# Project 3 Proposal: Airbnb

My plan is to use the Kaggle dataset found here: data.insideairbnb.com/united-states/dc/washington-dc/2015-10-03/data/listings.csv.gz that is a web scraped dataset from 10/3/2015.

My goal is to test various machine learning models to see if I can A) correctly predict price B) Find reasonable price segmentation among the offerings C) Identify areas where more revenue could be generated by hosts and Airbnb and optionally if there is time to do D) Web scrape a 2019 data set (or use someone else’s code) to compare the models built to use 2015 data vs. 2019 data.

My steps are going to be:

* Data Preprocessing & cleaning
* Initial exploration with linear regression to test feature selection and correlation among variables
* Then move to more formal models along
* The three-four regressions:  
   Logistic
* Lasso
* Ridge
* I want to run a decision tree on a slice of zipcode data to see if that can take the bias of location out of the picture.
* Random Forest
* KNN (attempt to identify price segmentation)
* Clustering (attempt to identify price segmentation)

My hypothesis is that Airbnb’s own price recommendation algorithm in use in 2015 was weak. Therefore, due to the Sherman Anti-Trust act, the newness of Airbnb and lack of both government regulation and Airbnb’s ability to predict and suggest prices, I expect to see more over and underpriced options done by hosts. I expect to find numerous opportunities to present price segmentation and revenue enhancement. Although, those opportunities likely exist because of the hosts setting their own price without data. Therefore, I expect to achieve fairly poor predictions from any ML. If I get the chance to create or obtain a 2019 dataset for time series comparison, I am almost the complete opposite results due to government regulation, growth and awareness of Airbnb and its’ ability to recommend more regularized price recommendations to their hosts.

Since none of the data accounts for photos placed on the offerings, I expect to see poor prediction results of price.