CSC5003

Where To Invest in the Paris region

MARIE Nordine - TELECOM SudParis

December 27, 2021

Abstract

By defining a Building Score and a GrandParisExpress Score we were able to classify the cities of the Paris region in order to determine the cities in which it seems interesting to invest. The results obtained are in line with the investment advice available on the Internet. Finally, a web application using these results has been designed to facilitate decision-making according to the profile of each user.

Table of Contents

1 Introduction

As a student soon to graduate, I was wondering where it would be most interesting to invest in the Paris region. This is the reason why I tried to classify the cities of this region by taking into account two parameters : a **Building score** predefined and the proximity of the city with **the future stations** of the *Grand Paris Express*.

2 Datasets

- **Dataset of the cities of Paris region** - data.gouv.fr (Open License)
  The dataset contains multiple informations on each city of the Paris region such as its postal code, INSEE code, population and GPS position.

- **Dataset of building permits** - data.gouv.fr (Open License)
  The dataset contains **the list of the french building permits since 2017** as well as multiple information concerning the construction associated to the building permit.

- **Location of the Grand Paris Express stations** - data.gouv.fr (Open License)
  The dataset contains the location of the future stations of the *Grand Paris Express*. The data uses the CC49 French coordinate system, so we will have to reproject them in GPS coordinates to fit the rest of the project.

- **GPS Shapefiles of the departments of Île-de-France** - data.gouv.fr (Open License)
  Shapefiles of the departments of Île-de-France that will be used to plot and represent graphically Paris region map.

3 Data pre-processing

3.1 Spark pre-processing :

First, **in Spark** with the building permit dataset, we **filter only the IDF building permit** thanks to the region code column. (NB : `IDF_regioncode = 11`) Then we **select only 3 columns** of the dataset : the INSEE city code, the actual start date of the building permit and the total number of home units created. (For example a building permit of a T3 flat creates 2 home units since there is 2 bedrooms in it.)

Then **we reduce by key** this result, where the key is the INSEE city code, the commutative operator the addition `(_+_)` and the reduced value the total number of home units created.

To have have more information on the city such as its population or its name, we have joined the precedent reduced data with the dataset of the cities of Paris region. Cities with a population less than 10.000 are excluded to maintain some reliability as sparsely populated cities have little year-round information available. Finally, we created a new column called `Building Score` which is defined as $BSc_{city} = \frac{HomeUnitsCreated_{city}}{Population_{city}}$ .After that **we export the result as a CSV file.**

3.2 Pre-processed data format :

We end up with a **pre-processed CSV file** with the following header :

INSEE code (String) — Postal code (String) — City Name (String) — Department (Integer) — Population (Integer) — Nb of Home Units created (Integer) — Building Score (Float) — GeoPoint ((Float,Float))

4 Clusterizations

4.1 Building score clusterization :

First, we classified among the Building Scores with a **K-Means algorithm and K=4** since **it minimized the silhouette index.** We can identify 4 building strategy among the Paris region cities :

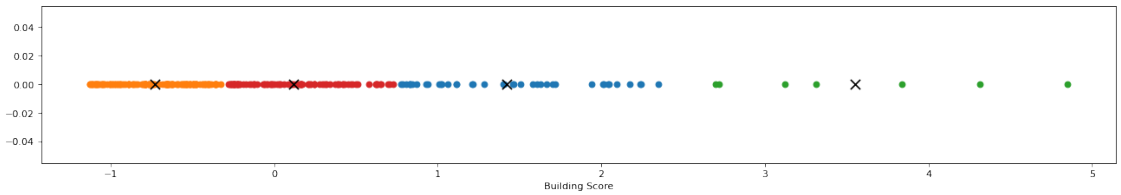

Figure 1: Building score K-Means clusterization

- Cities which don't aim to construct anymore.
  (Low Building score)
- Cities that build standardly compared to their population.
  (Medium Building score)
- Cities that build a lot compared to their actual population.
  (High Building score)
- And the last one are cities that drastically build compared to their population
  (Very High Building score)

On the next figure we can see the distribution of those geographically classified cities :

