

2021

Data Visualisation

Module number: DALT7016

Student Number: 19132761

MSc Course: Data Analytics

Word count: 2209

Table of Contents

1.	Dataset selection and justification.....	3
	Overview	3
	Domain Questions.....	4
2.	Data examination and transformations	4
3.	Analysing the data	7
	Interactive graph.....	7
	Percentage of Listing type vs Location.....	9
	Mean of Price by Property type	10
	Number of Reviews over time by Host	10
	The popularity of Airbnb in New York	12
	Regular Host/Super Host vs Price Over Time.....	13
4.	Shiny web app.....	13
	Home Page.....	13
	Map Page.....	15
	Neighbourhood Page	17
	Compare Page.....	18
5.	Summary	19
	Bibliography	19

1. Dataset selection and justification

Overview

Since 2008, Airbnb has assisted visitors and hosts in travelling in a more special and personalised manner. Thanks to its enthusiastic founder, Brian Chesky, the firm grew from a mattress to a multinational enterprise estimated at more than \$30 billion. The year 2020 was expected to be a watershed moment, when Airbnb would go public and announce the world's most valuable stock. Corona virus occurred, which was tragic, the pandemic devastated the tourism industry.

However, it cannot be denied that Airbnb has redefined Hospitality. The sharing economy and vacation rentals are among the hottest topics that have touched millions of lives worldwide. Airbnb is a worldwide marketplace that allows users to book or to offer lodging and tourism experiences. Airbnb has been instrumental in the hospitality service and currently operates in more than 191 countries. This service is accessible via their website or mobile apps.

This visualisation project aims to analyse and determine popular rental neighbourhoods in New York and to estimate how much it would cost if someone were planning a trip here. The visualisations are designed to explore popular neighbourhoods based on the traveller's preferences if they travelled or transited in New York and its neighbourhoods.

To acquire the most up-to-date data, this project used New York City's dataset from InsideAirbnb, which is an independent platform offering non-commercial datasets from Airbnb. The primary csv dataset is from New York in June 2020 with over 106 different dimensions. It was interesting sorting through the data and finding relevant dimensions

for the visualisations. The data is extracted into a secondary subset containing a number of listings per neighbourhood, which is used exclusively in the visualisations.

Domain Questions

Airbnb is most frequently used by tourists, so we explore the various factors that influence tourists' decisions on where to stay when travelling. The interactive map will use spatial visualisation to explore different variables from dataset and answer questions about price fluctuations and ratings through various locations of New York City.

The storyboard has been categorised into four categories:

- Overview of Airbnb
- Property analysis
- Pricing analysis
- Host analysis

Some of the target questions are:

- What is the average price by accommodation types?
- What kinds of accommodation types are more prevalent in the different New York neighbourhoods?
- Number of Reviews over time
- Host/Super Host vs Price over time

Link for Shiny web app: <https://maido.shinyapps.io/AirbnbNYC/>

2. Data examination and transformations

The dataset comprises three main tables: listings, reviews and neighbourhoods.

The listing data contains an immense amount of information about individual Airbnb listings in New York City. The data was relatively clean, need to be eliminated and made simple adjustments such as transforming value types, trimming and using text facet clustering to ensure it was formatted correctly. Some other filter, merge, combine etc., depend on each graph's purpose, will be included in RMarkdown.

- filter the dimensions
 - remove NA and \$ in price
 - remove the duplicate entry from the map

The rows in price columns containing null values are dropped to preserve all of the information when performing exploratory research. The symbol \$ in price also is removed.

```
# Maps
neighmap <- geojson_read("neighbourhoods.geojson", what = "sp")

# Remove duplicate entry from neighmap
neighmap <- neighmap[which(!duplicated(neighmap$neighbourhood)), ]
```

Breakdown of the different dimensions of listings which we selected for the purpose of domain questions:

- host_id (num): Host's ID
- host_since (date): Host since
- host_name (chr): Host's name
- name (chr): Name of listing
- host_is_superhost (logi): Super Host – true or false
- longitude (num): Longitude of the listing
- latitude (num): Latitude of the listing
- neighbourhood_group_cleansed (chr): Specific borough where the listing is located.
- neighbourhood_cleansed (chr): Specific neighbourhood where the listing is located.
- price (chr): Price in USD of renting the accommodation for one night.
- room_type (chr): Category of the room/accommodation type being listed.
- property_type (chr): Category of the property type being listed.
- review_scores_location (num): Review scores are based on Location.
- number_of_reviews (num): Number of reviews.

The reviews dataset: date (DateTime), comment (textual), listing_id (discrete), reviewer_id (discrete).

The neighbourhoods file is in geojson format, which can be used to produce New York City's interactive maps.

A quick look at the data:

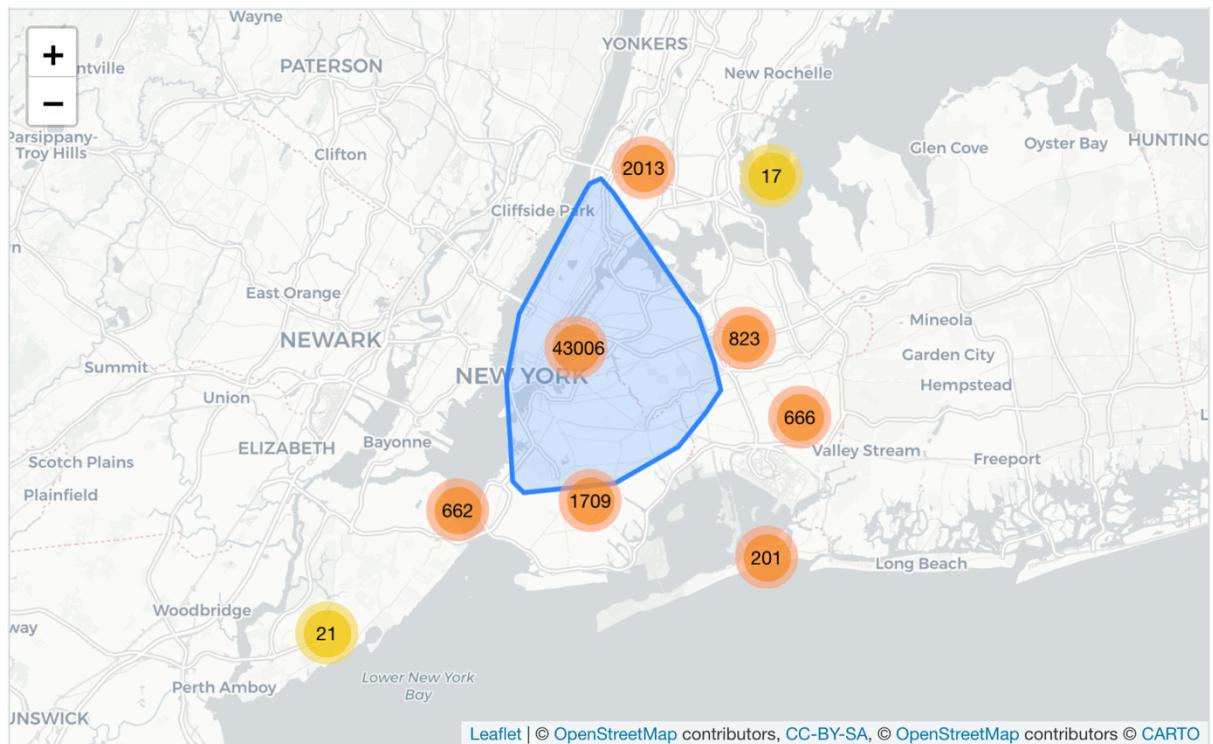
- 49,530 unique listings in New York City in total. The first listing was belonged to a host named Jenny and was created in August 2008 in Washington Heights, Manhattan.
- The listing prices vary from \$10/night to \$999/night, according to the records. The \$999 price tag listings are for Entire home/apt or Hotel room in Manhattan.
- Since then, guests have written over 1 million reviews.

3. Analysing the data

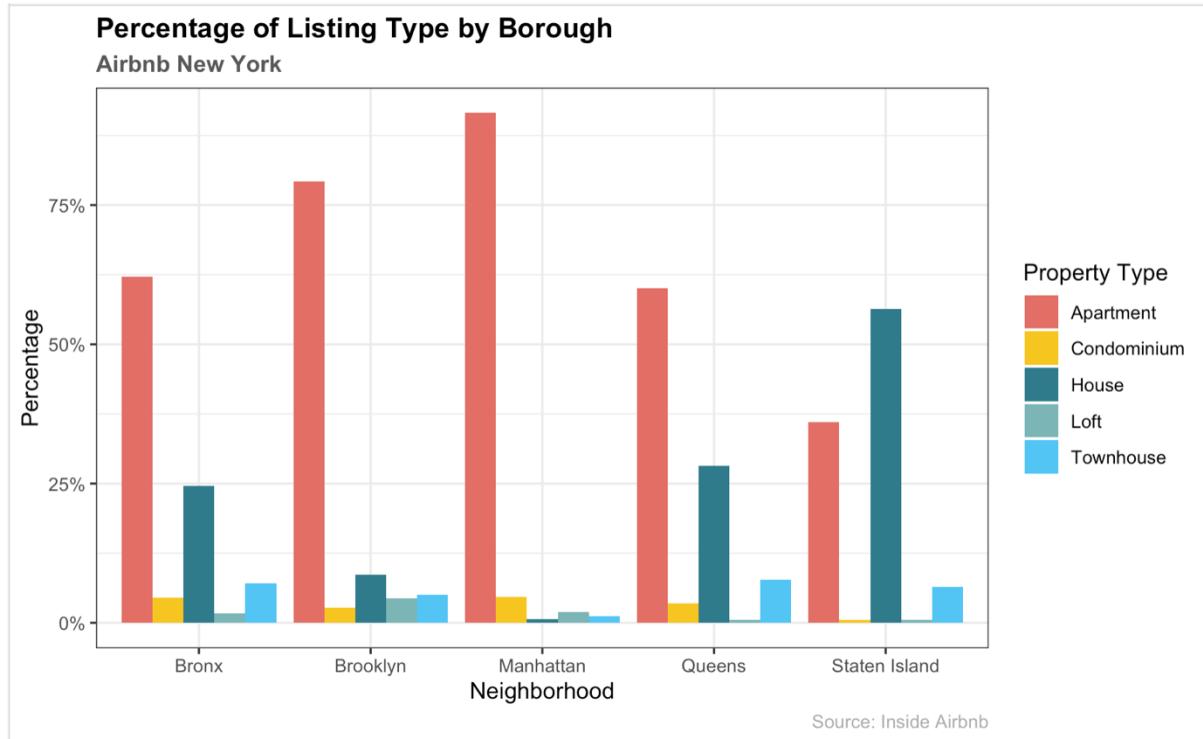
Interactive graph

This is a basic interactive chart with all New York City listings appearing in group style. We can click on clusters to see the listings they include. This provides a zoomed view that contains details about Listing Name, Hostname, Property type, Room type, Price of the property. This visualisation helps to explore every listing geographically. It provides an overall picture of how listings are spread across neighbourhoods. The map shows that the majority of the listings are clustered around Manhattan and Brooklyn, followed by Queens and The Bronx, and the lowest number of listings are in Staten Island.

The location ratings can be a reliable measure of the attractiveness of the neighbourhoods. Highly ranked neighbourhoods are more likely to be well connected (subway stations) or closer to city hotspots (Wall Street, Empire State, Times Square).



Percentage of Listing type vs Location

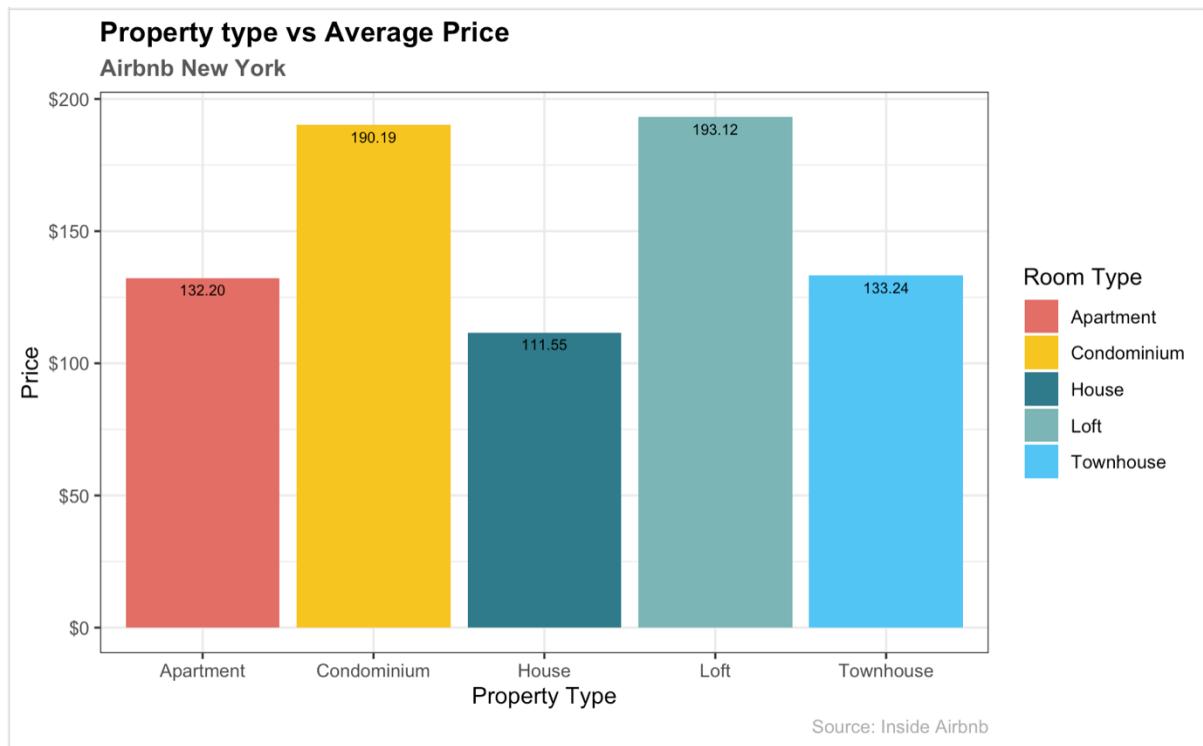


The chart displays the ratio of property type by borough.

- Except for Staten Island, all neighbourhoods have the most apartment listings. Staten Island has a higher proportion of 'Houses' than 'Apartments.' Since Staten Island is not highly populated and has plenty of space, this analysis seems logical. Followed by Condominiums which are also not as popular in Staten Island.
- House listings are the second most common property type, but not in Manhattan. Most properties in Manhattan are apartments (approximately 90%), and houses are the least common. Next is Brooklyn, with 80% apartment listings, followed by Queens with approximately 60% apartment listings.
- Queens and Bronx also have plenty of house listings. Queens has over 25% house listings, which is highest after Staten Island.

- Town House style properties are also common in New York: about 10% of listings in all boroughs except Manhattan.
- Loft style properties are only common in Brooklyn, where they constitute 10% of listings.

Mean of Price by Property type



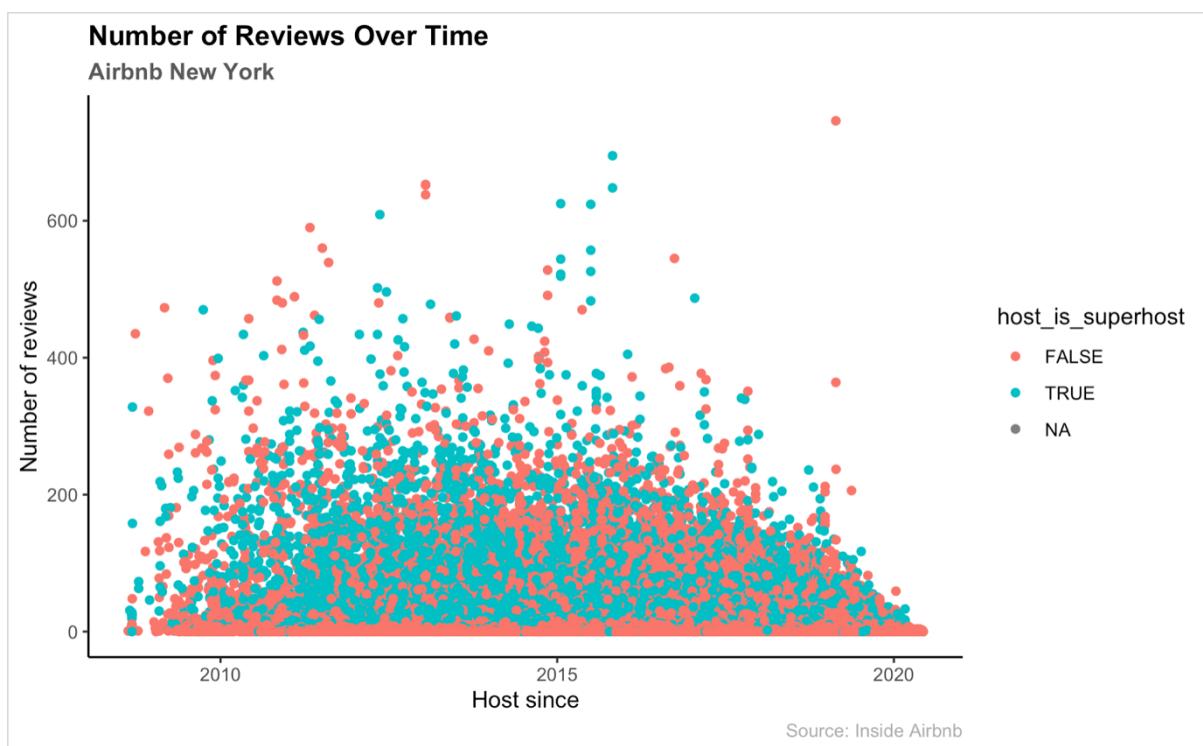
Follow by popularity, Loft and Condominium style listings are less popular and have the highest mean price, which almost double house style listings. Apartments style listings have a reasonable price. However, the prices correlate with Location, and what we see here that Staten Island and Queens are the cheapest, not that houses are cheaper than apartments.

Number of Reviews over time by Host

Airbnb lists its hosts into two categories - Super Hosts and Regular Hosts.

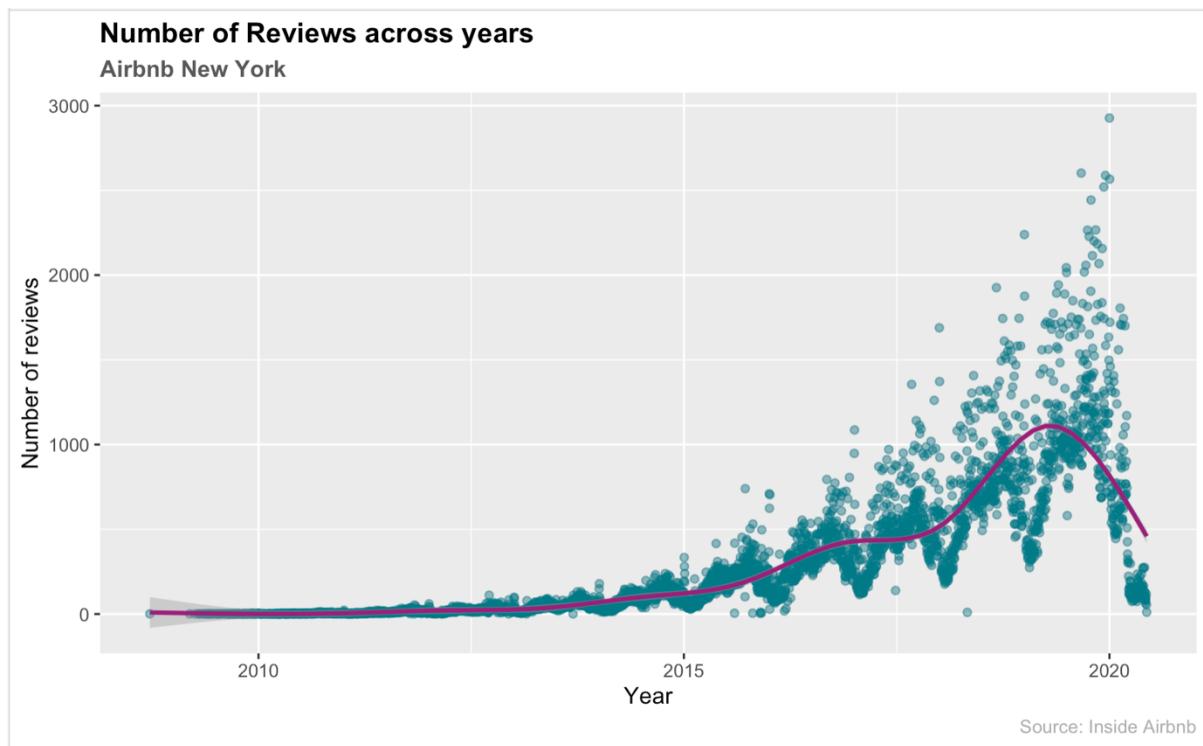
According to Airbnb policies (Airbnb, n.d.), to become a Super Hosts, four criteria should be met by the hosts:

- Provide a minimum of ten stays in a year.
- Maintain 90% response rate
- More than 80% 5-star reviews.
- Zero cancellations (with exceptions)



Here we have a scatterplot showing the number of reviews for hosts compared to when they became hosts. The colours separate hosts from Super Hosts. Hosts who are not Super Hosts (regular hosts) tend not to have many reviews. However, some hosts have many reviews though they are not Super Hosts.

The popularity of Airbnb in New York

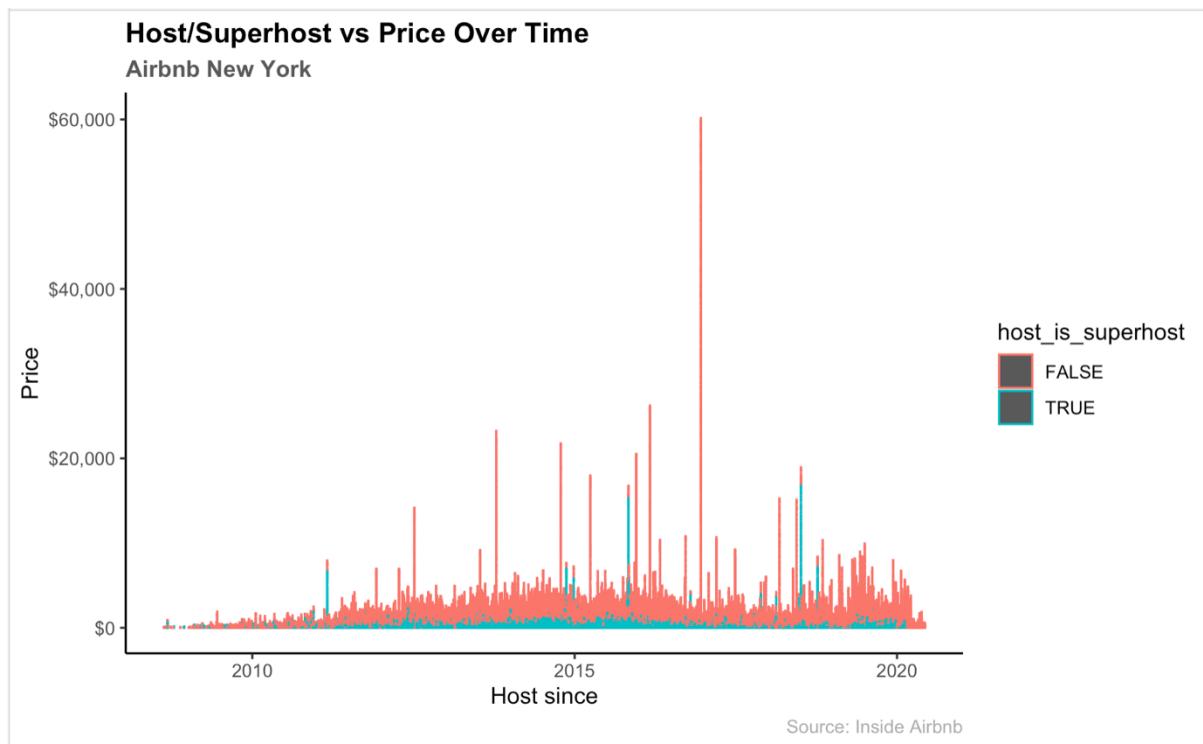


According to Airbnb, about half of all guests leave ratings for listings. As a result, analysing the review's number would give a fair evaluation of demand. The listings that have received ratings has risen exponentially every year, showing a growth in demand.

It is clear that the number of reviews/demands follows a seasonal trend. Every year demand peaks and falls, showing that some months are more guesses than others.

However, from the end of 2019, it started decreasing and became very little in 2020. It can be explained by the annual pattern where there are fewer reviews in the winter and the 2020 pandemic. Travelling is restricted and has a massive effect on tourism.

Regular Host/Super Host vs Price Over Time



As for the graph, it seems as though the people who have changed the most for Airbnb are Regular Hosts. Also, since 2014, hosts tend to set higher prices.

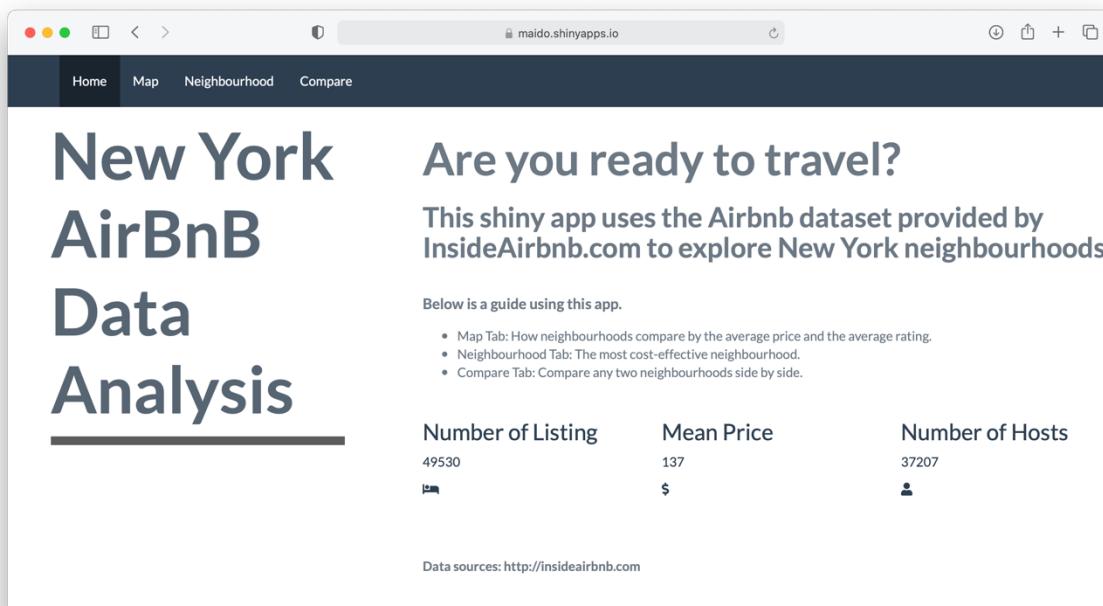
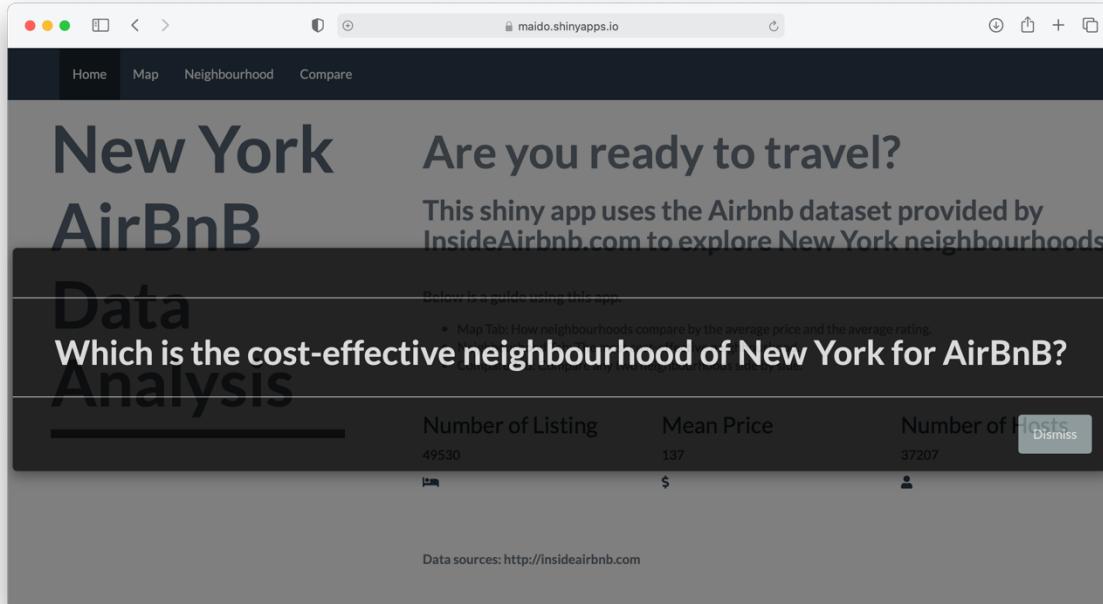
4. Shiny web app

Home Page

This page gives an overview of New York Airbnb and the Shiny web app. There are three tabs are combined:

- Map tab: How neighbourhoods compare in terms of average price and average review.
- Neighbourhood Tab: The most cost-effective neighbourhood.
- Compare Tab: Compare any two Neighbourhoods side by side.

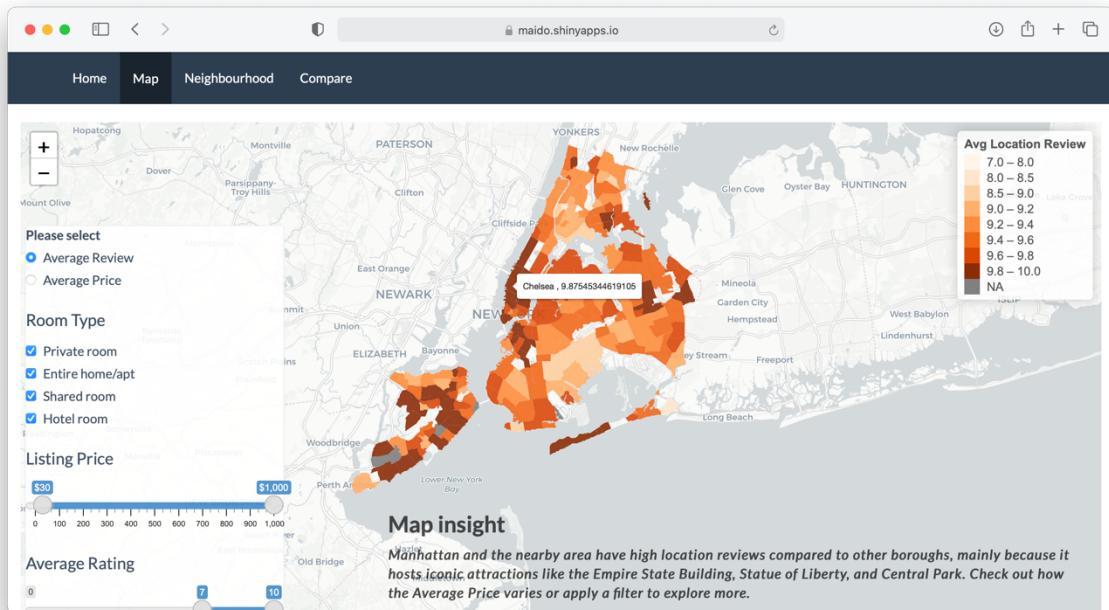
The number of listings until June 2020 in New York is 49.530, the mean price is \$137, and 37.207 unique hosts.



Map Page

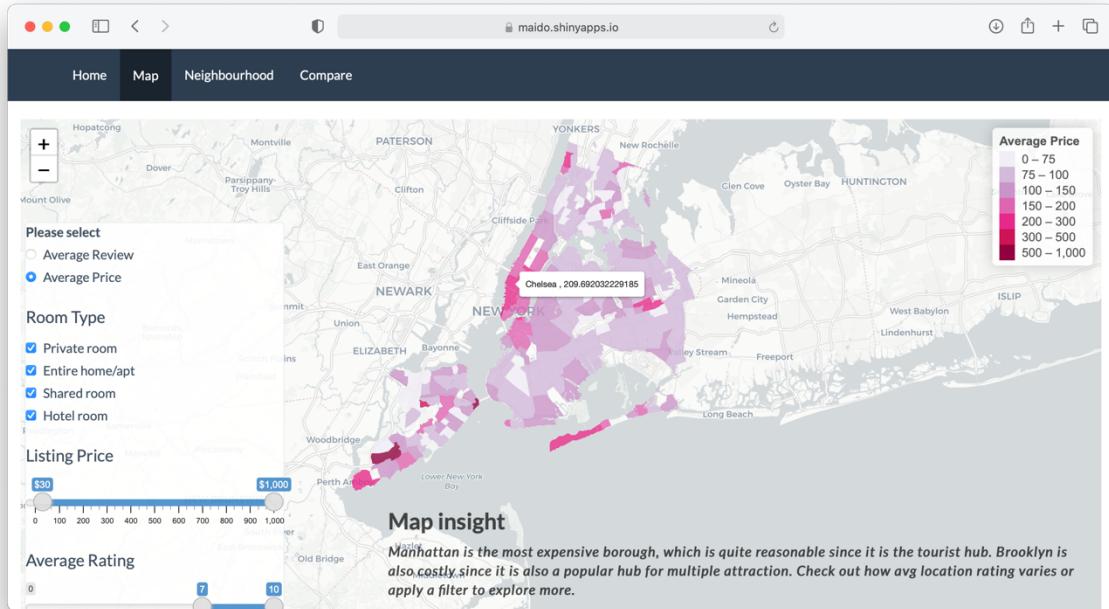
This page contains three panels:

- A set of radio buttons and sliders where we can select information which we want to filter, such as Room Type, Listing Price, Average Rating, Number of Reviews. There are two maps on this page, map by average review or average price.
- Map insight gives a short comment about the areas in New York City-based on Review or Price.
- Legend: Average Review has the average location review rates from under 7.0 to 10.0; some areas have unknown reviews. The average price has ranged from 0 to \$1000 a night.



The graph confirms the interactive graph assessment, Manhattan gets the highest place ratings for the downtown area. The places closest to the Goethals Bridge on Staten Island got the best place ratings. Some areas in Brooklyn where close to Manhattan

also have higher place scores. By observing the Brooklyn's metro lines, it's worth mentioning that the high-rated locations are correlated to metro lines. The same holds true for The Bronx, there are no train tracks.



This map is a continuation of the previous location ranking given by the neighbourhood map. It goes without saying that high-rated places are often likely to be expensive (demand versus supply).

Again, downtown Manhattan is the clear winner for high rents, which is also true for the Brooklyn neighbourhood near Manhattan.

The East Village area in Downtown Manhattan is an obvious exception, where both the Rental and Location ratings tend to be lower than the surrounding areas. It would be interesting to conduct two studies:

- Find areas with high ratings and low rents: The Mid Island on Staten Island (as mentioned in the preceding chart) is one of neighbourhood where prices are

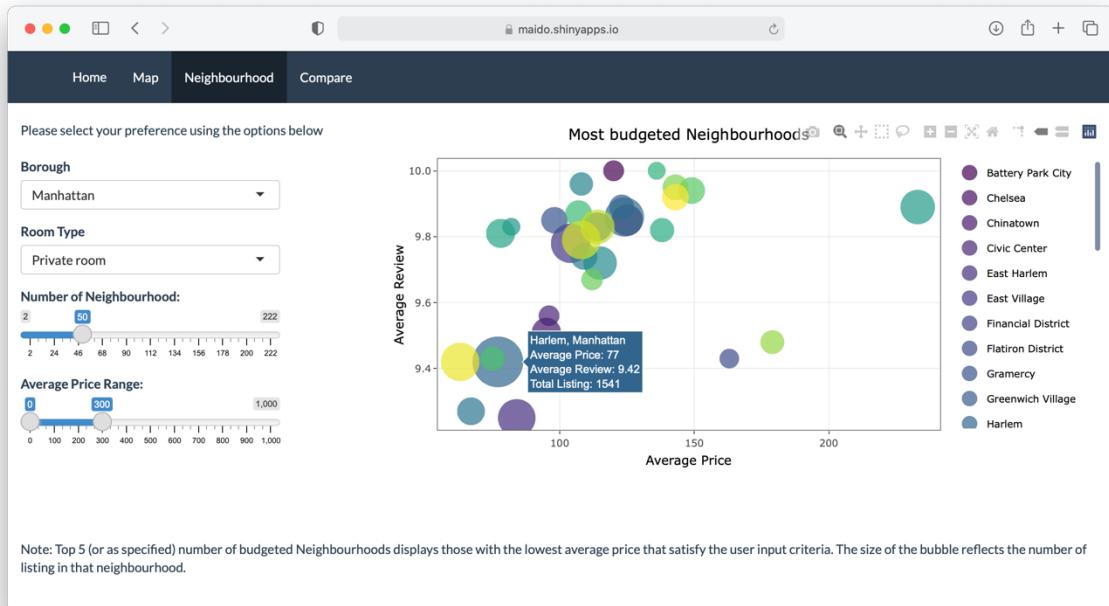
relatively low even though getting a high location score. Another great location is in Brooklyn near New York Bay.

- Look for low-rated areas with high prices: The Castle Hill neighbourhood on The Bronx has an extremely high price and in an inconvenient location.

Neighbourhood Page

This page contains a panel and a bubble graph:

- The radio button selects preference such as Borough, Room Type, Number of Neighbourhood and Average Price Range.
- Bubble Plot: The top number of budgeted Neighbourhoods displays those with the lowest average price that satisfy the user input criteria. The number of listings in that neighbourhood is reflected by the size of the bubble.



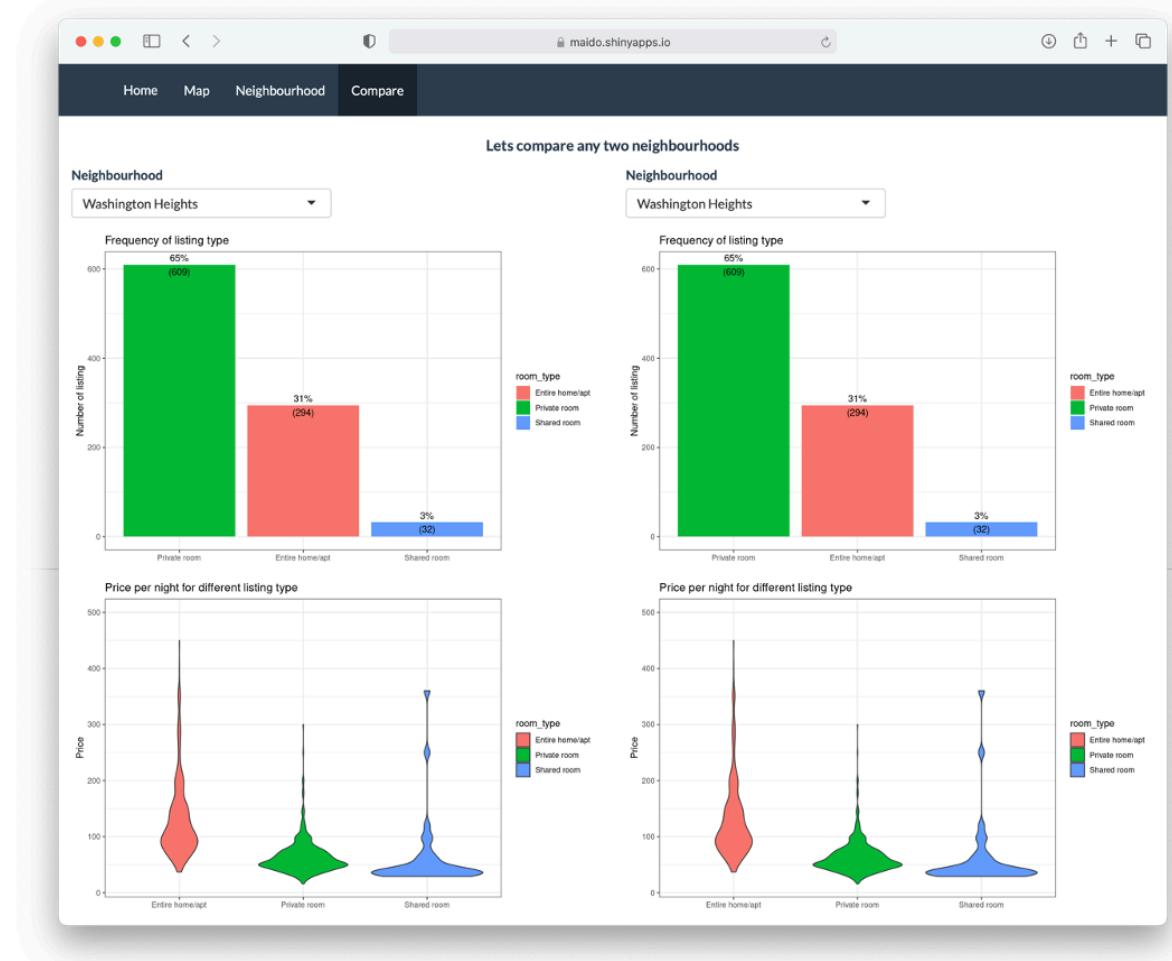
The bubble plot shows an example of private rooms in Harlem, Manhattan, where the average price is \$77/night, the average review is 9.42, and the number of listings is

1541. Harlem has the biggest bubble, indicating the neighbourhood has the highest number of listings for private rooms in Manhattan.

Compare Page

This page contains four graphs about the frequency of listing type and price per night for different listing type in two preferred neighbourhoods.

In this example, Washington Heights has no hotel room listings, and 65% of listings are for private rooms. However, in Murray Hill 79% of listings are for an entire home/apt and also has some hotel room listings. Most of the Washington Heights listings are lower than \$100/night whilst most in Murray Hill are higher than \$100/night.



5. Summary

The Airbnb dataset is a brilliant resource for further understanding New York's bustling rental environment. New York has proved to be one of Airbnb's fastest-growing markets, with 50,000 listings recorded in the last 12 years. The years between 2015 and 2020 were the Airbnb-boomed, when more and more people adopted Airbnb as a service provider; the number of reviews and listings are doubled every year. Downtown Manhattan and the surrounding areas of Brooklyn have by far the highest concentration of listings. Staten Island and the Bronx have fewer people who want to stay.

Bibliography

Airbnb. (n.d.). *How do I become a Superhost?* Retrieved from Airbnb:

https://www.airbnb.co.uk/help/article/829/how-do-i-become-a-superhost?_set_beve_on_new_domain=1618324242_Y2NhZmVhYjNjOTBi

Kirsten Grind, J. E. (2020, April 8). *Airbnb's Coronavirus Crisis: Burning Cash, Angry Hosts and an Uncertain Future*. Retrieved from The Wall Street

Journal.: <https://www.wsj.com/articles/airbnbs-coronavirus-crisis-burning-cash-angry-hosts-and-an-uncertain-future-11586365860>

Tuong, L. (2020, Jun 25). *Airbnb-Part A (Python-Visualization, Comparative Study, Regression)*. Retrieved from <https://towardsdatascience.com/airbnb-part-a-python-visualization-comparative-study-regression-7466c0cb5a1d>