SDSC 2023 London

Cloud Native Spatial Analysis with CARTO: Exploring the Impacts of Climate Change on Insurable Assets

Giulia Carella, PhD giulia@carto.com
Lucía García-Duarte, lgarciaduarte@carto.com

Context

Assets such as infrastructure, buildings, and natural resources are critical to human society and economic growth. Climate change can have severe consequences on these assets, including damage from extreme weather events, sea-level rise, and shifting precipitation patterns and if these impacts are not adequately assessed and managed, they can lead to significant financial losses, social disruption, and environmental degradation.

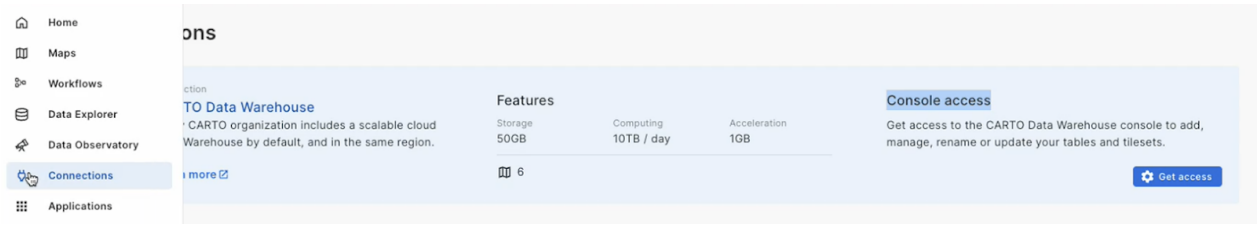
In this workshop, you will explore two use cases that demonstrate the practical application of CARTO's Analytics Toolbox for BigQuery in analyzing the impacts of climate change at the asset level. The first use case involves creating a composite score to measure the risk of extreme climate events for two major coffee shop brands in Great Britain. The second use case focuses on modeling the change in energy consumption by building in Great Britain under climate change. Through these use cases, you will learn how to effectively use CARTO's Analytics Toolbox for BigQuery to gain valuable insights and inform decision-making related to climate change impacts.

Set up you account before starting

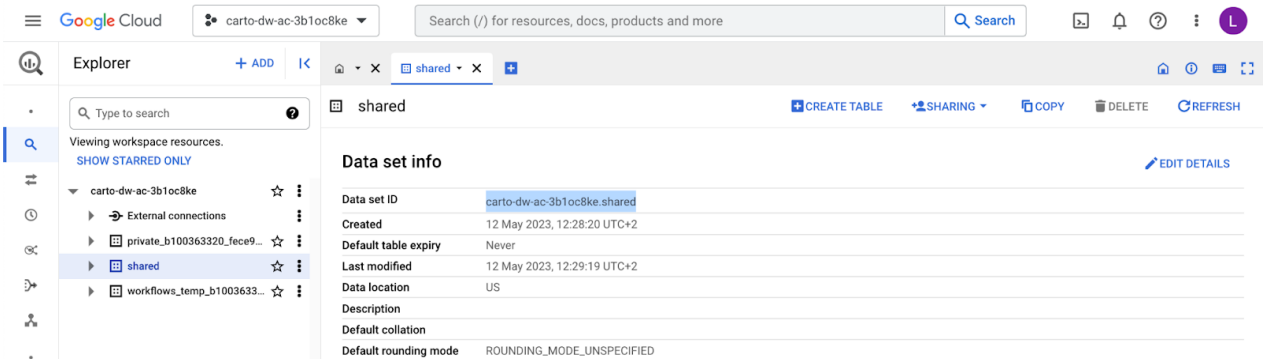
Before diving into the use cases, let us make sure that we have permissions to access our CARTO Data Warehouse through the BigQuery console:

1. Log in to your CARTO account.

1. Go to the *Connections* sections, located on the left side menu and click on *Get access*. Introduce the Google account you want to use.



2. Click on *Access console*, copy your qualified dataset name, and you are ready to go! Notice that:
    - You will be creating output tables in your own qualified dataset, so you will need to replace `<my-project>` by your project name in all upcoming queries.
    - You will be able to access the input data from this qualified dataset `cartobq.sdsc23_workshops`, which has only reading permissions.
    - All the outputs are also stored in `cartobq.sdsc23_workshops` in case you need to access them.



Create a composite score to represent the risk associated with extreme climate events in the period 2081-2100

We will start by creating a spatial composite score that represents the climate risk caused by severe weather occurrences under climate change. A spatial composite score is obtained by combining variables, scaled and weighted accordingly, into a meaningful indicator. Composite indicators can be used to measure complex and multidimensional concepts that otherwise are difficult to define, and cannot be measured directly, as for example the possible effects of climate change.

Prepare the input data

For this workshop, we will use data from different sources

- 2081-2100 averages for different climate indices measuring extreme temperature and precipitation conditions from The Climate Data Factory
- Sociodemographic, urbanity and POI-related information from CARTO Spatial Features
- Building footprints from Ordnance Survey to calculate built up areas as the total area covered by buildings

All the relevant data sources can be found in this public table where they have been previously joined:
`cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched`

Visualize the input data

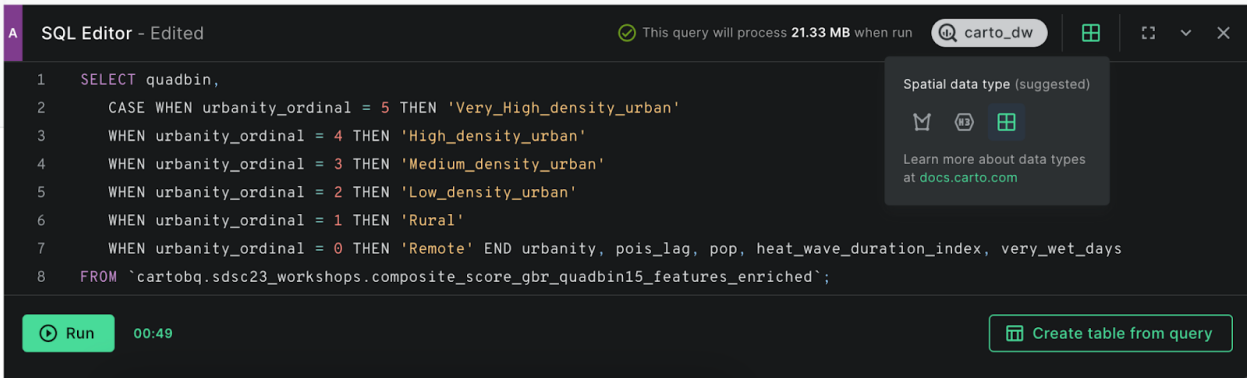
We will start by visualizing the data that will be used to create the composite score. The data are aggregated with a quadbin spatial index, with resolution 15. To create a map using CARTO, simply follow the next steps:

1. Go to the left-side menu to navigate the *Maps* section. Then click on *New map* on the right upper corner.

2. Select *Add source from* from the left bottom menu and click on *Custom Query (SQL)*, select your *CARTO Data Warehouse* connection and click on *Add source*.