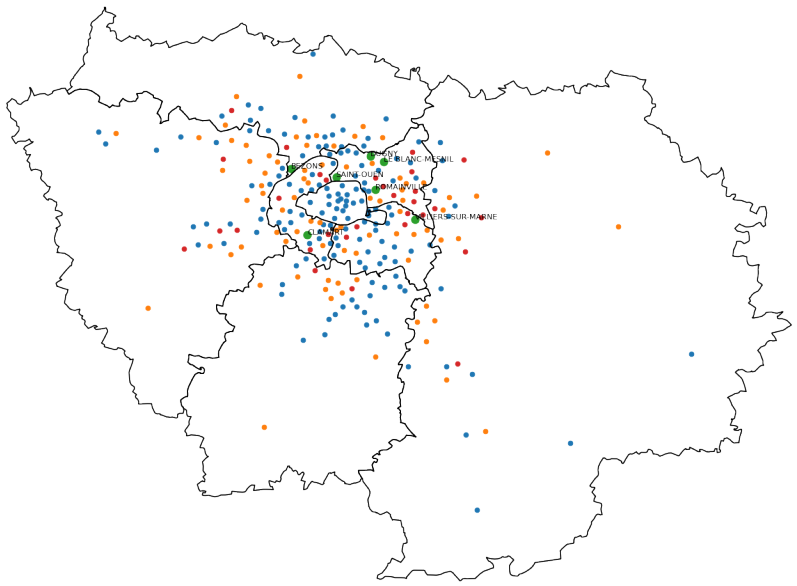

Figure 2: Paris region cities classified among their Building Scores

4.2 Grand Paris Express scores :

The *Grand Paris Express* (GPE) is a public transport network project consisting of four automatic metro lines around Paris, and the extension of two existing lines. **Disrupting at the same time the real estate in the Paris region.** It seems therefore very interesting to **study this parameter.**

On the next figure you can see **the position of the future stations of the GPE**.

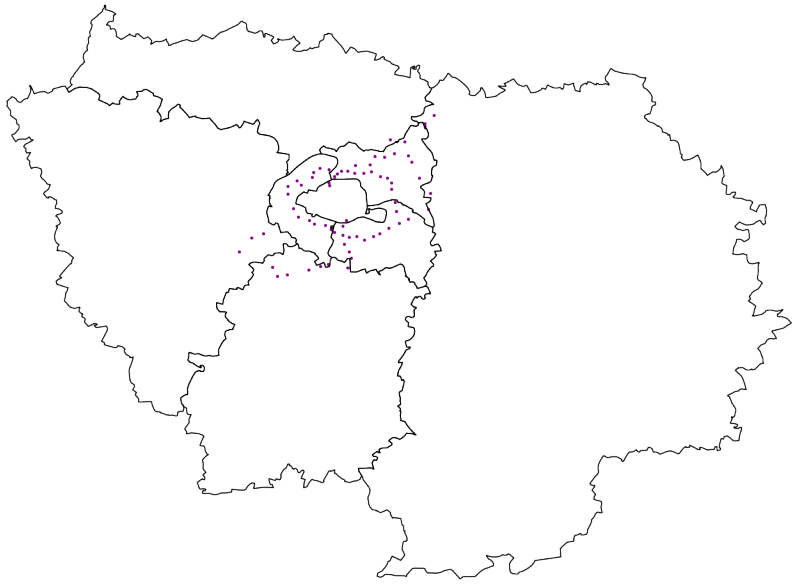

Figure 3: Grand Paris Express Stations

As you can see, these stations are **not intended to serve Paris but its suburbs.** Some cities around Paris will thus be able to benefit from a privileged access to the metro.

