SINGAPORE MANAGEMENT UNIVERSITY

IS630 – STATISTICAL THINKING FOR DATA SCIENCE

GROUP FINAL PROJECT REPORT

AIRBNB INSIGHTS

TEAM 5

Contents

1. Introduction

Airbnb has revolutionized and disrupted the traditional hospitality industry by allowing short to mid-term renting of properties to travellers. For many years, Airbnb had constantly edged out hotels by offering unique accommodations at a lower price. However, in today's market, finding a room on Airbnb at a cheaper rate than an equivalent 3-5-star hotel is getting increasingly challenging in part due to the many miscellaneous fees such as cleaning fees and parking fees, which many see as hidden fees. Therefore, to remain competitive it is imperative for Airbnb hosts to price their listings competitively to maximize occupancy rates, guest satisfaction and revenues while still taking into considerations the property type, location of the neighbourhood and amenities provided etc. From the travellers' perspective, understanding how the factors like locations, neighbourhood etc influence pricing is equally important. By gaining insights into the pricing dynamics relative to locations and amenities etc., travellers can reduce information asymmetry and make more informed data-driven decisions that ultimately enhance their experience.

2. Overall Concept

The analysis will examine Airbnb listings from three cities namely – Mexico, Paris and Tokyo to determine how various factors impact listing prices. Key questions include (but not limited to) – what factors most influence listing prices across different cities and neighbourhoods, which neighbourhoods offer the best value for money from the traveller's perspective, will be answered.

3. Data Sources

The datasets are obtained from insideairbnb.com. Specifically, the listing and review datasets for 3 cities. Key variables to be analysed are as follow:

Table 1

Variable	**Data Type**	**Description**
Price	Numerical	Price of Airbnb listing.
Neighbourhood	Categorical	Neighbourhood where listing is located.
Geographical location	Numerical	Longitudinal and Latitude of listing.
Amenities	Categorical	Type of amenities provided by the host.
Accommodation type	Categorical	Entire apartment/house, private rooms, shared rooms, hotel room.
Review score	Numerical	Review scores ranging from 0 to 5.
Open ended text reviews	Unstructured text data	Open-ended reviews left by guests.

4. Data Cleaning/ Preparation

a. Missing data. Rows with missing listing prices are removed.
b. Special characters. Rows with special characters due to the presence of non-English characters such as kanji are handled through normalization with the unicodedata.normalize function.
c. Currency exchanges. Converted all local currency to SGD.
d. Data transformation. Unstructured text data from the review datasets are transformed to categorical data types (i.e. Highly Positive, Positive, Neutral, Negative) via a two-step process. First step involves the transformation of unstructured text data to numerical data type via Natural Language Processing python library – Textblob. Next, the numerical polarity score obtained from Textblob are transformed to categorical data type via classification to be used in subsequent analysis.
e. Data extraction and grouping. A total number of 137 unique amenities were identified in the dataset. Amenities with similar meanings were grouped together (e.g. shampoo, bathtub, shower are all categorised as

‘Toiletries’) resulting in a total of 10 unique categories. These categories were then used as categorical data type for subsequent analysis.

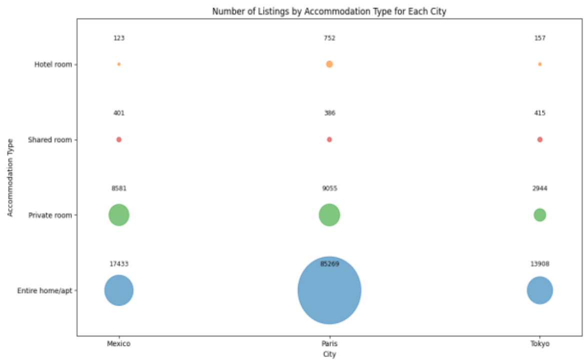
f. For the analysis of effect of presence of public transportation on mean listing price, a Python script was written to leverage the Overpass API to query the OpenStreetMap data for the following points of interest (POIs): metro stations, bus stops, and restaurants within a 500-meter radius from each listing. The API directly returns the number of these POIs. Metro stations data is also modified using pandas to set the count to 1 when the number of POI(s) was greater than or equal to 1. Additionally, the average number of POIs was calculated to understand the density of POIs across districts in different cities (see Appendix for details).

5. Descriptive Statistics

5.1 Distribution of Listings by Accommodation Types

Across cities, the most common accommodation type is Entire Home/Apartment followed by Private Room while Shared Rooms and Hotel Rooms are the least common (Fig 1). Paris has a notably high listing density at ~ 904.8 listings per km^2, followed by Tokyo (~26.3 listings per km^2) and Mexico (~17.8 listings per km^2). This indicates uniquely dense and competitive Airbnb market in Paris while Tokyo and Mexico with larger area have lower densities possibly due to stricter regulations and/or alternative accommodation options within the city.

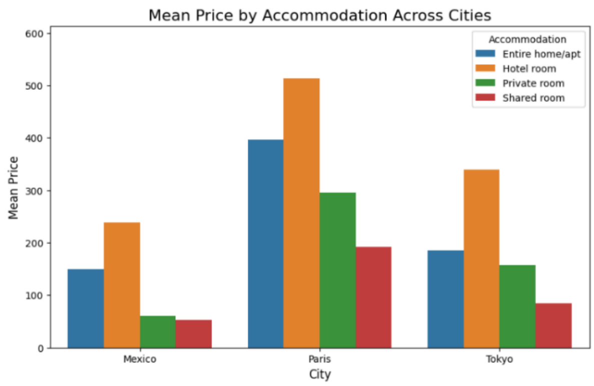
Fig 1



5.2 Distribution of listings by Accommodation Type with Price

Across all accommodation types, Mexico had the lowest listing prices followed by Tokyo and Paris (Fig 2). Airbnb guests should expect to pay more for certain accommodation type in a different city. For instance, with approx. S$200, you could rent an Entire home/Apartment in Tokyo, but the same amount would only afford a Shared Room in Paris. It’s also important to note that Airbnb has expanded its accommodation types to include ‘Hotel Room’ which is more like a traditional hotel and in this report, we will be taking this Hotel Room to be a proxy for traditional hotel.