3. Now you can type in your query to retrieve the variables you want to visualize. Make sure you are selecting quadbin as data type and click on *Run*.

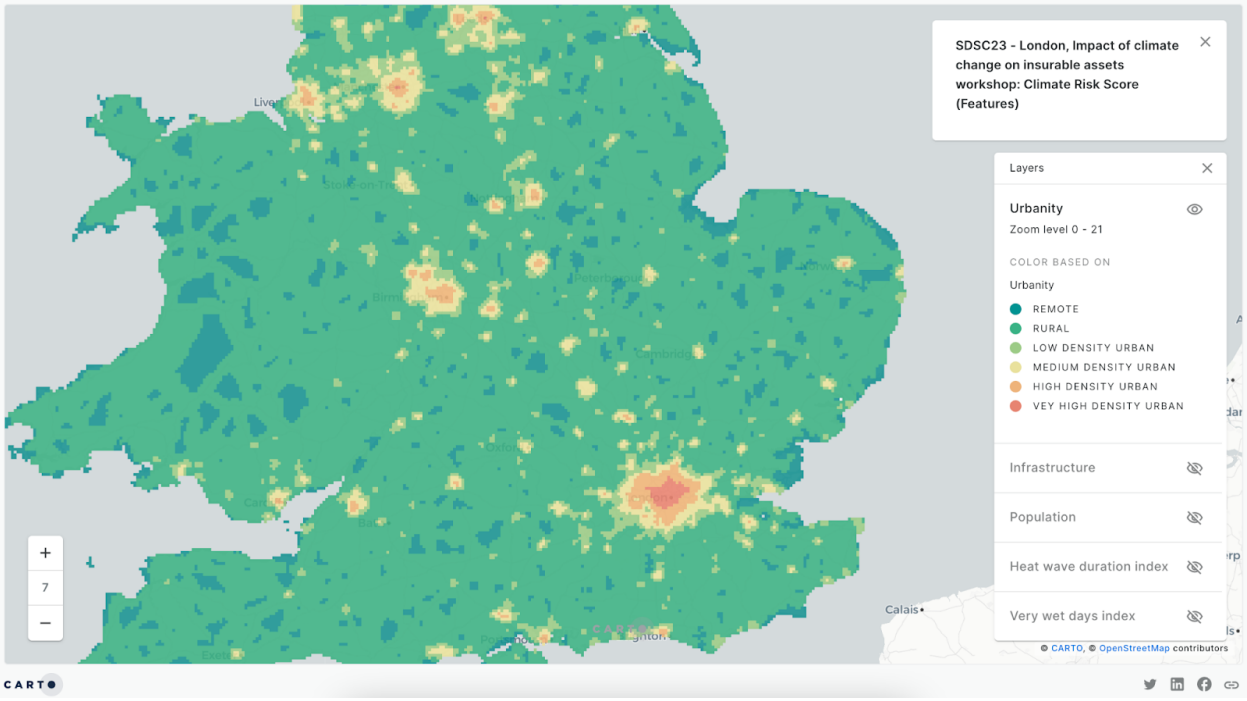
```
SELECT quadbin,
   CASE WHEN urbanity_ordinal = 5 THEN 'Very_High_density_urban'
   WHEN urbanity_ordinal = 4 THEN 'High_density_urban'
   WHEN urbanity_ordinal = 3 THEN 'Medium_density_urban'
   WHEN urbanity_ordinal = 2 THEN 'Low_density_urban'
   WHEN urbanity_ordinal = 1 THEN 'Rural'
   WHEN urbanity_ordinal = 0 THEN 'Remote' END urbanity, pois_lag, pop,
heat_wave_duration_index, very_wet_days
FROM `cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched`;
```



4. Next, style your map by selecting the *Layer 1* box that appears on the left side menu, by coloring the cells based on a variable that you want to visualize. For optimization purposes, you need to select an appropriate aggregation to correctly visualize the data at different zoom levels.

5. You need to create one layer per variable. To do this, simply select *Add layer* from the three dots menu (⋮) in the data source box.

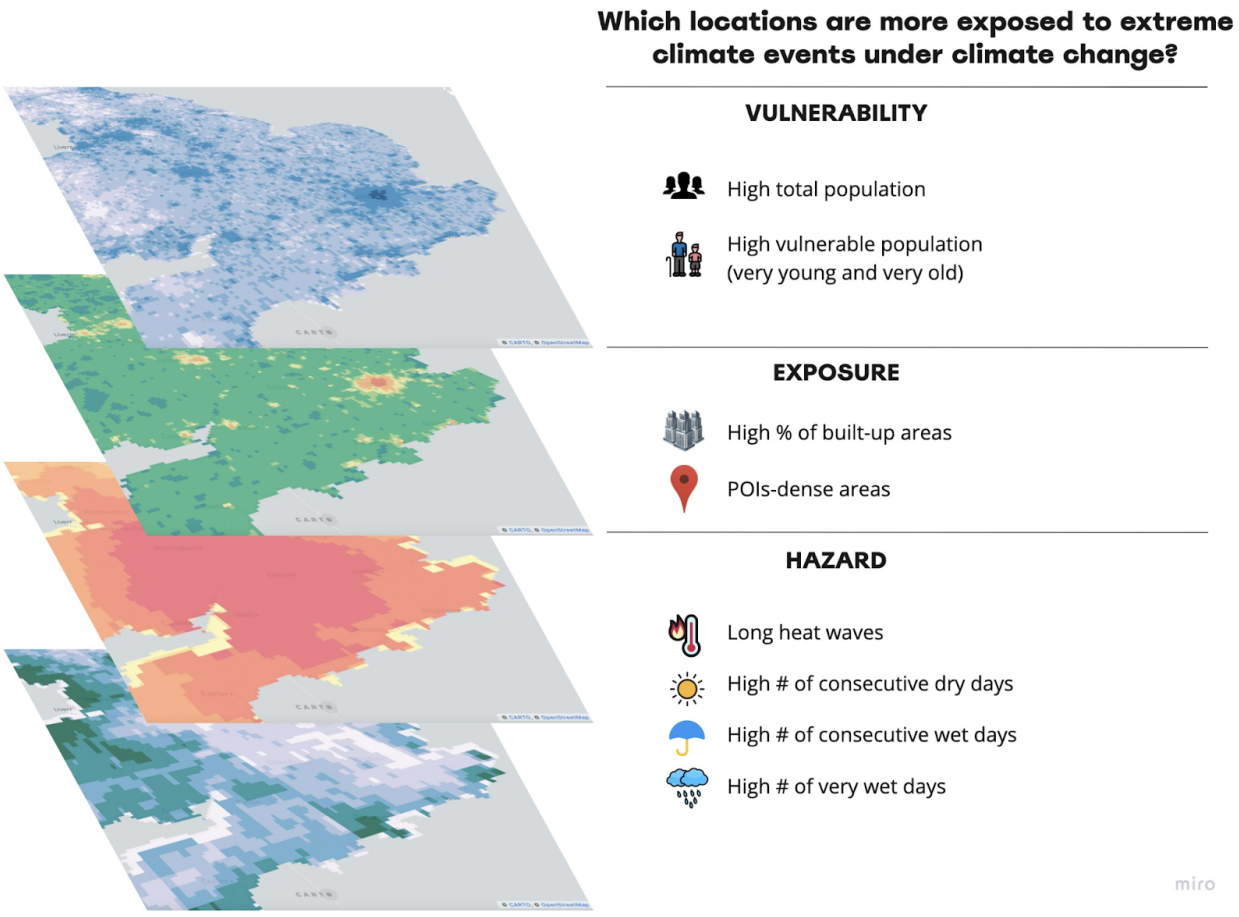
This map shows an example of the visualization of the input data.



