

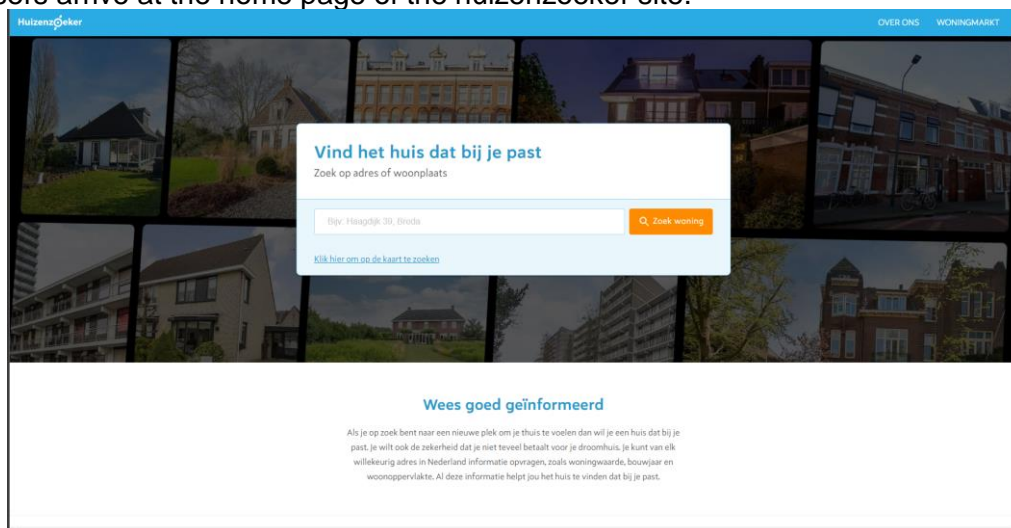
ODCM – Team project – Team 3 – Scraping the Dutch Housing Market – huizenzoeker.nl

The navigation path is defined by how an user would navigate the site in a web browser, and needs to be specified both with an entity (obtaining data on all instances of an entity) and across multiple entities (navigating from one instance of an entity to one or multiple instances of a related entity). For each single entity you extract data for, you need to extract data from multiple pages; e.g. by slightly modifying the instance's URL or via interaction (clicking, scrolling). Next, user behaviour is simulated to view how the content is dynamically loading.

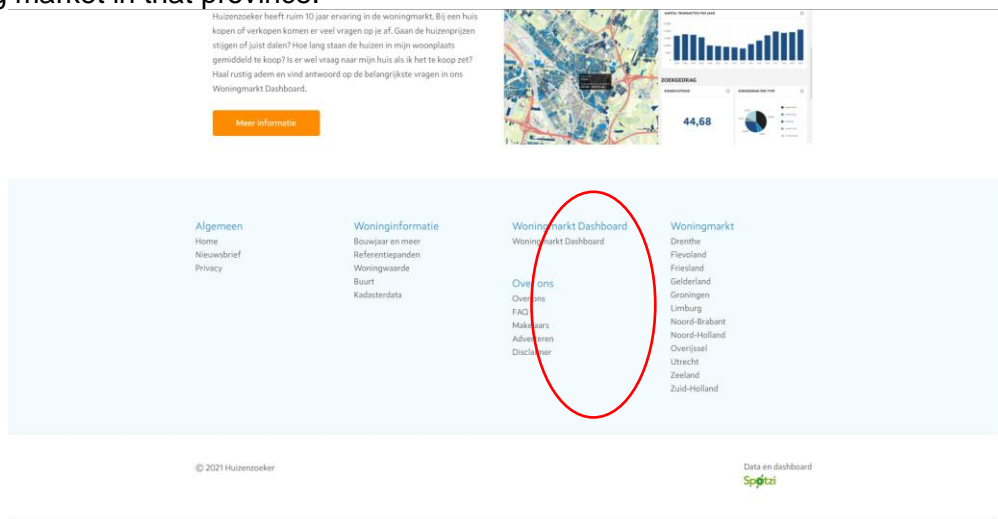
Our navigation path can be divided into several steps or 'pages' the user would visit to go from the general [huizenzoeker.nl/woningmarkt/](https://www.huizenzoeker.nl/woningmarkt/) page, to the [huizenzoeker.nl/woningmarkt/*province*](https://www.huizenzoeker.nl/woningmarkt/*province*) page, to go to the [huizenzoeker.nl/woningmarkt/*province*/municipality*](https://www.huizenzoeker.nl/woningmarkt/*province*/municipality*) page.

