Finding the right neighbourhood to move to in Paris

Francesco Siciliano,
14/07/2019

1. Background

- Paris, as any other large city, has a wide variety of neighbourhoods, each with its **peculiarities and points of interests** → it is not easy to determine which is the **best place** to live in when you are not familiar with it;

- This project aims at providing a user looking for a new **apartment** a **supporting tool** allowing to look for some zones that satisfies the user requirements, allowing to narrow down the research to fewer zones of the city.

- Knowing beforehand where your potential home is located is essential to be more confident in your choice and avoiding **troublesome situations** (e.g. leaving the house after a short time) when looking for new houses → help in making an **informed decision**.

2. Data Management

❖ Sources:

Open Data Paris:

- *Arrondissements* and quartiers data (location, names, postal codes etc.);
- Clean, ready to use .json and .geojson files;
- Used to locate quartiers and venues.

Foursquare API:

- Venues;
- Need to search through .json file output to retrieve venues categories and compare its occurrence frequencies;
- Used for clustering algorithm.

TomTom API:

- Traveling times, distances and routes;
- .xml format → need of scraping to retrieve data;
- Used to find best candidate in terms of traveling times.

3. Exploratory Data Analysis

- ❖ Search by category is too specific and does not narrow-down the search scope → clustering is used to perfom a partition of locations based on shared features;

- ❖ Large percentage of 'Restaurant' type-venues → Clustering bias, with large group based on restaurants → restaurant-type categories neglected in the search;

- ❖ 'Hotel' type-venues neglected, as not significant for the search of a full time accommodation;

4. Neighbourhoods Clustering

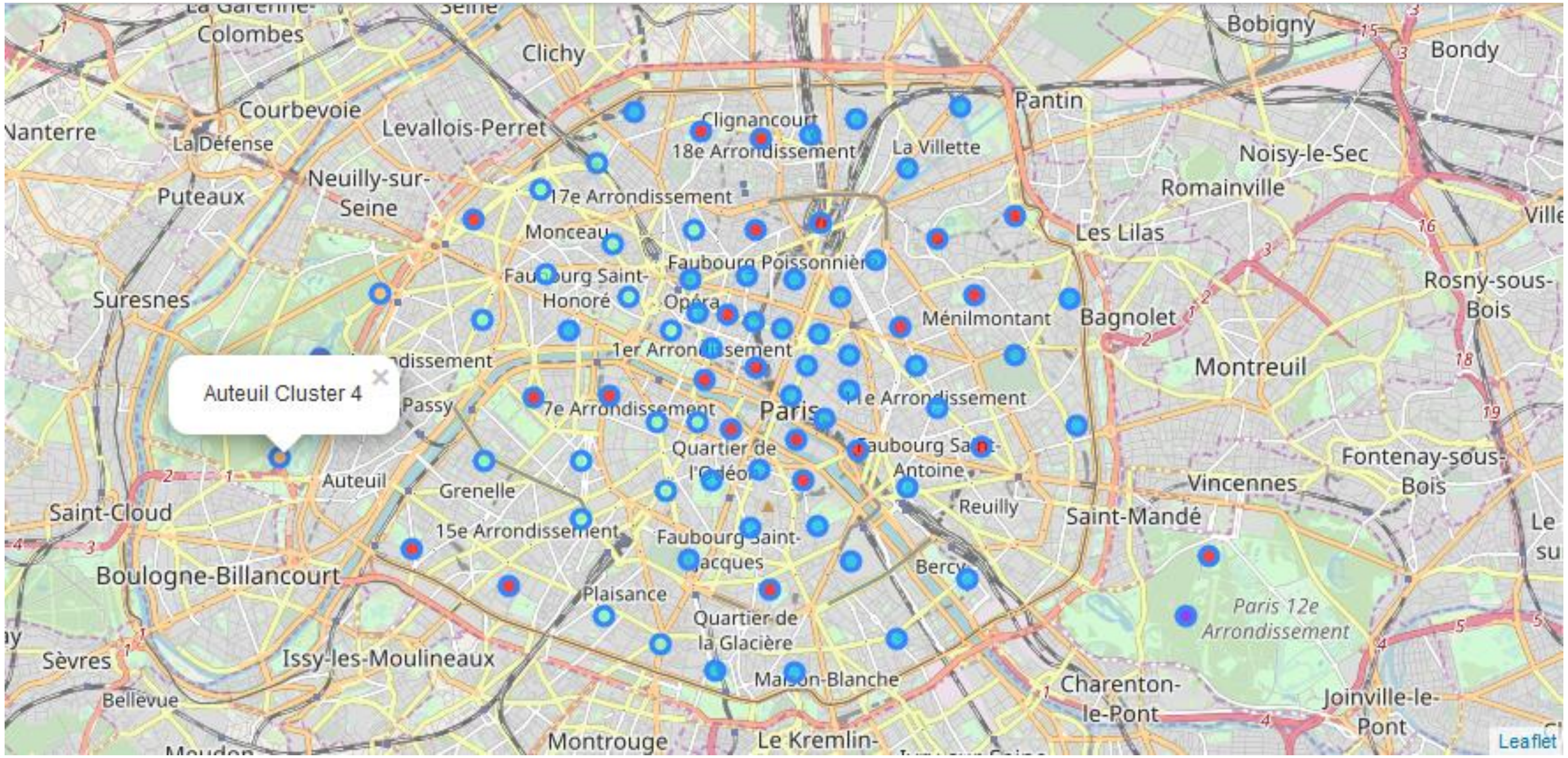
- *K-means clustering* algorithm;
- Partitioning in **5** mutually exclusive clusters;
- Venues sorted in one of the 5 clusters;
- Result are 5 **dataframes** grouping the different clusters;
- User has the possibility to choose among one of the cluster according to her preferences;

