



Munich Personal RePEc Archive

The Visible Host: Does Race guide Airbnb rental rates in San Francisco?

Kakar, Venoo and Franco, Julisa and Voelz, Joel and Wu,
Julia

San Francisco State University

10 March 2016

Online at <https://mpra.ub.uni-muenchen.de/78275/>

MPRA Paper No. 78275, posted 14 May 2017 16:35 UTC

The Visible Host: Does Race guide Airbnb rental rates in San Francisco?

Venoo Kakar ^{*} Joel Voelz [†] Julia Wu [‡] Julisa Franco [§]

June 18, 2016

Abstract

The surge in Peer to Peer e-commerce has increasingly been characterized by changing the online marketplace to a more personalized environment for the buyer and seller. This personalization involves revealing information on buyer reviews, pictures and biographical information on the sellers to reduce the perceived “purchase risk” or to facilitate trust with the buyers. However, this personalization has generated possibilities for discrimination in the online marketplace. In this paper, we examine the effect of host information available online (race, gender and sexual orientation etc.) on price listings on Airbnb.com in San Francisco. We find that Hispanic hosts and Asian hosts, on average, have a 9.6% and 9.3% lower list price relative to their White counterparts, after controlling for neighborhood property values, user reviews and rental unit characteristics. We don’t find any significant impact of gender and sexual orientation on price listings. Overall, our findings corroborate the presence of racial discrimination in the online marketplace.

Keywords: Airbnb, Discrimination, Race, Online marketplace

JEL: D40, D47, J15, J71

^{*}Department of Economics, San Francisco State University, E-mail: vkakar@sfsu.edu

[†]Department of Economics, San Francisco State University, E-mail: jvoelz@mail.sfsu.edu

[‡]Department of Economics, San Francisco State University, E-mail: jwu18@mail.sfsu.edu

[§]Department of Economics, San Francisco State University, E-mail: julisa92@mail.sfsu.edu

1 Introduction

Internet commerce has grown dramatically over the past decade, moving from an interesting niche to a mainstream component of both business and consumer markets with 2013 figures totaling over \$700 Billion of consumer purchases in the U.S. A growing sub-market is the area of Peer-to-Peer or P2P commerce. P2P e-commerce involves consumers or small craftspersons acting as sellers and buyers offering and buying everything from used goods (EBay/Craigslist) to new craft products (Etsy) to personal services (TaskRabbit) to rooms for rent (Airbnb).¹ Accompanying this growth and emergence of the P2P e-commerce market has been an evolution from the early anonymous arms-length transaction environment of the internet to a more personalized environment and purchase process where the goal is to make a personal connection between the buyer and seller. This personalization involves techniques such as buyer reviews, pictures and biographical information on the sellers to give potential buyers more information on the seller. These techniques attempt to reduce perceived purchase risk and create a personal social connection making the purchase from a stranger more palatable.

However, as P2P commerce has become more personal, it becomes less anonymous and so opens the possibility of various forms of discrimination by both buyers and sellers. This can occur because the race and gender of participants are frequently revealed through photos and biographical information. So, while P2P e-commerce has opened opportunities for minorities to participate in a growing market it has also generated questions about and possibilities for discrimination similar to face-to-face markets. Buyers now have the information to bypass e-commerce sellers based on race or gender in a manner similar to bypassing a brick and mortar store.

In this paper we address the question of whether there is evidence that information regarding the race of the Airbnb host affects the listing price of rooms in the San Francisco market. The assumption is that this could indicate potential price discrimination based on race against the hosts. Airbnb incorporates multiple rating techniques to help increase buyer confidence including reviews by previous guests and available social media links of the hosts. To personalize listings, they permit sellers offering room listings (the hosts) to provide both a picture and biographical/listing information. This allows potential renters to identify both the race and sex of the host. A representative sample of Airbnb listings in San Francisco was analyzed and the race of each host identified visually through their posted picture. After controlling for common factors, we find evidence of statistically significant price differences for racial minorities. We find that, White hosts are able to charge 9.2% more than Asian hosts and 9.6% more than Hispanic hosts holding all other listing factors constant such as location, room amenities and ratings.

These findings raise questions concerning the migration of racial discrimination to on-line markets, possible differing pricing strategies or business objectives of minorities, and policy issues regarding potential liability of companies such as Airbnb where business practices enable potential discrimination in on-line commerce.

¹At the end of 2014, Airbnb had 925,000 listings and over 25 million customers, [Todisco \(2015\)](#).

1.1 Related Literature

1.1.1 Discrimination in Traditional Housing markets

Discrimination against minorities in the rental and housing market is certainly not new. [Turner \(2013\)](#) found evidence of significant discrimination against minorities related to rental offerings and availability. The study conducted more than 8,000 paired tests in a nationally representative sample of 28 metropolitan areas. As reported, Hispanic renters learned about 12.5 percent fewer available units and were shown 7.5 percent fewer units than Whites. Asians learned about 9.8 percent fewer available units and were shown 6.6 percent fewer units than Whites. Also, [Bayer et al. \(2012\)](#) analyzed panel data covering over two million repeat-sales housing transactions from four metropolitan areas and found that Black and Hispanic homebuyers pay premiums of about three percent on average across the four cities, differences that are not explained by variation in buyer income, wealth or access to credit. While we did not use the techniques of these early studies, they provide a backdrop against which our study and others evaluating the emerging internet commerce space are evaluated.

1.1.2 Discrimination in the emerging P2P Commerce Market

The emergence of the internet and on-line commerce has created both opportunities and pitfalls. Early Internet commerce provided anonymity to both buyer and seller. This lack of personal information increased the arms length nature of the transaction and removed many opportunities to practice discrimination against a now unknown” buyer or seller.

However, in recent years the trend in internet commerce - and especially in the rapidly growing P2P market has been to increase the personalization of buyer and seller in an attempt to reduce the perceived risk of dealing with individuals instead of a commercial concern. An earlier study, [Pope and Sydnor \(2011\)](#) examined the effect a loan applicant’s personal information, including a picture on their acceptance rate for a personal loan at the P2P lending site Prospero. The study found that Black applicants had a 2.4-3.2 percentage points lower chance of getting funded, other factors held constant. The study further compared this to the average probability of getting funded, 9.3%. Blacks had a 30% decrease in the likelihood of funding. Additionally, they found that the interest rates offered to Blacks were 60-80 basis points higher than Whites with similar credit profiles. This was an example of supply side discrimination.