Step 1: Arriving at <https://www.huizenzoeker.nl/>

First users arrive at the home page of the huizenzoeker site.



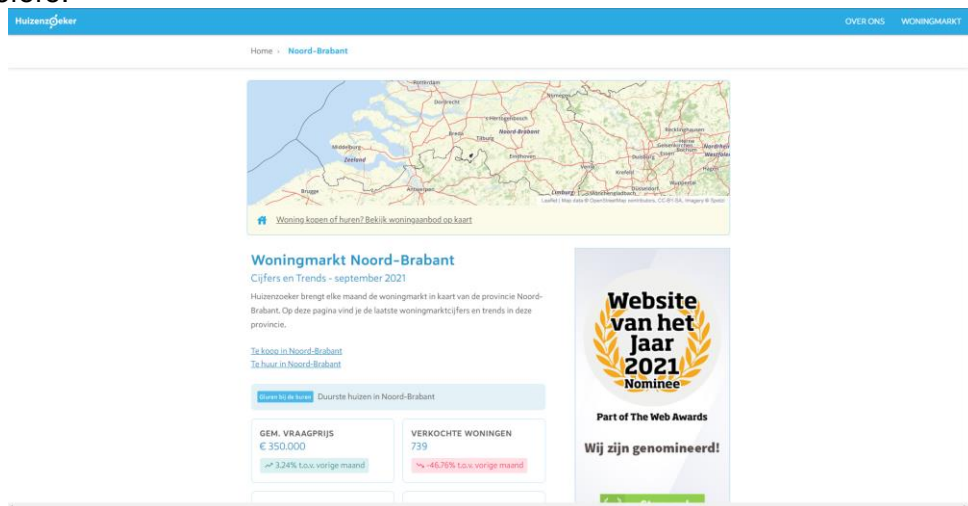
They then need to scroll down, where they will find in the bottom-right corner the 'Woningmarkt' section, which leads them to the Woningmarkt page for each respective province they decide to click on, which contains information on the current situation of the housing market in that province.



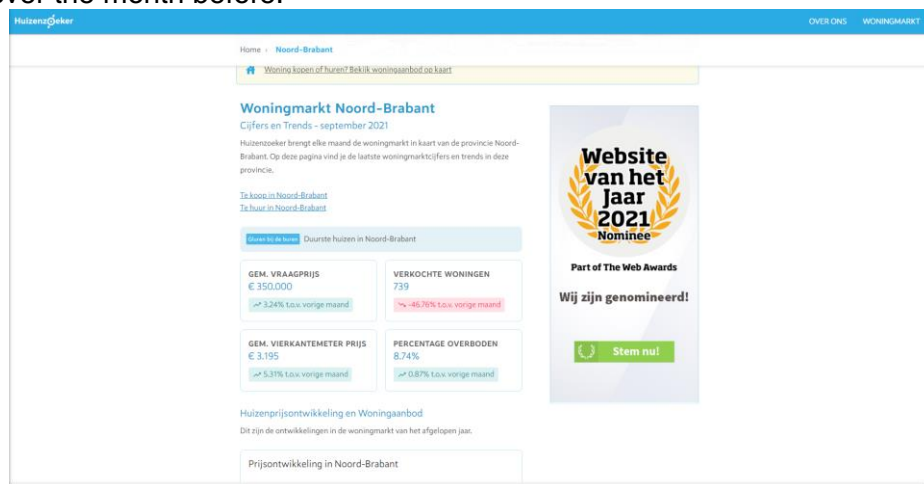
We will for now assume that the user clicks on the province 'Noord-Brabant'.

