Capstone Project

Airbnb Bookings Analysis

Team Members:

Abhishek Kulkarni

Animesh Mahajan

Nileshkumar Lavand

Prajwal Kulkarni

Business Overview

- Airbnb - “Air bed and breakfast”
- Presence in 220 countries
- It has 150 million customers worldwide
- Revenue : 4.8$ billion in 2019
- Business : Lodging, Hospitality
- This dataset consists of booking data at various airbnb’s throughout NYC in 2019

Problem Statement & Business Objective

The main objective of this exploratory data analysis (EDA) project is to understand the dynamics of Airbnb rentals in New York City (NYC) by analyzing various factors such as pricing, availability, neighborhood groups, room types, and customer feedback. By examining the average price, number of listings, availability, and customer reviews, we aim to gain insights into the pricing patterns, popularity of neighborhoods, customer satisfaction, and market opportunities. This analysis will provide valuable information for potential investors or hosts to make informed decisions and understand the key factors influencing the success and profitability of Airbnb rentals in NYC.

Business Objective

The business objective is to maximize occupancy rates, profitability, and guest satisfaction in the client's Airbnb rental business in New York City.

Methodology

Dataset First Look

```
# Dataset First Look
df.head()
```

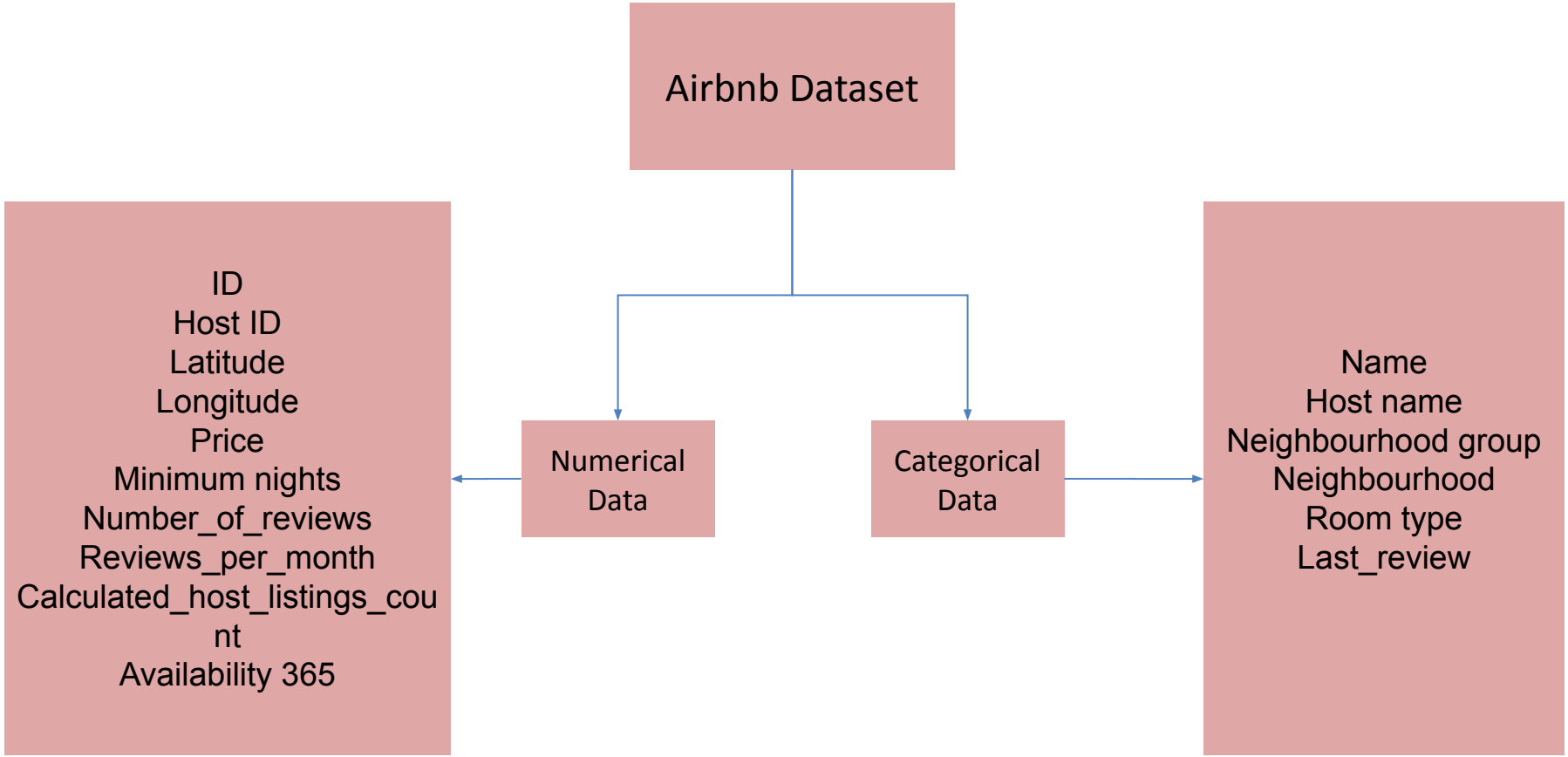
	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN	1	365
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	1	0

```
# Last 5 rows
df.tail()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	NaN	NaN	2	9
48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	NaN	NaN	2	36
48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN	NaN	1	27
48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	NaN	NaN	6	2
48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	90	7	0	NaN	NaN	1	23

48894 Rows and 16 Columns

Data Summary



Variables

Variable	Description	Datatype	Unique Values
Id	Unique ID of the Airbnb listing	int64	48895
name	Name of the Airbnb listing	object	47905
host_id	Unique Host ID	int64	37457
host_name	Name of the Host	object	11452
neighbourhood_group	Location	object	5
neighbourhood	Area	object	221
latitude	Latitude Range	float64	19048
longitude	Longitude Range	float64	14718

Variables

Variable	Description	Datatype	Unique Values
room_type	Type of Room listed	object	**3**
price	Price of listing	int64	674
minimum_nights	Minimum nights to be paid for	int64	109
number_of_reviews	Number of reviews	int64	394
last_review	Last date of review	object	1764
reviews_per_month	Number of reviews per month	float64	937
calculated_host_listings_count	Total count of listings of the host	int64	47
availability_365	Availability around the year	int64	366

Data Cleaning

```
# Check if all id values are unique
df['id'].nunique()
```

```
48895
```

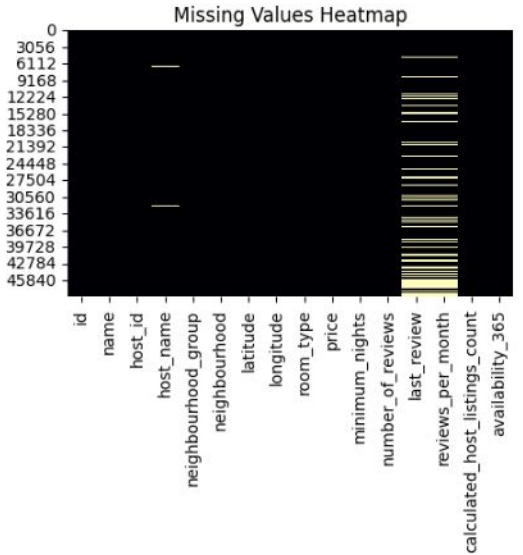
```
# Dataset Duplicate Value Count
duplicate_count = df.duplicated().sum()
print("Number of duplicate values:", duplicate_count)
```

```
Number of duplicate values: 0
```

- The dataset has no duplicate values
- The dataset has null values in name, host_name, number_of_reviews & last_review column

```
# Visualizing the missing values

plt.figure(figsize=(5, 3))
sns.heatmap(df.isnull(), cbar=False, cmap='magma')
plt.title('Missing Values Heatmap')
plt.show()
```



Data Wrangling

```
# Checking the number of rows with zero price values in our dataset.
df2[df2.price==0].shape
```

```
(11, 16)
```

- Here we can clearly see that these eleven entries need to be drop in order to get meaningful analysis so we will be dropping these entries where price equals to zero ("0").

