Real estate prices in Paris:
A multivariate study of the location criterion.

Stephanie Blanchet, Researcher
March 15, 2019

1. Introduction

The fourth most expensive metropolis in the world [1], Paris has seen its real estate prices quadrupled in 20 years. According to the Paris Notary Chamber [2], a new record should be reached in the spring of 2019 with an average of 9730 euros/m2, an increase of around 7% in one year. If the permanent soaring benefits all boroughs, all are not equal. Average prices currently range from 7930 euros/m2 in the "19th arrondissement" to 13150 euros/m2 in the "6th". At the neighborhoods scale, the gaps are even greater, with an average range of 7460 euros/m2 in "La Chapelle" (18th) to 17410 euros/m2 in "Odéon" (6th), nearly 15 times the net minimum income.

Real estate specialists generally consider that the location of a property is the primary criterion in estimating its market value. Benefits such as the number of services and shops, local amenities, public transport or schools are regularly cited to explain that one area is valued more than another. But what about the City of Lights? Does that kind of rationality also drive the invisible hand of the Parisian market?

This is the question this study proposes to answer by measuring the link between several variables related to location and the average price of real estate by borough. It should allow to establish a grid of parameters for the use of professionals such as real estate agencies and notaries, but also of any potential buyer wishing to understand what weighs in the price of a good according to its situation in the French capital, especially when this price appears to be inappropriate with the intrinsic qualities of the property in question (e.g. high price for a narrow area, obsolete facilities, no elevator, ...).

> *The problem is:*
>
> What can explain the differences in the price of real estate from one Parisian borough to another, all other things being equal?

2. Methodology

2.1 Data selection

The data were selected on criteria of reliability, transparency on their origin, date of last update, as well as on the presence of a division by borough of Paris, in order to gather various factors likely to explain the variability of the prices depending on the location.

Thus, official sources and open institutional or government databases have been favored:
- the French national institute for statistics and economic studies (INSEE),
- Paris Notary Chamber,
- Paris City Hall,
- the Regional Council of Ile-de-France (Greater Paris)
- the French Ministry of Education

FourSquare API provided user generated content (UGC) about popular venues in each borough.

Data content and use	Source	Information link	Update
1. Average real estate prices by borough - the target variable	Paris Notary Chamber	http://notairesdugrandparis.fr/fr/carte-des-prix#paris	January, 2019
2. Geodata - get the latitude and longitude of each Parisian borough so to send queries to the FourSquare API - also get areas to calculate the density per square kilometer of each independent variable	Paris City Hall	https://opendata.paris.fr/explore/dataset/arrondissements/information/	-
3. Population - calculate the population density per km2 of each borough	INSEE	https://www.insee.fr/fr/statistiques/fichier/3677781/dep75.pdf	January, 2019
4. Service offering - list and count the different categories of services available in each borough	Regional Council of Ile-de-France	https://www.data.gouv.fr/fr/datasets/les-services-aux-particuliers-par-commune-ou-arrondissement-base-permanente-des-equipements-idf/	March, 2016
5. Shop offering - list and count the different categories of stores and shops available in each borough	Regional Council of Ile-de-France	https://www.data.gouv.fr/fr/datasets/les-commerces-par-commune-ou-arrondissement-base-permanente-des-equipements-idf/	March, 2016
6. Popular venues - list and count the most common categories of venues in each borough	FourSquare API	https://fr.foursquare.com/	March, 2019
7. Municipal facilities - list and count the different categories of collective facilities in each borough	Paris City Hall	https://opendata.paris.fr/explore/dataset/equipements_de_proximite/information/	May, 2013
+ Rail transport (refers to 'Municipal facilities') _ list and count the rail transports (metro, tramway, train) available in each borough	Regional Council of Ile-de-France	https://data.iledefrance.fr/explore/dataset/gares-et-stations-du-reseau-ferre-dile-de-france-par-ligne/information/	November, 2018
8. Education offering (a) **Primary & secondary schools** - list and count public and private primary and secondary schools available in each borough	French Ministry of Education	https://data.education.gouv.fr/explore/dataset/fr-en-adresse-et-geolocalisation-etablissements-premier-et-second-degre/information/	February, 2019

(b) **Public higher education establishments** - list and count public higher education establishments available in each borough	Regional Council of Ile-de-France	https://data.iledefrance.fr/explore/dataset/implantations-des-etablissements-denseignement-superieur-publics/information/	October, 2017
(c) **Private higher education establishments** - list and count private higher education establishments available in each borough	Regional Council of Ile-de-France	https://data.iledefrance.fr/explore/dataset/idf_universites/information/	May, 2013
9. Social housing - list and count the different categories of social housing, built or acquired, in each borough	Paris City Hall	https://opendata.paris.fr/explore/dataset/logements-sociaux-finances-a-paris/information/	February, 2019

Table 1. Data sources and use

2.2 Variables and indicators

In this study, **6 groups of variables** were analyzed as possible predictors of real estate prices in Paris.

After data preprocessing and transforming stages, the **total number of features was reduced from 323 to 49** so to avoid too much granularity and to make the results readable[3]. They were then normalized on the basis of their density per square kilometer to allow the comparison between boroughs.

2.2.1 Private / public service offering

This dataset was reduced from 31 to 14 features listed below:

- Law enforcement offices
- Courts
- Tax offices
- Employment agencies
- Banks
- Automotive services
- Building crafts
- Personal care services
- Estate agencies
- Restaurants
- Post offices
- Veterinaries
- Laundry / Dry cleaning services
- Undertakers

2.2.2 Shop offering

This dataset was reduced from 23 to 8 features listed below:

- Supermarkets
- Convenience stores
- Food shops
- Personal goods shops
- Household goods shops
- Books and stationery shops
- Florists
- Gas stations

2.2.3 Popular Venues

Due to many duplicates and redundant categories - which is inherent to user generated content, this dataset was reduced from 195 to 10 features listed below:

- Popular hotels
- Popular bars, pubs & cafés
- Popular beauty care places
- Popular culture & entertainment places
- Popular easy food places
- Popular nightclubs
- Popular restaurants
- Popular shops & markets
- Popular touristic sites
- Popular sport activity places

2.2.4 Municipal facilities

This dataset was reduced from 41 to 4 attributes listed below:

- Collective daycares
- Community centers
- Local sport facilities
- Squares & gardens
- Rail stations (metro, tramway, city train)

2.2.5 Education offering

The primary and secondary education dataset was reduced from 18 to 2 categories that are "primary" and "secondary" and divided by status "private" vs "public"; public and private higher education datasets were treated as the two modalities of the category "higher education":

- Primary schools
- Secondary schools
- Higher education
- Public vs private

2.2.6 Social housing

This dataset was reduced from 15 to 7 attributes listed below, "via construction" and "via acquisition" being the two methods of social housing implementation:

- for dependent persons
- for families
- for young adults
- for poor people
- for migrants
- via construction vs via acquisition

2.3 Exploration and data analysis

The target variable i.e. real estate prices in Paris, is a continuous value which we seek to determine what makes it increase or decrease. Thus, for each group of explanatory variables, a correlation analysis was performed through both Pearson's r and Spearman's rho calculation [4].

2.3.1 Paris real estate prices by borough

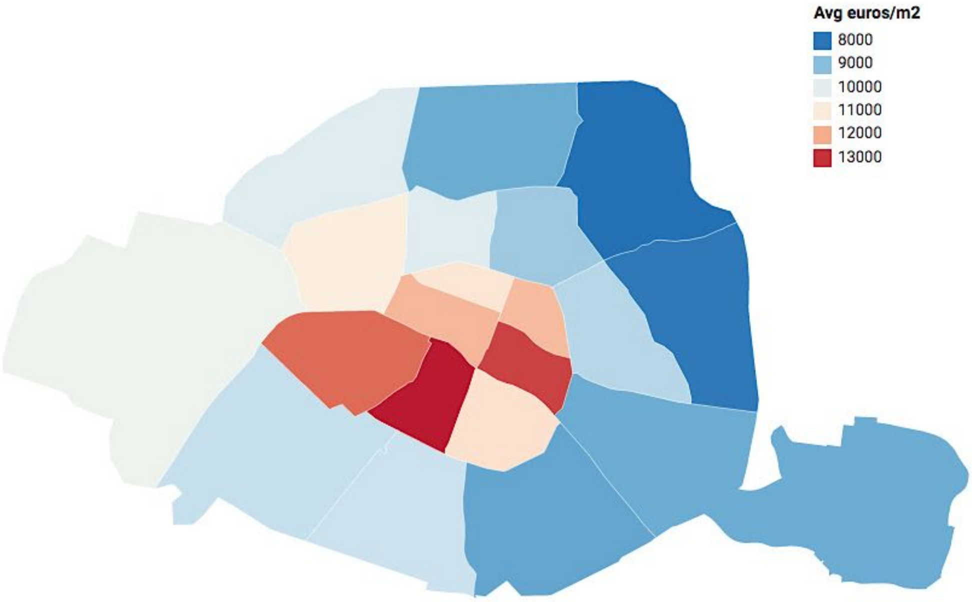


Figure 1. Paris real estate prices map

The most expensive boroughs (> 11000 euros/m2) are concentrated in the heart of Paris. The top 3 whose average price exceeds 12500 euros per square meter is composed of the 6th, 4th and 7th boroughs. Then come the 1st, 3rd, 5th and 2nd occupying the range from 11000 to 12000 euros/m2 (see figure 1).

In the west, the 8th and 16th hold the prices above 10000 euros/m2.

The range from 9000 to 10000 euros/m2 is distributed in the north-west with the 17th and 9th boroughs, in the South with the 14th and 15th boroughs, and around the hyper-center with the 11th and 10th boroughs.

The least expensive boroughs (< 9000 euros/m2) are located in the eccentric East, where we find the 19th, 20th, 13th, 18th and 12th.

2.3.2 Population density by borough

The northern half of Paris is the most populated (> 30000 inhabitants/km2), while the central zone sees the lowest concentration (< 10000 inhabitants/km2). With more than 40000 inhabitants/km2, the 10th borough holds the highest density population (see figure 2).

This seems to stick to the distribution of real estate prices, with the less expensive in the north and the most expensive in the center. A moderate negative correlation is observed (r = -0.47, p = 0.03 ; rho = -0.5, p = 0.02).

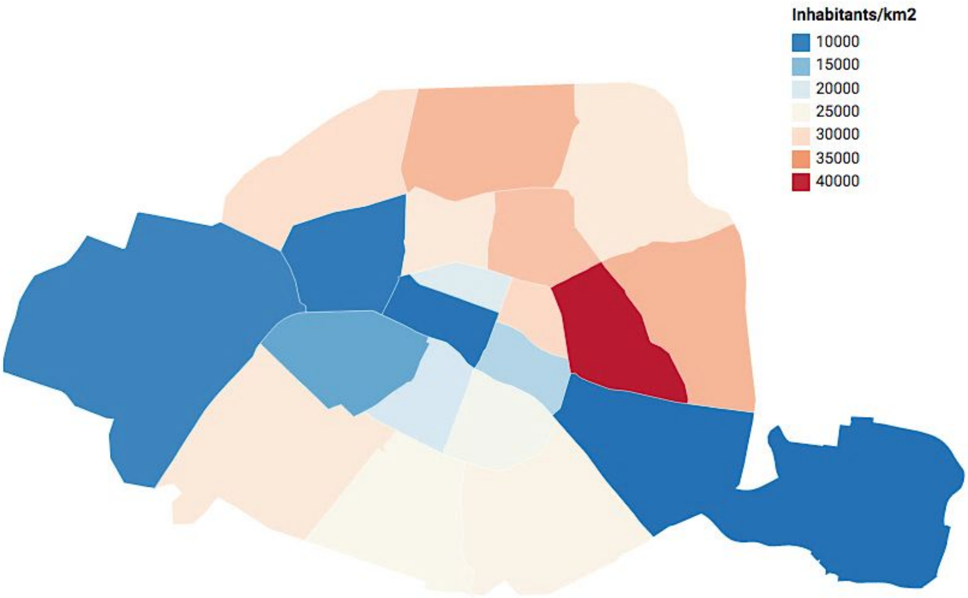


Figure 2. Paris population density map

2.3.3 Service offering analysis

On the northern outskirts of the Parisian center, the 2nd, 10th, 9th, 11th and 8th boroughs show the highest service density (> 1000/km2). The least well-endowed occupy the southern half. At the east and west ends, the 12th and 16th hold the lowest concentrations (see figure 3).

