

# Building Spatial Applications with Databricks and CARTO

ORGANIZED BY  databricks



**Isabel Pozuelo**

Spatial Analytics expert, CARTO  
ipozuelo@carto.com



**Kent Marten**

Staff Product Manager, Databricks  
kent.marten@databricks.com  
@datakmart

# Improving food Delivery Time estimations using Spatial Analytics

Since 2019, Brits have increased their spending on takeaways by 42%. Londoners are no exception, spending on average £781 per year each. However, competition amongst restaurants in the capital is high and food should arrive when promised. If it doesn't, the customer can demand refunds and take their business elsewhere.

Understanding the spatial nature of these problems is crucial. Poor delivery times could be affected by factors such as busy restaurants, congested streets or low availability of delivery drivers, all of which are inherently spatial. So what data will delivery businesses have to assess performance?

Start →



STORY MAP

# Optimizing Restaurant Deliveries

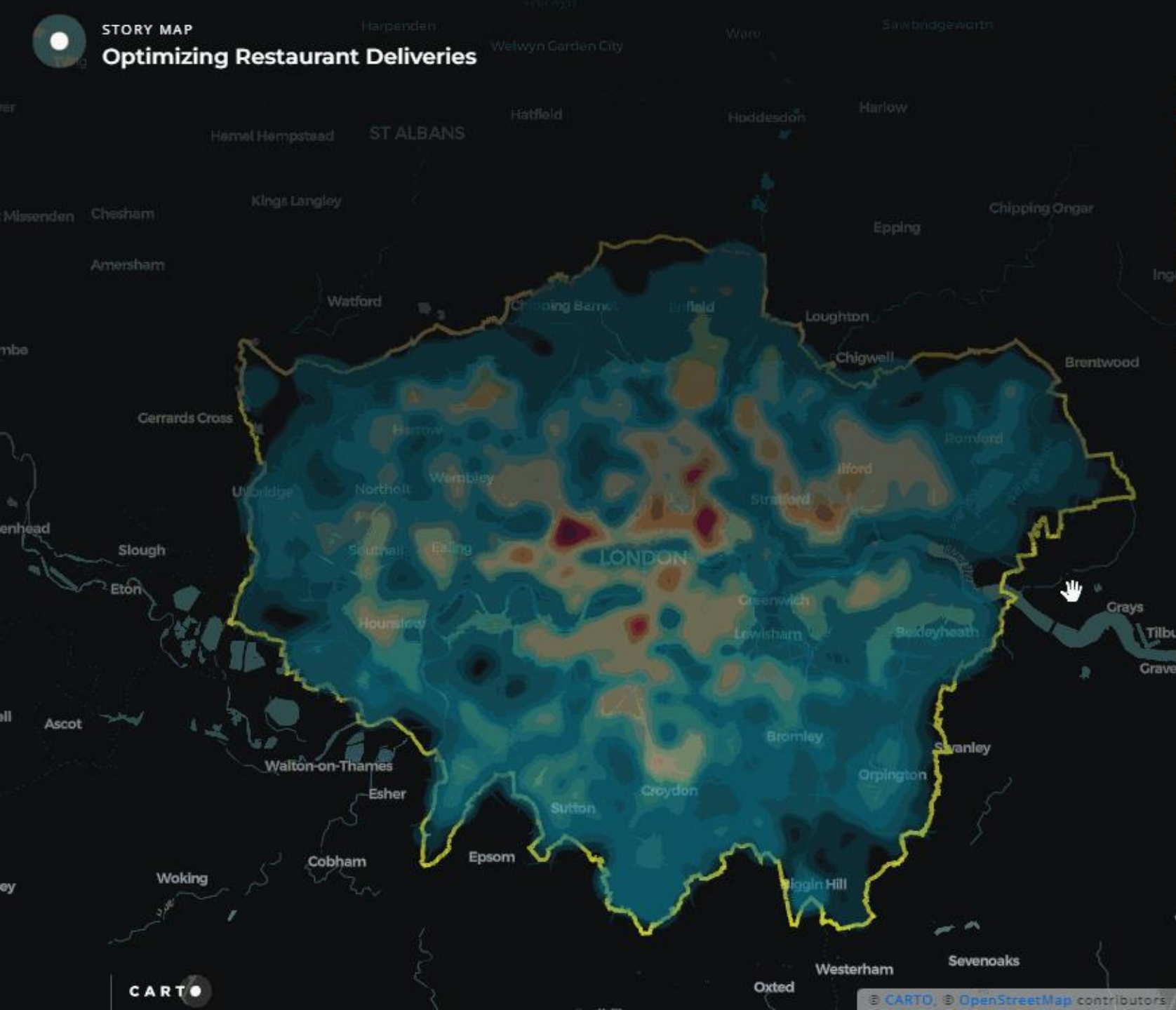


Photo by Rowan Freeman on Unsplash

MAP 2 OF 7

## The spatial side of takeaways

High performing delivery businesses understand the relationships between where people live (or sometimes work or socialize) and where they're ordering food from. Understanding spatial patterns of demographics is key to this.

Businesses are not just looking at the situation now, but constantly seeking out new markets. Identifying ideal market conditions in new cities is key to covering a greater percentage of the population.



Next →

CARTO

CARTO, OpenStreetMap contributors

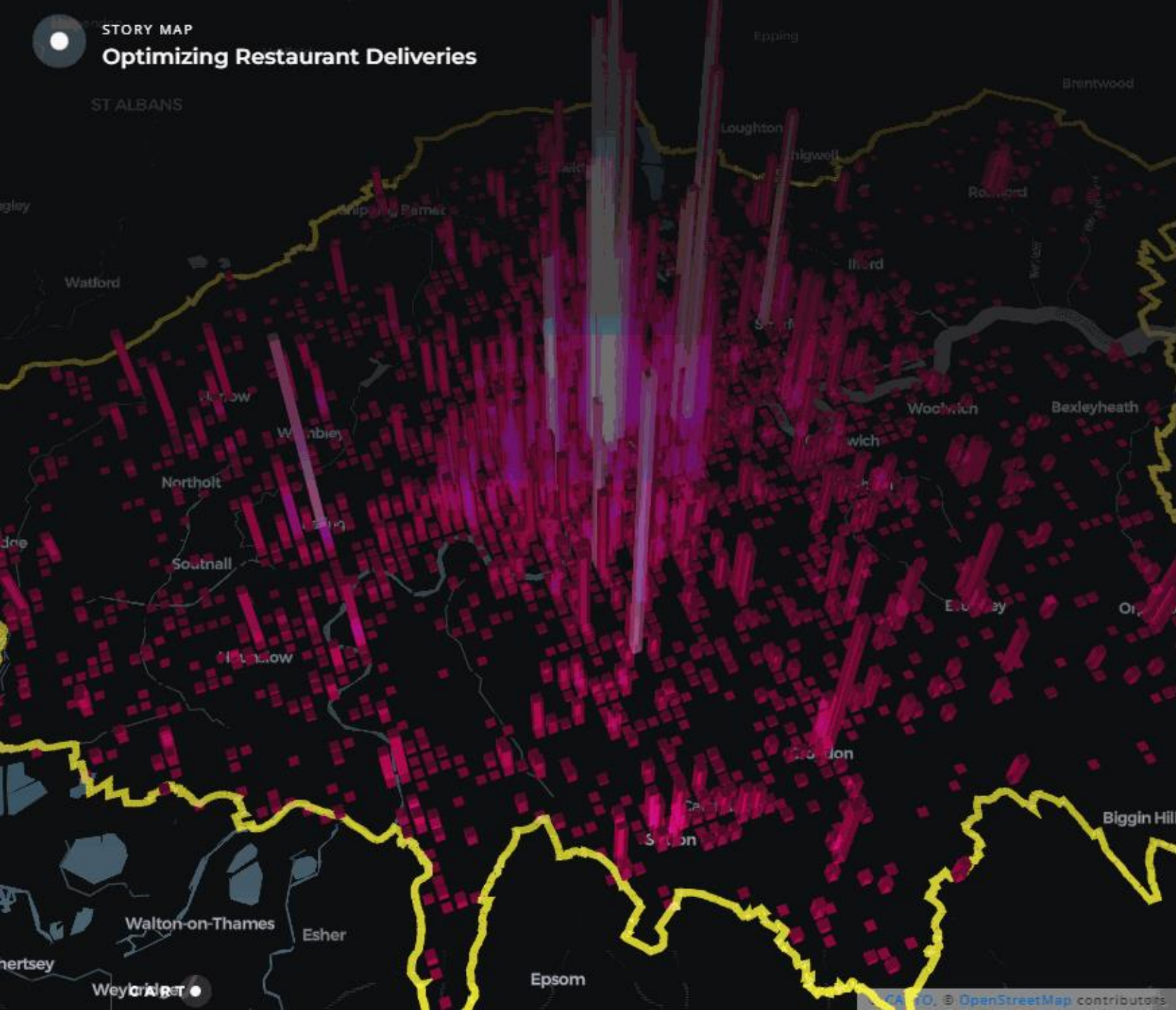


Photo by Bruno Martins on Unsplash

MAP 3 OF 7

## The geography of restaurants

In addition to demographics, the other key consideration is restaurants. According to OpenStreetMap data, there are around 13,000 catering establishments in the Greater London area - that's about 1 restaurant per 100 residents!

Key to successful delivery operations is understanding not just the number of restaurants and customers, but spatial patterns of the type of restaurants and customers. Do you have the right



Next →



Photo by Emmy Smith on Unsplash

MAP 4 OF 7

## Hungry customers don't stay happy customers

One of the things that matters most to the customer is when they get their food. So much of this is dependent on the routing decisions used to connect restaurants to their customers.