Another recent study, [Doleac and Stein \(2013\)](#) posted classified advertisements offering iPod Nano music players for sale on several hundred locally focused websites throughout the US. They observed the effect of the seller’s skin color on outcomes such as bid price and preference for a face-to-face delivery vs having the product mailed. They included a photograph of a dark -skinned (Black) or light-skinned (White) hand holding the item for sale and were able to vary the apparent race of the seller while fixing other sales and market characteristics. In addition to race they examined the effect of a social signal communicated by a tattoo on the seller’s hand. They used this option as a suspicious’ White control group.

Controlling for all common factors it was found that Black visible hand seller ads had 13% fewer responses and 18% fewer offers. The offers made to Black-handed sellers were on average 11% lower than those made to White-handed sellers. Black-handed sellers also garnered lower trust as they were 17% less likely to have the bidder’s name included in their

initial offer or inquiry. This study shows the impact of visual information that identifies the race of the seller, although it cannot conclude whether the captured effect is due to taste-based discrimination. This is similar to our study’s examination of the use of host pictures on Airbnb listings. In this case, it was an example of demand side discrimination as potential buyers bid lower on products being sold by minorities or life style groups (such as those with tattoos) that are discriminated against.

1.1.3 Discrimination in the P2P Room Rental Market of Airbnb

Two recent studies have focused on potential pricing effects of racial discrimination against minority hosts of Airbnb, the P2P commerce site for short term room rentals. The initial study by [Edelman and Luca \(2014\)](#) using the set of all Airbnb listings in New York City, identified the race of the hosts using their on-line listing photo. Controlling for all visible information on Airbnb listings regarding the quality of the listing they found statistically significant evidence that non-Black hosts charge approximately 12% more than Black hosts for an equivalent rental. They point out the possible unintended consequences of offering hosts the ability to post pictures for increased personalization on enabling discrimination.

All 3752 New York listings were downloaded and the offered rental price as well as all other listing variables coded. Amazon Mechanical Turk was used to categorize a host’s race based on their on-line photo as well as individually rate each listing for quality on a 1 seven-point scale.² Using the Airbnb rating variables (cleanliness, location, communication, check-in, etc) as well as the host race and listing quality, they find evidence of differences in listing price based on race.

Their findings showed a statistically significant difference between posted rent prices of Black hosts vs non-Black hosts. They also found that non-Black hosts earn roughly 12% more for a similar rental relative to Black hosts. Although Black hosts’ listings tended to be in less desirable locations and therefore could be expected on average to have lower prices, their study controlled for all attributes that are viewable by customers in making a rental decision. However, they were not able to determine the extent of taste-based vs statistical discrimination taking place. This study did not explicitly try to control for the quality of the property or the value of the neighborhood as a factor affecting the listing price and it focused only on Black/Non- Black hosts. Our study has tried to improve on this by using a normalized factor for property values in each neighborhood where a listing is located in addition to race, gender, couple and sexual orientation variables.

A follow up study to [Edelman and Luca \(2014\)](#) focused on potential racial discrimination against Asian Airbnb hosts in the city of Oakland in the California market, [Wang et al. \(2015\)](#). Although a minority group, Asians have a very different social, economic, and educational profile than Blacks. They have the highest median income of all racial groups, as well as the highest average test scores for college admission. Despite polling data where Asians indicated they were more satisfied than the general public with life, finances and the country’s direction there are indicators of negative bias against Asians.

The focus of their study was to answer the question of whether Asian minority group might experience covert racial discrimination on Airbnb. Using a sample of Airbnb listings

²Amazon Mechanical Turk is a service that provides workers to do large scale repetitive IT-related tasks

from a racially and economically balanced neighborhood they followed a methodology similar to the previous study in New York City. In addition to Airbnb listing variables, a host's race was determined from inspection of the listing photo. However, only two alternatives were used in their dataset: Asian or White. Any hosts that were not either Asian or White (or could not be determined) were excluded. Additionally, the number of observations were very small, 101, and as a result they needed to manipulate the form of the independent variables in order to compensate for strong right-skewed distributions.

Basing their listing price on a week's stay vs a single night to reduce price variability, they found a statistically significant price difference of 20% less per week for Asian hosts vs White hosts with similar listings. As with the New York study, the authors cannot say that the cause of this price difference is discrimination or whether it is based on differing price strategies or goals for Asian vs White hosts. They also acknowledge that an extension of their study would be to control for the quality of the listing location, something that was attempted in our study of the San Francisco market.

Another recent paper, also by [Edelman et al. \(2015\)](#) once again looked at the the Airbnb market for potential racial discrimination but this time focusing on the supply side. In contrast to their earlier study where they looked at discrimination aimed at Airbnb hosts, manifested in a lower listing price, presumably due to lack of demand at a price equivalent to non-minority listings - here they examined the rates of Airbnb host acceptance of a request from Black and White potential guests. Using 6400 Airbnb accounts in five different cities they sent guest requests for a room using identical data except for a distinctly White or Black name. Their results show that Blacks received a positive response 42% of the time compared to roughly 50% for White guests. This translates into a 16% difference between the two groups, consistent with the racial gap found in several other markets including taxicab rides, labor markets and on-line lending (earlier cited Prospero study). This recent work adds evidence to a growing structure and operation of P2P commerce markets.

2 Data Sources and Structure

We focus on the market for Airbnb listings in San Francisco. Our data set consists of cross sectional data that was available from [Inside Airbnb](#), a data collection project independent of Airbnb that compiles information of Airbnb listings for public use. The raw dataset was comprised of approximately 6,000 listings as of September 2, 2015. Since we utilized several guest rating categories as explanatory variables, we applied adjustments to the complete dataset to increase the accuracy of our sampling process and regression analysis. First, we restricted the listings to hosts with a verified identity (per Airbnb processes) and profile picture - a picture being essential to determining the host's gender and race. Second, to ensure that the user ratings of the Airbnb rentals and hosts were as reliable as possible, we only included listings with a minimum of 5 reviews (since we used several of the Airbnb review values as explanatory variables). These restrictions reduced the Inside Airbnb data set to 2,772 listings belonging to 2,161 unique hosts. Approximately 85 percent of hosts have only one listing, with a small number of hosts having 10 or more listings. This distribution of listings per host closely matched an earlier study by the San Francisco Chronicle using Airbnb data from May 2014, [Said \(2014\)](#).

Since we needed to individually verify the race and gender of the host for each listing, we created a smaller sample data set from the full Inside Airbnb dataset. Because of resource limitations, we drew a proportionally weighted sample of 800 observations from the Inside Airbnb dataset based on the number of listings per neighborhood based on zip code in the full dataset. Listings were randomly selected without replacement to create a sample with the same weighted neighborhood listing composition as the complete dataset. This was done to ensure that our sample reflects the distribution of Airbnb listings across the various San Francisco neighborhoods.

