

Report on Dutch cities

By Stefan Arends as Part of the Capstone project on “IBM Data Science Professional Certificate

Introduction

I am planning to move to the Netherlands, a small country in Europe. Because it is a small country, all cities are acceptable because of traveling distances. To have more insights of which city to select for moving to, I want to unlock the data available in Foursquare on analysing the different cities.

Based on the information I want to cluster the different cities with machine learning, so I have insight on similar cities within the Netherlands.

Problem statement:

Give insights in the characteristics of the 50 biggest cities in the Netherlands, based on venue information in Foursquare.

This can be input for selecting a (cluster of) cities which are candidate to move to.

Data

For this analysis the following data sources will be used:

1. **List of Dutch cities**, downloaded from the CBS, the Dutch Statistics Agency. Link: <https://www.cbs.nl/nl-nl/maatwerk/2019/31/kerncijfers-wijken-en-buurten-2019>
2. **Geo coordinates** of the major Dutch cities, downloaded from <https://simplemaps.com/data/nl-cities>
3. **Top venues** in each Neighbourhoods gathered from Foursquare by the API.

These data sources will be combined, to cluster the different Dutch cities.

Methodology

For the analysis the K-means clustering algorithm is used. In order to use this, several steps need to be taken:

- A list of 50 biggest cities need to be found. For this project a list of the CBS (Dutch statistics Agency) is available in excel.

- Because this does not contain geographical information about the location of the cities, this list must be combined with geographical information. This will be done by merging with another data source, containing latitude and longitude for around 400 Dutch cities.
- After combining the result is a list with the top 50 dutch cities, including latitude and longitude.
- With this geo-location, the Foursquare-api will be used to gather popular venues for each city.
- For each city the top 10 category of venues will be calculated, including the mean use of it. This top 10 categories can be seen as a characterization of the city.
- This characterization can be input for the clustering of cities, to assign similar cities to the same cluster.
- For this clustering the K-means machine learning is used.
- Based on the clustering the clusters of cities can be analysed and characterized, as input for making the choice.

Results

First result is merging of the cities dataset with the geo dataset. The first 5 cities are presented in the table below

	city	number_inhabitants	latitude	longitude
0	Amsterdam	862965	52.350000	4.916667
1	Rotterdam	644618	51.916667	4.500000
2	Den Haag	537833	52.083333	4.300000
3	Utrecht	352866	52.093813	5.119095
4	Eindhoven	231642	51.450000	5.466667

These cities can also be plotted on a map (without clustering yet)



Based on geo-location, for each city the top 10 venues can be calculated, with result (only 5 first cities are shown)

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alkmaar	Furniture / Home Store	Office	Design Studio	Professional & Other Places	Residential Building (Apartment / Condo)	Hardware Store	Parking	Electronics Store	Salon / Barbershop	Mattress Store
1	Almere	Office	Bridge	Building	Elementary School	Other Great Outdoors	Garden	School	Mental Health Office	Medical Center	Bus Stop
2	Alphen aan den Rijn	Office	Medical Center	Bus Stop	Greek Restaurant	Church	Building	Salon / Barbershop	Gym / Fitness Center	Gym	Playground
3	Amersfoort	Bar	Restaurant	Women's Store	Café	Other Nightlife	Shoe Store	Wine Bar	Coffee Shop	Clothing Store	Gay Bar
4	Amstelveen	Dentist's Office	Building	Office	Bakery	Drugstore	Salon / Barbershop	Martial Arts Dojo	Supermarket	Wine Shop	Miscellaneous Shop

By using K-means, this can be input for clustering the cities, with result below. Each color is representing another cluster.



For insight in the characteristics of (the cities in) the different clusters are shown by diving in the different clusters. This results are available in excel and also printed in the appendix



Discussion

If you dive into the different clusters, there is not much difference shown between the different clusters. For all clusters the category “Office” is the number one for most cities because of the amount of the Foursquare venues in this category.

However small differences between the clusters can be identified.

A personal interpretation of the clusters can be:

- Cluster 0: Working environment with offices and automotive
- Cluster 1: Working environment with medical services in place
- Cluster 2: Working environment most time situated in the “Randstad” the most densely populated region of the Netherlands
- Cluster 3: Studying environment
- Cluster 4: Living environment with furniture and home stores nearby.

Conclusion

For selecting a city to live in, this analysis gives limited information, because the differentiation between the different cities and clusters is small. Maybe only cluster 3 and 4 differs from the others, based on their profile. So if Education (cluster 3) or Furniture/home stores (cluster 4) are important items, these cities can be considered.

To get more information on the cities, further research is recommended. See next section for recommendations on that.

Recommendations

For gathering more information on the different cities, extra analysis is recommended. Possible options for that are:

- Filter out categories in Foursquare that are not relevant for making your choice (like “Office”)
- Use information in data source 1. A lot of information on the cities is kept in this dataset (like typical household, mean income, origins of inhabitants, age)
- Consider to unlock not only on city level, but also on Borough or Neighbourhoods.
- Use Google Places instead of Foursquare, because of the low usage of Foursquare.

