

MSE - 491 Marketing Analytics

Airbnb Dataset Analysis Report

Pawan Seth: 22110192

Deepak Soni: 22110068

Vansh Kumar: 22110280

Garima Nama: 22110084



IIT Gandhinagar
September 24, 2024

Introduction

Airbnb is a short- and long-term letting platform between hosts and renters found globally. As an intermediary, the platform earns revenue by collecting commissions from bookings. Since its launch, Airbnb has expanded at warp speed to provide millions of listings.

The data size under research is over 21,371 Airbnb listings from 79 neighbourhoods in Belgium and the Netherlands. They are observed to analyse how host behaviour, location, and room type affect their rental prices and availability.

In this report, we analyse several vital areas. First, we will conduct a Room Analysis to examine the distribution of available rooms based on country, neighbourhood, and type. This will help us identify trends in room availability and preferences across different regions.

Next, we will perform a Price Analysis, comparing room prices based on location, type, and duration of stay. This will allow us to understand how these factors influence rental costs and guide guests in making informed choices.

We will also conduct an International Comparison to analyse room prices and availability across various countries. This comparison will reveal global trends in the short-term rental market and highlight how location impacts pricing and accessibility.

Lastly, we will explore Correlations and Regression Models to identify and quantify relationships among factors affecting room prices and availability. This will enable us to derive data-driven insights for better decision-making in the rental market.

Data Description

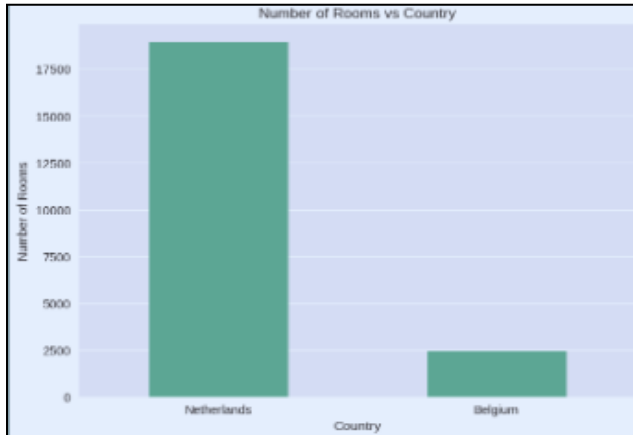
It has a total of 14 columns, and each column contains key information about each property, its host, pricing, customer feedback, etc.

| Column Name | About | Column Name | About |
|------------------|---|----------------------------------|--|
| A) ID | Unique identifier for each property | H) Room Type | Accommodation type. |
| B) Name | Property name | I) Price | Price per night |
| C) Host ID | A unique identifier for the host. | J) Minimum Nights | Minimum nights allowed for stay at this specific property. |
| D) Country | Dummy variable, 1 means Netherlands and 2 Means Belgium | K) No. of Reviews | Number of reviews on Air BnB for that property. |
| E) Neighbourhood | Locations of the rooms | L) Reviews Per Month | The average number of reviews per month received. |
| F) Latitude | Latitude of the property | M) Calculated Host Listing Count | Number of properties listed by the host on Airbnb. |
| G) Longitude | Longitude of the property | N) Availability 365 | Days the property was available last year. |

