

THE IMPLEMENTATION OF BIG DATA ANALYSIS IN REGULATING ONLINE SHORT-TERM RENTAL BUSINESS: A CASE OF AIRBNB IN BEIJING

J. Li *, F. Biljecki

Urban Analytics Lab, National University of Singapore, Singapore — (li_jialin00@u.nus.edu, filip@nus.edu.sg)

Commission VI, WG VI/4

KEY WORDS: Airbnb, real estate, Beijing, GIS

ABSTRACT:

With the fast expansion and controversial impacts of short-term rental platforms such as Airbnb, many cities have called for regulating this new business model. This research aims to establish an approach to understand the impact of Airbnb (and similar services) through big data analysis and provide insights potentially useful for its regulation. The paper reveals how Airbnb is influencing Beijing's neighbourhood housing prices through machine learning and GIS. Machine learning models are developed to analyse the relationship between Airbnb activities in a neighbourhood and prevailing housing prices. The model of the best fit is then used to analyse the neighbourhood price sensitivity in view of increasing Airbnb activities. The results show that the sensitivity is variable: there are neighbourhoods that are likely to be more price sensitive to Airbnb activities, but also neighbourhoods that are likely to be price robust. Finally, the paper gives policy recommendations for regulating short-term rental businesses based on neighbourhood's price sensitivity.

1. INTRODUCTION

Airbnb has sparked wide discussion on whether it brought gentrification to neighbourhoods and disrupted local community by driving out residents. This paper tries to investigate whether such disruptive impacts exist in Beijing, and if so, how could the city respond through regulations. Because of the city's relatively high housing prices (an average price per square meter of RMB60 588; around S\$12 088/ €7873/ US\$8766 according to the currency exchange rate at the time of the submission of this paper) and a high percentage (37%) of long-term renters (people.cn, 2018), this discussion may be more relevant in Beijing than in other Chinese cities. Short term rental industry is an emerging sector enjoying policy support in China. Unlike many cities around the world, there is no regulation against the short-term rental practice. Since Airbnb opened its division in China in 2016, the platform emerged as the first-tier online homestay applications in the country, with its highest market share in Shanghai, Beijing, and Chengdu (Airbnb, 2017).

The central two districts in Beijing — West and East — are the historic districts, together with four near-ring-road districts (Hai Dian, Chao Yang, Feng Tai, Shi Jing Shan; see Figure 1).

These districts make up the core districts, hosting functions including national governments, finance district, headquarter office, hi-tech industry park, scattered with residential land use. The surrounding districts are satellite towns, with yellow coloured districts locating in Ecological Conservation Area. Only the core districts were inside Beijing when the country gained independence in 1949, other districts were slowly assimilated into Beijing as the core districts are fully built up and expanded outwards. To this date, job opportunities, higher-ranking education institutions and medical facilities to a large extent still cluster in the core districts. A larger percentage of tourism attractions such as the Forbidden City, and historic palaces are also located in the core districts.

*Corresponding author

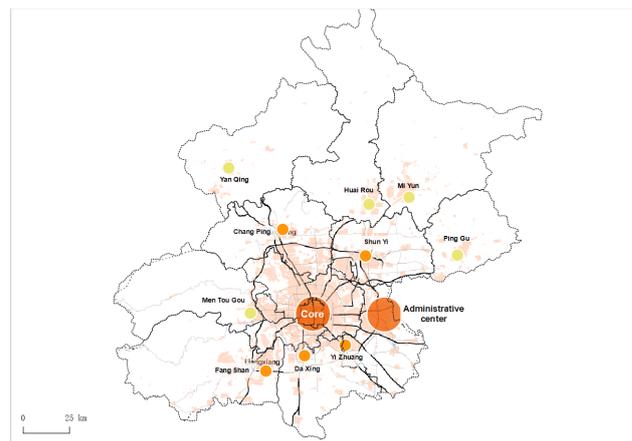


Figure 1. The concept plan of Beijing.

This research tries to draw a relationship between short-term home rental business and the impact on neighbourhoods. Thus we will focus on districts that are (1) more built up; and (2) have been part of Beijing for a longer time. The orange coloured districts and Men Tou Gou (coloured in yellow) will be the focus of this research (Figure 1).

2. RELATED WORK AND POLICIES

2.1 Background and regulations

Regulations differ across countries. In recent years, with an increasing number of local governments around the world implementing data sharing policy on Airbnb, the huge data pools are readily available for analyses. Cities could better understand the short-term rental process taking place within their boundaries and reveal the hidden dimensions of Airbnb activities (Bloomberg, 2018). In response to the controversial impacts the service had,

many local governments reached agreements with Airbnb to impose regulation policies directly on Airbnb platform, such as the maximum day cap, e.g. in London and France (Airbnb, 2019b,a). Taxes are also collected by Airbnb on hosts' behalf and remitted to relevant local tax authorities (Airbnb, 2019c,d). This form of cooperative regulation whereby governments and Airbnb negotiate rules and implement regulations together represents a new era in regulating digital economy and it is encouraging for the domain of research and policy making.

Airbnb is also accelerating its transformation towards working with local governments in terms of regulations. Its self-regulation finally extends to technical, market-based regulation toolkits such as a model split to gain landlord's approval to put up Airbnb listing, in return for a part of the rental income. Airbnb has also published home-sharing design advices for developers to minimising the potential for conflict between visitors and long-term residents. The advices include separate entrances for long term residents and tourists, carpeted lobbies to absorb the sound of wheeled suitcases, and exhibiting local art for an authentic 'sense of place'. In addition, it is interesting to note that there third parties that involved in the regulation process. There are start-up consultancies detecting short term rental violation by analysing Airbnb listing data (Kelly, 2016).

