Property market in Paris

Comparing prices vs neighborhood’s services

Problem

- Price per m^2 too high
- Prices in central neighborhoods can in some cases be 67% higher than in peripheral ones
- Is the price difference supported by the range and density of available services?
- Can we help people with a preview about what to expect when looking for property?

Data acquisition

- Neighborhood's information (identification and geographical) provided by Paris City Hall (downloaded from https://www.data.gouv.fr and https://opendata.paris.fr)
- Property price information provided by Paris Chamber of Notaries (scrapped from https://droit-finances.commentcamarche.com)
- Services information provided by FourSquare

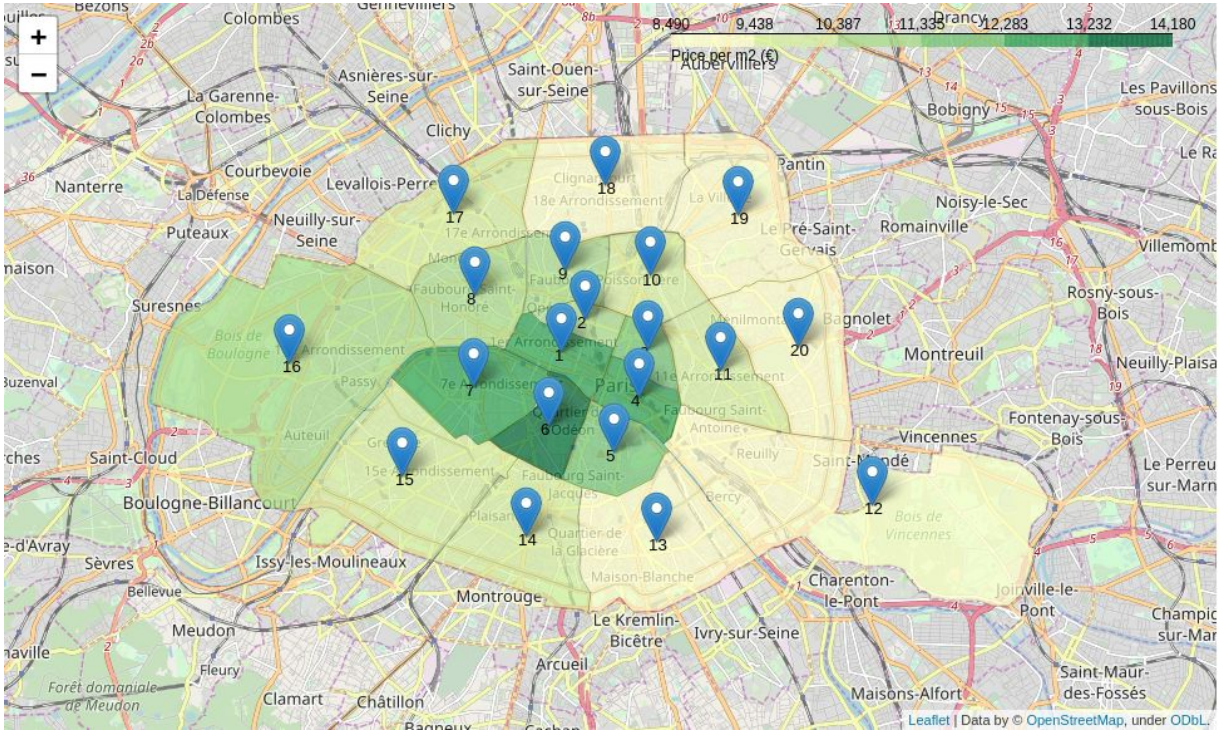
Data wrangling

- Final dataframe with 20 neighborhoods and 36 features
- Neighborhood identification and geographical information: 5 features
- Price information: 1 feature
- Services available: 14 features
- Services density (indexes): 15 features
- Clustering information: 1 feature

