

PRÁCTICA 5 AMD

TABLEAU, POWER BI, SUPERSET

Parga Martínez, Alberto

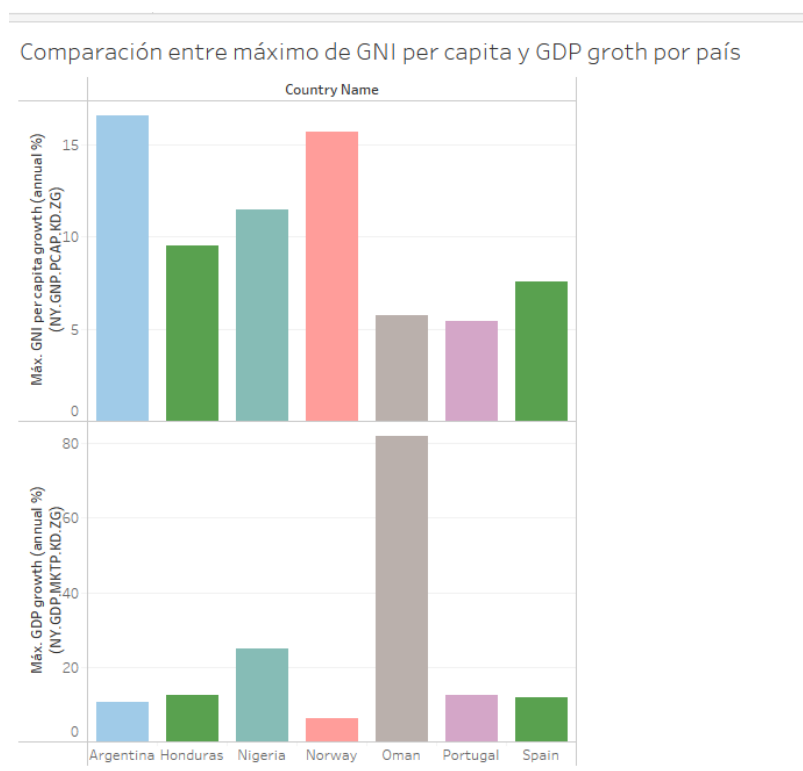
Pérez Galán, Adrián

GRÁFICAS DE DATOS DEL *World Bank Databank*

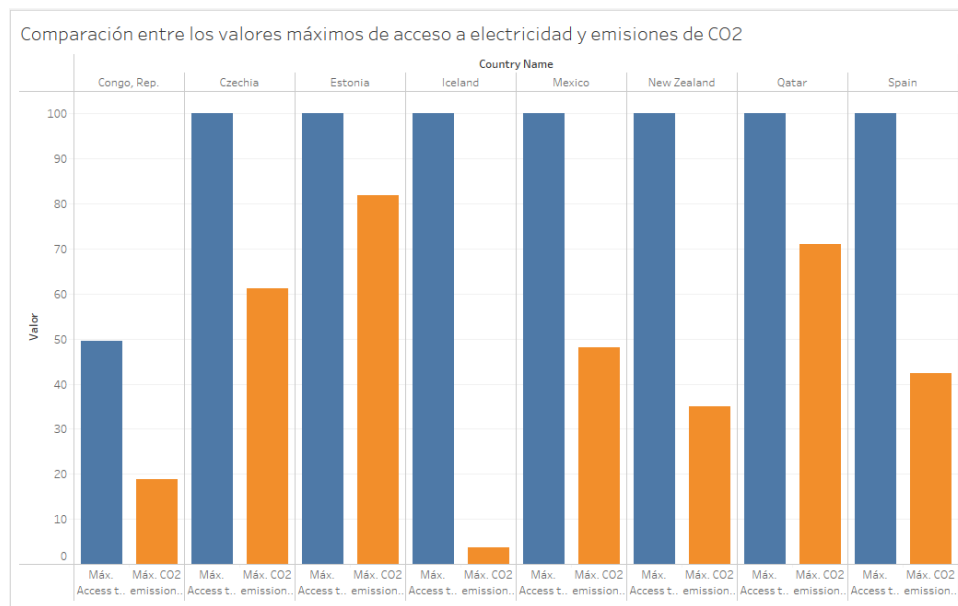
TABLEAU

Antes de crear las gráficas con esta herramienta, tuvimos que tratar los datos modificando ciertos valores que aparecían representados como '..' y que fueron sustituidos por '0'.

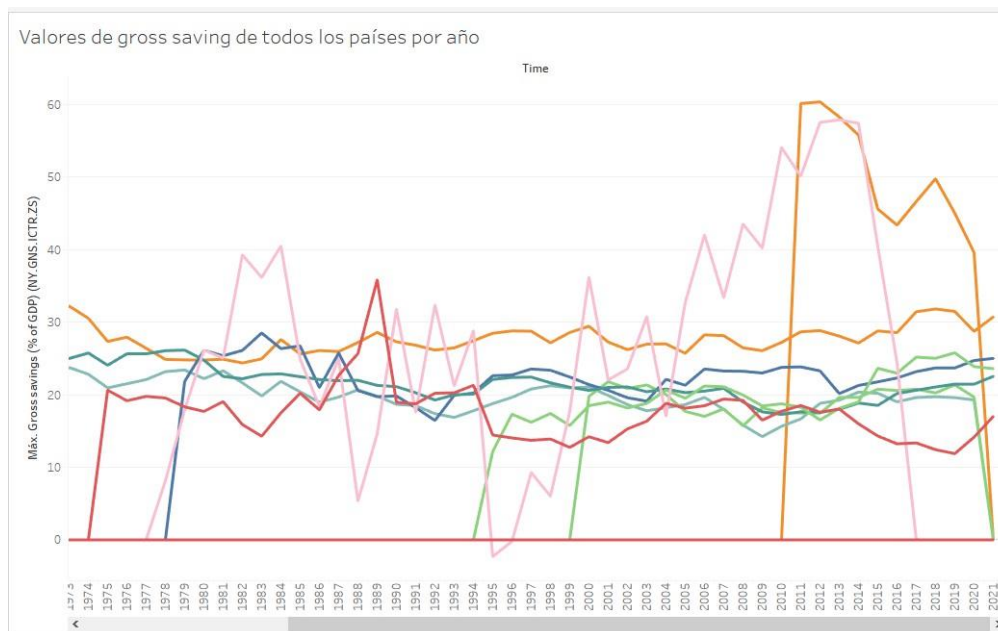
En este primer gráfico hacemos una comparación en dos gráficas de barras diferentes entre el GNI per capita y el GDP growth por país para ver si escogiendo 7 países de diferentes partes del mundo se puede encontrar una relación entre ambos términos. Viendo los resultados no se ha encontrado una relación con sentido.