Links

Gitbub location with all resources used: <https://github.com/StefanArends/CapStone>

- Python Code in Notebook:
<https://github.com/StefanArends/CapStone/blob/master/Clustering%20Dutch%20Cities.ipynb>
- Output exported to excel:
<https://github.com/StefanArends/CapStone/blob/master/output.xlsx>

Appendix 1

Cluster 0

city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Tilburg	0	Office	Building	Factory	Automotive Shop	Coworking Space	School	Transportation Service	Park	Doctor's Office	College Academic Building
Haarlemmermeer	0	Office	Building	Automotive Shop	Hardware Store	Factory	Furniture / Home Store	Restaurant	Pizza Place	Playground	Hobby Shop
Westland	0	Building	Office	Flower Shop	Automotive Shop	Convention Center	Dentist's Office	Bus Stop	Professional & Other Places	Bus Line	Distribution Center
Hoeksche Waard	0	Office	Building	Automotive Shop	Bus Stop	Dentist's Office	School	Library	College Classroom	Bus Line	Gym / Fitness Center
Schiedam	0	Office	Building	Factory	Soccer Field	Athletics & Sports	Automotive Shop	Garden	Lounge	Nightclub	Restaurant

Cluster 1

city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Eindhoven	1	Office	Residential Building (Apartment / Condo)	General College & University	Furniture / Home Store	Housing Development	Chinese Restaurant	Turkish Restaurant	Music Venue	Bar	Fast Food Restaurant
Groningen	1	Building	Office	Medical Center	Dentist's Office	Pizza Place	Tech Startup	Professional & Other Places	Residential Building (Apartment / Condo)	Bar	Bus Stop
Breda	1	Office	Bus Stop	Residential Building (Apartment / Condo)	Gas Station	Professional & Other Places	Cosmetics Shop	Bakery	Parking	Playground	Supermarket
Nijmegen	1	Office	School	Medical Center	Snack Place	Residential Building (Apartment / Condo)	Playground	Dentist's Office	High School	Tech Startup	Building
Haarlem	1	Hospital	Office	Medical Center	Building	Residential Building (Apartment / Condo)	Bus Stop	Physical Therapist	Pool	Medical Lab	Bar
Arnhem	1	Building	Office	Playground	Bakery	Doctor's Office	Bus Stop	High School	General Entertainment	Snack Place	Garden
Amersfoort	1	Bar	Restaurant	Women's Store	Café	Other Nightlife	Shoe Store	Wine Bar	Coffee Shop	Clothing Store	Gay Bar
Zaanstad	1	City Hall	Building	Platform	Office	Residential Building (Apartment / Condo)	Hotel	Salon / Barbershop	Bus Stop	Bus Line	Convenience Store
Den Bosch	1	Office	Bus Stop	Medical Center	Residential Building (Apartment / Condo)	Professional & Other Places	Dog Run	Gas Station	Discount Store	Building	Arts & Crafts Store

Zoetermeer	1	Art Gallery	Furniture / Home Store	Middle Eastern Restaurant	Office	Music Venue	Flower Shop	Restaurant	Diner	French Restaurant	Dentist's Office
Leiden	1	Residential Building (Apartment / Condo)	Bus Stop	Medical Center	Building	Bar	General College & University	Harbor / Marina	Trail	Playground	River
Leeuwarden	1	Office	Dentist's Office	Residential Building (Apartment / Condo)	Medical Center	Bar	Doctor's Office	Salon / Barbershop	Bus Line	Building	Recruiting Agency
Maastricht	1	General College & University	College Arts Building	Building	Residential Building (Apartment / Condo)	College Classroom	Office	Fraternity House	Bar	Tattoo Parlor	Restaurant
Dordrecht	1	Office	Gym / Fitness Center	Bus Stop	Building	Professional & Other Places	Soccer Field	Bike Shop	Veterinarian	Gas Station	Garden Center
Alphen aan den Rijn	1	Office	Medical Center	Bus Stop	Greek Restaurant	Church	Building	Salon / Barbershop	Gym / Fitness Center	Gym	Playground
Emmen	1	Bus Line	Office	Building	Bike Shop	Salon / Barbershop	Restaurant	Optical Shop	Pharmacy	Gym	Church
Sittard-Geleen	1	Office	Building	Bar	Residential Building (Apartment / Condo)	Other Great Outdoors	General Entertainment	Tea Room	Pub	Greek Restaurant	Italian Restaurant
Helmond	1	Dog Run	Bus Stop	Building	Park	Office	Fast Food Restaurant	Supermarket	Parking	Doctor's Office	Playground
Oss	1	Office	Bus Line	Building	Professional & Other Places	Residential Building (Apartment / Condo)	Fast Food Restaurant	Medical Center	Bar	Bus Stop	Salon / Barbershop
Amstelveen	1	Dentist's Office	Building	Office	Bakery	Drugstore	Salon / Barbershop	Martial Arts Dojo	Supermarket	Wine Shop	Miscellaneous Shop
Sneek	1	Clothing Store	Women's Store	Men's Store	Salon / Barbershop	Pharmacy	Furniture / Home Store	Drugstore	Ice Cream Shop	Toy / Game Store	Electronics Store

