New York City Airbnb and Housing Prices

Do More Expensive Neighborhoods Have More Expensive Airbnbs?

Molly Chiang

Goals and Background

The goal of this project was to analyze housing and Airbnb prices in New York City. New York City has one of the largest, most unique, and most analyzed housing/lodging markets in the world and is subject to a number of stereotypes. Through this project, I hope to shed some light on what ideas and theories about New York housing are false, and which hold some truth.

I addition to exploring and presenting data about housing and Airbnb prices, I posed the research question, do more expensive neighborhoods (in terms of housing) have more expensive Airbnb prices? I used a linear regression in an attempt to answer that question.

Data

The data for this project comes from two different sources: Zillow's research platform and InsideAirbnb.

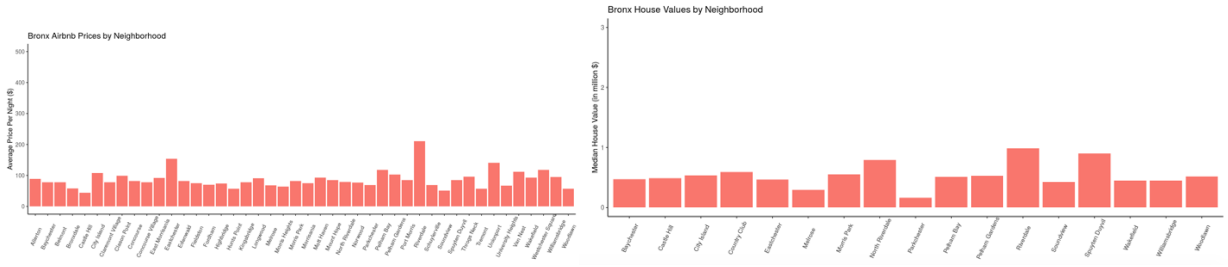
Zillow is an online real estate database. Zillow collects data on houses, apartments, condos, etc. that are for sale and for rent all across the United States. It tracks a number of variables (demographic, location, credit scores, etc.) including listed prices and sale prices which are used to make algorithms which create accurate estimates of house values. The data for this project comes from Zillow's 'Zillow Research'", platform, which is independent from Zillow's revenue center and aims to provide open, accurate data on the US housing market.

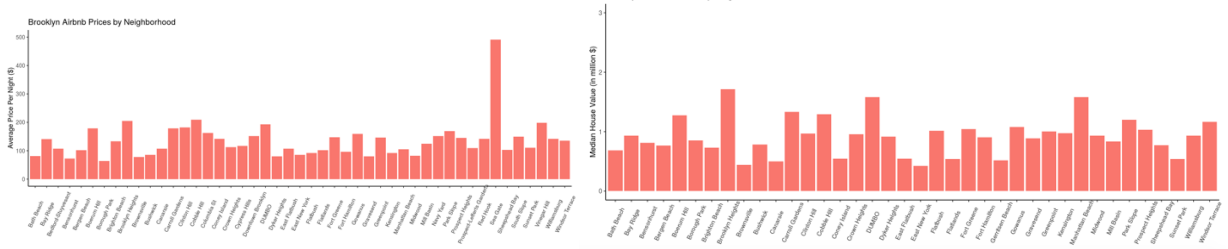
The specific data from Zillow used in this project is from September 30th 2019. The data contained information on this month's zhvi value for all homes (home, apartment, condo, etc.) on the neighborhood level. Thus, there was data on neighborhoods of metro areas all across the US (I, however, selected only NYC data to analyze). The zhvi (Zillow Home Value Index) value is Zillow's smoothed, time-dependent measure of the median estimated home value across a given region in USD.

