

PRÁCTICA 5

ALMACÉNS E MINARÍA DE DATOS

Álex Baquero Domínguez

alex.baquero@rai.usc.es

Hugo Gómez Sabucedo

hugo.gomez.sabucedo@rai.usc.es

Alejandro Aybar Cifuentes

alejandro.aybar@rai.usc.es

Índice

1. Panel de control sobre el crecimiento de la economía mundial y el consumo eléctrico.	3
a. Ejercicio en Tableau	5
Tabla sobre el porcentaje de uso de energías renovables en los países escandinavos/nórdicos (Dinamarca, Noruega y Suecia) y países del mediterráneo (España, Portugal y Francia).	6
Gráfico de barras sobre consumo energético según el continente: % de energía renovable y % de fósiles.	6
Mapa mundial con el porcentaje de consumo de combustibles fósiles en la década de 2010 hasta hoy.	7
Gráfico de barras del uso de la energía total en la década de 2010.	8
Gráfico de barras sobre el gasto mundial de la década del 2010	9
Gráfico de línea sobre el crecimiento del PIB de 1960 a 2021 en los diferentes continentes.	10
Mapa sobre el PIB per capita en Europa de 1990 al 2000.	13
Gráfico de línea comparativo entre Rusia y Estados Unidos sobre la deuda del gobierno central a partir de los 90s.	14
Panel de control	15
b. Ejercicio en Power BI Desktop	18
c. Ejercicio en Apache Superset	34
Tabla sobre el porcentaje de uso de energías renovables en los países nórdicos (Dinamarca, Noruega y Suecia) y países del Mediterráneo (España, Portugal y Francia).	35
Gráfico de barras sobre consumo energético según el continente: % de energía renovable y % de fósiles	35
Mapa mundial con el porcentaje de consumo de combustibles fósiles en la década de 2010 hasta hoy	36
Gráfico de barras del uso de la energía total en la década de 2010	36
Gráfico de barras sobre el gasto mundial de la década del 2010	37
Gráfico de línea sobre el crecimiento del PIB de 1960 a 2021 en los diferentes continentes	37
Mapa sobre el PIB per cápita en Europa de 1990 a 2000	38
Gráfico de línea comparativo entre Rusia y Estados Unidos sobre la deuda del gobierno central a partir de los 90	39
Panel de control	39
2. Conexión a Adventure Works y creación de vistas que muestren el número de clientes, ventas por divisa y otros temas de interés.	41
a. Ejercicio en Tableau	41
Tabla sobre el número de clientes por territorio	42
Numero de clientes por país	43
Número de Venta por Divisa	44
Numero de ventas por Trimestre	44
b. Ejercicio en Power BI Desktop	46

c. Ejercicio en Apache Superset	52
Tabla sobre el número de clientes por territorio	52
Número de clientes por país	52
Número de ventas por divisa	52
Número de ventas por trimestre	52

1. Panel de control sobre el crecimiento de la economía mundial y el consumo eléctrico.

Para comenzar, vamos a elegir los conjuntos de datos con los que vamos a trabajar. Para la economía mundial, escogemos los siguientes conjuntos de datos:

- Crecimiento del PIB per cápita
- PIB per cápita
- Deuda del gobierno central
- Gasto

Para el consumo eléctrico, elegimos los siguientes:

- Consumo Energético
- Consumo de Energía Renovable
- Consumo de Combustibles Fósiles
- Uso de energía
- Uso de energía por \$1000 de PIB

Para obtener información sobre el crecimiento de la economía mundial y el consumo eléctrico, crearemos los siguientes gráficos en el panel de control:

Consumo Eléctrico

- Tabla sobre el porcentaje de uso de energías renovables en los países escandinavos/nórdicos (Dinamarca, Noruega y Suecia) y países del mediterráneo (España, Portugal y Francia)

- Mapa mundial con el porcentaje de consumo de combustibles fósiles en la década de 2010 hasta hoy
- Gráfico de barras sobre consumo energético según el continente: % de energía renovable y % de fósiles.
- Gráfico de barras del uso de la energía total en la década de 2010

Economía Mundial

- Gráfico de barras sobre el gasto mundial de la década del 2010 hasta hoy.
- Gráfico de línea sobre el crecimiento del PIB de 1960 a 2021 en los diferentes continentes
- Mapa sobre el PIB per cápita en Europa de 1990 al 2000.
- Gráfico de línea comparativo entre Rusia y Estados Unidos sobre la deuda del gobierno central a partir de los 90s.

a. Ejercicio en Tableau

Para comenzar para que el Tableau lea bien los datos, tenemos que sustituir los valores “.”, que representan que no hay datos, por un espacio en blanco. También sustituimos el separador que viene predeterminado “,” por un “;” y cambiamos el formato de las columnas a un tipo de datos como número decimal, redondeando a 2 decimales.

Por otro lado, para que Tableau reconozca todos los puntos como separadores decimales, así como los valores como números decimales de forma predeterminada. Nos conectamos al archivo TXT, entramos en "Propiedades de archivo de texto" y en el menú de diálogo, cambiamos la configuración regional del archivo a Inglés.

The screenshot shows the Tableau interface with a file named 'res_Gasto.csv' loaded. A dialog box titled 'res_Gasto.csv' is open, displaying the following settings:

- Separador de campos: Punto y coma
- Calificador de texto: Automático
- Conjunto de caracteres: UTF-8
- Configuración regional: gallego (España)

The background shows a data table with the following columns:

	res_Gasto.csv	res_Gasto.csv	res_Gasto.csv	res_Gasto.csv
	Gasto 80s	Gasto 90s	Gasto 2000s	Gasto 2010-hoy
	NULL	NULL	NULL	NULL
	NULL	NULL	NULL	NULL
	NULL	25.645,000000	NULL	NULL
	NULL	NULL	NULL	NULL
	NULL	14.004.999.999.999.900.00...	NULL	NULL
	NULL	NULL	NULL	NULL
	24.691,000000	23.917,000000	NULL	NULL
	00000 35.579,000000	NULL	NULL	NULL
	NULL	NULL	NULL	NULL
	NULL	15.463.000.000.000.000.00...	NULL	NULL
	27.822,000000	25.251,000000	NULL	NULL

res_Gasto.csv

res_Gasto.csv

Separador de campos: Punto y coma
 Calificador de texto: Automático
 Conjunto de caracteres: UTF-8
 Configuración regional: inglés (Estados Unidos)

Campos

Tipo	Nombre de campo	Tabl...	Nom...
🌐	Country Name	res_G...	Count...
#	Gasto 60s	res_G...	Gasto ...
#	Gasto 70s	res_G...	Gasto ...
#	Gasto 80s	res_G...	Gasto ...
#	Gasto 90s	res_G...	Gasto ...
#	Gasto 2000s	res_G...	Gasto ...

#	#	#	🌐
res_Gasto.csv	res_Gas...	res_Gasto.csv	res_Gasto.csv
Gasto 80s	G...	Gasto 2000s	Gasto 2010-hoy
NULL	NULL	10.026.666.666...	12182,00
NULL	NULL	39,85000	41276363636363...
26,9210	50,88	58.068,00000	40,20
NULL	NULL	34.577,00000	34258181818181...
37,2960	39,20	36.685,00000	39875454545454...
10,7180	8,30	13.105,00000	15953636363636...
NULL	NULL	21,46000	21118888888888...
NULL	NULL	NULL	26814999999999...
NULL	285,18	28.919.999.999...	30348333333333...
NULL	13,12	20.732.00000	24757272727272...

Tabla sobre el porcentaje de uso de energías renovables en los países escandinavos/nórdicos (Dinamarca, Noruega y Suecia) y países del mediterraneo (España, Portugal y Francia).

A través del excel del uso de energía renovables, calculamos la media de las décadas y lo pasamos a porcentajes. Introducimos este valor en la columna y en la fila "country name". Finalmente, aplicamos un filtro seleccionado estos 6 países y poniendo un filtro de color dependiendo si el porcentaje del consumo renovable es mas alto o no.

Uso da enerxía renovable nos países nórdicos e mediterráneos.


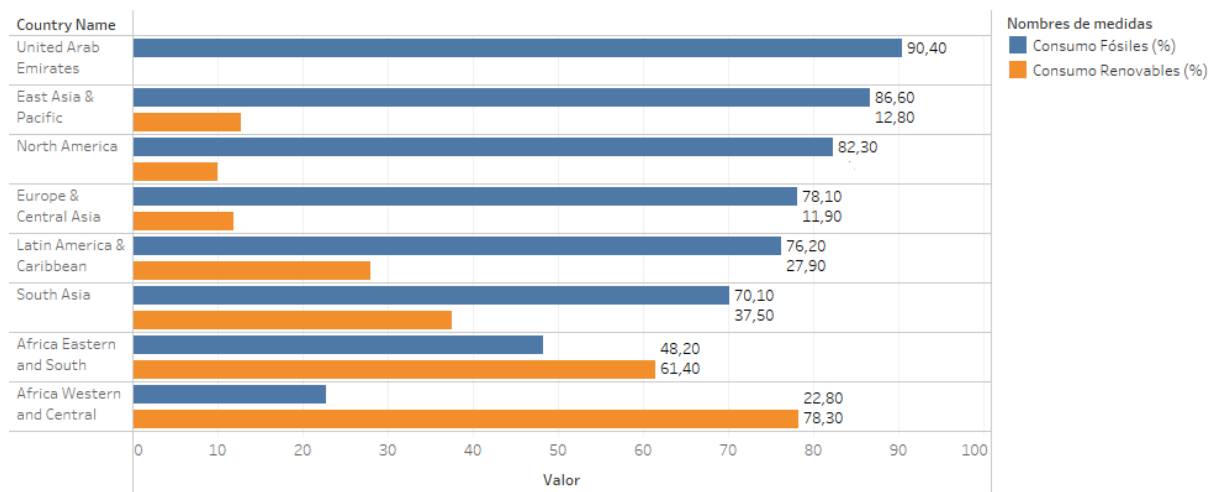
Country		Consumo Renovables(%)
Norway	58,71	 15,60 58,71
Sweden	49,33	
Denmark	30,16	
Portugal	27,15	
Spain	16,34	
Italy	15,60	

Gráfico de barras sobre consumo energético según el continente: % de energía renovable y % de fósiles.

Calculamos la media del consumo fosil y renovable y lo ponemos en la fila. En la Columna seleccionamos la columna “country name” del .csv y aplicamos un filtro para seleccionar solo los continentes. Por ultimo, aplicamos un filtro de color para diferenciar las dos barras y una etiqueta para ver el resultado exacto de cada consumo.

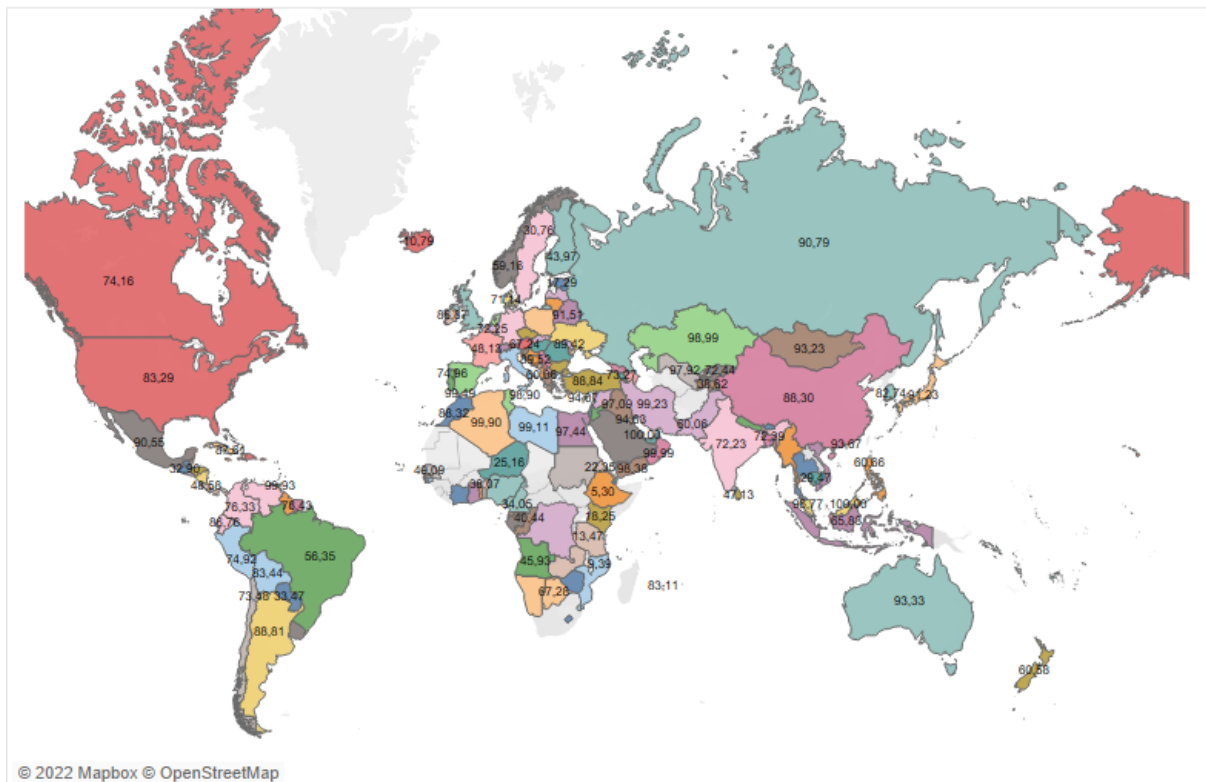
Energía Renovable vs Combustibles Fósiles



Mapa mundial con el porcentaje de consumo de combustibles fósiles en la década de 2010 hasta hoy.

Agrupamos los años de 2010 hasta hoy en un nuevo excel para tenerlos todo en una misma columna. Añadimos un función geográfica de país/región a la columna country name. De esta manera, Tableau genera un campo Latitud y un campo Longitud (*Latitud (generado)* y *Longitud (generado)*). Ponemos como columna la longitud y la fila como latitud, creándose un mapa. En marcas, aplicamos la columna country name por color y la suma de la década de los 2010.

Consumo mundial de combustibles fósiles do ano 2010 ata hoxe.



Mapa basado en Longitud (generado) y Latitud (generado). El color muestra detalles acerca de Country. Las marcas se etiquetan por suma de Decada 2010. La vista se filtra en Country, lo que conserva 179 de 221 miembros.

Gráfico de barras del uso de la energía total en la década de 2010.

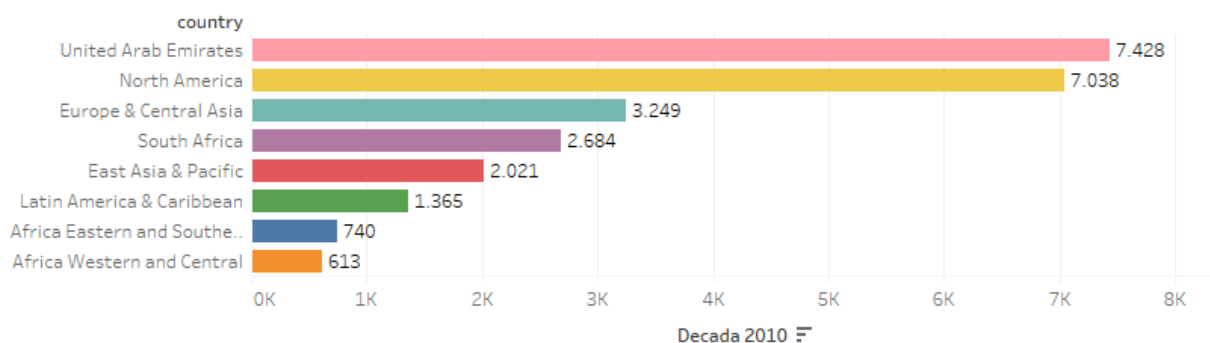
Agrupamos los datos de 2010 hasta hoy en un excel y los colocamos como una columna. Seleccionamos la columna country name como una fila e aplicamos un filtro escogiendo únicamente solo los continentes.

Columnas	SUMA(Decada 2010)
Filas	country

Finalmente, aplicamos una marca de color por pais y añadimos una etiqueta para que aparezca el número exacto de energía.



Uso da enerxía total na década dos 2010



Suma de Decada 2010 para cada country. El color muestra detalles acerca de country. Las marcas se etiquetan por suma de Decada 2010. La vista se filtra en country, lo que conserva 8 miembros.

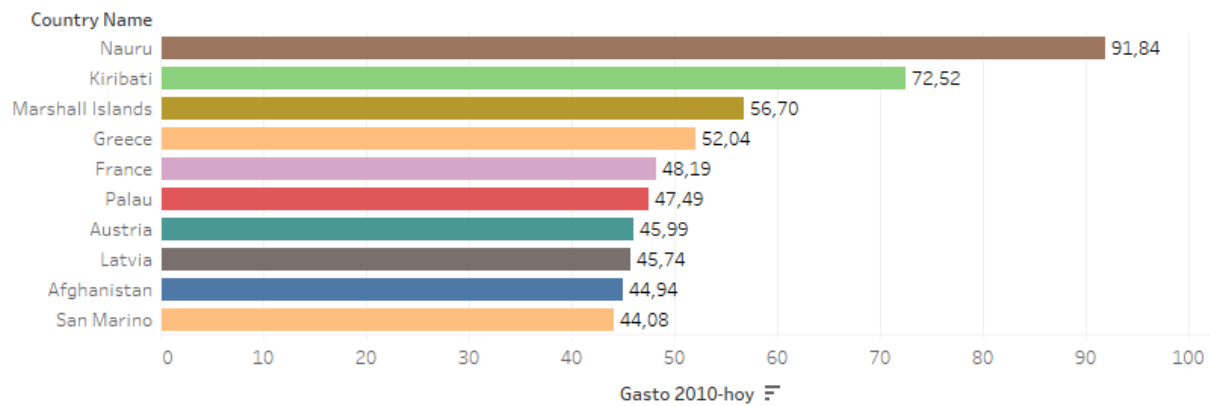
Gráfico de barras sobre el gasto mundial de la década del 2010

Para crear un gráfico de barras sobre el gasto mundial de la década del 2010 hasta hoy, editamos el archivo de gastos.csv para unir las columnas de cada año en décadas. Consiguiendo un archivo .csv que contenga los gastos de 60s,70s,80s etc...Para este caso utilizaremos la columna de la década de 2010.

Creamos colocando una dimensión país en el estante **Columnas** y una medida sum(gasto 2010-hoy) en el estante **Filas**. Filtramos esta medida para que no muestro países que no tengan datos, es decir, que gastos 2010-hoy en ese país sea null y le añadimos un color a cada país.

Mostramos el top 10 de los países con más gastos:

Gastos a nivel mundial na década dos 2010 ata hoxe.



