New York City Airbnb Design Study

Dataset

The source of the dataset is Kaggle, it contains 16 attributes and 48,895 items. The attributes consist of listing ID, name of listing, host id, host name, neighbourhood_group(borough), neighbourhood, latitude, longitude, price, room type, minimum nights, number of reviews, last review data, reviews per month, calculated host listings count, and availability year round. Each item is a unique listing within New York City.

Questions

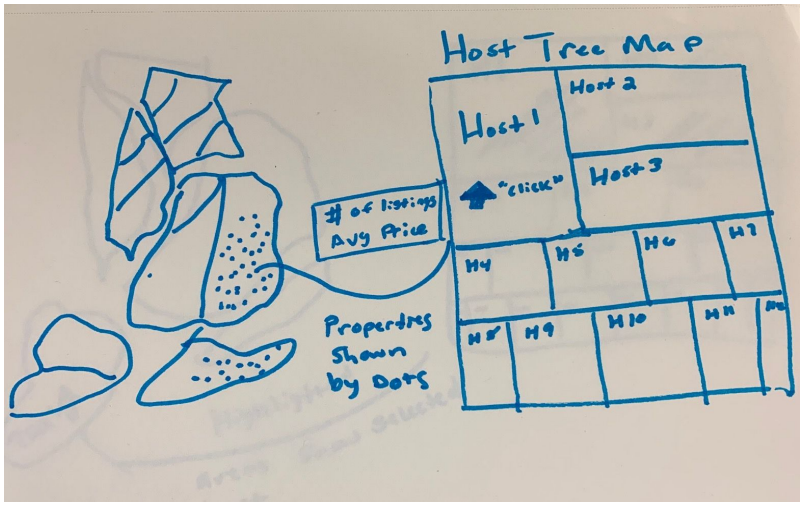
How does the amount of listings in a borough or neighbourhood effect pricing?

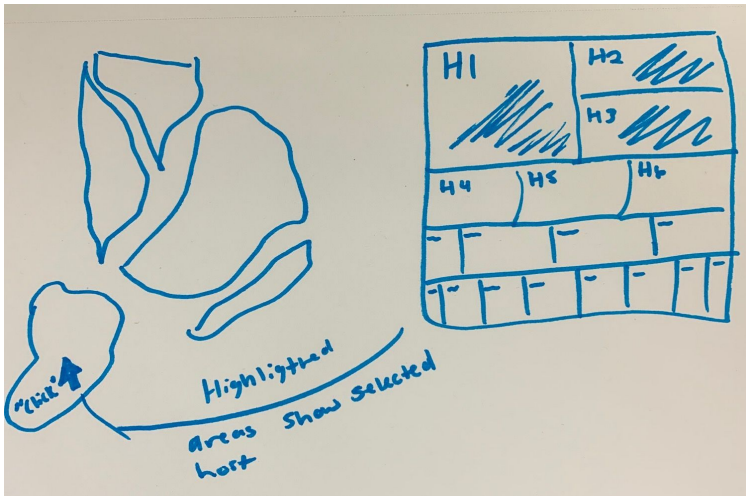
Do host who list multiple properties usually have their listings in the same area or are they scattered amongst NYC?

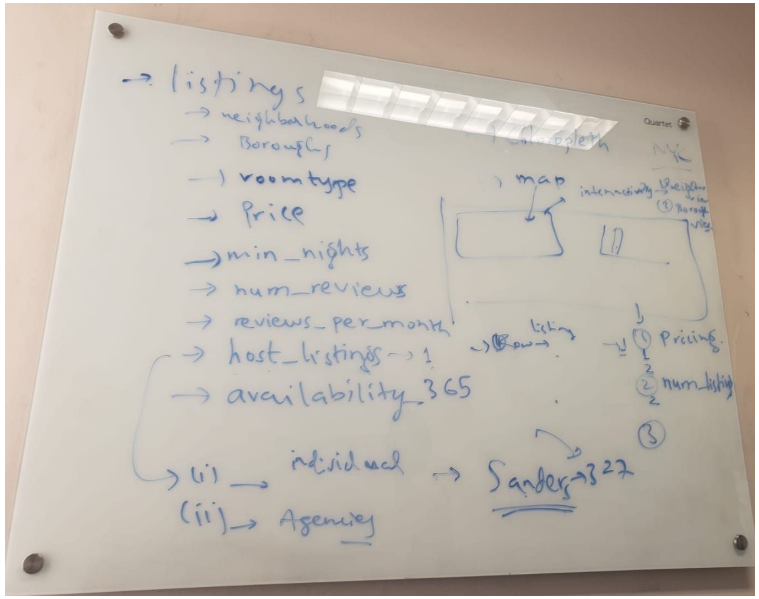
Design Process:

Before we began to construct our visualizations, data processing was required, this consisted of removing null values in the data, which resulted in 31,355 remaining objects. The attributes we decided to focus on to answer our questions were room type, pricing, neighbourhood,listing ID, name of listing, calculated host listings count, and host id. Our original approach to answer the questions was to create a multi view visualization, that consisted of a tree map that shows which host has the most properties by interactive with a choropleth map showing the number of pricing and listings in an area. Below is a sketch of what we first envisioned and our list of ideas.

Sketches:

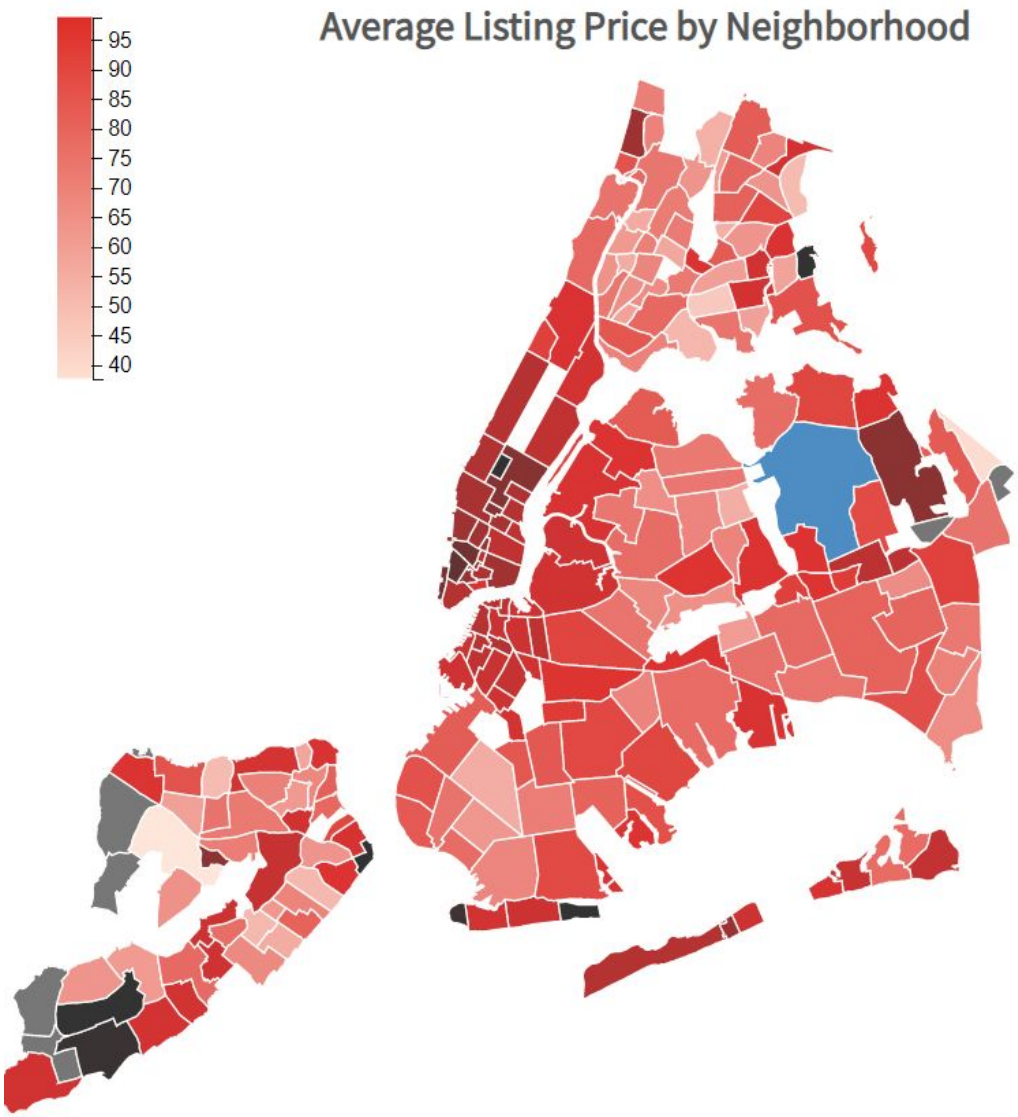

Host Tree Map
Host 1
Host 2
Host 3
"click"
# of listings
Avg Price
Properties Shown by Dots

