Insights On Our NYC Properties

AGENDA

- Objective
- Background
- Key findings
- Recommendations
- Appendix:
  - Data sources
  - Data methodology
  - Data model assumptions

OBJECTIVE

- To analyse the data provided for our NYC properties.
- To gather insights based on the data analysis.
- Provide a report based on observations to increase the revenue.

BACKGROUND

- The revenues have dropped in the past few months.
- This drop is majorly due to restrictions on travel
- As the restrictions are being lifted, we need to be prepared for the increasing influx of customers

A WIDE RANGE OF PARAMETERS ARE PROVIDED

- The data has information on hosts: their ID,name and platform details.
- Localities and neighbourhoods are given with coordinates.
- Room types are listed with their metrics: price, minimum number of nights, reviews and availability

```
AB_NYC_2019_Cleaned.head()
```

	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
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	19-10-2018	0.21	6	365
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	21-05-2019	0.38	2	355
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN	1	365
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	05-07-2019	4.64	1	194
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	19-11-2018	0.10	1	0

AN EXCEL SHEET WAS PROVIDED FOR THE DATA

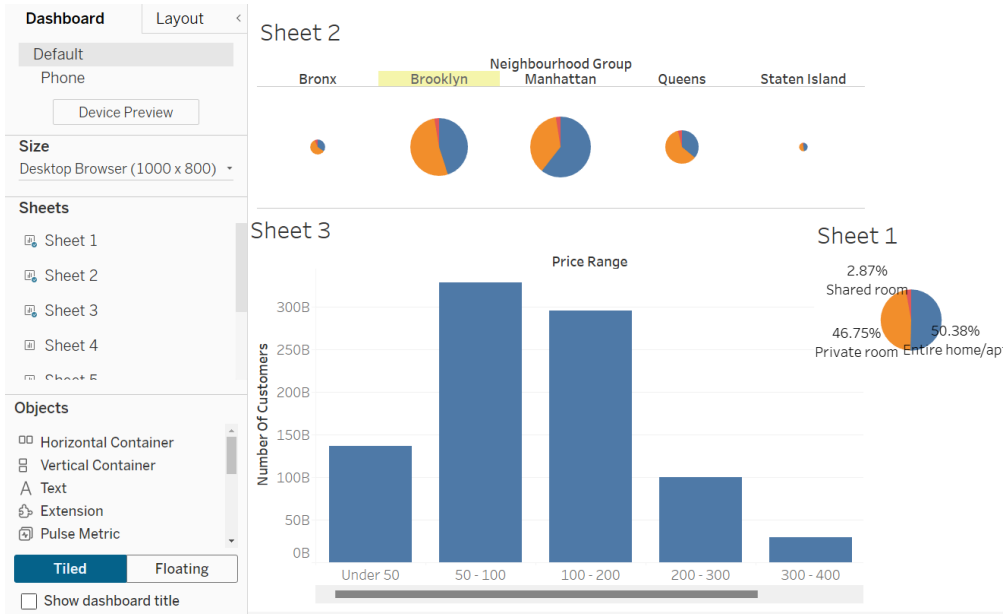
- Data was provided in a structured format of an excel sheet which could be worked upon in python and later used for visualisations
- 48895 entries are recorded, which is an ample amount of data to be studied and gather information from
- Missing values and datatypes were dealt with using python

```
