

Geneva House Renting Consultancy

Simone Lisi

Applied Data Science Capstone

IBM data science professional certificate

Coursera

Background

Geneva (Switzerland) is among the most expensive cities in the world to live in. Optimising assets' allocation is paramount for improving life quality. Rent, typically the largest single expense for individuals, is of utmost importance.

Goal

Helping newcomers to the city find a location to rent that satisfies their needs.

Key steps:

Getting insights on the essential services distributions around the city: Hospitals, schools, transports, parks, markets.

Identifying most relevant features determining the price of a property: structural properties? Nearby locations?

Data

How will we collect relevant data?

Available properties: by scraping the popular listing website [immoscout24.ch](https://www.immoscout24.ch) for a list of the available properties and their key features, noteworthy: position, price per month, surface.

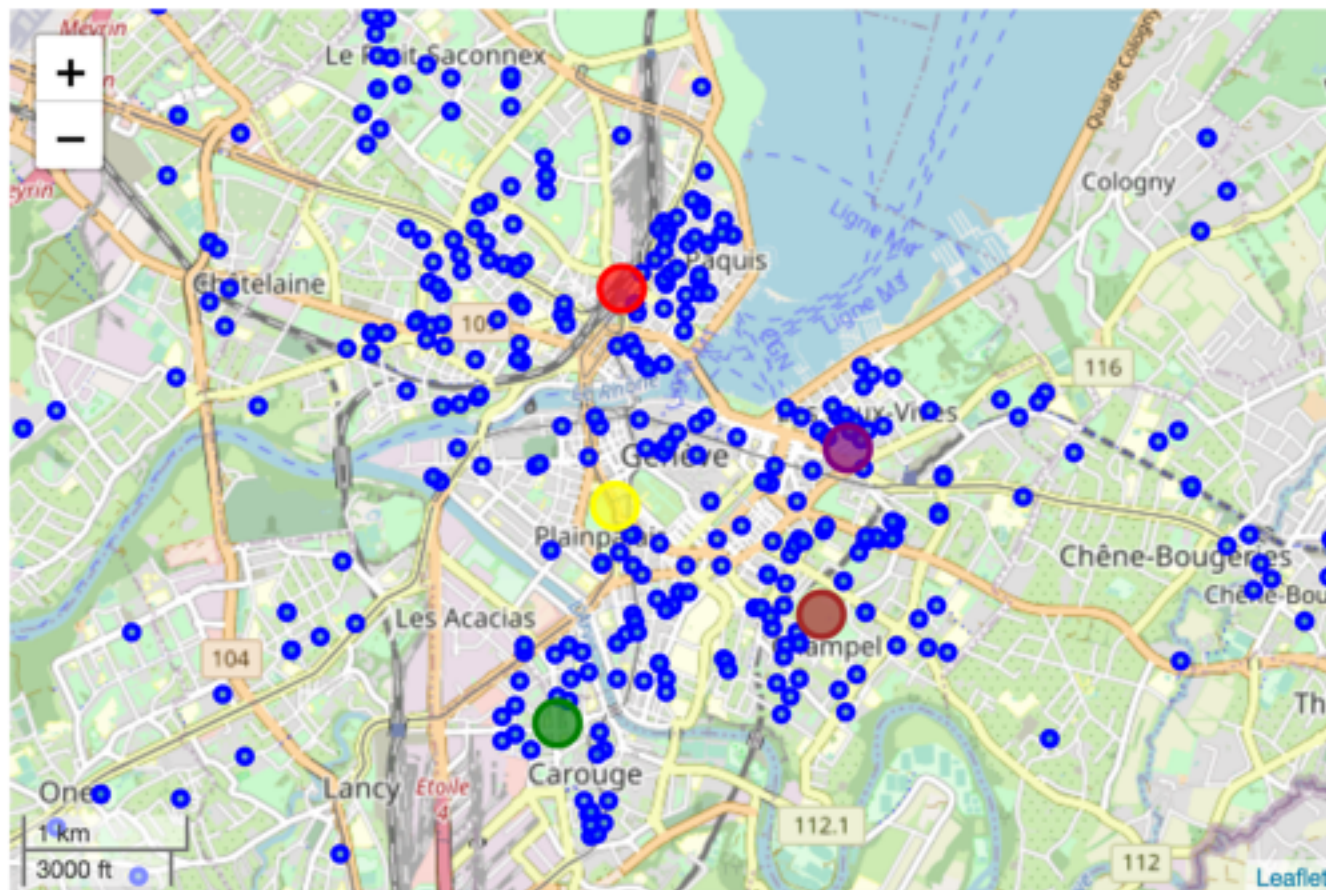
Essential services: by querying the foursquare.com database.

Nearby locations: by querying the foursquare.com database, looking for the top locations in a 500m radius within each property position.

Properties features and geo-distribution

	cityName	extraPrice	price	latitude	longitude	numberOfRooms	street	surfaceLiving	zip	pricePerRoom	pricePerSqm
6065071	Genève	350.0	2550.0	46.19363	6.16698	4.0	Rue Le Corbusier 18	85.0	1208	637.500000	30.000000
6046525	Genève	200.0	4200.0	46.19646	6.16090	5.5	Route de Malagnou 40E	110.0	1208	763.636364	38.181818
6063101	Genève	170.0	2810.0	46.20429	6.15891	4.0	Rue du 31-Décembre 16	60.0	1207	702.500000	46.833333
6054808	Genève	250.0	4750.0	46.19523	6.15391	7.0	Rue Marignac	165.0	1206	678.571429	28.787879
6037279	Le Lignon	200.0	2700.0	46.20313	6.09613	6.0	Avenue du Lignon	107.0	1219	450.000000	25.233645

