Tableau Public Link:
https://public.tableau.com/shared/TX263Q78X?:display_count=n&:origin=viz_share_link

SARC 5400 Final Project: Concrete Jungle Journeys

      For the topic of 'extreme,' I have chosen to explore different factors that may influence how hosts determine the listing prices of their New York City Airbnb rentals. The data is made available through Inside Airbnb, an organization of activists against the adverse impact of short-term rentals, and my source of inspiration is growing up in the infamous city while hearing constant complaints about the 'extreme' housing situation. My leading question is "what factors share a relationship with the rate which hosts choose to price their listings at?" This question can then be divided into sub-questions including what is the relationship between location (by boroughs/neighborhoods) and the average price, room type and pricing, and the popularity of the neighborhood (marked by total number of reviews/bookings) with average pricing per night.

      The interactive visual seeks to display a few of the most pertinent factors in determining the average cost per night when staying at any one rental. I chose to use a Tableau Dashboard to present my findings, and I did this because it is a simple and intuitive method to arrange the different levels I was looking to show. Utilizing a combination of both highlights and filters, my end product allows users to navigate the nuanced relationships within the different variables in a more friendly manner. By re-orienting how the audience views this collection of data, my goal is an attempt at exposing some interesting correlations with the pricing strategy of property owners that may not have been previously acknowledged from the data's raw form.

      In my interactive visualization, there are three main visual strategies: the map of the data, the treemap of neighborhoods, and the price distribution box plots. The map was selected to give the user a general overview of each data point. By offering a view into each borough, users can get an overall sense of where listings typically collect, how dense an area is, and specific details into individual rentals when hovering the mouse over. Additionally, this map serves as the main visual to reference when using other filters such as room type, pricing, and neighborhoods. As the main strategy, it is complex enough to offer multiple layers of information while still refraining from appearing overwhelming in respect to other methods I had considered. On the other hand, with a smaller role, the interactive tree map serves to display which neighborhoods within each borough of the city garners the most attention, using the number of reviewers per listing. This visual offers the percentage of total bookings made within each neighborhood, and its secondary role is to act as a button that filters the main map visual (click on it!). I chose a tree map to visualize which neighborhoods that other users may interest you because it is a straightforward visual that essentially processes as parts of a whole. The sizing of the nested rectangles is an easy way to grasp the message of the visualization when viewed, and it does not take away from the main visual much since it is not visually intricate. Lastly, the boxplot gives a final reading of the main question, how these different filters/factors change the average pricing. The boxplot can provide a holistic view of the spread of pricing within each borough.

      In the beginning, I started off with a very rough draft of different features I wanted to make sure I included so that users can freely explore the different data conditions that pertain to their particular interests. This included a map, a search filter, a slider, and a generic line graph or chart of some sort. When I sought to develop this idea, I ran into my biggest issue, the data available. I struggle to find data with all the variables I was interested in for my question of interest. Any data manipulation that could be done was extremely complex and difficult to