Suma de Gasto 2010-hoy para cada Country Name. El color muestra detalles acerca de Country Name. Las marcas se etiquetan por suma de Gasto 2010-hoy. La vista se filtra en Country Name suma de Gasto 2010-hoy. El filtro Country Name conserva 10 de 184 miembros. El filtro suma de Gasto 2010-hoy conserva solo los valores no NULL.

Gráfico de línea sobre el crecimiento del PIB de 1960 a 2021 en los diferentes continentes.

Para realizar un gráfico de líneas necesitamos una columna tipo fecha la cual no tenemos en los .csv descargados. Así que manipulamos el .csv del crecimiento del pib en python a través de read e writes para obtenerlo.

```

1 Country Name,Decade,Crecimiento PIB per cápita
2 Africa Eastern and Southern, "60s", "1,94 %"
3 Africa Eastern and Southern, "70s", "0,30 %"
4 Africa Eastern and Southern, "80s", "-0,33 %"
5 Africa Eastern and Southern, "90s", "-0,88 %"
6 Africa Eastern and Southern, "2000s", "1,68 %"
7 Africa Eastern and Southern, "2010-hoy", "-0,07 %"
8
9
10 Africa Western and Central, "0,86 %", "3,38 %", "-2,57 %", "-0,23 %", "3,33 %", "0,86 %"
11
12 East Asia & Pacific, "5,77 %", "3,06 %", "3,66 %", "3,05 %", "4,26 %", "4,08 %"
13
14 Europe & Central Asia, "2,64 %", "1,76 %", "1,25 %", "1,70 %", "1,24 %"
15
16 Latin America & Caribbean, "2,64 %", "3,61 %", "0,09 %", "1,09 %", "1,57 %", "0,84 %"
17
18 North America, "3,28 %", "2,08 %", "2,15 %", "1,96 %", "1,04 %", "1,34 %"
19
20 South Asia, "1,92 %", "0,66 %", "3,17 %", "3,22 %", "4,18 %", "4,17 %"
21
22 United Arab Emirates, "0,14 %", "-4,58 %", "-0,11 %", "-4,83 %", "0,81 %"
23

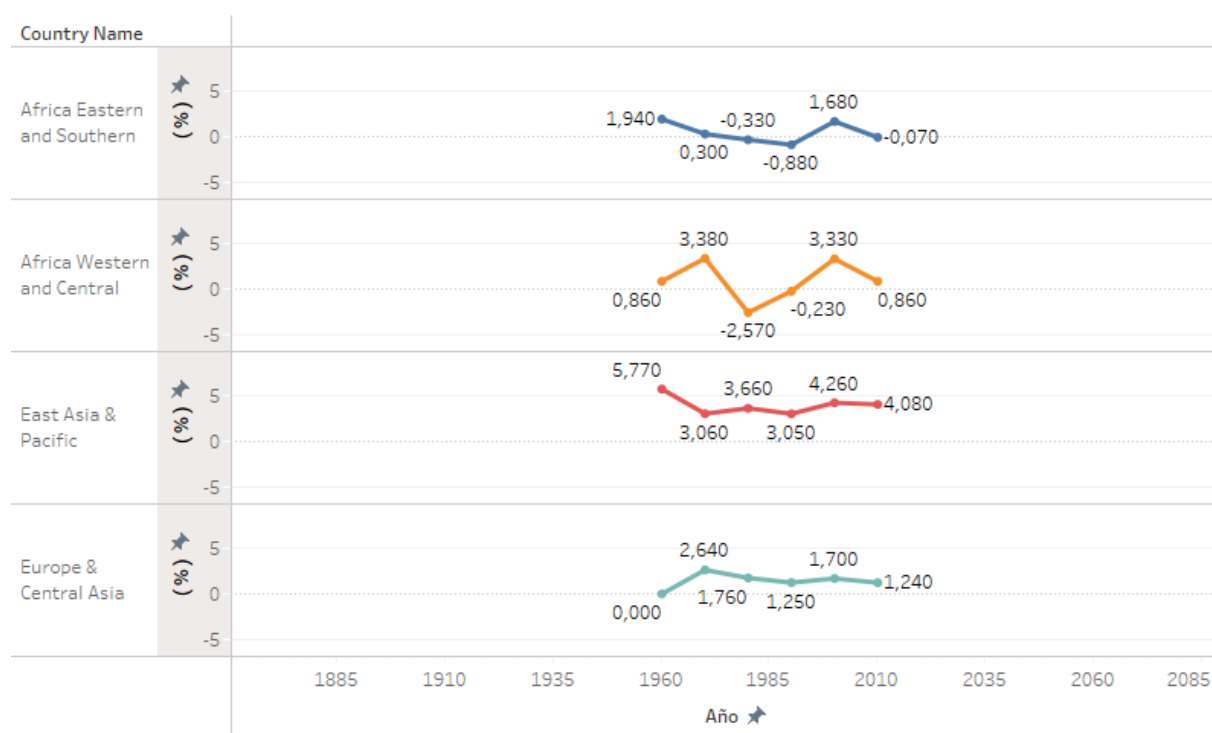
```

El resultado es algo así:

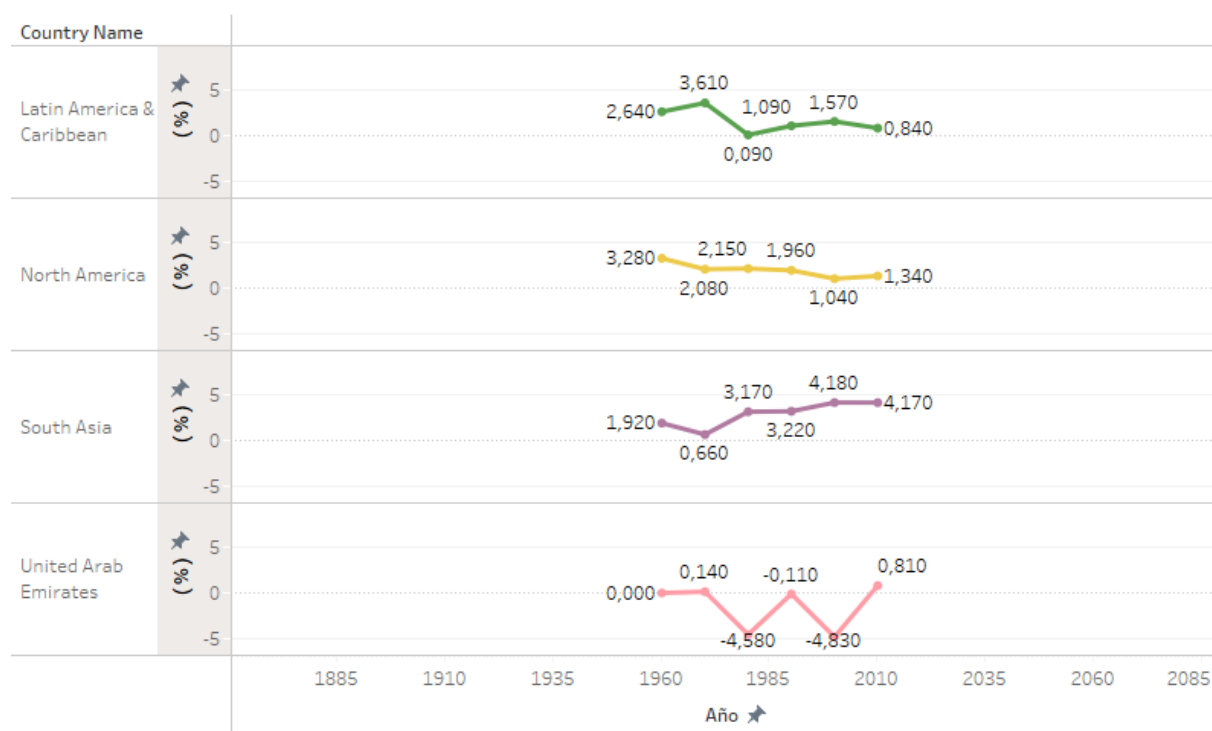
Country Name	Decade	Crecimiento PIB per cáp...
Africa Western and Central	01/01/1980	-2,57
Africa Western and Central	01/01/1990	-0,23
Africa Western and Central	01/01/2000	3,33
Africa Western and Central	01/01/2010	0,86
East Asia & Pacific	01/01/1960	5,77
East Asia & Pacific	01/01/1970	3,06
East Asia & Pacific	01/01/1980	3,66
East Asia & Pacific	01/01/1990	3,05
East Asia & Pacific	01/01/2000	4,26
East Asia & Pacific	01/01/2010	4,08
Europe & Central Asia	01/01/1960	0,00
Europe & Central Asia	01/01/1970	2,64

Añadimos en la columna la fecha de las décadas y en las filas: country name y la suma del crecimiento del pib per cápita (el cual está en porcentaje). Finalmente, aplicamos diferentes colores a cada país y una etiqueta para durante la línea ponga el porcentaje exacto

Crecemento do PIB dende 1960 ata hoxe.



Crecemento do PIB dende 1960 ata hoxe.



Mapa sobre el PIB per capita en Europa de 1990 al 2000.

Para crear un mapa, primero tenemos que aplicar una función geográfica de País/Region a la columna Country Name.

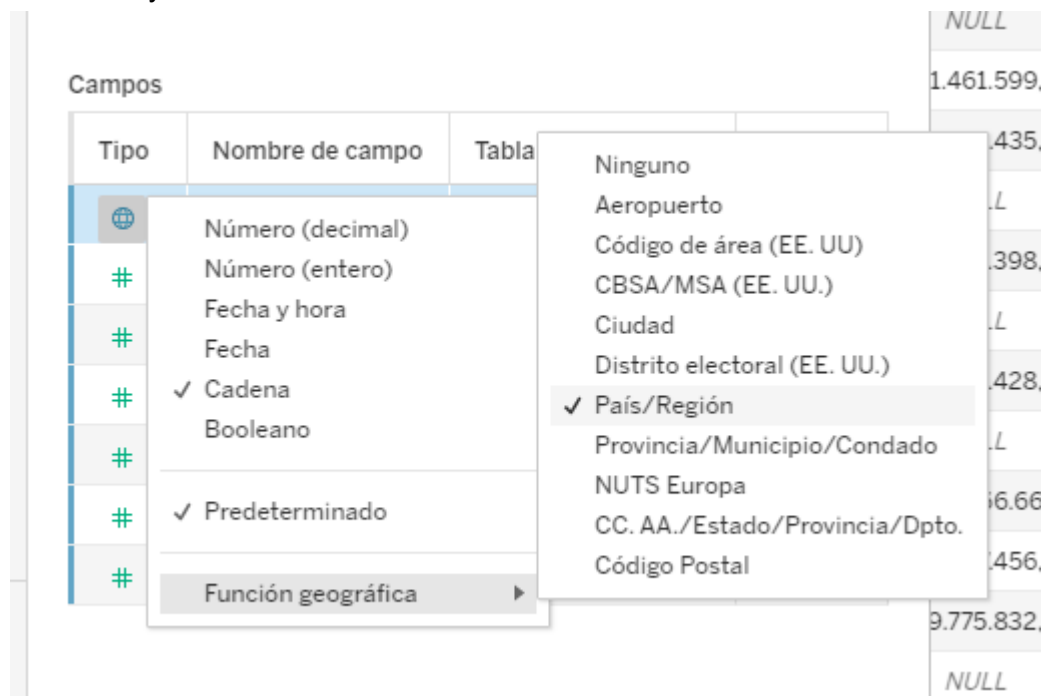
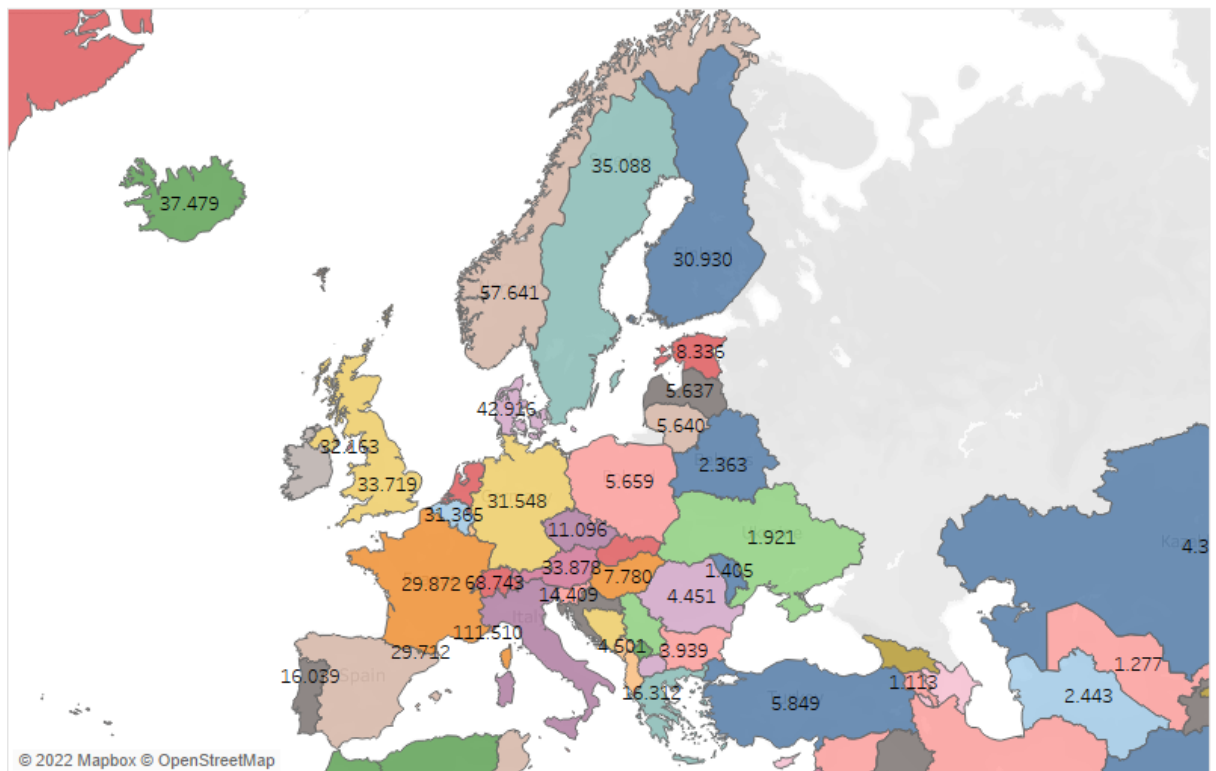


Tableau genera un campo Latitud y un campo Longitud (*Latitud (generado)* y *Longitud (generado)*). Esto se debe a que asignó funciones geográficas a los datos. Ponemos como columna la longitud y la fila como latitud, creándose un mapa. En marcas, aplicamos la columna country name por color y la suma de los años 90s (juntando antes en un excel todos los años que pertenecen a esa decada)

PIR per capita en Europa na década dos 90s



Mapa basado en Longitud (generado) y Latitud (generado). El color muestra detalles acerca de country name. Las marcas se etiquetan por suma de década 90. La vista se filtra en country name, lo que conserva 222 de 260 miembros.

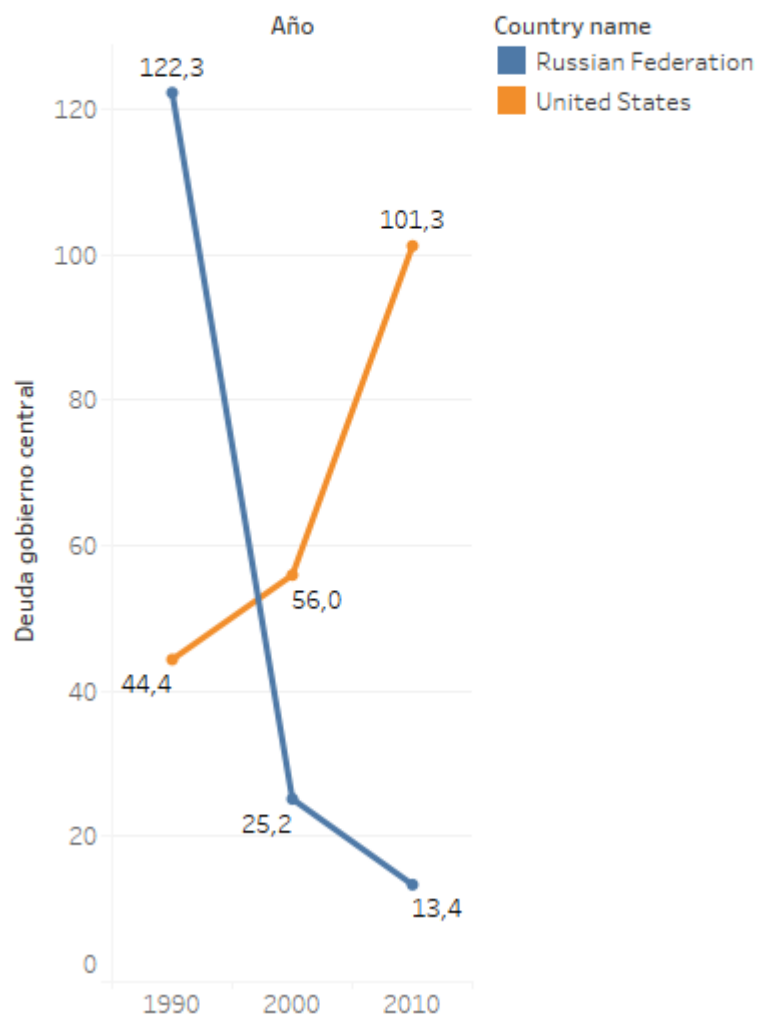
Gráfico de línea comparativo entre Rusia y Estados Unidos sobre la deuda del gobierno central a partir de los 90s.

Para empezar modificamos el .csv añadiendo una columna nueva que sea “año”, que será tipo Data. De esta manera, conseguimos el siguiente .csv:

```
deudaGobiernoCentralLinea.csv: cuaderno de notas
Ficheiro  Editar  Formato  Ver  Axuda
Country name;Decada;Deuda gobierno central
Russian Federation;1990;122.345
Russian Federation;2000;25.17375
Russian Federation;2010;13.38090909090909
United States;1990;44.357
United States;2000;55.980999999999995
United States;2010;101.26545454545455
```

Ponemos el año en la columna y en la fila la suma de la deuda del gobierno. Como siempre, añadimos un color para cada país y ponemos las etiquetas para los años 1990, 2000 y 2010,

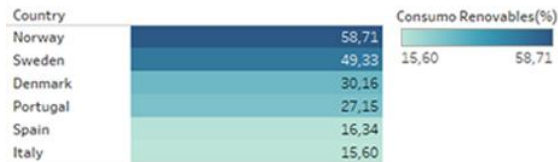
Gráfico comparativo de la deuda del gobierno central a partir de los 90s



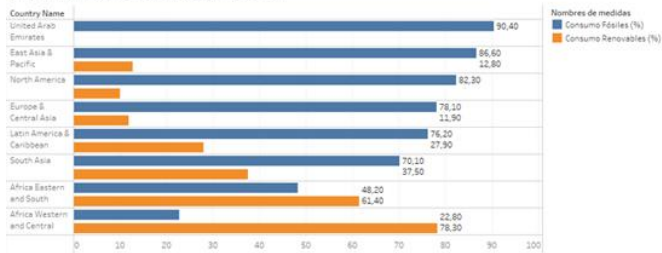
Panel de control

El panel de control de la consumo eléctrico quedaría de la siguiente forma:

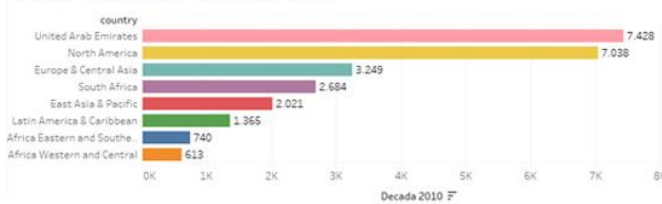
Uso da enerxía renovable nos países nórdicos e mediterráneos.



