

Measuring Equity and Bias in Student Housing in Koblenz: Affordability, Accessibility, and Satisfaction

Master's Thesis

in partial fulfillment of the requirements for
the degree of Master of Science (M.Sc.)
in Web and Data Science

submitted by
Kavya Hanumanthu

First supervisor: Prof. Dr. Frank Hopfgartner
Institute for Web Science and Technologies

Second supervisor: Dr. Ing. Stefania Zourlidou
Institute for Web Science and Technologies

Koblenz, September 2025

Statement

I hereby certify that this thesis has been composed by me and is based on my own work, that I did not use any further resources than specified – in particular no references unmentioned in the reference section – and that I did not submit this thesis to another examination before. The paper submission is identical to the submitted electronic version.

	Yes	No
I agree to have this thesis published in the library.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
I agree to have this thesis published on the Web.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
The thesis text is available under a Creative Commons License (CC BY-SA 4.0).	<input type="checkbox"/>	<input checked="" type="checkbox"/>
The source code is available under a GNU General Public License (GPLv3).	<input type="checkbox"/>	<input checked="" type="checkbox"/>
The collected data is available under a Creative Commons License (CC BY-SA 4.0).	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Koblenz, 23.09.2025
.....
(Place, Date)


.....
(Signature)

Note

- If you would like us to contact you for the graduation ceremony,
please provide your personal E-mail address:kavyahanumanthu.h@gmail.com
- If you would like us to send you an invite to join the WeST Alumni
and Members group on LinkedIn, please provide your LinkedIn ID :kavya-hanumanth

Zusammenfassung

Diese Studie untersucht die Wohnerfahrungen von Studierenden in Koblenz mit einem besonderen Fokus auf Bezahlbarkeit, Erreichbarkeit, Zufriedenheit und Gerechtigkeit. Es wird ein Mixed-Methods-Ansatz verwendet, der Umfragedaten, georäumliche Analysen und quantitative Modellierungen kombiniert, um zu bewerten, wie Wohnbedingungen das Wohlbefinden und das akademische Leben von Studierenden beeinflussen können.

Die Analyse prüft, ob Unterschiede in den Wohnergebnissen zwischen verschiedenen Studierendengruppen bestehen, und berücksichtigt die Möglichkeit, dass solche Unterschiede auf zugrunde liegende Benachteiligungen im Wohnsystem hindeuten. Sowohl objektive Indikatoren (wie Miete pro Quadratmeter, Entfernung zum Campus und Reisezeiten) als auch subjektive Perspektiven (Zufriedenheitsbewertungen und Sentiment-Analysen offener Antworten) werden einbezogen, um eine umfassende Bewertung zu ermöglichen.

Die Ergebnisse zeigen Unterschiede in den Wohnerfahrungen zwischen den Gruppen, mit Mustern, die auf ungleiche Bezahlbarkeit, Erreichbarkeit und Zufriedenheit hinweisen, insbesondere bei internationalen Studierenden. Diese Erkenntnisse verdeutlichen, dass Wohnerfahrungen sowohl durch strukturelle Faktoren wie Mietniveau, Zimmergröße und Lage als auch durch die individuellen Erfahrungen der Studierenden geprägt werden, was die Notwendigkeit unterstreicht, die Chancengleichheit im studentischen Wohnungswesen kritisch zu hinterfragen.

Die Studie kommt zu dem Schluss, dass gezielte Maßnahmen – wie die Ausweitung bezahlbarer Wohnheimplätze, die Verbesserung der Verkehrsanbindung und transparente Wohnungsvergabeprozesse – notwendig sind, um einen faireren und gerechteren Zugang zu studentischem Wohnraum im deutschen Hochschulkontext zu unterstützen und gleichzeitig ein Rahmenwerk zur Messung potenzieller Benachteiligungen bei Wohnergebnissen bereitzustellen.

Abstract

This study investigates student housing experiences in Koblenz, focusing on affordability, accessibility, satisfaction, and equity. A mixed-methods approach is employed, combining survey data, geospatial analysis, and quantitative modelling to assess how housing conditions may influence student well-being and academic life.

The analysis examines whether disparities in housing outcomes exist between different student groups and considers the possibility that such differences reflect underlying bias within the housing system. Both objective indicators (such as rent per square meter, distance to campus, and travel times) and subjective perspectives (satisfaction ratings and sentiment analysis of open-ended responses) are incorporated to provide a comprehensive assessment.

The results indicate variations in housing outcomes across groups, with patterns that may point to uneven affordability, accessibility, and satisfaction, particularly among international students. These insights highlight that housing outcomes are shaped both by structural factors such as rent levels, room size, and location, as well as by students' lived experiences, underscoring the need to critically assess equity in student housing.

The study concludes by emphasizing the importance of targeted measures—such as expanding affordable residences, improving transport connectivity, and ensuring transparent housing allocation processes—to support fairer and more equitable access to housing in the German higher education context, while providing a framework for measuring potential bias in housing outcomes.

Contents

1. Introduction	1
1.1. Motivation	2
1.2. Thesis Objective	4
2. Literature Review	7
2.1. Existed Studies	9
2.1.1. Affordability	9
2.1.2. Accessibility	11
2.1.3. Housing Satisfaction	12
2.1.4. Equity in Housing	13
2.2. Research Gap	14
2.2.1. Affordability	14
2.2.2. Accessibility	15
2.2.3. Housing Satisfaction	16
2.2.4. Equity in Housing	17
2.3. Thesis Research Questions	18
3. Theoretical Background	19
3.1. Machine Learning Algorithms	19
3.2. Evaluation Metrics	28
3.3. Statistical Tests	31
3.4. QGIS: Quantum Geographic Information System	33
4. Methodology	35
4.1. Data Collection	35
4.2. Data Preprocessing	37
4.2.1. Data Cleaning	37
4.2.2. Quantitative Data Processing	37
4.2.3. Geocoding of Residences	38
4.2.4. Qualitative Data Processing	38
4.3. RQ1: Affordability of Student Housing	38
4.3.1. Data and Variables	39
4.3.2. Analytical Strategy	39
4.3.3. Regression Models	39
4.4. RQ2: Accessibility of Student Housing	40
4.4.1. Data and Variables	41
4.4.2. Accessibility Measure	41

4.4.3.	Analytical Strategy	41
4.4.4.	Perceived and Behavioural Accessibility	42
4.5.	RQ3: Housing Satisfaction	43
4.5.1.	Sentiment Analysis	43
4.5.2.	Statistical Analysis of Likert Data	44
4.5.3.	Supervised Machine Learning	44
4.5.4.	Unsupervised Machine Learning	45
4.6.	Equity in Housing	46
4.6.1.	Data and Variables	46
4.6.2.	Statistical and Spatial Analysis	46
5.	Results and Discussion	48
5.1.	RQ1: Affordability of Student Housing	51
5.1.1.	Regression Models	52
5.2.	RQ2: Accessibility of Student Housing	55
5.2.1.	Regression Models of Accessibility	55
5.2.2.	DBSCAN Clustering of Student Housing	58
5.2.3.	Perceived Dimensions of Accessibility	60
5.3.	RQ3: Student Satisfaction	62
5.3.1.	Sentiment Analysis of Open-Ended Responses	62
5.3.2.	Recurring Concerns in Open-Ended Responses	64
5.3.3.	Structured Satisfaction Indicators (Likert-Scale)	66
5.3.4.	Predictive Modelling of Dissatisfaction	68
5.3.5.	Unsupervised Clustering of Housing Satisfaction	69
5.3.6.	Composite Satisfaction Index	71
5.4.	RQ4: Equity in Student Housing	72
5.4.1.	Equity Index	73
5.4.2.	Regression Analysis: Geographically Weighted Regression (GWR)	74
6.	Conclusions	78
6.1.	Limitations	79
6.2.	Future Work	80
	Appendices	88
A.	Enrollment Statistics by Semester	88
B.	Survey Questionnaire	89
C.	Qualitative Interview	99
D.	Summary Tables of Survey Results	100

1. Introduction

The availability and quality of student housing are fundamental determinants of student academic success, well-being, and general integration into university life. In university cities such as Koblenz, where higher education institutions attract a diverse and growing student population, the demand for suitable housing is both a persistent challenge and a critical opportunity to promote equity and inclusion. Research consistently demonstrates that students who reside in university-provided or well-supported housing options tend to achieve higher academic results, experience higher retention rates, and report higher satisfaction with their university experience compared to those living off campus or in less favourable conditions [9, 18]. On-campus students, for example, benefit from reduced commute times, increased access to campus resources, and stronger social networks, all of which contribute to improved academic achievement and personal development [51].

Despite these clear benefits, the landscape of student housing is marked by significant disparities. Factors such as affordability, accessibility, and quality of available accommodation vary widely not only between geographic regions but also between different demographic groups. These disparities can lead to inequities that hinder both academic and social success, particularly for students from lower socioeconomic backgrounds, international students, and those with specific accessibility needs [75, 64]. Financial constraints, limited availability of suitable housing, and challenges related to inclusivity and accessibility disproportionately affect these populations, often resulting in heightened stress, reduced sense of belonging, and compromised academic outcomes. In addition, the quality of housing, including aspects such as safety, security, and the presence of communal spaces, directly affects students' mental health and general satisfaction with their university life [20].

These inequities also raise the possibility that student housing outcomes may be influenced by underlying bias, whether in affordability, accessibility, or satisfaction. This thesis therefore approaches the analysis not only to document disparities but also to uncover whether systematic biases shape the experiences of different student groups.

Ensuring fairness in student housing is thus a multifaceted challenge that goes beyond simply providing adequate shelter. It involves promoting equity, inclusion, and well-being in a diverse student body. Fairness in this context encompasses three interrelated dimensions:

- **Affordability:** which determines whether students from various economic backgrounds can access suitable accommodation;
- **Accessibility:** which relates the physical distance to campus, the availability

of transport links, and the ease with which students can reach academic facilities and campus resources; and

- **Satisfaction:** which reflects students' perceptions of their living conditions and the extent to which their needs are met.

These factors are deeply intertwined and collectively shape students' academic performance, mental health, and overall university experience. The interplay between geographic and demographic dimensions further complicates the picture, as students' housing experiences are influenced not only by their personal circumstances but also by the broader context of the neighbourhoods and communities in which they live.

A comprehensive assessment of fairness in student housing must therefore analyze how affordability, accessibility, and satisfaction vary across geographic locations within Koblenz and among different demographic groups. By incorporating this perspective, the study frames fairness as more than differences in housing outcomes, focusing on the possibility that bias may be embedded in the housing system and seeking to uncover its presence and extent. Such an analysis is essential for identifying existing inequities, informing policy decisions, and guiding the development of more equitable housing practices.

In summary, this thesis will explore the critical role of student housing in shaping academic and social outcomes, with a particular focus on fairness across affordability, accessibility, and satisfaction. Through a detailed analysis of student housing in Koblenz, the study aims to contribute meaningful insights and recommendations for creating a more inclusive and supportive environment for the city's diverse student population.

1.1. Motivation

This study is motivated by the need to understand how these factors affect students' housing experiences and to identify potential sources of bias with respect to affordability, accessibility, satisfaction and inequity. It aims to answer the pressing questions of how different groups of students—defined by their gender, nationality, geographic location, etc., are disproportionately affected by housing challenges. The insights derived from this study could serve as a basis for creating more equitable housing allocation practices, not only benefiting students, but also informing housing solutions for other vulnerable populations. Addressing these challenges is not only a matter of fairness, but also an investment in fostering academic and societal equity.

In Koblenz, a city that hosts both the University of Koblenz and the Koblenz University of Applied Sciences, the demand for affordable, accessible, and satisfactory student accommodation is increasing, driven by a diverse and growing student population. The city offers a variety of housing options, studio apartments and shared

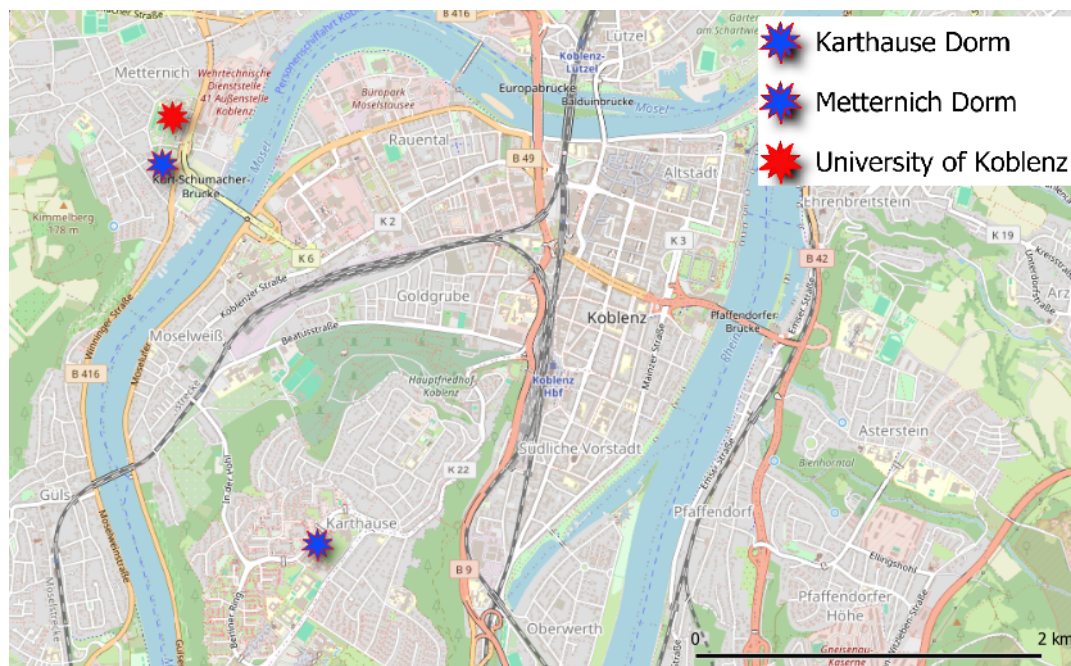


Figure 1.1.: Location of the University of Koblenz and nearby student dormitories on OSM (Open street Map) basemap.

apartments managed by the Studierendenwerk Koblenz, which are designed to provide affordable all-inclusive rents, proximity to campus, and a vibrant community life [67].

Despite these provisions, challenges persist. Students must navigate limited availability, eligibility restrictions based on age and enrollment, and the need to secure accommodation for entire semesters regardless of their arrival or departure dates [4]. Accessibility is also influenced by the location of student residences seen in Figure 1.1, with some halls within walking distance of the campus and essential amenities, while others require longer commutes [68]. Student satisfaction is shaped not only by the physical and financial aspects of housing but also by the quality of community engagement and support services, such as elected student hall councils and on-site caretakers, who facilitate integration and address resident needs [67].

A further challenge lies in the application process for student housing. Applications for shared apartments (Wohngemeinschaften, or WGs) and other residence types managed by the Studierendenwerk are typically submitted via an online portal or app like ImmoScout, WG-gesucht. While this digital approach streamlines the process for many, it can inadvertently create barriers for international students and those with limited German language proficiency. The application interfaces and instructions are often available primarily in German, making it difficult for non-German speakers to navigate the system, complete the necessary documentation, or understand eligibility criteria. As a result, some students experience rejections or

delays not due to lack of qualification, but due to language barriers and misunderstandings during the application process [67].

In addition, the limited availability of student residences means that demand frequently outstrips supply, particularly at the beginning of each semester. Many students, especially newcomers and international students, face long waiting lists and uncertainty about their housing status. This scarcity can force students to seek accommodation in the private market, where rents are often higher and access to support services is reduced.

Ensuring equitable housing quality is essential to promote diversity within university communities because inclusive and accessible living environments foster a sense of belonging, reduce feelings of isolation, and support the well-being of students from all backgrounds [3]. When universities provide equitable housing, addressing affordability, accessibility, and cultural sensitivity—they enable students of different socioeconomic statuses, identities, and abilities to participate fully in campus life and access the resources necessary for academic and personal success [3].

Equitable housing also encourages intercultural exchange, personal growth, and the development of strong social networks, all of which are critical to creating vibrant, innovative, and resilient university communities [8]. By prioritizing equity in student housing, universities not only fulfil a moral imperative but also invest in the success and happiness of their entire student body [3].

1.2. Thesis Objective

Understanding fairness in student housing requires a multifaceted approach that goes beyond simply providing accommodation. This study sets out to systematically examine the key dimensions that shape students' housing experiences in Koblenz, a medium-sized German city located at the confluence of the Rhine and Moselle rivers. As one of the oldest cities in Germany, Koblenz has a historically evolved urban structure that differs significantly from more recently developed university towns, with narrower streets, older housing stock, and infrastructural layouts not originally designed for today's levels of student mobility and demand. The city's distinctive geography, marked by its rivers and bridges, creates natural barriers that influence connectivity and commuting patterns. Unlike many other German cities built on flatter terrain, Koblenz lies in a valley with several residential areas situated on surrounding hills, further shaping travel routes and the accessibility of housing. At the same time, its limited student housing market presents additional challenges compared to larger metropolitan areas. By focusing on affordability, accessibility, and satisfaction, the research aims to uncover whether underlying biases exist across demographic and geographic groups, and how such biases contribute to inequities in housing outcomes. The following objectives guide the study's efforts to deliver a comprehensive and data-driven analysis of student housing in Koblenz.

1. To Analyze Housing Affordability:

- Examine the extent to which rent burden and cost per square meter vary among students, and assess whether affordability pressures disproportionately affect certain groups.
- Compare housing costs across accommodation types (shared flats, studios, university residences) and evaluate their alignment with quality and size, with particular attention to disparities across geographic zones (city center vs. suburbs) and demographic groups (domestic vs. international students).

2. To Assess Accessibility:

- Investigate geographic accessibility of student housing to the university campus and essential amenities, integrating both spatial measures and students' reported commuting behaviours. Explore whether perceptions of accessibility differ across groups, and whether such differences suggest possible barriers or disadvantages.
- Compare accessibility perceptions and preferences between domestic and international students, for example examining whether international students—who may attend fewer days per week—place greater value on proximity to campus and essential services.

3. To Evaluate Housing Satisfaction:

- Perform sentiment analysis on open-ended survey responses to quantify subjective housing experiences and generate sentiment scores that reflect perceptions of living conditions. Analyze both structured (Likert-scale) and unstructured (open-ended) survey responses to understand how students perceive their housing conditions. Assess whether systematic differences in satisfaction exist between groups, indicating potential inequities in housing quality and lived experiences.
- Evaluate housing satisfaction through Likert-scale survey items, for example comparing ratings of affordability, accessibility, safety, and social integration between domestic and international students to assess significant differences.
- Integrate findings from sentiment and Likert responses to provide a comprehensive comparison of housing satisfaction between the two groups and highlight disparities.

4. To Evaluate Equity in Housing:

- Integrate affordability, accessibility, and satisfaction into a composite equity framework to assess fairness in student housing outcomes. Apply statistical and spatial methods to test whether inequities are systematically distributed across student groups and neighbourhoods.

- Investigate disparities in perceptions and preferences, such as desired housing distance, participation in extracurriculars, and perceived impact on academic performance, with attention to differences between domestic and international students.

By pursuing these objectives, the study contributes to advancing equity in higher education by examining whether housing disparities in Koblenz reflect underlying biases and by providing evidence-based recommendations to foster fairer, more inclusive housing systems.

2. Literature Review

Germany has established itself in recent years as a major destination for international students, attracting a steadily growing and increasingly diverse student population. Since 2013/14, the number of international students has increased by more than 55 percent, including a 2.4 percent increase in 2024 [37]. This surge reflects several factors: tuition-free higher education, the expansion of English-language programs, and Germany's strong reputation for academic quality and post-graduation opportunities [45]. Internationalization is reshaping the higher education landscape, creating new opportunities for cultural exchange and collaboration while also placing additional demands on institutional infrastructure.

The decision to choose a particular university is shaped by both academic and non-academic factors. While program quality and institutional reputation remain important, practical considerations such as cost of living, safety, transportation, experiences of discrimination, lifestyle, immigration prospects, and the presence of social networks often play an equally decisive role [72]. Relatedly, access to part-time employment during studies and career opportunities afterwards constitutes a crucial dimension of university choice.

At the University of Koblenz, international enrolments reflect these national trends. The university offers a broad portfolio of programs, including international master's degrees taught in English across disciplines such as computer science, mathematical modelling and degree programs in natural sciences, and cultural studies. Between the winter semesters of 2019/20 and 2024/25, the number of international students nearly doubled, from 514 to 959 [70].

The graph in Figure 2.1 focuses on international students (*Bildungsausländer*), both in the overall university and within the ZFUW (Center for Distance Learning and Continuing Education) Master's programs. International enrolments have shown consistent growth, rising from 877 to 1,067 in total, and 786 to 959 in the Master's category. The observed growth in international enrollment aligns with global trends in the internationalization of higher education at the University of Koblenz, where flexible, digital, and modular Master's programs have gained increasing appeal—especially in the post-pandemic era.

The rising influx of students from countries such as India, China, Iran, and Turkey has diversified the student population and enriched intercultural exchange [32]. However, this growing diversity also presents new challenges to university infrastructure and student services, particularly in housing provision [21]. International students often face additional barriers, including limited access to subsidized residences, language and cultural difficulties in the private market, and financial constraints. These challenges are consistent with nationwide trends in Germany, where

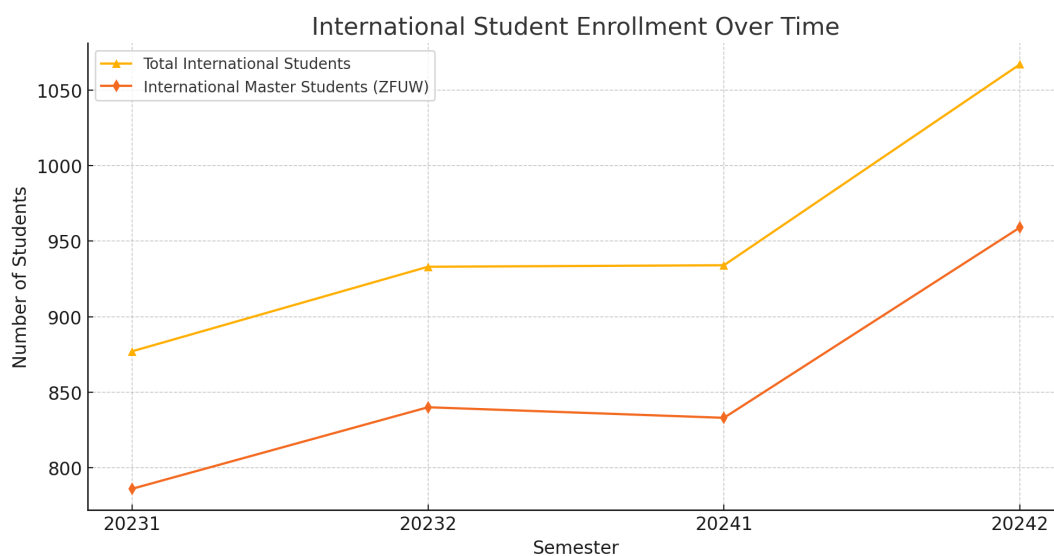


Figure 2.1.: International Student Enrollment Over Time (SoSe 2023 – WiSe 2024/25) Note: 20231 = Summer Semester 2023, 20232 = Winter Semester 2023/24. The last digit indicates the semester within the year (1 = summer, 2 = winter).

rising enrolments have strained already limited housing resources [54].

Affordable, accessible and inclusive accommodation is closely related to student academic performance and psychological health, particularly for international students who may face additional barriers to secure suitable housing. Research highlights persistent challenges in affordability, with many students experiencing high rent burdens, and accessibility, where language barriers, limited availability, and geographic disparities can hinder the ability to secure appropriate accommodation. These issues are compounded by the rapid growth in the number of international students, placing additional pressure on already limited housing resources [37]. Such patterns highlight the possibility that housing outcomes may not be evenly distributed and could reflect underlying bias that disadvantages some groups of students.

In parallel, greater attention is being paid to inclusivity in student housing. Universities are increasingly called upon to provide gender-inclusive, disability-accessible, and culturally sensitive accommodation options [3]. Yet, limited resources and physical space constraints continue to hinder implementation, leaving a gap between policy intent and practice [3].

The local housing market in Koblenz illustrates these broader issues. Available options include shared flats, studio apartments, and student residences, with wide disparities in affordability. Shared flats typically cost €400–450 per month, while studios range from €750 to €1,700, often exceeding a typical student budget [55]. Student residences managed by the Studierendenwerk Koblenz provide a more affordable alternative, with rents ranging from €280–335 per month depending on

room type (shared flats or single apartments) [67]. However, capacity is severely limited, with only 191 spots available. Each semester, fewer than 10 % of international students secure these places, with allocations averaging just 5 international and 3 Erasmus students [67]. While Catholic and Protestant dormitories also exist, they remain largely inaccessible to international students due to eligibility restrictions, including language proficiency and religious affiliation. Most students are therefore forced into the private rental market, where higher rents, deposits, utility bills, and lump-sum semester payments contribute to significant financial strain [4].

In sum, the literature indicates a growing misalignment between the rapid internationalization of higher education and the limited housing capacity available to students. In Koblenz, as in many German university towns and cities, this manifests in affordability pressures, accessibility barriers, and equity challenges that disproportionately affect international students. These dynamics highlight the increasing importance of housing as a central issue in ensuring academic success, social integration, and the long-term sustainability of internationalization in higher education.

