Exploring Paris to Make an Informed Relocating Decision

May 23rd 2020

> contents

- introduction and business problem formulation
- data sources
- methodology
  - visualizing districts and neighbourhoods
  - grouping rent data by district
  - visualizing rent and sales data
- results
  - is it better to buy or to rent ?
  - does population density affect real estate value ?
  - what are the most common venues for each neighborhood?
  - where are similar neighborhoods, and what do they have in common ?
- discussion
- conclusion

> introduction and business problem formulation

- Every year a number of people relocates to Paris, the French capital and global center of finance, fashion, arts, science, culture, and history.
- However, making an optimal choice when relocating can be overwhelming. Which district should be chosen ? Is it better to buy or to rent an apartment? A number of factors including renting vs. purchasing price, as well as characteristics of the neighborhood must be taken into account.
- This project aims to create a tool to help make informed decisions based on a comparison of available data : rent prices, sales prices, population density and location data using Foursquare.
- Different users will be drawn to different conclusions, according to their preferences and financial capabilities.

> data sources

A number of sources can be used to gather relevant information about Paris

- https://parisdata.opendatasoft.com/ : rental data (80 neighborhoods)
- https://www.data.gouv.fr/ : administrative districts (20 arrondissements).
- https://cadastre.data.gouv.fr/ : property sales data in 2019, per district.
- https://developer.foursquare.com/ : location data.
- https://fr.vikidia.org/ : population data.

The data was gathered using web-scraping, making calls to the Foursquare API, and downloading geojson files.

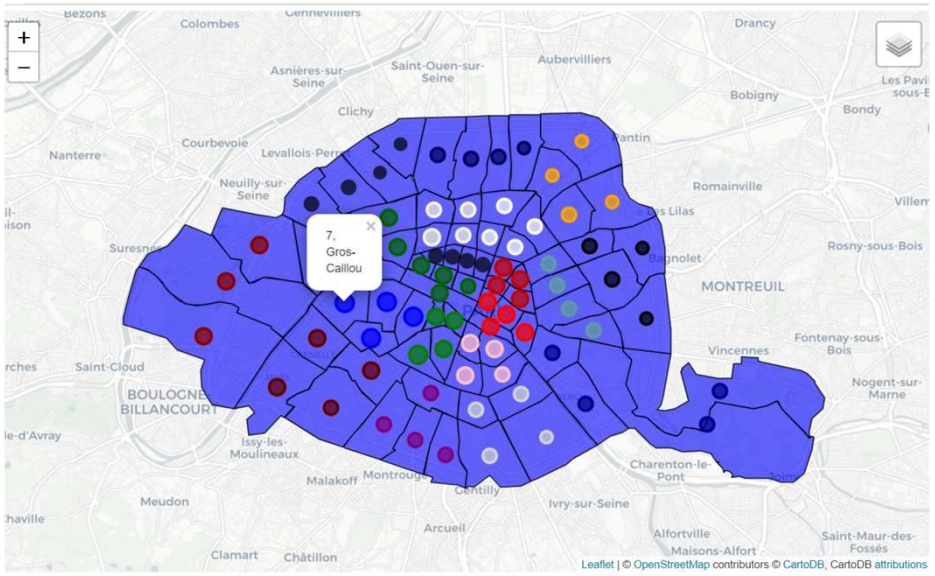
Minor data cleansing was done including adding renaming columns, adding district column names, getting rid of duplicates, deleting unnecessary columns.

> methodology

- The majority of our exploratory data analysis focused on using longitude and latitude information to make maps and other visuals in order to better understand where neighborhood, districts, and other points of interests were located.
- We utilized choropleth maps, dot maps, scatter plots, heat maps and content tables.
- We also compared real estate rent and sale values, which are supposed to be correlated with each other, identifying districts having significant deviations.
- Relevant data sets were available on a per neighborhood basis or on a per district basis. Reconciliation of the different dataset was necessary prior to analysis.

