UC Berkeley Data Open

Team 16: Sam Dixon, Jiazhong Mei, Chen Meng, Ari Pickar

July 9, 2019

1 Central Question

We entered into the competition with the hypothesis that there are two types of Airbnb hosts. The first class of hosts uses the service casually, subletting rooms when they're not in town or hosting travelers who just need a couch or air mattress to crash on. The other class are much more savvy, using the service to manage one or more rental properties as a hotel manager or landlord would, responding rapidly to seasonal changes in demand and other larger market forces.

We chose to attempt to differentiate these two classes of users by probing one metric of market savviness: price variability. We then used other information about the listings (average price, location, local real estate market value, dates of availability) and the listers themselves to explore what kind of properties each type of host is listing on the Airbnb market.

Airbnb has severely disrupted both the rental and hospitality industries, and gaining a better understanding of the various ways people are using the service will be essential to determining the future of the company.

2 Datasets and Tools Used

We made the most use of the calendar dataset, which gives the time series prices of each listing along with dates of availability. We joined this table with the listings and real estate tables when exploring the properties of these listings. All analysis was done in Python, making extensive use of the pandas package for data wrangling.

3 Results

3.1 Variability

The central metric to this analysis was the list price variability, defined simply as the standard deviation of the nightly list price over the length of the listing. Of the 48,035 listings with available time series pricing data, 16,466 (34.3%) had no change in price over the length of the listing (see Fig. 1). The remaining listings did change their prices, some quite significantly. Fig. 2 shows a histogram of the list price variability, normalized by the average nightly list price.

3.2 Hosts with Multiple Properties

We hypothesize that hosts using Airbnb to manage several properties are savvier, i.e. more likely to vary their list prices over the course of a listing. This hypothesis is supported by the data. In Fig. 3, we show the average list price volatility for hosts in different property number brackets. Hosts that list a single property tend to have much lower price volatility than hosts with dozens of properties. The trend in the means of these distributions remain similar regardless of whether or not we exclude the data of listings that have no change in price. We can also note that all hosts with more than 40 listings made changes to the prices of all of their listings.

3.3 Timing and Magnitude of Price Changes

The majority of price changes happened in a regular manner. As we can see from Fig. 4, there was a distinctive pattern, with a large number of price changes made on on Thursdays and Saturdays, as compared to the rest of the week. In addition, the magnitude of price changes increased significantly

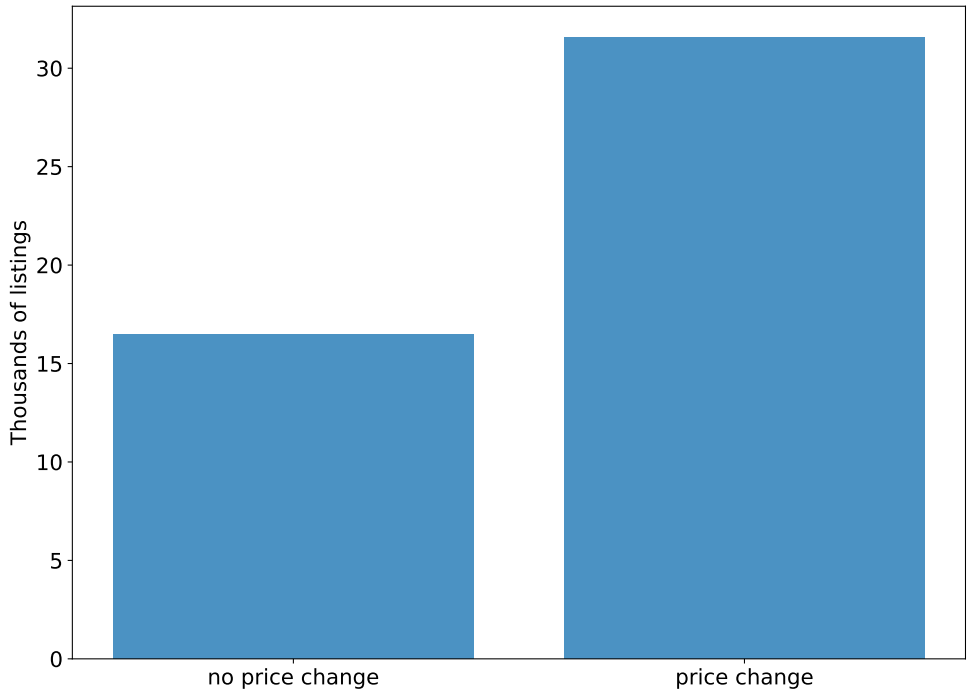

Figure 1: Listings with and without price changes over the length of the listing. 34.3% had no change in price over the length of the listing.