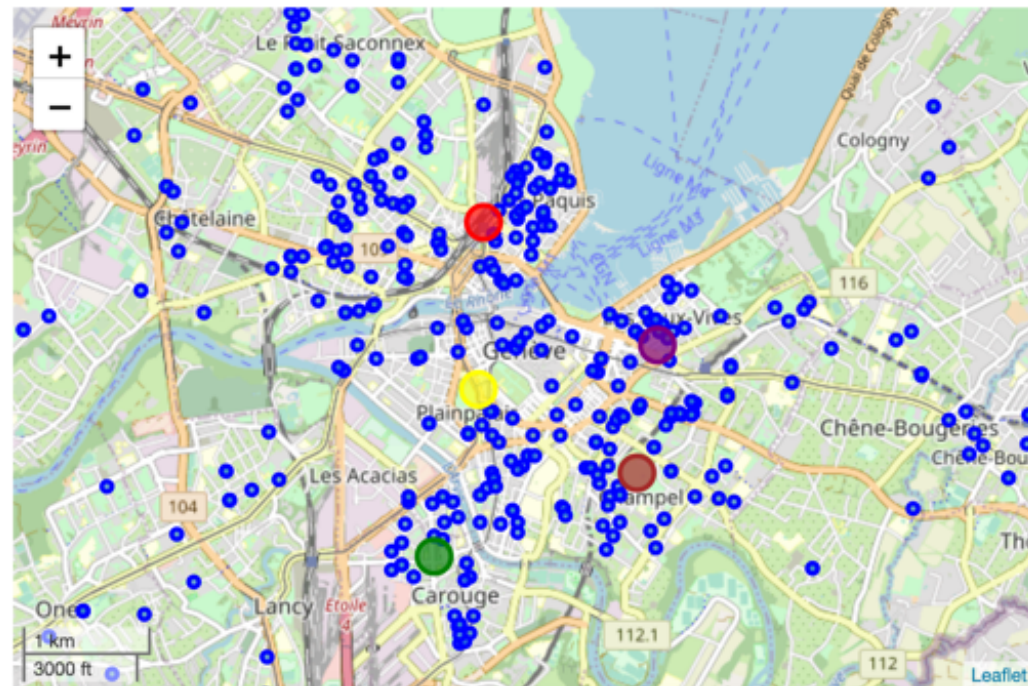
Properties



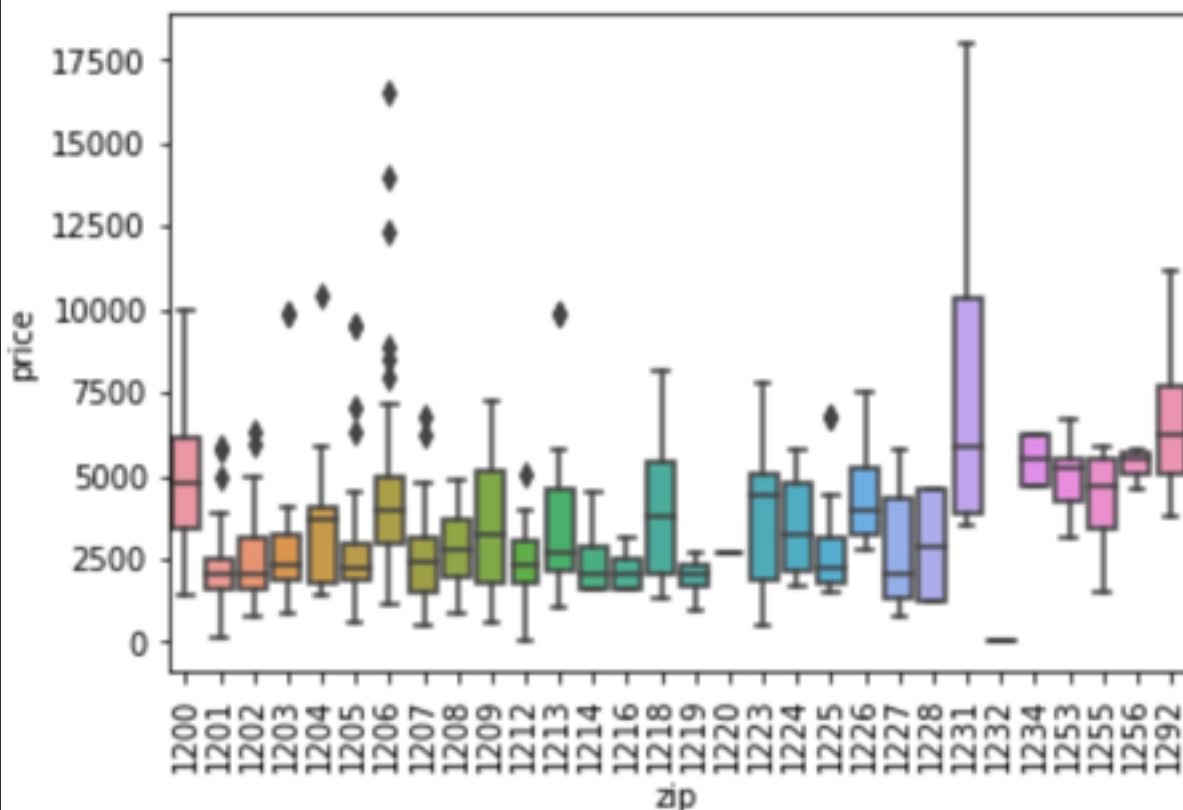
A map of the currently available properties (blue) and their location around the key spots: central railway station (red), the central neighbourhoods of Plainpalais (yellow), Les Eaux-Vives (purple), Champel (brown) and the nearby town of Carouge (green).

Price position dependency

Properties



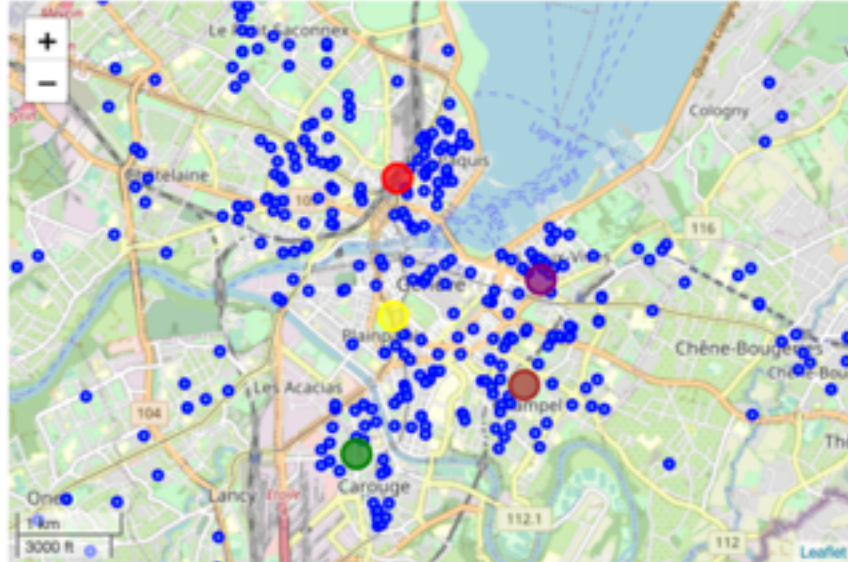
A map of the currently available properties (blue) and their location around the key spots: central railway station (red), the central neighbourhoods of Plainpalais (yellow), Les Eaux-Vives (purple), Champel (brown) and the nearby town of Carouge (green).



