Analysis of AirBnB Business for the City of New York and Future Steps for Expansion

-by

Srishti Shri

Sushil Kumar

Vishnuvardhan Pinjala

Agenda

- Improve Revenue Generation for AirBnB New York City
- Classify Customers on Preferences, Identify Profitable Neighbourhoods.
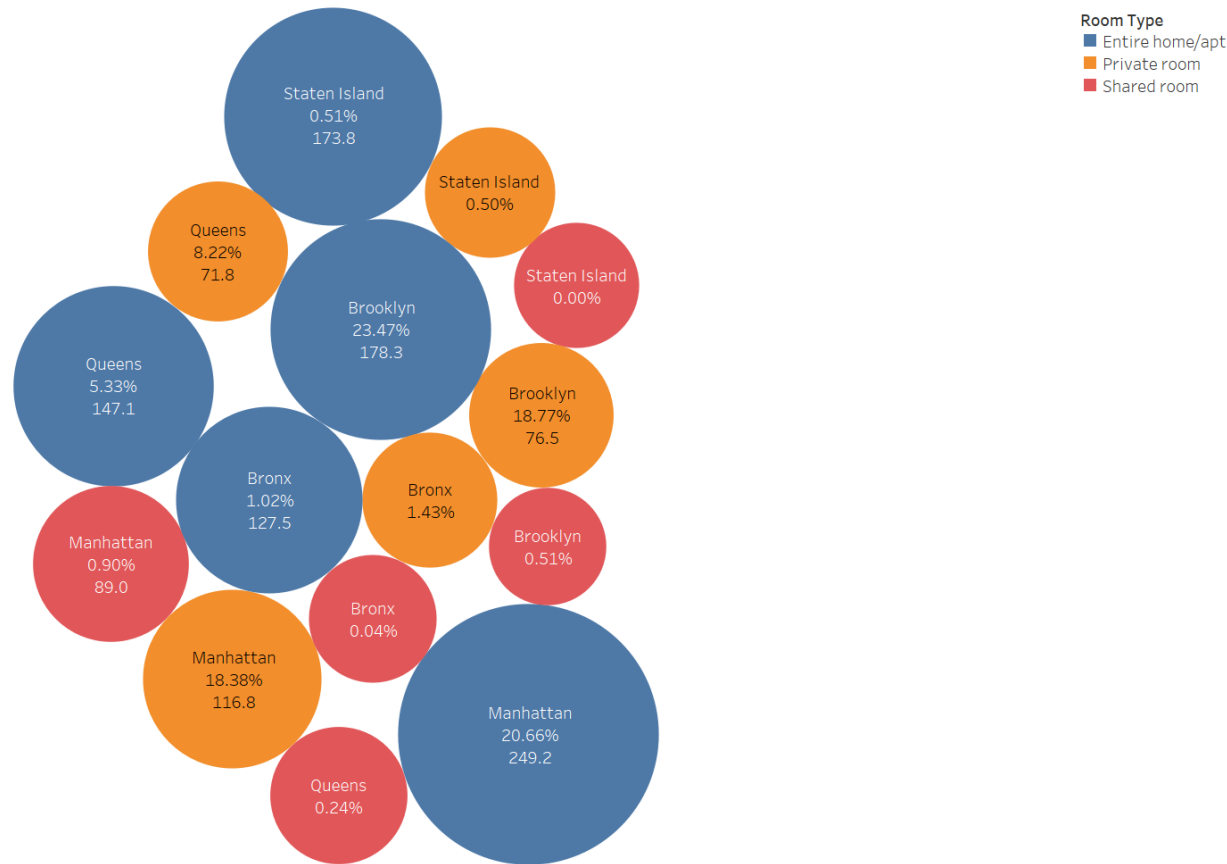- Steps to be taken to achieve said goal.

