

New York Airbnb Analysis

DETAILED PROJECT REPORT



Shubham Tembhurne

PROJECT DETAIL

Project Title	US Pollution Analysis
Technology	Business Intelligence
Domain	Travel and Hospitality Tech
Project Difficulty level	Advance
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel, Tableau

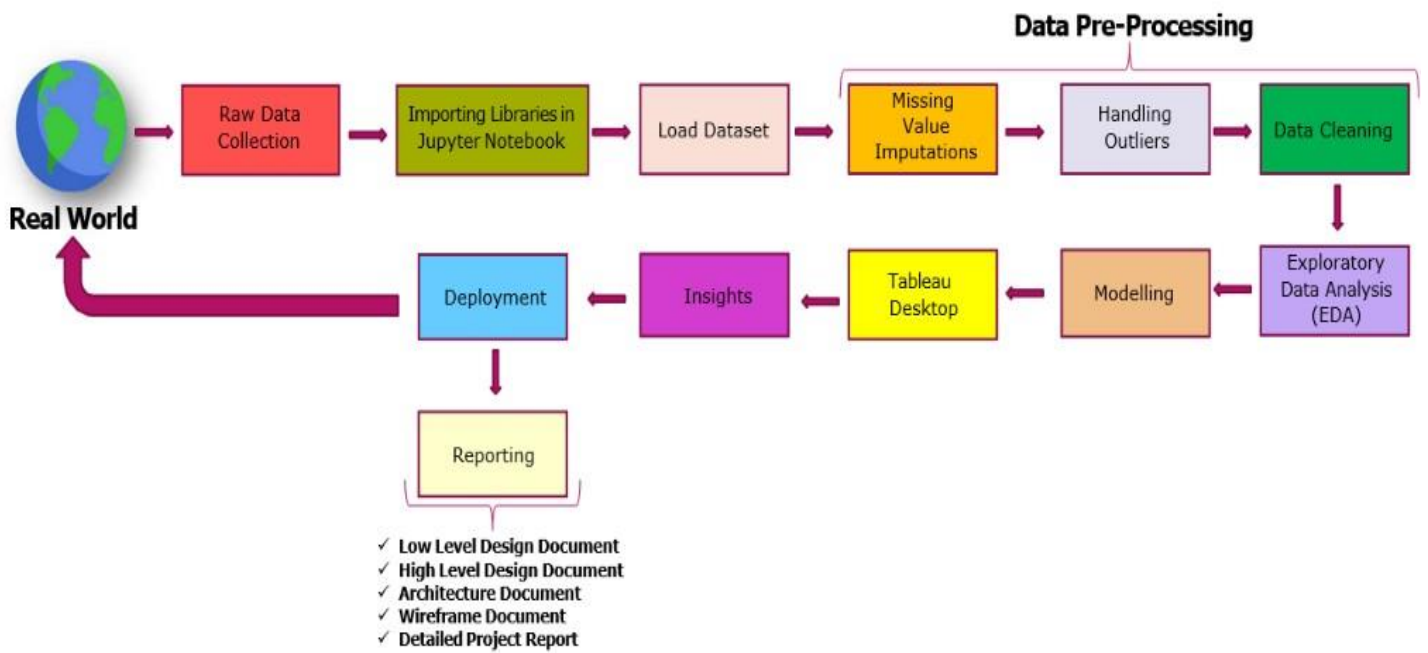
OBJECTIVE

- Airbnb, Inc. is San Francisco, California based company which act as a broker and provides an online marketplace for short-term homestays. The company charges a commission from each booking. Airbnb providing it premium service and experience to customers since 2008. Today Airbnb has millions of listings. These listings generate loss of data. Analyzing this data become crucial factor for the company. This data can be used for business decisions, marketing, implementations of initiatives, additional services and much more.

PROBLEM STATEMENT

- Given dataset include all information about host, listed properties, geographical location, prices reviews and all other required metrics. Analyze the given dataset makes different predictions and draw meaningful conclusion in order to grow the business. Also state what can we learn from different predictions.

Architecture



DATASET INFORMATION

This dataset has round 48895 listings and 16 Columns. It is mix between categorical and numeric values. Given dataset contains null values as blanks well which we have to consider while doing analysis. Last_review and reviews_per_month has more null values. There are 5 neighborhood group in which all listings located. Nearly 80-85% of listings located in Manhattan and Brooklyn. In Manhattan booking price is bit higher as compared to other neighbourhood groups. There are 3 kind of room type (i.e. Shared Room, Private Room, Entire home/Apt). Out of which Shared room are least preferred by the customer even after having less price for booking.

Columns Present in dataset are:

Id: Listing ID of the property

Name: Name of the listed property.

host_id: ID of the property owner.

host_name: Name of the property owner.

neighbourhood_group: Location at which property located.

Neighbourhood: Area in which property located.

Latitude: Latitude coordinate.

