Airbnb New York Data insights

By Atish, Devika, Sambit

Objective

- **Problem Statement:** Having a huge decline in revenue, Airbnb is looking for a growth in its revenue. It wants to be prepared to take the opportunity of the lifting of restriction.

- Approach :

To perform a thorough analysis of the data set provided

To have a clear data insights

Conclusion based on the analysis techniques and visualization

Importing libraries and reading Data

In [1]:
```
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt,seaborn as sns
%matplotlib inline
```

In [2]:
```
data= pd.read_csv('C:\\Users\\Admin\\Downloads\\AB_NYC_2019.csv')
data.head()
```

Out[2]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revie
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	2
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

Data Clean Up

```
In [3]: data.shape
```

```
Out[3]: (48895, 16)
```

```
In [4]: #Checking NULL values
        data.isnull().sum()
```

```
Out[4]: 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
        reviews_per_month               10052
        calculated_host_listings_count      0
        availability_365                    0
        dtype: int64
```

Removing null values

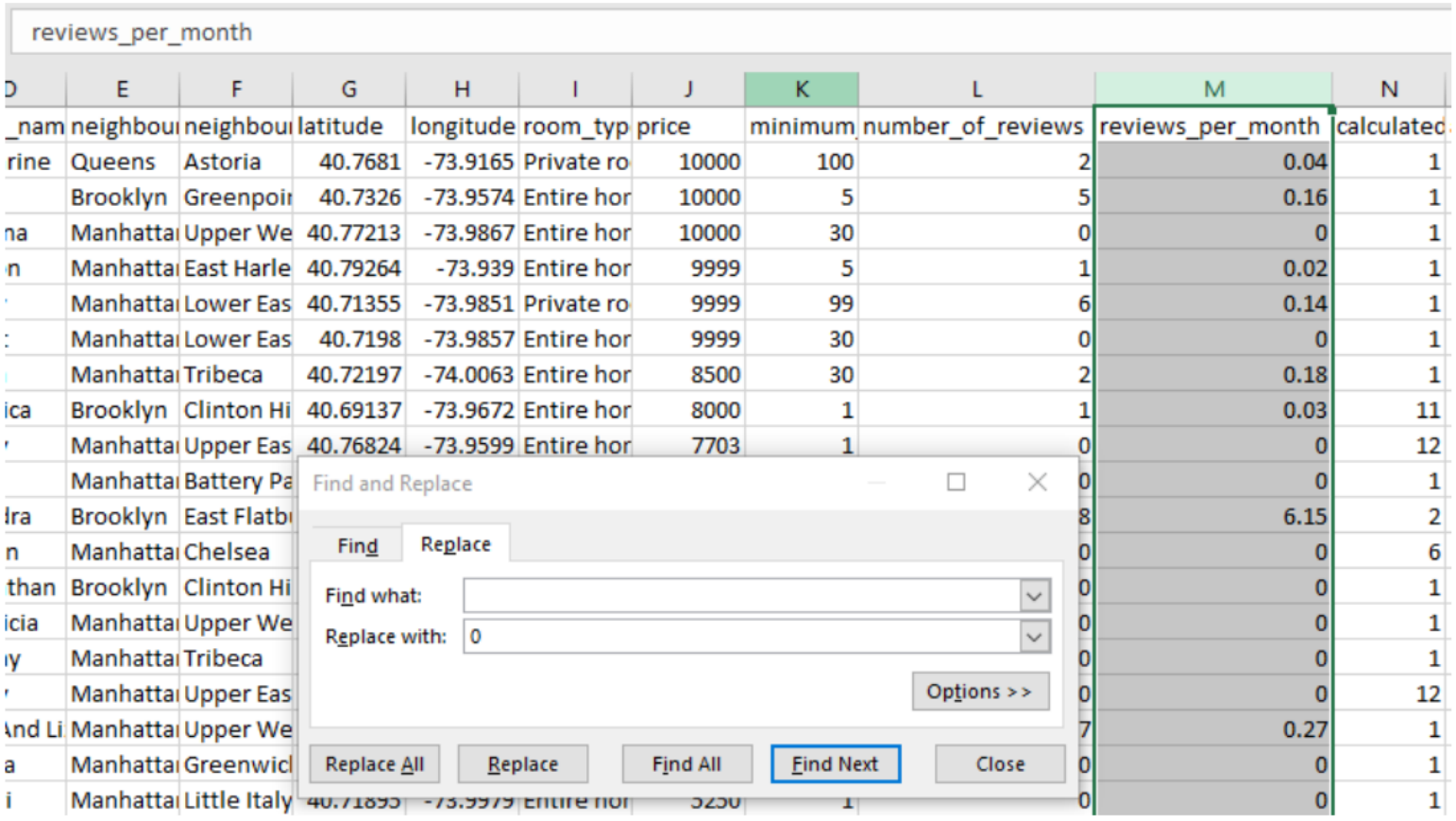
```
data= data[~data.name.isnull()]
```

```
data= data[~data.host_name.isnull()]
```

```
data.drop('last_review',inplace=True,axis=1)
```