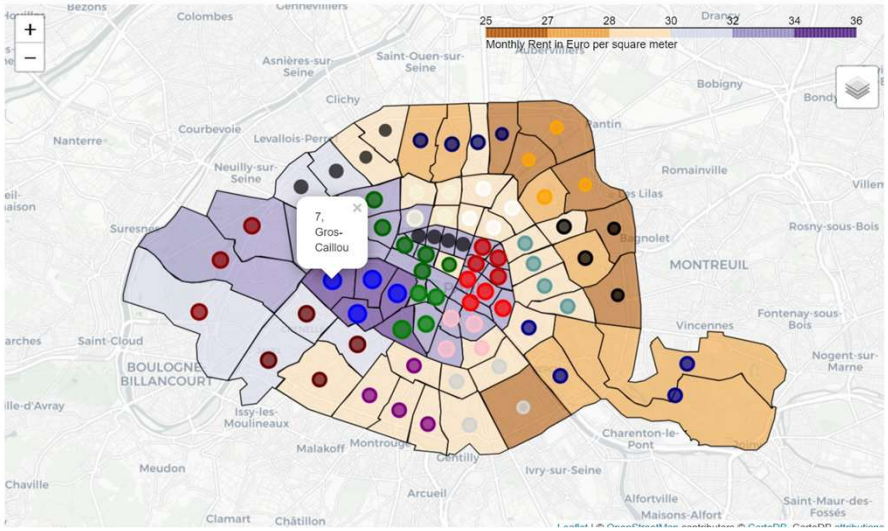
> visualizing districts and neighborhoods

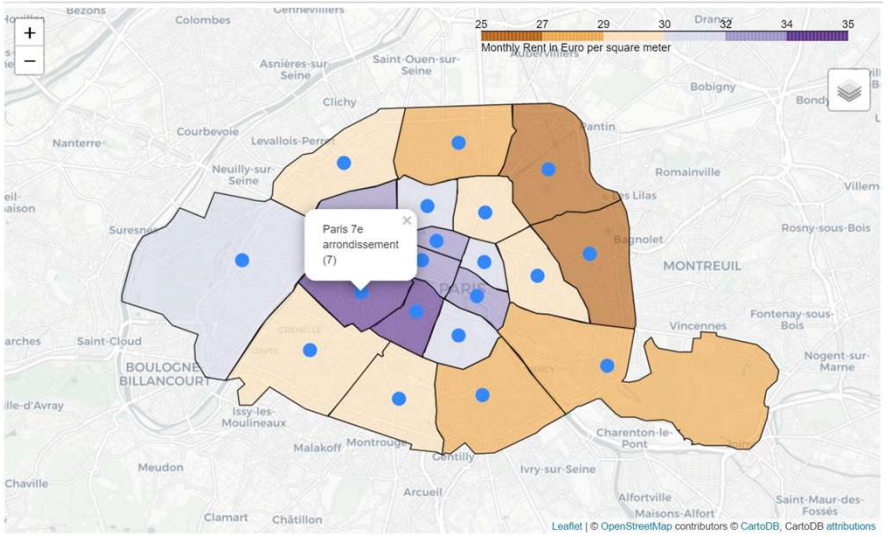
- First of all, we need to understand where districts and neighborhoods are located
- We used choropleth maps for rent data / sales data and dot maps to identify the district, with a specific color code per district.
- We can see that the Paris district numbers are arranged in a clockwise spiral from the center outwards.



> grouping rent data by district

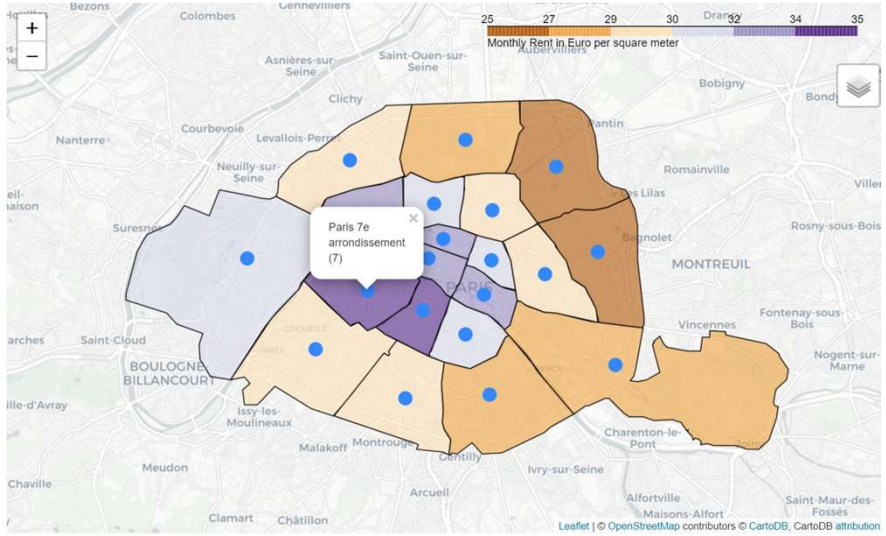
- Neighborhood rent data were grouped by district.
- This representation was necessary to compare with sales data, which are available on a district basis.

