

Business Intelligence Project

Apartment Rental Data Analysis

16.05.2025

Fallah Mohammad Mahdi, Muller Nicolai, Seibert Jakob, Seo Jiyoung



Agenda

01 Data Selection

02 BI Justification and Focus

03 Data Modeling

04 Data Processing

05 Data Analytics

06 Key Insights

07 Reflection

08 Work Distribution

01 Data Selection

Chosen Topic: Apartment Rental Market in Germany and Poland

- Both datasets share key variables such as size, rooms, rent, balcony and date.
- They can be easily integrated into a common star schema, enabling meaningful cross-country and time-based analysis.

Dataset Overview

1. Apartment rental offers in Germany		2. Apartment Prices in Poland	
Description	Rental offers from Immoscout24 in Germany	Rent offers from the 15 largest cities in Poland	
Number of Files	1	8	
Rows	268,850	69,920	
Columns	49	28	
Time Range	from the dates 2018-09-22, 2019-05-10 and 2019-10-08.	8 months(November 2023 and June 2024)	
Source	Kaggle (https://www.kaggle.com/datasets/corrieaar/apartment-rental-offers-in-germany?resource=download)	Kaggle (https://www.kaggle.com/datasets/krzysztofiamroz/apartment-prices-in-poland/data?select=apartments_pl_2024_03.csv)	

Fact Table Definition

- One entry: Rent for an apartment on a specific date in Germany or Poland, with a given living space size(m2), Number of rooms, balcony status(Yes/No).
- Measures: Rent price(transformed by purchasing power).

Dimensions

- Date
- Country
- Apartment
- Rent category

02 BI Justification

Analytical Objective

- The goal is to support data-driven decision making for both real estate investors and renters in Germany and Poland.
- Analyze how features like balcony, room count, and size affect rent, and compare PPP-adjusted price ranges between the two countries.

Analytical Questions

1. What is the minimum, maximum or average rent per country?
2. How do rent prices per square meter compare between Germany and Poland?
3. What influence does a balcony have on the rent of an apartment?
4. Does a balcony have a stronger influence on the rent price in Poland or Germany (if it has one)?
5. What influence does the number of rooms have on the rent? Is the relationship linear?
6. How do rent categories (e.g., <500, 500–1000) relate to number of rooms and apartment size?

Benefits

Investors Perspective:

- Gain insights into which apartment attributes (e.g., size, balcony, room count) most significantly impact rent prices
- Use this information to identify features that maximize rental income and ROI when investing in properties.

Renter Perspective:

- Understand which apartment features have the greatest influence on rental costs.
- Make smarter decisions by avoiding **unnecessary attributes** (e.g., "Do I really need a balcony?")

Cross-Country Market Perspective:

- Analyze how country-level differences affect housing prices

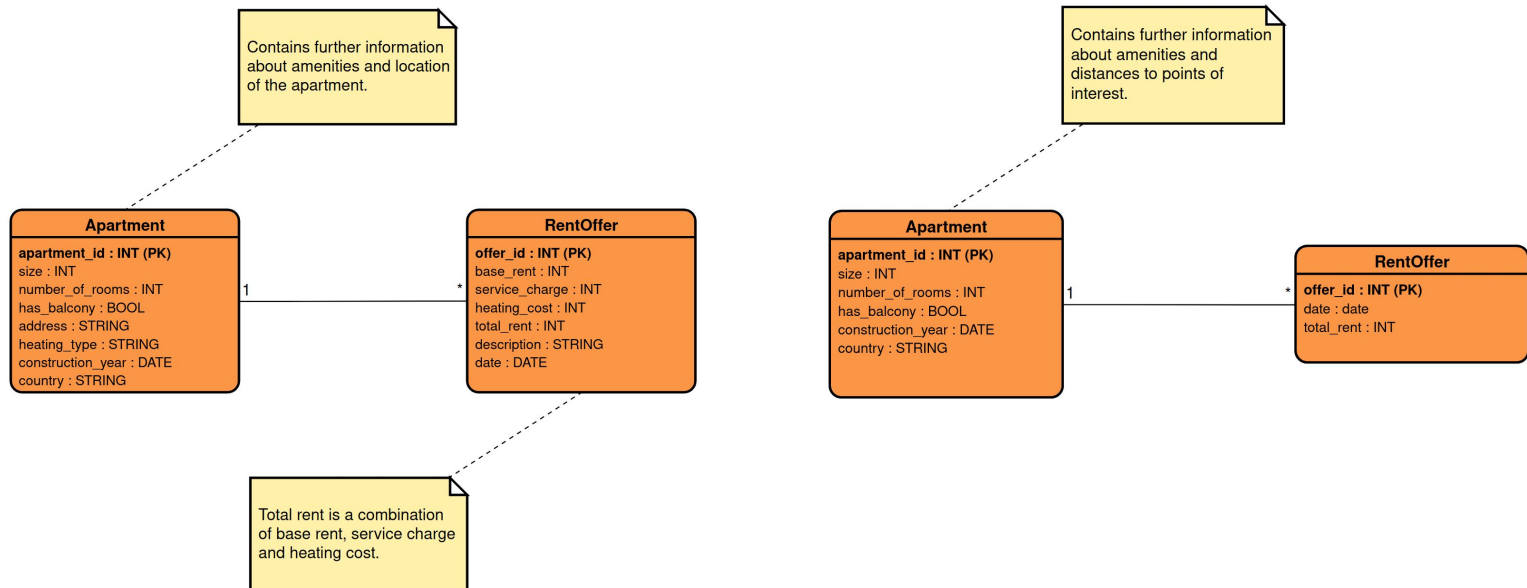
03 Data Modeling

Source Data

- Entity-Relationship Models show the original structure of two source datasets(normalized form, 3NF)

	1. Apartment rental offers in Germany	2. Apartment Prices in Poland
Columns	<ul style="list-style-type: none">Apartment: size, number_of_rooms, has_balcony, address, heating_type, construction_year, country.RentOffer: base_rent, service_charge, heating_cost, date, description.	<ul style="list-style-type: none">Apartment: Size, number_of_rooms, has_balcony, construction_year, country.RentOffer: date, total_rent.

ER model

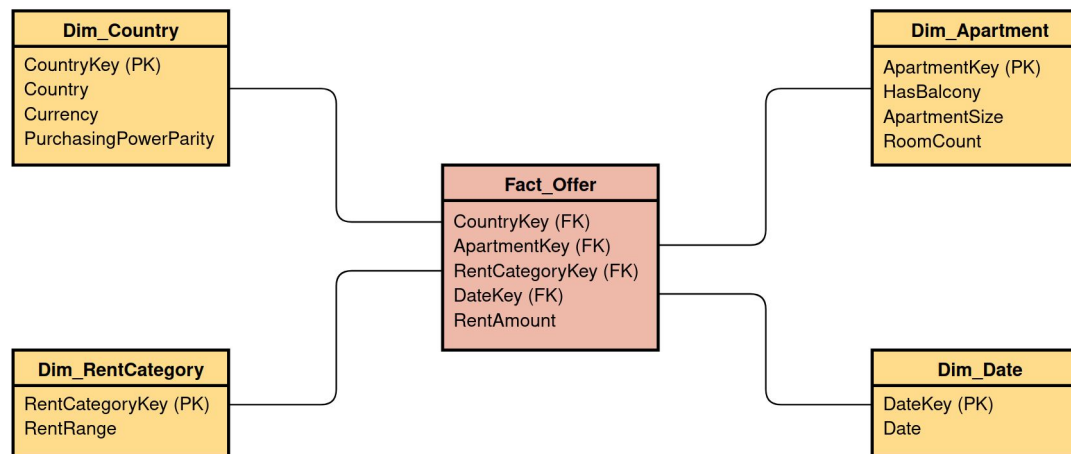


03 Data Modeling

BI Data

	1. Apartment rental offers in Germany	2. Apartment Prices in Poland
Columns Choices	<ul style="list-style-type: none">Features like balcony, room count, and apartment size were included because they are directly linked to the analytical questions, such as understanding their impact on rent and comparing pricing patterns.Variables such as heating type, description, or exact address were excluded since they do not contribute to answering the defined questions and are either inconsistent or unstructured across datasets.	
Used Columns	<ul style="list-style-type: none">Balcony(Boolean): indicating whether the apartment has a balconytotalRent(int): indicating price in €livingSpace(int): indicating size in sqmnoRooms(int): indicating number of roomsDate(date): indicating when the data was scraped	
Dimensions Description	<ul style="list-style-type: none">Country: Country-level metadata including name, currency, and purchasing power parity (PPP) factor for rent adjustment.Apartment: Apartment features that influence rental price, such as size, number of rooms and balcony.Rent Category: Classifies rent price into defined categories (e.g., <500, 500–1000) for aggregated analysis and visualization. This was implemented using Tableau.Date: Date information for each rental entry, enabling time-based analysis.	