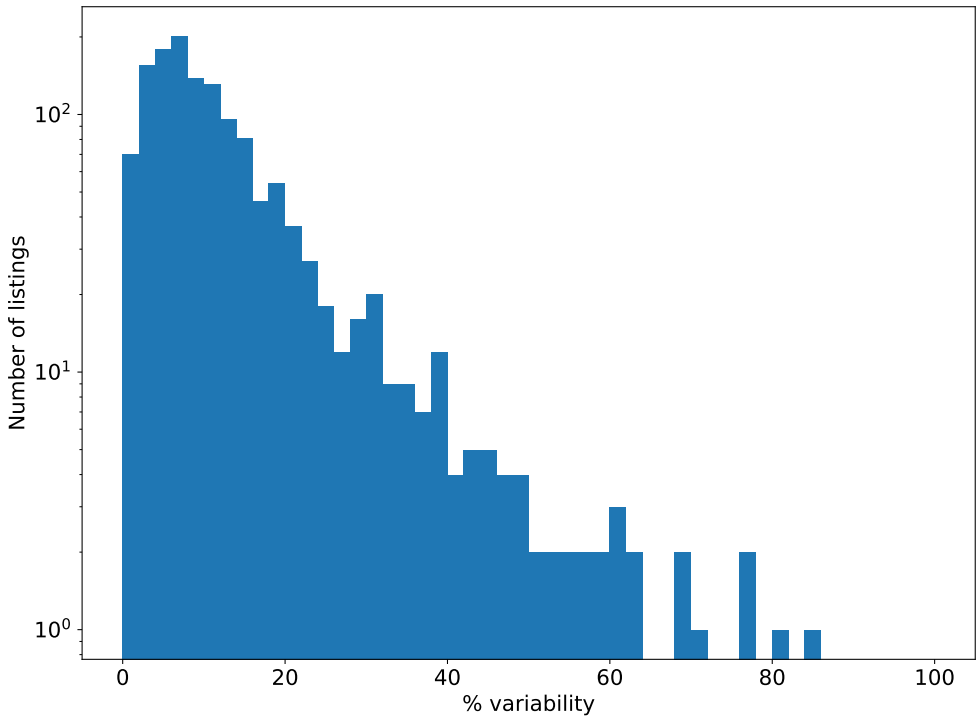

Figure 2: List price variability as percentage of average list price for listings that changed their prices over the length of the listing. The average normalized list price volatility is 12.4% with a median value of 9.1%

