

FINDING THE BEST LOCATION FOR AN APARTMENT IN PARIS

by Olivier Chauffert-Yvart - May 2020

As for many capitals of the world, finding an apartment in Paris is a genuine quest

- A lack of construction of new housing buildings and the ever growing attractiveness of the French capital make the **demand far higher than the offer**
- Consequently the prices of real estate were **multiplied by 4** between 2000 and 2020¹
- Looking for an apartment in Paris is a true hassle and the competition is fierce²
- Helping **individuals** find their new home and **real estate professionals** hunt the right place for their clients is a large business, addressed by many competitors
- Data science may help **saving time and energy** by targeting the right neighborhood before starting a research



Data acquisition, cleaning and preparation

In order to get information on Paris and its housing market, we used **4 data sources**:

- [Wikipedia](#) as a source for administrative details and demographic information
- [Data.gouv.fr](#), the open platform for French public data, driven by the Government, had a GeoJSON file of Paris
- [Foursquare API](#), to retrieve common venues for a given district of Paris, among a list of selected categories of venues
- [Le Bon Coin](#), ("The Good Corner"), a website similar to Craigslist or Gumtree, which lets individuals sell nearly anything, including houses and apartments

**20 districts
studied**

We aggregated the names, coordinates, borders, population and surface of each district in a single dataframe from Wikipedia and a French public source

**806 venues
retrieved**

Foursquare API was the perfect source for Parisian venues in the area of each district center, in order to create a profile of each district

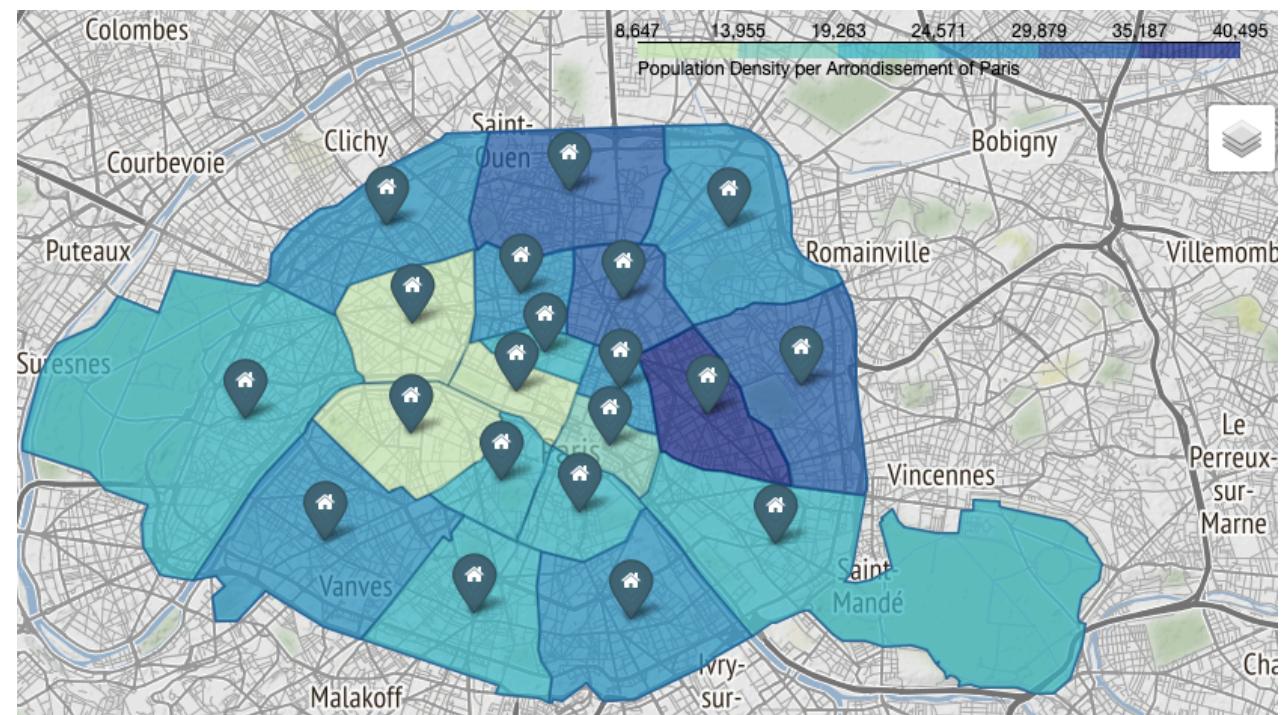
**136 offers
analysed**

Despite anti-bots security measures, we managed to retrieve enough apartments offers to conduct an analysis of the real estate market in Paris