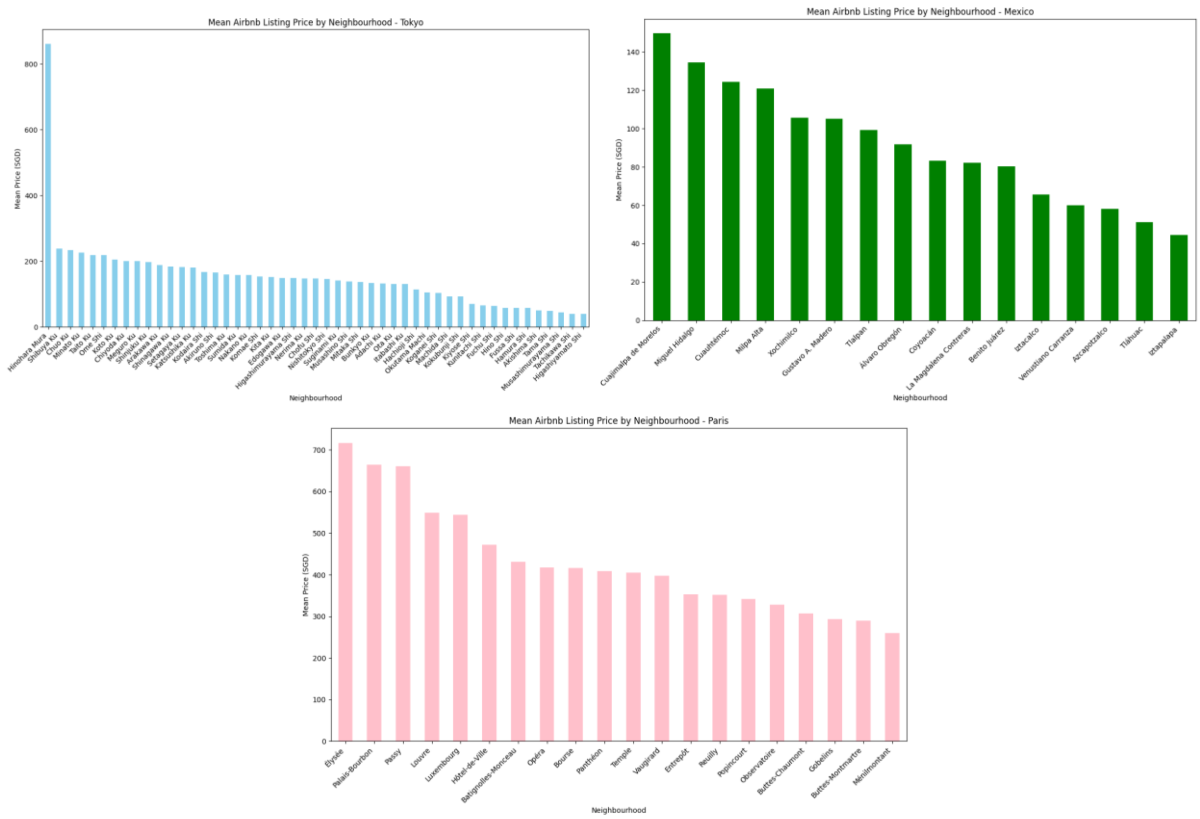
Fig 2



5.3 Distribution of listing by Neighbourhood with Price

Next, we looked at the mean price of listings vis a vis neighbourhood within the city (Fig 3). Unsurprisingly, certain neighbourhoods command a higher mean price compared to others. This suggests the need to explore possible factors such as proximity to city centre, availability of facilities like metro stations and restaurants, amenities provided and listing's review scores or perhaps a combination of these few factors, which will be investigated in subsequent sections.

Fig 3



5.4 Amenities Type by City

The bar chart displays the normalized frequency of each amenity category across the three cities. In general, amenities related to entertainment, housekeeping, comfort, safety, toilet equipping/ toiletries and kitchen appliances are the top 6 categories, suggesting that these types of amenities are most offered. The uniformity of the bars across the different cities suggest that these amenities are also seen as essential/standard features for listing regardless of the city. The lower normalised count for the remaining 4 categories suggests that accessibility features such as wheelchair access or step-free entrances are not widely offered which may limit options for guests with accessibility needs (Fig 4).

Fig 4

6. Inferential Statistics

6.1 Analysis of Mean Prices Across Accommodation Types and Across Cities Using Hypothesis Testing

As mentioned in section 5.2, there were observable differences in the mean prices across the accommodation types in different cities. Therefore, a two-way ANOVA test was conducted to determine if there is a statistically significant difference in mean prices across the accommodation types. A normality check was performed on the residuals to ensure that the ANOVA assumption of normality was not violated.

For Accommodation Types

H_0: µEntire Home/Apt = µPrivate Room = µShared Room = µHotel Room.

H_1: At least one accommodation type has significantly different mean price compared to others.

For Cities

H_0: µMexico = µParis = µTokyo.

H_1: At least one city has significantly different mean price compared to others.

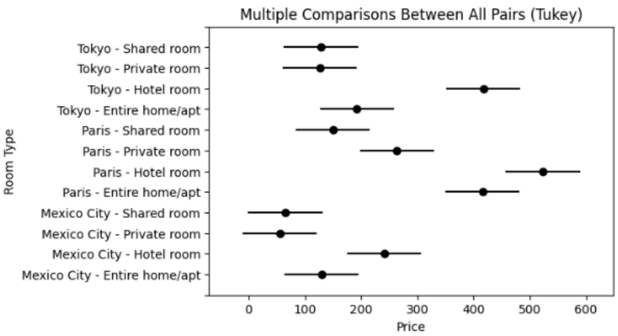
For Accommodation Types and Cities Interaction

H_0: There is no interaction between accommodation types and cities on the mean price.

H_1: There is interaction between accommodation types and cities on the mean price.

