# Tableau Project – Option2 – Airbnb

Joshua Ng – DA Flex – Jan. 20, 2025

## Project Goals:

The scope of this project was to show familiarity and proficient ability in working with Tableau and a given data set to graphically present meaningful correlations between the data. the visualizations that were created mainly pertained to location, cost, and listing reviews. The overarching question was whether there were any correlations between the host, the location, and the amenities (beds, room type, property type, cost) with the review ratings.

## Process:

Some initial assumptions were made in preparation to interpret the data. Namely that the "Prices" data listed in the dataset are PER NIGHT, and that the "Host Since" data listed also represents the initial date that the associated listing was put into service. Additionally, the data in the provided spreadsheet was cleaned prior to being connected to in Tableau. Once this was done, the data could loaded/connected to within Tableau.

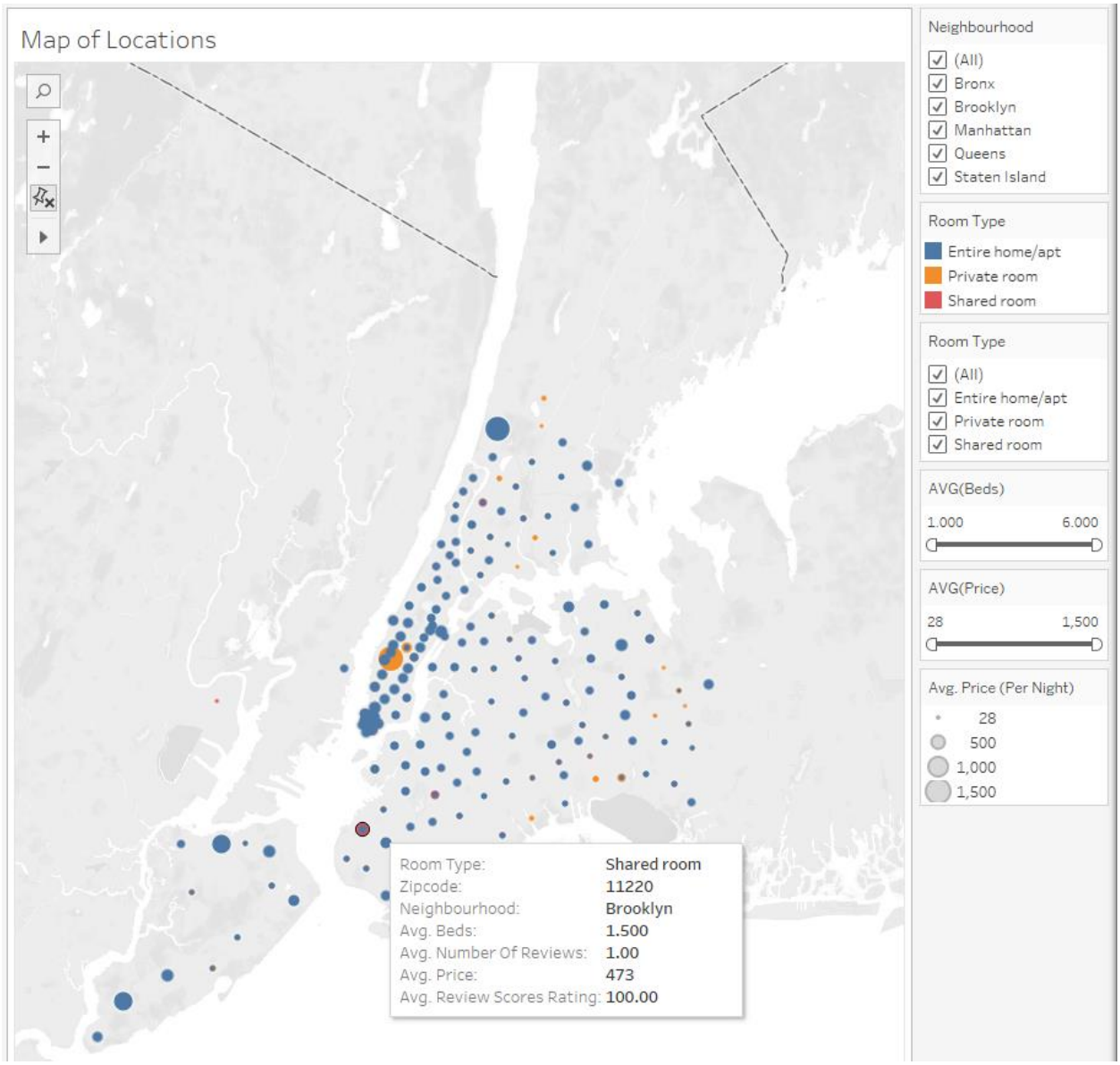
Once in Tableau, the data was analyzed to verify the different types that were present. Additionally, a second round of filtering within Tableau was performed to further streamline and make consistent the data to be interpreted.

