STORYTELLING CASE STUDY: AIRBNB, NYC

Data insights of Airbnb in NYC

CONTENTS

- OBJECTIVE
- BACKGROUND
- DATA PREPARATION
- AIRBNB DATA DESCRIPTION
- DATA ASSUMPTIONS – VARIABLES
- PROBLEM STATEMENT OF AIRBNB
- KEY FINDINGS – CHARTS
- APPENDIX – DATA METHODOLOGY
- CONCLUSION AND RECOMMENDATIONS

OBJECTIVE:

➢Airbnb is an online platform using which people can rent their unused accommodations.

➢During the covid time, Airbnb incurred a huge loss in revenue.

➢People have now started travelling again and Airbnb is aiming to bring up the business again and e ready to provide services to customers.

BACK GROUND

➢For the past few months, Airbnb has seen a major decline in revenue.

➢Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

➢So, analysis has been done on a dataset consisting of various Airbnb listings in New York.

AIRBNB DATA DESCRIPTION

The columns in the dataset are self-explanatory. You can refer to the diagram given below to get a better idea of what each column signifies.

Note: The price column contains the price/night.

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

Dataset Description

DATA ASSUMPTIONS - VARIABLES


```
Categorical Variables:
    - room_type
    - neighbourhood_group
    - neighbourhood

Continous Variables(Numerical):
    - Price
    - minimum_nights
    - number_of_reviews
    - reviews_per_month
    - calculated_host_listings_count
    - availability_365
- Continous Variables could be binned in to groups too

Location Varibles:
    - latitude
    - longitude

Time Varibale:
    - last_review
```

