A study on the impact of Chicago L Train (Elevated Train) accessibility on housing prices

Department of Urban and Regional Planning
Dasom Han

Abstract

Housing prices are subject to a multitude of influences, encompassing individual house characteristics, the surrounding environment, economic conditions, and policy shifts. This study, focused on Chicago, recognized the likelihood of regional variations in the factors impacting housing prices. An in-depth analysis was conducted to identify the significant determinants of housing prices in specific regions. Notably, this study hypothesized a substantial influence of the Chicago L(short for "elevated") train on housing prices, leading to separate examinations for areas within and beyond a 1-mile radius of the L train. The analytical tools employed included multiple regression and geographically weighted regression (GWR).

The findings revealed that, across all analysis methods, the individual characteristics of houses (such as the number of rooms, year of construction, parking availability, etc.) exerted the greatest influence on housing prices in both the areas within and outside the 1-mile radius of the L train. Additionally, variables related to the surrounding environment (such as distance to the station and proximity to parks) were identified as factors influencing housing prices, although their impact was comparatively less pronounced than that of individual housing characteristics.

Disparities emerged between the outcomes of the two analysis methods. Multiple regression analysis indicated a positive correlation between house price and proximity to the station, while GWR yielded an opposing result. A detailed examination by region highlighted that, with the exception of the northern part of Chicago, housing prices generally increased with greater distance from the station, or statistical significance was not confirmed. Moreover, when scrutinizing areas within and outside the 1-mile radius of the station separately, GWR did not exhibit enhanced explanatory power for houses situated more than 1 mile away from the station.

In summary, GWR generally produced superior analytical outcomes, yet this trend was not consistent across all regions. Consequently, this study corroborates the notion that the factors influencing housing prices vary by region and emphasizes that GWR may not uniformly offer superior explanatory power in all geographic areas.

1. Introduction

Chicago, the third-largest city in the United States, following New York and Los Angeles, stands renowned for its rich history, diverse culture, and economic significance. Its evolution into a major land transportation hub commenced in the mid-19th century, fostering a vital connection between the Great Lakes and the Mississippi River. According to the "Globalization and World Cities Research Network" (GAWC), a global organization dedicated to researching world cities and globalization, Chicago held the 19th position in the Global City Index as of 2020. Moreover, in 2018, it secured the world's highest gross domestic product (GDP) at $689 billion, establishing itself as the second-largest global financial center, trailing only behind New York.

Integral to Chicago's public transportation system is the iconic Chicago L(short for “elevated”) train, facilitating crucial connections between downtown Chicago and its suburbs. Its inception with the Green Line in 1892 has burgeoned into a network comprising 8 lines and 145 operational stations. Particularly renowned is the Loop line, encircling downtown Chicago, solidifying the L train's status as a signature mode of transportation.

The significance of public transportation, particularly for accessing downtown, where the majority of Chicago's industries converge, cannot be overstated. While it may bring certain inconveniences such as congestion and noise to residents in proximate areas, the prevailing sentiment leans towards its advantages. Public transportation substantially enhances local accessibility and streamlines commuting, making it an indispensable asset. Consequently, the accessibility to public transportation is anticipated to exert a positive influence on nearby housing prices.

This study aims to indirectly validate the efficacy of public transportation by examining the impact of Chicago's L train—a pivotal element in providing convenient access to downtown and connecting major attractions and regions—on housing values.

2. Previous research review

2.1 Literature review

The theory of land value and accessibility traces its origins to the Theory of isolated states (Thünen et al, 1826) and the Location of Land Use (Alonso, 1964). Both theories operate on the premise that the market is central, asserting that savings in transportation costs translate into rent. Consequently, they contend that locations closer to urban areas may incur higher land costs due to lower transportation expenses. Here, land cost refers to the surplus remaining after covering normal profits and production costs for the company. This implies that the more accessible a city is, the higher its land value. This elevated land value primarily stems from proximity to jobs and commercial facilities, coupled with reduced transportation costs, especially in locations near stations. Recent studies conducted worldwide support these theories.

Seo et al (2014) investigated the relationship between house prices in Phoenix, Arizona, and accessibility to highways and light rail using a hedonic price model. Results indicated a positive impact on housing prices from proximity to highway exits and light rail stations, with highways built underground having a relatively more positive effect on surrounding housing prices than above-ground highways.

Hess and Almeida (2007) employed a hedonic model to analyze the impact of station accessibility on housing prices for light rail in Buffalo. Homes within a one-mile radius of a light rail station exhibited a premium of 2-5% over the city's average home price. However, it was found that the physical characteristics of the house, such as the size of the house, whether it faces east or west, and the number of bathrooms, had more influence on the house price than accessibility to the light rail. There was a difference between the influence of each individual regression model for the 14 light rail stations analyzed, and the effect of proximity to the light rail showed a positive effect in places classified as high-income areas, while it showed a negative effect in places classified as low-income areas.

Dziauddin (2019) conducted an analysis of land values surrounding light rail stations in Greater Kuala Lumpur, Malaysia, utilizing geographically weighted regression (GWR). The results indicate a positive impact of light rail on residential real estate values, particularly in lower-middle and upper-middle income areas. However, this impact is not significant in high-income areas.

Tomal (2020) analyzed factors affecting rental prices for residential real estate in Cracow, Poland. Ordinary Least Squares (OLS), Spatial Autoregressive (SAR), and GWR were used for analysis. As a result of the analysis, all methods revealed that the physical characteristics of the house (total floor area, age, etc.) were the factors that had the greatest impact on the house price. In addition, the GWR-SAR model had the highest explanatory power, so it was concluded that GWR-SAR was the most appropriate model to study rent.

Tan et al. (2019) analyzed the effect of distance from subway stations on housing prices in the Wuhan, China. First, the results of hedonic analysis showed that the opening of subway stations had a greater impact on land prices in suburban areas than in the city center, and this influence was found to be 1600m for non-transfer stations and 4800m for transfer stations. In addition, as a result of analyzing the period before and after the opening of the subway using a difference-in-difference (DID) model, the treatment and control group showed similar trends in the period before the opening of the subway.

Yazdanifard et al. (2021) used the DID model to analyze the impact on residential real estate values in metro stations in Tehran, Iran. As a result of the analysis, the price of real estate closer to the station was higher, and the effect was greater in areas with less public transportation than in areas with good public transportation.

Han and Choi (2022) conducted an analysis of the relationship between housing and accessibility to light rail, employing a hedonic price model that focused on the Ui light rail in Seoul, Korea. The study distinguished between the east side, influenced by the existing catchment area (Line 4) other than the Ui light rail, and the west side, which had no such influence. The analysis revealed that housing prices in the west area increased the closer they were to the light rail station. However, in the east area, where the existing catchment area (Line 4) had an influence, the impact was different. It was found that the influence of the Line 4 was greater than Ui light rail, and housing prices increased the further away the light rail station was. In other words, the study confirmed that even within the same catchment area, the influence can manifest differently due to other factors.

McDonald and Osuji (1994) employed a hedonic model to assess the impact of the Midway Rapid Transit Line (Orange Line, opened in 1993) on the value of surrounding residential land, connecting Chicago's downtown (loop) and Midway Airport. Their analysis revealed that the value of residential areas within 0.5 miles from the station site was 17% higher than those located outside the station, indicating that the land market anticipated the opening of the Midway Line.

A subsequent study by McMillen and McDonald (2004) delved into the influence of the Midway Line on single-family home prices. Utilizing repeat sales and hedonic analysis with transaction data for homes within 1.5 miles of the Midway line, they found a negative correlation between the distance to the station and housing prices (The closer it is to the station, the higher the housing price). This negative correlation intensified as the opening of the Midway Line approached in 1990 and only slightly eased after its opening in 1996. The results suggested that public transportation had a significant impact on housing prices, with houses closer to the station and closer to the opening time experiencing a more pronounced effect.

In a more recent discussion, Wilcox and Micah (2023) explored the Chicago Transit Authority's (CTA) upgrade of its elevated trains. They argued that the CTA implicitly gentrified the "L" by allowing the market to unevenly distribute development around the stations. Despite being one of the few areas with direct government investment, the ridership on the Chicago Transit Authority's "L" elevated train system changed in ways reflecting gentrification-disinvestment patterns in Chicago's affluent North Side and the shift to less affluent southern regions. Multiple regression analysis supported this claim, revealing a positive correlation between the change in median income and the number of “L” riders for the city as a whole and for areas north of 0.5 miles from the nearest “L” station. The study concluded that, even as ridership declined on the south side, the north side experienced increased income, highlighting the "L" as an amenity facility where passenger growth paralleled income growth in the North.

The impact of urban railways on housing values has been a focal point for researchers and urban planners. Proximity to stations often correlates with higher home values due to the convenience of public transportation, a particularly crucial factor in diverse regions like Chicago. The concept of transit-oriented development (TOD), creating mixed-use developments near train stations to encourage walking, biking, and public transportation, has gained attention in urban planning. However, the impact of trains on home values can vary depending on the route or station characteristics. This complexity underscores the need for a comprehensive examination of the L train's impact on housing values, considering factors such as spatial heterogeneity, spatial dependence, and neighborhood characteristics. Thus, this study aims to examine the impact of the L train on housing values, considering these factors.