Neighborhoods overview

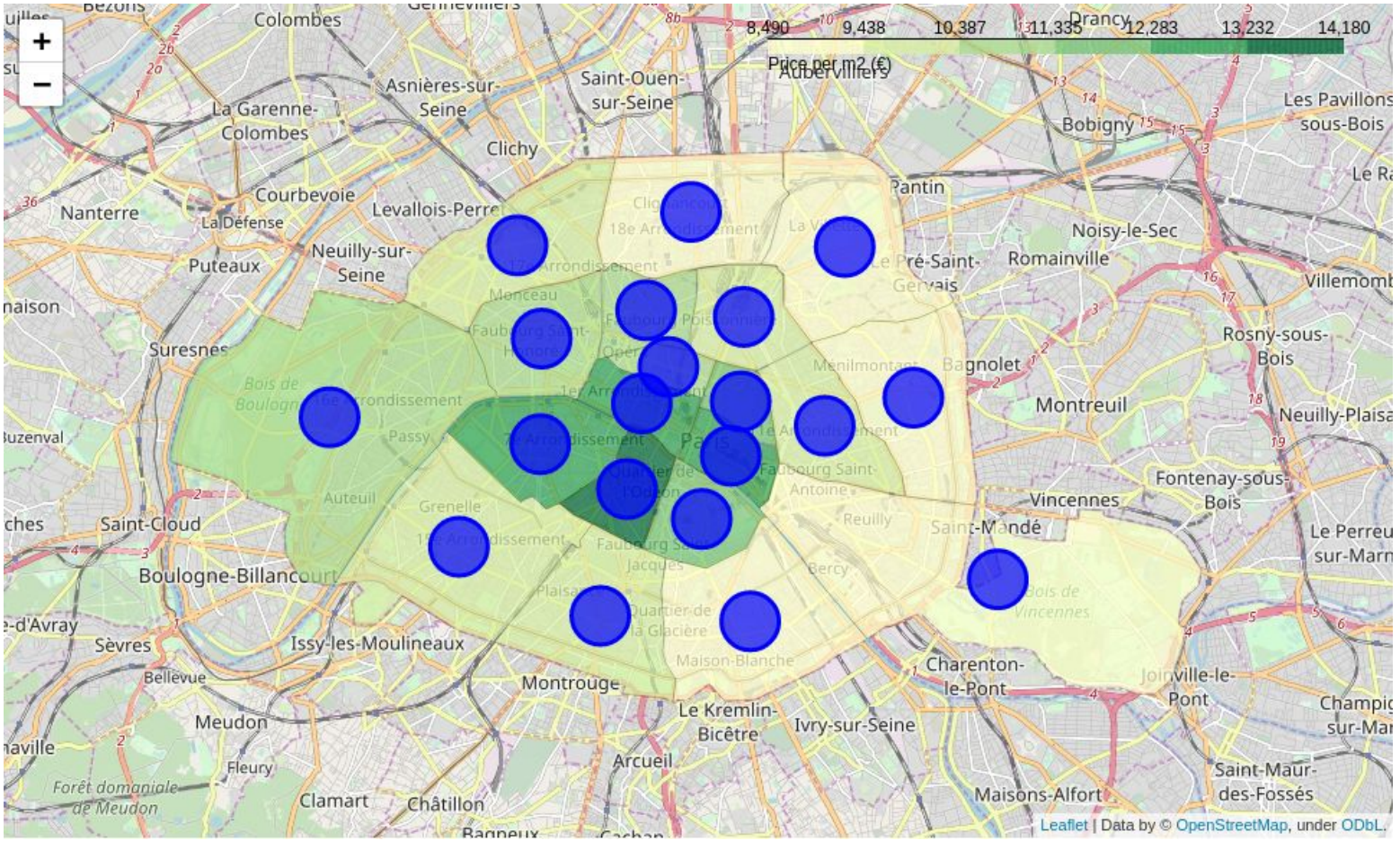
- Paris neighborhoods

Neighborhood	Name	PricePerM2
1	Louvre	12840
2	Bourse	11250
3	Temple	12260
4	Hôtel-de-Ville	12790
5	Panthéon	12140
6	Luxembourg	14180
7	Palais-Bourbon	13230
8	Élysée	11240
9	Opéra	10730
10	Entrepôt	9730
11	Popincourt	9980
12	Reuilly	9310
13	Gobelins	9060
14	Observatoire	10170
15	Vaugirard	10030
16	Passy	10680
17	Batignolles-Monceau	10210
18	Buttes-Montmartre	9360
19	Buttes-Chaumont	8490
20	Ménilmontant	8560



