

2020

# Toulouse demographic clustering



simone coniglio

23/03/2020

## Introduction

The goal of this project is to get some insight about the town I am living in: Toulouse (FR). In the very next future I could be interested in buying or selling an apartment so I would like to have an insight that could help me finding the best location. The example of the introduction of this module will be completed with a dataset that is freely available on insee website (even if it is in French). I am the main stakeholder of this study but also other people living here could be interested. Clustering will be based on demographic metrics such as the population density, the fraction of poor households, household average surface and fraction of households built after 1990. Looking at these parameters I could find region that are interesting for the stakeholder to buy a new house or apartment.

## DATA

The data employed to solve this problem were downloaded from The National Institute of Economic Studies INSEE: [www.insee.fr](http://www.insee.fr)

They consist of a csv file about squares on French territory related to 2015 that were collected thanks to taxes declaration.

Due to privacy limitation the data are aggregated on square with a variable surface. The size of the square can be adapted to ensure that at least 11 households are inside the square.

A full documentation about the attributes related to each square can be found here:

<https://www.insee.fr/fr/statistiques/4176290?sommaire=4176305>

Most important attributes are described here:

*Id\_carr\_n*: is a unique identification id related to each square, it gives the coordinates of the bottom left corner of the square and its size

*Ind*: gives the number of individuals living in the square

*t\_maille*: gives the size of the square in meters

*Men\_pauv*: gives the number of poor households in the square considering a Winsor's threshold

*Log\_ap90*: gives the number of households built after 1990

*Men\_surf*: gives the sum of households surface in the square

*Men*: gives the number of households in the square

Many other metrics were available, but were not used.

