Data-driven Occupancy Optimization to Mitigate Airbnb's Pressure on Copenhagen's Housing Market

---

Final Exam Hand In

Data Management and Visualization

Abid Hussain and Travis Greene

Students: Johan Schommartz, Sebastian Uedingslohmann, Tim Rößling

IDs: 175871, 175867, 175881

Pages: 21

Character Count: 41,329

Date: January 6, 2025

Tableau Public: `https://public.tableau.com/app/profile/tim.r.ling/viz/airbnb_17354823633220/NeighborhoodView?publish=yes`

GitHub: `https://github.com/tiro27cbs/Airbnb_Copenhagen`

Copenhagen Business School
Department of Digitalization
2000 Frederiksberg
rg.digi@cbs.dk

Executive Summary

Problem addressed. This project focuses on the challenges in Copenhagen's housing market and their impact on Airbnb. Cities like New York, Barcelona, and Paris have already imposed strict rules on short-term rentals to address the housing shortage. The aim here is to avoid such restrictions or bans. The business goal is not to grow the market by adding more listings but to maximize the occupancy of existing properties. Expanding the short-term rental market without negatively impacting the regular housing market is crucial, as stricter regulations might otherwise be imposed.

Data implications. First, data is cleansed to improve its quality by identifying missing values, outliers, errors, and modifying data types. Upon eliminating irrelevant data and enriching and transforming the given datasets, the final data model is developed. The data shows that the market is dominated by small apartments within a moderate price range. Listings are heavily concentrated around the city's central neighborhoods like Nørrebro, Versterbro, and Indre By, where both demand and prices tend to be higher than in peripheral neighborhoods, which highlights the pressure on central located properties. Since the data was scraped for a future time period, it only reflects booking patterns rather than actual occupancy, allowing only a relative analysis of success factors based on listing attributes. Nonetheless, clear booking trends emerge, such as increased occupancy at the start of the winter semester and during Christmas holidays. While the demand depends on the season, the overall pricing of the listings remains constant. Additionally, specific amenities and host attributes, such as faster response times and profile pictures, are shown to have a notable impact on occupancy. By identifying key attributes and trends, Airbnb can develop data-driven solutions to support hosts and optimize performance.

Business implications. This project delivers an interactive dashboard that provides a comprehensive overview of the Airbnb market in Copenhagen. The dashboard features a multipage design, providing an overarching market view and a detailed action summary to support data-driven decision-making. Its interactive and customizable visuals ensure usability while prioritizing clarity and avoiding information overload, enabling Airbnb to derive measures to improve occupancy. Airbnb could integrate real-time booking data to enable continuous monitoring of occupancy rates across different neighborhoods and time periods. This functionality allows hosts to receive dynamic notifications on current market trends and strategies for optimizing occupancy and pricing. Additionally, the dashboard offers valuable insights into how specific listing and host attributes influence occupancy rates, which should be communicated to hosts. These insights can help address underperforming neighborhoods and reduce the concentration of bookings in central areas. Airbnb should further support hosts by streamlining pricing adjustments, facilitating more effective use of monetary incentives to enhance occupancy rates.

Contents

1 Introduction

This project bridges two distinct yet interdependent roles: the data scientist and the business analyst. These roles come together in the development of an interactive Tableau dashboard designed to address a specific business problem. The data scientist focuses on understanding, cleansing, and analyzing data, as well as building a robust Tableau dashboard to present meaningful insights. The business analyst then interprets these insights within a business context to derive actionable decisions that address the predefined problem. Together, these perspectives form a comprehensive approach that integrates technical expertise with strategic business understanding. In this project, the dashboard is specifically tailored to provide an actionable overview for the analysts in the Airbnb's regional business development and public affairs teams.

The business problem addressed in this project is derived from Brian Chesky's public commitment to building sustainable and cooperative relationships with local communities by actively addressing their local challenges (Bloomberg, 2023). In Copenhagen, this involves maintaining business growth while mitigating the pressure on the local housing market by trying to reduce the conversion of private living space into Airbnbs. The primary objective of the dashboard is to support these goals by presenting options to optimize the utilization of Airbnb's existing listing supply. The utilization is measured by the occupancy rate, reflecting the proportion of nights booked relative to the total availability of a listing over a year. An occupancy rate increase allows for greater revenue without increasing the number of rental properties converted into short-term Airbnb listings, thereby reducing pressure on the housing market.

By highlighting areas with low occupancy rates, the dashboard provides actionable insights into which neighborhoods in Copenhagen have the greatest potential for optimization. Additionally, it equips the local team to implement instruments that increase market-wide occupancy and offers guidance on how to collaborate with hosts, encouraging them to adopt practices that boost occupancy. These insights allow Airbnb to optimize utilization were most needed to support sustainable growth and foster positive relationships with the Copenhagen's authorities.

In summary, this project integrates technical data analysis with business strategy. Through data preparation and cleansing, exploratory data analysis (EDA), and dashboard design, the data scientist ensures reliability, accessibility and understandability of insights. Meanwhile, the business analyst uses these insights to addresses Airbnb's business problem in Copenhagen and make actionable decisions, that reduce the pressure on the housing market. Together, these roles illustrate how data-driven decision-making can align with organizational objectives to achieve impactful business outcomes.

2 Data Preparation and Cleansing

Three datasets are provided for this project: calendar2024.csv, listings2024.csv, and reviews2024.csv. With all three datasets being linked through the listing_id and each contributing unique information.

The listings2024.csv dataset represents all active listings captured during the data scraping on June 29, 2024. This dataset provides all the details, like average price, rating, room type and location, around each listing, having 75 columns and 20.9 thousand rows. The calendar2024.csv dataset includes daily data on availability, price and booking restrictions for each listing within the timeframe from June 29, 2024 to June, 28 2025. It contains 7 columns and approximately 7.6 million rows, because each listing has around 365 rows. Lastly, the reviews2024.csv dataset contains information on all written reviews posted since Airbnb began its operations in Copenhagen in 2010 until the scarping date, comprising 6 columns and 366 thousand rows.

Together, these datasets offer a comprehensive view of Airbnb's operations in Copenhagen, while posing challenges related to data quality and temporal diversity. The varying time spans across the datasets necessitate careful handling to ensure consistency and extract meaningful insights.

To enable the EDA and dashboard creation, data cleaning and transformation is essential. Firstly, all columns were reviewed to ensure consistent and accurate data types. This included converting the price column in the listings2024.csv dataset from string format, containting a "$" sign into floats, representing DKK, and transforming the calendar2024.csv dataset's date column into a datetime format. Additionally, the spelling of neighborhood names in the listings2024.csv dataset was corrected, for example, correcting "Nrrebro" to "Nørrebro."

An initial data analysis revealed several quality issues requiring attention. Using a heat map, which can be seen in Figure 1, all missing data points in each column of the listings2024.csv dataset are identified. The heat map reveals that the columns neighbourhood_group_cleansed, calendar_updated, and license do not contain any data. Consequently, these three columns are deemed irrelevant for the data analysis and are being dropped.