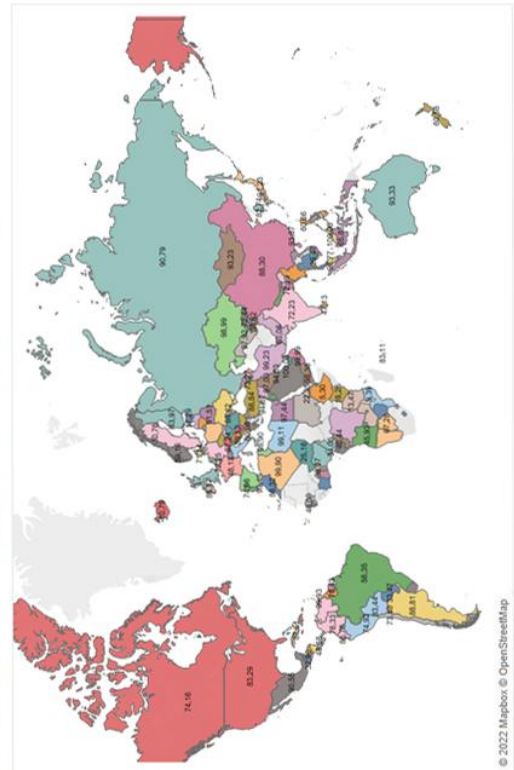
Energía Renovable vs Combustibles Fósiles



Uso da enerxía total na década dos 2010



Consumo mundial de combustibles fósiles do ano 2010 ata hoxe.

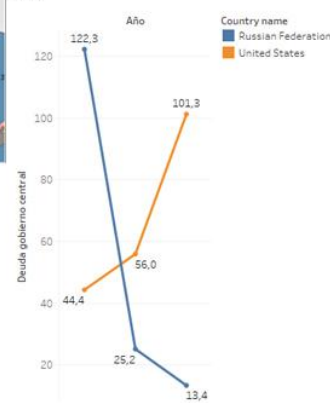


El panel de control del economía mundial sería el siguiente:

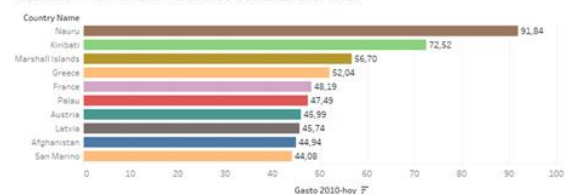
PIR per capita en Europa na década dos 90s



Gráfico comparativo de la deuda del gobierno central a partir de los 90s



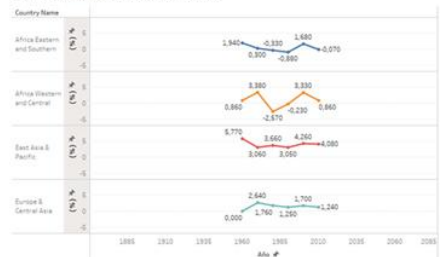
Gastos a nivel mundial na década dos 2010 ata hoxe.



Crecedo do PIB dende 1960 ata hoxe.



Crecedo do PIB dende 1960 ata hoxe.



b. Ejercicio en Power BI Desktop

1. Limpieza de datos

Comenzamos importando los datos a Power Bi. Tenemos 10 tablas que tenemos que importar por separado. Para cada una de ellas, nos apoyamos en la herramienta Power Query.

Hay que hacer una tarea de limpieza de datos considerable. Para todas las tablas, eliminamos las columnas de código de país, nombre de la serie y código de la serie. Marcamos la opción “Usar la primera fila como encabezado”.

Además, sustituimos los valores “..”, que representan que no hay datos, por dejar el campo vacío. A continuación, para los campos donde sí hay valores, sustituimos el separador por “,” en vez de “.”, para poder configurar el tipo de datos como número decimal. Por último, redondeamos a 2 decimales.

Veamos un ejemplo del antes y el después con la tabla “Uso de Energía”.

Antes:

	APC Country Name	APC Country Code	APC Series Name	APC Series Code	APC 1960 [YR1960]	APC 1961 [YR1961]
1	Afghanistan	AFG	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
2	Albania	ALB	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
3	Algeria	DZA	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
4	American Samoa	ASM	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
5	Andorra	AND	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
6	Angola	AGO	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
7	Antigua and Barbuda	ATG	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
8	Argentina	ARG	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
9	Armenia	ARM	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
10	Aruba	ABW	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
11	Australia	AUS	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE	3063.55427059293	3115.78708385004
12	Austria	AUT	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE	1546.26146801032	1554.0349059502
13	Azerbaijan	AZE	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
14	Bahamas, The	BHS	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
15	Bahrain	BHR	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
16	Bangladesh	BGD	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
17	Barbados	BRB	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
18	Belarus	BLR	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE
19	Belgium	BEL	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE	2519.49731954668	2570.81562308497
20	Belize	BLZ	Energy use (kg of oil equivalent per capita)	EG.USE.PCAP.KG.OE

Después:

	Country Name	1.2 1960 [YR1960]	1.2 1961 [YR1961]	1.2 1962 [YR1962]	1.2 1963 [YR1963]	1.2 1964 [YR1964]
1	Afghanistan	null	null	null	null	null
2	Albania	null	null	null	null	null
3	Algeria	null	null	null	null	null
4	American Samoa	null	null	null	null	null
5	Andorra	null	null	null	null	null
6	Angola	null	null	null	null	null
7	Antigua and Barbuda	null	null	null	null	null
8	Argentina	null	null	null	null	null
9	Armenia	null	null	null	null	null
10	Aruba	null	null	null	null	null
11	Australia	3063,55	3115,79	3172,97	3284,05	3349,41
12	Austria	1546,26	1554,03	1675,87	1824	1855,09
13	Azerbaijan	null	null	null	null	null
14	Bahamas, The	null	null	null	null	null
15	Bahrain	null	null	null	null	null
16	Bangladesh	null	null	null	null	null
17	Barbados	null	null	null	null	null
18	Belarus	null	null	null	null	null
19	Belgium	2519,5	2570,82	2810,06	3043,31	3021,65
20	Belize	null	null	null	null	null

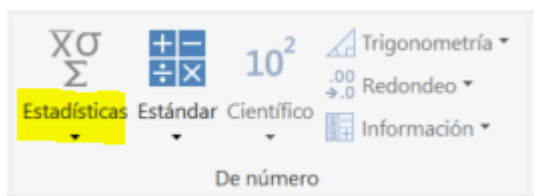
Si bien para los años 60 los datos son escasos, según avanzan los años los datos son abundantes, por lo que el conjunto nos será útil para el análisis.

Hacemos esta misma transformación en todas las tablas, ya que todas vienen con el mismo formato.

No obstante, llegamos a la conclusión de que tener decenas de columnas, cada una representando un año (1960, 1961, ..., 2020, 2021) hacía al conjunto de datos algo caótico y dificultaba el poder hacer análisis.

Por lo tanto, decidimos agrupar los años por décadas, de manera que las columnas tendrían serían, a partir de ahora, las siguientes: “1960s, 1970s, 1980s, 1990s, 2000s, 2010-hoy”.

Realizamos esta transformación en todas las tablas. El proceso consiste en seleccionar las columnas por décadas (ej. década de los 60 -> 1960-1969), y en la pestaña Agregar Columna de Power Query, seleccionar “Estadísticas -> Promedio”

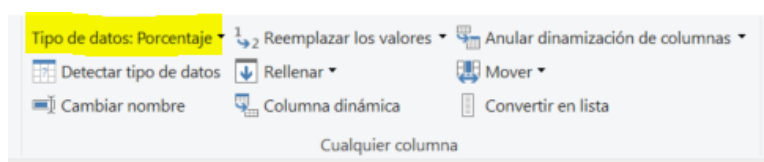
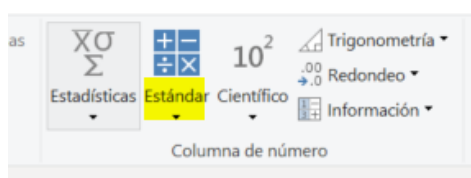


Al hacer esto con todas las décadas, las tablas nos queda así:

Country Name	1.2 PIB per cápita 60s	1.2 PIB per cápita 70s	1.2 PIB per cápita 80s	1.2 PIB per cápita 90s	1.2 PIB per cápita 2000s	1.2 PIB per cápita 2010-hoy
1 Afghanistan	null	null	null	null	367,364	547,653
2 Albania	null	null	1768,518	1437,02	2655,743	4094,849
3 Algeria	1944,067	2711,398	3281,382	2942,226	3570,25	4054,561
4 American Samoa	null	null	null	null	12481,715	11745,227
5 Andorra	null	35986,428	30754,574	29712,286	37681,63	35980,148
6 Angola	null	null	2442,191	1823,073	2375,289	2871,451
7 Antigua and Barbuda	null	5900,937	8875,066	12507,988	15179,607	14443,046
8 Argentina	7885,832	9775,832	9435,767	10061,676	11163	13298,717
9 Armenia	null	null	null	1112,59	2262,342	3668,621
10 Aruba	null	null	20288,225	26638,051	28685,986	27719,121
11 Australia	22080,81	28389,946	32901,932	39267,749	49637,168	56727,917
12 Austria	14342,628	21650,948	27362,183	33878,133	41294,121	44686,97
13 Azerbaijan	null	null	null	1652,501	2902,243	5294,086
14 Bahamas, The	25375,42	24720,216	31160,385	31605,55	35278,161	30889,824
15 Bahrain	null	null	18336,042	21584,903	22237,335	21455,025
16 Bangladesh	490,065	441,261	479,605	567,496	771,956	1312,822
17 Barbados	null	12222,928	14141,984	14543,546	16914,638	16263,595
18 Belarus	null	null	null	2363,34	3818,908	6063,234
19 Belgium	14183,105	20954,147	25751,635	31365,076	37943,104	41075,618
20 Belize	1311,615	1848,848	2443,195	3718,058	4674,045	4608,612

Además, nos fijamos en que hay varias tablas cuyos valores representan porcentajes, pero vienen expresadas como números decimales (ej. 76.34). Para pasarlas a porcentaje, dividimos el valor entre 100, y cambiamos el tipo de dato a “Porcentaje”. De esta manera, la transformación es: 76.34 -> 0,7634 -> 76,34%.

Para hacer esto, seleccionamos las columnas que nos interesan y nos apoyamos de las siguientes herramientas:



Aplicamos estos pasos a las tablas: Consumo Combustibles Fósiles, Consumo Energía Renovable, Gasto y Crecimiento PIB per cápita. Nos quedan así:

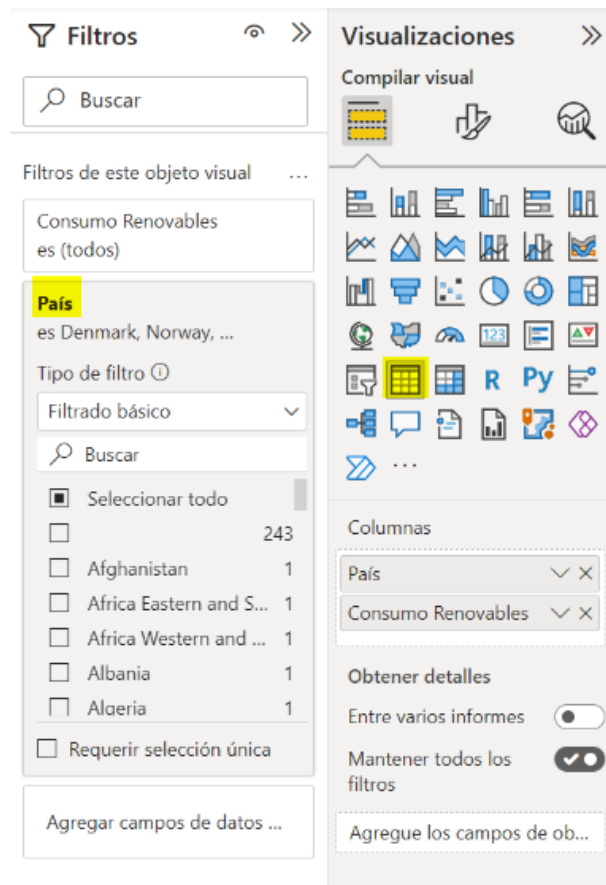
% Consumo Fósiles 60s	% Consumo Fósiles 70s	% Consumo Fósiles 80s	% Consumo Fósiles 90s	% Consumo Fósiles 2000s	% Consumo Fósiles 2010-hoy
null	null	null	null	null	null
null	75,83 %	79,40 %	55,08 %	63,60 %	60,06 %
null	99,23 %	99,76 %	99,81 %	99,76 %	99,90 %
null	null	null	null	null	null
null	null	null	null	null	null
null	18,59 %	19,51 %	24,14 %	32,69 %	45,93 %
null	null	null	0,00 %	0,00 %	null
null	92,17 %	89,37 %	87,78 %	88,38 %	88,81 %
null	null	null	79,52 %	68,25 %	73,27 %

Ya con los datos limpiados y organizados, vamos a proceder con los análisis del crecimiento económico y el consumo energético. Comenzamos con el consumo energético:

2. Análisis del Consumo Energético

Comenzamos con la tabla del % de energías renovables utilizado en países nórdicos en comparación con los países mediterráneos.

En Visualizaciones, seleccionamos un gráfico de tabla, y en las columnas añadimos los campos "País" y "Consumo Renovables". Además, en Filtros, añadimos un filtro básico sobre el campo "País", donde seleccionaremos los países de Noruega, Suecia, Dinamarca, Portugal, España e Italia. 3 nórdicos y 3 mediterráneos.

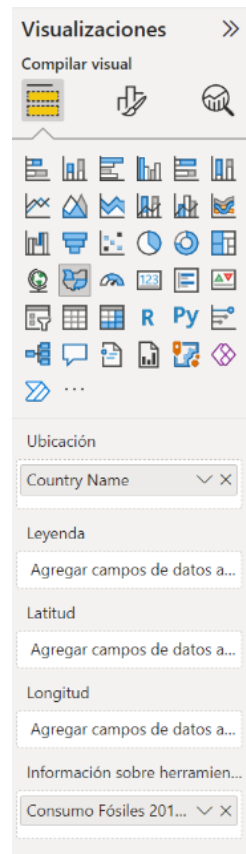


Le damos formato a la tabla, activando los bordes con un 15% de curvatura, cambiando la paleta de colores a un degradado de morados, y mostrando los porcentajes en color blanco.

País	Consumo Renovables
Norway	58,71 %
Sweden	49,33 %
Denmark	30,16 %
Portugal	27,75 %
Spain	16,34 %
Italy	15,60 %

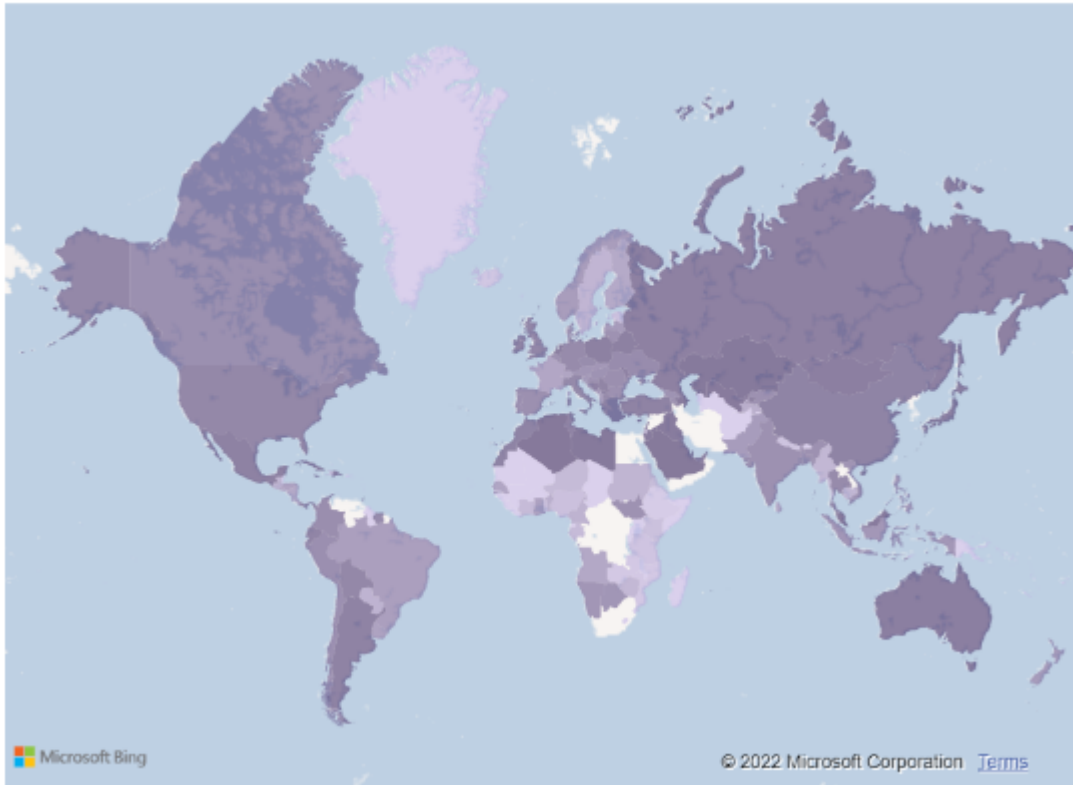
Continuamos con el mapa de % de consumo de combustibles fósiles. De nuevo en Visualizaciones, seleccionamos un mapa coroplético, al que añadimos, en Ubicación, el campo Country Name. Además, en Información sobre herramientas, añadimos el campo

Consumo Fósiles 2010-hoy, para que al situar el ratón sobre un país del mapa, se muestre el % de consumo de combustibles fósiles de este.



Le damos al mapa un estilo acorde al informe, con una escala de morados, donde un morado claro representa un bajo %, mientras que un morado oscuro representa un % alto.