perfect with such a large data set and such a short amount of time. My original idea involved hotel listings, however, after previewing the available public datasets, I settled with Airbnbs.
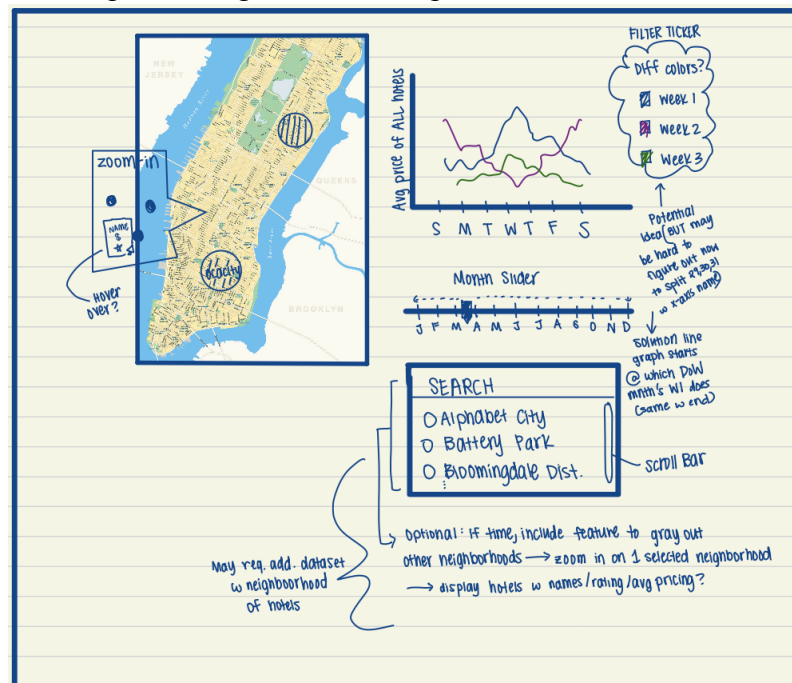

Moving onto my first prototype, another small issue I faced with the data set is that the number of reviews left for a listing is used in place of information on the number of overall past bookings. My intentions to reveal which neighborhood garners the most attention may be slightly off since leaving a review does not equate to the number of stays. For instance, some tenants may not have left any reviews or there may be a correlation between increasing review number for a place and decreasing experience/rating. Thus, this data was not entirely accounted for. Secondly, I faced difficulty when determining whether or not to keep the outlying prices. For instance, some listings would be listed as zero in place of currently unavailable housing situations. Another example is that some listings would be priced around $10,000 per night which doesn't seem realistic. Upon further investigation, I was able to discover that these listings were either listed absurdly high to discourage potential tenants from booking during a certain time period (i.e. a host taking time to improve a rental after a poor review) or for a luxurious rental that has a higher minimum nights stay required. Some peers supported keeping these outliers to play more into the topic of 'extreme.' While I did keep the outliers in the data, I removed the listings of zero dollars, and I also slightly tweaked the axis of the boxplot visualizations. Originally, the x-axis tick marks of the boxplot would automatically select the range depending on the data points of each borough. I ended up making the axis operate under a fixed range, and I also capped this specific visualization at 500. Other than the countable few listings offered at thousands, that was not the norm, even for outliers. I believed it would be easier to notice the slide up and down the axis of the box plot for each borough with a fixed axis. This brings attention to the drastic pricing happening in specific boroughs like Manhattan which can be more beneficial for the intention behind this project.

