Unlocking the Secrets of Homestays Pricing using immersive visualisations

Chen Yun
Singapore Management University
Singapore, Singapore
yun.chen.2020@scis.smu.edu.sg

Lim Pei Shyan
Singapore Management University
Singapore, Singapore
pslim.2021@scis.smu.edu.sg

Edwin Chong Tong Tong
Singapore Management University
Singapore, Singapore
edwin.chong.2021@scis.smu.edu.sg

Ng Mahn Yin
Singapore Management University
Singapore, Singapore
mahnyin.ng.2021@scis.smu.edu.sg

Justus Jonas Nau
Singapore Management University
Singapore, Singapore
jj.nau.2023@exchange.smu.edu.sg

***Abstract*—In this project, we conduct an in-depth analysis of homestay pricing in New York City using a comprehensive dataset from Airbnb. Our research explores a wide range of factors, including both the unique characteristics of homestays and the amenities in their surrounding neighborhoods, to uncover their impact on pricing trends. To gain a more comprehensive understanding, we integrate multiple datasets, broadening the scope of variables under scrutiny. We employ various data visualization techniques to present our findings in a clear and accessible manner.**

Keywords—Data Visualization, Data Analysis, Amenity, Tourism, Subway, Homestay, Pricing Factors.

I. Introduction

In an urban environment, travelers often seek cost-effective accommodation options beyond traditional hotels, with Airbnb being one of the options. Against the backdrop of high inflation and various economic uncertainty, factors that determine booking decisions become increasingly complex for those seeking justifications for their expenses, as evident by the review of their fee system to counter customer's concerns over transparency in 2021 [1]. While affordability remains to be the main concern for travelers, hosts are also struggling to maintain profitability amidst a far more competitive market for short-term stays. In 2022, majority of the top short-term rental markets had experienced low occupancy rates with some even hitting 0% towards the end of the year [2]. In other words, hosts should be looking to provide reasonably competitive rates and the appropriate level of service, especially those in areas of increasing supply. This project will look to leverage visualization in providing greater insight for hosts and travelers to enhance the experience for both parties by considering property listings, local attractions, and overall guest satisfaction.

II. Motivation

As homestays have emerged as a popular accommodation option for travelers since 2011 [3], their pricing dynamics remain intricate and multifaceted. This project aims to shed light on this complexity by utilizing the power of data visualization techniques. By combining the datasets, we will be able to take a deeper and more comprehensive look at the pricing determinants. Through this project, we aspire to empower both travelers and hosts with valuable information, as currently, there is a lack of such in-depth analysis. Ultimately, this project hopes to provide useful insights for travelers and hosts, allowing them to make informed decisions and contributing to the continued growth of the homestay ecosystem.

III. Literature Review

The most common usage of the Airbnb dataset is for individual analysis which focuses on the performance of individual Airbnb properties to maximize profit and to find the best places suitable to open a new profitable Airbnb properties. One of the examples is Awning, they provide analysis services to calculate and estimate the Airbnb profit margin [4]. These analysis services are often offered as a commercial service and tend to cater to property owners seeking profit optimization.

There is also a research paper which focuses on the attributes that influence Airbnb user`s experience on a dataset about Sydney [5]. This study found, that Airbnb guests typically consider the same accommodation features that are commonly associated with hotel stays. The majority of written feedback of users focuses on location, amenities and the host.

IV. Datasets

For this project, we will be utilizing datasets provided by Inside Airbnb [6], a website that offers enriched Airbnb datasets for various cities across different countries. Our focus will be on U.S. cities, specifically New York. We have obtained a dataset consisting of 39,453 listings and a comprehensive set of 75 columns. However, for the purposes of this project, we will concentrate on a select set of 9 columns as outlined in Table I.

TABLE I. AIRBNB DATASET

Column	Description	Data Type
Latitude	Geometry location	Number
Longitude	Geometry location	Number
Property Type	Type of property. E.g., entire unit, private room, condo	String
Bathrooms	Number of bathrooms	Number
Bedrooms	Number of bedrooms	Number
Amenities	Amenities within the property	String
Price	Price of the property per night	Number
Number of Review	Number counts of review given	Number
Review Score Rating	Average review rating score	Number

In addition to the Airbnb listing data, we have augmented our analysis with supplementary datasets related to urban development and facilities in New York. These datasets encompass tourist locations [7] and subway stations in New York [8], which will enable us to explore how property pricing is influenced by the surrounding neighborhood amenities and transportation options.

TABLE II. NEW YORK TOURIST LOCATIONS

Column	Description	Data Type
Name	The name of tourist attraction	String
Address	Address of tourist attraction	String
Zipcode	Zipcode of tourist attraction	Number

TABLE III. NEW YORK SUBWAY LOCATIONS

Column	Description	Data Type
Name	The name of subway station	String
Latitude	Geometry location	Number
Longitude	Geometry location	Number

V. TOOLS AND RESOURCES

For data preprocessing and exploratory data analysis, we will use **Python** through **Jupyter Notebook**, a web-based platform enabling the creation and sharing of documents containing live code, visualization, and narrative text. To enhance our analysis, we will use:

- **NLTK**: A toolkit offers various tokenizers to assist with text processing.
- **Pandas**: A popular library which aids in data manipulation and analysis during pre-processing and exploratory analysis.