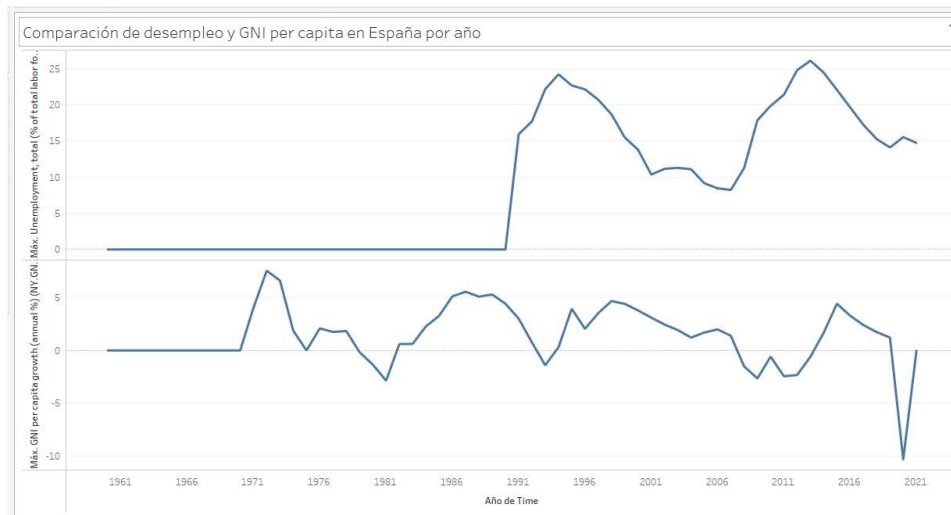
En la segunda gráfica buscamos la comparación entre el máximo de acceso a electricidad y las emisiones de CO2 de países de diferentes partes del mundo. Quisimos ver si podíamos encontrar alguna relación ya que el sector eléctrico es el mayor responsable de emisiones CO2 en casi todo el mundo



El tercer gráfico de líneas muestra por años el ahorro bruto de diferentes países.



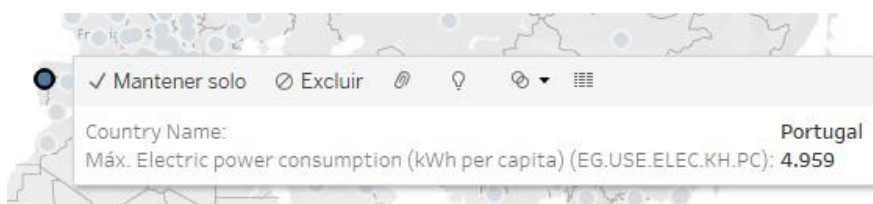
El siguiente gráfico es la comparativa de desempleo y el ingreso nacional bruto en España. La intención era ver y confirmar de que estos dos datos tienen una relación directa y es que se puede ver que en los valores máximos de desempleo es cuando menos ingreso per cápita hay en el país.



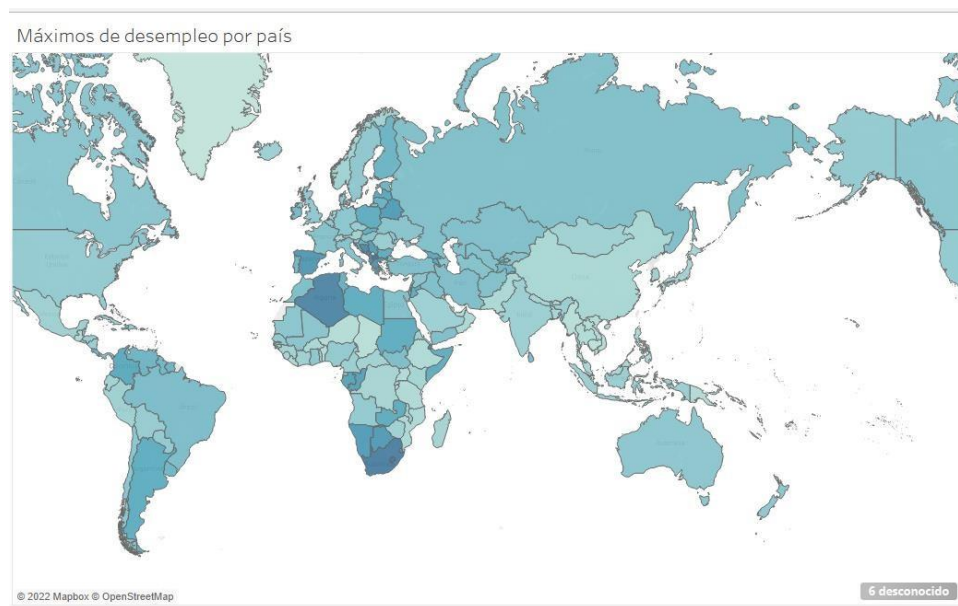
Mapa con el consumo máximo de electricidad por país:



Si pinchamos en uno de los círculos nos muestra el nombre del país y el máximo consumo de electricidad.



Con este mapa de calor cuanto más oscuro sea el color mayor será la tasa de desempleo del país.



Y por últimos una tabla comparando los valores medios de consumo eléctrico, GDP y GNI per capita.
Cuanto mayor sea el GDP mayor será el GNI per capita:

Country Name	Prom. Electric power ..	Prom. GDP (constant ..	Prom. GNI per capita (..
Afghanistan	0	4.559.026.187	9
Albania	828	5.140.460.905	1.162
Algeria	421	86.977.152.479	2.986
American Samoa	0	210.501.010	0
Andorra	0	1.701.151.214	0
Angola	71	31.074.468.656	789
Antigua and Barbuda	0	705.903.505	216
Argentina	1.240	351.653.691.538	10.132
Armenia	648	3.753.800.066	1.215
Aruba	0	1.377.801.007	439
Australia	6.275	727.718.151.850	36.314
Austria	4.785	249.314.181.329	17.544
Azerbaijan	842	15.895.522.046	86
Bahamas, The	0	7.974.832.662	17.005
Bahrain	10.068	12.375.363.245	7.040
Bangladesh	69	84.157.687.240	622
Barbados	0	3.152.578.832	255
Belarus	1.356	20.909.318.213	2.033
Belgium	5.032	300.737.208.002	16.486
Belize	0	787.784.034	2.474
Benin	34	5.545.367.315	639
Bermuda	0	5.120.357.305	13.755
Bhutan	0	634.628.529	820
Bolivia	254	16.328.674.196	1.608
Bosnia and Herzegovina	895	5.920.893.864	1.530
Botswana	625	5.777.304.750	3.002
Brazil	1.082	993.757.453.990	5.581

Power BI

En esta herramienta también aplicamos ciertas transformaciones a los datos: en particular, utilizar el editor de consultas de PowerBI para sustituir los valores “..” que usa el WDB como nulo, por nulls propiamente dichos, y cambiar el tipo de las columnas numéricas a “Número decimal”, para facilitar las agregaciones.

