Exploratory Data Analysis On Airbnb Bookings Analysis

TEAM NAME : New World
Team Members
1. Agam Singh 2.Jyoti Patel

ABOUT AIRBNB

- Airbnb (ABNB) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.Airbnb offers people an easy, relatively stress-free way to earn some income from their property.
- The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia.
- AIRBNB is an American San Francisco-based company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities, the booking platform is accessible via website and mobile app.
- The company acts as a broker and charges a commission from each booking, it does not own any of the listed properties.

CONTENTS

- ❖ Data summary

- ❖ Dataset cleaning

- ❖ EDA on dataset

- ❖ Conclusion

PROBLEM STATEMENT

1.What is the correlation between different variables ?

2.What is the distribution of Airbnb listings by borough in New York City?

3.Which hosts are the busiest and why?

4.Which Hosts are charging higher price?

5.what are the most popular types of properties available on Airbnb in New York City?

6.What can we learn from Data? (ex: locations, prices, reviews, etc)

7.What are the traffic difference among different areas and what could be the reason for it?

8.What is the average nightly price of Airbnb listings in New York City, and how does it vary by borough?

STEPS INVOLVED IN EDA

1) Exploratory Data Analysis
2) Null values Treatment
3) Data Exploration
4) Data Visualization
5) Standardization of features

DATA SUMMARY

Name of the Dataset————————————AIRBNB Booking Analysis

Number of variables/Columns ——————16

Number of observations/Row——————48895

 Duplicate rows ——————————————0 (0.0%)

Total size in memory——————————— 6.0 MB

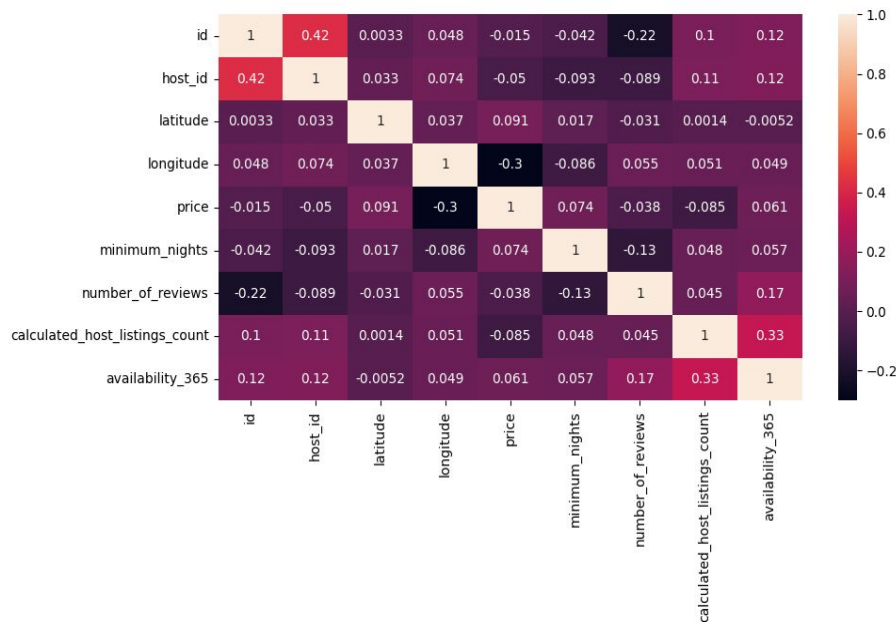## VARIABLE DATA TYPE

Data Type	Columns
Numeric - int64	0 id 2 host_id 9 price 10 minimum_nights 11 number_of_reviews 12 calculated_host_listings_count 13 availability_365
Numeric – float64	6 latitude 7 longitude
String - object	1 name 3 host_name 4 neighbourhood_group 5 neighbourhood 8 room_type

CLEANING DATASET

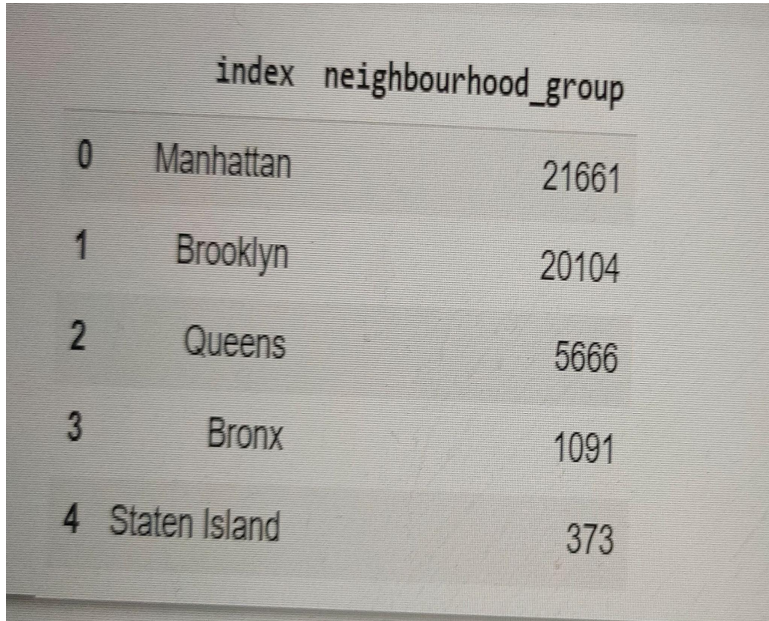| Columns	Total	Percent
last_review	10052	0.205583
reviews_per_month	10052	0.205583
host_name	21	0.000429
name	16	0.000327

