

# **Examining potential interesting real estate investment opportunities in the province Utrecht in the Netherlands**

**Author :** Arne van den Bosch

**Date :** 8/3/2021

Created as part of the final module of the IBM Data Science Professional Certificate Course – the ‘Capstone Project’

# **Table of contents**

- Business Problem**
- Description of the data**
- Methodology and results**
- Conclusion**

# Business Problem

- A real estate investment firm likes to expand its activities to province Utrecht (Netherlands)
- Assignment: Find potential areas in province Utrecht that are potentially interesting for the investment firm to purchase and resell property.
- The following criteria should be met for potentially interesting areas:
  - Stable increase in house price over at least the past 5 years
  - Current house price in the area should be relatively low
  - Current house price trend shows house prices will very likely increase
  - There is a variety of venues nearby, i.e. supermarkets, restaurants, shopping malls, etc
  - Additional factors that could be important:
    - Low crime rates / nuisance figures
    - Lower energy consumption
    - Low number of people using employment allowance
    - High number of organisations being nearby (employment opportunities)

# Description of the data

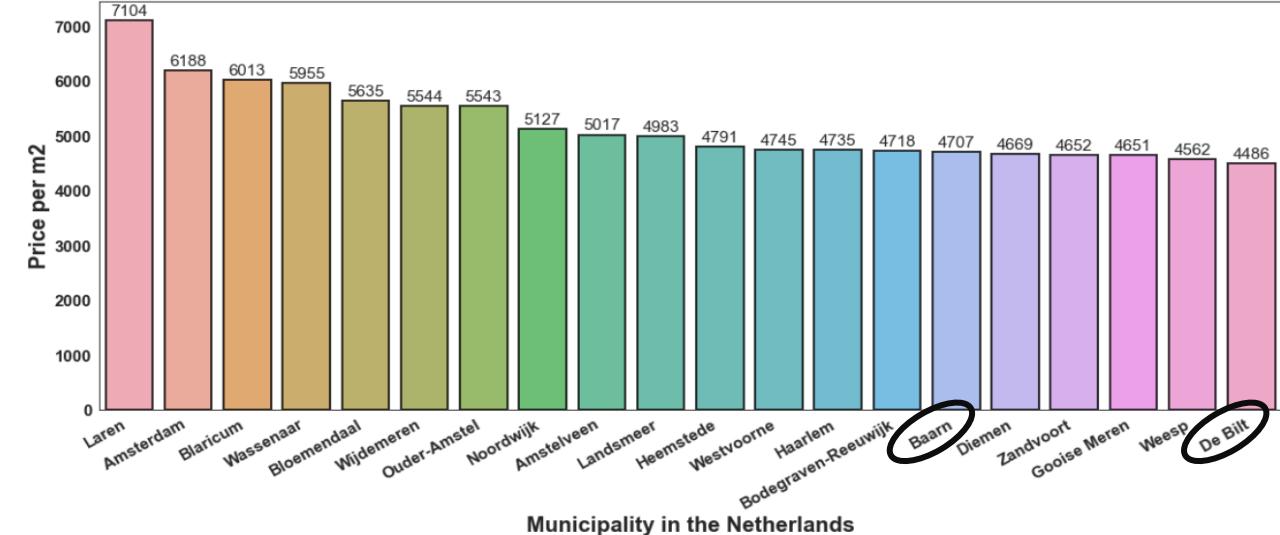
- **House prices per m<sup>2</sup>** : Excel file from [huizenzoeker.nl](http://huizenzoeker.nl)
  - Average house prices per m<sup>2</sup> for each municipality in the Netherlands
  - Data from September 2008 to February 2021
  - Pre-processing: replacing old municipalities, merging with geospatial data frame, filtering for municipalities in province Utrecht, transposing the data frame for looping purposes, renaming columns, adding and dropping columns
- **Geospatial data** : [Geojson file from hub.arcgis.com](http://hub.arcgis.com)
  - Polygon (geometry) data for each municipality in the Netherlands (2018)
  - Pre-processing: replacing old municipalities, merging with other datasets based on municipality column, converting polygon into centre points, latitude, longitude and coordinates in a tuple
- **Venue location data** : [Foursquare – API](http://Foursquare)
  - Venue latitude, longitude, venue name and venue category for each municipality within 5km radius from municipality centre
  - Maximum amount of venues per municipality was 100
  - Retrieving data using a function containing API request and Get request
  - Pre-processing: grouping, sorting, selecting top 10, dropping columns, inserting columns/arrays, clustering data using k-means
- **Municipality statistics** : [Central Bureau of Statics \(CBS for short\) - API](http://Central Bureau of Statistics (CBS for short) - API)
  - Nr of inhabitants, electricity/gas usage, avg income, %low income, nr of unemployment allowance, nr of companies
  - Data obtained via library cbsodata-library and ‘cbsodata.get\_data(‘table number’)’ and url ‘opendata.cbs.nl’
  - Pre-processing: column selection/renaming, removal leading/trailing white space, filtering, dropping, replacing old municipalities, merging with geospatial data,

# Methodology and results

House price per m<sup>2</sup> in Feb 2021 for the Netherlands

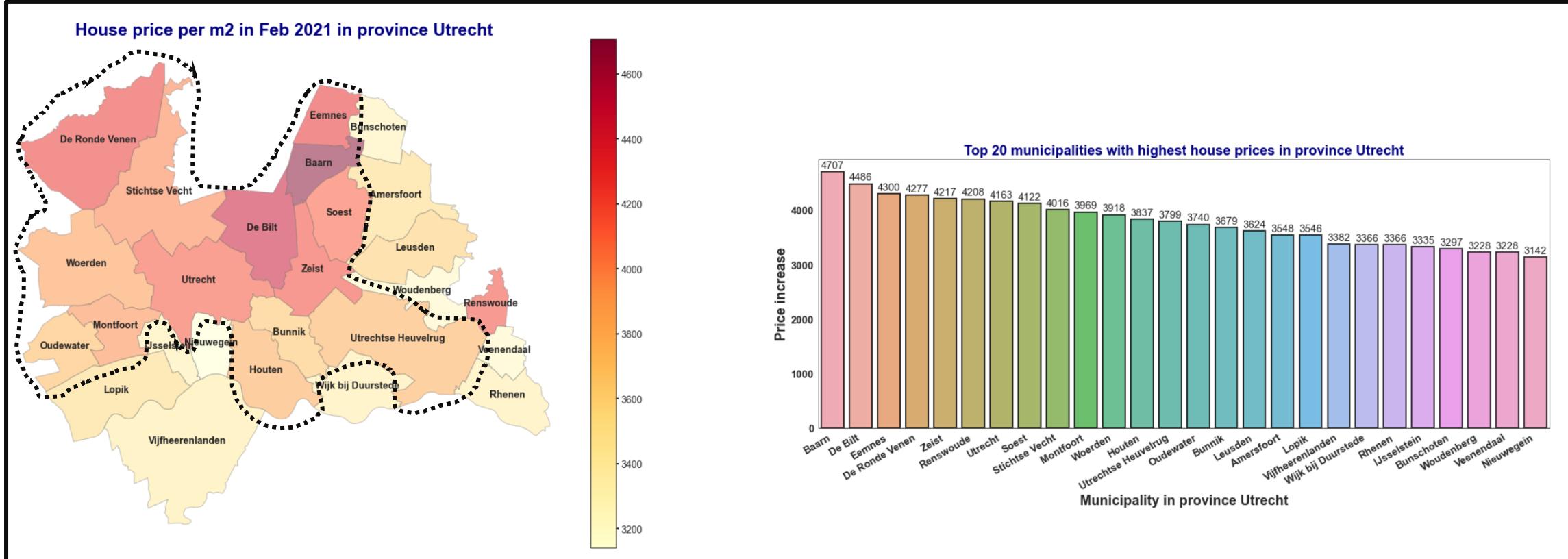


Top 20 municipalities with highest house prices in the Netherlands



- Price per m<sup>2</sup> obtained from Excel file “Huizenzoeker.nl”, geospatial data from geosjon file “arcgis.com”
- Province Utrecht (indicated on the map) is the 4<sup>th</sup> largest city in the Netherlands
- Utrecht is amongst the areas where house prices is highest (price range: €3,228 to €4,707 per m<sup>2</sup> as per feb. 2020)
- In the bar chart we see the 2 most expensive areas in Utrecht (Baarn, De Bilt) which are on place 15 and 20