2.2 Factors in understanding Airbnb's impact and prevalence

This research outlines a few matrices that are commonly used in the discussion of Airbnb (Airbnb, 2019e; Gurran and Phibbs, 2017).

Availability How many days of a year the listing could be rented out, where high availability may suggest the listing being dedicated to short-term rental use instead of residential use.

Percentage of entire home rentals in a city's listings Airbnb allows a variety of different rentals, e.g. renting out a shared room to an entire house. Entire home rentals may suggest that long-term residents are displaced for short term rental use.

Case of multi-listings Owners with multi-listings may likely be running a business in residential buildings. For example, out of the 8568 active listings in Beijing:

- the entire home listings account for 59.1% of total listings;
- 86.7% of listings are considered highly available on Inside Airbnb (available for more than 60 days per year); and
- 76.4% of listings have owners with multiple listings.

2.3 The economic impact of Airbnb

Proponents of Airbnb have highlighted the ways that Airbnb could benefit its customers. Besides offering cheaper alternatives to expensive hotels for tourists in tourist cities, Airbnb could add flexible income to help hosts make their own houses more affordable, either by adding to the mortgage payments or by covering part of the rents. The idea is challenged by a research, questioning that this potential income could only benefit those with adequate computer skills required to participate in the digital economy and

time required to be an Airbnb host, responding and catering to guests. This paper also argues that successfully attracting guests to stay at an Airbnb listing will depend on the individual characteristics of homes appearance, presentation, location, and facilities (Gurran and Phibbs, 2017). Nonetheless, it is acknowledged that in planning for short-term rental regulations, this type of lower-income families should be identified and 'regulatory responses should be designed to support these opportunities while minimizing impacts on neighbouring residents'.

2.4 The impact on housing, rental market, and neighbourhoods

Housing shortage and long-term rental availability have been at the centre of discussion on Airbnb's impact. Airbnb's response to these issues was by stressing on the possibility of allowing hosts to 'make ends meet, keeping residents in communities amid increasing living costs and income inequality' (Airbnb, 2015). This argument was confirmed by research. In a research on Sydney area, it is found that Airbnb hosts are earning rental income roughly equivalent to 10% to 19% of the median rents and mortgages (Gurran and Phibbs, 2017).

However, some papers stressed the impact on long-term rentals. The research by Gurran and Phibbs (2017) summarized findings submitted to the New South Wales Parliament by 212 local entities on the impact of Airbnb on neighbourhoods. They found that the short-term rental income from frequently listed Airbnb listings exceeds long-term rental income by more than AUD 600 per month in greater Sydney area (S\$570 / €370 / US\$413). The price difference is likely to encourage conversion from long-term rental to short-term rentals and creates upward pressure on Sydney's rents. The upward pressure was also mentioned in one study commissioned by Airbnb found that in San Francisco, Airbnb 'increases rents on average by about US\$19 a month' (SFBOS, 2015).

In addition, Gurran and Phibbs (2017) calculated the number of entire units removed from the long-term rental market by counting frequently available entire units on Airbnb. When zooming into popular tourist neighbourhoods where Airbnb listings concentrate such as Waverley and the city of Sydney, the number of whole-houses frequently available on Airbnb is more than three times the vacancy rate in Waverley. In the city of Sydney, a total of 1268 properties were available for short term rental on Airbnb, which is equivalent to 144% of the neighbourhood's vacant rental stock. This impact is studied by various other papers. A paper studying short term rental in Los Angeles found that 7316 rental units were removed from the city's rental market, with acute pressure in beachside neighbourhoods such as Venice, where 12.5% of the neighbourhood's residential units were listed on Airbnb (Lee, 2016). This suggests that Airbnb rentals have a sizeable impact on the availability of long-term rental housing with consequent pressure on rents.

Research has also argued that higher rents generated from short term rental activities could bring neighbourhood displacement and gentrification. There was a case in Portland where a landlord was accused of evicting long term tenants to rent the property as an Airbnb listing (Gurran and Phibbs, 2017). Residents cashing out on homes and leaving the gentrified neighbourhood are another possible scenario. The potential upward pressure on rents and resulting gentrification even has chain effects. There are reports on middle-income renters who were displaced from

the gentrified neighbourhoods and pushed into cheaper neighbouring communities, they in turn gentrify the cheaper neighbouring communities and passing down the chain effect (Lee, 2016). Meanwhile, the landlords who need the extra income from Airbnb to make ends meet are likely to be the ones who share their home with tourists. Landlords who could rent out the entire home for an extended period in a year are likely the ones who own multiple residences. Therefore, the argument of extra income for the community is only likely to hold in cases where one room is listed on Airbnb. The paper also points out a future scenario, where the earning from short term rental activities is capitalized into housing prices of the popular locations, and pushing up the housing price, eroding the affordability benefit and the extra income generated from renting on Airbnb. A case study on the impact of Airbnb in Barcelona found that new housing stock has been added to the neighbourhoods with strong tourists' presence and argued that Airbnb contributed to the housing demand in these areas.

