

Establishment of a coffee shop in Paris

Problem exposition

- Coffee places are staples of French culture and tourism, lots of French people are used to go a coffee place to take breakfast, enjoy a cup of coffee while reading newspaper or even work. Lots of tourists in Paris also go to visit coffee places.
- However, there is already an abundance of coffee places in Paris making it hard to enter the market.
- How can data help us in deciding where to establish a new coffee place?
- Which type of customers should we expect and target ?

Data acquisition

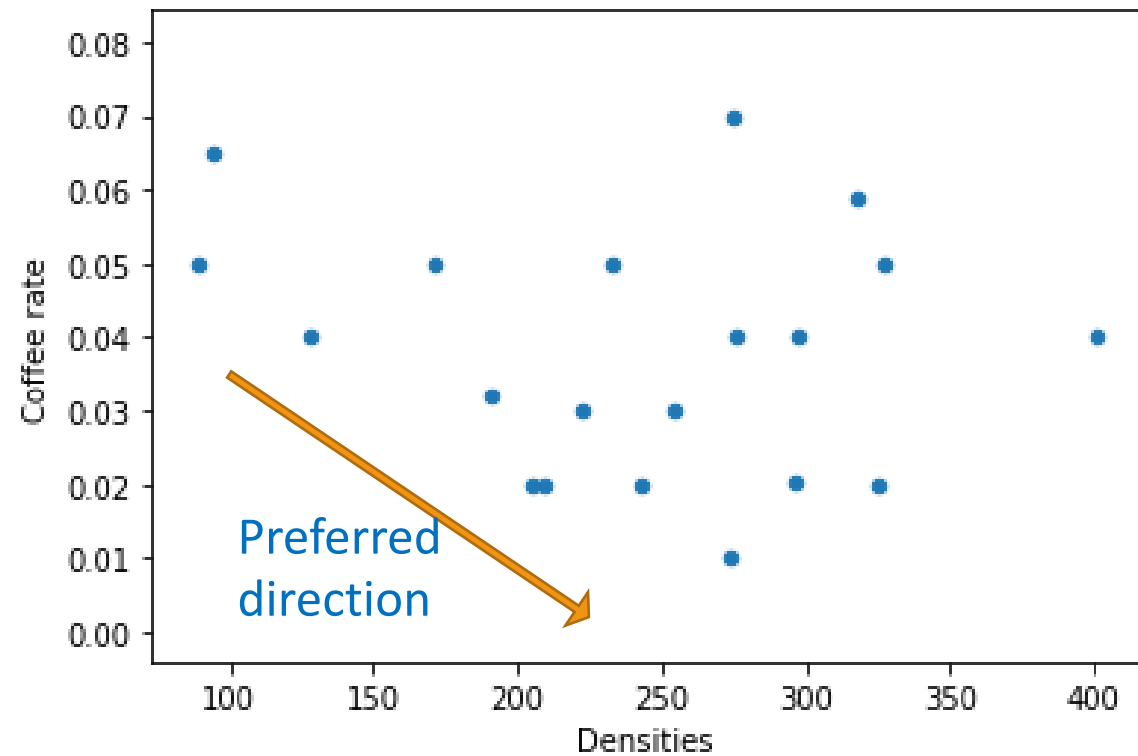
- Data extracted from INSEE databases (French national institute for statistical studies) for revenue and demographics
- Usage of Foursquare API for calculation of coffee shops presence rates

Features :

- Age bins for population and rate of population in each bin per district
- Median available revenue in each district
- Average density per district
- Coffee shops presence rate per district