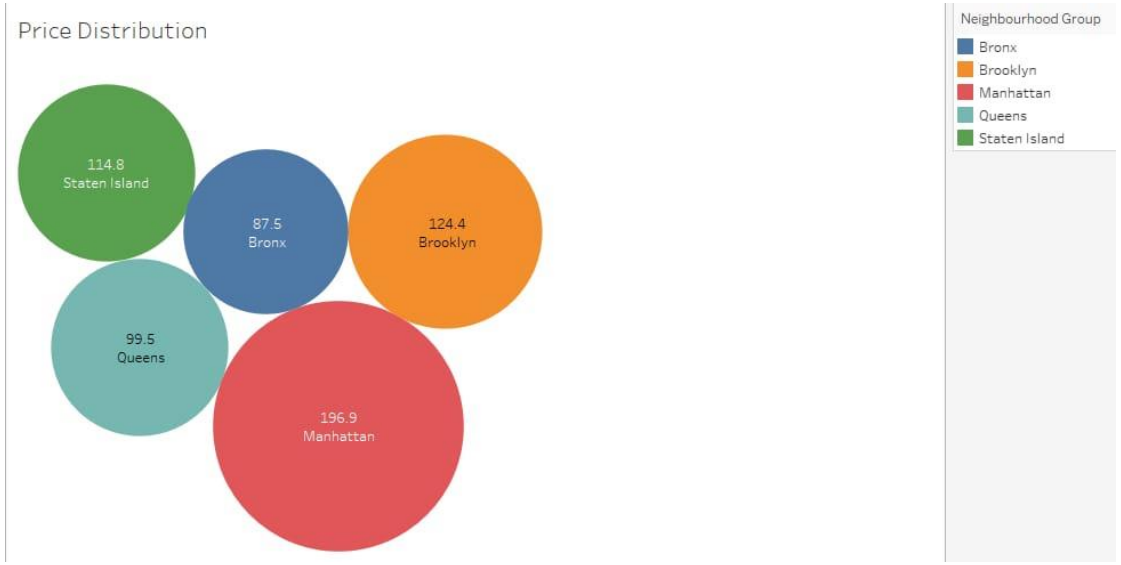
Data Manipulation

After exporting the data back to csv, some data manipulations were done in MS-Excel such as replacing null values with 0 in “reviews_per_month “column.