Step 2: Arriving at <https://www.huizenzoeker.nl/woningmarkt/noord-brabant/>

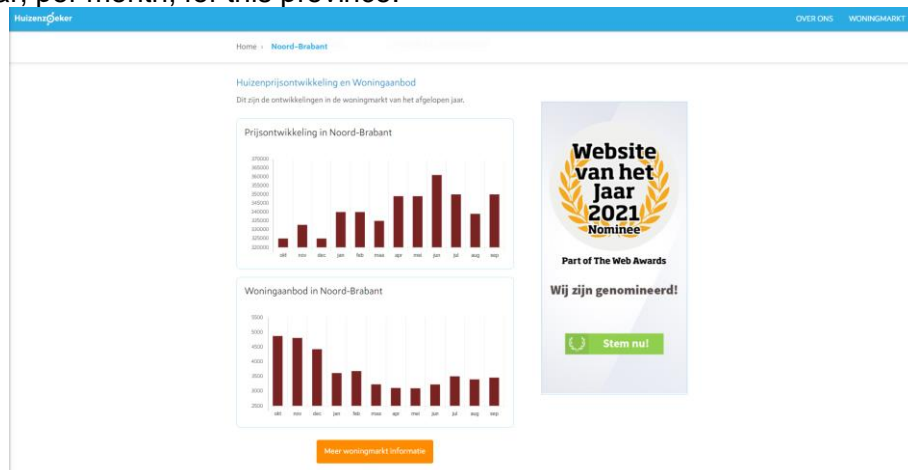
The user has now arrived at the woningmarkt page for the province Noord-Brabant. Here the user can find a lot of information about the current housing market situation in the province, for the previous month and how these figures have increased/decreased compared to the month before.



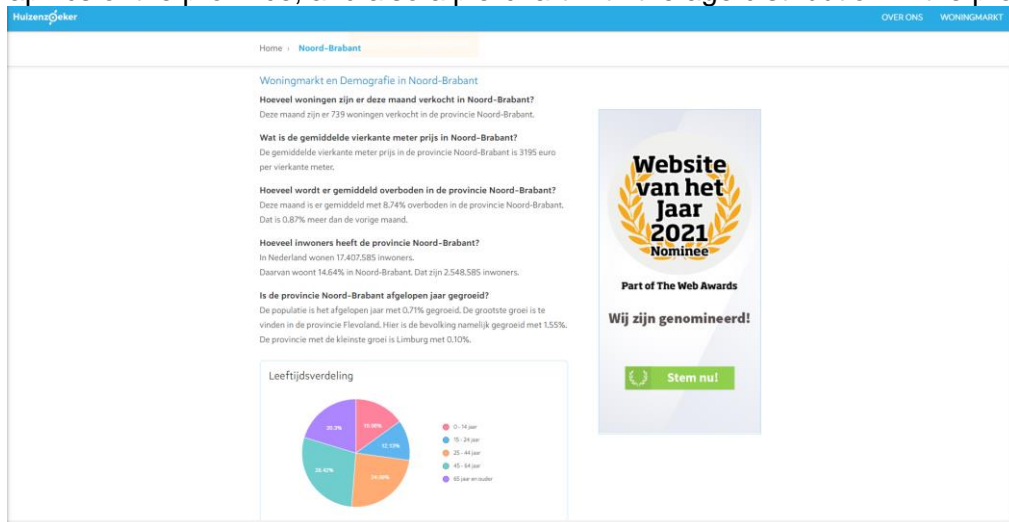
Firstly, the user will come across the trend data for the province, including the gemiddelde vraagprijs, the number of verkochte woningen, the average vierkantemeter prijs, and the percentage overboden. For each of these figures the user can see how these numbers have changed over the month before.



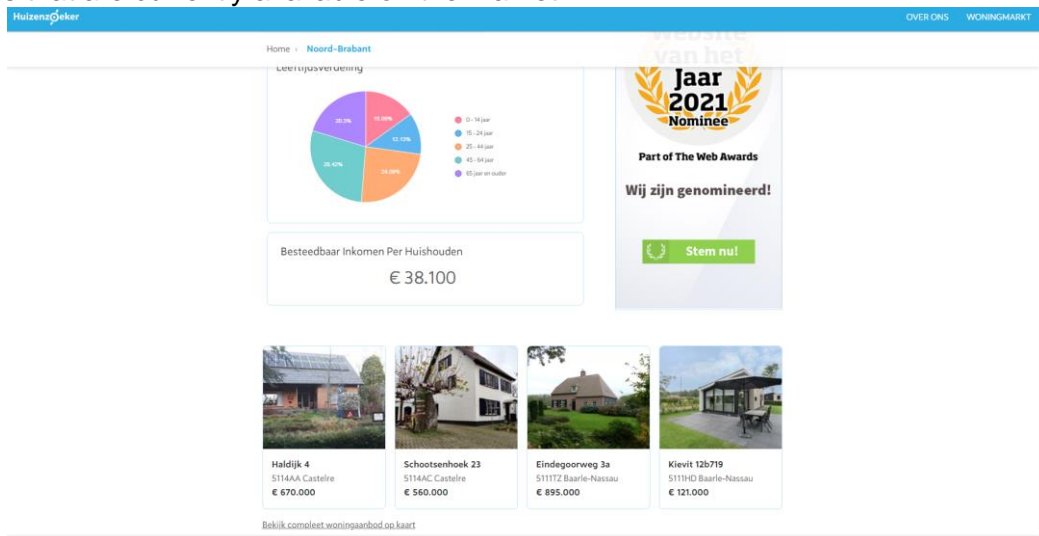
Secondly, the user sees some graphs displaying the developments in housing prices over the last year, per month, for this province.



