LONDON BOROUGHS HOUSE PRICE ANALYSIS

Data Analytics Project- NCG613

MSc in Data Science and Analytics

National University of Ireland, Maynooth

Maynooth, Co. Kildare, Ireland.

Department of National Centre for Geocomputation (NCG)

Maynooth University, Maynooth, Co. Kildare Ireland.

Professor – Dr Kevin Credit

Submitted by –

Anush Harish - 21250164
Madhushree Saravanan - 21250314
Naveen Ramprasad - 20253139
Subbalakshmi Narayan Prasad - 21250754

May - 2022

Contents

List of Figures

1. Introduction and Research Question

Over the past 22 years, several studies have examined the impact of public schooling on house prices. The location of a house is tied to local public schooling more often than not. And the value of this schooling is capitalized into the house price. Starting with (Black, 1999)[1] numerous studies have examined the effect of schools on house prices. But the house's value is not solely based on the quality of the school; it is also related to the distance of the nearest school from the house. Studies of local public goods, such as open space, have long examined how the value of an amenity change with distance from the house. In the context of schooling, the value of distance has received little attention (Des Rosiers, 2001) [2] (Thomas J. Kane, 2006)[3] (Owusu-Edusei, 2007) [4] (Stephen Gibbons, 2013). [5]

To analyse housing prices in any given nation, it is necessary to understand how or why the demand for purchasing houses in a specific location is determined. The concern of affordability in purchasing houses has always been debated. Many factors influence variation in the price of a house, such as a locality, transportation facilities, and access to facilities like schools and hospitals. London's housing markets have had several ups and downs over a longer time frame, with volatile house prices in London tending to exaggerate fluctuations in national house prices. Although drops in the ordinary property's actual (nominal) value are unusual, London has seen numerous instances of actual house price deflation since the Office for National Statistics (ONS) began collecting data in 1969. ONS produces and distributes a

---

[1] Black, S., 1999, Do Better Schools Matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics, 114, 577-599 *https://go-gale-com.jproxy.nuim.ie/ps/i.do?p=ITOF&u=nuim&id=GALE|A54955819&v=2.1&it=r&sid=summon*

[2] Des Rosiers, F., A. Lagana, and M. Theriault, 2001, Size and Proximity Effects of Primary Schools on Surrounding House Values, Journal of Property Research, 18, 149-168. *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

[3] Thomas J. Kane, Stephanie K. Riegg and Douglas O. Staiger ,2006, School Quality, Neighborhoods, and Housing Prices, *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

[4] Owusu-Edusei, K., M. Espey, and H. Lin, 2007, Does Close Count? School Proximity, School Quality, and Residential Property Values, Journal of Agricultural and Applied Economics, 39, 211-221, *https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/abs/does-close-count-school-proximity-school-quality-and-residential-property-values/F138798DDD9911837C1593E8B2EC8CD7*

[5] Gibbons Machin, and Silva,2013, Valuing school quality using boundary discontinuities, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119012000769?via%3Dihub*

wide range of data about the United Kingdom that may inform social and economic policies and present a picture of the country as its population changes over time. This is frequently done to make comparisons with various societies and economies possible. However, based on previous cycles ' patterns, there are no clear trends from price data alone to determine if London house prices are approaching a new peak and whether it will result in a levelling down or a more extreme downward adjustment.

As education is the pathway to any nation-building, the quality of education is the most valuable commodity. Several organizations and companies offer services, often at a cost, to provide detailed information on school attributes and availability throughout the UK, to help parents decide on locations where they are most likely to secure a place in a famous or oversubscribed school. The wide availability of these programs suggests that proximity to high-quality free schools is quite essential. It seems intuitive that the closer a house is to a school, the more desirable it would be to families with school-aged children due to commuting and safety concerns. Additionally, education provided by schools could enhance economic performance and reduce crime rates in the area nearby. Households who live close to a school may also have easy access to recreational facilities. So, the desirability of living closer to schools could increase the value of a residential property with a shorter distance to schools.

The question is, how much are people willing to pay, and how important is this aspect of distance to the nearest school in estimating housing prices when compared to more traditional variables? This paper uses hedonic analysis of various house characteristics like area, the number of rooms, type of houses, and the neighbourhood characteristics like distance to the city center, distance to the nearest transit station, distance to the nearest school, and distance to roads to analyse the impact on London housing prices. More importantly, it answers whether the distance to the nearest school impacts house prices.

Few papers in the existing literature concentrate on the topic of how distance to school is capitalized into housing values. A study conducted by (Des Rosiers, 2001) [6] was the first one

---

[6] Des Rosiers, F., A. Lagana, and M. Theriault, 2001, Size and Proximity Effects of Primary Schools on Surrounding House Values, Journal of Property Research, 18, 149-168. *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

that focused exclusively on this topic. (Owusu-Edusei, 2007) [7] and (Metz, 2015) [8] aimed at exploring the same topic. The results from these studies, in general, support that houses with a smaller than average distance from schools tend to have a higher value than houses with a greater than average distance from schools. When (Des Rosiers, 2001) and (Owusu-Edusei, 2007) measured distance to schools, they used some specific ranges of distance rather than measuring distance on a continuous basis. (Metz, 2015) also used specific ranges of distance in some of his estimations. The choices of distance ranges in these studies are somewhat arbitrary. We argue that these arbitrarily chosen distance ranges weaken the conclusions of these studies because estimation results could change if a various set of ranges of distance is chosen.

Another problem with using arbitrarily chosen distance ranges is that it ignores how values of houses may change within a certain distance. On the contrary, we use a continuous distance measure, which helps mitigate the problem of the sensitivity of estimation results to arbitrarily chosen ranges of distance. Thus, our study can generate more precise and objective results about the extension of capitalization of school distance into housing price. The second addition covered in this report is the inclusion of the Spatial lag component to introduce spatial dependence into the regression model. Ordinary least squares models are used to analyse the impact of various predictors on housing prices, and the model is built with and without the distance to the nearest school parameter to evaluate the impact of the predictor on housing prices in London.

Our analysis results show a high variance in house prices with increased prices when the school is closer to the house and lower prices with lower variance as the distance increases. It also shows that the distance to the nearest school is correlated with some of the other factors and is significant when included with other spatial features rather than being considered as an essential feature. This paper adds to the overall literature concerning school

[7] Owusu-Edusei, K., M. Espey, and H. Lin, 2007, Does Close Count? School Proximity, School Quality, and Residential Property Values, Journal of Agricultural and Applied Economics, 39, 211-221, *https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/abs/does-close-count-school-proximity-school-quality-and-residential-property-values/F138798DDD9911837C1593E8B2EC8CD7*

[8] Neil E. Metz,2015, Effect of Distance to Schooling on Home Prices, *https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=23574bb6-a2ad-49fb-a11d-8b1af89b2b6c%40redis*

and house prices, while its main contribution is to expand the small strand of literature focused on distance.

2. Literature Review

In the context of schooling, the value of distance has received little attention This paper uses hedonic analysis of house transactions to estimate the impact of distance to school on sale prices in the Denver Public School District. This paper also contributes to the overall literature on school and house prices. In contrast, a significant contribution of this paper is the expansion of the small literature strand focusing on distance. (Black, 1999) [9] and (William T. Bogart, 2000) [10] were among the first to estimate the quality of schools based on school boundaries and regression discontinuity designs. Several research has investigated the capitalization of school quality into housing prices after these initial works. (Des Rosiers, 2001)[11] and (Owusu-Edusei, 2007) [12] estimated distance to schools using particular ranges of distance rather than measuring distance continuously.

(Metz, 2015) [13] also used specific distance ranges in some of his calculations. In these trials, the distance ranges were chosen randomly. In his paper, he uses one mile as the range of distance to the nearest school as a preferred factor by house owners to explain that house owners within the one mile from schools have a higher impact on house prices and the houses greater than one mile have significantly less impact on the prices being influenced by school proximity. We contend that these arbitrary distance ranges weaken the outcomes of this

[9] Black, S., 1999, Do Better Schools Matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics, 114, 577-599 *https://go-gale-com.jproxy.nuim.ie/ps/i.do?p=ITOF&u=nuim&id=GALE|A54955819&v=2.1&it=r&sid=summon*

[10] William T. Bogart, Brian A. Cromwell, 2000, How Much Is a Neighborhood School Worth? *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119099921423?via%3Dihub*

[11] Des Rosiers, F., A. Lagana, and M. Theriault, 2001, Size and Proximity Effects of Primary Schools on Surrounding House Values, Journal of Property Research, 18, 149-168. *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

[12] Owusu-Edusei, K., M. Espey, and H. Lin, 2007, Does Close Count? School Proximity, School Quality, and Residential Property Values, Journal of Agricultural and Applied Economics, 39, 211-221, *https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/abs/does-close-count-school-proximity-school-quality-and-residential-property-values/F138798DDD9911837C1593E8B2EC8CD7*

[13] Neil E. Metz,2015, Effect of Distance to Schooling on Home Prices*, https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=23574bb6-a2ad-49fb-a11d-8b1af89b2b6c%40redis*

research because estimation results could differ depending on which set of distance ranges is used. Another issue with using arbitrarily chosen distance ranges is that it ignores how housing prices may change within a certain distance. Instead, we employ a continuous distance measure, which helps to mitigate the problem of estimation result sensitivity. The general consensus is that increased school quality is reflected in higher housing prices.

