UNLOCKING AIRBNB NYC

PRESENTATION - I

AANCHAL SAXENA
ANIKET SHAMBHARKAR
ARUNIMA BARMAN

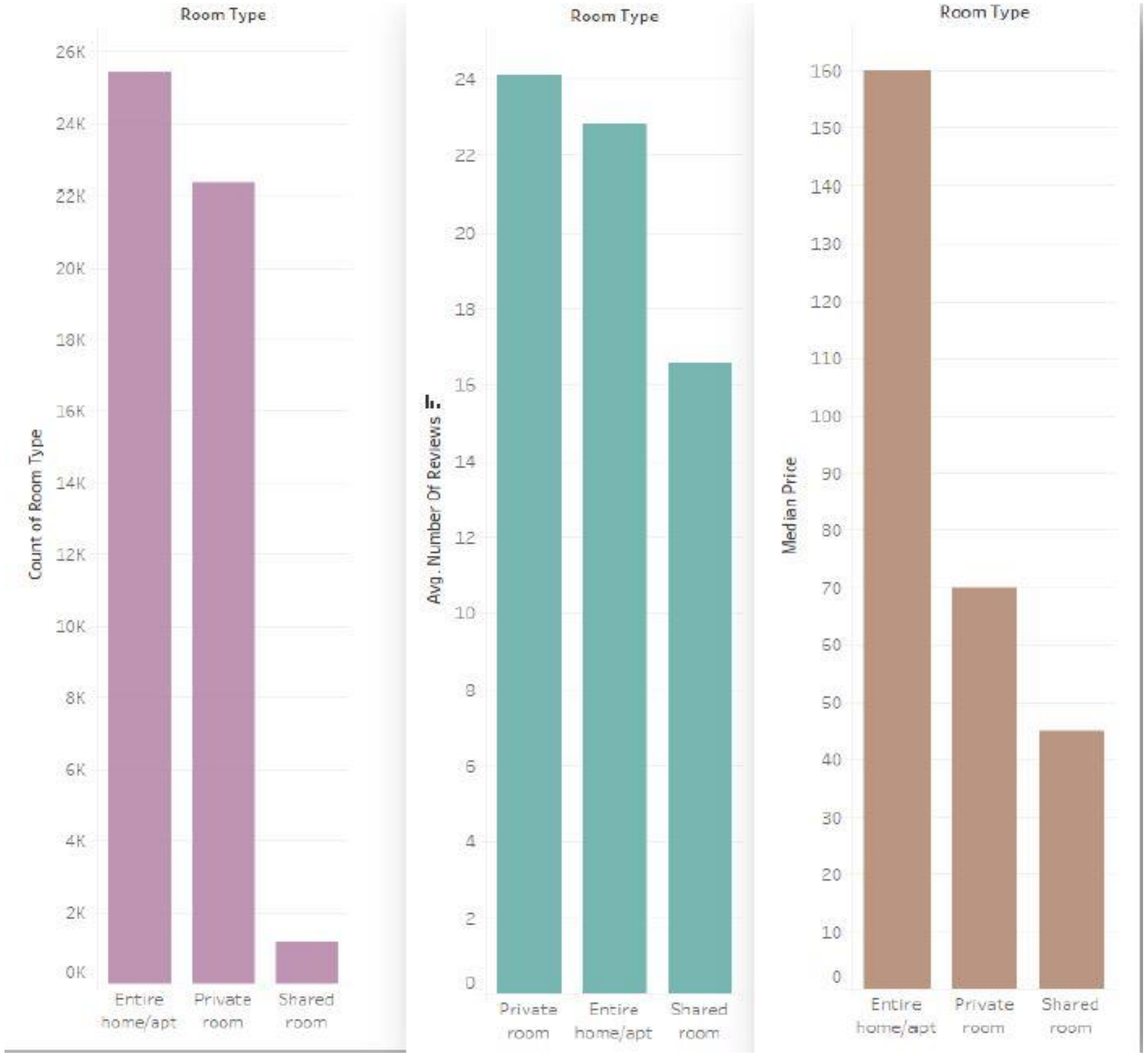
OBJECTIVE

- Understand the Airbnb market in NYC and do a thorough analysis
- Identify key factors influencing occupancy and pricing
- Provide actionable insights to increase revenue

