

The Battle of Neighborhoods

Finding the best area in Amsterdam to for a place to live for international assignees

Date: April 2, 2020

Author: Ronald de Lange

1. INTRODUCTION	2
1.1 BACKGROUND	2
1.2 PROBLEM	2
2. DATA ACQUISITION AND PREPROCESSING.....	3
2.1 DATA SOURCES USED	3
2.2 DATA ABOUT AMSTERDAM NEIGHBORHOODS FROM THE CITY OF AMSTERDAM.....	3
2.3 AMSTERDAM SAFETY DATA.....	4
2.4 LIST OF AMSTERDAM NEIGHBORHOODS	4
2.5 DATA CLEANING AND CONSOLIDATING.....	4
2.6 FOURSQUARE API	5
3. METHODOLOGY.....	6
3.1 EXPLORATORY ANALYSIS.....	6
3.2 CLUSTER ANALYSIS	7
4. RESULTS	9
4.1 BRINGING IT ALL TOGETHER BY INDEXING RESULTS	9
5. QUESTIONNAIRE.....	9
5.1 SETUP	9
5.2 SCORE FOR ATMOSPHERE	9
5.3 SCORE FOR OTHER CHARACTERISTICS	10
5.4 FINAL SCORE	10
6. FOOD FOR DISCUSSION	11
6.1 LIMITATIONS OF THE FOURSQUARE DATA	11
6.2 LIMITATION OF CLUSTERING	11
6.3 LIMITATION OF CHARACTERISTICS	11
6.4 RECOMMENDATION.....	11
7. CONCLUSION	12
7.1 OUTCOME FOR OUR RANDOM ASSIGNEE	12
7.2 OUTCOME FOR OUR RANDOM ASSIGNEE	13

1. Introduction

1.1 Background

I work in Amsterdam as a manager for an international company. Once every while we have international assignments within our organization. I often get the question from the assignee what the best place would be to look for a house or apartment in Amsterdam.

Most of the time I propose to jointly discuss the possibilities, with the map of Amsterdam in front of us.

1.2 Problem

It is not easy to find a place to live in Amsterdam. In order to be able to focus on a specific area, it would be beneficial to understand the characteristics of all the areas. This will enable drilling down on the areas which suits most. As part of this project I explore different neighborhoods of Amsterdam and find the best area to look for a place to live for international assignees.

From the assignee perspective, a lot of factors come into play when finding the best accommodation. This study will focus on the following characteristics of the neighborhoods in Amsterdam:

- the safety
- the general atmosphere
- the house density
- the number of households with children in the neighborhood
- the population density in the neighborhood

In order to make sure that every individual assignee will get a specific advice I included a questionnaire in the notebook where the assignee can rate the importance of specific characteristics to him/her. I will use this input to weight the different characteristics into a final score per neighborhoods. Based on the final scores, a top 5 of neighborhoods is advised. for this specific international assignee.

2. Data Acquisition and Preprocessing

2.1 Data Sources used

This project will use a couple of sources of data:

- *Dataset of Amsterdam Neighborhood characteristics*
- *Dataset with safety index for Amsterdam Neighborhoods*
- *List of Neighborhoods in Amsterdam*
- *Foursquare data via the API*

2.2 Data about Amsterdam neighborhoods from the city of Amsterdam

The foundation of the data used in this project is the dataset made available by the city of Amsterdam containing information about the different neighborhoods:

- *Data from the city of Amsterdam, with different characteristics of the Neighborhood*
- *source: Amsterdam Datastore*
- *url:*
<https://api.data.amsterdam.nl/dcatd/datasets/G5JpqNbhweXZSw/purls/NlraIEidgfbWGw>

The overview below is depicting the first 5 rows of this dataset. The full dataset consists of 805 columns (fields) and 19161 rows.

niveau	aanam	SD	sdnaam	gebiedcode15	gebiedcodenaam	gebiednaam	GSD	jaar	BEVTOTAAL	BEV_PROG2015	...	SK017_KWETS34
niveau												
1	Amsterdam	S	Amsterdam	STAD		Amsterdam	Amsterdam	NaN	2040.0	NaN	NaN	...
2	Stadsdelen	A	A Centrum		A	A Centrum	Centrum	NaN	2040.0	NaN	NaN	...
2	Stadsdelen	B	B Westpoort		B	B Westpoort	Westpoort	NaN	2040.0	NaN	NaN	...
2	Stadsdelen	E	E West		E	E West	West	NaN	2040.0	NaN	NaN	...
2	Stadsdelen	F	F Nieuw-West		F	F Nieuw-West	Nieuw-West	NaN	2040.0	NaN	NaN	...

