Hostel Takeover: The Effects of Airbnb Rentals on Real Estate Costs
Team 8: The Unsupervised Learners
Stuart Kirwan, Zach Labkovski, Jonathan Mager, Daniel Nagle, & Nik Patel

Introduction:

Housing affordability is a key issue with impacts on both individual and societal levels. Having access to affordable housing is important for an individual's financial stability as well as their health and well-being. High housing costs can force individuals to live in substandard or overcrowded housing which commonly lead to respiratory and stress-related illnesses. Additionally, high housing costs may lead to the displacement of long term residents or impact an employer's efforts to attract and retain workers, both of which impact community stability and economic growth. Among economists and policymakers, there is discussion as to whether Airbnb listings have contributed to the rise of housing costs. While the magnitude of Airbnb's impact is debated, this question is worth investigating in order to better inform policy decisions that affect housing affordability.

Problem Definition:

Although there have been studies purely on whether housing supply is affected by the growth of Airbnb in various markets, there has yet to be conclusive evidence presented as to whether Airbnb growth has had direct impacts on housing prices nationwide. As a result, we have chosen to study various large markets throughout the United States and pursue investigation into the effect of the number of Airbnb listings on median market value of single family homes, represented by Zillow's ZHVI (Zillow Home Value Index). The impacts of our studies, if they go as anticipated, may be overarching. Those who are interested in investing in real estate would wish to know whether their Airbnb properties cannot only generate sufficient income to pay off the acquisition expenses, but if the value of their property will also grow over time given the area. Prospective home buyers and current home owners can also use this information to forecast if the level of Airbnb activity in their neighborhood will affect the future value of their home. Moreover, if our analysis goes as expected, it could potentially force legislation at the local and federal government levels to limit Airbnb growth. Additionally, our interactive visualizations can provide insight on likely future variation in housing markets throughout the United States.

Literature Survey:

Most studies show that an increase in listings by either density or availability are also associated with an increase in housing prices as measured by selling price or rent.[2,5] There are some studies that use techniques such as interaction variables that suggest an excess in Airbnb listing density can actually drive down prices. A study using the Corelogic MLS platform of real estate data used these interaction terms to understand the pricing effects on zip codes during the COVID-19 pandemic[5]. In CA, prices dropped while prices increased in TX and NY, showing how geographic position is a key factor in a real estate analysis. One analysis of Austin, Texas's housing market uses geographically weighted regression to find that Airbnb listings are negatively correlated with an increase over time in land value[9], but did not properly account for differences in potentially confounding factors that differed throughout the city. Another study of Washington, DC found both in fixed-effects and pooled models that Airbnb listings are highly correlated with property price at a statistically significant level[15]; the fixed effects model accounted for unobserved time-invariant characteristics. A detailed study on London's housing market over a three-year

period found a significant increase in home value of £14.78 per square meter for every unit increase in Airbnb listing density. The study also held neighborhood features fixed such as income, employment, education, health, and crime rate[14]. A national approach found a positive correlation between Airbnb listings and higher house prices, with lesser effects in areas with higher owner-occupancy rate, further driving home the impact[1]. This positive connection between listing and home value was also found using an ordinary least squares regression on 80,000 listings in over 40 U.S. cities[8]. It was also seen that reliable public transportation had a positive influence, suggesting that we may want to consider listing's proximity to travelers transportation options. Another study found that proximity to a popular tourist destination did not affect the inflation of property value with high Airbnb usage[6]. Conversely, a study focusing on listings in Chicago among 40 other cities in the US found the impact the racial demographic of a neighborhood has on the Airbnb traffic. Asian and black hosts had significantly lower earnings when housing and neighborhood features were held fixed which proves that the demographic makeup of a neighborhood can also affect future housing market changes[12]. Another possible negative effect is derived from increased tourism. A study analyzing listings in Auckland, New Zealand along with a crime index rating produced a hedonic pricing model that showed a downturn in housing prices due to Airbnb density using an interaction term[4]. When considering Airbnb's prices, we have mentioned the home's value, which can be sourced from Zillow's "home value index." This metric responds quickly to changes in the real estate market and is available so that we can compare regional prices over various time periods[11]. A popular methodology for predicting the effect on market price is a regression model with Airbnb usage, demographics, and geographic data[2]. Another common data cleaning technique is to look at changes over years, not months, even though the data is available[13]. Additionally, a study has shown that city regulations on Airbnb significantly reduce short term rental listings and housing prices, thus providing additional evidence for the positive relationship between Airbnb listings and housing prices.[8] Attributing potential increases in market price to Airbnb usage could distract from other issues like gentrification and potential future decline in property value due to negative consequences[13]. One study controlled for income and education changes and found that this was an effective way to determine gentrification was not a driving factor[2]. The possible payoffs could be for cities that are debating banning Airbnb or throttling its usage to see the effect on a wide variety of other cities.

