

Statistical Analysis of San Francisco AirBnb Listing Prices and Review Ratings: Codebook

The Outliers 1

4/26/2020

About the Cases

We worked with data in the “listings.csv” file that was obtained through the website: <https://www.kaggle.com/jeploreitizo/san-francisco-airbnb-listings>. This dataset has been initially published at Inside Airbnb (<http://insideairbnb.com/get-the-data.html>). The cases were identified through publicly available information through the Airbnb website. There are 8,111 unique listings and 100 variables that contribute to each listing.

Table of Variables

Name	Description	Range/Code	Missing Data
Price	Price per Night in USD	0 to 1000	No Missing Data
Overall Rating	Rating out of 100	0 to 100	N/A
Latitude	Latitude of listing	37.7000 to 37.8000	No Missing Data
Longitude	Longitude of listing	-123.0000 to -122.0000	No Missing Data
Neighborhood	Neighborhood within SF	56 different categories	N/A

Why these Variables?

- Price: We will use Price as one of our main response variables.
- Overall Rating: We will use Overall Rating as one of our main response variables and will then use it as an explanatory variable to predict Price of listing. Please note, some listings do not have any ratings or reviews. The missing data is denoted with N/A.
- Latitude: The latitude of the listing is a value between 37.7000 and 37.8000 which is the latitude of the area that San Francisco spans.
- Longitude: The longitude of the listing is a value between -123.0000 and -122.0000 which is the longitude of the area that San Francisco spans.
- Neighborhood: along with the exact geographical coordinates, this categorical variable will help us explore how geographical proximities between listings can correlate with their prices and ratings.