

Toulouse demographic clustering

Simone Coniglio

Coursera Data Science Capstone project

Why demographic clustering?



To learn about unknown area where we could be interested in open a new business or buy a household for living.



To give a new insight on a city you think to know

Why Toulouse?



- ▶ Because it's where I am living!
- ▶ I am the main stakeholder after all!

Data collection



Data employed in this study were downloaded from : www.insee.fr



Data are related to squares all over the French territory



Attributes are presented in this documentation:
<https://www.insee.fr/fr/statistiques/4176290?sommaire=4176305>

Most relevant attributes

- ▶ *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



Square center coordinates where computed from the square ID name



A projection was applied to compute longitude and latitude of each square



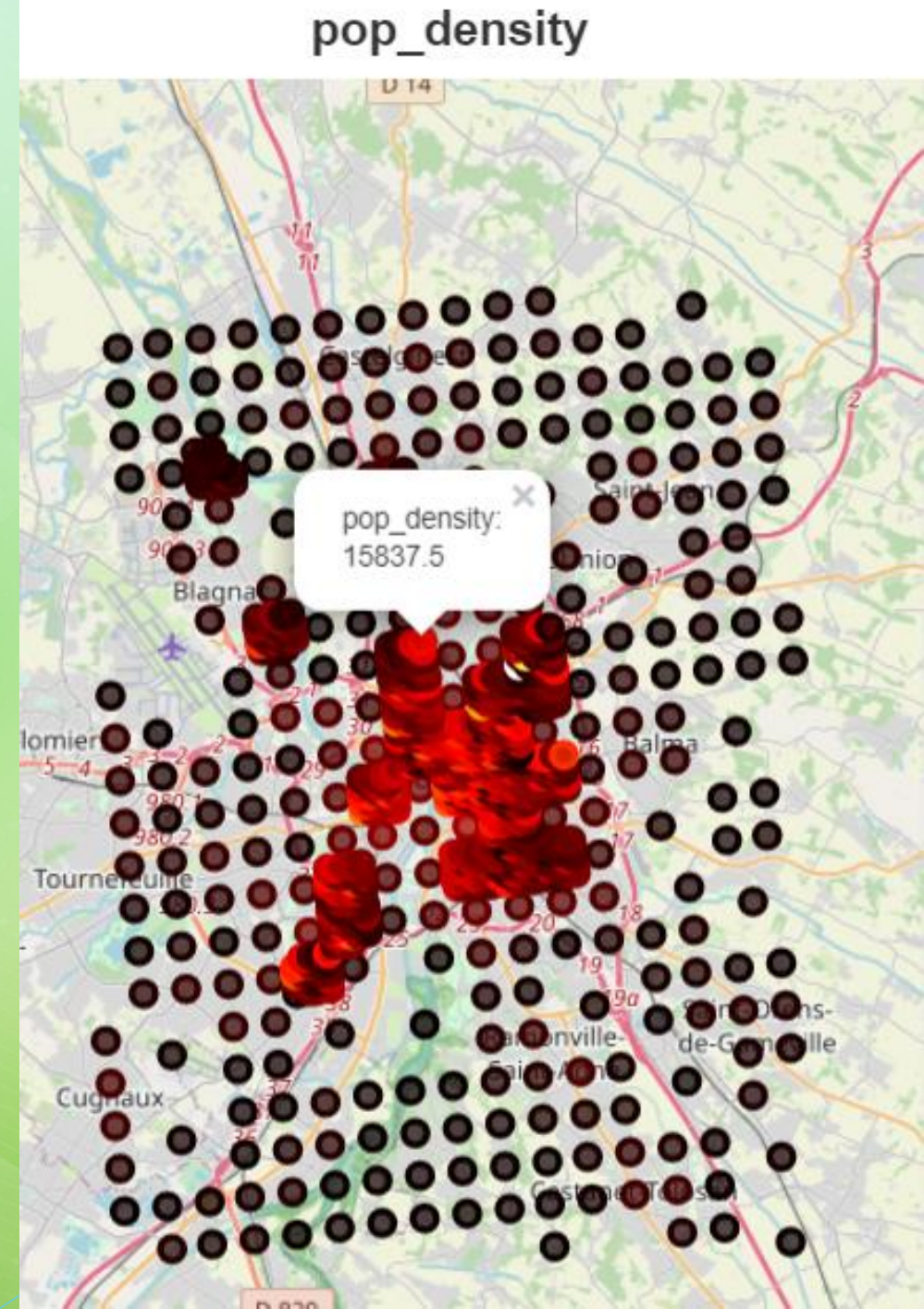
Squares in the nearby of Toulouse where selected for the computation of other interesting metrics



