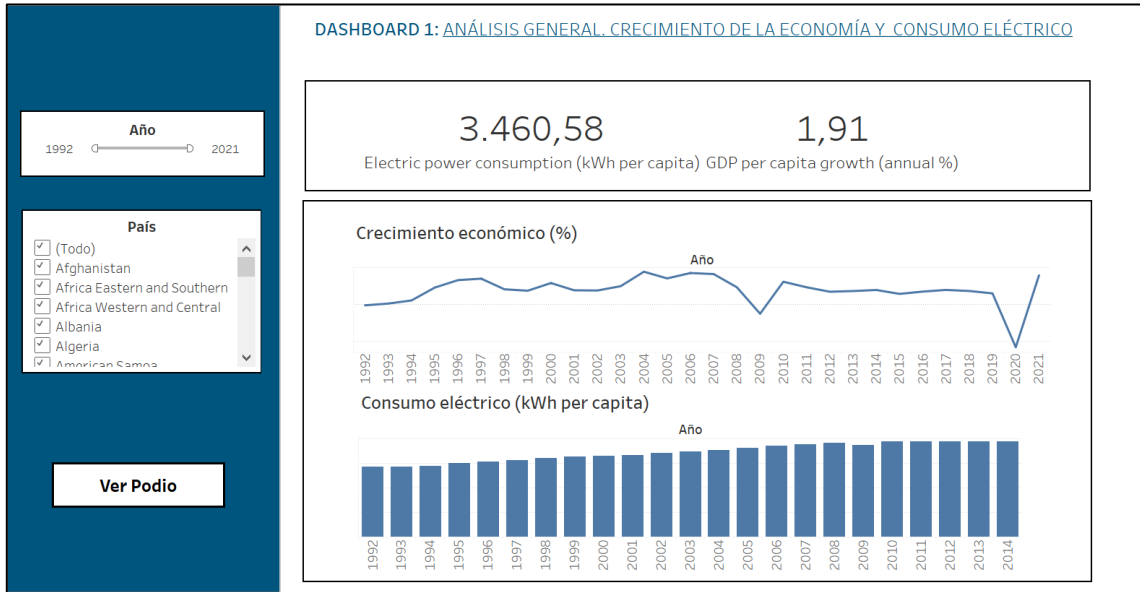
Cabe mencionar que, para el caso del consumo eléctrico, se han limitado los años hasta 2015 ya que a partir de ese año no se tenían datos de ningún país. De esta forma, las hojas para la creación de ambas gráficas quedarían de la siguiente forma.



Además, para la primera página del Dashboard, se incluirá también una gráfica de etiquetas en la que se muestra el promedio total del consumo y del crecimiento económico a lo largo de los años, para poder ver un resumen general de toda la información.



Tras la creación de estos tres diagramas, se podrá elaborar el primer Dashboard, donde aparecerán estas graficas más una sección en la que poder filtrar los resultados según un rango de años y países en concreto, funcionando de manera interactiva.



- DASHBOARD 2: RANKING POR CRECIMIENTO ECONÓMICO Y CONSUMO**

Para el resto de las páginas asociadas a la tarea de visualización a realizar, se han creado los rankings de resultados según el crecimiento económico y el consumo eléctrico. Para ello, se han definido 4 hojas con gráficas para cada variable, quedando un total de 8 diagramas. En ellos, se ha incluido un ranking ordenado con el promedio de cada país en cada una de las variables, tanto en orden descendente como ascendente, y una gráfica en la que únicamente se refleja un top 3 con los países que están en los tres primeros puertos de ese respectivo ranking.

Para la creación de los rankings, se ha utilizado la opción de Tableau para crear parámetros calculados. De esta forma, se ha utilizado la función Rank\_Dense para crear dicho ranking, indicando la opción ‘desc’ o ‘asc’ dependiendo de si se quiere hacer el ranking de primeros o de últimos respectivamente. La opción Dense, lo que hará será eliminar los valores repetidos y nulos.

<div>Ranking</div> <div>Los resultados se procesan en Tabla (a lo largo).</div> <div><code>RANK_DENSE([Promedio consumo], 'desc')</code></div>	<div>Ranking ult</div> <div>Los resultados se procesan en Tabla (a lo largo).</div> <div><code>RANK_DENSE([Promedio consumo], 'asc')</code></div>
--	---

Además, se han incluido las segmentaciones de datos necesarias en cada caso para permitir filtrar por variable a medir: países, en el caso de que se quiera ver un subconjunto específico, y años , para poder ver los países con mayor o menor valor en un periodo de tiempo concreto. De esta forma, las tablas para ver los rankings seguirían el siguiente formato, siendo el procedimiento de manera análoga para ambas variables y para los rankings de últimos y primeros.



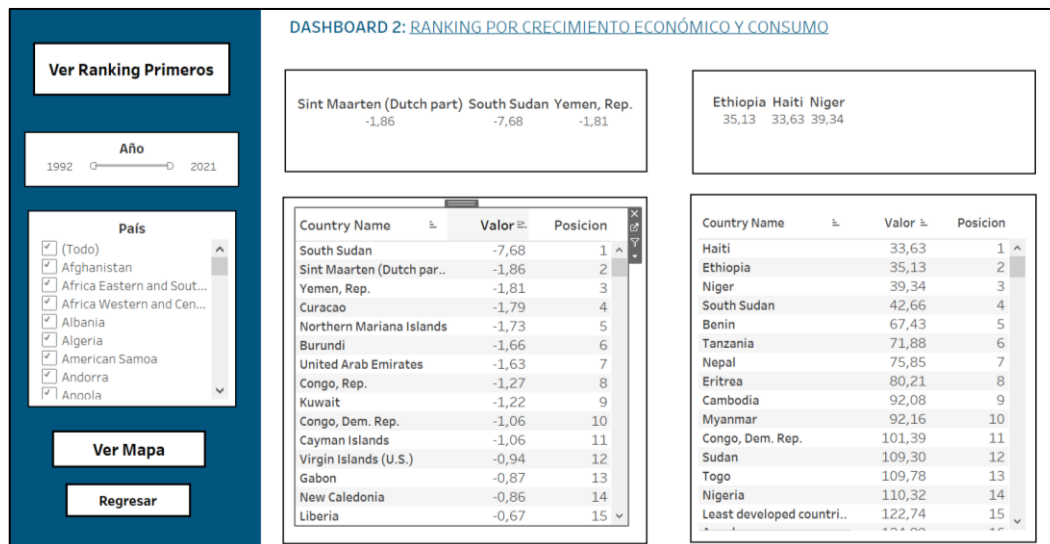
Páginas	Columnas	Nombres de medid...
Filtros	Filas	Country Name
Ranking	RankDesc (consumo)	
Acción (Countr..)		
Año		
Nombres de m...		
Series Name: Ele...		
Country Name		
Marcas		
Automático		
Color		
Tamaño		
Texto		
Detalle		
Valores de me...		
Valores de medidas		
PROM(Valores)		
Ranking		

Páginas	Columnas	
Filtros	Filas	
Acción (Countr..)	Top3Desc(consumo)	
Año		
Series Name: Ele...		
Ranking		
Marcas		
Automático		
Color		
Tamaño		
Texto		
Detalle		
Country Na...		
PROM(Valores)		

Así, se podrá crear el Dashboard correspondiente, en el cual se permitirá alternar entre el ranking de últimos y de primeros mediante un botón. Además, se incluirán en la parte izquierda los filtros de segmentación tanto de países como de años. Así mismo, se podrá visualizar el promedio de cada variable de todos los países, así como el puesto que ocupan en el ranking, apareciendo ordenados del primero al último, y viéndose en la parte superior el top 3 de países que ocuparían el podio.