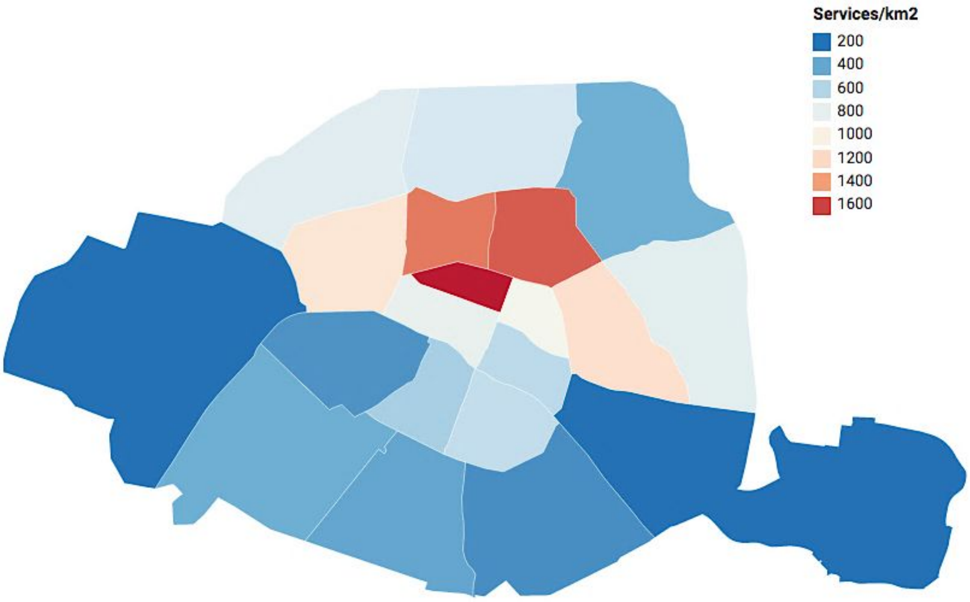


Figure 3. Paris services density map

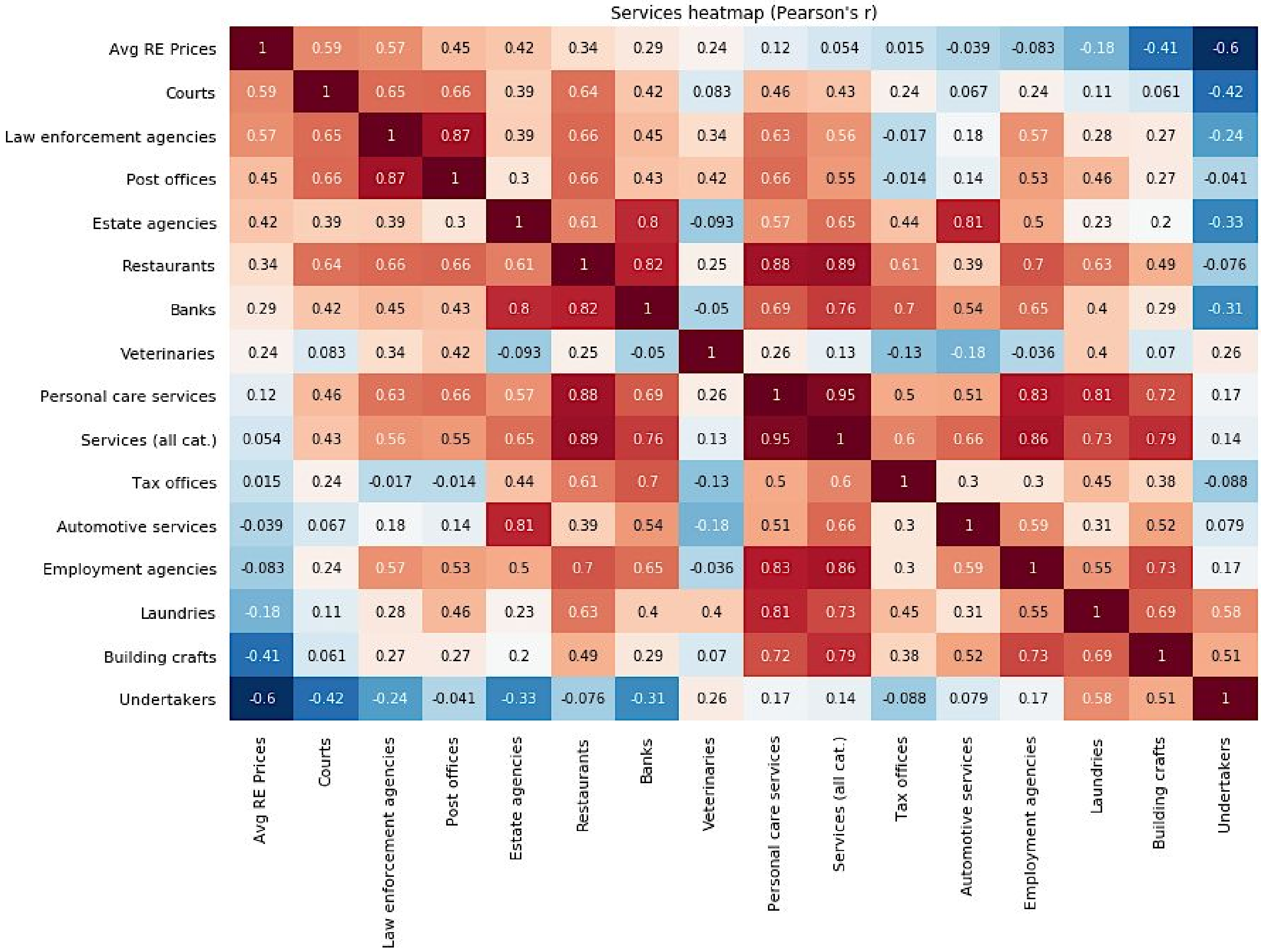


Figure 4. Service offering heatmap

No significant correlation was observed between the overall services offering per square kilometer and the average property prices in Paris (r = 0.05, p = 0.8 ; rho = 0.09, p = 0.7 | see figure 4).

However, some indicators show moderate correlations:

- Courts (r = 0.58, p = 0.006), law enforcement offices (r = 0.57, p = 0.008), post offices and estate agencies tend to make the prices higher

- while death care services and building crafts tend to make them lower.

2.3.4 Shop offering analysis

In the center and its proximity, the 2nd, 1st, 3rd, 6th and 9th boroughs that are the most expensive share the highest store concentration (> 548/km2).

Similarly to the distribution of services, the southern half sees the lowest shop density. The 16th and 12th again are the least well-endowed (see figure 5).

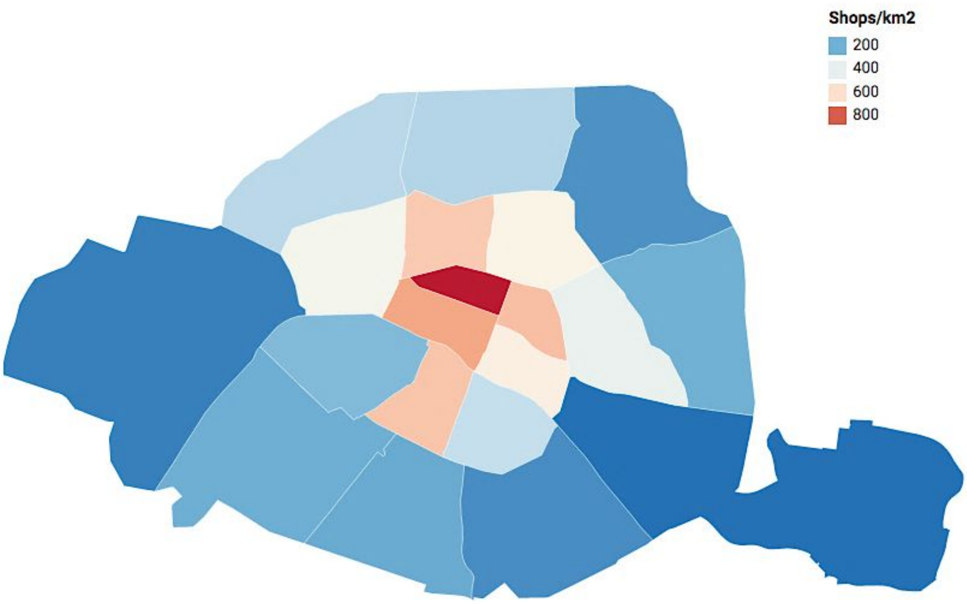


Figure 5. Paris shops density map

On the overall, the "shop offering" variable shows a moderate correlation with the average real estate prices in Paris ($r = 0.58$, $p = 0.006$ | see figure 6).

Within this category, the density of personal goods stores per square kilometer proves to have the most positive weight ($r = 0.6$, $p = 0.004$). Follow household goods stores ($r = 0.59$, $p = 0.005$), florists ($r = 0.59$, $p = 0.005$) and books and stationery shops ($r = 0.57$, $p = 0.01$).

Interestingly, the density of gas stations ($r = -0.52$, $p = 0.01$) and supermarkets ($r = -0.43$, $p = 0.01$) per square kilometer tends to make the prices lower. One reason may be that both cause undesirable car traffic in the borough which constitutes a nuisance for the quality of life.

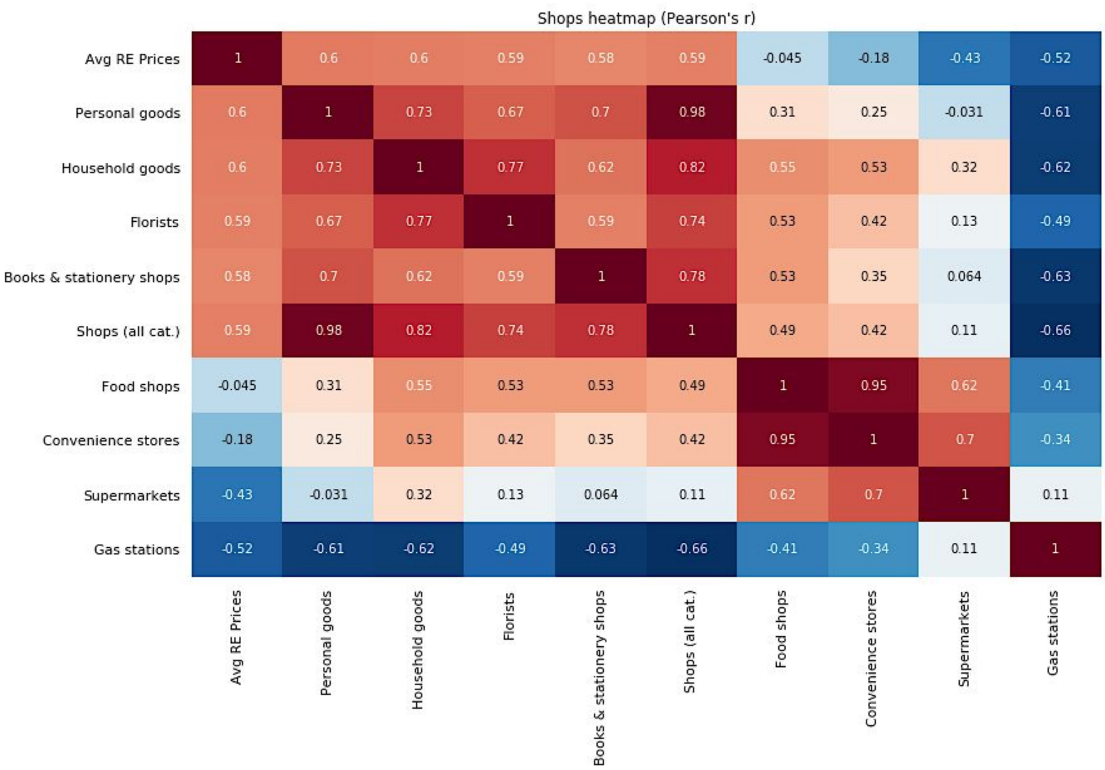


Figure 6. Shop offering heatmap

2.3.5 Popular venues analysis

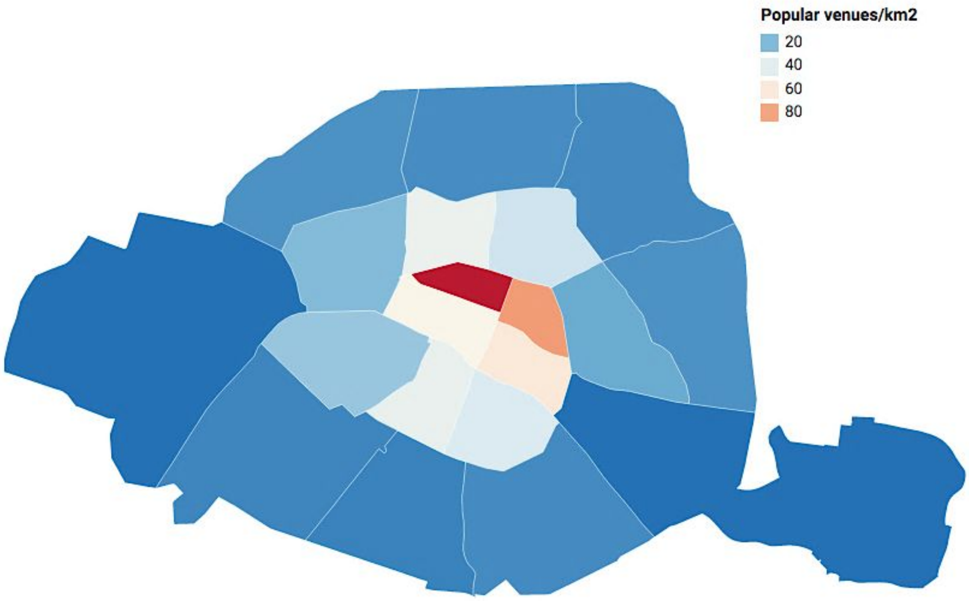


Figure 7. Paris popular venues density map

The heart of Paris and its proximity concentrate the most popular venues, the 2nd, 3rd and 4th being the top 3 boroughs (> 60/km2). The density falls below 10/km2 at the eccentric borders (see figure 7). A pattern that follows prices distribution.

