DSCI 5360 – Data Visualization

Summer 2020 / Final Project

Nishanth Reddy Dyapa (Student ID: 11398415)

Introduction:

AirBnB has seen a meteoric growth since its inception in 2008 with the number of rentals listed on its website growing exponentially each year. AirBnB has successfully disrupted the traditional hospitality industry as more and more travelers resort to AirBnB as their premier accommodation provider. New York City has been one of the hottest markets for AirBnB, with almost 50,000 listings as of November 2019. In this project, I performed an exploratory analysis of the AirBnB dataset sourced from the Inside AirBnB website (http://insideAirBnB.com/get-the-data.html) to understand the rental landscape in NYC through various static and interactive visualizations.

Following are a few questions that we aim to answer through our analysis:

1. What is the pricing landscape of AirBnB listings in New York City?
2. How the pricing is distributed across all the neighborhoods, across individual neighborhoods groups, and individual room types in NYC?
3. Which hosts are receiving most reviews and what is their corresponding average price and average availability in a year?
4. Which of the Hosts are most expensive AirBnBs in NYC?
5. Which neighborhoods have more listings?
6. As a potential customer/ an end user, where can a user get the best pricing and how can they quickly compare neighborhoods in terms of price and ratings?

Data Description:

The data is sourced from the Inside AirBnB website which hosts publicly available data from the AirBnB site. The website has detailed listings data in the city as of September 12, 2019. The dataset comprises of Host name (text), Host Id (ID), Type of Room (Categorical), price (continuous), longitude (continuous), latitude (continuous), neighborhood group (categorical), neighborhood (categorical), Reviews (continuous), Reviews per month(continuous) calculated host listings count(continuous), availability_365 (continuous) among others. The Dataset consists of 48896 unique listings in NYC in total.

All the listings in the NYC are divided into 5 neighborhood groups (Manhattan, Brooklyn, Queens, Bronx, and Staten Island). Also, there are 3 types of rooms in the dataset (Private Room, Shared Room, and Entire home/Apt)

Analysis and Conclusions:

In this section, we will detail our analysis to the questions of interest mentioned in the introduction and gain preliminary insights through exploratory data analysis and visualization.

Sheet 1 (Map) is the basic interactive map with all the AirBnB listings in New York City appearing in a clustered fashion. We can click on each listing to see details like neighborhood groups, Host Name, Type of room, Avg Price of the room, Total reviews. This visualization helps to explore every listing geographically. It gives the overall sense

of how the listings are distributed across neighborhood. We can see from the map that maximum listings are clustered around Manhattan and Brooklyn region, and the least number of listing are in Staten Island.

From sheet 2, 3 & 4 We can say that

1. Pie Chart shows us that entire Home/Apt style listings constitute about 52% of all the AirBnB listings in NYC.
2. Also, Entire Home/Apt style listings are highest in number in Manhattan and Brooklyn. Staten Island, Queens and Bronx have more 'Private room' style property than 'Apartments'.
3. In all the neighborhood groups, shared rooms constitute the least.
4. The maximum apartment style listings are in Manhattan, constituting 61% of all properties in that neighborhood group. Next is Brooklyn with 48% Apartment style listing followed by Queens with 37% Apartment style.
5. Queens and Bronx also have a greater number of private room listings.
6. Sheet 3 represents the count of AirBnBs in the different neighborhood groups. From the plot, we can easily visualize that Entire Home/Apartment are listed most near Manhattan while Private Rooms and Apartments near Brooklyn are nearly equal.

Price Distribution of AirBnB:

From sheet 5 (Price Distribution) we can analyze the following:

In order to analyze the price distribution, I have used calculated field to create a normally distributed curve using the below formula:

(1/ (sqrt ([St Dev] ^ 2*(2*3.14)))) * exp (- (([Price Bin Filled]-[Mean]) ^2 / (2*[St Dev] ^ 2)))

* [Price Bin Size]

* TOTAL (SUM ([Number of Records]))

With the added filters of Room type and Neighborhood group we can analyze the following:

1. The prices of listings range from $10 to $10,000 per night. Half of the listings in NYC are under **$110** a night.
2. The price distribution of AirBnB in **New York city** averages around **$153** per night depending upon the neighborhood.
    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **New York city** averages around **$212, $90, $70** per night respectively depending upon the neighborhood.
3. The price distribution of AirBnB in **Manhattan** Neighborhood group averages around **$197** per night.
    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **Manhattan** averages around **$249, $117, $89** per night respectively depending upon the neighborhood.
4. The price distribution of AirBnB in **Brooklyn** Neighborhood group averages around **$124** per night.
    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **Brooklyn** averages around **$178, $77, $51** per night respectively depending upon the neighborhood.
5. The price distribution of AirBnB in **Queens** Neighborhood group averages around **$100** per night.

    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **Queens** averages around **$147, $72, $69** per night respectively depending upon the neighborhood.
6. The price distribution of AirBnB in **Staten Island** Neighborhood group averages around **$115** per night.
    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **Staten Island** averages around **$174, $62, $57** per night respectively depending upon the neighborhood.
7. The price distribution of AirBnB in **Bronx** Neighborhood group averages around **$87** per night.
    a. The individual price distribution of AirBnB for Apartment, Private Room and shared room in **Bronx** averages around **$128, $67, $60** per night respectively depending upon the neighborhood.