2.2 Differences from previous studies

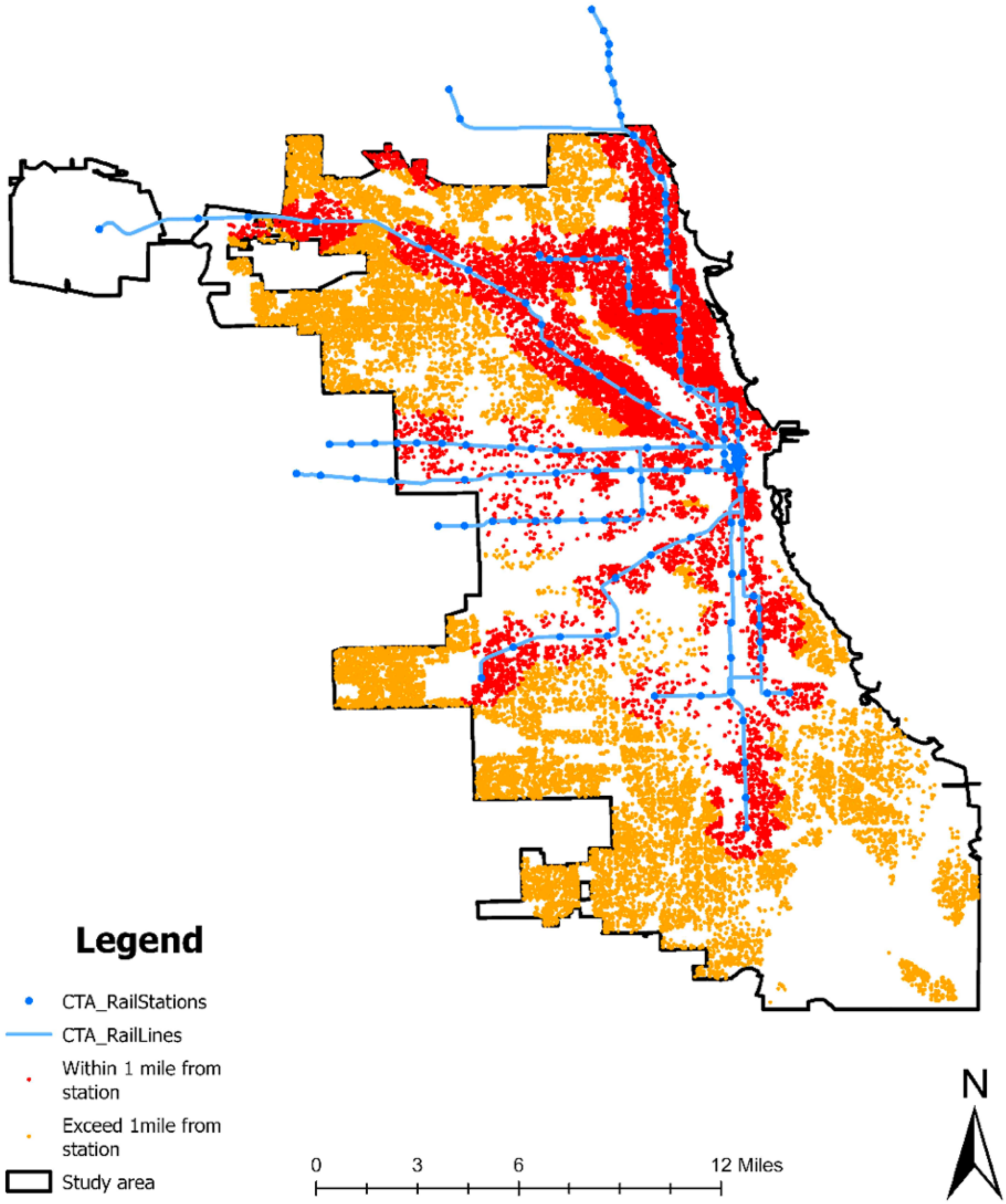
Most studies have utilized hedonic price models to capture the real estate value effects of rail transit, employing housing transaction prices or rental prices as dependent variables. (The hedonic price model is a type of multiple regression model, and multiple regression analysis is also called hedonic regression to emphasize the price estimation and interpretation method.) The hedonic model, a classical model in real estate economics, has faced criticism for its insensitivity to spatial heterogeneity and potential spatial dependence between variables (So, Tse, & Ganesan, 1997). Recent advancements in spatial analysis techniques have allowed researchers to consider spatial interactions, leading to the use of geographically weighted regression (GWR) to account for variations in the impact of rail transport across geographical areas. While some studies have employed both hedonic models and GWR to explain housing prices in other areas, such comprehensive analyses are rare for houses in the Chicago area. This study addresses this gap by additionally using GWR to examine housing prices in Chicago by region in more detail, systematically dividing the area into those within 1 mile of the station and those outside 1 mile.

3. Method description

3.1 Study area

Data from 47,914 houses traded in Chicago over the past three years(Nov.10.2020-Nov.09.2023) were collected via Zillow and analyzed by categorizing them into two groups: those located within 1 mile of the Chicago L train (27,887 cases) and those situated outside the 1-mile radius (20,027 cases). The distribution of real estate under scrutiny is illustrated in <Figure 1.>.

Figure 1. Distribution of housing transaction data in Chicago



3.2 Analysis techniques

The study categorized the variables for analysis into the distance from the station, housing characteristics, surrounding environment characteristics, crime, and population variables, aligning with the study's objectives. Data on infrastructure (parks, hospitals), geographic information, crime, and population were sourced from the Chicago Data Portal. It's noteworthy that the study utilized network distance, as opposed to Euclidean distance, for all distance variables. Unlike Euclidean Distance, which calculates straight-line distances between two points, network distance considers the shortest path, measured by the number of edges or the total weight of the edges between two nodes in the network. Despite the complexity of the network distance calculation method, it is considered to yield more accurate results, as indicated by previous studies (Mora & Marti, 2018; La Rosa, 2014). Therefore, this study employed network distance to measure the proximity to all facilities, with calculations executed through ArcGIS.

The initial phase of the research involved conducting descriptive statistical analysis to scrutinize transaction prices of the selected house samples and the distribution of major variables. Further investigation included independent sample t-tests and correlation analyses to ascertain price disparities contingent on distance from the station, housing characteristics, and surrounding environment features. Subsequently, the study applied sequential multiple regression analysis, Ordinary Least Squares (OLS), and Geographically Weighted Regression (GWR).

3.3 Composition of variables

The composition of variables was determined through a review of previous research, as outlined in <Table 1.> below. Among the variables, the presence of a parking lot and region(within or outside of 1 mile) were treated as dummy variables, while the ratio scale was used for all remaining variables. In the surrounding environment variables, middle school scores were excluded due to lack of data, with only elementary and high school scores being included.

Table 1. Variable type

Variable category	Variable name		Definition	Value type
Dependent variable	Transaction price (per square foot)		Actual transaction price divided by exclusive use area	$ (Dollar)
Independent variable	Distance to station		Distance to the nearest Elevated Train station (mile)	Ratio scale
	Housing characteristics	Area for exclusive use	Area for exclusive use of the target housing (square feet)	Ratio scale
		Number of beds	Number of beds	Ratio scale
		Number of baths	Number of baths	Ratio scale
		Built year	Year the house was built	Ratio scale
		Parking availability	Divided into no parking (0) and parking available (1) (ref.no parking)	(0, 1)
	Surrounding environment characteristics	Primary school rating	Based on research conducted by GreatSchools.org, an independent non-profit organization, It is based on four levels of the School Assessment Scale: Student Progress Rating or Academic Progress Rating, College Readiness Rating, Equity Rating, and Test Score Rating.	Ratio scale
		High school rating		Ratio scale
		Distance to park	Distance to the nearest park (mile)	Ratio scale
		Distance to general hospital	Distance to the nearest general hospital (mile)	Ratio scale
	Crime frequency		Indicates the number of crimes (only violent or more serious crimes) from 2001 to 2023 for the zip code belonging to the property.	Ratio scale
	Population density		Average 2020 population of the zip code in which the property belongs	Ratio scale
	Region		Within1 (0) & exceed1 (1) mile of station dummy (ref. within1 mile)	(0, 1)

4. Basic status analysis

4.1 Distribution of housing prices by Chicago L station

After calculating the transaction price per square foot for houses in Chicago, average prices were compared based on the nearest station to each home. As illustrated in <Figure 1.>, out of a total of 145 stations, those not belonging to Chicago were excluded, resulting in a total of 128 stations. The average housing price per square foot for these 128 stations is provided in Appendix 1. Examining each train line, the average housing price was higher at stations where multiple lines passed, with most of these stations located in downtown Chicago (Chicago Loop). When focusing on individual lines, the Brown Line, which has most of its stations downtown, exhibited the highest price per square foot at $288.66, while the Green Line (Englewood) had the lowest price at $161.16. On a station-by-station basis, downtown stations recorded higher prices, with State/Lake's price being the highest at $443.14. Conversely, housing prices at stations outside of Chicago were lower, particularly on the north side of Chicago, with the Garfield-Dan Ryan station having the lowest at $122.9.

4.2 Descriptive statistical analysis

The study's variables are detailed in <Table 2.>. The dependent variable, transaction price per square foot, averaged $253.1, with a minimum of $61 and a maximum of $975. The distance from the station ranged from 0.0031 to 8.7 miles, averaging 1.38 miles.

Turning to housing characteristics, the housing area exhibited a range of 80 to 17,769 square feet, averaging at 1732.6 square feet. The number of rooms varied from 0 to 14, with an average of 3, and the average number of bathrooms was 2.27. Construction years spanned from 1800 to 2024, with an average of 1960. Parking availability was denoted as a dummy variable with presence (1) and absence (0), and the average was 0.86, signifying that the majority of houses had parking available.

In consideration of surrounding environmental variables, ratings for primary and high schools ranged from 0 to 10, with primary schools averaging 5 points and high schools averaging 2.5 points. Distances to the nearest park ranged from 0 to 2.2 miles, averaging 0.36 miles, while distances to the nearest hospital ranged from 0.007 to 7.6 miles, averaging 1.64 miles. Crime frequency ranged from at least 69 to 16,616, averaging 5032.4, and the population ranged from 897 to 114,453, averaging 56,255.8.

Table 2. Descriptive statistics of key variables

Variable category	Variable name		Unit	Minimum	Maximum	Mean	Standard deviation	Median
Dependent variable	Transaction price (per square foot)		$(Dollar)	61	975	253.1	103.7	242.1
Independent variable	Distance to station		Mile	0.0031	8.7	1.38	1.38	0.78
	Housing characteristics	Area for exclusive use	Square feet	80	17,769	1,732.6	959	1,450
		Number of beds	Number of beds	0	14	3	1.26	3
		Number of baths	Number of baths	0	27	2.27	1.03	2
		Built year	Year	1800	2024	1960.1	37.7	1,958
		Parking availability	Dummy variable	0	1	0.86	0.35	1
	Surrounding environment characteristics	Primary school rating	Rating	0	10	5	2.33	5
		High school rating		1	10	2.5	2.17	1
		Distance to park	Mile	0	2.2	0.36	0.212	0.326
		Distance to general hospital	Mile	0.007	7.6	1.64	1.26	1.29
	Crime frequency		Number of crimes	69	16,616	5,032.40	4,021.70	3,579
	Population density		Number of populations	897	114,456	56,255.8	21,298.10	54,064
	Region		Dummy variable					
Number of data				**47,914**				

4.3 Price difference according to catchment area

The influence of the station is presumed to be significant on houses located in close proximity. While the definition of a catchment area varies across studies, it is generally considered to be within a maximum radius of one mile. To investigate the price disparity based on this 1-mile distance from the station, an independent samples t-test was conducted.

The results of the analysis are presented in <Table 3.>. The t-test revealed a significant difference in the variance between the two datasets. Moreover, it was observed that the average price per square foot for homes within 1 mile of the station was $88.87 higher. This finding suggests that the pricing characteristics of houses within 1 mile of the station differ from those located 1 mile away.

