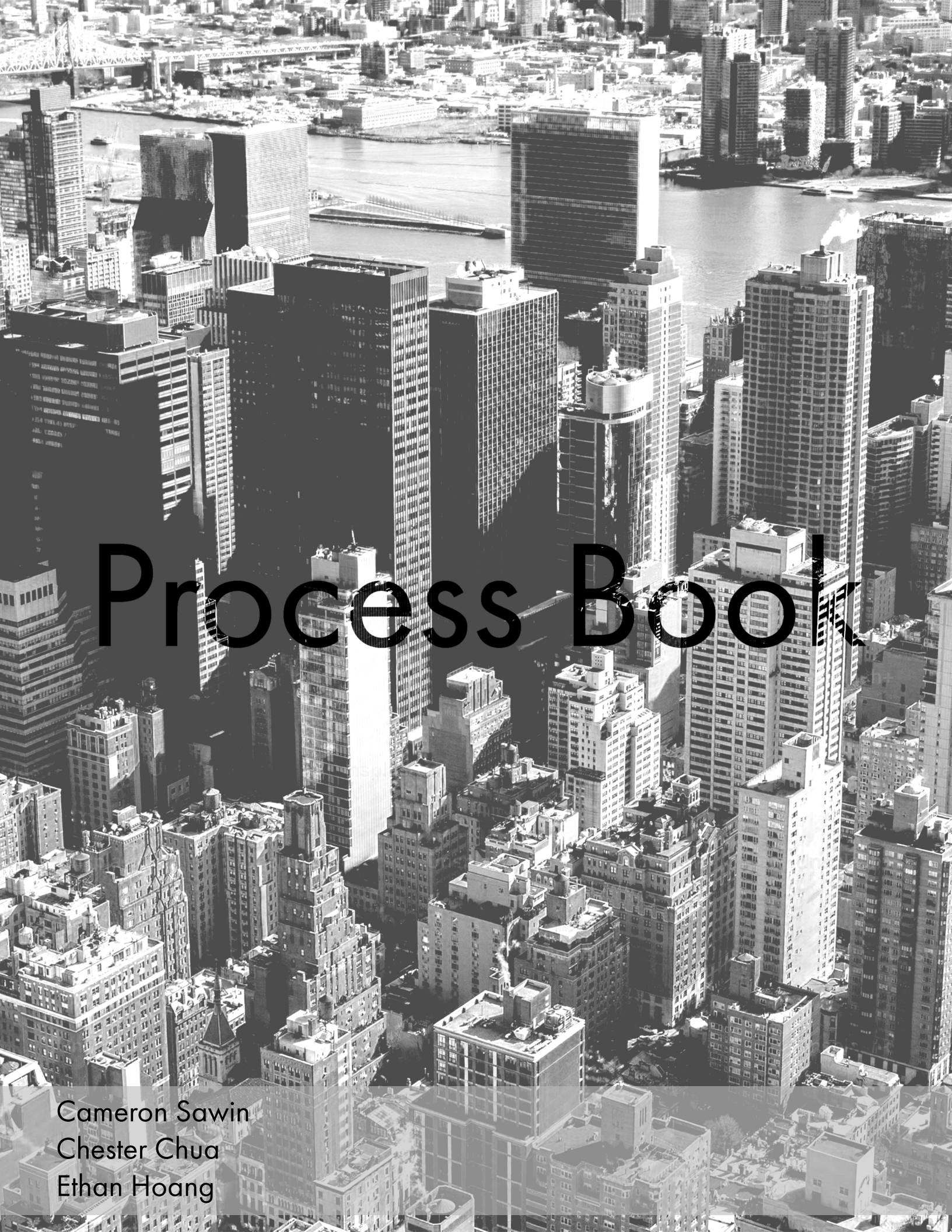
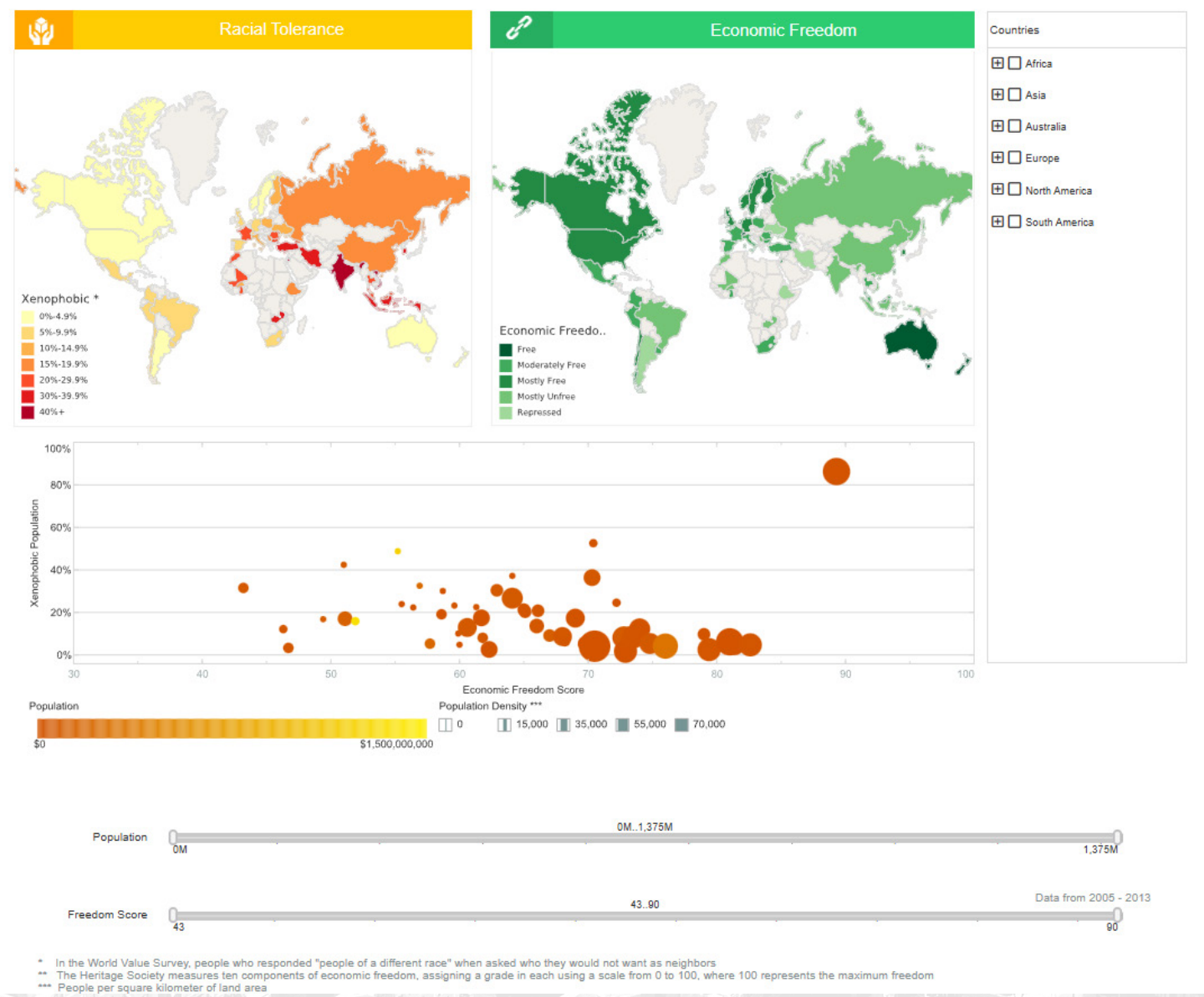


Process Book

Cameron Sawin
Chester Chua
Ethan Hoang

We want our dashboard to serve as an informational guide to users, such as potential AirBnB hosts who wish to accurately price their AirBnBs in a way that will help them visualize specific trends among other AirBnBs. By utilizing outlying factors found in other surrounding AirBnBs in New York, users can analyze the general trends/relationships created and make an informed decision to price their AirBnB accordingly.



https://www.inetsoft.com/evaluate/bi_visualization_gallery/page.jsp?dbIdx=102

Overview//Precedents

General: What patterns/relationships can we find within our dataset that will reveal important information about pricing AirBnBs in New York City?
Which boroughs have the highest average pricing of AirBnB listings?
How does the date of construction affect the price of the listing?
What neighborhoods within each of the five boroughs in NYC have the highest AirBnB ratings?

As our project evolved, we began to ask more questions and introduce other factors related to creating an accurately priced listing.

Some new questions include:
Are recently built AirBnBs priced higher than older ones?
Does listing a higher service fee for AirBnBs command a higher price listing overall?
Do differing room types affect the price of AirBnBs?

