Capstone Project

The Battle of Neighborhoods (Week 1)

Which neighborhoods are best suited for the elderly?

Problem and background

- The city of Utrecht has a lot of different neighborhoods. We want to group these neighborhoods so that we can gain more insight in which neighborhoods are more suited towards the elderly.
- This information will be useful for the elderly to choose where they would like to live. It might also be helpful for the city council to decide where to build more hospitals or other necessities for the elderly. It will also be helpful for general practitioners to decide on where to start a new practice.

Description of the data

We will be using data from the Central Bureau of Statistics in the Netherlands. We will use date from two different places. We will therefore have to clean the data correctly and merge them together.

One source will provide data on the amount of citizens and also on the amount of citizens of 65 years of age and older. By using this information we can calculate the percentage of citizens of 65 years of age and older in each neighborhood ourselves.

The other source will provide us with information on the average distance to a general practitioner. This is important for the elderly. It will also provide the city council and general practitioners on where there are opportunities.

We will locate the geographical coordinates of Utrecht by using geolocator. As the original data is in Dutch, we will have to translate it to English.

Links

https://www.cbs.nl/nl-nl/reeksen/kerncijfers-wijken-en-buurten-2004-2019

https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS&tableId=84463NED&_theme=401

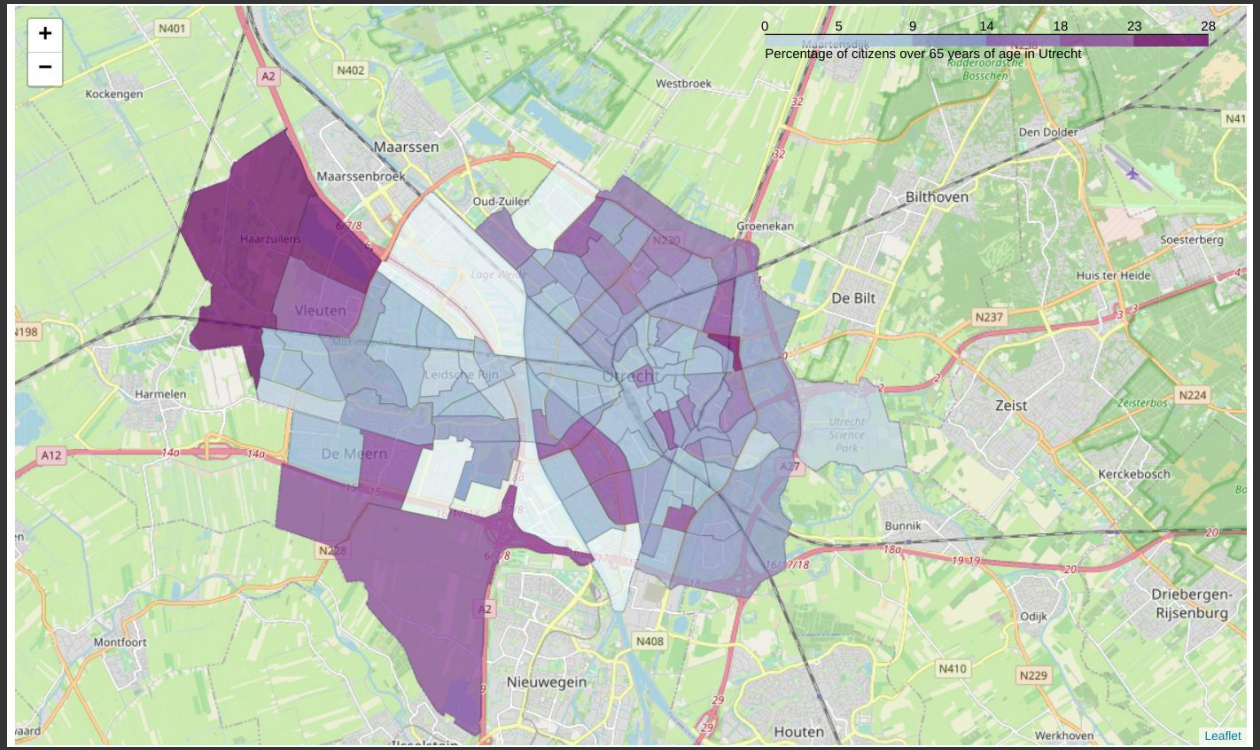
The first five rows of our cleaned dataframe