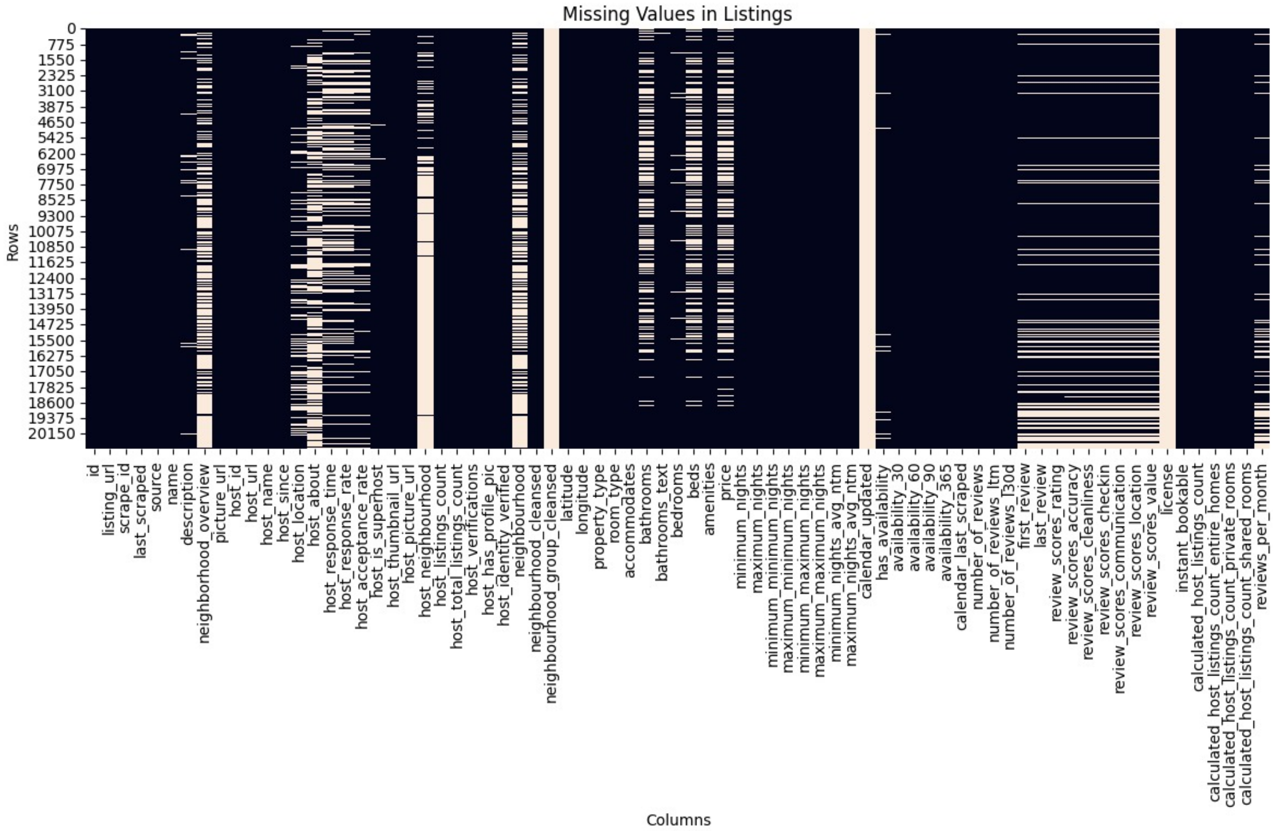

Figure 1: Boxplots of review scores subcategories.

Further inspection of the pricing data in the `listings2024.csv` dataset exposed significant outliers, which are defined as observations which are not consistent with the rest of the data and can skew the analysis (Chatfield, 1986). Sorting by descending price revealed extreme outliers such as a regular three-bed listing priced at 100,000 DKK per night. Manual verification of these listings via their URLs confirmed them as erroneous and misleading to the data analysis required to solve the business problem. Similarly, sorting the `calendar2024.csv` dataset by ascending price identified listings priced at 21 DKK per night, which was unrealistically cheap and inconsistent with the pricing data from the `listings2024.csv` dataset. The outliers on the low end of the price spectrum often differed by a factor of 7.46 between the two datasets, which coincidentally matches the exchange rate of DKK/USD on June, 29 2024. These inconsistencies indicate that low-outlier listings in the `calendar2024.csv` dataset are priced in USD instead of DKK, necessitating their exclusion from the dataset.

These outliers were removed using the interquartile range (IQR) measure. Listings with prices exceeding eight times the IQR from the 75th or 25th percentiles were removed. Using a larger multiplier like eight times the IQR allowed to target only the most extreme and clearly erroneous data points while preserving the integrity of the dataset. This approach aligns with objective of excluding outliers that significantly skew analysis and visualizations, without unnecessarily discarding valid data that falls outside standard thresholds but still reflects real-world phenomena. Additionally, the rows in the `listings2024.csv` dataset that are missing pricing data were dropped to ensure high data quality. The `calendar2024.csv` and `reviews2024.csv` datasets are accordingly filtered to exclude all rows in those two datasets that are associated with `listing_id`s not present in the cleaned `listings2024.csv` dataset.

To enhance the analytical value of the datasets, additional attributes were generated from the existing data. An `occupancy_rate` column was calculated by dividing the total number of booked days for each listing seen, retrieved from the availability column of in the `calendar2024.csv` dataset, by the number of total days available, expressed as a percentage. This metric was added to the `listings2024.csv` dataset via ID mapping, offering insights into the utilization of each listings.

Further data enrichment focused on amenities. The `amenities` column initially contained thousands of distinct entries, making it necessary to standardize these values to analyze the relationships between amenities and other metrics effectively. The dataset included 3,377 unique amenities, reduced from 4,446 after removing missing values and outliers. Using regular expressions, these were consolidated into 129 standardized amenities across all listings. After replacing the original values in the `amenities` column with these standardized entries, a new melted dataframe was created, containing each listing ID and its amenities in separate rows. Additionally, a third column was added to this dataframe, categorizing each amenity into one of eleven overarching categories.

Finally, a new dataframe was created containing each date from the `calendar.csv` dataset along with the corresponding average occupancy. This value was calculated using the `available` column by grouping the data by `date` and determining the average occupancy as the ratio of non-available listings to the total number of listings.

These processes significantly modified the three datasets. The `listings2024.csv` dataset was compressed to 73 columns and 13.5 thousand rows, with the `calendar2024.csv` dataset's number of rows also being reduced by 36% to only 4.9 million. The number of rows in the `reviews` dataset was only compressed by 28% to 264 thousand rows. While the datasets have been cleaned and enriched, the following EDA will reveal which columns are actually relevant for solving Airbnb's business problem and will end up in the final database.

3 Exploratory Data Analysis

3.1 Distributions and Descriptive Statistics

The EDA plays a vital role in gaining an initial understanding of the data and maximizing its value through discovering patterns and anomalies (Jebb et al., 2017). Upon cleansing and obtaining a first overview, the analysis aims to extract insights from the data that can guide further exploration and help address the business problem.

