**New York Airbnb Listings**

**Data Source: -**

The data for this analysis was obtained from the NYC Airbnb Overview dataset available on Kaggle. The dataset contains 48,895 rows (observations) and 16 columns (features). Each row represents a unique Airbnb listing in New York City. The columns provide details about the listing such as its name, host information, location coordinates, room type, pricing, review activity, and host listing count. This comprehensive dataset allows us to explore various aspects of the New York Airbnb rental market.

## **Column Dictionary for NYC Airbnb Overview Dataset**

## This table provides a detailed description of each column in the NYC Airbnb Overview dataset obtained from Kaggle:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Data Type** |
| id | Unique identifier for each Airbnb listing. | Integer |
| name | Title or name of the Airbnb listing. | String |
| host\_id | Unique identifier for the host who manages the listing. | Integer |
| host\_name | Name of the host who manages the listing. | String |
| neighbourhood\_group | Broad geographic area of the listing (e.g., Manhattan, Brooklyn). | String |
| neighbourhood | Specific neighborhood where the listing is located. | String |
| latitude | Geographic latitude coordinate of the listing. | Float |
| longitude | Geographic longitude coordinate of the listing. | Float |
| room\_type | Type of accommodation offered (e.g., Entire Apartment, Private Room). | String |
| price | Nightly price of the Airbnb listing. | Integer |
| minimum\_nights | Minimum number of nights required for a stay. | Integer |
| number\_of\_reviews | Total number of reviews received for the listing. | Integer |
| last\_review | Date of the most recent review received for the listing (if any). | Datetime |
| reviews\_per\_month | Average number of reviews received per month for the listing (if any). | Float |
| calculated\_host\_listings\_count | Total number of Airbnb listings managed by the host. | Integer |
| availability\_365 | Number of days the listing was available throughout the year (out of 365). | Integer |

**Data Preprocessing :-**

The Data has no duplicate values but few columns had missing values which are Name, Host\_name, last\_review and reviews\_per\_month

The name and the host name NA values have been replaced with Not Available whereas reviews\_per\_month NA values have been replaced with 0 and last review column missing values have left because. Listings with Na values in last\_review likely represent new Airbnb listings that haven't received any reviews yet. Replacing these Na values with a placeholder date would introduce artificial data and skew analysis of review trends or host performance.

The fact that last\_review and reviews\_per\_month Na values coincide strengthens the assumption of new listings. Since a new listing wouldn't have reviews, it wouldn't have a last review date or reviews per month.

reviews\_per\_month Na values with 0 makes sense because a new listing with no reviews would naturally have an average of 0 reviews per month. This imputation maintains data integrity and avoids introducing bias.

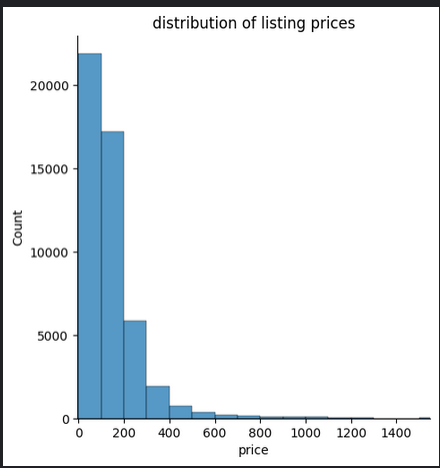
**Statistical Analyisis :-**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| price | 48895.0 | 152.72068 | 240.154170 | 0.0 | 69.00 | 106.00 | 175.00 | 10000.0 |
| minimum\_nights | 48895.0 | 7.029962 | 20.510550 | 1.0 | 1.00 | 3.00 | 5.00 | 1250.0 |
| number\_of\_reviews | 48895.0 | 23.274466 | 44.550582 | 0.0 | 1.00 | 5.00 | 24.00 | 629.0 |
| reviews\_per\_month | 48895.0 | 1.090910 | 1.597283 | 0.0 | 0.04 | 0.37 | 1.58 | 58.5 |
| calculated\_host\_listings\_count | 48895.0 | 7.143982 | 32.952519 | 1.0 | 1.00 | 1.00 | 2.00 | 327.0 |
| availability\_365 | 48895.0 | 112.78132 | 131.62228 | 0.0 | 0.00 | 45.00 | 227.00 | 365.0 |

