2019 Airbnb New York Dataset

Ryan Meston

Bellevue University

Abstract

The goal of this project is to explore the 2019 New York Airbnb dataset to see why rental prices are set how they are. I want to see if these properties are primarily vacation homes/second homes, or if there are multiple properties owned by companies using Airbnb's platform to rent properties. The questions I want to answer with this project are:

- Which hosts are the busiest and why?
- How much money do hosts make per year that have more than 50 properties?
- Are there any relationships between prices based on the neighborhood the rental unit is in?
- Are the units cheaper than hotel rooms in the area?
- Are the busiest locations around tourist hotspots or are they buy large corporations?

The dataset being used has longitude and latitude coordinates so I will be able to plot each rental units' location. This will also help in researching the neighborhood each rental unit is in. Once we have are rental units plotted, we will be able to see where the properties are located and what they are located by. This will also help in figuring out why they rental units are price the way they are.

Introduction and Background

The dataset I used for this project was found on Kaggle.com [1]. This dataset houses all of New York's 2019 Airbnb data. The row headings for the dataset are as follows:

- id - Shows a number instead of a name of the client who used Airbnb's service.
- name - Brief description of the rental property.
- host_id - Shows a number that correlates with the host_name.
- host_name - Name of the host or company that is renting out the property.
- neighbourhood_group - Showing the large neighborhood, the rental property is in.
- neighbourhood - A more detailed name of the neighborhood that the rental property is in.
- latitude - Latitude coordinate of the rental property.
- longitude - Longitude coordinate of the rental property.
- room_type - Showing whether the room is a home/apartment/private room.
- price - Showing the price of the rental property.
- minimum_nights - Showing the minimum number of nights stayed at the rental property.
- number_of_reviews - Showing the number of reviews of the rental property.
- last_review - Showing the date of the last review of the rental property.
- reviews_per_month - Showing how many reviews per month the rental property receives.
- calculated_host_listings_count - Showing the number of rental properties, the host_id has.
- availability_365 - Showing how many days a year, the rental property is available.

During my preliminary analysis of the dataset, I noticed that there were host ids with multiple rental properties. This led me to look into the names of these host ids and I found out that there are companies that will use Airbnb's platform to rent their properties. Companies that utilize Airbnb's platform will use Airbnb to rent their properties. These properties can be rented by clients wanting to

vacation or for people who need to stay for a month or longer on work assignments. There are multiple availabilities at these properties that allow flexibility for the client.

Methodology

The main method I used for this project is exploratory data analysis. I used the latitude and longitude coordinates to map out visuals that show where the rental units are located, and I also include the price and set my parameters so that I could see what rental properties cost the most. This allowed me to look into areas that had higher prices to see if there prices where high due to the region, they were in. Some of these areas are in parts of New York that have year-round tourism.

I used pandas for the majority of this project. I like the easy way that it allows me to look up statistics in the dataset. The first thing I did when I imported that dataset was to use df.info to get a quick overview of what I was working with. From here I was able to look up a few statistics that I thought where relevant to this project. I looked up mean, min, max, and standard for rental property pricing.