Star Schema



04 Data Processing

Data Transformation Steps

- Created reusable transformation pipeline in Python for reproducibility.
- The pipeline was written in Python with reusable logic, using loops and standardized cleaning steps.
- No manual cleaning was done in spreadsheets or Tableau. All transformations were handled programmatically.

1 Data Collection

- Downloaded raw data from Kaggle using kagglehub.
- Loaded 1 CSV file for Germany and 8 monthly rental files for Poland.
- All Poland rental CSVs were merged into one DataFrame after extracting dates from filenames.

2 Data Selection & Processing

- Selected only relevant columns from both datasets.
- Germany: Filled missing totalRent values using baseRent + serviceCharge + heatingCosts
- Poland: Converted rent from PLN to EUR (price * 0.23), adjusted to PPP (Purchasing Power Parity)

3 Data Integration

- Renamed and aligned column names across both datasets for consistency.
- Combined German and Polish data into a single dataframe.
- Unified structure: size, number_of_rooms, balcony, elevator, floor, construction_year, date, total_rent, country.

4 Data Cleaning

- Outliers (rent price < 300 or > 20,000).
- Non-integer room counts.
- Apartments with more than 10 rooms.

5 Star Schema Construction

- Constructed four dimension tables from the cleaned dataset(country, apartment, date, rent category).
- Created fact_offer table by joining the cleaned dataset with all dimension tables using foreign keys.
- Exported all tables as CSVs for loading into Tableau.

05 Data Analytics

Analytical Question 1. What is the minimum, maximum or average rent per country?

- While Poland shows a higher average and median rent than Germany, Germany has a much larger price range, with luxury listings skewing the maximum upward to €19,998.
- This highlights the higher price variance in Germany, whereas Poland appears more concentrated.

	Germany	Poland
Min Rent	€300,00	€301,30
Avg. Rent	€825,18	€894,54
Max Rent	€19.998,00	€5.290,00
Median Rent	€655,00	€729,10

Analytical Question 2. How do rent prices per square meter compare between Countries?

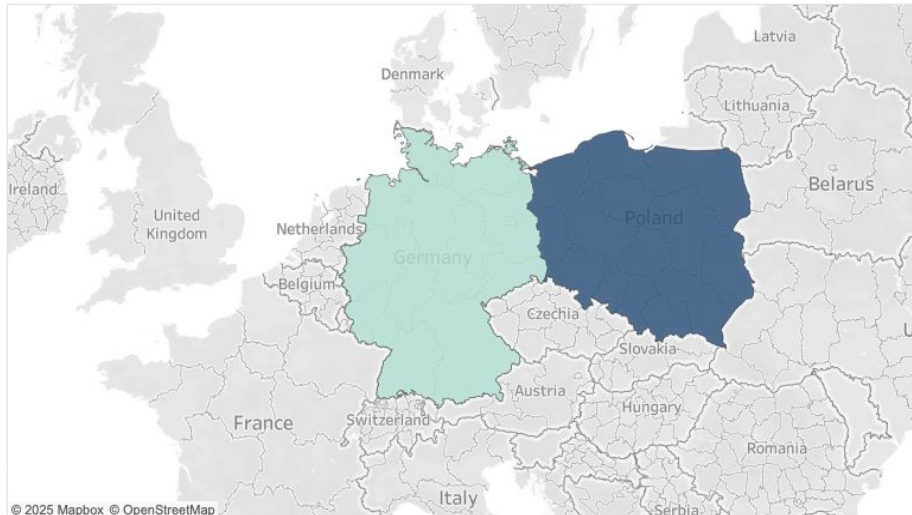
- The average rent per m² is higher in Poland than in Germany.