The impact from Airbnb on neighbourhoods are believed to vary. The New South Wales Parliament hearing mentioned in Gurrán and Phibbs (2017) found that for neighbourhoods that has a large presence of tourists, short-term rentals of residential homes are easier to be accepted (Rhodes, 2015). In metropolitan areas that are already troubled by traffic and parking congestion, increasing numbers of tourists are often less welcomed. The Leichardt City Council states that resident complaining about tourist behaviour increased from one in six months to a weekly occurrence with increasing number of Airbnb listings in the neighbourhood. One interesting thing to note is that resident complaints however revolve around a general feeling of unease about increasing presence of visitors in the neighbourhood, rather than tangible cases of noise or nuisances.

2.5 Quantifying the impact and activity of Airbnb

In studies assessing Airbnb's impact, the density of listings or reviews is frequently taken as a reflection of Airbnb activities. To create buffers around housing transactions and calculate the number of active Airbnb reviews or listings in the buffers at the time of transactions is the most frequently used methodology in empirical research on Airbnb. A paper studying the impact of Airbnb on housing prices in Amsterdam applied the buffering method and counted Airbnb reviews as a reflection of Airbnb activities in the buffered neighbourhoods (Bijl, 2016). The same buffering method has been employed in an empirical research conducted on New York, five buffers with radii of 150, 300, 500, 1000, and 2000 meters respectively were created around every transacted unit and counted Airbnb listings within the buffers as Airbnb activities.

To separate the impacts from Airbnb and other neighbourhoods' characteristics, other features reflecting intrinsic features of the transacted unit (such as size and number of bedrooms), building's intrinsic features (such as building age, lift access, number of rooms), neighbourhood amenities (such as school and area of interests), demographic characteristics, and other variables are usually included in the model.

In academic discussion revolving around Airbnb, there are increasing number applying GIS techniques to visualise the data produced from Airbnb, and a lot of them rely on the Airbnb open database project (Airbnb, 2019e). A research comparing the spatial patterns of hotels and peer-to-peer accommodations by Gutiérrez et al. (2017) visualised density maps to determine

the degree of concentration or dispersion of the types of accommodations, and then analysed locational patterns using spatial statistical indicators. In a study done by Garcia-Ayllon (2018), the findings are generated using GIS indicators and geostatistical analysis of spatial correlation and concluded the correlation between short term rental activities and gentrification in the cities of Madrid, Barcelona, and Palma de Mallorca. In this paper we are using some of the measures from previously published research.

3. DATA AND TOOLS

Before presenting the methodology and scope of research, we elaborate on the available data and tools for a better understanding of the possibilities. We also include a short exploratory analysis of the available datasets in order to give an overview.

3.1 Airbnb activity dataset

Airbnb does not disclose a lot of information that could be used in large-scale analyses. A source of data of Airbnb activities in Beijing is Inside Airbnb, an online data portal that continuously web scrapes and consolidates public information on Airbnb website for research purposes. The database has been the basis of several government policy recommendations in recent years, for example the *Analysis of the impact of short-term rentals on housing* by Budget and Legislative Analyst's Office of San Francisco in 2015 (SFBOS, 2015).

Among the datasets of Inside Airbnb, this research focused on two datasets, one describing listings in Beijing, and one including reviews left by guests who have stayed in a particular listing. The listing dataset contains 23 152 entries, each for an individual listing on Airbnb across the timespan and in the study area. Each entry has 97 columns, containing the following information: listing web page information (listing id, listing web link, web scrape id, scrape date, listing name, a summary of listing, description, neighbourhood overview, transit, access, house rules), host information (host id, host web link, host profile name, registered date, residence city, description, response time, response rate, review score, listings number under this host); geographic information (street, neighbourhood, city, state, zip code, country, latitude, longitude), listing features (property type, room type, number of bathrooms, number of bedrooms, number of beds, bed type provided, amenities, size of listing property), and Airbnb booking information (first reviewed date, listed price in RMB, service fee charged, availability of the listing on Airbnb, number of reviews accumulated, review score matrix, whether business trip ready, reviews received per month).

The listing dataset spans across nine years (from 2010-2018). During the timespan, the listing number in Beijing grew from a handful of listings in 2010, to 8568 in 2018, with Airbnb activities peaking in 2017 (see Figure 2).

The rich information associated with each listing provides an opportunity for a thorough understanding of Airbnb activities in Beijing. From the dataset, geographic location information, features such as property type (entire apartment listed or shared space), Airbnb booking information (such as price) are of particular interest to this research. The geographic information allows for an in-depth analysis of Airbnb's impact on the street block level. Property types of listings show the nature of Airbnb activities in Beijing – whether they are more of 'sharing of an under-utilised, vacant room of the house' as claimed by Airbnb, or more of

profit-driven, rental purpose. For example, in the year of 2018, entire room listings make up 61% of the total new addition to Airbnb listings in Beijing. Entire home listings or listings of all private rooms of a particular apartment are in effect informal hotel industry that is not identified nor regulated.

The review dataset contains the following information: the listing id of the reviewed listing, and review information (content of the review, reviewer date, reviewer ID, reviewer name). The review dataset contains 159 039 individual reviews from 2010-2018. Airbnb stays in Beijing grew from a handful of stays in 2010 to 34 919 stays in 2018, with historic highest number of stay at 41 956 in 2017. The Airbnb review dataset is matched to the listing dataset based on the unique listing ID.

The research selected 16 dimensions for the combined Airbnb database, covering: perspective listing information (listing id, property type, room type, square feet, price, number of reviews); review information (review date, reviewer id, reviewer name, comments); and location information (neighbourhood, zip code, latitude, longitude). After filtering out probable errors and outliers (e.g. those with price of zero), we obtained a subset of 95 692 entries, each representing a review left by the guest after his/her Airbnb stays. The category of the entire home rental was most preferred, followed by the 'private room' and 'shared room' in the past two years. Here we made an assumption that every guest leaves a review for the specific listing after their stay. In practice, the rate of the guest leaving a review is reported to be only above 70%, varying for different listings.