We can conclude that on an average,

1. Getting a shared room is cheapest near Bronx and Brooklyn and costliest near Manhattan
2. Getting a private room is cheapest near Staten Island and costliest near Manhattan
3. Cost of Entire Home or Apartment is Most Expensive near Manhattan whereas Bronx has the cheapest option.

Sheet 6 shows a tabular chart which shows the top N most popular neighborhoods. The ranking is given based on the total reviews the host has received. With the Top N parameter, we can select a dynamic integer to get our desired number of host list based on the reviews. This chart gives us the details about the neighborhood, number of reviews received, Avg. Price of the listing, Avg. Availability in a year, minimum and maximum nights a guest stayed in that listing.

From this chart we can observe that listings in Brooklyn and Manhattan received most reviews. If we are running out of time and want to find a budget friendly listing with many ratings, this chart will be very useful to narrow down the options. Brooklyn is a great choice for cheaper cost per night and higher rating score.

Sheet 7 shows geographical map consisting of the 30 most expensive AirBnBs in NYC. From this we can observe that there is no shared room in this category. Apartment room in Manhattan and Private room in Queens neighborhood group are having highest price of $10,000 per night.

Functions, Calculated fields, Parameters used:

I have used INDEX (), STDEV (), MEAN () functions and calculated fields Price Bin, Price Bin filled, Top N filter, Mean, Median, Normal Distn Curve and others to visualize the data.

I have also used 3 parameters.

Dashboard 1: AirBnB NYC stats

This dashboard summarizes the information already given in the above sheets which shows the distribution of listings across five different neighborhood groups depicted in geographical map and also the distribution of room types in NYC shows as a bar chart. This also consists if summary of the basic AirBnB data like percentage of different room types, Avg. Price of AirBnB rooms in NYC, Total No.of Reviews. We can also filter out the neighborhood groups to see the average price and total reviews of a host.

Dashboard 2: Room selection

For an end user/ customer, this dashboard helps to select a room based on many factors like they can choose a room to stay based on location, reviews, avg. price, avg. reviews per month. This helps user to select a room of his own choice. The filters help them to filter out the options which they do not need. This dashboard is one of the main dashboards among the three I built. We can be able to see top N most popular hosts and their avg price. The interactive map also shows the exact location of the listing selected.

Dashboard 3: Types of rooms

This dashboard allows user to select any number of neighborhoods with most listings in that neighborhood. This helps us in knowing the number of listings in that area, the number of reviews they are getting and the average price in that neighborhood.

Story: The story represented in helps customer/ Business user in making decision to select an AirBnB listings in NYC based on various criterion. This also helps them to see the listings which are not getting many reviews and not becoming popular. Many guests will give reviews based on the room cleanliness, hospitality etc.

Business Implications:

With a dataset of almost 49000 listings representing the AirBnB market in New York city in 2019, there is a need to draw insights from it which helps the end users or the management to make decisions on increasing profits, customer satisfaction, getting more reviews etc. New York being the busiest city in USA, guests who come to visit on business purpose and travel purpose, location makes one of the major impacts in choosing the stay. We can see that majority of the AirBnB listings are in Manhattan and Brooklyn which are the places where majority of business complexes and tourist spots and subways connecting to major part of the city are located.

The maximum number of rooms listed on Airbnb are private rooms and entire home and apartments and a very small number of shared rooms are listed on Airbnb. Maximum number of people love to stay at Manhattan and Brooklyn. The reason behind their popularity may depend upon the price, neighborhood, cleanliness and many more. This helps management to concentrate on increasing hosts in and around this area which improves the business. This also helps in identifying why the other neighborhoods are lacking popularity. If the company focuses more on increasing the listings, my suggestion is to increase the number of shared rooms because at a tourist place budget plays the main role for many travelers. A neatly maintained shared rooms with many good reviews will be the busiest room which brings us more revenue.

From the guest perspective, choosing a room of his choice will depend on budget, duration of stay, location of the stay and many more. This dashboard helps guests in considering all these factors and finally deciding a place to stay. For example, most travelers, backpackers and people on low budget like to stay in the shared rooms. They live on an average of 1-2 days as they keep on moving from one place to another. If this is the choice, the lowest priced shared room are present in Bronx and Brooklyn neighborhood. People love to stay in the entire home or apartments when they travel with the family or friends. They can prepare their own meals if they want in these apartments.

