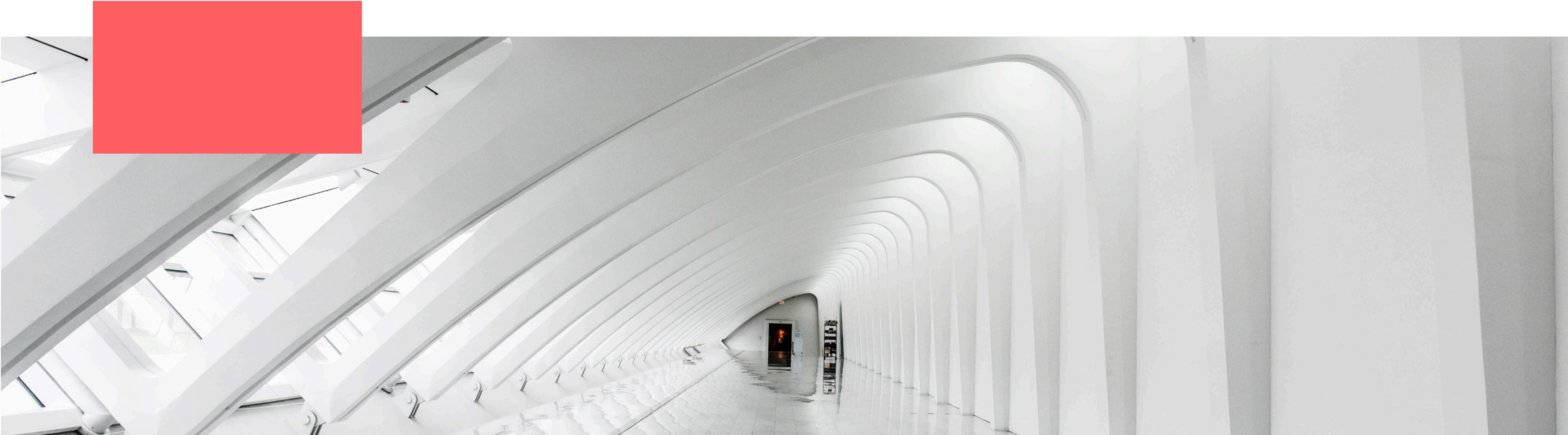




December 12, 2024

Property Revenue Analysis

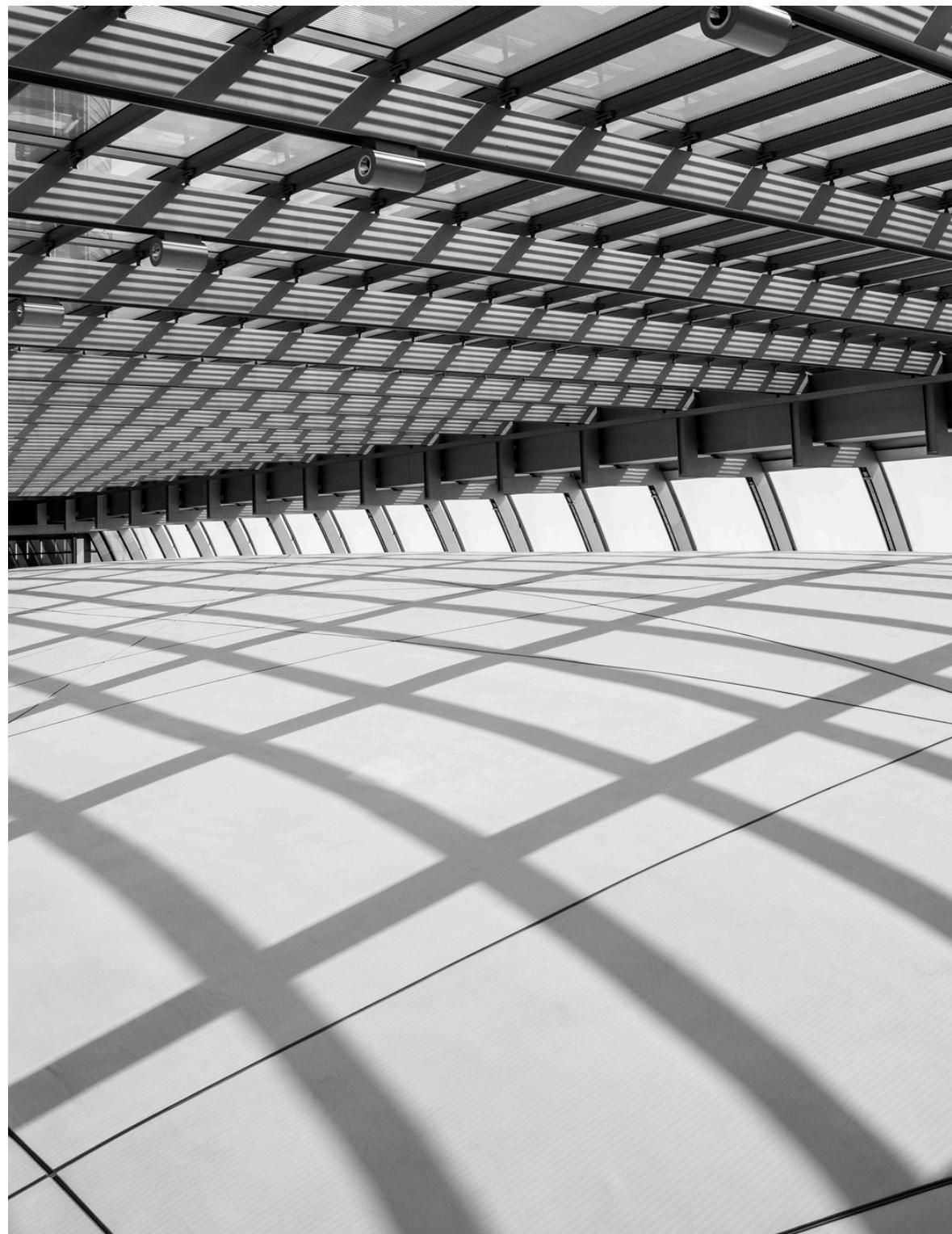
Jeffrey Fernandez, Jay Chang, & Jack Keaveny



Our Objective

To identify the **top 5 most influential property features** driving **Airbnb revenue** in **Spain** in order to optimize property management and maximize revenue.





Action Plan

1. Determine the **best city** to invest in within Spain
2. Identify the **best property types** to invest in
3. Determine what **property attributes** influence revenue the most (top 5)
4. Create **models** that predict which variables influence revenue the most



Exploratory Data Analysis (EDA)