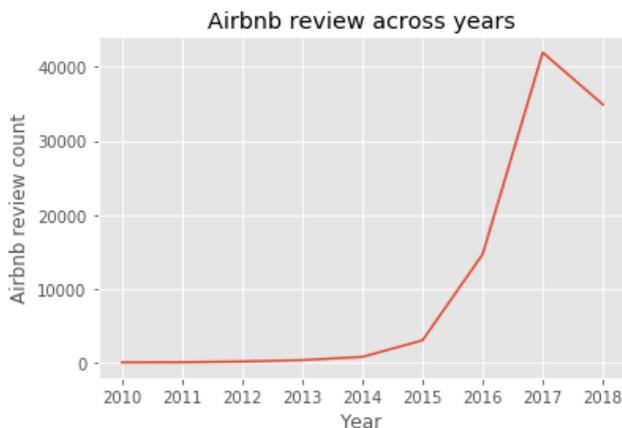


Figure 2. Airbnb reviews during years in the study area.

One thing to note about this dataset is the geocoded location of Airbnb listings. To protect the privacy of the hosts, location information for listings are not precisely available, error is intentionally induced in the coordinates (between 0 and 150 metres from their actual address). In addition, for listings in the same building which have the same coordinates, the geocoding process will assign each of them to different coded coordinates within the 150m buffer, so that may appear 'scattered' in the area surrounding the actual address. However, the geocoding process does not put Airbnb listings to the opposite side of a street, meaning they would only be coded to new coordinates within their original street blocks. And since our analysis focuses on neighbourhoods, a very precise location of each listing is not essential.

3.2 Housing price dataset

The historic housing price data of Beijing come from a property resale, rent, asset management company, LianJia. Estab-

lished in Beijing in 2001, the online portal holds the largest market share of property resale market in Beijing, at around 50%. This research extracted historic housing price data by web scraping past resale transactions from LianJia online portal. The acquired dataset contains 595 087 transactions occurring in 13 selected Beijing districts, from 2013 to 2018. Each entry stands for a single transaction, with 20 columns covering the following dimensions: portal information (property id, webpage, first listed price, transacted price, transacted date, on-market duration, listing price adjustment times, follower, viewed times), geographic information (residential project name, Baidu longitude, Baidu latitude, district), and property features (number of bedrooms, number of living rooms, size, price per square meter in RMB, floor, lift access, built year).

The resale transaction dataset is cleaned in a few steps: drop all basement transactions; drop transactions falling outside of the 13 selected districts; remove outliers by dropping all transactions with PPSM (price per square meter) falling below 100 RMB (around SGD 20); drop transaction with no floor information; drop transactions with no built year information. After cleaning the dataset (filtering out transactions that appear to be erroneous or unrealistic), we have 241 167 entries. Each entry stands for a single transaction occurred between 2013 to 2018. To account for the location specific effect and time specific effects, two sets of categorical variables, year (6 dummy variables) and district variables (13 dummy variables) are then created for the dataset. In addition, as floor information in the dataset is categorical in nature (whether the unit is on ground level / lower level / mid-level / high level / top level), categorical variables for floor level (5 dummy variables) are added into the dataset. A numeric variable, building age, is calculated from the 'built year' variable and added to the dataset. Resale housing price generally increased from 2013-2018 (Figure 3).

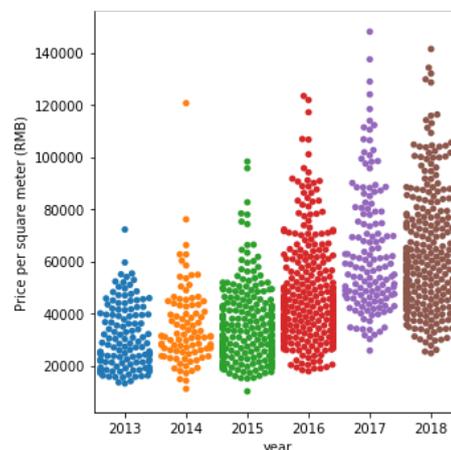


Figure 3. Resale price trend between 2013-2018 (sample of 0.05%).

3.3 Point of Interest (POI)

To analyse the impact of Airbnb activity on neighbourhood housing price, it is important to account for other locational attributes in the neighbourhood. Some common locational attributes, such as accessibility by public transport, public infrastructure, and amenities, have been shown to be influential factors in housing price studies (Diao et al., 2017).

The datasets describing points of interest come from the Beijing government geoinformation database (bjdata.gov.cn). The dataset contains a general description of the POI, such as name, three levels of industry classification, location, and respective coordinates. The POI datasets cover various aspects, including transport (MRT station, bus stops, railway stations, long-distance bus terminals, bus interchanges, airports), education (primary school, secondary school, college), tourism (tourism attraction), commercial (convenience store, Food & Beverage, commercial establishments), and business (office buildings serving as a proxy for job centres). One thing to note about the data for the business category is that only companies falling in the sub-category of “established companies” are included in the analysis. The selected companies are mostly headquarters which occupy sizable office space or industry space. These companies serve as a proxy for job clusters location, as the CBD alone occupies a large area with many other types of land use scattered across the area.

3.4 Tools

This research is conducted using the following tools: QGIS for mapping; Octopus in conducting Web scraping. Python (Numpy, Pandas, Matplotlib, Seaborn) is used for all parts of data analysis. The machine learning model is built with Scikit-learn.

