



airbnb

New York Listings' Data Analysis

- Pooja Agarwal

Problem Statement

The dataset and problem statement is provided by **International Institute of Information Technology, Bangalore (IIIT-B)** as part of *Advanced Certificate Program in Data Science*.

Suppose that you are working as a data analyst at Airbnb. Now that the covid-19 related restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change. The leaders at Airbnb want to gain valuable insights from the dataset containing various Airbnb listings in New York to boost revenue.

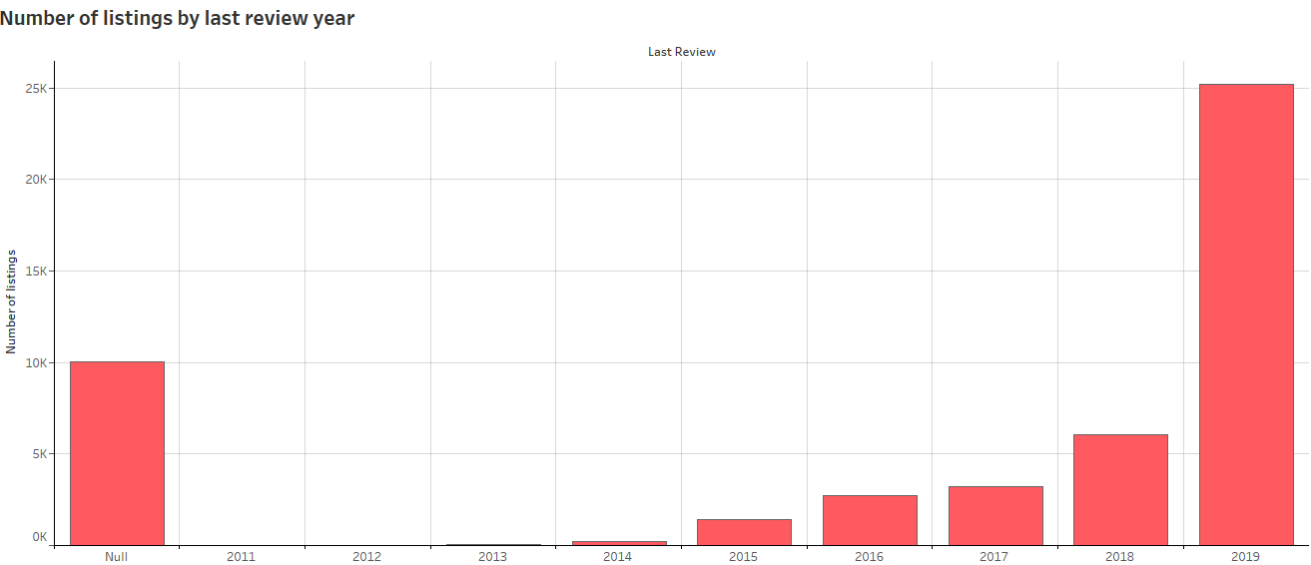
Analyse the dataset and prepare a report with the next best steps that Airbnb needs to take as a business. The report will be submitted to Heads of Acquisitions, Operations, and Marketing NYC who look after acquisition of the best properties, price negotiation, negotiating and providing support to improve the services the properties offer, and marketing and advertisements.

Dataset Description

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Note: The price column contains the price/night.

Listings By Last Review Year



(Fig. 1)

Insights

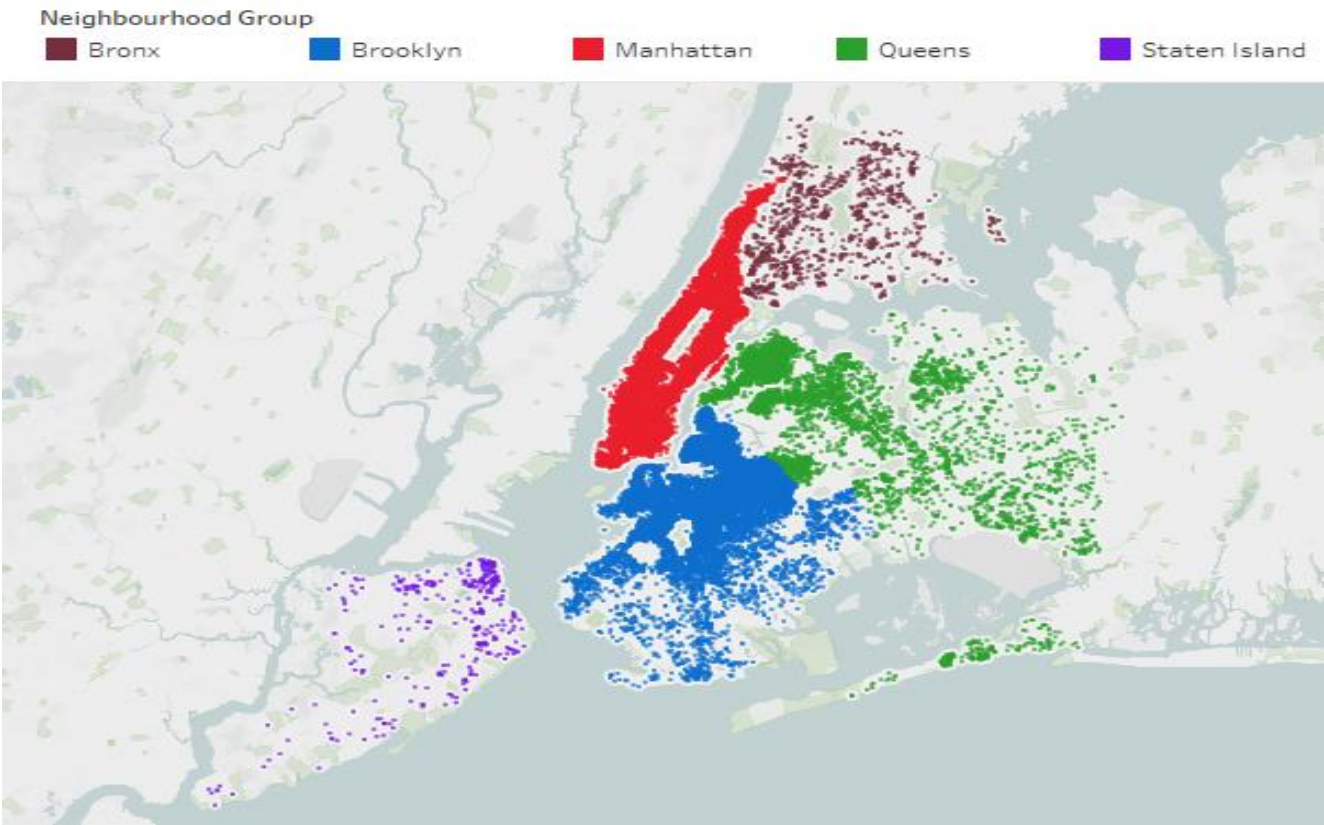
Among our dataset of 48,870 New York listings, only 51.5% received their most recent review in 2019, pre-pandemic. Around 27.8% of listings had their last review before 2019, and notably approximately 20.5% of listings lacked any reviews.