Derive the Climate Risk Score

The composite score is derived using a Hazard - Exposure - Vulnerability model, that models the Risk as the product of these three components:

- **Hazard**: a potentially destructive physical phenomenon. Two types of extreme events will be considered here as potential danger: long dry periods and extreme heat waves duration, and long wet periods and extreme precipitation events.
- **Exposure**: the assets that could be affected by the hazard. For this use case, urbanity, infrastructure (points of interest) and building presence are key to specify regions more exposed to climate change.
- **Vulnerability**: the likelihood of the damage due to the exposure to a hazard. For instance, an elderly person may be more vulnerable to the impacts of extreme weather because of age and some medications, which may change how well the body responds to heat. Also, extreme weather conditions can affect children's physical and mental health, and the fact that they depend more on adults can put them at greater risk.



To compute each score using CARTO Analytics Toolbox, we can run the queries presented in the following sections. For a more detailed overview of how to compute composite scores, here you can access a detailed guide to help you select the optimal parameters for your specific use case.

A composite score to represent climate change hazards

The **Hazard Composite Score** was derived using the CREATE_SPATIAL_COMPOSITE_UNSUPERVISED procedure from a set of climate indexes related to extreme weather events provided by our partner The Climate Data Factory, as 2081-2100 averages: the largest number of consecutive days with temperature + 5°C above the average during the 1981-2010 reference period (`heat_wave_duration_index`), the number of 5-consecutive dry (`consecutive_dry_days_index`) and wet (`consecutive_wet_days_index`) days periods, and the percentage of very wet days (precipitation > 95th percentile as computed for the 1981-2010 reference period) during wet days (`very_wet_days`).

The method requires to set several parameters, including:

- `scoring_method`: the `FIRST_PC` method computes the score as the first principal component of a Principal Component Analysis and it was selected to maximize the variation in the input data.
- `correlation_var:` the score will be positively correlated with the variable `heat_wave_duration_index`
- `return_range:` the score will be returned on a 0-to-1 scale, where 0 means less hazard and 1 states for more hazard.

```
CALL `carto-un`.carto.CREATE_SPATIAL_COMPOSITE_UNSUPERVISED(
-- Query to the input data
'''SELECT quadbin, heat_wave_duration_index, consecutive_dry_days_index,
consecutive_wet_days_index, very_wet_days FROM
`cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched`''',
-- Name of the unique identifier column
'quadbin',
-- Query to the output data
'<my-project>.shared.composite_score_gbr_quadbin15_H_risk',
-- Query to select the spatial scoring options
'''{
```

```
   "scoring_method":"FIRST_PC",
   "correlation_var":"heat_wave_duration_index",
   "return_range":[0.0, 1.0]
}'''
);
```

We can also measure the internal consistency of the variables used to derive the spatial composite score based on the strength of correlations between individual variables using the CRONBACH_ALPHA_COEFFICIENT functionality. Larger values of the coefficient mean higher internal consistency.

```
CALL `carto-un`.carto.CRONBACH_ALPHA_COEFFICIENT(
-- Query to the input data
'''SELECT heat_wave_duration_index, consecutive_dry_days_index,
consecutive_wet_days_index, very_wet_days FROM
`cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched`''',
-- Query to the output data
'<my-project>.shared.composite_score_gbr_quadbin15_H_risk_cronbach'
);
```

Visualize the hazard composite score by land cover class

Next, we will use CARTO's raster loader to load a raster of the land cover types in Great Britain, which can be downloaded from here and visualize the computed hazard composite score for each class.

First, we need to load the data into Google BigQuery, which natively does not yet support raster data. Thanks to CARTO Raster Loader, available as an open source Python library, we can load the land cover raster file into BigQuery directly within the GCP console or on your local machine, by following these steps:

1. Install the package with pip

```
pip install raster-loader
```

2. Convert the raster to a Quadbin raster which uses Quadbin cells as pixels, and therefore requires resampling into the Web Mercator projection used by the Quadbin spatial index

```
gdalwarp gb2020lcm1km_dominant_aggregate.tif\
   -of COG \
   -co TILING_SCHEME=GoogleMapsCompatible \
   -co COMPRESS=DEFLATE \
   -co RESAMPLING=MODE \
   gb2020lcm1km_dominant_aggregate_quadbin.tif
```

3. Load the raster to Bigquery

```
carto bigquery upload \
    --file_path gb2020lcm1km_dominant_aggregate_quadbin.tif \
    --project <my-project> \
    --dataset shared \
    --table composite_score_gbr_landcover_raster \
    --output_quadbin
```

Here you can also find a docker-compose.yml file which can simplify the setup steps.

