NYC Airbnb data Project report

Ist 652

**Courtney Zimmer-Bartels**

**PROJECT SCOPE**

The analysis will include various tabular reports. These tabular reports will show the percentage of Airbnb listing by NYC boroughs and room type. We will also be looking at the average price of room types and regions, as well as the average listing availability of each region. The data sets I have identified are both structured data sets. An aspect of this project that creates hesitation is the lack of experience working in python and with unstructured data. The focus will be in importing and analyzing the structured datasets identified above and using exploratory analysis to look for any connection between the rise in Airbnb rentals to a decrease in the apartment rental market.

### DATA SET

Airbnb, Inc. has quickly become the go-to marketplace for vacation rentals in the United States and abroad. Founded in 2007, Airbnb now claims they have on average 2 million people staying on Airbnb per night in over 1,000 cities globally. Airbnb provides data for download by city and per month/per year starting with January 2015. I exported the January dataset for 2015 – 2020, containing over 19 columns, containing information such as the listing id, host id, neighbourhood, price, availability and reviews. For my second data set, I utilized Streeteasy, a NYC only apartment listing website. The Streeteasy dataset included rental inventory for the 5 NYC boroughs for every month starting January 2010 – July 2020.

Data Source:

1. <http://insideairbnb.com/get-the-data.html>
2. <https://streeteasy.com/blog/data-dashboard/?agg=Total&metric=Inventory&type=Rentals&bedrooms=Any%20Bedrooms&property=Any%20Property%20Type&minDate=2010-01-01&maxDate=2020-06-01&area=Flatiron,Brooklyn%20Heights>

### Data Cleanup

The Streeteasy data set required some preprocessing before any analytics could be gleamed from the data. I first subsetted the csv file into 2 separate data frames, one containing only the borough rows and another with only the neighbourhood rows (neighborhoods within each of the 5 boroughs). For the neighbourhood dataframe, I chose to only look at the 10 neighborhoods that have the most amount of Airbnb rentals as of 2020. The next obstacle was that the data frames were structured so that the boroughs/neighborhoods are the rows and the dates are the columns. In order to group the years together, I transposed the rows and columns, so that my dates were now the rows. This allowed me to create new additional columns, one for Year and Month, taken from the YearMonth column.

### Analysis

**Has in the introduction of Airbnb to NYC negatively affected the rental market?**

1. **Exploratory Analysis**

Starting with the Airbnb datasets, the first step was to group each of the datasets and then ultimately merge them into a dataframe. The first grouping was to look at the average pricing by borough for each of the 6 years’ worth of Airbnb data. Looking at the results, there are fluctuations in the average price, but no significant jumps of any kind. Next, we added to our initial grouping, but adding room type in the group by. The results are to be expected in that ‘Entire home/apt’ are the most expensive to rent on Airbnb consistently over the years. Next, was to look at the Airbnb inventory by borough, room type and top 10 neighborhoods. Some very interesting points jump out when looking at the Airbnb inventory over time. Manhattan inventory decreased almost 13% between 2015 & 2020. Whereas Brooklyn and Queens both increased 5.4% and 5.7% respectively. The Bronx and Staten Island also increased their Airbnb inventory although at a rate slower than 2%. The last table to examine is the number of Airbnb listings per neighborhood, sorted by the top ten neighborhoods in 2020.

Our Streeteasy data only has rental inventory by borough or neighborhood, but due to our data cleanup we can now group our data by year and month. We can see how each borough rental inventory has changed over the year and despite Airbnb, the inventory has mostly increased in ever borough. We can also see that there is a spike in inventory in the summer months, which is what we would expect to see. The last table is filtered to show the ten neighborhoods we identified in the Airbnb data as having the most amount of Airbnb listings. There is some variation in inventory over the years for these specific neighborhoods, but no decreases that would suggest that Airbnb has any kind of impact on the NYC rental apartment inventory.

1. **Visual Analysis**

I created two bar graphs, the first showing the amount of Airbnb listings per borough, per year. You can see that every borough grew the amount of Airbnb listings. You can also clearly see that Brooklyn had the most amount of growth over the 6 years. The second bar graph represents the apartment rental inventory again by borough and year. Visually you can see that every borough experienced growth, with the exception of perhaps the Bronx. The first line graph is representing the Airbnb listing price by borough by year. Relatively stable, the only boroughs to experience any kind of fluctuation in pricing appear to be Staten Island and Manhattan. The last plot is showing the apartment listing inventory of the top 10 neighborhoods. You can see that none of the neighborhoods have a decrease in inventory over the years.

### Conclusion

As previously discussed in the analysis section, no significant correlation between the rise in Airbnb in NYC and a decrease in apartment rental inventory. Perhaps the strict laws in NYC to prevent short term rentals in apartment buildings have worked in preventing a loss of rental inventory for those who need it. Further study should look at other major cities to see if Airbnb had any effect on the real estate market, such as San Francisco, Seattle, Boston and Chicago.