For version control, collaboration, and file management within our team, we will utilize **GitHub**. We will develop an interactive web application for sharing data insights. To achieve this, we will leverage:

- **D3**: A JavaScript library for crafting dynamic and interactive data visualization, especially useful for creating complex customized visuals.
- **Leaflet**: An open-source JavaScript library for interactive maps
- **Ant Design**: A popular React UI library that provides a comprehensive set of design components and styles for building modern web applications.

VI. TASK

For this project, we will focus on understanding how different attributes contribute to the price formation of Airbnb listings. Here are the tasks we will perform:

A. How do common attributes affects Airbnb listing prices

In this task, we will conduct a comprehensive comparison of listing prices against common property attributes, including room types, room popularity, owner information, and review ratings, among others. This analysis aims to provide viewers with a fundamental understanding of how various common factors may influence the listing price. Exploring these correlations will offer valuable insights into the dynamics between common property attributes and Airbnb listing prices, benefiting both property owners and potential tenants in making informed decisions.

B. How do amenities affect Airbnb listing prices

Our focus in this task will be on delving into the amenities associated with listings. We aim to identify common amenities mentioned in most listings and selectively analyze some of the popular ones. Through in-depth analysis, we seek to understand precisely how each of these amenities affects the listing price. This exploration can assist property owners in identifying elements to include or avoid when preparing a property, ultimately influencing the listing price. Additionally, it provides insight for tenants on whether the requirement for certain amenities incurs a higher cost of staying.

C. How do nearby facilities affect Airbnb listing prices

In this task, we will explore the influence of external factors, such as nearby facilities—specifically tourism spots and subway stations—on Airbnb listing prices. Our objective is to visually represent the relationship between the proximity of these facilities and listing prices. In our project, 'nearby' is defined as a distance less than 1 kilometer between the listed property and the facilities. Through visualizations, we aim to uncover insights into whether the proximity of these facilities has a discernible impact on the pricing of listings.

VII. METHOD

A. Data Preprocessing

We utilized three datasets for our dashboard project: Airbnb Data, New York Subway Station Data, and New York Tourist Location Data. The Airbnb dataset comprises 39,453 rows and 76 columns, containing information about Airbnb listings in New York. The New York Subway Station Data includes 473 rows and 5 columns, providing details such as the station name, subway line affiliation, longitude, and latitude. Lastly, the New York Tourist Location Data consists of 347 rows and 5 columns, offering information on popular tourist attractions, including their address, zipcode, and coordinates (longitude and latitude).

For the Airbnb Data, we extracted 15 features from the 76 columns based on contextual knowledge of the homestay industry. These features include 'Name,' 'Host Id,' 'Price,' 'No. of reviews,' 'Rating,' 'Latitude,' 'Longitude,' 'City,' 'Property type,' 'Room type,' 'Accommodates,' 'Bathrooms,' 'Bedrooms,' 'Beds,' and 'Amenities.' Data manipulations were necessary for features such as 'No. of reviews,' 'Price,' 'Property type,' 'Bathrooms,' and 'Amenities.'

The 'amenities' column, for instance, required word tokenization to convert it into useful data. After analysis, we identified 5572 unique keywords. Sorting them by frequency, we retained amenities present in more than 10% of the overall listings, resulting in 66 unique amenities. To streamline the selection further, we handpicked the final amenities, resulting in a set of 23. These were chosen based on both frequency and business logic.

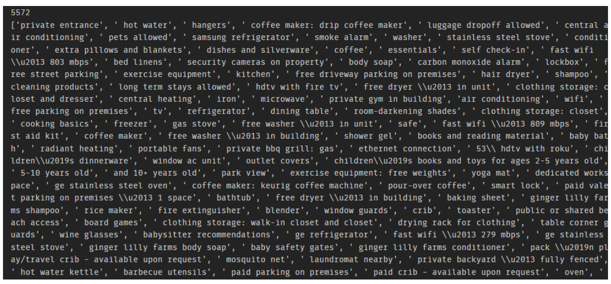

Fig. 1. List of amenities collected from amenities column

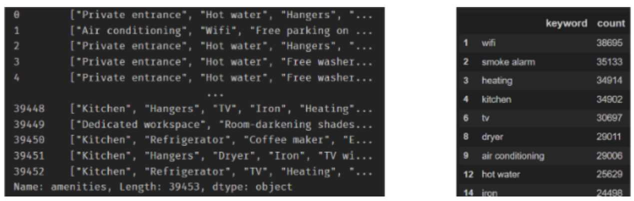

Fig. 2. List of amenties in each row vs hot encoding of amenities

Outliers, including listings with unusually high prices (up to $30,000 per night) and those with up to 50 bedrooms, were removed. For ‘bedrooms’, the dataset was refined to include listings of up to 5 bedrooms and for ‘price’, it is adjusted to only reflect those within a specified range.

