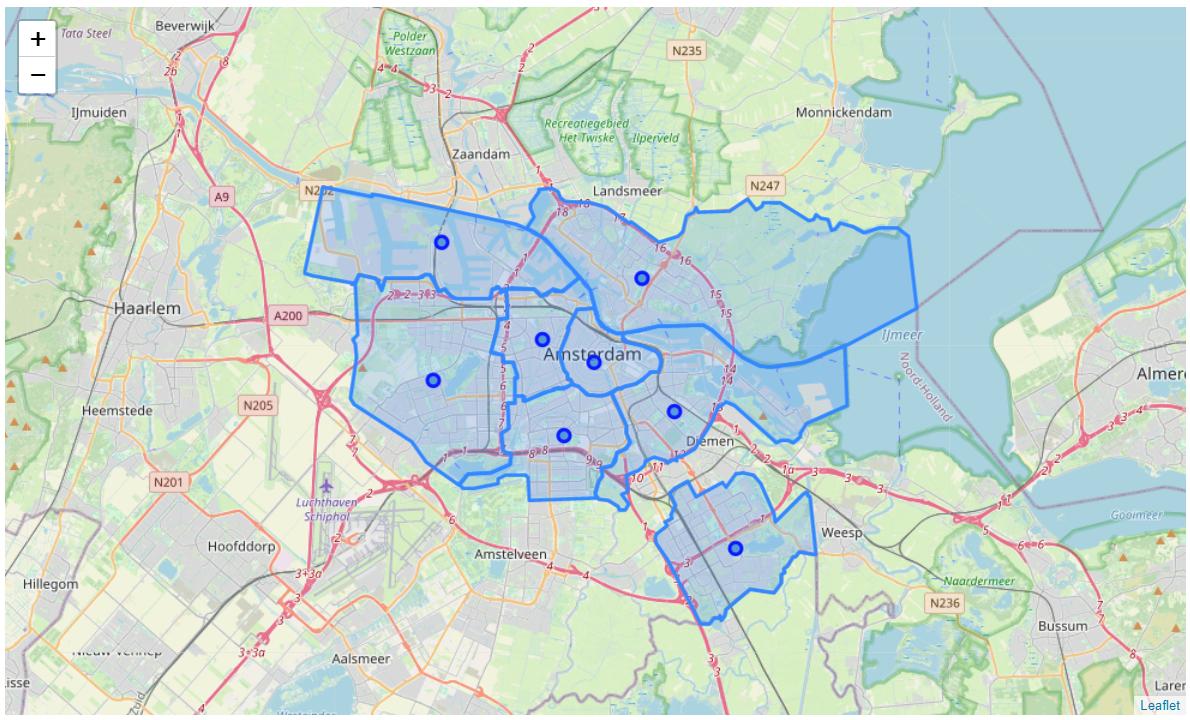
2021

Jojanneke Stoof

Coursera

1-1-2021

Characterisation of City Districts in Amsterdam for Prospective Residents



# Introduction / Business problem

Finding suitable accommodation (to buy or rent) can be quite difficult in an unfamiliar city, certainly when the city is quite large. Finding the perfect accommodation is not only a matter of finding a house of the right size, price, and lay-out, but is also heavily dependent on the neighbourhood where it is located and the amenities nearby. While housing sites may offer a clear description of the house itself and possibly a short description of its neighbourhood it would be much more convenient to first decide on a neighbourhood before starting to look at specific houses. Of course, one can decide to visit the city before starting the actual house hunting process to get a feel for the different neighbourhoods, but this is not always possible or preferred (e.g. due to high travel costs or long travel time, or currently due to travel-restrictions). An alternative might be to ask an estate agent for help, but this can be quite costly and is not always wanted since nowadays it is quite easy to buy or rent a house without the help of an estate agent. A tool which helps house hunters discover which neighbourhoods in their city of choice are best suited to their preferences may reduce the time needed to find a suitable house and may also improve satisfaction with their chosen accommodation.

# Data

To answer this problem I will make use of two data sources: Foursquare and the publicly available datasets provided by the municipality of Amsterdam (data.amsterdam.nl). Where Foursquare provides information on the location and type of venues, data.amsterdam.nl provides information on population statistics (e.g. age distribution, nationality, education, and income) as well as on housing (e.g. living area, and average value of housing).

For statistical purposes Amsterdam can be divided into eight districts (coordinates can be downloaded from: <https://maps.amsterdam.nl/open_geodata/> or <https://api.data.amsterdam.nl/gebieden/stadsdeel/03630000000016/>? ). The venue data from Foursquare will be categorised into six groups:

1. restaurants (including pubs, diners, food trucks, and fast food restaurants)
2. shopping (shops and stores, excluding supermarkets and grocery stores)
3. markets
4. parks (including lakes, trails, and nature preserves)
5. sport (including gyms, pools, and yoga studios)
6. culture (theater, museums, music venues, opera houses, and other art)

These data will be combined with socio-economic data of the eight districts of Amsterdam, as supplied by the municipality of Amsterdam (<https://data.amsterdam.nl/datasets/G5JpqNbhweXZSw/basisbestand-gebieden-amsterdam-bbga/>) on:

* age distribution
* nationality
* highest level of education
* income
* size of living area
* average value of housing