Nissewaard	1	Doctor's Office	Bridge	Snack Place	Medical Center	Housing Development	Building	Physical Therapist	Bus Stop	Office	Residential Building (Apartment / Condo)
Meerijstad	1	Building	Office	Bus Stop	Salon / Barbershop	Pizza Place	Furniture / Home Store	Doctor's Office	Jewelry Store	Clothing Store	Restaurant
Hengelo	1	Office	Gas Station	Building	Bar	Salon / Barbershop	Beer Garden	Hookah Bar	Supermarket	Snack Place	Bus Stop
Hengelo	1	Office	Gas Station	Building	Bar	Salon / Barbershop	Beer Garden	Hookah Bar	Supermarket	Snack Place	Bus Stop
Purmerend	1	Building	Gas Station	Bus Line	Bus Stop	Grocery Store	Supermarket	Event Space	Gym / Fitness Center	Garden	Lounge
Leidschendam-Voorburg	1	Office	Building	Residential Building (Apartment / Condo)	Salon / Barbershop	Event Space	Professional & Other Places	School	Veterinarian	Bus Stop	Dentist's Office
Gouda	1	Office	Medical Center	Doctor's Office	Hospital	Residential Building (Apartment / Condo)	School	Bus Stop	Snack Place	Church	Professional & Other Places

Cluster 2

city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Amsterdam	2	Office	Building	Residential Building (Apartment / Condo)	Coworking Space	Assisted Living	Bar	Non-Profit	Tech Startup	Playground	Meeting Room
Rotterdam	2	Office	Residential Building (Apartment / Condo)	Building	Bar	Bus Stop	Coworking Space	Restaurant	Business Service	Event Space	Housing Development
Den Haag	2	Office	Embassy / Consulate	Art Gallery	Salon / Barbershop	Dentist's Office	Restaurant	Residential Building (Apartment / Condo)	Miscellaneous Shop	Plaza	Coworking Space
Utrecht	2	Office	Salon / Barbershop	Snack Place	Clothing Store	Bar	General Entertainment	Tech Startup	Fast Food Restaurant	Shoe Store	Cocktail Bar
Almere	2	Office	Bridge	Building	Elementary School	Other Great Outdoors	Garden	School	Mental Health Office	Medical Center	Bus Stop
Apeldoorn	2	Office	Building	Medical Center	Electronics Store	Bank	Market	General Entertainment	Food Stand	Gas Station	Government Building
Ede	2	Office	Medical Center	Building	Elementary School	Other Great Outdoors	Non-Profit	Chiropractor	Dentist's Office	Housing Development	Professional & Other Places
Delft	2	Office	Building	Other Great Outdoors	Athletics & Sports	Campground	Playground	Dog Run	Medical Center	General Entertainment	Park
Venlo	2	Office	Music Venue	Bus Stop	Park	Gay Bar	Doctor's Office	Bakery	School	Gas Station	College Lab
Hilversum	2	Office	Furniture / Home Store	Building	Arts & Crafts Store	Dog Run	Construction & Landscaping	Gay Bar	Parking	Law School	Restaurant
Lelystad	2	Office	Medical Center	Building	College Administrative Building	College Classroom	Bridge	College Auditorium	Coworking Space	Gym / Fitness Center	Gym

Cluster 3

city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Enschede	3	College Classroom	Office	Building	Design Studio	Professional & Other Places	Residential Building (Apartment / Condo)	Tattoo Parlor	Bank	Rest Area	Restaurant
Zwolle	3	Office	College Classroom	Building	Parking	Platform	Coworking Space	Professional & Other Places	Bus Line	Tech Startup	Train Station
Heerlen	3	College Classroom	Office	College Auditorium	College Administrative Building	College Academic Building	Student Center	General College & University	College & University	College Cafeteria	Coworking Space

Cluster 4

city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Alkmaar	4	Furniture / Home Store	Office	Design Studio	Professional & Other Places	Residential Building (Apartment / Condo)	Hardware Store	Parking	Electronics Store	Salon / Barbershop	Mattress Store
Deventer	4	Automotive Shop	Furniture / Home Store	Bus Stop	Building	Dentist's Office	Office	Police Station	Miscellaneous Shop	School	Auto Dealership
Roosendaal	4	Office	Factory	Furniture / Home Store	Building	Automotive Shop	Design Studio	Professional & Other Places	Sandwich Place	Tech Startup	Fast Food Restaurant