To handle missing data, variables with missing data has been dropped, except the variable 'host_name' and 'name'.

➢ **What is the correlation between different variables ?**

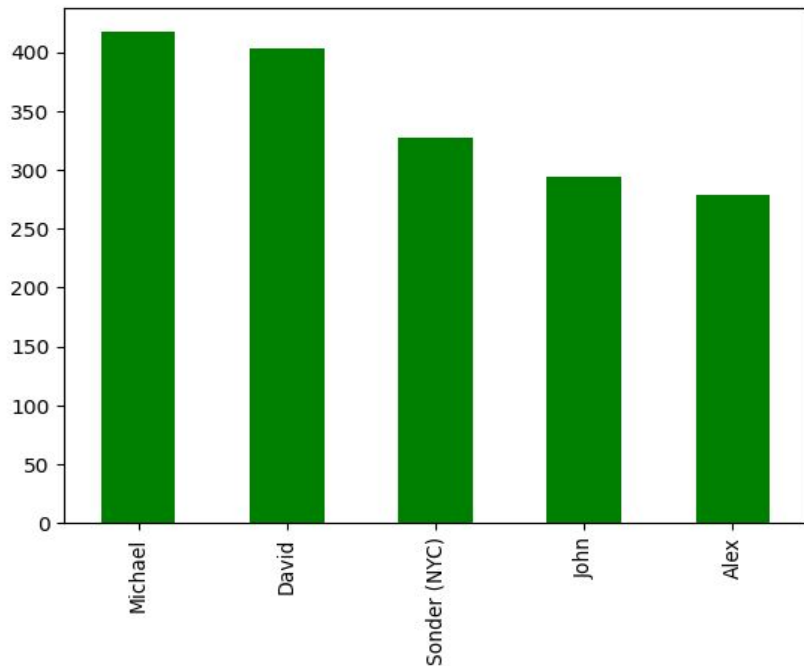*This data shows the correlations between different variables of the NYC .*

➢ **Check the distribution of Airbnb listings by borough in New York City.**

index	neighbourhood_group	
0	Manhattan	21661
1	Brooklyn	20104
2	Queens	5666
3	Bronx	1091
4	Staten Island	373

This data shows the distribution of Airbnb listings by borough in New York City .From this data we find that most of Airbnb listing by borough in NYC is in Manhattan ,followed by Brooklyn,Queens,Bronx and Staten Island.

➢ **Which hosts are the busiest and why?**

After analyzing the data we found Michael,David etc.are the busiest host as we can see from above graph.

Summary of the main variable('Price')