# Methodology and results

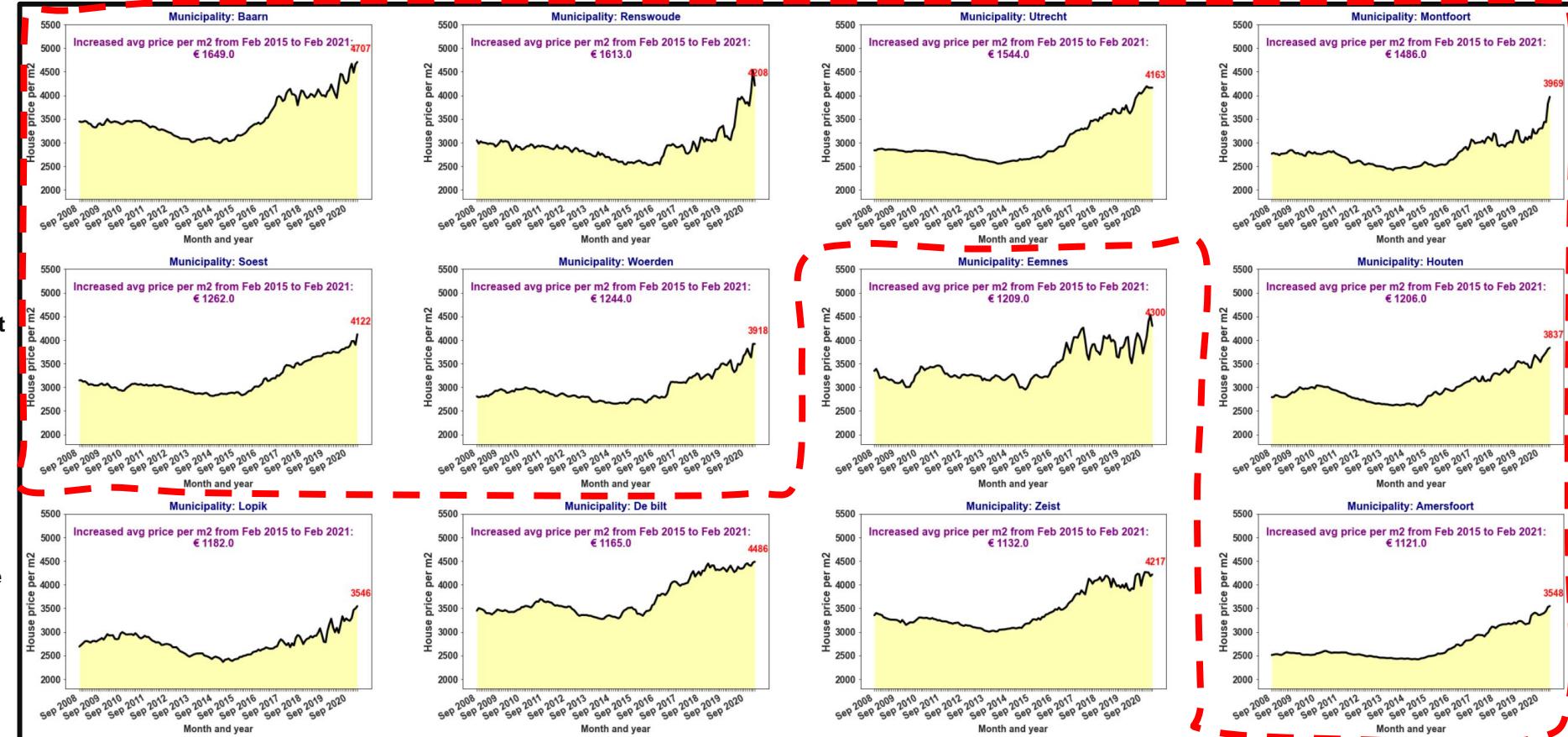


- The encircled area shows areas where house prices are average (orange) or in the higher spectrum (reddish)
- There are some municipalities with relatively low prices near the main city Utrecht, which are IJsselstein, Nieuwegein)
- Municipality Amersfoort, which the second largest city in province Utrecht, also has relatively low average house prices per m<sup>2</sup>.
- Other places with relatively low prices: Leusden, Bunschoten, Veenendaal, Rhenen, Woudenberg, Vijfheerenlanden, Lopik

# Methodology and results

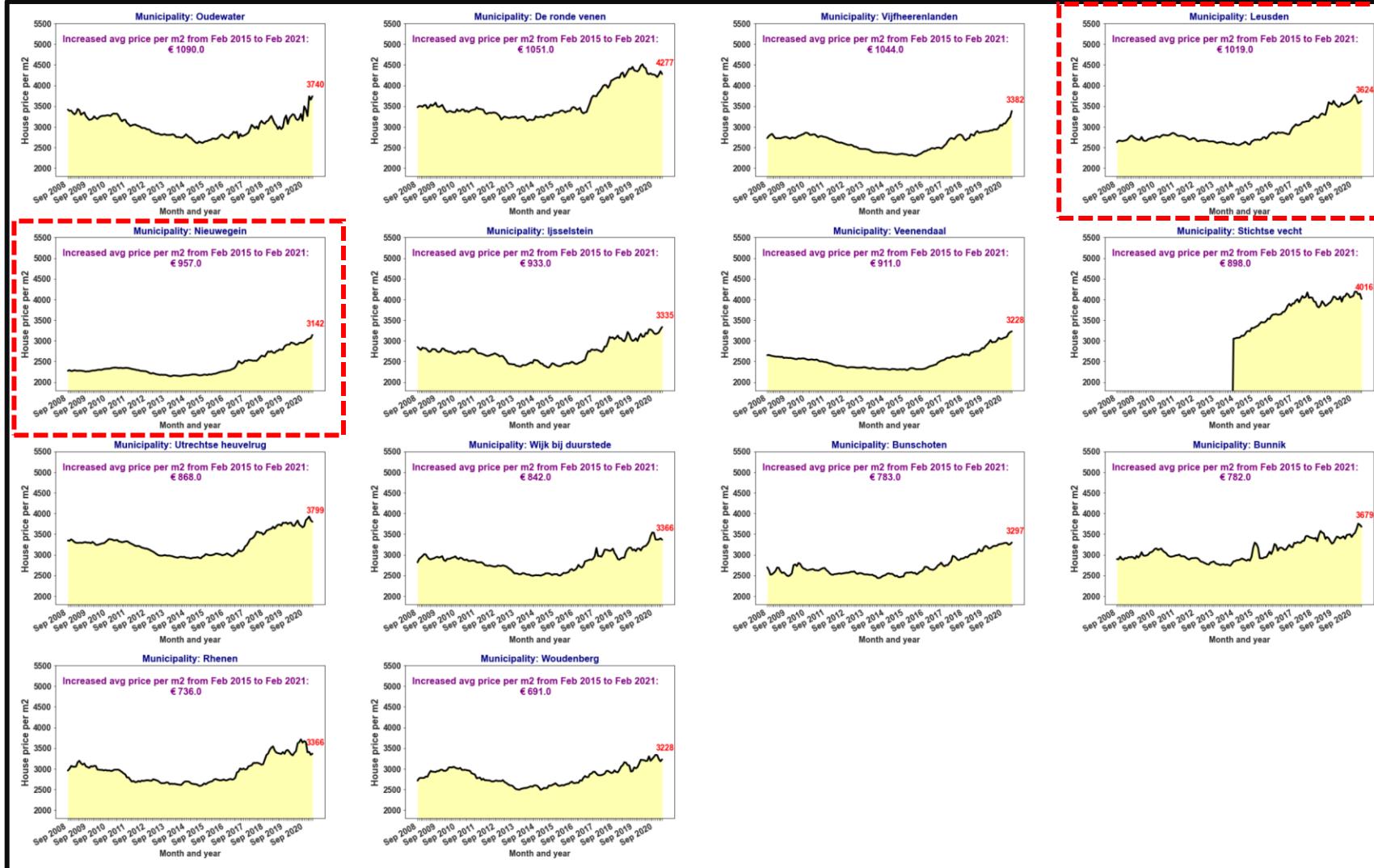
Line charts – showing trends for house prices per municipality in Utrecht

- Line charts were created using a loop function and the subplot functionality of matplotlib.
- The charts are placed in descending order of average house price increase per m<sup>2</sup>
- This slide shows the 12 municipalities with highest house price increase between February 2015 and February 2021.
- Most municipalities with highest house price increase are also the most expensive municipalities at the moment
- Looking purely at price increase over the past 6 years, then the **red encircled dashed municipalities** would be interesting areas to invest in. Areas where house prices seem to reach a plateau or are erratic have been omitted.
- A criterium from investment agency was combination of relative low current house price, but stable house price increase over past 6 years
- A few municipalities with relatively low house price, but reasonable house price increase and stable upward trend:
  - Amersfoort
  - Lopik
  - Nieuwegein (next slide)
  - IJsselstein (next slide)



