



# Capstone Project

## The Battle of Neighborhoods

Final report



A study of real estate in Val de Marne (France)

# A study of real estate in Val de Marne (France)

## Introduction

Buying a home for an individual is a complicated problem.

Real estate agencies can help and advise.

Even if the on-site visit is essential, it is possible to simplify the problem with the help of data science before a site survey.

Here we want to help Real Estate agencies and buyers to buy a property within a radius of 500 m from a metro or RER station located in the Val de Marne, a department bordering Paris,

Methods used here must allow to do easily the same kind of study on different areas or different purposes.

# Tools

- ❖ Python 3,6 on Jupyter Notebook with modules
  - ❖ Request
  - ❖ Pandas
  - ❖ Numpy
  - ❖ Matplotlib
  - ❖ Folium
  - ❖ Sklearn
  - ❖ Shapely

# Datas

Datas used for solve this problem

- ◊ Cities in Val de marne

```
url='https://public.opendatasoft.com/explore/dataset/code-insee-postaux-geoflar/download/?format=csv&q=94&refine.nom_region=ILE-DE-FRANCE&refine.nom_dept=VAL-DE-MARNE&timezone=Europe/Berlin&lang=fr&use_labels_for_header=true&csv_separator=%3B'
```

- ◊ Real Estate Sales in France for the year 2019 ( french Governement )

```
url='https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2019/departements/94.csv.gz'
```

- ◊ Metro and RER Stations

```
url='https://dataratp2.opendatasoft.com/explore/dataset/positions-geographiques-des-stations-du-reseau-ratp/download/?format=csv&timezone=Europe/Berlin&lang=fr&use_labels_for_header=true&csv_separator=%3B'
```

- ◊ Polygone shape for french Departments

```
url='https://raw.githubusercontent.com/gregoiredavid/france-geojson/master/departements.geojson'
```

# Methodology

## How the data have been used to solve the problem

### Data Cities in Val de marne

- ◊ We had to extract from the data a subset with demographic datas in this department and it is interesting to visualize the density in each city with a choropleth map
- ◊ Data Real Estate in Val de marne
  - ◊ The french government provides a database of all sales per year  
After cleaning it's possible to have the price in each city  
the aberrant values (special cases ) are dropped out.  
The data base is reduced only to houses and an average price per square meter and city is calculated  
A chloropeth map is done with this datas

## Metro and RER Stations in Val de Marne

The data base gives us all the stations in Ile de France , we have to clean it and keep only Metro and RER

A restriction to the department is done with the polyline of the department and the module shapely to know if the Station points are in the polygone.

A map with all the stations can then be drawn

## Fetching Venues with Fourquare

Foursquare allows us to know all different venues around the stations

## Segmentation with kmeans

A segmentation with Kmeans and cluster analysis allow us to better understand the environment for each train station .