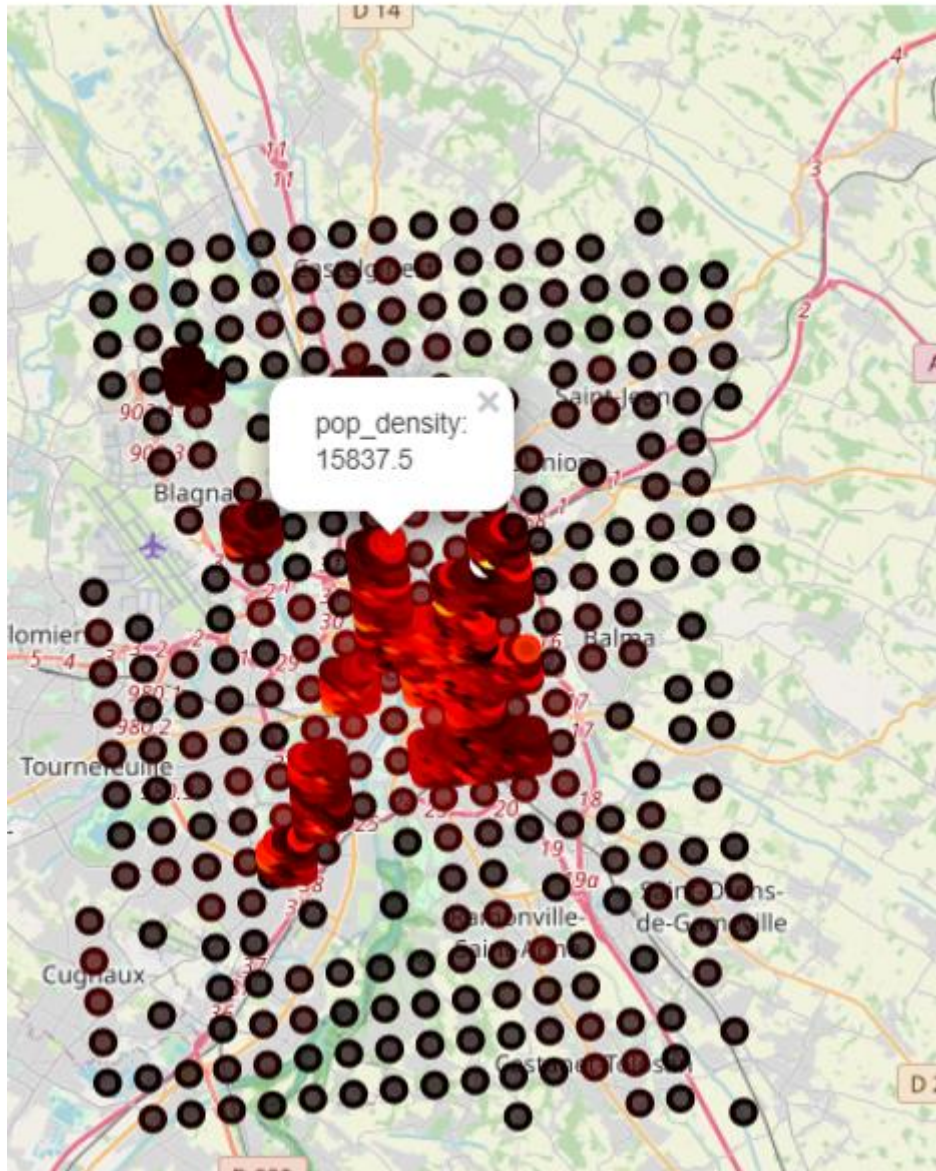
## Methodology

Since the square have different sizes some preprocessing were considered:

The population density defined as the number of individuals living inside a square per squared kilometer was computed as:

$$pop_{density} = \frac{Ind}{t_{maille}^2} \times 10^6 [Inhabitants \text{ per } Km^2]$$

## pop\_density



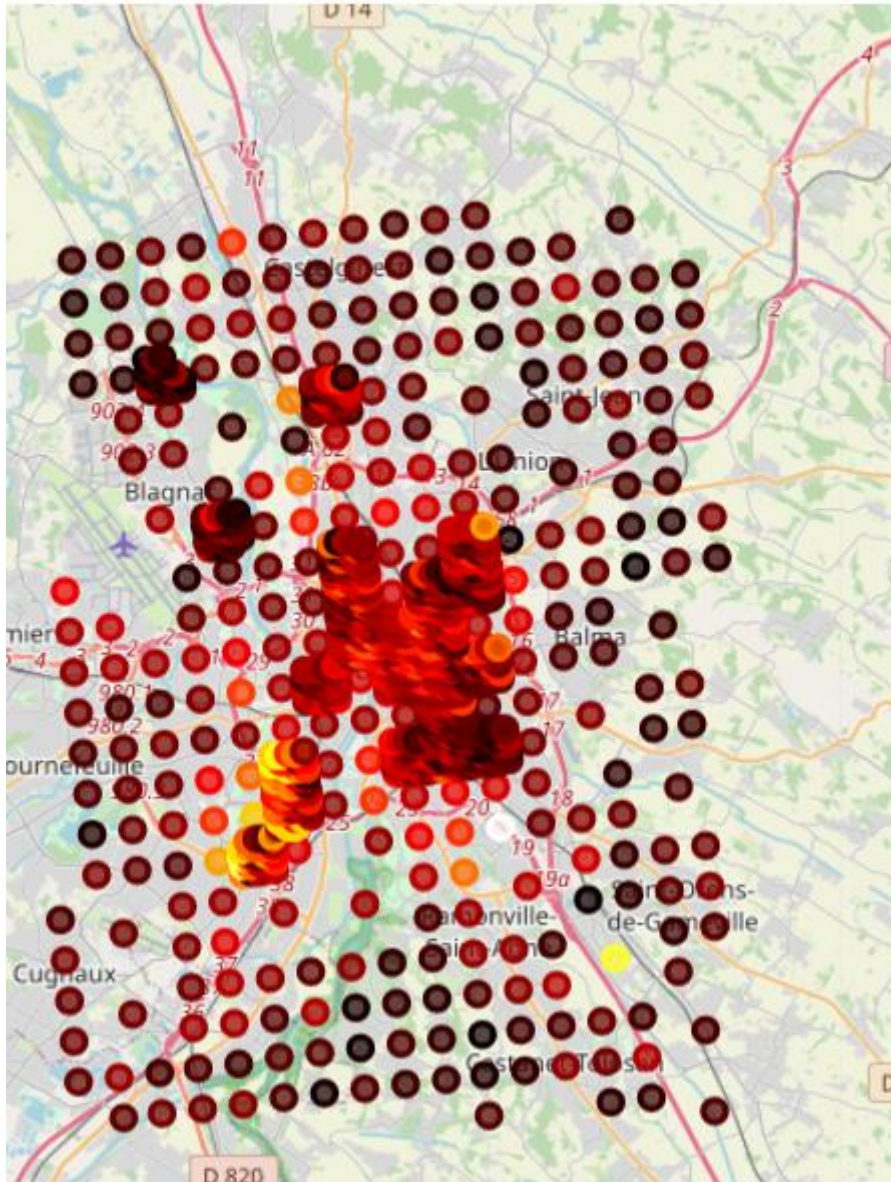
This parameter is important for me since a highly populated square is for me an important parameter for the choice where you are going to invest for your own household.