	neighborhood	total	over65yo	percentage_over65yo	distance_gp_km
0	Welgelegen, Den Hommel	1500	340	22.666667	0.7
1	Oog in Al	4370	435	9.954233	0.4
2	Halve Maan-Zuid	1465	220	15.017065	0.4
3	Halve Maan-Noord	1710	145	8.479532	0.7
4	Lombok-Oost	2290	195	8.515284	0.5

The first five rows of our cleaned dataframe look like this:

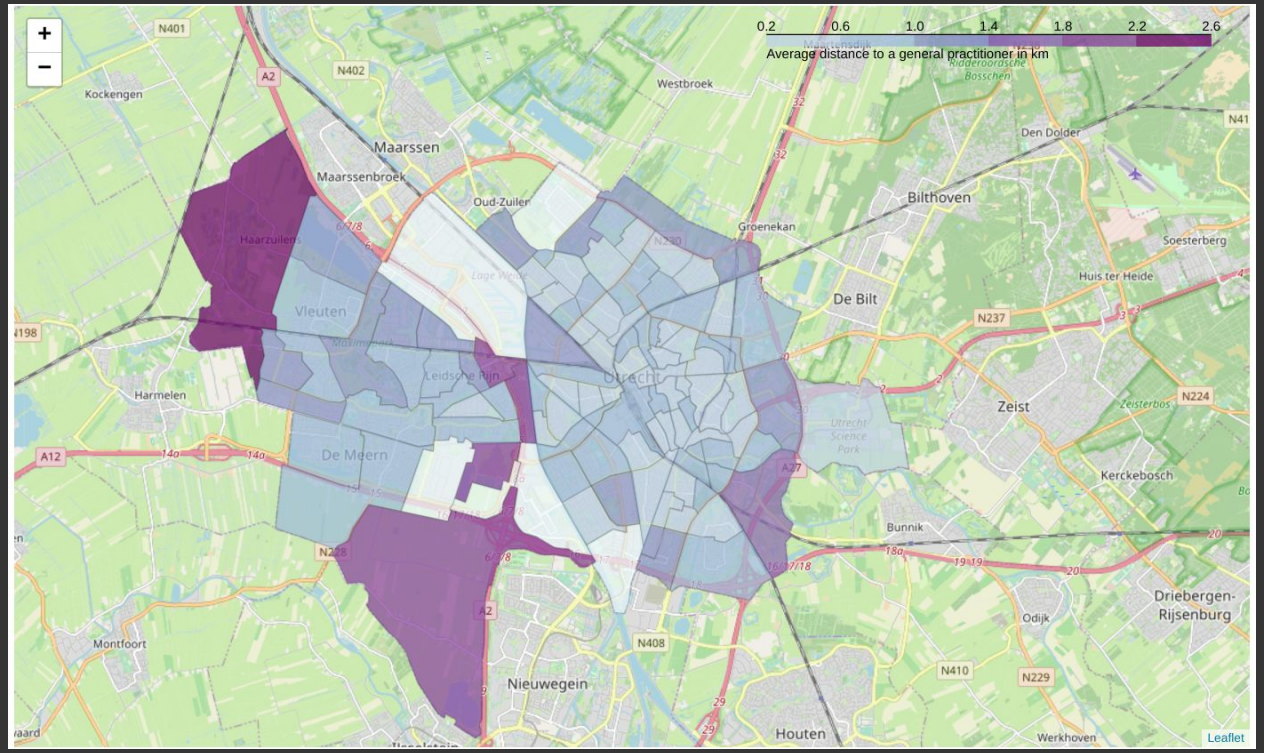
Choropleth map

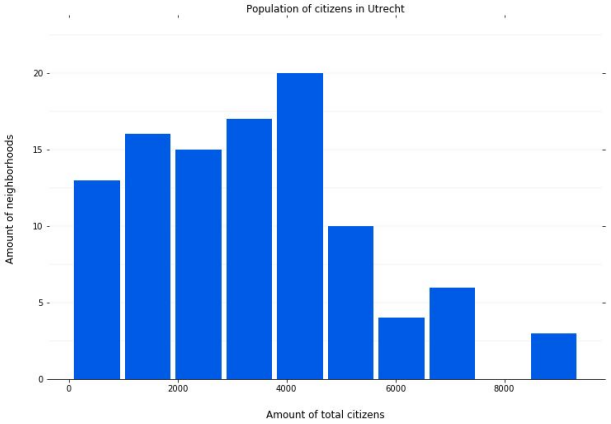
Percentage of citizens over 65 years of age in Utrecht

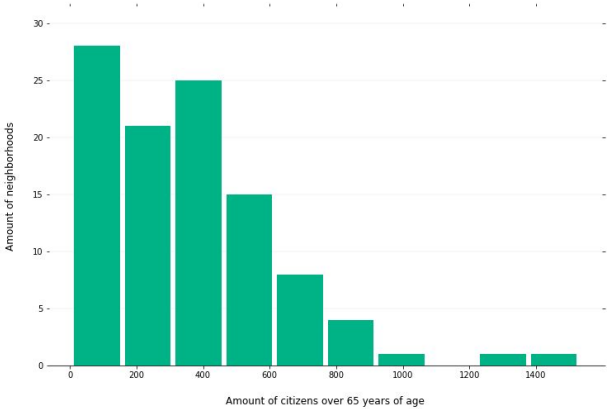


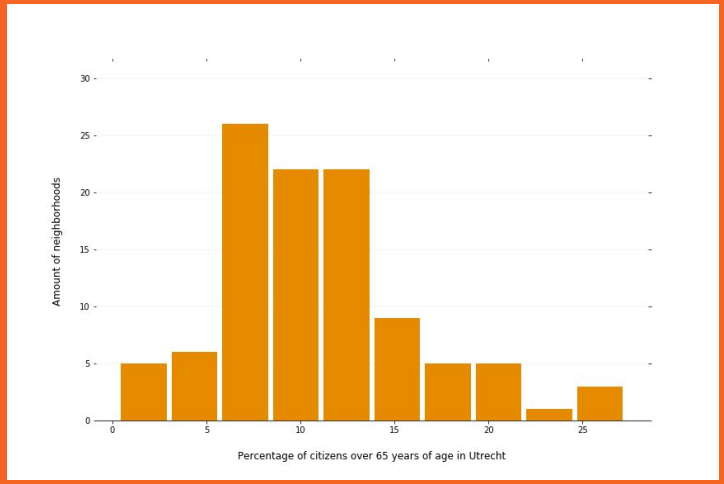
Choropleth map

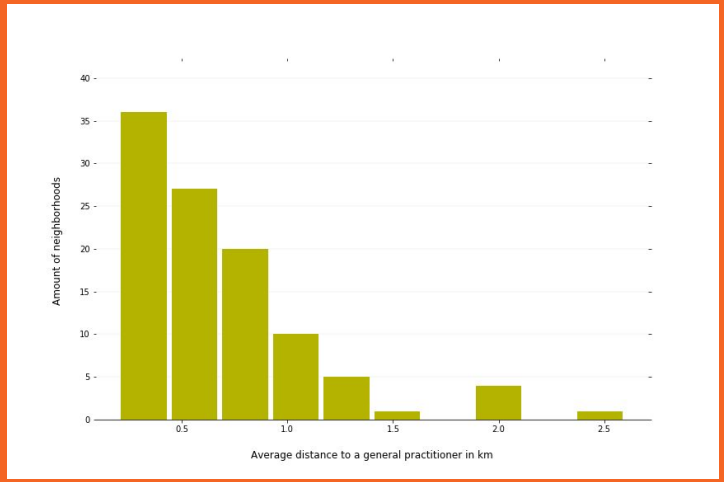
Average distance to a general practitioner in km



