

Open Data Visualization and Early Analysis

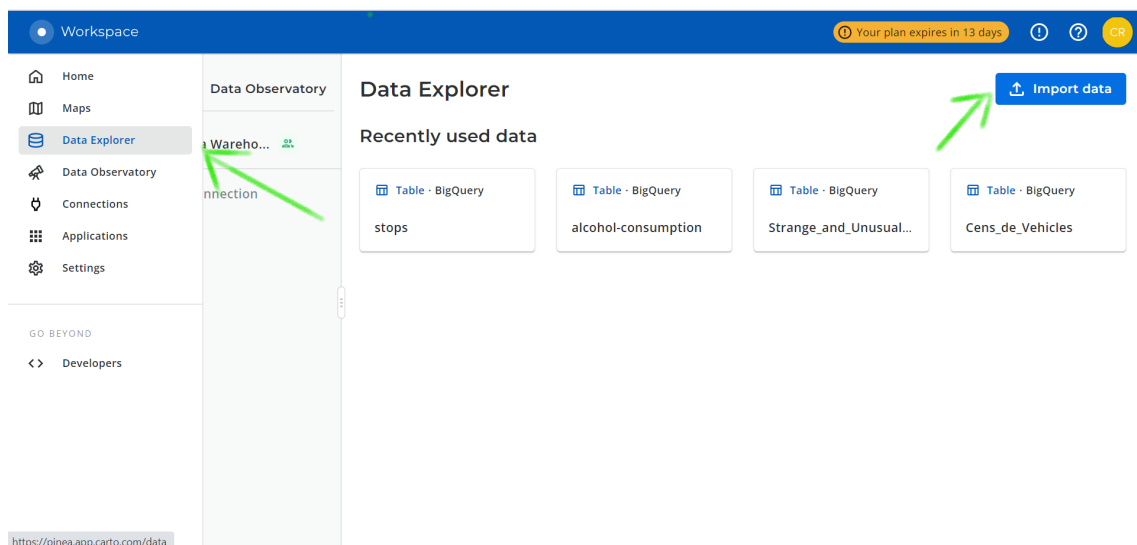
Assignment (CARTO)

Cristian Torres Ortega

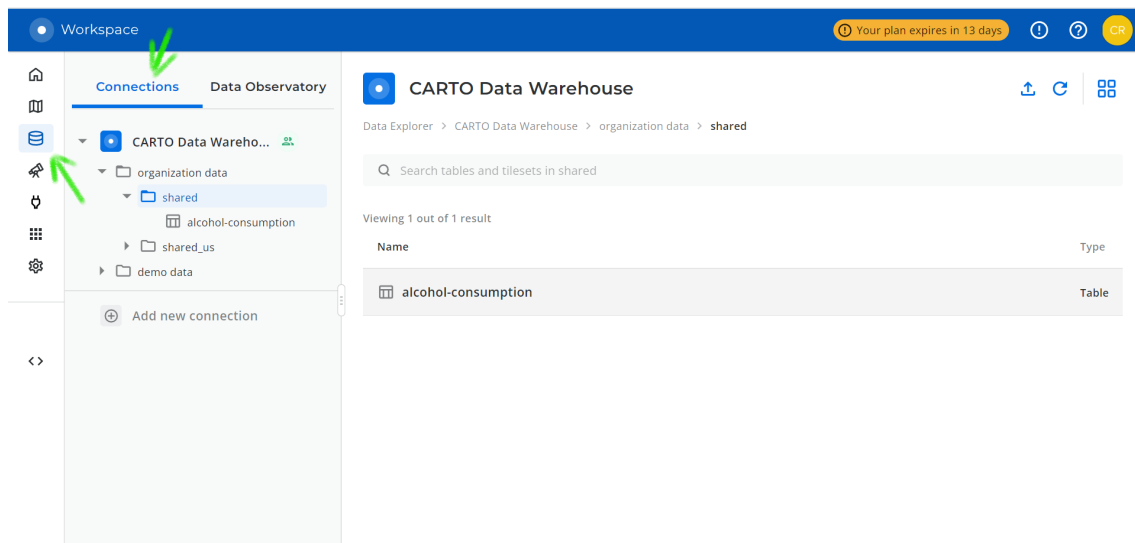
En este ejercicio se van a tratar 3 datasets extraídos del portal de Data Science y Machine Learning “Kaggle”. Por un lado, se tratará un dataset que recompila datos sobre el consumo de alcohol per cápita en 189 países. Otro dataset tratará sobre las 300 mejores universidades del mundo. Y finalmente, se estudiará un último dataset de los mayores artistas musicales superventas.

