# Predicting AirBnB Rental Rates

Trevor Isaacson, Jonathan Olavarria, Jasmine DeMeyer

12/10/2021

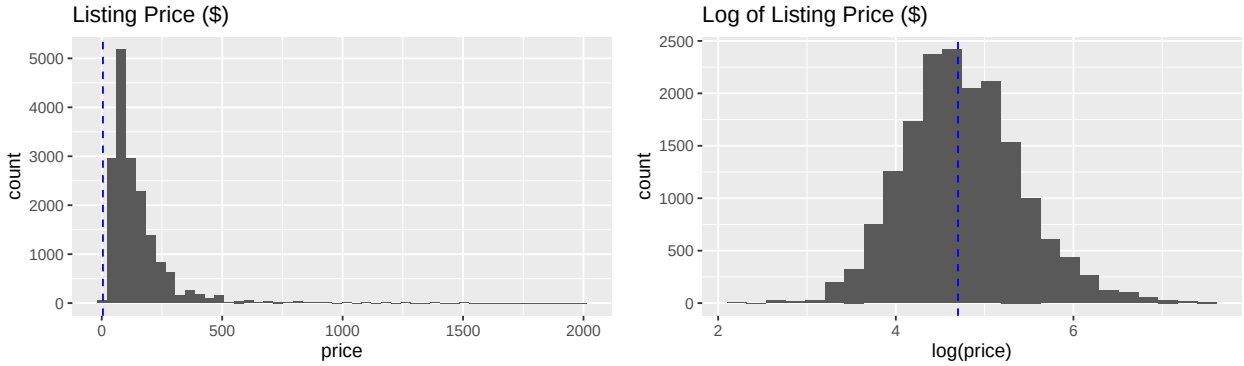
**Introduction:**

**The Data:** As a company, AirBnB is very open and transparent with the data they collect about their rental properties. They provide data about rental spaces in their system for cities and countries all over the world. Because of this, we were able to find a large dataset on Kaggle with AirBnB listings in major US cities including New York City, Los Angeles, San Francisco and others. The dataset available on Kaggle has over 74,000 entries and was used as a competition a few years ago. For the sake of time and processing, we trimmed our training data to about 17,500 entries and our test data to about 5,000 entries. We did this by taking a random sample of the provided training data. This allowed for easier access and faster processing while maintaining a large amount of data and individual AirBnB listings.

The original dataset contained 30 variables about each listing. Due to high correlations and lack of relevancy, our final dataset consisted of twenty-two variables. Those twenty-two variables can be split into four categories: property, location, host and host reviews.

Property includes:

- price: listing price




Note: Because the original price data is very heavily skewed, we needed to log transform the prices. This gives a normal distribution of prices allowing for better prediction capabilities and interpretability.

- property_type: defines the type of property listed
  - There are 21 different types including anything from apartments, houses, and condos to boats, cabins, hostels and even castles
- room_type
- accommodates

- bedrooms
- beds
- bed_type
- bathrooms

Location includes:

- city
- latitude
- longitude

Host includes:

- cancellation_policy
- cleaning_fee
- host_has_profile_pic
- host_identify_verified
- instant_bookable
- host_response_rate

Host Reviews:

- number_of_reviews
- review_scores_rating
- first_review_year
- last_review_year
- host_since_year

**Content:**

**Results:**

**References:**