Table 3. Price difference due to the influence of station

		N	Average	Standard deviation	t	p	Average difference
Division of area	**Within 1 mile**	27,887	290.27	106.2	108	0.000	88.87
	Exceed 1 mile	20,027	201.4	73.77			

5. Empirical analysis

5.1 Correlation between major variables

Before delving into causal analysis, a correlation analysis was performed to assess the relationships between the explanatory variables and the transaction price per square foot, the dependent variable. The analysis findings, detailed in <Table 4.>, indicate significant correlations between all independent variables and the transaction price per square foot.

Examining the specifics, the distance from the station demonstrated a negative correlation with the transaction price per square foot ($r = -0.374$, $p < 0.000$), implying that the closer the distance to the station is, the higher the housing price.

Concerning housing characteristics, all variables, except the number of beds, exhibited a positive correlation. Notably, the year of construction displayed the most substantial correlation with the transaction price ($r = 0.312$, $p < 0.000$).

In terms of surrounding environment characteristics, both primary and high school ratings exhibited positive correlations, while distances to parks and hospitals displayed negative correlations. Higher school ratings and closer proximity to facilities were associated with elevated housing prices. Among these, Primary school rating showed the highest correlation ($r= 0.307$, $p < 0.000$).

Crime variables and population factors displayed negative correlations, indicating that lower crime frequencies and population densities were linked to higher housing prices. Additionally, regional variables were negatively correlated with housing prices ($r = -0.447$, $p < 0.000$), signifying that homes within 1 mile of a station were more expensive compared to those situated 1 mile away.

The correlation analysis yielded results consistent with prior studies, with the exception of the number of bedrooms. The negative correlation observed in the number of bedrooms and housing prices can be attributed to the generally smaller size and fewer rooms in houses within 1 mile of the station compared to those located 1 mile away.

Table 4. Correlation between key variable

Variable category	Variable name		Correlation coefficient	p
Dependent variable	Transaction price (per square feet)		1	0.000
Independent variable	Distance to station		-0.374	0.000
	Housing characteristics	Area for exclusive use	0.069	0.000
		Number of beds	-0.160	0.000
		Number of baths	0.144	0.000
		Built year	0.312	0.000
		Parking availability	0.073	0.000
	Surrounding environment characteristics	Primary school rating	0.307	0.000
		High school rating	0.248	0.000
		Distance to park	-0.196	0.000
		Distance to general hospital	-0.269	0.000
	Crime frequency		-0.361	0.000
	Population density		-0.080	0.000
	Region		-0.447	0.000

- Dummy variables (Parking availability, Region): Spearman correlation coefficient
- All except dummy variables: Pearson correlation coefficient

5.2 Multiple regression analysis

5.2.1 Analysis of factors influencing overall housing prices in Chicago

In the examination of factors influencing housing prices in Chicago, various variables were considered, including distance from the station and housing characteristics such as area, number of bedrooms, number of bathrooms, year of completion, and parking availability. Subsequently, for hierarchical regression analysis, additional factors such as surrounding environmental elements, crime frequency, population, and regional variables were incrementally introduced. The regression equation provided below was employed, and the analysis results are presented in <Table 5.>.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + \epsilon$$

$X_1 - X_n$: $Distance\ to\ station\ variable, Housing\ characteristic\ variable,$
$Surrounding\ environment\ characteristics\ varaiable,$
$Crime\ frequency\ variabl, Population\ varaible, Region\ variable$

Table 5. Analysis result of factors affecting housing prices in Chicago

		Basic model			Environment model			Crime model			Population model			Region model		
		B	β	VIF	B	β	VIF	B	β	VIF	B	β	VIF	B	β	VIF
Distance to station(sqft)		-21.481***	-0.288	1.220	-17.732***	-0.215	1.620	-19.601***	-0.238	1.628	-19.555***	-0.238	1.628	-11.373***	-0.138	2.870
Housing characteristics	**Area for exclusive use**	-0.010***	-0.091	3.151	-0.016***	-0.144	3.542	-0.022***	-0.198	3.589	-0.022***	-0.199	3.591	-0.023***	-0.208	3.601
	Number of beds	-17.198***	-0.208	2.315	-11.964***	-0.142	2.590	-2.433***	-0.029	2.791	-1.964***	-0.023	2.864	-0.152	-0.002	2.922
	Number of baths	31.179***	0.310	3.379	31.218***	0.301	3.742	30.255***	0.292	3.744	30.171***	0.291	3.745	29.524***	0.285	3.750
	Built year	0.513***	0.186	1.178	0.489***	0.177	1.197	0.396***	0.143	1.214	0.394***	0.142	1.216	0.38***	0.137	1.219
	Parking availability (ref.no parking)	15.145***	0.050	1.117	16.728***	0.054	1.131	19.314***	0.062	1.132	19.414***	0.063	1.132	20.173***	0.065	1.133
Surrounding environment characteristics	**Primary school rating**				10.679***	0.231	1.187	5.989***	0.130	1.350	6.275***	0.136	1.441	6.108***	0.132	1.442
	High school rating				3.542***	0.071	1.221	3.446***	0.069	1.221	3.732***	0.075	1.300	3.000***	0.060	1.327
	Distance to park				-30.075***	-0.056	1.208	-26.896***	-0.050	1.209	-25.276***	-0.047	1.231	-19.019***	-0.036	1.248
	Distance to general hospital				-8.59***	-0.098	1.387	-8.089***	-0.092	1.388	-8.325***	-0.095	1.405	-7.535***	-0.086	1.416
Crime frequency								-0.008***	-0.295	1.368	-0.008***	-0.283	1.666	-0.007***	-0.271	1.684
Population density											0.0001***	-0.025	1.480	0.0001***	-0.018	1.487
Region (ref.=within 1 mile)														-35.016***	-0.156	3.069
Constant		-736.9***			-737.871***			-506.431***			-499.808***			-477.902***		
N		47914														
F		2306***			1900***			2270***			2080***			1990***		
R^2		0.243			0.338			0.401			0.402			0.410		
Adjusted R^2		0.243			0.338			0.401			0.401			0.409		

*- Dependent variable: Transaction price per square feet ($ Dollar) - *p<0.05, **p<0.01, ***p<0.001*

First, in the basic model, the influence on housing transaction prices was analyzed using a regression analysis model that considered only the distance to the station and housing characteristic variables. The resulting R^2value, indicating the model's explanatory power, was 0.243. Notably, all variables in this model exhibited significant effects. Specifically, distance to the station (-), smaller area (-), fewer bedrooms (-), more bathrooms (+), more recent construction year (+), and parking lots availability (+) were associated with higher transaction prices for houses.

Subsequently, the model was expanded to include variables related to surrounding environmental characteristics. This adjustment significantly increased the R^2value to 0.338, highlighting the meaningful contribution of environmental variables. Higher scores for primary and high schools (+) and closer proximity to parks and hospitals (-) were associated with higher housing transaction prices.

Further analysis incorporating crime frequency as a variable revealed a substantial increase in the model's explanatory power (R^2= 0.401). This indicated that crime frequency was a potent explanatory factor, with lower crime rates (-) correlating with higher housing prices.

Upon introducing population density, the model's R^2value remained unchanged, suggesting no significant impact compared to the model excluding population considerations.

Lastly, an analysis including dummy variables within 1 mile and 1 mile outside of the station demonstrated a modest increase in the R^2value by 0.401 to 0.409. Notably, the 1-mile dummy variable with the station displayed a negative correlation, indicating higher housing prices within 1 mile from the station, and was found to have the most significant impact among the variables. Considering the change in the R^2value, the significance of the change in F, and the significance of sequentially input variables, it can be concluded that region model, which represents the final step, is the most suitable for predicting housing prices. Furthermore, with VIF values consistently below 5 in all models, no multicollinearity issues are apparent.

Nevertheless, the outcomes indicating that smaller house areas and fewer rooms are associated with higher prices contradict prior research. This inconsistency is likely attributed to the concentration of small 1-bed or studio-type homes in downtown Chicago, where property values are inherently elevated. As a result, the next section will aim to thoroughly examine these phenomena and evaluate the influence of the catchment area. This will entail dividing the sample into areas within 1 mile of the station and 1 mile outside and conducting regression analysis to gain a more profound understanding of the factors influencing home prices in each specific area.

5.2.2 Analysis of factors influencing housing prices within 1 mile of Chicago L Train Station

A significant proportion of homes situated within 1 mile of the Chicago L train station are concentrated downtown, in contrast to those located a mile away. However, certain models revealed a VIF value exceeding 5, indicating multicollinearity. Consequently, a reanalysis was undertaken, excluding the housing area for exclusive use variable due to its high VIF value and relatively limited influence. The results of this analysis are presented in <Table 6.> (See Appendix 2 for the analysis before exclusion). Except for the distance variable from the park, all other factors exhibited significant effects. The VIF values for all variables were less than 5, suggesting the absence of multicollinearity issues.

Across all basic models, distance to the station (-) correlated with higher housing prices, while variables related to housing characteristics, such as fewer bedrooms (-), more bathrooms (+), more recent completion (+), and availability of parking lot (+), were associated with higher prices. In the surrounding environmental model, the R^2 value increased by 0.072, indicating that the inclusion of surrounding environment variables was meaningful. There was no substantial difference in impact compared to the basic model.

For the crime model, the R^2 value increased by 0.072, demonstrating the meaningful contribution of the crime variable. Notably, there was no significant deviation from the surrounding environment model, except for a rise in the p-value for the distance to the park. In the population model, the R^2 value remained constant, revealing that the population variable input had no discernible effect, and the distance variable to the park was not significant.

In summary, homes within 1 mile of the Chicago L Train station exhibit higher prices when closer to the station, with the distance to the station variable having the most significant impact among all variables. Moreover, coefficient values confirm that housing characteristic variables, such as the number of bedrooms, bathrooms, and parking availability, exert a more substantial influence on housing prices compared to overall surrounding environmental characteristics.