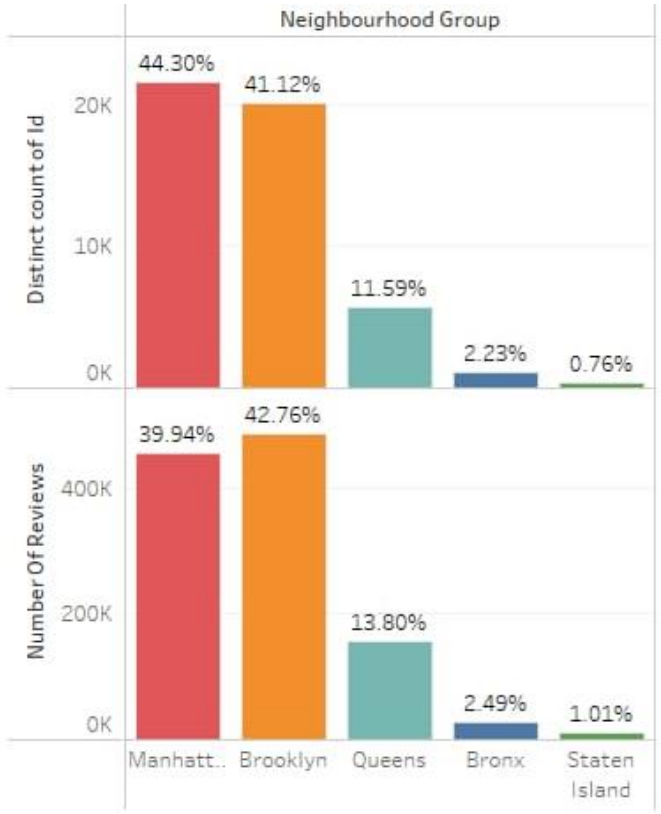
Pricing in the neighborhoods

- Manhattan has the highest price tag and Bronx has the lowest price tag.



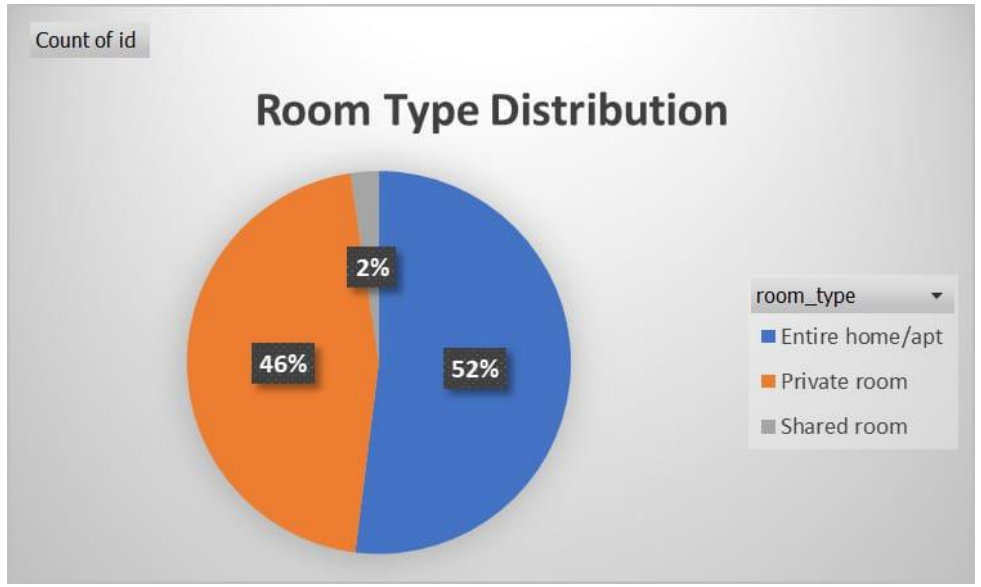
Most contributing neighborhoods

- “Manahattan” has the maximum number of properties where as “Staten island” is has the minimum.
- Maximum review received from the guest is from “Brooklyn” neighborhood.
- “Manhattan” and “Brooklyn” together have more than 80% listing in comparison to other neighborhoods.