In general, we initially started out by asking broad questions about the data; we wanted to identify certain patterns and relationships we could find in the data without having them mesh together into a full "story", which was:

How can users find patterns and trends in the data as a way to help them accurate price AirBnBs listings based on their surrounding area?.

However, as we received feedback, we not only changed our questions to fall more in line with the narrative we wanted to tell, but we also began to ask how those questions related to each other. As will become clear in our dashboard, we wanted these questions to overlap with each other. That way, we could have our visualizations intersect with each other through interactions, while also exploring distinct attributes that could reasonably affect the price listings of AirBnBs.

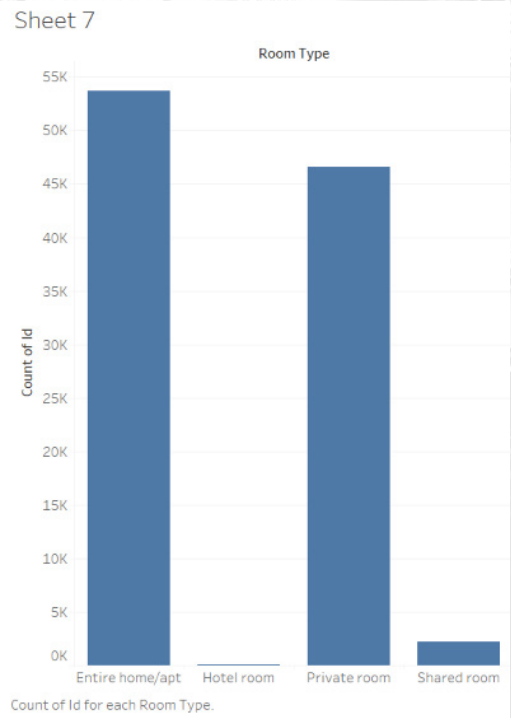
Questions

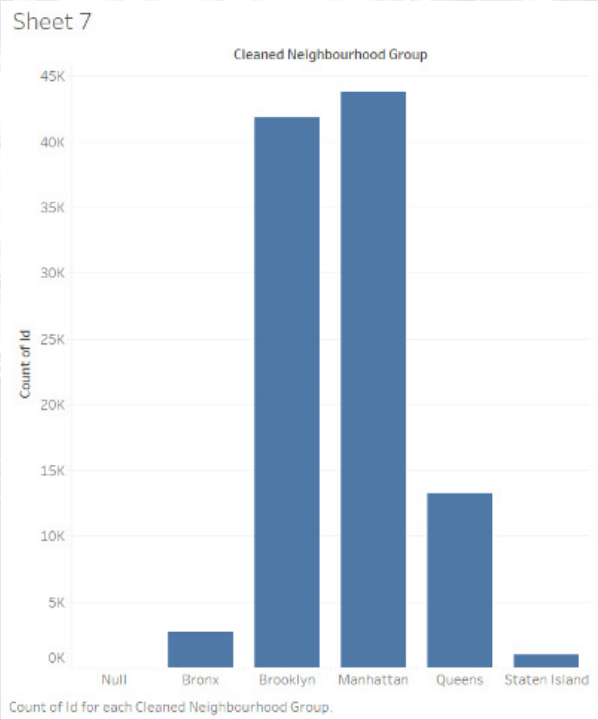
Attribute	Raw Type	Type	Description
id	integer	categorical	Airbnb's unique identifier for the listing.
name	text	categorical	Name of the listing.
host_id	integer	categorical	Airbnb's unique identifier for the host/user.
host_identity_verfied	text	categorical	Shows whether the host's identity has been verified or not by listing either "unconfirmed" or "verified".
host_name	text	categorical	Name of the host. Usually just the first name(s).
neighborhood_group	text	categorical	Borough the listing is in (such as Manhattan or Queens).
neighborhood	text	categorical	Neighborhood the listing is in.
lat	numeric	quantitative	Latitude value for where the listing is located. Uses the World Geodetic System (WGS84) projection.
long	numeric	quantitative	Longitude value for where the listing is located. Uses the World Geodetic System (WGS84) projection.
country	text	categorical	Country the listing is located in.
country_code	text	categorical	Country code of country.
instantly_bookable	boolean	categorical	Either true if guests can automatically book the listing without the host being required to accept the request, or false if not. This is an indicator of a commercial listing.
cancellation_policy	text	ordered	Either strict, moderate, or flexible based on the host's policies on cancellation.
room_type	text	categorical	Type of room the listing contains. These values can be Entire home/apt, Private room, Shared room, or Hotel room.
construction_year	integer	quantitative	Year that the property was built.
price	numeric	quantitative	Price of the listing per night.
service_fee	integer	quantitative	Service fee of the listing.
minimum_nights	integer	quantitative	Minimum number of nights a customer must stay at the property.
number_of_reviews	integer	quantitative	Total number of reviews the listing received.
last_review	date	quantitative	Date that the last review of the listing was received on.
reviews_per_month	numeric	quantitative	Average number of reviews the listing receives per month.
review_rate_number	integer	ordered	Average rating left in reviews of the listing.
calculated_host_listings_count	integer	quantitative	Number of listings that the host has in total.
availability_365	integer	quantitative	Availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
house_rules	text	categorical	Host's written rules to the guest for the listing.
license	text	categorical	The license/permit/registration number.

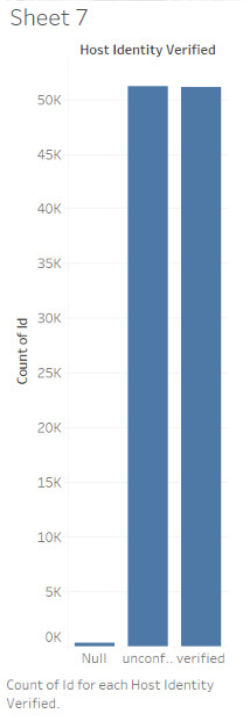
https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata

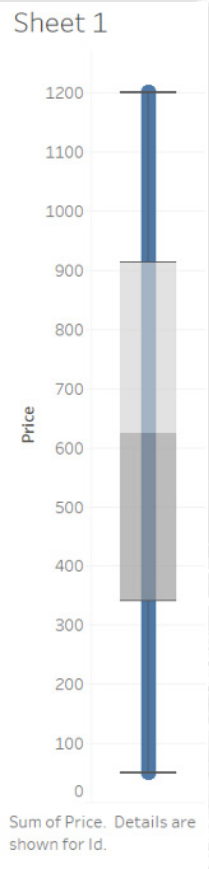
Our dataset measured several important factors regarding AirBnBs around New York City. The dataset, sourced from Kaggle, contains 26 attributes measuring around 60,000 items in total. This dataset contained a rich mix of quantitative, categorical, and ordered attributes, which we imported into Tableau to generate a comprehensive table. Finally, we cleaned up the data by standardizing the syntax of all items. For example, we got rid of a few duplicate rows and the fixed the spellings of boroughs for certain items.