Recommendations

- Investigate non-reviewed listings:** Prioritize understanding why nearly half of the listings (48.5%) had no 2019 reviews or none at all. Reach out to hosts of these listings and initiate host surveys, focus groups, and feedback forms to gather insights.
- Support and incentivize hosts for review engagement:** Conduct host workshops, guide on improving guest experience, provide review request templates, and introduce automated review collection reminders. Additionally, offer incentives to hosts who actively seek reviews.
- Encourage and incentivize guests to leave reviews:** Implement automated thank you messages with review reminders after the guest’s stay. Incentivize guests by offering rewards, discounts, or exclusive perks to share their experiences.

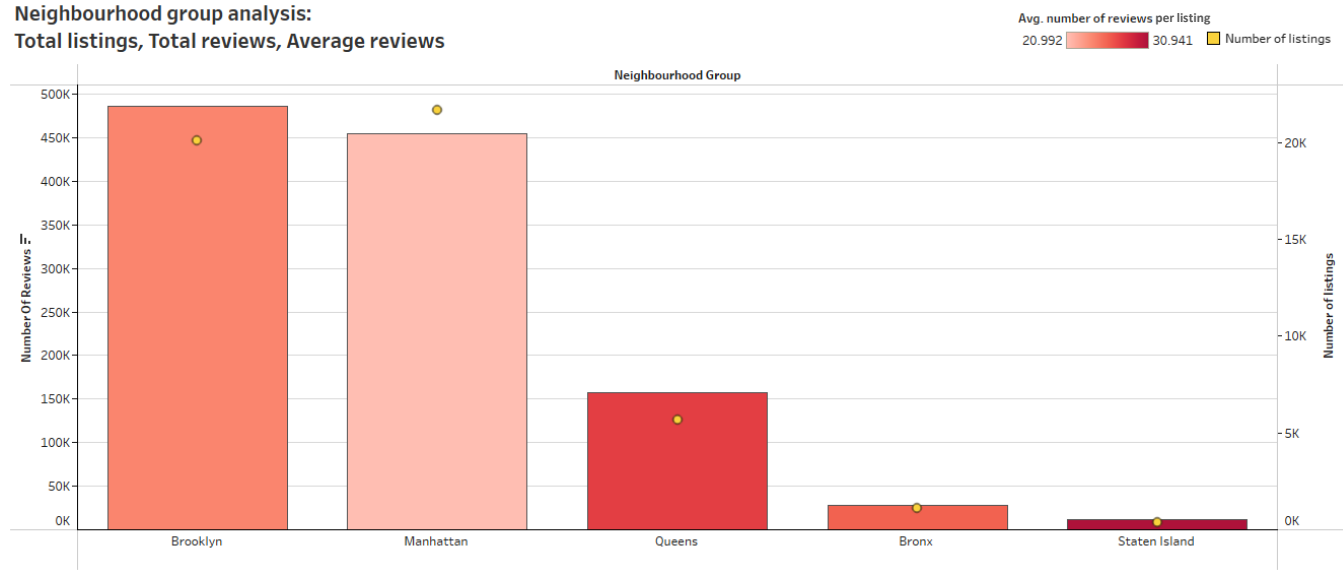
Popular Neighbourhood Groups

Listings' location map



(Fig. 2)

Neighbourhood group analysis:
Total listings, Total reviews, Average reviews



(Fig. 3)

Insights

Manhattan and Brooklyn can be considered the most popular neighbourhood groups because they have the highest number of listings and reviews. Meanwhile, Queens stands out for its potential growth, boasting high average reviews per listing, substantial listings and reviews.