2.1. Existed Studies

The rising demand for tertiary education and the increasing mobility of students across borders, the challenge of securing affordable, accessible, and satisfactory housing has gained global attention. Prior studies emphasize that student accommodation is not merely a matter of shelter but a factor that shapes academic performance, social integration, and overall well-being [4]. Yet, within this body of research, questions of fairness are often implicit rather than explicit, leaving scope to investigate whether observed disparities reflect structural or demographic bias in housing outcomes.

2.1.1. Affordability

Affordability is one of the most persistent challenges in student housing research. It is generally understood as the relationship between housing costs and household resources, often measured using the rent-to-income ratio or cost per square meter of living space. The “30 percent rule” introduced in [35] remains widely used for defining housing as unaffordable when costs exceed 30 percent of income. Some scholars extend this threshold to 40 or 50 percent in urban contexts where housing demand is particularly acute [66].

However, in the context of student housing, the use of income data is problematic, as students may lack consistent earnings or rely on family support. Alternative metrics, such as rent-per-unit area or rent relative to monthly expenditure, have been proposed to bypass the need for direct income data while still providing insights into affordability disparities [42].

In the context of higher education, affordability is not merely an economic concern but a key determinant of student persistence, academic performance, and well-

being [5]. The studies indicate that first-generation college students and international students are particularly at risk of housing insecurity, frequently encounter challenges such as overcrowded accommodation, frequent relocations, or in severe cases, homelessness. Such difficulties can significantly disrupt their academic experience and limit their ability to engage fully in university life [5]. International students are especially vulnerable, often excluded from subsidized accommodation and faced with limited rental histories, language barriers, and higher reliance on the private market [12].

European studies confirm the scale of affordability pressures. The European Commission (2022) reported that nearly two-thirds of students in Europe experience financial stress due to housing costs, with over 20 percent constantly worried about their ability to meet expenses. In Germany, the Deutsches Studentenwerk Sozialerhebung highlights that rent is the single largest expense for students, consuming an average of 35–40 percent of monthly budgets, with international students frequently paying even higher shares [16]. In Koblenz, where the estimated monthly living cost (including rent, food, and leisure) is around 850–950 Euros, rent alone can account for more than half of a student's monthly budget [55].

The difference between accommodation types is particularly stark. A place in a Studierendenwerk dormitory costs only around €280–335 per month, which corresponds to roughly 30 percent of a student's average budget. By contrast, a room in a shared flat (WG) typically ranges from €400–450—already 45–50 percent of monthly expenses—while private studios often exceed €750, pushing rent to nearly 70 percent of the budget. Students who cannot secure a dormitory spot therefore face significantly higher financial burdens, often forcing them to make difficult choices, such as working longer hours, reducing their course load, or even reconsidering their educational plans altogether. These disparities raise the possibility of bias in affordability: if certain groups, such as international students, systematically encounter higher rents or smaller rooms, this reflects not just market pressures but unequal outcomes embedded in housing systems [43].

Financial strain from unaffordable housing can impact mental health, increase stress, and limit students' ability to participate in campus life and social activities [55]. Students facing high housing costs may experience frequent moves, overcrowded living conditions, or delayed payments, all of which can disrupt their studies and hinder their academic success [65]. The literature further suggests that long commute times—often a result of seeking cheaper accommodation farther from campus—are linked to lower well-being and academic engagement [55].

Current studies also indicate that these burdens are not evenly distributed. International students are disproportionately affected, often paying higher rents for smaller rooms, facing limited access to subsidized housing, and experiencing greater housing insecurity compared to domestic peers [54]. In Koblenz, Studierendenwerk dormitories admit international students only for a single semester, whereas domestic students may remain for longer durations, which further limits stability for those coming from abroad. This situation illustrates how affordability challenges intersect

with institutional housing policies, shaping equity and inclusion in student housing.

Overall, the literature demonstrates that affordability is not merely an economic metric but also a social and academic one, with consequences for equality of opportunity, financial well-being, academic performance, mental health, and social participation. Rent-to-income ratios, cost per unit area, and residual income measures consistently show that students are highly vulnerable to rent burdens. Yet, gaps remain in understanding how these affordability pressures vary across student groups within small to mid sized German university cities like Koblenz, and how they interact with accessibility and satisfaction to shape the broader student experience.

2.1.2. Accessibility

Accessibility has long been recognized as a key dimension of housing quality, shaping both residential choice and well-being. In [30] the authors defined accessibility as the ease of reaching desired opportunities, a concept later expanded in [27], who highlighted land-use, transport, temporal, and individual components. In student housing research, accessibility refers not only to the proximity of accommodation to campus but also to essential amenities such as supermarkets, healthcare, and public transport, all of which influence academic and social life.

Geographic accessibility, defined as the ease with which individuals can reach essential services and opportunities, is also considered a fundamental aspect of spatial justice [53]. Within this framework, demographic parity examines whether different student groups have equal access to favourable outcomes, such as proximity to campus or shorter commute times [15]. Prolonged commutes have been linked to negative consequences including stress, fatigue, reduced academic performance, and weaker participation in campus activities [74]. Campus location further shapes accessibility: students on suburban campuses, typically more distant from city centres and with limited transport options, engage in fewer but longer trips to self-improvement destinations, whereas students at urban campuses benefit from shorter, more frequent trips facilitated by higher accessibility and transport service density [29].

Spatial analysis tools, particularly Geographic Information Systems (GIS), have been widely applied to assess patterns of accessibility and identify disparities in housing allocation [6]. Such tools allow researchers to propose targeted interventions, including transport improvements or the development of affordable housing nearer to campuses. While accessibility is often measured quantitatively through network distances or travel times, qualitative approaches are increasingly valued for capturing students' lived experiences of location, safety, and convenience, thereby enriching our understanding of how accessibility impacts well-being [62].

International students are often especially sensitive to accessibility challenges. Barriers such as limited information, language difficulties, and exclusion from institutional housing constrain their choices, sometimes forcing them into poorly con-

nected or less desirable neighbourhoods [12] [77]. This suggests that accessibility outcomes may not be evenly distributed and could point toward bias in how housing opportunities are realized across student groups.

In smaller German university cities or towns like Koblenz, the absence of U-Bahn or S-Bahn systems and reliance on buses can amplify such disparities, particularly for students in peripheral or elevated neighbourhoods where service frequency is low. While Koblenz has relatively strong bicycle infrastructure — including sign-posted cycling routes along the Rhine and Moselle, a network of dedicated bike lanes of approximately 90 km, and bicycle parking [71] — cycling is less feasible for longer distances, steep terrain, or in poor weather. Safety concerns further shape accessibility experiences, with female and first-year students reporting stronger preferences for centrally located housing due to perceptions of security [23].

The evidence suggests that accessibility is not simply about distance, but a multi-dimensional construct encompassing transport, affordability trade-offs, behavioural patterns, and perceptions. Despite extensive study in metropolitan contexts, smaller cities such as Koblenz remain under-explored, where geographic features and infrastructural limitations may amplify accessibility disparities. These disparities are further shaped by differences between domestic and international students: while many German students are able to commute using privately owned cars or bicycles, international students often rely almost exclusively on public transport. The high costs and administrative hurdles of obtaining a German driver's licence further reinforce this dependence, making accessibility particularly challenging for international students.

2.1.3. Housing Satisfaction

Housing satisfaction, while often overlooked in housing equity research, is an essential component of understanding fairness in housing systems, particularly in the context of student populations. It refers to the subjective perception of individuals regarding the adequacy of their living conditions, encompassing factors such as housing quality, safety, access to amenities, and overall comfort [4]. While affordability and accessibility are critical measurable dimensions, satisfaction provides an equally important, though subjective, lens for evaluating housing equity.

Numerous studies have identified the factors that shape housing satisfaction. These include physical attributes of housing (size, quality of construction, and maintenance), neighbourhood characteristics (safety, cleanliness, and access to public services), and personal circumstances such as demographics and household composition [69]. In the student housing context, additional dimensions such as proximity to campus, availability of social spaces, and internet connectivity play a pivotal role [46]. Research consistently shows that inadequate housing, such as overcrowded or poorly maintained accommodation, contributes to dissatisfaction, stress, and reduced well-being, particularly among international students and those with limited housing alternatives [77]. Conversely, housing that meets or exceeds expectations

enhances satisfaction and positively influences academic performance and social integration [36].

Traditional methods for measuring housing satisfaction, such as structured surveys, provide valuable quantitative insights but often fail to capture the nuances of subjective experience. To address this, researchers have increasingly applied sentiment analysis and other natural language processing techniques to open-ended survey responses or online housing reviews. These methods enable the detection of patterns in emotional expression and perceptions, offering richer insights into the lived experiences of students [1]. These strands of research highlight that satisfaction outcomes may not be distributed evenly across groups, pointing to the possibility that bias is present in how housing adequacy is experienced by students.

In the student housing context, the research highlights that housing satisfaction acts as a bridge between material housing conditions and student well-being. Incorporating satisfaction into broader evaluations of affordability and accessibility is crucial for developing equitable housing policies and improving quality of life in university contexts.

2.1.4. Equity in Housing

Equity in housing is typically framed as a multidimensional concept, encompassing economic, spatial, and experiential aspects of fairness in access to adequate housing [53]. Within urban and housing studies, equity extends beyond the mere provision of housing units to consider whether individuals have comparable access to adequate, affordable, and satisfactory housing conditions. This perspective highlights the interplay between structural factors, such as rent levels and infrastructure, and subjective experiences, such as perceptions of comfort, security, and inclusion [10]. For the purposes of this study, three contributors—affordability, accessibility, and satisfaction—are examined as the key dimensions shaping equity in student housing.

In the student housing context, equity is rarely treated as a stand-alone dimension but rather emerges from the intersection of the above mentioned dimensions. Prior research shows that inequities frequently align with demographic factors such as income, nationality, and gender, as well as geographic location. For instance, domestic and international students may face very different housing outcomes due to language barriers, eligibility for subsidies, or unequal access to institutional housing [12, 77]. Similarly, students living in peripheral neighbourhoods often experience a compound disadvantage: higher commuting costs and times, lower satisfaction with housing quality, and limited opportunities for social integration [29]. These conditions can also affect class attendance and participation, with longer commutes and unstable housing arrangements linked to reduced academic engagement and performance. However, few studies have explicitly framed these inequities as the potential outcome of bias in housing systems. Recognizing this possibility highlights the importance of moving beyond descriptive disparities to examine whether

unfair processes or hidden disadvantages shape student housing equity.

To study such disparities, researchers increasingly rely on spatial and statistical methods, including clustering and Geographically Weighted Regression [59] to identify localized or group-specific disparities. These approaches reveal that inequities are rarely uniform but vary across neighbourhoods and subpopulations. Nevertheless, relatively few studies systematically integrate these three dimensions into a single equity framework, especially in German university cities where rising international enrolments intersect with limited housing capacity.

The literature establishes that equity provides a unifying perspective on housing fairness. By combining objective measures of affordability and accessibility with subjective measures of satisfaction, equity assessments offer a comprehensive understanding of how housing systems reproduce or mitigate inequalities. However, there is still opportunity for more research, especially in university cities in Europe, as there is a noticeable gap in the integrated frameworks that are yet underdeveloped in the context of student housing.

These studies underscore that housing is more than a material resource; it is a foundation of educational opportunity and social participation. While the three dimensions have often been examined individually, their interconnections highlight the need for a more integrated perspective in evaluating fairness in student housing. This multidimensional framing forms the basis for the present study and leads to the identification of research gaps explored in the following section.

2.2. Research Gap

Despite extensive attention to student housing in the literature, several critical questions remain unanswered. Research has established that affordability, accessibility, and satisfaction are central to the student experience, but little is known about how these dimensions interact to produce equitable or inequitable outcomes. Existing work tends to focus on broad trends in national or metropolitan contexts, leaving smaller to mid-size university cities and towns, where resources are more constrained and student populations increasingly diverse, largely overlooked. Moreover, few studies directly compare the experiences of domestic and international students, even though their housing challenges and expectations often differ substantially. These gaps highlight the need to investigate not only disparities but also the possibility that bias influences student housing outcomes.

2.2.1. Affordability

In addition, much of the existing literature on affordability tends to emphasise national averages or metropolitan contexts [13]. While these perspectives are valuable for highlighting broad trends, they often obscure localised pressures in smaller or mid-sized university cities, where housing supply is constrained, institutional accommodation is limited, and the private rental market dominates. Research on

“studentification” in European contexts [63, 38] shows that smaller towns or cities are particularly vulnerable to rent inflation when student numbers grow, as demand quickly outpaces supply. This often forces students into difficult trade-offs: accepting smaller or lower-quality housing at disproportionately high cost, working longer hours to cover rent, reducing their course load, or even reconsidering their studies [31, 34].

Moreover, relatively few studies systematically investigate how affordability burdens vary across accommodation types. University residences are generally considered more affordable and stable, yet limited in capacity, while shared flats and private studios often present higher costs but varying degrees of quality and space [46]. The lack of comparative analysis across these housing categories limits our understanding of the differentiated financial pressures students face. It also raises the possibility that such pressures are not neutral but may reflect underlying bias in how affordability outcomes are distributed among students.

Another overlooked dimension is demographic variation, particularly between domestic and international students. International students are more likely to be excluded from subsidised or institutional housing and may encounter additional barriers such as language, rental guarantor requirements, or limited knowledge of the housing market [12, 77]. This positions them at greater risk of housing insecurity and higher rent burdens [43]. However, most affordability studies aggregate students into a single category, neglecting these important demographic differences.

These gaps highlight the need for context-sensitive affordability research that goes beyond national averages to examine how costs are distributed across housing types and demographic groups in university cities and towns. Addressing this gap is essential to capture the unequal financial pressures shaping student experiences and to inform policy responses that promote equity in higher education.

2.2.2. Accessibility

Accessibility has been widely acknowledged as a central determinant of student housing outcomes, linking residential choice with academic participation, social integration, and overall well-being. Much of the existing research, however, has concentrated on metropolitan regions with well-developed transport systems [29]. These contexts provide valuable insights into how proximity to campuses and efficient transport links influence students’ quality of life, but they offer limited applicability to smaller or mid-sized university cities, where infrastructure is less extensive and students may depend on a single transport mode.

In such towns, accessibility challenges are shaped not only by distance but also by geographic and infrastructural constraints. In cities like Koblenz, bus networks provide the primary means of transport, with frequency and coverage varying considerably across neighbourhoods. Yet few studies systematically analyse how these localised constraints affect students’ commuting times, campus engagement, or access to essential services such as healthcare, supermarkets, and social spaces.

Moreover, accessibility research has generally prioritised quantitative measures of distance and travel time, while giving less attention to behavioural and perceptual dimensions. Factors such as how frequently students attend campus, how much value they place on proximity, and how safe they feel during commutes significantly shape the lived experience of accessibility [69]. These aspects remain under-explored, particularly in comparative analyses of domestic and international students, who may differ in both commuting behaviours and expectations of location. This leaves open the possibility that differences in accessibility outcomes may reflect underlying bias rather than neutral variation.

While spatial analysis tools such as GIS and network modelling are increasingly applied in urban studies to evaluate accessibility [6], their use in student housing research, particularly in smaller European university cities, remains limited. Most studies continue to rely on broad measures of distance or travel time, overlooking the potential of advanced open-source tools such as OSMnx, which allow for fine-grained analysis of street networks, travel routes, and connectivity patterns. Incorporating such methods offers the possibility of identifying localized inequities that traditional measures may obscure.

Existing research does not sufficiently capture how accessibility is experienced in smaller university cities, nor how behavioural and perceptual dimensions (e.g., commuting preferences, attendance patterns) intersect with geographic constraints. These gaps are especially relevant in contexts where international and domestic students may value or experience accessibility differently. Addressing this requires both methodological innovation and context-specific analysis. By applying network-based approaches such as OSMnx alongside survey data, this study contributes to filling this gap and advances understanding of how accessibility shapes equitable student housing outcomes.

2.2.3. Housing Satisfaction

Housing satisfaction is recognised as a key indicator of residential quality, capturing subjective perceptions of adequacy, comfort, and well-being. Studies in the student housing context have examined the role of physical attributes (size, quality, maintenance), neighbourhood characteristics (safety, cleanliness, access to services), and institutional factors (proximity to campus, availability of social spaces, internet connectivity) in shaping satisfaction levels [4, 46]. However, much of this research has been conducted in non-European or metropolitan settings.

A further limitation of the existing literature lies in its overreliance on structured survey methods such as Likert-scale responses. While these provide useful insights, they are restricted in their ability to capture the full range of students' lived experiences. open-ended survey data remain underexplored. Although NLP approaches such as sentiment analysis and topic modelling have been successfully applied in transport and urban studies [24], and in the analysis of open-ended survey responses [58], they are rarely used in student housing research. This creates

a clear methodological gap that the present study addresses by applying sentiment analysis to student survey narratives.

Satisfaction research has tended to treat student populations as relatively homogeneous, with limited comparative analysis of domestic and international students. This is a significant omission given that international students frequently face unique challenges such as language barriers, online applications for rental space, limited access to institutional housing, and greater dependence on the private rental market, which may directly affect their satisfaction [12]. Without systematically examining these differences, existing studies risk overlooking important disparities in housing experiences that influence academic success and social integration. Such disparities also point to the possibility that housing satisfaction outcomes may be shaped by bias, a dimension that current research has yet to explicitly address.

In summary, existing research on housing satisfaction provides valuable insights into the factors shaping student perceptions, yet it remains limited in scope, methodology, and inclusivity. What is missing is a nuanced, comparative, and methodologically innovative understanding of how different groups of students experience satisfaction in diverse housing contexts. Addressing this gap is essential for building a more comprehensive picture of student housing and for informing policies that promote equitable and supportive living environments.

2.2.4. Equity in Housing

Equity in housing is commonly discussed in relation to distributive justice, where fairness is evaluated not only by the availability of housing but also by how opportunities and burdens are distributed across different groups [53, 10]. Within the student housing literature, however, equity is seldom treated as an explicit object of study. Instead, researchers typically examine dimensions such as affordability, accessibility, or satisfaction separately, making it difficult to draw conclusions about fairness in housing outcomes overall. This siloed approach overlooks the fact that equity is inherently multidimensional and emerges through the interplay of economic, spatial, and subjective experiences.

A further limitation lies in how equity is operationalised. Existing studies often reduce questions of fairness to single indicators, such as rent-to-income ratios or commuting distances. While valuable, these measures cannot capture the full spectrum of disparities that students face. For instance, a student may have relatively affordable rent but live in poorly maintained housing far from campus, resulting in inequitable outcomes when viewed more holistically.

Few studies attempt to combine objective housing indicators with subjective measures of satisfaction in a unified framework. As a result, important disadvantages particularly those experienced by diverse student groups which remain hidden. This lack of an integrative framework represents a clear integration gap in the literature, and addressing it is essential for understanding fairness in student housing. Equally, the absence of studies that explicitly consider the possibility of bias means

that current research may underestimate the extent of unfairness in housing outcomes.

By synthesizing these areas, the present study situates its research questions within established theoretical and empirical debates. While previous works provide valuable insights into affordability, accessibility, satisfaction, and equity, few have systematically integrated these perspectives into a single framework, particularly in the context of European university cities. This study seeks to bridge that gap by drawing on survey data, geospatial analysis, and advanced statistical modelling to provide a comprehensive understanding of student housing experiences and their implications for equity and student well-being.

2.3. Thesis Research Questions

The present study investigates the housing experiences of university students in Koblenz, with particular attention to issues of affordability, accessibility, satisfaction, and equity. Building on insights from prior research and the identified gaps in the German student housing context, the study is guided by the following research questions:

- Do students experience affordability disparities in housing across different demographic groups and geographic regions?
- Do students experience disparities in geographic accessibility to campus and essential amenities across different demographic groups?
- How can student survey data be analysed using machine learning and sentiment analysis to identify trends in housing satisfaction and uncover recurring concerns related to fairness in student housing experiences?
- How can geographical and demographic data be combined with sentiment analysis to assess equity in student housing in Koblenz among diverse student groups?

3. Theoretical Background

Understanding student housing experiences requires not only descriptive and statistical approaches but also the application of advanced computational methods. In recent years, machine learning (ML) has emerged as a valuable tool for analyzing complex, multidimensional datasets in the social sciences, including housing and urban studies. Unlike traditional statistical models, ML algorithms can capture nonlinear patterns, perform classification and clustering tasks, and integrate diverse data types such as survey responses, geospatial information, and textual feedback. This flexibility allows researchers to identify hidden structures in the data, predict outcomes, and provide more nuanced insights into affordability, accessibility, satisfaction, and equity.

The following subsections introduce the machine learning algorithms and evaluation metrics employed in this research, explaining their theoretical foundations and relevance to the research questions.

3.1. Machine Learning Algorithms

Ordinary Least Squares (OLS) Regression

Ordinary Least Squares (OLS) regression is one of the most widely used methods in statistical modeling and machine learning, particularly suited for understanding relationships between a continuous dependent variable and one or more independent variables. The central idea is to fit a linear function that best represents the relationship between predictors and outcomes by minimizing the discrepancies between observed and predicted values. OLS is not only computationally efficient but also provides interpretable results, making it especially relevant in applied social science research such as housing studies.

Formula The basic regression equation can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i \quad [78] \quad (3.1)$$

where y_i is the dependent variable (e.g., rent per square meter),

β_0 is the intercept,

β_j are the regression coefficients showing the effect of predictors x_{ji} (e.g., student status, location, room size),

and ϵ_i is the error term.

To estimate the coefficients, OLS minimizes the sum of squared residuals, defined as the difference between the observed and predicted values. This principle can be formally expressed as:

$$\hat{\beta} = (X'X)^{-1}X'y \quad [78] \quad (3.2)$$

Equation 3.2 the closed-form OLS estimator, which provides the best linear unbiased estimates of the regression coefficients.

In the context of this study, OLS regression was applied to analyze affordability and accessibility dimensions of student housing. For instance, Equation 3.1 was specified with rent per square meter as the dependent variable and predictors such as room size, geographic location, and student status as independent variables. The coefficients estimated in this model quantify how much each characteristic contributes to differences in housing costs, making it possible to detect systematic disparities between groups. In addition, accessibility measures (e.g., travel distance to campus) were modeled using the same framework to assess whether demographic or structural variables significantly influenced commuting patterns.

A graphical representation of the OLS method is shown in Figure 3.1. The scatter-plot illustrates the relationship between rent and room size, with the regression line representing the fitted values. The vertical distances between the observed points and the line correspond to residuals, which OLS minimizes to achieve the best fit.

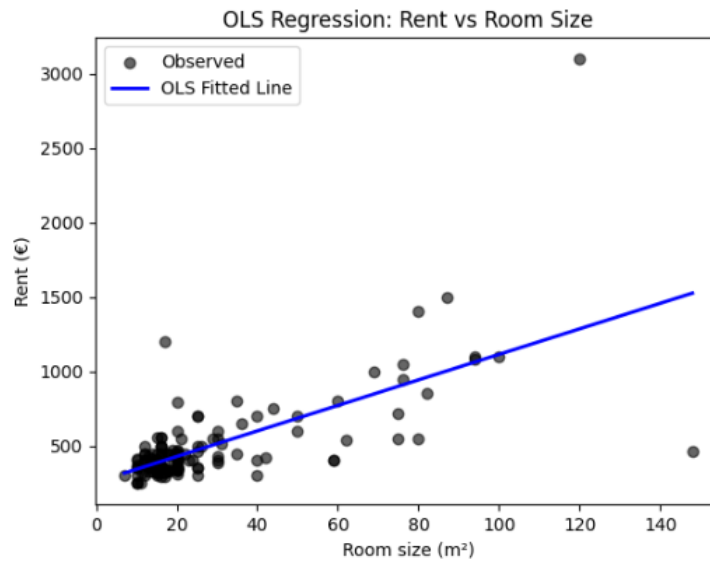


Figure 3.1.: Illustration of OLS regression applied to housing affordability: rent (y-axis) against room size (x-axis), with the fitted regression line.

While OLS is a powerful and interpretable method, its validity relies on key assumptions, including linearity, independence of errors, homoscedasticity (constant variance of errors), and normally distributed residuals. These assumptions were checked during the analysis to ensure the robustness of results. Nonetheless, OLS provides an essential foundation for this study by enabling both explanatory and

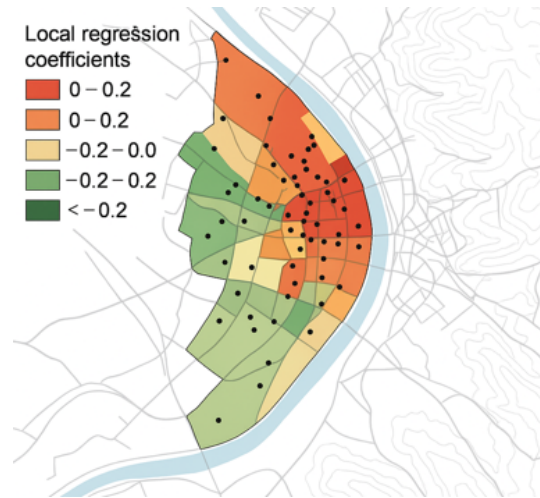


Figure 3.2.: Illustration of Geographically Weighted Regression (GWR). Local regression coefficients are mapped as a choropleth, showing how the strength and direction of relationships vary across geographic space.

inferential insights into the affordability and accessibility of student housing.

Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) extends the classical OLS framework by allowing regression coefficients to vary across geographic locations. Instead of estimating a single global set of parameters for all observations, GWR calibrates a separate local regression for each location using spatially weighted data from its neighbors. This approach captures spatial heterogeneity, meaning it reveals how the strength and direction of relationships change depending on where observations are situated.

Formula

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i), x_{ki} + \epsilon_i \quad [22] \quad (3.3)$$

Where:

- y_i is the dependent variable at location i (e.g., rent per square meter),
- (u_i, v_i) are the spatial coordinates of location i ,
- $\beta_k(u_i, v_i)$ are location-specific regression coefficients,
- x_{ki} are predictor variables (e.g., room size, student status),
- ϵ_i is the residual error.

In the GWR framework see Equation 3.3, each predictor is associated with a location-specific coefficient $\beta_k(u_i, v_i)$. The resulting values represent how the strength and direction of the relationship between predictors (e.g., room size, student status, or accommodation type) and the outcome variable vary across space. High positive local coefficients indicate neighbourhoods where the predictor strongly increases the dependent variable (e.g., higher rent per square meter), while negative coefficients suggest locations where the same predictor has a weaker or opposite effect as shown in the 3.2. By mapping these coefficients, it becomes possible to identify geographic disparities and evaluate equity in terms of affordability and accessibility

Logistic Regression

Logistic regression is a statistical method for analyzing relationships where the dependent variable is categorical, most often binary (e.g., yes/no, 0/1). Unlike ordinary least squares (OLS), which predicts continuous outcomes, logistic regression estimates the probability of an event occurring and ensures that predictions remain bounded between 0 and 1 [33]. This makes it especially useful in the social sciences where outcomes of interest are frequently dichotomous.

Formula

The model is based on the logistic, or sigmoid, function, which maps a linear combination of predictors onto probabilities:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}, \quad [33] \quad (3.4)$$

Where:

- p_i : Probability that the event occurs (e.g., rent burden = 1).
- $\frac{p_i}{1-p_i}$: Odds of the event occurring.
- \ln : Natural logarithm used to linearize the probability.
- β_0 : Intercept term, baseline log-odds when all predictors are zero.
- β_j : Coefficients showing the change in log-odds of the outcome per unit change in predictor X_j .

Logistic regression is often illustrated through the *sigmoid curve*, which shows how the probability of an event gradually increases from 0 to 1 as the predictor variable changes. Figure 3.3 demonstrates this concept with binary outcomes plotted against the fitted logistic curve [52]. In this study, the approach is applied to the affordability dimension, where the outcome of interest is rent burden, coded as 1 if a student spends more than half of their monthly budget on accommodation and 0 otherwise. The model therefore enables assessing how demographic, structural, and spatial characteristics influence the likelihood of students experiencing affordability stress.

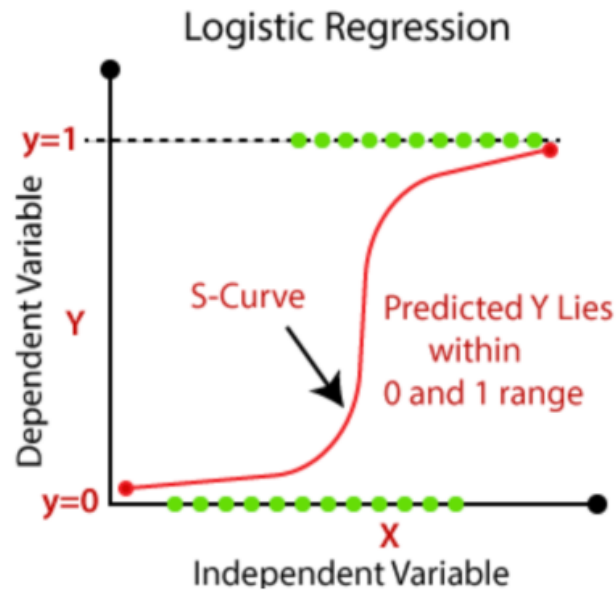


Figure 3.3.: Illustration of the logistic regression curve (adapted from [52]).

Random Forest Classifier

Random forests are an ensemble learning method introduced by Breiman (2001) that combine multiple decision trees to improve predictive accuracy and reduce overfitting [7]. Unlike a single decision tree, which may be sensitive to small variations in the data, random forests aggregate the predictions of many trees built on bootstrapped samples and random subsets of features. This approach makes them highly effective for classification tasks involving complex, non-linear relationships.

Algorithm

1. Draw B bootstrap samples from the training data.
2. For each bootstrap sample, grow an unpruned decision tree:
 - At each node, randomly select m predictors from the full set of p predictors.
 - Choose the best split among the m predictors.
 - Split the node into two child nodes.
3. Repeat the process until the forest of B trees is grown.
4. For classification, each tree casts a “vote” for a class, and the majority vote determines the final prediction.

Random forest classification is often illustrated by showing how multiple decision trees are combined to form an ensemble. Each tree provides a classification,

and the forest aggregates these votes to produce a final prediction. Figure demonstrates this concept, where different trees split the data in varied ways but their combined output yields a more robust and accurate classification [7]. In this study, the approach was applied to the dimension of satisfaction. Specifically, the random forest classifier was used to categorize student responses on Likert-scale satisfaction questions, identifying patterns across demographic, structural, and geographic factors. By leveraging the ensemble of trees, the method helps to reveal which characteristics most strongly influence students' housing satisfaction.

K-Means Clustering

K-means is one of the most widely used unsupervised machine learning algorithms for partitioning data into meaningful groups. Unlike regression or classification methods, which predict specific outcomes, K-means seeks to identify inherent structures in the data by grouping observations into k clusters such that members of the same cluster are more similar to each other than to those in other clusters [40].

Algorithm

The K-means algorithm can be summarized in the following steps:

1. Choose the number of clusters k .
2. Randomly initialize k centroids.
3. Assign each observation to the nearest centroid.
4. Update centroids by computing the mean of all assigned points.
5. Repeat steps 3–4 until convergence (no significant change in assignments).

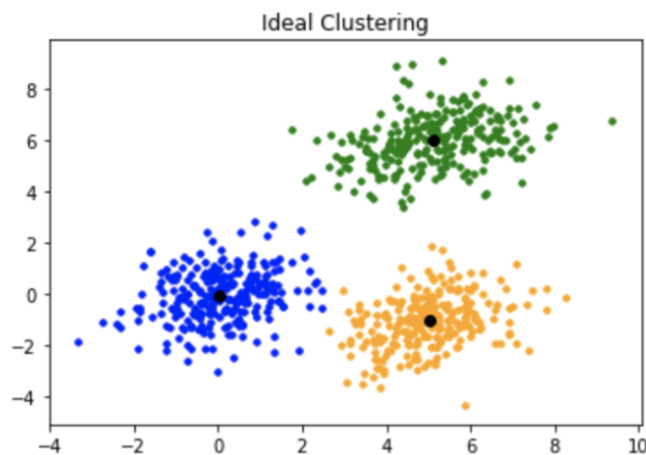


Figure 3.4.: Illustration of K-means clustering with centroids (adapted from [26]).

K-means is often illustrated by plotting data points in two dimensions and showing how the algorithm iteratively refines cluster assignments until stable groups emerge. Figure 3.4 demonstrates this process, where data points are partitioned into clusters around their centroids [26]. In this study, K-means clustering was applied to the *satisfaction dimension* to explore patterns in student survey responses. By grouping students according to their satisfaction ratings across multiple housing-related factors, the method provides insights into distinct subgroups—such as highly satisfied, moderately satisfied, or dissatisfied students—and highlights the multidimensional nature of housing experiences.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN is an unsupervised machine learning algorithm. Unlike K-means, which requires the number of clusters to be specified in advance, DBSCAN identifies clusters based on data density. It is particularly effective at discovering clusters of arbitrary shape and handling noise (outliers) [19].

Algorithm

The DBSCAN algorithm can be summarized as follows:

1. Select two parameters: ε (radius of the neighborhood) and *MinPts* (minimum number of points required to form a dense region).
2. For each point, retrieve its ε -neighborhood:
 - If the neighborhood contains at least *MinPts*, the point is labeled as a *core point*.
 - If the point is within the neighborhood of a core point, it is a *border point*.
 - Otherwise, the point is labeled as *noise*.
3. Clusters are formed by connecting core points and their neighbors, expanding iteratively until no further points can be added.

DBSCAN is often illustrated by plotting data points and showing how dense areas are grouped into clusters, while sparse areas are treated as noise. Figure 3.5 demonstrates this process, where clusters of arbitrary shapes are discovered without pre-specifying their number [25]. In this study, DBSCAN was applied to the accessibility dimension to identify clusters of students with similar commuting patterns while isolating outliers with exceptionally high travel distances or times.

nlptown BERT-based Sentiment Model

The nlptown/bert-base-multilingual-uncased-sentiment model is a transformer-based language model fine-tuned for sentiment analysis across multiple languages. It classifies input text on a 1–5 star scale, where lower scores correspond to negative sentiment and higher scores correspond to positive sentiment [47]. This makes it suitable

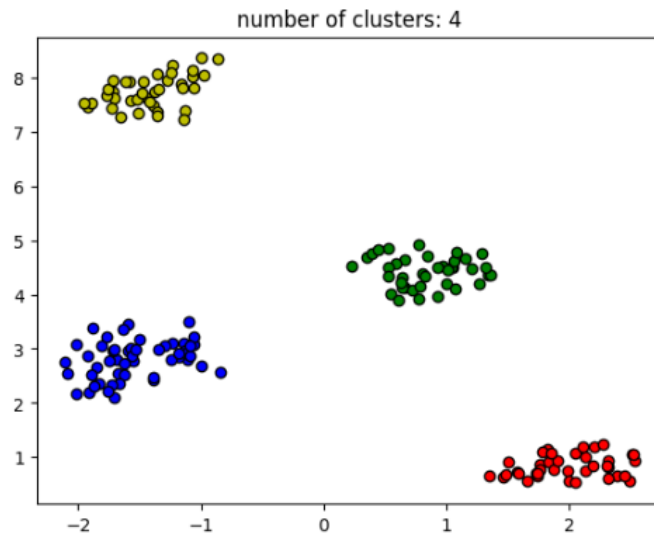


Figure 3.5.: Illustration of DBSCAN clustering with noise points (adapted from [25]).

for analyzing open-ended survey responses, reviews, or textual feedback in multi-lingual contexts.

Pseudocode

```
function SentimentClassification(text):
    load pretrained BERT sentiment model
    load tokenizer
    tokens = tokenize(text)
    outputs = model(tokens)
    prediction = argmax(outputs)
    if prediction in [1, 2]:
        return "Negative"
    else if prediction == 3:
        return "Neutral"
    else if prediction in [4, 5]:
        return "Positive"
```

To illustrate the process, open-ended responses from the student housing survey were passed through the model, and their predicted sentiment categories were assigned:

- *“I was fortunate enough to get an affordable single apartment in one of the most central neighborhoods of Koblenz. I consider myself incredibly lucky.”* → **Score 5 – Positive**

- *“Actually it is a very calm, peaceful and quiet place where I am living, but the problem is with the rent. Actually, the rent is a bit high for me.”* → **Score 3 – Neutral**
- *“Student housing is very bad. First you don’t find it or you have to compromise a lot to get an apartment.”* → **Score 1 – Negative**

These examples demonstrate how the model converts unstructured text into structured categories of positive, neutral, and negative sentiment, enabling quantitative analysis of qualitative feedback.

BERTopic

BERTopic is a topic modeling technique that leverages transformer-based embeddings (e.g., BERT) and clustering algorithms to generate coherent topics from text data [28]. Unlike traditional topic modeling methods such as Latent Dirichlet Allocation (LDA), BERTopic creates document embeddings using pre-trained language models, reduces dimensionality (commonly with UMAP), and then applies clustering (e.g., HDBSCAN) to identify dense groups of semantically similar documents. The result is a set of interpretable topics represented by their most frequent words.

Pseudocode

The general workflow of BERTopic can be summarized in pseudocode as follows:

```
function TopicModeling(text_documents) :
    # Step 1: Generate embeddings
    embeddings = TransformerModel.encode(text_documents)

    # Step 2: Reduce dimensionality
    reduced_embeddings = UMAP.reduce(embeddings)

    # Step 3: Apply clustering
    clusters = HDBSCAN.fit(reduced_embeddings)

    # Step 4: Extract topics
    topics = extract_keywords_per_cluster(clusters, text_documents)

    return topics
```

When applied to open-ended survey responses, BERTopic groups semantically similar answers into clusters that represent distinct topics. In this study, BERTopic was applied to open-ended survey responses to uncover recurring themes in students’ housing experiences. This enabled a more systematic understanding of qualitative feedback by organizing diverse statements into coherent and interpretable categories.

3.2. Evaluation Metrics

To ensure the robustness of these models, appropriate evaluation metrics and complementary statistical tests are equally important. Evaluation measures not only quantify model performance but also help assess whether identified patterns are meaningful, reliable, and generalizable. By combining ML algorithms with statistical metrics, the present study advances beyond descriptive reporting to a systematic and evidence-based analysis of student housing.

Silhouette Score

The Silhouette Score is an evaluation metric commonly used to assess the quality of clustering results. It measures how well each observation fits within its assigned cluster compared to other clusters, thereby capturing both cohesion (similarity within a cluster) and separation (difference between clusters). A score close to +1 indicates that the observation is well-matched to its cluster and poorly matched to neighboring clusters, while a score near 0 suggests overlapping clusters, and negative values indicate potential misclassification.

Formula

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \quad [60] \quad (3.5)$$

where:

- $a(i)$ = the average distance between observation i and all other points in the same cluster (cohesion),
- $b(i)$ = the minimum average distance between i and all points in other clusters (separation), and
- $s(i)$ = silhouette score of observation i , ranging from -1 to +1.

In this research, the Silhouette Score was applied to evaluate clustering results from algorithms such as K-Means and DBSCAN. As defined in Equation 3.5, the score quantifies both cohesion ($a(i)$) within a cluster and separation ($b(i)$) from other clusters, providing a standardized measure of clustering quality. Higher silhouette values indicated that clusters represented distinct profiles of student housing outcomes, while lower values suggested overlapping or less interpretable structures.

Cohen's d

Cohen's d is an effect size metric commonly used to quantify the magnitude of differences between two groups. Unlike significance tests that only indicate whether a difference exists, Cohen's d provides a standardized measure of how large that difference is, expressed in terms of standard deviations. This makes it particularly useful for interpreting the practical importance of group comparisons. Cohen's d is

defined in Equation 3.6, where the difference between the means of two groups (\bar{X}_1 and \bar{X}_2) is standardized by the pooled standard deviation s_p .

Formula

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p} \quad [11] \quad (3.6)$$

where:

- \bar{X}_1 and \bar{X}_2 are the sample means of the two groups, and
- s_p is the pooled standard deviation

The pooled standard deviation s_p , shown in Equation 3.7, combines the variability of both groups into a single denominator by weighting their variances according to sample size::

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad [11] \quad (3.7)$$

with s_1, s_2 representing the standard deviations and n_1, n_2 the sample sizes of the two groups.

In this study, Cohen's d was used to evaluate the strength of differences between domestic and international students in affordability, accessibility, and satisfaction outcomes. For instance, when comparing average rent per square meter or reported satisfaction scores, a larger effect size (Equation 3.6) indicated that the difference was not only statistically significant but also practically meaningful. Following Cohen's guidelines, values around 0.2 were considered small effects, around 0.5 medium effects, and 0.8 or higher large effects, enabling a standardized interpretation of equity disparities in student housing.

Confusion Matrix

To evaluate the performance of classification models, a confusion matrix is commonly used as a baseline representation. It summarizes the predictions of a classifier by comparing predicted labels with the actual labels. The matrix is structured into four outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as shown in Table 3.1.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 3.1.: Confusion matrix structure for classification outcomes.

This matrix forms the foundation for computing evaluation metrics such as Accuracy, Precision, Recall, and the F1 Score, which provide different perspectives on the effectiveness of classification models.

Accuracy

Accuracy is a fundamental evaluation metric for classification models, measuring the proportion of correctly predicted instances out of the total number of instances. It provides an overall indication of model performance, although it may be less informative when class distributions are highly imbalanced. Accuracy is defined in Equation 3.8, where the number of true positives (TP) and true negatives (TN) are compared against the total number of observations.

Formula

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad [44] \quad (3.8)$$

Here, TP denotes true positives, TN true negatives, FP false positives, and FN false negatives.

For instance, if the model correctly predicts whether 85 out of 100 students are rent-burdened or not, the accuracy would be 0.85. A high Accuracy score therefore indicates that the model correctly classifies the majority of cases overall, combining both positive and negative outcomes.

Precision

Precision is an evaluation metric that measures the proportion of correctly predicted positive cases out of all cases that were predicted as positive. It indicates how reliable the model is when it classifies an instance as belonging to the positive class, making it particularly important in contexts where false positives carry a high cost. Precision is defined in Equation 3.9, where the number of true positives (TP) is compared to the sum of true positives and false positives (FP):

Formula

$$\text{Precision} = \frac{TP}{TP + FP} \quad [56] \quad (3.9)$$

For example, if the model predicts 50 students as rent-burdened and 45 of them actually are, the precision would be 0.90, meaning most predicted cases are correct. A high Precision score indicates that most of the instances classified as positive by the model are indeed correct, meaning the model makes few false positive errors.

Recall

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive cases that the model correctly identifies. It captures the ability of a model to detect all relevant instances, making it particularly important when missing positive cases carries significant consequences. Recall is defined in Equation 3.10, where the number of true positives (TP) is compared to the sum of true positives and false negatives (FN):

Formula

$$\text{Recall} = \frac{TP}{TP + FN} \quad [56] \quad (3.10)$$

As an example, if 60 students are truly rent-burdened and the model correctly identifies 48 of them, the Recall would be 0.80. A high Recall score shows that the model captures most actual positive cases, meaning fewer truly rent-burdened students are overlooked.

F1 Score

The F1 Score is an evaluation metric that combines Precision and Recall into a single value by calculating their harmonic mean. It provides a balanced measure that is particularly useful when dealing with imbalanced datasets, where optimizing only Precision or only Recall may not give a complete picture of model performance. The F1 Score is defined in Equation 3.11, where Precision (P) and Recall (R) are combined:

Formula

$$F1 = 2 \times \frac{P \times R}{P + R} \quad [56] \quad (3.11)$$

For instance, if a model's Precision is 0.90 and Recall is 0.80, the F1 Score would be 0.85. A high F1 Score indicates that the model achieves both strong Precision and strong Recall, maintaining a balance between avoiding false positives and minimizing false negatives.

Together, Accuracy, Precision, Recall, and the F1 Score were used in this study to evaluate classification models such as Logistic Regression and Random Forest, ensuring that predictions of student housing outcomes were both reliable and comprehensive.

3.3. Statistical Tests

In addition to machine learning evaluation metrics, statistical tests were employed to compare group differences and examine associations in the data, ensuring that the observed patterns were statistically robust.

Welch's t -test

Welch's t -test is a statistical test used to determine whether there is a significant difference between the means of two groups. Unlike the classical Student's t -test, Welch's version does not assume equal variances or equal sample sizes, making it more robust and suitable for applied social science research.

Formally, the test statistic is given in Equation 3.12, where the difference between sample means is standardized by a denominator that accounts for unequal variances:

Formula

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad [73] \quad (3.12)$$

Here, \bar{X}_1 and \bar{X}_2 represent the sample means of the two groups, s_1^2 and s_2^2 are the corresponding variances, and n_1 and n_2 are the sample sizes.

For example, when comparing the average rent per square meter between domestic and international students, Welch's t -test evaluates whether the observed difference in means is statistically significant even if the two groups differ in size and variance. A larger t value relative to its reference distribution suggests that the difference is unlikely to be due to chance, while a low p -value provides evidence of a meaningful disparity in housing outcomes.

Mann–Whitney U Test

The Mann–Whitney U test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to assess whether two independent groups differ in their distributions. Unlike the t -test, it does not assume normally distributed data, making it suitable for ordinal data or continuous data that deviate from normality.

Formally, the test statistic U is defined in Equation 3.13, where R_1 denotes the sum of ranks for group 1:

Formula

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad [41] \quad (3.13)$$

Here, n_1 and n_2 represent the sample sizes of the two groups, and R_1 is the sum of ranks assigned to the observations in group 1 after combining and ranking all observations. A corresponding statistic U_2 can be computed for group 2, and the smaller of the two values is typically used for significance testing.

For example, when comparing self-reported satisfaction scores between domestic and international students, the Mann–Whitney U test evaluates whether the distributions of responses differ, regardless of whether they follow a normal distribution. A low p -value from this test would suggest that differences in satisfaction are unlikely to have arisen by chance.

Chi-square Test

The Chi-square (χ^2) test is a non-parametric statistical test used to assess whether there is a significant association between two categorical variables. It compares the observed frequencies in a contingency table with the frequencies that would be expected if the variables were independent.

Formally, the Chi-square statistic is defined in Equation 3.14, where the squared difference between observed and expected frequencies is standardized by the expected frequency:

Formula

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad [2] \quad (3.14)$$

Here, O_{ij} represents the observed frequency in cell i, j of the contingency table, and E_{ij} is the expected frequency for that cell under the assumption of independence.

dence. The statistic follows a Chi-square distribution with degrees of freedom determined by the number of categories.

For example, when examining whether student status (domestic vs. international) is associated with type of accommodation (shared apartment, student dormitory, private rental), the Chi-square test evaluates whether the observed distribution differs from what would be expected under independence. A significant result would suggest that student status and housing type are not independent, but rather systematically related.

3.4. QGIS: Quantum Geographic Information System

QGIS is an open-source geographic information system widely used for visualizing, managing, and analyzing spatial data. It provides a flexible environment to integrate survey-based datasets with secondary spatial information such as road networks, administrative boundaries, and points of interest. In research contexts, QGIS is particularly valuable because it allows the creation of high-quality thematic maps, supports spatial statistics, and offers compatibility with a wide range of geospatial data formats.

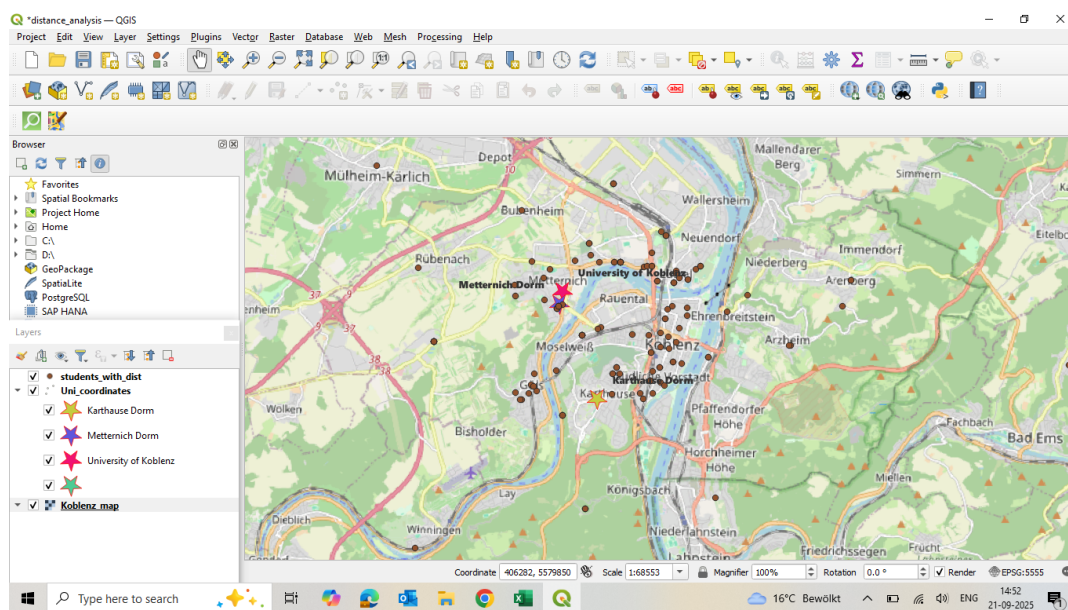


Figure 3.6.: Spatial distribution of student residences in Koblenz, visualized with QGIS.

In this study, QGIS was employed to map student residences and visualize their proximity to key amenities such as the university campus, transport hubs, and supermarkets. Figure 3.6 shows student housing locations (brown points) alongside key reference points such as the University of Koblenz and student dormitories (star

markers), enabling analysis of proximity and accessibility. In addition, clustering results such as those obtained from DBSCAN were mapped with distinct colors and symbols, allowing spatial patterns of student housing equity to be interpreted more intuitively. QGIS served both as an analytical and visualization tool, providing spatial insights that complemented statistical modeling and enhancing the interpretability of findings by translating abstract data patterns into geographically meaningful representations.

This chapter has outlined the theoretical underpinnings of the machine learning algorithms, evaluation metrics, statistical tests and QGIS tool employed in the study. By integrating regression, classification, clustering, and natural language processing models with appropriate performance measures and inferential tests, the research establishes a comprehensive methodological foundation. This combination allows for both predictive and explanatory analysis, capturing the economic, spatial, and subjective dimensions of student housing. Together, these approaches provide the analytical depth required to investigate complex patterns and ensure that the findings are both methodologically sound and empirically robust.