% Consumo de Combustibles Fósiles

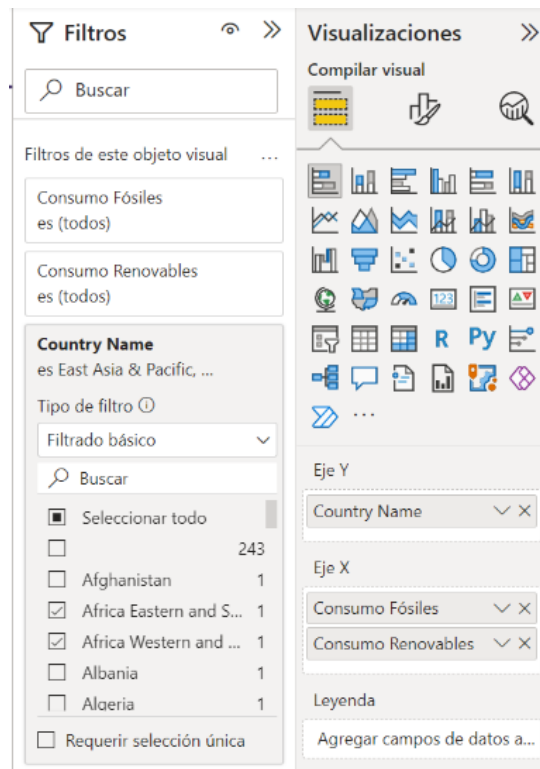


El siguiente gráfico a insertar es un gráfico de barras que muestre el % de consumo de combustibles fósiles y el % de consumo de energías renovables.

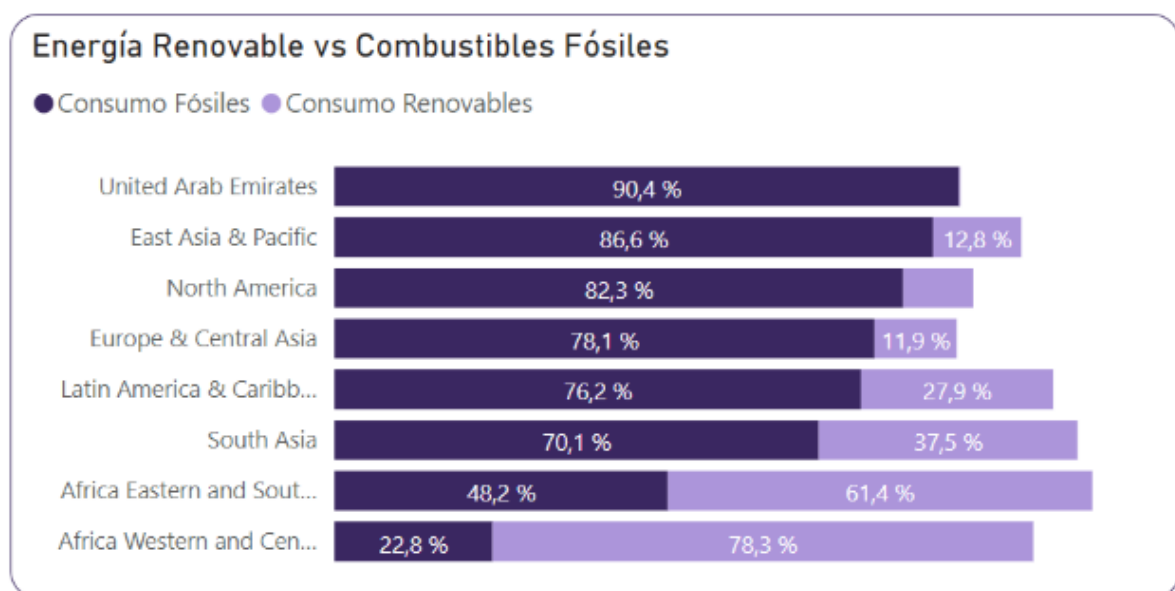
Seleccionamos un gráfico de barras apiladas, y añadimos:

- En el eje Y: el campo País
- En el eje X: los campos Consumo Fósiles y Consumo Renovables

Además, establecemos un filtro sobre el campo País, de manera que nos quedemos solo con los continentes:



Configuramos el estilo acorde al informe, quedando el gráfico de la siguiente manera:



Por último, introducimos el gráfico de barras del Uso Total de Energía. Añadimos en el eje Y el campo país, y en el eje X el campo "Uso Energía Total". Además, en Información sobre herramientas, ponemos los campos Consumo Renovables y Consumo Fósiles, para que se muestren al posar el ratón sobre una barra concreta.

Para este gráfico, mantenemos el filtro de los continentes y aplicamos el estilo correspondiente:

Filtros
»

Filtros de este objeto visual ...

- Consumo Fósiles 2010... es (todos)
- Consumo Renovables 2... es (todos)
- Country Name**
es East Asia & Pacific, ...
- Uso Energía Total 201... es (todos)
- Agregar campos de datos ...

Filtros de esta página ...

- Agregar campos de datos ...

Filtros de todas las páginas ...

- Agregar campos de datos ...

Visualizaciones »

Compilar visual

Eje Y

Country Name
▼ ×

Eje X

Uso Energía Total 201...
▼ ×

Legenda

Agregar campos de datos a...
×

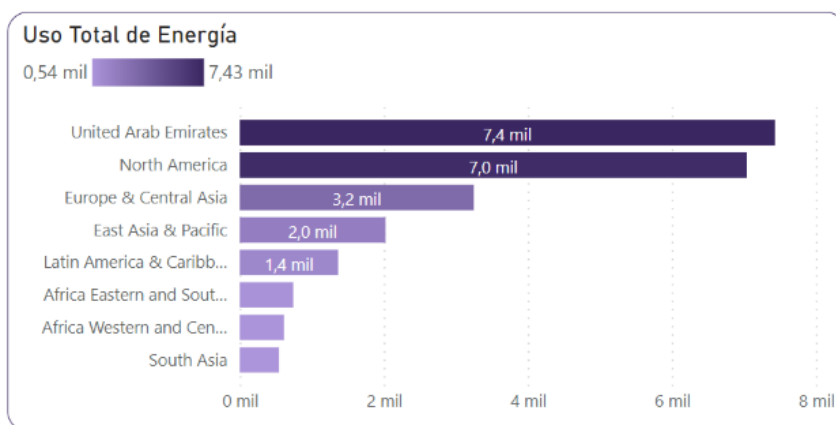
Múltiplos pequeños

Agregar campos de datos a...
×

Información sobre herramientas

Consumo Renovables ...
▼ ×

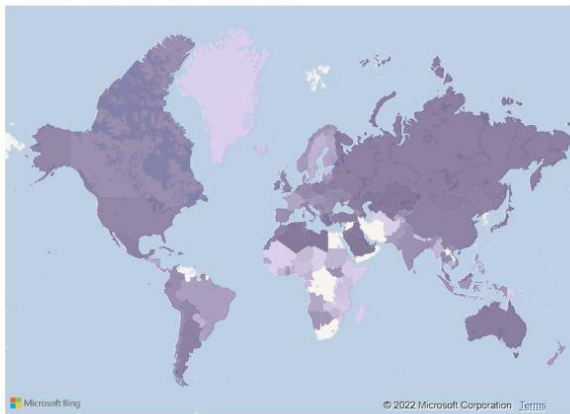
Consumo Fósiles 201...
▼ ×



Por lo tanto, nuestro informe sobre el Consumo Energético queda de la siguiente manera:

Consumo Energético 2010-2021

% Consumo de Combustibles Fósiles

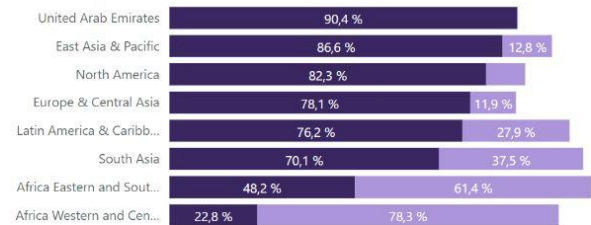


Energías Renovables - Nórdicos vs Mediterráneos

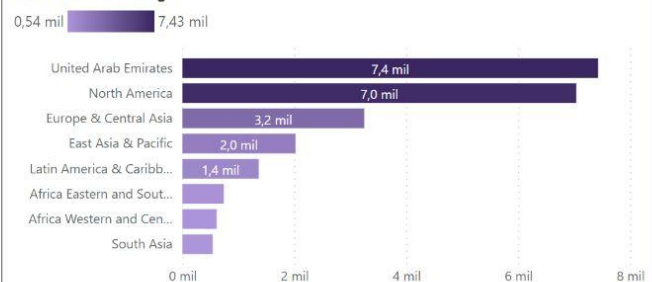
País	Consumo Renovables
Norway	58,71 %
Sweden	49,33 %
Denmark	30,16 %
Portugal	27,75 %
Spain	16,34 %
Italy	15,60 %

Energía Renovable vs Combustibles Fósiles

● Consumo Fósiles ● Consumo Renovables



Uso Total de Energía

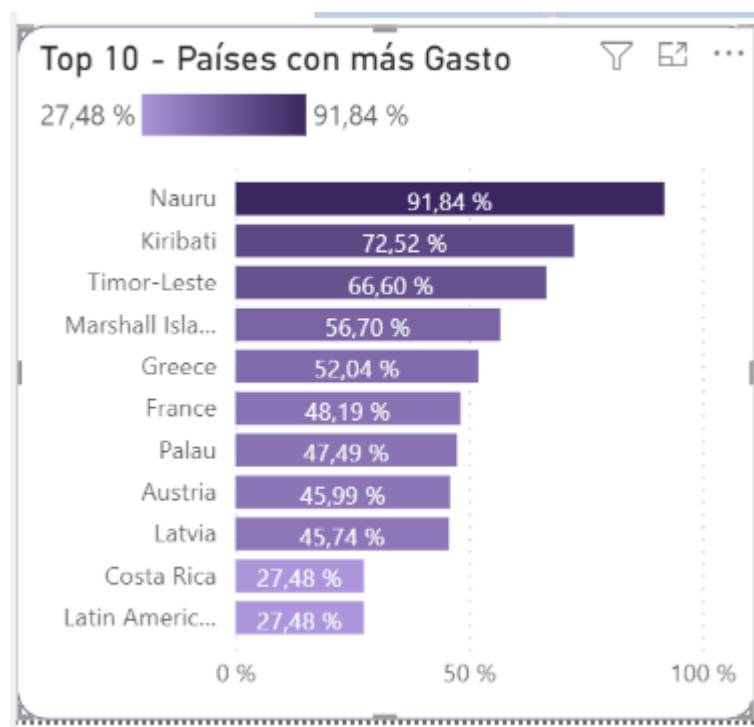
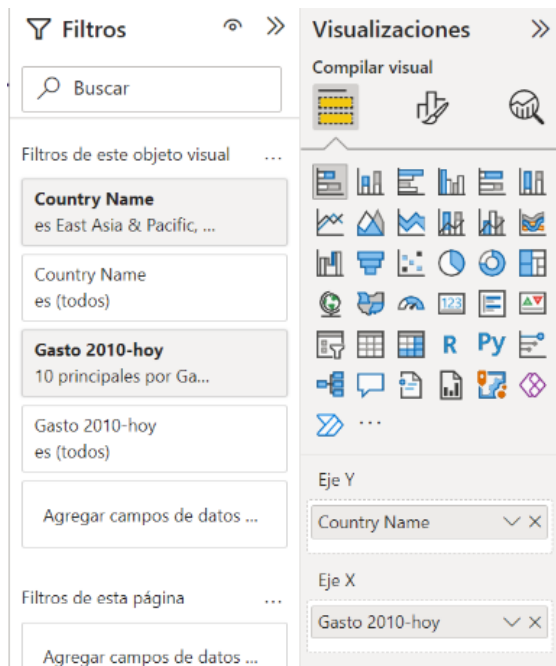


2. Análisis del Crecimiento de la Economía

El primer gráfico que insertaremos será el gráfico de barras en el que mostraremos el Top 10 de países con mayor % de gasto en la última década. Seleccionamos:

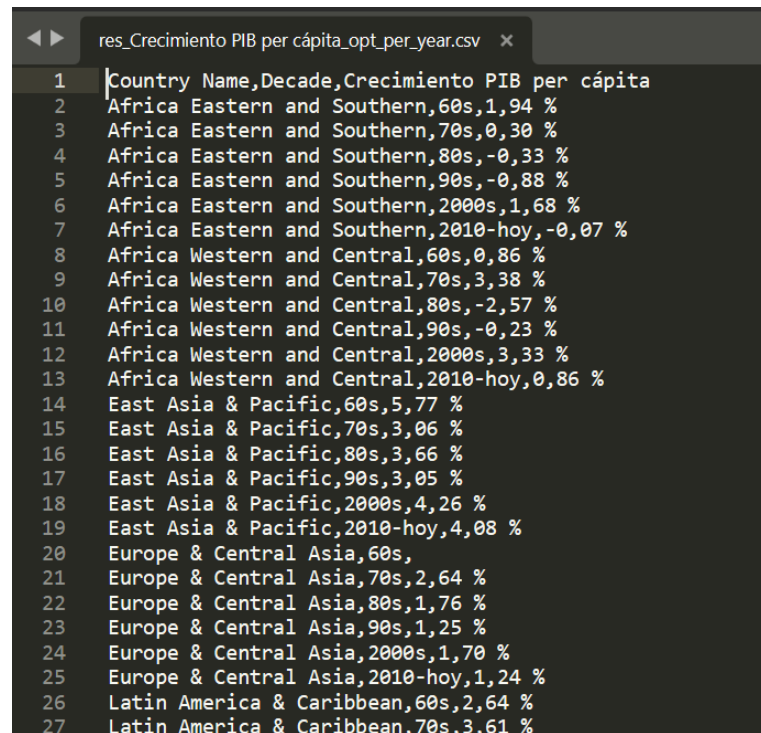
- Eje Y: nombre del país
- Eje X: Gasto 2010-hoy

Además, establecemos como filtro un Top 10 superior según el Gasto.



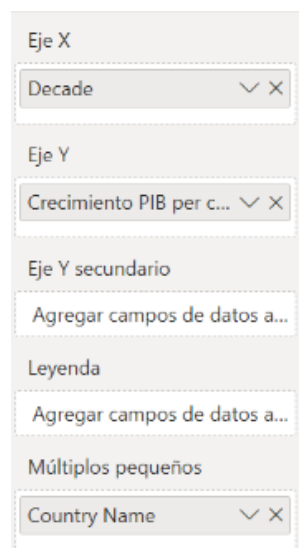
El siguiente gráfico es un gráfico de línea con el crecimiento del PIB en los distintos continentes a lo largo de las décadas. Para realizar este gráfico, hemos tenido que realizar una modificación en el CSV, de manera que las décadas aparezcan representadas en las filas, en vez de en las columnas, para así poder agrupar por décadas.

Para ahorrarnos el filtro, lo hemos realizado solo para los continentes. Por lo tanto, el nuevo fichero CSV queda así:



1	Country Name,Decade,Crecimiento PIB per cápita
2	Africa Eastern and Southern,60s,1,94 %
3	Africa Eastern and Southern,70s,0,30 %
4	Africa Eastern and Southern,80s,-0,33 %
5	Africa Eastern and Southern,90s,-0,88 %
6	Africa Eastern and Southern,2000s,1,68 %
7	Africa Eastern and Southern,2010-hoy,-0,07 %
8	Africa Western and Central,60s,0,86 %
9	Africa Western and Central,70s,3,38 %
10	Africa Western and Central,80s,-2,57 %
11	Africa Western and Central,90s,-0,23 %
12	Africa Western and Central,2000s,3,33 %
13	Africa Western and Central,2010-hoy,0,86 %
14	East Asia & Pacific,60s,5,77 %
15	East Asia & Pacific,70s,3,06 %
16	East Asia & Pacific,80s,3,66 %
17	East Asia & Pacific,90s,3,05 %
18	East Asia & Pacific,2000s,4,26 %
19	East Asia & Pacific,2010-hoy,4,08 %
20	Europe & Central Asia,60s,
21	Europe & Central Asia,70s,2,64 %
22	Europe & Central Asia,80s,1,76 %
23	Europe & Central Asia,90s,1,25 %
24	Europe & Central Asia,2000s,1,70 %
25	Europe & Central Asia,2010-hoy,1,24 %
26	Latin America & Caribbean,60s,2,64 %
27	Latin America & Caribbean,70s,3,61 %

Teniendo los datos de esta manera, ahora sí podemos realizar el gráfico de líneas. Seleccionamos los siguientes campos:



Eje X

Decade

Eje Y

Crecimiento PIB per c...

Eje Y secundario

Agregar campos de datos a...

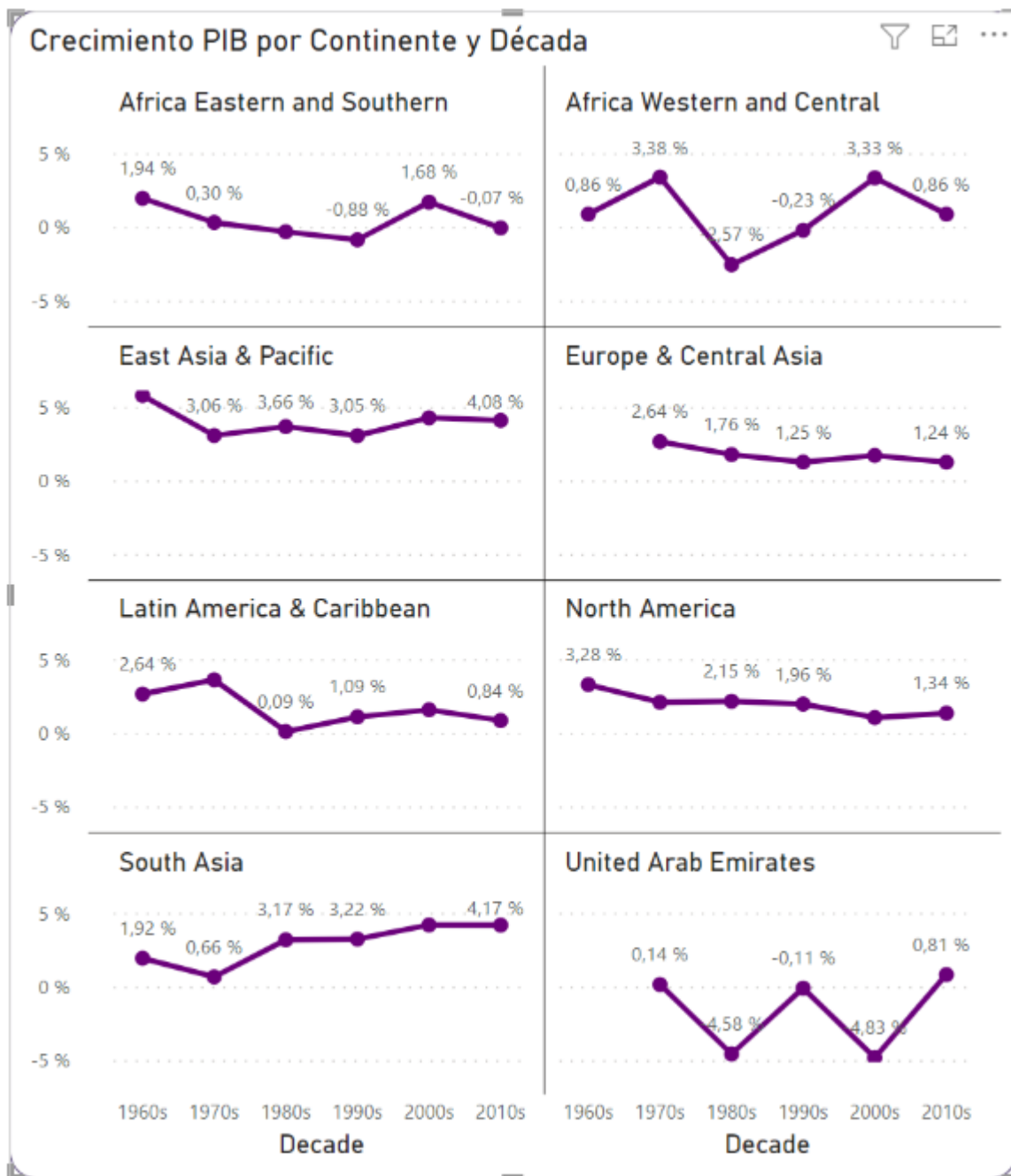
Leyenda

Agregar campos de datos a...