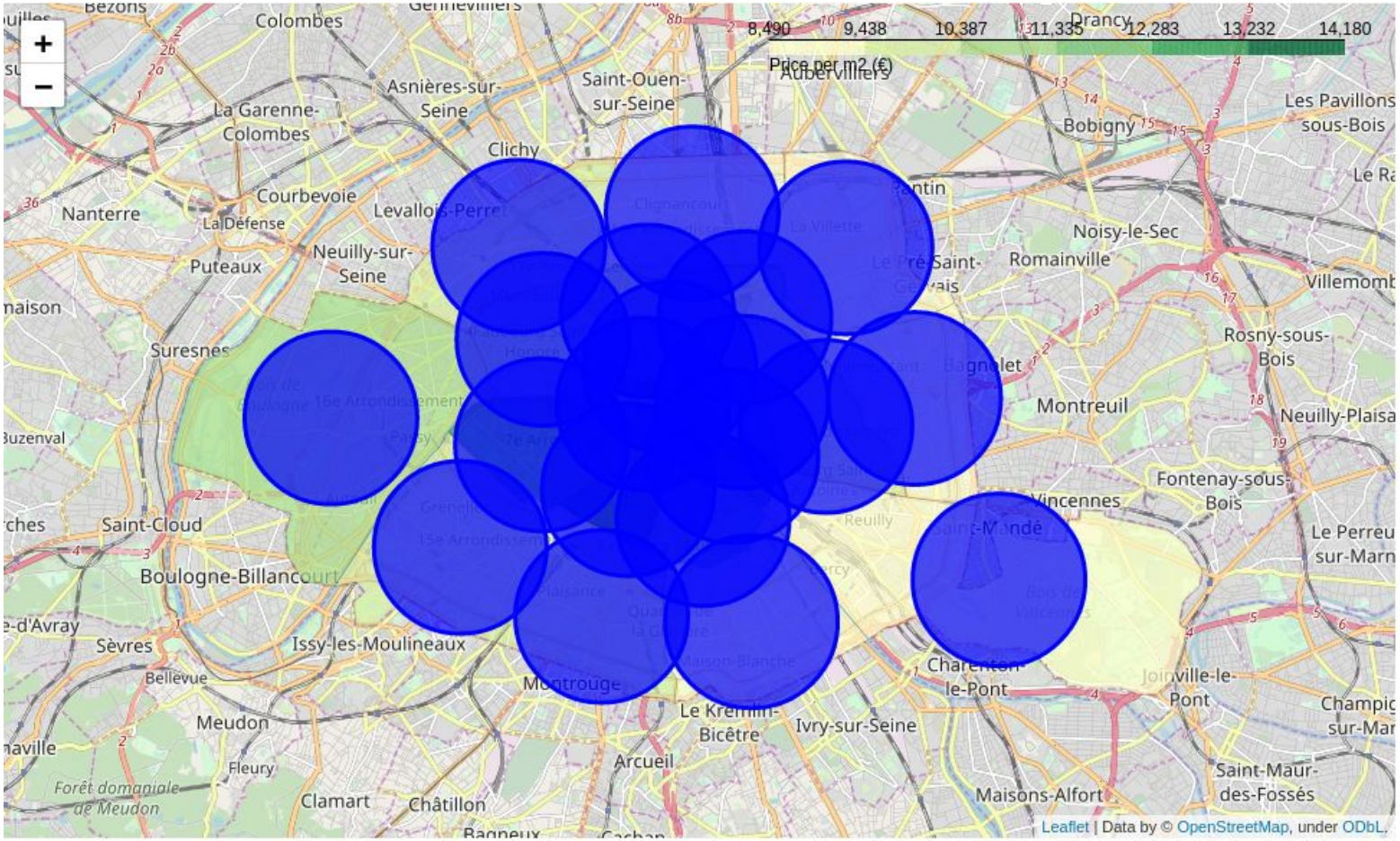
Defining the search area

- Small radius => lack of depth for large neighborhoods



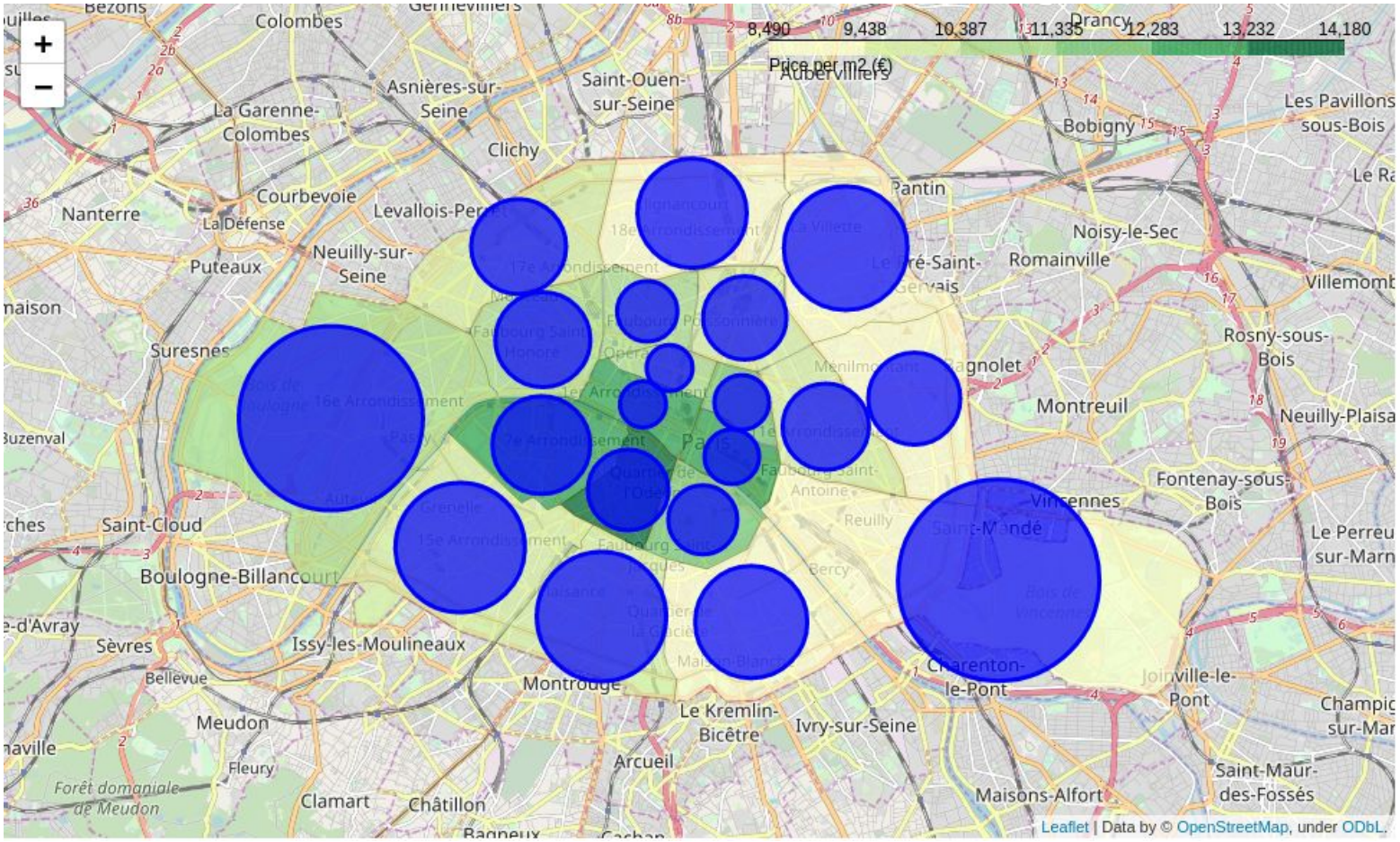
Defining the search area

- Big radius => overlapping of search areas



Defining the search area

- Dynamic radius => better coverage, no overlapping



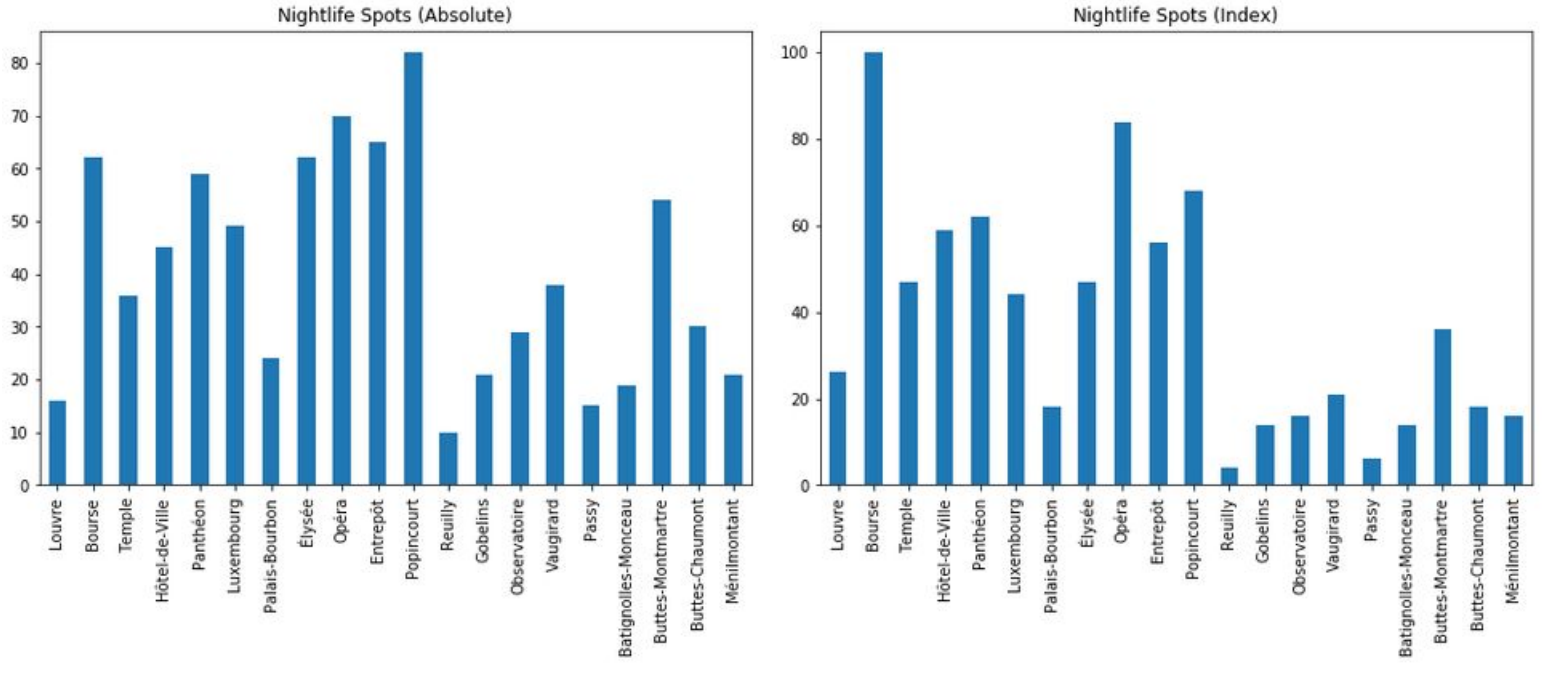
Searching criteria

- Dynamic radius for each neighborhood
- Limit to 100 findings per category per neighborhood
- Categories used:

Security	Mobility	Education	Entertainment	Services
Police Station	Parking	College & University	Arts & Entertainment	Post Office
Fire Station	Transport	School	Food	Shop & Service
Medical Center			Nightlife Spot	Spiritual Center
			Outdoors & Recreation	

