IIIT-B , Jan 2024

Storytelling Case Study: Airbnb, NYC

PPT-1

- Pallav Rajput

AGENDA PAGE

Objective

Data Lifecycle

Analysis methods

Conclusion

Objective

The revenue of Airbnb has significantly decreased during the last few months. We want to make sure that we are ready for this shift now that the limitations have begun to loosen and people are beginning to travel more.

Data Lifecycle

1. Importing 2. Missing value treatment 3. Creating features[binning]

01.

```
port warnings
#warnings.filterwarning("ignore")
import numpy as np, pandas as pd, matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
air = pd.read_csv("AB_NYC_2019.csv")
air.head()
```

```
air.isnull().sum()
```

```
id                                0
name                             16
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
last_review                   10052
                              10052
                                  0
availability_365                  0
dtype: int64
```

Collapse Output

02.

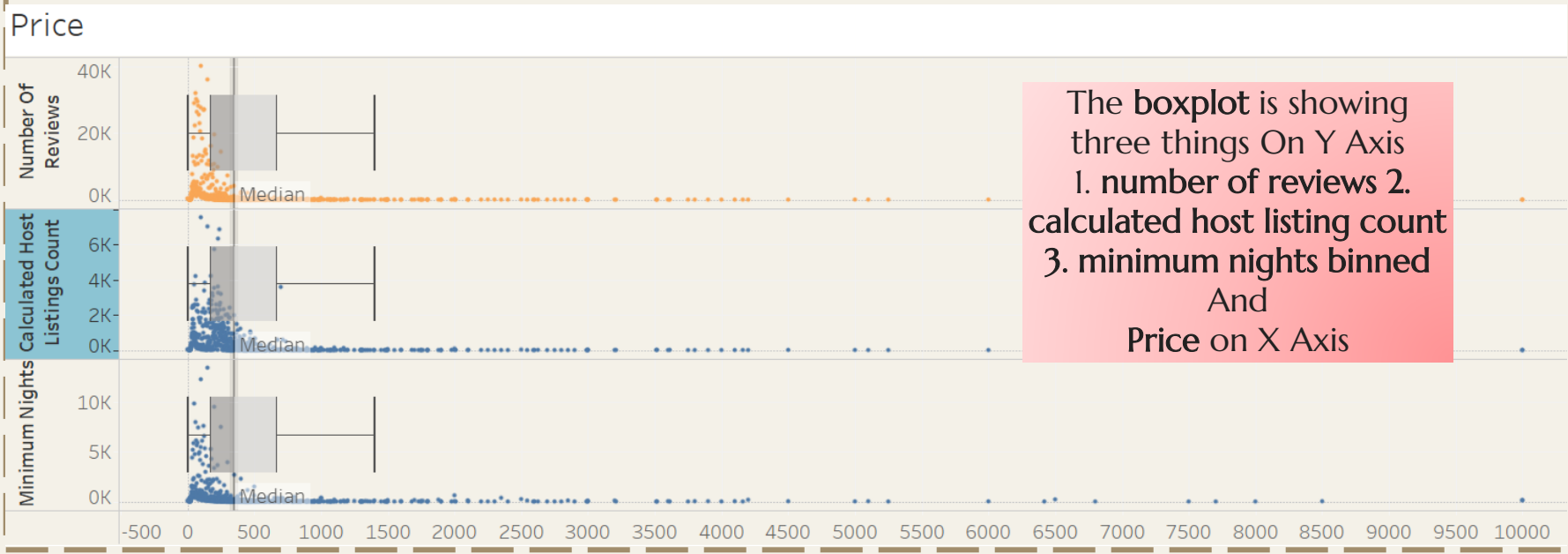
```
[17]: air.fillna({'reviews_per_month':0}, inplace=True)
```

03.