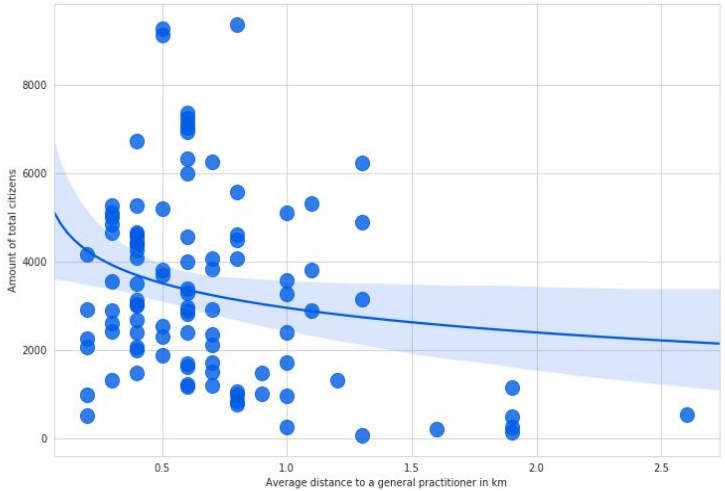

Population of citizens in Utrecht
Amount of neighborhoods
Amount of total citizens
0
5
10
15
20
0
2000
4000
6000
8000


Amount of neighborhoods
Amount of citizens over 65 years of age
0
5
10
15
20
25
30
0
200
400
600
800
1000
1200
1400

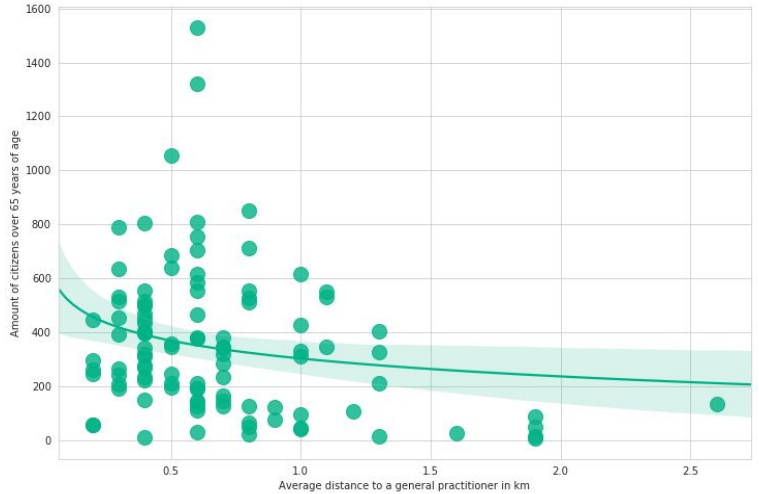

Amount of neighborhoods
Percentage of citizens over 65 years of age in Utrecht
0
5
10
15
20
25
30
0
5
10
15
20
25


Amount of neighborhoods
Average distance to a general practitioner in km
0
5
10
15
20
25
30
35
40
0.5
1.0
1.5
2.0
2.5