```
# Removing rows having zero price values from our dataset.
df2 = df2[df2['price'] !=0]
```

```
# Since both 'name' and 'host_name' has very less missing values as compared to the
# Entire dataset we can fill these missing values in both with unknown & no_name respectively!

df2['name'].fillna('unknown',inplace=True)
df2['host_name'].fillna('anonymous',inplace=True)
```

```
# Checks if any null values still exist!
df2[['host_name','name']].isnull().values.any()
```

```
False
```

Data Wrangling (Contd)

```
# Filling null values in reviews_per_month column with 0
df2['reviews_per_month'] = df['reviews_per_month'].fillna(0)
```

```
# Drop the last_review column
df2.drop(['last_review'], axis=1, inplace=True)
```

```
# Checking change in shape
df2.shape
```

```
(48884, 15)
```

Previous shape was (48895, 16) and the shape of the new data frame created after cleaning is (48884, 15) Hence correctly removed 11 rows where the price value was 0

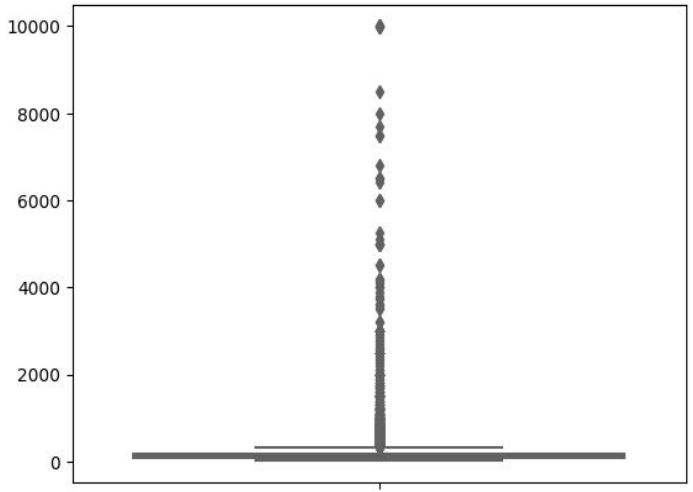
```
# Checking the changes
df2.isnull().sum()
```

```
id                                0
name                              0
host_id                           0
host_name                         0
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
reviews_per_month                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

Data Wrangling (Contd)

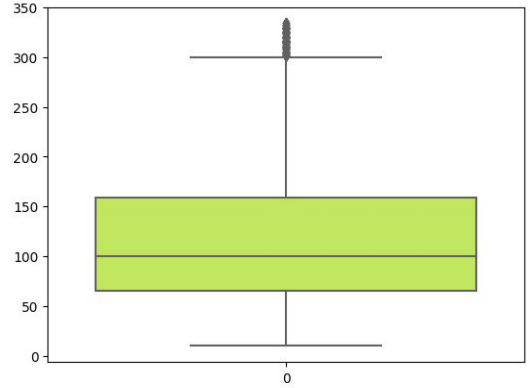
```
sns.boxplot(df2['price'], color='#cafe48')
```

```
<Axes: >
```



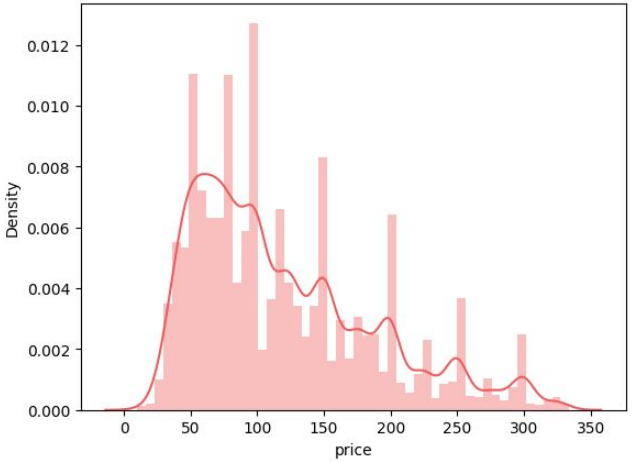

```
sns.boxplot(df2['price'], color='#cafe48') #the boxplot for price now looks good without outliers!
```

```
<Axes: >
```

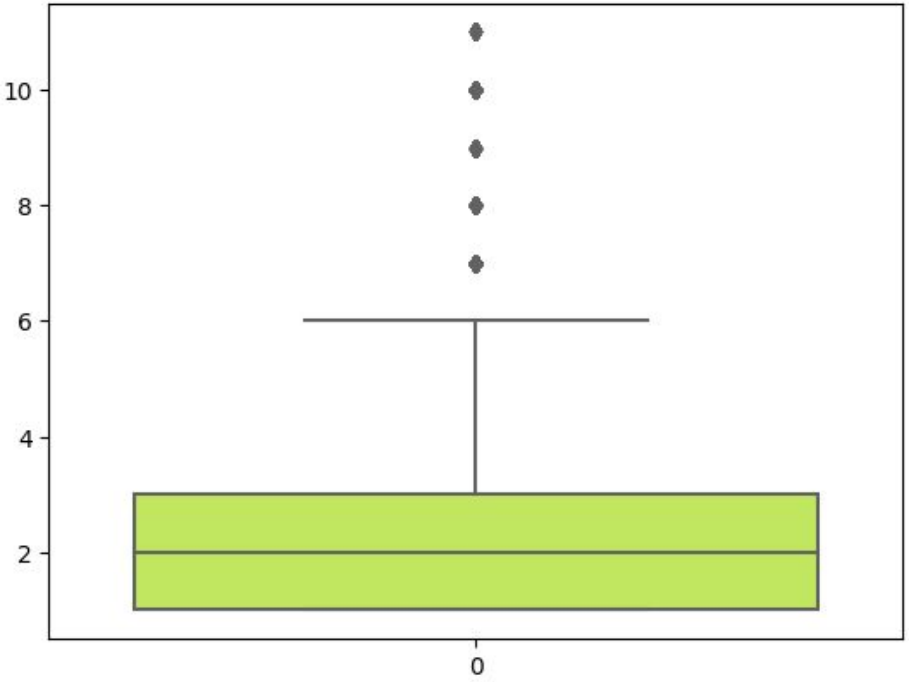


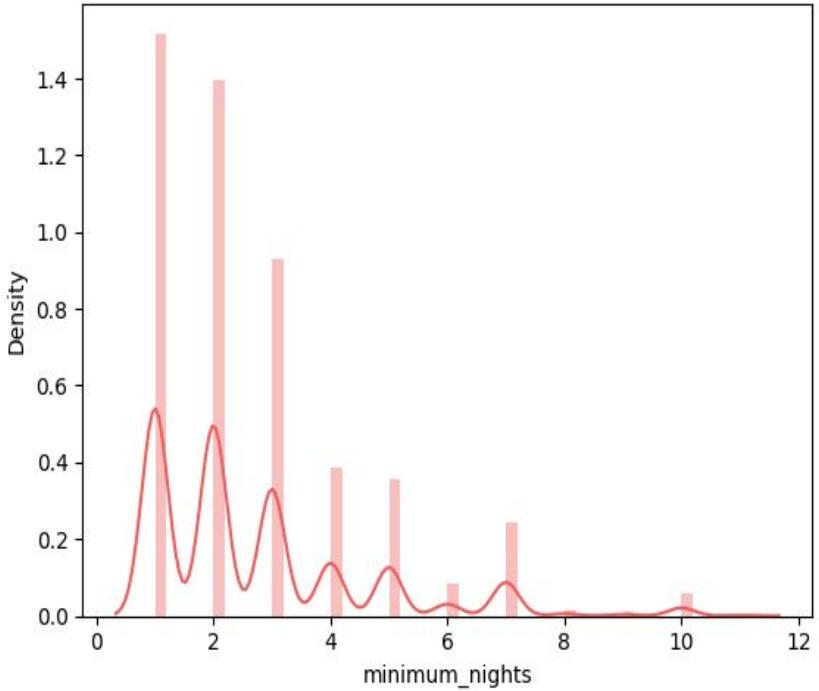

```
sns.distplot(df2['price'], color='#f05d5e')
```

```
<Axes: xlabel='price', ylabel='Density'>
```