We will **try to quantify this proximity** with these future stations in the form of a GPE Score thanks to the **haversine distance** noted d_{hv} (which measures the distance as the crow flies between two GPS points)

Empirically we defined the GPE Score as :

$$GPEScore_{city} = \sum_{station} exp\left(-\frac{d_{hv}(city, station)}{2}\right)$$

Since it provides a natural and coherent distribution of the GPE Score around the stations according to the following colormap :

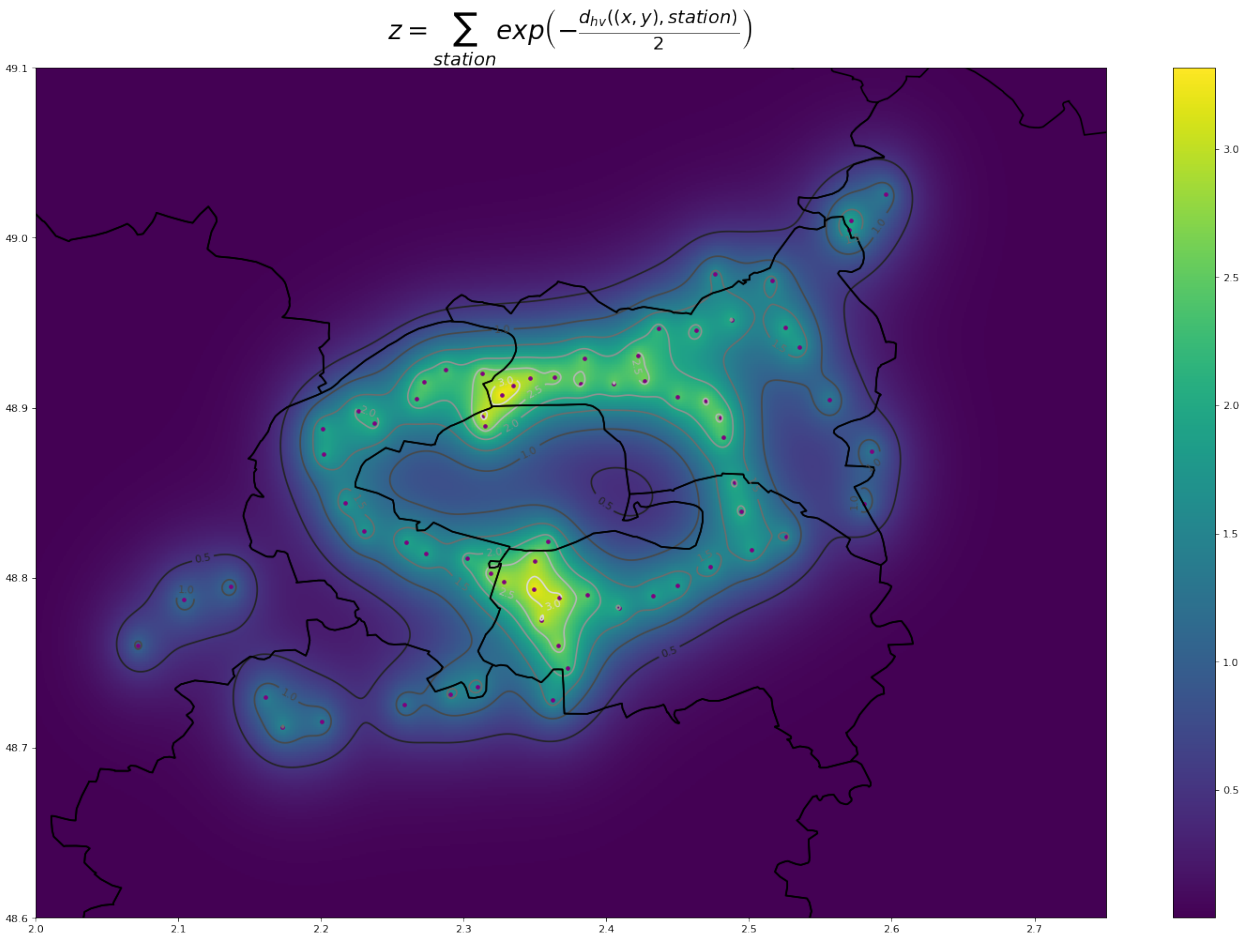

Figure 4: GPE Score Colormap

4.3 Building and GPE Score clusterization :

We then made a **2-dimensional 4-clusterization** based on the Building score and the GPE score below is the city 2D-cluster plot.

