# Introduction:

Housing prices in Utrecht, Netherlands are high, this causes trouble for many people. A lack of transparency in pricing only makes this worse. People need to have the right information available when making these purchase decisions. However, for many it is unclear why prices are much higher in some areas than in others. Should this affect your purchase decision?
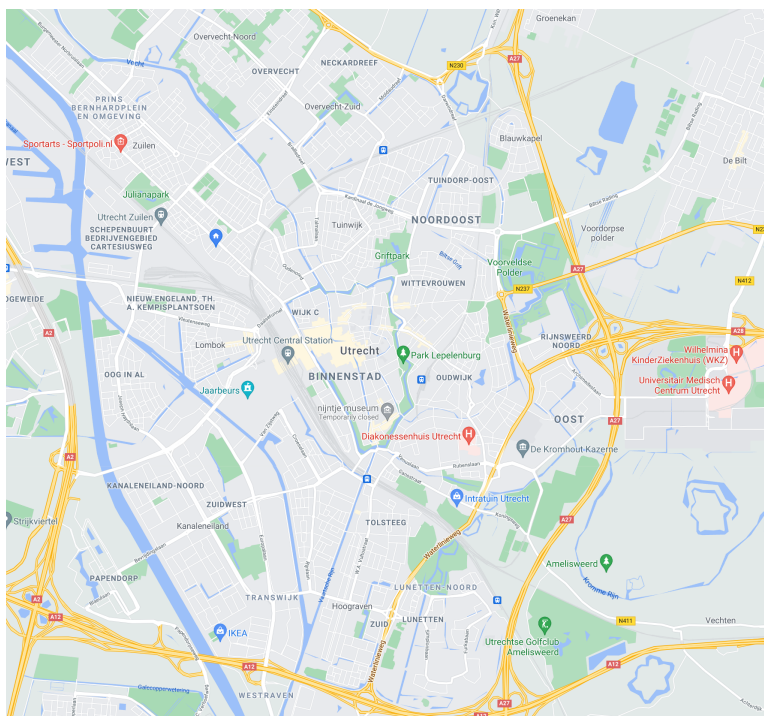
The beautiful city of Utrecht, Netherlands

Breaking down the price of housing in Utrecht allows people to understand this phenomena and act accordingly. Because neighbourhood data is used in this analysis, it could help when looking for new and upcoming areas to live or invest in. In this assignment I will combine Utrecht housing price data, Utrecht neighbourhood data, and Foursquare API data to compare housing prices in neighbourhoods in Utrecht.

Apart from the benefits to the Utrecht community, this information can help the government understand various factors influencing housing pricing on a local level which can lead to improvements in housing policies to combat the rise of unaffordable housing.

(Data: next page)

Data:

The data used in this analysis is from 3 sources. Utrecht neighbourhood data, Utrecht housing price data, and Foursquare API.

Utrecht neighbourhood data. The Dutch Central Bureau for Statistics (CBS) has a vast amount of data available on neighbourhoods. I have used the data set on neighbourhoods to find information on the number of inhabitants, children, addresses, neighbourhood size, and more. This data is available on various government levels. Initially, the idea was to use this data on 'Buurt' level (which is essentially a sub-neighbourhood), because utrecht has more than a 100 of those. However because of the lack of housing price data on such a small level of granularity, I had to make the choice to go with 'Wijk' level (large/parent neighbourhood). In the methodology section I will discuss the implications of that choice (significantly less data points).

	ID	WijkenEnBuurten	Gemeentenaam_1	SoortRegio_2	Codering_3	IndelingswijzigingWijkenEnBuurten_4	AantalInwoners_5	Mannen_6	Vrouwen_7	k_0Tc
0	14564	Wijk 01 West	Utrecht	Wijk	WK034401	1	29270	14210	15055	
1	14576	Wijk 02 Noordwest	Utrecht	Wijk	WK034402	1	45255	22060	23195	
2	14590	Wijk 03 Overvecht	Utrecht	Wijk	WK034403	1	34295	17200	17095	
3	14601	Wijk 04 Noordoost	Utrecht	Wijk	WK034404	1	39680	18780	20895	
4	14613	Wijk 05 Oost	Utrecht	Wijk	WK034405	1	32080	14905	17170	
5	14627	Wijk 06 Binnenstad	Utrecht	Wijk	WK034406	1	19165	9815	9345	
6	14639	Wijk 07 Zuid	Utrecht	Wijk	WK034407	1	27895	13685	14210	
7	14648	Wijk 08 Zuidwest	Utrecht	Wijk	WK034408	1	38620	19430	19190	
8	14656	Wijk 09 Leidsche Rijn	Utrecht	Wijk	WK034409	1	41290	20685	20605	
9	14672	Wijk 10 Vleuten-De Meern	Utrecht	Wijk	WK034410	1	49795	24465	25330	

10 rows × 118 columns

General Neighbourhood data

The next dataset is on housing prices in Utrecht's neighbourhoods. This data is from 2019, as the data from 2020 was not available at the time of this analysis. The data only includes transactions from 2019. The dataset was scraped from the web using the standard pandas scraping functionality (see notebook). This dataframe is merged with the already existing dataframe containing general neighbourhood data.

	Neighbourhood	Price_2019Q3	Price_per_m2
0	West	341.181	4.219
1	Noordwest	309.690	4.015
2	Overvecht	274.954	2.908
3	Noordoost	445.016	4.788
4	Oost	507.770	4.887
5	Binnenstad	434.747	5.091
6	Zuid	327.887	3.802
7	Zuidwest	313.324	3.661
8	Leidsche rijn	415.235	3.673
9	Vleuten-De Meern	428.002	3.562

House prices table in euros

The last dataset is coming from the Foursquare API. A central point was picked in each neighbourhood, and the coordinates of those points were added into the already existing dataset. Next, using the Foursquare API, all venues in a radius of 700m were identified. The 700 meter radius was chosen because 1) neighbourhood size is roughly 1.4km diameter, and 2) city centre reached the API return limit of 100 venues, so a bigger area would require multiple API calls. The returning values are transformed so that it represents the number of venues in that neighbourhood/area and appended to the existing dataset.

Latitude	Longitude	Venue
52.1091	5.0668	4
52.1108	5.0902	24
52.1180	5.1080	24
52.1035	5.1351	17
52.0840	5.1547	7
52.0889	5.1175	100
52.0641	5.1245	15
52.0758	5.1003	20
52.0954	5.0457	19
52.1002	5.0021	6

Neighbourhoods coordinates and number of Foursquare venues

Using these datasets, I can create variables that are believed to be more meaningful than the variables already in the dataset:

Venues	= Count API results
Children_percentage	= Children / Inhabitants
Address_density	= Addresses / Landsurface
Population_density	= Inhabitants / Landsurface
Square_meters_2019	= Price / Price per square meter