# Airbnb Pricing and Neighborhood Characteristics in San Francisco

## Yanjie LUO and Mizuki KAWABATA

**Abstract**: Airbnb is a prominent peer-to-peer platform that accelerates the rapid expansion of the sharing economy. This article analyzes the relationships between Airbnb pricing and neighborhood characteristics in San Francisco. The relationships are examined with hedonic pricing models at the block group levels. We collect spanning all Airbnb listings in San Francisco and related neighborhood characteristics including transportation accessibility, restaurant sufficiency and demographic attributes from Yelp.com, Census Bureau and SFMTA as study datasets. We find that adding neighborhood characteristics into the analysis model is helpful to enhance the reliability of models while different categories of transportation, population and housing characteristics have different significance and effects in specific analysis models. The results are expected to help us develop tourism management policies.

**Keywords**: Airbnb, neighborhood characteristics, hedonic pricing models, San Francisco

### 1. Introduction

In recent years, urban tourism has encountered massive growth and has become an essential affair in many metropolises. Simultaneously, new technologies and the innovative development of Internet bring some sorts of new service such as Uber, Airbnb, and Mobil-cycling, with the benefit of widespread scales of population which are able to mutually make use of specific inventory based on fee structure (Zervas et al., 2017). The fundamental phenomenon belongs to one component of "the sharing economy."

Airbnb is a prominent example of the sharing economy. Since launching in 2008 in San Francisco, CA, Airbnb has grown to become one of the largest single tourism accommodation distribution platforms in the world, with 2 million listings and 60 million guests (Airbnb, 2018). Airbnb, on one hand, encourages tourists to "live like a local," insinuating that it can enhance emoluments to local hosts and industries without auxiliary jeopardies required on neighbors and regional communities. Critics, on the other hand, declare that Airbnb has capacitated rental accommodations to infiltrate residential neighborhoods, which makes conflicts increasing between guests and regional residents, dislodging permanent house units in high-demand cities and accelerating affordability pressures for low-income strata (Gurran et al., 2017; Zervas et al., 2017). Furthermore, the surrounding neighborhood characteristics such as crime have a specific influence on the location and the price decision of Airbnb accommodations (XU and Pennington-Gray, 2017).

In this study, we emphasize the question of whether there is evidence that the neighborhood characteristics of the Airbnb accommodations affect the listing price of Airbnb rental accommodations in San Francisco. The purpose is to indicate potential price determinants based on neighborhood characteristics in contrast to internal factors. The relationships are examined with hedonic pricing models at the block group levels. The results are expected to help us develop tourism management policies.

### 2. Methodology

We use hedonic pricing models to examine the effects of a group of related characteristics, especially neighborhood characteristics on Airbnb listing price in comparison with related studies. We look at which characteristics can elucidate Airbnb listings prices, and therefore investigate users' valuation for specific attributes. As characteristics derive utility and create a basement for users to act when the attributes are presented, the market price of an Airbnb listing is supposed to be a function represented by associated characteristics. We specify the hedonic pricing function of Airbnb listings as follows:

$$\ln Price_i = \alpha + \beta P_i + \gamma H_i + \delta R_i + \tau N_i + \varepsilon_i \quad (1)$$

The $P$ vector contains variables on the features of physical characteristics of Airbnb listings. The $H$ vector is composed of characteristics which are related to Airbnb hosts. The $R$ vector contains categorical variables on reputation characteristics from guests, while the $N$ vector contains variables on neighborhood characteristics related to each Airbnb accommodations spatially.

### 3. Data

We created a dataset of 6,624 Airbnb listings in 579 neighborhoods of San Francisco in the U.S. as of August 2018 (Figure 1). We chose San Francisco as our study area for two reasons. Firstly, as the spiritual birthplace of Airbnb, San Francisco is a historic city that undergoes serious obstacles with various racial inhabitants and international tourists related to tourism and crime. Secondly, many existing studies such as Kakar et al. (2016, 2018) have estimated various aspects of Airbnb in San Francisco from diverse perspectives empirically. Table 1 describes the variables used in the hedonic pricing model, with ln*Price* as the dependent variable. The Airbnb listing price data are derived from the Inside

Yanjie LUO, Ph.D. Student, Graduate School of Economics, Keio University, 2-15-45, Mita, Minato-ku, Tokyo, Japan
Phone: +81-80-5728-2893
E-mail: luoyanjie1992@gmail.com

Airbnb website (http://insideairbnb.com/). We extracted physical characteristics, host characteristics, and reputation characteristics from Inside Airbnb data of San Francisco. Besides, we collected the following neighborhood characteristics from open datasets.

