

GeoPython 2025

MACHINE LEARNING WITH OPEN GEOSPATIAL VECTOR DATA

Kamil Raczycki

2025-02-25





TUTORIAL OUTLINE

Geospatial vector data

Embeddings & SRAI library

OSM and Overture Maps data

Practical examples



Kamil Raczycki

kamilraczycki.com

 /raczyckikamil

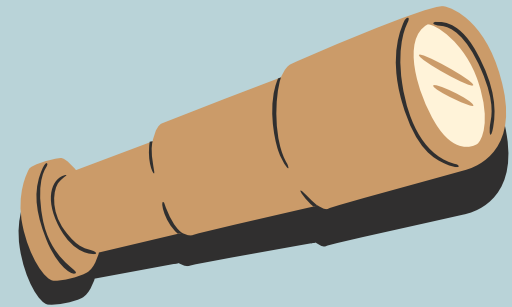
 /raczeq

ABOUT ME

Background: MSc in Data Science, part of the Kraina.AI research group.

Open source contributor: co-developer of **SRAI** library, maintainer of **QuackOSM** and **OvertureMaestro** libraries.

Currently working as a Senior Data Scientist specializing in geospatial analyses.

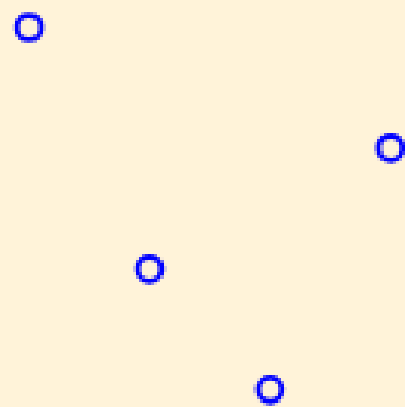


GEOSPATIAL VECTOR DATA

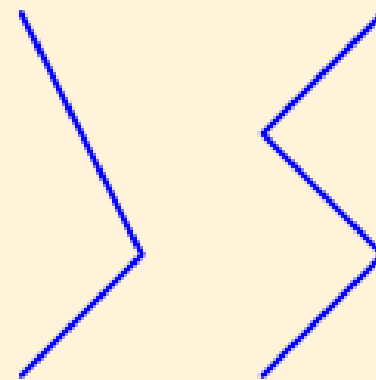


Types of geospatial vector data

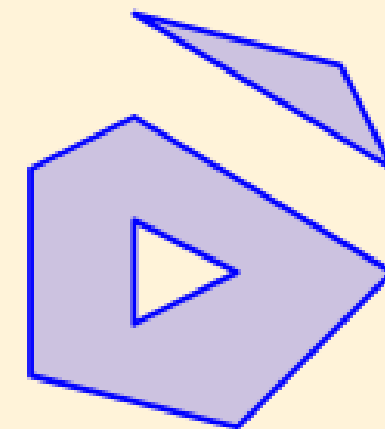
Points



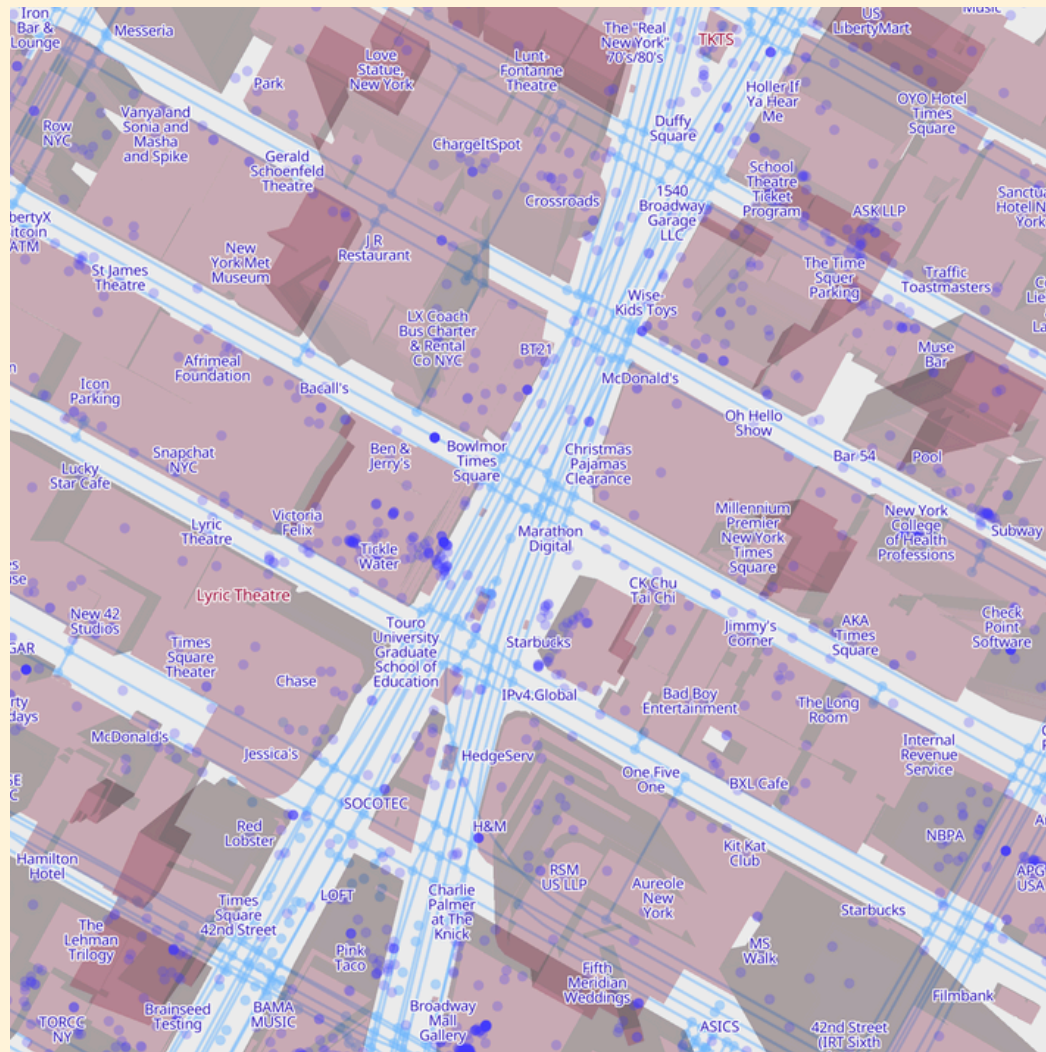
LineStrings



Polygons



VECTOR VS RASTER (SATELLITE)



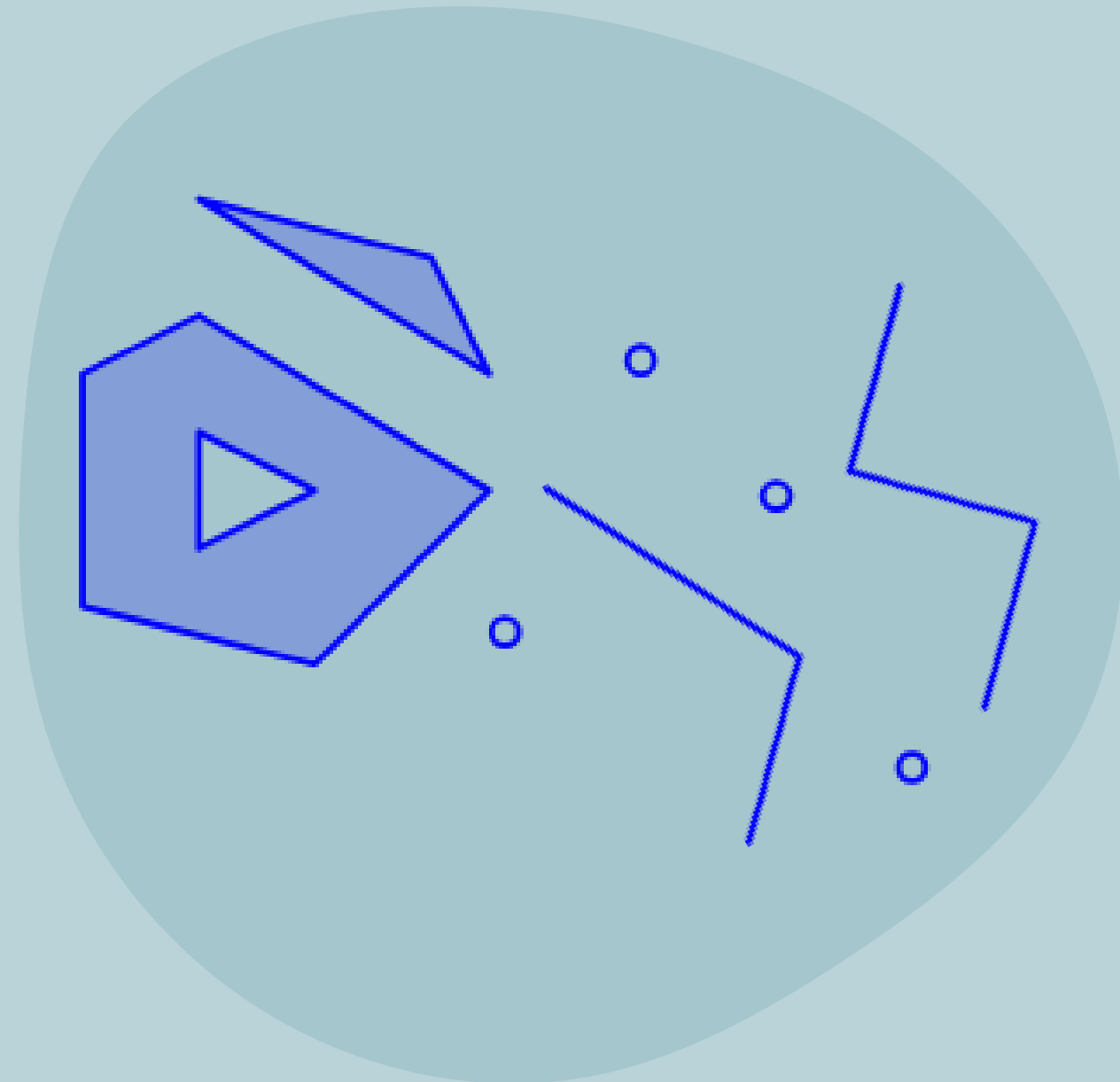
Vector data is great for modelling inside dense urban regions with rich and precise data.

Raster data is great for modelling large-scale phenomena (weather patterns, vegetation index).



HOW CAN VECTOR DATA BE USED IN ML MODELS?

Machine Learning is optimized for working with numbers.