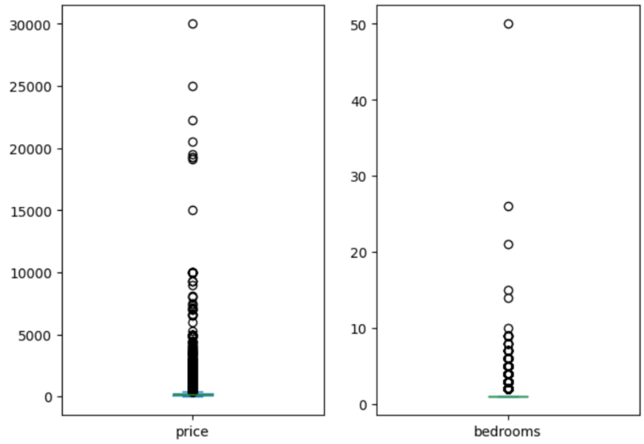

Fig. 3. Boxplot of price and bedrooms before removing outliers

The New York Train Station and Tourist Location datasets were both leveraged by utilizing their coordinates, which were geocoded from zip codes for the Tourist Location dataset.

B. Major Challenges – Missing Data

A challenge we encountered with the processing of our dataset is the appearance of a large percentage of missing values in some of the important features, predominantly in the ‘Bedrooms’ and ‘Review Score Rating’ columns which made up 42.8% and 26% of the data respectively.

Variable Name	Missing Percentage
id	0.00
name	0.00
host_id	0.00
latitude	0.00
longitude	0.00
property_type	0.00
room_type	0.00
neighbourhood	41.92
accommodates	0.00
bathrooms_text	0.13
bedrooms	42.82
beds	1.53
amenities	0.00
price	0.00
number_of_reviews	0.00
review_scores_rating	25.96
city	0.00
dtype: float64	

Fig. 4. Missing data in columns (in percentage)

For the ‘bedrooms’ column, they were reasoned to be empty due to their property types being ‘private room’ or ‘entire rental unit’. For ‘entire rental unit’ listings with empty values, we decided to impute the values based on the median of that property type.

property_type	id	bedrooms
Private room in rental unit	9148	0
Entire rental unit	2499	0
Private room in home	2174	0
Private room in townhouse	1092	0
Private room in condo	516	0
...	...	...
Private room in dorm	1	0
Private room in castle	1	0
Private room in camper/rv	1	0
Entire home/apt	1	0
Private room in kezhan	1	0

65 rows × 2 columns

Fig. 5. Count of rows with missing bedrooms by ‘Property types’

While analyzing the ‘property_type’ column, we also discovered that there were many different types of properties listed and many of them were unique and had less than 10 occurrences. As these data only take up a small proportion of our dataset, we decided to remove them to simplify our data processing in the future. We then decided to remove listings that have less than 10 listings, which resulted in up to 40 unique property types and 111 rows of data to be removed.

For ‘Review Score Rating’ it is surmised that it is due to the lack of reviews left by a notable amount of people and thus the empty values were replaced with ‘no reviews’ to improve clarity.

C. Major Challenges – Lack of Time Series Data

The data collected from Inside Airbnb provides valuable insights into pricing and listing details at a specific point in time during the scripting process. However, it's important to note that this dataset lacks information on listing variations across different time frames. Consequently, our analysis is limited, preventing us from conducting a comprehensive examination of price changes over different months or time periods.

D. Major Challenges – Technical Depth in D3

Fig. 6. Map with coloured cluster cricles

As a team new to D3, we embarked on a steep learning curve to acquire various D3-related skills within a short timeframe. Consequently, the team may not have fully harnessed the advantages of D3. One notable example is the implementation of colored circles on the map. While the original intent was to display the density of listings within the region, technical challenges prevented us from implementing a sequential color map, which would have been a more suitable approach for this specific case. Despite this limitation, the team remains committed to refining our D3 skills for future enhancements.

VIII. Visual Analysis Results and Discussions

A. *Scatter plot of price vs different factors*

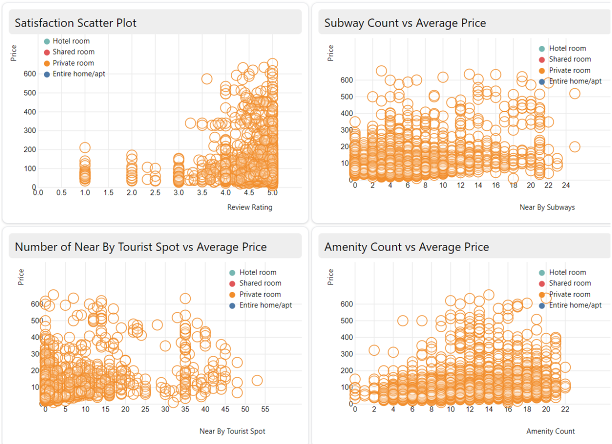

Fig. 7. Relationships between review ratings, subway accessibility, tourist spots, amenities, and pricing

Fig. 7 consists of a few scatter plots, which were utilized for the visual identification of the distribution and trends that may emerge between the variables and assess any possible correlations which is crucial in understanding how they change in relation with one another, as well as easily isolate any outliers, which provides insights into their relationships.

