

Analysis of Airbnb Rental Prices in European Cities

Thomas Lin

Department of Statistics, University of Connecticut

December 16, 2023

Introduction

With its presence in over 220 countries and 7 million listings, Airbnb stands as a transformative force in the sharing economy. Offering a diverse range of accommodations, from urban lofts to countryside cottages, Airbnb encapsulates the essence of curated travel experiences. For hosts, pricing becomes a delicate equilibrium, shaping the competitiveness and profitability of their listings. Travelers, amidst numerous choices, navigate pricing alongside considerations like location and amenities.

Beyond individual transactions, Airbnb's pricing dynamics echo through urban planning and community dynamics, posing a nuanced challenge for policymakers. Our research delves into the intricate tapestry of Airbnb pricing, unraveling the determinants influencing this dynamic marketplace. We aim to contribute to the academic understanding of the sharing economy and provide practical insights for hosts, travelers, and policymakers navigating the evolving travel and hospitality landscape.

Background

Airbnb's expansive growth, with over 7 million listings across 220 countries, has reshaped the global travel and lodging sector. The platform's ubiquity underscores its transformative impact, offering diverse accommodations from modest apartments to luxurious villas.

Within this dynamic marketplace, pricing strategies exhibit considerable variability. Hosts, acting as independent entrepreneurs, employ diverse approaches influenced by factors such as location, property type, and seasonal demand. Location plays a pivotal role, with properties in different regions commanding distinct pricing structures.

Property type further shapes Airbnb pricing, reflecting the diverse range of accommodation styles available on the platform. Seasonal demand introduces complexity, with travel patterns influencing booking volume and optimal pricing strategies for hosts throughout the year.

Existing studies highlight the impact of property attributes, location characteristics, and host behavior on pricing. However, a comprehensive, data-driven exploration of Airbnb pricing dynamics is needed. This research project aims to fill this gap by delving into pricing trends and uncovering patterns within the Airbnb dataset. The goal is to provide actionable insights for hosts, travelers, and policymakers navigating the dynamic realm of short-term accommodations.

Specific Aims

This research project is designed to achieve the following specific aims:

- 1. Data Collection:** Gather and curate a comprehensive dataset of Airbnb listings in European cities, incorporating detailed information on property attributes, location characteristics, host profiles, and pricing details.
- 2. Data Analysis:** Apply advanced analytical techniques to identify primary factors influencing Airbnb pricing in European cities. Investigate relationships between pricing and property type, location, occupancy rates, and seasonality variables.
- 3. Predictive Models:** Develop robust predictive pricing models specific to European cities to estimate Airbnb rental prices based on identified determinants. Empower hosts and travelers with tools to make informed decisions.

Data Collection

To ensure the integrity and relevance of the dataset, this study will utilize the Kaggle dataset containing Airbnb rental data for multiple European cities. The dataset includes various features that can significantly contribute to understanding the factors influencing European Airbnb rental prices. The relevant columns from the dataset include:

- **realSum:** Total price of the listing
- **room_type:** Type of room offered (e.g., private, shared, entire home/apt)
- **room_shared:** Whether or not the room is shared
- **person_capacity:** Maximum number of people allowed in the property
- **host_is_superhost:** Whether or not the host is a super host (boolean value)
- **multi:** Whether it's for multiple rooms or not
- **biz:** Whether it's for business use or family use
- **dist:** The distance from the city center
- **metro_dist:** The distance from the nearest metro station
- **guest satisfaction overall:** Overall satisfaction rating from guests
- **Cleanliness rating:** Rating for cleanliness
- **Bedrooms:** Number of bedrooms

Methods

The analysis employs a comprehensive approach utilizing selected features from the Kaggle dataset to identify primary factors influencing Airbnb pricing in European cities:

1. **Feature Selection:** Select relevant features, such as realSum, room_type, host_is_superhost, dist, metro_dist, etc., based on their importance and relevance to the research goals.

2. **Exploratory Data Analysis (EDA):** Conduct exploratory data analysis to understand the distribution of critical variables, identify potential outliers, and gain insights into the relationships between features.
3. **Regression Analysis:** Utilize multiple regression techniques to model the relationship between pricing and selected independent variables. Identify significant factors influencing Airbnb pricing in European cities.
4. **Machine Learning Models:** Implement advanced machine learning algorithms to build predictive pricing models specific to European cities. These models will provide a granular understanding of how various factors affect Airbnb pricing in the European market.
5. **Geospatial Analysis:** Conduct geospatial analysis by plotting distance variables (dist, metro_dist) with respect to latitude and longitude. This will indicate geographical locations where businesses could benefit from higher occupancy rates based on neighborhood proximity and help tackle seasonal variations.
6. **Correlation Analysis:** Use correlation matrices to identify strong correlations between variables. This will help establish relationships across different features and guide decisions on which parameters to consider based on one another.

