Capstone Project

The Battle of Neighborhoods (Week 1)

Which neighborhoods are best suited for the elderly?

Instructions

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

1. *In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea.*
2. *In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?*

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

A description of the problem and a discussion of the 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.

A description of the data and how it will be used to solve the problem

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.

- 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

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.

The databases and cleaned dataframe

Our first uncleaned imported database looked like this:

	gwb_code_10	gwb_code_8	regio	gm_naam	recs	gwb_code	ind_wbi	a_inw	a_man	a_vrouw	a_00_14	a_15_24	a_25_44	a_45_64	a_65_oo	a_ongeh	a_geh
0	NL00	0	Nederland	Nederland	Land	NL00	.	17282163	8581086	8701077	2739819	2131944	4255450	4840946	3314004	8374133	670
1	GM0003	3	Appingedam	Appingedam	Gemeente	GM0003	.	11721	5720	6001	1718	1238	2376	3491	2898	4981	
2	WK000300	300	Wijk 00	Appingedam	Wijk	WK000300	1	11720	5720	6000	1715	1235	2375	3490	2895	4980	
3	BU00030000	30000	Appingedam-Centrum	Appingedam	Buurt	BU00030000	1	2370	1135	1230	230	225	500	715	700	985	
4	BU00030001	30001	Appingedam-West	Appingedam	Buurt	BU00030001	1	3035	1505	1530	480	350	570	990	640	1285	

Our second uncleaned imported database looked like this:

	Unnamed: 0_level_0	Unnamed: 1_level_0	Unnamed: 2_level_0								
	Onderwerp	Onderwerp	Onderwerp	Utrecht	Wijk 01 West	Welgelegen, Den Hommel	Oog in Al	Halve Maan-Zuid	Halve Maan-Noord	Lombok-Oost	Leidseweg en omgeving
0	RegioaanduidingGemeentenaam	RegioaanduidingGemeentenaam	naam	Utrecht	Utrecht	Utrecht	Utrecht	Utrecht	Utrecht	Utrecht	Utrecht
1	Soort regio	Soort regio	omschrijving	Gemeente	Wijk	Buurt	Buurt	Buurt	Buurt	Buurt	Buurt
2	Codering	Codering	code	GM0344	WK034401	BU03440111	BU03440112	BU03440113	BU03440114	BU03440121	BU03440122
3	Indelingswijziging wijken en buurten	Indelingswijziging wijken en buurten	code	.	1	1	1	1	1	1	1
4	Gezondheid en welzijnHuisartsenpraktijkAfstand...	Gezondheid en welzijnHuisartsenpraktijkAfstand...	km	06	05	07	04	04	07	05	06

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

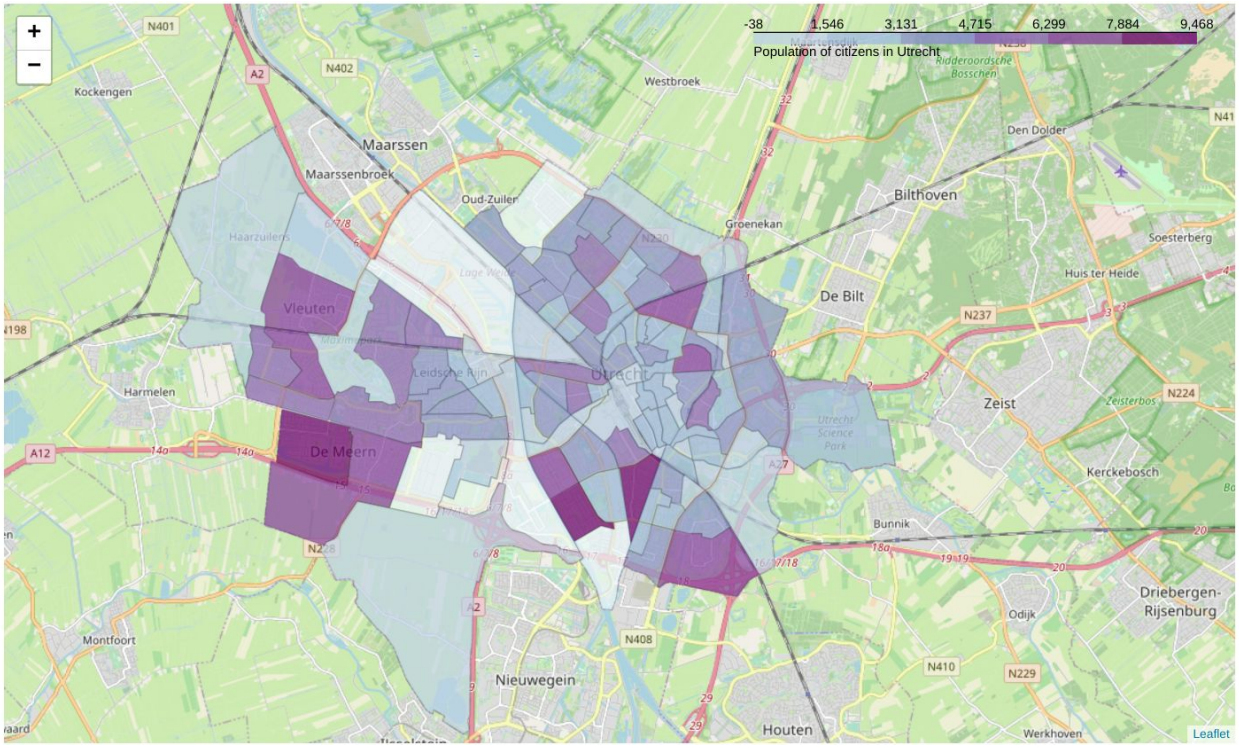
	neighborhood	total	over65yo	percentage_over65yo	distance_gp_km
0	Welgelegen, Den Hommel	1500	340	22.666667	0.7
1	Oog in Al	4370	435	9.954233	0.4
2	Halve Maan-Zuid	1465	220	15.017065	0.4
3	Halve Maan-Noord	1710	145	8.479532	0.7
4	Lombok-Oost	2290	195	8.515284	0.5