Múltiplos pequeños

Country Name

Como podemos observar, hemos añadido el nombre del país (en este caso, del continente) a Múltiplos Pequeños, para ver el análisis de la gráfica de línea por separado. Por lo tanto, nos queda de la siguiente forma:



A continuación, introducimos un mapa coroplético que refleja el PIB per cápita en los años 90 en Europa.

Cogemos los siguientes campos:

Ubicación

Country Name ▼ ×

Leyenda

Agregar campos de datos a...

Latitud

Agregar campos de datos a...

Longitud

Agregar campos de datos a...

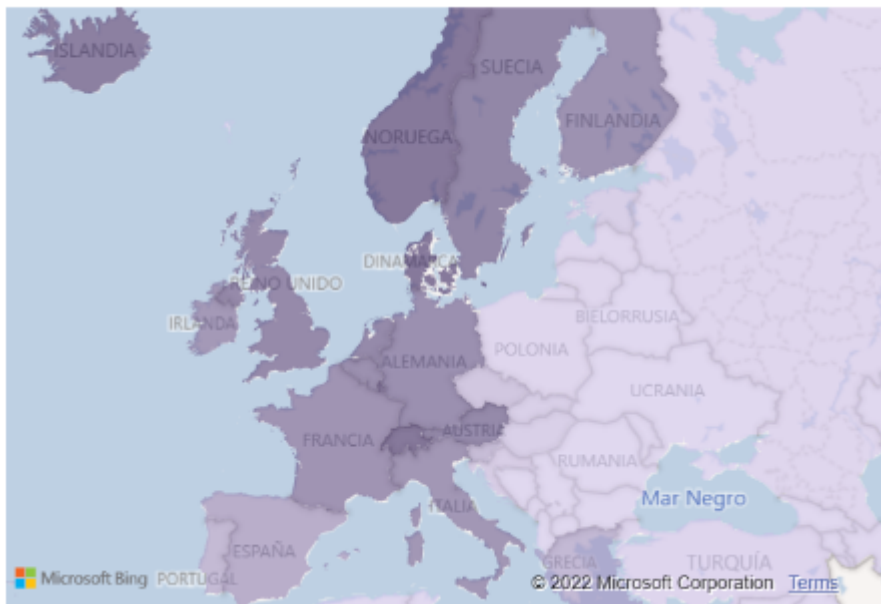
Información sobre herramien...

PIB per cápita 90s ▼ ×

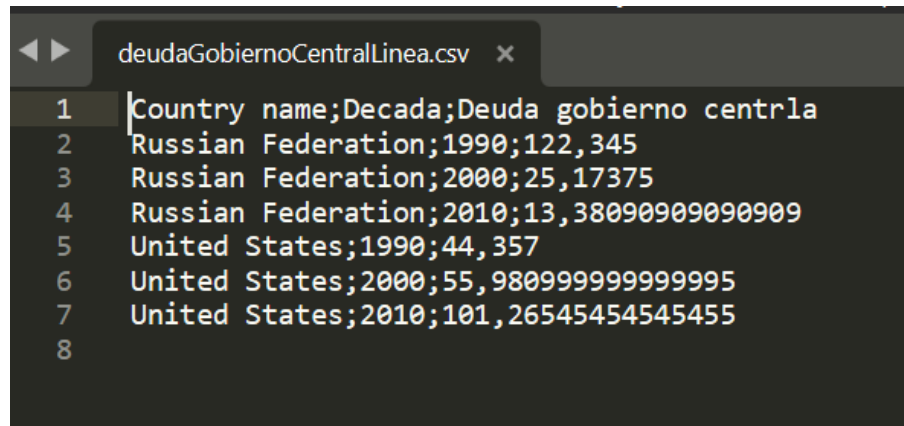
El PIB per cápita en Información sobre herramientas nos mostrará dicho dato de manera interactiva cuando podemos el ratón sobre un país concreto.

Seleccionamos una escala de morados según el PIB per cápita, y acercamos el mapa para que muestre la parte de Europa:

PIB per Cápita - Años 90 - Europa

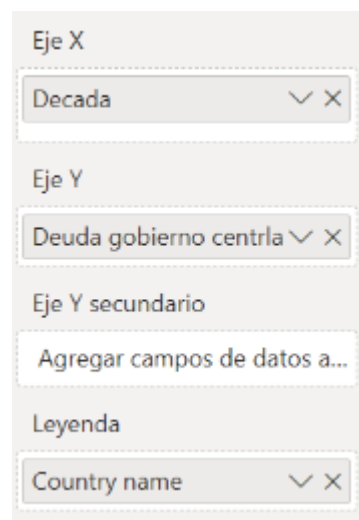


Por último, vamos a insertar un gráfico de líneas para ver la evolución de la deuda del gobierno central entre EE.UU. y Rusia. Para este gráfico, hemos tenido que modificar también el CSV, de una manera similar a como lo hicimos para el crecimiento del PIB. Nos queda de la siguiente forma:



```
1 Country name;Decada;Deuda gobierno centrla
2 Russian Federation;1990;122,345
3 Russian Federation;2000;25,17375
4 Russian Federation;2010;13,38090909090909
5 United States;1990;44,357
6 United States;2000;55,980999999999995
7 United States;2010;101,26545454545455
8
```

Con los datos así estructurados, procedemos a insertar el gráfico de líneas, cogiendo los siguientes campos:



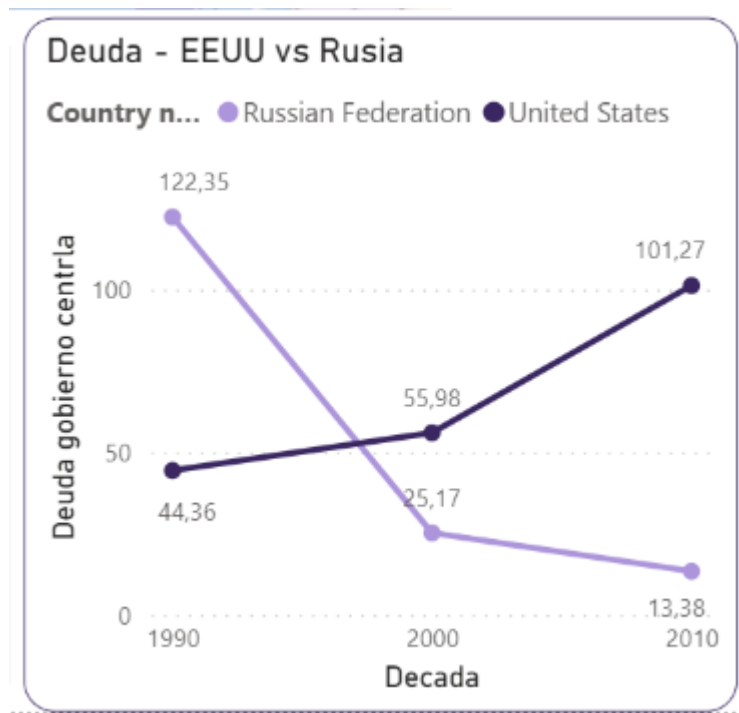
Eje X
Decada

Eje Y
Deuda gobierno centrla

Eje Y secundario
Agregar campos de datos a...

Leyenda
Country name

El gráfico nos queda, por lo tanto, así:



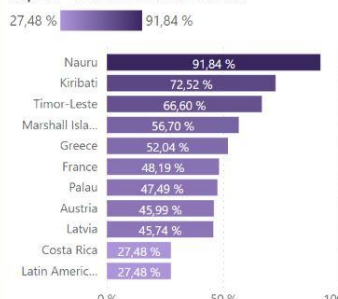
Por lo tanto, el informe completo del Crecimiento Económico es el siguiente:

Crecimiento de la Economía Mundial

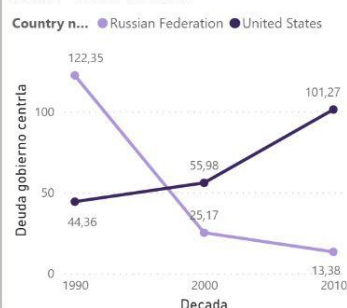
PIB per Cápita - Años 90 - Europa



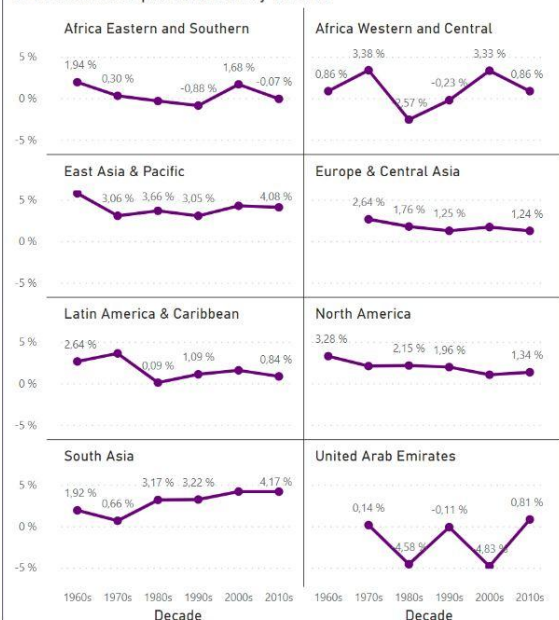
Top 10 - Países con más Gasto



Deuda - EEUU vs Rusia



Crecimiento PIB por Continente y Década



c. Ejercicio en Apache Superset

En lo que respecta a Apache Superset, para importar los datos con los que trabajaremos, tenemos que crear primero una conexión a una base de datos. En primer lugar, crearemos dicha base de datos en postgres, denominándola WDBank. Para conectarnos, en Settings, debemos elegir la opción de “Database Connection”. Al estar ejecutando Apache Superset en un contenedor de Docker, debemos ver la IP desde la cual está conectada Docker, y añadirla a la lista de hosts permitidos, tanto en el archivo postgresql.conf como en el archivo pg_hba.conf. Esta IP será la que indiquemos en el apartado de host. El puerto será el que usa Postgres por defecto, el 5432. En Database name indicaremos el nombre de la base de datos, y a continuación introduciremos el usuario con el que nos conectaremos ('postgres') y la contraseña que tengamos definida para dicho usuario.

Una vez hecho esto, podremos empezar a cargar los archivos CSV como datasets. Al igual que en los apartados anteriores, hemos tenido que realizar pequeños cambios en la configuración de los archivos (eliminar los '..0 y cambiar las , por ;), y hemos agrupado todos los datos por décadas calculando, para cada década, el promedio de todos los datos de cada año de forma individual, de forma que el trabajo se simplifique. Para cargar los datos en Apache Superset, tenemos que elegir, en las opciones de seguridad avanzada de la base de datos, la opción de permitir la subida de archivos a la base de datos. Así, nos aparecerá, en la parte superior derecha, la opción de “Upload CSV to database”. De esta forma, sólo tendremos que subir todos los archivos CSV a su correspondiente tabla. Elegiremos el nombre de tabla, seleccionaremos el archivo de nuestro ordenador y la base de datos a la que lo subiremos, pondremos el delimitador como ';' y haremos que, si la tabla existe, la reemplacemos (es decir, borrarla y crearla de nuevo).

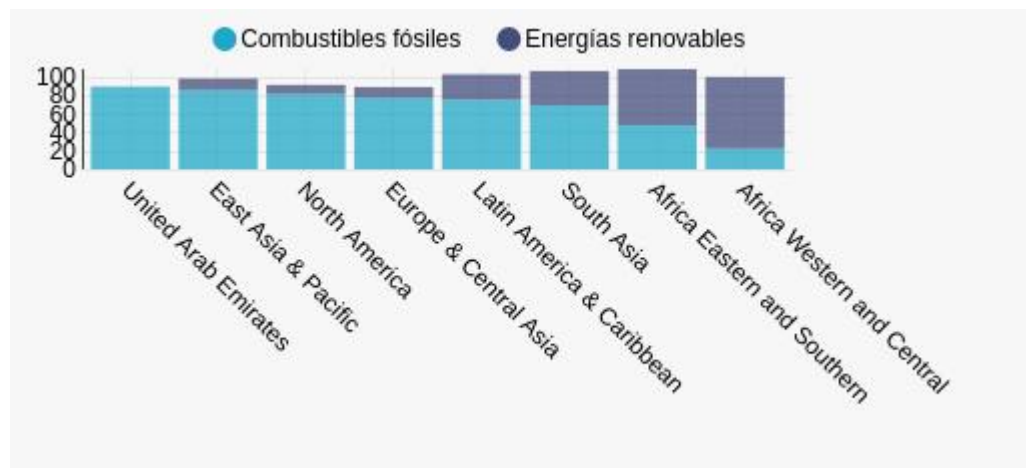
Con todos los datos importados, podemos crear los dos dashboards y empezar a trabajar en los gráficos. El primer dashboard contendrá los gráficos relacionados con el consumo energético (siempre en la década de 2010 en adelante): consumo de energías renovables en los países nórdicos vs. mediterráneos, consumo de combustibles fósiles, uso total de energía y uso de energía renovables vs. combustibles fósiles. El segundo dashboard contendrá los gráficos relacionados con el crecimiento de la economía: el top 10 de países con más gasto en la década de 2010 en adelante, el crecimiento del PIB por continente y década, la deuda del gobierno de EEU vs. Rusia y el PIB per cápita en Europa en los años 90.

Tabla sobre el porcentaje de uso de energías renovables en los países nórdicos (Dinamarca, Noruega y Suecia) y países del Mediterráneo (España, Portugal y Francia).

País	% Renovables
Norway	58.712
Sweden	49.333
Denmark	30.164
Portugal	27.749
Spain	16.34
Italy	15.597

Para este gráfico, cogemos los datos de Consumo de energías renovables, y creamos una tabla. En “query mode”, seleccionaremos raw data. En columns, seleccionaremos el nombre del país y los datos de la década de 2010, para los cuales hemos definido, previamente, en el dataset, las etiquetas que se mostrarán en la tabla, que serán país y % renovables. En “Filters”, seleccionaremos la columna país y la opción IN, y pondremos los nombres de los 6 países que nos interesan. Ordenamos los datos de mayor a menor, y ya tenemos el gráfico.

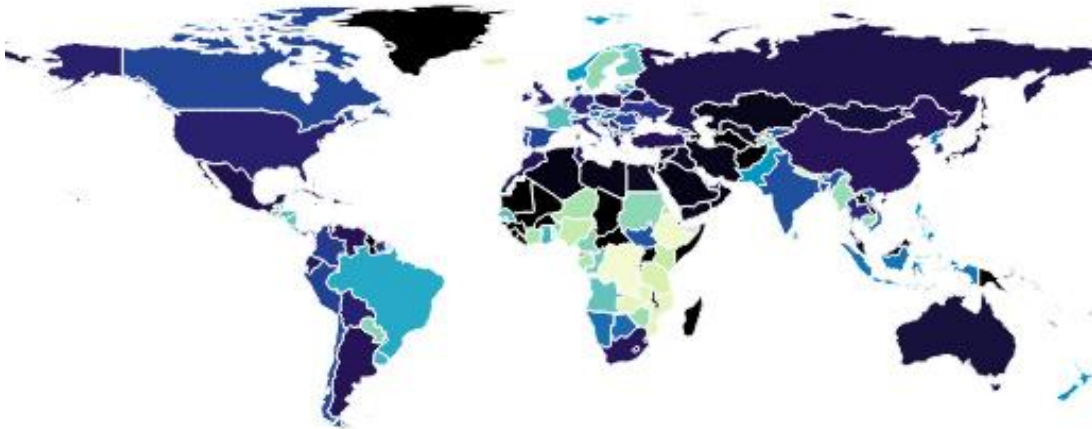
Gráfico de barras sobre consumo energético según el continente: % de energía renovable y % de fósiles



Para crear este gráfico, además de las modificaciones realizadas a los archivos CSV, tenemos que combinar, en un mismo archivo, los datos de combustibles fósiles y energías renovables. Para ello, simplemente hacemos, para cada década, una columna denominada Renovables1960, y otra denominada Fosiles1960, y hacemos esto para cada década. Con estos datos en el dataset, elegimos un gráfico de barras y seleccionamos como métricas, por una parte, la columna Fosiles2010 (etiquetada como “Combustibles fósiles”) y, por otra, la columna Renovables 2010 (etiquetada como “Energías renovables”). En la función de

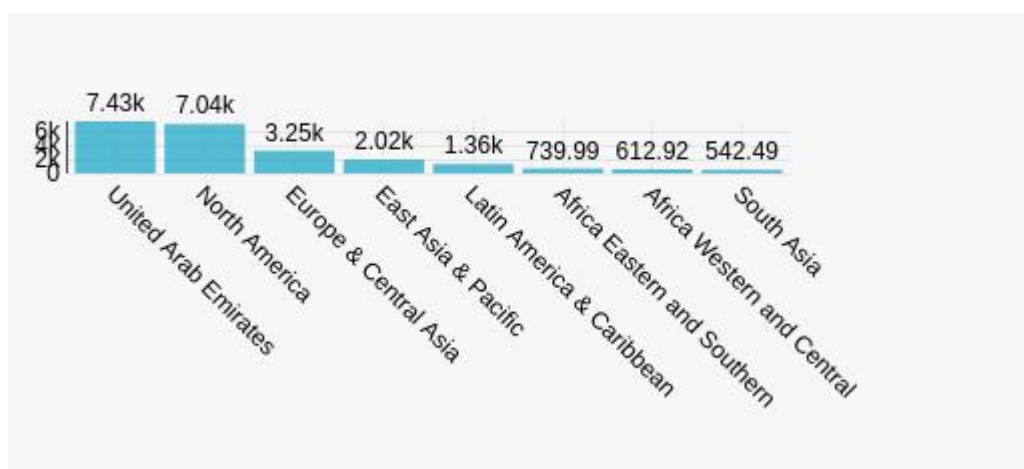
agregación, elegimos sum, aunque sólo haya un dato por cada región y década. Para que las barras estén apiladas, en la sección Customize, debemos seleccionar “Stacked bars”. En dimensions, seleccionaremos “Country name”. Para quedarnos sólo con los datos de los continentes, aplicaremos un filtro. En este caso, como en el CSV tenemos el código del país, filtraremos por esa variable, para que sea más sencillo, los códigos asociados a cada uno de los continentes. Respecto al diseño del gráfico,

Mapa mundial con el porcentaje de consumo de combustibles fósiles en la década de 2010 hasta hoy



En este caso, partimos del dataset que contiene el porcentaje de uso de combustibles fósiles. Elegimos un mapa del mundo, donde representaremos los datos usando la columna “Country code” (para lo cual elegiremos, en “Country field type”, la opción de código ISO cca3). Como metric, elegimos la columna de la década de 2010, y como función de agregación, AVG.

