Effects of Residents’ Background on Home Sales in Philadelphia

Julia Dreifuerst

Fall 2021

**Introduction**

Buying a home is considered a standard milestone in most peoples’ lives. People save up for years, go to school, work hard at their jobs to earn raises, and take out loans, all so they can afford to own their own property. There is a lot that goes into the process of finding the right house alone, before even getting into the process of buying it. Some of these things the buyer has to consider are location, price range, size of the home, age of the home, safety of the area, what the neighbors are like, if there are any issues with the home like mold or wood rot, finding a good realtor, and so on.

Home prices are always changing based on supply and demand in the area they are located in. Redfin reported that in Philadelphia, in October 2021, the median home price was $259,000, which was up by 7.9% from the year prior, although fewer homes were sold in the current year. This essentially means that the demand may have increased for buying a home in Philadelphia in the past year, while the supply did not. The buyer’s background matters a lot in a situation like this where supply may be limited. Factors of the buyer’s background that matter may be things like education level, income, credit score, and other outstanding loan balances, for example.

The housing market is always changing as time goes on and various factors occurring in the world affect the economy which filter down into determining home prices. This study is focused on the effects of median household income and educational background on the median home prices in Philadelphia, to determine whether there is a relationship between these variables, and if they have the same effects. This study will evaluate a sample of 3,899 houses sold in Philadelphia in 2015 in 25 different zip codes, and build a regression explaining whether the median household income and educational background have the same effect on the home sales. The median home sale price in 2015 will be used as a dependent variable because the median sale price is a strong indicator of home sales performance.

**Background**

When it comes to researching the topic of home sales, I came to find it is not exactly a widely researched subject. The housing market is often fluctuating, and the trends can look very different each decade, and it is especially variable based on location. There are a few studies that have been done in the real estate industry on the relationship between days on the market and home sales, household debt and home prices, and the strategies that can be used to enhance property values in Philadelphia. These topics were chosen because they may provide some insight into the potential influences on home prices and how they operate.

In regard to financial background in relation to home prices, a study done by Jacobsen and Naug (2004) focused on the influences of a growing household debt in relation to housing prices. This study was done in response to the increasing household debt that is related to developments in the housing market and a decline in interest rates. They analyzed the growing debt trends using an empirical model. A lot more often than not, when people buy a home, they finance the home with a loan. If housing prices spike, then debt growth occurs simultaneously. Households can also increase their debts when they accrue loans to finance consumption and investment with the increased value of their dwellings as collateral. This study models households’ domestic gross debts measured by the C2 credit indicator. It was found that only a small part of the housing stock changes hands every year, meaning if the housing prices stabilize, they will still sell for higher than they were bought for. Household debt may continue to increase though because higher house prices can result in higher final wealth and borrowing conditions that are better for many households. Overall, the main takeaway from this study is that when house prices increase, debt increases with it.

Another study that was done focused on public investment strategies and how they affect neighborhoods and home values in Philadelphia done by Wachter and Gillen at the Wharton School at University of Pennsylvania (2006). This looks into the potential influences on property values and which public investments can lead to increased home prices. The types of investments they looked at specifically, were place-based public investment strategies such as neighborhood greening, vacant land management, public safety, school improvements, commercial-corridor improvements, and access to transit with the purpose of examining the economic impacts on the properties that would result. In the methodology, they measured impact through a neighborhood quality evaluation tactic developed by the Wharton School Geographical Information Systems (GIS) laboratory for the City of Philadelphia. The GIS uses a spatial database that combines value and attribute data with geographic data along with information on public service areas and specific neighborhood interventions. The main dependent variable measured was the change in home values in Philadelphia. The results showed that when the home was near a commercial corridor in excellent condition, it increased the home value 23%, when streetscapes were improved home value increased 28%, and if there was an adjacent vacant lot, the home value decreased 20%, but increased 17% if the lot was stabilized and greened. They also found an overall steady trend of increasing home prices in the past 25 years in Philadelphia.