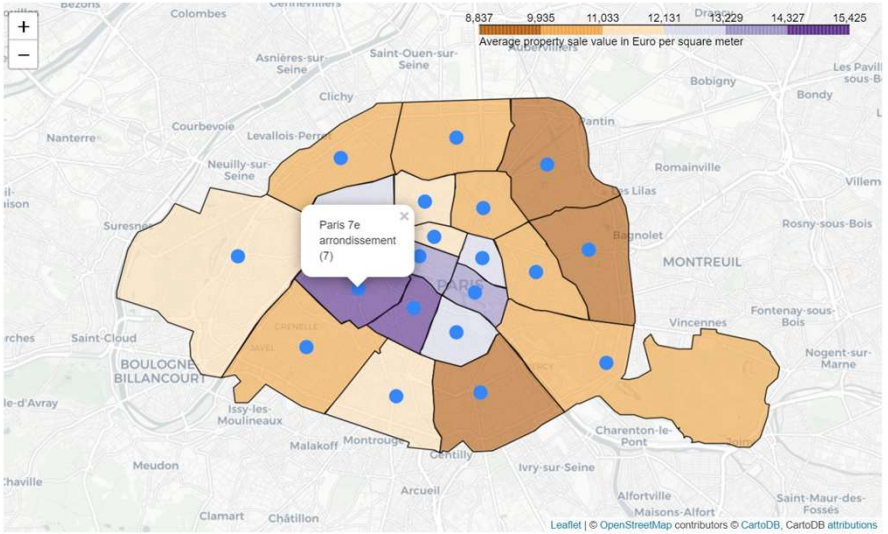




> visualizig rent and sales data

- A consistent trend is observed in rent and sales data, with central and western districts being generally more expensive than eastern districts.
- A more detailed analysis is necessary to make an informed decision of renting vs. buying.