For each listing, we manually viewed the Airbnb host’s online profile page and, based on the included picture of the host and biographical information, categorized the host’s gender, race, whether the host was a couple, and the host’s sexual orientation (gay or not), if it could be reasonably determined based on host’s biography. For interracial couple hosts where one partner was White, we categorized the listing under the race of the non-White host. Listings of interracial couples where both partners were non-White were uncommon and were removed from our sample. When the host’s characteristics were unidentifiable (e.g., the host’s photo did not include any people), the listing was removed from the sample. After this categorization process, we were left with 715 listings belonging to 588 unique hosts. The dummy variables female, Black, Asian, Hispanic, couple and Gay were created using the values we obtained from a direct observation of the host’s listing site.

To incorporate the quality of the neighborhood as a factor influencing the listing price, both in terms of customer desirability and in terms of owner costs that could be reflected in pricing, we used data on the price per square foot (cost) based on recent property sales in each neighborhood. Property costs/sq ft were obtained from [Trulia.com](https://www.trulia.com). This was done for all property types in each Airbnb listing neighborhood. See Airbnb neighborhood list in Appendix. We used the value of all property types and not just single family homes or condominiums to reflect the fact that Airbnb listings can include single family homes, condominiums and commercial buildings. For each neighborhood in our sample, we calculated a z-score to show how many standard deviations each neighborhood was above or below the mean average cost per square foot. This z-score was then used as an explanatory variable in our model for each listing/neighborhood. The variables used in our paper are specified in Tables 1 and 2.

The baseline race category was White; baseline gender category was male. Single was the baseline category for the couple dummy variable, and non-gay/not-specified for the gay dummy variable. For the dummy variables created from the Inside Airbnb data, Shared Room’ was the baseline room type (as opposed to Private Room or Whole Apartment) and non-superhost was the default for the Superhost variable. Using one of the possible race, gender, couple/single and gay categories as the base case or control eliminated the potential of perfect multicollinearity for those variables.

3 Model

3.1 Dependent Variable

We used log of price per unit listing as our dependent variable. This enabled us to relate the effects of our independent variables to percentage changes in the listing price. This is significantly more revealing since a listing price will vary due to the number of rooms, bathrooms, etc and using log of price allows us to see the effect of the race variables on listing price in percentage terms, vs absolute terms.

3.2 Independent Variables

We categorize our independent variables into four major categories, namely,

1. *Host features*: consisting of dichotomous variables on the host's race (White, Hispanic, Black, other); gender (male, female); whether the host was a couple, and the host's sexual orientation (gay or not).
2. *Rental listing features*: We expected the listing price to be directly affected by specific attributes of the rental which would objectively add or subtract to the perceived value of the unit. These are variables quantifying the number of bedrooms, bathrooms, whether rental is a whole apartment, has a private room and maximum number of guests accommodated. They all seemed likely to have a direct and positive impact on listing price.
3. *User Reviews*: The other Airbnb variables included in the model are categorical involving reviews from previous guests and therefore less reliable or subject to interpretation. However, it seems that they are still likely to have an impact on a potential guests evaluation of the listing and quality of their stay. Consequently user reviews for cleanliness, communication (responsiveness of the host to the guest) and overall value were included in the model. The final Airbnb variable included was the designation of Superhost by Airbnb. This captures a signal by Airbnb of a minimum level of quality of the host, and we expected it to influence a rental decision or act as a filter for potential guests. All the user review variables and the Superhost designation were expected to have positive effects on the listing price.
4. *Neighborhood value*: As foreshadowed, this variable is a z-score constructed to reflect the neighborhood value and captures how many standard deviations each neighborhood was above or below the mean average cost per square foot. This can proxy for customer desirability or demand and owner costs. The listing price should be influenced by the value of the neighborhood both on the renters side, such that a nicer neighborhood would be more valuable to the renter, and from the host's perspective, a more expensive neighborhood might incur higher costs, both, to purchase and maintain the property and therefore require a higher rental rate. Our final model includes both a neighborhood value variable and the square of the value.

We estimate the following specification:

$$\log(price_i) = \alpha + \delta H_i + \beta R_i + \tau U_i + \zeta N_i + \epsilon_i \quad (1)$$

The H vector contains information on the features of host i ; the R vector is composed of rental listing features; the U vector contains categorical information on user reviews and the N vector contains the z-score and the squared value of the z-score reflecting neighborhood values as described above. Table 3 presents the summary statistics for all the variables used in the estimation.

4 Results

Table 4 represents the estimation results from four different models were estimated.

1. Model 1: includes host’s biographical information as explanatory variables
2. Model 2: includes Model 1 and user reviews variables
3. Model 3: includes Model 2 and rental unit features
4. Model 4: includes Model 3 and neighborhood values

Our initial parsimonious models 1 and 2 do not perform well in explaining the variation in listing prices on Airbnb as reflected in their very low R-squared of 1% and 7%. Model 3 has an R-squared of 68%, which makes it a good predictive model for Airbnb listings in San Francisco. We suspect that this is strongly related to the suggested price feature on Airbnb’s website, which suggests a listing price to the host based on the listing information that the host enters about the rental unit features. Model 4 has an R-squared of approximately 73% which means that all our explanatory variables are able to explain 73% of the variation in listing prices on Airbnb. With the exception of some of the race or gender variables, all of our independent variables are statistically significant. For the discussion below, we will refer to Model 4.

Our expectation was that we would find a pricing differential in listing prices based on race. In fact, our model does predict a statistically significant 9.6% lower list price for Hispanic hosts and a 9.3% lower list price for Asians vs the control group of White, single, male hosts, while keeping all other explanatory factors constant. While the coefficient for Blacks indicated a 2.3% lower listing price (the sign being as expected) vs the White control group, it was not statistically significant. Our assumption is that this occurred due to the low number of observations of Black hosts in our data set (only 12) which accounted for a mere 1.6% of the observations compared to the San Francisco population percentage of 6.1%. While the Black race of the host was not a significant explanatory variable, it did predict a negative price effect as in [Edelman and Luca \(2014\)](#) on Airbnb listings. However, their study was focused specifically on potential discrimination against Blacks and only used two race categories, Black vs non-Black. Our model included more racial groups and other host features and was based on a sample size of 715 observations. In addition, we chose San Francisco as the focus of the study because it has very different racial demographic than New

York City (San Francisco is 6.1% Black; New York City is 15% Black) and also because we were interested in examining the possible effect on listing prices of other minority categories as well as possible gender effects.³

The signs of the coefficients for many of our race and gender variables reflected our expected results. We generally expected minorities (e.g. Asians and Hispanics) to price their listings below the market price of non-minority hosts, which was in fact the result we found. Our model's predicted effect on listing price of being an Asian host of -9.3% is consistent with but less in absolute values than that found by the recent Oakland Airbnb study, [Wang et al. \(2015\)](#). Their study found that Asian hosts earn on average 20% less than White hosts per week. However, there are significant differences between our data set and methodologies. The Oakland study used a very small sample size (100 observations), in a carefully selected neighborhood with a balanced Asian and White population and based listing price on a weekly cost vs a daily cost in an attempt to reduce price volatility. Our study used a larger neighborhood weighted sample of 715 observations representing the entire San Francisco market, daily prices, and included variables to reflect the quality of specific neighborhoods. Additionally, the Oakland study, like the New York City study, only looked at two race categories. The Oakland study only used Asian/White as a variable of interest without the additional race, gender, couple and Gay variables in our model.

