Finding the best location for an apartment in Paris

by Olivier Chauffert-Yvart - May 2020

1.   Introduction

1.1. Background

Paris is the capital of France and ranks second among the most visited cities in the world[1]. As the economic heart of the country, it is also the most populated French city with **2.19 millions** inhabitants as of 2017[2]. Its surface divided in **20 districts** - called *arrondissements* in French - is quite small compared to other capitals: **105.40 km²**. As a Parisian looking for a new place to live, I wondered if data science could help me find the best location for my next apartment, by leveraging open data and information available freely on the Internet.

1.2. Problem

Paris' attractiveness in terms of tourism and economic dynamism, lined with a low increase of new buildings, have caused a **sharp increase in prices** in the last 20 years and make it very difficult to find a place to live. Whether you want to rent a flat or buy an apartment, the competition with other candidates will be fierce.
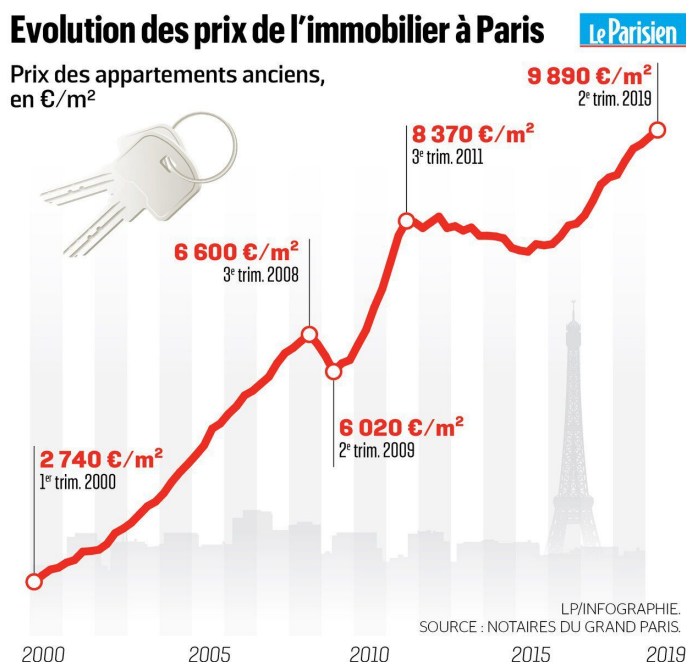
Real estate price evolution (€/m²) between 2000 and 2019 in Paris, France[3]

As you can see on this graph, the increase trend from year 2000 is quite steep. The mean price was 2740€ per square meter (m²) in Q1 2000, and it has now exceeded 10.000€/m² in 2020[3].

The Parisian market knew two recessions, consequences of the 2008 financial crisis, but they are already part of history. Maybe the Covid-19 pandemic will cause a new decrease of prices.

In this context, how could someone leverage data to choose the best location and find his or her next apartment?

[1] Business Insider: "*The 19 most visited cities around the world in 2019*"

[2] Wikipedia: dzd

[3] Le Parisien: "*10 000 euros le m2 à Paris : l'évolution du prix de l'immobilier dans la capitale en 5 graphiques*"

1.3. Interest

The project I chose to develop may be a source of interest for two categories of people:

a. **Individuals** planning to buy an apartment in Paris
b. **Real estate professionals** who have a mandate to find an apartment for their client, with specific research criteria about the location

2. Data acquisition, cleaning and preparation

2.1. Sources

As mentioned above, I tried to find free sources of information on the Internet. My research of data aimed to get a better understanding of **the city** of Paris, its **population,** its **districts**, the repartition of **venues I like to go to**, and its **housing market**.

A. **Wikipedia**[4] as a source for administrative details and demographic information;
B. **Data.gouv.fr**[5], the open platform for French public data, driven by the Government;
C. **Foursquare API**[6], to retrieve the most common venues for a given district of Paris, among a list of selected categories of venues I enjoy;
D. **Le Bon Coin**[7], ("The Good Corner"), a website similar to Craigslist or Gumtree, which lets individuals sell nearly anything, including houses and apartments.

2.2. Cleaning & Preparation

The cleaning and preparation operations were the most time-consuming of the project, as data sources were not always well formatted nor reconciled.

A. For Wikipedia, the districts were listed in a table with population and inhabitants density along history; I kept only the most recent population count, which was from 2017, and calculated the corresponding density (which was missing). I also added some information needed to reconcile these data with the other datasets of the project.
B. On the open data platform of the French government, I found a GeoJSON file describing the districts of Paris. Two districts had their centers in the middle of the woods, which was not appropriate for a housing research. I managed to modify these centers in more urban areas.
C. The Foursquare API was leveraged to retrieve data about the most common venues in each district, among a list of categories I selected according to my preferences: bars, nightclubs, comedy clubs, concert halls, music festivals and breweries.
D. Last but not least, I chose to scrape housing offers on Le Bon Coin website, which aims to match supply and demand, to get a real-life vision of the flats available for acquisition in Paris. I had to define research criteria, like my budget or the minimum number of rooms, identify the profile of a result page to extract a list of offers, and do the same process to extract useful information out of each offer page (140+). The most challenging part was to avoid being blocked by anti-bots cybersecurity measures in place: I made a code running slowly, to mimic the behaviour of a regular human Internet user.

---

[4] Wikipedia: List of Paris districts
[5] Data.gouv.fr: Arrondissements
[6] Foursquare API: Developers portal
[7] Le Bon Coin: French classified ads website

2.3. Aggregation

In order to use all these data simultaneously in *pandas* dataframes, I decided to use the **postal codes** and **district numbers** of each *arrondissement* as matching keys. It was essential to reconcile demographic, geographic, social and commercial information I had to get a holistic view of each district and choose wisely were I would establish my next residence in the future.

The complete report will include the following sections:

3. Methodology section

It represents the main component of the report where will be discussed and described any exploratory data analysis that I did, any inferential statistical testing that I performed, if any, and what machine learnings were used and why.

4. Results section

Where the results will be discussed.

5. Discussion section

6. Observations

Any observations I noted and any recommendations I can make based on the results will take place in this section

7. Conclusion section