Population of citizens in relation to average distance to a GP



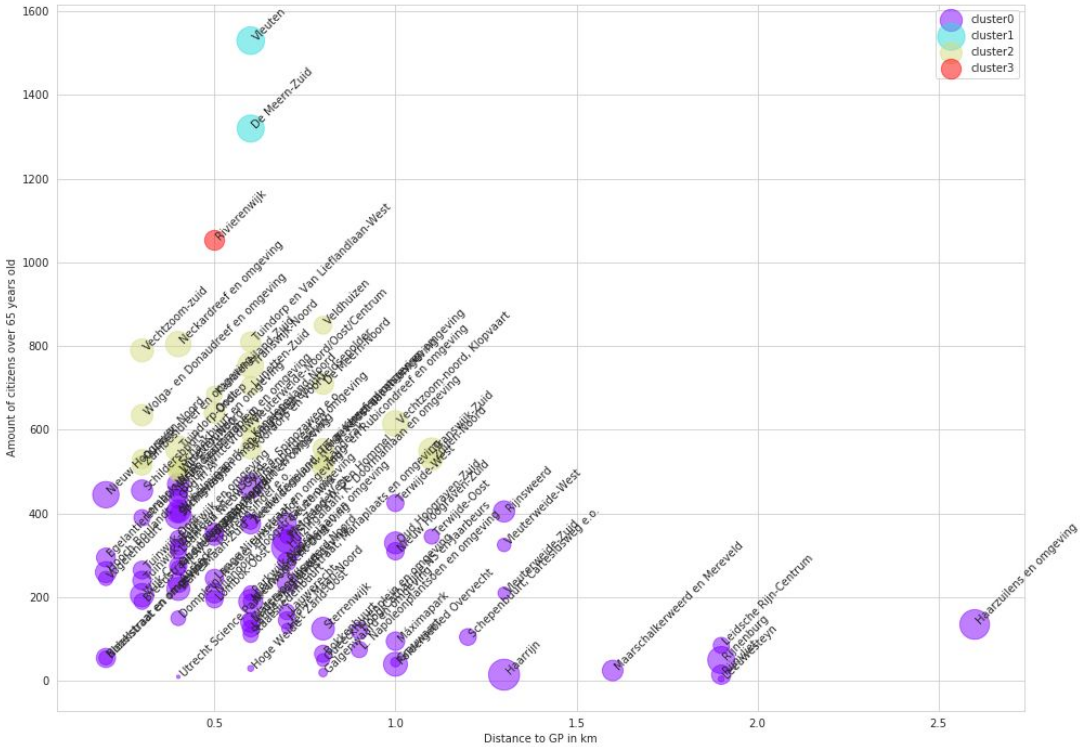
Population of citizens over 65 years of age in relation to average distance to a GP



Results section

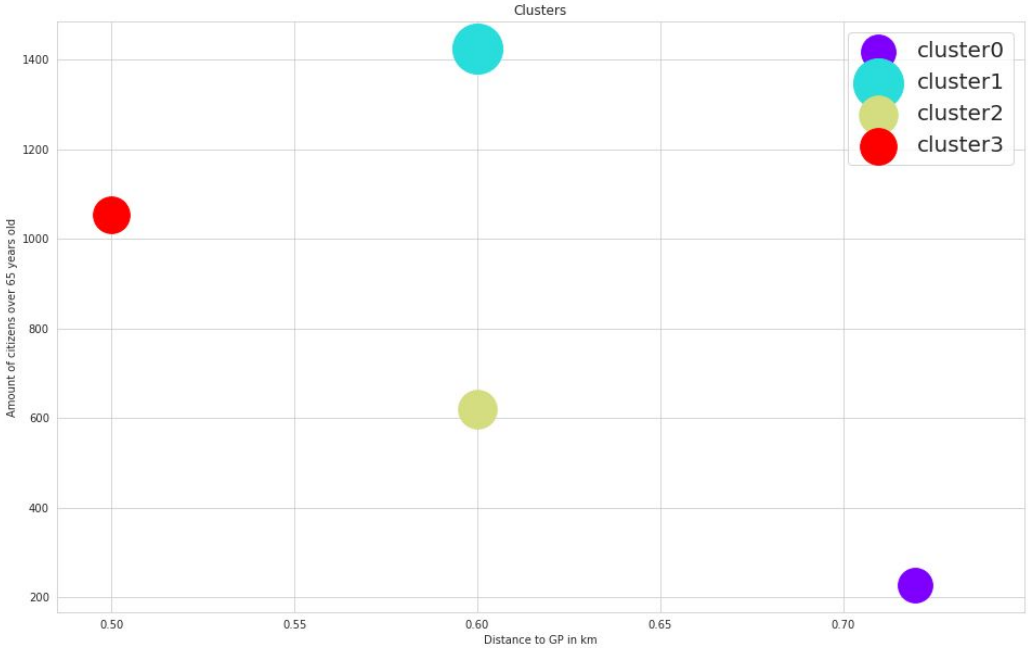
We have used the 'AgglomerativeClustering' function from scikit-learn library to cluster the dataset. The AgglomerativeClustering performs a hierarchical clustering using a bottom up approach. We have used an average linkage criteria, which minimizes the average of the distances between all observations of pairs of clusters.

