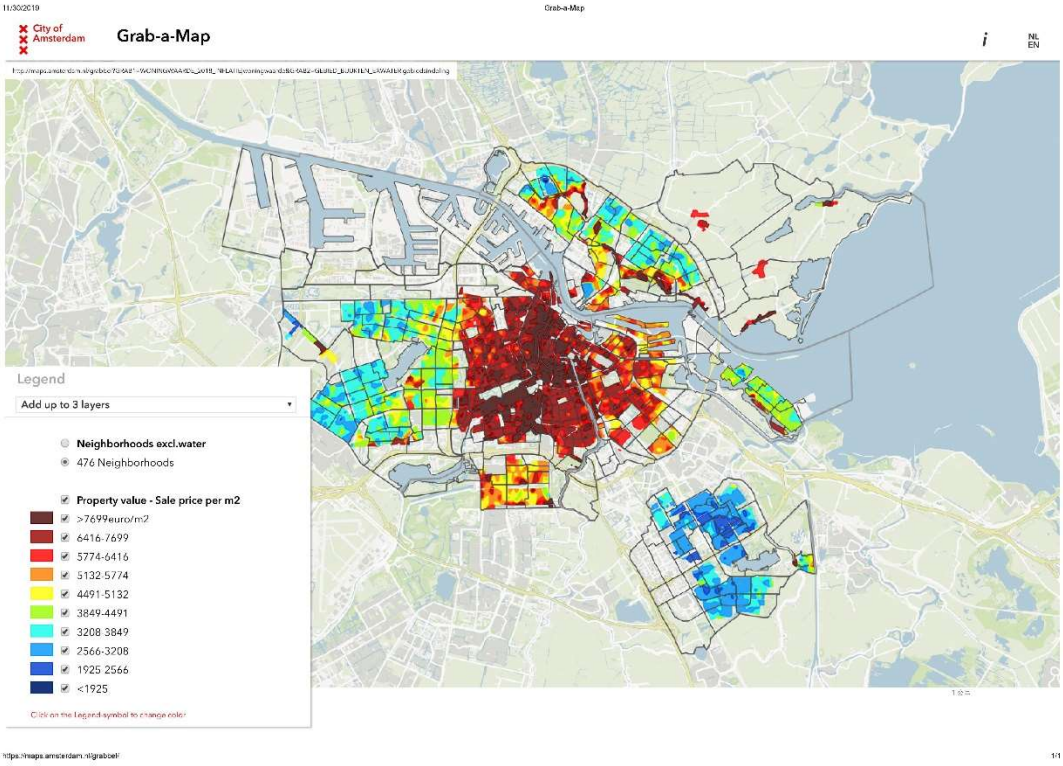
The Battle of Neighborhood

Introduction

The housing price in Amsterdam has been dramatically increasing since 2016, causing by a huge shortage of newly built houses. Although the government has come up with a handful of solutions to tackle the problem, the shortage would still last for a few years. With the median housing price being around €448,000, investing in the real estate market, whether for appreciation or living purpose, will still be a crucial decision for most people. Looking at the following map of neighborhoods distribution and average housing prices, most urban neighborhoods have above price of €5,000 per m2. However, there are still some rural neighborhoods with more appealing price. Therefore, this project tries to find answers to a two-part question: can we find undervalued areas within the urban area, and if not, can we identify possible substitutes in rural areas with similar characteristics with urban ones.



Data

- **The City of Amsterdam**

The City of Amsterdam website has a Maps Data section, which is the accessible and central entry point for anyone looking for objective, reliable and current data and information about the city. This database contains both information of neighborhood division and housing price. Unfortunately, the provider of housing price data is a private company, so the data are neither gathered on a neighborhood level nor modifiable. Consequently, the methodology section will show how the problem will be dealt with.

- **Foursquare**

Foursquare is a location technology platform dedicated to improving how people move through the real world. The platform built a massive dataset of accurate location data. The function this paper employs for analysis is to find popular venues around a certain location.

Methodology

- Housing Price Estimation

Following the data section, to estimate housing price within a neighborhood, the price ranges within a neighborhood are averaged. Taking "Oostoever Sloterplas" as an example, this neighborhood has 3 price ranges in total, namely 3208-3849, 3849-4491, 4491-5132[1].

[1] Please note the price ranges are not the exact values for the price records, the values are stored in the .dbf file within a .shp folder.

Then the price estimation for neighborhood is the average 4170. Apply the same algorithm the every neighborhood, and exclude those neighborhoods without any housing price recordes.

