TOKYO AIRBNB DATA ANALYSIS

Luis Alejandro Vargas Ramos

BUSINESS PROBLEM

Rental platforms like Airbnb in Japan are hoping for a boost from a new law that will take effect next month before the expected surge in demand for the 2020 Tokyo Olympics, experts warn that it could actually hamper businesses from short term. Currently, anyone who rents a room runs the risk of "breaking the law", short-term rentals will be legalized on June 15, clearing a legal gray area. The new law also introduces new restrictions, discouraging many who rent rooms to tourists through Airbnb or similar platforms. People who rent their houses or apartments will have to register their accommodation with the local authorities, the new law limits total overnight stays to 180 days per year. The new legislation also allows local authorities to impose their own restrictions. The city of Kyoto, for example, has said it will only allow rentals in residential areas between mid-January and mid-March, the lowest season of the year. Jake Wilczynski, Airbnb spokesman for Asia-Pacific, told AFP that the new laws are a "clear sign that Japan is welcoming the idea of short-term rentals for individuals." Many hosts have canceled reservations or simply removed their accommodations from the platform.

THE OPPORTUNITY

Despite the new restrictions, there is a large potential market for short-term rentals as the country prepares for Tokyo 2020 and the Rugby World Cup in 2019. Airbnb rents have skyrocketed in recent years, driven by a surge in tourism and a surprising lack of hotel infrastructure. With around 60,000 houses and apartments, Airbnb dominates the vacation rental market in Japan, even though it lags far behind many countries in Europe - France, for example, has 450,000 listings. Demand is poised to rise, as Japan is targeting an influx of 40 million visitors in 2020 when it hosts the Summer Olympics, up from 29 million last year.

THE DATA

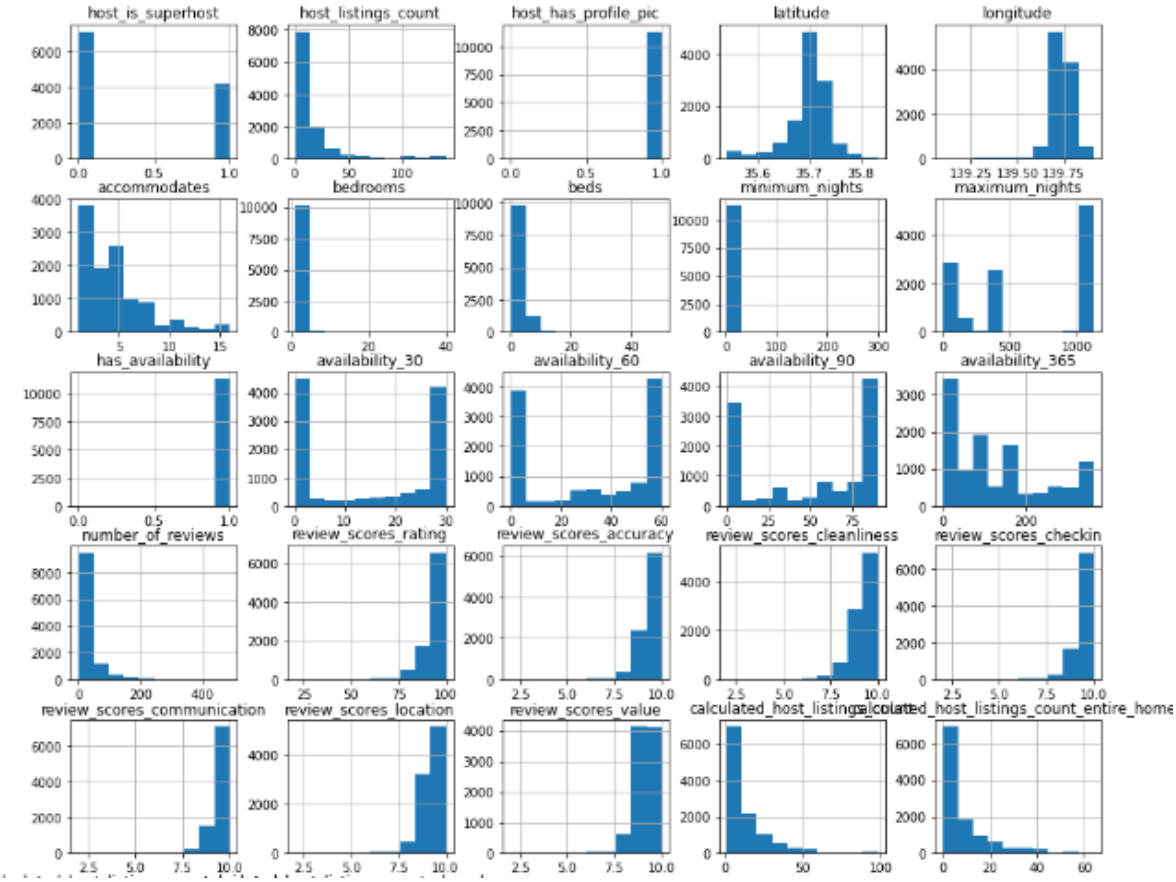
Airbnb does not release any data on the listings in its marketplace, a but separate group named Inside Airbnb scrapes and compiles publicly available information about many cities Airbnb's listings from the Airbnb web-site. For this project, their data set scraped on December 29, 2020, on the city of Tokyo, Japan, is used. It contains information on all Airbnb listings that were live on the site on that date (almost 10,000). Here's the database:

http://insideairbnb.com/tokyo/

http://insideairbnb.com/get-the-data.html

DESCRIPTION OF THE VARIABLES

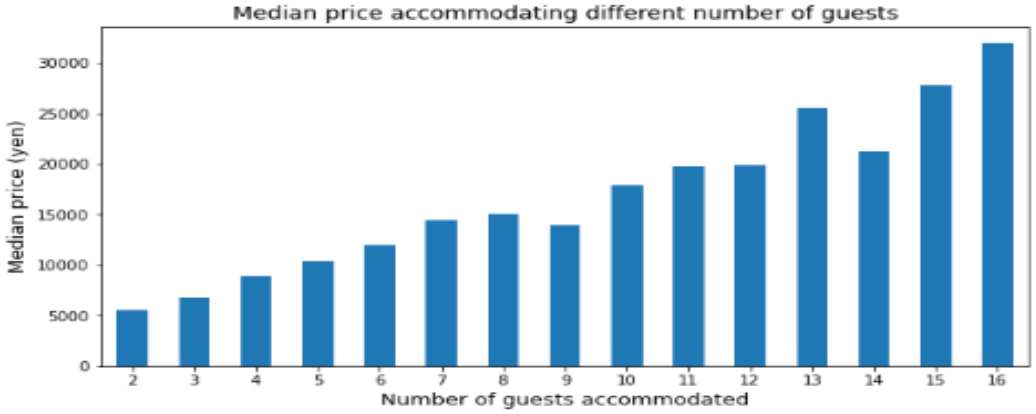
- host_response_rate: proportion of messages that the host replies to.
- host_acceptance_rate: proportion of customers that the host accept.
- host_is_superhost: whether or not the host is a superhost.
- host_listings_count: how many listings the host has in total.
- host_identity_verified: whether or not the host has been verified with id.
- neighbourhood_cleansed: the neighbourhood the property is in
- property_type: type of property, e.g. house or flat.
- room_type: type of listing, e.g. entire home, private room or shared room.
- accommodates: how many people the property accommodates.
- bathrooms_text: type of bathroom.
- beds: number of beds.
- amenities: list of amenities.
- price: nightly advertised price (the target variable).
- maximum_nights: the maximum length of stay.
- availability_30: how many nights are available to be booked in the next 30 days.
- availability_60: how many nights are available to be booked in the next 60 days.
- availability_90: how many nights are available to be booked in the next 90 days.
- availability_365: how many nights are available to be booked in the next 365 days.
- number_of_reviews: the number of reviews left for the property.
- review_scores_rating: guests can score properties overall from 1 to 5 stars.
- review_scores_accuracy: guests can score the accuracy of a property's description from 1 to 5 stars.
- review_scores_cleanliness: guests can score a property's cleanliness from 1 to 5 stars.
- review_scores_checkin: guests can score their check-in from 1 to 5 stars.
- review_scores_communication: guests can score a host's communication from 1 to 5 stars.
- review_scores_location: guests can score a property's location from 1 to 5 stars.
- review_scores_value: guests can score a booking's value for money from 1 to 5 stars.
- license: type of license of the property
- instant_bookable: whether or not the property can be instant booked (i.e. booked straight away, without having to message the host).
- calculated_host_listings_count: Count of total listing of the host
- calculated_host_listings_count_entire_homes: Count of total homes listings of the host
- calculated_host_listings_count_private_rooms: Count of total private rooms listings of the host
- calculated_host_listings_count_shared_rooms: Count of total shared rooms listings of the host
- reviews_per_month: calculated field of the average number of reviews left by guest each month