	Arrondissement	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	1	0	Café	Plaza	Exhibit	Coffee Shop	Historic Site
1	1	0	Pizza Place	Wine Bar	Café	Clothing Store	Bakery
4	2	0	Bistro	Wine Bar	Salad Place	Plaza	Creperie
13	4	0	Ice Cream Shop	Bakery	Wine Bar	Plaza	Coffee Shop
14	4	0	Plaza	Gastropub	Historic Site	Bakery	Park
19	5	0	Bakery	Bar	Wine Bar	Museum	Bistro
20	6	0	Plaza	Wine Bar	Bistro	Café	Bar
24	7	0	Café	Coffee Shop	Ice Cream Shop	Bistro	Pizza Place
25	7	0	Plaza	Café	History Museum	Art Museum	Cultural Center
35	9	0	Bakery	Bistro	Coffee Shop	Music Venue	Bar
37	10	0	Wine Shop	Coffee Shop	Theater	Grocery Store	Food & Drink Shop
42	11	0	Bar	Bakery	Wine Bar	Bistro	Coffee Shop
43	11	0	Bar	Café	Wine Bar	Bistro	Coffee Shop
44	12	0	Plaza	Playground	Sports Club	Recreation Center	Café
49	13	0	Bar	Park	Bakery	Indie Movie Theater	Gaming Cafe

Dataframe related to cluster 0.