The currently available data do not show meaningful price differences among different areas around the city.

Services distribution

Properties



Transports



Parks



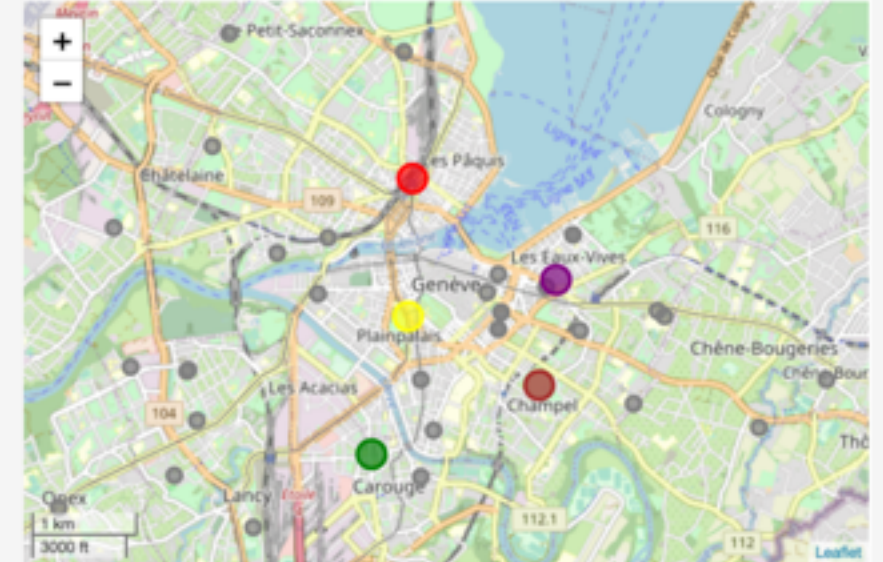
Hospitals



Markets



Schools

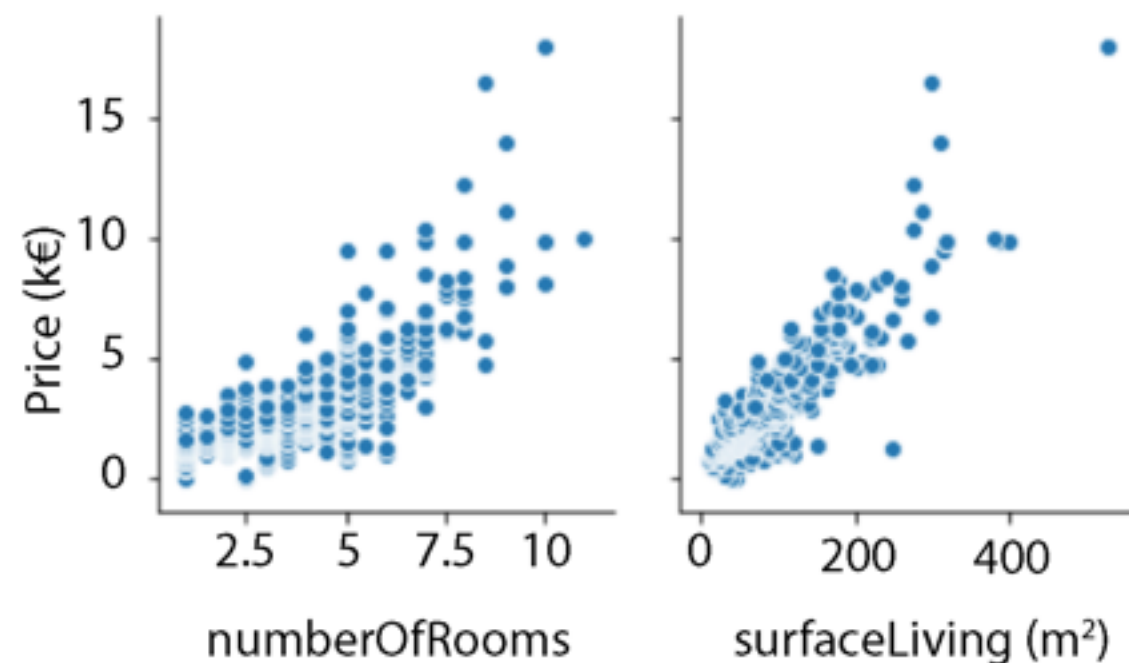


Public transportations efficiently connects the outskirts to the city centre. Parks, markets and schools are scattered all around the city. Medical centres can be found mainly in the city centre.

Key insight n°1: the relatively small area of the city, the even distribution of services and the capillary public transport network could explain the seemingly homogenous (or low fluctuation) in the prices around the city

