# Choosing the perfect neighborhood

Pierre Beylard

11 May 2020

# Introduction

## Problem background

Considering some time limitation and Foursquare API requests limit, we will focus our work on the city of Bordeaux in France.

Bordeaux metropole is constituted of 28 towns dispatched on an area of 578 square kilometers[[1]](#footnote-1). The city is close to the Atlantic Ocean on the west part of France

Bordeaux is a city with a relatively rapid demographic growth around 1.5% per year during the previous 7 years. In 2017, Bordeaux had 760 000 inhabitants with an objective of 1 million inhabitants in 2030.

To achieve this objective, new neighborhoods are constructed. Furthermore, transport within the city are developed.

Of course, this growth is not only natural, and immigration plays an important role: A lot of people are starting to arrive in the city from other French cities. This growth also attracts new companies with an average of 2000 new jobs created per year thanks to these new companies[[2]](#footnote-2).

Nowadays, in Bordeaux, when we talk about real estate business, it seems clear that is at the center of a lot of expectations:

* For project owners:
  + The choice of a neighborhood to implement a business will be critical for its future success (competition, potential customers, attractivity of the neighborhood…) ;
  + When it comes to invest in a house, the decision of the neighborhood must be thoroughly deliberate as it will impact people quality of life (activities in the neighborhoods, school, shops, parks, distance to their work…) .
* For investors, this choice will have a direct impact on their investments and the profitability.

In other terms, real estate has a direct financial impact combined with a strong social impact on population lives.

## Objectives – Analytical approach

Our goal here is to provide insight on the different neighborhoods that will help decider to choose the correct place to invest.

Choosing a correct business emplacement will depend on several aspects: the type of business, the target, the population density, the competition, price per square meter of the local …).

Choosing a correct place to live will depend also on several aspects : the age of the buyer, the family structure (single, couple, kids…) , their hobbies, the place of their work, commodities, transport services, price per square meter, type of housing facilities ….

Finally, investors will be mainly interested in the capacity of the borrowers to pay of their loan, but they will also be interested in the potential price trends of the neighborhood in order to secure their investment.

In order to achieve our goal of showing relationships between neighborhoods, a descriptive approach will be conduct. We will aggregate neighborhoods in clusters depending on the following information:

* Real estate price
* Most common type of real estate properties (apartments, houses…)
* Principal venues of the neighborhoods

# Data

## Data requirements & collection:

As told above, we will need some critical information to construct our model:

* **Foursquare API** to find the main venues of given Bordeaux Metropole borough or towns
* **OpenData** [[3]](#footnote-3) Bordeaux provides some accurate information on Bordeaux Metropole towns, as location of borough and towns in GeoJSON format. Other information as locations of main public transport stop or companies are available. But we will use Foursquare to achieve this objective.
* **DVF tool** [[4]](#footnote-4) is a dataset provide by the French government regrouping the real estate transactions intervened during the last five years on the French metropolitan territory and the DOM-TOM, except for Alsace-Moselle and Mayotte. This dataset is not exhaustive as it regroups the information only on the transactions passed during the 5 last years. Even if all the information will not be available, it will give us a pretty good insight on the structure of Bordeaux real estate market ( *price, house/flat surface area, number of rooms in the transaction, address of the building…*.). in order to access this information, we have two choices: download txt files per year or use an unofficial API. We will use the API.

# Methodology:

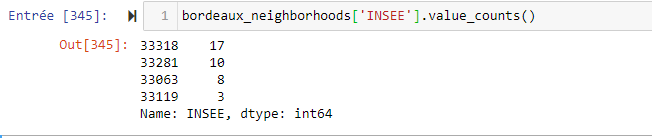
## Data understanding and preparation:

We divided our work in 3 parts:

* Obtain the correct and accurate coordinates of the neighborhoods constituting the metropole.
* Obtain information on real estate market
* Obtain local venues for each neighborhood.

### Bordeaux Neighborhoods :

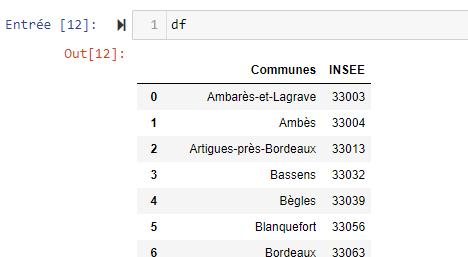
When started to work on the coordinates of neighborhoods, we rapidly faced an issue on accuracy of the data. In France, every town is represented by a unique INSEE Code. As we know that the Bordeaux metropole is constituted of 28 towns, we count the different number of INSEE codes contained in our dataset to see if we were accurate:



We only had 4 towns on a total of 28.

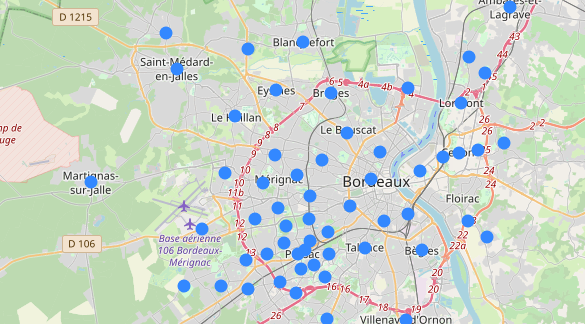
We had to find other datasets to have an accurate vision of metropole towns and neighborhoods.