Longitude: Longitude coordinate.

room_type: Type of the room (Entire Home/ Appt, Private Room, Shared Room)

Price: Price in Dollars.

Minimum_nights: Amounts of minimum night stay at property

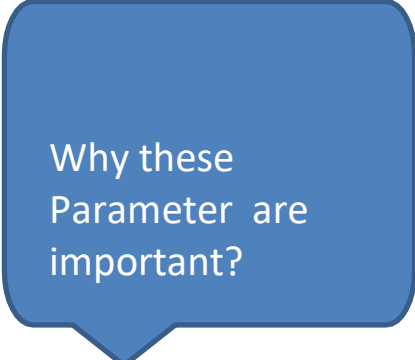
number_of_reviews: No. Of reviews

last_review: last review on which date.

reviews_per_month: Numbers of reviews per months.

calculated_host_listings_count: Count of properties listed on that host.

availability_365: Number of days when listing is available for booking



Why these
Parameter are
important?

1.Location Metadata

Latitude, Longitude: Help uniquely identify the monitoring location.

Id, Name, host_id, host_name, neighbourhood_group, Neighbourhood:
Provide human-readable location information for spatial analysis, policy targeting, and regional trends.

Importance: These are crucial for geospatial analysis and identifying pollution hotspots.

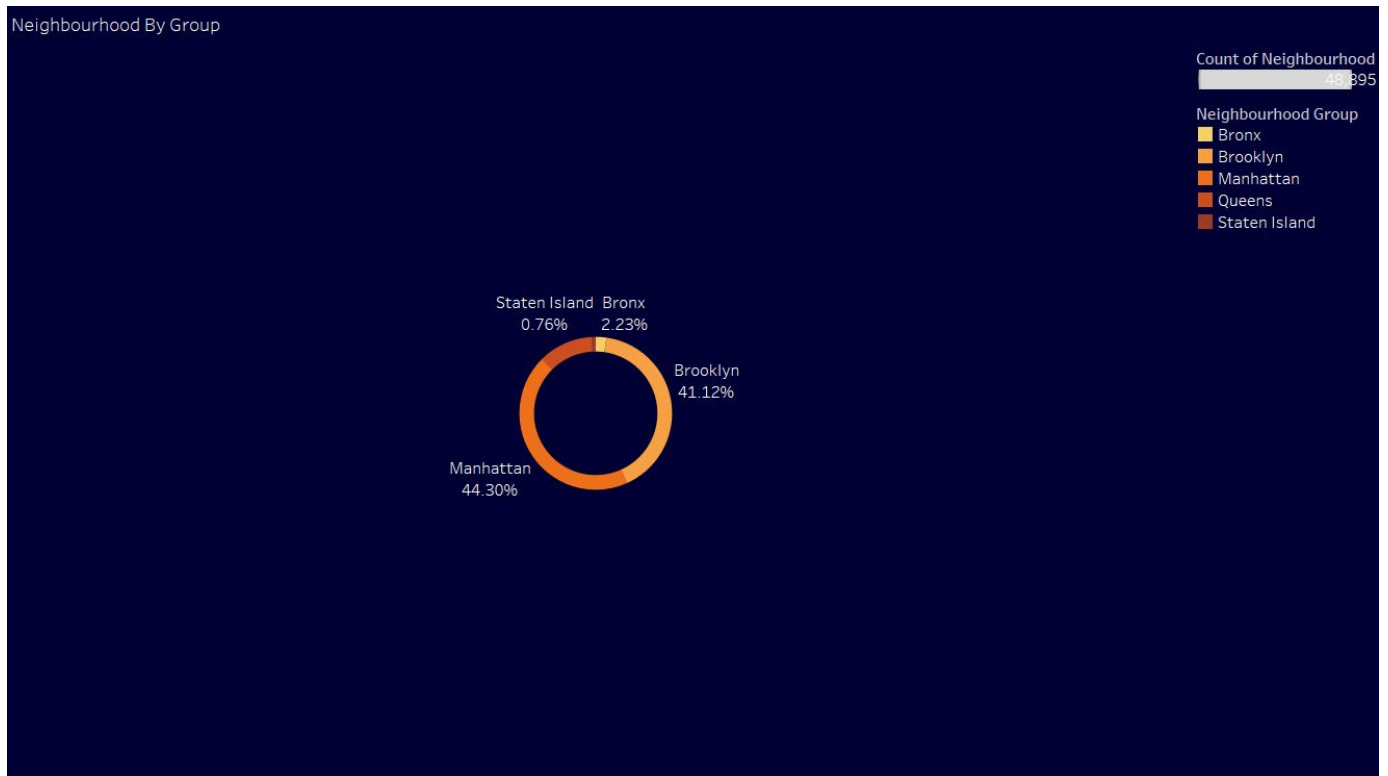
2.Date

last_review: Specifies the date of observation.

Importance: Enables temporal analysis to detect trends, seasonal patterns, or year-over-year changes in reviews, booking, number of customers.

