CAPSTONE PROJECT
EDA on Airbnb Dataset

CONTRIBUTOR:
- Yogesh K

CONTENTS

- **Objective**
- **Problem Statement**
- **Data Summary**
- **Data Cleaning**
- **Data Analysis, Visualization & Insights**
- **Challenges Faced**
- **Conclusion**

AI

OBJECTIVE

To perform data-driven research and harness insights that will help in realizing trends and increasing the market value and revenue of Airbnb.

PROBLEM STATEMENT

Derive key business insights using Airbnb's NYC data to boost business decisions that will enhance profitability, market expansion and consumer experience and extract information about traffic and prices in different areas.

DATA SUMMARY

- Airbnb dataset : ~49000 observations , 16 features

- Numeric features : 8 , Categorical features : 2

- Data is specific to New York and nearby suburbs

- Data spanning across 2017 through 2019

- Important features : host_id ,neighbourhood_group, neighbourhood, room_type,   price, minimum_nights, availability_365 , reviews_per_month

DATA CLEANING

- last_review = 10052 null

- reviews_per_months = 10052 null

- replaced null for reviews_per_month with 0

- no changes in last_review

- No changes in calculated_host_listing_counts

Null vs Not Null percentage

NULL
2.6%

97.4%
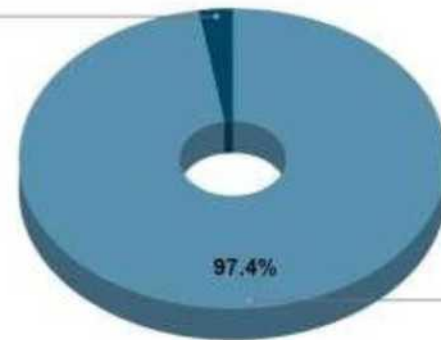
Not Null
97.4%

NUMERICAL FEATURES:

| price	This feature represents the price of a particular listing.
minimum_nights	This represents the minimum number of nights spent by a person in a listed Airbnb
availability_365	Shows availability of the listed Airbnbs in days
host_id	represents a unique id corresponding to a host
latitude, longitude	latitude and longitudinal values of listed Airbnbs
reviews_per_month	This represents total reviews received per month
number_of_reviews	Total number of reviews received till date

AI

DATA ANALYSIS

AI

CATEGORICAL FEATURES:

| neighbourhood_groups	room_type	minimum_nights_category
•Manhattan •Brooklyn •Queens •Staten Island •Bronx	•Private room •Entire home/apt •Shared room	•1 day •2-3 days •few days •1 week •less than a month •1 month •more than a month •more than half a year •less than a year

MOST NUMBER OF AIRBNB LISTINGS:

- **5 unique neighborhood groups**

- **Manhattan and Brooklyn have the most listed properties**

- **Manhattan and Brooklyn belong to New York's city center**
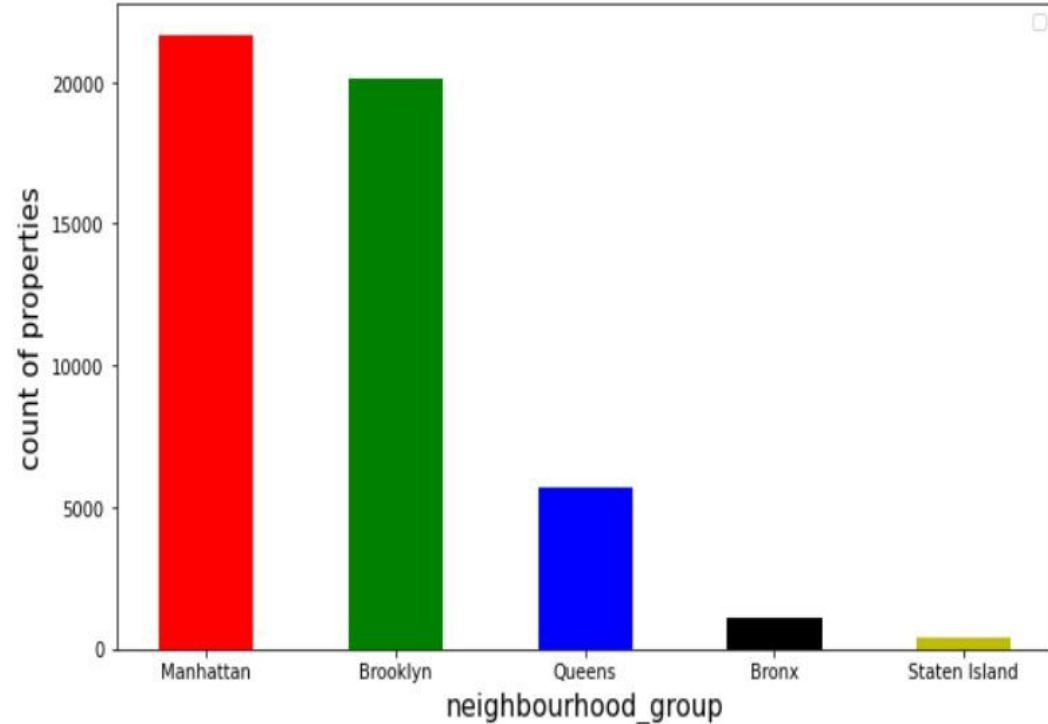
HOST LISTINGS IN AIRBNBS ACCORDING TO ROOM TYPE :