The data was last scraped on June 29, 2024, as indicated in the `last_scraped` column. The `calendar2024.csv` file consists of availability data for these listings, covering a period from June 29, 2024, to June 29, 2025, derived from the `date` column. It is important to note that the availability data was scraped at a specific point in time and does not reflect the actual availabilities from the scrape date onwards. Therefore, any conclusions drawn from the occupancy data should be limited to comparing listings, rather than interpreting real occupancy trends. The `reviews2024.csv` file contains a total of 366,636 reviews, with the earliest review dating back to July 25, 2010 and the most recent ones recorded on the day of the scrape.

Looking at the distribution of `price` in the `listings2024.csv` dataframe using the `describe()` method provided by the library `pandas`, it is shown that the average price per night is 1,329.88 DKK with a standard deviation of 712.39. The lowest price per night is 100.00 and the highest is 6,857.00 DKK, indicating a broad price range. Notably, half of the listings have a nightly price within the range 900.00 to 1,551.00 DKK, showing a high concentration of the distribution. To put these prices into a context, the following figure shows the distribution of room types and number of bedrooms per listing.

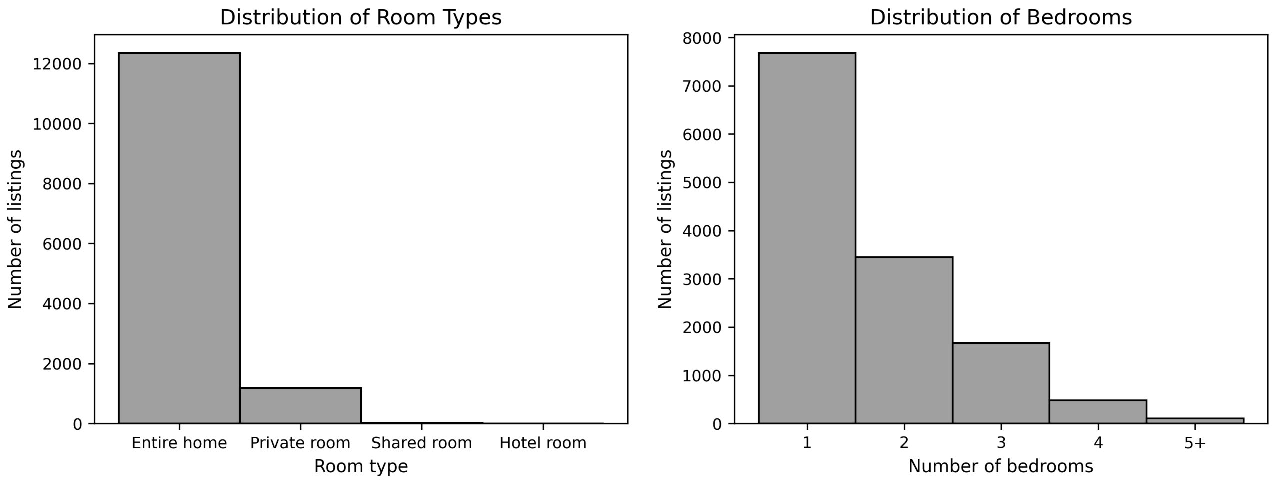

Figure 2: Distribution of room types and number of bedrooms.

The majority of listings are entire homes or apartments with one or two bedrooms. These listings tend to have a lower price compared to larger properties with three or more bedrooms, which contributes to most prices being concentrated at the lower end of the price range. As a result, the price distribution is left-skewed, with relatively few listings priced above the third quartile. This highlights the dominance of smaller, more moderately priced properties in the market.

Another important factor that influences the price distribution is the location distribution of the listings. By examining the geographical distribution of listings and the average price per night across neighborhoods, further insights into how different areas contribute to overall market trends can be gained.

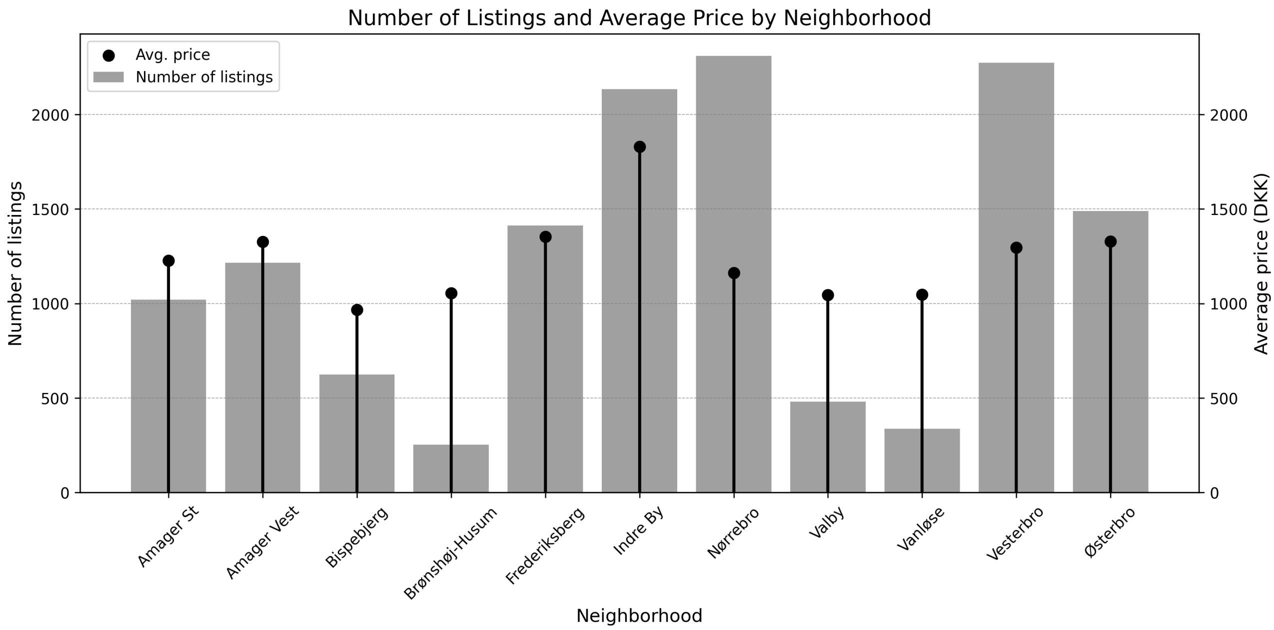

Figure 3: Geographical distribution and average price by neighborhood.

From the diagram, it can be seen that listings are mostly concentrated in central neighborhoods, such as Nørrebro, Vesterbro, and Indre By, rather than in peripheral areas. The average price is highest in these neighborhoods, which reflects their appeal. Indre By stands out with the highest average price, while the overall price variation across neighborhoods is relatively small, with a standard deviation of 238.72 DKK per night.

While price and location play an important role in shaping the Airbnb market, guest reviews offer insights into the overall quality of the listings. Reviews not only reflect guest satisfaction but can also indicate whether higher prices in popular neighborhoods meet customer expectations. The `review_scores_rating` column of the `listings2024.csv` dataset shows an average review score of 4.83, with a standard deviation of 0.26 and a median of 4.91. It is evident that this metric is skewed to the right with the overall review score being high. Hence, it is difficult to measure the success of certain attributes of the Airbnb listings based on the review score. Given this, it is apparent that the scores in the review subcategories must also be distributed right-skewed, as illustrated in Figure 4. This implies that the review score cannot provide insights to explain the price distributions of the listings.