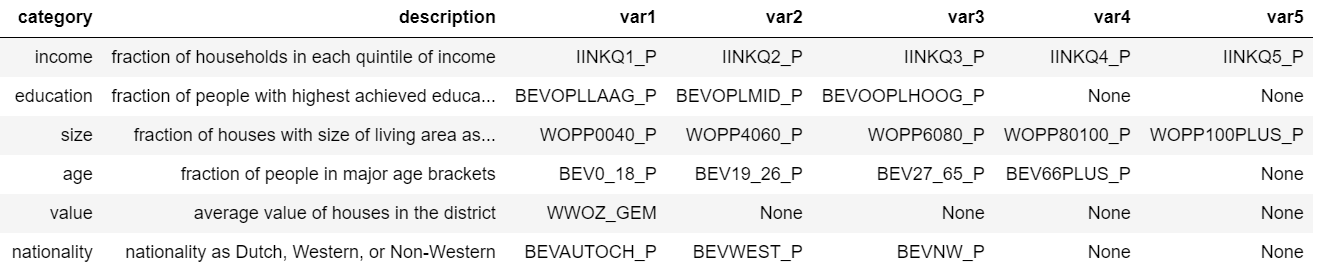
Together, this information will be used to describe the neighbourhoods and make recommendations for prospective house buyers/movers.

# Methodology

### Exploratory data analysis

The data used in this project comes from two sources: the municipality of Amsterdam and Foursquare. While the dataset on the socioeconomic background and housing of people in Amsterdam is very elaborate, not all information is necessary or useful for somebody looking for a new place to live. The first step was therefore to remove all columns from the table that did not pertain to the variables of interest. Additionally, only the data for 2018 were selected for analysis. While for some variables data was also available for the years 2019 and 2020, 2018 was the most recent dataset that was virtually complete. In addition, most variables were given as both a raw number (e.g. number of people in a certain age category) as well as a fraction. Since fractions are easier to interpret, the raw numbers could also be removed from the dataset. The resulting dataframe consisted for 8 rows (i.e. the 8 districts) and 23 columns (i.e. the 6 categories of interest, divided over 23 variables). A summary of these categories/variables can be found in Table 1.

**Table 1. Summary of the 23 variable/6 categories from the Amsterdam dataset.**

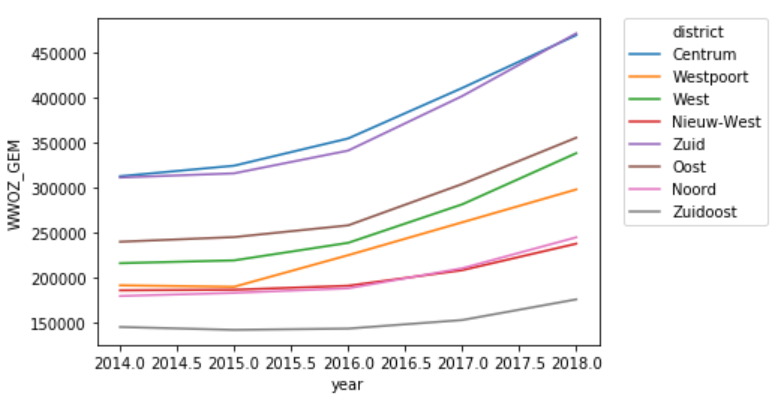
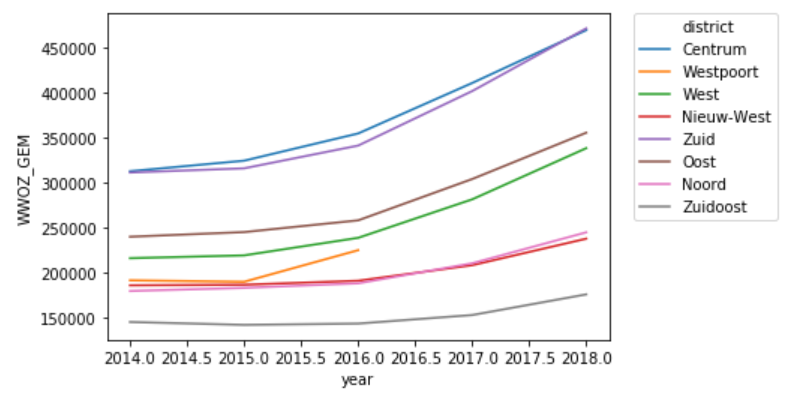


The variables are ordered from low to high (e.g. IINKQ1\_P is the lowest income group and IINKQ5\_P the highest). When there are less than 5 variables per categories the remaining cells are entered as None-type.

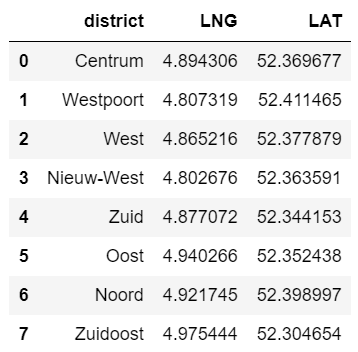
#### Missing data

For Westpoort three variables appear to be missing (IINKQ4\_P, IINKQ5\_P, WWOZ\_GEM). Values for IINKQ4\_P and IINKQ5\_P are given for 2012 (where the values are 7% and 6%, respectively), but are missing for all other years. From this we can conclude that either these fractions (or rather, the raw numbers on which these fractions are based) are not known for all other years, or the numbers are negligible. Since the fractions for 2012 are very low, it was decided to set these values at 0 for any further analysis. Data for the average house value (WWOZ\_GEM) are available from 2014 for all districts but are missing from 2017 onwards for Westpoort (see Figure 1). From 2015 onwards, house values have been rising in all districts although the slope differs slightly per district. To estimate the house values of Westpoort in 2018 the following formula was used:

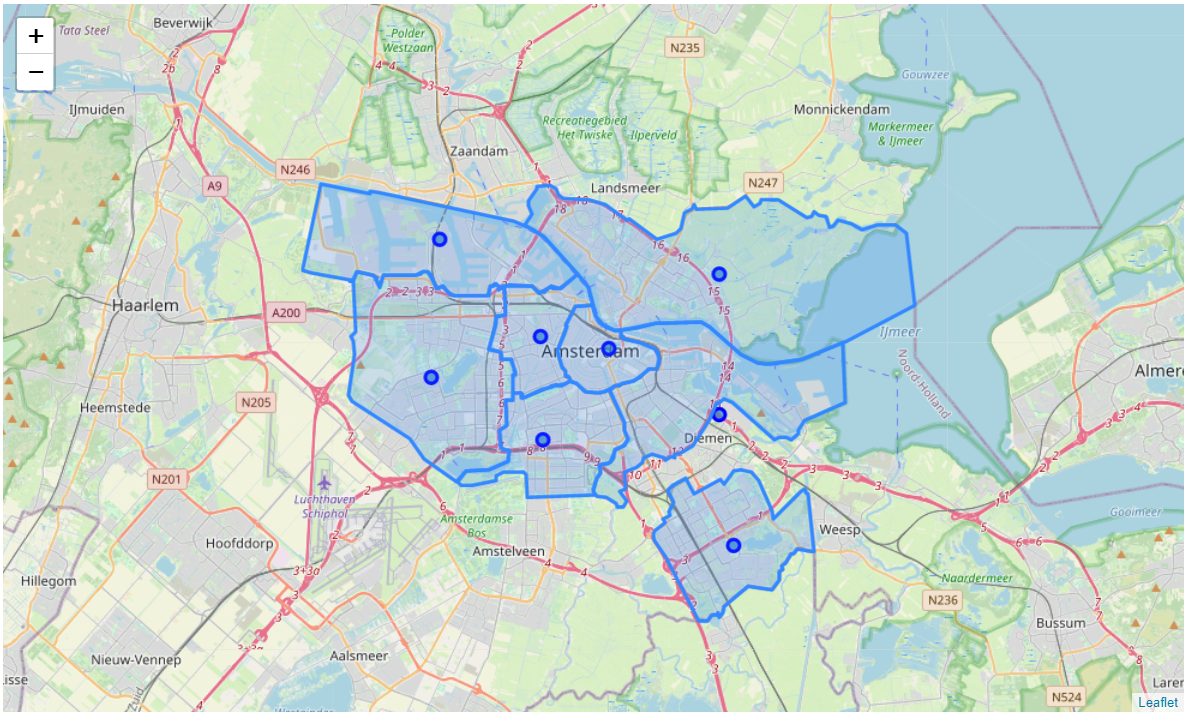
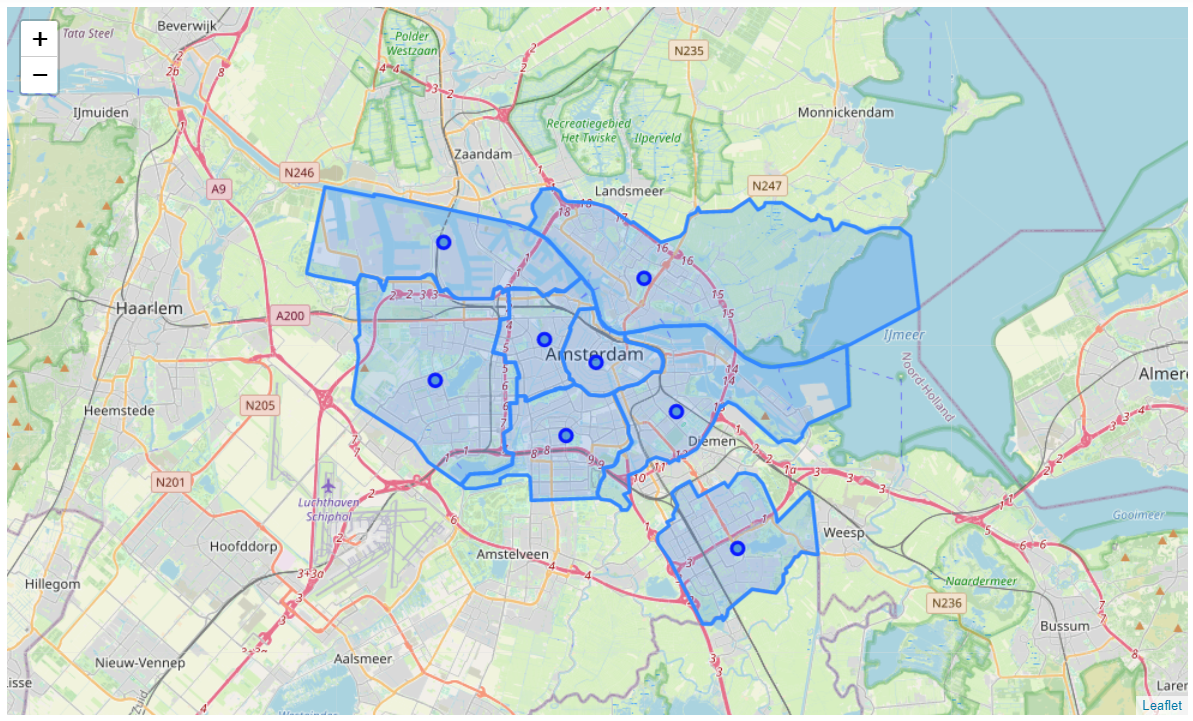
Or, in words, calculate the average ratio of value increase from 2016 to 2018 for the other districts and multiply the house values of Westpoort in 2016 with this ratio to estimate the house values of Westpoort in 2018.