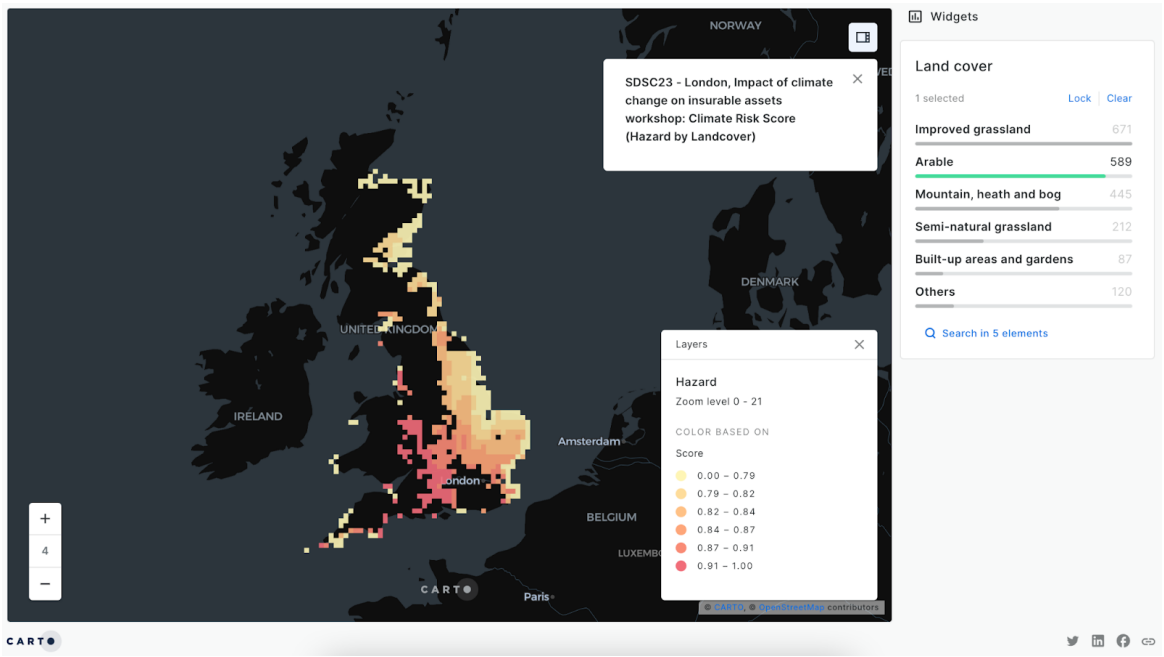
We can then run the following query using the RASTER_ST_GETVALUE procedure to return each pixel and the associated land cover class.

```
CALL `carto-un`.carto.RASTER_ST_GETVALUE(
'cartobq.sdsc23_workshops.composite_score_gbr_landcover_raster',
NULL,
NULL,
'<my-project>.shared.composite_score_gbr_landcover_raster_value'
);
```

Finally, we can visualize the hazard composite score by land cover class as shown in this map. To create your own, run the following query and visualize it.

```
CREATE OR REPLACE TABLE
`<my-project>.shared.composite_score_gbr_quadbin15_H_risk_landcover` AS (
-- Query to resample the resolution of the land cover data (quadbin 14 to 15)
WITH landcover AS (
```

```
   SELECT qb15 quadbin, band_1_uint8
   FROM `<my-project>.shared.composite_score_gbr_landcover_raster_value`,
   UNNEST (`carto-un`.carto.QUADBIN_TOCHILDREN(quadbin, 15) ) qb15
)
-- Query to join the hazard composite score to the land cover data and assign a
class label
SELECT hazard.*, CASE landcover.band_1_uint8
                 WHEN 1 THEN 'Broadleaf woodland'
                 WHEN 2 THEN 'Coniferous woodland'
                 WHEN 3 THEN 'Arable'
                 WHEN 4 THEN 'Improved grassland'
                 WHEN 5 THEN 'Semi-natural grassland'
                 WHEN 6 THEN 'Mountain, heath and bog'
                 WHEN 7 THEN 'Saltwater'
                 WHEN 8 THEN 'Freshwater'
                 WHEN 9 THEN 'Coastal'
                 ELSE 'Built-up areas and gardens'
                 END as ukceh_landcover_class
FROM `<my-project>.shared.composite_score_gbr_quadbin15_H_risk` hazard
LEFT JOIN  landcover
ON hazard.quadbin = landcover.quadbin
WHERE band_1_uint8 > 0
);
```



Procedures used:

- CREATE_SPATIAL_COMPOSITE_UNSUPERVISED
- CRONBACH_ALPHA_COEFFICIENT
- RASTER_ST_GETVALUE

A composite score to represent the assets exposure to climate change hazards

Next, we'll compute the **Exposure Composite Score** by selecting variables that represent infrastructure and urban areas as well as the following parameters:

- `scoring_method`:with the `CUSTOM_WEIGHTS` method, the score is computed by first scaling each input variable and then aggregating them according to user-defined scaling and aggregation functions and individual weights. In this case equal weights will be assigned to each
- `scaling`: the scaling function was set to `RANKING` to compute the percent rank since one of the input variables is ordinal (`urbanity_ordinal`). Notice that categorical variables need to be first converted to ordinal.
- `aggregation`: the `LINEAR` aggregation function will aggregate the variables linearly, which means that high values in one variable compensate for low values in others

```
CALL `carto-un`.carto.CREATE_SPATIAL_COMPOSITE_UNSUPERVISED(
-- Query to the input data
```

```
'''SELECT quadbin, pois_lag, urbanity_ordinal, buildings_area FROM
`cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched`''',
-- Name of the index column
'quadbin',
-- Query to the output data
'<my-project>.shared.composite_score_gbr_quadbin15_E_risk',
-- Query to select the spatial scoring options
'''{
   "scoring_method":"CUSTOM_WEIGHTS",
   "scaling": "RANKING",
   "aggregation": "LINEAR",
   "return_range":[0.0, 1.0]
}'''
);
```

Procedures used:

- CREATE_SPATIAL_COMPOSITE_UNSUPERVISED

A composite score to represent the vulnerability to climate change hazards

Finally, we will compute the **Vulnerability Composite Score** using a similar approach as for the Exposure Score, taking into account the total and vulnerable population and by setting the following parameters:

- `weights:` by giving a weight of 0.2 to the total population (`pop`), the remaining 0.8 is split into the other variables (`pop_under_15` and `pop_65_and_over`), which will imply a greater influence over the final score
- `aggregation:` with the `GEOMETRIC` aggregation high values in one variable won't be fully compensated with low values in others and balance between the input variables is rewarded, which is preferable when representing vulnerability . Notice that the input variable values must be strictly positive, so we will remove those cells where data is below a close-to-0 threshold

```
CALL `carto-un`.carto.CREATE_SPATIAL_COMPOSITE_UNSUPERVISED(
-- Query to the input data
'''SELECT quadbin, pop_under_15, pop_65_and_over, pop FROM
```

```
`cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_features_enriched` WHERE
pop_under_15 > 0.001 AND pop_65_and_over > 0.001 AND pop > 0.001 ''',
-- Name of the index column
'quadbin',
-- Query to the output data
'<my-project>.shared.composite_score_gbr_quadbin15_V_risk',
-- Query to select the spatial scoring options
'''{
  "scoring_method":"CUSTOM_WEIGHTS",
  "scaling": "MIN_MAX_SCALER",
  "aggregation": "GEOMETRIC",
  "weights": {"pop":0.2},
  "return_range":[0.0, 1.0]
}'''
);
```

