

Perceived Ethnicity and Hosts' Approval of Private Room Vacation Rentals

Team: Blake Bormes, Maddie DiLullo Byrne, Joe Houck, Raul Merino

Introduction

Does perceived ethnicity impact the ability to book listings on Misterb&b? The American Fair Housing Act has been fighting discriminatory offline housing markets since 1968. However, platforms like Airbnb and Misterb&b have operated in a legal grey area when it comes to discrimination. Stories from the hashtag movement AirbnbWhileBlack on social media, BLM, and academic experiments showcase user discriminatory effects on the platform. Driven by these studies and experiences, our study aims to better understand the potential biases that users experience within the LGBTQ+ platform Misterb&b. Since the biggest minority within America is the Hispanic population, our study aims to further the analysis of previous studies on discrimination towards African Americans and understand bias present towards Hispanic people when booking on Misterb&b in four major US regions: the South, East, Midwest, and West. Our study seeks to answer the question:

Are seemingly hispanic profiles less likely to be approved by a host than seemingly white profiles?

Background

A similar approach has been attempted in measuring discrimination towards African Americans on Airbnb. In *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*¹ Edelman, Luca and Svirsky conducted an experiment that found that “applications from guests with distinctively African American names are 16% less likely to be accepted relative to identical guests with distinctively white names.”

In their experiment, the authors inquired about the availability of roughly 6,400 listings on Airbnb across five cities. Furthermore, they created guest accounts that differ by name but are otherwise identical. This led to a grouping of response types and measurement. Approvals were documented as a 1 and all other responses (i.e., denial, no response, etc.) were listed as a 0.

After reading this study, our team determined a further application of the study to understand bias against hispanic people would derive great insights, seeing as

¹ Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics*, 9 (2): 1-22.
DOI: 10.1257/app.20160213

hispanics are the largest minority in the US, and that there is evidence that they already face discrimination in the housing market².

Our team leveraged the LGBTQ+ platform Misterb&b instead of Airbnb because of specific interest in discrimination within the LGBTQ+ community, as there is a lack of experimental research done into that particular topic.

Most previous research on discrimination within the LGBTQ+ community was done through surveys. The results of these studies found that People of Color, particularly Black men, report higher levels of Racial/Ethnic stigma in LGBT spaces³, report higher levels of discrimination when applying for a job or interacting with the police than their white counterparts⁴, and experience sexual racism within the LGBTQ+ community⁵.

Hypothesis

Our hypothesis is that a seemingly Hispanic name and profile photo will result in a lower host approval rating than a seemingly white name and profile photo.

Based on the existing data on discrimination against Hispanics in the housing market, as well as the survey reports of discrimination against People of Color within the LGBTQ+ community, we believe we will find a significant difference in the approval rates between both groups in Misterb&b.

Comparison of Potential Outcomes

Our participants were 30 male profiles, with differences in perceived ethnicity as the difference in individuals. To illustrate, 15 accounts in the control group that had the most common names listed by the SSA⁶ as a representation of names likely to be perceived as white. 15 accounts in the treatment group had names that are likely to be perceived as Hispanic, found on babycenter.com⁷.

All profiles in the treatment group had the same hispanic-looking profile photo, while all profiles in the control group had the same white-looking profile photo. We procured our

² Christensen, Peter and Sarmiento-Barbieri, Ignacio and Timmins, Christopher. 2021. "Racial Discrimination and Housing Outcomes in the United States Rental Market". National Bureau of Economic Research - <https://www.nber.org/papers/w29516>

³ McConnell, Elizabeth A. and Simons-Rudolph, Ashley. "People of Color Experience Discrimination Within LGBT Spaces" Community Psychology - <https://www.communitypsychology.com/people-of-color-experience-discrimination-within-lgbt-spaces/>

⁴ Discrimination in America: Experiences and Views of LGBTQ Americans. Harvard T.H. Chan School of Public Health, Robert Wood Johnson Foundation, National Public Radio 2017. https://www.rwjf.org/content/dam/farm/reports/surveys_and_polls/2017/rwjf441734.

