



MONASH

University

Data Visualization Project

Navigating Amsterdam's Airbnb Landscape: Insights for Travelers



Student name: Bao Ngoc Truong

Student ID: 34301054

Applied session number: 02

Teaching Associate's name: Jeffery Liu

Table of Contents

Introduction	2
Problem Description:.....	2
Motivation:	2
Research Questions:	2
Design Process	2
Sheet 1	3
Sheet 2.....	4
Sheet 3.....	5
Sheet 4	6
Sheet 5.....	7
Implementation	9
Technical Implementation	9
Interactive Narrative Visualisation Implementation.....	10
Using the Implementation.....	14
Conclusion	15
Bibliography	16
Appendices.....	16

Introduction

Problem Description:

Following the Covid-19 pandemic, the resurgence of Airbnb operations has significantly transformed the tourist accommodation market in Europe, with short-term rentals now making up about 25% of the market share (Pascoe, 2023). This trend is particularly notable in Amsterdam, where 75% of available properties are entire homes, averaging a price of €241 per night (BNR, Berg, 2023). As travelers navigate this changing market, finding well-priced accommodation becomes crucial. This report explores and provides insights into the Airbnb market for Amsterdam from 2013-2023. It aims to help travelers understand pricing variations and identify successful hosts based on responsiveness and listing quality.

Motivation:

As an avid traveler planning a future trip to Amsterdam, I am particularly drawn to the appeal of Airbnb accommodations. It is important to balance affordability with quality, and Airbnb presents a promising solution. Beyond budget considerations, Airbnb supports my preference for engaging with local communities and finding hospitable hosts who can enhance my travel experience with their guidance and assistance (Airbnb, 2020).

Research Questions:

What factors drive pricing variations in Amsterdam's Airbnb listings?

What attributes characterize successful Airbnb hosts (e.g., responsiveness, listing quality)?

Design Process

One of the frameworks I used throughout the design process is the what-why-how Tamara Munzner's visualisation framework (2014). The datasets I used for the project include tabular and spatial data. The tabular dataset for analyzing Airbnb listings in Amsterdam spans from 2009 to 2023 and is available in a CSV file. It contains 8,739 rows and 75 columns, providing detailed information on listings and hosts. Attributes include property types, room setups, amenities, prices, availability, review counts, ratings, dates, licensing statuses, and calendar updates. This dataset is a rich resource for understanding Airbnb dynamics in Amsterdam, with both nominal and numerical data types allowing for in-depth analysis. Extending to December 12, 2023, it offers valuable insights into trends over the past decade.

The second dataset I used is a comprehensive geospatial dataset from April 2022. Initially in JSON format, it was then restructured into a Simple Feature Collection (sf object) for analysis and visualization. It includes attributes such as unique neighborhood codes (Gebiedcode), neighborhood names (Gebied), district codes (Stadsdeelcode), district names (Stadsdeel), and areas in square meters (Oppervlakte_m2). The 'geometry' field contains polygon geometry in the WGS 84 coordinate reference system, defining geographical boundaries for visual mapping. This refined dataset enables the visual analysis of price variations across neighborhoods, offering insights into the spatial distribution of average price.

In terms of why question, my project has multiple reasons. Some visualizations are used for discovering new insights about the Airbnb market to answer two research questions. For instance, successful host attributes will be derived from comparing elements and visually identifying patterns. Another reason is for presenting. Presenting

tasks require mapping elements and visual variables to create a view. This means separating the data visualization into topics and arranging them according to the narration to tell the story or demonstrate the insights visually. Another presentation task is to include filters or interactive elements to select a subset of elements and manipulate a view. To answer my research questions about hosts and price, having some kind of price and host filters are necessary to help tie the whole visualization together.

The last question to answer is how I want to make the visualization look like. Firstly, thinking about how I can encode the data into the visualization and visual elements helps me come up with new visualization ideas. Then, the next step is to think about how to connect these visualizations with each other through manipulating different elements. To ensure the visualization demonstrates the intended insights, facets are helpful to split the project to multiple views. Another thing I also consider is how filtering and aggregation can help break down the dataset into data that can be used for visualization. From the framework, I have a better idea of how data work together to work out Sheet 1.

Sheet 1

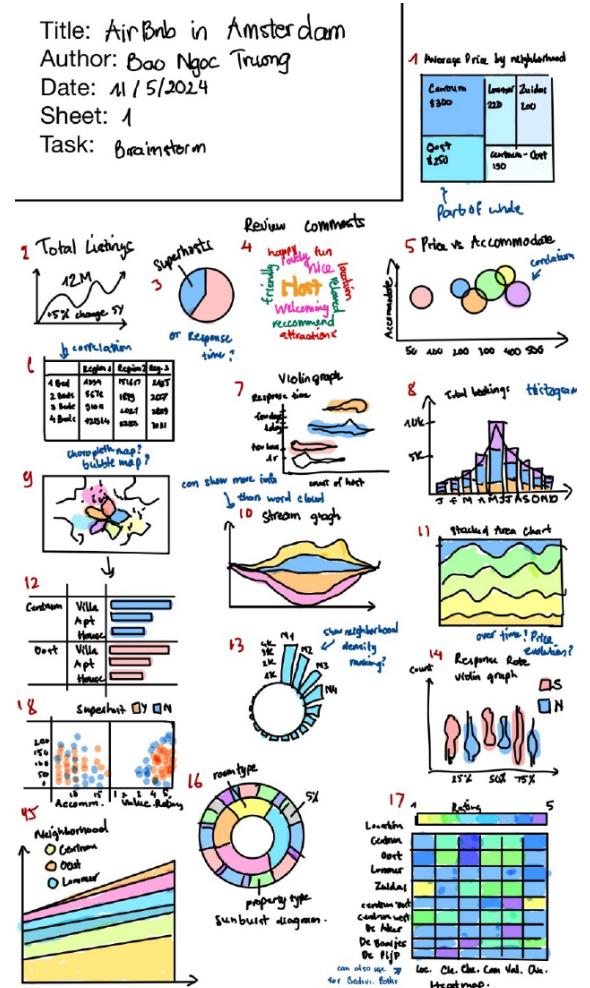
Sheet 1 serves as a brainstorming tool to expand the design space and generate numerous potential visualization designs, emphasizing quantity. In this project, I employed sketching as the primary design methodology.

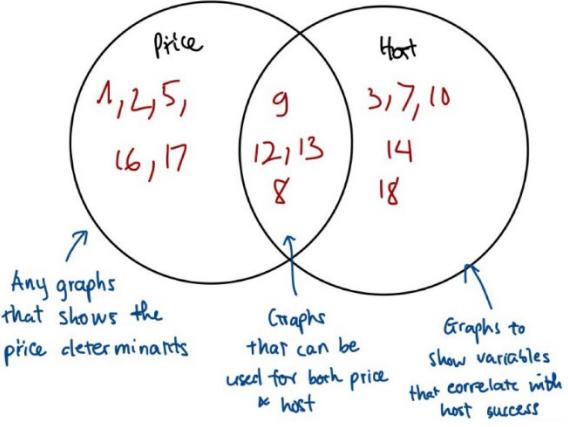
The task encompassed five stages, beginning with an ideation phase where I sketched as many ideas as possible. Consequently, I produced 18 sketches of various graph types, which I believed would enhance the visualization and aid in addressing my research questions. These included: (1) tree map, (2) line graph, (3) pie chart, (4) word cloud, (5) bubble plot, (6) correlation table, (7) violin graph, (8) histogram, (9) bubble map, (10) stream graph, (11) stacked area chart, (12) divided horizontal line chart, (13) circular bar plot, (14) grouped violin chart, (15) area chart, (16) donut chart, (17) heat map, and (18) scatter plot. These graph types were selected due to the tabular nature of my Airbnb listing dataset, which comprises predominantly categorical and quantitative data.

The histogram was chosen as the sole univariate graph for price distribution, given its centrality as a dependent variable in my analysis. It also enables the identification of a normal distribution post outlier removal. The remaining visualizations were bivariate or multivariate. Many multivariate graphs incorporate color saturation and hue as additional visual variables, in line with Jacques Bertin's principles in "Semiology of Graphics" (2011).

Overlaid graphical elements are a common feature in proposed graphs such as area plots, stacked bar charts, stream graphs, and donut charts.