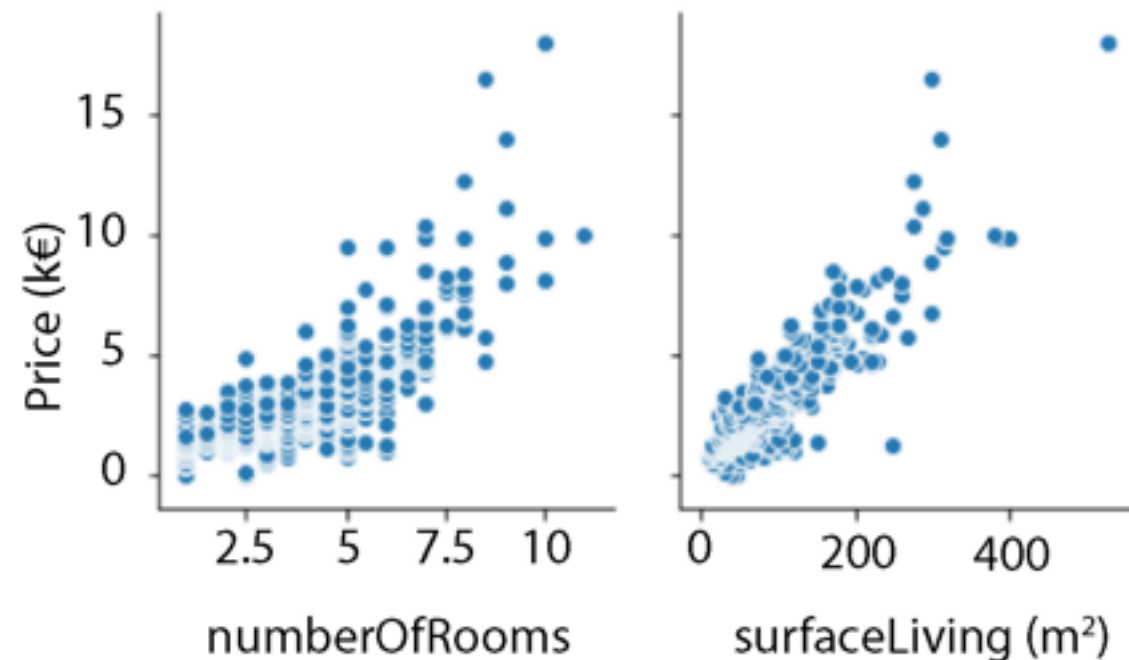
Best price predictor

	cityName	extraPrice	price	latitude	longitude	numberOfRooms	street	surfaceLiving	zip	pricePerRoom	pricePerSqm
6065071	Genève	350.0	2550.0	46.19363	6.16698	4.0	Rue Le Corbusier 18	85.0	1208	637.500000	30.000000
6046525	Genève	200.0	4200.0	46.19646	6.16090	5.5	Route de Malagnou 40E	110.0	1208	763.636364	38.181818
6063101	Genève	170.0	2810.0	46.20429	6.15891	4.0	Rue du 31-Décembre 16	60.0	1207	702.500000	46.833333
6054808	Genève	250.0	4750.0	46.19523	6.15391	7.0	Rue Marignac	165.0	1206	678.571429	28.787879
6037279	Le Lignon	200.0	2700.0	46.20313	6.09613	6.0	Avenue du Lignon	107.0	1219	450.000000	25.233645



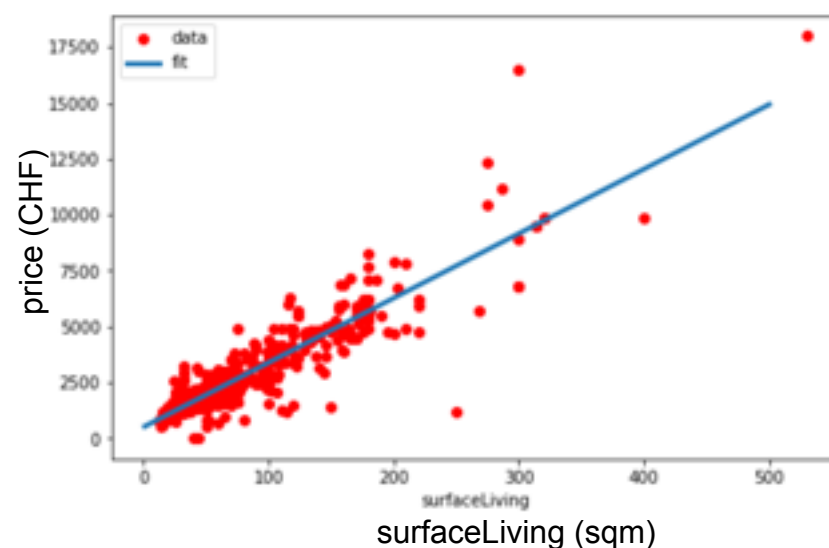
According to available informations per property, size is the best price predictor. Both n° of rooms (numberOfRooms) and surface (surfaceLiving) are good candidate as predictor. The latter offers the most straightforward model: a liner price vs surface dependence

Best price predictor

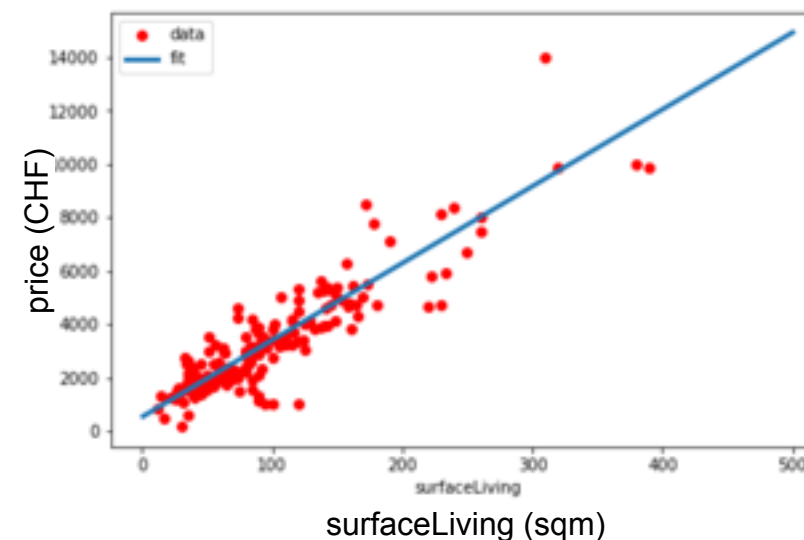


Among the available informations per property, size is the best price predictor. Both n° of rooms (numberOfRooms) and surface (surfaceLiving) are good candidate as predictor. The latter offers the most straightforward model: a liner price vs surface dependence

Train set



Test set



Intercept: 650 CHF; Slope: 27 CHF/sqm

R2-score: 0.77

Key insight n°2: the price of a property is mainly determined by its surface.