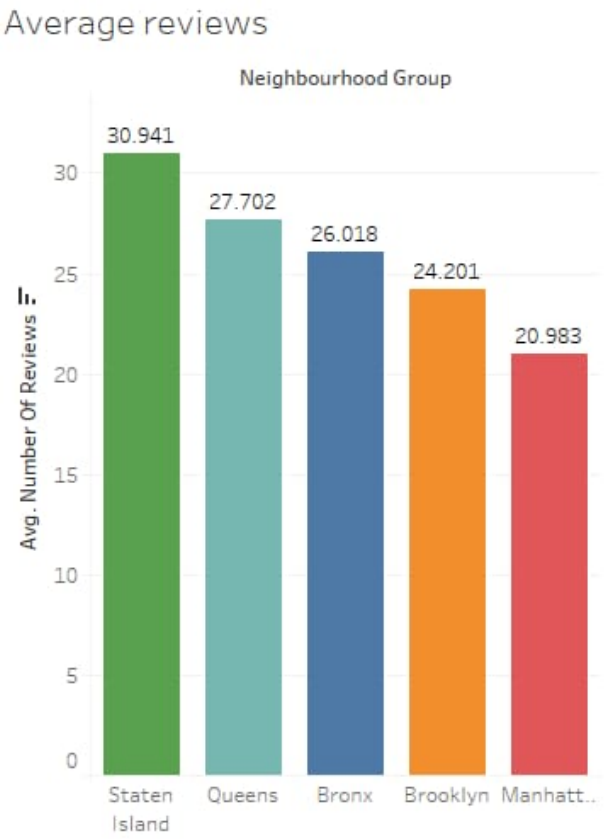
Room type in neighborhoods

- “Entire home/Apartment” type properties are maximum in the neighborhoods.
- “Shared room” type properties are less in the neighborhoods.