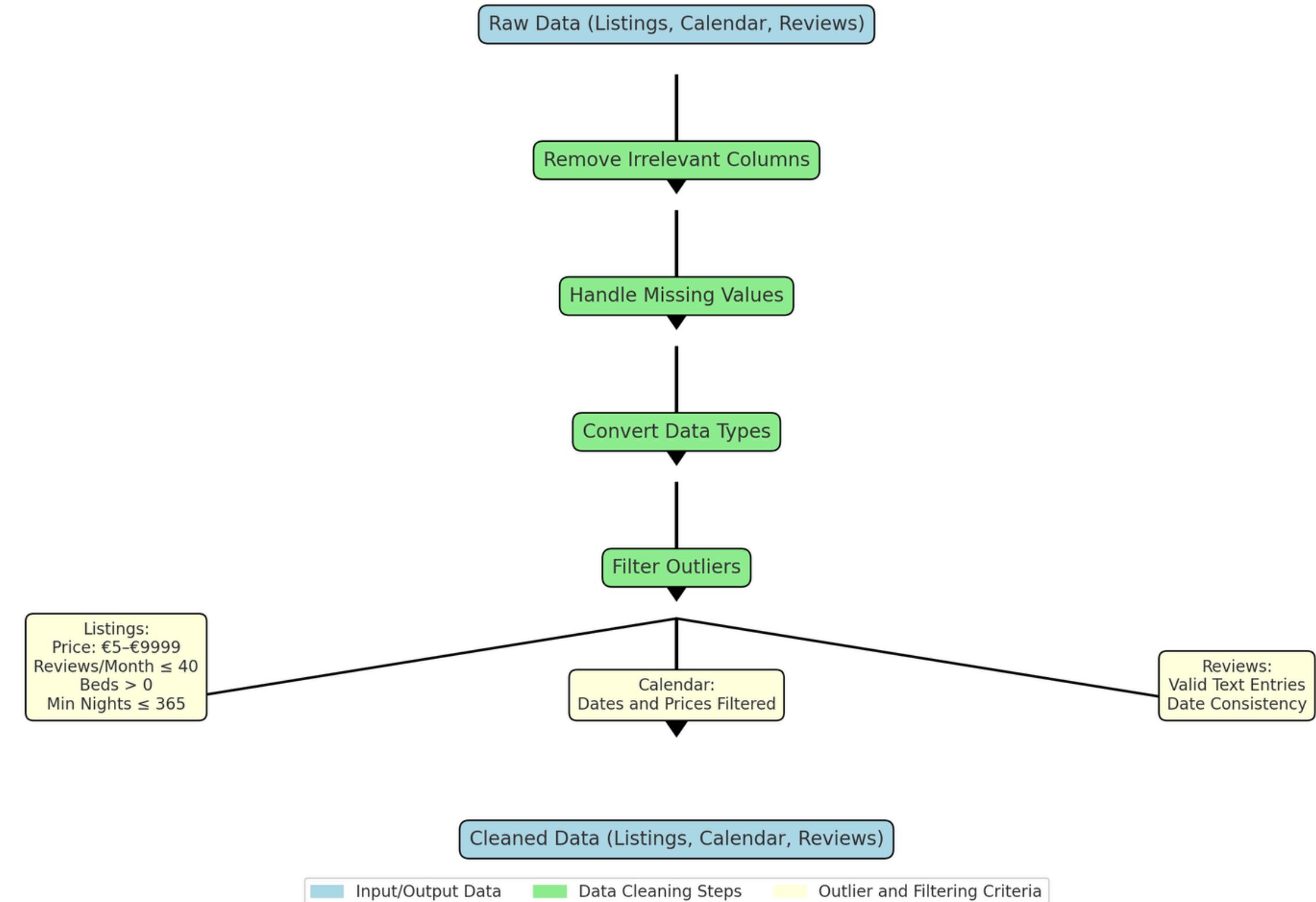
Data Loading

Name	↑
Barcelona	
Madrid	
Malaga	
Mallorca	
Menorca	

Name	↑
calendar.csv.gz	👤
listings.csv	👤
listings.csv.gz	👤
neighbourhoods.csv	👤
neighbourhoods.geojson	👤
reviews.csv	👤
reviews.csv.gz	👤

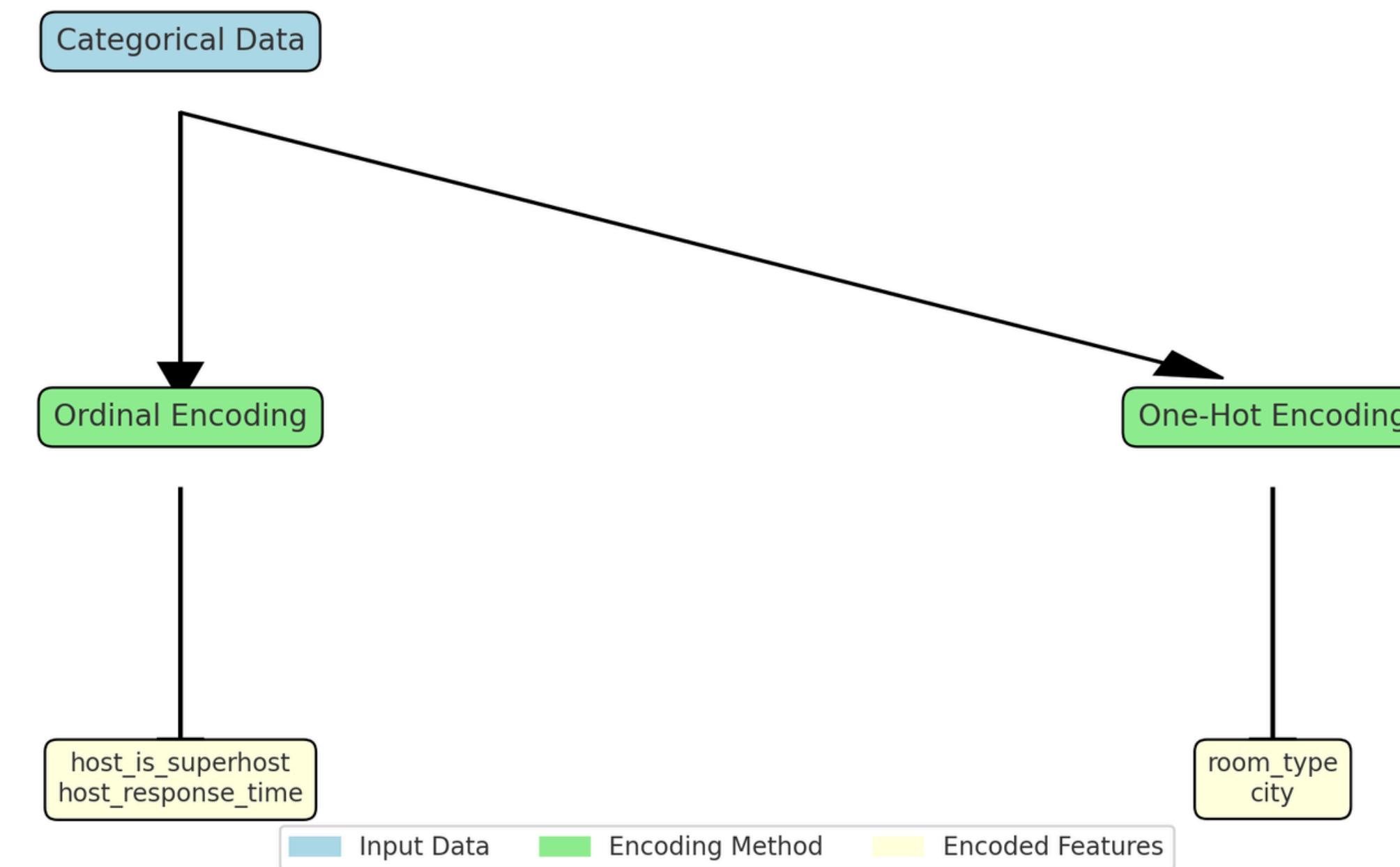
- Iterated through folders containing city-specific .csv.gz files.
- Created a dictionary (dataframes) to store listings, reviews, and calendar data for each city in Spain.
- Ensured proper error handling for missing files, parsing issues, and non-.csv.gz files.

