

Implying community vitality index through housing factors

By Mark Li

AUTHOR BIO

Mark Li, a 16-year-old Junior at Avon High School, CT, excels in mathematics, computer science, and physics. Mark's passion extends to data science and social science research, where he combines machine learning techniques with social sciences to explore complex issues. Beyond academics, he enjoys absorbing new knowledge and finds fulfillment in guiding his peers. In the future, Mark aspires to tackle societal well-being or dilemmas by incorporating AI design as a generalized calculator for medicine, making a meaningful impact on the world's challenges.

ABSTRACT

The Community Vitality Index (CVI) is pivotal in assessing community well-being, informing decision-making processes, encouraging community participation, and efficiently distributing resources to enhance residents' overall quality of life. Leveraging contemporary housing data and sophisticated multivariate regression algorithms allows us to train a machine-learning model for a real-time vitality index. This enables prompt recognition and response to evolving community dynamics, all without delays.

Keywords: *Real-Time, Community Vitality Index, Machine Learning, Multiple Regression Analysis, Housing data, Census*

INTRODUCTION & BACKGROUND

Persistent disparities have existed in various aspects of life within American communities. To combat these inequalities, a concerted effort has involved collaboration with academic societies, policymakers, consulting firms, governmental agencies, and non-profit organizations.

Academic societies: Dr. LaPlante at the University of Southern Maine and her team defined 'community vitality' as the well-being that attracts residents, identifying key indicators, including aesthetics, infrastructure, services, recreation, and transportation, while considering social and economic factors.

Think Tank: The Economic Innovation Group, a bipartisan public policy think tank, releases its Distressed Communities Index (DCI) to gauge the economic well-being of U.S. communities and highlight disparities nationwide. The DCI categorizes zip codes by evaluating factors that include adult education levels, property rates, median income ratios, and more.

Consulting Institutes: Fourth Economy, a consulting firm, has developed a comprehensive community index that evaluates vital metrics of flourishing communities. This index comprises 20 factors, such as gross domestic product per employee, the per capita ratio of college students, local food availability, net migration, and other elements. This evaluation is conducted county-wide.

Nonprofit organizations: DataHaven, supported by the Hartford Foundation for Public Giving, creates the Greater Hartford Community Well-being Index, which includes economic, health, and educational well-being indicators applicable to Hartford and its neighboring towns.

Community indices typically rely on data from the 5-year U.S. census survey, which limits their capacity to capture real-time community dynamics. Furthermore, some of these indices operate at the county or metropolitan level, potentially lacking the detailed community-level insights that local governments require.

This paper proposes incorporating a house price-based index into assessing community vitality. House prices can reflect a community's overall health, including investments, local economic conditions, and government policies. They also provide insights into a region's economic development and urban vibrancy, with higher housing prices often correlating with improved living conditions that benefit residents' physical and mental well-being. Moreover, house price fluctuations can significantly impact economic growth and consumer spending, with sharp declines potentially triggering recessions.

Two main categories of factors influence house prices: those directly linked to the physical structure of the house, such as square footage (referred to as "location-independent factors"), and those tied to the overall health of the community, like crime rates (referred to as "location-dependent factors"). By focusing on the factors related to location, we can create an index for comparing community vitality. The resulting index will also provide real-time insights since housing data is updated with every transaction, such as buying or selling a house.

The Indiana State Vitality Index was developed by combining weighted factors like "income, education, [and health]" (referred to as IUPPI). However, this dataset doesn't cover all the elements influencing community vitality. To comprehensively understand all factors related to communities, this

paper suggests using an exclusion-based method, which involves filtering out irrelevant location-dependent factors from housing prices.

RELATED WORKS

How a community index is calculated commonly: A CVI can be directly calculated by amortization. According to Benson, the primary factor of price appreciation is the location of the home, essentially, the location-dependent factors. Closeness to work, services, or recreational areas increases the CVI. “Homes adjacent to natural resources like parks and open spaces hold an 8%-20% higher value than comparable properties”(University of Washington). This quote demonstrates the value of natural resources in improving community vitality. Crime rate is also a significant factor that impacts a community's vitality. A study by Florida State University [10] shows that a 10% increase in violent crimes within a neighborhood reduces house values by 6% in Miami-Dade County, Florida. These studies indicate the importance of safety and a low crime rate in relation to community vitality.