Descriptive Statistics

	realSum	person_capacity	cleanliness_rating
Min.	34.78	2.000	2.000
1st Qu.	148.75	2.000	9.000
Median	211.34	3.000	10.000
Mean	279.88	3.162	9.391
3rd Qu.	319.69	4.000	10.000
Max.	18545.45	6.000	10.000

Table 1: Descriptive statistics for key variables (Part 1).

	guest_satisfaction_overall	bedrooms
Min.	20.00	0.000
1st Qu.	90.00	1.000
Median	95.00	1.000
Mean	92.63	1.159
3rd Qu.	99.00	1.000
Max.	100.00	10.000

Table 2: Descriptive statistics for key variables (Part 2).

Correlation Matrix

	realSum	person_capacity
realSum	1.00	0.20
person_capacity	0.20	1.00

Table 3: Correlation matrix for selected variables (Part 1).

	cleanliness_rating	guest_satisfaction_overall
cleanliness_rating	1.00	0.71
guest_satisfaction_overall	0.71	1.00

Table 4: Correlation matrix for selected variables (Part 2).

bedrooms	
bedrooms	1.00

Table 5: Correlation matrix for selected variables (Part 3).

Graphs and Visuals

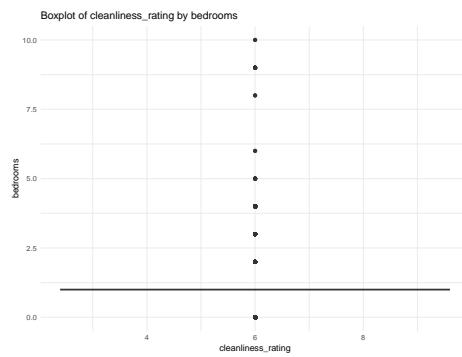


Figure 1: Cleanliness Rating vs. Bedrooms Boxplot

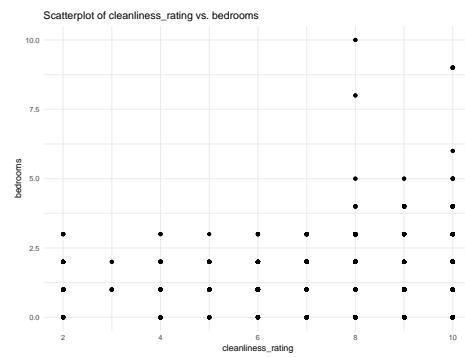


Figure 2: Cleanliness Rating vs. Bedrooms Scatterplot

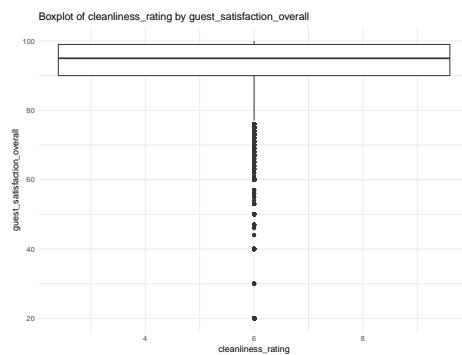


Figure 3: Cleanliness Rating vs. Guest Satisfaction Overall Boxplot

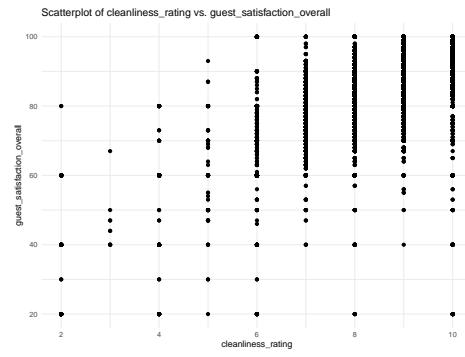


Figure 4: Cleanliness Rating vs. Guest Satisfaction Overall Scatterplot

Figure 5: Graphs and Visuals

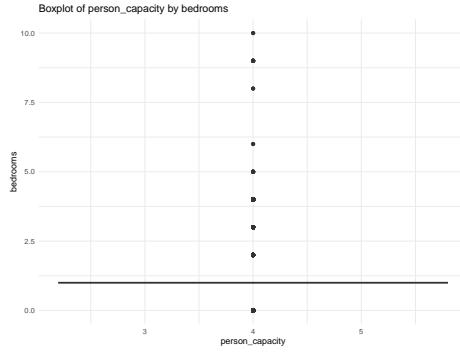


Figure 6: Person Capacity vs. Bedrooms Boxplot

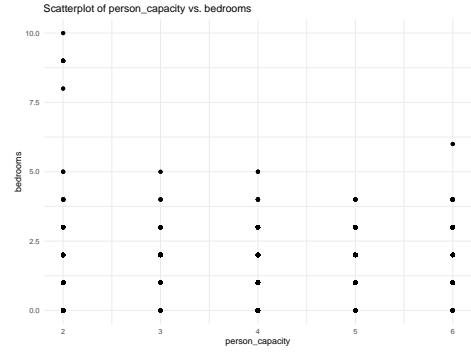


Figure 7: Person Capacity vs. Bedrooms Scatterplot

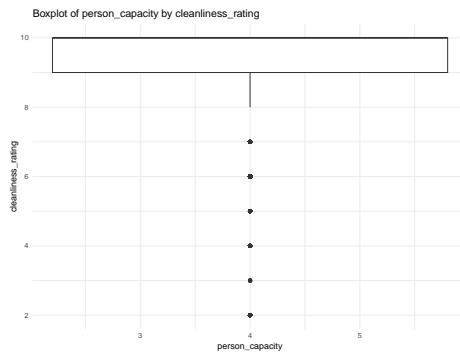


Figure 8: Person Capacity vs. Cleanliness Rating Boxplot

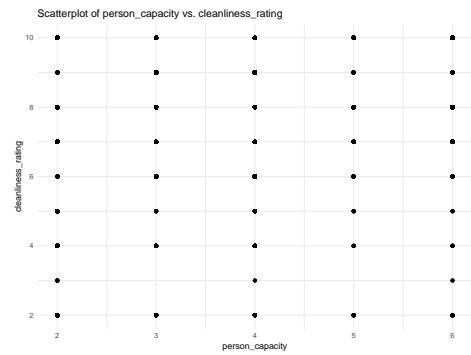


Figure 9: Person Capacity vs. Cleanliness Rating Scatterplot

