CARTO

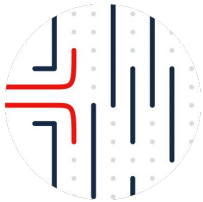
Using CARTO for Spatial Data Science

FOLLOW @CARTO ON TWITTER

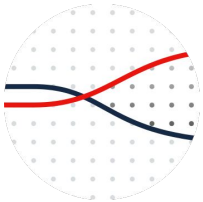
CARTO

Overview

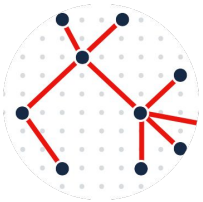
CARTO turns location data into business outcomes in 5 key steps:



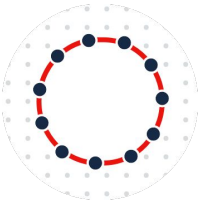
Data Ingestion & Management



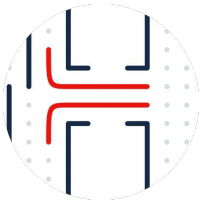
Data Enrichment



Analysis



Solutions & Visualization



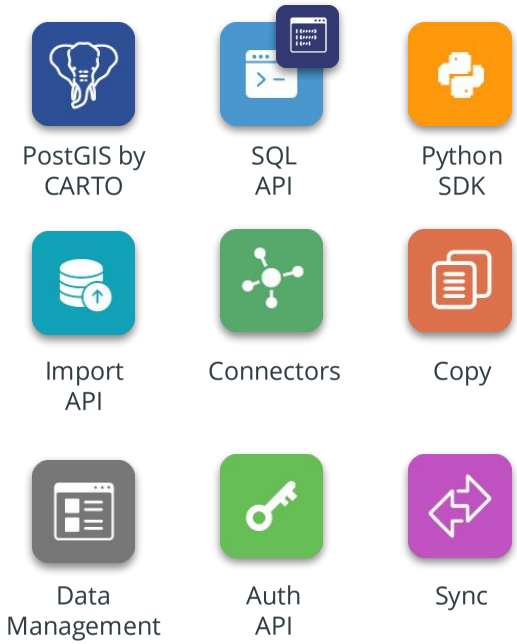
Integration

1. Data Ingestion & Management

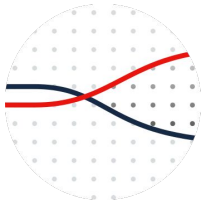
- Spatial database with multiple ways to connect and manipulate your data
- Dynamic data in the cloud
- Fully managed database with automatic backups and regular upgrades
- Enterprise data sharing and access across CARTO

Wide support for geospatial formats (inc. Shapefiles, KML, KMZ, GeoJSON, GPX, OSM, GeoPackage, GDB, CSV, Excel or OpenDocument).

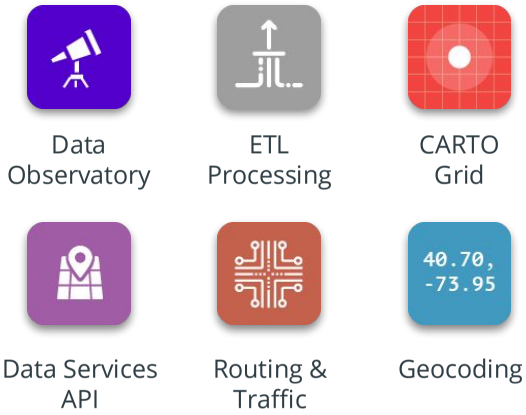
Plug ready database connectors (ArcGIS Server, DB Connectors via APIs (MySQL, PostgresSQL, Microsoft SQL Server, Hive on request)).