Focusing on the ‘Private Rooms’ property type, the Satisfaction scatter plot in Fig. 7 compares the distribution of the review ratings against price, reveals that most of the data points have predominantly consistent high ratings, with only a few sporadic instances with cheaper pricing associated with lower ratings.

In general, there appears to be limited correlation between satisfaction and price, although there might be slight variability towards the lower end of the pricing spectrum, suggesting that pricing alone does not strongly signal customer satisfaction and other factors like amenities and location play contributory roles.

Examining the remaining three graphs, further illustrates the weak correlation between subway proximity, tourist spot density and amenity count against average price. The data points exhibit a fairly even distribution across the entire pricing range, regardless of the magnitude of the respective variables, particularly the subway and tourist spot counts. Plausible explanations for these observations could stem from the context of New York as the location of interest, where a well-connected public transport system extends beyond the immediate proximity of train stations. The prevalence of cab-hailing, for instance, could potentially account for the lack of a clear trend between subway count on price. With respect to the amenity count, while it is no means definitive indicator of price, there is a subtle inclination for higher amenity count exerting some influence of commanding a higher fee.

Importantly, these conclusions are not exclusive to private homes; when examined on the scatter plots, ‘entire homes and apartments’ property type mirror the same trends as well. These are useful in addressing analysis Task A. It is crucial for property owners to understand these nuanced correlations for setting competitive prices and maximising property value.

B. *Wordle Cloud for Amenities*

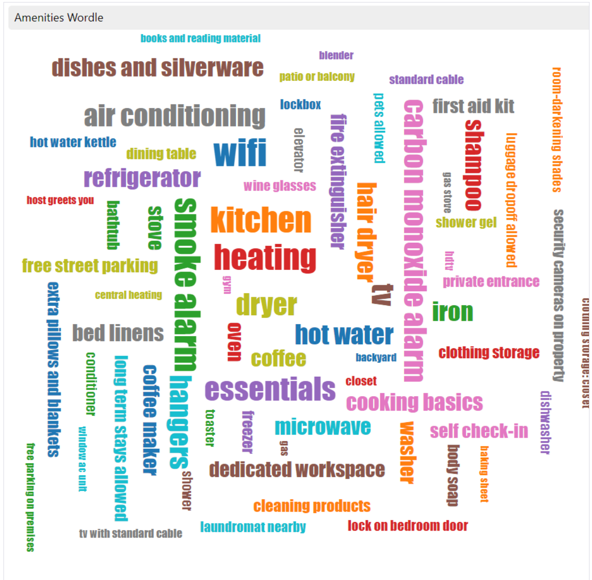

Fig. 8. Wordle visualising frequency of amenities mentioned in reviews

To vividly portray the frequency of amenities mentioned in reviews in an engaging manner, we employed a wordle-style word cloud. This visual representation offers a quick and intuitive glimpse into the most common terms, utilizing visual hierarchy for easy interpretation. The word cloud not only serves its purpose functionally but is also aesthetically pleasing. Additionally, to optimize space utilization, we implemented different text formatting, both vertical and horizontal, considering the varying word lengths of different amenities.

This word cloud aids in understanding Task B. By assessing the font size of each word, it is evident that amenities such as Wi-Fi, kitchen, heating, smoke alarm, and TV are frequently mentioned in the listings. This is helpful to address analysis Task B as it highlights the key amenities to focus on.

C. *Bar charts of Amenities vs Price*

Fig. 9. Price comparisons depending on presence of amenity and ranking of key amenities.

This amenity dashboard aims to offer significant insights into how various amenities affect pricing, as outlined in Task B. Presenting the bar graphs in small multiples/trellis charts facilitates comparisons across numerous variables/categories, reducing visual clutter and enhancing readability. This approach allows for a detailed examination of a specific subset of data, in this case, focusing solely on amenities. This is especially valuable when dealing with a large number of different categories, ensuring that important details are not lost. Accompanied by the added coloured labels, this method effectively conveys data and its variances, particularly in terms of percentage changes, with a single glance.

Highlighting the impact of amenities on pricing in Fig. 9, a noteworthy observation is the significantly higher price of properties inclusive of a TV. This finding is intriguing, given the prevailing notion that TVs are becoming more obsolete due to changing viewing habits favoring streaming services and personal devices over conventional television entertainment. This raises questions about the perceived value of having a TV within the property. Nevertheless, this apparent surprise could be addressed by the assumption that TVs are a standard amenity and a fundamental part of the package at upscale accommodations.