* **Price:**
  + The average price is ~$153 per night, but there's a significant standard deviation (~$240), indicating a wide range of prices.
  + The minimum price is $0, while the maximum is a very high $10,000 suggesting that might need further investigation.
  + The median price is ~$106, providing a more representative value of the central tendency.
* **Minimum Nights:**
  + The average minimum stay requirement is about 7 nights.
  + There's a minimum of 1 night required by some listings, while others have a much higher minimum.
* **Number of Reviews:**
  + The average listing has around 23 reviews, but the standard deviation is high (44.55), suggesting some listings have many reviews while others have very few (minimum is 0).
* **Reviews per Month:**
  + I replaced missing values with 0, resulting in an average of 1.37 reviews per month. This might be slightly skewed if some new listings haven't had a chance to accumulate reviews yet.
  + The median (0.72) might be a more indicative value for established listings.
* **Calculated Host Listings Count:**
  + The average host manages around 7 listings on Airbnb.
  + There's a high standard deviation (32.95), indicating some hosts have many listings while others have just one.
* **Availability:**
  + Listings are available, on average, for about 113 days throughout the year. This could be due to seasonal rentals or hosts taking their listings offline for personal use.
  + The minimum availability is 0 days, and the maximum is the full year (365 days).

**Vizualization :-**

**1)**



In the above plot the x-axis represents the price, and the y-axis shows the count.

Here are some specific observations from the graph:

* The most frequent price range falls between $200 and $400 per night.
* There are a significant number of listings priced below $200 per night.
* The distribution appears skewed to the right, as there are more expensive listings than very cheap listings

2)

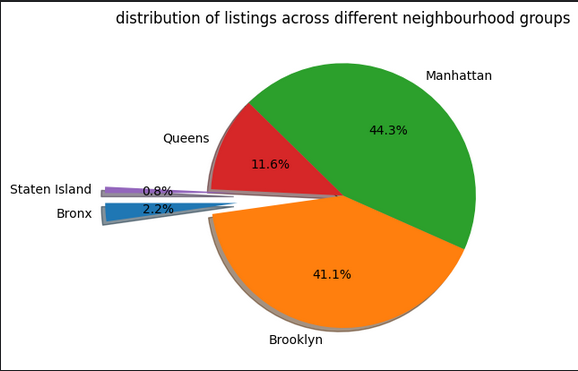


the distribution of room types by neighborhood in a chart.The x-axis shows the neighborhood group (Brooklyn, Manhattan, Queens, Staten Island, Bronx), and the y-axis shows the count. There are three bars for each neighborhood group, colored blue, orange, and green, which likely represent the different room types (Entire Home/Apt, Private Room, Shared Room).

Here are some observations from the graph:

* **Accommodation Mix by Borough:** The distribution of room types appears to vary across the boroughs. we can make some general observations. For instance, Manhattan seems to have a higher proportion of entire homes/apartments compared to brooklyn, which might have more private rooms than manhattan.
* **Limited Shared Rooms:** It appears that shared rooms are not as prevalent as entire homes/apartments and private rooms across all boroughs, based on the height of the green bars.

**3)**



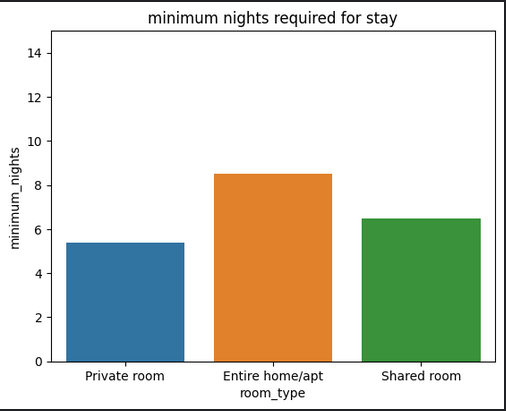
The image depicts a pie chart showing the distribution of Airbnb listings across different borough groups in New York City.