Thirdly, there is some text that more elaborately explains the housing market and demographics of the province, and also a pie-chart with the age distribution in the province:



Then we see the average disposable income per household in the province and some houses that are currently available on the market.



Then finally at the bottom of the page, we see an overview of the different municipalities of the province, which links us to the figures on the Woningmarkt page for each municipality individually:

The screenshot shows a grid of municipalities in Noord-Brabant, organized in a 4x4 layout. The municipalities listed are:

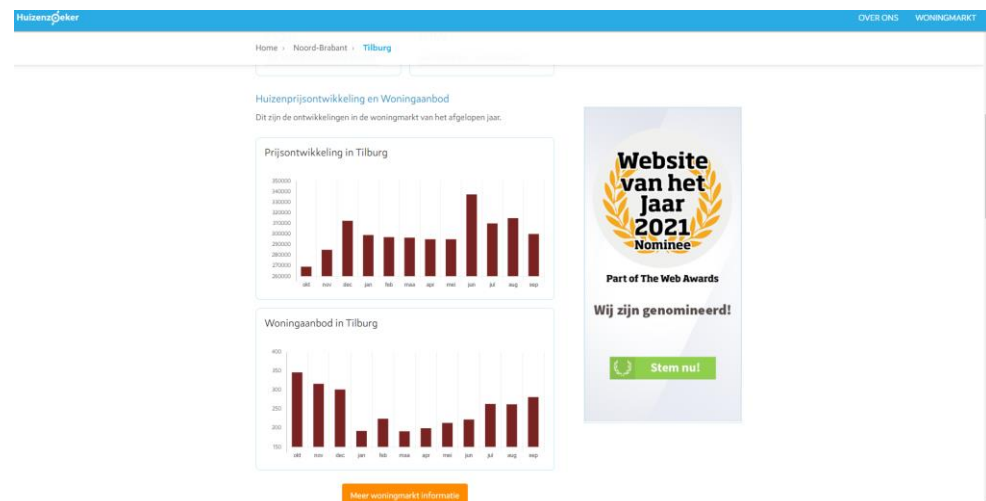
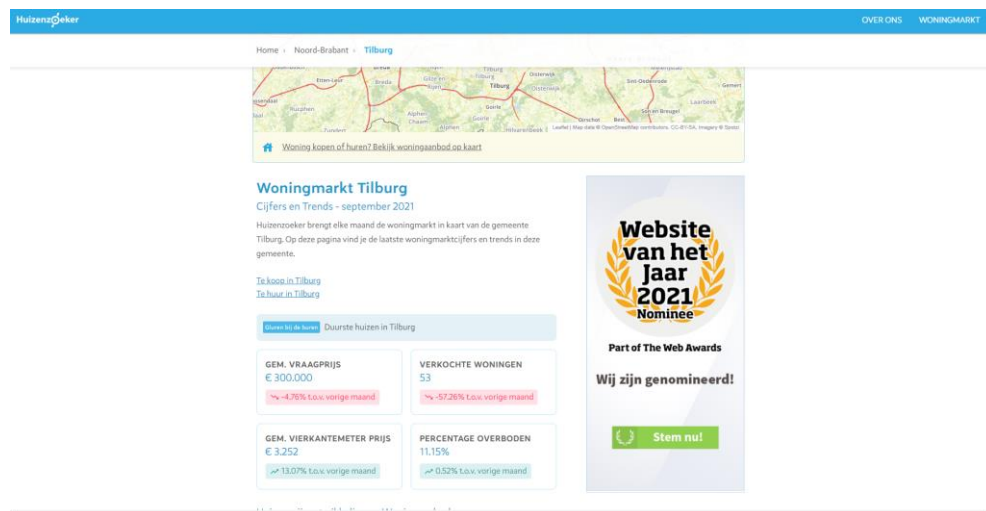
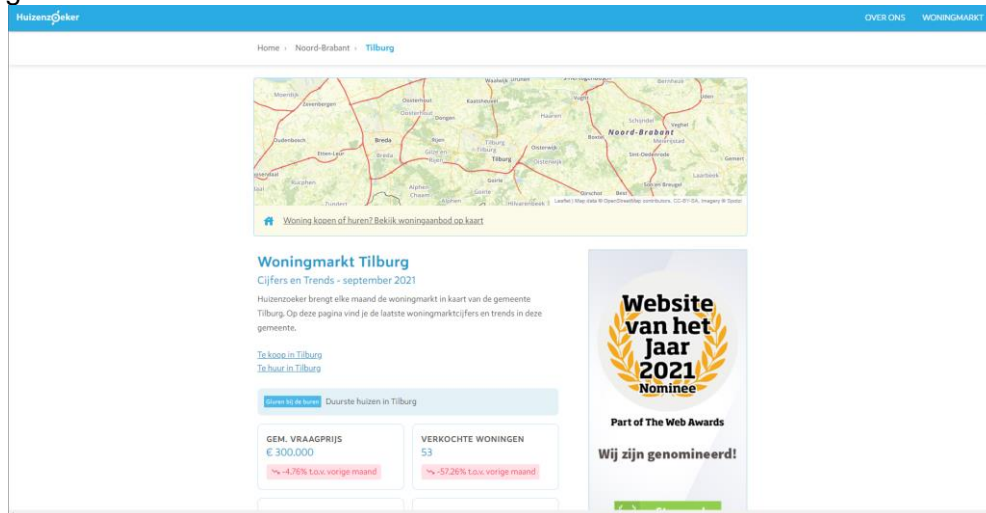
Alphen-Chaam	Altena	Asten	Baarle-Nassau
Bergeijk	Bergen op Zoom	Bernheze	Best
Bladel	Boekel	Boxmeer	Boxtel
Breda	Cranendonck	Cuijk	Deurne
Dongen	Drimmelen	Eersel	Eindhoven
Etten-Leur	Geertruidenberg	Geldrop-Mierlo	Gemert-Bakel
Gilze en Rijen	Gorinxe	Grave	Halderberge
Heeze-Leende	Helmond	Heusden	Hilvarenbeek
Laarbeek	Landerd	Loon op Zand	Meierijstad
Mill en Sint Hubert	Moerdijk	Nuenen, Gerwen en Ned...	Oirschot