4. Methodology

The methodological approach of this study is designed to systematically investigate student housing experiences in the city of Koblenz, Germany, with a particular focus on affordability, accessibility, and equity. To achieve this, the research combines survey data with geospatial and statistical analyses, ensuring that both subjective perceptions and objective measures are considered. The survey provides insights into students' socio-demographic characteristics, housing choices, and satisfaction levels, while geocoded data enables the assessment of spatial accessibility to universities and essential amenities. Statistical models and hypothesis tests are then applied to evaluate differences across groups and to identify key determinants of housing outcomes. This mixed-methods design ensures robustness by integrating quantitative rigour with spatial contextualisation, thereby capturing the multifaceted nature of housing equity.

4.1. Data Collection

The primary data for this study was collected through a structured survey administered to students of the University of Koblenz residing in different parts of the city. To ensure inclusivity, the questionnaire was prepared in both English and German and distributed over a two-month period through multiple channels. These included online forums such as WhatsApp groups, student associations coordinated via the International Relations Office (IRO), and university communication networks such as Outlook. In addition, the survey link was distributed during classes to encourage student participation. The instrument was designed to capture a broad spectrum of information, including demographic characteristics, housing affordability, accessibility, and satisfaction, while also incorporating open-ended questions to enable sentiment analysis.

As shown in Figure 4.1, the final sample consisted of 159 students, of which 89 were international and 70 were domestic. This distribution highlights the diversity of the student body in Koblenz and provides a balanced basis for analyzing disparities across demographic groups.

The survey was divided into five main sections:

- Demographic Information: Email, study program (bachelor or master), gender, age, semester, student status (domestic vs. international), nationality, languages spoken, and housing status (on or off campus).

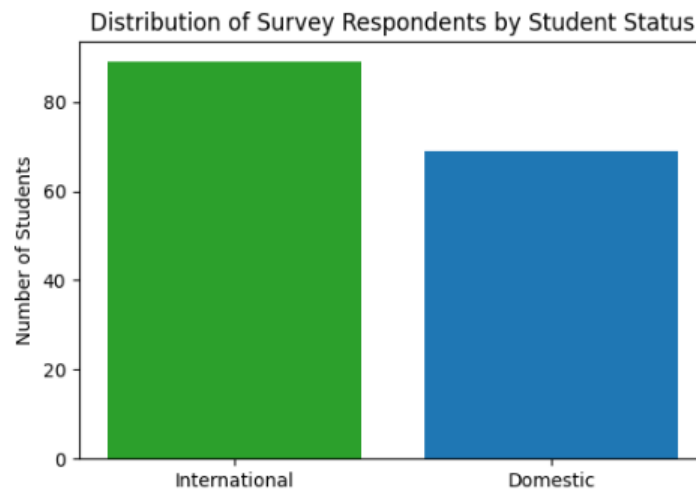


Figure 4.1.: Distribution of survey respondents by student status (Domestic vs. International).

- **Housing Search and Experience:** University support in finding accommodation, number of applications submitted before securing housing, key factors influencing accommodation choice, and reflections on what students value and the challenges they face in their current housing.
- **Affordability:** Room size, rent, type of accommodation (shared or single dwelling), number of sharing, and a subjective affordability rating.
- **Accessibility:** Address (street name and postal code of residence), primary mode of transport, commute time to campus, and proximity to essential amenities (e.g., grocery stores, gyms, and healthcare facilities) rating.
- **Satisfaction:** Likert-scale questions (1–5) on housing quality, maintenance, safety, and overall satisfaction, along with an open-ended question to collect additional qualitative insights.

In addition, the survey contained several other questions, such as factors influencing the choice of accommodation, budget considerations, and related housing experiences, which will be discussed in later sections.

By combining quantitative measures (e.g., Likert scales, numerical inputs such as rent and commute time) with qualitative responses (open-ended reflections on housing experiences), the survey followed a mixed-methods approach. This design ensured the statistical comparability of responses while also capturing deeper insights into students' lived experiences of housing in Koblenz.

4.2. Data Preprocessing

Once the survey responses were collected, several preprocessing steps were undertaken to prepare the dataset for statistical and spatial analysis. These steps ensured the accuracy, consistency, and usability of both the quantitative and qualitative information.

4.2.1. Data Cleaning

- Responses were first screened for completeness and consistency.
- Common placeholder responses such as “none,” “nil,” “not applicable,” “keine” or blank values were standardized and coded as zeros or with the mode values, depending on the context.
- Since the survey was conducted in both English and German, responses were mapped to a common English version to ensure consistency (e.g., “Ja/Yes” → Yes, “Nein/No” → No, “Vielleicht/May be” → May be).
- Dependent categorical responses (e.g., factors influencing accommodation choice, challenges in current accommodation) were also standardized across languages.
- The two datasets (English and German) were then merged into one consolidated file for further analysis.

Duplicate checks could not be performed since participants’ email addresses were masked to protect identity and ensure anonymity. However, the final dataset was carefully reviewed for completeness and structural consistency before moving into quantitative and qualitative preprocessing.

4.2.2. Quantitative Data Processing

- Numeric variables such as semester, rent, room size, and commute time were normalized into consistent units (e.g., rent per square meter, commute time in minutes) to ensure comparability across respondents.
- The number of accommodation requests and housing interviews, which were provided as textual responses with special characters, were converted into numbers and assigned to appropriate bins for comparability.
- Likert-scale responses (1–5) were already provided in numeric form and therefore did not require further cleaning.

These transformations ensured that all quantitative variables were directly comparable across respondents and prepared for subsequent statistical modelling.

4.2.3. Geocoding of Residences

- To maintain respondent privacy, students provided only street name and postal code rather than full addresses.
- Any entries marked with “privat” or “private” as the address were removed, as these did not provide valid spatial information.
- Street names were standardized by correcting common variations (e.g., endings such as “str” or “strasse” were uniformly converted to “straße”).
- The cleaned street names were then appended with the postal code and “Koblenz, Germany” to ensure compatibility with geocoding tools.
- These standardized details were used to generate approximate latitude and longitude coordinates, enabling spatial analysis of student housing relative to the university and city amenities.

This process ensured a balance between data utility for spatial analysis and protection of student anonymity.

4.2.4. Qualitative Data Processing

- Open-ended survey responses were analyzed using the nlptownBERT pre-trained sentiment analysis model, which is capable of handling multilingual text.
- Each response was passed into the model, which automatically assigned a sentiment score ranging from 1 (very negative) to 5 (very positive).
- This approach provided a systematic and scalable method for evaluating subjective housing experiences, ensuring that both English and German responses were assessed on a comparable numerical scale.

Through a combination of language harmonization, quantitative standardization, sentiment scoring of qualitative inputs, and geocoding of residence data, the survey responses were transformed into a comprehensive and structured dataset. This preprocessing ensured that the data were consistent, comparable, and ready for subsequent statistical and spatial analyses while preserving participant anonymity.

4.3. RQ1: Affordability of Student Housing

This study adopts a cross-sectional, quantitative design based on survey data collected. The primary aim is to examine affordability disparities in student housing

across demographic groups (Domestic vs International students) and spatial divisions of the city. The analysis is explanatory in nature, using both statistical inference and multivariate regression modelling to examine whether disparities exist across groups and to explore the factors that may account for them.

4.3.1. Data and Variables

The principal dependent variable is cost per square meter (€/m²), calculated as the total monthly rent (including utilities) divided by room size. This measure captures affordability relative to space, which is particularly relevant in the student rental market where room sizes vary substantially. A supplementary binary measure of rent burden is also used, coded as one if the student reports spending more than 50 percent of their monthly budget on housing, and zero otherwise.

Independent variables are grouped into three categories. First, demographic differences are captured by a binary indicator of student status (*is_student*), coded as 0 for Domestic and 1 for International students. Second, structural housing features are represented by room size (in square meters), number of sharers (the number of co-residents parsed from survey responses), and accommodation type (shared room, shared apartment, single apartment, dormitory). Third, geographic variation is captured by a region variable constructed from geocoded student addresses. Both unsupervised clustering (K-means) and grid-based latitude/longitude binning were tested; however, to avoid sparse categories, the final model collapses these into a binary indicator distinguishing central from non-central locations.

4.3.2. Analytical Strategy

The analysis proceeded in three stages. First, descriptive statistics and visualizations were produced to illustrate initial differences in affordability. Group means of cost per square meter were compared between Domestic and International students, and boxplots were used to visualize the distributions. The proportion of students exceeding the 50 percent rent burden threshold was also calculated.

Second, inferential statistical tests were conducted to formally assess group differences. Welch's t-test was used to compare mean affordability between Domestic and International students, and Cohen's d quantified the effect size of the difference. For the binary rent burden outcome, a simple logistic regression was estimated with rent burden as the dependent variable and student status as the sole predictor.

4.3.3. Regression Models

Third, multivariate regression modeling was undertaken to separate direct demographic effects from structural and geographic influences. Three nested Ordinary Least Squares (OLS) regression models were estimated.

- **Model 1:**

$$Rent \sim student_status + room_size + accom_code$$

- **Model 2:** examined raw disparities by regressing cost per square meter on student status only.

$$\text{cost_per_m}^2 \sim \text{student_status}$$

- **Model 3:** introduced structural controls by adding room size, number of sharers, and type of accommodation.

$$\text{cost_per_m}^2 \sim \text{student_status} + \text{room_size} + \text{num_sharing} + \text{accom_code}$$

- **Model 4:** added geographic controls by including the central versus non-central regional indicator.

$$\text{cost_per_m}^2 \sim \text{student_status} + \text{room_size} + \text{num_sharing} + \text{accom_code} + \text{region_binary}$$

- **Model 5:** used logistic regression with rent burden as the binary outcome, estimated as a function of student status only.

$$\text{rent_burden}^2 \sim \text{student_status}$$

Comparing the coefficient of student status across these models allows for distinguishing between direct demographic bias (persistent even after controls) and structural or systemic bias (attenuated once controls are included).

Several steps were taken to ensure the robustness of results. Outliers in cost per square meter were identified through boxplots and residual diagnostics, and models were re-estimated excluding extreme values. Alternative definitions of region were tested, including K-Means clustering with different cluster sizes and quantile-based grids. The binary central versus non-central definition was ultimately adopted for stability and interpretability. Nonlinear effects of room size were also explored by introducing quadratic terms, though the relationship between room size and cost per square meter remained approximately linear.

4.4. RQ2: Accessibility of Student Housing

This research investigates how geographic accessibility to the university campus and essential amenities differs across student demographics. In particular, the analysis examines disparities between domestic and international students—as well as other subgroups—in relation to their proximity to campus, transportation hubs, and everyday services such as supermarkets and healthcare facilities. By integrating geospatial analysis with survey responses, the study seeks to capture both the objective spatial distribution of student housing and the subjective perceptions of accessibility, thereby providing a comprehensive assessment of how location shapes student experiences.

4.4.1. Data and Variables

The combined dataset of survey responses and secondary spatial data. The survey provided the geographic coordinates (latitude and longitude) of student residences, demographic characteristics including student status (domestic vs. international) and gender, self-reported travel times, and modes of transport used to commute to the university. In addition, the survey included perception-based questions on accessibility, such as whether students believed living closer to campus would affect their attendance, academic performance, or extracurricular participation, the number of days per week they attended university, and their preferred residential distance from campus.

Secondary data on essential services—including supermarkets, healthcare facilities, and transit stops—was retrieved from OpenStreetMap (OSM) using the OSMnx library. Campus coordinates were collected from official university sources. These datasets were integrated with the survey data to produce a comprehensive base for analyzing both objective and subjective accessibility outcomes.

4.4.2. Accessibility Measure

Accessibility was measured in two primary ways.

- Euclidean (straight-line) distances between each student’s residence to the campus building and to the Koblenz Hauptbahnhof were computed as a baseline measure of proximity. However, this measure has clear limitations, as straight-line distance does not reflect actual travel routes, which in practice vary due to the street network. In the context of Koblenz—a city shaped by rivers and dependent on bridge connectivity—the discrepancy between straight-line and real travel distances can be particularly pronounced.
- Network-based driving distances were calculated using OSMnx street networks, representing the shortest path along the road network from each student residence to campus. Distances to the nearest supermarket, healthcare facility, and transit stop were calculated via nearest-neighbour spatial joins in GeoPandas.

To capture broader spatial patterns, DBSCAN clustering [19] was applied to the latitude and longitude coordinates of student housing. This density-based method identified distinct geographic clusters of student residences, enabling assessment of whether domestic and international students were unevenly distributed across the urban area.

4.4.3. Analytical Strategy

The analysis was conducted in multiple stages. Descriptive statistics (means, medians, and standard deviations) were calculated for distances to campus and ameni-

ties, disaggregated by student status and gender, to provide an overview of accessibility patterns. Inferential tests were applied to compare groups: Welch’s t-tests for continuous measures (distances, travel times), Mann–Whitney U for skewed distributions, and Chi-square for categorical perception responses. Regression models were estimated to test whether accessibility outcomes were systematically associated with demographic characteristics and travel behaviours.

- **Model 1:** distance to campus as the dependent variable, with student status and travel time as predictors, to examine whether background differences persisted after accounting for reported commuting effort.

$$dist_to_campus \sim student_status + travel_time$$

- **Model 2:** travel time as the dependent variable, predicted by student status, driving distance, and mode of transport, to evaluate how transport mode shapes the distance–time relationship.

$$travel_time \sim student_status + dist_to_campus + mode_transport$$

- **Model 3:** because transport modes were unevenly distributed, a restricted regression was run only for bus commuters (n = 106), the only subgroup with sufficient sample size for robust inference.
- **Model 4:** the days per week at university were modelled as a function of student status, gender, and distance to campus, connecting objective accessibility to behavioural frequency.

$$days_per_week \sim student_status + dist_to_campus + gender$$

4.4.4. Perceived and Behavioural Accessibility

Complementing the objective metrics, survey responses were analyzed to capture the perceived and behavioural dimensions of accessibility. These included items asked whether students perceived that living closer would affect attendance, performance, or extracurricular involvement. Perception-based variables were compared across groups using chi-square tests of independence, while attendance frequency was modeled with an OLS regression mentioned in Model 4. This mixed approach ensured that disparities in accessibility were not only measured in terms of physical proximity, but also in terms of students’ lived experiences and expressed needs.

Spatial analysis and visualization were conducted using QGIS [57], which provided an interactive environment for integrating survey data with secondary spatial datasets. Student housing coordinates and campus locations were imported as separate layers. Distances and spatial joins computed in Python were cross-validated and displayed in QGIS to ensure consistency. Cluster maps from the DBSCAN algorithm were overlaid on city basemaps to highlight concentrations of student

housing. Beyond mapping, statistical outputs were visualized using boxplots and stacked bar charts to compare distributions of distances, travel times, and categorical responses across demographic groups. These visualizations made the results interpretable by combining spatial context with group-level comparisons.

By combining Euclidean and network-based distance calculations, clustering, regression modeling, and categorical tests, the study examined both physical and perceived accessibility disparities. The use of QGIS for mapping, alongside statistical modeling in Python, ensured that accessibility was analyzed within both a spatial and a behavioural framework. This design distinguishes between disparities arising from geographic disadvantage and those shaped by differing perceptions of accessibility.

4.5. RQ3: Housing Satisfaction

This study employed a mixed-methods analytical framework, combining quantitative and qualitative survey analysis with machine learning and natural language processing (NLP) to investigate fairness in student housing experiences. The approach integrated (i) Likert-scale survey responses on housing satisfaction dimensions, (ii) open-ended text responses analyzed with multilingual sentiment analysis and topic modeling, and (iii) both supervised and unsupervised machine learning models to uncover trends and recurring concerns. The multi-method design was chosen to ensure both statistical rigor (quantitative tests of differences between groups) and rich contextual insights (textual analysis of lived experiences), enabling triangulation of findings [14].

4.5.1. Sentiment Analysis

To quantify satisfaction expressed in open-ended survey responses, sentiment analysis was implemented using the `nlptown/bert-base-multilingual-uncased-sentiment` model from Hugging Face's transformers library [76]. This model is a variant of BERT [17] fine-tuned for sentiment classification across multiple languages, including English and German, which made it appropriate for the multilingual nature of the survey data. The model was accessed through the Hugging Face pipeline API, using the following components:

- **Tokenizer (`AutoTokenizer`):** converted raw text into subword tokens recognizable by the model.
- **Model (`AutoModelForSequenceClassification`):** applied pretrained weights to classify text into one of five sentiment categories.
- **Pipeline (`pipeline("sentiment-analysis")`):** provided an end-to-end interface for generating sentiment predictions.

Each open-ended response was passed directly to the pipeline without additional preprocessing, as the model's tokenizer internally handled text normalization and encoding. The output was a probability distribution over five sentiment classes (from 1 = very negative to 5 = very positive). For each response, the most likely label was extracted and mapped onto a 1–5 numeric scale, representing the respondent's sentiment-derived satisfaction score. These scores were then aggregated at the respondent level and compared across groups (Domestic vs. International) to identify systematic disparities in expressed satisfaction. This approach provided a quantitative measure of subjective experiences, complementing structured Likert ratings and enabling integration into the composite satisfaction index. To further explore fairness concerns, responses with negative sentiment (scores ≤ 2) were isolated, and keyword frequency analysis was performed. In addition, BERTopic [28] was applied to cluster these responses into recurring thematic concerns, thereby complementing quantitative Likert findings with qualitative insights.

4.5.2. Statistical Analysis of Likert Data

Parallel to sentiment analysis, Likert-scale items measuring housing satisfaction were analyzed using inferential statistics. The survey included multiple dimensions such as affordability, room size, accessibility, safety, facilities, cleanliness, expectations, and overall satisfaction. Each item was coded on a five-point ordinal scale, where higher values represented greater satisfaction. Although Likert data is ordinal, it is common in housing and social research to treat aggregated scores as approximately interval-scaled for group comparison [48]. Independent-sample Welch's t-tests were conducted to compare domestic and international students across each satisfaction dimension, as this test does not assume equal variances [61]. To assess the practical significance of differences, Cohen's d effect sizes were computed and classified following Cohen's [11] guidelines: $d < 0.2$ negligible, 0.2–0.5 small, 0.5–0.8 medium, >0.8 large. In addition, a composite satisfaction index was constructed by averaging the mean Likert score with the sentiment-derived score, yielding a more holistic indicator of student housing satisfaction that incorporated both structured and unstructured data. Group differences on this composite measure were also tested using Welch's t-test.

4.5.3. Supervised Machine Learning

To move beyond group mean comparisons and identify the most important predictors of dissatisfaction, supervised learning techniques were applied. A binary outcome variable was defined: students reporting an overall satisfaction score ≤ 3 were categorized as dissatisfied (1), while those above this threshold were coded as satisfied (0). Two models were implemented.

- First, Logistic Regression was used to estimate the relationship between dissatisfaction and multiple predictor variables, including affordability, room size,

facilities, expectations, and safety. Logistic regression was selected for its interpretability, as coefficients can be directly translated into the odds of being dissatisfied [33].

- Second, a Random Forest Classifier [7] was trained to capture non-linear relationships and rank feature importance. This ensemble method constructs multiple decision trees and aggregates their predictions, providing robust estimates and relative variable importance scores. To address class imbalance (fewer dissatisfied students), class weights were applied.

Supervised models (Logistic Regression and Random Forest) were employed to identify predictors of dissatisfaction. Logistic regression coefficients provided interpretable estimates of direction and strength, while Random Forest ranked features by their relative importance.

4.5.4. Unsupervised Machine Learning

To explore latent groupings of students based on their housing satisfaction profiles, unsupervised clustering was employed. The analysis used the KMeans algorithm [39], which partitions observations into k clusters by minimizing within-cluster variance. Prior to clustering, all Likert-scale satisfaction variables were standardized using z -scores to ensure comparability across dimensions, as the original items were measured on the same ordinal scale but could still exhibit different variances. The optimal number of clusters was determined using the silhouette coefficient [60], which evaluates clustering quality based on cohesion (similarity within clusters) and separation (dissimilarity between clusters). Once the optimal k was identified, KMeans was fitted to the standardized data to generate cluster memberships. These cluster assignments were subsequently cross-tabulated with student demographic variables (domestic vs. international status) to examine whether specific groups of students were disproportionately represented in particular housing experience clusters.

By combining sentiment analysis, statistical testing, and machine learning, this study addressed a multi-layered approach. Sentiment analysis captured recurring concerns in students' own words, while Likert tests quantified disparities across specific housing dimensions. Supervised models identified which features most strongly predicted dissatisfaction, and unsupervised clustering revealed well defined clusters. The composite satisfaction index further validated that disparities persist across both structured and unstructured measures. Together, these methods allow student survey data to be analyzed for trends in housing satisfaction and to investigate whether recurring disparities raise concerns about fairness.

4.6. Equity in Housing

This study adopts a multi-dimensional approach that integrates demographic, geographic, and perceptual data to assess equity in student housing. Unlike earlier analyses that treated affordability, accessibility, and satisfaction separately, this stage combines them into a unified framework to evaluate whether certain student groups—particularly international students—experience systematic disadvantages. By employing statistical modelling, sentiment analysis, and spatial techniques, the methodology ensures that both objective conditions and subjective experiences are incorporated. This integrated design provides a robust basis for examining whether equity in student housing is distributed evenly across groups and locations in Koblenz.

4.6.1. Data and Variables

To examine equity in student housing, the cleaned survey dataset, containing demographic information (domestic or international status, gender), housing characteristics (rent, room size, location coordinates), accessibility indicators (distance and commute time to campus), and measures of housing satisfaction. Rent per square meter (€/m²) was considered as the primary indicator of affordability. Accessibility was represented through the objective distance from each residence to campus, while satisfaction was represented by a composite score integrating Likert-scale responses and sentiment-based evaluations, ensuring that both structured and unstructured perspectives were incorporated.

4.6.2. Statistical and Spatial Analysis

Multiple linear regression (MLR) was applied to estimate the influence of demographic status, distance to campus, and room size on affordability and satisfaction. MLR is a standard method for identifying independent effects while controlling for confounding variables [78]. To compare overall fairness across groups, a composite equity index was constructed by standardizing affordability, accessibility, and satisfaction scores, following composite indicator methodology used in social equity research [50]. Group differences were then tested using independent-sample t-tests.

To capture spatial heterogeneity, Geographically Weighted Regression (GWR) was employed, a method that relaxes the assumption of global stationarity and allows coefficients to vary across space [22]. Implemented via the `mgwr` Python package [49], GWR was used to model affordability as a function of demographic status, distance, and room size, identifying localized areas of inequity.

The findings were supported through a set of visualization techniques. Radar charts were used to compare equity profiles of domestic and international students across the three dimensions. Correlation heatmaps illustrated systemic relationships between affordability, accessibility, and satisfaction factors. Boxplots presented the distribution of the equity index by subgroup.

To conclude this chapter outlined the mixed-methods framework used to study student housing in Koblenz. A bilingual survey provided both quantitative and qualitative data, which were cleaned, standardized, geocoded, and prepared for analysis. Statistical modeling, hypothesis testing, geospatial techniques, and sentiment analysis were applied to examine affordability, accessibility, satisfaction, and equity. Together, these methods form a rigorous foundation for the results presented in the next chapter.

5. Results and Discussion

This chapter integrates the presentation of findings with their interpretation in relation to the study's objectives. The preceding chapters established the importance of examining affordability, accessibility, satisfaction, and equity with attention to the possibility of bias in student housing outcomes. In this chapter, the analysis presents the empirical results and identifies where such biases are evident, how they manifest across groups, and what implications they hold for fairness in student housing.

By combining empirical evidence with critical discussion, this chapter moves beyond descriptive reporting to consider the implications of the findings. Visualizations, statistical outputs, and thematic insights are presented to illustrate disparities across student groups, while the discussion interprets these patterns in light of theoretical and practical concerns. This integrated structure ensures that the chapter remains both data-driven and analytically meaningful.

To set the stage for the subsequent analyses, it is useful to first examine disparities in the housing application process itself. Access to accommodation begins well before affordability or location are considered, and the number of applications students must submit provides a clear indicator of structural barriers. If international students are systematically required to submit more applications than their domestic peers, this points to unequal access and heightened competition from the very outset. Figure 5.1 illustrates this initial bias in the housing search.

