

Dive Into Boston and Seattle Airbnb Data

A data investigation using Airbnb listings data for Boston and Seattle

Introduction

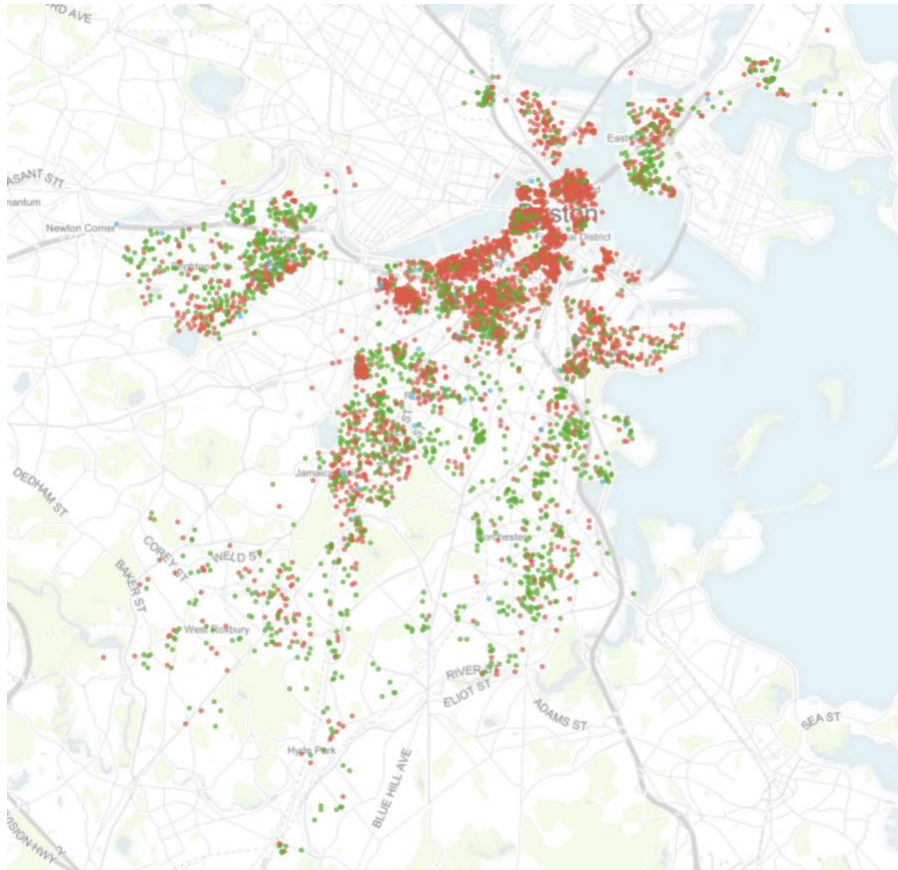
Are you wondering how Airbnb host set up rental price for their apartment/house? What can we discover from Boston and Seattle Airbnb dataset?

There are several questions I found interesting before diving into the data, hope we can figure them out by the end of the blog:

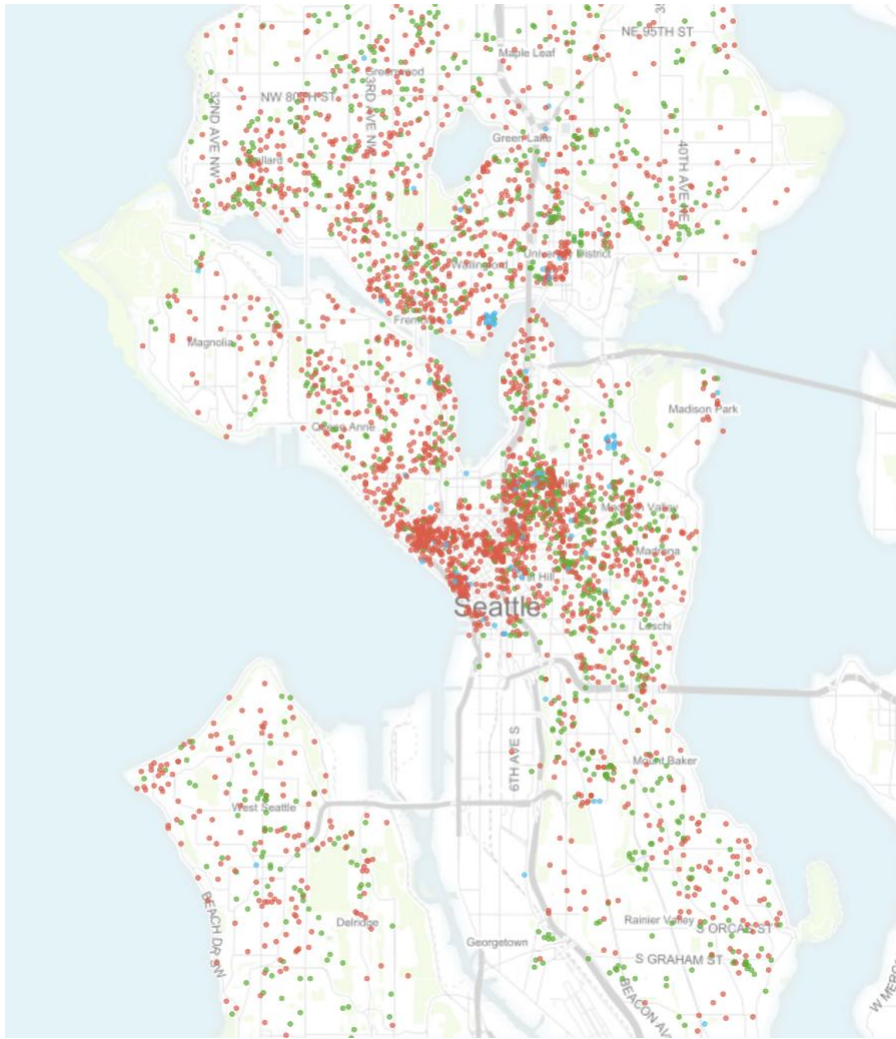
- *Is there any noticeable difference between Seattle and Boston Airbnb?*
- *What are the most important features to estimate Airbnb rental price?*
- *What are the top amenities people needs most?*

We might be able to answer these questions based our own understanding, but let's also look at the answer from data.

Quick Glance at Seattle and Boston



Boston Airbnb Listings



Seattle Airbnb Listings

From the data provided by mid August 2018, Boston has 6036 listings with an average of \$184/night while Seattle has 8494 listings with an average of \$152/night.

To give a more detailed sense of price range, 75% of Boston listings lies below \$219/night while 75% of Seattle listings lies below \$189/night which suggests the rental price for Airbnb in Seattle is slightly cheaper than in Boston.

The most expensive list in Boston is \$3999/night while in Seattle the most expensive one is \$5400/night.

Airbnb hosts can list entire homes/apartments (red), private (green) or shared rooms (blue). In Boston, hosts listed their room type as 62.2%, 36.6% and 1.2% respectively for the above three room types. While in Seattle the room type percentage is 66.6%, 30.4% and 3.1% respectively.

Let's take a deeper look at data and try to use them to estimate the rental price of Airbnb in Boston and Seattle.

Predicting Rental Price in Boston and Seattle

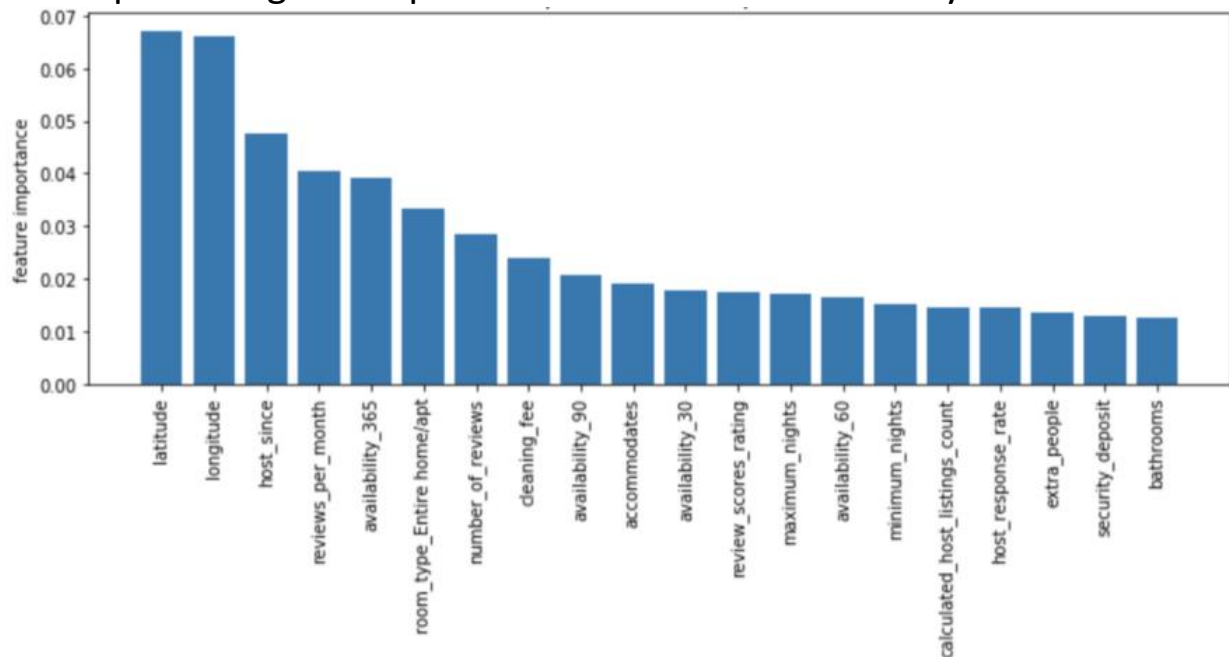
As 99% of the listing price are below 500, so I drop the rows above 500 to get a more stable prediction. Missing values were filled by median value or most frequent value based on some other related features.

