

# A Kind Guide to Understanding the German Apartment Rental Market :

Exploring Price and Regional Differences



Project Team 14  
Sangzun Park

## Abstract

The objective of this project is to analyze apartment rental prices in Germany. Not only does it look at the apartment prices across different regions of Germany, but it also examines the various factors that contribute to the formation of these prices. However, another aim of the project is to help people become more familiar with the area of Germany through this process. For this reason, the focus has been on analyzing simple yet core factors, and on creating visualizations that are more likely to be memorable, rather than complex and varied analyses.

# **Introduction**

Since the era of hunting and gathering, housing has always been a critical issue for humanity. With the advent of settled life, the importance of housing has come to occupy an overwhelming proportion of human life. Reflecting this aspect, today's housing costs are one of the largest expenses for modern people. Many of our dreams, in other words, a significant number of life's goals, are related to homes, and many of the difficulties we face now are also connected to housing. Therefore, I wanted to analyze a topic related to housing, and rental prices seemed to be the most appropriate data. The reason for not choosing the pure price of houses is that I wanted to avoid speculative elements as much as possible. Moreover, rental price data can be a window into broader societal issues, such as affordability crises, gentrification, and demographic changes. By analyzing this data, we can uncover the underlying causes of these issues and propose informed solutions. It's not just about numbers and trends; it's about understanding the realities of people's lives and the communities they inhabit.

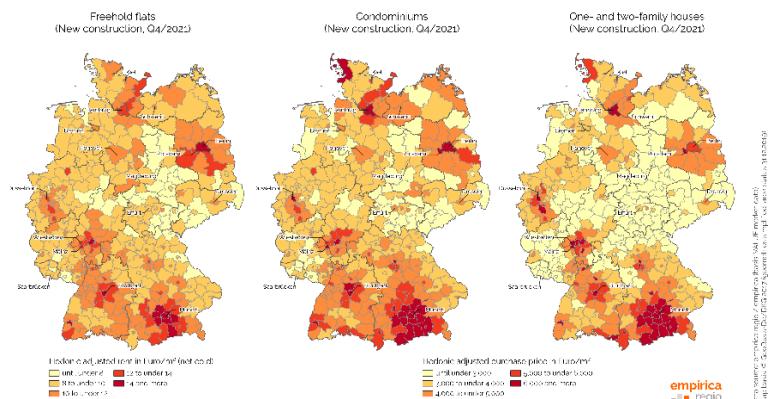
## **Background**

Germany is not only one of the most influential countries in Europe but also in the world. It has the largest population in Europe after Russia and boasts the fourth-largest economy globally. Considering the size of the country, these are notable figures. However, to many in East Asia and the United States, Germany remains a relatively unfamiliar country. Beyond a few stereotypical images, our knowledge about Germany is as distant as the geographical distance between the nations.

There are several reasons why I chose Germany to analyze my motivation concerning housing and rental costs. Firstly, I wanted to exclude areas where real estate speculation is intensified, to focus solely on the original value of housing. Secondly, the region needed to be generally stable in terms of industry, culture, and welfare. Thirdly, it had to be a region with a relatively long history of settlement; in other words, a land with a history. Lastly, it needed to be a relatively unfamiliar area. By analyzing a less familiar region, we can shed our preconceptions and focus on the essence. It also becomes more interesting and memorable. Just as we may not remember a tree in front of our house, but we will always remember the baobab trees seen during a trip to Madagascar. Europe, with its relatively long history, satisfied these conditions, and among it, Germany was the most industrially developed region.

## **Objectives**

My goal is to enhance visualization that aids in analysis. There are a few visualizations available for analyzing German apartment prices. Since this analysis is for North Americans and Asians, I believed there needed to be more gentle assistance regarding German geography during the visualization process. While preparing this proposal, I explored many related analyses. From familiar pie charts to geo graphs, there were numerous studies, but most were not intuitive for those unfamiliar with Germany. If someone is not European or preparing for immigration, they might not find such data very engaging. In other words, many materials neglected the visualization behind the technical visualization. I want to strengthen this aspect so that viewers can quickly understand and remember the analysis for a long time.