(Machin, 2008)[14] and (Phuong Nguyen, 2011)[15] provide comprehensive reviews of the school quality capitalization literature . (Des Rosiers, 2001) [16] and (Owusu-Edusei, 2007) [17] on the other hand, primarily focus on the distance to school and its nonlinear relationship with housing price. Des Rosiers, Lagana, and Theriault (2001) study the impact of distance to elementary schooling on house value, considering nonlinear distance effects. Their study revealed a congestion effect near schools. Those within 1000 feet of a school suffer from traffic and noise associated with the school and are thus valued lower than houses further away. In addition, they discovered that the ideal distance between a house and elementary school in terms of house value is between 1,000 and 1,500 feet. (Thomas J. Kane, 2006)[18] and (Stephen Gibbons, 2013). [19] are two studies that examine the impact of school quality and include a control variable for distance from house to school. These papers acknowledge that controlling for distance is significant for calculating the impact of school quality on house price. However, they pay little attention to this result and the potential nonlinear nature of

---

[14] Stephen Gibbons and Stephen Machin, 2008, Valuing school quality, better transport, and lower crime: evidence from house prices, *https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=764d9dc6-5bf2-4b81-9ca9-2714562012a8%40redis*

[15] Phuong Nguyen-Hoang , John Yinger , 2011, The capitalization of school quality into house values: A review, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S1051137711000027?via%3Dihub*

[16] Des Rosiers, F., A. Lagana, and M. Theriault, 2001, Size and Proximity Effects of Primary Schools on Surrounding House Values, Journal of Property Research, 18, 149-168. *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

[17] Owusu-Edusei, K., M. Espey, and H. Lin, 2007, Does Close Count? School Proximity, School Quality, and Residential Property Values, Journal of Agricultural and Applied Economics, 39, 211-221, *https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/abs/does-close-count-school-proximity-school-quality-and-residential-property-values/F138798DDD9911837C1593E8B2EC8CD7*

[18] Thomas J. Kane, Stephanie K. Riegg and Douglas O. Staiger , School Quality, Neighborhoods, and Housing Prices, *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

[19] Gibbons Machin, and Silva,2013, Valuing school quality using boundary discontinuities, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119012000769?via%3Dihub*

value for distance between house and school. While most existing research focuses on the effect of school quality on housing prices (Black, 1999)[20] (Zabel, 2002) [21] (Lucas, 2004)[22] (David Brasington, 2006) [23] (Machin, 2008)[24] (Phuong Nguyen, 2011)[25] (Machin S. , 2011) [26]; (Bonilla-Mejía, Lopez, & McMillen, 2020) [27], this report does not consider another important school factor that may affect housing prices: distance to a school. There is a typical question house buyers consider while buying a house is it close to the nearest school.

Close proximity to a school may also provide convenient access to recreational facilities (The Empirical Economics Letters, 17(2): (February 2018). As a result, the desire to live closer to schools may lead to an increase in the value of a residential property with a shorter distance to schools. The effect of school distance on residential housing prices is investigated in this paper. Only a few papers in the existing literature focus on how school distance is factored into housing value. Moreover, amongst the ones that focus on distance, the research has aimed to find the nonlinear relationship between house prices and distance to school and to find the impact of house prices by considering ranges of distance in the form of dummy variables. In this report, we will be looking at the linear relationship of various house

[20] Black, S., 1999, Do Better Schools Matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics, 114, 577-599 *https://go-gale-com.jproxy.nuim.ie/ps/i.do?p=ITOF&u=nuim&id=GALE|A54955819&v=2.1&it=r&sid=summon*

[21] Thomas A. Downes and Jeffrey E. Zabel, 2002, The impact of school characteristics on house prices, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119002000104?via%3Dihub*

[22] David N. Figlio and Maurice E. Lucas, 2004, What's in a Grade? School Report Cards and the Housing Market, *https://www-jstor-org.jproxy.nuim.ie/stable/3592944?pq-origsite=summon&seq=1*

[23] David Brasington, Donald R. Haurin, 2006, EDUCATIONAL OUTCOMES AND HOUSE VALUES: A TEST OF THE VALUE ADDED APPROACH, https://onlinelibrary-wiley-com.jproxy.nuim.ie/doi/abs/10.1111/j.0022-4146.2006.00440.x

[24] Stephen Gibbons and Stephen Machin, 2008, Valuing school quality, better transport, and lower crime: evidence from house prices, *https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=764d9dc6-5bf2-4b81-9ca9-2714562012a8%40redis*

[25] Phuong Nguyen-Hoang , John Yinger , 2011, The capitalization of school quality into house values: A review, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S1051137711000027?via%3Dihub*

[26] Stephen Machin, 2011, Houses and schools: Valuation of school quality through the housing market, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0927537111000601?via%3Dihub*

[27] Bonilla, Lopez and McMillen 2020, House prices and school choice: Evidence from Chicago's magnet schools' proximity lottery, https://onlinelibrary-wiley-com.jproxy.nuim.ie/doi/full/10.1111/jors.12447

characteristics and spatial neighbourhood characteristics on London house prices, along with an extension to analyse the impact of distance to the nearest school as a continuous variable.

3. Data, Study area and Method

To estimate the impact of distance to schooling on housing sales, data is collected on school locations and their catchment areas, housing characteristics, house prices of each individual house in London. Existing urban and neighbourhood characteristics are also considered in the model, for which the data gathered contains the distance to centre of London, distance to the nearest transit station and the distance to the nearest road. Figure 1 contains the sample of school data collected along with the type of school along with the gender-based count of students accommodated by each school. There is a total of 3889 school points in the dataset. The housing data considered in Figure 2 shows the housing points of London listed in 2011. Figure 2 shows the various housing characteristics of each house which include the number of rooms, age, area of the house, etc.

The data of schools considered for this analysis involves all types of schools, i.e., primary, secondary, and high schools. In this report, we will be considering the distance of each house to the nearest school regardless of what school it is. The nearest distance is considered to filter out the closest school to each house. This distance measure is considered for the analysis to explain the variation in housing prices across London (Figure 3) Yellow depicts close distances and green depicts distances of houses that are farther away from school.

Figure 4 shows the school area density variations across boroughs in London. The colours in the map represent the area-based density of schools by boroughs. The school density plot Figure 5 indicates a heavily right-skewed distribution which means that the mean is more significant than the median as there are quite a few boroughs with higher school density. Analysis was conducted to choose the best way to bucket the area density amongst three typical classification schemes to understand more about the data distribution. We looked at Equal Interval, Mean-Standard deviation, and Fisher-Jenks schemes. The basic concept of the equal interval scheme is that each bin contains an equal width (w) of the attribute value for a specified number of bins(k). Figure 6 shows the splits classified by Equal intervals. Each of the bins has the exact width of w = 1.4. This value of k = 5 corresponds directly to the default histogram displayed above. In order to check for outliers, the Mean-Deviation scheme helps. For the default k = 5, the standard definition is to set the upper and lower bins as above and below the mean by two standard deviations (respectively).

In contrast, the fourth and second bins are set to within one standard deviation of the attribute, and the middle bin straddles the attribute mean. Any values larger or smaller than two standard deviations from the mean are placed into the upper and lower bins. Figure 7 depicts the split results where we see that two boroughs are considered outliers when looking at school point densities concerning areas of boroughs. However, the mean-standard deviation approach would be helpful if we had a normal distribution, but in this case, it is right skewed. Finally, to minimize the sum of absolute deviations around class means (ADCM) (Figure 8), i.e., to increase within-group similarity as much as possible, the analysis covers the Fisher-Jenks classification scheme (Figure 4).

Global spatial autocorrelation is a significant part of spatial analysis. For example, it measures the extent to which a value of a particular house is dependent on the house prices of its neighbours. The most basic version of assessing this correlation is the "join-count" statistic, which measures spatial association in terms of a binary variable. The traditional case is enumerated in terms of "black" and "white" cells (as on a chessboard); in this case, we will re-categorize our mean price per square foot variable in terms of "high" values (i.e., "black") over £6,500 per square meter. The split in the distribution of cells is represented by a scatter plot (Figure 9), and the resulting map of spatial autocorrelation is shown in (Figure: spatial autocorrelation). From the map, there seems to be a clear pattern of spatial autocorrelation: black cells tend to be located near other black cells, and white cells near one another (i.e., there is a stark cluster of cells with high values of mean price of houses near the London city center). The join count statistic tests this in a very straightforward way: by counting up the number of various types of joins (i.e., black-black, black-white, and white-white) observed and comparing those counts to the null hypothesis "expected" the number of counts based on the assumption that black and white cells are randomly distributed across the region (this assumption is called complete spatial randomness or CSR). Possible joins are evaluated based on the given spatial weight's matrix, and the hypothesis confirms spatial autocorrelation between black-black and black-white cells.