Procedures used:

- CREATE_SPATIAL_COMPOSITE_UNSUPERVISED

Combining the sub-scores

To get the final **Climate Risk Score** we will use again the CREATE_SPATIAL_COMPOSITE_UNSUPERVISED procedure, this time also by specifying how to discretize the resulting score:

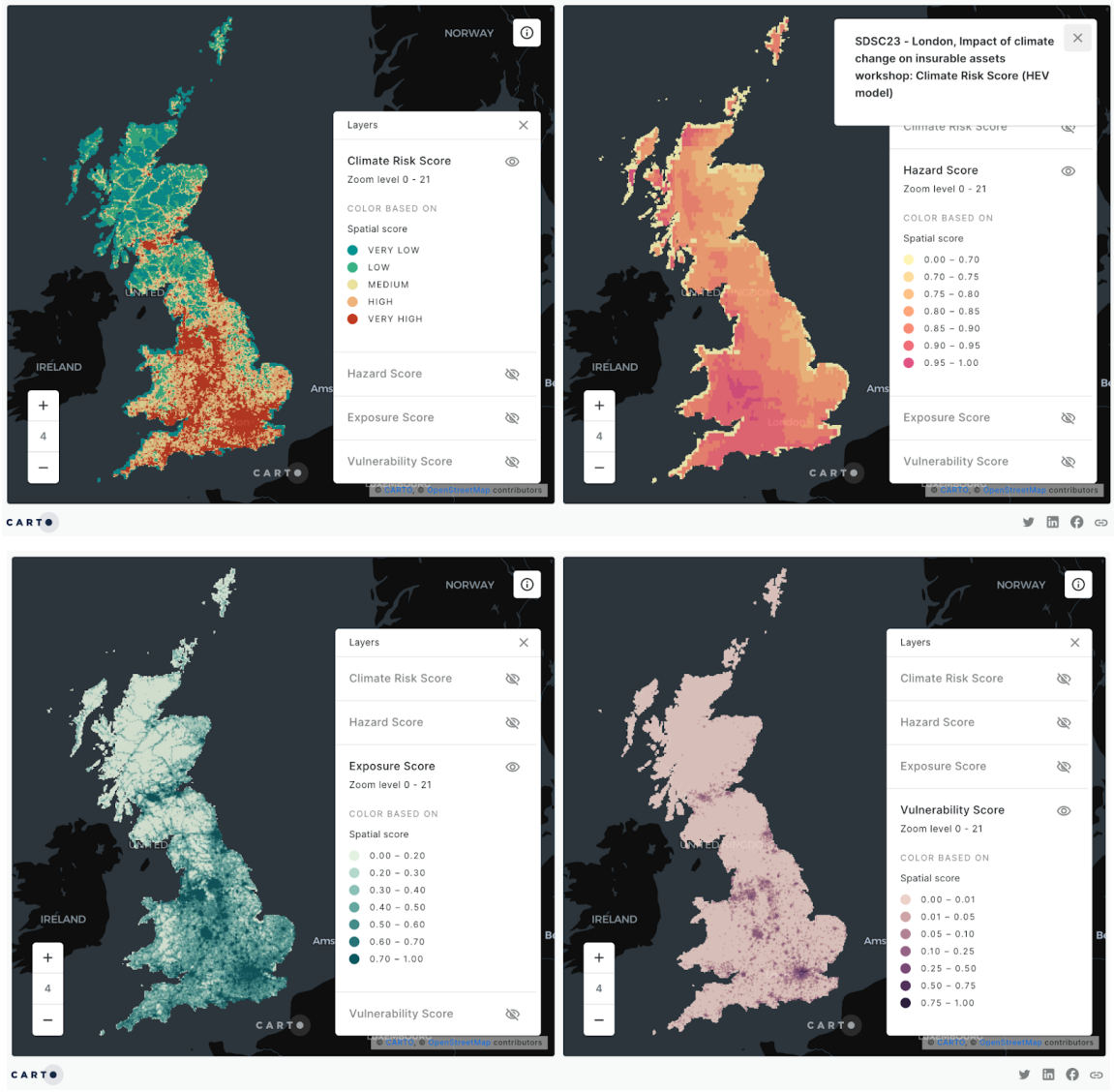
- `bucketize_method`: the `quantiles` method is used to categorize the score into `nbuckets` (5) quantile-based breaks.

```
CALL `carto-un`.carto.CREATE_SPATIAL_COMPOSITE_UNSUPERVISED(
-- Query to the input data
'''
 SELECT h.quadbin, h.spatial_score as hazard, e.spatial_score as exposure,
v.spatial_score as vulnerability
 FROM `<my-project>.shared.composite_score_gbr_quadbin15_H_risk` as h
 LEFT JOIN `<my-project>.shared.composite_score_gbr_quadbin15_E_risk` as e
 ON h.quadbin = e.quadbin
 LEFT JOIN `<my-project>.shared.composite_score_gbr_quadbin15_V_risk` as v
```

```
 ON h.quadbin = v.quadbin
''',
-- Name of the index column
'quadbin',
-- Query to the output data
'<my-project>.shared.composite_score_gbr_quadbin15_R_risk_quantiles',
-- Query to select the spatial scoring options
'''{
 "scoring_method":"CUSTOM_WEIGHTS",
 "aggregation": "LINEAR",
 "scaling": "MIN_MAX_SCALER",
 "return_range":[0.0, 1.0],
 "bucketize_method": "quantiles",
 "nbuckets": 5
}'''
);
```