Both gender and sexual orientation were found to be insignificant. Female hosts represented a fairly large number of total observations, 265 of 715 or 37% and gay host listings were only 35 or 4.8% of the sample set vs a recently reported gay population for San Francisco of 6.2%, [Newport and Gates \(2015\)](#). We speculated that a female host may have an effect on listing price as a result of female guests having a preference for staying with a female host. However, since we have no data to observe the demand side of the market, it is unknown what the distribution of potential guests is based on gender who are shopping Airbnb listings or actually renting a listing. Therefore, any potential gender effects are difficult to identify. While the number of gay host listings is close to being representative of the San Francisco demographics (4.8% vs 6.2%) it is still a small number of observations which may be the reason its insignificance.

The other Airbnb variables such as bedrooms, bathrooms, and being titled a "Superhost" proved to be consistent with our intuition and were found to positively affect the listing price. Only the overall review score value, although being statistically significant, had a sign opposite of what we expected, negative vs positive and is puzzling. This could be occurring because it is seen as too subjective to be reliable or that the term value may somehow have a low quality or cheap connotation. Since neither of the previous Airbnb studies - New York City or Oakland, included this variable in their model, we do not have a reference point against which to judge this result.

Our use of the neighborhood value as reflected by the constructed z-scores and a their squared value proved to be highly significant. Our intuition was that the quality or value of the neighborhood as expressed by real estate prices should have an impact on listing price. This could have an impact on price from both the buyer and sellers side. On the buyers side, a higher quality neighborhood should be seen as more desirable and therefore justify a higher rental price and on the host side, a higher-priced neighborhood probably

³Population Demographics for New York 2016 and 2015

translates into higher mortgage and maintenance costs and should be reflected in required rental rates. The positive sign for area normalized and negative sign for area normalized sq in Model 4 is consistent with a diminishing return to neighborhood value. It makes intuitive sense that at some point, the prices of the neighborhood real estate may have diminishing value to renters as the neighborhood may be more residential and further away from typical tourist destinations or may reflect prices that are more a function of property values than the specific amenities a guest may have access to. Based on the model and final regression, the inflection point for neighborhood value is the marginal effect of area normalized can be found by taking the first derivative of our model with respect to area normalized and setting it equal to zero: $0.163 - 2(0.036) \text{ Area normalized} = 0$ and $\text{max area normalized} = 2.26$.

Since area normalized is expressed in standard deviations, this translates to a value of \$1675/sq ft. The highest valued neighborhoods in our study were Presidio and Pacific Heights at \$1573/sq ft. Therefore, no neighborhood in our study was above the inflection point for diminishing marginal return on value.

The economic impact of a reduced listing price of Asian or Hispanic hosts can be significant. Assuming an average San Francisco Airbnb listing price of \$160, which assumes a mixture of single rooms and whole apartment offerings, an occupancy rate of 47% and an average price difference of -9.5% for Asian and Hispanic hosts, translates into a yearly revenue gap of \$2607.⁴ While variables such as occupancy rates and average room rates are volatile, this rough estimate on average still represents a significant economic impact.

While our results tend to support an assumption of potential racial discrimination, we must remember that we are trying to infer discrimination or intent from the demand side that manifests itself in a lower equilibrium price on the supply side for minority hosts. There are several other possible explanations for the price differences. As mentioned previously, the study by [Ikkala and Lampinen \(2015\)](#) presented alternative motivations of Airbnb hosts based on both social and economic factors. It is possible that minority hosts are pricing their listings low in order to increase the pool of interested guests and therefore maximize their occupancy rate and revenue - an exercise in profit maximization, or they could value the larger pool of potential guests as a way to be more selective in who they decide to rent to. On the other hand, White hosts may be pricing high in order to create a self-selection process of potential guests that better meet the characteristics of guests they wish to have in their residence and engage socially with higher income, professional, etc.

What is clear is that the internet P2P commerce market with its trend toward increased personalization is developing the same set of complex issues - from potential racial, gender, discrimination - to competitive pricing strategies that have always been evident in the brick and mortar market. Complicating the P2P commerce market, especially in an area such as the Airbnb short term room rental market, are the dual motivations of social experience and economic gain. As the market defined by the Airbnb space continues to grow at a nearly 100% rate per year and quickly approaches the size of the three largest hotel chains, it is clear that a better understanding of the dynamics of this market is needed, [Mudallal \(2015\)](#).

⁴[Airbnb vs Hotels: A Price Comparison.](#) and [Mashvisor.com](#)

4.1 Sensitivity analysis

We reviewed each Airbnb host’s profile to determine whether the host listed their social media accounts (Facebook, LinkedIn, and/or Google Plus). [Edelman and Luca \(2014\)](#) found that having a LinkedIn profile had a significant positive relationship with listing price in New York city. However, when designing our model, none of the social media platforms had a significant relationship with listing price and they were not jointly significant. Therefore, they were not included in our final model. It is our opinion that the rapidly rising expectation of people being active on social media may have turned this into a of course people are on social media and so negated any impact on listing price.

We also tried incorporating average monthly rent per neighborhood to represent neighborhood cost and desirability. However, when designing the regression model, the explanatory variable did not show a strong relationship with listing price.

We also postulated that female hosts might be seen as more valuable than males hosts due to stereotypes concerning cleanliness of the rooms, etc and the fact that female guests might have a preference to stay in a female-hosted listing. We re-estimated the model by interacting gender with cleanliness to examine this hypothesis but found that the effect was insignificant.