# Methodology and results

## House prices (continued)

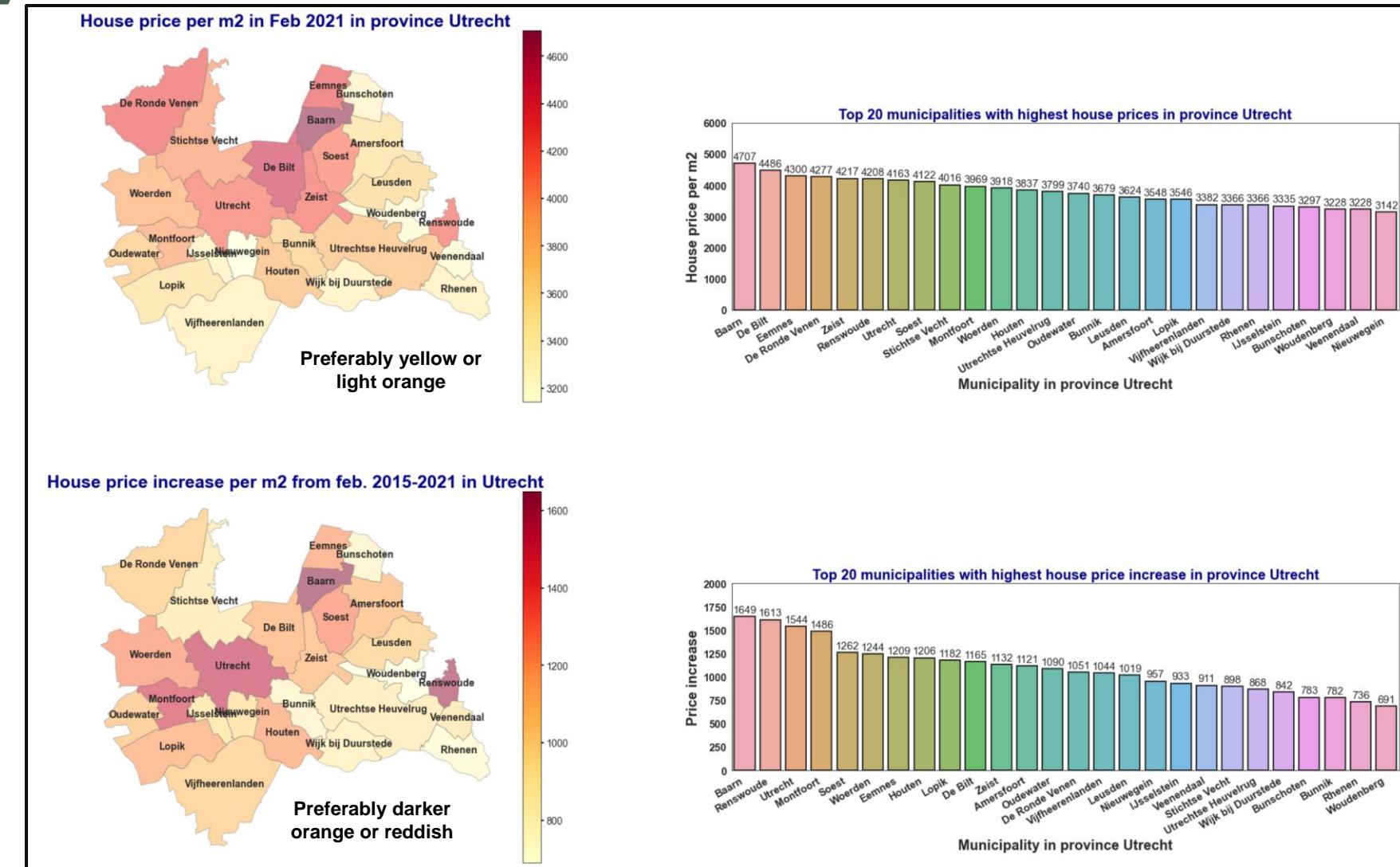


# Methodology

Another choropleth map was created for Increase in house price per m<sup>2</sup> over the past 6 years. The two choropleth maps could be compared to see municipalities with low house prices but high increase in house price per m<sup>2</sup>. The following municipalities could be selected that met both criteria:

- Montfoort
- IJsselstein
- Nieuwegein
- Amersfoort
- Woerden

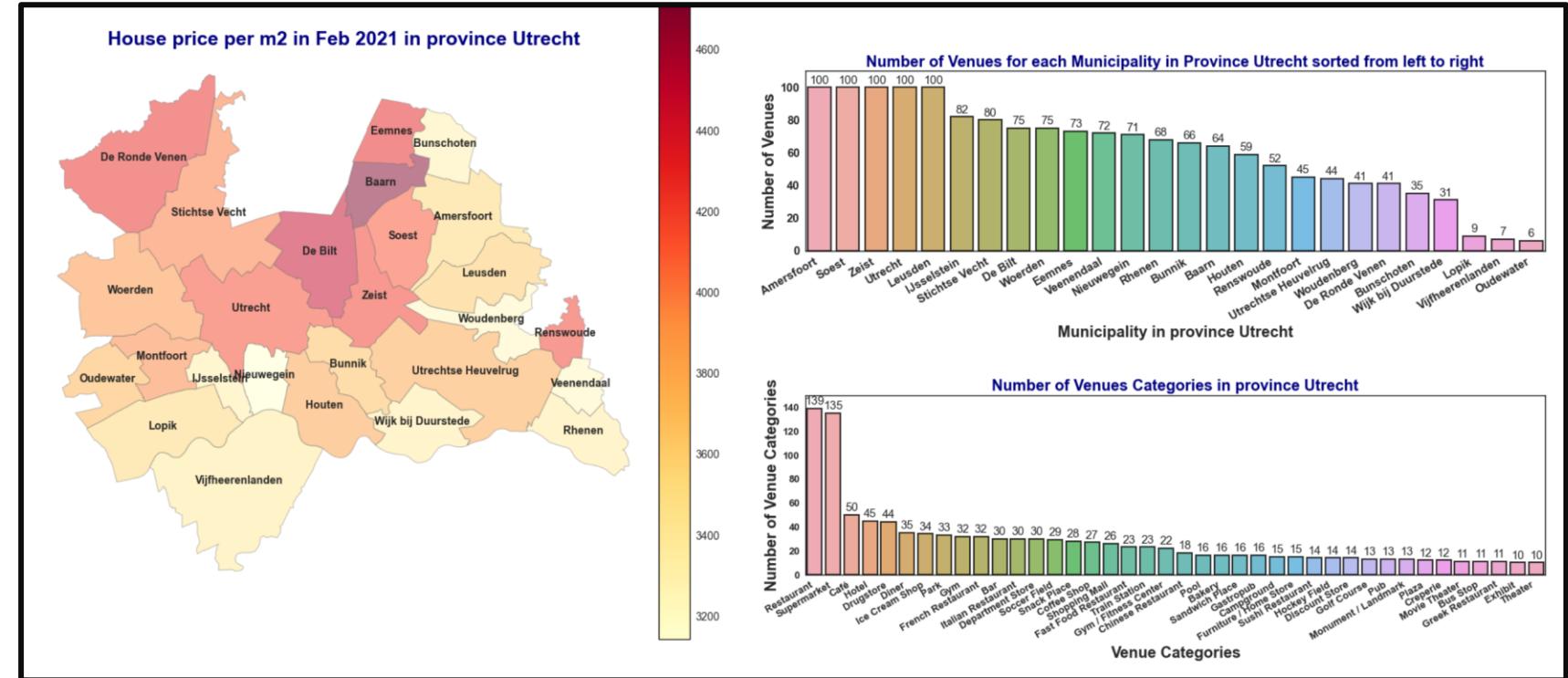
Looking for areas with low house prices and high house price increase



# Methodology and results

Looking for areas with low house price, but high venue count

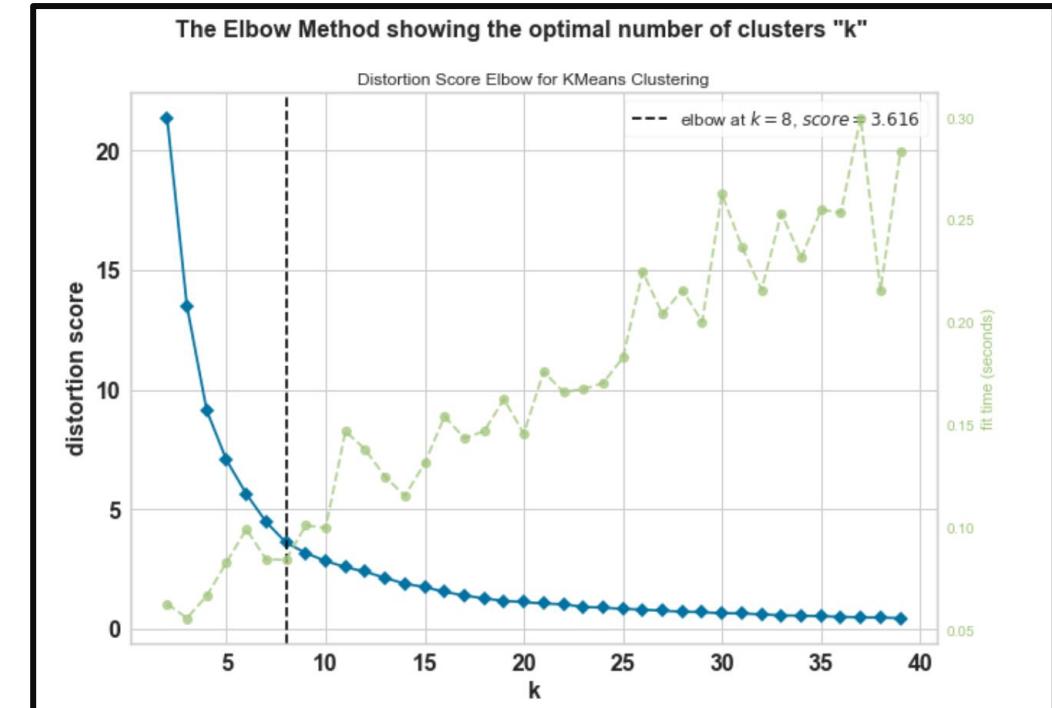
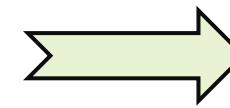
- Using the Foursquare API, venue location data was collected including, latitude and longitude of the venue, venue name and venue category
- Venues were collected for each municipality based on a 5km radius around the centre point (midpoint) of each polygon of the municipality
- The maximum number of venues that was collected per municipality was set to 100. That is why the first 5 municipalities in the upper bar chart stop at 100.
- Comparing the upper bar chart with the choropleth map with house prices, the following becomes clear:
  - Areas with higher house prices do not necessarily have more venues
  - A few areas with lower house prices but high venue count are for instance:
    - Amersfoort
    - Leusden
    - IJsselstein
    - Woerden
    - Veenendaal
    - Nieuwegein



# Methodology and results

Applying the ‘elbow’ method to find the optimum amount of clusters ‘k’ to split the dataset in

