

Airbnb Listings in San Francisco

Introduction

About Data Set

The data “Inside Airbnb” site is sourced from publicly available information from the Airbnb site. Link: <http://insideairbnb.com/get-the-data.html>. This data set list all the Airbnb Listing Information for a bunch of cities. There are three tables: “listings”, “reviews”, and “calendars”:

- Listings: For each record in the data set, it represents one listing on the Airbnb and contains information about location, average price, availability within 1 year and the number of reviews. It contains listing id and host id. One host can have multiple listings.
- Reviews: Records the review id and review time for each listing
- Calendars: Records the availability and price from 2019 March to 2010 March

In order to get the geometry of each neighborhood, I also joined this data set with a shapefile found here:

<https://data.sfgov.org/Geographic-Locations-and-Boundaries/SF-Find-Neighborhoods/pty2-tcw4>

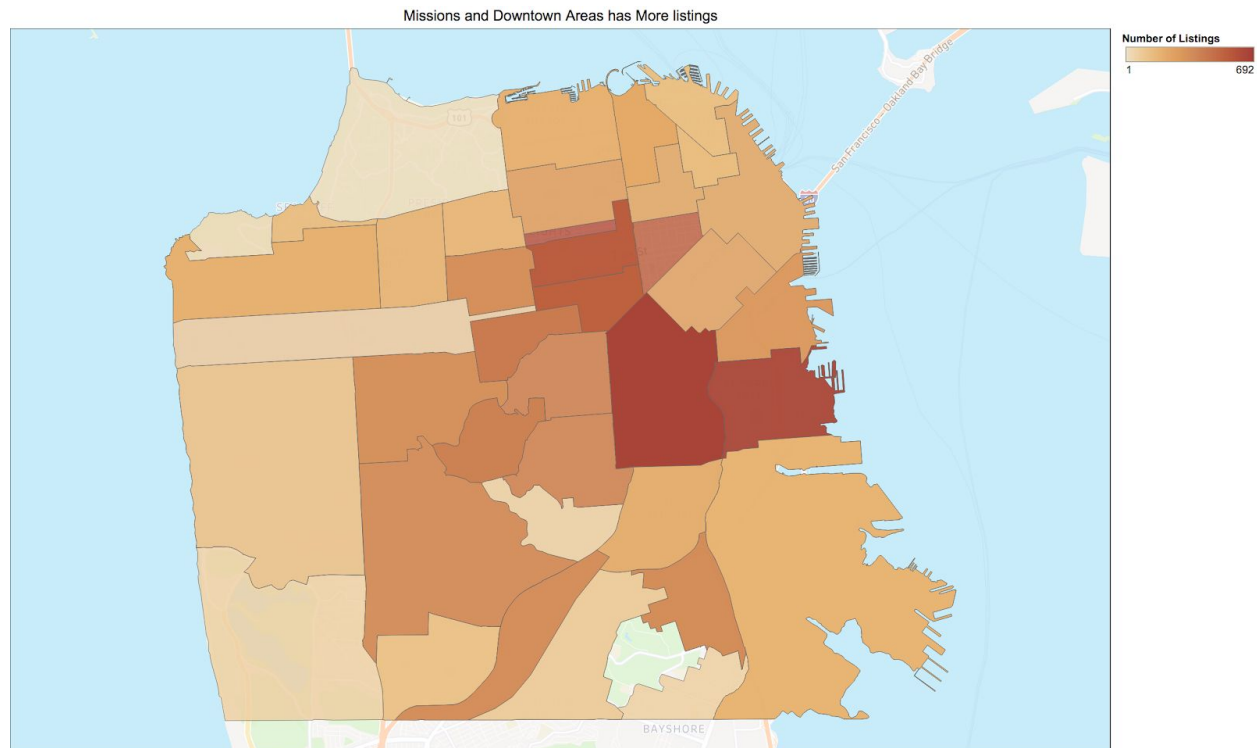
Motivation

I am a big fan of traveling. When I travel to a new place, I like living in Airbnb because of the cheaper price and nicer hosts. With the growth of Airbnb, more and more people are willing to become a host and more listings are available. There are more than 6 million listings available worldwide. There are 102% increase in the number of senior hosts over 60 since 2017. I found that some of the hosts become full-time hosts and run more than 1 listings. The Airbnb listings are experiencing commercialization and hosts choose their location and set their price intentionally. To narrow down the scope of the problem, I only looked into the Airbnb Listings in San Francisco, to see whether there is a pattern of Airbnb listings' location, room type, price, etc.

