

Selection of the best place to open a lab

Project for IBM applied Data Science capstone Giovanni Manfredi

2020/10/19

Goal

Find the best place to open a new laboratory in Genova

Customer

Generic Health Company

Requirements

- must be near public transportation
- must be near universities
- must be near health institutes
- lower costs are preferable

Used data

Venues data

FourSquare

The data contained all the venues near a specific area comprising precise location, venue category, and venue principal information.

Public transport stops

Genova municipality

<http://dati.comune.genova.it/dataset/amt-trasporto-pubblico-nel-comune-di-genova/resource/8dab9db4-c0b1-4146-8e3a-29d9c91d08ef>

Each point of the dataset comprised a name of the stop, its coordinates, the municipality, and other information.

Real estate data

Italian Observatory of real estate market

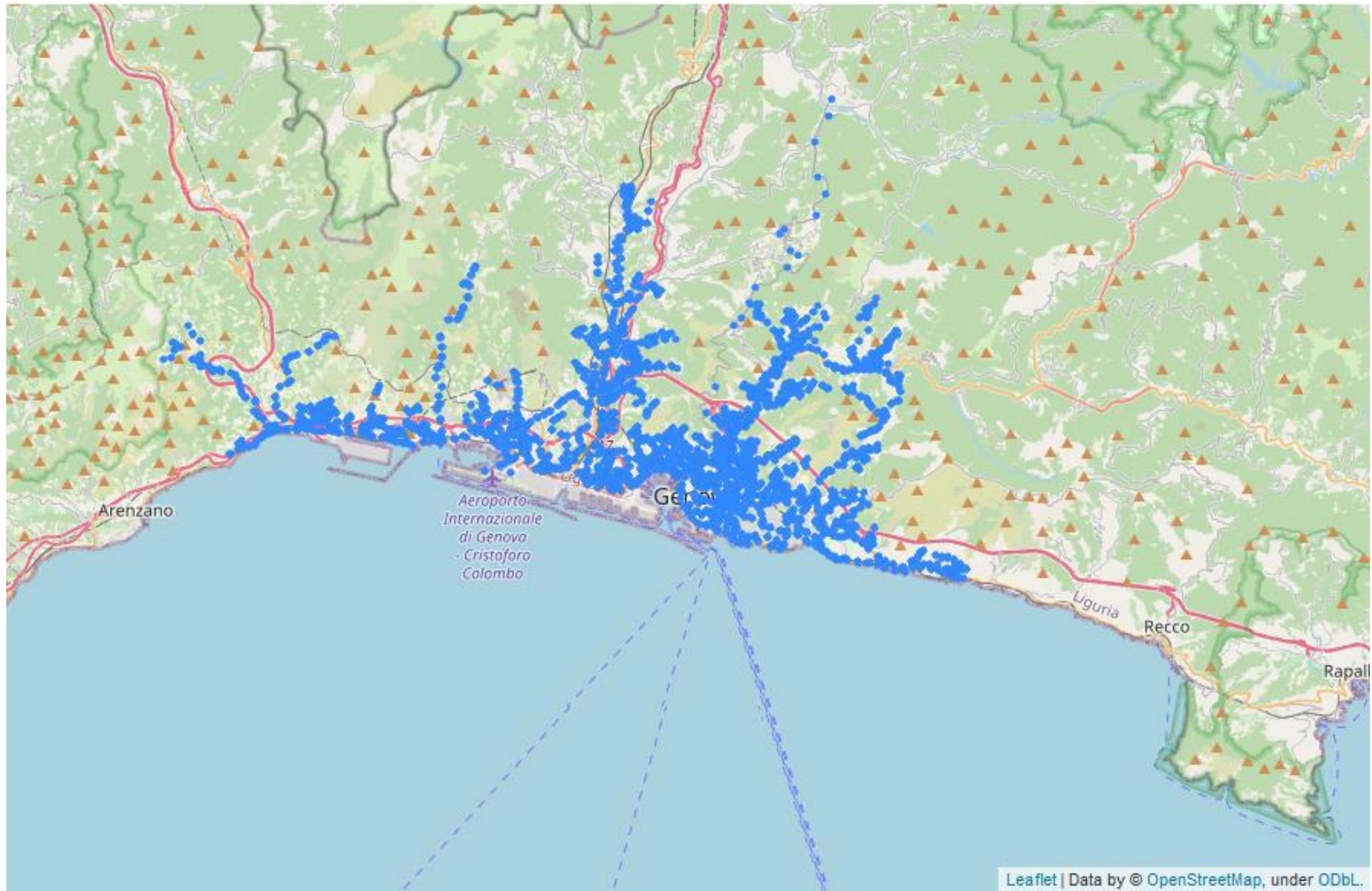
<https://www.agenziaentrate.gov.it/portale/web/guest/aree-tematiche/osservatorio-del-mercato-immobiliare-omi>