AB_NYC_2019_Cleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              48895 non-null  int64
 1   name                            48879 non-null  object
 2   host_id                         48895 non-null  int64
 3   host_name                       48874 non-null  object
 4   neighbourhood_group             48895 non-null  object
 5   neighbourhood                   48895 non-null  object
 6   latitude                        48895 non-null  float64
 7   longitude                       48895 non-null  float64
 8   room_type                       48895 non-null  object
 9   price                           48895 non-null  int64
 10  minimum_nights                  48895 non-null  int64
 11  number_of_reviews               48895 non-null  int64
 12  last_review                     38843 non-null  object
 13  reviews_per_month               38843 non-null  float64
 14  calculated_host_listings_count  48895 non-null  int64
 15  availability_365                48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

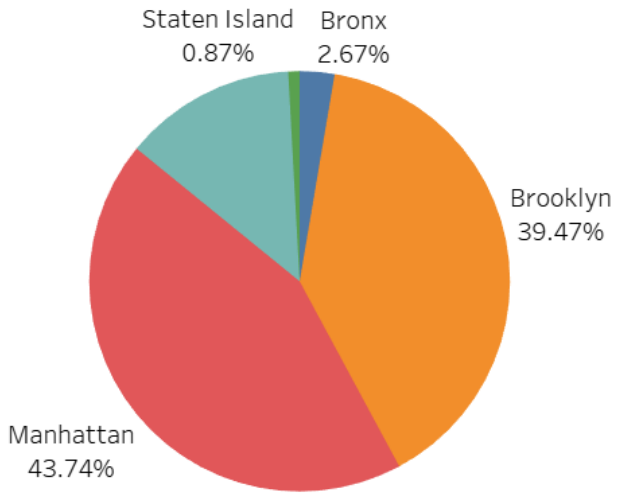
VISUALISATION IN TABLEAU FOR THE EDA

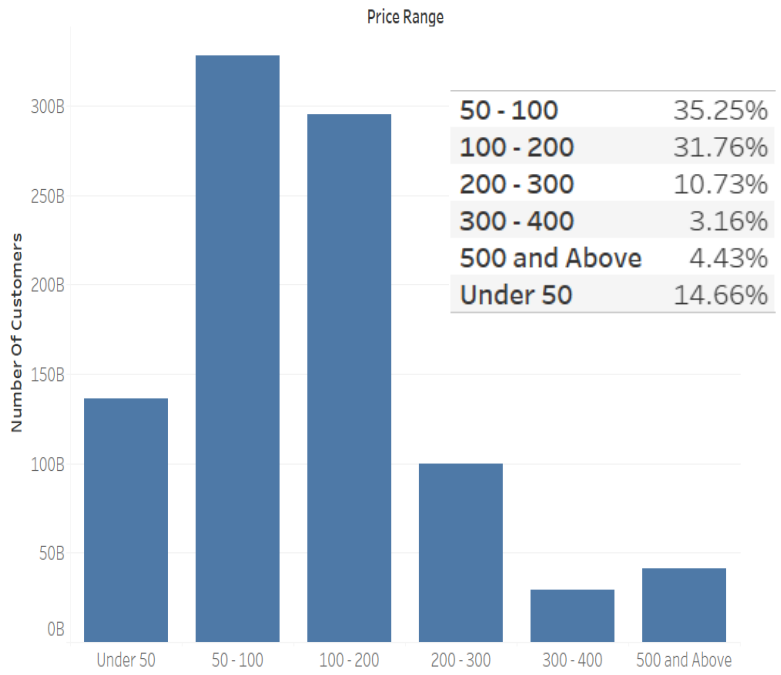
- The cleaned data was saved in another excel sheet and imported to tableau.
- Tableau was used to carry out the EDA, wherein all the parameters and metrics were analysed easily with the available functionalities.
- Suitable charts and graphs were used based on the data to prepare the visualization.



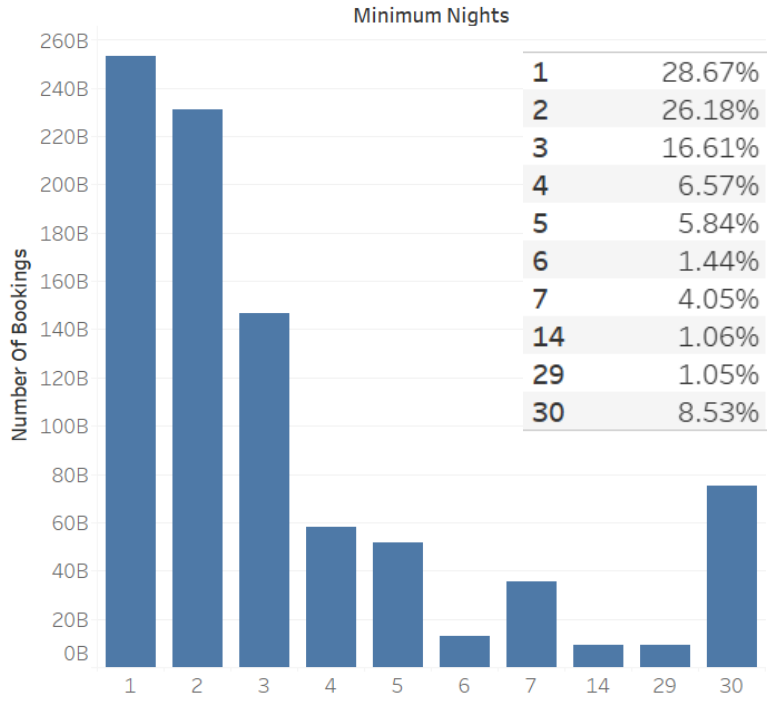
MAJOR INSIGHTS DRAWN FROM THE DATASETS

- Manhattan and Brooklyn attribute to more than 80% of the total bokings
- Prices less than $200 also bring more than 80% bookings
- Lesser minimum nights attract more customers due to the flexibility and ease of choice, followed by a month and then a week.





