**EXECUTIVE SUMMARY**

Property that has shared spaces, and share bathrooms have lower listing prices. Hosts with more experience tend to charge a higher price. The population of the city also negatively impacts the listing price due to fierce competition. White hosts do not have any bias on pricing.

Regarding the demand for properties, there is a higher demand for hosts that make multiple listings, have long experience, are of high age, are male, and have properties located in big cities. There is a lower demand for properties that are not rated, have share areas/bathrooms, include cleaning fees, and are owned by black hosts. There tends to have less demand for 5-star rating properties, which could be explained by their higher price.

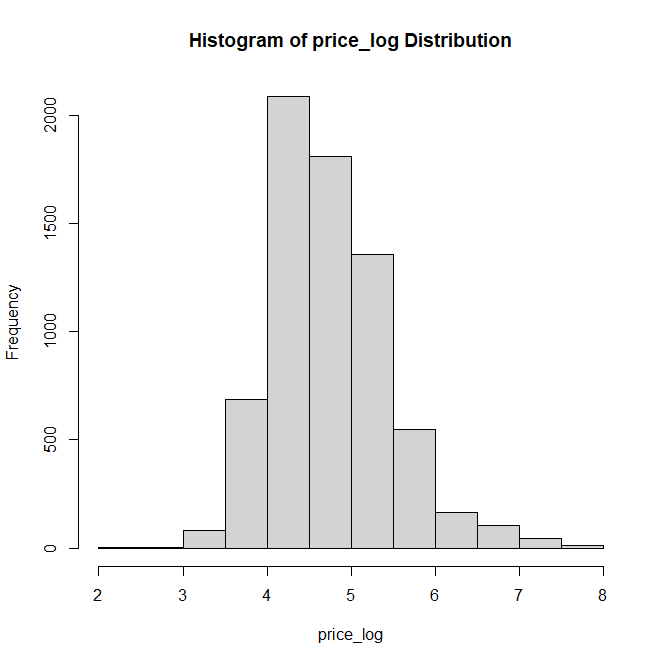
Host acceptance rates are also impacted by guest gender and race. White female guests have a 44% higher chance to be accepted by hosts, while that of black females is -35%. Hosts likely to say yes to white male guests (9% higher chance) than black male guests, who are 18% less likely to be accepted.

**ANALYSES**

**Goal 1: Examine what factors affect Airbnb listing prices**

**Outcome variable selection**

Chart, histogram

Description automatically generated 

* The variable ‘price’ follows power law distribution which is not suitable for running regression.
* The variable ‘log\_price’ follows a normal distribution, we use this variable to predict the outcome variable.

**Initial regression model**

|  |  |
| --- | --- |
|  | Observation on variables:   * ‘Shared\_property’ has a negative impact on the price of the listing house. * ‘Log\_host\_experience’ implies that hosts with more experience charge higher prices. * ‘Log\_population’ negatively impacts price, which could be due to competition in the big city   The Adjusted R-squared indicates that the model explain 0.4474 of the ‘log\_price’ |

**Model diagnostics**

|  |  |
| --- | --- |
|  | Residual vs fit and Residual vs Leverage charts all look good.  Normal Q-Q: most parts of the data follow the normal distribution while some data deviate away from the line, indicating that the data is skewed or has heavy tails.  Residuals vs leverage chart indicates that there are some outliers but not significantly impact the model (given the cook’s distance line didn’t show up in the chart) |

**Multicollinearity**

|  |  |
| --- | --- |
| **VIF stepwise with the threshold of 5**  There is 1 variable that has a collinearity problem with the host\_race\_white | **Lasso model with lambda = 10**  Beta of all variables equal zero which indicate that all variables are eligible for the model |
| **Output:** | **Output:** |

**Regression model without “host\_race\_white” variable model**

|  |  |
| --- | --- |
|  | **Observation**  The adjusted R-squared stay the same, indicates that the “host\_race\_white” does not help improve model performance |