Table 6. Analysis of factors affecting housing prices within 1mile of L train in Chicago

		Basic model			Environment model			Crime model			Population model		
		B	β	VIF	B	β	VIF	B	β	VIF	B	β	VIF
Distance to station(sqft)		-64.745**	-0.144	1.043	-49.963***	-0.109	1.069	-26.008***	-0.057	1.106	-24.014***	-0.052	1.157
Housing Characteristics	**Number of beds**	-23.106***	-0.271	2.627	-17.266***	-0.201	2.992	-5.566***	-0.065	3.248	-5.147***	-0.060	3.313
	Number of baths	34.560***	0.369	2.616	28.717***	0.306	2.967	21.163***	0.225	3.056	21.024***	0.224	3.062
	Built year	0.500***	0.196	1.146	0.478***	0.188	1.155	0.406***	0.160	1.166	0.404***	0.159	1.167
	Parking availability (ref.no parking)	14.126***	0.048	1.154	15.723***	0.054	1.162	18.195***	0.063	1.163	18.031***	0.062	1.164
Surrounding environment characteristics	**Primary school rating**				8.403***	0.193	1.150	4.308***	0.099	1.271	4.608***	0.106	1.400
	High school rating				2.179***	0.049	1.196	1.711***	0.039	1.197	2.065***	0.047	1.370
	Distance to park				-14.688***	-0.021	1.051	-8.230*	-0.012	1.053	-7.421	-0.011	1.056
	Distance to general hospital				-21.96***	-0.139	1.181	-9.127***	-0.058	1.272	-9.338***	-0.059	1.277
Crime frequency								-0.012***	-0.336	1.554	-0.012***	-0.325	1.893
Population density											-0.0001***	-0.025	1.760
Constant		-689.912***			-677.256***			-497.093***			-492.285		
N		27,887											
F		959***			792***			1030***			942		
R^2		0.159			0.231			0.304			0.304		
Adjusted R^2		0.159			0.231			0.303			0.304		

*- Dependent variable: Transaction price per square feet ($ Dollar)   - *p<0.05, **p<0.01, ***p<0.001*

5.2.3Analysis of factors influencing housing prices 1 mile outside Chicago L Train Station

Houses situated 1 mile outside the Chicago L Train station are in areas less likely to be directly influenced by the station compared to those within a 1-mile radius. The analysis results are presented in <Table 7.>. In the basic model, all variables, except the number of bedrooms, demonstrated a significant effect. Moreover, both the crime frequency model and the population model exhibited significant results for all variables, except for the high school rating variable. Overall, the findings closely resembled those of the model in <Table 5.>, which analyzed the entire Chicago area, with the exceptions of the high school rating variable, number of beds variable, and population density variable.

For houses located 1 mile away from the L train, an increase in the number of beds corresponded to higher housing prices. Although the population variable showed a positive correlation, the magnitude of the correlation was minimal. Similar to the analysis of the overall area in <Table 5.>, it was reaffirmed that, in the residential area outside the L train, the primary variables influencing housing prices were housing characteristics. Notably, unlike the analysis model within 1 mile of the L train <Table 6.>, it was observed that the impact of distance from the station was diminished.

Table 7. Analysis of factors affecting housing prices 1 mile outside Chicago's L train

		Basic model			Environment model			Crime model			Population model		
		B	β	VIF	B	β	VIF	B	β	VIF	B	β	VIF
Distance to station(sqft)		-7.994***	-0.147	1.032	-8.200***	-0.147	1.122	-11.253***	-0.202	1.170	-10.141***	-0.182	1.195
Housing characteristics	**Area for exclusive use**	-0.040***	-0.430	2.000	-0.053***	-0.510	1.890	-0.055***	-0.528	1.895	-0.054***	-0.518	1.901
	Number of beds	-0.464	-0.007	1.637	4.530***	0.070	1.675	6.825***	0.105	1.694	5.754***	0.089	1.712
	Number of baths	27.353***	0.324	2.124	26.197***	0.298	1.991	25.892***	0.295	1.991	26.178***	0.298	1.992
	Built year	0.266***	0.104	1.118	0.360***	0.145	1.174	0.256***	0.103	1.202	0.250***	0.101	1.202
	Parking availability (ref.no parking)	22.428***	0.091	1.043	24.121***	0.098	1.047	25.156***	0.102	1.048	22.588***	0.092	1.055
Surrounding environment characteristics	**Primary school rating**				12.246***	0.331	1.207	7.204***	0.195	1.500	6.167***	0.167	1.550
	High school rating				-1.641***	-0.029	1.075	0.104	0.002	1.090	-0.325	-0.006	1.093
	Distance to park				-14.974***	-0.047	1.055	-15.467***	-0.048	1.055	-21.42***	-0.067	1.077
	Distance to general hospital				-3.522***	-0.077	1.242	-3.681***	-0.080	1.243	-2.295***	-0.050	1.301
Crime frequency								-0.004***	-0.302	1.436	-0.005***	-0.363	1.678
Population density											0.001***	0.146	1.352
Constant		-308.891***			-531.698***			-277.436***			-288.847***		
N		20,027											
F		513***			521***			638***			627***		
R^2		0.141			0.279			0.342			0.358		
Adjusted R^2		0.141			0.279			0.342			0.358		

*- Dependent variable: Transaction price per square feet ($ Dollar) - *p<0.05, **p<0.01, ***p<0.001*

5.3 Spatial Autocorrelation

As the hedonic model does not account for spatial autocorrelation, initial step involves employing the Global Moran's I to assess the spatial correlation of housing prices in Chicago. Moran's I index is a statistical analysis technique designed to identify spatial dependence and ascertain the presence of spatial correlation in specific regions. The formula for Moran's I is as follows:

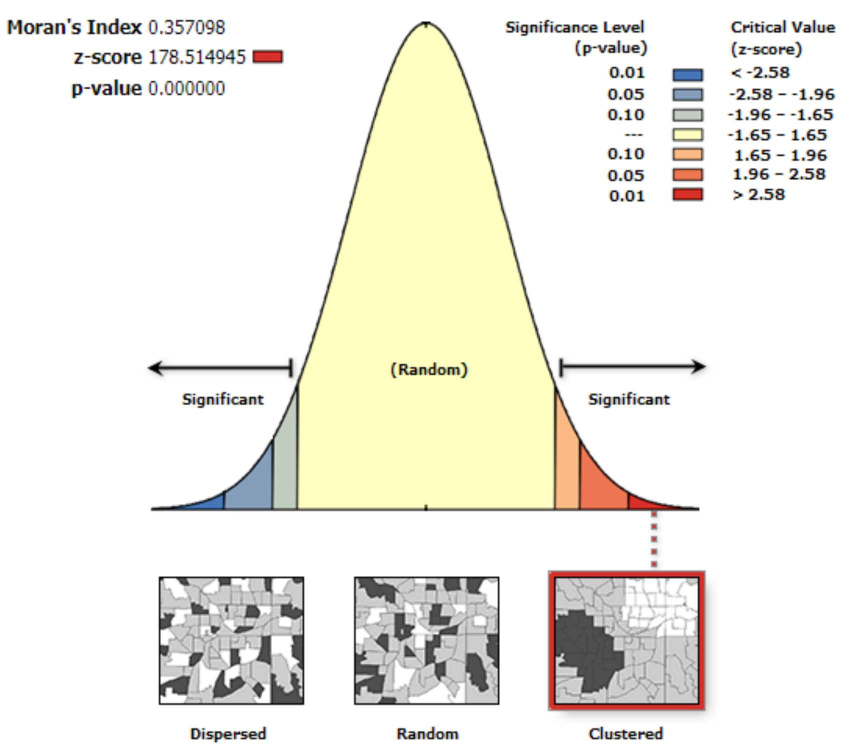
$$I = \frac{n}{\sum_i^n \sum_j^n w_{ij}} \frac{\sum_i^n \sum_j^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_j - \bar{x})}$$

$$I_i = \frac{x_i - \bar{X}}{m_2} \sum_j w_{ij} (x_j - \bar{X}) \qquad m_2 = \frac{\sum_i (x_j - \bar{X})^2}{n}$$

n is the total number of observations (spatial objsects).
x_i is the attribute value of feature *i*
x_j is the attribute value of feature j
X is the mean of this attribute

The outcomes of the Moran's I index measurement are presented in <Figure 2.>. Specifically, the Moran's I index yields a positive value of 0.357. This indicates a confirmed spatial correlation among housing prices in Chicago. It is crucial to note that the global Moran's I index solely gauges spatial correlation and does not discern local effects. Consequently, in the subsequent session will undertake Geographically Weighted Regression (GWR) analysis. This analytical approach will enable us to identify and examine the localized impacts on housing prices, providing a more nuanced understanding beyond the global spatial correlation measured by Moran's I index.

Figure 2. Global Moran's I index of Chicago housing



Given the z-score of 178.514945, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

5.4 Geographically Weighted Regression Analysis (GWR)

Hedonic pricing models have historically served as a robust tool for real estate valuation, endorsed by numerous researchers. However, their applicability to spatially heterogeneous data has been questioned. This arises from the assumption inherent in the hedonic price model that the geographic area remains uniform in the relationship between dependent and independent variables. Acknowledging this limitation, this study aims to enhance analysis by incorporating Geographically Weighted Regression (GWR) based on the hedonic price model.

To establish a foundation for GWR analysis, Ordinary Least Squares (OLS) analysis was conducted, yielding the correct model for GWR. The detailed results of this OLS analysis can be found in Appendix 3.

Following the methodology employed in the earlier hedonic model analysis, GWR was applied to the entire area and subsequently subdivided into regions within 1 mile of the station and those beyond 1 mile. The formula for GWR is outlined as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k}^{p} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i$$

(u_i, v_i) are the geographical coordinates.

This approach enables us to account for spatial variations and investigate localized impacts on housing prices, providing a more nuanced understanding compared to the traditional hedonic model.

5.4.1 Analysis of factors influencing overall housing prices in Chicago

To address multicollinearity, several variables—region dummy (within 1 mile or outside), housing area, primary school and high school rating, crime frequency, and population density—were excluded from the overall model. The analysis results are presented in <Table 8.>. Despite the removal of several variables due to multicollinearity in comparison to the hedonic model, the model's explanatory power(R^2) significantly increased to 0.629. However, the AICc also rose, suggesting a decrease in model fit.

Upon closer inspection of the variables, akin to the hedonic model, the surrounding environment variable demonstrated the highest explanatory power. Notably, a key deviation from the hedonic model is that, in this context, the farther the distance from the station and the greater the number of beds, the higher the housing price. This divergence highlights the nuanced influence of these factors on overall housing prices in Chicago, underscoring the importance of accounting for specific regional dynamics in the analysis.