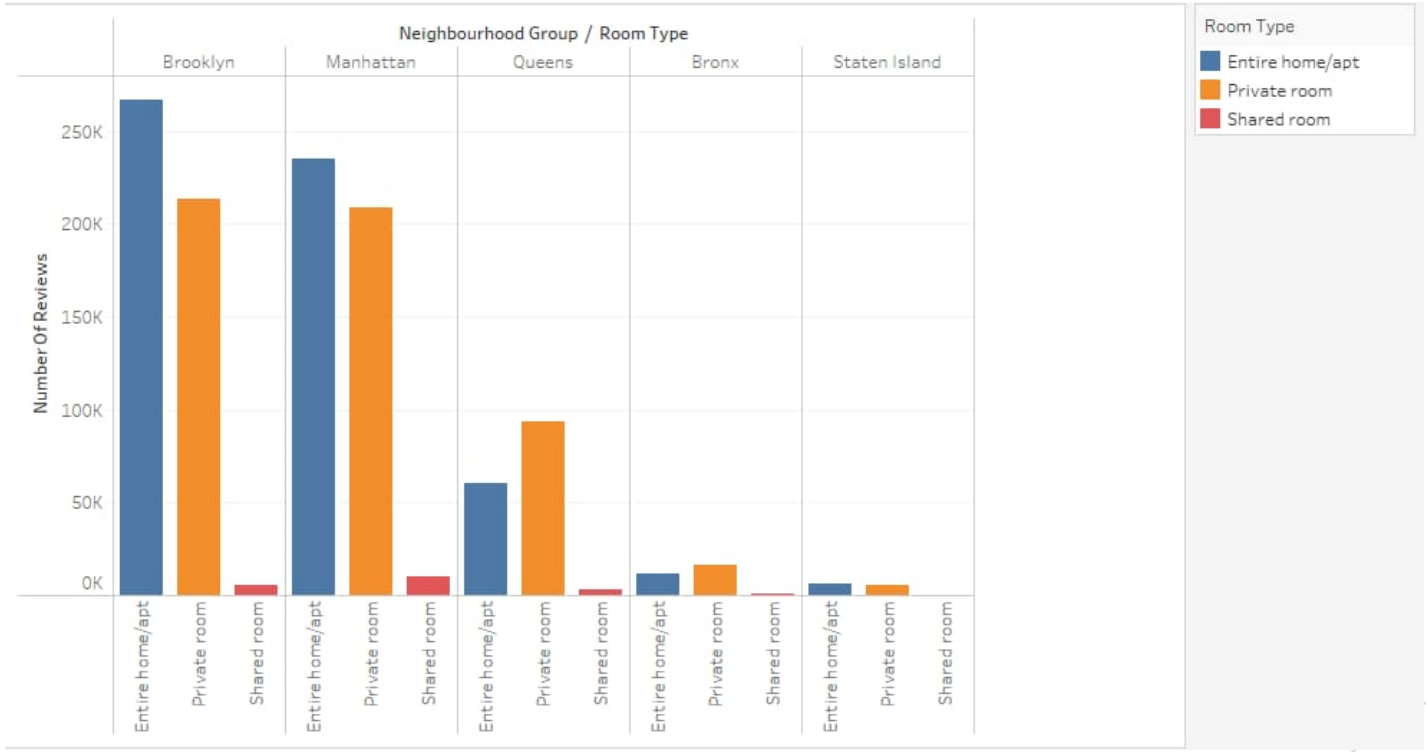
Reviews for neighborhoods

- Average no. of reviews received from “Staten Island” neighbourhood is the maximum among all neighbourhoods followed by Queens.
- It indicates the demand is more in “Staten Island” and “Queens” even through the count of properties are less.



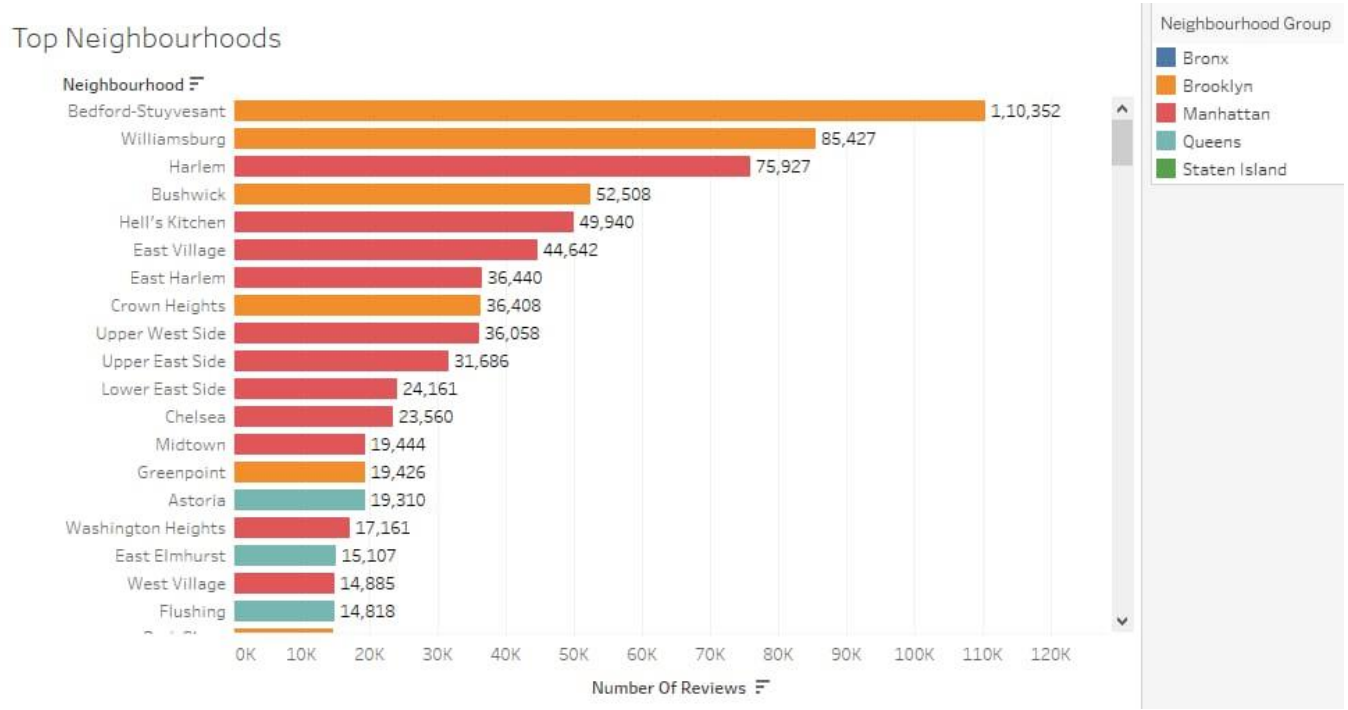
Reviews for neighborhoods

- “Shared rooms” received the least review due to the less availability in neighbourhood.
- “Brooklyn” and “Manhattan” have more reviews for “Entire home/Apartment” type properties.