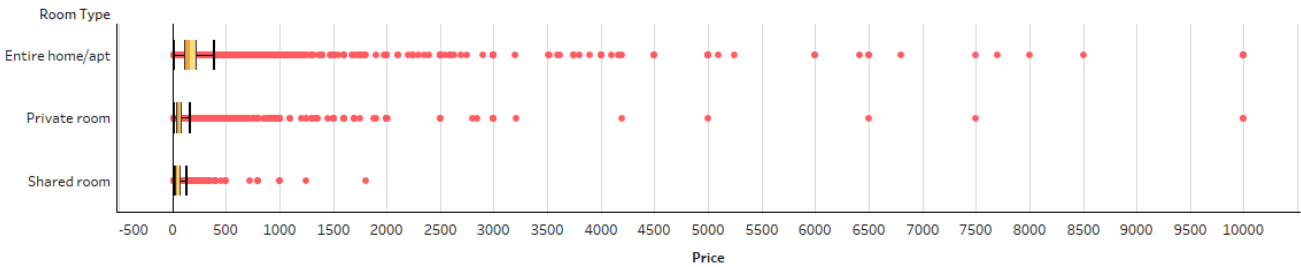
- **Manhattan** has the largest share of listings at 44.3% and yet, maintains the lowest average reviews per listing (around 21), indicating a compelling need for improving customer experience.
- **Brooklyn**, accounting for 41.1% of total listings, leads with 42.7% of total reviews. Also, its average reviews per listing stand at 24
- **Queens** representing only 11.6% of total listings, and shines with its substantial total reviews and notably higher average reviews per listing (around 28). It signifies high demand, robust guest satisfaction and review engagement, and immense scope for further growth and expansion.
- The **Bronx and Staten Island**, respectively represent 2.2% and 0.8% of total listings, 2.5% and 1.01% of total reviews; and maintain an average of 26 and 31 reviews per listing.

Recommendations

- **Improve customer experience and review engagement in Manhattan:** Support hosts in Manhattan to deliver quality services and effectively engage in review collection by providing hosts with hospitality training, offering customer service resources, implementing quality assurance and host recognition programs.
- **Strategic expansion and targeted marketing in Queens:** Channel resources into expanding Airbnb's presence in Queens and Staten Island, given their remarkable potential for growth. Implement precise marketing strategies tailored to these neighbourhoods. Encourage hosts to maintain exceptional guest experiences and active review collection.
- **Targeted marketing and review engagement in Brooklyn, Bronx, and Staten Island:** In Brooklyn and the Bronx, deploy targeted marketing strategies to attract guests actively. Support and encourage hosts to engage proactively with guests and cultivate reviews.

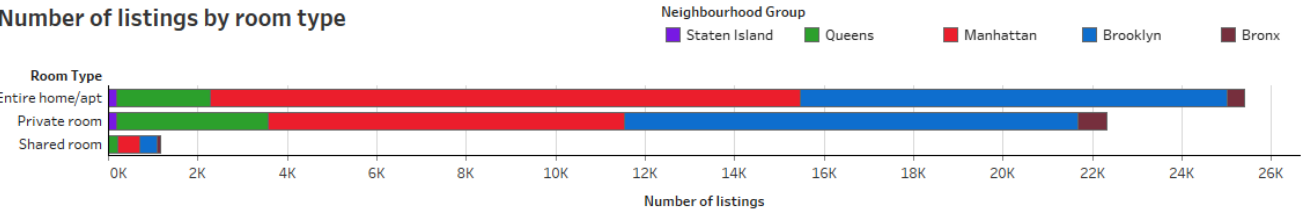
Room Type

Listings' price distribution by room type



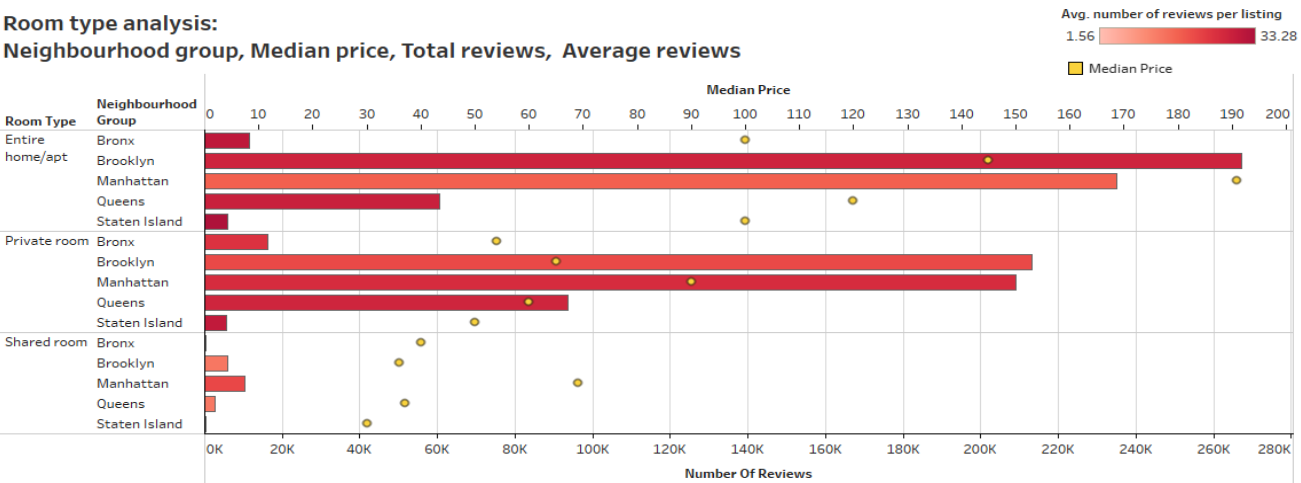
(Fig. 4)