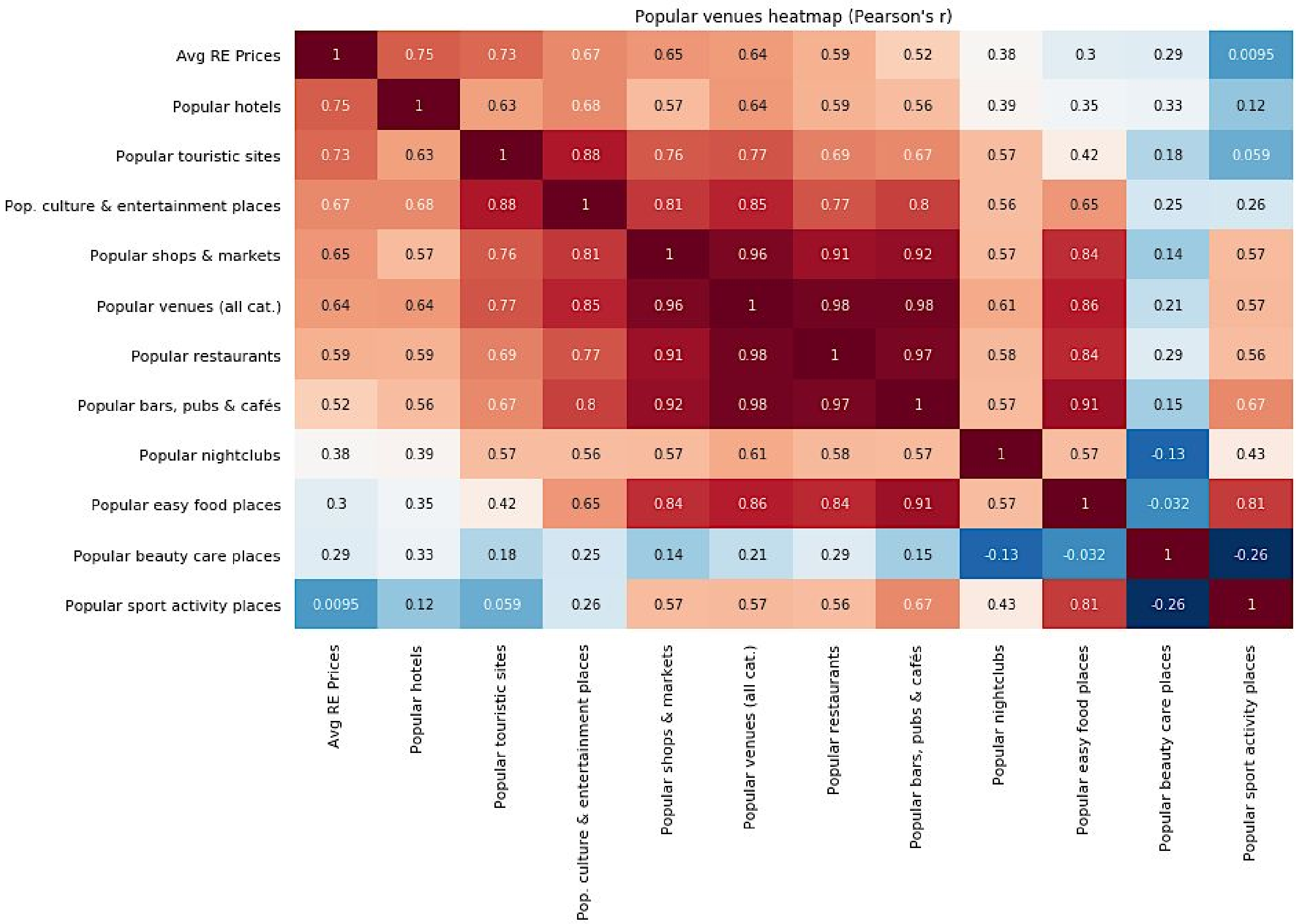


Figure 8. Popular venues heatmap

The "popular venues" category brings together the factors with the most positive weight in the average property prices in Paris (see figure 8).

The densities of popular hotels (r = 0.75, p = 0.0001) and touristic sites (r = 0.73, p = 0.0003) per square kilometer are indeed the two most positive regressors among the 49 indicators analyzed in this study. Follow popular culture and entertainment places (r = 0.67, p = 0.001), shops and markets (r = 0.65, p = 0.002), restaurants (r = 0.59, p = 0.005) and bars, pubs and cafés (r = 0.52, p = 0.01) also make the prices higher.

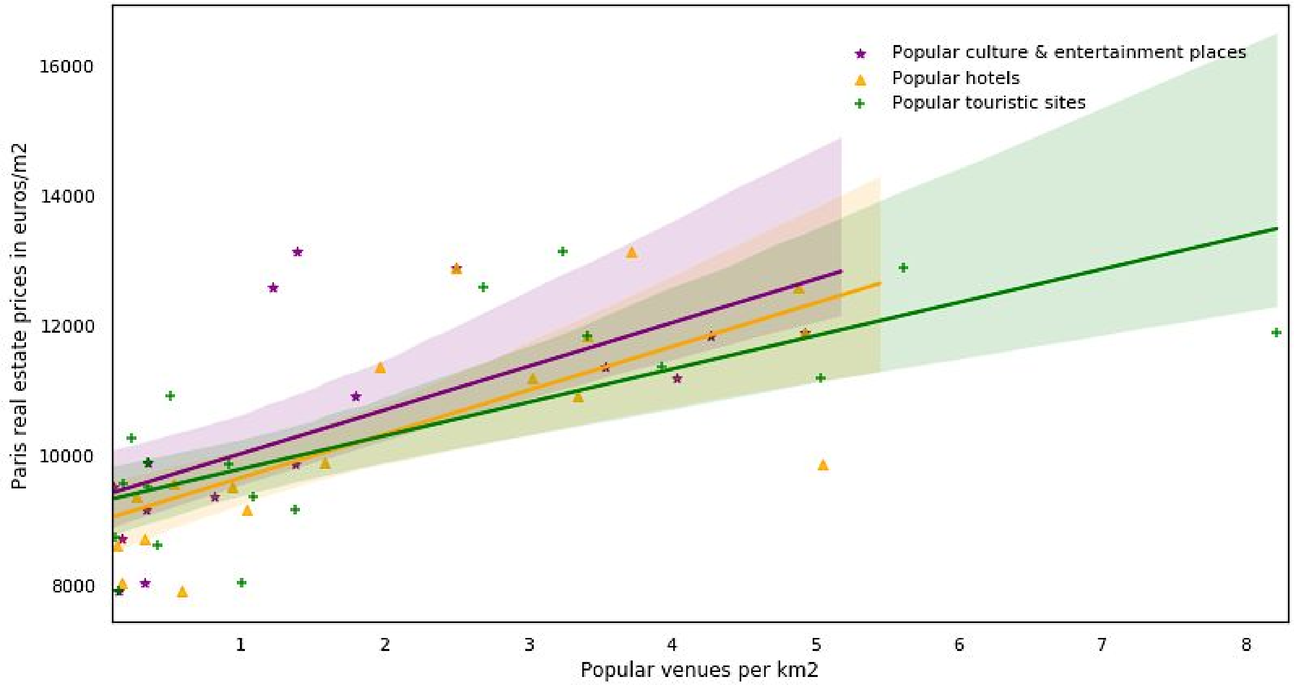


Figure 9. Popular venues regression plot

2.3.6 Municipal facilities analysis

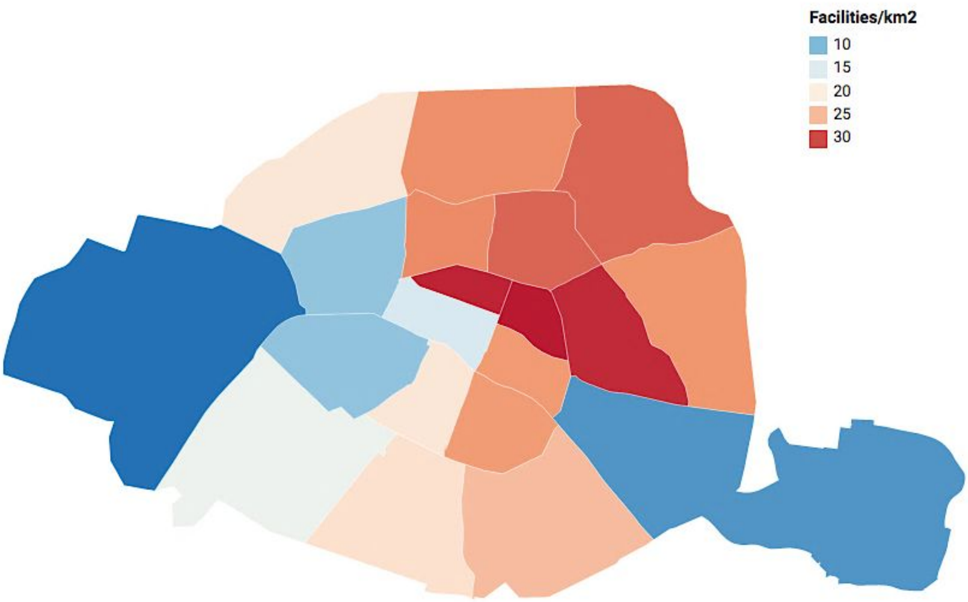


Figure 10. Paris municipal facilities density map

The municipal facilities are relatively well distributed throughout the Parisian territory. But again, the 16th and 12th boroughs are distinguished by their very lower density (see figure 10).

On the overall this category proves to have no impact on the average property prices in Paris. Interestingly, if it had one, it would be rather negative (r = -0.15, p = 0.5 | see figure 11).

Indeed, one indicator emerges: the density of local sport facilities (e.g. physical education grounds, gymnasiums) per square kilometer which tends to make the prices lower (r = -0.51, p = 0.02).

It can be noted that municipal facilities assemble attributes that benefit to family life (e.g. daycare, playground, etc.) and/or day-to-day (e.g. transport). As such one can expect them to be valued. The quite cold (but not significant) aspect of this group of variables raises hypotheses about the Parisian sociology. Further research could indeed focus on the profile of borough residents by examining variables such as family and employment status.

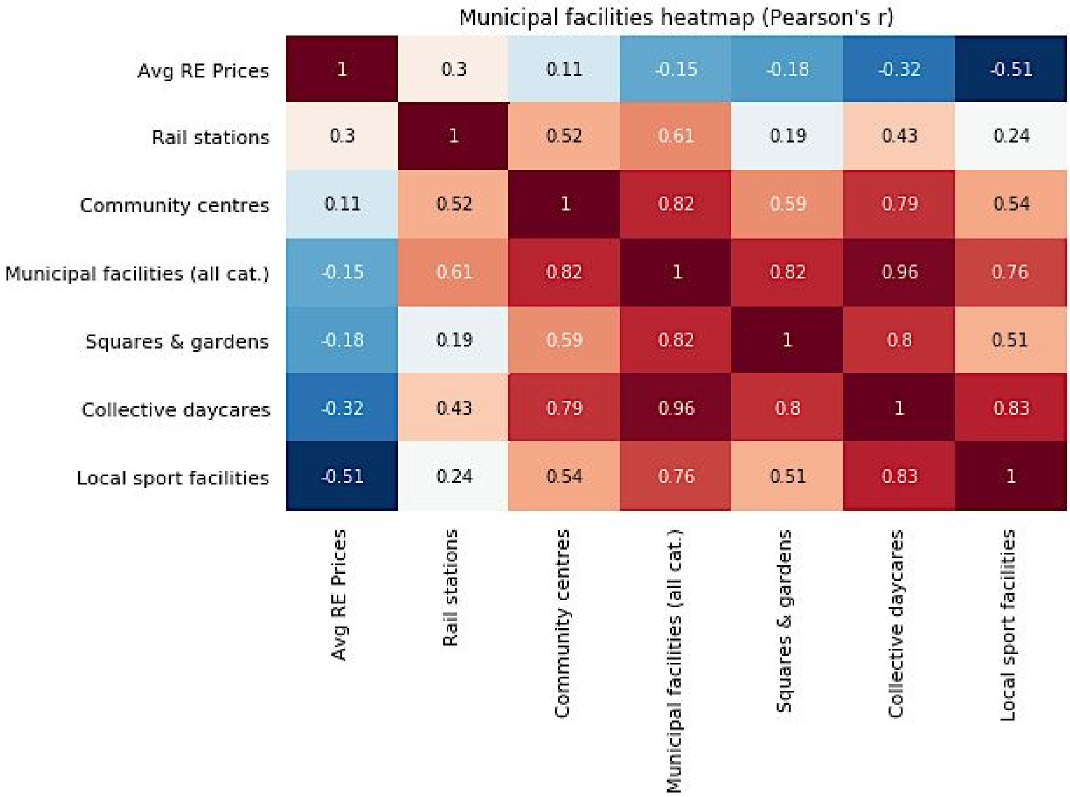


Figure 11. Municipal facilities heatmap

2.3.7 Education offering analysis

On the overall, the "education offering" category does not show any correlation with the target either (r = 0.29, p = 0.2 | see figure 13).

Only does the density of private education establishments per square kilometer tend to make the average property prices in Paris higher (r = 0.51, p = 0.02). This is mainly due to the high concentration found in the 6th borough (> 18/km2) which is also the most expensive one (13150 euros/m2 | see figure 12).

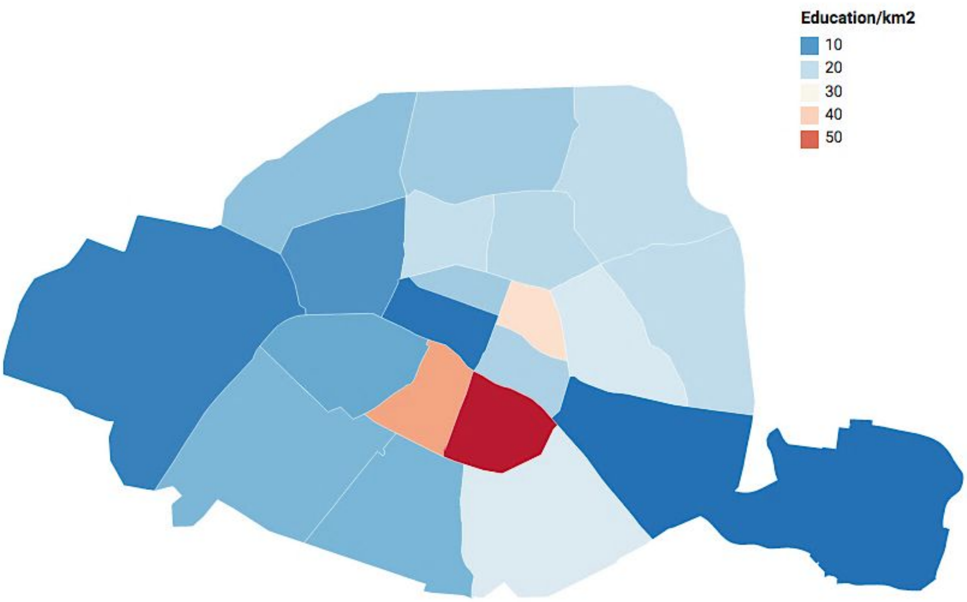


Figure 12. Paris education offering density map

In line with the absence of correlation observed with municipal facilities, the fact that education offering does not weigh in real estate prices in Paris may tell something from a sociological point of view.