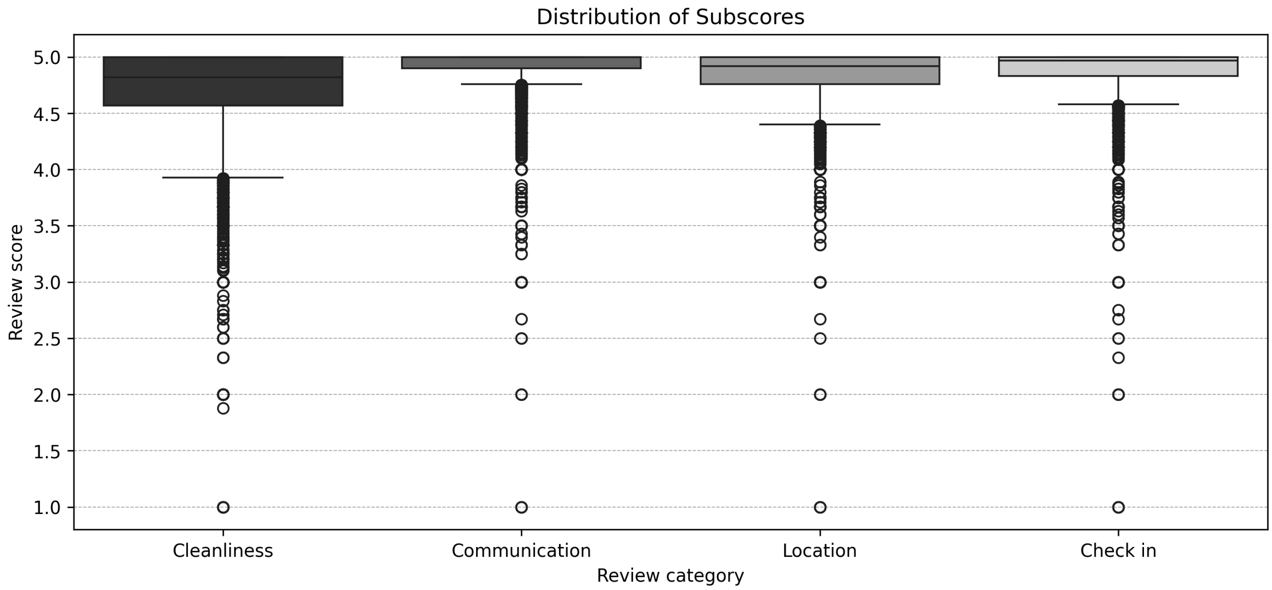

Figure 4: Boxplots of review scores subcategories.

While the data on reviews provides insights into the overall quality of listings, it is important to delve deeper into the occupancy data to get a deeper understanding of the business problem.

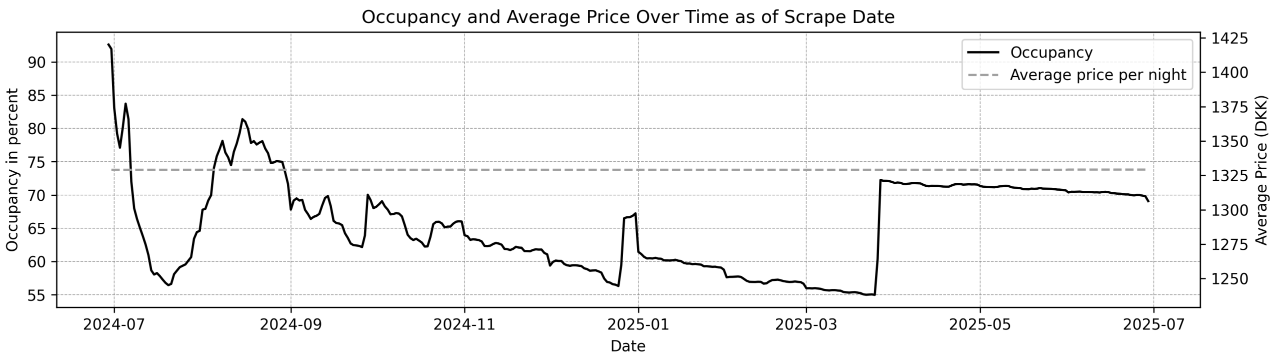

Figure 5: Occupancy and average nightly price over the year.

The average occupancy for the next 365 days is 65.4% on June 29, 2024. The graph shows that occupancy was high on the scrape date, exceeding 90%. It then declines to around 57% by the end of July before gradually increasing as the semester begins. This rise could be linked to new and exchange students temporarily staying in Airbnbs. Peaks in occupancy are visible during the autumn and Christmas holidays, which reflects higher demand during holiday seasons. Despite these seasonal changes in demand, the average nightly price remains constant at 1,329.00 DKK throughout the year. Additionally, the trend of the graph shows occupancy decreasing over time, likely because these dates are farther from the scrape date. The drop in availability of nearly 20% in early April may indicate that hosts limit bookings far in advance.

To better understand how to improve the occupancy of Airbnb listings, it is important to look at descriptive statistics on booking requirements and review patterns. Analyzing the `minimum_nights` column in the `listings2024.csv` dataframe reveals that the average minimum stay required is just under four nights. The third quartile is four nights, indicating that while most listings have relatively low minimum stay requirements, there are significant outliers with much higher requirements. Additionally, the majority of listings allow stays of 365 days or more, showing that night restrictions are stricter at the lower end than at the upper end. Since there is no data on the average duration of bookings, the `reviews_per_month` column provides a reasonable proxy for understanding booking patterns. On average, listings receive 0.98 reviews per month, with a standard deviation of 1.31. Although the proportion of guests who leave reviews is unknown, it is reasonable to assume that a significant number of guests do provide feedback. From this, it can be derived that the overall guest turnover in listings is relatively low, which may be due to bad occupancy or the prevalence of long-term stays.

3.2 Key Findings

The first part of the EDA focused on understanding the distributions and descriptive statistics of the data, providing a foundation for analyzing the Airbnb market in Copenhagen further. In this section, the analysis shifts to identifying key insights and examining relationships within the data. A particular focus will be on understanding factors that influence occupancy, as improving this metric is central to the business case. To explore the relationship between occupancy and two key metrics, price and review scores, scatterplots were used to identify potential patterns.