The ratio of poor households was computed:

$$part_{poor} = \frac{Men_{pauv}}{Men}$$



part\_poor

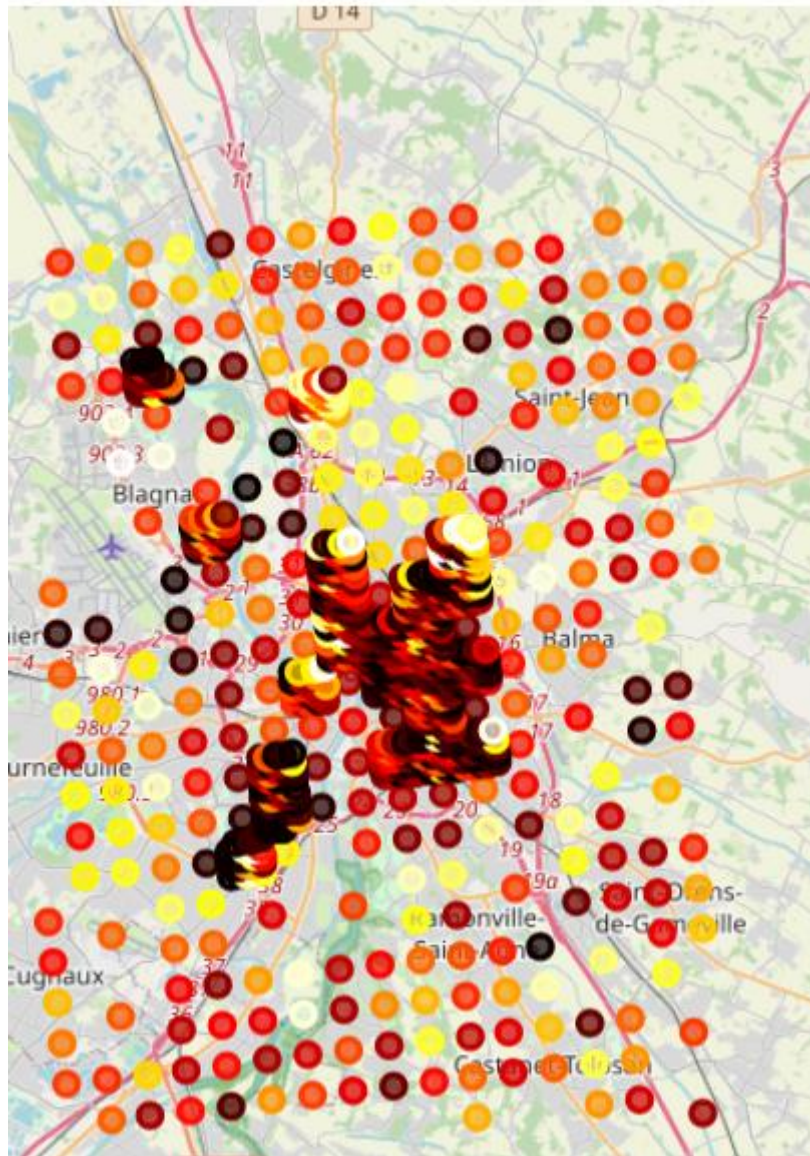


This parameter can give an idea also of the kind of population leaving in a square can also be correlated with the age of the population.