Future Steps to Raise Revenue Generation

- Acquire more Entire home/apt and Private Rooms in Brooklyn

- Decrease the availability of Entire home/apt and Private Rooms in Brooklyn

- Increase the availability of Entire home /apt and Private Rooms in Manhattan

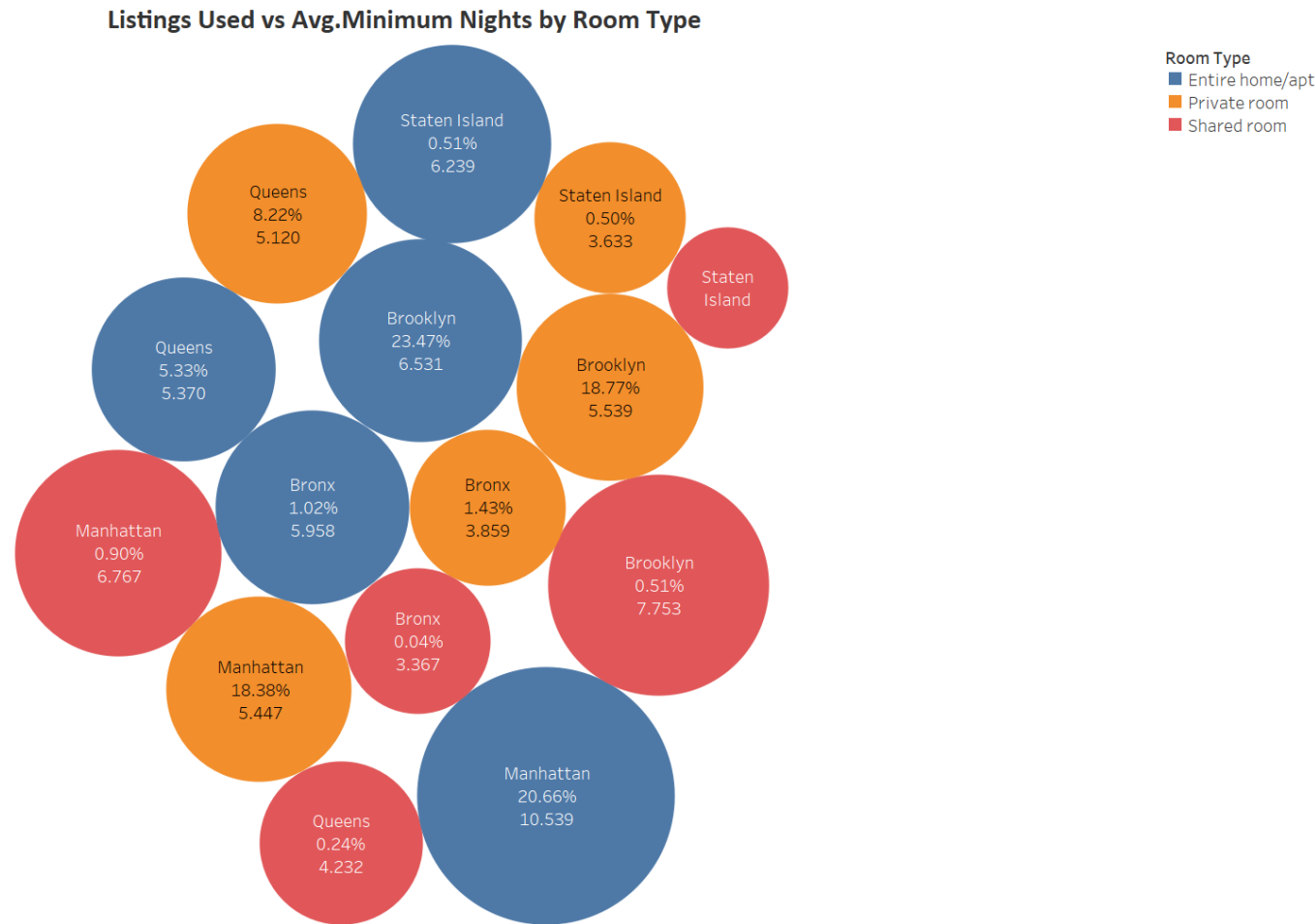- Let go of Listings of type Shared Rooms in all Districts.

Customer Preferences

Listings Used vs Avg.Price by Room Type

- The Chart here depicts variation in pricing of listing type across Neighbourhoods.

- From here we can gather information that despite high prices customers prefer **Entire homes** followed by **Private Rooms**.

- **Brooklyn** leads as preferred Neighbourhood.

Contd.

Listings Used vs Avg.Minimum Nights by Room Type

Room Type
- Entire home/apt
- Private room
- Shared room

Staten Island
0.51%
6.239

Queens
8.22%
5.120

Staten Island
0.50%
3.633

Staten Island

Brooklyn
23.47%
6.531

Queens
5.33%
5.370

Brooklyn
18.77%
5.539

Bronx
1.02%
5.958

Bronx
1.43%
3.859

Manhattan
0.90%
6.767

Brooklyn
0.51%
7.753

Manhattan
18.38%
5.447

Bronx
0.04%
3.367

Manhattan
20.66%
10.539

Queens
0.24%
4.232

Neighbourhood Group, % of Total Number Of Reviews and average of Minimum Nights. Color shows details about Room Type. Size shows average of Minimum Nights. The marks are labeled by Neighbourhood Group, % of Total Number Of Reviews and average of Minimum Nights.