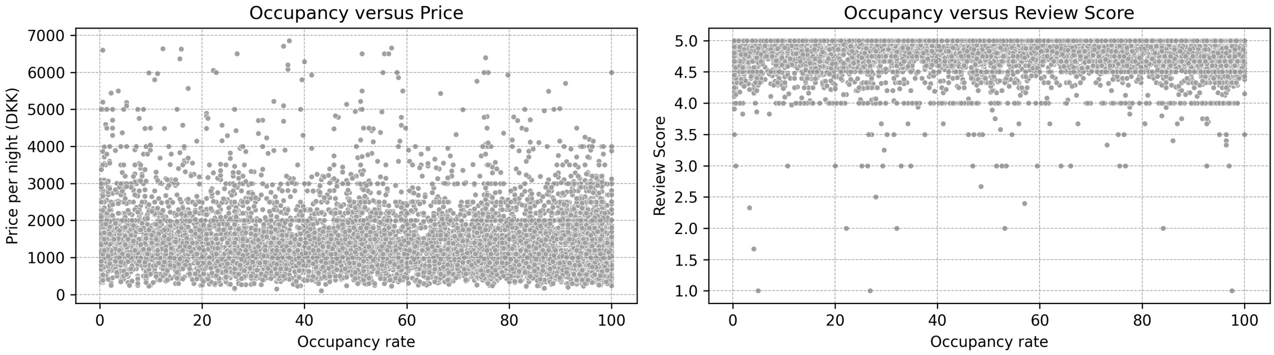

Figure 6: Relation between occupancy rate, price, and review score.

With a correlation coefficient of just -0.16, it is evident that the price does not have a significant effect on the occupancy. This suggests that diverse supply segments serve different demand segments in the market. For example, there are both affordable listings with a high demand and more expensive accommodations that still find their customer base and show nearly the same occupancy rate as more affordable listings. Moreover, the data suggests that the overall market is not very price sensitive and customers do not necessarily look for the cheapest option but may prioritize other factors such as reviews, location, or amenities. However, this does not mean that each customer segment is entirely indifferent to price when viewed in isolation.

The graph on the right-hand side of Figure 5 illustrates that there is no apparent relationship between the occupancy rate and review scores. It is important to note that the review scores are highly concentrated at the upper end of the range, with most listings achieving very high ratings and a minimal standard deviation. This suggests that the general level of satisfaction among guests is consistently high, making it challenging to draw any meaningful conclusions about occupancy rates based on the overall review scores. Consequently, the occupancy rate cannot be inferred from review scores in this context.

The data indicates no clear link between price or review scores and occupancy. This is likely because the market is not price-sensitive, and most listings have high ratings. This suggests that guests may prioritize other attributes. Exploring geographical differences in occupancy rates could provide valuable insights.

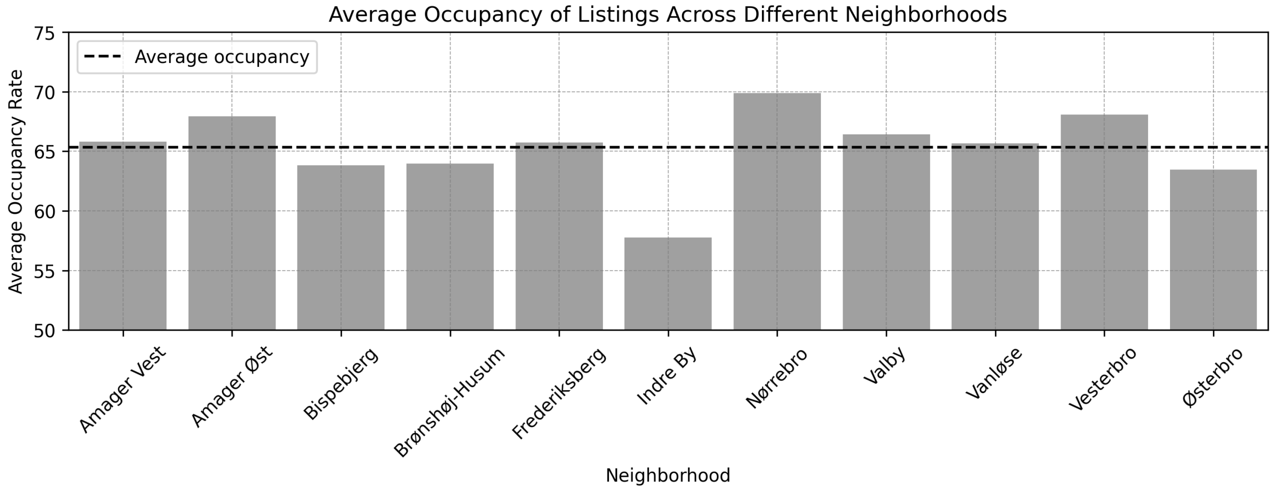

Figure 7: Occupancy across neighborhoods.

The chart indicates that certain neighborhoods, such as Nørrebro and Vesterbro, experience higher relative demand, as their occupancy rates surpass the overall average. In contrast, neighborhoods like Brønshøj-Husum and Valby show relatively lower occupancy, suggesting less demand for listings in these areas. However, since the data reflects a future observation period starting from the scrape date and extending 365 days ahead, these patterns represent relative rather than absolute demand. For example, the low occupancy rate in central neighborhoods like Indre By might not reflect low actual demand but rather the tendency of these listings to be booked less far in advance, leaving more availability closer to the scrape date. The consistent overall average occupancy rate of 65 percent serves as a benchmark for evaluating neighborhood performance within this context.

After analyzing the impact of reviews, price, and location on occupancy, the next step is to examine which attributes guests are looking for when choosing a listing. By focusing on the amenities provided by accommodations, this analysis aims to identify which features are most valued by customers and how they contribute to the success of a listing. These insights can help better understand guest preferences and suggest ways to improve occupancy.

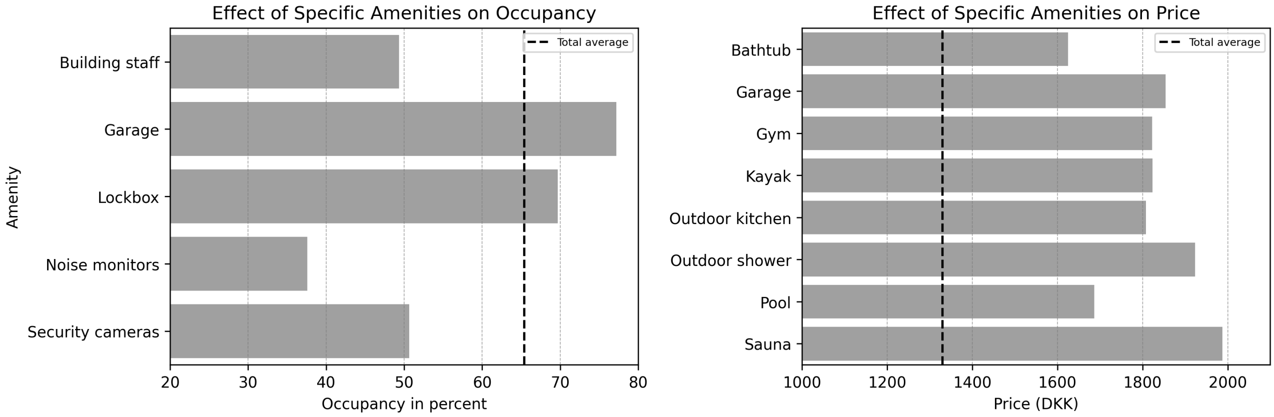

Figure 8: Effect of selected amenities on occupancy and price.