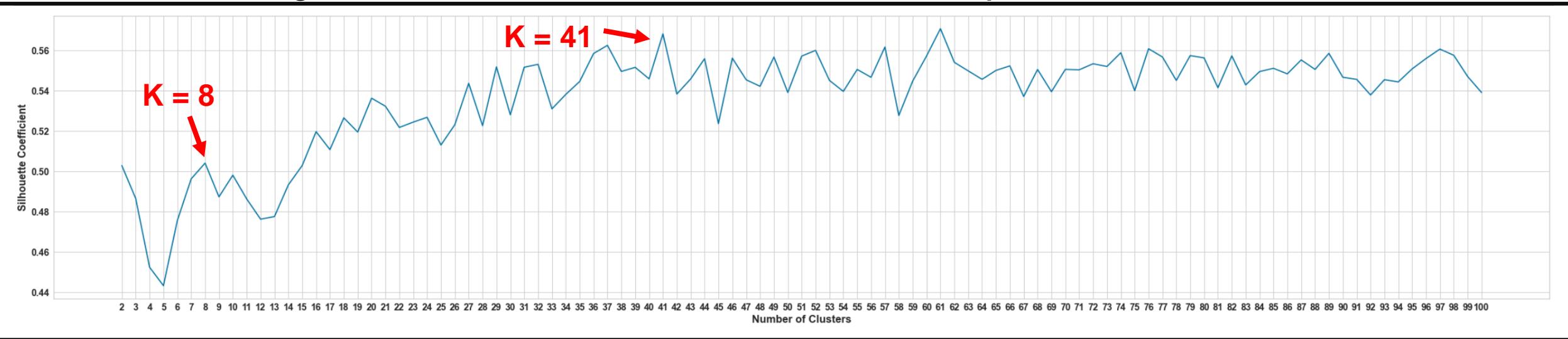
	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kasteel De Hooge Vuursche	52.203667	5.245899	Monument / Landmark
1	Paleis Soestdijk	52.194061	5.280125	Palace
2	Kasteel Groeneveld	52.218676	5.255431	Monument / Landmark
3	Hoeve Ravenstein	52.216226	5.255728	Farm
4	Vuursche Bos	52.203181	5.234696	Trail
...	...	...	...	...
1241	Knooppunt Lunetten	52.055450	5.144516	Intersection
1242	Busstation Lekbrug Oost	51.990818	5.082772	Bus Station
1243	De Lek	51.989969	5.083416	Beach
1244	Sushi & Meer	52.064542	5.135597	Sushi Restaurant
1245	Sporthal Lunetten	52.063064	5.142305	Gym
1246 rows × 4 columns				



- A selection of venue coordinates (red circled section above) was applied for applying the K-means algorithm
- Objective was to find and visualize clusters of venues, where each cluster represents areas where venues are near each other
- Putting the resultant clusters as coloured markers on top of a choropleth map could give some insight about amount and type of venues in expensive versus less expensive areas
- First, the elbow method was applied and an optimum amount of clusters was found to be  $k = 8$ .

# Methodology and results

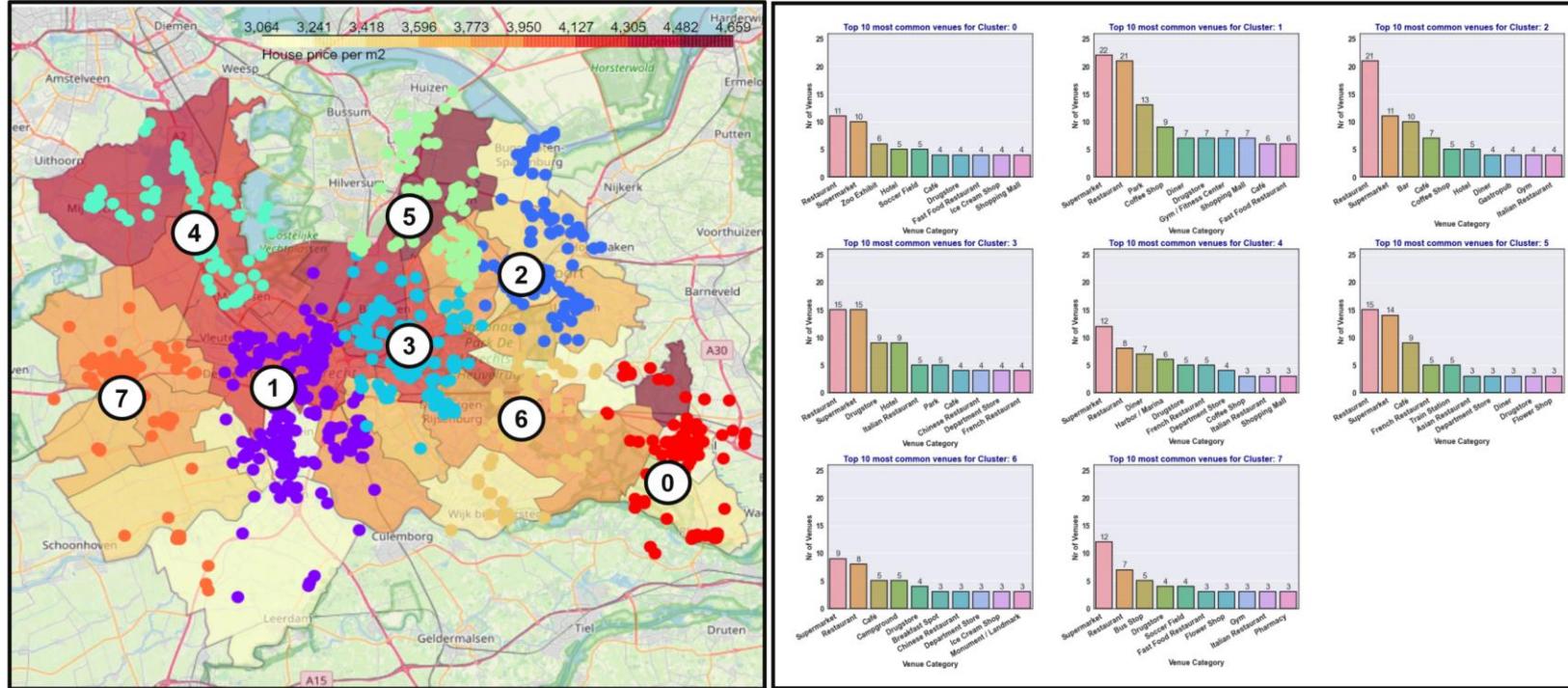
Plotting a “Silhouette Coefficient” in a line-chart to find the optimum number of clusters ‘k’



- I had some difficulties in interpreting the Silhouette plot. I read that a higher Silhouette Coefficient would generally be better, but in all examples I saw there was one obvious peak in the chart in contrast to what is shown in above chart. The chart also starts with a decreasing trend.
- I had selected an amount of clusters of 41, but later discovered that after re-running the syntax in the notebook, the updated charts changed, except for the first part. Therefore I doubted the reliability of the outcome.
- Eventually I choose K = 8 as the number of clusters, relying more on the elbow method on previous slide.

# Methodology and results

Making a choropleth map with the venue as coloured markers on top



- With K = 8, the K-means algorithm was applied and the resultant clusters were inserted in the original data frame with venue information
- Using the folium library, the clusters were put on top of an interactive folium map.
- For each of the clusters, the top 10 most common venues were plotted in bar charts as shown next to the choropleth map on the right
- The resultant clusters turned out to be not particularly useful in identifying interesting investment areas, however – the visualization of the markers on the map in combination with the choropleth map was helpful.
- In cluster area 2 (dark blue markers) we see that there are relatively a lot of venues, but house prices are low- to average.
- The south part of cluster area 1 (purple markers) also show an area with a lot of venues, but where house prices are relatively low.
- These low-priced areas near the city, where there are multiple venues near each other are potentially interesting areas.

# Methodology

- The clustering was also applied to a dataset in which the Venue Category was filtered for outdoor drinking and eating facilities. Popular living areas have a variety of these type of venues and putting these on a map could help identify potential interesting investment areas.
- Applying the 'elbow' method to find the optimum number of clusters 'k' resulted in the same amount of clusters and a similar trend as before. The optimum amount was found to be k = 8.

## Filtering the list with venues for outdoor eating and drinking facilities

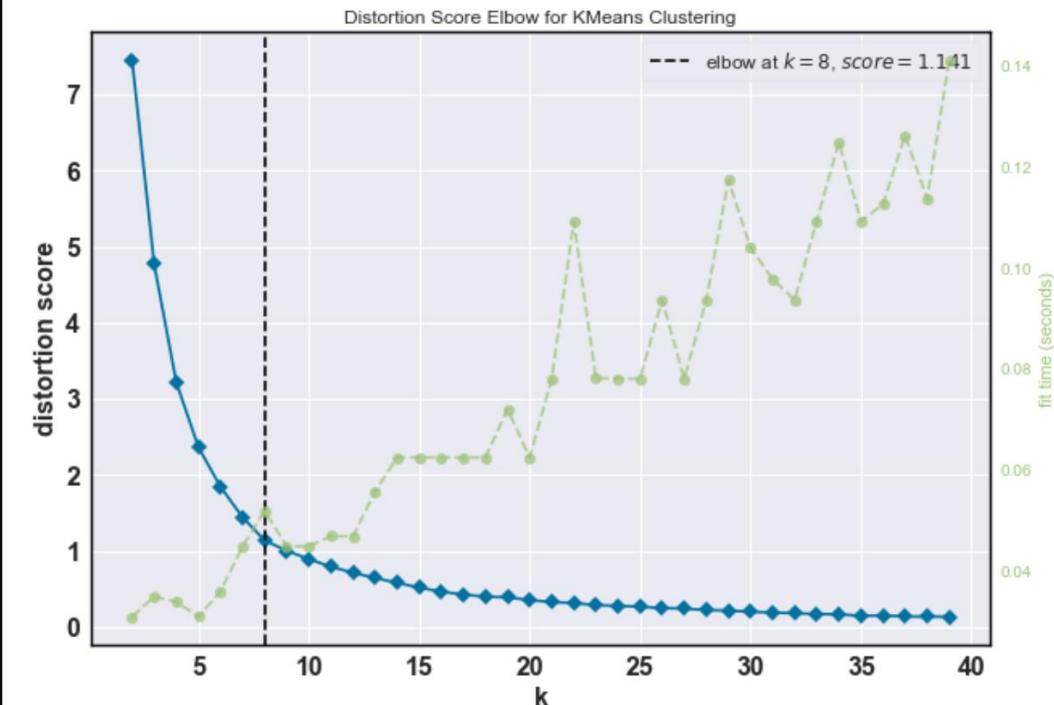
Make a list of venues related to drinking and eating venues outdoor and use the list to select relevant venues from above datafram

```
In [74]: List_of_Venues_Eat_or_Drink = ['Restaurant','Café','bistro','brasserie','joint','Cafetaria','Coffee','Bar','Diner','Food','Ice Cream','Pizza','Pub','Salad','Sandwich','Snack','Wine','Drink']
Nearby_Eat_and_Drink_Venues_Utrecht = Cluster_df[Cluster_df['Venue Category'].str.contains('|'.join(List_of_Venues_Eat_or_Drink))].reset_index(drop = True)
Nearby_Eat_and_Drink_Venues_Utrecht.head(5)
```