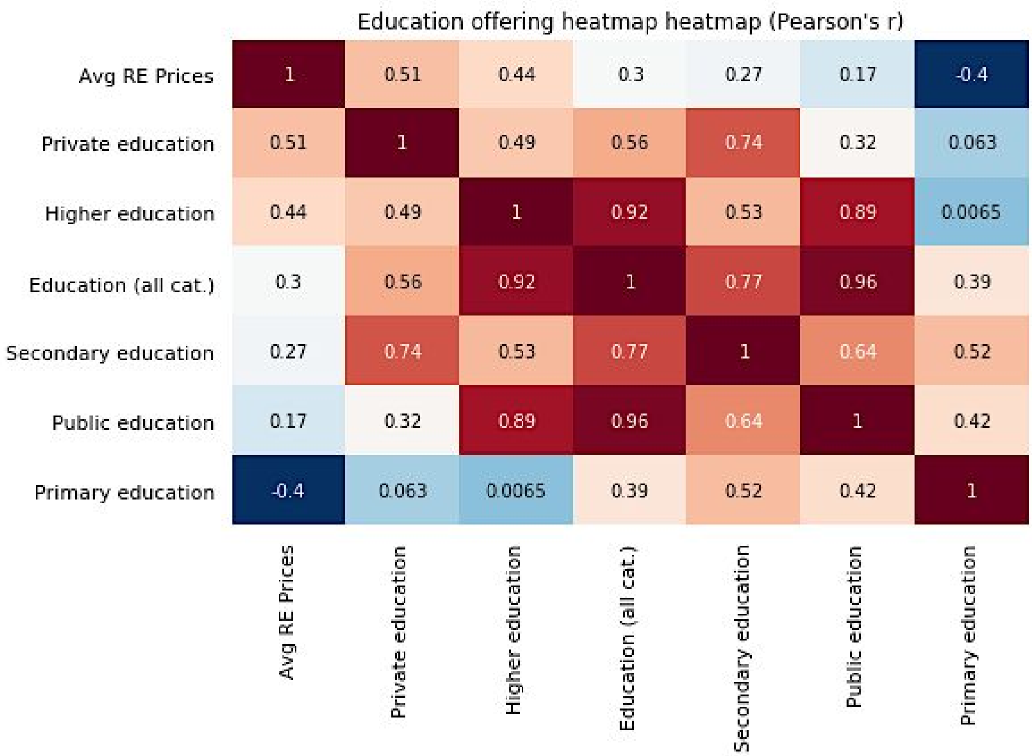


Figure 13. Education offering heatmap

2.3.8 Social housing analysis

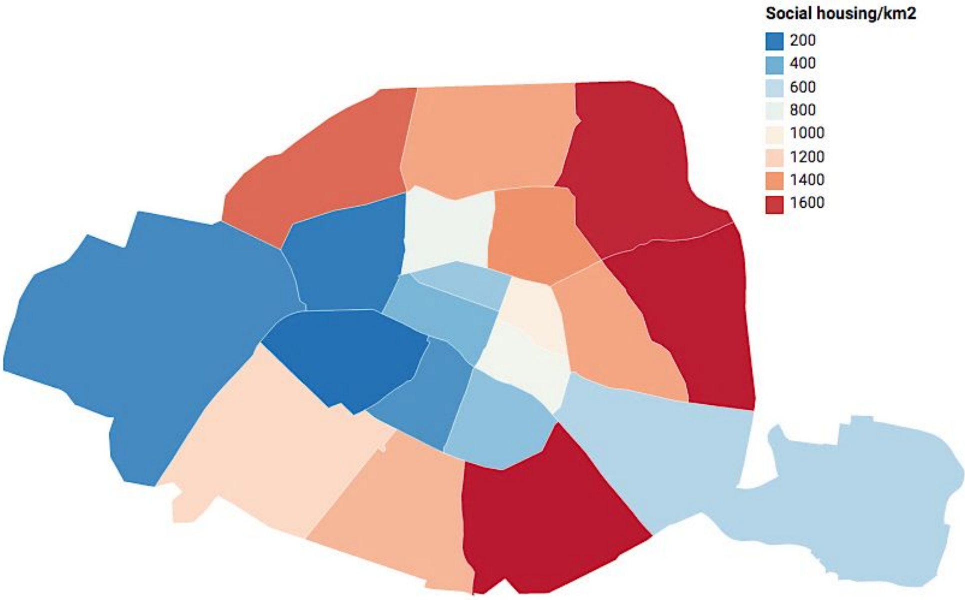


Figure 14. Paris social housing density map

A clear pattern raises that the most expensive boroughs, which are concentrated in the heart of Paris, are "preserved" from social housing (see figure 14). Around, the 20th, 13th, 19th and 17th borough see the highest social housing density (>15000/km2).

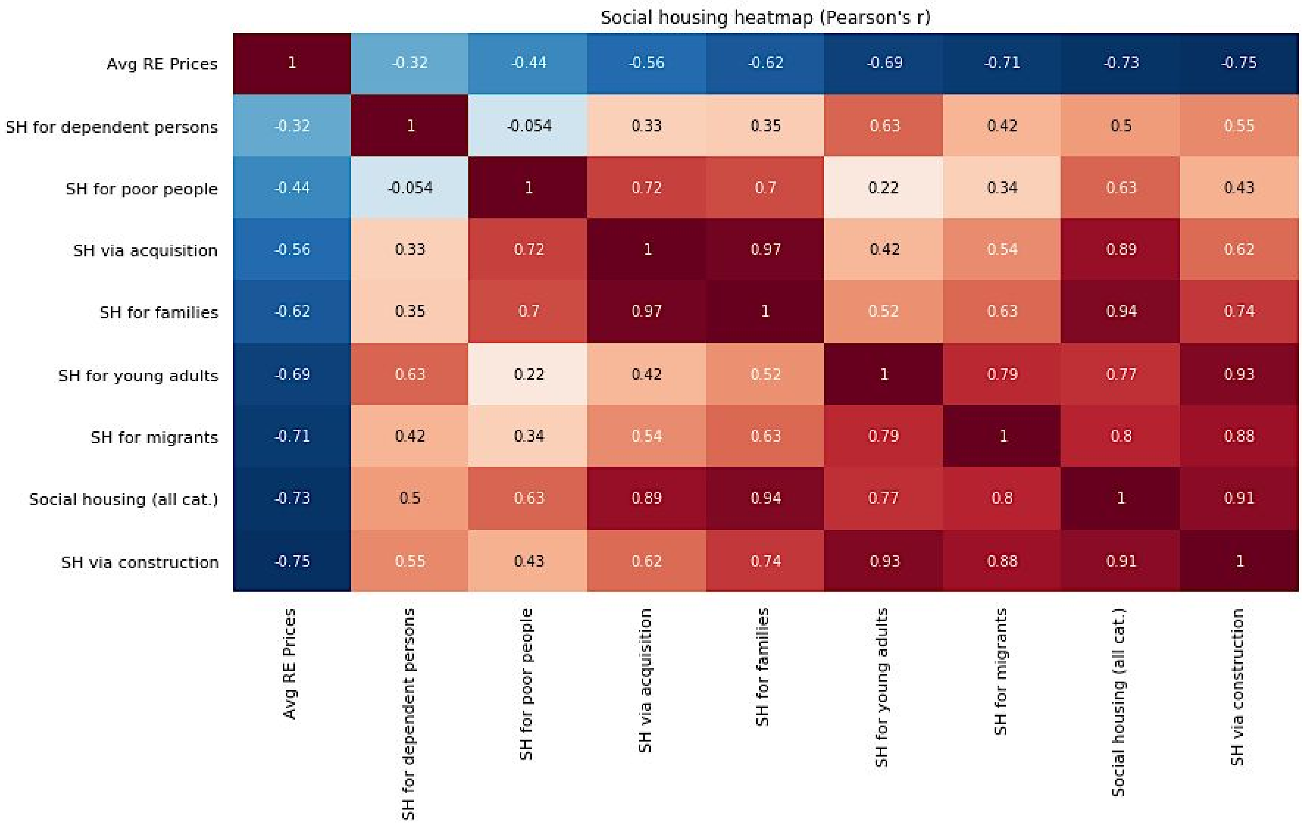


Figure 15. Social housing heatmap

This category indeed includes the factors with the most negative weight in the average property prices in Paris (see figure 15).

In particular, regardless of the type of beneficiaries (family, dependent persons, young adults, poor people, migrants), the more social housing built (as opposed to acquired: r = -0.55, p = 0.01) per square kilometer, the lower the market value (r = -0.75, p = 0.0001 | see figure 16).

Looking at the target population, we can see that the density of social housing for migrants shows the most negative correlation (r = -0.7, p = 0.0005), followed by social housing for young adults (r = -0.69, p = 0.0008). To a lesser extent, social housing for families also makes the prices lower (r = -0.61, p = 0.003).

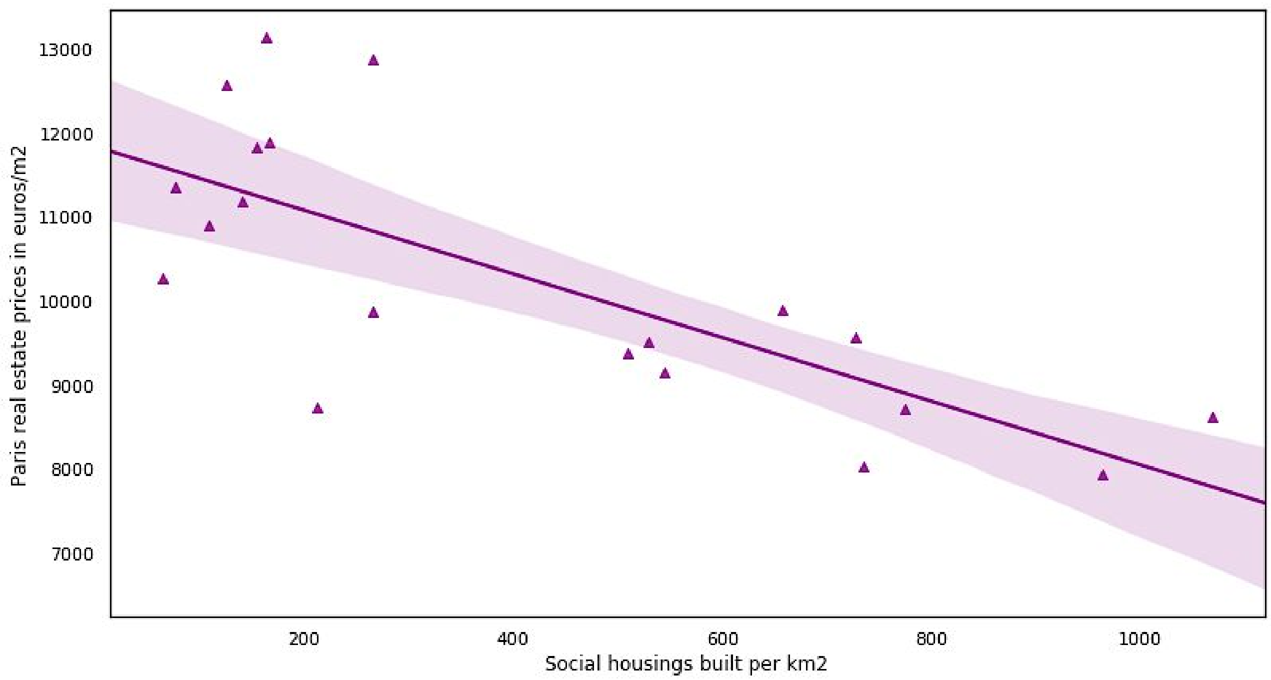


Figure 16. Social housing via construction regression plot

2.4 Modeling

2.4.1 Regression

Multiple Linear Regression and Random Forest Regression algorithms were chosen for modeling as first-line given that we want to examine which variables have an effect and how on a continuous value, i.e. real estate prices [4]. Given the small size of the dataset, a limited number of features was tested.

Feature selection was driven by the strength of correlations between the different indicators and the target (see table 2 in appendix). To gradually refine the models, indicators were tested in descending order of correlation, using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R2) as the evaluation metrics.

The acceptable limit error was set to 5% of the lowest average price (7930 euros/m2) i.e. 396.5 euros, which represents 3% of the highest average price (13150 euros/m2).

A. 2-features model

a) *Popular touristic sites* and social housing for migrants has been added to the initial 3-features model (a) in which configuration the ML algorithm performed better.

As a result, the ML model actually saw an increase in accuracy, providing the lowest root mean squared error compared to the previous three models i.e. 5,1% which is slightly above the 5% limit error, as well as a mean absolute error of 4,5% (RMSE = 407.6, MAE = 358.4, R2 = 0.94 | see figure 24). In contrast, the performance of the RF model decreased significantly (RMSE = 729.5, MAE = 550.4, R2 = 0.82 | see figure 25).