Proposed Method:

Although many of the approaches mentioned above were effective in their own rights, combining the more successful of the approaches can yield more valuable insights. To this point, we downloaded time series data for various metropolitan areas throughout the United States containing their Zillow Home Value Index (ZHVI) at monthly intervals from January of 2000 to February of 2023. Additionally, we have downloaded all available cities' data from Inside Airbnb, a website that webscrapes Airbnb listings for various cities both in the United States and throughout the world. With that data, we went through and calculated the number of active listings for each city at all corresponding dates to the Zillow data using the python pandas library to calculate those summary statistics. In order to do that, we had to define dates of first review and most recent review as de facto boundaries for when a listing is considered "active". We have also gone through and begun to incorporate demographic and economic features to make sure that we are accounting for the possibility of external features impacting home values, rather than attributing all changes to changes in number of Airbnb listings. Additionally, because of the lack of availability of monthly (or for that matter annual) data at a more granular city level, we downloaded CPI data (as an inflation/cost of living measure), unemployment rate data (as a general measure of economic conditions), and federal funds rates (strongly impact mortgage interest rates, which affect housing demand) from the

Federal Reserve Economic Data Page on the St. Louis Federal Reserve's website. With all our data downloaded, we proceeded to model the relationship using various regression-based methods and build an interactive Tableau dashboard that allows the user to view a map with Airbnb listings and ZHVI, along with various graphs from city selection. The modeling techniques and visualization methods we used are described below.

Modeling Techniques:
1. Multiple Linear Regression
    The primary aim of using multiple linear regression was to investigate the impact of Airbnb listings on ZHVI. Additionally, federal funds rate, CPI and unemployment rate were included in this analysis as independent variables to control for these relevant economic features. Within each of our city datasets, each independent variable was scaled to assist with the interpretation of coefficient estimates in the regression output. To perform scaling, the mean value of the column was subtracted from each value and then divided by the column's standard deviation. The multiple linear regression was then implemented by using the Ordinary Least Squares (OLS) method from Python's statsmodels module. Next, to check for multicollinearity the variance inflation factor (VIF) was analyzed for each independent variable within each city dataset and compared to the maximum of 10 or (1/(1-R-squared)). If the VIF exceeded this derived value, it was determined that multicollinearity was not an issue. This process was completed using the variance_inflation_factor function from Python's statsmodels module. Next, the regression estimates and p-values were analyzed to determine the individual impact of each independent variable on ZHVI. Finally, to evaluate the performance of the multiple linear regression, R-squared and MSE measures were used to examine goodness of fit and to check how close predicted ZHVI values were to the actual values.

2. First Difference Model
    One concern with building a basic multiple linear regression model is that there may be a correlation between number of Airbnb listings and local home values, but that relationship may not be linear. As an alternate approach to multiple linear regression, the first difference model was also implemented to allow for some time based considerations (with direct comparison to the previous period) while still accounting for all exogenous factors. The primary aim of this model was to investigate the relationship between the differences in ZHVI from one period to the next and the differences in number of active Airbnb listings, among other difference variables. To account for confounding variables, federal funds rate, CPI and unemployment rate were also included in the model. Again, the independent variables were scaled in a similar manner to the multiple regression model and the first difference model was implemented using the Ordinary Least Squares (OLS) method from Python's statsmodels module. To check for multicollinearity, the VIF was calculated for each independent variable within each city and analyzed in a similar manner to the multiple regression model. Next, the regression estimates and p-values were analyzed to determine the relationship between the differences in number of active Airbnb listings and the differences in ZHVI from one period to the next. Finally, to evaluate the performance of the first difference model, R-squared and MSE measures were used to examine goodness of fit and to check how close predicted values were to the actual values.

