Airbnb Data Analysis - Excel

GREETINGS

Hello, It's Pernel.

I was very delighted to work on this project as it aligned perfectly with my passion for finding innovative solutions to business challenges. This opportunity allowed me to design and engage in an innovative sector, explore business constraints, leverage my data skills, and satisfy my curiosity by applying a data-driven approach at every step.

I hope that each reader will enjoy reading this report as much as I enjoyed creating it.

Stay Hysinghts !

NB: *It is important to understand that all data gathered in this context are influenced by specific circumstances so that it may not accurately represent reality.*

Summary

ABOUT THE PROJECT

Business Description

Airbnb is a global online marketplace that connects hosts offering accommodations with guests seeking short-term lodging experiences. Founded in 2008, Airbnb has revolutionized the travel and hospitality industry by allowing individuals to rent out their homes, apartments, or unique spaces. The platform facilitates transactions, provides secure payment options, and enables both hosts and guests to rate and review each other.

Airbnb operates as a two-sided marketplace, with key stakeholders including hosts (property owners), guests (travelers), and the platform itself. The business model is primarily commission-based, earning fees from both hosts and guests for each booking.

ABOUT THE PROJECT

Project Overview

The dataset includes details about hosts, locations, pricing, and key metrics, aiming to uncover trends in guest preferences, pricing strategies, and market dynamics across various European cities. The analysis helps stakeholders make informed decisions about the accommodation, expenses, and preferred room types in specific areas. It also guides travellers them in finding ideal locations based on specific criteria like price or hosts' reputation, making trip planning easier for those unfamiliar with the destinations.

Questions

1. Which factors have the greatest impact on the pricing of listings in various European cities ?
2. How do guest preferences differ from one city to another ?
3. Which areas within the cities experience the highest demand, and what features contribute to their appeal for guests ?
4. What market trends and seasonal variations are observed in occupancy rates and pricing across different cities ?
5. What insights can be drawn about future booking trends and potential revenue expansion for hosts ?

PROJECT WORKFLOW

ABOUT DATASET

We created our dataset by gathering information from various Airbnb listings in cities, with a focus on certain cities that have exhibited significant activity from the company. We selected features that aligned with our analysis goals. Initially, the raw data was organized by region and split into time periods (weekdays and weekends) before being combined into a single comprehensive dataset for better analysis and understanding.

The dataset includes about 51,707 listings with 22 columns, featuring both categorical and numerical data. The cities represented are Amsterdam, Athens, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome, and Vienna. The Booking prices range from $34.78 to $18,545.45 per night (Note: The pricing format on Airbnb is $/night, which has been considered in this analysis).

There are three types of room categories : Shared Room, Private Room, and Entire home/Apt. Notably, Shared Rooms are the least popular, despite being generally cheaper.

Please Take a look at the metadata file to explore all the features

CLEANING & PREPARATION

SUMMARY

- *Missing Data Handling:* No value found
- *Handling Duplicates :* No Duplicates streaming found
- The *data types* are appropriate for the content (e.g., int64 for integers, float64 for floating-point numbers, bool for boolean values, and object for categorical data like room_type)
- *Standardization:* All boolean Data has been turned into numerical values to better analysis process
- *Features:* room_shared and room_private have been delete because their information is already present on room_type
- *Outlier Treatment:* Some data observed in booking amount field are too high but at this point of the analysis, we can tell exactly that if there are outliers or not.