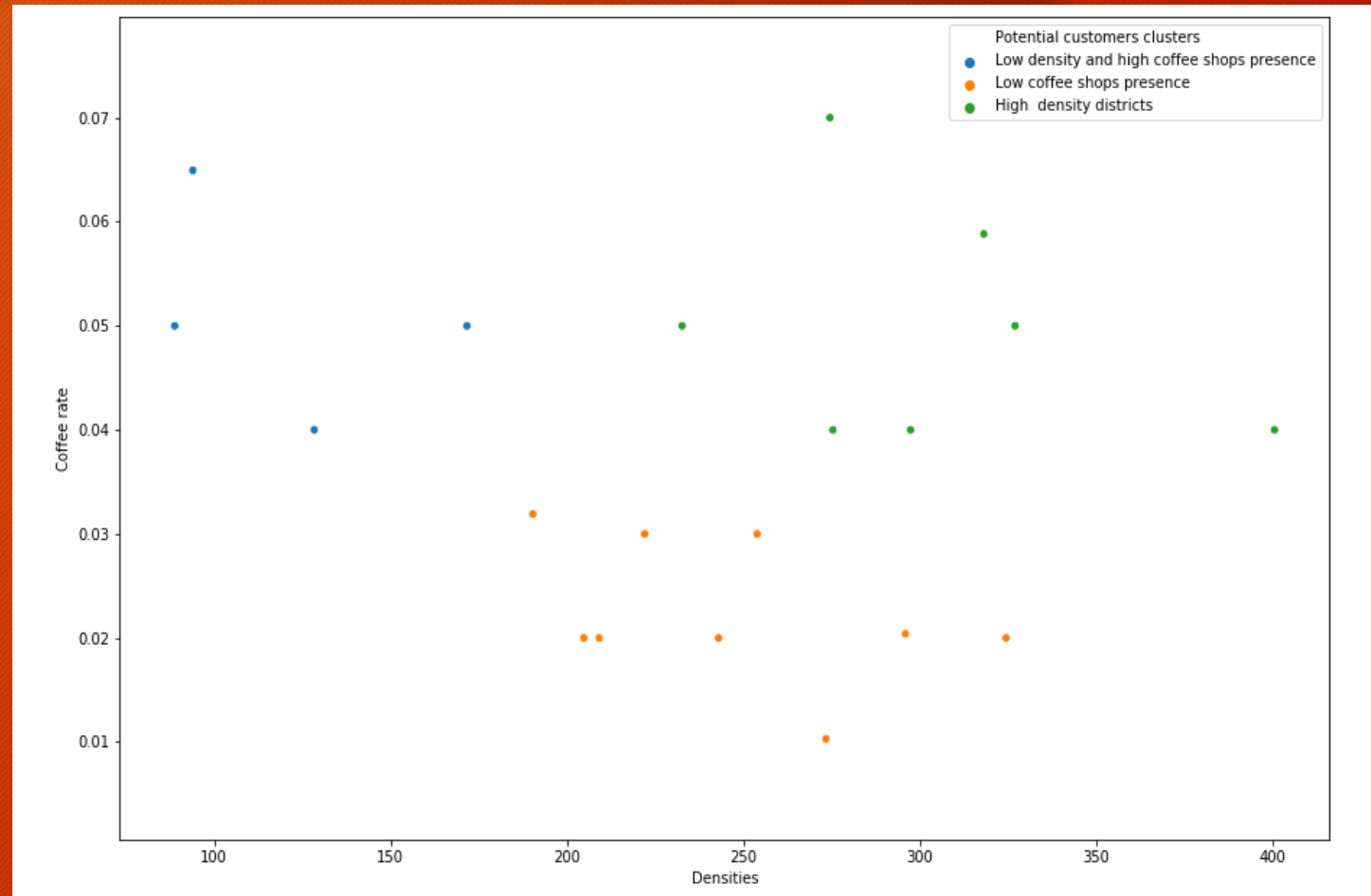
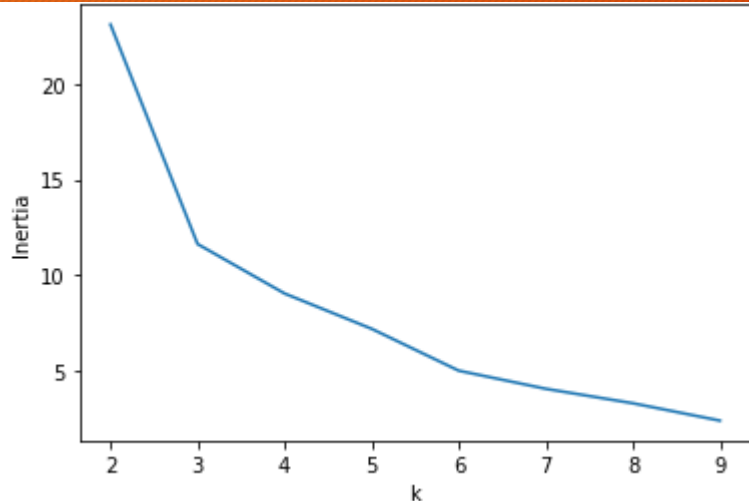
Best district selection based on population density and market saturation

- We would prefer to establish a new coffee place in a district with fewer coffee places, to have reduced competition
- We would also prefer a higher population density for a potentially larger effective customer base