Fig 1: Geographic map showing the distribution of AirBnB listings in NYC

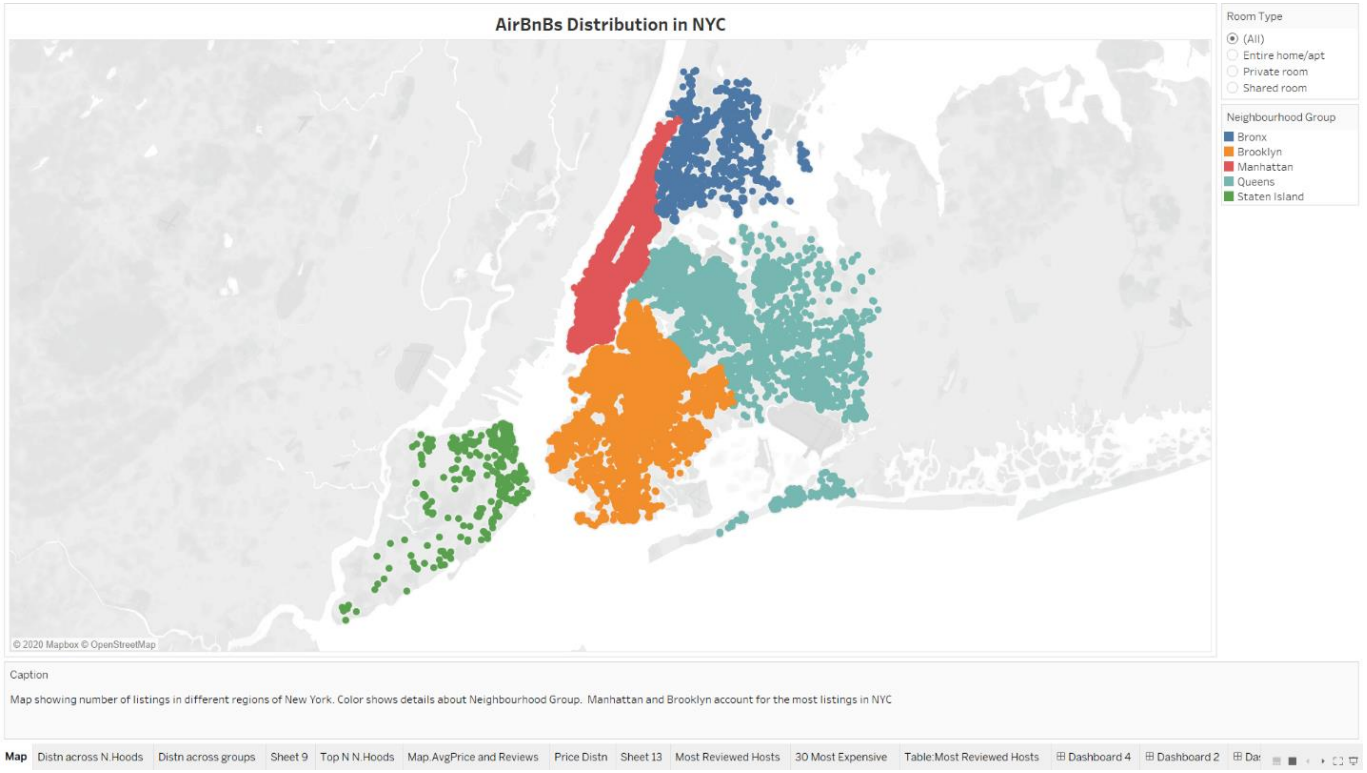


Fig 2: Pie Chart showing the percentages of types of rooms in NYC and also in individual neighborhood groups

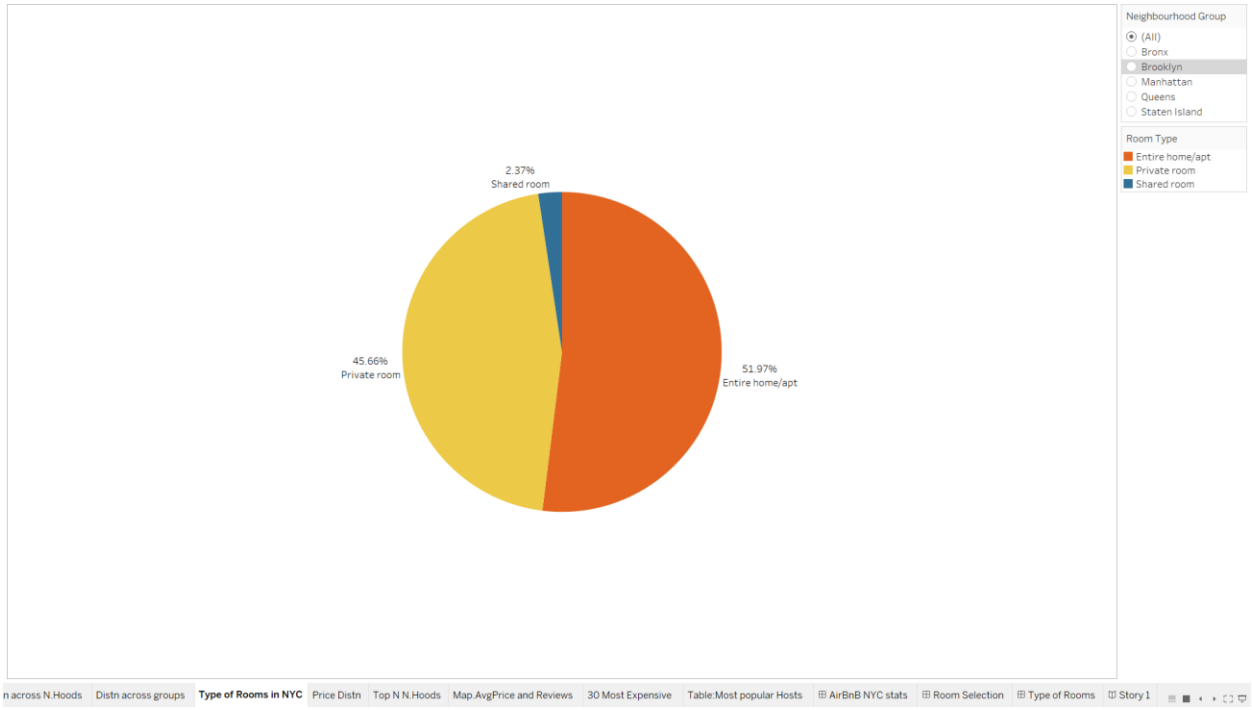


Fig 3: Bar Chart showing the number of each type of room in each neighborhood group in a single view.