Number of listings by room type



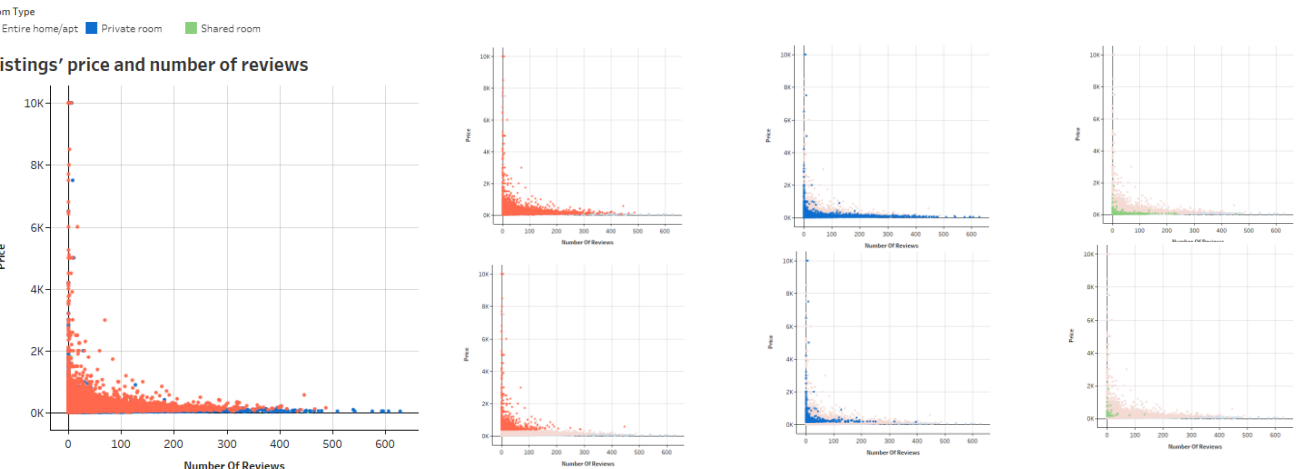
(Fig. 5)

Room type analysis:
Neighbourhood group, Median price, Total reviews, Average reviews



(Fig. 6)

Listings' price and number of reviews



(Fig. 7)

Insights

- The most common listing type is 'Entire home/apt' with 25,398 listings, followed closely by 'Private rooms' at 22,315 listings. 'Shared rooms' constitute a smaller segment, totalling 1,157 listings, and with Staten Island having just 9 listings in this category. *(Fig. 5)*
- Notably, except for Manhattan, all other four neighbourhood groups predominantly offer 'Private rooms'. *(Fig. 5)*
However, room type 'Entire home/apt' has higher total reviews in Manhattan, Brooklyn, and Staten Island; while 'Private room' has higher total reviews in Queens and the Bronx. *(Fig. 6)*
- Entire home/apt in Manhattan, Private rooms in Brooklyn, and Shared rooms in all neighbourhood groups have significantly lower average reviews per listing. *(Fig. 6)*
- In comparison to other neighbourhoods, 'Shared rooms' in Manhattan receive higher total reviews and higher average reviews per listing, indicating a high demand and potential for growth and expansion. *(Fig. 6)*
- The price distribution of listings is skewed. The interquartile range and median price for each room type are as follows *(Fig. 4)*:

Room Type	Interquartile Range	Median Price
Entire home/apt	120 to 229	160
Private rooms	50 to 95	70
Shared Rooms	33 to 75	45

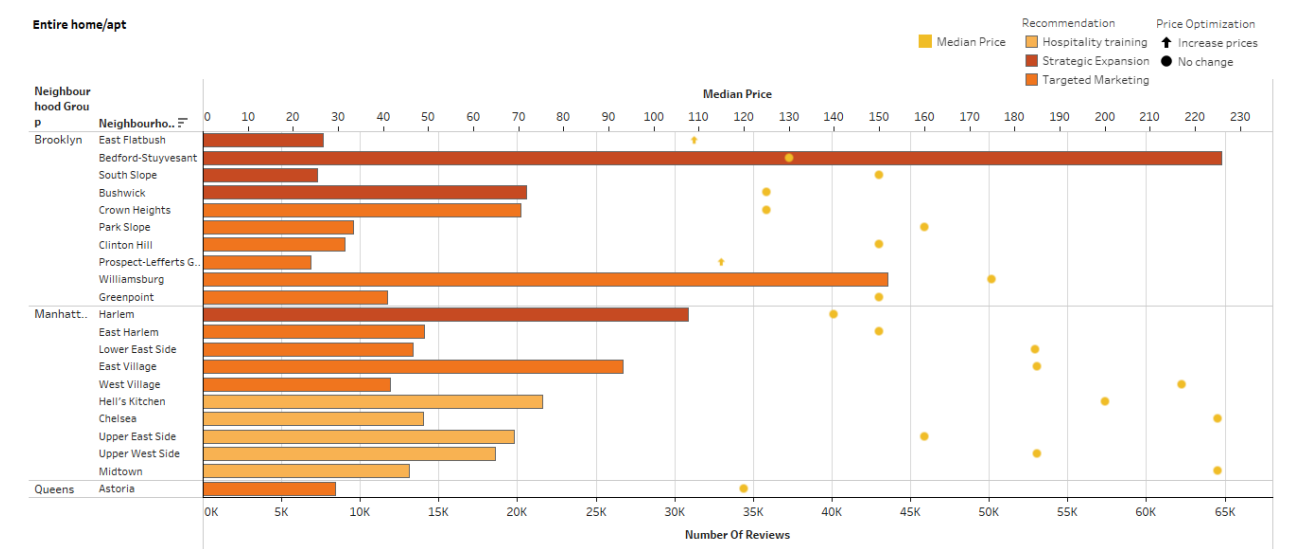
- In Manhattan, the median prices for all room types *(Fig. 6)* consistently exceed the median prices calculated for those room types across all neighbourhoods *(Fig. 4)*, highlighting Manhattan's generally higher room prices.
- There is a trend that as listings' prices increase, the number of reviews tends to decrease . This trend becomes particularly pronounced for listings with prices exceeding the upper whisker value in the 'price distribution' for the respective room type. *(Fig. 4 & 7)*.
(The upper whisker value for 'Entire home/apt' is 392, for private rooms is 160, and for shared rooms is 130).
Most listings with over 300 reviews are private rooms. *(Fig. 7)*