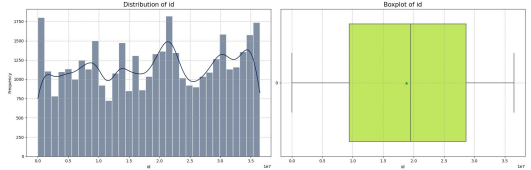


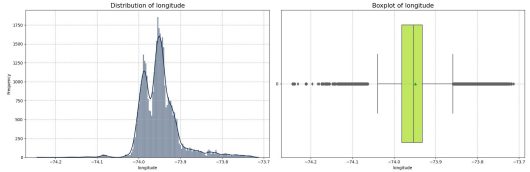
Data Wrangling (Contd)

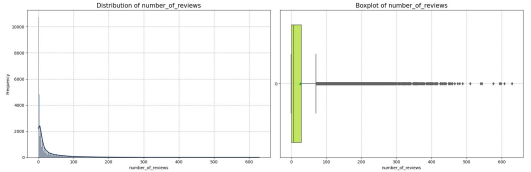


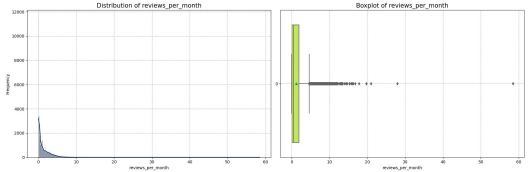


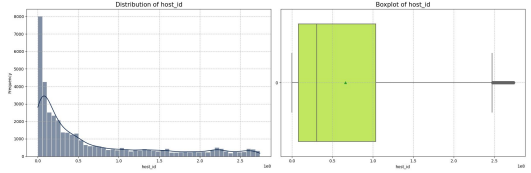
Data Visualization

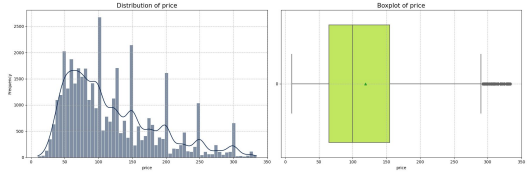


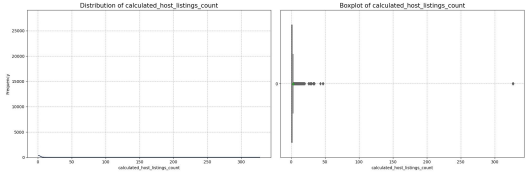


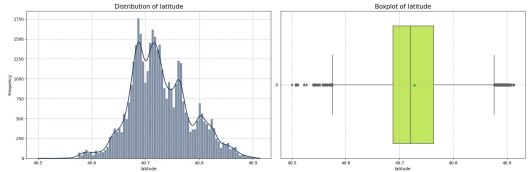


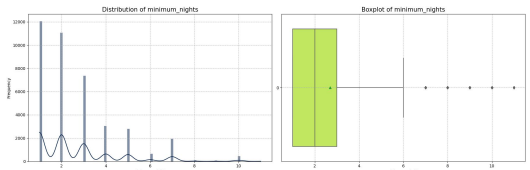


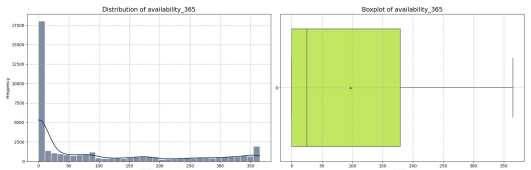




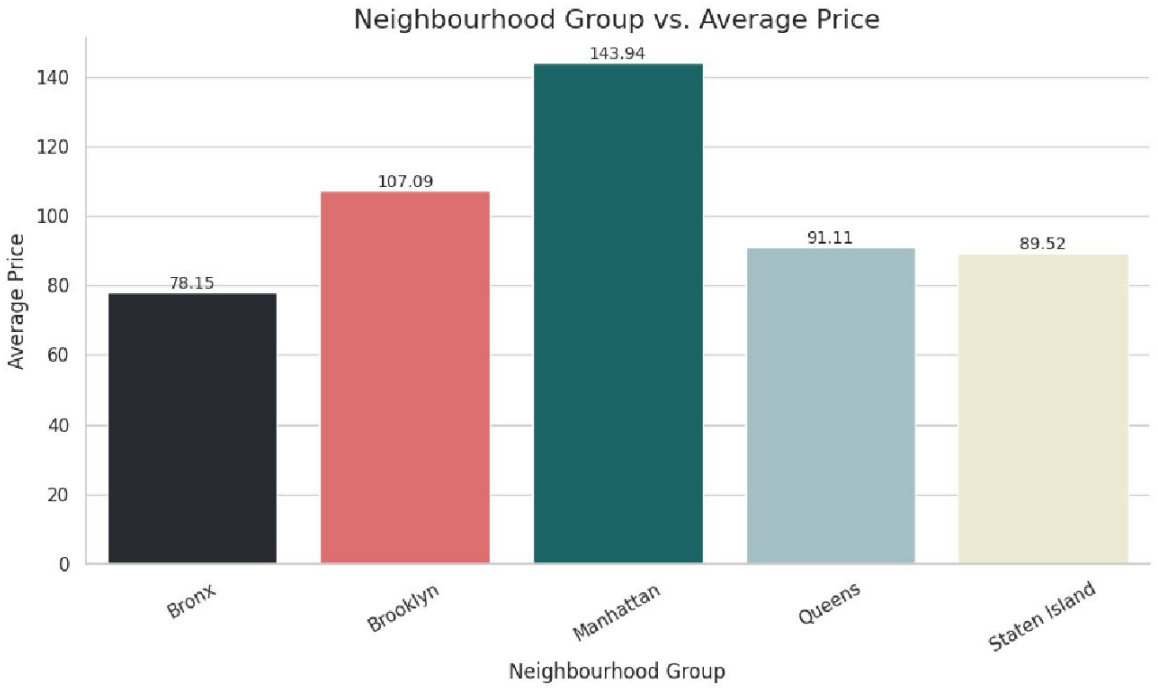




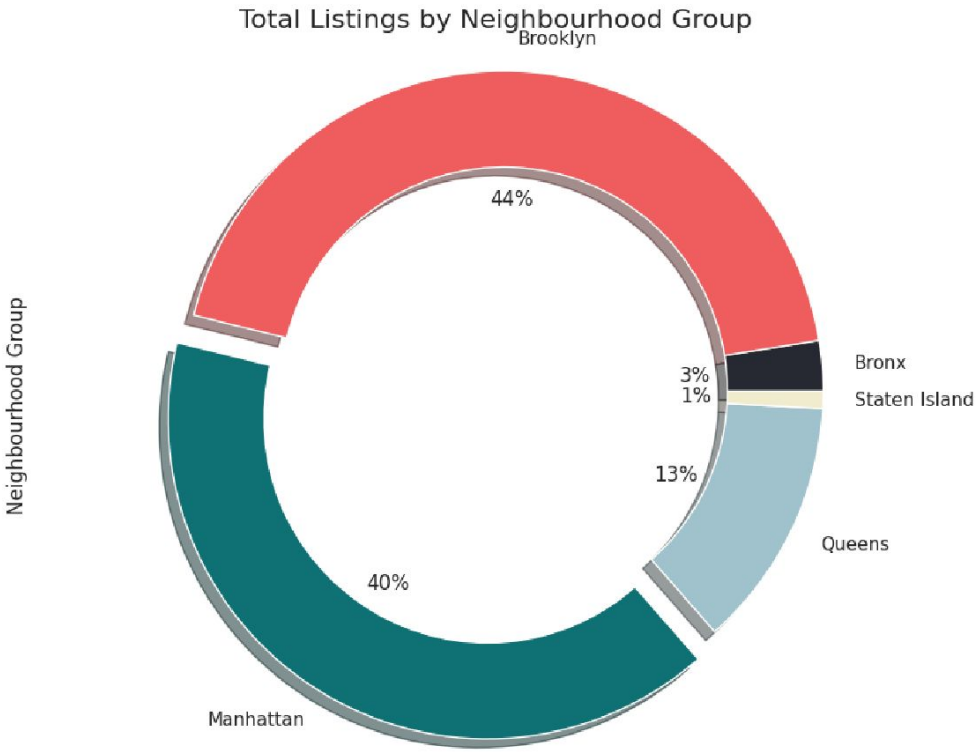




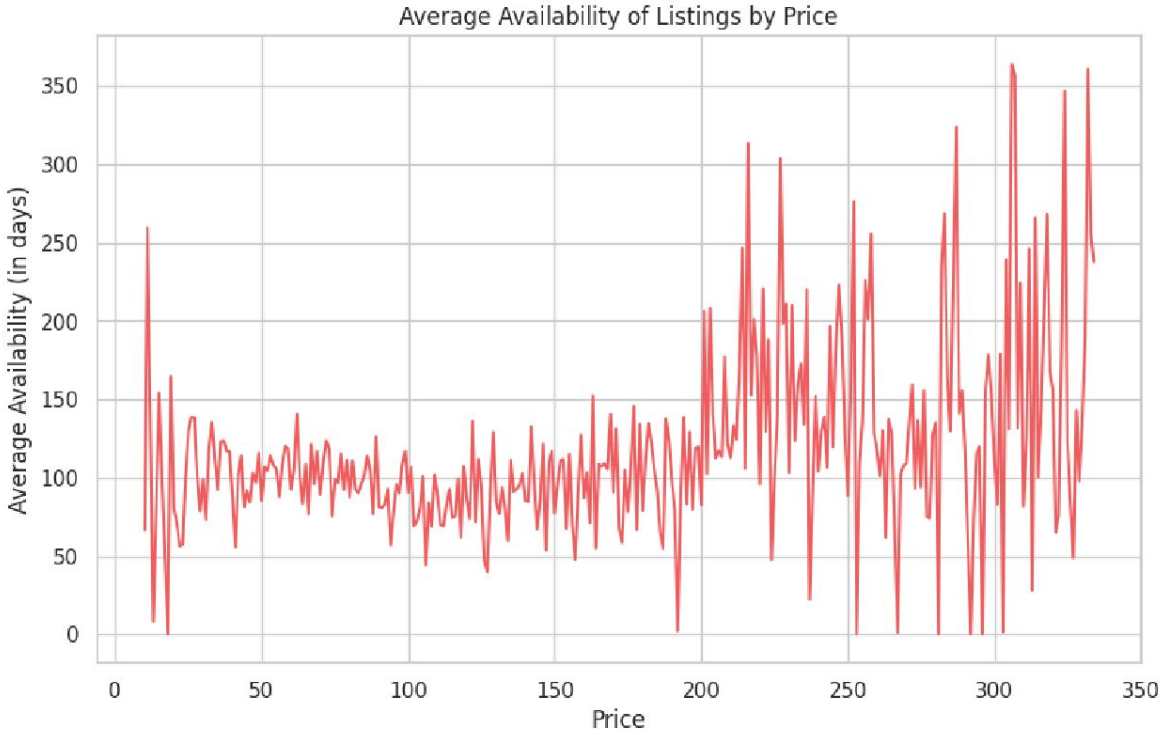
Data Visualization