Min Nighs Binned

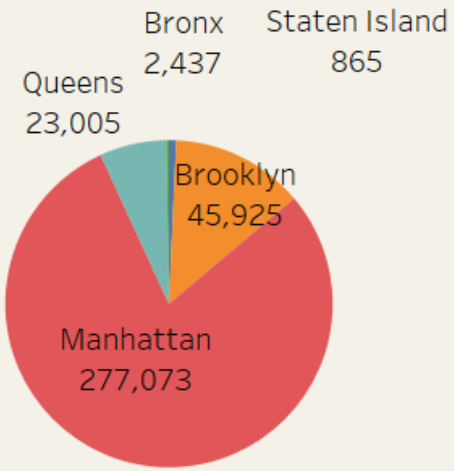
```
if [Minimum Nights] = 1 then '1'
elseif [Minimum Nights] = 2 then '2'
ELSEIF [Minimum Nights] = 3 then '3'
elseif 4<=[Minimum Nights] and [Minimum Nights]<=6 then '4-6'
elseif 7<=[Minimum Nights] AND [Minimum Nights]<=14 then'1Wk-2Wk'
ELSEIF 15<=[Minimum Nights] and [Minimum Nights]<=28 then'2Wk-4Wk'
ELSE '1Mn/>'
END
```

Analysis



The **boxplot** is showing three things On Y Axis
1. **number of reviews** 2. **calculated host listing count** 3. **minimum nights binned**
And
Price on X Axis

Distribution of total host listing count



01.

Calculated Host Listings Count
349,305

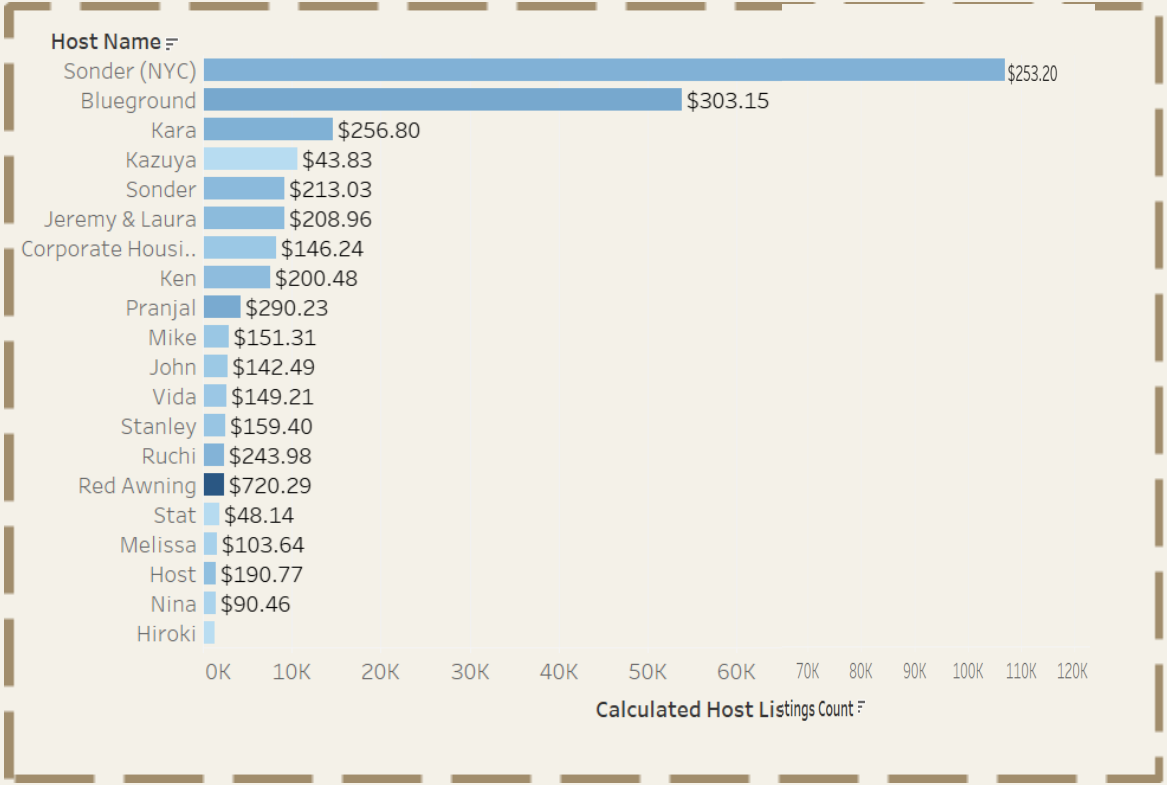
The total calculated hosting count is **349,305** out of which the Manhattan has a maximum number of listing counts with **277,073**.