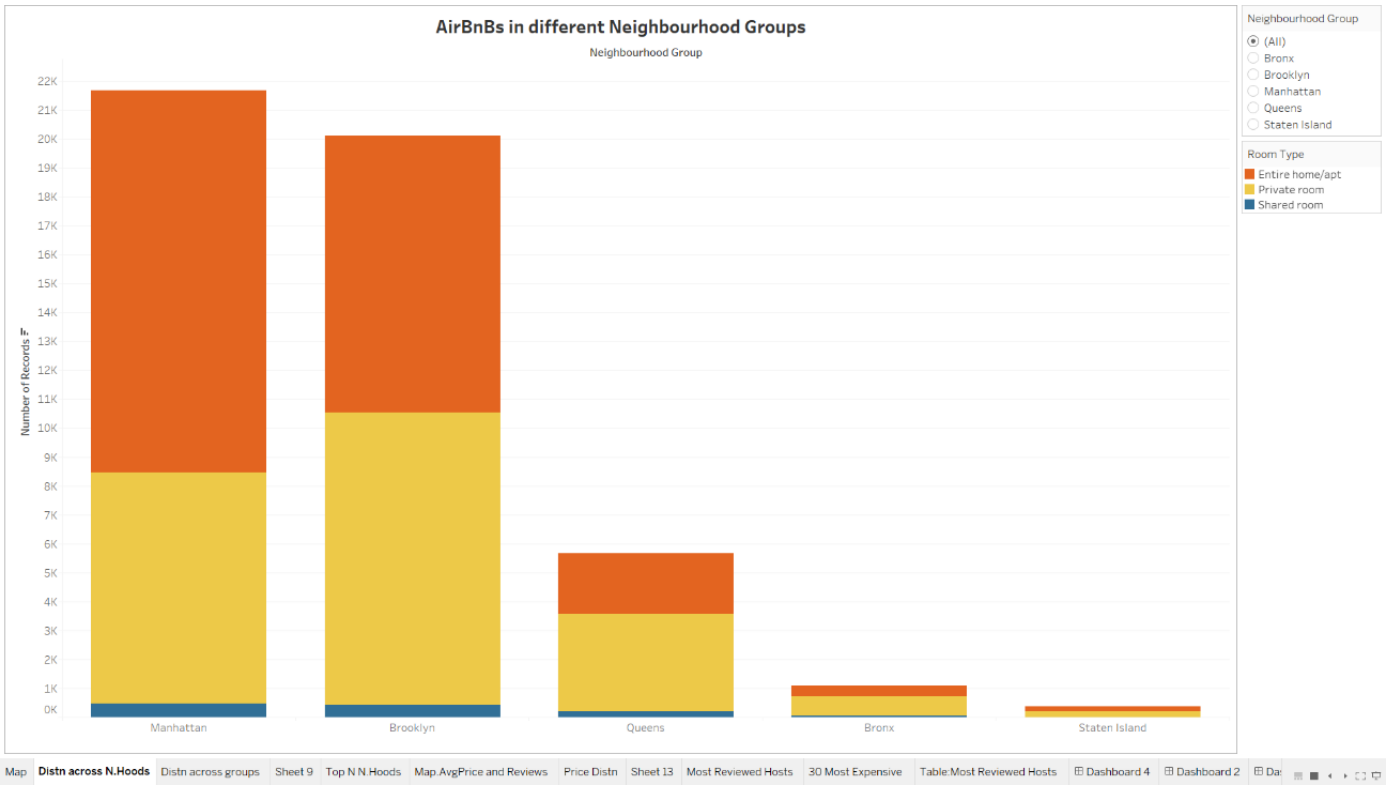


Fig 4: Bar Chart showing the number of each type of room in each neighborhood group in a single view.