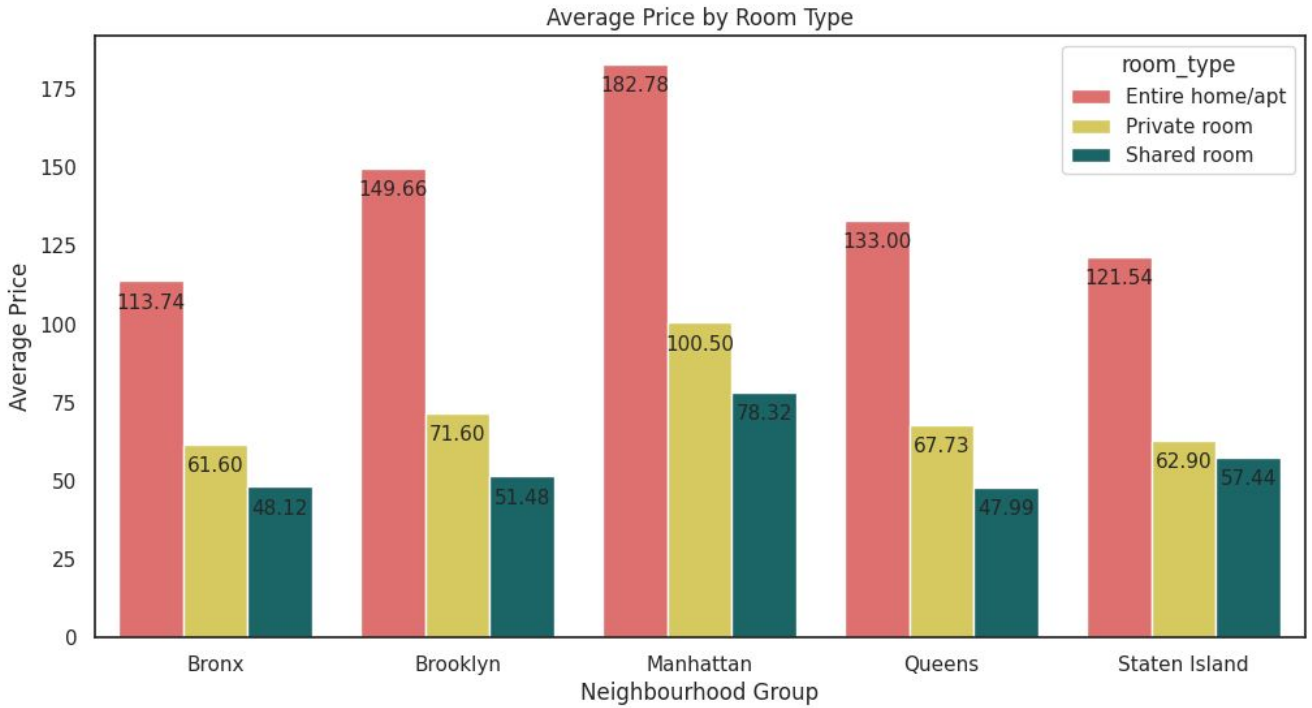
Data Visualization



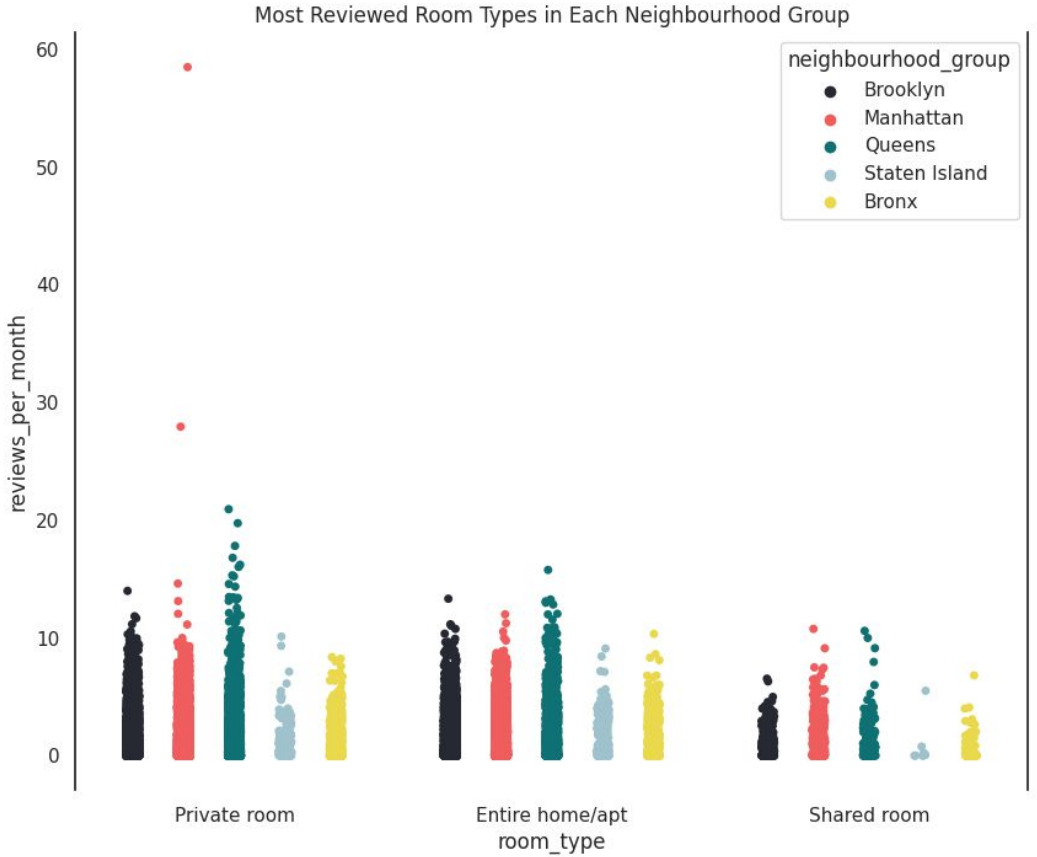
Data Visualization



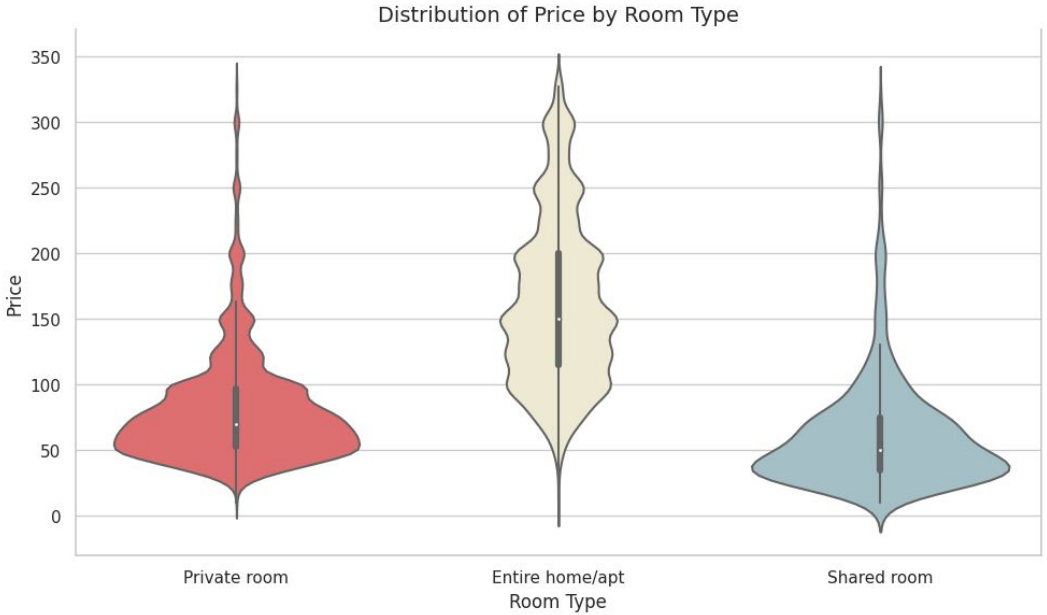
Data Visualization

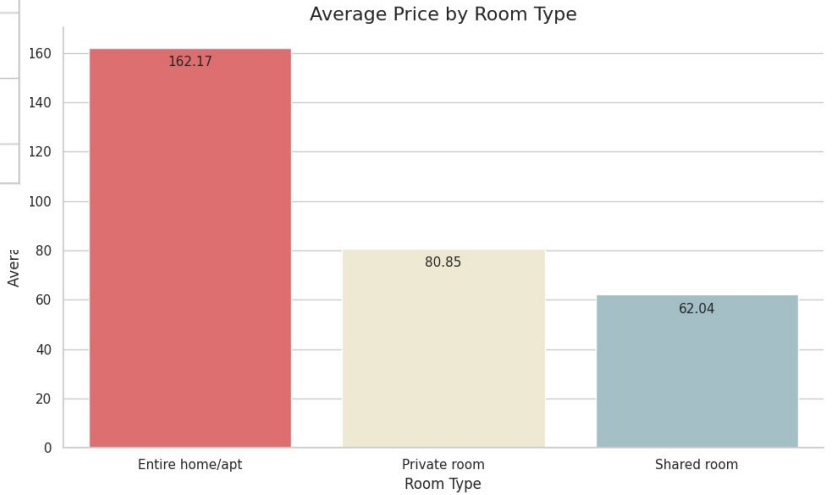


Data Visualization

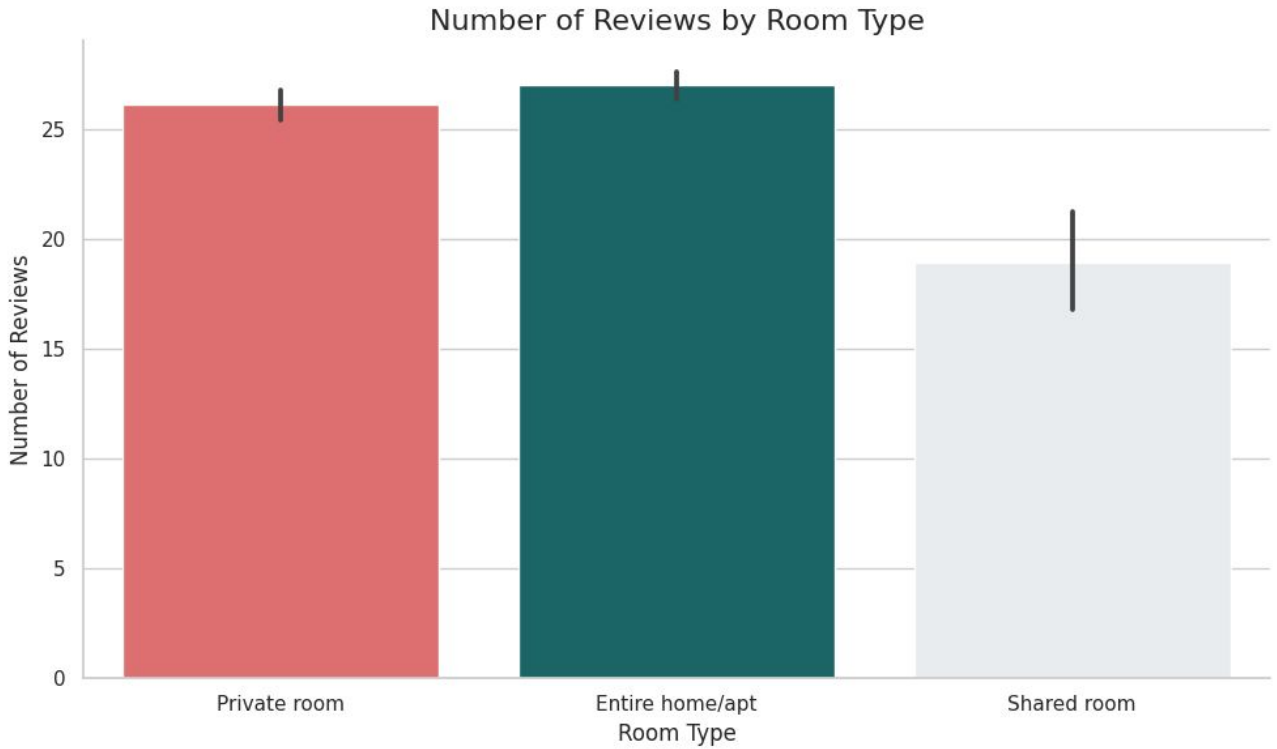


Data Visualization

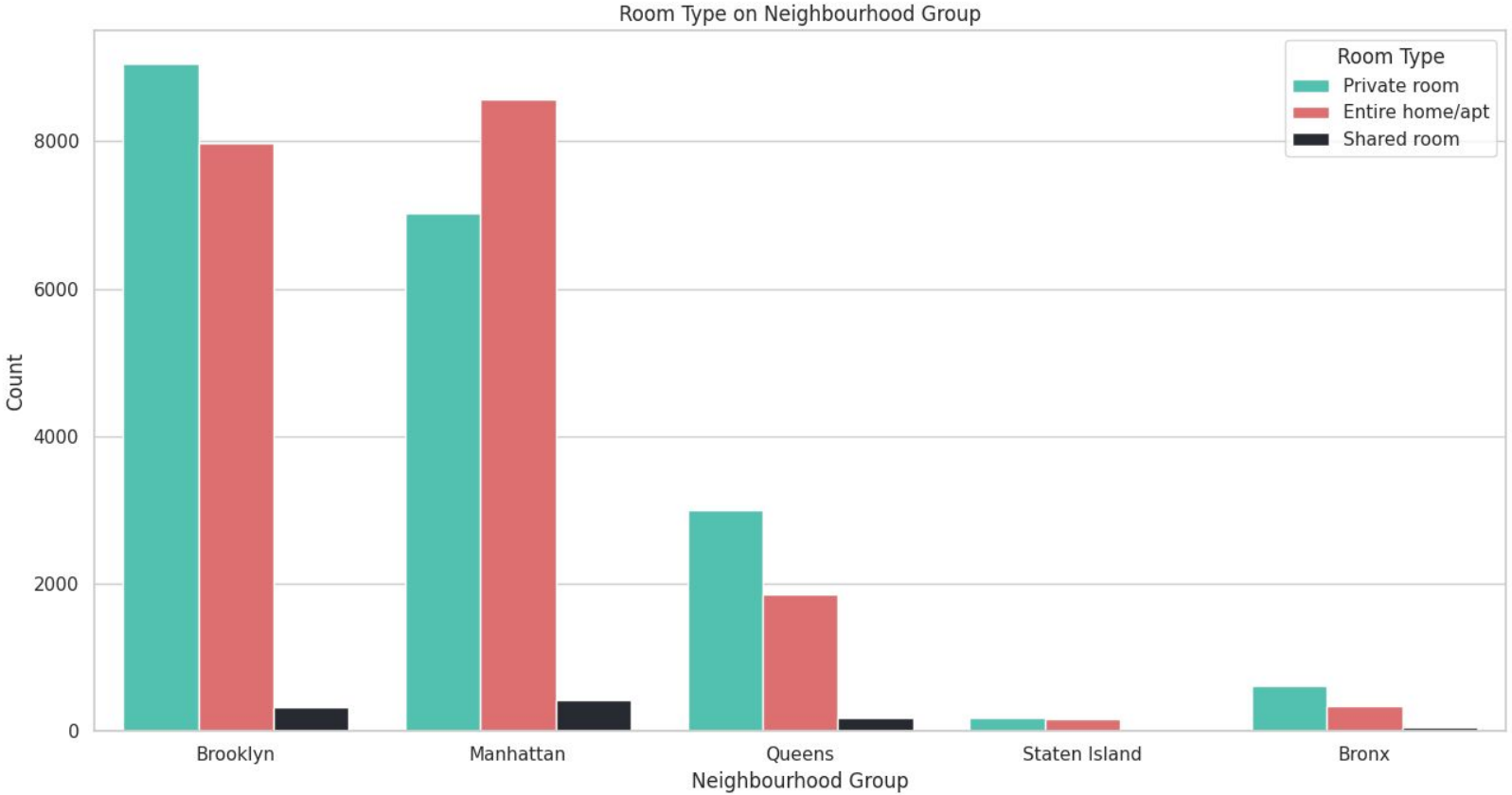




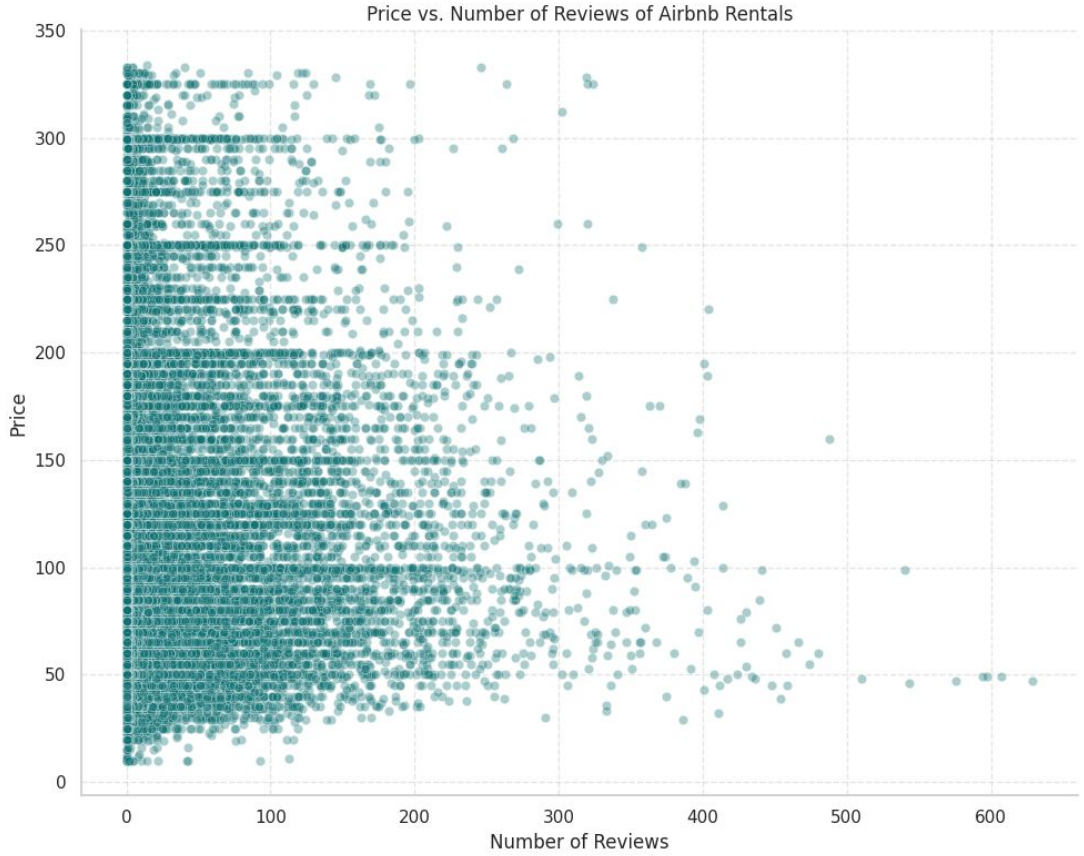
Data Visualization



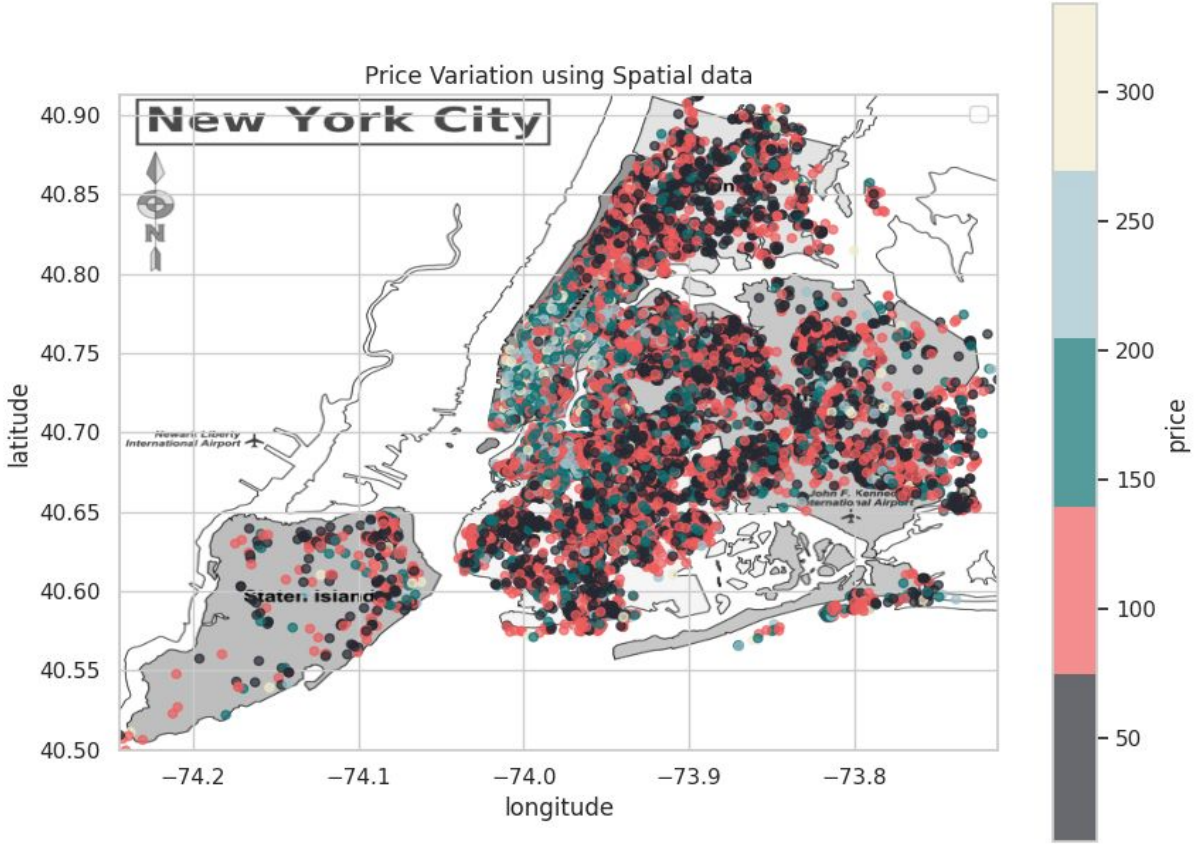
Data Visualization



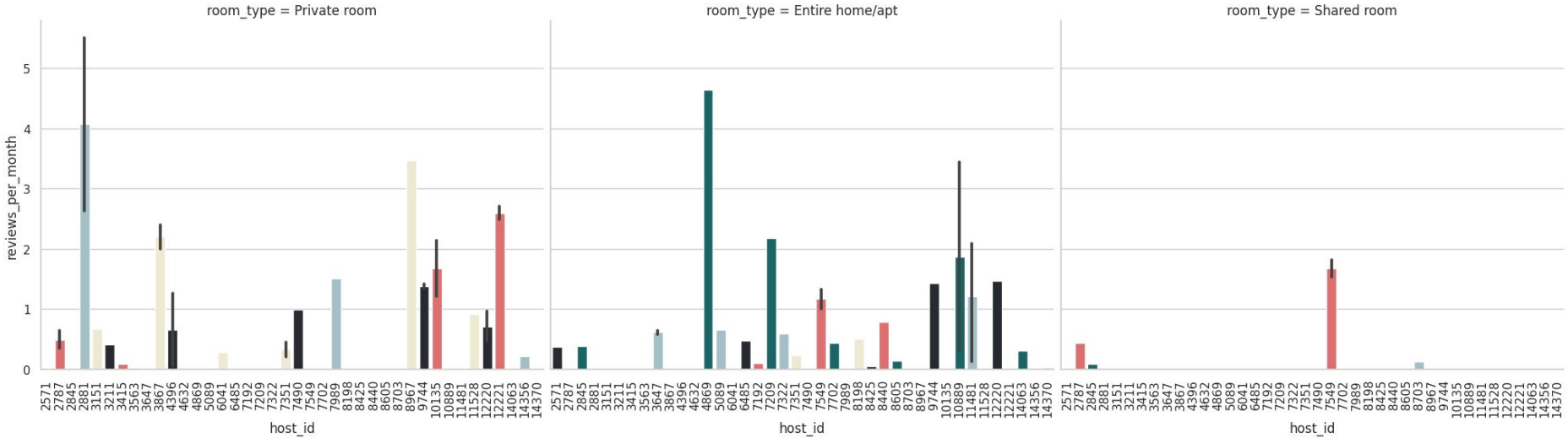
Data Visualization



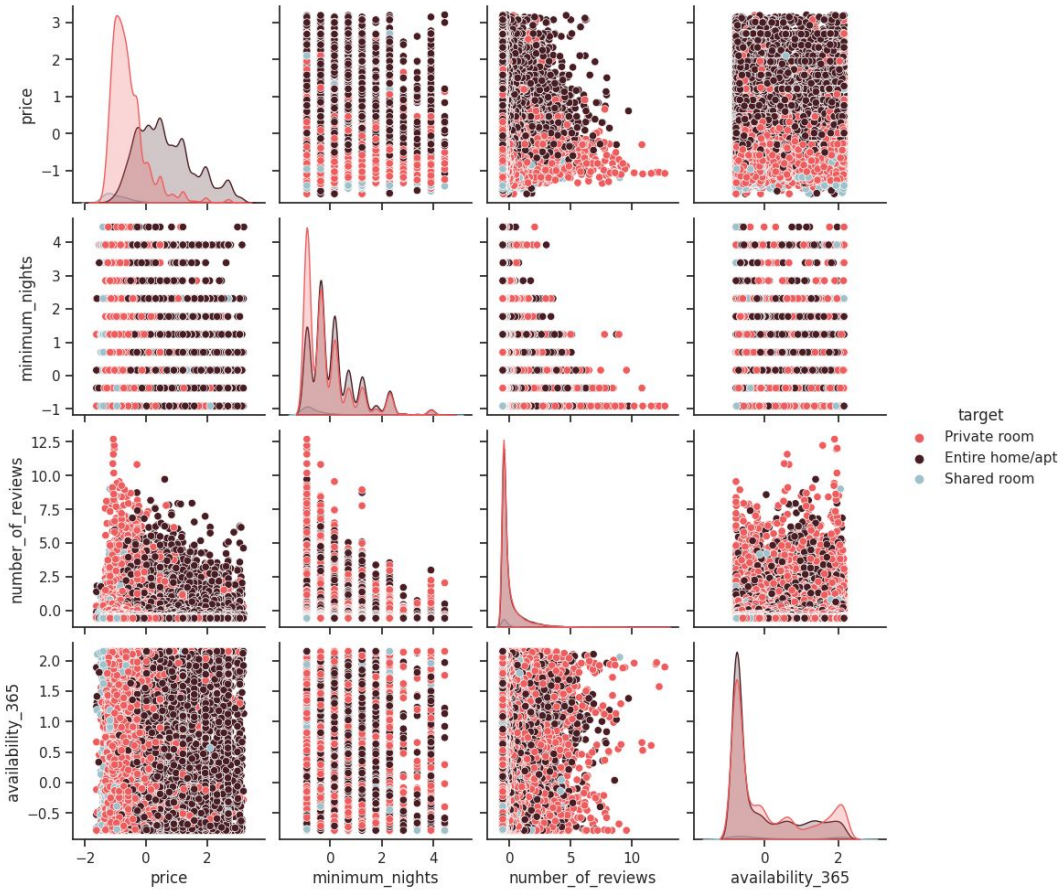
Data Visualization



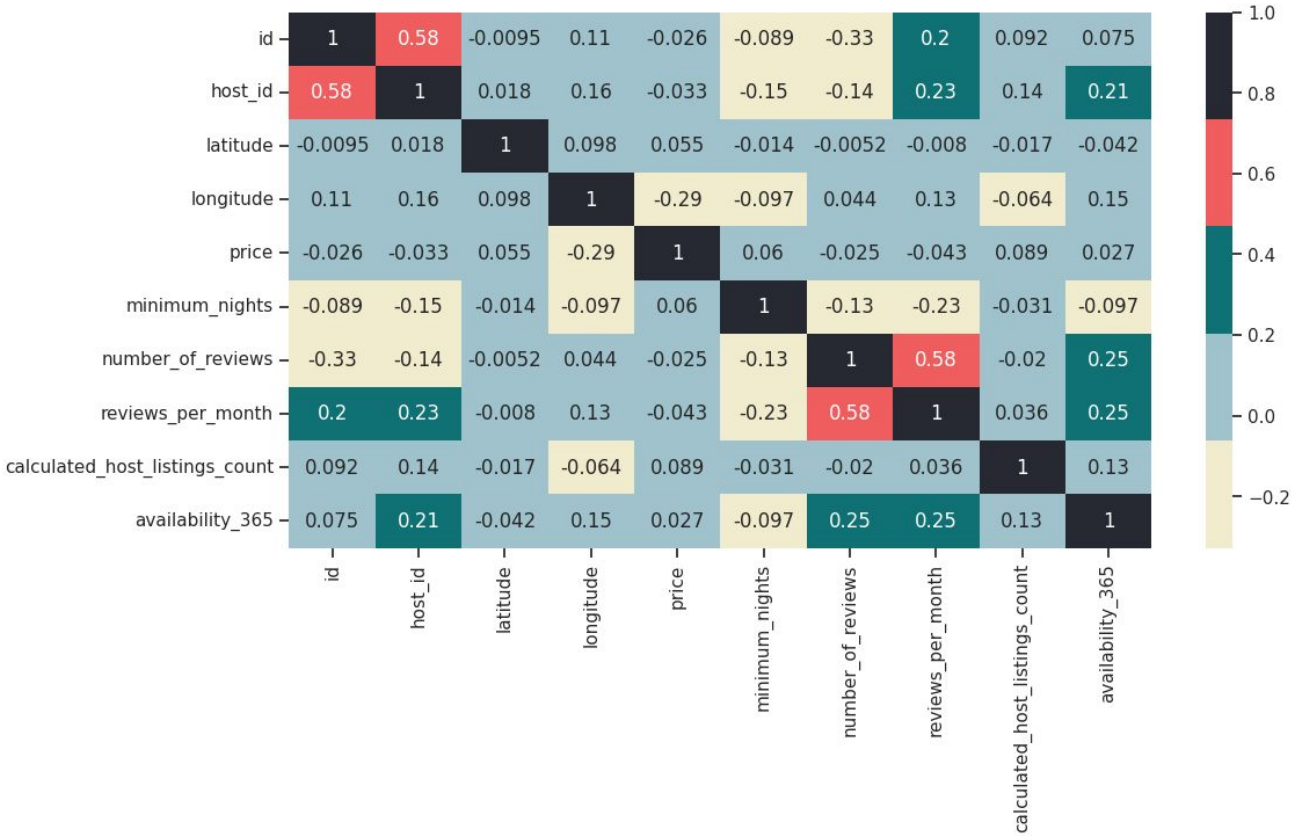