We had initially tested for model misspecification and performed the Regression equation specification error test (RESET). We learned that our model is missing some non-linear terms, which reinforced our decision to include the squared z-scores into the model. The negative coefficient sign of the squared z-score intuitively reflects the potential for diminishing marginal returns of the property value of each neighborhood, which confirms our economic intuition about the variable. We also ran a Breusch-Pagan/Cook-Weisberg test for heteroskedasticity and test was positive, suggesting that the error variance may depend on one or more explanatory variables and is not constant. Since we confirmed the presence of heteroskedasticity,, we report the reporting robust standard errors for our estimation output for each model.

5 Conclusions, limitations and future research

This paper attempts to examine the effect of host features such as race, gender, sexual orientation on Airbnb price listings in San Francisco. After controlling for the rental listing features, user reviews and neighborhood values, we find evidence of price differentials in listing prices based on race. Relative to the control group of White hosts, we find a 9.6% lower list price for Hispanic hosts and a 9.3% for Asians. Our sample of 715 listings closely mimics the demographic of San Francisco. We also find that rental listing features, user reviews and neighborhood value positively affect price listings.

Our study is based on a sample of 715 listings since host features like race, gender, sexual orientation (wherever observed), whether the host is a couple, were manually recorded by going to each host profile. We recognize that this is a smaller sample as compared to [Edelman and Luca \(2014\)](#), a study that did a similar analysis on New York City Airbnb listings for potential racial discrimination and outsource the data collection to Amazon Turk services. However, in contrast to [Edelman and Luca \(2014\)](#) and [Wang et al. \(2015\)](#), we improved

the model by including explanatory variables such as the gender of the host, different racial groups (White, Black, Hispanic, Asian and other), sexual orientation and value of the various neighborhoods.

A further extension would be to improve on this by creating a composite or collection of variables to capture other factors impacting an areas desirability. Examples of this include things such as crime rate, walkability ratings (now being used for real estate ratings), Yelp neighborhood restaurant reviews and others.⁵

We would also suggest looking at a sample data set with a larger number of gay listings to investigate the impact of gay hosts on listing price. Going into our study we speculated that there might be a positive price effect for gay host listings in San Francisco given the high percentage of gay population and the consequent attraction of the city as a travel destination for gays. However, we feel our sample did not have a large enough number of gay listings to be significant - although the sign for its coefficient was positive and intuitive.

Another improvement would be to examine price differences through time so that listing prices reflect both seasonality through the timing of major tourist, social events and holidays. Room prices are highly volatile based on the time of year and the patterns of tourism and vacations. This could have significant impact on the prices for different types of room offerings, different neighborhoods and hosts. For example, room demand in the Castro neighborhood of San Francisco during the Gay Rights Parade could be expected to increase significantly and the the price differential for a gay host listing could be positive.

Additionally, given the lack of visibility about the motivations of Airbnb hosts to price the way they do, it would be interesting to conduct interviews with minority hosts to try and understand how and why they price the way they do. In our study and previous studies, the best that can be done is to try to infer the motivations of hosts to reduce price due to lack of demand based on potential racial discrimination. However, as mentioned earlier, it is a distinct possibility that minority hosts are pricing lower either with an economic motive to maximize occupancy and revenue or in a social motive to maximize the pool of interested guests from which to pick those they are more comfortable with.

Another area for additional study would be to examine a very different demographic or geographical market. While we picked San Francisco partly because it had different demographic than New York City. San Francisco has a considerably higher Asian population and lower Black population. However, there are still significant similarities between the two markets that exist. Both, New York City and San Francisco have majority White populations and very urban, sophisticated markets. It would be useful to look at markets where a racial group such as Asians or Blacks are a majority. Examples would include Atlanta, GA which is 54% African American and Alhambra, CA which is 53% Asian.⁶ Of course, future research could always look at Airbnb listings in cities with diverse populations to examine the effect of race on price listings, even internationally.

As this market continues to expand, it is increasingly important to understand the dynamics that are in play and how different demographics - including race, gender and sexual orientation are affecting prices.

⁵[Walkscore.com](https://www.walkscore.com)

⁶[City of Atlanta demographics](#) and [City of Alhambra demographics](#)

6 Appendix

Table 1: Airbnb variables used

Name	Type	Description	Source
review scores cleanliness	categorical	1-5 star rating of cleanliness	Inside Airbnb
review scores communication	categorical	1-5 star rating of communication quality	Inside Airbnb
review scores value	categorical	1-5 star rating of overall value	Inside Airbnb
Private room	dummy	1=Private room, 0 otherwise	Inside Airbnb
whole apt	dummy	1=whole apt, 0 otherwise	Inside Airbnb
bedrooms	quantitative	number of bedrooms	Inside Airbnb
bathrooms	quantitative	number of bathrooms	Inside Airbnb
accommodates	quantitative	number of guests accommodated	Inside Airbnb
superhost	dummy	meets Airbnb quality std, 1=superhost, 0 otherwise	Inside Airbnb

Table 2: Variables created

Name	Type	Description	Source
female	dummy	1=female, 0 otherwise	Airbnb host's profile
Black	dummy	1=Black, 0 otherwise (white=control group)	Airbnb host's profile
Hispanic	dummy	1=Hispanic, 0 otherwise	Airbnb host's profile
Asian	dummy	1=Asian, 0 otherwise	Airbnb host's profile
other race	dummy	1=other_race, 0 otherwise	Airbnb host's profile
couple	dummy	1=couple, 0 otherwise (single=control)	Airbnb host's profile
Gay	dummy	1=Gay, 0 otherwise	Airbnb host's profile
z-score	quantitative	1=Neighborhood value, 0 otherwise	Trulia
z-score squared	quantitative	captures diminishing return of neighborhood value	Trulia

Table 3: Summary Statistics

	mean	sd	median	min	max
Cost					
price (\$)	203.87	145.87	165.00	40.00	1250.00
log(price)	5.15	0.56	5.11	3.69	7.13
Host characteristics					
female	0.37	0.48	0.00	0.00	1.00
White	0.74	0.43	0.00	0.00	1.00
Black	0.02	0.13	0.00	0.00	1.00
Hispanic	0.06	0.24	0.00	0.00	1.00
Asian	0.14	0.35	0.00	0.00	1.00
other race	0.04	0.19	0.00	0.00	1.00
couple	0.26	0.44	0.00	0.00	1.00
Gay	0.05	0.22	0.00	0.00	1.00
User reviews					
review scores cleanliness	9.49	0.68	10.00	4.00	10.00
review scores communication	9.83	0.41	10.00	8.00	10.00
review scores value	9.28	0.57	9.00	7.00	10.00
Rental features					
private room	0.39	0.49	0.00	0.00	1.00
whole apt	0.59	0.49	1.00	0.00	1.00
bedrooms	1.31	0.84	1.00	0.00	6.00
bathrooms	1.24	0.52	1.00	0.00	4.00
accommodates	3.21	2.03	2.00	1.00	15.00
superhost	0.23	0.42	0.00	0.00	1.00
Neighborhood value					
z-score	0.05	0.74	0.03	-1.89	1.94