- Venues Collection

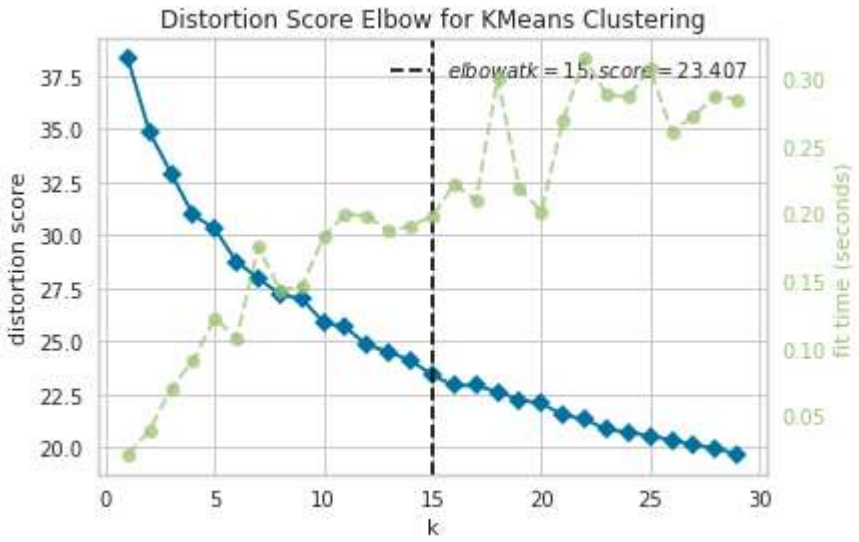
For each neighborhood, the most popular venues within in five hundred meters radius are collected from Foursquare database. The limit of the number of venues is set to 100.

- Clustering

After collecting the venues, there are 354 categories of venues in total, and the occurrence frequency are calculated as input for the clustering algorithm. The following table serves as an example of how the inputs look like.

	Neighborhood	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	…
Alexanderplein e.o.	0.00	0.00	0.00	0.00	0.00	…
Amstelveldbuurt	0.00	0.00	0.00	0.00	0.00	…
Amsterdamse.Bos	0.00	0.00	0.02	0.00	0.02	…
…	…	…	…	…	…	…

This project then uses K-Means clustering method to assign labels to the neighborhoods. After examining the best group number using elbow plot, a group number of 15 is used.



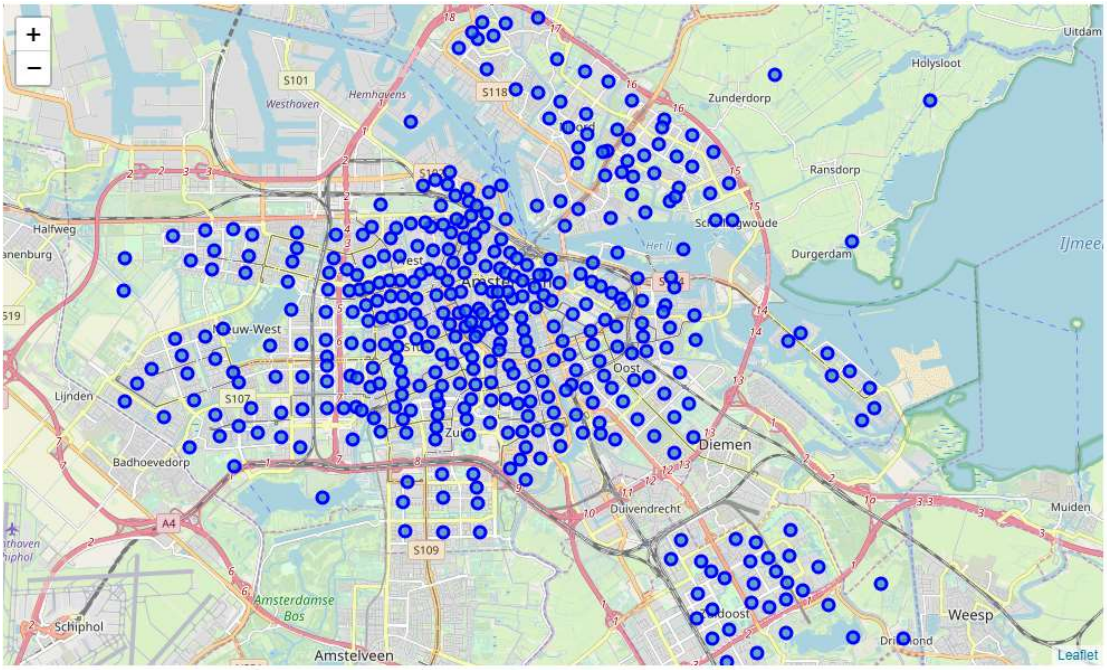
- Valuation

After clustering, the top and bottom 10 percentile are selected. The top 10-percentile neighborhoods are potentially overvalued, and bottom 10-percentile neighborhood are potentially undervalued. One thing to be noted is that clusters with less than ten neighborhoods are excluded.