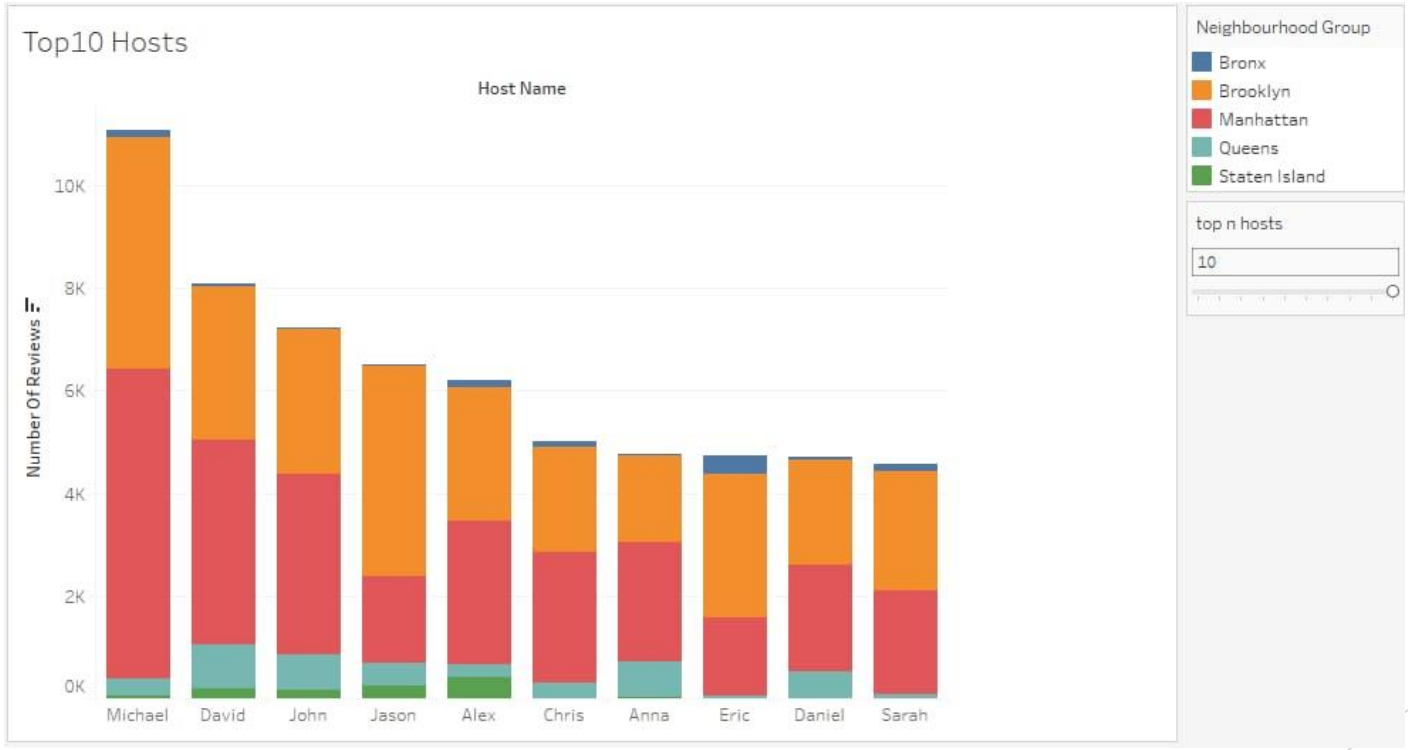
Reviews for neighborhoods

- “Bedford-Stuyvesant” has maximum number of reviews as part of “Brooklyn” neighborhood.
- Top 15 reviews belong to “Brooklyn” and “Manhattan” neighborhood.



Host matters the most

- “Michael” is the top host in “Brooklyn” and “Manhattan” neighbourhood.



Conclusion

- Data insights are derived based on the data set provided.
- Visualizations are created for the stakeholders perusal.
- Review scores play a vital role in this analysis.