While these factors are essential, they only constitute a fraction of the comprehensive variables required to create a strong vitality index. Other crucial factors, including population size, median household income, and various socioeconomic indicators, substantially impact the overall vitality index. However, these data points are typically updated every five years during the census, resulting in significant delays between updates. Additionally, gathering and updating this extensive dataset is both resource-intensive and time-consuming. Furthermore, collecting and updating this extensive dataset is both time-consuming and labor-intensive. In contrast, a vitality index centered on housing prices, excluding location-independent variables, benefits from real-time updates through fresh housing transaction data.

A particular case highlights the importance of the community index: Zillow's 2021 loss of more than \$400 million was associated with its failure to consider the vitality index, additionally, its neglect of rapid inflation during that period and the overall housing market's stagnation (What happened at Zillow? CNET). However, according to Purdue, population was the most significant factor impacting the CVI. The COVID-19 pandemic, which led to more people working from home due to increased remote work, strongly influenced community vitality.

PREVIOUS RESEARCH

The Indiana government employed the grid-based partitioning method, specifically INDIANA-GSS (Grid Scanning System), to develop a vitality index for its communities. However, the index values generated by INDIANA-GSS were inaccurate for certain regions characterized by low house density and sparse populations. The paper adopts a digressing approach known as locational clustering to address this limitation.

According to the Indiana government, most smaller and larger counties have an inaccurate vitality index value at 49.34% accuracy, in contrast to the 80.00% accuracy threshold. These counties behave similarly to outliers, where population-related factors are in the 1st and 4th quartile in the regression process and too extreme for the regression model to produce the vitality index. By normalizing the population density in each subdivision, the clustering method generates a functional index of those areas within the previous density distribution's 2nd and 3rd quartile, ensuring there are no outliers concerning population density.

Although clustering processes the data to detect outliers high-dimensionally, the data could form a circular cluster and negatively affect the model's accuracy. The following section describes using HDBSCAN and K-Means++ to fix the circularization issue.

According to Azemlu, their data cleaning used a similar cluster algorithm to process the data used in the regression. In this paper, the algorithm cleaned “cluster data using different methods such as K-means [and] DBSCAN.” (House Price...Comparative Study) In the model, DBSCAN eliminated some outliers, which are data points in a region with a data density of below 0.1 points per square unit, and the K-means clustering improved R^2 value to accuracy above 70% by finding houses with similar values in each factor, and comparing house prices. However, upon using clustering, the cleaned data points in the price calculation model had a tendency to resemble a round-shaped cluster, which considerably impacted the regression's accuracy and decreased the R^2 value by 18.34%. Data circularization occurs in multidimensional datasets, where outliers are excluded through clustering to lower regression accuracy.

Preventing the data cluster from becoming elliptical was solved in the Price Determination model outlined in this paper. When the clusters become elliptical, accuracy decreases significantly by 28.12%. In this paper, to obtain the accuracy and reliability model, a function for preprocessing data was created to remove outliers based on location-independent factors in the three steps below.

1. HDBSCAN clustering produces a membership score for every data point. All data points with a membership score of 0.7 or higher were excluded, creating a lenient and controllable outlier removal system.
2. Linear regression analysis estimates the coefficient for each variable in the housing dataset.
3. The slope values defined two parallel hyperplanes, and the y-intercept was adjusted to encompass 65% of the data within the range of hyperplanes; other data was dropped.

For Hartford County locational clustering graphed by Matplotlib below, K-means was paired with K-means++ to find the optimal number of clusters¹. K-means identified outliers via the longitude and latitude of houses and clusters to control location-dependent factors.