CLEANING & PREPARATION

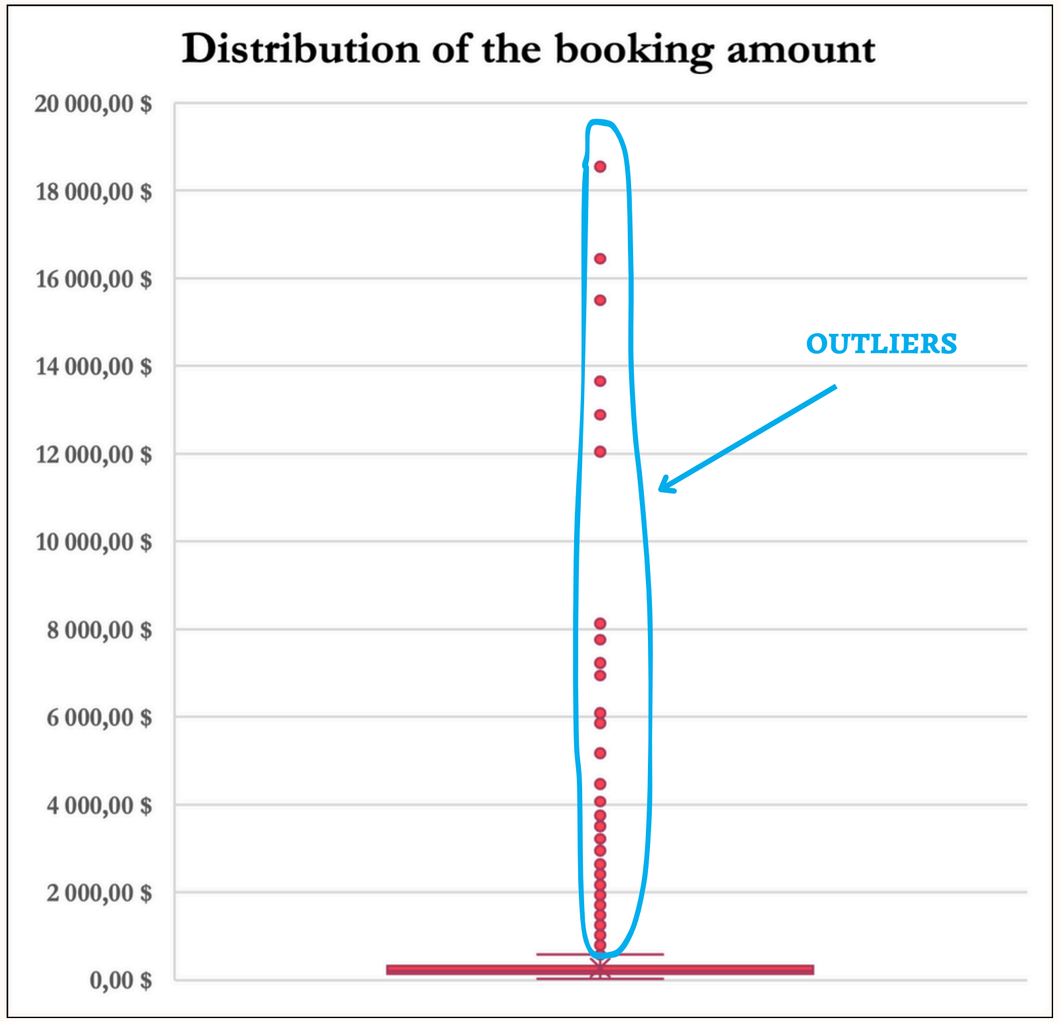
HANDLING OUTLIERS

The boxplot of booking amount suggests that the distribution is extremely right-skewed.

Mean	279,88
Median	211,34
Std	327,95

The descriptive stats observed suggests that the price mosty concentrate around 211$ (as median represents 50% of the data).

The surrounded data points might be considered as outliers. The values like 18,000$ are *extremely abnormal (outlier)* for a one night renting price. These values are probably due to data reporting error.



CLEANING & PREPARATION

HANDLING OUTLIERS

After visual inspection (Boxplot Interpretation), we used an Interquartile Range (IQR) method to detect the values considerate as Outliers (considered around 7 % of the data set).

IQR Stats

IQR	**170,94**
1st quartile	**148,75**
3rd quartile	**319,69**
Lower Bound	**-107,66**
Upper Bound	**576,11**

TRIM 5%	**255,07**
TRIM 10%	**246,14**

To handle these, we test the methods below :

Trimming : This technique result of losing precious data. Moreover, our data are right-skewed, that is not recommended for trimming.

Log Transformation : Can't really help in the visualization process (May be useful in case of data prediction).

Mean/Median Imputation : Still inappropriate because our data set is large, the outliers vary greatly (simple imputation might oversimplify the data and lead to inaccurate results).

We finally agreed that *removing* the outliers is the best option.