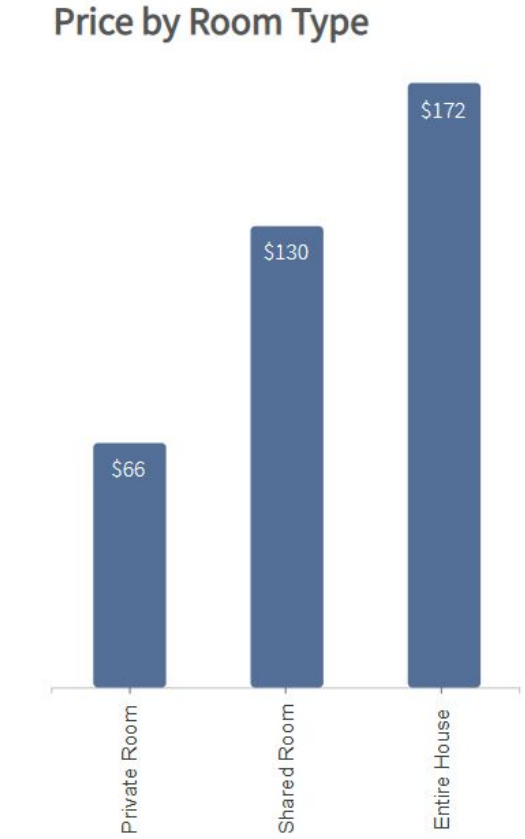

H1
H2
H3
H4
H5
H6
"click"
Highlighted
areas
Show selected
host


listings
neighbourhoods
Boroughs
room type
Price
min_nights
num_reviews
reviews_per_month
host_listings → 1
availability_365
map
Pricing
individual
Sanders
(i)
(ii) Agencies

First Iteration Results:

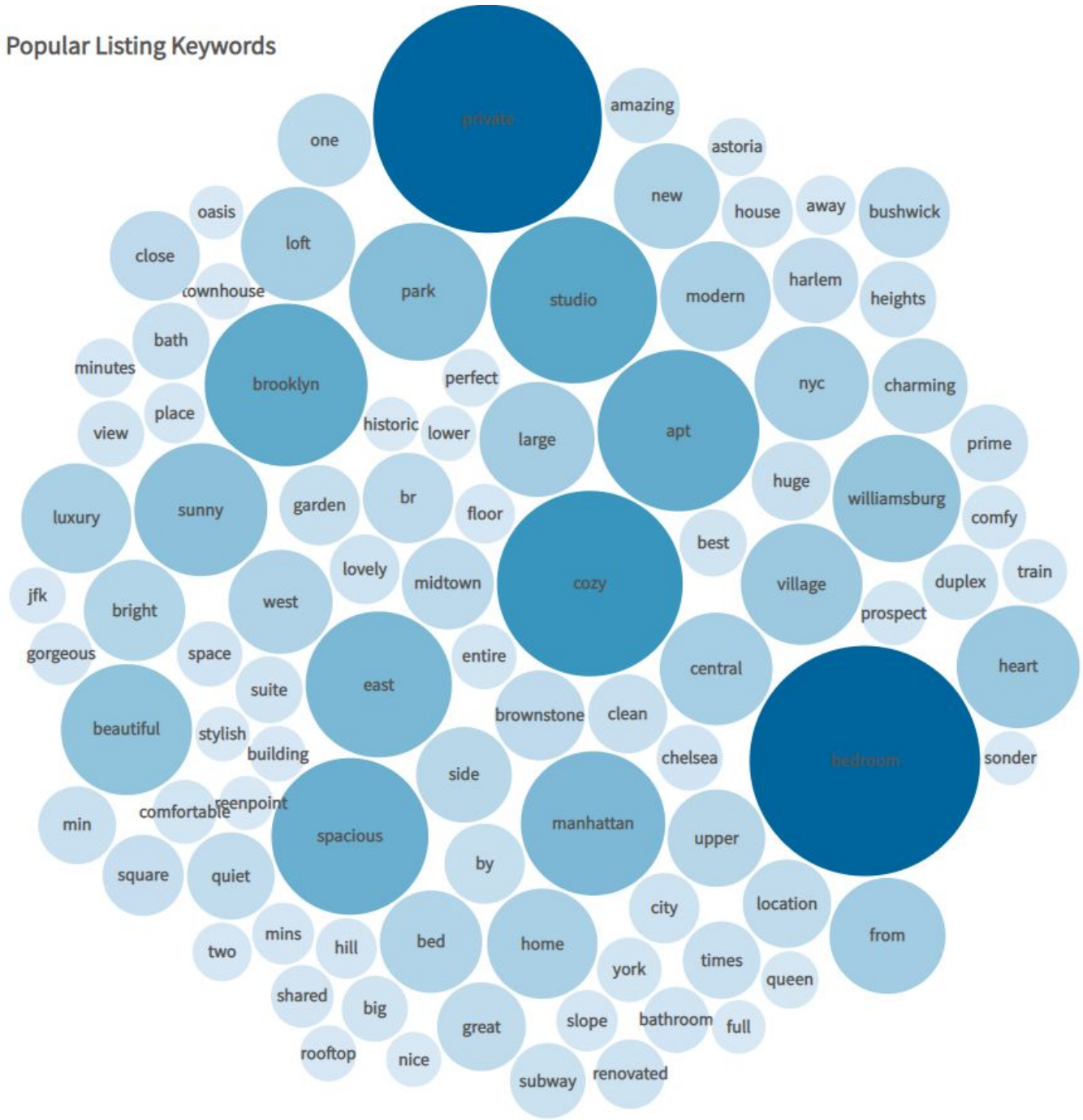
Once we began constructing the viz, it resulted in us designing a choropleth map that showed the neighborhoods of NYC. You could select a neighborhood on the map and it would display the average prices of three categories of listings, private rooms, shared rooms, and entire houses in a bar graph. Once you hover over the bar in the bar graph it shows the amount of listing that are available in that category. We also decided to show a bubble chart that contains keywords of the names of listings and a treemap that shows host who own 10 or more listings. The first iteration of the design is shown below. Each visualization was shown separately on the webpage except the choropleth map and the bar graph.