Visualizations were then built according to the data's categorical features. Relationships with meaningful information that could be gleaned were used to generate graphical representations of their correlations. These were pondered and informed the formation of the Dashboard that would be created using these Worksheets.

## Result:

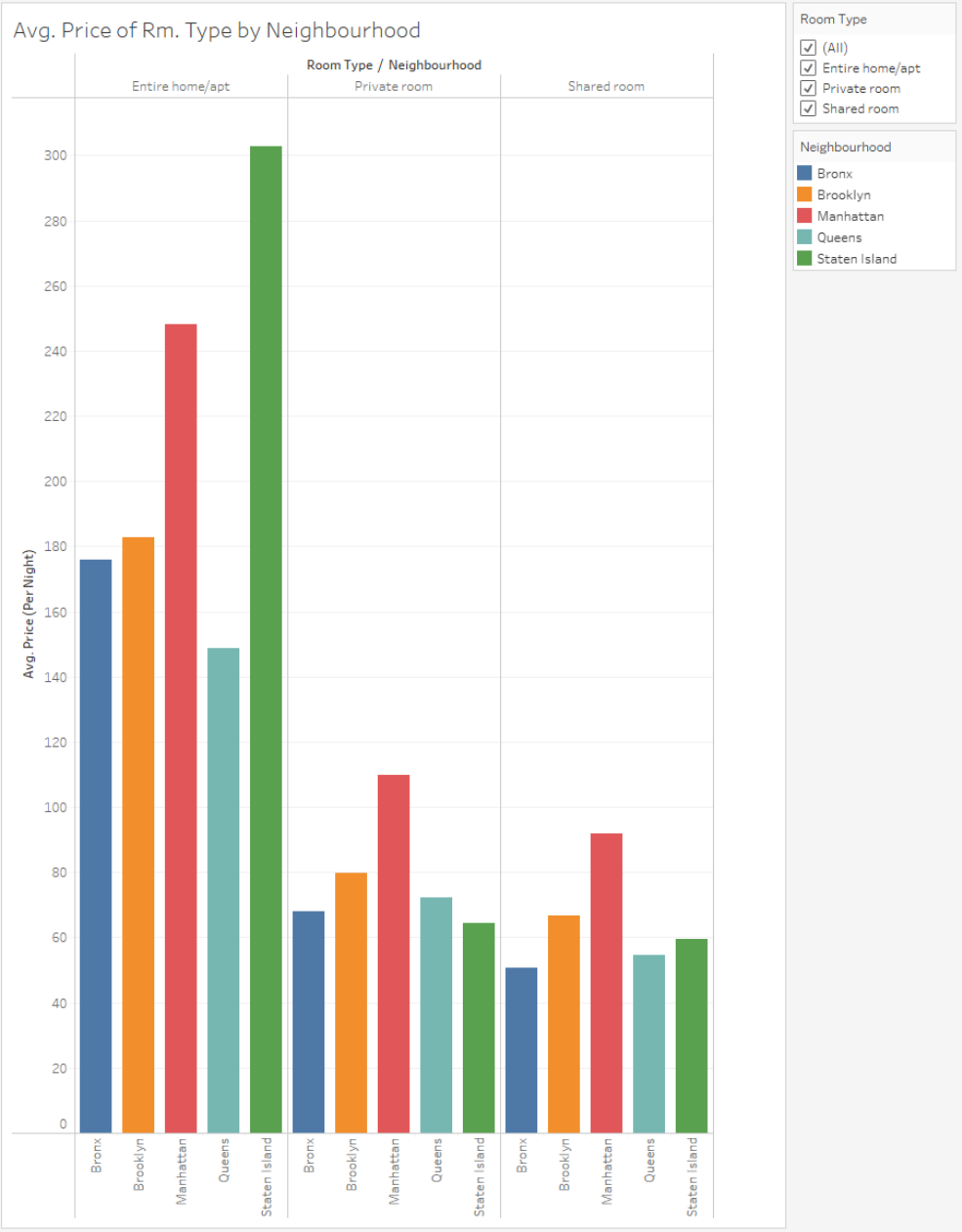
Five visualizations were created to convey and correlate the information in the dataset; they are as follows:

## Map of Locations




The first graphic was a map of New York City, with each listing location based on Zip Code (as no specific coordinates or addresses of each property were provided/available) represented. Categorical filters such as the Neighbourhood they were in, the types of rooms offered, number of beds, and average cost per night were also applied to allow for a flexible viewing experience. This map would provide summarized info on the listings in the area with specifics listed on a Tooltip when hovering a cursor over the location dot, as pictured.

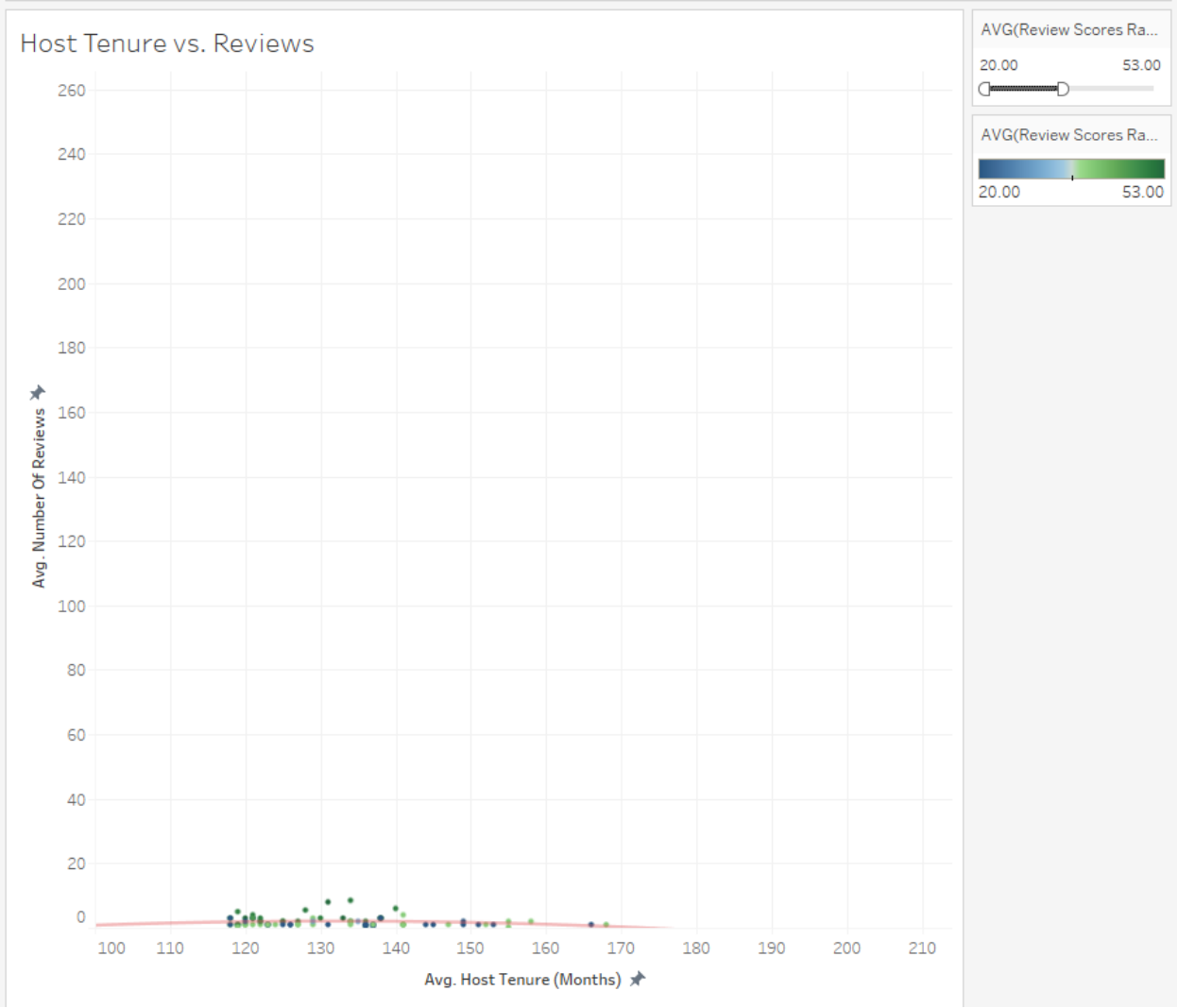
## Avg. Price per Night by Room Type & Neighbourhood