Methodology section

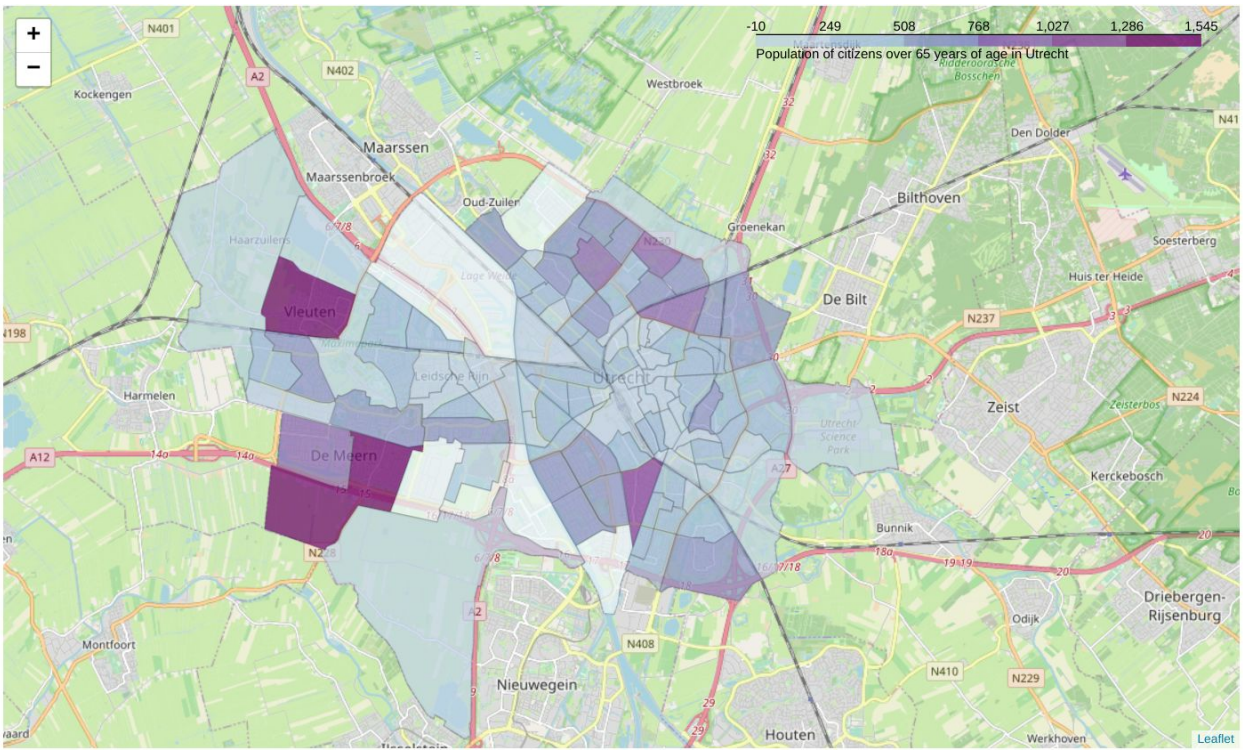
This represents the main component of the report where we discuss and describe the exploratory data analysis that we did, the inferential statistical testing that we performed, and what machine learnings were used and why.