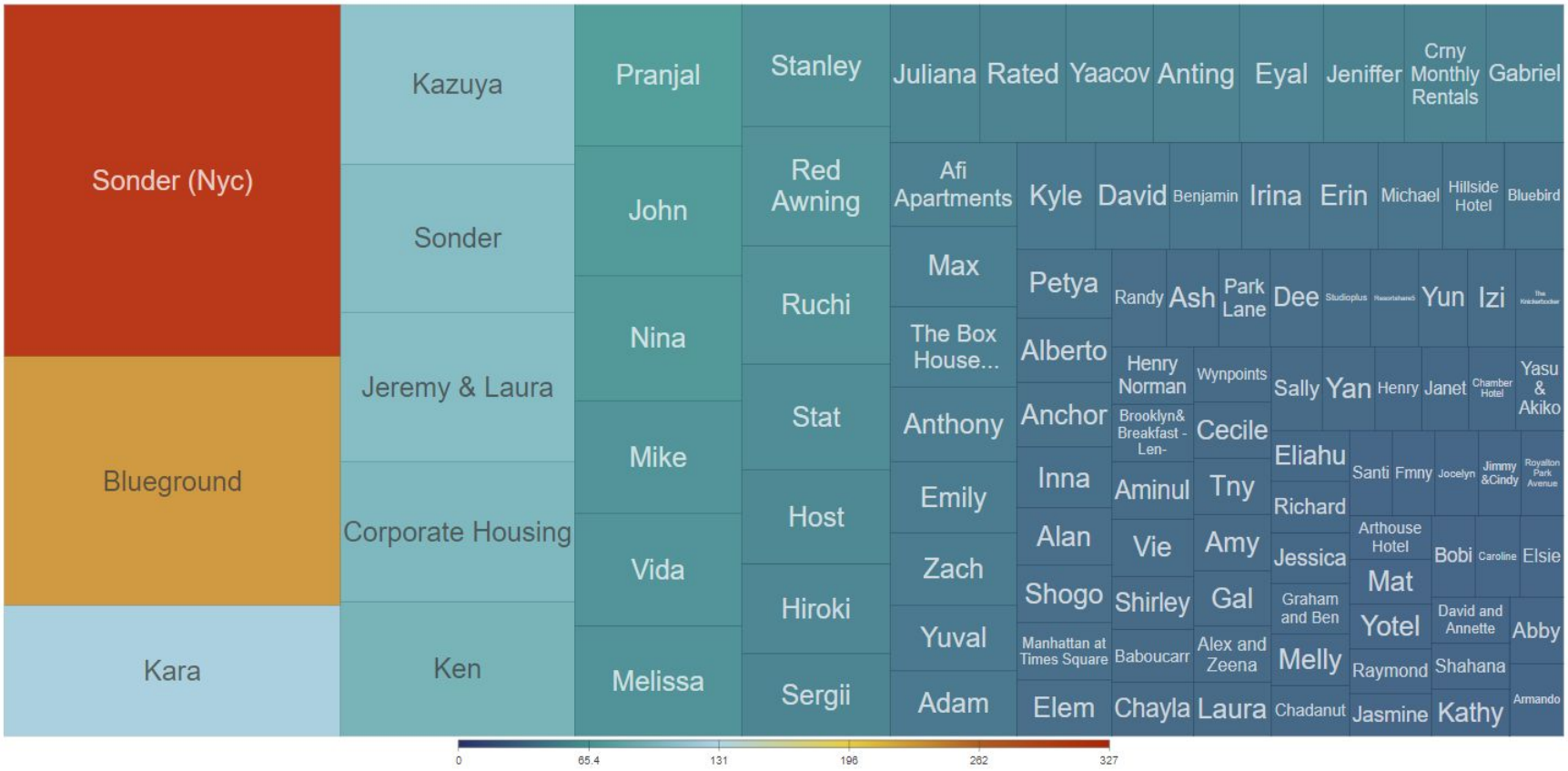




0-365 days

Availability Selected is 0-365 days


Popular Listing Keywords
private
amazing
one
astoria
new
house
away
bushwick
oasis
loft
close
park
studio
modern
harlem
heights
townhouse
bath
minutes
brooklyn
perfect
nyc
charming
place
view
historic
lower
large
apt
prime
huge
williamsburg
luxury
sunny
garden
br
floor
comfy
best
lovely
midtown
cozy
village
duplex
train
jfk
bright
west
prospect
gorgeous
space
entire
central
heart
suite
east
brownstone
clean
beautiful
stylish
chelsea
bedroom
sonder
building
side
greenpoint
comfortable
min
spacious
manhattan
upper
square
quiet
by
location
city
from
two
mins
hill
bed
home
york
times
queen
shared
big
great
slope
bathroom
full
rooftop
nice
subway
renovated


Sonder (Nyc)
Blueground
Kara
Kazuya
Sonder
Jeremy & Laura
Corporate Housing
Ken
Pranjal
John
Nina
Mike
Vida
Melissa
Stanley
Red Awning
Ruchi
Stat
Host
Hiroki
Sergii
Juliana
Rated
Yaacov
Anting
Eyal
Jeniffer
Crny Monthly Rentals
Gabriel
Afi Apartments
Kyle
David
Benjamin
Irina
Erin
Michael
Hillside Hotel
Bluebird
Max
Petya
Randy
Ash
Park Lane
Dee
Studioplus
Yun
Izi
The Box House...
Alberto
Henry Norman
Wynpoints
Sally
Yan
Henry
Janet
Chamber Hotel
Yasu & Akiko
Anthony
Anchor
Brooklyn& Breakfast - Len-
Cecile
Eliahu
Santi
Fmny
Jocelyn
Jimmy &Cindy
Royalton Park Avenue
Emily
Inna
Aminul
Tny
Richard
Zach
Alan
Vie
Amy
Jessica
Arthouse Hotel
Bobi
Caroline
Elsie
Shogo
Shirley
Gal
Graham and Ben
Mat
Yuval
Manhattan at Times Square
Baboucarr
Alex and Zeena
Melly
Yotel
David and Annette
Abby
Adam
Elem
Chayla
Laura
Chadanut
Raymond
Shahana
Jasmine
Kathy
Armando
0
65.4
131
196
262
327

Design Rationale:

Cholrepth Map

The colors on the map are a single hue, that highlights blue when a neighborhood is selected. The bar chart changes when the neighborhood is selected, changing the height and value of each bar. The bar chart also shows the average price of the room types available. When a neighborhood is hovered over the name of the neighborhood is displayed in a tooltip. The scale to the left, is a color scale of the pricing in the area. We also decided to incorporate a drop down of the availability that of each listing. Categorized by the amount of days its available from 0 to 365. The map and bar chart are interactive with each other.

Bar Graph

The bar graph was used to compare the room types by price. Values are shown by height, and displayed by text. The hue used is blue to correlate with the selected area of the map. By hovering over the bars this allows users to view the amount of listings that fit that room type description.