A third study done by Tucker, Zhang and Zhu called Days on Market and Home Sales (2012) in the SSRN Electronic Journal looks at the relationship between the number of days a home is on the market and how this affects home sales after Massachusetts implements a law that prevents the realtors from taking the listing down and putting it back up to reset the number of days the house is on the market. They looked specifically at single-family home sales in 20 towns along the Massachusetts-Rhode Island border, because Rhode Island did not implement this law while Massachusetts did, making Rhode Island a good option for a control. They found that the houses listed in Massachusetts right before the policy was put in place showed a buyer reaction that led to the homes being reduced by $21,500 and lengthened the days on market by 20 days compared to Rhode Island homes. They also found that the overall sales prices went down by about $11,000 after the policy was implemented, although a positive effect was that the average days a listing was on the market was shortened by 18 days. The hardest hit homes were the ones that were listed just before the policy, because both the buyers and the sellers were sort of caught off guard by the policy, they dropped the prices the most and the time houses were on the market was also longer than usual.

**Methodology**

This study is looking at whether the independent variables median household income and educational background, specifically the percentage of residents in the same zip code that have a bachelor’s degree or higher, have the same effect on the dependent variable median home sales price. In the models, the independent variable educational background will be identified as “% of Residents with a BA or Higher”, and the dependent variable median home sales price will be represented as “Median Sales Price”. First, a t-test will be performed to understand whether there is a difference between the two variables and their effects on median home sale price. Then, the correlation matrix will be calculated to determine which independent variables in the study have impacts on one another and the strength of the relationship they have with the dependent variable. Lastly, a linear regression analysis will be done to try to explain the potential relationship independent variables have on the dependent variable. The simple linear regression model will be this:



**Sample Size and Gathering Data**

The population of interest in this study is Philadelphia homes that were for sale in 2015. The sample is of 3,899 homes for sale in Philadelphia in 2015, in 25 different zip codes. The data was gathered from the cleaned data sets posted on the GSB 519 D2L page for this project.

**Variables Measured**

The variable chosen to represent the dependent variable in this study is median sales price, because sales price is a strong indicator of the status of the housing market and tends to be influenced by a variety of factors. The relevant independent variables are median household income and percentage of residents with a BA or higher. The average median household income per area was $42,228 with a standard deviation of $18,392. In regard to the percentage of residents with a BA or higher, the average was 30.28% with a standard deviation of 24.44%. For the dependent variable, the mean sales price was $187,121, and with a standard deviation of $124,066, meaning there was probably a decent amount of variation in sales price among the zip codes.

The other independent variables involved, population, mean days on the market, units sold, and hotness of the area, may also have an effect on the median home sales price. This would be expected to be a positive relationship for population, units sold, and hotness, and an inverse relationship for mean days on the market between these variables and the median home sales price. The average population per zip code is 31715, with a standard deviation of 17,146. This means that each data point comes from areas that may vary a good amount in the population associated with the home sales in that zip code. The average number of days a house was on the market was 80 days, with a standard deviation of 20 days, approximately. The average number of units sold per zip code was 162 with a standard deviation of about 125, meaning there was also a lot of variation here with how many houses were sold per area. Also, the mode for hotness in the area is 0, meaning more homes are located in less hot areas.

**Data Analysis**

The t-test for paired two sample for means was run using the Data Analysis Tools on Excel. These were the results:

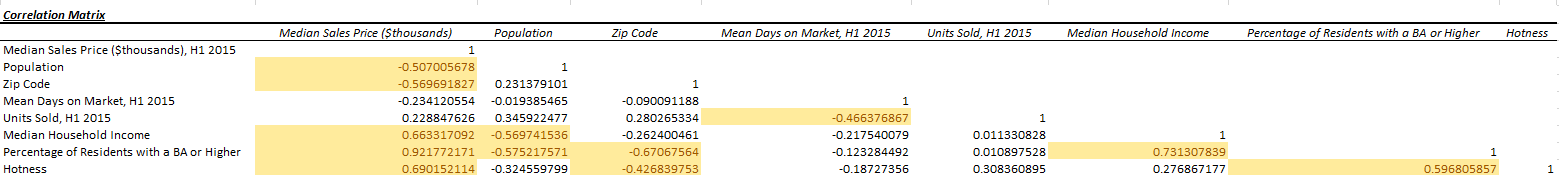
Table

Description automatically generated

This test was performed to determine whether there is a difference in the mean median household income and percentage of residents with a BA or higher. The hypotheses for this test were:

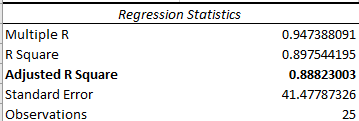
The results of this test, which was done with an alpha of 0.05, show a t-statistic of 11.4829, a two-tail t-critical value of 2.0639, and a p-value of 0.00. Since the p-value is less than the alpha (0.00<0.05), and the t-statistic is larger than the critical value (11.4829>2.0639), we reject the null hypothesis, meaning there is a difference between the means.