DATASET: Consumo de alcohol per cápita en 2016

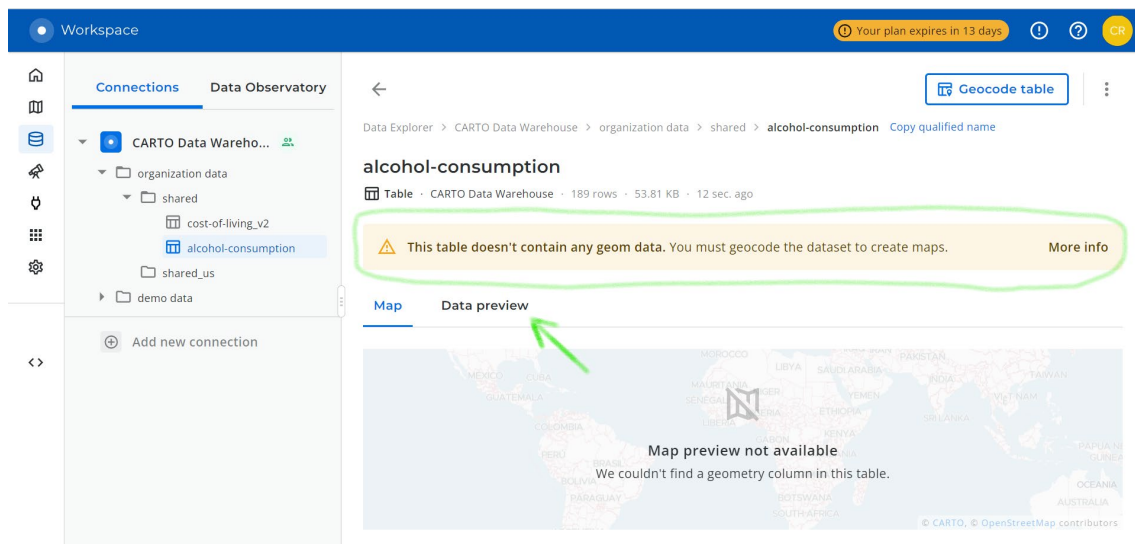
Una vez se tiene descargado el dataset se procede a importarlo en Carto para poder crear el mapa. Para ello, en la página de inicio de Carto, clicamos en “Data Explorer” y luego en “Import Data”.



Una vez importados los datos, en la ventana de data explorer, en la pestaña de conexión, se busca la carpeta en la que hemos guardado el dataset.