Figure 10: Graphs and Visuals

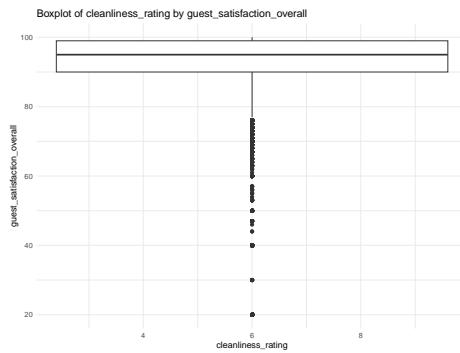


Figure 11: Cleanliness Rating vs. Guest Satisfaction Overall Boxplot

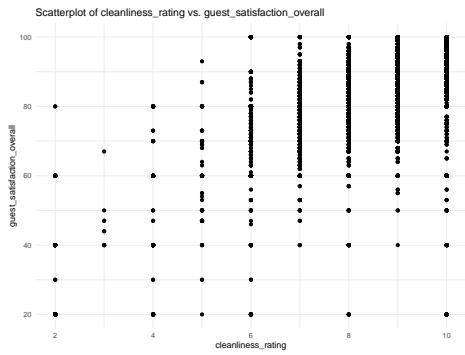


Figure 12: Cleanliness Rating vs. Guest Satisfaction Overall Scatterplot

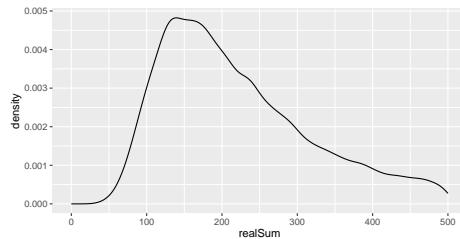


Figure 13: Density Graph

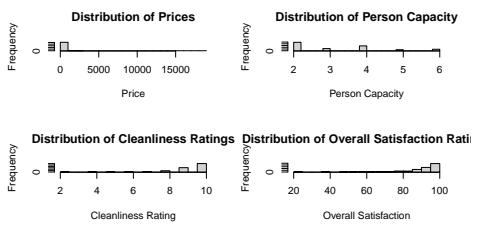
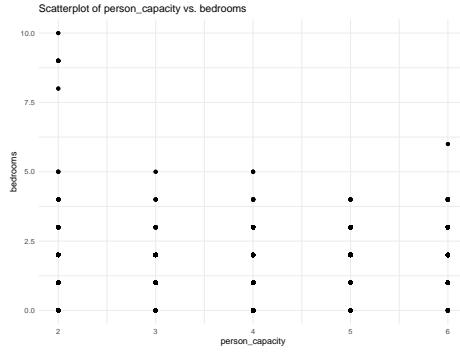
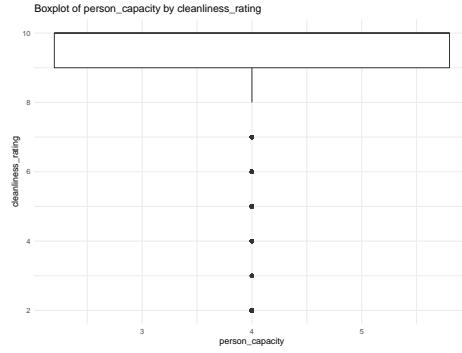


Figure 14: Histograms AirBnB

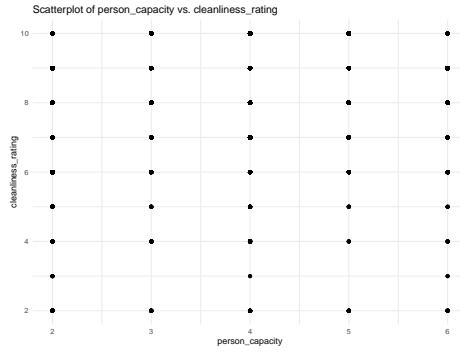
Figure 15: Additional Graphs



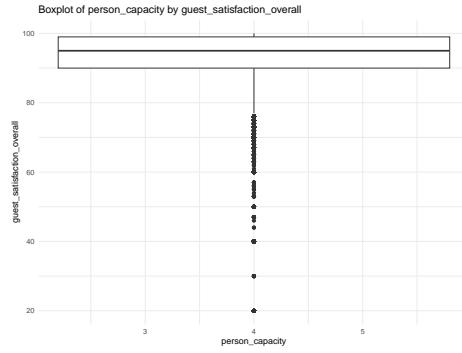
(a) Person Capacity vs. Bedrooms
Scatterplot



(b) Person Capacity vs. Cleanliness Rating Boxplot

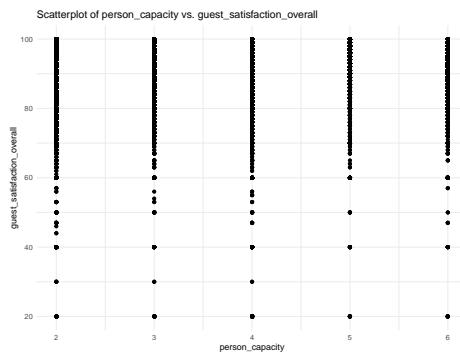


(c) Person Capacity vs. Cleanliness Rating Scatterplot



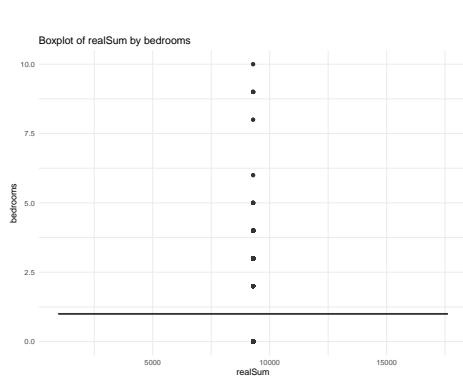
(d) Person Capacity vs. Guest Satisfaction Overall Boxplot

