January 6, 2024

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Airbnb's listings occupancy analysis: strategies to maximize revenue

Report

Data Management and Visualization

Program: MSc in Business Administration and Data Science

Authors: Michele Daconto (176193), Federico De Marinis (176194), Alessio Desideri (176184), Ludovico Francia (177119)

Teacher(s): Abid Hussain, Travis Greene

Number of characters: 40,872

Number of pages: 18

Link to dashboard: here

Academic Year 2024/2025

Executive Summary

Airbnb Denmark wants to leverage the data available in the files *calendar2024.csv*, *reviews2024.csv* and *listings2024.csv* to increase the revenue generated from short rentals in the city of Copenhagen.

The data science team with this report wanted to analyze strategies to increase the Occupancy Rate of listings over the next 365 days. An increase in this value means a higher number of nights booked and thus an increase in Gross Booking Value, one of the main performance metrics recognized by Airbnb.

After cleaning the given data using Python – Pandas library more specifically – and creating an analytical schema with PostgreSQL, four main variables closely related to Occupancy Rate were identified: the price and the rating of listings, the response time of hosts, and the ability to book instantly.

Based on the findings, the data scientists propose the Airbnb management to adopt a strategy based on the following approaches:

1. Raising awareness among hosts about price choice, the importance of reviews and guest communication;
2. Implementing new features within the platform to directly influence these factors that affect Occupancy Rate.

Following these recommendations would allow Airbnb, in addition to increasing the number of nights booked, to increase the value added by both hosts and guests.

Airbnb indicates within its shareholder letter (Q3 2024) two key business metrics: *Nights Booked* and *Gross Booking Value*. By leveraging the recommendations of this report, Airbnb's management team has the opportunity to understand how to impact these two metrics that are key to Airbnb's business growth in the city of Copenhagen.

Keywords: Airbnb – Database Construction – Dashboard Design – Occupancy Rate – Identified Strategies

Table of contents

1. Introduction

1.1 The Airbnb’s business model

Founded as a start-up in San Francisco in 2008, Airbnb has revolutionized the short-term rental market by creating a platform that connects property owners (hosts) with travelers seeking accommodation. The company's success stems from its ability to offer travelers a more flexible and diverse alternative to traditional hotels, while allowing hosts to monetize otherwise under-utilized spaces. By comprehending a wide range of properties - from budget rooms to luxury villas - the platform appeals to a broad audience with varying preferences and needs.

Since its inception, Airbnb has experienced exponential growth. In December 2020, the company went public with an initial public offering (IPO) that was widely regarded as a significant success, managed by a consortium of investment banks led by Morgan Stanley and Goldman Sachs. Further, by January 2024, Airbnb's revenues had reached nearly $9 billion (Q3 2024). Beyond its remarkable financial milestones, the company has increasingly gained social relevance by substantially changing the way people travel.

The cornerstone of the company’s success is its platform business model. In such a model, value is created by facilitating connections between individuals (hosts and guests in this case), and revenue is derived from service fees. Similar companies that have adopted this platform model include Glovo, Trivago, Booking.com and many others.

Further, Airbnb's emphasis on technological innovation has positioned it not only as an economic leader in the industry, but also as a technological pioneer. For example, it was the first platform to introduce qualitative distinctions between hosts through the 'superhost' designation.

Airbnb generates revenue primarily through variable commissions on bookings made through its platform. Hosts are typically charged around 3% of the booking total, while guests pay a service fee of between 14% and 20%. In addition, the company has diversified its revenue streams by offering services such as Airbnb Experiences (e.g. creative workshops or cooking classes hosted by locals), partnerships, advertising, cancellation fees, and premium tools such as Airbnb Plus and Host Tools.

1.2 Airbnb in the Denmark area

Since 2009, Airbnb has played a significant role in Denmark's hospitality industry, especially in Copenhagen. After Stockholm, it is the Nordic city with the highest number of Airbnb listings per capita, with visitors representing 45% of the country's total tourist population. The platform offers thousands of listings to accommodate the approximately 4 million tourists who visit the capital each year. Outside of Copenhagen, the company has a notable presence in Aarhus, although the scale is significantly smaller - around a fifth of the listings in the 'Hygge Capital'.

1.3 The datasets

As part of Airbnb's data science team, we were given three CSV (comma-separated values) datasets: *listings2024.csv*, *calendar2024.csv* and *reviews2024.csv*. We haven't used the third dataset for our analysis. Since the data begins in late June 2024, let's assume we are in that period.

The main data source of the project was the first dataset, which provides detailed information about each listing, including the host, property details, availability and various ratings collected over time.

The information within *calendar2024.csv* was also key to our conclusions. The file focuses almost solely on the previsions of availability of each listing over the next 365 days, using a boolean variable (*availability*) to indicate whether a listing is available on a given day. It is structured so that each listing has 365 rows - one for each day of the year - with the date and availability status being the only variables that change. Interestingly, the price variable, which is also included, remains the same for each listing every day of the year.

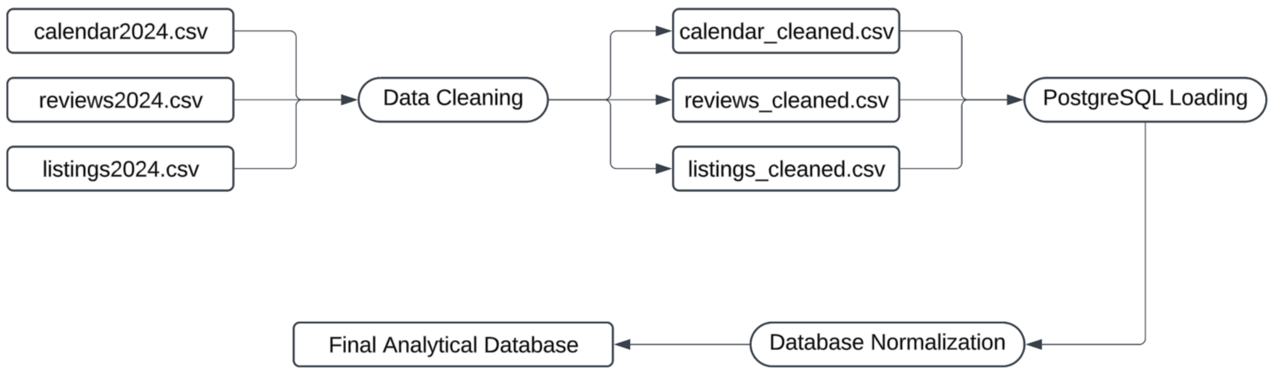
1.4 The idea and how we arrived to it

The unusual stability of the price variable across all days of the year prompted us to further investigate the pricing strategies used by hosts. Our aim was to understand how Airbnb could optimize the efficiency of its business and, in turn, maximize occupancy across all listings. This focus on occupancy was further validated by the company's Q3 2024 letter to shareholders, where the two key business metrics identified - *Gross Booking Value (GBV)* and *Nights and Experiences booked* - are substantially expression of the occupancy performance.

To explore these insights, we have developed a dashboard with two key views. The first one enables management to understand the negative correlation between price and occupancy. In particular, it highlights how reducing the percentage of commissions charged by Airbnb could have a volume effect, benefiting the business as a whole. The second view, which offers an alternative approach if the pricing strategy is deemed too intrusive, provides data to support the idea that Airbnb could still increase occupancy by focusing on host education. In this sense, in fact, campaigns aimed at training or raising awareness about best practices could enable hosts to effectively increase their Occupancy Rates.