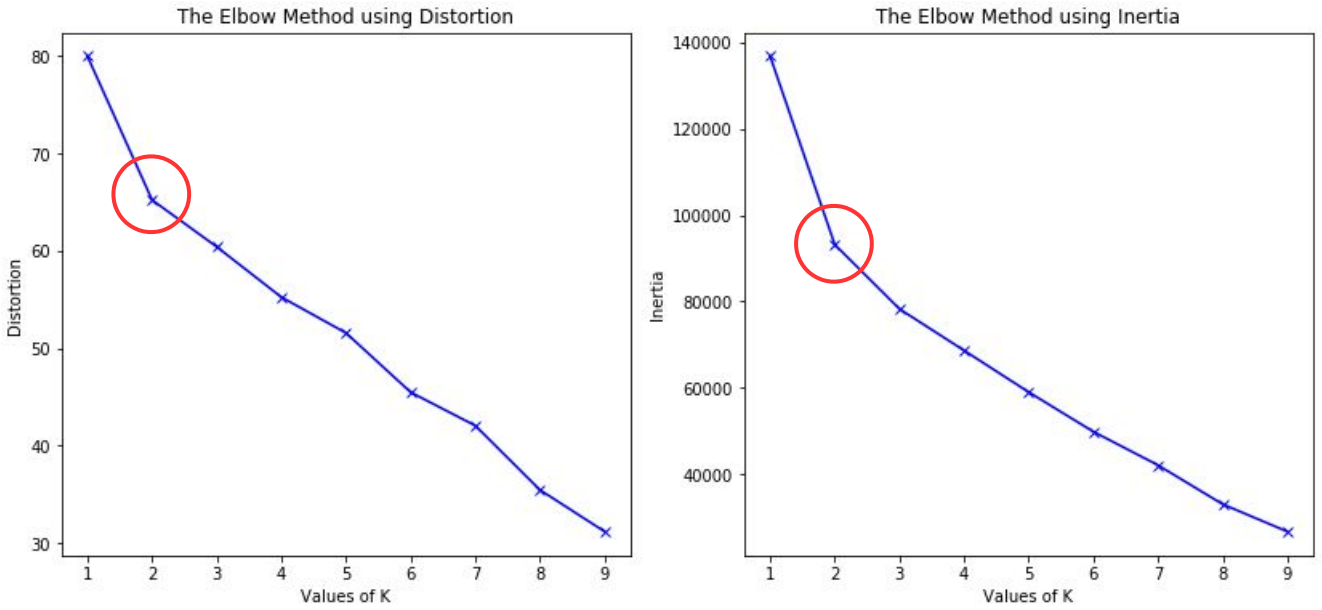
Normalizing data

- Difficult to compare absolute numbers (4 ATMs seems better than 2 ATMs, but it’s not if we’re comparing a 1000 m^2 area against another with 100 m^2)
- Solution: divide the number of observations per search area size and normalize it to a 0 - 100 range