Title: Airbnb in Amsterdam
Author: Bao Ngoc Truong
Date: 11/5/2024
Sheet: 1
Task: Brainstorm



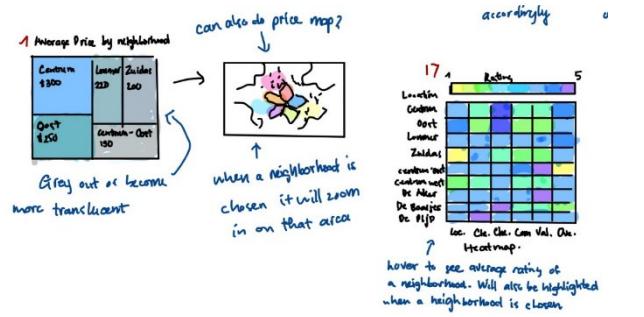


During the subsequent stage of filtering, I eliminated graphs with redundant or overly similar concepts, resulting in the omission of graphs 6, 11, and 15. Graph 6 was excluded because I intended to visualize beds and baths, which I deemed more effective as a heat map. The area chart (graph 11) and another area chart (graph 15) were discarded as they did not align with my research questions regarding price evolution.

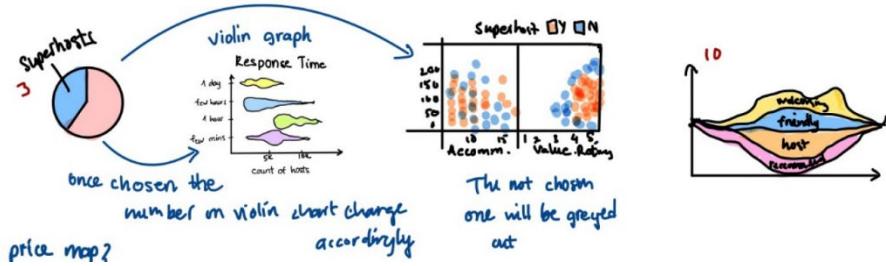
I then categorized the remaining graphs based on their ability to support my two research questions concerning price and host variation factors. Some graphs were eliminated due to their limited utility in visualizing the relevant variables.

In the combine and refine stage, I organized the graphs into groups to create a cohesive narrative. I experimented with making the price map interactive, enabling users to filter by neighborhood on the price tree map and view corresponding data on the neighborhood map. Additionally, a heat map was considered to show the correlation between neighborhoods and their listing review scores across various criteria.

For the second research question regarding successful host attributes, defined by Superhost status, I utilized a pie chart to display the percentage of Superhosts. Users can interact with the chart to filter subsequent graphs by



Superhost status. The response time violin chart and grouped scatter plot showing accommodation capacity and value rating can be filtered based on the pie chart selection. Finally, I included stream graphs to depict commonly mentioned host attributes, providing qualitative insights beyond numerical data.

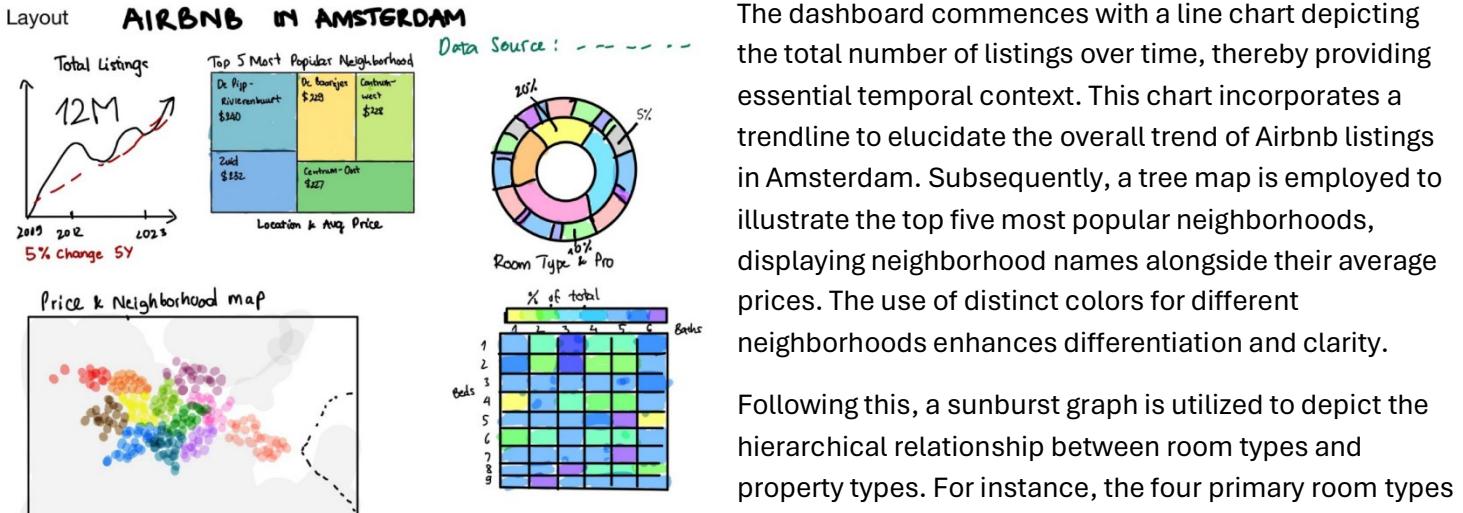


After refining and combining the visualizations, I evaluated their coherence and ability to support the research questions. Although the interactive elements did not fully address the research questions, they enhanced my understanding of the relationships between variables and the effectiveness of the visualizations. I recognized the importance of adding explanatory text and context to ensure clarity and coherence in the narrative.

This comprehensive design process, grounded in theoretical principles, resulted in a well-structured and insightful dashboard that effectively communicates the findings and engages the intended audience.

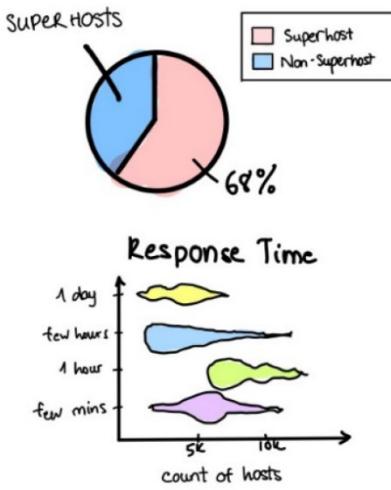
Sheet 2

Sheet 2 constitutes the initial layout of a dashboard meticulously designed to provide an exhaustive visual analysis of Airbnb listings in Amsterdam. This sheet leverages insights generated during the brainstorming phase in Sheet 1, integrating diverse visualization techniques to address two primary research questions: the dynamics of price and the attributes of hosts within the Airbnb market. The layout is methodically structured to guide users through the data in a logical and coherent manner.

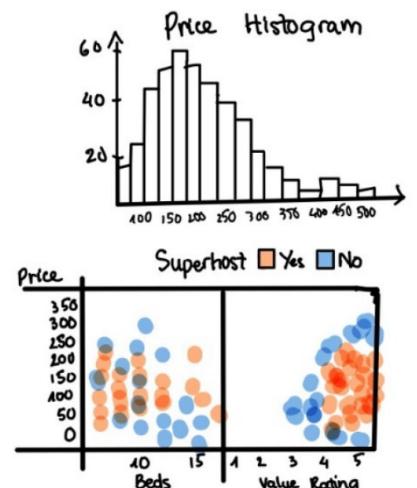


subdivided into various property types, thereby clarifying the hierarchical nature of the data. The adoption of a consistent color palette across visualizations ensures a uniform aesthetic, while contrasts are employed to highlight key data points.

Additionally, a combination map of price and neighborhood is incorporated, with a heat map positioned adjacent to display the distribution of listings by the number of beds and baths. The heat map employs a sequential color scale (e.g., viridis palette) to effectively represent density and frequency. A price histogram is also included to illustrate the distribution of price ranges, thus providing viewers with a comprehensive overview of Airbnb prices in the market.



For the second research question concerning host attributes, a pie chart is utilized to display the proportion of Superhosts. This chart includes interactive elements, allowing users to filter data based on host type. Upon selection of a specific host type, related graphs, such as the violin chart of response time and the grouped scatter plot of the number of beds and value rating, dynamically update accordingly.

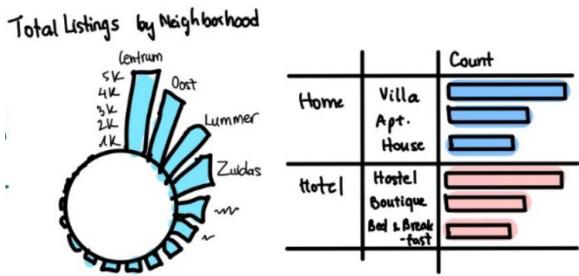


In conclusion, the design choices for Sheet 2 are informed by theoretical principles, emphasizing consistency in design, appropriate color palettes, and the effective use of visual variables. The narrative flow and interactive elements are meticulously crafted to engage users and provide profound insights into the Airbnb market in Amsterdam.