⁵ Plummer, Mary Dianne. 2007. "Sexual racism in gay communities: negotiating the ethnosexual marketplace" University of Washington.

<https://digital.lib.washington.edu/researchworks/handle/1773/9181>

⁶ <https://www.ssa.gov/oact/babynames/decades/century.html>

⁷

https://www.babycenter.com/baby-names/most-popular/100-most-popular-hispanic-baby-names-of-2011_10363639

profile photos by using a site that offers AI-adjusted photos⁸ to create profile pictures to avoid any elements of recognition and privacy protection. We refreshed the site so we could gather approximately 10-15 sample profile photos that we felt could be appropriate for our treatment and control groups, and then voted amongst our research group as to which photos we thought suited our treatment & control demographic.

The ROXO comparison for the treatment and control groups are as follows:

Treatment: R X O

Control: R X O

The R - Randomization is the assignment of a participant to a host's stay. The X - intervention is the participant sending a message to the host requesting a stay, and the O - observation is whether or not the host approved or denied the stay request, didn't respond to the request, or asked for more information.

The observations for the treatment and control groups will be directly compared to each other to understand the difference in types of responses and amount of approvals between the two groups.

Randomization Process

We first created a spreadsheet that included more than 180 "private room" listings. These listings are in the United States, specifically sampling hosts from each part of the country (i.e. Midwest, East, South, and West). When creating this list, we randomly selected listings for private rooms in our 4 regions that all required host approval to book and ensured there were no listings that had the same host. We also randomized whether a host received a message from the control (seemingly white) or treatment (seemingly hispanic) group, by randomly assigning each of the listings to one of our assigned profiles. We blocked by region, so each of those had half of their listings assigned to control and half to treatment. We also performed a few balance checks and found no significant difference between both groups in terms of price per night, rating nor the number of listings that had been booked before. We did however find that there was a difference in the average number of bookings between both groups, however the medians were very similar, so we decided this randomization was sufficiently balanced.

Table 1: Pre-treatment covariate values for both control and treatment groups.

| Covariate | Control (White) | Treatment (Hispanic) |
|----------------------------|-----------------|----------------------|
| Avg. Price per Night (USD) | 150.5889 | 151.0667 |

⁸ <https://this-person-does-not-exist.com/en>

| | | |
|----------------------------------|------------|----------|
| Avg. Rating (1-5) | 4.93245921 | 4.92058 |
| Has Been Booked Before | 67.78% | 76.78% |
| Avg. Number of Bookings | 11.43333 | 17.95556 |
| Median Number of Bookings | 5 | 6 |

Additionally, we wanted to perform this study for a specific persona: a gay male that was either white or hispanic and between the age of 25-30. We randomized the ages assigned to each of the 30 profiles, while ensuring equal representation of ages.

Treatment

Keeping all else equal, the treatment is the change in profile name and profile picture.

In order to create the Misterb&b profiles, we collectively chose the values that we would keep constant for all of our accounts: last name (Caucasian: Jones, Hispanic: Garcia), home city (Dallas), pronouns (He/Him), gender identity (Male), LGBTQ+ community connection (Gay). We also kept the message to hosts the same. The message read, "Hi I'm looking to visit [insert city name] from August 19 to 21. Could I book a stay at your listing?" To avoid flagging Misterb&b of our creation of fake profiles, we randomized the profile ages within a limited range (25 to 30 years), but this was not information that was offered to the hosts during the time of reviewing our requests. We also opted to make the profile photos private to avoid members of the Misterb&b community from realizing that we had multiple profiles that had identical profile photos.



Images: Profile photos used for our treatment (left) and control (right) groups.

