Prediction of Rental Prices Based on Craigsist Listings:

Business Case

The market for home rentals in Seattle is very dynamic. The rental prices have varied significantly from year to year the last few years (5% last year).

The prices for rentals are relevant to many stakeholders such as:

* Landlords and Tenants: trying to figure out the fair market price for properties
* The City: in evaluating the low-income thresholds and allocating budget to low income resident services and other city planning.
* Prospective home owners evaluating the net present value of buying versus renting.

These stakeholders make important decisions based on their understanding of the rental market and they would benefit from having a model that can be easily updated,

The Data

The data was collected from Craigslist. Craigslist is a very popular site for rental classifieds and I assumed that it is representative for the rental market in Seattle.

The data was scraped using Beautiful Soup. I obtained 3000 listings, which included features such as number of bedrooms, bathrooms, area, dog allowed, cat allowed, latitude, longitude, number of images in the listing, zip code, w/d in unit, laundry in building, carport, and garage.

The data was cleaned to delete duplicates, listings without price and outliers.

The final set contained 1742 listings, with an average price of $1874/month and price range of $600 to $6995/month. The mean listing included 1.5 bedrooms and 1.3 bathrooms.

Several linear regression models were compared. A model was selected based on the higher cross-validated adjusted R2.

The model selected was a polynomial grade 2 model, which was estimated using LassoCV. The input variables were normalized. Its cross-validated adjusted R2 was 0.6 and the mean error was 15%.

The relative effect of the variables in the final price can be examined by looking at the coefficients of the linear regression, because the variables were normalized. Thus, there are a lot of insights that could be drawn from the relative value of the regression coefficients.

The errors show a conical shape. Therefore, there may be other solutions outside the linear regression space that could yield a better result.

Conclusions:

Although the model selected only explains 60% of the variation of price and the mean error is about 15%. The model is very simple and easy to implement and considered acceptable given the few featured included.