Although housing prices are observed, the underlying characteristics determining housing prices are unobserved or implicit. Like most existing research, our empirical analysis is conducted based on the hedonic method developed by (Rosen, 1974) [28]. (Rosen, 1974) showed that the implicit prices or "hedonic prices" of house characteristics or neighbourhood attributes could be revealed by regressing housing prices on the measures of these characteristics. An estimated coefficient on a characteristic variable represents an extra price that would be paid to an additional unit of that variable. To analyse the effects of different characteristics on housing prices, our basic estimation hedonic model is:

Model 1- $Y = X1\beta1 + X2\beta2 + \upsilon$

Y = price which has been log-transformed

$X1$ = vector of housing characteristics (e.g., number of rooms, no of beds),

$X2$ = vector of the neighbourhood and environmental characteristics (e.g., distance to the nearest school, distance to the city center, distance to the nearest transit station, etc.), and

υ = error term

The first term of this equation is straightforward - the basic assumption is that the price of a house will be significantly related to how large it is, how many rooms it has, and what kind of house it is. The second term looks at the influence of neighbourhood characteristics and their impact on house prices.

Before diving into the hedonic model, Figure 10 shows the distribution of house prices. As it is skewed, log transformation is applied to the prices to provide a better fit (Figure 11). A similar transformation is applied to the area parameter as well. The models are built with and without the nearest school distance to analyse the impact of the nearest school to a house. A quadratic term of the nearest school distance is used to identify the behaviour of the parameter. The scatter plot (Figure 12) shows the raw relationship between the log(price) and the nearest distance to school from each house in London. The model results suggest signs of residual spatial structure in the residuals. To explicitly suggest the need for spatial regression, we have considered the spatial weights matrix(w) to calculate the residual Moran's I and a

[28] Sherwin Rosen, 1974, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *https://www-jstor-org.jproxy.nuim.ie/stable/1830899?pq-origsite=summon&seq=1*

few Lagrange Multipliers for Spatial lag, spatial error, and spatial lag + error models. From the Figure 13, all of our spatial diagnostic tests are significant, suggesting significant spatial dependence in the model specifications.

To introduce the spatial dependence in regression Model 1, spatial lags of the X variable are included in the model. The enhanced model looks like this:

Model 2- $Y = X\beta + WX\gamma + \epsilon$

Where W is the spatial weights matrix and γ is the set of spatial lag coefficients for the independent variables. This model can be estimated using OLS because the spatial lags do not violate any basic assumptions (as other spatial models do) - they are additional exogenous variables. The model is developed with and without the nearest school distance parameter to analyse how much variation in London house prices is explained by the nearest school distance

4. Results

The Figure 14 and hypothesis tests confirm that the independent variables used to explain the variance of house prices correlate with the neighborhood's parameter values. Several specifications are estimated to explore the relationship between distance to assigned public schooling and house price. Ordinary least square models are built in stages by adding house and neighbourhood parameters to analyse the influence of the parameters on the housing prices. For better linear fit, the area and price parameters are log-transformed, and the first group of housing characteristics is modelled to explain the variability of housing prices. The factors include property type, number of rooms, and the house area. Figure 15 represents the results of the model fit. The adjusted R-Square observed is 0.467, and all the parameters are significant.

Before adding the spatial parameters, residual variation is verified to be significant by neighbourhood (Figure 16). Neighbourhoods on the left of the scale have consistently lower observed values of housing prices than the model would predict based on their raw housing characteristics (i.e., negative residuals), and neighbourhoods on the right have consistently higher values than the model would predict (i.e., positive residuals). It appears that a lot of the boroughs on the high and low end of the residual spectrum are close to one another in space. It is verified by creating a distance-based spatial weights matrix and generating spatial lags of the residuals. In Figure 17, the regression line shows a significant positive correlation between the model's residuals in a given point and the residuals of the model in its neighbours, thereby confirming that the prediction errors are clustering.

Therefore, we introduced the spatial components to the existing model to control spatial heterogeneity. The new model included the independent parameters of distance to the city center, distance to the nearest transit station, and distance to the nearest road for every house. These additional spatial neighbourhood parameters explained more of the variation in London house prices (Figure 18) as the adjusted R square is 0.71. Adding in the nearest school distance parameter turned out to be insignificant(Figure 19). The nearest distance to school parameter as a raw feature does not capture any more variance in the prices that the existing variables fail to capture. However, adding in the quadratic term of distance to school confirms that the effect of distance to school is a convex graph where the house prices have a higher variance in price when the distance to school is low and the effect of price gradually reduces as the distance increases. (Figure 20) This is indicated by the positive value (0.000002) observed in the squared term and a negative value (-0.0002) observed in the distance term. (Figure 12). In order to introduce spatial dependence directly into the regression specifications, we have considered adding the spatial lag of variables as independent parameters for the hedonic model. The results of this model (Figure 21) show an increase in adjusted r square value, i.e 0.7559. All the variables, including the nearest distance to school, are now significant. This says that the nearest school distance is correlated with some of the other spatial lag variables and is significant when included in the model. This indicates an indirect impact on London house prices.

Thus, the results say that the hypothesis of a linear relationship between the distance to school and house prices has been rejected as the raw variable of distance to school is insignificant. However, the nearest school distance has an indirect effect on prices and is significant when the neighbourhood parameters are included. We also see high variability in house prices when the distance to school is less, but the houses farther away from school seem to have lower prices.

5. Figures

URN	SCHOOL_NAM	TYPE	PHASE	ADDRESS	TOWN	STATUS	GENDER	WARD_NAME	LSOA_NAME	LA_NAME	
135155	Ayesha Siddiqa Girls School	Other Independent School	Not applicable	165-169 The Broadway	Southall	Open	Girls	Southall Broadway	Ealing 026C	Ealing	
140492	Beis Medrash Elyon	Other Independent School	Not applicable	233 West Hendon Broadway	London	Open	Boys	West Hendon	Barnet 036F	Barnet	
141411	Big Creative Independent School	Other Independent School	Not applicable	Silver Birch House	Walthamstow	Open	Mixed	Higham Hill	Waltham Forest 014C	Waltham Forest	
142336	Wetherby Senior School	Other Independent School	Not applicable	100 Marylebone Lane	London	Open	Boys	Marylebone High Street	Westminster 011B	Westminster	
100042	St Mary's Kilburn Church of England Primary Sc...	Voluntary Aided School	Primary	Quex Road	London	Open	Mixed	Kilburn	Camden 020C	Camden	http://www.stmarykilburn.cam

WEBLINK	AGE	Primary	map_icon_l	POSTCODE	NoofBoys	NoofGirls	NumberOfPu	PerFSM	Tele
None	11 - 19	0	2	UB1 1LR	0	85	84	0.0	na
None	11 - 16	0	2	NW9 7DG	80	0	78	0.0	na
None	15 - 16	0	2	E17 5SD	5	5	9	0.0	na
None	11 - 16	0	2	W1U 2QB	0	0	0	0.0	na
http://www.stmarykilburn.camden.sch.uk/	3 - 11	1	2	NW6 4PG	115	110	227	29.1	0207 3726565

Figure 1- The sample of school data

postcode	price	dateoftransfer	propertytype	oldnew	duration	categorytype	recordstatus	year	tfarea	numberrooms	priceper	pcd	pcd2
CR53EZ	795000.0	2012-03-15	D	Y	F	A	A	2012.0	250.0	7.0	3180.000000	CR5 3EZ	CR5 3EZ
CR53EZ	1250000.0	2015-12-18	D	N	F	A	A	2015.0	279.0	9.0	4480.286738	CR5 3EZ	CR5 3EZ
CR53EZ	1000000.0	2011-12-22	D	Y	F	A	A	2011.0	302.0	9.0	3311.258278	CR5 3EZ	CR5 3EZ
CR53EZ	970000.0	2015-12-04	D	N	F	A	A	2015.0	272.0	9.0	3566.176471	CR5 3EZ	CR5 3EZ
CR53EZ	995000.0	2016-10-18	D	N	F	A	A	2016.0	288.0	7.0	3454.861111	CR5 3EZ	CR5 3EZ