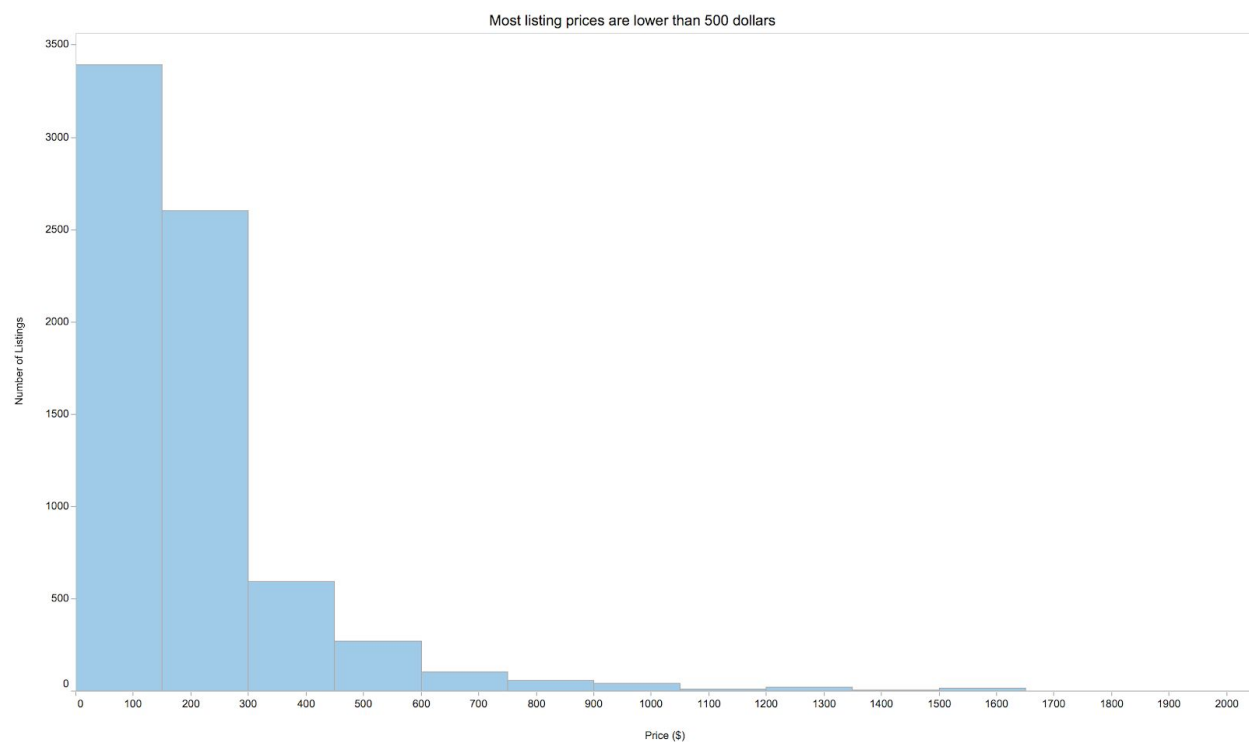
Summary of Data & Storyline

In this section, I will use various visualization to summarize data and uncover patterns in this data set.