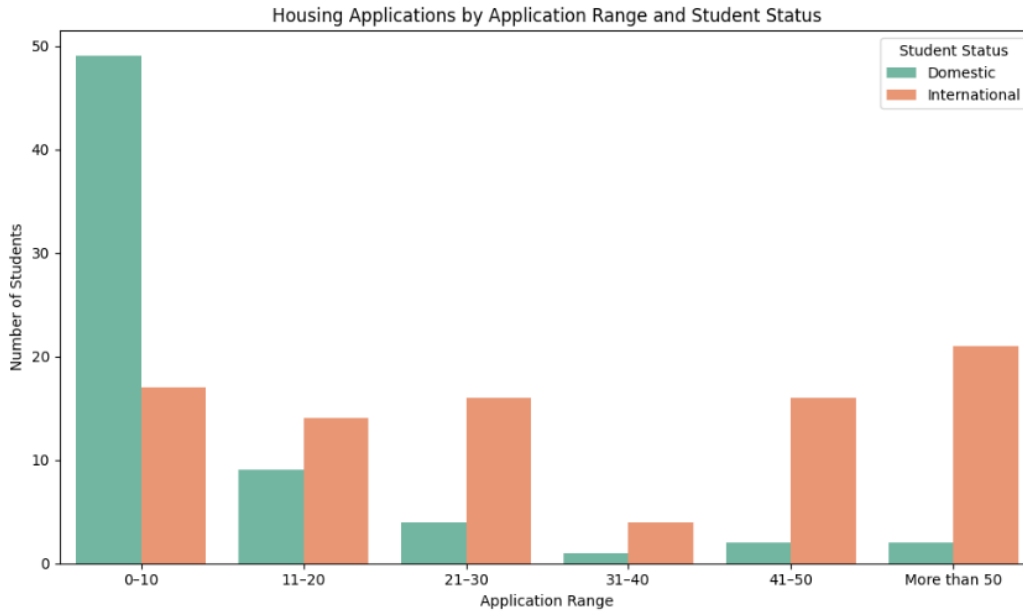


Figure 5.1.: Housing applications by domestic and international students across application ranges.

The distribution of housing applications reveals a striking divergence between domestic and international students. Domestic students are heavily concentrated in the lowest range (0–10 applications), indicating that most were able to secure accommodation after only a few attempts. In contrast, international students appear disproportionately in the higher ranges, with many reporting 20 or more applications and a significant share submitting over 50. This pattern suggests that international students face greater obstacles in accessing housing, such as limited networks, language barriers, or discriminatory practices, which force them to expend considerably more effort in their search. The imbalance observed here highlights an early manifestation of structural bias that precedes and compounds subsequent challenges of affordability, accessibility, and satisfaction.

While the number of applications reflects the effort required to secure housing, an equally important dimension is whether students had any real choice between multiple accommodation options. For some, the process ends with several viable alternatives, allowing them to select housing that best fits their needs. For others—especially those facing structural barriers—the lack of alternatives means being compelled to accept whatever option becomes available, even if it is far from ideal. Figure 5.2 illustrates this contrast between domestic and international students.

Availability of Alternative Accommodation Options by Student Status

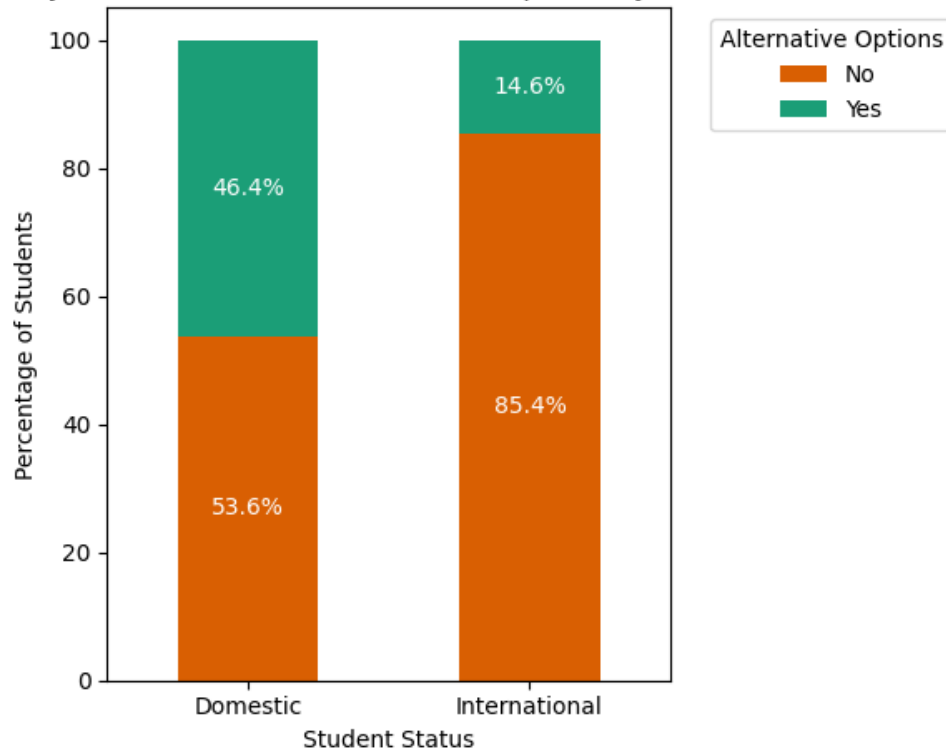


Figure 5.2.: Availability of alternative accommodation options among domestic and international students.

The results reveal that domestic students are nearly evenly split, with around 47% reporting that they had more than one housing option to choose from. In contrast, only 14.6% of international students reported having such alternatives, with the overwhelming majority forced to accept the only option available to them. This lack of choice underscores a deeper structural disadvantage: while domestic students retain a degree of agency in their housing decisions, international students are often compelled to compromise, accepting accommodation that may be less affordable, less accessible, or lower in quality due to limited opportunities.

Beyond the challenges of securing accommodation, it is also important to understand what criteria students prioritize when making housing decisions. Examining these choice factors highlights the underlying trade-offs between affordability, location, and quality that frame the analyses in the following sections.

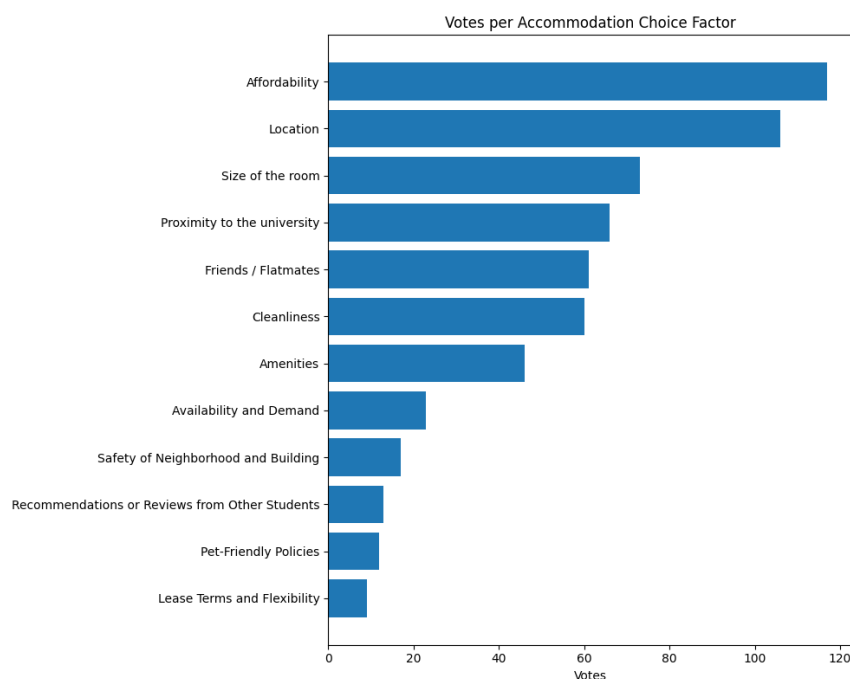


Figure 5.3.: Key factors influencing accommodation choice among students.

The results show that affordability and location are the most influential factors shaping student housing choices, followed by room size and proximity to the university. Elements such as facilities, social environment, and landlord reputation were mentioned less frequently. These priorities emphasize the centrality of financial and spatial considerations in shaping housing outcomes and directly connect to the affordability and accessibility dimensions explored in RQ1 and RQ2. With this context in place, the analysis now turns to RQ1, which examines affordability in student housing and investigates whether systematic disparities exist across groups.

5.1. RQ1: Affordability of Student Housing

Affordability represents a central dimension of equity in student housing. It captures both the absolute costs students face and the relative burden of these costs compared to their financial resources. For this analysis, affordability was examined through multiple lenses:

- (i) absolute rent levels,
- (ii) rent per square meter ($\text{€} / \text{m}^2$), capturing the relative cost of housing adjusted for size, and
- (iii) rent burden, defined as spending more than 50 percent of monthly income on housing.

5.1.1. Regression Models

To examine affordability disparities in student housing, a series of regression models (1-5) were estimated, each capturing a different dimension of the rental market. Model A begins with absolute rent, assessing whether student status, accommodation type, or room size explains variation in costs. Models B through D shift the focus to affordability per square meter (€/m²), progressively introducing housing structure and geographic location to determine whether disparities arise from demographic or structural factors. Finally, Model E uses a logistic specification to evaluate rent burden, defined as spending more than 50 percent of one's budget on rent, in order to capture the lived financial strain of housing costs. Together, these models provide a layered understanding of how affordability is shaped by both market characteristics and student demographics as depicted in Table 5.1.

Model	Dependent Variable	R ² /Pseudo R ²	Adj. R ²	t/z-value	Key Finding
Model 1	Rent (absolute)	0.461	0.449	t = 9.894 (Room Size)	Room size highly significant; student status and accommodation type not significant.
Model 2	Cost per m ²	0.065	0.045	t = 2.147 (Student Status)	International students pay significantly higher €/m ² .
Model 3	Cost per m ²	0.318	0.298	t = -7.096 (Room Size)	Room size strongly significant; student status weakens (n.s.); smaller rooms costlier per m ² .
Model 4	Cost per m ²	0.303	0.289	t = -5.633 (Room Size); t = 2.646 (Status)	Room size remains significant; international status regains significance; region not significant.
Model 5	Rent Burden (Logit)	0.024	-	z = 2.231 (Student Status)	International students more likely to face rent burden (>50% of budget).

Table 5.1.: Regression Models for Affordability

Model 1 (absolute rent) demonstrates that room size is the dominant predictor of rental cost, with each additional square meter adding roughly €8.6 to rent. Student status (domestic vs. international) and accommodation type did not show significant effects at this stage.

Model 2 focusing on €/m², identifies a demographic disparity: international students pay on average €3.62 more per square meter than domestic peers, even after controlling for housing type and sharing arrangements. As visualized in Figure 5.4,

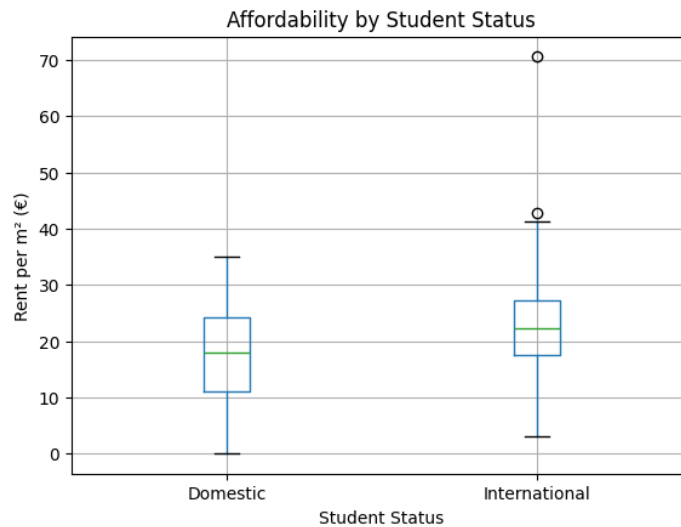


Figure 5.4.: Affordability comparison by student status

international students' €/m² costs are not only higher on an average but also more dispersed, reinforcing the statistical disparity highlighted in Model 2.

Model 3 building on Model 2, introduces room size as an additional control. Once room size is accounted for, the previously significant effect of student status disappears, while room size emerges as a highly significant predictor. This indicates that smaller rooms are disproportionately more expensive on a per-square-meter basis, and international students' concentration in such units largely explains the affordability gap observed in Model 2.

Model 4 which adds geographic location as a further control, shows that room size remains the strongest predictor of rent per square meter, confirming that smaller rooms are systematically more expensive. Importantly, once geography is included, the coefficient for international student status becomes significant again, indicating that international students pay higher rents per square meter than domestic peers even after accounting for both room size and location. By contrast, geographic factors themselves do not exert a significant influence on affordability in this sample, suggesting that inequities are primarily rooted in structural housing characteristics (such as room size) and demographic differences (domestic vs. international), rather than spatial location within the city.

Model 5 uses logistic regression to capture rent burden (>50% of income). The results show that international students are significantly more likely to experience rent burden, meaning that affordability disparities manifest as concrete financial pressure on their everyday living conditions.

In Figure 5.5 international students are visibly more likely to be rent-burdened than domestic students, reinforcing the logistic regression results from Model E.

The regression results reveal three distinct dimensions of bias in student housing

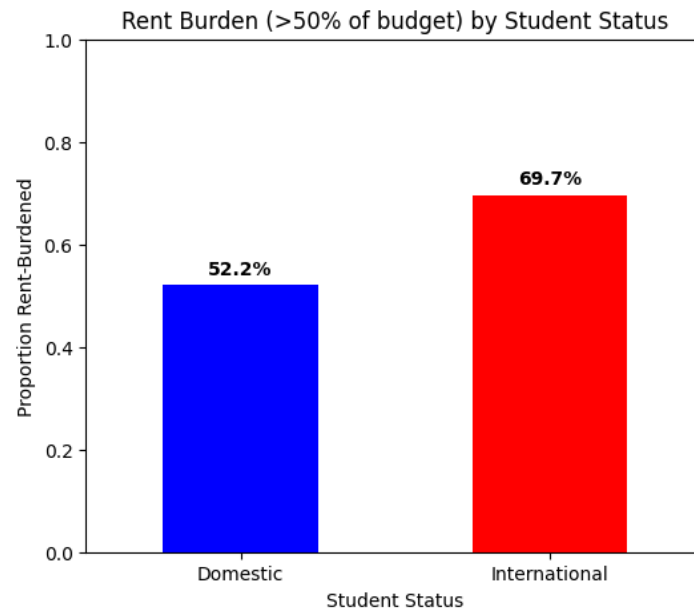


Figure 5.5.: Proportion of students rent-burdened (>50% of monthly budget) by status

affordability:

- **Structural bias:** Smaller rooms are disproportionately expensive on a per-square-meter basis, and international students are overrepresented in such units. This pattern emerges consistently in Models 1, 3, and 4, where room size is the most significant predictor of affordability outcomes.
- **Demographic bias:** International student status itself is associated with higher costs. In Model 2, international students pay significantly more per m²; in Model 4, status regains significance even after accounting for room size; and in Model 5, international students are more likely to exceed the rent-burden threshold. These results highlight a persistent demographic inequity beyond structural market factors.
- **Geographic bias:** Location effects were tested in Model 4 and found to be statistically insignificant. Central versus peripheral housing did not systematically influence affordability, indicating that geographic disparities are not a major driver of inequity in this sample.

Together, these findings suggest that affordability inequities are primarily driven by structural and demographic factors, rather than spatial location.

Collectively, these models reveal that affordability inequities in Koblenz student housing are both structural and demographic. Structural bias arises because smaller

rooms, which are disproportionately occupied by international students, command a per-square-meter premium. Demographic bias emerges because even after accounting for room size, international student status remains a predictor of higher rents and greater rent burden. Geographic location does not significantly shape affordability, underscoring that inequities are embedded within the housing market rather than urban geography.

These results align with broader findings in student housing research: international students, lacking local knowledge and networks, are more likely to be funneled into high-cost, small-scale units, amplifying their financial strain. The combination of regression results and visual evidence underscores the need for targeted housing support, particularly the expansion of affordable dormitory spaces and advisory services to help international students avoid the structural traps of the private rental market.

5.2. RQ2: Accessibility of Student Housing

Accessibility was assessed as both an objective and perceived dimension of student housing. Objective measures were derived from geographic coordinates of student residences, distances to the university campus, and proximity to public transport hubs. Perceived measures were drawn from survey questions on commute times, transport modes, and whether students believed living closer to campus would improve their attendance or academic performance. By combining spatial analysis with statistical comparisons, this section examines how accessibility differs across student groups, with a focus on disparities between domestic and international students.

5.2.1. Regression Models of Accessibility

To examine the structural factors shaping accessibility, a series of regression models were estimated. These models tested how distance to campus, travel time, transport mode, and student status interact to explain objective measures of accessibility, as well as how these factors influence academic participation. The results are summarised in Table 5.2.

- **Model 1: Distance to Campus** This model confirmed that commute time is a reliable predictor of physical distance, validating the use of self-reported travel times as a proxy. Student status, however, was not significant, indicating that international and domestic students are not systematically separated in terms of how far they live from campus.
- **Model 2: Travel Time – All Students** When predicting commute time, distance was by far the strongest factor, while student status again showed no significant effect. This suggests that differences in commute duration are primarily explained by where students live rather than demographic status.

Model	Dependent Variable	R ²	Adj. R ²	Key Predictors	Key Findings
Model 1	Distance to Campus	0.196	0.186	Travel Time (p < 0.001); Student Status (n.s.)	Distance strongly predicted by commute time; no systematic difference by student status.
Model 2	Travel Time (All Students)	0.323	0.315	Distance (p < 0.001); Student Status (n.s.)	Commute time explained primarily by distance; no significant group difference once distance is controlled for.
Model 3	Travel Time (Bus Users)	0.384	0.373	Distance (p < 0.001); Student Status (n.s.)	Among bus users, distance predicts commute length; no significant difference between international and domestic students.
Model 4	Attendance (Days per Week)	0.088	0.077	Student Status (p < 0.001); Distance (n.s.)	International students attend on average 0.85 fewer days/week; distance not significant.

Table 5.2.: Regression models of accessibility (RQ2).

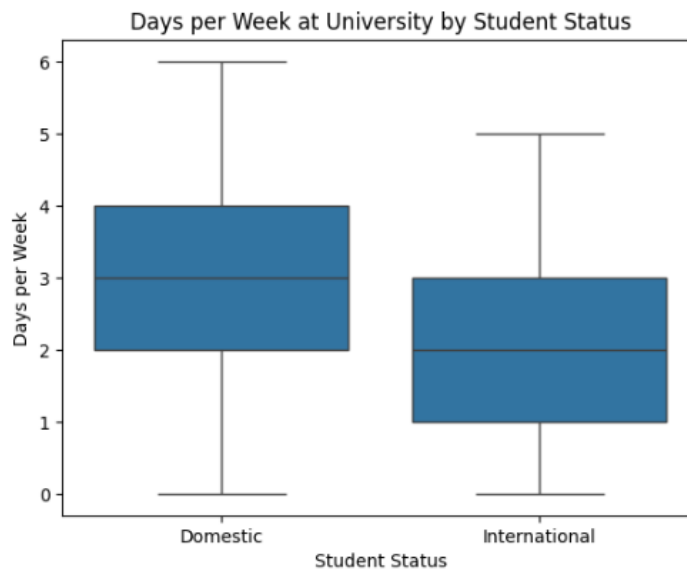


Figure 5.6.: Days per week at University by student status

- **Model 3: Travel Time – Bus Users** Restricting the analysis to bus users produced a similar pattern: travel times scaled with distance, and no systematic difference was observed between domestic and international students once distance was controlled for. This indicates that inequities do not lie in the functioning of the bus system itself, but rather in where students reside.
- **Model 4: Attendance Frequency** Unlike the distance and travel time models, attendance frequency revealed a demographic disparity. International students reported attending significantly fewer days per week than domestic students, independent of distance. This suggests that accessibility barriers for international students extend beyond geography into patterns of academic participation. Figure 5.6 illustrates this difference. Domestic students typically attended 3–4 days per week, while international students clustered lower, often attending only 2 days. The distribution also shows a longer tail of low attendance among internationals, reinforcing the regression result that reduced participation is a demographic rather than geographic phenomenon.

“Across the models, a clear distinction emerges: objective accessibility measures (distance, travel time) do not show demographic bias once distance is controlled, whereas academic participation does. This implies that although international and domestic students may live at comparable geographic distances from campus, international students face additional barriers that reduce their effective access. These barriers extend beyond distance itself and include greater financial strain from higher housing costs, limited flexibility in commuting due to heavier reliance on public buses rather than private cars or bikes, language and administrative hurdles in nav-

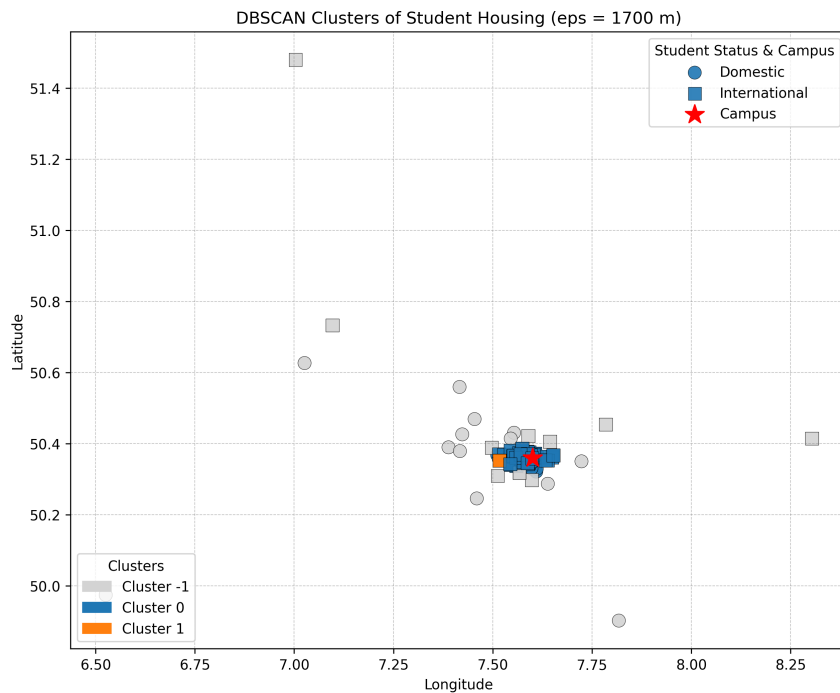


Figure 5.7.: DBSCAN clusters of student housing

igating local transport systems, and broader integration challenges that may discourage frequent campus attendance or involvement in extracurricular activities.

5.2.2. DBSCAN Clustering of Student Housing

To complement the regression models, DBSCAN clustering was applied to the geocoded residence coordinates in order to identify spatial patterns of student housing. The clustering was performed with a neighborhood radius (eps) of 1700 meters, which produced two main clusters (Cluster 0 and Cluster 1) as well as several noise points that did not belong to any cluster.

Figure 5.7 shows the DBSCAN clusters in relation to the university campus. The majority of student residences fell into two clusters located around the central campus area, with a mix of domestic and international students in both. This indicates that most students live in broadly similar geographic zones close to the university.

However, a number of noise points (shown in grey) were identified outside of the main clusters. Many of these correspond to students living at considerable distances from campus, often in peripheral or isolated areas. Importantly, international students were disproportionately represented among these noise points, suggesting that they are more likely to live in less connected or more distant housing locations. The DBSCAN results were complemented with QGIS mapping of all student residences, the campus (Figure 5.8). This representation shows the spatial layout of

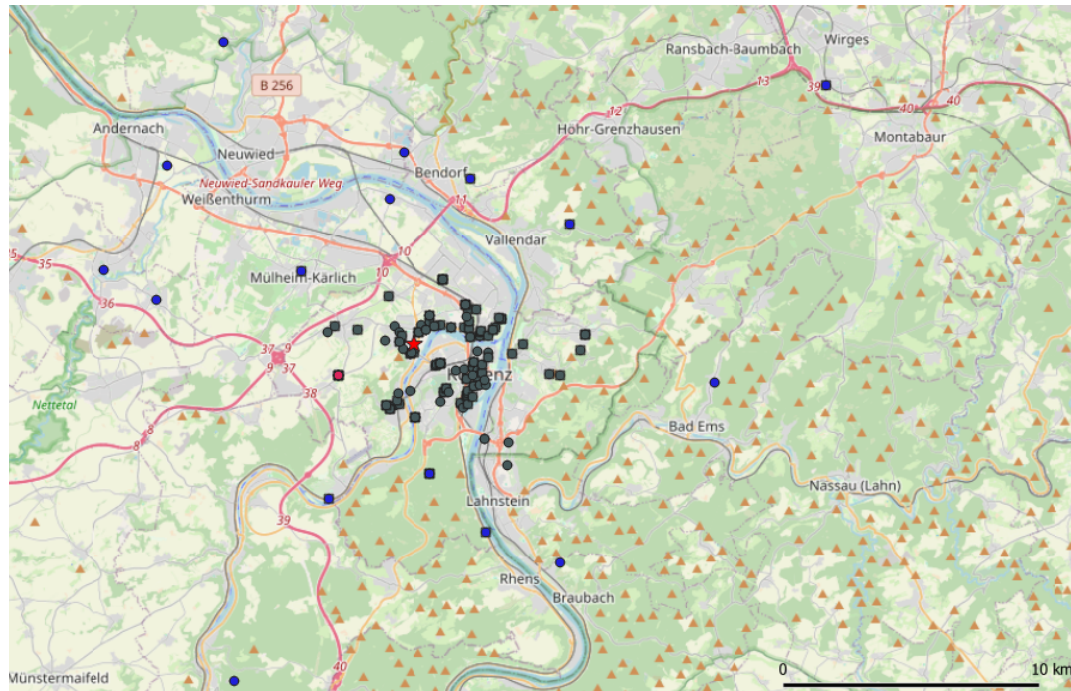


Figure 5.8.: QGIS spatial representation of student residences in Koblenz on OSM (OpenStreetMap) basemap

Koblenz and confirms the clustering results: domestic students are more concentrated in central neighbourhoods within walking and cycling distance of the university, while international students are more frequently found in peripheral areas such as Karthause, pffafendorf. Importantly, some international students reside in locations that are poorly connected to the main transport network, contributing to longer commute times and reduced accessibility.