A new map gathering all the datas is then proposed with legend

## Analysis

① A manual Analysis is then realised to understand the clustering done by kmean

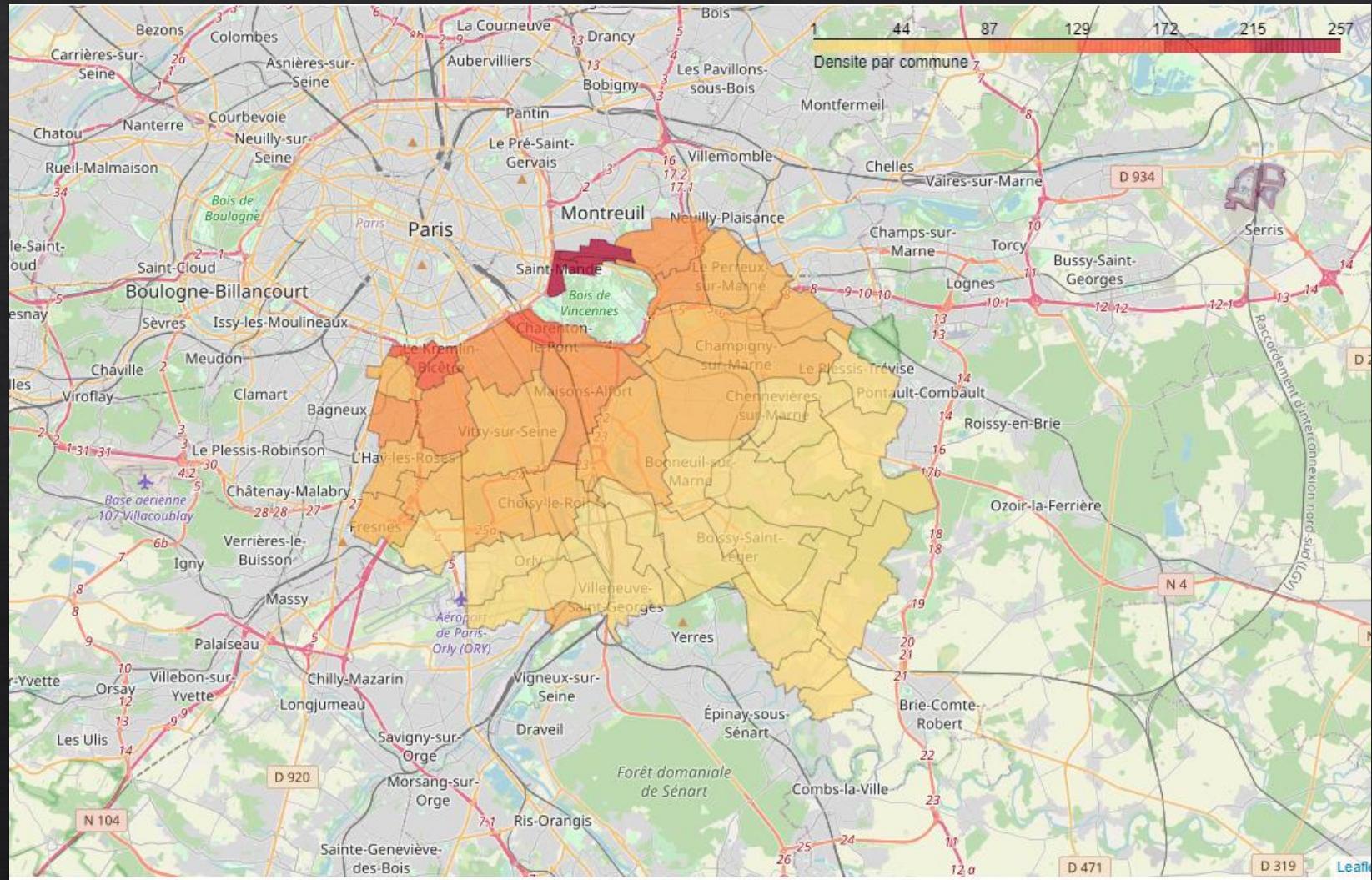
① This analysis allows us to give a type to each cluster

- ① Cluster 1 Italian – Wineshops
- ① Cluster 2 Asian, Wine-shop, Farmers market
- ① Cluster 3 Gym - Park - Fast Food
- ① Cluster 4 Diversity
- ① Cluster 5 Fast food - Farmers market - Creperie