Ver Ranking Últimos	DASHBOARD 2: RANKING POR CRECIMIENTO ECONÓMICO Y CONSUMO	
Año	Equatorial Guinea Bosnia and Herzegovina China	Iceland Norway Bahrain
1992 2021	12,69 9,75 8,50	32.588,29 24.347,81 19.789,24
País	Country Name Valor Posición	Country Name Valor Posición
<input checked="" type="checkbox"/> (Todo)	Equatorial Guinea 12,7 1	Iceland 32.588,29 1
<input checked="" type="checkbox"/> Afghanistan	Bosnia and Herzegovina 9,8 2	Norway 24.347,81 2
<input checked="" type="checkbox"/> Africa Eastern and Sout...	China 8,5 3	Bahrain 19.789,24 3
<input checked="" type="checkbox"/> Africa Western and Cent...	East Asia & Pacific (excl.. 7,1 4	Canada 16.376,09 4
<input checked="" type="checkbox"/> Albania	East Asia & Pacific (IDA .. 7,1 5	Kuwait 15.587,27 5
<input checked="" type="checkbox"/> Algeria	Myanmar 6,9 6	Finland 15.301,32 6
<input checked="" type="checkbox"/> American Samoa	Nauru 6,2 7	Luxembourg 15.140,39 7
<input checked="" type="checkbox"/> Andorra	Iraq 5,5 8	Sweden 15.105,06 8
<input checked="" type="checkbox"/> Angola	Vietnam 5,5 9	Qatar 13.772,47 9
	Guyana 5,3 10	North America 13.428,01 10
	Lithuania 5,2 11	United States 13.104,67 11
	Albania 5,0 12	United Arab Emirates 10.905,66 12
	Late-demographic divid.. 4,9 13	Australia 10.062,09 13
	Latvia 4,8 14	New Zealand 9.300,34 14
	Lao PDR 4,7 15	Post-demographic divid.. 8.692,79 15
		High income 8.584,60 16
		Japan 8.050,02 17
Ver Mapa		
Regresar		

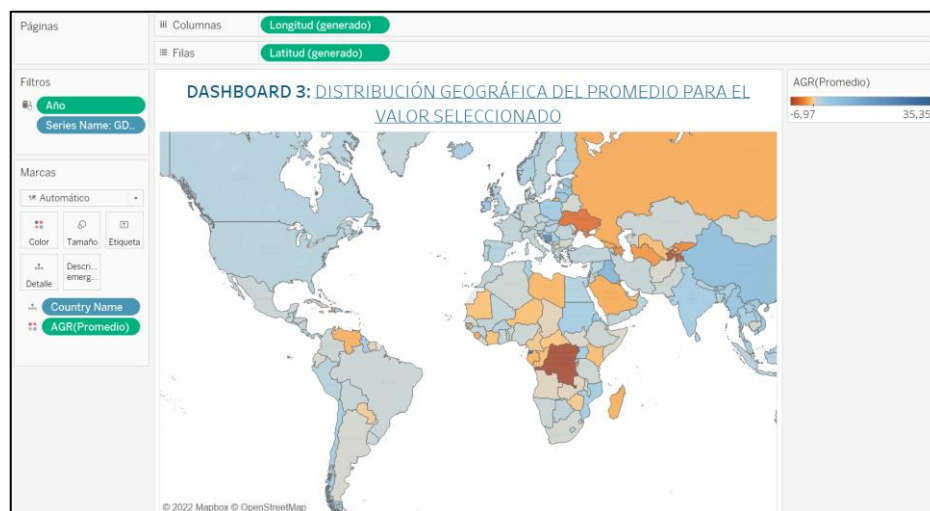
Ranking de primeros



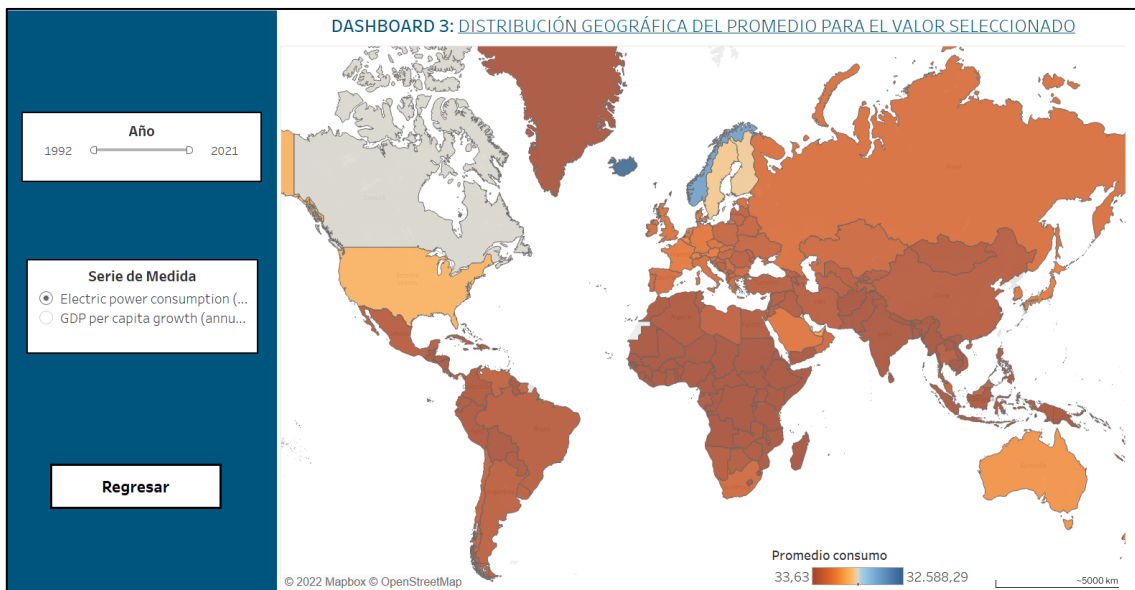
Ranking de últimos

- **DASHBOARD 3: DISTRIBUCIÓN GEOGRÁFICA DEL PROMEDIO PARA EL VALOR SELECCIONADO**

Finalmente, para la creación del Dashboard que representa el mapa mundial con los resultados de cada país, se ha utilizado la herramienta de Tableau para incluir mapas. Estos se elaboran con los nombres de los países obtenidos de los datos importados. De esta forma, se incluye en las medidas el promedio de los valores, y se incorporan filtros de la variable a medir, para permitir ver las dos existentes sobre el mapa, permitiendo escoger los años concretos sobre los que se quieren ver los resultados. Consecuentemente, la hoja Tableau asociada a esta creación sería el siguiente:



En el Dashboard correspondiente, se han incluido segmentaciones para poder cambiar entre las dos variables sobre las que se hacen las medidas, de forma que se aprecia como se va modificando el mapa en cada caso. Además, también está la elección del rango de igual manera que en los Dashboards anteriores.



## • DASHBOARDS ADVENTURE WORKS

A continuación, se utilizará la base de datos “Adventure Works” para realizar una serie de vistas que permitan visualizar alguna información relevante, como el número de clientes distribuidos geográficamente o las ventas efectuadas por cada divisa.

Para ello, será necesario realizar una conexión con la base de datos desde Tableau. De esta forma, dentro de un nuevo Dashboard, se seleccionará la opción del menú superior “Datos”, y se indicará “nueva fuente de datos”, donde se escogerá la opción de **postgresql** ya que será el tipo de base de datos a incluir. A continuación, se incluye la configuración que permitirá realizar la conexión desde la aplicación a la BBDD.