Choropleth maps

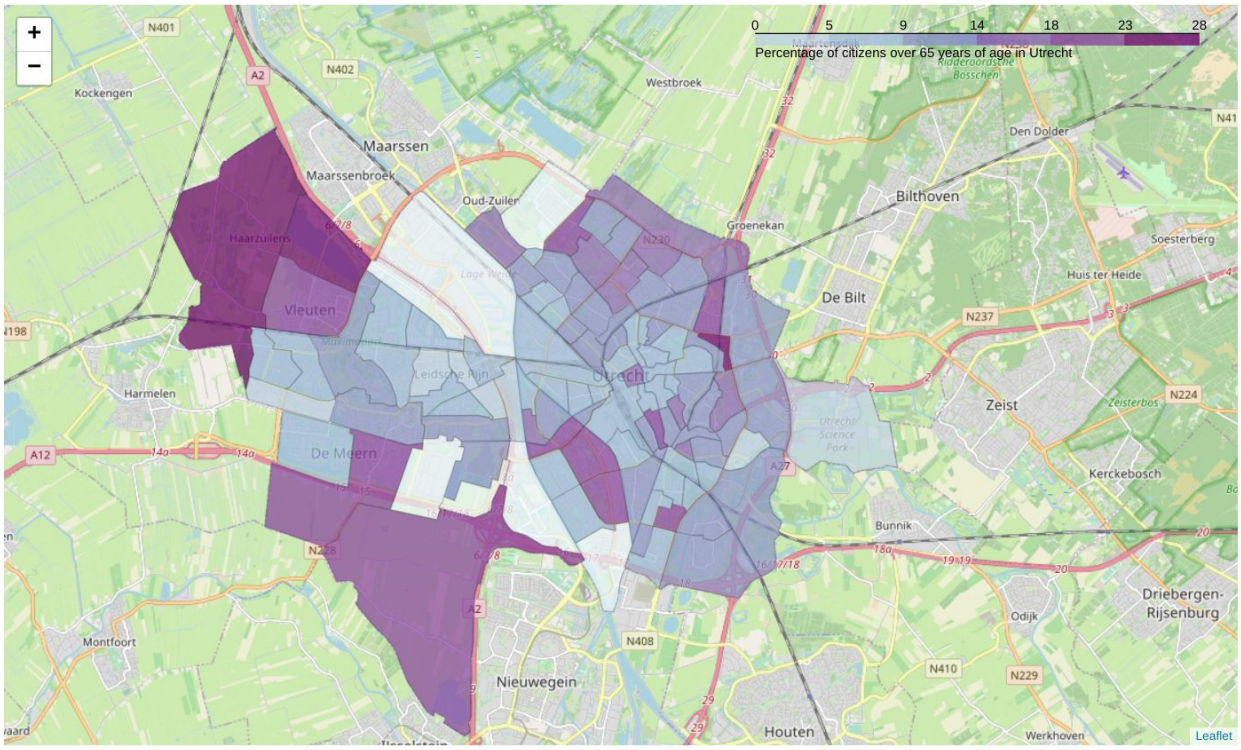
Population of citizens in Utrecht



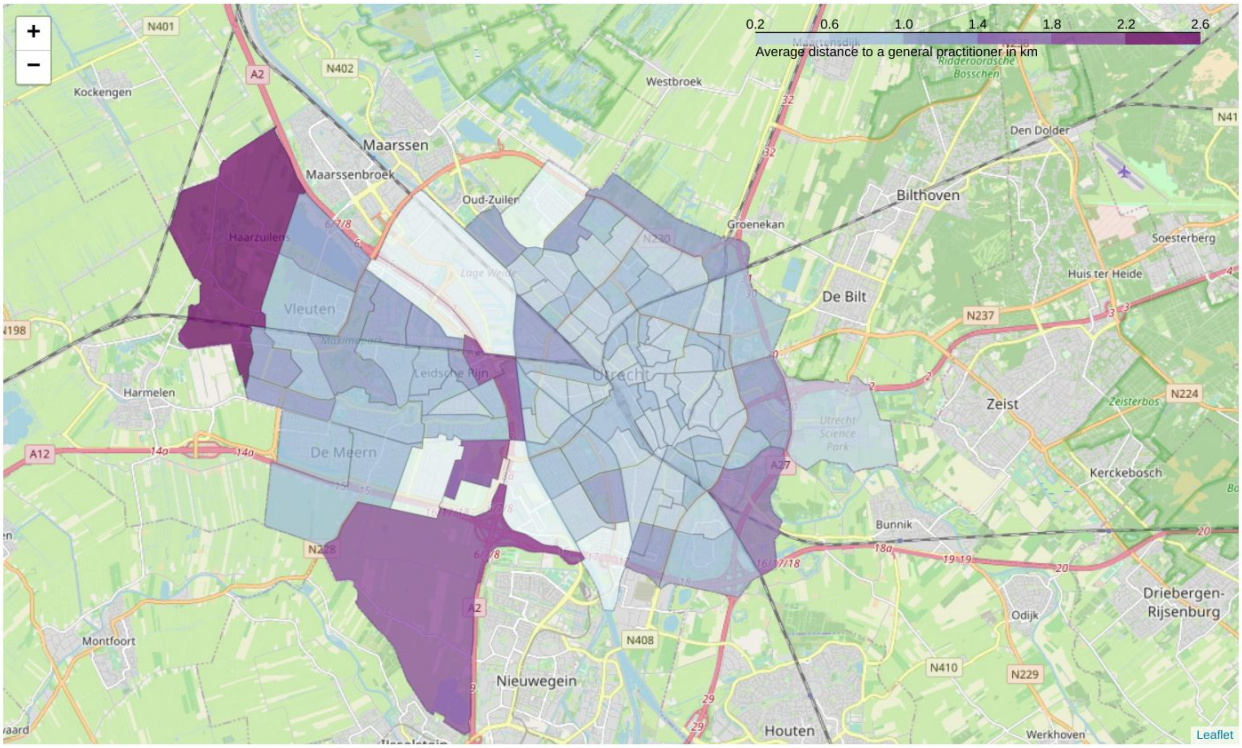
Population of citizens over 65 years of age in Utrecht



