

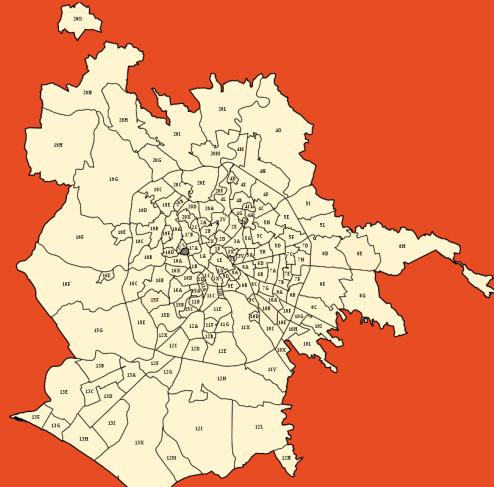
Katarzyna Anna Parys

Neighborhood Analysis for Real Estate Investing in Rome

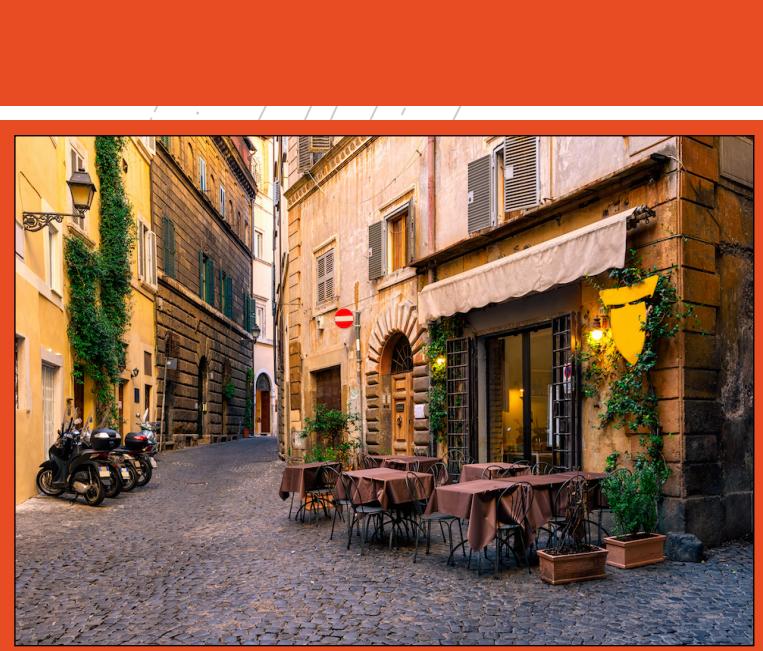
CapStone Project of IBM Data Science Certification