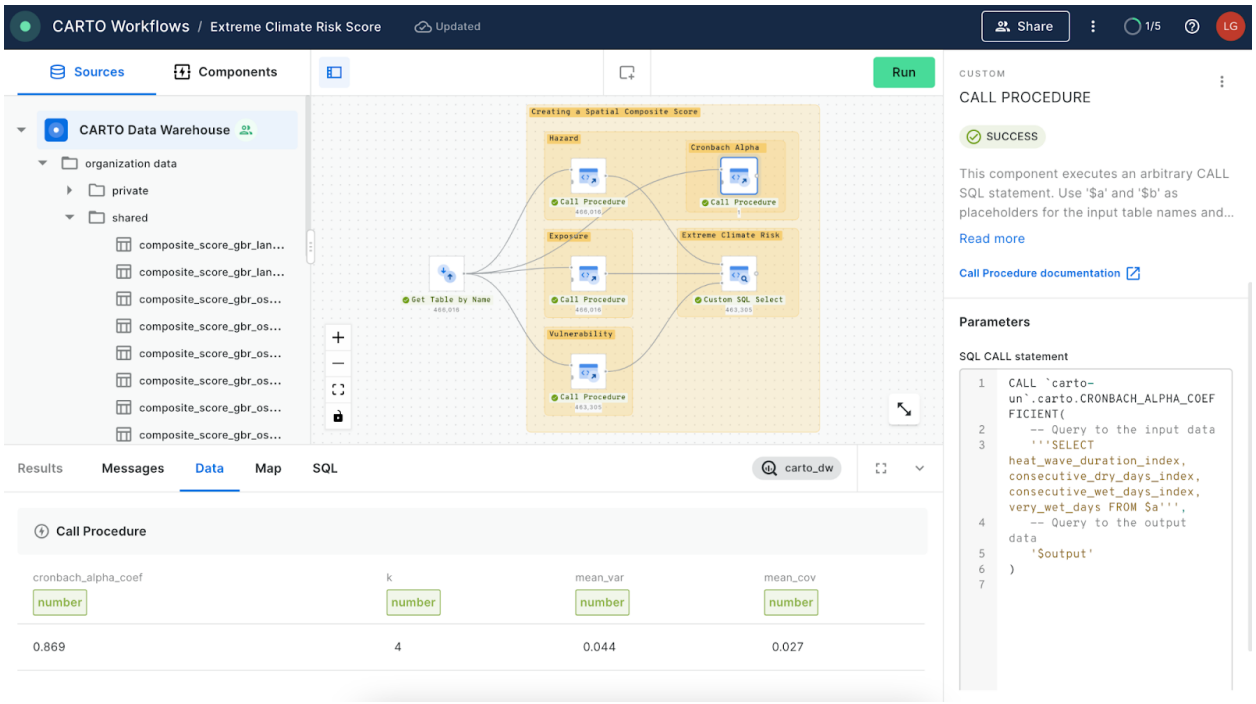
We can finally assign a label to the discretized score and visualize the result on a map by creating the following table.

```
CREATE OR REPLACE TABLE
`<my-project>.shared.composite_score_gbr_quadbin15_R_risk_quantiles` AS (
  SELECT *, CASE WHEN spatial_score = 1 THEN 'Very Low'
    WHEN spatial_score = 2 THEN 'Low'
    WHEN spatial_score = 3 THEN 'Medium'
    WHEN spatial_score = 4 THEN 'High'
    ELSE 'Very High'
    END spatial_score_cat
  FROM `<my-project>.shared.composite_score_gbr_quadbin15_R_risk_quantiles`
);
```



Use CARTO Workflows to compute the Climate Risk Score with a reproducible and repeatable analysis

While spatial SQL provides a common language that can be used to perform spatial analysis and harness the parallel processing power of data warehouses, some queries might not look very accessible in practice. CARTO Workflows is built on the premise of making this analytic power more accessible for users of any skill level or role. Here you can find an example of using Workflows to compute the Climate Risk Score.



Procedures used:

- CREATE_SPATIAL_COMPOSITE_UNSUPERVISED
- CRONBACH_ALPHA_COEFFICIENT

Which stores are more at risk under climate change?

Let's now use the computed Climate Risk Score to characterize the risk at the asset level, by comparing the market penetration of two of the main coffee shop brands in Great Britain (Starbucks and Costa Coffee) in areas where the Climate Risk is High or Very High.

To do this we can use the Merchant universe matching analysis available in the Consumer Packaged Goods (CPG) module of CARTO's Analytics Toolbox to generate a mapping from our current universe (the list of Starbucks and Costa Coffee merchants) to the total universe (a larger set of potential merchants, usually from third party data) and extract insights from it.

To simulate the current universe, we jittered the locations of Starbucks and Costa Coffee shops from the OpenStreetMap dataset available in Google BigQuery. To simulate the total universe, we took instead the exact locations of all coffee shop

brands available in Great Britain from the same dataset. We can then visualize the current and total universe in this map.

First, we run the UNIVERSE_MATCHING procedure in the Analytics Toolbox, which performs a fuzzy match between the two datasets provided, using the venues' spatial position as a first filter and the venues' name as the second filter.

```
CALL `carto-un`.carto.UNIVERSE_MATCHING(
-- Current universe arguments
'''
  SELECT a.*
  FROM `cartobq.sdsc23_workshops.composite_score_gbr_osm_starbucks_costa` a
  JOIN `cartobq.sdsc23_workshops.composite_score_gbr_quadbin15_R_risk_quantiles` b
  ON ST_CONTAINS(`carto-un`.carto.QUADBIN_BOUNDARY(b.quadbin), a.geom)
  WHERE a.osm_name = "Starbucks" AND b.spatial_score_cat IN ('High','Very High')
''',
'osm_id',
'osm_name',
'geom',
-- Total universe arguments
'cartobq.sdsc23_workshops.composite_score_gbr_osm_cafes',
'osm_id',
'osm_name',
'geom',
-- Universe matching arguments
'<my-project>.shared.composite_score_gbr_osm_starbucks_costa_universe_matching_resu
lts_starbucks',
'''{
-- Hard limits on number of neighbors and distance
  "max_neighbors": 60,
  "max_distance": 1500,
-- Weights used to compute the weighted average based on proximity and text
similarity
  "weights": {"text_similarity": 0.7, "proximity": 0.3}
}'''
);
```

Finally, using the UNIVERSE_MATCHING_REPORT procedure we can create a filtered table using a minimum similarity acceptable for each pair, create an expansion universe table, including all the rows in the expansion universe that are matched over a user-defined minimum similarity threshold, and create a report which includes the market penetration of the current universe.

```
CALL `carto-un`.carto.UNIVERSE_MATCHING_REPORT(
-- Total universe arguments
'cartobq.sdsc23_workshops.composite_score_gbr_osm_cafes',
'osm_id',
-- Universe matching results
'<my-project>.shared.composite_score_gbr_osm_starbucks_costa_universe_matching_results_starbucks',
-- Report arguments
'<my-project>.shared.composite_score_gbr_osm_starbucks_costa_universe_matching_starbucks',
'''{
  -- Minimum similarity threshold
  "min_similarity": 0.7,
}'''
);
```