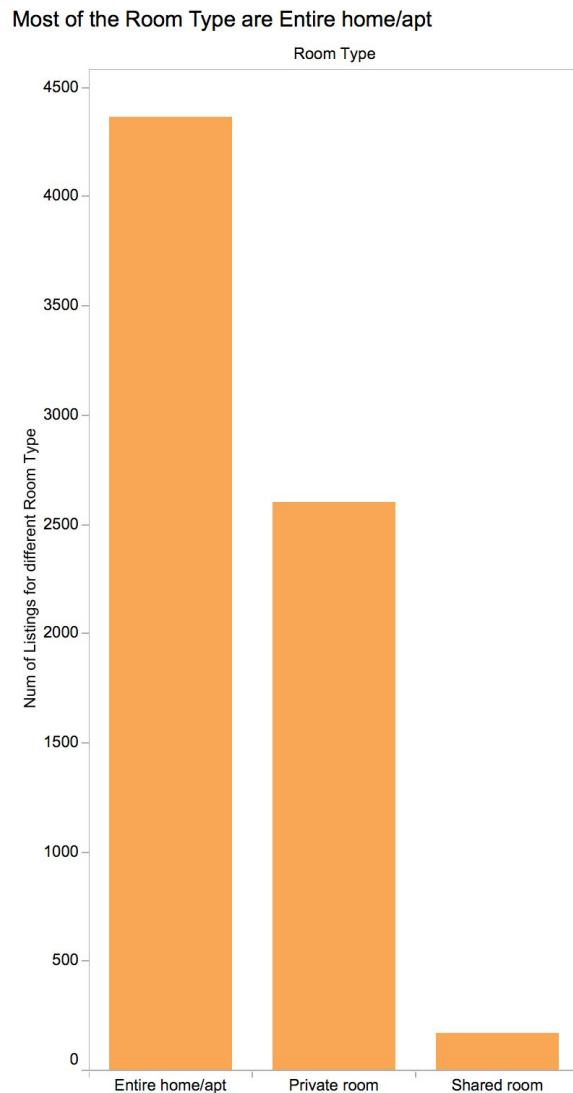
First of all, I use Choropleth Map plots the location of listings. Most of the listings are located in the Mission and downtown area in San Francisco.



According to the histogram, the distribution of average prices has a long right tail. I found most of the average prices for the listings are lower than 500 dollars.



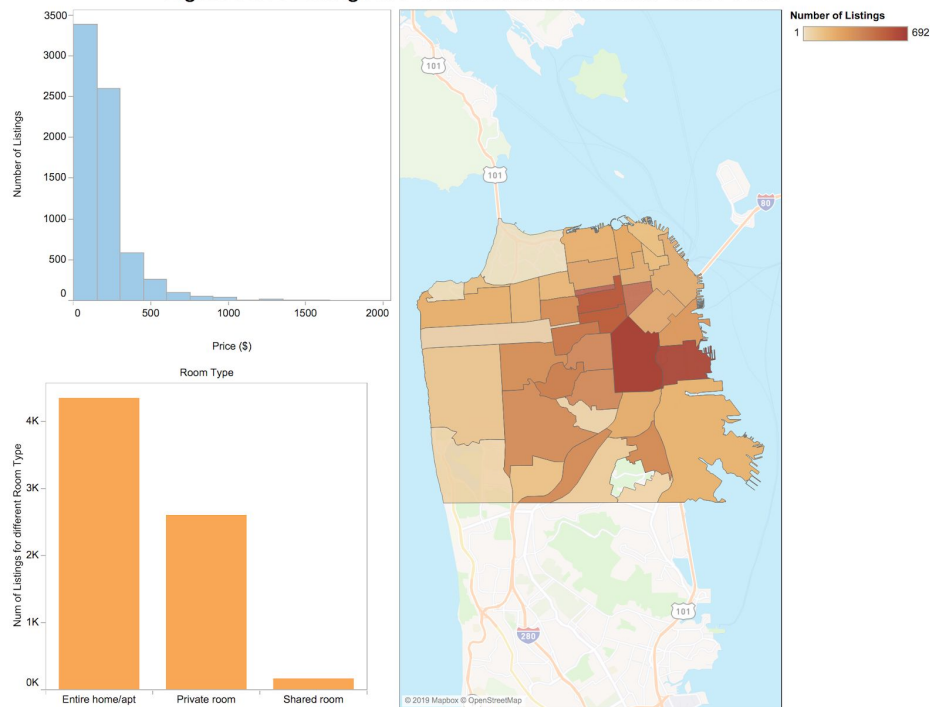
As for the room type, for the listings in San Francisco, most of them are entire rooms/apt. We can see from the plot below:



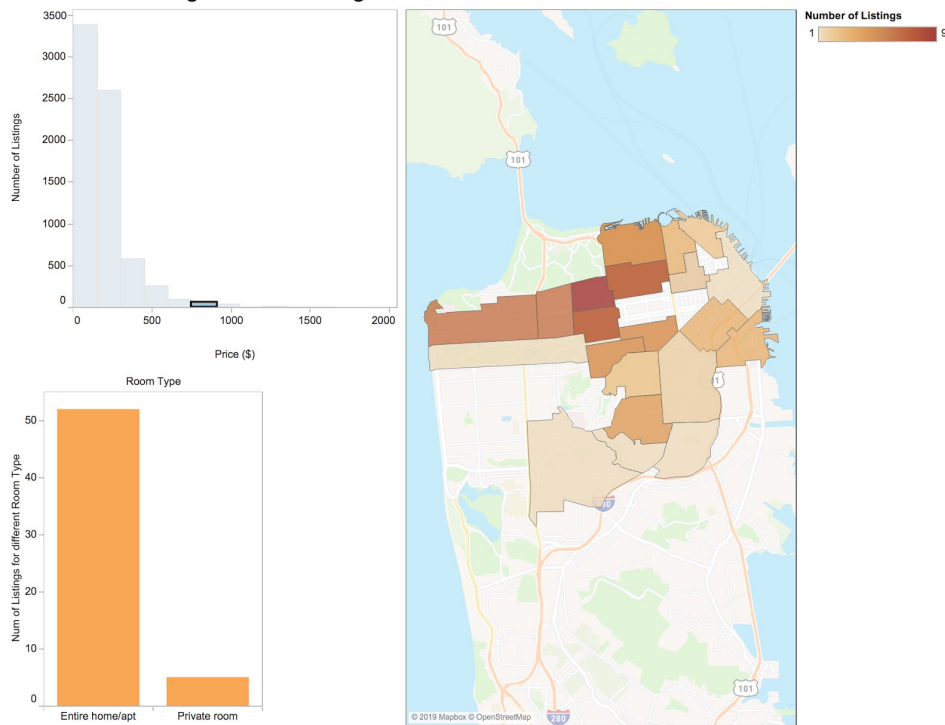
To explore the relationship between price, location and room type, I created an interactive dashboard which can be found in the link below:

https://10ay.online.tableau.com/t/ireneyang218/views/DataVizFinal/InteractiveDashboard?iframeSizedToWindow=true&:embed=y&:showAppBanner=false&:display_count=no&:showVizHome=no&:origin=viz_share_link