Data Cleaning



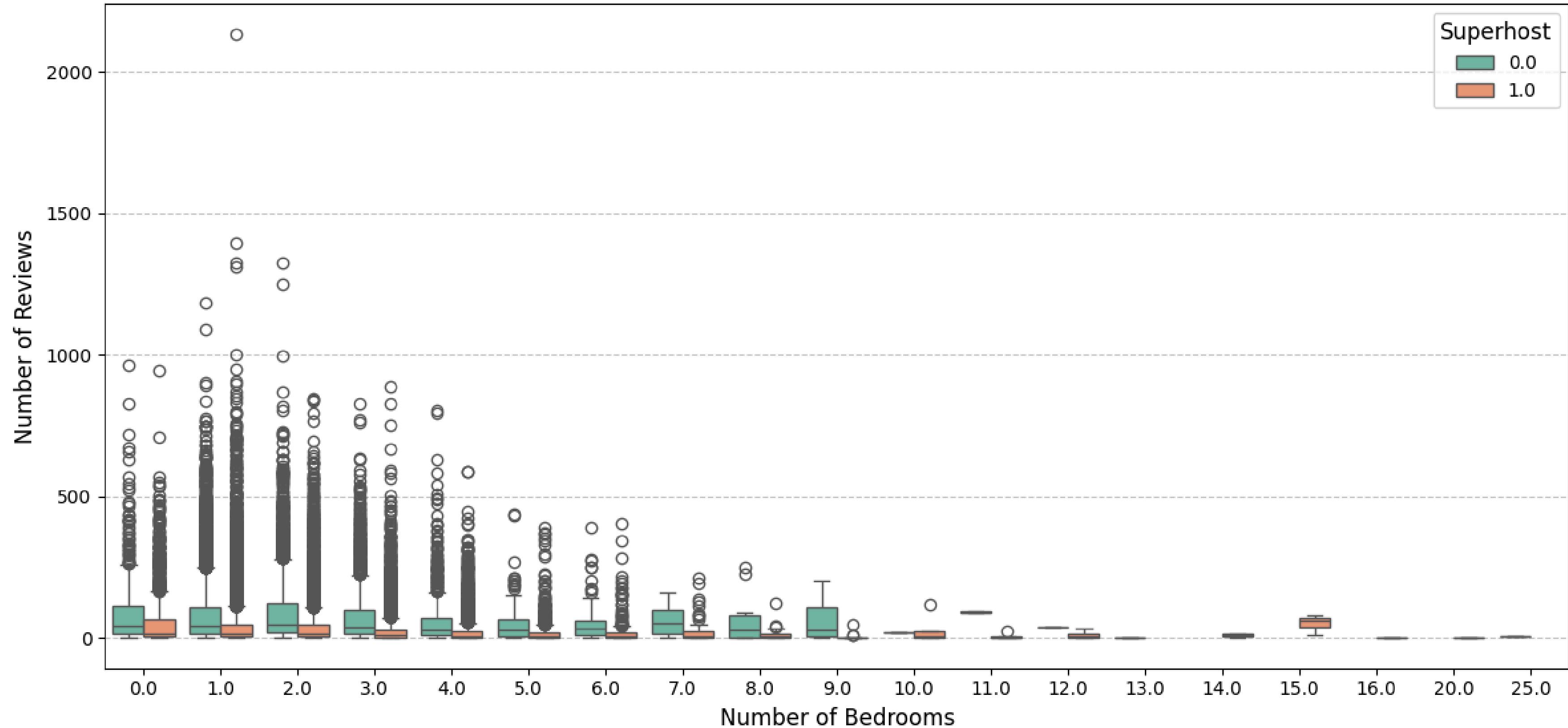
Categorical Variable Encoding

`listings2`: DataFrame of all cleaned listings data



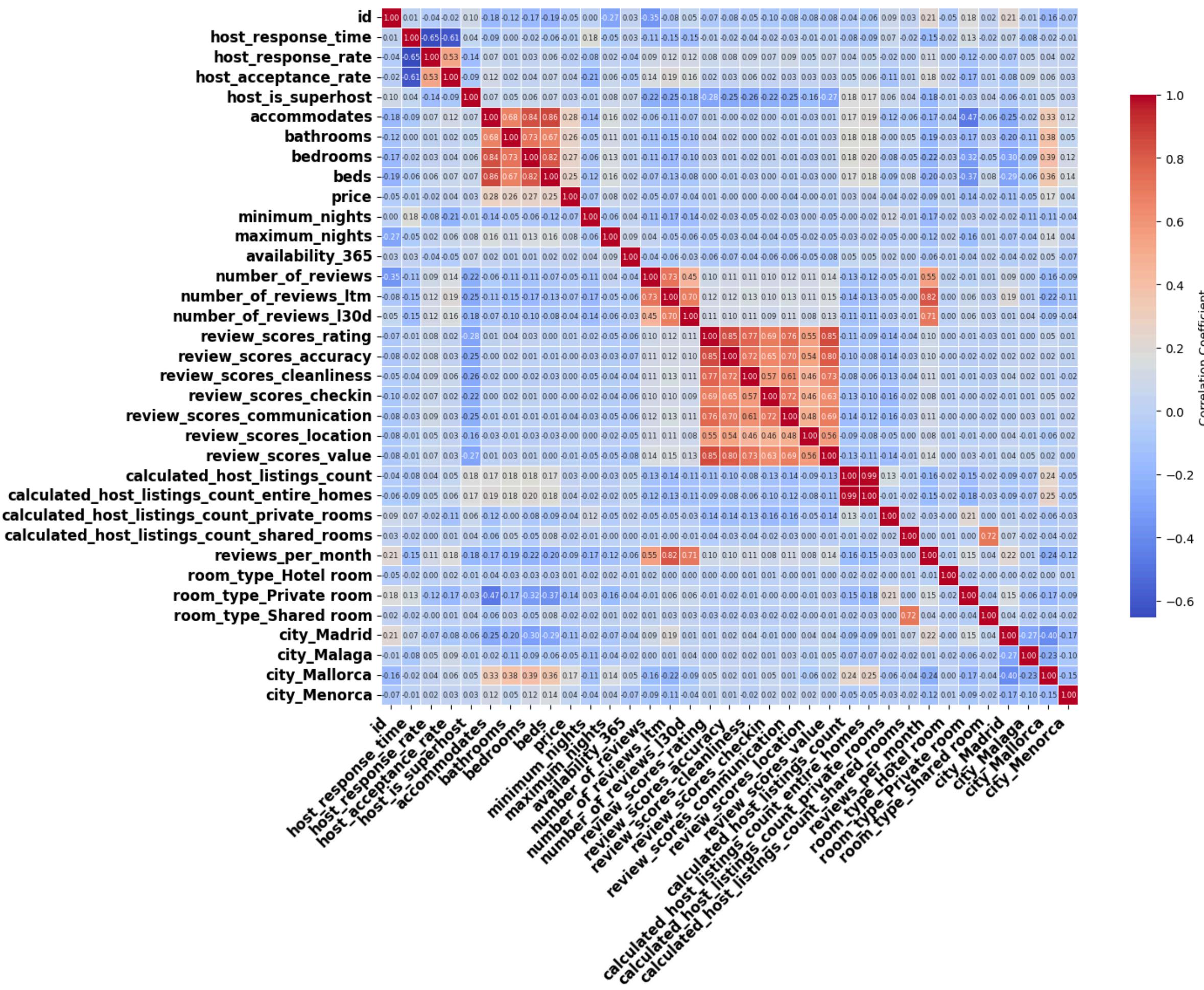


Distribution of Review Numbers by Bedrooms and Superhost Status



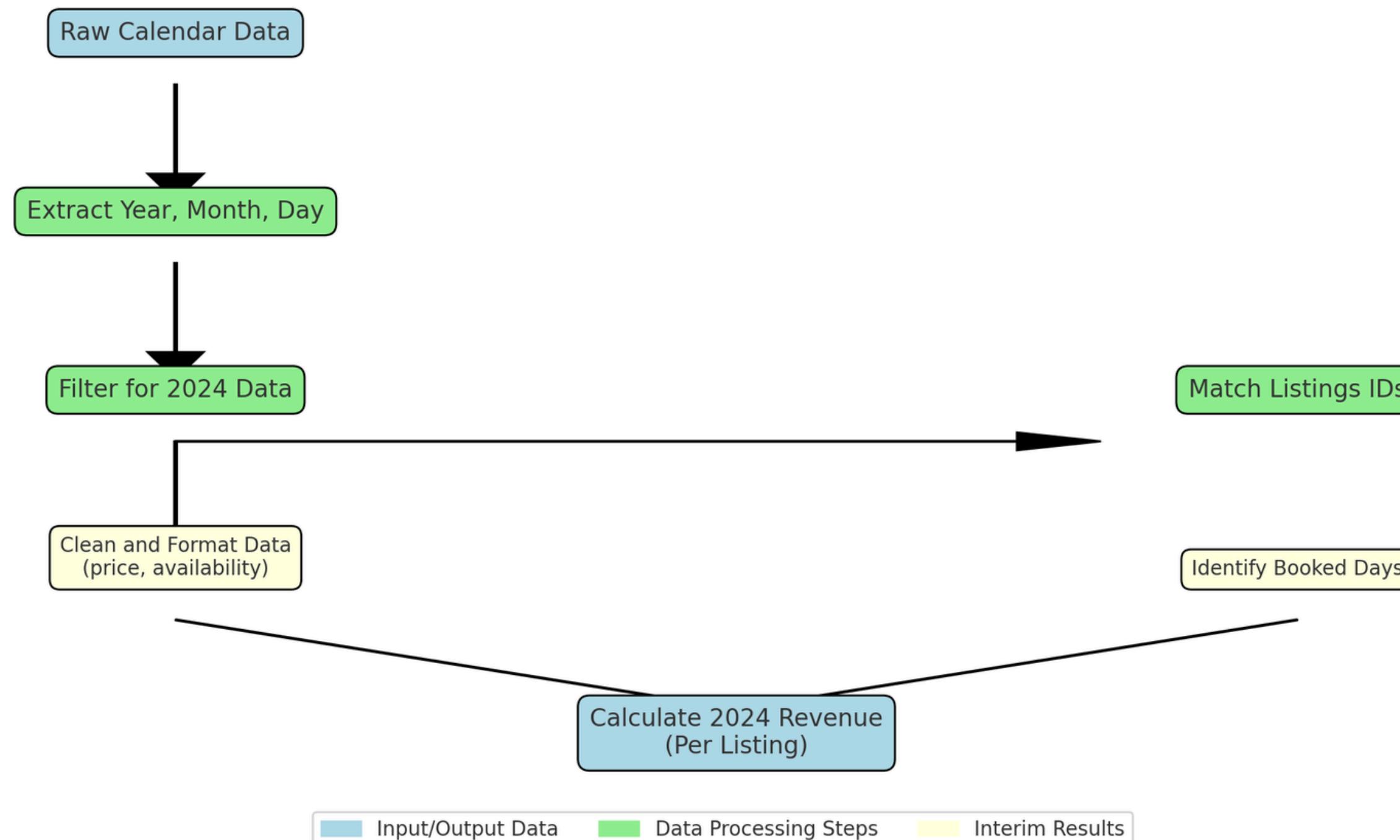
Correlation Heatmap of Numerical Attributes