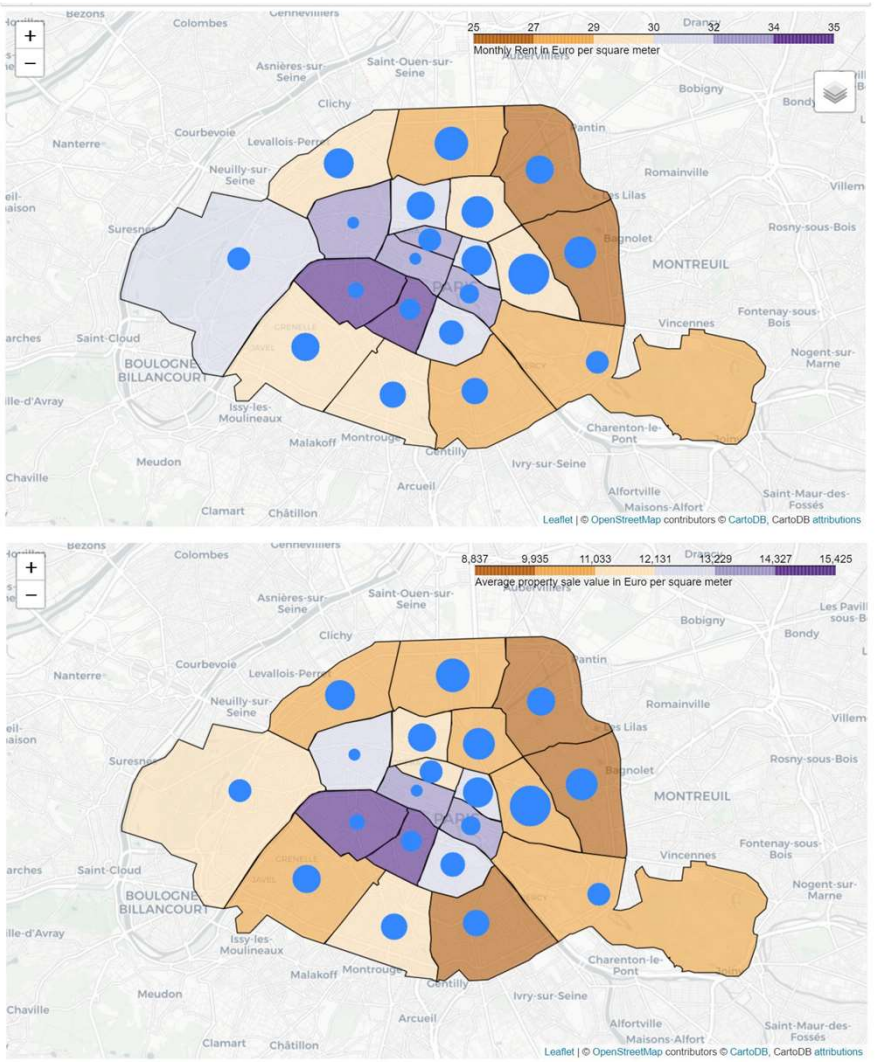


> is it better to buy or to rent ?

- Rent value per unit surface and Sales value per unit surface are linearly correlated, as one could expect.
- Above the trendline, it seems more reasonable to rent than to buy an apartment. For instance, such is the case for District n. 6, which has the highest sales value per square meter.
- Below the trendline, it seems more reasonable to buy than to rent. For instance, such is the case for district n. 15, 16 and 2, which have a relatively low sales value compared to the rent value.
- This analysis is only preliminary. Before making any buying decision, it should be completed with a more exhaustive investigation concerning financial aspects such as one's capacity to obtain a mortgage, interest rates, trend analysis on sales data in the past year and the forecasted overall economic outlook in Europe, France and Paris.
- Also, the peculiar characteristics of the neighborhood should be taken into account.

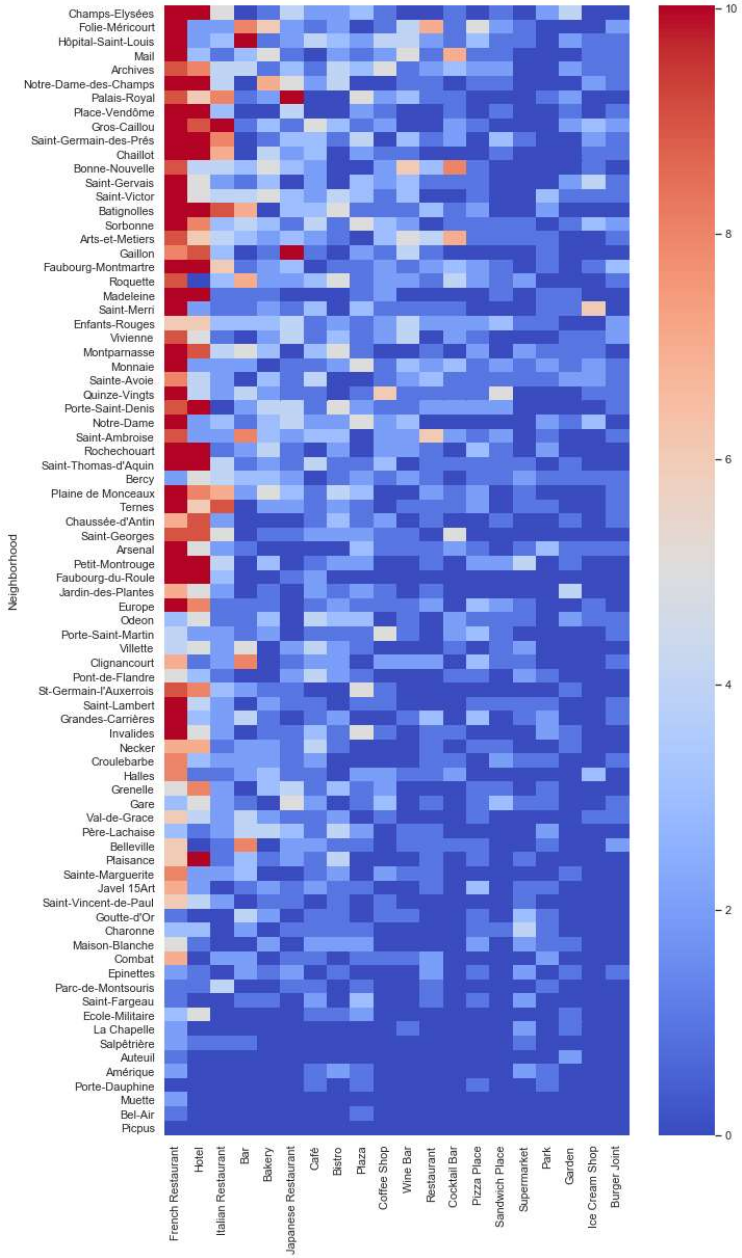
> does population density affect real estate value ?