Source of Variation	df	Sum of Squares	Mean Squares (MS)	F-value	P-value
Accommodation Type	3	1295.870	431.957	891.448	**0.000**
City	2	74.878	37.439	1409.477	**0.000**
Accommodation Type and City (Interaction)	6	301.182	50.197	73.570	**4.33e-^{88}**
Residuals	4788	2087.222	0.437	NaN	NaN

Fig 5



The low p-values for both Accommodation Type (0.000) and City (0.000) indicate that these 2 factors individually have a statistical significance on the mean price. The low p-value of 4.33e-^{88} suggest that the interaction between Accommodation Type and City is statistically significant. Since they are significant factors, we conducted Tukey HSD to investigate which specific group differ significantly (Fig 5). Using 'Hotel Room' type on Airbnb as a proxy for traditional hotels, we find that both single room types (Shared and Private Room) are statistically lower in price by comparison, suggesting that Airbnb listings are in general still more affordable than traditional hotels.

6.2 Analysis of the Effect of Location (Proximity to City Centre) on Mean Price

To analyse the effect of location and its effect on mean price, we looked at the proximity of the listing to its respective city centre[1]. A sample of 1000 listings per city are randomly selected for the analysis. Haversine formula was used to calculate the distance between each listing to the city centre. Finally, an average distance to city centre for all listings from each neighbourhood is calculated to determine a ranking for the top 5 neighbourhoods. We took a step further to perform Simple Linear Regression to explain the effect of change in the independent variable (distance to city centre) on the dependent variable (price).

Table 2

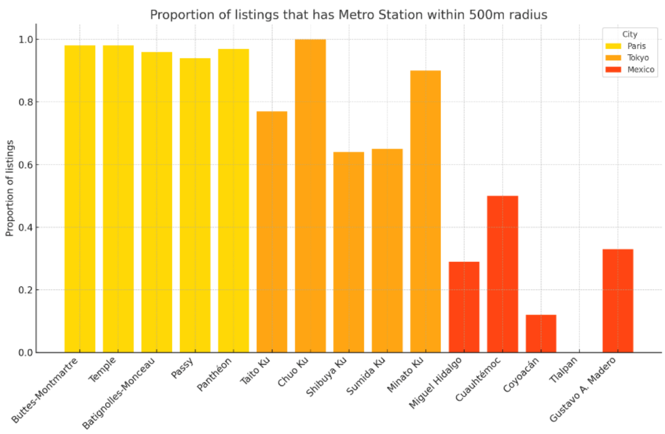
[1] Paris: 48.8534° N, 2.3488° E, Mexico: 35.0844° N, 106.6504° W, Tokyo: 35.667°N, 139.739°E

City	Neighbourhood	Price to Distance Ratio (Inverse)	P-value	Rank	Significant?	R^2
Paris	Bourse	934.57	0.514	1	N	NA
	Hôtel-de-Ville	757.57	0.002	2	Y	0.348
	Luxembourg	469.26	0.450	3	N	NA
	Louvre	400	0.000	4	Y	0.009
	Temple	245.09	0.000	5	Y	0.004
Tokyo	Fuchu Shi	128.20	0.014	1	Y	**0.734**
	Chuo Ku	51.81	$1.97e^{-06}$	2	Y	0.011
	Shibuya Ku	48.54	$1.04e^{-15}$	3	Y	0.064
	Kita Ku	20.533	$5.60e^{-05}$	4	Y	0.151
	Taito Ku	29.23	0.871	5	N	NA
Mexico	Cuajimalpa de Morelos	0.0865	0.191	1	N	NA
	Tlalpan	0.076	$1.46e^{-11}$	2	Y	0.080
	Cuauhtémoc	0.073	0.022	3	Y	0.027
	Iztapalapa	0.0624	$2.08e^{-05}$	4	Y	0.102
	Miguel Hidalgo	0.061	$6.82e^{-18}$	5	Y	0.005

Table 2 shows the top 5 neighbourhoods within cities ranked by the price to distance ratio. It also showed that certain neighbourhoods, despite being closer to the city centre, showed no statistical significance suggesting proximity to city centre may not be a statistical significance on listing price. Other factors[2] like amenities etc could be influencing the price and should be considered instead. The linear regression analysis of the 11 neighbourhoods with significant p-values showed that only Fuchu Shi in Tokyo had a high R^2 value of 0.734, indicating 73.4% of the variation in the mean price of Fuchu Shi can be attributed by the variation in distance to city centre. Apart from simple linear regression, correlation analysis could also be performed to obtain these results.

6.3 Analysis of the effect of presence of public transportation on mean price

Fig 6



To analyse the effect of the presence of public transportation (metro stations) have on mean listing price, five neighbourhoods each with sample size of 600 listings were randomly selected. For each of the 600 listings, we looked at the proportion of listings within the respective neighbourhoods that had metro stations within a 500m radius (Fig 6).

An upper tail test was conducted with the following hypothesis:

H_0: The mean price of listings near metro stations is less than or equal to the average price of listings not near metro stations.
H_1: The mean price of listings near metro stations is greater than the mean price of listings not near metro stations.

Table 3

Paris	Neighbourhood	Buttes-Montmartre	Temple	Batignolles-Monceau	Passy	Panthéon
	P-value	0.126	0.137	0.333	0.309	0.117

[2] Other models with non-linear relationships such polynomial relationship could also be explored.

Mexico	Neighbourhood	Miguel Hidalgo	Cuauhtémoc	Coyoacán	Tlalpan	Gustavo A. Madero
	P-value	0.290	0.158	**0.00374**	(see Note 1)	0.328
Tokyo	Neighbourhood	Taito Ku	Sumida Ku	Chuo Ku	Minato Ku	Shibuya Ku
	P-value	**0.00161**	0.381	(see Note 2)	**0.0211**	0.454

Note 1: There are zero metro stations within a 500m radius of all listings in Tlalpan.
Note 2: Every listing in Chuo Ku has at least 1 metro station.

The hypothesis test (Table 3) revealed that in Paris, all five neighbourhoods had p-values >0.05. Therefore, H_0 is not rejected, and it indicates there is no significant difference in mean price between listings near and not near metro stations. However, in Mexico and Tokyo, 1 and 2 neighbourhoods respectively, showed p-value <0.05, hence rejecting H_0, suggesting proximity to metro stations influence prices in those areas.