The charts show that certain amenities have a noticeable impact on both occupancy rates and price. The total average occupancy rate is 65.85%, and when grouping by amenities, the standard deviation of the average occupancy rates is 6.29. This indicates that the average occupancy varies notably between listings equipped with certain amenities. For instance, garage and lockbox are associated with occupancy rates higher than the average, suggesting that these features are attractive to guests. On the other hand, amenities like noise monitors are linked to lower occupancy rates, which might indicate a more limited appeal. Regarding price, premium amenities such as sauna, outdoor shower, and pool are linked to significantly higher prices compared to the overall average, which could be explained through their luxurious nature. Notably, some amenities, like a garage, positively influence both occupancy and price, making them particularly valuable for hosts. Due to the significant right-skewness of the distribution, it was not found useful to analyze the interdependency between amenities and the review scores.

Building on the previous analysis, the focus now shifts to understanding how specific host attributes impact the occupancy and review scores of listings.

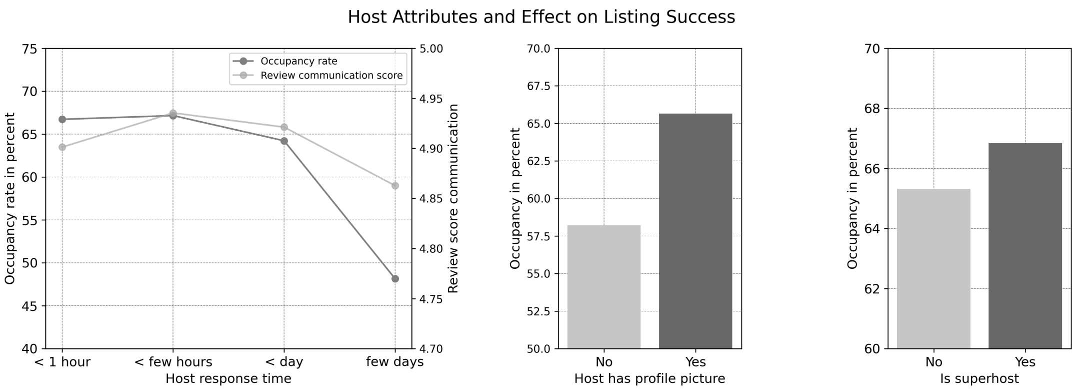

Figure 9: Effect of host attributes on occupancy and review scores.

Hosts with faster response times by average have higher occupancy rates and better communication review scores. Listings with a response time under a few hours perform best, while delays of a few days lead to lower occupancy and review scores. Furthermore, hosts with a profile picture have an average occupancy rate of about 65%, compared to 55% for those without, suggesting that guests feel more comfortable booking with hosts who provide a personal touch. Similarly, superhost listings, which reflect consistent reliability and positive guest experiences, achieve occupancy rates of around 68%. This value exceeds the average occupancy of non-superhosts by 6% and emphasizes how trust-building elements influence the appeal and success of a listing.

The EDA provided valuable insights into the Airbnb market in Copenhagen. In particular, the geographical analysis and the impact of specific amenities on average price and occupancy offer opportunities for improving the performance of existing Airbnb listings. Additionally, the analysis highlights that small adjustments, such as adding a profile picture or improving response time, can significantly enhance a listing's appeal to potential guests.

4 Data Model

It was not found necessary to set up and design a database in Postgres to fulfill the objective. Instead, a data model was developed to include only relevant information to address the business problem. Following the data cleaning and EDA steps, a normalized data structure was selected for organizing the data and eliminating redundancy. Figure 10 provides an overview of the resulting data model.

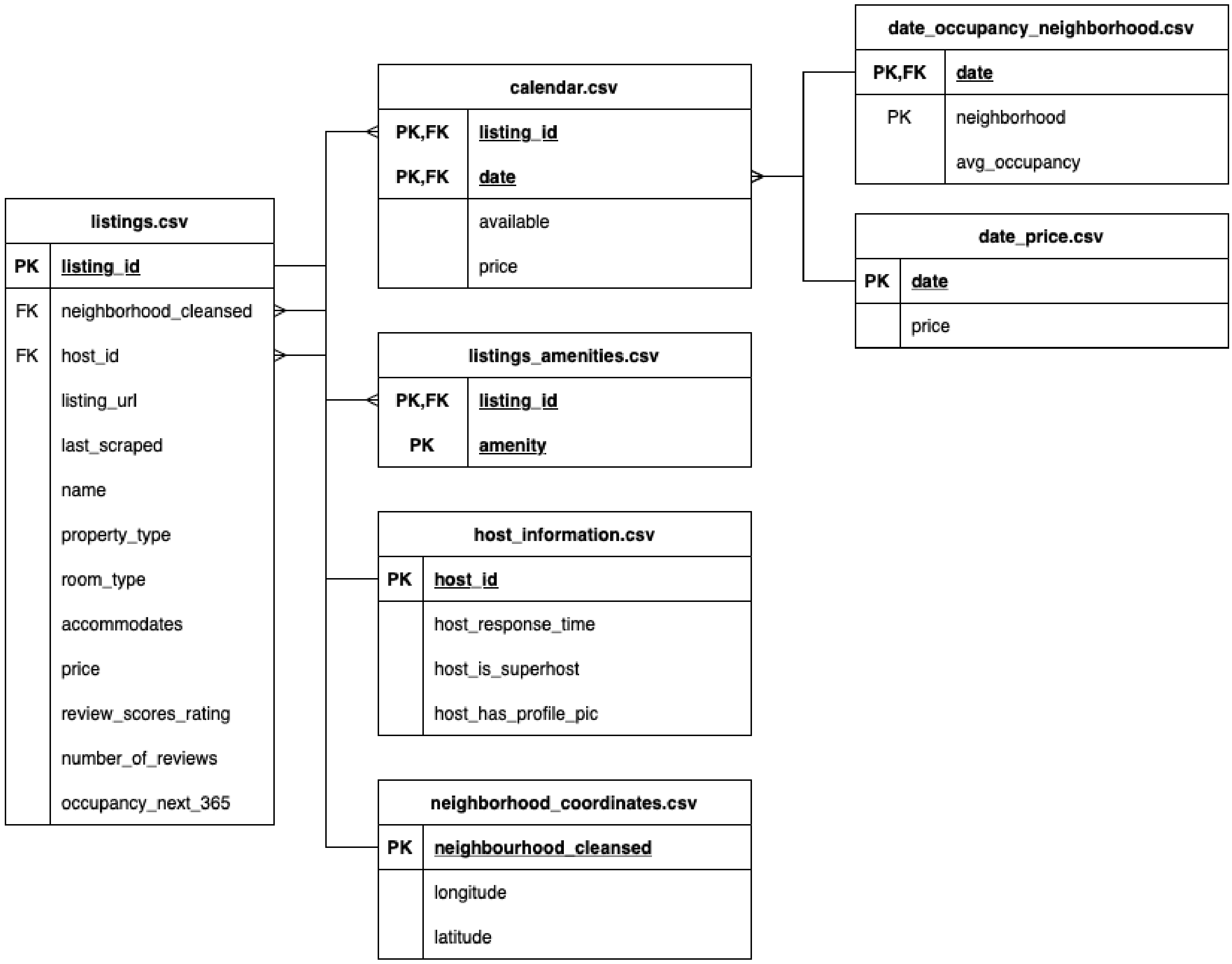

Figure 10: Data model with relationships and keys.