4. METHODOLOGY

4.1 Building experiment data

To measure the impact of Airbnb on neighbourhood housing prices, the Airbnb dataset is merged with the historic housing price dataset with QGIS. The research quantifies the impact of Airbnb by creating a 1 km buffer around each transacted unit as a ‘local neighbourhood’ and measuring Airbnb activities in that spatial extent as a proxy (using the variables mentioned in Section 3.3).

The numeric variables of all the neighbourhood buffers are plotted in Figure 4 to show their frequency distribution. Variable of zero value across all neighbourhood is identified and deemed irrelevant in influencing the neighbourhood’s housing price, therefore abandoned from further analysis (e.g. airport variable).

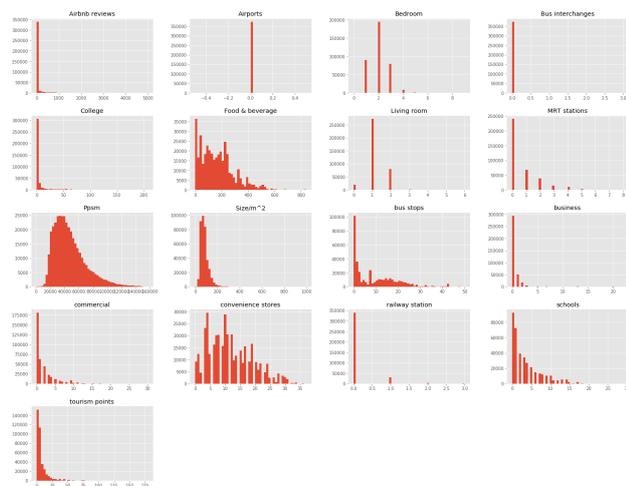


Figure 4. Numeric variables histogram plots.

Figures 5 and 6 investigate whether there exists a relationship between Airbnb activities in the neighbourhood and the price of

housing in the neighbourhood. One percent of neighbourhood buffer dataset is randomly selected to a sample and plotted in Figure 5. The figure shows transacted housing prices with respect to transaction year. The plot suggests that, in general, neighbourhood’s transaction prices increase across the years, with a larger number of Airbnb-occurring neighbourhoods falling in the upper portion of the price range of every year. Figure 6 shows the number of Airbnb activities in the neighbourhood with respect to the transacted housing price. It suggests that, among the high-price neighbourhoods, there is a higher percentage of them having Airbnb activities.

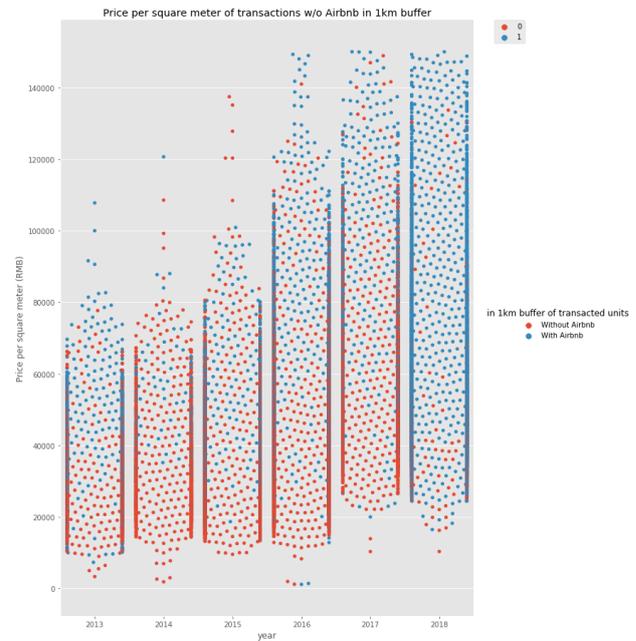


Figure 5. Unit resale price w/o Airbnb in 1km buffer.



Figure 6. Number of Airbnb review in 1km buffer against unit resale price.

Figure 6 shows the number of Airbnb activities in the neighbourhood with respect to the transacted housing price. It indicates that, among the high-price neighbourhoods, there is a higher percentage of them having Airbnb activities.

4.2 Machine learning models

We trained three common machine learning models to find the model of the best fit. The performance matrix measuring the models is root-mean-squared-error (RMSE) and the coefficient

of determination (R-squared) is given in Table 1. RMSE indicates the sum of the squared errors in the model's predictions and suggests the model's predictive capabilities. R-squared indicates how well the models fit the training and validation data sets. It measures to what percentage could the variance be explained by the model. After evaluating its performance, the model that yields minimum RMSE and an R-squared value closer to 1 would be the model of best fit and selected for further analysis.

Table 1. Performance of the models.

Machine learning model	R-squared	RMSE
Multiple linear regression	0.74	138 358 865
Gradient boosting	0.8899	138 210 310
Random Forest	0.9178	44 295 551

Random Forest model outperforms both the Gradient Boosting and Multiple linear Regression, motivating us to focus on it. The research manually tested several parameters and found the R square result to be the best when the following parameters are set. The 'n estimator' is 700: The 'n estimator' is the parameter to specify the number of trees. The 'min samples leaf' = 2: The 'min samples leaf' is the minimum number of samples that must be present in each node of the tree. For an overview of the performance, Figure 8 shows the predicted price compared to the actual price in the test set.