Avg. Rent Price p SQM

€ 11,24  € 16,03

	Germany	Poland
Avg. Rent per m2	€11,24	€16,03

Rent Price p size



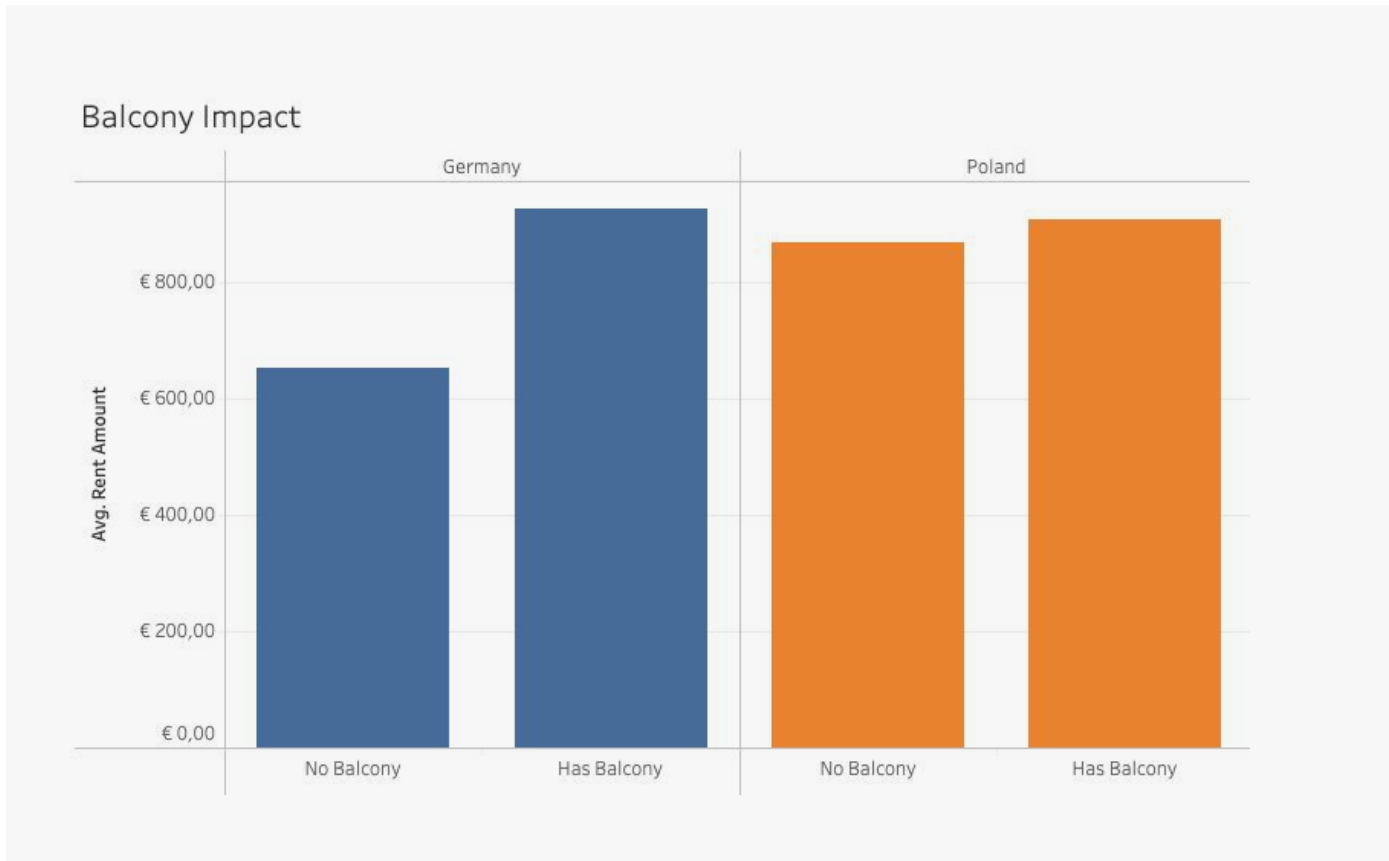
05 Data Analytics

Analytical Question 3. What influence does a balcony have on rent?

- In Germany, having a balcony leads to a significant increase in rent.
- In Poland, the difference exists but is much smaller.

Analytical Question 4. Does a balcony have a stronger influence on the rent in Poland or Germany?

- This suggests that balconies are more highly valued in Germany than in Poland.



