Effects of property attributes on housing prices in Milwaukee County from 2002-2018

Michael J. Scanlon

Marquette University

Completed as a final project for COSC 4610/5610

Abstract

Throughout Milwaukee’s history, we’ve seen several wildly different communities form through poor practices such as redlining and other oppressive old policies. These communities typically have had residual poverty and high crime-rates plaque them. When looking at residential property across all of Milwaukee, how do these communities differ in terms of value and why? How can we show that inherently there is less property value in these communities than other communities? Using supervised learning algorithms, we can analyze what gives properties value in different aldermanic districts in Milwaukee.

Keywords: Multiple regression, Housing Prices, Linear regression

**1. Introduction**

The idea of using supervised learning to predict the prices of homes based on features in general is simple. However, goal within this paper, is to examine the divide in Milwaukee (wealth, poverty, and crime are a few examples of factors that separate areas from each) and the inherent effect on their value due to these factors in the area using these models within context about individual districts. Essentially determining what factors lead to homes in different areas losing or having value at all.

Linear regression models and multiple regressions are a strong method to predict y given x. These are the two methods of supervised learning models that we will be using to explore the data we are working with.

**2. Related Works**

There is plenty of research related to what we are discussing here and some general findings. Milwaukee has a high poverty and unemployment rate when compared to the rest of the state of Wisconsin. When looking at analyzing homes in general, the consensus found is that homes with more finished area and are larger tend to have more value.

**3. Approach**

The general problem we started with was which attribute can we use to build the best model to predict sale price of a residential unit. As testing began it became clear that the factors in the dataset where not enough to build a strong enough model and especially not valid enough to begin to answer our question.

So, as we began to see results that weren’t super promising, we started to think about what it was we were trying to show, the difference in separate areas of Milwaukee. The dataset included the aldermanic district which each house resided in as well as more specific neighborhood numbers. The neighborhood numbers were too specific and there were far too many to test in reasonable time. But there are 15 districts in Milwaukee all representing rather different sections of the city.

Map

Description automatically generated

Picture 1

The new problem that became the focus of model building and testing was how could we show that each district was distinct in value for reasons beyond basic attributes. The data was then broken up into 15 different dataframes where each entry was from the same district. Using these new dataframes, we began to test linear regressions using the square feet of finished area in the property.

**3.1 Linear regression models**

Linear Regression (eq. 1)

Multiple Regression (eq. 2)

To begin working with these models we can first examine the equations and what we are doing. A regression model is an equation created using distance from each existing data point to attempt and create a best fit line of sorts to predict the value of y given any x. The variable represents the intercept of that line and , are the slopes of the line, or lines in the case of multiple regression. (Note. in the associated code our models display as b and as w)

As built up to in the first part of this section, we used a regression model to show the sale price as finished square feet increased for the two most expensive districts (3 and 14) and the two least expensive districts (1 and 15) determined using mean sale price. We used R-scores to determine which models were the most accurate.

**4 Experiment and Materials**

The dataset used was acquired from (source 1) and represents all sale data in Milwaukee from 2002-2018 of all properties. To process the data, we first removed all property types that were not residential from the dataset, since they are not relevant to our problem, as well as removing columns containing irrelevant data that made it hard to read. Then we cut down any extreme outlying properties that would not give our models any beneficial data and skew R-scores. Finally, we removed entries with null values in our target fields.

We used my personal PC running windows 10 to run Jupyter Notebook software, where we did all testing and documentation of progress through code and header blocks in the notebook file. We split the data into train and test sets for each regression using $ train\_test\_split(X, y, random\_state = 0) $.

**4.1 Results**

Chart, scatter chart

Description automatically generated

Graph 1

District 3 is the wealthiest district, $288k average sale, we are going to look at and we can see that the regression is strong from the graph alone. The R-squared score for this regression was 0.569 on the test set and 0.644. These are very strong scores that prove beyond reasonable doubt that we have strong correlation between the finished square feet and sale price in this district.

Chart, scatter chart

Description automatically generated

Graph 2

District 14 was our second wealthy district with an average sale of $168k. This regression was not as strong as District 3, but it is still worth considering as the R-squared score was 0.226 for training and 0.16 for test. Which is significantly lower than 3, but significant compared to the next two regression we’ll look at.

Chart, scatter chart

Description automatically generated

Graph 3

District 1, $70k average sale, had a strong test result, with R-squared at 0.15. However, a lower training score of 0.118 might suggest that the test was disproportionately represented by the model.

Chart, scatter chart

Description automatically generated

Graph 4

District 15 had the lowest average sale price at $65k. It also had the worst R-squared scores at 0.079 for test and 0.064 for training sets. These scores clearly display that there are more pressing factors when it comes to sale price despite the attributes having a high correlation.

**4.2 Analysis**

The results we can see are a strong indication of the point that there’s an outlying factor when it comes to values of property in Milwaukee. The R-scores show us that most places at least are relatively related to the size and recency of the home, but we can see that something isn’t right in district 15.

After additional research we were able to find that district 15 correlated very much to a high crime area. Districts 14 and 3 were both very low crime and located on the lake. We are able to deduce that because of this, the safety of an area is also a large role in determining the price of property.

Map

Description automatically generated

Picture 2 (source 2)

This is a map showing the highest crime areas in Milwaukee. Dark blue is high crime and descending into light/empty squares where there is low-no crime. District 15 is dead center of the map, as we saw in picture 1, and districts 3 and 14 are lakeside and have far less crime.

**5. Conclusion**

The relevance of this work resides in the reasons for and damage that crime can cause in an area. Milwaukee housing loses significant value in inner-city areas with high crime-rates, which is severly damaging to the residents and infrastructure located there. Since crime is often a problem of residual poverty and oppression in some cases, it is a cycle of low value property resulting in poor residents and high crime.

So we answered the question what effects home value in Milwaukee, in safer and less prosecuted districts, we can see that homes have proportional value to their immediate quality. But, in over-policed and socially discriminated communites we see that homes are inherently less valuable.

Works Cited , sources 1 and 2 respectively

*Property Sales Data - 2002-2018 Master File - City of Milwaukee Open Data Portal*, data.milwaukee.gov/dataset/property-sales-data/resource/f083631f-e34e-4ad6-aba1-d6d7dd265170.

“Milwaukee, WI Crime Rates.” *NeighborhoodScout*, www.neighborhoodscout.com/wi/milwaukee/crime.