We first scraped a governmental [[5]](#footnote-5)page to check which towns were not included in this first data set by taking only the town name(communes) and its associated INSEE code:



Then, we found another dataset [[6]](#footnote-6)containing all the towns in the county of Gironde (where Bordeaux Metropole is located). We add only the towns of the metropole that were not in the first data set thanks to the INSEE Code (present in the first dataset and scraped from the webpage).

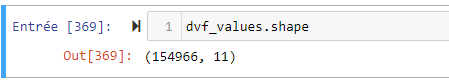
Thanks to these three steps we had an accurate dataset of Bordeaux metropole and neighborhoods.



### Real estate information:

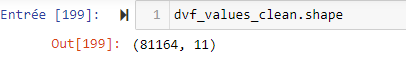
Here again, we perform some basics data analysis to explore the dataset:

* See the shape of the data frame

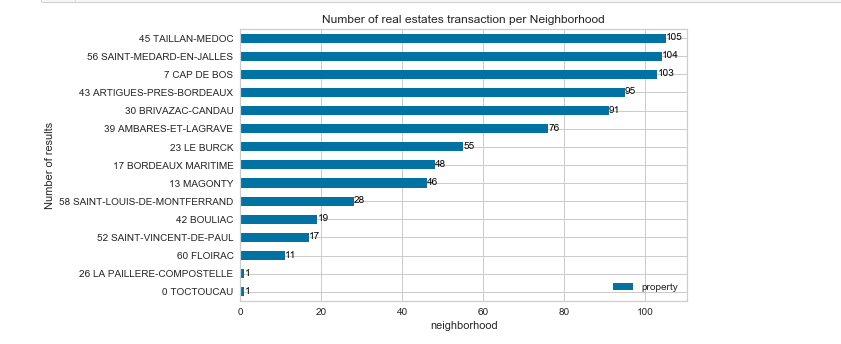


* Check presence of null values:

Our purpose is to obtain an average price per square meter. To achieve this, we need to drop all potential rows without information on ‘area’. After achieving that, we still have more than 80 000 results.



* Check the relevance of data, with this amount of results for only 62 neighborhoods, we may think that the data is accurate. However, after a short analysis, we found that 2 neighborhoods had only one result:

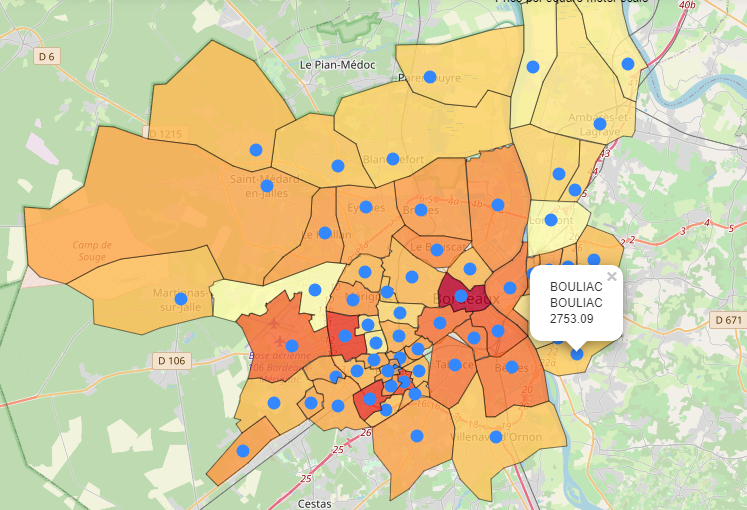


We replaced the result average price per square meter of LA PAILLERE-COMPOSTELLE and TOCTOUCAU as one result is not enough to confirm the potential price of the neighborhood. We used the average price of their town to calculate the new price for these neighborhoods.

To finally prepare the real estate data, we have operated two last modifications:

1. We have calculated the average price per square meter per transaction
2. We have grouped the data per neighborhood by calculated the median price per square meter per neighborhood and keeping only the most recurrent property type per neighborhood:

By using the geometry data obtained in the previous data set we have displayed the information of price per square meter and most common property type on a choropleth map:



Which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

## Data preparation

## Modeling

# Results

## Evaluate

where you discuss the results.

# Discussion

where you discuss any observations, you noted and any recommendations you can make based on the results.

# Conclusion

where you conclude the report.

1. Source : <http://www.bordeaux.fr/p287/bordeaux-en-chiffres> [↑](#footnote-ref-1)
2. Source : <https://www.20minutes.fr/bordeaux/2248127-20180403-bordeaux-projets-entreprises-metropole-creent-plus-plus-emplois> [↑](#footnote-ref-2)
3. <https://opendata.bordeaux-metropole.fr/> [↑](#footnote-ref-3)
4. <https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/> [↑](#footnote-ref-4)
5. <https://www.insee.fr/fr/metadonnees/cog/intercommunalite-metropole/EPCI243300316-bordeaux-metropole> [↑](#footnote-ref-5)
6. https://github.com/gregoiredavid/france-geojson/tree/master/departements/33-gironde [↑](#footnote-ref-6)