Bubble Chart

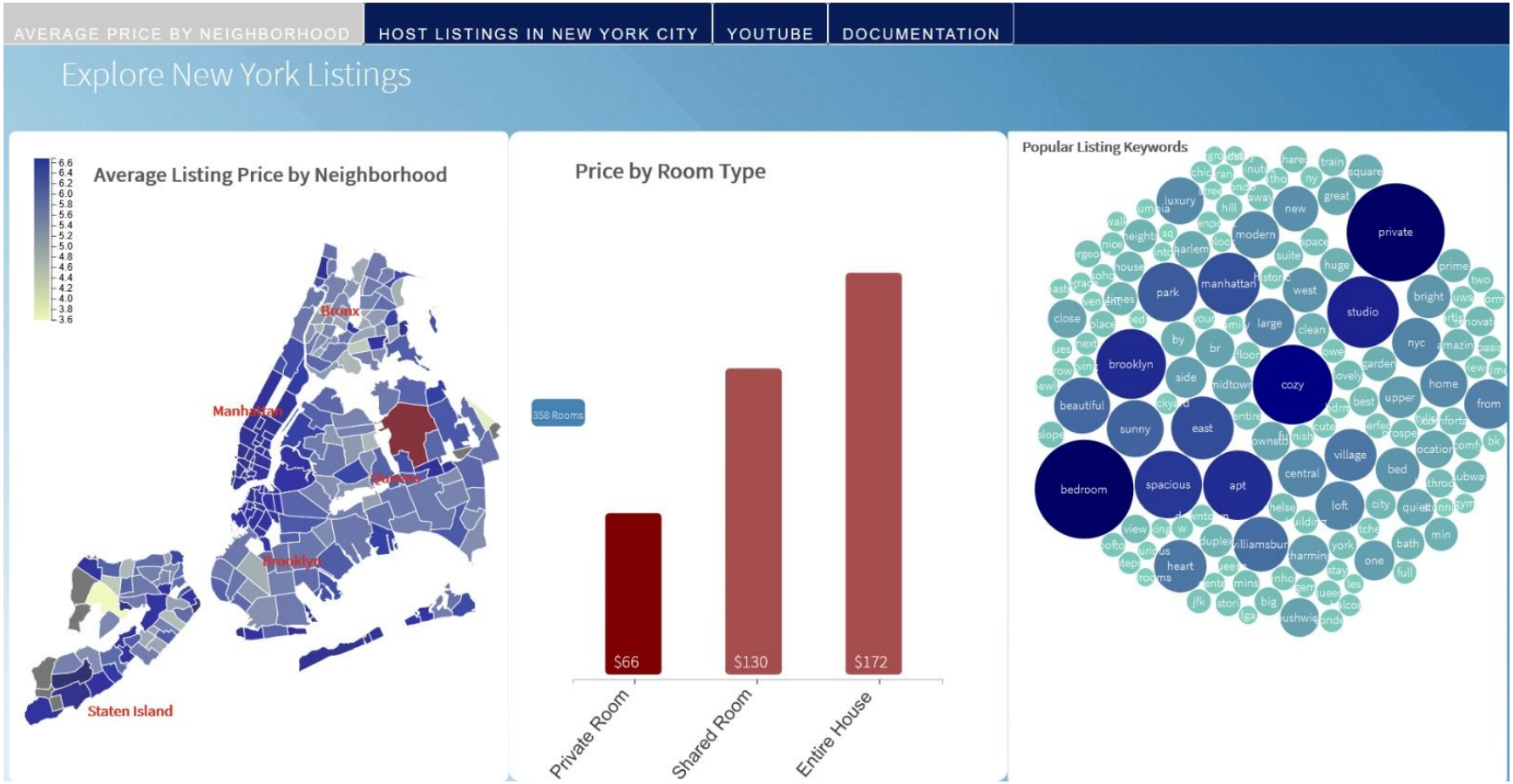
The bubble chart was an addition that we provided to give users more insight into the type of listings in NYC, the keywords show some of the type of listings available and the most common words used. The viz uses a single blue hue, the value of the bubble increases the size of the area and darkness. The viz doesn't have any interaction with other maps but provides a visual reference for common key words.