PostgreSQL

General SQL inicial

Servidor  
localhost

Puerto  
5432

Base de datos  
Adventureworks

Autenticación  
Nombre de usuario y contraseña

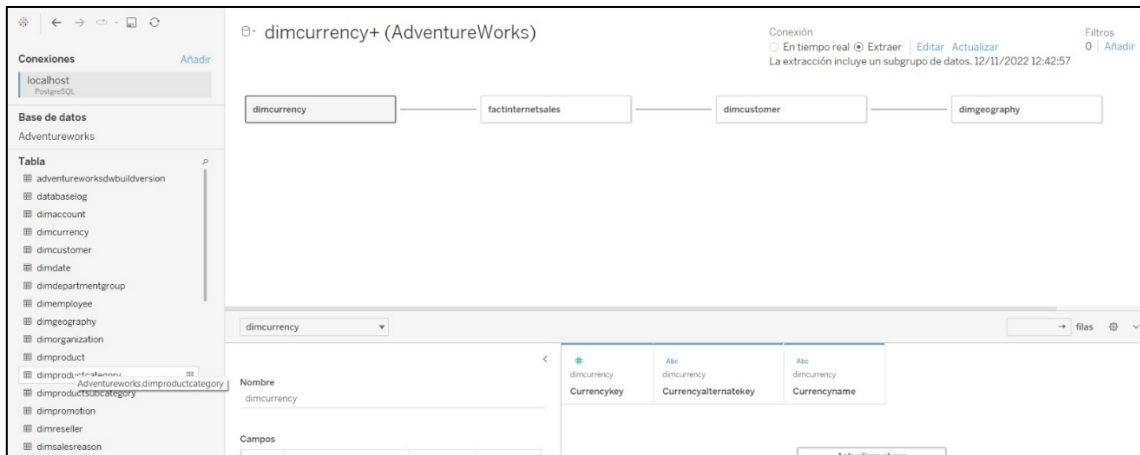
Nombre de usuario  
postgres

Contraseña  
\*\*\*\*\*

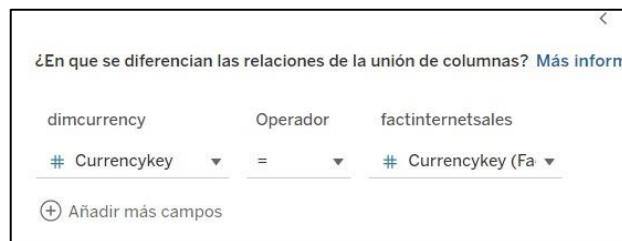
☐ Requiere SSL

Iniciar sesión

Tras realizar esto, la base de datos se habrá cargado dentro de la aplicación Tableau, en donde se podrá apreciar a mano izquierda una vista con todas las tablas que esta contiene. De esta forma, se seleccionarán aquellas sobre las cuales se querrá extraer información para representar los distintos diagramas, arrastrándolos al panel central.

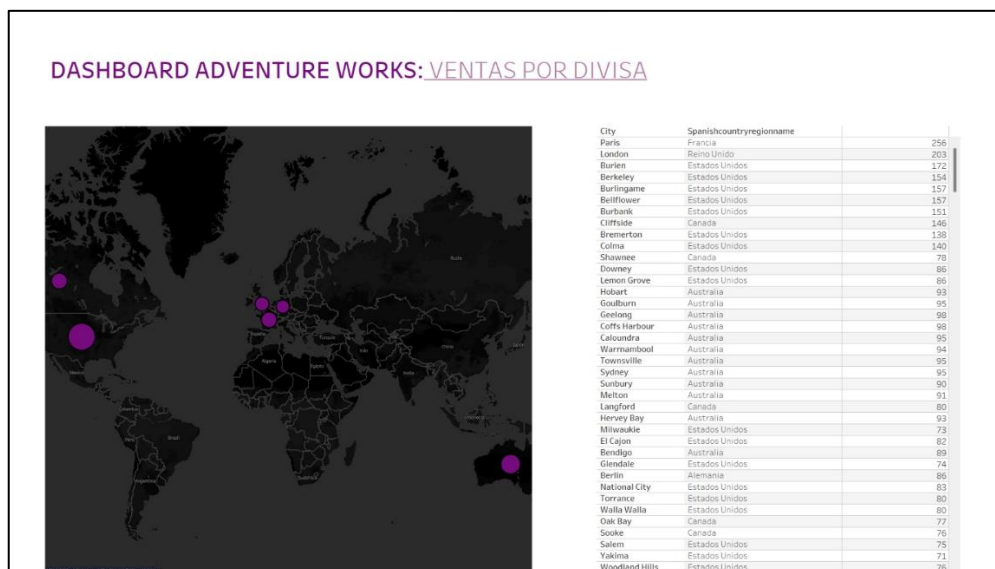


Además, cabe mencionar que, para realizar la unión de las tablas de manera correcta, será necesario establecer las relaciones de claves de cada una, especificando aquellos atributos que funcionarán como tal (se explicará en más detalle en la realización de este ejercicio en Power Bi).



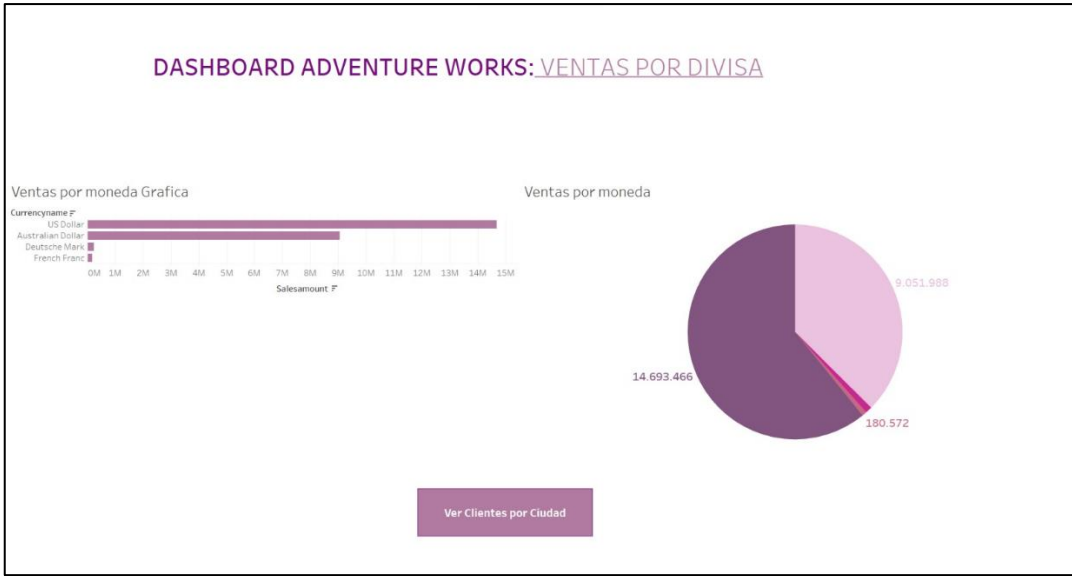
Finalmente, una vez se tienen todas las tablas que contiene la información que se pretende utilizar, se procederá a crear los distintos Dashboards.

En primer lugar, se muestra un panel con el número total de clientes por ciudad, representando esta información en una tabla con cada una de las ciudades registradas, y un mapa en el que se pueden ver los distintos países y su número de ventas totales, realizando la suma de cada una de las ciudades de la tabla.



Además, se ha creado otro Dashboard en el que se representa, mediante un gráfico de barras horizontal y un gráfico circular, los valores de ventas por cada divisa y en qué proporción se

distribuyen dichas ventas entre ellas, donde se puede ver que la gran mayoría de ellas son producidas a través del dólar estadounidense y el dólar australiano.



## 6. Power Bi

Para la elaboración de los diferentes Dashboards asociados a esta herramienta se parte de las visualizaciones anteriormente modeladas en Tableau. Además, para poder realizar las diferentes gráficas, se parte también del fichero Excel anteriormente mencionado con los datos limpios.

Al igual que para el resto de las herramientas utilizadas en esta práctica, el formato en el que se disponen los datos dificulta la realización de representaciones que busquen analizar el impacto y la variación de la información a lo largo de los años.

Por ello, antes de cargar los datos se pulsa sobre el botón de transformar datos para acceder al editor de Power Query. Esta herramienta incorporada en Power Bi se utiliza para manipular los datos cargados, adaptándolos para que su uso sea cómodo y correcto. En este caso, se seleccionan las columnas asociadas a los años y, pulsando sobre el botón derecho, se selecciona la opción de “Anulación de dinamización de columnas”, obteniendo la siguiente tabla:

Country Name	Series Code	Año	Valor
Trinidad and Tobago	NY.GDP.PCAP.KD.ZG	2014	-1,48
Serbia	NY.GDP.PCAP.KD.ZG	2014	-1,13
Turkiye	NY.GDP.PCAP.KD.ZG	2014	3,17
Monaco	NY.GDP.PCAP.KD.ZG	2014	5,98
Guinea	NY.GDP.PCAP.KD.ZG	2014	1,30
New Zealand	NY.GDP.PCAP.KD.ZG	2014	2,11
Turkmenistan	NY.GDP.PCAP.KD.ZG	2014	8,28
Guinea-Bissau	NY.GDP.PCAP.KD.ZG	2014	-1,67
Lao PDR	NY.GDP.PCAP.KD.ZG	2014	6,02
Isle of Man	NY.GDP.PCAP.KD.ZG	2014	5,78

A continuación, se selecciona la opción de “Usar la primera fila como encabezado” y, por último, se comprueba que el tipo de dato asociado a cada columna es el correcto. Finalmente, se pulsa sobre cerrar y aplicar; obteniendo así la tabla que se va a emplear para realizar las diferentes gráficas.