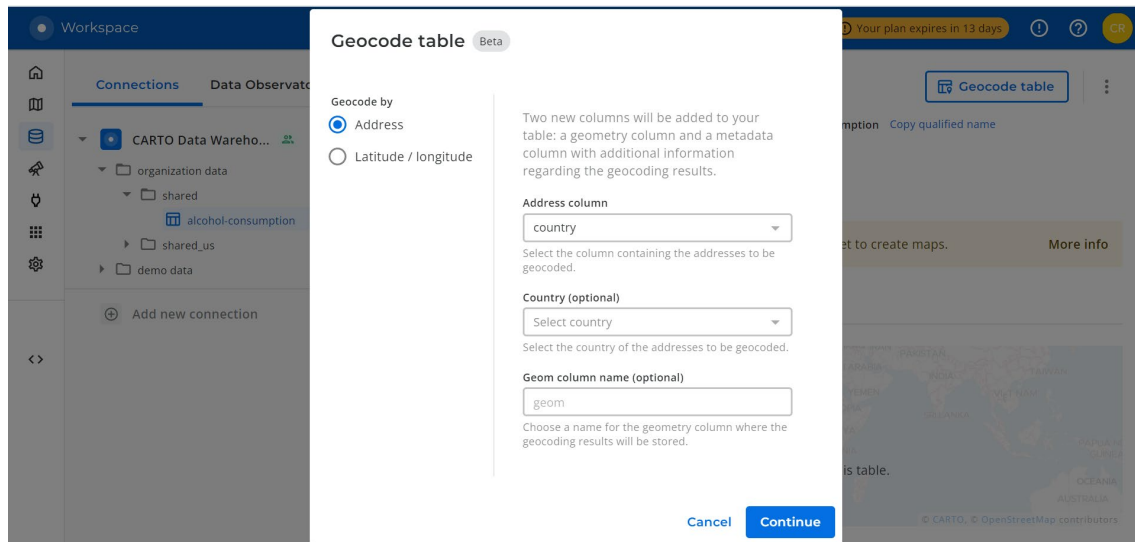


A continuación, clicamos en el dataset “alcohol-consumption”, donde se observa el siguiente mensaje de advertencia:

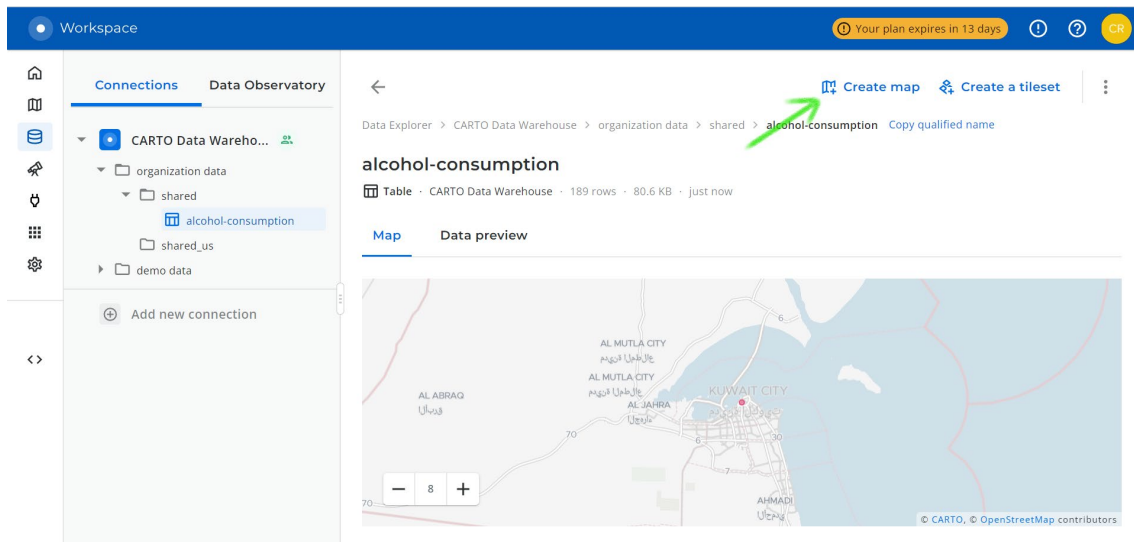


Los datos del dataset pueden verse clicando en “Data preview”. Este dataset contiene una columna dedicada a la localización de los registros, el problema es que los datos que almacena dicha columna es de tipo “String”, donde se almacena los nombres de cada país, pero no tiene ninguna columna que haga referencia a la geolocalización de los mismos para que Carto pueda montar el mapa. Pero esto no es problema ya que haciendo uso del algoritmo de Carto, es capaz de leer las celdas de cada país y crear