Before we dive into the details of the dashboard, however, let us first explain how we cleaned and organized the data we were given.

2. From raw data to structured data



To prepare the data we have available for subsequent analysis, we performed several operations that are depicted in Figure 5, that can be found in the appendix. The data contained in the three csv files were each subjected to an initial exploration and cleaning process. The data were then uploaded to PostgreSQL, normalized and used to build an analytical database.

2.1 Exploring and cleaning the calendar dataset

The file *calendar2024.csv* provides data on the availability and prices of each listing in the city of Copenhagen for a period of 365 days from 2024-06-29 to 2025-06-29. A detailed description of each variable can be found in Figure 2, located in the Appendix. Before conducting any kind of analysis, the data were explored and cleaned using a Jupyter notebook (available here) to ensure their accuracy.

After loading the file into a pandas dataframe we checked for missing values. The *adjusted_price* column exclusively counted missing values and, thus offering no useful information, we decided to remove it. The *minimum_nights* and *maximum_nights* columns each had only one missing value that was present in the same record. Since the dataset contains a large number of records we decided to drop this record.

We ascertained that the dataset did not have redundancy by checking for duplicate records.

We assigned each attribute in the dataset the appropriate format. *Listing_id* remained unchanged. Values in *date* were converted to datetime format. The values in *available* were transformed into the boolean values True (if t) and False (if f). We removed the dollar sign and thousand separators from values contained in *price* and then converted them to float format. Values in *minimum_nights* and *maximum_nights* were converted to integer format to facilitate subsequent calculations.

We verified that of all 20,909 listings, only 2 listings had less than 365 days of data, so we decided to remove all records associated with these listings.

Since the prices are expressed in Danish kroner, we wanted to rename the *price* column to *price_dkk* to be clearer.

In analyzing the distribution of prices, which varies from a minimum of DKK 21 to a maximum of about DKK 100,000, we implemented corrections to the extreme values to try to correct any errors in scraping. After analyzing the prices on the Airbnb portal for listings in Copenhagen, prices of listings below 300 DKK (~40€) were changed to 300 DKK, and those above 37,303 DKK (~5,000€) were changed to 37,303 DKK.

We observe that the prices of each listing do not change and remain the same during the 365-day period. However, the values of *minimum_nights* and *maximum_nights* change on different days.

2.2 Exploring and cleaning the reviews dataset

The file *reviews2024.csv* contains data on reviews made by users about the listings where they stayed overnight. Each review is associated with an id that identifies it, a date of publication, a specific listing, and a specific reviewer. A detailed description of each variable can be found in Figure 3, located in the Appendix. Again, we explored and cleaned the data using a Jupyter notebook (available here).

After loading the file as a pandas data frame, we found the presence of 17 missing values in the *comments* column. Since this attribute is the one that gives us the most information in this dataset, we decided to remove the 17 records with missing values. The resulting dataset is therefore free of additional missing data.

We verified the uniqueness of the reviews using the *id* column (review IDs), which in fact has no duplicate values (the number of unique values is equal to the length of the dataset).

The values in the *date* column were converted to datetime format while the other attributes remained unchanged.

The distribution of values in *date* shows us that we have available reviews published in the date range from 2010-07-25 to 2024-06-29.

In order to allow for greater readability and to prepare the texts of the reviews for possible later analysis, we made some transformations. Break tags were replaced with spaces, multiple spaces were collapsed into a single space and apostrophes were removed.

2.3 Exploring and cleaning the listings dataset

The file *listings2024.csv* contains data regarding listings published on the Airbnb portal such as price, host and property characteristics. A detailed description of each variable can be found in Figure 4, located in the Appendix. Again, data exploration and cleaning was conducted using a Jupyter notebook (available here).

As a first operation we replicated the transformations applied to the *price* attribute of the *calendar2024.csv* file: the *price* attribute was renamed *price_dkk* and prices were capped below 300 DKK to 300 DKK and above 37,303 DKK to 37,303 DKK.

The values of the attributes *host_response_rate* and *host_acceptance_rate*, which were stored as percentages, were changed to float format after removing the percentage symbol.

The values of the attributes *last_scraped*, *host_since*, *calendar_last_scraped*, *first_review* and *last_review* were converted to datetime format while the values of *host_is_superhost*, *host_has_profile_pic*, *host_identity_verified*, *has_availability* and *instant_bookable* were converted to boolean values.

The columns *neighborhood_group_cleansed*, *calendar_updated* and *license* contained only missing values and were therefore removed. Missing values of *beds* were replaced with values, if any, of *bedrooms* following the assumption that if a listing has *x* bedrooms then it will definitely have at least *x* beds.

7,313 listings of 20,909 listings have a missing value of *price_dkk*. These missing values will be handled later.

The *neighborhood_cleansed* attribute had five spelling errors in the neighborhood names that we had to correct. The *neighborhood* column contained redundant data already contained in the variable *neighborhood_cleansed* and therefore was removed.

The missing values in *bathrooms* were filled in for the most part by extracting the numeric values contained in the variable *bathrooms_text*. We then constructed a new boolean variable *is_bathroom_shared* that takes the value True if the word "shared" is contained in the *bathrooms_text* column, False otherwise.

Finally, to increase the readability and usability of the text contained in the variables *description*, *neighborhood_overview* and *host_about* we removed the break tags and compressed the multiple spaces into a single space.

2.4 Handling missing prices in listings dataset

One of the major problems found in the *listings2024.csv* file is that of the 20,909 listings, as many as 7,313 have a missing price. To solve this problem, we leveraged the data contained in the *calendar2024.csv* file. Previous analysis of this file revealed that each listing is associated with a price that remains constant over the year. This fact allowed us to make the following assumption: if a listing with a missing price from the *listings2024.csv* file is present in the *calendar2024.csv* file with a non-zero price, then we can reliably extract these prices and insert them into the *listings2024.csv* file.

We used a Jupyter notebook (available here) to perform these transformations.

We first verified that the 7,313 listings with missing prices were present in the file *calendar2024.csv* with non-zero price values.

Finally, for each listing, the corresponding price was extracted from the *calendar2024.csv* file, which was then substituted for the missing value.

These operations allowed us to reduce the missing values of the *price_dkk* attribute from 7,313 to 0 without having to delete any records from the *listings2024.csv* file.

2.5 Relational database

Once the data cleaning phase was finished, we decided to transform the data into a structured database using PostgreSQL and pgAdmin. Each file was loaded into separate tables while respecting the original data structure.

Using a relational database allowed us to query the available data more quickly, access the information of interest more easily, and most importantly, this approach makes future analysis developments scalable, which will be possible due to the ability to integrate new data more easily.

The tables loaded into the database were subjected to a normalization process and then linked to the dashboards built on Tableau.

2.5.1 Data normalization

Once the files with clean data were loaded into the *reviews*, *calendar* and *listings* tables of the database, we proceeded to carry out a process of normalization of the data. This process allowed us to reduce the redundancy and ambiguity of the information (also allowing us to optimize the space used), reduce the risk of inconsistencies in future data updates, and improve the performance of the queries, which by operating on smaller tables will be executed in less time.

The *reviews* table violated the second and third normal forms: the *reviewer_name* attribute was not fully dependent on the primary key *id* and had a transitive dependency with the *reviewer_id* attribute. To normalize this table, we removed *reviewer_name* from *reviews* and created the new table *reviewers* which includes the attributes *reviewer_id* and *reviewer_name*.

The table *calendar* was not modified since it is already normalized to the third normal form.