Clustering properties by nearby venues

1) Collect the nearby venues for each property (a sample for one of the properties is shown here)

2) Find all categories and subcategories associated to each venue

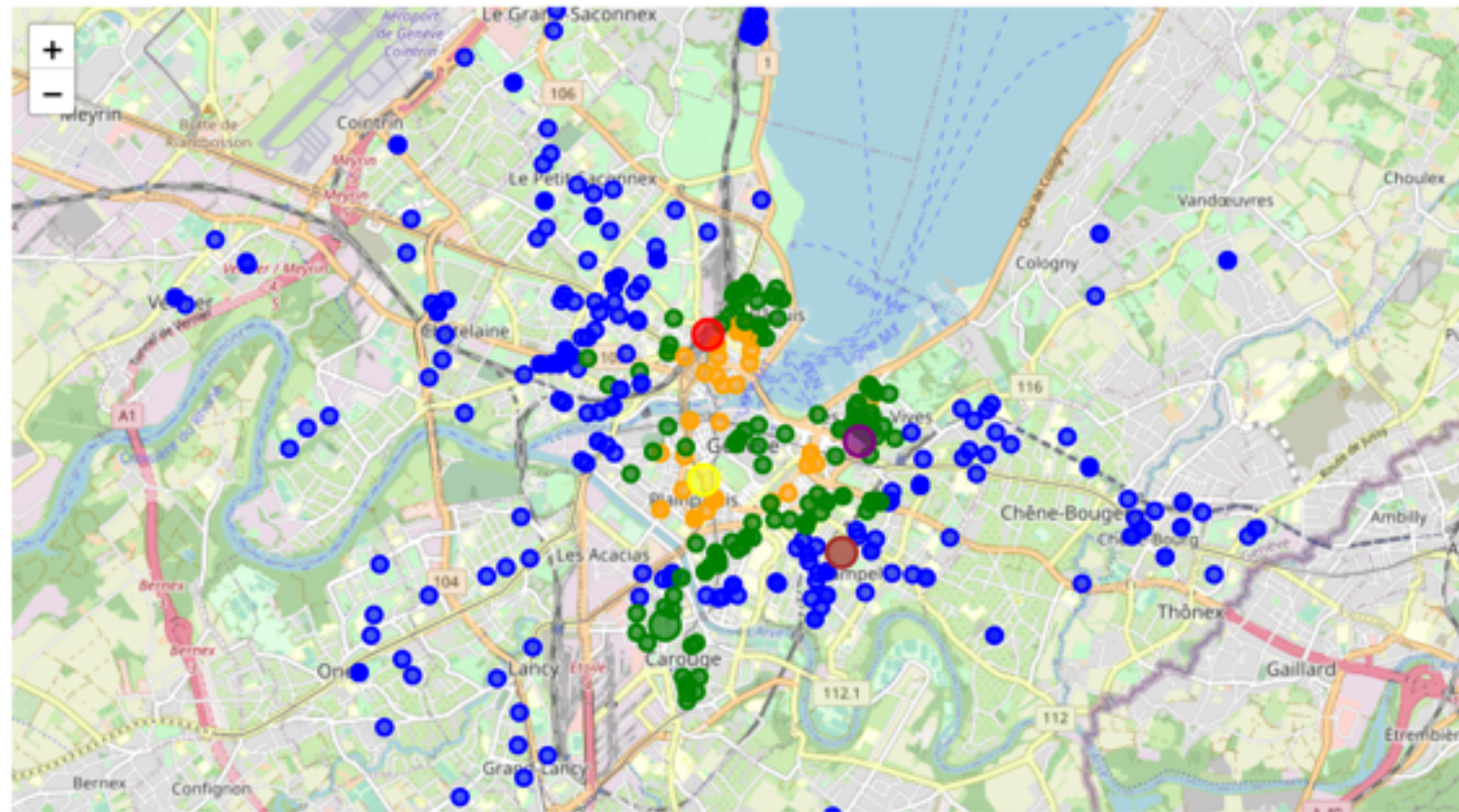
	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	categoryId	Lv0	Lv1	Lv2	Lv3
0	46.19589	6.14108	Marché de Plainpalais	46.197183	6.141728	Farmers Market	4bf58dd8d48988d1fa941735	Shop & Service	Food & Drink Shop	Farmers Market	None
1	46.19589	6.14108	Inglewood	46.194008	6.140400	Burger Joint	4bf58dd8d48988d16c941735	Food	Burger Joint	None	None
2	46.19589	6.14108	Café-chocolaterie Berger	46.195047	6.142642	Tea Room	4bf58dd8d48988d1dc931735	Food	Tea Room	None	None
3	46.19589	6.14108	Mangia Bene	46.195303	6.140693	Italian Restaurant	4bf58dd8d48988d110941735	Food	Italian Restaurant	None	None
4	46.19589	6.14108	Tartare&Co	46.196738	6.142011	French Restaurant	4bf58dd8d48988d10c941735	Food	French Restaurant	None	None

3) add to our properties database the number of venues for each category and use these features for k-means clustering (n_kmeans = 3)