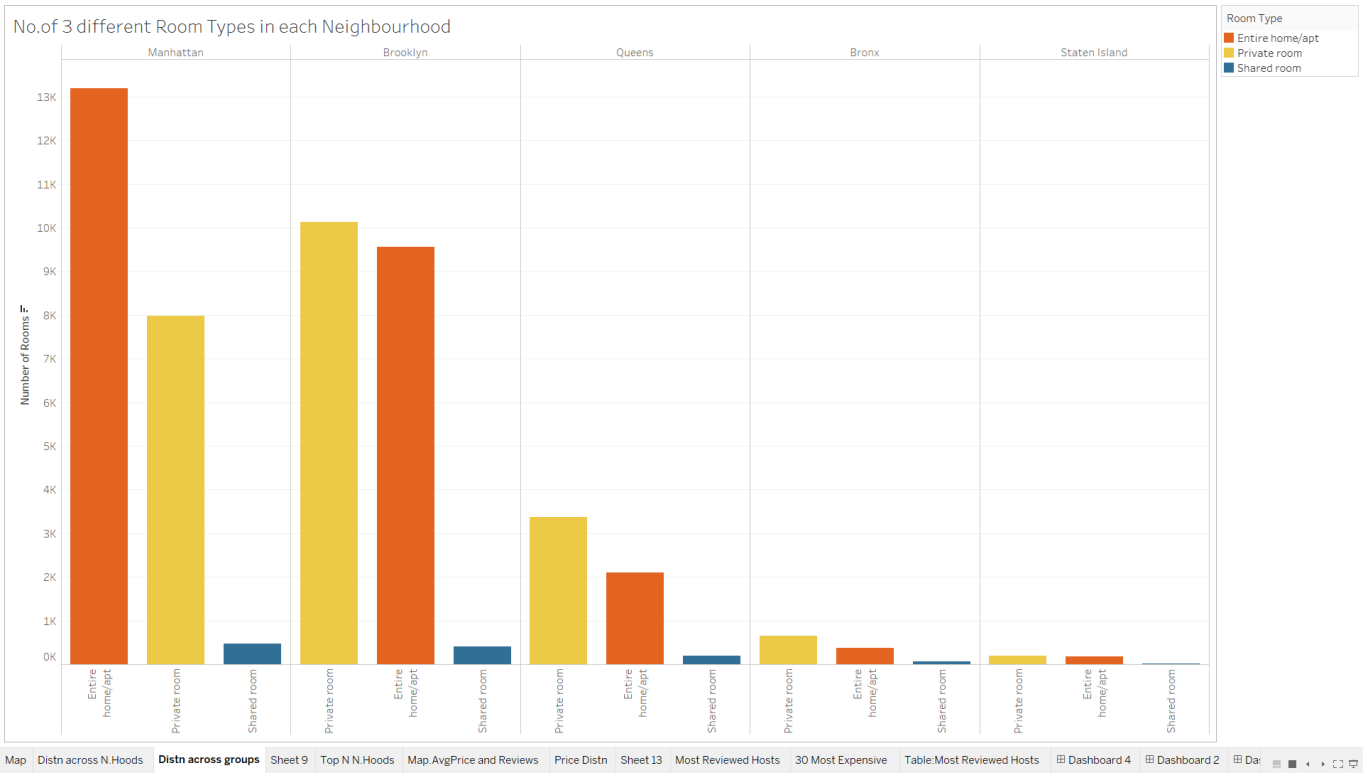


Fig 5: Price distribution of all type of rooms in all neighborhood