Exploratory Data Analysis

Report on Airbnb Dataset Analysis

Research Question:

What insights can we derive from the Airbnb dataset to enhance pricing strategies and user experience?

Introduction:

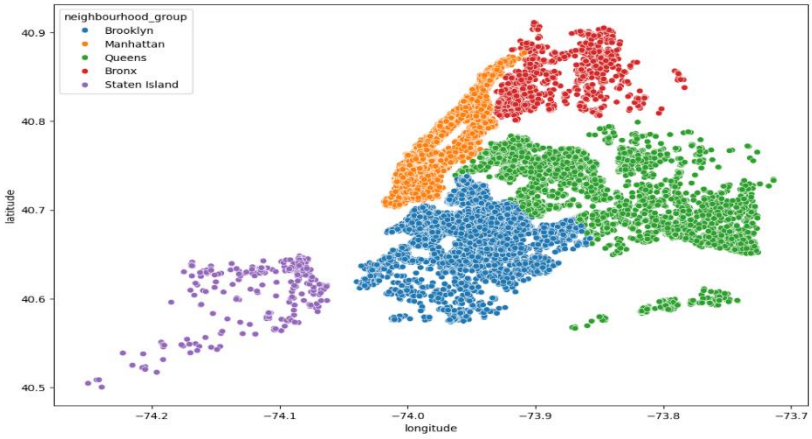
The Airbnb dataset, comprising 48895 rows and 22 columns, presents an opportunity to delve into factors influencing pricing and user satisfaction within the platform. Through exploratory data analysis (EDA), we aim to uncover patterns, address data quality issues, and glean actionable insights.

Exploratory Data Analysis:

We now need to determine the dataset's row and column counts. It's also important to count the number of distinct values in each column. We'll also apply the unique for this. Thus, there are 22 columns and 20758 rows.

Scatter Plot to show the neighbourhood group based on Latitude and Longitude:

To create the location's cluster, we will plot the same latitude and longitude on a scatter plot.

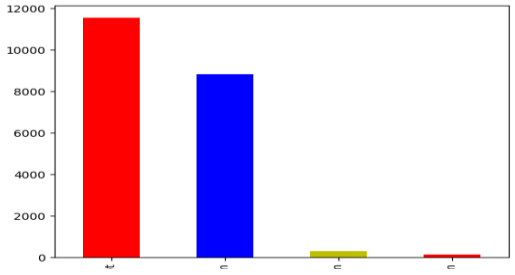


Unique Values

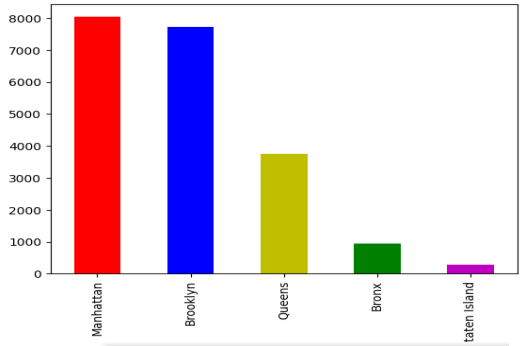
Understanding and analysing the unique values is crucial since it sheds light on the data and the user's preference for a specific option. We'll take Neighborhood group and Room Type. The dataset contains three distinct types of rooms and five different neighborhood groups, as shown below.

```
Unique value for room_type are : ['Private room' 'Entire home/apt'
'Hotel room' 'Shared room']
Unique value for neighbourhood_group are : ['Brooklyn' 'Manhattan'
'Queens' 'Bronx' 'Staten Island']
```

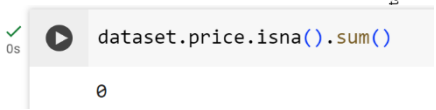
Room Types and Neighbourhood Group: By combining the data, we will first examine the room type distribution. It is evident from the below that private and apartment data exceeds that of shared rooms. Shared accommodations are generally less expensive and might be especially helpful for frequent travellers who transfer cities. Even with the little data on shared rooms, we will continue to search for as much information as we can.



Exploration of Neighbourhood Group: Now let's investigate the neighborhood group to view the distribution of data. According to the list below, Manhattan and Brooklyn appear to have more listings than Queens, the Bronx, and Staten Island.