Previously acquired and cleaned by OnData civic hacking group

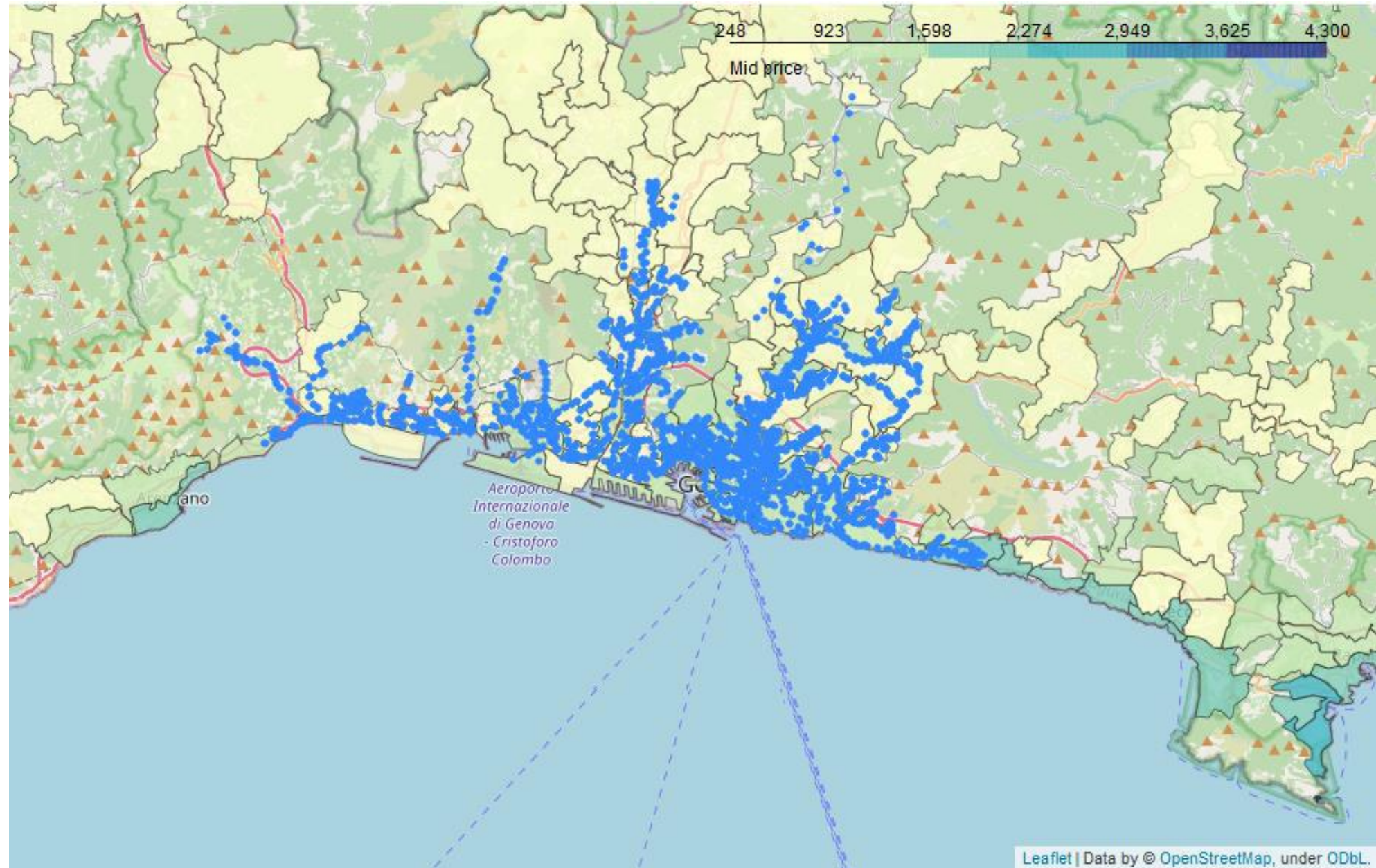
<http://blog.ondata.it/i-dati-sulle-quotazioni-immobiliari-dellagenzia-entrate/>

Each point of the dataset is related to a specific geographical zone and included a minimum purchase and rent price, a maximum purchase and rent price, plus other information.