Sheet 3

Sheet three represents a refined iteration of the dashboard design, organized into three primary sections: an overview of the Airbnb market, an analysis of price variation factors, and the attributes of successful hosts. This

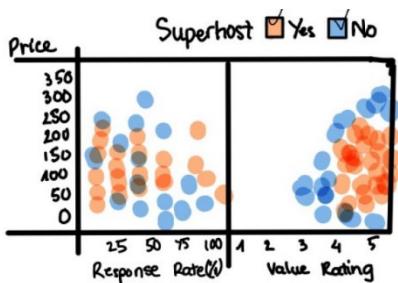
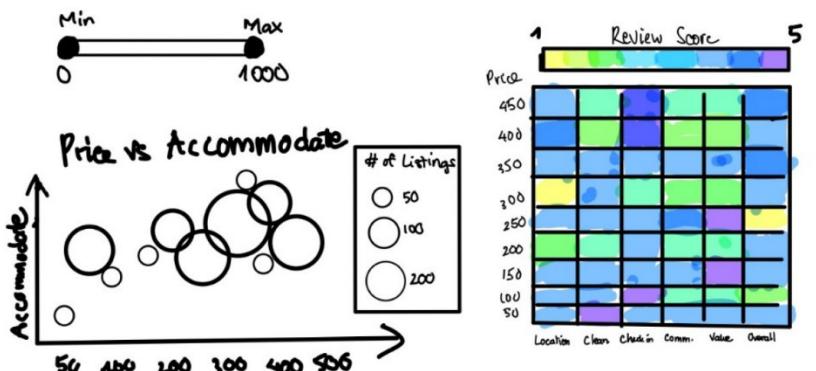
structure adheres to Munzner's what-why-how framework, addressing the "what" (data types), "why" (user goals), and "how" (visual encoding and interaction) for each segment.



The design employs clear, sans-serif typography to ensure readability and establish a visual hierarchy that guides users seamlessly through the narrative. To visualize the general distribution of total listings, a circular bar plot was selected for its efficacy in depicting neighborhood distributions, sorted from highest to lowest counts. This choice leverages the human visual system's proficiency in detecting differences in length and position.

For the distribution of room and property types, a grouped bar chart is used to facilitate easy comparison of categories. The introduction of a price range slider as an interactive element enables users to dynamically adjust the price range, with changes reflected instantaneously in related visualizations.

A heatmap, repurposed to display review scores across different price ranges and criteria (such as cleanliness and communication), employs color hue to convey variations, leveraging principles of preattentive processing for quick data comprehension. Additionally, a price and accommodation capacity bubble map is included to illustrate the positive relationship between these variables. The size and color of the bubbles encode information about listings, facilitating rapid understanding of complex data relationships.



For the analysis of successful host attributes, a grouped scatter plot is employed to show the distribution of response rate and value rating by host type. A word cloud is also used to visualize commonly mentioned traits of hosts, providing qualitative insights into host characteristics.

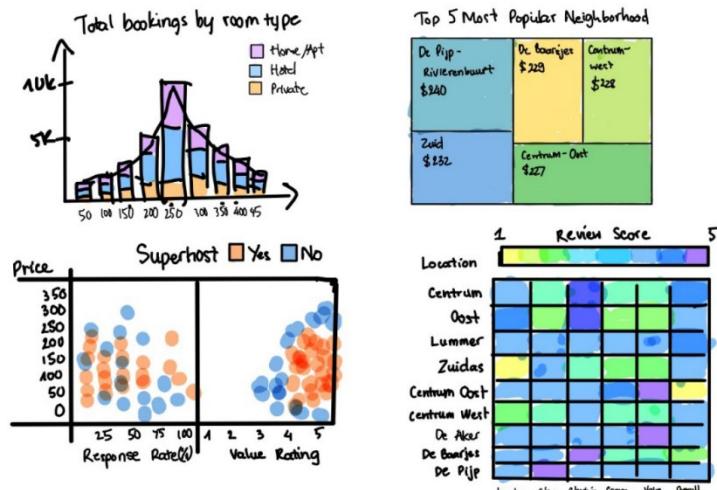
These visual and interactive elements are carefully chosen to create a cohesive and engaging narrative that enables users to explore the data intuitively and draw meaningful insights about Airbnb listings and hosts in Amsterdam.

Sheet 4

Consistency was maintained throughout the dashboard by using a cohesive color palette, intuitive layout, and uniform interaction mechanisms. The viridis color palette was chosen for its perceptual uniformity and accessibility to colorblind users. Consistent interactions, such as hover effects and clickable elements, ensured an intuitive user experience, allowing users to explore the data seamlessly.

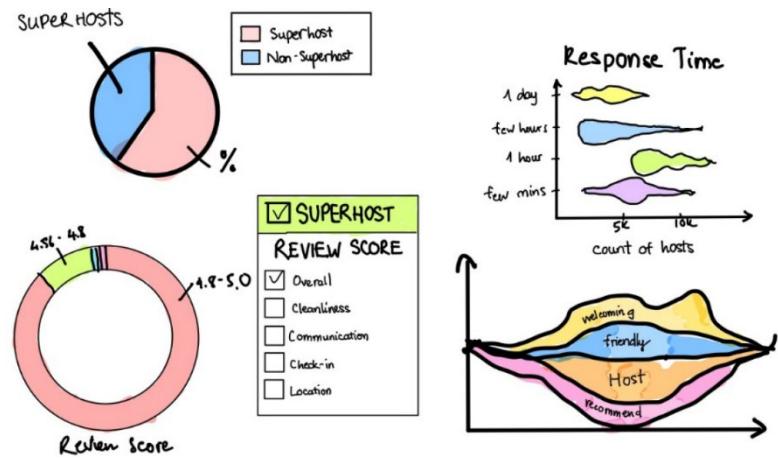
AIRBNB IN AMSTERDAM

Layout



Sheet 4 is the last initial design of the dashboard. For visualizing price distribution, I started with a histogram chart color-coded by room type. This choice leverages the human visual system's ability to detect differences in hue and facilitates a clear understanding of price variations across different room types. A tree map was used to highlight the most popular neighborhoods and their average prices, providing an intuitive visual summary of neighborhood popularity and pricing. To explore the correlation between location and review scores, a heatmap was employed. This visualization employs color hues to depict differences in review scores, leveraging the principles of preattentive processing.

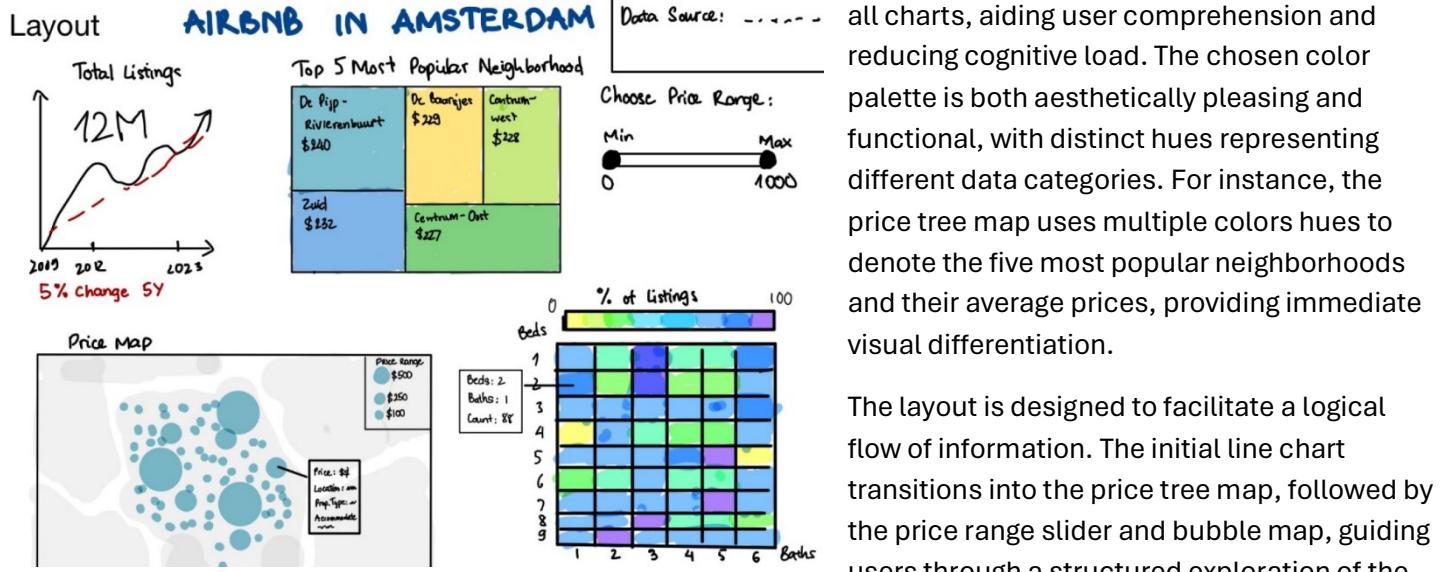
For the second research question, a grouped scatter plot was used to show the distribution of value ratings and response rates across different host types. This visualization allows for easy comparison and trend identification. Interactive elements, such as the pie chart and donut chart, were incorporated to enhance user engagement. The pie chart shows the distribution of superhosts among all hosts, and clicking on the superhost section filters other maps related to host attributes. The review score donut chart provides detailed insights into review scores for different criteria, while the violin chart illustrates the distribution of response times, leveraging the human visual system's ability to interpret shapes and distributions effectively. The stream graph, displaying the frequency of commonly mentioned host attributes, was chosen for its ability to show trends over time in a visually engaging manner. This choice aligns with the explanatory genre of narration, guiding users through the data while allowing for exploratory analysis.