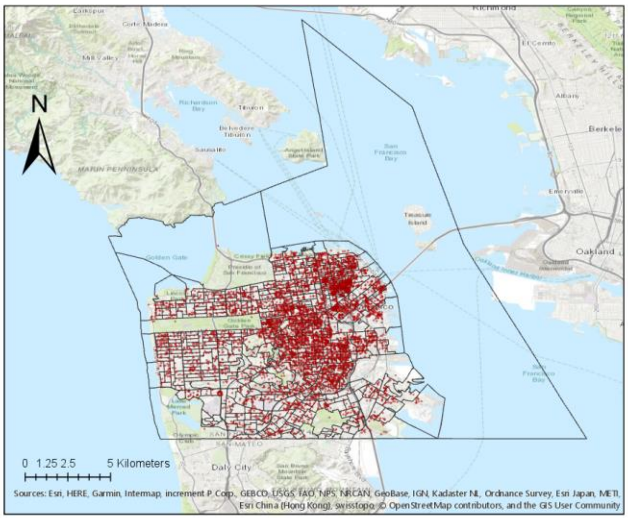

Figure.1 Airbnb accommodations in San Francisco

Firstly, we collected crime data from DataSF (https://datasf.org/opendata/) provided by the San Francisco City Government, which dispensed crime heat map and related detail datasets from 2003. Crime detail data contain information of crime types, day of week, date, time, district, resolution, address, and geological locations both latitude and longitude. We chose all crime types and time of crime data during the last 3 years to apply into analysis model. Secondly, population at block levels came from Census Reporter, based on 2016 ACS (American Community Survey) 1-year Estimates from the U.S. Census Bureau.

In comparison with previous studies, we emphasize location characteristics analysis more precisely rather than select the distance to city hall or the geological center of the city in previous studies. We categorize SFMTA (San Francisco Municipal Transportation Agency) transit stops into different groups by transportation types such as muni metro, muni bus, cable car, historical street car. We assembled data of Top 30 most popular attractions from Trip Advisor and batched these locations into geographic coordinates by Google Maps API and Python 3.7.0. We calculated the walking distance from each Airbnb accommodation to the nearest transit stop/ attraction by Find Nearest Analysis, supported by ArcGIS Online. Find Nearest Analysis follows paths and roads that allow pedestrian traffic and finds solutions that optimize travel time. The walking speed is set to 5 kilometers per hour.

In this research, we applied restaurants data into our analysis model by extracting open dataset from Yelp.com, known as the most popular crowd-sourced review forum in the US. We calculated the number of restaurants located within the walking distance of 5 minutes and 10 minutes by Summarize Nearby Analysis, supported by ArcGIS Online.