* **Borough Breakdown:**
  + Manhattan (41.1%) - This borough has the highest percentage of listings, represented by the green slice.
  + Brooklyn (44.3%) - This borough has the second-highest percentage, represented by the orange slice.
  + Queens (11.6%) - Represented by the red slice.
  + Staten Island (0.8%) - Represented by the purple slice (smallest).
  + Bronx (2.2%) - Represented by the light blue slice.

**Insights from Distribution:**

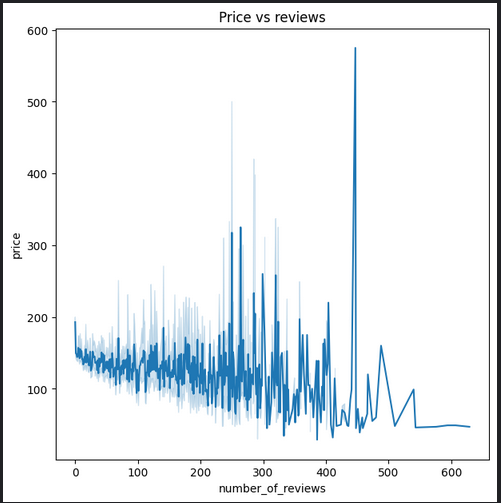
* **Manhattan and Brooklyn Dominance:** The pie chart reveals that Manhattan and Brooklyn have the most Airbnb listings, collectively accounting for over 85% of the listings represented in this data. This could be due to several factors, such as:
  + Tourist attractions and business districts concentrated in these boroughs.
  + A higher concentration of suitable rental properties (apartments, lofts) compared to other boroughs.
* **Lower Availability in Other Boroughs:** The significantly lower percentages for Queens, Staten Island, and the Bronx suggest a lower availability of Airbnb listings in these areas. This doesn't necessarily mean there are no listings, but the concentration is considerably lower compared to Manhattan and Brooklyn.

**4)**



* The graph compares the minimum nights required for stay across three different room types: Private room, Entire home/apt, and Shared room.
* **Higher Minimums for Private Rooms:** The minimum nights required for private rooms appear to be the highest, followed by entire homes/apartments and then shared rooms. This suggests that hosts might set stricter minimum stays for private rooms, possibly due to wanting a longer booking duration or targeting weekend stays.
* **Shared Rooms Most Flexible:** Shared rooms appear to have the most flexible booking options, with the lowest minimum nights required. This could be because shared rooms are generally priced lower and cater to budget-conscious travelers who might be open to shorter stays.

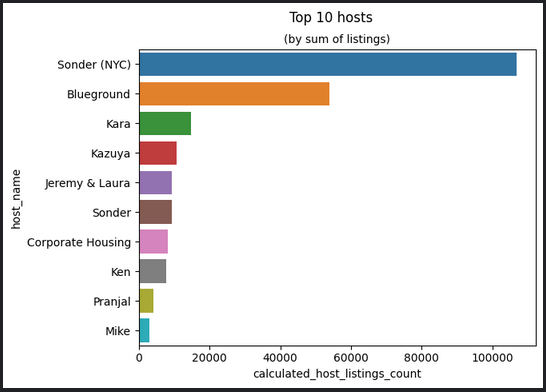
5)



line chart showing a slight downward slope with price on the y-axis and number of reviews on the x-axis suggests a **weak negative correlation** between price and review count

* **Price Doesn't Guarantee Reviews:** Listings with higher prices don't necessarily translate to a significant increase in the number of reviews. Guests might be leaving reviews for various reasons beyond just price (e.g., exceptional amenities, unique location, or negative experiences)
* **Value Focus:** There's a possibility that budget-conscious travelers (who tend to book lower-priced listings) might be more motivated to leave reviews, either positive (highlighting good value) or negative (warning others about potential drawbacks).