The ratio of new hoseholds as:

$$part_{new} = \frac{Log_{ap90}}{Men}$$

part\_new



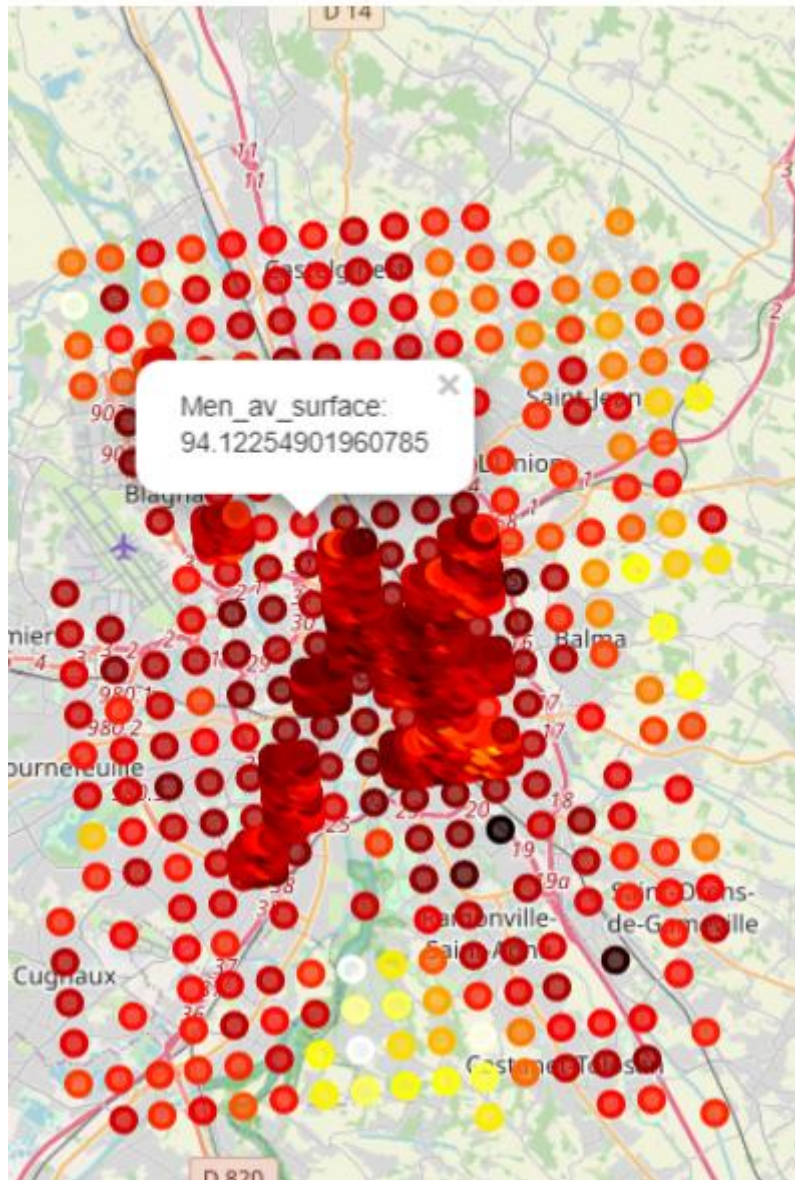
This parameter is important to give an idea of the age of the building inside the square.

The average surface of households:

$$Men_{avsurf} = \frac{Men_{surf}}{Men}$$



## Men\_av\_surface



That is of course important to know when looking for an apartment or a house.

Now that we have these feature four the square around Toulouse, we want to generate a clustering of squares based on such parameters.

The reason of that I am at the beginning of my research and I do not know all neighbors of Toulouse. To limit my research, I want to find out the kind of neighbors that one can find in this beautiful city to then select the places where I could be interested in finding a place.

K-mean clustering was adopted using 3 clusters. A scaling of data was considered before computing the clusters. The number of feature and of cluster in this methods was selected in order to have a consistent and simple explanation on the final results.