3. Time series model for control

While the regression techniques previously mentioned effectively investigate the relationship between Airbnb listings and ZHVI, baseline time series models were also implemented to investigate the trends observed in ZHVI. One important check that needs to be made before fitting our ZHVI data to a SARIMAX model is that the data is stationary, which means that the data process is consistent over the sample. This was checked by using the augmented Dickey-Fuller test which tests for stationarity by taking a lag of the data. Once this step was completed and the models fit stationarity, we are able to fit each city's ZHVI to a SARIMAX model, which is an seasonal autoregressive integrated moving average with exogenous factors. This is determined through a minimization of the Akaike Information Criterion (AIC) of various ARIMA parameters (p,d,q) and SX (P,D,Q,cycle length). In most cases the models selected used a SARIMAX(1,1,1)(1,1,1,12) with each of the following properties:

- AR(1) - current value depends on preceding value
- I(1) - first difference of values
- MA(1) - use a moving average of previous months value
- S(1,1,1) - seasonal versions of ARIMA values
- S(12) - length of the cycle which in our case is months in a year

This analysis was completed using statsmodels time series analysis method which can fit, optimize, and predict the ZHVI for each city's timeframe of data. As a check of each model's fit, a root mean square error (RMSE) was calculated. The range of RMSE for all cities was (646.31, 4836.15), which represents a very encouraging predictor for data that is in the magnitude of 100 thousand.

Data Visualization:

1. Tableau
   Using the results of our models, we created an interactive dashboard using Tableau that aims to allow users to create their own conclusions about the impact of Airbnb listing numbers on housing markets, as reflected by Zillow Home Value Index (ZHVI). With Tableau's automated calculations of latitude and longitude for various cities, we plotted all the cities in our dataset on a US map, with size of the cities' markers determined by number of Airbnb listings and color on a red-blue scale representing average ZHVI. In addition to the map, we also included two graphs that allow the user to explore trends in the studied cities; we have a dual-axis line chart of Airbnb listings and ZHVI over time, along with a dual-line chart of ZHVI compared with our time-series model-based predictions. Using filters and actions, when a city (or multiple cities) is selected on the map, the three graphs will be filtered to reflect the selected city/cities. This will make it as interactive as possible and allow users to look at the big picture as well as singular data points using only GUI. There is also a Dashboard-level filter that will allow the user to select a range of dates that impacts the range of data points for both the map (used to calculate average of ZHVI and number of Airbnb listings for color and size of the markers) and for the data points included in the graphs.

Experiments/Evaluation:

All of our ordinary least squares (OLS) models with scaled data were able to explain at or above 94.5% of the variation in each model, which signifies the strong linear fit that the number of listings as well as

national economic factors influence on Zillow Home Value Index (ZHVI). All coefficients are statistically significant which means that their values are statistically different from 0. The unemployment rate is inversely related to increases in home value, except for in the case of New York City, which could be related to the large difficulties in building homes in the area as well as skyrocketing rental prices. The magnitude of the 4 coefficients was led by the consumer price index in all cases with the number of listings being second highest, and the federal funds rate and unemployment rate alternating for the smallest magnitude between the 17 cities. While it is intuitive that the CPI will be a strong indication of housing prices, it is encouraging that the Airbnb listings are also large influencers in the prediction of home value as well.