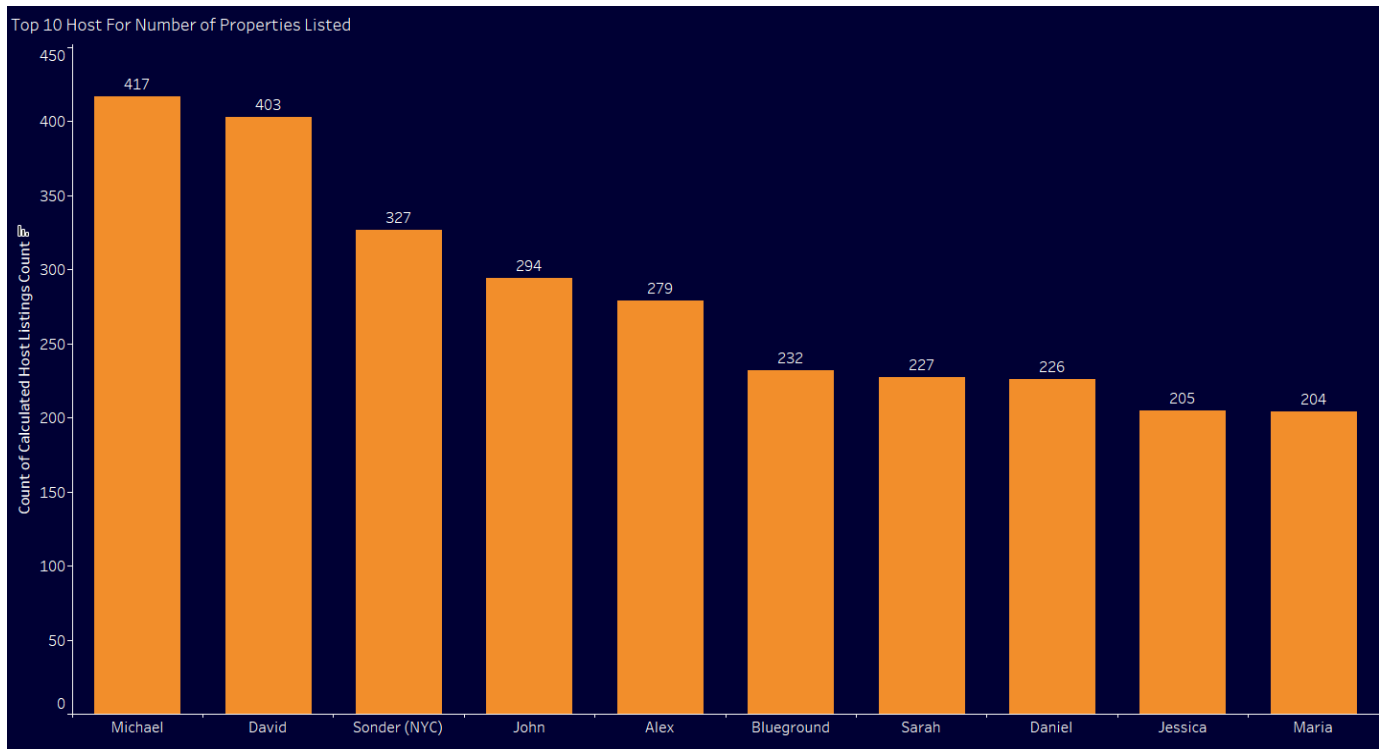
INSIGHTS

1. In which Neighborhood group there is maximum number of properties listed?



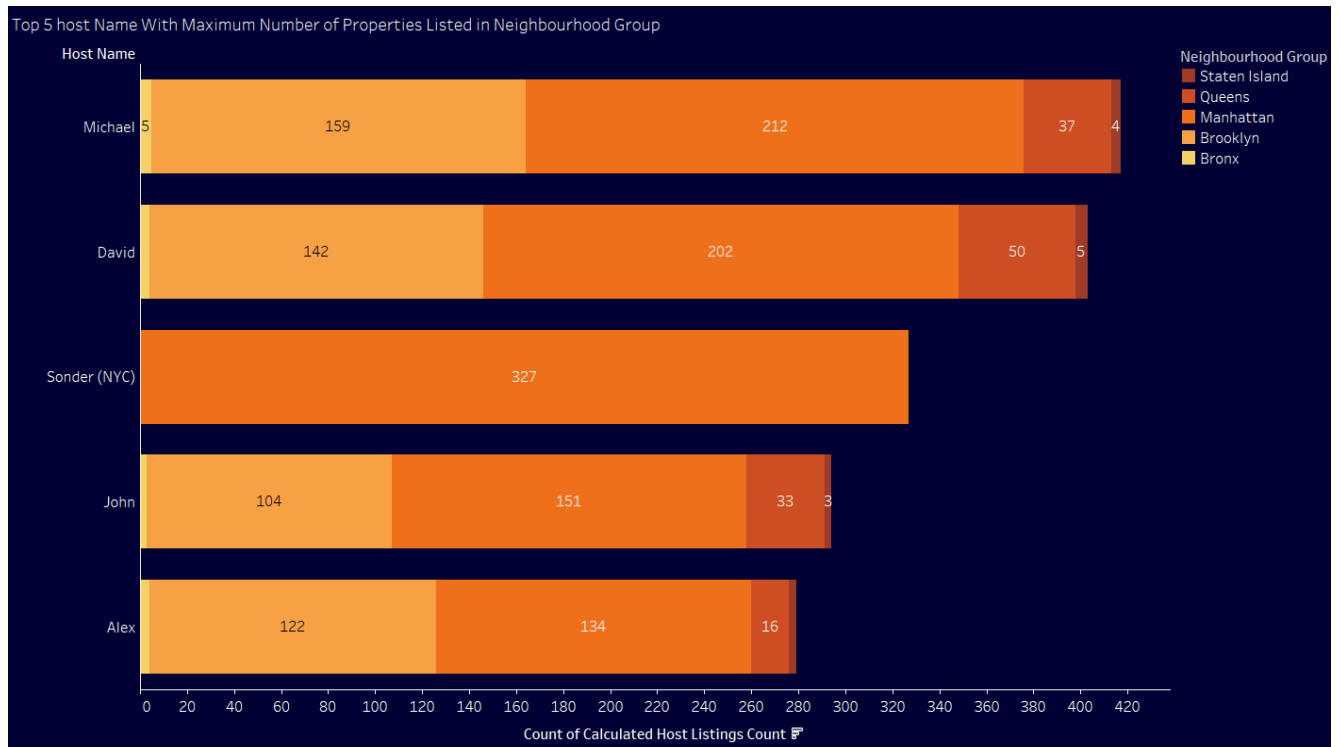
- The maximum number of properties in neighborhood group is Manhattan with 44.40% of property count.

2. Which host has maximum number of properties listed?