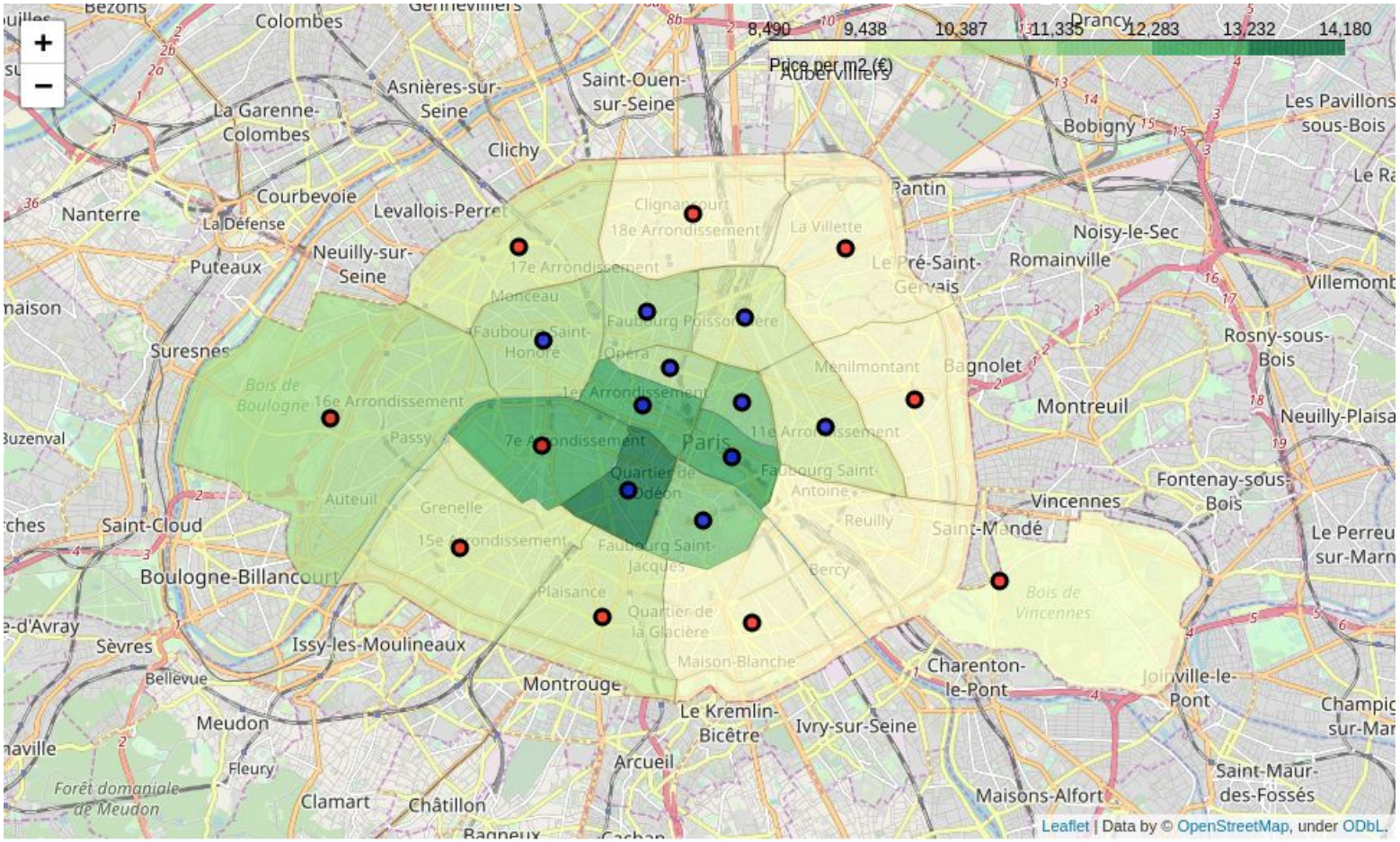
Clustering neighborhoods

- How many clusters should K-means work with? Two

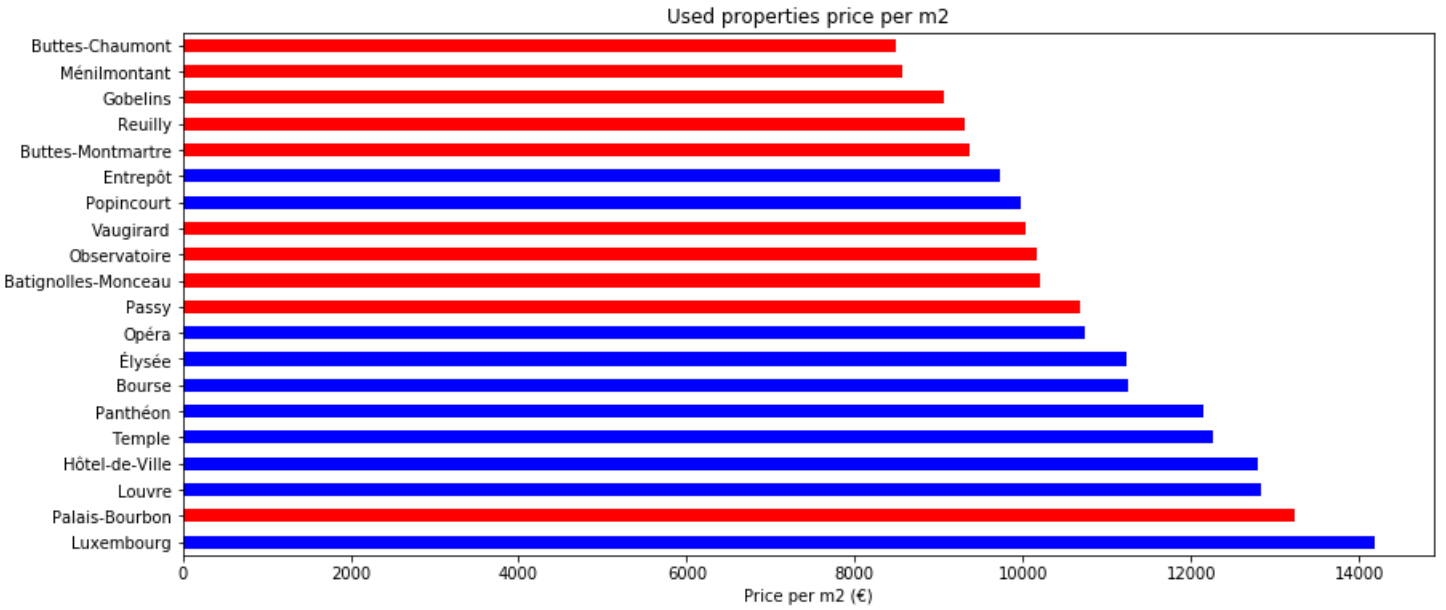