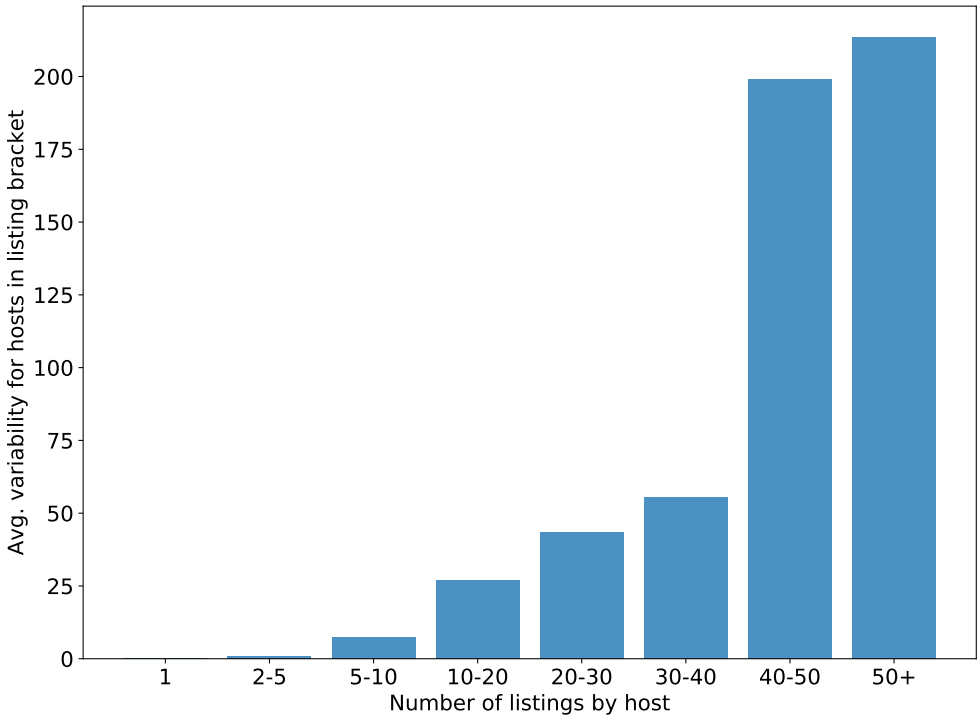

Figure 3: Average list price variability for users with differing numbers of listed properties. The clear trend here is that users with multiple property listings have much more variability in their prices. All hosts with more than 40 listed properties change the prices of each listing at some point.

on Thursdays and Saturdays, averaging a price decrease of $-28.38 on Thursdays, and an increase of $26.80 on Saturdays. This differed from our initial hypothesis, which thought that there would be an increase in price for Thursday and Friday, as those days are what are usually considered to be more likely to be rented. This led us to the conclusion that the majority of the price changes are targeted to entice consumers, and that the Airbnb market in its current state, has more people that are trying to rent out their house than those looking to rent. In addition, we found that the price of the average rental decreased by $-6.31 dollars week over week, meaning that rentals got cheaper, the longer they were on the market. This demonstrates the saturation of the market, as the number of available places increased, the necessity to stay competitive meant that the price of each place went down.

3.4 Listing Availability

We may intuit that casual users list their properties for much shorter times than superusers. We tested this by looking at the number of days the listings were available. Fig. 6 shows the distribution of number of days the listings in the dataset were available over the year.

There are interesting spikes at 1-2 day, 1 month, 3 month, 6 month, and 1 year points. These seem to be popular lengths of time to rent a room, most likely reflecting that hosts listing vacation homes (for 3/6 months) or apartments for a weekend (2 days).

We can also look at the list price variability of as a function of the number of available days. We have normalized the list price variability by the number of days the listing was available to account for the statistical effect of having more data points available to calculate the standard deviation. We can see small increases in the price variability around the 3 and 6 month marks. This would tell us that hosts listing a property for 3 or 6 months are changing their prices across this listing.

The variability in listings available for all 365 days of the year is quite small. One possible explanation for this could be that users listing for this length of time are looking to act as landlords, fixing rent per month for a lease period of one year.

Interestingly, there is much more variability in the price for hosts listing homes for less than

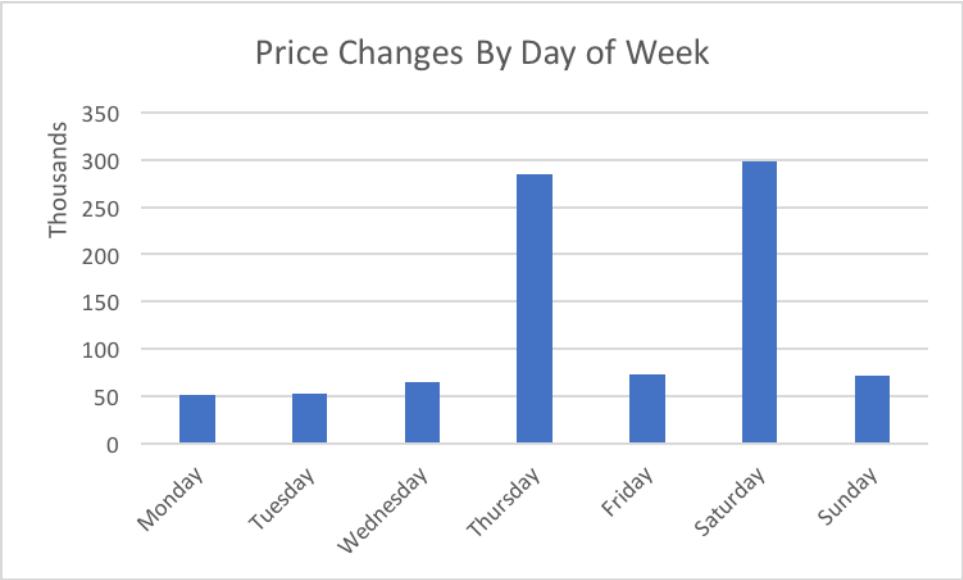

Figure 4: The number of thousands of price changes, broken down by day of week. The majority of price changes were concentrated on Thursdays and Saturdays, while the rest of the days have a relatively even number of price changes. This cyclic weekly nature shows deliberate effort on the part of listers to alter the market for their listings.