➡ From This chart it is clearly inferred that Minimum Nights has no bearing towards Customer's influence in bookings.

Adjustment of Booking Availability

Listings Used vs Availability by Room Type
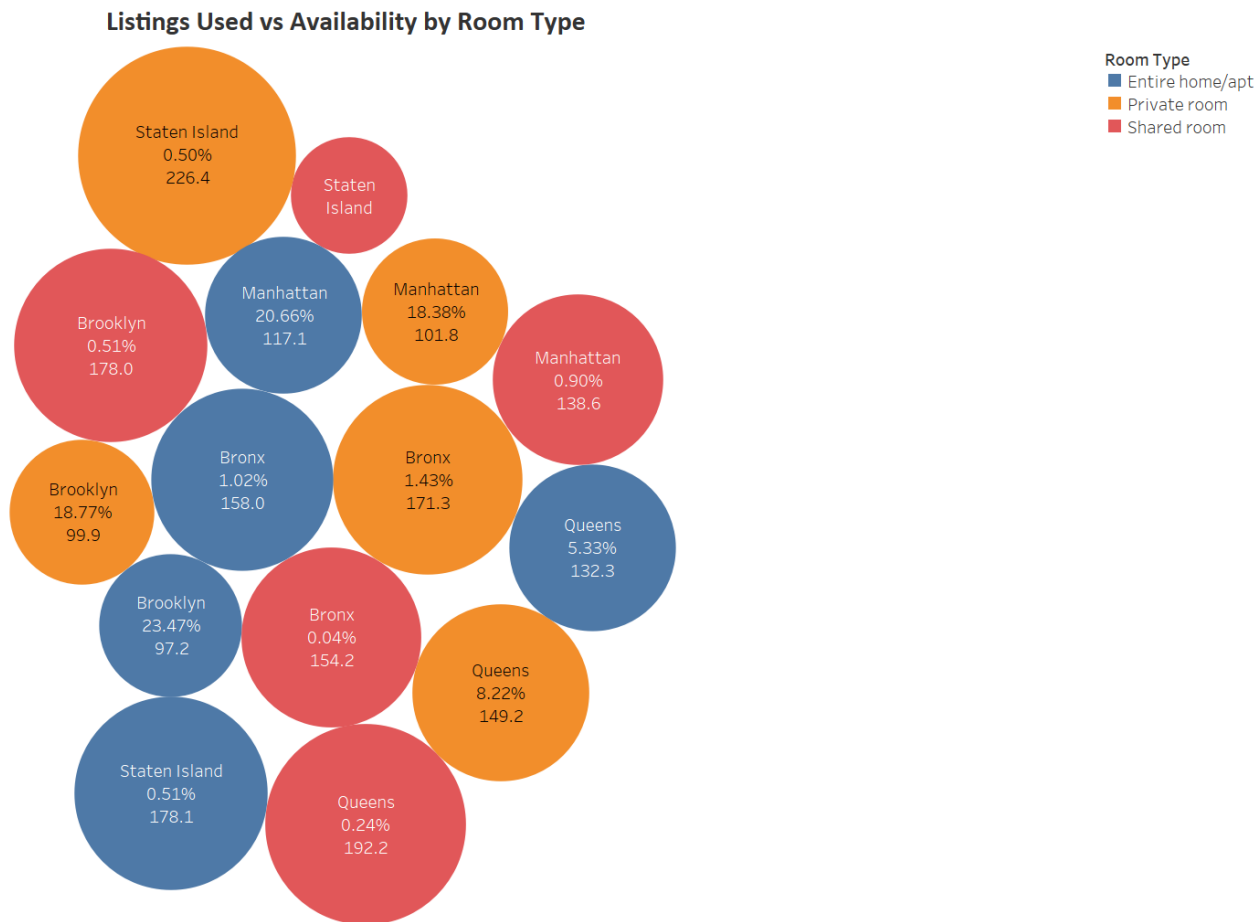
Neighbourhood Group, % of Total Number Of Reviews and average of Availability 365. Color shows details about Room Type. Size shows average of Availability 365. The marks are labeled by Neighbourhood Group, % of Total Number Of Reviews and average of Availability 365.