Expanding the analysis to include bus stops revealed a high density of bus stops across all five Parisian and Tokyo neighbourhoods. This distribution of bus stops across listings may explain the lack of significance observed in the influence of metro station proximity on price in these cities. In Mexico, however, this expanded definition resulted in additional 2 neighbourhoods showing statistically significant difference in mean price between listings near and not near public transportation (metro stations and bus stops) as shown in Table 4.

Table 4

Neighborhood	Miguel Hidalgo	Cuauhtémoc	Coyoacán	Tlalpan	GustavoA. Madero
P-value	0.133	**0.02**	**0.013**	-	**0.00037**

6.4 Analysis of effect of reviews and review scores on the mean price

Under this section, we analyse whether there are differences in the mean listing price based on the 1) open-ended review received by the listing and 2) review scores received by the listing.

Fig 7

H_0: Mean prices of listings across all polarity classes are the same.
H_1: Mean prices of listings of at least 1 polarity class is different.

For the first analysis, a random sample of 100 reviews were analysed for sentiment using TextBlob revealed high guest satisfaction across cities. In Paris, listings with 'Highly Positive' and 'Positive' sentiments tend to command higher prices, suggesting a link between positive sentiments and premium pricing. Conversely, in Mexico, positive sentiments are common for lower-priced listings indicating that even budget options receive favourable reviews. Overall, 'Highly Positive' and 'Positive' categories span a wide price range, implying positive reviews may support premium pricing (Fig 7). Given non-normally distributed residuals, a Kruskal-Wallis test was used to assess statistical difference in the mean prices between the different polarity classes.

Table 5

City	P-value from Kruskal-Wallis test	Conclusion
Paris	0.258	P-value is more than 0.05. Hence, H_0 is not rejected. There is insufficient evidence to suggest that there is a difference between the mean prices across different polarity classes.
Tokyo	0.240	
Mexico	0.102	

For the second analysis, we extracted the top 100 scorers and bottom 100 scorers that had at least 30 reviews. The scatter plot showed that all the top 100 scorers across cities have a perfect score of 5.0 while the bottom 100 scorers have scores ranging from 3.4 to 4.4 (Fig 8). Next, a two-population Z-test were conducted for each city with the following hypothesis:

H_0: µTop 100 Review Score = µBottom 100 Review Score
H_1: µTop 100 Review Score ≠ µBottom 100 Review Score

Fig 8

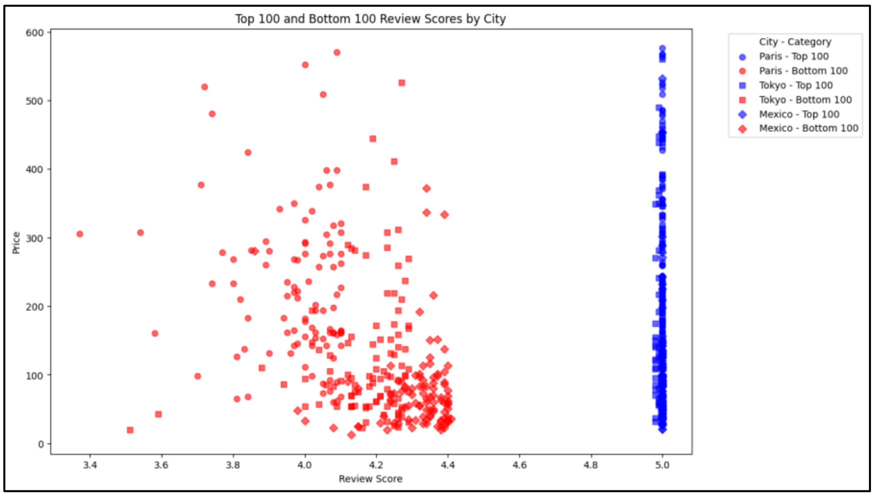


Table 6

City	P-value from Z-test	Conclusion
Paris	$7.948e^{-09}$	P-value is less than 0.05. Hence, H_0 is rejected. There is sufficient evidence to suggest that there is a difference between the mean prices of the top 100 scored listings and bottom 100 scored listings.
Tokyo	$1.064\ e^{-08}$	
Mexico	0.000	

With the rejection of H_0, we are interested to find out what are the differentiating factors between the top 100 and bottom 100 to explain the differences in mean prices which will be addressed in Section 6.5.

6.5 Amenities x Reviews Analysis

As mentioned in the previous section, there are significant difference between the mean listing price of the top 100 listings and the bottom 100 listings based on review scores. Hence, further analysis (two-way ANOVA) was conducted between the Basic and Premium amenities with top 100 and bottom 100 listings based on review scores. Similarly, a normality check was done on the residuals to ensure that the ANOVA assumption of normality is not violated.

For Amenity Type
H_0: µBasic Amenities Only= µ(Basic + Premium Amenities) = µ(No Basic Amenities) .
H_1: At least one amenities type has significant different mean price compared to others.

For Review Groups
H_0: µTop 100 listings = µBottom 100 listings
H_1: Significant difference in prices between top or bottom rated listings

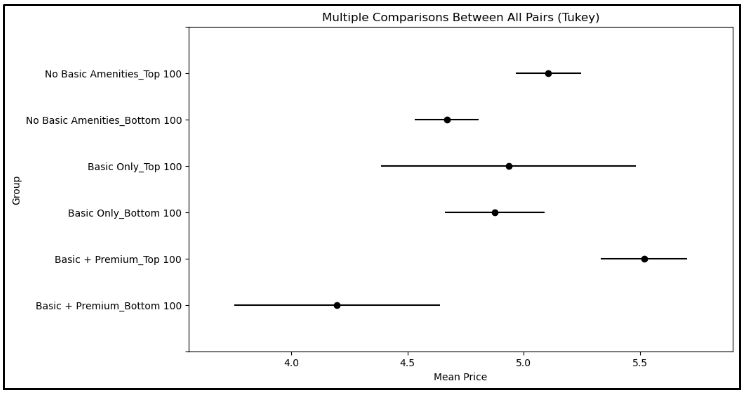
For Amenities Type and Review Interaction
H_0: There is no interaction effect between amenity groups and reviews on mean prices.
H_1: There is interaction effect between amenity groups and reviews on prices.