Higher Price Listings are Concentrated at Golden Gate Areas



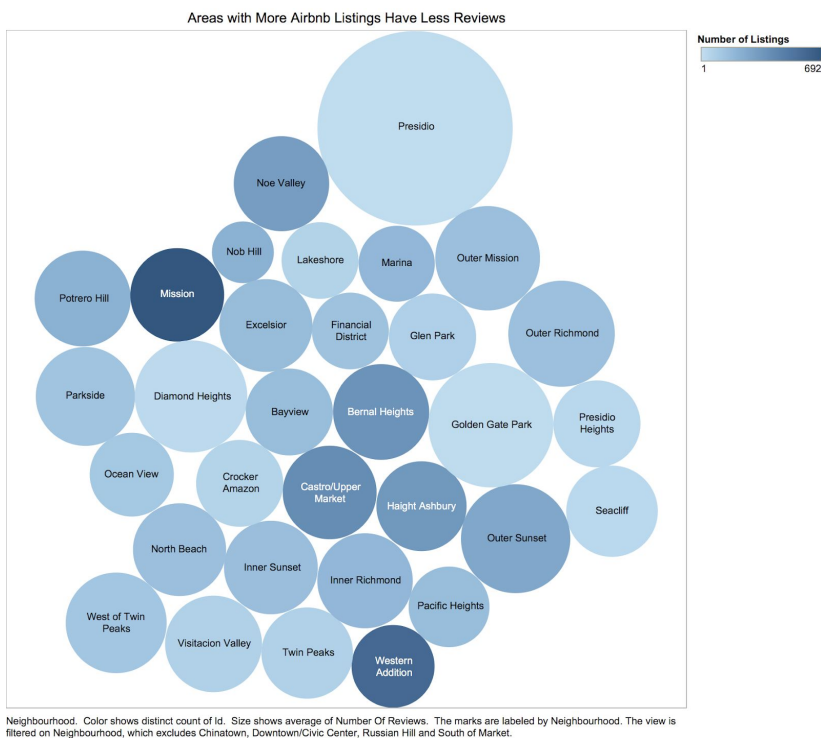
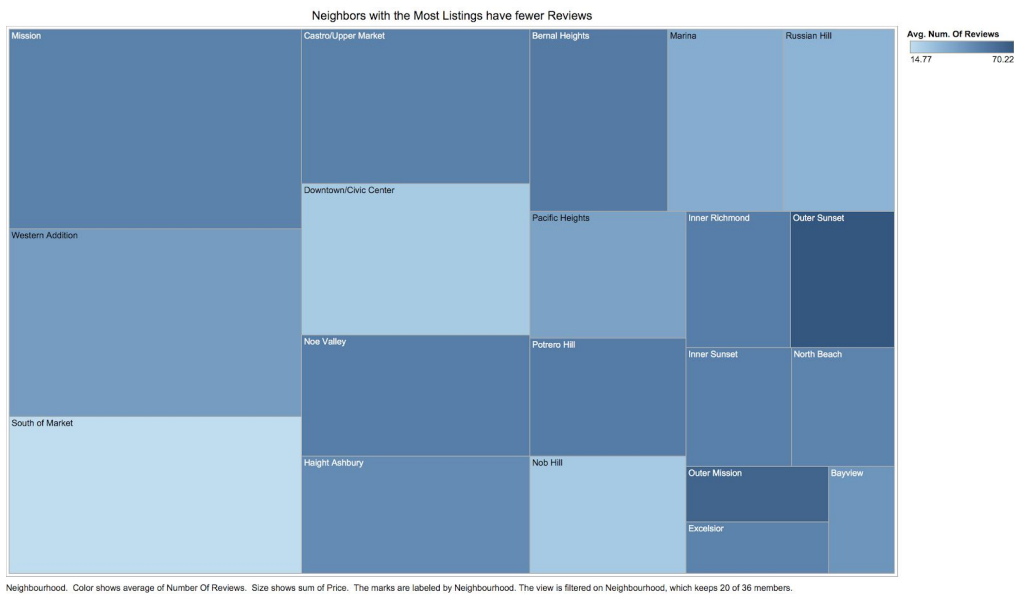
Higher Price Listings are Concentrated at Golden Gate Areas



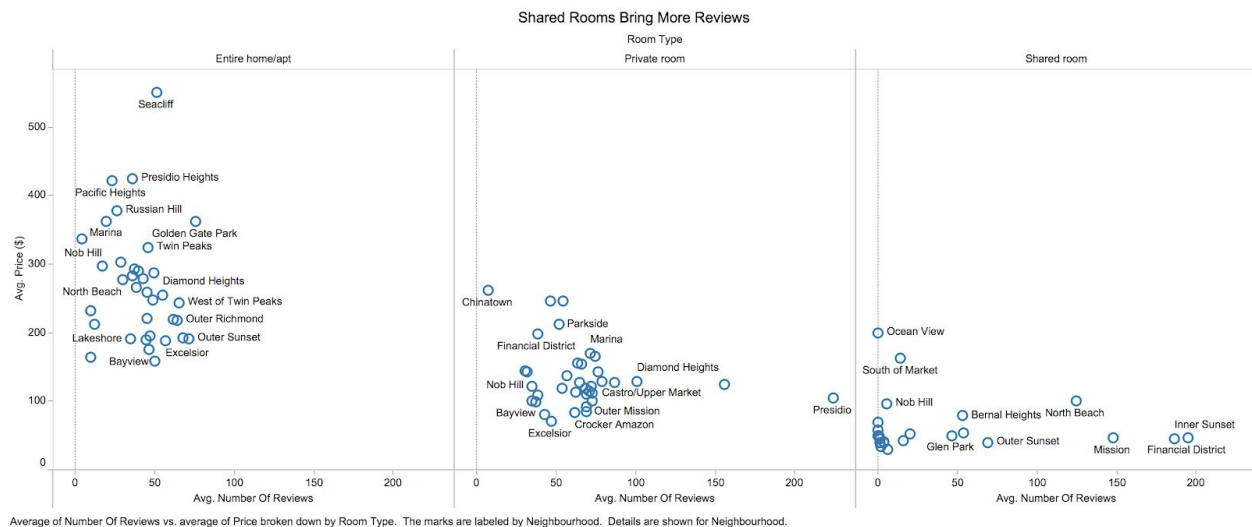
According to this plot, I found that the listings with higher prices are more likely to locate in the golden gate area. These apartments are more likely to have a better view and most of them target to family travelers who want to enjoy the great view have their own space.