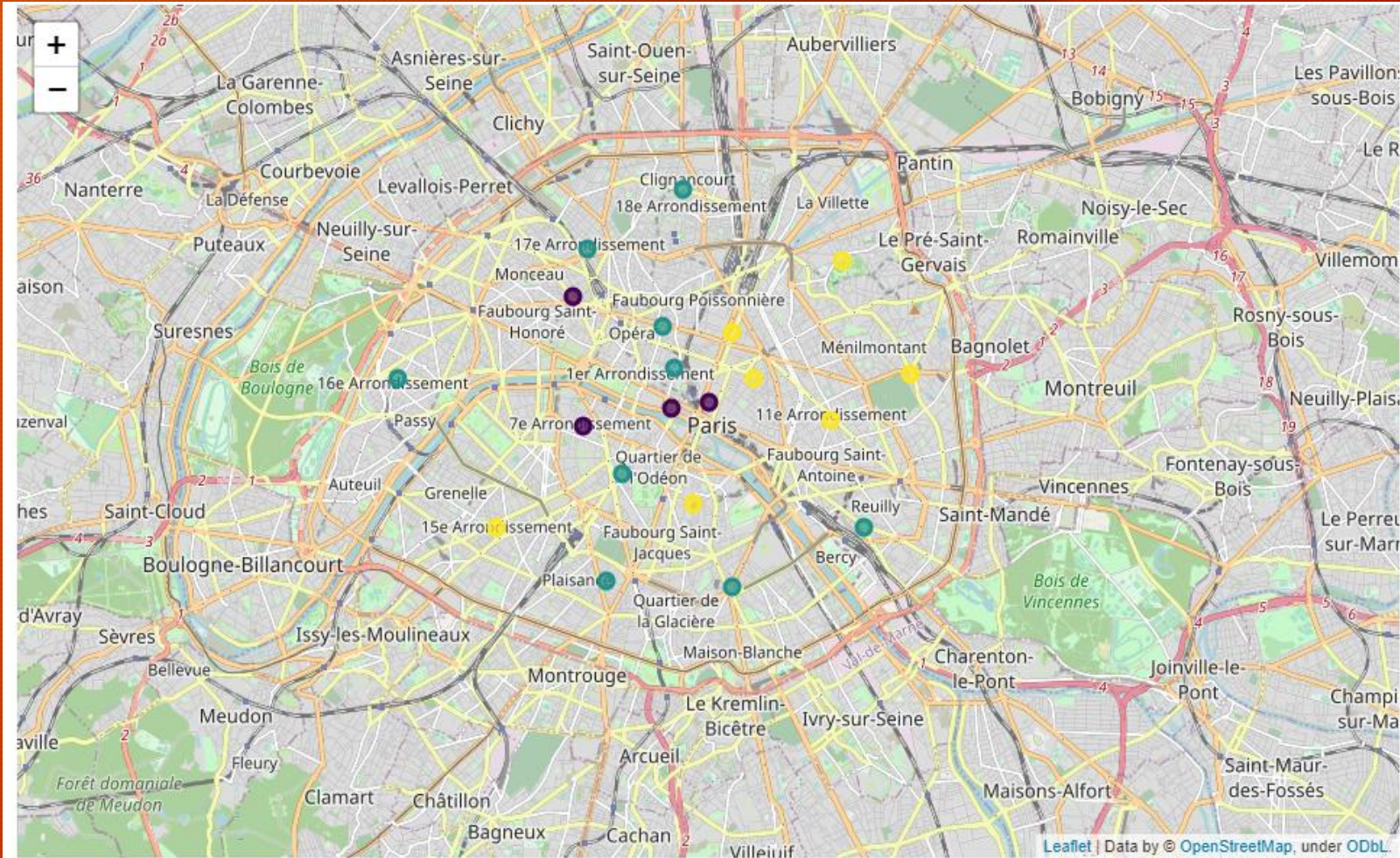


Best district selection based on population density and market saturation

- Kmeans clustering applied to scaled densities and coffee rates
- Inertia versus number of clusters shows that an optimal number of clusters could be three (elbow method)