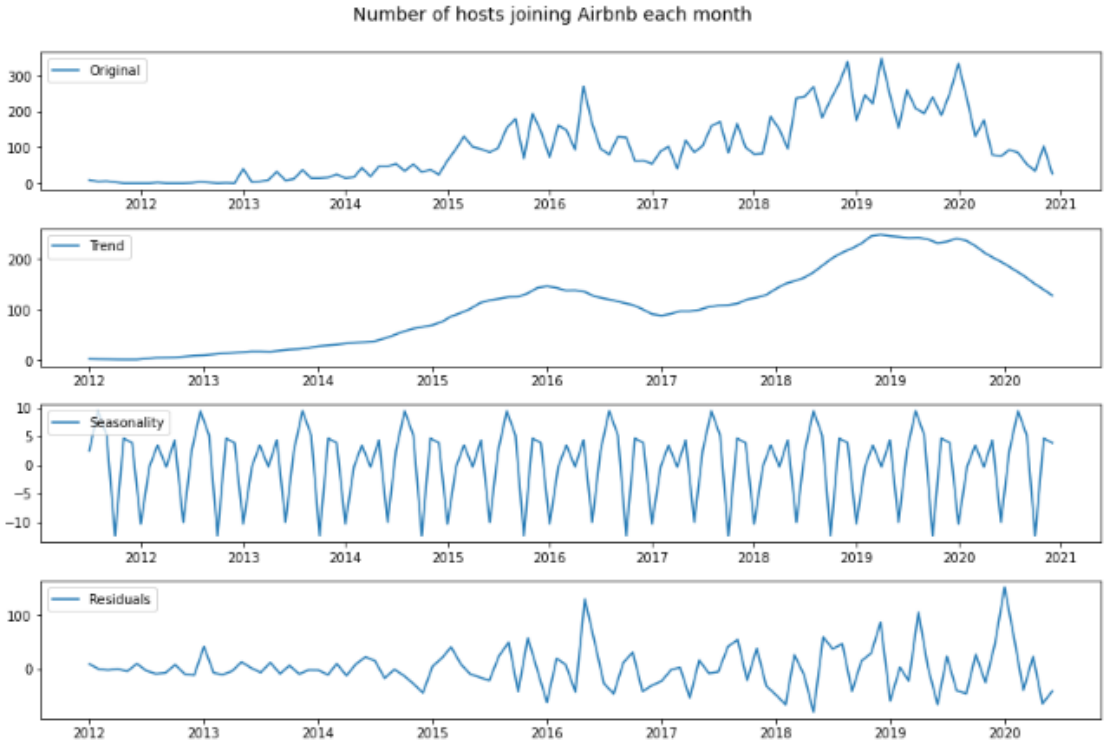


EXPLORATORY DATA ANALYSIS

Time is an important factor to consider in a model when we would like to predict prices or trends. A time series is a series of data points ordered in time. In a time series, time is often the independent variable and the goal is usually to make a forecast for the future. There are also other questions that come to play when dealing with time series. For example: Is there any seasonality to the price? Is it stationary? Even though we may not be able to include this aspect into our model, it is good to explore it to be aware of it and be able to make recommendations for future research. Thus, in this section, we will explore this aspect of the data.

For Airbnb prices, a high level of seasonality is expected due to the characteristics of the market. In October, room rental price rises considerably. It is an extremely popular event and much of the rented property available will have been taken up due to the number of people who attend each year. Tokyo has lots of other events throughout the year, mainly focused between March and December.