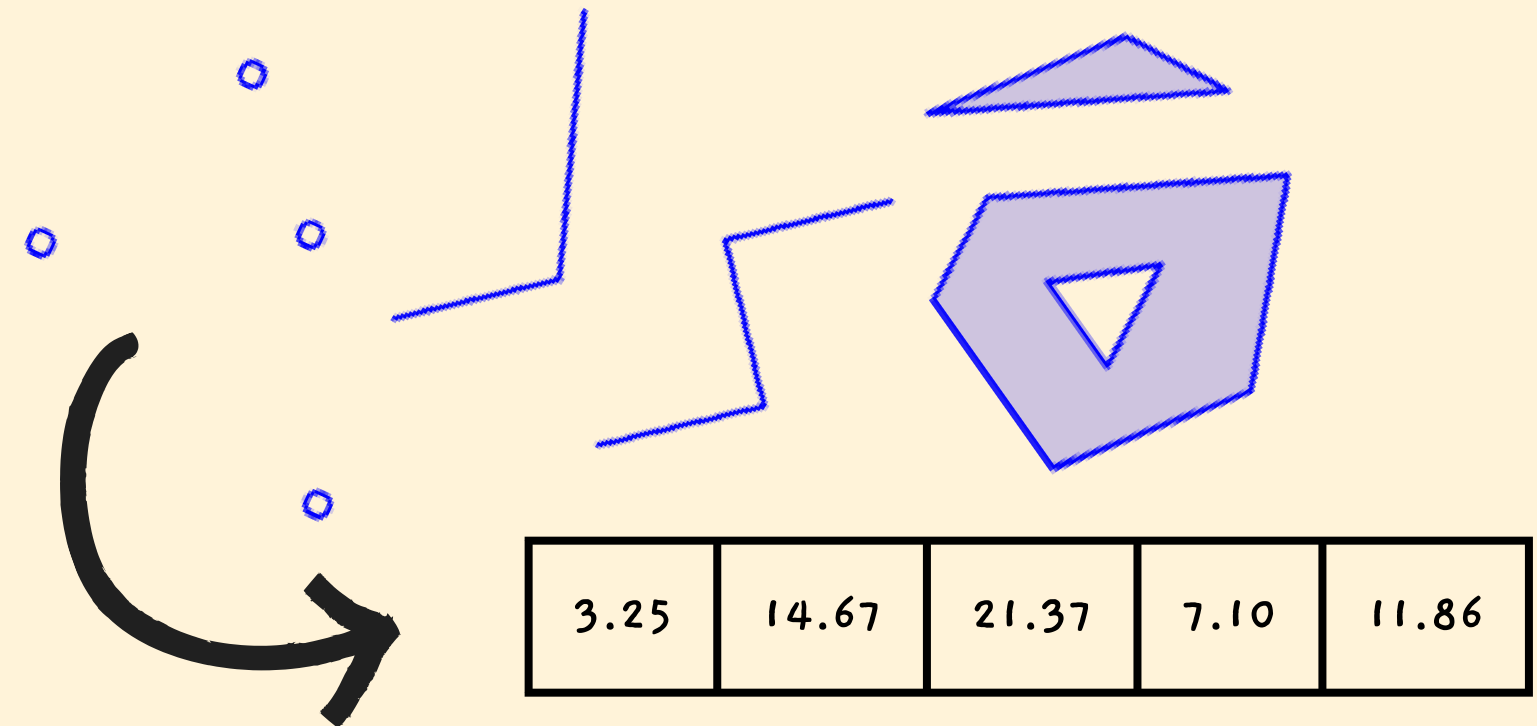
Maybe I'll put
POINT (7.4228
43.72985)
as a string into the
model?

ANSWER



EMBEDDINGS!

Embeddings are representations of values or objects like text, images, audio or **geospatial data**.