In London the average delivery time is 25.4 minutes. Improving on this drives customer satisfaction, and is something businesses should be striving to do - but how?



Next →

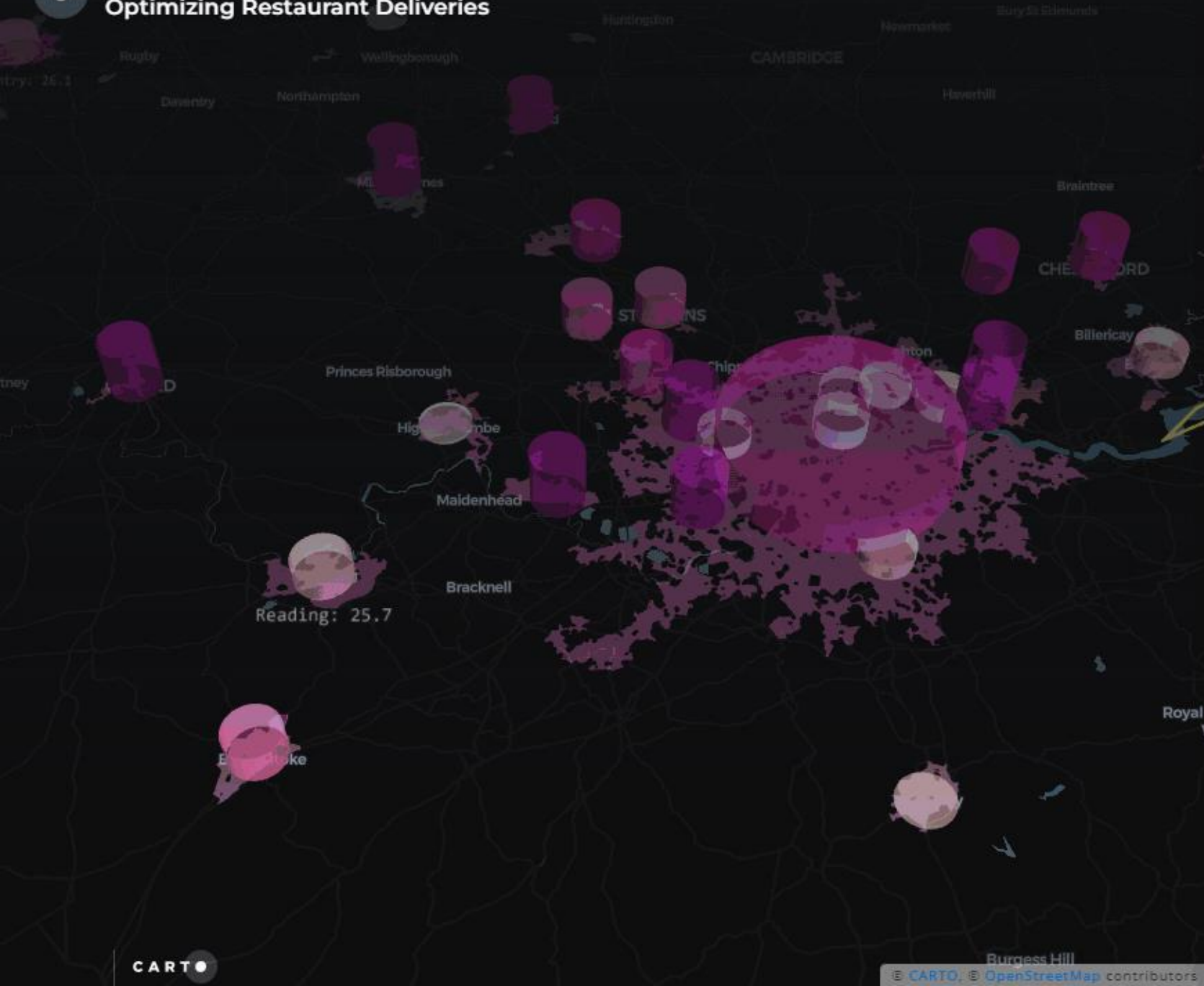


Photo by Egor Myznik on Unsplash

MAP 5 OF 7

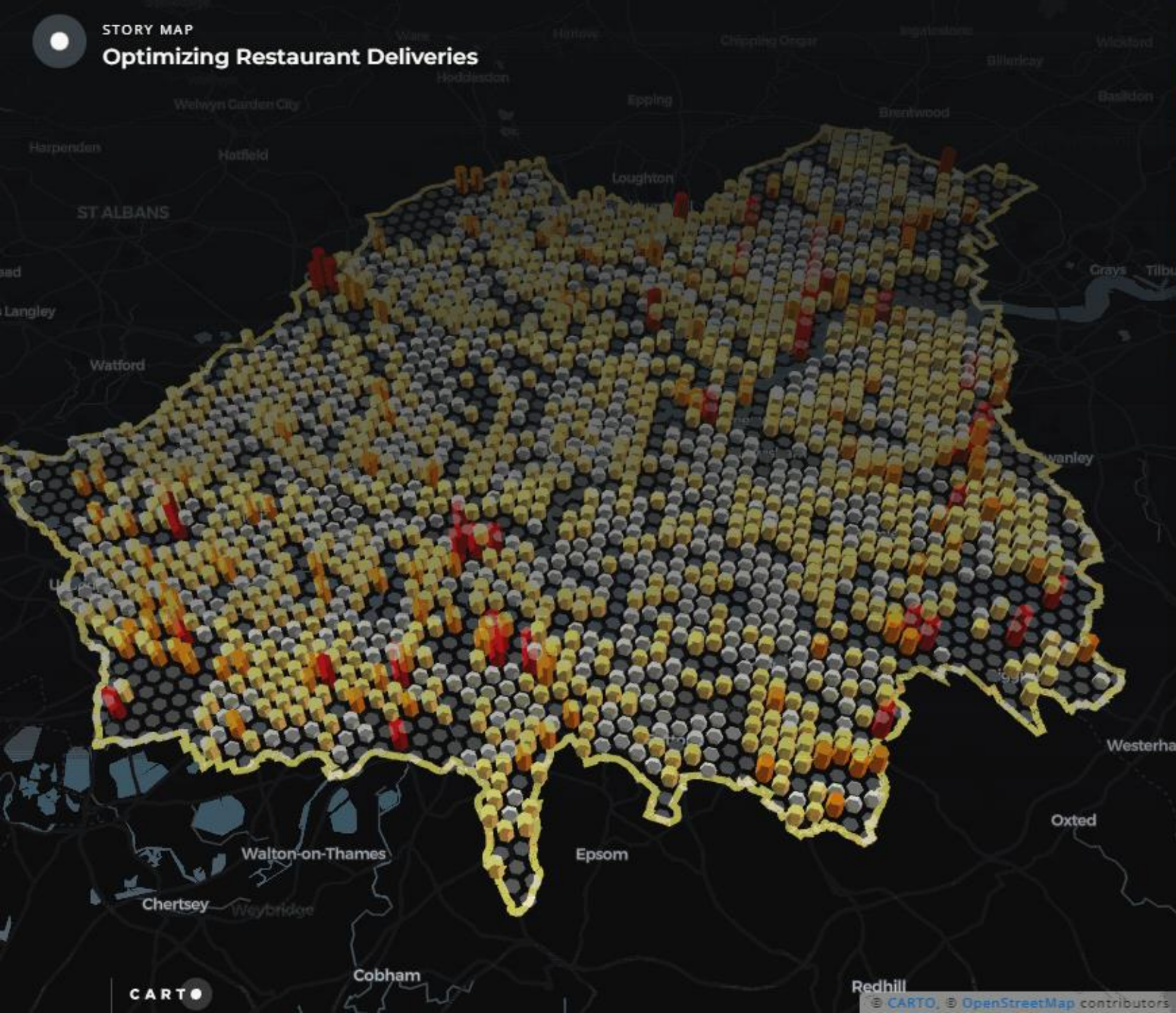
## What is the geography of my business?

One of the biggest blockers driving spatial-based insights is looking at the wrong geography. An example of this would be analyzing order delivery times at a city level, like this.

Imagine you're a takeaway delivery driver. Now imagine all the things that could delay your order. Your local area could be extra busy. A restaurant could complete it late. There could be extra traffic, or a road closure. You could struggle to find your customer's address. But your colleague in the next



Next



About



Share

Photo by Richard Bell on Unsplash

MAP 6 OF 7

## A fresh perspective

If we aggregate and analyze data to a spatial index like H3, then we can start to see a story. We can see areas of spatial outliers and patterns, with red and orange points on the map have higher delivery time.

This is the first step in a delivery company's spatial strategy! They can use this intelligence to begin to drive decision making, such as where to engage more restaurants or delivery drivers.