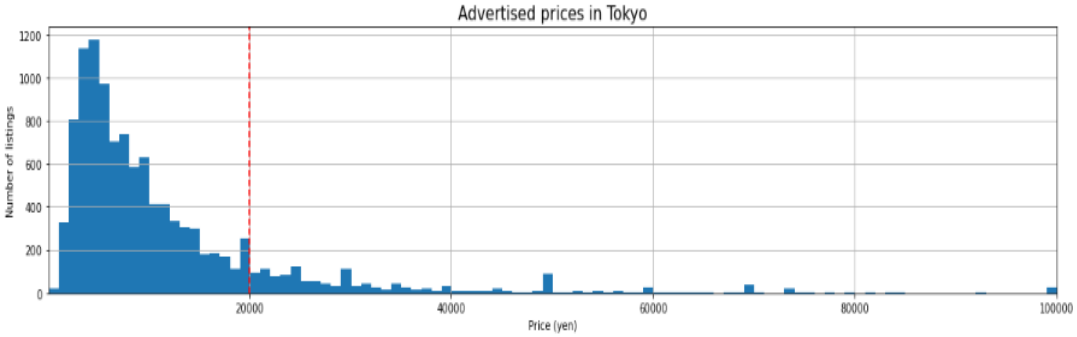




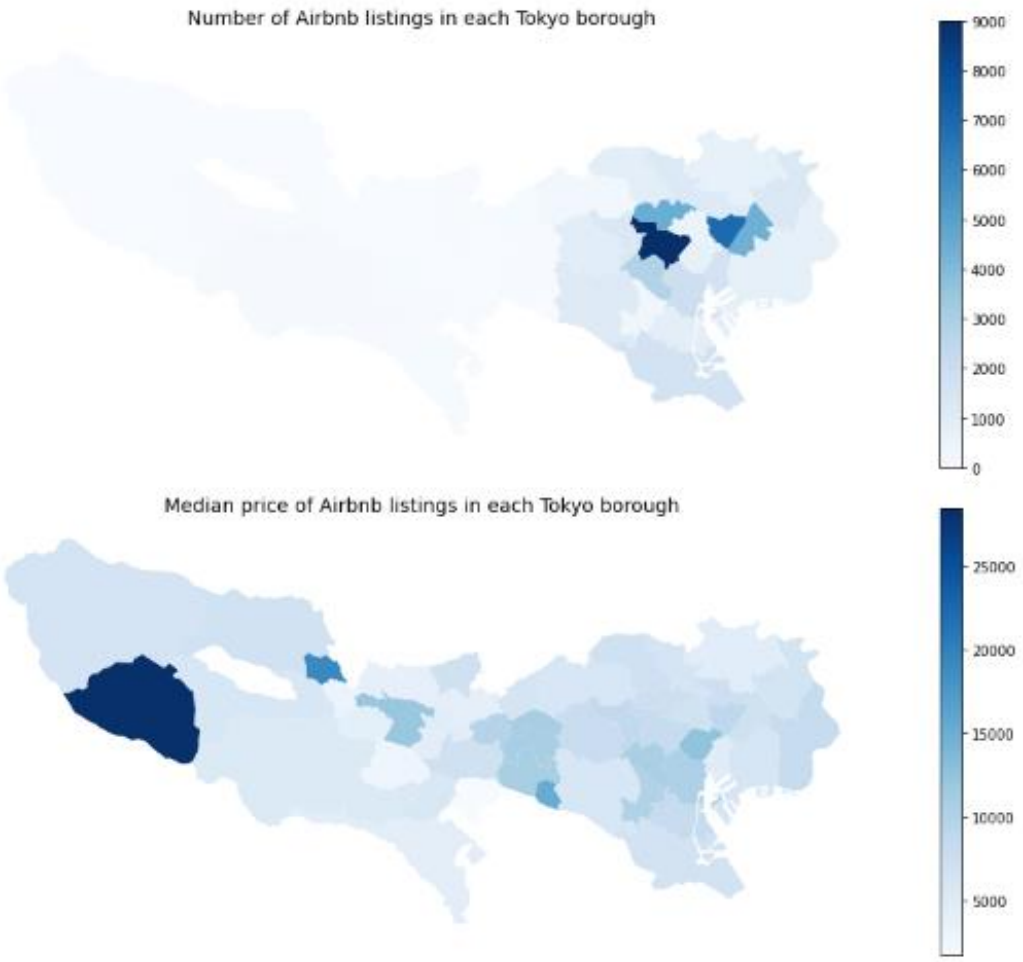
We can see above that there is a clear seasonality. Every year, you see a peak towards hosts joining around the middle of the second half of the year (summer/fall), and the lowest points are the beginning and the end of each year. There is a big peak in the number of hosts

joining Airbnb between 2014-2015 and 2015-2016. Indeed, there has been a fast growth of Airbnb since middle 2014. This was around the time when Airbnb became increasingly popular for short-term lease, as a way to get around local legislation and taxation.

Another important pattern to observe is the number of listings per host. There are a number of professional Airbnb management companies which host a large number of listings under a single host profile. However, there is no consistent upwards trend in the average number of properties managed by each host.



NEIGHBOURHOOD



MULTY COLLINEARITY