The `listings.csv` dataframe was cleaned by removing all columns that were not necessary for further analysis. Detailed descriptions of the columns can be found in the provided data dictionary. Notably, the `occupancy_next_365` column provides the occupancy rate for each listing over the next 365 days, starting from the `last_scraped` date. Similarly, the `calendar.csv` dataframe was reduced to include only relevant information and now consists of four columns. Based on this dataframe, the `date_occupancy_neighborhood.csv` file was created, containing each combination of date and neighborhood along with the average occupancy of the observed listings. Additionally, `date_price.csv` was generated to include the average nightly price of listings for each date, with both dataframes linked through the `date` column. The `listings_amenities.csv` dataframe provides detailed information about the amenities associated with each listing in the form of a melted dataframe. Furthermore, the `host_information.csv` file linked to `listings.csv` via the `host_id`, contributes to reducing redundancy. Lastly, the `neighborhood_coordinates.csv` file includes the geographical coordinates of each neighborhood. The initially provided `reviews2024.csv` dataframe was not utilized for further analysis beyond the EDA, as it was deemed irrelevant to addressing the business problem. The resulting data frames were imported into Tableau and serve as the foundation for the dashboard, which will be discussed in the following section.

5 Dashboard

5.1 Visuals and Business Insights

The dashboard serves as a strategic tool to support Airbnb's business objectives by aligning data-driven insights with the company's commitment to fostering sustainable relationships with local communities. It visualizes key metrics, such as occupancy rates across neighborhoods and host characteristics, providing actionable insights to optimize occupancy. This subchapter explains how the visuals support business objectives, highlights key insights, and offers practical guidance, while the second subchapter explores the scientific rationale behind the design choices and features, ensuring the dashboard effectively translates data into actionable insights.

The following dashboard is divided into the overview page and the action summary page. The first page maps Airbnb listings and summarizes performance metrics, focusing on occupancy. The second page provides insight and recommendations for price optimization, besides key findings on host characteristics, which can be used as a baseline for host management and potential incentives.

The first page contains three key visuals: the KPI table, geographic map of listings, and the boxplot for occupancy by neighborhood. These align with the business objectives of Airbnb: optimizing occupancy rates, improving utilization of existing listings, and minimizing the impact on the local housing market.

The KPI table shows key indicators such as the number of listings, reviews, ratings, price, and occupancy. Where the average occupancy for the period from June 29, 2024 to June 29, 2025 is 65%, it does show that there is some room for growth. With targeting of the neighborhoods that have a relatively low occupancy, Airbnb could optimize revenues without needing extra listings, thus helping to reduce the pressure on the local housing market and foster sustainable growth.

The geographic map offers an interactive view of the number of listings across Copenhagen. The differently sized and colored circles represent the concentration of listings in each neighborhood, enabling users to assess where listings are most prevalent. This map helps identify high-density areas like Indre By, Nørrebro, and Vesterbro, which have larger numbers of listings.

The occupancy boxplot gives a more detailed view of how the occupancy rate varies across neighborhoods, showing the central tendency (median) and spread (quartiles) of the occupancy rates for each neighborhood. This visualization clearly shows that neighborhoods like Nørrebro (70%) and Vesterbro (68%) are doing well in terms of occupancy, while neighborhoods like Indre By (58%) and Østerbro (63%) have room for improvement.

Taken together, these charts provide a rich insight into the performance of various neighborhoods. Central areas such as Nørrebro, Vesterbro, and Amager Øst present higher occupancy rates and more listings, indicating higher demand. Areas such as Indre By and Østerbro have a high number of listings but relatively low occupancy rates, which may suggest that targeted interventions could be beneficial in increasing the occupancy rate in these neighborhoods.

By providing a clear understanding of where listings are concentrated and which neighborhoods have lower occupancy, Airbnb can target specific neighborhoods to enhance occupancy by optimizing pricing, all while maintaining positive relationships with local communities.

The second page of the dashboard focuses on providing actionable insights for both Airbnb and hosts to optimize occupancy in Copenhagen.

What Airbnb Can Do. This section includes a dual-axis line chart of overall occupancy rate and average listing prices in Copenhagen between June 29, 2024, and June 29, 2025. The combined visualization of these indicators provides Airbnb with a clear view of occupancy trends alongside price fluctuations, enabling actionable insights for pricing optimization and host incentivization. While occupancy and price are not directly correlated across all price segments as pointed out during EDA, targeted price adjustments and pricing strategies can impact occupancy and revenue within specific price segments (Cahyani et al., 2021). For example, during periods of low occupancy, Airbnb could reduce commissions or encourage hosts to lower prices to attract more bookings, thereby enhancing both occupancy rates and potential revenue. Conversely, during periods of high occupancy and rising prices, Airbnb might explore opportunities to increase commissions or advise hosts to adjust their pricing strategies to capitalize on high demand, further enhancing revenue potential.

What Hosts Can Do. This section provides insights into actions that hosts can take to increase their occupancy rates. The first visual is a bar chart comparing occupancy rates for hosts with profile pictures versus those without. Hosts with profile pictures see an 8% higher occupancy rate, highlighting the importance of having a complete and trustworthy profile.

The second visual is a column chart showing the correlation between response time and occupancy. Hosts who respond to inquiries within an hour (67%) or within a day (64%) see significantly higher occupancy rates than those who take several days to respond (48%). This underlines the importance of quick communication for improving booking rates.

The final visual is a column chart that displays the best and worst performing amenities in relation to occupancy rates. Amenities like a garage (77%), games (70%), and a children's playroom (73%) are highly appreciated by guests and are associated with above-average occupancy. Conversely, amenities like security cameras (51%), building staff (49%), and noise decibel monitors (38%) are linked to lower occupancy rates. These insights offer hosts guidance on which amenities to prioritize or reconsider to increase their listing's attractiveness and occupancy.

Together, these visuals equip both Airbnb and hosts with valuable information to optimize occupancy in Copenhagen.

5.2 Dashboard Design and Visuals

The design of the dashboard is critical to ensuring its usability, effectiveness, and alignment with business objectives. Drawing on established principles from scientific research and human-computer interaction, effective dashboards should prioritize clarity by avoiding visual clutter (Few, 2006). Incorporating features such as interactivity, grouping of visuals, and multi-page layouts helps manage complexity and ensures that information is both accessible and actionable (Sarikaya et al., 2019).

The Airbnb dashboard has a multipage structure, featuring an executive summary and an action summary. This layout allows users to first gain an overarching view of the Copenhagen Airbnb market, focusing on occupancy—the primary area for optimization. The second page delves into detailed information, enabling users to base actionable recommendations on in-depth insights.