155 Neighborhoods

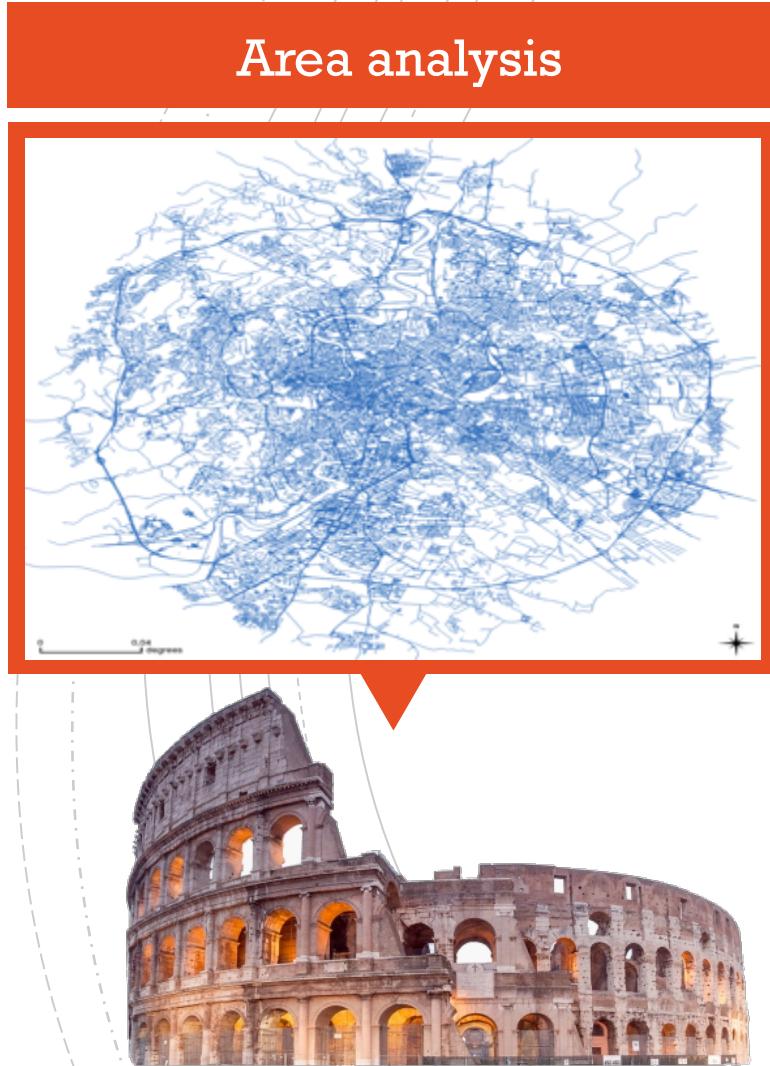


- The project is focused on the neighborhoods in the City of Rome, a territory with 2,822,981 residents, the third most populous city in Europe



- The prices of the properties for sale vary considerably across the neighborhoods, from 1,600 €/sqm in the Castelverde suburbs up to 9,100 €/sqm in Piazza di Spagna
- The distance from the city center isn't the only factor influencing property prices





- The goal of this study is to use open data about the neighborhoods to categorize them depending on a series of factors, such as density of population, traffic, percentage of foreigners, distance from the city center and the number of significant venues in a radius of 1 km from the center of the neighborhood

Supporting decisions



- Knowing which characteristics of a neighborhood affect property prices in a given area could be useful for all participants in real estate markets, especially investors, to understand and predict property price trends.
- In particular, the Department of the Treasury, which manages the public sector real estate in Italy, could benefit from the project