Colors are used to represent each metric. A hotter color represent a higher value of the presented metric

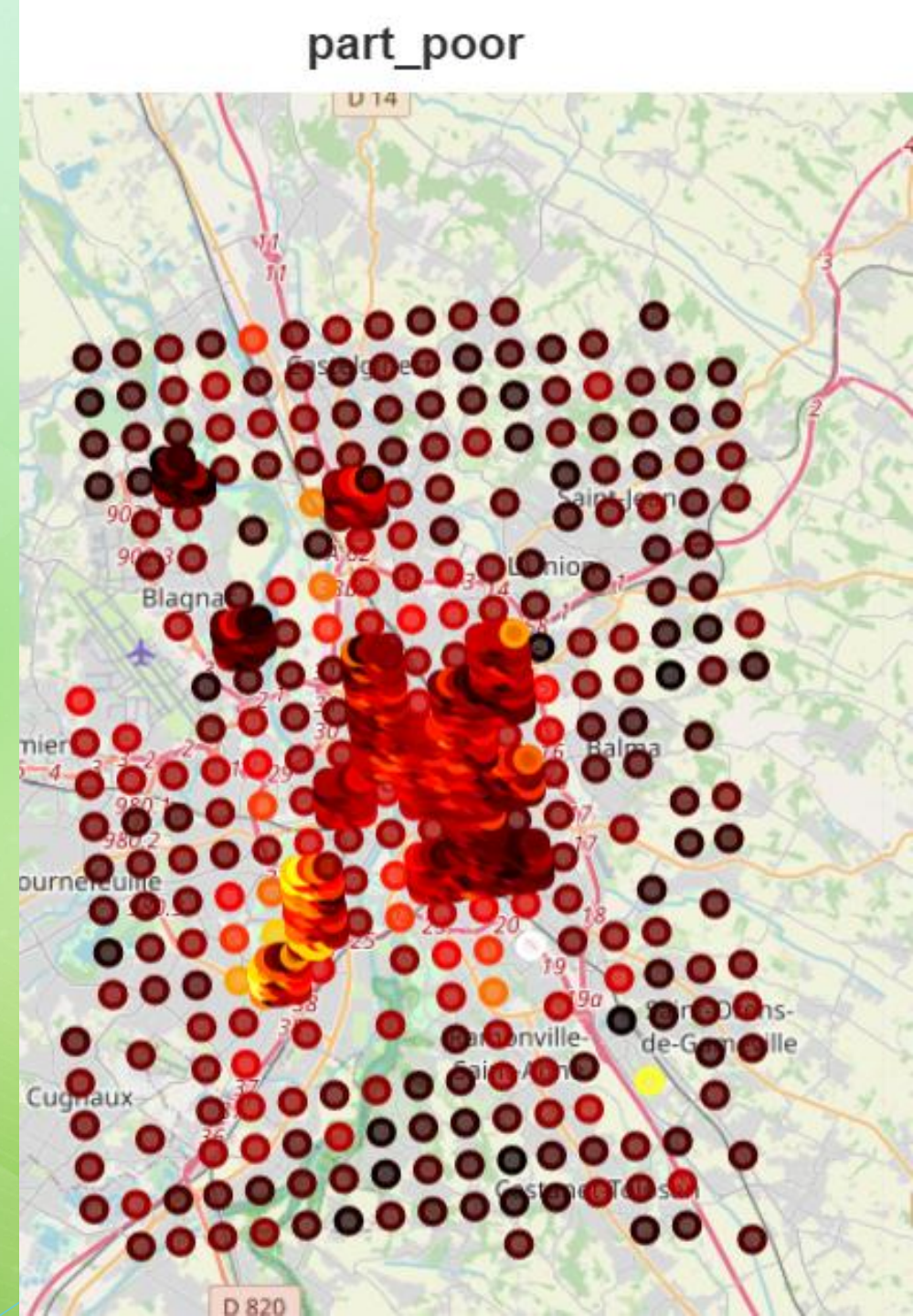
Methodology Part-2

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