Variable Categories

PROBLEM STATEMENT OF AIRBNB

- For the past few months, Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

- The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue. Our responsibility is to provide valuable insights to aid in decision making.
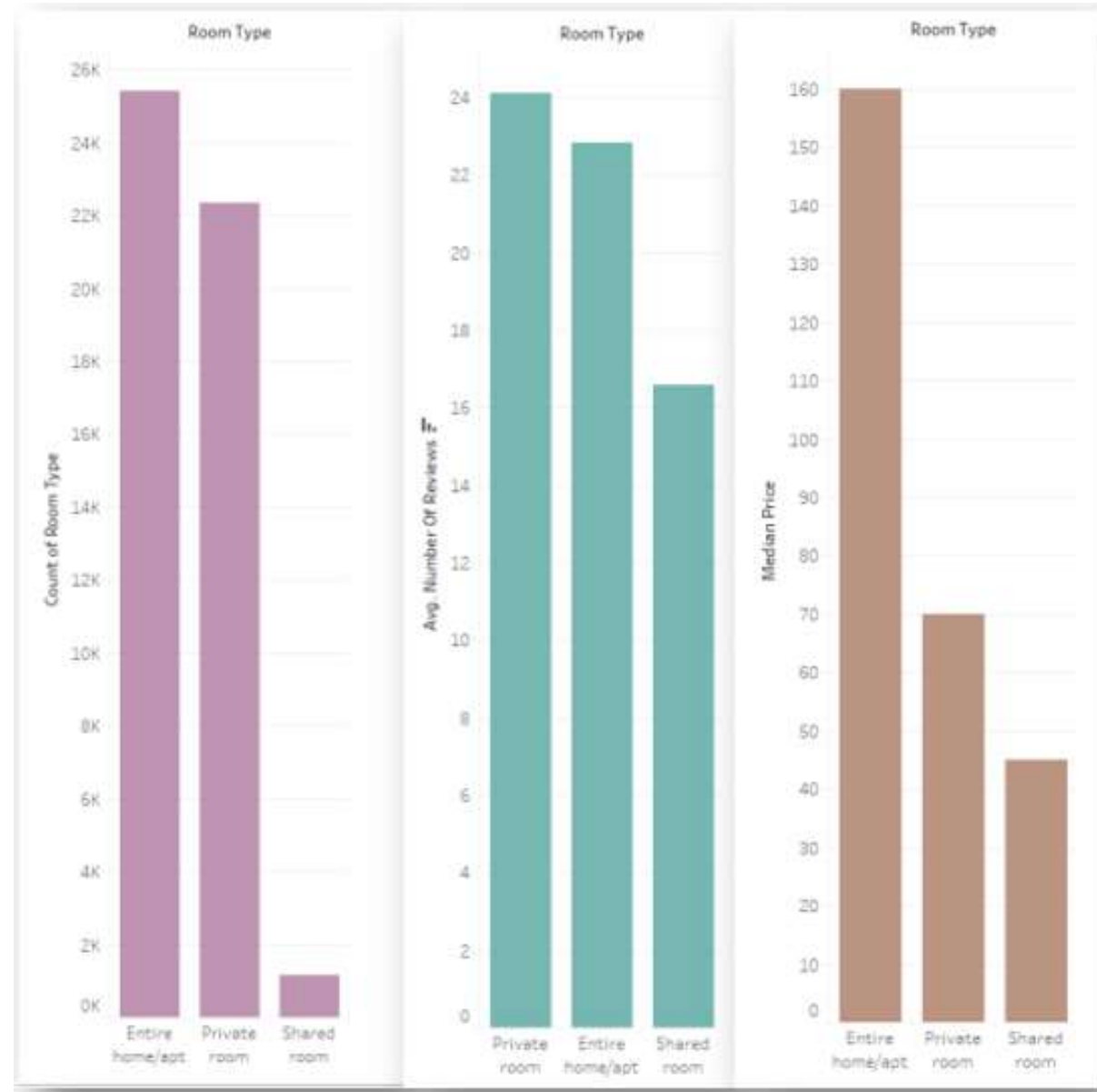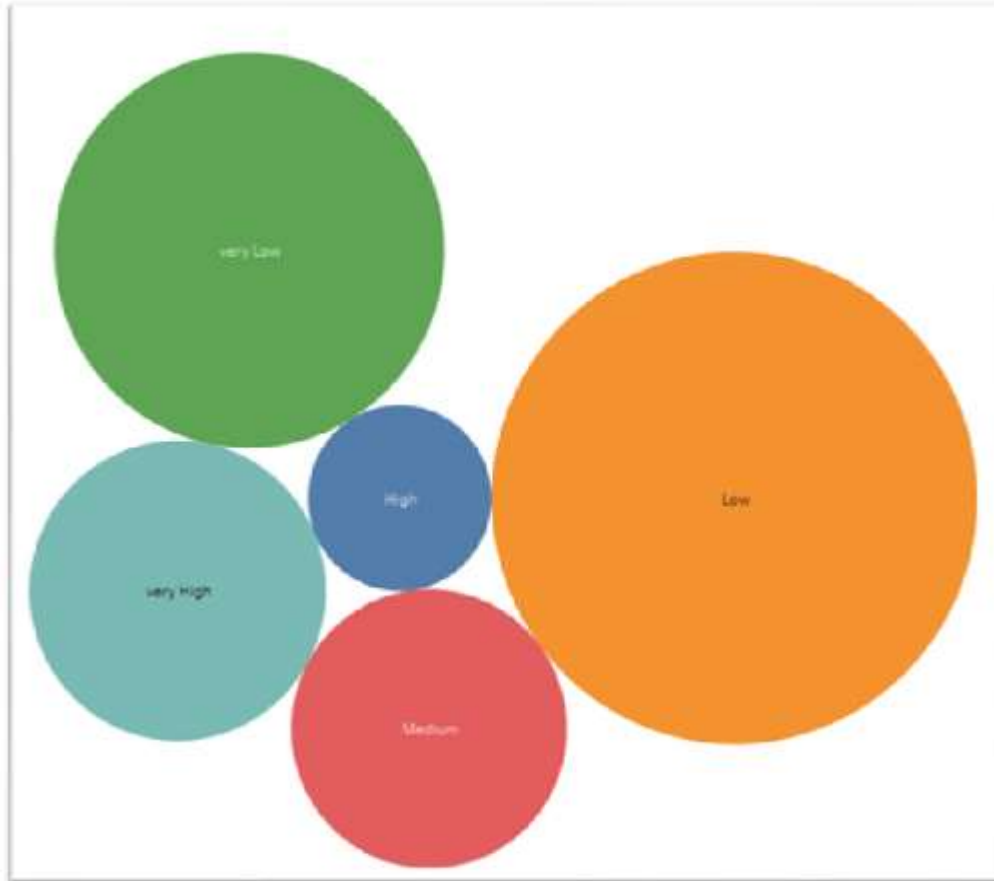
EVERY HOST MATTER
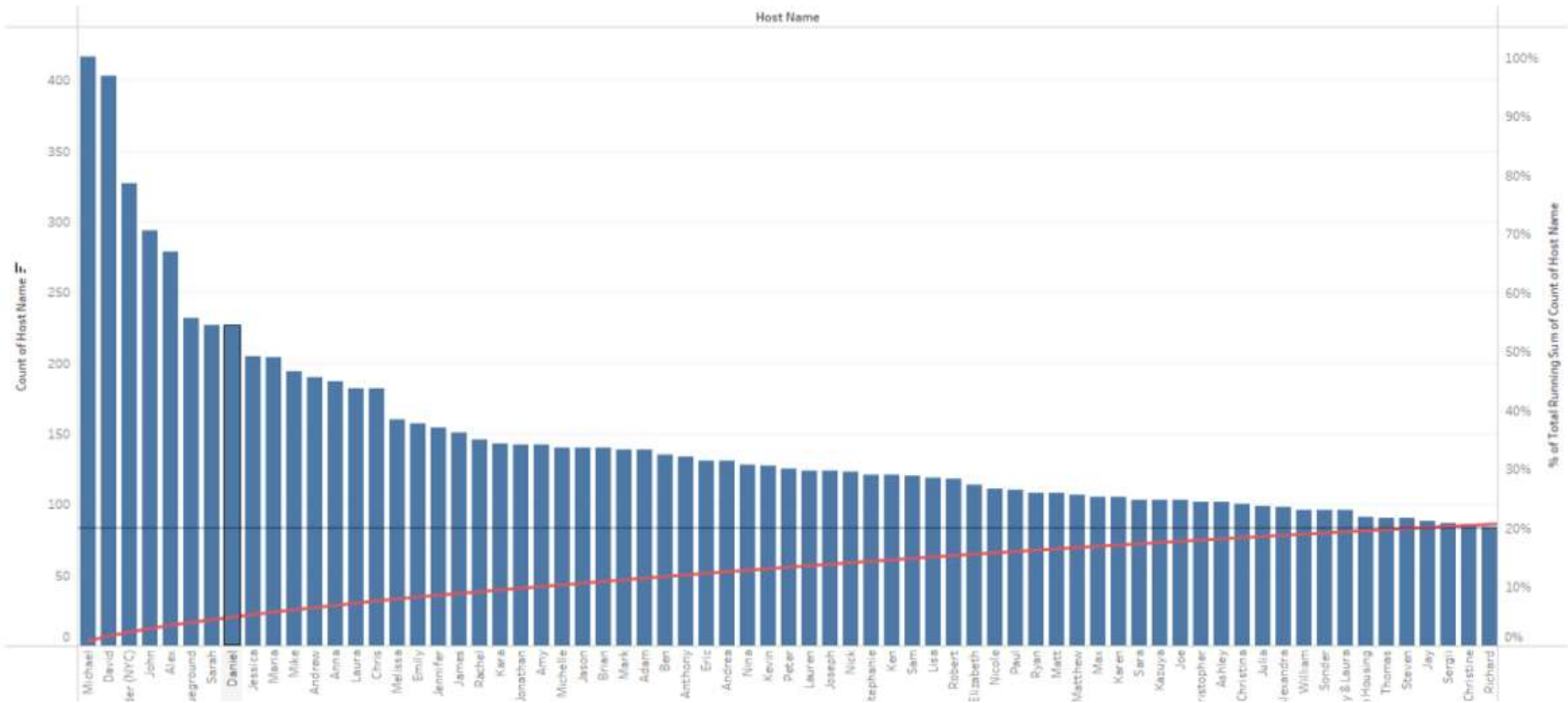
- The top 60 hosts only make up 20% of the total host count!