5. Data Visualization (1/3)

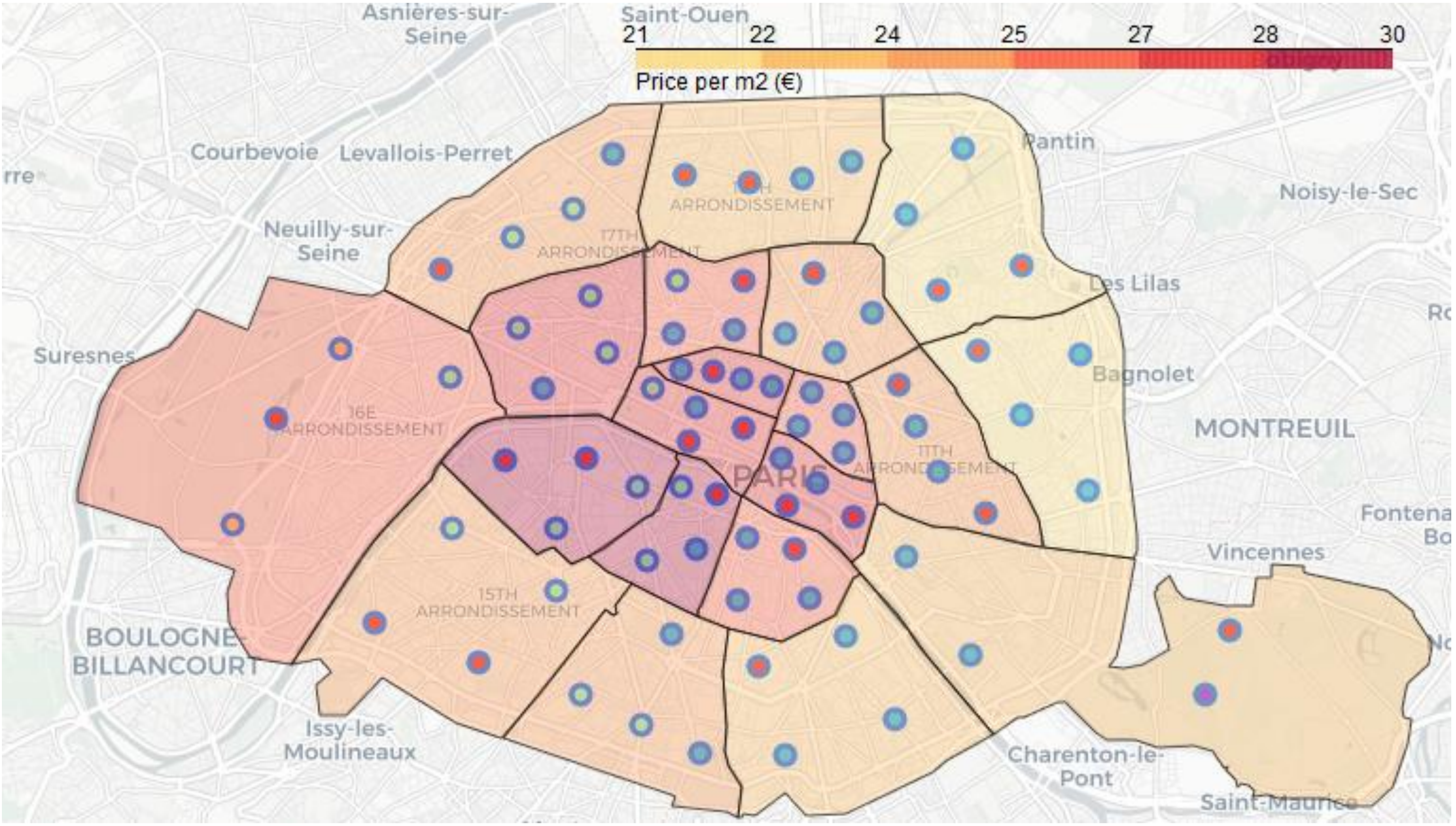
- ❖ Each quartier belongs to a specific cluster;
- ❖ Evenly spread throughout Paris.



Map of Paris with clusters elements. By clicking on each element, the administrative quartier name and cluster label of belonging are shown.

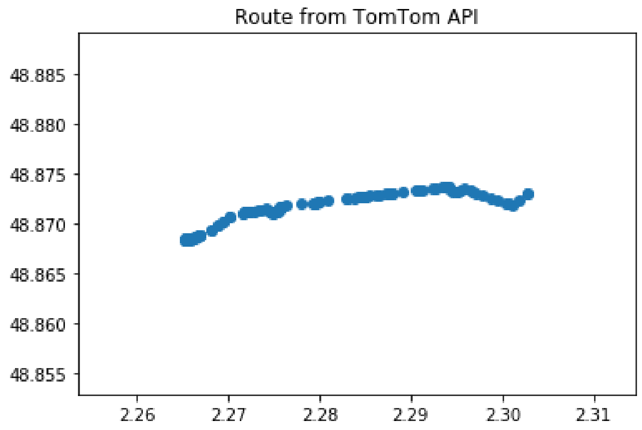
5. Data Visualization (2/3)

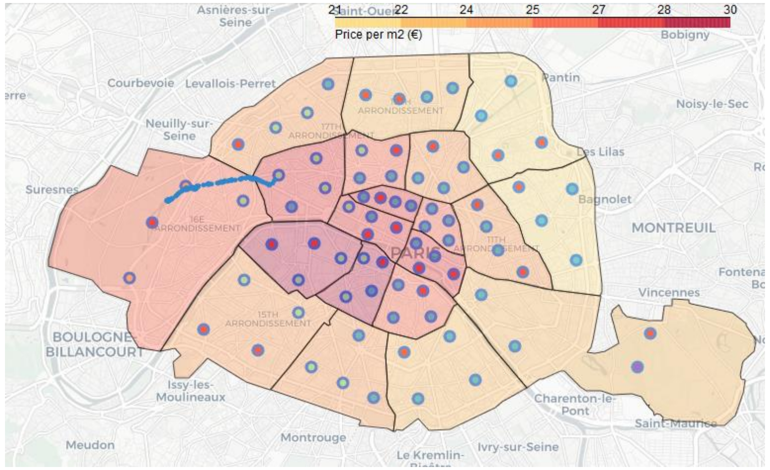
- ❖ Heatmap showing **average rent prices** per *arrondissement,*
- ❖ Intuitive and fast way to **keep user aware** of the prices of the arrondissement he will be looking at;
- ❖ No in depth information of prices → Comparative understanding.