**Figure 1. Average house values districts Amsterdam (2014-2018). Left:** original data. **Right:** original data plus estimation for Westpoort.

Next, the geographical data of the 8 districts, including district centre and district borders, were obtained from [www.data.amsterdam.nl](http://www.data.amsterdam.nl). A basic clean up was performed to get rid of unnecessary data and the column names were translated to English. While this dataset contains the coordinates of each district border in the shape of a list of lists within the table that can be used to plot the district borders in a folium map (after some transformation since the list of lists is saved as a string) it proved to be easier to download a JSON file with the district border coordinates (<https://api.data.amsterdam.nl/gebieden/stadsdeel/03630000000016/>?) and save this as a GeoJSON file. The resulting folium map of Amsterdam can be seen in Figure 2. Each district centre coordinate is given by a blue circle while the districts themselves are marked with blue borders and a blue hue. While each circle is supposed to mark the centre of each district one can see that this is not always the most obvious location. This is especially clear for district Oost where the marker falls outside the district itself due to the horse shoe shape of the district. As the Foursquare call later in this project will make use of these centre district coordinates to draw a circle in which to query the venues present these coordinates were changed to capture a more representative section of each district (Figure 2 right). The new coordinates were selected using Google Maps (Table 2).

**Table 2. Adjusted marker coordinates for the centre of each district.**

**Figure 2. Map of Amsterdam with its eight districts.** **Left:** The blue circle markers indicate the coordinates of each district centre as given by the municipality of Amsterdam. **Right:** Adjusted markers that are more representative of the district centres.

Next, using these adapted coordinates, a Foursquare query was performed for all the districts with a limit of 900 calls and a radius of 1000m. The calls were grouped by district and categorised based on venue type. Since the query may result in different venues at different timepoints a flexible list was used to group the venues into categories to deal with queries in which a venue-type is not present. The lists used can be found in Table 3. The ‘restaurant’ category contains all places where one can sit down to eat, as well as food trucks and take-away places. However, shops where one can buy food have been excluded. Examples of excluded venues include bakeries, chocolate shops, and pastry shops. The ‘shopping’ category includes all shops, with the exception of supermarkets and other food stores. The ‘market’ category is self-explanatory and includes all markets. The ‘parks’ category is not limited to parks alone, but also includes other forms of nature that can be found in a city. The ‘sport’ category includes all outdoor and indoor sport facilities. Lastly, the ‘culture’ category includes all venues that are related to culture, ranging from churches to museums and from opera houses to nightclubs.

**Table 3. List of venues per category**

|  |  |
| --- | --- |
| Category | List of venues |
| Restaurant | 'Argentinian Restaurant', 'Asian Restaurant', 'BBQ Joint', 'Beer Bar', 'Beer Garden', 'Bistro', 'Breakfast Spot', 'Burger Joint', 'Café', 'Caribbean Restaurant', 'Chinese Restaurant', 'Cocktail Bar', 'Coffee Shop', 'Comfort Food Restaurant', 'Creperie', 'Diner', 'Dutch Restaurant', 'Ethiopian Restaurant', 'Falafel Restaurant', 'Fast Food Restaurant', 'Food Court', 'Food Truck', 'French Restaurant', 'Fried Chicken Joint', 'Friterie', 'Gastropub', 'Greek Restaurant', 'Indian Restaurant', 'Indonesian Restaurant', 'Italian Restaurant', 'Japanese Restaurant', 'Juice Bar', 'Lebanese Restaurant', 'Mediterranean Restaurant', 'Middle Eastern Restaurant', 'Pizza Place', 'Pub', 'Restaurant', 'Salad Place', 'Sandwich Place', 'Scandinavian Restaurant', 'Seafood Restaurant', 'Snack Place', 'Steakhouse', 'Sushi Restaurant', 'Thai Restaurant', 'Turkish Restaurant', 'Vegetarian / Vegan Restaurant', 'Vietnamese Restaurant' |
| Shopping | 'Arts & Crafts Store', 'Bookstore', 'Bridal Shop', 'Clothing Store', 'Department Store', 'Discount Store', 'Drugstore', 'Electronics Store', 'Flower Shop', 'Furniture / Home Store', 'Gift Shop', 'Paper / Office Supplies Store', 'Pharmacy', 'Record Shop', 'Shopping Mall', 'Shopping Plaza', 'Smoke Shop', 'Sporting Goods Shop', 'Toy / Game Store', 'Women\'s Store' |
| Markets | 'Farmers Market', 'Fish Market', 'Market' |
| Parks | 'Lake', 'Park', ‘Trail’, ‘Nature Preserve’ |
| Sport | 'Arcade', 'Athletics & Sports', 'Baseball Field', 'Bowling Alley', 'Dance Studio', 'Golf Course', 'Gym', 'Gym / Fitness Center', 'Gym Pool', 'Gymnastics Gym', 'Playground', 'Pool', 'Soccer Field', 'Spa', 'Sports Club', 'Stadium', 'Tennis Court', 'Yoga Studio' |
| Culture | 'Art Museum', 'Arts & Entertainment', 'Church', 'Comedy Club', 'History Museum', 'Movie Theater', 'Museum', 'Music School', 'Music Venue', 'Nightclub', 'Opera House', 'Public Art', 'Theater' |