Table.1 Descriptions of research variables

| Variables | Decriptions | Sources |
|---|---|---|
| ***~Airbnb~*** | | |
| **lnPrice** | Logarithm of daily price of Airbnb accommodation | Inside Airbnb |
| ***~Physical characteristics~*** | | |
| **PropertyType1** | Dummy variable of combined accommodation types: 1=Apartment, Serviced Apartment, condominium and loft, 0=others | Inside Airbnb |
| **PropertyType2** | Dummy variable of combined accommodation types : 1=Bangalow, Guest House, House, Tiny House, Town House, Tree House, Villa, Cabbin and Cottage, 0=others | Inside Airbnb |
| **Accommodations** | Number of people that can be accommodated | Inside Airbnb |
| **Bathrooms** | Number of bathrooms for a given listing | Inside Airbnb |
| **Bedrooms** | Number of bedrooms for a given listing | Inside Airbnb |
| **Wireless Internet** | Dummy variable of offer wireless Internet access. | Inside Airbnb |
| **Breakfast** | Dummy variable of offer breakfast. | Inside Airbnb |
| **Free parking** | Dummy variable of offer free parking. | Inside Airbnb |
| ***~Host characteristics~*** | | |
| **Superhost** | Dummy variable of whether a given listing's host is super host, 1 =super host and 0 = regular host | Inside Airbnb |
| **Identity verified** | Dummy variable of whether identity of a given listing's host is vertified , 1 =vertified, 0= not vertified | Inside Airbnb |
| **Cancellation** | Dummy variable of strictness of cancellation policy, with values of 1 = flexible, 0 = modeproportion and strict | Inside Airbnb |
| **Host Verifications** | Number of verification options (email, phone, Facebook, Google,LinkedIn, etc.) disclosed online for a given listing's host | Inside Airbnb |
| **Host long** | days since became Airbnb listing's host | Inside Airbnb |
| ***~Reputation characteristics~*** | | |
| **Review Overall Rating** | Customer-geneproportiond review scores of the overall ratings for a given listing | Inside Airbnb |
| **Review Scores Accuracy** | Customer-geneproportiond review scores of accuracy of accommodation description for a given listing | Inside Airbnb |
| **Review Scores Cleanliness** | Customer-geneproportiond review scores of cleanliness for a given listing | Inside Airbnb |
| **Review Scores Checkin** | Customer-geneproportiond review scores of the overall ratings for a given listing | Inside Airbnb |
| **Review Scores Communication** | Customer-geneproportiond review scores of check-in for a given listing | Inside Airbnb |
| **Review Scores Location** | Customer-geneproportiond review scores of location for a given listing | Inside Airbnb |
| **Review Scores Value** | Customer-geneproportiond review scores of value of Airbnb listing's accommodation for a given listing | Inside Airbnb |
| ***~Neighborhood characteristics~*** | | |
| **Transit1: Bus** | Walking distance (miles) to the nearest bus stop | SFMTA |
| **Transit2: Cable Car** | Walking distance (miles) to the nearest Cable Car station | SFMTA |
| **Transit3: Metro** | Walking distance (miles) to the nearest Metro station | SFMTA |
| **Transit4: Street Car** | Walking distance (miles) to the nearest Historical Streetcar station | SFMTA |
| **5minsrest** | Number of restaurants within 5-minute walking distance (miles) from Airbnb accommodation | Yelp open dataset |
| **10minsrest** | Number of restaurants within 10-minute walking distance (miles) from Airbnb accommodation | Yelp open dataset |
| **Attractions** | Walking distance (miles) to the nearest main tourist attractions | TripAdvisor |
| **Populatiom Density** | Density of inhabitants at block group levels in 2016 | U.S. Census Bureau (2016) |
| **Crime Density** | Density of crime occurrence during last 3 years(From 8/6/2015 through 8/6/2018) | DataSF |
| **Young Proportion** | Percentage of people aged between 18 and 34 years | U.S. Census Bureau (2016) |
| **Income** | Logarithm of median household income | U.S. Census Bureau (2016) |
| **Employed Proportion** | Percentage of employed population in labor force residents | U.S. Census Bureau (2016) |
| **High Education Proportion** | Percentage of residents aged 25 years old with B.A. and higher educational degrees | U.S. Census Bureau (2016) |
| **Tenure Proportion** | Percentage of occupied housing units | U.S. Census Bureau (2016) |
| **House value** | Logarithm of median house value | U.S. Census Bureau (2016) |
| **Vacancy Proportion** | Percentage of vacancy | U.S. Census Bureau (2016) |
| **Single Proportion** | Percentage of single residents | U.S. Census Bureau (2016) |
| **White Proportion** | Percentage of non-hispanic white inhabitants | U.S. Census Bureau (2016) |
| **Black Proportion** | Percentage of non-hispanic black inhabitants | U.S. Census Bureau (2016) |
| **Asian Proportion** | Percentage of non-hispanic asian inhabitants | U.S. Census Bureau (2016) |
| **Hispanic Proportion** | Percentage of hispanic inhabitants | U.S. Census Bureau (2016) |

## 4. Results

Table 2 presents the estimation results of the following five models extended from function (1):

Model 1: $\ln Price_i = \alpha + \beta P_i + \varepsilon_i$

Model 2: $\ln Price_i = \alpha + \beta P_i + \gamma H_i + \varepsilon_i$

Model 3: $\ln Price_i = \alpha + \beta P_i + \gamma H_i + \delta R_i + \varepsilon_i$