- The host has maximum number of properties listed are Michael, David, Songer (NYC), John, Alex, Blueground, Sarah, Daniel, Jessica, Mari

3. Which host has maximum properties listed in neighborhood groups having maximum properties listed?



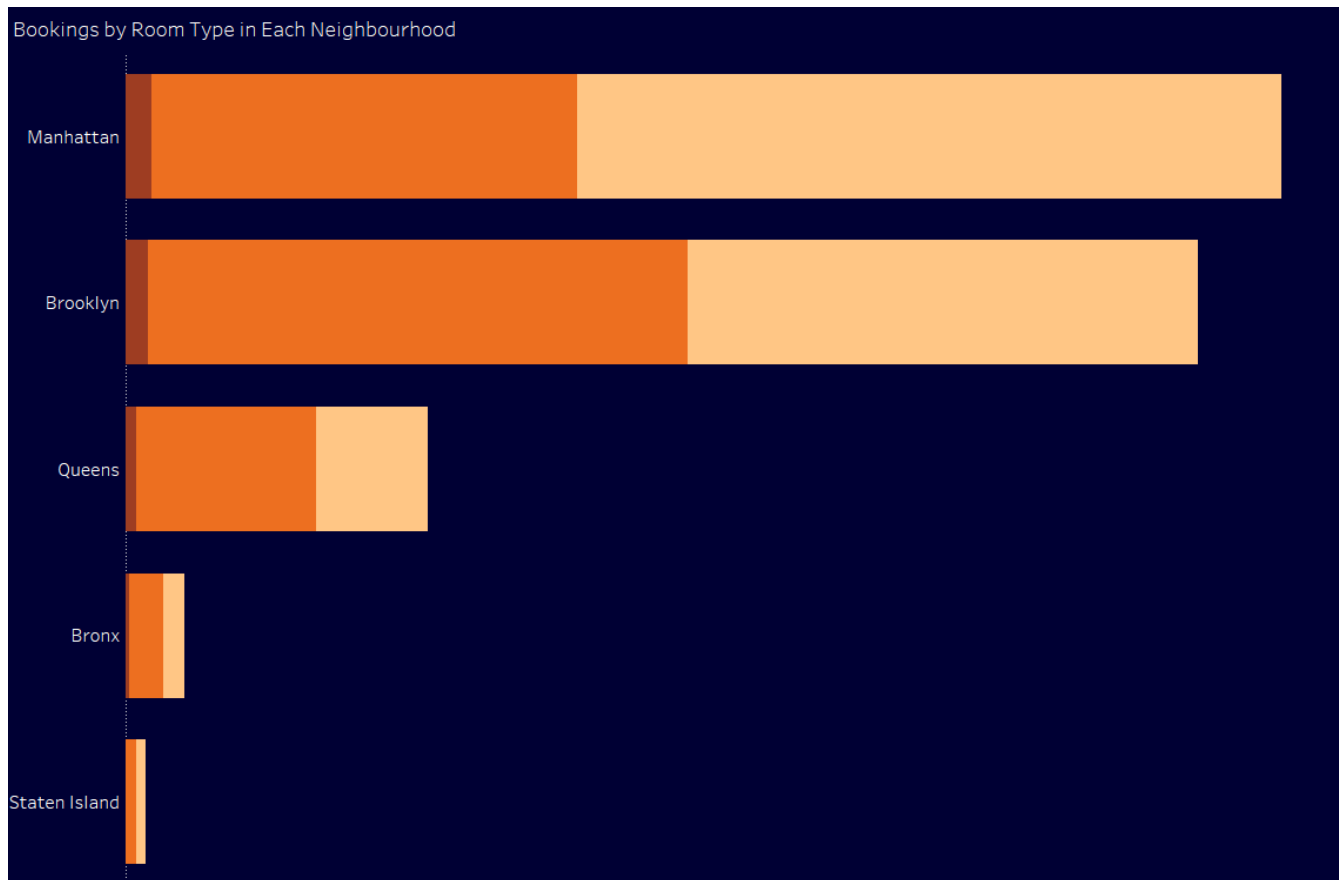
- The host which has the maximum properties listed is Sonder (NYC) in Manhattan.