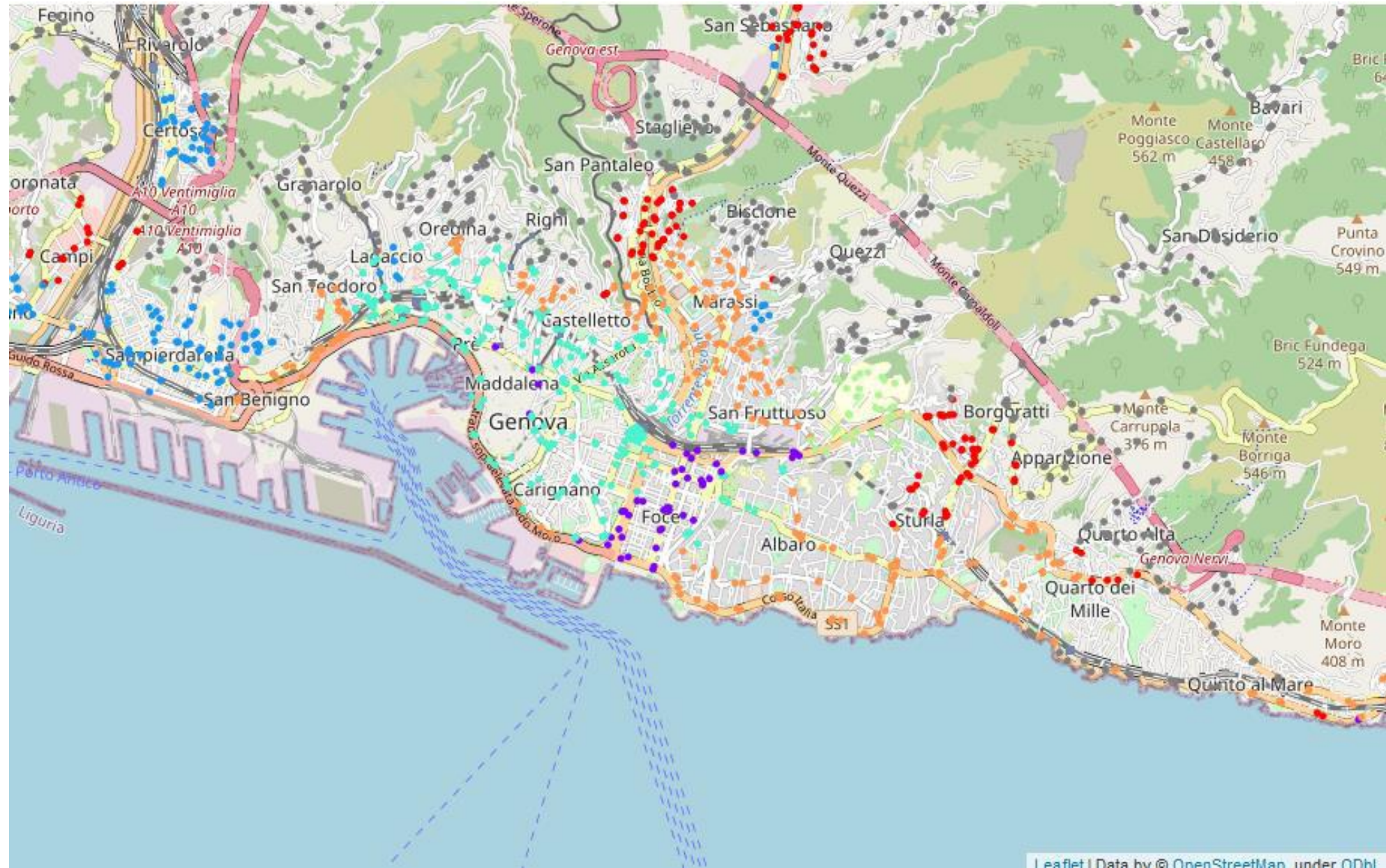
Genova public transport stops form an efficient mesh of all the municipal territory