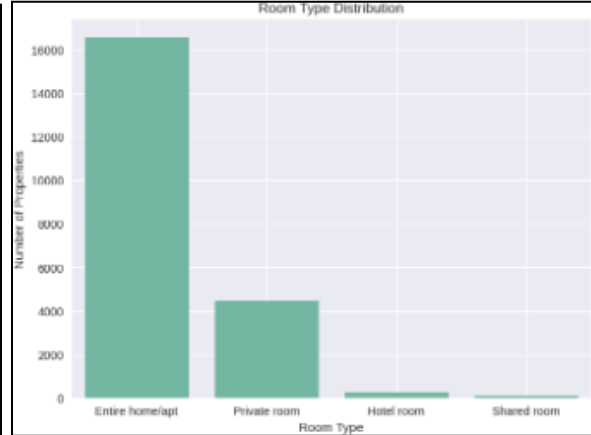
Analysis

1. Room analysis:

- Distribution of rooms



Number of Rooms in Both Countries

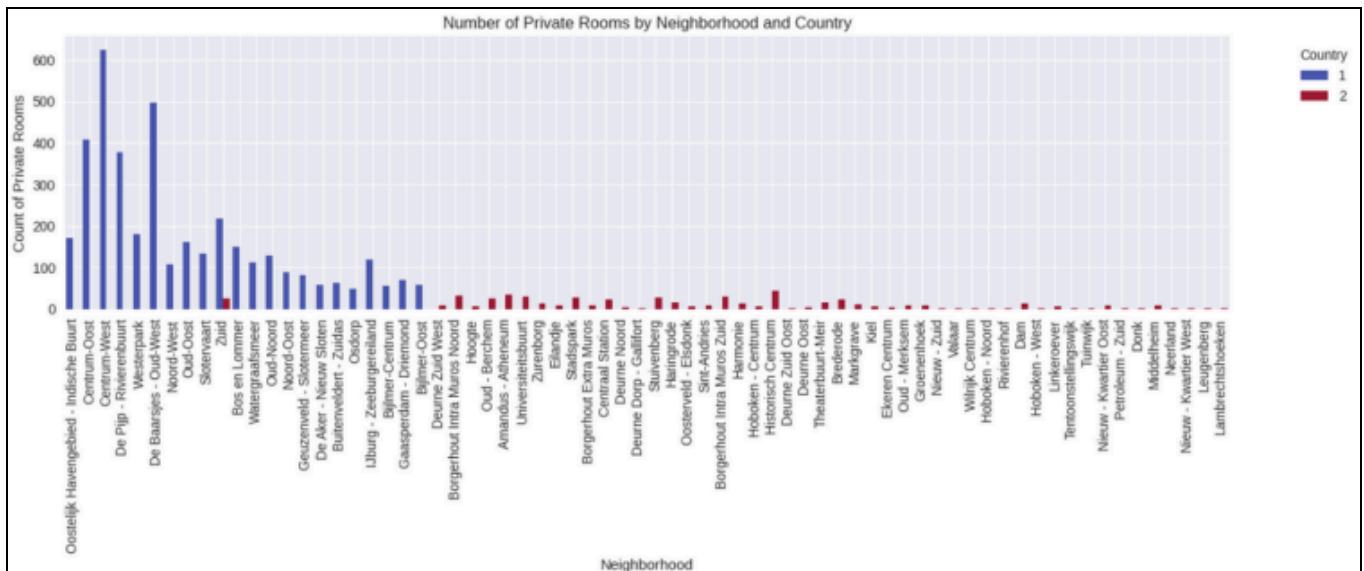


Room Type Distribution\

Entire Home/Apt - **16570**, Private Rooms - **4466**, Hotel Room - **240**, Shared Room - **95**

As we can see from the plots, the Netherlands has more rooms than Belgium. In both countries, the highest rent is for the entire property, followed by private rooms, hotel rooms, and shared rooms.

- Room Type Distribution:



Quantity of private rooms in the neighbourhood for both Countries

The Netherlands generally has higher prices than Belgium, likely due to more tourist-heavy cities like Amsterdam, where demand exceeds supply. Additionally, entire homes/apartments are pricier than private or shared rooms, reflecting the greater privacy and space they offer.

