Achievement 6 Project Brief: Immoscout Germany rent

Context

The rental market is a dynamic and ever-evolving sector, and the wealth of data available on rental websites presents a valuable opportunity for in-depth analysis. I believe that by harnessing this data, we can uncover valuable insights that benefit a wide range people. One of them being myself because I'm interested on moving myself, which equals to me having fun conducting this Analysis.

Data Source

This data is external from the company Immoscout 24 a part of the well-known search engines for all types of sort like example Autoscout 24(it's the same company). Immoscout has an API which you have to get access to, but because I'm short on time. I went on Kaggle and downloaded a dataset that has been scraped from their API. The only issues with this data could be missing information, wrong values and incorrect information that is being collected by the search engine. It can be wrong because it's based on inputs from customers who use the site.

Immoscout has only 3 datasets available for download, at all times and I got an older that came from their official website, which is open source and that needed some correction.

Data Profile

Immoscout data: The data is from 2018-2020 and includes information about everything that is important for a tenant looking for a home.

Variable	description	time-variant or invariant	structured or unstructured	qualitative or quantitaive	nominal / ordinal or discrete / continuous
regio1	name of the state	time-invariant	structured	qualitative	nominal
serviceCharge	additional cost	time-variant	structured	quantitative	continuous
heatingType	what type of heating is used	time-variant	structured	qualitative	nominal
newlyConst	is the building new yes or no	time-variant	structured	qualitative	binary
balcony	is there a balcony yes or no	time-invariant	structured	qualitative	binary
totalRent	combined rent with extra cost	time-variant	structured	quantitative	continuous
yearConstructed	building year constructed	time-invariant	structured	quantitative	discrete
hasKitchen	has kitchen yes or no	time-variant	structured	qualitative	binary
cellar	has cellar yes or not	time-invariant	structured	qualitative	binary
baseRent	rent without extra cost	time-variant	structured	quantitative	continuous
livingSpace	living space in sqm	time-invariant	structured	quantitative	continuous
condition	condition of the flat	time-variant	structured	qualitative	nominal
petsAllowed	are pets allowed yes, no or maybe	time-variant	structured	qualitative	nominal
lift	are there lifts in the building yes or no	time-invariant	structured	quantitative	binary
baseRentRange	rent range by bins	time-variant	structured	quantitative	ordinal
typeOfFlat	what type of flat is it	time-invariant	structured	qualitative	nominal
noRooms	rooms of the flat	time-invariant	structured	quantitative	discrete
floor	the floor of the flat	time-invariant	structured	quantitative	discrete
garden	is there a garden yes or no	time-variant	structured	qualitative	binary
regio2	district or city	time-invariant	structured	qualitative	nominal
regio3	neighborhood	time-invariant	structured	qualitative	nominal
year	year of registration	time-variant	structured	quantitative	discrete

Data limitations and ethics

The data can't be biased because it's just collection information about what is registered on the website. It's collected frequently but only three datasets are available at the same time which limits broad information over time. So, a timeline might not be as accurate as it could be, which skews the accuracy of some discoveries that are going to be made.
Manual errors when the data is being input, from the user, is something that will happen for sure. There might some missing values or different version of an answer but all in all it will be categorized correctly.

Questions to explore

- What is the average monthly rent in a specific city or neighborhood?
- How has the average rent in a particular city or region changed over the past year?
- Are there significant price differences between furnished and unfurnished rentals?
- What is the average square footage of rental properties in different neighborhoods?
- Are there correlations between property age and rental prices?
- How does the heating type impact rent prices?
- Can we predict future rent prices in a specific area based on historical data?
- What factors contribute to fluctuations in rental prices, and can we build predictive models?

These are some questions I would like to answer if possible, I might find more question along the way of the Analysis

Source of data

https://www.immobilienscout24.de/