Finally, we had to modify the structure of the table *listings* which did not comply with the normalization rules. The primary key of the table is *id* and therefore we expect each attribute to depend completely on this key. Some data, for example *host_name*, depended on other attributes and not on the primary key. Therefore, we created the following tables: *hosts* which contains information about hosts, *web_scrape* which stores details about web scraping, and finally *availability* and *listings_reviews_summary* which because they contain data that changes frequently over time, we thought it appropriate to isolate from the other listings information.

All SQL queries used to transform the three initial CSV files into the final analytical database are available here. The final result of the data transformation process is represented in the entity-relationship diagram, refer to Figure 1 of the Appendix. The diagram illustrates the structure of the constructed analytical database by showing the main entities, their attributes and the relationships between the different tables.

3. Interpretation: disclaimer and limitations

This chapter outlines the assumptions, interpretations, and limitations of the analysis conducted in this report. By clarifying how the availability data was treated and highlighting the potential shortcomings of our approach, we aim to provide transparency and set a proper context for interpreting the findings.

3.1 Interpretation of availability data

The availability column in the dataset represents whether a listing is available for booking on a specific day, marked as either True or False:

- True: Indicates that the listing is available for booking but not currently generating revenue.
- False: Indicates that the listing is unavailable for booking, which we interpreted as the listing being rented and generating revenue.

However, this interpretation simplifies reality. In practice, there are scenarios where a listing may appear as unavailable (False) but is not actually generating revenue. This discrepancy often occurs due to the booking time cap, where hosts limit how far in advance guests can book their listings. For example, a host may only allow bookings within a 1- or 3-month window, even if the listing is unoccupied for the remaining months.

Similarly, a host may block availability for personal reasons, such as renovations or vacations, without any intention of renting the property during that time.

For the purpose of this analysis, we are ignoring these scenarios and making the simplifying assumption that any listing marked as False is rented and generating revenue. We acknowledge that this assumption may lead to an overestimation of Occupancy Rates.

Why does this matter? To illustrate why this assumption is significant, consider the following example: imagine a listing with a very low rating (1 out of 5 stars). Suppose the listing is not booked for the next month, and the host has set a booking cap of 1 month in advance. For the remaining 11 months, the listing appears as False in the dataset, making us assume it is fully booked and generating revenue. However, at the end of the year, the listing might have been completely unoccupied, earning no revenue at all.

This example highlights the potential for misinterpretation when using future availability data as a measure of occupancy performance, particularly for listings with restrictive booking policies or low guest demand.

3.2 Data source and accessibility

While the *listings*2024.csv dataset offers valuable insights, it lacks the granularity and historical context of Airbnb's internal data, such as actual booking histories, guest preferences, or host-specific performance metrics.

Access to past performance data, such as actual Occupancy Rates and revenue figures, would greatly enhance the depth and accuracy of this analysis. However, due to the limitations of the available data, our findings are derived solely from the future availability information and should be interpreted with caution.

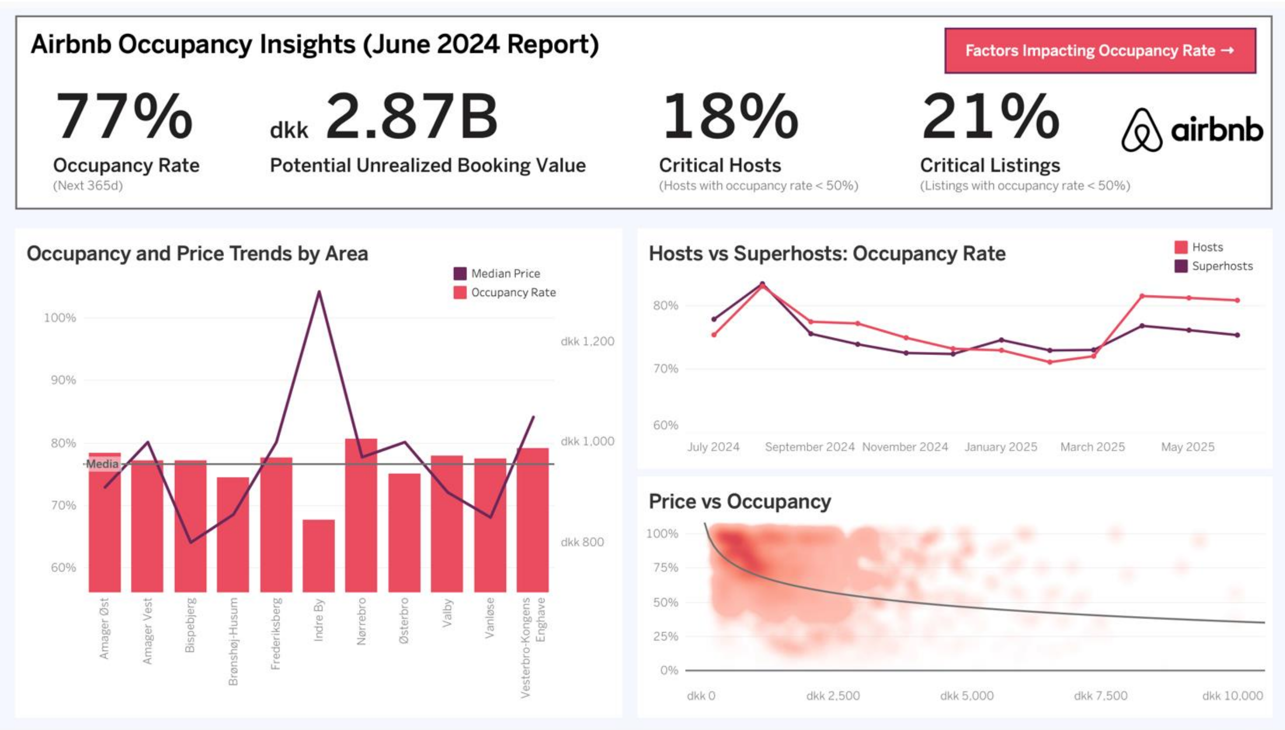
3.3 Other limitations

Beyond the nature of the dataset and the interpretation, there are broader contextual factors that could influence the results of this analysis:

- **Market competition:** listings' performance is not only influenced by their internal characteristics but also by external competition. Competing platforms or other rental options in the same area could impact Occupancy Rates and pricing strategies.
- **Economic factors:** macroeconomic conditions, such as inflation, tourism trends, or changes in consumer spending habits, could also play a significant role in shaping booking behaviors and availability patterns.

While these external factors are beyond the scope of this analysis, they are important considerations for Airbnb when interpreting the findings and implementing any recommendations for them and for the hosts.

3. First View – Airbnb Occupancy Insights



The primary objective of the first dashboard view (Figure 6 in the Appendix) is to provide the management with insights into the overall Occupancy Rate on the platform, highlighting areas for improvement. In particular, it seeks to analyze the correlation between Occupancy Rates and apartment price levels. In addition, the dashboard showcases the seasonality effect in the number of booked nights throughout the year.

The data show clear trends, distinguishing low season months such as December and February, when the Occupancy Rate falls below the average of 77%, from high season months such as August and April, when the rate exceeds the average. This variation is obviously due to tourists' preference for warmer months.

By presenting these insights, the dashboard provides the management team with the tools to make informed decisions about pricing strategies during low occupancy periods.

4.1 Key Metrics

At the top of the dashboard, strategically placed to grab the management's attention first, are four fundamental statistics that highlight areas where Airbnb has the potential to improve. This placement enables the management to quickly identify Key Performance Indicators (KPIs) that require strategic action.