[<https://www.empirica-regio.de/en/tags/housing-market/>]

For instance, the above graph is a heatmap chart commonly seen regarding German rental costs. While it is sufficient in providing the necessary information, it may present as a foreign image to those not well-acquainted with Germany. In other words, it may not be very memorable. An interactive topographical map that includes not just German regions but also surrounding areas might be more memorable. Moreover, since most people require information on major cities, other methods might be more appropriate than a visually overwhelming heatmap. Hence, I have aimed for open regional exploration and simplicity.



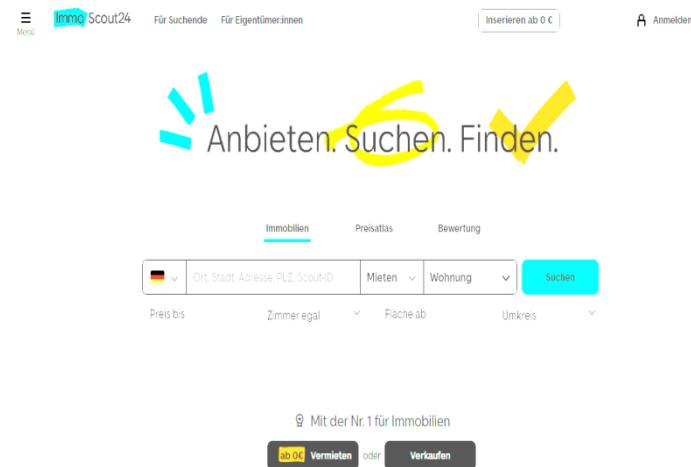
[<https://lukinski.com/stuttgart-buy-rent-house-apartment-property-square-meter-price/>]

This is another example. The above bar chart is simple but intuitively easy to understand. However, its simplicity also makes it easily forgettable. For instance, merely changing the title of the above bar chart could leave most people unable to discern whether it represents data from Germany or Singapore. I believe this is the downside of simplicity. Therefore, I aimed to compensate for this drawback by creating charts that are simple yet varied enough to be memorable.

# Process

## Part 1 EDA

The dataset I chose is from Immoscout24, the largest real estate platform in Germany. Since it is data scraped directly from the platform, it required thorough data cleansing before use.



[<https://www.immobilienscout24.de/>] [<https://www.kaggle.com/datasets/corrieaar/apartment-rental-offers-in-germany/data>]

The EDA (Exploratory Data Analysis) took more time than anticipated, primarily because the dataset I chose comprised 49 columns. I spent considerable time deliberating which features to select among usable data columns (clean data), columns that needed to be made usable, and unusable data columns. During this process, I also experienced trial and error in changing the sub-themes of the analysis.

A screenshot of a Jupyter Notebook titled 'Part 1 / EDA'. The code in cell [1] imports various Python libraries: pandas, matplotlib, numpy, sys, and scipy.stats. Cell [2] shows the command to read the CSV file 'lesso\_data.csv'. Cell [3] displays the first few rows of the DataFrame and its summary statistics. Cell [4] prints the count of NaN values across all columns. The output for cell [3] is a table showing the count of non-null values for each column: region (269550), serviceCharge (261952), heatingType (233994), serviceContract (233994), telekomHybridUploadSpeed (49020), balconyConst (268850), balcony (268850), pictureCount (268850), picture (268850), telekomUploadSpeed (235492), yearConstructed (230380), year (230380), and noFurnishedPlaces (268850). The output for cell [4] is a table showing the total count of NaN values for each column: region (0), serviceCharge (6909), heatingType (44896), serviceContract (3219), telekomHybridUploadSpeed (223830), balconyConst (0), balcony (0), pictureCount (0), picture (1832), and telekomUploadSpeed (268850).

Eventually, I realized that it was impossible to select all the columns for use in one go. Therefore, I planned to go through several iterations of discarding unnecessary columns and refining the dataset. The cleansing process included modifying NaN values, handling outliers, analyzing correlations, and performing KDE (Kernel Density Estimation) analysis. All these steps were centered around the key column: the total rent price.



## Part 2 Setting for Geo map

I designated this stage as Part 2 because creating a geographical map visualization was still a challenging and unfamiliar task for me. After researching various methods, I ultimately decided to use Geo pandas. Additionally, I chose to use the German geographic information data "plz-5stellig.shp" from a German postal code search website for this purpose.