Table 4: Estimation Results

Variables	(1) log(price)	(2) log(price)	(4) log(price)	(4) log(price)
Host Features				
female	0.02 (0.05)	-0.01 (0.05)	-0.01 (0.03)	-0.01 (0.03)
black	-0.33** (0.16)	-0.29** (0.14)	-0.03 (0.09)	-0.03 (0.08)
hispanic	-0.25*** (0.08)	-0.25*** (0.08)	-0.14*** (0.05)	-0.10** (0.04)
asian	-0.12* (0.07)	-0.16** (0.07)	-0.14*** (0.03)	-0.09*** (0.03)
other race	-0.09 (0.09)	-0.08 (0.09)	0.08 (0.07)	0.09 (0.07)
couple	0.03 (0.05)	-0.02 (0.05)	-0.07** (0.03)	-0.04 (0.03)
gay	-0.21** (0.10)	-0.18** (0.09)	0.02 (0.05)	0.03 (0.05)
User Reviews				
review scores cleanliness		0.13*** (0.04)	0.07*** (0.02)	0.07*** (0.02)
review scores communication		0.11* (0.06)	0.11*** (0.03)	0.10*** (0.03)
review scores value		-0.15*** (0.05)	-0.06** (0.03)	-0.08*** (0.03)
superhost		0.16*** (0.05)	0.14*** (0.03)	0.12*** (0.03)
Rental features				
private room			0.88*** (0.11)	0.86*** (0.11)
whole apt			1.32*** (0.11)	1.27*** (0.11)
bedrooms			0.16*** (0.02)	0.18*** (0.02)
bathrooms			0.14*** (0.03)	0.12*** (0.03)
accommodates			0.06*** (0.01)	0.06*** (0.01)
Neighborhood value				
z-score				0.16*** (0.01)
z-score squared				-0.04*** (0.01)
Constant	5.18*** (0.04)	4.23*** (0.63)	2.30*** (0.37)	2.49*** (0.35)
Observations	715	715	715	715
R-squared	0.03	0.07	0.68	0.73

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1: Sample Host demographics vs. Market demographics

Compared with SF demographics, host sample contained significantly fewer minorities