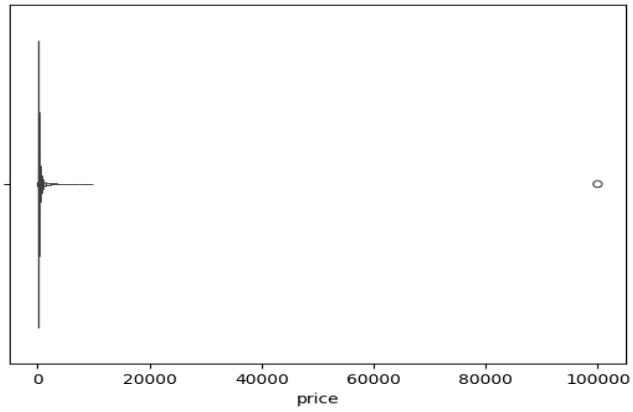


Price Exploration: We will investigate to see whether the price column contains any null values, but based on the information below, it appears that there are none that need to be addressed.



Let's quickly review the pricing information. It is evident from the summary statistics that the Price falls between 0 and 180. However, there is also a charge that can go up to $10,000. We cannot rule this out as an anomaly because price variations occur in a variety of circumstances. The location, kind of room, neighborhood, season, and other variables all affect the pricing. Additionally, as we can see from the chart below, there are a few values that have a value of 0. These could be the result of dynamic pricing or the decision to withhold the price from Airbnb.

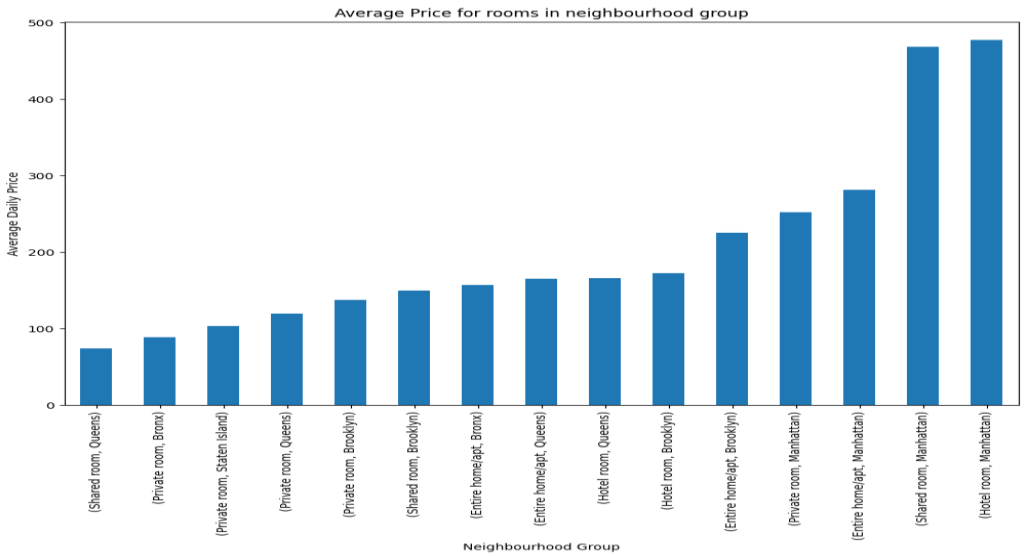
To see how the data is distributed for high price ranges throughout all regions, we can produce a boxplot.



Average Room Rent for Locality:

The price is the most crucial factor for any traveller as it determines the trip's budget. We shall therefore calculate the average cost per night in the section below. We will investigate various room types and calculate the average nightly stay based on the neighborhood group. In every place, renting an apartment is always more expensive than staying in a shared or private room. This is the case since the family rents out the entire room for a pleasant stay, where privacy is a key consideration. On the other hand, travellers who move around a lot and don't want to spend a lot of time in one spot tend to prefer to stay in shared rooms.

Thus, it is evident from the plot that:



The least costly place to stay per night is on Staten Island in a shared room; the most expensive place to stay per night is in Manhattan renting a whole apartment or house.

The typical cost of in addition to being far more expensive than other private rooms in the neighborhood, Manhattan's shared accommodations are also very pricey. This demonstrates that staying in Manhattan is more expensive than in any other place. According to a comparison of neighborhood groups based on room type, Bronx is the most affordable place to stay. While renting an apartment is not the most affordable option on Staten Island, sharing a room is. This might be because it's in the ideal position to escape the bustle of the city and spend time with family.