December 12, 2024



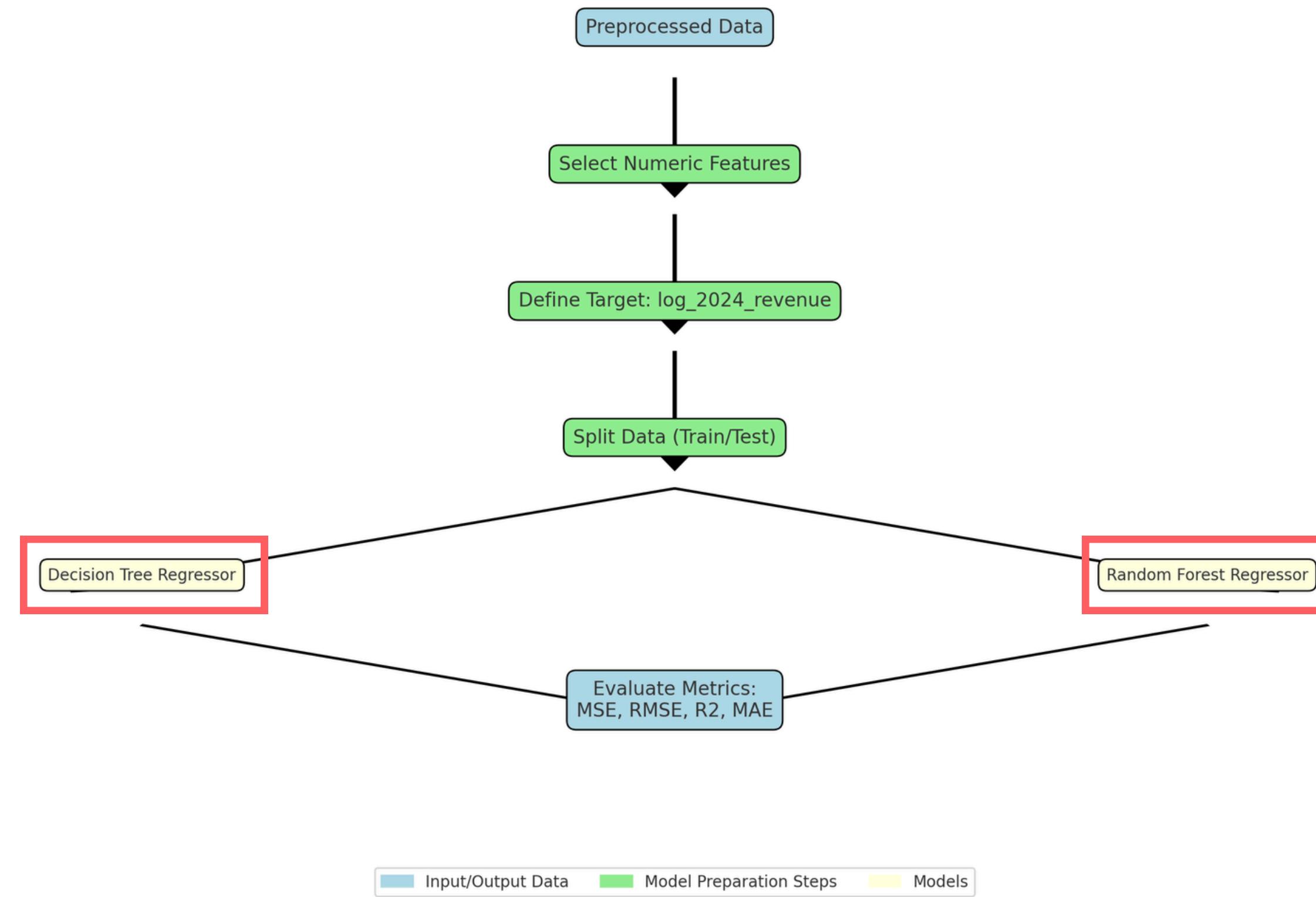
Revenue Calculations

calendar2: DataFrame of all cleaned calendar data



Model and Feature Selection

Models



Feature Selection:

- Selected numeric predictors relevant to the target (`log_2024_revenue`) while excluding identifiers and highly correlated features like `price` and `availability_365`.
- Ensured features are scaled appropriately for the model requirements.

Model Selection:

- Decision Tree Regressor:**
 - Simplicity in interpreting splits and feature importance.
 - Risk of overfitting if the tree grows too deep.
- Random Forest Regressor:**
 - Aggregates multiple decision trees to reduce overfitting and improve generalization.
 - Robust against noisy data with better predictive performance.

Model Results



Model Selection for Analysis



Random Forest vs. Decision Tree Performance:

Random Forest significantly outperforms the Decision Tree

- Higher R-squared: 0.52 vs. 0.01.
- Lower Mean Squared Error (MSE) and Mean Absolute Error (MAE), indicating better predictive accuracy.
- Stronger ability to capture underlying patterns in the data.

Interpretability:

Decision Tree offers simplicity and transparency

- Useful for analyzing feature importance and identifying key revenue drivers.
- Insights into how features influence predictions.

Conclusion:

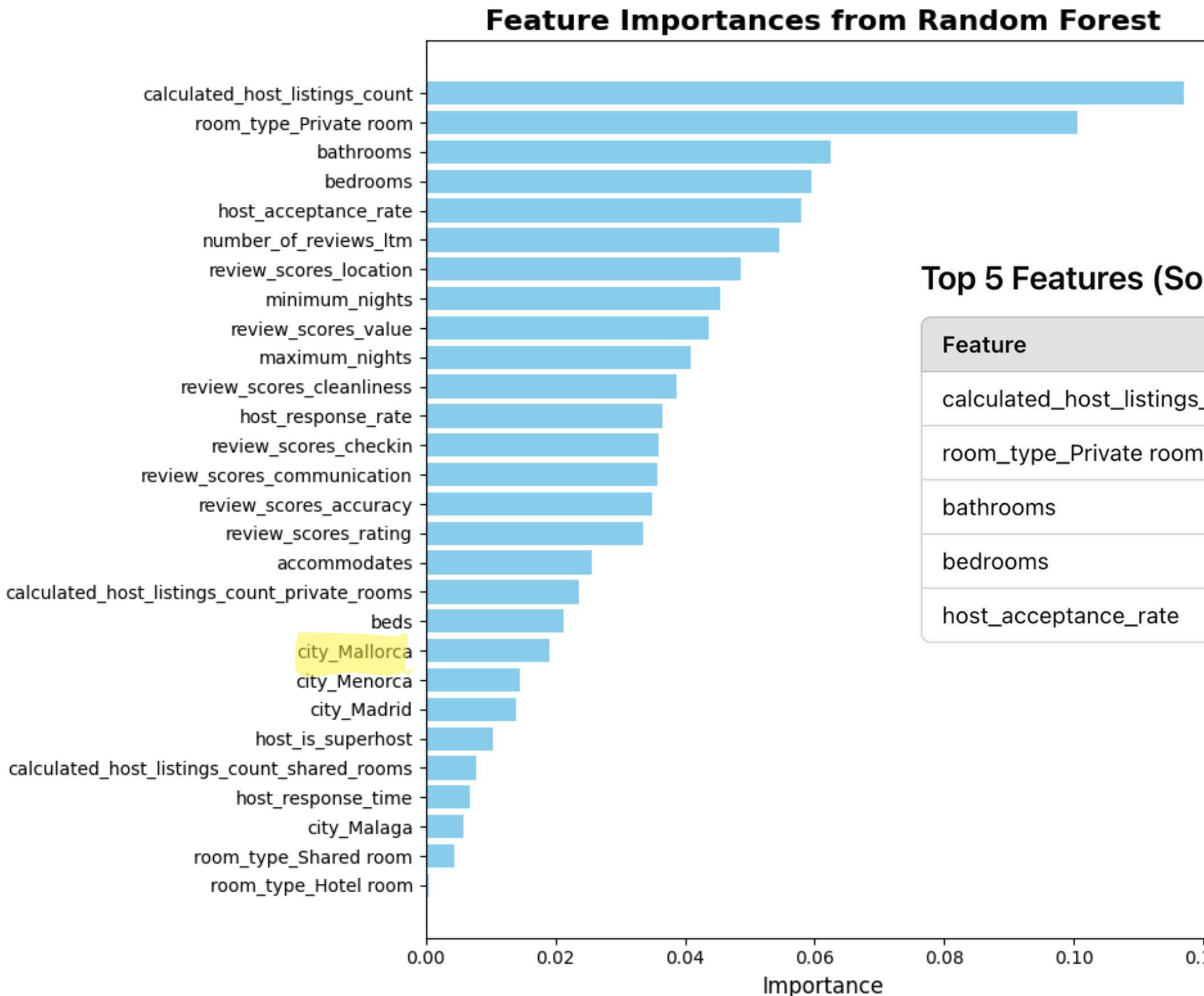
Random Forest for accuracy.

Decision Tree for interpretability and strategic insights.

