Housing Prices and Park Proximity

A Study of the Impact of Parks on House Prices in King County, WA

August 24, 2020

By: R. Nate Mease, CFA

Housing Prices and Park Proximity

Introduction

This project will examine the relationship between housing prices and parks. Real estate professionals are always looking for factors that influence prices. It is a cliche that the most important thing in real estate is location, location, location. I hope to examine how the location of a house relative to popular parks may impact pricing. This project would help real estate agents, real estate appraisers, single family rental investors, and prospective buyers and sellers in understanding the value of a particular house. In addition, the project could help city planners with park placement and management.

Parks, particularly popular ones, are attractive features for a neighborhood. They provide space for a variety of outdoor activities including exercise, picnics, and enjoying nature. While parks are not typically a feature that prospective house buyers look for, they provide real benefits when they are nearby and well maintained. I hope to examine the impact that popular parks have on housing prices in King County, Washington. My hypothesis is that popular parks would increase the value of houses in close proximity.

Data

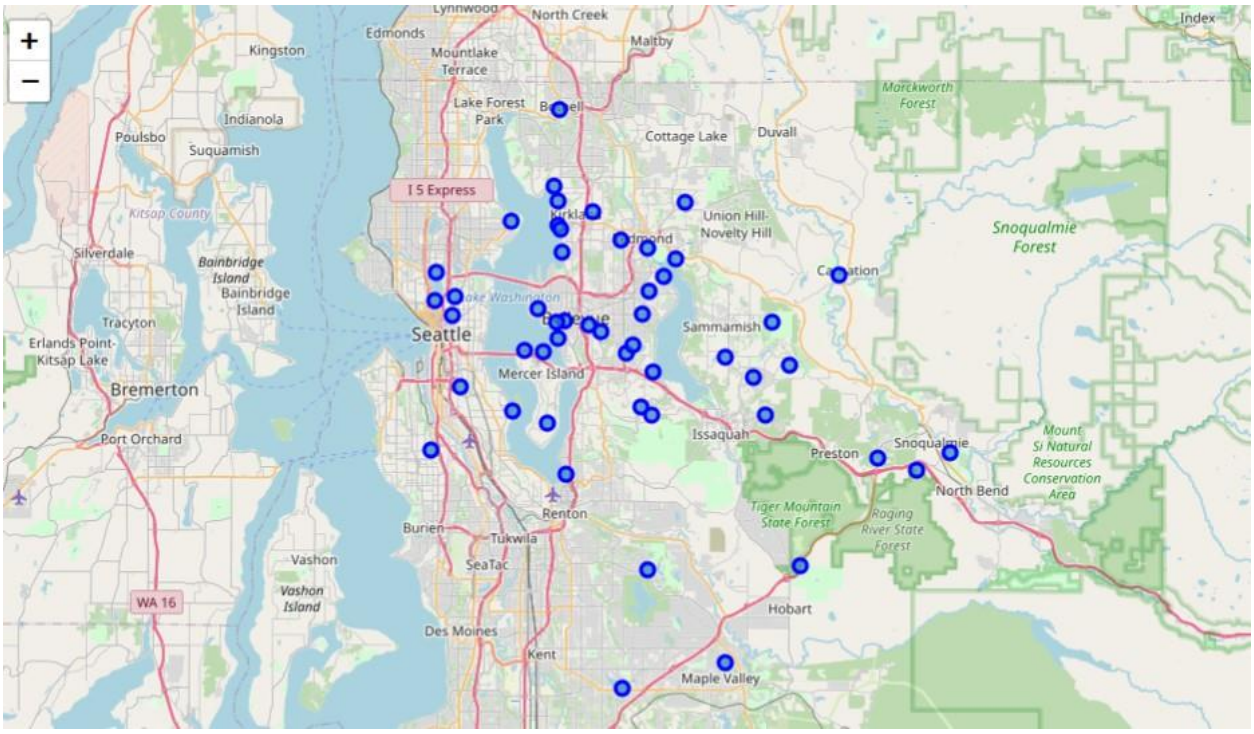
In a previous class we examined the housing data for King County, Washington and how various features of the house itself impacted prices. The housing data from that project included the latitude and longitude of the houses as well as the prices. Foursquare provides information on a variety of venues including parks. I will retrieve the location of the popular parks in King County, Washington from Foursquare.