**Goal 2: Analyze how prices and other characteristics impact the total number of guests.**

**Poisson regression “total\_guests” ~ “log\_price”**

|  |  |
| --- | --- |
| **Result** | **Interpretation** |
| **Observation:**  Price negatively impacts demand. One unit increase in log\_price will reduce demand by 3.4% | |

**Poisson regression “total\_guests” ~ all other characteristics variables**

|  |  |
| --- | --- |
| **Multivariate Poisson regression of “log\_total\_guest”** | **Interpretation**    **Observation:**   * Some of the characteristics that greatly improve the demand are multiple listings, long host experience, host age, host gender M/MF/MM, and locations in the big city. * Some of the characteristics that reduce the demand are price, not\_rate, shared property, shared bathroom, cleaning fee, and host race black. * Interestingly, a five-star rating reduces the demand, which could due to the high price of high-rating properties. |

**Compare model fit between one variable and multivariable Poisson regressions**

|  |  |
| --- | --- |
|  | **Observation:**  Multivariable Poisson regression better explain the impact of other variables on demand |

**Goal 3: Examine whether there is discrimination of host response towards gender and race**

**Probability of host accepting the request by guest gender and race**

|  |  |
| --- | --- |
| **Results**: | **Observation**:   * Gender: Female customers have higher acceptance rates than male customers. * Race: Black customers have lower acceptance rates than white customers * Hosts tend to favor white female customers and are less likely to accept the request by black male customers |

**Logistic regression to explain the outcome variable (host\_response\_yes) using guest characteristics**

|  |  |
| --- | --- |
| Result: | **Observation**:   * Guest\_race\_white has higher acceptance change than that of their counter parts. * Guest\_gender\_male is less likely to be accepted by the hosts |

**Logistic regression includes all explanatory variables**

|  |  |
| --- | --- |
| Result: | The goodness of fit: |
| **Observation**:   * The regression results confirm that the hosts do accept guests based on gender and race. * The Pseudo R2 value is low, indicating that the model is not robust. |

**Impact of four gender-race groups on host acceptance rate**

|  |  |
| --- | --- |
| **Result:** | **Impacts on host acceptance**: |
| **Observation**:   * White female guests have a 44% higher chance to be accepted by hosts, while that of black females is -35%. * Hosts likely to say yes to white male guests (9% higher chance) than black male guests, who are 18% less likely to be accepted. | |

**Description of the datasets**

1. **Airbnb\_listing.csv**

* Listing\_id: index for listing

***Outcome variables***

* Price: listing price (prices may vary across dates, consider this price as the representative price or historical average price).
* Log\_price: log scaled price
* Total\_guests: total number of guests who have booked the property in the past
* Log\_total\_guests: log scaled total guests

***Listing characteristics***

* Shared\_property: =1 if the property is shared; =0 if the property is private
* Shared\_bathroom: =1 if bathroom(s) are shared; =0 if bathroom(s) are private
* Strict\_cancellation: =1 if the cancellation rule is strict
* Has\_cleaning\_fee: =1 if has cleaning fee in addition to the price quoted
* Multiple\_listings: =1 if the host has multiple listings on Airbnb
* Not\_rated: =1 if the listing has not been rated
* Five\_star\_property: =1 if the listing has been rated as five stars

***Host characteristics***

* Host\_age: {young, middle, old, unknown}
* Host\_gender: {female, male, female-female, male-male, female-male, unknown}
* Host\_race: {asian, black, hispanic, white, multiracial, unknown}
* Log\_host\_experience: log scaled host experience (experience is measured as number of months the host has been on Airnbn)

***Location characteristics***

* City: listing city
* Log\_tract\_listing: log scaled number of listings in the corresponding census tract
* Log\_population: log scaled total population in the corresponding census tract

1. **Airbnb\_request\_response.csv**

* Listing\_id: index for listing

***Guest characteristics***

* Guest race: {black, white}
* Guest gender: {female, male}

***Host response***

* Host\_response\_yes: =1 if the host accepts the request