Percentage of citizens over 65 years of age in Utrecht

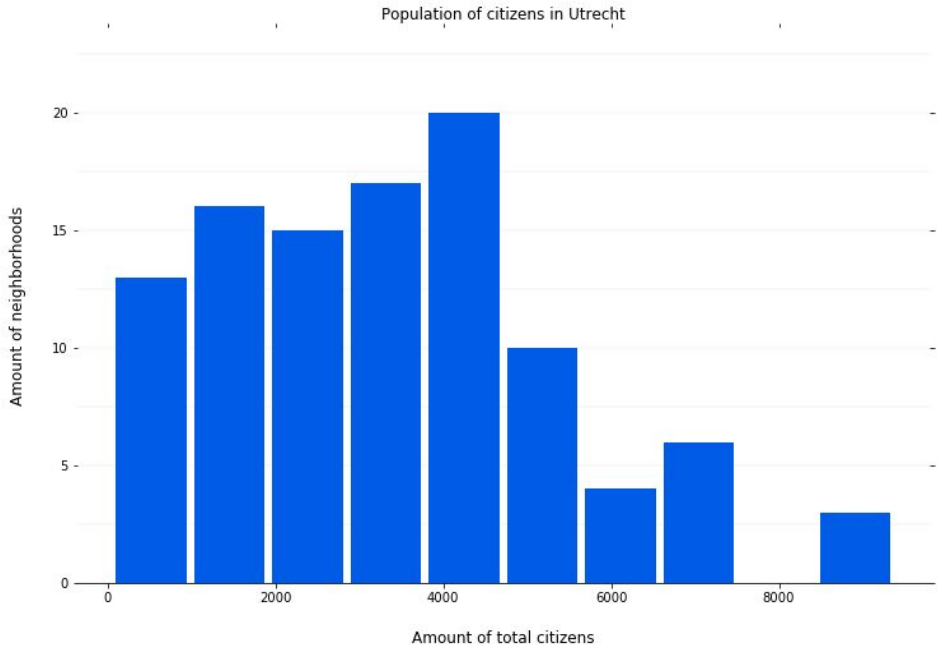


Average distance to a general practitioner in km

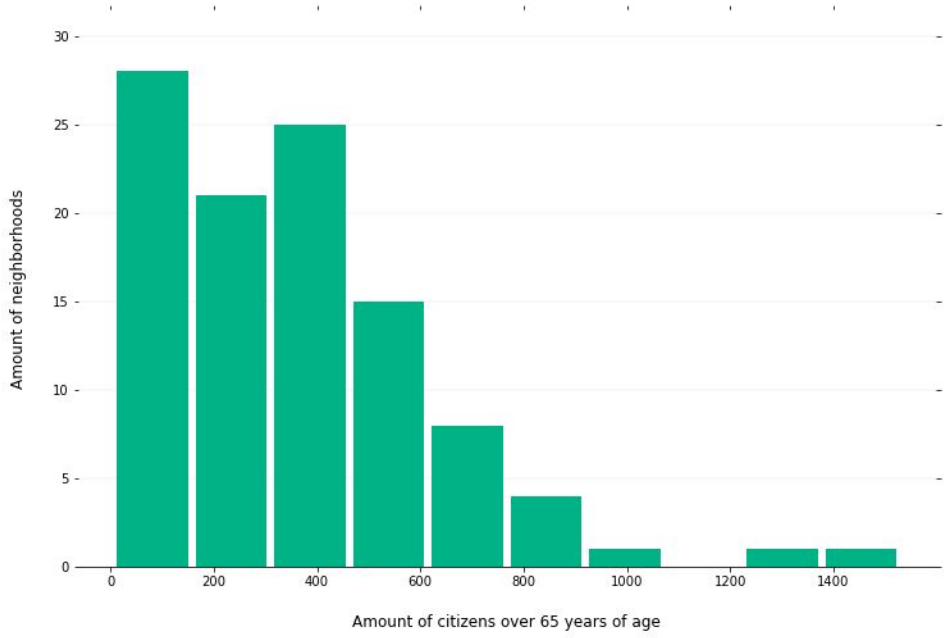