### Analysis

For the main analysis the six categories from the socio-economical and housing dataset were examined by means of bar charts and hierarchical clustering. The hierarchical clustering was used which districts were most similar. Combined with the bar charts, these clusters could then be given a name based on the most distinctive features. For age, nationality, education level, and size of living area a cut-off of 15 was used. For the average house value a cut-off of 80,000 was used. Additionally, clustering was also performed for the venue data, using a cut-off of 60. Visualisation of the venue data was performed by means of a heatmap.

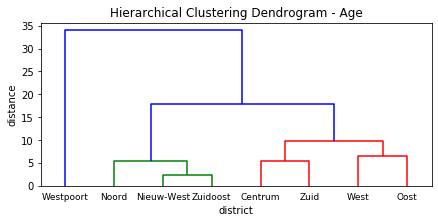
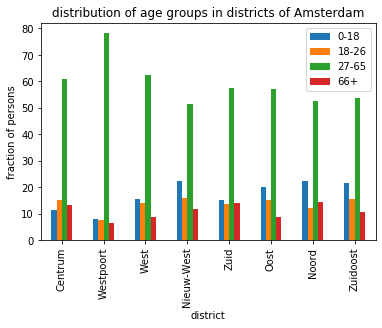
# Results

### Distribution of age groups in the districts of Amsterdam

The age groups are based on the four main age groups in society:

* 0-18: children
* 18-26: young adults/professionals
* 27-65: working population
* 66+: pensioners

The 27-65 group is the largest group in every district (Figure 3). This makes sense since it is also the largest age group in the general population of the Netherlands. Within Amsterdam, three clusters can be described:

* **Working**: high fraction of people within the working age combined with low fractions of people in the other three age groups. This cluster only consists solely of the district Westpoort.
* **Young families**: relatively high fraction of children. This cluster consists of the districts Noord, Nieuw-West, and Zuidoost.
* ******Mixed**: a relatively even distribution of children, young adults, and pensioners. This clusters consists of the districts Centrum, Zuid, West, and Oost.

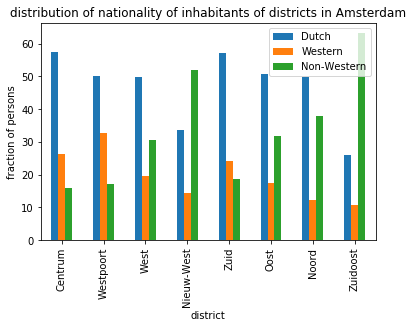
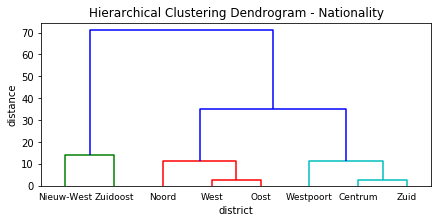
**Figure 3. Distribution of age group in the districts of Amsterdam.** **Left**: bar chart the fraction of people in each age groups versus the districts. **Right**: hierarchical clustering results in three categories: working (blue), young families (green), mixed (red).

### Distribution of nationality of inhabitants of the districts of Amsterdam

Nationality was defined by the study based on the migration background of the parents of the inhabitants and not on official nationality of the inhabitants themselves (there is probably a better word than nationality for this in English, but if there is I don’t know it). Inhabitants of whom one or both parents were born outside the Netherlands were defined as being either of Western or non-Western parentage. In case both of the parents were born abroad the nationality of the mother was leading. Western countries were defined as all countries in Europe (excluding Turkey and the Netherlands), North-America, Oceania, Indonesia, and Japan. Non-Western countries were defined as all countries in Africa, Latin-America, and Asia (excluding Indonesia, Japan, Suriname, Netherlands Antilles, Turkey, and Morocco).

The bar chart immediately highlights the international community of Amsterdam; in none of the districts is more than 60% of the inhabitants of Dutch parentage and there are even districts where this number drops to less than 40% (Figure 4). Overall, the distribution of Western and non-Western nationalities is comparable, though within the districts there are stark differences.

Hierarchical clustering found three clusters:

* Non-Western: Over half of the inhabitants are over non-Western parentage and the remaining inhabitants are primarily of Dutch parentage. This cluster consists of the districts Nieuw-West and Zuidoost.
* Mixed: Dutch is the predominant nationality, but the difference between the fraction of inhabitants of Western and non-Western nationality is small. This cluster consists of the districts Noord, West, and Oost.
* Dutch/Western: while Dutch is again the predominant nationality, there are clearly more inhabitants of Western parentage than of non-Western parentage.

