Student living: mapping the market of student room listings

Scraping the Kamernet Website

Jibbe Beerens - 2099674
Lieke Geudens - 2121205
Lingwei Zhao - 2113685
Victoria Hsu -2127346

Data documentation template adapted by Hannes Datta.
Originally based on Gebru, Morgenstern, Vecchione, Vaughan,
Wallach, Daumeé, and Crawford. (2018). Datasheets for Datasets.

1. Motivation

1.1 *What primary research question, business problem, or knowledge gap motivated the creation of this dataset? How does the dataset offer insights into new phenomena, contribute to developing new models, or streamline gathering essential information? Why is this dataset valuable to the broader research community or industry stakeholders? Please provide a description about your research context.*

The contemporary Dutch student housing market is in crisis. Many students cannot find a room and are living at home longer and longer. International students are requested to return to their country. The supply in the market is increasingly decreasing while supply is increasing. This allows housing providers to charge the highest price for their accommodation. This makes it almost impossible to find a good room for the normal student. We as students at Tilburg University are also affected by this current crisis. For this reason, interest has arisen to conduct research into the current student market and to see which cities are still accessible to students looking for a room.

This dataset was created to gain more insight into the Dutch housing market with focus on the prices. The price per square meter in the different cities is taken into account in the dataset. In addition, various variables are examined, such as whether the price of the room includes or excludes utility costs, how many roommates there are, etc.

By doing this we mostly address the pathway of Improving Measurement. We scrape precise measurements on a large scale of which we make analyses and therefore we provide inferences about the student housing marketing trends. Other researchers can use this dataset for other purposes as well. Firstly of course, to understand and identify trends and patterns in the housing market, such as seasonal fluctuations in prices or differences in pricing in different areas. Also, policymakers could use this dataset to address housing affordability issues. They can for example use the data to assess the effectiveness of existing housing policies and design targeted interventions to support vulnerable populations.

1.2 *the various websites and APIs you assessed relevant to your data context, why did you choose your specific data source? Discuss the research fit, efficiency of resource use, and any other factors that made it emerge as the best choice. What extraction methods did you consider, and why did you choose the method you used? Were alternatives to web scraping evaluated? How did you ensure the scope of your data context was appropriate to maintain validity and identify any other valuable information that might be relevant? Please motivate why you selected this particular data source.*

List of our considered data sources:

Source	**Scraping or API**	**Link**
Housing Anywhere	Web scraping	https://housinganywhere.com
Housing Anywhere API documentation	API	https://developers.housinganywhere.com/
Kamer	Web scraping	https://www.kamer.nl/
Kamernet	Web scraping	https://kamernet.nl/en
Directwonen	Web scraping	https://directwonen.nl/kamers-huren/nederland
Room	Web scraping	https://www.room.nl/en/offerings/to-rent#?gesorteerd-op=prijs%2B&locatie=Regio%2BAmsterdam

There are a few other websites aimed at the student housing market, such as housinganywhere.com, which also includes kamernet.nl. But because Housinganywhere is mainly focused on international students and therefore does not only focus on the Dutch market, it was decided to scrape the data from Kamernet.nl. This provides insight into the Dutch market as it currently is for students looking for a room. Besides Kamernet, there are not many student housing websites that are popular, which means that the amount of room listings is a lot lower than that of Kamernet. This makes it not representative of the Dutch student housing market.

Because Kamernet.nl does not have an API available, only the web scraping technique was used to obtain the data.

To ensure the scope of the data context's validity, we looked at factors that impact the housing market greatly now, which are mostly the size of the room in combination with the price. This enables us to see which cities still charge a reasonable price for a living space and which cities don't.

1.3 *Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?*

This dataset was created by the students Jibbe Beerens, Lingwei Zhao, Victoria Hsu and Lieke Geudens of group 8 of the course Online Data Collection and Management at Tilburg University in winter 2024. The instructor for this course is Hannes Datta.

1.4 *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

There has been no funding or grant for the creation of this dataset.

2. Composition

2.1 *What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

Each instance in the dataset represents a room listing of a certain city. In total Kamernet features all the student cities in the Netherlands, but there are also rooms for rent in smaller villages nearby these cities.