GradientBoostingRegressor was used as classifier for both datasets and a five-fold GridSearchCV was applied to find the best hyperparameter for the classifier.

On the other hand, 1/5 of the preprocessed dataset was used as test data and the remaining 4/5 are used to train machine learning model.

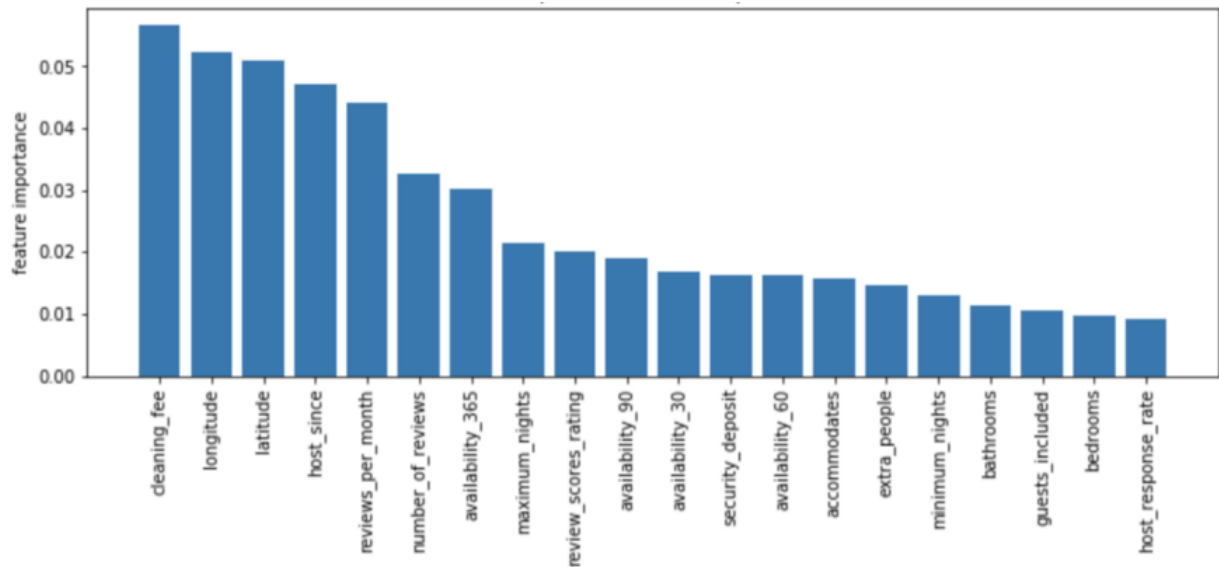
What are the Most Important Features to Predict Rental Price?

I listed top 20 features for both cities which plays the most important role in predicting rental price. Let's take a look how they look like.



Top 20 Features to predict Boston Rental Price

It is no surprise that *latitude* and *longitude* are one of the top important features, as rental price are highly depending on the location of the property. *Host_since* is a feature represents when the host start their journey with Airbnb, if a host start airbnb earlier they probably had more experience attracting people and know how to make their guest more comfortable. *Reviews_per_month* and *number_of_reviews* tell us people are willing not only stay there but also leave a comment, which means they enjoy the stay. Some other features such as *accommodates*, *security_deposit* and *bathrooms* are also among the top list.