## Results:

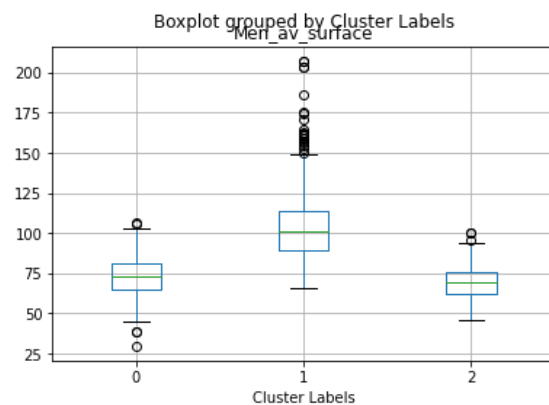
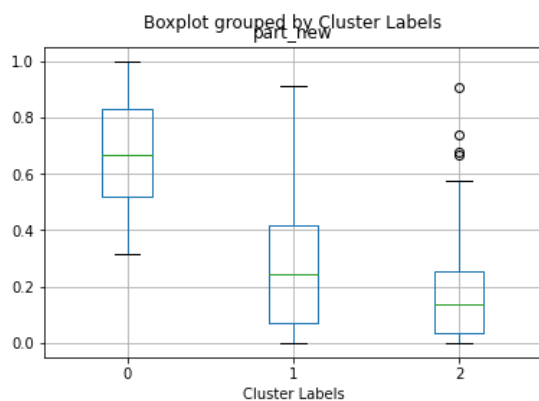
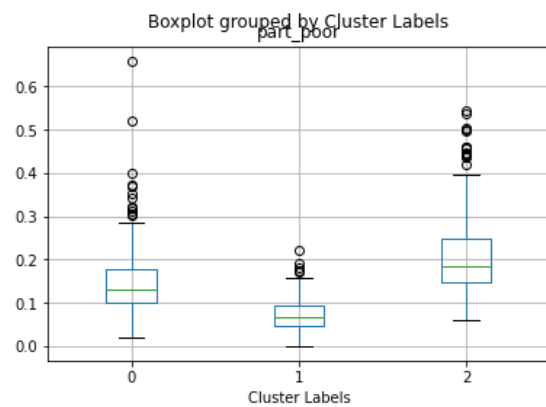
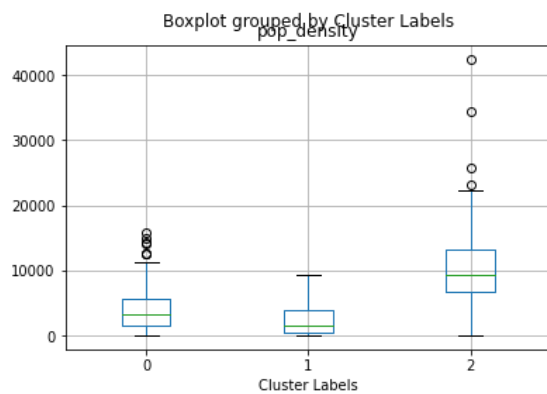
A nice way of getting an insight from the nature of each cluster is by looking at the statistical distribution of each one of its attributes.

Cluster 2 has a large population density and ratio of poor households, moreover this also have smaller and older households. Smallest incomes, older households, cheaper deals could be interesting for students, young couples, or modest family.