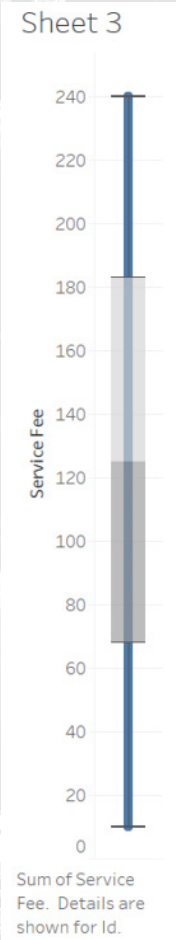
Data

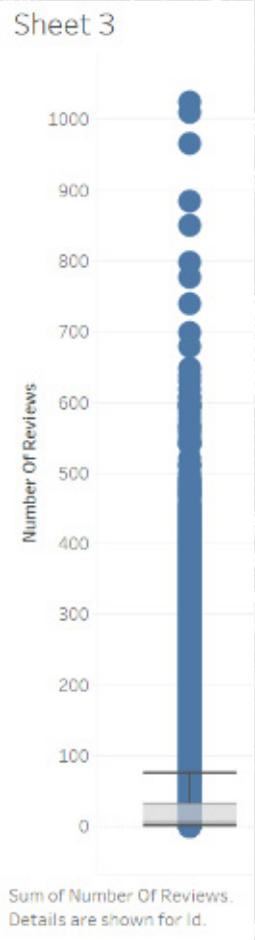






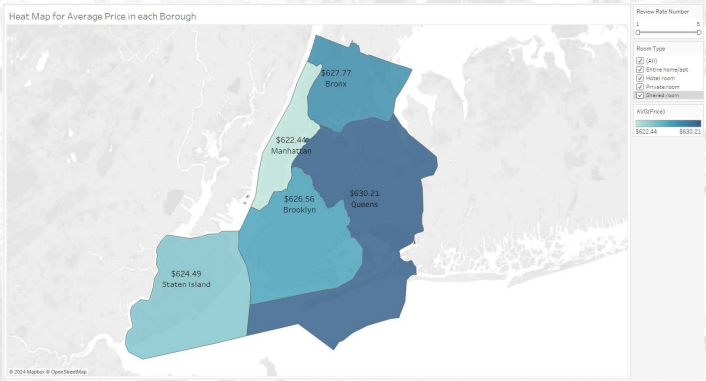


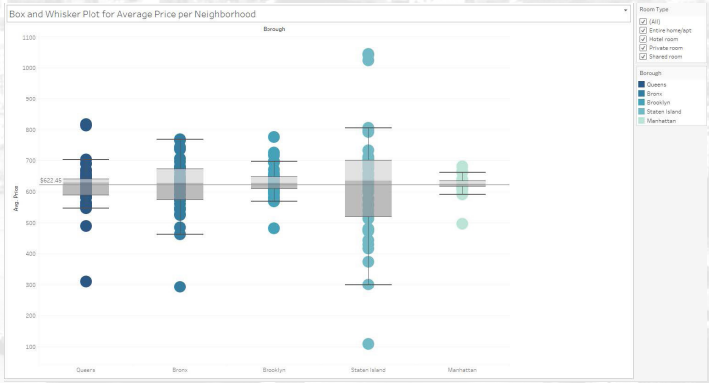




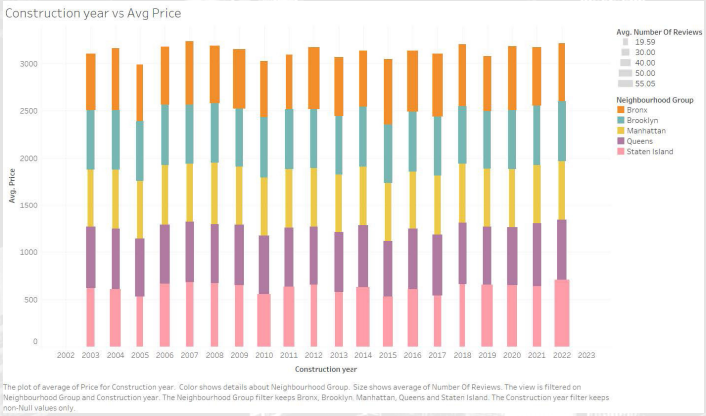
We measured the distribution of quantitative attributes to assess variability and frequency. Additionally, we used a boxplot to compute the median, minimum, and maximum of the quantiative attributes

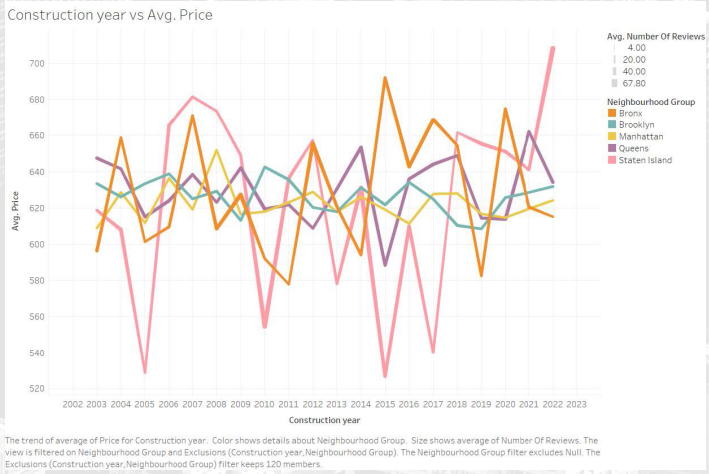
Data





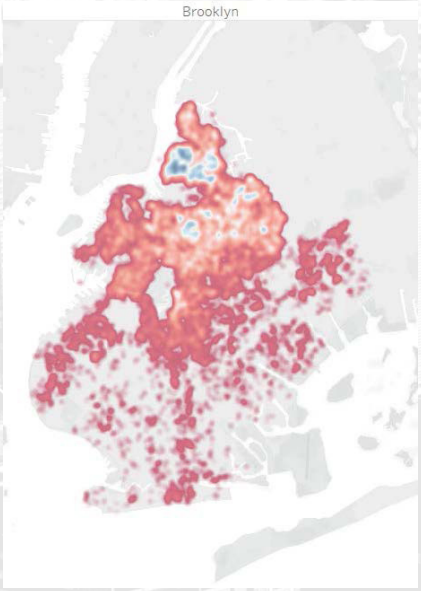
Map and boxplot exploring boroughs and their average prices

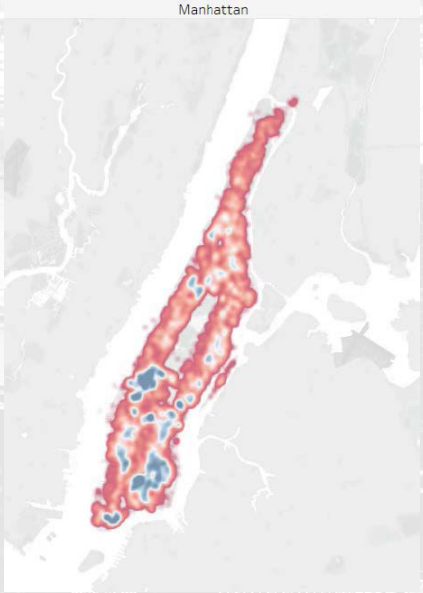


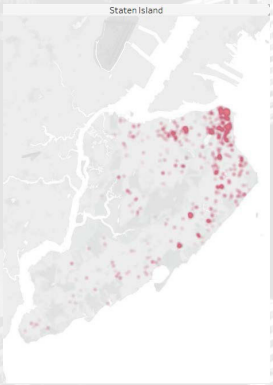


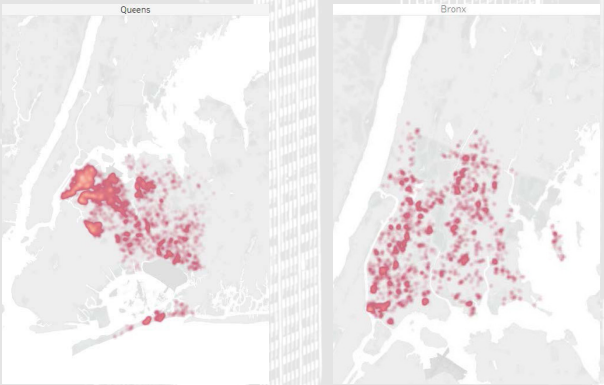
Stacked barchart and line graph exploring the relationship between date of construction and price