- From this chart we can see that **Brooklyn** has **lowest availability** for bookings.

- It can also be understood that non-preferred listings are available for the most time yet have failed to attract customers.

- So further *__decreasing availability in Brooklyn and increasing availability in Manhattan__* will nudge customers towards Manhattan without affecting booking in Brooklyn.

Acquiring Further Listings
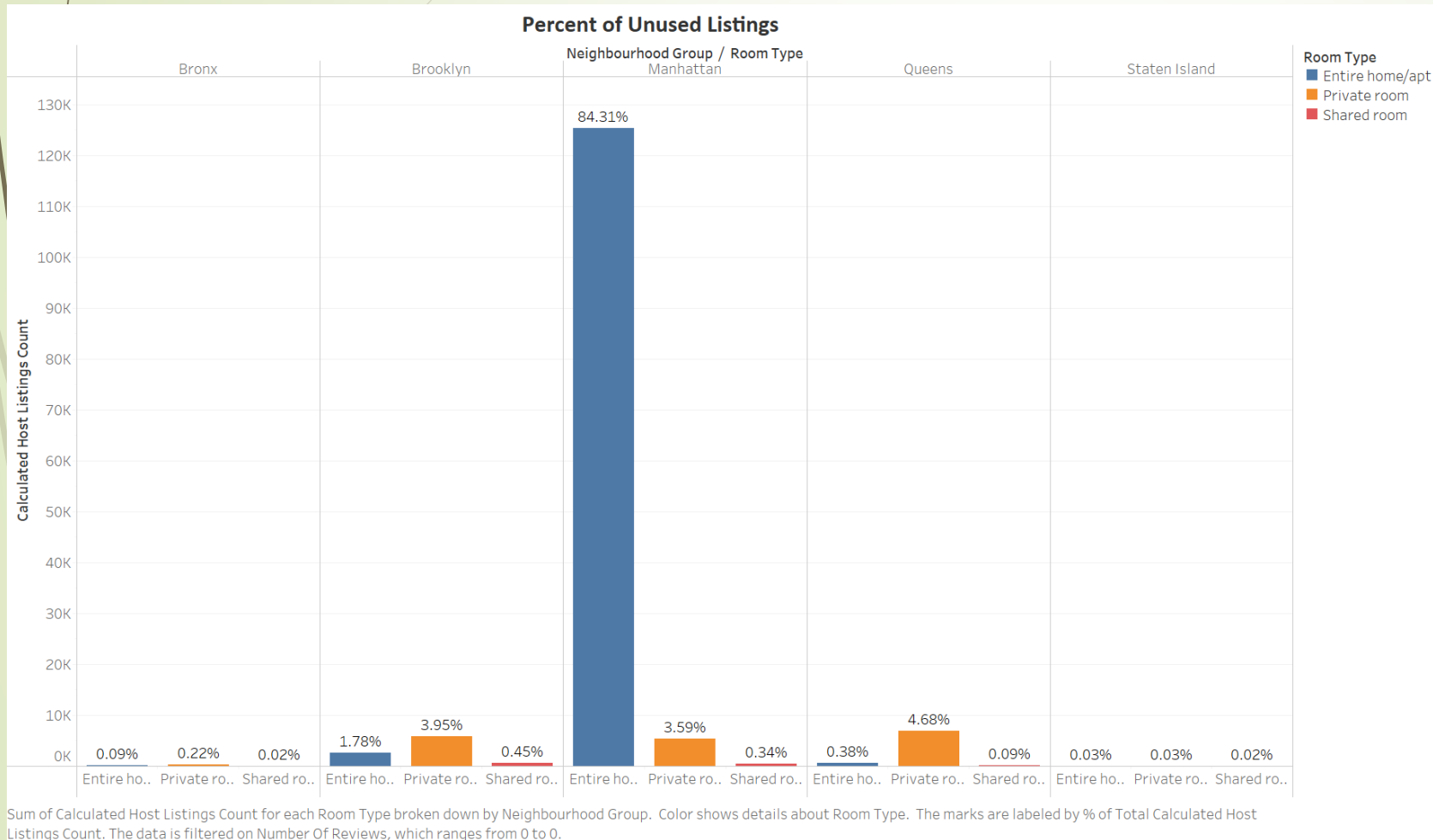
We can see rom this analysis that **_over 120K listings of Entire homes in Manhattan have not generated any revenue_**.