1. **(Estimated) Occupancy Rate for the next 365 days and Unrealized Booking Value**

   The first two metrics are directly aligned with Airbnb's Key Business Indicators (KBIs) as referenced in the company's financial reports (*Gross Booking Value* [GBV] and *Nights and Experiences Booked,* as described in the Q3 2024 report).

   An Occupancy Rate of 77% implies an unoccupancy of 23%. This level of unoccupancy translates into an unrealized value of approximately DKK 2.87 billion. While a precise calculation is difficult to perform due to the high volatility of the commission percentages charged by Airbnb to hosts and guests (which depend on several factors), an *estimated* average commission of 15% suggests that the company is losing more than DKK 431 million in annual potential revenue due to these unoccupied properties.

   These figures emphasize the importance of optimizing occupancy to improve revenue generation.

2. **Critical Hosts (%) and Critical Listings (%)**

   According to the previsions, about 18% of hosts have an Occupancy Rate below 50%, and 21% of listings remain unbooked for more than half of the year. We have highlighted these two statistics because we believe Airbnb should take some kind of action in raising awareness among hosts about the factors preventing them from increasing bookings for their properties.

   Anyway, these issues are analyzed in more detail in the second view of the dashboard, which includes graphs illustrating the relationship between occupancy and various variables such as host response time and the availability of instant bookings for a listing.

4.2 Graph 1 – Hosts vs Superhosts and seasonality

The line graph (Figure 7 in the Appendix) shows the average occupancy of properties hosted by regular hosts (red) and super hosts (purple) throughout the year. Although the calendar for each property includes data for the last two days of June 2024, we have chosen not to include them in the graph. This decision was made because the occupancy data for this month would have been significantly less reliable than the others, making the June 2024 rate equal to 93%.

We expected different results from this graph than we got.

Firstly, we were quite surprised to see that the Occupancy Rate of superhosts is not always higher than the one of regular hosts. In fact, it is often lower. We attribute this to the likelihood that superhosts' properties may be more expensive and that superhosts may be more selective in managing their bookings.

Secondly, we expected a more pronounced seasonal effect, particularly in the colder months of December and January. This leads us to view the rental market in Copenhagen as a competitive and stable one, where price plays a crucial role for tourists - especially given that the city consistently ranks among the most expensive in Europe.

4.3 Graph 2 – How Prices influence Occupancy

The heatmap (Figure 8 in the Appendix) is very intuitive: it shows the (negative) relationship between price and occupancy. The included logarithmic trend line clearly shows that as the price of a property (x-axis) increases, its occupancy (y-axis) decreases.

This result is exactly what we expected, and a chart like this reinforces our hypothesis: by lowering the price level of listings, we can increase their occupancy.

4.4 Graph 3 – Overview of Price and Occupancy per Area

In this graph (Figure 9 in the Appendix), we have created an overview to allow users of the dashboard to quickly grasp the price-occupancy situation in all neighbourhoods within the metropolitan area of the city. It is a two-axis graph: the percentage on the left axis represents the Occupancy Rate, while the values on the right axis indicate the median price per night.

We considered the median to be more reliable than the average in this case, as it is more robust to outliers (some properties have significantly higher nightly prices than others).

The neighbourhoods are all fairly close to the average Occupancy Rate of 77%, with Norrebrø being highly occupied (81%) and Indre By (the city-centre district) having a much lower availability (68%). These Occupancy Rates for Indre By and Norrebrø are easily explained by the fact that Indre By is by far the most expensive area in Copenhagen, while Norrebrø offers excellent value for money.

The phenomenon observed in Indre By did not surprise us much, as an article published by Airbnb in September 2019 states that three out of four guests choose to stay outside the city centre. This also supports our idea of applying a small additional percentage discount for less popular neighbourhoods (those with fewer listings) to benefit from the volume effect.

Finally, this chart confirms us that price and occupancy are highly negatively correlated.

5. Second View – Factors impacting Occupancy Rate

The purpose of this dashboard is to identify and analyze key factors that can impact Occupancy Rates independently of pricing. Even though, as we have seen in the previous dashboard, pricing adjustments are a direct way to influence occupancy, there are cases where maintaining the current pricing structure is necessary due to market conditions or host preferences. This dashboard explores alternative elements, such as review scores, host response times, and booking modes, to uncover actionable insights that hosts, and Airbnb can leverage to boost Occupancy Rates without altering prices.

5.1 Key Metrics

To set the foundation for the analysis, we start with three key metrics displayed as BANs. These metrics highlight the potential impact of three factors – review scores, host response time, and booking modes – on Occupancy Rates. Each metric quantifies how these factors can influence occupancy, providing a basis for the deeper insights highlighted by the dashboard's graphs.

1. **Impact of Review Scores (+10%)**
   Listings with review scores above 4 exhibit a 10% higher average Occupancy Rate compared to those with review scores below 4. This finding underscores the critical role of guest satisfaction in attracting future bookings. High ratings not only reflect positive guest experiences but also enhance the trust and appeal of listings.

2. **Impact of Host Response Time (+23%)**

Listings where hosts respond to inquiries within a day show a 23% increase in Occupancy Rates compared to those where the response time exceeds 24 hours. This suggests that prompt communication fosters trust and convenience, making guests more likely to book.

3. **Impact of Booking Mode (-3%)**
Listings without the *Instant Book* feature exhibit a 3% higher average Occupancy Rate compared to those with this option. This may indicate a preference among some guests for more interactive communication with hosts before finalizing a booking.

These BANs highlight the significance of factors beyond pricing in driving Occupancy Rate. Next, we delve into each graph to better understand the data and the actionable insights they offer.

5.2 Graph 1 – Rating vs Occupancy

This graph (Figure 11 in the Appendix) visualizes the relationship between the average listing Occupancy Rate (y-axis) and the review score rating (x-axis). Each data point represents a listing, with the density of points displayed using a heatmap-style representation. The darker the colour, the higher the concentration of listings with similar Occupancy Rates and review scores. Additionally, a trend line is plotted to show the general direction of the relationship between the two variables.

The graph reveals a clear positive correlation between higher review ratings and higher Occupancy Rates. Listings with review scores above 4 consistently achieve Occupancy Rates above the overall average, while those with lower ratings exhibit significantly lower Occupancy Rates.

This suggests that well-rated listings are more attractive to guests, reinforcing the importance of maintaining high-quality service and guest satisfaction to maximize occupancy. Hosts should focus on improving key aspects that compose the overall rating of their listings in order to improve their Occupancy Rates.

More specifically, the review score is composed of various subcategories, including accuracy, cleanliness, check-in process, communication, location, and more. While it might be tempting (and amusing) to suggest that hosts move their listings closer to the city centre to improve their location rating, this is clearly impractical. Instead, Airbnb could focus on actionable strategies to help hosts improve ratings in areas they can control.

For example:

- Cleanliness: Encourage hosts to hire professional cleaning services or implement stricter cleaning protocols to enhance guest satisfaction.
- Communication: Provide tools or training for hosts to respond promptly and effectively to guest inquiries.

- Check-in Process: Suggest the adoption of user-friendly, flexible check-in methods, such as keyless entry or self-check-in options.
- Accuracy: Guide hosts in ensuring that listing descriptions, photos, and amenities are accurate and up to date.

By focusing on these controllable factors, Airbnb can help hosts improve their overall review scores, leading to higher Occupancy Rates and better guest experiences. Additionally, offering workshops, webinars, or even a "best practices" guide tailored to these areas could be a valuable resource for hosts.