AGENDA

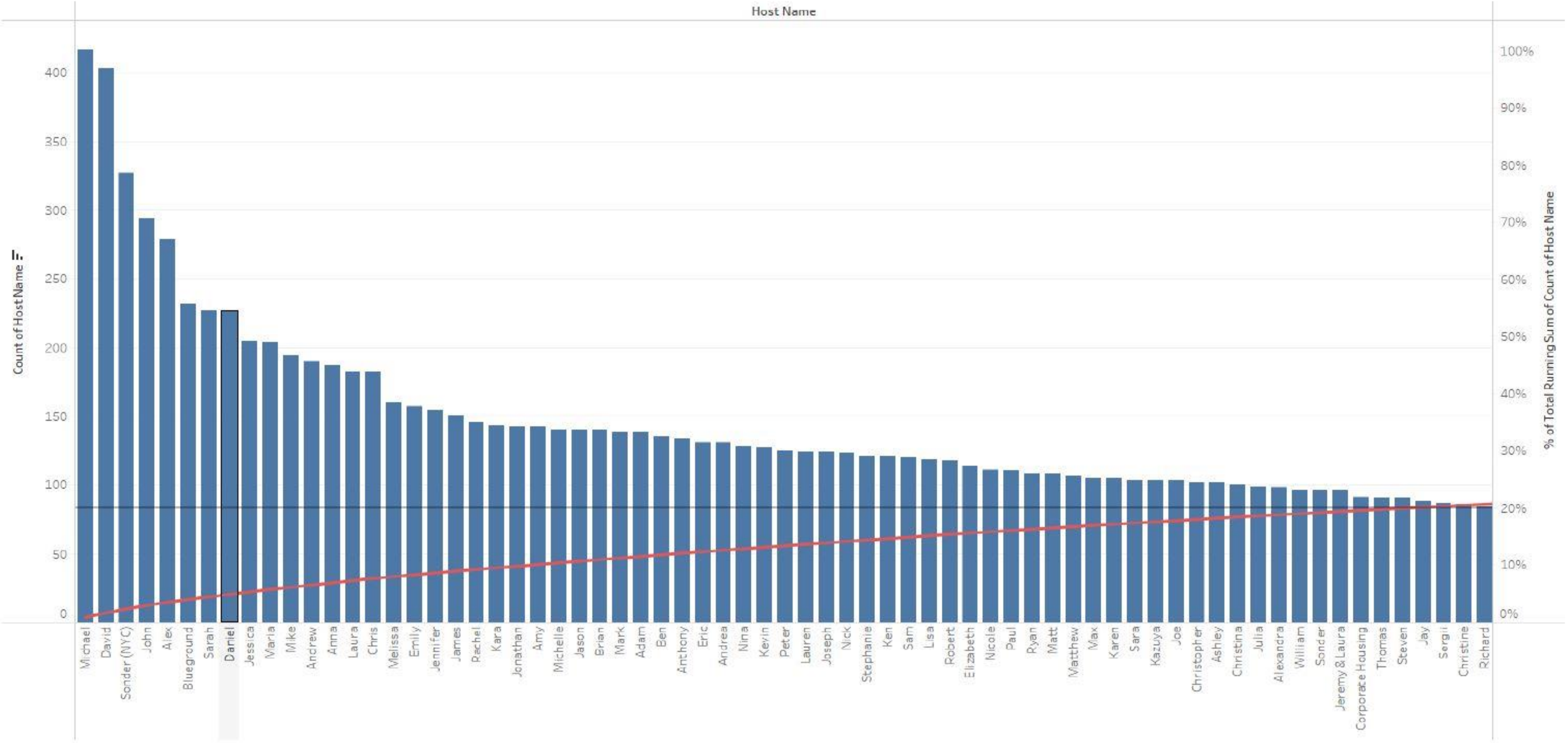
- Insights
  - Problems with Shared Rooms
  - Every Host Matters
  - Most Contributing Neighborhoods/Minimum Night Categories
  - Effect of Minimum Nights on Reviews
- Conclusion
- Appendix
  - Data Sources/Methodology
  - Steps Performed during the Analysis