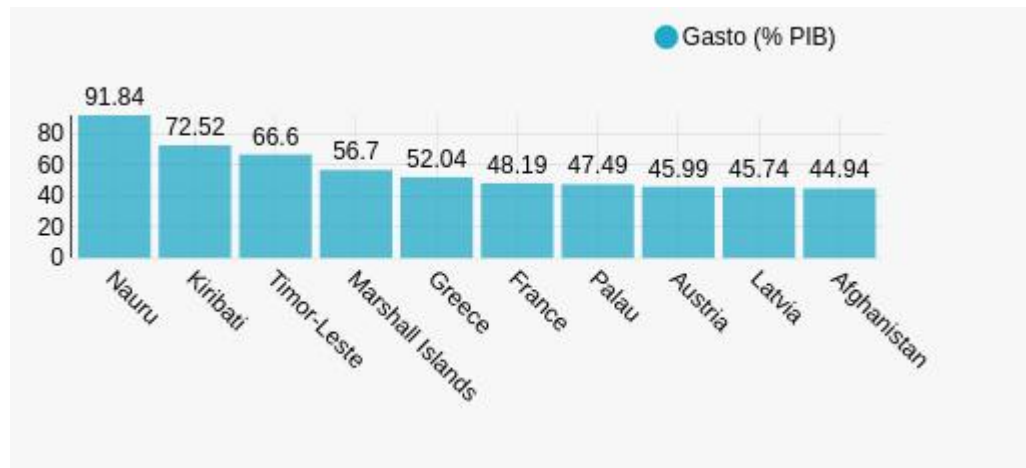
Gráfico de barras del uso de la energía total en la década de 2010



Para este gráfico, elegimos los datos de uso total de energía por décadas. De nuevo, en Metrics, elegimos la columna de la década de 2010 y SUM como función de agregación, y en el apartado de Filters, ponemos los mismo que en el anterior gráfico de barras donde

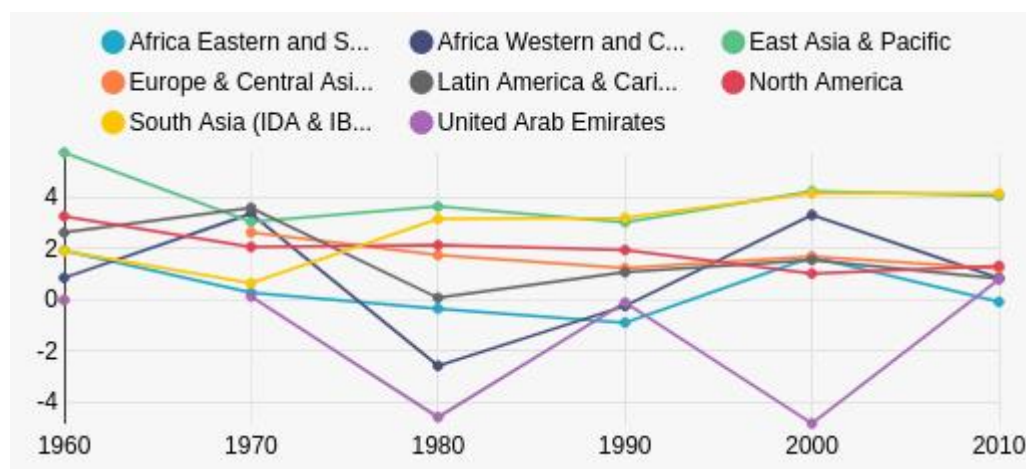
representamos el porcentaje de consumo de energías renovables y fósiles. De la misma forma, en dimensions, pondremos el nombre del país (en este caso, será el del continente).

Gráfico de barras sobre el gasto mundial de la década del 2010



Para este gráfico, el proceso es distinto al anterior. Ya que desde la sección de gráficos, Apache Superset no dispone de una opción para seleccionar directamente los 10 valores más altos (ya que estos están en columnas, no en filas), hemos de realizar una consulta, y guardar su resultado en un dataset, creando de esta forma una tabla virtual. La consulta es sencilla: seleccionamos el nombre del país y los datos de la década de 2010 de la tabla donde tenemos almacenados los datos, eliminando en el proceso los valores nulos (es decir, WHERE Datos2010 IS NOT NULL), agrupamos por el nombre del país, ordenamos por el gasto de forma descendente, y cogemos los 10 primeros valores (con LIMIT 10). Así, sólo nos queda hacer un gráfico de barras a partir de este dataset virtual, donde elegiremos en metrics la columna con los datos del gasto y en dimensions el nombre del país.

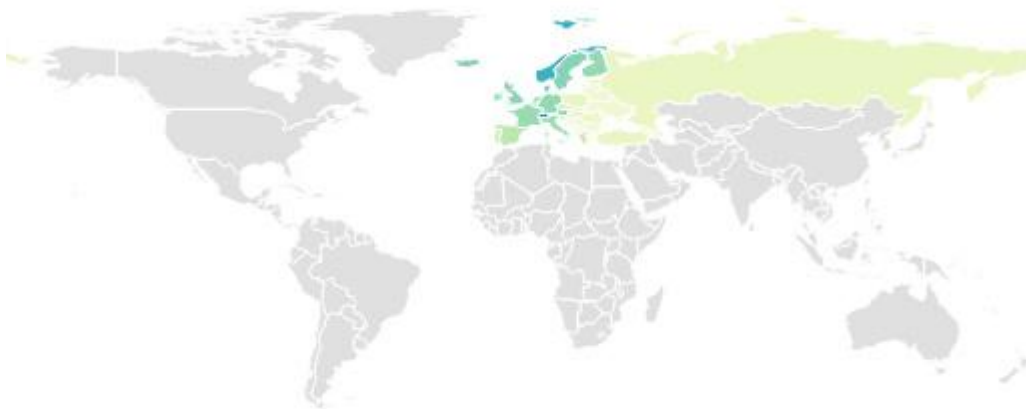
Gráfico de línea sobre el crecimiento del PIB de 1960 a 2021 en los diferentes continentes



Para este gráfico, tenemos que agrupar, mediante una consulta a la base de datos, los datos de los continentes, de forma que en cada fila del archivo CSV que usamos para crear el dataset tengamos el país (continente), la década, y el porcentaje de crecimiento del PIB.

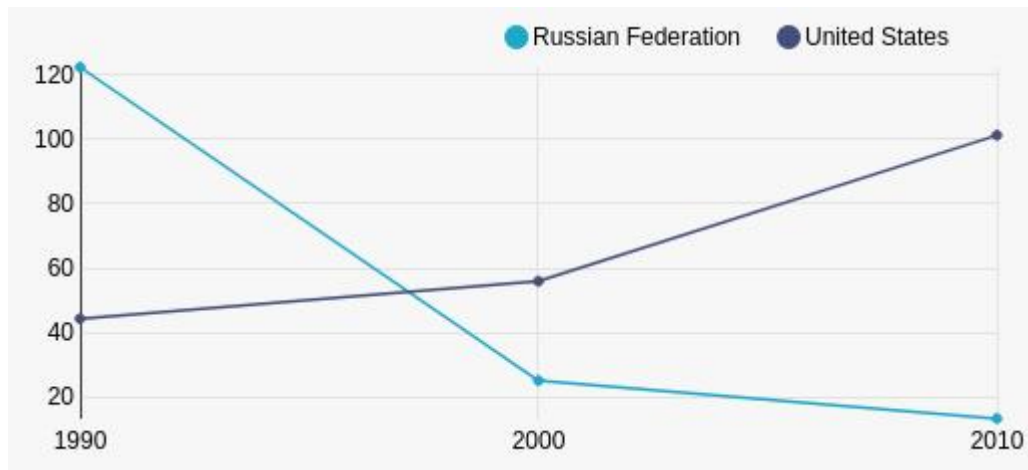
Con estos datos cargados como un nuevo dataset, elegimos como “Time column” la columna que contiene datos de la década, y en “Time Grain” seleccionamos “original value”. En Query, seleccionamos la columna de la deuda, y aplicamos SUM como agregación. Por último, en Dimensions, seleccionamos Country name. Cabe destacar que Apache Superset nos permite mostrar los puntos para cada uno de los valores en el gráfico, pero en la versión utilizada y estándar, no permite mostrar el valor en cada uno de estos puntos.

Mapa sobre el PIB per cápita en Europa de 1990 a 2000



Para realizar este mapa, partimos del dataset que contiene los datos del PIB per cápita agrupados por década. Apache Superset no nos permite crear un mapa personalizado donde elijamos nosotros los países (sólo nos deja crear un mapa para un único país, mostrando sus subdivisiones, pero no un mapa del mundo con sólo algunos países). Por lo tanto, tenemos que elegir la opción de un mapa del mundo. Para saber qué país es cual en el mapa, haremos igual que en el anterior mapa, seleccionando la columna Country Code y, en Country Field Type, cca3. La métrica serán los datos de la década de 1990, a la cual aplicaremos la función SUM para que Apache nos deje representar los datos. Para elegir sólo los países de Europa, ya que partimos del dataset que contiene todos los países del mundo, tendremos que hacer un filtro. Para hacerlo más cómodo, filtraremos los datos por el código del país, por lo que tendremos que introducir en el filtro los 53 países que representaremos.

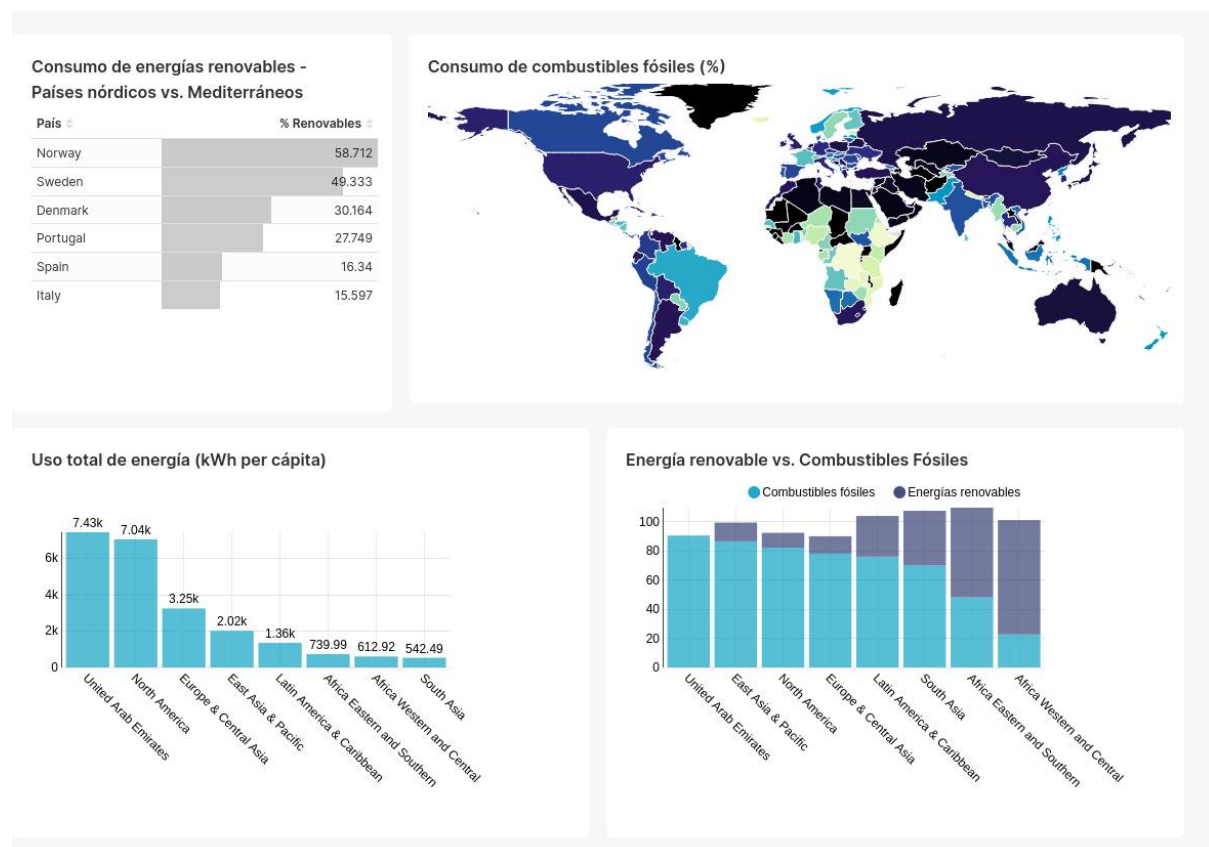
Gráfico de línea comparativo entre Rusia y Estados Unidos sobre la deuda del gobierno central a partir de los 90



Para este gráfico, tenemos un problema similar al anterior gráfico de líneas: tenemos que tener una línea para cada país y década. Esto lo podemos hacer con el SQL Query Tool, lo cual, además, nos guarda los datos en un dataset virtual. Con estos datos, creamos un “Line Chart”, donde la “Time Column” será la década; la métrica, la deuda; y las dimensiones, el nombre del país.

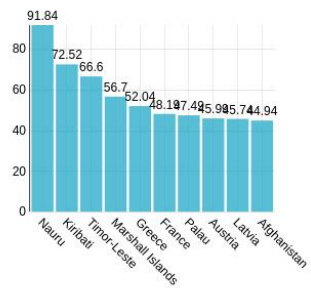
Panel de control

De esta forma, el panel de control de los datos de consumo de energía quedaría así:

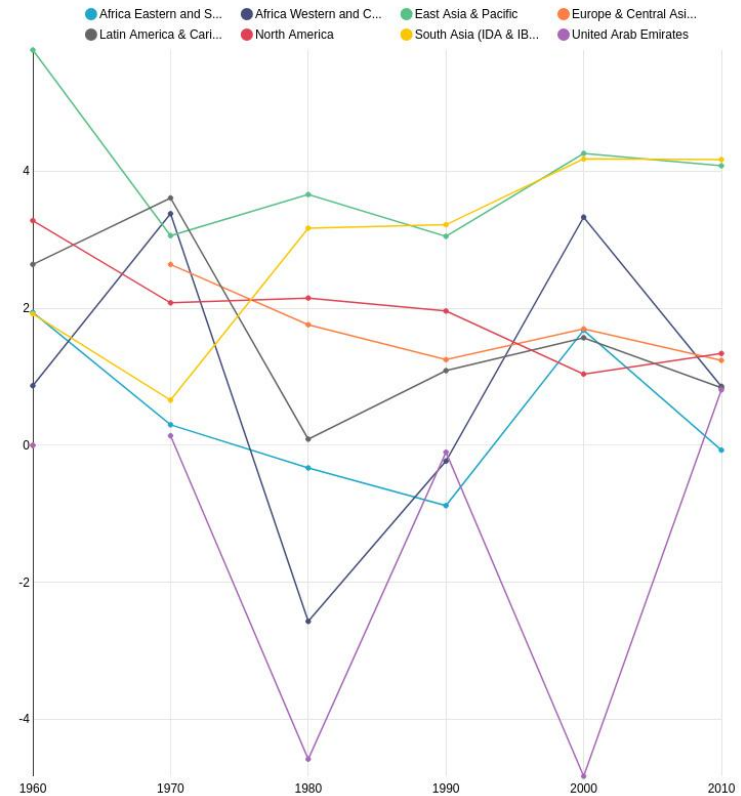


Por otra parte, el panel de control de los datos de crecimiento económico sería el siguiente:

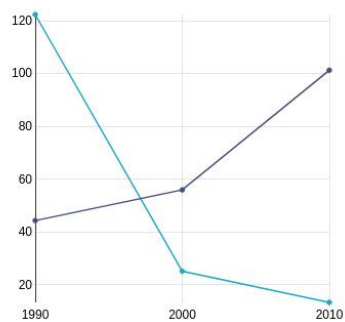
Top 10 países con más gasto



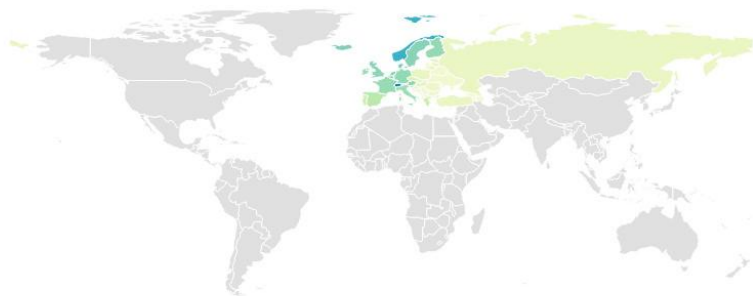
Crecimiento del PIB por continente y década