The Staten Island shows the least listing count with only **865** listing.

This is how the minimum nights were binned using the tableau calculated fields.

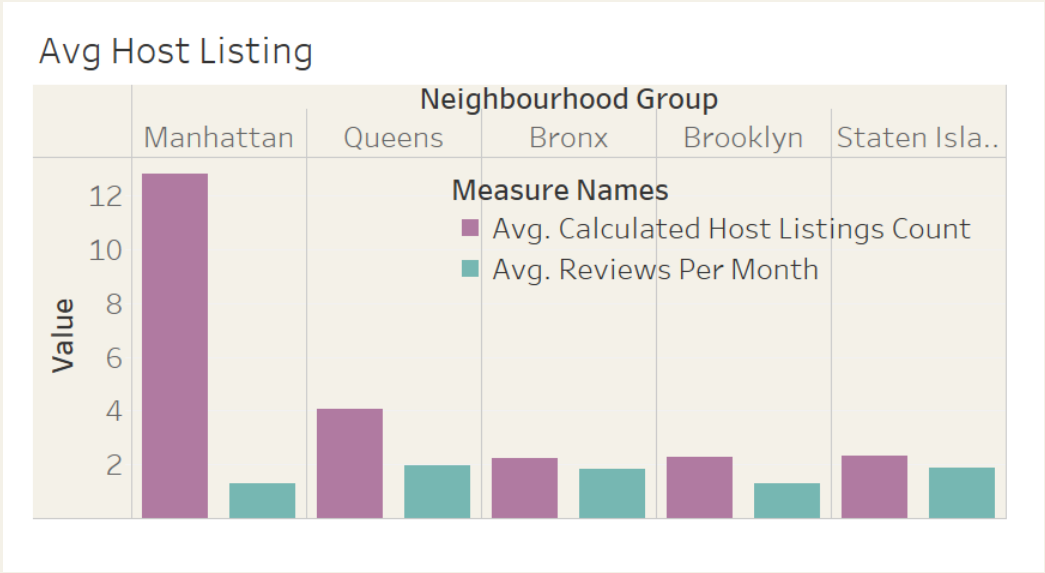
- 1 Day = 1
- 2 Days 2
- 3 Days = 3
- 4-6 days = 4-6
- 7 – 14 days = 1week – 2weeks
- 15 - 28 days = 2weeks – 4weeks
- 28 days – 365 days = 1 month or more

Top 20 most listed



The bar chart shows the top 20 host name sorted descending by Listing. Where the **Sonder(NYC)**, **Blue ground** and **Kara** are among the top 3.