Table 7

	df	Sum Square	Mean Square	F	PR(>F)
Amenities Groups	2	14.855	7.427	10.355	0.000
Review Groups	1	32.804	32.804	45.733	0.000
Amenities and Review Groups	2	11.951	5.975	8.330	0.000
Residual	592	424.643	0.717	NaN	NaN

With all p-values <0.05, H_0 are all rejected, indicating significant difference in listing prices based on amenity type and rating. The interaction effect was also significant showing impact of amenities on prices may depend on review scores (Table 7). From the Tukey chart, for top 100 listings, mean prices differ significantly between listings with 'Basic Only' and 'Basic + Premium' amenities. In contrast, no significant price difference was found for amenities type in bottom 100 listing, indicating guests may be unwilling to pay for premium amenities when listing quality is low (Fig 9).

Fig 9



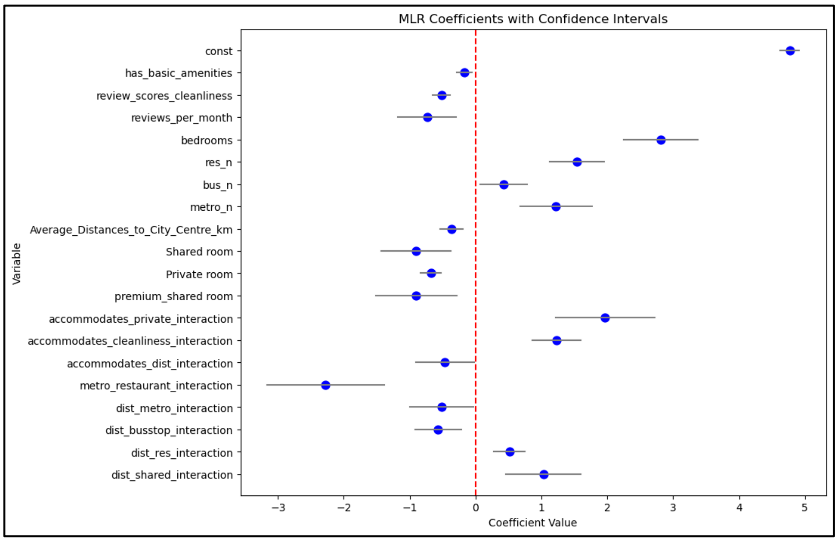
Note: The listing prices were normalised to allow comparison across different cities, reduce the influence of extreme values (outliers), and reduce skewness of the data.

6.6 Multiple Linear Regression (MLR) Analysis

We used a MLR model with log-transformed price (log_price_sgd) as the dependent variable to identify key drivers of Airbnb listing prices. Independent variables included listing attributes like amenities, review scores, room type, and proximity to local POIs. Interaction terms were added to capture how these factors interact, providing a more nuanced view of their combined effects on listing prices. The **adjusted R^2 of 0.605** indicates that **60.5%** of the variance in log-transformed prices is explained by the model, suggesting a **moderate fit**. Stepwise selection was adopted to achieve the best MLR model. A normality check was also performed to ensure the normality assumption was not violated. Similarly, Variance Inflation Factor (VIF) was also applied, based on VIF values of not more than 10, to avoid instances of multi-collinearity.

The multiple linear regression analysis (Fig 10) reveals several key factors influencing Airbnb listing prices. Capacity and privacy emerge as significant drivers, with listings that accommodate more guests or offer private rooms generally achieving higher prices due to their appeal to families and groups. Cleanliness and space are also valued; larger accommodations with higher cleanliness ratings can command higher prices, especially for group and family stays. Location factors, notably distance from the city centre, inversely impact pricing as listings farther away tend to be less in demand. Although proximity to metro stations and bus stops enhances listing appeal, this effect diminishes as distance from the city centre increases, suggesting that central locations are perceived as more accessible overall. Interaction terms highlight nuanced effects: for example, listings farther from the city centre benefit from nearby restaurants to counterbalance distance-related pricing drops, and shared rooms see varied impacts based on specific amenity-location combinations, reflecting a complex interplay between room type and location-based amenities. Overall, these insights guide pricing strategies that emphasize location, capacity, and convenience to maximize revenue potential.

Fig 10



7. Key Findings & Insights

The analysis identifies key factors impacting Airbnb pricing across cities, accommodation types, and amenities. City and accommodation type play a major role, with listings in Paris being generally more expensive and Mexico more affordable. Entire homes and hotel rooms typically command higher prices, especially in high-cost cities like Tokyo and Paris. Proximity to city centres generally exert an influence on the price across cities while accessibility to transportation option such as metro stations and bus stops does not display much statistical significance. Higher review scores and premium amenities, particularly in top-rated listings, correlate with elevated prices, underscoring the value perception tied to quality and added amenities. The multiple linear regression further reveals from a broader perspective, positive influences from larger accommodation size, transit proximity, and cleanliness, while factors like distance from the centre and shared/private room types reduce prices. Interactions between amenities and location highlight how specific features can help sustain pricing in less central areas.

8. Recommendation for Travellers

Overall, Private and Shared Room across all cities provide good value compared to Entire Home/Apartment and Hotel Room as guests are expected to pay a significantly lower price. Travellers with flexible destination (and barring other costs for flights, food etc) may consider Mexico as it has the lowest room costs across all accommodation types. Travellers who value proximity to city centres and metro stations could consider neighbourhoods in Table 3 and Table 4 respectively that are labelled not significant to leverage the proximity without paying a premium. Next, travellers who value review scores should expect to pay a premium for top scoring listing vs a listing that is not. However, it is also worthy to note that the general sentiment across all listings is skewed towards the positive side. Hence, there are still some value finds at a lower budget and travellers seeking premium features can consider choosing lower rated listings to enjoy premium amenities without the price tag.