Concrete Jungle Journeys
Navigating the (Extreme) NYC Airbnb Listing Prices
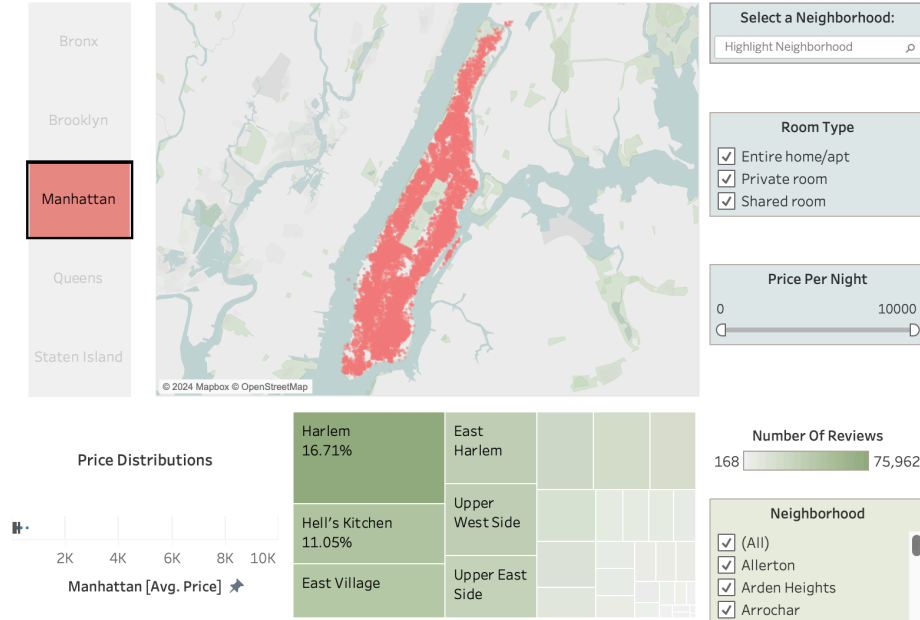
The other minor issue I encountered when working with the data is that there are so many data points, which is expected being a large data set. However, this made the data points displayed on the map difficult to read. I wanted to work more with utilizing color and sizing of the data point to reveal something interesting, but this made the overall view of congested boroughs like Manhattan too confusing. Adjusting the point opacity and sizing was also difficult in that decreasing these attributes made less dense boroughs like Staten Island seem bare. This required a lot of playing around with the aesthetic feature of the data points.
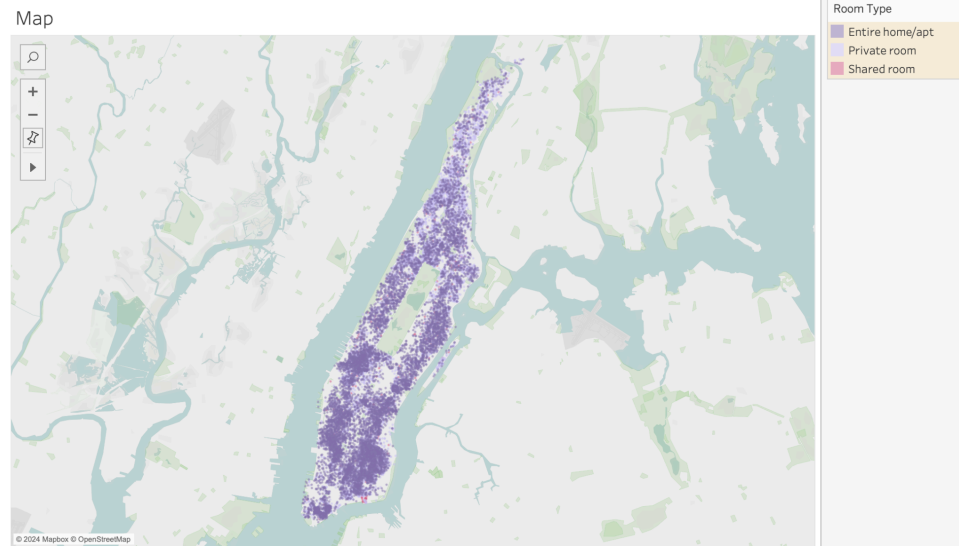
Lastly, I conducted some research on the existence of short-term rentals (listings with a minimum number of stays under 30 days) in New York City and came to find out that many listings are considered under city laws. New York City only allows short-term rentals if hosts register with the city and are physically present during the rental. While Airbnb offers an innovative approach to home-sharing, the influx of rental properties under a certain few

individuals directly contributes to the negative housing situation of the city. These hosts would buy out multiple rooms or housing options to rent out to city visitors, making it harder for locals to fight the payment increases. Thus, the city enacted the condition that the host must be physically also present during the tenant's stay. I included a button to filter short and long-term rental stays, but I explored this further when organizing the static visual.

Ultimately, I discovered that the variety of factors I sought to explore did, in varying degrees, influence pricing. The borough buttons filter all the visualizations to only present data from the selected borough. Unsurprisingly, you can observe that Manhattan has the greatest spread in pricing distribution with the Upper East Side exhibiting higher rates of outliers beyond the upper extreme. In addition, while entire homes and apartments are priced at higher levels, private and shared rooms are priced generally the same on average.