

## **Introduction/ Business problem**

As part of Coursera IBM Data Science Professional Certificate Capstone project, we were asked to cluster neighborhoods in Toronto. The main idea was to get a list of neighborhoods in Toronto, find their geographical coordinates and use them as input to the Foursquare API, from where we can find the top venue categories in each neighborhood e.g: restaurants, hotels, shops etc. Using the frequency of venue categories in each neighborhood, we used the k-means clustering algorithm to cluster neighborhoods into segments with similar venue categories. So similar to this exercise I want to identify clusters of neighborhoods in Amsterdam, the Netherlands that would be suitable for families with children.

From the Foursquare API I will retrieve venue categories and additionally I will use datasets about public population indicators to help me identify residential neighborhoods that are suitable for families with children.

In the end I want to compare my results with a findings from web/blogs that are analyzing best Amsterdam neighborhoods for families to determine how my analysis compares with the reality.

The result of the analysis would be identification of clusters of neighborhoods that would be best for families with children. The benefit of knowing these results is that families who are looking for buying or renting residence in Amsterdam might choose to live in neighborhoods that are family friendly and want to avoid the most popular neighborhoods that are mentioned frequently because those neighborhoods might be overcrowded and real estate prices might be high due to the high demand. The interesting part would be to find similar neighborhoods that can offer comparable life quality and that are less known and this and thus might prove a better choice for families.

## Data

This section introduces the datasets that will be used and their sources.

Based on definition of our problem, factors that will influence our decision are:


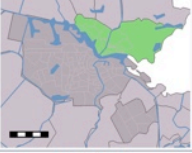


- Amsterdam data related to the Boroughs/Neighborhoods
- The corresponding geographical coordinates latitude and longitude of each Boroughs/Neighborhoods

### Initial Data

The initial table i've found on Wikipedia was a summary one with all boroughs and their corresponding neighborhoods [https://en.wikipedia.org/wiki/Boroughs\\_of\\_Amsterdam](https://en.wikipedia.org/wiki/Boroughs_of_Amsterdam) :

#### List of boroughs [\[ edit \]](#)

Since 2010, there are eight boroughs.<sup>[\[8\]](#)[\[9\]](#)</sup>

Borough ↕	Area ↕	Population ↕	Population density ↕	Location (in green) ↕	Neighbourhoods ↕
<a href="#">Centrum</a> (Centre)	8.04 km²	86,422	13,748/km²		Binnenstad, Grachtengordel, Haarlemmerbuurt, Jodenbuurt, Jordaan, Kadijken, Lastage, Oosterdokseiland, Oostelijke Eilanden, Plantage, Rapenburg, Uilenburg, Westelijke Eilanden, Weteringschans
<a href="#">Noord</a> (North)	49.01 km²	94,766	2,269/km²		Banne Buiksloot, Buiksloot, Buisklootmeer, Floradorp, Kadoelen, Molenwijk, Nieuwendam, Nieuwendammerdijk en Buisklooterdijk, Oostzanerwerf, Overhoeks, Tuindorp Nieuwendam, Tuindorp Oostzaan
<a href="#">Nieuw-West</a> (New West)	32.38 km²	151,677	4,478/km²		Geuzenveld, Nieuw Sloten, Oostoever, Osdorp, Overtoomse Veld, Sloten, Slotermeer, Slotervaart
<a href="#">Oost</a> (East)	30.56 km²	135,767	7,635/km²		IJburg, Indische Buurt, Eastern Docklands, Oud-Oost, Watergraafsmeer

This initial information I've found on Amsterdam boroughs and neighborhood was great for having an overview of the administrative organization and comparing the boroughs by area and population. Since in Amsterdam I've seen there are only 8

boroughs it makes no sense to cluster the boroughs since they are so few but rather find information and geographical coordinates on neighborhood level.

But before finding data on neighborhood level I've used the Geopy library with the Nominatim geolocator service to get the geographical coordinates of all the boroughs:

	Borough	Area	Population	Population density	Neighbourhoods	Population %	Latitude	Longitude
0	Centrum Amsterdam, NL	8.04 km <sup>2</sup>	86422	13,748/km <sup>2</sup>	Binnenstad, Grachtengordel, Haarlemmerbuurt, J...	0.102280	52.373730	4.895691
1	Noord Amsterdam, NL	49.01 km <sup>2</sup>	94766	2,269/km <sup>2</sup>	Banne Buiksloot, Buiksloot, Buikslotermeer, FI...	0.112155	52.401739	4.915352
2	Nieuw-West Amsterdam, NL	32.38 km <sup>2</sup>	151677	4,478/km <sup>2</sup>	Geuzenveld, Nieuw Sloten, Oostoever, Osdorp, O...	0.179510	52.363777	4.813812
3	Oost Amsterdam, NL	30.56 km <sup>2</sup>	135767	7,635/km <sup>2</sup>	IJburg, Indische Buurt, Eastern Docklands, Oud...	0.160680	52.356608	4.930577
4	West Amsterdam, NL	9.89 km <sup>2</sup>	143842	15,252/km <sup>2</sup>	Frederik Hendrikbuurt, Houthaven, Spaarndammer...	0.170237	52.372748	4.889846
5	Westpoort Amsterdam, NL	10 km <sup>2</sup>	192	10/km <sup>2</sup>	Westpoort	0.000227	52.411468	4.800573
6	Zuid Amsterdam, NL	17.41 km <sup>2</sup>	144432	9,349/km <sup>2</sup>	Apollobuurt, Buitenveldert, Hoofddorppeleinbuur...	0.170935	52.339194	4.874232
7	Zuidoost Amsterdam, NL	22.08 km <sup>2</sup>	87854	4,391/km <sup>2</sup>	Bijlmermeer, Venserpolder, Gaasperdam, Driemond	0.103975	52.310514	4.960695