A navigation button, positioned in the bottom-right corner of the visual, facilitates easy access to the second page. On the second page the user will also find a navigation button to go back to the executive summary. The dashboard is highly interactive, allowing users to filter and navigate its content intuitively. The first page features three key visuals, which can be filtered by neighborhood, room type, or property type to tailor the information to the user's needs. Additionally, users can customize the geographic map view through a toggle button, choosing between a neighborhood-level overview or a detailed listings display. The second page allows the user to filter the date range and desired neighborhood for the line chart enabling users to create optimization strategies for occupancy specifically by date and neighborhood.

By structuring the dashboard this way, it avoids overwhelming users with excessive details upon initial interaction. This ensures a seamless and user-friendly experience. The design prioritizes usability, customization, and user-centric principles. It also actively mitigates the risk of information overload, a common phenomenon that can distract rather than focus decision-makers' attention (Yigitbasioglu & Velcu, 2012, p. 42). The dashboard allows users to access detailed information through tooltips, which provide additional insights and filtering options when needed. For example, the map displays tooltips with information about individual listings and more advanced distributions of room types. The line charts also feature tooltips offering optimization tips based on trends in occupancy. Overall, the design follows Ben Shneiderman's three steps for effective visualization: overview first, zoom and filter, and details on demand (Shneiderman, 2003).

To enhance user experience, the dashboard employs color effectively. The main color scheme is based on the Airbnb logo, creating a consistent and recognizable visual identity throughout the dashboard. The neighborhood legend uses a color palette to improve clarity, helping users quickly identify and differentiate between areas. On the action summary page, red and green colors are used to represent trends in occupancy. Red highlights lower occupancy or negative trends, while green indicates higher occupancy or positive trends. This color scheme follows the concept of analogical reasoning, where familiar associations with red and green (e.g., traffic lights) make it easier for users to understand the data and make quick decisions (Patterson et al., 2014).

6 Limitations

The datasets used in this project come with several limitations that affect how in-depth and accurate the analysis can be, especially given the goal of optimizing occupancy without adding pressure to existing housing market issues.

One key problem is the lack of historical booking data, which would have been helpful in establishing benchmarks and identifying long-term trends in occupancy. Since this information is missing, the analysis had to rely on future availability patterns, which limits the ability to assess past performance. Additionally, about a third of the pricing data in the listings dataset was missing, making it harder to draw reliable conclusions about pricing. There was also an issue with the `prices` column in `calendar.csv`, as it appeared to include listings with different currencies, which made price comparisons more complicated and affected data consistency.

Another limitation is that the dataset does not provide information about the average length of stays, which could have offered useful insights into seasonal patterns and neighborhood-specific behavior. Moreover, reviews help to understand guest satisfaction, however, they tend to be very positive, which makes it difficult to detect variations in performance.

Lastly, the absence of real-time booking data is a factor that affects the usefulness of the dashboard. Without this, it is not possible to monitor occupancy trends in real time or respond quickly to changes.

7 Conclusion

This project successfully uses the technical rigor of data science to generate actionable business insights that address Airbnb's challenge of optimizing occupancy in Copenhagen to mitigating pressures on the local housing market. A comprehensive Tableau dashboard has been developed that empowers Airbnb's regional business development and public affairs teams to make data-driven business decisions and foster constructive relationships with Copenhagen's authorities.

From a data science perspective, the foundation of this project was built on thorough data preparation and data cleansing. Data inconsistencies, such as extreme pricing outliers and incomplete records, were addressed to ensure high-quality insights. Through EDA, the current price inelasticity on Airbnb's marketplace and the influence of specific host attributes and amenities on occupancy rates were uncovered. These insights serve as the basis for the recommendations outlined in the dashboard.

From a business perspective, the Tableau dashboard provides actionable insights that directly address Airbnb's business problem. The EDA shows that neighborhoods like Indre By, Østerbro, and Brønshøj-Husum have the lowest occupancy rates, with Indre By having the most immediate need for optimization. To improve occupancy in these areas, Airbnb should implement a dynamic pricing strategy that adjusts prices based on demand fluctuations seen in the future occupancy rates. During periods of low demand, prices can be reduced to attract more bookings, while in high demand periods such as holidays or major events, prices can be increased to balance demand with supply and maximize revenue. This approach not only maximizes revenue but also optimizes the utilization of existing listings to ensures that the housing market is not further strained by the addition of new short-term Airbnb rentals.

Beyond pricing, Airbnb can engage hosts directly to improve occupancy rates through targeted recommendations. The three key actions are: (1) ensuring hosts add profile pictures, as listings with personal touches show higher trust and booking rates, (2) encouraging hosts to respond faster, which significantly enhance the likelihood of bookings, and (3) promoting family-friendly amenities like children's playrooms or board games, which cater to the important audience of traveling families. These adjustments are low-cost and easy to implement, yet can have a substantial impact on listing utilization.

The Tableau dashboard also supports Airbnb's public affairs team by providing a clear narrative that can be communicated towards local communities and authorities. This transparency and constructive cooperation can help build trust and foster sustainable relationships with all stakeholders in Copenhagen.

In summary, this project equips Airbnb with the tools and insights to navigate the challenges of Copenhagen's housing market by maximizing utility of existing supply through dynamic pricing, and targeted host engagement. This balanced approach aligns Airbnb objective of sustainable business growth with community needs, ensuring that Airbnb remains a leading player in Copenhagen's rental market.

8 References

Cahyani, N. P. E. Y., Astawa, I. K., & Triyuni, N. N. (2021). Increasing Room Occupancy and Room Revenue through Price Decision Strategy. International Journal of Glocal Tourism, 1(1), 1–12. https://doi.org/10.58982/injogt.v1i1.12

Chatfield, C. (1986). Exploratory data analysis. European Journal of Operational Research, 23(1), 5–13. https://doi.org/10.1016/0377-2217(86)90209-2.

Few, S. (2006). Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media, Inc.

Inside Brian Chesky's Ultimate Airbnb. (2023, June 23). [Broadcast]. https://www.bloomberg.com/news/videos/2023-06-23/inside-brian-chesky-s-ultimate-airbnb

Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. Human Resource Management Review, 27(2), 265–276. https://doi.org/10.1016/j.hrmr.2016.08.003.

Patterson, R. E., Blaha, L. M., Grinstein, G. G., Liggett, K. K., Kaveney, D. E., Sheldon, K. C., Havig, P. R., & Moore, J. A. (2014). A human cognition framework for information visualization. Computers & Graphics, 42, 42–58. https://doi.org/10.1016/j.cag.2014.03.002.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What Do We Talk About When We Talk About Dashboards? IEEE Transactions on Visualization and Computer Graphics, 25(1), 682–692. IEEE Transactions on Visualization and Computer Graphics. https://doi.org/10.1109/TVCG.2018.2864903.

Shneiderman, B. (2003). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In B. B. Bederson & B. Shneiderman (Eds.), The Craft of Information Visualization (pp. 364–371). Morgan Kaufmann. https://doi.org/10.1016/B978-155860915-0/50046-9.

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. International Journal of Accounting Information Systems, 13(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002.