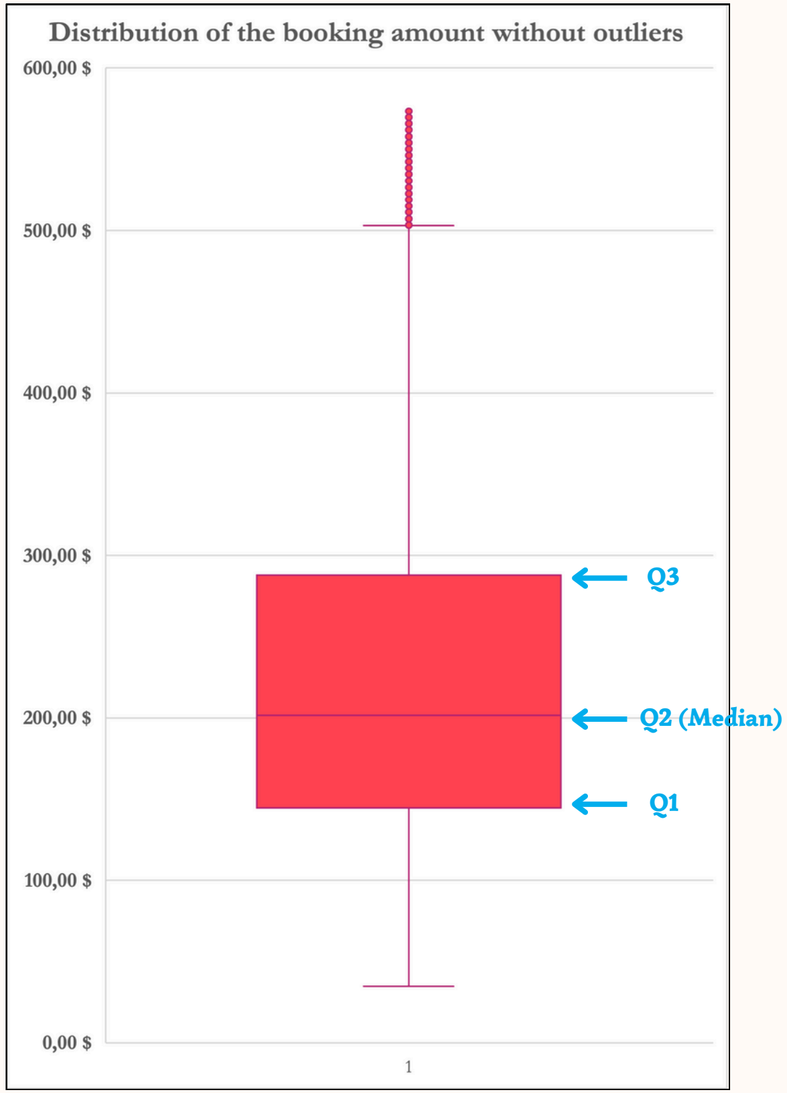
CLEANING & PREPARATION

AFTER HANDLING OUTLIER ANALYSIS

The data is more readable, considered as slightly right-skewed (median is closer to Q1). This indicate that most of the airbnb are approximatively booked at 200$/night.

The IQR (range Q1 to Q3) suggests that most airbnb are clustered around the 144 $ - 288 $ range.

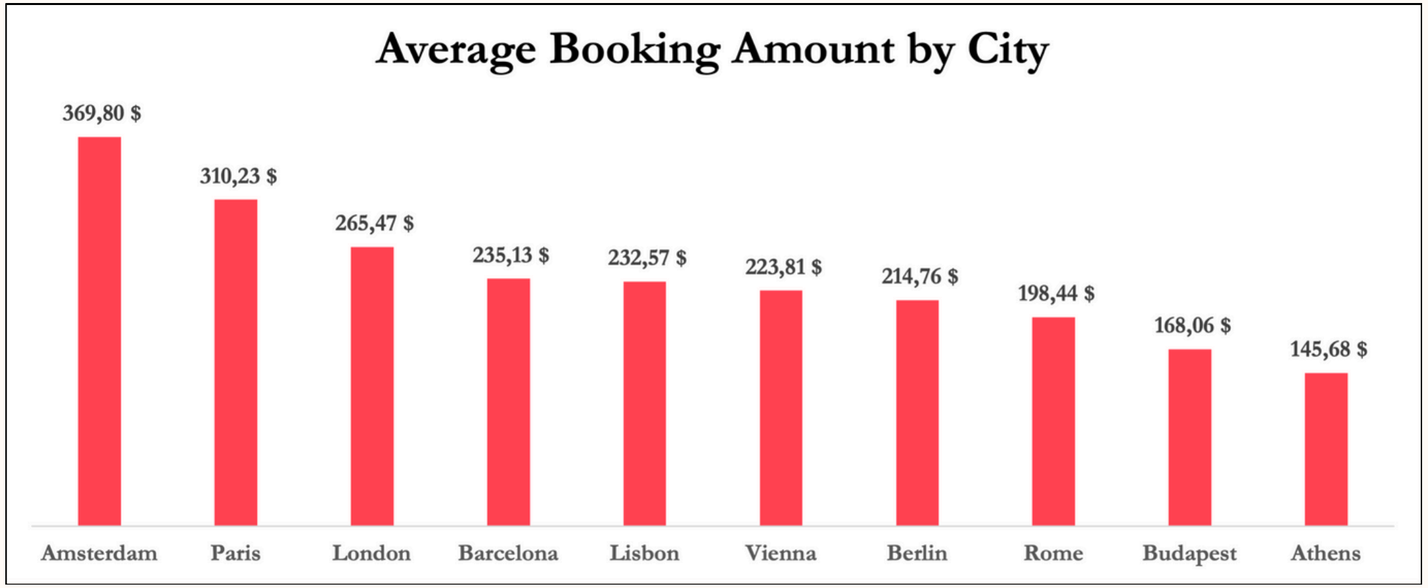
The presence of higher amount may indicates exceptional performances or may due to other conditions like room_type or person_capacity.