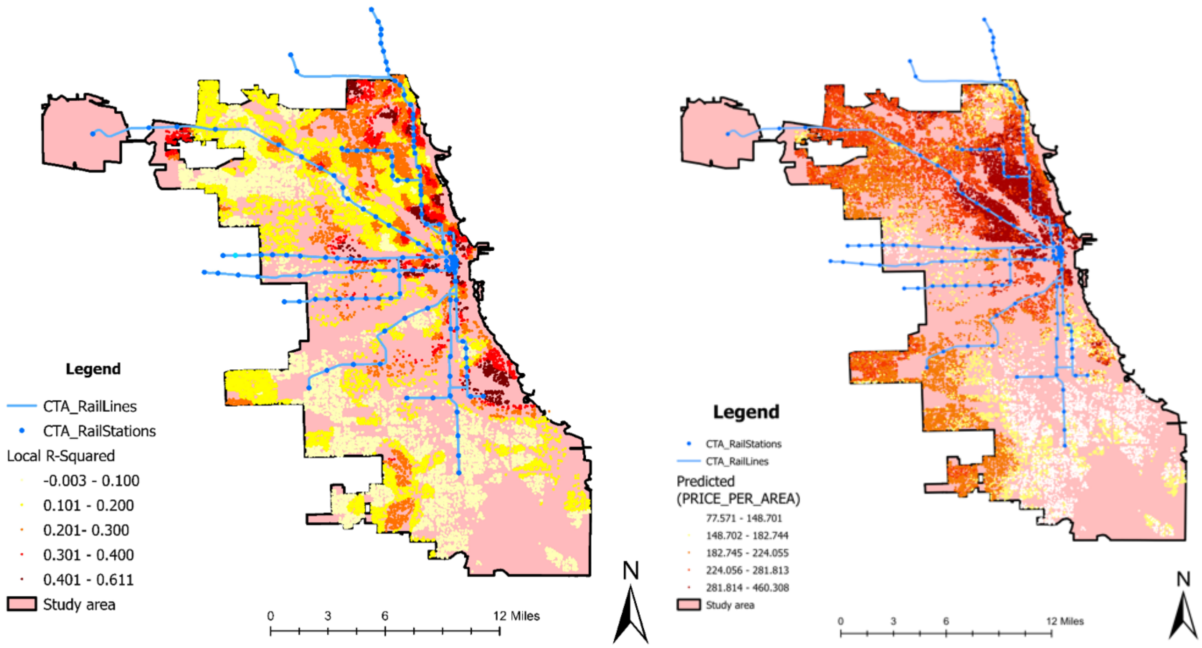
Table 8. Analysis of factors affecting Chicago housing prices(Comparison of Regression analysis and GWR)

		Multiple Regression			Geographically Weighted Regression		
					Coefficient(β)		
		B	β	VIF	Min	Max	Mean
Distance to station(sqft)		-11.373***	-0.138	2.870	-838.2	737.5	6.1
Housing characteristics	**Area for exclusive use**	-0.023***	-0.208	3.601			
	Number of beds	-0.152	-0.002	2.922	-32.4	33	2.7
	Number of baths	29.524***	0.285	3.750	-44.6	79	3.4
	Built year	0.38***	0.137	1.219	-1.24	7	0.137
	Parking availability (ref.no parking)	20.173***	0.065	1.133	-103.1	67.5	9.6
Surrounding environment characteristics	**Primary school rating**	6.108***	0.132	1.442			
	High school rating	3.000***	0.060	1.327			
	Distance to park	-19.019***	-0.036	1.248	-666.5	1284.5	-1.46
	Distance to general hospital	-7.535***	-0.086	1.416	-864	476.2	-3.7
Crime frequency		-0.007***	-0.271	1.684			
Population density		0.0001***	-0.018	1.487			
Within 1 & exceed 1 dummy (ref.=within 1 mile)		-35.016***	-0.156	3.069			
Constant		-477.902***			-13834.2	2627.4	-44.6
N		47,914			44,109		
AICc		434,299			492,361.949		
R^2		0.410			0.646		
Adjusted R^2		0.409			0.629		

*- Dependent variable: Transaction price per square feet ($ Dollar) - *p<0.05, **p<0.01, ***p<0.001*

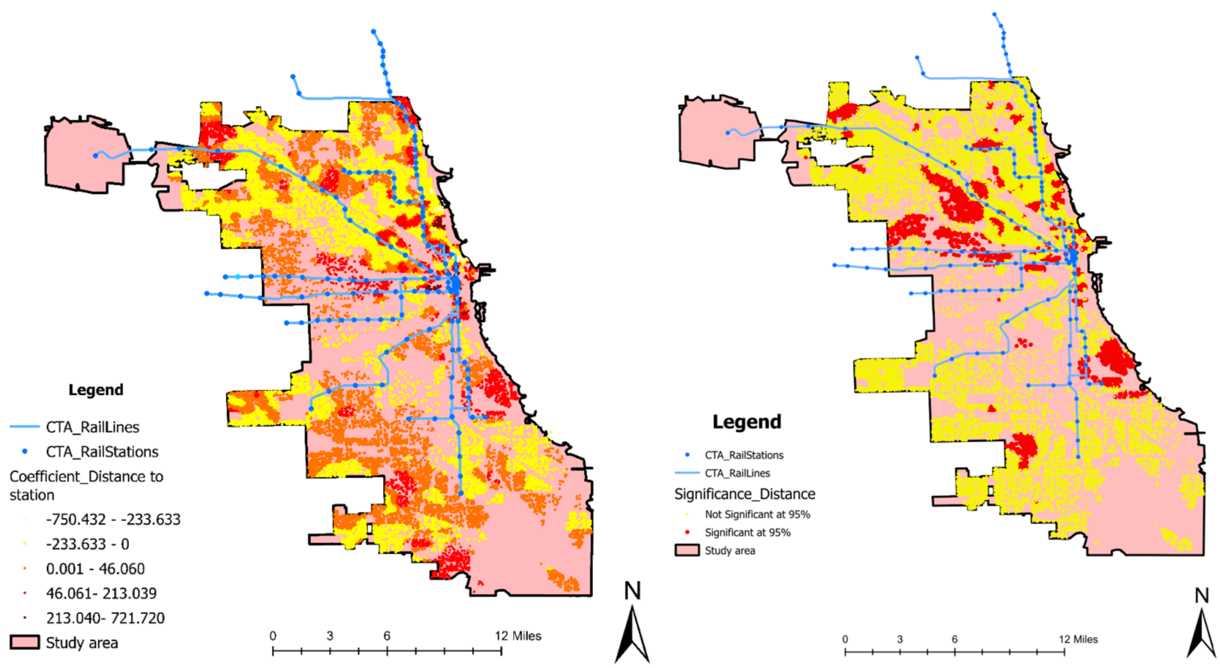
Looking into the details through the map, the information is illustrated in <Figure 3.>. From the left, utilizing the R^2map, The explanatory power of the regression model was observed to be higher in the eastern region near the river and in certain areas near the airport. Conversely, regions with low R^2values, particularly in the west, indicate that housing prices cannot be fully predicted by the explanatory variables. Regarding predictions, it becomes apparent that housing prices are elevated in areas closer to the station, particularly in downtown and the northern regions of Chicago.

Figure 3. R-square and prediction throughout Chicago



Moving forward, let's examine the impact of the distance variable from the station. The results are depicted in <Figure 4.> below.

Figure 4. Local coefficient estimates and statistical significance of estimates for "distance to station".



Upon scrutinizing the coefficient for the relationship between housing prices and distance to the station, it became evident that the impact of distance varied unevenly across the entire region. The findings revealed that, in many areas, particularly north of Chicago, housing prices increased as the distance to the station decreased. However, in certain regions near downtown, the opposite trend was observed, with housing prices rising as the distance from the station increased. The statistical significance of this relationship was also noted to differ by region.

Contrary to the conclusion drawn by the hedonic model, which suggested higher housing prices closer to the station, the GWR analysis produced a contrasting result. According to the GWR analysis, housing prices were higher farther away from the station. This disparity underscores the complexity of regional influences and their significance. The GWR model, while more intricate, provides compelling statistical results by suggesting that the impact and significance of distance vary for each region.

In the upcoming session, to delve deeper into this influence, the sample is divided into areas within 1 mile and beyond the station, followed by a GWR analysis.

5.4.2 Analysis of factors influencing housing prices within 1 mile of Chicago L Train Station

In the model focusing on areas within 1 mile of the Chicago L Train station, the explanatory power(R^2) saw a significant increase to 0.596. However, the AICc also rose, indicating a decrease in the model fit. Examining the variables in <Table 9.>, akin to the hedonic model, a higher number of baths, recent construction, parking lot availability, proximity to hospitals were associated with higher housing prices. In contrast, a different result from the hedonic model emerged, indicating that the farther the distance from the station and the greater the number of rooms, the higher the housing price.

Table. 9. Analysis of factors affecting housing prices within 1mile of L train in Chicago (Comparison of Regression analysis and GWR)

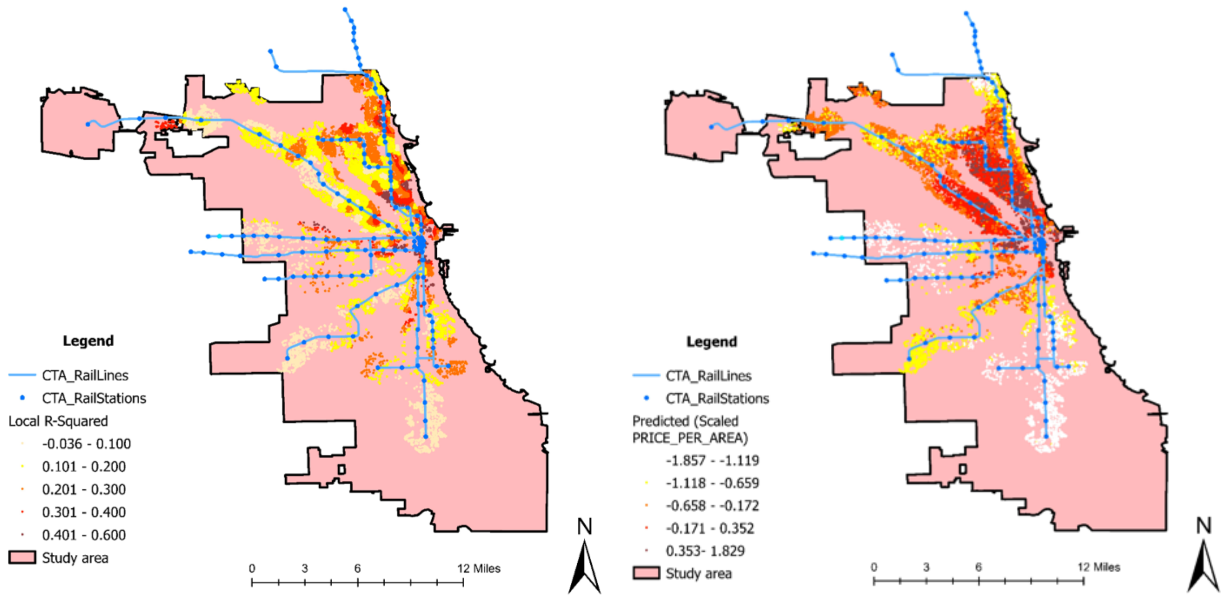
		Multiple Regression Analysis			Geographically weighted regression		
					Coefficient(β)		
		B	**β**	**VIF**	**Min**	**Max**	**Mean**
Distance to station(sqft)		-24.014***	-0.052	1.157	-905.2	608.2	5.5
Housing Characteristics	**Number of beds**	-5.147***	-0.060	3.313	-34.5	32.6	5.1
	Number of baths	21.024***	0.224	3.062	-47.5	79.4	6.7
	Built year	0.404***	0.159	1.167	-0.85	6.7	0.276
	Parking availability (ref.no parking)	18.031***	0.062	1.164	-82.9	64.2	9.1
Surrounding environment characteristics	**Primary school rating**	4.608***	0.106	1.400			
	High school rating	2.065***	0.047	1.370			
	Distance to park	-7.421	-0.011	1.056			
	Distance to general hospital	-9.338***	-0.059	1.277	-710.2	519.6	-5.3
Crime frequency		-0.012***	-0.325	1.893			
Population density		0.0001***	-0.025	1.760			
Constant		-492.285			-13185.4	2232	-283
N		27,887			25,398		
AICc		280,510			286,855.242		
R^2		0.304			0.615		
Adjusted R^2		0.304			0.596		

*- Dependent variable: Transaction price per square feet ($ Dollar) - *p<0.05, **p<0.01, ***p<0.001*

Looking into the details through the map, as illustrated in <Figure 5.>, a higher explanatory power of the regression model, based on the R^2map, is discerned in the northern area centered on downtown and some regions near the airport.