5.3 Graph 2 – Occupancy and Host Response Time

This bar chart (Figure 12 in the Appendix) represents the relationship between Occupancy Rates (y-axis) and the response time of hosts to guest inquiries (x-axis). The categories of response time range from *a few days or more* to *within an hour*, and the chart highlights the average Occupancy Rates for each category. A horizontal line represents the overall average Occupancy Rate as a benchmark.

The chart clearly demonstrates a strong correlation between faster response times and higher Occupancy Rates. Listings where hosts respond within a day achieve an average Occupancy Rate of 72%, compared to just 49% for those responding in *a few days or more*. This 23% increase underscores the significant impact that host responsiveness has on guest decisions. Faster responses likely instill confidence and trust in potential guests, improving the likelihood of booking.

Given the crucial role of response time, Airbnb could encourage hosts to enhance their responsiveness by leveraging technology. For instance:

- AI-Powered Chatbots – Host could integrate AI chatbots to handle initial guest inquiries and provide immediate answers to frequently asked questions. These chatbots could bridge the gap until the host is available to personally join the conversation.
- Hybrid communication model – Combining AI efficiency with human interaction could offer the best of both worlds, ensuring guests feel heard and valued while receiving swift responses.
- Push notifications and reminders – Airbnb could implement automated reminders for hosts, prompting them to respond promptly to new inquiries.

These strategies would improve guest satisfaction in terms of time of responsiveness but also help increase Occupancy Rates across listings, benefiting both hosts and Airbnb.

5.4 Graph 3 – Occupancy and Booking Mode

This last bar chart (Figure 13 in the Appendix) explores the relationship between Occupancy Rates (y-axis) and the type of booking mode (x-axis): Not Instant Booking and Instant Booking. The chart highlights the average Occupancy Rates for each category.

The data reveals that listings without the Instant Booking feature achieve a slightly higher average Occupancy Rate (approximately 77%) compared to listings with Instant Booking enabled (around 74%). While the difference is not significant (around 3%), it suggests that guests may prioritize interacting with hosts before committing to a booking. This preference could be attributed to guests feeling more secure when communicating directly with the host, as it allows them to confirm the authenticity of the listing and reduce concerns about potential scams or misrepresentation.

Although Instant Booking offers convenience, Airbnb could focus on providing tools to enhance communication between guests and hosts for listings without this feature. Possible strategies include:

- Verified messaging systems: Encourage hosts and guests to utilize Airbnb's platform for secure and transparent communication, offering reassurances about the authenticity of listings.
- Trust-building features: Introduce features that emphasize host credibility, such as badges for verified hosts or tools that highlight detailed reviews and response times.
- Optional video calls: Allow hosts and guests to engage in optional video calls through the platform, providing an additional layer of trust.

By emphasizing safety and trust in the booking process, Airbnb can cater to guest preferences for interaction without undermining the convenience of Instant Booking. This balanced approach could help maintain and even improve Occupancy Rates.

6. Further ideas and strategies

This section presents additional ideas and strategies that could complement the findings of our analysis. These concepts are aimed at refining the approach to analysing and optimizing Airbnb's pricing and occupancy trends while considering alternative perspectives for a more comprehensive understanding of the data.

6.1 Interpreting the concept of overpricing

The concept of "overpriced" listings requires careful interpretation within a localized context. If listings with low Occupancy Rates are consistently underperforming when compared to similar listings in the same cluster (e.g., listings in the same neighborhood or with similar features), this may indicate that the pricing is a significant barrier. However, pricing is not always the sole factor.

For example, consider a cluster of properties that accommodates a high number of guests. These properties tend to have higher prices due to their size, amenities, or unique features. If the entire cluster exhibits low Occupancy Rates, it could suggest underlying issues such as limited market demand for large properties, unfavorable seasonal trends, or even competition from alternative accommodations (e.g., hotels).

This insight underscores the importance of contextualizing pricing strategies. Instead of solely focusing on reducing prices, Airbnb could work with hosts to explore other factors, such as improving the quality of amenities, enhancing listing descriptions, or targeting specific customer segments. Identifying overpriced listings should involve a comprehensive evaluation of both individual performance and the broader cluster performance, ensuring that any recommendations align with market realities.

6.2 Underpricing

While identifying overpriced listings is critical, there is an equally important opportunity to address listings that are potentially underpriced. Within each cluster, some listings may offer lower-than-average prices compared to the cluster mean, achieving high Occupancy Rates. These listings represent opportunities for revenue optimization.

The strategy would involve the following steps:

- **Defining the cluster price range:** Calculate the average rental price for each cluster and set a target price range (mean ± a specified margin). The range represents a practical benchmark for aligning prices within a cluster.
- **Identifying underpriced listings:** Listings priced significantly below the cluster's target price range would be flagged as underpriced.
- **Adjusting pricing:** Airbnb could recommend increasing the price of these listings to bring them closer to the cluster average. Assuming no changes in market demand or external competition, these listings would still remain attractive due to their relatively low pricing within the cluster.

The broader goal of this strategy is to narrow the distribution of listing prices within each cluster. By creating a bell curve that is more concentrated around the cluster mean, Airbnb could help stabilize pricing dynamics while improving overall revenue. For both overpriced and underpriced listings, this approach aims to optimize prices without disrupting market equilibrium.

6.3 Alternative analytic opportunity

One of the limitations of this analysis was the reliance on long-term availability data (365 days). However, an alternative analytical approach could focus on short-term booking behaviors by analyzing availability also within shorter timeframes (e.g., 1, 2, or 3 months). Indeed, short-term availability data could provide valuable insights into seasonal trends, last-minute booking patterns, and the impact of booking time caps. For example:

- Listings with low Occupancy Rate in long terms such as a year, but high Occupancy Rate for the next 1–3 months may indicate higher demand during seasonal period or presence of booking time cap.
- Alternatively, properties with significant availability also within short future periods might highlight underperformance due to factors like poor marketing, high prices, or competition.

While these short-term trends would complement the insights gained from long-term data, the primary goal of this project was not to explore host-imposed restrictions or booking time caps. Instead, the focus remained on understanding patterns in performance represented by the assumption made over long-term availability. Future analyses could delve deeper into these alternative perspectives to offer a more holistic understanding of host and guest behaviors on the Airbnb platform.

7. Conclusion

This report explored strategies to enhance Occupancy Rates for Airbnb listings in Copenhagen, focusing on two distinct approaches: optimizing pricing dynamics and leveraging non-pricing factors. The first dashboard view highlighted the inverse relationship between pricing and occupancy, showcasing how strategic price adjustments could significantly improve booking rates and maximize revenue. The second view delved into factors beyond pricing, such as host response times, guest review scores, and booking modes, revealing their notable influence on occupancy.

Our analysis underscores the critical role of balancing pricing strategies with non-pricing enhancements to optimize performance. By addressing these factors, Airbnb and its hosts can better align their offerings with guest preferences, fostering improved booking outcomes. While the analysis was based on available data of the three given datasets with inherent limitations, the insights provide a strong foundation for understanding key drivers of Occupancy Rates and guiding future decision-making processes.

