TYLER DAE DEVLIN | DANIEL KUNIN | JONG HA LEE | DANIEL XIANG

BITTER SUITE

HOW AIRBNB AFFECTS LOCAL RESIDENTS

Abstract

In the following report we investigated the influence of Airbnb on landlords across the country and especially in the Bay Area. We looked at prices of specific properties when being rented to local residents and also the dollar amount to be gained by instead renting through Airbnb. The results of the regression techniques suggested the presence of significant incentives for landlords to switch to Airbnb. However, we found that there was not enough evidence in either the provided datasets or supplementary data sets found online to conclude that landlords acted upon this incentive to evict their current residents.

Introduction

SINCE ITS FOUNDING less than a decade ago in 2008, Airbnb has completely transformed the market for short-term accommodation. Whereas running a traditional hotel is a large-scale and costly operation, anyone with a spare mattress and a little extra floor space can become a successful Airbnb host. The company's stated mission is to provide a "marketplace for people to list, discover, and book unique accommodations around the world." According to their self-reported figures, Airbnb appears to be delivering on their promise: 31% of guests in San Francisco wouldn't have visited the city for as long or at all if it weren't for Airbnb, and 79% say that Airbnb makes them more likely to return to the city [3].

It is crucial to note that Airbnb's non-host user base consists of tourists and non-local visitors. The company's enormous growth in recent years speaks to the fact that Airbnb does in fact provide a useful and highly valuable service for travelers. But it is somewhat less clear how this new service affects local residents of a city. One report by the *Los Angeles Times* describes how some apartment owners have evicted their tenants in order to convert their properties into Airbnb locations [4]. The rationale is that landlords stand to make a lot of money by charging nightly prices to tourists rather than monthly prices to long-term residents. Indeed, it can be argued that Airbnb's success as a business lies in the fact that it occupies a price niche in between long-term housing and traditional hotels (Fig. 1).

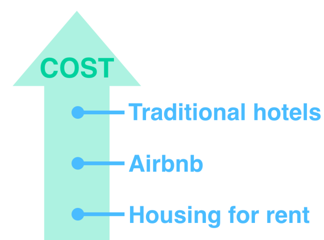

Fig. 1: The cost of staying at an Airbnb lies between that of staying at a hotel and that of renting an apartment.

Although this has benefits for travelers who can take advantage of Airbnb's lower prices, it also creates an incentive for landlords to stop providing affordable housing for local residents and instead market their properties to tourists who are willing to pay more. The *L.A. Times* article provides evidence that such an incentive exists in at least a few cases, but in the absence of a more comprehensive data-driven analysis, there is no way to quantify whether this has the potential to be a problem at a larger scale.

Nontechnical Summary

Main Questions

Our analysis can be broken down into two main questions

1. *Does the presence of Airbnb in west coast cities encourage landlords to evict current tenants in order to offer full time Airbnb locations?*

2. *If such an incentive exists, is there evidence to suggest that landlords are in fact evicting their tenants?*

Key Findings

We found that as a result of being able to charge higher prices for rent through Airbnb, there was indeed an incentive for landlords in the San Francisco area to prefer Airbnb to standard rent. The linear regression gave projected profits that landlords would gain by switching from traditional renting to renting via Airbnb. The expected earnings from switching to Airbnb were over $300. The key predictors we chose for the regression included numbers of rooms as well as median income of the surrounding area. The next step was to look into the other data sets to see if landlords actually acted upon this incentive, which led to the next part of our solution.

To discover possible real-world evidence of landlords taking advantage of this pricing discrepancy, we sought to explore whether landlords were unfairly evicting residents in order to offer Airbnb locations. This implied possibly a high number of locations which had both evictions as well as Airbnb listings shortly thereafter. Though we found 8,577 unique Airbnb locations (multiple listings may have the same location if within the same building/apartment) and 6,104 unique Eviction locations, analysis indicated there were only 13 locations that both were listed as an Airbnb property and also filed eviction notices within a year of their first posting on Airbnb, an insignificant number considering that evictions may not have been necessarily due to our proposed hypothesis: landlords unfairly displacing tenants to gain further profits from Airbnb. As a result, we could not find evidence of landlords evicting residents in order to offer Airbnb locations.