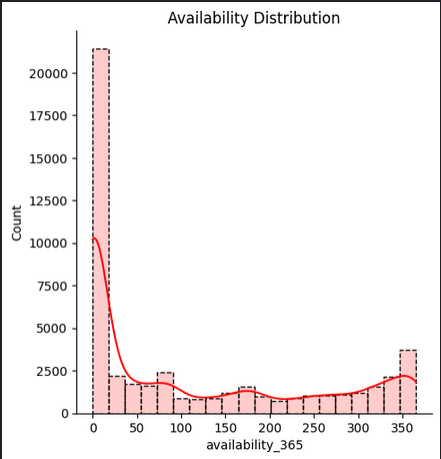
6)



The bar chart depicts the top 10 hosts based on their listing the insights that can be derived are:-

* **Most Prolific Hosts:** The y-axis shows the host names, and the x-axis represents the total number of listings each host has. Sonder (NYC) appears to be the host with the most listings, followed by Blueground, Kara, Kazuya, and Jeremy & Laura. These top hosts likely manage a significant number of Airbnb properties in the data set.
* **Concentration of Listings:** The varying lengths of the bars on the x-axis indicate an uneven distribution of listings among these top hosts. Sonder (NYC) has a considerably higher number of listings compared to the other hosts displayed.

**7)**



The Distribution chart shows the availablity of listings in the New york with availablity\_365 on the x-axis ranging from 0 to 350 and on the y-axis is the count Based on the chart it shows that the more number of listing are available for less number of days and vice versa This is because

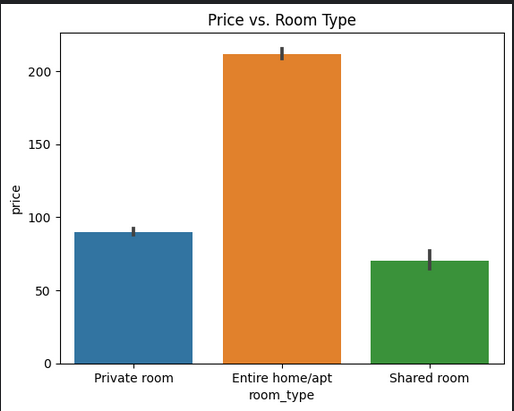
**Shorter Stays in Demand:**

* Travelers seeking short getaways or weekend trips might be more prevalent, leading to a higher number of listings catering to those needs.

**Booking Strategies:**

* Some hosts might prefer to keep their listings flexible by offering shorter availabilities initially. They can then adjust the availability to longer stays closer to the date if no shorter bookings come in. This strategy allows them to capture last-minute bookings or accommodate travelers with flexible plans.

**8)**



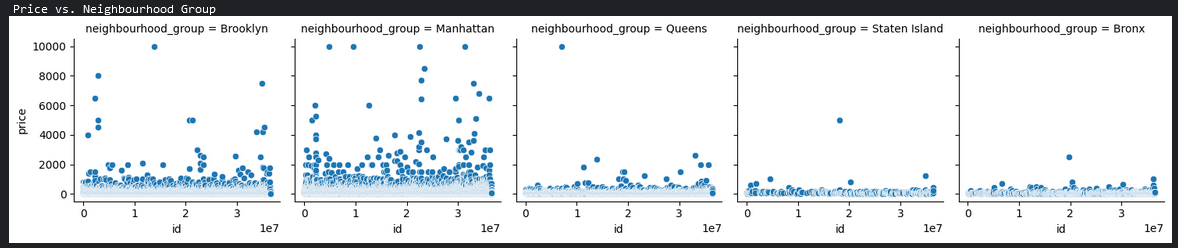
**Price Trends by Room Type:**

* Generally, entire homes/apartments appear to be the most expensive room type, followed by private rooms (blue) and shared rooms (green). This aligns with expectations as entire homes offer more space and privacy, justifying a higher price.
* **Price Variation Within Room Types:** There's also variation in price within each room type. Some entire homes/apartments might be priced similarly to high-end private rooms, while other private rooms could be priced closer to shared rooms. Factors like location, amenities, and size can influence price within each room type.