Model 4: $\ln Price_i = \alpha + \beta P_i + \gamma H_i + \delta R_i + \tau N_i + \varepsilon_i$

Model 5: $\ln Price_i = \alpha + \beta P_i + \gamma H_i + \delta R_i + \tau N_i + \varepsilon_i$

The adjusted R-squared in Model 1 is 0.402, indicating that the physical characteristics explain 40.2 percent of the variation in the market price of Airbnb listings. By adding host characteristics to physical characteristics (Model 2), the adjusted R-squared increased by nearly 1.5 percentage points to 41.7 percent. Then by including reputation characteristics and neighborhood characteristics, the adjusted R-squareds of Models 3-5 increased by 3 and 8 percentage points.

In Models 1-3, property types, accommodations, bathrooms, and bedrooms are positively and significantly associated with the market price of the Airbnb listings. However, property type 2 is not significant in Models 4 and 5. The number of bathrooms has negative and significant effects on the Airbnb prices, indicating that the market price declines with more bathrooms. Airbnb users in San Francisco might prefer choosing compact accommodations with fewer bathrooms. The significant and positive effects of bedrooms are observed in Models 1-5, which mean that the market price would increase with more bedrooms. These results suggest that like Airbnb mission statement "To live in this world where you can be home"(Airbnb, 2018), fewer bathrooms and more bedrooms would make more connected spaces for guests and accommodations with right size would be more popular for guests in Airbnb rental market. The characteristics of wireless internet and breakfast are almost nonsignificant in Models 1-5, which are different from the findings of Wang and Nicolau (2017). In Models 4 and 5, free parking has significantly positive effects on the market prices, coinciding with research expectation and in comparison with Gibbs et al. (2018), parking only is associated with significant price increases in 2 cities of 5 study areas while it is significant when models include neighborhood characteristics in this study.

The host characteristics are included in Models 2-5. The identity verified and the number of host verification options exhibit significantly positive effects, while flexible cancellation policy shows statistically negative effects on the Airbnb price. We can deduce that Airbnb accommodations with higher prices would hold more strict cancellation policy when users make booking reservations. The time since became Airbnb hosts is only significant in Models 2 and 3. Presented in Models 3-5, Super host characteristic has significant positive effects on market price as well as other host characteristics, which means that Super host verification would increase the security and reliability of accommodations emotionally and positively act on Airbnb prices. (Chen and Xie, 2017; Xie and Mao, 2017; Guttentag et al., 2018)

Except for review scores of accuracy and location, the other characteristics of reputation are all significant in Models 3-5. Review scores of overall rating and cleanliness exhibit significant and positive effects on the market prices, while scores of check-in, value and communication show significant and negative effects. It is supposed to be cited that customer reviews have been emphasized by tremendous studies investigating demand effects (Panda et al., 2015; Ert et al., 2016). Nevertheless, the effects of some customer reviews in impacting the market price of Airbnb listings in our study are faint. To explain such an insignificant effect, it is mentioned in Panda et al. (2015) that diverged from physical characteristics, social interactions are a post hoc yardstick of Airbnb's performance, which cannot be predicted by guests prior to their reservations. While guests would envisage that social interactions with hosts and neighbors could be more intensive and intimate during their stays, their expectation can only be satisfied afterward. As a consequence, it may intensify an incidental effect on their valuation of Airbnb accommodations in advance.

Furthermore, a growing proportion of professional hosts in Airbnb market may have handicapped the host-guest interactions and communications (Edelman et al., 2014, 2017). When social interactions are inordinately commercialized to only seek for billeting mass tourists and making profits, the advantage of social interactions that Airbnb asserted as its concept may subside.