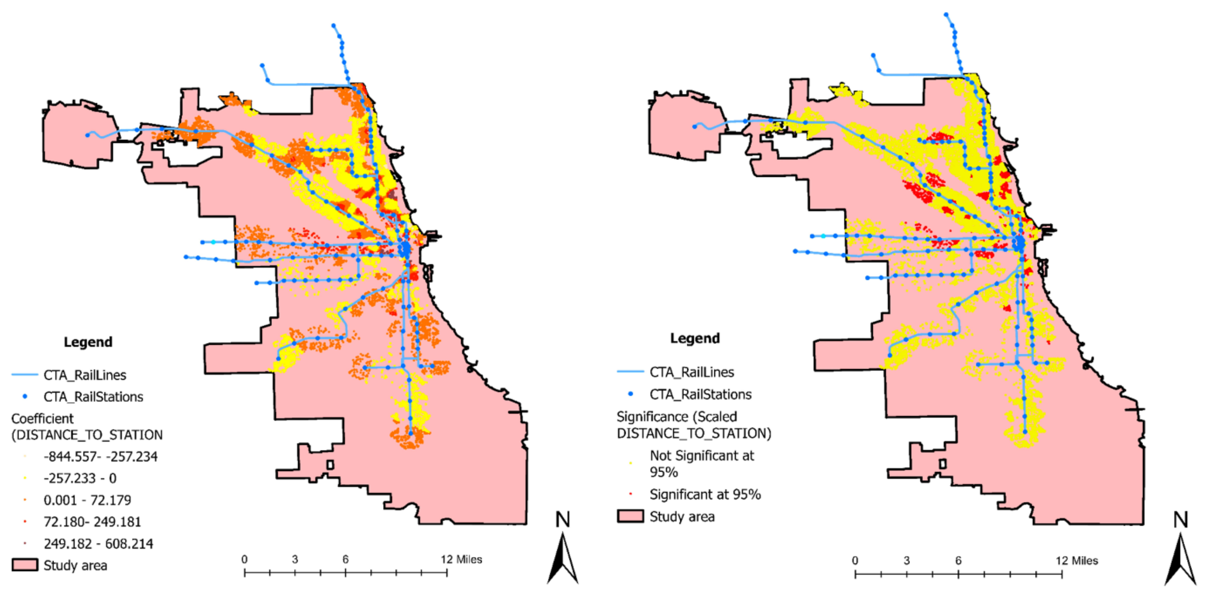
Furthermore, as one moves away from downtown, fewer locations in the northern region exhibit lower R^2values. However, in the western and southern regions, R^2decreases with the increasing distance from downtown, resulting in reduced explanatory power. In essence, except for the northern area, housing prices in regions distant from downtown cannot be entirely predicted using explanatory variables. Regarding predictions, similar to the overall regional analysis, it becomes evident that housing prices are higher in areas closer to the station within the downtown and northern areas centered on downtown Chicago.

Figure 5. R-square and prediction within 1mile of L train in Chicago



Moving forward, let's delve into the impact of the distance variable from the station. The results are illustrated in <Figure 6,> below. Upon scrutinizing the coefficient for the relationship between housing prices and distance to the station, it becomes evident that the influence of distance varies unevenly across the entire region. In the overall analysis and across numerous areas in the north, the closer the distance to the station, the higher the housing price. Moreover, statistical significance is confirmed in certain areas, particularly those centered around downtown and the northern region.

Figure 6. Local coefficient estimates and statistical significance of the estimates for "distance to station" within 1 mile of the Chicago L Train.



5.4.3 Analysis of factors influencing housing prices 1 mile outside Chicago L Train station

The analysis results are presented in <Table 10.>. For the model outside 1 mile of the Chicago L Train, the explanatory power (R^2) increased, akin to both the overall model and the model within 1 mile. However, the increase in R^2 was modest, rising only 0.041. Consequently, there was no significant difference compared to the hedonic model, even though AICc increased notably. This suggests that GWR might not be the most suitable analysis for explaining housing prices when compared to the hedonic model.

Examining the variables, the results were contrary to the hedonic model, except for the parking availability variable and the distance to the hospital. These findings indicate that GWR does not offer improved explanatory power for houses located 1 mile outside the Chicago L train station.

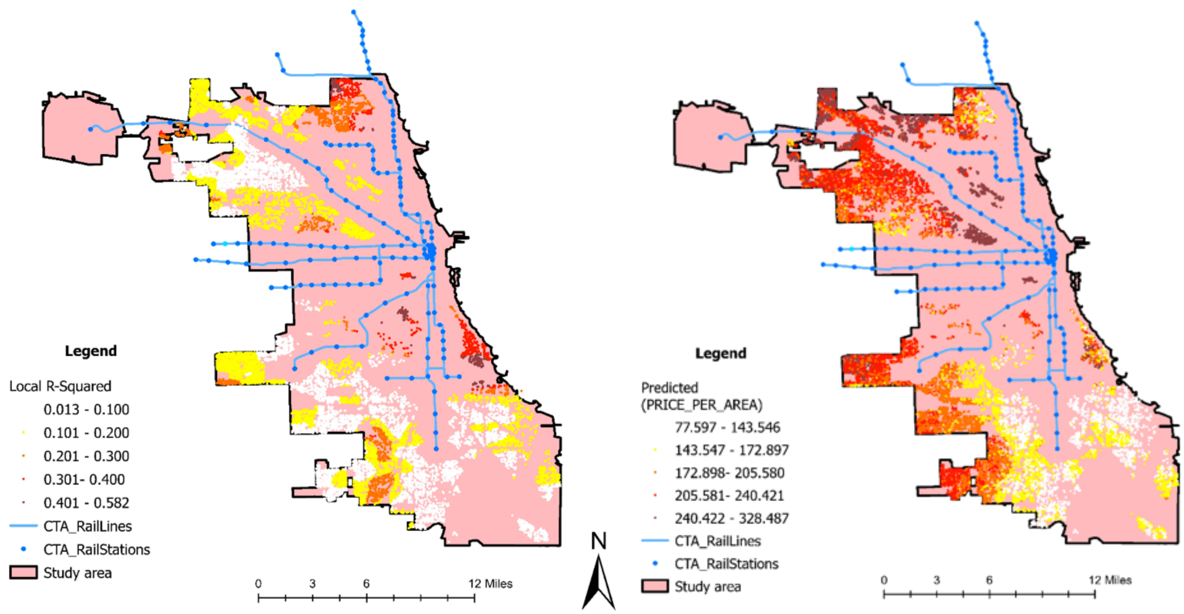
Table 10. Analysis of factors affecting housing prices 1 mile outside Chicago's L train(Comparison of Regression analysis and GWR)

		Multiple Regression Analysis			**Geographically weighted regression**		
					Coefficient(β)		
		B	**β**	**VIF**	**Min**	**Max**	**Mean**
Distance to station(sqft)		-10.141***	-0.182	1.195	-205.6	162.4	3.8
Housing characteristics	**Area for exclusive use**	-0.054***	-0.518	1.901			
	Number of beds	5.754***	0.089	1.712	-19.9	24.6	-0.455
	Number of baths	26.178***	0.298	1.992	-22.7	35.1	-1.22
	Built year	0.250***	0.101	1.202	-1.75	1.74	-0.039
	Parking availability (ref.no parking)	22.588***	0.092	1.055	-103.1	67.4	9.6
Surrounding environment characteristics	**Primary school rating**	6.167***	0.167	1.550			
	High school rating	-0.325	-0.006	1.093			
	Distance to park	-21.42***	-0.067	1.077	-208	149.1	2
	Distance to general hospital	-2.295***	-0.050	1.301	-186.7	189	-0.390
Crime frequency		-0.005***	-0.363	1.678			
Population density		0.001***	0.146	1.352			
Constant		-288.847***			-3253	3655.6	255.4
N		20,027			18,712		
AICc		148,026			205,081		
R^2		0.358			0.426		
Adjusted R^2		0.358			0.399		

*- Dependent variable: Transaction price per square feet ($ Dollar) - *p<0.05, **p<0.01, ***p<0.001*

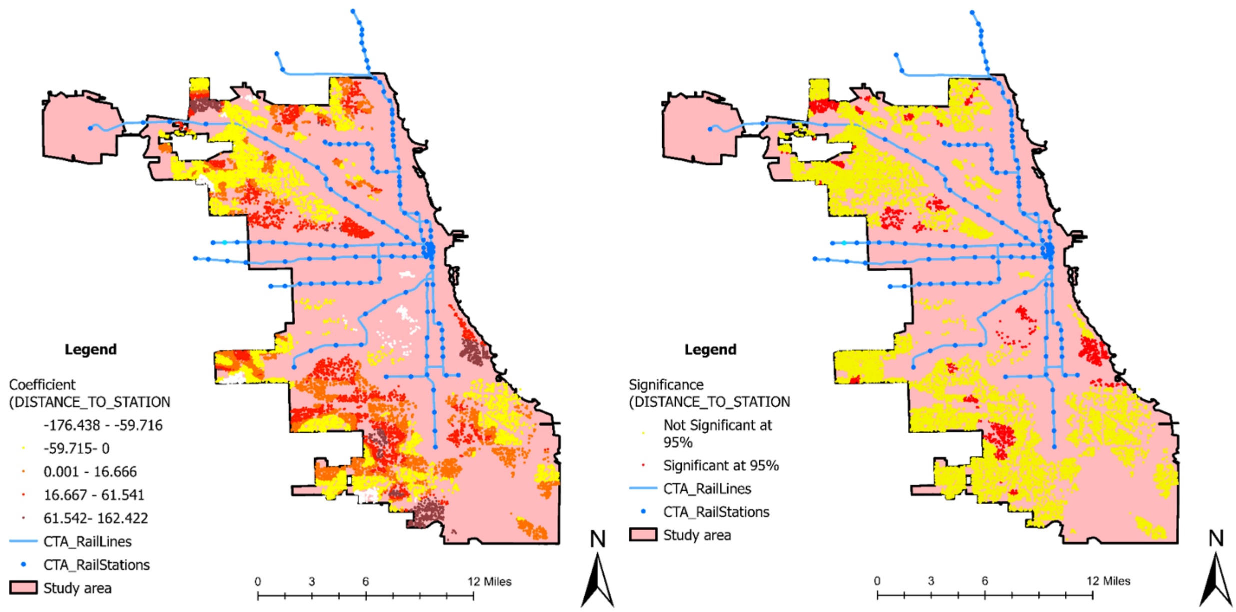
Looking into the details through the map in <Figure 7.>, the R^2map indicates a better explanatory power of the regression model primarily in some outlying areas. Regarding predictions, it becomes evident that housing prices in the west area are higher.