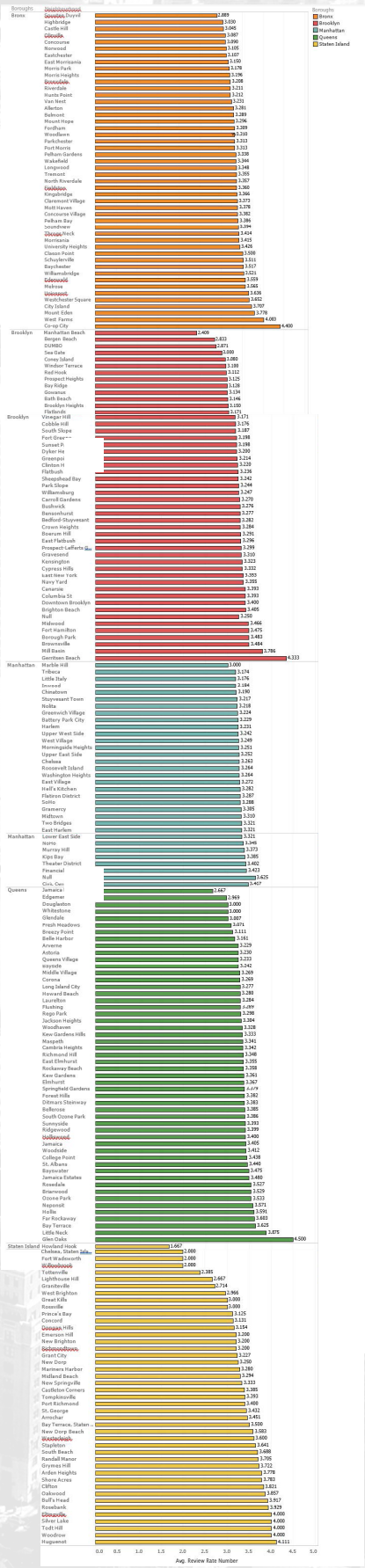
Exploratory Data Analysis





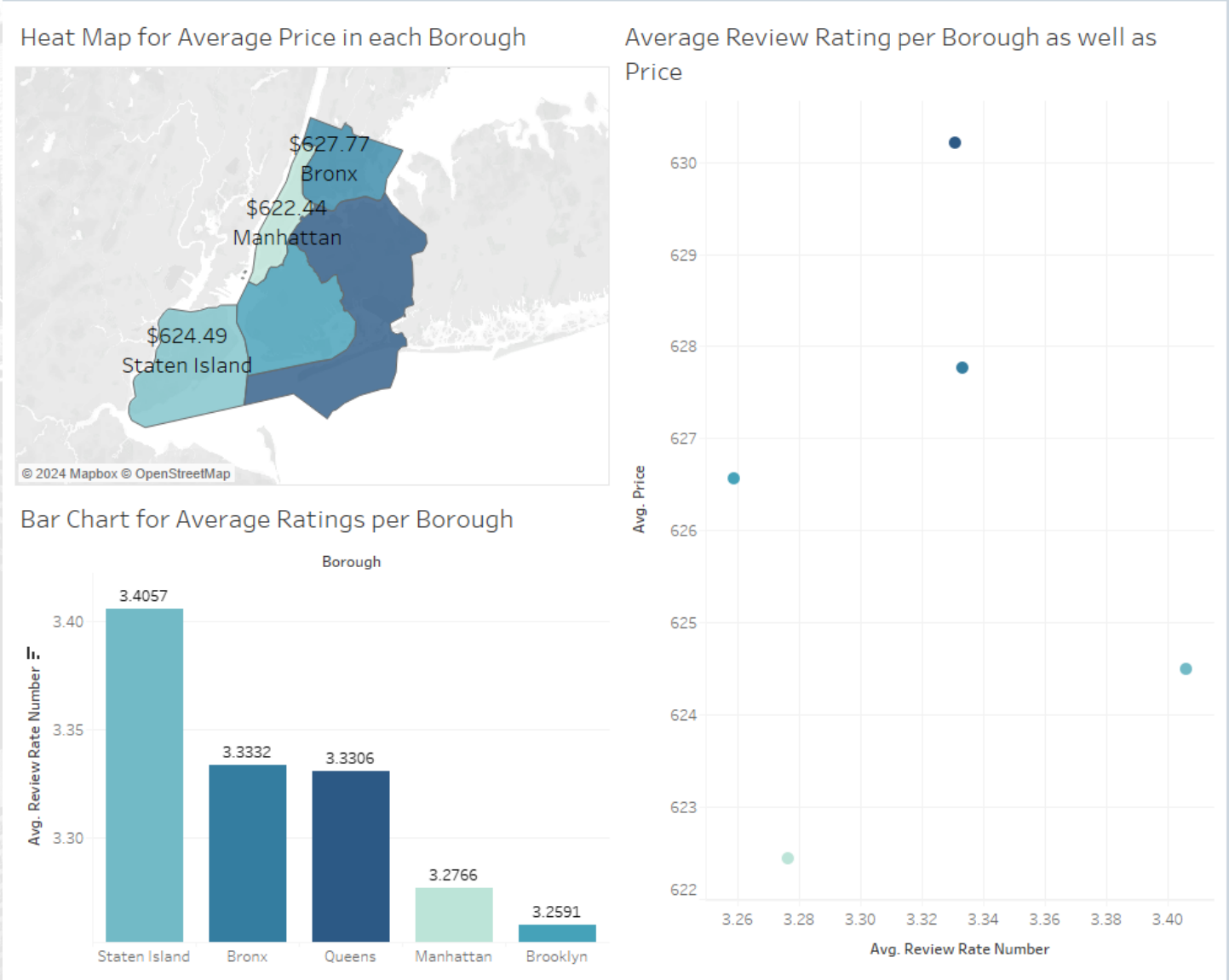






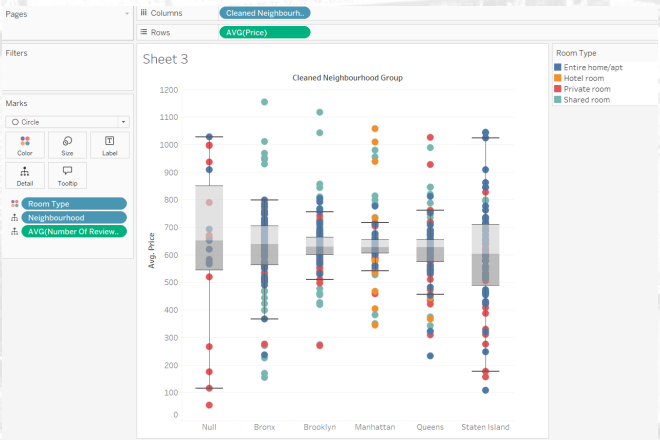
Map and barchart exploring individual neighborhoods and their average ratings across all five boroughs

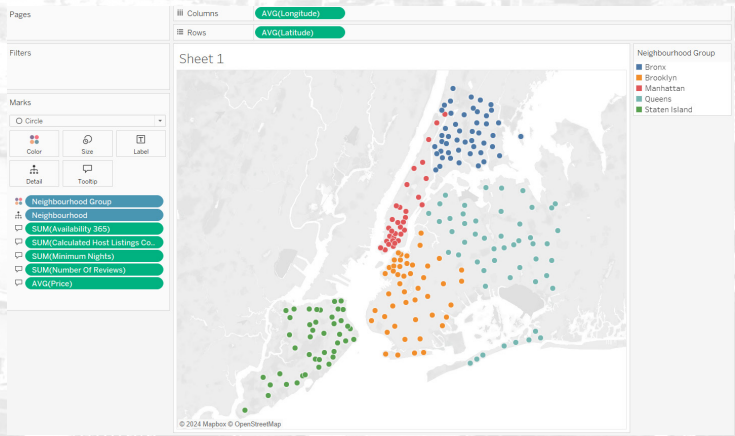
Exploratory Data Analysis



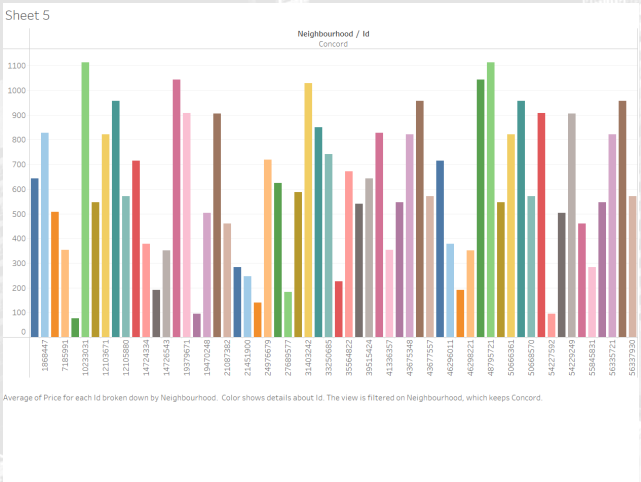
This dashboard differed from our first, as it was our first attempt to refine the visualizations into being tools to filter from one main visualization. These three visualization related to each other at the borough level and allowed for patterns to be discovered at that level.