groups in NYC

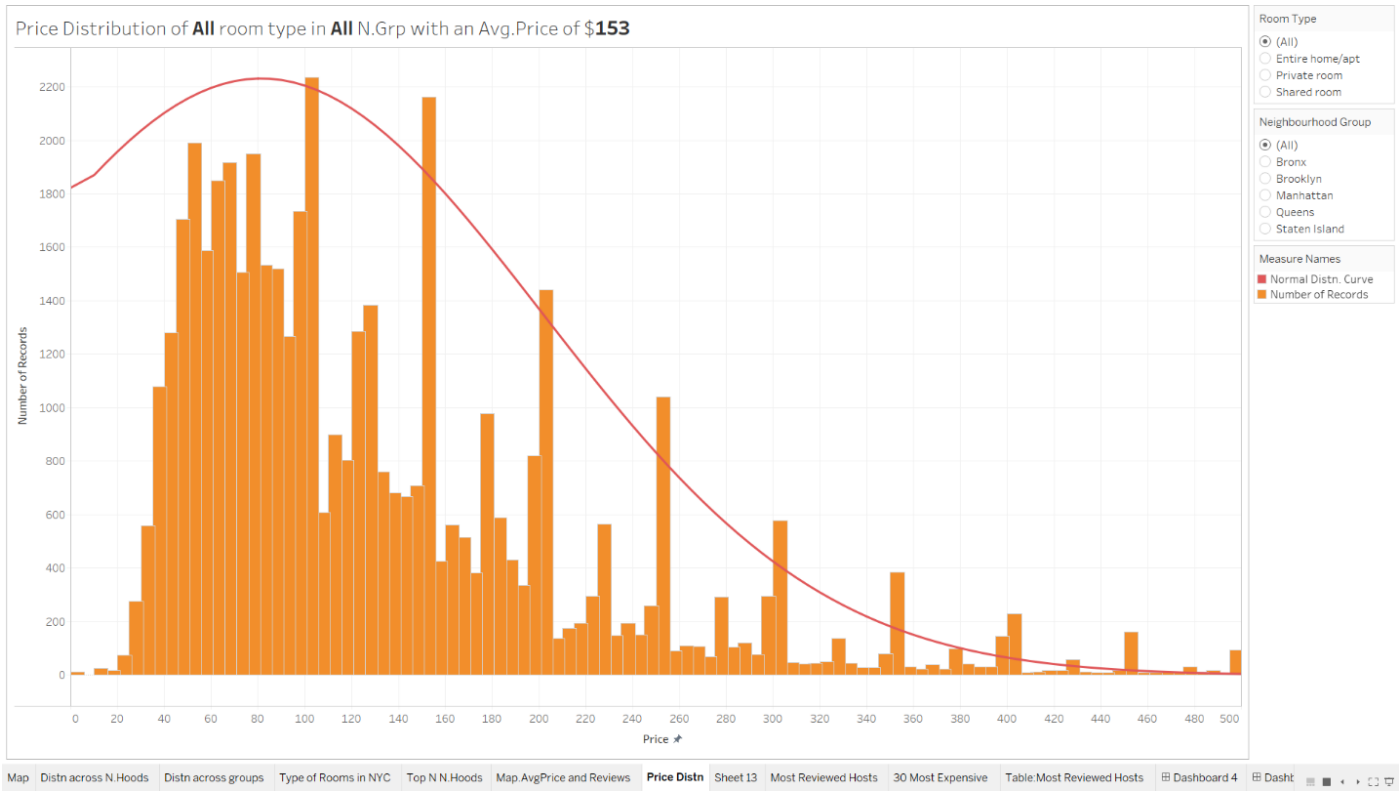
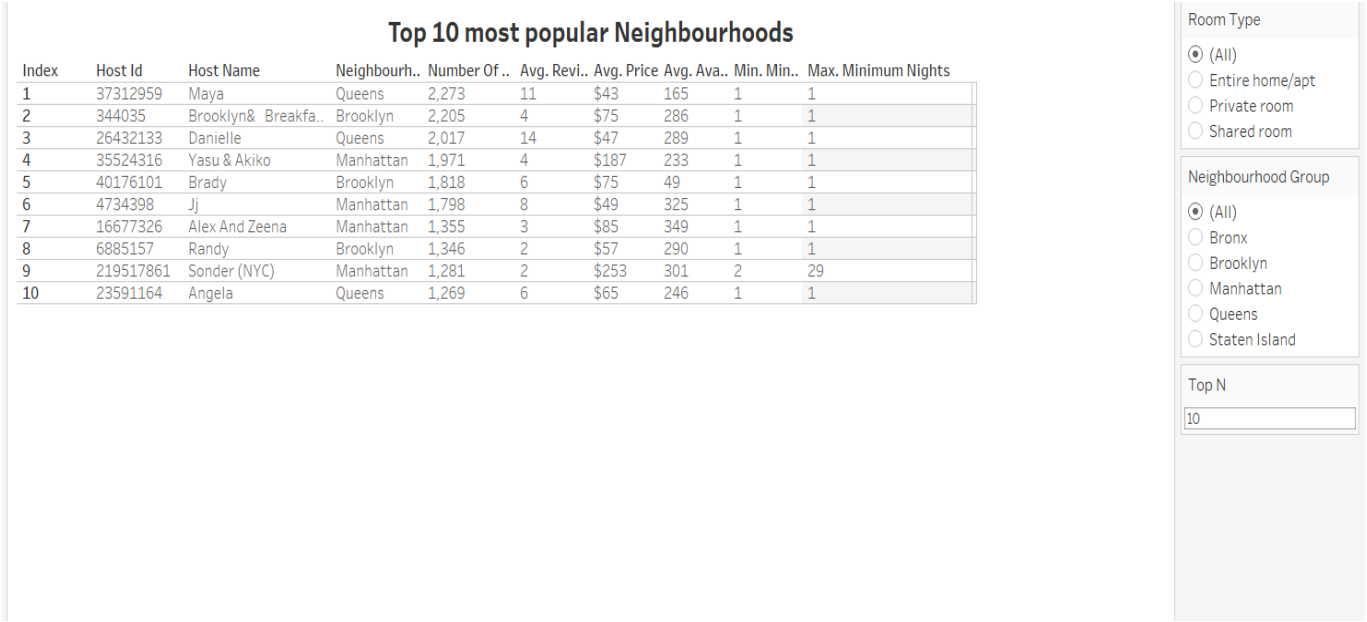