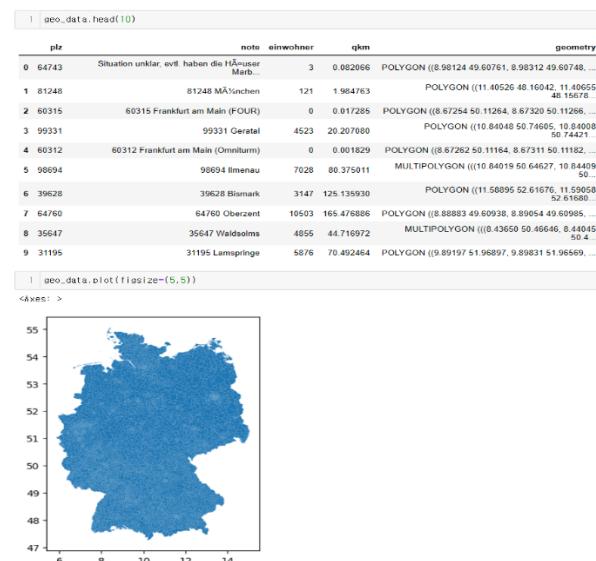
The screenshot shows a search interface for German postal codes. At the top, there's a header with the text "Postleitzahlen Deutschland" and a sub-instruction: "Hier finden Sie Informationen von Städten und Gemeinden aus dem gesamten Bundesgebiet. Zur Postleitzahlensuche können Sie entweder das Suchfeld verwenden oder innerhalb der Bundesländer ihren gesuchten Ort wählen. Neben der Postleitzahl bekommen Sie zusätzlich die Telefonvorwahl, das Autokennzeichen und die Lage mit Koordinaten auf einer Karte angezeigt." Below this is a section titled "Suche in Bundesländer" with a list of states and their abbreviations:

Baden-Württemberg	Niedersachsen
Bayern	Nordrhein-Westfalen
Berlin	Rheinland-Pfalz
Brandenburg	Saarland
Bremen	Sachsen
Hamburg	Sachsen-Anhalt
Hessen	Schleswig-Holstein
Mecklenburg-Vorpommern	Thüringen

To the right of the list is a map of Germany divided into postal code regions, with major cities like Berlin, Hamburg, and Munich labeled. Below the map is a button labeled "Map öffnen". At the bottom of the page are navigation links: "Ortsverzeichnis", "Umkreissuche", "PLZ Entfernung", and "Downloads".

[<https://www.suche-postleitzahl.org/>]

To use Geopandas effectively, it was crucial not only to install Geopandas itself but also to install the surrounding packages in the correct order. I installed Shapely, GDAL, Fiona, and pyproj, ensuring compatibility with my operating system and Python version. Following this, I familiarized myself with the basic usage of Geopandas in Python. This step was vital for ensuring that I could handle geographic data accurately and efficiently in the subsequent stages of my analysis.



## Part3 Visualization

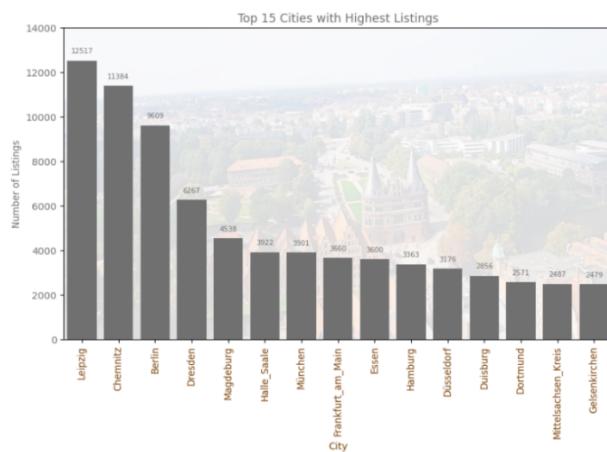
My goal for visualization was to maximize creativity within the scope of my Python skills. The visualization consisted of five main areas :

1. Vertical & Horizontal Bar Charts: For the main themes, I chose the simple yet time-tested method of bar chart visualization. Most of the data centralizes around rental prices (grouped by rental price).
2. Interactive Map Chart: As a first complement to the vertical & horizontal bar charts, I opted for responsive map charts. These allow people to freely explore the map within the chart, enhancing engagement and understanding.
3. Animation Chart: As a second device to complement the bar charts, I incorporated animation charts. This added a point of interest to the analysis that might otherwise become monotonous.
4. Creative Chart: For the final part of the visualization analysis, I deliberately avoided looking at references and instead relied on my imagination to create unique charts.