Table.2 Estimated results model

| Variables | Model1 | Model2 | Model3 | Model4 | Model5 |
|---|---|---|---|---|---|
| *~Physical characteristics~* | | | | | |
| **PropertyType1** | 0.189*** | 0.193*** | 0.168*** | 0.024* | 0.024* |
| **PropertyType2** | 0.124*** | 0.173*** | 0.188*** | -0.007 | -0.009 |
| **Accommodations** | 0.141*** | 0.139*** | 0.139*** | 0.135*** | 0.135*** |
| **Bathrooms** | -0.049*** | -0.038*** | -0.036*** | -0.057*** | -0.056*** |
| **Bedrooms** | 0.219*** | 0.215*** | 0.211*** | 0.214*** | 0.214*** |
| **Wireless Internet** | 0.048 | -0.051*** | -0.049 | -0.042 | -0.041 |
| **Breakfast** | -0.004 | -0.001 | -0.005 | -0.014 | -0.014 |
| **Free parking** | 0.010 | -0.010 | -0.001 | 0.071*** | 0.071*** |
| *~Host characteristics~* | | | | | |
| **Superhost** | | 0.111 | 0.105*** | 0.100*** | 0.100*** |
| **Identity verified** | | 0.035*** | 0.032** | 0.029** | 0.028** |
| **Cancellation** | | -0.144** | -0.157*** | -0.159*** | -0.159*** |
| **Host Verifications** | | 0.000*** | -0.015*** | -0.016*** | -0.016*** |
| **Host long** | | -0.015*** | 0.000** | 0.000 | 0.000 |
| *~Reputation characteristics~* | | | | | |
| **Review Overall Rating** | | | 0.003** | 0.003* | 0.003* |
| **Review Scores Accuracy** | | | 0.000 | 0.008 | 0.008 |
| **Review Scores Cleanliness** | | | 0.120*** | 0.119*** | 0.119*** |
| **Review Scores Checkin** | | | -0.113*** | -0.075*** | -0.075*** |
| **Review Scores Communication** | | | -0.059*** | -0.043*** | -0.043*** |
| **Review Scores Location** | | | 0.075*** | -0.008 | -0.008 |
| **Review Scores Value** | | | -0.06*** | -0.032*** | -0.032*** |
| *~Neighborhood characteristics~* | | | | | |
| **Transit1: Bus** | | | | 0.375*** | 0.369*** |
| **Transit2: Cable Car** | | | | -0.095*** | -0.093*** |
| **Transit3: Metro** | | | | 0.027 | 0.029 |
| **Transit4: Street Car** | | | | 0.023* | 0.023* |
| **5minsrest** | | | | 0.000 | |
| **10minsrest** | | | | | 0.000 |
| **Attractions** | | | | -0.024** | -0.023* |
| **Populatiom Density** | | | | 0.000*** | 0.000*** |
| **Crime Density** | | | | 0.000 | 0.000 |
| **Young Proportion** | | | | 0.139* | 0.143* |
| **lnincome** | | | | 0.000 | 0.000 |
| **Employed Proportion** | | | | 0.221 | 0.226 |
| **High Education Proportion** | | | | 0.238*** | 0.238*** |
| **Tenure Proportion** | | | | 0.065* | 0.064* |
| **lnvalue** | | | | 0.000 | 0.000 |
| **Vacancy Proportion** | | | | 0.275** | 0.270** |
| **Single Proportion** | | | | -0.008 | -0.012 |
| **White Proportion** | | | | 0.171* | 0.176* |
| **Black Proportion** | | | | 0.073 | 0.075 |
| **Asian Proportion** | | | | -0.083 | -0.085 |
| **Hispanic Proportion** | | | | 0.109 | 0.107 |
| **Observations** | 6624 | 6624 | 6624 | 6624 | 6624 |
| **Adjusted R2** | 0.402 | 0.417 | 0.441 | 0.498 | 0.498 |

(Note: ***$p < 0.01$, ** $p < 0.05$ and * $p < 0.1$)