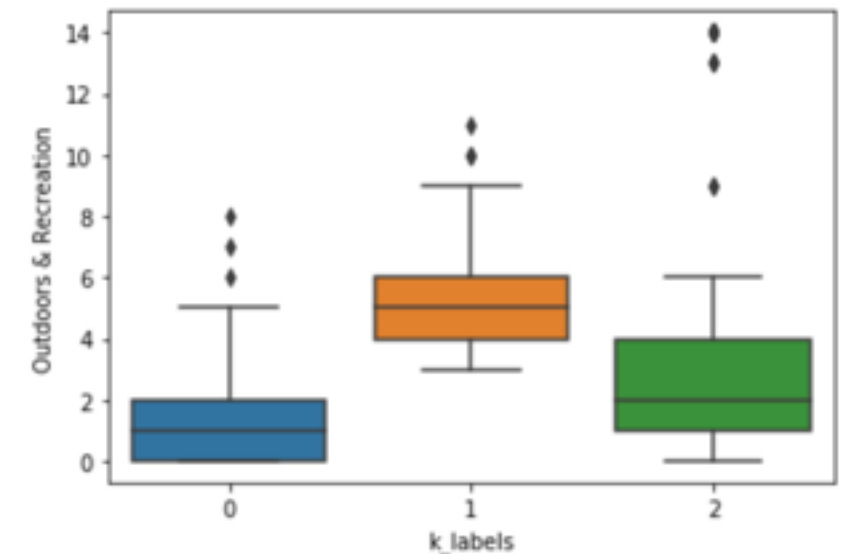
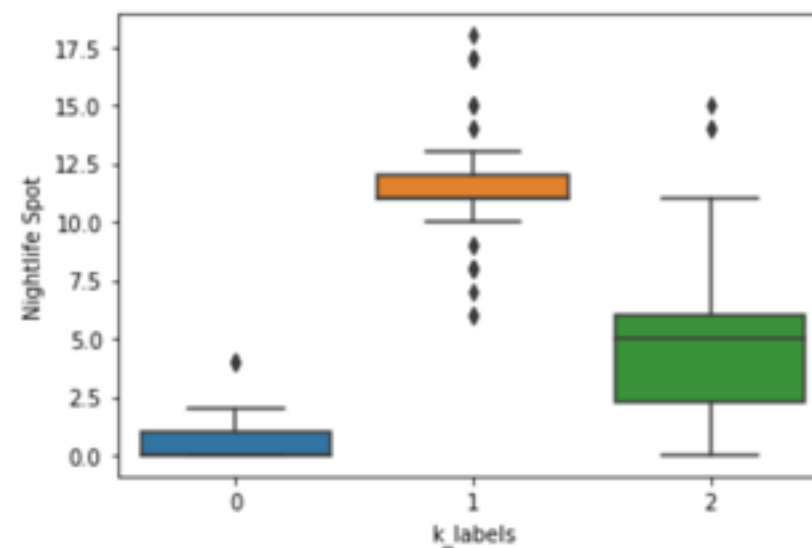
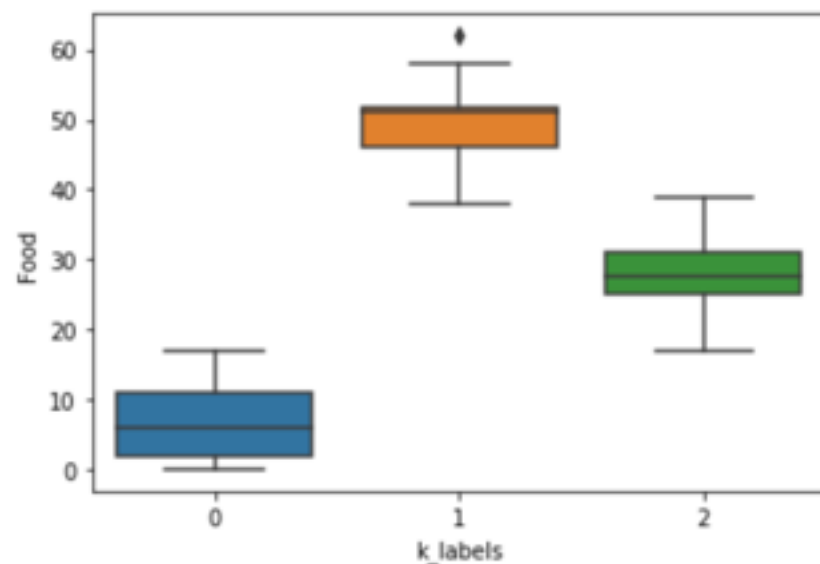
	Food	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot	Shop & Service	Travel & Transport	Professional & Other Places
6038912	19	5	5	8	2	5	0
6054808	19	2	2	1	2	0	0
6037279	3	1	0	0	1	0	0
5912116	9	1	1	1	5	3	1
5937875	6	2	1	1	1	4	0

