SDSC 2022 New York

Powering Cloud-based Spatial Analytics For Retail with CARTO [short guide for visualization]

Context

Identifying an optimal location for a new store is not always an easy task, and we often do not have enough data at our disposal to build a solid model to predict potential revenues across an entire territory. In these cases, managers rely on different business criteria in order to make a sound decision for their expansion strategy. For example, they rely on defining their target market and segmenting population groups accordingly in order to locate the store closer to where the target market lives (e.g. areas with a great presence of youngsters).

In this example, we are going to use CARTO's Analytics Toolbox for BigQuery to explore good locations to open a new Pizza Hut restaurant in Honolulu, Hawaii. To do that, we will run a couple of different spatial analyses.

Area of study

We will start by defining an area of interest for our study. For that, we define a buffer of 5 km around downtown Honolulu.

```
-- We use the ST_BUFFER to define a 5 km buffer centered in Honolulu
SELECT ST_BUFFER(ST_GEOGPOINT(-157.852587, 21.304390), 5000) AS geom
```

Functions used:

- ST_BUFFER

Find and visualize all Pizza Huts in Honolulu

Next, we will get all Pizza Hut restaurants in Honolulu. For that, we use OSM POI data, which is publicly in CARTO's Spatial Data Catalog.

```sql
SELECT
tag.value AS brand, geometry AS geom,
FROM
`cartobq.docs.sdsc_honolulu_osm_planet_nodes` d,
--
`carto-do-public-data.openstreetmap.pointsofinterest_nodes_usa_latlon_v1_quarterly_v1` d,
UNNEST(all_tags) as tag
WHERE ST_CONTAINS(ST_BUFFER(ST_GEOGPOINT(-157.852587, 21.304390), 5000), geometry)
AND ((tag.value in ("Pizza Hut") AND tag.key = 'brand'))
```

Note that we are using the table `` `cartobq.docs.sdsc_honolulu_osm_planet_nodes` `` that contains all POIs in the area of study instead of the OSM table (commented in the code). We are using this table so that:

- Attendees don't need to create a new connection.
- Queries run faster.

Get the right data

Visualize area of study polyfilled

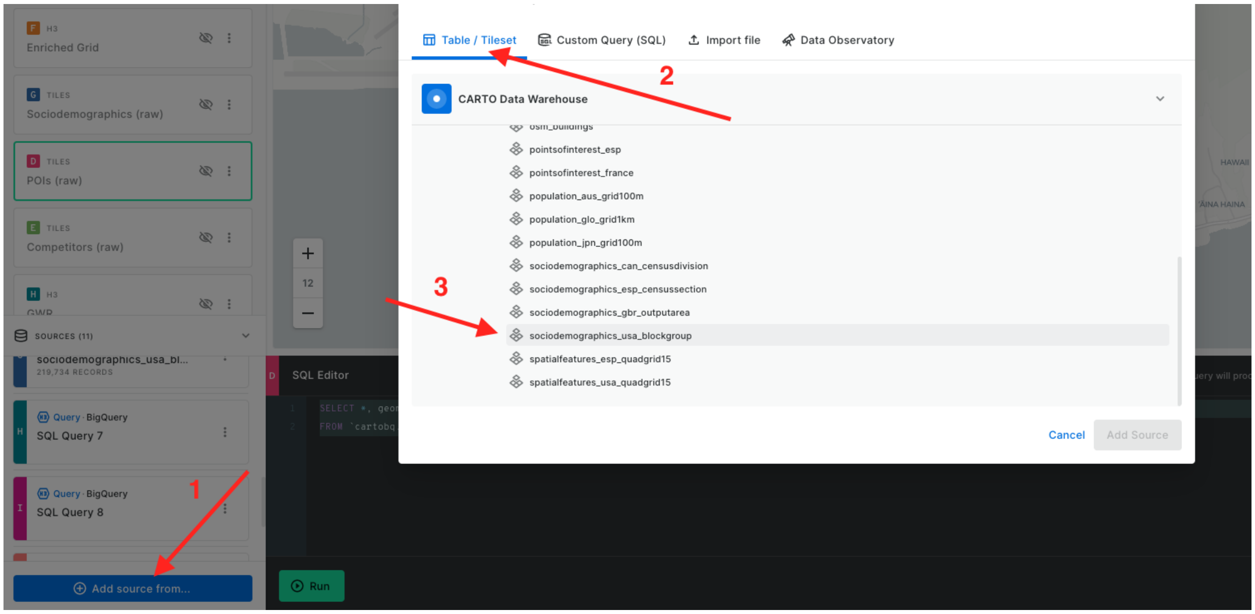
```sql
SELECT h3
FROM UNNEST(`carto-un`.carto.H3_POLYFILL(ST_BUFFER(ST_GEOGPOINT(-157.852587,
21.304390), 5000), 10)) h3
```