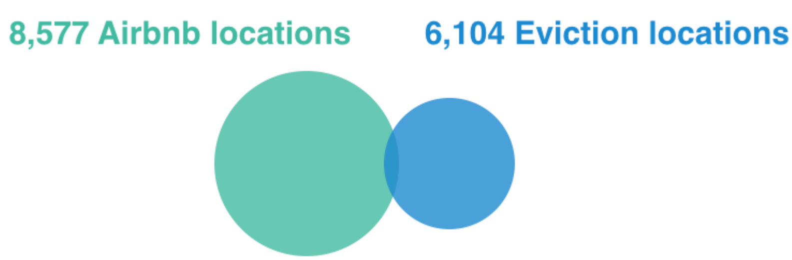

Fig. 2: We found that very few Airbnb locations in San Francisco from 2006 - 2017 had a previous eviction history.

Technical Summary

Exploring the Data

Perhaps the most important step in data analysis is to make sure you understand the data. Only then can you make an informed decision about what kinds of questions can reasonably be answered with the provided dataset. Thus, we began by discussing the meanings of the various columns in each dataset, resolving possible ambiguities by examining the data directly. We also noted how the different data sets might be integrated using common keys. For example, the variables `id` and `listing_id` could be used to link the `listings` and `calendar` datasets (through an appropriate join operation), while the `zipcode` variable was useful in tying the Airbnb listings to demographic and socioeconomic data.

We then proceeded to a more direct form of exploratory data analysis. During this process, our guiding principle was to "make lots of ugly graphs quickly." For example, we created:

- histograms to understand the distributions of single variables, paying special attention to skewness and outliers.
- scatter plots (including scatter plot matrices) to understand collinearity of variable pairs.
- box plots to understand how distributions split across levels of a categorical variable.

We were more interested in efficient exploration rather than effective exposition. Fig. 3 below shows just one of the many scatter plots we made in this phase of our project.

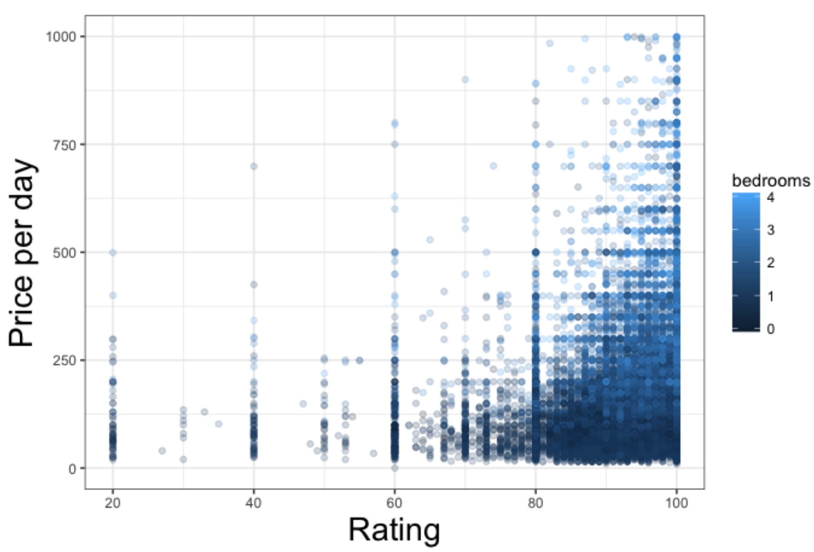

Fig. 3: The scatter plot to the left illustrates the weak positive association between the rating of a listing and its pricer per day, with each point colored by the number of bedrooms.

Fig. 4 is a histogram showing the price distribution of the Airbnb listings. Note that the histogram is truncated at \$1000 even though about 1.3% of entries in the `listings` dataset were priced above

$1000. We found that such listings overwhelmingly corresponded to "unusual" special event offerings, such as wedding venues. Given that these listings constituted a small portion of the data set and were not in line with the typical use case of Airbnb's service, we excluded them from analysis.

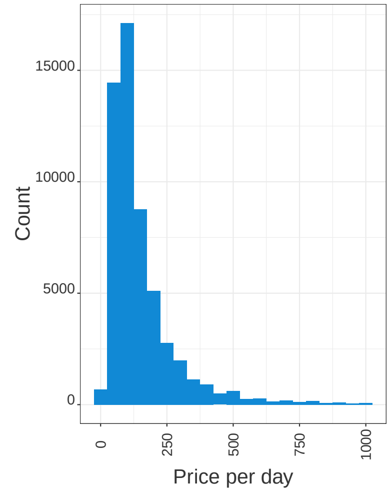

Fig. 4: A histogram of price per day of booking an Airbnb, thresholded at prices below $1000.