Figure 2 - The housing points of London

Figure 3 - Distance to nearest school

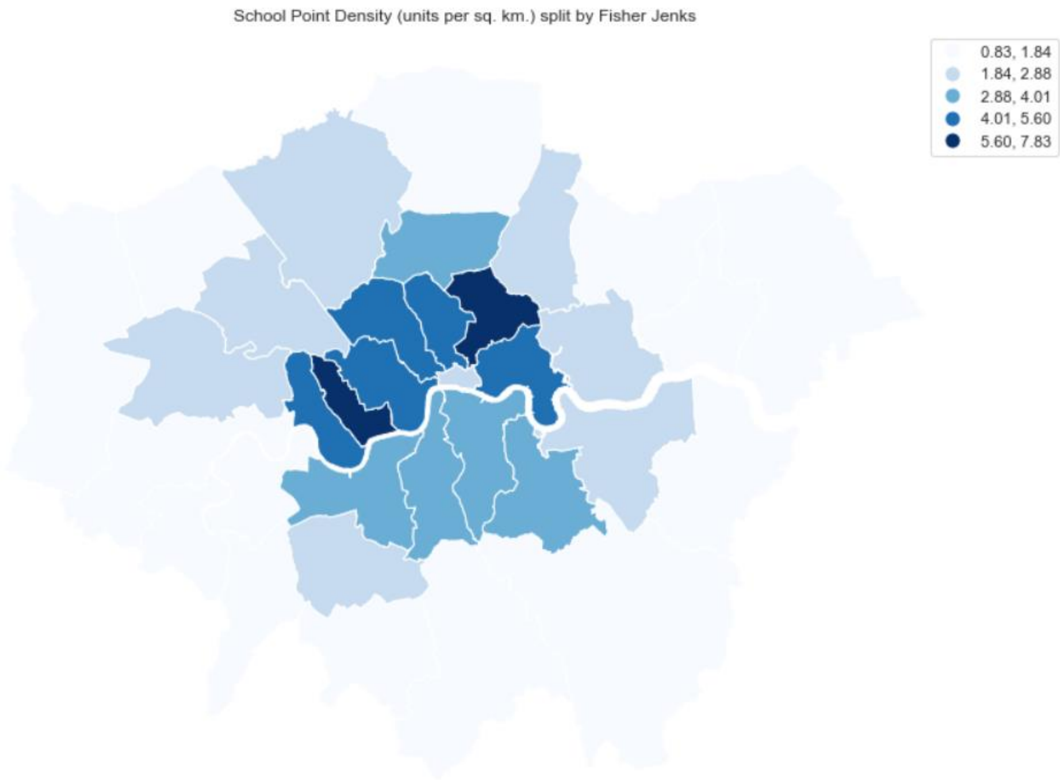

Figure 4- The school area density variations across boroughs in London.

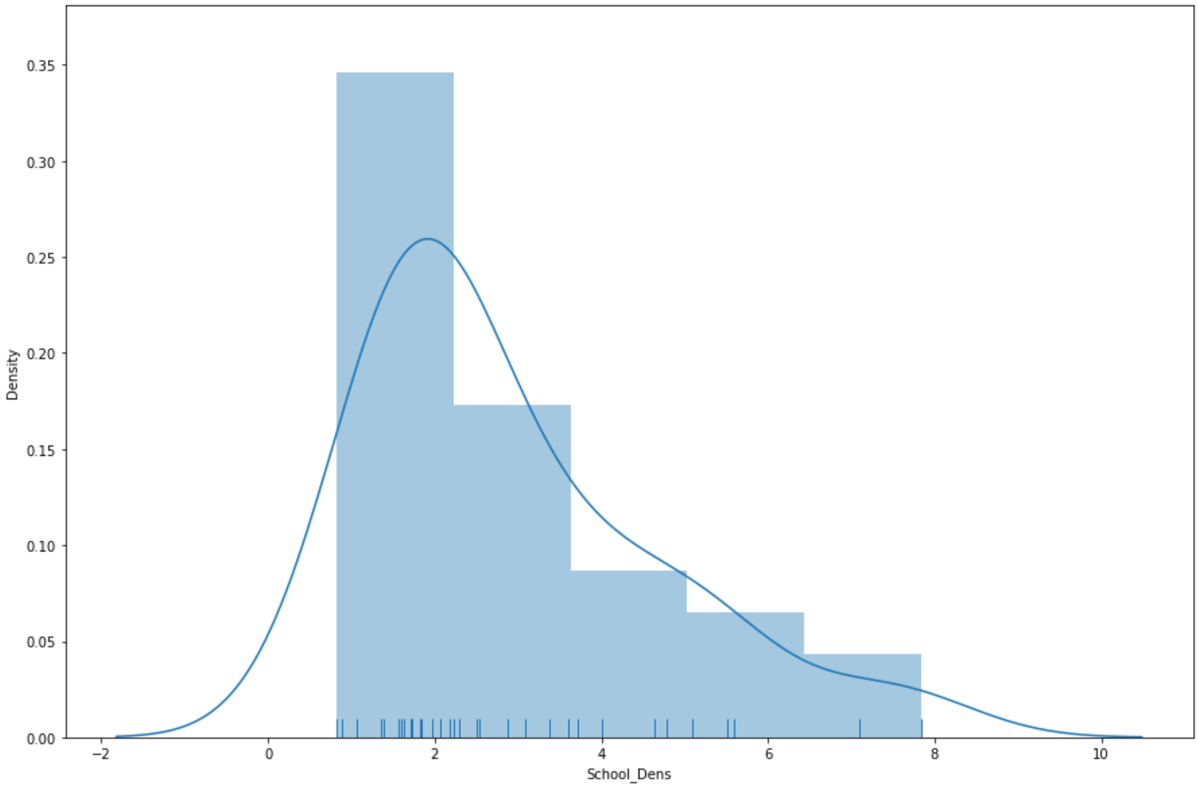

Figure 5 - The school density plot

```
EqualInterval

  Interval      Count
--------------------
[0.83, 2.23] |    16
(2.23, 3.63] |     8
(3.63, 5.03] |     4
(5.03, 6.43] |     3
(6.43, 7.83] |     2
```

Figure 6- The splits classified by Equal Interval

```
StdMean

   Interval       Count
----------------------
( -inf, -0.64] |     0
(-0.64,  1.14] |     3
( 1.14,  4.69] |    24
( 4.69,  6.47] |     4
( 6.47,  7.83] |     2
```