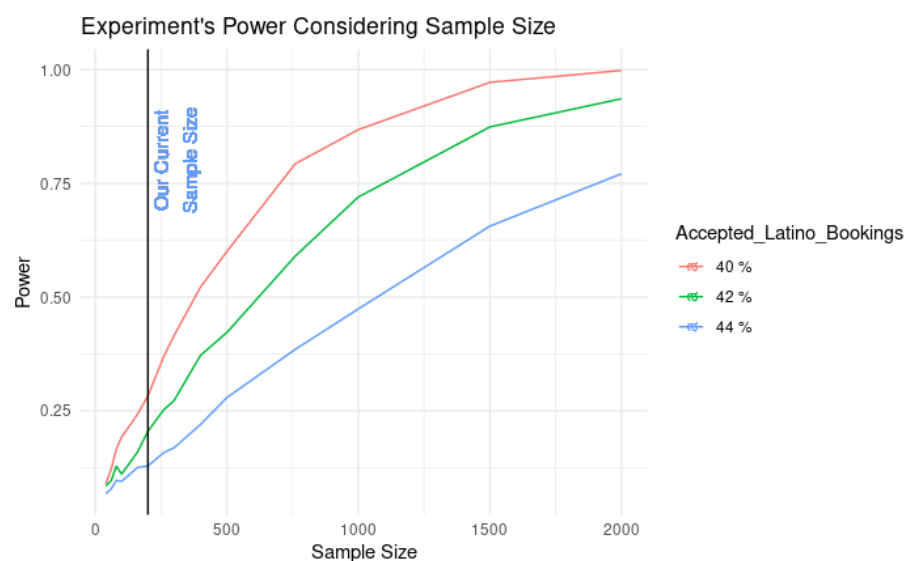
Consort

For this study, we limited our study participants to those who had a “room share” listing for a “private room” available on Misterbandb.com between August 5th to 7th within the United States. We also only used listings managed by individuals, rather than corporations such as hotels. Furthermore, we avoided properties that had an instant booking option. We made these decisions because our experiment relied on the social perception and response of our profiles which was unlikely to occur with these listing types.

Power Calculation

When we conducted our original assessment of our experiments power, we found that if our analysis was going to be similar to *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*⁹, we were likely going to be running an underpowered experiment. When modeling our treatment effect to reflect theirs for our power analysis, we found that our experiment would need approximately 1500 data observations or data points in order to be a sufficiently powered experiment.

Power Assessment: As shown in the image below, our power analysis indicated that our experiment was considerably underpowered in our sample size, regardless of the spectrum of acceptance rate of Hispanic users. We assumed a 50% approval rate for White profiles, and tested three possible scenarios for Hispanic profiles, with 44%, 42% and 40% approval rates.



⁹ Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics*, 9 (2): 1-22.
DOI: 10.1257/app.20160213

Figure 1: Required sample size for our power analysis scenarios

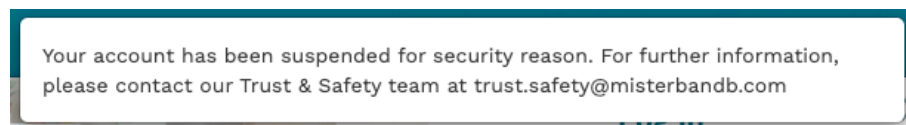
Although we were aware of our underpowered experiment, we decided to continue our experiment with the sample size that we had originally intended. The reasons why we decided to maintain the same sample size was due to a few factors:

1. Limited number of listings available

Unlike Airbnb, Misterb&b has considerably fewer bookings worldwide, and we found that in different geographical areas (i.e. Midwest) it was considerably more difficult to find listings and hosts that fit our criteria. This could be due to political beliefs (i.e. potentially fewer people who are openly part of the LGBTQ+ community) or decreased familiarity or popularity of the site in those areas.