ANALYSIS & VISUALIZATION

MARKET POSITIONING

This chart tells the story of how different European cities position themselves in the Airbnb market.



The data suggests that travelers heading to Amsterdam or Paris should prepare for higher accommodation costs, while those visiting Athens or Budapest might find more budget-friendly options. This observation can be attributed to factors like tourism demand, economic conditions or local market dynamics and house amenities.

ANALYSIS & VISUALIZATION

SATISFACTION ANALYSIS

This heatmap displays the average guest satisfaction ratings across various European cities.

The color gradient from red to green identifies cities with lower to higher satisfaction scores. Despite lower booking prices, cities like Athens and Budapest still excel in guest satisfaction potentially offering better value for money.

On the other hand, higher-priced cities like London and Paris, while still performing well, may need to address specific areas to enhance guest experiences further.

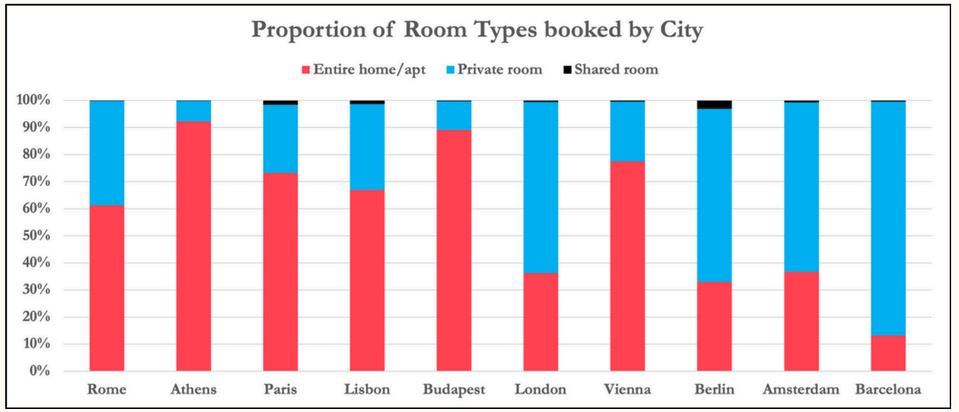
Cities	Avg_guest_satisfaction
London	90,60154717
Lisbon	91,04950843
Barcelona	91,24442736
Paris	91,84463128
Rome	93,10234899
Vienna	93,73700833
Amsterdam	94,1119403
Berlin	94,33067227
Budapest	94,60276382
Athens	94,99162543