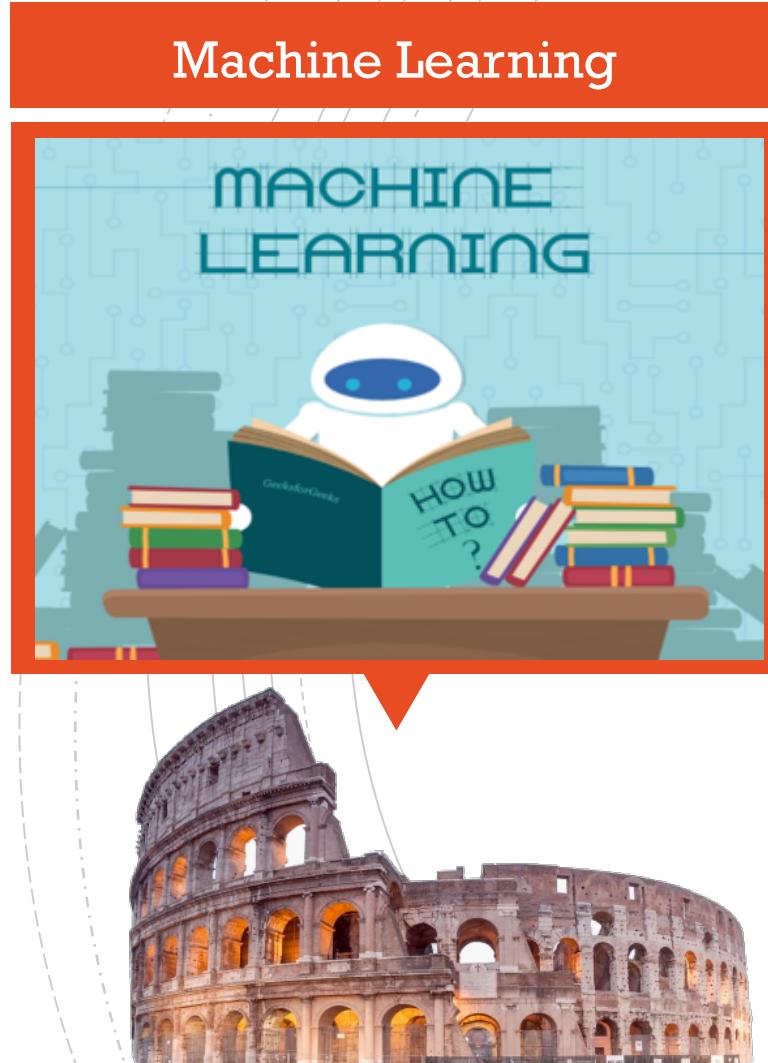


Data Sources

Data	Source	Last updated
House prices per square meter in Rome	https://www.immobiliare.it/mercato-immobiliare/lazio/roma/	Daily updates
List of the neighborhoods (urban zones)	https://www.comune.roma.it/web-resources/cms/documents/Elenco_Z_Urbanistiche_rg_A.pdf	2019
Area of the urban zones	https://www.comune.roma.it/web/it/roma-statistica-territorio.page	31.12.2019
Population of the urban zones	https://www.comune.roma.it/web/it/roma-statistica-popolazione1.page	31.12.2020
Foreign population of the urban zones	https://www.comune.roma.it/web/it/roma-statistica-popolazione1.page	31.12.2020
Top 100 venues in a 1000 meters range of the centre of each urban zone, in 10 categories	FourSquare API	Daily updates
Pollution levels (as a rough indicator of traffic)	http://www.arpalazio.net/main/aria/sci/basedati/bolletini/2021/BA192021.pdf	16.05.2021
List of the air monitoring stations in Rome with coordinates	http://dati.lazio.it/catalog/it/dataset/rete-di-monitoraggio-della-qualita-dell-aria/resource/0c9d32b8-06ed-4bb0-8727-9954c6d703f2	2020