Figure 7. R-square and prediction 1 mile outside Chicago's L train



Moving forward, the focus of study shifts to the impact of the distance variable from the station. The findings are depicted in <Figure 8.> below. Upon scrutinizing the coefficient related to the connection between housing prices and distance to the station, a notable observation emerges—the influence of distance is not uniform throughout the entire region. Interestingly, areas exhibiting high statistical significance are also dispersed randomly across the studied area.

Figure 8. Local coefficient estimates and statistical significance of estimates for "distance to station". 1 mile outside Chicago's L train



6. Conclusion and implications

Chicago stands out among American cities for its low dependence on vehicles, with an average annual miles per driver of 9946.14 (compared to the all-county average of 11620.78) and an average annual miles per vehicle of 10657.81 (versus the all-county average of 11130.59) (National Household Travel Survey, 2020). This trend is attributed to the robust public transportation system in Chicago, encompassing not only the Chicago L but also an extensive network of bus lines and Amtrak services. Reflecting these results, both the hedonic model and ordinary least squares (OLS) analysis affirm a positive correlation between proximity to the station and higher housing prices. However, a geographically weighted regression (GWR) analysis presents a counterintuitive finding: the farther the distance from the station, the higher the housing prices. Examining this in detail, in the analysis of the entire region, areas with higher housing prices were distributed unevenly as the distance to the station decreased. However, in the analysis specifically within 1 mile from the station, areas with higher housing prices were predominantly located in the northern region as the distance to the station decreased. In essence, this suggests that the influence of the Chicago L Train is more significant in the northern area compared to the southern region.

In Chicago, income levels are higher on the north side than on the south side, and public transportation tends to be more effective in high-income areas (Seo et al., 2014; Hess & Almeida., 2007). This trend is primarily attributed to the fact that high-income areas often boast better developed and managed transportation infrastructure, including public transport systems. Additionally, demand for public transportation tends to be higher in high-income areas due to limited parking space and convenience.

In the case of Chicago, public transportation is more advanced in the north than in the south, and the larger population in the north contributes to increased demand for public transportation. Conversely, in the south, accessibility to downtown has only recently improved with the expansion of the Red Line, and demand for public transportation remains low. Consequently, in the northern region, the distance to the station may be a significant factor when purchasing a home, reflecting its impact on housing prices. Furthermore, the importance of accessibility to downtown Chicago in certain outlying areas is evident, as reflected in the GWR analysis results.

Through GWR analysis, detailed regional analyses were conducted, resulting in an up to two-fold increase in the explanatory power of the model. Each explanatory variable better aligns with common sense (e.g., The more rooms and parking spaces there are, the higher the housing price.), enhancing the overall quality of the analysis. However, multicollinearity led to the exclusion of many variables during GWR analysis, limiting the ability to discern detailed effects between variables. Additionally, the decrease in the goodness of fit (increase in AICc) outside 1 mile of the station suggests that GWR analysis may not be superior in all regions, contrary to previous research (Dziauddin., 2019; Tomal., 2020).

Despite these insightful findings, the study has limitations. First, due to data collection constraints, the analysis only covers the past three years. Collecting at least 10 years of data, focusing on recently opened stations, could provide a more comprehensive understanding of effects before and after their opening. Moreover, for a more accurate assessment of public transportation effects, analyzing rental prices rather than housing prices would be appropriate. This is because, in the case of Chicago, like most areas in the United States, owner-occupied homes (generally traded homes) tend to be located in the suburbs, far from public transportation, and rental housing tends to be located close to public transportation. However, insufficient rental data within the last three years (less than 500 data) led to its exclusion from the analysis. If more rental data is collected and the influence of stations is identified through rental data, different results may be obtained.

Currently, the Chicago Transit Authority (CTA) is planning the expansion of the Red Line and Purple Line. While plans for the Orange Line and Yellow Line extensions are not confirmed, they are under consideration. According to the study results, the impact of train station in the less developed south is less effective than in the relatively developed north. Therefore, achieving the desired effect of train line expansion in the southern area may require complex investments in various infrastructure facilities alongside train line expansion. It is my aspiration that this study will serve as a valuable reference for future public transportation expansion plans. I anticipate that the insights garnered from this research will prompt further investigations and studies, contributing to a more comprehensive understanding of related factors and facilitating informed decision-making in urban development and transportation planning.

References

Thünen Johann Heinrich von Wartenberg C. M. & Hall P. (1966). Von thünen's isolated state : an english version of der isolierte staat. Pergamon press.

Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Cambridge, MA and London, England: Harvard University Press.

Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: A spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography*, *41*, 53-62.

Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban studies*, *44*(5-6), 1041-1068.

Dziauddin, M. F. (2019). Estimating land value uplift around light rail transit stations in Greater Kuala Lumpur: An empirical study based on geographically weighted regression (GWR). *Research in Transportation Economics*, *74*, 10-20.

Tomal, M. (2020). Modelling housing rents using spatial autoregressive geographically weighted regression: A case study in Cracow, Poland. *ISPRS International Journal of Geo-Information*, *9*(6), 346.

Tan, R., He, Q., Zhou, K., & Xie, P. (2019). The effect of new metro stations on local land use and housing prices: The case of Wuhan, China. *Journal of Transport Geography*, *79*, 102488.

Yazdanifard, Y., Talebian, M., & Joshaghani, H. (2021). Metro station inauguration, housing prices, and transportation accessibility. *Journal of Transport and Land Use*, *14*(1), 537-561.

Han, D., & Choi, C. G. (2022). A Research on the Influence of the Ui-Sinseol Light Rail Transit on Housing Price: Focusing on the Apartments near the Ui-Sinseol Light Rail Transit. *Journal of Korea Planning Associatio, 57(2),* 108-126.

McDonald, J. F., & Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. *Regional science and urban economics*, *25*(3), 261-278.

McMillen, D. P., & McDonald, J. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics*, *32*(3), 463-486.

Wilcox, M. (2023). Major “L”: How the Chicago Transit Authority Gentrified its Elevated Trains.

So, H. M., Tse, R. Y., & Ganesan, S. (1997). Estimating the influence of transport on house prices: evidence from Hong Kong. *Journal of property valuation and investment*, *15*(1), 40-47.

Mora-Garcia, R. T., Marti-Ciriquian, P., Perez-Sanchez, R., & Cespedes-Lopez, M. F. (2018). A comparative analysis of manhattan, euclidean and network distances. Why are network distances more useful to urban professionals?. *International Multidisciplinary Scientific GeoConference: SGEM*, *18*(2.2), 3-10.

La Rosa, D. (2014). Accessibility to greenspaces: GIS based indicators for sustainable planning in a dense urban context. *Ecological Indicators*, *42*, 122-134.

Web sites

GAWC Global City Index as of 2020 https://www.lboro.ac.uk/microsites/geography/gawc/

Chicago's housing transaction data, https://www.zillow.com

National Household Travel Survey 2022, https://nhts.ornl.gov/

Chicago Transit Authority, https://www.transitchicago.com/

Appendix

Appendix 1. Prices of housing near Elevated Train station

Nearest station	Average Price per station	Line	Average Price per line
Addison-O'Hare	$248.83	Blue Line	$264.38
Austin-Congress	$168.00		
Belmont-O'Hare	$251.94		
Chicago/Milwaukee	$316.20		
Clinton-Congress	$313.52		
Cumberland	$222.17		
Damen/Milwaukee	$336.85		
Division/Milwaukee	$328.46		
Forest Park	$139.64		
Illinois Medical District	$223.68		
Irving Park-O'Hare	$232.92		
Jackson/Dearborn	$236.67		
Jefferson Park	$223.66		
LaSalle	$279.84		
Logan Square	$266.93		
Montrose-O'Hare	$232.48		
Racine	$347.57		
UIC-Halsted	$346.95		
Washington/Dearborn	$293.70		
Cicero-Congress	$134.46	Blue Line (Congress)	$188.35
Kedzie-Homan	$137.39		
Pulaski-Congress	$145.12		
Western-Congress	$209.82		
California/Milwaukee	$280.42	Blue Line (O'Hare)	$275.83
Grand/Milwaukee	$353.52		
Harlem-O'Hare	$250.06		
Western/Milwaukee	$303.21		
Addison-Ravenswood	$340.22	Brown Line	$288.66
Damen-Ravenswood	$298.92		
Francisco	$250.90		
Irving Park-Ravenswood	$333.45		
Kedzie-Ravenswood	$229.93		
Kimball	$223.66		
Montrose-Ravenswood	$297.39		
Paulina	$321.33		
Rockwell	$277.11		
Southport	$347.13		
Western-Ravenswood	$261.84		
35-Bronzeville-IIT	$171.59	Green Line	$197.13
43rd	$178.24		
47th-South Elevated	$180.24		
51st	$187.71		
Cermak-McCormick Place	$290.24		
Cottage Grove	$146.97		
Garfield-South Elevated	$204.92		
Halsted/63rd	$128.72		
Indiana	$173.73		
King Drive	$139.57		
Ashland/63rd	$161.16	Green Line (Englewood)	$161.16
Austin-Lake	$145.94	Green Line (Lake)	$202.66
California-Lake	$256.77		
Central-Lake	$178.16		
Cicero-Lake	$183.51		
Conservatory-Central Park	$202.95		
Harlem-Lake	$207.25		
Kedzie-Lake	$239.60		
Laramie	$167.70		
Pulaski-Lake	$170.97		
35th/Archer	$211.74	Orange Line	$214.19
Ashland-Midway	$232.04		
Halsted-Midway	$251.28		
Kedzie-Midway	$182.83		
Midway Airport	$213.77		
Pulaski-Midway	$223.32		
Western-Midway	$169.73		
18th	$291.60	Pink	$251.43
California-Douglas	$204.78		
Central Park	$163.90		