Figure 2: Airbnb Listing price by Race

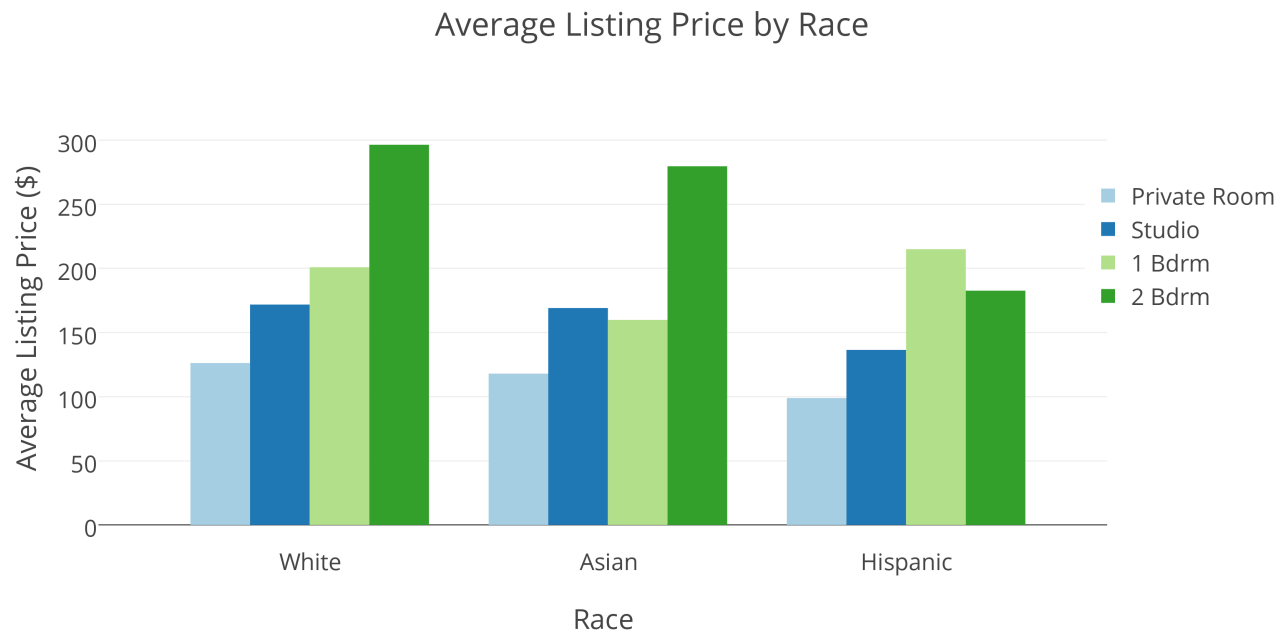


Figure 3: Market occupancy demographics

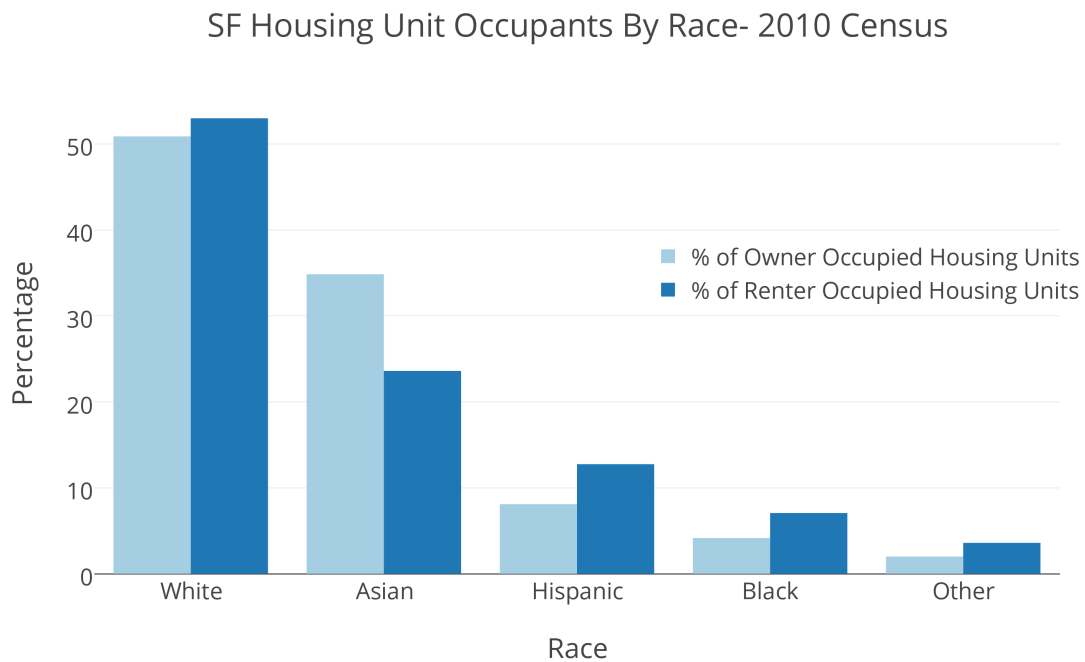


Figure 4: Airbnb sample distribution by neighborhood and race

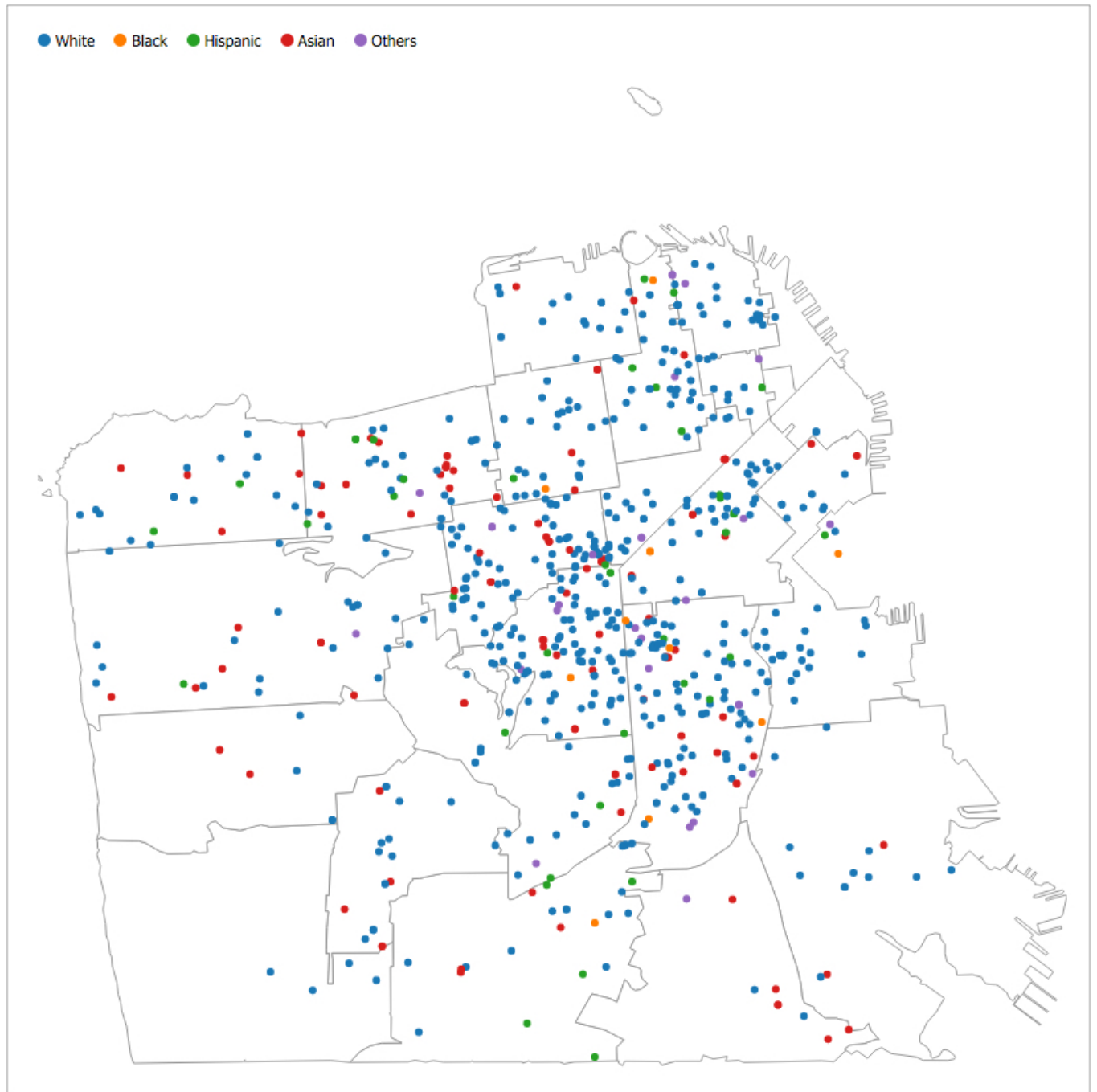


Figure 5: Airbnb neighborhood value heatmap

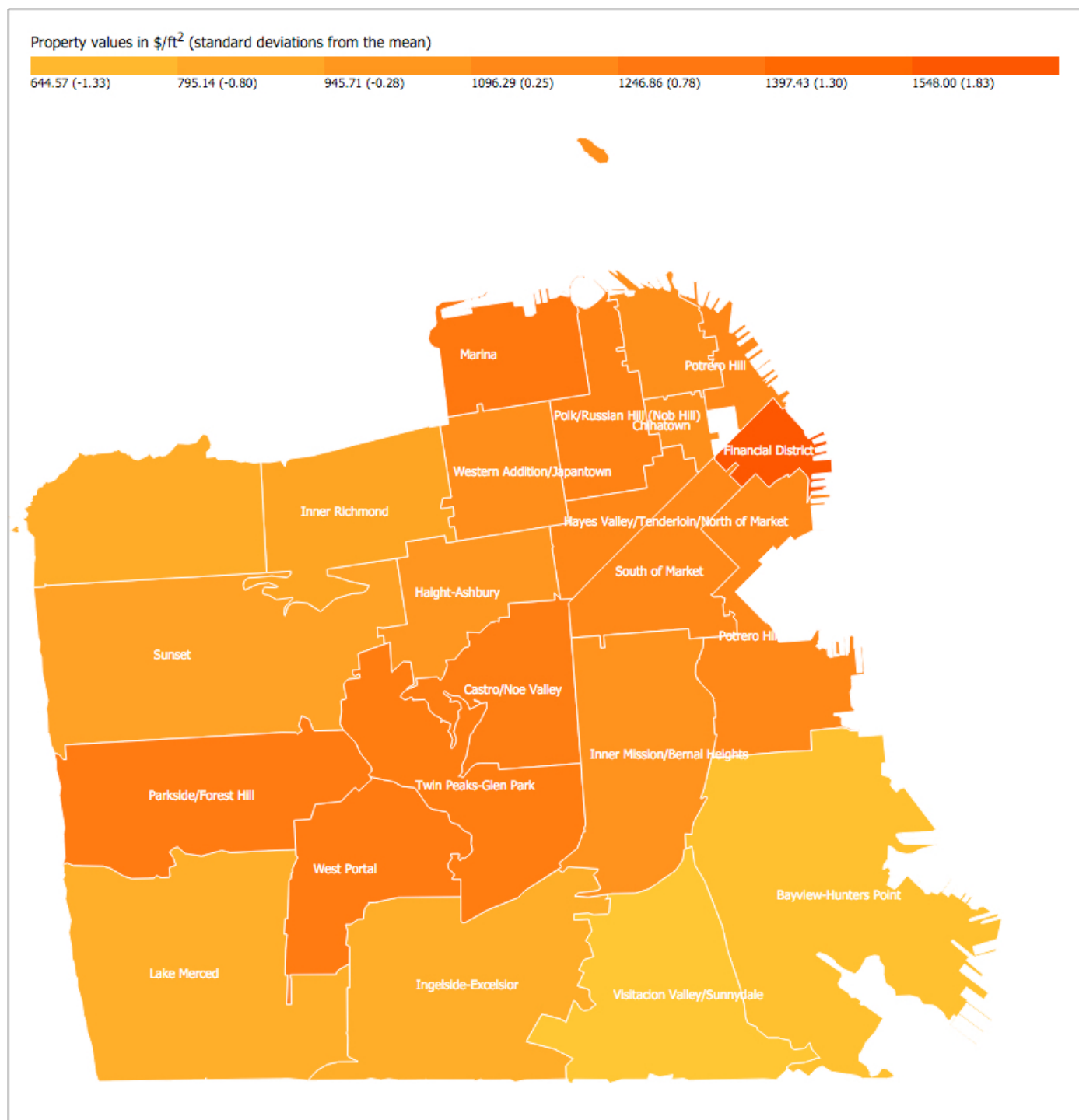


Figure 6: Airbnb listing example

The image shows a screenshot of an Airbnb listing for a "Just Remodeled 2 Bedroom Luxury Suite" in Los Angeles. The listing is priced at \$222 per night. The listing photo shows a modern living room with a large window, a red patio umbrella, and a black leather sofa. The host photo shows a man in a blue shirt. The listing description includes details about the suite, such as the number of bedrooms, bathrooms, and the location. The listing also includes a section for reviews from prior guests, showing a 4.9 rating and several positive reviews.

Listing Photo

Host Photo

Listing Description

Reviews from prior guests

References

- CaseyResearch. <https://www.caseyresearch.com/articles/sudden-rise-peer-peer-p2p-commerce>.
- Inside Airbnb. <http://insideairbnb.com/san-francisco/>.
- Statista, The Statistics Portal. <http://www.statista.com/statistics/239372/us-b2c-e-commerce-volume-since-2006/>.
- The possibilities of digital discrimination: Research on e-commerce, algorithms and big data - Journalist's Resource Journalist's Resource.
- Bayer, P., M. D. Casey, F. Ferreira, and R. McMillan (2012). Estimating Racial Price Differentials in the Housing Market. Technical report, National Bureau of Economic Research.
- Bertrand, M. and S. Mullainathan (2003). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Technical report, National Bureau of Economic Research.
- Constantinides, E. (2004). Influencing the online consumer's behavior: the Web experience. *Internet research* 14(2), 111–126.
- Doleac, J. L. and L. C. Stein (2013). The visible hand: Race and online market outcomes. *The Economic Journal* 123(572), F469–F492.
- Edelman, B. G. and M. Luca (2014). Digital discrimination: The case of airbnb. com. *Harvard Business School NOM Unit Working Paper* (14-054).
- Edelman, B. G., M. Luca, and D. Svirsky (2015). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *Harvard Business School NOM Unit Working Paper* (16-069).