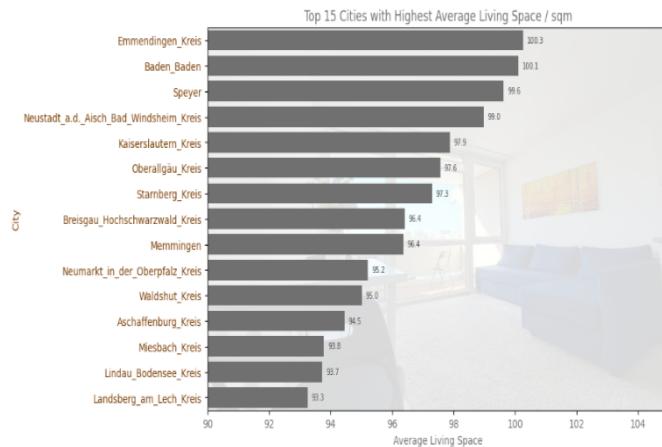
## Results and insights

Vertical Bar: For the main analytical theme, I applied vertical bar charts. In exploring various chart options for the main subject, I found that no method was as quick and effective in grasping data as bar charts. Particularly, pie charts appeared confusing when there were many variables or when the values of different variables were close to each other. My final candidates were bar charts, 2D scatter plots, and line charts. After several experiments, I chose bar charts, believing that people instinctively measure something from bottom to top. Moreover, bar charts are already well-established and proven in data visualization.

I added memorable elements, particularly in design, to these charts. Initially, I experimented with surrounding the plot with vibrant images or photos, but this seemed to diminish the focus on data analysis. After several attempts, I settled on placing a faint background behind the plot. The color of the bars also went through dozens of trials. The final choice was a combination that complemented the theme and background, without detracting from the essence of the analysis. Lastly, I differentiated the text at the bottom. I found that text in an unfamiliar word or language, when displayed in black, was hard to distinguish. Therefore, I changed the color of the text on the horizontal axis to a shade that both blended with other elements and provided differentiation.

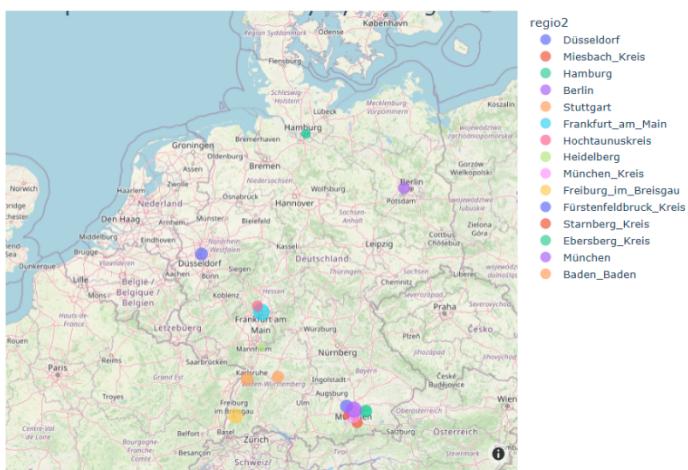


Horizontal Bar: For the sub-themes, I applied horizontal bar charts. This choice was intended to create a distinction from the main theme, thereby stimulating the memory of viewers. For consistency, I utilized the same elements that were finalized for the vertical bars in these horizontal bars as well. This approach helped maintain a cohesive design across different sections of the visualization, while still providing enough variation to keep the audience engaged and attentive to the different aspects of the analysis.