For property owners, this underscores the balance required meeting the traditional expectations of consumers as well as adapting to evolving trends. It emphasises the need to understand and navigate customer preferences to make informed decisions regarding inclusion of amenities and pricing strategies.

D. *Price vs Review Rating*

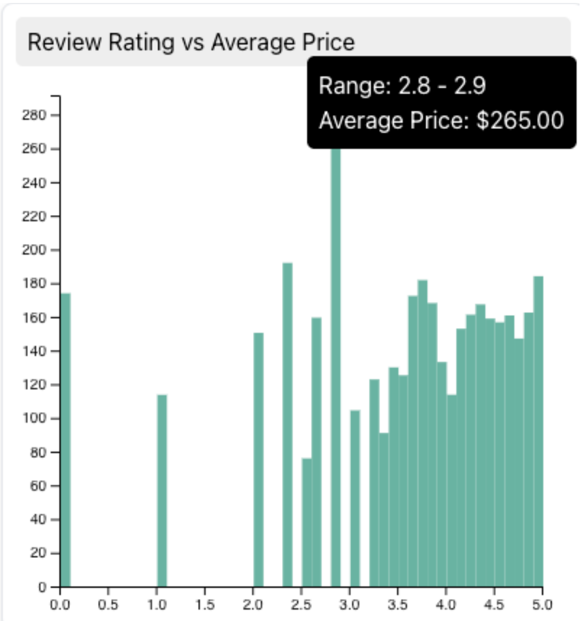

Fig. 10. Relationship between average price and review ratings

Fig. 10 presents the dispersion of average price for the respective review rating which is associated with Task A that provides an overview of how review ratings on listing affects the price.

The use of a narrow bar chart allows for high level of data density or categories, in this case the full range of the review ratings, together with effective space allocation. The notable peak in average price review rating 2.8 – 2.9 could once again, reflect the customers' expectations at pricey accommodations, like the anticipation of elevated service standards and amenities when paying a premium. Consequently, reviews will be less forgiving when expectations aren't met. However, it needs to be acknowledged that without accounting for property type and accommodation capacity, these conclusions may oversimplify this relationship between price and ratings.

E. *Room Type Distributions*

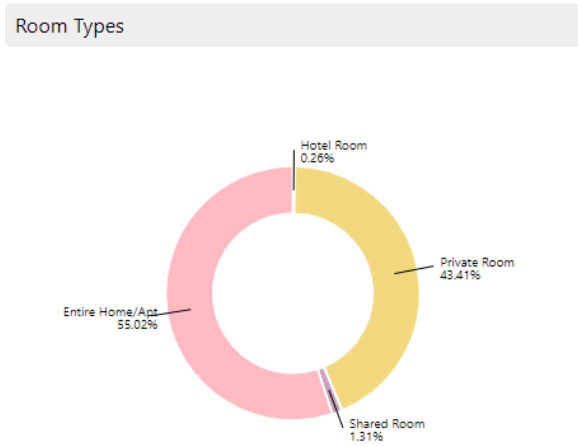

Fig. 11. Distribution of different property types

The donut chart used aids in efficient comprehension of the composition of property types that make up the dataset, by emphasising the relative percentages and leveraging the limited categories. This somewhat helps to discover the finding for Task A as the distribution of the room types may affect the listing price. In this instance, by drawing attention the minimal representation that hotels make up in the total number of properties in relation to the other property types, it renders us the clarity of price peak at review rating 2.8 – 2.9 being predominantly composed of hotel costs.

F. *Price vs Review Rating*

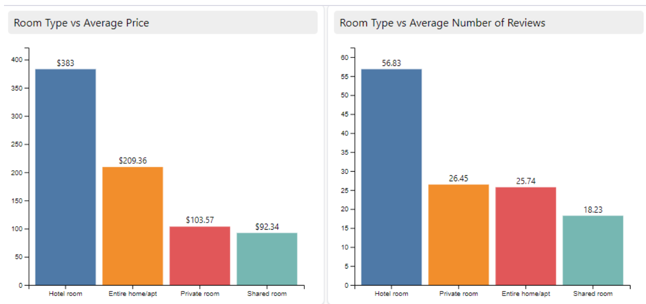

Fig. 12. Comparisons across room types with average price and number of reviews.

These bar charts further help us to understand our analysis Task A. The inclusion of the above bar and donut graphs in Fig. 11 and 12, enriches the interpretation of Fig. 10, illustrating the distinct price and review count variation across the different property types, with 'Hotels' commanding the highest prices at almost $400 despite accounting for only a negligible 0.26% of all properties. In contrast, 'Entire homes and Apartments' emerge as the next most expensive category but make up a significant proportion of all properties along with 'Private homes' at 55% and 43% respectively, as well as similar average review counts at around 25 per listing.