----------Statistical Summary----------

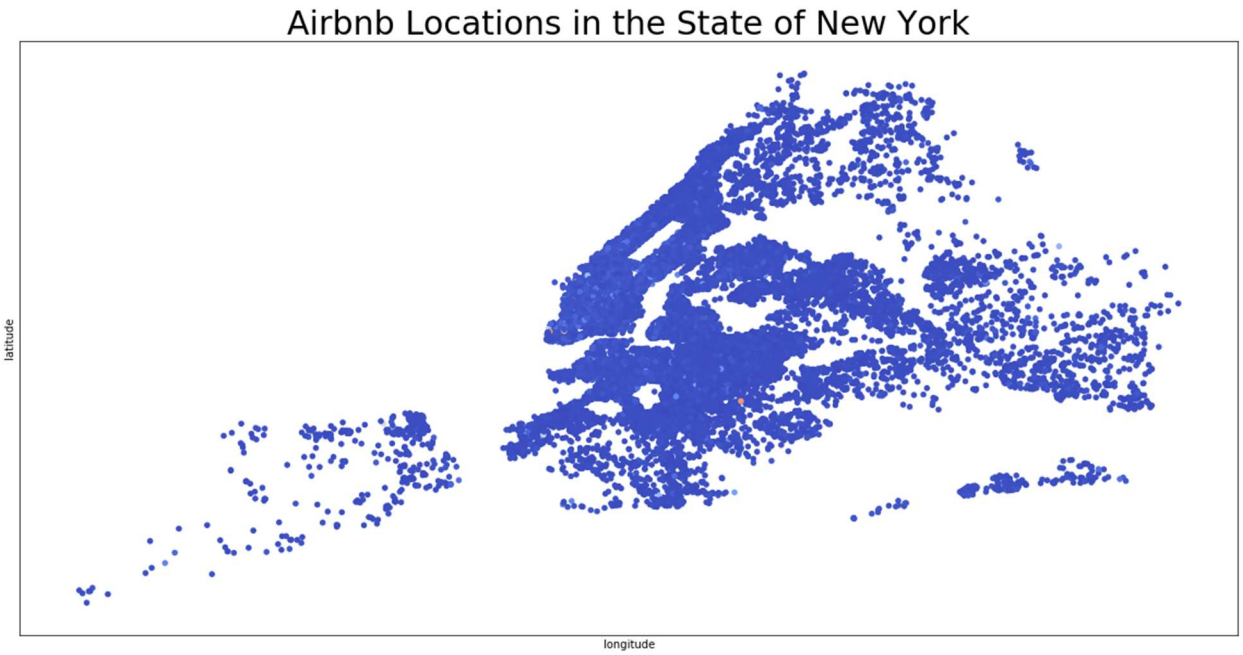
The average price per rental unit is: $152.72.

The minimum price per rental unit is: $0.00.

The maximum price per rental unit is: $10,000.00.

The standard price per rental unit is: $240.15.

After I gathered this information, I decided to plot my data using matplotlib. This allowed me to locate each rental property by longitude and latitude. I also included the price of the property and included cmap of coolwarm to show the higher priced properties as reds and the more affordable properties as blues.



I also wanted to look at certain data within the set that would allow me to look at total dollars from rental properties in 2019, total nights units were stayed in, number of reviews, and reviews per month.

----------2019 New York Airbnb Data----------
Total Dollars from Rental Properties: $7,467,278.
Total Nights Units were Stayed in: 343,730.
Total Number of Reviews: 1,138,005
Total Reviews per Month: 53,340.

One host_id that stood out immediately to me was 219517861 because they have 327 rental properties. 219517861 is also known as Sonder. Sonder is a company that uses Airbnb's platform to showcase their rental properties. Sonder is a company that allows you to rent from one night to as long as you'd like. This company has clients that will be using their platform for vacations, and it also allows them to reach clients who are doing temporary work or extended vacations. The average price for a Sonder rental property is $253 a night. The least expensive rental unit is $100 a night and the most expensive unit is $699 a night.

----------2019 Sonder Data----------
Total Dollars from Rental Properties: $82,795.
Total Nights Units were Stayed in: 4,353.
Total Number of Reviews: 1,281.
Total Reviews per Month: 397.

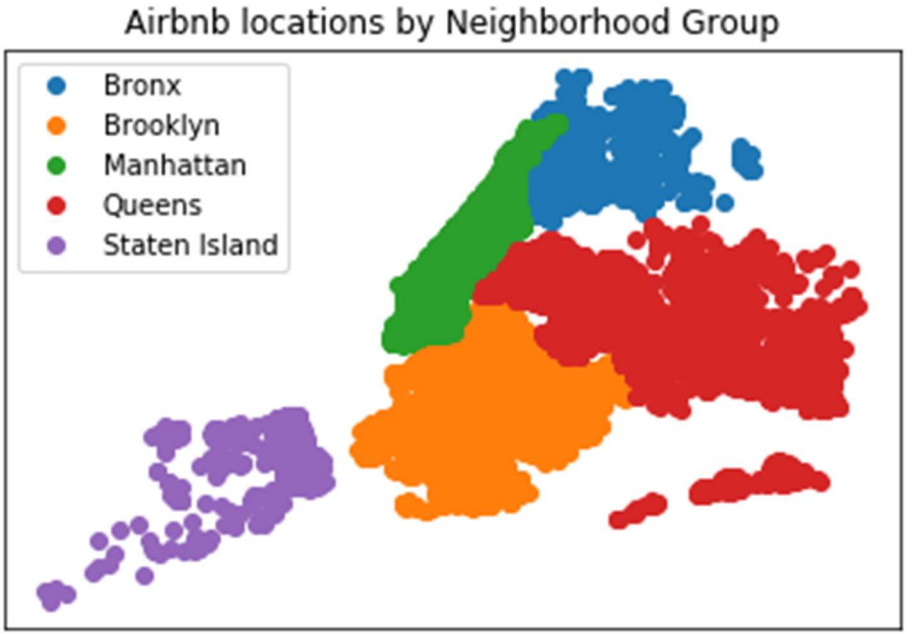
I wanted to look into more of these companies but do to time constraints I decided to focus on the top three, Sonder, Blueground, and Kara. Like Sonder, Blueground is a company that allows you to rent on your own terms. You have the options of renting month to month or longer if needed. This Airbnb client would be more of a temporary worker or someone who is going on a long getaway and wants to be able to stay in a nice size rental unit that is furnished and close to amenities throughout the city. Blueground has 232 rental properties available in the state of New York.The average rental unit for Blueground is $303 per night. The minimum rental unit is $184 per night and the most expensive rental unit is $481 per night.