	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Smaecken van Hamelink	52.211656	5.291049	Ice Cream Shop
1	Restaurant Vuur	52.202181	5.247596	Restaurant
2	The Golden Coffee Box (Boot koffie)	52.211594	5.288511	Coffee Shop
3	Iussalon Bemer	52.187868	5.283528	Ice Cream Shop
4	Cosa cucina & wine bar	52.211967	5.284038	Italian Restaurant

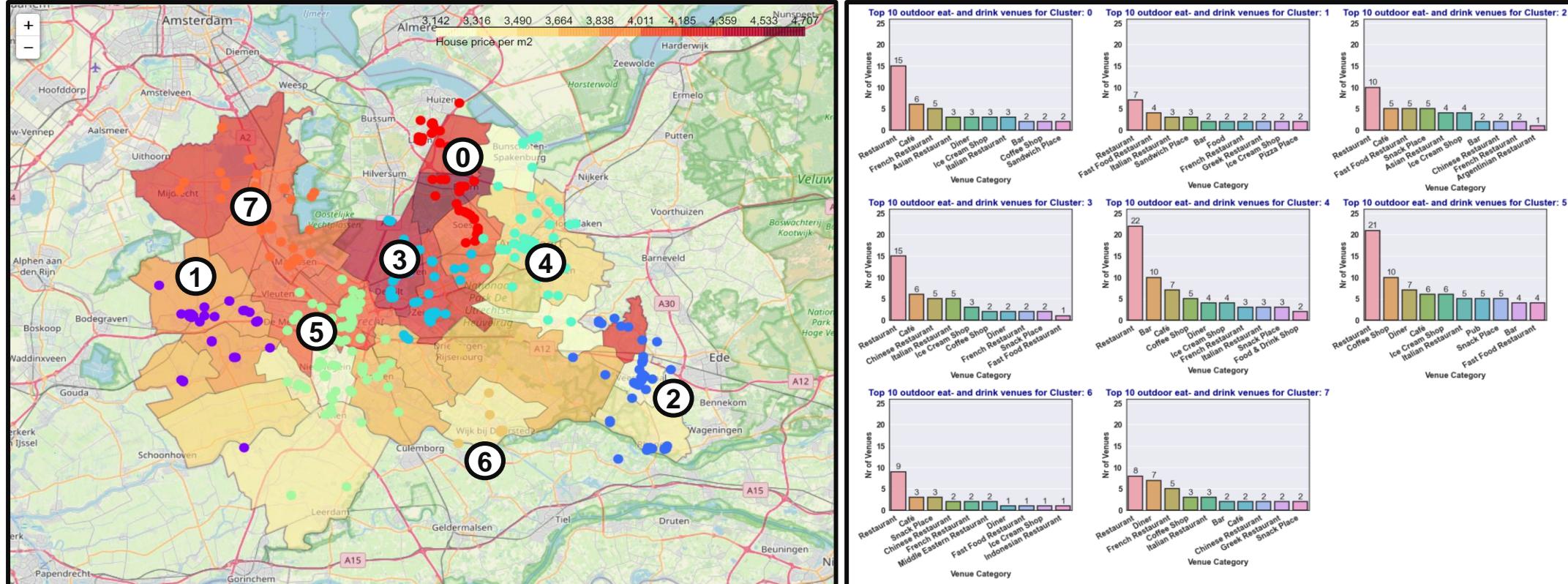


The Elbow Method showing the optimal number of clusters "k"



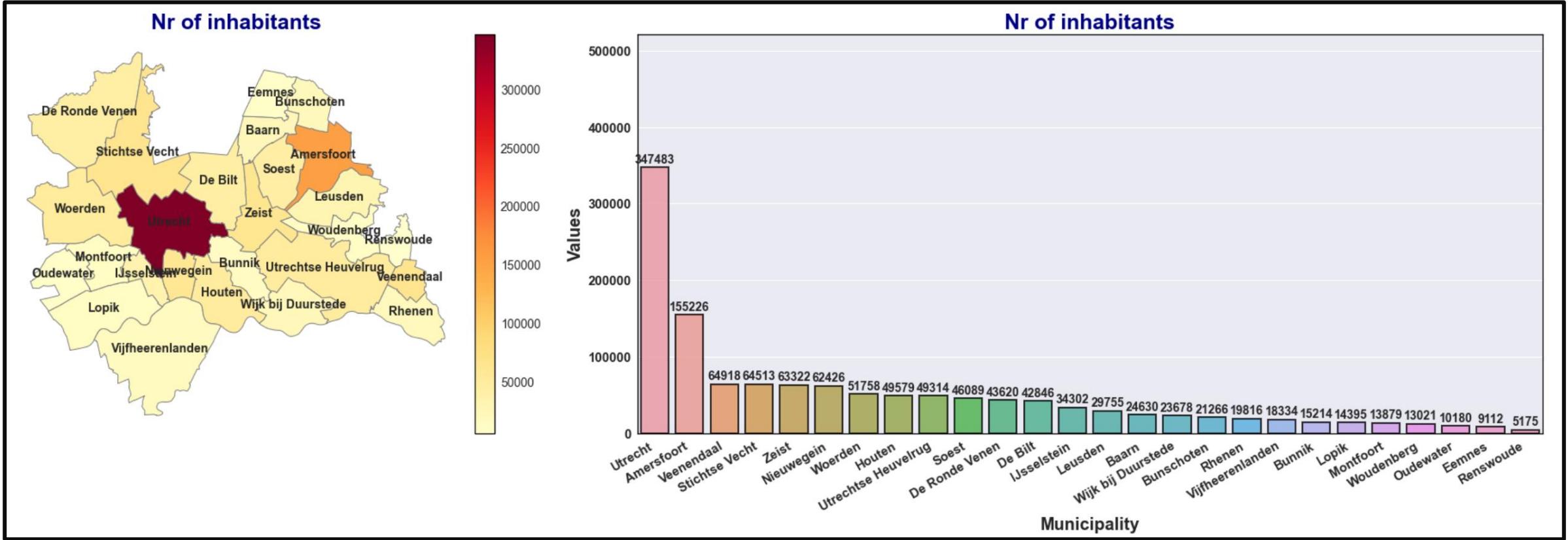
# Methodology and results

## Venues for outdoor eating and drinking venues put on a choropleth map



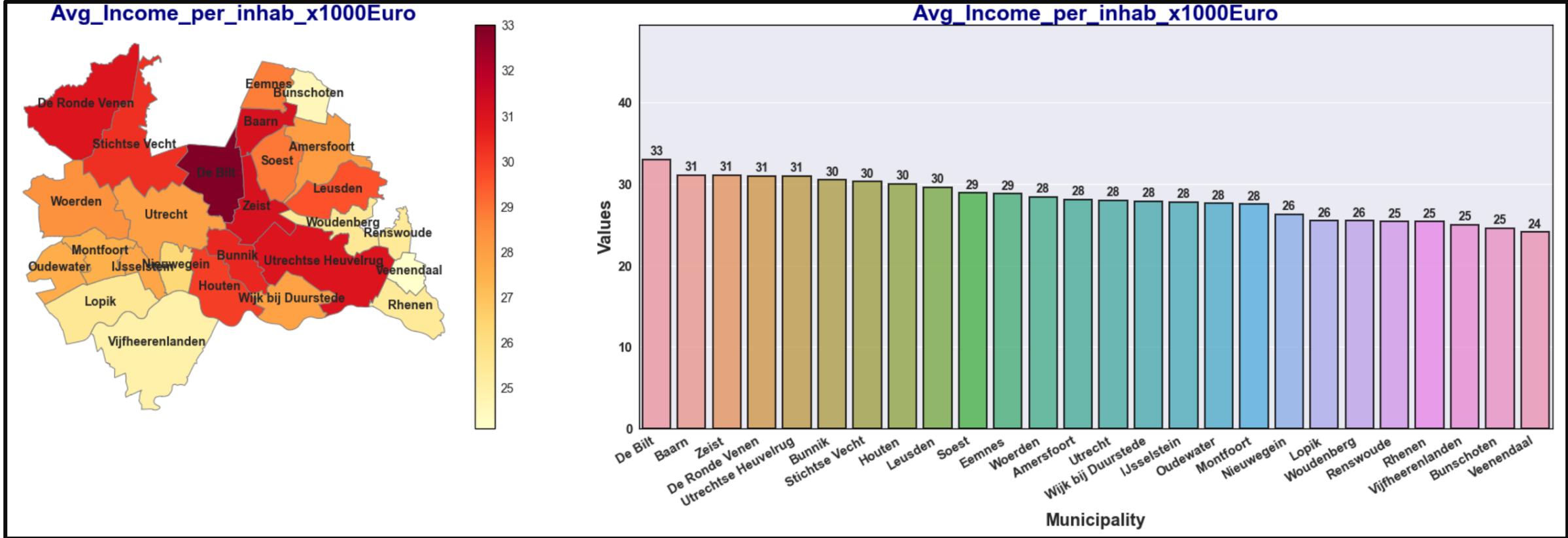
- The choropleth map with outdoor drinking and eating venues is shown above. Similar clusters are formed as before with 8 clusters.
- The amount of venues in cluster 6 has reduced significantly. Only 4 venues are visible on the map now.
- Areas that were earlier already seen as potentially, also seem to be interesting on this map:
  - In cluster 4, there are several eating and drink venues present and house prices in that area are relatively low.
  - In the lower part of cluster 5, house prices are also relatively low, but amount of outdoor eating and drinking facilities is high.