Clustering results

Clustered Properties

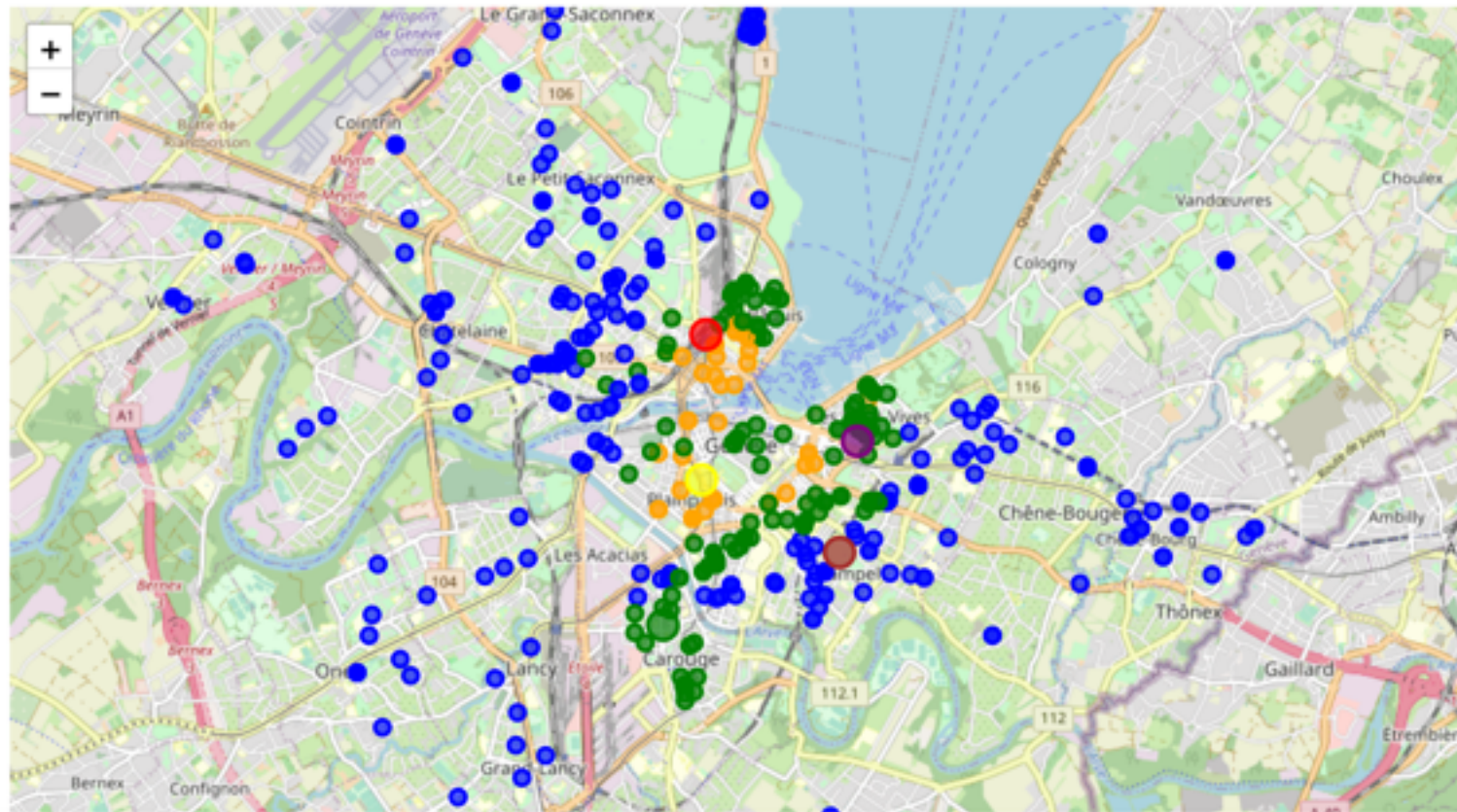


Properties in cluster n°1 (orange) have significantly higher number of nearby venues in the 'Food', 'Nightlife Spot', and 'Outdoors & Recreation' categories, followed by the ones in cluster n°2 (green).

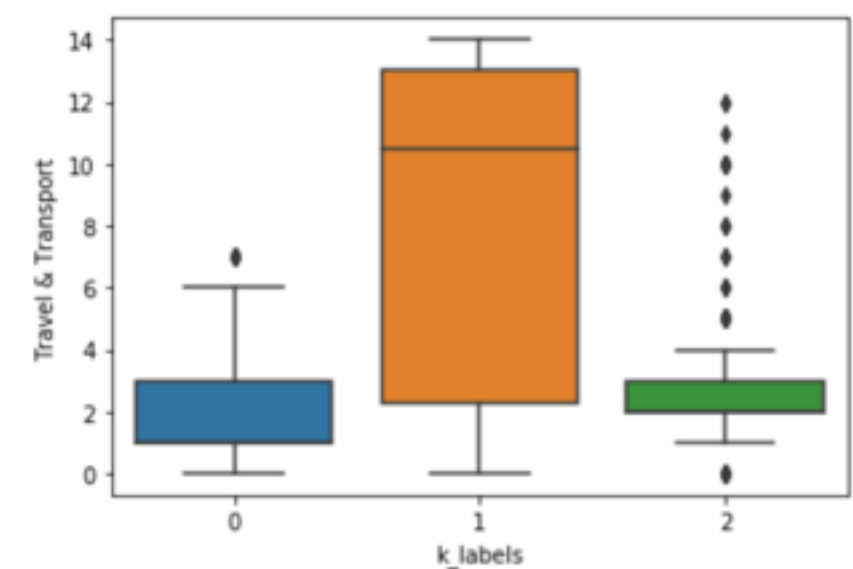
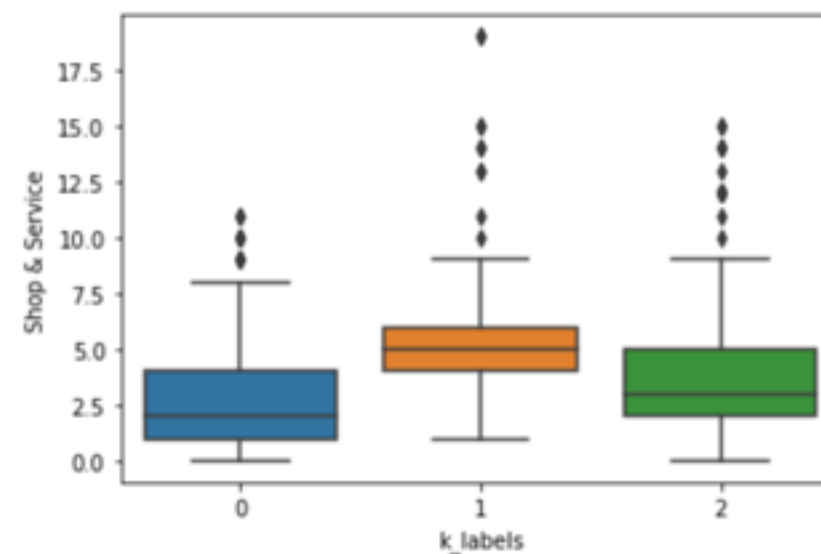
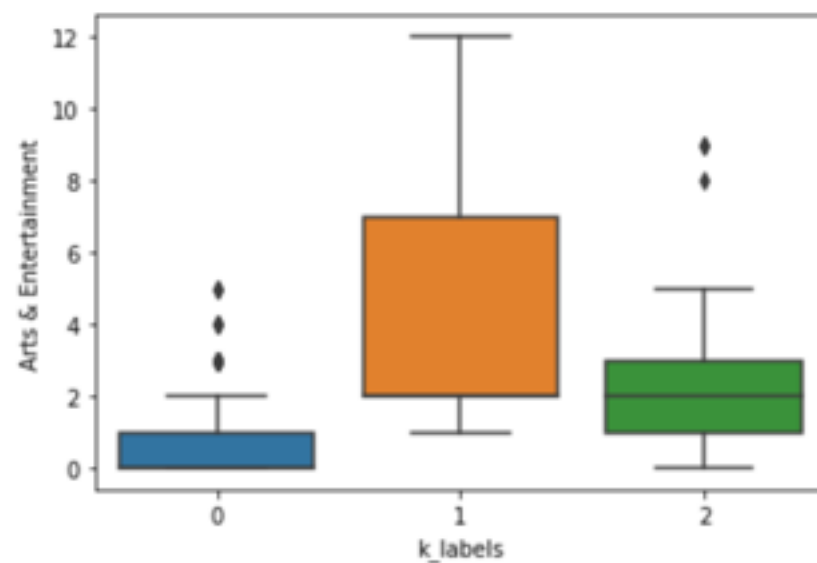