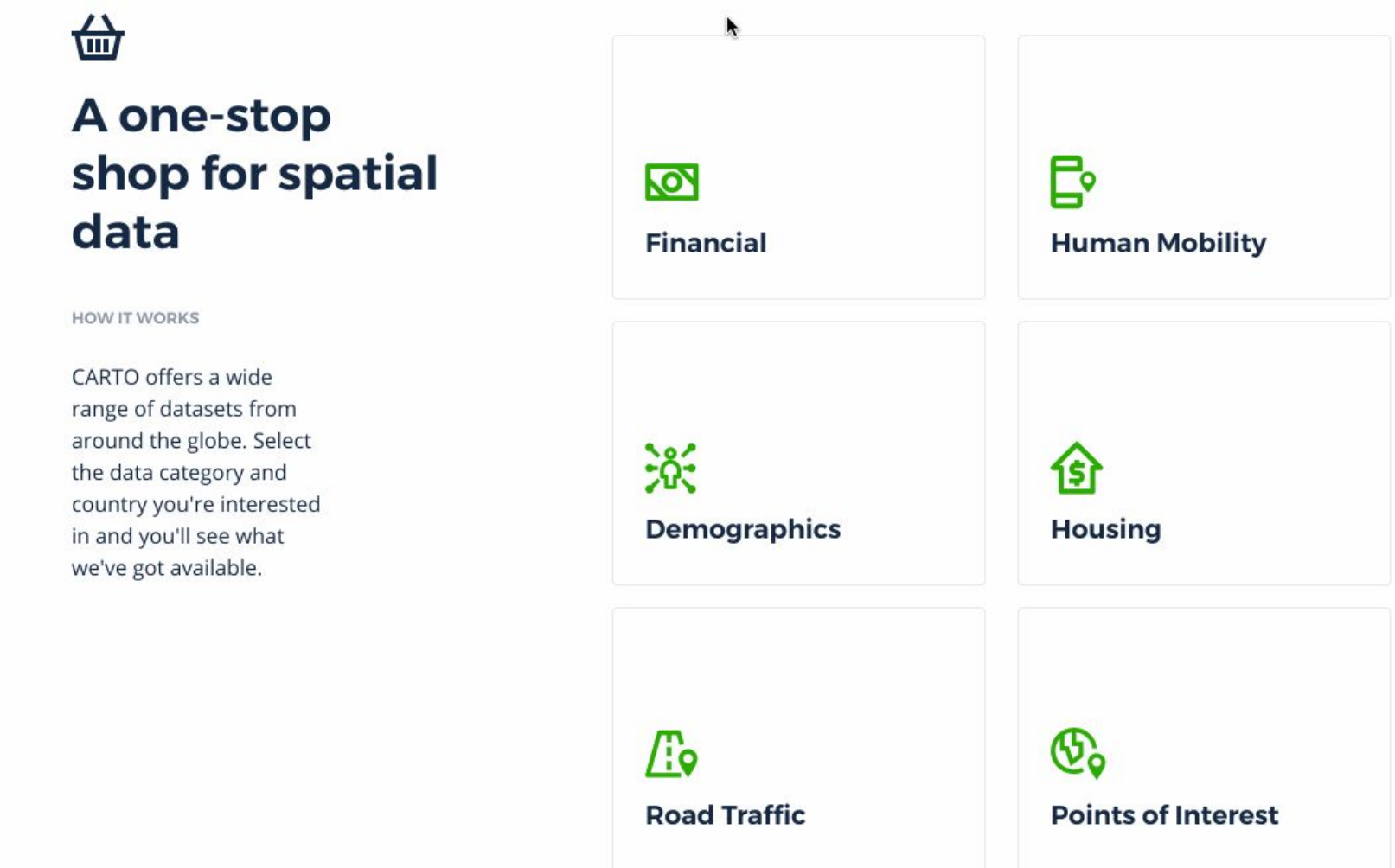


2. Data Enrichment



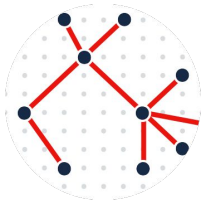
- Save time in gathering spatial data, augmenting your existing data with demographics from across the globe
- Create locations from addresses and understand travel time all from within CARTO
- Develop robust ETL processes and update mechanisms so your data is always enriched
- Premium data to understand and analyze deeper trends and behavior





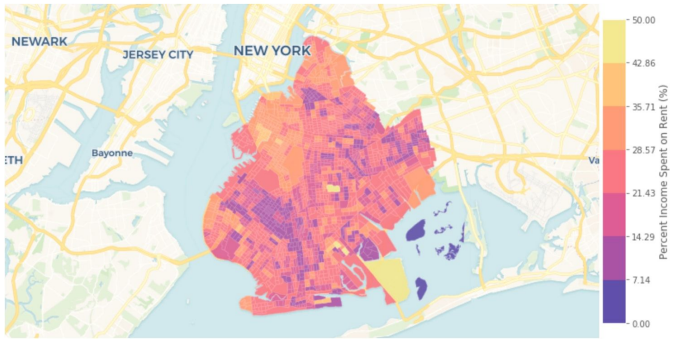
https://carto.com/platform/location-data-streams/

3. Analysis



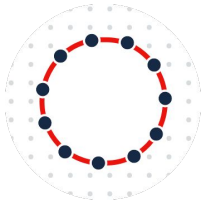
- Bring maps and data into your Data Science workflows and the Python data science ecosystem with **CARTOframes**
- Exploratory analysis embedded in CARTO as simple **SQL calls** for clustering, outliers analysis, time series predictions, and geospatial weighted regression
- Use the power of PostGIS and our APIs to productionalize analysis workflows in your CARTO platform

```
from cartoframes import Layer, styling
cc.map(layers=Layer('demographics_kmeans_out',
                    color={'column': 'percent_income_spent_on_rent_2011_2015',
                           'scheme': styling.sunset(7,'equal'),
                           'legend': 'Percent Income Spent on Rent (%)'}));
```