REGION EMBEDDINGS



 Ocean

 Mountain


 Lake

 Treasure



				
1		1		
	2		1	
				1

1 mm = 50 km





srai

SRAI LIBRARY

**Spatial Representations for
Artificial Intelligence**

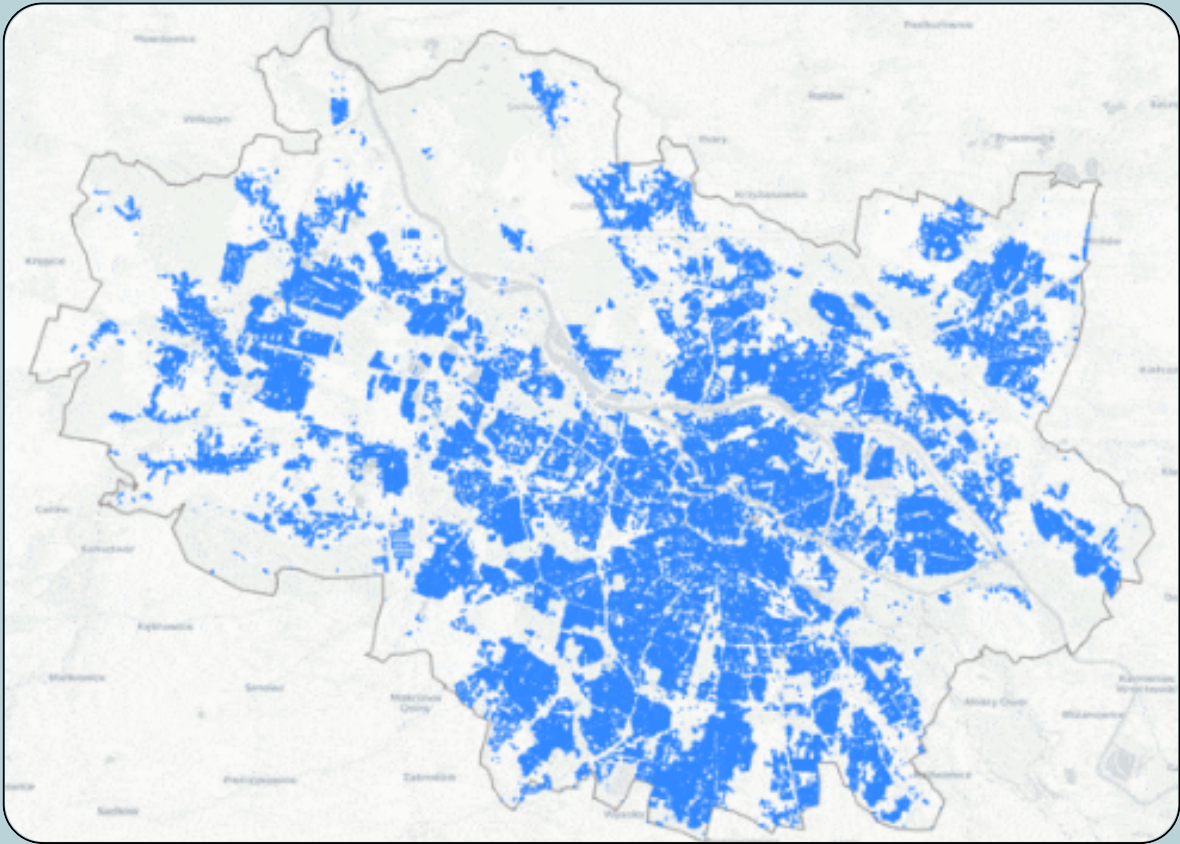
The library created for geospatial
feature engineering and generating
region embeddings.