Recommendations

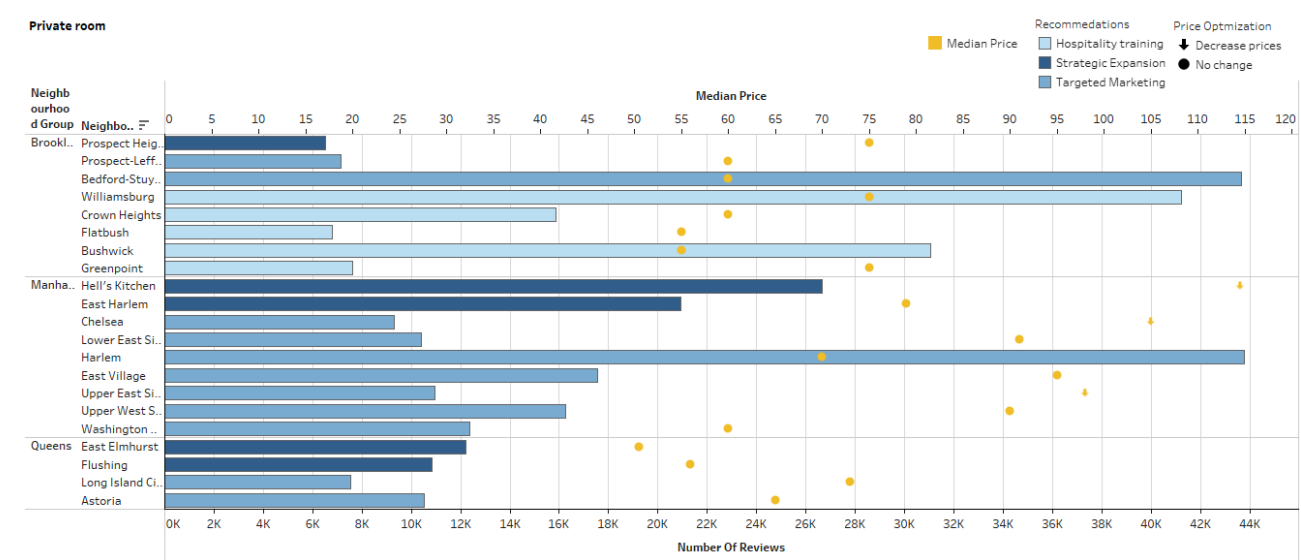
- **Focus on quality service and active review engagement:** Support hosts of 'Entire home/apt' in Manhattan, 'Private rooms' in Brooklyn, and 'Shared rooms' in all neighbourhood groups with delivering consistent and quality services, maintaining cleanliness standards, and actively soliciting reviews. This can be achieved through hospitality training, resource provision for review collection (e.g., review request templates), regular inspections, quality assurance programs, host recognition initiatives, and the introduction of automated review reminders for both hosts and guests.
- **Targeted marketing and incentives for choosing 'Shared rooms':** Launch strategic marketing campaigns to promote and position 'Shared rooms' as affordable, value-for-money options that offer opportunities for social interactions, catering to budget-conscious travellers. Incentivize guests by offering rewards, discounts, or exclusive perks for choosing 'Shared rooms' and sharing their experiences via reviews.
- **Support hosts in Manhattan with price optimization:** Help hosts in Manhattan (specially of 'Entire home/apt' given its low average reviews per listing) optimize their listing prices, striking the right balance between being competitive and reasonable. This involves introducing clear pricing guidelines and advanced dynamic pricing tools, enabling hosts to adjust rates in real-time based on market demand, seasonality, and local events.
- **Strategic expansion and targeted marketing of 'Shared rooms' in Manhattan:** Allocate resources to expand the availability of 'Shared rooms' in Manhattan and implement marketing campaigns to promote it. Given the high number of 'Entire home/apt' listings in Manhattan, which typically command high median prices but receive lower average reviews per listing, consider encouraging a portion of these 'Entire home/apt' hosts to convert their listings to 'Shared rooms'. Provide guidance, support, and incentives to facilitate this transition.
- **Strategic expansion and targeted marketing for Queens:** Channel resources into expanding the availability of 'Entire home/apt' and 'Private rooms' in Queens. These room types not only have a substantial number of listings and total reviews but also maintain high average number of reviews per listing, suggesting high demand, and growth potential. Implement precise marketing strategies tailored to this location and room types. Encourage hosts to maintain exceptional guest experiences and active review collection.
- **Quality assurance and incentives for premium accommodations:** Implement quality assurance programs to ensure that hosts offering premium accommodations (listings priced higher than \$392 for 'Entire home/apt', \$160 for 'Private rooms', and \$130 for 'Shared rooms') deliver exceptional guest experiences, personalized attention, and excellent service. Also encourage these hosts to proactively cultivate reviews. Additionally, introduce incentives or promotions for guests who book and provide reviews for higher-priced listings.