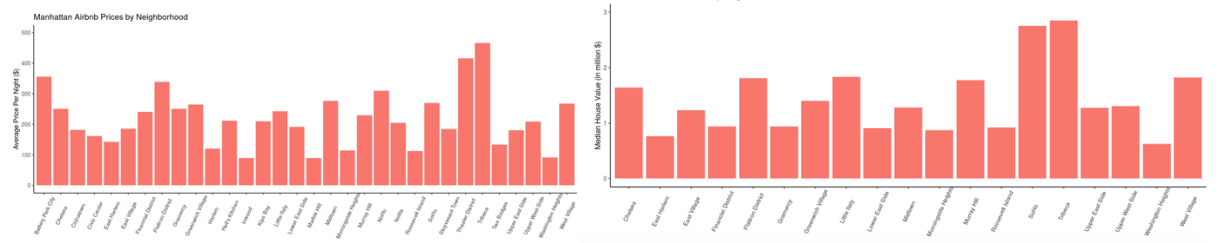
Airbnb is a website that allows users to arrange and offer lodging. The company is one of the biggest lodging platforms, with big implications on local hospitality industries. The data used for this project comes from 'Inside Airbnb' which is a set of data (independent from Airbnb the company) created to help people explore how Airbnb is really being used around the world. The specific data for this project contains host, lodging, location, price, and review information for airbnb's in New York City. I also used data from Inside Airbnb that contained shapefiles for all the neighborhoods in NYC in order to plot results on a map.

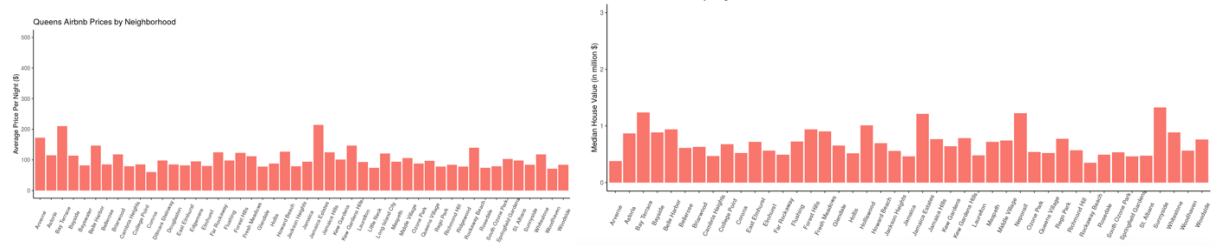
Methods and Results

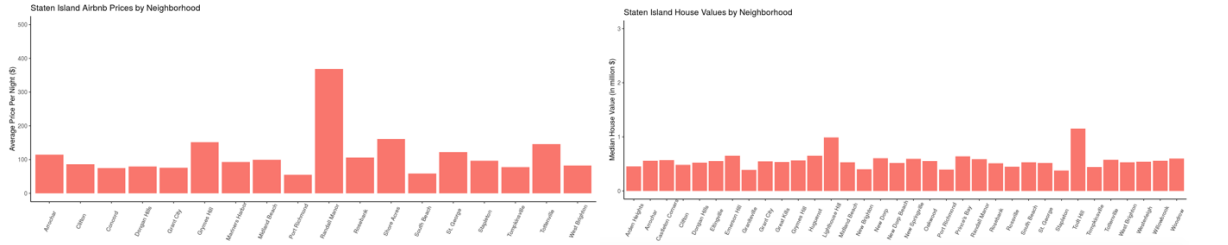
In order to analyze my data, I first simply made bar graphs showing the neighborhoods of the different boroughs and their average Airbnb prices per night and median home values. In order to create these graphs, I first imported my data, cleaned it (selecting only relevant variables), and modified it slightly for best presentation. The resulting graphs are shown below. Each graph also had a set y axis (for all Airbnb prices and for all home values) in order to more easily compare neighborhoods.