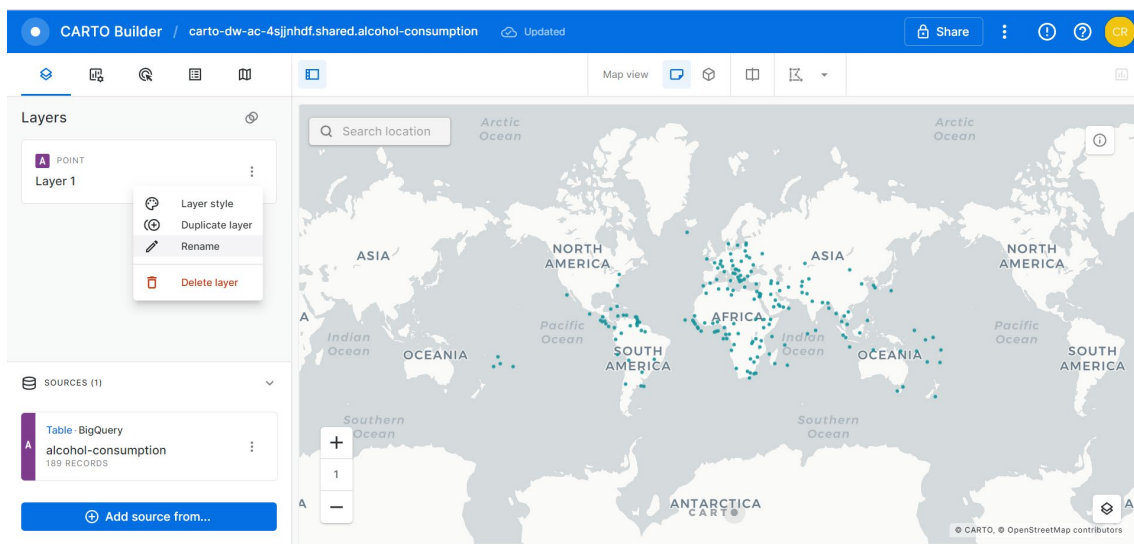
otra columna con información sobre la localización de estos países que si puede ser interpretada por Carto para realizar la creación del mapa. Para ello, se clic en “Geocode table” que se encuentra en la esquina superior derecha y se abrirá la siguiente ventana:



Se selecciona “Address” ya que el dataset trae el nombre del país, si el dataset contuviera dos columnas, una para la latitud y otra para la longitud, se elegiría la otra opción. Luego, en “Address column” se debe de elegir la columna del dataset que contiene los nombre de los países. El campo “Country” es opcional cumplimentarlo, se debe de hacer en el caso de que el dataset contenga información de ciudades o calles, entonces se debería de elegir el país al que pertenece esa información para ayudar al algoritmo. Para finalizar la geocodificación, se puede otorgar si se desea un nombre a la columna que va a ser creada automáticamente, este nombre se introduciría en el apartado “Geom column name”. Por defecto dicha columna aparecerá con el nombre “geom”. Finalmente se clic en “continue” y luego en “geocode”. Se desplegará la siguiente pantalla:



A continuación, se clic en crear mapa. Se ha elegido este dataset con la finalidad de observar en que países se consume más cerveza, vino y licores. A priori, se espera que sea Alemania el país que mas cerveza consuma, debido a su fama. Igual pasa con Francia con el vino o Rusia con los licores. Para realizar esta observación, se adaptará el mapa con puntos calientes, para que sea fácilmente identificable los países que más consumen alcohol en los diferentes casos. Para ello:

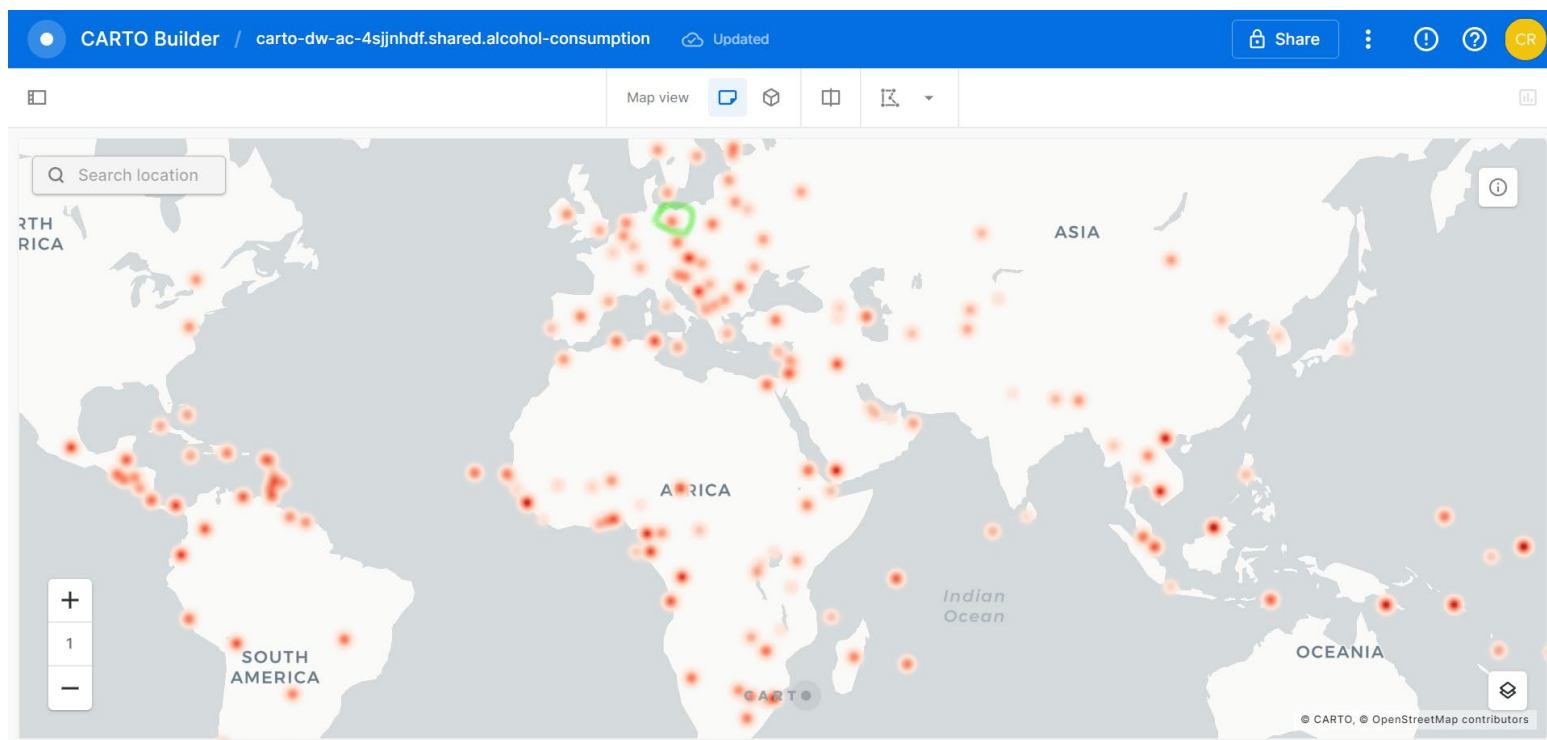


Se renombra la capa con el nombre de “Puntos calientes” y se clicla en ella. Una vez dentro de esta capa se añade la siguiente configuración:

- En “Lower type” se elige “Heatmap”.
- En “Color”, se elige en “type” la opción “Singlehue” con la escala de rojos que más se adapte.
- En “Radius” se selecciona el valor de “20”.
- En “Weight base don” se selecciona las columnas en las que se desea que los puntos se basen, en este caso de estudio serían las siguientes columnas “Beer_percentage”, “Wine_percentage” y “Spirits_percentage”.