Together, the DBSCAN clustering and QGIS mapping highlight that accessibility inequities are better captured in spatial distribution than in the regression results. The regression models showed that distance is the dominant predictor of commute time, with no systematic differences between domestic and international students once distance is controlled. However, the clustering and mapping reveal that international students are more frequently represented among outlier residences located further from campus or in peripheral neighbourhoods.

This suggests that while regression did not detect a significant demographic effect in average commute times, spatial methods uncover patterns of dispersion and marginality that are not visible in linear models. In other words, the inequity lies less in day-to-day commute efficiency and more in the structural reality that international students are more often housed in isolated or poorly connected areas.

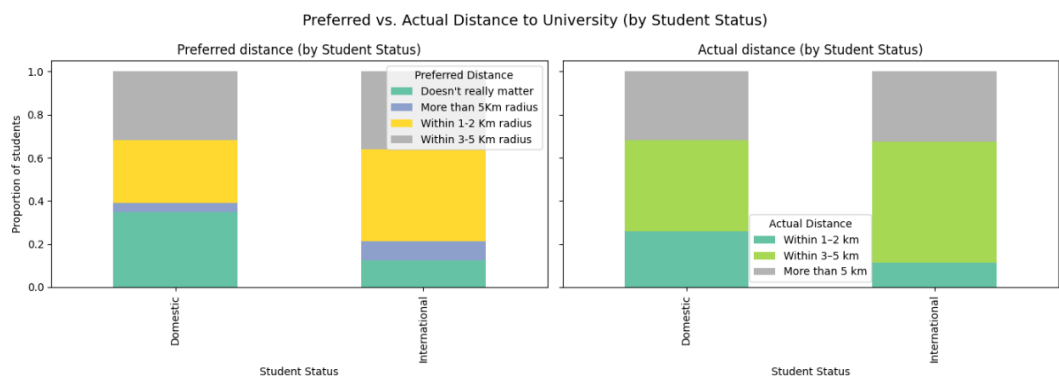


Figure 5.9.: Preferred vs. actual distance to university by student status

5.2.3. Perceived Dimensions of Accessibility

In addition to objective measures such as distance and travel time, the survey captured students' subjective perceptions of accessibility and its influence on academic participation. These indicators provide insight into how students experience their commute and whether they believe proximity to campus matters for their studies.

Preferred vs. Actual Distance.

Figure 5.9 compares students' preferred distance to campus with their actual residential locations. Domestic students were more likely to indicate that distance "does not really matter" or that living within a 3–5 km radius was sufficient. In contrast, international students expressed a much stronger preference for living within 1–2 km of campus. However, their actual residences tell a different story: both domestic and international students are most frequently located within the 3–5 km range, but international students are underrepresented in the 1–2 km range and overrepresented in the 3-5 km+ category. This mismatch reveals an aspiration–outcome gap, where international students want to live closer but are less able to secure central accommodation.

Perceptions of Accessibility

Figures 5.10 and 5.11 illustrate how students perceive the role of proximity in their academic engagement. The heatmaps reveal a striking divergence between groups. For attendance, a majority of domestic students (56.5%) stated that proximity would not matter, while only few felt it would help. By contrast, nearly three-fifths of international students (58.4%) reported that proximity would improve attendance, with just 12% indicating no effect. For academic performance, the pattern is similar. Almost half of domestic students (49.3%) believed distance made no difference, compared to just over a third (34.8%) who expected improvements. Among

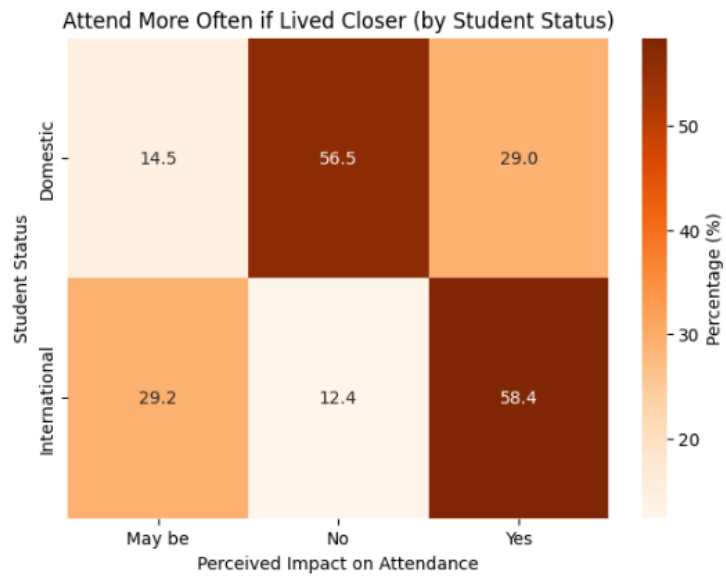


Figure 5.10.: Perceived impact of proximity on attendance, by student status

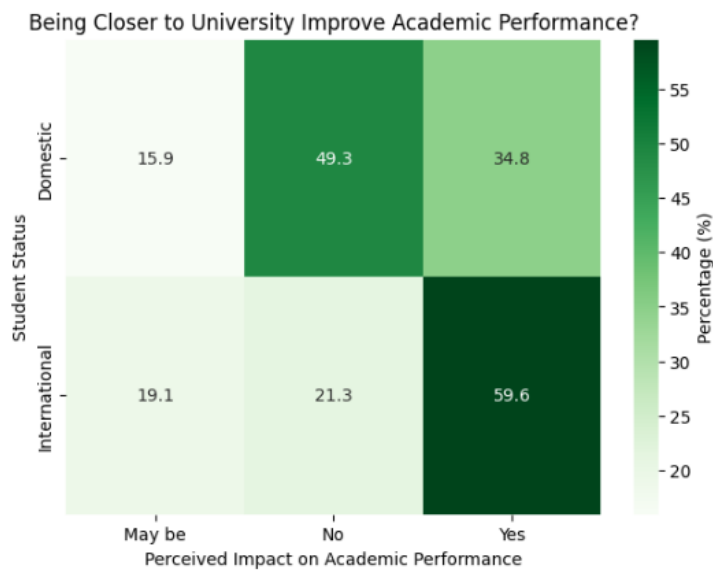


Figure 5.11.: Perceived impact of proximity on academic performance, by student status

international students, however, nearly 60% reported that proximity would enhance performance, while only one-fifth (21%) said it would not. Together, these figures demonstrate a consistent perception gap: domestic students are more likely to view proximity as irrelevant, whereas international students strongly associate it with better attendance and academic outcomes.”

The regression models showed that commute time is overwhelmingly explained by network distance, with no significant demographic effect once distance is controlled. Spatial clustering (DBSCAN) and QGIS mapping, however, revealed that international students are more often dispersed into peripheral or isolated residences, making them overrepresented among outliers. Perception-based measures deepen this picture: international students expressed a strong preference for living closer to campus but are disproportionately housed further away, and they were also far more likely to state that proximity would improve their attendance and academic performance. Taken together, these results suggest that while objective accessibility measures appear similar across groups, the lived experience of distance is more burdensome for international students, who face both structural barriers to securing central housing and heightened sensitivity to the academic consequences of distance. This indicates that accessibility inequities in Koblenz are less about raw geography and more about the interaction of spatial distribution, housing allocation, and student status.

5.3. RQ3: Student Satisfaction

Satisfaction is a subjective yet essential dimension of student housing equity. Unlike affordability or accessibility, it reflects the lived experience of students, encompassing perceptions of housing quality, comfort, and fairness. To capture this dimension comprehensively, both structured (Likert-scale ratings) and unstructured (open-ended responses) survey data were analysed using natural language processing, sentiment analysis, and machine learning methods. This multi-layered approach enables the identification of systematic disparities in housing satisfaction across domestic and international students.

5.3.1. Sentiment Analysis of Open-Ended Responses

To capture students’ subjective housing experiences, open-ended survey responses were analysed using the `nlptown/bert-base-multilingual-uncased-sentiment` model. This multilingual model assigns a sentiment score from 1 (very negative) to 5 (very positive), ensuring both English and German responses were evaluated consistently. The analysis provides an additional layer of insight beyond structured survey items, highlighting how students describe their housing situations in their own words.

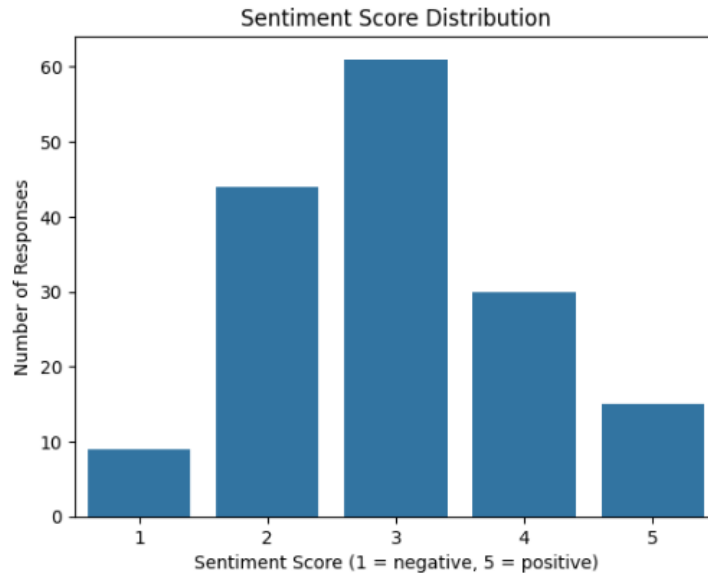


Figure 5.12.: Overall Sentiment Distribution

Overall Sentiment Distribution

The distribution of sentiment scores across all responses is shown in Figure 5.12. The most common category was score 3 (neutral), followed by moderately negative scores (2) and moderately positive scores (4). Extreme responses (1 = very negative, 5 = very positive) were relatively less frequent. This pattern indicates that most students described their housing experiences in balanced or mixed terms rather than in strongly positive or negative ways.

At the same time, the distribution highlights that a non-negligible share of students reported dissatisfaction, as reflected in the proportion of responses with scores of 1 or 2. These lower scores suggest that for many students, housing presents notable challenges. The nature of these challenges is explored further through topic modelling of the open-ended responses, which provides insight into the specific concerns driving more negative sentiment.

Average Sentiment Score by Student Status

When disaggregated by student status as shown in Figure 5.13, clear differences emerge. Domestic students reported a higher average sentiment score **3.14** (SD = 1.13, 95% CI [2.87, 3.42], N = 69), reflecting experiences that trend toward neutral-to-positive. In contrast, international students averaged **2.84** (SD = 0.93, 95% CI [2.65, 3.04], N = 89), indicating a leaning toward dissatisfaction. Although the overall distributions are centred near the neutral category, the lower mean and tighter confidence interval for international students demonstrate a consistent shift toward

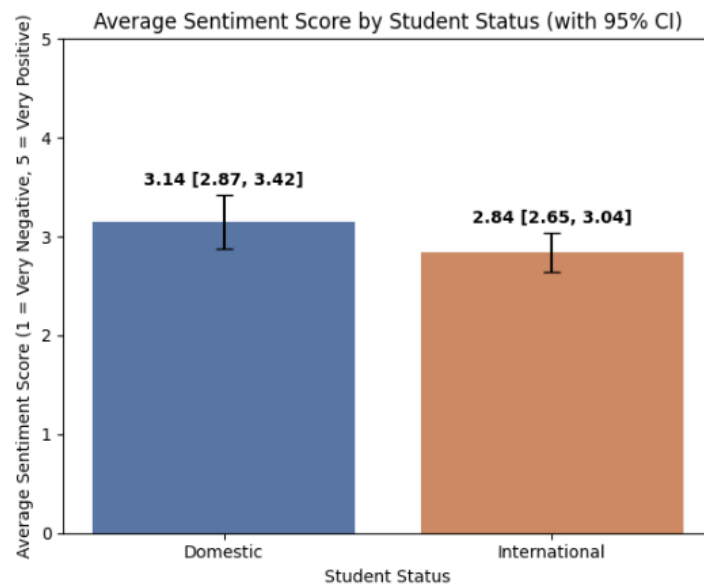


Figure 5.13.: Average Sentiment Score by Student Status

less positive sentiment. The limited overlap between the confidence intervals further highlights the robustness of this group difference.

This divergence provides an important foundation for the next stage of analysis, where the content of negative responses is examined in detail to identify the specific concerns underlying these differences.

5.3.2. Recurring Concerns in Open-Ended Responses

To complement the sentiment distribution analysis, the open-ended survey responses were further examined to uncover recurring themes in students' narratives. This was done in two steps: first, by extracting frequently used keywords from negative responses, and second, by applying topic modelling (BERTopic) to cluster responses into coherent themes.

Keyword Frequency

The most common words appearing in negative and neutral responses are shown in Figure 5.14. These include both German and English terms, reflecting the multi-lingual nature of the survey.

- **Cluster 0 keywords** such as Koblenz, find, student, wohnung (apartment), housing, accommodation, finden (to find), difficult point to the difficulty of securing affordable housing in Koblenz. They highlight both the competitive housing search and the barriers faced by international students.

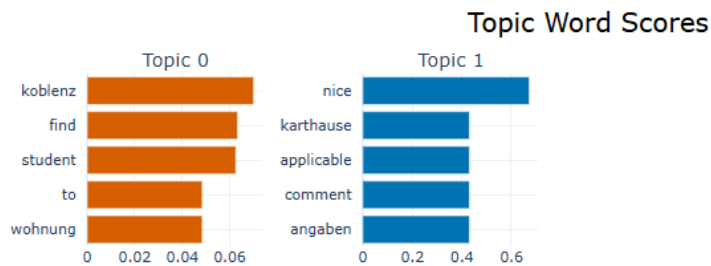


Figure 5.14.: Recurring Keywords

- **Cluster 1 keywords** such as nice, karthause (student neighbourhood), applicable, angaben (details), groß (large), esg (student dorm), keine (none), ok, ist reflect more neutral or positive experiences. These terms suggest that while many students struggled, others expressed satisfaction with particular locations (e.g., Karthause) or dormitory arrangements (ESG).

Together, these keyword patterns show that while the dominant narrative focuses on affordability and access challenges, a smaller group of students described more neutral or even positive experiences.

Topic Modelling Results

To move beyond individual keywords, BERTopic was applied to cluster responses into coherent themes. Two dominant topics emerged:

- **Topic 0: Housing Search and Affordability (Count = 147 responses)** This was the largest cluster, dominated by words such as koblenz, find, wohnung, housing, difficult. Students described the stressful and competitive process of securing accommodation, particularly for international students. One noted:

"It's very difficult to find any accommodation in Koblenz, especially as an international student. After applying for 50+ rooms you get reply for only 5 and from that 4 say 'our vibe doesn't match', so we are left with hardly any options."

Others connected affordability directly to availability:

"I was not able to find something affordable, as landlords overcharge for (officially) shared apartments... Also there are way too few small and affordable apartments available in Koblenz."

- **Topic 1: Neutral or Positive Experiences (Count = 12 responses)** A smaller cluster grouped words such as nice, karthause, esg, groß, ok. These responses suggested that some students found housing that met their expectations, often in student neighbourhoods or dormitories. For example:

"Not applicable, ESG on Karthause is nice"

In addition to the issues highlighted in Topic 1, individual comments also raised housing quality and maintenance concerns. For example:

"For the last 2 years there is a fungus issue because of water leakage... I wrote to the housing company so many times but till now there is no update, and now it's affecting my health."

A smaller number of responses also described severe challenges in the safety and social environment of shared housing. One international student, for instance, recounted experiences of discrimination and unsafe living conditions, noting prolonged stress and lack of landlord support. While such cases were not common, they underscore the vulnerability of some students when institutional or managerial interventions are absent.

The keyword analysis, topic clusters, and illustrative comments show that housing dissatisfaction is rooted in recurring and concrete challenges. The largest group of responses pointed to the difficulty of finding affordable housing in Koblenz, particularly for international students navigating a competitive rental market. Other comments highlighted housing quality and maintenance issues, such as unresolved repairs and health hazards, while a small but important set of students described problems with the safety and social environment of shared living.

At the same time, a minority of responses reflected more neutral or positive impressions, often linked to specific neighbourhoods or to dormitory settings. This contrast suggests that student housing experiences are heterogeneous: while some students find arrangements that meet their needs, others—especially international students—face persistent barriers that shape their overall sense of dissatisfaction.

While open-ended responses provide rich insights into students' lived experiences, they remain qualitative and uneven in detail. To complement these narratives, the survey also included a battery of Likert-scale questions covering different dimensions of satisfaction, such as affordability, room size, facilities, and overall experience. Analysing these structured responses allows for a more systematic comparison between domestic and international students and provides statistical evidence to confirm whether the patterns observed in the open-ended responses—particularly around affordability, housing quality, and expectations—are consistently reflected across the broader sample.

5.3.3. Structured Satisfaction Indicators (Likert-Scale)

In addition to open-ended responses, the survey included 17 Likert-scale items covering diverse aspects of the student housing experience, from affordability and room size to safety, facilities, and overall satisfaction. Independent-samples t-tests were conducted to compare domestic and international students on each indicator, with Cohen's *d* used to assess effect sizes. Table 5.3 summarises the results.

The results reveal clear disparities across several core dimensions of satisfaction. Domestic students reported higher satisfaction in areas such as access to accommodation, room size, affordability, overall experience, and whether housing met their

Question	t-stat	p-value	Significant?	Cohen's d	Effect Size	Direction
Ease of Accommodation	6.950	0.0000	Yes	1.132	Large	Domestic > International
Transport	-4.004	0.0001	Yes	-0.622	Medium	International > Domestic
Room size	3.791	0.0002	Yes	0.602	Medium	Domestic > International
Overall experience	3.874	0.0002	Yes	0.611	Medium	Domestic > International
Expectations	3.146	0.0020	Yes	0.510	Medium	Domestic > International
Noise level	-3.128	0.0021	Yes	-0.497	Small	International > Domestic
Impact	-2.422	0.0167	Yes	-0.391	Small	International > Domestic
Affordability	2.303	0.0227	Yes	0.368	Small	Domestic > International
Safe Neighbourhood	-1.458	0.1470	No	-0.237	Small	International > Domestic
Proximity to university	1.217	0.2257	No	0.195	Negligible	Domestic > International
Proximity to amenities	1.026	0.3067	No	0.162	Negligible	Domestic > International
Responsiveness	0.924	0.3573	No	0.150	Negligible	Domestic > International
Cleanliness	-0.906	0.3665	No	-0.144	Negligible	International > Domestic
Facilities	0.873	0.3843	No	0.140	Negligible	Domestic > International
Building condition	-0.388	0.6983	No	-0.061	Negligible	International > Domestic
Proximity to bus stop	0.148	0.8823	No	0.024	Negligible	Domestic > International
Room-mate interaction	0.079	0.9371	No	0.013	Negligible	Domestic > International

Table 5.3.: Independent-samples t-test results for Likert-scale satisfaction questions.

expectations. These findings suggest that domestic students not only find it easier to secure housing but are also more likely to perceive it as adequate and aligned with their needs.

By contrast, international students are disadvantaged in terms of affordability and housing quality, they report somewhat more positive experiences with transport accessibility and, paradoxically, noise and housing impact. This could reflect differences in expectations or reliance on public services among international students.

For the remaining indicators — including safety of neighbourhood, proximity to campus and amenities, responsiveness of landlords, cleanliness, facilities, building condition, proximity to bus stops, and roommate interaction — no statistically significant differences were found. Effect sizes for these items were negligible, suggesting broadly similar experiences across groups in these dimensions.

Overall, the Likert results show that satisfaction disparities are concentrated in critical areas linked to affordability, access, and housing adequacy, rather than being evenly spread across all aspects of student housing. These structured findings reinforce the themes already observed in the open-ended responses, where students—particularly internationals—frequently expressed concerns about difficulty in securing affordable housing, smaller room sizes, and unmet expectations. The alignment of qualitative and quantitative evidence strengthens the conclusion that international students face systemic disadvantages in the Koblenz housing market.

While the Likert-scale comparisons highlight where domestic and international students differ in their reported satisfaction, these results remain descriptive. They show which dimensions diverge between groups but do not explain which factors are most influential in shaping dissatisfaction overall. To address this, predictive models were employed.

5.3.4. Predictive Modelling of Dissatisfaction

To move beyond descriptive comparisons, predictive models were used to identify the structural drivers of housing dissatisfaction. Dissatisfaction was operationalised as a binary outcome, coded as 1 if students rated their overall housing experience as 3 or below, and 0 otherwise. Two complementary models were applied: logistic regression and random forest classification.

Logistic Regression Results

Class	Precision	Recall	F1-score
Satisfied (0)	0.91	0.91	0.91
Dissatisfied (1)	0.80	0.80	0.80
Accuracy	0.88		

Table 5.4.: Classification performance of logistic regression model for predicting dissatisfaction.

The logistic regression model achieved an overall accuracy of 88%, with balanced performance across both satisfied and dissatisfied classes (Table 5.4). Precision, recall, and F1-scores were strong for both groups, indicating that the model reliably distinguishes between satisfied and dissatisfied students.

The coefficients provide further insight into the drivers of dissatisfaction. The most influential negative predictors were expectations (-0.68), facilities (-0.53), affordability (-0.46), and room size (-0.46), followed by responsiveness (-0.27). These values indicate that lower ratings in these areas significantly increase the likelihood of a student being dissatisfied. Other factors such as neighbourhood safety and proximity to bus stops showed positive coefficients, but their magnitudes were smaller and less meaningful in the overall model. These results showcase that dissatisfaction is primarily driven by structural housing conditions — when expectations are unmet, facilities are inadequate, or rooms are small and unaffordable, students are much more likely to report negative overall experiences.

These findings align closely with the patterns observed in the Likert-scale t-tests. While domestic students generally scored higher on room size, expectations, and facilities, international students rated these aspects lower, placing them at greater risk of dissatisfaction. Responsiveness of landlords also emerged as an important factor, reflecting student frustration with slow or inadequate responses to maintenance requests and other housing problems.

Random Forest Feature Importance

The random forest model confirmed and extended these insights by ranking the relative importance of each predictor in classifying dissatisfied students. The feature importance plot (Figure 5.15) highlights the same dimensions as central drivers—facilities,

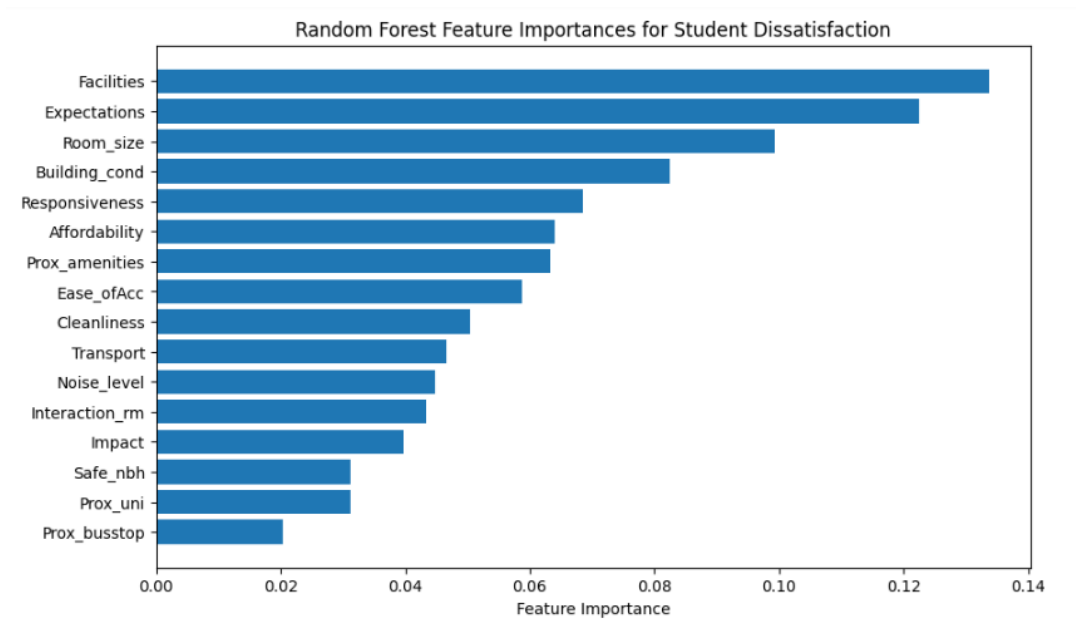


Figure 5.15.: Random Forest model

expectations, room size, and responsiveness—with affordability also gaining prominence in the non-linear model.

Unlike logistic regression, which quantifies direction and strength of association, the random forest provides a robust ranking of variables even in the presence of complex interactions. The prominence of affordability here reinforces its salience in both sentiment analysis and open-ended responses, even if its effect was smaller in the Likert comparisons.

The two models demonstrate that dissatisfaction is not evenly distributed across all aspects of student housing but is concentrated in a small set of critical dimensions. Facilities, expectations, and room size consistently emerge as the strongest predictors, with affordability and landlord responsiveness also playing meaningful roles. These findings provide strong evidence of structural inequities: international students, who systematically score lower on these high-importance dimensions, are disproportionately represented among dissatisfied respondents.