The design choices were heavily influenced by the intended audience, including travelers who are planning to book an Airbnb in Amsterdam, or someone who plans to become an Airbnb host. Visualizations were selected and designed to ensure that users with varying levels of data literacy could navigate and interpret the insights effectively. The use of interactive elements and clear visual encoding aimed to engage users and provide them with actionable insights.

Sheet 5

Sheet 5 is the final design of the layout. The first chart, a line chart showing total Airbnb listings in Amsterdam, serves as an introductory overview, leveraging the familiarity and simplicity of line charts to convey temporal trends effectively. This chart sets the stage for more detailed analyses by providing a clear, high-level view of market growth over time. The design ensures consistency by using similar color schemes and visual styles across



all charts, aiding user comprehension and reducing cognitive load. The chosen color palette is both aesthetically pleasing and functional, with distinct hues representing different data categories. For instance, the price tree map uses multiple colors hues to denote the five most popular neighborhoods and their average prices, providing immediate visual differentiation.

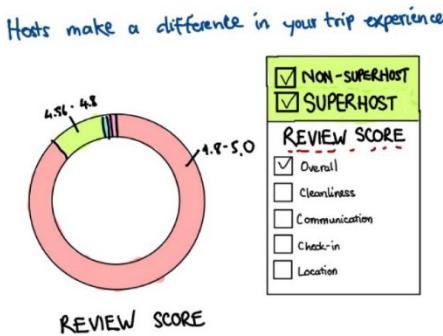
The layout is designed to facilitate a logical flow of information. The initial line chart transitions into the price tree map, followed by the price range slider and bubble map, guiding users through a structured exploration of the

data. Clear headings and consistent typography ensure readability and guide users through the narrative. The use of a price range slider allows users to interactively filter the data, enhancing engagement and allowing for personalized exploration of price variations.

Each visual element is carefully selected to represent specific aspects of the data. The heatmap, for instance, shows the correlation between the number of beds and baths and the average price, utilizing color intensity to depict relationships and trends. This aligns with Munzner's why-how framework, focusing on revealing patterns and correlations through effective visual encodings. For the second research questions about host variation factors, I chose a donut chart to show different reviews scores for various criteria. The graph is also filtered by a tick box table that can filter data of host types and review score types. The grouped scatter plot of response rate and value rating of listings for hosts is also included to show the comparison between superhosts and non-superhosts. Lastly, the violin chart is added to show the response time of

The dashboard employs a hybrid narrative style, combining explanatory and exploratory elements. Initial charts provide explanatory insights into overall trends, while interactive components like the price range slider and bubble map offer exploratory opportunities, allowing users to delve deeper into specific areas of interest. This combination enhances user engagement and caters to different user preferences and analytical needs.

In designing the visualizations, I considered aspects of the human visual system, such as the ability to quickly process visual patterns and the effectiveness of spatial arrangements in conveying information. The use of familiar chart types (e.g., line charts, heatmaps) leverages users' existing knowledge, facilitating quicker understanding. The inclusion of interactive filters and hover-over details further supports user engagement by providing additional layers of information without cluttering the visual space.

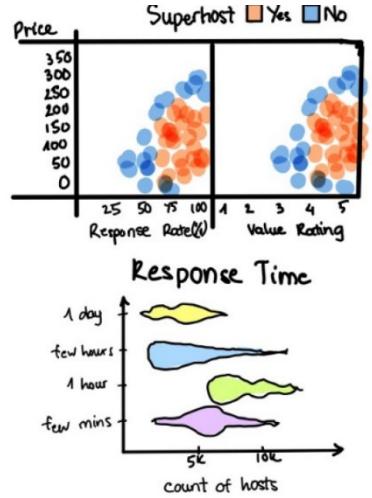


For the second research question about host variation factors, I chose a donut chart to show different review scores for various criteria, supported by a tick box table to filter data by host types and review score types. This choice is justified by the need for clarity and the ability to compare categorical data effectively. The grouped scatter plot of response rate and value rating provides a direct comparison between superhosts and non-superhosts, utilizing spatial positioning to highlight differences. The violin

chart, showing response time distribution, adds another layer of analysis, illustrating the spread and central tendencies of response times.

Aggregation is employed in most graphs, summarizing data to highlight key trends and insights. This approach simplifies complex datasets, making them more accessible and interpretable. For instance, the line chart includes a regression analysis to illustrate the overall trend of listings over time, providing a clear visual representation of market dynamics.

The design process was heavily influenced by the intended audience, which includes both casual users and more analytical viewers. The combination of explanatory and exploratory elements caters to a broad audience, ensuring that both high-level insights and detailed analyses are accessible. Interactive features like the price range slider and tick box filters enhance user engagement, allowing for tailored exploration based on individual interests.



Implementation

Technical Implementation

To develop my Data Visualization Project, I utilized an array of libraries, including Plotly, Shiny, Leaflet, Datasets, ggplot2, Lubridate, dplyr, treemapify, Viridis, shinydashboard, ggiraph, geojsonsf, sf, data.table, and R.utils. Each library was selected for its specific capabilities in data manipulation and interactive visualization, contributing to the overall efficacy of the final design.

In response to feedback received during the presentation of my final design, I integrated narrative texts to enhance the comprehensibility of the visualizations. These annotations provide contextual explanations, thereby facilitating a deeper understanding of the insights. Additionally, to improve navigation, I organized the dashboard into three distinct tabs: an overview of the Airbnb market in Amsterdam, an analysis of price variation factors, and an examination of successful host attributes. This tabbed structure aligns with Munzner's what-why-how framework, effectively delineating the types of data, user goals, and visualization methods for each section.

Several adjustments were made to refine the visualization design, ensuring greater coherence and insightfulness. Initially, the line chart depicting the total number of listings was enhanced by adding a second line representing the average price over time. This dual-line chart offers a more nuanced depiction of market dynamics, illustrating the concurrent trends in supply and pricing.

To further investigate price variation factors, I incorporated a histogram of prices color-coded by room type, as proposed in Sheet 4. This histogram not only reveals the overall price distribution but also highlights variations across different room types, providing a clearer segmentation of the market.

The scatter plot illustrating the relationship between price and accommodation capacity from Sheet 3 was included to address the first research question. The addition of a linear regression trendline visualizes the correlation between these variables, aiding in the identification of significant patterns.

For geographical insights, the initial bubble map was replaced with a combination of choropleth and bubble maps. This hybrid approach effectively conveys both the spatial distribution of listings and the average prices within

neighborhoods. Given the inconsistencies in the dataset, extensive data wrangling was undertaken to ensure accurate matching of neighborhood names between the tabular and geospatial data.

	Gebiedcode	Gebied	Stadsdeelcode	neighbourhood	neighbourhood_cleansed
21	GT23	Bijlmer-Oost	T		Oud-Oost
20	GT22	Bijlmer-Centrum	T		Geuzenveld - Slotermeer
22	GS25	Weesp, Driemond	S		Oostelijk Havengebied - Indische Buurt
19	GN20	Noord-Oost	N		De Pijp - Rivierenbuurt
13	GN19	Oud-Noord	N		Oostelijk Havengebied - Indische Buurt
18	GN18	Noord-West	N		De Baarsjes - Oud-West
17	GM17	IJburg, Zeeburgereiland	M	Amsterdam, Noord-Holland, Netherlands	Centrum-Oost
16	GM16	Watergraafsmeer	M	Amsterdam, Noord-Holland, Netherlands	Slotervaart
15	GM15	Indische Buurt, Oostelijk Havengebied	M		De Baarsjes - Oud-West
14	GM14	Oud-Oost	M	Amsterdam, Noord-Holland, Netherlands	Bos en Lommer
12	GK13	De Pijp, Rivierenbuurt	K		Centrum-Oost
11	GK12	Buitenveldert, Zuidas	K		De Baarsjes - Oud-West
10	GK11	Oud-Zuid	K	Amsterdam, Noord-Holland, Netherlands	Oud-Oost
9	GF10	Slotervaart	F		Centrum-West
8	GF09	De Aker, Sloten, Nieuw-Sloten	F		Bijlmer-Oost

The design of the review score filter was optimized by converting it from a tick box list, as initially implemented in Sheet 5, to a dropdown menu in the final design. This adjustment enhances the layout's clarity and tidiness while improving user interaction by providing a more intuitive filtering mechanism.