Popular Neighbourhoods^(*)

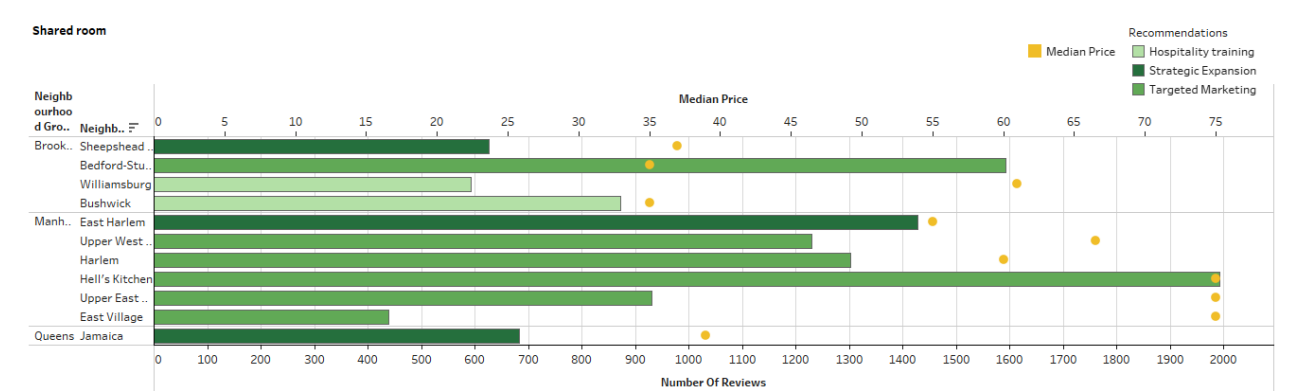
Median Price, Total reviews, Recommendations based on average reviews



(Fig. 8)



(Fig. 9)



(Fig. 10)

Recommendations^(*2)



Entire home/apt



Private rooms



Shared rooms



Price optimization – Increase prices

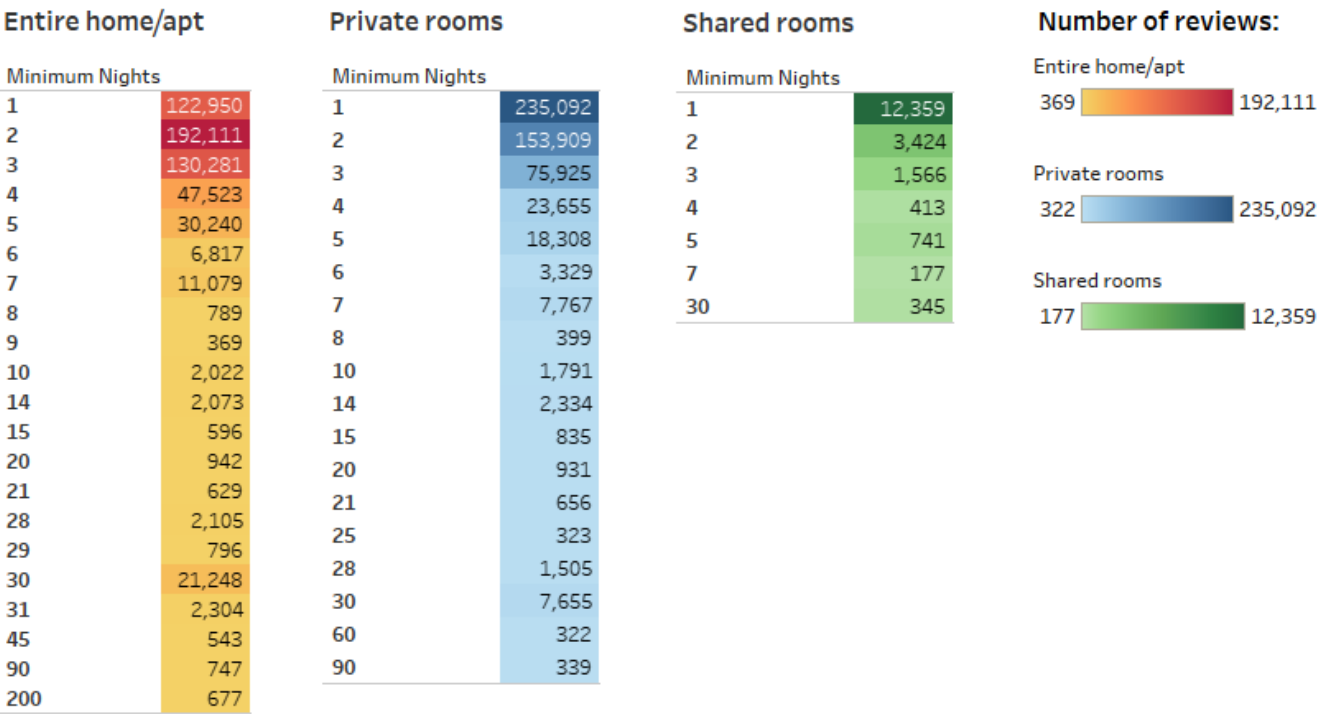


Price optimization – Decrease prices

Neighborhood Groups	Strategic expansion, and targeted marketing	Targeted marketing, and review collection support	Hospitality training, quality control measures, review collection support
Brooklyn	<div>Bushwick</div> <div>South Slope</div> <div>Bedford-Stuyvesant</div> <div>East Flatbush</div> <div>Prospect Heights</div> <div>Sheepshead Bay</div>	<div>Greenpoint</div> <div>Williamsburg</div> <div>Prospect-Lefferts Gardens</div> <div>Clinton Hill</div> <div>Park Slope</div> <div>Crown Heights</div> <div>Bedford-Stuyvesant</div> <div>Prospect-Lefferts Gardens</div> <div>Bedford-Stuyvesant</div>	<div>Bushwick</div> <div>Flatbush</div> <div>Greenpoint</div> <div>Crown Heights</div> <div>Williamsburg</div> <div>Bushwick</div> <div>Williamsburg</div>
Manhattan	<div>Harlem</div> <div>East Harlem</div> <div>Hell's Kitchen</div> <div>East Harlem</div>	<div>East Harlem</div> <div>Lower East Side</div> <div>East Village</div> <div>West Village</div> <div>Hell's Kitchen</div> <div>Lower East Side</div> <div>Washington Heights</div> <div>Harlem</div> <div>East Village</div> <div>Upper East Side</div> <div>Upper West Side</div> <div>Chelsea</div> <div>East Village</div> <div>Hell's Kitchen</div> <div>Harlem</div> <div>Upper East Side</div> <div>Upper West Side</div>	<div>Midtown</div> <div>Upper East Side</div> <div>Upper West Side</div> <div>Chelsea</div>
Queens	<div>East Elmhurst</div> <div>Flushing</div> <div>Jamaica</div>	<div>Astoria</div> <div>Astoria</div> <div>Long Island City</div>	