Map of Paris with arrondissements colored according to their price range (the price is referred to the cost an apartment in terms of €/m2).

5. Data Visualization (3/3)

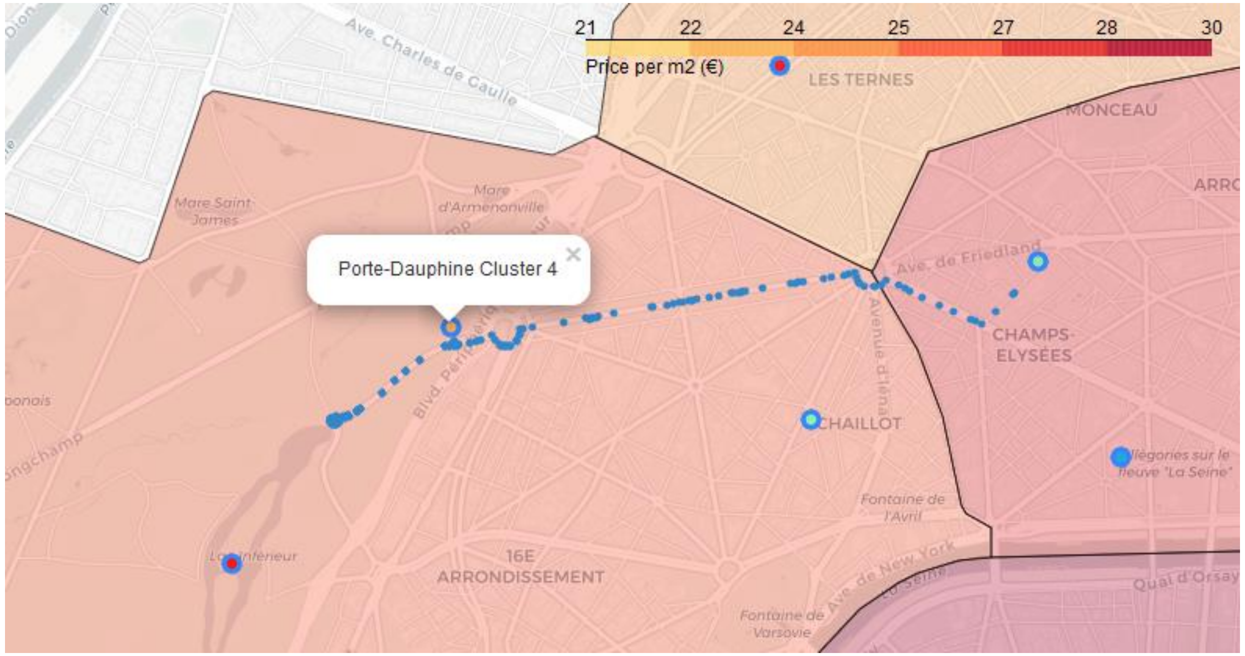




- **Cluster** and **workplace position** are chosen by user;

- **Iteration** through cluster in order to find location with **shortest traveling time** to reach workplace using TomTom API queries;

- Ideal candidate found → retrieve route → overlap on Paris Map.



6. Conclusion and future directions

- The research succeeded in providing a classification of the city of Paris;
- Most common websites for looking for apartments provides information about apartments only → This tool can be useful as support in a more in dept-research → User can have an initial understanding of the neighbourhood the apartment is located in;
- Highest effectiveness in large cities, with high density of venues where it is necessary to perform clustering to narrow down search scope → in case of smaller cities search can be modified in order to perform single category-based filtering;
- Several potential developments:
  - GUI to ease the use,
  - Machine learning algorithms to create a recommender system to find common searches and hints at most popular settings,
  - Use an API that allows to compute the traveling time not only by car, but also by other means, e.g. public transports,
  - Use of multiple sources to derive the venues and create a comparison (Foursquare may be used less in Europe than in the US, for instance).