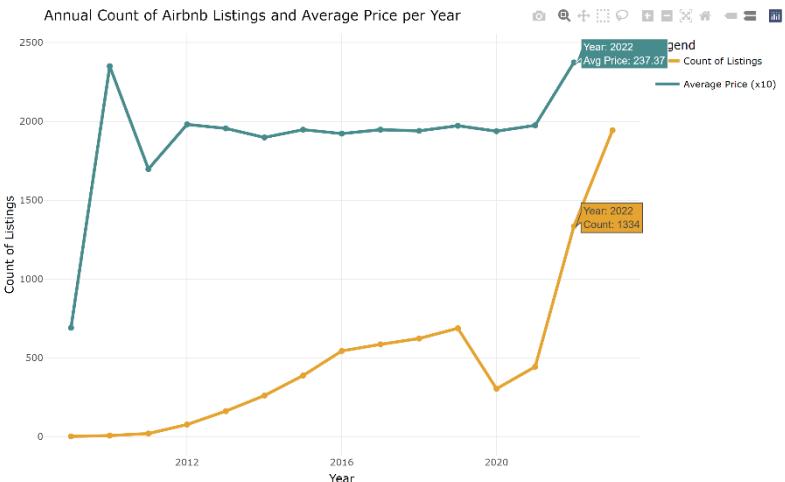
To ensure a coherent narrative flow, The layout of the visualization has also been adjusted from a single view dashboard to multiple views with multiple tabs for different purposes. The first tab is the introduction for the Data Visualization Project and overview of the Airbnb market, second tab for price variation factors, third tab for successful host attributes, and last tab for data source summary. In addition, the graphs were carefully repositioned, grouping related visualizations together. For instance, the top 5 most popular neighborhood tree map is placed alongside the choropleth map showing the average price and density of listings. Utilizing a sequential color scale, this choropleth map effectively demonstrates the density and price variations between different neighborhoods.

For the analysis of host attributes, a pie chart was included to illustrate the proportion of superhosts. This interactive element allows users to filter data based on host type, dynamically updating the violin chart of response time and the grouped scatter plot of beds and value rating. This interactivity enhances the exploratory capabilities of the dashboard, facilitating the discovery of insights regarding host performance.

In summary, the final design integrates a variety of visualization techniques and interactive elements to provide a comprehensive and insightful analysis of the Airbnb market in Amsterdam. Each visualization was meticulously selected and positioned to support the research questions, ensuring that the dashboard delivers a clear and coherent narrative to its users.

Interactive Narrative Visualisation Implementation

Based on the insights from "Genres of Narrative Visualization" by Edward Segel and Jeffrey Heer (2010), my dashboard strategically combines elements from both slideshow and comic strip formats to deliver a sophisticated visual narrative. This approach allows for the seamless integration of multiple frames, facilitating clear segmentation of graphs by different topics. The text font used throughout the dashboard is Sans serif to keep it clean and modern looking.

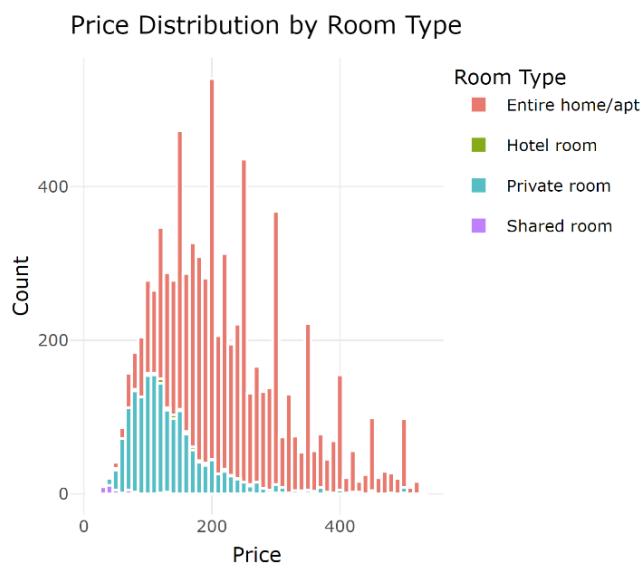


In the final implementation, I initiated a line chart illustrating the total number of listings, augmented by an additional line depicting the average price trend over the years. This dual-line presentation offers users a comprehensive overview of the Airbnb market in Amsterdam, with interactive features enabling them to explore specific data points. Notably, post-COVID-19 trends reveal significant peaks in both price and listings, reinforcing the narrative's key insights.

Following this overview, users can engage in a detailed price analysis, examining various factors

influencing price fluctuations such as accommodation capacity, room type, and neighborhood. This structured approach empowers users to gain comprehensive insights into the factors driving Airbnb prices in Amsterdam, facilitating informed decision-making.

Following an introduction to the overarching market dynamics, users are provided with a structured approach to engage in a comprehensive price analysis. This analytical endeavor aims to dissect the multitude of factors influencing price variation, encompassing accommodation capacity, room type, bed and bath count, and neighborhood location. By following this structured analytical schema, users can develop a holistic understanding of the diverse determinants that shape Airbnb pricing dynamics within the Amsterdam locale.

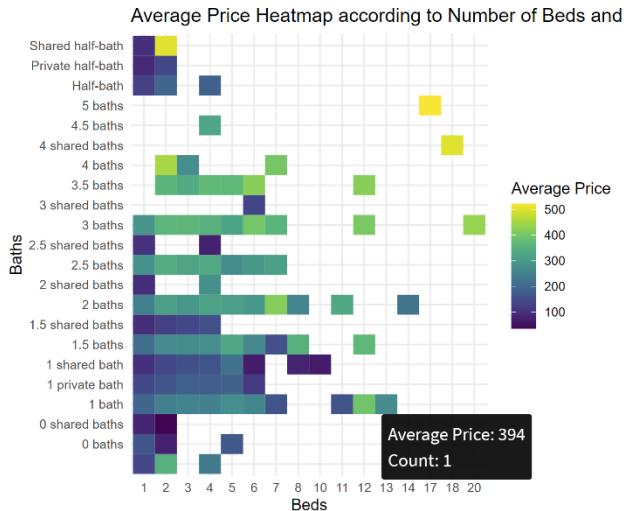


Given that price serves as the primary dependent variable of the analysis, a price range slider featuring adjustable upper and lower limits has been incorporated. This interactive feature empowers users to customize the price range applied to the histogram and tree map tabs. Notably, the slider's price range is meticulously determined post-removal of outliers from the dataset, ensuring the integrity of the data visualization. Specifically, listings with prices exceeding \$1000 are excluded to mitigate the impact of outliers on the data analysis outcomes.

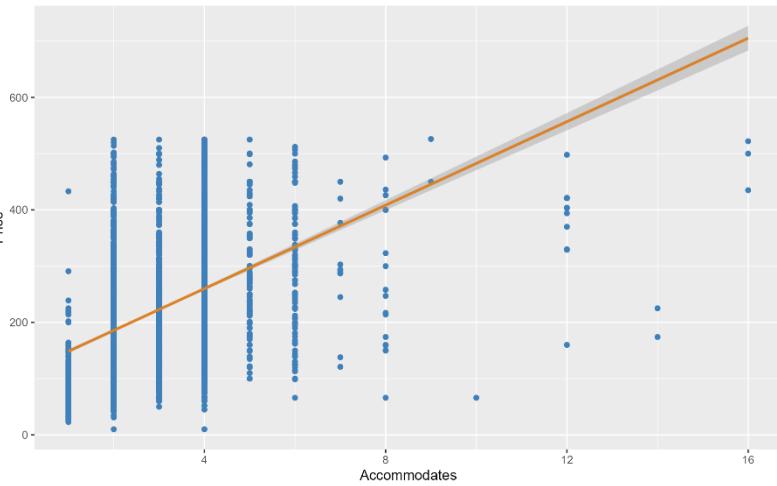
Commencing the price analysis tab is the histogram depicting price distribution alongside room type categorization. This graphical representation illuminates the distribution of prices across different room types, with emphasis placed on identifying the most prevalent room type, notably, entire home/apartment.

emphasis placed on identifying the most prevalent room type, notably, entire home/apartment.

The subsequent visualization implemented is a heat map, illustrating pricing variations based on the combination of beds and bathrooms. This graphical representation reveals a direct correlation between the number of beds and bathrooms and the corresponding price. Notably, the heat map also provides insights into the distribution of property sizes concerning baths and beds. While properties can accommodate up to 20 beds, the maximum limit for bathrooms is capped at 5. Additionally, viewers gain insights into the number of bathrooms available for a specific number of beds, with, for instance, one-bed properties offering up to three bathrooms. By hovering over any box on the heatmap, viewers can access information regarding the number of listings and their average prices for properties with specific bed and bathroom combinations.



