Have COVID-19 Vaccinations Impacted US Tourism?

Kasha Muzila, Stephanie He & Hannah Gross

W203 Statistics for Data Science ~ Summer 2021

Contents

Please note that due to extended wait times for rendering output pdf, we have commented out the code for several tables and plots, and used images of the figures instead. Should you like to run the tables and plots directly from this Rmd file, please uncomment the code above each image.

1   Introduction

Since its founding in 2008, Airbnb has skyrocketed in popularity as an alternative to traditional hotel stays. The company provides travelers an authentic experience at their destination by connecting them with locals, who are willing to rent out their homes and apartments. By 2011, Airbnb had expanded their listings to 220 countries and regions, surpassing seven million listings by 2021[1]. However, the world-wide COVID-19 pandemic caused nations to close their borders to stop the spread of the virus. As a result of limited international and intra-national travel, Airbnb suffered a 72% decline in revenue and over 1,800 job cuts[2]. Because Airbnb has become such a popular company in the tourism industry, we decided to use Airbnb availability data as a proxy to measure tourism levels within the United States.

Starting in January 2021, COVID-19 vaccinations have been distributed across the US. By April 2021, the United States' Center for Disease Control and Prevention (CDC) began to relax travel restrictions[3]. As international and intra-national borders began reopening, Airbnb and the rest of the travel industry are beginning to recover.

This paper seeks to identify a causal relationship between Airbnb availability and state-level vaccination percentages. We believe that as a state's fully and partially vaccinated populations grow, more people will be willing to travel again within the United States, and the availability of Airbnb listings will decrease. Our research question is as follows:

Does a higher vaccinated state population percentage decrease Airbnb availability within popular tourist destinations within that state?

We believe that investigation of causal relationships between COVID-19 and tourism will shed light on important information to help the tourism industry recover from the COVID-19 pandemic. We also believe that this research may be useful for helping to quell the pandemic by linking the aspects of the pandemic to the interests of the tourism industry. Finally, we believe that studying the relationships between COVID-19 vaccinations and state policy may be useful to lawmakers in states whose economies rely heavily on tourism.

The rest of the paper is structured as follows. The Data Description section states each dataset we used and provides a brief background about each set. The Exploratory Data Analysis section describes the background of the variables used in our models as well as any transformations made and the relationships between these variables. The Statistical Model section states our models and their differences. The Results section discusses the statistical significance of each model output as well as presents our overarching findings. The Assumptions section evaluates each model by both large sample assumptions and additional CLM assumptions. The Omitted Variables section reports variables that were omitted from our study but may cause bias in our results. Finally, our Conclusion and Limitations sections summarize our major findings and potential opportunities for continued research.

---

[1]Deane, "2021 Airbnb Statistics: Usage, Demographics, and Revenue Growth"
[2]Fortune.com, "Airbnb's IPO filing reveals huge COVID impact"
[3]AJMC, "A Timeline of COVID-19 Vaccine Developments in 2021"

2   Data Description

To operationalize our research question, we utilized three data sources: Inside Airbnb[4], the CDC's COVID-19 Vaccinations in the United States, Jurisdiction[5] and the Boston University School of Public Health's COVID-19 US State Policy Database[6]. We joined the three datasets by state, giving us our final dataset of 104,095 observations.

2.1   Inside Airbnb

Originally designed by Murry Cox, Inside Airbnb is an independent, non-commercial set of tools and data. Inside Airbnb provides users with Airbnb listing data that was scraped directly from the Airbnb website at specific points in time. The primary goal of Inside Airbnb is to facilitate public discussion on if hosts are renting out residential properties permanently as hotels, as opposed to sharing the primary residence in which they live "occasionally"[7]. While the intent of the dataset is outside the scope of this analysis, we believe the broad methodology employed in data collection makes this a valid dataset to derive insights regarding tourism.

Inside Airbnb only provides data for about 30 different locations within the United States. For the purpose of our analysis, we gathered all of the available US listings data. Pertinent listing information in this dataset includes: the availability over the next 60 days, the type of room, the cumulative rating, the price, and whether the host is a Super-Host.

To ensure the quality and appropriateness of the Inside Airbnb listings data, we made two adjustments to the dataset. First, we found that 27,486 of the Airbnb listings in the dataset have zero availability over the next 365 days. We believe it is highly unlikely that 27,486 Airbnb listings have been completely booked for the next 365 days by guests. Rather, it is far more likely the availability of these listings were blocked by the host. Therefore, we removed these listings from the dataset because we believe that the hosts of these listings do not intend to rent out their listings in the coming months. Second, we found that 865 Airbnb listings impose a minimum of nights greater than 90 nights. We believe that it is highly likely that the hosts of these listings are using the number of required minimum nights to deter travelers from booking their listing. Therefore, we removed these listings from our dataset, because we do not consider these listings to be valid for our analysis.

2.2   COVID-19 Vaccinations in the United States, Jurisdiction

Our secondary data source, COVID-19 Vaccinations in the United States, Jurisdiction, provides official information from the CDC regarding COVID-19 vaccinations within the United States and is updated daily. The data represents all vaccine partners, including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities[8]. Pertinent vaccination information in this dataset consists of vaccination rates by state and age demographics.

2.3   COVID-19 US State Policy Database

Our third data source is the COVID-19 US State Policy Database created and maintained by the Boston University School of Public Health. The database tracks the dates when each US state implemented and revoked new social safety net, economic, and physical distancing policies in response to the COVID-19 pandemic. In the interest of our analysis, we used physical distancing policies, including the date of restaurant reopening and the date face mask mandates were ended.

---

[4]Cox, "Inside Airbnb"

[5]CDC, "COVID-19 Vaccinations in the United States,County"

[6]Raifman, Julia, Nocka, Kristen, Jones, David, Bor, Jacob, Lipson, Sarah, Jay, Jonathan, . . . Galea, Sandro. COVID-19 US State Policy Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-11-02. https://doi.org/10.3886/E119446V40 "link"

[7]ibid

[8]ibid

3   Exploratory Data Analysis

See Appendix A for codebook.

3.1   Outcome Variable: Airbnb Availability (T)

`availability_60`

To operationalize tourism, our outcome concept, we decided to use Airbnb listing availability over the next 60 days as our outcome variable. This variable reports the number of days a listings has neither been booked by a traveler or blocked by the host.

In addition to the 60 day availability variable, Inside Airbnb also provides availability information for the next 30, 90 and 365 days for each listing. We decided to use the 60 day variable for a number of reasons. First, our data was scraped between March and May 2021. This time period is about 60-90 days before people tend to take summer vacations (~June - August). We also decided to use 60 days instead of 90 days because we believe that the uncertainty with the pandemic will likely make people book Airbnb listings closer to their vacation date.

Distribution of Airbnb Availability of the Next 60 Days
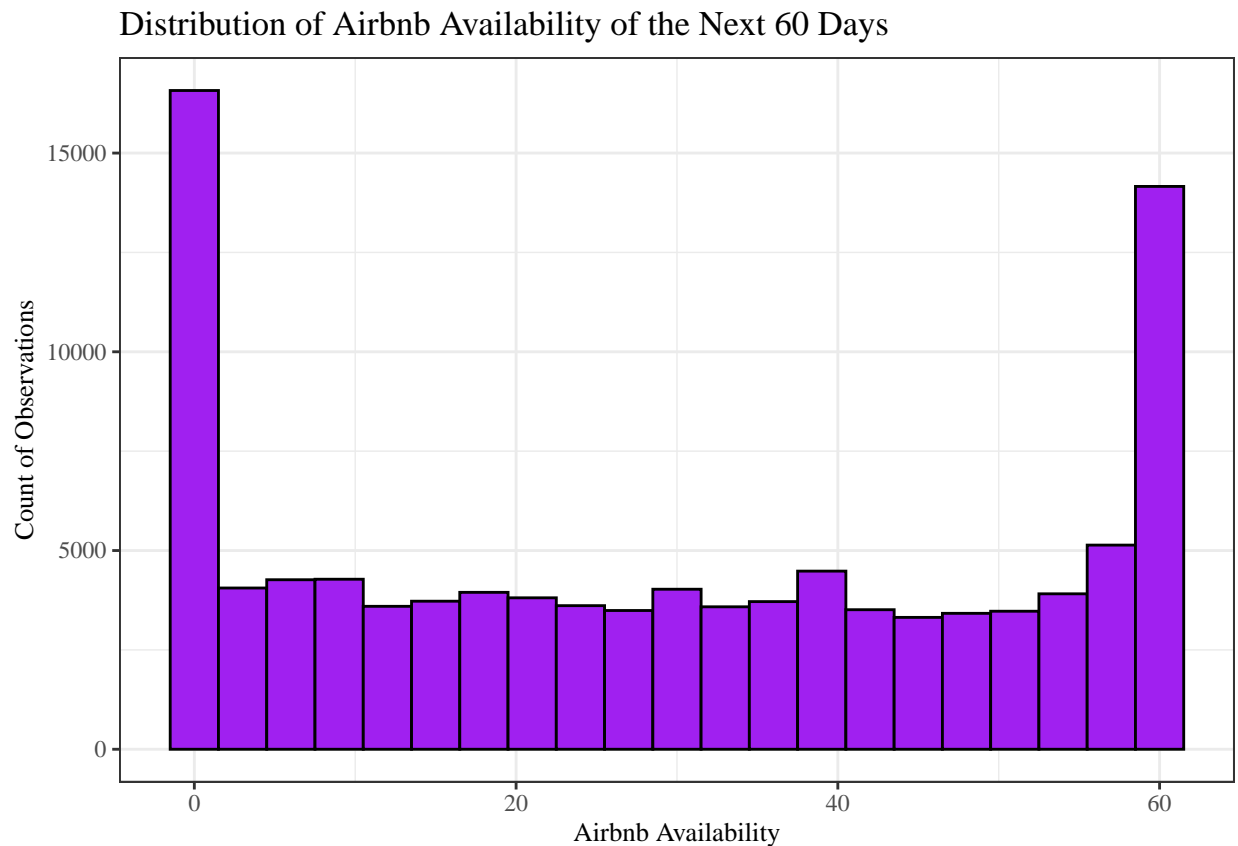

Figure 1: Airbnb Availability Over the Next 60 Days

Our final dataset contains 13,214 (~13%) listings with zero availability and 10,085 (~9.7%) that are completely available for the next 60 days. The rest of the listings are evenly split across the two extremes (see Figure 1).
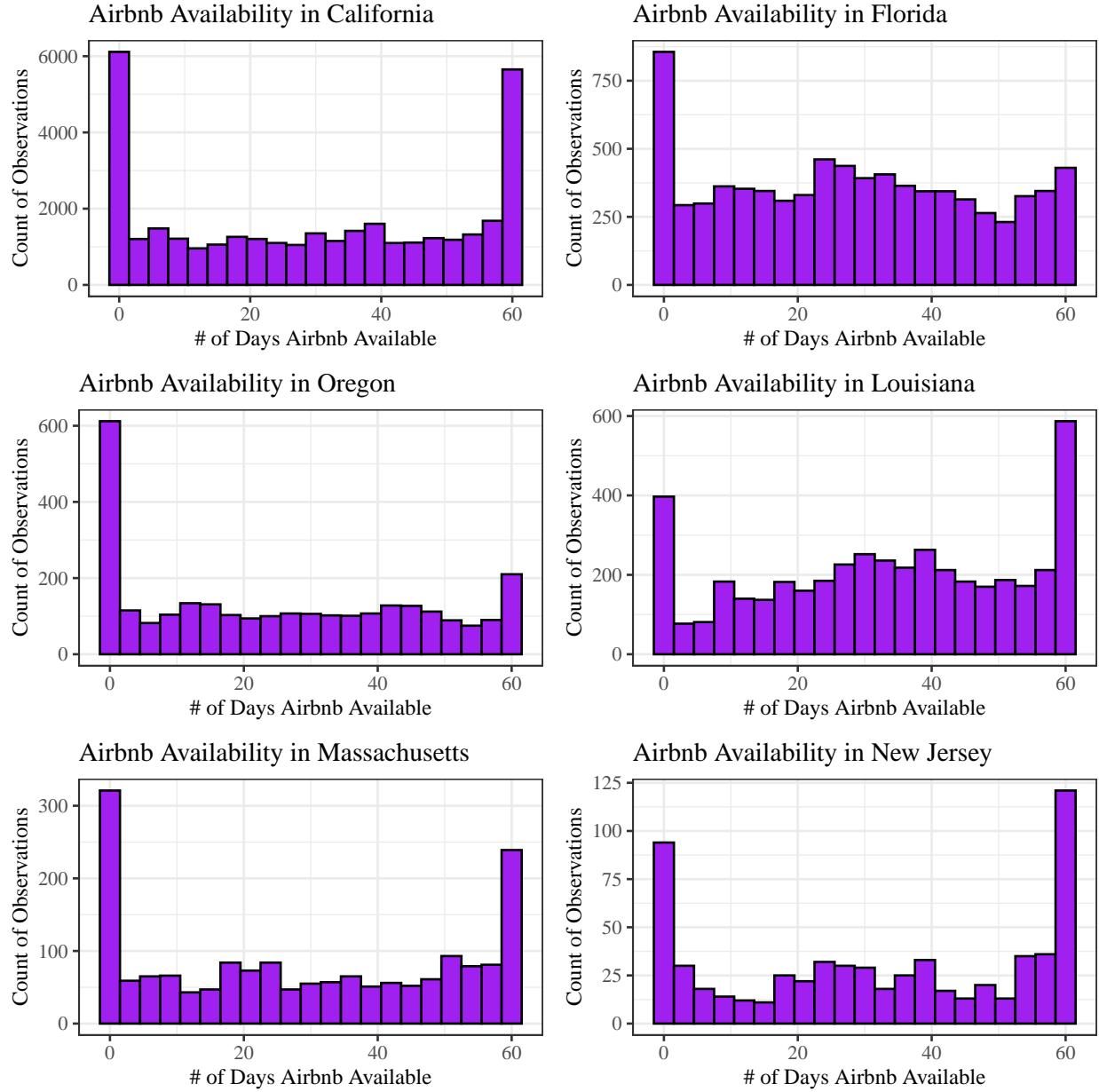
Figure 2: Availability Histograms by State

In addition, we found that the distribution of availability for each state was more or less proportional to that of the entire dataset (see Figure 2). Therefore, the distribution of Airbnb availability does not have any skewness toward any particular state. We believe that the large amount of listings with completely open availability in the next 60 days makes sense because we expect potential travelers to be more hesitant about traveling during the pandemic. Similarly, it makes sense that hosts may be blocking their calendars to stop people from booking their listing due to COVID-19. As such, we believe it is appropriate to use this data for our outcome variable without transformation.

3.2 Feature of Interest: Vaccinated Population Percentage (V)

`administered_dose1_recip_18plus_pct`

To operationalize vaccinated population percentage, we used the percentage of people vaccinated with at least one dose of a COVID-19 vaccine within that state for age demographics 18 years+, reported in the CDC's COVID-19 Vaccinations in the United States, Jurisdiction dataset.

There are several reasons why we believe this variable is a good indicator for the vaccination rate for a particular state. First, we believe people who receive the first dose of vaccine are likely to receive the second dose in the near term. Additionally, many other studies regarding COVID-19 only look at adult populations because this group is at higher risk[9]. We believe focusing our attention on the adult population is consistent with COVID-19 research.

We generated two methodologies for joining the COVID-19 Vaccination dataset with the Inside Airbnb listings data. First, we filtered the vaccine dataset by a single date and then joined by state. This method resulted in only 18 unique values, significantly limiting the information in our main feature of interest.

We replaced this initial method with an alternative that joins the datasets by state and the `last_scraped` date of each Airbnb listing. We modified the `last_scraped` date by subtracting 14 days from each observation. We chose 14 days because it takes approximately 14 days for a vaccine to become effective and also for people to take their second dose. We also expect that Airbnb bookings must have been booked some time before the listings' `last scraped` date, so we must use a date further in the past to try to approximate the time when bookings may have occurred. We believe this second method allows us to better approximate the conditions under which Airbnb bookings were made. This second method also allowed us to add more variation within our main variable of interest. After this adjustment, there are 122 unique values within this variable.
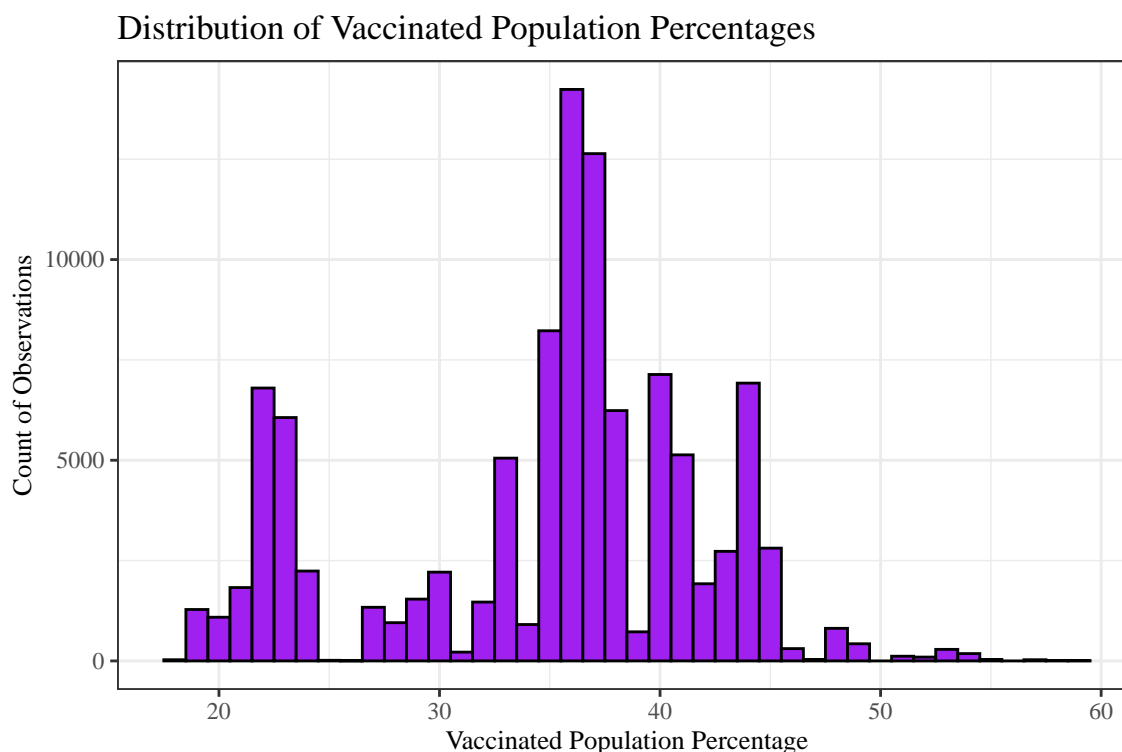

Figure 3: State-Level Vaccinated Population Percentages

---

[9]Leidman E, Duca LM, Omura JD, Proia K, Stephens JW, Sauber-Schatz EK. COVID-19 Trends Among Persons Aged 0–24 Years — United States, March 1–December 12, 2020. MMWR Morb Mortal Wkly Rep 2021;70:88–94. DOI: http://dx.doi.org/10.15585/mmwr.mm7003e1external icon. "link"

Figure 3 shows the distribution of our key variable of interest, vaccinated population percentage. The variable fails to produce a normal distribution and clusters around a few values (22%, 36%, 45%, etc) for a few reasons. Firstly, the variation calculated from the second methodology derives the vaccination percentage 14 days prior to a listings' `last_scraped` date and the vaccinated population percentage grows gradually over time. This means that no significant vaccination rate changes occurred during the 14 days. Additionally, the Airbnb listings are also concentrated in certain states. In our dataset, the top three states - California, New York and Florida account for more than 50% of all listings (see Appendix B).

3.3 Additional Features

Based on AirDNA's research, a specialized short-term rental analytics website with a focus on Airbnb[10], key factors that impact Airbnb listing occupancy rates during normal times, which is close to our outcome variable, are pricing, Super-Host status, and ratings. Below we discuss additional features we have incorporated in our own models that may impact our outcome variable.

3.3.1 Listing Price per Night (log transform) (P)

`lprice`

Price shows the daily price of an Airbnb listing in USD. Values range from $10.00 to $24,999.00 with median at $120.00 and mean at $186.80.
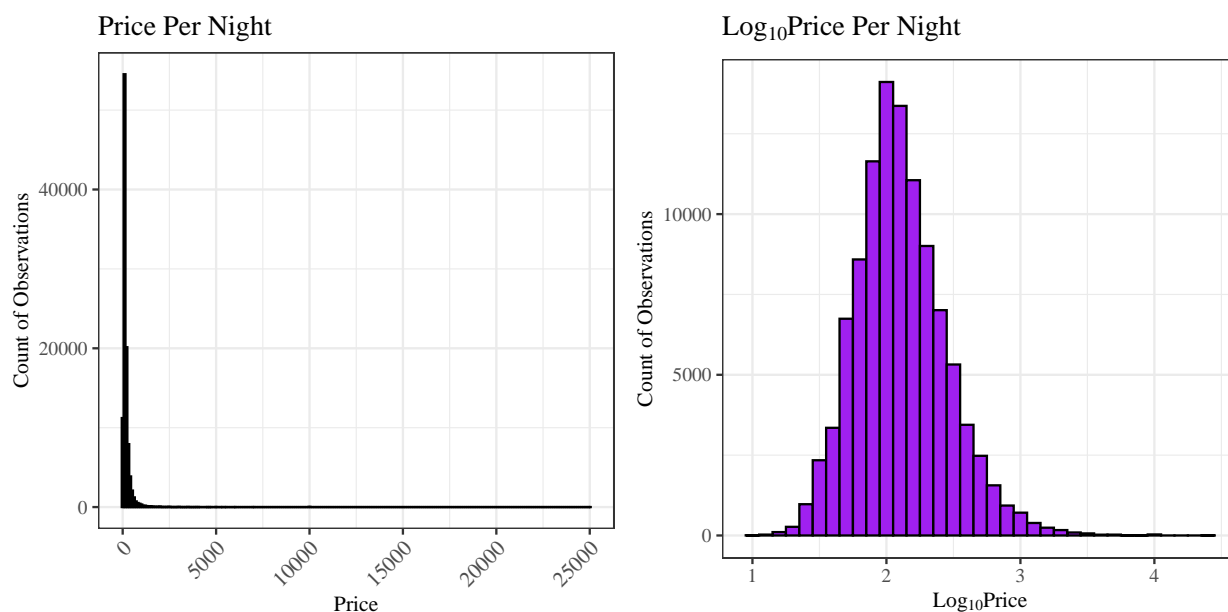
Figure 4: Listing Price per Night Variable Comparison

Based on the histogram of price (Figure 4), we view it is reasonable to use the log base 10 transformation as people focus on price increase in terms of percentage instead of dollar amount.

3.3.2 Maximum Capacity (A)

`accommodates`

The Inside Airbnb dataset provides the maximum capacity of each listing. The range of values for `accommodates` is between 1 to 16 with median at 4.

---

[10] AirDNA, "How to Optimize Your Airbnb Occupancy Rate with Data Analytics"

3.3.3 Listing Rating (S)

`review_score_value`

The Inside Airbnb dataset gives us information on the cumulative rating of each Airbnb listing. This variable scores on a scale between one and ten. The values in this variable range from 2 to 10 with mean and median at 9 and 10 respectively. Most of the review scores are between 8 and 10 with ~59% at a score of 10. Because both hosts and guests rate each other, we view the highly skewed distribution of this variable as reasonable[11].

3.3.4 Super-Host (SH)

`host_is_superhost`

Super-Hosts are experienced hosts who provide an outstanding example for other hosts and extraordinary experiences for their guests. A Super-Host badge is displayed on the webpage of these host's listings[12]. The Inside Airbnb dataset provides a Boolean variable to indicate a Super-Host. We believe that Super-Host status is closely related to a listings' desirability. There are 44,015 Super-Host listings in our final dataset.

3.3.5 Property Type (PT)

`room_type`

All Airbnb listings are grouped into four different room types; entire place, hotel room, private room, and shared room. Entire place functions as a house or apartment, and typically includes a kitchen, bedroom, bathroom, and a dedicated entrance that is not shared with others. Hotel rooms are regular hotel rooms that meet Airbnb's listing standards. Private rooms function as a private room for sleeping with access to shared spaces, such as a kitchen and/or living room area. Shared rooms function as a sleeping area that a traveler would share with other guests.

The most popular room types in our dataset are the entire home/apartment (~72%) and private room (~26%). The number of shared rooms and hotel rooms are small, accounting for ~2% of listings.

3.3.6 Number of Days Since the End of Face Mask Mandates

`date_diff_fm_end`

To calculate this variable, we obtained the most recent date the face mask mandate ended for each state from COVID -19 US State Policy Database as of July 14th 2021. If the state did have a mask mandate that ended before a listing's last scraped date, we got the difference between the end date of the mask mandate and the listing's last scraped date. If the state had a mask mandate that has either not ended or ended after a listing's last scraped date, the variable's value should be zero. In the case where a state did not have a mask mandate, the variable's value is also zero. We believe face mask mandates will significantly impact people's willingness to travel, so we believe it is appropriate to include this variable in our model.

In terms of the data, most of the states have not lifted the face mask mandate. Over 90% of values in this variable are zero, and only one state, Texas, has lifted the face mask mandate.

3.3.7 Number of Days Since Restaurants Reopened

`date_diff_end_rest`

Similar to the variable above, we obtained the most recent date restaurants were reopened for each state from COVID -19 US State Policy Database as of July 14th, 2021. The variable is calculated based on the days between the restaurant last reopen date and the `last_scraped` date of the listing. We believe that the number of days restaurants have been reopened will have a large impact on tourism within that state. There is a larger range of values in this variable than in the face mask policy variable. However, the values are sporadic. Over 35% of data points are zero, ~15% between 40-100 days and rest around 300-360 days.

---

[11] Airbnb, "Reviews for stays"
[12] Airbnb, "What is a Superhost?"
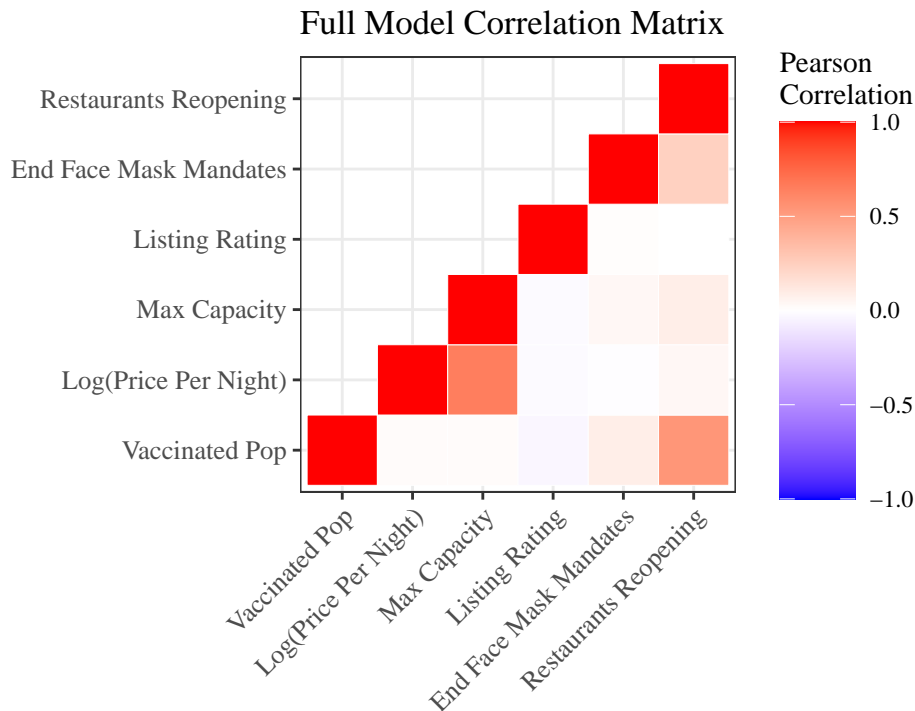
3.4 Relationships Amongst Variables

Figure 5: Full Model Correlation

Figure 5 shows the crosswise correlation between each feature in our models. Most variables are independent of each other, as indicated by the light-colored squares. However, there are three areas where we detect collinearity. First, we can see that there is a fairly strong positive relationship between our variable of interest and the number of days since restaurants were reopened. There also seems to be a less strong but important positive relationship between the number of days since restaurants were reopened and the number of days since the end of face mask mandates. We are cognizant that the two state policy variables may have a collinear relationship with a state's vaccinated population percentage, because the state is more likely to loosen social safety policies if there is a high vaccination rate within the state[13]. Despite this potential collinearity, we believe that state policy is an important variable to include in our causal models. Finally, there is a relatively high positive correlation between price and maximum capacity of a listing. We have decided to keep the maximum capacity variable because we believe it will add more detail to our models.

---

[13]Apnews.com, "More states ease lingering virus rules as vaccine rates rise"

4 Statistical Models

To explore our research question, we built four models to examine the impact of our key variable, vaccinated population percentage (V).

Model Name	Model Code	Formula
Restricted Model	`model_one`	$$f(T) = \beta_0 + \beta_1 V$$
Less Restricted Model - Airbnb	`model_two`	$$f(T) = \beta_0 + \beta_1 V + \beta_2 P + \beta_3 A \\ + \beta_4 SH + \beta_5 S + \beta_6 PT(Hotel) \\ + \beta_7 PT(Private) + \beta_8 PT(Shared)$$
Less Restricted Model - State Policy	`model_three`	$$f(T) = \beta_0 + \beta_1 V + \beta_2 FM + \beta_3 RS$$
Full Model	`model_four`	$$f(T) = \beta_0 + \beta_1 V + \beta_2 P + \beta_3 A \\ + \beta_4 SH + \beta_5 S + \beta_6 PT(Hotel) \\ + \beta_7 PT(Private) + \beta_8 PT(Shared) \\ + \beta_9 FM + \beta_{10} RS$$

Our null hypothesis is $H_0 : \beta_1 = 0$. We will reject the null hypothesis if the resultant p-value of the coefficient of our variable of interest is less than .05.

5 Results

Table 1: Relationship Between Airbnb Availability and Vaccinated Population Percentage

	Airbnb Availability			
	Model 1	Model 2	Model 3	Model 4
Vaccinated Pop	−0.014	−0.052***	0.045***	−0.024*
	(0.009)	(0.009)	(0.011)	(0.010)
Listing Rating		−2.271***		−2.257***
		(0.073)		(0.073)
Log(Price)		9.733***		9.632***
		(0.272)		(0.272)
Max Capacity		−0.355***		−0.338***
		(0.030)		(0.030)
Super-Host		−4.942***		−4.955***
		(0.133)		(0.133)
Property Type-Hotel		8.685***		8.725***
		(0.834)		(0.834)
Property Type-Private		13.239***		13.193***
		(0.172)		(0.173)
Property Type-Shared		24.850***		24.745***
		(0.530)		(0.530)
End Face Mask Mandates			−0.062***	−0.026**
			(0.008)	(0.008)
Restaurants Reopening			−0.005***	−0.002***
			(0.001)	(0.001)
Constant	29.735***	31.646***	28.581***	31.121***
	(0.321)	(0.917)	(0.344)	(0.924)
Observations	104,095	104,095	104,095	104,095
R^2	0.00002	0.102	0.002	0.102
Adjusted R^2	0.00001	0.102	0.002	0.102
Residual Std. Error	21.481 (df = 104093)	20.359 (df = 104086)	21.464 (df = 104091)	20.355 (df = 104084)

Note: *p<0.05; **p<0.01; ***p<0.001

Based on the model outputs, three of our four models show statistically significant coefficients for our variable of interest.

5.1   Model 1 ~ Restricted Model
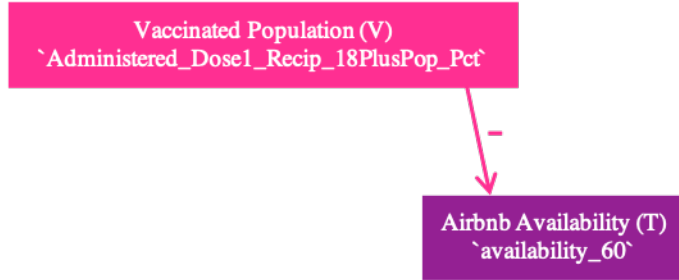
$$f(T) = \beta_0 + \beta_1 V$$

Figure 6: Model 1 Causal Graph

Our restricted model produced the least statistically significant result for our vaccination percentage coefficient out of all of our models. This is an important result because it tells us that despite the vast amount of observations we supplied to our model, ultimately it was unable to reject the null hypothesis with a p-value of 0.139. This result may give us reason to doubt any direct causal impact vaccinated population percentage has on Airbnb availability. However, our model also produced an adjusted R^2 value of 1.141e-05. This means that our restricted regression failed to capture well variation within features incorporated in this model.

5.2   Model 2 ~ Less Restricted Model - Airbnb

$$f(T) = \beta_0 + \beta_1 V + \beta_2 P + \beta_3 A + \beta_4 SH + \beta_5 S + \beta_6 PT(Hotel) + \beta_7 PT(Private) + \beta_8 PT(Shared)$$
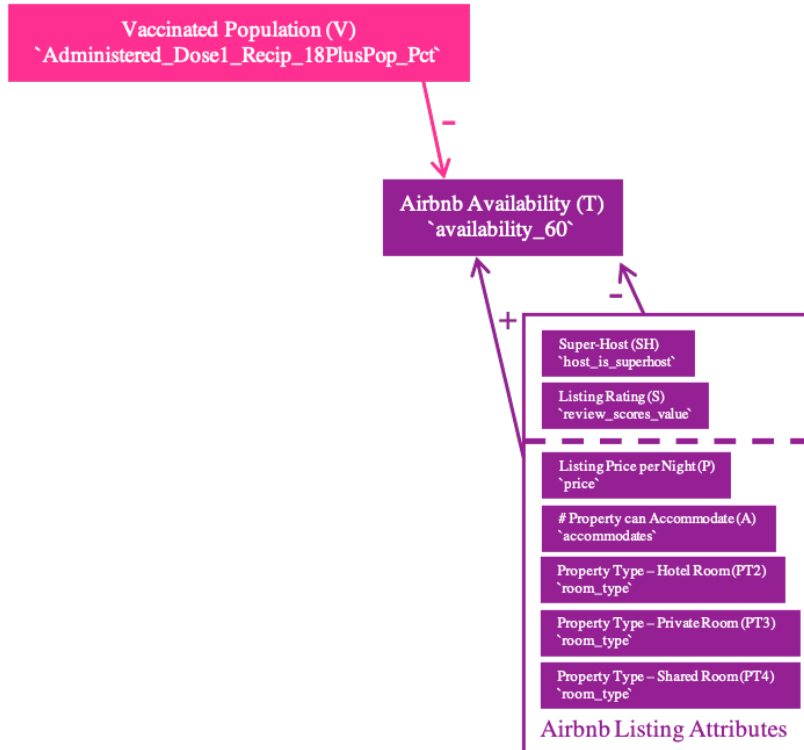

Figure 7: Model 2 Causal Graph

Model 2 produced the smallest coefficient of all of our models (-0.0521). The model successfully rejected the null hypothesis with a very highly significant p-value of 2.09e-09. And the sign of the variables are consistent with our causal theory. However, the practical significance of our result is questionable. The coefficient tells us that, holding others constant, a 20% increase in the vaccinated population percentage will likely cause a listing's availability to decrease by one day. Despite the very high statistical significance of Model 2, it appears that the impact of state-level vaccination percentage is very small. Additionally, the adjusted R^2 value is 0.102. While this is a definite improvement from Model 1, the variables in Model 2 do not explain a significant portion of variance within our input variables.

5.3 Model 3 ~ Less Restricted Model - State Policy

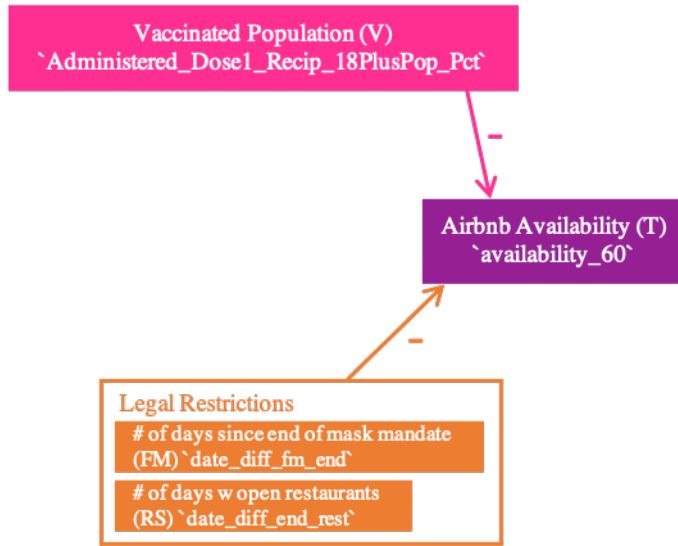$f(T) = \beta_0 + \beta_1 V + \beta_2 FM + \beta_3 RS$

Figure 8: Model 3 Causal Graph

Model 3 produced the largest coefficient of all of our models (0.045) for the vaccination percentage variable. The model rejected the null hypothesis with a very highly significant p-value of 3.97e-05 and the robust standard error of the coefficient is 0.011.

Another important aspect to note is that the sign of the vaccination percentage coefficient changed in Model 3. This is likely due to the high amount of positive correlation between our variable of interest and the state policy variables (see Figure 5). This model is important because it suggests a potentially stronger relationship between state policy and Airbnb listing availability. Additionally, the adjusted R^2 value is 0.002. While this still is an improvement from Model 1, it is much less than that of Model 2. This indicates that Model 2 is summarizing the variance within the data better than Model 3.

5.4   Model 4 ~ Full Model

$$f(T) = \beta_0 + \beta_1 V + \beta_2 P + \beta_3 A + \beta_4 SH + \beta_5 S + \beta_6 PT(Hotel) + \beta_7 PT(Private) + \beta_8 PT(Shared) + \beta_9 FM + \beta_{10} RS$$
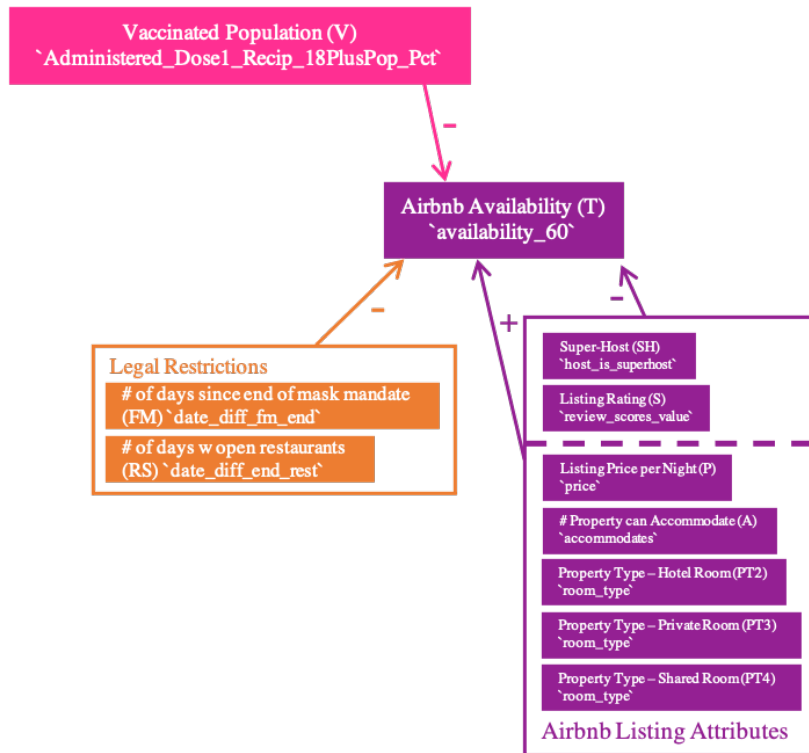

Figure 9: Model 4 Causal Graph

Model 4 rejected the null hypothesis with a significant p-value of 0.022. However, it also produced the most problematic robust standard error of all of our models (0.0104) for the vaccination percentage variable coefficient (-0.024). This means that our model is failing to capture variance specifically within our variable of interest. Interestingly, the adjusted R^2 value is 0.102, which is the same as that of Model 2. While we do see a lack of improvement in adjusted R^2 value, by including all features, Model 4 provides the most comprehensive representation of the real world of all our models.

6   Model Assumptions

As the vaccinated population percentage has 122 unique values and the full dataset has 104,095 listings, our model should fulfill large sample assumptions. We will discuss whether our model satisfies the large sample model assumptions as well as the additional assumptions of the Classical Linear Model ("CLM").

6.1   Large Sample Model Assumptions

6.1.1   IID

Inside Airbnb draws listings from 28 cities of 18 states in the United States. This list includes the most popular tourist destination cities in the US. In addition, because of the vast amount of data provided for each city, we are comfortable that the listings are representative of all the listings within each city. Having said that, we believe there may be a clustering effect given that a few states account for the majority of the listings (see Appendix B). This is also consistent with AirDNA's study result as of June 2019[14].

---

[14]CDN, "The Most Popular Cities for Airbnb in the U.S."

6.1.2 Unique BLP/No Perfect Collinearity

There is no perfect collinearity as no variables were dropped from the model.

6.2 Additional CLM Assumptions

6.2.1 Linear Conditional Expectation

To assess linear conditional expectation we evaluate a plot of the residuals vs fitted values for each model. To confirm linear conditional expectation there should be a relatively straight line at zero.
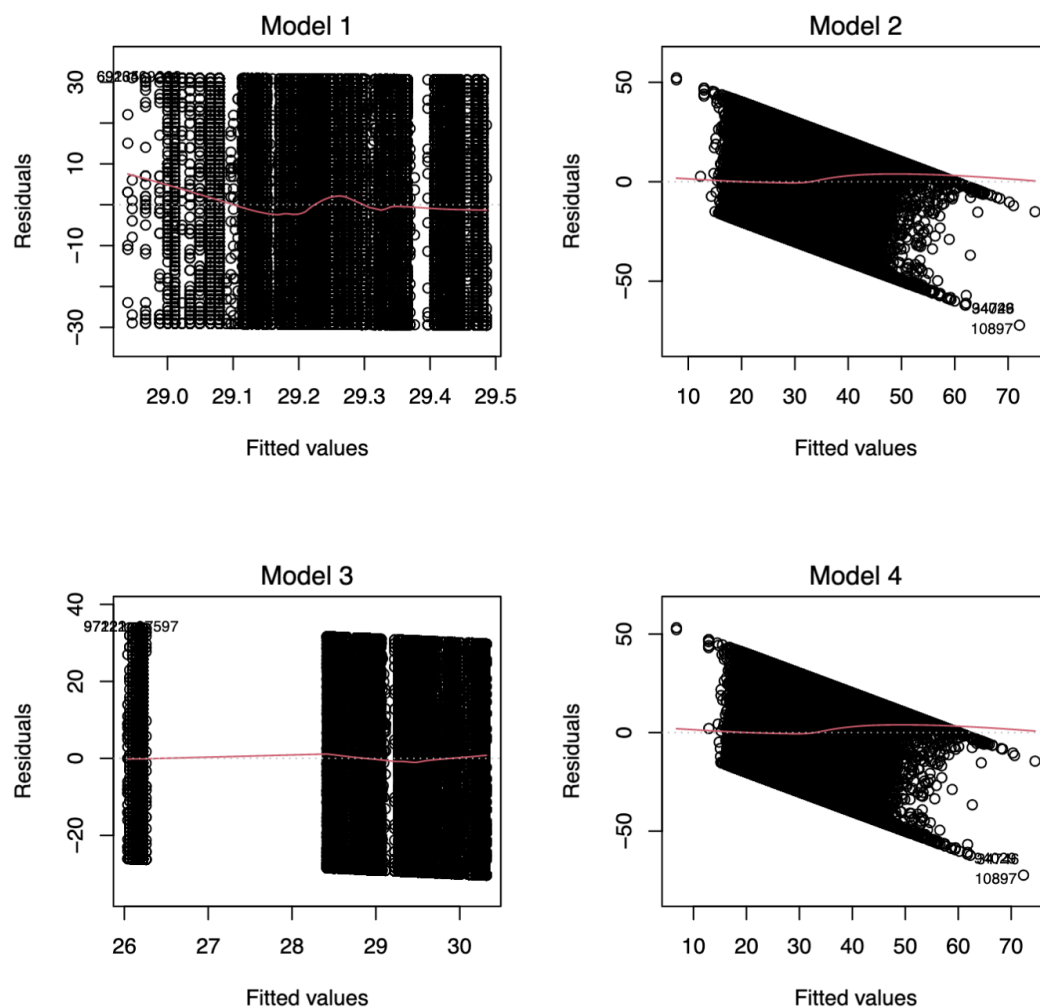

Figure 10: Residuals vs Fitted Plots

Model 3 has the most flat line at zero, with Models 2 and 4 as close seconds. Because we have fulfilled the first three assumptions of the CLM, we can be sure that the coefficients in Models 2-4 are unbiased.

6.2.2 Homoskedasticity

There are three methods to assess homoskedasticity in each model. First, we can leverage the residual vs fitted plots (above) to look for constant variance around each line. It appears that the variance around each line is inconsistent, giving us reason to doubt the presence of homoskedasticity. Second, we can look for a straight line in the scale-location plot (below).
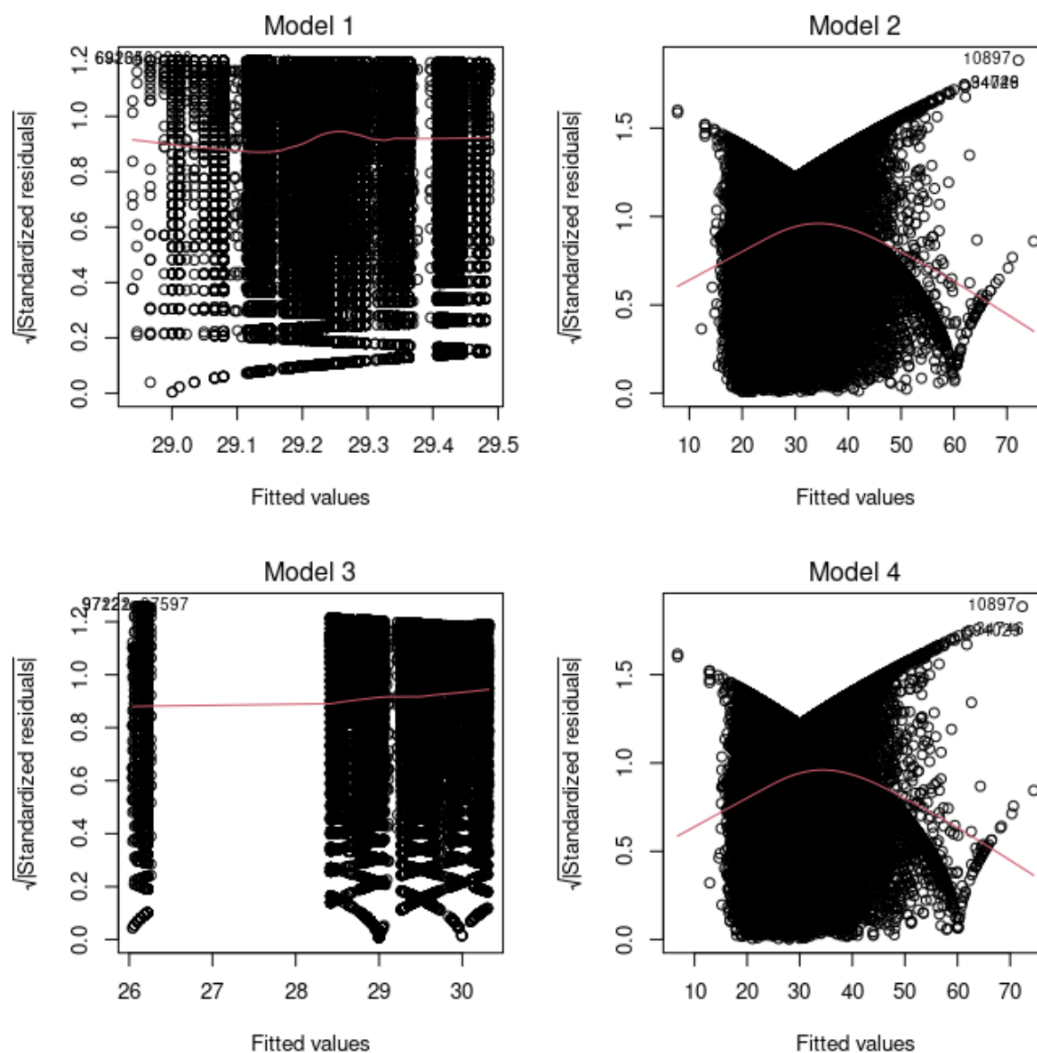
Figure 11: Scale-Location Plots

It appears that Model 3 produces the flattest line of all models, with Model 1 as a close second. The lines for Models 2 and 4 seem to have a consistent bow shape. Finally, the Breusch-Pagan test is a valid statistic which tests the following null hypothesis: the data is homoskedastic. However, because our data is very large, the Breusch-Pagan test produces highly statistically significant results that are not supported by the other two assessment methods.

Regardless of whether homoskedasticity is truly in our model, we can overcome any lack of homoskedasticity in the data by using robust standard errors.

17

6.2.3 Normally Distributed Errors

To assess for normally distributed errors, we can simply plot the histogram of each model's residuals and look for a normal distribution.

Figure 12: Residuals Plots

It appears that none of the model residuals have a particularly normal distribution.

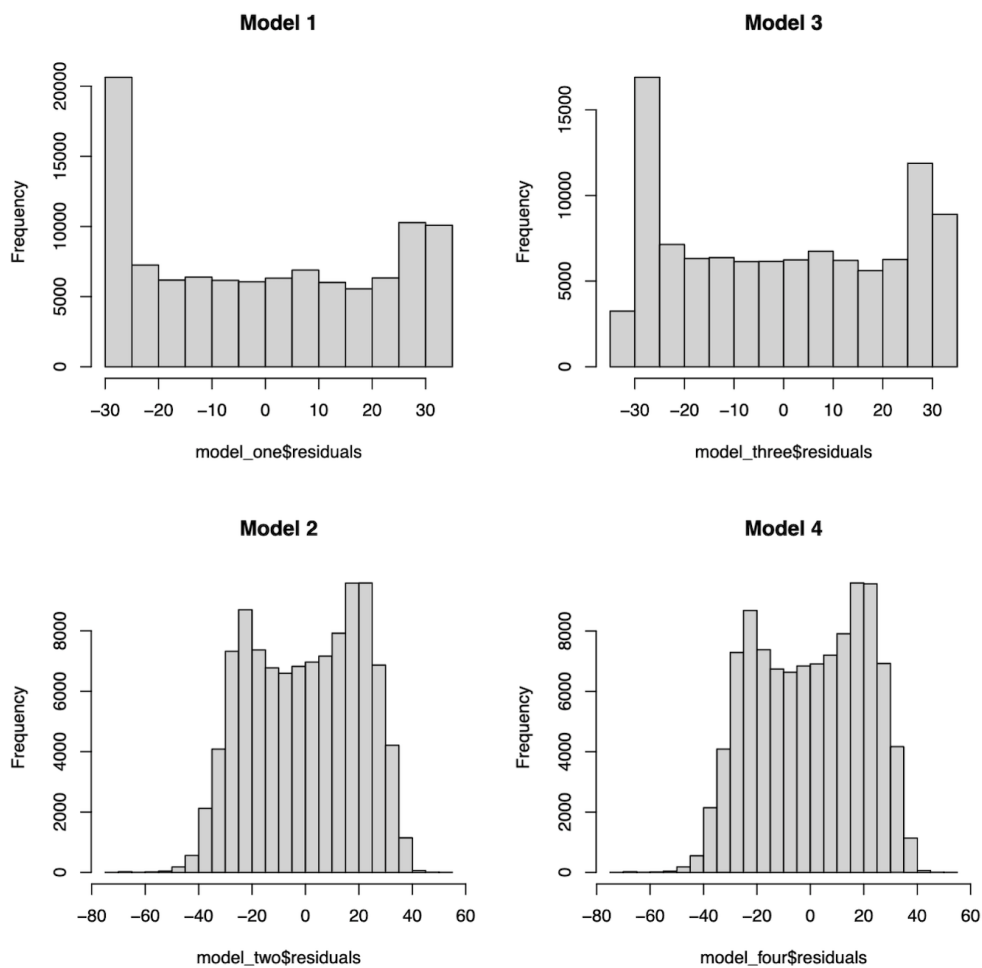Again, even though our models' ability to fulfill the strict assumptions of the CLM are questionable, our models take a large number of observations, allowing us to use the large sample assumptions to confirm the asymptotic properties of our model variables.

7   Omitted Variables Discussion

See Appendix C for our imagined 'True Model'

While we tried to include every variable available that may impact a listing's 60 day availability, we believe there are additional variables not included that may cause our estimated coefficients for vaccinated population percentage to be biased. Please note we are using the full model coefficient (-0.024) to assess omitted variable bias.

7.1   Natural Disasters (N)

Tourists may choose to avoid certain cities if there is a high risk of extreme weather. For example, wildfires in California or hurricanes in Florida. We believe that natural disasters would increase the 60 day availability and also decrease the vaccinated population percentage. When we conduct the thought experiment, we find bias moving away from zero, suggesting that our estimated coefficient for vaccinated population percentage may be overestimating the negative impact of vaccinated population percentage on Airbnb availability.

7.2   Gas and Flight Prices (G)

Gas and flight prices may impede tourists' ability to travel in-state. We believe that increased gas and flight prices should increase Airbnb availability as well as decrease the vaccinated population percentage. When we conduct the thought experiment, we find bias moving away from zero, suggesting that our estimated coefficient for vaccinated population percentage may be overestimating the negative impact of vaccinated population percentage on Airbnb availability.

7.3   International COVID Restrictions (TR)

Recently we have seen an increase in COVID-19 cases, largely due to unvaccinated populations and increasingly contagious variants. Because of this, we have seen increased COVID-19 restrictions from European countries, such as Spain, Greece and France[15]. We expect that increased international COVID-19 restrictions will decrease Airbnb availability within the United States and will increase the vaccinated population. When we conduct the thought experiment, we find bias moving away from zero, suggesting that our estimated coefficient for vaccinated population percentage may be overestimating the negative impact of vaccinated population percentage on Airbnb availability.

7.4   Traveler Sentiment (TS)

Due to the COVID-19 pandemic many people have felt unsafe to travel. Recently, it appears that travelers have felt increasingly pessimistic about travel safety due to the new Delta variant[16]. As travel sentiment increases, we expect less available Airbnbs and an increased vaccinated population. When we conduct the thought experiment, we find bias moving away from zero, suggesting that our estimated coefficient for vaccinated population percentage may be overestimating the negative impact of vaccinated population percentage on Airbnb availability.

7.5   Employment (E)

Fluctuating levels of employment is a reality that should be accounted for. It appears that employed Americans are taking less and less vacations[17]. Because of this, we believe that high levels of employment will increase Airbnb availability. Additionally, we believe that those who are employed are more likely to get vaccinated. When we conduct the thought experiment, we find bias moving towards zero, suggesting that our estimated coefficient for vaccinated population percentage may be underestimating the negative impact of vaccinated population percentage on Airbnb availability.

---

[15]New York Times, "The New Reality of a European Trip: 'Things Are Going to Change'"

[16]Destinationanalysts.com, "Update on American Travel Trends & Sentiment—Week of July 26th"

[17]Harvard Business Review, "The Data-Driven Case for Vacation"

8 Conclusion

Our study was made to explore our causal theory that a state's vaccinated population decreases Airbnb availability with that state. Even though the majority of our models presented highly statistically significant results, the estimated coefficients produced were minute. Therefore, we are not yet confident that a causal relationship between these two variables exists.

9 Limitations

We acknowledge there may be several limitations to our study.

First, we believe that measuring a state's vaccinated population percentage may be too broad a unit of measurement. For example, in California there is lots of intra-state travel from San Francisco to Los Angeles, two cities with very different rates of COVID-19 vaccinations. Because we measured vaccinations on the state level, our study fails to account for the nuances of intrastate travel and may not be precise enough to accurately estimate any causal relationship within larger states. Future studies may wish to collect vaccination data on the county or city level to improve precision of measurement for our key variable of interest.

Secondly, the clustering effect within our data may violate the IID assumption, causing our study to potentially lack generalizability. Future research may wish to utilize stratified random sampling methods to gather Airbnb listings data.

Thirdly, our study was limited by our use of each listings' `last_scraped` date to approximate the date of a listings' bookings. We acknowledge that this was an imprecise methodology despite our best efforts. To improve the relevancy of the dates we pull for vaccinated population percentage, it would be hugely beneficial to collect data on Airbnb bookings, rather than listings. This would allow us to make a better judgement on traveler behavior when they decide to book a listing.

Fourth, we believe that the combined effect of clustered data by state and dates caused a large portion of our data to lack sufficient variation. We believe that inconsistent variance in some of our variables caused our adjusted R^2 values to be relatively small. Furthermore, the omitted variables may have more explanatory power to the outcome variables than we currently identified, this may change our model result significantly. We recommend that future researchers collect more data over a longer period of time to add more variation. to the data.

Fifth, Airbnb is only one company in the entire tourism industry. Certain cities, such as Las Vegas, may favor traditional hotels over Airbnb. Furthermore, Airbnb tends to be used most frequently on the east and west coasts of the United States[18]. Therefore using Airbnb data may exclude important tourism information for states in the middle and south of the US.

Finally, the study does not explicitly take into consideration the recent change of trends in Airbnb usage. According to a recent Airbnb blog, given the impact of COVID-19, Airbnb stays show increasing trends for living, working and connecting safely away from home more than tourism. If this trend applies broadly to our Airbnb data set, that would challenge our original thesis to use Airbnb availability as an indicator to tourism.

Despite our inclusive results, we still believe that continued investigation of causal relationships between COVID-19 and tourism could yield important information to help the tourism industry recover from the effects of the pandemic. Furthermore, we believe it may be possible to leverage this causal relationship to encourage the tourism industry to put their support behind COVID-19 vaccination programs and other state policies. We also believe that continued study focusing on a causal relationship between COVID-19 vaccinations and state policy may be useful to lawmakers in states whose economies rely heavily on tourism.

---

[18]ibid

10   References

AirDNA, "How to Optimize Your Airbnb Occupancy Rate with Data Analytics"

Airbnb, "Reviews for stays"

Airbnb, "What is a Superhost?"

AJMC, "A Timeline of COVID-19 Vaccine Developments in 2021"

Apnews.com, "More states ease lingering virus rules as vaccine rates rise"

Cox, "Inside Airbnb"

CDC, "COVID-19 Vaccinations in the United States,County"

CDN, "The Most Popular Cities for Airbnb in the U.S."

Deane, "2021 Airbnb Statistics: Usage, Demographics, and Revenue Growth"

Destinationanalysts.com, "Update on American Travel Trends & Sentiment—Week of July 26th"

Fortune.com, "Airbnb's IPO filing reveals huge COVID impact"

Harvard Business Review, "The Data-Driven Case for Vacation"

Leidman E, Duca LM, Omura JD, Proia K, Stephens JW, Sauber-Schatz EK. COVID-19 Trends Among Persons Aged 0–24 Years — United States, March 1–December 12, 2020. MMWR Morb Mortal Wkly Rep 2021;70:88–94. DOI: http://dx.doi.org/10.15585/mmwr.mm7003e1external icon. link
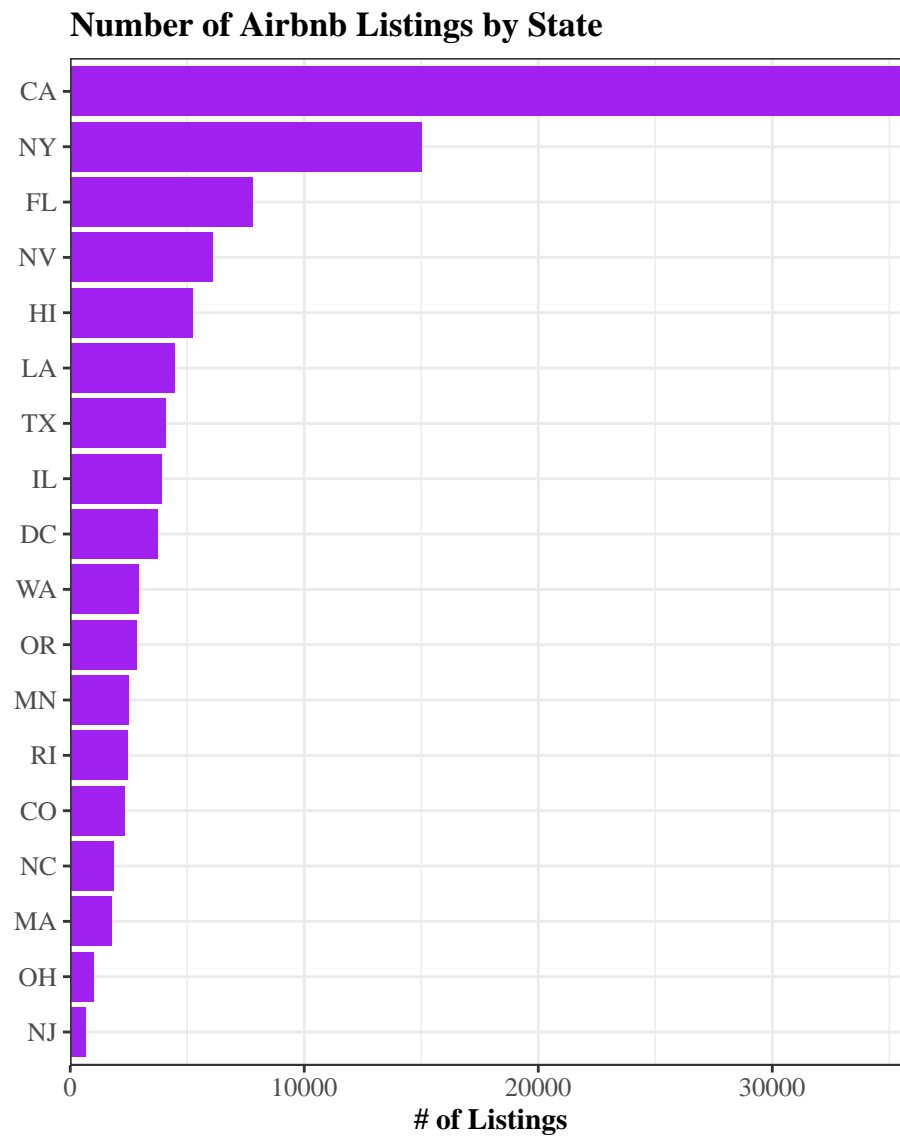
New York Times, "The New Reality of a European Trip: 'Things Are Going to Change'"

Raifman, Julia, Nocka, Kristen, Jones, David, Bor, Jacob, Lipson, Sarah, Jay, Jonathan, . . . Galea, Sandro. COVID-19 US State Policy Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-11-02. https://doi.org/10.3886/E119446V40 link

A   Codebook

Variable Name	Variable Code	Description	Data Type
Airbnb Availability (T)	`availability_60`	The availability of the listing days in the future as determined by the calendar	Numeric (int)
Vaccinated Population (V)	`administered_dose1-recip_18plus_pct`	Vaccination rate of the state 14 days prior to the Airbnb scraped date	Numeric (%)
Listing Rating (R)	`review_scores_value`	Ratings from guests	Numeric (int)
Price Per Night (P)	`lprice`	Log of daily price in USD	Numeric (float)
Super-Host (SH)	`host_is_superhost`	Whether the host is a superhost	Boolean
Property Type (PT)	`room_type`	Entire home/Apt, Private Room, Shared Room or Hotels	Categorical
Maximum Capacity (A)	`accommodates`	Maximum capacity of listing	Numeric (int)
Number of Days Since the End of Face Mask Mandates (FM)	`date_diff_fm_end`	Days between the listing's last scrapped date and face mask policy removal date	Numeric (int)
Number of Days Since Restaurants Reopened (RS)	`date_diff_end_rest`	Days between the listing's last scrapped date and restaurant open date	Numeric (int)

B   Airbnb Listings by State

Number of Airbnb Listings by State

C   True Model?

Vaccinated Population (V) `Administered_Dose1_Recip_18PlusPop_Pct`

Traveler Sentiment (TS)

International COVID Restrictions (TR)

Employment (E)

Natural Disasters (N)

Gas/Flight Prices (G)

Airbnb Availability (T) `availability_60`

Legal Restrictions
# of days since end of mask mandate (FM) `date_diff_fm_end`
# of days w open restaurants (RS) `date_diff_end_rest`

Super-Host (SH) `host_is_superhost`

Listing Rating (S) `review_scores_value`

Listing Price per Night (P) `price`

Property can Accommodate (A) `accommodates`

Property Type – Hotel Room (PT2) `room_type`

Property Type – Private Room (PT3) `room_type`

Property Type – Shared Room (PT4) `room_type`

Airbnb Listing Attributes