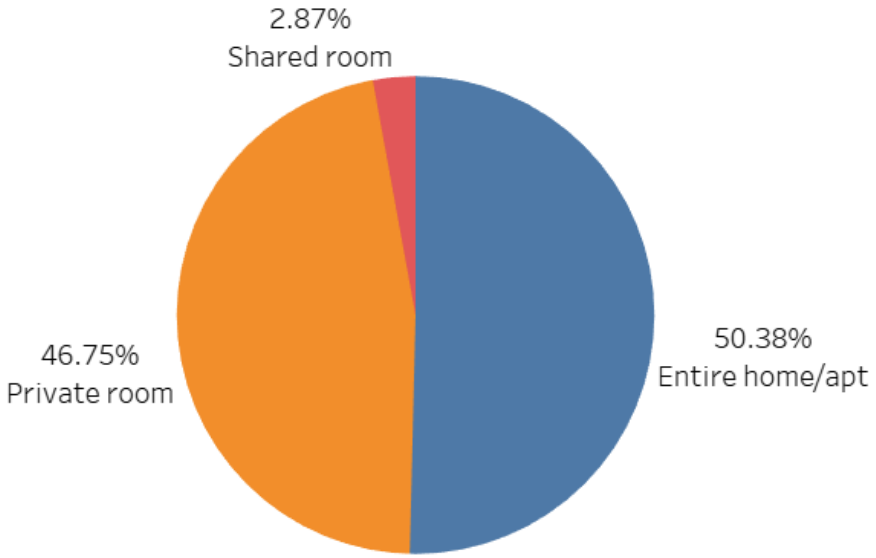
50 - 100	35.25%
100 - 200	31.76%
200 - 300	10.73%
300 - 400	3.16%
500 and Above	4.43%
Under 50	14.66%

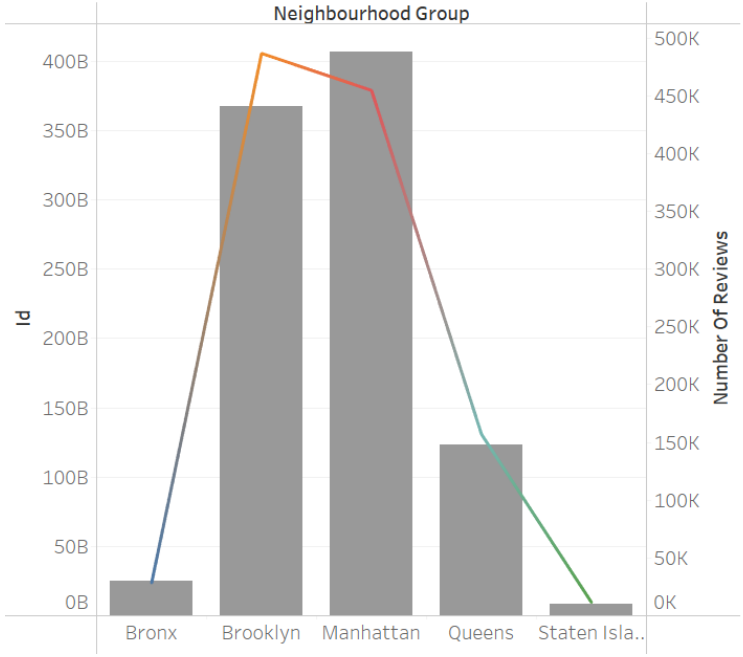


1	28.67%
2	26.18%
3	16.61%
4	6.57%
5	5.84%
6	1.44%
7	4.05%
14	1.06%
29	1.05%
30	8.53%

PROBLEM AREA: SHARED ROOMS AND REVIEW RATINGS

- Shared rooms are the least preffered with a very less percentage of bookings.
- Number of reviews are also very less compared to the number of bookings
- Review ratings are missing from the data which should be gathered to provide more information on customer satisfaction.





Conclusions and Recommendations

- The available data has provided sufficient information to take actions on it.
- Insights and Visuals are clear enough to be presented to the stake holder to give them an understanding and overview.
- Data can be collected for the review rating which would give a better understanding of customer satisfaction and quality of services

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

The attribute names and what they describe is provided in the table

APPENDIX - DATA METHODOLOGY

- We conducted a thorough analysis of the Airbnb Data. The process included:
  - Collection of data in the form of excel file.
  - Cleaning the data and handling missing values using python and its libraries.
  - Setting up the correct datatypes for the provided entries.
  - Importing the data in Tableau to create charts and visuals
  - Preparing visuals and charts to carry on EDA.
  - Presenting the observations in the form of a report for the understanding of stackholders

APPENDIX - DATA ASSUMPTIONS

```
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
```