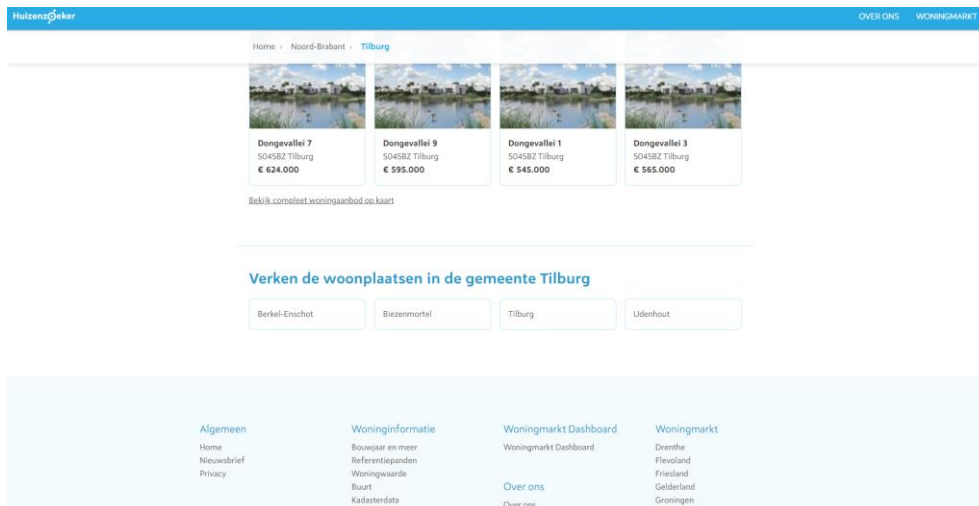
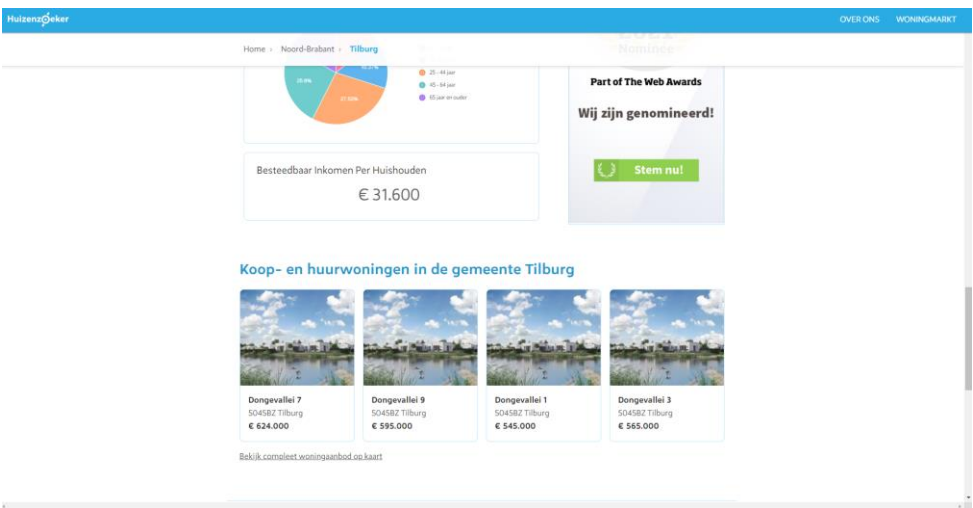
For now we assume that the user will go the municipality 'Tilburg'.