Starting with a demographic analysis of the Parisian population distribution

The center of Paris is less populated as many of its famous monuments, parks, squares and large avenues are located there – leaving less space to housing buildings

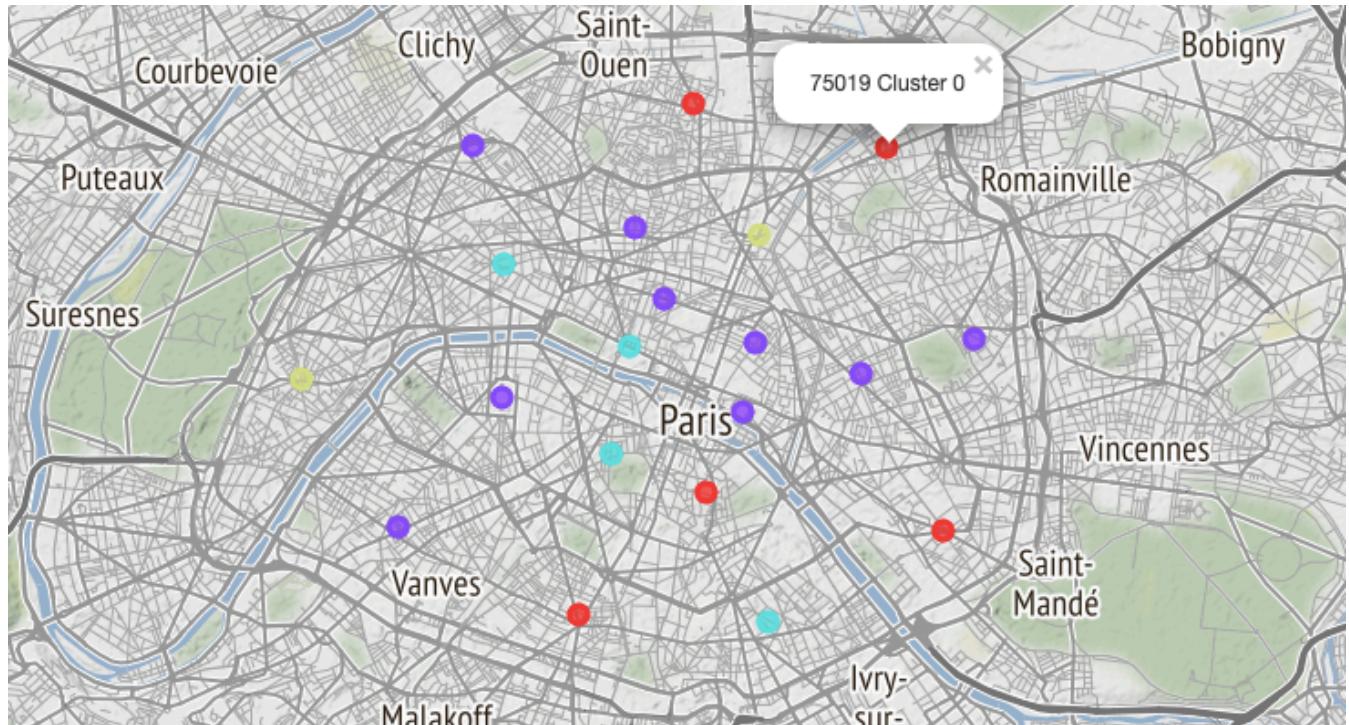
- The **1st, 7th and 8th arrondissements** (districts) are the least dense: this is where you will find the Musée du Louvre, the Jardin des Tuileries, Concorde square, the Palais de l'Elysée, the lawns and golden dome of the Invalides, and the Arc de Triomphe. Iconic landmarks mean higher prices, more parks mean less space for buildings
- The **north-east** districts seem globally denser (more blue than green) than the south-west half of the city
- The **11th arrondissement** is the densest of all (in dark blue) with a population density over 40.000 inhabitants per square kilometer