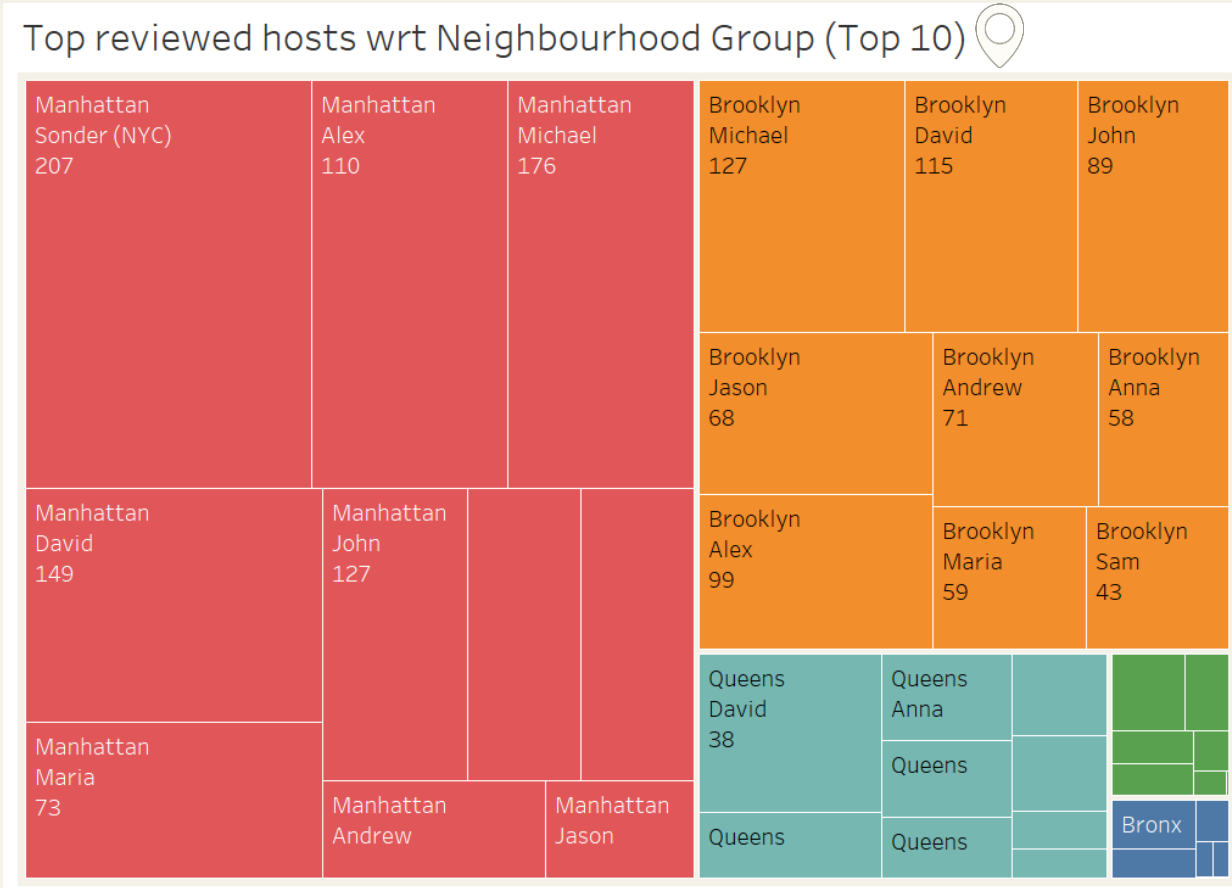
Red Awning listed 15th however is the most expensive in top 20 most listed, followed by **Blue ground** and **Pranjal**.



Calculated host listing and reviews per month

Manhattan has the highest average calculated host listing count however the reviews are quite low as compared to it.

Top reviewed hosts wrt Neighbourhood Group (Top 10)



S

STRENGTHS

It has been easy to perform the data cleaning because they were not major missing values and outliers

WEAKNESSES

W

The study has not seen any potential weaknesses

O

OPPORTUNITIES

The aim is discusses strategies for acquiring hosts, customer categorization, targeting neighborhoods, pricing ranges, identifying popular localities and strategies for attracting unpopular properties.

THREATS

T

The study has not seen any potential Threats

APPENDIX
Dataset Description & Variable Categories

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

```
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
```

Conclusion

01.

Characteristics

Numerous characteristics in the dataset are used to draw strong, important findings.

02.

Presentations

The presentations for the stakeholders might make use of a wide range of visuals.

03.

Modeling

It is possible to create a clustering machine learning model to find collections of related items in datasets with two or more varied quantities.

04.

In order to support the subsequent analysis, the data collecting team should gather information regarding review scores.

THANK YOU!

Pallav Rajput
+91-9888143002
ribhu.s18@gmail.com