Histograms

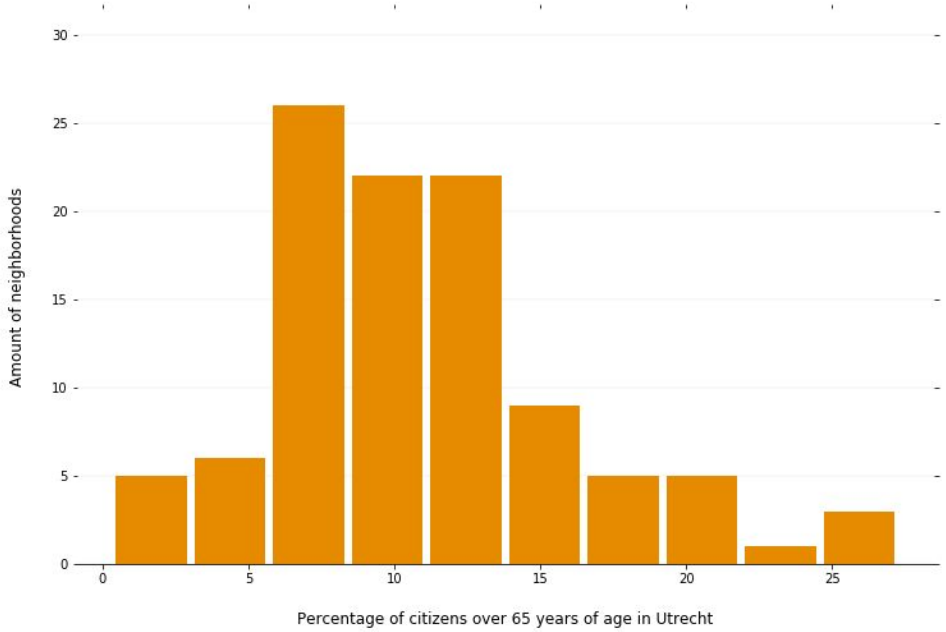
Population of citizens in Utrecht



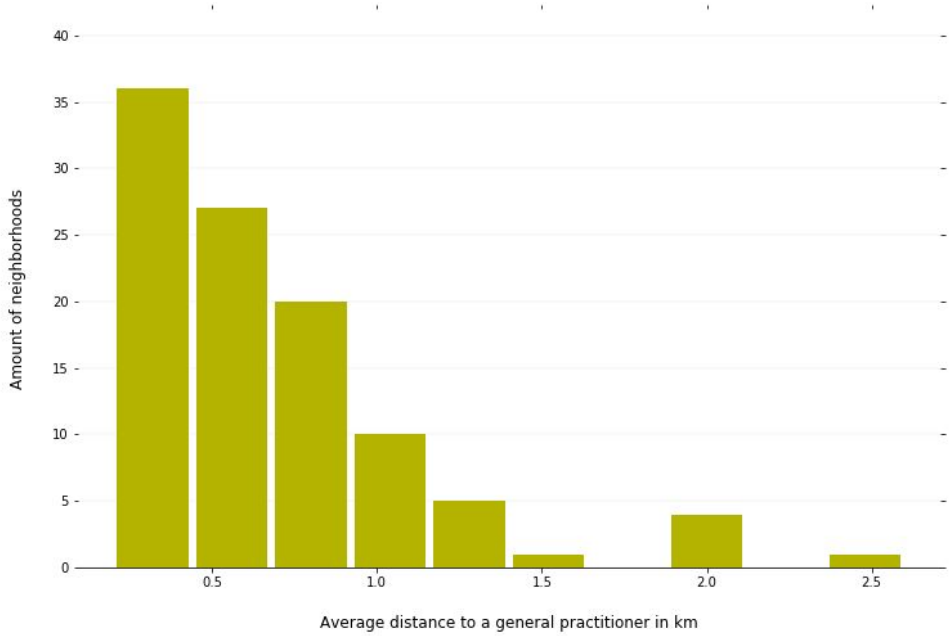
Population of citizens over 65 years of age in Utrecht