----------2019 Blueground Data----------
Total Dollars from Rental Properties: $70,331.
Total Nights Units were Stayed in: 7,470.
Total Number of Reviews: 29.
Total Reviews per Month: 6.

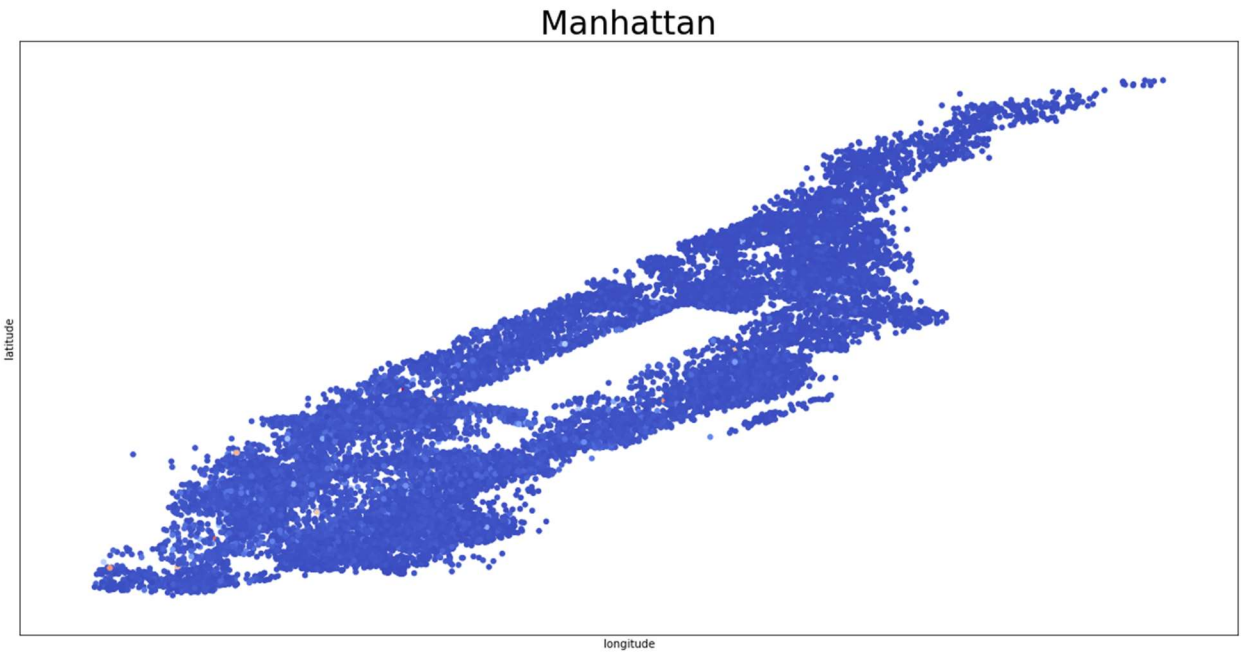
I did not find a lot of information on the internet about Kara. I can not tell if this is a company or a person (maybe even a small group of investors renting properties) renting these properties. What I do know is that they have 121 rental properties, and the average rental unit is $277 per night. The minimum price of a rental unit is $109 per night and the most expensive rental unit they have goes for $1,170 a night.

----------2019 Kara Data----------
Total Dollars from Rental Properties: $33,581.
Total Nights Units were Stayed in: 3,767.
Total Number of Reviews: 65.
Total Reviews per Month: 3.