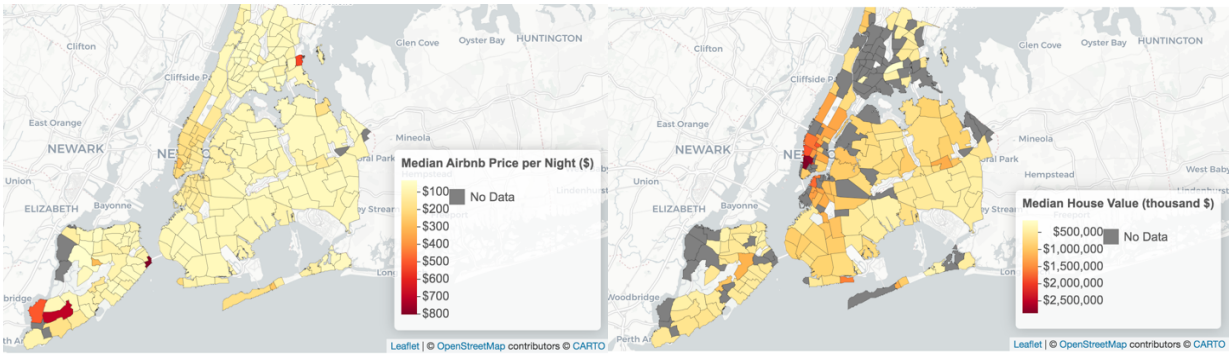








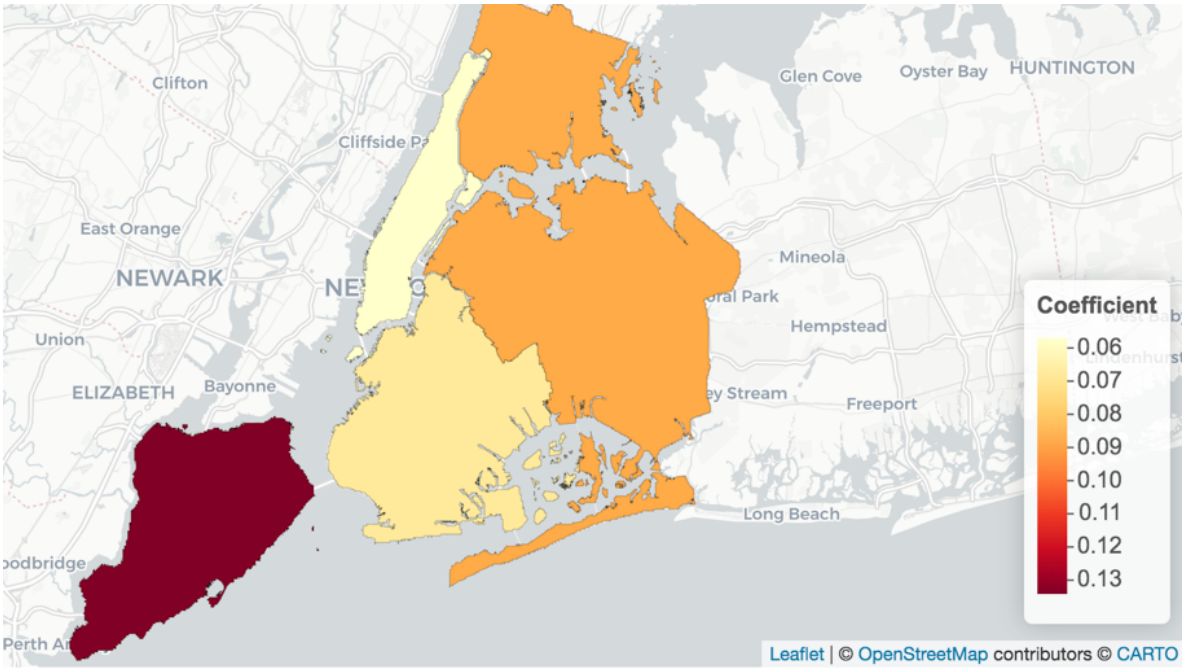
In addition to these bar graphs, I created two heat maps, visually showing the neighborhoods of the city and their corresponding median Airbnb price per night values and median home values (according to the Zillow index.