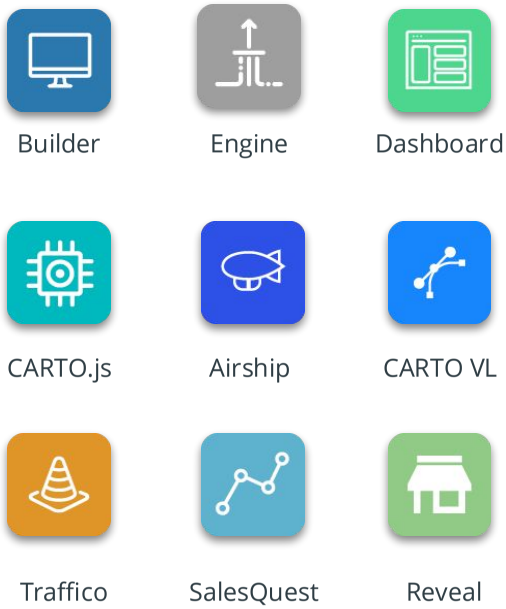


CARTO Frames | Analysis API | SQL API | Python SDK

4. Solutions & Visualization



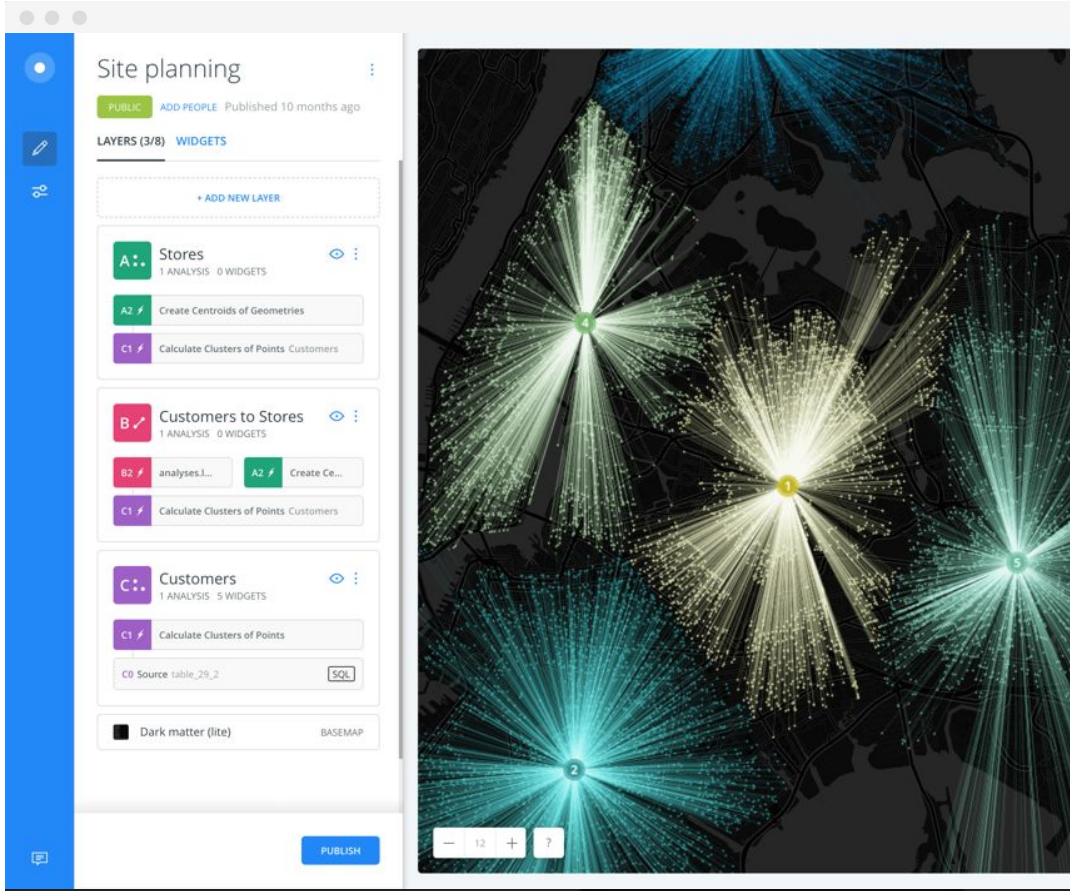
- Develop and build custom applications with a full suite of frontend libraries, using Airship and CARTO VL.
- Work with CARTO's Professional Services and Support team as and when you need it.
- Create lightweight, intuitive dashboards for simple sharing of insights across your organization.



Builder

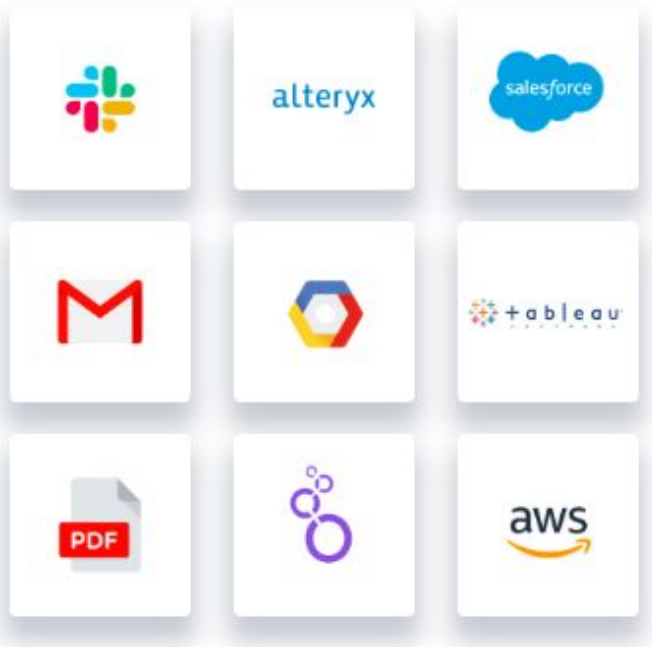
A web-based drag-and-drop tool for analysts and business users to discover and predict key insights from location data

- Performant at scale
- Self-service widget powered dashboard
- Enrich your location data with the Data Observatory



5. Integration

- Using CARTO's APIs and SDKs, connect your analysis into the places that matter most for you and your team.
- Work with our Professional Services team for custom configurations or developments.

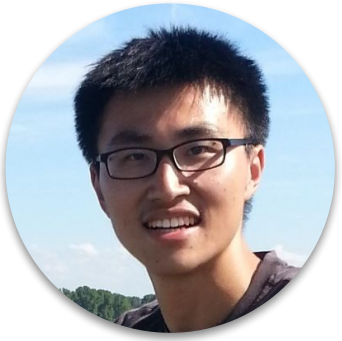


CARTO

Our Data Science Team

Andy Eschbacher

Data Scientist



Dongjie Fan

Data Scientist

Giulia Carella

Data Scientist

Miguel Álvarez García

Data Scientist

Álvaro Arredondo

Data Scientist

Alejandro Polvillo Hall

Data Scientist

What do we do?