9. Recommendations for Hosts

Hosts in central locations can maximize revenue by emphasizing privacy and convenience offered by private rooms or entire homes. For listings further from the centre, focusing on nearby amenities like public transportation and dining options may enhance appeal. Cleanliness of the listing may not be a strong price driver on its own, however, coupled with a larger room type show positive influence on the prices, as travellers in bigger groups such as families are likely to view cleanliness as essential. Host that can accommodate big groups should emphasize on cleanliness as one of their selling points. Additionally, offering basic amenities with selected premium features, such as accessibility options, could address a gap in the market and attract a broader range of travellers at competitive prices.

10. Limitations

This analysis faces several limitations. Firstly, the sentiment analysis tool, TextBlob, may lack the sophistication needed to capture nuanced sentiments in reviews, potentially impacting accuracy in polarity classifications. Additionally, the absence of seasonal data restricts our ability to assess how pricing might fluctuate across different times of the year, which is essential in markets with varying tourist demand. Similarly, without booking rate data, we cannot further evaluate how a listing's popularity influences pricing dynamics, limiting insights into demand elasticity for Airbnb listings.

11. Future Works

This EDA only scratches the surface and further studies could be conducted to address the limitations through various approaches. Comparing Airbnb demand with that of traditional hotels could highlight unique competitive advantages or preferences in the lodging market. Conducting demographic analysis would help identify key customer groups more inclined toward Airbnb, aiding targeted marketing efforts. A seasonal occupancy study could also be valuable in understanding demand variations and optimizing pricing strategies accordingly. Furthermore, refining the model by focusing on significant variables and simplifying interactions could improve interpretability and reduce the risk of overfitting, while still explaining a substantial portion of price variance.

12. Conclusion

In conclusion, there are a multitude of variables that shapes the listing price. This study offers an analysis of Airbnb listing prices across multiple cities, examining the impact of various factors such as accommodation type, location proximity, amenities, and review characteristics. Through descriptive, inferential, and multiple linear regression analyses, we uncovered significant insights into how these factors influence pricing. Key findings indicate that location near central areas, availability of metro and bus access, and certain room types elevate prices, while factors like distance from the city centre and shared accommodation types lead to lower pricing. Additionally, high cleanliness ratings and essential amenities bolster pricing potential, highlighting the value travellers place on quality and convenience. While challenges such as limited sentiment analysis accuracy and lack of seasonal data suggest areas for improvement, this analysis lays a strong foundation for data-driven recommendations. For travellers, value identification strategies help optimize booking choices based on budget and preference, while hosts can leverage insights to refine pricing and enhance listing appeal. Future research aimed at addressing current limitations could further strengthen Airbnb’s market strategies and pricing models, offering continued value to both travellers and hosts.

Appendix

Additional Analysis for Amenities

Boxplot below shows the types of amenities available in each listing are observed to have difference in the mean listing price (Fig A1). A two-way ANOVA test and Tukey was conducted to statistically validate whether there are significant differences in prices across the amenity groups (Fig A2).

Fig A1

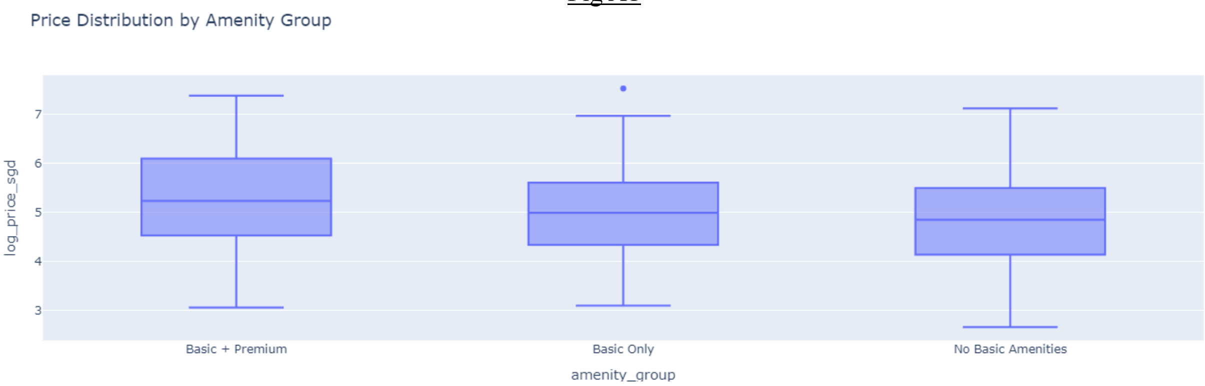
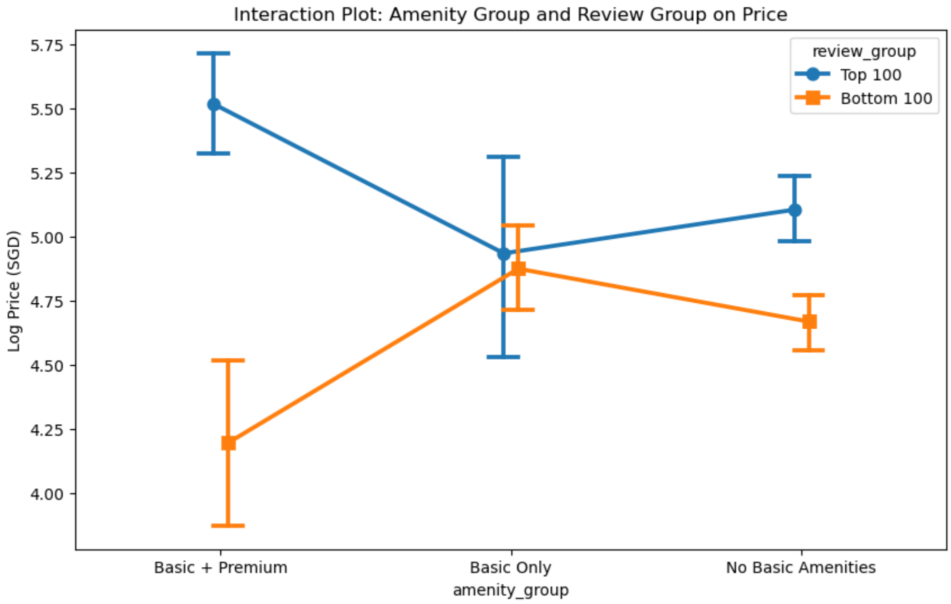


Fig A2



Top 100 Listings:

For listings in the Top 100 review group (blue line), the price is highest when both Basic + Premium amenities are offered, and it gradually decreases as the level of amenities reduces. The presence of more amenities is positively correlated with higher prices in well-reviewed listings, indicating that guests are willing to pay a premium for full amenities in highly-rated listings.