For further analysis all the information not needed for this project have been filtered out. This resulted in a table consisting of 8 relevant columns (fields) and 100 rows as shown below.

gebiedcode15	sdnaam	gebiednaam	jaar	BEVTOTAAL	BEVDICHT	BEVHHMKIND_P	WDICHT
A00	A Centrum	Burgwallen-Oude Zijde	2020.0	4465.0	12814.0	7.9	8524.0
A01	A Centrum	Burgwallen-Nieuwe Zijde	2020.0	4134.0	7374.0	6.9	5282.0
A02	A Centrum	Grachtengordel-West	2020.0	6440.0	14382.0	12.7	9621.0
A03	A Centrum	Grachtengordel-Zuid	2020.0	5436.0	10457.0	14.0	6708.0
A04	A Centrum	Nieuwmarkt/Lastage	2020.0	9703.0	13969.0	13.2	8952.0

2.3 Amsterdam safety data

The next piece of data used is the dataset containing the so-called safety index per neighborhood in Amsterdam. This data is also pulled from the formal data made available by the City of Amsterdam. The safety index is depicting the safety of a neighborhood based on for instance the amount of particular crimes. The lower the index, the more save the neighborhood is. A neighborhood with an index around 100 is considered to be relative save.

The data is available for a couple of years, but latest data is from 2018. I decided to use the 2018 data for this project.

		Wijken	2014	2015	2016	2017	2018
gebiedcode15							
A00	A00	Burgwallen-Oude Zijde	76.0	50.0	82.0	112.0	74
A01	A01	Burgwallen-Nieuwe Zijde	77.0	78.0	70.0	67.0	85
A02	A02	Grachtengordel-West	35.0	56.0	38.0	42.0	70
A03	A03	Grachtengordel-Zuid	77.0	57.0	40.0	73.0	65
A04	A04	Nieuwmarkt/Lastage	64.0	71.0	75.0	68.0	68

2.4 List of Amsterdam neighborhoods

From the formal datasets from the city of Amsterdam I was able to obtain a table with the different neighborhoods including the geographical data. This can be used to plot the data on maps as part of the analysis.

	LNG	LAT
NeighbourhoodID		
A00	4.896898	52.372560
A01	4.893655	52.374241
A02	4.887073	52.373251
A03	4.892358	52.365285
A04	4.904794	52.371762

2.5 Data cleaning and consolidating

The above 3 mentioned datasets have been cleaned and consolidated based on the unique neighborhood ID, which was available in all of the datasets. As a final step I renamed the column headers into meaningful headers.

	Borough	Neighbourhood	Year	Population_Total	Population_Density	HouseHolds_with_Children_Percentage	House_Density	SAFETY_INDI
NeighbourhoodID								
A00	Centrum	Burgwallen-Oude Zijde	2020.0	4465.0	12814.0	7.9	8524.0	74
A01	Centrum	Burgwallen-Nieuwe Zijde	2020.0	4134.0	7374.0	6.9	5282.0	85
A02	Centrum	Grachtengordel-West	2020.0	6440.0	14382.0	12.7	9621.0	70
A03	Centrum	Grachtengordel-Zuid	2020.0	5436.0	10457.0	14.0	6708.0	65
A04	Centrum	Nieuwmarkt/Lastage	2020.0	9703.0	13969.0	13.2	8952.0	68

