

# New York Airbnb Data

Vivek Varma Konduru  
Gowtham Sai Chinta  
Prithvi Panchineni  
Vivek Kondrakunta

April 8, 2020

Airbnb is an online marketplace for arranging or offering lodging, primarily home stays, or tourism experiences. It acts as a broker, receiving commissions from each booking. The data-set describes the listing activity and metrics of New York City on the Airbnb platform for the year 2019. New York City has been one of the hottest markets for Airbnb, with close to 50,000 listings as of December 2019. This means over 40 homes are being rented out per square km in New York City on Airbnb! One can perhaps attribute the success of Airbnb in NYC to the high rates charged by the hotels, which are primarily driven by the exorbitant rental prices in the city. This drives to do research and gain insights such as factors affecting the price of the property, the role of the property title in attracting customers and so on. This data-set has 16 features related to location, reviews, property details and prices.

A multitude of data and interrelationship among the features with each other influences the price, which leads to heteroskedasticity. Dealing with this problem and minimizing the error we can predict the price of a property using regression technique when a host wants to list a property based on the neighborhood, room type, minimum number of nights to stay and availability of the property in a year. Naming the property is as important as providing good hospitality for the customers. Most of the time fancy names tend to get more attention from the customers. By doing text analysis we can find which words in the titles are attracting the customers. Also, by applying the classification method we can build a recommendation system that makes it easy for the users to find the best property in their desired location based on the neighborhood, room type and number of nights they want to stay. Using clustering we can categorize the neighborhoods based on minimum revenue generated.