**Figure 4. Distribution of the nationality of the inhabitants of the districts in Amsterdam, based on the country of birth of their parents.** **Left:** bar chart of the nationality versus the districts. **Right**: hierarchical clustering results in three clusters: non-Western (blue), mixed (green), Dutch/Western (red)

### Distribution of education level among inhabitants of the districts in Amsterdam

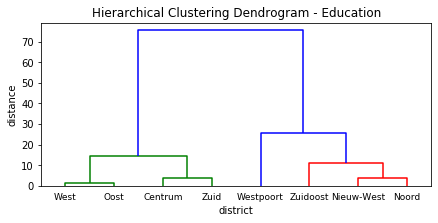
The Dutch education system is split in three stages: primary school (age 4-12), secondary school (age 12-16/18, depending on the school level), and tertiary education. Tertiary education consists either of vocational training (e.g. hairdressers, farmers, car mechanics), higher vocational training (HBO, e.g. teachers, nurses, laboratory technicians), and university (e.g. lawyers, doctors, scientists). The study defines three levels of education, based on the highest level of education a person has completed:

* Low: primary school or lowest level of secondary school
* Middle: highest two levels of secondary school or vocational training
* High: higher vocational training or university

The average distribution of education level of the entire country is more or less evenly distributed between the three levels: 28% low, 39% middle, and 31% high (source: <https://www.onderwijsincijfers.nl/kengetallen/onderwijs-algemeen/hoogst-behaald-opleidingsniveau>).

Hierarchical clustering found three clusters (Figure 5):

* High: characterised by a large fraction (>50%) of highly educated inhabitants. The districts in this cluster are Centrum, West, Zuid, and Oost.
* Middle: characterised by a large fraction (>50%) of inhabitants with a middle-level education. This cluster only contains the district Westpoort.
* Mixed: characterised by a relatively even distribution of all education levels, this cluster is most similar to the general population of the Netherlands. This clusters contains the districts Nieuw-West, Noord, and Zuidoost.



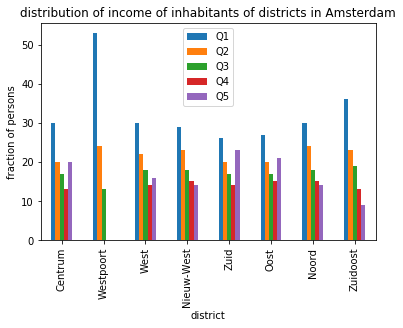
**Figure 5. Distribution of education levels (categorised as low, middle, and high) of inhabitants of the districts in Amsterdam**. **Left**: bar chart. **Right**: hierarchical clustering results in three clusters: high (green), middle (blue), and mixed (red).

### Distribution of income levels of inhabitants of the districts in Amsterdam

The income levels in the Amsterdam dataset were defined as the spendable income, meaning the gross income minus social contributions and other affordable transfers (e.g. alimentation) and income taxes. Data was only presented for districts where there were more than 10 households representing that quantile. The missing data for Westpoort is therefore most likely due to a very low number of households with a high income.

The income ranges for each quantile are as follows:

* Q1: <€20,294
* Q2: €20,294 – €29,015
* Q3: €29,015 – €40,516
* Q4: €40,516 – €56,512
* Q5: >€56,512

Due to the missing data for Westpoort no hierarchical clustering could be performed. For each district, most inhabitants fall in the first or second quantile, and the fraction of inhabitants decreases as the income rises (Figure 6). Interestingly, there are three districts (Centrum, Zuid, and Oost) where the fraction of inhabitants with the highest income is considerably larger than the fractions on Q3 and Q4. 

**Figure 6. Distribution of income of inhabitants of districts in Amsterdam**. Income is grouped in quantiles, with Q1 being the lowest income group and Q5 the highest.

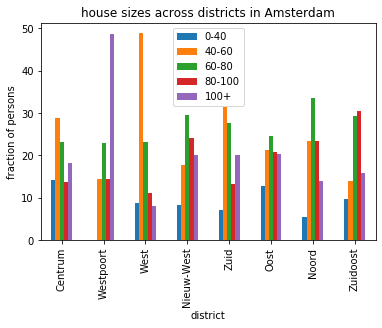
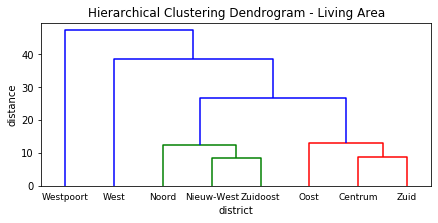
### Distribution of living area (house) size in the districts of Amsterdam

The house sizes were categorised in five groups in the dataset:

* <40m2
* 40m2 to 60m2
* 60m2 to 80m2
* 80m2 to 100m2
* >100m2

The majority of the houses fall in the three mid-range groups, but there are some striking differences between districts (Figure 7). Hierarchical clustering found four clusters:

* Large: more than 45% of the houses are larger than 100m2. The only district in this cluster is Westpoort.
* Small: more than 45% of the houses are in 40m2 to 60m2 size range. The only district in this cluster is West.
* Mid-range: more than 70% of the houses are in the combined 40m2 to 100m2 size range. The districts Noord, Nieuw-West, and Zuidoost are in this cluster.
* Mixed: a relatively even distribution of all size ranges. The districts Oost, Centrum, and Zuid are in this cluster.