Percentage of citizens over 65 years of age in Utrecht

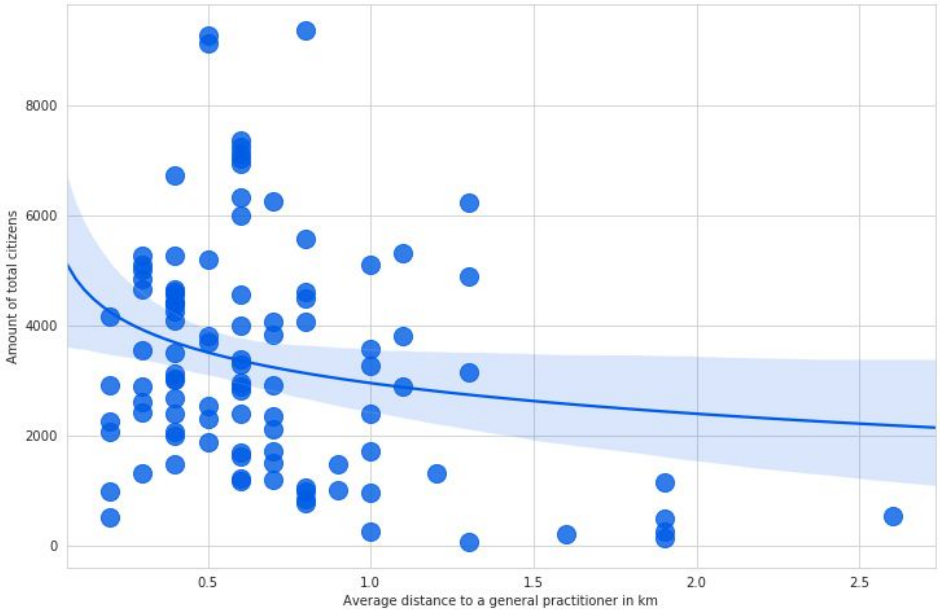


Average distance to a general practitioner in km