Cicero-Douglas	$306.22		
Damen	$262.92		
Kedzie-Douglas	$162.80		
Kostner	$160.53		
Polk	$256.91		
Pulaski-Douglas	$161.54		
Western-Douglas	$270.96		
47th-Dan Ryan	$169.97	Red Line	$222.32
63rd-Dan Ryan	$127.04		
69th	$129.05		
79th	$133.29		
87th	$149.12		
95th/Dan Ryan	$168.41		
Addison-North Main	$288.31		
Argyle	$239.44		
Berwyn	$266.66		
Bryn Mawr	$257.73		
Cermak-Chinatown	$273.02		
Chicago/State	$362.59		
Clark/Division	$333.59		
Garfield-Dan Ryan	$122.90		
Grand/State	$380.64		
Granville	$202.90		
Harrison	$249.28		
Jackson/State	$320.40		
Jarvis	$177.30		
Lake/State	$353.86		
Lawrence	$258.81		
Loyola	$195.14		
Monroe/State	$270.01		
Morse	$195.39		
North/Clybourn	$386.56		
Roosevelt/State	$297.42		
Sheridan	$260.64		
Sox-35th-Dan Ryan	$240.43		
Thorndale	$211.76		
Dempster-Skokie	$267.05	Yellow Line	$268.63
Oakton-Skokie	$269.26		
Adams/Wabash	$409.97	Brown, Orange, Pink, Purple (Express), Green	$439.87
State/Lake	$443.14		
Washington/Wabash	$440.32		
LaSalle/Van Buren	$337.32	Brown, Orange, Pink, Purple (Express)	$298.65
Library	$207.38		
Washington/Wells	$298.28		
Clark/Lake	$362.94	Brown, Orange, Pink, Purple (Express), Green, Blue	$362.94
Armitage	$435.49	Brown, Purple (Express)	$364.10
Chicago/Franklin	$379.62		
Diversey	$345.64		
Merchandise Mart	$372.00		
Sedgwick	$369.23		
Wellington	$311.92		
Ashland-Lake	$320.06	Green (Lake), Pink	$354.80
Clinton-Lake	$369.12		
Morgan	$395.11		
Roosevelt/Wabash	$338.16	Orange & Green Lines	$338.16
Main	$163.72	Purple Line, Evanston Express	$163.72
Wilson	$243.85	Purple, Red Line	$243.85
Belmont-North Main	$312.65	Red, Brown, Purple (Express)	$346.06
Fullerton	$369.01		
Howard	$199.74	Red, Yellow, Purple, Evanston Express	$199.74

Appendix 2. Multiple Regression Analysis_Within 1 mile_include area

		Basic model			Environment model			Crime model			Population model		
		B	**β**	**VIF**	**B**	**β**	**VIF**	**B**	**β**	**VIF**	**B**	**β**	**VIF**
Distance to station(sqft)		-63.428***	-0.141	1.050	-50.423***	-0.110	1.072	-26.709***	-0.058	1.107	-24.739***	-0.054	1.158
Housing characteristics	**Area for exclusive use**	0.007***	0.071	4.350	-0.004**	-0.041	4.958	-0.014***	-0.137	5.081	-0.014***	-0.137	5.081
	Number of beds	-24.578***	-0.289	2.887	-16.349***	-0.191	3.333	-2.059*	-0.024	3.699	-1.649	-0.019	3.762
	Number of baths	30.037***	0.321	4.649	31.295***	0.333	5.214	29.581***	0.315	5.219	29.433***	0.313	5.225
	Built year	0.502***	0.196	1.147	0.476***	0.187	1.157	0.396***	0.156	1.170	0.394***	0.155	1.171
	Parking availability (ref.no parking)	14.277***	0.049	1.154	15.652***	0.054	1.162	18.044***	0.062	1.163	17.882***	0.062	1.164
Surrounding environment characteristics	**Primary school rating**				8.467***	0.194	1.156	4.378***	0.100	1.272	4.675***	0.107	1.400
	High school rating				2.222***	0.050	1.199	1.841***	0.042	1.199	2.191***	0.049	1.372
	Distance to park				-15.204***	-0.022	1.053	-9.738*	-0.014	1.054	-8.937*	-0.013	1.058
	Distance to general hospital				-22.421***	-0.142	1.207	-10.223***	-0.065	1.285	-10.43***	-0.066	1.290
Crime frequency								-0.013***	-0.348	1.592	-0.012***	0.337	1.933
Population density											0.0001***	-0.025	1.760
Constant		-692.9***			-673.552***			-478.205***			-473.479***		
N		27887											
F		806***			714***			957***			879***		
R^2		0.16			0.231			0.307			0.308		
AdjustedR^2		0.16			0.231			0.307			0.307		

Appendix 3. Ordinary Least Square Regression (OLS)

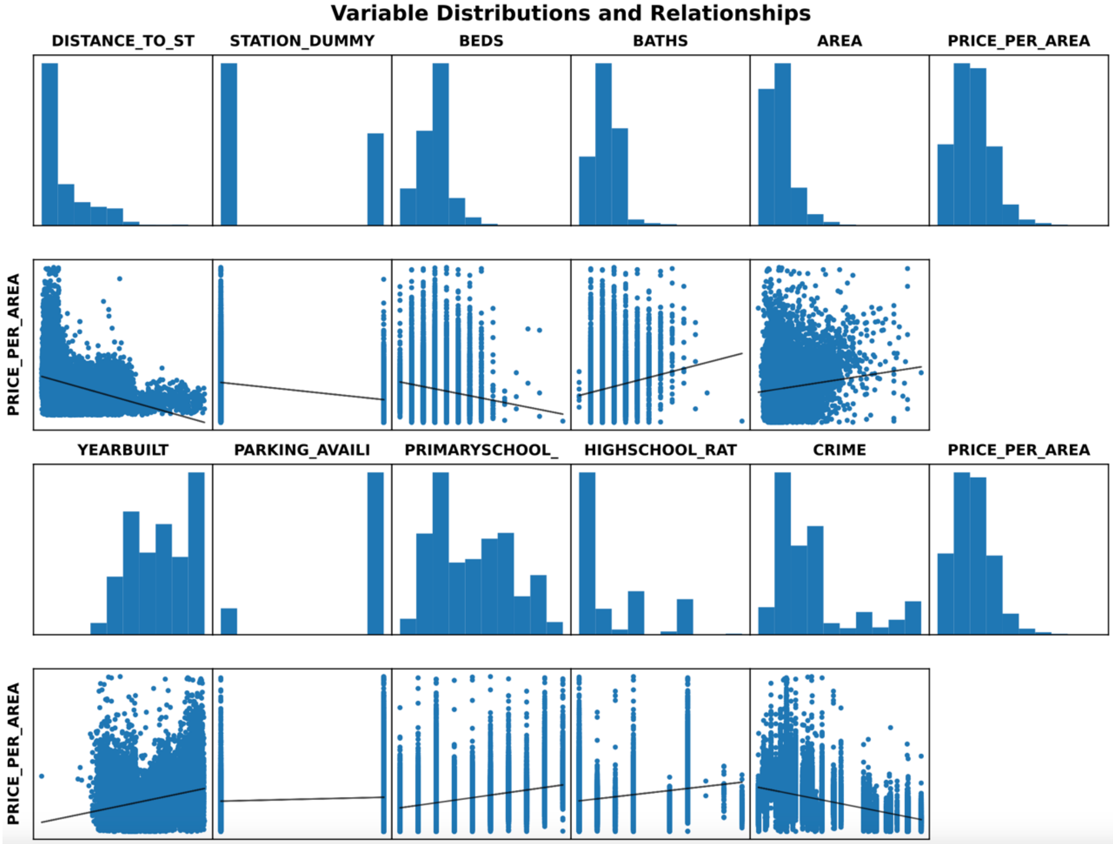
Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-477.694413	23.895353	-19.991101	0.000000*	26.411572	-18.086558	0.000000*	--------
DISTANCE_TO_	-11.379658	0.555330	-20.491686	0.000000*	0.440706	-25.821437	0.000000*	2.869934
STATION_DUMM	-35.042049	1.563111	-22.418143	0.000000*	1.376153	-25.463781	0.000000*	3.068413
BEDS	-0.146235	0.574757	-0.254428	0.799170	0.622194	-0.235030	0.814187	2.919819
BATHS	29.527668	0.798152	36.995058	0.000000*	0.895298	32.980836	0.000000*	3.734741
AREA	-0.023186	0.000841	-27.566110	0.000000*	0.001220	-19.005216	0.000000*	3.593499
YEARBUILT	0.379994	0.012174	31.214186	0.000000*	0.013535	28.074244	0.000000*	1.218644
PARKING_AVAI	20.144902	1.316233	15.304966	0.000000*	1.409192	14.295359	0.000000*	1.132568
PRIMARYSCHOO	6.103150	0.220799	27.641259	0.000000*	0.234691	26.004993	0.000000*	1.442400
HIGHSCHOOL_R	2.995334	0.227913	13.142423	0.000000*	0.234811	12.756381	0.000000*	1.327171
CRIME	-0.007229	0.000138	-52.480573	0.000000*	0.000117	-62.044513	0.000000*	1.684355
TOTAL_POPULA	-0.000089	0.000025	-3.591426	0.000344*	0.000023	-3.830784	0.000139*	1.487375
DISTANCE_TO_	-7.538834	0.414506	-18.187510	0.000000*	0.364260	-20.696278	0.000000*	1.415409
DISTANCE_TO_	-19.045036	2.371408	-8.031109	0.000000*	2.230286	-8.539279	0.000000*	1.247794

OLS Diagnostics

Input Features	Sort_transaction data	Dependent Variable	PRICE_PER_AREA
Number of Observations	37214	Akaike's Information Criterion (AICc)['d']	434278.191263
Multiple R-Squared['d']	0.409728	Adjusted R-Squared['d']	0.409522
Joint F-Statistic['e']	1986.293436	Prob(>F), (13,37200) degrees of freedom	0.000000*
Joint Wald Statistic['e']	31364.684092	Prob(>chi-squared), (13) degrees of freedom	0.000000*
Koenker (BP) Statistic['f']	1058.078838	Prob(>chi-squared), (13) degrees of freedom	0.000000*
Jarque-Bera Statistic['g']	56046.689859	Prob(>chi-squared), (2) degrees of freedom	0.000000*

Variable Distributions and Relationships



Variable Distributions and Relationships (Cont.)