Analysis

Feature Analysis - Random Forest

Features

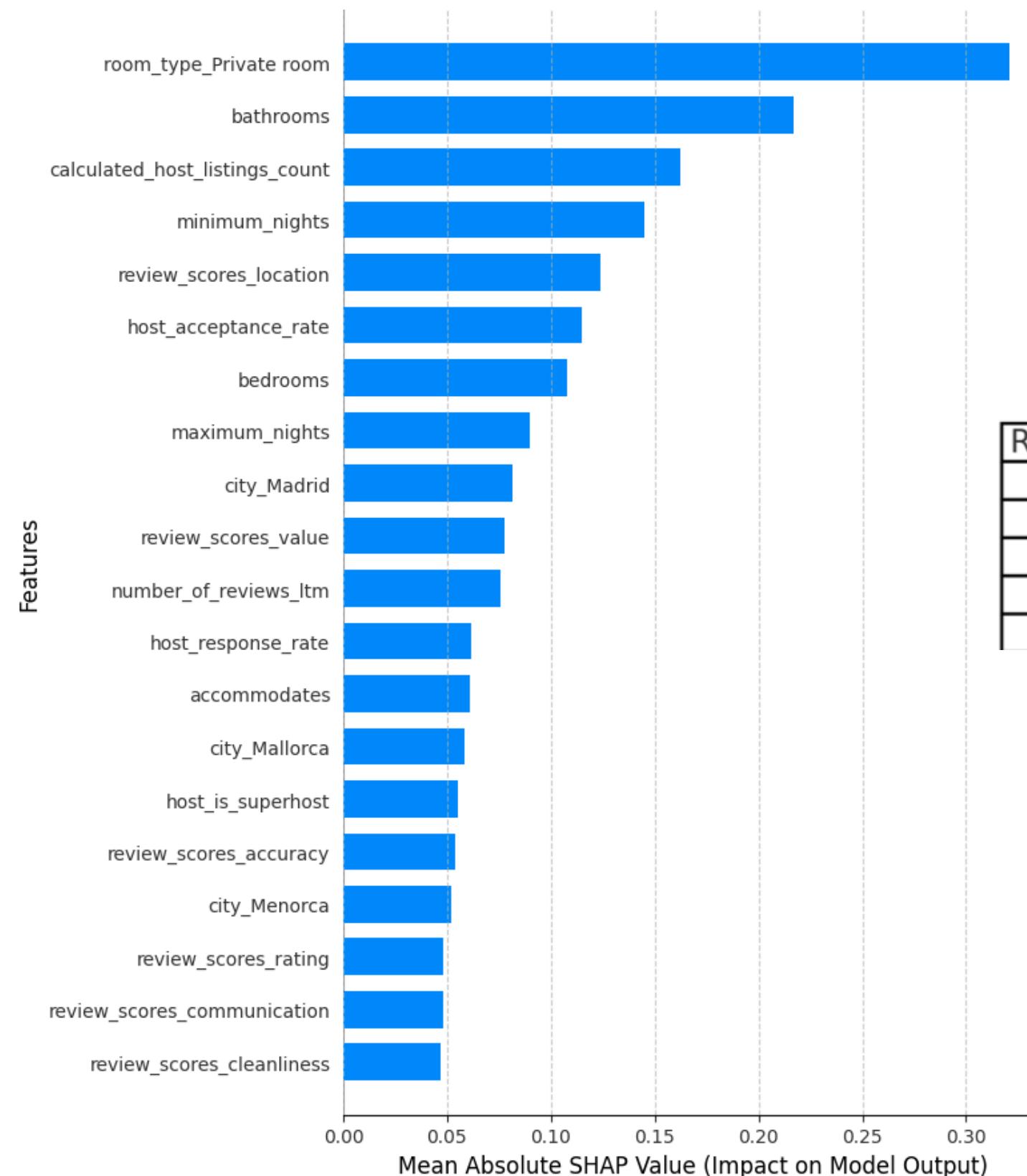


Top 5 Features (Sorted by Importance)

Feature	Importance
calculated_host_listings_count	0.116888
room_type_Private room	0.100459
bathrooms	0.062410
bedrooms	0.059465
host_acceptance_rate	0.057918

Feature Analysis - Decision Tree

Decision Tree Model Feature Importance: SHAP Summary (Bar Plot)

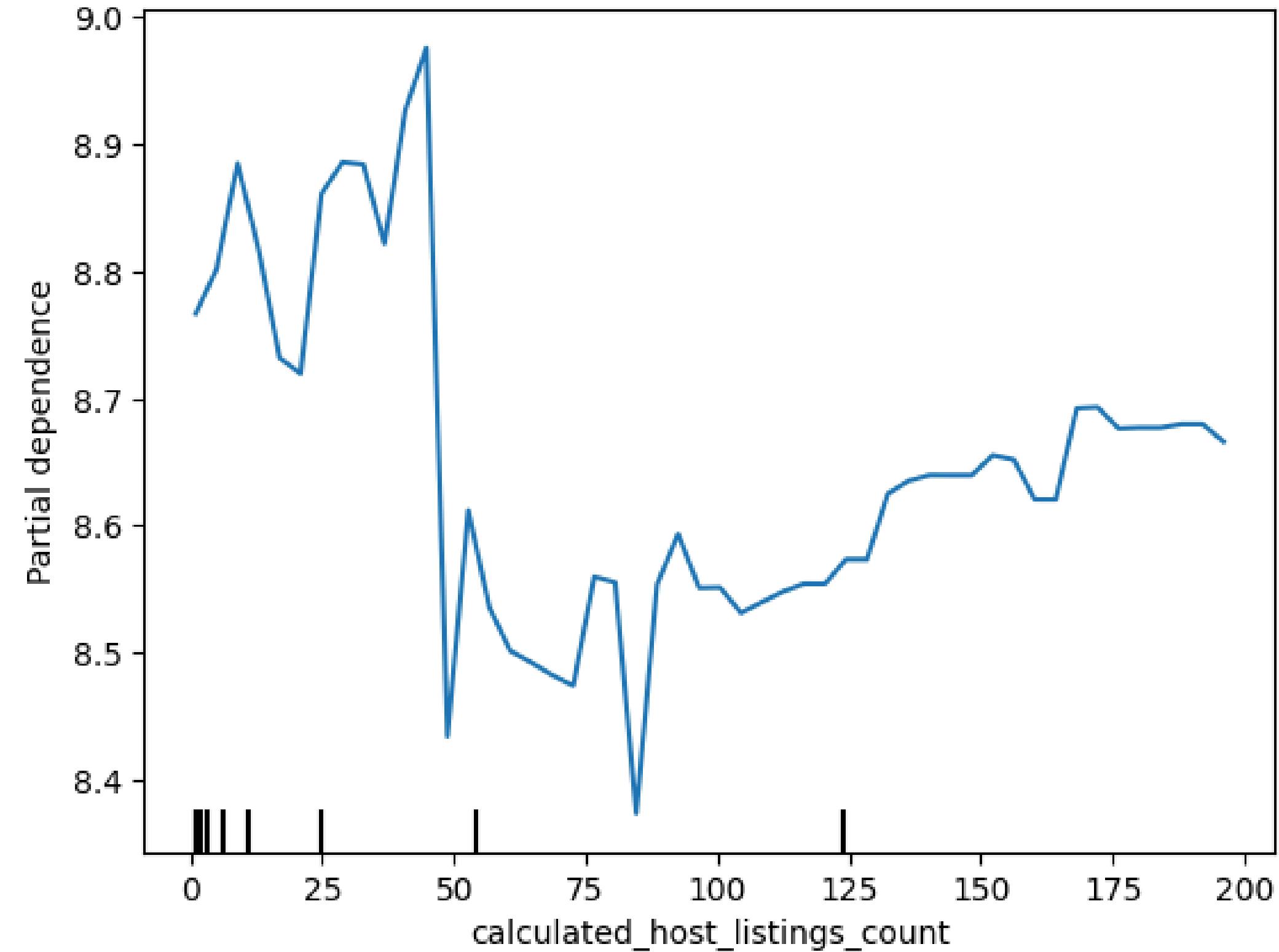


Rank	Feature	Description
1	room_type_Private room	Indicates the room type is a private room.
2	bathrooms	Number of bathrooms in the property.
3	calculated_host_listings_count	Total number of listings a host manages.
4	minimum_nights	Minimum number of nights required for a booking.
5	review_scores_location	Guest rating of the property's location.