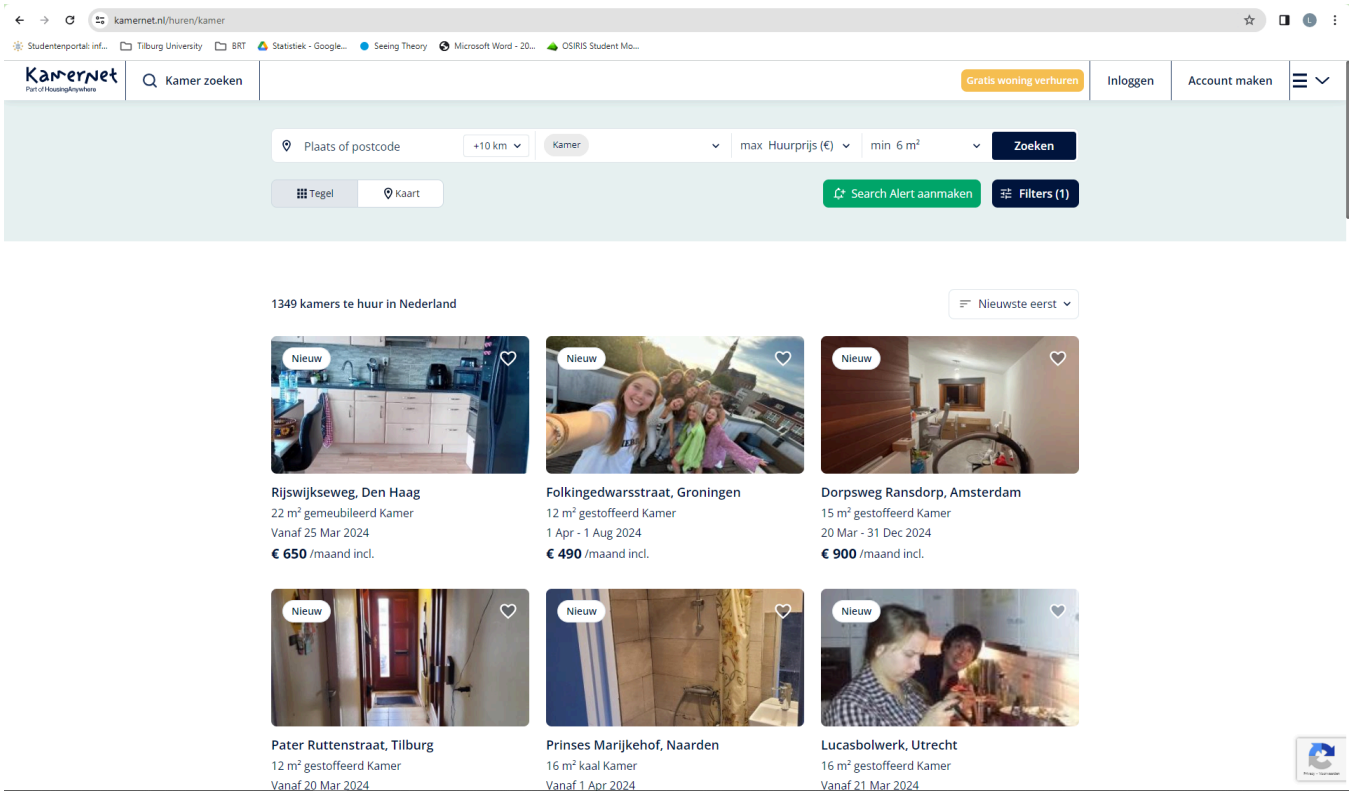


Image 1: Page where all rooms are scraped from

For each room listing the information price, square meters, including or excluding utilities, address and room details are scraped.