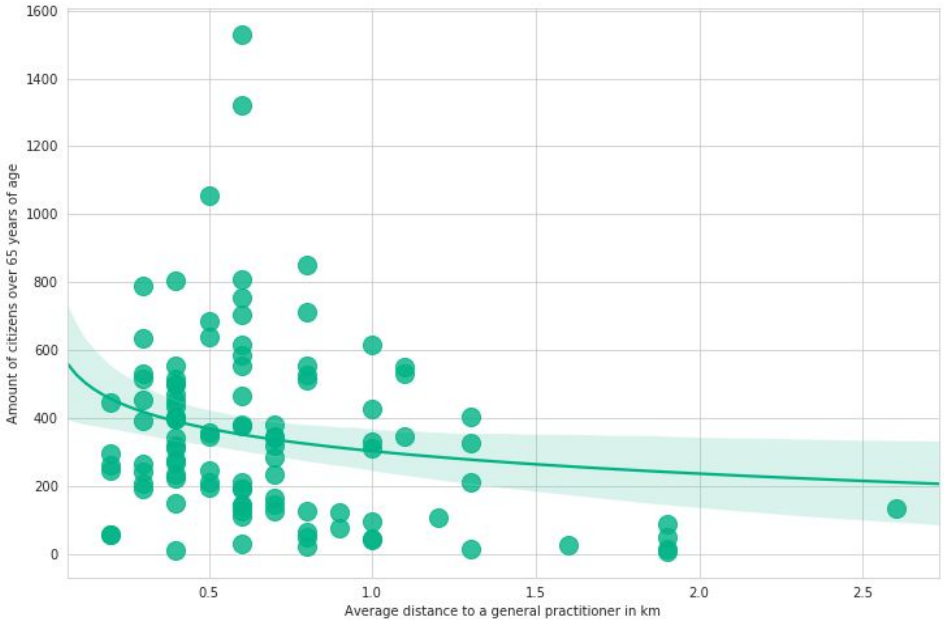


Regressionplots

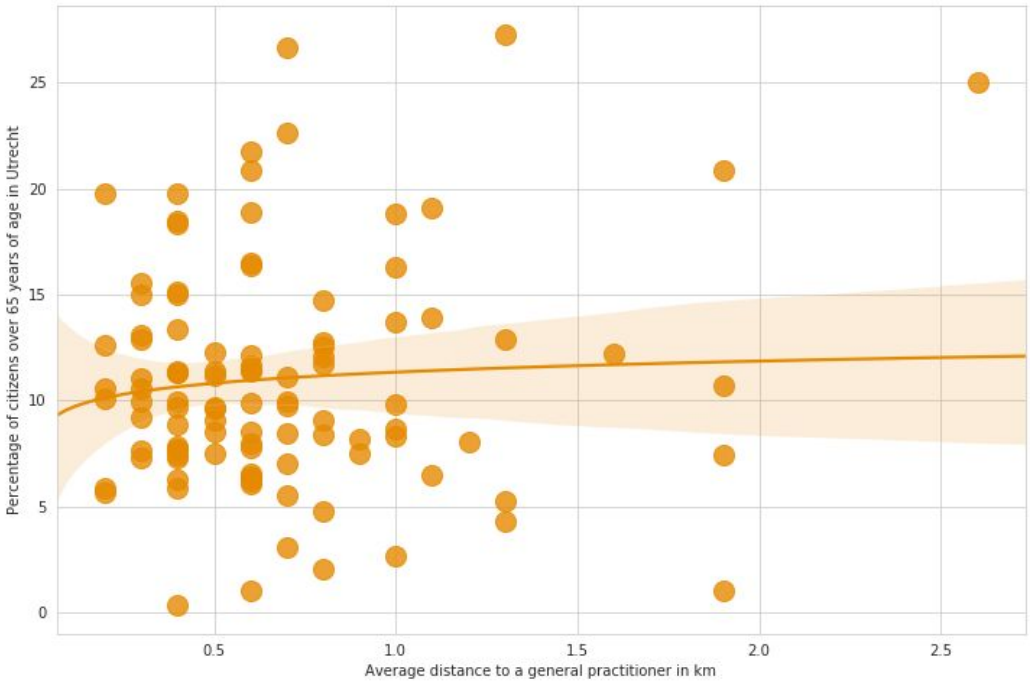
Population of citizens in relation to average distance to a GP