- Hanson, A. and Z. Hawley (2011). Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics* 70(2), 99–114.
- Hanson, A., Z. Hawley, and A. Taylor (2011). Subtle discrimination in the rental housing market: Evidence from e-mail correspondence with landlords. *Journal of Housing Economics* 20(4), 276–284.
- Ihlanfeldt, K. and T. Mayock (2009). Price discrimination in the housing market. *Journal of Urban Economics* 66(2), 125–140.
- Ikkala, T. and A. Lampinen (2015). Monetizing network hospitality: Hospitality and sociability in the context of AirBnB. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1033–1044. ACM.
- Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp. com. *Com* (September 16, 2011). *Harvard Business School NOM Unit Working Paper* (12-016).

- Mikians, J., L. Gyarmati, V. Erramilli, and N. Laoutaris (2012). Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pp. 79–84. ACM.
- Morton, F. S., F. Zettelmeyer, and J. Silva-Risso (2003). Consumer information and discrimination: Does the internet affect the pricing of new cars to women and minorities? *Quantitative Marketing and Economics* 1(1), 65–92.
- Mudallal, Z. (2015). Airbnb will soon be booking more rooms than the worlds largest hotel chainse. <http://qz.com/329735/airbnb-will-soon-be-booking-more-rooms-than-the-worlds-largest-hotel-chains/>.
- Newport, F. and G. J. Gates (2015). San francisco metro area ranks highest in lgbt percentage. *Gallup*.
- Pope, D. G. and J. R. Sydnor (2011). Whats in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human Resources* 46(1), 53–92.
- Said, C. (2014). Window into Airbnbs hidden impact on S.F. - San Francisco Chronicle. <http://www.sfchronicle.com/business/item/Window-into-Airbnb-s-hidden-impact-on-S-F-30110.php>.
- Todisco, M. (2015). Share and Share Alike? Considering Racial Discrimination in the Nascent Room-Sharing Economy. *Stanford Law Review Online* 67, 121.
- Turner, M. A. (2013). Housing discrimination against racial and ethnic minorities 2012: Executive summary.
- Wang, D., S. Xi, and J. Gilheany (2015). The Model Minority? Not on Airbnb. com: A Hedonic Pricing Model to Quantify Racial Bias against Asian Americans. *Technology Science*.
- Zettelmeyer, F., F. M. S. Morton, J. Silva-Risso, et al. (2001). Consumer Information and Price Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities? Technical report, Yale School of Management.