Bottom 100 Listings:

For listings in the Bottom 100 review group (orange line), the pattern is almost reversed: listings with Basic + Premium amenities have the lowest prices on average, while those with Basic Only or No Basic Amenities tend to have slightly higher prices. This suggests that, for poorly reviewed listings, the presence of premium amenities does not justify higher prices. It might even imply that guests are less willing to pay for premium amenities if the overall quality (as indicated by reviews) is low.

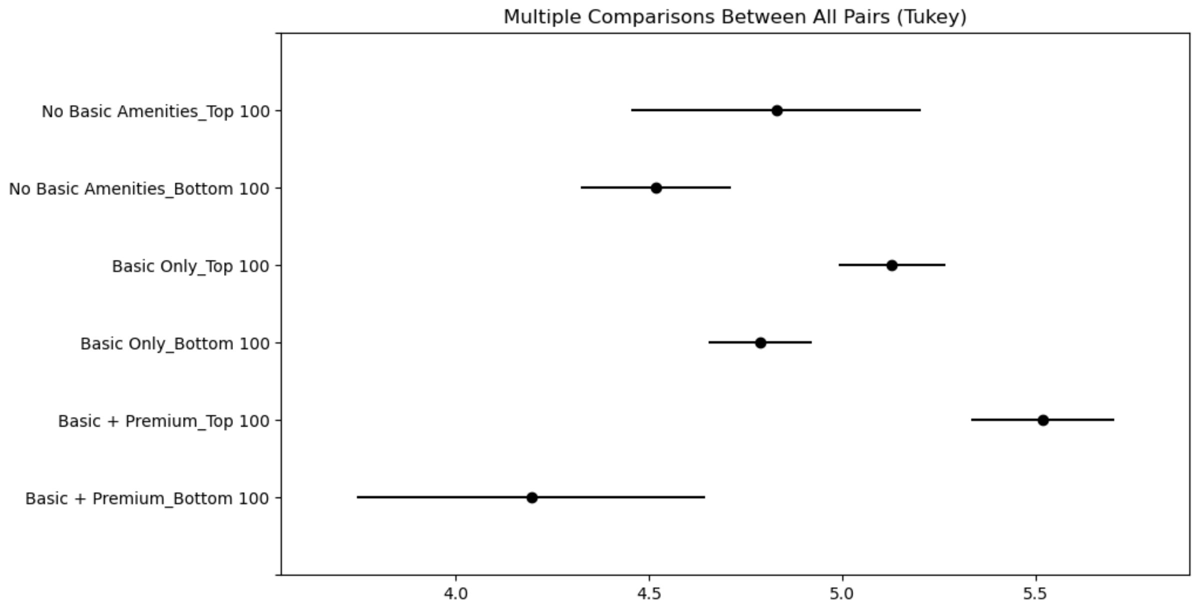
Interaction Effect:

There is a clear interaction between review group and amenity group. The lines cross, indicating that the impact of amenities on price depends on the review group. For highly-rated listings, having full amenities boosts prices, while for poorly rated listings, the same amenity boost does not lead to higher pricing and might even correlate with lower prices.

Summary:

High review scores amplify the value of amenities, allowing hosts to command higher prices for listings that provide both basic and premium amenities. For low review scores, the presence of additional amenities does not increase the perceived value, and listings with premium amenities may even be priced lower than expected. This could indicate that guests prioritize reviews over amenities when considering lower-rated listings. The analysis was followed with a Tukey's Range Test to gain deeper understanding of the findings (Fig A3).

Fig A3



The key insights are summarized as follows.

a. Top 100 vs. Bottom 100 Listings: Listings in the Top 100 review group consistently have higher mean prices across all amenity categories compared to the Bottom 100 review group. This is evident from the separation between the groups for each amenity category. The largest price difference is seen in the Basic + Premium amenities group, where the Top 100 listings are priced significantly higher than the Bottom 100.

b. Effect of Amenity Levels: Within each review group (Top 100 and Bottom 100), prices increase as the level of amenities increases. Basic + Premium amenities in the Top 100 group has the highest mean price, while No Basic Amenities in the Bottom 100 group has the lowest mean price. This indicates that the combination of high review scores and comprehensive amenities (both basic and premium) commands the highest prices.

c. Statistical Significance: Other than the listings with No Basic Amenities, the confidence intervals are generally distinct for Top and Bottom groups, suggesting significant differences between them for each amenity level.

Additional Analysis of Proximity of Listing to Public Transportation

The following shows the scatter plot (Fig A4, A5 and A6) returned by the API showing the occurrence of public transportation systems i.e. metro stations and bus stops.