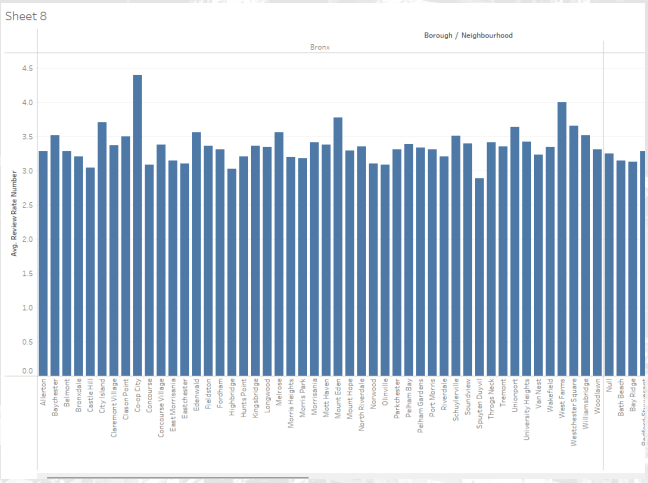
Design Evolution





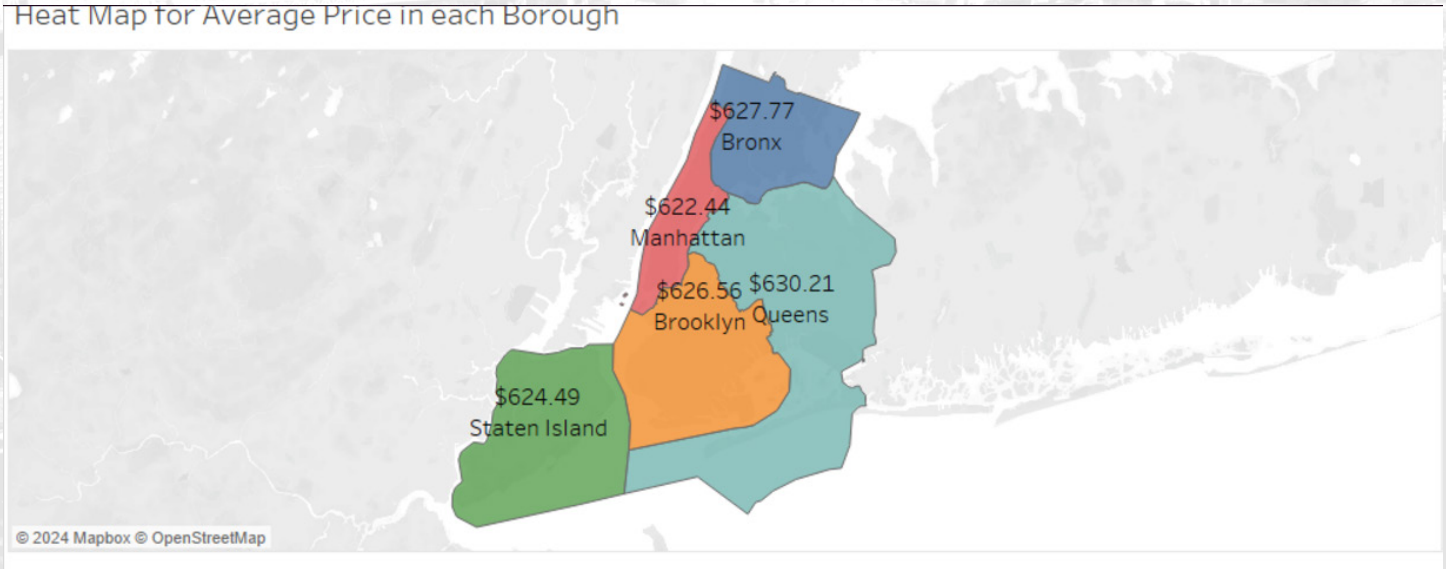
We then tried utilizing different visualizations styles to create more "complex" visualizations. The boxplot combined borough and room type to generate a distribution of average prices, allowing for easier recognition of possible trends/patterns in the data. The map, on the other hand, generated a comprehensive list of all neighborhoods to be used as a point of filtration for subsequent visualizations.

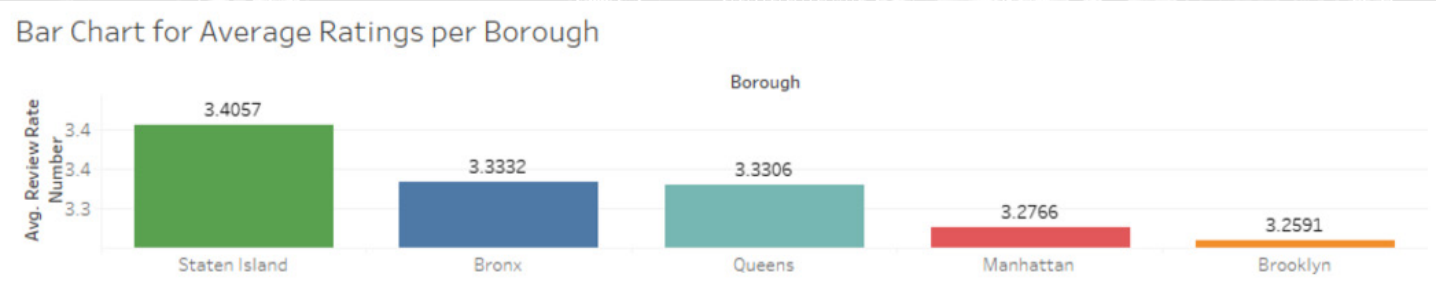


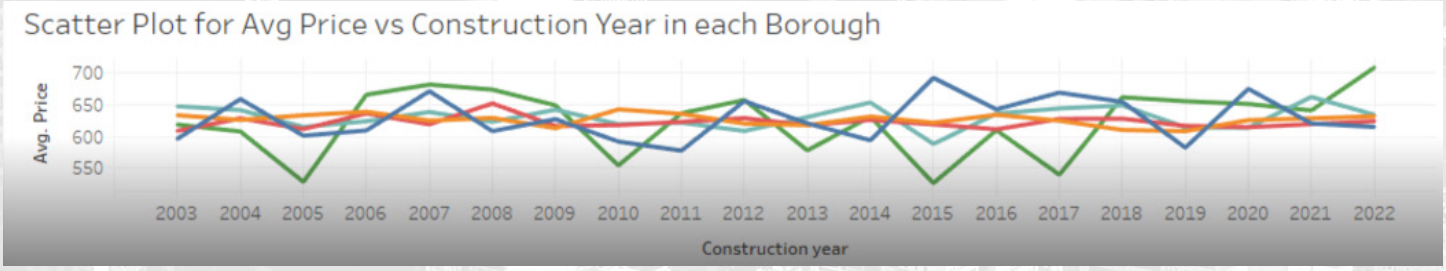


We also experimented with formatting, trying to get our neighborhood-leve bar chart to fit within the bounds of our dashboard. These bar charts reflect the sample size of a visualization filtered by a single borough, which was used as a benchmark to denote the approximate size of the barchart when filtered by neighborhood in specific boroughs.

Design Evolution

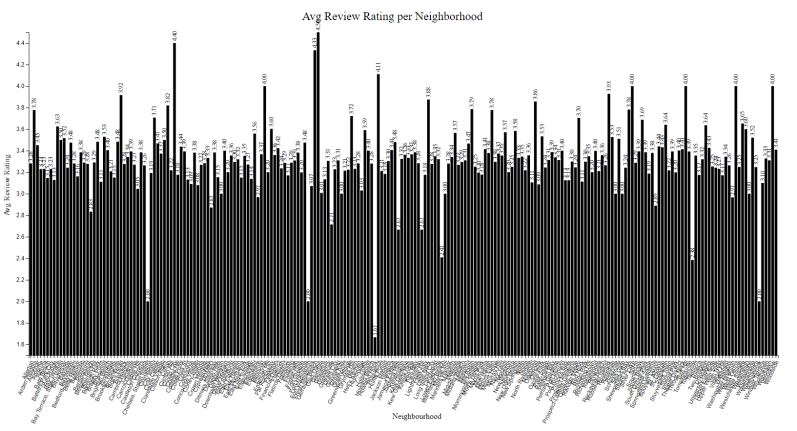




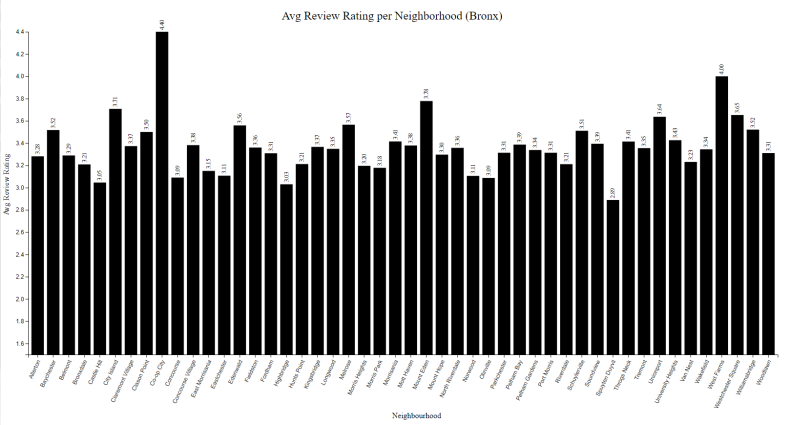


This third dashboard retained a few aspects of our first dashboard but modified others; for example, we wanted to provide a level of filtration provided at the borough level, so we kept the map largely the same. However, we changed the old bar charts to reflect that level of filtration better; the line graph and new bar chart would filter down to show neighborhoods only in the borough selected by the user; so now, unlike our old dashboard which was static and largely unrelated to each other, this dashboard was more streamlined and provided a basic level of interaction with each visualization.