Partial Dependence Plot

calculated_host_listings_count

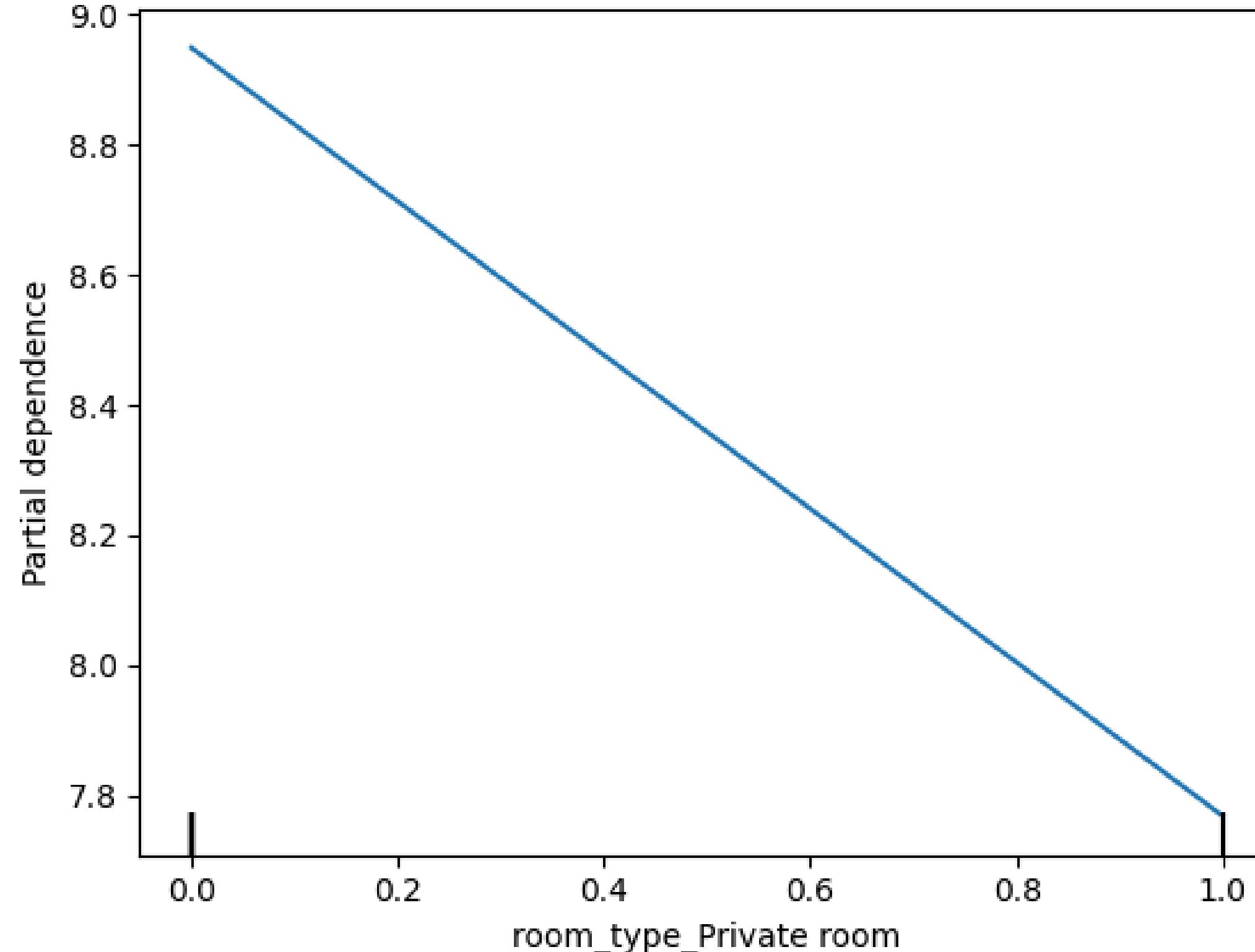
- Complex non-linear relationship; confounding factors
- Moderate number of listings might be associated with higher revenue
- Excessive number of listings can lead to fluctuations and even a decline
- Large portfolio requires careful resource allocation and may face challenges related to maintaining high-quality service and managing competition



Partial Dependence Plot

room_type_Private room

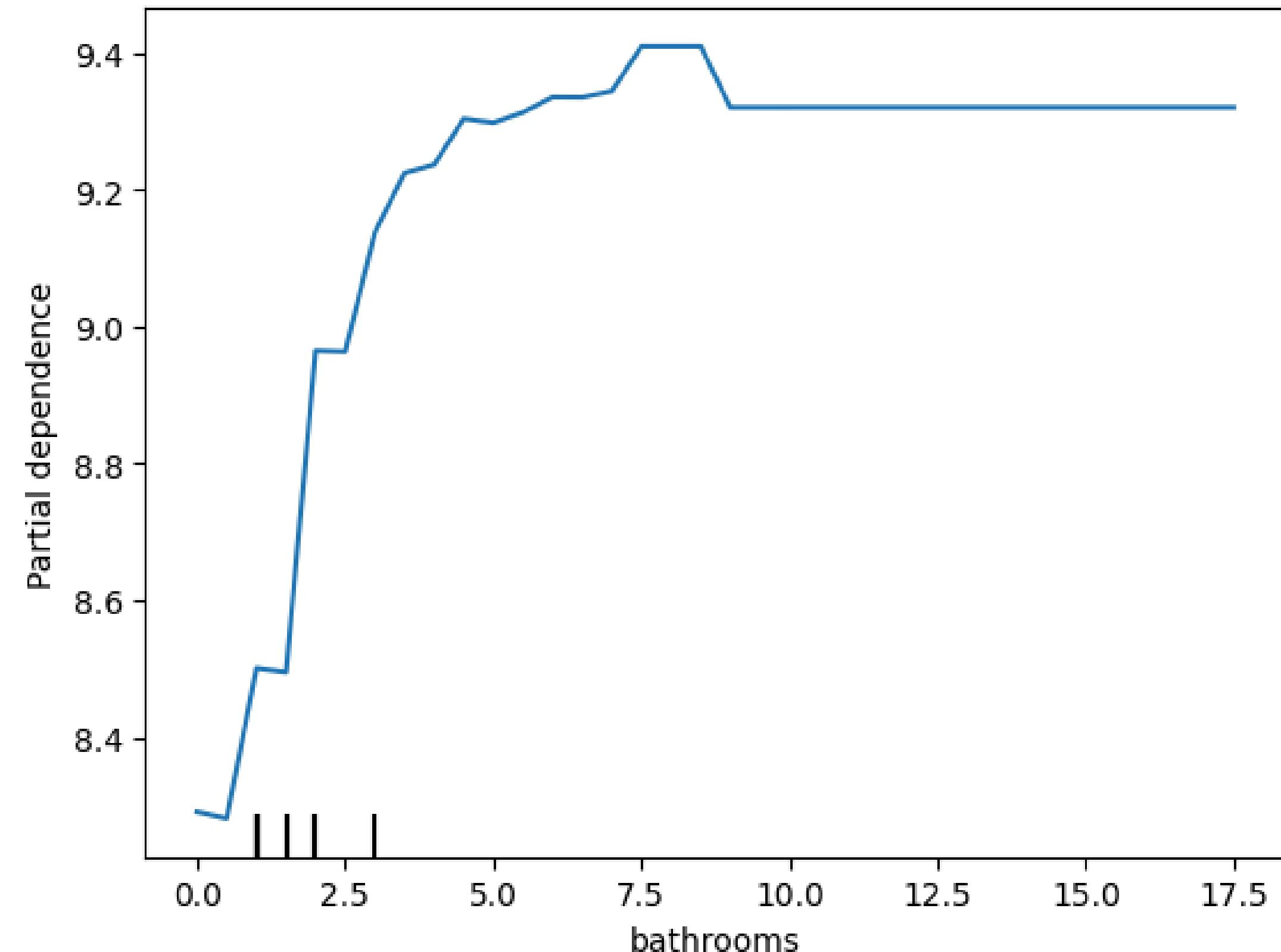
- Negative association
- Properties listed as "private rooms" might be associated with **lower revenue** or less favorable outcomes compared to other room types (eg, entire homes)



Partial Dependence Plot

bathrooms

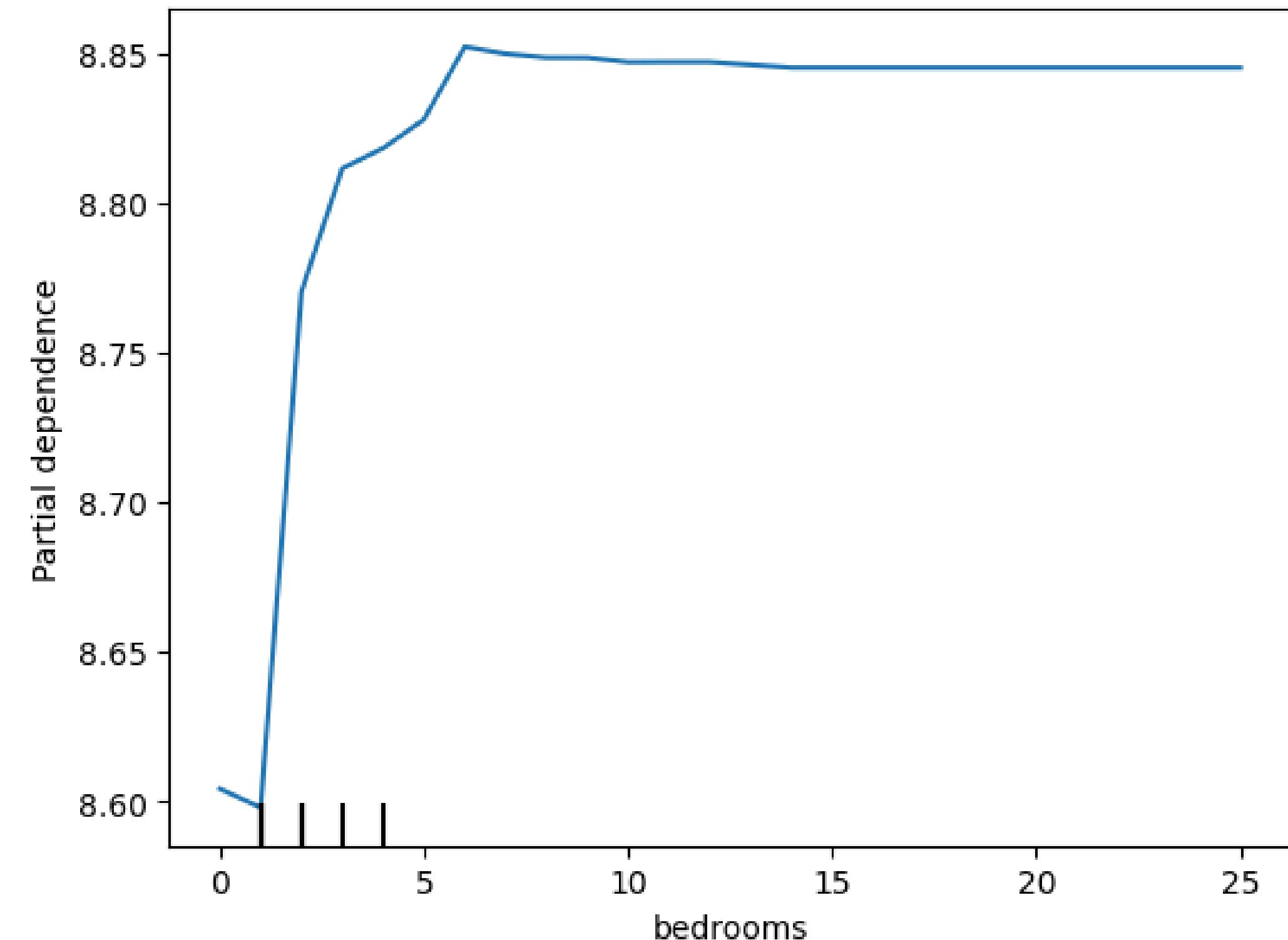
- Positive association
- Properties with more bathrooms tend to have **higher predicted revenue**
- Diminishing returns, with the initial increase in revenue being more pronounced for the first few bathrooms