2.6 Foursquare API

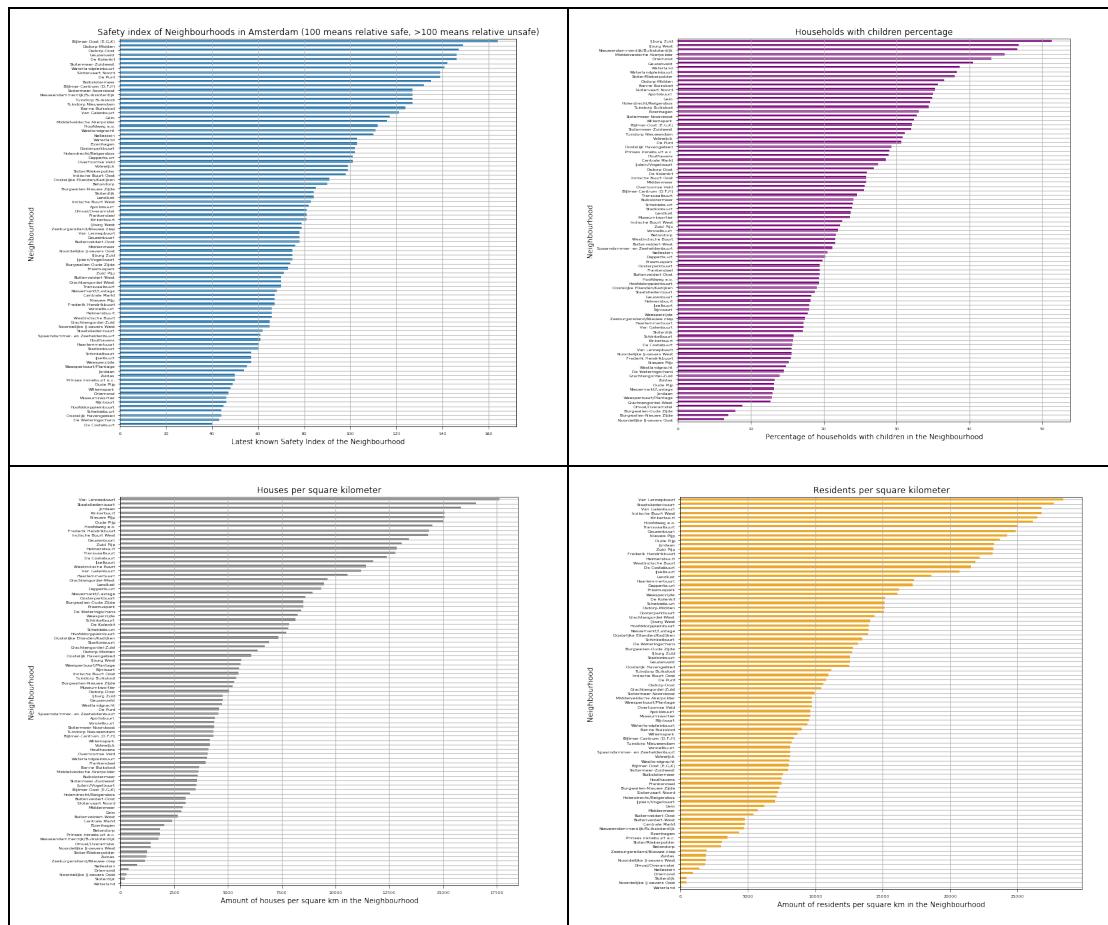
Finally, Foursquare API was used to call the top 50 popular venues in each neighborhood. This was done using the ‘explore’ function of requesting URL. The acquired data is shown below and consist of 2567 rows and 7 columns.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Burgwallen-Oude Zijde	52.37256	4.896898	Rosalia's Menagerie	52.371678	4.899174	Cocktail Bar
1	Burgwallen-Oude Zijde	52.37256	4.896898	De Koffieschenkerij	52.374043	4.898427	Coffee Shop
2	Burgwallen-Oude Zijde	52.37256	4.896898	Sofitel Legend The Grand Amsterdam	52.371093	4.895410	Hotel
3	Burgwallen-Oude Zijde	52.37256	4.896898	Wynand Fockink	52.372301	4.895253	Liquor Store
4	Burgwallen-Oude Zijde	52.37256	4.896898	Dam	52.372824	4.893702	Plaza

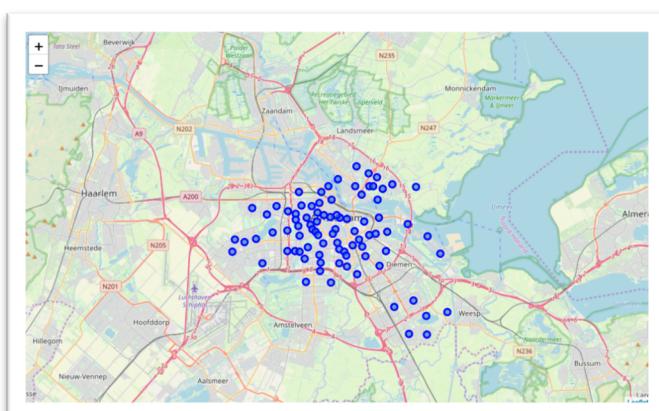
3. Methodology

3.1 Exploratory Analysis

In order to explore the data some visualizations were created to show how different neighborhoods 'score' for specific characteristics like safety, percentage of children, house density and population density.