Problems With Shared Rooms



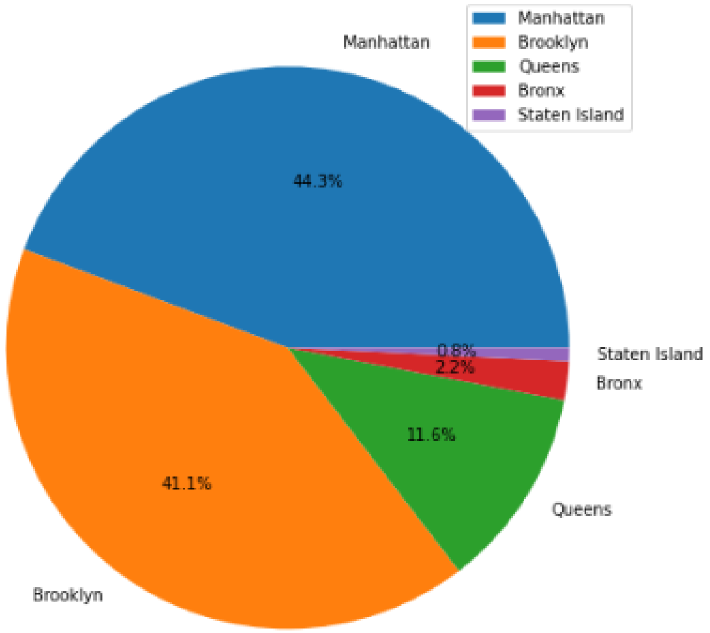
- Shared rooms only account for 2 % of the total types of rooms.
- They are less likely to be reviewed.
- Median rates for shared rooms are significantly lower.

Every Host Matter



The Top 60 hosts constitute only 20 % of the total host count

Most Contributing Neighborhoods



Here are some key points about this observation:

Popularity: Manhattan and Brooklyn are the most popular areas for Airbnb listings, likely due to their central locations and vibrant neighborhoods.

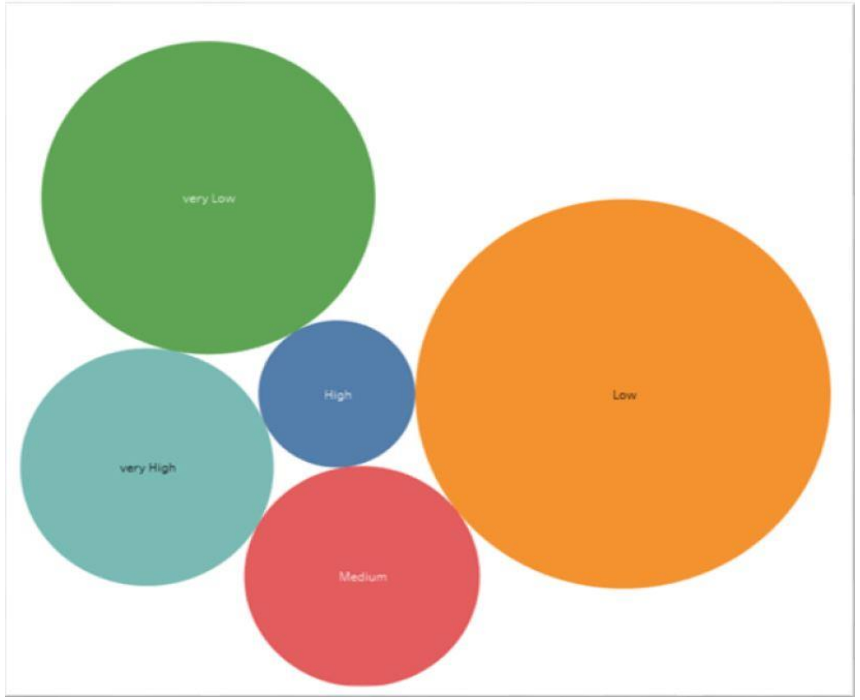
Demand: High demand for accommodation in these areas is driven by their cultural attractions, business centers, and tourist spots.

Diverse Options: These neighborhoods offer a diverse range of properties, from luxury apartments in Manhattan to charming brownstones in Brooklyn, catering to various preferences and budgets.

Accessibility: Both Manhattan and Brooklyn are well-connected by public transportation, making them convenient for visitors.

These factors contribute to the high concentration of Airbnb listings in Manhattan and Brooklyn.