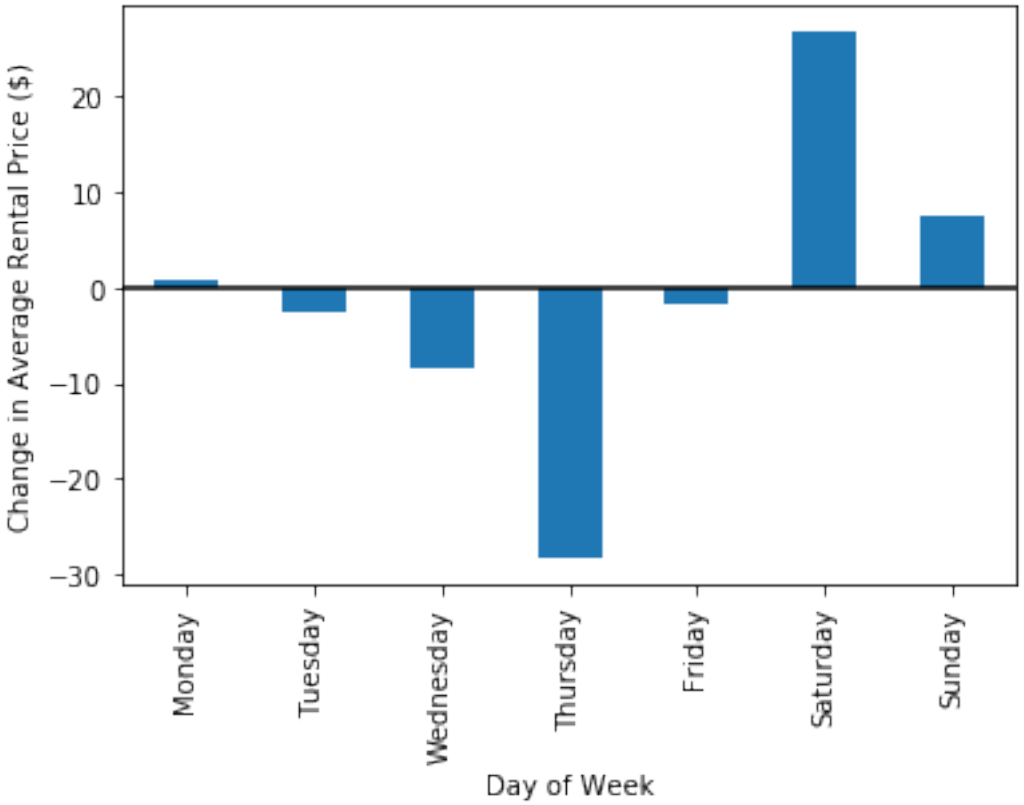

Figure 5: The average amount of change in rental price, by day of week. This chart also demonstrates the weekly cyclic nature on the part of those listing their houses to alter the market. The $-28.38 decrease on Thursday is almost fully made up by the $26.80 increase on Saturday, with the rest of the week demonstrating little average change in rental price.

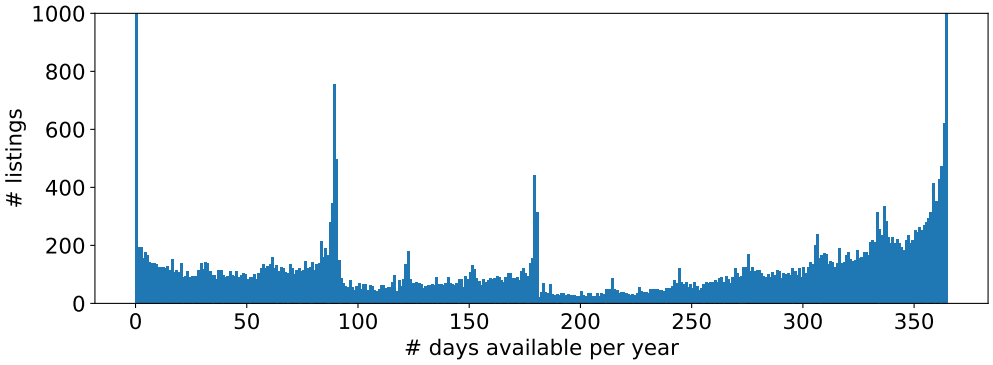

Figure 6: Number of days each listing was available per year. Notice the spikes at 3 and 6 months.