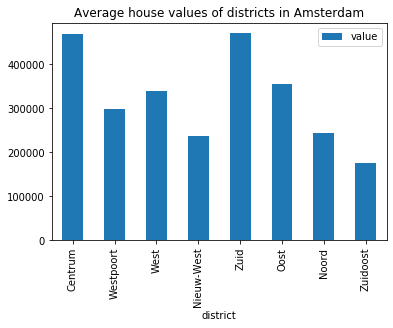
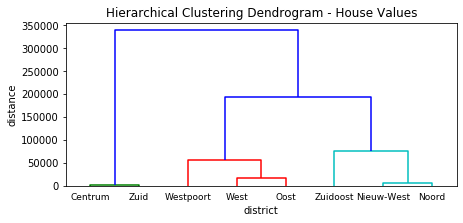


**Figure 7. Distribution of house size across districts of Amsterdam. Left**: bar chart. **Right:** hierarchical clustering reveals four clusters: large (blue left), small (blue right), mid-range (green), mixed (red).

### Distribution of average house values in the districts of Amsterdam

House values were given as averages of the entire district and range between €175,000 and €475,000 (Figure 8). Hierarchical clustering found three clusters:

* High: characterised by an average house value of over 400k euros. The districts in this cluster are Centrum and Zuid.
* Mid-range: characterised by an average house value ranging between 250k and 400k euros. This cluster contains the districts Westpoort, West, and Oost.
* Low: characterised by an average house value below 250k euros. This cluster contains the districts Zuidoost, Nieuw-West, and Noord.



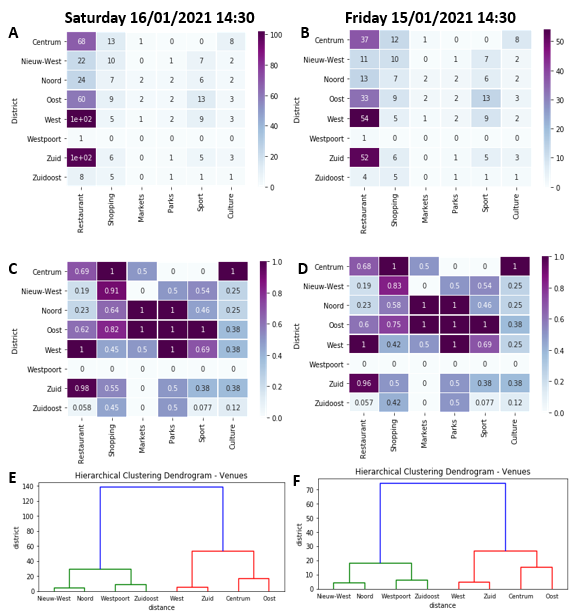
**Figure 8 Average house values of districts in Amsterdam. Left**: bar chart. **Right**: hierarchical clustering reveals three clusters: high (green), mid-range (red), and low (mint).

### Distribution of venues in the districts of Amsterdam

Foursquare returned 509 venues in all of Amsterdam (on Saturday 16/01/2021 14:30), distributed over 152 venue types. Of the 509 venues, 326 venues could be grouped in of the six predefined categories (Figure 9). Most venues were grouped in the ‘Restaurant’ category, at a distance followed by ‘Shopping’ and ‘Sport’. However, the Foursquare data differs per query/time of query as can be seen in Figure 9. Despite the considerable difference between Figure 9A and 9B, these differences are nihil when normalised and do not affect the clustering.

Hierarchical clustering reveals two main clusters:

* Few venues: approximately half (or less) the number of venues in the other cluster. This cluster contains the districts Nieuw-West, Noord, Westpoort, and Zuidoost.
* Many venues: This cluster contains the districts West, Zuid, Centrum, and Oost.



**Figure 9. Distribution of venues in the districts of Amsterdam, based on Foursquare location data. A, B:** heatmap of the raw number of venues per category and district. **C, D:** heatmap of normalised venue data per category and district. **E,F:** hierarchical clustering based on the venue data finds two clusters: few venues (green), many venues (red).

# Discussion

Foursquare venue data and socio-economic/geographical data provided by the municipality of Amsterdam can be used to inform potential house buyers which districts of Amsterdam are most suited to their needs and preferences. However, these data sources also have several drawbacks.

Firstly, the socio-economic data is linked to the districts and not to individual locations/households within the district. While this makes the data anonymous and therefore publicly accessible, it also prevents more detailed analysis. For example, it is not possible to perform correlation analysis between distribution of house value and house size to investigate whether more expensive (valuable) houses are also larger or whether this is linked to the location of the house. Should one want to investigate these kind of relations further, it is also possible to obtain the data cluster per neighbourhood (of which there are 100) and might be better suited for correlation analysis.

Secondly, the Foursquare data is very changeable and might not be an accurate reflection of the venues in Amsterdam. While the query was made for 900 venues, only approximately 500 were returned, of which the most were restaurants. Of course there are many restaurants in Amsterdam (1,150 restaurants and 1,515 cafés and bars) there are more shops (5,517). It therefore seems that the number of shops is underrepresented while the number of markets (6 in both queries and 32 in reality) is highly overrepresented (source: <https://amsterdam.org/nl/feiten-en-cijfers.php>) .

# Conclusion

Based on the previous analyses we can make the following characterisations of each district.

### Centrum

Centrum is the city centre and is therefore unsurprisingly characterised by a high number of restaurants, shops, and culture-related venues. The house sizes are mixed and vary from very small to very large, though house prices are generally high. The population is generally highly educated and consists of people of all ages. Most inhabitants are of either Dutch or Western parentage.