Step 3: Arriving at the <https://www.huizenzoeker.nl/woningmarkt/noord-brabant/tilburg/>

The user has now arrived at the woningmarkt page for the municipality Tilburg in the province Noord-Brabant. Here the user can find more specific information on the current housing market in the municipality Tilburg for the last month, and how these figures again have increased/decreases compared to the month before.

Similarly to the province woningmarkt pages, the municipality woningmarkt page shows similar figures:





Then at the bottom of the page under 'Verken de woonplaatsen in Tilburg', the site lists the different residences per municipality.

A visual summary of our navigation path:



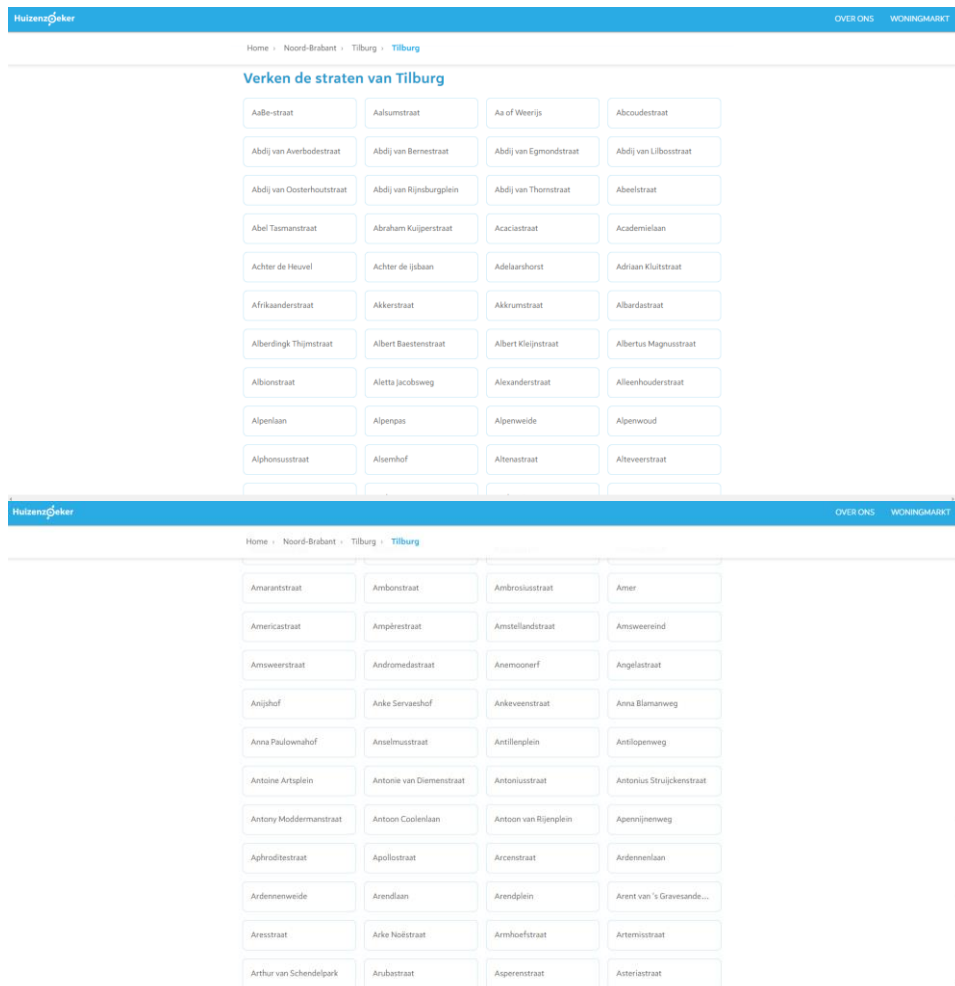
*If you try to type the ‘`www.huizenzoeker.nl/woningmarkt/`’ URL directly into the search bar it redirects you to the home page of Huizenzoeker.nl; there is thus no main page for the woningmarkt part of the site specifically.