Figure 16: Additional Graphs

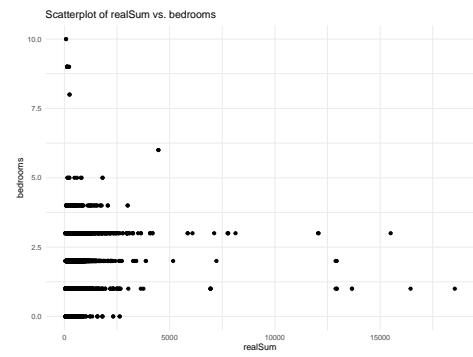


(a) Person Capacity vs. Guest Satisfaction Overall Scatterplot

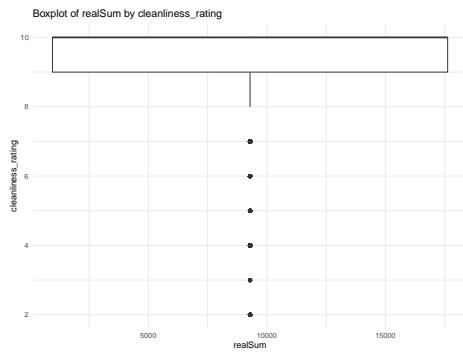
Figure 17: Additional Graphs



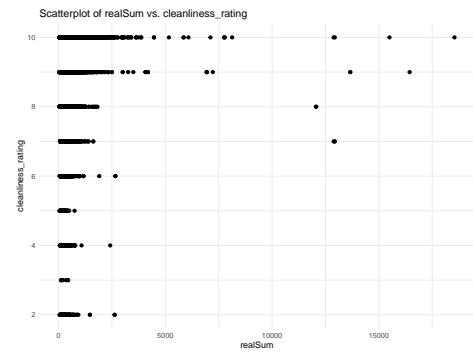
(a) Real Sum vs. Bedrooms Boxplot



(b) Real Sum vs. Bedrooms Scatterplot

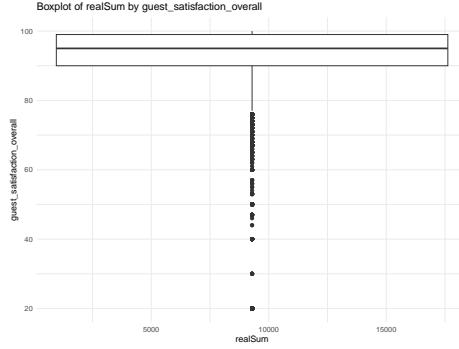


(c) Real Sum vs. Cleanliness Rating Boxplot

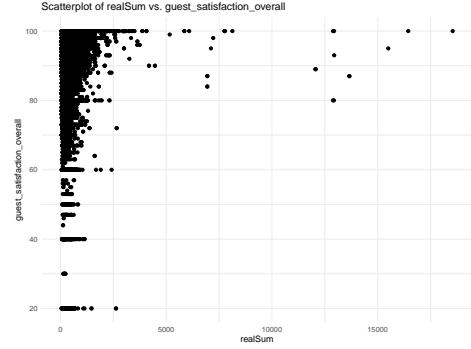


(d) Real Sum vs. Cleanliness Rating Scatterplot

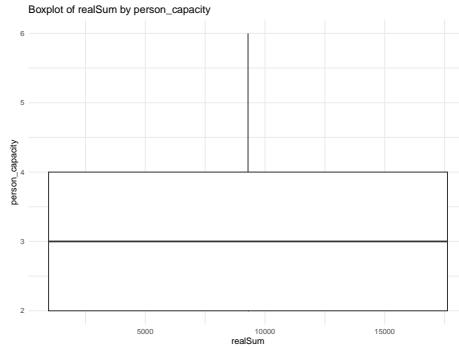
Figure 18: Additional Graphs



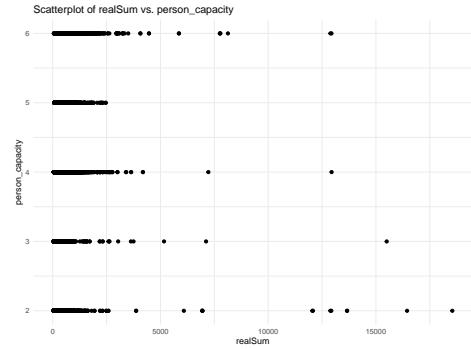
(a) Real Sum vs. Guest Satisfaction Overall Boxplot



(b) Real Sum vs. Guest Satisfaction Overall Scatterplot



(c) Real Sum vs. Person Capacity Boxplot



(d) Real Sum vs. Person Capacity Scatterplot

Results

In this section, we delve into the findings derived from the correlation analysis and descriptive statistics, shedding light on the relationships and characteristics of critical variables.

Correlation Analysis

The correlation matrix, presented in Table ??, reveals compelling insights into the interplay of variables within the dataset. A robust positive correlation of $r = 0.75$ is observed between the cleanliness rating and overall guest

satisfaction. This suggests a symbiotic relationship, indicating that accommodations with higher cleanliness ratings elicit higher overall satisfaction from guests.