As you can see in above overviews, different neighborhoods have very different characteristics. In the final analysis these will be weighed based on the input from the questionnaire. In below picture you will find an overview of all neighborhoods plotted on the map of Amsterdam:



3.2 Cluster Analysis

Afterwards, K-means clustering was conducted in order to group the neighborhoods according to what venues they have using Foursquare data, in order to depict the ‘atmosphere’ of each neighborhood.

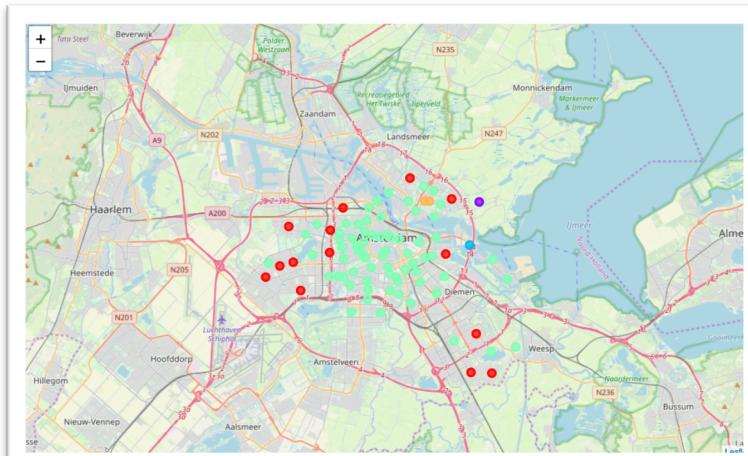
As the first step of cluster analysis, one hot encoding was conducted to give binary values to each venue categories.

Neighbourhood	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Atl:
0 Burgwallen-Oude Zijde	0	0	0	0	0	0	0	0	0	0	0	0	0
1 Burgwallen-Oude Zijde	0	0	0	0	0	0	0	0	0	0	0	0	0
2 Burgwallen-Oude Zijde	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Burgwallen-Oude Zijde	0	0	0	0	0	0	0	0	0	0	0	0	0
4 Burgwallen-Oude Zijde	0	0	0	0	0	0	0	0	0	0	0	0	0

Then, the data was grouped by neighborhood names to find out how many venues of each category exist in the neighborhood within the top 50 venues. The category counts were altered to frequency of how often the category appears among others. This leads to a list of most common venue categories in each neighborhood as follows.

Neighbourhood	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts Crafts Store	Enterta
0 Apollobuurt	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
1 Banne Buiksloot	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
2 Betondorp	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
3 Bijlmer-Centrum (D,F,H)	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
4 Bijlmer-Oost (E,G,K)	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
5 Buikslotermeer	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
6 Buitenveldert-Oost	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.055556	0.00	0.000000	0.
7 Buitenveldert-West	0.00	0.00	0.00	0.00	0.00	0.000000	0.000	0.000000	0.00	0.000000	0.
8 Burgwallen-Nieuwe Zijde	0.00	0.00	0.00	0.00	0.00	0.020000	0.020	0.020000	0.02	0.000000	0.

Based on the venue categories, K-means clustering was conducted to group the boroughs into 5 different clusters based on their similarity. The color dots below represent different clusters.



After observing each cluster and the characteristics they possess, names have been given for each cluster that best depicts their characteristics.