One possible inference could be that the private rooms experiencing an average price of approximately 100, implies that tenants seeking to balance affordability and a reasonable standard of service/amenities which in turn translate to high ratings, even if they are not luxury accommodations.

G. Listing and Facility Distribution Map

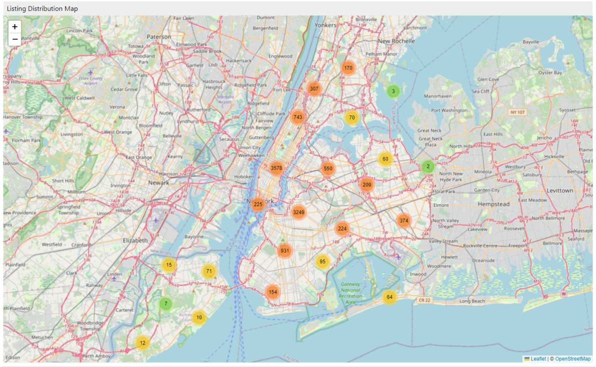

Fig. 13. Overview of Airbnb listing distributions in New York

We employed a map which can easily showcases the distribution of the Airbnb listings and facilities across New York, which allows us to quickly grasp the concentration and spread of accommodation in specific areas. This helps us to perform our analysis task C which focus on how nearby facilities affect the listings price. The interactive elements (zoom in/out, panning) allows users to focus on a particular district or make comparisons of the listing densities between neighbourhoods.

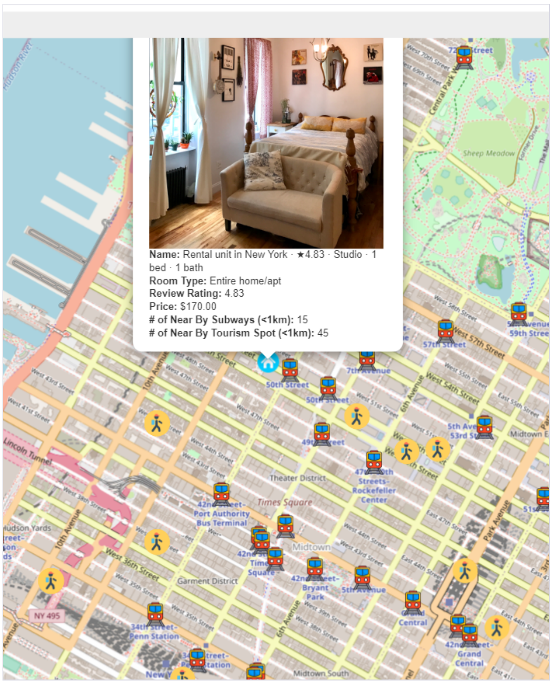

Fig. 14. Selected Airbnb Listing

With the option to select a specific listing, it gives tenants the ability to conduct detailed evaluation of the property and its characteristics by considering location, the surrounding facilities according to their set requirements, and make comparative assessments with other alternatives.

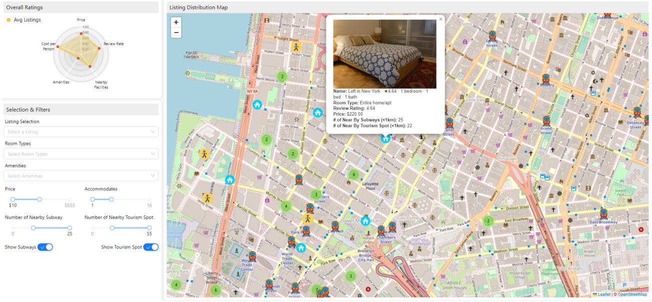

Fig. 15. Map with Filter and Selector

The usage of filters enables tenants to narrow their desired preferences, simplifying an efficient location-based search of finding several suitable. This not only saves time and effort of sifting through irrelevant options but reduces information overload, improving the usability of the platform, particularly in densely populated areas. Property owners can similarly also leverage the filters to pinpoint properties that share the same demographic appeal and acquire perspective of features that can be improved for enhanced competitiveness.

H. Overall Radar Chart

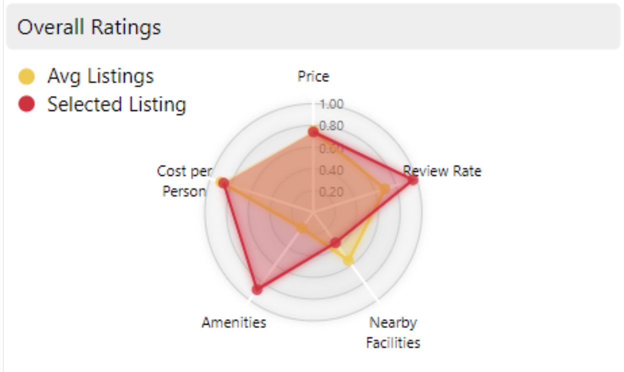

Fig. 16. Radar Chart Comparsion

Utilizing a radar chart allows for a comprehensive comparison of multiple factors simultaneously, addressing analysis Task A, B, and C. The visual representation facilitates a quick assessment of how selected listings fare across different dimensions compared to the average. The team focused on five key measurements: price, review rating, the number of nearby facilities, the number of amenities, and cost per person. To enhance comparability, a standardization method was employed, mapping original values to a scale of 0 to 1. A value of 0 indicates the least recommended or worst performance, while 1 represents the most recommended or best performance.