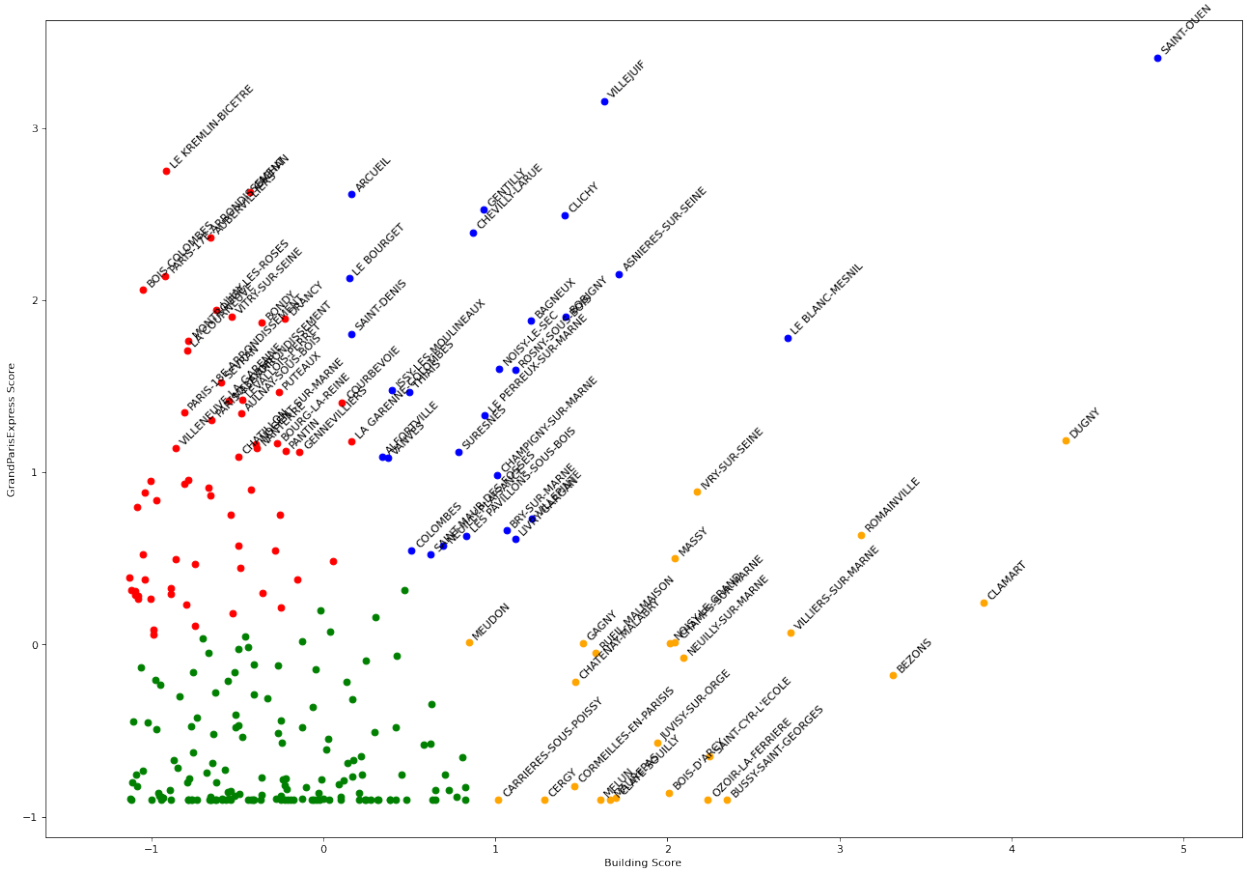

Figure 5: Building score and GPE score Clusterization

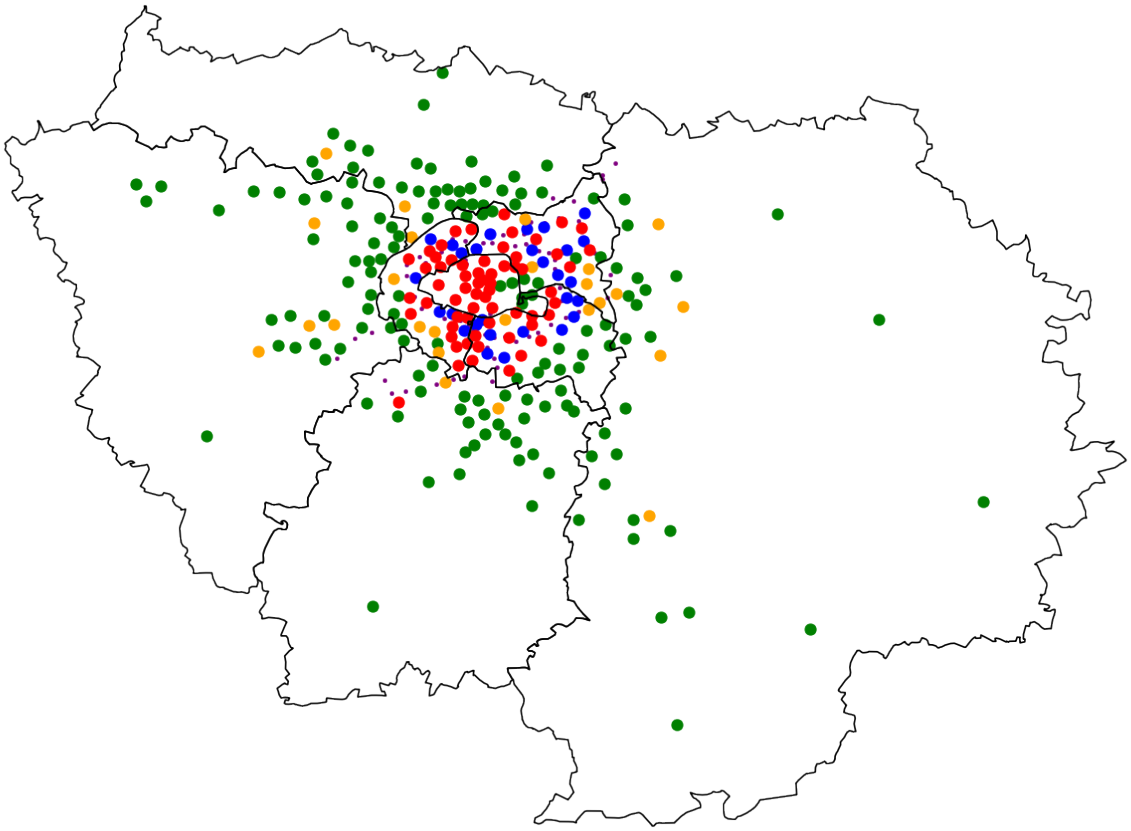

Figure 6: Paris region cities 2D-classified

Geographically, we can describe the **4 classes** :

- The blue class represents the cities that have **a great GPE Score and a medium to high Building score.** They are on the Parisian crown near to multiple Grand Paris Express future stations.
- The red class represents cities that have **a great GPE Score but which don't have a high Building score.** They are the boroughs of Paris or the cities on the Parisian crown.
- The orange class represents cities that have **a medium to high Building score and a low to medium GPE Score.** They are distributed in a disparate manner around Paris
- The green class represents cities with **a low Building and GPE Score.** Most of the cities around Paris are in this class. They represent **the least interesting cities to invest in.**

5 Results

5.1 Analysis

So naively according to this classification the **cities with the highest potential are those in the blue class.** However, it should be taken into account that **if the Grand Paris Express Score is static** (Since there would be no new stations planned), **the Building score can evolve over time**. Therefore, **red class cities should not be overlooked** for our real estate investment projects, since they may translate as blue class cities due to the presence of the future GPE stations.

Besides, if we look at the **articles of real estate investment advice** such as [**1**] [**2**] or [**3**], **the recommended cities are mostly blue class cities** (or red class if not) which seems to validate our model.

5.2 Dash web application

Finally, **to facilitate the investment decision**, a web dash application has been designed to **visualize the data more easily and to match the evolution of the price per square meter to each city** by hovering it. Thus, users will be able to search for a city that is not only interesting in terms of investment but also fits their budget.