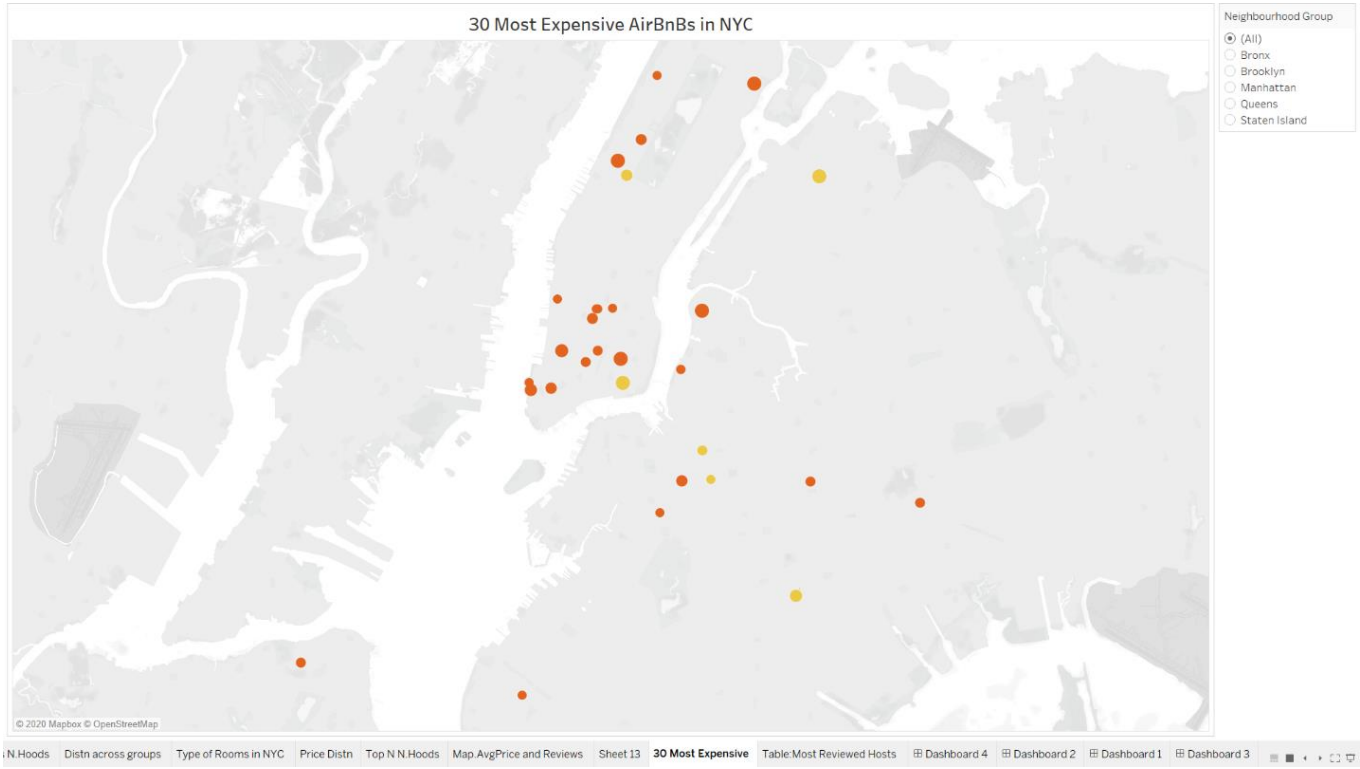
Fig 6: This visualization showing the Top N most popular neighborhoods based on the reviews each host got.



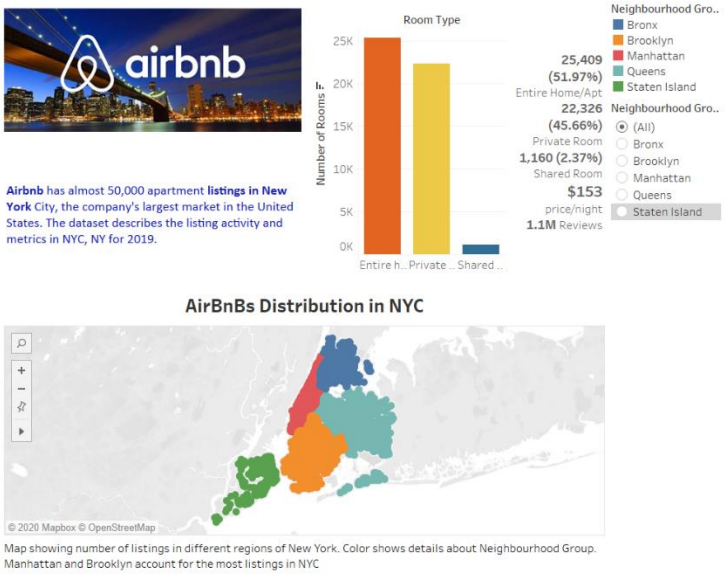
Top 10 most popular Neighbourhoods

Index	Host Id	Host Name	Neighbourh..	Number Of ..	Avg. Revi..	Avg. Price	Avg. Ava..	Min. Min..	Max. Minimum Nights
1	37312959	Maya	Queens	2,273	11	$43	165	1	1
2	344035	Brooklyn& Breakfa..	Brooklyn	2,205	4	$75	286	1	1
3	26432133	Danielle	Queens	2,017	14	$47	289	1	1
4	35524316	Yasu & Akiko	Manhattan	1,971	4	$187	233	1	1
5	40176101	Brady	Brooklyn	1,818	6	$75	49	1	1
6	4734398	Jj	Manhattan	1,798	8	$49	325	1	1
7	16677326	Alex And Zeena	Manhattan	1,355	3	$85	349	1	1
8	6885157	Randy	Brooklyn	1,346	2	$57	290	1	1
9	219517861	Sonder (NYC)	Manhattan	1,281	2	$253	301	2	29
10	23591164	Angela	Queens	1,269	6	$65	246	1	1

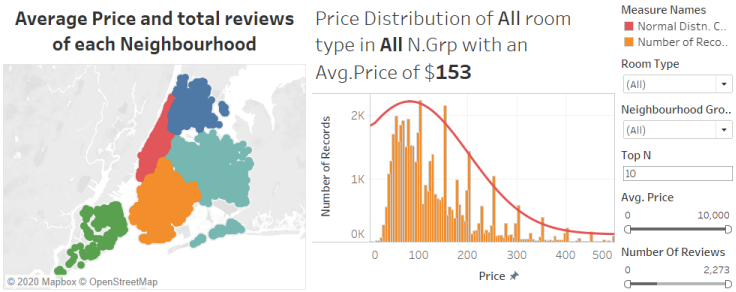
Fig 7: This visualization showing the top 30 most expensive AirBnBs in NYC and their reviews.