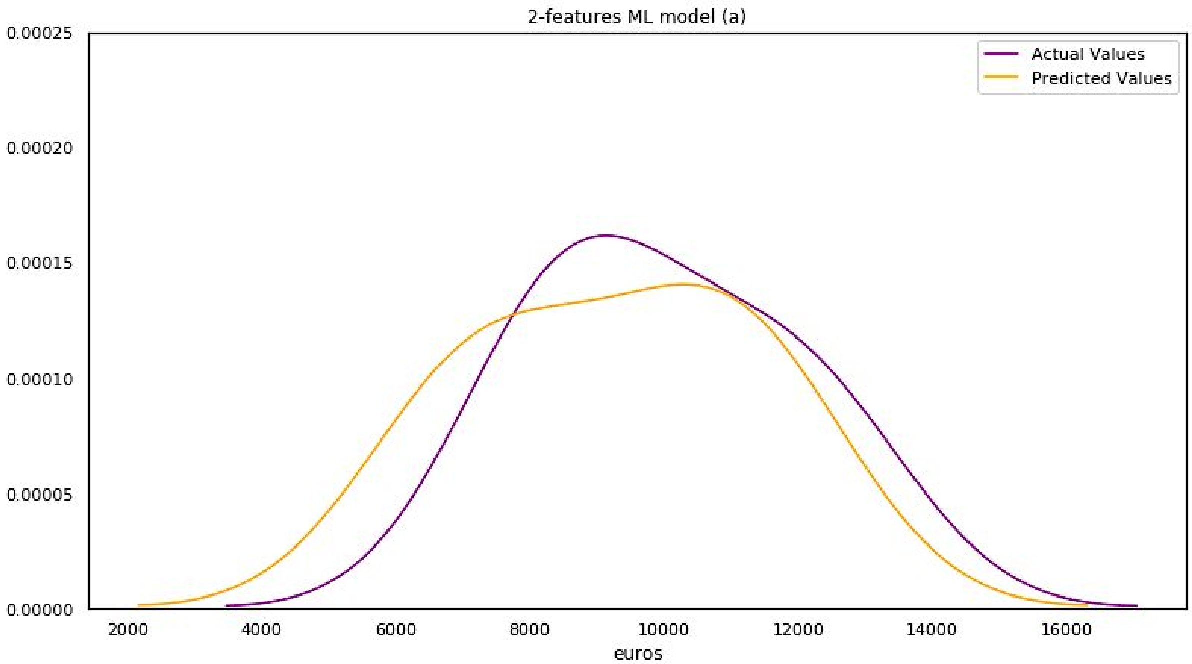


Figure 17. 2-features model (a) with Multilinear Regression

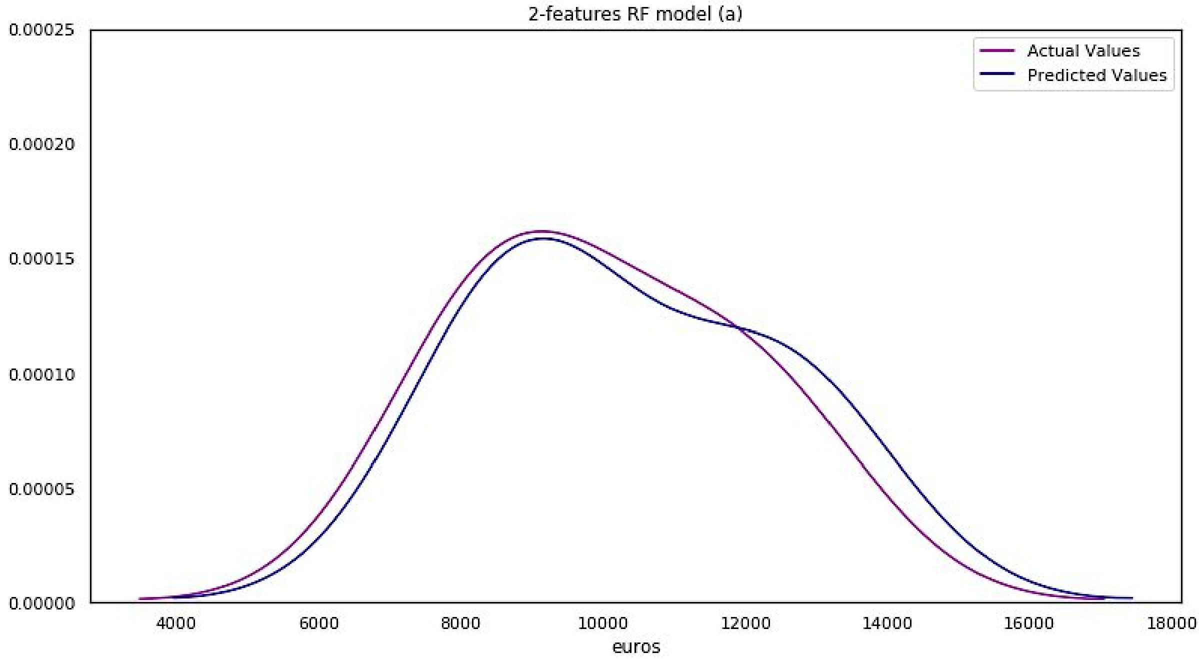


Figure 18. 4-features model (a) with Random Forest Regression

b) Here *popular culture & entertainment places* has been added to the initial 3-features model (a) in which configuration the ML algorithm performed better.

As a result, the ML model actually saw an increase in accuracy, providing the lowest root mean squared error compared to the previous three models i.e. 5,1% which is slightly above the 5% limit error, as well as a mean absolute error of 4,5% (RMSE = 407.6, MAE = 358.4, R2 = 0.94 | see figure 24). In contrast, the performance of the RF model decreased significantly (RMSE = 729.5, MAE = 550.4, R2 = 0.82 | see figure 25).

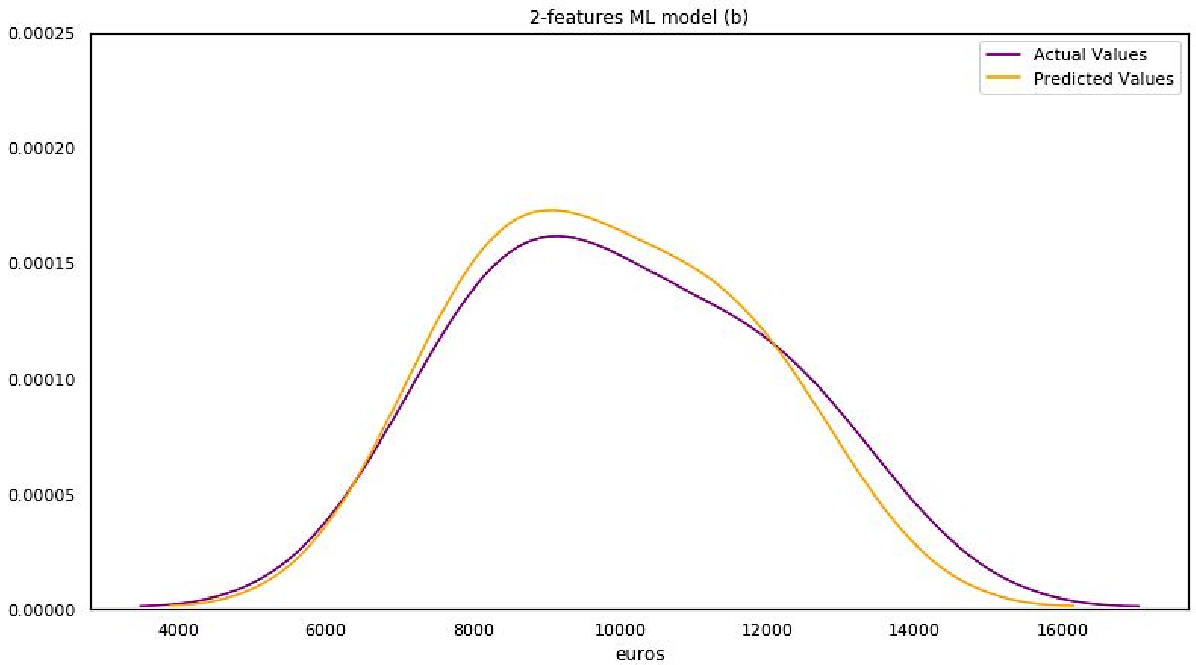


Figure 19. 2-features model (b) with Multilinear Regression

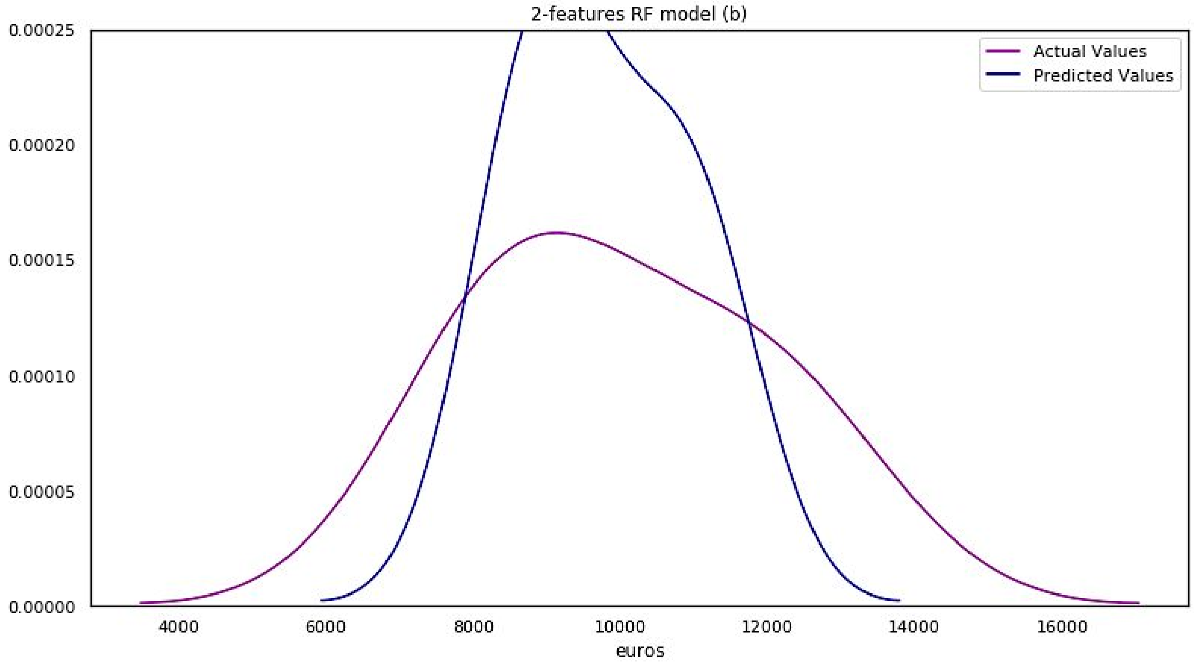


Figure 20. 2-features model (b) with Random Forest Regression

B. 3-features model

a) The 3 features presenting the highest correlation with the target variable were tested in first place: *popular hotels*, *popular touristic sites* and *social housing built*.

Both the Multilinear (RMSE = 538.07, MAE = 430.4, R2 = 0.9) and Random Forest (RMSE = 554.53, MAE = 415.1, R2 = 0.89) models demonstrated quite good performance (see figure 18 and 19), confirming the weight of these variables in the prices of real estate in Paris, yet above the limit error of 5%.

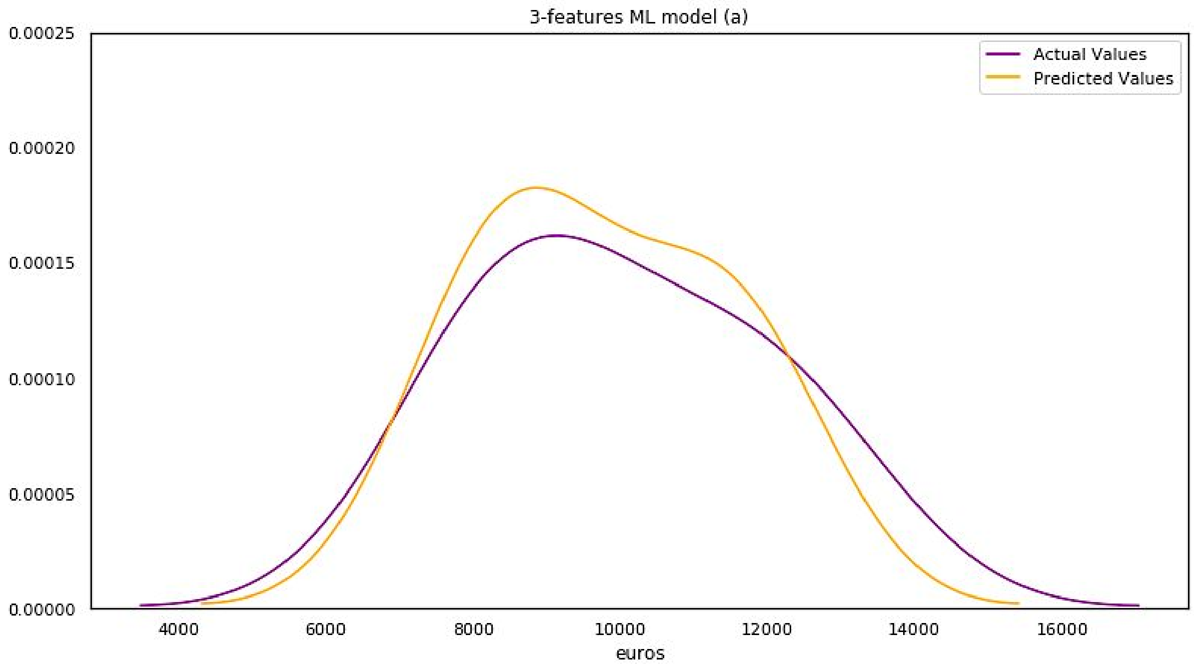


Figure 21. 3-features model (a) with Multilinear Regression

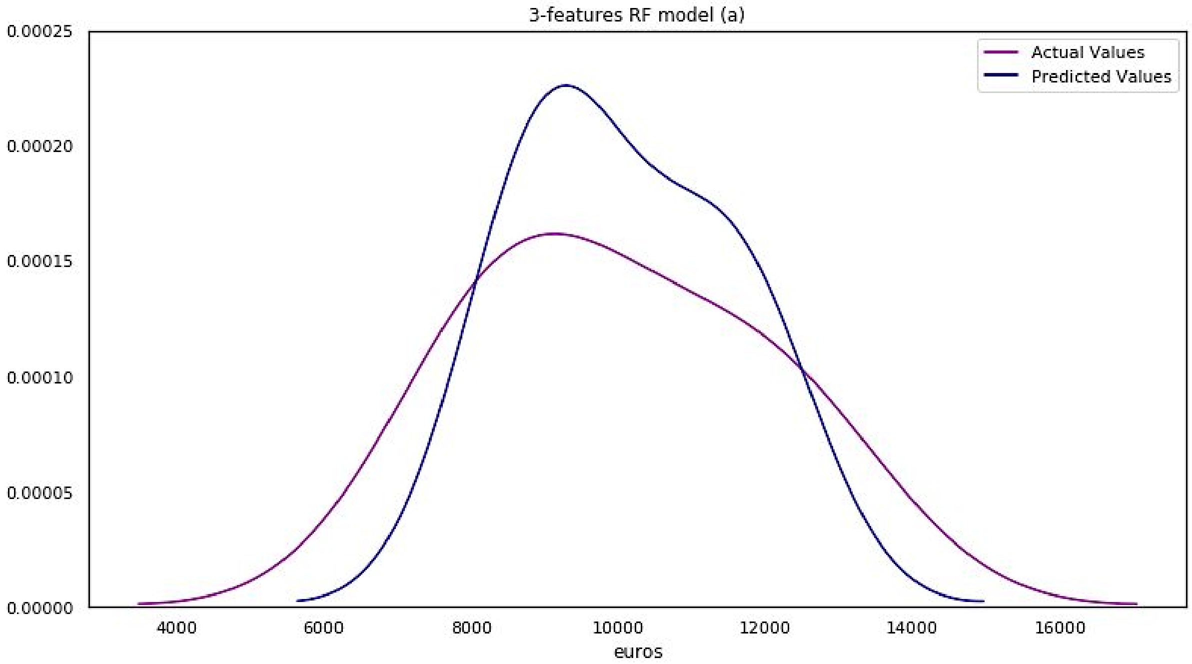


Figure 22. 3-features model (a) with Random Forest Regression

b) *Social housing built* was then replaced by *social housing for migrants*.