This approach benefits tenants by providing increased confidence in rental choices through easy identification of strengths and weaknesses in contrast to the average. Additionally, it offers owners greater foresight into various aspects, empowering them to make strategic decisions aligned with market demands.

IX. DEMO OF THE VISUAL ANALYTICS SYSTEM

As mentioned earlier, our visual analytics system is entirely powered by D3 and is accessible through this link: https://main.d797d423fvl0f.amplifyapp.com/. For local access, please ensure you have Node.js installed (version 18.0 or higher). Once you have all the prerequisites, navigate to the source code folder, and execute the following terminal commands:

1. npm install
2. npm start

The first command installs the necessary packages and dependencies required to run the visual analytics system. The second command launches the system locally and can be accessed via http://localhost:3000/ by default. The system comprises four distinct dashboards: Tenant, Property Owner, Amenity, and Map.

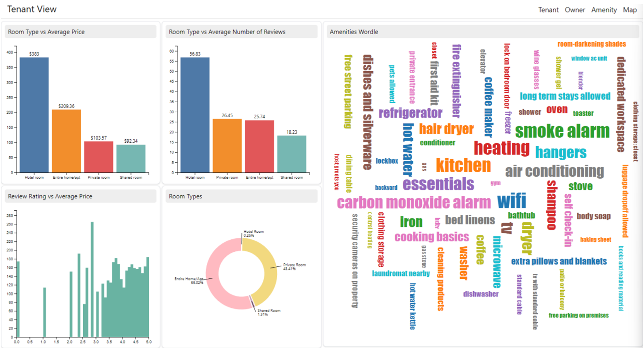

Fig. 17. Tenant Dashboard

The Tenant dashboard provides an overview of the price and popularity of different room types. Tenants can view information such as the proportion of various room types, the difference in average listing price between room types, their review ratings, as well as the average number of reviews per room type. An amenities wordle is also available for tenants to see a brief overview of the popular amenities available in the listings.

This information can help tenants to access the proportion of various room types available in the Airbnb listings, their average listing price comparisons, the quality of the room types, as well as an overview of the popular amenities in the listings, which can help tenants identify specific features that they may want to prioritize.

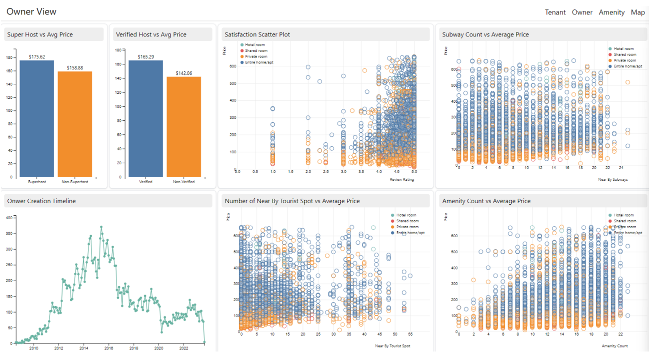

Fig. 18. Property Owner Dashboard

The Owner Dashboard offers insights into the distributions of various attributes that may impact the listing price. Owners can see evaluate the benefits a superhost and verified host status by comparing against their average listing price, as well as the effects of the number of amenities and subways can have on the listing price.

Fig. 19. Amenity Owner Dashboard

The Amenity dashboard provides detailed visual analytics on popular amenities and their corresponding prices. Owners can have a general overview of the possible effects on a listing's price with and without a specific amenity, as well as the necessary amenities present in most listings.

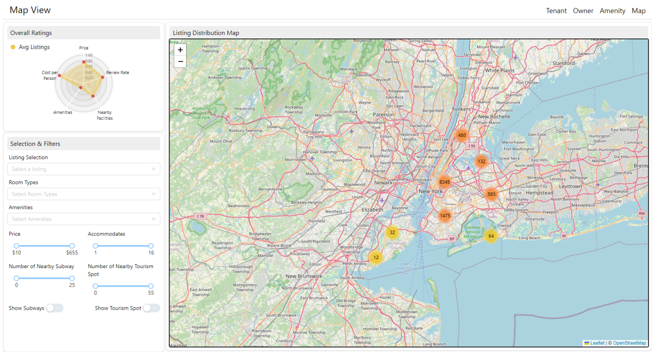

Fig. 20. Map Dashboard

Lastly, the Map dashboard includes a map that displays the distribution of different listings. The map can be enlarged to show up to details of specific listings. There are also multiple filters and selectors for tenants to refine their choices, such as filtering by price, nearby subways, tourism spots, room types, and amenities. You can also toggle to display subways and tourist spots in the map.