Clustering results

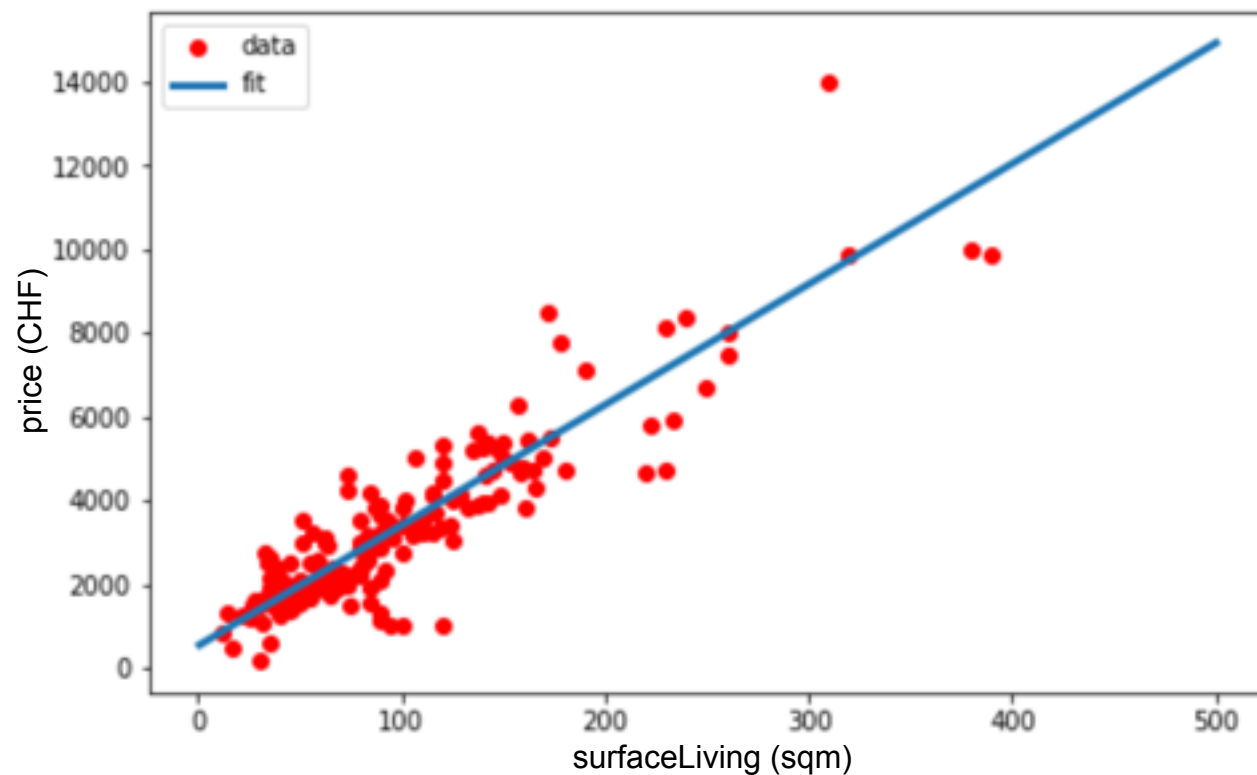
Clustered Properties



A similar pattern, with less differences among the clusters, is found in the remaining categories

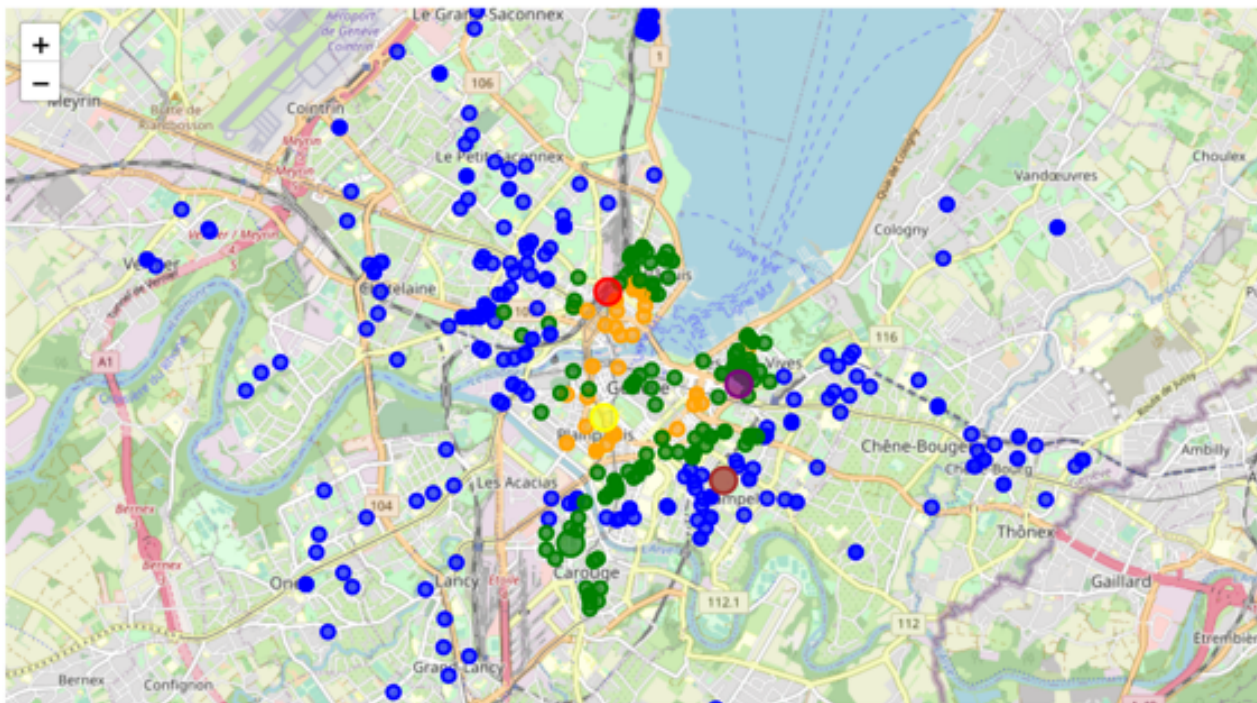


Advices for our customers



1) A clear guideline for determining whether a property is being offered for a fair price.

Clustered Properties



2) The rental price for properties throughout the city is rather homogeneous. The choice of where to live should be mainly driven by personal preferences. We would suggest a customer looking for a lively neighbourhood to look for properties in clusters n°1 and n°2, while cluster n°0 would be more befitting for people looking for quieter places.

Perspectives

Regular data collection over time will provide a deeper understanding of the rental prices in Geneva.

Information that could not be detected, e.g. a slight price differences between neighbourhoods, in the present analysis could become evident (or definitively ruled out) with a more voluminous dataset.

Thank you for your kind attention!