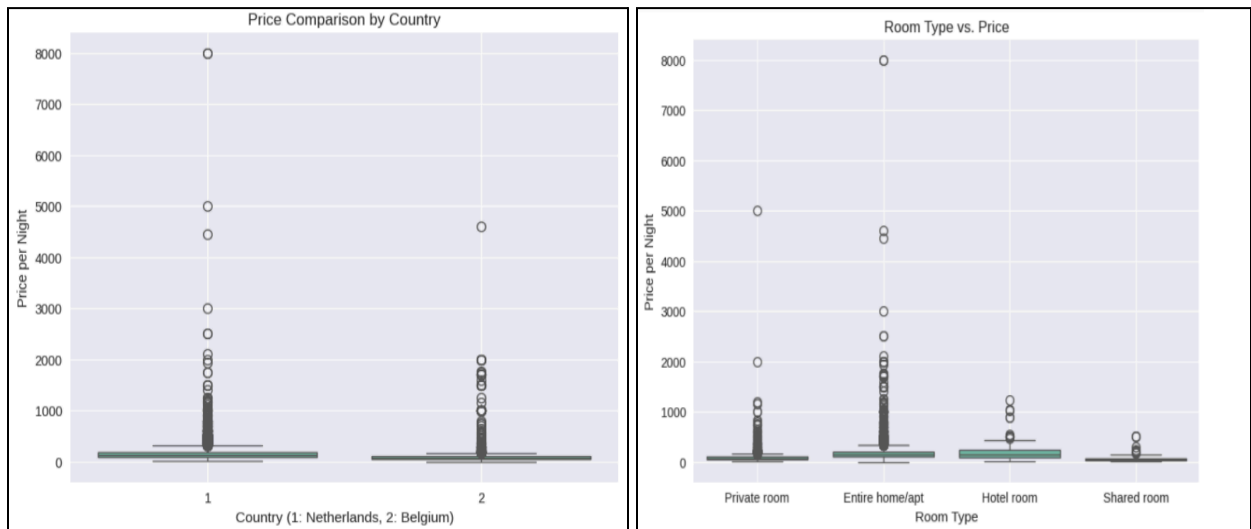


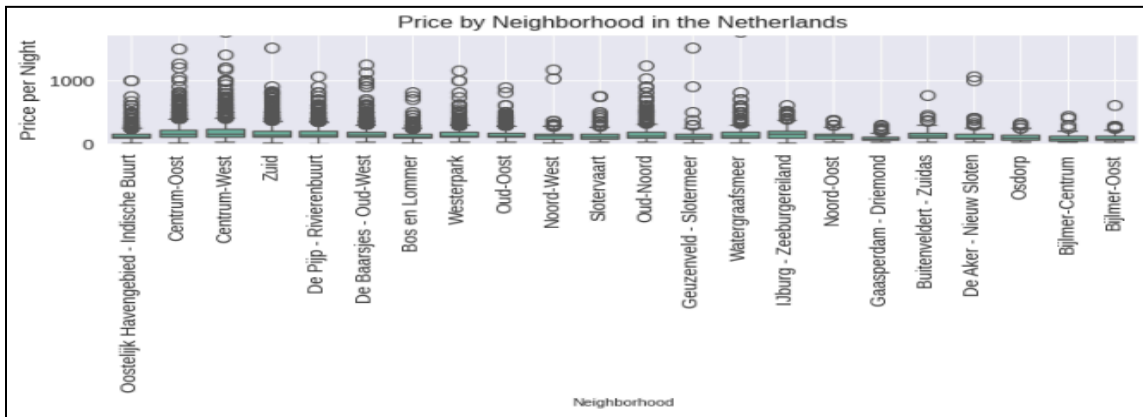
Fig 5: Country vs. Price

Fig 6: Room Type vs. Price

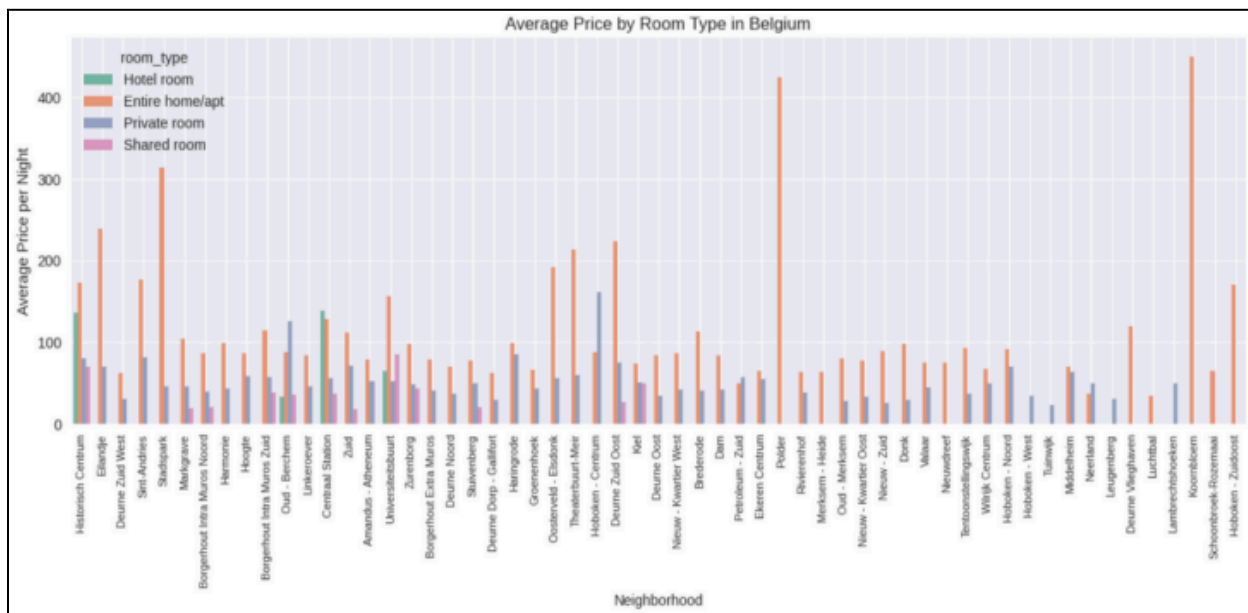
The plots illustrate price comparisons based on room type and location in both countries. The data reveals that entire homes/apartments command the highest nightly rates, followed by private rooms, hotel rooms, and shared rooms, which offer the most economical option for solo travellers. Additionally, the analysis indicates that the Netherlands generally has higher prices than Belgium, likely due to its tourist-heavy cities where demand exceeds supply.