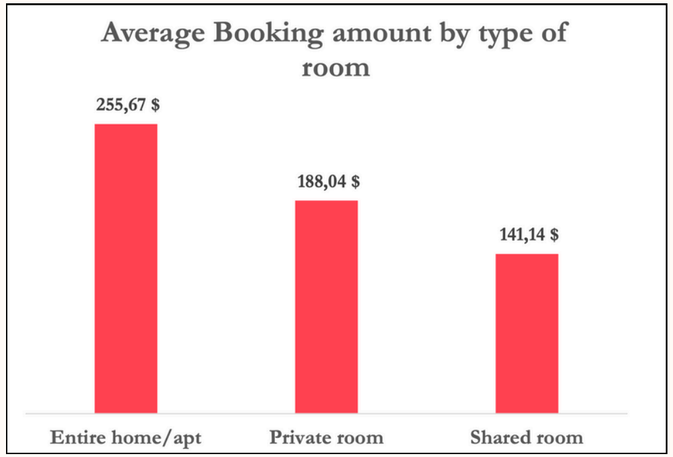
ANALYSIS & VISUALIZATION

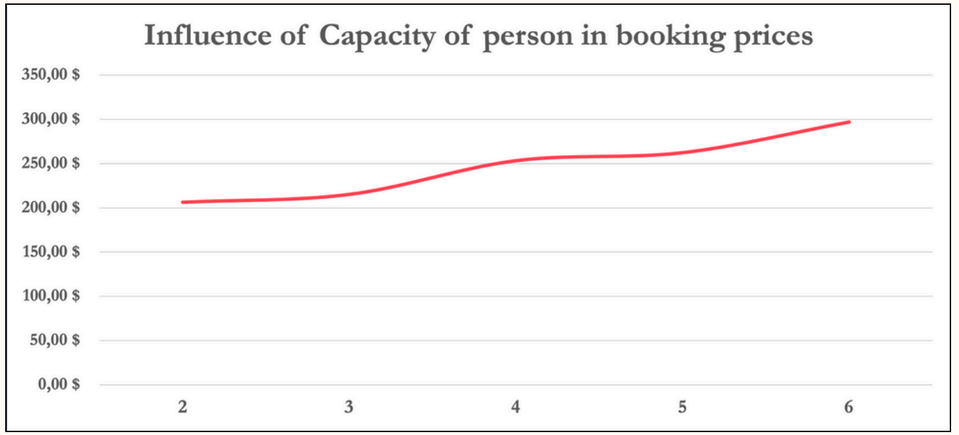
PRICE AND ROOM TYPE ANALYSIS

The average costs associated with different types of accommodations booked suggests that reserving an entire home tends to be more costly compared to booking a private or shared room.
Furthermore, the line chart reinforces the observation that as the capacity for guests increases, the prices tend to rise correspondingly.

Shared Rooms analysis indicates that people rarely prefer taking this type of accommodations despite the cheapest pricing.