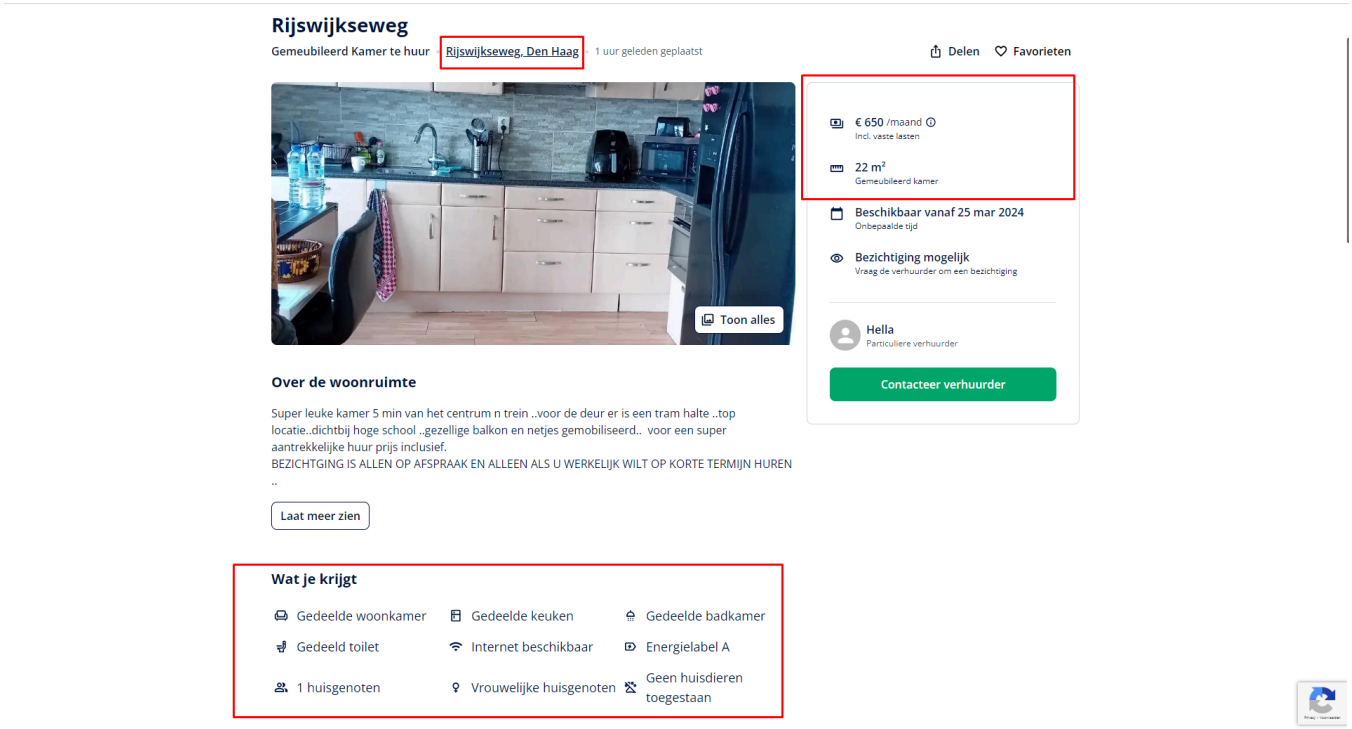


Image 2: Details of a singular room scraped

2.2 *How many instances are there in total (of each type, if appropriate)?*

To be able to make representative and valid statements about the datasets all the student rooms that are listed on the website will be scraped. As of March 20 2024, these are in total 1,350 instances divided over 75 pages. This is a momentary observation as the number of listings can change over time.

2.3 *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified (e.g., to mitigate algorithmic interference). If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).*

Yes, all instances were scraped available on Kamernet. So no sample was used.

2.4 *At what frequency was the data collected?*

The scraper ran every 24 hours at the same time everyday to make sure to catch all of the new listings being posted on Kamernet in the future. This will also give many duplicates in the dataset, but these are removed afterwards.

In total the scraper had to scrape details of 1350 rooms. Assuming it takes 5 seconds to scrape a single room it will take 6750 seconds to scrape all the rooms. Transferring this into hours and therefore dividing 6750 by 3600, the scraper will take 1,875 hours to finish scraping all the information, which is very feasible. In reality we noticed that the scraper needed around 1 hour to finish.

2.5 *What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features/operationalized variables? In either case, please provide a description.*

For each instance the following data is collected:

Variable	**Description**	**Type of data**	**Example values**
Room_id	ID of the room to remove duplicates	Unprocessed text	2210794
Price	Price of the room per month	Unprocessed text	650
Included	If the price is included or excluded utility costs	Unprocessed text	Incl. vaste lasten
Area	The number of square meters of the room	Unprocessed text	22 m2
City	The address of the listing, which contains street name + city	Unprocessed text	Den Haag

Details	The details of the room, which contains: which parts of the property are shared, if there is internet, the energy label of the property, how many housemates there are, the gender of the housemates and if pets are allowed.	Unprocessed text	- Gedeelde woonkamer - Gedeelde keuken - Gedeelde badkamer - Gedeeld toilet - Internet beschikbaar - Energie label A - 3 huisgenoten - Vrouwelijke huisgenoten - Geen huisdieren toegestaan
Time of extraction	Date of extraction to give structure to the dataset	Unprocessed text	1710935911

The date of extraction is added to each instance because the amount of listings will differ over time. It gives structure to the dataset, and it helps with deleting the duplicates. Another reason for adding the date of extraction is that it can help future researchers by investigating trends over time.

2.6 *Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? For example, if you submit multiple data files, can they be merged meaningfully? If so, please describe how these relationships are made explicit.*

All of the listings are connected to a certain city which can be used to create relationships. For example: 'The student rooms in Amsterdam are way more expensive than the student rooms in Tilburg.' Also, the same goes for the street name. Listings that have the same street name in the same city can be linked together.

2.7 *Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they ex- isted at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

The dataset relies on the external source Kamernet.nl, because it scrapes the information from there. But otherwise the dataset does not have to do with external sources. Every time data is added to the

dataset, a timestamp is added, so the dataset is basically self-contained. If a more recent version of the data is needed in the future, the scraper will have to be run again and a new data set will have to be created. The Kamernet website will have to continue to exist for this.