Data Visualization



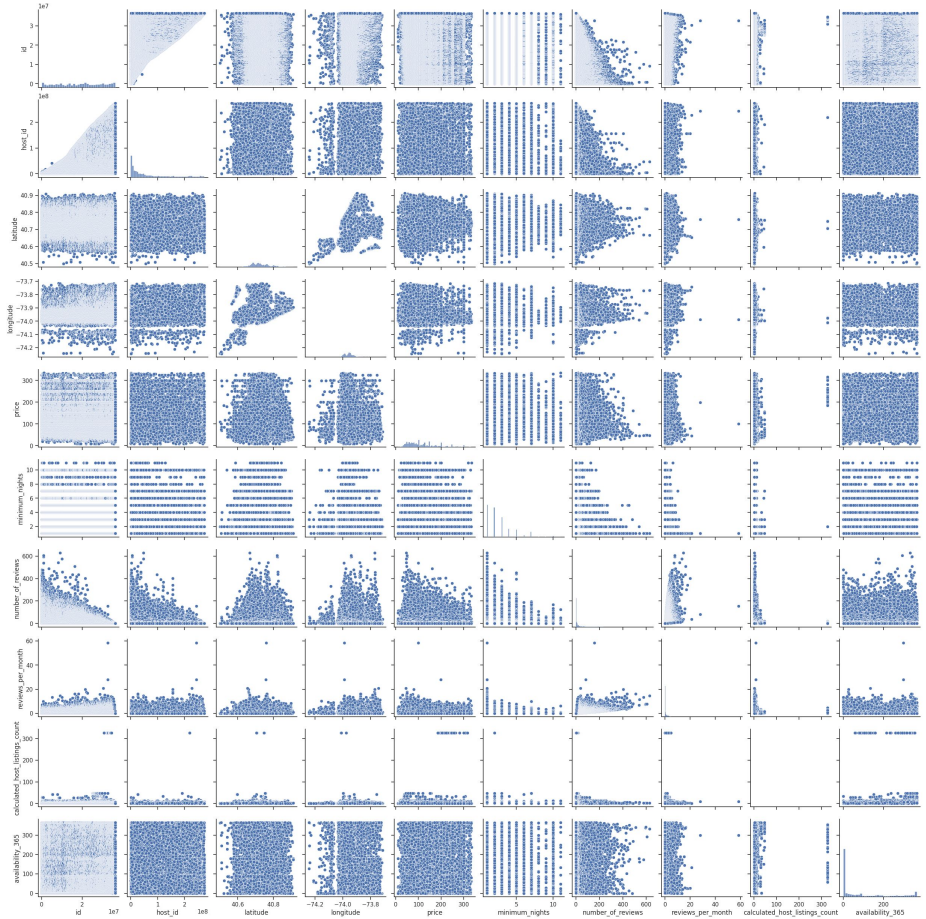
Data Visualization



Data Visualization



Data Visualization



Conclusion

1. The 2019 Airbnb NYC project revealed a significant growth in the number of listings in New York City, indicating a booming short-term rental market.
2. Manhattan emerged as the borough with the highest concentration of Airbnb listings.
3. Pricing analysis showed that Manhattan consistently had the highest average nightly rates, while the outer boroughs offered more affordable options.
4. Entire homes and apartments were the most common types of listings, indicating a preference for private accommodations among hosts and guests.