- **CARTOframes**

  We help to build CARTOframes, a great visualization and data enrichment tool

- **CARTO Data Observatory**

  We help to scope and manage CARTO Data Observatory, a set of premium datasets from both 3rd party providers and curated by CARTO

- **Spatial Data Science**

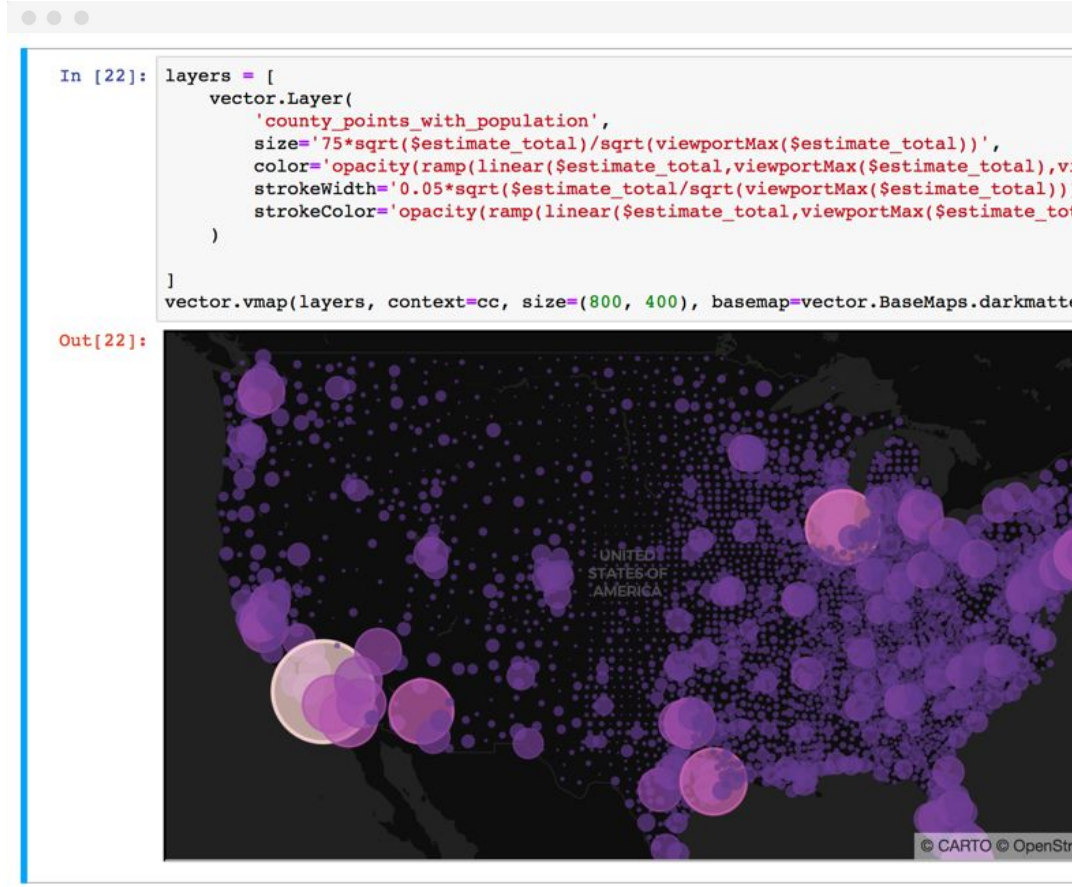
  We develop and deploy different types of analysis

- **Research and Development**

  We research and develop new methods, also in collaborations with research institutions

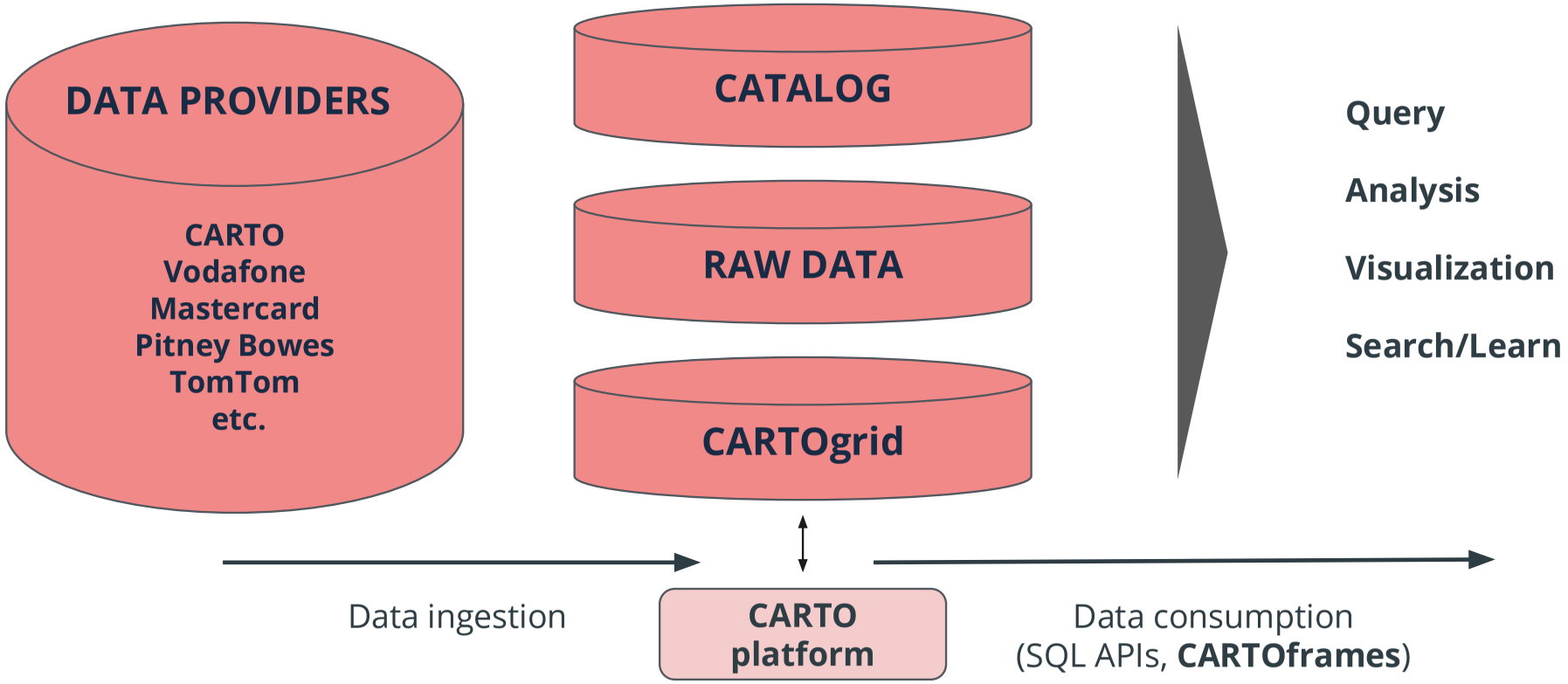
What is CARTOframes?