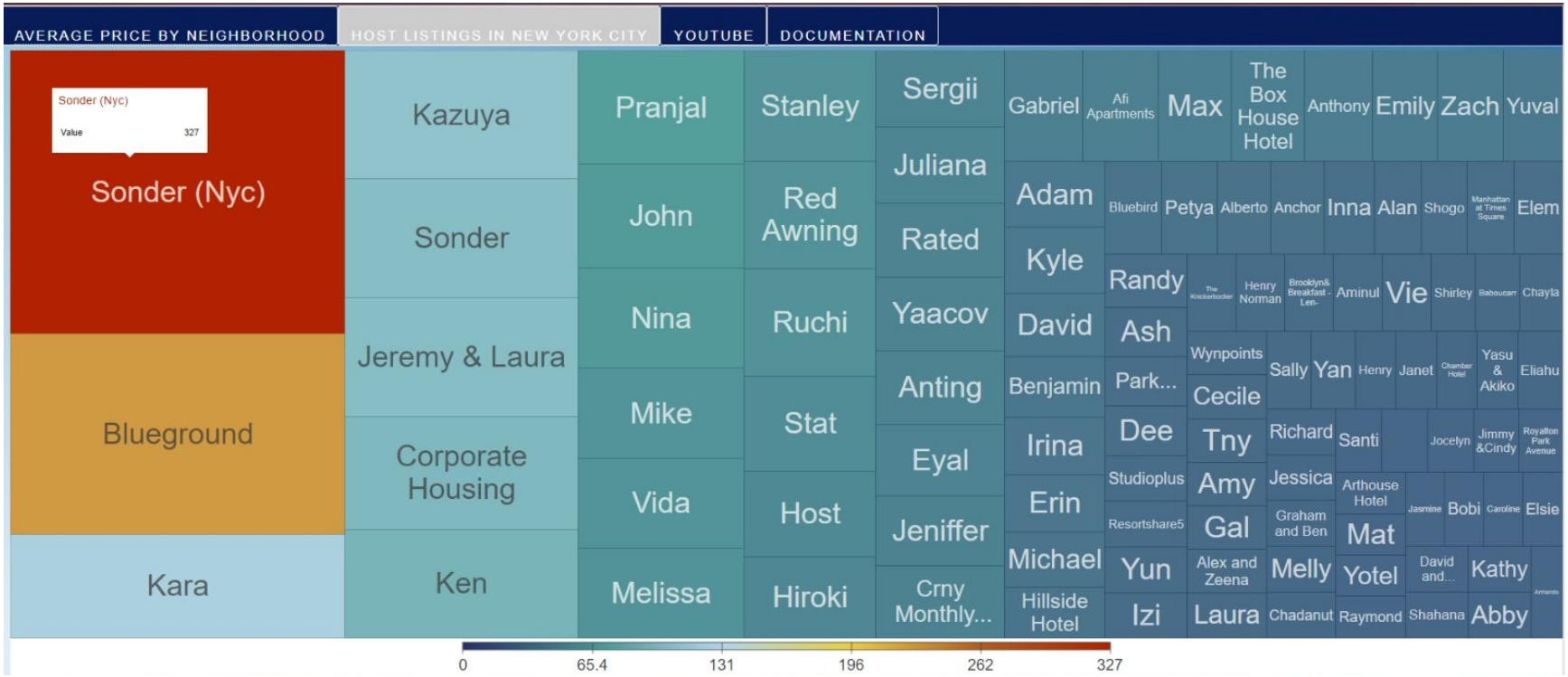
Treemap

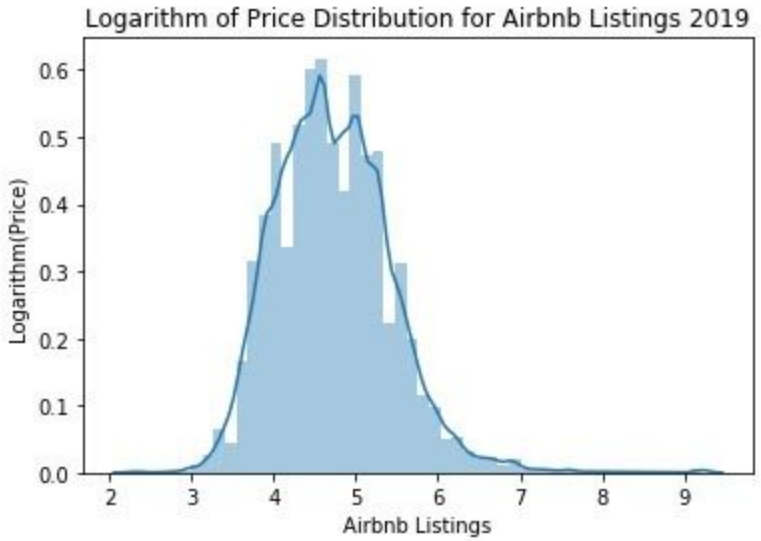
The treemap turned out to be a good decision to see who owned the majority of the listings because users can quickly see the hierarchy of the number of host listings. To constructed the treemap we limited it to host that own ten or more properties, this decision was made because the treemap didn't display 31,355 objects clearly, so by making this adjustment there were 115 remaining objects. We decided to use a multi hue color scale with dark blue being the 0 and red being 327. The number 327 was chosen because it is the highest value in the dataset. Value controls the area, size, and hue within the visualization. The treemap isn't interactive with other maps but there is an on hover transition that displays a tooltip showing how many properties are owned by each host, and text is used to display the name of each host.