2. Cybersecurity issues with account authentication

Another barrier that we encountered was the security measures imposed for our account creation. When creating the Misterb&b accounts, we needed to provide methods of authentication including a unique email and phone number for each profile. We found that after creating 20 (of our originally planned 40 accounts) we started experiencing problems with using Gmail as our email platform, and eventually we started to resort to using email addresses and phone numbers from friends and family to complete the account creation. Another hurdle we encountered was before we sent messages to hosts, Misterb&b started automatically flagging some of our accounts as fraudulent. Based on our final efforts, we think that the main element that prompted them to disable the accounts we were creating was the photo that we were uploading to create the profiles, however we are not certain.



Pictured: The fraudulent activity alert we would receive when creating our accounts

3. Time and Resources

Another constraint of our experiment was trying to determine what would be a reasonable amount of time and effort to find suitable listings that fulfilled our criteria. As mentioned, we found that the United States didn't have enough listings to fulfill our power requirements, so we would have had to either broaden our scope of listings (i.e. not just private rooms) or broaden our scope to international listings. Additionally, we were able to use a script to send our messages to the listings, but the process of account creation for email, phone

number verification, and Misterb&b account creation was time consuming, and if we found we needed to create more profiles to conduct this experiment, we would need dedicated resources.

After conducting our experiment, we believe that our experiment had more power than we original anticipated for a few reasons:

1. Adding photos to our profiles & requests

Within the aforementioned study, they conducted their experiments with only the change in name for their profiles, whereas we needed to include photos in order for the Misterb&b profiles to be configured. While we are not able to account for the coefficient of power that the photo provided, it seems safe to assume that including the photos in our experiment likely made our profiles seem more realistic, and improved our response percentage compared to the original study.

Analysis and Results

All 180 messages were sent on a Sunday night over a month before the requested dates through an automated selenium script that logged into all 30 accounts and sent 6 messages from each account to each of the listings according to our randomization.

Only 1 host was unable to be reached, as the listing was booked in between the sampling and when the messages were sent, which caused our script to fail in that case. Since we were unable to measure their response, we consider it as attrited and it won't be part of our analysis.

In the table below it can be seen how many messages were successfully sent for each group, and how they were distributed across regions.

Table 2: Count of messages sent by region and group.

| Region | Control (Seemingly White) | Treatment (Seemingly Hispanic) |
|----------|------------------------------|-----------------------------------|
| West | 23 | 23 |
| Mid-west | 21 | 20 |
| East | 25 | 24 |
| South | 21 | 22 |
| Total | 90 | 89 |

The variables we used to categorize our findings were responded (boolean of whether the host responded to the inquiry), response (if the host responded positively, neutrally, or negatively), and the response time (number of days the host took to respond). To categorize the response of the host, we utilized the preapproval function that Misterb&b offered, which allowed hosts to unconditionally approve or decline the request. If a host responded to the inquiry requesting further information or clarification, we categorized these responses as neutral.