Figure 7- The results of the split

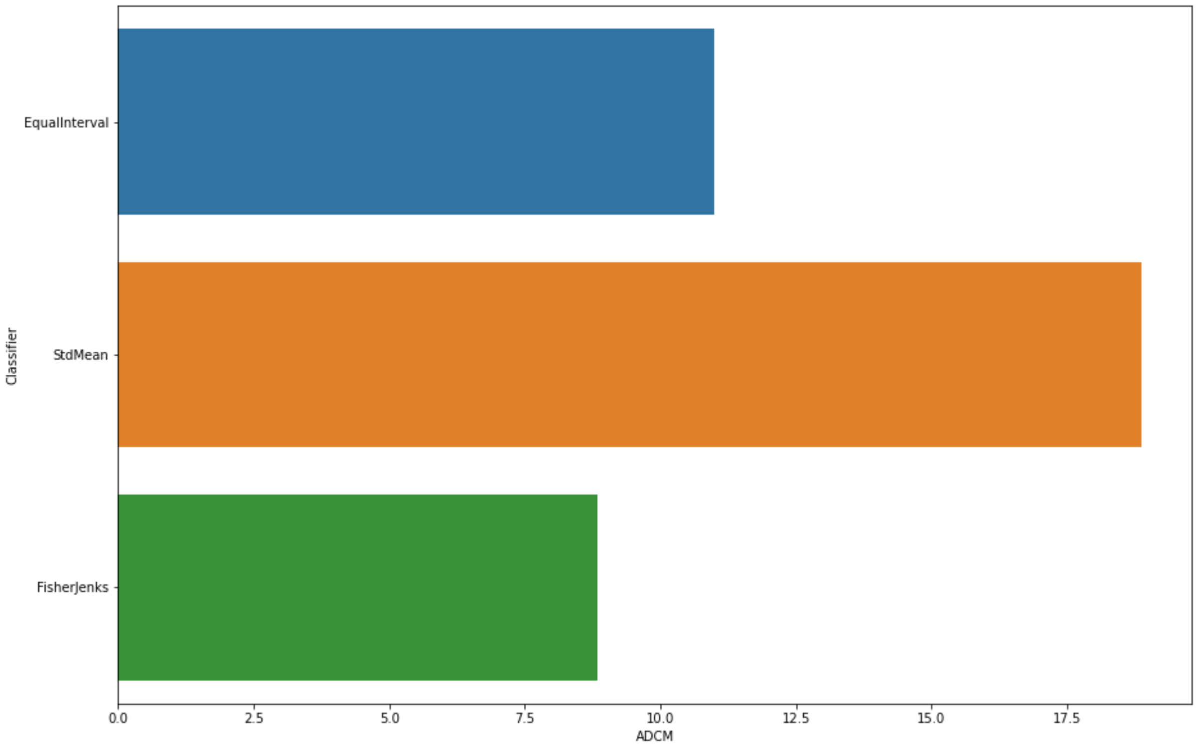

Figure 8- The graph of ADCM vs Classifier

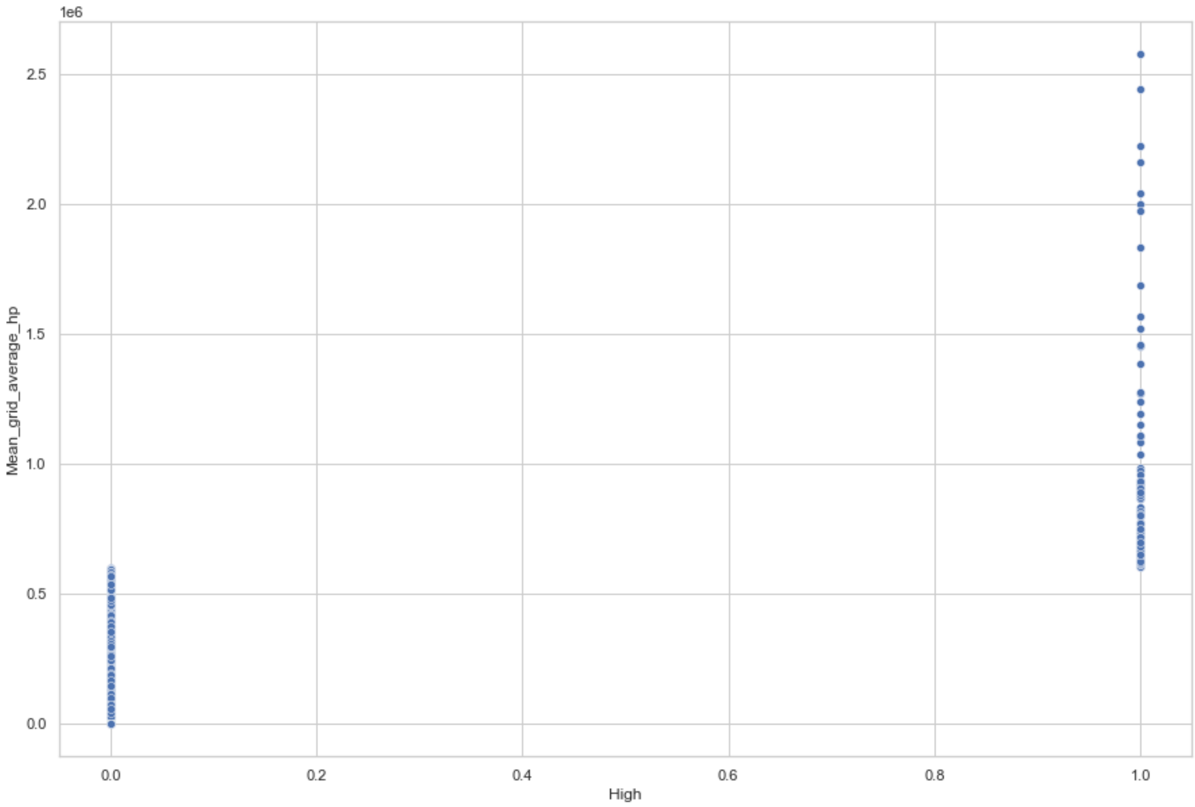

Figure 9- Scatter plot

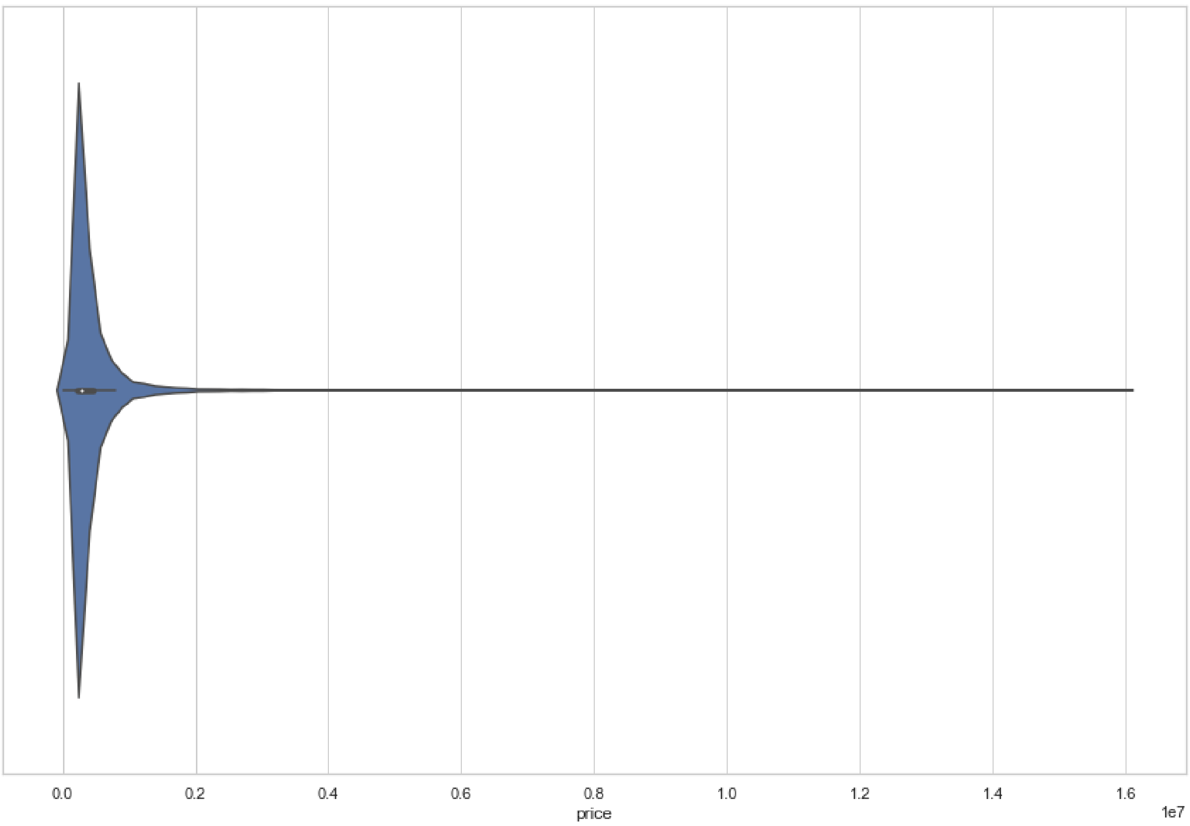

Figure 10- The distribution of house prices.