# Methodology and results



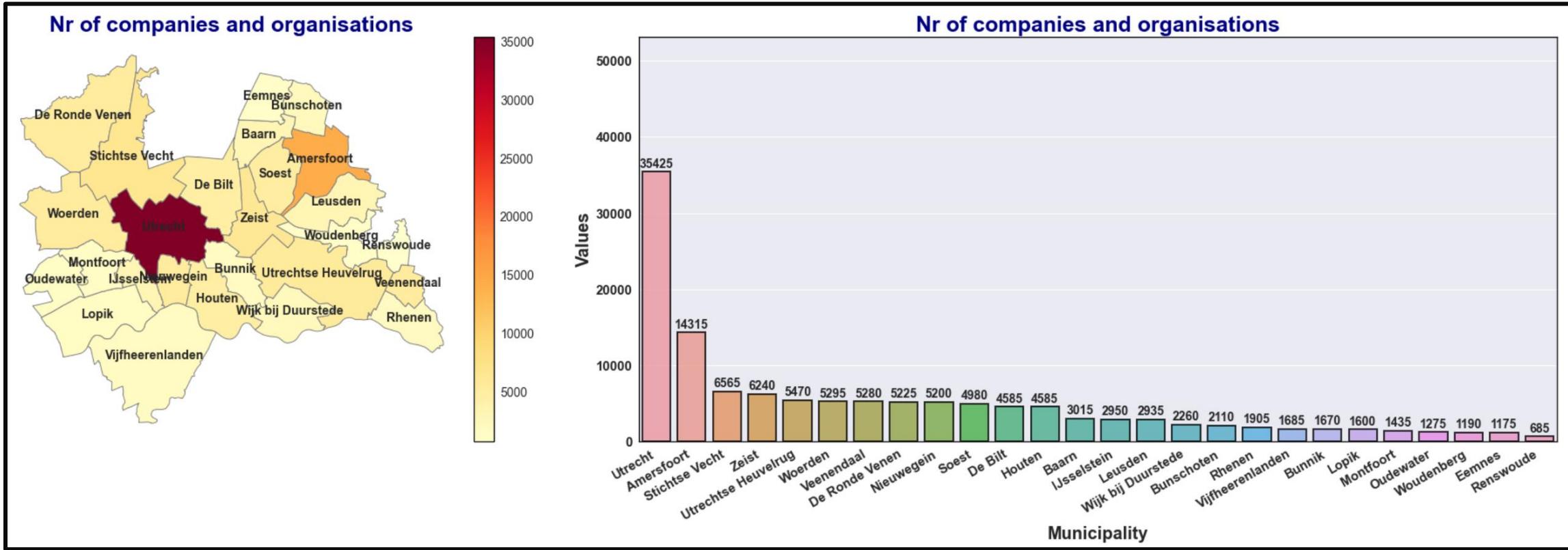
- As significant amount more people live in the 2 biggest cities Utrecht and Amersfoort compared to other municipalities.
- The two cities are popular areas to live in, which could be taken into consideration when selecting an area to invest in property.

# Methodology and results



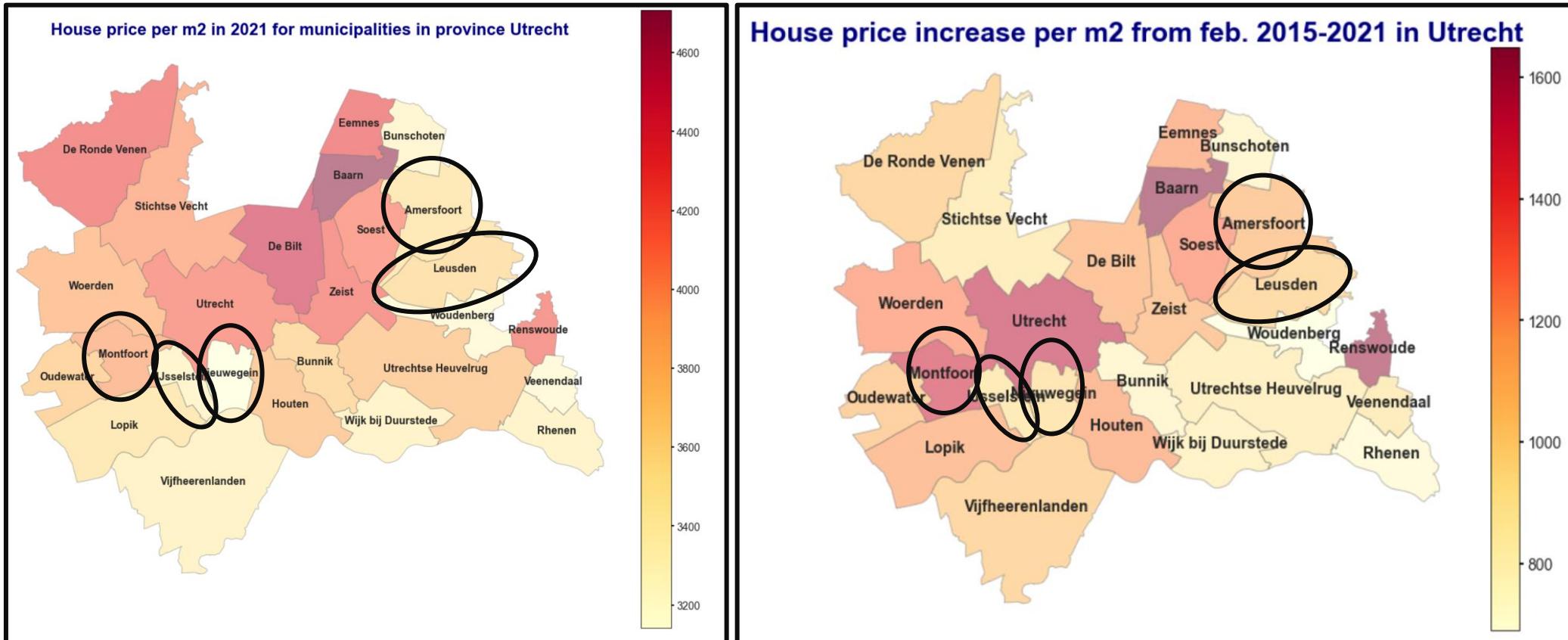
- Average income seems to be generally higher in areas where house prices are also high, with the exception of Amersfoort, Leusden and Utrechtse Heuvelrug. These areas have low house prices but relatively high average income.
- People with higher average income could drive up the house prices in the area. For instance, when people reconstruct their house.

# Methodology and results



- **Utrecht and Amersfoort have a lot of companies and organisations.**
- **This is another reason to look for investment opportunities for places near or in Utrecht or Amersfoort, as employment opportunities are higher.**

# Conclusion

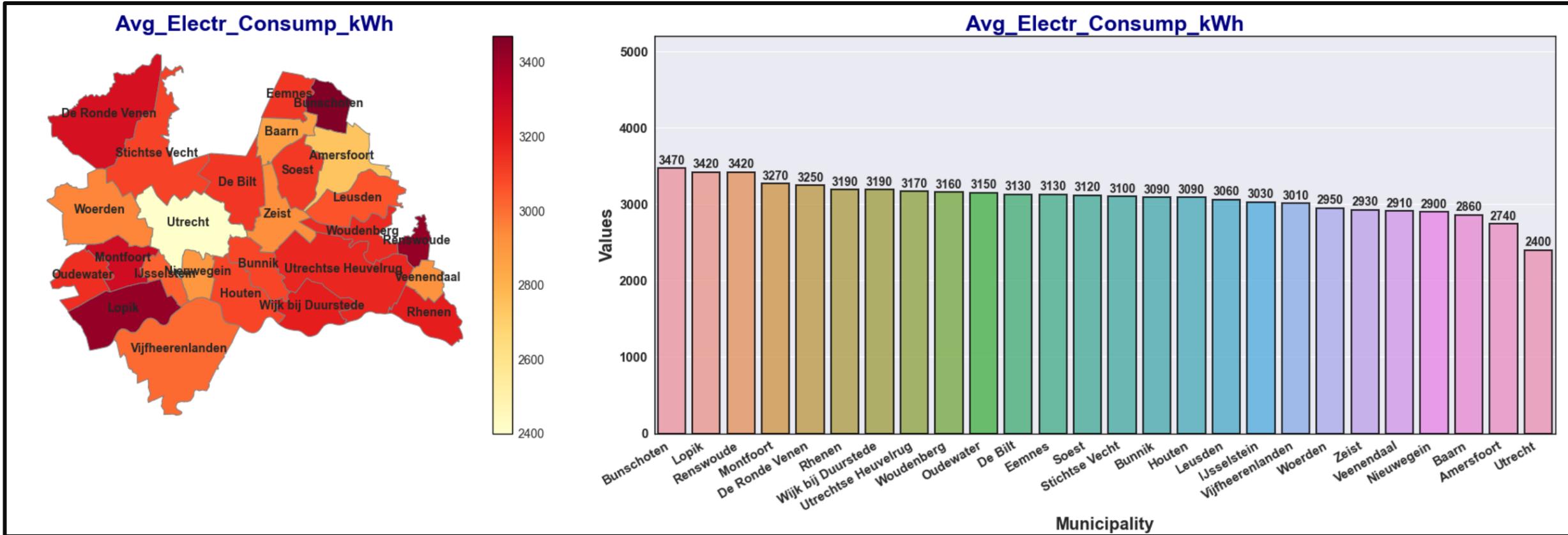


- Putting all the information together, the above encircled municipalities have been selected as potentially interesting areas to invest in real estate, thereby considering the different criteria of the real estate investment firm.
- The encircled municipalities are in areas with lower house prices, have had a stable increase in average house price, have a variety of venues that are close together and are near or in the main cities. Purchasing and reselling houses in these areas will likely yield good profit.