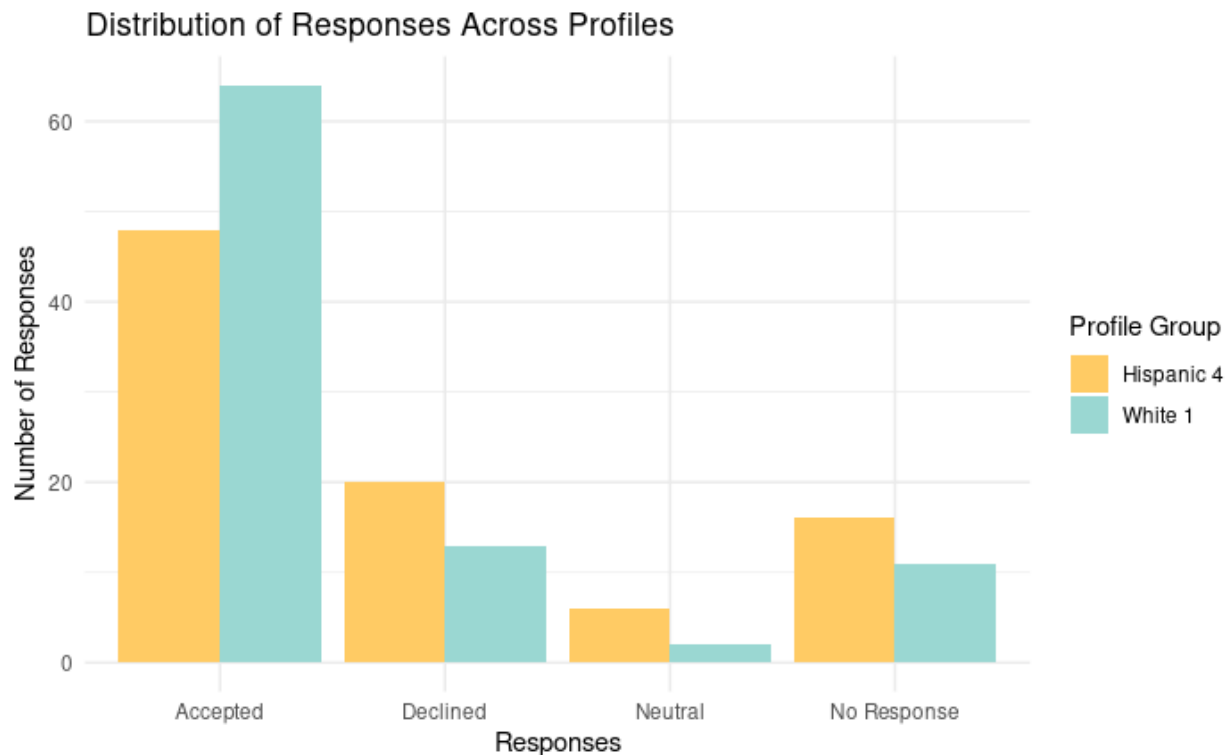


Figure 2: Distribution of Responses across both groups

As described in the figure above, those in the Control Group (seemingly white) received substantially more approvals than the Treatment Group (seemingly hispanic). To illustrate more closely, the Control Group received a ~71.1% approval rating (64 of 90 messages) and the Exposed Group received a ~52.8% approval rating (47 of 89 messages). Furthermore, the Treatment Group received more rejections, more neutral responses, and fewer responses.

In order to properly test whether this difference was statistically significant, we constructed several models that intended to explain this phenomenon. The main difference between the models we constructed was our outcome variable. For all of them, we will control by region, price per night, and number of bookings.

Regression 1: Approval vs Everything Else

Our initial model focuses solely on whether each message got a positive response or not. In this model, rejections, neutral responses, and no responses would all be considered a negative response. This coding of the responses is similar to the Airbnb study we analyzed prior to the experiment and supports the assumption that generally a host would want to accept a reservation at their stay and would accept the request if they wanted to accept the requestor staying at their place.

| Yes vs Everything else models | | |
|-------------------------------|--------------------------------|------------------------|
| | <i>Dependent variable:</i> | |
| | said_yes | |
| | (1) | (2) |
| treatment | -0.208*** (0.072) | -0.199*** (0.071) |
| price_night | 0.0002 (0.0005) | |
| region_stateMidwest | 0.008 (0.104) | |
| region_stateSouth | 0.135 (0.103) | |
| region_stateWest | 0.134 (0.098) | |
| Number_Bookings | 0.001 (0.001) | |
| has_booked_before | 0.158* (0.085) | 0.188** (0.079) |
| Constant | 0.494*** (0.116) | 0.584*** (0.073) |
| Observations | 179 | 179 |
| R ² | 0.089 | 0.065 |
| Adjusted R ² | 0.052 | 0.055 |
| Residual Std. Error | 0.474 (df = 171) | 0.473 (df = 176) |
| F Statistic | 2.391** (df = 7; 171) | 6.165*** (df = 2; 176) |
| Note: | * p<0.1; ** p<0.05; *** p<0.01 | |

We can see from our initial regression (1) that the treatment is statistically significant, with a p-value of 0.00408 and an estimated effect size of -0.208, meaning that by having a seemingly hispanic profile rather than a white seeming one, you are about 20.8% less likely to be pre-approved through the Misterb&b.