As a result, the RF model saw a sharp increase in performance with a mean absolute error of only 3,9% (MAE = 310.1, RMSE = 519.8, R2 = 0.9 | see figure 20). In contrast, the performance of the ML model decreased significantly (RMSE = 828.9, MAE = 598.9, R2 = 0.77 | see figure 21).

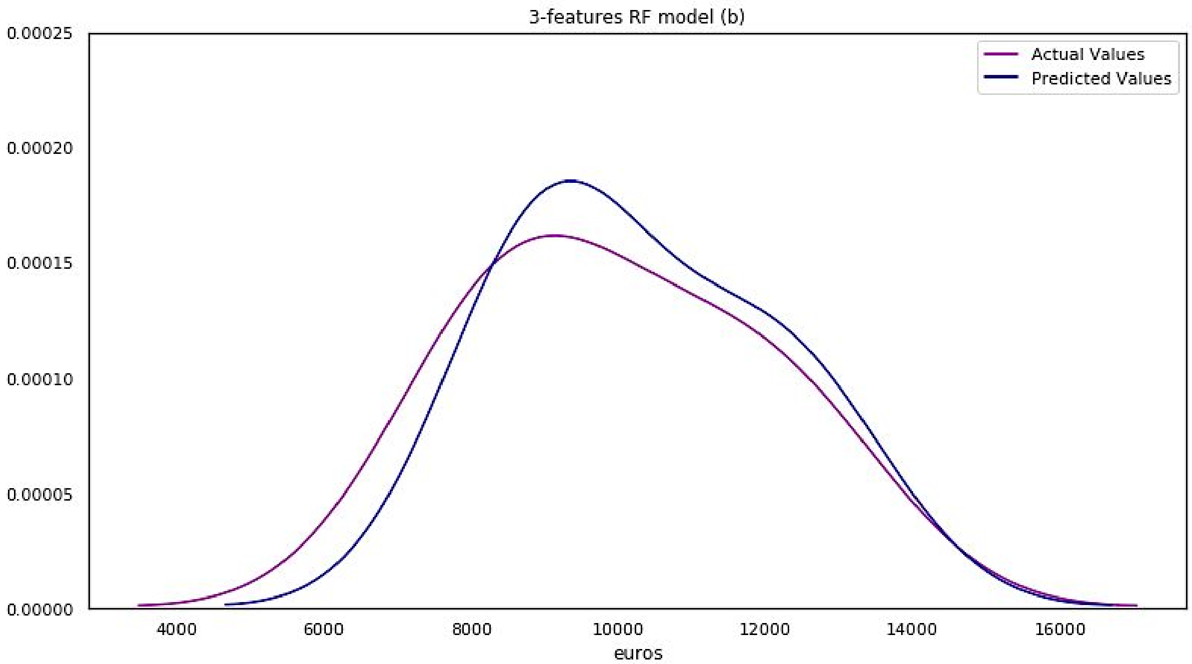


Figure 23. 3-features model (b) with Random Forest Regression

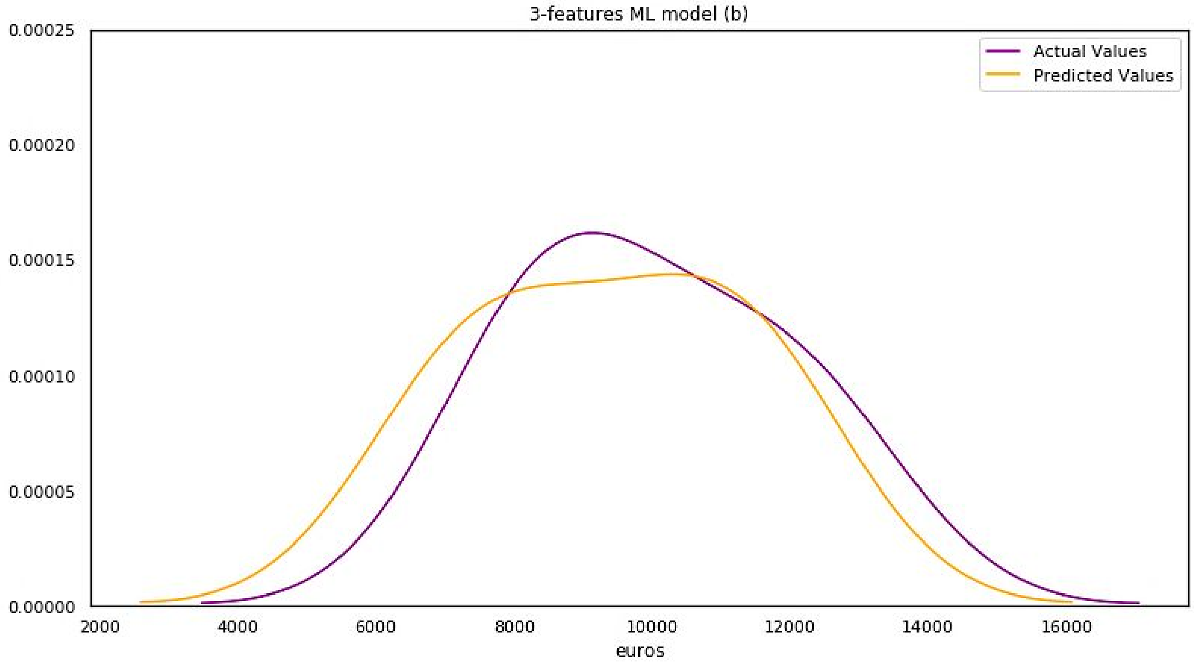


Figure 24. 3-features model (b) with Multilinear Regression

c) In order to correct the persistent tendency towards overestimation shown by the RF model, *popular hotels* has been replaced here by *popular culture and entertainment places.*

The RF model then provided a better root mean squared error value than previously, indicating a decrease in larger errors, and still a mean absolute error below 5% (RMSE = 461.5, MAE = 339.5, R2 = 0.92 | see figure 22). The performance of the ML model however continued to decrease significantly (RMSE = 1045.1, MAE = 791.5, R2 = 0.63 | see figure 23).

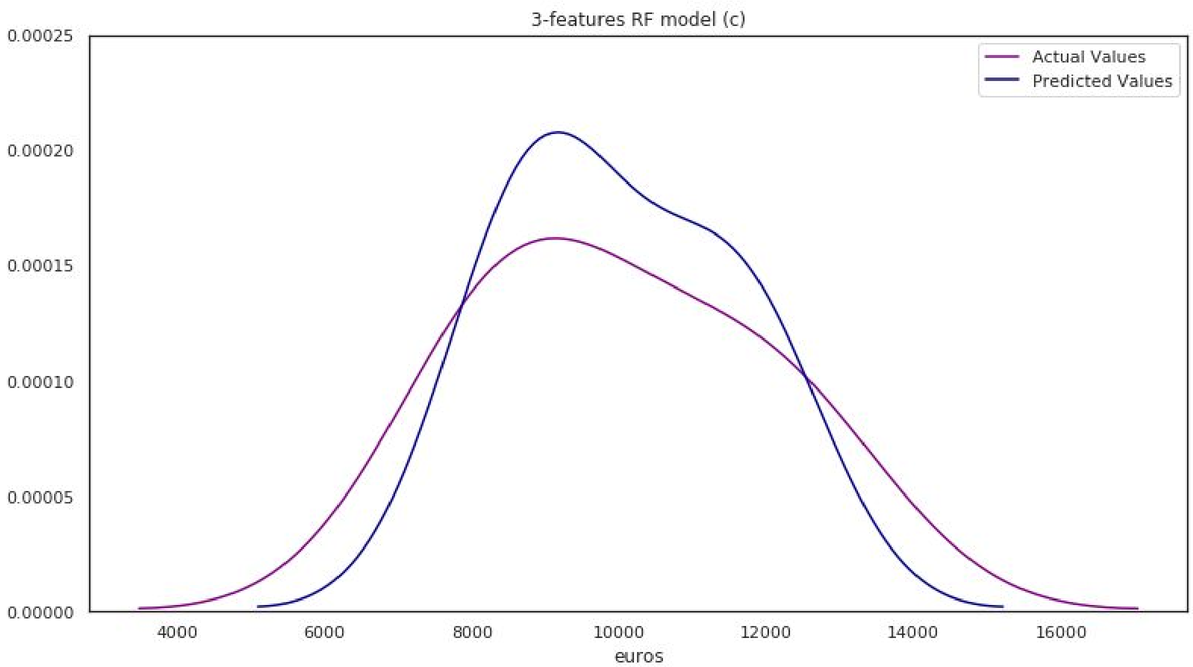


Figure 25. 3-features model (c) with Random Forest Regression

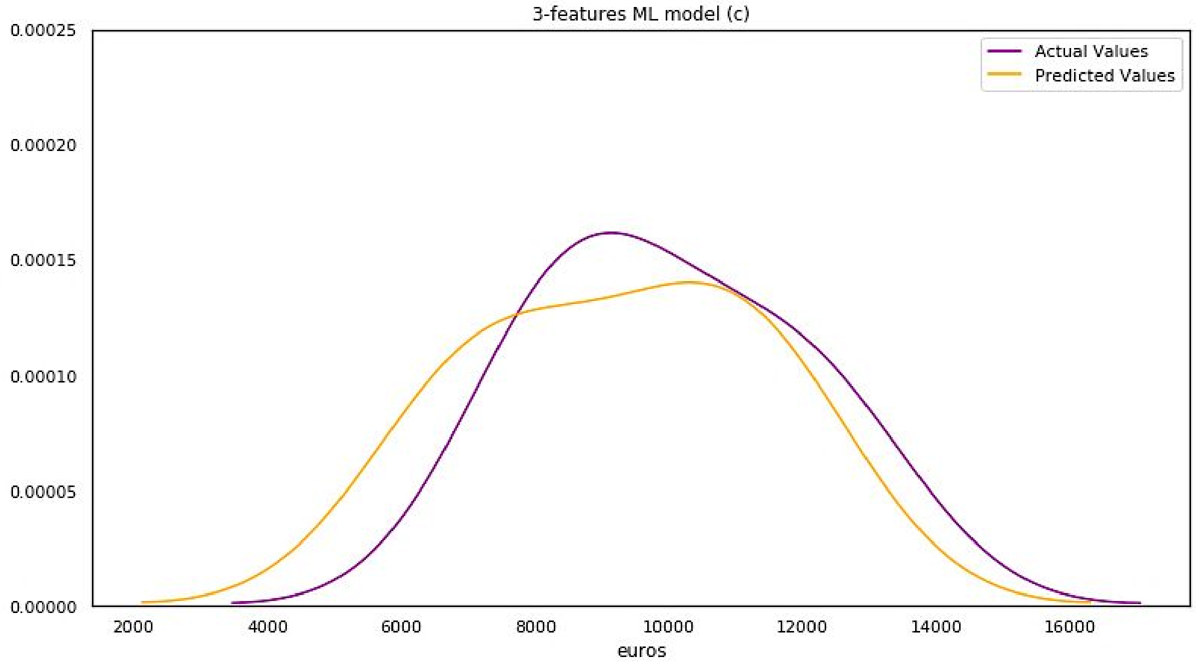


Figure 26. 3-features model (c) with Multilinear Regression

		Multilinear		**Random forest**	
Model	**Metrics**	*test*	*train*	*test*	*train*
2-features model (a) Popular touristic sites Social housing for migrants	RMSE	1044.6	837.4	503.5	347.1
	MAE	792	700.5	**395.9**	**303.03**
	R2	0.63	0.68	**0.91**	**0.94**
	R2 adj	0.26	0.63	0.83	0.93
2-features model (b) Popular hotels Social housing built	RMSE	448.6	966.2	836.4	544.1
	MAE	392.8	688.3	661.1	395.2
	R2	0.93	0.58	0.76	0.86
	R2 adj	0.86	0.51	0.53	0.84
3-features model (a) Popular hotels Popular touristic sites Social housing built	RMSE	538.07	836.7	554.5	409.6
	MAE	430.4	694.07	**415.1**	**333.03**
	R2	0.9	0.69	**0.89**	**0.92**
	R2 adj	0.61	0.60	0.90	0.58
3-features model (b) Popular hotels Popular touristic sites Social housing for migrants	RMSE	828.9	808.6	519.8	324.1
	MAE	598.9	705.5	**310.1**	**283.86**
	R2	0.77	0.71	**0.9**	**0.95**
	R2 adj	0.08	0.63	0.63	0.94
3-features model (c) Popular touristic sites Popular culture & entertainment places Social housing for migrants	RMSE	1045.1	837.03	**461.5**	316.3
	MAE	791.5	701.5	**339.5**	**256.3**
	R2	0.63	0.69	**0.92**	**0.95**
	R2 adj	-0.46	0.60	0.71	0.94

Table 3. Summary of regression model performance

2.4.2 Classification

In a second step, classification algorithms were used to highlight features that would have escaped the previous analysis. This stage confirmed the importance of the variables formerly identified, that are *social housing* and *popular venues*.