THE PROBLEMS OF SHARED ROOMS

➢ Median rates for shared rooms are significantly lower.

➢ They are less likely to be reviewed.

➢ Shared rooms only accounts for 2% of the total types of rooms.
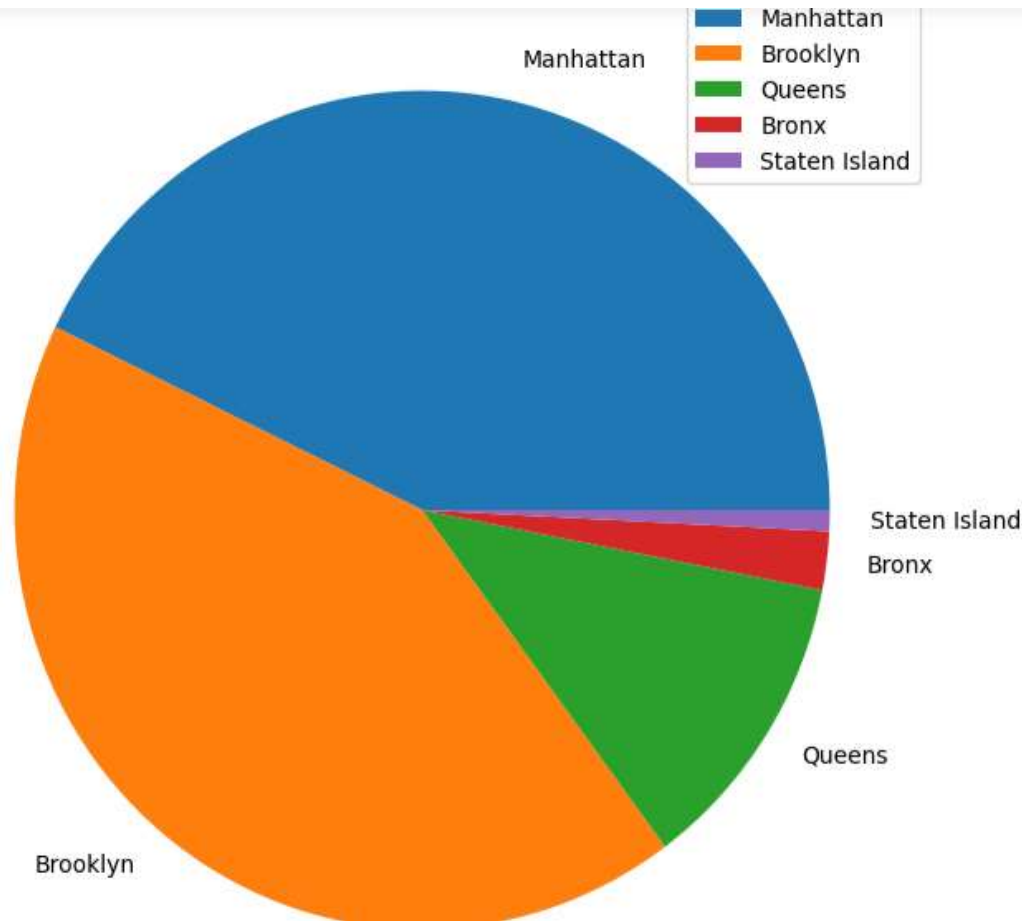
MINIMUM NIGHT CATEGORIES

Minimum night category percentages

Low	48.280192
very Low	26.014930
very High	14.997444
Medium	12.960425
High	5.747009