count 48895.000000
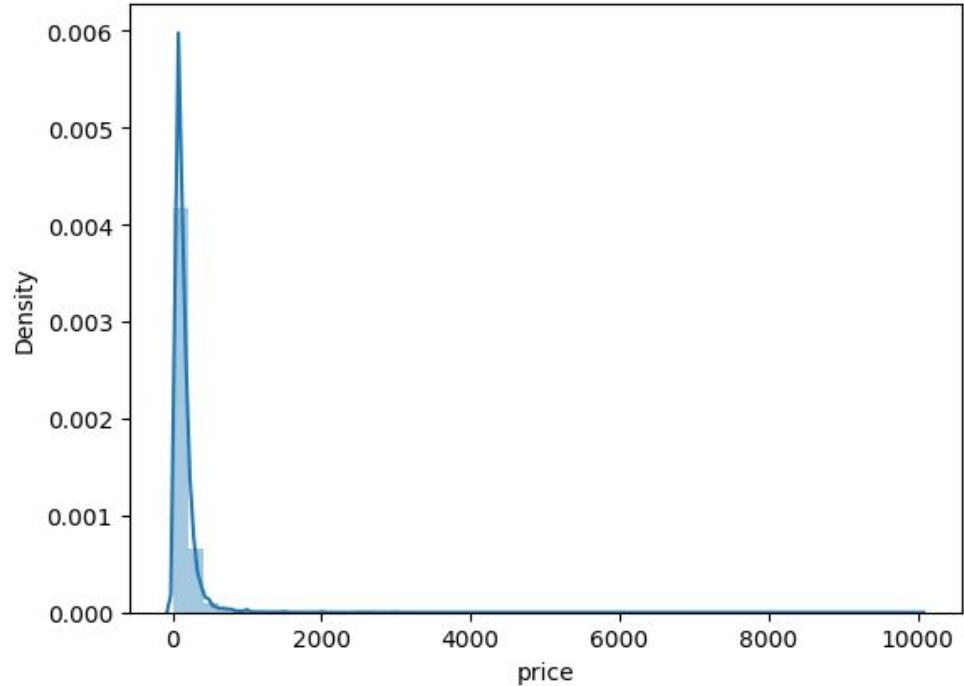
mean 152.720687

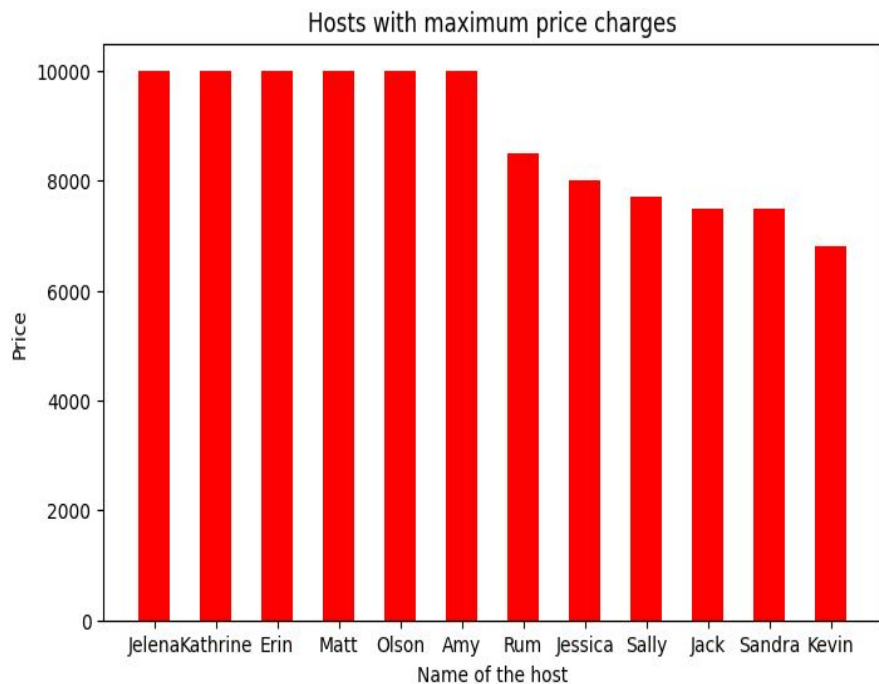std 240.154170

min 10.000000

25% 69.000000

50% 106.000000

75% 175.000000

max 10000.000000

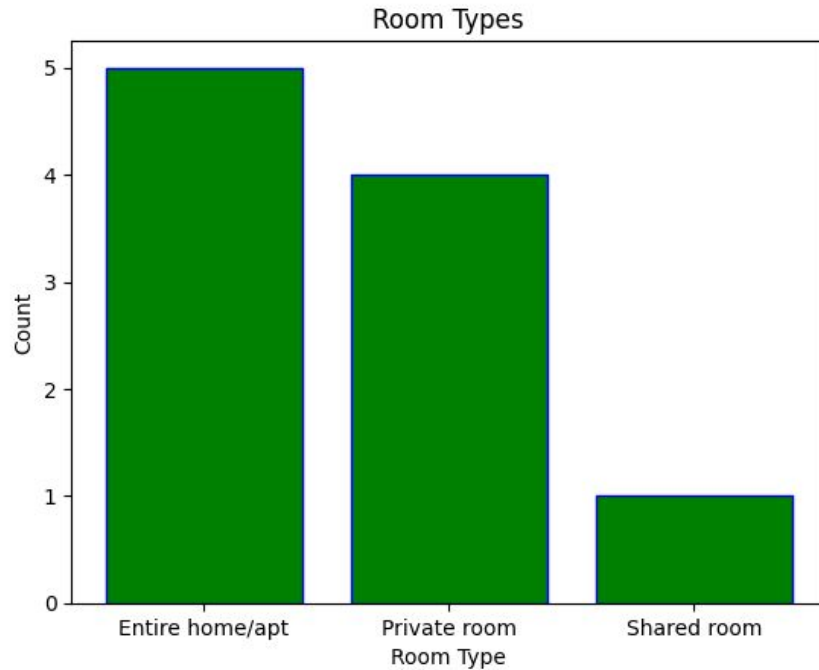This data shows that more bookings are done at the price range of 10 - 500