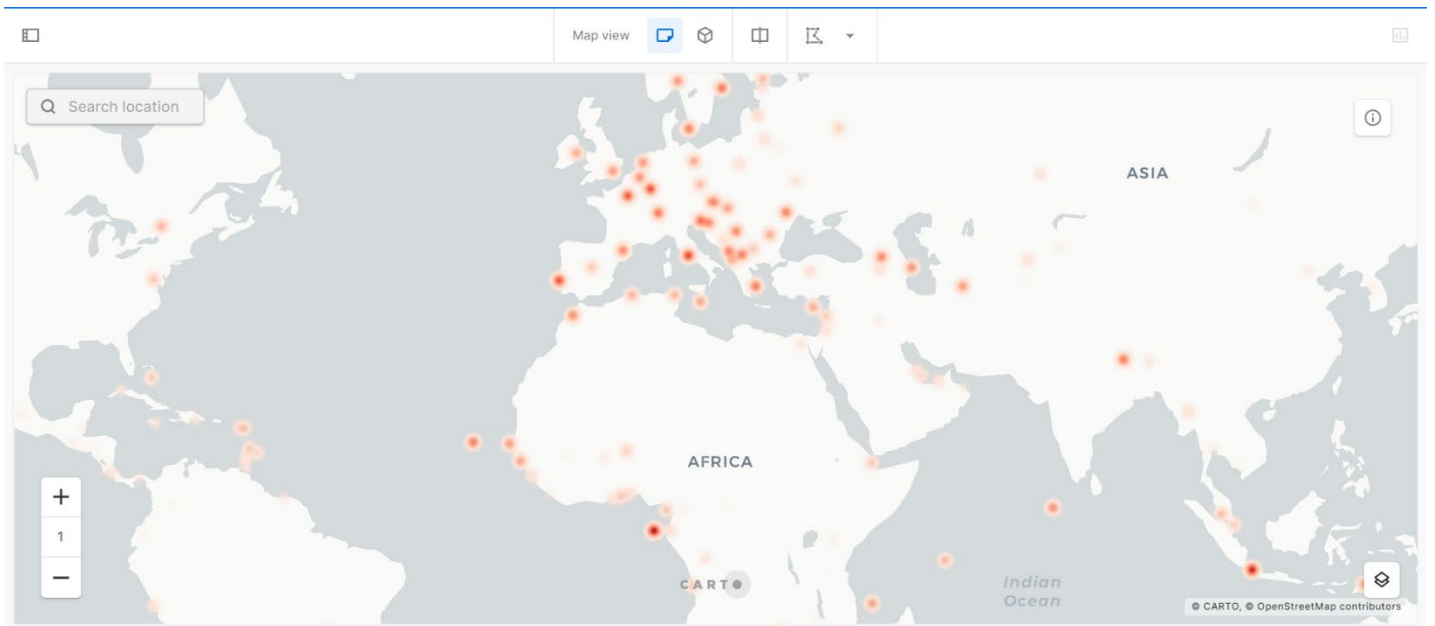
Los resultados obtenidos serían:

Beer percentage



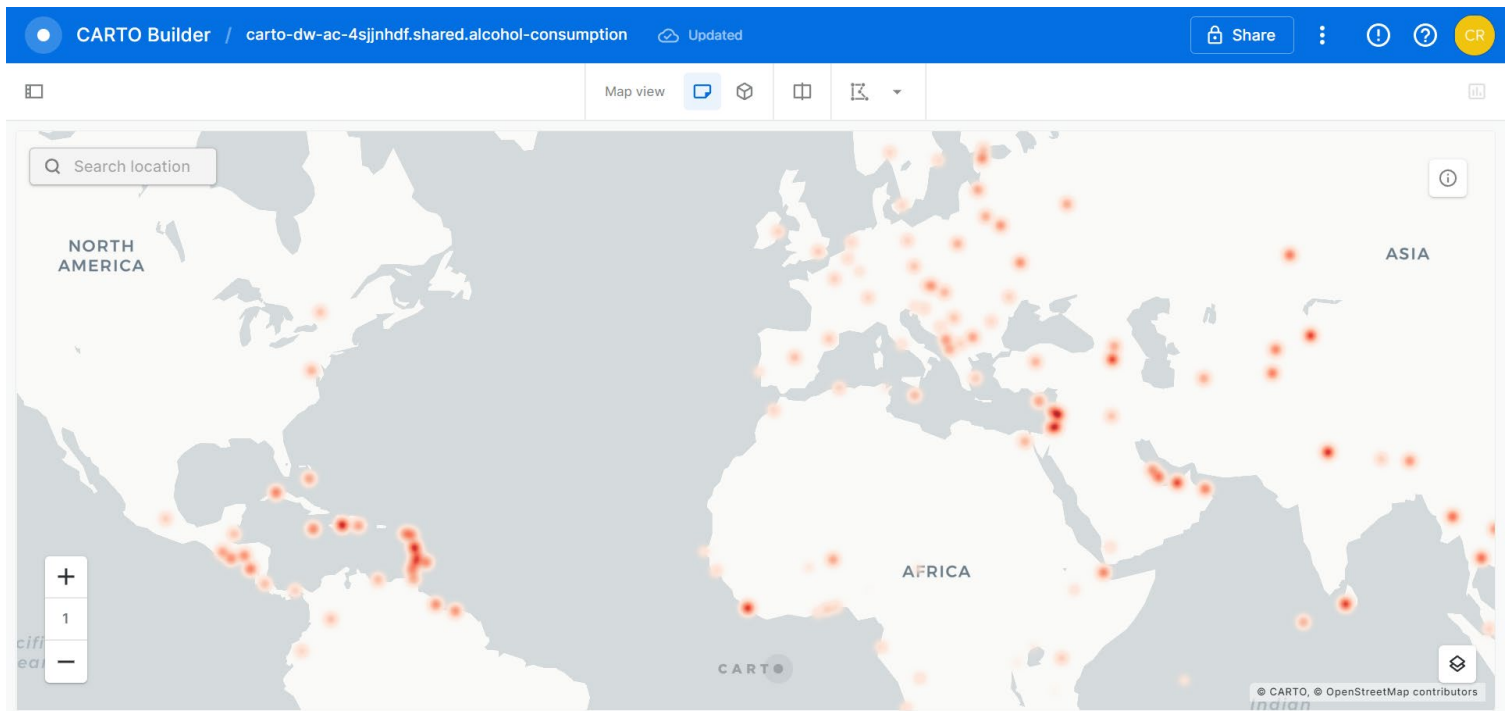
El punto rodeado de color verde es Alemania, como se observa y para sorpresa, no se trata del país donde más cerveza se consume, ya que tanto al norte de Sudamérica, costa oeste de África y el sur del continente asiático, se observan países más candentes que Alemania.