Bonus: As an added feature to our Python scraping project, we scraped some more data for a subset of the streets in the residence of Tilburg. Namely, the data on the street-pages would have given us some extra insights into people’s considerations when deciding on where to live, which could have really enriched our dataset. Unfortunately, we were not able to scrape all streets in the Netherlands (as this simply would be too much data to parse, which would take ages). Therefore, we then decided to focus on the data of the streets of the residence Tilburg. However, pagination appeared to be more difficult than anticipated to apply to this site, as there is not a button ‘Next page’ instead you have to click on the next number. Additionally, when we click on the next page number, the URL of the page does not change, therefore many other methods to loop through multiple pages of streets would not work either. Finally, through specifying specific xpaths leading to the buttons for the next pages, we were able to scrape all data we were looking for the individual streets for all residences of the Netherlands, in theory. In practise, as discussed before, this would take too much time to parse as there would be too many streets to scrape, which the scope of our project unfortunately did not allow for.. But we worked too hard on this code to not include it, therefore we would still like to show how we would scrape the information for the last 200 listed streets of the residence of Tilburg below!

Bonus step 4: Arriving at `https://www.huizenzoeker.nl/woningmarkt/noord-brabant/tilburg/tilburg/`

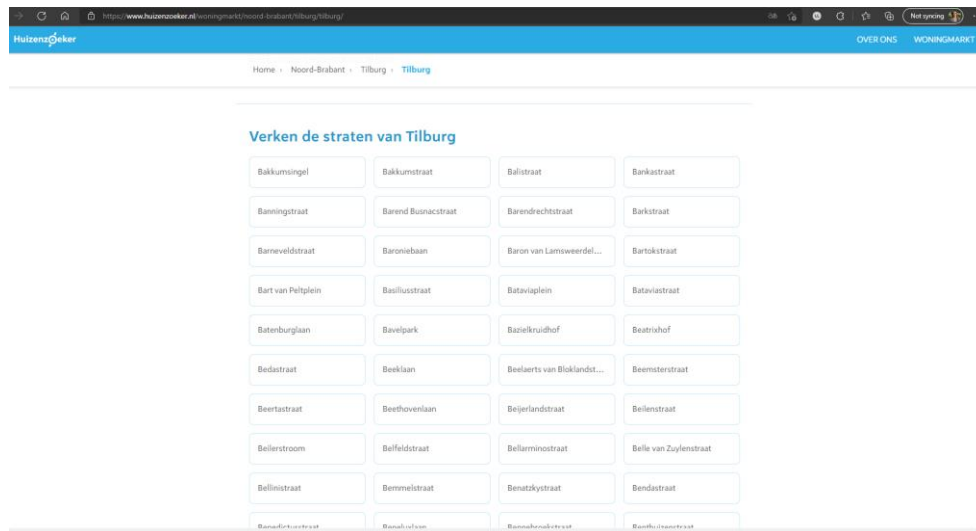
The user has now arrived at the woningmarkt page for the residence Tilburg, in the municipality Tilburg, in the province Noord-Brabant. Here the user can find the same statistics as for the municipality Tilburg, but more specific for the residence Tilburg. Therefore, no additional screenshots are included for this part of the page.

Then, at the bottom of the page under ‘Verken de straten van Tilburg’ we find a list of all streets in the residence of Tilburg.