Review Volume

Apart from location and prices, the Number of reviews is an important element for clients to evaluate the Airbnb listing and make their choice. It can show the satisfaction of previous clients to the Airbnb listings. The treemapping and bubble map below are sized by the number of listings in that location and colored by the average number of reviews for each listing. They shows that some neighbors with the top number of listings have fewer reviews.



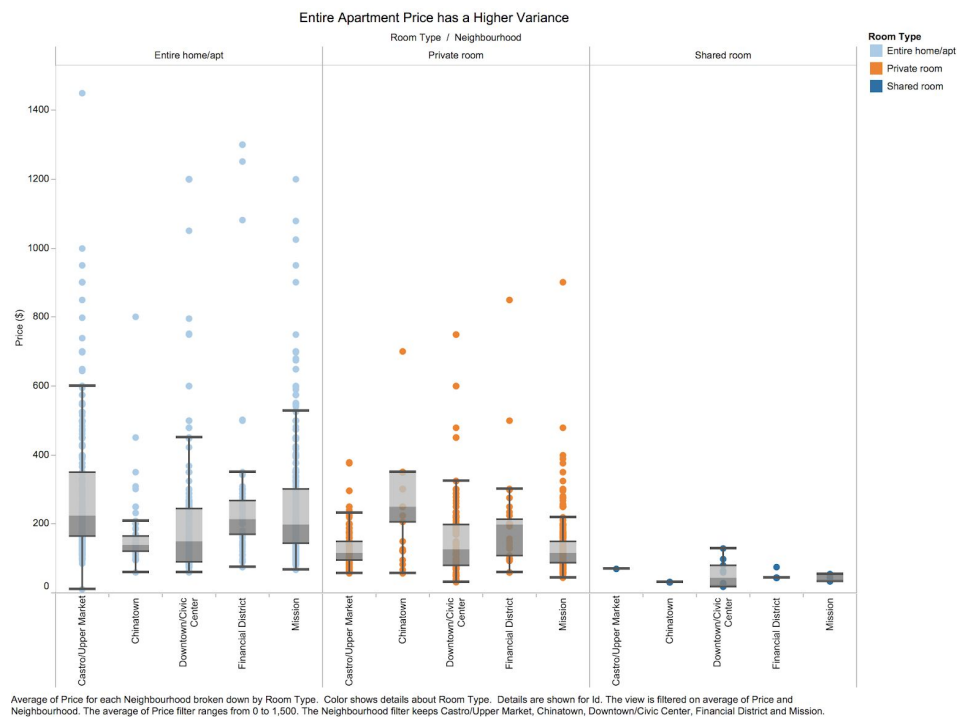
Room type and prices are also important elements to consider for both hosts and clients. It might be related to review volume.



In this plot, I want to investigate what's the relationship among room type, number of reviews and prices. Each point here represents the aggregation of the listings within one neighbor. Y-axis represents the average price, the x-axis represents the average number of reviews for the listing within that neighbor. Neighbor represents the predefined area, like Mission, South of Market, Financial District, etc. According to the plots, I found the shared room is much cheaper than the entire apartment but similar to a private room. In the meantime, the shared room tends to bring more reviews, especially those shared room located in the Financial District. The average number of reviews are around 200. The reason might be those guests who are willing to share a room with others tend to be more outgoing and willing to share their experience as well.

According to the tree mappings, we found that the Mission, Western Addition, South of Market, Castro/Upper Market, and Downtown/Civic Center are neighborhoods that have the most number of listings. Let's focus on these places and investigate the relationship between price, room type.