Además, se utilizaron dos datasets distintos: el primero se utilizó para el grueso de los ejercicios, y es equivalente al CSV utilizado en el resto de las herramientas; el segundo es un dataset específico de la vista en Python, que contiene datos sobre los porcentajes de producción eléctrica que suponen las diferentes industrias energéticas para los diferentes continentes disponibles en el WDB.

La primera vista realizada es una tabla que contiene, en el rango 2007-2014, datos sobre todos los países. Estos datos son promedios para el crecimiento del PIB, de la RNB per cápita, del ahorro bruto, del IPC, del desempleo, y del acceso a la energía eléctrica. Vemos que en ciertos países faltan datos, incluso eligiendo un rango de años razonable.

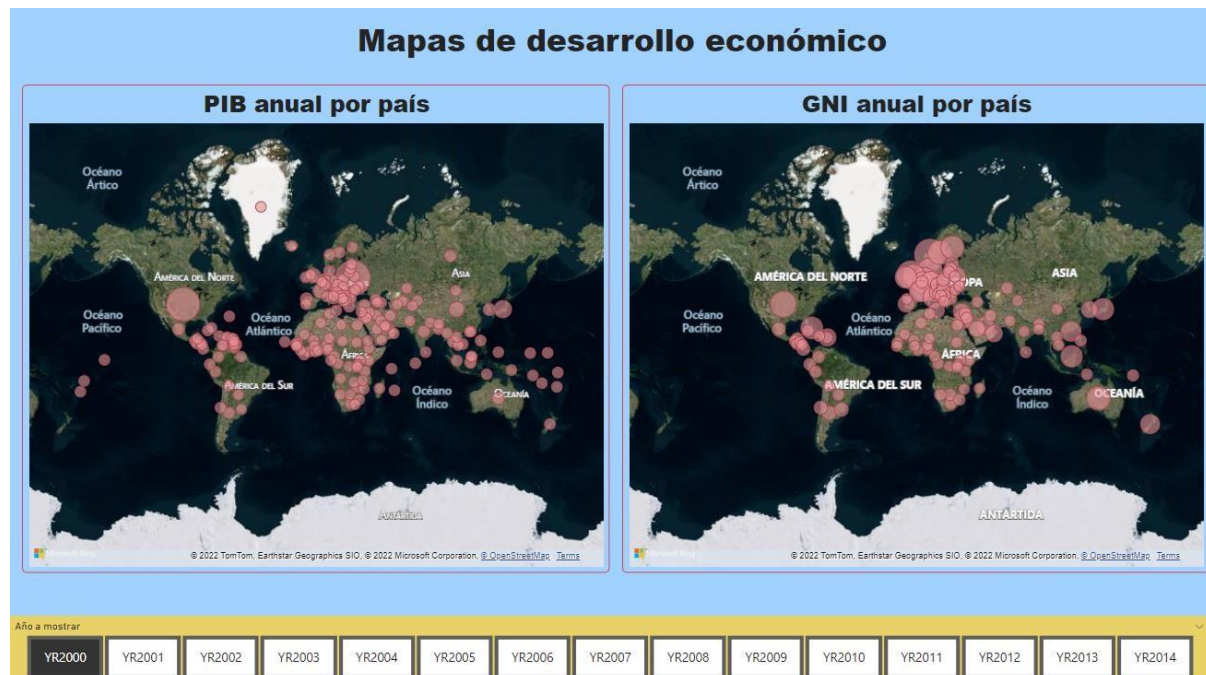
Esta tabla no tiene un objetivo analítico concreto, sino más bien servir de vista global para estos parámetros (podemos, por ejemplo, comprobar que el IPC promedio es superior a 100, por lo que en general se ha dado inflación a nivel mundial), y también servir de referencia: es sencillo buscar el país que queremos y obtener todos sus estadísticos, o ordenar la tabla por uno de ellos si queremos hacer un “ranking”.

Promedios de factores para todos los países en el rango [2007-2014]						
Country Name	Crecimiento del PIB promedio	Crecimiento de RNB per cápita promedio	Ahorro bruto promedio	IPC promedio	Desempleo promedio	Acceso a la electricidad (%) promedio
Afghanistan	9.38			109.79	11.22	54.75
Albania	3.41	3.61	19.10	101.01	14.70	99.98
Algeria	2.99	1.31	49.57	103.86	10.77	98.95
American Samoa	-1.22					
Andorra	-2.16					100.00
Angola	6.59	3.99	35.21	106.66	6.28	35.84
Antigua and Barbuda	-0.74		10.60	101.76		98.95
Argentina	2.77	1.87	17.96		7.68	98.63
Armenia	3.44	3.65	17.13	102.38	16.38	99.47
Aruba	-0.09		7.00	100.58		99.17
Australia	2.87	1.98	23.08	101.33	5.18	100.00
Austria	0.94	0.25	26.36	102.18	4.94	100.00
Azerbaijan	7.62	8.69	44.37	100.83	5.51	99.86
Bahamas, The	-0.10	-1.72	23.19	101.24	12.89	100.00
Bahrain	4.61	0.16	34.85	100.23	1.15	100.00
Bangladesh	6.09	4.98	38.45	106.73	4.22	57.96
Barbados	-0.92		6.57	103.60	10.34	100.00
Belarus	4.57	4.41	29.05	171.67	6.41	100.00
Belgium	1.18	0.28	25.45	101.89	7.78	100.00
Belize	2.11	-0.04	9.91		8.30	89.14
Benin	4.58	1.68	13.76	102.07	1.68	33.17
Bermuda	-2.46		38.91			100.00
Bhutan	7.86	5.73	33.85	108.03	3.18	79.54
Bolivia	5.10	3.17	25.26	106.17	2.45	87.22
Bosnia and Herzegovina	1.60	2.33	9.46	101.00	26.80	99.86
Botswana	5.18	3.99	38.55	104.30	18.08	51.13
Brazil	3.50	2.58	17.86	104.65	7.40	99.02
Total	3.33	2.10	23.71	104.34	8.12	80.60

La siguiente visualización son un par de mapas. En ellos, colocamos el PIB del país, y su RNB (renta nacional bruta) per cápita. El objetivo de esta visualización es observar donde se concentran los valores más altos de estos estadísticos, y si existe alguna correlación entre ellos. Además, se añadió la posibilidad de seleccionar el año para el que se quiere visualizar (entre 2000 y 2014, para tener buena variedad de datos, ya que si se tomasen años muy tempranos los países menos desarrollados no habría tomado medidas).