2.8 *In collecting this data, how did you ensure research validity is balanced with the technical feasibility of collecting the data as well as any associated legal or ethical risks?*

During the collection of the dataset we did not face many validity or legal issues. For example there were no algorithmic biases when collecting the data. There are also no legal or ethical risks with this dataset, because no personal or sensitive information is collected from the website. Only the public data of the rooms is scraped. All information needed to develop the database can therefore also be obtained from the website. This requires iterating through all pages, but there is no limit to this. Also regarding the sample size there are no real challenges as all the rooms are scraped from the website.

This makes it easy to balance the validity with the technical feasibility, because everything within the scope of this project is still feasible. The only challenge of this project is dealing with duplicates. On every page are 5 advertisement rooms, which are the same 5 rooms on every page. These duplicates are removed by using the unique room_id.

The data is ultimately stored in a JSON file to promote readability and because the data can be more easily analyzed and cleaned in R.

2.9 *Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

This dataset does not contain any confidential data, as there is no user-related information extracted from the website. All data scraped in the dataset is publicly available on the website.

2.10 *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

The dataset does not contain any data that can be offensive, insulting or threatening, as it only contains information about student rooms.

2.11 *Does the dataset relate to people? If not, you may skip the remaining questions in this section.*

Because no user information has been scraped the dataset does not relate to people and no statements about people can be done using this dataset. Only conclusions about student rooms or any information related to the student rooms can be drawn.

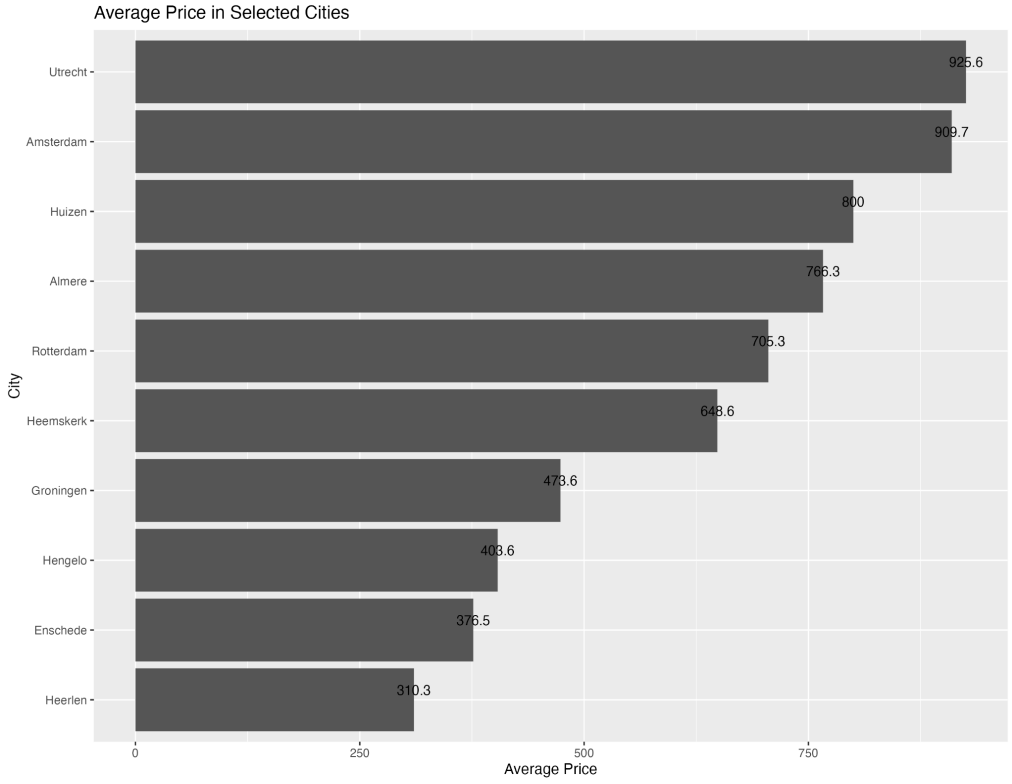
3. Data inspection

3.1 *Please provide meaningful summary statistics and plots. For example, the number of units per entity, means/SD for continuous variables, or frequency distributions for categorical variables. This part of the documentation is intended to illustrate (the richness of) the collected data.*