Scatter Plot of Price and Accommodates



While the heat map offers a broad understanding of the correlation between price and accommodation capacity, a scatter plot is employed to delve deeper into this relationship. Notably, the presence of a linear regression trend line indicates a positive correlation between these variables. This trend line aids users in estimating the appropriate price for an Airbnb accommodating a certain number of individuals. However, the widening grey area towards larger accommodation capacities suggests a diminishing strength in the relationship with price. Consequently, individuals booking larger accommodations for larger groups of travelers can anticipate slightly lower prices.

The subsequent pair of graphs shed light on the influence of location on Airbnb listings' pricing. The first visualization is a tree map displaying the top five most sought-after neighborhoods and their respective average nightly prices.

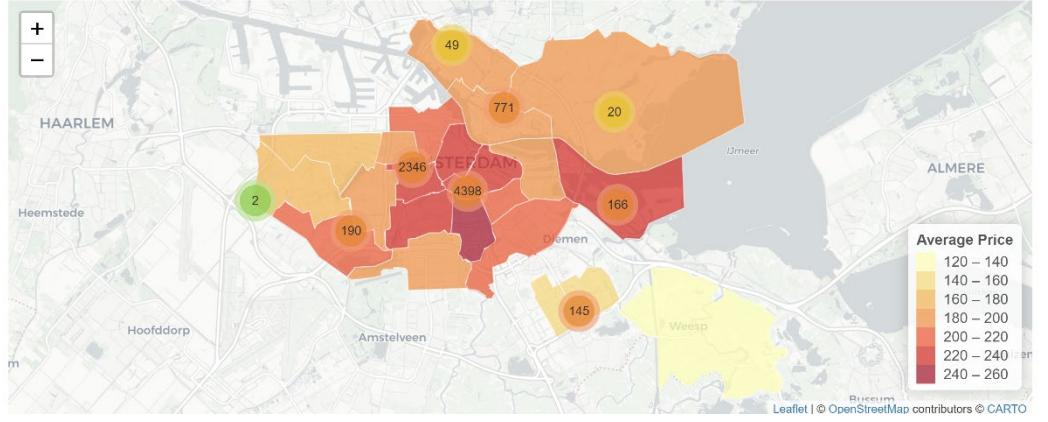
This information holds significance for travelers, as popular neighborhoods typically boast a wealth of tourist attractions and points of interest. While selecting an ideal neighborhood is paramount, cost considerations also play a crucial role. Hence, the tree map enables viewers to effortlessly compare prices across popular neighborhoods. Disparities in pricing between neighborhoods can offer insights into the amenities and surrounding environments. For instance, while most popular neighborhoods likely offer proximity to city attractions, pricier locales such as De Pijp and Rivierenbuurt may offer enhanced amenities and more desirable surroundings, such as local eateries, well-maintained streets, and tranquil nighttime atmospheres (AmsterdamWonderland, 2018).



The final graph delves deeper into the insights gleaned from the tree map, offering a combined view akin to a choropleth and bubble map.

This visualization not only delineates neighborhood boundaries but also pinpoints the precise locations of all listings, providing a more accurate distribution of properties on the map.

The choropleth component illustrates varying price points across different regions, with pricing largely correlating with property location. Utilizing a sequential color gradient, the map delineates distinct price ranges, effectively representing numerical data. Typically, properties situated farther from the city center command lower prices. Consequently, as one moves away from the urban core, both the price and density of Airbnb listings tend to decrease.

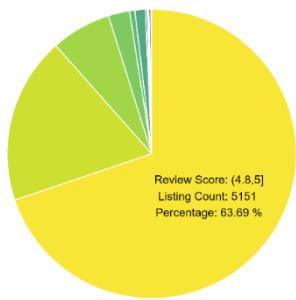


Select Review Score:

Select Review Score:

Review Score Rating

Superhosts tend to have better review score than non-superhosts in general. Successful host tend to have good review rating scores for their listings. In fact, listings under superhosts always have higher percentage of excellent rating of more than 4.8/5.0. The most contrasting review score category between two host type is review score for cleanliness. While there are 71.52 % of listings from Superhosts have review score over 4.8 stars, only around half of the listings (49.65 %) by non-superhosts are scored 4.8 and over.

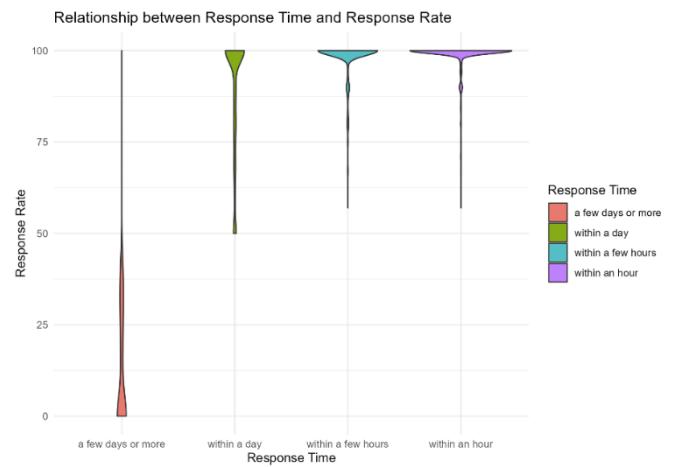


Transitioning to the subsequent tab, which delves into the second research question concerning the common attributes of successful hosts, I've integrated two interactive filters to enhance viewer engagement and comprehension of the graph insights. These filters are crafted to serve as user tasks for both data and view specification, aligning with Heer and Shneiderman's taxonomy of tools aimed at facilitating the seamless and adaptable utilization of visualizations (2012). Firstly, a tick box filter allows viewers to toggle between different host types, facilitating comparisons between Superhosts and non-Superhosts. Secondly, a dropdown menu enables viewers to select the type of review score they wish to visualize on the pie chart below.

The pie chart illustrates the distribution of review scores across various host types. By leveraging both the Superhost filter and the review score dropdown menu, viewers can juxtapose the review

scores of listings belonging to Superhosts against those of non-Superhosts. Notably, a compelling insight emerges regarding the cleanliness review score, wherein Superhosts notably outperform their non-Superhost counterparts. This finding resonates with an internal Airbnb survey, which reported that listings with a cleanliness rating of 4.8 or higher experienced a 20% increase in bookings compared to those with lower cleanliness ratings (The Economist, 2020).

Subsequently, the violin chart elucidates the relationship between response rate and response time. These metrics serve as pivotal indicators of a successful Airbnb host, directly reflecting the host's communication prowess and



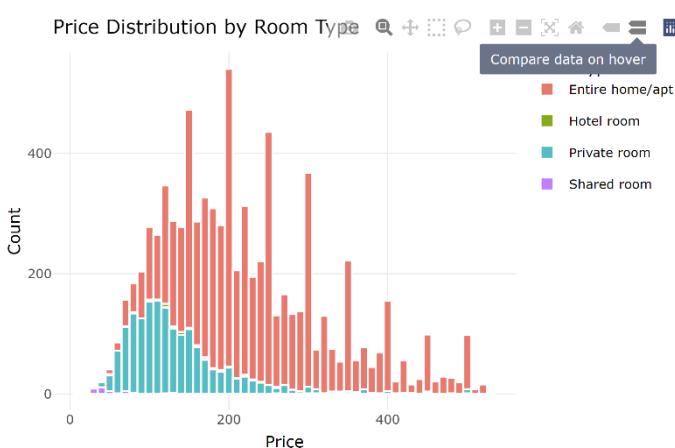
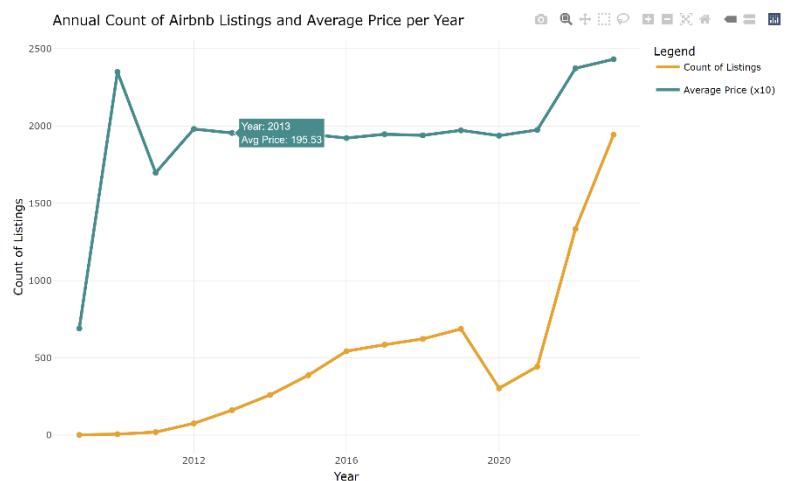
attentiveness to guests' needs. Generally, hosts who respond within a day or less achieve a response rate exceeding 50%. For Superhosts, maintaining a response time of within an hour typically results in a response rate ranging from 75% to 100%. Conversely, a lack of response after several days often indicates an absence of subsequent communication. Interestingly, this trend is particularly pronounced among superhosts with a response rate falling below 25%.

