

# Programming 1 - HFJJ VAARTSTRA 352664

## An analysis into how urbanity has influence on once health.

This document was created alongside the file FINAL-ASSIGNMENT-352664.ipynb and is intended to give more background and information to go with the python file.

### 1. The region and the domain category the datasets are about

The data analysis was done on townships on the province of Groningen. The analysis takes on:

- Groningen, Het Hogeland, Delfzijl, Midden-Groningen, Oldambt, Pekela, Stadskanaal, Veendam, Westerkwartier.

Note: The township Delfzijl has merged with Appingedam and Loppersum as of the 1<sup>st</sup> of January in 2021 and has since been called Eemsdelta. The coordinates found originated from before this and is thus only for Delfzijl. The mental health data found from the GGD does include the newly formed township of Eemsdelta. For the sake of visualising, the name Delfzijl has been chosen to ensure no further confusion when looking at the visualized map.

The domain the analysis looks into is mental health. A mental health survey results page was found from the GGD Groningen and was used as mental health data set.

### 2. The research question

Taking the above two mentioned data source resulted in the following research question "Does urbanity have an influence on once mental health?". Which researches the correlation between several mental health aspects (divided into age groups of junior and adults) and the urbanity of their hometown.

### 3. The data storage and processing approach

The data used originates from different sources:

Data source	Data type	Location
Addresses per km2	Webpage (html text)	<a href="https://allecijfers.nl/gemeente/groningen/">https://allecijfers.nl/gemeente/groningen/</a>
Mental health questionnaires	CSV file	<a href="https://ggdgroningen.incijfers.nl/dashboard/dashboard/psychische-gezondheid/">https://ggdgroningen.incijfers.nl/dashboard/dashboard/psychische-gezondheid/</a>
Township coordinates (middle points)	Webpage	<a href="https://www.coordinatenbepalen.nl/coordinates/620-delfzijl">https://www.coordinatenbepalen.nl/coordinates/620-delfzijl</a>
Township coordinates (borders)	Geojson	<a href="https://data.overheid.nl/dataset/5581-gemeentegrenzen-nederland-gegeneraliseerd-vlak-bestand">https://data.overheid.nl/dataset/5581-gemeentegrenzen-nederland-gegeneraliseerd-vlak-bestand</a>

One additional source was used to determine urbanity which is mentioned in the dataframes.

- Determining urbanity (source webpage) <https://www.cbs.nl/nl-nl/nieuws/2019/44/meeste-afval-per-inwoner-in-minst-stedelijke-gemeenten/stedelijkheid>

The webpages, which provide html text, were handled by using a get command from python. This scraped the entire webpage and put it in an html text in python after which the needed information was found with the 'find' method from python.

The CSV file was downloaded, loaded in and finally reshaped and reformatted to proper formatting into a dataframe.

The Geojson was loaded as a json file, handled as not the entire file was needed, but was kept as a json file.

#### **4. The analysis approach**

To determine which plots to use and what would be showing off the data that would be useful, the data was firstly analysed manually. Checking on missing data, on what the data represented and how the data was structured. E.g. some indications were negatively based and some positively, thus this needed to be handled before any visualisations were made.

#### **5. The data visualisation approach**

##### *Bar plots*

I have chosen to firstly visualise data in bar plots to see how the data was divided per indication and per urbanity. The latter one to give a first insight into if any indicators differed in the rural townships from the urban townships.

##### Scatterplots and heatmap

Scatterplots were made to see if I could visualise any correlations between indicators and addresses per km<sup>2</sup> (ascending). The scatterplots were based on the original data.

The heatmap has the same function, to see if any indicators have correlations with urbanity. Except for the heatmap, absolute correlated numbers were used which were calculated with the correlation method from python.

##### Map

The map was created to visualise the scores of the mental health survey and to clearly see if these differ per township and to be more specific, if indicators differ on a townships urbanity.