### Nieuw-West

Nieuw-West is one of the newer districts as its name suggests (‘nieuw’ = ‘new’) and is characterised by the presence of many young families with children. The district has a large non-Western population with a mixed level of education similar to that of the entire population of the Netherlands. House values are low compared to the rest of the city, though this does not mean that the houses are small since the majority of the houses are in the mid-range. The number of restaurants, shops, and culture-related venues is relatively low, though there are multiple locations for sport.

### Noord

Noord is the most northern district of Amsterdam and a large area of the district consists of farmland. The city part of the district is characterised by mid-range size houses which are relatively cheap. The population is sports large fractions of people of Dutch, Western, and non-Western parentage and contains many young families with children. Education levels are mixed. There are few venues.

### Oost

Oost is comparable to Centrum in terms of the number of restaurants and shopping venues, but it contains considerably more venues for sport. Just as in Centrum, the house size vary from very small to very large though the house values are lower in the mid-range. The population is mixed with large fractions of people of Dutch, Western, and non-Western parentage with a mixed education level. The population spans people of all age groups.

### West

The West district is located adjacent to the city centre and could be seen as its continuation. There are a lot of restaurants, shops, and culture-related venues. However, compared to the city centre, housing is less expensive (mid-range) though overall smaller than in the city centre. Most inhabitants are highly educated and are both of Dutch and international parentage. The population is also diverse in terms of age containing substantial groups of children and pensioners.

### Westpoort

Westpoort is located at the edge of Amsterdam and is characterised by a lot of industry which is reflected by the lack of venues in this district. House prices are in the mid-range, but house size are generally large. This district contains most people of the working population (80%), of which a large fraction has a mid-level education. Inhabitants are mainly of Dutch or Western parentage.

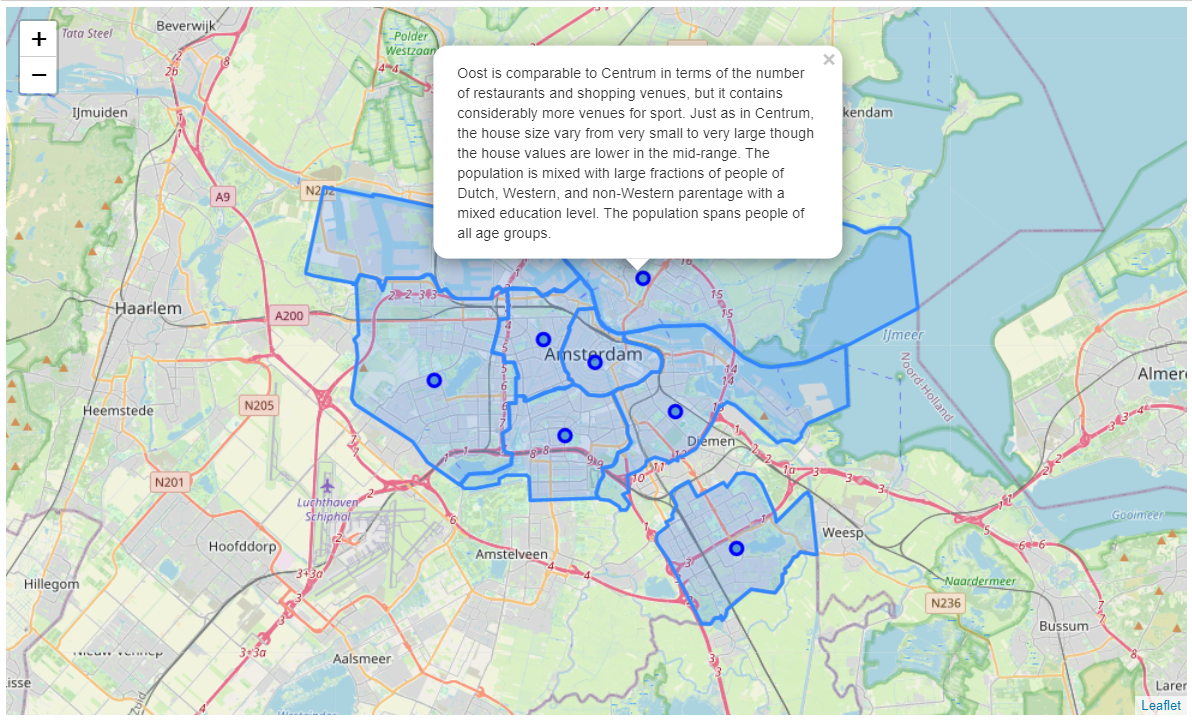
### Zuid

Zuid is known as the financial district of Amsterdam. With a swift connection to the major airport of the Netherlands, the presence of the campus of one of the two universities in the city, as well as many restaurants and culture-related venues it is not surprising that this is an expensive place to live. Similarly, it is unsurprising that most inhabitants have a high level of education. House sizes range from very small to very large and offer therefore space for people of all age groups. The majority of the population is of either Dutch or Western parentage.

### Zuidoost

Zuidoost is separated from the main city by the town Diemen. This district hosts one of the city’s universities but boasts few other venues. However, the house values are low while the house size are in the mid-range making this an affordable yet attractive place to live. Most inhabitants are of non-Western parentage and there is a large fraction of young families with children.

### Final remarks

The descriptions of the districts above have been added to an interactive map where the district borders and district centres have been marked (Figure 10). Clicking on the marker for a district centre results in a pop-up screen with the description of the district, thus providing a quick and easy overview of the different districts.

**Figure 10. Map of Amsterdam with eight districts and pop-ups with a description of each district.**