- Python package
- To be used in Jupyter Notebooks
- Built for Data Scientists
- Part of CARTO Analysis stack
- Connector to our platform
- Viz (for free) using vector maps



https://carto.com/developers/cartoframes/

What is CARTO Data Observatory?



Spatial Data Science

- **Inference and predictive modelling**

  Why we see more orders in some areas than in others? How can we predict the income for new stores?

- **Clustering and similarity analysis**

  How can we quantify similarities between locations?

- **Logistic optimization and territory management**

  What is the optimal set of routes for a fleet of vehicles to traverse in order to deliver to a given set of customers? How do you maintaining equity between sales people/delivery centers?

- **Network analysis**

  Graph analysis (Origin-Destination matrices) from GPS signals

CARTO

Success stories

vodafone

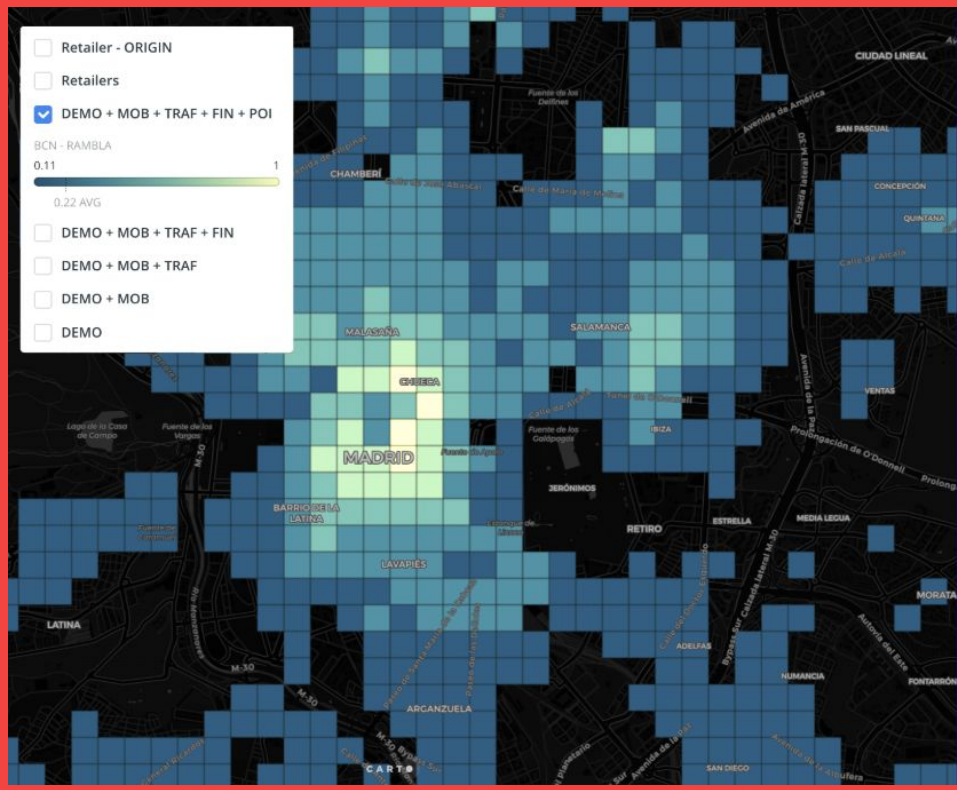
Vodafone Analytics: Twin Areas

As part of Vodafone Analytics, CARTO built a model to identify areas on a 250x250m grid with similar behavior patterns based on Vodafone footfall data.

The similarity score is modeled based on:

- Distance between cells is calculated with a L2 norm on a Principal Component (PC) space.
- Uncertainty due to missing values and dimension of PC space is tackled following an ensemble probabilistic approach.
- Similarity Score = Continuous Rank Probability Skill Score (CRPSS)

By enriching the data with other sources this model can be used for Site planning, Investment Analysis, etc.