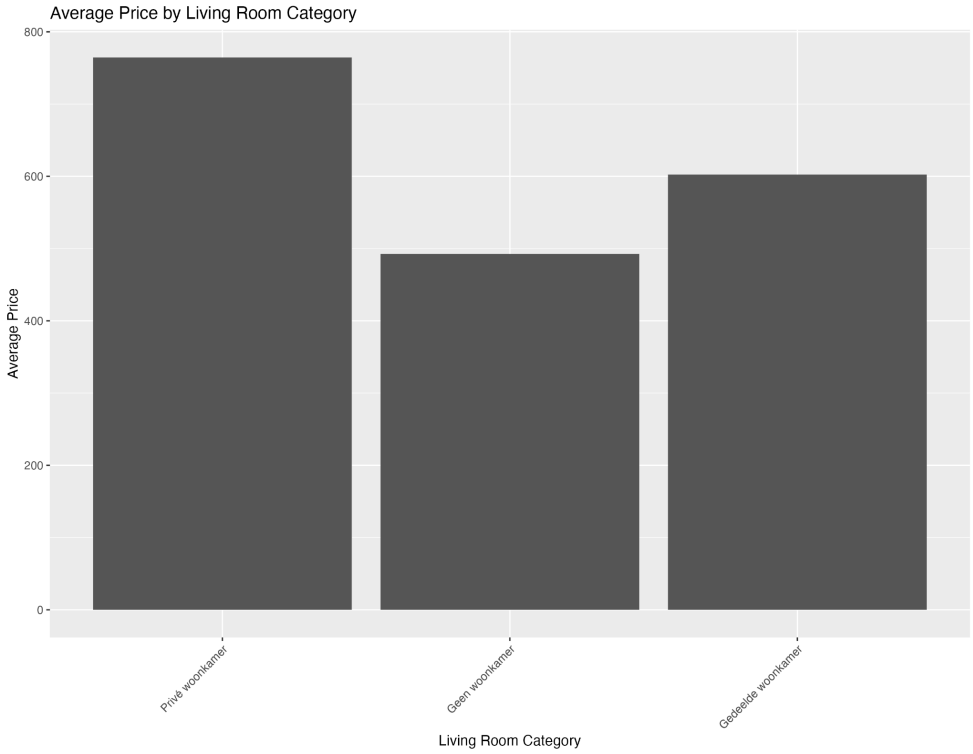
The table depicts the statistical data that was collected on the listings on Kamernet.nl. It shows the top 10 among 138 cities that have the most number of listings on the housing website.

	city	n
	<chr>	<int>
1	Enschede	260
2	Amsterdam	199
3	Utrecht	126
4	Groningen	125
5	Almere	93
6	Rotterdam	93
7	Heerlen	86
8	Hengelo	85
9	Heemskerk	76
10	Huizen	75

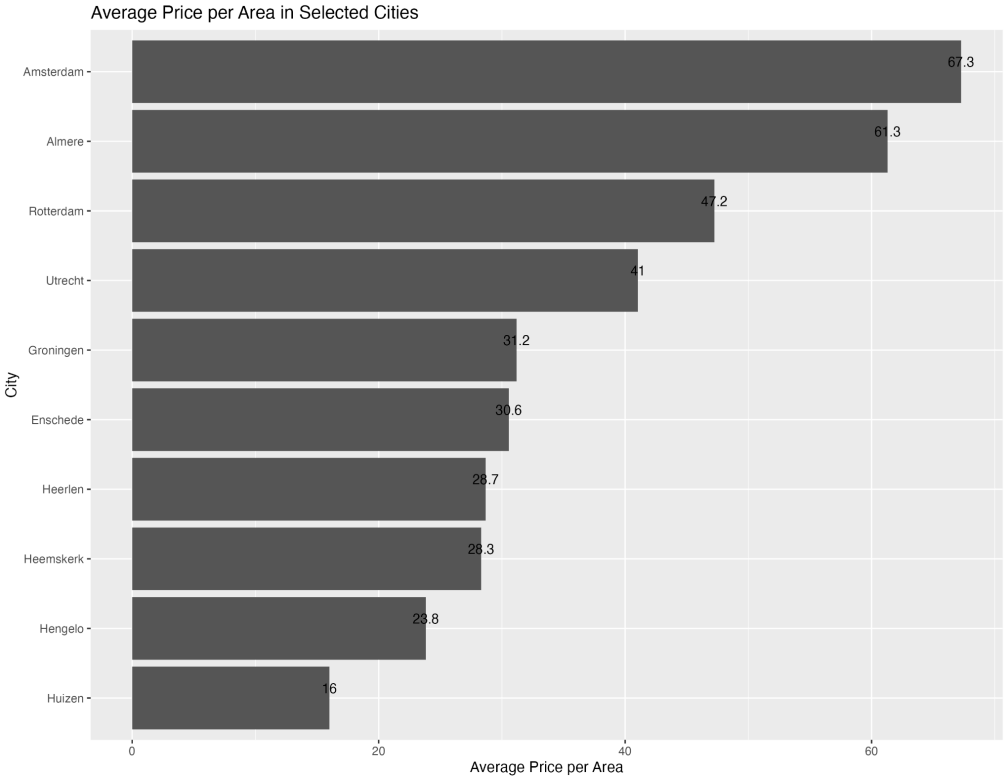
The chart presents the average price per top 10 cities listed on Kamernet.nl in descending order.