Design Evolution



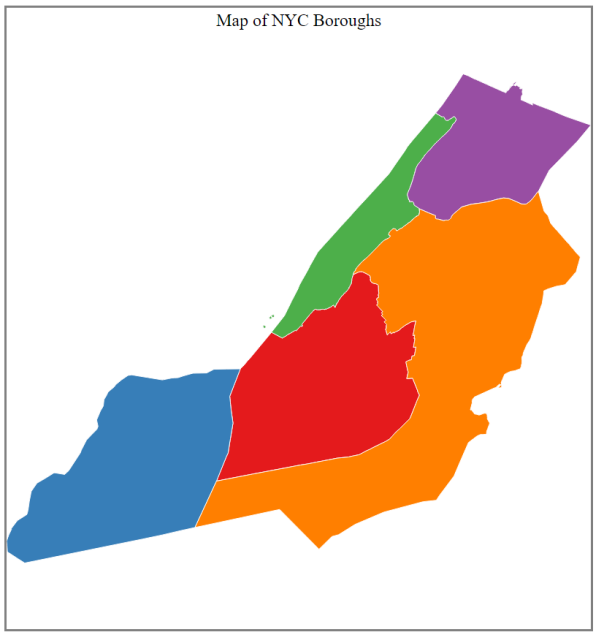
D3 Review Neighborhood Rating Barchart, unfiltered



D3 Review Neighborhood Rating Barchart, filtered

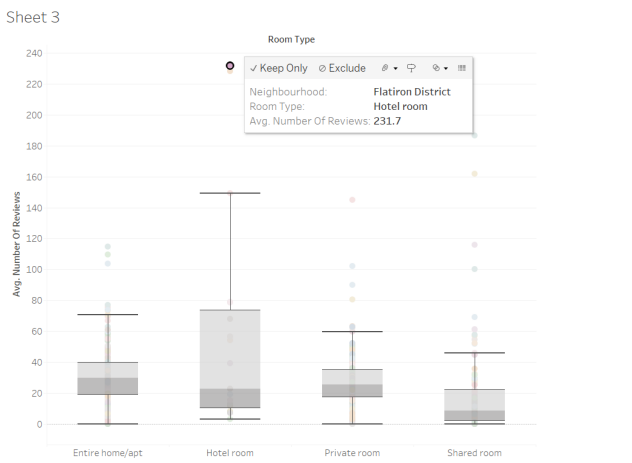
After receiving feedback from our third dashboard, we decided to pursue visualizations that were less aggregated and broader in terms of scope. These visualizations, collected at the individual level, provide and level of complexity that provides not just full details of the entire dataset but also allows users to drill down and filter the data to the level they want (borough, neighborhood, etc.). By creating these visualizations at individual level, we can create complex designs that can still be used to show patterns and trends in specific spots as designated by the user.