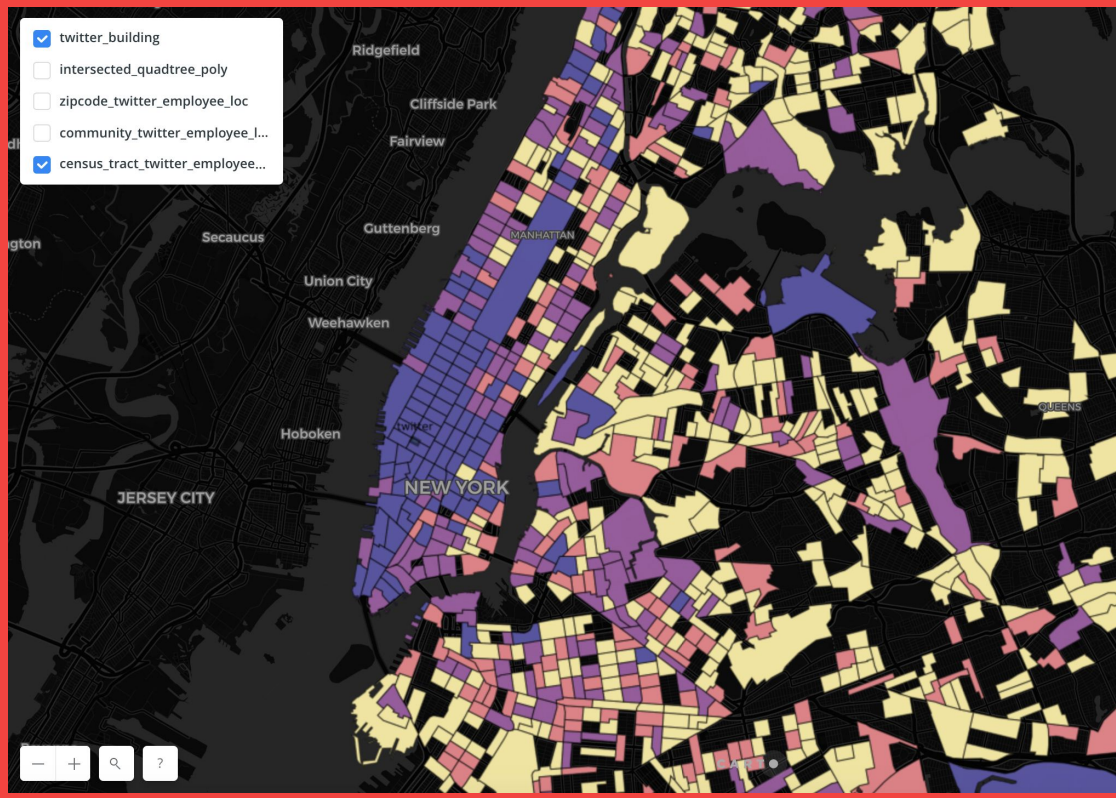


HWE

HWE uses CARTO to find optimal spots for different business purposes from Marketing to Real Estate.

Using Unacast foot traffic GPS data they are able to get the home locations of people visiting a selected POI, aggregated at different sizes of grid cells. This is achieved by:

- Identifying the quadkeys intersecting the POI building
- Using GPS data to identify people staying or visiting the POI based on the number of pings and average accuracy of target users
- Using GPS data to bactrack user's home locations based on pings during night hours

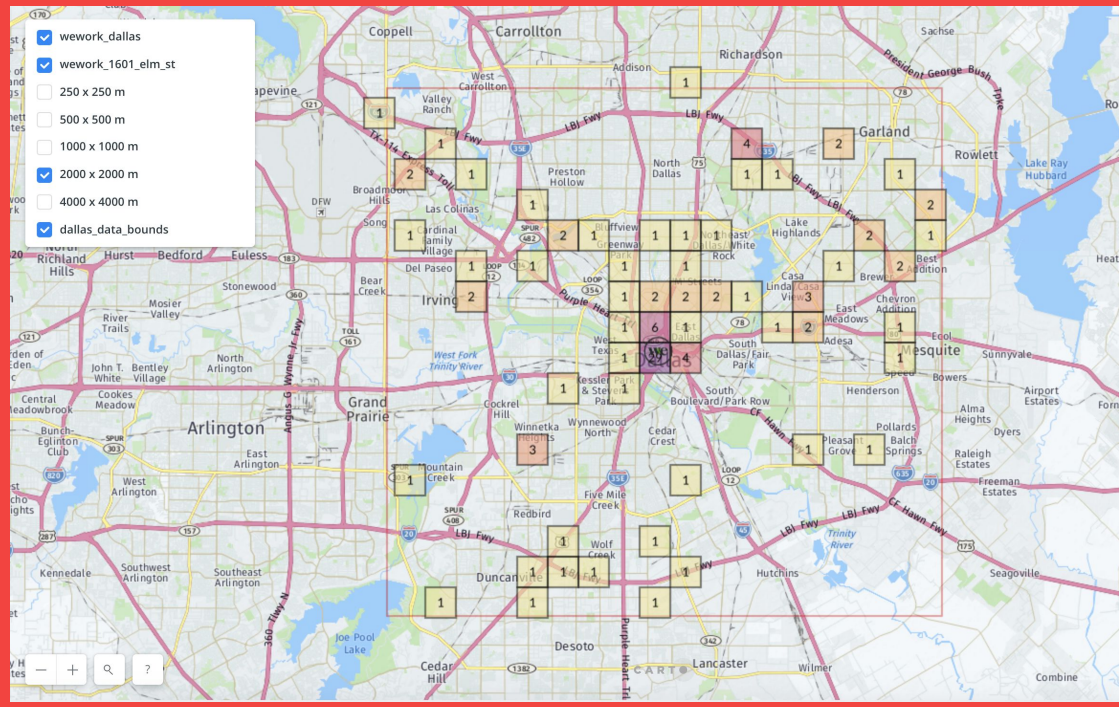


wework

WeWork uses CARTO to identify optimal spots to open new coworking spaces or move existing ones.

Using Unacast foot traffic GPS data they are able to get the home locations of people working at or visiting a selected POI (usually a work location). This is achieved by:

- Identifying the quadkeys intersecting the POI building
- Using GPS data to identify people working or visiting the POI based on the number of pings during working hours & days, and average accuracy of target users
- Using GPS data to bactrack user's home locations based on pings during night hours