- Cluster 0: Busy area (Coffee shops, Clothing stores, Nightclubs)

NeighbourhoodID	Borough	Households_with_Children_Percentage	House_Density	SAFETY_INDEX	LNG	LAT	Cluster_Labels	Common_Venue	Common_Venue	Common_Venue	Common_Venue
E36	E West	17.1	204.0	84.0	4.853234	52.389607	0	Nightclub	Park	Bar	Cafe
E39	E West	25.9	7853.0	146.0	4.841405	52.376981	0	Park	Restaurant	Bar	Cafe
F78	F Nieuw-West	40.5	4746.0	146.0	4.802849	52.378967	0	Bus Stop	Supermarket		
F81	F Nieuw-West	26.9	5045.0	147.0	4.806869	52.358827	0	Clothing Store	Turkish Restaurant		
F82	F Nieuw-West	36.5	6379.0	149.0	4.794732	52.356743	0	Turkish Restaurant	Park		

- Cluster 1: Healthy area

NeighbourhoodID	Borough	Households_with_Children_Percentage	House_Density	SAFETY_INDEX	LNG	LAT	Cluster_Labels	1st Most Common_Venue	2nd Most Common_Venue	3rd Most Common_Venue
N73	N Noord	38.7	37.0	103.0	4.979187	52.392644	1	Bike Trail	Athletics & Sports	Zo Exit

*! Cluster 2: Harbor area
amsterdam_merged.loc[amsterdam_merged['Cluster_Labels'] == 2, amsterdam_merged.columns[0] + list(range(5, amsterdam_merged.shape[1]))]*

- Cluster 2: Harbor area

NeighbourhoodID	Borough	Households_with_Children_Percentage	House_Density	SAFETY_INDEX	LNG	LAT	Cluster_Labels	1st Most Common_Venue	2nd Most Common_Venue	3rd Most Common_Venue
M34	M Oost	17.4	1141.0	79.0	4.969712	52.368609	2	Harbor / Marina	Gas Station	Café

- Cluster 3: Lively area (Pubs, Bars, Hotels)

NeighbourhoodID	Borough	Households_with_Children_Percentage	House_Density	SAFETY_INDEX	LNG	LAT	Cluster_Labels	1st Most Common_Venue	2nd Most Common_Venue
A00	A Centrum	7.9	8524.0	74.0	4.896989	52.372560	3	Bar	Hotel
A01	A Centrum	6.9	5282.0	85.0	4.893655	52.374241	3	Bar	Coffee Shop
A02	A Centrum	12.7	9621.0	70.0	4.887073	52.373251	3	Hotel	Café
A03	A Centrum	14.0	6708.0	65.0	4.892358	52.365285	3	Italian Restaurant	Hotel
A04	A Centrum	13.2	8952.0	68.0	4.904794	52.371762	3	Bar	Hotel

- Cluster 4: Quiet area

NeighbourhoodID	Borough	Households_with_Children_Percentage	House_Density	SAFETY_INDEX	LNG	LAT	Cluster_Labels	1st Most Common_Venue	2nd Most Common_Venue	3rd Most Common_Venue
N63	N Noord	34.4	5407.0	127.0	4.929079	52.393208	4	Campground	Ice Cream Shop	Pharm
N64	N Noord	46.6	1788.0	127.0	4.932702	52.393161	4	Bar	Tennis Court	Fi

4. Results

4.1 Bringing it all together by indexing results

In order to be able to validate a neighborhood against the wishes of the person looking for a place to live, the differentiating characteristics should be normalized in order to be able to weight them by the specifics being looked for by the international assignee. In the next step we give every characteristic an index by rating them against the results of the other neighborhoods. (We do this for all characteristics except for atmosphere).

Neighbourhood	SAFETY_INDEX	HouseHolds_with_Children_Percentage	House_Density	Population_Density	Cluster_Labels	Safety	Children	Houses	Population
Burgwallen-Oude Zijde	74.0	7.9	8524.0	12814.0	3	0.725806	0.964444	0.517483	0.482517
Burgwallen-Nieuwe Zijde	85.0	6.9	5282.0	7374.0	3	0.637097	0.986667	0.701802	0.298198
Grachtengordel-West	70.0	12.7	9621.0	14382.0	3	0.758065	0.857778	0.455114	0.544886
Grachtengordel-Zuid	65.0	14.0	6708.0	10457.0	3	0.798387	0.828889	0.620729	0.379271
Nieuwmarkt/Lastage	68.0	13.2	8952.0	13969.0	3	0.774194	0.846667	0.493149	0.506851

5. Questionnaire

5.1 Setup

All the data is now in place to weight the items against the wishes from the person looking for a place to live in Amsterdam. In order to get these wishes a questionnaire is created. The answers for every item will be a rating between 0(agree) and 5(disagree). The following form consists of the following questions:

Please rate the following questions between 0 and 5
Where 0 means the item does not influence your preferred area
and 5 means it does heavily influence your preferred area.
Do you like an area with many children:5
Do you like an area with a high safety index:5
Do you like an area with a lower house density:4
Do you like an area with a lower population density:4
Do you like an area with many health facilities like gyms:2
Do you like an area with many pubs and restaurants:0
Do you like an busy area with for instance many coffee shops and clothing stores:0
Do you like a quiet area with for instance many yoga facilities:0
Do you like to live in an area with traditional industries:0

Above example is prefilled with the data we will use as input being used in the remainder of the analysis/document.