## Amsterdam Municipality Data

As explained before my goal was to find different data source where i would have information on neighborhood level as i will use neighborhoods further in order to cluster them and identify similar neighborhoods good for families. So I did a lot of investigations in order to find proper data and finally I came across the official Amsterdam Municipality data .

I've extracted the boroughs data to cross check with the initial data I've found on wiki:

OBJECTNUMMER	Stadsdeel_code	Stadsdeel	Opp_m2	WKT_LNG_LAT	WKT_LAT_LNG	LNG	LAT	Unnamed: 8
0	1	A Centrum	8043500	POLYGON((4.932973 52.3704,4.932942 52.370539,4...	POLYGON((52.3704 4.932973,52.370539 4.932942,5...	4.903712	52.373297	NaN
1	2	B Westpoort	28991600	POLYGON((4.885861 52.39937,4.882702 52.401695,...	POLYGON((52.39937 4.885861,52.401695 4.882702,...	4.807319	52.411465	NaN
2	3	E West	10629900	POLYGON((4.895084 52.388684,4.894675 52.389933...	POLYGON((52.388684 4.895084,52.389933 4.894675...	4.865216	52.377879	NaN
3	4	F Nieuw-West	38015500	POLYGON((4.850498 52.364232,4.850459 52.365189...	POLYGON((52.364232 4.850498,52.365189 4.850459...	4.802676	52.363591	NaN
4	5	K Zuid	17274000	POLYGON((4.914989 52.342139,4.914945 52.342421...	POLYGON((52.342139 4.914989,52.342421 4.914945...	4.866063	52.341721	NaN
5	6	M Oost	30594900	POLYGON((5.039059 52.354569,5.038812 52.358098...	POLYGON((52.354569 5.039059,52.358098 5.038812...	4.967149	52.350438	NaN
6	7	N Noord	63828800	POLYGON((5.079164 52.388647,5.074264 52.413831...	POLYGON((52.388647 5.079164,52.413831 5.074264...	4.967446	52.399439	NaN
7	8	T Zuidoost	22113700	POLYGON((5.021546 52.302451,5.021464 52.303129...	POLYGON((52.302451 5.021546,52.303129 5.021464...	4.975444	52.304654	NaN

And then I've extracted the important data I needed on neighborhood level :

Buurt_code	Buurt	Buurtcombinatie_code	Stadsdeel_code	Opp_m2	WKT_LNG_LAT	WKT_LAT_LNG	LNG	LAT
F81d	Calandlaan/Lelylaan	F81	F	275360.0	POLYGON((4.800801 52.355175,4.809055 52.356842...	POLYGON((52.355175 4.800801,52.356842 4.809055...	4.809697	52.355701
F81e	Osdorp Zuidoost	F81	F	519366.0	POLYGON((4.818583 52.357519,4.818622 52.356295...	POLYGON((52.357519 4.818583,52.356295 4.818622...	4.811344	52.353736
F82a	Osdorp Midden Noord	F82	F	215541.0	POLYGON((4.786657 52.362712,4.795326 52.364434...	POLYGON((52.362712 4.786657,52.364434 4.795326...	4.791792	52.362071
F82b	Osdorp Midden Zuid	F82	F	258379.0	POLYGON((4.788293 52.359736,4.796917 52.36148,...	POLYGON((52.359736 4.788293,52.36148 4.796917,...	4.793781	52.358831
F82c	Zuidwestkwadrant Osdorp Noord	F82	F	240774.0	POLYGON((4.790209 52.356207,4.799258 52.358027...	POLYGON((52.356207 4.790209,52.358027 4.799258...	4.795597	52.355521

Since the data came from the official municipality of the city the column names are in Dutch so I've translated and renamed them in my final table.

I've decided to use the geographical coordinates available in this data since they came from the official municipality API and considering there are 481 neighborhoods would make no sense to use the same function I've created before for boroughs to loop through the list of names and extract them from Geopy library since they are already present in this data.

## Additional data

Additional data related to the population/housing by neighborhoods will also be extracted from Amsterdam Municipality official API.