Podemos ver que existe cierta relación entre los valores: los RNB más altos parecen concentrarse en los estados del centro y norte de Europa, que también tienen PIB moderadamente altos (algo

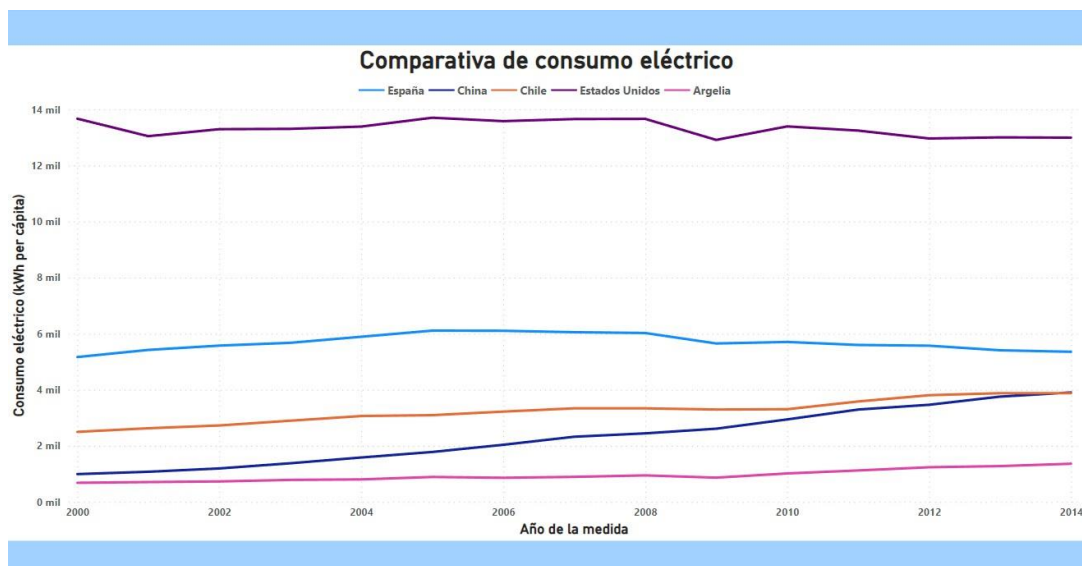
esperable en una zona que suele considerarse como altamente desarrollada); mientras que los países con bajo PIB por lo general comparten un RNB per cápita bajo. Existen ciertas excepciones: por ejemplo Estados Unidos, que siendo el mayor PIB no presenta una RNB muy destacable.



La siguiente gráfica es una comparativa del consumo eléctrico, medido en kWh per cápita, a lo largo del período 2000-2014. Para no llenar la gráfica con cientos de líneas y hacerla incomprensible, se seleccionaron 5 países que se consideran representativos:

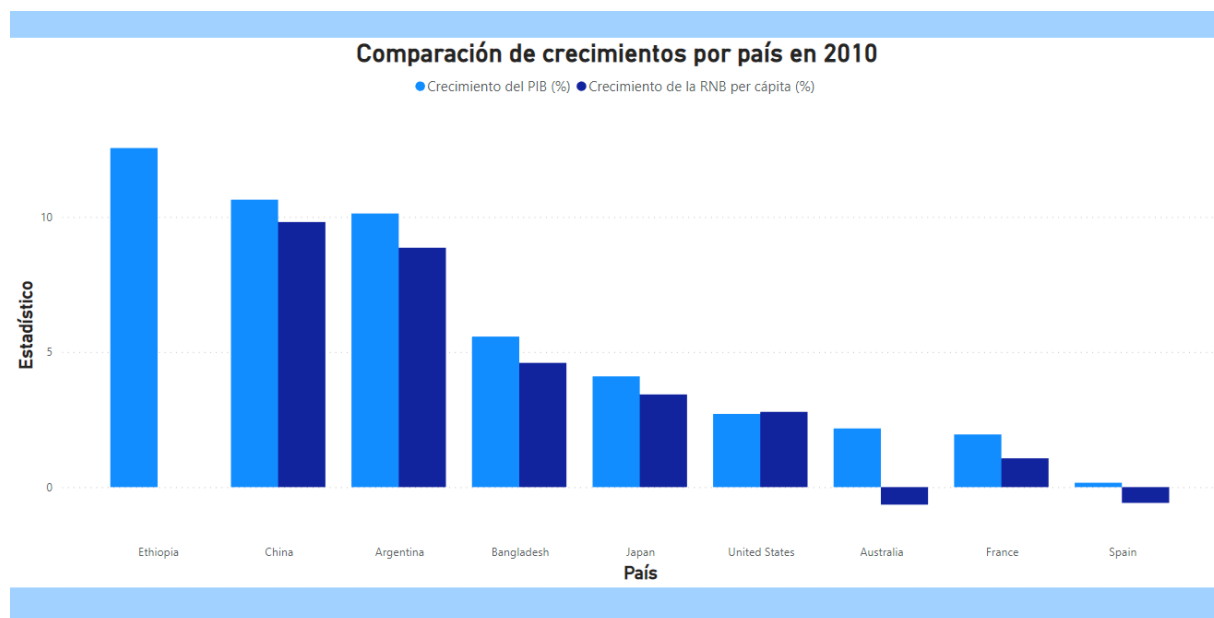
- España: por proximidad personal, y como país desarrollado, pero no especialmente puntero.
- Estados Unidos: como país muy desarrollado y puntero en consumo e industria.
- China: como país que se ha estado desarrollando de forma explosiva en los últimos años.
- Chile: como país no privilegiado, pero tampoco totalmente pobre.
- Argelia: como país en vías de desarrollo.

Podemos ver que la gráfica más o menos ofrece resultados acordes a lo esperable. Estados Unidos presenta un consumo enorme, muy por encima de los demás, propio de un país muy industrializado y cuyos habitantes disponen de muchas comodidades tecnológicas. España le sigue muy por detrás, como se esperaría de un país ya desarrollado pero que no se encuentre en el “top”; mientras que detrás de ella Chile y China están casi empatados, y Argelia es el último con gran diferencia. De estos últimos datos, lo más interesante es ver la evolución relativa de China y Chile: ambos parecen encontrarse en desarrollo, pero el país asiático crece a más o menos el doble de ritmo, lo que encaja en nuestra idea preliminar.