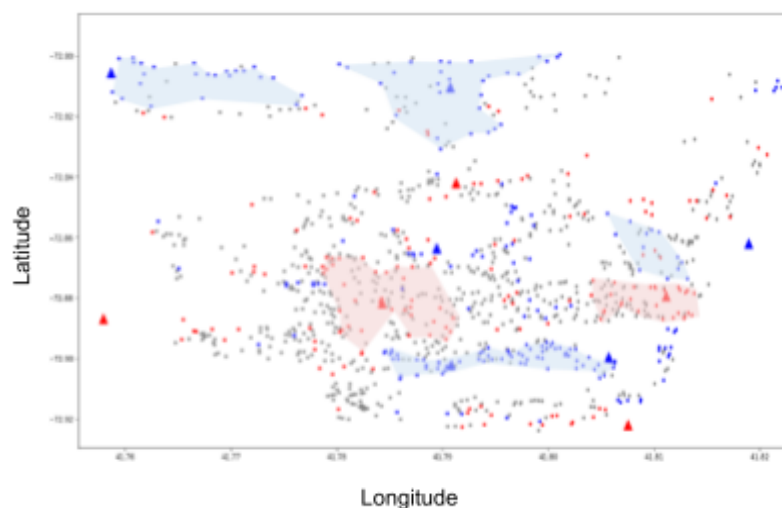


Figure 1. Housing Clusters in Hartford County, CT

¹ Kmeans++ identified four 6 high priced clusters, 5 low-priced clusters, and 10 medium-priced clusters

Figure 1 reflects low and high house prices. Red denotes house prices below 350,000 dollars, and blue indicates house prices above 750,000. The shaded regions represent actual clusters of houses. Triangles represent potential clusters of houses.

METHODOLOGY & DATA PREPARATION

Location-independent and location-dependent are two overarching parts of the regression model.

Location-independent: Square Footage, Square Footage of Heated Area, Lot Size, Number of Rooms, Number of Bedrooms, Number of Bathrooms, The Number of Garages, Whether the house has a pool, Flooring material, External Siding material, Heating fuel type.

Location-dependent: Crime Rate, Police Count, Number of Nearby Highways, Number of Nearby Local Roads, Number of Nearby Service Roads.

All houses missing important information were dropped from the dataset in the file. The categorical variables were split into several categories of binary variables for usage in machine learning algorithms.

MODEL-PD (PRICE DETERMINATION)

SciKit-Learn is the primary machine-learning Python library, including linear regression and logistic regression packages that provide classification, clustering, and regression algorithms.

MODEL-PD is an agglomerative hierarchical regression model to determine housing prices and create a CVI normalized for the migration classification model. Initially, the regression model used square footage as an independent variable and house prices as a dependent variable, partitioned by the year built, controlling time in terms of decades. Increasing the location-independent dataset from two dimensions to five also turns the regression model into a multiple linear regression model to control location-dependent variables. Next, adding K-means accounted for the location, and adding HDBSCAN removed outliers in the location-independent dataset. Then, MODEL-PD split the dataset into binary and non-binary variables and converted categorical variables into a series of binary variables. Finally, an additional proportional weighting splits the house price between these two types of variables. MODEL-PD development consists of two interactions.

Variable Selection

1. Square Footage as an independent variable and house prices as a dependent variable were chosen for the linear regression with OLS because they were the most strongly correlated, at 0.6092.
2. Segmenting the year built and controlling the sold year improves the correlation to 0.6792.

The following graph depicts the time-segmented relationship:

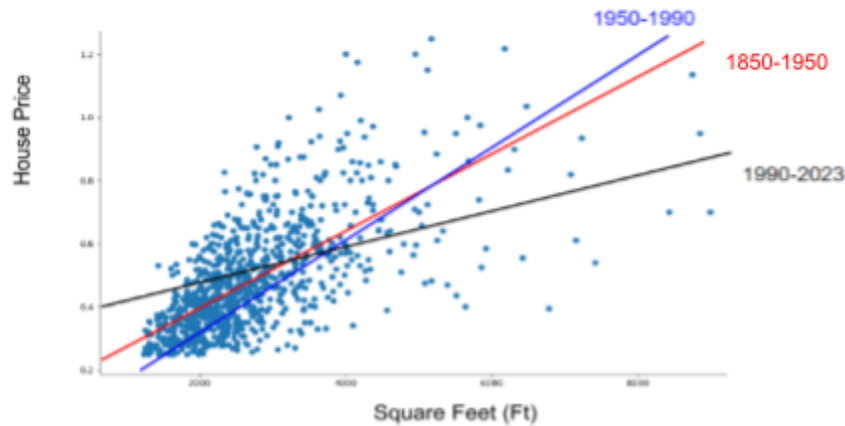


Figure 2. Year Partitioned Comparison of Sq. Footage and House Price

The R^2 value of the data segments was originally 0.6792 ± 0.3059 . Introducing location-independent variables decreased the R^2 value and variance to 0.5381 ± 0.2101 .

Data Cleaning

1. HDBSCAN clustering removed noise and outlier points, and K-Means clustering found clusters of houses with similar prices and attributes, whose means were subtracted respectively. The accuracy improved to 0.7302 ± 0.2113 .
2. Additional categorical and binary location-independent variables were added to help HDBSCAN identify noisy points more efficiently by increasing the distance between data points.

Model Selection

1. A hierarchical regression method processed the categorical, binary, and continuous numeric variables separately.
2. Adding Crime Rate measured the effects of a location-dependent variable. The coefficient of the Crime Rate was consistent across all houses and time series, suggesting that location-dependent variables can be calculated separately from location-independent variables, indicating a proportional weighting to calculate the community index.
3. By weighting proportion, the price was split into 71.42% for location-dependent price determination and 28.58% of the price for calculating the CVI.
4. To adjust for inflation from COVID-19, housing prices increased by 30% since 2019.
5. Next, the CVI was normalized into the final index by the following formula:

$$\text{Unprocessed Community Index} / \text{Population in Community} \times 1000 - 1000$$

The minimum index calculated was 1192, so subtracting 1000 would keep the index in the positive range.

MODEL-PD Results

The table below lists all the independent variables and coefficients for location-independent factors used in the regressive location-dependent variable analysis.

Table 1: Independent variables and coefficients

Variable		Coefficient	Confidence ²
Y-Intercept		-129,352	N/A
Numerical Variables (71.2%)			
Square Footage		2618*	0.088
Square Footage of Heated Area		149*	0.075
Lot Size		125*	0.018
Number of Rooms		1724*	0.047
Number of Bedrooms		1392	0.715
Number of Bathrooms		1102	0.692
Roads	Number of Nearby Highways ³	-5161	0.654
	Number of Nearby Local Roads	4353	0.476
	Number of Nearby Service Roads	1043	0.142
Location Independent Factors	Crime Rate (Normalized by a factor of 1000) ⁴	-4.76	0.276
	Police Count	-102	0.445
Binary/Categorical Variables (28.8%)			
Has a Pool		19983*	0.045
Flooring Information	Has Hardwood Floor	19276*	0.098
Exterior Siding Information	Exterior Siding Includes Wood or Other	-912	0.158
	Exterior Siding Includes Brick or Stone	15407	0.466
Heating Utility Information	Heating Uses Natural Gas or Oil	93	0.539
	Heating Uses Electric power	-463	0.572

² The margin for statistical significance is 0.1 or less. The value used is a P-value. Categorical and Binary variables are handled separately in the regression model.

³ Road type is defined by the standards in the OSM API: <https://wiki.openstreetmap.org/wiki/Key:highway>

Nearby is defined as being within a 5 km radius of the house.

Highways are roads that are motorway, trunk, or primary.

Local Roads are secondary or tertiary

Service Roads are service, residential, or living streets.

⁴ Crime and police count are used as a factor that shows the effect of a location dependent variable on the price.

MODEL-MC (MIGRATION CLASSIFICATION)

MODEL-MC is a multivariate logistic regression model that predicts the trend in population migration and validates the accuracy of the CVI. The normalized CVI was created by taking 28.58% of the total actual price because this percentage produced the best correlation between location-independent factors in MODEL-PD. Model-MC then performed regression using the CVI to produce the NMR⁵. Population data was collected from government census sources (census.gov), and MODEL-PD calculated the CVI. The logistic regression model used the NMR as the dependent variable and CVI as the independent variable to determine whether the NMR would increase or decrease. According to Cato, the vitality index and NMR reflect the collection of location-dependent variables. (Immigration and Community Vitality) Population changes also reflected these variables and established a causal relation between the vitality index and the net migration rate by logistic regression.