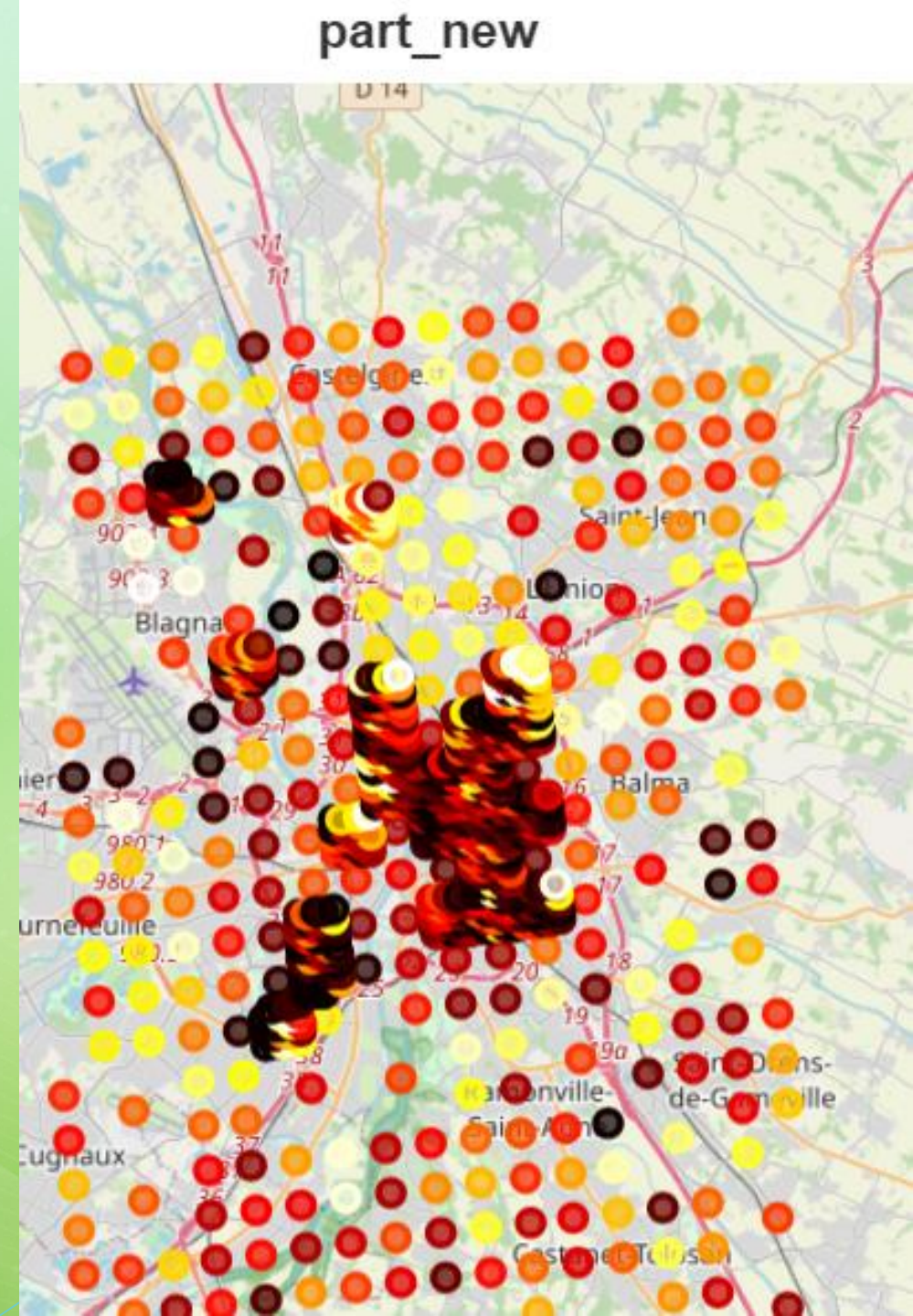
Methodology Part-3

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



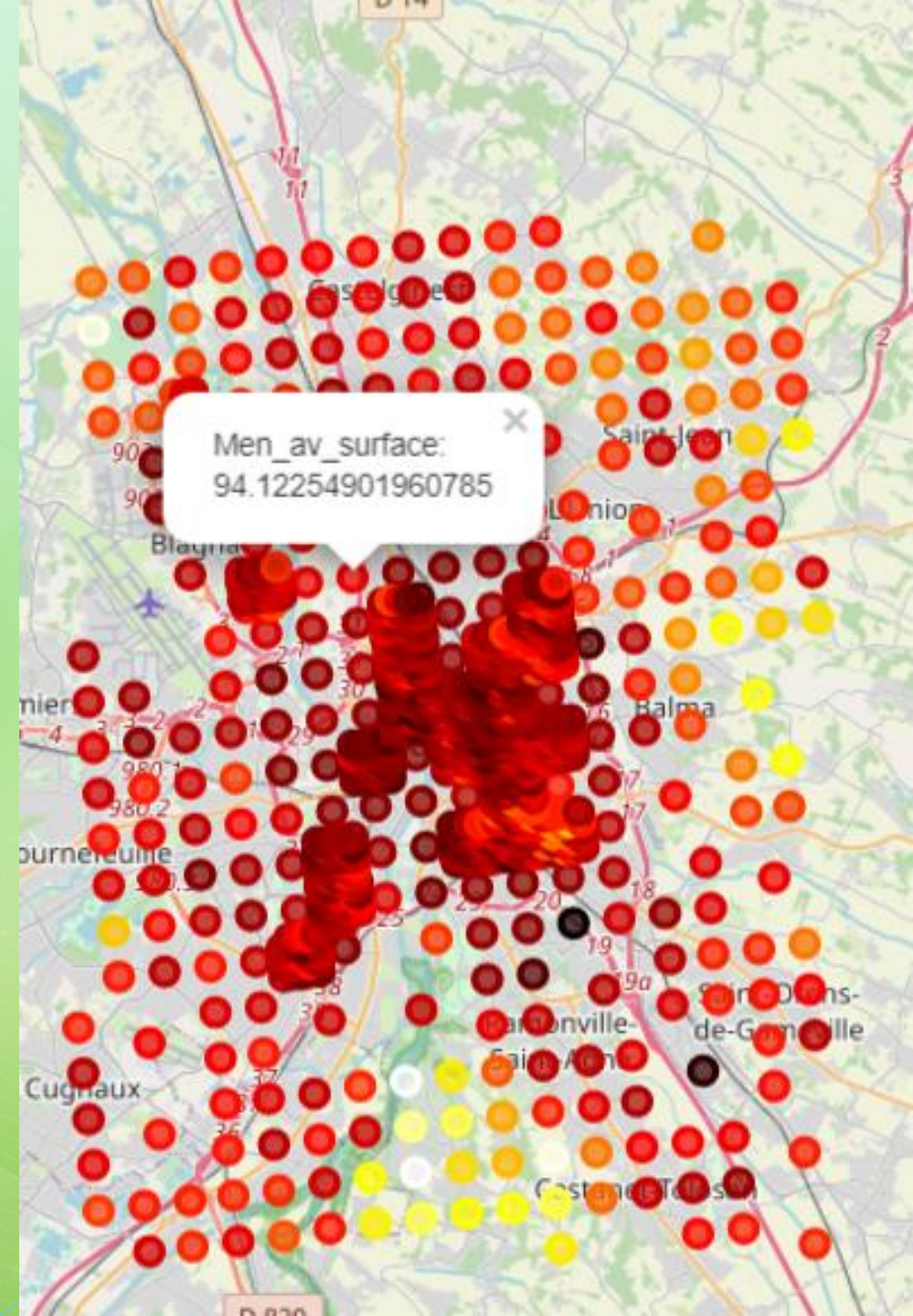
Methodology Part-4

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