The first difference models represent far worse explanations of variation of the data compared to OLS, with most difference models explaining around 25% of variation. These models do address some stationarity issues that can occur in the data that OLS is not able to capture. Listings are a significant predictor for all cities in this difference model, but what is most interesting is the variation among the economic rates between cities. For west coast cities, the federal funds rate (FFR) is insignificant, which means it is not useful in determining the ZHVI. However, all other cities have a highly significant inverse relationship with the federal funds rate coefficient. This means that in non-west coast cities an increase in FFR would lead to a decrease in overall housing prices and intuitively lower home values.

Finally, the time series model aims to model the ZHVI through a SARIMAX model which utilizes 4 significant properties of time series analysis: regression on past values, integrating to create covariance stationary processes, moving averages, and seasonality controls. Each cities' model is able to accurately reflect trends occurring with a root mean square error never returning larger than 4836.15, which suggests that our weakest model is only on average off by about $5,000. When we recall that most of the ZHVI is calculated in one hundred thousand or even millions of dollars, these models are very accurate.

To evaluate the performance regression based models, we primarily used R^2 which will measure the goodness of fit for our regression models. In other words, R^2 provided us with a metric that represents the proportion of variance in ZHVI that is explained by our independent variables (Airbnb listings and economic factors). The R^2 values ranged from 0.945 to 0.989 for the multiple linear regression models for all cities. For the first difference models, the R^2 values ranged from 0.188 to 0.440 for all cities. Additionally, the partial effects of Airbnb listings on ZHVI for both multiple linear regression and first difference regression were investigated to provide a measure of the isolated impact of Airbnb listings on ZHVI while controlling for the effects of the other variables. This estimate was significant for each city's multiple linear regression. For the first difference model, 16/17 of the regression estimates for Airbnb listings were statistically significant.

The interactive visualization tool was developed to emphasize usability and accessibility of information, regardless of a user's familiarity with real estate terminology. We used a Google form survey to assess how user friendly our dashboard, asking for a rating of 1 to 5 as a response. We had 26 responses and the mean and median rating is presented (mean, median) for each question below:

1) How understandable was the first graph, ZHVI and AirBnB listings over time? (3.654, 4)
2) How understandable was the second graph, Predicted versus Actual ZHVI (Time Series Model)? (3.692, 4)
3) How intuitive was it using the date filter? (4.077, 4)

4) How intuitive was the city map to use? (4.038, 4)
5) How useful did you find the information presented in the dashboard? (3.385, 3)
6) How intuitive was the dashboard to use as a whole? (4.154, 4)

Conclusions and Discussion:

Housing affordability has been shown to be a key issue with impacts on both individual and societal levels. The rise of short term rental services like Airbnb has sparked conversations among economists and policymakers regarding the effects of such services on local housing markets. The present study contributes a novel perspective to this debate by combining both interactive visualizations and robust statistical models to examine the relationship between Airbnb listings and home values as measured by the Zillow Home Value Index.

The results from the multiple linear regression analysis reveal a significant positive relationship between Airbnb listings and home values even when controlling for other relevant economic factors such as federal funds rate, CPI and unemployment rate. These findings are especially relevant for policymakers interested in improving housing affordability within their respective communities. Should policies regulating short term rental services be implemented, future studies should examine the effects these regulations have on home values. This research should consider leveraging statistical methods such as paired t-tests or CUSUM tests to examine the effect of limiting Airbnb listings on ZHVI. While the present study demonstrates a significant relationship between these two variables, the proposed future studies will provide key insight into whether there is a causal relationship. Home buyers may also opt to stray away from communities that would be more susceptible to AirBnB listings if they do not wish to be surrounded by tourists frequently.

The time-series SARIMAX model developed to forecast ZHVI for each city demonstrated strong performance for all cities. This robust forecasting model can be used by policymakers to estimate future home values in their community. This insight will provide policymakers with time to prevent (or at least to better prepare for) the potential negative consequences of increased housing costs such as the displacement of long term residents or the failure to attract and retain workers for local businesses. This tool is also helpful for prospective property purchases where buyers can ensure the value on their real estate or their homes will increase over time if they can expect an increase in ZHVI.