If we query the report we can compare Starbucks and Costa Coffee market penetration in high climate risk areas

```
(SELECT 'Starbucks' AS osm_name, * FROM
`<my-project>.shared.composite_score_gbr_osm_starbucks_costa_universe_matching_starbucks_report`)
UNION ALL
(SELECT 'Costa Coffee' AS osm_name, * FROM
`<my-project>.shared.composite_score_gbr_osm_starbucks_costa_universe_matching_costa_report`)
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	osm_name	current_universe	total_universe	matched_universe	expansion_universe	market_penetration
1	Costa Coffee	485	3490	402	3125	0.11518624641833...
2	Starbucks	258	3490	188	3329	0.05386819484240...

Procedures used:

- UNIVERSE_MATCHING
- UNIVERSE_MATCHING_REPORT

Derive the Area-of-Applicability of a Machine Learning model used to make predictions for the years 2081-2100 of the energy consumption of each building in Great Britain

The use of Machine Learning (ML) methods has gained immense popularity for spatio-temporal mapping, thanks to their remarkable ability to capture complex and nonlinear relationships. However, a major drawback of these algorithms is that they can only be effectively used on new data if they are similar to the training data. Given that spatio-temporal mapping involves making predictions for new predictor properties, it becomes crucial to determine the reliability of the prediction model for a particular area. To address this issue, we can compute the Area of Applicability (AOA) as the region where the model predictions can be trusted when the predictions are extrapolated outside the training space (i.e. where the estimated cross-validation performance holds).

Query the model SHAP values

We'll start from a pre-trained ML model of synthetic annual per capita energy consumption data for 2019 for each Middle Layer Super Output Area (MSOA). We can query the model SHAP values, to identify the relative importance of each model predictor.

```
SELECT *
FROM `cartobq.sdsc23_workshops.energy_consumption_engwal_msoa_2019_shap`
```

In this model we are using only two predictors: the total annual income and the sum of the Heating and Cooling Degree Days (HDD and CDD respectively) in 2019. The

latter was provided by our data partner The Climate Data Factory, and represents a proxy of the energy consumption for heating and air conditioning usage.

Next, we use this model to make predictions for the years 2081-2100 on a quadbin grid with resolution 15. Here we will make the assumption that the total annual income for each grid point will not have changed compared to present conditions and focus instead on the effect of the change in the heating and cooling degrees days variable under climate change. The forecast for the heating and cooling degrees days was provided by The Climate Data Factory as the annual mean for 2081-2100.

```
SELECT * FROM ML.PREDICT(
  MODEL `cartobq.sdsc23_workshops.energy_consumption_engwal_msoa_2019_xgb`,
  (SELECT quadbin, hdd_cdd_sum, total_annual_income
  FROM `cartobq.sdsc23_workshops.energy_consumption_engwal_quadbin15_2081-2100`)
)
```

Estimate the Area-of-Applicability

Given the SHAP values of the pre-trained model, the Area-of-Applicability (AOA) procedure computes a Dissimilarity Index (DI) for each new data point used for prediction as the multivariate distance between the model covariates for that point and the nearest training data point. To identify those new points that lie in the model AOA, the DI is compared using a threshold obtained as the (outlier-removed) maximum DI of the training data derived via cross-validation: for each training data point the DI is computed as the distance to the nearest training data point that is not in the same (spatial) cross-validation fold with respect to the average of all pairwise distances between all training data.

To compute the Area-of-Applicability (AOA) of the model when this is used to derive estimates for the 2081-2100 period, we can run this query.

```
CALL`carto-un`.carto.AREA_OF_APPLICABILITY(
  -- Query to the training data
  '''SELECT msoa_code AS geoid, hdd_cdd_sum, total_annual_income  FROM
`cartobq.sdsc23_workshops.energy_consumption_engwal_msoa_2019`''',
  -- Query to the data used for prediction
  '''SELECT quadbin AS geoid, hdd_cdd_sum, total_annual_income FROM
```

```
`cartobq.sdsc23_workshops.energy_consumption_engwal_quadbin15_2081-2100`''',
  -- Query to the model SHAP values
  '''SELECT * FROM
`cartobq.sdsc23_workshops.energy_consumption_engwal_msoa_2019_shap`''',
   -- Name of the column of the geographic unique identifier
  'geoid',
  -- Output prefix
  '<my-project>.shared.energy_consumption_engwal_quadbin15_2081-2100_AOA',
  -- Options
  '''{
    --- The distance function
    "distance_type":"GOWER",
    ---The cross-validation strategy: in this example we are using a spatial
cross-validation strategy, where the training data is split using a multivariate
method (Principal Component Analysis + K-means clustering) to specify sets of
similar conditions based on the input covariates
    "threshold_method":"ENV_BLOCKING_KFOLD",
    "nfolds":5,
    --- The scale factor used to define the threshold, analogue to Tukey's fences k
parameter for outlier detection
    "outliers_scale_factor":1.5,
    --- Whether we want to normalize the DI between [0,1]
    "normalize_dissimilarity_index":True
  }'''
);
```

Procedures used:

- AREA_OF_APPLICABILITY

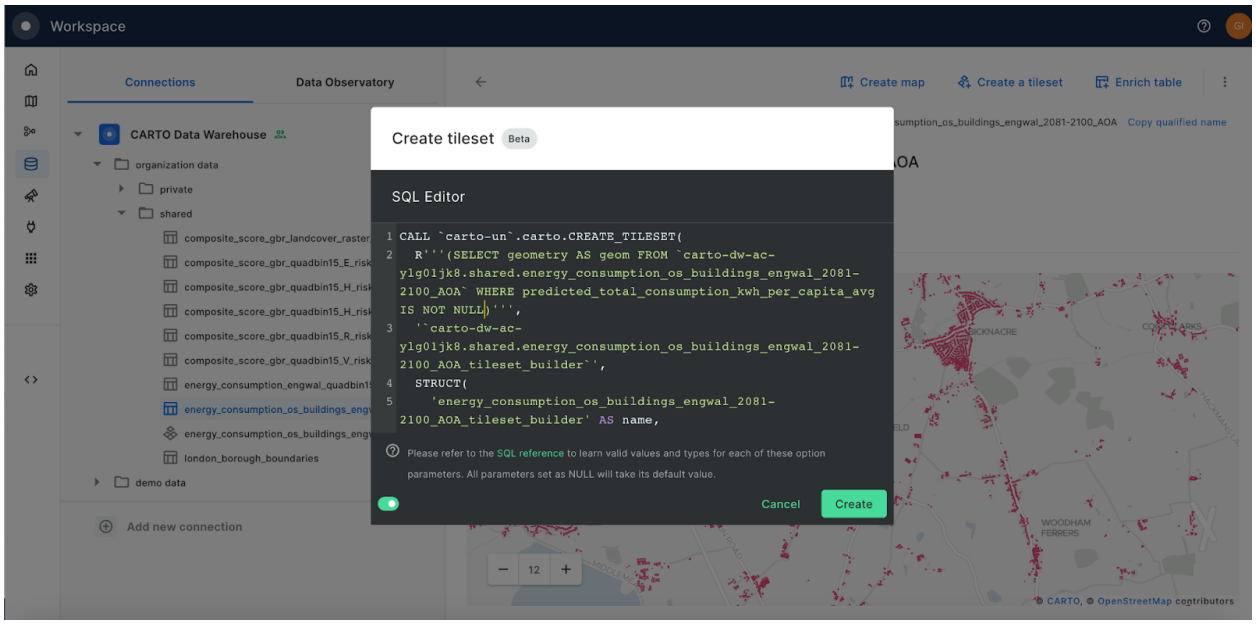
Enrich building footprints within the model Area-of-Applicability with the model estimates

Finally, we can derive the predicted energy consumption per capita by building footprint, by averaging over each polygon the model predictions within the AOA using the enrichment capabilities of CARTO's Analytics Toolbox.