To that end, 4 price classes was created with the aim of balancing the number of observations in each sample:

- price < 9000 euros/m2 (n=5)
- price < 10000 euros/m2 (n=6)
- price < 11500 euros/m2 (n=4)
- price > 11500/m2 (n=5)

Popular touristic sites + social housing for migrants:
SVM accuracy = 0.8
svclassifier = SVC(kernel='poly', degree=2, gamma='scale', C=100)

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         2
           1       1.00      1.00      1.00         1
           2       0.50      1.00      0.67         1
           3       0.00      0.00      0.00         1

   micro avg       0.80      0.80      0.80         5
   macro avg       0.62      0.75      0.67         5
weighted avg       0.70      0.80      0.73         5
```

Popular hotels + social housing built:
SVM accuracy = 0.6
svclassifier = SVC(kernel='poly', degree=2, gamma='scale', C=100)

```
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         2
           1       0.50      1.00      0.67         1
           2       0.50      1.00      0.67         1
           3       0.00      0.00      0.00         1

   micro avg       0.60      0.60      0.60         5
   macro avg       0.50      0.62      0.50         5
weighted avg       0.60      0.60      0.53         5
```

A. 21 features

21 features among the moderate to strong regressors (see table 2) fed the algorithms. In this configuration, *social housing* was reduced to its binary modalities 'via construction' vs 'via acquisition', as well as the *education offering* which was presented according to its 'private' vs 'public' modalities.

a) Decision tree

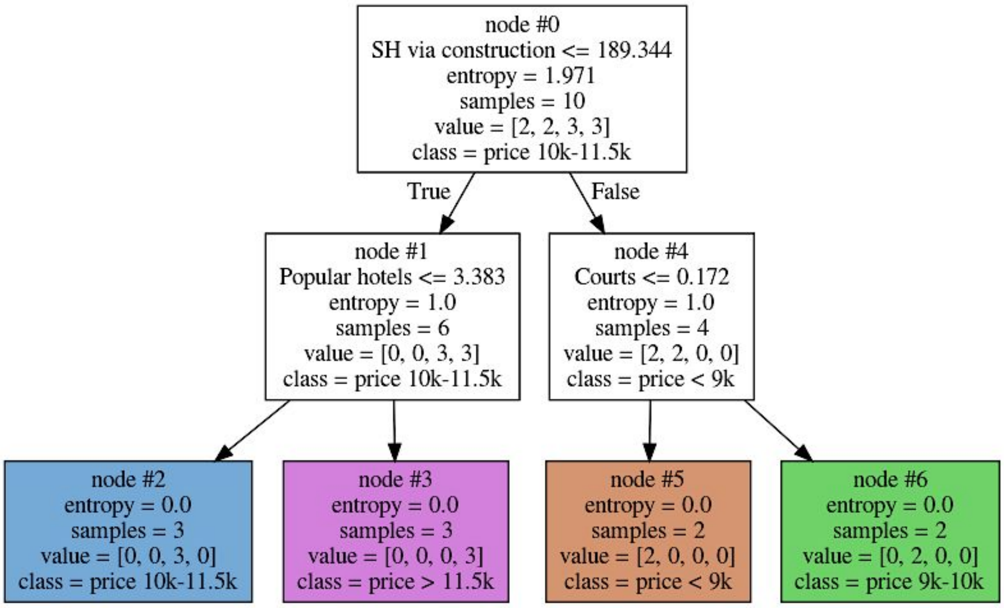


*Figure 27. Decision tree * 21 features*

Best accuracy decision tree: Jaccard score = 0.8, F1 score = 0.79

The decision tree identified the threshold of 189 social housing units built per square kilometer as a critical level above which average prices are below 10000 euros/m2.

As a rank 2 condition, a density of more than 3 popular hotels/km2 determines the greater price class, that is above 11500 euros/m2.

b) Random forest tree

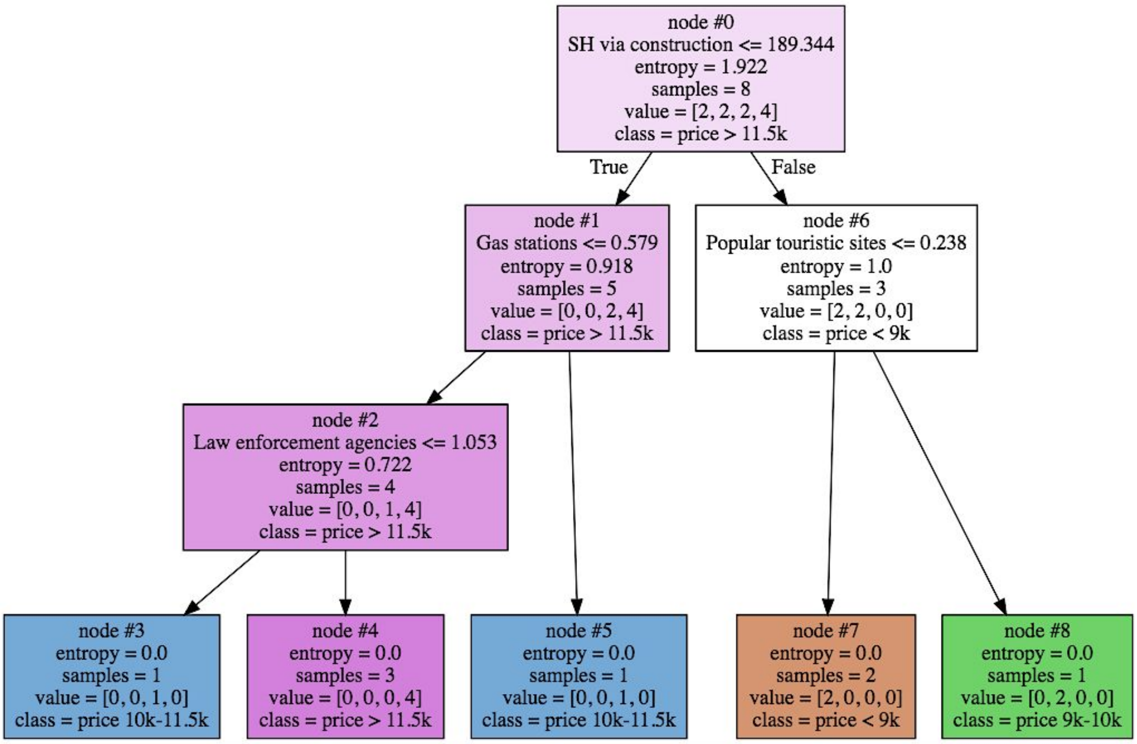


*Figure 28. Random forest tree * 21 features*

Accuracy score = 0.7

	precision	recall	f1-score	support
< 9k	0.75	1.00	0.86	3
9k-10k	1.00	0.50	0.67	4
10k-11.5k	0.50	1.00	0.67	1
> 11.5k	0.50	0.50	0.69	2
micro avg	0.70	0.70	0.70	10
macro avg	0.69	0.75	0.67	10
weighted avg	0.78	0.70	0.69	10

The random forest generated a more complex model but also identified the threshold of 189 social housing units built/km2 as a critical level and defined it straightaway as the key for the highest price class.

A density of more than 2 popular touristic sites/km2 separates the lower prices ranges.

B. 23 features

In this configuration, the *social housing* category was presented according to the strongest modalities of its different types of beneficiaries – 'for migrants', 'for young adults', 'for families'; the education offering was also considered according to the age group addressed – 'primary', 'secondary', 'higher'.

a) Decision tree

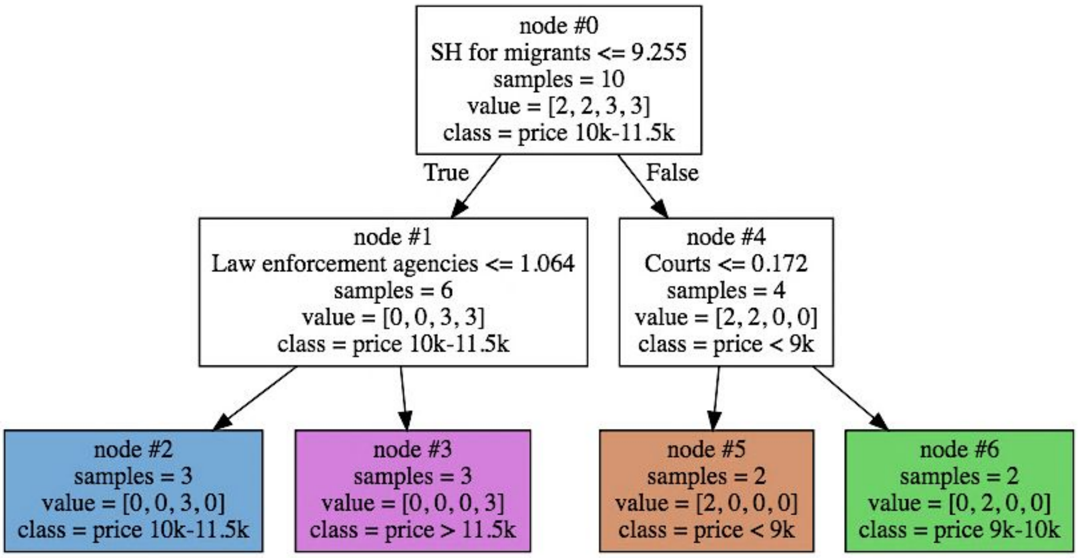


*Figure 29. Decision tree * 23 features*

Best accuracy decision tree: Jaccard score = 0.8, F1 score = 0.8

Here, the decision tree highlighted the threshold of 9 social housing units for migrants per square kilometer as a critical level above which average prices go under 10000 euros/m2.

As a rank 2 condition, a density of more than 1 law enforcement agency/km2 determines the greater price class.

b) Random forest tree

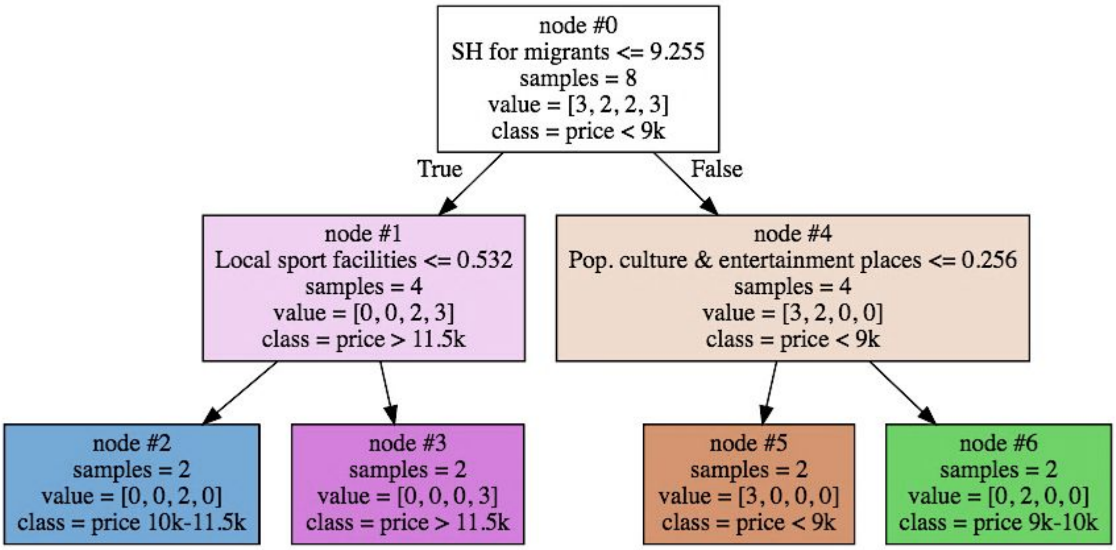


*Figure 30. Random forest tree * 23 features*

Accuracy score = 0.6

	precision	recall	f1-score	support
< 9k	0.60	1.00	0.75	3
9k-10k	1.00	0.25	0.40	4
10k-11.5k	0.50	1.00	0.67	1
> 11.5k	0.50	0.50	0.50	2
micro avg	0.60	0.60	0.60	10
macro avg	0.65	0.69	0.58	10
weighted avg	0.73	0.60	0.55	10

3. Results and discussion

Touristic and cultural attractions feed the price

Within the "popular venues" category (r = 0.64, p = 0.002), **the density of popular hotels** (r = 0.75, p = 0.0001), **touristic sites** (Pearson's r = 0.73, p = 0.0003) and **culture and entertainment venues** (r = 0.67, p = 0.001) appear to contribute significantly to the price per square meter.

In short, the more touristic a borough is, the higher the property price.

Although the "shop offering" category shows only a moderate correlation (r = 0.58, p = 0.006), it is yet consistent with the shops and markets popularity indicator. In particular, the density of personal goods shops per square kilometer is not without effect (r = 0.6, p = 0.004).

Unsurprisingly, the weight of this type of parameter reflects the hedonistic nature of the real estate market, especially in big cities.

However, it must be emphasized that popularity metrics being by definition generated by humans, they are likely to fluctuate with trends. In addition, it would be good to cross the data provided by FourSquare with those from other platforms of the same type to ensure the representativeness of the data processed in this study.

A cheaper market comes with social mix

On the other hand, all the features pertaining to the "social housing" category clearly make the real estate prices lower (r = -0.73, p = 0.0002). In particular, **the density of social housing built** in a given borough show the strongest negative correlation (r = -0.75, p = 0.0001).

One can intuitively assume that the larger the borough, the more room there is for social housing construction, which would explain why some boroughs have more than others. The moderate correlation (r = 0.42, p = 0.06) between area and the total of social housing built does not allow to confirm that hypothesis. So, if it is not the area, we can assume that the decision to build social housing in one borough rather than another would be driven by the willing to preserve the prestige of certain boroughs and, consequently, their market value.

What is clear, however, is that the more the social housing construction, the higher the population (r = 0.84, p = 0.0). This is even more obvious when all social housing modalities (social housing for migrants, for young adults, for poor people, for families, for dependent persons) are grouped (r = 0.92, p = 0.0). Then, does a link exist between population density and real estate prices? As we saw earlier (see chapter 2.3.2), the answer is yes but moderate (r = -0.47, p = 0.03): a slight tendency that the more populated a borough, the lower the price.

Services, municipal facilities and education offering are out of the game

Finally, contrary to what is commonly put forward to explain the weight of the location of the property, the overall service offering (r = 0.05, p = 0.82), municipal facilities (r = -0.15, p = 0.52) and education offering (r = 0.29, p = 0.2) per square kilometer demonstrate no effect.

Only can we observe some moderate correlations: densities of courts (r = 0.58, p = 0.006), law enforcement offices (r = 0.57, p = 0.008) and private schools (r = 0.51, p = 0.02) tend to make prices higher.

In contrast, local sports facilities (r = -0.51, p = 0.02) appear to have a negative impact. Overall, the quite cold map presented by municipal facilities indicators raises hypotheses on Parisian sociology. Further research may indeed focus on the profile of residents by borough by examining variables such as family and professional status.

Three – four factors decide the price

At the modeling stage (Multiple Linear Regression, Random Forest Regression), four predictors proved to be together the more explanatory: 'Popular hotels', 'Popular touristic sites', 'Social housing built' within which 'Social housing for migrants'. With regard to the multilinear model metrics, 'Popular culture & entertainment places' may be added as a positive regressor.