- **Maximum listing room types: Entire home/apt , private room types**

- **Manhattan and Brooklyn have the most number of listings**

- **Very few listings in Queens, Staten Island and Bronx , but they also follow the same trend**
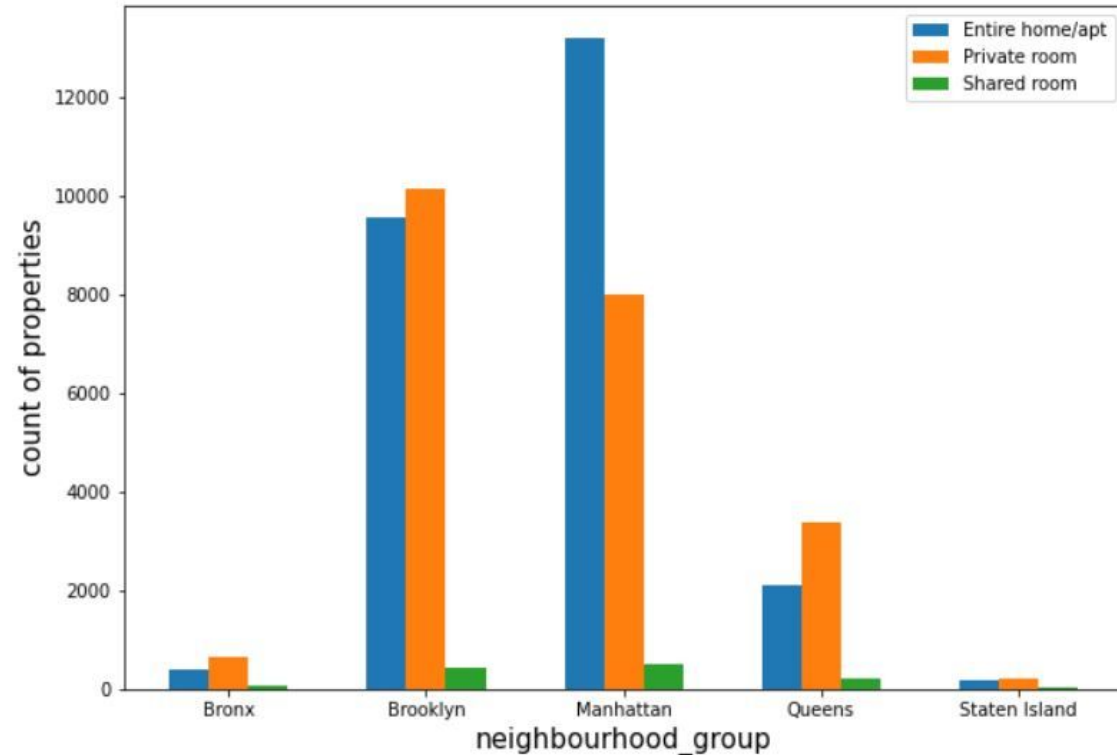
RATES OF AIRBNB LISTINGS ACCORDING TO ROOM TYPES OVER DIFFERENT NEIGHBORHOOD GROUPS:

- **Rate of Entire home/apt room type is the highest in all the neighborhood groups**

- **Entire home/apt room types have the maximum listed price**

- **This makes sense because an entire home will cost more.**
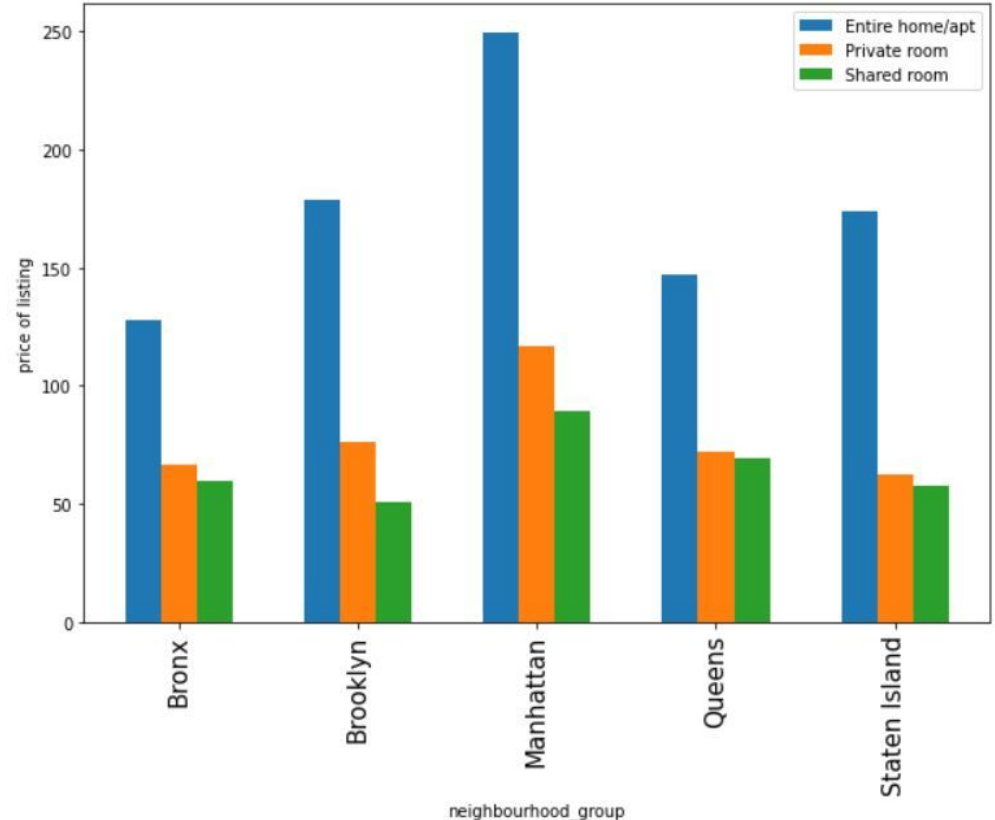
PROPERTIES UTILIZED VS MONTHS IN A YEAR:

- Most footfall of customers is during May, June, and July

- A spike in June due to summer vacation

- Marketing team can propose discounts in off season months to drive consumer engagement, and hence profitability
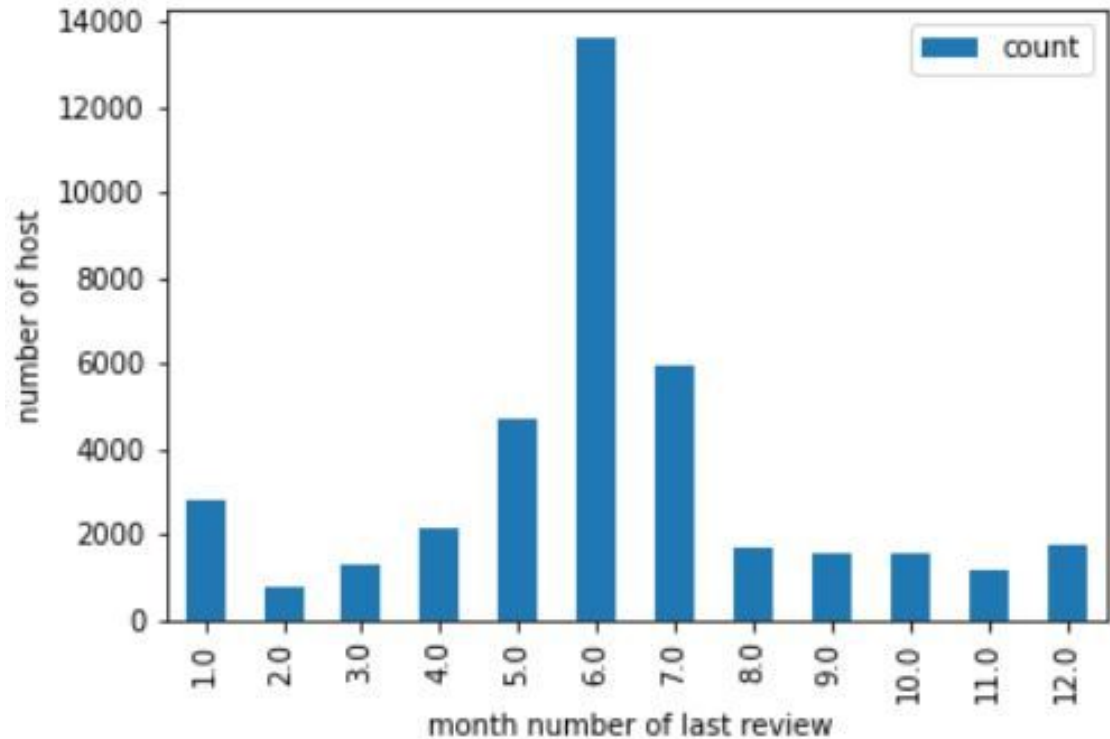
MINIMUM NUMBER OF NIGHTS SPENT BY CUSTOMERS IN AN AIRBNB:

- Maximum people book Airbnbs between 1 to 3 days

- A reason for this could be tourists and people visiting New York City for business trips.
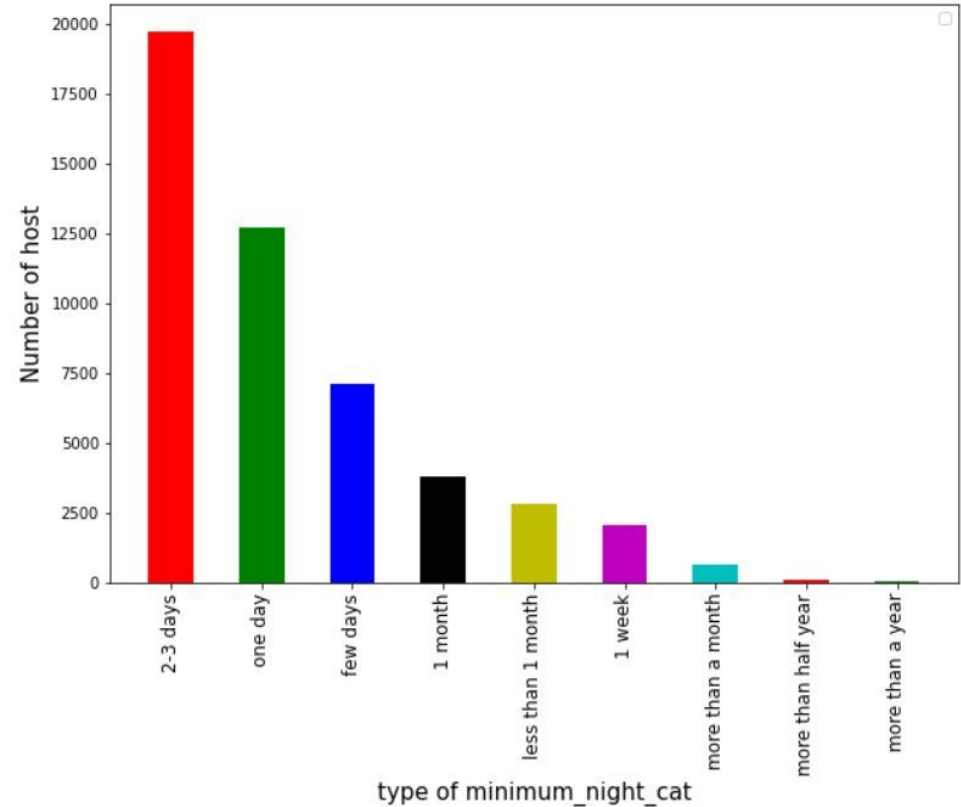
IDENTIFYING GAPS IN SUPPLY AND DEMAND:

- **Availability is less, implying they have full occupancy**

- **Onboarding of more Airbnbs to increase supply as the demand in these locations surpass the supply**

- **Conversion of shared rooms to other room (refer previous plot)**

	index	neighbourhood_group	neighbourhood	room_type	availability_365	host_id	price
0	295	Manhattan	Morningside Heights	Entire home/apt	0.0	150	130.0
1	290	Manhattan	Marble Hill	Entire home/apt	0.0	5	90.0
2	282	Manhattan	Kips Bay	Private room	0.0	112	99.0
3	33	Bronx	Fieldston	Private room	0.0	7	64.0
4	157	Brooklyn	Cobble Hill	Private room	0.0	22	81.0
5	270	Manhattan	Greenwich Village	Private room	0.0	99	110.0
6	268	Manhattan	Gramercy	Shared room	0.0	5	69.0
7	171	Brooklyn	Downtown Brooklyn	Entire home/apt	0.0	52	161.0
8	260	Manhattan	East Village	Shared room	0.0	23	75.0
9	259	Manhattan	East Village	Private room	0.0	660	95.0
10	172	Brooklyn	Downtown Brooklyn	Private room	0.0	31	85.0
11	201	Brooklyn	Greenpoint	Private room	0.0	476	75.0
12	212	Brooklyn	Navy Yard	Entire home/apt	0.0	10	150.0
13	245	Manhattan	Battery Park City	Private room	0.0	21	100.0
14	244	Manhattan	Battery Park City	Entire home/apt	0.0	48	225.0
15	240	Brooklyn	Williamsburg	Private room	0.0	1997	75.0
16	296	Manhattan	Morningside Heights	Private room	0.0	192	75.0

INTRODUCING SOME DISCOUNTS IN LOW DEMAND AREAS:

- **Occupancy in these locations are relatively less**

- **Discounts can be given to lure customers to these locations**

	index	neighbourhood_group	neighbourhood	room_type	availability_365	host_name	price
0	178	Brooklyn	East Flatbush	Shared room	365.0	24	34.5
1	444	Queens	South Ozone Park	Shared room	365.0	8	30.0
2	28	Bronx	Eastchester	Private room	364.0	6	68.0
3	437	Queens	Ridgewood	Shared room	361.5	10	31.5
4	27	Bronx	Eastchester	Entire home/apt	358.5	6	202.5
5	137	Brooklyn	Borough Park	Shared room	357.0	16	35.0
6	511	Staten Island	Randall Manor	Private room	355.0	9	55.0
7	362	Queens	Corona	Shared room	354.0	24	27.5
8	507	Staten Island	Port Richmond	Private room	353.0	6	48.5
9	17	Bronx	Clason Point	Private room	349.0	11	70.0
10	452	Queens	Sunnyside	Shared room	343.0	25	30.0
11	140	Brooklyn	Brighton Beach	Shared room	343.0	7	50.0
12	350	Queens	Bellerose	Private room	342.5	10	70.0
13	534	Staten Island	West Brighton	Entire home/apt	342.0	5	99.0
14	510	Staten Island	Randall Manor	Entire home/apt	340.0	9	109.0
15	527	Staten Island	Stapleton	Private room	340.0	17	55.0

CHALLENGES FACED

- Interpretation of all the features and deciding which ones will be helpful in our analysis.

- Understanding the dataset and deciding what features we can convert into categorical features, apart from the existing categorical features.

- Handling NaN values for last_review feature.

- Filtering out important trends and visualizations from the unimportant ones.

- Choosing the type of visualization to convey data in a better way

CONCLUSION

- Manhattan and Brooklyn are the most crowded suburbs

- Listing rates across all the 5 unique neighborhood groups follow a similar trend for all room types. (Refer Slide 10)

- Customers prefer to stay in Entire home/apt or private room types

- Identifying supply demand gap to boost business decisions

- Highest footfall of customers is in the months of May, June and July

- Converting shared room type Airbnbs in Manhattan and Brooklyn to other room types will improve revenue generation.

- Onboarding more Airbnbs of the Entire home/apt and private room types might help in increasing the revenue.