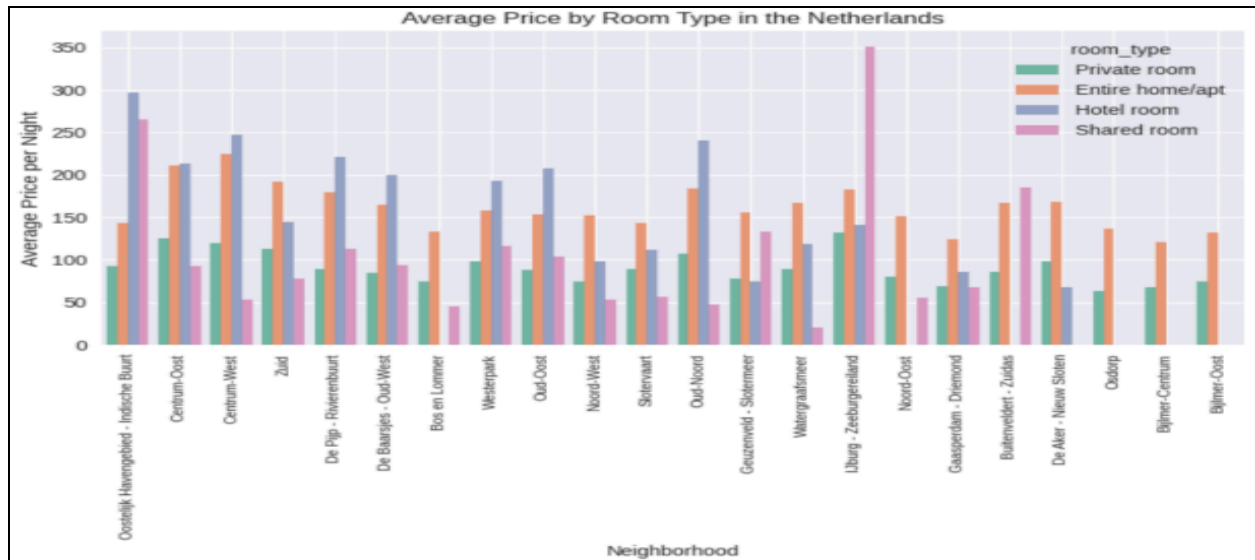
Neighbourhood Location vs.Price





This plot compares prices across different neighbourhoods, showing how location affects pricing. Some neighbourhoods may command higher prices based on their proximity to tourist attractions or city centres, which is key for decision-making around pricing strategy. Prices tend to spike in popular tourist districts, business hubs, or areas close to major attractions compared to residential districts on the outskirts.





- Entire homes/apartments are the highest priced, with shared rooms as the budget option, indicating demand for premium accommodations in central areas.
- Belgium shows extreme price spikes in some areas, with entire homes and hotel rooms commanding high prices, highlighting strong demand for luxury stays.
- The Netherlands has a generally higher price compared to Belgium. This could be due to the Netherlands having more tourist-heavy cities (e.g., Amsterdam) where demand surpasses supply, thereby driving prices higher.
- Entire homes/apartments are priced higher than private or shared rooms. This price difference can be attributed to the increased privacy and space provided by entire homes.

Effect of Outliers:

Netherlands

Before Outlier Removal: **€157.96** , After Outlier Removal: **€137.66**

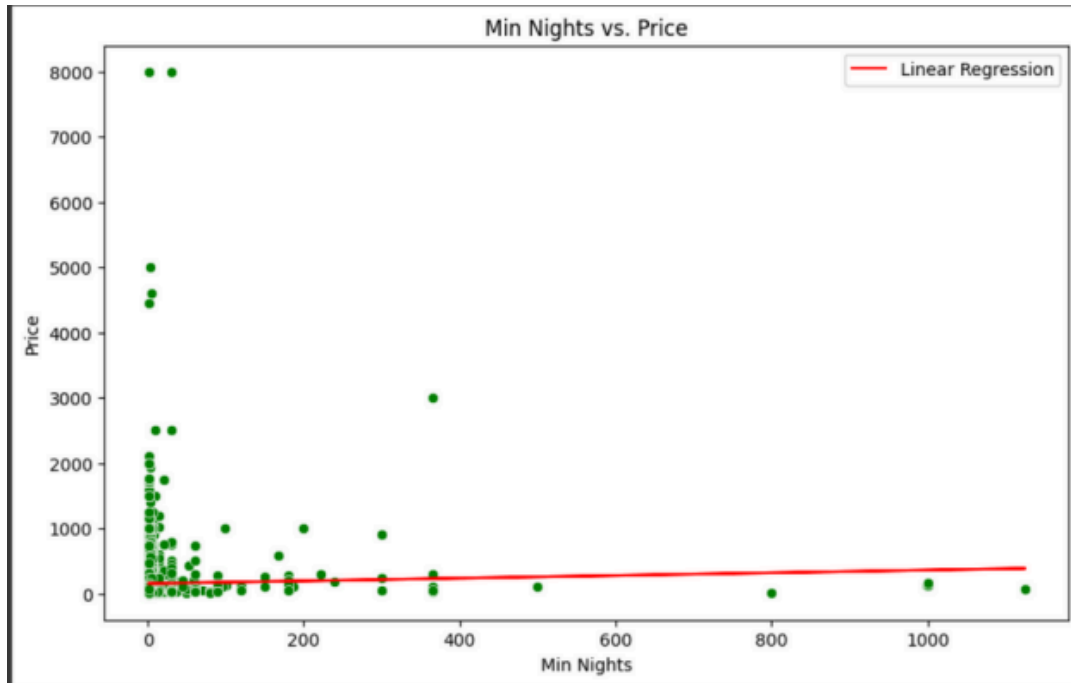
Outliers notably inflated the average price. Once these outliers were removed, the average price decreased by about €20, highlighting that specific properties with excessively high prices distort the data.