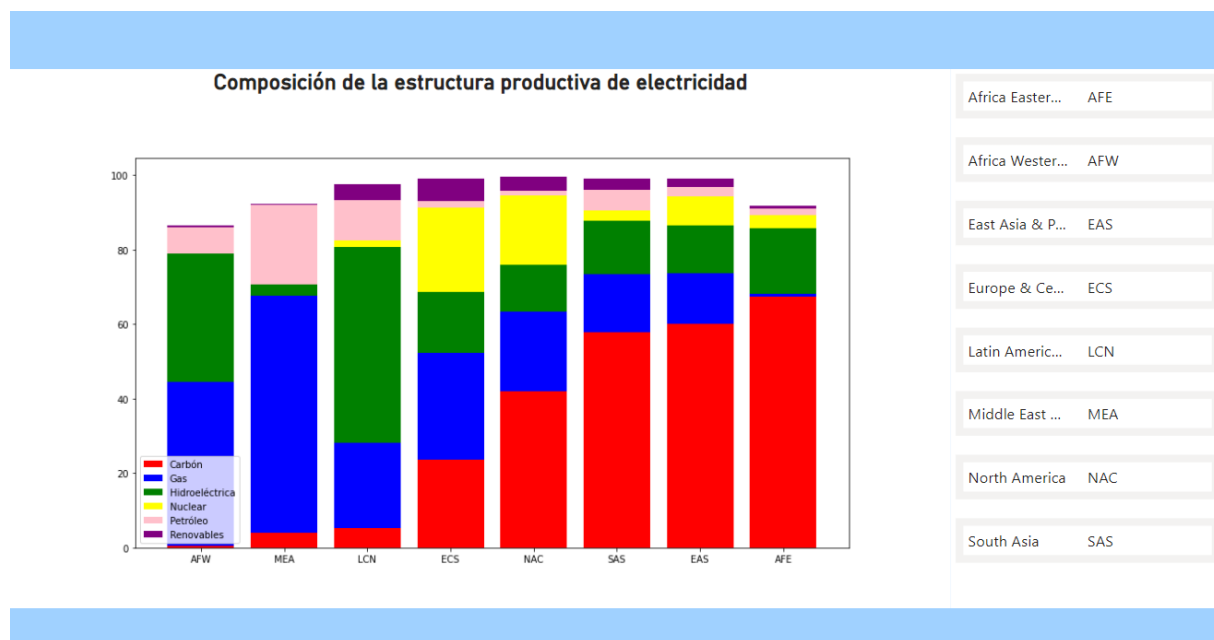
El siguiente gráfico muestra una comparación entre el crecimiento porcentual del PIB y de la RNB per cápita en diversos países, tomando datos del año 2010. En general, estos datos deberían estar relacionados, dado que una mayor producción dentro del país (PIB) debería repercutirse en unos mayores ingresos para sus habitantes (RNB per cápita).

Como vemos, esto es así excepto en dos países, Australia y España. Esto podría sugerir que en 2010 estos dos países perdieron ayudas económicas externas que venían recibiendo anteriormente, lo que se reflejó en una menor ganancia para sus habitantes incluso si lo producido dentro de las fronteras del país aumentó.



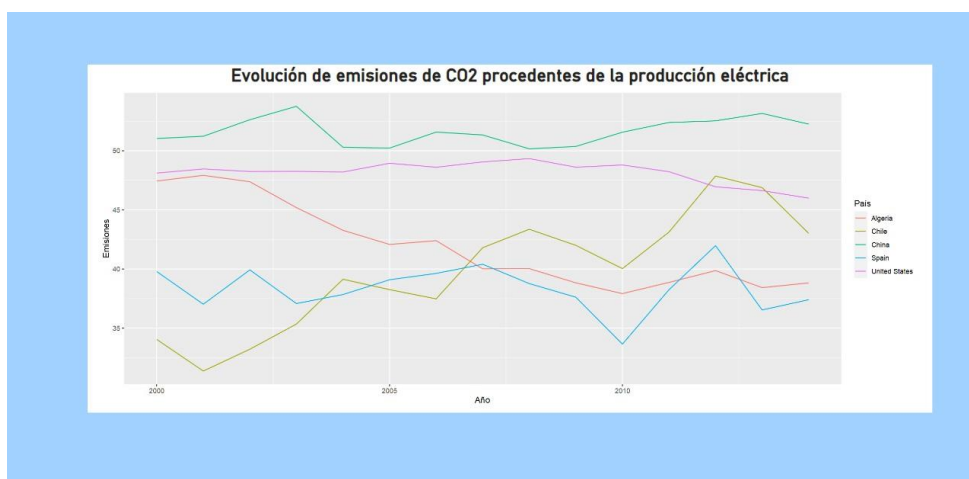
El siguiente gráfico fue hecho a través de Python, y refleja la estructura de producción de la energía eléctrica en diversas agregaciones de países. Dado que el dataset original no incluía estas agregaciones, como se comentó en la introducción de esta sección se importó un segundo dataset. Todos los datos son del año 2010.

Es difícil sacar datos relevantes de este análisis, pero podemos observar algunos detalles, como por ejemplo que las zonas que más dependen de la energía nuclear son las zonas tradicionalmente percibidas como ricas: Norteamérica y Europa. Esto tiene sentido, ya que la energía nuclear es costosa y requiere de importantes avances técnicos. También podemos ver que las renovables escasean en las zonas tradicionalmente pobres, como ambas agrupaciones africanas y Oriente Medio; lo que es razonable, ya que estas zonas preferirán fuentes de energía más eficientes, aunque contaminen.

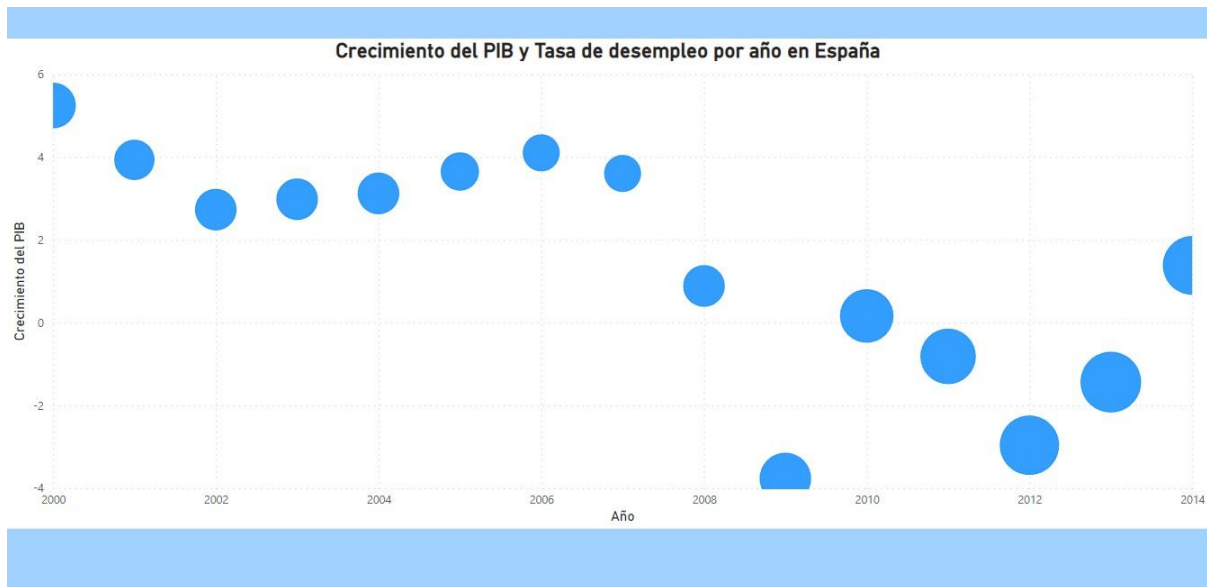


El siguiente gráfico compara, en el período usual (2000-2014), las emisiones de CO₂ achacables a producción de electricidad y calefacción (medidas en % del total de consumo de combustibles fósiles). Se usan los mismos países que se utilizaron para la gráfica de consumo eléctrico, por motivos similares (representan diversos perfiles de nación). Se produjo utilizando R.

