Exploring neighbourhoods in Overijssel, the Netherlands

Younjung Choi  
Applied Data Science Capstone, Coursera

# Introduction

## Problem statement

Finding a house in big cities in the Netherlands is getting challenging to both young local people and internationals in the Netherlands. House price in such cities like Amsterdam and Utrecht has rapidly increased. Moreover, due to the lack of property supplies and the resulting competitive bidding processes, the home seekers are troubled.

Overijssel, located in the east side of the Netherlands, could be an alternative choice to the home seekers because its lower housing prices, as well as its less competitive housing market. However, due to the insufficient information on its neighborhoods, Overijssel does not strongly attract home seekers. Therefore, this paper aims providing a GIS-based information presenting clustered neighbourhoods in Overijssel, the Netherlands

## Intended readers

This paper is expected to help two kinds of stakeholders, home seekers who are willing to live outside of the Dutch major cities and real estate agencies working for Overijssel.

# Data

This study obtained the following three kinds of data from APIs and a website:

1. the postal codes of neighbourhoods in Overijsel (https://www.metatopos.eu/overijssel2.html),
2. the latitude and longitude of each of these postal codes (Bing Map API, https://www.bingmapsportal.com/Account/Register), and
3. the information on venues located in around these neighbourhoods such as names, business type, and location data (Foursquare API, https://developer.foursquare.com/).

The instructions for the coding procedures were mainly provided by the Coursera course “Applied Data Science Capstone”, as well as other online materials.

# Methodology

As illustrated in Figure 1, the methodology consists of four steps. Firstly, we scrapped the names of cities and neighborhoods in Overijssel, and the postal codes of each of the neighbourhoods from a website using Python libraries of ‘requests’ and ‘BeautifulSoup’. To make a data frame consisting of these data, we employed Pandas and Numpy libraries. Secondly, for visualizing the neighbourhoods in a map we obtained their location data by means of the Bing Map API and the geocoder library, then the visualization was done using the folium module. Thirdly, we got the information on maximum 100 venues located within 500m away from each of the neighbourhoods from the Foursquare API. Next, we made a new data frame that shows the top 10 venues for each neighbourhood. Lastly, to categorize the neighbourhoods by means of the top 10 venues, we employed a machine learning algorithm ‘KMeans’. The KMeans algorithm allowed neighbourhoods to be categorized into 5 clusters.

Figure 1 Methodology

# Result

The five clusters of neighbourhoods are presented in Figure 2. The first cluster in red comprised neighbourhoods whose top one venues were mainly supermarket. The third cluster in blue has most neighbourhhods (84 neighbourhoods).

Figure 2 Result

Map

Description automatically generated

# Discussion

The five clusters did not have significant differences. Real estate agencies could say the two cities, Enschede and Zwolle, have the most neighbourhoods; however, it might be complicated for them to introduce other cities or neighbourhoods having distinguished characteristics in terms of venues. The KMean algorithm might be not suitable for this reason.

# Conclusion

This study showed the two cities Enschede and Zwolle have the most neighbourhhoods. However, it was complicated to distinguish neighbourhoods by means of their venues.