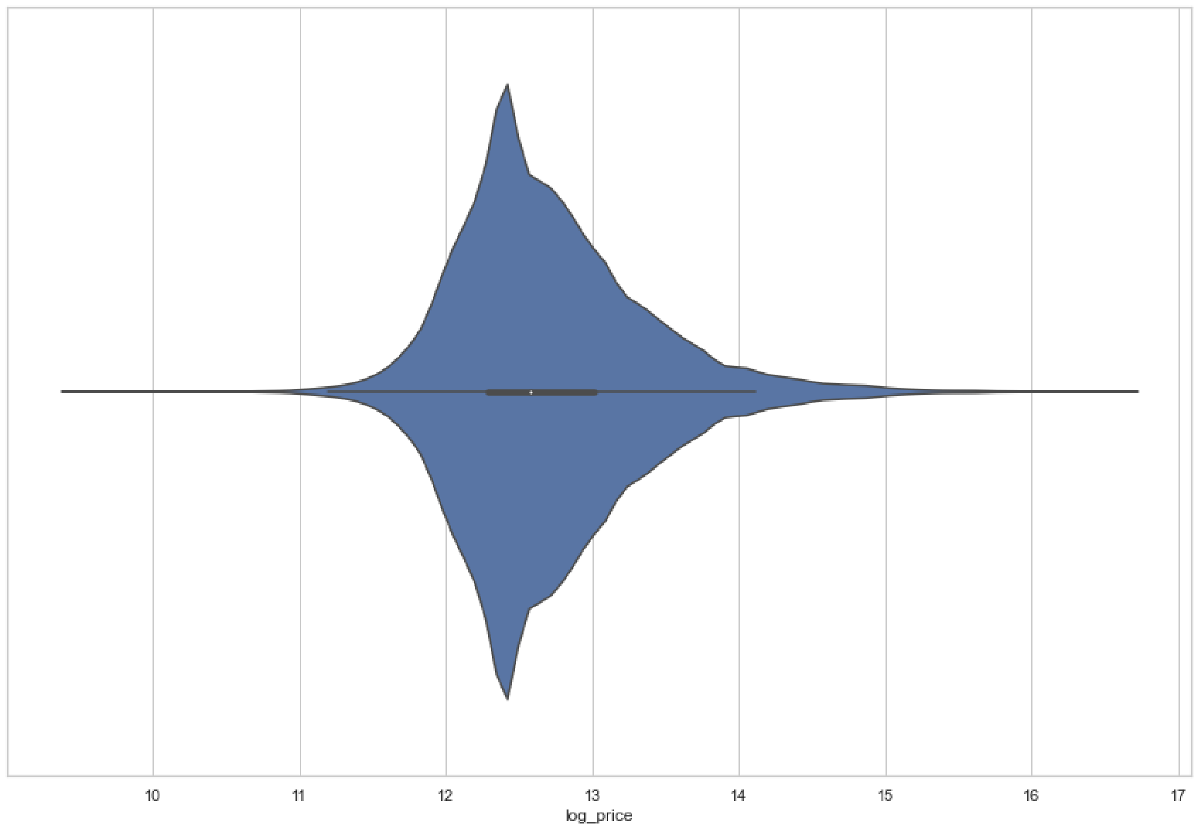

Figure 11- Log transformation of price

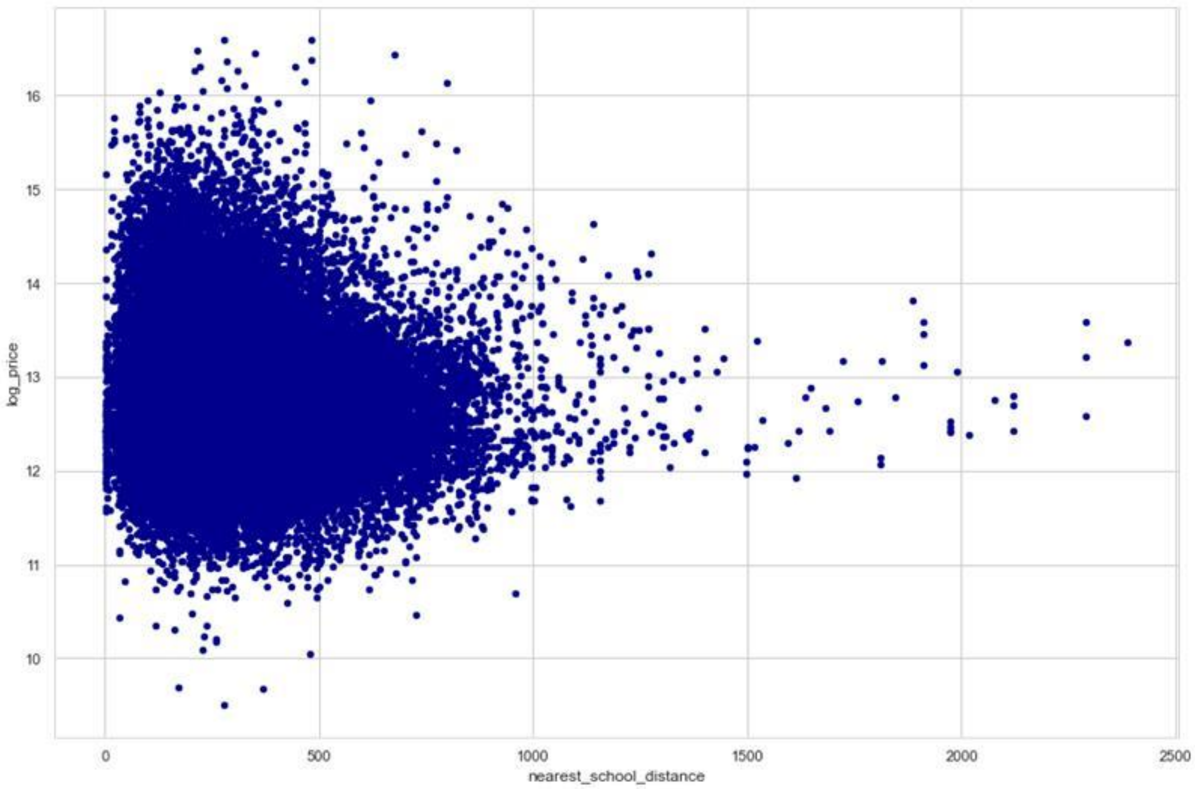

Figure 12- Relationship between the log(price) and the nearest distance to school

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           73.872

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2      16349.498        0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               11       6375.225        0.0000
Koenker-Bassett test             11       2962.800        0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)             0.5922        937.420        0.0000
Lagrange Multiplier (lag)         1       7156.879        0.0000
Robust LM (lag)                   1       5712.654        0.0000
Lagrange Multiplier (error)       1     871758.614        0.0000
Robust LM (error)                 1     870314.389        0.0000
Lagrange Multiplier (SARMA)       2     877471.267        0.0000
```

Figure 13- Spatial Diagnostics

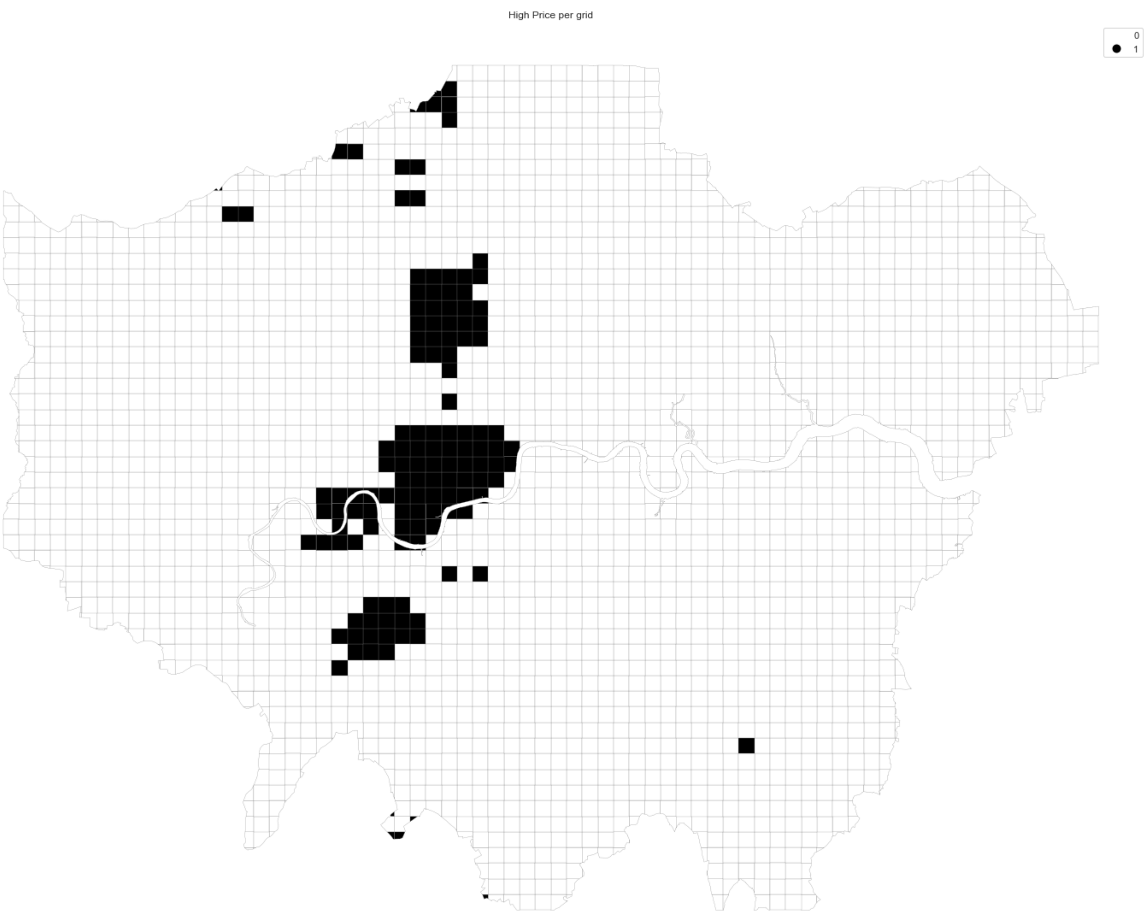

Figure 14 - Spatial Autocorrelation

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       71733
Mean dependent var  :     12.6945                Number of Variables   :           6
S.D. dependent var  :      0.6363                Degrees of Freedom    :       71727
R-squared           :      0.4677
Adjusted R-squared  :      0.4677
Sum squared residual:   15458.531                F-statistic           :  12606.6256
Sigma-square        :       0.216                Prob(F-statistic)     :           0
S.E. of regression  :       0.464                Log likelihood        :  -46737.177
Sigma-square ML     :       0.216                Akaike info criterion :   93486.354
S.E of regression ML:      0.4642                Schwarz criterion     :   93541.438

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       7.9585305       0.0356006     223.5507217       0.0000000
            log_area       1.0176671       0.0094878     107.2608530       0.0000000
         numberrooms       0.0379068       0.0023346      16.2368312       0.0000000
               Flats       0.2669421       0.0048301      55.2660932       0.0000000
            Detached       0.0943900       0.0071864      13.1345075       0.0000000
                 New       0.0710447       0.0168918       4.2058789       0.0000260
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           45.585

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2        5864.555           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                5        3732.302           0.0000
Koenker-Bassett test              5        2441.119           0.0000
================================ END OF REPORT =====================================
```