None of the regions, price per night or number_bookings appear to have statistical significance. However, “has been booked before” has a statistically significant effect. This dummy variable defines whether or not the stay has previously been booked on Misterb&b. It is significant at the 10% significance level with a p-value of 0.06381. Its

coefficient indicates that if a listing has been booked before, then the probability of being pre-approved increases by 15.8%.

We ran a second regression (2) with just those two covariates (being in treatment and whether a listing has been booked before), and now “has been booked before” is statistically significant at the 5% level of significance, with a p-value of 0.01867.

We can hypothesize that “has been booked before” is significant because these hosts historically approved and hosted guests, while listings with no previous bookings may be more selective because they’re new to the platform or are not particularly active. On the other hand, one could argue that listings with no bookings are newer, and may be more willing to accept reservations as hosts are eager to acquire initial reviews and earnings. Ultimately, the data here encourages the former hypothesis.

Regression 2: Removing Neutral Responses

As part of our experimental design, we did not respond to hosts who made any inquiries regarding our potential stay at their listing, as we didn’t want to add more variability to our results by potentially responding differently to each of them. This however meant that those entries were not as representative of real interactions users may have had, as in our previous model we coded them as negatives, when in reality they could have been turned into positive responses.

| Removing neutral responses | | |
|-------------------------------------|----------------------|------------------------|
| | Dependent variable: | |
| | said_yes | |
| | (1) | (2) |
| treatment | -0.184** (0.073) | -0.174** (0.072) |
| price_night | 0.0002 (0.0005) | |
| region_stateMidwest | 0.097 (0.108) | |
| region_stateSouth | 0.145 (0.104) | |
| region_stateWest | 0.122 (0.098) | |
| Number_Bookings | 0.001 (0.001) | |
| has_booked_before | 0.181** (0.086) | 0.205** (0.080) |
| Constant | 0.471*** (0.117) | 0.590*** (0.073) |
| Observations | 170 | 170 |
| R ² | 0.082 | 0.063 |
| Adjusted R ² | 0.042 | 0.052 |
| Residual Std. Error | 0.467 (df = 162) | 0.465 (df = 167) |
| F Statistic | 2.059* (df = 7; 162) | 5.646*** (df = 2; 167) |
| Note: *p<0.1; ** p<0.05; *** p<0.01 | | |

Similarly to our previous model, both the treatment and the “has been booked before” control are statistically significant. The effect size of the treatment went down to -0.184 and its p-value went up to 0.01229, while the “has been booked before” control is now even more statistically significant, with a p-value of 0.03605.

This new model does not really provide us with any new information, but it does show that the effect of the treatment persists even if we remove the neutral responses.

Regression 3: Let’s just look at the Approvals vs Rejections

As mentioned before, Misterb&b gives hosts a 24 hour window to respond to requests. If they do not respond within that time frame, then the request expires. We gave hosts the opportunity to respond for a week before categorizing them as a “No Response”, as we were more interested in their intention than their ability to respond within 24 hours.

Roughly 14.5% of hosts never responded during the one-week window. This failure to respond included messaging, rejecting or pre-approving the request. This lack of response can be either interpreted as hosts never seeing our message, or as them choosing not to respond, the latter potentially interpreted as a negative response.