Clustering neighborhoods

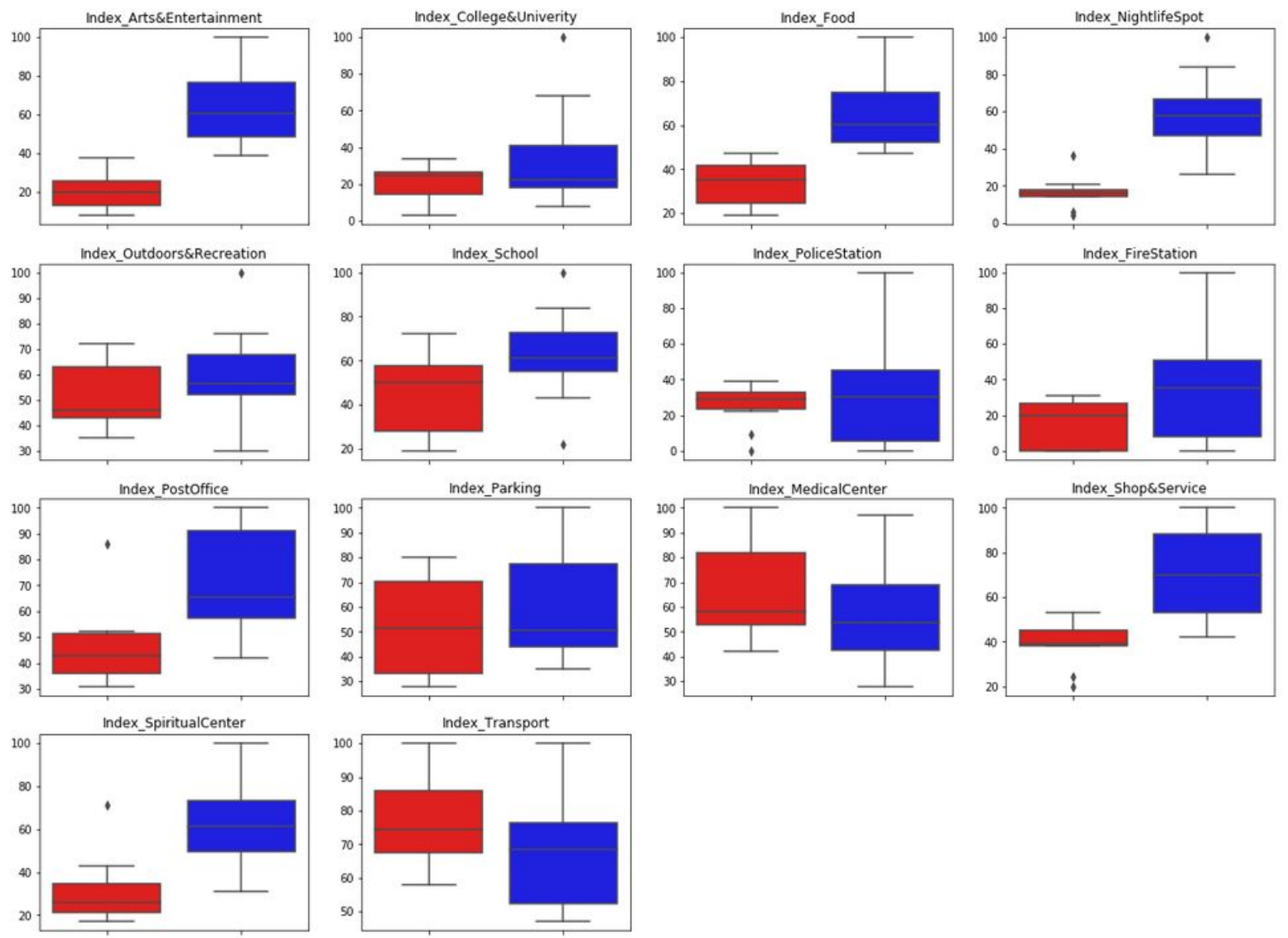
- Clusters almost classify neighborhoods as inner and outer