# Remaining visualisations

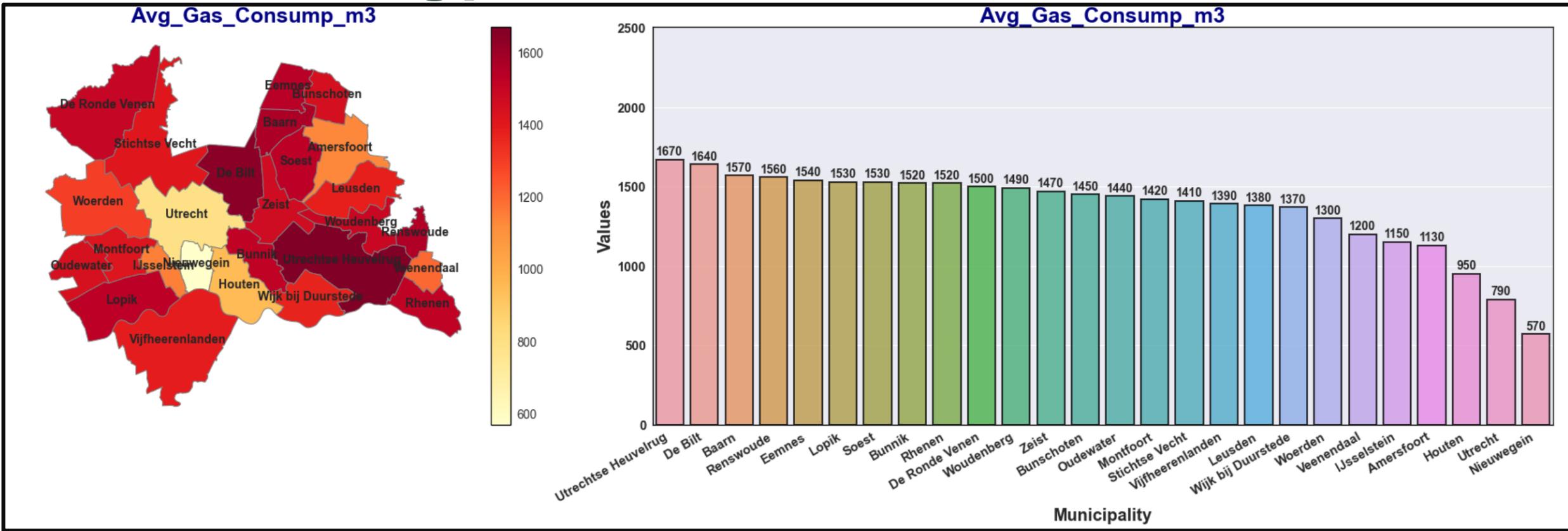
**The following slides show remaining visualisations which where not used in the selection of potential interesting areas to invest in property**

# Methodology and results



- Electricity consumption seems lower in cities Utrecht and Amersfoort. An article on the internet clarified that this is caused by the number of smaller apartments in cities that consume less energy and which causes the average energy consumption for both gas and electricity to be lower than areas outside cities with relatively more free-standing houses that consume more energy.
- The choropleth map turns out to be not very useful for selecting potential interesting areas to invest in property.

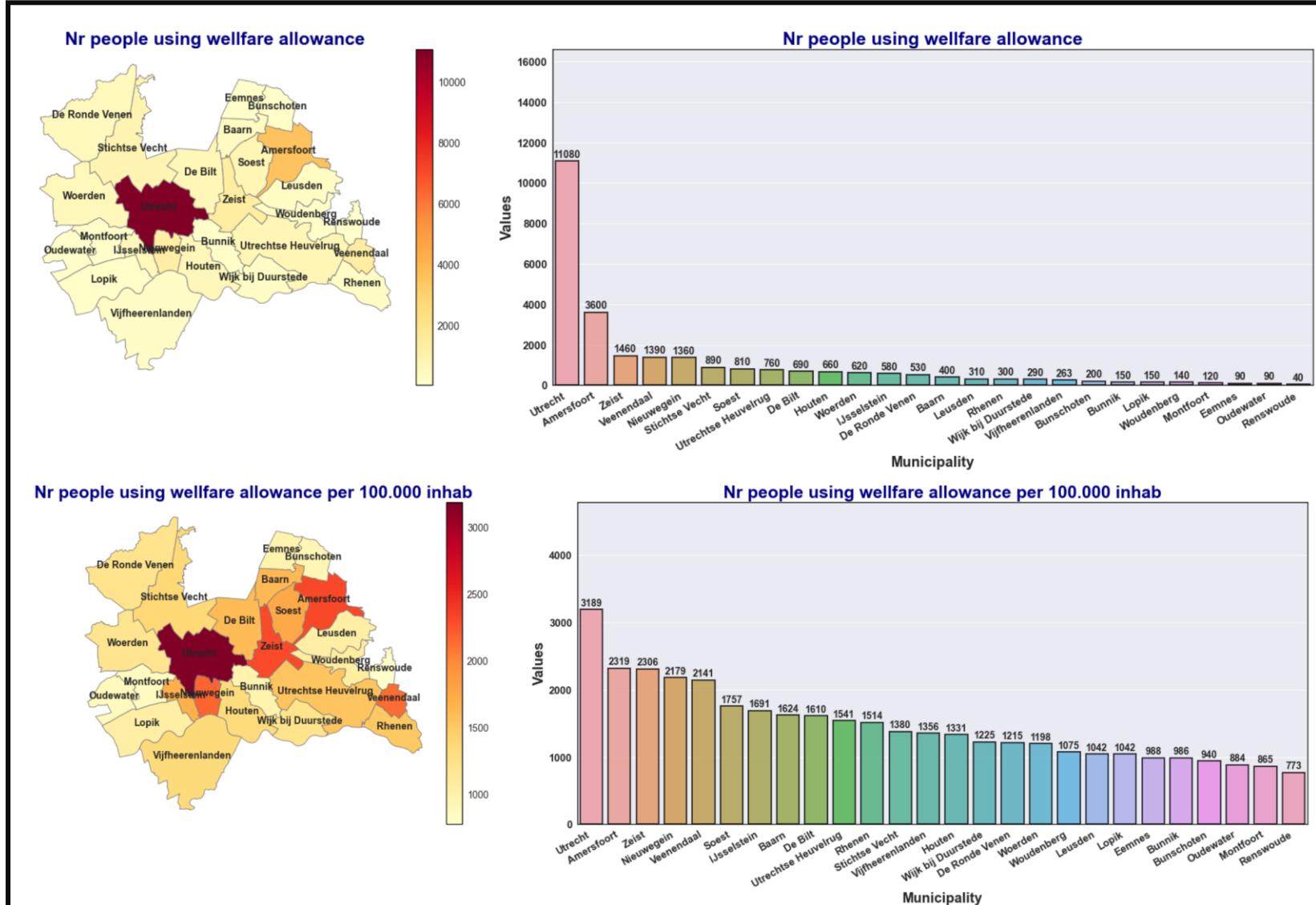
# Methodology and results



- Same comment applies as written down on previous slide.