Es difícil extraer conclusiones de esta visualización, principalmente porque el propio estadístico tiene múltiples interpretaciones posibles. Podemos analizar tendencias: China y Estados Unidos se mantienen en un dato alto, por lo que podríamos considerar que esto supone alta industrialización en el país; mientras que Algeria ha bajado en picado mientras Chile subía prácticamente al mismo ritmo. España está en un nivel estable pero más bien bajo, lo que podría apoyar la idea de la industrialización (no siendo España un país que destaque en esta área).



La última vista muestra el crecimiento del PIB en España por año, y mediante el tamaño de la burbuja indica la tasa de desempleo. Este gráfico buscaba mostrar relación entre el número de parados y el crecimiento de la producción del país, y parece que así es: los años con mayor tasa de desempleo (mayor burbuja) tienen tasas de crecimiento del PIB más bajas, o incluso negativas, salvo un outlier en 2014. Esto tiene sentido: cuantos menos trabajadores hayan en el país, más difícil será aumentar la producción interna del mismo.



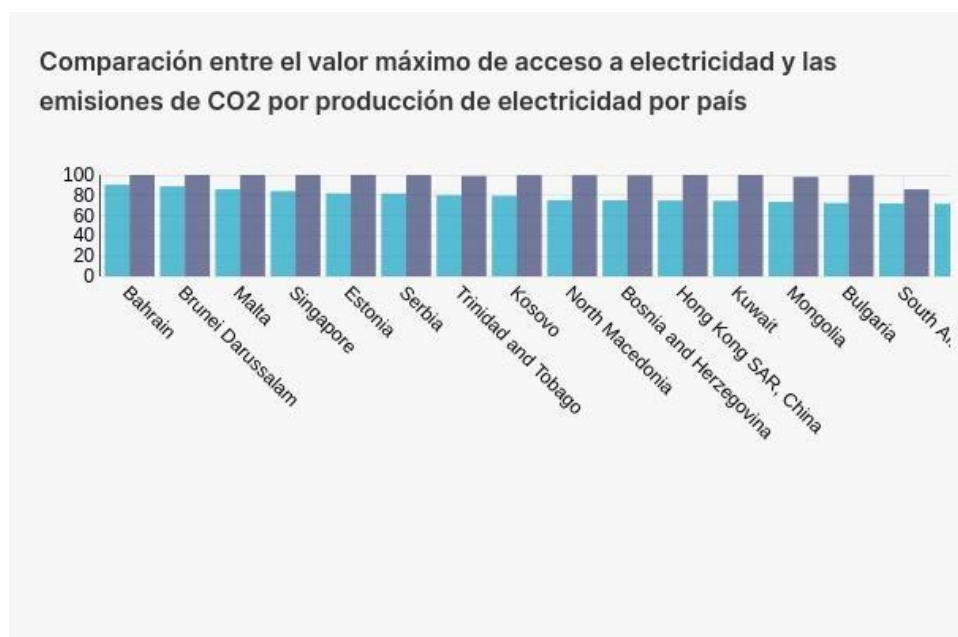
Apache Superset

Antes de crear las gráficas con esta herramienta, tuvimos que tratar los datos modificando ciertos valores que aparecían representados como '..' y que fueron sustituidos por '0'.

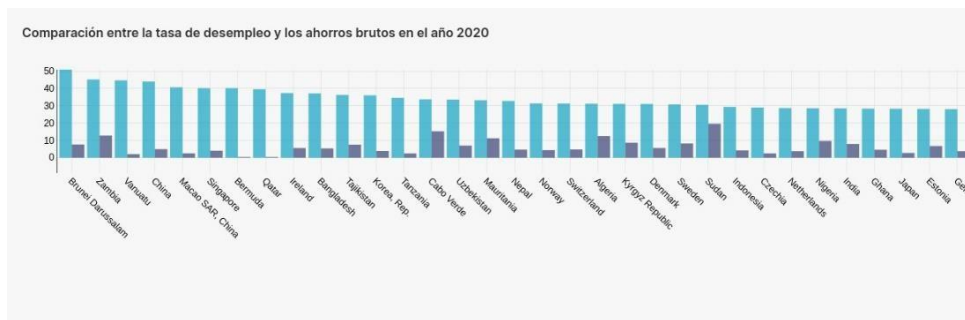
El primer gráfico creado fue una tabla en la que se puede observar la relación entre los valores medios de consumo de energía, GDP y GNI per cápita de cada país. Creamos esta tabla porque nos parecía interesante comprobar que impacto tiene el consumo eléctrico de un país en su economía.

Comparación de los valores medios de consumo eléctrico, GDP y GNI per cápita			
Country Name	AVG(Electric_power_consumption_kWh)	AVG(GDP)	AVG(GNI_per_capita)
Last Updated: 09/16/2022	N/A	N/A	N/A
Data from database: World Development Indicators	N/A	N/A	N/A
N/A	N/A	N/A	N/A
Norway	17.5k	2338	45.4k
Iceland	17k	6.61B	806.6
Canada	11.8k	548B	27.9k
Luxembourg	11.1k	311B	48.4k
Sweden	10.6k	315B	33.1k
Bahrain	10.1k	12.4B	704k
North America	9.41k	11.7T	34.5k
Finland	9.16k	154B	18.3k