The regression models explain 90% to 94% of the average price variation, with a mean absolute error from 358.4 to 310.1 euros according to the model, which represents only 4,5% and 3,9% error regarding the lowest price and 2,7% and 2,3% regarding the highest price.

Of course, there is still room for improvement as the models do not include intrinsic attributes of the property (area, modernity, …). But the present study was not meant to build a comprehensive model.

Limitations

The study was based on the average prices by district reported by Paris Notary Chamber, not on property prices observed individually. While these averages do reflect differences in market value from one borough to another, they inevitably entail a loss of information. At the same time, they avoid outliers, which are highly undesirable when it comes to small data sets. Here lies another limitation: the number of observations in the present study was indeed 20, that is the number of boroughs in Paris. This calls for caution in interpreting modeling performance. However, this does not call into question the strong correlations observed (both with Pearson's r and Spearman's rho) nor the descriptive interest of the study.

4. Conclusion

This study was meant to shed light on the differences in the price of real estate from one Parisian borough to another. After comparing 6 groups of variables forming a set of 49 features, 5 significant regressors emerged: on one hand, the popularity of hotels, touristic sites and culture and entertainment places increases the market value of a property; on the other hand, social housing and in particular when it is built or addressed to migrants reduces it. Contrary to what is generally presented as an explanation of what makes a property more expensive, services, municipal facilities and education offering showed no impact.

Hence, considering the location criterion in Paris, it appears that the high price is paid for the least social mix and the most touristic and cultural attractions, not for attributes that benefit to family life or make day-to-day easier. A similar pattern may govern other metropolises in the world such as New York and London, which could be the subject of another study.

To go further, variables such as income, age, education level, family and professional status should also be examined to capture the sociological dimension of the Parisian market.

Sources and additional readings

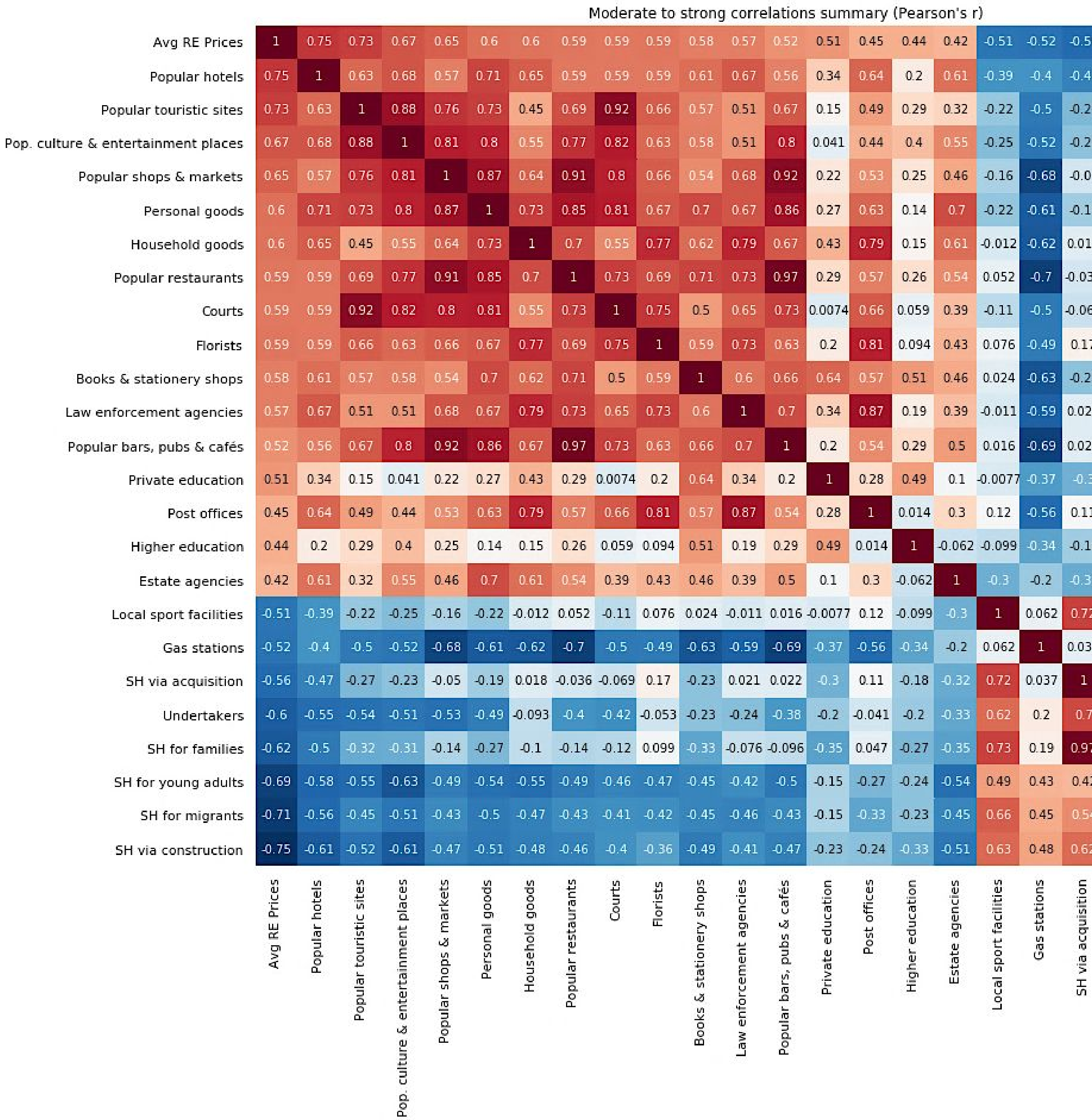
All details of data preparation and feature engineering are provided with the Jupyter notebook 1:
https://github.com/StfBlanchet/real_estate_study/blob/master/2.Paris_Boroughs_Study_Analysis%26Modeling_revised.ipynb

All details of data analysis and modeling are provided with the Jupyter notebook 2:
https://github.com/StfBlanchet/real_estate_study/blob/master/2.Paris_Boroughs_Study_Analysis%26Modeling.ipynb

Appendix

Variables	**Pearson's r**	**p**	**Spearman's rho**	**p**
POSITIVE CORRELATIONS				
Popular hotels per km2	**0.75**	0.0001	**0.72**	0.0003
Popular touristic sites per km2	**0.73**	0.0003	**0.7**	0.0005
Popular culture & entertainment places per km2	**0.67**	0.001	**0.73**	0.0003
Popular shops & markets per km2	**0.65**	0.002	**0.64**	0.002
Personal goods stores per km2	**0.6**	0.004	**0.65**	0.001
Household goods stores per km2	**0.6**	0.005	**0.6**	0.005
Popular restaurants per km2	0.59	0.005	**0.65**	0.002
Courts per km2	0.59	0.006	**0.64**	0.002
Florists per km2	0.59	0.006	0.55	0.01
Books & stationery shops per km2	0.58	0.008	**0.64**	0.002
Law enforcement agencies per km2	0.57	0.008	0.58	0.007
Private education establishments per km2	0.51	0.02	0.41	0.07
Popular bars, pubs & cafés per km2	0.5	0.02	0.52	0.02
Higher education establishments per km2	0.44	0.05	0.59	0.006
Post offices per km2	0.44	0.05	0.5	0.02
Estate agencies per km2	0.42	0.06	**0.64**	0.002
Banks per km2	0.28	0.2	0.59	0.006
NEGATIVE CORRELATIONS				
Social housing construction per km2	**-0.75**	0.0001	**-0.74**	0.0002
Social housing for migrants per km2	**-0.7**	0.0005	**-0.86**	0.0
Social housing for young adults per km2	**-0.69**	0.0008	**-0.8**	0.0
Social housing for families per km2	**-0.62**	0.003	**-0.63**	0.003
Undertakers per km2	**-0.6**	0.005	-0.57	0.008
Social housing via acquisition per km2	-0.55	0.01	-0.56	0.009
Gas stations per km2	-0.52	0.02	-0.49	0.02
Local sport facilities per km2	-0.51	0.02	-0.52	0.02
Social housing for poor people per km2	-0.44	0.05	-0.5	0.03
Supermarkets per km2	-0.43	0.06	-0.44	0.05
Building craft per km2	-0.4	0.8	-0.42	0.06
Collective daycare per km2	-0.32	0.2	-0.45	0.05
Social housing for dependent persons per km2	-0.3	0.2	-0.41	0.07

Table 2. Moderate to strong correlations with Paris average real estate prices


Moderate to strong correlations summary (Pearson's r)
Avg RE Prices 1 0.75 0.73 0.67 0.65 0.6 0.6 0.59 0.59 0.59 0.58 0.57 0.52 0.51 0.45 0.44 0.42 -0.51 -0.52
Popular hotels 0.75 1 0.63 0.68 0.57 0.71 0.65 0.59 0.59 0.59 0.61 0.67 0.56 0.34 0.64 0.2 0.61 -0.39 -0.4
Popular touristic sites 0.73 0.63 1 0.88 0.76 0.73 0.45 0.69 0.92 0.66 0.57 0.51 0.67 0.15 0.49 0.29 0.32 -0.22 -0.5
Pop. culture & entertainment places 0.67 0.68 0.88 1 0.81 0.8 0.55 0.77 0.82 0.63 0.58 0.51 0.8 0.041 0.44 0.4 0.55 -0.25 -0.52
Popular shops & markets 0.65 0.57 0.76 0.81 1 0.87 0.64 0.91 0.8 0.66 0.54 0.68 0.92 0.22 0.53 0.25 0.46 -0.16 -0.68
Personal goods 0.6 0.71 0.73 0.8 0.87 1 0.73 0.85 0.81 0.67 0.7 0.67 0.86 0.27 0.63 0.14 0.7 -0.22 -0.61
Household goods 0.6 0.65 0.45 0.55 0.64 0.73 1 0.7 0.55 0.77 0.62 0.79 0.67 0.43 0.79 0.15 0.61 -0.012 -0.62
Popular restaurants 0.59 0.59 0.69 0.77 0.91 0.85 0.7 1 0.73 0.69 0.71 0.73 0.97 0.29 0.57 0.26 0.54 0.052 -0.7
Courts 0.59 0.59 0.92 0.82 0.8 0.81 0.55 0.73 1 0.75 0.5 0.65 0.73 0.0074 0.66 0.059 0.39 -0.11 -0.5
Florists 0.59 0.59 0.66 0.63 0.66 0.67 0.77 0.69 0.75 1 0.59 0.73 0.63 0.2 0.81 0.094 0.43 0.076 -0.49
Books & stationery shops 0.58 0.61 0.57 0.58 0.54 0.7 0.62 0.71 0.5 0.59 1 0.6 0.66 0.64 0.57 0.51 0.46 0.024 -0.63
Law enforcement agencies 0.57 0.67 0.51 0.51 0.68 0.67 0.79 0.73 0.65 0.73 0.6 1 0.7 0.34 0.87 0.19 0.39 -0.011 -0.59
Popular bars, pubs & cafés 0.52 0.56 0.67 0.8 0.92 0.86 0.67 0.97 0.73 0.63 0.66 0.7 1 0.2 0.54 0.29 0.5 0.016 -0.69
Private education 0.51 0.34 0.15 0.041 0.22 0.27 0.43 0.29 0.0074 0.2 0.64 0.34 0.2 1 0.28 0.49 0.1 -0.0077 -0.37
Post offices 0.45 0.64 0.49 0.44 0.53 0.63 0.79 0.57 0.66 0.81 0.57 0.87 0.54 0.28 1 0.014 0.3 0.12 -0.56
Higher education 0.44 0.2 0.29 0.4 0.25 0.14 0.15 0.26 0.059 0.094 0.51 0.19 0.29 0.49 0.014 1 -0.062 -0.099 -0.34
Estate agencies 0.42 0.61 0.32 0.55 0.46 0.7 0.61 0.54 0.39 0.43 0.46 0.39 0.5 0.1 0.3 -0.062 1 -0.3 -0.2
Local sport facilities -0.51 -0.39 -0.22 -0.25 -0.16 -0.22 -0.012 0.052 -0.11 0.076 0.024 -0.011 0.016 -0.0077 0.12 -0.099 -0.3 1 0.062
Gas stations -0.52 -0.4 -0.5 -0.52 -0.68 -0.61 -0.62 -0.7 -0.5 -0.49 -0.63 -0.59 -0.69 -0.37 -0.56 -0.34 -0.2 0.062 1
SH via acquisition -0.56 -0.47 -0.27 -0.23 -0.05 -0.19 0.018 -0.036 -0.069 0.17 -0.23 0.021 0.022 -0.3 0.11 -0.18 -0.32 0.72 0.037 1
Undertakers -0.6 -0.55 -0.54 -0.51 -0.53 -0.49 -0.093 -0.4 -0.42 -0.053 -0.23 -0.24 -0.38 -0.2 -0.041 -0.2 -0.33 0.62 0.2
SH for families -0.62 -0.5 -0.32 -0.31 -0.14 -0.27 -0.1 -0.14 -0.12 0.099 -0.33 -0.076 -0.096 -0.35 0.047 -0.27 -0.35 0.73 0.19
SH for young adults -0.69 -0.58 -0.55 -0.63 -0.49 -0.54 -0.55 -0.49 -0.46 -0.47 -0.45 -0.42 -0.5 -0.15 -0.27 -0.24 -0.54 0.49 0.43
SH for migrants -0.71 -0.56 -0.45 -0.51 -0.43 -0.5 -0.47 -0.43 -0.41 -0.42 -0.45 -0.46 -0.43 -0.15 -0.33 -0.23 -0.45 0.66 0.45
SH via construction -0.75 -0.61 -0.52 -0.61 -0.47 -0.51 -0.48 -0.46 -0.4 -0.36 -0.49 -0.41 -0.47 -0.23 -0.24 -0.33 -0.51 0.63 0.48
Avg RE Prices
Popular hotels
Popular touristic sites
Pop. culture & entertainment places
Popular shops & markets
Personal goods
Household goods
Popular restaurants
Courts
Florists
Books & stationery shops
Law enforcement agencies
Popular bars, pubs & cafés
Private education
Post offices
Higher education
Estate agencies
Local sport facilities
Gas stations
SH via acquisition