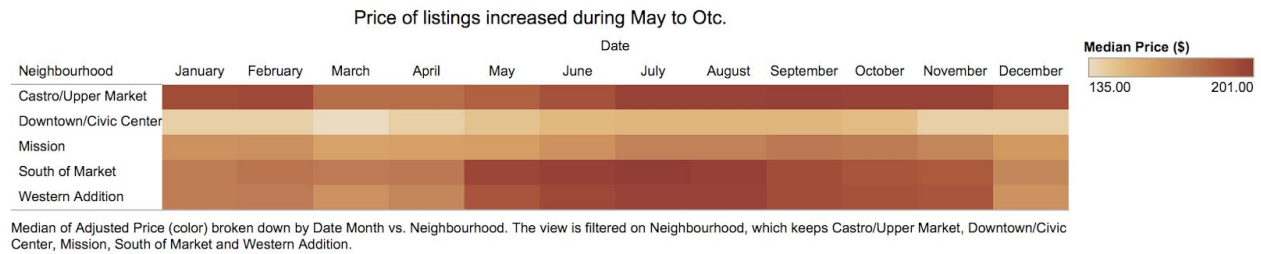
According to the bar plot and box plot below, we can see that, in general, the price of the entire apartment is higher than the other two. But China Town is an exception, where the private room has a higher average price. Besides, the price of the entire apartment also has a higher variance than the other two. It might be related to the higher variance of quality and size of the house.



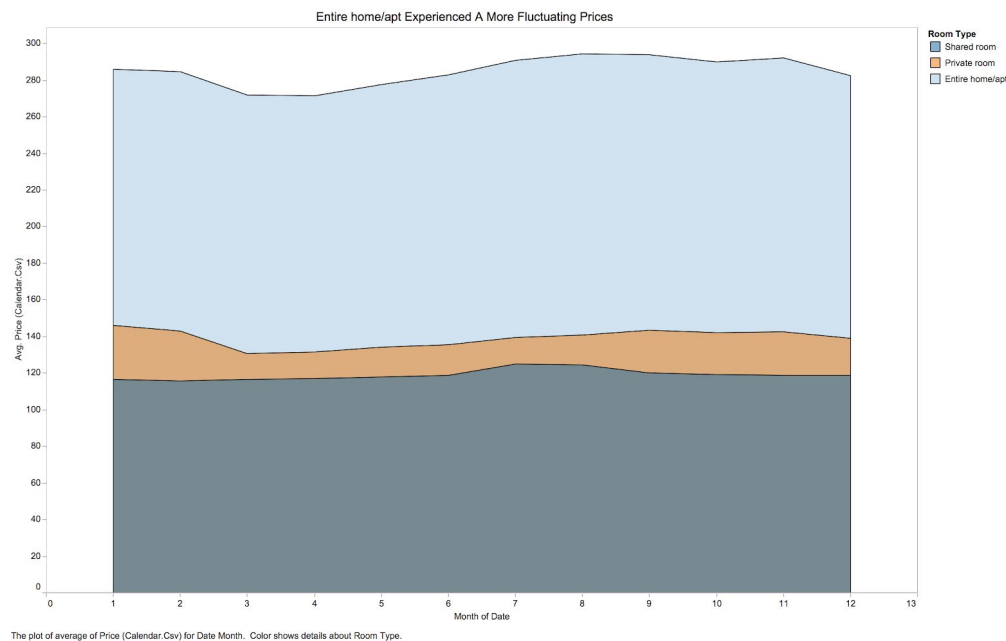
Prices & Month in Year

Prices are one of the most important element for hosts to consider if they want to make a profit from the Airbnb business. It's highly related to location, room type and time.

According to the heatmap below, Castro/Upper Market and South of Market have a higher price in general. All of these areas experience an increase in prices from May to October.



The stack area plots also indicate the same trend. But in addition, it shows that the entire home/apt shows more fluctuated price change.



Hosts with Multiple Listings

As I mentioned in the introduction, it's very common that one host runs several listings. It indicates some hosts actually run the Airbnb as a business which might violate the short term rental law. The connection map connects the listings that run by same hosts and color by the number of listings per host. I found that most of this kind of hosts run their listings in the Northwestern area.



Summary & Conclusion

In this project, I visualize the Airbnb Listings in San Francisco and found that the Northwestern area has the most number of listings; golden gate bridge area has the most expensive listing. Also, it's very common the hosts run the Airbnb as a business and run several listings at the same time.

Appendix & Github

https://github.com/IreneYang218/DataViz_Airbnb