Belgium

Before Outlier Removal: **€128.09** , After Outlier Removal: **€72.09**

The impact is even more pronounced in Belgium, where the average price fell by around €56. This indicates that several highly-priced properties did not accurately represent the general market conditions

Applying Linear Regression On Price



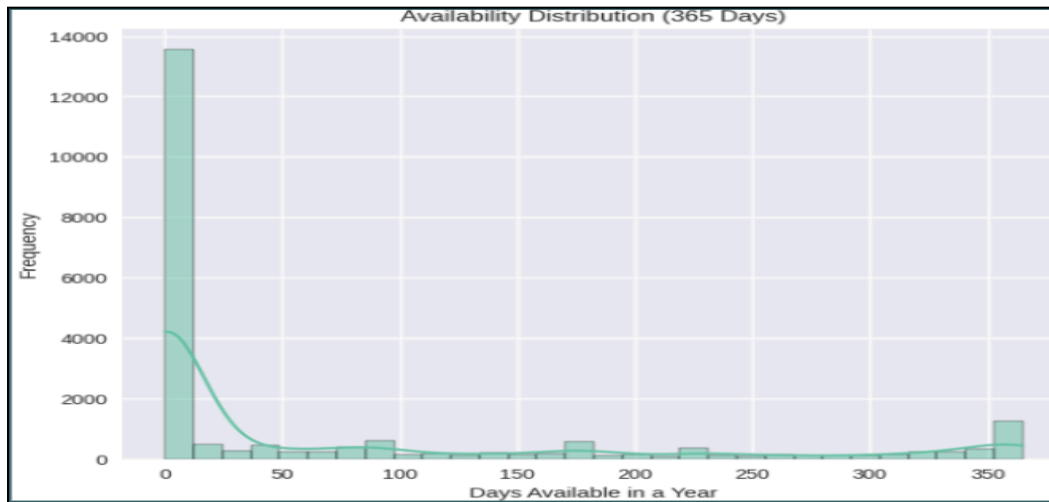
The linear regression line is almost flat, indicating a very weak or negligible positive correlation between the minimum number of nights and the price. This suggests that the minimum number of nights required has a minimal impact on the price of the listings.

Most listings with higher prices (above 1000) are clustered around lower minimum night values (close to 0). This implies that expensive listings tend to have lower minimum stay requirements.

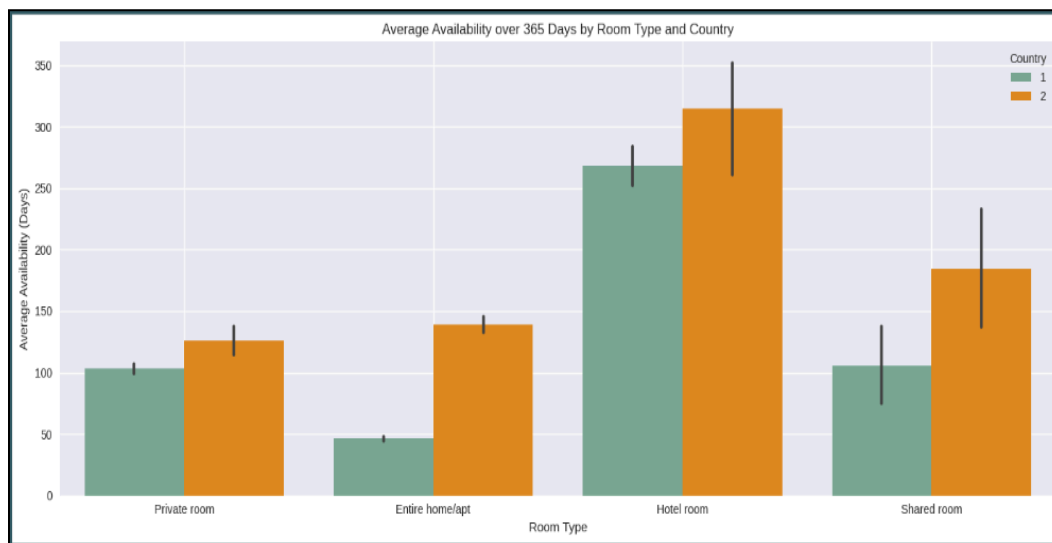
Summary of Price Analysis:

- Analysing price differences based on room type and country can help Airbnb adjust its pricing strategies. If the Netherlands consistently has higher prices, Airbnb can market Belgium as a more budget-friendly option while adjusting prices based on demand in premium locations.
- Airbnb can determine where to encourage hosts to list their properties or focus promotional campaigns by identifying which neighbourhoods have the highest prices and demand.
- Recognising and eliminating outliers produces a more precise representation of average market prices. This information could assist Airbnb in avoiding misleading potential customers or hosts by inaccurately estimating market trends using extreme listings.

3. Availability Analysis:



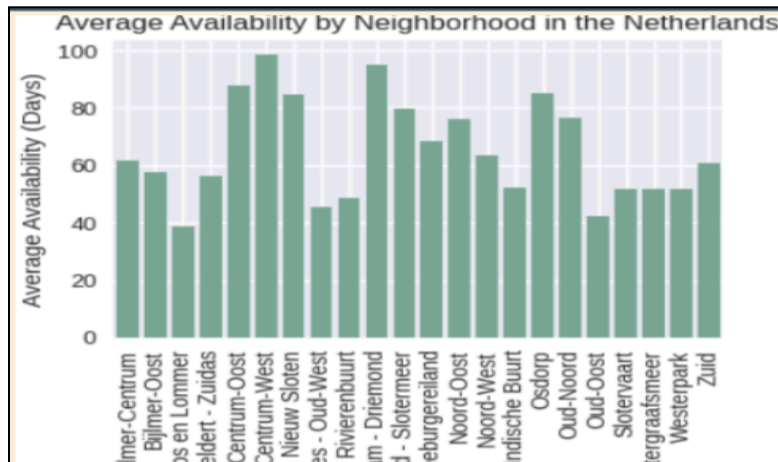
Frequency of Availability of rooms with days



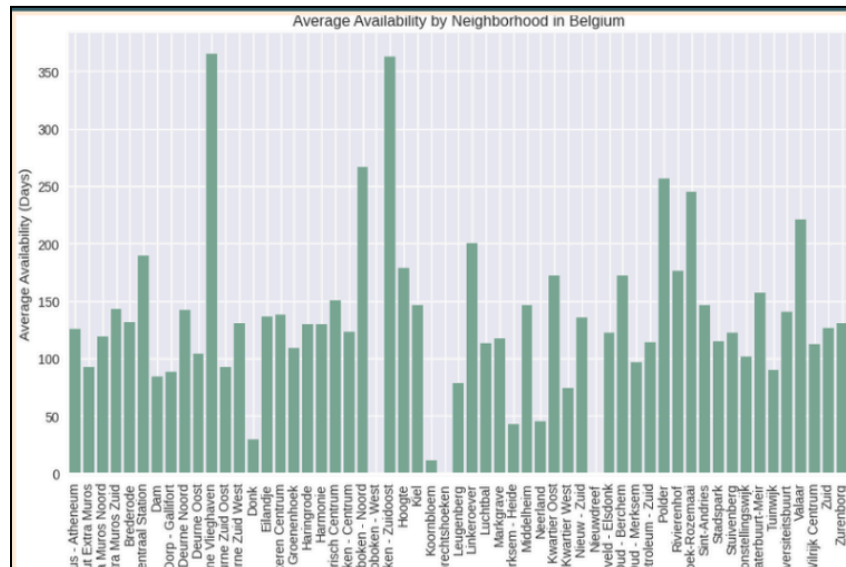
Availability of type of rooms in different countries

Rooms that are booked for shorter durations of time (1-3 days) have higher booking frequencies as opposed to the rooms placed for longer durations. Though the Netherlands has a broader dispersion of rooms, it still has more available listings than in Belgium.

Thus, understanding these availability patterns is crucial for managing supplies and analysing market saturation. For example, rooms available in the long term could mean less demand for these places or over-supply within those areas, demanding better pricing or marketing efforts to enhance booking rates. Conversely, if most neighbourhood areas experienced high booking rates, this would be considered a demand buoy, and thus, Airbnb would have to add more listings or, as a last resort, hike prices in those areas.