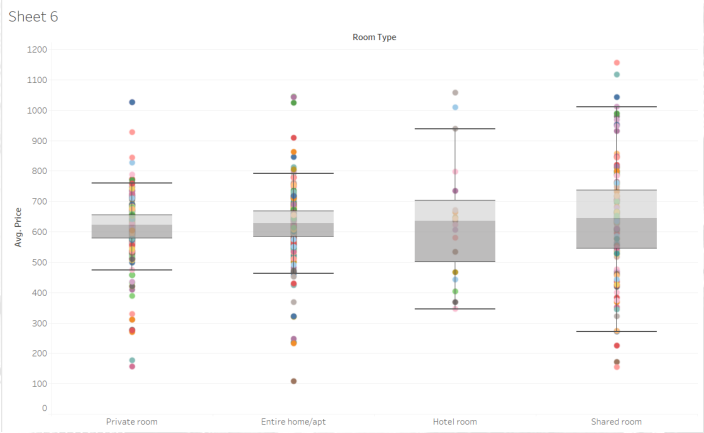
D3 Individual Listing Map



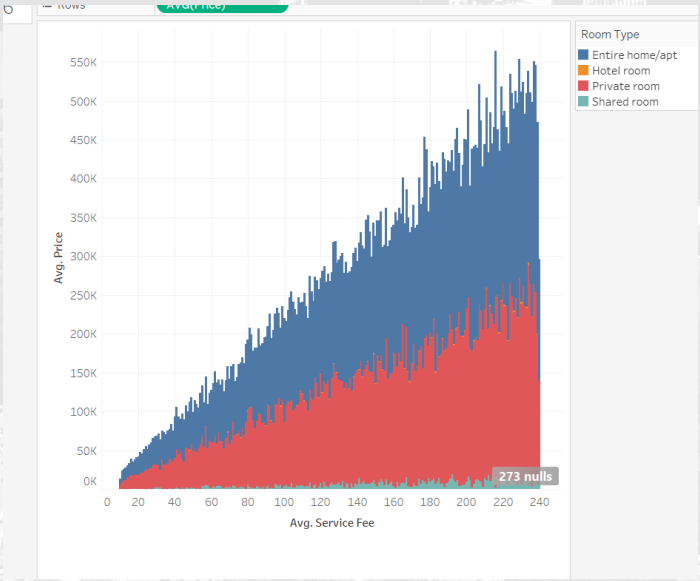
D3 Borough Map

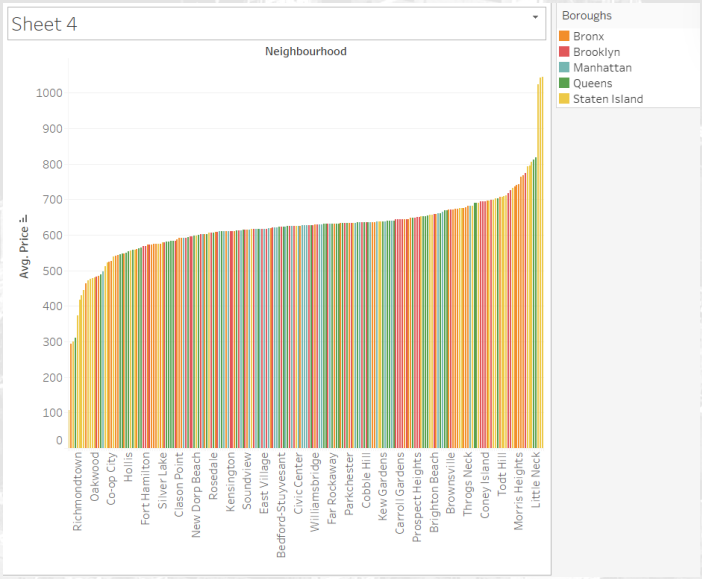
Design Evolution



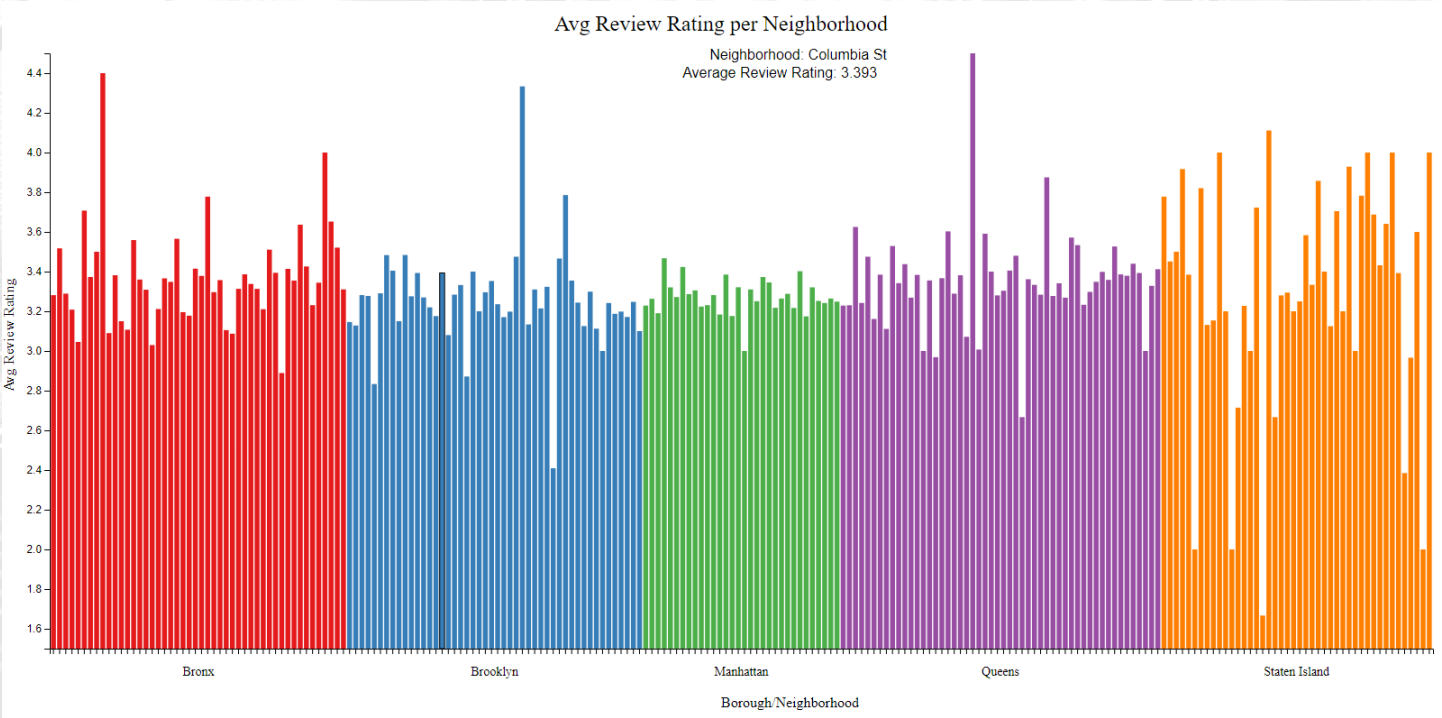


Along with having a central, more complex visualization at individual listing level, we began to refine subsequent visualizations by not only introducing different attributes but arranging them in ways that can induce specific trends and relationships users can find. While the boxplot on the left was made to show the examine data points sorted by "popularity" (number of reviews) among room types, the boxplot on the right was made to show the spread and inform the user the range of values they should charge by room type.

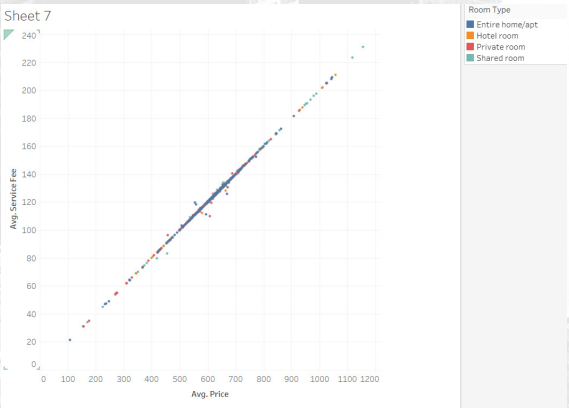


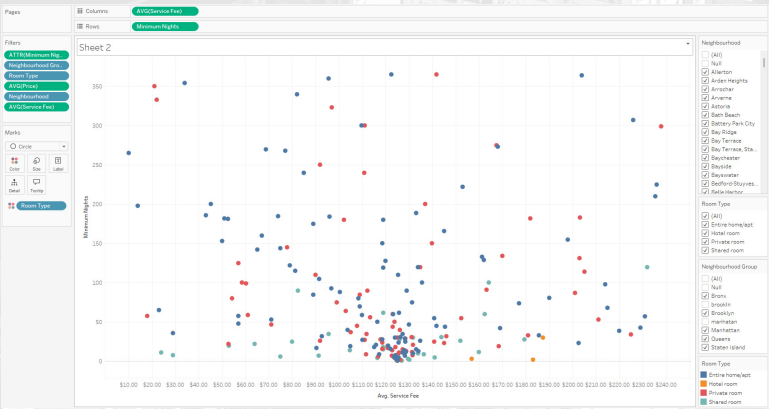


Design Evolution



This was our attempt to organize the the D3 review rating barchart into a refined quality for our final dashboard. This new visualization better organizes the data while retaining the finer level of granularity of the original D3 barchart. It also provides basic interaction via mouseover events detailing each bar.

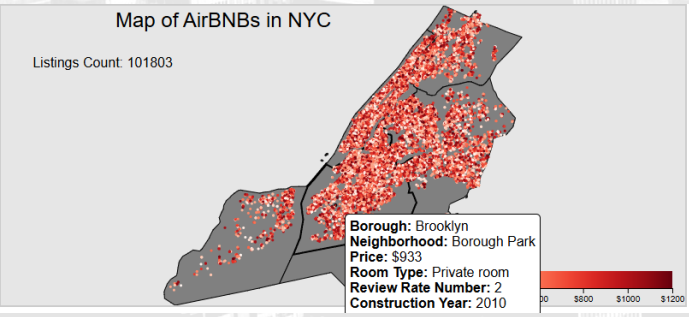




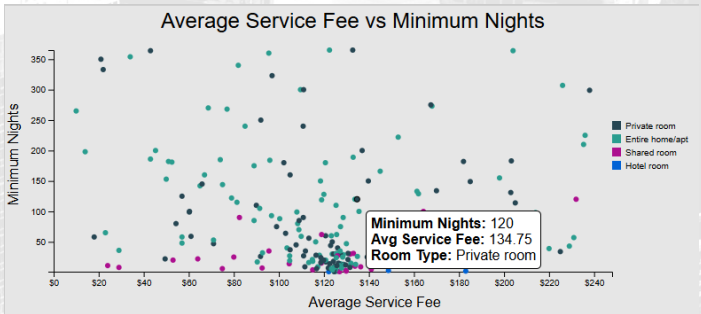
These visualization also try to utilize new attributes to discover distinct patterns within its own set of data. In other words, we tried finding visualizations that, when filtered, could produce paterns and relationships that can be compared and contrasted with the same visualization filtered differently. While the scatterplot on the right shows a clear pattern iterated across all room types, the scatterplot on the left contains a central pattern that potentially differs in distribution/trend iterated across all room types.

Design Evolution

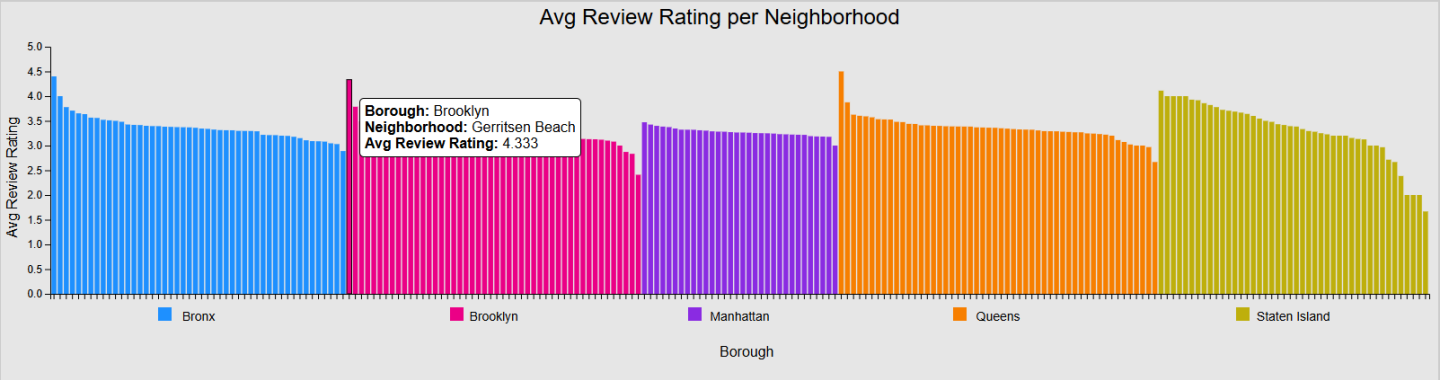
For all 3 visualizations, tooltips allow users to view detailed information about each visualization's attributes. For example, hovering over the individual points on the map reveals a tooltip explaining that specific listing's price, borough/neighborhood, room type, rating, and date of construction. Additionally, the map allows for users to click on specific boroughs, which will filter the scatterplot and barchart in order to view the data at a finer granularity (all the neighborhoods contained in that specific borough), and the barchart can be ordered to show the highest to lowest ratings across all 5 boroughs combined.