Next →

# Customers are successful today

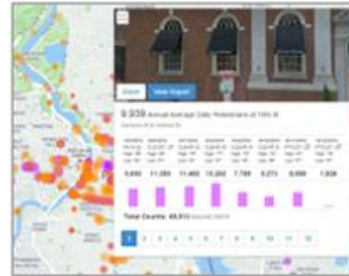
Across all industries + common patterns

## Fraud and Abuse



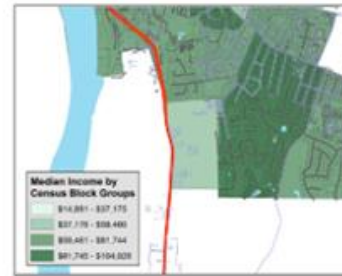
Detect patterns of fraud and collusion (e.g. claims fraud, credit card fraud)

## Retail



Site selection, urban planning, foot traffic analysis

## Financial Services



Economic distribution, loan risk analysis, predicting sales at retail investments

## Healthcare



Identifying disease epicenters, environmental impact on health, planning care

## Disaster Recovery



Flood surveys, earthquake mapping, response planning

## Defense and Intel



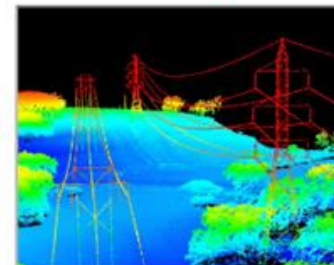
Reconnaissance, threat detection, damage assessment

## Infrastructure



Transportation planning, agriculture management, housing development

## Energy



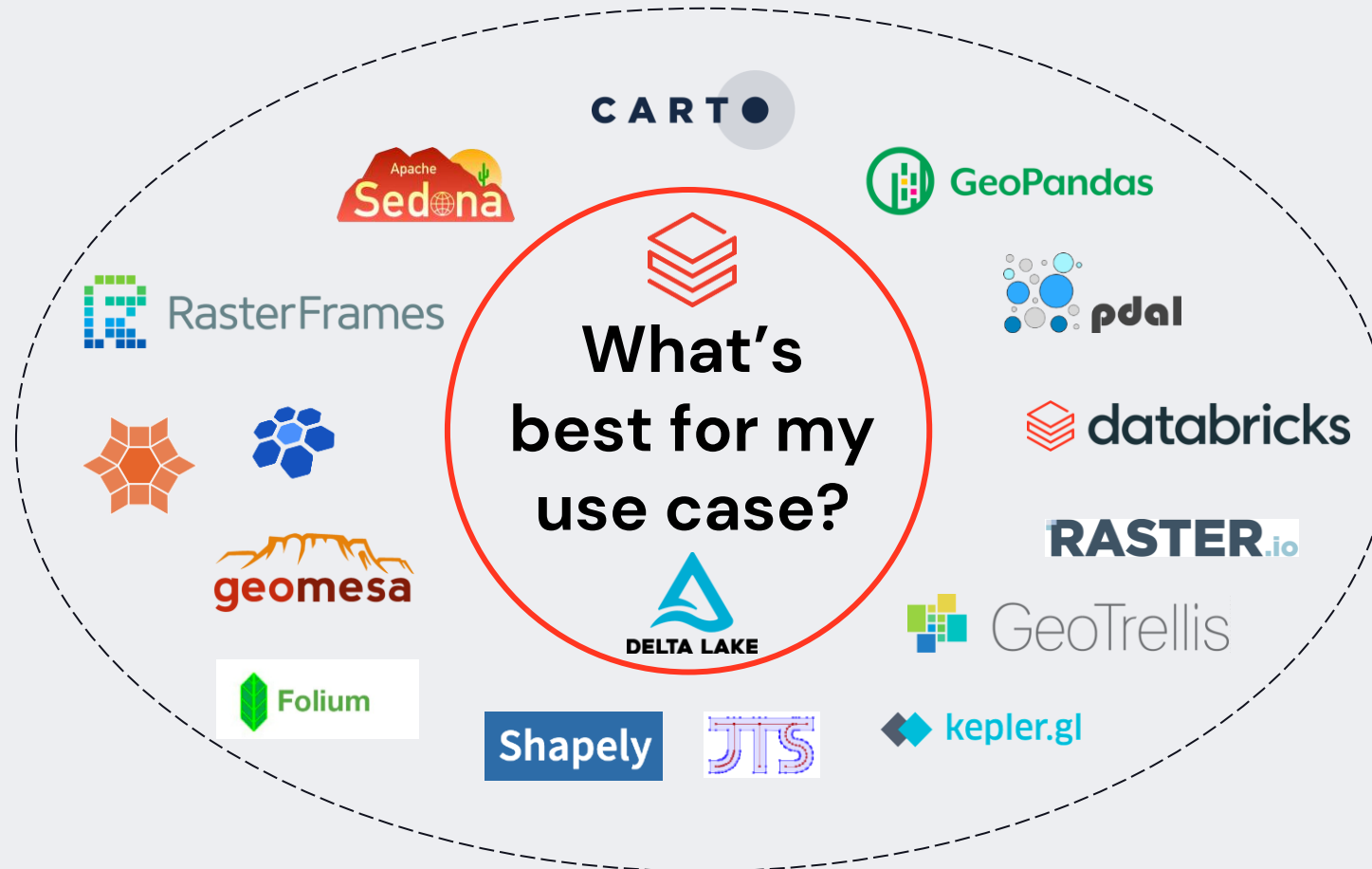
Climate change analysis, energy asset inspection, oil discovery





# Broad Ecosystem

Choose your own adventure for geospatial processing



# Partnering with CARTO

Get more value out of your location data

- **Fast Spatial Analysis.**  
Familiarity with Spatial SQL
- **Fast visualization of large datasets.**  
Large-scale visualization capabilities
- **Great Data sharing capabilities.**  
Integration with Data Observatory



# Product Roadmap

## Best in-class geospatial capabilities

How do we decide what to build?

- Customer feedback and signals
- Partner alignment + integration
- Simplicity and performance → End to End



# Native H3 support

## Supercharge geospatial processing

### Why H3?

- Broad use for spatially aggregating or joining
  - Where to locate a hospital or store?
  - Where will forest fires occur?
  - And much more!

### How is H3 beneficial?

- **Efficient** storage + **Fast** spatial joins + **Easy** to visualize

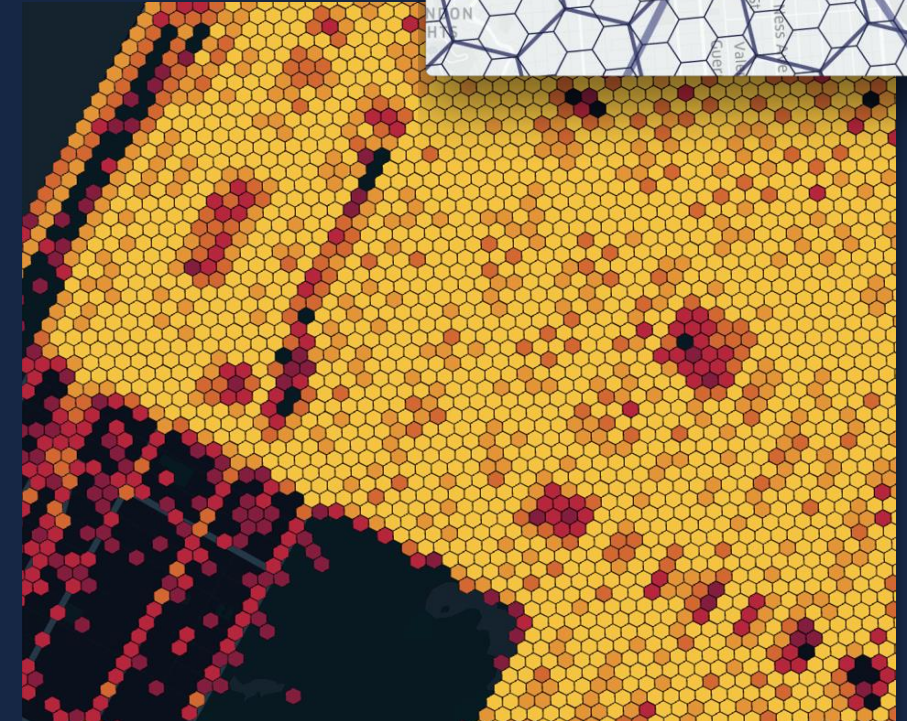
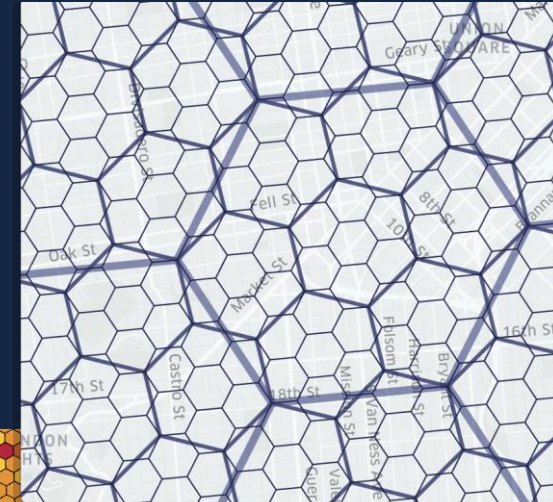
### When will it be available?

- Preview is underway
- GA planned for next quarter

### Will this work in CARTO Builder?

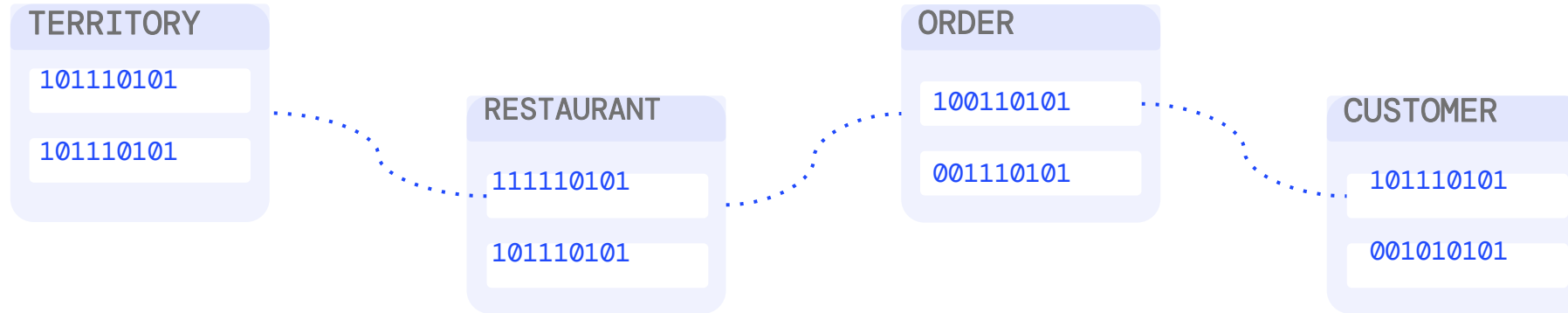
- Yes + will launch with integration with CARTO

H3 cells, across multiple resolutions.  
Supported from global scale to street level.