Optimal Minimum Nights As Per Room Type^(*3)

Optimal minimum nights for:



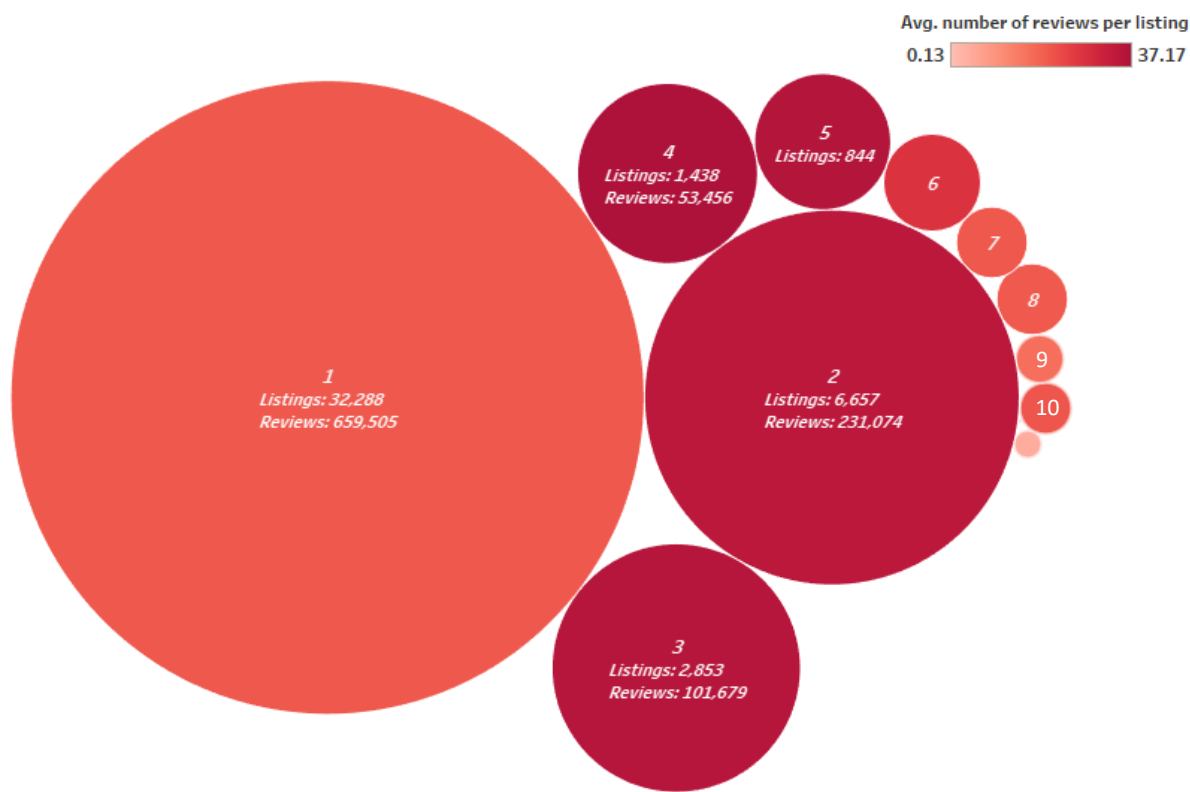
(Fig. 11)

Recommendations

Encourage hosts to align minimum nights for their listings with the optimal minimum nights determined for the respective room type.

Common Host Listing Count^(*4)

Analysis of common host listing count: Number of listings, Total reviews, Average reviews



(Fig. 12)

Insights

- **Host listings count:** Hosts with a single listing (host listing count: 1) are the most common, representing a significant portion of the Airbnb host community, with 32,288 listings in this category. Hosts with 2 to 10 listings also contribute substantially to the platform, with thousands of listings in each category/host listing count.
- **Review trends:** Listings with host listing count '1' have lower average reviews per listing, while those with 2-5 listings achieve the highest average reviews. Beyond five listings (host listing count: 6 -10), there is evidence of diminishing returns in terms of average reviews per listing.
- **Optimal Portfolio Size:** Hosts with two to five listings (host listing count: 2-5) tend to maintain a balance between the number of listings and guest reviews. This demonstrates their ability to manage multiple properties effectively while maintaining guest satisfaction, positioning them uniquely for increased profitability.

Recommendations

- **For single listing hosts (host listing count: 1):** Offer support and training for enhanced hosting skills, specially to new hosts. Provide educational resources and training on creating appealing listings, offering excellent hospitality, maintaining cleanliness standards, managing guest expectations, and actively soliciting reviews. Encourage portfolio expansion, aiming for 2-5 listings.
- **For hosts with 2-5 listings (host listing count: 2-5):** Encourage to continue to prioritize quality and guest satisfaction. Implement efficiency tools and automation. Utilize features like automated messaging, booking management systems, and dynamic pricing tools.
- **For hosts with more than 5 listings (host listing count: 6-10):** Provide advanced tools and resources to efficiently handle large number of listings. This may include advanced management software, specialized marketing assistance, guest relations training programs, dedicated support hotline, advanced pricing tools, tiered rewards system, exclusive webinars for experienced hosts. Also Implement strict quality control measures and regular audits to ensure quality standards are met.