4. What is the average price in different properties listed?

Average Price by Room Type	
Room Type	
Entire home/apt	\$211.8
Private room	\$89.8
Shared room	\$70.1

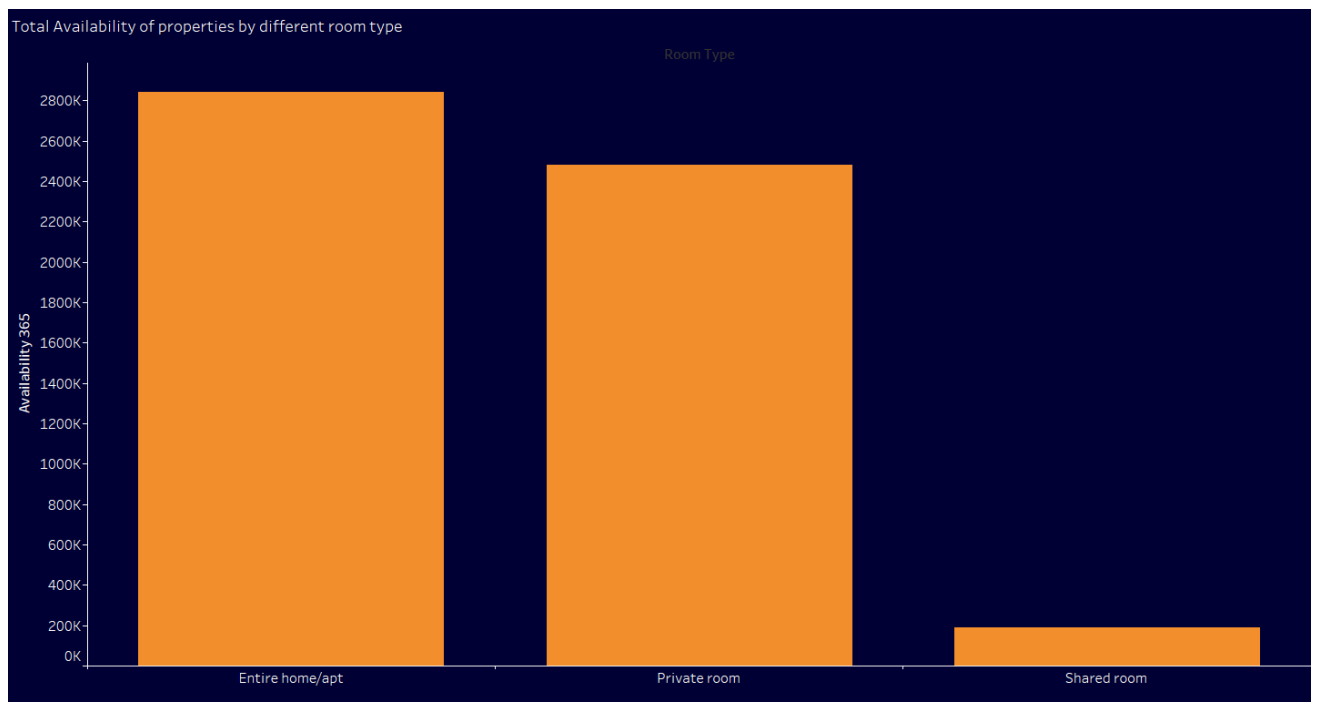
- The Average price for Entire home/apt is \$211.8, Private room - \$89.8 and \$70.1
- Entire home/apt is most expensive on average.