References

- Airbnb, *Airbnb in Copenhagen: 3 out of 4 guests stay outside the city centre*, Airbnb News, (n.d.). https://news.airbnb.com/airbnb-in-copenhagen-3-out-of-4-guests-stay-outside-the-city-centre/
- Airbnb, *Q3 2024 shareholder letter,* Airbnb, 2024. https://s26.q4cdn.com/656283129/files/doc_financials/2024/q3/Airbnb_Q3-2024-Shareholder-Letter_Final.pdf
- Airbnb, *Airbnb's impact in Copenhagen*, Airbnb News, 2019. https://news.airbnb.com/wp-content/uploads/sites/4/2019/06/Report-Airbnb%E2%80%99s-Impact-in-Copenhagen.pdf
- Emre Guerbuezer, *Airbnb in Copenhagen, Denmark: The Short-Term Rent Gap, Gentrification & the Housing Market*, Lund Universiteit (Department of Human Geography), 2021. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9049686&fileOId=9049692

Appendix

Figure 1 - ER model of the database

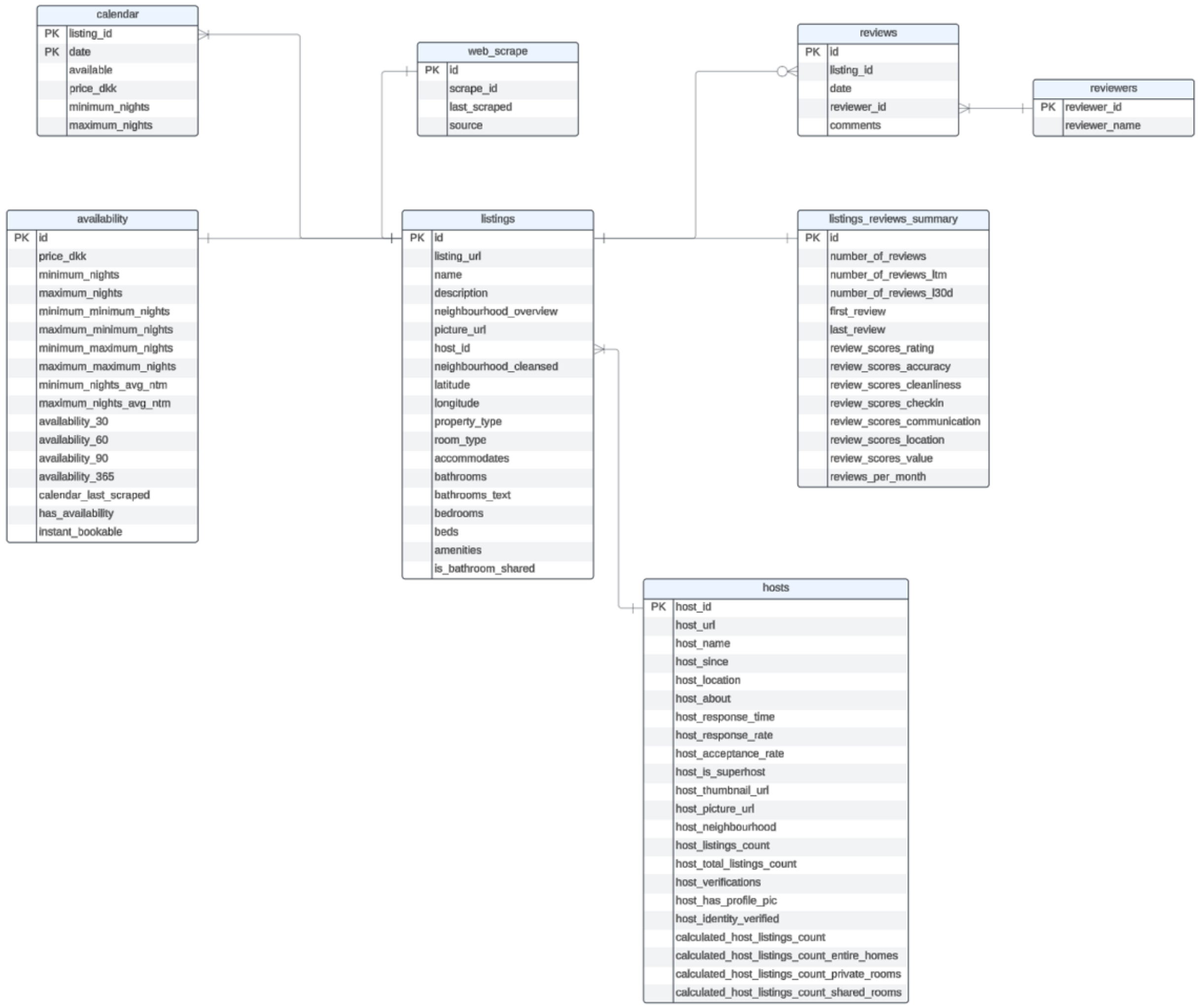


Figure 2 - calendar2024.csv data dictionary

	calendar2024.csv	
	listing_id	Unique identifier of the listing.
	date	The specific date of the calendar entry.
	available	Whether the listing is available on the specified date.
	price	The price of the listing in the specified date.
	adjusted_price	-
	minimum_nights	The min number of nights required for a booking.
	maximum_nights	The max number of nights allowed for a booking.

Figure 3 - reviews2024.csv data dictionary

reviews2024.csv	
listing_id	Unique identifier for the listing.
id	Unique identifier for the review.
date	The date the review was posted
reviewer_id	Unique identifier for the reviewer.
reviewer_name	Name of the reviewer.
comments	Text content of the review.

Figure 4 - listings2024.csv data dictionary

listings2024.csv	
id	Unique identifier for the listing.
listing_url	URL of the listing's page.
scrape_id	Unique identifier for the scraping session.
last_scraped	The date when this listing was scraped.
source	Source of the listing data.
name	Name of the listing.
description	Detailed description of the listing.
neighborhood_overview	Host's description of the neighbourhood.
picture_url	URL of the listing's main picture.
host_id	Unique identifier for the host.
host_url	URL of the host's profile page.
host_since	The date the host account was created.
host_location	Location of the host.
host_about	Host's self-description.
host_response_time	Average response time of the host.
host_response_rate	Percentage of messages responded to by the host.
host_acceptance_rate	Percentage of booking requests accepted by the host.
host_is_superhost	Wheter the host is a "Superhost"
host_thumbnail_url	URL of the host's thumbnail image.
host_picture_url	URL of the host's profile picture.
host_neighbourhood	Host's neighborhood.
host_listings_count	Number of active listings managed by the host.
host_total_listings_count	Total number of listings ever managed by the host.
host_verifications	Methods used to verify the host's identity.
host_has_profile_pic	Indicates whether the host has a profile picture.
host_identity_verified	Indicates whether the host's identity is verified.
neighbourhood	Neighborhood where the listing is located.
neighbourhood_cleansed	Neighborhood as defined by open digital shapefiles.
neighbourhood_group_cleansed	Neighborhood group as defined by open digital shapefiles.
latitude	Latitude of the listing's location.
longitude	Longitude of the listing's location.
property_type	Type of property.
room_type	Type of room offered.
accommodates	The maximum capacity of the listing.
bathrooms	Number of bathrooms available in the listing (old scrapes).
bathrooms_text	Number of bathrooms available in the listing (new scrapes).
bedrooms	Number of bedrooms in the listing.
beds	Number of beds in the listing.
amenities	List of amenities offered in the listing.
price	Price per night in the listing (local currency)..
minimum_nights	The min number of nights required for a booking.
maximum_nights	The max number of nights allowed for a booking.
minimum_minimum_nights	Minimum of the minimum nights required (in the next 12 month).
maximum_minimum_nights	Maximum of the minimum nights required (in the next 12 month).
minimum_maximum_nights	Minimum of the maximum nights allowed (in the next 12 month).
maximum_maximum_nights	Maximum of the maximum nights allowec (in the next 12 month).
minimum_nights_avg_ntm	Average minimum nights required (in the next 12 months).
maximum_nights_avg_ntm	Average maximum nights allowed (in the next 12 months).
calendar_updated	-
has_availability	Wherer the listing has availability.
availability_30	Number of days the listing is available in the next 30 days.
availability_60	Number of days the listing is available in the next 60 days.
availability_90	Number of days the listing is available in the next 90 days.
availability_365	Number of days the listing is available in the next 365 days.
calendar_last_scraped	Date the calendar was last scraped.