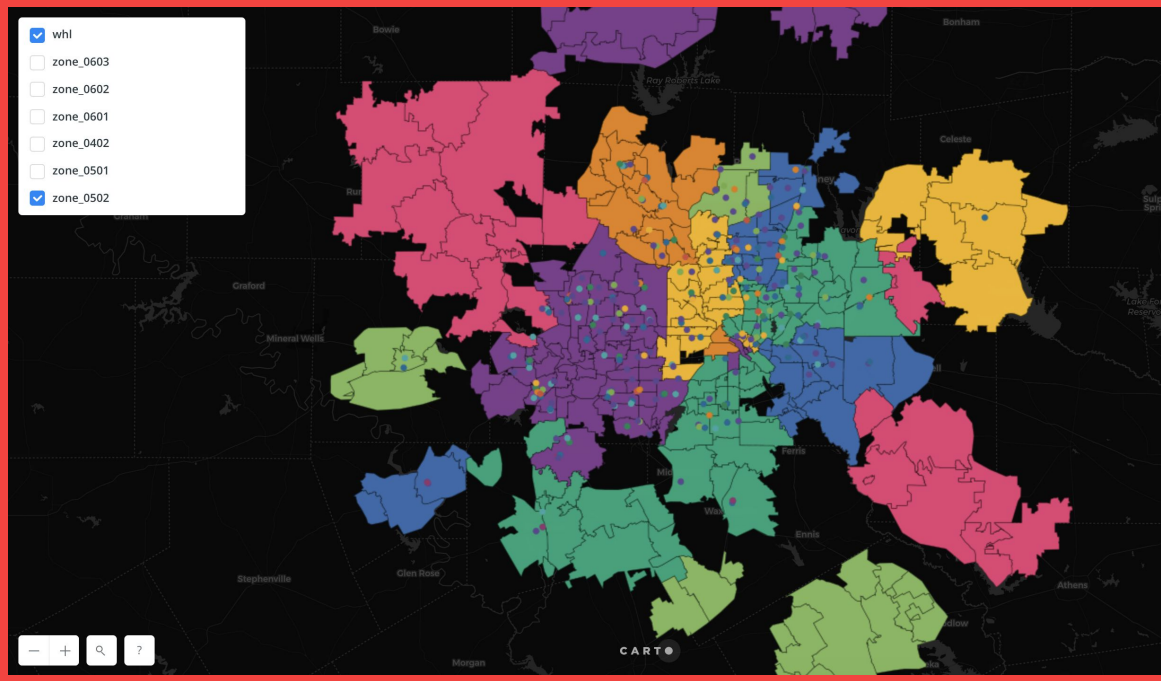


instacart

Instacart uses CARTO to build balanced delivery zones while maintaining high service levels.

Using the locations of the retailers they deliver from, a city's zip codes, and a list of service level criteria, a graph algorithm is applied whose targets are:

- To balance delivery zones making compact and similar size areas by minimizing the maximum crossing distance of each delivery zone
- To maximize service level by trying to satisfy as many service level criteria as possible

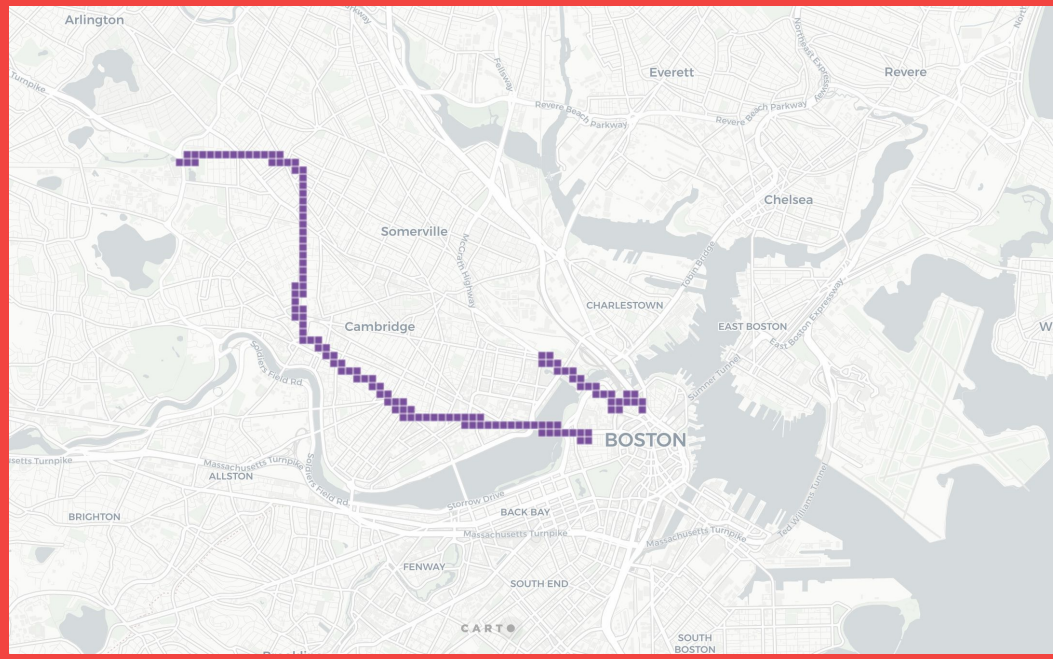


ARCADIS

Arcadis uses CARTO to analyze urban mobility data and transform it to produce meaningful aggregated data to enrich further analyses.

Using Unacast GPS data they are able to build an Origin-Destination (OD) matrix for every 30-minute time window with the number of people traveling by different means of transportation. This is achieved by:

- Using GPS data to identify people passing by target locations.
- Using the speed and acceleration to infer the means of transportation.