Data Cleaning

From the perspective of Hadley Wickham's tidy data analysis framework [2], the provided datasets were already in fairly good shape – rows corresponded to observations while columns represented variables. This does not mean, however, that there was no cleaning work to be done. Missing values needed to be properly encoded, for example. In the `listings` data, the entries of the `price` column were strings containing both dollar signs and commas; these special characters needed to be removed and the resulting strings needed to be converted to integers. The `zipcode` column in the same dataset was a complete mess. Some zip codes used the 5-digit format whereas others used the 9-digit format. Sometimes zip codes included decimal points, e.g. the code 94131 was represented as 94131.0. Finally, there were some zip codes that did not even correspond to west coast towns.

These are all factors that could potentially lead to significant errors in downstream analyses, so we were sure to address them early on. The `dplyr` package in `R` was extremely useful in this respect. The full details of our entire cleaning procedure is documented in our attached code.

The Model

Suppose you are a landlord and are considering whether to convert your property to an Airbnb rather than continue to rent to your current tenants. In order to make this decision, you need to estimate how much money you can expect to make in a month using Airbnb. If this amount is larger than the rent you currently charge, you have an incentive to make the switch. Thus we define the *potential monthly earnings* P of the property $\mathbf{h}$ as the revenue you can earn by converting your property into a full-time Airbnb listing. Naturally, P is a function of $\mathbf{h}$.

$$P(\mathbf{h}) \doteq \text{Potential Monthly Earning (PME) of property } \mathbf{h}.$$

We represent a household by a feature vector $\mathbf{h} \in \mathbb{R}^4$, whose entries include

- the constant term 1 (for the bias term),

- the median income of zipcode in which the property is located,
- the number of bedrooms of the property, and
- the number of bathrooms respectively.

Using the `listings` dataset, we ran a linear regression on these variables in order to predict the potential monthly earnings of a full-time Airbnb location. Thus we assume the PME as a function of the features is of the form

$$P(\mathbf{h}) = \mathbf{w}^T \mathbf{h} + \varepsilon$$

where $\mathbf{w} \in \mathbb{R}^4$ is the coefficients vector and the random vector ε is the error. We used the standard ordinary sum-of-squares loss function,

$$L(\mathbf{w}) = \|P(\mathbf{h}) - \mathbf{w}^T \mathbf{h}\|^2.$$

We opted not to include ℓ^1 or ℓ^2 regularization terms since the number of variables compared to the number of samples was small, as we only used 3 predictors [1].

Minimizing the sum of squared errors on the `listings` dataset yields a set of estimated coefficients $\hat{\mathbf{w}}$. The resulting coefficients of the linear regression are shown in the table below.

Linear Regression Results				
Predictor	Estimate	Std. Error	t-value	$\Pr(> \|t\|)$
(Intercept)	$-8.54 \cdot 10^2$	$3.58 \cdot 10^1$	-23.90	$< 2 \cdot 10^{-16}$
Bedrooms	$1.63 \cdot 10^3$	$1.49 \cdot 10^1$	109.16	$< 2 \cdot 10^{-16}$
Bathrooms	$9.05 \cdot 10^2$	$2.02 \cdot 10^1$	44.74	$< 2 \cdot 10^{-16}$
Median Income	$1.78 \cdot 10^{-2}$	$4.46 \cdot 10^{-4}$	39.99	$< 2 \cdot 10^{-16}$

The p-values in the fourth column suggest that all three features are statistically significant. The F-statistic (not shown) allows us to reject the null hypothesis that all coefficients are identically zero. The positive coefficients make sense in the context of the problem: more bedrooms and bathrooms result in higher PME, and properties in wealthier neighborhoods can charge higher nightly prices.

We then used these estimated weights to predict the monthly rent of a sample of property data obtained from the Zillow website [6]. The predicted monthly rent was compared to the actual monthly rent listed on Zillow and we found that landlords could indeed make significantly more money by solely renting their property through Airbnb. According to our model, we found that they could expect to earn an average of $368.30 more each month by switching from traditional renting to renting through Airbnb.

This finding is highly significant in that it suggest that *there is in fact a sizable incentive for landlords to convert their rental properties to full-time Airbnb listings*.

Evictions in San Francisco

Having confirmed that there are incentives for landlords to switch from a classic renting scheme to becoming Airbnb hosts, we attempted to find evidence of this trend in data. Using a dataset of eviction notices filed in San Francisco from 1999 to 2017 [1] and the given "listings" dataset of Airbnb properties we sought to find the number of San Francisco Airbnb properties which had a history of eviction. If landlords are unfairly evicting their tenants to earn more profits on Airbnb, then we would find a large number of Airbnb properties with eviction histories. Surprisingly, we found that very few Airbnb locations had a previous eviction history as summarized in the following table of summary statistics.