➢ **Which Hosts are charging higher price?**

Hosts with maximum price charges

Hear graph represent top 12 Hosts who are charging maximum price. (1.Jelena Kathrin,2.Erin,3.Matt,4.Olson,5.Rum, 6.Jessica,7.Sally,8.Jack,9.Sandra,10. Kevin)
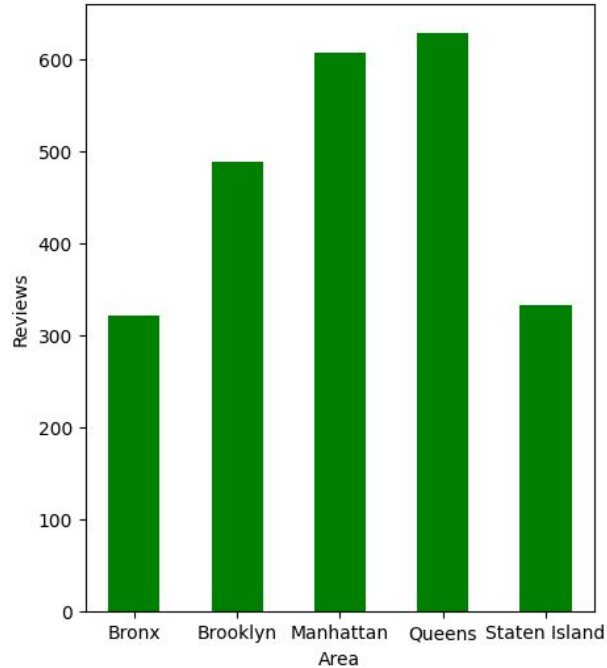
● Maximum Price is 10000 USD by Jelena Katherine

➢ **What are the most popular types of properties available on Airbnb in New York City?**

Entire home/apt is the highest number of room types overall and prices are high in the brooklyn and Manhattan for entire home/apt.
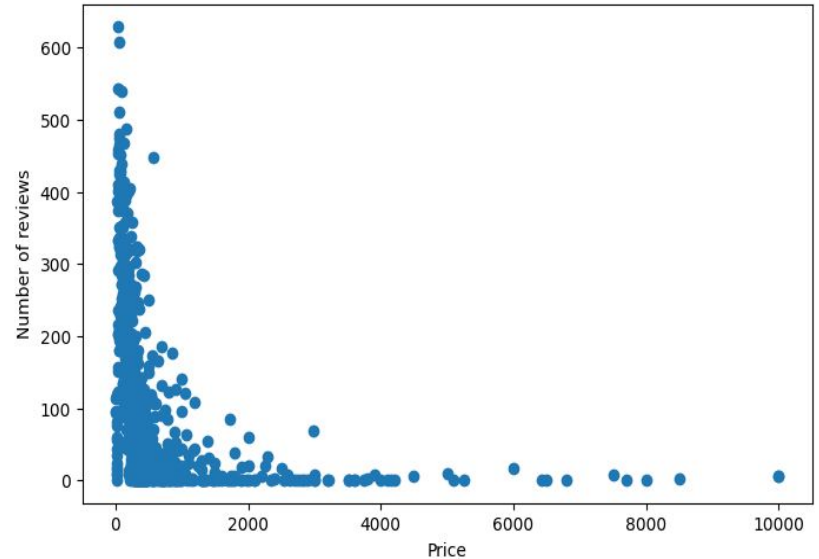
➢ **What can we learn from Data? (ex: locations, prices, reviews, etc)**

Number of Reviews in terms of area

From this visualization we can say that most number of people like to stay in less price and their reviews are higher in those areas.

Number of Reviews VS Price

Conclusion

Throughout this dataset, based on the findings of the EDA project on Airbnb booking analysis, we can conclude that the Sonder(NYC) and Blueground hosts have listed the highest number of listings in Manhattan. The most common room type preferred by visitors is Entire home/apt, and prices are relatively high in Brooklyn and Manhattan for this room type. Visitors also tend to prefer staying in rooms with lower prices and higher reviews. The busiest hosts are those who list their room types as Entire home and Private room, and their reviews are generally higher. The top 10 hosts charging the maximum price have properties located in Manhattan, Brooklyn, and Queens. Manhattan has the highest number of Entire home/apt listings, while Brooklyn, Queens, and Bronx have more private room types. This analysis can help inform future strategies for hosts and guests to optimize their Airbnb experiences.