Fig A4

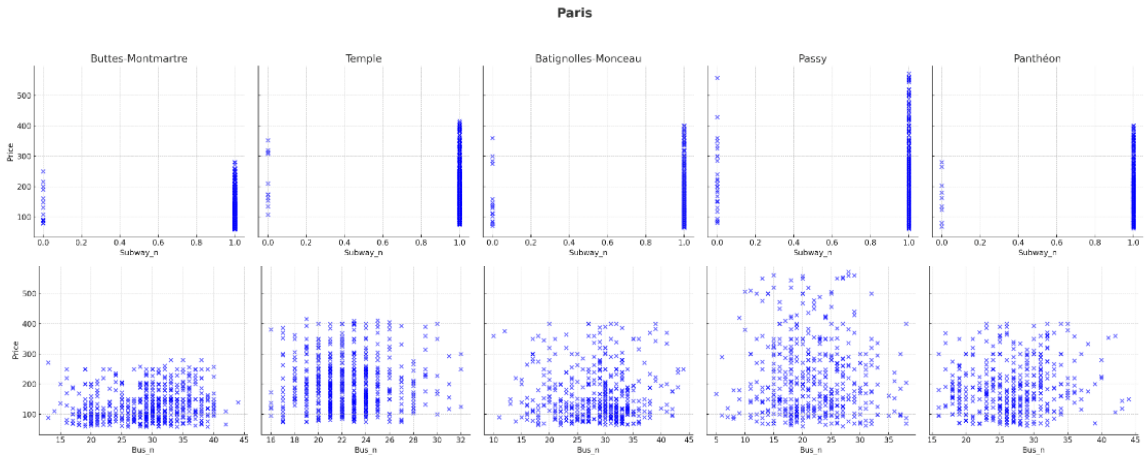


Fig A5

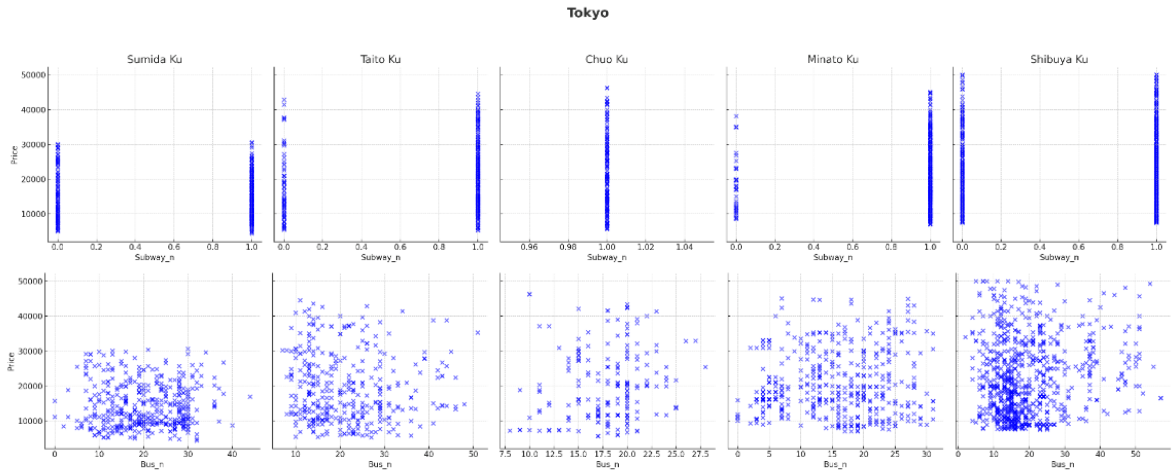
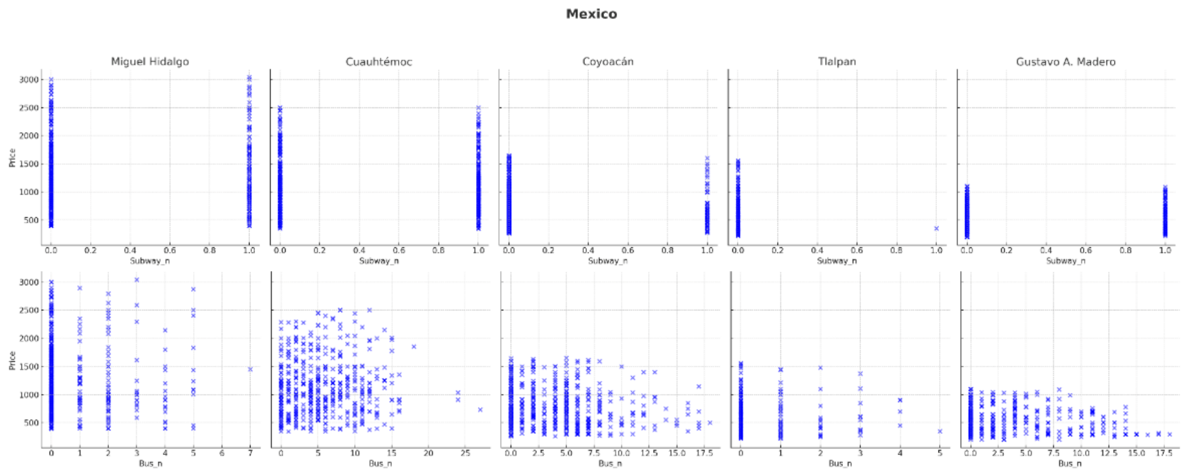


Fig A6



Additional Analysis of The Effect of Presence of Restaurant on Mean Price

We analysed the effect of the presence of restaurant within a 500m radius of the listing in the same five neighbourhood and its mean price. An upper-tail hypothesis test is conducted with the following hypothesis:

> H_0: The mean price of listings with restaurants in its vicinity is less than or equal to the mean price of listings without restaurants in its vicinity.
> H_1: The mean price of listings with restaurants in its vicinity is more than the mean price of listings without restaurants in its vicinity.

Table A1.

Paris	Neighbourhood	Buttes-Montmartre	Temple	Batignolles-Monceau	Passy	Panthéon
	P-value	**0.0101**	**0.0171**	0.208	0.16	**0.003**
Mexico	Neighbourhood	Miguel Hidalgo	Cuauhtémoc	Coyoacán	Tlalpan	Gustavo A. Madero
	P-value	0.163	**$3.969e^{-06}$**	**$3.345e^{-08}$**	0.191	**0.013**
Tokyo	Neighbourhood	Taito Ku	Sumida Ku	Chuo Ku	Minato Ku	Shibuya Ku
	P-value	**0.0011**	0.275	**0.0006**	**$4.663e^{-06}$**	**$9.4303e^{-14}$**

The hypothesis testing (Table A1) showed that each city has at least two out of five neighbourhood has p-value < 0.05 hence rejecting H_0. There is sufficient evidence in those neighbourhoods, listings that have restaurant in its vicinity has a higher mean price than listings without.