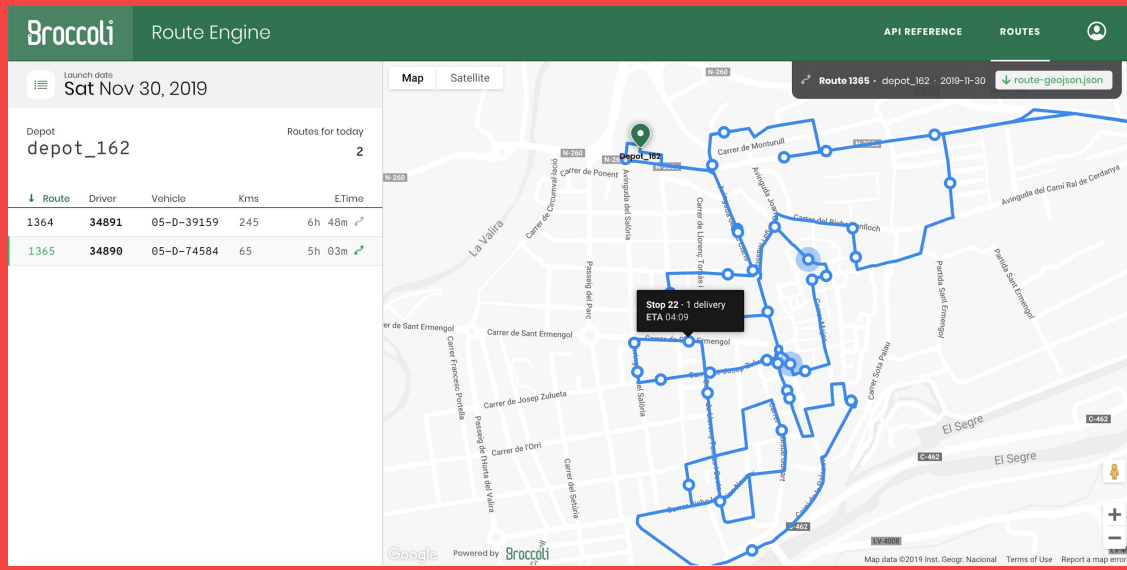


sorigué

Sorigué uses CARTO to optimally plan and manage the routes for the collection and disposal of urban waste.

Using the locations of the collection points and the disposal plants, optimal routes are calculated to minimize fleet and the costs associated with running this service by:

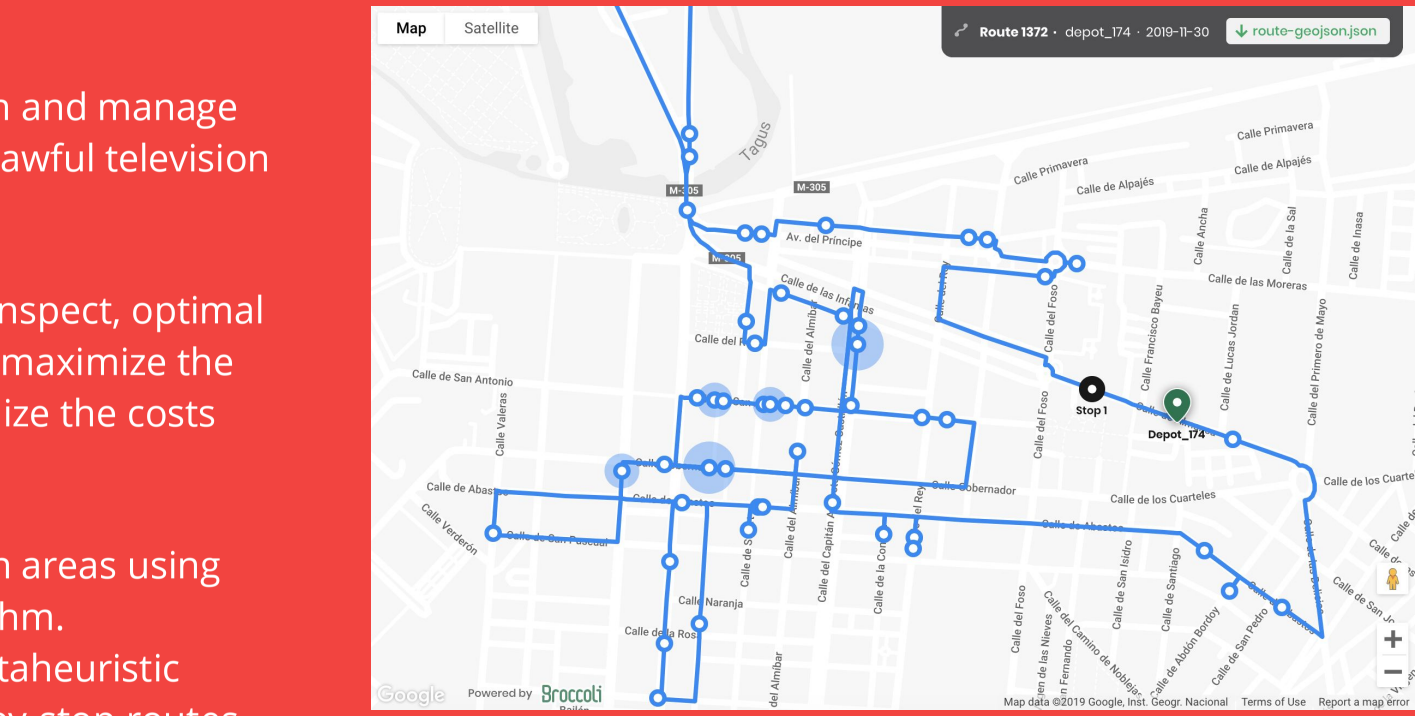
- Designing balanced collection areas using a balanced k-means clustering algorithm.
- Applying a multi-objective metaheuristic algorithm to design the stop-by-stop routes.



La Liga uses CARTO to optimally plan and manage their inspector's routes to detect unlawful television broadcasting of football matches.

Using the locations of the places to inspect, optimal multimodal routes are calculated to maximize the number of places visited, and minimize the costs associated with these routes by:

- Designing balanced inspection areas using the DBSCAN clustering algorithm.
- Applying a multi-objective metaheuristic algorithm to design the stop-by-stop routes.



CARTO

Demos

https://github.com/CartoDB/workshop_UCL-CASA

- Visualizing spatial data with CARTOframes
- Compute measures of spatial dependence
- Model continuous spatial processes
- Model discrete spatial processes
- Optimization and the Travelling Salesman problem