- These choropleth maps represent property value, with the blue lables proportional to population density.
- The most expensive districts in terms of property value are also the least populated.
- A number of reasons including accessibility to services and proximity to tourist and business venues could explain this trend.
- In order to gain more insight, location analysis is necessary.



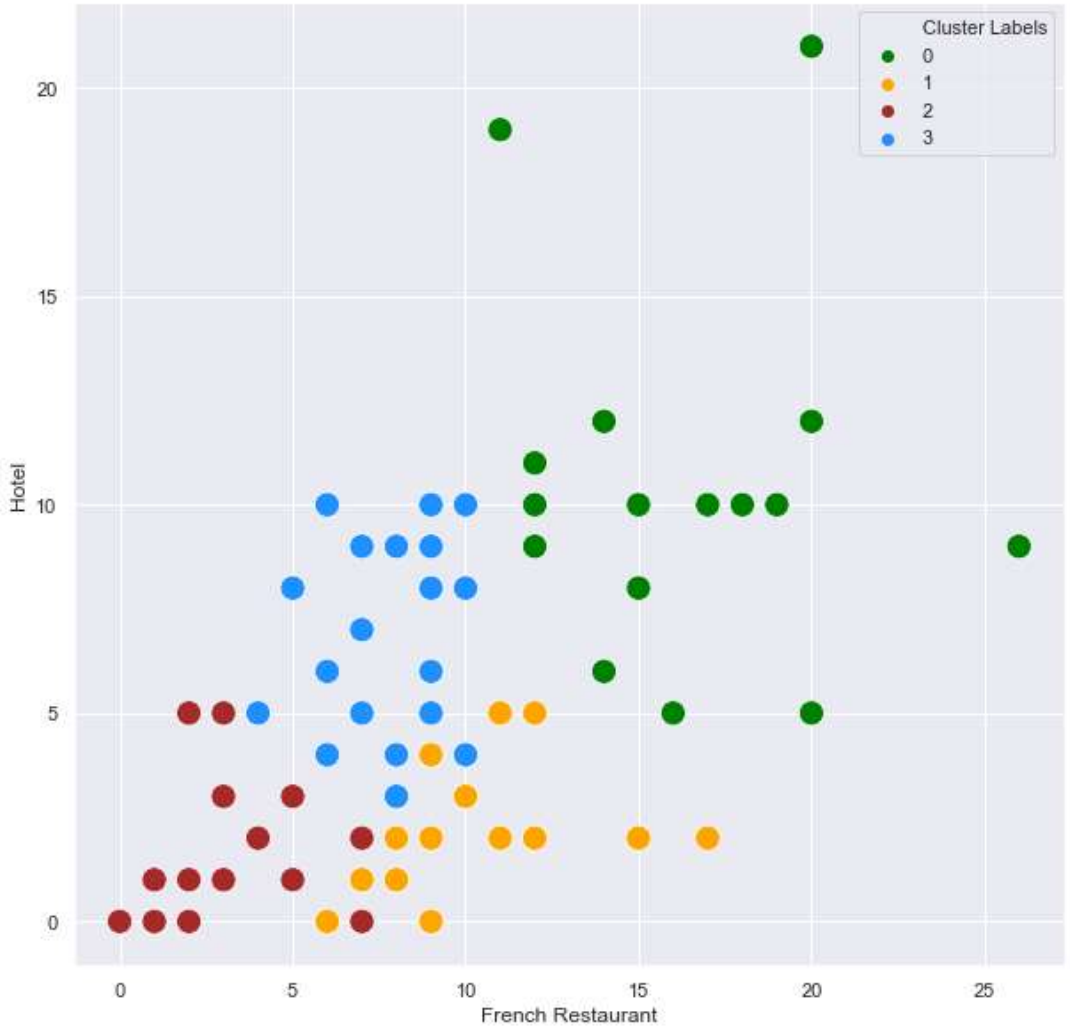
> what are the most common venues ?