Italian real estate observatory data covers the territory to large extent



Detail of clustering in quarters



Score to find the best place

$$S = \frac{\sum_i w_i \cdot n_i}{\ln(p)}$$

Where

S = score of the row of the table

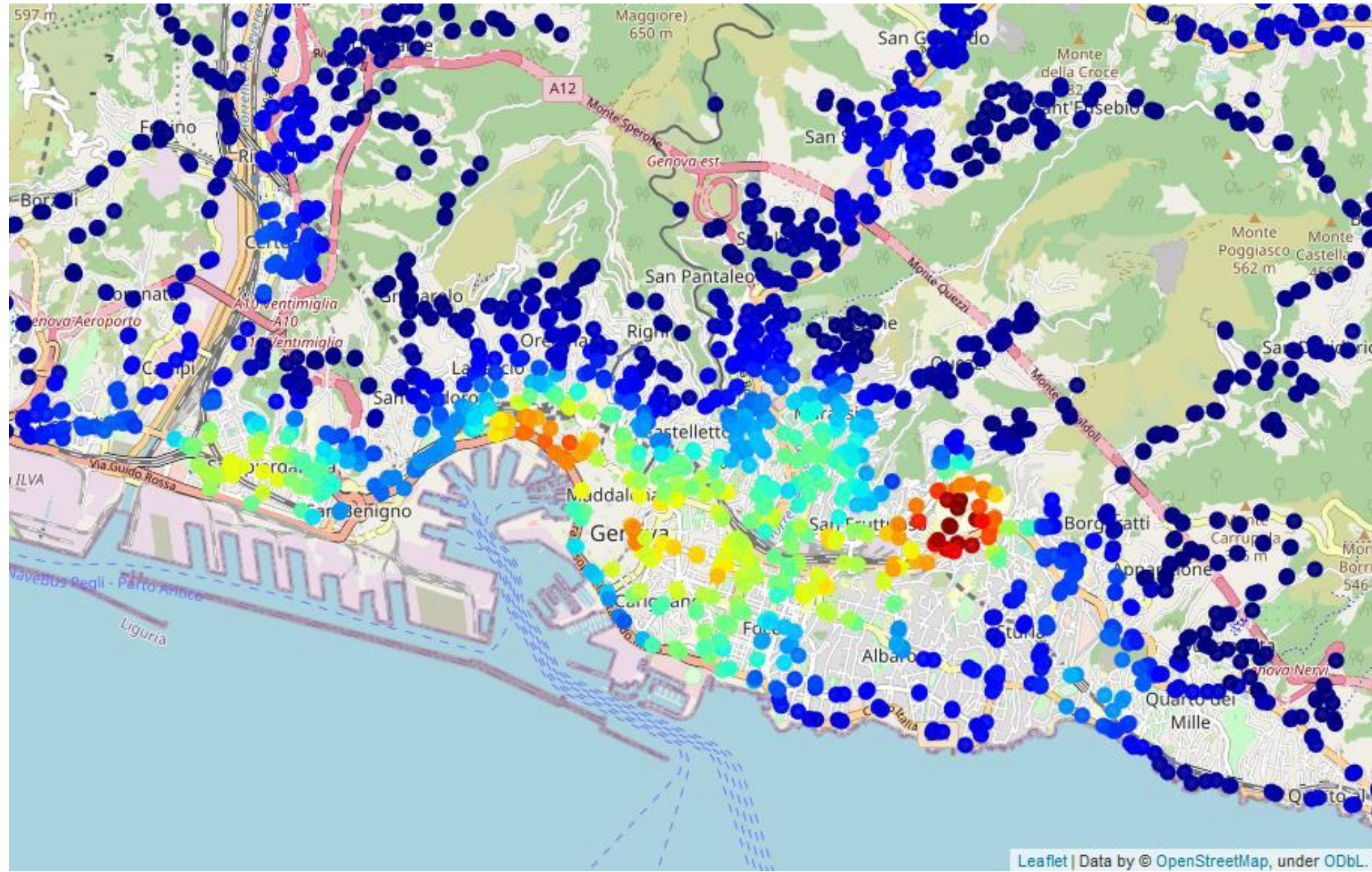
n_i = number of venuest of the i-th category of the row

w_i = weight of the i-th category in the score

p = mid-price of the row of the table

Category	w_i
Airports	0.5
Conventions	0
Entertainment	0
Fitness	0
Food	1
Government	0
Health	10
Hotels	0.5
Nightlife	0
Outdoor activities	0
Residence	1
School	1
Shops	0.5
Transports	2
University	7
Work activities	1
Entertainment	0