```
CALL `carto-un`.carto.ENRICH_POLYGONS(
-- Input query
R'''
SELECT * FROM `cartobq.sdsc23_workshops.energy_consumption_gbr_os_buildings`
WHERE ST_AREA(geometry)>1
```

```
''',
-- Name of the geometry column in the input query
'geometry',
-- Query to the data that will be used to enrich the polygons provided in the input
query
R'''
SELECT `carto-un`.carto.QUADBIN_BOUNDARY(quadbin) AS geom,
predicted_total_consumption_kwh_per_capita
FROM ML.PREDICT(
  MODEL `cartobq.sdsc23_workshops.energy_consumption_engwal_msoa_2019_xgb`,
  (SELECT quadbin, hdd_cdd_sum, total_annual_income
  FROM `cartobq.sdsc23_workshops.energy_consumption_engwal_quadbin15_2081-2100` a
  JOIN `<my-project>.shared.energy_consumption_engwal_quadbin15_2081-2100_AOA`b
  ON a.quadbin = CAST(b.geoid AS INT)
  WHERE is_in_area_of_applicability = True)
)
''',
-- Name of the geometry column in the data query
'geom',
-- Enrichment variable and aggregation function
[('predicted_total_consumption_kwh_per_capita', 'avg')],
-- Fully qualified name of the output table
[``<my-project>.shared.energy_consumption_engwal_os_buildings_2081-2100_AOA`']
);
```

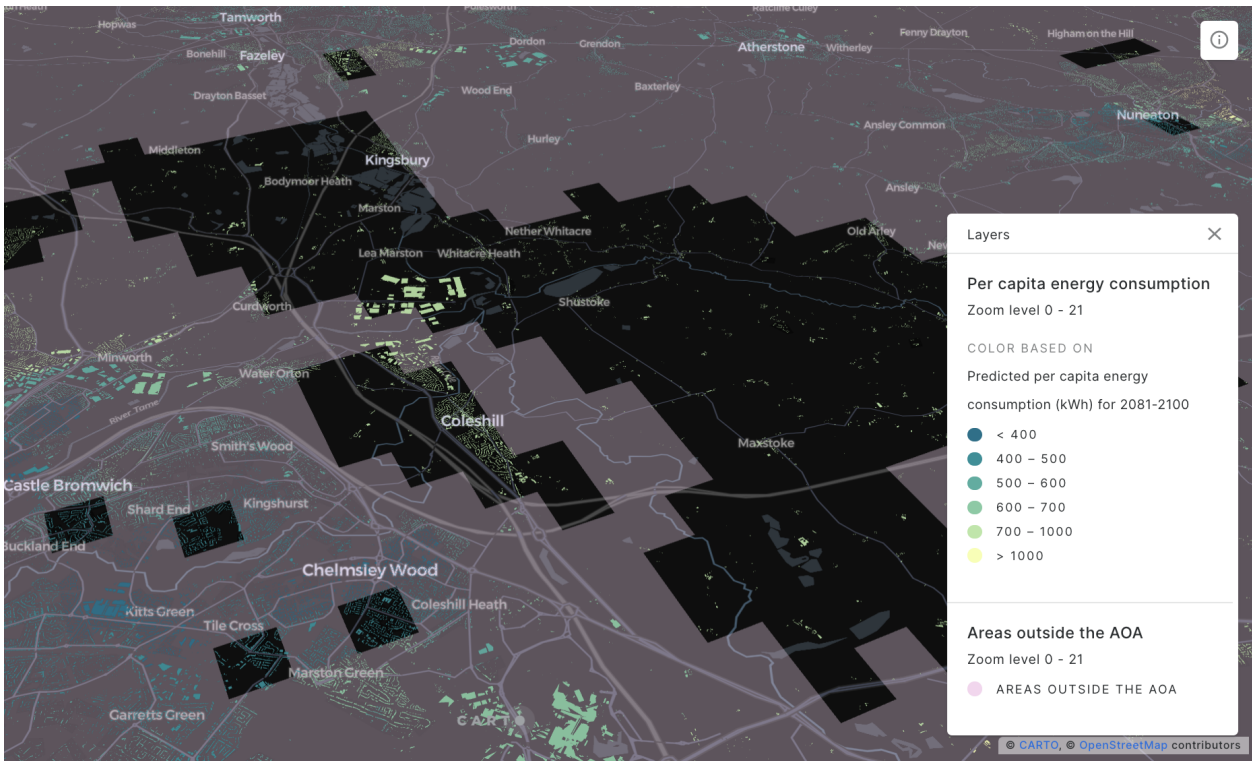
To visualize the results on a map, we can create a fast and scalable visualization using CARTO's Analytics Toolbox tiler directly from the Data Explorer.



Or by running the following query

```
CALL `carto-un`.carto.CREATE_TILESET(
-- Input query
R'''
(SELECT geometry, predicted_total_consumption_kwh_per_capita_avg AS
predicted_total_consumption_kwh_per_capita
FROM `<my-project>.shared.energy_consumption_engwal_os_buildings_2081-2100_AOA`
WHERE predicted_total_consumption_kwh_per_capita_avg IS NOT NULL)
''',
-- Fully qualified name of the output table
''``<my-project>.shared.energy_consumption_engwal_os_buildings_2081-2100_AOA_tileset`''',
-- Options
STRUCT(
  "Energy consumption per capita 2081-2100" AS name,
  "This tileset shows the predicted energy consumption per capita for 2081-2100. Only buildings within the model Area-of-Applicability are shown. The building footprints were derived from Ordnance Survey" AS description,
  NULL AS legend,
  0 AS zoom_min,
  18 AS zoom_max,
```

```
  "geometry" AS geom_column_name,
  NULL AS zoom_min_column,
  NULL AS zoom_max_column,
  1024 AS max_tile_size_kb,
  "RAND() DESC" AS tile_feature_order,
  true AS drop_duplicates,
  R'''
    "custom_metadata": {
      "version": "1.0.0",
      "layer": "os_buildings_AOA"
    }
  ''' AS extra_metadata
));
```



Procedures used:

- ENRICH_POLYGONS
- QUADBIN_BOUNDARY
- CREATE_TILESET