Here we see the population of citizens over 65 years of age in relation to average distance to a general practitioner. The size of the blobs represents the percentage of citizens over 65 years of age in that particular neighbourhood.


cluster0
cluster1
cluster2
cluster3
Veuten
De Meern-Zuid
Rivierenwijk
Vechtzoom-zuid
Wolga- en Donaudreef en omgeving
Veldhuizen
Vechtzoom-noord, Klopvaart
Maarschalkerweerd en Mereveld
Leidsche Rijn-Centrum
Harzuilens en omgeving
Haarrijn
Amount of citizens over 65 years old
Distance to GP in km
0
200
400
600
800
1000
1200
1400
1600
0.5
1.0
1.5
2.0
2.5

Discussion section

In the plot below we can see more clearly where the centroid of each cluster is. We can differentiate several clusters of neighborhoods. Cluster0 is a larger cluster with relatively few citizens of over 65 years of age and a large spread in distance to a general practitioner. Cluster2 has somewhat more citizens of over 65 years of age situated more close to a general practitioner. Cluster 3 is situated closest to a general practitioner compared to the other cluster. There are a lot of citizens of over 65 years of age living there, although when looking at the size of the blob we note that the percentage of citizens of over 65 years of age is not that high. Cluster 1 has a lot of citizens of over 65 years of age living there. When we look at the size of the blob the percentage of citizens of over 65 years of age living there is also large. Cluster 1 is situated somewhat less close to a general practitioner, although the difference in distance is minimal. We can see that clearly when looking at our previous plot.


Clusters
cluster0
cluster1
cluster2
cluster3
Amount of citizens over 65 years old
1400
1200
1000
800
600
400
200
0.50
0.55
0.60
0.65
0.70
Distance to GP in km

Conclusion section

Elderly who would like to live in a neighborhood with lots of other citizens of over 65 years of age and who would like to live relatively close to a general practitioner may prefer the neighborhoods in cluster1. A lot of citizens of over 65 years of age live here, both in numbers and percentually. Elderly who prefer to live closer to a general practitioner could opt for cluster 3. Although less citizens of over 65 years of age live there, both in numbers and percentually. Elderly who prefer to live percentually with more citizens of over 65 years of age, but in relatively quiet neighborhoods with less total citizens, and who don't mind distance to a general practitioner could opt for neighborhoods in cluster0.

The city council might want to decide to build more hospitals or other necessities for the elderly in the neighborhoods of cluster1. Lots of of citizens of over 65 years of age live here, both in numbers and percentually.

General practitioners who want to start a new practice might also choose for cluster1, as elderly over there might prefer to live closer to one. A general practitioner might also opt for a neighborhood in cluster0. There are not many people living in some of those neighborhoods. However, percentually there are living a lot of citizens of over 65 years of age and the distance to a general practitioner is relatively high.

Thank you for your attention 😊