Next, the correlation of each of the independent variables in relation to the dependent variable was considered. The results were as follows:



The critical value for this correlation matrix was calculated to be 0.3961. Using conditional formatting in Excel, the highlighted cells show where the absolute value of the correlation coefficients are greater than the critical value, meaning they are significant. Furthermore, this means those variables that are highlighted are believed to have impact on each other.

The last test that was run was a regression analysis. This regression model was initially run with all seven independent variables, and it was found that the only statistically significant variable was the percentage of residents with a BA or higher with a p-value of 0.0015 and an alpha of 0.05, but median household income had a high p-value of 0.9758. Also, the adjusted R square value for this initial regression was 87.65%, meaning this percentage of the variation in median home sales price can be determined by this model. Also, the F significance was 0.00 which is less than 0.05, meaning the model is a better fit with independent variables than without. For the second regression analysis, the independent variables zip code and hotness were deleted. The results of this showed an increase in the Adjusted R Square value to 87.99%, and the independent variable units sold became statistically significant at a p-value of 0.0116. In the third round the independent variables population and mean days on the market were deleted. This left the independent variables units sold, median household income, and percentage of residents with a BA or higher. The Adjusted R Square value increased to 88.32% in this regression, but the median household income still had a high p-value at 0.8110 while the other two independent variables remained statistically significant. So, for the fourth regression analysis, and the most relevant one, the independent variable median household income was taken away, and the results were the following:



Table

Description automatically generated

Table

Description automatically generated

In this final model, the independent variables that remained were percentage of residents with a BA or higher and units sold in 2015. This regression resulted in all the independent variables being statistically significant, except for the intercept. The Adjusted R Square was 88.82%, which tells us that 88.82% of the variation in median sale price of homes can be explained by the units sold and percentage of residents with a BA or higher. Also, the F test tells us that the model is still significant from the first regression. The intercept shows us that the median sales price of a home would be $10,628 if all of the independent variables were set to zero. This can be true because the value is positive, and the independent variables contribute to making a house worth a lot more than that. The p-value of the percentage of residents with a BA or higher is 0.00, so this proves the hypothesis partially true in that educational background has an effect on the home sale prices in the area. The equation for this is:



This equation shows that if there was a one unit increase in units sold, a $220 increase in median home sales price would result, other things equal. This value makes sense, because when other units are sold, there is less supply, which tends to drive up price. Also, a one percent increase in percentage of residents with a BA or higher results in a $4,670 increase in median home sales price, other things equal. This value also sounds accurate because more educated individuals tend to have higher incomes and are therefore able to spend more on a home.

The last important part of this regression analysis is to look at the residual scatterplots to check the assumptions of the model. These are the residual plots for the two independent variables:

The residual plots show no signs of heteroskedasticity or multicollinearity.

**Conclusion**

This study was conducted to better understand the relationship between the residents’ background and the effects it might have on home sales prices in Philadelphia. The sample consisted of 3,899 houses that were sold in Philadelphia in 2015. The independent variables involved items that tend to affect the dependent variable, which was median home sales price. The independent variables were population, mean days on the market, units sold, median household income, percentage of residents with a BA or higher, and the hotness of the area, although this study focused on the effects of median household income and educational background. The first test that was performed was a t-test in which the results suggested the alternative hypothesis was true in that there was a difference between the mean percentage of residents with a BA or higher, and the median household income. The correlation test was performed next and showed results that suggested the median home sales price and the median household income and percentage of residents with a BA or higher may have effects on each other. Interestingly, in the regression analysis, it was found that the median household income was not a statistically significant variable, and the final regression model was best without that variable. The variables that were found to be statistically significant in the most relevant regression model were units sold and percentage of residents with a BA or higher. Also, in this model, 88.82% of the variation in median home sales price can be explained by the two final independent variables. The hypothesis that median household income and percentage of residents with a BA or higher have an effect on median home sales price help up to only be partially true since median household income was not found to be statistically significant. It is also important to consider that a shortcoming of this study is that it was only focused on a few thousand homes in the city of Philadelphia alone, so conducting a wider study that looks at more homes in other geographic regions would be a good way to continue researching this topic. A suggestion for future research would be to also take into account more variables like safety of the neighborhood, proximity to resources like shopping centers, or the number of HOAs in the area, which may also have influences on the median home prices in an area.