number_of_reviews	The number of reviews the listing has.
number_of_reviews_ltm	The number of reviews the listing has (last 12 months).
number_of_reviews_l30d	The number of reviews the listing has (last 30 days).
first_review	Date of the first review.
last_review	Date of the last review.
review_scores_rating	Overall rating score from reviews.
review_scores_accuracy	Accuracy rating score from reviews.
review_scores_cleanliness	Cleanliness rating score from reviews.
review_scores_checkin	Check-in process rating score from reviews.
review_scores_communication	Communication rating score from reviews.
review_scores_location	Location rating score from reviews.
review_scores_value	Value-for-money rating score from reviews.
license	License number.
instant_bookable	Whether the listing can be instantly booked.
calculated_host_listings_count	Number of listings the host has in the current scrape.
calculated_host_listings_count_entire_homes	Number of entire homes the host has in the current scrape.
calculated_host_listings_count_private_rooms	Number of private rooms the host has in the current scrape.
calculated_host_listings_count_shared_rooms	Number of shared rooms the host has in the current scrape.
reviews_per_month	Average number of reviews per month.

Figure 5 - From raw data to structured data flowchart

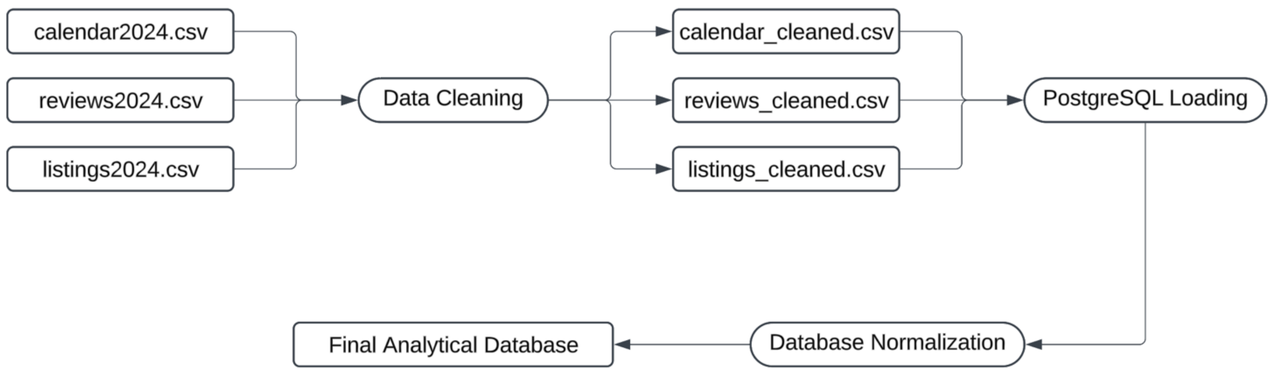


Figure 6 - Airbnb Occupancy Insights (First View)