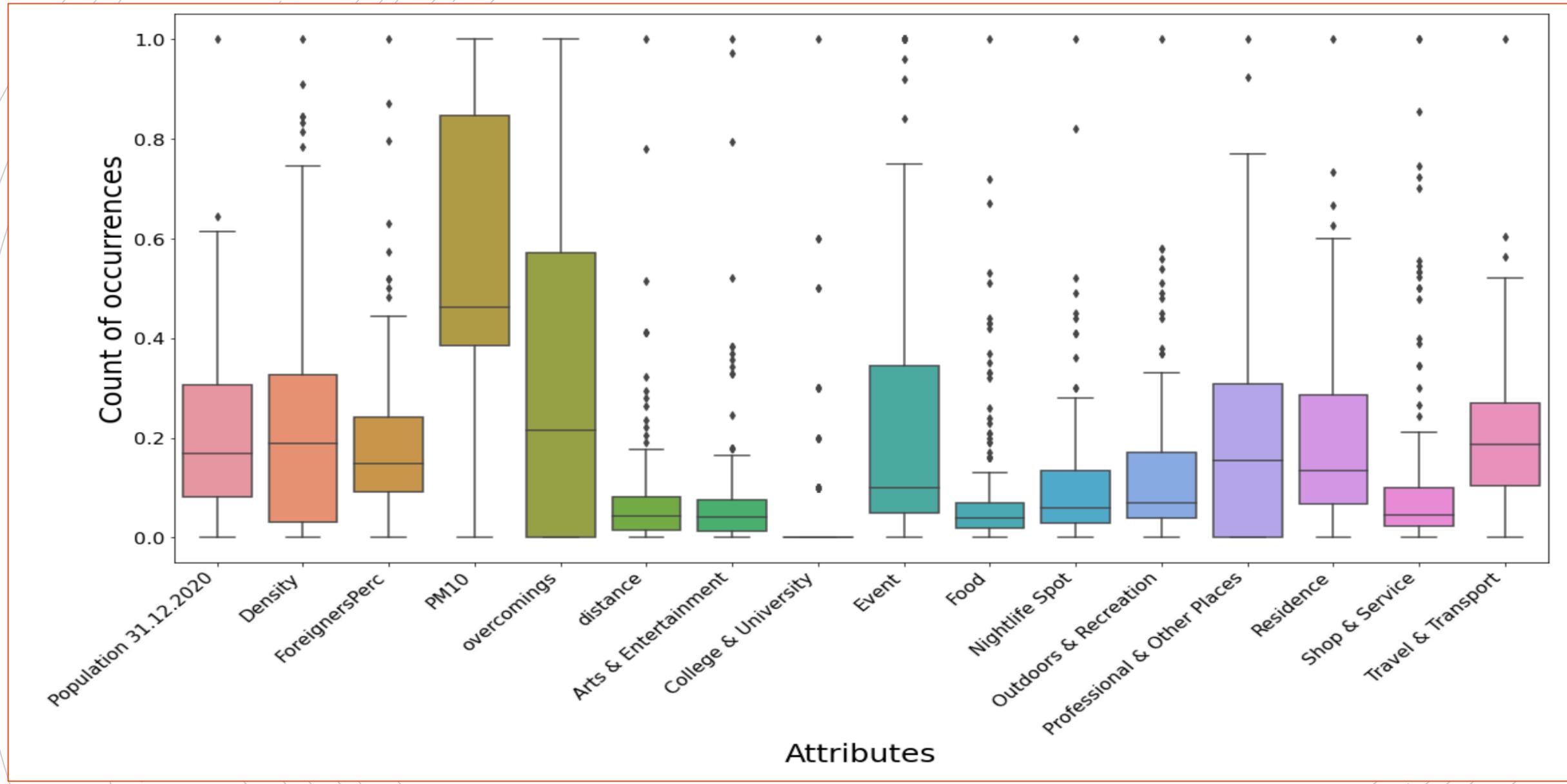


- For each neighborhood, from an address in its center, the latitude and longitude will be derived, and those in turn will be used to retrieve the FourSquare data and venues
- The match with population, density and pollution data will be done on the basis of the neighborhood

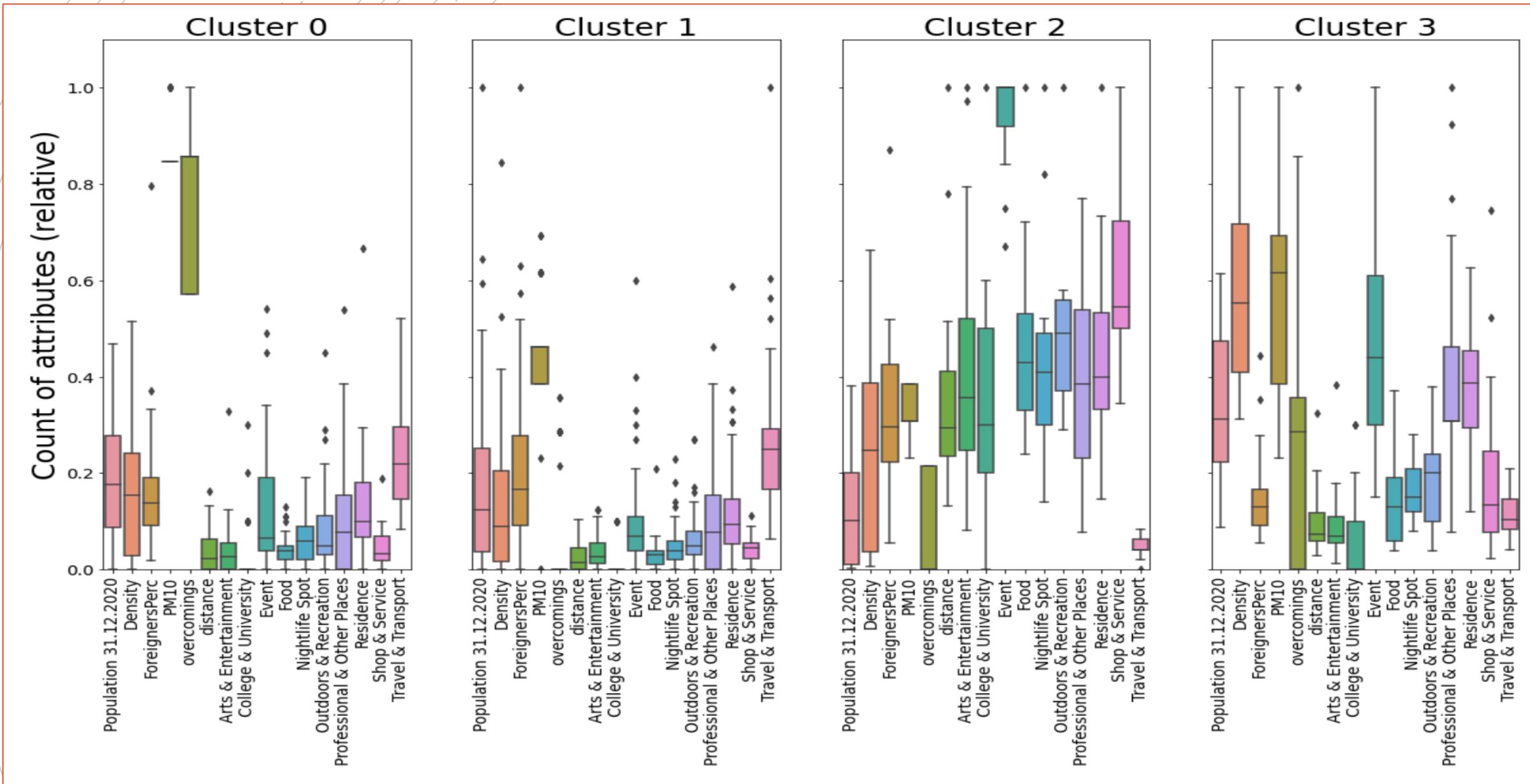


- The features will be normalized for better manipulation of the clustering algorithm, and then a clustering will be performed with 4 clusters suggested by the elbow method
- The clusters will be analysed, using both boxplots, means and maps