X. FUTURE WORK

As we navigate the next step in our exploration of homestay pricing, we've identified areas for growth and improvement in our work. We would like to create a more interactive and engaging experience for users, building upon the foundation of our current visualisations.

A. Adding dynamic interaction

One significant aspect of our future work involves adding more dynamic interactions to our visualisation, to allow more hands-on exploration, making the dashboard enjoyable for both hosts and travelers alike, enhancing their overall user experience.

B. Incorporating meaningful animation

In addition to increasing interactivity, we would like to incorporate meaningful animations into our visualisation dashboards. By doing so, we hope to make it more accessible and engaging with the users.

C. Acquiring time-series data

On top of that, we would like to acquire time series data on property occupancy rates throughout the year. This way, we will be able to track the seasonal changes within the homestay industry, helping the users to understand how different times of the year shape the market.

D. *Global expansion*

Last but not least, while we have been focusing on New York City due to time constraints, we are setting our sights on a global exploration of homestay pricing, investigating cities worldwide. We aim to uncover how amenities influence Airbnb pricing on a global scale, contributing to a broader understanding of the homestay landscape globally.

XI. CONCLUSION

This project embarked on a journey to unravel the intricacies of the rapidly evolving homestay industry, where the hosts and travelers are concerned about the pricing dynamics influenced by both tangible and intangible.

Our effort centered around using visualisations as a tool, and not merely for the presentation of data, but more as a dynamic interface fostering engagement. Both New York hosts and travelers can use the dashboard to gain meaningful insights.

For hosts, the dashboard reveals intriguing insights, challenging conventional assumptions. While there appears to be limited correlation between price and satisfaction, our findings highlight nuanced relationships. Surprisingly, the availability of subways and tourist spot density does not strongly correlate with average price. However, the inclusion of amenities such as TV significantly impacts listing prices, whereas the absence of free street parking doesn't necessarily lead to a lower listing price. These nuanced observations provide owners with actionable insights, allowing them to tailor their properties to meet guest expectations, striking a balance between meeting guest needs and adapting to the evolving landscape.

On the other hand, the guests can use the respective features to look at average room type and price, as well as having an interactive map to know more about the listings.

In conclusion, our project not only sheds light on the intricate dynamics of homestay pricing but also equips both hosts and travelers with a powerful tool for navigating this dynamic landscape. As we conclude this phase, the insights gleaned pave the way for further refinement, ensuring that our tool continues to evolve, providing even more tailored and relevant information for all stakeholders in the ever-evolving world of homestays.

XII. PROJECT SCHEDULE

For the project to run smoothly, we proposed and followed the following schedule:

TABLE IV. PROJECT SCHEDULE

Week & Date	**Task**
Wk 6 – 21 Sep	Project Proposal Draft
Wk 7 – 28 Sep	Project Proposal Submission
Wk 8 – 5 Oct	Exploratory Data Analysis
Wk 9 – 12 Oct	Create visualizations
Wk 10 -19 Oct	Combine and fine tune visualizations
Wk 11 – 26 Oct	Prepare project report and presentation
Wk 12 – 2 Nov	Project Presentation & Report Submission

REFERENCES

[1] J. Korn, "Airbnb to make prices clearer after user complaints of hidden fees | CNN Business," CNN, 07 Nov 2022. [Online]. Available: https://www.cnn.com/2022/11/07/tech/airbnb-cleaning-fees/index.html .

[2] M. McCluskey, "Why Airbnbs Are Getting Harder to Rent Out," Time, 02 Nov 2022. [Online]. Available: https://time.com/6223185/airbnbs-empty-short-term-rentals/. [Accessed 25 Sep 2023].

[3] 4. S. Deane, "2022 airbnb statistics: Usage, demographics, and revenue growth,," Stratos Jet Charters, Inc, 2022. [Online]. Available: https://www.stratosjets.com/blog/airbnb-statistics/.

[4] D. Shirshikov, "Airbnb Profit Margins 101: A Comprehensive Guide to Estimating Your Earnings," Awning, 22 July 2022. [Online]. Available: https://awning.com/post/profit-margin-airbnb.

[5] Mingming Cheng, Xin Jin, "What do Airbnb users care about? An analysis of online review comments," International Journal of Hospitality Management , 01 May 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0278431917307491 .

[6] I. Airbnb, "Inside Airbnb," Inside Airbnb, [Online]. Available: http://insideairbnb.com/ .

[7] Kaggle, "348 New York Tourist Locations," Kaggle, 21 Mar 2018. [Online]. Available: https://www.kaggle.com/datasets/anirudhmunnangi/348-new-york-tourist-locations .

[8] N. O. Data, "Subway Stations," NYC Open Data, 2023. [Online]. Available: https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49 .