Because of that uncertainty, we wanted to construct a final model that only looked at unambiguous responses: either the host pre-approved the request or they rejected it.

| Approvals vs Rejections | | |
|-------------------------|-----------------------------|-----------------------|
| | <i>Dependent variable:</i> | |
| | said_yes | |
| | (1) | (2) |
| treatment | -0.148** (0.070) | -0.139** (0.069) |
| price_night | 0.0005 (0.001) | |
| region_stateMidwest | 0.067 (0.106) | |
| region_stateSouth | 0.074 (0.100) | |
| region_stateWest | 0.016 (0.095) | |
| Number_Bookings | 0.003* (0.002) | 0.003** (0.001) |
| has_booked_before | 0.076 (0.086) | |
| Constant | 0.632*** (0.121) | 0.792*** (0.051) |
| Observations | 144 | 144 |
| R ² | 0.071 | 0.053 |
| Adjusted R ² | 0.023 | 0.040 |
| Residual Std. Error | 0.417 (df = 136) | 0.413 (df = 141) |
| F Statistic | 1.475 (df = 7; 136) | 3.961** (df = 2; 141) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 | |

Even though removing non-responses reduced our sample size and the power of our experiment, we still found the treatment (seemingly Hispanic) to have a statistically significant effect on the probability of getting pre-approved by the host, rather than being rejected.

What is interesting about this third model is that unlike the previous two, the “Has been booked before” dummy variable is no longer statistically significant. By removing neutral and non-responses, it seems knowing whether a booking has been booked before is no longer relevant.

In its place, however, we see that the number of bookings (the number of bookings previously made at the stay) is now statistically significant. The probability of a request being approved increases by 0.3% for each time that listing has been booked before. This effect makes sense, as listings with fewer bookings are probably more selective than those with more bookings.

Conclusion

As we can see in all three regressions, the estimated effect of having a seemingly hispanic profile (instead of a seemingly white profile) is statistically significant independent of how we measure our outcome variable.

Across all of our regressions, the estimated treatment effect was as high as 20.8% when looking at Approvals vs Everything Else, and as low as 13.9% when looking at just Approvals vs Rejections. In either case, our results were not only statistically significant, but also practically significant.

Our data shows that it is more difficult for a user with a hispanic name and picture to be approved when requesting a stay at a private room in Misterb&b, when compared to a user with a white name and picture.

Additional Research & Studies

To further extend this research study, we could leverage a larger sample size of listings within specific regions to identify if there's a statistically significant effect within regions. Our data showed that Southern and Western regions tended to have a higher approval rate of Hispanics, although not statistically significant. Another possibility is to change the format of the study to include African American-perceived names rather than Hispanic names.

One key limitation to this study was the availability of listings to choose from. Certain cities had very limited booking options, so identifying stays within regions proved challenging. A second limitation to our study is leveraging the same profile photo for each participant in the treatment (seemingly hispanic) or control (seemingly white) group. If there was a general dislike for the Hispanic photo, our experiment would not

have identified any general distaste for the picture. Rather, our experiment shows a negative sentiment towards Hispanics booking.

Ethical considerations regarding participants are a concern and potential limitation of the study. One can reasonably argue that the study did not receive the hosts' consent to participate, misled hosts and wasted their time. However, we designed the study to limit the negative impact to participants while pursuing our goal of measuring potential discrimination and injustice. To do this, we only sent one message to each host rather than having a conversation to limit time wasted. Second, the sent message is only asking for pre-approval, rather than making a reservation and canceling the booking. Third, we sent the message a month before the potential booking, giving the host plenty of time to find someone else to rent the listing. Finally, one should consider the alternative. If we alerted participants ahead of time, we may not be able to measure the hosts' natural, subconscious actions because they know they're being monitored.

In order to fight discrimination within the Misterb&b platform, the platform could consider removing the option for hosts to view a requestor's name and profile photo, and only view "clean" requestor reviews from other hosts that don't mention names or other characteristics that could be tied to a person's ethnicity. This would ensure hosts don't discriminate against minority requestors when accepting stay requests.