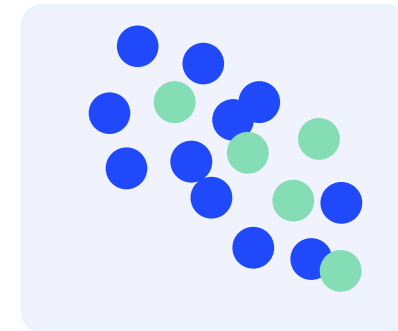
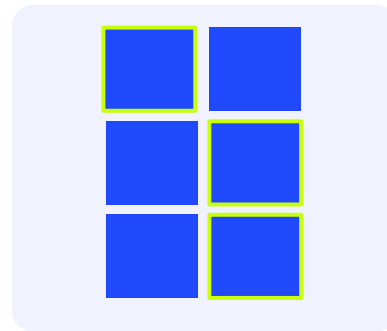
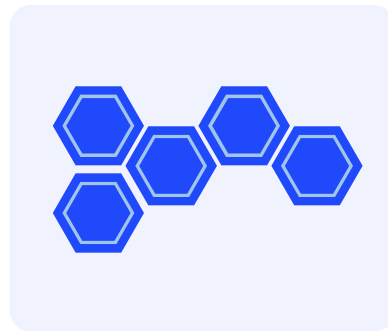
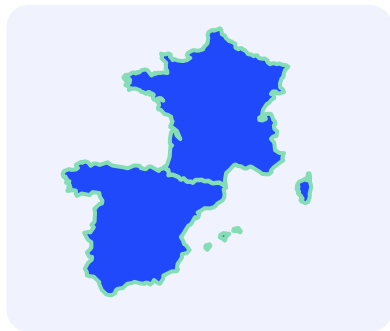


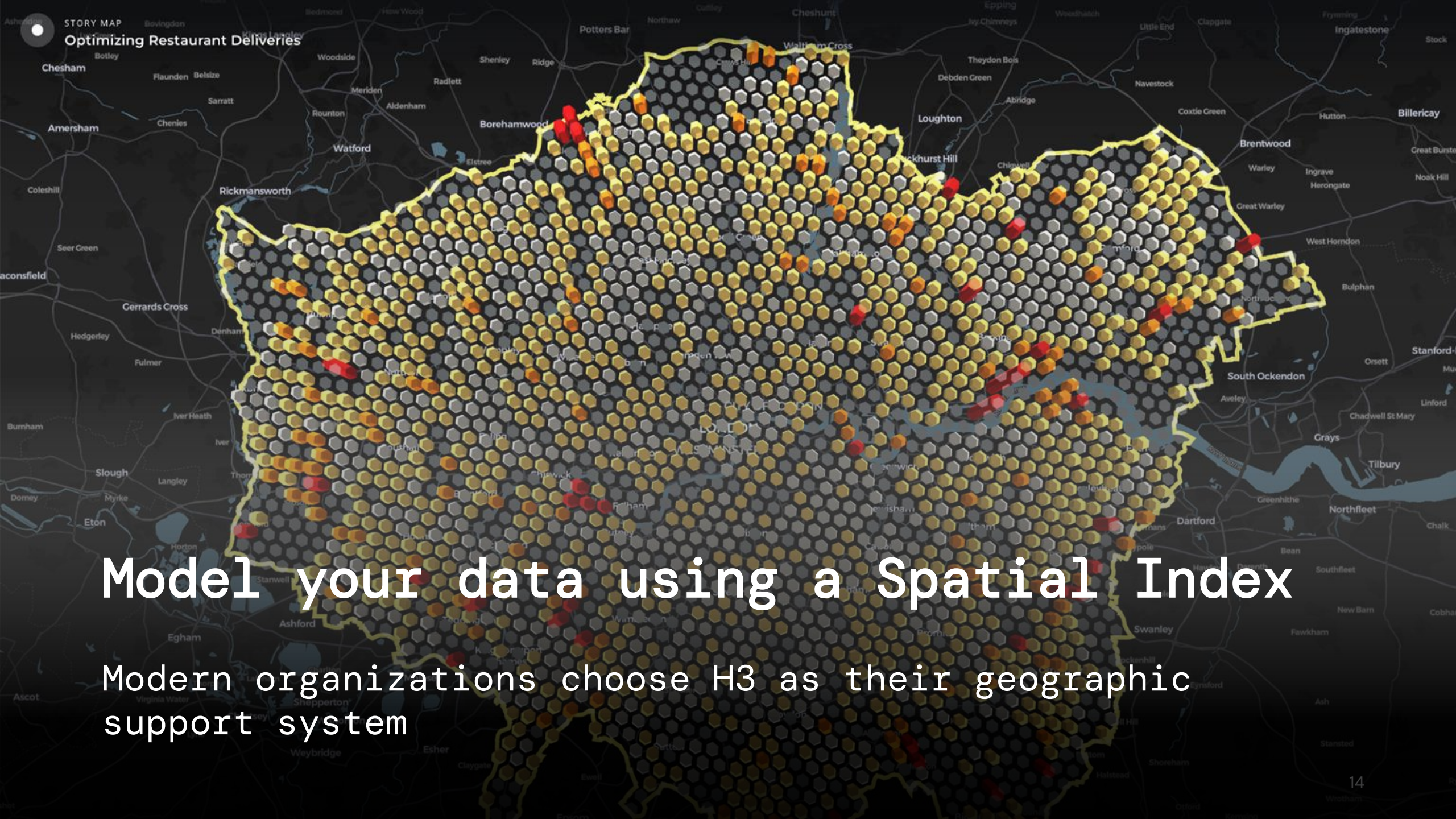
Rideshare pick-up locations in New York City  
in a Databricks Notebook.

# How you model your data?



What geographic support you use matters a lot



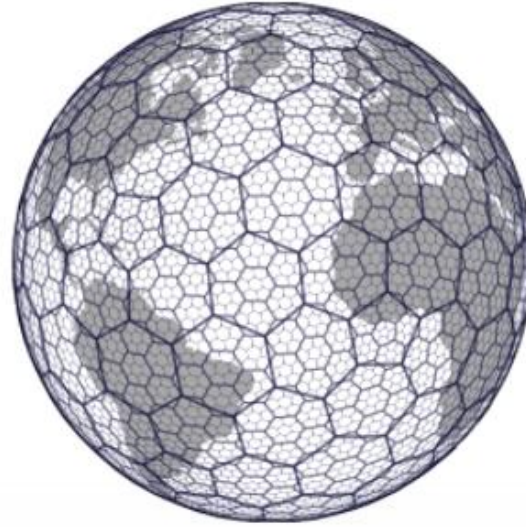


# Model your data using a Spatial Index

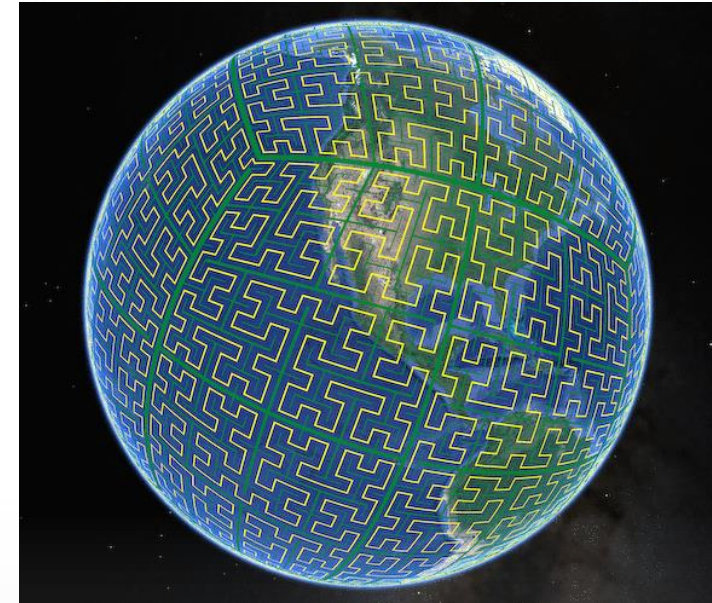
Modern organizations choose H3 as their geographic support system

(0,0)	(1,0)	(2,0)	(3,0)	(4,0)	(5,0)	(6,0)	(7,0)
(0,1)	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)	(7,1)
(0,2)	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)	(7,2)
(0,3)	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)	(7,3)
(0,4)	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)	(7,4)
(0,5)	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)	(7,5)
(0,6)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)	(7,6)
(0,7)	(1,7)	(2,7)	(3,7)	(4,7)	(5,7)	(6,7)	(7,7)

Quadkey ([source](#))