5. However, shared rooms were also available, catering to different budget and accommodation preferences.
6. Host demographics revealed that the majority of hosts in NYC were individuals who rented out their primary residences, suggesting a trend of using short-term rentals for additional income.
7. The presence of hosts with multiple listings highlighted the emergence of professional hosts and potential concerns around commercialization.
8. It is important to note that the project's findings were based on data of only 2019, and subsequent developments in regulations may have further shaped the market dynamics.

Solution to Business Problems

1. Optimize pricing strategy based on factors like location, property type, and amenities to maximize occupancy rates and profitability.
2. Enhance the quality of listings by improving descriptions, photos, and amenities to attract potential guests.
3. Prioritize excellent customer service to ensure a positive guest experience and encourage positive reviews.
4. Maintain consistent property availability, especially during peak seasons and events, to maximize occupancy rates.
5. Monitor competitor activity to stay competitive in the local market and adjust strategies accordingly.
6. Implement special offers, discounts, or loyalty programs to attract and retain guests.
7. Stay compliant with local regulations and legal requirements for short-term rentals.
8. Seek guest feedback to identify areas for improvement and enhance the overall guest experience.
9. Foster positive community relations by being respectful, responsible, and considerate to maintain good community relationships.

By following these suggestions, the client can optimize their operations, attract more guests, and achieve their business objectives of maximizing occupancy rates, profitability, and guest satisfaction.

Thank You