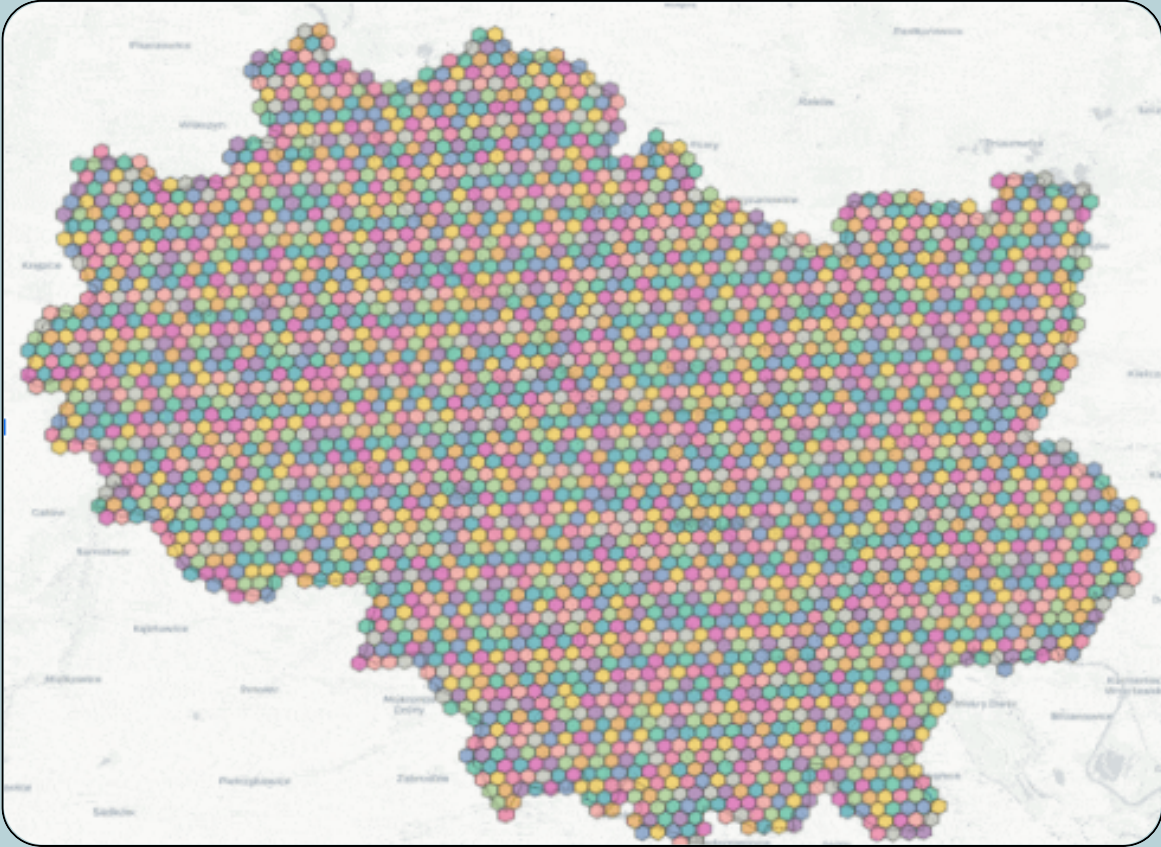
Define area
of interest



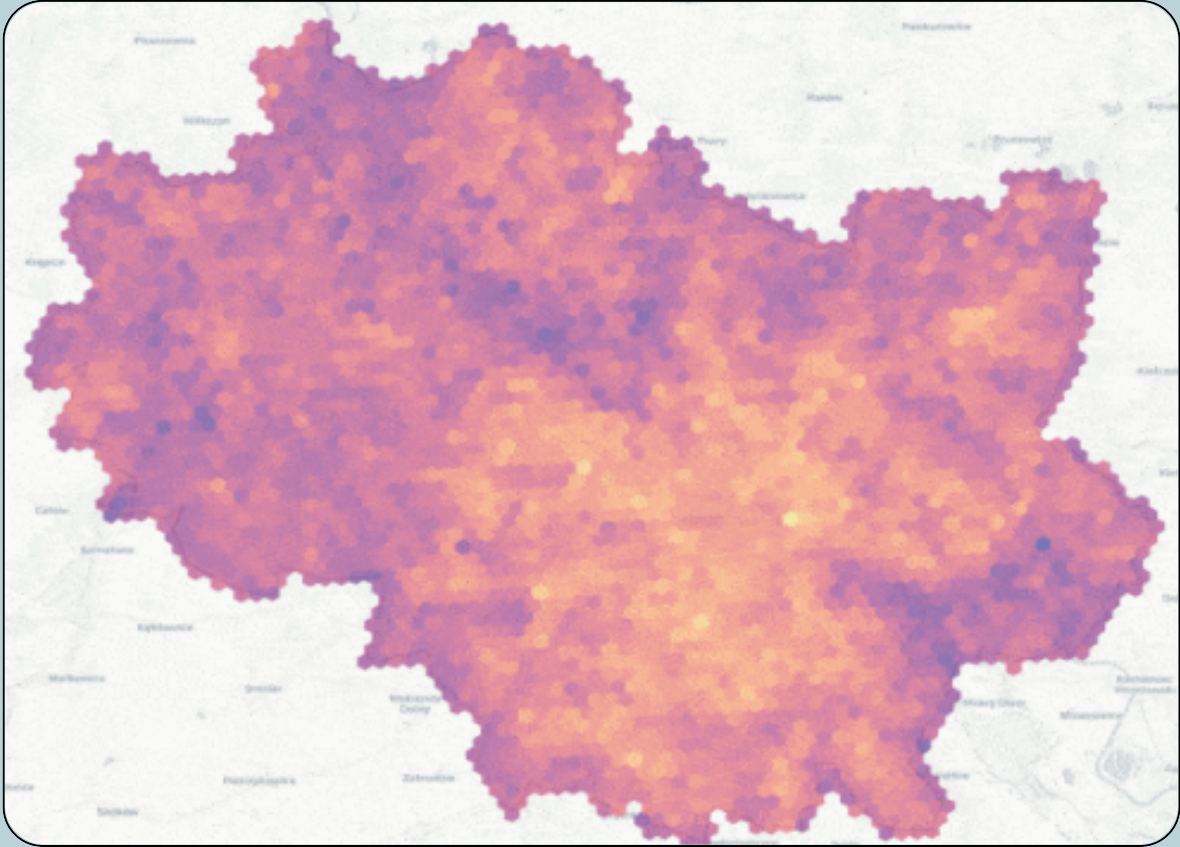
Download
data

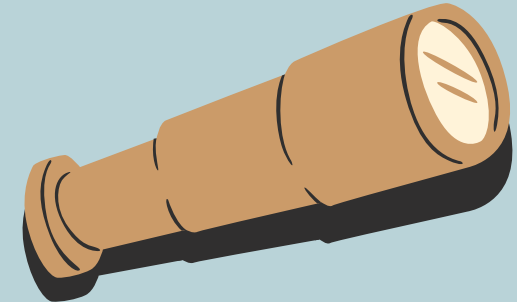


Regionalize



Create
embeddings





DIFFERENT FEATURES



What can you find inside the SRAI library?

Regionalizers:

H3, S2, Voronoi,
Admin Boundaries

Loaders:

OpenStreetMap,
OvertureMaps,
GTFS

Embedders:

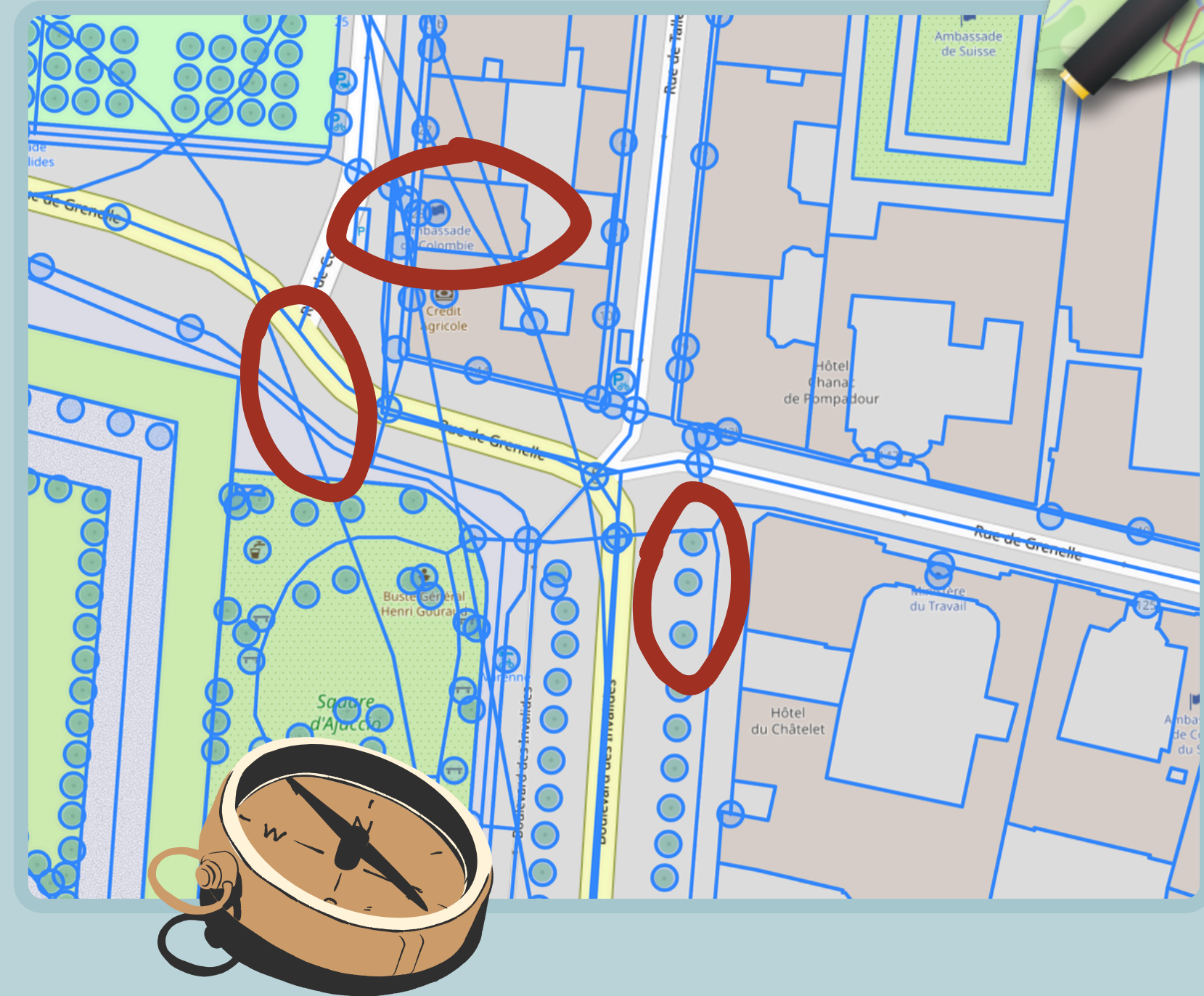
Count, Contextual
Count, Hex2Vec,
GeoVex, GTFS2Vec,
Highway2Vec

OPENSTREETMAP DATA

OSM data consists of 3 data elements:

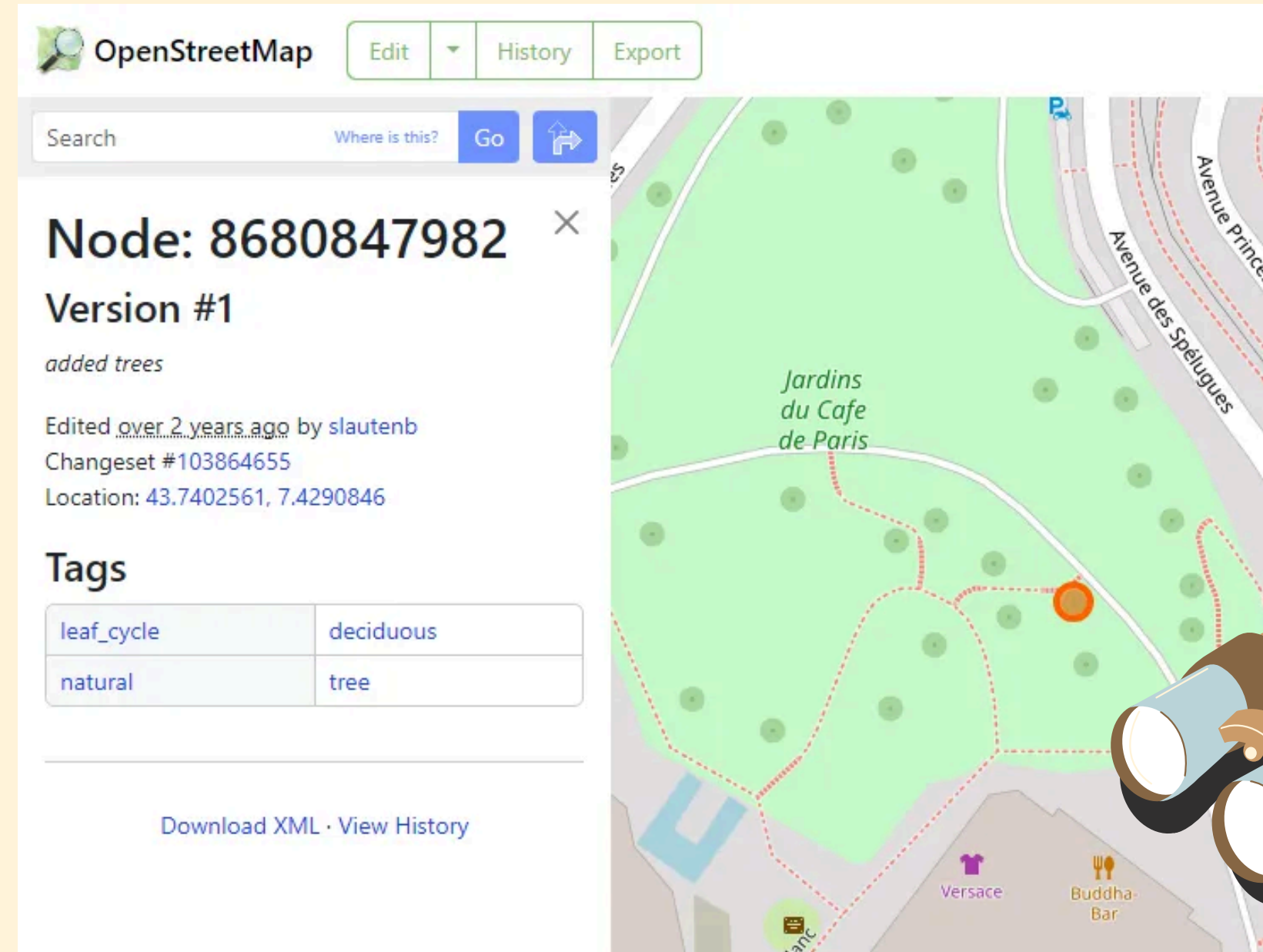
- Nodes
- Ways
- Relations

Only nodes have coordinates (lat / lon).
Everything else has to be reconstructed.



OPENSTREETMAP DATA

There are also tags containing information about the object.



The screenshot displays the OpenStreetMap interface. At the top, there's a search bar and navigation buttons (Edit, History, Export). The main content area shows details for a specific node:

- Node: 8680847982** (with a close button)
- Version #1**
- added trees*
- Edited over 2 years ago by [slautenb](#)
- Changeset #[103864655](#)
- Location: [43.7402561, 7.4290846](#)

Below this information is a table of tags:

Tags	
leaf_cycle	deciduous
natural	tree

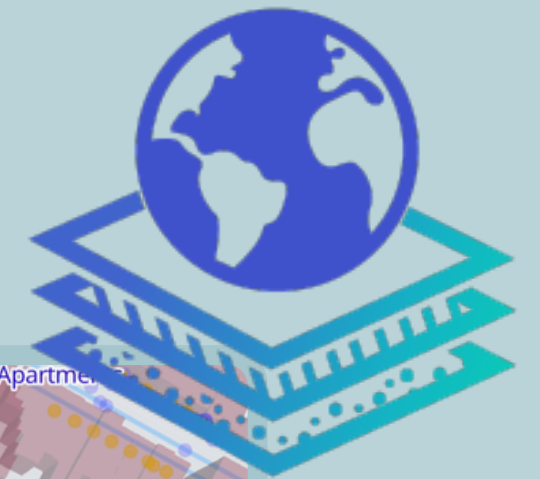
At the bottom of the details panel, there are links for [Download XML](#) and [View History](#).

The map on the right shows a green area labeled "Jardins du Cafe de Paris" with several tree icons. A red dot indicates the location of the selected node. Surrounding streets include "Avenue des Spélugues" and "Avenue Princes". Other nearby points of interest like "Versace" and "Buddha-Bar" are also visible.

OPENSTREETMAP DATA

feature_id		geometry	aeroway	amenity	building	healthcare	historic	landuse
node/1269618045		POINT (21.08914 52.24265)	None	restaurant	None	None	None	None
node/1269640653		POINT (21.03589 52.20203)	None	fast_food	None	None	None	None
node/1269640664		POINT (21.03967 52.20219)	None	None	None	None	None	None
node/1270633772		POINT (20.9571 52.32017)	None	None	None	None	None	None
node/1270689059		POINT (21.06246 52.13417)	None	None	None	None	None	None
...	
way/1358329361	POLYGON ((21.06142 52.24099, 21.06143 52.24099...		None	toilets	toilets	None	None	None
way/1358329962	POLYGON ((21.03105 52.25706, 21.03107 52.25703...		None	None	None	None	None	garages
way/1358332342	POLYGON ((20.92473 52.20373, 20.9247 52.20367,...		None	parking	None	None	None	None
way/1358332343	POLYGON ((20.92505 52.20357, 20.92513 52.20359...		None	parking	None	None	None	None
way/1358376631	POLYGON ((20.96831 52.25629, 20.96831 52.2563,...		None	parking	None	None	None	None