We estimated the effects of neighborhood characteristics at the block group levels in Models 4 and 5 which contained the number of restaurants within 5- or 10-minute walking distance from Airbnb accommodations. Like most previous studies such as Gibbs et al. (2018), Gutiérrez et al. (2017), and Teubner et al. (2017) which focus on Airbnb pricing decision applied by the hedonic pricing models, as spatial and geographical attributes, the location is generally defined as the distance to city hall or city center simply, as well as accessibility of transportation is habitually ignored or simplified. According to our research purpose acknowledged as “explore the relationship between Airbnb pricing and neighborhood characteristics,” we classified and processed these neighborhood characteristics. Consequently, the walking distance to the nearest MUNI bus stop and historical streetcar significantly and positively affects the market price. However, the walking distance to the nearest cable car stop and popular attractions show the significant negative effects on the market price. As well as attractions, three cable car lines in San Francisco City are located in the area from Market St. to Fisherman's Wharf, which is generally viewed as the center of City. It can be conjectured that Airbnb users would like to choose those accommodations away from clamorous city center rather than preferring accessibility of transportation. The effects of population density are significant and feebly negatively associated with the market price. Young and high education characteristics exhibit significantly positive effects, as well as the proportion of owner-occupied housing units. The areas which assemble young and higher educational population would be more popular and reflect recognition into the market price. The significance of the white proportion characteristic also embodies this inference. Out of curiosity, vacancy proportion characteristics show strongly significantly positive effects on the market price. This result suggests that vacancy supply can positively act on consumer valuation of Airbnb listings.

## 5. Conclusion

This article analyzes the relationships between Airbnb pricing and neighborhood characteristics in San Francisco. The relationships are examined with the hedonic pricing models at the block group levels. By examining such comprehensive factors that influence prices of Airbnb listings, this research provides hosts with insights as to potential strategies they could implement to increase their revenues. The findings could also be useful to Airbnb when suggesting prices to hosts, in addition to underscoring for Airbnb the need to further educate hosts. The analysis results also empathize some aspects of Airbnb that are alarming from policy perspectives managed by government and public authorities, the cooperative relationship across various regional industries and identified rental accommodation supervision strategies. The research also has various limitations. As the further proposal, it is necessary to estimate more precisely the transportation and restaurants characteristics in other metropolitan areas and the data scraping effort is involved in collecting data across multiple time periods for multiple markets to compare different circumstances chronologically.


## Acknowledgments

This work was supported by Keio University Doctorate Student Grant-in-Aid Program.


## References


Airbnb, 2018. About us. <https://press.airbnb.com/about-us/>

Chen, Y., & Xie, K., 2017. Consumer valuation of Airbnb listings: a hedonic pricing approach. *International journal of contemporary hospitality management*, **29(9),** 2405-2424.

Edelman, B., Luca, M., 2014. Digital Discrimination: The Case of Airbnb.com. *Harvard Business School NOM Unit Working Paper*, **14,** 054.

Edelman, B., Luca, M., and Svirsky, D., 2017. Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, **9(2),** 1-22.

Ert, E., Fleischer, A., and Magen, N. 2016. Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, **55,** 62-73.

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., and Goodwill, A., 2018. Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, **35(1),** 46-56.

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),** 80-92.

Gutiérrez, J., García-Palomares, J. C., Romanillos, G., and Salas-Olmedo, M. H., 2017. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*, **62,** 278-291.

Guttentag, D., Smith, S., Potwarka, L., and Havitz, M., 2018. Why tourists choose Airbnb: a motivation-based segmentation study. *Journal of Travel Research*, **57(3),** 342-359.

Kakar, V., Franco, J., Voelz, J., and Wu, J. 2016. Effects of host race information on Airbnb listing prices in San Francisco. *Munich Personal RePEc Archive,* 69974

Kakar, V., Voelz, J., Wu, J., and Franco, J. 2018. The Visible Host: Does Race Guide Airbnb Rental Rates in San Francisco?. *Journal of Housing Economics***, 40,** 25-40.

Panda, R., Verma, S., and Mehta, B., 2015. Emergence and acceptance of sharing economy in India: understanding through the case of Airbnb. *International Journal of Online Marketing (IJOM)*, **5(3),** 1-17.

Teubner, T., Hawlitschek, F., and Dann, D., 2017. Price Determinants on Airbnb: How Reputation Pays off in the Sharing Economy. *Journal of Self-Governance & Management Economics*, **5(4).**

Wang, D., and Nicolau, J. L., 2017. Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb. com. *International Journal of Hospitality Management*, **62,** 120-131.

Xie, K., and Mao, Z., 2017. The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management*, **29(9),** 2240-2260.

XU, Y. H., and Pennington-Gray, L., 2017. Explore the spatial relationship between Airbnb rental and crime. *Tourism Travel and Research Association: Advancing Tourism Research Globally.* **5.**

Zervas, G., Proserpio, D., and Byers, J. W., 2017. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of Marketing Research*, **54(5),** 687-705.