Visualize POIs (raw)

```sql
SELECT *, geometry AS geom FROM `cartobq.docs.sdsc_honolulu_osm_planet_nodes`
```

Visualize sociodemographics (raw)

We'll add a new Tileset visualization in 3 simple steps.



Visualize final enriched grid

```
SELECT * FROM `cartobq.docs.sdsc_honolulu_aos_enriched`
```

Analysis

Moran's I

We start by running some descriptive analyses.

We are interested in knowing if the variables selected show spatial autocorrelation. This can be a very powerful analysis because if our variable of interest shows high spatial autocorrelation, that means:

- We can find homogeneous local areas more easily
- We have more flexibility for selecting the exact location for the new store

```
SELECT `carto-un`.carto.MORANS_I_H3(input_data, 2, 'exponential')
FROM (
  SELECT ARRAY_AGG(STRUCT(h3, pop_15_34_pct)) AS input_data
```

```
  FROM `cartobq.docs.sdsc_honolulu_aos_enriched`
)
```

Functions used:

- [MORANS_I_H3]

Geographically Weighted Regression

Next, we'd like to know how income per capita is related to the other sociodemographics variables.

For Pizza Hut, higher income does not necessarily translate into higher sales.

```
SELECT * FROM `cartobq.docs.sdsc_honolulu_gwr`
```

Functions used:

- [GWR_GRID]

Commercial hotspots

Next, we would like to identify areas that meet Pizza Hut requirements, i.e., locations with large populations aged 15-34 and far from existing Pizza Hut restaurants.

In order to identify these locations, we use the Commercial hotspots functionality available in the AT. This functionality identifies areas with values that are significantly higher than the average.

```
SELECT index as h3, combined_gi, p_value
FROM `cartobq.docs.sdsc_honolulu_commercial_hotspots`
WHERE p_value < 0.05
```

Functions used:

- [COMMERCIAL_HOTSPOTS]

Local outlier factor

As next step, we'll visualize Pizza Hut competitors and compute the local outlier factor to identify those that are very close to one another and those far from the other, to visualize where it would be more interesting to open a new restaurant.

Visualize competitors

```
SELECT CAST(id AS STRING) AS id , tag.value, geometry as geom
FROM `cartobq.docs.honolulu_planet_nodes` d,
UNNEST(all_tags) as tag
WHERE ST_CONTAINS(ST_BUFFER(ST_GEOGPOINT(-157.852587, 21.304390), 5000), geometry)
AND ((tag.value in ('fast_food', 'restaurant') AND tag.key = 'amenity'))
```

Compute LOF

```
-- We get all amenities tagged as restaurants or fast_food POIS in Honolulu
WITH fast_food AS (
SELECT CAST(id AS STRING) AS id , tag.value, geometry as geom
FROM `cartobq.docs.honolulu_planet_nodes` d,
UNNEST(all_tags) AS tag
WHERE ST_CONTAINS(ST_BUFFER(ST_GEOGPOINT(-157.852587, 21.304390), 5000), geometry)
AND ((tag.value IN ('fast_food', 'restaurant') AND tag.key = 'amenity'))
),
-- We calculate the Local Outlier Factor in order to identify restaurants without
competition.
lof_output as (
SELECT `carto-un`.carto.LOF(ARRAY_AGG(STRUCT(id,geom)), 5) AS lof FROM fast_food
)
SELECT lof.geoid, lof.geo as geom, lof.lof
FROM lof_output, UNNEST(lof_output.lof) AS lof
```

Functions used:

- [LOF]

Twin Areas

Next, we'd like to identify areas in the city that are similar to the best performing store located in cell `8a464b96b817fff`.

We'll run this analysis using the twin areas functionality that given an origin location (best performing store in this case) and a set of variables (external variables we have used to characterized the grid), identifies areas that are similar to the origin. For further detail, take a look at this blog post.

```
SELECT * FROM `cartobq.docs.sdsc_honolulu_twin_scores`
```

Functions used:

- FIND_TWIN_AREAS