The interactive visualization tool illustrates the trends in Airbnb listings, ZHVI and predicted ZHVI over time for each city. This visualization tool can be especially useful for prospective homebuyers to understand current home values in an area, the presence of Airbnb listings, and their interaction. Not only is the presence of Airbnb listings significantly related to home values, rental properties often have a large impact on the culture of local communities and raise questions regarding safety. Future research should investigate the relationship between short term rental properties like Airbnb, crime and the perception of safety in local communities. Such research is especially relevant for policymakers and urban planners interested in attracting young professionals with families to their cities. While the current visualization tool does not address crime, it does illustrate the volume of Airbnb listings and provides a rough measure for how Airbnb listings might impact a prospective homebuyer's residential experience.

Contributions:
All team members have contributed a similar amount of effort.

Works Cited:

1. Barron, Kyle, et al. "The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb." *Marketing Science*, vol. 40, no. 1, 2021, pp. 23–47., https://doi.org/10.1287/mksc.2020.1227.

2. Barron, Kyle, et al. "The Sharing Economy and Housing Affordability: Evidence from Airbnb." *Social Science Research Network,* 2018., https://doi.org/10.1145/3219166.3219180

3. Benitez-Aurioles, Beatriz, & Tussyadiah, Iis. "What Airbnb does to the housing market." *Annals of Tourism Research,* vol. 90, 2020. https://doi.org/10.1016/j.annals.2020.103108

4. Cheung, Ka Shing & Yiu, Chung Yim. "The paradox of airbnb, crime and house prices: A reconciliation". *Tourism Economics*, 2022. https://doi.org/10.1177/13548166221102808

5. D'Lima, Walter, et al. "COVID-19 and housing market effects: Evidence from U.S. shutdown orders." *Real Estate Economics,* vol. 50, no. 2, 2022, pp. 303-339. https://doi.org/10.1111/1540-6229.12368.

6. Garcia-López, Miquel-Àngel, et al. "Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona." *Journal of Urban Economics,* vol. 119, 2020., https://doi.org/10.1016/j.jue.2020.103278

7. Horn, Keren, & Merante, Mark. "Is home sharing driving up rents? Evidence from Airbnb in Boston." *Journal of Housing Economics,* vol. 38, 2017, pp. 14-27. https://doi.org/10.1016/j.jhe.2017.08.002

8. Jiao, Junfeng & Bai, Shunhua. "An empirical analysis of Airbnb listings in forty American cities." *Cities,* vol. 99, 2020., https://doi.org/10.1016/j.cities.2020.102618

9. Jiao, Junfeng, et al. "Land Value Impacts of Airbnb Listings on Single-Family Homes in Austin, Texas, USA." *International Journal of Housing Markets and Analysis*, vol. 15, no. 2, 2021, pp. 392–410., https://doi.org/10.1108/ijhma-01-2021-0008.

10. Koster, Hans, et al. "Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles." *Journal of Urban Economics,* vol. 124, 2021. https://doi.org/10.1016/j.jue.2021.103356.

11. Kuchler, Theresa, et al. "Chapter 6 - Housing Market Expectations." *Handbook of Economic Expectations*. 2023, pp. 163-191., https://doi.org/10.1016/B978-0-12-822927-9.00013-6

12. Marchenko, Anya. "The impact of host race and gender on prices on Airbnb". *Journal of Housing Economics,* vol. 46, December 2019. https://doi.org/10.1016/j.jhe.2019.101635

13. Sheppard, Stephen, & Udell, Andrew. "Do Airbnb properties affect house prices?" *Williams College Department of Economics Working Papers,* vol. 3, 2016, pp 1-45.

14. Todd, James, et al. "Assessing the impacts of Airbnb listings on London housing prices". *Urban Analytics and City Science,* vol. 49, 2022, pp. 206-222. https://doi.org/10.1177/23998083211001836

15. Zou, Zhenpeng. "Examining the Impact of Short-Term Rentals on Housing Prices in Washington, DC: Implications for Housing Policy and Equity." *Housing Policy Debate*, vol. 30, no. 2, 2019, pp. 269–290., https://doi.org/10.1080/10511482.2019.1681016.