The bar chart shows the relationship between the listing living room category (shared vs. private vs. none) to the prices



This bar chart displays the average price per square meter in the top 10 most expensive cities for housing on the Dutch website Kamernet.net.



3.2 *Is any information missing from individual instances? If so, please describe why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information but might include, e.g., redacted text.*

Some housing details (column "details" in dataset) may be missing, such as:

- descriptions of shared living spaces (living room, kitchen, bathroom),
- Internet availability,
- roommate information (number and gender),
- energy efficiency ratings,
- and pet policies.

This information is optional for landlords to include, so it might not be available for every listing. Additionally, landlord contact details are often only accessible after logging into Kamernet, therefore no private information has been collected to avoid privacy and legal matters.

room_id	price	included	area	city	street	details	time of extraction
2210576	975	Incl. vaste lasten	13	Amsterdam	Ottho Heldringstraat	['Geen woonkamer', 'Gedeelde keuken', 'Gedeelde bad...	1710935522
2143851	475	Incl. vaste lasten	9	Maasland	De Herenwei	['Gedeelde woonkamer', 'Gedeelde keuken', 'Gedeelde...	1710935523
2210571	350	Incl. vaste lasten	16	Ede	Maanderweg	['Geen woonkamer', 'Gedeelde keuken', 'Gedeelde bad...	1710935527
2210572	460	Excl. vaste lasten	20	Arnhem	Spijkerstraat	['Geen woonkamer', 'Gedeelde keuken', 'Gedeelde bad...	1710935530
2206186	695	Incl. vaste lasten	12	Rotterdam	Ceintuurbaan	['Gedeelde woonkamer', 'Gedeelde keuken', 'Gedeelde...	1710935531
2208544	850	Incl. vaste lasten	12	Eindhoven	Orpheuslaan	['Geen huisdieren toegestaan']	1710935533
2203234	320	Incl. vaste lasten	7	Enschede	Bultsweg	['Gedeelde woonkamer', 'Gedeelde keuken', 'Gedeelde...	1710935534
2210296	290	Incl. vaste lasten	10	Heerlen	Meezenbroekerweg	['Geen woonkamer', 'Gedeelde keuken', 'Gedeelde bad...	1710935534

Finally, on rare occasions, there would be listings where the prices are missing, likely due to the changing HTML combined with human error - erroneous price entry. These missing values are therefore disregarded and assigned "Missing data".


```
303  [{"room_id": "2210129", "error": "Missing data"}]
304  [{"room_id": "2209972", "price": " 503", "included": "Incl. vaste lasten", "area": "16 m²", "city": " Delft", "street"
305  [{"room_id": "2210137", "price": " 800", "included": "Incl. vaste lasten", "area": "12 m²", "city": " Purmerend", "str
306  [{"room_id": "2209198", "price": " 1.400", "included": "Excl. vaste lasten", "area": "11 m²", "city": " Amsterdam", "s
307  [{"room_id": "2209971", "error": "Missing data"}]
```

4. Collection Process

4.1 *Can you describe your technical extraction plan in such a way that another researcher or team could replicate your data collection process?*

- First, we intuitively approached Kamernet.nl by selecting each city category. This means we have to create a function to loop through cities, then subsequently use Selenium to run through each page, and then use the 'For loop' to collect all rooms' URL and additional details. However, to make the extraction process easier, we decided to navigate to the Kamernet.nl/huren page, which contains comprehensive listings of available rooms on the Kamernet platform.

- Second, since we found the page that contains all rooms, we decided to use Selenium, which can automatically navigate through the website. To accomplish this, we decided to use Selenium to build code that clicks on the "Next Page button" which goes through all pages.

- Third, we made a list and used the 'While Loop' function to capture the URLs of each room page. After that we built the For loop function to capture individual page information and parse our extracted data into JSON file, ensuring the retrieval of all necessary details for further analysis.

4.2 *Why did you choose a particular data extraction technology over others (for example, why did you opt for Selenium over BeautifulSoup for website scraping, or a specific package instead of self-coded requests for APIs)?*

Since Kamernet has no public API, we use web scraping. For our choice of data extraction method, we used both Selenium and BeautifulSoup. Since iterating through pages on Kamernet works in a dynamic way, and not just with an ordinary url change, we had to use Selenium. Selenium can automate browser actions like clicking the "Next Page Button" to go through all the pages. For collecting the individual room data, we chose to use BeautifulSoup since Selenium is no longer needed. Moreover, we can extract any HTML element and text or attributes within it. Both data extraction technologies aligned with our project's needs.