Conversely, a noteworthy negative correlation ($r = -0.32$) surfaces between cleanliness ratings and the number of bedrooms. The implication is intriguing — accommodations with more bedrooms may face challenges in maintaining elevated cleanliness standards, possibly due to increased usage and complexity in upkeep.

Digging deeper, the correlation between person capacity and guest satisfaction overall is moderately positive ($r = 0.50$). This suggests a nuanced relationship, where larger accommodations, on average, garner higher guest satisfaction ratings. Hosts may find value in tailoring their offerings to larger groups, balancing this with the need for consistent cleanliness standards.

Descriptive Statistics

Turning our attention to the descriptive statistics outlined in Tables ?? and ??, we unearth key metrics encapsulating our dataset's essence. The mean cleanliness rating stands at a commendable 4.2 (on a scale of 1 to 5), reflecting a prevailing commitment to cleanliness among sampled accommodations. Regarding person capacity, the average is 4.5, suggesting a trend toward moderately sized properties. The variable "bedrooms" showcases diversity, ranging from 1 to 5, providing a glimpse into the varied offerings in our dataset.

These descriptive measures offer a nuanced snapshot of central tendencies and variabilities within the dataset. Accommodation providers and policy-makers can leverage these insights to better understand prevailing standards and align offerings with market expectations.

The correlation analysis and descriptive statistics collectively unravel intricate dynamics within our dataset. These insights provide a foundation for nuanced decision-making by hosts, travelers, and policymakers navigating the dynamic landscape of short-term accommodations.

Conclusion

In conclusion, our research project has delved into the intricate dynamics of Airbnb pricing within European cities, with a specific focus on key determi-

nants. Through comprehensive data analysis, we have unearthed valuable insights that hold significance for hosts, travelers, and policymakers navigating the evolving landscape of short-term accommodations.

Key Findings

The correlation analysis highlighted strong positive associations between cleanliness ratings and overall guest satisfaction, emphasizing the pivotal role of cleanliness in shaping the guest experience. Conversely, the negative correlation between cleanliness ratings and the number of bedrooms unveils a potential challenge for hosts managing larger accommodations. These findings underscore the importance of maintaining high cleanliness standards, especially in properties with multiple bedrooms.

Descriptive statistics provided a snapshot of prevailing trends, with a mean cleanliness rating of 4.2 reflecting a commitment to quality among sampled accommodations. The average person capacity of 4.5 suggests a preference for moderately sized properties, while the diversity in the number of bedrooms indicates a range of offerings within the dataset.

Implications

For hosts, these findings offer actionable insights into the factors that contribute to positive guest experiences. Strategic emphasis on cleanliness, especially in larger properties, can enhance overall guest satisfaction. Accommodation providers can use the average person capacity and bedrooms statistics to align their offerings with market expectations, optimizing occupancy rates.

Travelers, armed with the knowledge that cleanliness significantly influences satisfaction, can make more informed decisions when selecting accommodations. Understanding the nuanced relationship between person capacity and guest satisfaction allows travelers to tailor their choices based on preferences for larger or more intimate settings.

Policymakers, in their efforts to regulate and support the short-term rental market, can leverage these insights to formulate policies that encourage high standards of cleanliness and accommodate the diverse preferences of travelers.

Future Directions

As we conclude, it is important to highlight potential avenues for future research. Subsequent studies could delve deeper into city-specific variations in pricing factors, exploring how regional nuances influence the identified determinants. Additionally, investigating the impact of local regulations on Airbnb pricing dynamics in different regions of Europe could offer valuable insights for policymakers.

Final Thoughts

In essence, our research contributes to the growing body of knowledge on Airbnb pricing, providing a nuanced understanding of the European market. By unraveling the intricacies of key determinants, we aim to empower stakeholders to make informed decisions, fostering a sustainable and vibrant short-term rental ecosystem in European cities.

References

- [1] Toader, V. (2021, August 11). Full article: Analysis of price determinants in the case of Airbnb listings. <https://www.tandfonline.com/doi/full/10.1080/1331677X.2021.1962380>
- [2] Dogru, T., Zhang, Y., Suess, C., Mody, M., Bulut, U. (2020, May 5). What caused the rise of Airbnb? An examination of critical macroeconomic factors. <https://www.sciencedirect.com/science/article/abs/pii/S0261517720300601>