Feedback:
We presented this visualization in class and received recommendations on our design. The recommendations were to use a multi hue color scale for the cholrepth map and to make it a logarithmic scale . Another suggestion was to make sure each visualization could be seen at the same time, to show the relationship between each viz on one page.

Final Iteration Results:




AVERAGE PRICE BY NEIGHBORHOOD
HOST LISTINGS IN NEW YORK CITY
YOUTUBE
DOCUMENTATION
Sonder (Nyc)
Value 327
Sonder (Nyc)
Blueground
Kara
Kazuya
Sonder
Jeremy & Laura
Corporate Housing
Ken
Pranjal
John
Nina
Mike
Vida
Melissa
Stanley
Red Awning
Ruchi
Stat
Host
Hiroki
Sergii
Juliana
Rated
Yaacov
Anting
Eyal
Jeniffer
Crny Monthly...
Gabriel
Afi Apartments
Max
The Box House Hotel
Anthony
Emily
Zach
Yuval
Adam
Bluebird
Petya
Alberto
Anchor
Inna
Alan
Shogo
Manhattan at Times Square
Elem
Kyle
Randy
The Knickerbocker
Henry Norman
Brooklyn& Breakfast - Len-
Aminul
Vie
Shirley
Baboucarr
Chayla
David
Ash
Wynpoints
Sally
Yan
Henry
Janet
Chamber Hotel
Yasu & Akiko
Eliahu
Benjamin
Park...
Cecile
Irina
Dee
Tny
Richard
Santi
Jocelyn
Jimmy &Cindy
Royalton Park Avenue
Erin
Studioplus
Amy
Jessica
Arthouse Hotel
Jasmine
Bobi
Caroline
Elsie
Michael
Resortshare5
Gal
Graham and Ben
Mat
Yun
Alex and Zeena
Melly
Yotel
David and...
Kathy
Armando
Hillside Hotel
Izi
Laura
Chadanut
Raymond
Shahana
Abby
0
65.4
131
196
262
327



Conclusions:
When we began, we wanted to analyze the three most influential factors when users of airbnb searched for accommodations, host information, pricing, and location. After our first iteration and feedback we made changes to improve the comprehension of the visualization. We combined most of the visualizations but because we used a different package to create the treemap it didn't integrate with the choropleth map, bubble chart, and bar chart. We gave the choropleth map a multi hue color scale of blue and yellow and changed the scale to logarithmic. We changed the color of the bar chart to the same color of the selected neighborhood to show the correlation between each viz. The hue of the bubble chart was changed to a similar blue hue to the choropleth map to keep the theme of the entire visualization. There were no changes to the treemap. Once our changes were made we believe our visualization did a better job answering our research questions. The questions that our designs focused on gave us some insight about which areas are most expensive, it confirmed our assumptions about which room type is the most expensive, what keywords host used the most to market their properties and which host had the most listings.

Most Expensive Area: Manhattan, which can be seen on the map because it is the darkest borough.

Most Expensive Room Type: From greatest to least, Entire home, Private Room, Shared Room
Most Used Keywords: From greatest to least, Private, Bedroom, Cozy
Host with the Most Listings: Sonder, Blueground, Kara

Once the results were synthesized we found correlation between the most used listing names and the pricing of private rooms. It was also interesting to see that the host who had the most listings turned out to be organizations and not individuals which resulted in them having properties in multiple boroughs.

Next Steps:
We would integrate data on the users most searched listing names, to compare with our bubble chart that displays the most commonly used listing name by the host, to find any correlation. We would also create the treemap using a different package to ensure each viz could be seen at once, and making them interactively cross functional.