Additionally, we will provide a list of average rates for each kind of hotel in each neighborhood so that travelers may make budget-conscious plans.

List of Average Price per night based on the neighbourhood group:

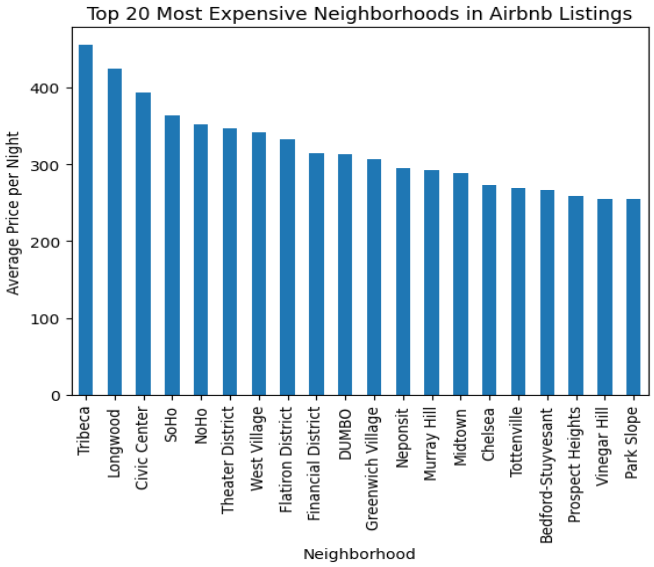
List of Average Price per night based on the neighbourhood group

room_type	neighbourhood_group	price
Entire home/apt	Bronx	156.750000
	Queens	165.218750
	Brooklyn	224.703390
	Manhattan	280.666667
Hotel room	Queens	165.714286
	Brooklyn	172.000000
	Manhattan	476.780822
Private room	Bronx	88.166667
	Staten Island	102.888889
	Queens	118.964072
	Brooklyn	137.162338
	Manhattan	252.100186
Shared room	Queens	74.111111
	Brooklyn	149.400000
	Manhattan	468.250000

High-Priced Neighborhood:

We have looked through the location group thus far, but we haven't found every neighborhood. Finding the location with the biggest price difference for the number of nights a traveller spends is crucial in terms of cost. Let's investigate more to gain more insight.

Thus, we will now chart the priciest neighborhood, and we will only chart the top 15 and bottom 15 neighborhoods in terms of average price. This will assist a tourist in selecting the best neighborhood for his budget.

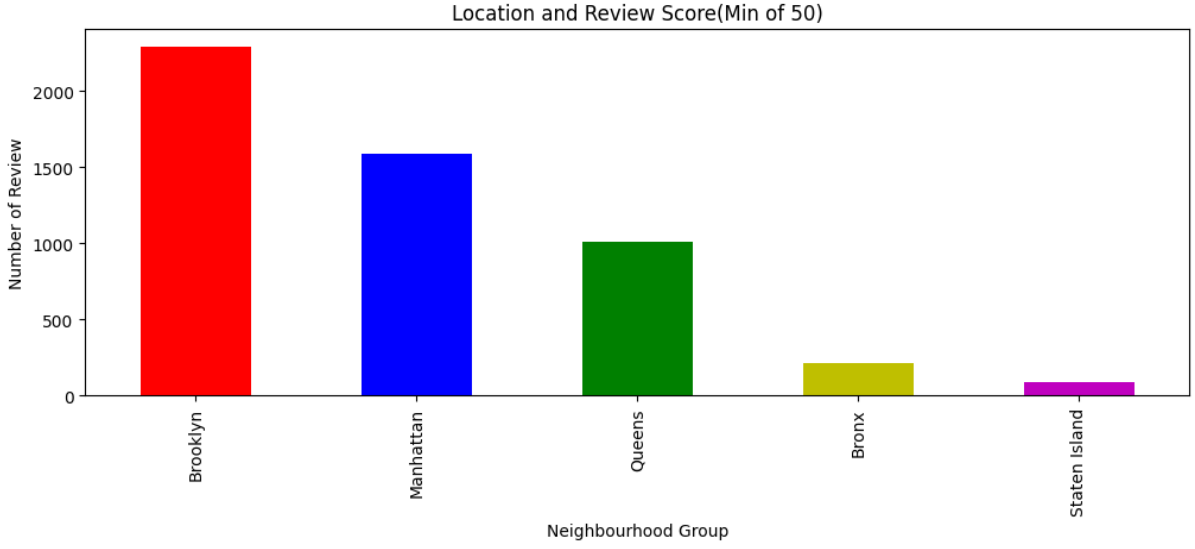


Thus, based on the plot below, Tribeca is the costliest neighborhood. On the other hand, lodging in Park Slope is the least expensive.