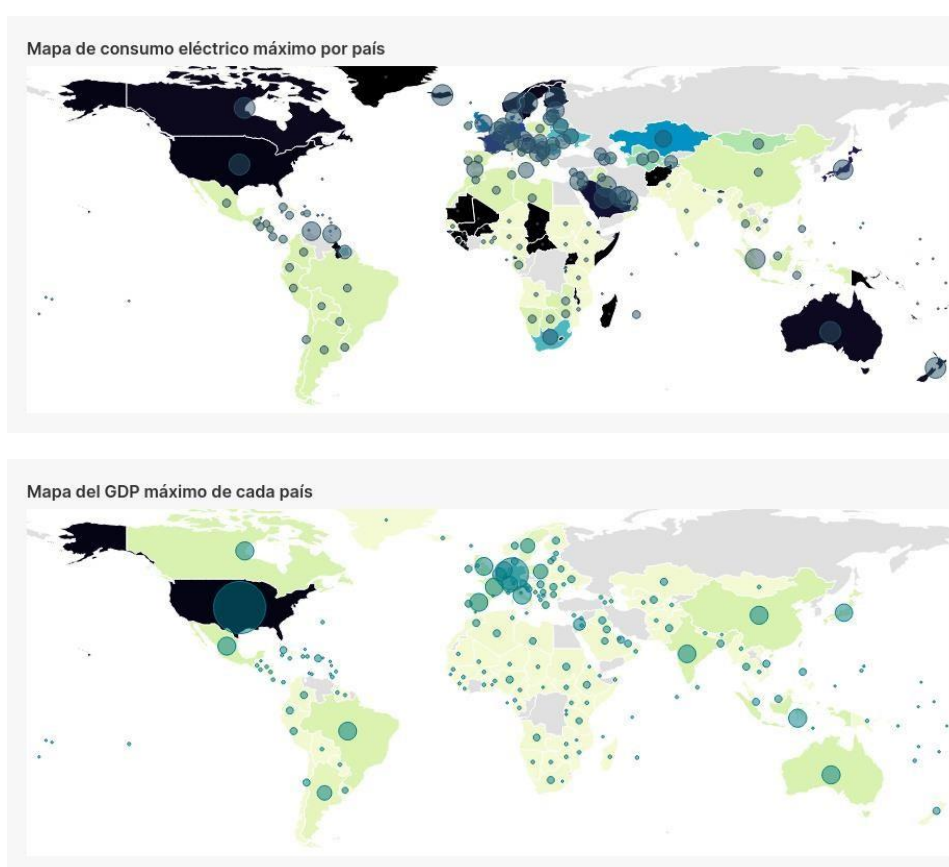
El segundo gráfico es un diagrama de barras que compara los valores máximos de % de acceso a electricidad y emisiones de CO2 por producción de electricidad por país. Gracias a este gráfico se puede aproximar que países producen su propia electricidad (mediante medios que emiten CO2) y que países usan fuentes renovables o compran la electricidad a otro productor.



El tercer gráfico es otro diagrama de barras que compara los valores de desempleo y ahorros brutos de los países en el año 2020. Creamos este diagrama para comprender el impacto que tiene el nivel de desempleo en los ahorros de un país.



El cuarto y quinto diagrama son mapas del mundo que muestran por un lado el consumo eléctrico máximo por país y por otro el valor máximo de GDP de cada país. Esta comparativa nos ofrece otra forma de comprobar el impacto del consumo eléctrico en la economía de cada país.



El último gráfico es un diagrama de líneas que muestra la variación del crecimiento del GDP y del GNI per cápita de España desde 1960. En este gráfico se puede observar claramente el impacto de la crisis económica del 2008 o la pandemia de la COVID-19 en nuestro país.

Crecimiento del GDP y del GNI per cápita de España desde 1960

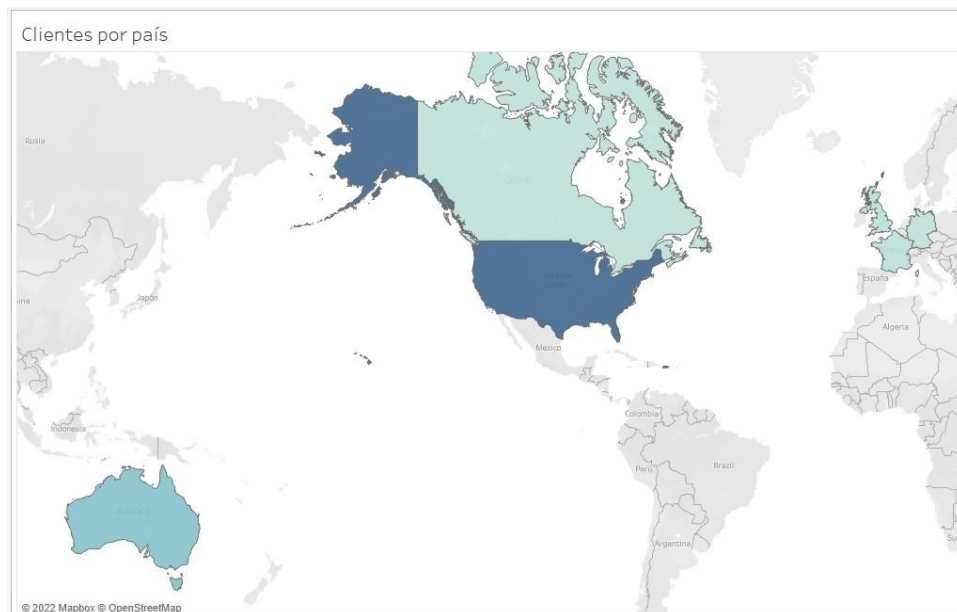


GRÁFICAS DE DATOS DE *Adventure Works*

TABLEAU

Cientes por país

countryregi..	
AU	3.665
CA	1.791
DE	1.852
FR	1.884
GB	1.991
US	8.637



Ventas por divisa

currency	
AUD	6.843
CAD	4.067
EUR	5.295
GBP	3.219
USD	12.041

POWER BI

Nota: tanto para PowerBI como para Tableau, la base de datos AdventureWorks utilizada es ligeramente distinta a la empleada en el ejercicio de Superset. Esto se debe a que los equipos Windows de los miembros que realizaron estas partes no eran compatibles con el fichero sql de backup utilizado para restaurar AdventureWorks en Linux. En estos equipos se usó un fichero de instalación basado en CSVs (el presente en la Práctica 1 de la asignatura), cuya BD es algo diferente. Por tanto, aunque el espíritu del almacén de datos es el mismo, los nombres de las relaciones y los valores pueden cambiar ligeramente.

Para llevar a cabo este trabajo en la herramienta PowerBI, fue necesario un tratamiento de datos específico de cada ejercicio.

Para los ejercicios sobre clientes/región se importaron los datos de PostgreSQL directamente, trayéndonos las tablas sales.customer y sales.salesterritory. PowerBI es lo suficientemente avanzado como para traerse las relaciones del modelo al área de trabajo, por lo que elaborar el mapa/la tabla pedida es sencillo: si usamos como ubicación/columna el territorio de la venta, usando el campo countryregionname, y como tamaño de burbuja/columna un recuento distintivo de los IDs de cliente, PowerBI asocia las tablas a través del ID de territorio y realiza el cálculo correctamente:



Como último detalle del ejercicio, en la importación de datos desde Postgres se tuvo que modificar el countryregionname asociado a Alemania para que fuese DEU y no DE, ya que de otra manera PowerBI no lo reconoce como un código de país válido, y no coloca sus datos en el mapa.