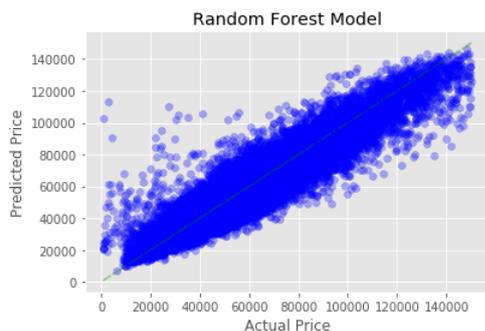


Figure 7. Histogram of sensitivity score.

4.3 Sensitivity Analysis

Sensitivity is defined as the change in predicted value with respect to a change in model attributes. In this research, sensitivity is defined as the percentage change in the predicted housing price when the Airbnb activity in the neighbourhood is changed by a unit:

$$Sv_x = \frac{Pr_{part} - Pr_{orig}}{Pr_{orig}} \quad (1)$$

First the housing price is predicted using the actual values of Airbnb activities (i.e. the number of reviews) in their neighbourhoods. This is denoted by Price original (Pr_{orig}). Second, the values of Airbnb activities in neighbourhoods in the test dataset are increased by one unit, in this case, one review. Third, use the model we trained to predict housing prices given the new Airbnb activity values in their respective neighbourhoods. This is denoted by Price new (Pr_{part}). Finally, the 'sensitivity value' for each neighbourhood housing price is computed, by dividing the change in prediction values by the original prediction price.

The sensitivity value indicates how many times the new prediction value is comparing to the original prediction values. For example, a sensitivity value of 0.03 suggests the new predicted unit housing price (PPSM) is 1.03 times the value of originally predicted unit price, given one extra Airbnb stay in the neighbourhood.

5. IMPLICATIONS OF THE RESULTS

For every neighbourhood in the dataset, the numbers of Airbnb activities were increased by 1 and its unit housing price sensitivity is computed. The value of sensitivity is how many times the new prediction value is comparing to the original prediction values, when Airbnb activity in the neighbourhood is increased by 1 unit.

The sensitivity results suggest that there are three possible responses from neighbourhoods to increasing Airbnb activities (Figure 8). Around half of the neighbourhoods are expected for an increase in neighbourhood housing values, possibly due to speculating investment in the neighbourhoods and resulting gentrification problem. It could also be a result of price competition for limited housing resources, in view of more houses being converted to short term rental use. On the other hand, some neighbourhoods are more resilient to Airbnb activities, with no change in new predicted price (sensitivity = 0). The result also suggests that around 40% neighbourhoods might experience a less than proportionate price drop. For example, a sensitivity score of -0.03 suggests the new predicted price is 99.7% of the original predicted price. The price change is possibly due to nuisance from swarms of tourists in local neighbourhoods.

The sensitivity results are then geocoded to their unique house ID and mapped out using its coordinates, shown in the form of a colour coded map in Figure 9. The blue colour dots in the map implies negative values of price sensitivity, which suggest housing price drops at the location when an Airbnb stay is spent in the neighbourhood. The red colour dots imply a positive price sensitivity and indicate a housing price increase when an Airbnb stay occurs in the neighbourhood. The darker the red colour, the larger percentage increase on housing price.

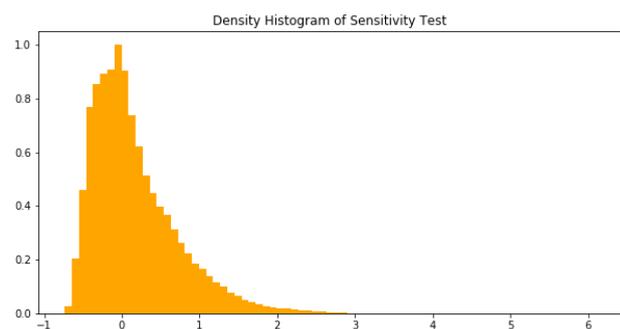


Figure 8. Histogram of sensitivity score.

On the city level, we can conclude that the neighbourhoods in the outskirts districts in Beijing are more price sensitive to an increase in the number of Airbnb activities, compared to the central districts (Figure 9). Neighbourhoods in the central districts, on the other hand, are less price sensitive to Airbnb activities. This might be due to the fact that housing prices are already high in central districts. There is less room for a price increase in the face of Airbnb activities.

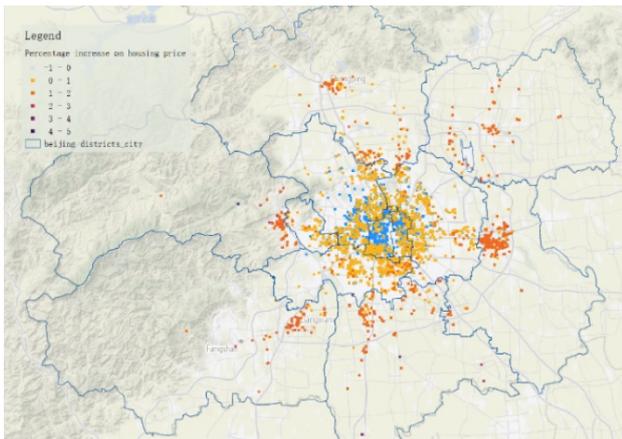


Figure 9. Sensitivity score of neighbourhoods in the study area in Beijing.