5. What is the most preferred room type in every neighborhood groups?



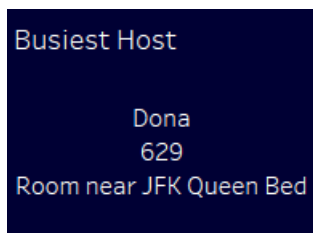
- Entire home/apt is most common in Manhattan and Brooklyn, while private rooms are also popular.

6. Total availability of properties having different room type?



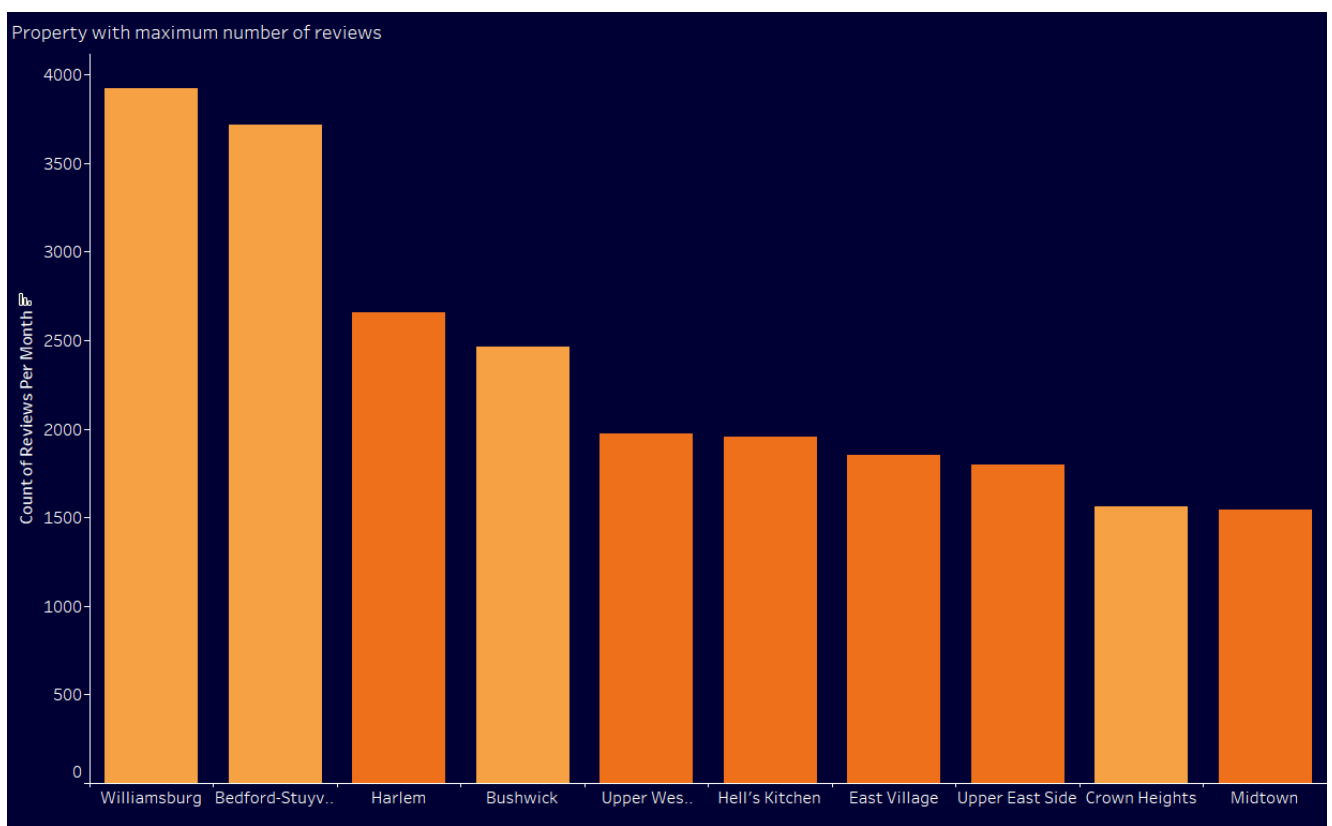
- The Entire home/pt have the higher total availability.

7. Which one is the busiest host?



- Dona's property "Room near JFK Queen Bed" has the most reviews.