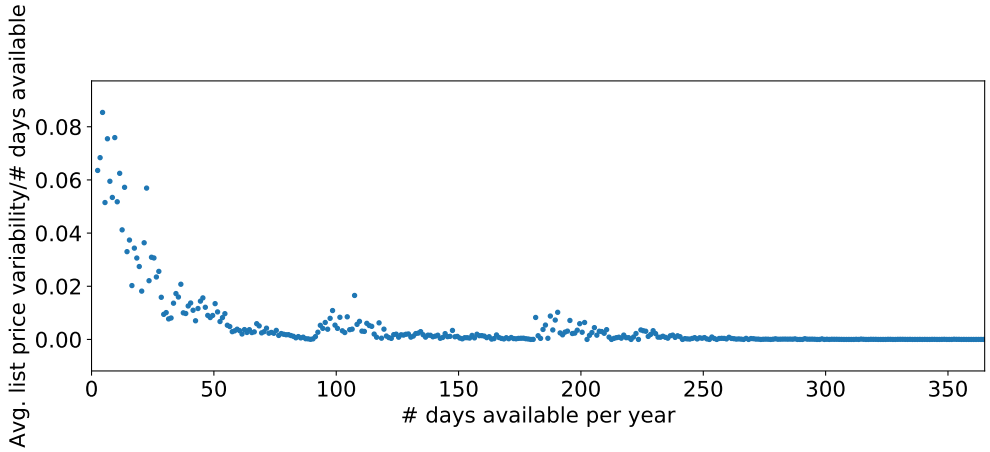

Figure 7: List price variability normalized by number of days the unit is available for rent. Note the increases in variability at the 3 and 6 month points

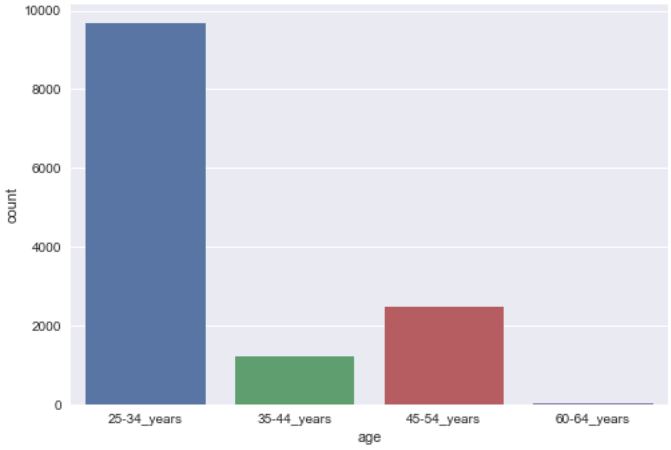

Figure 8: Number of properties listed by dominating age group.

60 days. This runs counter to our initial hypothesis, since it seems to indicate that very short-term listings have a more volatile price. More investigation into the cause of this is necessary, but preliminary explanations could include the impact of large events (e.g. music festivals, sporting events) on supply and demand of short-term rentals.

3.5 Listing Location and Demographics

We inspected the relationship between list price variability and the demographics of the city that the property is located. We represented the demographics of each city by its largest age group, largest income group, and its mean household income. Since the distribution of price variability is extremely right skewed and have some very large outliers, we instead use log price variability when examining its relationship with demographics.

We first focused on age. From Fig. 8, we see that most properties are listed in relatively young cities, where their largest age group is 25-34. Then, we relate to the distribution of log price variability of those properties for each dominating age group. Fig. 9 shows that 45-54 age group has the largest mean (30.38) and spread compared to other groups. It implies that the properties in cities with a dominating age group of 45-54 are making larger changes to their pricing.

