Clustering Amsterdam Neighborhoods Buurt or Wijk

CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS

YI LI, LIYI201809@GMAIL.COM

Introduction

In **Amsterdam**, we can basically classify all neighborhoods by **two regional types**:

- **Buurt:** a typical neighborhood or a localized community within a city
- **Wijk:** a more residential area with a lot of old and/or historical family houses.

In this project, I would like to investigate

- Whether these two types of region have a similar **distribution of venues** ?
- Backwardly, can we identify the region type of a neighborhood by **clustering** them based on their venues?

Data Source

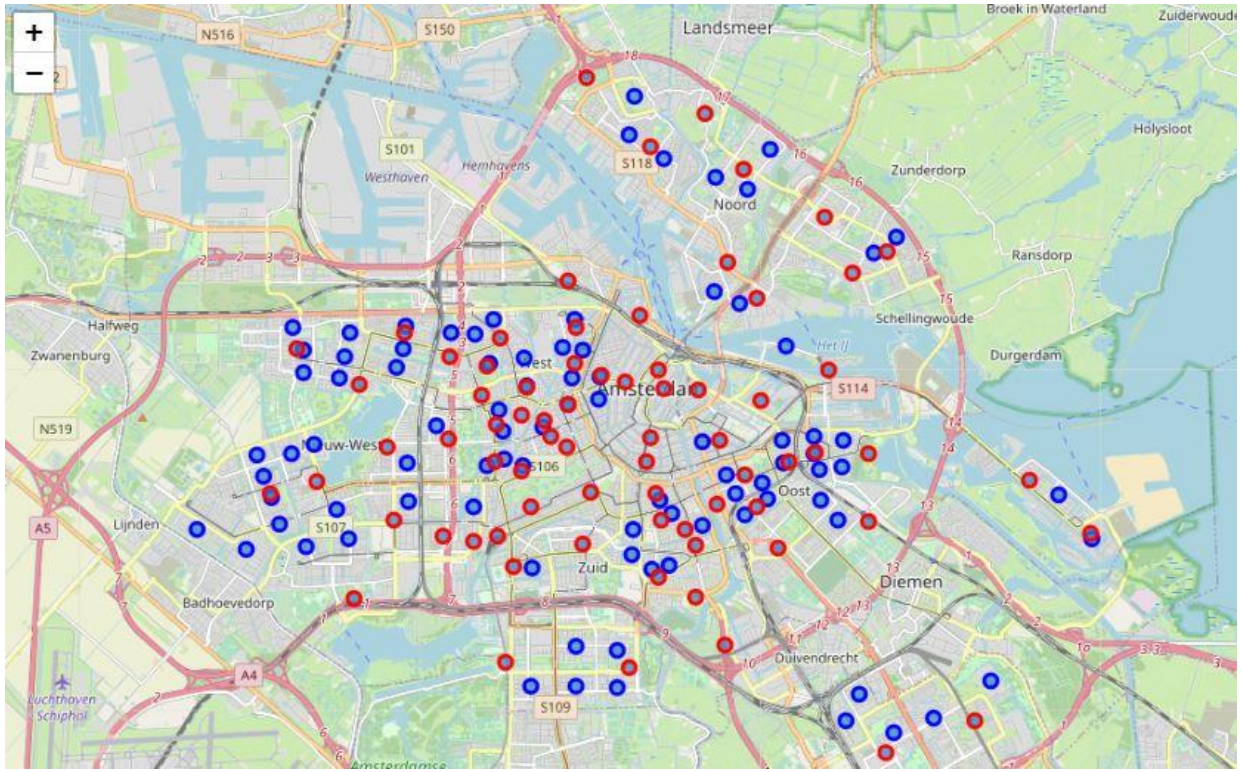
- Statistics Netherlands CBS, a Dutch governmental institution
- districts and neighborhoods of Amsterdam in 2016
- consist of 583 columns(features) and 4019 rows(neighborhoods)