The second visualization was a side-by-side bar graph to show the average price per night of each room type within each neighbourhood. This format was selected to better compare at a glance what the pricing trends may be, as impacted by these 3 factors. As one would expect, the Entire Home/Apt. Room Type is the more expensive option, with the ones on Staten Island being by far the costliest. However, it is also interesting to note that Manhattan, regardless of the Room Type, consistently remains one of the most, if note the most, expensive neighbourhoods. This comes as no surprise as it is quite literally the central core of New York City and so this outcome is as expected.

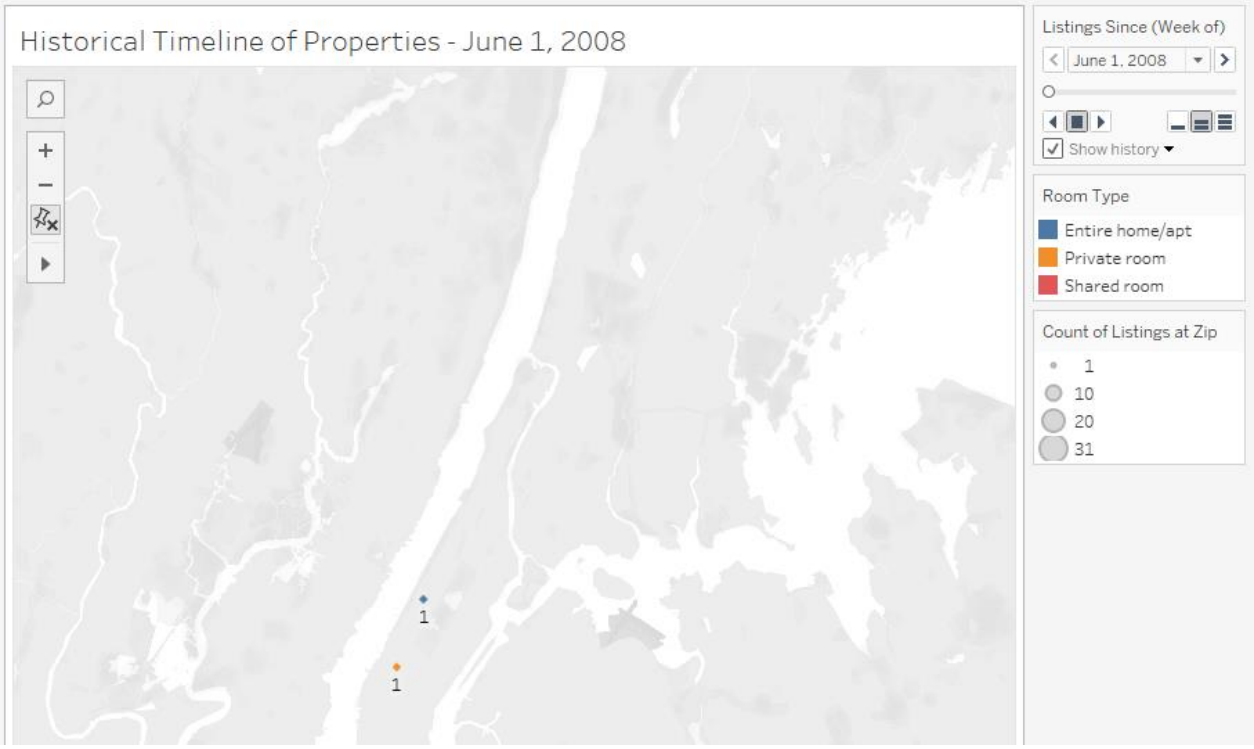
## Hosts Tenure Vs. Reviews


The third visualization was a scatter plot comparing the hosts' tenure as Airbnb hosts with their reviews (both number of, and review scores). This was to try to glean a trend as to whether longer tenured hosts would have more, and higher rated, reviews. Based on the data shown, it certainly seems like this is the case. Hosts who have been hosting Airbnb locations longer tended to have higher overall review rating scores, in addition to more reviews in general. This intuitively does make sense as it would likely mean more people have rented their listings. The trendline confirms this as well, showing an upward curve in number of reviews as the Hosts' Tenure increased.




What is also interesting – as can be seen here – is that the lower scoring reviews DOES also tend to skew towards Hosts who have had a shorter tenure. One feature of this graphic is a colour-range that was used to denote the ratings from each review. When using the slider to filter down the ratings only to those that were ~50/100 or below, the scatter plot only populates on the lower half of the length of Host Tenure. This conclusion can be seen quite easily by making use of this visualization's filtering slider capabilities.