Minimum Night Categories

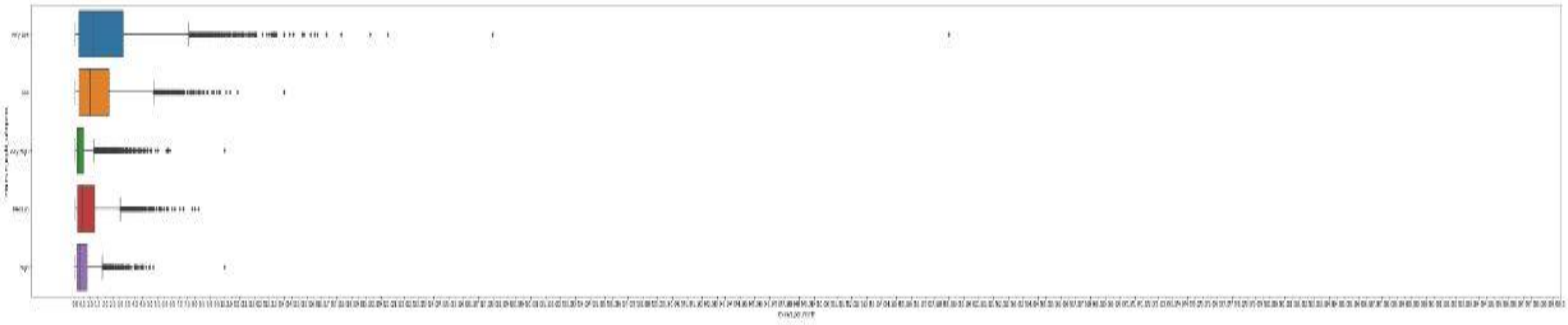


Minimum night category percentages

Low	40.280192
very Low	26.014930
very High	14.997444
Medium	12.960425
High	5.747009

Low Category in minimum_night contributes 40% only.

Effect Of Minimum Night On Reviews



Customer's are more likely to leave reviews for low number of minimum nights

Adjustments in the existing properties to make it more customer-oriented ?

minimum_nights should be on the lower side to make properties more customer-oriented

CONCLUSION

- Significant insights have been derived from various attributes in the dataset.

- A wide range of visuals have been included in the presentations for stakeholders.

- The data collection team should gather data on review scores to enhance future analyses.

- A clustering machine learning model can be developed to identify groups of similar objects in datasets with multiple variables.

APPENDIX

DATA SOURCES

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

METHODOLOGY

- Conduct a thorough analysis of New York Airbnb Dataset.
- Clean the data set using python.
- Derived necessary features.
- Used group aggregation & other statistical methods.
- Created charts and visualizations using univariate and bivariate analysis in Python.

Steps Performed during the Analysis

1. Importing Libraries & Reading The Data

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

```
airbnb = pd.read_csv("AB_NYC_2019.csv")
airbnb.head()
```

2. Data Exploration & Variable Identification

```
#checking what are the variables here:
airbnb.columns
```

```
Index(['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'],
      dtype='object')
```

```
#basic information about the dataset
airbnb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
```

3. Creating Features

3.1 Categorizing The "availability_365" Column Into 5 Categories

```
def availability_365_categories_function(row):
    """
    Categorizes the "minimum_nights" column into 5 categories
    """
    if row <= 1:
        return 'very Low'
    elif row <= 100:
        return 'Low'
    elif row <= 200 :
        return 'Medium'
    elif (row <= 300):
        return 'High'
    else:
        return 'very High'
```

3.2 Categorizing The "minimum_nights" Column Into 5 Categories

```
def minimum_night_categories_function(row):
    """
    Categorizes the "minimum_nights" column into 5 categories
    """
    if row <= 1:
        return 'very Low'
    elif row <= 3:
        return 'Low'
    elif row <= 5 :
        return 'Medium'
    elif (row <= 7):
        return 'High'
    else:
        return 'very High'
```

3.3 Categorizing The "number_of_reviews" Column Into 5 Categories

```
def number_of_reviews_categories_function(row):
    """
    Categorizes the "number_of_reviews" column into 5 categories
    """
    if row <= 1:
        return 'very Low'
    elif row <= 5:
        return 'Low'
    elif row <= 10 :
        return 'Medium'
    elif (row <= 30):
        return 'High'
    else:
        return 'very High'
```

By categorizing, we are able to better understand relationships and connections between things and better communicate our findings.