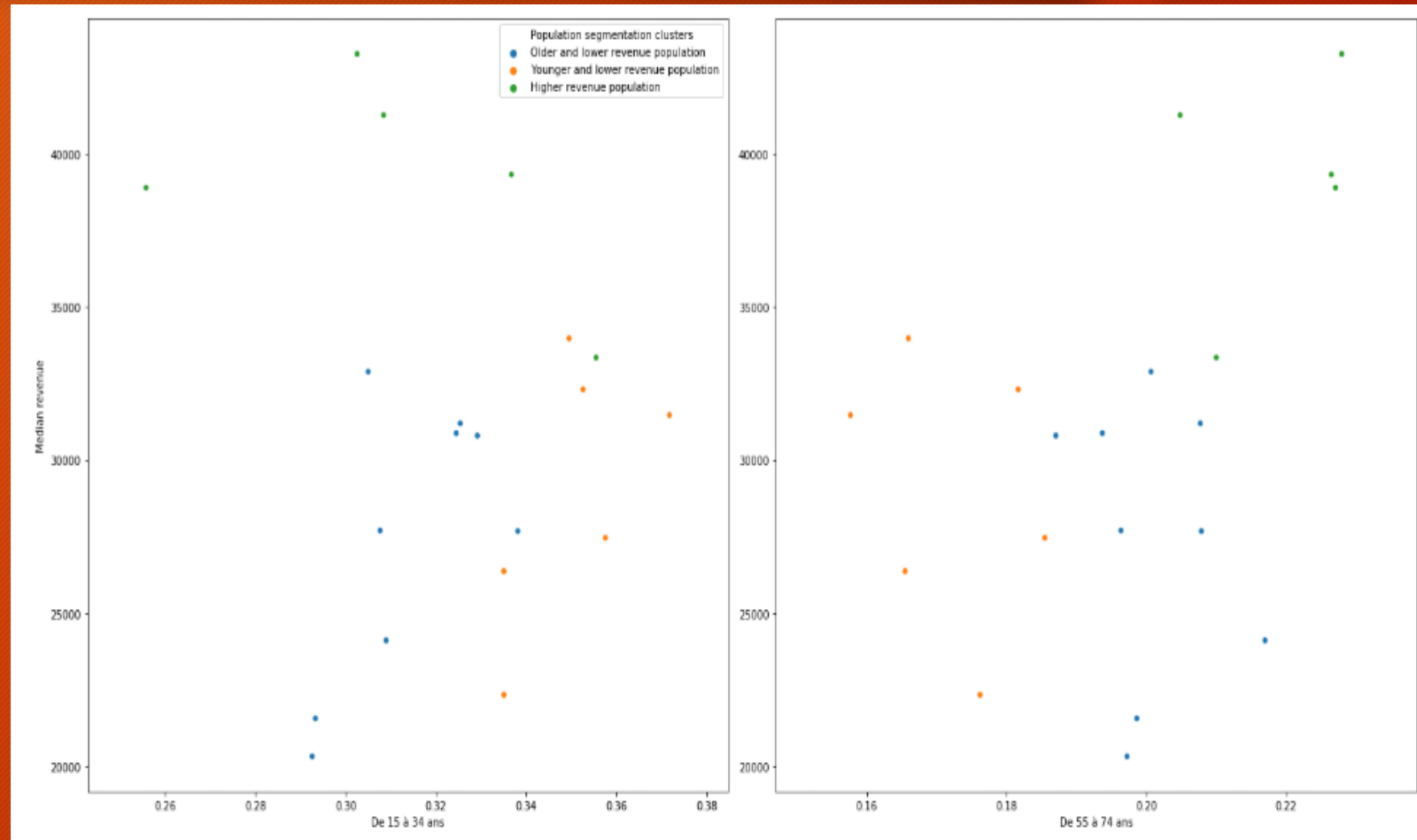
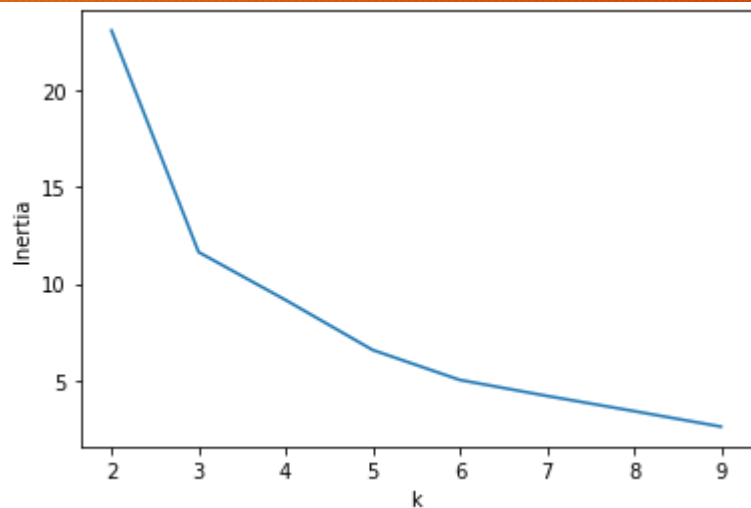


- High density districts
- Low coffee shops presence districts
- Low density and high coffee shops presence districts