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



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

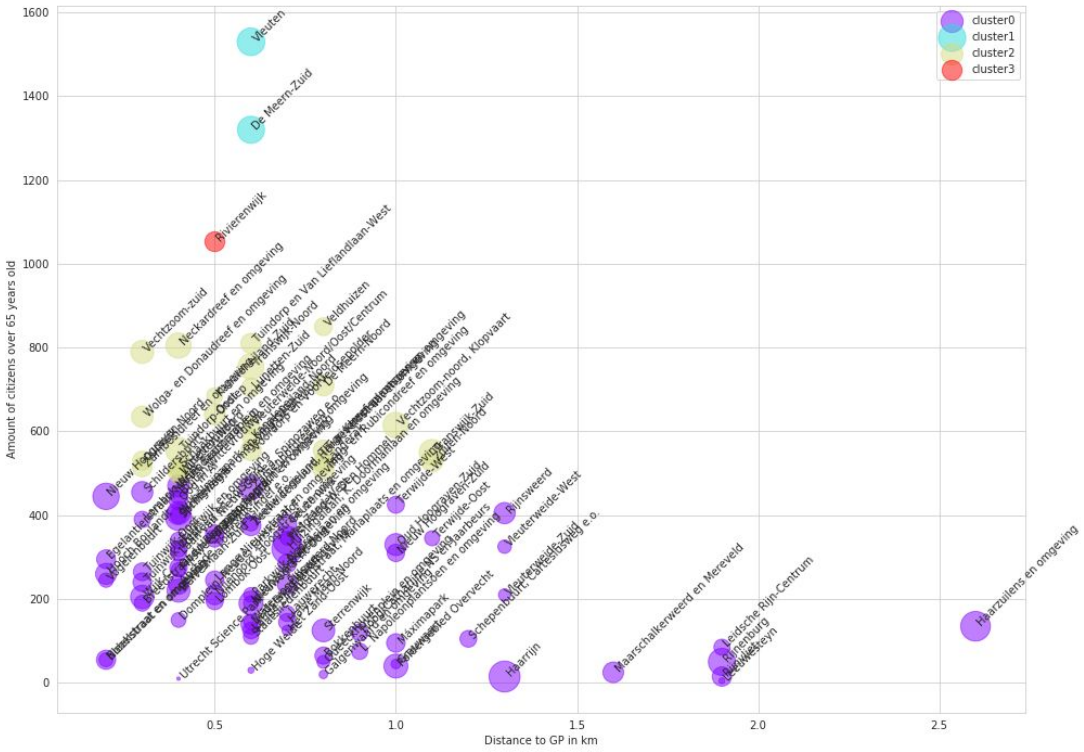


Results section

In this section the results are discussed.

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.

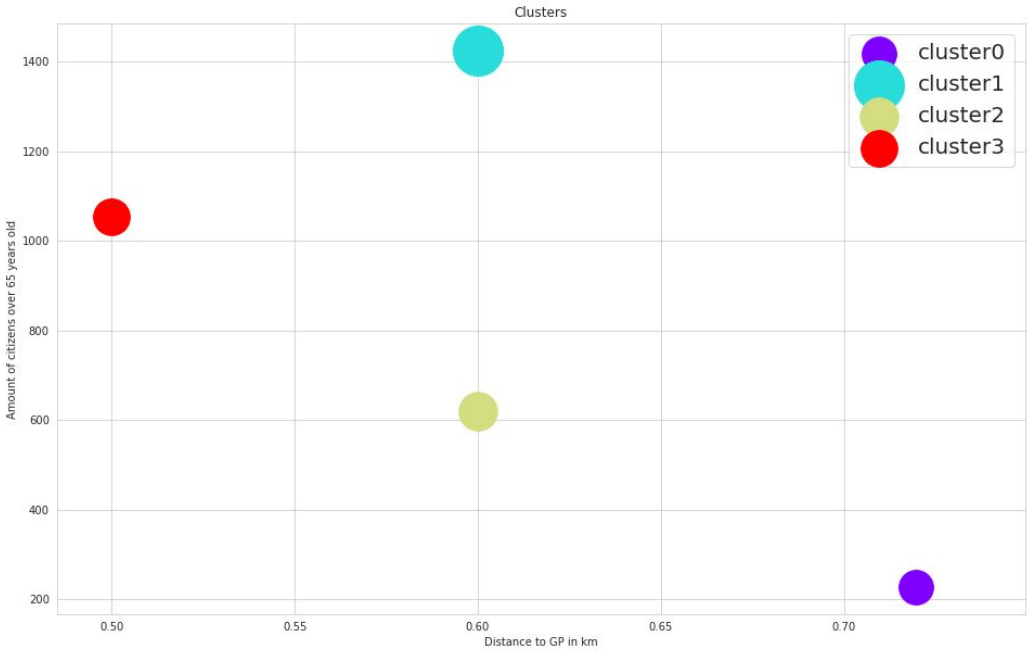
Here we see the population of citizens over 65 years of age in relation to average distance to a general practitioner. The size of the blobs represents the percentage of citizens over 65 years of age in that particular neighbourhood.



Discussion section

In this section we will discuss the observations we noted. We will also make recommendations based on the results.

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.



Conclusion section

In this section we will conclude the report.

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.