4.3 *If you encountered technical challenges during the scaling of your data collection, how did you resolve them? Please provide a clear explanation of the debugging process.*

One technical challenge during the scaling of the project was that some (8 out of 1932) website pages were formatted completely differently from all other pages. This caused the Python code to crash, since it could not find the page elements. We had to build a mechanism (with 'try' and 'except') that made sure the code kept running, even if an error occurred. The JSON object would be replaced with "error: missing data".

4.4 *What technical obstacles did you face during the data extraction process, and how did you overcome them?*

- We Failed to use Selenium at first. We used the driver.find_element(By. XPATH) method to locate the Next page button. However, since XPath is hard to locate from Kamernet and also it can be more brittle if the website has added or removed elements, we then decided to use driver.find_element(By.Class_NAME).find_elements(By.TAG_NAME) not only it is easier to locate but also a safer option. In the end, it allowed us to facilitate automation.

- While extracting pages, we found out that rooms from the Featured Top Ads section had been duplicated since they appeared on every page. To prevent duplication in the data set, we also extracted the room id. This room id can be found in the url of the room's page.

4.5 *What measures or monitoring systems were in place to ensure and validate the quality of the extracted data? Can you describe how these monitoring systems functioned?*

We do not have a monitoring system to validate and ensure our extracted data.

4.6 *Can you specify the infrastructure you used for the deployment and execution of your data collection?*

We use the local laptop MacBook Pro with 16 GB memory to execute the data collection.

4.7 *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created. Please provide meaningful summaries, possibly referencing timestamps from log files.*

The final data set was collected from March 20th till March 24th, 2024.

4.8 *Where was the data stored during the collection process? Detail any specific storage mechanisms or locations you utilized.*

During the data collection process, we decided to store our extracted data in JSON files, which were stored locally on our computers. The reasons we chose JSON since it provides several advantages: It not only offers a structured format for organizing and storing data, making the data comprehendible but also supports diverse data types, providing flexibility in representing complex data structures.

4.9 *Who was involved in the data collection process (e.g., students, crowd workers, contractors), and how were they compensated (e.g., how much were crowd workers paid)?*

Only four students from the MSc Marketing Analytics program at Tilburg University participated in the data collection process. These students engaged in the task without monetary compensation; instead, they received hands-on experience and in-depth training in the practice of web scraping.

4.10 *Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

There was no formal ethical review process conducted in this project. However, to examine whether Kamernet is scrapable or not, we looked into robots.txt, which showed that Kamernet did not explicitly prohibit web scraping. Furthermore, the data we opted for collecting did not acquire any sensitive individual information.

```
User-agent: *
Allow: /
User-agent: *

Disallow: /SearchRooms/GetLocations #July 2016 - prevent internal urls from being crawled
Disallow: /SearchRooms/GetRooms #July 2016 - prevent internal urls from being crawled
Disallow: /ajax/ #July 2016 - prevent internal urls from being crawled
Disallow: /SearchRooms/GetAdverts #July 2016 - prevent internal urls from being crawled
Disallow: /changelanguage/ #July 2016 - prevent internal urls from being crawled
Disallow: /*searchtenants #May 2017 - prevent crawling of unauthorized pages
Disallow: /*searchalladverts #May 2017 - prevent crawling of unauthorized pages
Disallow: /*displayroomadvert #May 2017 - prevent crawling of unauthorized pages
Disallow: /SearchRooms/CreateAlert #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/SearchRooms/CreateAlert #May 2017 - prevent crawling of unauthorized pages
Disallow: /account/alerts #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/account/alerts #May 2017 - prevent crawling of unauthorized pages
Disallow: /PropertiesApi/GetResultCount #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/PropertiesApi/GetResultCount #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/my-account #May 2017 - prevent crawling of unauthorized pages
Disallow: /mijn-account #May 2017 - prevent crawling of unauthorized pages
Disallow: /mijn-advertenties #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/my-adverts #May 2017 - prevent crawling of unauthorized pages

Disallow: /mijn-berichten #added on March 3, 2021 after HA recommendation
Disallow: /en/my-messages #added on March 3, 2021 after HA recommendation

Disallow: /account/instellingen #May 2017 - prevent crawling of unauthorized pages
Disallow: /en/account/settings #May 2017 - prevent crawling of unauthorized pages
Disallow: /EN/SearchTenants/GetTenants
Disallow: /en/searchTenants/gettenants
Disallow: /en/searchrooms/getrooms
Disallow: /en/test/
Disallow: /test/
Disallow: /studentenhuisprofiel/
Disallow: /en/studenthouseprofile/
Disallow: /backoffice.kamernet.nl #added on March 3, 2021 after HA recommendation

Disallow: /*searchtenants #May 2017 - prevent indexation of unauthorized pages; changed to disallow on March 3, 2021 after HA recommendation
Disallow: /*searchalladverts  #May 2017 - prevent indexation of unauthorized pages; changed to disallow on March 3, 2021 after HA recommendation
Disallow: /*displayroomadvert  #May 2017 - prevent indexation of unauthorized pages; changed to disallow on March 3, 2021 after HA recommendation
Sitemap: https://kamernet.nl/sitemap.xml
```