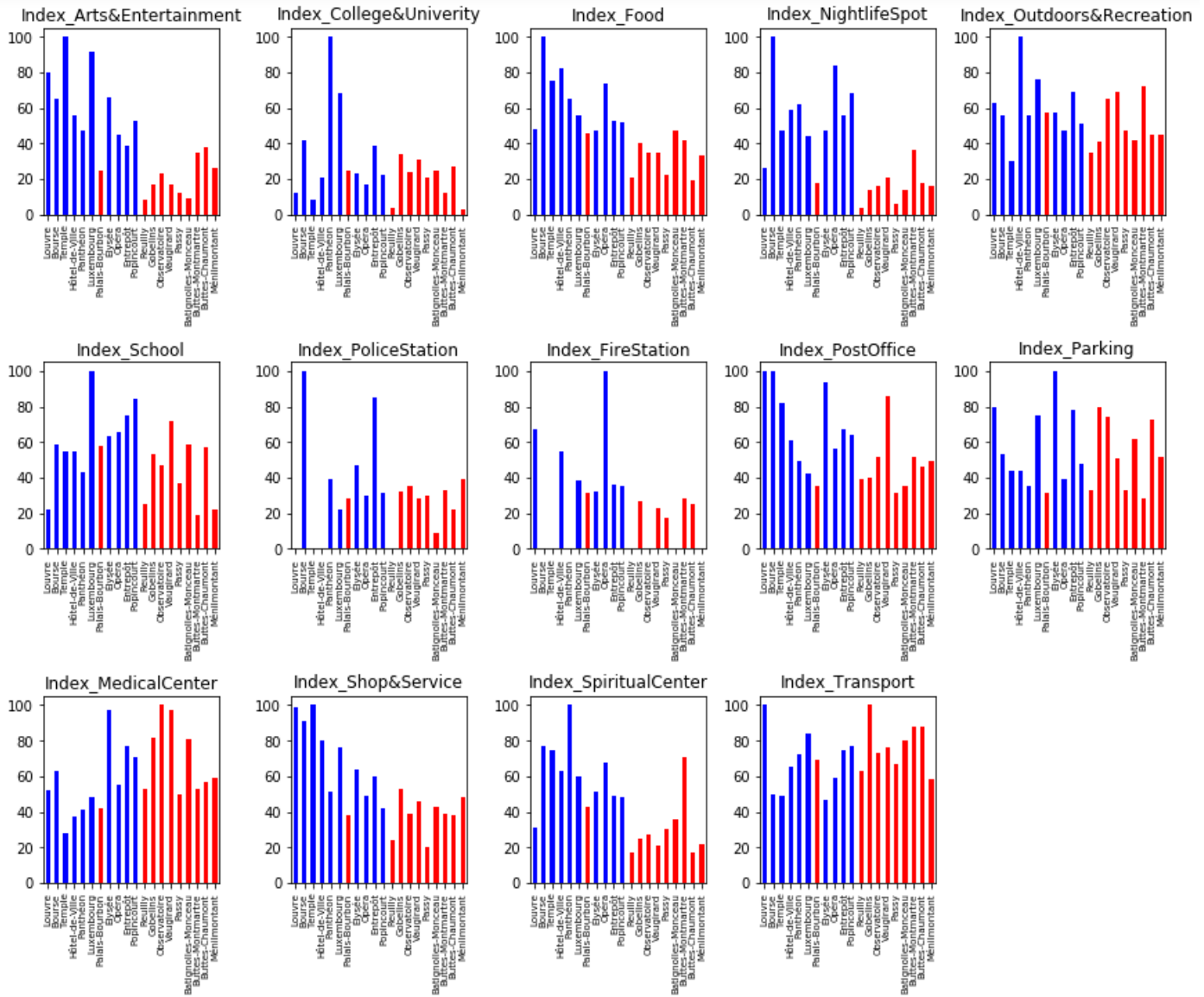
Clusters and property price



Comparing clusters

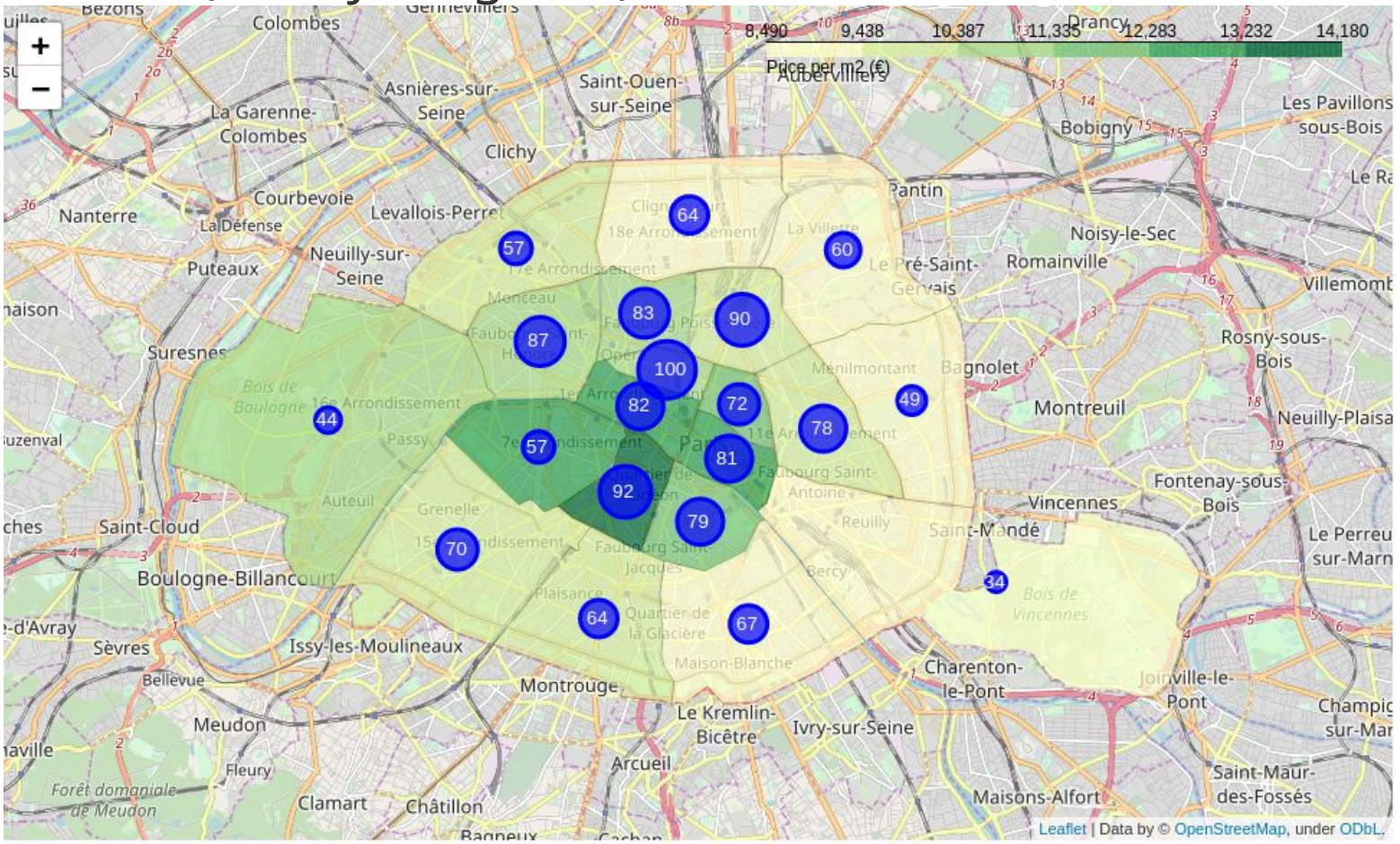


Comparing neighborhoods



Global index

- Build a global index (0-100) that summarizes all the categories at once (evenly weighted)



Service and price ranking

Neighborhood	Global_Index	Global Index Rank	PricePerM2	Price Rank
Louvre	82	6	12840	18
Bourse	100	1	11250	14
Temple	72	10	12260	16
Hôtel-de-Ville	81	7	12790	17
Panthéon	79	8	12140	15
Luxembourg	92	2	14180	20
Palais-Bourbon	57	16	13230	19
Élysée	87	4	11240	13
Opéra	83	5	10730	12
Entrepôt	90	3	9730	6
Popincourt	78	9	9980	7
Reuilly	34	20	9310	4
Gobelins	67	12	9060	3
Observatoire	64	13	10170	9
Vaugirard	70	11	10030	8
Passy	44	19	10680	11
Batignolles-Monceau	57	17	10210	10
Buttes-Montmartre	64	14	9360	5
Buttes-Chaumont	60	15	8490	1
Ménilmontant	49	18	8560	2

Neighborhood top choices

- **Bourse** is ranked 1st on the global index with 100, but when it comes to price it's just the 7th most expensive
- **Entrepôt** is ranked 3rd on the global index with 90, but it's only the 7th less expensive

Underrated neighborhoods

- **Palais-Bourbon** has a low global index of 57 (16th) but it's the 2nd most expensive. The low index is due to almost ⅓ of the neighborhood being occupied by monuments, governmental buildings and gardens, thus penalizing the index
- **Passy** has a very low global index of 44 (19th) but it's the most expensive peripheral neighborhood (10th most expensive globally). The low index is due to more than ½ of its area being occupied by the biggest green area in Paris
- **Reuilly** has the worst global index (34), but it was heavilly penalized by having more than ½ of its área occupied by a green area and because it has an highly irregular form, the radius excluded a considerable urban area

Conclusions

- This study can help understand the price variation along neighborhoods
- It provides guidelines by comparing density of services available in each neighborhood
- An improvement over this model would be the exclusion of green areas in the index calculus and an optimization of the radius for some neighborhoods (it would highly increase the complexity)
- An add-on would be receiving category's weights as an input before calculating the indexes to fit every customer needs