[1] The city of San Francisco provides hundreds of open datasets on the website DataSF. This dataset can be found at https://data.sfgov.org/Housing-and-Buildings/Eviction-Notices/5cei-gny5

	Airbnb	Eviction	Both	Both (<1 yr)
Locations	8,577	6,104	76	13
Total Units/ Notices	8,577	18,726	209	50

Fig. 5: We found 8,577 unique Airbnb locations and 6,104 unique eviction locations in San Francisco from 2006 - 2017. Only 76 locations were both an Airbnb property and an eviction location. Further only 13 of these locations filed eviction notices within a year of their first posting on Airbnb.

To obtain this result, we first filtered the "listings" dataset to only the Airbnb listings with San Francisco zip codes, and filtered the eviction dataset to unique geographic locations determined by (latitude, longitude). We then merged these datasets based on the latitude and longitude of the listings and eviction notices. In order to obtain reasonable results with this merge we rounded location data to four decimal places of a degree, or roughly 30 feet. Lastly, using the merged dataset of Airbnb listings with an eviction history we filtered to find only the locations that had evictions notices within a year previous to becoming Airbnb properties.

Interestingly, of the 76 Airbnb locations that had an eviction history 21 of them were owned by hosts that owned multiple Airbnb properties. And of the 13 locations with an eviction notice within a year of their first posting, 5 of them were owned by hosts that owned multiple properties. Both these percentages are significantly higher than the roughly 18 percent of Airbnb hosts that own multiple prop-

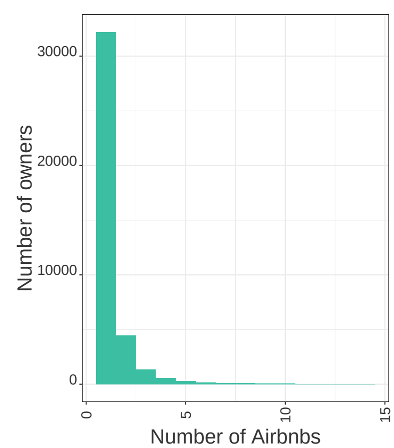

Fig. 6: A histogram counting number of owners who have 1 to 15 Airbnb properties.

erties in the general population, which can be seen in Fig. 6.

Conclusion

81% of guests chose Airbnb so they can "live like a local", but according to our analysis, they've begun to live in the San Francisco area *instead* of the locals. Our initial suspicions turned out to be somewhat correct, but we also didn't find evidence to support the hypothesis that landlords were evicting residents in order to turn their property into an Airbnb rental home. As shown in the Linear Regression, the potential monthly earnings from renting via Airbnb exceeded the earnings from the traditional renting process by $368.30 on average.

As a clear advantage to earning more profits by leasing out property on Airbnb rather than regular rent leasing, we hypothesized that some landlords may unfairly evict tenants to capitalize on the extra revenue opportunity. However, an analysis on comparing and intersecting the eviction data and the Airbnb property data in San Francisco determined there was insignificant evidence that landlords were evicting tenants unfairly.

As our next steps, we would like to further investigate hosts which own multiple properties and is a full-time Airbnb property leaser. We believe these hosts could possibly still be a factor in affecting housing markets for local residents and hiking prices to capitalize on the revenue difference.

References

[1] James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. New York. Springer, 2013.

[2] Wickham, Hadley. "Tidy Data." *The Journal of Statistical Software*, 59, 10 (2014).

[3] Airbnb. "Overview of the Airbnb Community in San Francisco." San Francisco, san-francisco.airbnbcitizen.com. Accessed 9 Sept. 2017.

[4] Poston, Ben. "L.A. apartment owners charged with allegedly evicting tenants, then renting their units via Airbnb." *Los Angeles Times*, 20 June 2016.

[5] (PSC), Michigan Population Studies Center. "Zip Code Characteristics: Mean and Median Household Income." Michigan Population Studies Center - Institute for Social Research, Accessed 9 Sept. 2017.

[6] Zillow, Inc. "Rental Listings in San Francisco CA - 1,642 Rentals." Zillow, www.zillow.com/homes/for_rent/San-Francisco-CA/. Accessed 9 Sept. 2017.