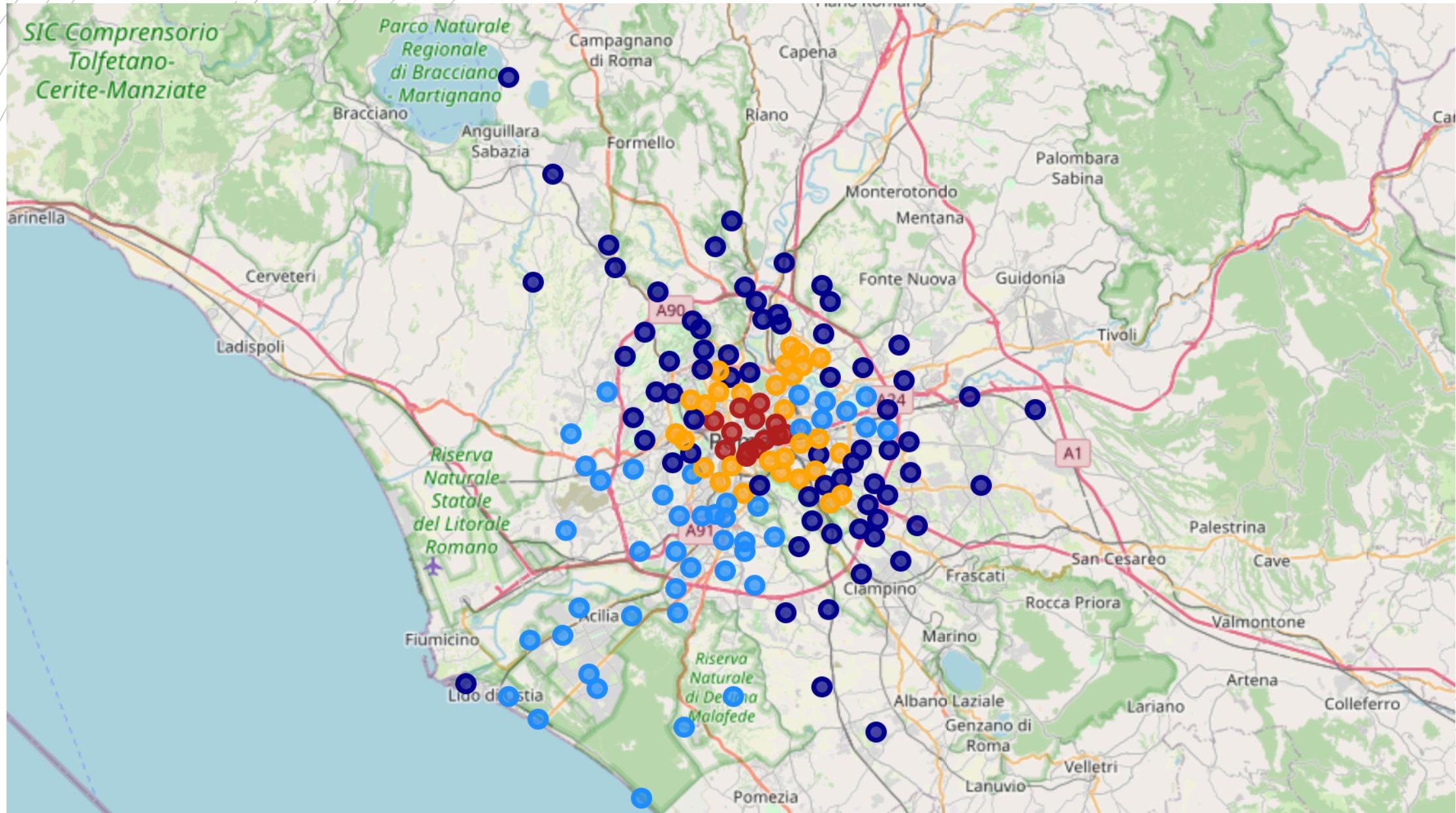
Results – Categories distribution among neighborhoods



Results – Characteristics of the four clusters



Results – Map of the computed clusters



Results



- Clusters 0 and 1 have similar distributions, with Cluster 1 denoting slightly lower values for all features but the percentage of foreign-born residents. The only significant difference between Clusters 0 and 1, in addition to the percentage of foreign-born residents, is air pollution: the neighborhoods in Cluster 0 had higher levels of PM10 and more days with PM10 level exceeding the safety threshold of 50 $\mu\text{g}/\text{m}^3$.