- The 80 neighborhood in Paris were analysed using Foursquare API. A total of 296 different venue categories were found. For the quantitative analysis we used the first 20 most popular venue categories for each neighborhood.
- Neighborhoods and venues were ordered by total number of venues.
- Overall, the most common location in paris is French Restaurant, followed by Hotels and Italian Restaurants.
- A heatmap is a compact and efficient way to identify relevant venues. For instance, if you are looking for a Japanese restaurant, the two neighborhoods to go to are Palais Royal and Gaillon.



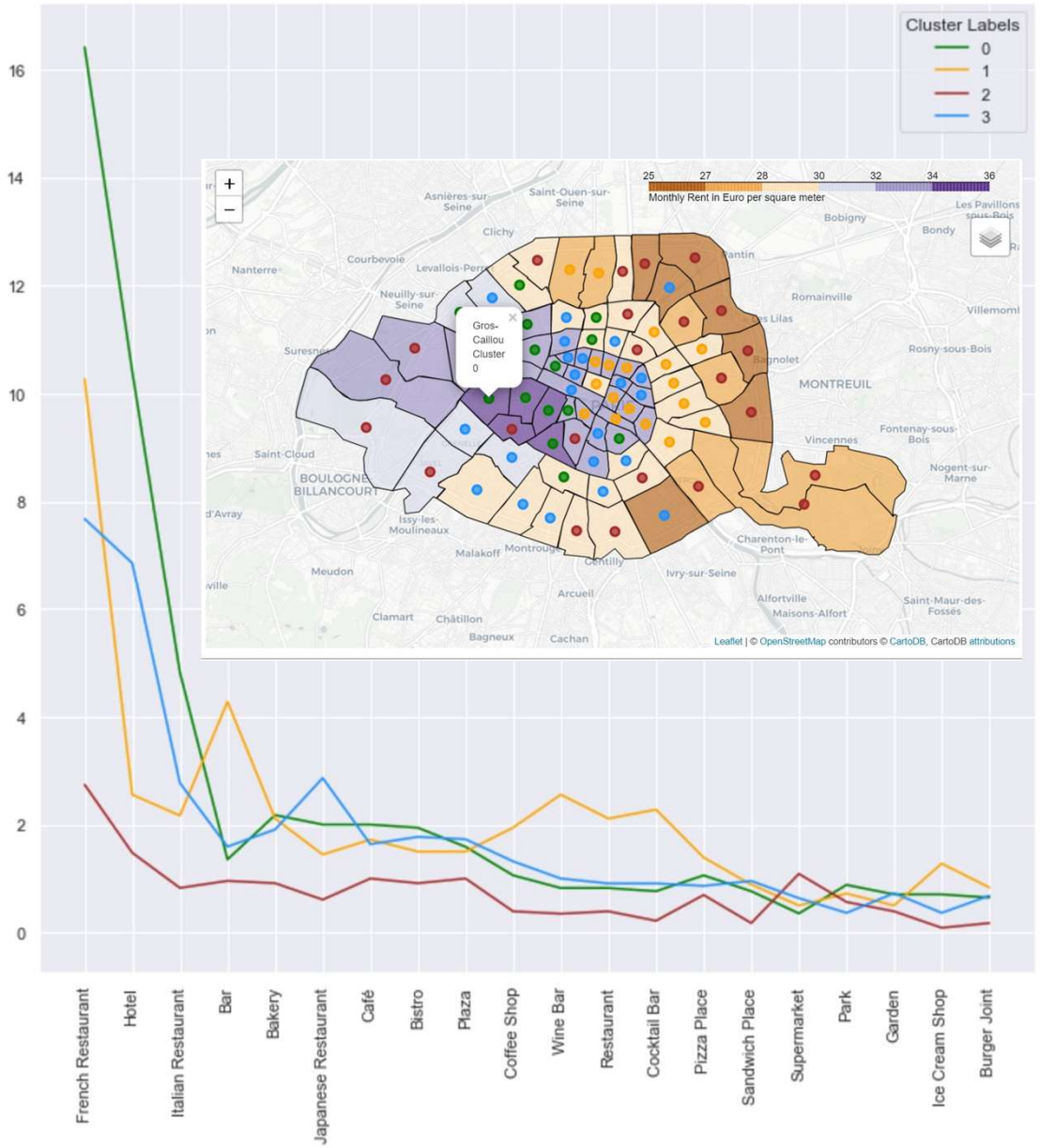
> where are similar neighborhoods, and what do they have in common ?