Para la realización de las diferentes visualizaciones, se crean una serie de medidas DAX. Este es un lenguaje de programación que se utiliza en Microsoft Power BI para crear columnas calculadas, medidas y tablas personalizadas. Así, se han definido las siguientes:

- promedio consumo = `CALCULATE(AVERAGE('Datos traspuestos'[Valor]), 'Datos traspuestos'[Series Code]="EG.USE.ELEC.KH.PC")`
  - Básicamente, calcula mediante la función “average” el promedio del consumo asociado a la tabla. Para ello, se realiza un filtrado de los datos de entrada utilizando como campo el código de serie asociado al consumo.
- Promedio de crecimiento económico = `CALCULATE(DIVIDE(AVERAGE('Datos traspuestos'[Valor]),100), 'Datos traspuestos'[Serie Code]="NY.GDP.PCAP.KD.ZG")`
  - Al igual que en el caso anterior, se calcula el promedio de crecimiento económico. En este caso, se añade además la conversión correspondiente del cálculo a un formato de porcentaje.
- contador Usuarios = `count('public dimcustomer'[firstname])`
  - Este cálculo se corresponde a los DashBoards elaborados sobre la base de datos de Adventure Works. Hace referencia al recuento del número de usuarios asociados a la tabla dimcustomer. Para realizar dicha cuenta se hace uso de la columna firstname.

A continuación, se relatan el proceso de elaboración de los diferentes Dashboards:

- **DASHBOARD 1: ANÁLISIS GENERAL. CRECIMIENTO DE LA ECONOMÍA Y CONSUMO ELÉCTRICO**

En este informe se recoge un análisis general sobre los datos recogidos en el fichero Excel. En primer lugar, se llevan a cabo un gráfico de líneas y otro de barras apiladas para mostrar respectivamente el crecimiento económico y el consumo eléctrico a lo largo de los años.

Para ello, se añade sobre cada uno de los objetos visuales un filtrado por el número de serie, representando así en el eje X los años y en el eje Y la columna ‘Valor’ sobre la que se aplicará el filtro en la sección Filtros (lateral derecho).

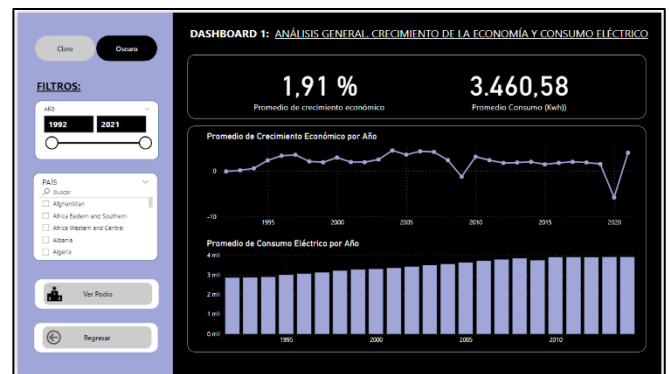
Se incorporarán además dos gráficas de tarjetas para mostrar el promedio total asociado a cada uno de los valores representados a lo largo de los años. Para este tipo de tarjetas se hace uso de las medidas DAX anteriormente descritas.

Para mayor visibilidad de los datos, se incorporan dos gráficos de segmentación en el lateral izquierdo para poder visualizar los datos en función de los años y países.

Además, se incluyen una serie de botones para añadir interactividad a la gráfica:

- **Claro/Oscuro:** permite cambiar los colores del informe según las preferencias del usuario. Para realizar esta tarea se ha duplicado la página asociada al Dashboard principal y se han modificado manualmente los colores para, finalmente, aplicar a los botones de Claro / Oscuro la opción de navegar entre páginas. Por último, se oculta la página de oscuro para que al publicar el informe en Power Bi solo aparezca la primera de las páginas.
- **Ver podio:** este botón se ha creado para poder navegar a la página del DashBoard asociada al podio (se explicará en la siguiente sección).
- **Regresar:** se define un botón que permita volver a la página anteriormente visitada.

Finalmente, se le aplica el estilo deseado a las diferentes gráficas y botones presentes en el informe, modificando los colores, los bordes, el fondo... y añadiendo formas e iconos para llevar a cabo un Dashboard más atractivo.



## • DASHBOARD 2: RANKING POR CRECIMIENTO ECONÓMICO Y CONSUMO

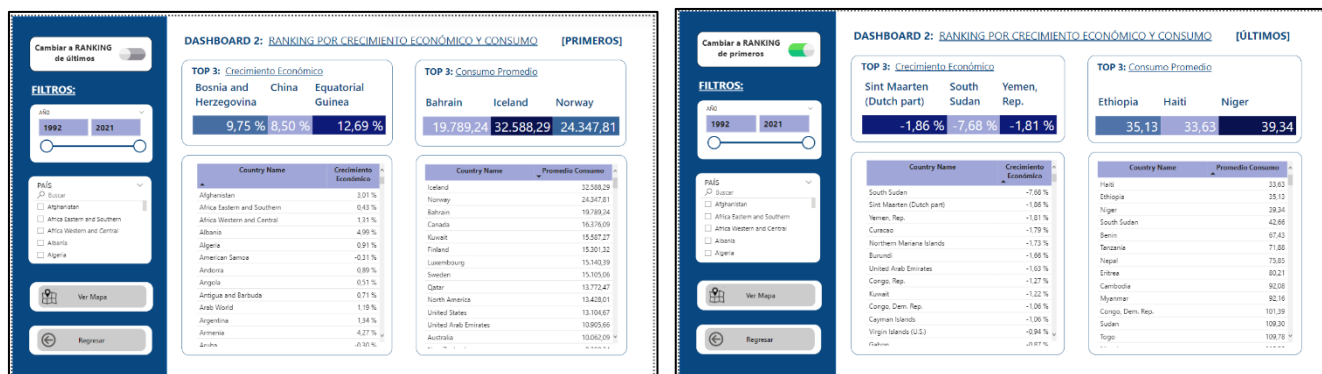
Para este segundo Dashboard se ha decidido crear un ranking para poder visualizar los valores de crecimiento económico y de consumo asociado a cada país. De este modo, se han elaborado dos tablas sobre las que se representan los países y las medidas DAX elaboradas para cada métrica.

Además, se han creado dos matrices para crear un TOP 3. Para su elaboración se han seguido los siguientes pasos:

1. Introducir en las columnas el valor de country name y en los valores la medida DAX asociada a cada top.
2. Asignar un filtro de figura sobre el que se define la opción de TOP N, empleando como valor la medida DAX correspondiente. Se escoge top 3 superior o inferior según sea el caso de primeros o de últimos.
3. Definir un formato condicional para representar mediante una escala de colores las posiciones dentro del ranking.

Al igual que para el anterior informe se incluyen una serie de gráficos de segmentación de datos para poder filtrar y varios elementos interactivos que permiten navegar entre las distintas páginas. Así se introduce un botón para cambiar el ranking de los primeros de la lista a los últimos, un icono para navegar fácilmente al siguiente Dashboard (Mapa) y un botón de regreso a la anterior visualización.