Results



- Clusters 2 and 3 present definitely higher values for almost all examined features. The density of population is much higher in these neighborhoods, with the mean of 6496 residents/square km for Cluster 2 and 14653 for Cluster 3
- There are also more venues in all categories, especially in Cluster 2. Also, Cluster 2 neighborhoods have better air quality and a higher mean percentage of foreigners (22,2% compared with 12,5% for Cluster 3, 12,4% for Cluster 0 and 15,4% for Cluster 1)



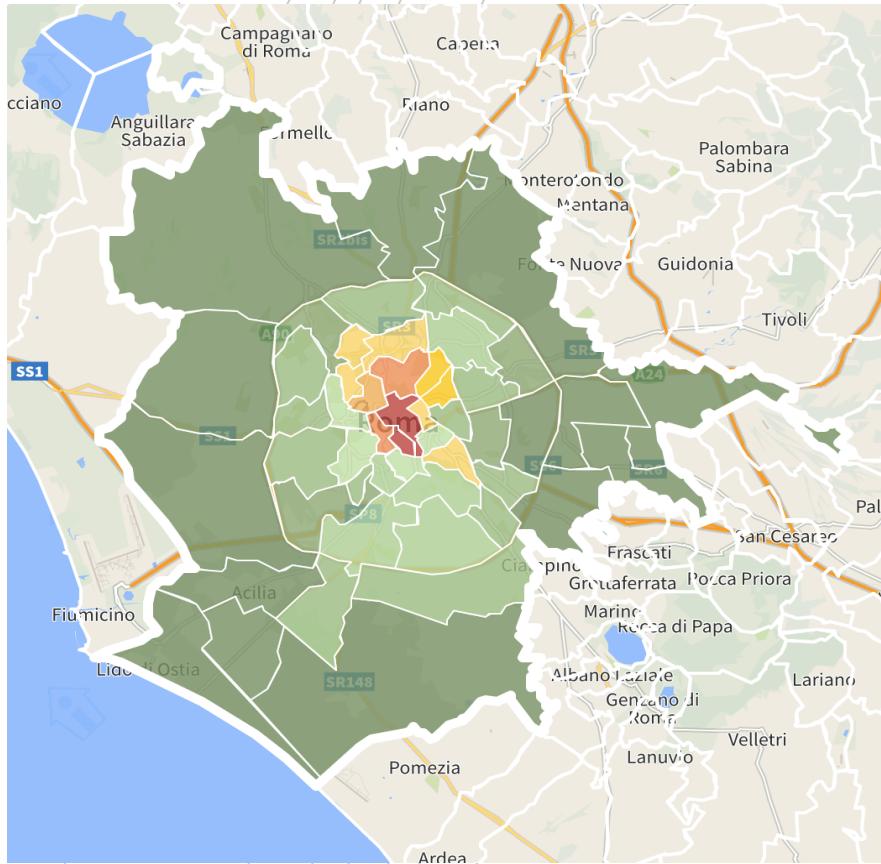
Results



- The highest house prices per square meter can be found in the neighborhoods of the strict center of the city that form Cluster 2
- The mid-range prices can be found in the neighborhoods adjacent the city center that form Cluster 3
- The lowest house price averages belong to clusters 0 and 1, whose neighborhoods are located further away from the city center



Results



- The neighborhoods distribution matched broadly the price ranges given at www.immobiliare.it, as shown on the map

Conclusions



- The project made use of websourced data and a non-commercial FourSquare account
- Notwithstanding these limited resources, it proved strong insights into the neighborhoods
- Using a commercial FourSquare account could give a more granular description of venues and increase the understanding of the neighbourhoods

Conclusions



- The clustering model showed that house prices depend on the density of population, the traffic (air pollution), the presence of venues as shops, restaurants, cinemas etc. and the distance from the city center
- Observing the indicators and their trends over time could enhance property valuation and help investors predict in which neighborhoods the prices are likely to increase or decrease

Further Developments



- To make the model more accurate, some other features could be added, like public transport availability, income statistics or crime statistics.
- An interesting development would be also the use of Google Street View images and deep learning models for image recognition to assess the structural condition of buildings in each area.





THANKS!