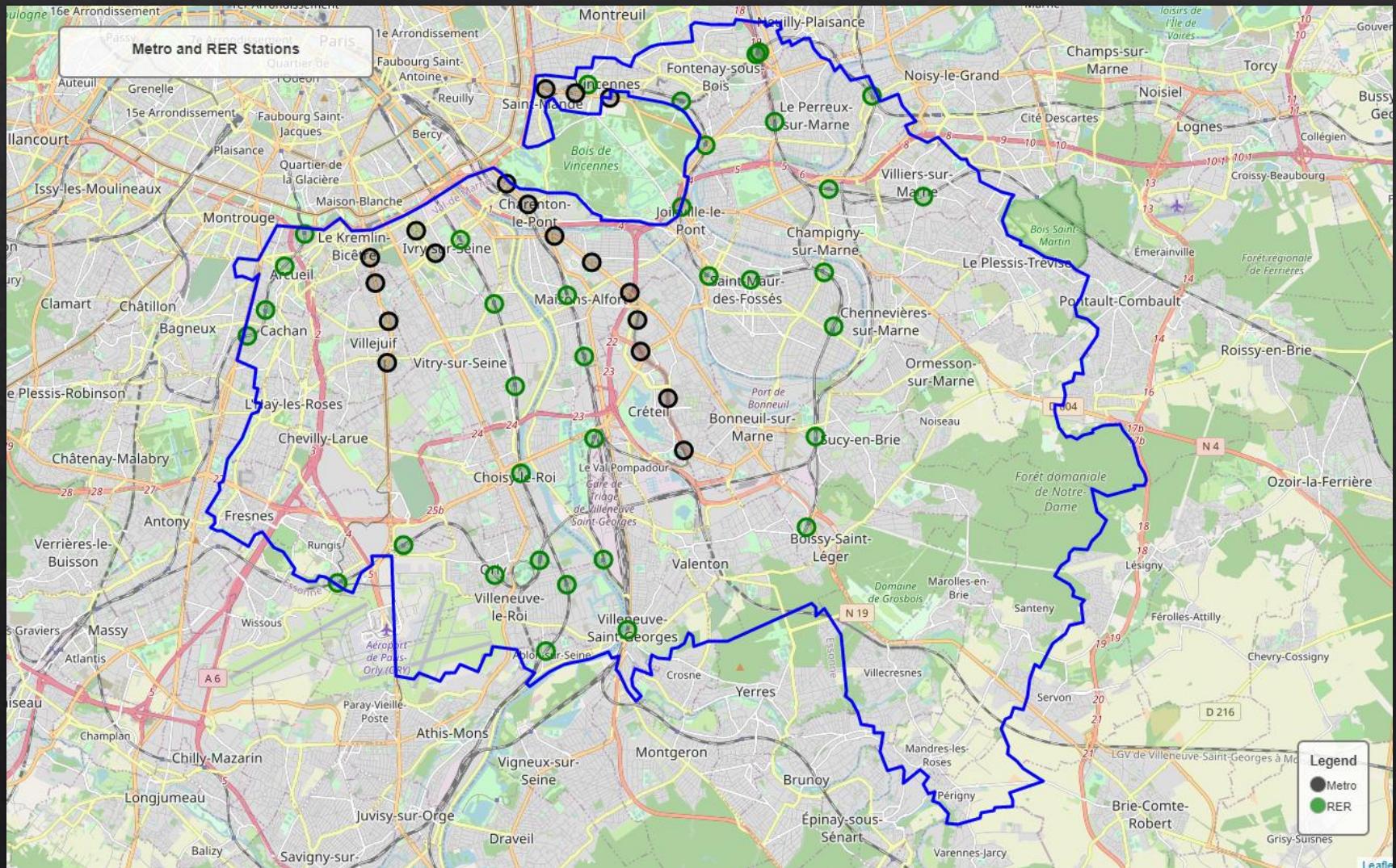
# Result Section

In this section are presented the different maps and graphs resulting from this Analysis