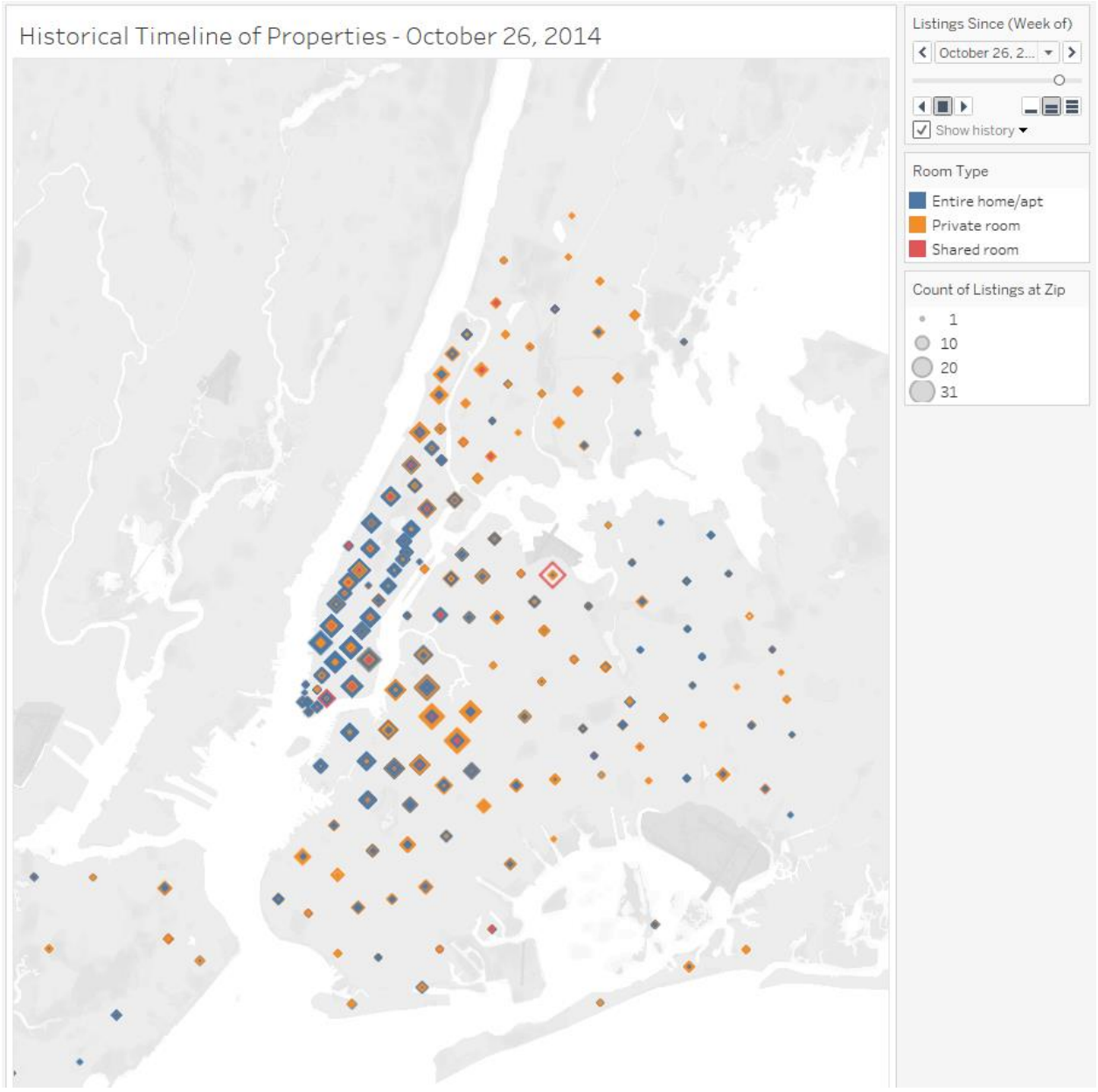
## Historical Timeline of Properties



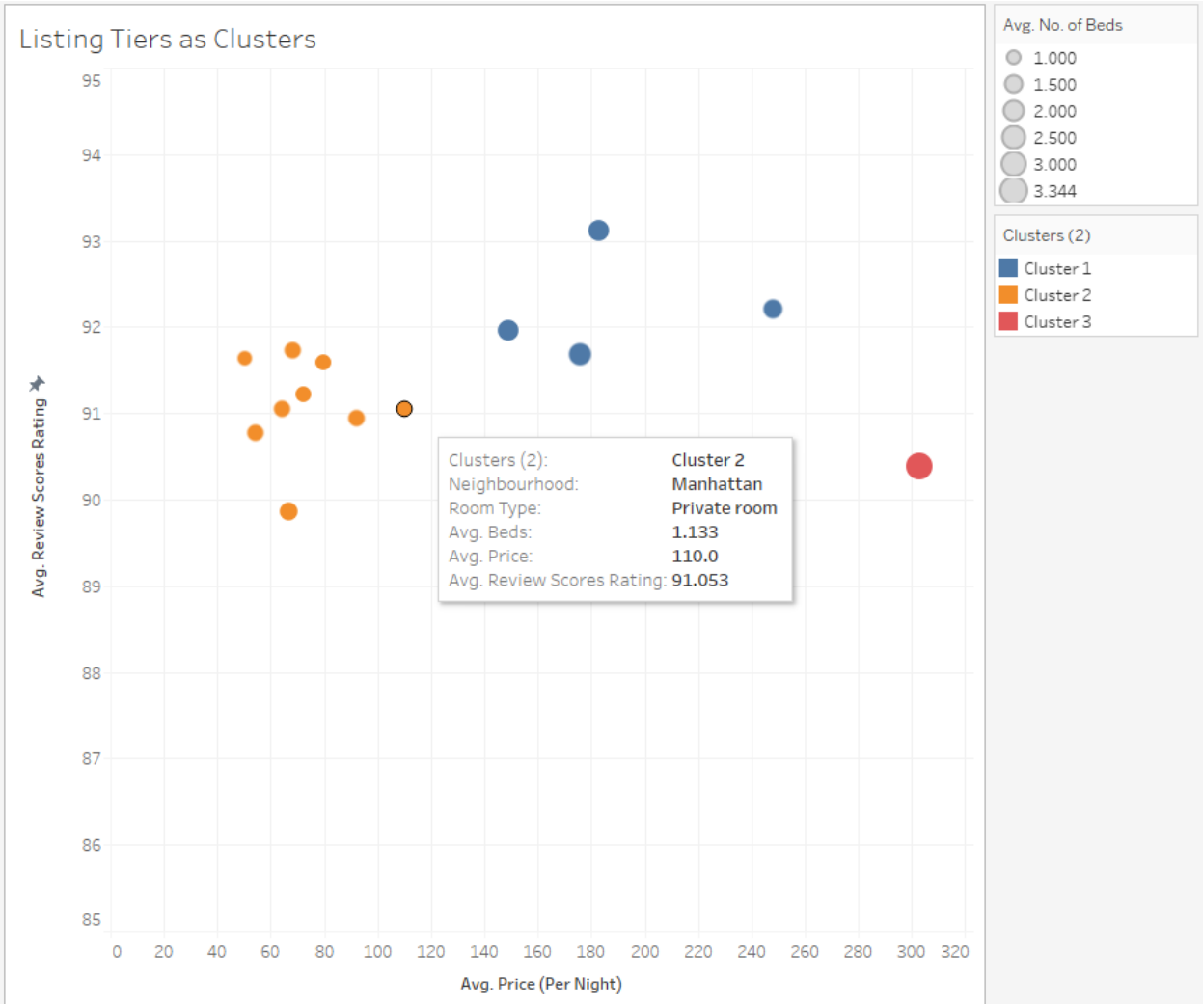

The fourth visualization was an animated historical timeline superimposed onto a map of the NYC area. This was to show the number of Airbnb listings that have been put into service over time. At a glance, this can easily tell which areas, zip code locations, and room types became more popular to host, over time.

For example, the image above shows the map at the start of the timeline (June 1, 2008). There seems to be only 2 Airbnb listings at this time – 1 “Entire Home/Apt” type unit, and 1 “Private Room” type unit in the Manhattan area.

However, as time goes on towards the end of our data timeline (Oct. 26, 2014), many other units have come into service since (please see image below). Not only are they in various other neighbourhoods at this point, a much larger variety of room types is available across NYC as well. The size of the markers also denotes the “density” or the number of listings at that zip code (again, as no specific addressed or coordinates were available). This was achieved by tracking the listings for each zip code in a rolling distinct count of the names of the listings.