4. Fixing Columns

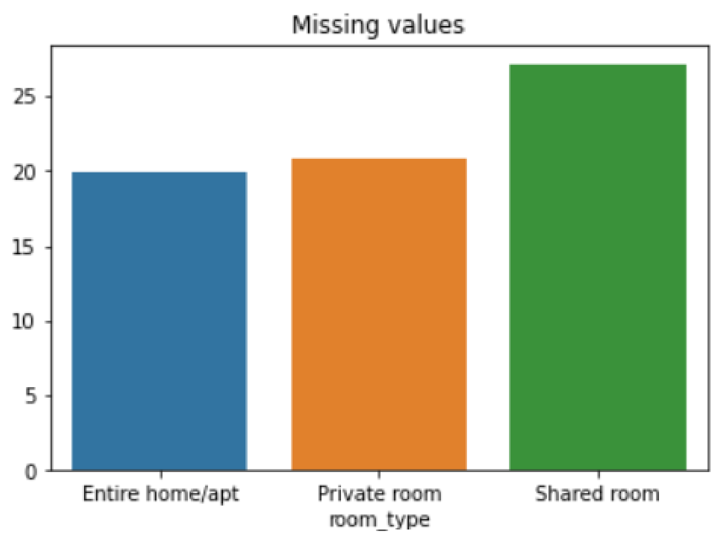
Fix: The **"reviews_per_month"** column is currently of object dtype ; changing it to **"datetime64"** would be more appropriate.

```
airbnb.last_review = pd.to_datetime(airbnb.last_review)
airbnb.last_review
```

```
0        2018-10-19
1        2019-05-21
2               NaT
3        2019-05-07
4        2018-11-19
            ...
48890           NaT
48891           NaT
48892           NaT
48893           NaT
48894           NaT
Name: last_review, Length: 48895, dtype: datetime64[ns]
```

There are no more D-types to be fixed and data does not contain inconsistencies such as shifted columns, which is need to align correctly. The columns necessary for the further analysis are also derived.

5. Missing Values Analysis



Detailed Analysis of Missing Values:

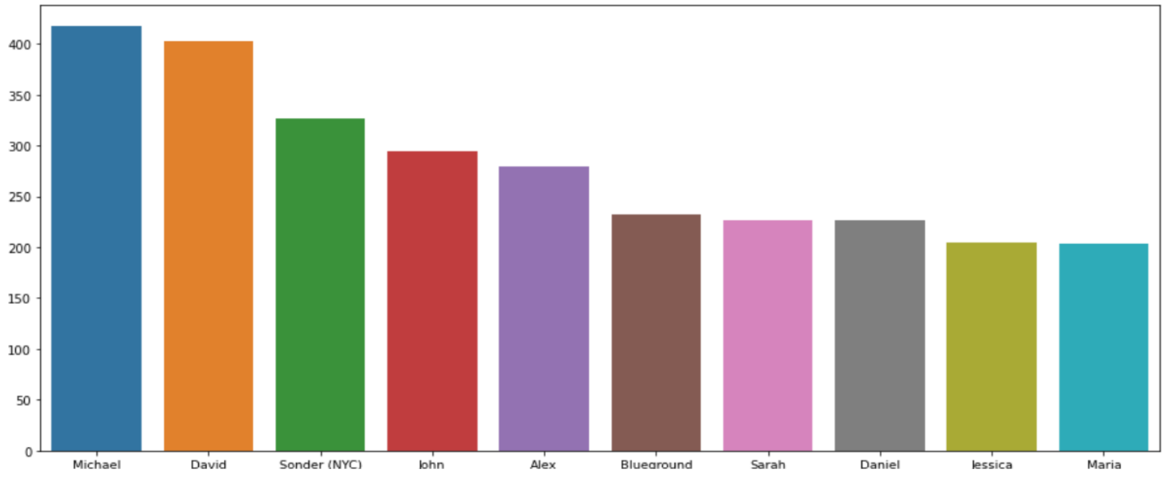
1) Higher Pricing with Missing 'last_review':
Observations indicate that listings without a 'last_review' tend to have higher prices.

2)Lower Review Frequency for Shared Rooms:
Shared rooms are less likely to receive reviews compared to other room types.

3)High Prices Correlate with Fewer Reviews:
Listings with higher prices tend to have fewer reviews, suggesting that guests may be less inclined to leave feedback for more expensive stays.

6. Univariate Analysis

```
# Top 10 host's
plt.figure(figsize=(16,7))
sns.barplot(x = airbnb.host_name.value_counts().index[:10] , y = airbnb.host_name.value_counts().values[:10])
plt.show()
```



7. Bivariate & Multivariate Analysis

Finding the correalations

```
plt.figure(figsize=(10,8))
sns.heatmap(data = airbnb[numerical_columns].corr())
plt.show()
```

' 'availability_365_categories', 'price_categories' and 'reviews_per_month'

```
pd.DataFrame(airbnb.groupby(['availability_365_categories','price_categories']).reviews_per_month.mean())
```