

## Group Assignment 1

### Team members

1. Vismitha Narayanaswamy
2. Rhythm Agarwal
3. Victoria Shang
4. Wesley Webb

## Data repository

[Airbnb - Listings — Data hub \(opendatasoft.com\)](https://opendatasoft.com/datasets/airbnb-listings)

We found our dataset on AirBnB listings from opendatasoft.com. This is a data sharing software that “allows companies, local authorities, and public administrations to organize, share, and visualize any type of data (as tables and graphs), as well as make data available via APIs.” ([Wikipedia.org/wiki/OpenDataSoft](https://en.wikipedia.org/wiki/OpenDataSoft))

## Research questions

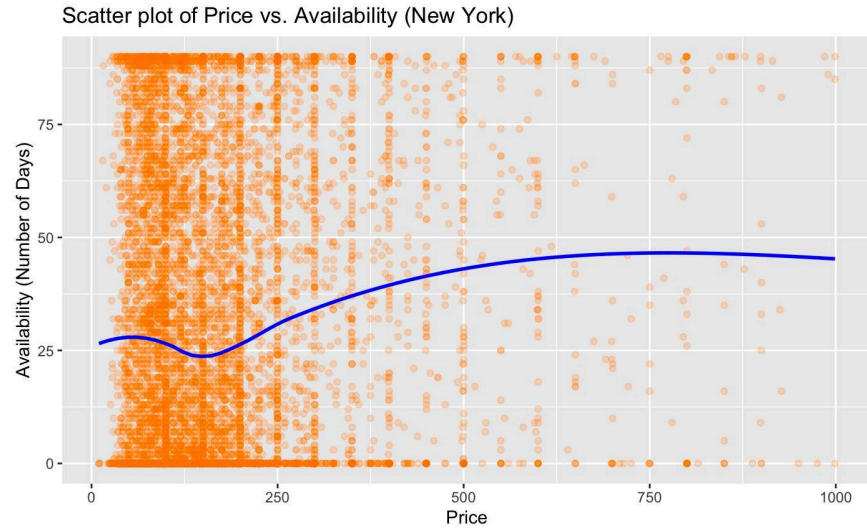
What are the most important factors or combinations of factors that result in a popular AirBnB listing, using availability as our guide for popularity? This will give insights into what property owners can do to make their listings more attractive and have their property booked more often.

## Fit a preliminary model. What are your `y` and what are your `x` variables?

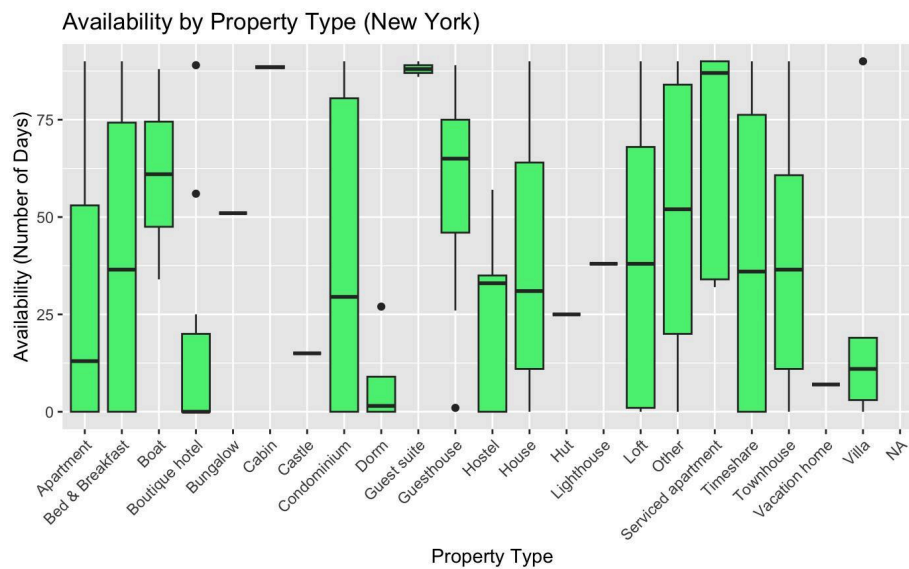
First we will examine NYC to determine the most influential factors that result in a successful listing. Our Y variable will be the 90 day availability (lower number meaning that it is more booked for that period). Our X variables will be Property Type, Room Type, Amenities, Bedrooms, Bathrooms, Price, Cleaning Fee, Features (Host is superhost),

## Make 2 graphs using ggplot exploring the data. Can you convince us you'll be able to answer your questions with the data you have access to?

We plotted a graph of availability vs price, to understand the relationship between the two and how prices affected availability. The availability increases as the prices increase though not very steeply as opposed to what we had anticipated. The lower prices airbnbs definitely have less availability meaning that they are booked more often.



We also plotted a graph of availability vs Property type, to understand if there was a certain type of property that people preferred. We did find some properties that had distinguishing features, such as serviced apartments which had more availability versus dorms which had very little availability.



We need to look at more data and plots to understand the differentiating factors and more precisely the relation between them. For example, we can definitely see that dorms are sold out more because of their lower prices, but then what about boutique

hotels? These are the types of questions that we need to look into and answer through this project.

**Describe in words what future analyses or datasets you might bring into your analysis.**

We may add or subtract different X variables to see if we are able to find better attributes that people find more worthwhile and result in higher average bookings per time period. This may vary based on the city being examined so we may also look at which cities value which variables.

<http://insideairbnb.com/new-york-city>