Historical Timeline of Properties - October 26, 2014
Listings Since (Week of)
October 26, 2...
Show history
Room Type
Entire home/apt
Private room
Shared room
Count of Listings at Zip
1
10
20
31

## Listing Tiers as Clusters




The fifth visualization was a cluster scatter plot. This made use of the clustering technique for listings to better categorize the "tiers" of the Airbnb rooms. It clustered number of beds, price, and review scores to compare with the average price per night to graphically show correlations to the factors that may impact the popularity and the price.

As seen at a simple visual glance, these do seem to correlate. “Cluster 2” is a collection of listings with relatively high review scores, but lower per night costs, and lower number of beds. These would put them as a more “budget” option. “Cluster 1” is middling and so would be considered a “mid-tier option”, and “Cluster 3” would be considered a “luxury” option. This does make intuitive sense as the listing(s) in “Cluster 3” are represented with a larger marker (meaning more beds per listing) and at a higher cost. It’s interesting to note, however, that the review score isn’t necessarily higher for such an Airbnb location. Rather the “mid tier” options tend to have higher review scores overall, relative to the other two.

# Challenges & Future Goals

The main challenge with this project was becoming familiar with Tableau's capabilities and making use of it to answer the questions that its graphics would be best suited to do so. The UI of Tableau makes sense but is not immediately intuitive due to the sheer scope of the capabilities that it holds as a graphical data representation tool.

If there was more time, I would use it to experiment more with Tableau, to see what other ways its features can be combined and used in tandem to represent more robust, flexible, and nuanced visualizations. Additionally, while the visualizations looked presentable, I would like to explore how best to tweak the representations beyond the defaults to really tailor the graphical theming to a specific audience.