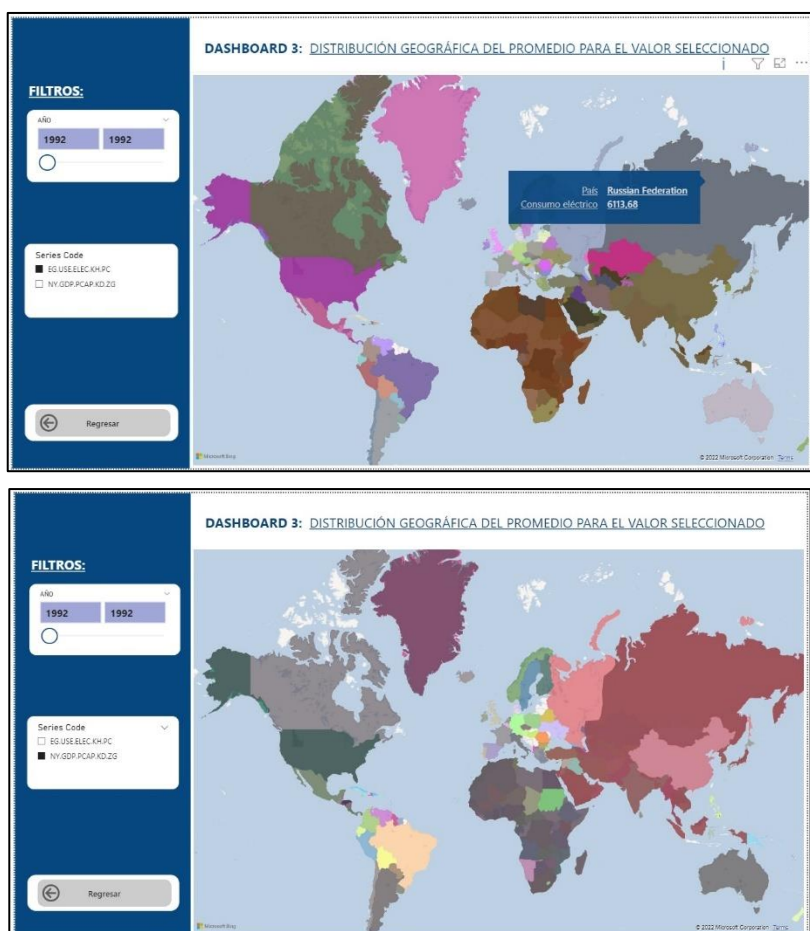


- DASHBOARD 3: DISTRIBUCIÓN GEOGRÁFICA DEL PROMEDIO PARA EL VALOR SELECCIONADO**

Con el siguiente Dashboard se busca llevar a cabo una representación geográfica de los valores manejados a lo largo de las representaciones realizadas (promedio de la columna valor).

Dado que las gráficas de mapas en este programa afectan negativamente al rendimiento, se opta por realizar una única representación, cuyos valores pueden modificarse fácilmente seleccionando la serie deseada en la segmentación de datos correspondiente.

Además, se incorpora la posibilidad de llevar a cabo un filtrado por años para poder realizar análisis más específicos.





- **Representaciones en R y Python**

Para Power Bi se solicita llevar a cabo al menos una representación en lenguaje Python y otra en lenguaje R. De este modo, es necesario descargar una serie de paquetes para garantizar el funcionamiento de las mismas. En lo relativo a las gráficas Python, es necesario descargar dicho lenguaje en el equipo. Para hacerlo se hace uso del gestor de paquetes Chocolatey, mediante el comando PowerShell **choco install python**.

Además, dado que se va a hacer uso de las librerías Pandas y matplotlib se instalan por medio de los siguientes comandos pip: **pip install matplotlib, pip install pandas**.

Una vez instalados todos los archivos necesarios, solo queda indicar la ubicación de instalación de python en los ajustes de Power Bi (menú de creación de scripts de python). A continuación, se procede a elaborar las gráficas deseadas.

En este caso, se lleva a cabo un Dashboard en el que se incluyen, en la parte superior tres gráficas elaboradas en Power Bi referentes al mínimo, promedio y máximo de consumo a lo largo de los años, y en la parte inferior las mismas representaciones creadas con python.

Para desarrollar visualizaciones en este lenguaje se escoge el icono “Py” en la sección de visualizaciones. A continuación, se arrastran los valores que se desean graficar (años y cálculo de la columna valor). Por último, se introduce el script que grafique los datasets introducidos:

```
#dataset = pandas.DataFrame(Año,Valor)
#dataset = dataset.drop_duplicates()
# Pegue o escriba aquí el código de script:
import matplotlib.pyplot as plt
import pandas as pd
dataset.plot(kind='line',x='Año',y='Valor', color='#893395',linewidth=4)
plt.show()
```

En esta página se opta por aplicar un **filtro de página** para establecer el número de serie asociado al consumo. Además, se incluye una segmentación de datos para filtrar por años y un botón con el logo de R para acceder a las representaciones de este lenguaje.

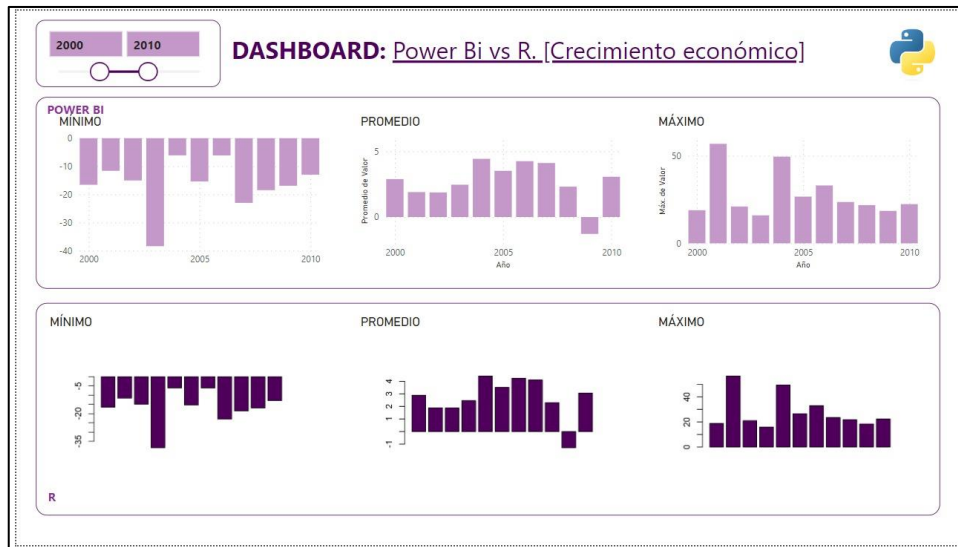
---

En relación a las gráficas de R, se sigue la misma estrategia que para python; representando esta vez el crecimiento económico. Para poder realizar las gráficas es necesario descargar R desde la web oficial e incluir su ruta en la configuración de Power Bi.

De este modo, se elaboran tres gráficos de columna representando mínimo, promedio y máximo del crecimiento económico; realizando la comparación entre la elaboración en Python y en Power Bi. El código R necesario es el siguiente:

```
# dataset <- data.frame(CountryName, Valor)
# dataset <- unique(dataset)
# Pegue o escriba aquí el código de script:
barplot(dataset$Valor,dataset$Año,col="#50005C")
```

Así mismo, se incluye un gráfico de segmentación y un acceso directo a las visualizaciones de Python. El resultado obtenido es el siguiente:



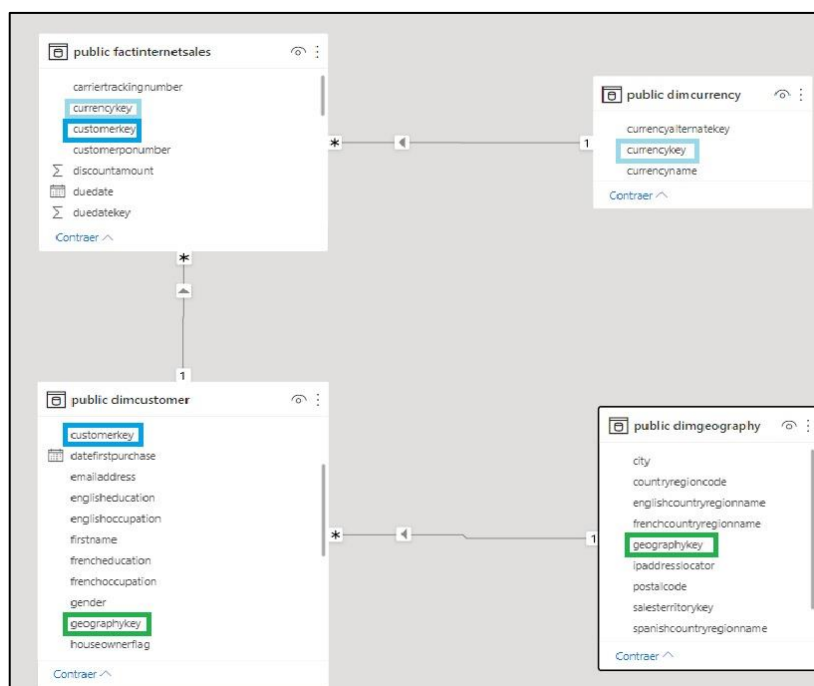
- **Gráfico de Burbujas**

A continuación, se realizará un gráfico de burbujas en el que se representará

- **DASHBOARDS Adventure Works**

Para continuar con las tareas a realizar en Power Bi, se pide hacer uso de la base de datos Postgres **Adventure Works**. Para ello, se recupera la base de datos utilizada en la práctica 1 y se lleva a cabo la conexión pulsando Obtener datos > Base de datos PostgreSQL. A continuación, se definen los parámetros de configuración (host: localhost, puerto: 5432, base de datos Adventure Works...)