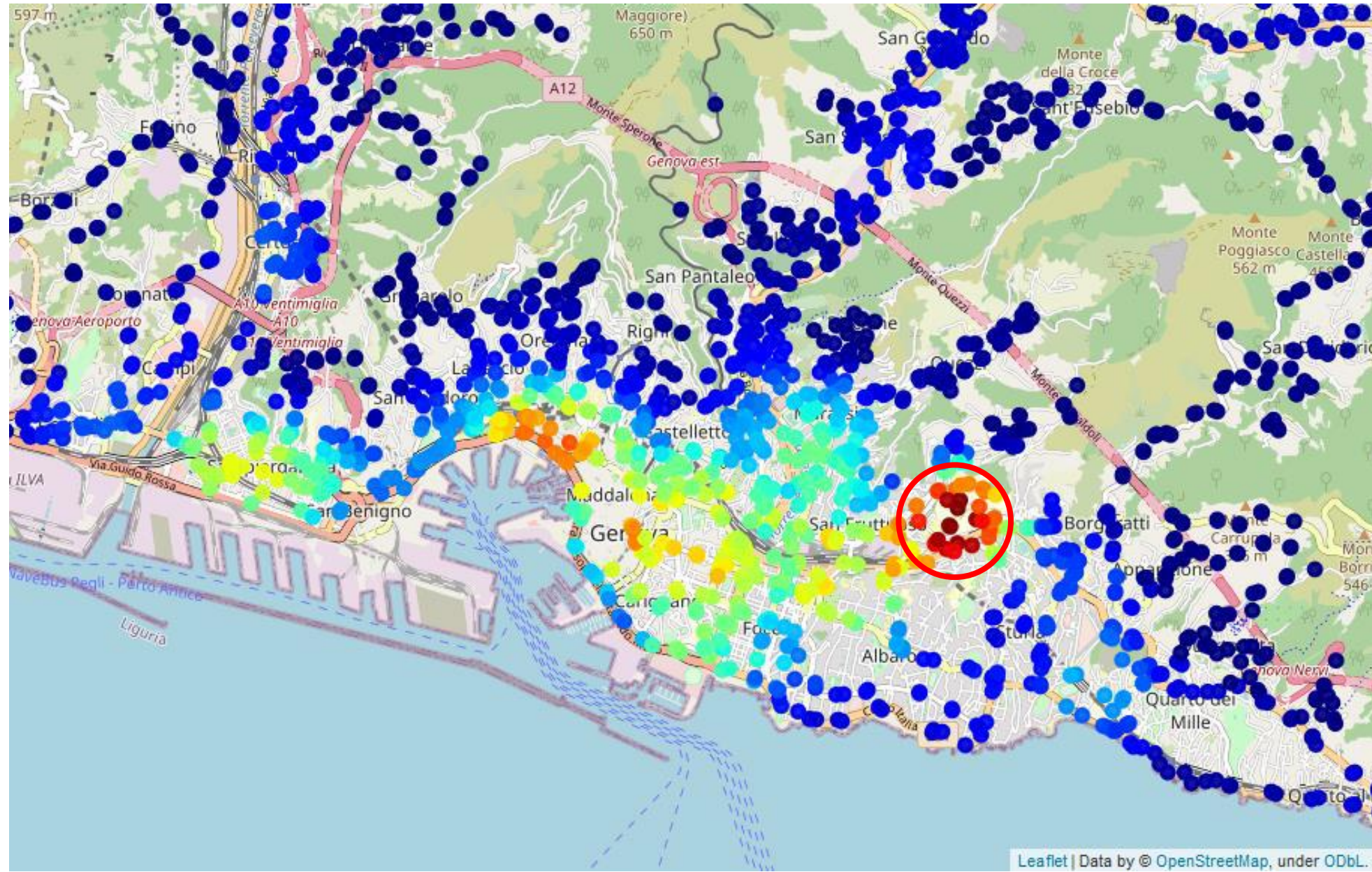
Map of the scores for each area



Selected zone



San Martino hospital area



Conclusion

Given the data supplied, the best place is **San Martino Hospital area**

Future developments

- Enhance real estate prices resolution scraping data from specialized sites.
- Enhance the quality of collected data.
- Improve the definition of the score to find the best place to build a laboratory.
- Train the score parameters on existing labs of the customer.