Dashboard 1: Shows the overall stats of the AirBnBs in NYC

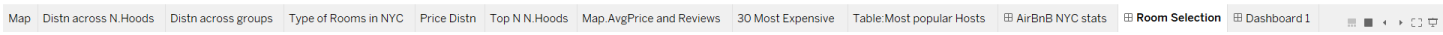


Dashboard 2: Showing the price distribution and the average price, reviews for the results which we filtered.

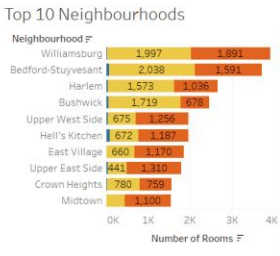


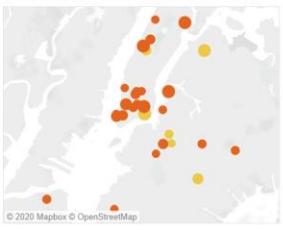
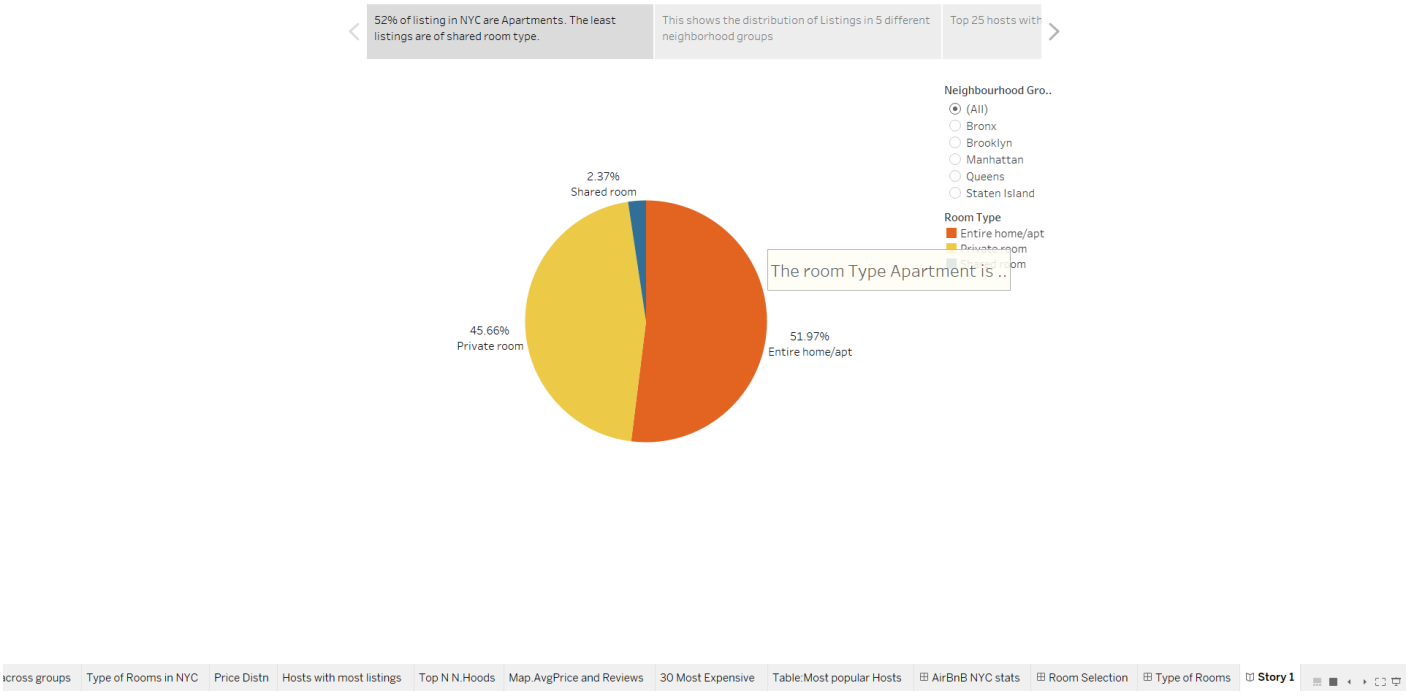
Top 10 most popular Hosts

Index	Host Id	Host Na..	Neighbo..	Number Of ..	Avg. Revi..	Avg. Price	Avg. Ava..	Min. Min..	Max. Minin
1	37312959	Maya	Queens	2,273	11	$43	165	1	1
2	344035	Brooklyn..	Brooklyn	2,205	4	$75	286	1	1
3	26432133	Danielle	Queens	2,017	14	$47	289	1	1
4	35524316	Yasu & A..	Manhatt..	1,971	4	$187	233	1	1
5	40176101	Brady	Brooklyn	1,818	6	$75	49	1	1
6	4734398	Jj	Manhatt..	1,798	8	$49	325	1	1
7	16677326	Alex And ..	Manhatt..	1,355	3	$85	349	1	1
8	6885157	Randy	Brooklyn	1,346	2	$57	290	1	1
9	2195178..	Sonder (..	Manhatt..	1,281	2	$253	301	2	29
10	23591164	Angela	Queens	1,269	6	$65	246	1	1



Dashboard 3: Type of rooms, their count and their location in NYC





Story: Story of AirBnB in NYC showing the important facts.