Availability of rooms in different neighbourhoods in Belgium

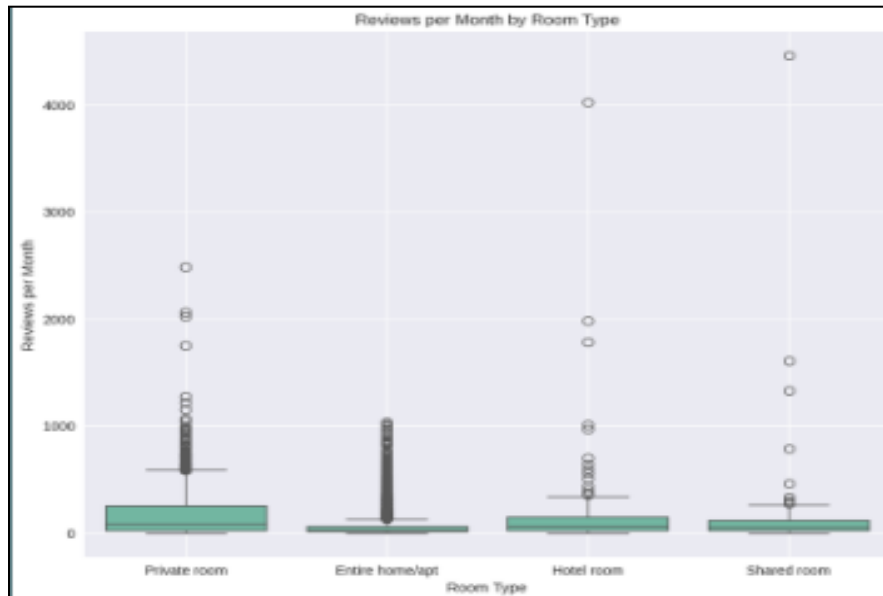


Availability of rooms in different neighbourhoods in The Netherlands

These graphs represent the availability of rooms in different neighbourhoods in both the Netherlands and Belgium. The above graph focuses on the Netherlands, showing the average daily availability for various neighbourhoods. We can observe that neighbourhoods like Centrum-west have higher availability, indicating lower occupancy rates.

The below graph displays a similar analysis for Belgium, where Hoboken shows significant availability, indicating that these areas might have less demand or are less attractive to visitors. This analysis shows that higher availability (over 180 days) might suggest lower demand or less desirable property types or locations. Conversely, rooms available for fewer days are likely in high-demand areas.

4. Review Analysis:



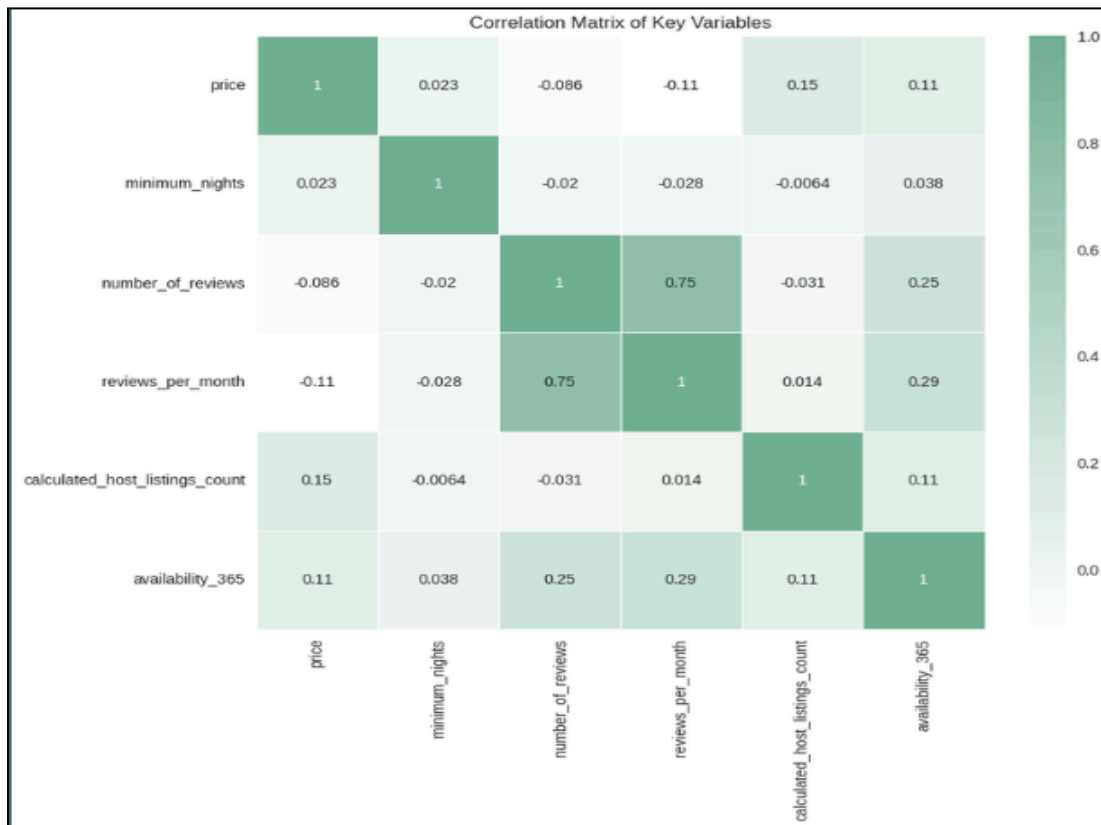
Private rooms receive the highest number of reviews per month, indicating strong demand compared to entire homes, hotel rooms, and shared rooms, which have fewer reviews.