Finally, this graph offers an alternative perspective by comparing the response rate and the review score for value across hosts, including superhosts. The visualizations distinctly illustrate that superhosts generally exhibit more consistent positive attributes, such as response rate and value rating, whereas non-superhosts display greater variations in the quality of their listings and response rates. Consequently, it can be inferred that the designation of superhost serves as a reliable indicator of a successful host.

Using the Implementation

To utilize the application, viewers must utilize the RStudio application and install all required libraries to ensure seamless operation. While the necessary library packages are integrated within the application codes, it is imperative for viewers to update all packages to ensure their RStudio environment is current.

For accurate interpretation of the price line on the line graph, it is important to note that the position of the price line on the y-axis is scaled to be ten times higher than the actual price. This adjustment aids viewers in visualizing the price line more effectively. However, the values displayed when hovering over any point on the price line remain accurate.





The choropleth map provides viewers with a broad overview of the average price across different neighborhoods. However, upon clicking any region within the map, viewers gain the ability to zoom in and explore specific properties, thus obtaining accurate nightly price details for individual Airbnb listings. This enhanced functionality empowers viewers to delve deeper into the pricing dynamics of specific locations, facilitating a more nuanced understanding of the variability in prices across different areas.

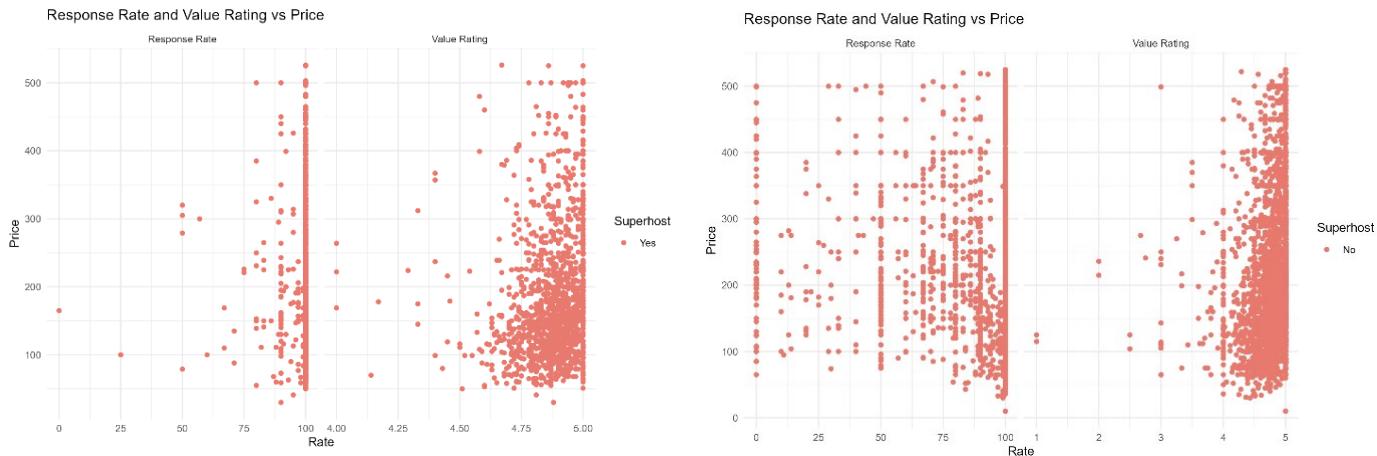
In addition to geographical insights, the dashboard incorporates superhost filters to enable viewers to compare various attributes associated with different host types. By utilizing these filters, viewers can narrow down their focus to a specific host type, such as superhosts, and explore the distribution of response rates and value ratings within this subgroup. This feature allows viewers to discern potential differences in the performance metrics of different host types, shedding light on the attributes that contribute to host success within the Airbnb ecosystem.

Superhost Filter

Filter Hosts:

Superhosts

Non-Superhosts



Moreover, by toggling between different host types, viewers can gain valuable insights into how factors like response rates and value ratings vary among superhosts compared to non-superhosts. This comparative analysis enables viewers to identify trends and patterns that may be indicative of successful hosting practices, thereby informing their decision-making processes when selecting accommodations on Airbnb.

Conclusion

In this project, I successfully created a comprehensive data visualization to explore and present insights about the Airbnb market in Amsterdam. Utilizing a variety of libraries including plotly, shiny, leaflet, ggplot2, and others, I crafted an interactive dashboard that elucidates key trends and factors affecting Airbnb listings. The final design, refined based on feedback, includes four main views: an overview of the Airbnb market, an analysis of price variation factors, an exploration of attributes of successful hosts and a dedicated page for data sources.

Throughout the project, several key insights emerged. For instance, adding an average price line to the total listings line chart helped illustrate price trends over time. Implementing a histogram of prices by room type provided clarity

on price distribution and room type prevalence. Additionally, the scatter plot with a linear regression trend line between price and accommodation capacity revealed a positive correlation. The combination of a choropleth and bubble map effectively demonstrated spatial price variations and precise listing locations.

In reflecting on the project, I've learned a great deal about the intricacies of data wrangling and the importance of a clear, narrative-driven approach to data visualization. One area for improvement would have been to secure the necessary datasets earlier in the process, allowing more time for refinement.

Looking ahead, I see potential for further enhancements. Future work could include integrating more dynamic filtering options and incorporating additional data sources to enrich the analysis. Additionally, exploring advanced visualization techniques such as animations or UI effects using CSS packages could provide even more engaging and insightful representations of the data.

Overall, this project has been an enlightening experience in creating a narrative visualization that not only presents data but also tells a compelling story about the Airbnb market in Amsterdam.

Bibliography

AmsterdamWonderland. (2018, May 13). *The best area to stay in Amsterdam with family - a neighbourhood by neighbourhood guide*. Amsterdam Wonderland. <https://amsterdamwonderland.com/where-to-stay-in-amsterdam-a-neighbourhood-by-neighbourhood-airbnb-guide/>

The Economist. (2020, December 8). *Airbnb guests seek out cleaner properties in the pandemic*. The Economist. <https://www.economist.com/graphic-detail/2020/12/08/airbnb-guests-seek-out-cleaner-properties-in-the-pandemic>

Munzner, T. (2014). *Visualization analysis & design*. Crc Press, Taylor & Francis Group.

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. <https://doi.org/10.1109/tvcg.2010.179>

Bertin, J. (2011). *Semiology of graphics : diagrams, networks, maps*. Esri Press.

Airbnb. (2020). *How much does Airbnb charge Hosts?*. Airbnb. <https://www.airbnb.com.au/resources/hosting-homes/a/how-much-does-airbnb-charge-hosts-288>

Berg, E. (2023). *Airbnb makes a comeback: the number of sleeping places doubled*. BNR. <https://www.bnr.nl/nieuws/economie/10527215/airbnb-maakt-een-comeback-slaapplaatsen-verdubbeld>

Pascoe, R. (2023). *Airbnb is back in Amsterdam and some councillors want more rules*. Dutch News. <https://www.dutchnews.nl/2023/10/airbnb-is-back-in-amsterdam-and-some-councillors-want-more-rules/>

Heer, J., & Shneiderman, B. (2012). Interactive Dynamics for Visual Analysis. *Queue*, 10(2), 30. <https://doi.org/10.1145/2133416.2146416>

Appendices

Appendix 1: Sheet 1

Title: AirBnb in Amsterdam
Author: Bao Ngoc Truong
Date: 11/5/2024
Sheet: 1
Task: Brainstorm

Centrum	\$300
Centraal	220
Zuidas	200
Oost	\$150
Centrum - Oost	150

Filter:
 Only the graphs that supports the questions will be chosen.
 As some graphs also has the same function, the ones with better visualization will be chosen instead.

1, 2, 3, 5, 7, 9, 10, 12, 13, 14, 16, 17, 18

Categorize

Price	1, 2, 5, 16, 17
Host	9, 12, 13, 14, 18
Overlap	3, 7, 10

Any graphs that shows the price determinants
 Graphs that can be used for both price & host
 Graphs to show variables that correlate with host success

Combine and refine

Once chosen the number in violin chart change accordingly

The not chosen one will be grayed out

can also do price map?