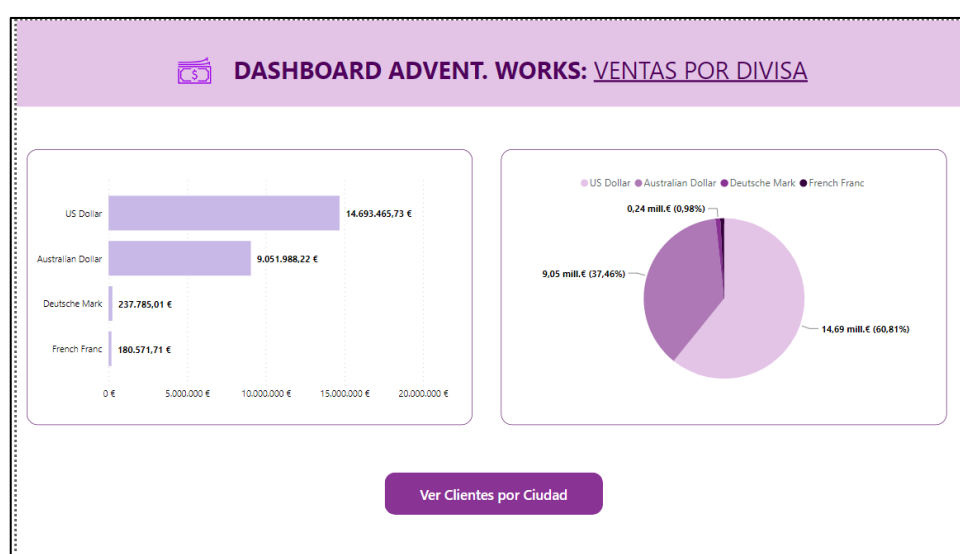
Una vez establecida la conexión se seleccionan las tablas a utilizar y se cargan los datos. Como se ve en la vista de **modelo**, las relaciones entre las tablas se construyen automáticamente de forma correcta mediante las claves de cada una de las tablas. Así, se concluye que las gráficas se pueden elaborar sin ningún problema de relaciones entre atributos. No obstante, en otras herramientas como Tableau es necesario tener en cuenta el orden en el que se pueden relacionar las distintas tablas; indicando explícitamente los campos a utilizar para establecer la relación:



Una vez importados los datos se elaboran dos Dashboards con el fin de comprender los datos almacenados en la BBDD sobre la que se está trabajando.

En el primero de ellos se recogerán dos gráficas referentes a la **suma** del importe de ventas por divisa. Los campos a representar son “**currencyname**” de la tabla “dimcurrency” y “**salesamount**” de “factinternetsales”.

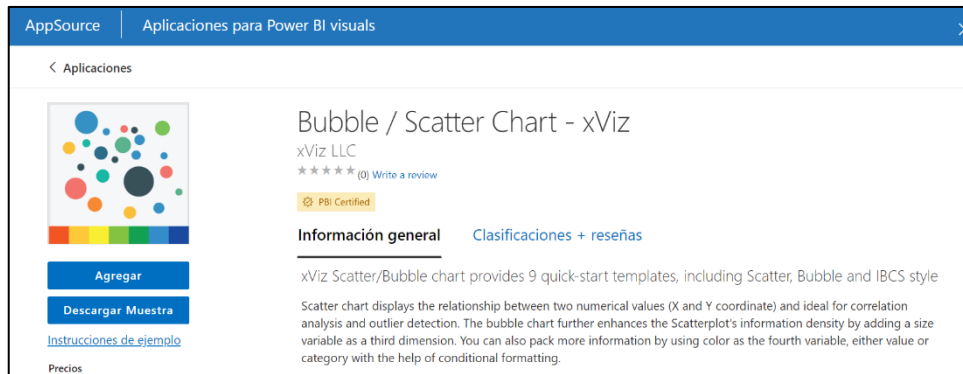
De este modo, se incorporarán filtros para eliminar todos aquellos valores que contengan la subcadena “**TRIAL**” pues son datos que no se desean representar. Además, se incluye un botón para poder acceder al resto de visualizaciones asociadas a esta base de datos. El informe obtenido es el siguiente:



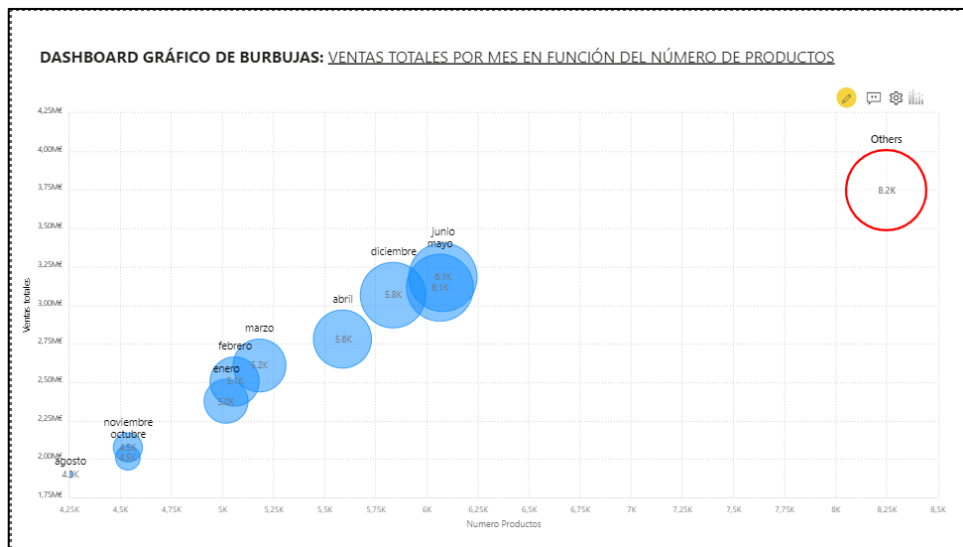
En relación a la última de las vistas asociadas a la BBDD Adventure Works, se lleva a cabo una visualización en la que se representarán los campos “**spanishcountryregionname**” y “**city**” de

la tabla dimgeography y la medida DAX **contador de usuarios**. Nuevamente, se introduce un filtro de página que elimine las entradas con la subcadena “TRIAL”. En este dashboard se incluyen un mapa coroplético que muestre en los globos el recuento de usuarios, una tabla que recoja el número de clientes por ciudad y país y un gráfico de segmentación que permita filtrar por país.

Para concluir con la sección de Power Bi se procede a llevar a cabo un gráfico de burbujas. Para ello, se pulsa sobre los 3 puntos ubicados en la zona de gráficos y se busca bubble en la tienda. Se escoge utilizar el siguiente componente; para ello, se pulsa en agregar:



Tras crear un nuevo Dashboard en el que se incluirá este gráfico de burbujas, se seleccionarán los datos a partir de los cuales se elaborará dicho diagrama. En concreto, se ha decidido que las burbujas representen los meses en los que se realizan las ventas, midiendo el número total de ventas en función del número de productos. De esta forma, aquellos meses con mayor número de ventas estarán simbolizados por una burbuja de mayor tamaño, tal y como se puede ver en el siguiente diagrama:



Para la realización del gráfico ha sido necesario crear una nueva columna calculada sobre la propia tabla (pulsando con el botón derecho sobre la misma en la sección de Campos). Esta servirá para identificar el número de productos; realizando a división de salesamount entre unit price. Así, finalmente se utilizarán las columnas: Número de productos, salesamount y orderdate(month).