Lower-priced listings (under €500) receive the most reviews, highlighting that affordability significantly drives guest interaction and engagement. On the other hand, High-priced listings (over €500) see fewer reviews, suggesting lower turnover and niche appeal.

Entire homes and hotel rooms attract fewer reviews, likely due to more extended stays or different guest preferences.

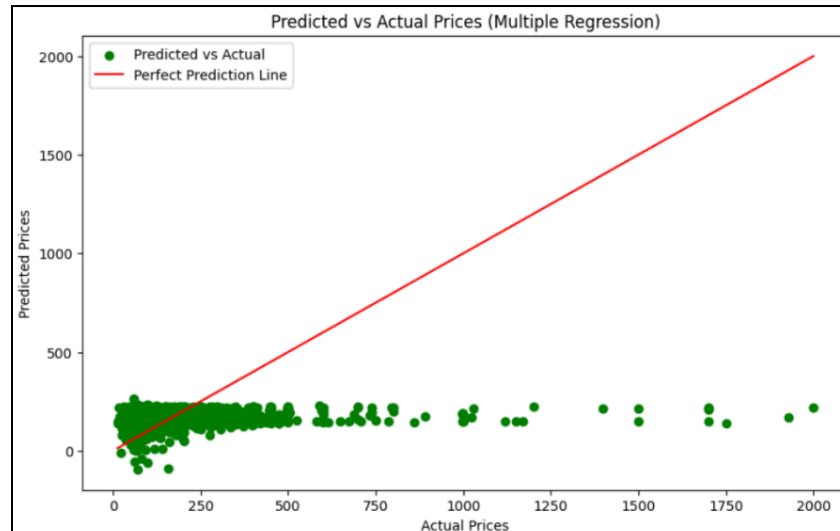
Correlation And Regression Model



Correlation Matrix of Dataset

The correlation analysis can help identify whether properties with more extended availability tend to have higher or lower prices. This would provide insights into pricing elasticity—do lower-priced rooms remain available longer, 0.11 between the price and availability. This is not a strong correlation and from this, we can make other decisions too.

This analysis can help AirBnB understand whether they should adjust pricing for properties that stay available for long periods. For example, if a negative correlation exists, lowering prices could help fill these properties more quickly, while a positive correlation might imply these properties are under-utilized despite high demand.



- Linear regression model demonstrates signs of underfitting, particularly in higher price ranges, considering more advanced models like decision trees, random forests, or gradient boosting could uncover non-linear relationships between features and price for improved prediction accuracy.
- The low explanatory power of the models implies that Airbnb needs to refine its approach by considering more complex factors such as proximity to attractions, host experience, or seasonality. Incorporating these features into the model would likely improve its predictive accuracy, helping AirBnB better understand the drivers behind price variability.

Overall Conclusion

This analysis provides critical insights into room distribution, pricing, and availability across Belgium and the Netherlands, helping AirBnB understand market dynamics better. If a traveller comes to either of the places seeking private rooms, Centrum-West emerges as the top location, while De Baarsjes is ideal for those looking for apartments. Entire homes/apartments and hotel rooms command the highest prices, especially in high-demand neighbourhoods like Centrum-West and Oosterparkbuurt - Indische Buurt, whereas shared rooms remain the most budget-friendly option.

Airbnb can focus more on these areas, giving this info to the users so they can have a good overview of the place. By identifying patterns in room type popularity and pinpointing areas where prices are inflated or skewed by outliers, this report can directly inform strategic decisions on pricing, room supply management, and future market expansion.

In summary, the data-driven insights, combined with recommendations for improving model performance, can help AirBnB optimise both customer satisfaction and revenue by tailoring its offerings based on actual market conditions.

This analysis provides a foundation for making data-driven decisions that enhance both customer satisfaction and revenue growth by better aligning AirBnB's offerings with market demand.