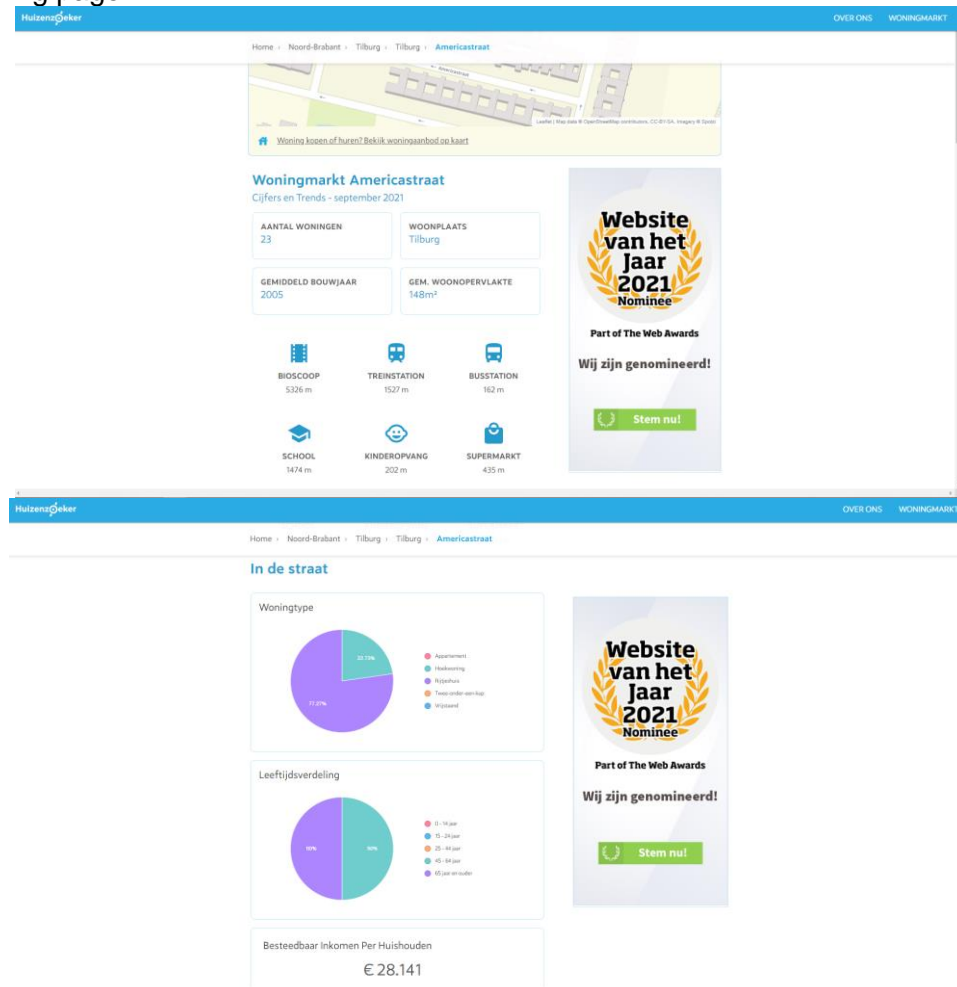
Here we illustrate that indeed there is no difference between the URL of page 1 and page 2:





Bonus step 5: Arriving at <https://www.huizenzoeker.nl/woningmarkt/noord-brabant/tilburg/tilburg/americastaat/>

If the user would then click on the street 'Americastaat' on page 1; the user would end up at the following page:



For which we scraped the information about the number of houses, the average construction year of the houses, the average living space, the disposable income per household, but also the distances to the closest cinema, train station, bus station, school, day-care and grocery store.

Huizenzoeker						OVER ONS	WONINGMARKT
Home > Noord-Brabant > Tilburg > Tilburg > Americastraat							
Alle huizen van Americastraat							
HUISNR	POSTCODE	TYPE	WONEN	BOUWJAAR	STATUS		
1	5036CK	Hoekwoning	182 m²	2005	-		
2	5036CK	Hoekwoning	166 m²	2005	-		
3	5036CK	Tussenwoning	182 m²	2005	-		
4	5036CK	Tussenwoning	166 m²	2005	-		
5	5036CK	Tussenwoning	122 m²	2005	-		
6	5036CK	Tussenwoning	166 m²	2005	-		
7	5036CK	Tussenwoning	122 m²	2005	-		
8	5036CK	Tussenwoning	166 m²	2005	-		
9	5036CK	Tussenwoning	122 m²	2005	-		
10	5036CK	Tussenwoning	171 m²	2005	-		
11	5036CK	Tussenwoning	133 m²	2005	-		
12	5036CK	Tussenwoning	166 m²	2005	-		
13	5036CK	Tussenwoning	115 m²	2005	-		
14	5036CK	Tussenwoning	166 m²	2005	-		

In that case, the navigation path would look more like this:

