To Airbnb or Not to Be:
Analysis of Airbnb vs. Long-Term Rentals in Manhattan

MGT 6203 Project Final Report

Team 45

Xiaoling Zhuang; xzhuang34; xzhuang34@gatech.edu

Saikishore Kukunoor; skukunoor3; skukunoor3@gatech.edu

Grace Yu; dyu35; dyu35@gatech.edu

Robert Bennett; rbennnett67; rbennett67@gatech.edu

Andrew Li; ali315; ali315@gatech.edu

Github Repository: https://github.gatech.edu/MGT-6203-Spring-2023-Canvas/Team-45

Georgia Institute of Technology

Atlanta, GA

April 16th, 2023

Abstract
People often ponder when choosing between a traditional long-term or the Airbnb short-term rental model. But is there a correct choice? To answer this, we used Airbnb and long-term rental data in Manhattan and performed different methodologies: two-sample t-test, regression analysis, stepwise selection and lasso regression. We found that there is a statistical difference in rental prices between the two rental options. In general, long-term rental generates higher income than short-term ones. However, we also need to consider the individual characteristics of the property. For both rentals, bedrooms, bathrooms, and the physical aspects of the apartment (entire home/apt and square footage) were important features/predictors of rental price. The variables doorman and elevator were significant to Airbnb rental price, while having a washer and dryer was significant to long-term rental price. Through lasso regression, we were able to see that neighborhoods are a key driving variable for rental price. The neighborhoods were broken down into 3 categories: ones that generated the most revenue (ex. West Village), neutral revenue (ex. Civic Center), and the least revenue (ex. Harlem). We used the VIFs of the features in our datasets to check for multicollinearity and used lasso regression to address it. We concluded this piece by conveying our roadblocks, suggestions for future study, and key takeaways.
Keywords: long-term rental, Airbnb, investor, owner, data analysis

Introduction

Our inspiration for this project stems from the article "New York Now Has More Airbnb Listings Than Apartments for Rent." Airbnb, the online peer-to-peer platform for rentals, has grown exponentially over the years and has taken over the rental market. Airbnb hosts can create a steady revenue stream through the platform, and many have created a functional business model. In comparison to Airbnbs, which are usually rented out in short periods of time to leisure travelers, traditional long-term rentals are often rented out to residential tenants in periods of over 30 days. Airbnbs and long-term rentals both have benefits and costs in terms of revenue for property owners. Benefits of renting out Airbnbs include the ability to frequently adjust rental price based on demand, but long-term rentals provide a steadier income stream with less likelihood of occupancy. The purpose of this project is to provide insights comparing the revenue of Airbnbs vs. long-term rentals and to conclude which rental type will pay out a higher revenue while identifying key attributes and amenities that influence rental price.

Literature Review

We consulted several scholarly research papers to study how others have analyzed rental prices of Airbnbs and long-term rentals. An analysis of Los Angeles Airbnbs and long-term rental prices concluded that Airbnb may not be the profitable choice when considering occupancy rates compared to long-term rentals (Houghton, 2020). Houghton uses a linear and semilog pricing model and concludes that specification and occupancy rates lead to less revenue in Airbnb when compared to a long-term rental. However, these results may be location specific, and other attributes may draw a different conclusion. In looking at the New York rental market, Coles et al, looked at differences in short term rental revenue versus long term revenue based on usage patterns and neighborhoods. The findings were that short-term rentals do not appear to be as profitable as long-term rentals. However, short-term rentals are more profitable than long-term rentals in middle-income neighborhoods (Coles et al, 2017). There is also the consideration that Airbnb has driven up long-term rental prices in neighborhoods (Horn and Merante, 2017). Using a fixed effect model, Horn and Merante find that one standard deviation increase in Airbnb leads to 0.4% increase in long-term rental asking prices. Because of easy accessibility to the short-term market through Airbnb, many traditional long-term rentals are turning to Airbnb which causes the supply of long-term rentals to run low and raise rental prices (Horn and Merante, 2017). Even so, homeowners and investors are still investing in the short-term rental market due to the revenue potential. Our project focuses on comparing revenue of Airbnb and long-term rentals while considering key amenities that determine asking rental price.

Initial Hypothesis

The null hypothesis for our project is that there is no significant difference between the revenue generated by Airbnb and that of a long-term rental. The alternative hypothesis is that there is a statistical difference between the revenue generated by Airbnb and long-term rental. Our group predicts that we will reject the null hypothesis and that there is a statistically significant difference between the income from the two rental choices.

Overview of Data

1. **Understanding The Data**

We obtained the datasets: the dataset from OpenDataSoft has 18,909 rows of Airbnb rentals in Manhattan; the dataset from StreetEasy includes a sample of 3,539 long-term rentals in Manhattan. The datasets share some common variables, for example, number of bedrooms, number of bathrooms, has_washer_dryer, neighborhood, monthly_revenue, etc. Each dataset also contains unique variables specific to the rental type it belongs to. For instance, the Airbnb dataset has the variables Daily Price and Cleaning Fee; the Long-term Rental dataset has the variables Building Age and Floor Number.

2. **Data Cleaning/Transformation**

There are three final datasets. First, the unioned dataset is the unioned result of the shared variables from the two rental datasets, where we encoded a new variable called "Rental Type": the Rental_Type equaled One is the Airbnb datapoint; otherwise, it's the Long-Term rental. The second and third datasets are the cleaned Airbnb and Long-term rental datasets respectively. We retained the important variables from each dataset to perform individual analysis as needed.

3. **Exploratory Data Analysis**

In Figure 1, the distributions of prices for Airbnb and Long-Term listings are both right-skewed, with a higher mean for long-term listings. All Summary statistics for both data, including price, are displayed in Table 1.

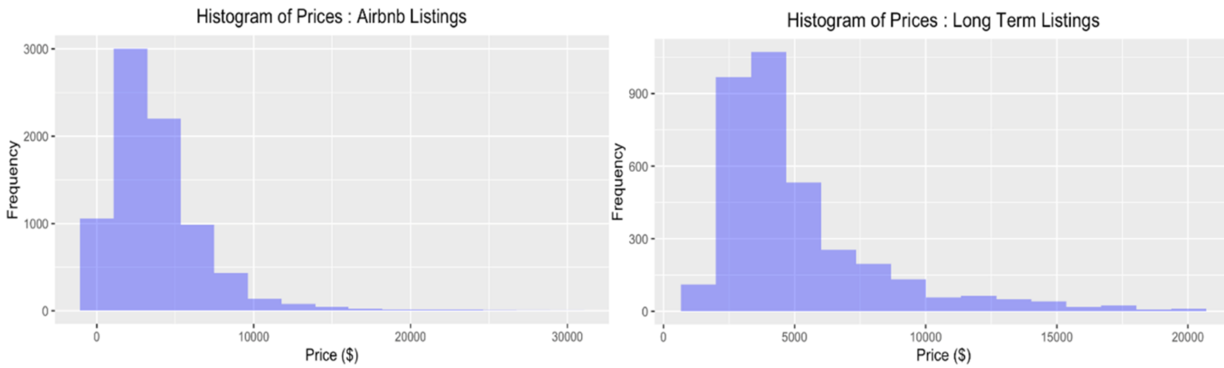


Figure 1 - Distribution of Prices for all Airbnb listings and all Long-Term listings (Histograms)

Note that all categorical variables in the data are binary, with values of 0 (for No) or 1 (for Yes). We first observe that for any such binary variable of this form, say 'C', its calculated mean (expectation) is insightful:

$$\mathbb{E}[C] = \sum_{c} c \cdot p(C = c) = 1 \cdot P(C = 1) + 0 \cdot P(C = 0) = \boxed{P(C = 1)}$$

as it represents the proportion for which the categorical variable is equal to "Yes". Thus the 'mean' calculations for variables in Table 1, rows 3-8, can be interpreted accordingly. Features such as door staff, patios, gyms, and dishwashers were more common in long-term listings. Laundry facilities and elevators were more common among Airbnb listings. The mean value for bedrooms was lower among Airbnb listings, due to a higher frequency of rental options such as studios, private rooms, and shared spaces. This occurs for bathrooms as well. Several economical Airbnb listings in the data may only offer shared bathrooms between separate units, which were categorized as 'num_bathrooms = 0'.

	mean		**std**		**min**		**max**	
rental_type	**Long-term**	**Airbnb**	**Long-term**	**Airbnb**	**Long-term**	**Airbnb**	**Long-term**	**Airbnb**
bedrooms	1.35	1.08	0.97	0.67	0.00	0.00	5.00	6.00
bathrooms	1.37	1.09	0.60	0.33	0.00	0.00	5.00	5.00
has_washer_dryer	0.16	0.41	-	-	0.00	0.00	1.00	1.00
has_doorman	0.28	0.15	-	-	0.00	0.00	1.00	1.00
has_elevator	0.29	0.41	-	-	0.00	0.00	1.00	1.00
has_dishwasher	0.19	0.00	-	-	0.00	0.00	1.00	1.00
has_patio	0.06	0.00	-	-	0.00	0.00	1.00	1.00
has_gym	0.17	0.09	-	-	0.00	0.00	1.00	1.00
price	5138.94	3840.48	3162.82	3136.65	1300.00	0.00	20000.00	29970.00

Table 1 - Summary Statistics for all Long-Term Listings and all Airbnb Data, before formation of the Union dataset.

For neighborhood-level geospatial analysis, a GeoJSON file was obtained from the New York City Department of City Planning's Census Metadata page. The file contained 262 polygons that cover all five boroughs of New York City. The 116,082 points that bound the polygons were consolidated and cross referenced with a list of neighborhoods and their respective boroughs for proper labeling. The data was filtered based on the label to include only Manhattan. The unioned dataset was used for the geospatial analysis, which follows in Figure 2:

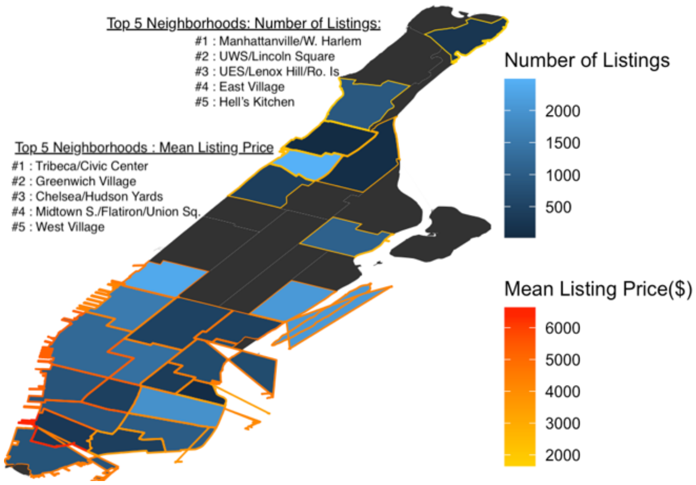


Figure 2 - Quantity and price of Manhattan listings- Union Dataset (both Airbnb and Long-Term listing data)

Top 5 neighborhoods with the most Airbnb Unit Vacancy in a 365 day period:		
Rank:	**Neighborhood:**	**Pct. Days Vacant:**
1	Flatiron District	44.88
2	Inwood	43.82
3	Harlem	40.97
4	Marble Hill	40.73
5	Little Italy	39.38
Top 5 neighborhoods with the lowest Airbnb Unit Vacancy in a 365 day period:		
Rank:	**Neighborhood:**	**Pct. Days Vacant:**
1	Morningside Heights	19.21
2	Roosevelt Island	25.7
3	Upper West Side	26.03
4	Stuyvesant Town	27.43
5	Kips Bay	27.71

Table 2 - Average vacancy percentage by neighborhood (Airbnb listings), top 5 most and least occupied by neighborhood.

Uptown neighborhoods, farthest from most tourist attractions and large corporate headquarters, were usually priced the lowest. One literature review component (Shokoohyar et al., 2020) emphasizes the importance of neighborhood selection in the rental strategy. Harlem has the most listings overall, but is one of the least expensive neighborhoods. Downtown neighborhoods are generally the most expensive.

Using the 365 Day availability column from the Airbnb data, the percentage of days rented is calculated for each listing. It's worth noting that some neighborhoods that have a high number of listings overall also exhibit high percentages of vacancy amongst its short-term listings. This could indicate saturation in a particular neighborhood. The issue of vacancy is potentially a major obstacle in assessing the efficacy of choosing short-term over long-term. For Airbnb to be more lucrative than a long-term rental, the revenue should exceed the monthly rental price for a comparable long-term rental in the area. The averages of all prices sorted by neighborhood are used. For Airbnb listings, the average monthly revenue estimate $\overline{MRE}$was calculated as follows:

$$\overline{MRE} = (NightlyRate) \cdot \left[1 - \frac{PercentDaysVacant}{100}\right] \cdot 365$$

Neighborhoods' average MREs generally did not exceed the average long-term listing prices for those neighborhoods. Little Italy is the only exception. Across all neighborhoods, the average vacancy would need to decrease by 27% for MRE's to meet or exceed the same neighborhood's average long-term monthly rental rate.

The correlation matrix for all variables in the unioned data is included in Figure 3, exclusive of neighborhoods. On average, an increase in the number of bedrooms is followed by an increase in the number of bathrooms. Some premium amenities such as dishwashers, door staff, patios, and gyms were less common among Airbnb data. If these premium amenities are present, they are more likely to occur together across all rental types. An assessment of the correlation between predictors is also a key prerequisite of models discussed in later sections. We did not immediately deem any combinations of predictors to be problematic based on the correlation matrix alone.

	Bedrooms	Bathrooms	W&D	Doorman	Elevator	Dishwasher	Patio	Gym	Rental Type	Price
Bedrooms	1	0.52	0.05	-0.01	-0.04	0.06	0.03	-0.01	-0.13	0.40
Bathrooms	0.52	1	0.08	0.06	0.03	0.12	0.08	0.07	-0.25	0.43
W&D	0.05	0.08	1	0.27	0.46	0.05	0.00	0.24	0.19	0.08
Doorman	-0.01	0.06	0.27	1	0.49	0.20	0.09	0.48	-0.13	0.14
Elevator	-0.04	0.03	0.46	0.49	1	0.11	0.03	0.38	0.08	0.06
Dishwasher	0.06	0.12	0.05	0.20	0.11	1	0.20	0.20	-0.40	0.08
Patio	0.03	0.08	0.00	0.09	0.03	0.20	1	0.08	-0.21	0.04
Gym	-0.01	0.07	0.24	0.48	0.38	0.20	0.08	1	-0.10	0.10
Rental Type	-0.13	-0.25	0.19	-0.13	0.08	-0.40	-0.21	-0.10	1	-0.15
Price	0.40	0.43	0.08	0.14	0.06	0.08	0.04	0.10	-0.15	1

Table 3 - Correlation Matrix: All Variables in Union Dataset. Some variable names have been abbreviated to ensure ease of readability.

Overview of Modeling

1. Methodology and Discussion of Code

Three distinct models are used to tackle our problem statement. First, Model 1a and Model 1b are used to determine whether Airbnb and Long-term rental prices are statistically different. Two-Sample t-test compares the mean of rental price between Airbnb or long term; Model 1b regression analysis validates the significant variables to the rental price. Second, Model 2: forward selection is performed to evaluate cross-validated R2 values of combined variables. Finally, we use Model 3 Lasso regression for feature selection.

Model 1: Using the Unpaired two-samples t-test method and Regression analysis that statistically proves the two rental types are different.

a. Unpaired Two-Samples t-test

At the 0.05 level of significance, where μ_1= mean of Airbnb Price, μ_2= mean of Long-term Rental Price.

$$H_o = \mu 1 = \mu 2$$
$$H_a = \mu 1 \neq \mu 2$$

Assumptions: Rental prices are independent of each other. The Shapiro-Wilk normality test is used to test whether each of the rental types follows a normal distribution. The p-value of f-test is =.5165 which means there is no significant difference between the variances of the two rental types. All assumptions are met; therefore, we can use two-sample t-test to determine significant difference between Airbnbs and long-term rentals.

To run the two-sample t-test, we use a Shapiro test for normality. We then test for homogeneity using the var.test with price as our dependent variable and rental_type as our independent variable. Then, we perform a two-sample t-test with function t.test on our dependent variable of price with var.equal set to TRUE. The output will help us determine if there is a statistical difference on price between the two rental types.

b. Regression Analysis

We first evaluate the variables that have significant impacts on the rental price on each rental type. We use lm () function in R, lm(price ~., data = airbnb) and lm(price ~., data = long_term), to see the coefficients of the various attributes (bedrooms, bathrooms, etc.). Then, we run a regression model on the unioned dataset, lm(price ~., data = union_data.), to check if rental_type is significant to the price.

Model 2: We implement forward stepwise selection on the Airbnb data, the long-term rental data, as well as the unioned dataset between both. Our thought process behind this approach is to use the significant variables identified from the regression analysis to determine the combination of most significant variables at each step.

In the code, we compare the cross-validated R^2 values at each step to see how much of a change occurs as the combination of significant variables is adjusted. After scaling the data and creating dummy variables from categorical variables, the first step is to predict which variable (by itself) would have the highest predictive power on price by evaluating each individual variable's cross-validated R^2 regressing on price. For the following step, the rest of the variables are evaluated and the best 2 variable combination (along with the variable chosen at the first step) is then selected at the step. This process repeats until cross-validated R^2 sees a decrease at the next best variable addition.

Model 3: Lasso regression is performed on the three datasets. Lasso is slightly more robust than the stepwise regression since it uses a global approach as opposed to a greedy stepwise approach. We utilize it to determine the feature importance of Airbnb and long-term units by evaluating the magnitude of coefficients as well as investigating which unimportant features have their coefficients become 0, as Lasso regression performs feature selection.

In the code, we first look at VIF (Variance Inflation Factors) of each X variable to see if there is any multicollinearity present. Lasso regression tends to mark correlated features with smaller coefficients even if it's important (it will just choose 1). After scaling and pre-processing the data, we adjust the parameter of alpha for the Lasso regression, which controls the strength of the L1 regularizations used to penalize and restrict coefficients. Finally, we run the Lasso regression a specified alpha value and evaluate the results.

2. Results

Model 1:

After taking a random sample of the groups (Airbnb vs. Long-term Rentals) and running a two-sample t-test, we reject the null hypothesis and say with 95% confidence that there is a statistical difference on rental price between the two-rental types with a p-value of 2.2×10^{-16}. The mean of long-term rentals is $5062.12 and the mean for Airbnbs is $3923.35.

The regression testing on each rental type tells us: significant factors affecting rental prices vary depending on the rental type. In either rental type, both bedroom and bathroom are significantly related to rental prices. An additional bedroom is associated with an increase of the rental price of Airbnb by $1300.84 monthly and Long-term rental by $786.46 monthly. This is confirmed by research (Shokoohyar et al., 2020) which identified that properties with more bedrooms and bathroom are more likely to have higher returns. As for amenities, variables doorman and elevator are significant to Airbnb rental price, while washer and dryer are significant to Long-term rental price. The Multiple R-squared value from long-term rental type of 0.6768 was higher than that seen for the Airbnb rental type .2685. The higher R-squared value for long-term suggests it is a better fit for the model. Also, the p-values are statistically significant for both the Airbnb and long-term rental from our regression model.

The union regression model accounts for 24.5% of the variance in the price outcome. The p-value of the model is less than 0.05; and the p-value for rental type specifically is 1.40e-07, which is also below alpha(0.05). The coefficient for rental type is –314.87, which means the monthly rental price goes down $314.87 on average when it's Airbnb, with respect to the base case of long-term rental.

Regression Models	**Model Summary**
Airbnb	Residual standard error: 2685 on 18870 degrees of freedom Multiple R-squared: 0.2685, Adjusted R-squared: 0.267 F-statistic: 182.3 on 38 and 18870 DF, p-value: < 2.2e-16
Long-term Rental	Residual standard error: 1807 on 3504 degrees of freedom Multiple R-squared: 0.6768, Adjusted R-squared: 0.6736 F-statistic: 215.8 on 34 and 3504 DF, p-value: < 2.2e-16
Union Dataset	Residual standard error: 2760 on 22438 degrees of freedom Multiple R-squared: 0.2451, Adjusted R-squared: 0.2448 F-statistic: 809.6 on 9 and 22438 DF, p-value: < 2.2e-16

Table 4 - Regression Model Summary

Model 2:

Stepwise forward selection identifies the variables that lead to the highest cross-validated R^2 increase. The features with * indicate that the variable has a negative impact on price. For example, the table and image below indicate that Harlem and Washington Heights have an adverse impact on property income.

Ranking	AirBnb Only	Long-Term Rental Only	AirBnb + Long-Term Rental Comb.
1	Entire Home/Apt	Square Footage	Bathrooms
2	Availability (30 days) *	Building Age	Bedrooms
3	Bedrooms	Bathrooms	Harlem *
4	Bathrooms	Harlem *	Washington Heights *
5	Harlem *	Washington Heights *	Has a Doorman

Table 5 - Forward Selection Variables Ranking

As shown above, the most important features for the standalone Airbnb and long-term rental apartments involve the physical aspects of the apartment (entire home/apt and square footage).

Since “entire home/apt” often relates to the size of the unit, this indicates that the physical aspect of the apartment is the best predictor in its revenue. Across all three datasets, there is a common theme of apartments in Harlem generating less income (along with Washington Heights apartments in Long-term rental apartments and the unioned dataset).

Model 3:

The lasso regression strengthens our understanding of important features that determine a higher revenue. We find the features that the lasso regression find important are generally in line with what stepwise regression identifies. For example, the number of bathrooms remains one of the most important predictors across the three datasets. If a unit is located in Harlem, then it has the most negative impact on revenue across the three datasets. There are also many variables with a coefficient of zero that are eliminated in the datasets, many of which happen to be cities. It means the cities that are eliminated by the lasso regression have minimal to no impact on revenue.

The lasso regression allows us to explore neighborhoods that would generate the most, neutral, and least revenue. Below are neighborhoods for each:

Highest Revenue Neighborhoods	Neutral Revenue Neighborhoods	Least Revenue Neighborhoods
West Village	Civic Center	Harlem
Chelsea	East Village	Washington Heights
Midtown	Little Italy	Morningside Heights
Greenwich Village	Upper West Side	Inwood
SoHo	Gramercy Park	Roosevelt Island

Table 6 - Lasso Regression Output using the unioned dataset with alpha = 25

The VIFs of the features show that there is not much multicollinearity present in the unioned and long-term rental dataset. The Airbnb-only dataset includes some perfectly correlated variables that are shown to be insignificant, so we use lasso regression to eliminate features and address this multicollinearity.

Unexpected problems & Challenges

Since we don’t have data on expenses for each Airbnb and apartment listing, such as taxes, utility and cleaning fees, and other administrative costs, we’re unable to calculate profits and instead must opt for revenue instead as a proxy. Additionally, our data only covers New York City Airbnbs and apartments, so our findings are only applicable to apartments in NYC and cannot be extrapolated to other cities.

Moreover, regarding model building, there are many features present in the data that do not have enough predictive power on monthly revenue. Many variables are correlated with each other, so we needed to account for multicollinearity within our variables. Using the results of the VIF analysis, we were able to see which variables are highly correlated with each other and which aren’t. We combatted the multicollinearity problem by 1) using a forward selection and evaluating R^2 adjusted and CV, which penalize for redundant variables and 2) incorporating regularization techniques such as lasso regression to perform shrinkage on unimportant variables.

Another challenge that we faced was not having a wider intersection of variables available between the AirBnb and long-term rental dataset. For example, the long-term rental dataset's most important feature from the forward selection was "square footage," and this field isn't available in the AirBnb dataset. It would be helpful to see how variables crucial for one rental type stacked up against the variables in the other rental type to see if certain variables uniquely favored a certain rental method.

Suggestions for Further Study

The current study focuses on the comparison between short-term AirBnb listings and traditional long-term rentals strictly from a revenue standpoint. For a comprehensive view, models that consider net operating income may be more realistic. The current Airbnb data from OpenDataSoft contains listings from hosts with a wide range of business goals. Some hosts effectively sublet their units to generate additional income, while others are real estate investors whose listings are exclusively occupied by short term renters. The classification of these participants is not well defined.

In future studies that focus on real estate investors, it may be necessary to limit the input of short-term data strictly to participants in this category. Collecting expense data from financial reports can be time consuming and costly, and not all investors may be willing to disclose this information. A randomized survey can be conducted among hosts within the desired category, and regression models could be used to predict how much a host can expect to spend on business expenses, based on location and other listing attributes. There is a high barrier for entry into New York City's residential real estate market due to competition and high property values. For location-flexible investors, similar studies could be conducted in other cities to estimate the most profitable property locations.

As competitors continue to enter the industry, similar studies involving their listing data may be beneficial. Some of these include Vrbo, Flipkey, and Homes & Villas. Airbnb continues to be the dominant player in their segment of the short-term rental market. Many new competitors are subsidiaries of longstanding reputable brands in the travel industry such as Expedia, TripAdvisor, and Marriott International. Property owners that list with these services may benefit from increased exposure to a much wider audience.

Conclusion

The findings allow us to reject the null hypothesis and state there is a statistical difference in rental price between the two rentals. The number of bathrooms and bedrooms are major predictors of price across our three datasets. An additional bedroom is associated with an increase in rental price of an Airbnb by $1300.84 monthly and long-term rental by $786.46 monthly. Amenities such as dishwashers, door staff, patios, and gyms are less common in

Airbnb. Doorman and elevator are significant variables to Airbnb rental price, while having a washer and dryer is significant to long-term rental price.

From an investor perspective, our results suggest that the traditional long-term type brings more rental income. The unioned regression model suggests, on average, monthly rental price goes down $314.87 for an Airbnb compared to a long-term rental. Investing in a long-term rental containing more bedrooms, bathrooms, and a washer and dryer should provide a high income. More importantly, the property's neighborhood is the key driver for rental revenue. Table 6 shows which neighborhoods go for the highest, neutral, and lowest rental revenue. An investor should ideally invest in a long-term rental in a high revenue yielding neighborhood (ex. Midtown).

For a property owner, personal preferences play a large factor. Preferences in location, affordability, and features can help an owner make the right decision. The monthly rent of Airbnb is cheaper, but the addition of amenities can increase rent at a higher rate than it does for a long-term rental. The lasso regression implies that owners who are willing to pay high rent to live in West Village and Chelsea will likely do so, since those neighborhoods generate the most rental revenue.

Our findings won't apply to every rental situation; there are methods to further enhance the scholarly discussion on this debate. To name a few: calculating profit, using data outside of NYC, and having a wider intersection of variables can provide more accurate results. More related research ideas to explore can be found in the further study section.

Works Cited

Barron, K., Kung, E., & Proserpio, D. (2021). The effect of home-sharing on house prices and rents: Evidence from AirBnb. Marketing Science, 40(1), 23–47. https://doi.org/10.1287/mksc.2020.1227

Bivand, Roger, et al. “Rgdal: Bindings for the “Geospatial” Data Abstraction Library.” R-Packages, 16 Sept. 2021, CRAN.R-project.org/package=rgdal.

Bivens, J. (2019). (rep.). Economic Policy Institute. Retrieved March 23, 2023, from https://www.epi.org/publication/the-economic-costs-and-benefits-of-airbnb-no-reason-for local-policymakers-to-let-airbnb-bypass-tax-or-regulatory-obligations/

Coles, Peter A. and Egesdal, Michael and Ellen, Ingrid Gould and Li, Xiaodi and Sundararajan, Arun, Airbnb Usage Across New York City Neighborhoods: Geographic Patterns and Regulatory Implications (October 12, 2017). Forthcoming, Cambridge Handbook on the Law of the Sharing Economy, Available at SSRN: https://ssrn.com/abstract=3048397 or http://dx.doi.org/10.2139/ssrn.3048397

Horn, Keren and Mark Merante, “Is Home Sharing Driving Up Rents? Evidence from Airbnb in Boston,” Journal of Housing Economics, 2017.

Houghton, R. B., & Cook, J. H. (2020). The profitability of AirBnbs relative to long-term rentals in Los Angeles (Thesis).

New York City Department of Planning (2020). Neighborhood Tabulation Areas (NTAs). NYC Department of City Planning. (2020). https://www.nyc.gov/site/planning/data-maps/open-data/census-download-metadata.page

Shokoohyar, S., Sobhani, A. and Sobhani, A. (2020), "Determinants of rental strategy: short-term vs long-term rental strategy", International Journal of Contemporary Hospitality Management, Vol. 32 No. 12, pp. 3873-3894. https://doi.org/10.1108/IJCHM-03-2020-0185

Appendix

Regression Model Variable Summary

Variables	airbnb		long_term	
	estimate	p.value	estimate	p.value
(Intercept)	899.85	0.01	-70.24	0.72
bedrooms	1300.84	0.00	786.46	0.00
bathrooms	1693.29	0.00	3021.59	0.00
has_washer_dryer	24.41	0.60	201.75	0.04
has_doorman	622.49	0.00	31.13	0.76
has_elevator	258.71	0.00	142.55	0.18
has_dishwasher	-2226.65	0.24	-43.19	0.64
has_patio	-517.02	0.74	-88.97	0.51
has_gym	305.02	0.00	-63.19	0.57
neighborhoodChelsea	728.08	0.04	1301.04	0.00
neighborhoodChinatown	-839.53	0.02	958.44	0.00
neighborhoodCivic Center	-579.63	0.28	-890.42	0.18
neighborhoodEast Village	-216.51	0.53	-44.13	0.86
neighborhoodFinancial District	-734.68	0.04	-229.24	0.27
neighborhoodFlatiron District	684.79	0.12	1060.27	0.00
neighborhoodGramercy Park	-158.46	0.67	213.77	0.46
neighborhoodGreenwich Village	1154.31	0.00	1162.18	0.00
neighborhoodHarlem	-1998.72	0.00	-2122.98	0.00
neighborhoodHell's Kitchen	-170.84	0.62	-2150.62	0.00
neighborhoodInwood	-2585.76	0.00	-2704.65	0.00
neighborhoodKips Bay	162.74	0.65	-975.88	0.36
neighborhoodLittle Italy	-394.43	0.37	-706.68	0.44
neighborhoodLower East Side	-634.04	0.07	-561.60	0.09
neighborhoodMarble Hill	-2346.87	0.14	-2342.81	0.20
neighborhoodMidtown	432.70	0.21	-149.49	0.42
neighborhoodMorningside Heights	-1773.43	0.00	-1388.02	0.01
neighborhoodMurray Hill	-176.54	0.64	444.97	0.48
neighborhoodNoHo	511.25	0.26	-1327.82	0.11
neighborhoodNolita	214.40	0.57	2257.77	0.00
neighborhoodRoosevelt Island	-2401.33	0.00	-28.02	0.98
neighborhoodSoHo	661.91	0.07	1711.34	0.00
neighborhoodStuyvesant Town	-1017.99	0.06	-252.86	0.19
neighborhoodTheater District	-247.71	0.53	-206.53	0.28
neighborhoodTribeca	742.15	0.07	-2806.28	0.00
neighborhoodTwo Bridges	-1436.20	0.00	1244.12	0.00

Lasso Regression for Unioned Dataset

Variable:	Coefficient:	Variable:	Coefficient:	Variable:	Coefficient:	Variable:	Coefficient:
bedrooms	850.748	Chinatown	-72.808	Little Italy	0	Stuyvesant Town	-25.36
bathrooms	819.639	Civic Center	0	Long Island City	0	Theater District	3.056
has_washer_dryer	19.377	East Village	0	Lower East Side	-77.681	Tribeca	118.852
has_doorman	154.222	Financial District	-11.27	Manhattanville	0	Two Bridges	-32.723
has_elevator	101.285	Flatiron District	87.933	Marble Hill	0	Upper East Side	0
has_dishwasher	-1.191	Gramercy Park	0	Midtown	43.423	Upper West Side	0
has_patio	0	Greenwich Village	157.494	Morningside Heights	-179.822	Washington Heights	-401.442
has_gym	16.086	Hamilton Heights	-18.341	Murray Hill	0	West Village	185.082
rental_type	0	Harlem	-643.597	NoHo	11.084		
Battery Park City	0	Hell's Kitchen	0	Nolita	19.232		
Central Park South	37.841	Inwood	-230.313	Roosevelt Island	-74.104		
Chelsea	208.409	Kips Bay	47.172	SoHo	125.196		

Lasso regression coefficients:
bedrooms: 850.748
bathrooms: 819.639
has_washer_dryer: 19.377
has_doorman: 154.222
has_elevator: 101.285
has_dishwasher: -1.191
has_patio: -0.000
has_gym: 16.086
rental_type: 0.000
Battery Park City: 0.000
Central Park South: 37.841
Chelsea: 208.409
Chinatown: -72.808
Civic Center: -0.000
East Village: -0.000
Financial District: -11.270
Flatiron District: 87.933
Gramercy Park: 0.000
Greenwich Village: 157.494
Hamilton Heights: -18.341
Harlem: -643.597
Hell's Kitchen: 0.000
Inwood: -230.313
Kips Bay: 47.172
Little Italy: -0.000
Long Island City: -0.000
Lower East Side: -77.681
Manhattanville: -0.000
Marble Hill: 0.000
Midtown: 43.423

Morningside Heights: -179.822
Murray Hill: -0.000
NoHo: 11.084
Nolita: 19.232
Roosevelt Island: -74.104
SoHo: 125.196
Stuyvesant Town: -25.360
Theater District: 3.056
Tribeca: 118.852
Two Bridges: -32.723
Upper East Side: -0.000
Upper West Side: -0.000
Washington Heights: -401.442
West Village: 185.082

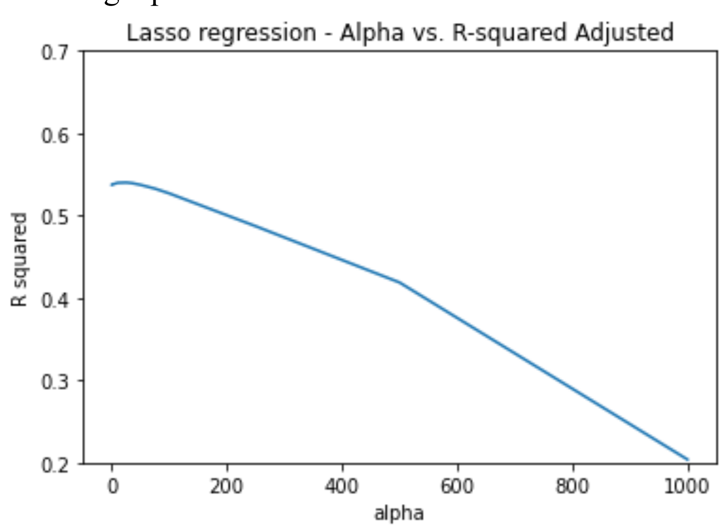
VIF for Unioned Dataset

Variable:	VIF:	Variable:	VIF:	Variable:	VIF:	Variable:	VIF:
bedrooms	1.394654	Chinatown	1.398981	Little Italy	1.094611	Stuyvesant Town	1.040076
bathrooms	1.48228	Civic Center	1.042564	Long Island City	1.002113	Theater District	1.186465
has_washer_dryer	1.36052	East Village	2.956708	Lower East Side	2.007723	Tribeca	1.284758
has_doorman	1.619184	Financial District	1.53517	Manhattanville	1.000317	Two Bridges	1.059664
has_elevator	1.669008	Flatiron District	1.16702	Marble Hill	1.003198	Upper East Side	2.840209
has_dishwasher	1.275869	Gramercy Park	1.330835	Midtown	2.452884	Upper West Side	3.162615
has_patio	1.070699	Greenwich Village	1.411201	Morningside Heights	1.411625	Washington Heights	1.895183
has_gym	1.41725	Hamilton Heights	1.007111	Murray Hill	1.277594	West Village	1.830064
rental_type	1.688656	Harlem	4.839633	NoHo	1.08818		
Battery Park City	1.151402	Hell's Kitchen	2.650341	Nolita	1.284566		
Central Park South	1.022266	Inwood	1.23547	Roosevelt Island	1.060942		
Chelsea	2.192218	Kips Bay	1.446168	SoHo	1.399435		

	VIF Factor	feature
0	1.394654	bedrooms
1	1.482280	bathrooms
2	1.360520	has_washer_dryer
3	1.619184	has_doorman
4	1.669008	has_elevator
5	1.275869	has_dishwasher
6	1.070699	has_patio
7	1.417250	has_gym
8	1.688656	rental_type
9	1.151402	Battery Park City
10	1.022266	Central Park South

11	2.192218	Chelsea
12	1.398981	Chinatown
13	1.042564	Civic Center
14	2.956708	East Village
15	1.535170	Financial District
16	1.167020	Flatiron District
17	1.330835	Gramercy Park
18	1.411201	Greenwich Village
19	1.007111	Hamilton Heights
20	4.839633	Harlem
21	2.650341	Hell's Kitchen
22	1.235470	Inwood
23	1.446168	Kips Bay
24	1.094611	Little Italy
25	1.002113	Long Island City
26	2.007723	Lower East Side
27	1.000317	Manhattanville
28	1.003198	Marble Hill
29	2.452884	Midtown
30	1.411625	Morningside Heights
31	1.277594	Murray Hill
32	1.088180	NoHo
33	1.284566	Nolita
34	1.060942	Roosevelt Island
35	1.399435	SoHo
36	1.040076	Stuyvesant Town
37	1.186465	Theater District
38	1.284758	Tribeca
39	1.059664	Two Bridges
40	2.840209	Upper East Side
41	3.162615	Upper West Side
42	1.895183	Washington Heights
43	1.830064	West Village

Choosing alpha – Unioned dataset



Airbnb Only – Lasso Regression

Lasso regression coefficients:
bedrooms: 763.673
bathrooms: 502.692
has_washer_dryer: 34.821
has_doorman: 142.397
has_elevator: 131.381
has_dishwasher: -0.000
has_patio: -0.000
has_gym: 102.425
guests_included: 223.496
extra_people: -0.000
min_nights: -40.074
max_nights: -0.000
availability_30: -1259.364
rental_type: 0.000
Battery Park City: 0.000
Chelsea: 147.188
Chinatown: -24.581
Civic Center: -0.000
East Harlem: -193.416
East Village: -0.000
Financial District: -21.596
Flatiron District: 28.291
Gramercy: 0.000
Greenwich Village: 99.825
Harlem: -339.197
Hell's Kitchen: 61.369
Inwood: -103.304
Kips Bay: 26.588
Little Italy: 0.420
Lower East Side: -45.831
Marble Hill: -0.000
Midtown: 128.801
Morningside Heights: -133.417
Murray Hill: 8.208
NoHo: 18.853
Nolita: 69.783
Roosevelt Island: -51.120
SoHo: 94.425
Stuyvesant Town: -7.071
Theater District: 24.009
Tribeca: 86.913
Two Bridges: -9.013
Upper East Side: -12.048
Upper West Side: -0.000
Washington Heights: -190.210
West Village: 147.214
Apartment: -159.943
Bed & Breakfast: -0.000
Boat: 0.000
Boutique hotel: -0.000
Bungalow: -0.000

Cabin: 0.000
Castle: 0.000
Condominium: 0.000
Dorm: -10.829
Guest suite: -0.000
Guesthouse: -0.000
Hostel: -22.400
House: 0.000
Hut: -0.000
Lighthouse: 0.000
Loft: 0.000
Other: 7.528
Serviced apartment: 0.000
Timeshare: 24.389
Townhouse: 0.000
Vacation home: 6.955
Villa: -1.601
Entire home/apt: 1003.120
Shared room: -11.745
Airbed: -10.772
Couch: -0.000
Futon: -0.000
Pull-out Sofa: 0.000

AirBnb only – VIF

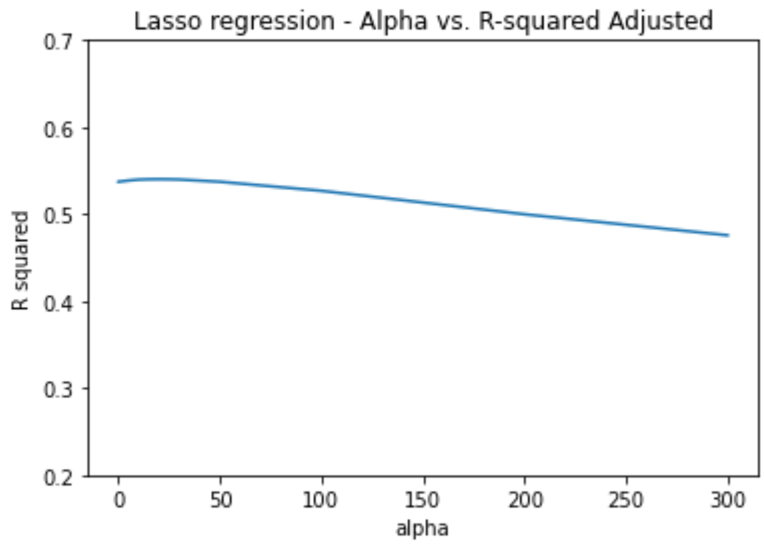
As shown below, there are numerous variables that are perfectly collinear with others

	VIF Factor	feature
0	1.851	bedrooms
1	1.306	bathrooms
2	1.365	has_washer_dryer
3	1.516	has_doorman
4	1.623	has_elevator
5	1.006	has_dishwasher
6	1.003	has_patio
7	1.386	has_gym
8	2.255	beds
9	1.592	guests_included
10	1.211	extra_people
11	1.015	min_nights
12	1.001	max_nights
13	7.868	availability_30
14	31.667	availability_60
15	19.031	availability_90
16	2.143	availability_365
17	0.000	rental_type
18	inf	Battery Park City
19	inf	Chelsea
20	inf	Chinatown
21	inf	Civic Center
22	inf	East Harlem
23	inf	East Village
24	inf	Financial District
25	inf	Flatiron District

26	inf	Gramercy
27	inf	Greenwich Village
28	inf	Harlem
29	inf	Hell's Kitchen
30	inf	Inwood
31	inf	Kips Bay
32	inf	Little Italy
33	inf	Lower East Side
34	inf	Marble Hill
35	9007199254740992.000	Midtown
36	inf	Morningside Heights
37	inf	Murray Hill
38	inf	NoHo
39	inf	Nolita
40	inf	Roosevelt Island
41	inf	SoHo
42	inf	Stuyvesant Town
43	inf	Theater District
44	inf	Tribeca
45	inf	Two Bridges
46	inf	Upper East Side
47	inf	Upper West Side
48	inf	Washington Heights
49	inf	West Village
50	inf	Apartment
51	inf	Bed & Breakfast
52	inf	Boat
53	inf	Boutique hotel
54	inf	Bungalow
55	inf	Cabin
56	inf	Castle
57	inf	Condominium
58	inf	Dorm
59	inf	Guest suite
60	inf	Guesthouse
61	inf	Hostel
62	inf	House
63	inf	Hut
64	inf	Lighthouse
65	inf	Loft
66	inf	Other
67	inf	Serviced apartment
68	9007199254740992.000	Timeshare
69	inf	Townhouse
70	inf	Vacation home
71	inf	Villa
72	4503599627370496.000	Entire home/apt
73	9007199254740992.000	Private room
74	inf	Shared room
75	inf	Airbed
76	inf	Couch
77	inf	Futon
78	inf	Pull-out Sofa

79 inf Real Bed

AirBnb only – Choosing alpha



Long-Term Rental Only – Lasso Regression

Lasso regression coefficients:
bathrooms: 625.887
has_washer_dryer: 7.542
has_doorman: -0.000
has_elevator: 0.000
has_dishwasher: -0.000
has_patio: -0.000
has_gym: 0.000
size_sqft: 2150.538
min_to_subway: -45.181
floor: 268.702
building_age_yrs: -256.037
no_fee: -26.690
rental_type: 0.000
Battery Park City: 0.000
Central Harlem: -234.988
Central Park South: 72.858
Chelsea: 160.728
Chinatown: -0.000
East Harlem: -122.030
East Village: 49.235
Financial District: -30.955
Flatiron: 135.156
Gramercy Park: 72.213
Greenwich Village: 104.602
Hamilton Heights: -84.870
Inwood: -80.661
Little Italy: 0.000
Long Island City: -13.300

Lower East Side: -0.000
Manhattanville: -0.000
Midtown: 0.000
Midtown East: -85.337
Midtown South: -0.000
Midtown West: -21.790
Morningside Heights: -34.189
Nolita: 19.415
Roosevelt Island: -10.685
Soho: 184.869
Stuyvesant Town/PCV: 0.000
Tribeca: 156.384
Upper East Side: -0.000
Upper West Side: 33.954
Washington Heights: -197.161
West Harlem: -19.627
West Village: 176.457

Long-term rental only – VIF

	VIF Factor	feature
0	1.139	Central Park South
1	1.316	Central Harlem
2	1.633	Midtown
3	3.263	bathrooms
4	1.442	building_age_yrs
5	1.958	has_gym
6	2.975	bedrooms
7	1.061	Morningside Heights
8	2.889	Midtown East
9	1.430	floor
10	1.015	Chinatown
11	1.317	Soho
12	1.025	Long Island City
13	1.191	no_fee
14	1.035	Nolita
15	1.019	Stuyvesant Town/PCV
16	1.304	Greenwich Village
17	1.016	Little Italy
18	2.390	has_doorman
19	1.377	has_washer_dryer
20	3.556	Upper West Side
21	1.411	has_dishwasher
22	1.218	Washington Heights
23	1.544	Battery Park City
24	1.181	Lower East Side
25	1.516	Flatiron
26	1.378	Midtown South
27	1.278	West Village
28	4.034	size_sqft
29	1.131	East Harlem
30	1.045	Inwood
31	2.137	Midtown West
32	1.017	West Harlem

33	1.048	has_patio
34	1.577	Tribeca
35	1.670	Chelsea
36	1.364	East Village
37	1.069	Hamilton Heights
38	2.449	Financial District
39	1.005	Manhattanville
40	3.370	Upper East Side
41	2.539	has_elevator
42	1.200	min_to_subway
43	1.023	Roosevelt Island
44	1.219	Gramercy Park

Long-term rental only – Choosing alpha