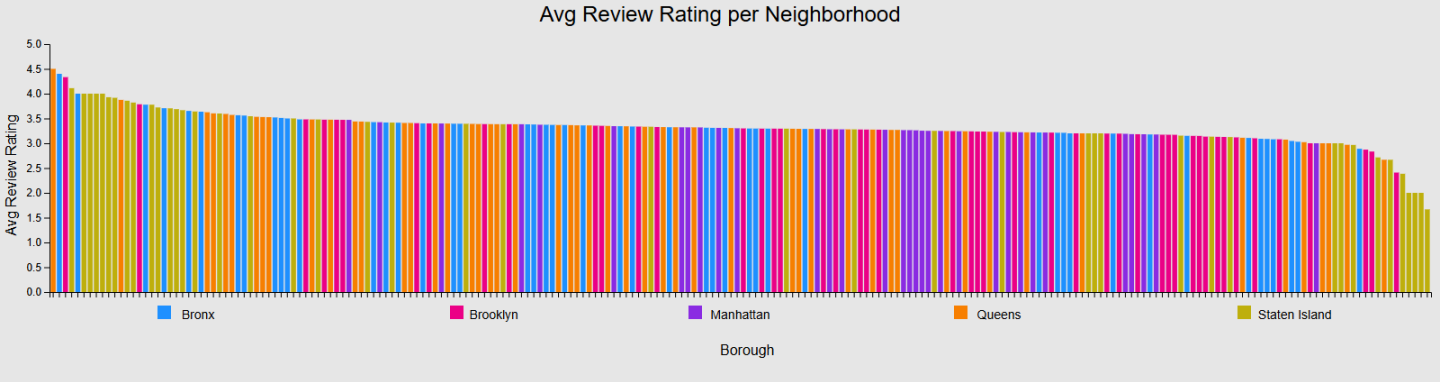
Tooltip for Heatmap



Tooltip for Scatterplot



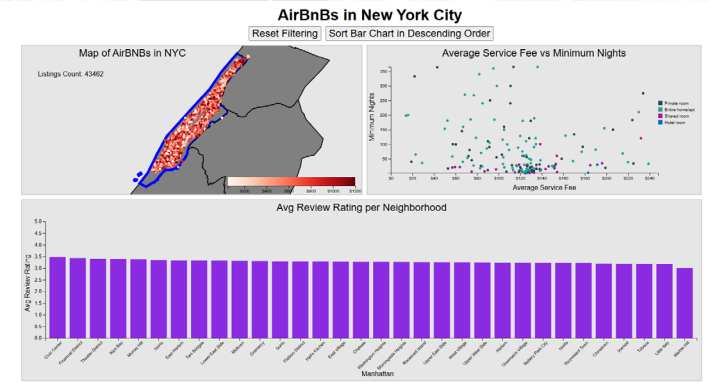
Tooltip for Barchart



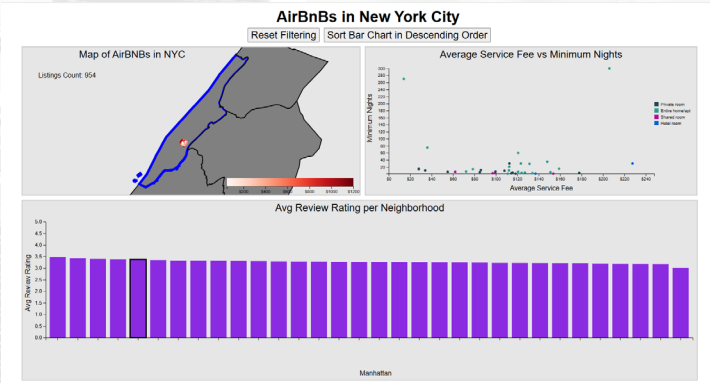
Ordering Neighborhoods by Descending Order, Across all Boroughs

Implementation

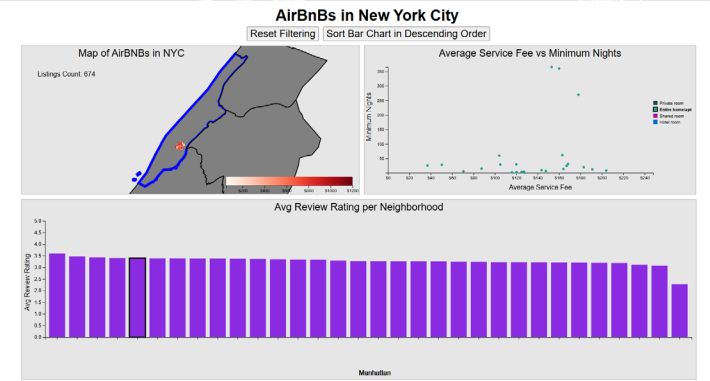
Once the bar chart and scatterplot update accordingly, users can access even finer levels of detail. Specific neighborhoods within individual boroughs can be filtered from the barchart, filtering the map to show individual listings in said neighborhood, while selecting specific room types from the scatterplot filters both the bar chart and map to show only listings with that specific room type.



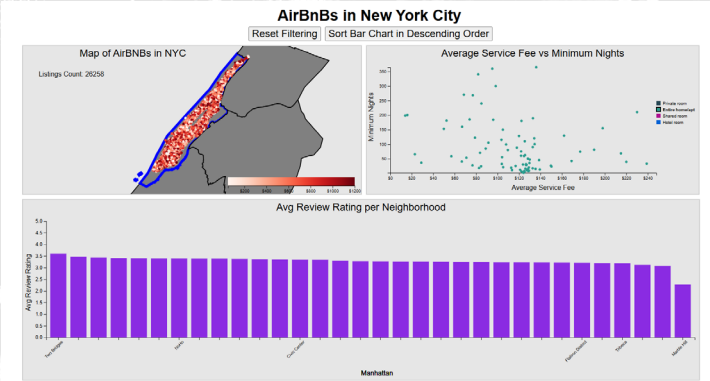
Filter by Specific Borough



Filter by Specific Neighborhood



Filter by Roomtype by Specific Neighborhood



Filter by Roomtype by Specific Borough

The intent behind all these interactions is to allow users to view the data at all 3 levels of granularity (overall data, individual boroughs, individual neighborhoods) and view the trends and relationships among each level. Each visualization has some element of filtration to allow users to recognize patterns among the data, especially when compared with data at a similar level of granularity. For example, the interactivity that lets user view all neighborhoods in one borough lets you compare and contrast the data between every single neighborhood in attributes like ratings and price. Similarly, selecting between room types allows the same level of comparison for every single item across the 3 levels of granularity. In summary, these interactions provide users with enough layers of interactivity to explore the dataset along different attributes, in order to not only laterally compare across data types, but also find patterns and trends among them.

Implementation

Through our visualizations, we were able to identify several patterns and relationships supported by our visualizations. For example, we learned through our bar charts that Staten Island as a whole has higher average ratings than the other 4 boroughs. The scatterplot also shows that, for private rooms and homes/apts, the average spread for pricing service fees based off minimum nights is around $120 to $140 and 0 to 50 nights, respectively, before fanning out. Conversely, the range of values for shared rooms and hotels are much more dispersed. Finally, the map shows that the neighborhoods around Manhattan, Brooklyn, and Queens are the most populated with AirBnBs.

These visualizations helped answer our questions greatly, as they allow users of the dashboard to make informed decisions when pricing their AirBnBs. Through the map we can clearly see not just the distribution of prices within neighborhoods around NYC, but also the frequency of the listings, showing users which neighborhoods are "prime real estate" for creating AirBnBs. The bar chart shows the correlation between neighborhoods and their ratings, which is useful for determining the list price based on how well rated the user's surrounding neighborhood is. Finally, the scatterplot is useful for determining how the user much the user should charge their service fee based on how many nights they are willing to let guests stay, depending on the room type offered.

Our visualization works well for the most part, as they serve their main function in informing users who use it while also being a great observational tool in general for visualizing the different characteristics of New York AirBnBs. It also delivers these characteristics in clear, compelling ways, in a manner that is visually digestible to the average person. Although our main visualization (the heatmap) is very crowded with points, we were able to filter those points down into meaningful patterns via the bar chart and scatterplot. Using those two visualizations as "support" visuals, they enhance the map as a method of plotting the prices in and around NYC, which helps to create meaningful patterns within the data shown by the bar chart and scatterplot.

One way we could improve on the dashboard is by creating a finer level of granularity with map. We could've tried to make the borders for each neighborhood within each borough, allowing users to physically separate the clusters of points between each other when selecting an individual neighborhood. Additionally, we could implement further sorting features in the bar chart, where we could allow for options such as sorting by ascending order or even by borough or neighborhood name. This could potentially allow for more avenues of discovering patterns and relationships in the data when comparing boroughs or neighborhoods. Finally, we could have possibly even introduced an entirely new visualization in our dashboard to further showcase the different attributes of the AirBnBs. For instance, we were very interested in showing how construction dates could potentially affect listing prices, so we could have utilized that as a way to connect it and the map to show possible correlation. Or we could've explored something else entirely, such as the frequency of reviews per month as they could denote "popularity" of the individual AirBnB. In general, including perhaps one more smaller visualization could really help tie in our dashboard altogether.

Evaluation