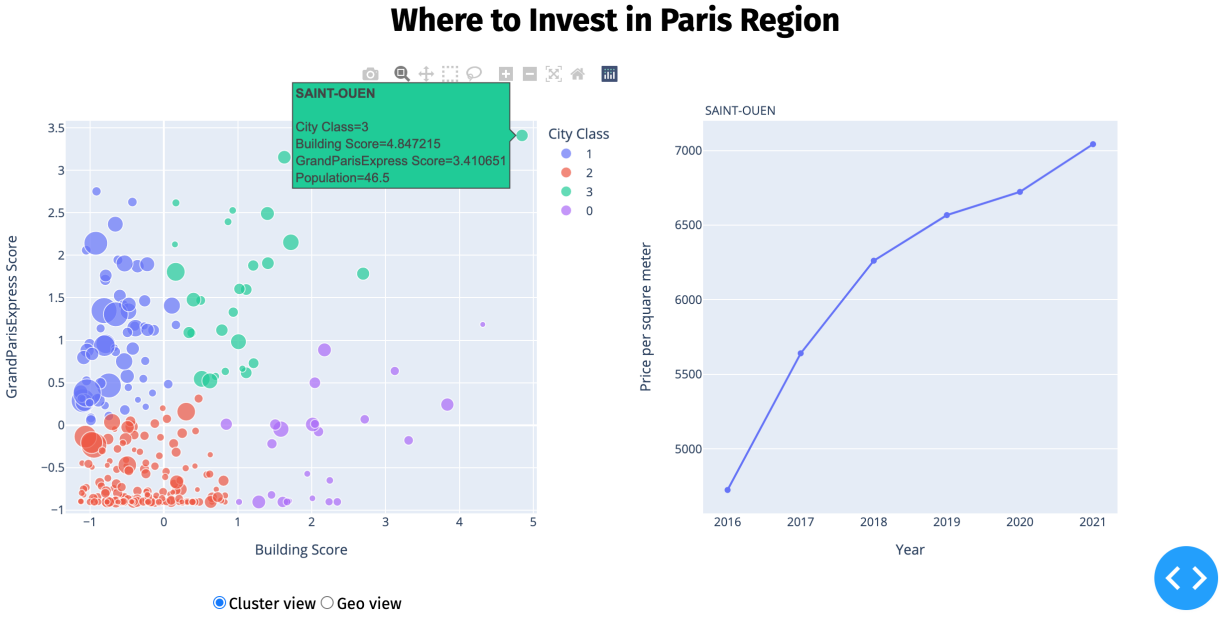

Figure 7: Where to Invest in Paris region Dash web application

NB : The dash app can be accessed by **clicking here**, or by manually deploying it from the source code (cf. `README.md` at the root directory of the source code)

6 Conclusion

To conclude, our model, along with its web application, **provides a tool to facilitate real estate investment choices.** We have chosen to **focus on 2 parameters** that we have defined before: the *Building Score* and the *GrandParisExpress Score*. These two parameters have allowed us to have satisfactory results because they have been **validated by many real estate investment articles**.

However, to improve our model it would seem necessary to **take into account the renovations in the Building score**, because many cities in the near Paris and the districts of Paris do not necessarily have space to build and renovate more than they build, which makes them equally interesting cities to invest. Thus some cities categorized as red (i.e. high GPE score and low Building score) should have been classified as blue (i.e. high GPE score and high Building score).

References

[1] **GRAND PARIS : Le top 10 des villes où investir**- Cheval Blanc Patrimoine, chevalblanc-patrimoine.fr, 2021
https://www.chevalblanc-patrimoine.fr/guides/grand-paris-le-top-10-des-villes-ou-investir/

[2] **Grand Paris : 12 villes où investir**- Mickael ZONTA, investissement-locatif.com, 2020
https://www.investissement-locatif.com/ile-de-france/12-villes-ou-investir-grand-paris.html

[3] **Grand Paris : Top 10 des villes et quartiers où investir** - Eric CHATRY, jerevedunemaison.com, 2018
https://www.jerevedunemaison.com/blog-immobilier/grand-paris-top-10-villes-quartiers-investir