## 7. Apache Superset

Para la realización de las visualizaciones asociadas a esta aplicación es necesario llevar a cabo una serie de pasos previos:

- En primer lugar, no es posible partir del fichero excel original dado que en esta plataforma no existe la opción de pivotar las columnas. Por ello, se opta por extraer los datos de Power Bi y modificarlos para poder realizar las tareas necesarias en esta plataforma.
- Así, para exportar los datos se crea una tabla básica en Power Bi sobre la que se introducen todos los campos del fichero pivotado. Posteriormente, se pulsa sobre los tres puntos y se selecciona la opción de exportar datos.
- Antes de inportar el csv generado a Superset, es necesario realizar un último cambio para el que se hará uso de Microsoft Excel. Es necesario sustituir de la columna valor todas las comas por puntos para que el intérprete de esta plataforma escoja el tipo de dato adecuado para cada variable.
- Otro de los cambios que es necesario describir está relacionado con el campo Año del fichero importado. A la hora de crear el dataset es importante indicar que es una variable temporal con formato %Y para que el sistema permita realizar todas las gráficas deseadas.

A continuación, se realizarán los ejercicios anteriores, pero utilizando la aplicación Apache superset.

Una de las particularidades de esta aplicación es que, de igual manera que en Tableau, es necesario crear cada una de las gráficas de manera independiente para posteriormente incluirlas en un dashboard. De esta manera, se procederá a la creación de cada uno de los Dashboards realizados en las anteriores aplicaciones.

Para este primer Dashboard, se incluirán dos gráficas de etiquetas para reflejar el promedio general de cada variable, además de una gráfica de barras para el consumo eléctrico y un diagrama de línea para el porcentaje de crecimiento económico de cada año.

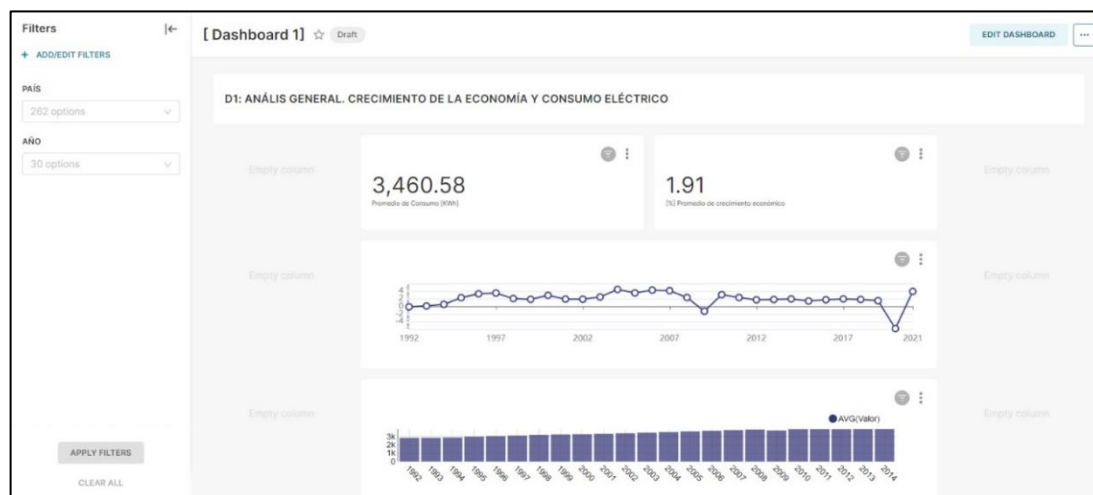
De esta manera, para crear cada gráfica, se seleccionará el botón “más” y se escogerá el tipo de gráfica deseado. Además, se escoge el Dataset sobre el que se va a actuar, que en este caso es el asociado al fichero csv que contiene la información con la que se está trabajando. A continuación, una vez cargado el Dataset, de igual forma que en el resto de las aplicaciones se incluye cada valor de la tabla en el campo en el que se quiere representar la información en la gráfica.

Consecuentemente, para este primer Dashboard se han creado dos gráficas de etiquetas, incluyendo la media de los valores de cada variable, y una gráfica de barras para el consumo eléctrico, así como otra gráfica para el GDP por años. Además, de igual manera que en los otros paneles de control, se incluyen en la parte izquierda de la página unos filtros para poder seleccionar tanto el rango de años como los países sobre los que se desea mostrar los resultados.

Para la creación de los Dashboards, se ha seleccionado la opción “layout element” para incluir elementos que permitan modificar el diseño del mismo. Consecuentemente, para poder distribuir la información que se muestra, se han creado columnas para poder centrar las distintas gráficas y posicionarlas correctamente. Además, se han incluido otros elementos como cuadros de texto para incluir el título del Dashboard. Finalmente, para poder eliminar los títulos de cada gráfica, se ha editado el Dashboard para modificar el css e introducir el siguiente comando:

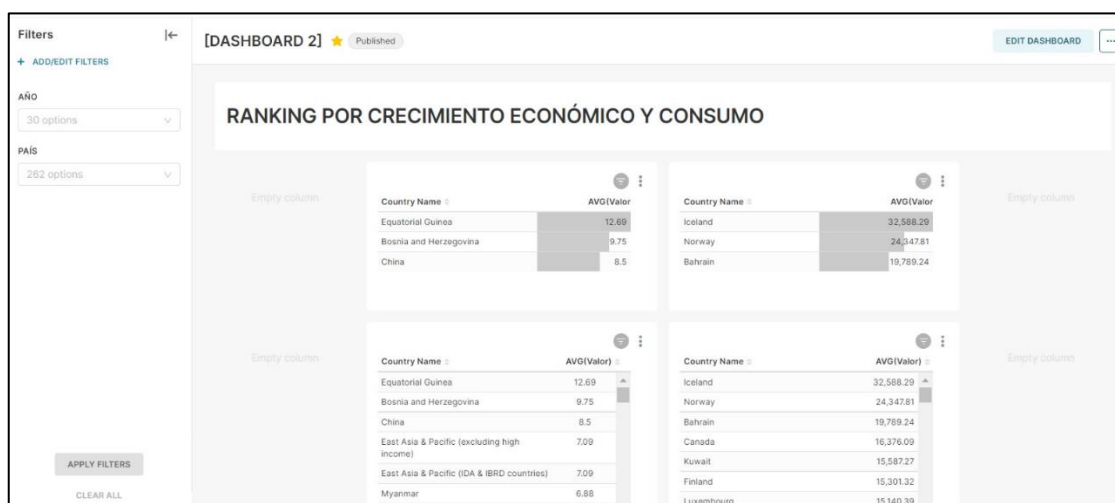
```
.header-title > .editable-title {  
    display:none;
```

}



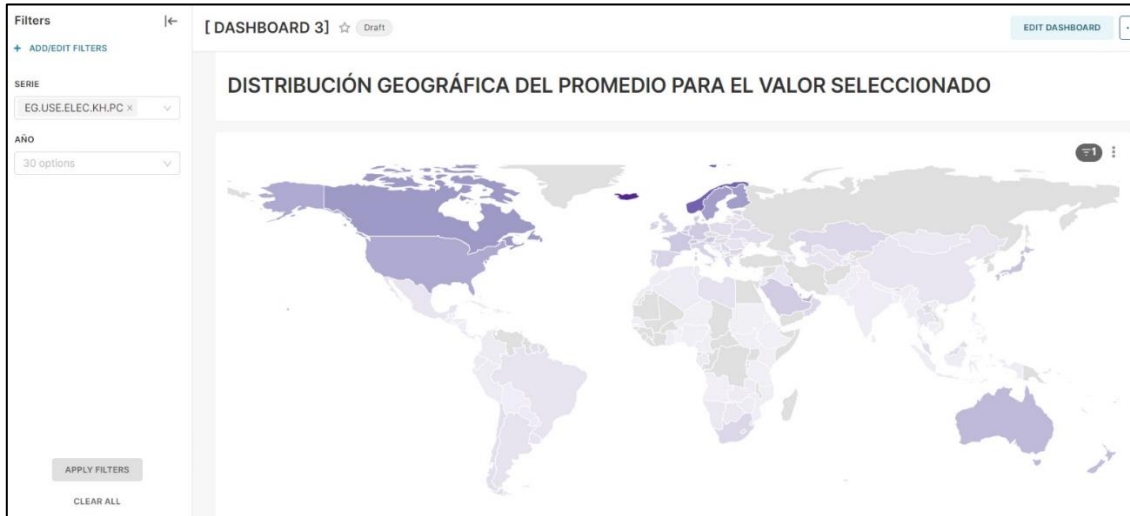
Para la creación del Dashboard en el que se muestra el ranking de los países con más consumo eléctrico y mayor porcentaje de crecimiento eléctrico anual, se han creado 4 tablas, donde dos de ellas incluyen la lista de todos los países junto con el valor de cada variable, ordenados de mayor a menor, y donde las otras dos representan el top 3 de países con mayores valores.