So based on the Customer Preferences seen earlier **_Further Acquisitions of Entire Homes should be targeted in Brooklyn_**.
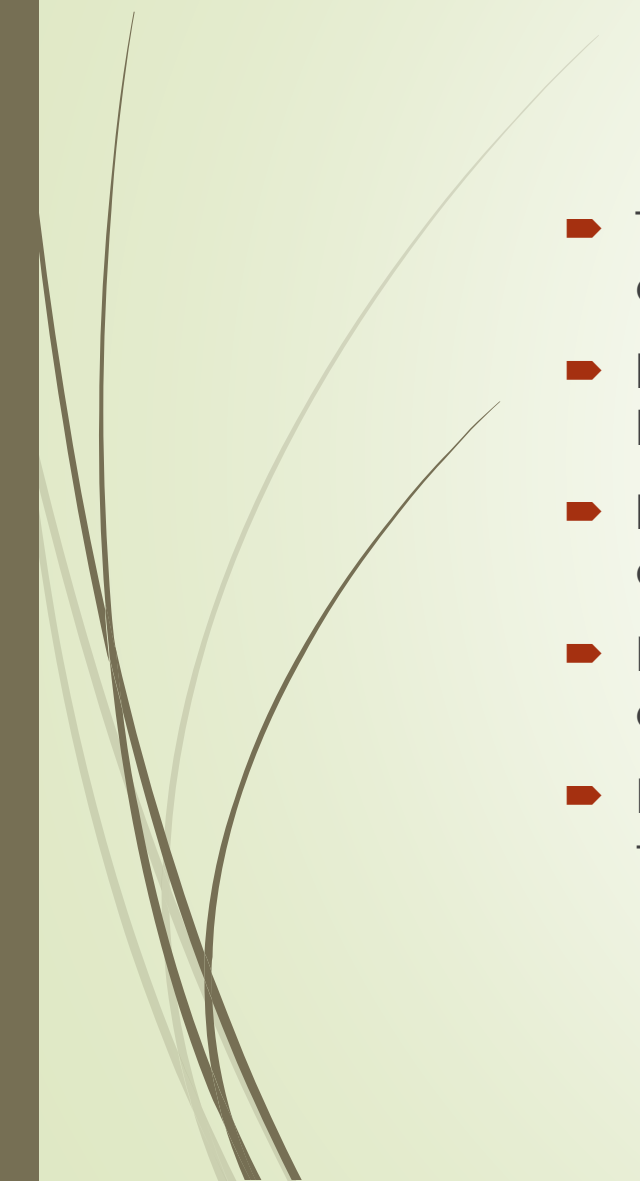
Revenue Analysis