Wine_percentage



Aquí se observa como el consumo de vino es mayor en Europa que en los demás continente, con Portugal, Francia, Italia y Luxemburgo liderando.

Spirits_percentage

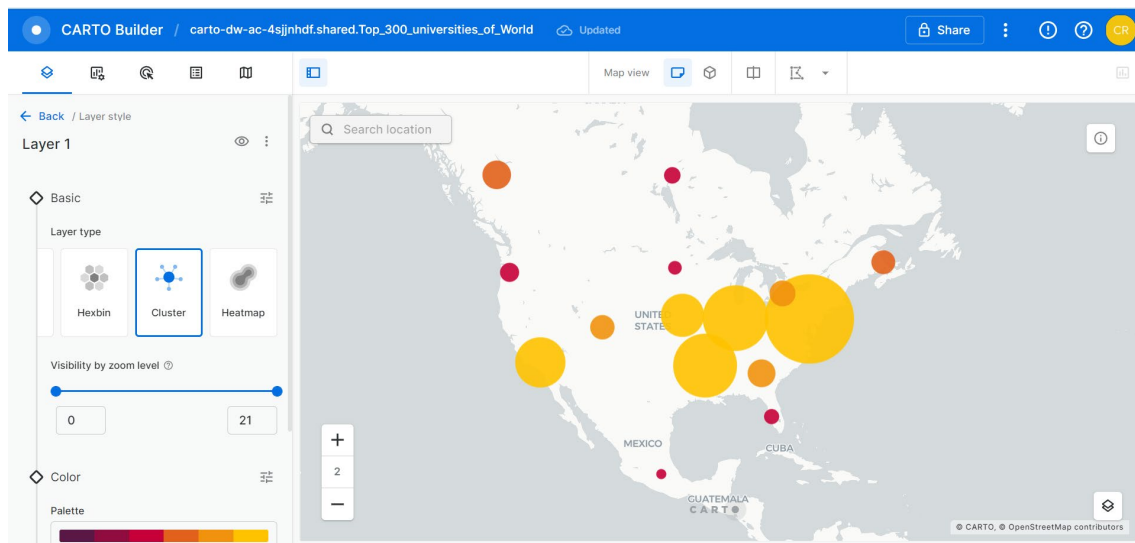


En cuanto a licores, se observa como el norte de Sudamérica y Asia Occidental son los dominantes. Los países de estos territorios parecen tener un consumo más alto en licores que Rusia.

Para el estudio de los siguientes dos datasets, se va a obviar la explicaciones del propio Carto y de la configuración de los puntos calientes, pues la manera de obrar es muy parecida a lo anterior descrito.

DATASET: 300 mejores universidades en el mundo en 2022

Para el estudio de este dataset, se ha realizado mediante cluters en vez de puntos calientes. Para realizar el cambio, seleccionamos “cluster” en “Layer Type”, y ya estaría preparado.



Luego ajustamos el mapa para veamos todos los continentes y los clusters se agrupan.

Como se observa, estos países han sido Estados Unidos y Inglaterra.