**Possible Interpretations:**

* **Space and Privacy:** The price difference reflects the value placed on space and privacy. Entire homes/apartments offer the most of both, followed by private rooms, and then shared rooms.
* **Amenities:** Entire homes/apartments might often come with additional amenities like kitchens or laundry facilities, contributing to a higher price point.
* **Location:** Location can significantly affect price. Entire homes/apartments in prime locations might command premium prices compared to private or shared rooms in less central areas.

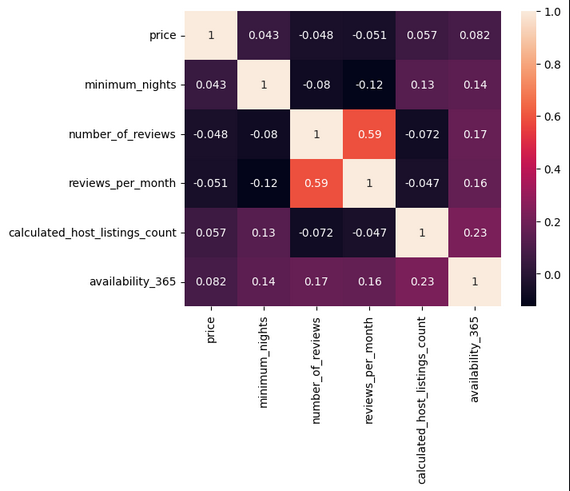
**9)**



The above plot is a Facet Grid which shows Price range of each listing in each neighbourhood the Key insights which can be noticed from this image are :-

* The Brooklyn and Manhattan are pricier than compared to other neighbourhoods this may be due to their location and accessibility to nearby places.
* Other neighbourhood such as queens, Bronx and Staten island have low price listings These boroughs might have a more residential feel with fewer high-end amenities or trendy shops. While they offer their own unique character and charm, they might not hold the same appeal for renters willing to pay top dollar for a vibrant lifestyle.
* Manhattan and Brooklyn These boroughs boast a higher concentration of high-rise apartments, studios, and smaller living spaces. This type of housing caters to a broader range of renters, including young professionals or those seeking a more urban lifestyle, potentially driving up average prices.
* Queens, Bronx, and Staten Island These boroughs might have a larger share of single-family homes, multi-family homes, and larger apartments. While these options can be attractive to families or those seeking more space for their money, they might not be as readily available in Manhattan and Brooklyn, affecting the overall price distribution.

**Correlation analysis :-**



matrix reveals interesting connections between various features of Airbnb listings. Here's a breakdown of some key insights:

**Price Correlations:**

* **Weak Correlation:** Price has a weak positive correlation with minimum\_nights (0.0428). This suggests a slight tendency for higher priced listings to have a minimum night requirement, but it's not a strong relationship.
* **Weak Negative Correlation:** Price has a weak negative correlation with both number\_of\_reviews (-0.0479) and reviews\_per\_month (-0.0506). This is counterintuitive, but it could be due to several factors. For example, some very expensive luxury rentals might be new and have few reviews yet.
* **Weak Positive Correlation:** Price has a weak positive correlation with calculated\_host\_listings\_count (0.0575) and availability\_365 (0.0818). This suggests a slight tendency for hosts with multiple listings and higher year-round availability to offer listings at slightly higher prices.

**Number of Reviews Correlations:**

* **Strong Positive Correlation:** The number of reviews (number\_of\_reviews) has a strong positive correlation with reviews\_per\_month (0.5894). This makes sense, as listings with a high total number of reviews tend to have a higher average number of reviews per month.

**Other Correlations:**

* **Weak Positive Correlation:** minimum\_nights has a weak positive correlation with calculated\_host\_listings\_count (0.1280) and availability\_365 (0.1443). This suggests that hosts with multiple listings or higher year-round availability might set slightly higher minimum night requirements.
* **Weak Positive Correlation:** calculated\_host\_listings\_count has a weak positive correlation with availability\_365 (0.2257). This suggests that hosts managing multiple listings tend to keep them available for rent for a larger portion of the year.
* **Strength of Correlations:** The correlations are mostly weak, so the relationships are not very strong