5.2 Score for atmosphere

Based on the input of the last 5 questions of the questionnaire (remember, these were the venue clusters), a simple rating for atmosphere is calculated per neighborhood. If a neighborhood is part of Cluster 2 (healthy area), then in above example the rating for atmosphere for that neighborhood will be 2/5 is 0.4

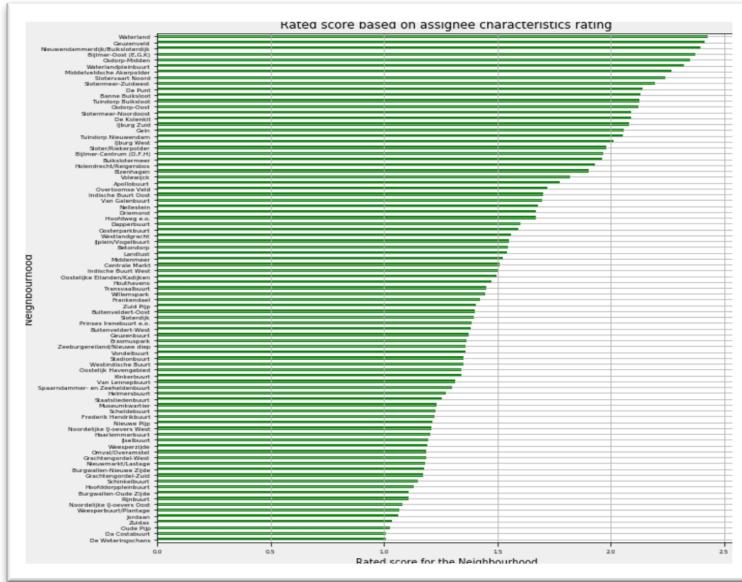
5.3 Score for other characteristics

For the other characteristics (percentage children, population density, house density, safety), the results per neighborhood are weighted with the answers from the questionnaire. For example, if a specific neighborhood has a safety index of y, then that safety index will be multiplied by the result of the questionnaire. In above example that would mean y multiplied by 1 (as the respondent answered with 5, very important to him/her)

5.4 Final score

The final score per neighborhood is calculated as simple sum of the different weighted indexes per characteristic, as shown in below table and related chart.

	Neighbourhood	Population	Rating_Atmosphere	Rating_Safety	Rating_Children	Rating_Houses	Rating_Population	Score
79	Waterland	-0.000000	0.4	0.508065	0.720000	0.800000	-0.000000	2.428065
30	Geuzenveld	0.267724	0.0	0.854839	0.760000	0.585821	0.214179	2.414839
73	Nieuwendammerdijk/Buiksloterdijk	0.099551	0.0	0.701613	0.895556	0.720359	0.079641	2.397168
82	Bijlmer-Oost (E,G,K)	0.197567	0.0	1.000000	0.573333	0.641947	0.158053	2.373333
32	Osdorp-Midden	0.360566	0.0	0.879032	0.671111	0.511547	0.288453	2.350143



6. Food for discussion

6.1 Limitations of the Foursquare data

To obtain the Foursquare data, a free personal account is used. This limits the amount of data which can be obtained from Foursquare and as a result it obviously limited the in-depth analysis possible. The methods used though still apply and can be used on larger datasets when available.

6.2 Limitation of clustering

After clustering the venue data from foursquare it appears that there was a large variance in the amounts of neighborhoods per cluster. This either means that most of the neighborhoods in Amsterdam are very similar or it is just a result of the limited Foursquare data available.

The different clusters were all given an atmosphere characteristic. This categorization is done by me and as such not 100% objective. What I consider to be characteristics of a busy area can be perceived differently by other individuals.