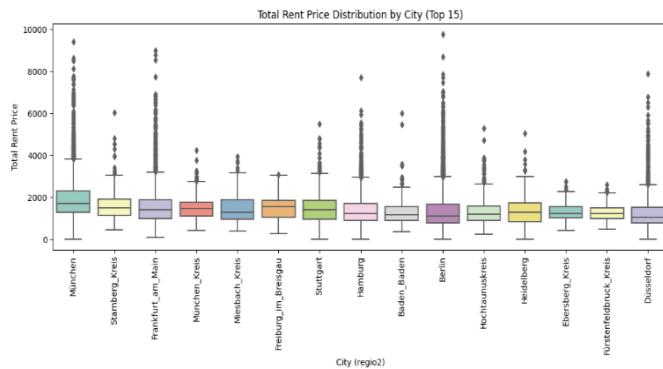
Responsive Geo Map: This was the most technically challenging part of the data visualization process. I loaded the Shapefile using Geopandas and merged it with the postal code information of the cities. Then, I extracted latitude and longitude from the geometry centroids and implemented the final responsive map using open-street-map. Additionally, I marked each city with circular markers of differentiated sizes based on their rankings. This geo chart made it easy to identify the actual locations of the top cities. Most importantly, it allowed for comparison with surrounding landscapes. People can conduct a more comprehensive analysis when they have the context of the surrounding environment.

Interestingly, as I increased the number of analyzed cities, I discovered that most cities with high rental costs were clustered in the southwestern part of Germany. This area corresponds to the old West Germany, adjacent to France and the Alps. Exceptions to this pattern were Hamburg and Berlin. This insight was particularly intriguing as it revealed regional economic patterns and historical influences on modern housing costs.

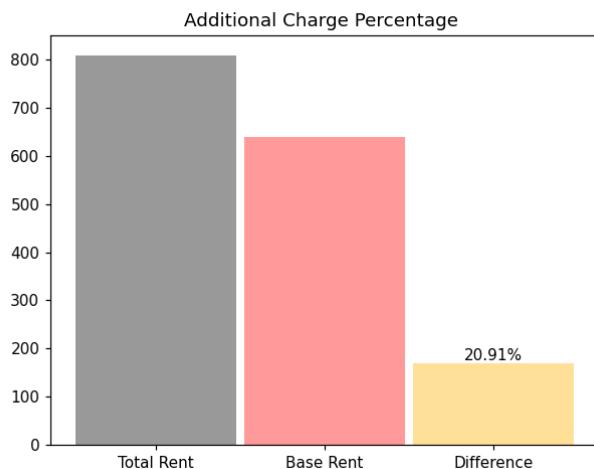


Box Plot: For those interested in more detailed information, I visualized statistical

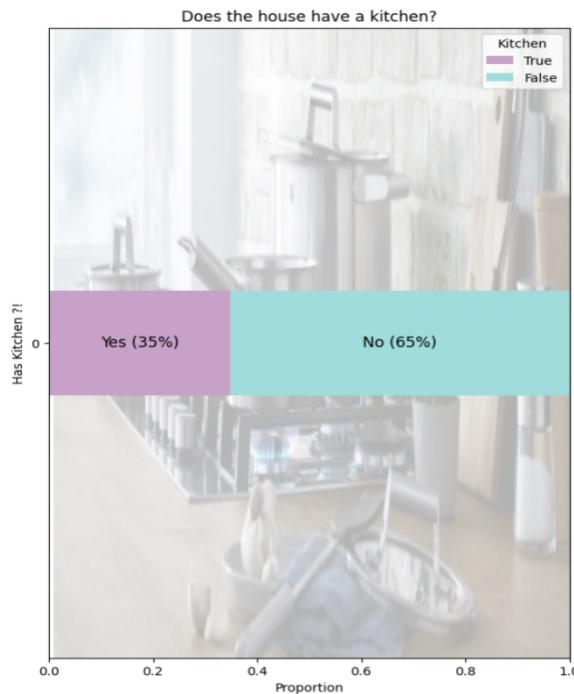
information about the cities using box plots, which allow for easy comparison. Box plots are particularly effective in showing the distribution of data, highlighting the median, quartiles, and potential outliers. This choice was aimed at providing a deeper understanding of the variation and spread of rental costs within each city. It offered a clear visual representation of the data's range and anomalies, thereby catering to viewers who sought a more nuanced and comprehensive view of the housing market in different German cities.



**Animated Bar:** To avoid monotony in the analysis and to make it more engaging, I represented additional costs, apart from the rental costs, using animations. To execute this, the `%matplotlib notebook` command is required. This command is used for dynamic rendering in Jupyter Notebooks, allowing for interactive animations within the notebook environment.