Revenue Generated Yearly
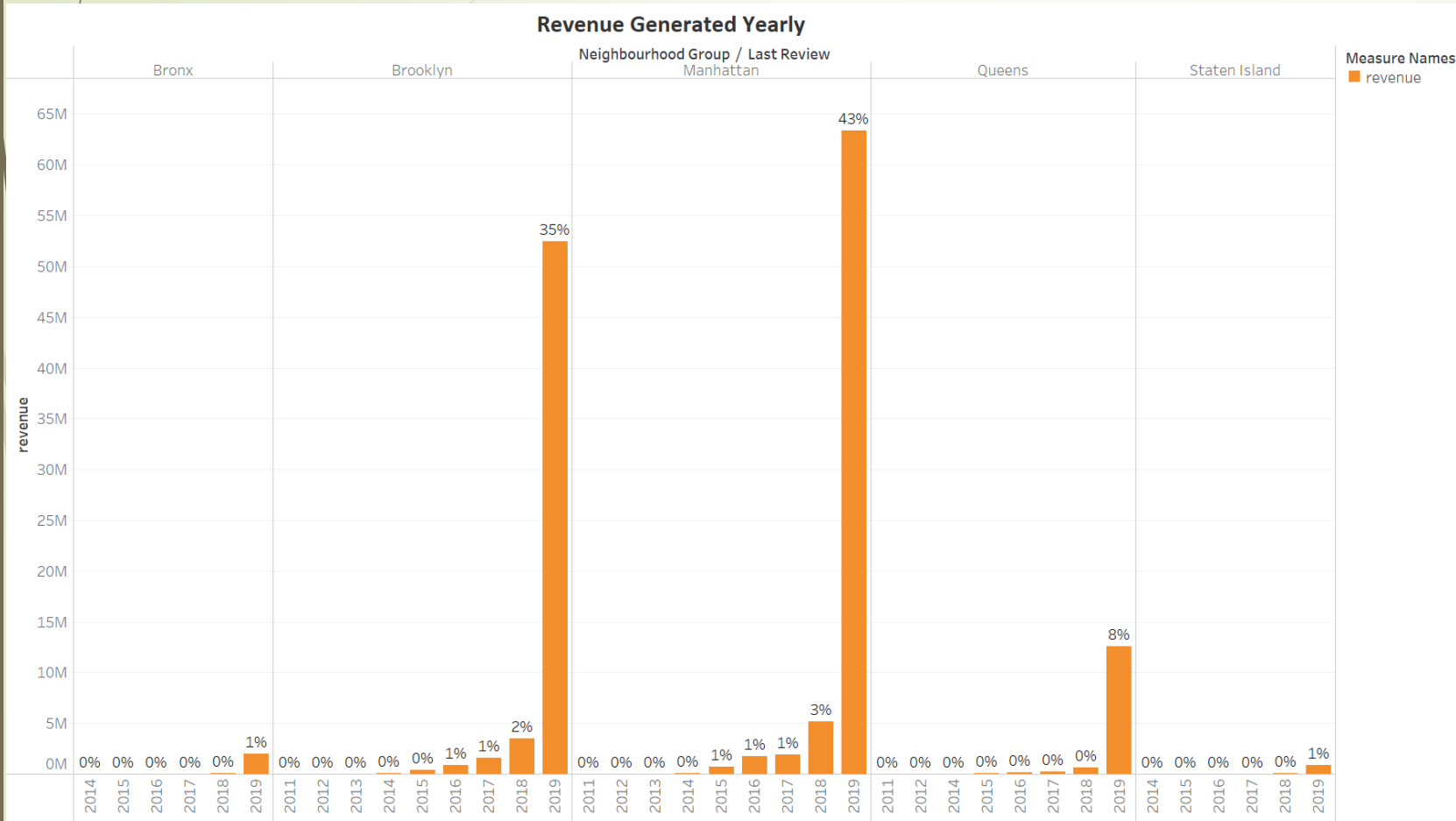
Revenue for each Last Review Year broken down by Neighbourhood Group. Color shows details about revenue. The marks are labeled by % of Total revenue. The view is filtered on Last Review Year, which excludes Null.

- From this graph it is seen that majority of revenue generated in 2019.

- Brooklyn and Manhattan have lion's share in revenue generation.

- Brooklyn and Manhattan have remained as highest revenue generators every year.

Conclusions

- To improve Revenue for upcoming years focus heavily in Neighbourhoods of Brooklyn and Manhattan.

- In Brooklyn acquire more Entire Homes and decrease the availability for bookings.

- In Manhattan acquiring further listings to be prevented and increase in availability for booking.

- Nos of Listings for shared room to be decreased as they are unappealing to customers.

- Reasons for appeal in Neighbourhoods of Bronx, Staten Island and Queens to be investigated as Data insufficient for Analysis.

Appendix

- **DATA CLEANING**

  - In the Dataset columns (name and host_name) contain blanks. These has been filled with Unnamed and Noname respectively as these datapoints need not be unique.

  - Also columns reviews_per_month and last_review contain blanks. Further analysis shows that they coincide with column number_of_reviews having data entry as 0. So they can be left as is.

- **DATA WRANGLING**

  - Column Price is taken as price per booking instead of per night.

  - Column Number_of_Reviews is taken as customers and assumed that all customers have posted reviews so Number of Reviews is equated as number of customers.

  - New Column Revenue Generated by multiplying data point in Number_of_Reviews with data point in price. Logic being evident from previous points.

- **CHARTS**

  - Bar Charts provide the best visuals for most of the insights gathered from the data.

  - Wherever dual-axis charts are used both axis have been synchronized.

  - Labels on the Bars in Bar charts display data as percent of total.

  - In all Bubble charts the size of the bubble displays the varying characteristics like average price, average availability and average minimum nights.

  - In all Bubble charts, color has been used to differentiate between room type, while labels contain neighbourhood name, percent of total visitors for that grouping and varying characteristics.

  - In time series data Null data display has been excluded as they inform about unvisited listings.

  - For looking into data about unused listings a filter of 0 number of reviews was set to fetch only listings data where number of reviews is set as zero.