Image 3: Robots.txt of Kamernet.nl

4.11 *Does the dataset relate to people? If not, you may skip the remaining questions in this section.*

There is no user-related information extracted during the collection of the data.

5. Preprocessing, cleaning, labeling

5.1 *Did you perform any pre-processing during the data extraction process? If yes, please provide specific examples and explain the reasoning behind each on-the-fly pre-processing step.*

To get to the final dataset we made use of several preprocessing steps. In order to only extract the data that we need for analysis we make use of the 'get_text()' method provided in the BeautifulSoup package. Furthermore we made use of the '.split[]' method. The Kamernet HTML code of a given room's page, the street and city were in the same text element. Since we want to be able to treat both individually for later analysis, we split them up into two separate elements. This method was also used to extract the room id from the room's URL. The room id is needed to filter out duplicates when preparing the data. There are about six rooms (out of approximately 1350) that have a different website format, we did not find a logical reason for this. For those rooms the HTML code looks different, and the data cannot be found If the room's data cannot be found, it will be replaced with "error: missing data" right after the room id.

5.2 *After collecting the data, what additional pre-processing steps were undertaken?*

After collecting all of the data there were no additional pre-processing steps undertaken.

5.3 *Were any measures implemented to ensure privacy, such as anonymizing user data? Please describe the methods used.*

Since there is no personal data being collected but only data about rooms there were no privacy concerns or actions undertaken.

5.4 *How did you address and clean out any implausible or erroneous observations in the dataset?*

Since statistical analysis was a project requirement, we used RStudio for data preparation, aggregation, and analysis. Occasionally, the dataset contained missing values (NAs), which we filtered out in Rstudio. Additionally, the scraper included UTF-8 euro symbols and square meter signs. We removed these in RStudio, as they would have prevented the analysis from running in the program.

5.5 *Did you modify the data structure for long-term storage, like rearranging the dataset or renaming columns for clarity? If so, provide details on these changes and their rationale.*

Not applicable.

5.6 *What potential threats or biases could arise from your pre-processing steps? Please elaborate on any risks associated with the modifications made to the data and how they might impact the dataset's integrity or utility.*

Splitting the address into street and city using the ',' could potentially not account for variations in address formats of format changes. This might cause the address to be split up wrongfully.

Automatically marking data as 'missing data' without manual verification may lead to incorrect classification and loss of data.

5.7 *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

Not applicable.

5.8 *Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.*

We used RStudio for data preparation, aggregation, and analysis. The Rscript used for this process is included in the attached zip files (src\reporting\descriptives.R). To run the script, you'll need to have R installed (RStudio is optional). Please refer to TilburgScienceHub for the installation guide.

6. Uses

6.1 *Has the dataset been used for any tasks already? If so, please provide a description.*

The dataset has not been used for any major tasks yet. The dataset is solely collected for the purpose of the project of the course: online Data Collection and Management, at Tilburg University.

6.2 *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

Not applicable.

6.3 *What (other) tasks / research projects could the dataset be used for? Provide a set of potential research questions or ideas for research projects.*

The dataset could potentially answer the following questions:

- What factors most significantly influence the student room's price?
- How do availability and price of student rooms change over time?
- Are the energy label and price of a room related?
- What percentage of student rooms are actually affordable for average students?
- Is there any evidence of price discrimination based on the gender that the landlord seeks as a new resident?
- To what extent does the amount of housemates affect the price per square meter of student rooms?
- To what extent does the distance of the residence to the city center affect the accomodation's price per square meter?
- What different clusters of student rooms are there?
- How does the time of the year affect rental price and availability of student accommodations?
- How does the housing market for students differ from the normal housing market in terms of availability and price per square meter?

6.4 *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?*

Not applicable.

6.5 *Are there tasks for which the dataset should not be used? If so, please provide a description.*

Not applicable since there are no legal or ethical concerns that apply to this dataset.