We can interpret this from the point of view of hosts as well as tenants. For hosts under this age group, they tend to be more knowledgeable to market and demand because they are more likely to do this as a business or simply have more insights. Therefore, these hosts would feel more confident in making large changes to their listing prices to accommodate the variability in demand. On the other hand, since the majority of the residents in those cities are middle age people, this could imply high population mobility due to commuting work. And this could lead to higher variability in house rental demand, and thus its price.

Next, we focused on the city dominating income groups. Again, we counted the number of properties in each income group (see Fig. 10). There are only 4 properties in the 25000-34999 income group, so the bar of that category is not visible. It also shows that most properties are in the cities with dominating income between 25000-34999.

However, when we plotted the distribution of their price variability (see Fig. 11), we noticed that the properties in the lowest income group have the highest mean (31.76) and largest spread. This shows that the listing price of the properties in lower income cities tends to change more dramatically, while properties in higher income cities have a relatively more stable pricing.

To summarize, we saw that the variability of property prices is most significant among middle age, low income cities. This provides us some insight on the people both on the demand side and the supply side. In general, cities of middle ages and low incomes tend to have a very fluctuating demand and supply on Airbnb house rental. And the cause of such fluctuation is either incorporated into the living style of the residents or major events in those cities.

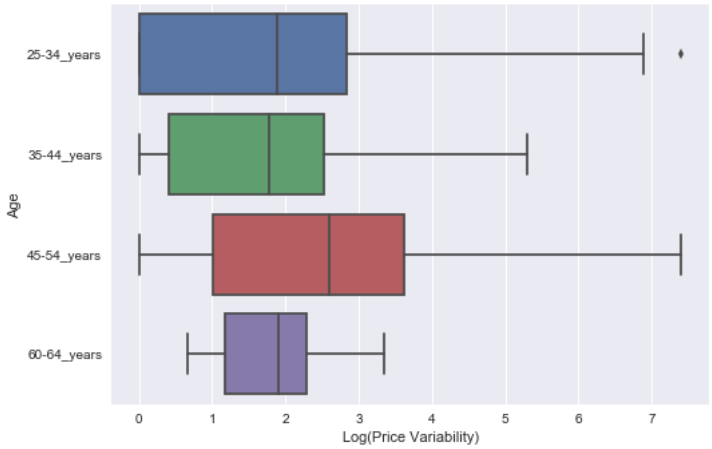

Figure 9: Distribution of log price variability by dominating age group.

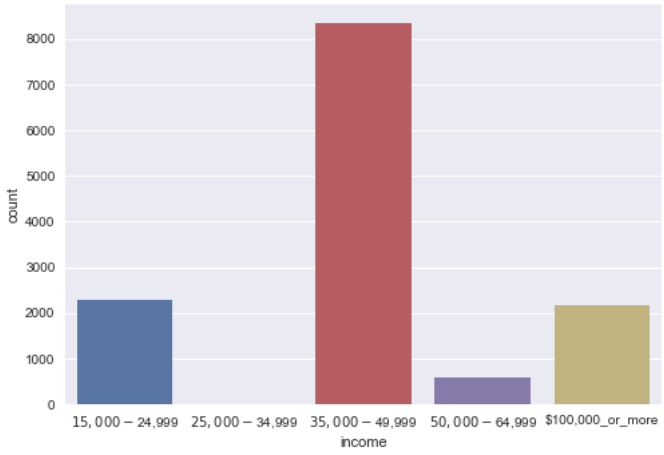

Figure 10: Number of properties listed by dominating income group.

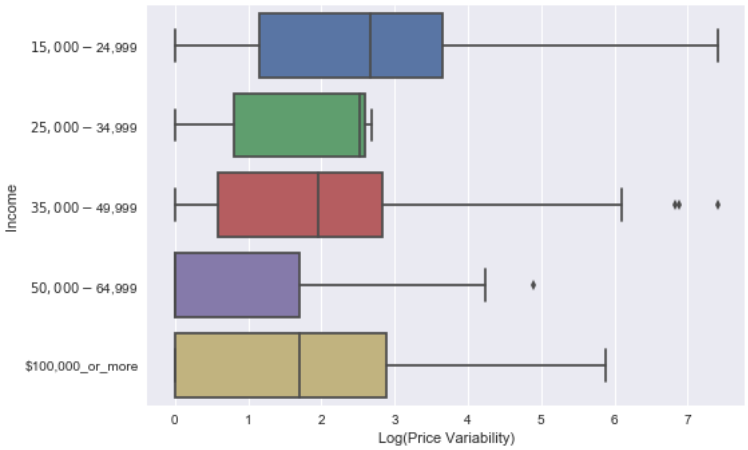

Figure 11: Distribution of log price variability by dominating income group.