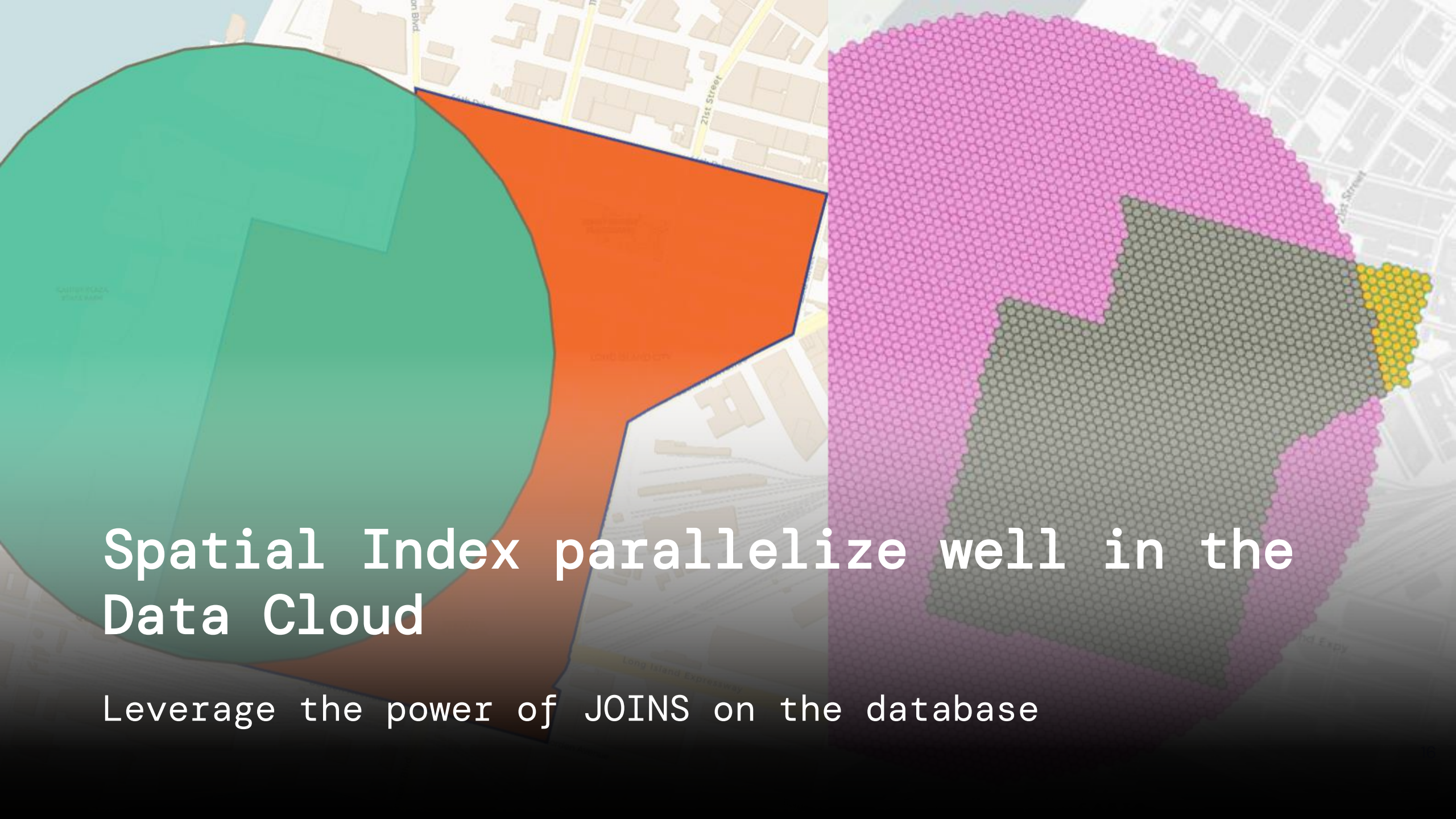
Uber's H3 ([source](#))



S2 ([source](#))

# Geospatial Hierarchical Indexes

Different strategies to partition the space into discrete grids



# Spatial Index parallelize well in the Data Cloud

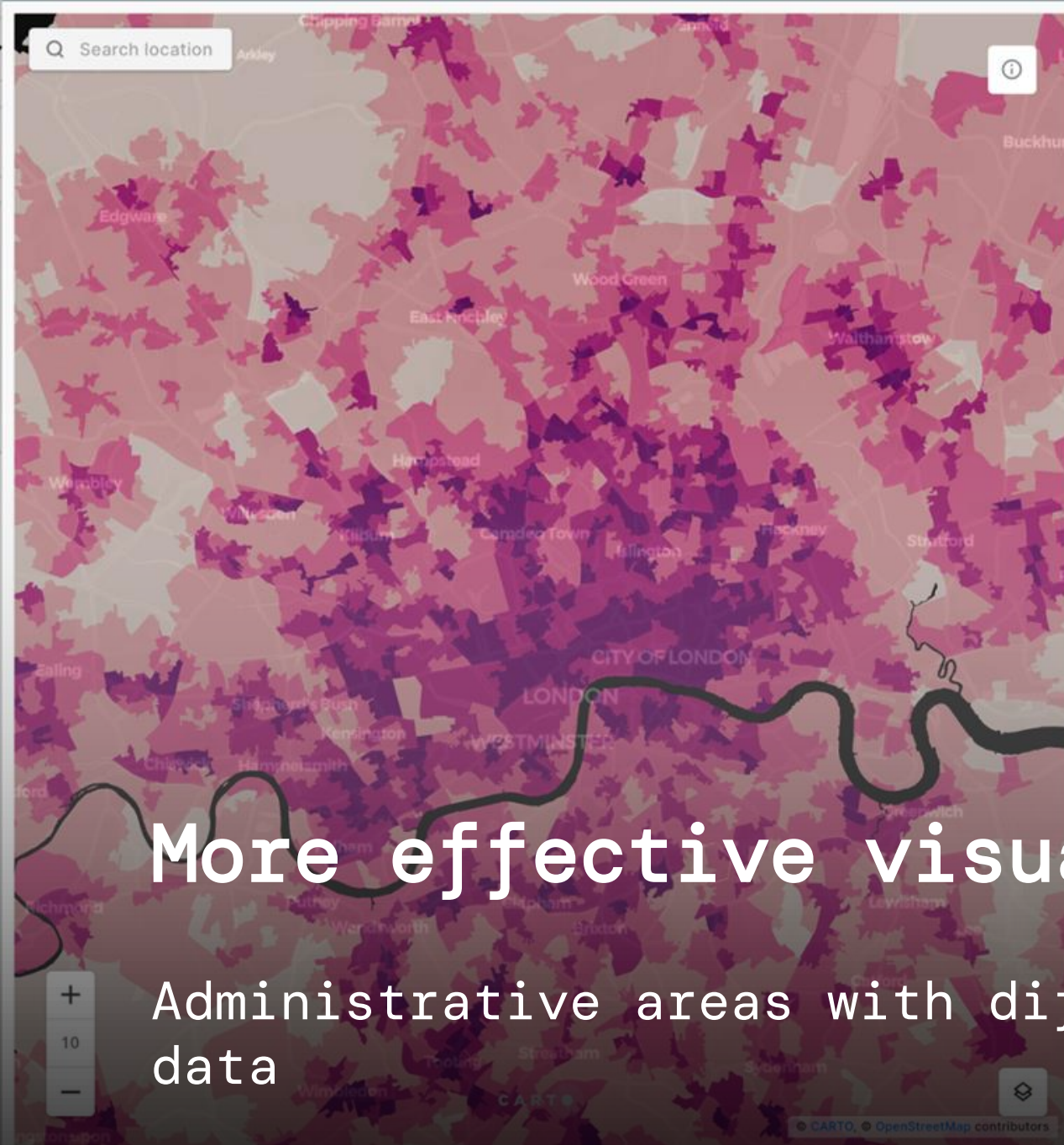
Leverage the power of JOINS on the database





JOIN with 3rd party data in a common  
geographic support system

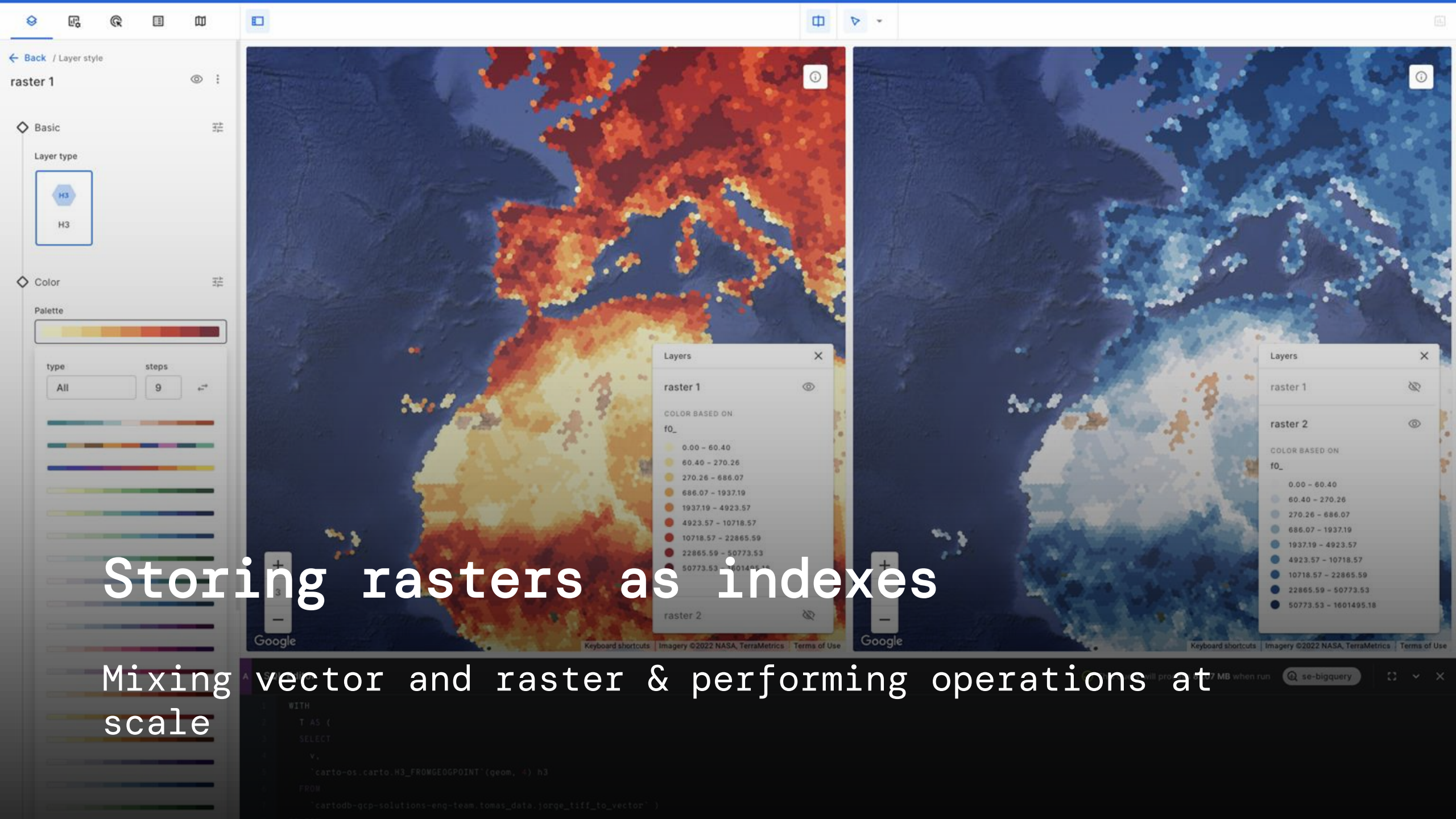
CARTO Spatial Features - Demographics, urbanity,  
climate, elevation...