Figure 15 - The results of the model fit with house parameters

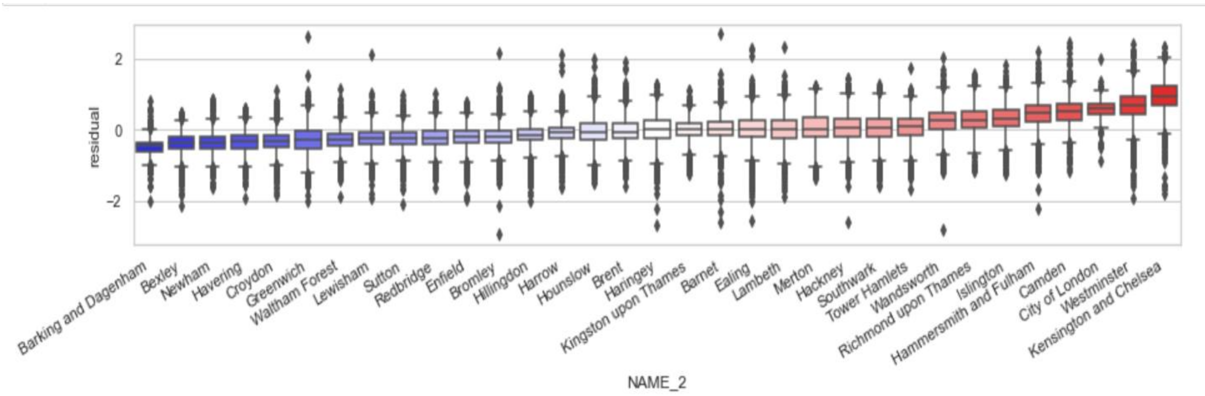

Figure 16- Residuals by borough

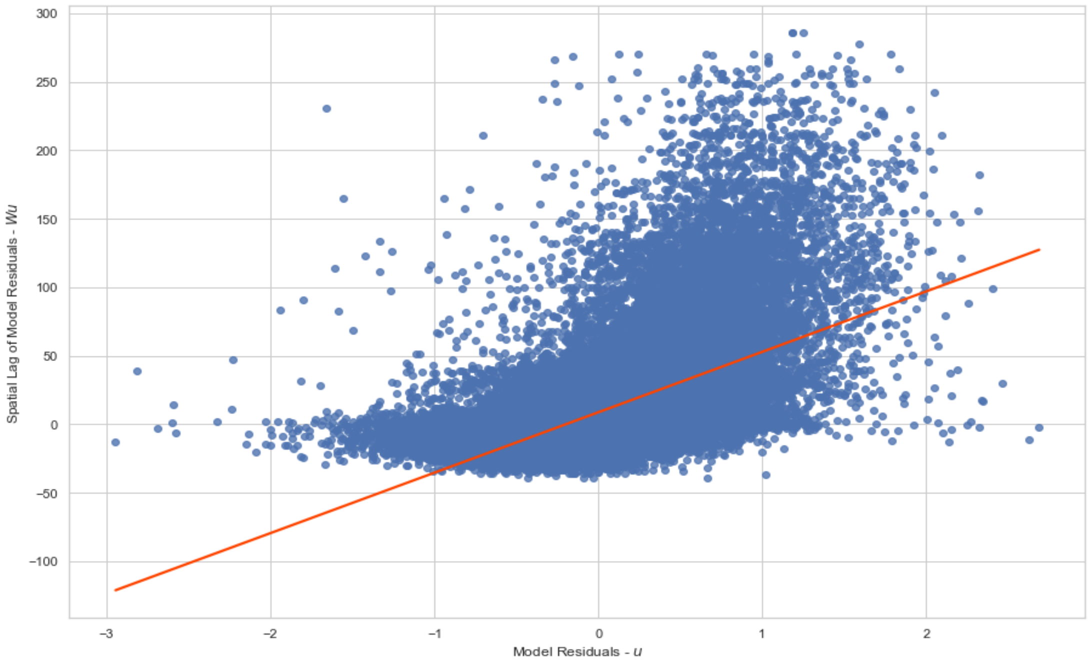

Figure 17- Moran's I Plot

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       71733
Mean dependent var  :     12.6945                Number of Variables   :          11
S.D. dependent var  :      0.6363                Degrees of Freedom    :       71722
R-squared           :      0.7109
Adjusted R-squared  :      0.7108
Sum squared residual:    8397.437                F-statistic           :  17633.5631
Sigma-square        :       0.117                Prob(F-statistic)     :           0
S.E. of regression  :       0.342                Log likelihood        :  -24850.203
Sigma-square ML     :       0.117                Akaike info criterion :   49722.406
S.E of regression ML:      0.3421                Schwarz criterion     :   49823.394

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.6775519       0.0294317     362.7908731       0.0000000
            log_area       0.7215081       0.0073023      98.8056228       0.0000000
         numberrooms       0.0391151       0.0017625      22.1932785       0.0000000
               Flats      -0.0441289       0.0039977     -11.0386934       0.0000000
            Detached       0.2425130       0.0059230      40.9443739       0.0000000
                 New       0.1289278       0.0142611       9.0404986       0.0000000
            Dist_0KM      -0.0000476       0.0000003    -160.9923518       0.0000000
           Dist_Road       0.0000146       0.0000012      12.2401171       0.0000000
        Dist_Transit      -0.0000125       0.0000025      -5.0172716       0.0000005
      Dist_OpenSpace      -0.0001115       0.0000093     -12.0446220       0.0000000
              DEPRHH      -1.2849534       0.0094852    -135.4700110       0.0000000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           70.056

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2        16347.592         0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               10         6360.697         0.0000
Koenker-Bassett test             10         2956.132         0.0000
================================ END OF REPORT =====================================
```

Figure 18- The results of the model fit with house and neighborhood parameters

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       71733
Mean dependent var  :     12.6945                Number of Variables   :          12
S.D. dependent var  :      0.6363                Degrees of Freedom    :       71721
R-squared           :      0.7109
Adjusted R-squared  :      0.7108
Sum squared residual:    8397.420                F-statistic           :  16030.3336
Sigma-square        :       0.117                Prob(F-statistic)     :           0
S.E. of regression  :       0.342                Log likelihood        :  -24850.131
Sigma-square ML     :       0.117                Akaike info criterion :   49724.263
S.E of regression ML:      0.3421                Schwarz criterion     :   49834.431

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.6771287       0.0294265     362.8400330       0.0000000
            log_area       0.7214631       0.0073057      98.7529968       0.0000000
         numberrooms       0.0391109       0.0017625      22.1907164       0.0000000
               Flats      -0.0441849       0.0040008     -11.0439041       0.0000000
            Detached       0.2423715       0.0059218      40.9283686       0.0000000
                 New       0.1289262       0.0142597       9.0412961       0.0000000
            Dist_0KM      -0.0000476       0.0000003    -156.8851296       0.0000000
           Dist_Road       0.0000146       0.0000012      12.2254784       0.0000000
        Dist_Transit      -0.0000126       0.0000025      -5.0378559       0.0000005
      Dist_OpenSpace      -0.0001114       0.0000093     -12.0369575       0.0000000
              DEPRHH      -1.2846980       0.0095236    -134.8956610       0.0000000
nearest_school_distance       0.0000031       0.0000077       0.3953641       0.6925754
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           73.872

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2        16349.498          0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               11        6375.225           0.0000
Koenker-Bassett test             11        2962.800           0.0000
================================ END OF REPORT =====================================
```

Figure 19- Model fit with nearest school distance parameter

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       71733
Mean dependent var  :     12.6945                Number of Variables   :          13
S.D. dependent var  :      0.6363                Degrees of Freedom    :       71720
R-squared           :      0.7116
Adjusted R-squared  :      0.7115
Sum squared residual:    8376.729                F-statistic           :  14745.3276
Sigma-square        :       0.117                Prob(F-statistic)     :           0
S.E. of regression  :       0.342                Log likelihood        :  -24761.645
Sigma-square ML     :       0.117                Akaike info criterion :   49549.291
S.E of regression ML:      0.3417                Schwarz criterion     :   49668.640