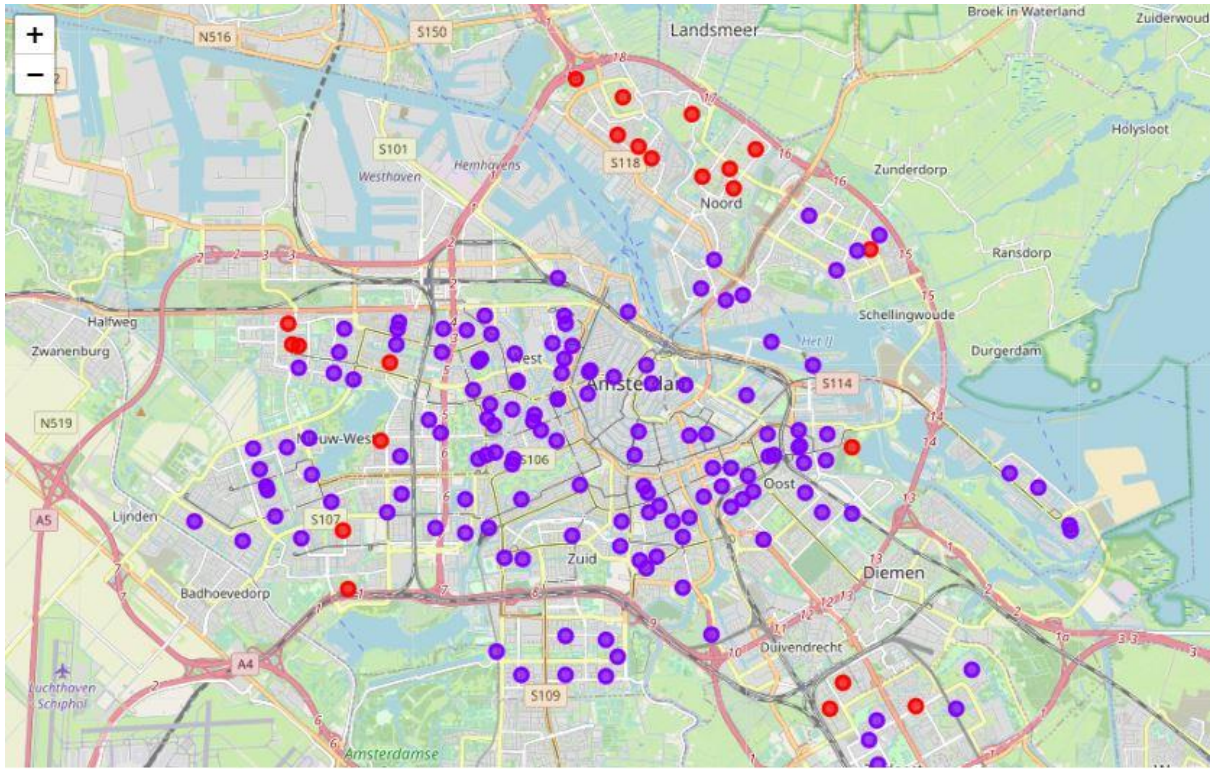
Methodology

- Data preparation: Neighborhood, Region Type, Population, Latitude & Longitude
- Merge with Venue features: Foursquare API
- K-means Clustering: k = 2
- Evaluation: accuracy scores and F-1 score

Results

- Geographical distribution: Buurt/Wijk vs. Clusters on Venues





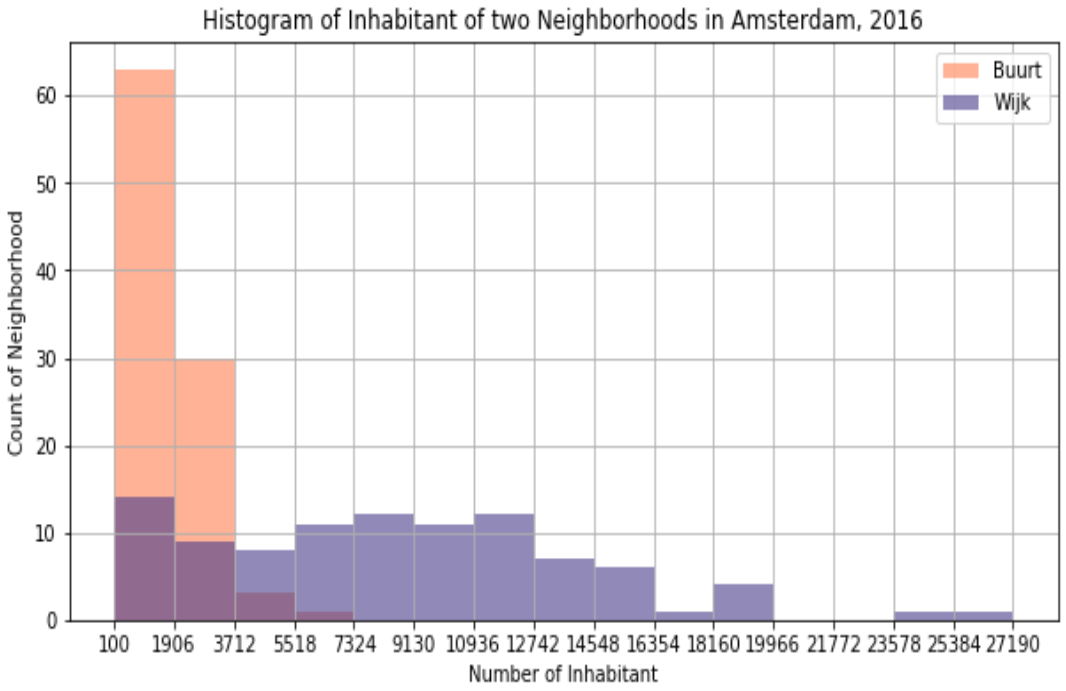
Result

Evaluation methods compare clustering labels (1/0), and the ground truth label converted from region_type feature, Buurt/Wijk

- Accuracy: 0.5342
- F-1 Score: 0.612

Discussion

- From the geographical results
  - Buurt and Wijk are pretty mixed together with each other
  - clusters derived based on venue information are clearly separated.
- From the evaluation results
  - that the clustering labels are not well match the ground-truth labels
- Recommendation
  - Instead of venue information, some other feature can be more indicative for classify Buurt/Wijk. For example: inhabitant



Conclusion

- Buurt and Wijk are geographically mixed, the distributions of their venue categories are also quite similar. It is hard to know/cluster the region types of a neighborhood based on its venue information
- Some other features might be more indicative to check the region types (Buurt or Wijk). Thus, building some inferential statistical models can help to improve; however, this is out of the scope of this projects.

THANK YOU!