Partial Dependence Plot

bedrooms

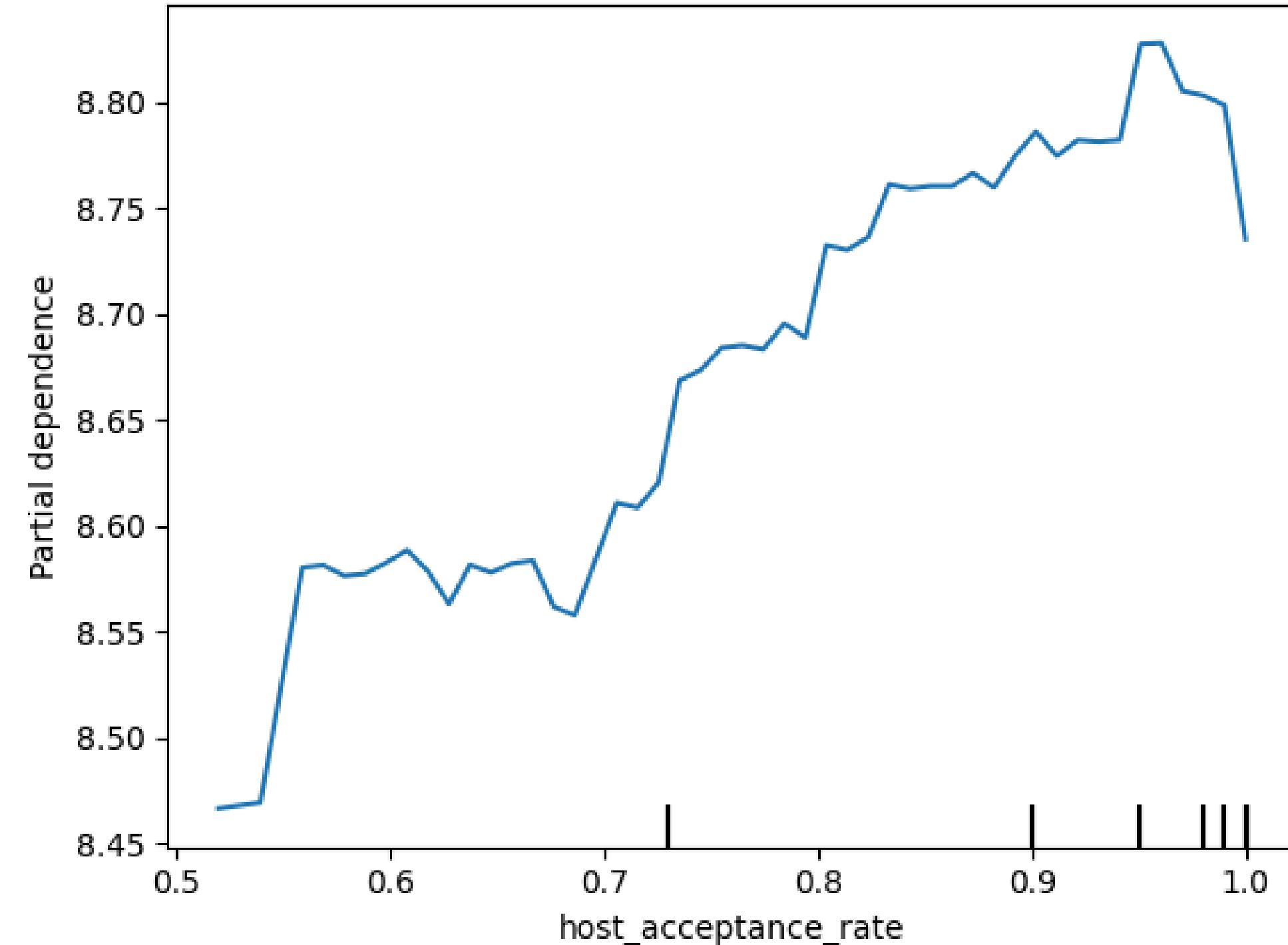
- Positive association
- Properties with more bedrooms tend to have **higher predicted revenue**
- Diminishing returns, with the initial increase in revenue being more pronounced for the first few bedrooms



Partial Dependence Plot

host_acceptance_rate

- Positive association
- Properties with higher host_acceptance_rate tend to have **higher predicted revenue**
- Diminishing returns over time, likely due to potential scams by hosts or other confounding factors

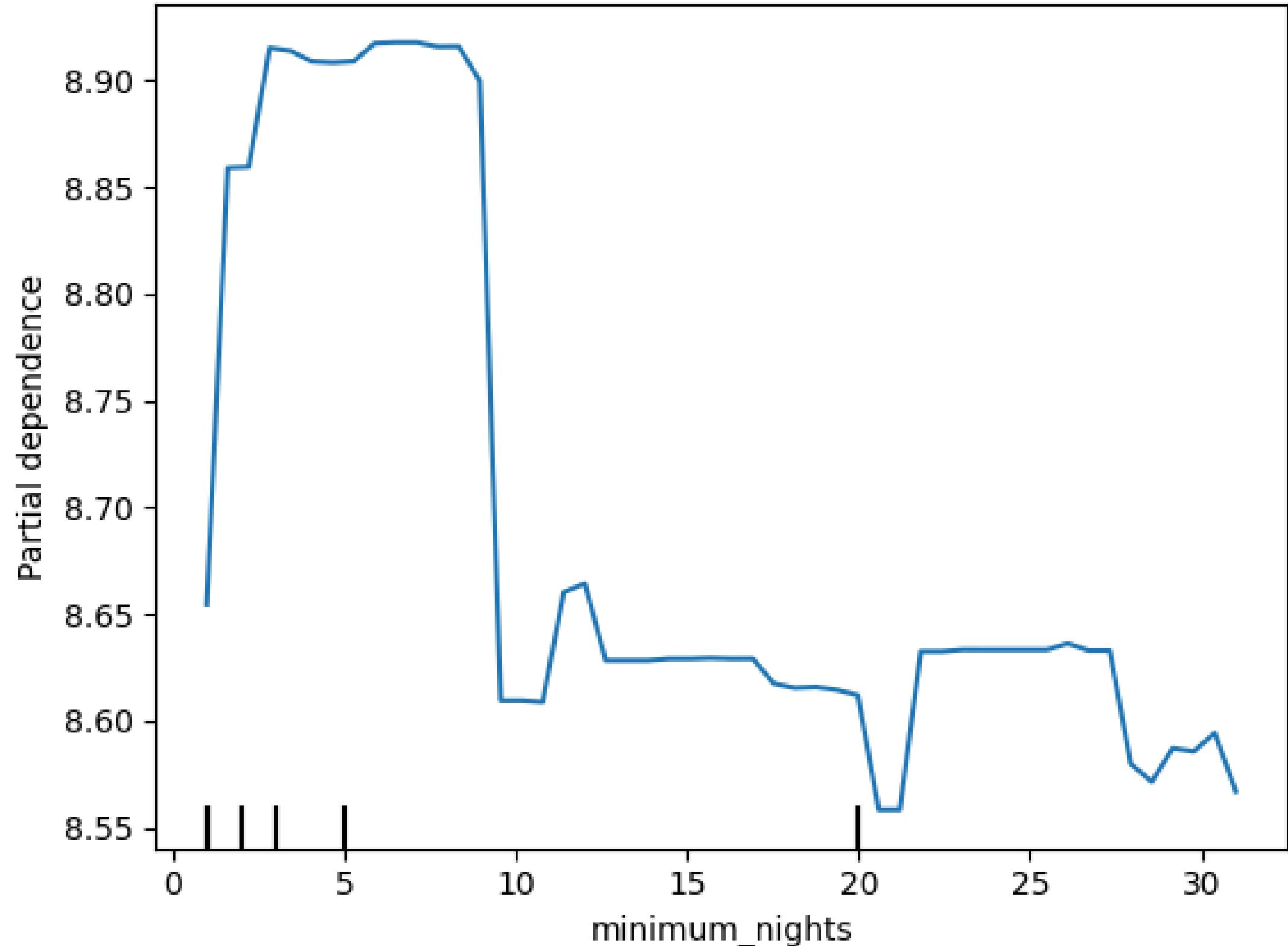


Partial Dependence Plot

December 12, 2024

EXTRA: minimum_nights

- Complex non-linear relationship
- Moderate minimum stay might be associated with higher revenue
- Excessive minimum stay can lead to fluctuations and even a decline