Using the latitude and longitude of both the parks and houses in the two data sets we can calculate the distance each house is from a particular park. We will use the popularity of a park on Foursquare as a measure of those that are attractive to potential buyers. With distance from popular parks and house price data we should be able to determine if proximity has an impact on housing prices.

Methodology

The housing data was only for King County, Washington, which includes Seattle and the surrounding areas. There were 21,613 housing records in total. The parks data was downloaded from Foursquare, with a limit of 50 parks in order of popularity. The 50 park limit was chosen to include enough popular parks near all of the housing areas utilized, but to also eliminate parks that are less attractive to prospective house buyers.

Below is a map of the parks selected.



Because location and house size have such a large impact on house prices I wanted to control for these factors, to the extent possible. The median house price for the entire housing data set was $450,000. The data set was sorted to find the zip codes with average house prices between $400,000 and $500,000 to provide a large enough sample size while also limiting the influence of location (e.g. a zip code with $2 million average house price).

The data set was further sorted to limit the influence of house size. The median house size for the full housing data set was 1,910 sq ft. Houses from the subset of zip codes above where then filtered to only include those ranging from 1,800 – 2,100 sq ft. The resulting data set included 516 houses across 14 zip codes within King County, Washington.

The next step in the analysis was to compare the distance from each house to the closest park in the subset of popular parks. To do this the geopy.distance.distance(coordinates 1, coordinates2) was utilized to measure the distance from each house to each park and then take the minimum distance for each house.

The minimum distances from a park for each house were then compared to housing prices using a regression analysis. The median minimum-distance from a popular park was 4.2 kilometers while the average was 5.2 kilometers.