The convergence of logistic regression and random forest results strengthens confidence in these conclusions. The models show that dissatisfaction is driven less by location or proximity measures and more by the quality, adequacy, and affordability of housing—areas where international students face structural disadvantages.

5.3.5. Unsupervised Clustering of Housing Satisfaction

To complement the regression and classification approaches, an unsupervised machine learning method was applied to the Likert-scale satisfaction responses. K-

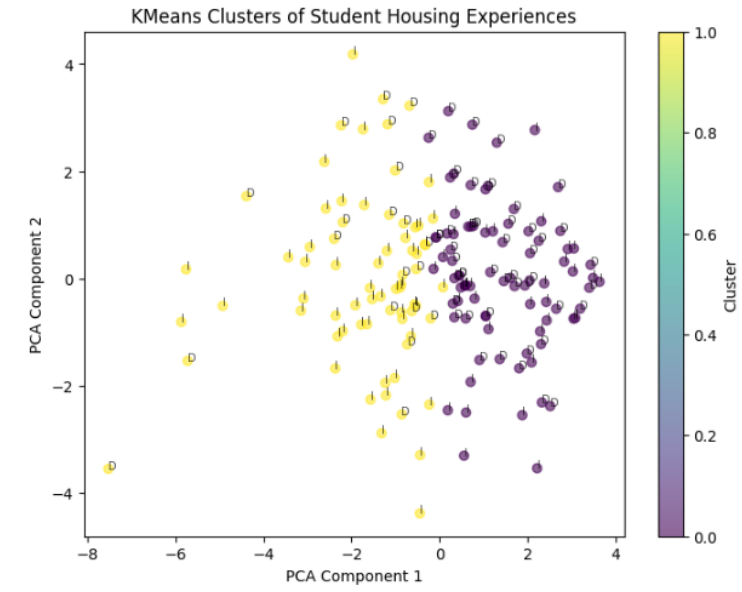


Figure 5.16.: K means cluster representation

means clustering was selected to identify naturally occurring groups of students based on their overall patterns of ratings, without imposing predefined categories such as domestic vs. international. This approach allows for a more exploratory analysis of whether distinct satisfaction profiles emerge from the data.

The Likert-scale variables were first standardized to ensure comparability across indicators with different ranges. The K-means algorithm was then applied, and the optimal number of clusters was determined using the silhouette score, which balances within-cluster cohesion and between-cluster separation. The analysis indicated that a two-cluster solution best fit the data (in Figure 5.16, capturing the heterogeneity of student housing experiences while maintaining interpretability).

- **Cluster 0: Struggling Students** Students in this group reported systematically lower ratings across most dimensions. Their scores for room size, facilities, responsiveness, and expectations were particularly low, with overall satisfaction averaging just above 3 on the 5-point scale. This cluster therefore reflects a group of students facing significant housing challenges, characterised by inadequate conditions, unmet expectations, and limited satisfaction. Importantly, international students were heavily overrepresented in this cluster, making up approximately 72% of its membership.
- **Cluster 1: Well-Housed Students** By contrast, the second cluster displayed consistently higher ratings across all indicators. These students rated facilities, room size, and cleanliness more positively, and their overall satisfaction averaged around 4.4. This group reflects those whose housing conditions met or

exceeded expectations, providing comfort and stability for their academic life. Domestic students were more prevalent in this cluster, comprising roughly 57% of its members.

The clustering results highlight that student housing experiences in Koblenz are not uniform but polarized into two broad profiles: those who are struggling and those who are well-housed. Crucially, the over representation of international students in the low-satisfaction cluster demonstrates that disparities are not only evident in group averages but also embedded in the underlying structure of the data. This unsupervised approach therefore reinforces the conclusion that international students systematically face less favourable housing conditions.

The advantage of clustering lies in its independence from predefined categories: even without specifying “domestic” and “international,” the algorithm grouped students in a way that naturally revealed inequities between these groups. This strengthens the argument that housing satisfaction disparities are robust and consistent across multiple analytical strategies.

5.3.6. Composite Satisfaction Index

To integrate insights from both structured survey items and unstructured open-ended responses, a composite satisfaction index was constructed. The measure combined standardized Likert-scale ratings with sentiment scores from the nlptown multilingual BERT model, producing a single index that captures both the quantitative and qualitative dimensions of student housing experiences.

The composite scores reveal a consistent gap between domestic and international students. Domestic students recorded a mean index value of 3.43, while international students averaged 3.05. Although both groups cluster around the midpoint of the scale, the difference of 0.38 points is statistically significant ($t = 3.46$, $p < 0.001$). Figure 5.17 visualises this difference, showing the mean composite scores with 95% confidence intervals. The minimal overlap between the intervals underscores the robustness of the disparity. In other words, even when combining structured ratings with sentiment derived from open-text comments, international students continue to express systematically lower satisfaction with their housing situation.

The composite index provides strong converging evidence that housing satisfaction inequities are both quantitative and qualitative in nature. Domestic students are more likely to evaluate their housing positively in structured ratings, and their narratives also lean more neutral-to-positive. International students, by contrast, are disadvantaged in both domains, with lower Likert ratings and more negative sentiment in open-ended responses.

This combined approach highlights the value of integrating methods: while Likert items capture consistency across the sample, sentiment analysis reveals the emotional undertone of students’ lived experiences. Together, they paint a comprehensive picture of housing satisfaction in Koblenz, confirming that international

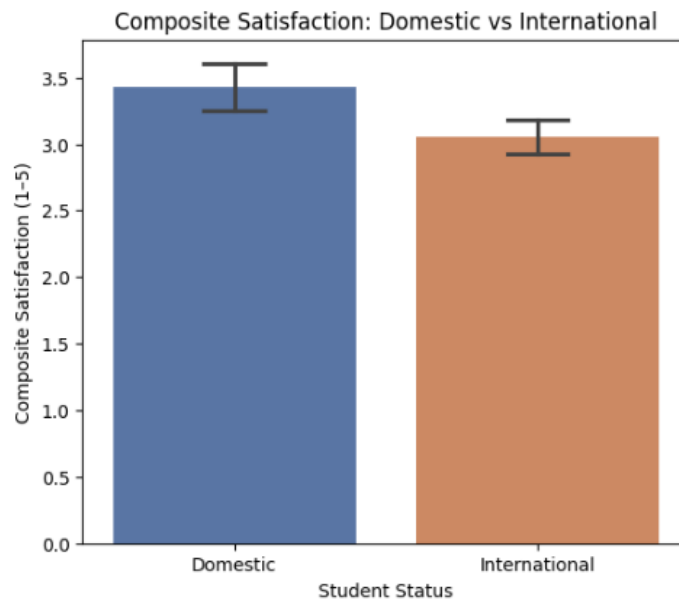


Figure 5.17.: Composite sentiment score by student status

students are disproportionately burdened by affordability pressures, smaller living spaces, and unmet housing expectations.

Across multiple methods, this study reveals clear disparities in housing satisfaction that disadvantage international students. Sentiment analysis and open-ended responses pointed to recurring frustrations with affordability, small rooms, and poor facilities, while structured Likert-scale comparisons confirmed that domestic students consistently rated these dimensions more positively. Predictive models reinforced these findings, with expectations, facilities, affordability, and room size emerging as the strongest predictors of dissatisfaction. Unsupervised clustering further demonstrated that international students are disproportionately concentrated in the low-satisfaction group, and the composite satisfaction index confirmed that these inequities persist even when structured and unstructured measures are combined. Taken together, the evidence shows that both demographic bias (linked to student status) and structural bias (linked to housing quality and affordability) shape the lived housing experiences of international students in Koblenz.

5.4. RQ4: Equity in Student Housing

To assess overall fairness in student housing, a composite equity index was constructed, combining affordability (inverse of cost per m²), accessibility (inverse of distance to campus and reported travel time), and housing satisfaction. This standardized index captures multiple dimensions of housing experience simultaneously, allowing systematic inequities between domestic and international students to be

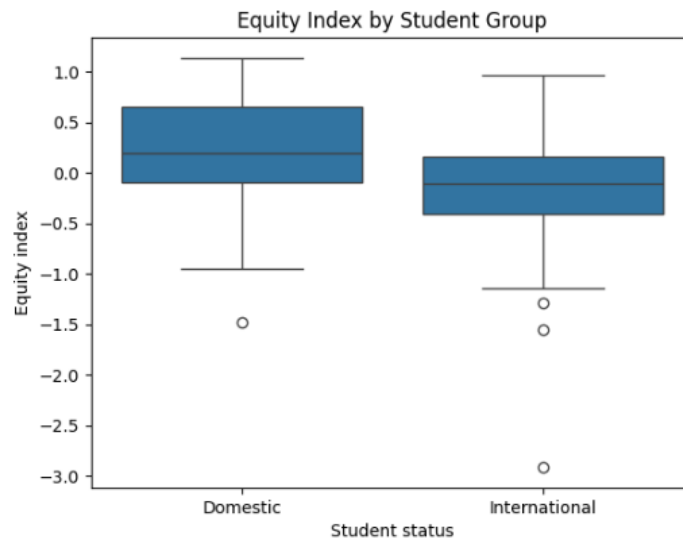


Figure 5.18.: Equity Index by student status

evaluated.

5.4.1. Equity Index

The distribution of the composite equity index highlights systematic differences between domestic and international students in terms of housing outcomes. On average, domestic students achieve a mean equity score of 0.215, whereas international students record a mean of -0.165 . This difference of approximately 0.38 standard deviations indicates a moderate but meaningful gap in housing equity between the two groups. In practical terms, domestic students tend to benefit from more favourable conditions across affordability, accessibility, and satisfaction, while international students are consistently disadvantaged. This overall disparity reflects the combined influence of structural and locational factors and already signals that equity in student housing is unevenly distributed across groups. Figure 5.18 (Box-plot) illustrates this divergence clearly. Domestic students not only achieve higher median equity index values but also exhibit a tighter distribution clustered around positive scores, suggesting more stable and favorable housing outcomes. In contrast, international students show a lower median and a wider spread, with several cases extending into negative values. This skew highlights that a subset of international students experience particularly severe disadvantages, reinforcing the notion of structural inequity rather than random variation. The differences become even more apparent when examining Figure 5.19 (Radar Chart), which compares domestic and international students across the three constituent dimensions of the equity index. Domestic students score consistently higher across affordability, accessibility, and satisfaction, with the gaps most pronounced in affordability and satisfaction.

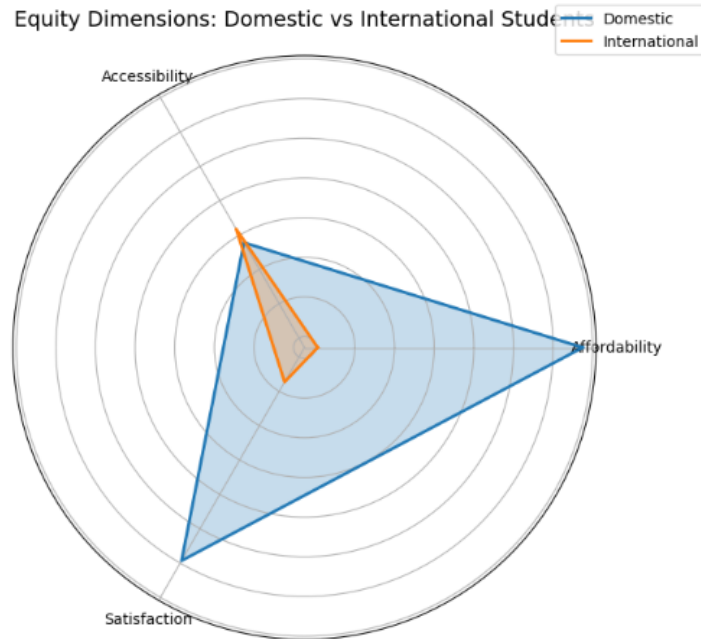


Figure 5.19.: Equity Index radar chart by student status

The affordability disparity suggests that international students are more burdened by high rents relative to room size, while the satisfaction gap underscores a subjective dimension of disadvantage that extends beyond measurable costs and distances. Accessibility shows a narrower difference, yet international students still lag behind, pointing to constraints in commuting efficiency and proximity. Together, the radar chart confirms that inequity is not isolated to a single dimension but emerges across all facets of the housing experience.

Figure 5.20 (Correlation Heatmap) provides insight into how these dimensions interact. The low correlations between affordability, accessibility, and satisfaction indicate that each captures a distinct aspect of housing equity. For instance, higher satisfaction is not necessarily tied to lower costs or shorter distances, and students who secure affordable housing may still report dissatisfaction due to other qualitative factors. This independence validates the construction of a composite equity index: combining dimensions is necessary to reflect the multifaceted nature of housing experiences rather than reducing equity to a single indicator.

5.4.2. Regression Analysis: Geographically Weighted Regression (GWR)

To assess the determinants of housing equity and capture how inequities vary across space, a Geographically Weighted Regression (GWR) model was estimated. Unlike

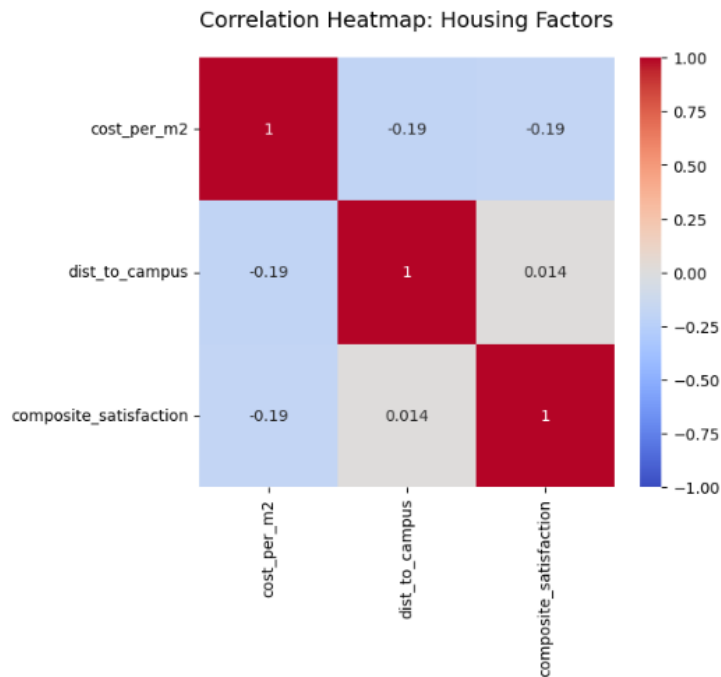


Figure 5.20.: Correlation Heatmaps of Housing Factors

standard global regressions, which assume coefficients are constant across all observations, GWR allows relationships to vary locally depending on geographic context. This approach is particularly relevant for student housing, where affordability, accessibility, and satisfaction conditions may differ significantly between central and peripheral neighborhoods. **Equity Model:**

$$equity_index \sim student_status + room_size + rent + dist_to_campus + travel_time + region_binary$$

The global regression (OLS baseline) explained about 28% of variation in the equity index (Adj. $R^2 = 0.278$). Results showed that being an international student is associated with significantly lower equity scores ($\beta = -0.236$, $p = 0.008$), even after accounting for room size, rent, travel time, and regional context. Structural housing factors such as room size ($\beta = +0.014$, $p < 0.001$) and rent ($\beta = -0.001$, $p < 0.001$) emerged as consistent predictors, while longer travel times ($\beta = -0.008$, $p = 0.011$) also reduced equity. In contrast, network distance to campus and regional binary classification were not significant predictors at the global level.

Once spatial heterogeneity was introduced through GWR, model performance improved substantially: Adjusted R^2 increased to 0.583. This indicates that inequities are not uniformly distributed across Koblenz, but instead vary according to location. For example, the penalty associated with being an international student ranged from -0.377 in certain central, high-demand areas (indicating strong disadvantage) to $+0.463$ in a few peripheral neighborhoods where the effect was neutral

or slightly positive. Similarly, the effect of travel time fluctuated across the city: in some areas, longer commutes strongly reduced equity, while in others they were less detrimental, suggesting differences in transport accessibility and connectivity. Region effects also varied locally, with certain regions amplifying inequity and others showing negligible influence.

Across space, structural predictors remained stable. Larger room sizes consistently improved equity (local β range = 0.011 to 0.027), while higher rents universally reduced it (local $\beta = -0.001$). These stable effects indicate that, regardless of location, affordability and housing quality are fundamental determinants of equity, while accessibility and international status interact more dynamically with geographic context.

The GWR analysis confirms that inequities in student housing are systematic yet spatially uneven. International students consistently record lower equity scores than their domestic peers, and this disadvantage is most pronounced in central, high-demand neighborhoods while diminishing in more peripheral areas. The composite equity index—integrating affordability, accessibility, and satisfaction—shows that inequity is not confined to higher rents or longer commutes but also extends to students' lived experiences of housing quality. Structural factors such as room size and rent exert stable effects everywhere, whereas accessibility (travel time) and locational context (region) display more spatial variability, reflecting differences in transport infrastructure and neighborhood characteristics. Together, these findings demonstrate that international students face both material and experiential disadvantages in their search for adequate accommodation, confirming that student housing inequity is multidimensional and geographically contingent.

In aggregate, the findings across all four research questions reveal consistent patterns of bias in student housing outcomes, particularly disadvantaging international students.

- **RQ1 (Affordability):** Results show a demographic bias, with international students paying higher €/m², a structural bias, as smaller rooms are disproportionately more expensive, but no evidence of geographic disparities between central and peripheral areas.
- **RQ2 (Accessibility):** While objective measures of distance and travel time show no gaps, a perceptual bias emerges as international students place greater importance on proximity for attendance, performance, and extracurricular participation.
- **RQ3 (Satisfaction):** Evidence shows both demographic bias (linked to student status) and structural bias (linked to affordability and housing quality), resulting in lower satisfaction levels and more negative experiences among international students.
- **RQ4 (Equity):** The composite equity index and GWR confirm a multidimensional and spatially uneven disadvantage, with inequities most pronounced

for international students in central, high-demand neighbourhoods, though weaker or neutral in peripheral areas.

In sum, these findings demonstrate that student housing inequities in Koblenz are systemic: while affordability and satisfaction reveal clear demographic and structural biases, accessibility inequities emerge more in perceptions than in raw measures, and overall equity reflects a multidimensional and spatially contingent disadvantage for international students. Taken together, the results confirm that the possibility of bias raised in earlier chapters is supported by empirical evidence. These insights form the basis for the concluding chapter, which reflects on their broader implications, acknowledges limitations, and outlines directions for future research and policy.

6. Conclusions

The conclusion of this study begins by returning to the problems most frequently reported by students themselves. As shown in Figure 6.1, high rent, maintenance difficulties, and other recurring issues dominate the student housing experience. These challenges highlight the seriousness of inequities in housing, demonstrating that the difficulties students face are not individual exceptions but collective patterns requiring systemic change.

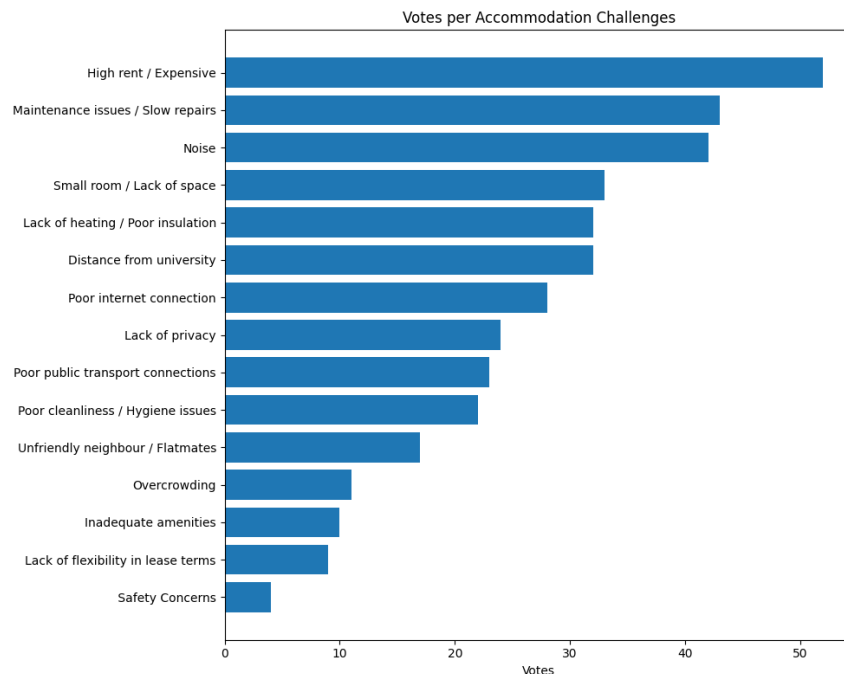


Figure 6.1.: Key challenges reported in student accommodation.

These concerns provide the context for interpreting the empirical results of this research. By analyzing affordability, accessibility, satisfaction, and equity in detail, the study demonstrates how the challenges identified by students align with systematic disparities across demographic, structural, and spatial dimensions. While previous chapters highlighted the possibility that bias might influence housing outcomes, the analyses presented here uncovered clear evidence that inequities are not randomly distributed but systematically disadvantage international students. By combining survey data with geospatial methods and statistical modelling, the study provides

a comprehensive assessment of student housing outcomes and demonstrates how inequities operate across structural, demographic, and geographic dimensions.

The analysis of affordability (RQ1) revealed a clear demographic bias: international students face systematically higher rents per square meter, while a structural bias emerges as smaller rooms are disproportionately expensive. Accessibility (RQ2) showed no major disparities in raw distance or travel time, but international students reported feeling more strongly constrained by location, underscoring how reliance on public transport and limited commuting flexibility shape their experiences. Housing satisfaction (RQ3) further confirmed demographic differences, as international students consistently expressed lower satisfaction in both survey ratings and open-ended responses, citing recurring concerns about cost, comfort, and integration. Finally, the composite equity index and Geographically Weighted Regression (RQ4) demonstrated that these inequities are multidimensional and spatially uneven, with disadvantages most pronounced in central, high-demand areas of the city. Taken together, the results highlight that both structural factors (such as room size and affordability) and demographic status (international versus domestic) jointly drive unequal housing outcomes in Koblenz.

Together, these findings demonstrate that the possibility of bias raised at the outset of the study is supported by empirical evidence. The research uncovered patterns of disadvantage that align with both material housing conditions (rent, room size, location) and subjective experiences (satisfaction, perceptions of accessibility). These results contribute to the literature by showing that inequities in student housing are systemic rather than incidental, and that they are shaped by both measurable structures and lived experiences.

The study carries important policy implications. Expanding affordable housing options, improving transport connectivity, and ensuring greater transparency in housing allocation are critical steps toward addressing the inequities identified here. More broadly, the findings highlight the importance of monitoring student housing not only in terms of costs and distances but also in terms of perceptions, satisfaction, and fairness.

The following sections discuss the limitations of this research and outline opportunities for future work, emphasizing the need for larger samples, longitudinal perspectives, and the inclusion of institutional actors in order to build a more comprehensive understanding of student housing equity.

6.1. Limitations

This study is not without limitations. First, the sample size (approximately 150–160 responses) limits the generalizability of the findings. While the dataset was sufficient for the statistical and spatial analyses conducted, it may not fully capture the diversity of student housing experiences across all demographic subgroups. Smaller subgroups, such as specific nationalities, gender intersections, or levels of study, could not be examined in depth due to insufficient representation. A larger sam-

ple would allow for finer-grained comparisons and greater confidence in subgroup-level inferences.

Second, the study design is cross-sectional, capturing housing conditions and perceptions at a single point in time. As a result, the analyses cannot establish causality or track how inequities evolve over the course of students' academic trajectories. Housing experiences are likely to change across semesters—particularly as students transition from initial accommodation to more permanent housing arrangements—and longitudinal data would provide deeper insights into the dynamics of inequity over time.

Third, the construction of the equity index involves methodological choices that may influence the results. In this study, affordability, accessibility, and satisfaction were weighted equally to reflect their collective importance in shaping housing equity. However, alternative weighting schemes, informed by theoretical frameworks or empirical evidence, could shift the balance of findings. For instance, affordability may play a more dominant role in shaping overall equity in markets with very high rents, while accessibility could matter more in cities with limited public transport infrastructure. Future work could test the sensitivity of results to different weighting approaches.

Fourth, the measures of accessibility rely on reported travel times and calculated distances, both of which have inherent limitations. Travel times were self-reported and may be subject to recall bias, while distance measures did not fully capture qualitative aspects of accessibility such as safety, convenience of routes, or reliability of transport services. These experiential aspects of accessibility may play an equally important role in shaping students' perceptions of equity.

Finally, although this thesis integrates quantitative survey data with geospatial analysis, it does not include perspectives from institutional stakeholders such as housing providers, landlords, or university administrators. These actors play a critical role in shaping housing allocation and rental market dynamics, and their absence means that the analysis reflects primarily the demand side of the housing equation. Incorporating supply-side perspectives could provide a more comprehensive view of the structural drivers of inequity.

6.2. Future Work

Future research should expand this framework beyond Koblenz to other German and European university towns, enabling comparative analysis of student housing inequities in different urban and institutional contexts. Larger cities with more competitive housing markets may exhibit stronger affordability disparities, while smaller university towns may reveal different accessibility dynamics. Comparative work would also allow researchers to test the generalizability of the equity index across contexts, shedding light on whether the patterns observed in Koblenz reflect broader trends or are locally specific.