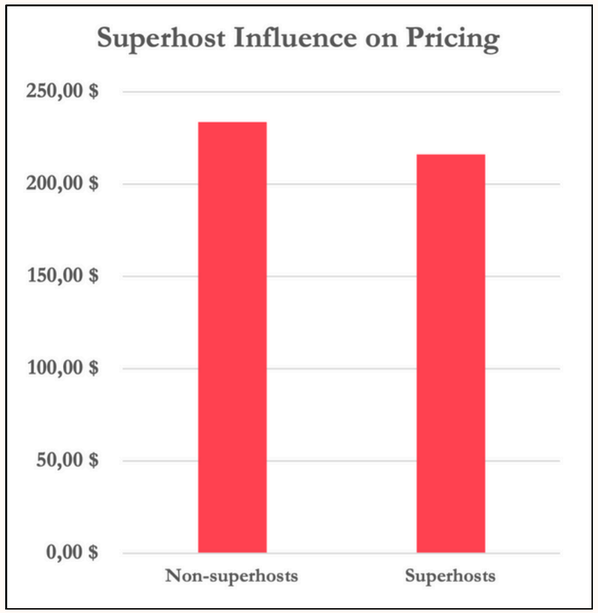


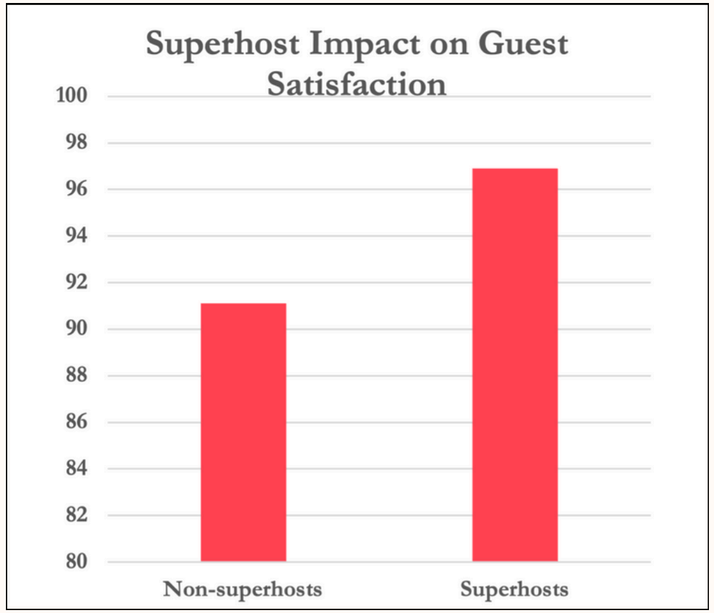


ANALYSIS & VISUALIZATION

SUPERHOST IMPACT ANALYSIS

Contrary to the assumption that Superhosts would set higher prices owing to their esteemed reputation, the data indicates that Non-Superhosts, on average, charge more.





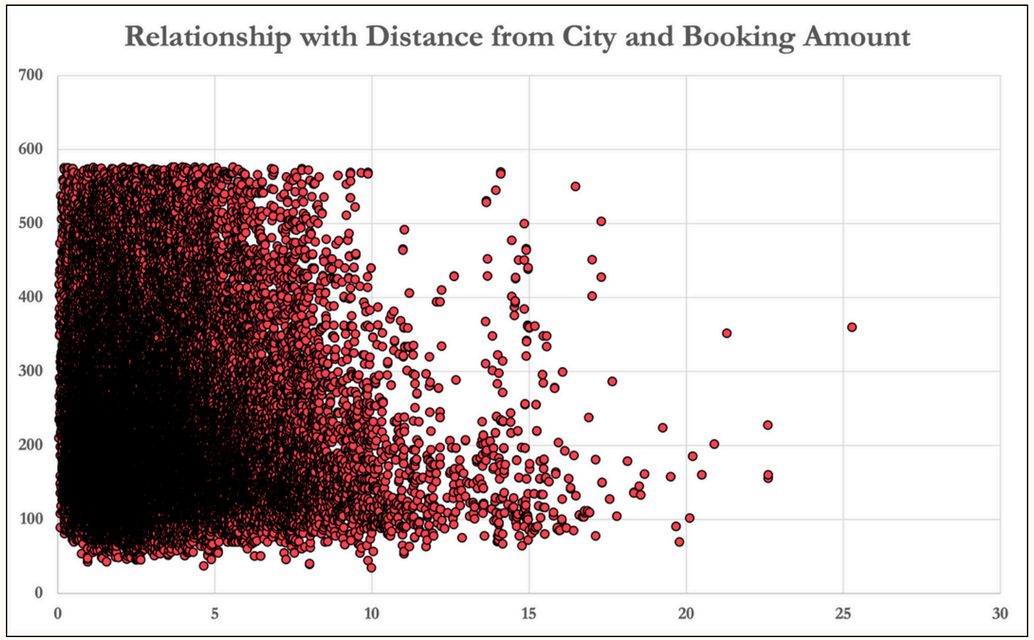
This may suggest that Superhosts focus on alternative priorities, such as achieving higher occupancy rates or enhancing customer satisfaction, rather than solely maximizing their nightly rates.

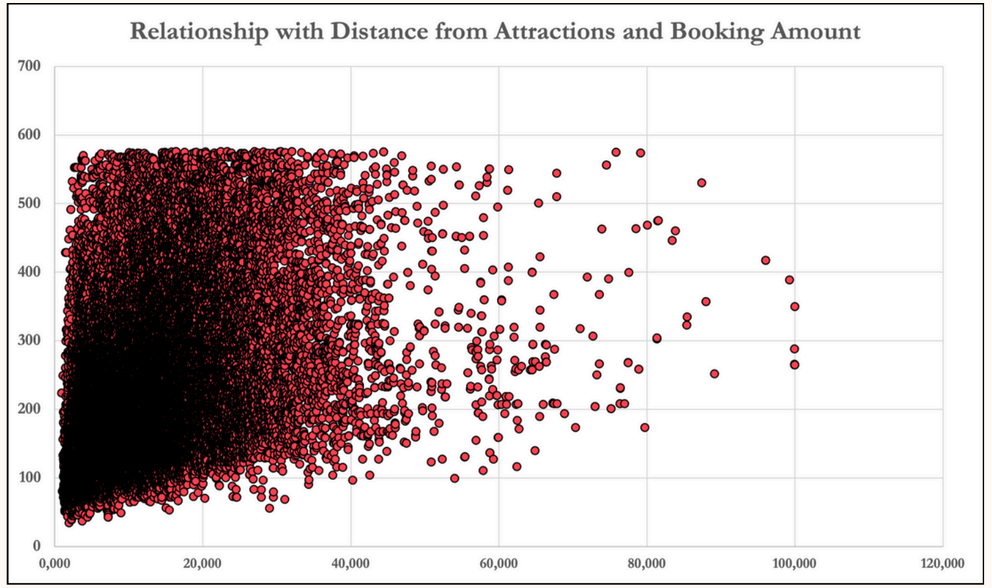
Evaluating the influence of Superhost status reinforces the notion that guests is greater satisfied when their Airbnb host is a Superhost.

ANALYSIS & VISUALIZATION

LOCATION AND ACCESS ANALYSIS

These scatter plots reveals a dense concentration of listings near principal city assets, with booking amounts varying widely within this group.



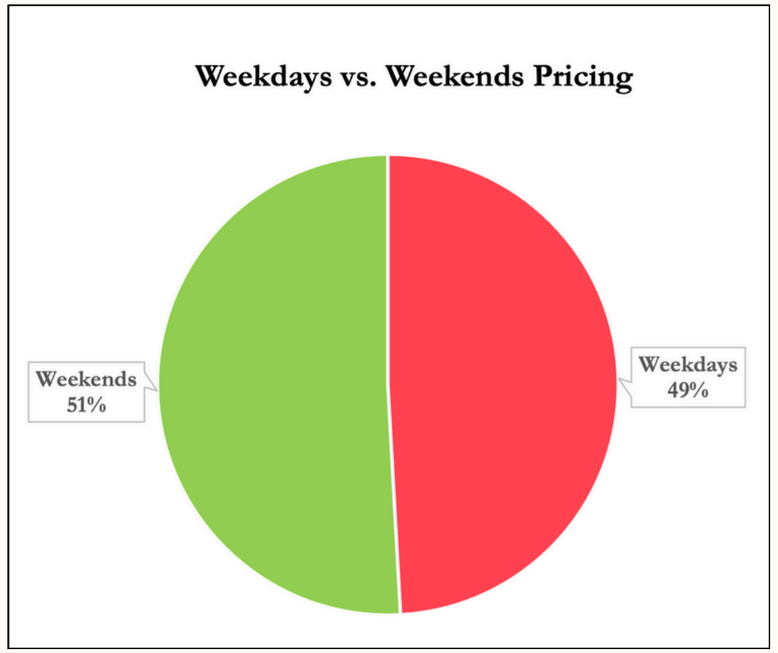


Interestingly, even at greater distances, some listings still achieve high booking prices, suggesting that proximity to attractions is not the only factor influencing pricing.

ANALYSIS & VISUALIZATION

TIME PERIOD ANALYSIS

This pie chart illustrates the distribution of pricing differences between weekdays and weekends.



While weekends traditionally are seen as more valuable booking days, the data suggests only a slight advantage in pricing, pointing towards a market where weekday and weekend demand may be closer than expected.

Actually, we don't have this data but we rather thinking that a deep dive in specific time periods or season analysis could reveal patterns to support more dynamic pricing variation of weekdays and weekends.

Please Take a look at the dashboard section on Excel file to explore the report

RECOMMANDATIONS

RECOMMANDATIONS

Optimize Pricing Strategies

- Raise the prices incrementally at lower-satisfaction cities, while keeping up with high level quality that drives guest satisfaction
- Spend a portion of the increased revenue on improving guest experience
- Dynamic pricing tools – which raise or lower items based on demand, season and what's going in your neighborhood
- Especially in cities with low booking volume, offer special packages or prices for stays of a longer duration
- Think about increasing the number of listings in areas with budget friendly pricing

Improve Guest Satisfaction

- Conduct detailed surveys or feedback analysis to pinpoint common issues such as cleanliness, check-in experience, or accuracy of listings. Use this information to develop targeted improvements.
- Introduce a quality control checklist for hosts in cities with lower satisfaction scores
- Provide training and resources to hosts to help them improve their service levels.
- Increase the Visibility of Superhost Listings

Thank you for your attention

By Pernel Avougnassou

pernelavougnassou@gmail.com