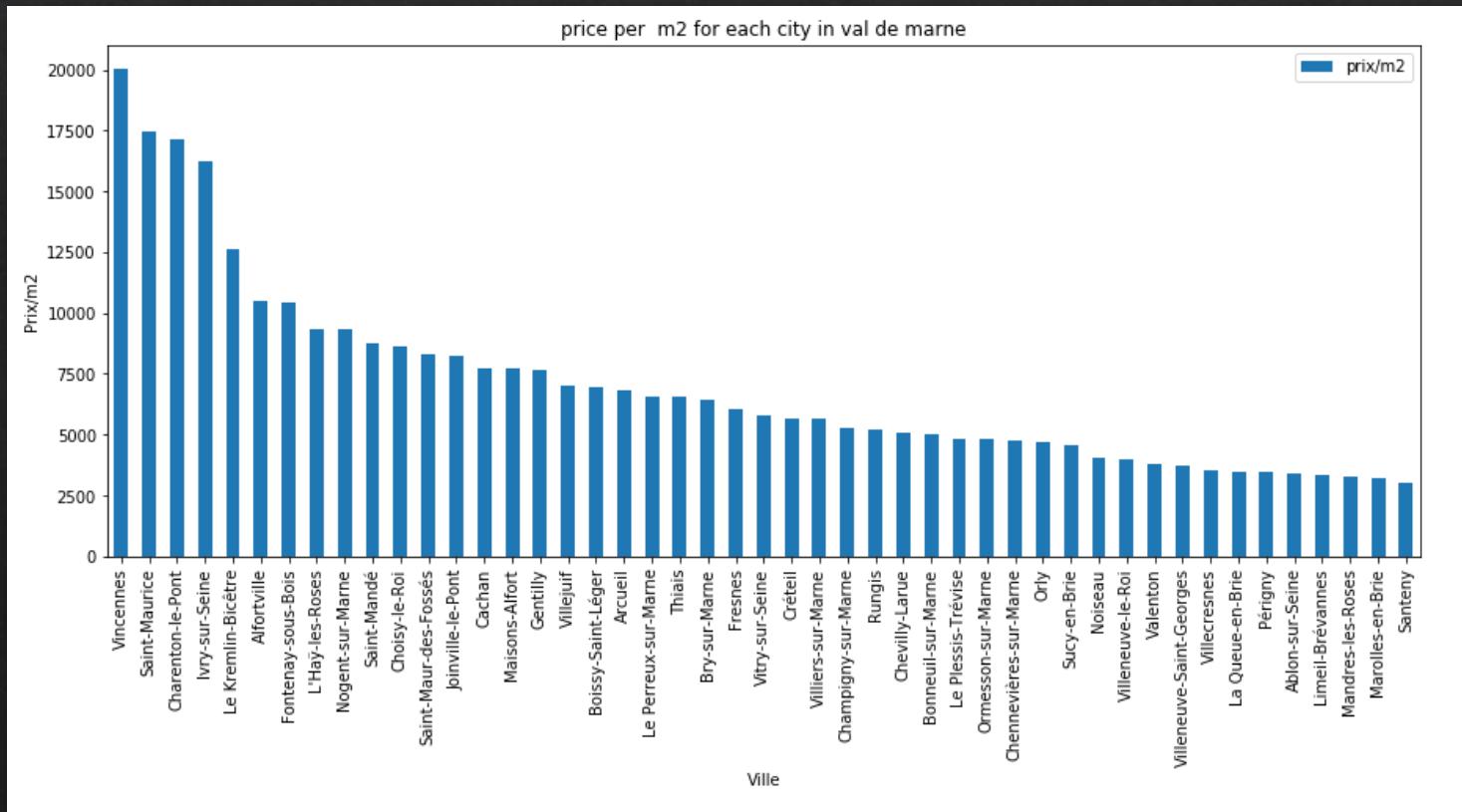
# Population Density in Val de Marne



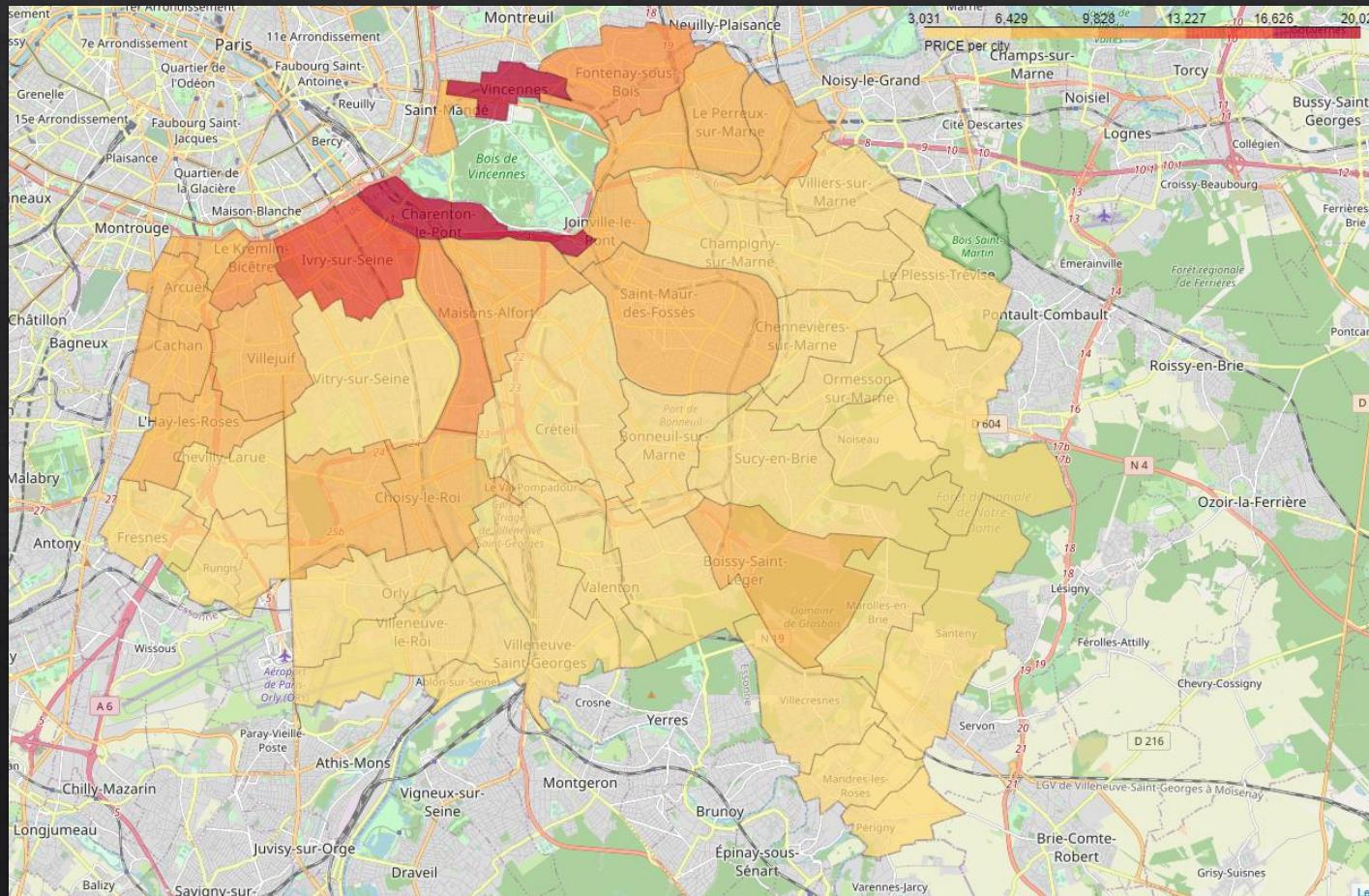
# Metro and RER Stations



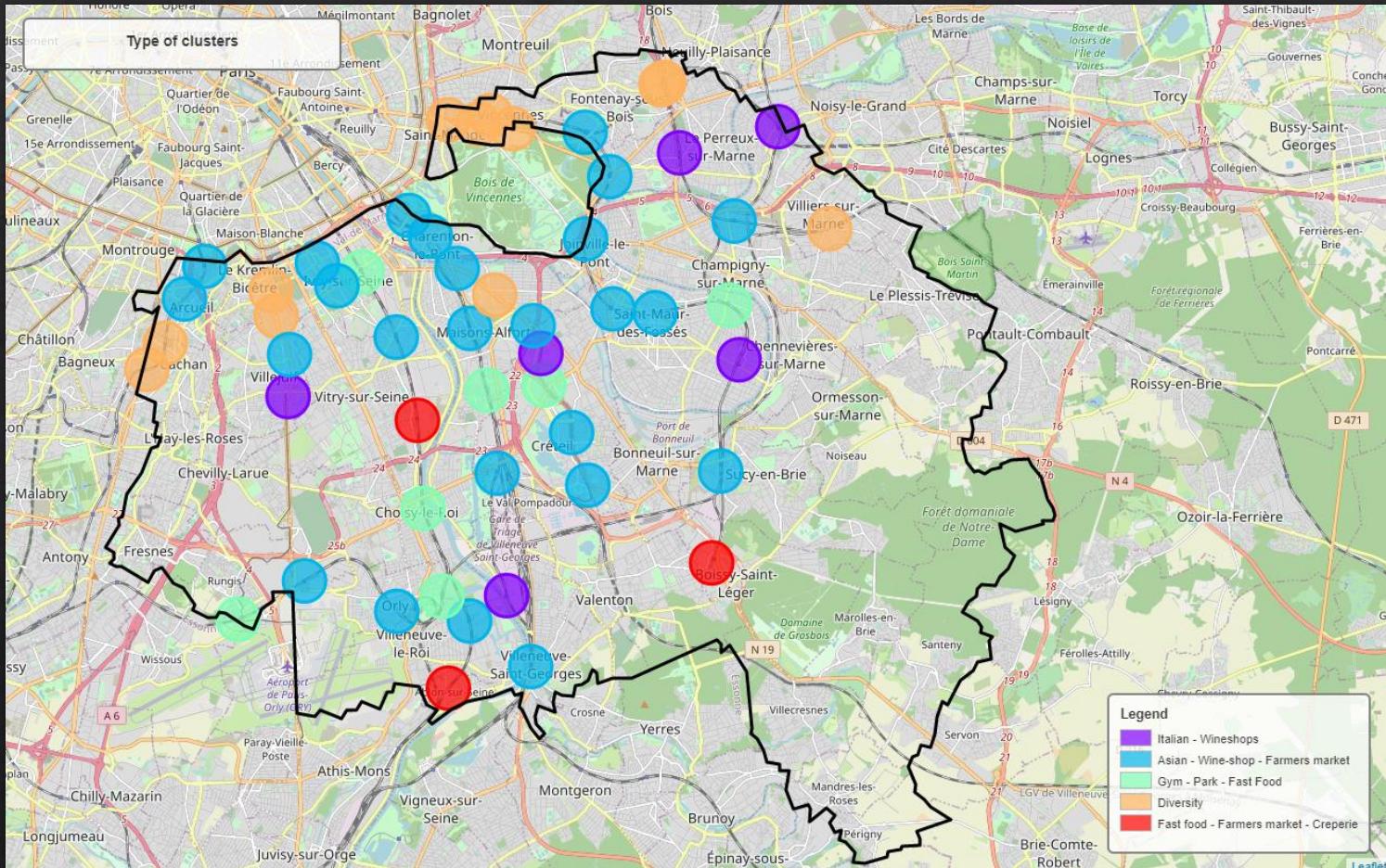
# Price per m<sup>2</sup> for each city in Val de Marne



# Price per city



# Type of Clusters



This analysis  
allows us to  
give a type to  
each cluster

- Cluster 1 Italian – Wineshops
- Cluster 2 Asian, Wine-shop, Farmers market
- Cluster 3 Gym - Park - Fast Food
- Cluster 4 Diversity
- Cluster 5 **Fast food - Farmers market - Creperie**

# Discussion

- ❖ To make this study meaningful, we had to restrict the categories presented by foursquare. All of these neighborhoods essentially have bakeries, cafes and supermarkets. I needed to eliminate them. Segmentation is also quite difficult to understand and relate to on field reality.
- ❖ In other studies it would be possible to study different segmentation to refine the study.
- ❖ On the technique, in addition to the courses and the handling of pandas, I learned for example to manage bases a little different from the bases of the course with ; separators, to separate XY coordinates, to use a geojson polygon to geographically separate data with shapely , or to add legends to folium maps.

# Conclusion

- ❖ This study provides a better understanding of the real estate segmentation in the department we have chosen.
- ❖ The methods applied here can be used for other French departments with relatively few adaptations.