8. Which property has maximum number of reviews?



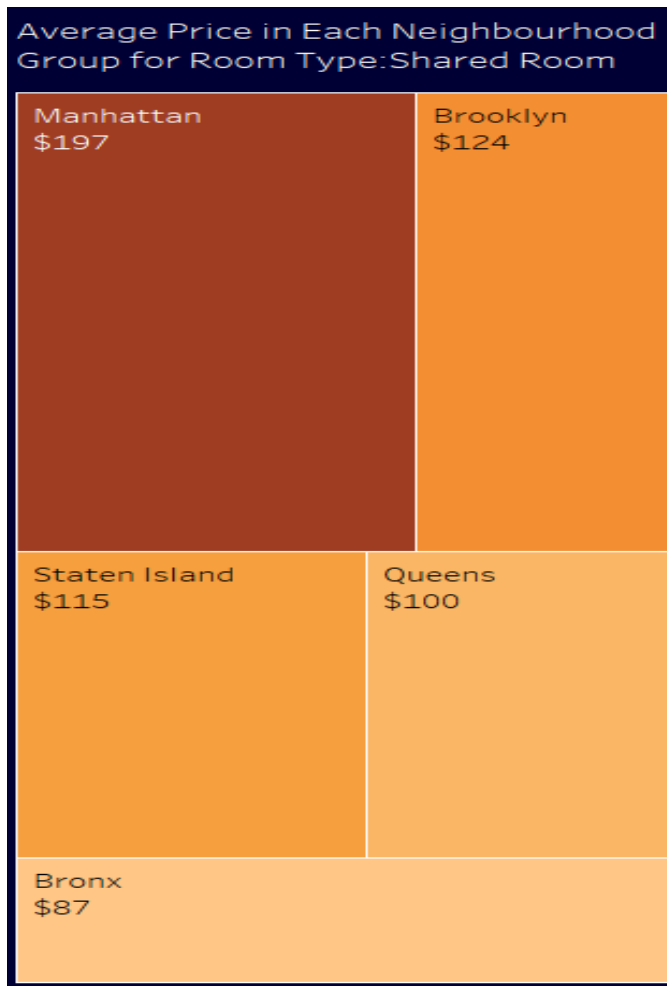
- Williamsburg in Brooklyn have the highest number of reviews

9. What are the average reviews per month by room type and neighborhood group?

Average Reviews per Month by Room Type And Neighbourhood Group					
	Bronx	Brooklyn	Manhattan	Queens	Staten Isl..
Entire home/apt	1.9577	1.2304	1.0384	1.7566	1.8743
Private room	1.4607	1.1427	1.2962	1.6960	1.5481
Shared room	1.1728	0.9664	1.5156	1.3364	1.0600

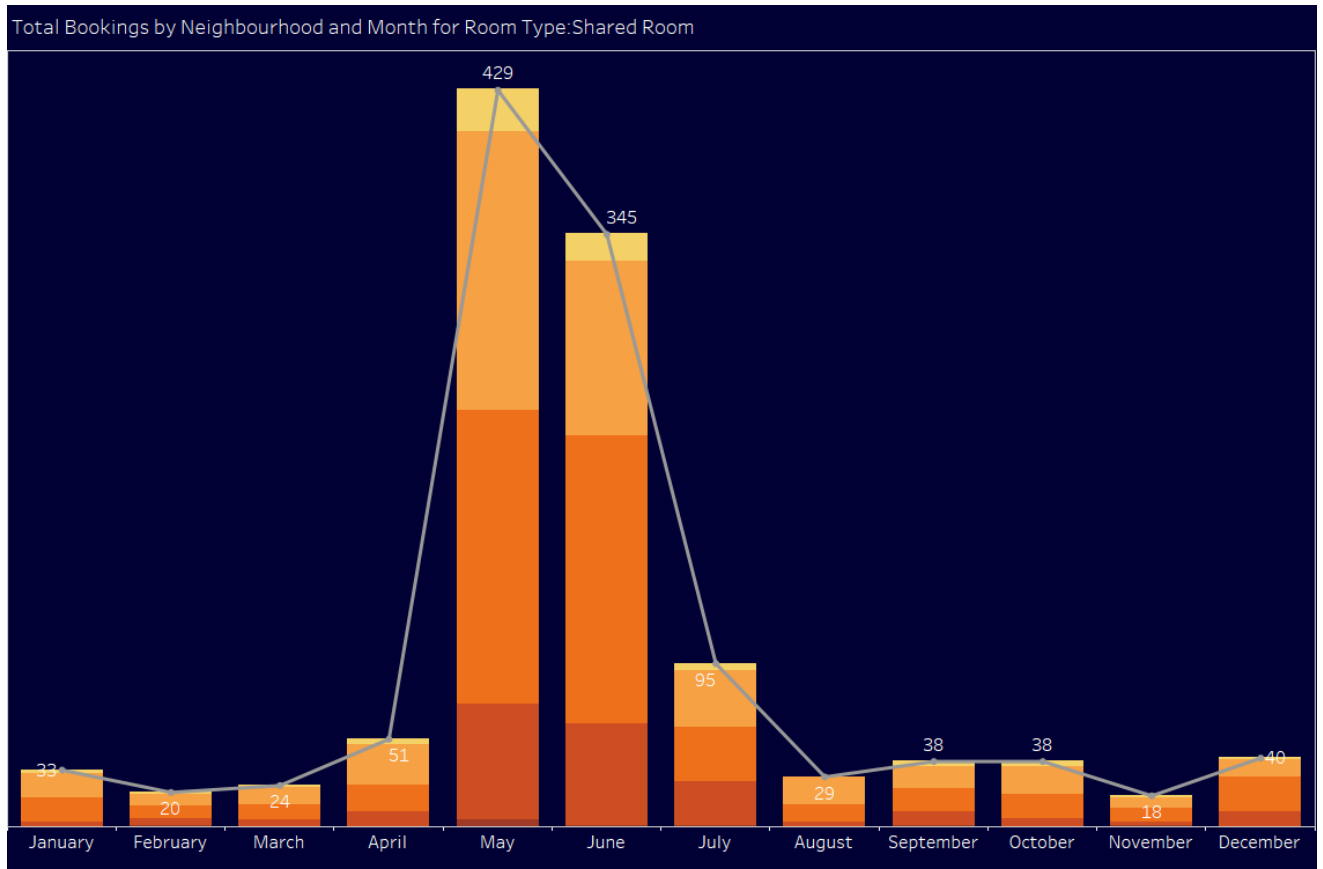
- We can see almost 1 or 2 reviews per month in different room type and with respect to neighborhood group