Place and Evaluation Points:

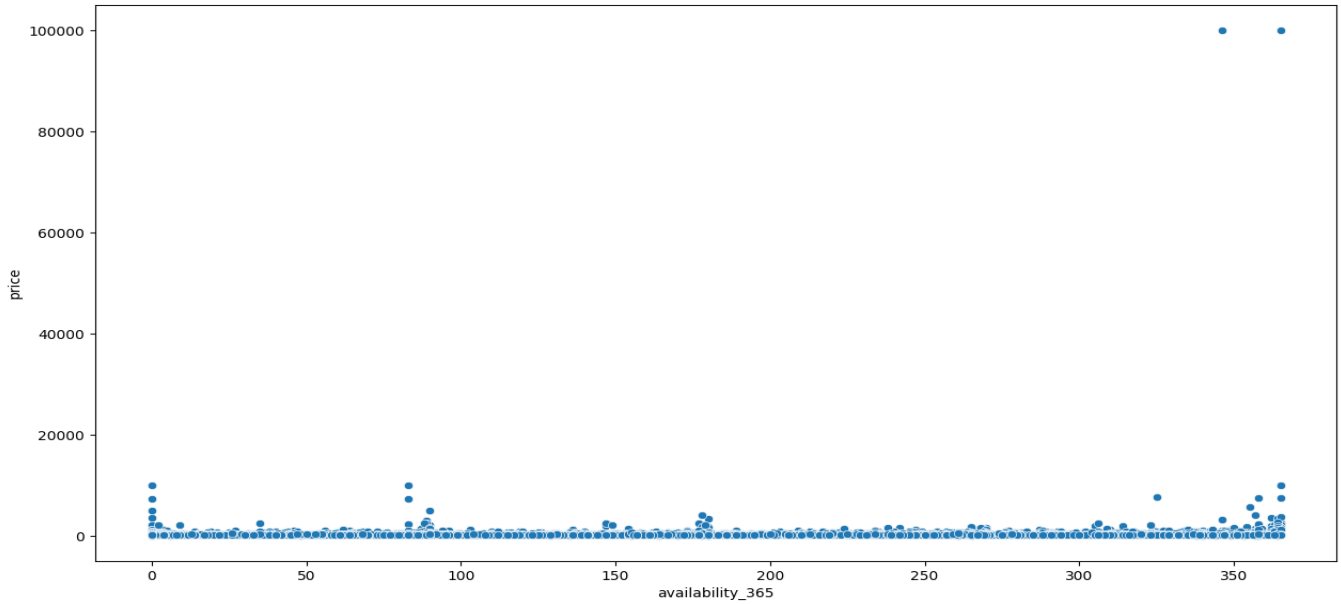
These days, one of the key factors in any online activity is the review. This provides tourists with a wealth of information about a specific location and allows them to choose their online booking mood. An inexpensive hotel with a poor reputation may discourage a traveller from making a reservation, and an expensive hotel with the best reviews may overcharge a visitor beyond his initial expectations. Thus, we shall try to determine how each neighborhood is performing in relation to the review. Given the limited facts under consideration, we shall make every effort to ascertain as much as we can.

First and foremost, we will only consider applicants who have received more reviews.



Plot Cost according to 365-Day Availability

To determine whether there has been a price increase depending on availability, we will create a scatterplot, from which it is difficult to deduce anything. However, it appears that the price goes up to $10,000 with 365-day availability.



Data Imputation:

Filled missing values in 'service fee' using estimated values based on correlation with 'price'. Addressed missing values in 'minimum nights' by filling with mean.

Imputed missing values in 'host_identity_verified', 'country', 'country code', 'instant_bookable', 'cancellation_policy', 'review rate number', 'calculated host listings count', and 'availability 365' using mode or median as appropriate.

Conclusion:

By conducting comprehensive data analysis and preprocessing, we have enhanced the Airbnb dataset's usability and quality. These insights can empower decision-making processes related to pricing strategies, marketing initiatives, and overall user experience enhancement on the Airbnb platform.