# More effective visualizations

Administrative areas with different sizes can distort data



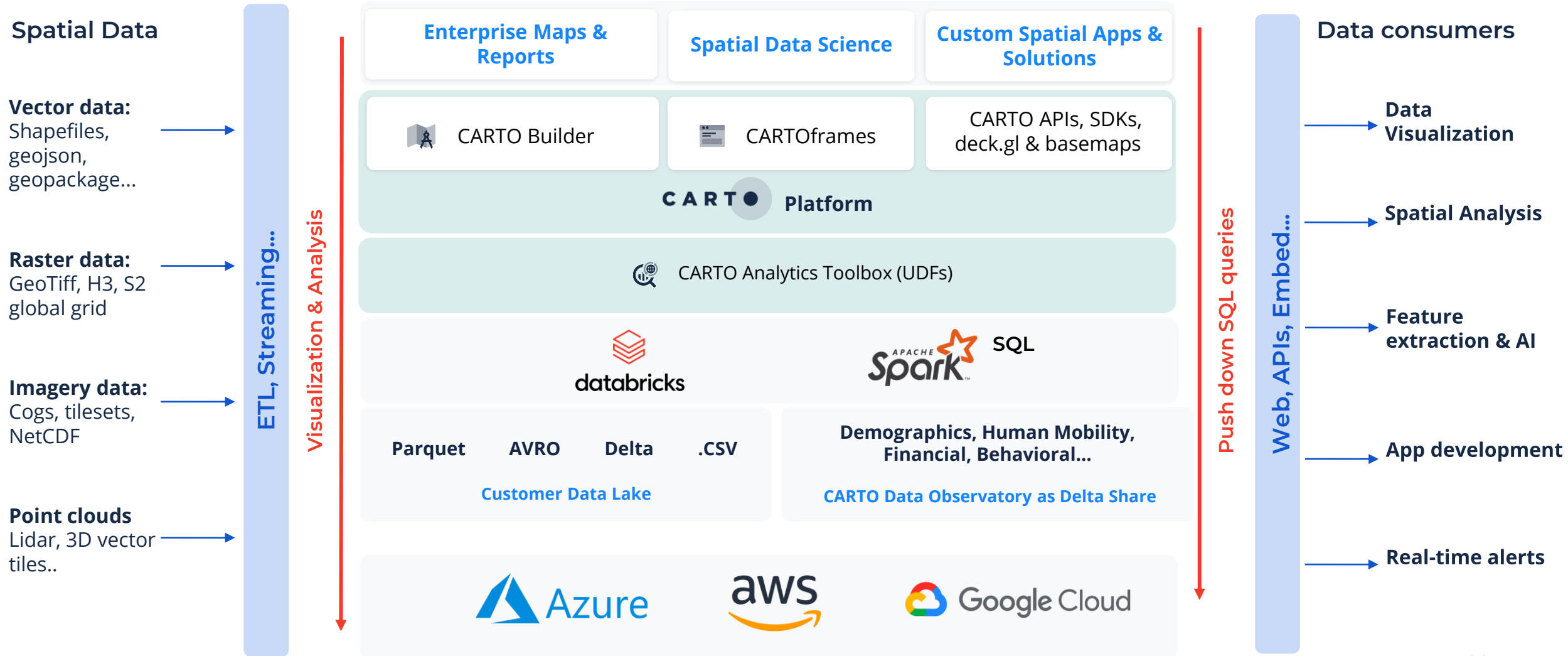


Storing rasters as indexes

Mixing vector and raster & performing operations at scale

```
WITH  
T AS (  
  SELECT  
    v,  
    'carto-os.carto.H3_FROMGEOPOINT'(geom, 4) h3  
  FROM  
    'cartodb-gcp-solutions-eng-team.tomas_data.jorge_tiff_to_vector' )
```

# CARTO + databricks



# Demo: How to build an Spatial Application?

Detached

Visualize and analyze a table of 200M points is imposible using standard tools

Cmd 3

```
1 select count(*) from customers_200M;
```

SQL

▶ (2) Spark Jobs

Table

Data Profile

	count(1) ▲
1	195723428

Showing all 1 rows.



Command took 0.43 seconds -- by alberto+databricks-temp@cartodb.com at 21/6/2022, 20:08:30 on



Detached



Markdown

## Analyzing with h3

Instead we can analyze our customers with h3, let's change the spatial support from lat/lng to h3

Cmd 5

```
1 CREATE TABLE customers_h3 as
2   SELECT h3_longlatash3(long, lat, 8) as h3, count(*) as total_customers
3   FROM customers_200M
4   GROUP BY h3
```

▶ (5) Spark Jobs


Query returned no results


Command took 55.46 seconds -- by alberto+databricks-temp@cartodb.com at 22/6/2022, 11:52:26 on H3 Preview




# Welcome to CARTO

Get started by creating stunning maps and bringing your spatial data to life! Enrich your analysis with our Data Observatory, build powerful apps, and more — you'll be soon turning your location data into powerful insights.

 **Getting started**  
Bring your data and create your first map

 **Access tutorials**  
Solve interesting use cases following our guides

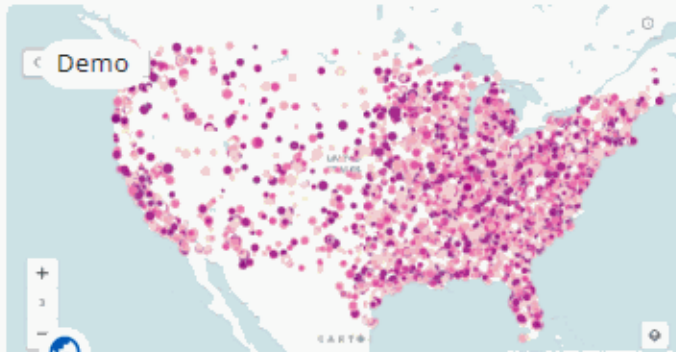
 **Check our documentation**  
Unlock the full potential of CARTO technology

## Start with your spatial analysis

[View all](#) →



Find the best place to create a store near the customers



Monitor retail store performance



Analyzing Airbnb ratings in Los Angeles



Create map

### Data Explorer

#### Connections

- databricks
  - cluster
    - carto\_dev\_data
    - data\_ai\_summit
      - avg\_eta\_per\_city\_buffer...
      - avg\_eta\_per\_city\_labels...
      - customers\_200m
      - customers\_h3**
      - customers\_h3\_enriched
      - customers\_h3\_enriched...
      - daily\_results\_length\_si...
      - daily\_results\_simplified...
      - daily\_simulation\_04022...
      - gb\_border
      - glovo\_simulation\_no\_ge...
      - london\_boundary
      - london\_h3\_kpi\_census\_2
      - osm\_poi\_restaurant\_1

Table · Databricks

### customers\_h3

Where: databricks > cluster > data\_ai\_summit [Copy qualified name](#)