- The 80 neighborhood in Paris were clustered into 4 groups based on the most popular venues, using K-means.
- As an example, the first two most significant venues, French Restaurants and Hotel, are considered here. The scatter plot represents each neighborhood with a circle having the color of the cluster it belongs to.
- Based on centeroid analysis we could define a name cluster :
  - Cluster 0: Exclusive : Glamour and Tourism
  - Cluster 1: Going-out
  - Cluster 2: Residential
  - Cluster 3: Business



> discussion

- A map with location and characteristics of each cluster, superposed to the real estate value choropleth, is presented. Based on this representation, we can immediately recognize subgroups. For instance, cluster 2 “Residential” can be further divided into East (less wealthy) and West (more wealthy).
- All data were aggregated in a single dataset, that can be further used to make optimal choices, given a decision vector which captures the weight of each criterion.
- More data could be aggregated to this dataset in order to further refine the choice: crime rate, distance from work, characteristics of the family to be relocated (e.g. availability of special schools for children), ethnic and sociological characteristics of the neighborhood, distance from significant cult-related venues, financial capability, etc. This enriched dataset could pave the way for an online service providing an overall real estate scoring based on its location.
- While the incorporation of more data in our model could provide additional insight, this goes beyond the scope of the present project.



> conclusion

- Several criteria come into play when relocating. This analysis took into account property value, population density and location analysis to provide relevant criteria to make an informed relocation choice.
- All data were aggregated in a single dataframe, that can be used to make optimal choices, for a given a decision vector. This dataset was used to create a visual representation of the location and characteristics of each cluster, superposed to the related real estate value.
- Based on this representation, when relocating to Paris, an informed choice can be made very easily.
- Further work can expand on this analysis to include other relevant criterias. The enriched dataset could be further developed into an online service providing multi-criteria real estate scoring based on location.

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	...	11th Most Common Venue	12th Most Common Venue
0	2	Amérique	French Restaurant	Supermarket	Bistro	Plaza	Pool	Bed & Breakfast	Café	Park	...	Smoke Shop	Exhibit
1	3	Archives	French Restaurant	Hotel	Coffee Shop	Italian Restaurant	Art Gallery	Bistro	Bar	Japanese Restaurant	...	Tapas Restaurant	Cocktail Bar
2	1	Arsenal	French Restaurant	Hotel	Park	Seafood Restaurant	Plaza	Italian Restaurant	Gastropub	Boat or Ferry	...	Cocktail Bar	Vegetarian / Vegan Restaurant
3	1	Arts-et-Metiers	French Restaurant	Cocktail Bar	Hotel	Wine Bar	Italian Restaurant	Restaurant	Bar	Coffee Shop	...	Chinese Restaurant	Moroccan Restaurant
4	2	Auteuil	Tennis Court	Garden	Stadium	Sporting Goods Shop	Outdoors & Recreation	Botanical Garden	Racecourse	Museum	...	English Restaurant	Ethiopian Restaurant