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.7174117       0.0295867     362.2378405       0.0000000
            log_area       0.7204296       0.0072942      98.7673519       0.0000000
         numberrooms       0.0392767       0.0017603      22.3124524       0.0000000
               Flats      -0.0456305       0.0039994     -11.4093093       0.0000000
            Detached       0.2380689       0.0059206      40.2104040       0.0000000
                 New       0.1271292       0.0142920       8.8950965       0.0000000
            Dist_0KM      -0.0000473       0.0000003    -155.8974358       0.0000000
           Dist_Road       0.0000140       0.0000012      11.7235930       0.0000000
        Dist_Transit      -0.0000168       0.0000025      -6.6326647       0.0000000
      Dist_OpenSpace      -0.0001169       0.0000092     -12.6838495       0.0000000
              DEPRHH      -1.2856980       0.0095229    -135.0110187       0.0000000
nearest_school_distance      -0.0002073       0.0000205     -10.0937719       0.0000000
sq_nearest_school_distance       0.0000002       0.0000000      10.8248218       0.0000000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           76.270

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       16345.786           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               12        6494.106           0.0000
Koenker-Bassett test             12        3017.252           0.0000
================================ END OF REPORT =====================================
```

Figure 20- Model fit with quadratic term of distance to nearest school

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :   log_price                Number of Observations:       71733
Mean dependent var  :     12.6945                Number of Variables   :          17
S.D. dependent var  :      0.6363                Degrees of Freedom    :       71716
R-squared           :      0.7559
Adjusted R-squared  :      0.7559
Sum squared residual:    7089.294                F-statistic           :  13880.5995
Sigma-square        :       0.099                Prob(F-statistic)     :           0
S.E. of regression  :       0.314                Log likelihood        :  -18776.539
Sigma-square ML     :       0.099                Akaike info criterion :   37587.079
S.E of regression ML:      0.3144                Schwarz criterion     :   37743.151

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.3740995       0.0288024     360.1814797       0.0000000
            log_area       0.6725990       0.0069869      96.2660941       0.0000000
         numberrooms       0.0436703       0.0016775      26.0322942       0.0000000
               Flats      -0.1455764       0.0039564     -36.7949315       0.0000000
            Detached       0.1952959       0.0057386      34.0319053       0.0000000
                 New       0.1324888       0.0134912       9.8204004       0.0000000
            Dist_0KM      -0.0000298       0.0000004     -82.2206286       0.0000000
           Dist_Road       0.0000138       0.0000011      12.5707866       0.0000000
        Dist_Transit      -0.0000051       0.0000022      -2.3288105       0.0198719
      Dist_OpenSpace      -0.0001977       0.0000086     -22.8590512       0.0000000
              DEPRHH      -0.9926831       0.0103514     -95.8987061       0.0000000
nearest_school_distance       0.0000502       0.0000072       6.9534678       0.0000000
          w_log_area      -0.0031712       0.0000783     -40.4780262       0.0000000
       w_numberrooms       0.0026298       0.0000670      39.2456113       0.0000000
             w_Flats       0.0100204       0.0001458      68.7106283       0.0000000
          w_Detached       0.0091699       0.0004091      22.4129767       0.0000000
               w_New      -0.0021108       0.0005092      -4.1449250       0.0000340
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           96.339

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2      34946.604         0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               16       8808.333         0.0000
Koenker-Bassett test             16       3255.540         0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.4235      674.370         0.0000
Lagrange Multiplier (lag)         1      27391.253         0.0000
Robust LM (lag)                   1        217.956         0.0000
Lagrange Multiplier (error)       1     445891.777         0.0000
Robust LM (error)                 1     418718.480         0.0000
Lagrange Multiplier (SARMA)       2     446109.733         0.0000

================================ END OF REPORT =====================================
```

Figure 21- Spatial Lag Model

6. Discussion and Conclusion

We see a higher variance in prices of houses closer to school and lower variance with low prices for houses farther away. This says that households are concerned with locating a school nearby more often than not. This paper shows that Distance to school is a significant factor that affects housing prices when included with various neighborhood parameters. However, a continuous distance to the nearest school parameter as a raw feature does not significantly correlate with London house prices. Compared to the previous research, our study analyses the impact of Distance to school continuously rather than arbitrarily chosen ranges of Distance to school. The fitted model is the hedonic model with linear regression, which separates this analysis from the existing ones that use non-linear and quantile, regression models. All the schools in London are given equal weightage, and there is no differentiation between the type of schools (e.g., primary, secondary, and high schools) or the quality of education. The results also aim to prove that there is spatial autocorrelation between the London house prices and various other spatial parameters such as Distance to the city center, Distance to the nearest transit station, and Distance to the nearest road have an impact on house prices. It is also significant in considering house parameters such as the area of the house, number of rooms, and type of the house.

This study contributes to the existing literature by stressing the importance of considering the distance from a house to school and school quality when households and policymakers value a house. These results have some straightforward policy implications. Since the homeowner's value is closer to schooling, care should be taken in the placement of schools. Boundaries should be drawn to minimize the overall Distance to homes within the school district. Several recent studies have focused on the quality of schools to address the price variations in houses across London. This paper opens up to possibilities of exploring questions around how the prices might vary with the interaction between the Distance to school parameter and the quality of school parameter. This paper might also enable researchers to consider various non-linear models to fit the parameters discussed. There could also be merit in splitting the schools into Primary, Secondary and high school categories to analyse the impact of prices. It might also be worth exploring ranges of Distance to the nearest school to better understand its impact on London house prices. Also, further research is needed to consider what features of a school make a "good school" (at least to parents).

References

1. Black, S., 1999, Do Better Schools Matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics, 114, 577-599 *https://go-gale-com.jproxy.nuim.ie/ps/i.do?p=ITOF&u=nuim&id=GALE|A54955819&v=2.1&it=r&sid=summon*

2. Des Rosiers, F., A. Lagana, and M. Theriault, 2001, Size and Proximity Effects of Primary Schools on Surrounding House Values, Journal of Property Research, 18, 149-168. *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

3. Thomas J. Kane, Stephanie K. Riegg and Douglas O. Staiger ,2006, School Quality, Neighborhoods, and Housing Prices, *https://www-jstor-org.jproxy.nuim.ie/stable/42705497?pq-origsite=summon&seq=1*

4. Owusu-Edusei, K., M. Espey, and H. Lin, 2007, Does Close Count? School Proximity, School Quality, and Residential Property Values, Journal of Agricultural and Applied Economics, 39, 211-221, *https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/abs/does-close-count-school-proximity-school-quality-and-residential-property-values/F138798DDD9911837C1593E8B2EC8CD7*

5. Gibbons Machin, and Silva,2013, Valuing school quality using boundary discontinuities, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119012000769?via%3Dihub*

6. Neil E. Metz,2015, Effect of Distance to Schooling on Home Prices, *https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=23574bb6-a2ad-49fb-a11d-8b1af89b2b6c%40redis*

7. William T. Bogart, Brian A. Cromwell, 2000, How Much Is a Neighborhood School Worth? *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119099921423?via%3Dihub*

8. Stephen Gibbons and Stephen Machin, 2008, Valuing school quality, better transport, and lower crime: evidence from house prices, *https://web-p-ebscohost-com.jproxy.nuim.ie/ehost/pdfviewer/pdfviewer?vid=0&sid=764d9dc6-5bf2-4b81-9ca9-2714562012a8%40redis*

9. Phuong Nguyen-Hoang, John Yinger , 2011, The capitalization of school quality into house values: A review, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S1051137711000027?via%3Dihub*

10. Thomas A. Downes and Jeffrey E. Zabel, 2002, The impact of school characteristics on house prices, *https://www-sciencedirect-com.jproxy.nuim.ie/science/article/pii/S0094119002000104?via%3Dihub*

11. David N. Figlio and Maurice E. Lucas, 2004, What's in a Grade? School Report Cards and the Housing Market, *https://www-jstor-org.jproxy.nuim.ie/stable/3592944?pq-origsite=summon&seq=1*

12. David Brasington, Donald R. Haurin, 2006, EDUCATIONAL OUTCOMES AND HOUSE VALUES: A TEST OF THE VALUE-ADDED APPROACH, *https://onlinelibrary-wiley-com.jproxy.nuim.ie/doi/abs/10.1111/j.0022-4146.2006.00440.x*

13. Bonilla, Lopez and McMillen 2020, House prices and school choice: Evidence from Chicago's magnet schools' proximity lottery, *https://onlinelibrary-wiley-com.jproxy.nuim.ie/doi/full/10.1111/jors.12447*

14. Sherwin Rosen, 1974, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *https://www-jstor-org.jproxy.nuim.ie/stable/1830899?pq-origsite=summon&seq=1*