Table Size

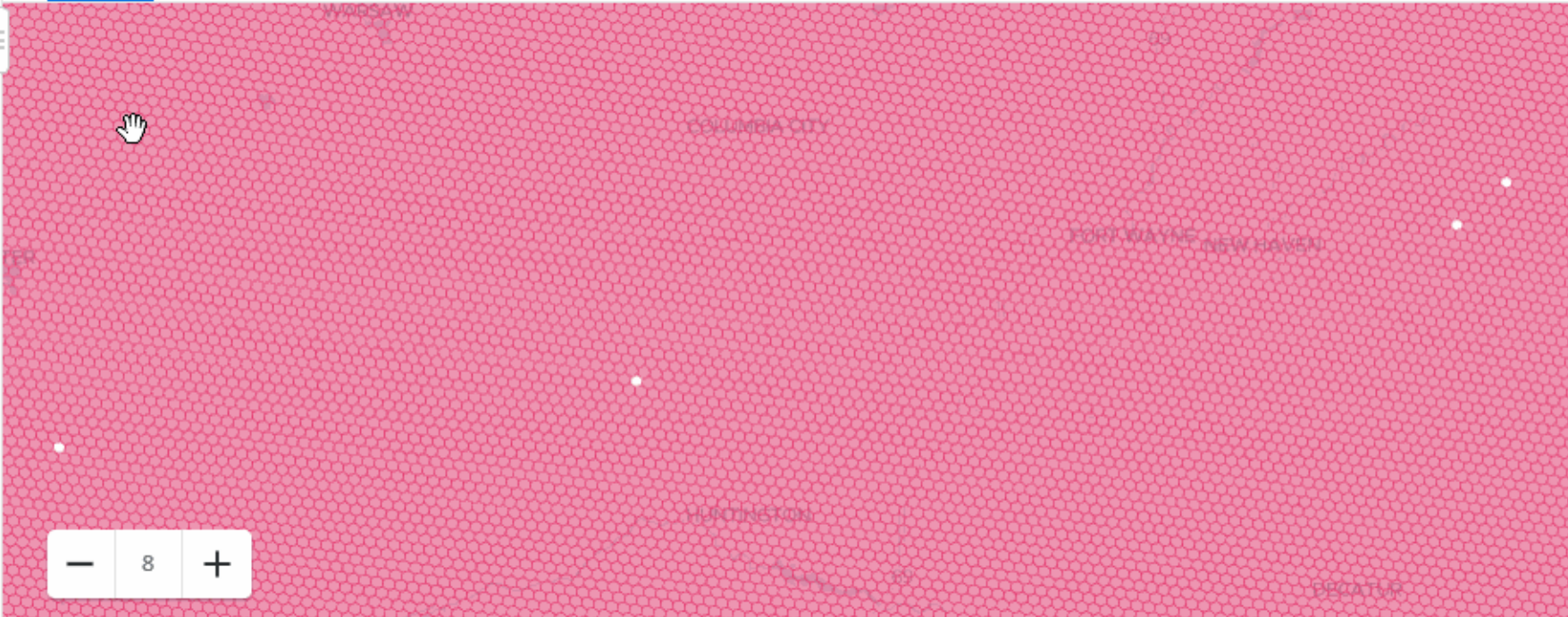
41.78 MB

Number of Rows

6.3M

Map

Data preview



Data Observatory

Detached



Markdown ▾

We're going to enrich our customers dataset with some extra data from CARTO Data Observatory. As we're working with the same geospatial support everything is quite easy, just a simple join

Cmd 7

```
1 CREATE TABLE customers_h3_enriched as
2   SELECT c.*,sf.population, male, female, food_drink
3   FROM customers_h3 c
4   INNER JOIN `data_observatory`.derived_spatialfeatures_usa_h3int_res8_v1_yearly_v2 sf ON
   c.h3 = sf.h3
```

▶ (6) Spark Jobs

Query returned no results

Command took 7.87 seconds -- by alberto+databricks-temp@cartodb.com at 22/6/2022, 11:52:31 on H3 Preview

Cmd 8

```
1 OPTIMIZE customers_h3_enriched ZORDER BY h3
```

▶ (11) Spark Jobs



### Data Explorer

Create map

#### Connections

- databricks
  - cluster
    - carto\_dev\_data
    - data\_ai\_summit
      - avg\_eta\_per\_city\_buffer...
      - avg\_eta\_per\_city\_labels...
      - customers\_200m
      - customers\_h3
      - customers\_h3\_enriched**
      - customers\_h3\_enriched...
      - daily\_results\_length\_si...
      - daily\_results\_simplified...
      - daily\_simulation\_04022...
      - gb\_border
      - glovo\_simulation\_no\_ge...
      - london\_boundary
      - london\_h3\_kpi\_census\_2
      - osm\_poi\_restaurant\_1

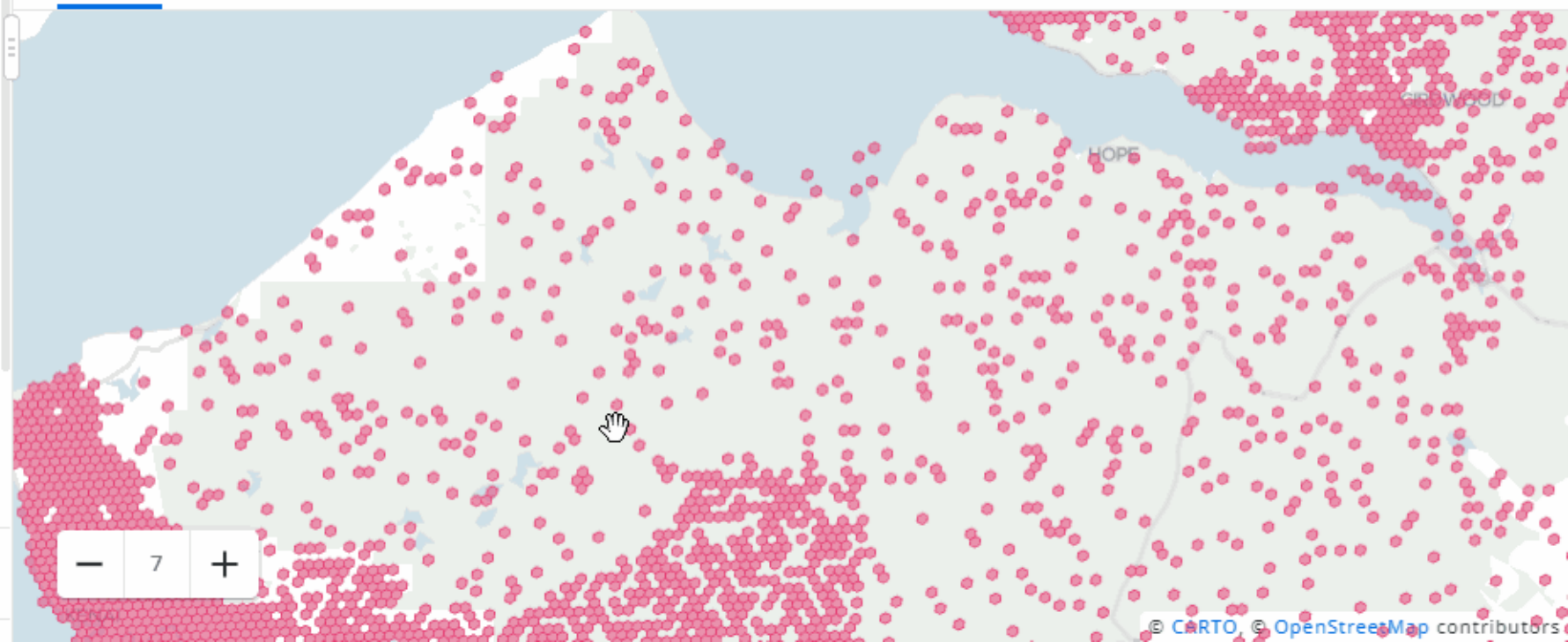
Table · Databricks

### customers\_h3\_enriched

Where: databricks > cluster > data\_ai\_summit [Copy qualified name](#)