**Works Cited**

Jacobsen, D. H., & Naug, B. E. (2004). What Influences the Growth of Household Debt? *Norges Bank Publication Series*. Retrieved 2021, from <https://norges-bank.brage.unit.no/norges-bank-xmlui/handle/11250/2504442>.

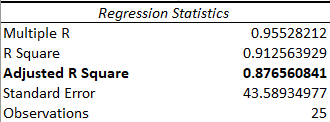
Redfin Corporation. (2021). *Philadelphia Housing Market: House Prices & Trends | Redfin*. Redfin. Retrieved November 12, 2021, from <https://www.redfin.com/city/15502/PA/Philadelphia/housing-market>.

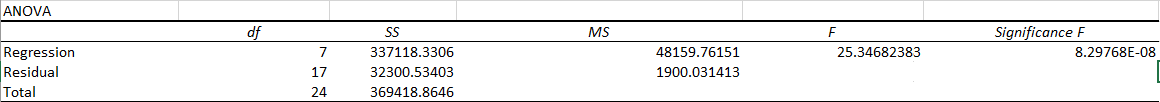
Tucker, C., Zhang, J., & Zhu, T. (2012). Days on market and home sales. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1481321

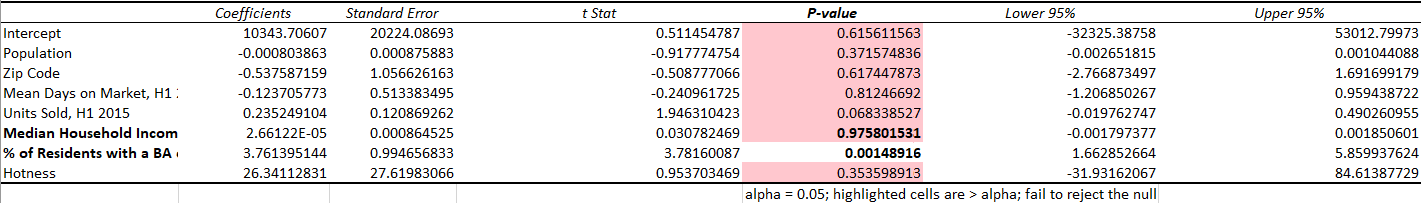
Wachter, S. M., & Gillen, K. C. (2006). Public investment strategies: How they matter for neighborhoods in Philadelphia. *unpublished report of the Wharton School of the University of Pennsylvania*.

**Appendix**

Regression 1 (all independent variables included)



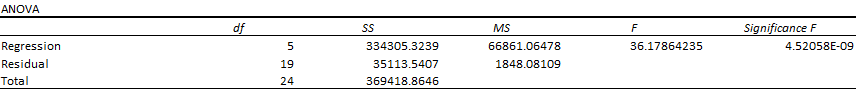




Regression 2 (deleted zip code and hotness)

Text

Description automatically generated



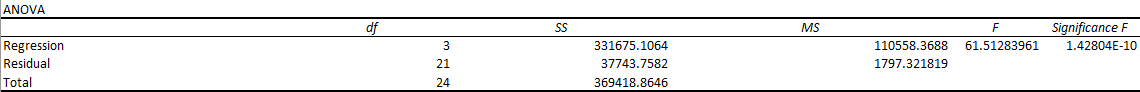
Graphical user interface, application

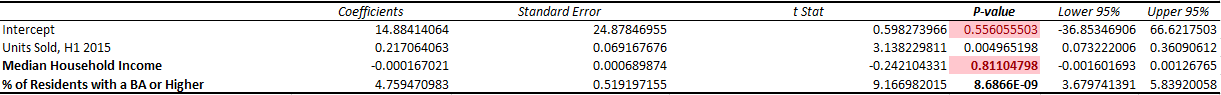
Description automatically generated

Regression 3 (deleted population and mean days on market)

A picture containing text

Description automatically generated





Descriptive Statistics