6.3 Limitation of characteristics

For this analysis a limited amount of specific neighborhood characteristics has been used. There are many more characteristics of a neighborhood which can be taken into account when choosing for a perfect place to live.

6.4 Recommendation

As it looks like that many neighborhoods in Amsterdam have similar characteristics based on our analysis, I would recommend to slightly broaden the area to analyze. By taking the areas around Amsterdam into account the individual clusters might differ more from each other which will lead to a more specific advice with respect to a matching area/neighborhood.

7. Conclusion

7.1 Outcome for our random assignee

With the analysis we are able to identify neighborhoods with a good fit with the wishes from the international assignee who wants to find a place to live in Amsterdam. We have for this example collected the input from a random assignee. The ratings of importance this assignee gave to the different characteristics of a neighborhoods are reported below. The final overview is based on these ratings.

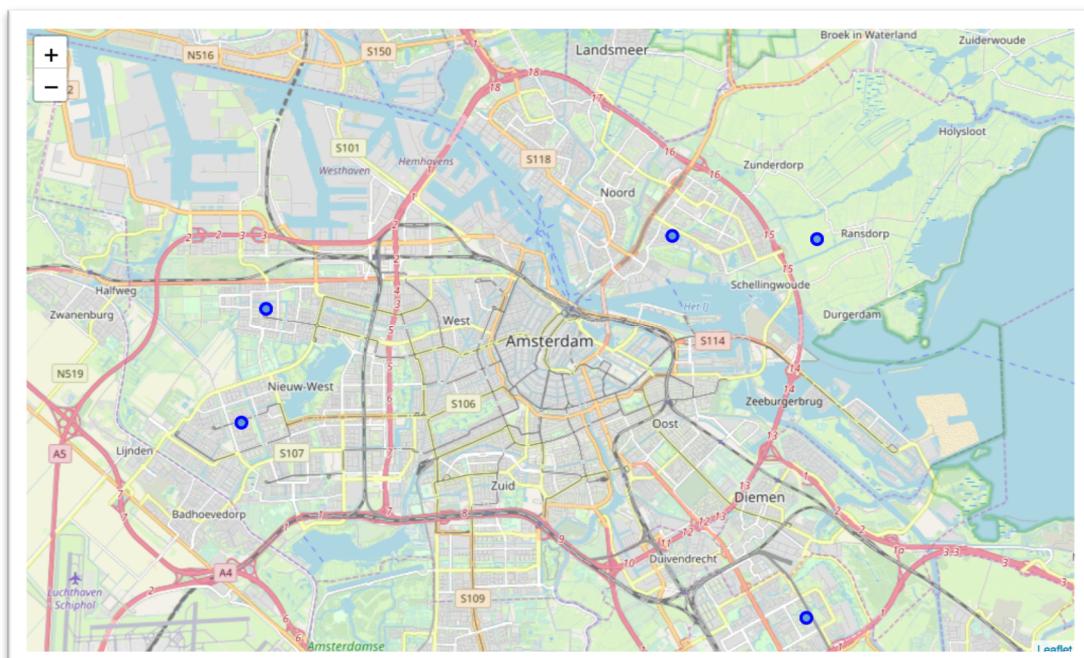
The random assignee gave the following ratings of importance (on a scale of 5):

- Importance to live in a busy area: 0
- Importance to live in an area with health facilities: 2
- Importance to live in an area nearby the harbour: 0
- Importance to live in a lively(shops etc) area: 0
- Importance to live in a quiet area: 0
- Importance to live in a safe area: 5
- Importance to live in an area with a lot of children: 5
- Importance to live in an area with a low population density: 4
- Importance to live in an area with a low house density: 4

Based on this input a top5 of neighborhoods is calculated:

borough	neighbourhood
N73	N Noord
F78	Geuzenveld
N64	N Noord
T94	Bijlmer-Oost (E,G,K)
F82	Osdorp-Midden

These neighbourhoods can be shown on the map of Amsterdam as per below view:



7.2 Outcome for our random assignee

As can be seen in the overview above, the neighborhoods with the best fit given the input from our random assignee are particularly outside the Amsterdam center area. This seems to be a perfect fit with the importance this assignee gave to the different characteristics a neighborhood should have. This random assignee was looking for a save and quiet neighborhood, with a high percentage of children and a low density of houses and population. This are typically the areas outside the center, which is indeed depicted in above overview.

=====END=====