Cabe mencionar que, a diferencia de las otras dos aplicaciones utilizadas en este informe, Apache Superset no incluye una función que permita realizar un ranking directamente, por lo que para realizar esto, se han creado tablas ordenadas por valores directamente, y en el caso del top 3, se ha limitado el número de filas a 3 en cada caso. Finalmente, se ha incluido en la parte izquierda filtros para poder seleccionar países concretos y rangos de años específicos. De esta manera, el Dashboard de rankings quedaría de la siguiente forma:



Para la creación del Dashboard asociado al mapa, se ha utilizado la herramienta gráfica de Superset para representar este tipo de visualizaciones. Sin embargo, un detalle a mencionar es que se ha tenido que ajustar la configuración de esta gráfica para que los nombres de los países se asocien directamente. Para ello, en la opción country file type, se ha incluido el código ISO 3166-1 alpha-2 (CCA2). Finalmente, se crea la gráfica del mapa introduciendo las columnas de valores y de países para cada variable.

Una vez creada la gráfica, se incluye en el nuevo Dashboard, añadiendo los filtros correspondientes a la izquierda para poder modificar el mapa entre el consumo eléctrico y el crecimiento económico de cada país, así como un filtro de rango de años.



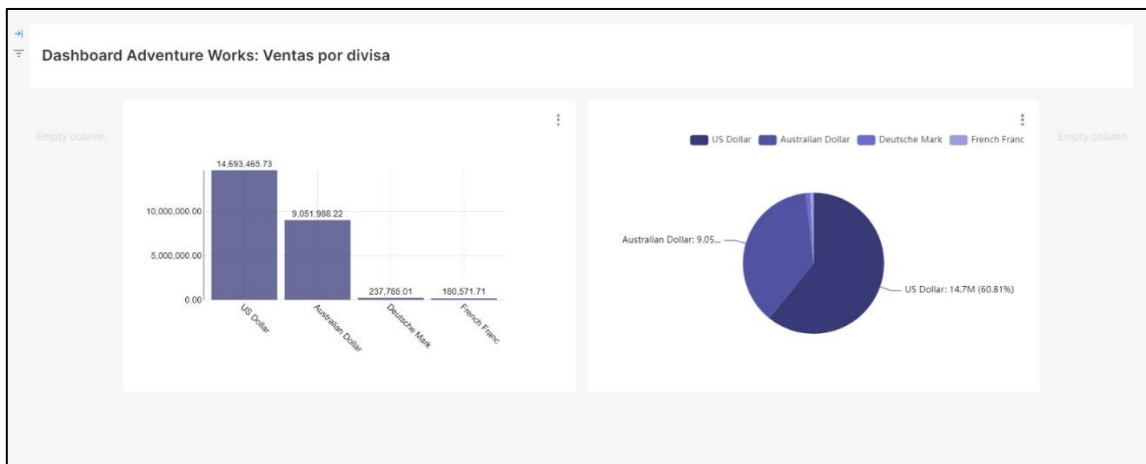
Por último, se finaliza con esta herramienta describiendo los diferentes ejercicios realizados a partir de la Base de Datos de Adventure Works.

En primer lugar, es necesario indicar que no es posible realizar gráficas a partir de valores ubicados en dos tablas diferentes. Por ello, es necesario llevar a cabo una serie de consultas SQL que permitan unir las tablas deseadas. De este modo, para el primero de los Dashboard asociados a esta parte de la tarea, se introduce la siguiente consulta en el SQL lab (la cual, posteriormente es exportada como dataset para ser empleada en la elaboración de gráficas).

```
Select *,cast(subquery.salesamount as numeric) as totalCasteado
  from (SELECT factinternetsales.*, dimcurrency.*
        FROM factinternetsales
        INNER JOIN dimcurrency on dimcurrency.currencykey =
factinternetsales.currencykey) AS subquery
where currencyname not like '%TRIAL%'
```

La función de dicha consulta SQL se basa en seleccionar todas las entradas asociadas al JOIN y realizar un casteo del total de ventas a tipo de dato numérico. Esto es necesario porque Superset no es capaz de gestionar correctamente las variables de tipo Money. Además, se aprovecha la necesidad de realizar una consulta para llevar a cabo un filtrado previo de los datos eliminando aquellas filas innecesarias que posean el patrón 'TRIAL'. Esto permite reducir el número de filas a cargar; lo cual será importante dado que el software utilizado solo es capaz de gestionar un número limitado de tuplas.

Además, es de interés comentar que este software no es capaz de mostrar las etiquetas asociadas a todos los datos de las gráficas (debido a su limitado tamaño). No obstante, es posible ver los valores de los mismos pasando por encima de las áreas deseadas con el ratón.



En lo relativo al segundo de los informes, la consulta a realizar es la siguiente:

```
Select *
  from (SELECT dimcustomer.*, dimgeography.*
        FROM dimcustomer
        INNER JOIN dimgeography on dimgeography.geographykey =
dimcustomer.geographykey) AS subquery
where subquery.city not like '%TRIAL%' and spanishcountryregionname
not like '%TRIAL%';
```

Su funcionamiento es muy similar al del caso anterior; no obstante, aquí no es necesario llevar a cabo ninguna conversión entre tipos de datos y es necesario añadir la condición de filtrado tanto para las ciudades como para los países.



## 8. Dificultades encontradas

En primer lugar, se ha presentado la dificultad de que los datos, tal y como venían en el fichero Excel importado desde la base de datos, presentaban un formato que dificultaba en gran medida su manejo, de forma que, al representarse los años en columnas, poder hacer cálculos que involucren los años sería imposible.

De esta forma, lo que se ha hecho es, mediante las herramientas de cada aplicación, realizar un pivoteo de las columnas de los años, de forma que, para cada país, se cree una fila por cada año



de información. Esto permite realizar los cálculos de promedios de los valores de manera sencilla, pudiendo utilizar un campo “año” de manera directa.

Otra dificultad que se ha presentado en la realización de la práctica ha sido el hecho de realizar la conexión de la base de datos a Apache Superset (en lo relativo a las tareas asociadas a la base de datos Adventure Works), tal y como se ha explicado en el apartado correspondiente.

Para la realización de las visualizaciones en Python y R mediante Power Bi, han surgido numerosas dificultades por la forma en la que este programa se relacionaba con las versiones de las otras aplicaciones. De esta forma, la visualización de las gráficas realizadas con Python y con R funcionan dependiendo de en qué aplicación se abra el fichero Power Bi.

## **9. Conclusiones**

Mediante esta práctica se ha profundizado en el aprendizaje de las herramientas de Tableau, Power Bi y Apache Superset para la realización de diagramas visuales para representar información de manera fácil y entendible. De esta forma se ha podido ver el funcionamiento de estas y las distintas posibilidades que presentan para realizar gráficas, tablas, mapas, y cualquier otro tipo de diagrama visual, así como el funcionamiento interactivo en la aplicación de filtros para poder manejar los resultados de la información en tiempo real.

Consecuentemente, se ha podido comprobar la facilidad que presentan estas aplicaciones para realizar análisis y representación de datos, muy útiles para casos donde se tenga grandes volúmenes de información y la extracción de conclusiones sea complicado. Además, las facilidades de limpieza y modificación de los datos, incluyendo la realización de cálculos y medidas.

Por otro lado, mediante la realización de esta práctica se ha podido ver el funcionamiento de las tres aplicaciones en su manejo con bases de datos, importando la información necesaria y realizando diagramas y gráficos de igual manera que con ficheros Excel o csv.

## **10. Horas de Trabajo**

26 horas.