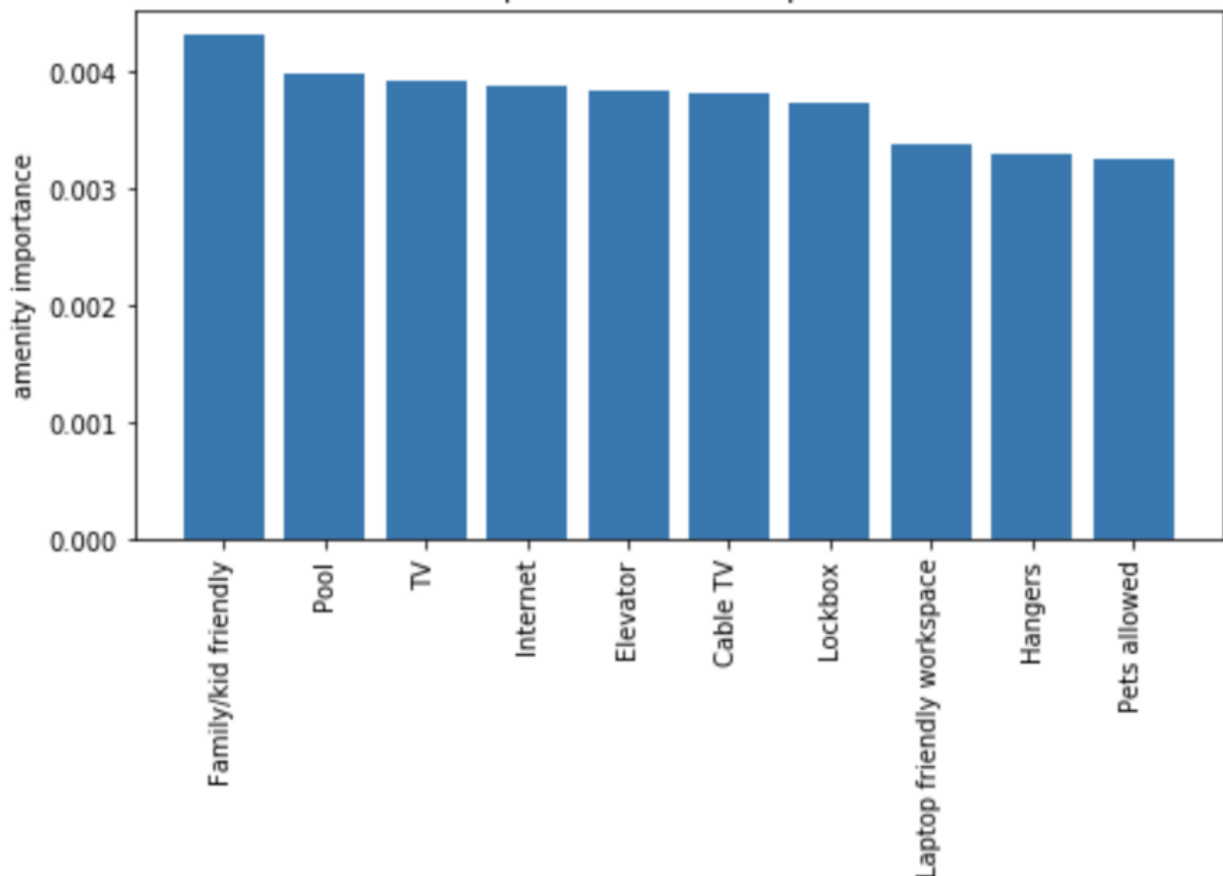


Top 20 Features to predict Seattle Rental Price

In Seattle, features are more equally important. There are some new features such as *host_response_rate*, *extra_people* and *guest_included*. But most of the top features are shared with Boston machine learning model. We can see *cleaning_fee* is the most important feature here, as more cleaning needs result from larger room, probably more bathrooms and more services.

What are the Top Amenities to Pay Attention?

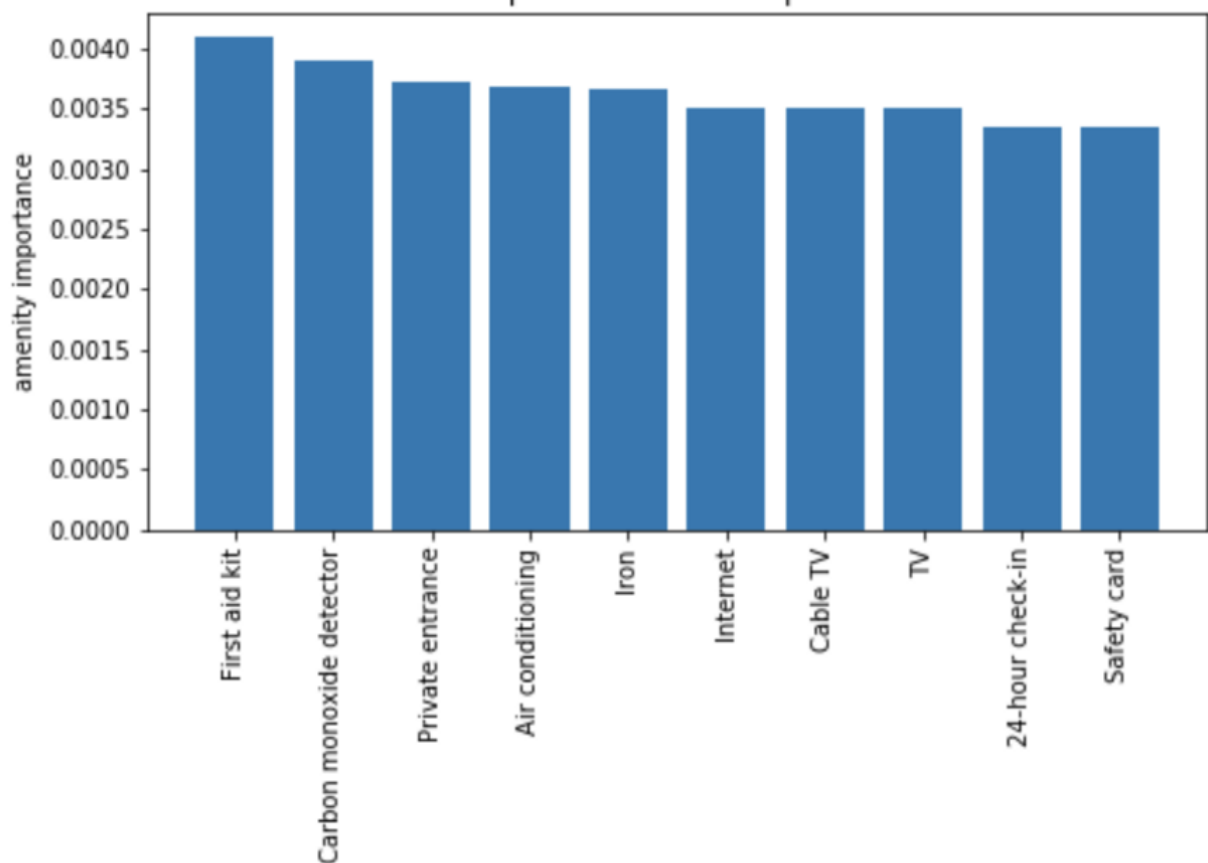
After training machine learning model, we can also check how amenities influence the price prediction/how important are they.



Boston Top 10 Important Amenities

In Boston, amenities such as TV, Internet lockbox and hangers are important as expected. Family/kid friendly is also a feature which most traveller will consider. We can see if you have a pool in your property, you might have more confident to raise your rental price.

One interesting phenomenon is Airbnb has more amenities listed in Seattle (168) than Boston (120).



Seattle Top 10 Important Amenities

Seattle hosts have some extra amenities than Boston's such as ***Amazon Echo, Heated Floors, Formal Dining Area and Memory Foam Mattress***. Noting that they are very host dependent and not a lot of hosts provide these amenities, so they might not influence the prediction a lot. We can see the importance of amenities are not as high as other features in the last section. As there are more than 100 amenities provided by the hosts and most of them are always provided as they are needed by most guest.

In Seattle, the most important amenities seem to be *First aid kit*, while TV and Internet are also in the top 10 as they should be.

Conclusion

In this blog, we dive into the most recent Airbnb Boston and Seattle dataset and found many interesting phenomena. Here I listed what we have done:

- *We gathered the Boston and Seattle Airbnb data, and compare the two dataset.*
- *We established a machine learning model to predict the rental price for both cities.*
- *We looked at the feature importance of the trained model and check if they make sense.*
- *We list all the important amenities to get a better feeling how host can make more money by providing better services to meet customers' need.*