Gray out or become more translucent

when a neighborhood is chosen it will zoom in on that area

Violin graph
 Response Time
 count of hosts

Accom. 1 2 3 4 5
 Value Rating 1 2 3 4 5

10

1 Average Price by neighborhood

Centrum	\$300
Centraal	220
Zuidas	200
Oost	\$150
Centrum - Oost	150

Review comments
 friend
 happy, today, fun
 nice
 Host
 Welcoming
 recommended
 attention

Part of whole

5 Price vs Accommodation

Accommodation

12

13

14

15

16

17

18

Question

+ Does it answer the research questions?
 Somewhat. Might need to add more texts for explanation & setting contexts.

AIRBNB IN AMSTERDAM

Layout

Total Listings: Line chart showing growth from 2019 to 2023, reaching 12M with a 5% change. **Top 5 Most Popular Neighborhood**: Bar chart showing De Pijp - Rivierenbuurt (\$240), De Torenjes (\$229), Centrum - west (\$228), Zuid (\$222), and Centrum - Oost (\$227) with their average prices.

Price & Neighborhood map: Bubble map where size represents price and color represents neighborhood.

Super Hosts: Pie chart showing 68% Superhosts and 32% Non-Superhosts.

Response Time: Box plot showing response times ranging from few mins to 1 day.

Data Source: - - - - -

Room Type % Pro: Donut chart showing distribution of room types.

% of total: Heatmap showing the percentage of total listings by number of beds (1-9) and baths (1-6).

Price Histogram: Histogram showing price distribution across different price ranges.

Superhost: Scatter plot showing Price vs. Beds for Superhosts (orange) and Non-Superhosts (blue).

Title: Airbnb in Amsterdam
Author: Bab Ngoc Truong
Date: 11/5/2024
Sheet: 2
Task:

Operations	Action	Result
+ Hover at any boxes on the heatmap w specific number of beds + baths	+ Show count of listings .	
+ Choose a neighborhood on the tree map	+ All other neighborhood will greyed out. Apply neighborhood filter on other graphs.	
+ Click the pie chart to choose to see just Superhost or Non-Superhosts	+ The map will zoom in on the chosen neighborhood	
+ Click on one of the bubbles on the map	+ The other graphs will apply the same filter	
	+ Show price, property type number of accommodation	

FOCUS

Before: Neighborhood map

After Price Map

(1) Once a neighborhood is chosen, the rest will grey out
(2) The map zoom in & become bubble map
(3) Show the number of beds & baths & avg price of the chosen neighborhood

Discussion

1. Pros:

- Interactive visualizations help explore data with higher engagement
- Linked visualization → gain insights about relationships between variables.
- Customization to user preferences
- Connection between price & hosts

2. Cons:

- User might find it overwhelming especially for people who are not familiar with visualization dashboard.
- The dashboard with linked interaction might take longer processing time with large datasets.
- User with disability or color blindness might have a hard time interpreting the visualization
- Overwhelming amount of information
- Host & price visualizations are not strongly connected

Layout

AIRBNB IN AMSTERDAM

Data Source: -----

Choose Price Range:

Min: 0 Max: 1000

Price vs Accommodate

of Listings: 50, 100, 200

Accommodate

Price

Review Score

Location: Clean, Checkin, Comm., Value, Overall

Superhost: Yes, No

Price

Response Rate(%)

Value Rating

Review comments

friendly, happy, fun, nice, Host, location, Polite, recommend, Welcoming, attractions

Title: Airbnb in Amsterdam
Author: Bao Ngoc Trung
Date: 1/5/24
Sheet: 3
Task:

Operations

Action	Result
+ Click on a bar in the circular barplot	→ All graph filter to that neighborhood
+ Choose room type using the divided bar chart	→ filter room type on all graphs
+ Choose how many people need it for accommodation	→ filter number of people who will accommodate
+ Slide two ends of the slider bar to set max & min for price range	→ apply price minimum & max to all graphs

Focus

Total Listings by Neighborhood

Dost

Centrum, Dost, Lummel, Zuidas

Count

Villa, Apt., House

Hostel, Boutique, Bed & Breakfast

Accommodate

Price

Review Score

Location: Clean, Checkin, Comm., Value, Overall

Apply filter for room type

filter the chosen price range and number of accommodation

Discussion

Pros:

- Interactive visualizations help explore data with higher engagement
- Linked visualization → gain insights about relationships between variables.
- Customization to user preferences

Cons:

- User might find it overwhelming especially for people who are not familiar with visualization dashboard.
- The dashboard with linked interaction might take longer processing time with large datasets.
- User with disability or color blindness might have a hard time interpreting the visualization
- Overwhelming amount of information
- Host & price visualizations are not strongly connected
- Wordcloud not linked to anything

AIRBNB IN AMSTERDAM

Data Source: - - - - -

Layout

Total bookings by room type

Room Type	Bookings
Home/Apt	~10K
Hotel	~5K
Private	~5K

Top 5 Most Popular Neighborhood

Neighborhood	Price
De Pijp - Rivierenbuurt	\$340
De Baarsjes	\$283
Centrum-West	\$282
Zuid	\$282
Centrum-Oost	\$227

Price

Superhost Yes No

SUPER HOSTS

REVIEW SCORE

- Overall
- Cleanliness
- Communication
- Check-in
- Location

Review score

Response Time

FOCUS

① Choose a price range & property type

② Choose location

③ Ratings of properties in chosen location will be highlighted. The rest are greyed out

Title: Airbnb in Amsterdam
Author: Bao Ngoc Truong
Date: 1/5/24
Sheet: 4
Task:

Operations

Action	Result
Click on a bar in the histogram	→ All graphs filter to that price range & room type
Click on one of the boxes in tree map	→ filter neighborhood on all graphs ↓ Non-chosen ones will be greyed out
Hover over one of the boxes on heat map	→ Show average rating
+ click the pie chart to choose to see just Superhost or Non superhosts	→ The other graphs will apply the same filter

Discussion

1. Pros

- Interactive visualizations help explore data with higher engagement
- Linked visualization → gain insights about relationships between variables.
- Customization to user preferences

2. Cons:

- User might find it overwhelming especially for people who are not familiar with visualization dashboard.
- The dashboard with linked interaction might take longer processing time with large datasets.
- Users with disability or color blindness might have a hard time interpreting the visualization
- Overwhelming amount of information
- Host & price visualizations are not strongly connected
- Unlinked usage of stream graph for comments about host

<p>Layout</p> <p>AIRBNB IN AMSTERDAM</p> <p>Total Listings: 12M (5% change 5Y)</p> <p>Top 5 Most Popular Neighborhoods:</p> <ul style="list-style-type: none"> De Pijp - Rivierenbuurt: \$240 Oe. tourrijer: \$229 Centrum-west: \$228 Zuid: \$232 Centrum-Oost: \$227 <p>Data Source: -----</p> <p>Choose Price Range: Min 0 Max 1000</p> <p>Price Map: Price range: \$500, \$250, \$100. Labels: Price: \$8, Location: Zuid, Prop Type: Apartment.</p> <p>% of Listings vs Beds: 0-100% vs 0-9 beds.</p> <p>Beds: 2, Bath: 1, Count: 88.</p> <p>Price vs Superhost: Price (0-350) vs Superhost (Yes/No).</p> <p>Response Rate vs Value Rating: Response Rate (25-100) vs Value Rating (1-5).</p> <p>Response Time: 1 day, few hours, 1 hour, few mins, 5k-10k count of hosts.</p> <p>Review Score: 4.8-5.0, 4.5-4.8.</p> <p>NON-SUPERHOST vs SUPERHOST filters.</p> <p>REVIEW SCORE filters: Overall, Cleanliness, Communication, Check-in, Location.</p> <p>Operations:</p> <p>Action</p> <ul style="list-style-type: none"> + Click on one of the box in the tree map + Slide two ends of the slider bar to set max & min for price range + Click on data points on visualizations + Tick/Untick the boxes to see only either Superhosts or Non superhosts + Hover over data points on visualizations <p>Result</p> <ul style="list-style-type: none"> → All graphs filter to that neighborhood → apply price minimum & max to all graphs → Show location, property type and number of people it can accommodate → The other graphs will automatically adjust. → Show labels (count of host or listings) 	<p>Title: Airbnb in Amsterdam Author: Bao Ngoc Truong Date: 12/5/24 Sheet: 5 Task:</p>
<p>Focus:</p> <ol style="list-style-type: none"> Once a neighborhood is chosen, the rest will grey out The map zoom in to show the price distribution of specific neighborhood Show the number of beds & baths & avg price of the chosen neighborhood <p>Before: Neighborhood Map</p> <p>After: Price map</p> <p>Detail</p> <ul style="list-style-type: none"> - Data processing & cleaning is needed for quality & reliability of data - Statistical analysis methods such as regression analysis or clustering may be employed to identify patterns - Visualization algorithms including interactive maps, charts & graphs, selected filters will be used for visualization. - R-Shiny will be used to build visualization - Data Storage & processing capabilities should be sufficient to handle the volume of data & computational requirements of algorithms used in the dashboard. 	