District clustering based on population segmentation on age and revenue

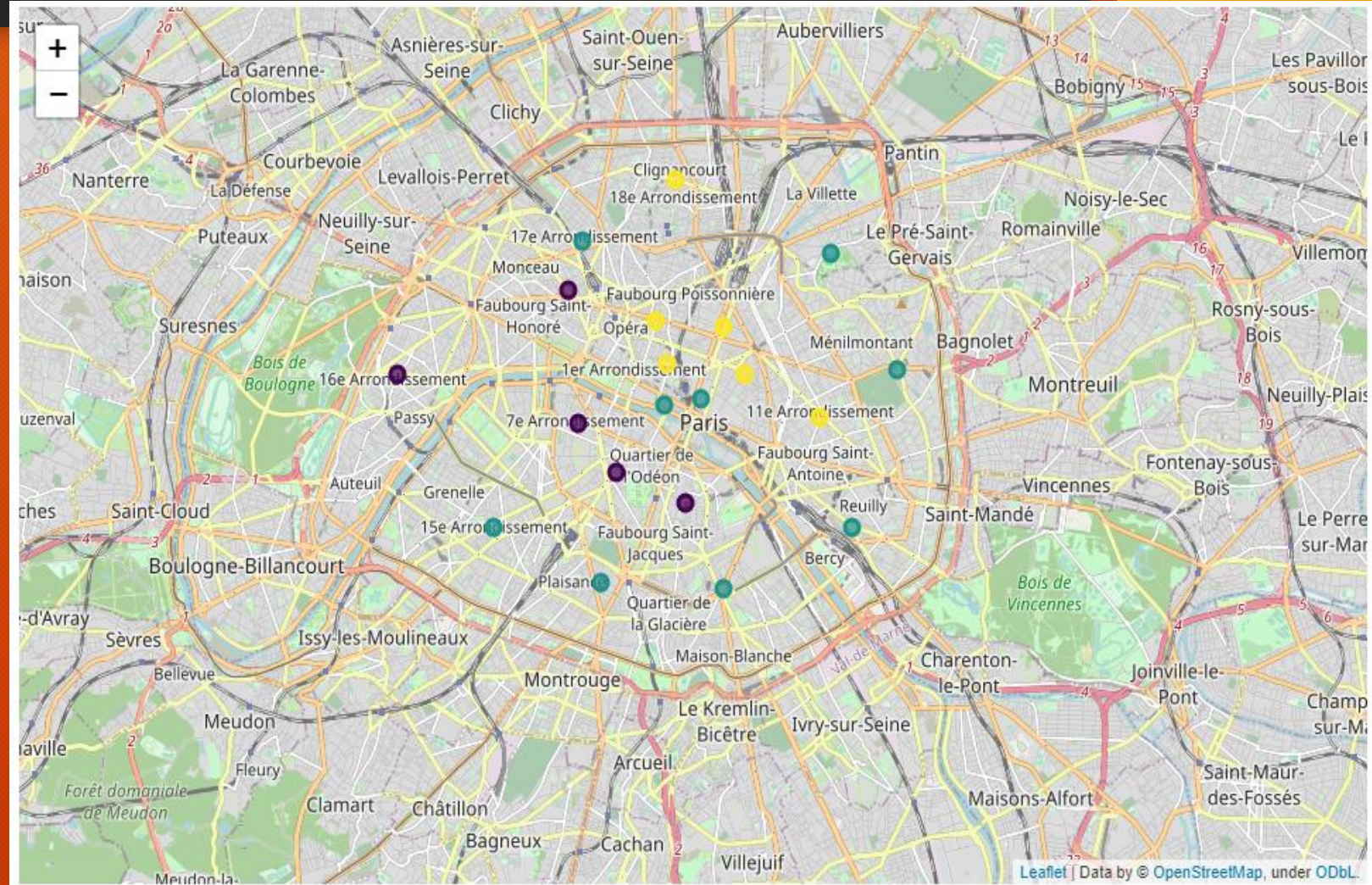
- Kmeans clustering applied to scaled rates of population in age bins and median revenue
- Inertia versus number of clusters shows that an optimal number of clusters could be three (elbow method)



District clustering based on population segmentation on age and revenue

- The three clusters help in defining a target audience, more « chics » coffee places in higher revenue zones or places aimed towards younger audience in other districts

Districts with a younger and lower revenue population ●
Districts with an older and lower revenue population ●
Districts with a higher revenue population ●



Conclusion

- These clusterings exercises help us define some interesting zones for establishment of a new coffee place, or district segmentation based on local population characteristics
- A further study should take more characteristics into account, such as tourism heatmaps