- Low category in minimum night feature contributes 40 %

MOST CONTRIBUTING NEIGHBOURS

5.4 neighbourhood_group
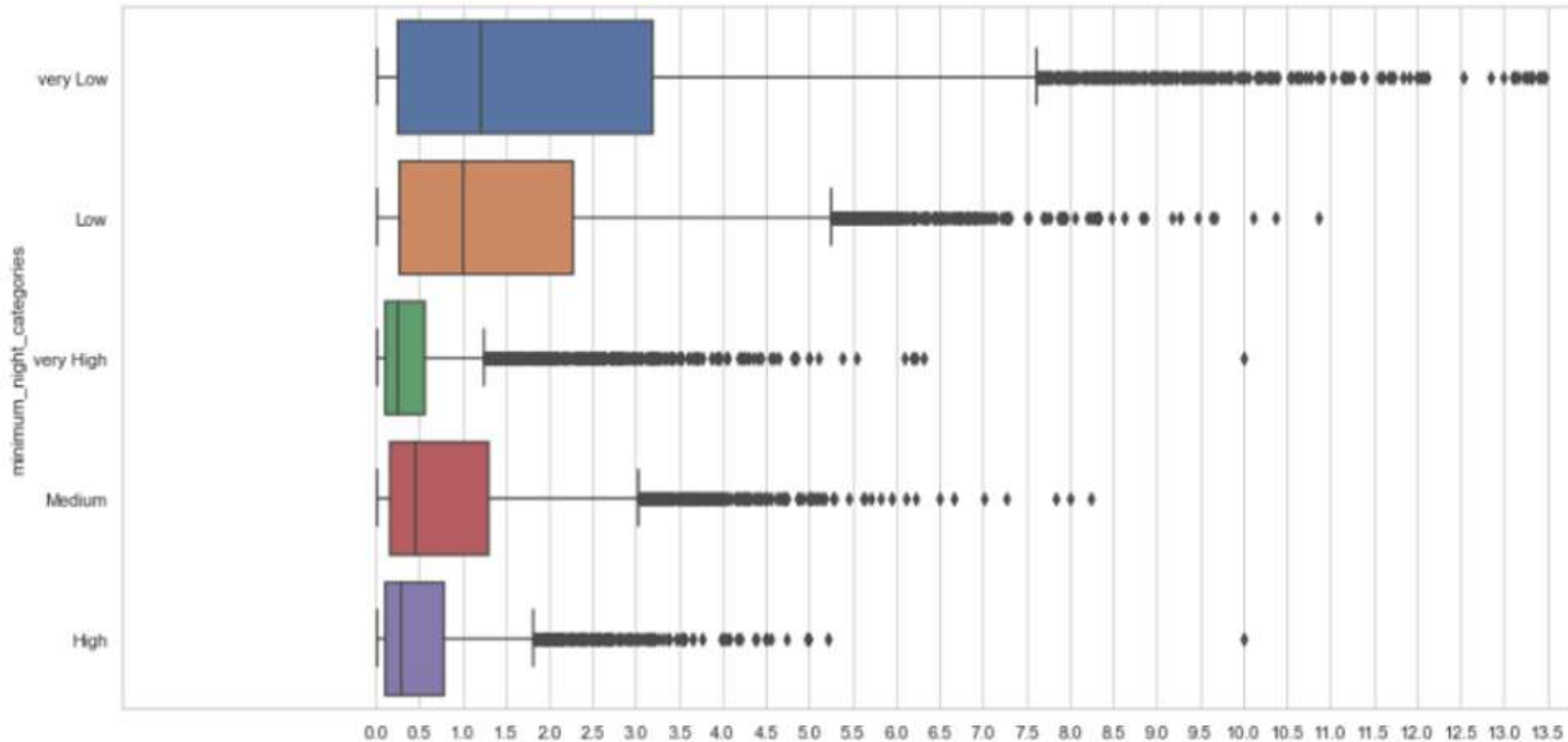
```
Airbnb_data1.neighbourhood_group.value_counts(normalize=True)*100
```

```
Manhattan        42.814456
Brooklyn         42.345638
Queens           11.777131
Bronx             2.253935
Staten Island     0.808841
Name: neighbourhood_group, dtype: float64
```

- 81 % of the listing are Manhattan and Brooklyn neighborhood group
- Staten Island has the lowest contribution.

EFFECT OF MINIMUM NIGHT CATEGORIES

- Customers are more likely to leave reviews for lower number of minimum nights.

DATA METHODOLOGY

➢ Conducted a through analysis of New York Airbnbs Dataset.

➢ Cleaned the data set using python.

➢ Derived the necessary features.

➢ Used group aggregration , pivot table and other statistical methods.

➢ Created charts and visualization using Tableau.

CONCLUSION AND RECOMMENDATIONS

❖Shared rooms need to be inspected upon.

❖More than 80% of the listings are Manhattan and Brooklyn neighbourhood group .

❖Minimum nights threshold should be on the lower side to make properties more customer-oriented.

❖The cumulative contribution of all hosts is better than a few hosts doing well.