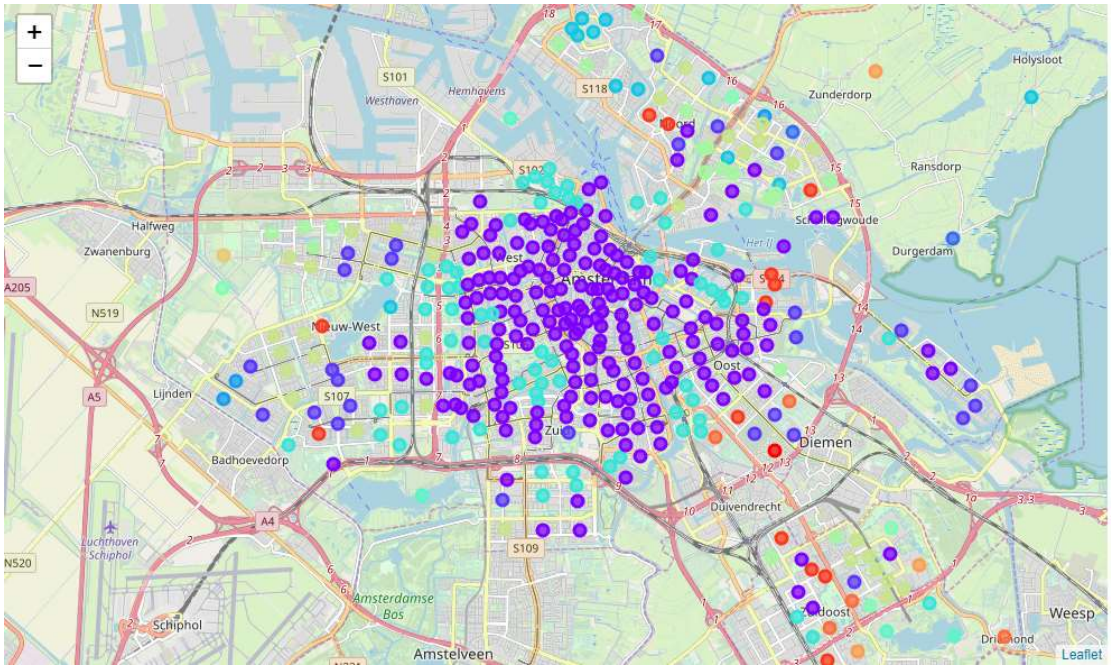
Results

- Clustering

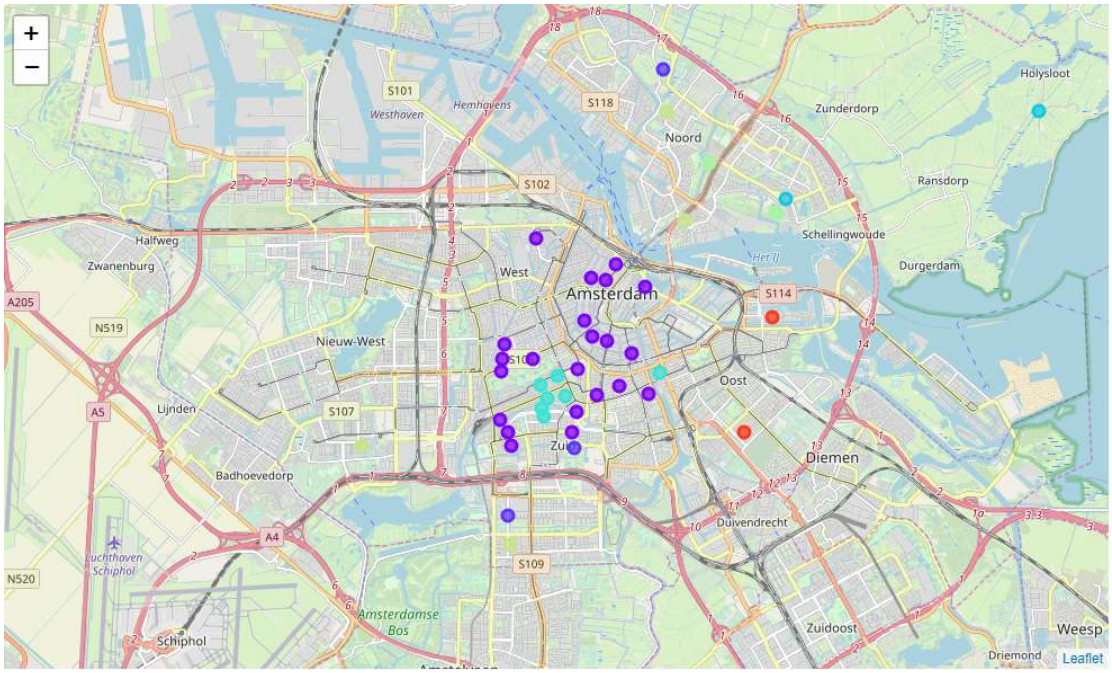
Before Clustering:



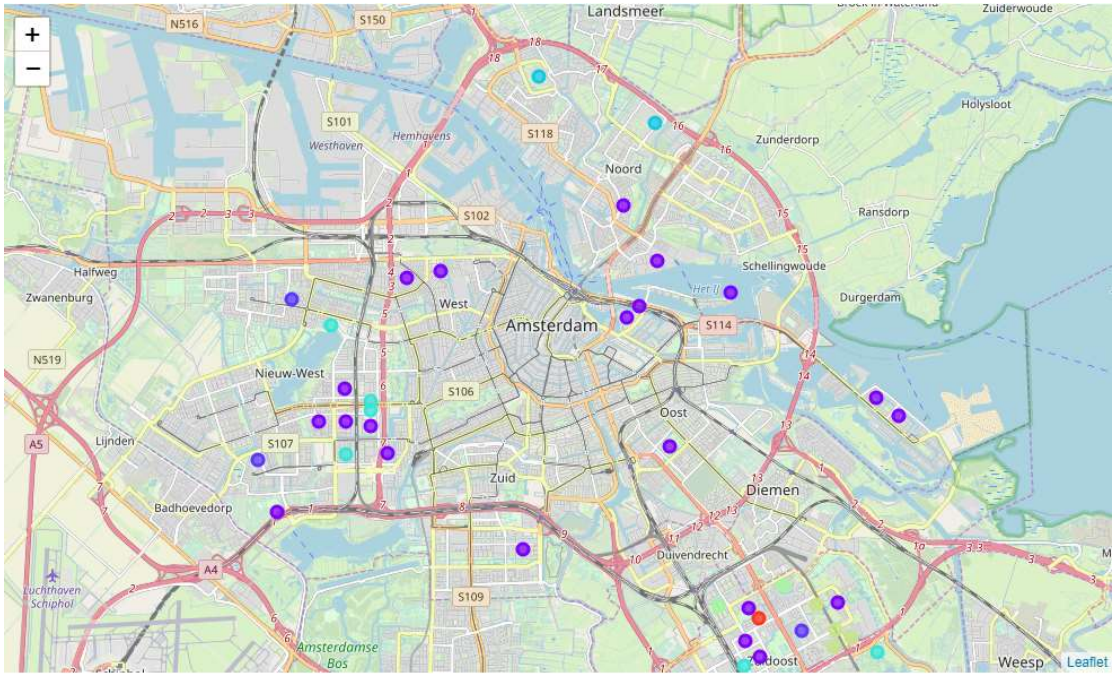
After Clustering:



- Highest 10 percent



- Lowest 10 percent



Discussion

- Homogeneity and effectiveness of clustering

If we compare the maps after clustering with the first housing price graph, we can find that the clustering algorithm captures the general trend of housing price. With purple labels stand for the

most dense and convenient but most expensive neighborhoods. The fact that the majority of the neighborhoods are labeled purple means that Amsterdam is already a highly developed city. Therefore, most neighborhood already have great accessibility to varieties of facilities.

- Limited Opportunities in Downtown

Looking at the distributions of top and bottom 10-percentile neighborhoods within each cluster, most overvalued neighborhoods fall in urban area. This is foreseeable with the on-going real estate booming. Most urban neighborhoods are over-heated, and it is not the best timing to invest.

- Potential Targets in Rural Area

However, potential investment opportunities still exist in rural area. Form the lowest 10 percentile graph, some countryside neighborhoods share similar characteristics with downtown ones, but the average housing prices are just as half as their urban siblings. Although we might have neglected some factors such as transportation, the results do provide some guidance on how to find competitive substitutes away from the city center.

Conclusion

Briefly, this project provides a meaningful clustering approach to group neighborhoods in Amsterdam city and captures some general characteristics that influence housing price. The clustering results confirm that few investing opportunities exist in urban area of the city, but it is possible to find good substitutes in the rural areas with far less average price.

Appendix

Link to the Notebook:

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/08e6c54e-0d8b-4943-9381-1028daaeb345/view?access_token=57e96f05e601d697fc52df971f63db1865b53020a077b051530d67901da6b3de

Neighborhood Data:

https://maps.amsterdam.nl/gebiedsindeling/?LANG=zh

Housing Price Data:

https://maps.amsterdam.nl/woningwaarde/?LANG=zh