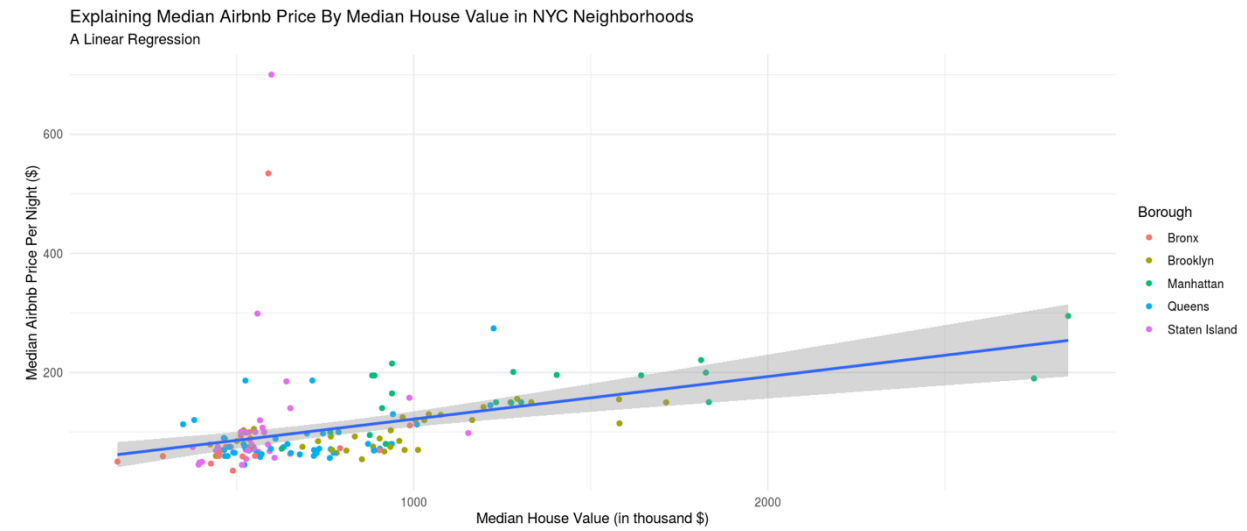
Looking at this data confirmed some stereotypes about New York City Pricing. For example, Manhattan had the highest median home values and the highest average Airbnb prices, and Staten Island had the lowest. However, it also revealed some new surprising information, like that the two neighborhoods with the highest median Airbnb prices were in Staten Island!

Statistical Analysis

A statistical analysis was run in an attempt to answer my research question and explain Airbnb prices by median home value prices in a neighborhood of NYC. A simple linear regression between Airbnb price per night and median home value was used. The regression was run in the different boroughs of the city. The coefficients ranged from .06 (Manhattan) to .13 (Staten Island) and were plotted on a heat map shown below.



For the overall regression, all the neighborhoods were used, and the overall coefficient was .07. The graph below shows the results of the overall regression. The coefficients of these regressions indicate that, according to this data, if the median home value of a neighborhood increases by $1000 (one unit of measurement in this analysis), Airbnb price per night should increase by the coefficient value in dollars. The overall coefficient of .07 thus indicates a slight positive correlation of home values and Airbnb prices throughout New York City.



Coefficient	5th Percentile	95th Percentile	R Squared
0.07	0.05	0.10	0.14

In addition to the coefficient value, I also displayed the 5th and 95th percentile values and r-squared value for the overall regression. The 5th and 9th percentile values indicate the credible/prediction interval for my coefficient value and thus an indication of uncertainty. These values tells us that there is a 90% chance that the true value for the predicted unit change in Airbnb price associated with one unit increase in home value falls within that range. Since the range (.05-.10) is fully positive, we can be pretty sure that there is at least some positive correlation between the variables. The r-squared value gives a measure of the fit of this model to the actual data. The value of .14 is relatively small, but since this is the only model that we have tested for this data, this value alone doesn't tell us much (perhaps this data just can't be modeled very well).

Implications

In the end, my analysis did provide some insight on my research question, indicating at least some positive correlation between home value and Airbnb prices in a neighborhood. With this information, the purpose of creating this model was as a potential predictor of Airbnb prices in different neighborhoods. I believe it could help predict what Airbnb prices would be in a new neighborhood or city that just started allowing the lodging company to be used, for example. The results of this project indicate that using the model that was created could give some

prediction interval on prices based on home values in the area, but including additional variables in the model would help to increase certainty."