Additionally, encourage hosts with multiple listings to provide rewards, perks, or exclusive offers to customers for booking multiple listings from their portfolio.

***Note:**

1. Defining ‘Popular Neighbourhoods’ for each room type:

To define ‘Popular Neighbourhoods’ for each room type (entire home/apt, private rooms, shared rooms), the following procedure was followed:

- Selected a room type and compiled the neighbourhoods with listings for that specific room type.
- Calculated the total 'number/sum of reviews' for the room type within these neighbourhoods and determined the 90th percentile value from the 'number of reviews' distribution.
- Neighbourhoods with reviews equal to or greater than the 90th percentile were marked as 'Popular Neighbourhoods' for that room type.

Repeated the above process independently for each room type, resulting in a distinct list of ‘Popular Neighbourhoods’ for each room type.

In total, 29 neighbourhoods are identified as ‘Popular Neighbourhoods’ from our dataset encompassing 221 neighbourhoods.

Room Type	90th percentile value for 'number of reviews' distribution
Entire home/apt	6738.6
Private rooms	6533
Shared rooms	421.6

2. Recommendations for ‘Popular Neighbourhoods’:

A. Determine the ‘average reviews per listing distribution’ for popular neighbourhoods of the room type.

Room Type	Average reviews per listing distribution	
	25 th percentile value	75 th percentile value
Entire home/apt	18.2	29.4
Private rooms	21	34.3
Shared rooms	18.7	31.7

If ‘average reviews per listing’ of the popular neighbourhood for the room type	Recommendation
< 25 th percentile value of ‘average reviews per listing distribution’	Hospitality training, quality control measures, review collection support
>= 25 th percentile and < 75 th percentile value of ‘average reviews per listing distribution’	Targeted marketing, and review collection support
>= 75 th percentile value of ‘average reviews per listing distribution’	Strategic expansion, and targeted marketing

B. Consider ‘price distribution of room type’ from *Fig. 4*.

If ‘Median price’ of popular neighbourhood for the room type	Price optimization recommendation for popular neighbourhood for a room type
< 25 th percentile price value of respective room type from <i>Fig. 4</i>	Increase Prices
> 75 th percentile price value of respective room type from <i>Fig. 4</i>	Decrease Prices

3. Determining ‘optimal minimum nights’ for each room type:

(The column ‘minimum_nights’ in the dataset has 101 unique values ranging from 1 to 365)

First, listings are filtered based on a particular room type. Then, for each unique value in the ‘minimum_nights’ column pertaining to the specific room type, the corresponding ‘number of reviews’ is calculated.

The 75th percentile value for the ‘number of reviews distribution’ is as follows:

Room Type	Entire home/apt	Private rooms	Shared rooms
75th percentile value for the ‘number of reviews distribution’	361	322	109

Thus, minimum nights with number of reviews equal to or greater than the respective 75th percentile value are considered as optimal minimum nights for that specific room type.

4. Defining ‘Common Host Listing Count’

(The column ‘calculated_host_listings_count’ in the dataset has 47 unique values ranging from 1 to 327).

The ‘Common Host Listing Count’ threshold is established by the ‘number of occurrences’ for each unique value in ‘calculated_host_listings_count’ column in the dataset. It is found that the 75th percentile value in the ‘number of occurrences distribution’ is 210.

As a result, ‘calculated host listings count’ with an occurrence equal to or greater than 210 is considered as ‘Common Host Listing Count’.

Data Cleaning

1. Columns:

A. reviews_per_month

There is ambiguity in the calculation and interpretation of the 'reviews_per_month' column. For instance, the column reports a 'reviews_per_month' value of 4.44 for a listing with 53 reviews and 2.12 for a listing with 260 reviews, making it challenging to discern a clear and meaningful pattern. Thus, to simplify the dataset and improve its interpretability, the 'reviews_per_month' column is removed.

B. availability_365

There is lack of clarity regarding the meaning and calculation method of the 'availability_365' column, along with a significant number of listings, i.e., 17,533 listings, displaying '0' for availability_365. The prevalence of '0' values raises questions about its accuracy and relevance. Therefore, to enhance the dataset's transparency and simplify data analysis, the column 'availability_365' is removed.

2. Understanding Null Values:

Within the dataset, null values are observed in the 'name,' 'host_name,' and 'last_review' columns. It is noteworthy that in cases where 'last_review' contains null values, the simultaneous presence of 'number_of_reviews' as '0' indicates that these listings did not receive any reviews. Therefore, the null values in the 'last_review' column are aligned with listings that had no reviews.

3. Rows:

A. Price

Within the dataset, a total of 11 listings exhibit a 'price' of '0.' Geographically, these listings are distributed as follows: 1 in the Bronx, 1 in Manhattan, and 9 in Brooklyn. Given the unusual nature of these price values, it is presumed that these listings contain invalid price information. The combined total number of reviews for these listings is 377. Thus, rows with details of these listings are removed from the dataset as they may not accurately represent typical listings or meaningful data for analysis.

B. minimum_nights

Among the listings, 14 were identified with 'minimum_nights' exceeding 365, exhibiting values such as 366, 370, 400, 480, 500, 999, 1000, and 1250. These listings accounted for a mere 74 reviews in total. Thus, rows with details of these listings are removed from the dataset.

THANK YOU