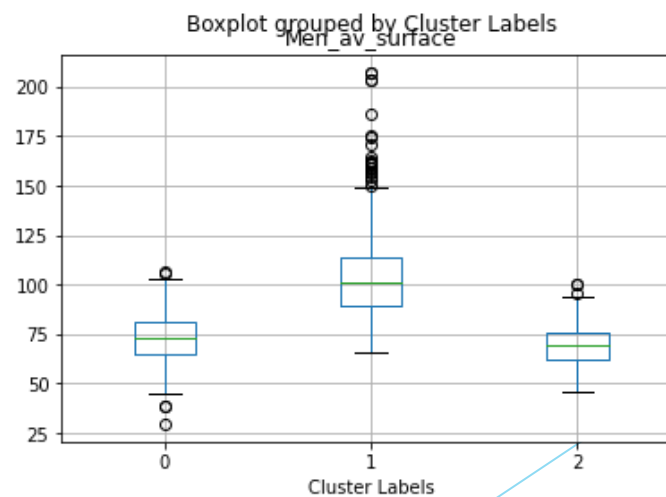
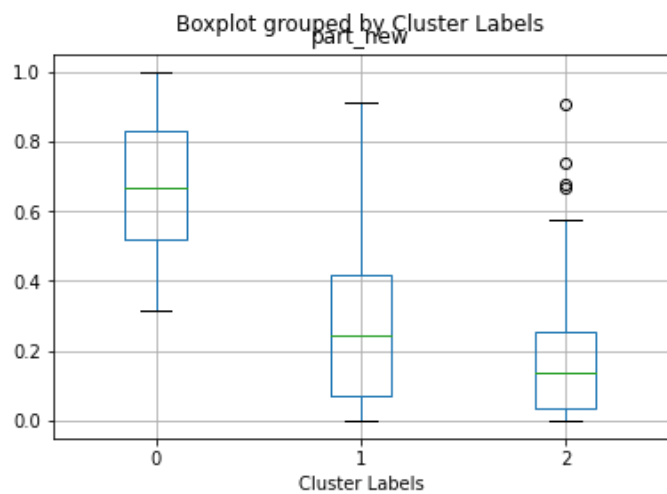
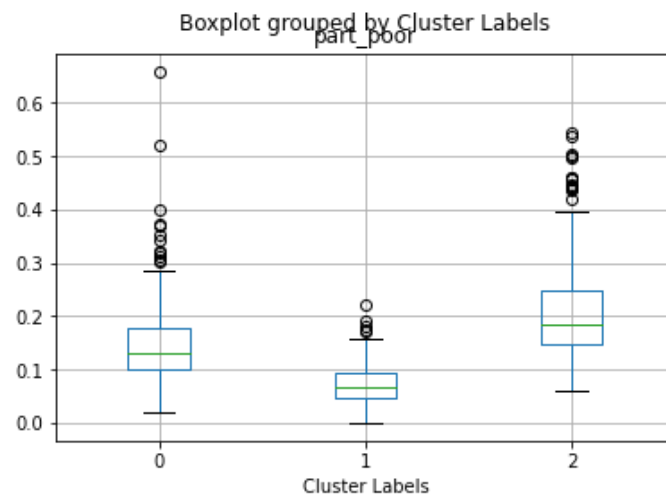
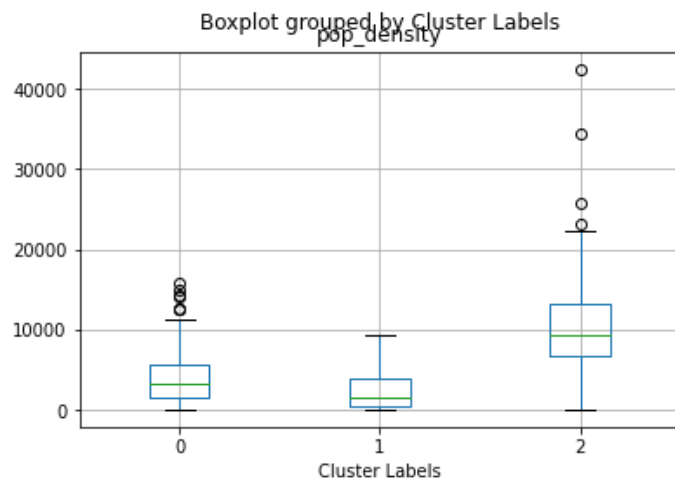


Methodology Part-5

► $Men_{av_{surf}} = \frac{Men_{surf}}{Men} [m^2]$



Nature of clusters



Cluster distribution

Conclusions

- ▶ Toulouse squares can be divided in 3 clusters:
 - ▶ 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.