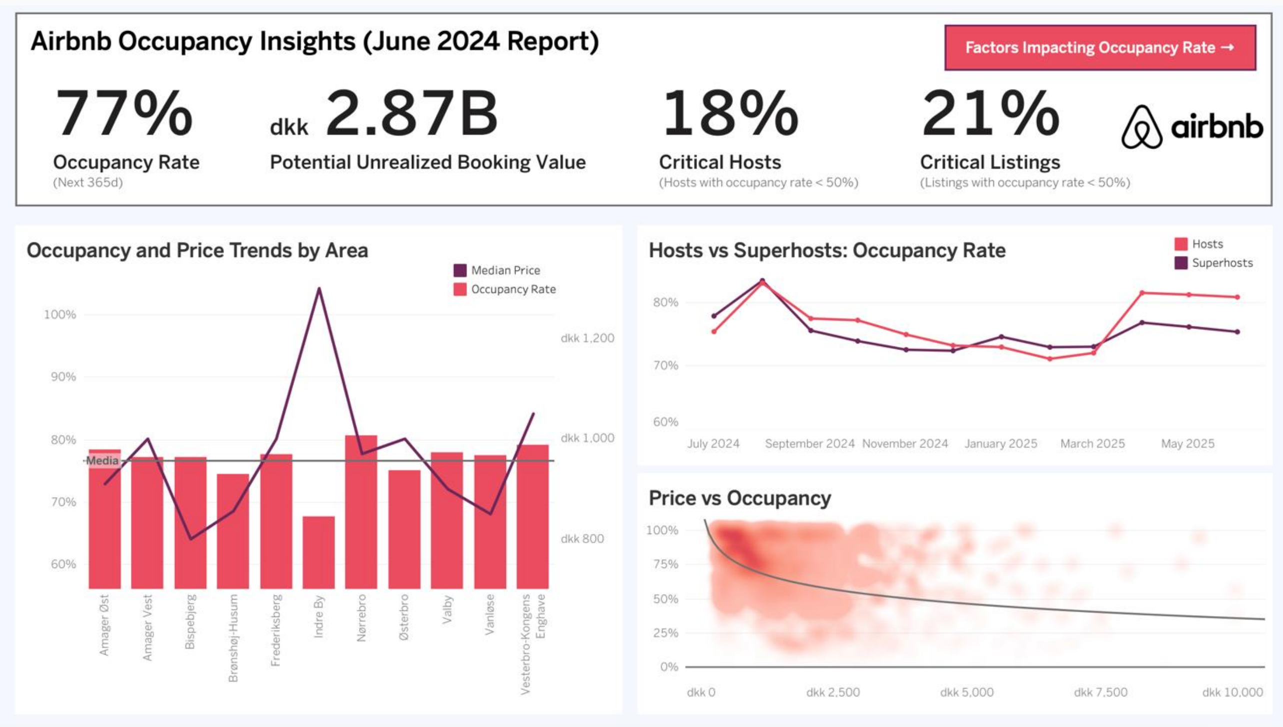


Figure 7 - Hosts vs Superhosts and Seasonality

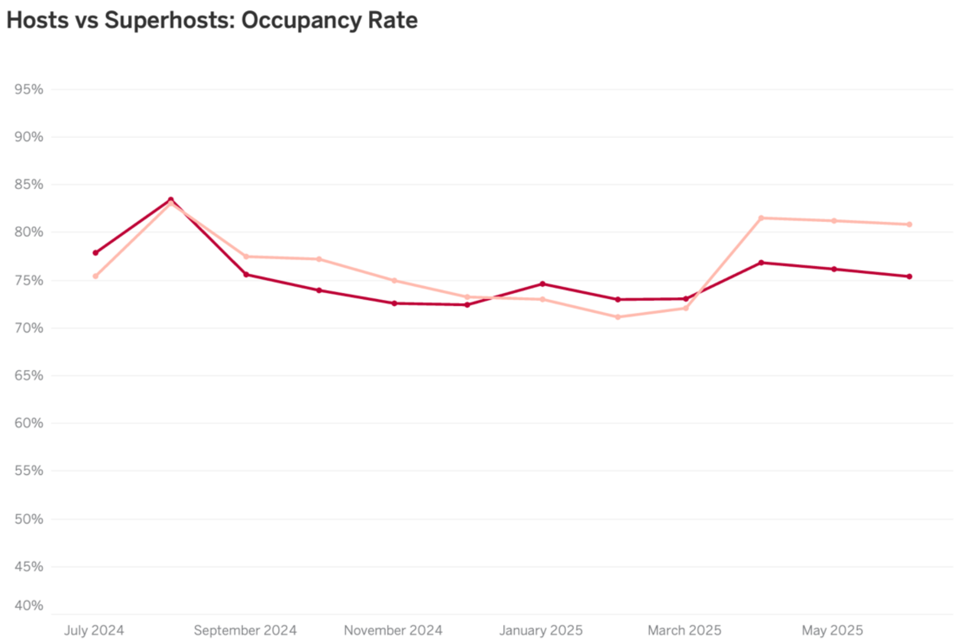


Figure 8 – How Prices influences Occupancy

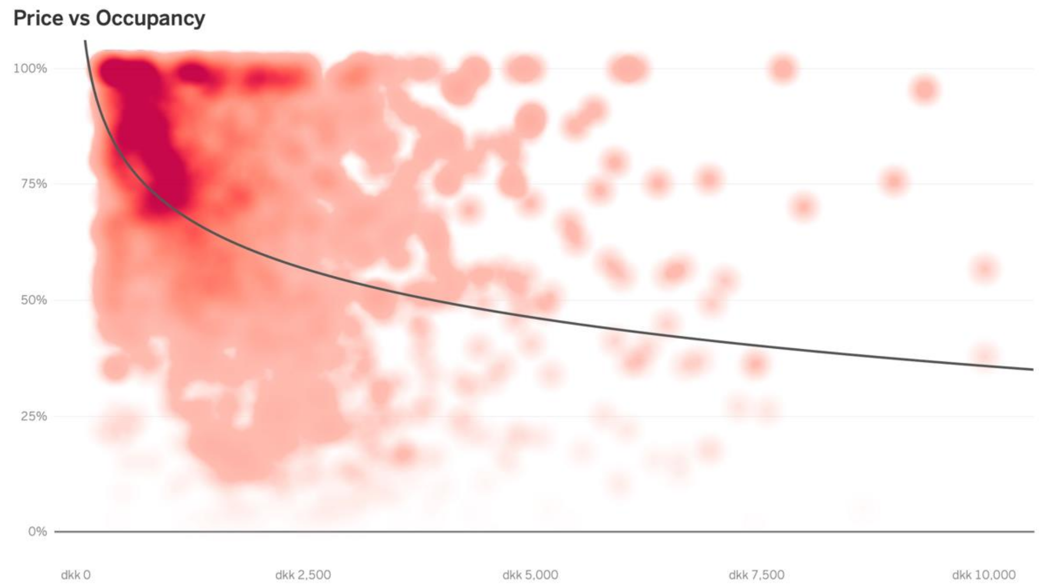


Figure 9 - Overview of Price and Occupancy per Area

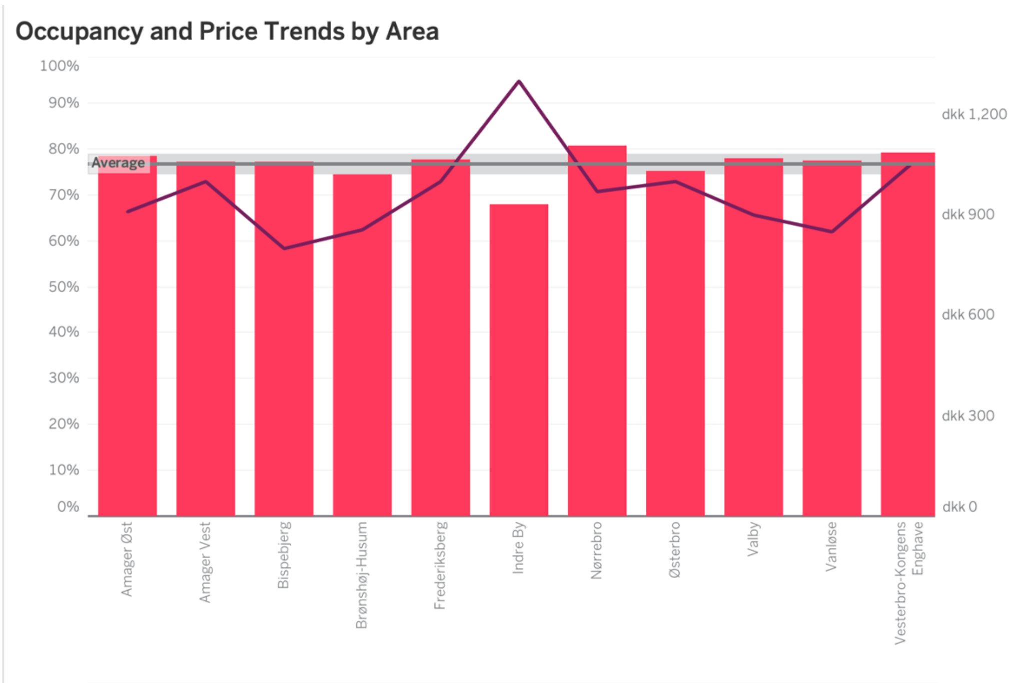


Figure 10 - Factors impacting Occupancy Rate (Second View)

Figure 11 - Rating vs Occupancy Trend

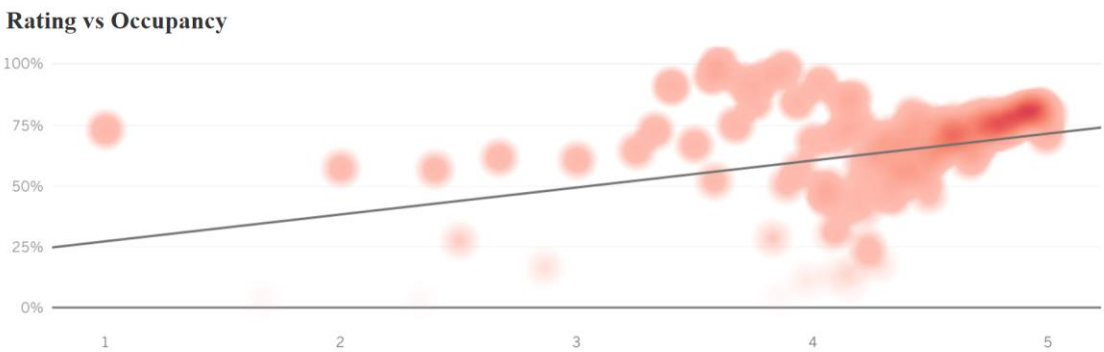


Figure 12 - Occupancy vs Host Response Time Trend

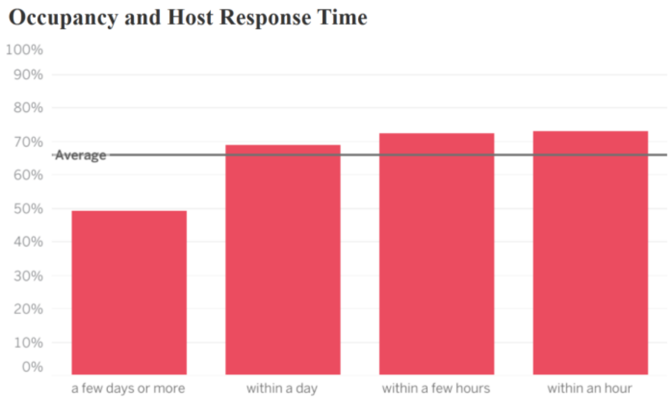


Figure 13 - Occupancy vs Booking Mode Trend