The results identified neighbourhoods that are more price sensitive to increasing Airbnb activities but not benefiting from the tourism activities, and thus in need for regulations. For example, Da Xing is an urban neighbourhood that are highly price sensitive to increasing Airbnb activities (per increase in Airbnb review leading to an increase on unit price by 4-5%). Yet a closer examination of the neighbourhood locations reveals that highly price sensitive neighbourhoods mostly locate in highly built up area, along the axial roads connecting to core districts (Figure 10). Airbnb activities in these neighbourhoods are thus less likely to be attracted here by the local neighbourhoods in Da Xing, but a mere expansion effect of Airbnb activities from core districts and a result of tourists seeking for cheaper alternatives in nearby suburb districts. As a result, housing price of these neighbourhoods might increase rapidly with the increasing number of Airbnb, while local businesses are less likely to benefit from the swarm of tourists, as the tourism activities are limited to accommodation and less localized. In addition, the potential price increase is likely to drive out long term renters who also live in these locations. With a large percentage of job opportunities and high housing prices in the core districts, many long term residents in Beijing find cheaper alternatives in accessible locations in suburbs. Airbnb activities in these popular nearby locations compete for rental stock with long term renters, which may lead to displacement of the latter. Local governments could consider implementing policies in these neighbourhoods to lower the potential upward rent pressure and restrain full conversions of residential units to short term rental units. Regulations such as maximal available days could be implemented on Airbnb for listings in such neighbourhoods.

The results also identified neighbourhoods that are price sensitive to increasing Airbnb activities but at the meantime benefiting from the tourism activities. For example, for neighbourhoods that are sub-centers of periphery districts, away from major roads connecting to core districts, the tourism behaviours are likely to be localised. The subcenters in Chang Ping and Shun Yi fall into such category (Figure 11). Regulation for Airbnb activities in these locations could be more paced, focusing on preventing conflicts between local residents and tourists. For example, the capital improvements identified by Airbnb to reduce tourist nuisance in local neighbourhoods could be allowed with the consent of local housing estates, such as separate entrances for long term residents and tourists, carpeted lobbies to absorb the sound of wheeled suitcases, local art for an authentic 'sense of place'.

The research also acknowledges that in some neighbourhoods the price sensitivity results does not imply real consequence and thus could be more liberal in terms of regulation. For example, in neighbourhoods that are located in less built up area of periphery districts (Figure 12), although the price sensitivity scores are relatively high, it is less likely to displace any renters. Rather, Airbnb activity in the neighbourhoods might suggest strong tourism potential and localized tourism behaviours. Short term rental activity in this type of location is likely to generate tourism revenue and benefit local communities, generating income for local governments. It fits into the national development strategy, which seeks to tap on the benefit of sharing economy in tourism industry and maximize the economic growth booster especially for rural areas.

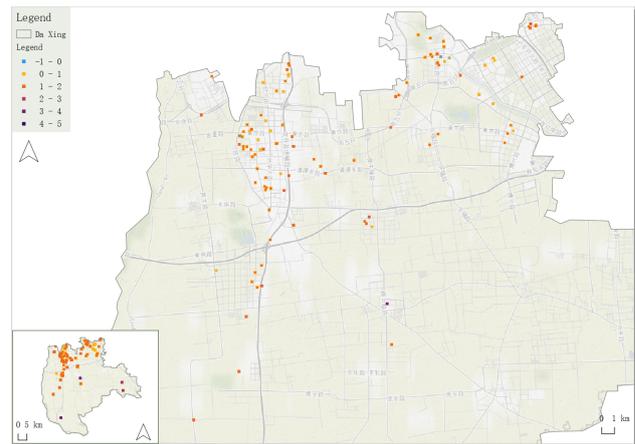


Figure 10. Sensitivity score of neighbourhoods in Da Xing.

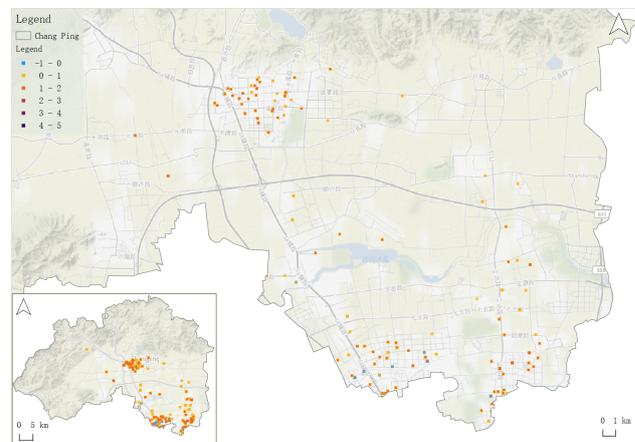


Figure 11. Sensitivity score of neighbourhoods in Chang Ping.

6. CONCLUSION

Our research had two goals. First, by the assumption that the impact of Airbnb would be reflected in the change of neighbourhood housing prices, this research assessed the impact of Airbnb on neighbourhood through price in Beijing. Second, by geo-locating the impacted neighbourhoods, this research tried to use big data analysis and GIS as a toolkit for exploring the regulation of short-term rental services such as Airbnb on a smaller scale. Our paper suggests that with the pool of big data produced alongside the usage of Airbnb platform, urban planners could measure location-specific impacts, and regulate the drawbacks of these disruptive

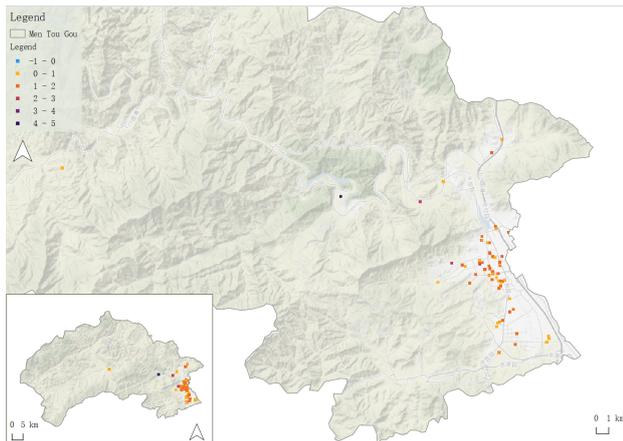


Figure 12. Sensitivity score of neighbourhoods in Men Tou Gou.

technologies in a location-specific scale, which does not entirely rule out new business models yet limits the potential harmful impacts.

The first goal is achieved by building machine learning models to predict the price change, where the Random Forest algorithm generated the highest predictive power and was selected for further analysis. A sensitivity test was then conducted for neighbourhood housing prices. The sensitivity score of each neighbourhood is then mapped to analyse the spatial pattern of neighbourhood's sensitivity. This method successfully located neighbourhoods that are highly price sensitive to Airbnb activity and made possible for location-based regulations.

One major limitation of the research lies in the access to data. As Airbnb does not have any disclosed official data, all data on Airbnb activity in Beijing comes from third party websites. The limited data led to some assumptions in the research process, for example, the proxy we used to measure Airbnb activity is the number of Airbnb reviews, while ideally, the number of bookings per listing is the accurate measure to Airbnb activities in neighbourhoods. In addition, the housing dataset we used in the study is also not a full representation of resale transactions in Beijing.

Future research could be improved by using a more comprehensive database once it becomes available. This may be viable for government agencies as Airbnb discloses information with some municipalities. In addition, future research could focus on specific response to Airbnb, for example, the sense of 'outsider intrusion' felt by local communities. The length of stay could be an indicator that give future research a fresh perspective.

ACKNOWLEDGEMENTS

We thank the reviewers for their comments. This article is based on the master thesis of the first author, carried out in 2019 at the National University of Singapore.

References

Airbnb, 2015. Shared Opportunity: How Airbnb Benefits Communities. https://www.ftc.gov/system/files/documents/public_comments/2015/05/01740-96152.pdf. Retrieved on 31 May 2019.

Airbnb, 2017. *airbnb爱彼迎在中国持续投入: 优化房源品质、全面升级客户服务, 深化与各城市的合作*. <https://press.airbnb.com/zh/investing-in-our-future-quality-customer-service-and-partnerships-in-china/>. Retrieved on 31 May 2019.

Airbnb, 2019a. Night Limits in Amsterdam and London: Frequently Asked Questions. <https://www.airbnb.com.sg/help/article/1628/night-limits-in-amsterdam-and-london-frequently-asked-questions>. Retrieved on 31 May 2019.

Airbnb, 2019b. Night Limits in France: Frequently Asked Questions. <https://www.airbnb.com.sg/help/article/2108/night-limits-in-france-frequently-asked-questions>. Retrieved on 31 May 2019.

Airbnb, 2019c. Occupancy tax collection and remittance by Airbnb in Germany. <https://www.airbnb.com.sg/help/article/2285/occupancy-tax-collection-and-remittance-by-airbnb-in-germany>. Retrieved on 31 May 2019.

Airbnb, 2019d. What is VAT and how does it apply to me? <https://www.airbnb.com.sg/help/article/436/what-is-vat-and-how-does-it-apply-to-me>. Retrieved on 31 May 2019.

Airbnb, I., 2019e. How is Airbnb really being used in and affecting the neighbourhoods of your city? <http://insideairbnb.com>. Retrieved on 31 May 2019.

Bijl, V., 2016. The effect of Airbnb on house prices in Amsterdam. Master's thesis, University of Amsterdam (UvA), MSc Business Economics: Real Estate Finance & Corporate Finance.

Bloomberg, 2018. Airbnb to Share Information With Authorities on Guests in China. <https://www.bloomberg.com/news/articles/2018-03-29/airbnb-to-share-information-with-authorities-on-guests-in-china>. Retrieved on 31 May 2019.

Diao, M., Fan, Y. and Sing, T. F., 2017. A new mass rapid transit (MRT) line construction and housing wealth: Evidence from the circle line. *SSRN Electronic Journal*.

Garcia-Ayllon, S., 2018. Urban Transformations as an Indicator of Unsustainability in the P2P Mass Tourism Phenomenon: The Airbnb Case in Spain through Three Case Studies. *Sustainability* 10(8), pp. 2933.

Gurran, N. and Phibbs, P., 2017. When tourists move in: How should urban planners respond to airbnb? *Journal of the American Planning Association* 83(1), pp. 80–92.

Gutiérrez, J., García-Palomares, J., Romanillos, G. and Salas-Olmedo, M., 2017. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*.

Kelly, H., 2016. Meet the Airbnb police. <https://money.cnn.com/2016/10/01/technology/airbnb-police/index.html>. Retrieved on 31 May 2019.

Lee, D., 2016. How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harvard Law & Policy Review* 10(1), pp. 229–253.

people.cn, 2018. 北京房客平均11个月换一次房最怕房东突然说卖房. <http://society.people.com.cn/n1/2017/0302/c1008-29117385.html>. Retrieved on 31 May 2019.

Rhodes, K., 2015. Adequacy of the regulation of short-term holiday letting in New South Wales: Submission to parliamentary inquiry. Technical report, NSW Government, Sydney, Australia.

SFBOS, 2015. Policy analysis report. Technical report, City And County Of San Francisco Board Of Supervisors Budget And Legislative Analyst. <https://dig.abclocal.go.com/kgp/PDF/265259701-Airbnb.pdf>.