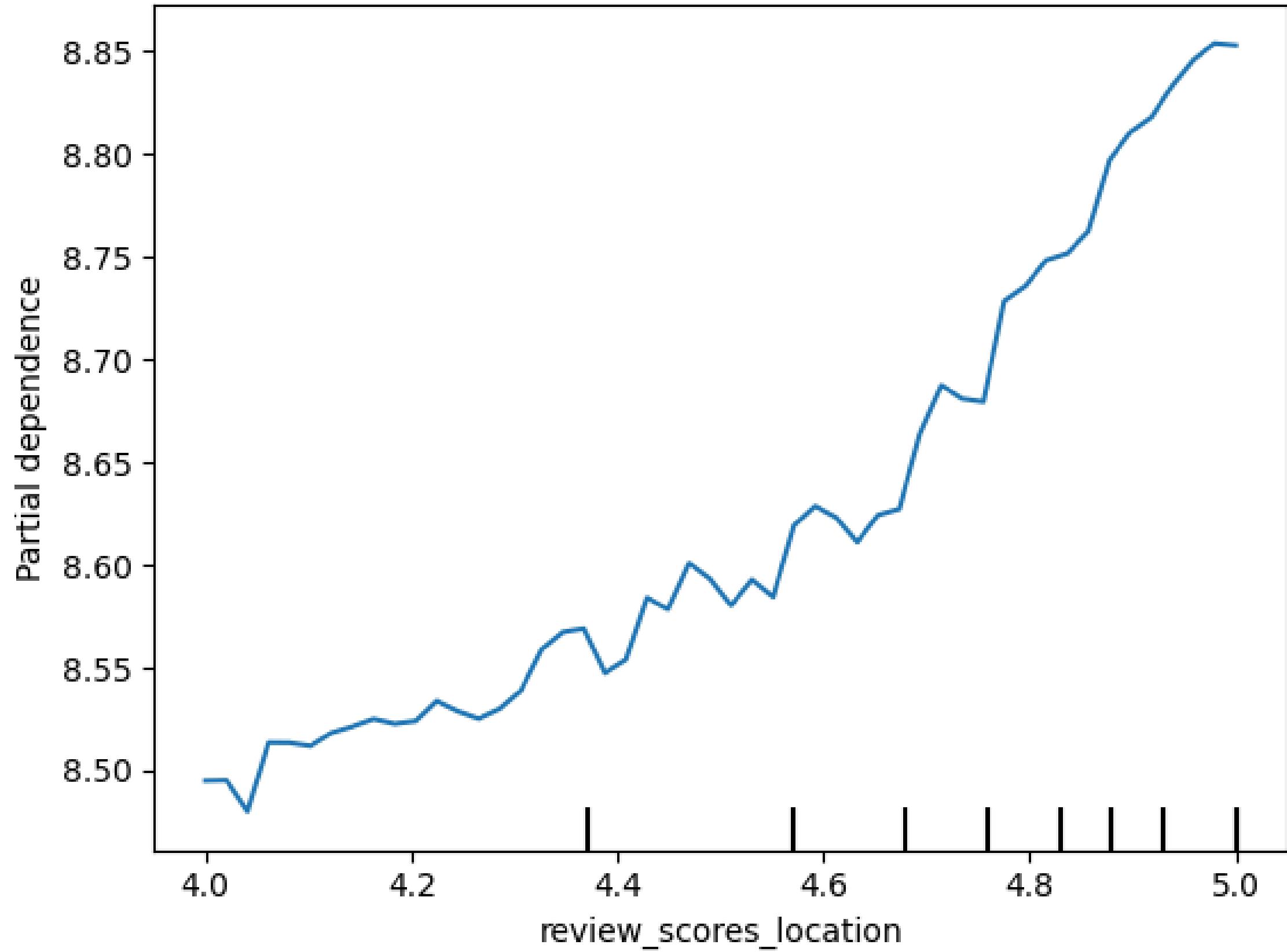


Partial Dependence Plot

December 12, 2024

EXTRA: review_scores_location

- Strong, positive relationship
- Higher location scores tend to have **higher predicted revenue**
- Guests value well-located properties



Conclusion

Conclusion

For investors to maximize Airbnb revenue, these findings should be used to refine property features, pricing strategies, and listing descriptions to maximize revenue.

Properties, Regulations

Mallorca

Ideal City

Entire house or apartment is best

Private Room

Worst Room Type

Best # Bedrooms

3-5 Bedrooms

House or Large Apt.

The more, the better

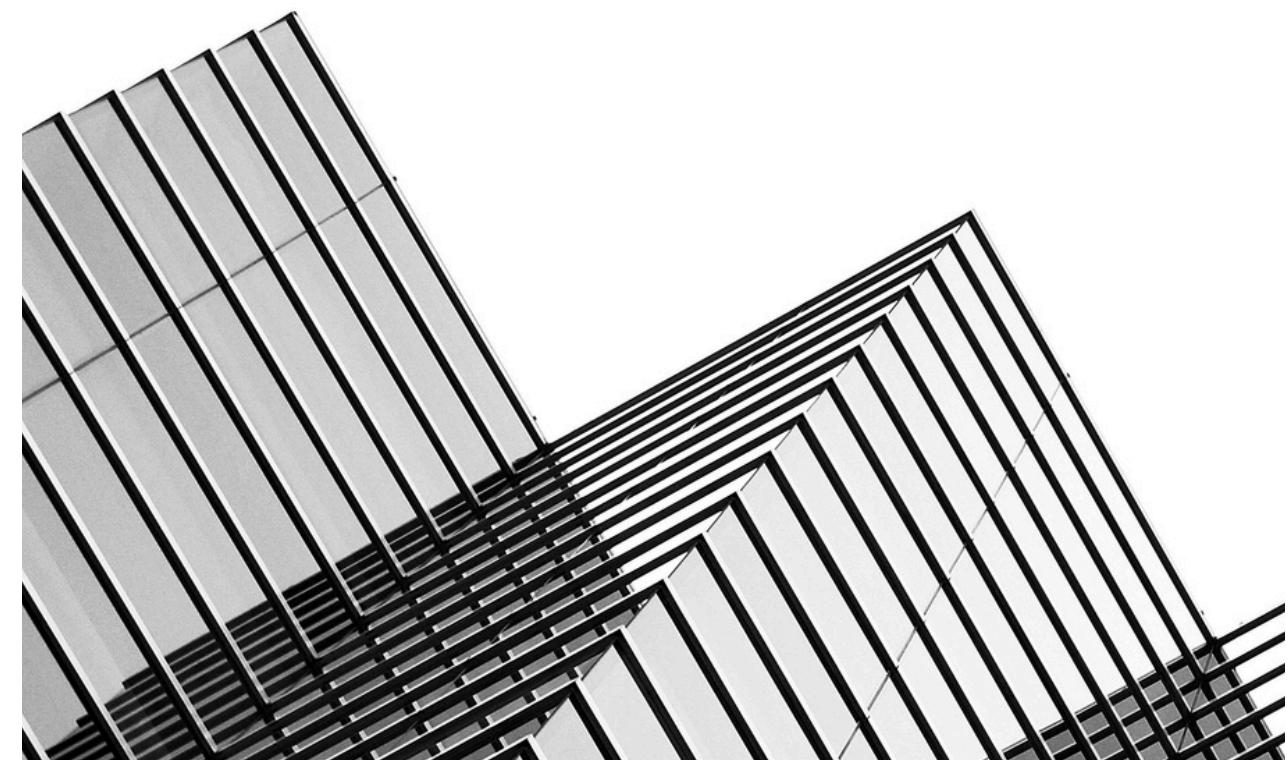
↑ Bathrooms

Compensate for Room Type

Attract More Bookings

Moderate

Minimum Stay Policy
& Portfolio Size



Next Steps



Cost Analysis

Integrate property price data from Zillow to understand net profit from purchase costs and rental revenue potential.

Sentiment Analysis

Extract customer sentiment from reviews to identify high-impact amenities. Highlight amenities most frequently associated with positive reviews (cleanliness, Wi-Fi quality, etc.)

Review Analysis

Analyze the trend of review volumes (rising or falling) over the past 12-24 months. Identify property types with consistently improving or declining reviews.

Geospatial Analysis

Map Airbnb properties along major tourist attractions, transportation hubs, and nightlife spots to identify zones with high-performing properties (based on reviews, occupancy rates, and/or revenue).



Thank You!

Any Questions?