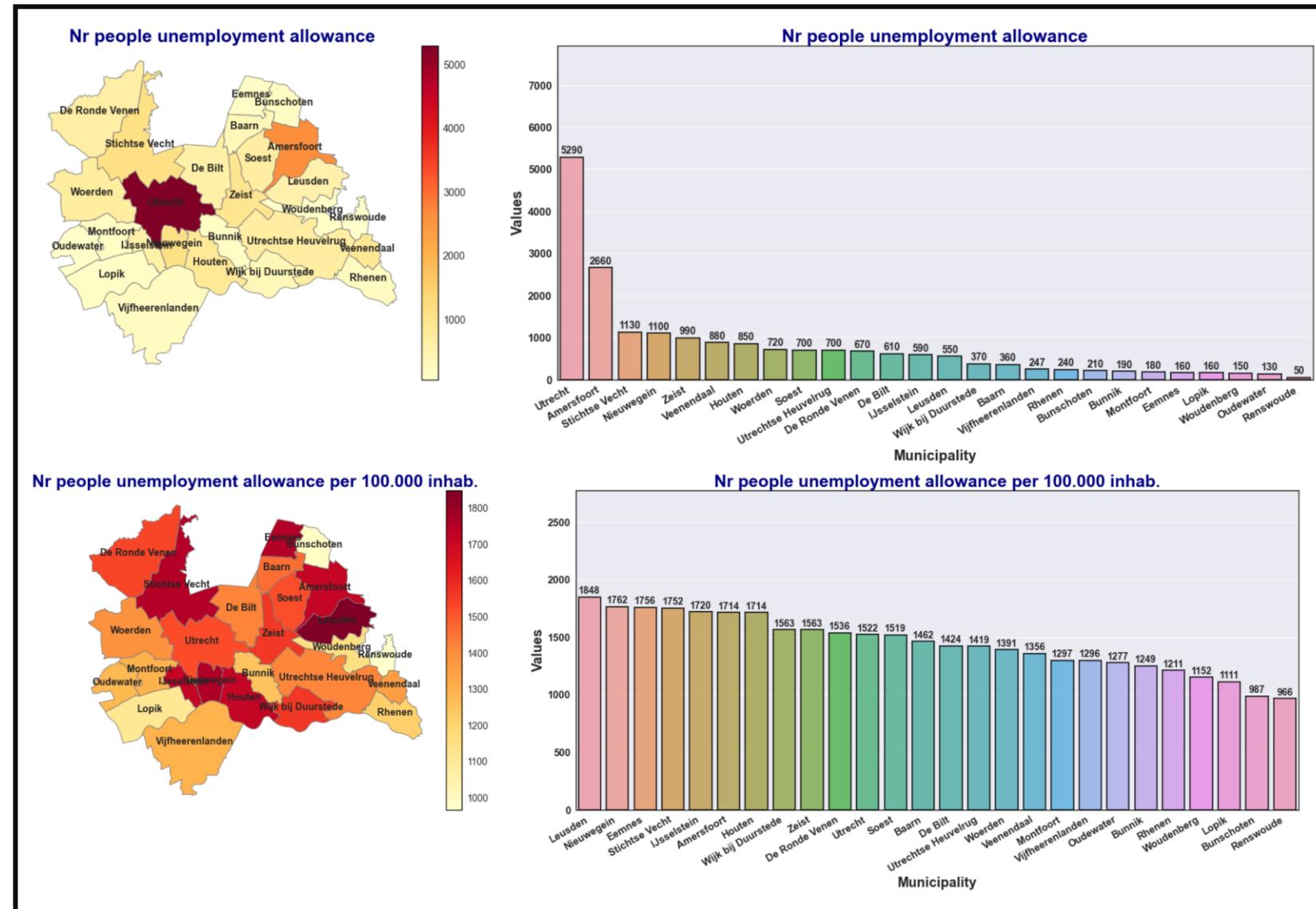
# Methodology

- Looking at total numbers (total count)  
Utrecht and Amersfoort seem to have the highest number of people using welfare allowance (financial support of government on top of income). However – those cities have a lot more inhabitants.
- For comparison the total numbers have been turned into relative numbers, where the total count was divided by the number of inhabitants and multiplied by 100,000.
- Looking at the lower choropleth map, we can see that also on a relative scale, Utrecht and Amersfoort have more people using a welfare allowance. Also in the municipalities Nieuwegein, Zeist and Veenendaal the relative number of people using welfare allowance seems to be higher.

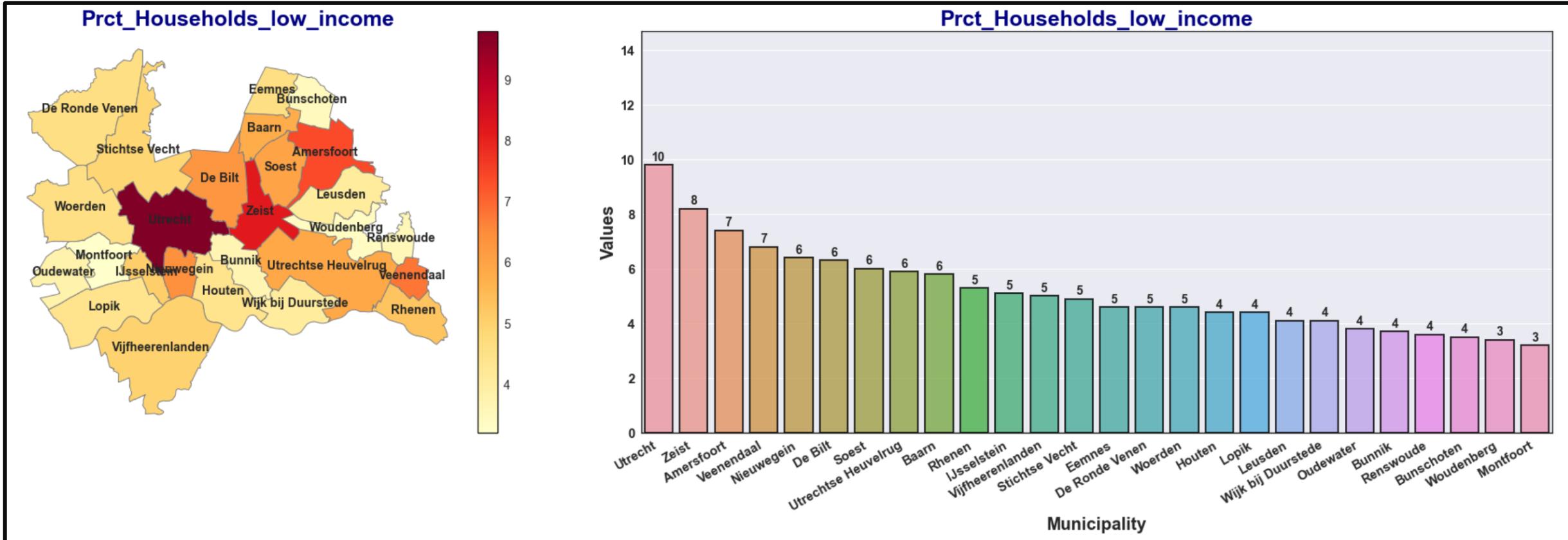


# Methodology

- The relative number (per 100,000 inhabitants) for people using an unemployment allowance seems to be in the higher spectrum for most of the municipalities.
- Areas with the lowest number of inhabitants using unemployment allowances are: Renswoude, Bunschoten, Lopik and Woudenberg.



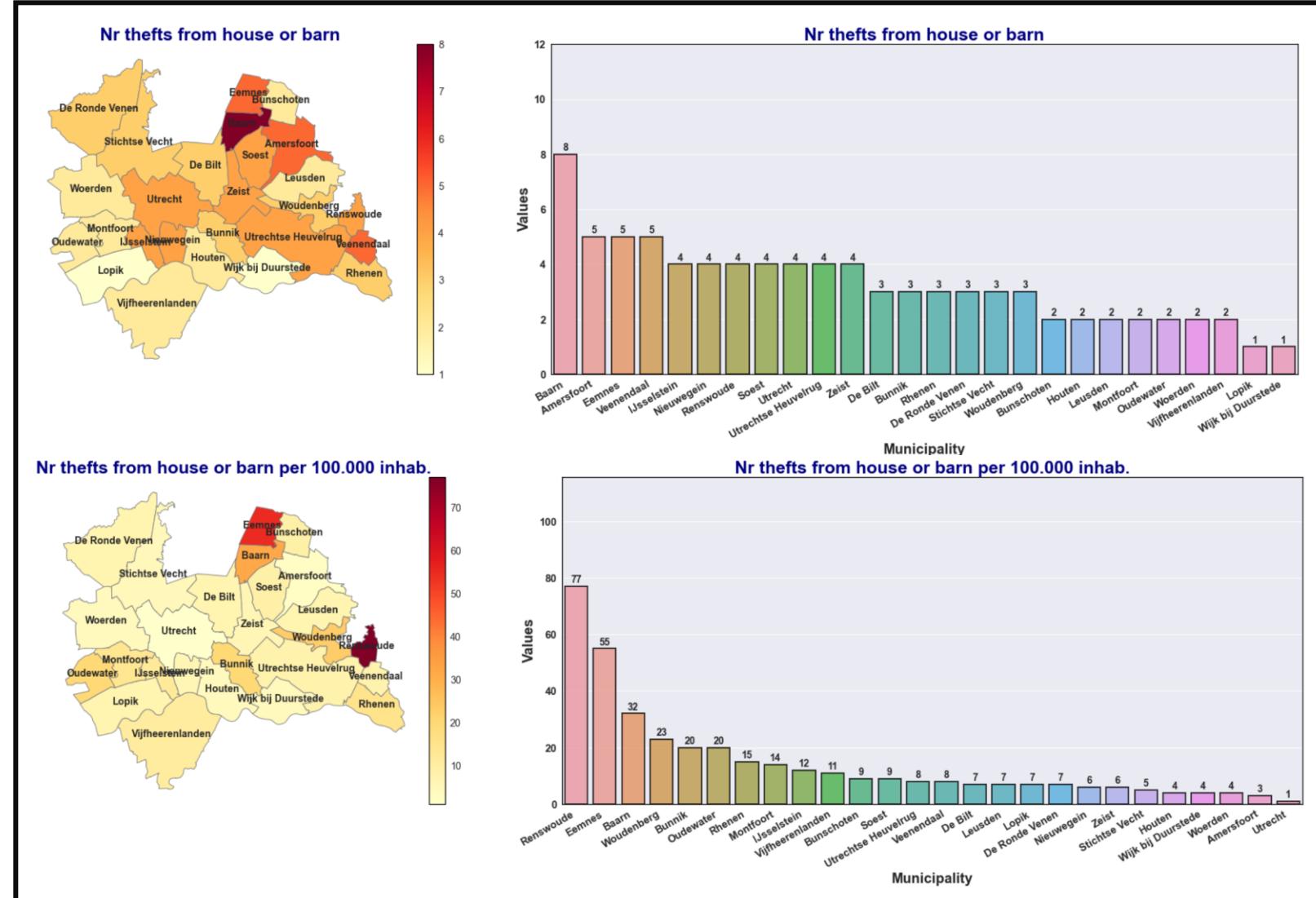
# Methodology



- Earlier we saw that Amersfoort was one of the areas with relatively high average income, but in contrast to that there also seems to be a higher percentage of households with low income.
- Other municipalities that have higher percentages for households with low income are: Utrecht, Zeist and Veenendaal

# Methodology

- Comparing the number of thefts with the amount of inhabitant the amount of registered thefts seem to be very low.
- Looking a total amount of thefts, then Baarn, Amersfoort, Eemnes, Veenendaal and IJsselstein seem to be on top of the list.
- Relative numbers show higher number of thefts for Renswoude, Eemnes, Baarn and Woudenberg.



# Methodology