Results

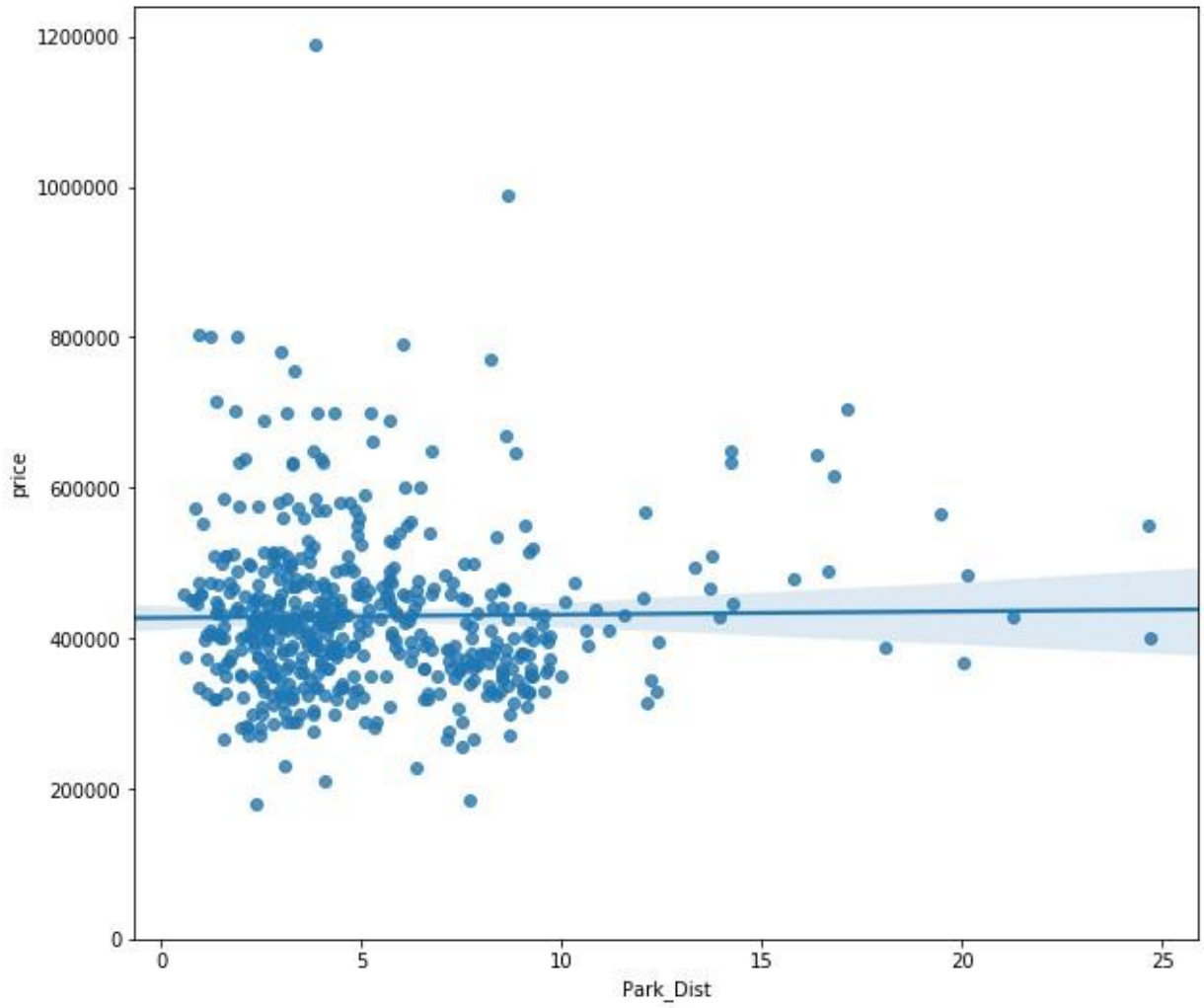
A correlation and regression analysis was done on minimum park distances and housing prices as well as a regression analysis. The resulting correlation matrix is below.

	Park_Dist	price
Park_Dist	1.000000	0.014167
price	0.014167	1.000000

The resulting regression equation is:

Price = 427,129.99 + 430.26(Park Distance)

A scatterplot with the regression line:



The R-squared for the regression was 0.0002 and the mean squared error was 11630000687.38.

A second regression was doing utilizing training and testing splits of the data. The results of that data are as follows:

number of test samples : 155
number of training samples: 361
Test Rsq: -0.008305838895816375
Train Rsq: 0.0014874621868116167

A cross validation test was conducted with a mean cross validation score of -0.0068.

Discussion

The correlation of the minimum park distances and prices is very low at 0.014. The regression line above is also very flat. The R-squared is low for both the initial regression and the regression utilizing training and testing data. This indicates that that proximity to parks is not an influential factor in housing prices in King County, Washington.

The median and average distances from parks were higher than expected at 4.2 km and 5.2 km, respectively. It could be that limiting the data to areas with house prices closer to the median house price also limited the impact of parks on pricing. In other words, it may be that houses in the more expensive zip codes are closer to popular parks but were eliminated, while median to average priced homes are too far from popular parks for the distance to have significant impact on pricing of those particular homes.

An additional analysis of houses in zip codes with higher and lower prices could yield additional insights into the influence of park distance on house prices. Analysis with an expanded number of parks may also provide additional information by reducing the distance of houses from parks.

Conclusion

This analysis indicates that houses priced in the median to average range are not significantly impacted by the distance from popular parks. Those focused on houses in this price range may not find this to be a valuable feature for potential buyers. It is possible that the analysis limited the impact by only including zip codes with average prices close to the median for King County, while limiting parks to the most popular, which may lie in higher priced zip codes.