Deuda-EEUU vs Rusia



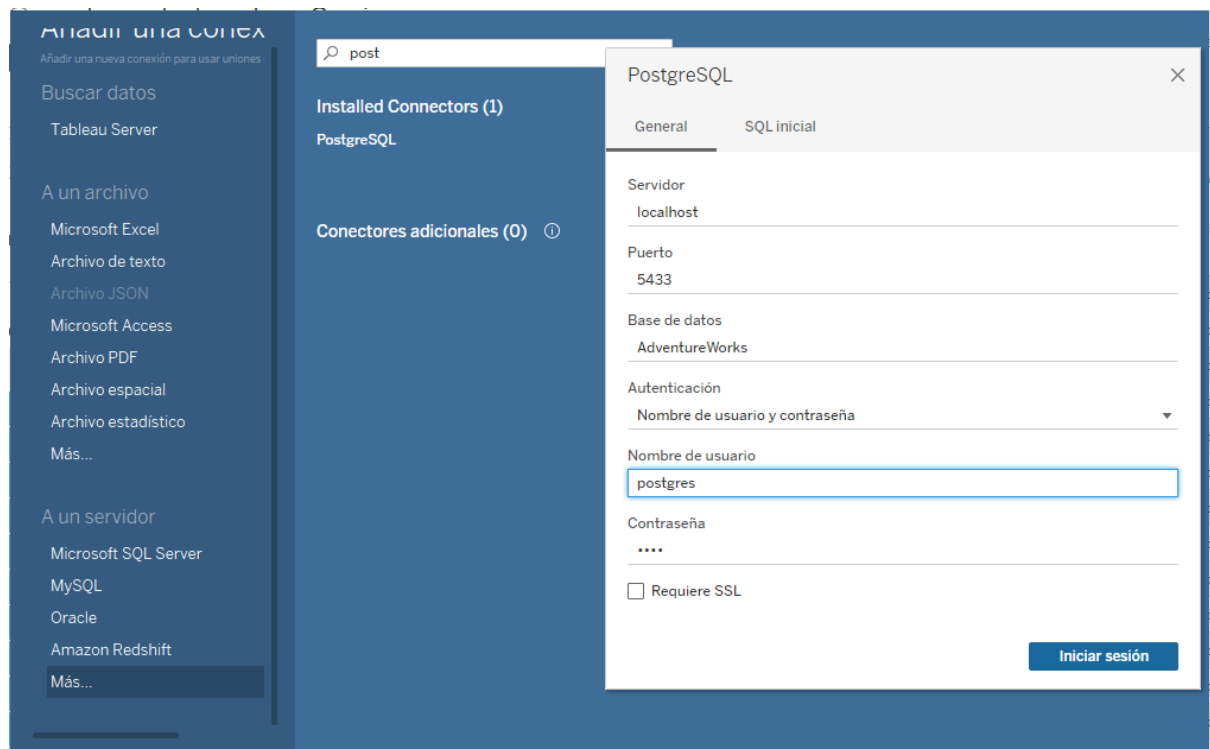
PIB Per Cápita-Europa-Años 90



2. Conexión a Adventure Works y creación de vistas que muestren el número de clientes, ventas por divisa y otros temas de interés.

a. Ejercicio en Tableau

Para empezar nos conectamos a la base de datos, clicando a “Añadir una conexión” -> A un servidor -> Postgresql. Introducimos los datos relacionados con mi base de datos:



Esperamos a que se ejecute la consulta y se crea una fuente de datos de la base de datos:

AdventureWorks

Conexión: ☒ En tiempo real ☐ Extraer

Filtros: 0 | Añadir

salesorderheader

¿Necesita más datos?

salesorderheader 25 campos 31465 filas 100 filas

Nombre	salesorderheader
Campos	
Tipo	Nombre de campo
#	salesorderid
#	revisionnumber
📅	orderdate
📅	duedate
📅	shipdate
#	status
Abc	onlineorderflag

salesorderid	revisionnumber	orderdate	duedate	shipdate
43.659	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.660	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.661	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.662	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.663	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.664	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.665	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.666	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.667	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.668	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/
43.669	8	31/05/2011 00:00:00	12/06/2011 00:00:00	07/06/

Nos conectamos a las tablas customer, salesorderheader y currency. Hacemos una combinación de tablas:

territoryid (salesorderheader.csv) = territoryid (customer.csv)
customerid (salesorderheader.csv) = customerid (customer.csv)
rowguid (salesorderheader.csv) = rowguid (customer.csv)
....

Procedemos a realizar la creación de las vistas a partir de estos datos:

Tabla sobre el número de clientes por territorio

Ponemos el territorio y continente en las columnas. Hacemos un filtro de los territorios elegidos. Finalmente, seleccionamos la opción recuento para el número de clientes.

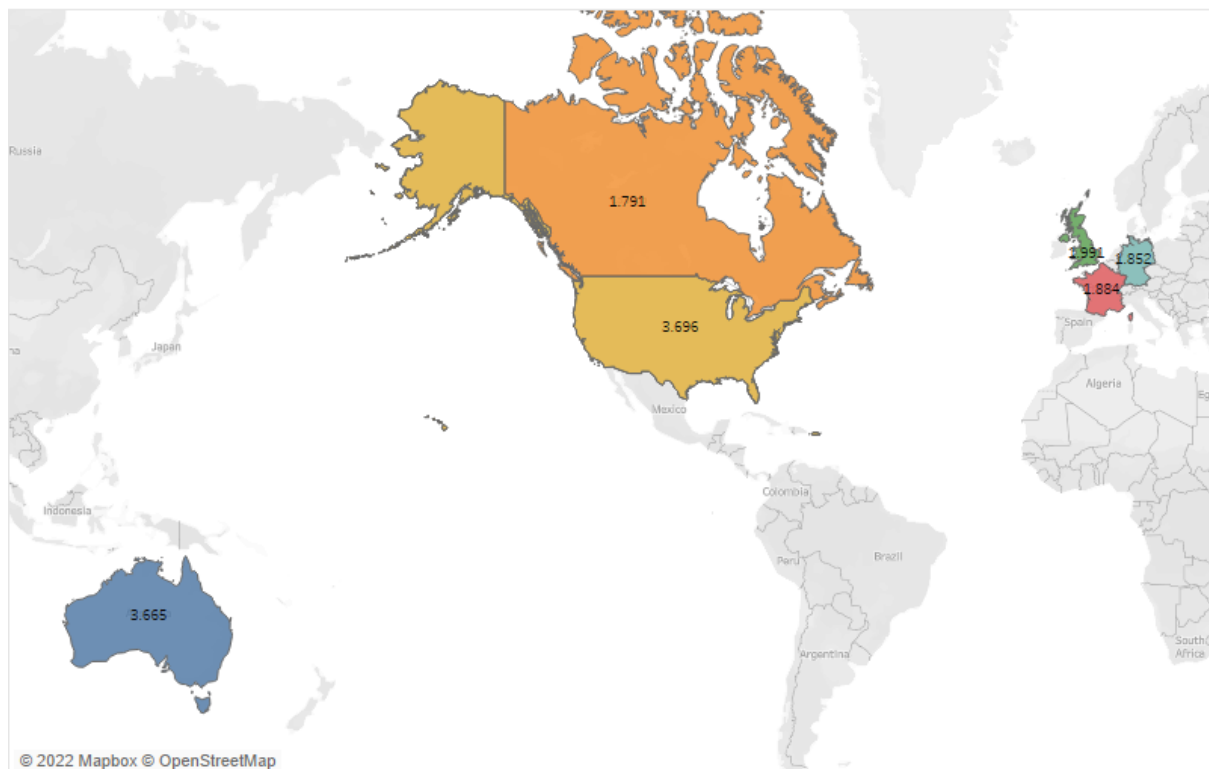
Numero de clientes por territorio

Territorio	continente	
USA Southwest	North America	4.696
USA Southeast	North America	3.696
Australia	Pacific	3.665
United Kingdom	Europe	1.991
France	Europe	1.884
Germany	Europe	1.852
Canada	North America	1.791
USA Central	North America	132
USA Northeast	North America	113
Total general		19.820

Numero de clientes por país

Convertimos la columna territorio a una función geográfica por país/región. Seleccionamos la longitud generada de columna y la latitud generada por filas. Hacemos un recuento de clientes y lo añadimos como una etiqueta. Aplicamos un filtro de los territorios seleccionados y le ponemos un color a cada uno.

Numero de clientes por país

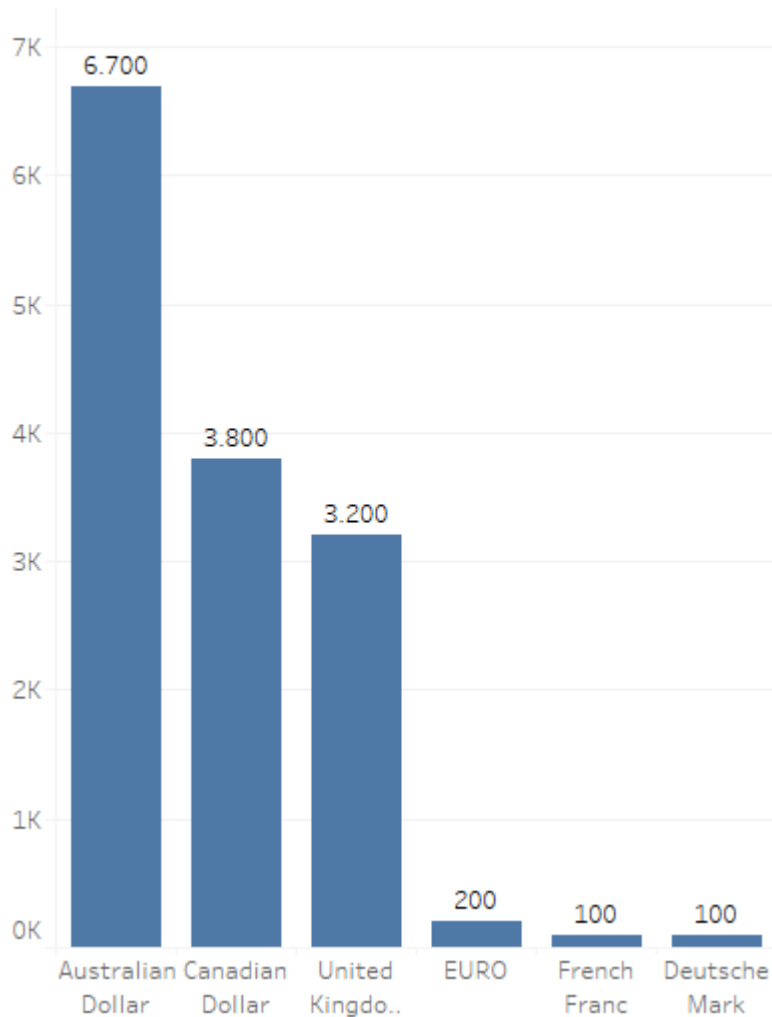


Hacemos un recuento de customer ir por country

Número de Venta por Divisa

Seleccionamos el número de venta, lo ponemos en la fila y aplicamos un sum. Luego, ponemos la divisa en la columna.

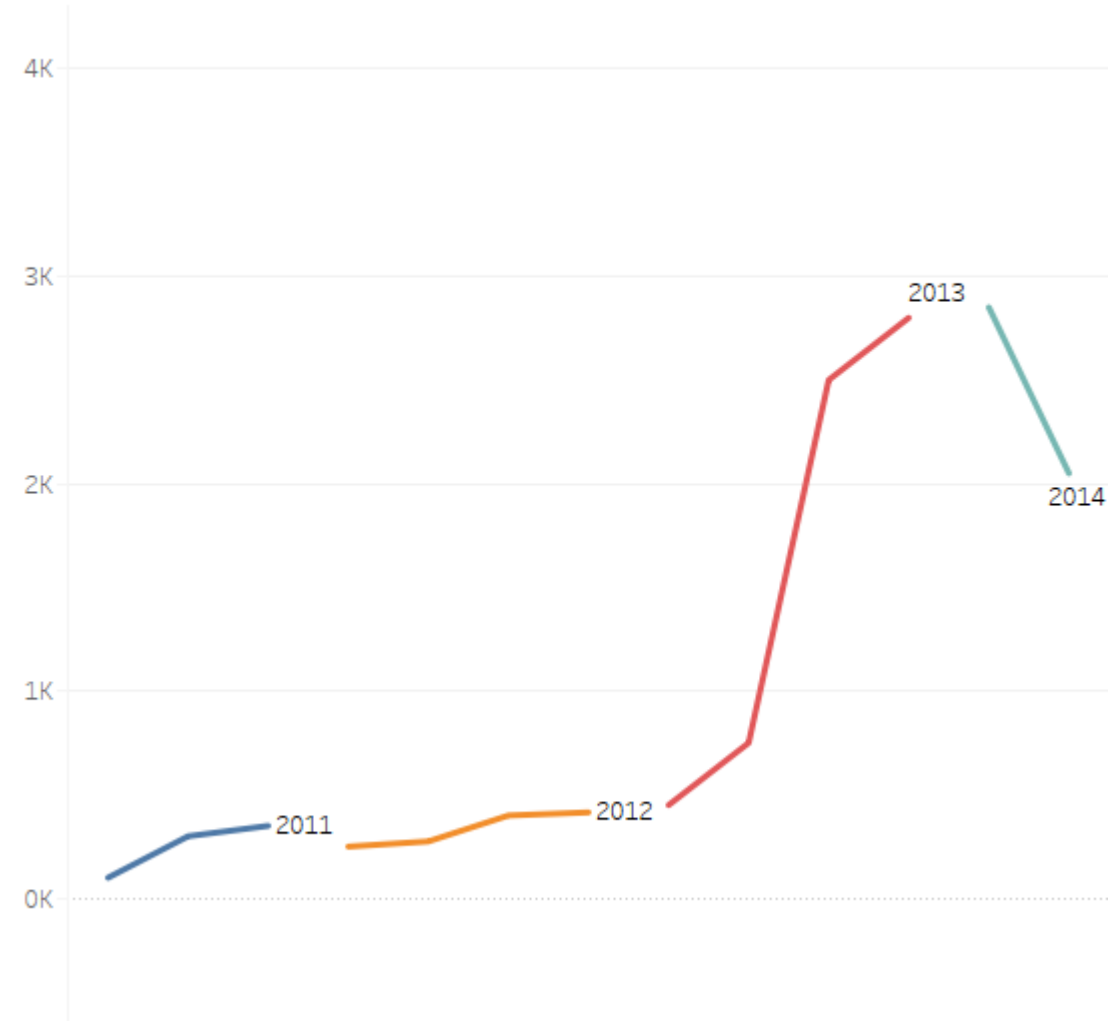
Número de venta por divisa



Numero de ventas por Trimestre

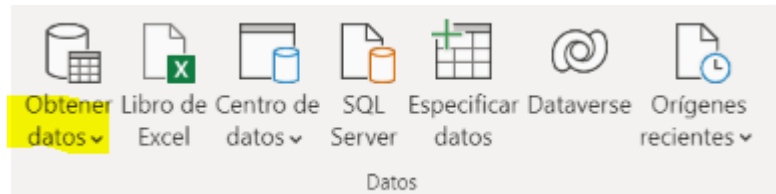
Ponemos las salesorderid en la fila y seleccionamos la opcion de recuento. Despues creamos una jerarquia de año y trimestre y lo ponemos en columna. Finalmente, aplicamos un filtro de años por color.

Numero de ventas por trimestre

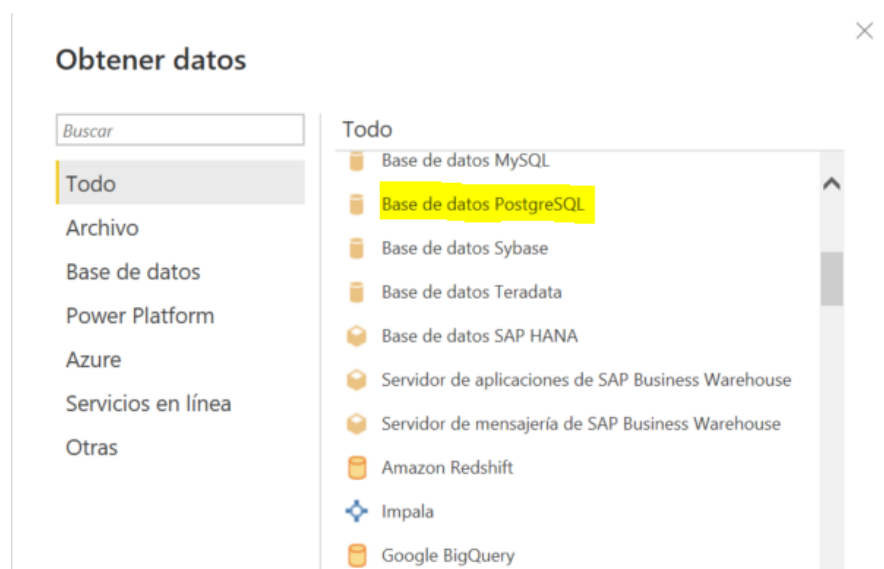


b. Ejercicio en Power BI Desktop

Para conectarnos a la base de datos de Adventure Works, que tenemos guardada en PostgreSQL, acudimos a la zona Datos -> Obtener Datos de Power BI:



Pinchamos en la opción Base de datos PostgreSQL:



Introducimos los datos de nuestro servidor, y la base de datos a la que nos queremos conectar:

Base de datos PostgreSQL

Servidor

127.0.0.1

Base de datos

Adventureworks

Modo Conectividad de datos ⓘ

☒ Importar

☐ DirectQuery

▸ Opciones avanzadas

En nuestro caso, vamos a utilizar la versión que contiene 11 schemas, ya que dispone de más información de clientes y ventas que nos será útil para el análisis:

Con los datos ya cargados, introducimos en primer lugar una tabla con el número de clientes por territorio, que vendrá acompañada de un mapa coroplético.

Utilizamos la tabla 'customer', donde hacemos uso de la herramienta "Combinar Consultas" de Power Query para obtener el nombre y el grupo del territorio a partir del ID del territorio. La tabla con la que combinamos es la de 'salesterritory'.

Combinar

Seleccione una tabla y las columnas coincidentes para crear una tabla combinada.

customer

customerid	personid	storeid	territoryid	rowguid	modifieddate	rowid
1		934	1	3f5ae95e-b87d-4aed-95b4-c3797afcb74f	12/09/2014 11:15:07	USA
2		1028	1	e552f657-a9af-4a7d-a645-c429d6e02491	12/09/2014 11:15:07	USA
7		930	1	03e9273e-b193-448e-9823-fe0c44aed78	12/09/2014 11:15:07	USA
3		642	4	130774b1-db21-4ef3-98c8-c104bcd6ed6d	12/09/2014 11:15:07	USA

salesterritory

territoryid	name	country	countryregioncode	group	salesytd	saleslastyear	costytd
1	USA Southeast	USA	US	North America	78871867882	32986944938	
2	USA Northeast	USA	US	North America	24021768476	36071489371	
3	USA Central	USA	US	North America	3072175118	32050140767	
4	USA Southwest	USA	US	North America	1,05109E+11	53665757098	

Tipo de combinación

Externa izquierda (todas de la primera, coincidencias...)

☐ Use las coincidencias aproximadas para comparar la combinación.

► Opciones de coincidencia aproximada

✓ La selección coincide con 19820 de 19820 filas de la primera tabla.

Aceptar

Cancelar

Los campos seleccionados para la tabla son:

Columnas		
Territorio	▼	×
Continente	▼	×
Número Clientes	▼	×

De manera que la tabla queda así:

Territorio	Continente	Número Clientes
USA Southwest	North America	4696
USA Southeast	North America	3696
Australia	Pacific	3665
United Kingdom	Europe	1991
France	Europe	1884
Germany	Europe	1852
Canada	North America	1791
USA Central	North America	132
USA Northeast	North America	113
Total		19820

El mapa coroplético, por su parte, mostrará de forma gráfica el contenido de la tabla. Seleccionamos para él los siguientes campos:

Ubicación

country ▼ ✕

Leyenda

Agregar campos de datos a...

Latitud

Agregar campos de datos a...

Longitud

Agregar campos de datos a...