Another approach I took was to plot longitude and latitude by rental property price based of off the neighbourhood_group. Below we can see the rental properties color coded by which neighborhood they are apart of.



I decided to map each neighborhood group to see each of the Airbnb locations in more detail. Manhattan was my favorite to plot because I thought it was really cool seeing all of the Airbnb locations and also seeing the white rectangular strip of Central Park in the middle of Manhattan.



It is hard to see in this picture but in my code book you can see the heatmap showing a few darker orange and red spots indicating higher priced rentals. I was also surprised at the consistency of rental prices around the area as well. There was one price that threw me off at a zero price, I was not able to see why that was listed at zero. The average price of a rental unit in Manhattan is $196. The most expensive place to stay for one night is $10,000.

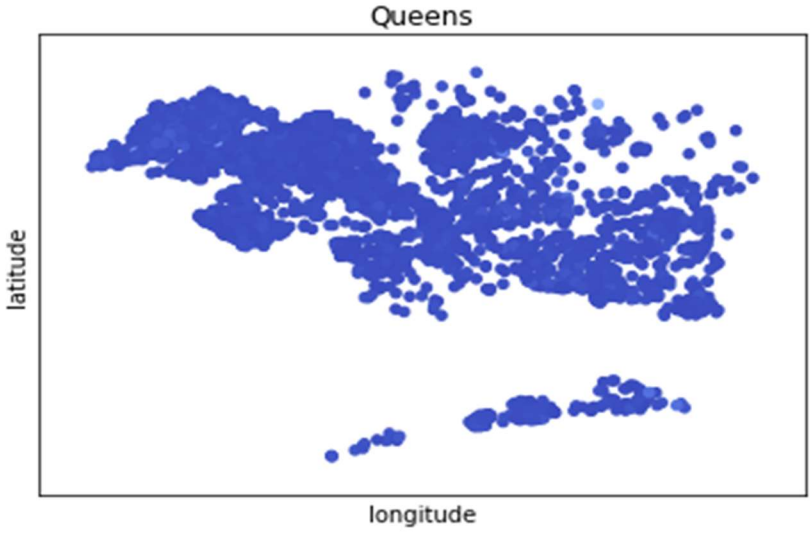
----------2019 Manhattan Data----------

Total Dollars from Rental Properties: $4,264,527.

Total Nights Units were Stayed in: 185,833.

Total Number of Reviews: 454,569.

Total Reviews per Month: 21,158.



In Queens, the average price per night is $99. The most you will pay per night at a rental unit in Queens is $10,000.

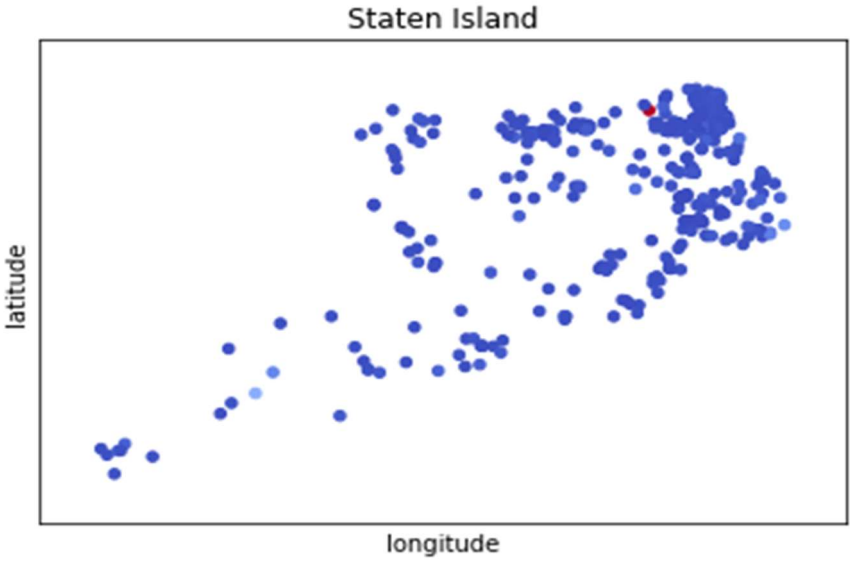
----------2019 Queens Data----------

Total Dollars from Rental Properties: $563,867.

Total Nights Units were Stayed in: 29,358.

Total Number of Reviews: 156,950.

Total Reviews per Month: 8,879.



In Staten Island the average price per night for a rental unit is $114. The most expensive rental units have $5,000 per night price tag.

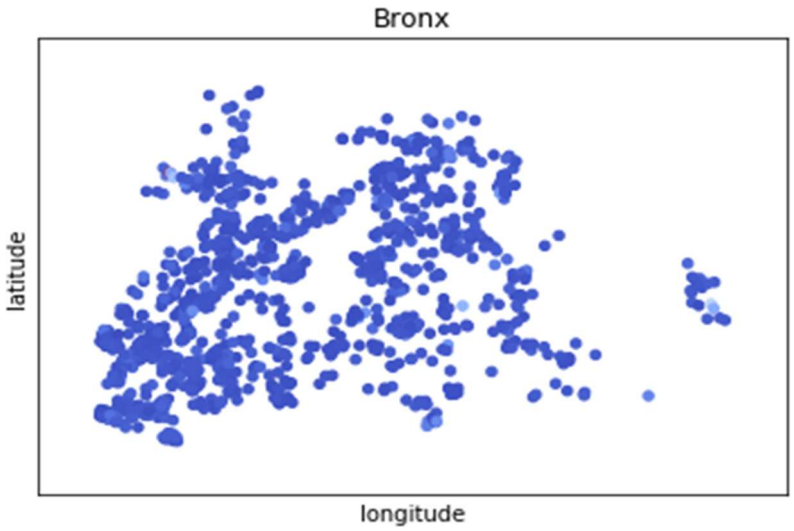
----------2019 Staten Island Data----------

Total Dollars from Rental Properties: $42,825.

Total Nights Units were Stayed in: 1,802.

Total Number of Reviews: 11,541.

Total Reviews per Month: 587.



Bronx data shows us that their average rental unit per night is $87. The most expensive rental property per night is $2,500.

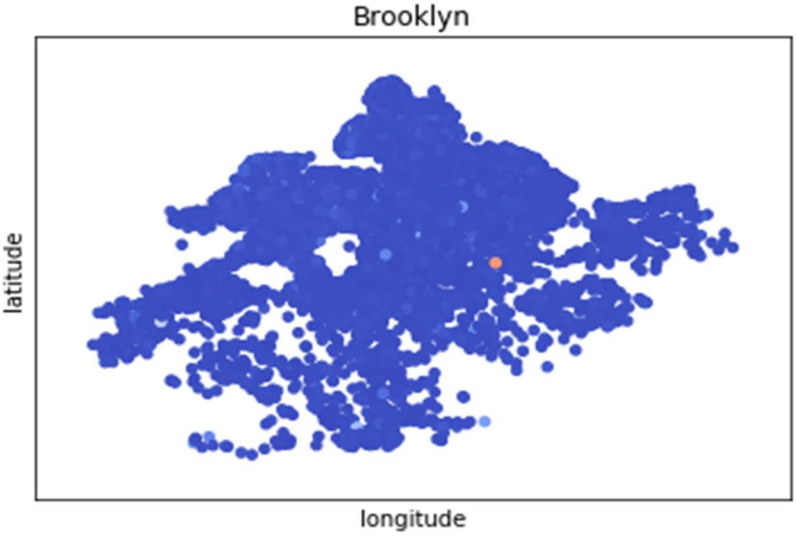
----------2019 Bronx Data----------

Total Dollars from Rental Properties: $95,459.

Total Nights Units were Stayed in: 4,976.

Total Number of Reviews: 28,371.

Total Reviews per Month: 1,609.



Brooklyn's average price per night is $124. Brooklyn's highest priced rental unit is $10,000.

----------2019 Brooklyn Data----------

Total Dollars from Rental Properties: $2,500,600.

Total Nights Units were Stayed in: 121,761.

Total Number of Reviews: 486,574.

Total Reviews per Month: 21,104.

Results

Below are the questions and the answers that I found during the exploratory analysis of the 2019 Airbnb New York Dataset:

- Which hosts are the busiest and why?
    - This question is answered with the three busiest host.
    - The three hosts I looked into where the top three, Sonder, Blueground, Kara.
        - These hosts are the busiest because they are companies using Airbnb's platform to have people stay at their rental properties.
- How much money do hosts make per year that have more than 50 properties?
    - This question is now how much money the top three hosts make per year with more than 50 rental properties.
        - Sonder: 327 rental properties.
            - Total Dollars from Rental Properties: $82,795.
        - Blueground: 232 rental properties.
            - Total Dollars from Rental Properties: $70,331
        - Kara: 121 rental properties.
            - Total Dollars from Rental Properties: $33,581.
- Are there any relationships between prices based on the neighborhood the rental unit is in?
    - After analyzing the dataset and seeing the prices for the rental units I would say there is not a lot of difference in price based on the neighborhood. There seemed to be a lot of prices that stayed consistent with prices from all areas. There were some places that had higher priced rental properties and that was due to location and size of properties.
- Are the units cheaper than hotel rooms in the area?
    - The average cost of overnight hotel accommodations in New York is $238.
    - The average price of overnight Airbnb accommodations in New York is $153.
        - It is on average $85 cheaper to look for Airbnb accommodations vs hotel accommodations when you plan a visit to the state of New York.
- Are the busiest locations around tourist hotspots or are they buy large corporations?
    - New York State is a tourist destination, and the prices of their accommodations reflect that. The pricing is consistent and standard throughout the state. You can see that there is some price increase when you start getting to more tourist spots in New York. Overall,

the number of rental properties that are in New York state it is hard to determine from my project if the busiest locations are around tourist attractions or corporations.

References

1) Shin, Terence. December 2020. KDnuggets. '14 Data Science projects to improve your skills'. https://www.kdnuggets.com/2020/12/14-data-science-projects-improve-skills.html
2) Inside Airbnb. http://insideairbnb.com/get-the-data.html
3) NYC The Official Guide. 2006-2021 NYC & Company Inc. https://www.nycgo.com/
4) New York State. The Official Website of New York State. https://www.ny.gov/
5) Sonder. 2021 Sonder Holdings Inc. https://www.sonder.com/
6) Blueground. 2021 Blueground. https://www.theblueground.com/
7) Big Endian Data. June 27, 2017. 'How to plot data on maps in Jupyter using Matplotlib, Plotly, and Bokeh'. https://www.bigendiandata.com/2017-06-27-Mapping_in_Jupyter/
8) ArcGIS for Power BI. https://doc.arcgis.com/en/power-bi/design/add-data-to-the-map.htm
9) Rosenberg, Zoe. August 4, 2017. 'New York's most and least affordable neighborhoods.' https://ny.curbed.com/2017/8/4/16099252/new-york-neighborhood-affordability
10) Office of the New York State Comptroller. April 2021. 'The Tourism Industry in New York City'. https://www.osc.state.ny.us/reports/osdc/tourism-industry-new-york-city#:~:text=New%20York%20City%20hosted%2066.6,reduction%20(see%20Figure%201).
11) Travel Budget for New York City. https://www.budgetyourtrip.com/united-states-of-america/new-york-city

Appendix

For this project I also wanted to add an interactive visual where you could see each Airbnb location and hover over the location to see the rental properties price. I reviewed quite a few documents on how to do that and I tried to over lay the coordinates over a map of New York state. I was not able to get it to where I wanted for this project, but I was able to get the coordinates roughly over the map. A completed visual would have given a client the opportunity to look at a map of where they want to stay and see how much prices are in that region. This would also allow for quicker analysis on the business end of companies when they want to review the price and the number of nights their rental unit was stayed in.

In this project I wanted to look into more of the top number of units a host had. I tried a few different techniques and was not able to come up with anything in time to present that I felt comfortable with.

Another thing I found interesting looking through this dataset was the number of minimum priced units at zero dollars. I do not know if these where simply units that were not available at the time or if there was an error in listing prices.

I do wish after going through the data and my questions I would have taken the time to try some machine learning to see if I could utilize that in anyway to see what rental property prices might look like in coming years.

10 questions.

1. How many companies utilize Airbnb's platform to rent properties in cities?
2. How many different cities are currently utilizing Airbnb's platform?
3. What is Airbnb's platform like on a world scale?
4. Does Airbnb charge companies the use their platform a fee?
5. How can we predict future dollars on rental properties based off a previous year's numbers?
6. How many rental properties are second homes?
7. Can you tell the success of an area based off rental property reviews?
8. What is the outlook of rental properties after Covid 19?
9. How will Airbnb and other companies get people attracted to Covid hotspot cities for vacation?
10. How long will it take for rental properties and companies to recover from the Covid 19 pandemic?