Para la siguiente vista, ventas por divisa, no se importó directamente la BD desde Postgres, sino que se efectuó una consulta SQL específica para facilitar el diseño del objeto visual:

```

SELECT soi, idt, countryregioncode, currency
FROM
    (select sa.soh.territoryid as idt, sa.soh.salesorderid as soi
    from sa.soh inner join sa.st on sa.soh.territoryid=sa.st.territoryid) as sq1
    inner join
    (select sa.st.countryregioncode, sa.st.territoryid as idt, sa.crc.currencycode as currency
    from sa.st inner join sa.crc on sa.st.countryregioncode=sa.crc.countryregioncode) as sq2
    using (idt)

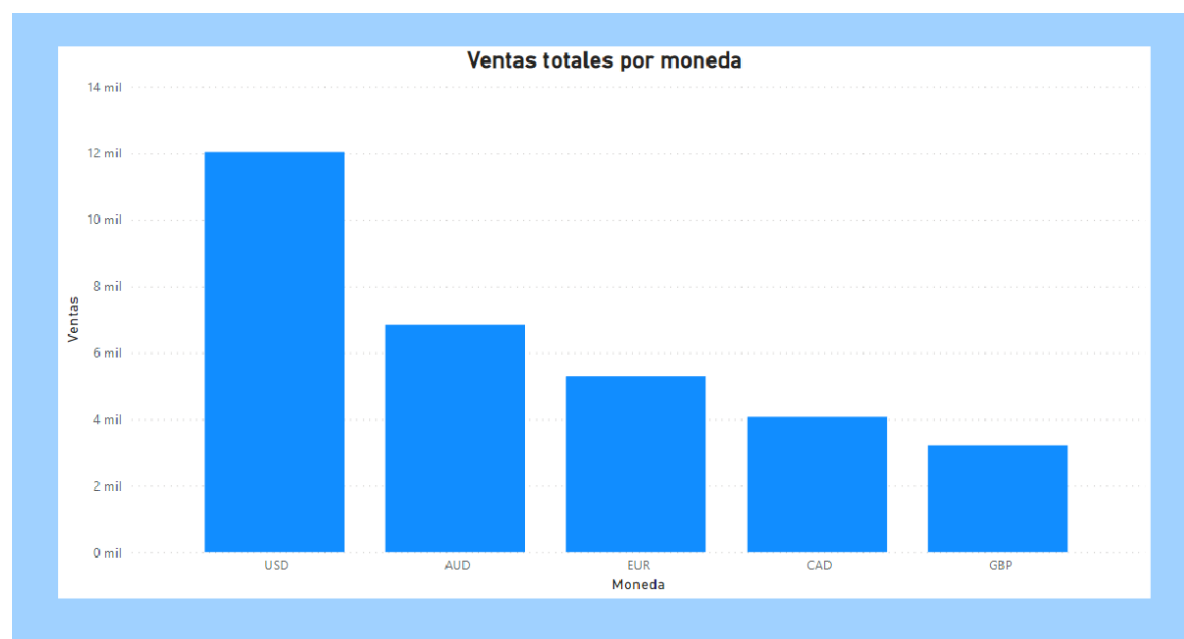
ORDER BY soi

```

La consulta es compleja si no se conocen los schemas de la BD: en esencia, vamos a extraer el conjunto de los Sales Order IDs, con su ID de territorio asociada, el código de país de ese territorio, y las monedas que maneja ese país según la BD. Por tanto, tenemos para cada venta su ubicación y la divisa que se usa en la misma. Una vez importamos los datos a PowerBI, debemos excluir un par de monedas que generarían resultados duplicados: estas son la FRF y la DEM (el franco francés, y el marco alemán). Esto no nos hace perder ventas: por como esta construida la consulta, para esos Sales Order IDs teníamos dos filas, una con FRF/DEM, y otra con EUR (porque esos países en teoría manejan ambas divisas).

Consideramos que la exclusión directa de estas monedas no es preocupante dado que, si ordenamos las ventas por fecha, todas son posteriores a 2011, cuando el franco y el marco llevaban años sin usarse.

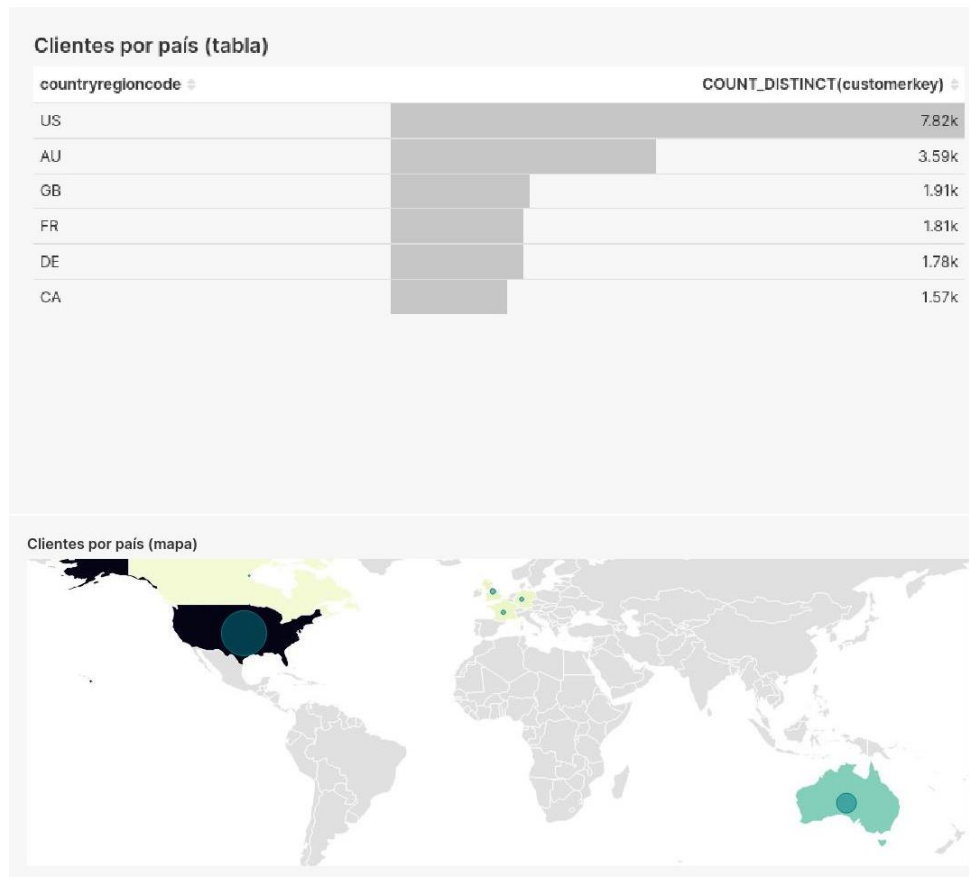
Una vez hemos filtrado así los datos, ya sólo tenemos que elegir una vista y colocar la currency, y un recuento distintivo de los Sales Order IDs:



Los datos obtenidos en ambos gráficos son consistentes: dado que la mayoría de las clientes son estadounidenses, tiene sentido que la mayoría de ventas se producirán en dólares americanos. Lo siguen los dólares australianos, que serían la segunda área con más clientes, y así sucesivamente.

APACHE SUPERSET

Para obtener los datos de clientes decidimos agruparlos por países:



Para los datos de ventas por divisas creamos dos diagramas distintos, uno para las ventas a través de internet y otro para las ventas de reseller.





Como gráfico adicional decidimos mostrar el porcentaje de empleados que están asignados a cada departamento.