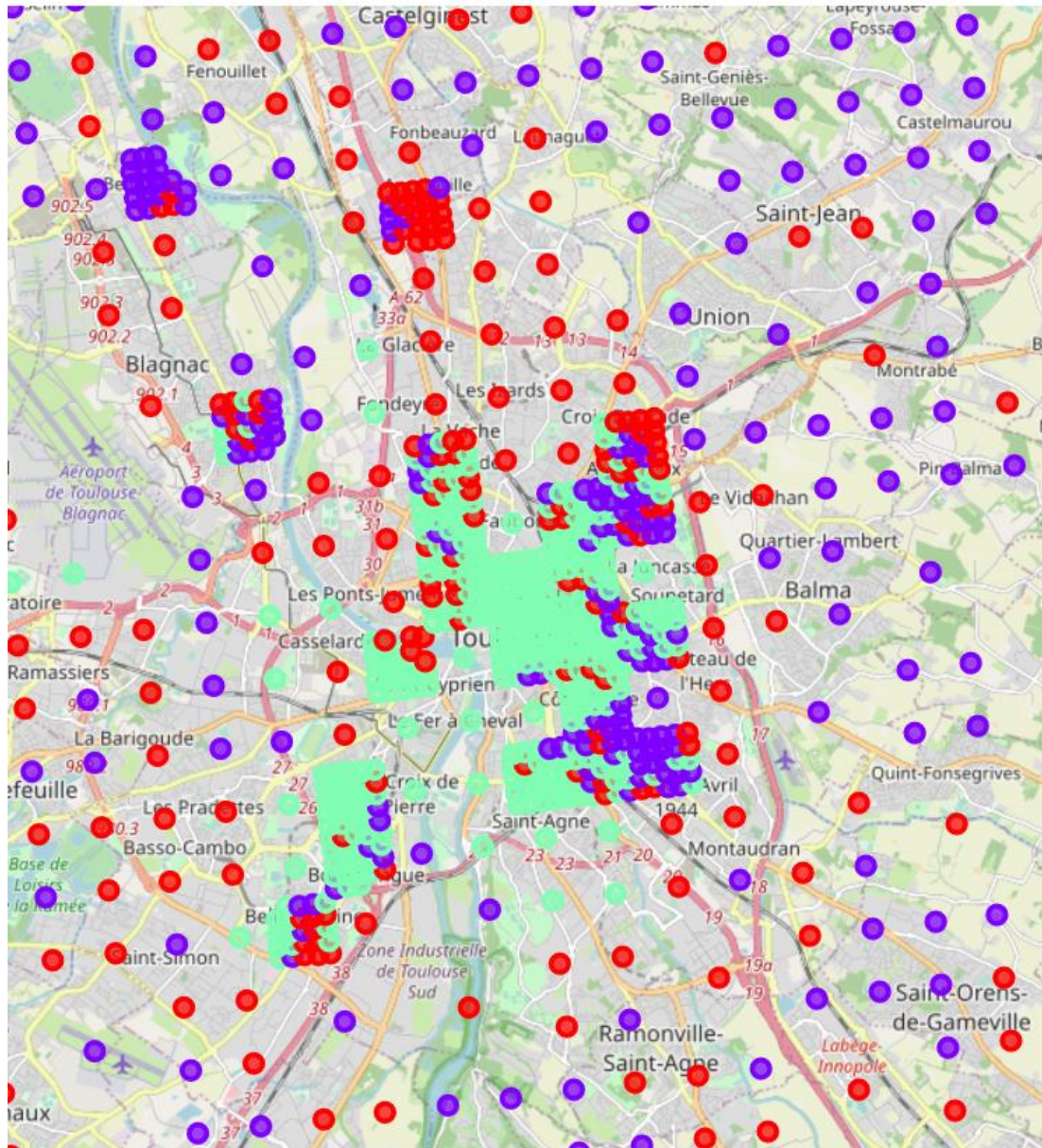
Cluster 1 on the other hand has the lowest population density and ratio of poor households, the largest average surface and middle a dispersion of new household ratio. Large house in residential region, could be interesting for rich, old couple with children.

Cluster 0 middle population, higher ratio of newer household, a large dispersion or ratio of poor households which means that the population is still varied in this kind of neighborhood a smaller population density and a middle surface.

That could be a quite good deal for my case!



## Toulouse Demographic Clusters



### Discussion

Giving a look at the distribution of clusters in the city we can see that most squares in cluster 2 are near the city center.

The squares of cluster 1 are in the south that is interesting and far from the city center.

The squares from cluster 0 are spread mostly far from the city center.

### Conclusion

It is possible to group Toulouse squares in 3 cluster based on relevant demographic indices. I hope you enjoyed this report.