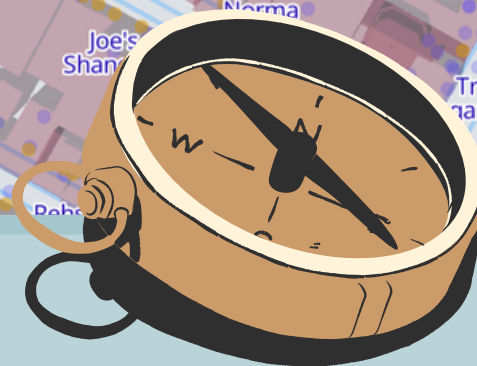
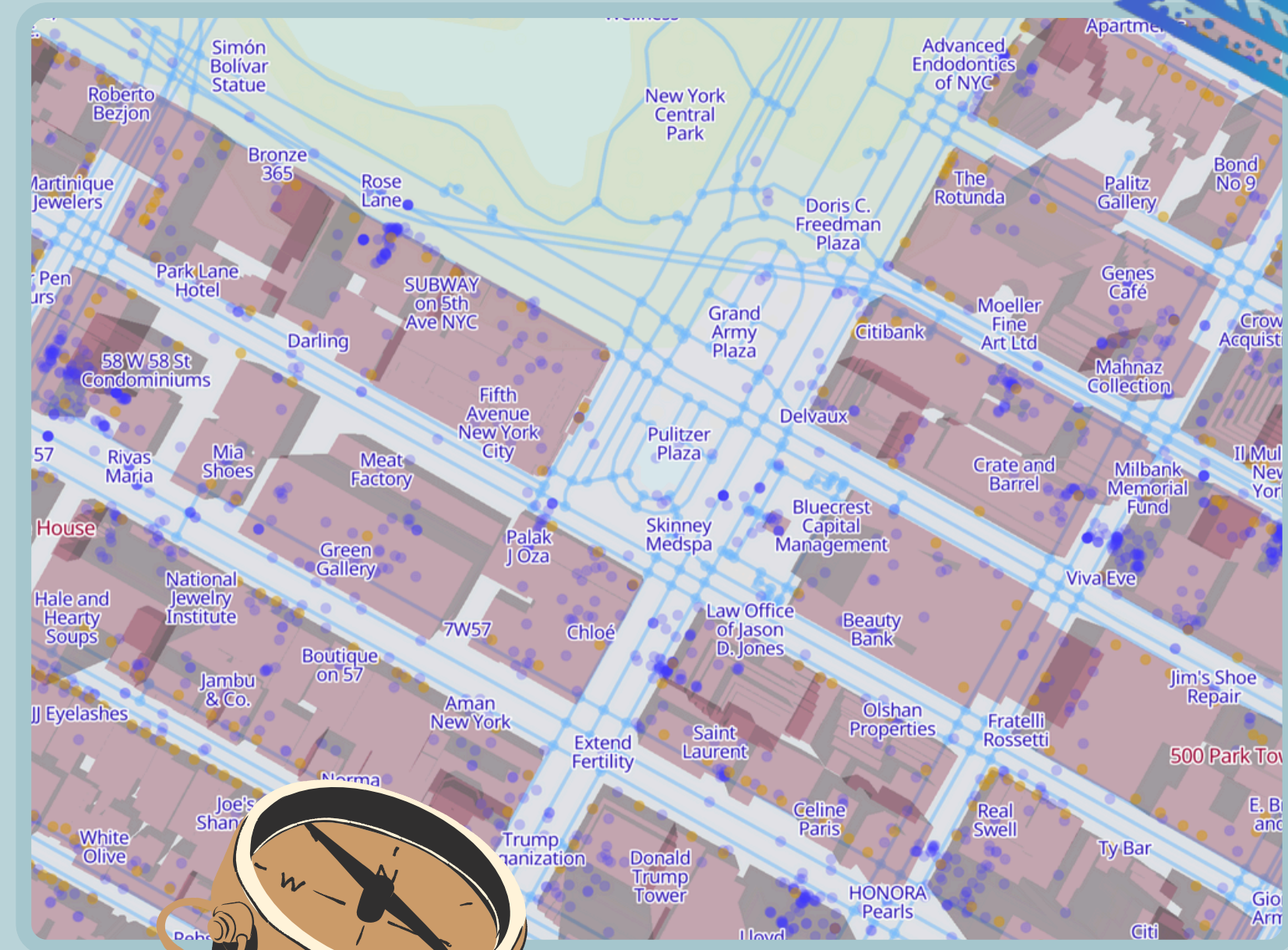
OVERTURE MAPS DATA



Overture Maps data has a well-defined schema.

Data is distributed as GeoParquet files every month.

Datasets are partitioned using theme/type pairs.



OVERTURE MAPS DATA

	geometry	version	sources
id			
08019fffffffff047dfb89a3c4b14d	LINESTRING (-6.33486 52.25478, -6.33887 52.256...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
08019fffffffff047be35289e98060	LINESTRING (-6.33657 52.25383, -6.34007 52.256...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
084194adffffffff047fff2976eb9e18	LINESTRING (-0.18519 51.49496, -0.1847 51.4948...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
084194adffffffff046fb689e755c681	LINESTRING (-0.07534 51.51346, -0.0755 51.5128...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
089194ad14d7ffff047f373a11370670	LINESTRING (-0.12 51.50069, -0.12012 51.5007, ...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
...
086194ad37ffff046b7f6a4e072d20	LINESTRING (-0.07889 51.50727, -0.084 51.50705...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
086194ad37ffff046bfcd3b02fef16	LINESTRING (-0.0968 51.50877, -0.09455 51.5089...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
082194ffffffff047f4f9d64e32154	LINESTRING (-0.10492 51.50439, -0.10161 51.504...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
082194ffffffff046b7f449df5cef3	LINESTRING (-0.14366 51.49584, -0.14358 51.496...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
081197ffffffff047dfab323e684b3	LINESTRING (3.18966 51.34097, 3.19028 51.34133...	0	[{'property': '', 'dataset': 'OpenStreetMap', ...
2387 rows × 20 columns			



PRACTICAL EXAMPLES

London population density

Bicycle sharing stations locations

Airbnb listing prices

Tutorial materials

kraina-ai/srai-tutorial



A tutorial for the SRAI library

5
Contributors

13
Issues

22
Stars

4
Forks



kraina-ai/srai-tutorial at geopython2025

A tutorial for the SRAI library. Contribute to kraina-ai/srai-tutorial development by creating an account on GitHub.

 GitHub



Every notebook can be
run on **Google Colab!**






Kamil Raczycki

THANK YOU!

kamilraczycki.com

 /raczyckikamil

 /raczeq