Información sobre herramien...

Recuento de custome... ▼ ✕

Al pasar el ratón sobre un país, se mostrará el número de clientes de éste.

Aplicamos una escala de morados en función del número de clientes, de manera que el mapa se vea de la siguiente forma:



Continuamos con un gráfico de columnas apiladas que reflejará el número de ventas por divisa. De forma parecida a como hemos hecho en el gráfico anterior, vamos a utilizar la herramienta “Combinar Consulta”, para poder ver, en la tabla ‘salesorderheader’, el nombre de la divisa a partir de su ID. Combinamos, por tanto, con la tabla ‘currency’.

Combinar

Selecione una tabla y las columnas coincidentes para crear una tabla combinada.

salesorderheader

idressid	shiptoaddressid	shipmethodid	creditcardid	creditcardapprovalcode	currencyrateid	subtotal
20425	20425	1	11150	335093Vi57504	2	33996
23967	23967	1	9497	1234632Vi48985	15	35784
16025	16025	1	1988	835095Vi10342	15	33745
27345	27345	1	7205	1035101Vi37381	15	33996

currencyrate

currencyrateid	currencyratedate	fromcurrencycode	tocurrencycode	averagerate	endofdayrate	modified
1	31/05/2011 0:00:00	USD	ARS	1	10002	31/05
2	31/05/2011 0:00:00	USD	AUD	15491	155	31/05
3	31/05/2011 0:00:00	USD	BRL	19379	19419	31/05
4	31/05/2011 0:00:00	USD	CAD	14641	14683	31/05

Tipo de combinación

Externa izquierda (todas de la primera, coincidencias...

☐ Use las coincidencias aproximadas para comparar la combinación.

Opciones de coincidencia aproximada

☒ La selección coincide con 13976 de 13976 filas de la primera tabla.

Aceptar

Cancelar

De nuevo en Power Query, escogemos los siguientes campos:

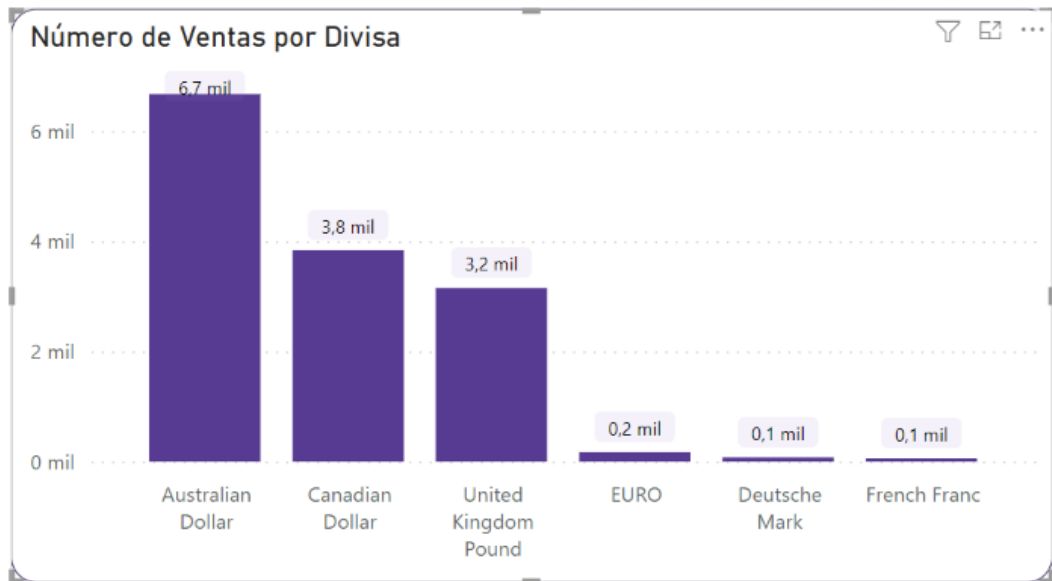
Eje X

currencyname

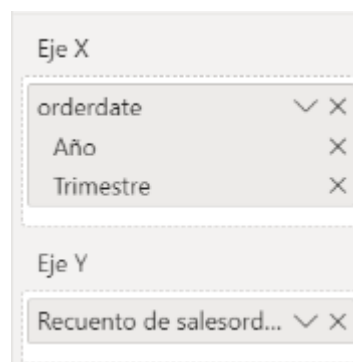
Eje Y

Recuento de salesord...

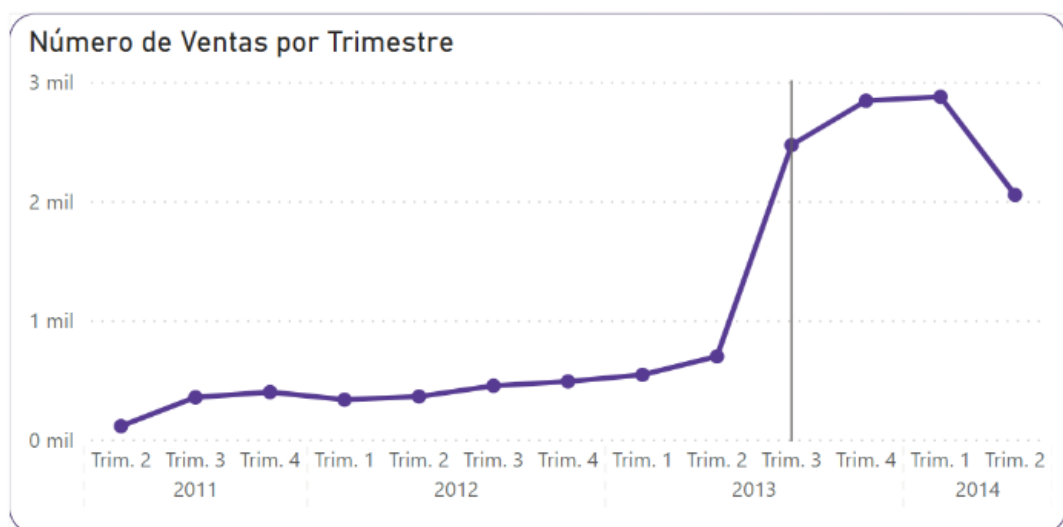
Quedando el gráfico de la siguiente manera:



Por último, hemos decidido implementar un gráfico de línea que refleje el número de ventas por trimestre. Aprovechamos, para esto, la jerarquía de fechas del conjunto de datos:



Mostramos los datos de la siguiente manera en el gráfico:



c. Ejercicio en Apache Superset

En el caso de los ejercicios en Superset con esta base de datos, nos conectaremos de forma análoga a como hicimos en el anterior apartado en este programa. Como ya tenemos los datos en la base de datos, no es necesario crear directamente ningún dataset, ya que podemos emplear la herramienta del SQL Lab para crear una dataset virtual, y usar esos datos para hacer los distintos gráficos. De esta forma, el panel de control que obtenemos es el siguiente:



Tabla sobre el número de clientes por territorio

Territorio	Continente	Número Clientes
Southwest	North America	4696
Australia	Pacific	3665
Northwest	North America	3520
United Kingdom	Europe	1991
France	Europe	1884
Germany	Europe	1852
Canada	North America	1791
Southeast	North America	176
Central	North America	132
Northeast	North America	113

Para realizar este gráfico, hemos realizado la siguiente consulta. Obtenemos los datos de la tablas de clientes y de ventas por territorio, y los agrupamos con un join por el id del territorio. Nos quedamos con los datos del nombre del territorio, el grupo (continente al que pertenecen) y la suma de los ids de los clientes como el total de clientes.

```
SELECT te.name, te.group, count(cu.customerid) AS NumeroClientes
FROM sales.customer AS cu JOIN sales.salesterritory AS te ON
cu.territoryid=te.territoryid
GROUP BY te.group, te.name
```

Para elaborar el gráfico, elegimos desde la herramienta de SQL Lab la opción del gráfico, y seleccionamos la tabla. Elegiremos el modo “Raw records”, y en las columnas seleccionaremos los campos name, group y numeroclientes (podemos renombrarlos para que reciban los nombres que queremos).

Número de clientes por país



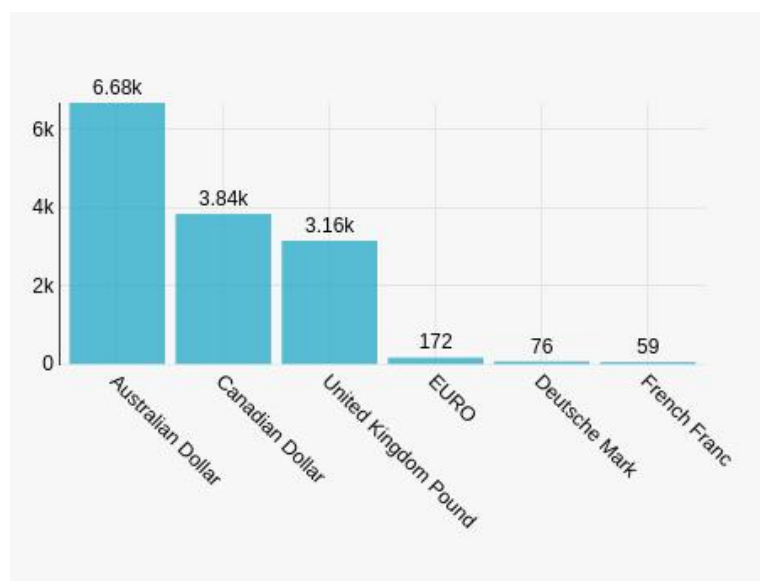
Para este gráfico, la consulta realizada es la siguiente. En este caso, obtenemos los datos igual que en el anterior apartado, con la diferencia de que en vez de seleccionar los nombres de los países, seleccionaremos su código, agrupando también los datos por el

código de país. De esta forma, podremos obtener el total de clientes que hay en cada país, y representarlo fácilmente en el mapa.

```
SELECT te.countryregioncode, count(cu.customerid) AS
numeroclientes
FROM sales.customer AS cu JOIN sales.salesterritory AS te ON
cu.territoryid=te.territoryid
GROUP BY te.countryregioncode
```

Para el gráfico, seleccionamos el mapa del mundo; en el campo “Country column” pondremos la columna con los códigos de los países, y en “Country Field Type”, cca2. Para la métrica, seleccionaremos la columna numeroclientes, y aplicaremos el agregador AVG.

Número de ventas por divisa

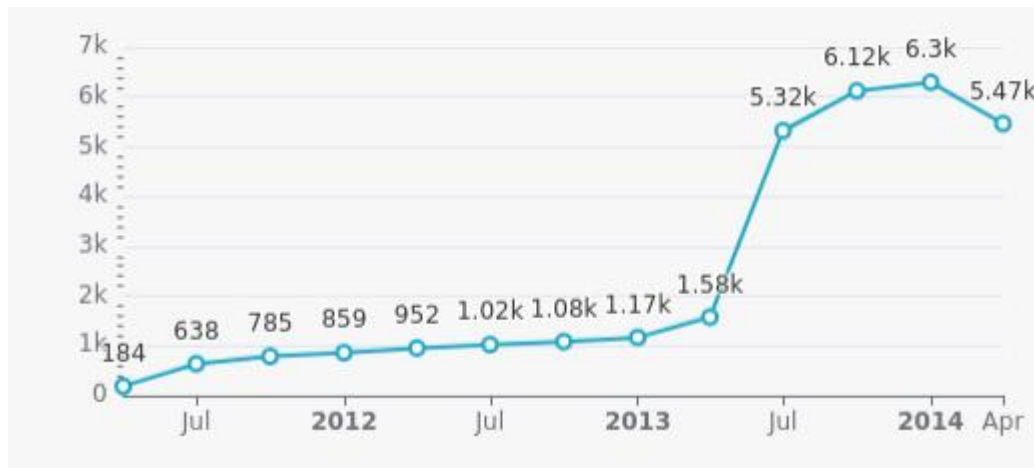


Para este gráfico, la consulta es más compleja. Seleccionamos los datos de tres tablas: salesorderheader, con datos de los pedidos; currencyrates, con datos del cambio de divisas; y currency, con datos sobre las monedas. Para las dos primeras tablas, hacemos el join en el id de cambio de divisas; para las dos últimas, en el código de divisa. Esto lo hacemos así porque, en la tabla de ventas, tenemos la referencia del intercambio de divisas, y en dicha tabla sólo se almacena el código, no el nombre. Seleccionamos, entonces, el código de divisa (por si fuese necesario), el nombre, y contamos el número total de ventas usando el atributo salesorderid.

```
SELECT cur.tocurrencycode, cu.name, count(sa.salesorderid) as
totalventas
FROM sales.salesorderheader AS sa
JOIN sales.currencyrates AS cur on
sa.currencyratesid=cur.currencyratesid
JOIN sales.currency AS cu on cu.currencycode=cur.tocurrencycode
GROUP BY cur.tocurrencycode, cu.name
```

En lo que respecta al gráfico, elegiremos un gráfico de barras, donde la métrica será el total de ventas, usando como agregador AVG; y las dimensiones, el nombre de la divisa. Además, en customize, activaremos la opción de mostrar los valores de cada barra.

Número de ventas por trimestre



Para este último gráfico, la consulta es la siguiente. Seleccionamos de la tabla la fecha, y contamos el número de ids de las ventas para obtener el total de ventas. Esto nos producirá como resultado un dataset en el que, para cada día, tenemos el total de ventas de ese día.

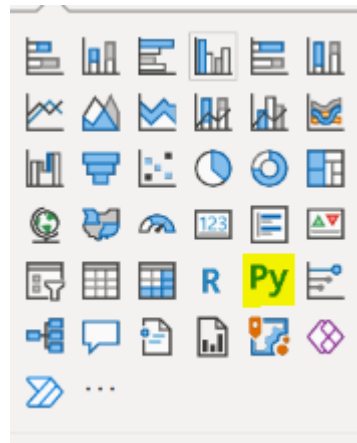
```
SELECT orderdate as date, count(salesorderid) AS totalventas
FROM sales.salesorderheader
GROUP BY date
```

Para crear el gráfico, seleccionaremos en el dataset, para la columna date, isTemporal, de forma que podamos usarlo como Time Column. Así, al seleccionar el gráfico de líneas, podemos seleccionar esta columna como Time Column, y elegir un Time Grain de Quarter, para que los datos se agrupen por trimestre. Sólo tenemos que seleccionar totalventas como metric, y tendríamos listo nuestro gráfico.

3. Visualización realizada en Python

Para la visualización en Python con Power Bi, vamos a mostrar un gráfico de barras del % de descuento que tienen las distintas ofertas de Adventureworks.

En Power Bi, en Visualización, introducimos un objeto visual de Python:



Y seleccionamos los campos que vamos a utilizar:



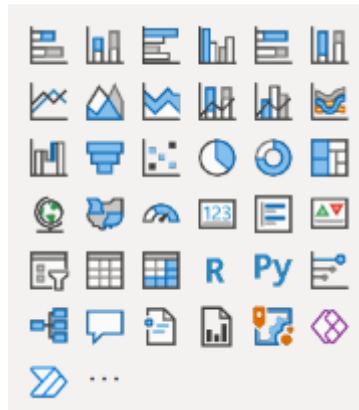
En la zona del script, escribimos en código Python la ejecución de una gráfica de barras a partir del conjunto de datos dado:

```
6 # Pegue o escriba aquí el código de script:
7
8 import matplotlib.pyplot as plt
9 dataset.plot(kind='bar', x='description', y='discountpct')
10 plt.show()
```

Por motivos de la configuración de Power Bi, la aplicación Power Bi Desktop presenta problemas a todos los usuarios que intentan mostrar un objeto visual de Python. Por lo tanto, no ha sido posible que nos muestre el gráfico en sí, si bien todos los pasos han sido seguidos como en la guía de Microsoft Power Bi.

4. Visualización realizada en R

La visualización en R con Power Bi va a contener el número de ventas por territorio en el último años. Añadimos un objeto visual de R:



Al que añadimos los campos involucrados:

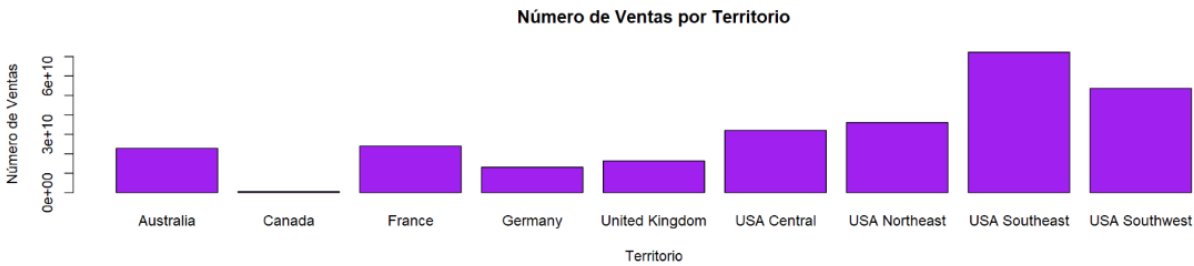


En la parte del script, indicamos que se muestre el número de ventas del último año, pasando como nombres de las barras los nombres de los territorios. Además, modificamos el nombre del gráfico en sí, de los ejes, el color de las barras y el borde de éstas.

```
6 # Pegue o escriba aquí el código de script:
7
8 sales <- table(dataset$saleslastyear)
9 |
10 barplot(dataset$saleslastyear, names.arg=dataset$name,
11         main="Número de Ventas por Territorio",
12         xlab="Territorio", ylab="Número de Ventas",
13         col="purple", border="black"
14 )
```

Con todo ello, el gráfico se ve de la siguiente manera:

Gráfico con R

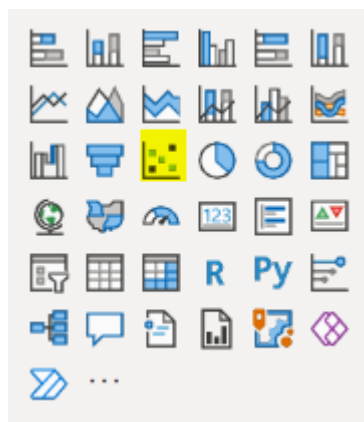


5. Gráfico de Burbujas

Por último, introducimos un gráfico de burbujas sobre el conjunto de datos del Crecimiento de la Economía Mundial.

El gráfico de burbujas va a presentar la correlación entre las variables de % de consumo energético y % de consumo de combustibles fósiles. Además, el tamaño de la burbuja en sí representará el Uso de Energía Total.

Insertamos, por tanto, un gráfico de dispersión:



Para evitar un elevado número de burbujas, vamos a aplicar el filtro de solo continentes. En los campos, seleccionamos los siguientes:

Valores
Agregar campos de datos a...
Eje X
Consumo Renovables ... ∨ ×
Eje Y
Consumo Fósiles 201... ∨ ×
Leyenda
Country Name ∨ ×
Tamaño
Uso Energía Total 201... ∨ ×

Por lo tanto, nuestro gráfico de burbujas queda así representado:

