Blog Post: Airbnb and Housing in Hong Kong

Julia Dragu

2023-11-30

Is the Airbnb Vacation Rental Market a Factor In Hong Kong's Housing Crisis?

As the world population continues to increase, housing crises are rising all over the world. The city with by far the highest degree of housing shortages is Hong Kong, China. However, Hong Kong is also the city with the highest level of tourism. As such, it becomes an interesting setting for an exploration of short-term rental prices.

Airbnb is a short-term rental company which specializes in vacation rentals. Open data can be found with detailed listings data from most major cities across the world, including Hong Kong.

For this exploration, the following will be covered:

1. General data exploration to better understand the factors at play, via data visualizations

2. Trends in Hong Kong neighbourhoods and the listing price distribution across the city

3. Trends in the affluance of hosts *This includes exploring whether hosts are foreigners or not, which is an interesting point of conversation when it comes to vacation rentals in areas with housing crises. A well-known example in the United States is Hawaii.

4. Predictors of listing price *This is an additional machine learning exploration

The purpose is to determine potential trends in the vacation rental market in cities such as Hong Kong in which its own residents struggle to find housing.

We will begin by loading the data, which was downloaded from Inside Airbnb, which is an initiative intented to make Airbnb data open and transparent to the public, including customers and hosts alike. The data was last updated on 17 September, 2023. As listed on the Inside Airbnb website, "The data behind the Inside Airbnb site is sourced from publicly available information from the Airbnb site. The data has been analyzed, cleansed and aggregated to faciliate public discussion." That said, the data will still require further cleaning, which can be followed below.

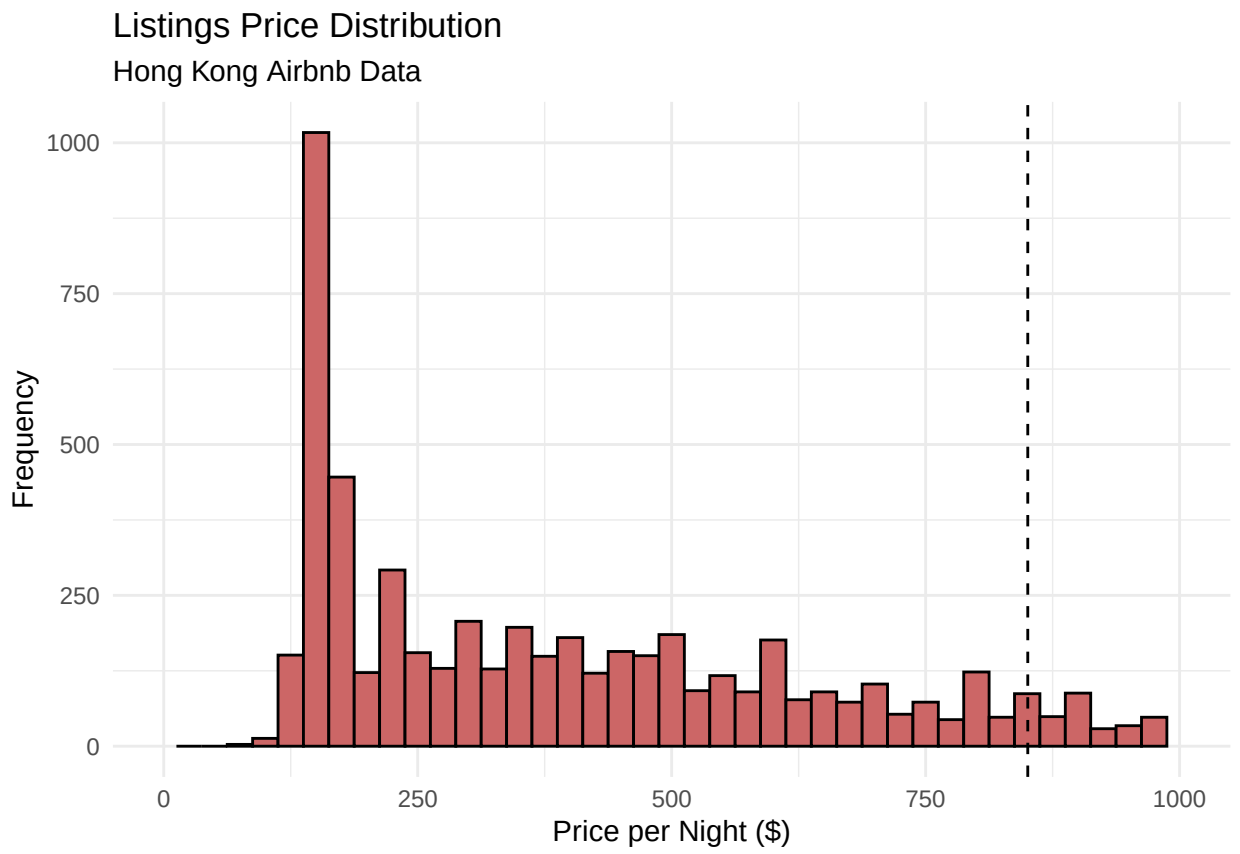
0. Data Set-Up

Although the data has been pre-cleaned, it requires further cleaning for the purpose of this study. To begin with, there are many variables; some of which, such as date of last data scrape, are unnecessary to the task at hand. As such, these variables have been omitted from the dataset. In addition, once the graphs have been plotted, categorical variables were then encoded as dummy variables.

1. Data Exploration - Visualizations

Price Distribution

Let us first explore the price distribution. This will give us a good estimate of the vacation rental market in Hong Kong. For reference, as found via InterNations, the average monthly rent for permanent residents of Hong Kong in the year 2023 is between 1,500 to 2,500 USD. This comes out at an average of approximately 67 USD per night.

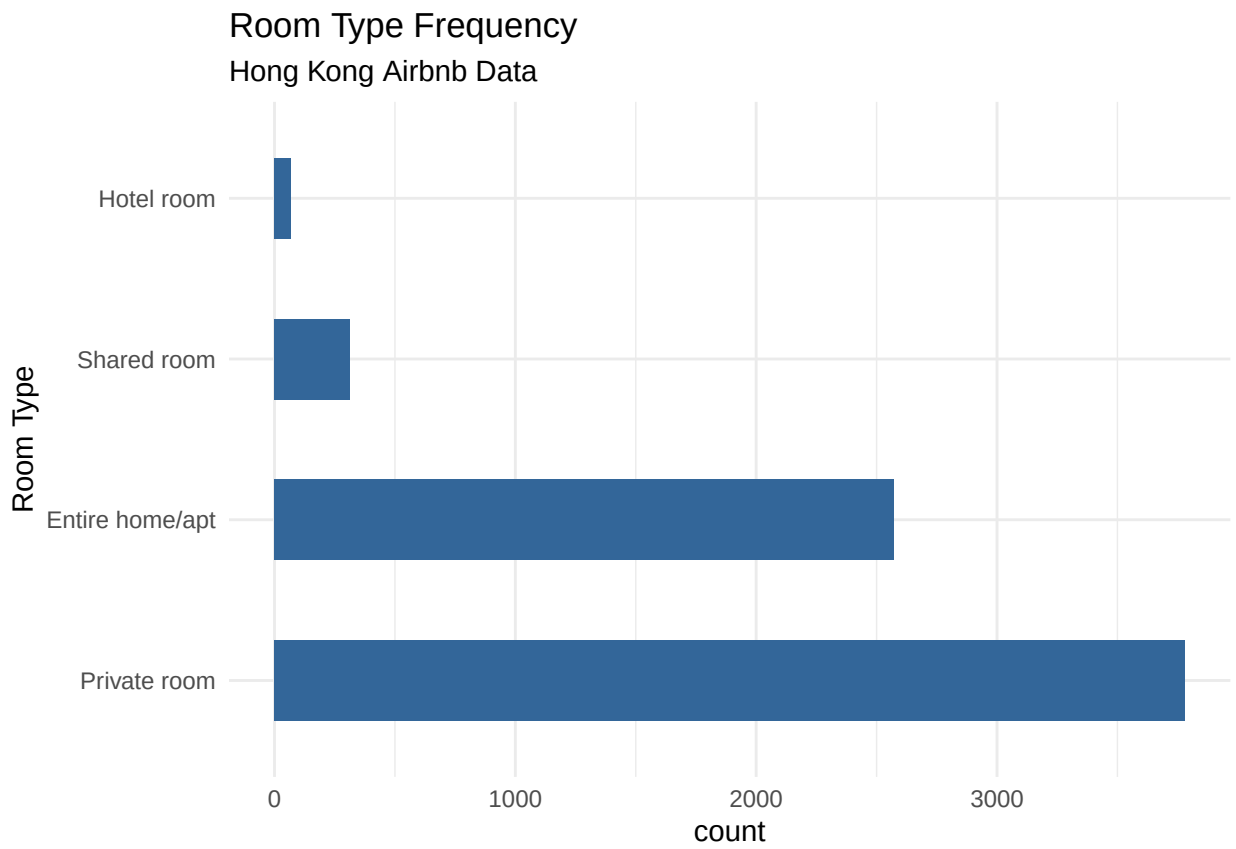



```
## [1] 850.5464
```

We can see that, in fact, the average listing price per night is approximately $850. This is despite the fact that the distribution seems to be skewed to the left, in which most prices are less than $500, and many are listed at approximately $100 per night. As such, we can assume that the data is skewed by outliers. Still, the average nightly price far exceeds the average rent price for a permanent resident of Hong Kong. This suggests that high demand from tourists of wealthier countries may play a role in the housing crisis.

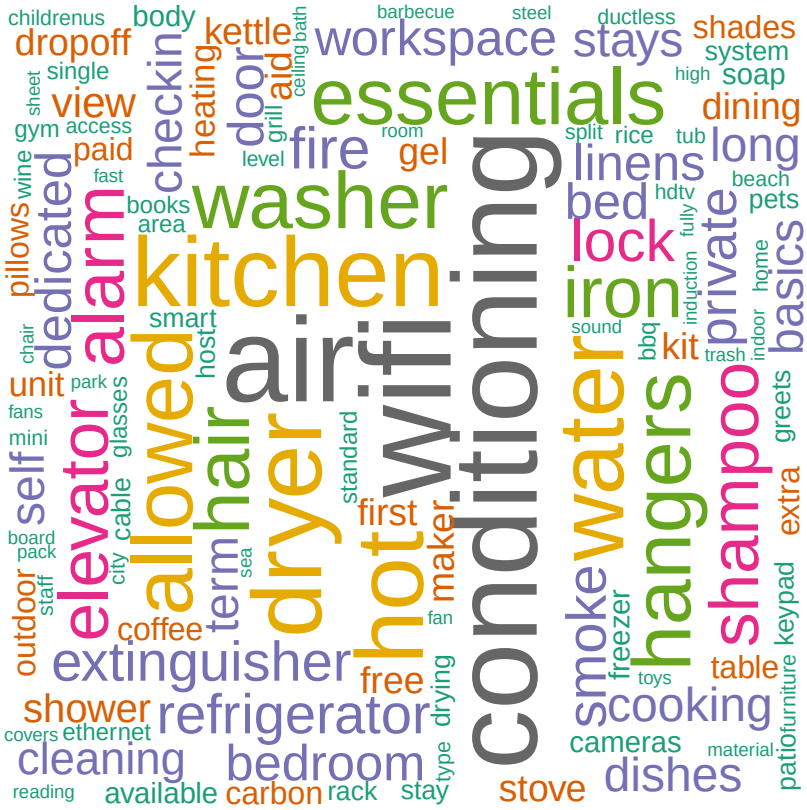
Listing Property Types

Now let us explore the types of properties listed. We will use the room_type variable instead of the property_type variable for the sake of simplicity, as there are far fewer levels to the variable in comparison. This will be done using a simple vertical bar chart with frequency of each room type in the listing data.



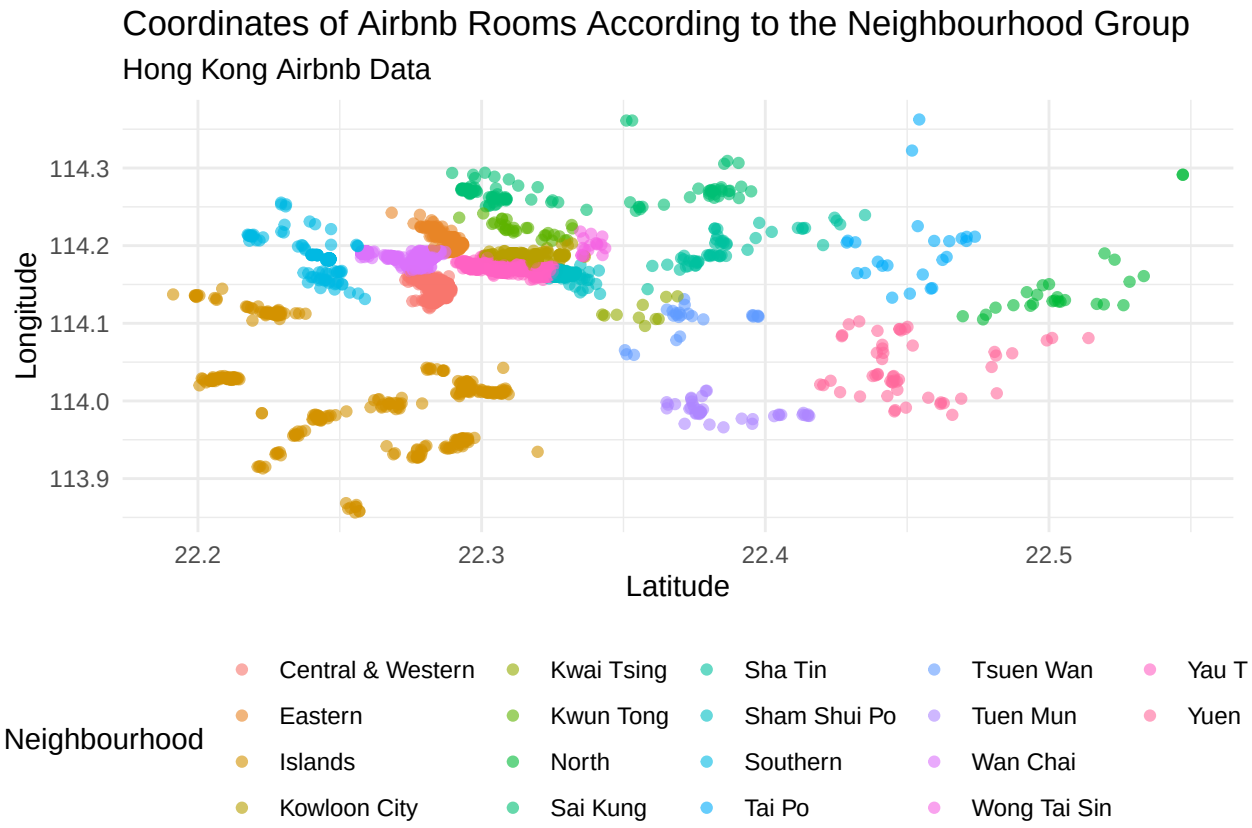
Our findings suggest that the majority of listings can be categorized as private rooms, with the next most frequent category being entire homes or apartments.

```
## Loading required package: RColorBrewer

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##     annotate
```



2. Trends in Neighbourhoods

Now let us explore the distribution of listings across latitude and longitude coordinates, grouped by neighbourhood. We will do this using a scatterplot in a manner similar to a heatmap, only in this case it is color coordinated by neighborhood and distributed by latitude and longitute.



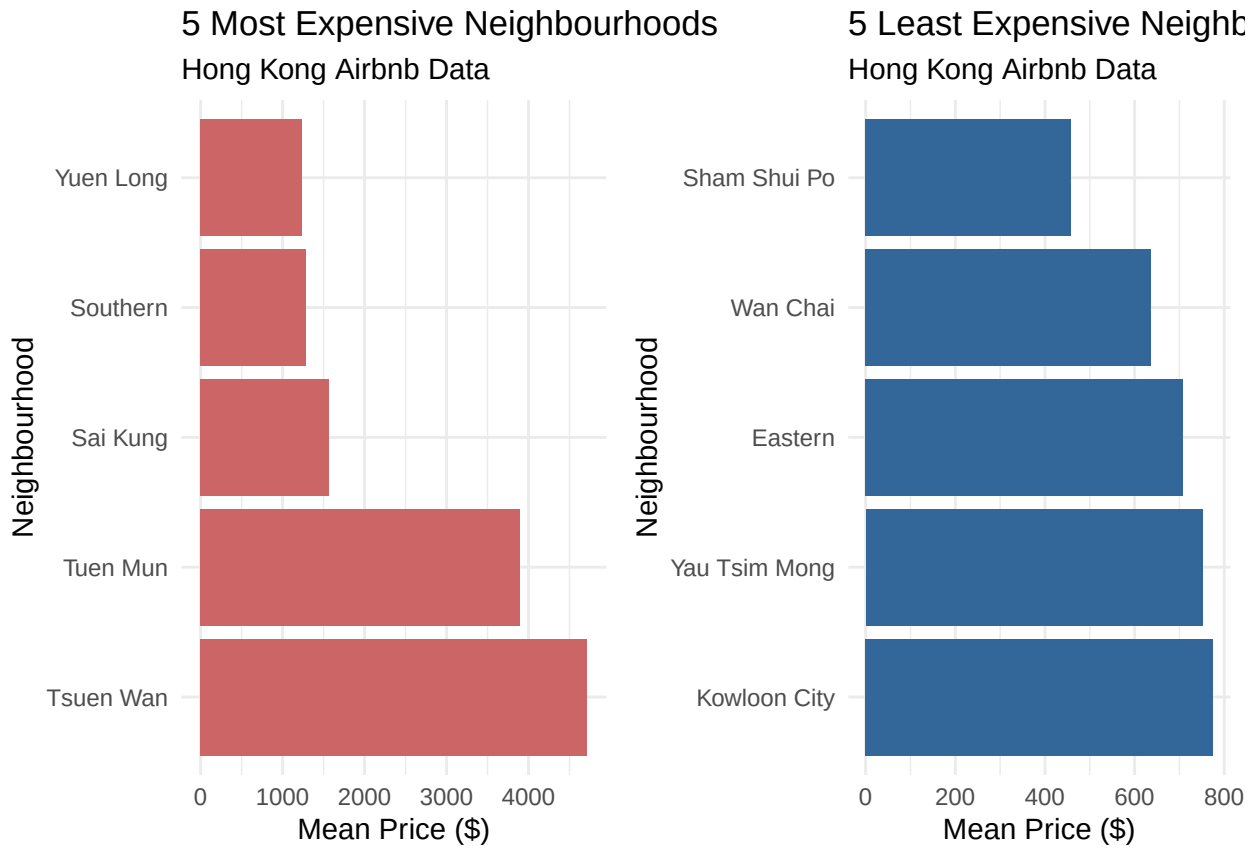
Based on the graph above we find that the most listings can be found clustered at the following neighbourhoods: Islands, Eastern, Central and Western, Wan Chai, Wong Tai Sin, and Kwai Tsing, among a few others. That being said, this heatmap-style graph is only meant to be used as a preliminary exploration of the neighbourhood clusters. Let us now explore the neighbourhoods at a greater depth by looking at average price in the most and least expensive neighbourhoods.

First beginning with the most expensive neighbourhoods:

We see that Islands, Kwai Tsing, Wong Tai Sin, and Southern–all with high clusters of listings–are each among the most expensive neighbourhoods.

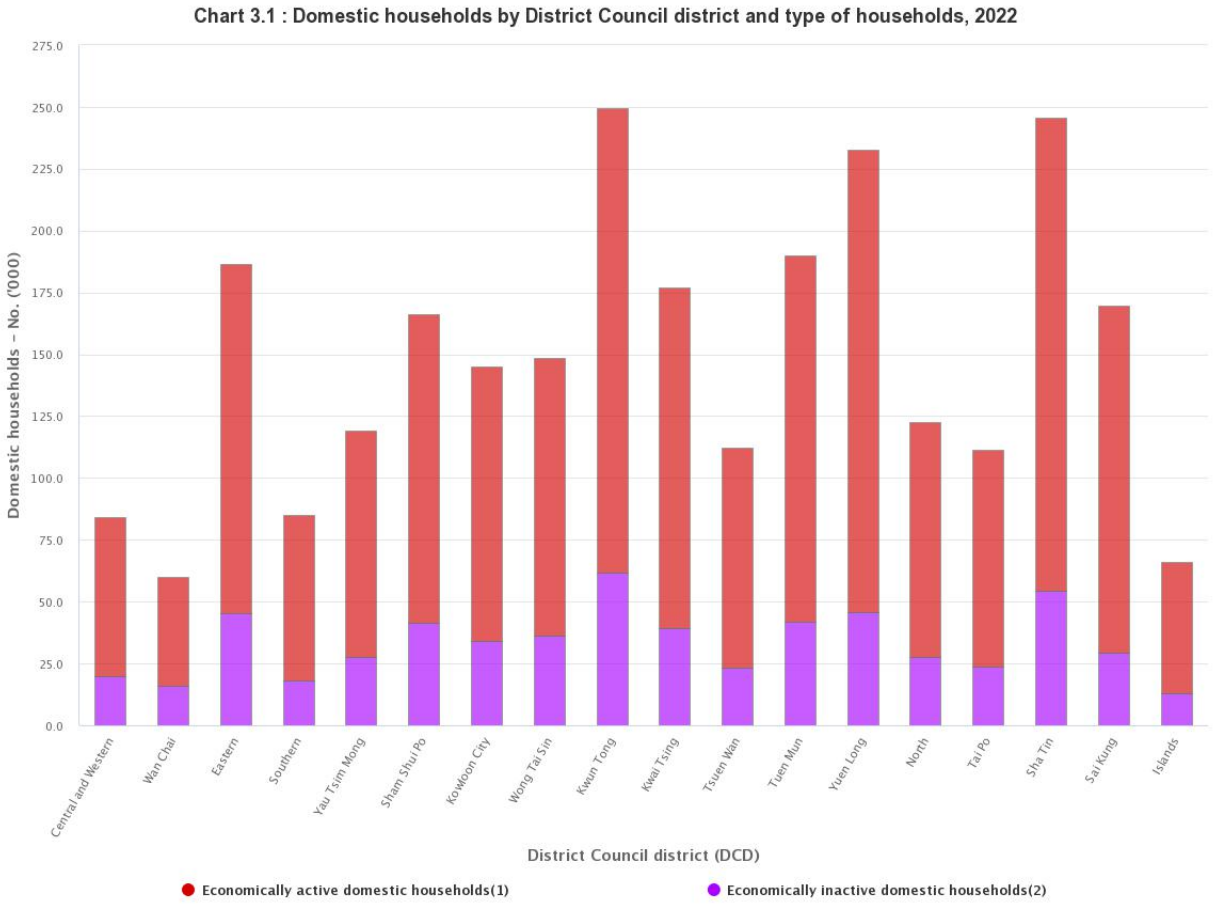
Let's explore the neighbourhoods by average price once more, now looking at the lease expensive neighbourhoods.

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```



Here we see some of the other neighbourhoods with high clusters of listings: Wan Chai, Eastern, and Kwai Tsing.

Now let us explore the real-word implications of these findings. Below, a chart can be found from the Census and Statistics Department detailing average number of households per neighbourhood.



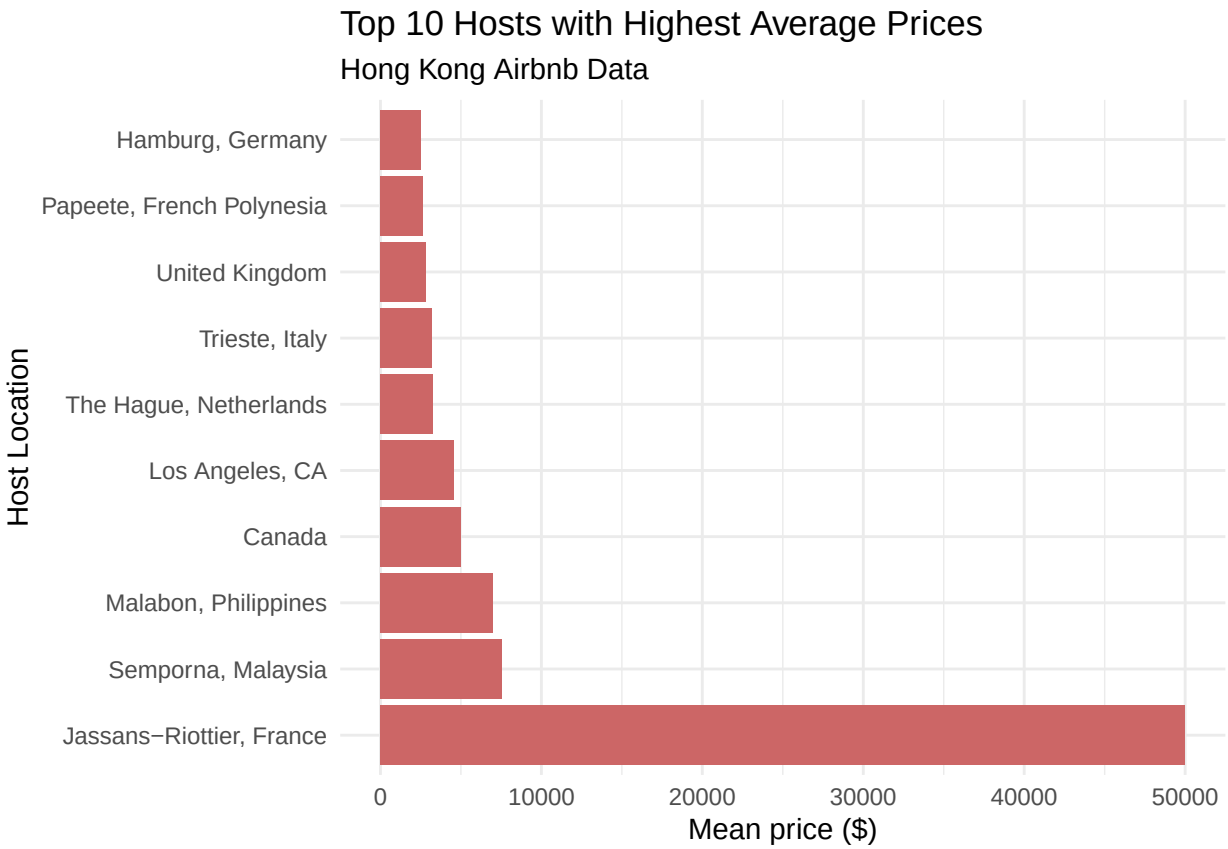
1 Economically active household is a domestic household with at least one member (excluding foreign domestic helpers) being economically active.

2 Economically inactive household is a domestic household with all members (excluding foreign domestic helpers) being economically inactive (e.g. home-makers, retired persons not available for work or not seeking work and those below the age of 15, etc).

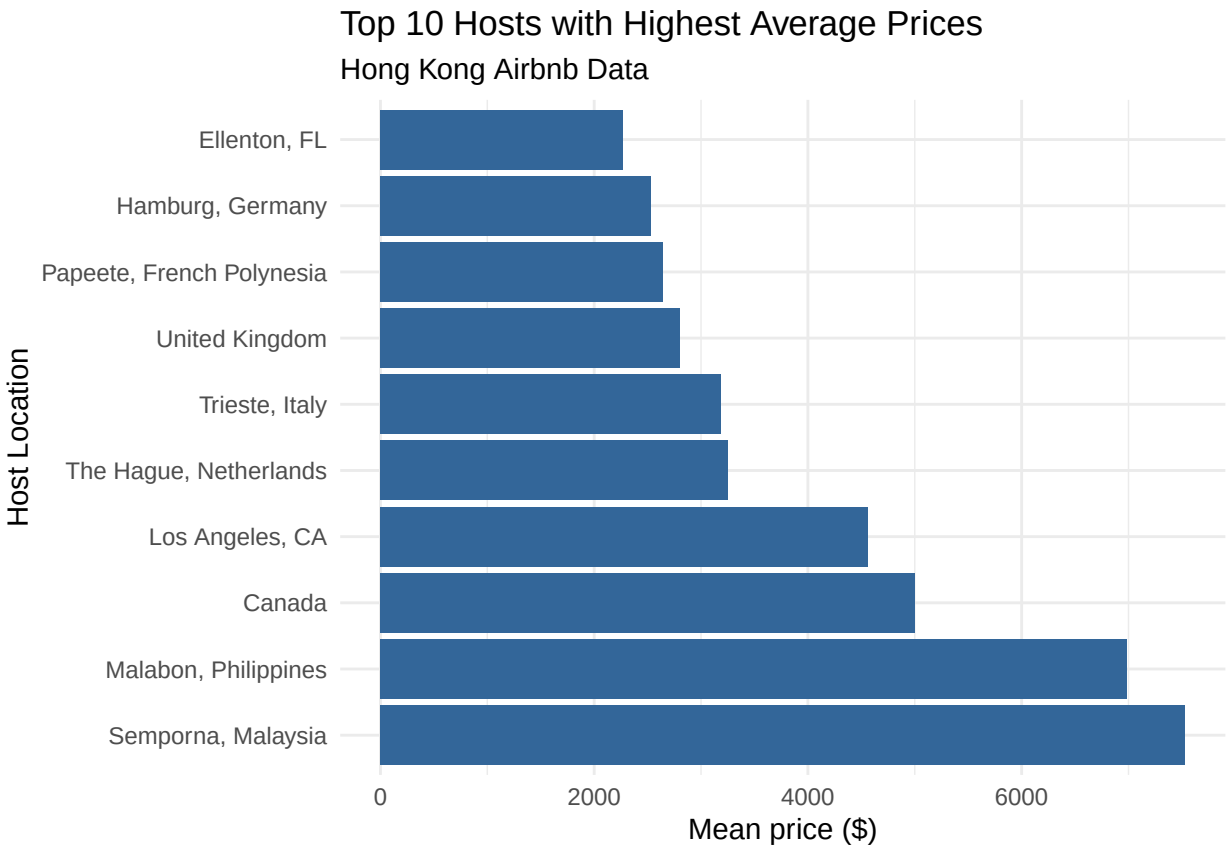
From the above chart, we see that Eastern and Kwai Tsing are two of the most heavily-populated neighbourhoods. Recall that we found they are both two of the cheapest neighbourhoods for Airbnb vacation rentals. This is a phenomenon that would be interesting to explore further in regards to the effects of the vacation rental market on housing crises across the world.

3. Trends in the affluance of hosts

Now we know which neighbourhoods have more expensive listings. However, the locations in which the hosts are from could also be a confounding variable in terms of listing price. Perhaps, for example, foreigners are more likely to be hosts than others. As such, let us explore the distribution of locations from which hosts come from. We will do this using a bar chart detailing the relationship between average listing prices and a host's country of origin.



Indeed, our guess was correct; the most expensive listings are owned by people foreign to Hong Kong, with the addition of one particularly wealthy host from Jassans-Riottier, France. Let us look at the data again with the top 10 hosts, omitting the host from Jassans-Riottier.



Now let us begin examining the relationships between data. We will start with a basic regression exploring the relationship between the listing price and the number of listings a host has in total. This is based on the reasonable assumption that, the more money a host has, it is likely that they own more vacation rental properties.

price

host_listings_count

-2.0963***

(0.2166)

N

6,735

R2

0.0137

Notes:

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

In fact, we find that there is a negative relationship between listing price and the number of listings a host has! It seems that, in reality, the more vacation rental properties a host has, the cheaper each listing will be.

4. Predictors of listing price

As a data science student, I wanted to explore a more complex scenario, in which we identify some of the predictor variables of listing price. This way, we can also explore confounding variables and reflect on what data may be missing. To do this in a manner similar to concepts covered in Econometrics, we will use a multi-linear regression as opposed to other Machine Learning Algorithms.

All the variables used in the regression were found using backwards elimination, which is a method in which we begin by using every variable in the regression and–one by one–eliminating the least significant variable until all are significant.

price

room_type_entire

552.9185***

(90.6002)

host_total_listings_count

-0.9505***

(0.2232)

accommodates

197.2758***

(21.2648)

maximum_nights

-0.3597***

(0.0873)

availability_30

15.1325***

(3.5715)

neighbourhood_cleansed_Islands

-500.4484***

(178.7812)

neighbourhood_cleansed_Tuen_Mun

2,400.0470***

(486.5980)

neighbourhood_cleansed_Wan_Chai

-235.8713**

(100.0169)

neighbourhood_cleansed_Yau_Tsim_Mong

-224.4585**

(91.7275)

N

6,735

R2

0.0473

Notes:

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

We find that the most significant predictors of price in Hong Kong are whether the vacation rental is an entire house or apartment, the total number of listings a host has, how many guests the listing can accomodate, the maximum number of nights offered, how many days the rental is available over the next 30 days, and whether the rental is found in the following neighbourhoods: Islands, Tuen Mun, Wan Chai, or Yau Tsim Mong.

If we recall, Islands and Tuen Mun are some of the most expensive neighbourhoods, and Yau Tsim Mong and Wan Chai are some of the cheapest. This is certainly an interesting point that could be further explored in additional analyses, as well as something can can be replicated in cities with similar situations of high levels of tourism with concurrent housing crises. Such cities include Sydney, Australia and New York, USA, as examples.

5. Conclusion

Perhaps the greatest concern raised in this exploration was the fact that the most expensive rentals are overwhelmingly owned by foreigners to Hong Kong. An additional step which would help solidify concerns would be to explore profits. After all, a listing can be very expensive, but perhaps few customers are ever renting it out. As such, a listing that is much cheaper by comparison may be a "market leader" in a given neighbourhood.

Indeed, the additional variable of profits would be extremely useful for additional analysis. Yet, in the given data, we have found that neighbourhoods and the general affluence of the hosts are important variables for consideration. Vacation rental type, which is covered by the variable room_type, as well as size, covered by accommodates (meaning how many guests a rental can accommodate) are two more obvious variables of importance, but this exploration supported the belief.

With additional data and time, it is worth exploring datasets covering income levels and general census data about the long-term and permanent residents of Hong Kong as a point of comparison.