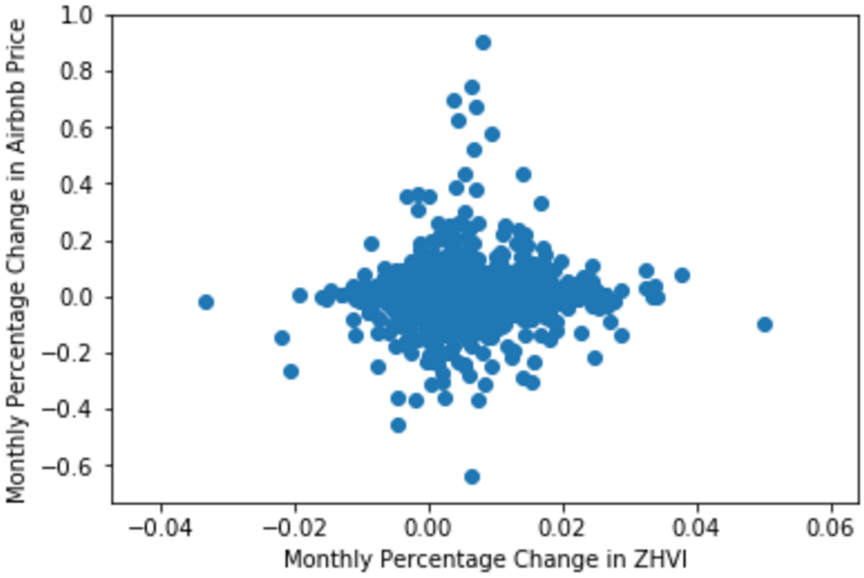

Figure 12: Relationship between monthly ZHVI change and monthly Airbnb listing price change

3.6 Correlation with Real Estate Price Change

We are also interested in examining the effects of real estate price change on the list price variability. Our hypothesis is that hosts change the prices of Airbnb listing depending on the bigger housing environment that they are located in.

To test for this, we first calculate the monthly change time series in the average Airbnb listing in each zip-code region and the corresponding monthly change time series in real estate index (specifically ZHVI). Next, we calculate the correlation between the changes to determine the relationship between two. Since most of the zip-code regions only have 12 data points, which are not enough to establish a valid statistical test, we decide to combine all the time series into a large dataset.

To make the graph more readable, we have eliminated outliers and restrict the y axis to -1 to 1. Surprisingly, the above graph (see Fig. 12) shows that there is not a noticeable correlation between changes in real estate index ZHVI and changes in Airbnb listing price. The calculated correlation is only 0.040922, which is not significant.

The result shows that the volatility in this market is more susceptible to short-term changes happening on a weekly basis as discussed in the previous sections; it is not much affected by the long-term changes in the housing market. Due to time limit, we did not have the time to do a similar analysis on the relationship between changes in Airbnb listing price and changes in ZRI. However, since rental houses are more a closer substitute for Airbnb listing price than home houses, we expect the correlation to be higher for the changes in ZRI. It would also be interesting to see the correlation between hotel pricing changes and Airbnb pricing changes.

4 Conclusions

We were able to begin investigating the differences in user behavior in setting prices for listings. We found:

- 34.3% of users make no change to the price of their listing for the entire time it is listed.
- There is a population of "superhosts" who use Airbnb to manage multiple properties, and these hosts are much more likely to vary the price of their listings
- "Superhosts" tend to list their properties for 3 or 6 months, as evidenced by increases in mean variability in units rented for this time frame
- Variability of property prices is most significant among middle age, low income cities

- Changes in daily rental price was cyclic, with the price decreasing every week on Thursday, increasing back up on Saturday, and staying relatively the same the rest of the week.
- Long-term changes in home market does not have strong effects on Airbnb listing price changes

These investigations give us insights into the different types of users Airbnb gives a platform to. In the coming years, this knowledge will be necessary in determining how Airbnb will navigate the growing intersection of the rental and hospitality sectors.