Table Size	Number of Rows
166.35 MB	6.2M

Map Data preview



Data Observatory



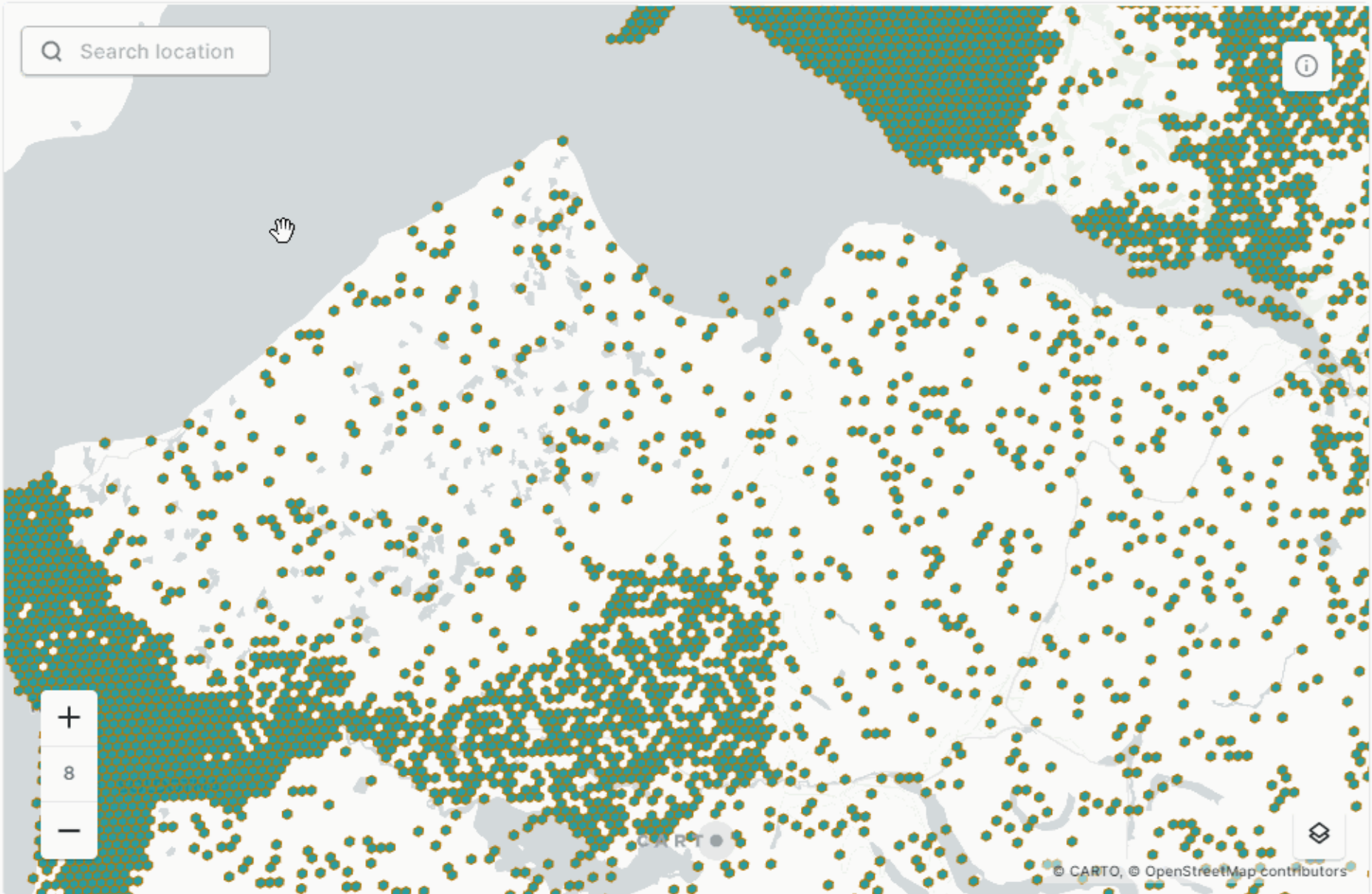
Layers

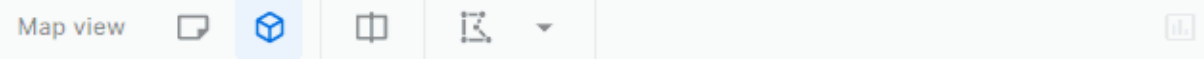
A H3  
Layer 1

SOURCES (1)

A Table · Databricks  
customers\_h3\_enriched  
6,224,143 RECORDS

+ Add source from...



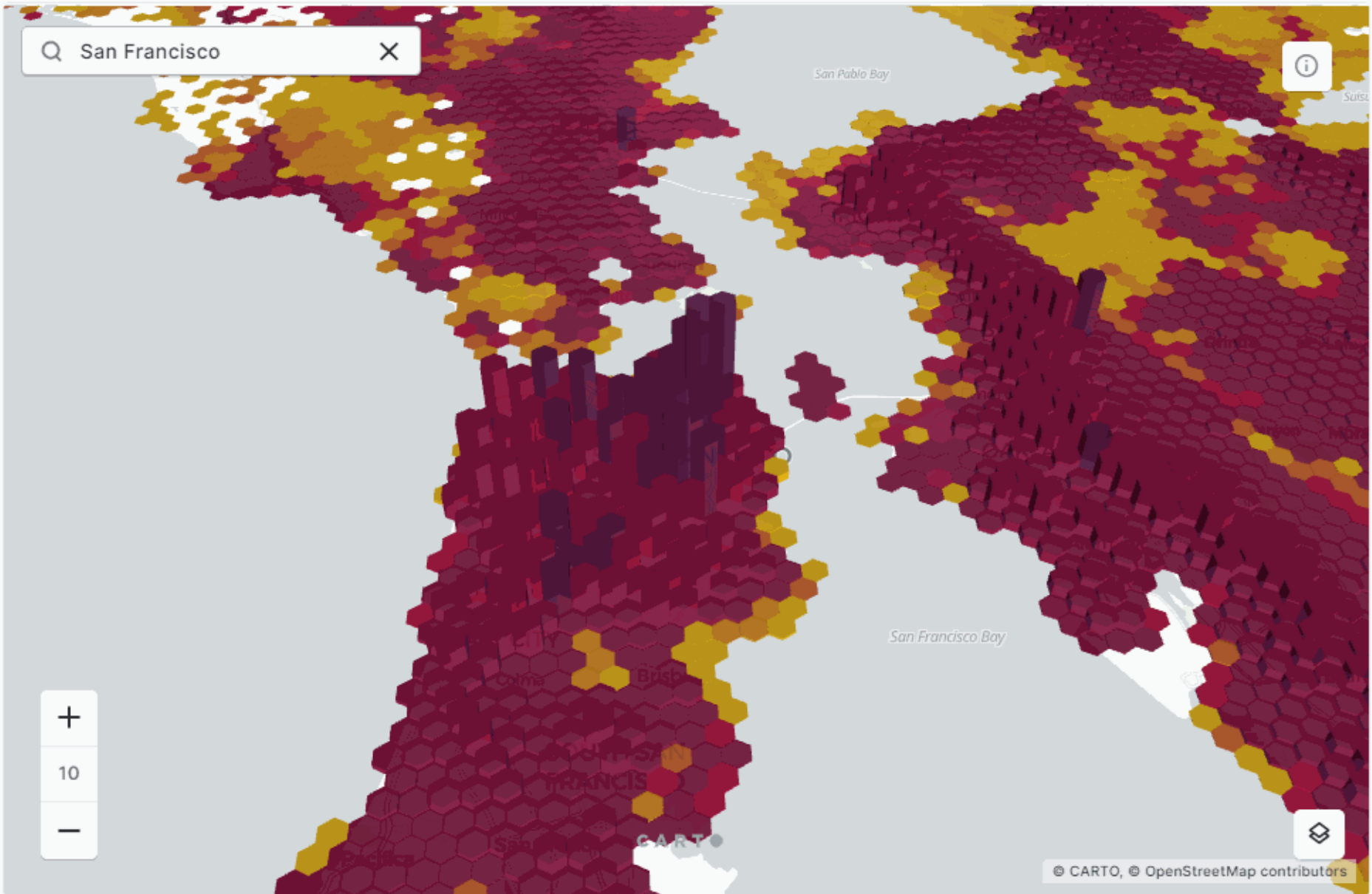


Widgets



No widget added yet

New widget



### Publish



#### Sharing options



Private (default)

Only you can view and edit



Organization

Visible to 4 users in your organization ai-summit



Public map

Visible to anyone with the public share link



#### Published map settings



Cancel

Save

### Widgets

FORMULA

Population

FORMULA

Customers

HISTOGRAM

Population

HISTOGRAM

Customers

Population

2M

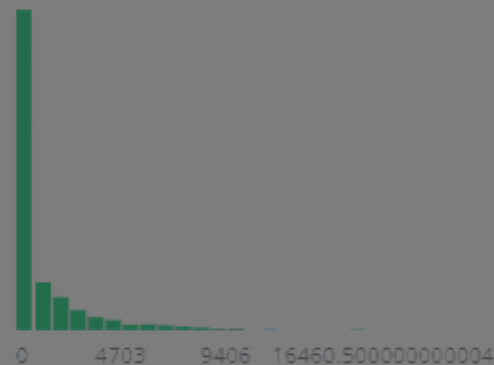
Customers

1278561

Population

All selected

1.5k



Customers

All selected

# How to get access?

Sign up for a free 14 day trial:

<https://carto.com/signup/>

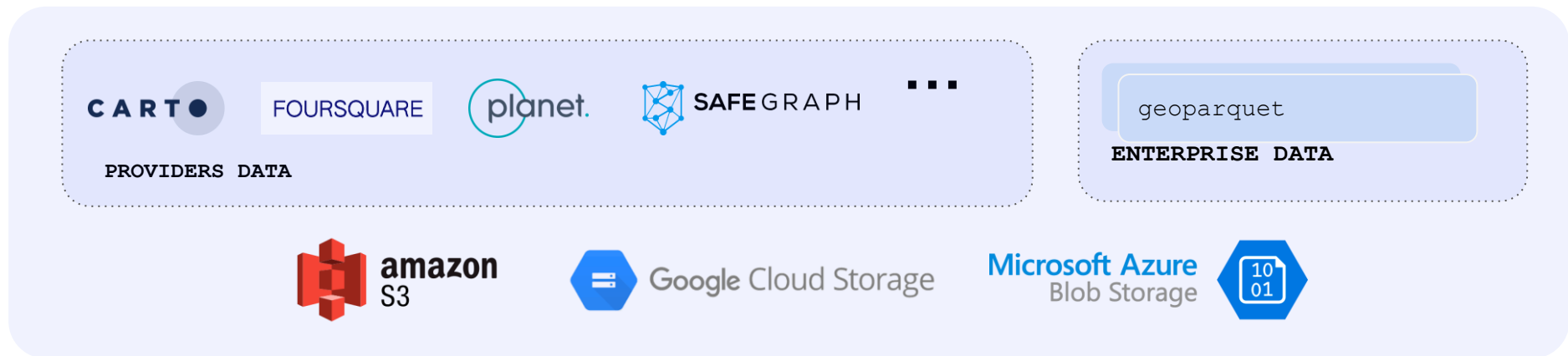


# Geoparquet as a standard storage layer for geo

COMPUTING  
ENGINES &  
LIBRARIES



CLOUD  
STORAGE



# A community driven initiative

Under the umbrella of the Open Geospatial Consortium



Open  
Geospatial  
Consortium.



[github.com/opengeospatial/geoparquet](https://github.com/opengeospatial/geoparquet)

# Stop by to chat:

**CART** ●

booth 633



**databricks**

Find Kent

**DATA+AI**  
**SUMMIT 2022**

# Thank you



**Isabel Pozuelo**

Spatial Analytics expert, CARTO  
ipozuelo@carto.com



**Kent Marten**

Staff Product Manager, Databricks  
kent.marten@databricks.com  
@datakmart