The battle of neighborhoods: Paris edition

After selecting specific kinds of venues, we used the **k-means algorithm** to define **4 clusters** of districts

- **Cluster 0** in red unites mostly bars and wine bars
 - **Cluster 1** in purple is the largest with 9 districts mixing bars and entertainment venues like concert halls and comedy clubs
 - **Cluster 2** in light blue is more about nightclubs, mainly located at the heart of Paris (1st, 6th, 8th and 13th arrondissements)
 - **Cluster 3** in green is the smallest, with 2 districts: bars and comedy clubs are most common venues there
- **Clusters 0 and 1 fit our interests best**



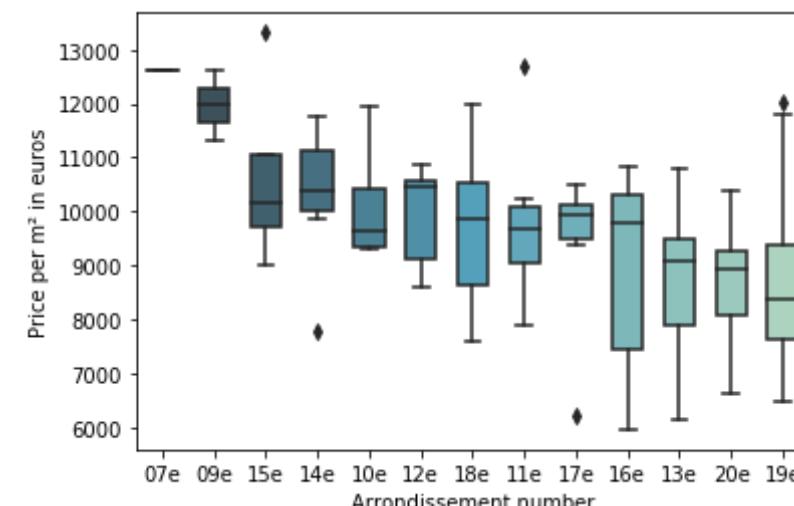
Scraping apartment offers was no small matter

Data is the new oil¹, and those who have data tend to protect it: we mimicked the behavior of a human Internet user to bypass anti-bots security measures and retrieve genuine apartment offers

- Among the 140 offers retrieved, we kept **136** after data cleaning, located in **13** out of the 20 arrondissements of Paris
 - The best KPI to compare districts in terms of housing offers was the price per square meter (€/m²): we aggregated the data in order to perform such a comparison
 - Box plots suit the visualization need, as they let display the statistical distribution of prices: mean, minimum, maximum and outliers
- **13th, 20th and 19th arrondissements are the cheapest in terms of mean price per square meter**

19th and 20th arrondissement belong to clusters 0 and 1 respectively. We will focus our apartment research on the **19th arrondissement** as it is the cheapest of both.

PostalCode	Price	Surface	Price_per_m ²	
12	75019	885,000 €	75 m ²	11,800 €
15	75019	865,000 €	72 m ²	12,013 €
41	75019	760,000 €	90 m ²	8,444 €
44	75019	875,000 €	104 m ²	8,413 €
59	75019	680,400 €	105 m ²	6,480 €
61	75019	735,000 €	88 m ²	8,352 €
79	75019	820,000 €	98 m ²	8,367 €
80	75019	880,000 €	88 m ²	10,000 €



Conclusion and future directions

This project was a good opportunity to reuse knowledge and competencies acquired during this Applied Data Science module, as well as developing a Proof of Concept on a use case for the real estate industry

- K-means clustering, data scraping and geodata visualization techniques were applied
- Multiple data source were identified and aggregated
- A future extension of this use case could multiply the number of sources and widen the scope of research, without limiting the research criteria: price, number of rooms, venues categories, city, etc.