**New Experimental Bar Chart:** To conclude the analysis with a creative touch, I experimented with a novel chart design. I wanted to represent an interesting and lesser-known fact about Germany in a fun and accessible way. During the analysis, I discovered that, unlike in other countries, it is common in Germany for people to move their kitchens with them when they relocate. I visualized this unique aspect in a simple yet instantly comprehensible form. I utilized the entire horizontal axis at the center of the screen for this visualization. This design approach not only alleviated the tension of the analysis but also was intended to leave a lasting positive impression of the overall study. By highlighting this unusual cultural practice in Germany, the chart not only conveyed information but also provided an engaging story, adding depth and context to the data being presented.



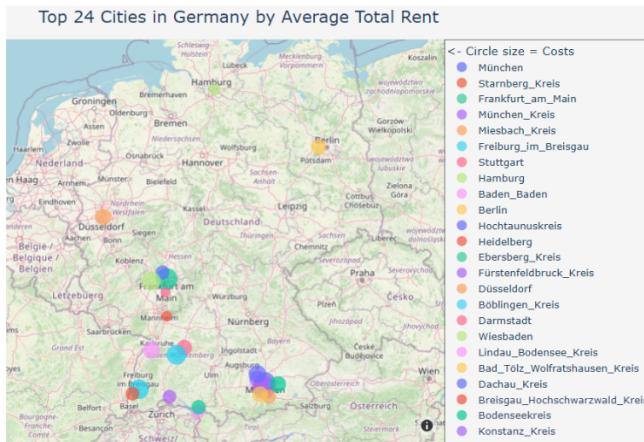
## Conclusion

Before taking this course, I had a vague notion that visualization was important. However, it was through attending the course, and especially after starting this project, that I came to understand even a simple-looking graph involves a lot of thought and calculation. For instance, the bar chart I used for the main topic required me to consider numerous elements and repeat experiments to perfect this seemingly simple graph. Additionally, I reconfirmed the significant role of design and color in visualization.

However, I had to make several modifications to the charts based on feedback from peers. Firstly, I changed all labels in the vertical & horizontal bar charts to the same color. My earlier experimental attempts were perceived as unfamiliar discomfort rather than a positive stimulus by most people. Other changes included adjusting font sizes, modifying some colors, and rotating text labels on the horizontal axis. This learning experience highlighted the intricacies of data visualization, the importance of user-friendly design, and the need for ongoing adjustments based on feedback. It underscored the balance between creativity and clarity in presenting data in a way that is both engaging and informative.



Additionally, modifications were needed in the map visualization as well. For instance, I overlooked small but crucial details such as the order of items in the legend and the meaning of the circle sizes. These elements, though minor, are essential for understanding the meaning and implications of the visualized data. The feedback highlighted the importance of not only presenting data in an aesthetically pleasing way but also ensuring that every aspect of the visualization is intuitive and informative, aiding in the accurate interpretation of the data.



Through feedback, I was able to address the elements I had misunderstood or omitted. This process taught me the importance of testing and feedback for a finished product. Fortunately, there was a lot of positive response to my last experimental bar chart. People appreciated how it effectively conveyed information that was previously not widely known.

If I were to continue this project, my first step would be to focus on acquiring more data. The original dataset for this project was large, but there weren't many significant or usable columns. If such data becomes available, I would like to collect about 10 feature columns and represent them using a Sankey chart. In this chart, the thickness of each branch converging on rental cost would represent the magnitude of each feature column. Additionally, I would like to animate the trend of rental cost fluctuations over five-year periods on the map chart. This representation would require more technical preparation. Overall, this project not only enhanced my skills in data visualization but also underscored the continuous need for learning, adapting, and innovating in the field of data analysis.

## Reference

<https://en.wikipedia.org/wiki/Germany>

[https://en.wikipedia.org/wiki/Geography\\_of\\_Germany](https://en.wikipedia.org/wiki/Geography_of_Germany)

<https://www.empirica-regio.de/en/tags/housing-market/>

<https://lukinski.com/stuttgart-buy-rent-house-apartment-property-square-meter-price/>

<https://www.immobilienscout24.de/>

<https://www.kaggle.com/datasets/corrieaar/apartment-rental-offers-in-germany/data>