05 Data Analytics

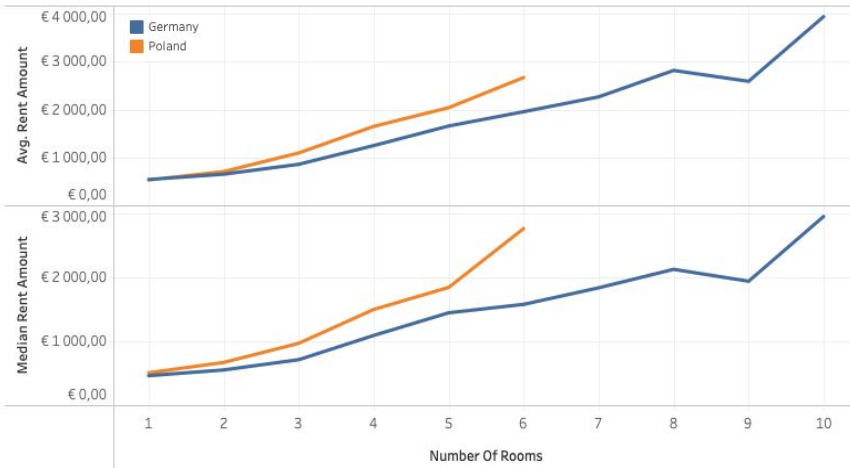
Analytical Question 5. What influence does the number of rooms have on the rent?

- In both countries, average and median rent increase with the number of rooms, indicating a generally positive linear relationship.
- However, in Germany, there is a slight drop in rent at 9 rooms, suggesting either data sparsity or a deviation from the linear trend in larger properties.

Analytical Question 6. How do rent categories relate to number of rooms and apartment size?

- As rent increases up to around €8,000, both room count and apartment size show a strong positive linear correlation.
- However, in higher rent ranges (between €11K and €16K), the number of rooms plateaus, while size continues to increase, suggesting that luxury apartments tend to prioritize spaciousness over the number of rooms.
- Above €8K, this correlation becomes weaker which is likely due to limited data volume in the highest price segments.

Rent per Number of Rooms



Rent Categories



06 Key Insight

Key Insight Summary

1. Country-Level Rent Comparison

- Germany has a wider price range (up to €19,998), but a lower median rent (€655) compared to Poland (€729).
- Suggests presence of luxury outliers in Germany and a more concentrated market in Poland.

2. Rent per Square Meter

- Surprisingly, Poland has a higher avg. rent per m² (€16.03) than Germany (€11.24), possibly due to smaller apartment sizes inflating the ratio.

3. Balcony Effect

- Having a balcony increases rent explicitly in Germany, while in Poland, the effect is less pronounced.
- Indicates higher value placed on balcony in the German market.

4. Number of Rooms vs Rent

- Rent rises almost linearly with room count in both countries up to 5–6 rooms.

5. Rent Categories vs Size and Rooms

- Rent categories from €0K–8K show a strong positive correlation with both room count and apartment size.

Conclusion

This analysis highlights how apartment features influence rental prices differently across countries, and how pricing patterns evolve in higher rent segments.

These insights are useful for

- **Investors**, to optimize for high-ROI property features
- **Renters**, to evaluate trade-offs in features vs cost
- **Policy makers**, to understand housing market dynamics between two EU countries

07 Reflection

What worked well

- Successfully built a reusable data pipeline in Python.
- Designed a clear star schema that supports flexible analysis in Tableau.
- Answered all analytical questions with interactive visualizations.
- Generated insights aligned with the initial objectives.

What could be improved

- With more time-series data and larger volume, insights could be more explicit and statistically robust.
- With data from additional European countries, the analysis could offer more meaningful cross-country comparisons.
- With more data and time, the model could be extended to forecast future rent trends

Questions encountered

- We weren't sure about whether the properties of the Apartment dimension should be split into separate ones. We decided to unify them into a single dimension.

Github Repository

https://github.com/mahdyfalah/housing_business_intelligence

Data source

<https://www.kaggle.com/datasets/corrieaar/apartment-rental-offers-in-germany?resource=download>

https://www.kaggle.com/datasets/krzysztofjamroz/apartment-prices-in-poland/data?select=apartments_pl_2024_03.csv

08 Work Distribution

Fallah Mohammad Mahdi

- Contributed to part 1 data selection, topic selection, definition of dimensions.
- Contributed to part 2 BI justification, definition of analytical questions.
- Led part 4 data processing and python pipeline development.

Muller Nicolai

- Contributed to part 1 data selection, topic selection, definition of dimensions.
- Contributed to part 2 BI justification, definition of analytical questions.
- Led part 3 data modelling and schema design.

Seibert Jakob

- Contributed to part 1 data selection, topic selection, definition of dimensions.
- Contributed to part 2 BI justification, definition of analytical questions.
- Led part 5 data analytics using Tableau visualizations.

Seo Jiyoung

- Contributed to part 1 data selection, topic selection.
- Contributed to part 5 and 6 describing insights based on Tableau visualizations.
- Created and structured the PowerPoint report for submission.