I've found in the available data that there are some indicators I could use for my further analysis, this information would be related to population, housing, public space, safety, education.

There is a table with definition of columns and I had a look there and only selected some variables related to the topics mentioned above:

	Variabele	Topic area	Label_1	Definition
0	BEVTOTAAL	Population	Population	Number of people registered in Amsterdam on Ja...
19	BEVDICHT	Population	Population density	Number of residents per square kilometre of land.
48	BEVNSTEDELING_P	Population	New urbanites (%)	Percentage of the population classified as 'ne...
63	BEVPAARMKINDHH_P	Population	Households: % couple with children	Percentage of households: couple with children...
178	WDICHT	Housing	Housing density	Housing density: Number of homes per square ki...
191	WKOOP_P	Housing	Property: % Owner-occupied	Percentage of addresses registered as property...
211	WWOZ_M2	Housing	Average house-value for tax purposes per m2	Average house-value per m2 as determined by th...
230	LBUURT_R	Housing	Neighbourhood: Satisfaction with neighbourhood...	Average answer to the question: \r\nHow satisf...
240	ORGROEN_R	Public space	Green spaces (1-10)	Average answer to the question: \r\nWhat is yo...
241	ORAANBODSPELEN_R	Public space	Playing facilities (1-10)	Average answer to the question: \r\nWhat is yo...
277	VVEILIGH_I	Safety	Safety index total	The Safety Index describes the safety of an ar...
669	OSCHBAO	Education	Primary education: schools	Number of schools for primary education.

Then i had to extract from the data with all kind of indicators only these columns, by doing so I've found out that not all these selected columns are available on neighborhood level unfortunately. So for example the public space, safety and

education related columns are only available on total not on this more granular level of neighborhoods...but still the information on population and housing would be helpful and I will use it further.

variabele	Neighborhood code	Population density	New urbanities %	Couples with children %	Population	Housing Density
0	A	13985.0	39.3	8.9	87310.0	8806.0
1	A00a	18447.0	47.2	3.3	1100.0	12947.0
2	A00b	8249.0	48.5	2.7	728.0	6051.0
3	A00c	25874.0	44.6	4.8	1613.0	16057.0
4	A00d	6562.0	57.5	5.3	351.0	4917.0

I've created a final data with combined info of neighborhoods, boroughs, geographical coordinated and additional indicators on population and housing:

Neighborhood	Neighborhood code	Borough code	Area m2	Latitude	Longitude	Borough	Borough Area m2	Borough Latitude	Borough Longitude	Population density	New urbanities %
Calandlaan/Lelylaan	F81d	F	275360.0	52.355708	4.809697	Nieuw-West	38015500	52.363591	4.802676	4735.0	7.0
Osdorp Zuidoost	F81e	F	519366.0	52.353736	4.811344	Nieuw-West	38015500	52.363591	4.802676	8282.0	17.6
Osdorp Midden Noord	F82a	F	215541.0	52.362078	4.791792	Nieuw-West	38015500	52.363591	4.802676	13403.0	3.3
Osdorp Midden Zuid	F82b	F	258379.0	52.358838	4.793781	Nieuw-West	38015500	52.363591	4.802676	14779.0	6.5
Zuidwestkwadrant Osdorp Noord	F82c	F	240774.0	52.355523	4.795597	Nieuw-West	38015500	52.363591	4.802676	13789.0	9.8

## Venue categories

Venue categories and their type and location in every neighborhood will be obtained using **Foursquare API**.

Example nearby venues (500 m) for one neighborhood:

nearby_venues				
	name	categories	lat	lng
0	Toko Bandung	Indonesian Restaurant	52.354358	4.810843
1	Enfes	Turkish Restaurant	52.354057	4.810545
2	Sportcentrum Caland	Gym / Fitness Center	52.354371	4.807132
3	De Meervaart	Theater	52.358970	4.807311
4	TK Maxx	Clothing Store	52.359155	4.805335
5	Action	Discount Store	52.358802	4.804731
6	Winkelcentrum Osdorpplein	Shopping Mall	52.358236	4.806964
7	Kruidvat	Drugstore	52.359389	4.806192
8	De Serre	Snack Place	52.358866	4.804710
9	Kruidvat	Drugstore	52.359695	4.806452
10	Snackbar 88	Snack Place	52.354488	4.812280

In the end will have a table with all neighborhoods and their nearby venue categories from Foursquare API

List of best neighborhoods for families according to web/ blog posts :  
Diamantbuurt, Rivierenbuurt, Stadionbuurt, Apollobuurt, Buitenveldert, Amstelveen, Eastern Docklands, Oostenlijke Havengebied, Indische Buurt, Transvaalbuurt, Singel, Prinsengracht, Herengracht, Keizersgracht, Jordaan, Osdorp, Sloten, Slotervaart, de Baarsjes , Bos en Lommer, Westerpark, Spaarndammerbuurt, Noord.

I will add a flag for these neighborhoods in the final data where we have all the info for Amsterdam's neighborhoods.