10. What is the average price for each neighborhood group for room type shared room?



- The Average price starts from Manhattan (\$197) highest in all and Brooklyn (\$124), Staten Island (\$115), Queens (\$100), Bronx (\$87) in each neighborhood group for room type shared room.

11.What are the total bookings by neighborhood and month for room type shared room?



- We have the highest number of bookings in the month of May and June for room type shared room.
- Manhattan have maximum booking in May and June month, 171 in May and 167 in June.
- Staten island have the least booking in may month ,4 and none in June.

12.What are the top 10 host by reviews?

Top 10 Host by Reviews			
Host Name	Neighbourhood	Avg. Price	
Arnie	Sunset Park	\$35	1
Catherine	Long Island City	\$90	21
Eric	Clinton Hill	\$69	1
Jennifer	Midtown	\$162	46
John	Bensonhurst	\$76	55
	Gravesend	\$114	41
	Kensington	\$149	9
Letha M.	Bedford-Stuyvesant	\$60	2
Loli	Bedford-Stuyvesant	\$59	390
Nataraj	Fresh Meadows	\$59	4
Tasos	Williamsburg	\$95	1
Teedo	Bedford-Stuyvesant	\$182	27

- Arnie, Catherine, Eric, Jennifer, John, Letha M, Loli, Nataraj, Tasos, Teedo are the top host with maximum reviews.

KEY PERFORMANCE INDICATOR(KPI)

1. Maximum number of properties listed in neighborhood groups
2. Host with Maximum number of properties listed.
3. Host with Maximum number of properties listed having maximum properties listed.
4. Average price in different properties listed.
5. Most preferred room type in every neighborhood groups.
6. Total availability of properties having different room type.
7. Properties with maximum number of reviews.
8. Average review per month by room type and neighborhood group.
9. Average price for each neighborhood group for room type shared room.
10. Total bookings by neighborhood and month for room type shared room.

CONCLUSION

- Manhattan dominates the market with 44.4% of total listings, making it the most active neighborhood group.
- It also records the highest average price (\$197) for shared rooms and peak bookings in May and June, confirming it as a high-demand, high-value area.
- Williamsburg (Brooklyn) stands out for having the most reviews, signaling high guest engagement and popularity in that sub-region
- Sonder (NYC) is the top-performing host in terms of property count, especially in Manhattan, while Dona's property has the most reviews, showing strong guest interest and trust.
- Other top-reviewed hosts (Arnie, Catherine, Eric, etc.) indicate potential brand ambassadors or candidates for host training programs.
- Entire home/apartment is both the most expensive (\$211.8 average) and widely available, especially in Manhattan and Brooklyn.
- Private rooms (\$89.8) and shared rooms (\$70.1) offer budget-friendly alternatives and also show steady booking patterns.
- May and June are the peak booking months, especially for shared rooms, highlighting a clear seasonal demand.
- Consistent review patterns (1–2 reviews per month) across listings and neighborhoods help customers judge reliability and performance.
- The data aids in identifying affordable yet well reviewed options ideal for new travelers unsure where to stay.
- Areas like Staten Island show low booking activity only 4 bookings in May and none in June, which identifies them as potential growth opportunities with the right incentives.
- A well-informed customer base is more likely to book confidently, increasing platform credibility and bookings.

Q & A

Q1) What's the source of data?

Ans) The Dataset was taken from Kaggle

Q2) What was the type of data?

Ans) The data was a combination of numerical and Categorical values.

Q3) What's the complete flow you followed in this Project?

Ans) Refer slide 5th for better Understanding

Q4) What techniques were you using for data?

Ans) -Removing unwanted attributes

-Visualizing relation of independent variables with each other and output variables

-Removing outliers

-Cleaning data and imputing if null values are present.

-Converting Numerical data into Categorical values.

Q6) What were the libraries that you used in Python?

Ans) I used Pandas, NumPy, Matplotlib,
and Seaborn libraries in Pandas.

THANK YOU