Longitudinal studies would provide an opportunity to examine how inequities

evolve over time. Many international students face particular challenges when they first arrive, such as limited knowledge of the rental market or barriers in navigating housing systems, but may gradually secure better housing as they integrate socially and academically. A longitudinal design could capture these dynamics and distinguish between short-term transitional disadvantages and long-term structural inequities that persist throughout a degree program.

Intersectional approaches also deserve attention. While this study focused primarily on differences between domestic and international students, inequities are rarely uniform within these categories. Gender, income, and migration background may intersect with student status in complex ways, producing layered disadvantages. For example, female international students may place even greater emphasis on safety and accessibility. Future studies should therefore disaggregate groups to explore within-group variation and highlight where inequities are compounded.

An avenue for future research lies in modelling students' intention to attend more frequently if housed closer to campus. Using variables such as current attendance days, distance, and transport mode, predictive models (e.g., logistic regression) could estimate the likelihood of increased attendance, providing deeper insights into the behavioural impacts of housing location.

Integrating institutional and market-level data would further strengthen the explanatory power of this model. University housing allocations, private rental market statistics, and policy frameworks could help clarify whether inequities are driven primarily by demand-side disadvantages (e.g., income, language barriers, student status) or supply-side restrictions (e.g., limited student dorm capacity, discriminatory rental practices). Combining survey-based equity indices with administrative data could also inform more policy-relevant recommendations, highlighting where interventions such as subsidized housing, targeted support for international students, or transport investments would be most effective.

Future research that extends this framework across different settings and perspectives will advance a more holistic understanding of student housing inequities, helping to connect structural analysis with lived experiences and guiding more targeted policy responses.

In conclusion, this thesis demonstrates that student housing inequities in Koblenz are systemic, multidimensional, and spatially uneven, with international students persistently disadvantaged across affordability, accessibility, and satisfaction. Addressing these inequities requires not only structural reforms to reduce costs and improve accessibility but also a recognition of the lived experiences that shape how students perceive and engage with their housing. Only by combining structural, spatial, and experiential perspectives can more equitable student housing policies be developed and implemented.

Bibliography

- [1] I. A. Addo. Assessing residential satisfaction among low-income households in multi-habited dwellings in selected low-income communities in accra. *Urban Studies*, 2016.
- [2] A. Agresti. *Statistical Methods for the Social Sciences*. Pearson, 2018.
- [3] Amber Student. Importance of diversity and inclusion in us student housing. Available at: <https://amberstudent.com/news/post/importance-of-diversity-and-inclusion-in-us-student-housing>, 2023. Accessed: 22 September 2025.
- [4] D. Amole. Residential satisfaction in students housing. *Journal of environmental psychology*, 2009.
- [5] L. Barreto. In the shadows of higher education: Housing insecurities among college students. Master's thesis, Master of Arts in Sociology, Northeastern University, 2023.
- [6] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, environment and urban systems*, 2017.
- [7] L. Breiman. Random forests. *Machine learning*, 2001.
- [8] Campus Advantage. Rethinking student housing operations: Fostering communities beyond walls. <https://campusadv.com/rethinking-student-housing-operations-fostering-communities-beyond-walls/>, 2024. Accessed: 22 September 2025.
- [9] V. Chang, L. Liu, Q. Xu, T. Li, and C.-H. Hsu. An improved model for sentiment analysis on luxury hotel review. *Expert Systems*, 2023.
- [10] D. Clapham. *The meaning of housing: A pathways approach*. policy Press, 2005.
- [11] J. Cohen. *Statistical power analysis for the behavioral sciences*. routledge, 2013.
- [12] F. L. Collins. Researching mobility and emplacement: Examining transience and transnationality in international student lives. *Area*, 2012.
- [13] E. Commission, D.-G. for Economic, F. Affairs, C. Frayne, A. Szczypińska, B. Vašíček, and S. Zeugner. *Housing market developments in the euro area – Focus on housing affordability*. Publications Office of the European Union, 2022.

- [14] J. W. Creswell and V. L. P. Clark. *Designing and conducting mixed methods research*. Sage publications, 2017.
- [15] A. Delbosc and G. Currie. Transport problems that matter—social and psychological links to transport disadvantage. *Journal of Transport Geography*, 2011.
- [16] Deutsches Studentenwerk. 21. sozialerhebung des deutschen studentenwerks: Wirtschaftliche und soziale lage der studierenden in deutschland 2021, 2022.
- [17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019.
- [18] K. Eagan, E. B. Stolzenberg, J. J. Ramirez, M. C. Aragon, M. R. Suchard, and S. Hurtado. The american freshman: National norms fall 2014. *Los Angeles: Higher Education Research Institute, UCLA*, 2014.
- [19] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, 1996.
- [20] G. W. Evans, N. M. Wells, and A. Moch. Housing and mental health: a review of the evidence and a methodological and conceptual critique. *Journal of social issues*, 2003.
- [21] Findence. Accommodation challenges faced by students in germany. <https://findence.com/accommodation-challenges-faced-by-students-in-germany/>, 2025. Accessed: 22 September 2025.
- [22] A. S. Fotheringham, C. Brunson, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, 2002.
- [23] M. Garmendia, J. M. Coronado, and J. M. Ureña. University students sharing flats: When studentification becomes vertical. *Urban Studies*, 2012.
- [24] Y. Ge, A. Collins, and N. Sipe. Mining public opinions on transport projects using twitter data: Sentiment analysis and topic modeling. *Journal of Transport Geography*, 64:373–384, 2017.
- [25] GeeksforGeeks. Dbscan clustering in ml. <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>, 2023. Accessed: 22 September 2025.
- [26] GeeksforGeeks. K-means clustering algorithm in machine learning. <https://www.geeksforgeeks.org/machine-learning/ml-k-means-algorithm/>, 2023. Accessed: 22 September 2025.

- [27] K. T. Geurs and B. Van Wee. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 2004.
- [28] M. Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.
- [29] J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares. Transit accessibility and social inequities: A multidimensional approach. *Journal of Transport Geography*, 2011.
- [30] W. G. Hansen. How accessibility shapes land use. *Journal of the American Institute of planners*, 1959.
- [31] S. Heath and E. Cleaver. *Young, free and single?: Twenty-somethings and household change*. Springer, 2003.
- [32] P. Hoffmeyer-Zlotnik, J. Grote, G. N. C. P. for the European Migration Network (EMN), F. O. for Migration, and Refugees. Attracting and retaining international students in germany. Working paper 85, Federal Office for Migration and Refugees, 2019.
- [33] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant. *Applied logistic regression*. John Wiley & Sons, 2013.
- [34] P. Hubbard. Geographies of studentification and purpose-built student accommodation: leading separate lives? *Environment and planning A*, 2009.
- [35] J. D. Hulchanski. The concept of housing affordability: Six contemporary uses of the housing expenditure-to-income ratio. *Housing studies*, 1995.
- [36] S. Kaya and F. Erkip. Satisfaction in a dormitory building: The effects of floor height on the perception of room size and crowding. *Journal of Environmental Psychology*, 2001.
- [37] B. Kelmendi. Higher education in germany key trends & statistics 2025 - study in germany for free. *Study in Germany for Free*, 2024.
- [38] T. Kenna. Studentification in ireland? analysing the impacts of students and student accommodation on cork city. *Irish Geography*, 2011.
- [39] S. Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 1982.
- [40] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1967.

- [41] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 1947.
- [42] J. N. V. O. Martijn B.W. Kobus and P. Rietveld. Student commute time, university presence and academic achievement. *Regional Science and Urban Economics*, 2015.
- [43] K. McKee, A. M. Soaita, and J. Hoolachan. “Generation Rent” and the Emotions of Private Renting: self-worth, status and insecurity amongst low-income renters. *Housing Studies*, 2019.
- [44] T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [45] Munich Eye. Increase in international students at german universities. <https://www.munich-eye.com/>, 2025. Accessed: 22 September 2025.
- [46] N. U. M. Najib, N. A. Yusof, and Z. Osman. Measuring satisfaction with student housing facilities. *American Journal of Engineering and Applied Sciences*, 2011.
- [47] NLPTown. Bert multilingual uncased sentiment model. <https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment>, 2020. Accessed: 22 September 2025.
- [48] G. Norman. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education*, 2010.
- [49] A. Octaviani, T. M. Oshan, W. Li, and A. S. Fotheringham. Mgwr: Multiscale geographically weighted regression for python. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019.
- [50] OECD and European Commission Joint Research Centre. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, 2008.
- [51] E. T. Pascarella and P. T. Terenzini. *How College Affects Students: A Third Decade of Research. Volume 2*. ERIC, 2005.
- [52] S. Pattabiraman. Logistic regression: An overview. <https://www.techicrumbs.com/logistic-regression-overview>, 2023. Accessed: 22 September 2025.
- [53] R. H. Pereira, T. Schwanen, and D. Banister. Distributive justice and equity in transportation. *Transport reviews*, 2017.

- [54] J. Pineda, J. Kercher, S. Falk, T. Thies, H. H. Yildirim, and J. Zimmermann. Accompanying international students in germany to academic success. Technical report, DAAD – German Academic Exchange Service, 2022. Accessed: 22 September 2025.
- [55] E. Play. Student housing rooms and apartments for rent in koblenz. *Erasmus Play*, 2025.
- [56] D. M. W. Powers. Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2011.
- [57] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2023.
- [58] S. Rafaeli, D. R. Raban, and G. Ravid. Mining textual responses in surveys: A natural language processing approach to open-ended questions. *International Journal of Social Research Methodology*, 2019.
- [59] P. A. Rogerson and A. S. Fotheringham. *The SAGE Handbook of Spatial Analysis*. Sage, 2008.
- [60] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 1987.
- [61] G. D. Ruxton. The unequal variance t-test is an underused alternative to student’s t-test and the mann–whitney u test. *Behavioral Ecology*, 2006.
- [62] R. Shields. Student mobility and access to higher education: Implications for equity. *Higher Education*, 78(5):761–777, 2019.
- [63] D. Smith. The politics of studentification and(un) balanced’urban populations: Lessons for gentrification and sustainable communities? *Urban Studies*, 2008.
- [64] R. A. Smith and N. G. Khawaja. A review of the acculturation experiences of international students. *International Journal of intercultural relations*, 2011.
- [65] L. Sotomayor. Lack of affordable housing affects student mental health, promotes social hierarchies. <https://www.yorku.ca/gradstudies/web7/2022/02/11/lack-of-affordable-housing-affects-student-mental-health-promotes-social-hierarchies/>, 2022. Accessed: 22 September 2025.
- [66] M. E. Stone. What is housing affordability? the case for the residual income approach. *Housing policy debate*, 2006.
- [67] Studierendenwerk Koblenz. Wohnanlagen. <https://www.studierendenwerk-koblenz.de/de/wohnen/wohnanlagen>, 2025. Accessed: 22 September 2025.

- [68] Study in Germany. Studierendenwohnanlage simmerner strasse in koblenz dormitory finder study in germany. <https://www.study-in-germany.de/en/plan-your-studies/preparation/finding-accommodation/dormitory-finder/detail/simmerner-strasse-koblenz>, 2025.
- [69] J. Thomsen. Home experiences in student housing: About institutional character and temporary homes. *Journal of Youth studies*, 2007.
- [70] Universität Koblenz. Numerous new international students at the university of koblenz. <https://www.uni-koblenz.de/de/international/numerous-new-international-students-at-the-university-of-koblenz>, 2023. Accessed: 22 September 2025.
- [71] Visit Koblenz. Cycling in koblenz and the region. <https://www.visit-koblenz.de/en/region/cycling>, 2023. Accessed: 22 September 2025.
- [72] R. Wang and M.-L. Tseng. Evaluation of international student satisfaction using fuzzy importance-performance analysis. *Procedia - Social and Behavioral Sciences*, 2011.
- [73] B. L. Welch. The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, 1947.
- [74] J. Wiers-Jenssen. Student satisfaction: Comparing expectations and experiences of international students. *Journal of Studies in International Education*, 2013.
- [75] P. Wilcox, S. Winn, and M. Fyvie-Gauld. It was nothing to do with the university, it was just the people: the role of social support in the first-year experience of higher education. *Studies in Higher Education*, 2005.
- [76] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in natural language processing: system demonstrations*, 2020.
- [77] B. Wong and T. Webb. Exploring international students’ housing experiences. *Journal of International Students*, 10(4):934–953, 2020.
- [78] J. M. Wooldridge. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 2016.

A. Enrollment Statistics by Semester

Semester Naming

- 20231 - Summer Semester 2023 (SS 2023)
- 20232 - Winter Semester 2023/24 (WS 2023/24)
- 20241 - Summer Semester 2024 (SS 2024)
- 20242 - Winter Semester 2024/25 (WS 2024/25)

Köpfe gesamt (ohne Beurl., ohne Exmatr.)				
	20231	20232	20241	20242
Gesamt	8830	9485	8917	9582
Master mit ZFUW	3702	4043	3864	4294

Table A.1.: Total Number of Students

Bildungsausländer*: Köpfe gesamt (ohne Beurl., ohne Exmatr.)				
	20231	20232	20241	20242
Gesamt	877	933	934	1067
Master mit ZFUW	786	840	833	959

Table A.2.: Total Number of Foreign Students

B. Survey Questionnaire

Koblenz student housing survey

Hello all, I'm Kavya Hanumanthu, an MSc student at the University of Koblenz. I am conducting this survey to analyse the state of student housing in Koblenz for both domestic and international students, I kindly request you to participate. Participation is entirely voluntary. This data will solely be used for my thesis purpose. The data will be removed and no other parties will be able to access it after the completion of my thesis. You will be making a substantial contribution to this study's success. It will take about 5 -7 minutes to finish the survey. I appreciate your interest in taking this survey.

Demographic Information

The following items capture demographic background information of participants.

1. Email address
2. What is your current level of study?
 - Bachelor's Degree
 - Master's Degree
 - Erasmus student
3. Which semester are you in? (Specify the semester in number)
4. Gender
 - Female
 - Male
 - Non-binary
 - Other
5. Age
 - Below 18
 - 18 - 21

- 22 - 25
- 26 - 29
- 30 - 34
- 35 and Above

6. What is your student status?

- Domestic
- EU (non-German)
- International

7. Nationality

8. Languages you know (proficiency - B2 or C1 level)

- Deutsch
- English
- Other

9. You are residing at

- On campus
- Off campus

10. Provide your address with just street name and zip code.

Survey Section: Affordability

This section includes questions related to students' housing costs and related topics of affordability.

11. What type of accommodation do you live in currently?

- Shared room
- Shared apartment
- Single apartment
- Live with family
- Dormitory
- Other

12. How many people do you share your accommodation with?
13. What is the total monthly amount you pay for your accommodation, including rent and all utilities (e.g., electricity, water, heating, internet)? Please specify the amount in Euros.
14. Over 50% of your monthly budget is spent on rent?
 - Yes
 - No
15. Mention the dimensions of the room you are currently living in (in square meters)

Survey Section: Accessibility

This section includes questions related to students' accessibility from their residences to the campus and their perceptions on accessibility.

16. What is your primary mode of transportation to campus?
 - Walk
 - Bike
 - Car
 - Bus
 - Train
 - Other:
17. How long does it typically take you to travel from your residence to campus? (Provide in minutes)
18. If you travel by bus, how many minutes does it take you to walk to the nearest bus stop from your house? (Provide in minutes)
19. How many days per week do you go to the university?
 - 0 days
 - 1 day
 - 2 days
 - 3 days

- 4 days
- 5 days
- 6 days
- 7 days

20. Is your frequency of travel to the university dependent on the distance from your house to the university?

- Yes
- No
- May be

21. Would you attend university more often if you live close to the campus?

- Yes
- No
- May be

22. Do you think being closer to campus would improve your academic performance

- Yes
- No
- May be

23. Would you participate in more extracurricular activities on campus/student clubs if you lived closer to the campus

- Yes
- No
- May be

24. How far would you like to live from the university?

- Within 1-2 Km radius
- Within 3-5 Km radius
- More than 5 Km radius
- Doesn't really matter

Survey section: Housing information

This section includes questions related to students' choice of housing and related challenges.

25. Did you find an accommodation before arriving to Koblenz?
 - Yes
 - No
 - Not applicable
26. How many requests did you had to send to find an accommodation? (Provide an approximate number if you are not sure)
27. How many apartments did you have to visit or interview for before getting your current accommodation?
28. Did you have any assistance from the university in finding your current accommodation?
 - Yes
 - No
29. Have you changed houses in Germany?
 - Yes
 - No
30. If you have changed houses, how many houses have you changed so far?
31. For what reasons have you changed the house for? (Choose up to 3)
 - Too far from university/work
 - Lack of cleanliness
 - Problems with room-mates
 - Contract terminated by the owner
 - Rent too high
 - Better amenities in the new accommodation
 - Moving to a larger/smaller apartment
 - Personal reasons
 - End of lease or sublet contract period

- Noise disturbance
 - Bad neighbourhood
 - Not applicable
 - Other:
32. Do you think being able to communicate in German will help you find a place in Koblenz?
- Yes
 - No
 - May be
33. Did you have alternative accommodation options when choosing your current place to live?
- Yes
 - No
34. Would you like to continue living in your current accommodation?
- Yes
 - No
 - May be
35. Would you recommend your current housing to other students?
- Yes
 - No
 - May be
36. What were the top factors influencing your choice of accommodation?(Choose up to 3)
- Size of the room
 - Affordability
 - Proximity to the university
 - Location
 - Cleanliness
 - Amenities
 - Safety of Neighbourhood and Building
 - Friends / Flatmates

- Availability and Demand
- Recommendations or Reviews from Other Students
- Lease Terms and Flexibility
- Pet-Friendly Policies
- Other:

37. What do you appreciate most about your current accommodation? (Choose up to 3)

- Size of the room
- Affordable Rent
- Proximity to the university
- Amenities
- Cleanliness
- Safety
- Flatmates
- Management / Owner responsiveness
- Quiet environment
- Location
- Good Internet Connection
- Access to Public Transport
- Flexibility of Lease Terms
- Other:

38. What are the major challenges you experience with your current accommodation? (Choose up to 3)

- High rent / Expensive
- Small room / Lack of space
- Poor cleanliness / Hygiene issues
- Noise
- Lack of privacy
- Inadequate amenities
- Poor internet connection
- Unfriendly neighbour / Flatmates
- Distance from university

- Poor public transport connections
- Overcrowding
- Maintenance issues / Slow repairs
- Lack of heating / Poor insulation
- Lack of flexibility in lease terms
- Other:

Survey Section: Satisfaction

This section presents Likert-scale questions on students' satisfaction with housing and their overall living experience. All the below questions are answered on a scale of 1-5.

39. How easy was it to find an accommodation in Koblenz?
 Very difficult (1) (2) (3) (4) (5) Very easy
40. How would you rate the size and spaciousness of the room you currently live in?
 Very insufficient (1) (2) (3) (4) (5) Very sufficient
41. How affordable do you find your current accommodation?
 Not affordable (1) (2) (3) (4) (5) Very affordable
42. How would you rate the efficiency of the public transport in Koblenz?
 Very Poor (1) (2) (3) (4) (5) Excellent
43. How would you rate the proximity of your housing to the university?
 Very Far (1) (2) (3) (4) (5) Very Close
44. How would you rate the proximity of your house to the nearest bus stop?
 Very Far (1) (2) (3) (4) (5) Very Close
45. How would you rate the proximity of your housing to the nearby amenities (e.g., Gym, Supermarket, Health care, Restaurants)?
 Very Far (1) (2) (3) (4) (5) Very Close

46. How would you rate the overall quality of the facilities (internet, heating, water, Wi-Fi, laundry, kitchen facilities, etc.) offered in your current accommodation?
- Very Poor (1) (2) (3) (4) (5) Excellent
47. In your current housing situation, how responsive do you think the management or landlord is in resolving your problems or issues?
- Does not respond (1) (2) (3) (4) (5) Responds quickly
48. How would you rate the building condition of your current accommodation?
- Very Poor (1) (2) (3) (4) (5) Excellent
49. Please rate the noise level in your locality according to how much it impacts your academic performance?
- Very Noisy (1) (2) (3) (4) (5) Very Quiet
50. How secure do you feel to interact with your room-mates in current accommodation?
- Very Unsafe (1) (2) (3) (4) (5) Very Safe
51. How safe do you feel in your neighbourhood?
- Very Unsafe (1) (2) (3) (4) (5) Very Safe
52. How would you rate the level of interaction with others in your dorm or shared apartment?
- Very Inactive (1) (2) (3) (4) (5) Very Active
53. How would you rate the cleanliness in your apartment?
- Very Dirty (1) (2) (3) (4) (5) Very Clean
54. How do you think your living arrangements impact your overall university experience and academic achievement?
- Does not impact (1) (2) (3) (4) (5) Impacts a lot
55. To what extent do you agree with the following statement: 'My housing meets my expectations for student accommodation'.
- Strongly Disagree (1) (2) (3) (4) (5) Strongly Agree

56. How pleased are you with your overall experience living in your current accommodation?

Very dissatisfied (1) (2) (3) (4) (5) Very satisfied

57. In your own words, how would you describe your overall experience with student housing in Koblenz? Please share any positive aspects or challenges you have encountered in 20-100 words.

C. Qualitative Interview

Interview with International Relations Office

This interview is being conducted by Kavya Hanumanthu, an MSc student at the University of Koblenz, as part of the research project titled “Measuring Equity and Bias in Student Housing in Koblenz: Affordability, Accessibility, and Satisfaction”. The aim is to gather insights and perspectives on the roles of the Accommodation Team and the International Relations Office in addressing the needs of international students living in student housing. Participation is entirely voluntary, and you may choose not to answer any question or withdraw from the interview at any time. The information collected will be used exclusively for academic purposes within this study. Your contributions will help to build a deeper understanding of how your team supports international students and will inform future efforts to ensure equitable and high-quality student housing.

Thank you for your valuable time and input.

1. How many students get accommodations at the dormitories?
2. How many international students can get the dormitory facilities and for how long?
3. In your experience, what support services or resources do you offer to assist international students in their transition to living in student housing?
4. How often do you hear complaints from students with respect to accommodation troubles?
5. Have you received complaints have with regards to accommodation unavailability?

Follow up Question: When did this Onboarding program start?

Follow up Question: What made you start this program?

6. Could you share any collaborations or partnerships you have established with external organizations or stakeholders to improve the planning, management, and overall student housing experience for international students?
7. Can you tell me available methods for engaging with international students to understand their living experiences?

Follow up Question: Should this student engagement program be improved?

D. Summary Tables of Survey Results

This appendix presents summary tables corresponding to the figures and key analyses reported in the thesis. The tables provide underlying counts and percentages for visualizations such as housing applications, alternative options, rent burden, accommodation choice factors, challenges, and sentiment scores. Only these aggregated summaries are included; the full raw survey dataset is not reproduced here.

Table D.1.: Survey respondents count

Category	Frequency	Percent
Domestic	69	43.7
International	89	56.3
Total	158	100

Table D.2.: Distribution of housing applications submitted by students

Application Range	Count	Percent
0–10	66	41.8
11–20	23	14.6
21–30	20	12.7
31–40	5	3.2
41–50	18	11.4
More than 50	23	14.6
Total	155	100

Table D.3.: Availability of alternative accommodation options by student status

Student Status	Count		Percent	
	No	Yes	No (%)	Yes (%)
Domestic	37	33	52.9	47.1
International	76	13	85.4	14.6
Total	113	46	71.1	28.9

Table D.4.: Factors influencing accommodation choice among students

Factor	Votes	Percent
Affordability	117	21.8
Location	106	19.7
Size of the room	73	13.6
Proximity to the university	66	12.3
Friends / Flatmates	61	11.3
Cleanliness	60	11.2
Amenities	46	8.6
Availability and Demand	23	4.3
Safety of Neighborhood and Building	17	3.2
Recommendations or Reviews from Other Students	13	2.4
Pet-Friendly Policies	12	2.2
Lease Terms and Flexibility	9	1.7
Total	603	100

Table D.5.: Rent burden (over 50% of monthly budget spent on rent) by student status

Student Status	Count		Percent	
	No	Yes	No (%)	Yes (%)
Domestic	34	36	48.6	51.4
International	27	62	30.3	69.7
Total	61	98	38.4	61.6

Table D.6.: Challenges reported in student accommodation

Factor	Votes	Percent
High rent / Expensive	52	15.0
Maintenance issues / Slow repairs	43	12.4
Noise	42	12.1
Small room / Lack of space	33	9.5
Distance from university	32	9.2
Lack of heating / Poor insulation	32	9.2
Poor internet connection	28	8.1
Lack of privacy	24	6.9
Poor public transport connections	23	6.6
Poor cleanliness / Hygiene issues	22	6.3
Unfriendly neighbour / Flatmates	17	4.9
Overcrowding	11	3.2
Inadequate amenities	10	2.9
Lack of flexibility in lease terms	9	2.6
Safety Concerns	4	1.2
Total	352	100

Table D.7.: Distribution of sentiment scores from student housing responses

Sentiment Score	Count	Percent
1 (Very Negative)	9	5.7
2 (Negative)	44	27.8
3 (Neutral)	61	38.6
4 (Positive)	30	19.0
5 (Very Positive)	14	8.9
Total	158	100