TESTING MODEL-PD USING MODEL-MC

The CVI and population rate change trained MODEL-MC and produced a confusion matrix that proves the index accurately predicts population rate change and validates the index. The confusion matrix compares predicted population rate change with census migration information recorded below. The accuracy of the CVI calculated by MODEL-PD is 0.7123 ± 0.1092 , which is sufficiently precise and accurate for use with MODEL-MC, which had an accuracy of 0.9064. With the location-dependent factors, logistic MODEL-MC was accurate, with this confusion matrix:

Table 2: Model MC with hierarchical predicted CVI

Model MC with hierarchical predicted CVI	Census Information Indicates Migration Rates Increased	Census Information Indicates Migration Rates Decreased
Predicted Migration Rates Increase	True Positive: 64 (49.6%)	False Positive: 8 (6.2%)
Predicted Migration Rates Decrease	False Negative: 4 (3.1%)	True Negative: 53 (41%)

According to the National Academy Press, 0.906 of the values produced were correct, showing that the MODEL-MC demonstrated sufficient accuracy for community population migration prediction, with an acceptability threshold set at 81%.

Compared to the multiple logistic method of predicting the trend in migration, the number of correct results was considerably more significant, as shown in the following confusion matrix.

Table 3: Model MC with Multivariate logistic CVI

⁵ Net Migration Rate can be calculated at the difference between Influx and Outflow

Model MC with Multivariate logistic CVI	Census Information Indicates Migration Rates Increased	Census Information Indicates Migration Rates Decreased
Predicted Migration Rates Increase	True Positive: 29 (25%)	False Positive: 26 (22.4%)
Predicted Migration Rates Decrease	False Negative: 23 (19.8%)	True Negative: 38(32.7%)

MATHEMATICAL EXPLANATION

The two models are meant to (1) find the CVI, and (2) validate the CVI by comparing predicted NMR with census NMR. The location-independent factors and the housing price are compared in MODEL-PD. The relationship can be represented as the formula below, where $f(x)$ cumulates location-independent factors into a price, $g(x)$ cumulates location-dependent factors into a price, P is the final house price, and e is the error:

$$P = f(\text{Location-independent factors}) + g(\text{Location-dependent factors}) + e$$

The Cumulative Location-independent factors are the price from location-independent factors. Cumulative Location-dependent factors are the unprocessed CVI. Error is composed of unaccounted location-independent factors. The Price is the housing price.

MODEL-PD uses a proportioning algorithm to split Price between CVI and location-independent price, and exclude errors caused by unaccounted location-independent factors, so the CVI is unaffected by location-independent factors. The proportioning algorithm finds the optimal distribution of price value between the proto-CVI and location-independent regression with respect to the regression R^2 value.

Proportioning splits the Price for MODEL-PD and location-dependent factors for CVI determination.

With the Price split, the equation is as follows:

$$\begin{aligned} \text{Price} - \text{Cumulative Location Independent Factor Index} \\ = \text{Cumulative Location-dependent factors} + \text{Error} \end{aligned}$$

The Cumulative Location-dependent factor is the unnormalized CVI, which is then scaled. According to “Immigration and Community Vitality,” the “[NMR] acts as a community indicator,” as many location-dependent factors, like the CVI, influence it.

$$\sum \frac{|O_i - E_i|^2}{|E_i|}$$

EXPLANATION OF POSSIBLE FACTORS

Location-Independent: Location-independent factors are directly proportional to the house price that will increase as construction costs.

Location-Dependent: By predicting the house price in the future, it is possible to calculate the CVI, as the amortized location-independent variables are unlikely to change significantly in a decade in a well-developed community.

Examination of two specific clusters

Table 4: Details of two clusters with different location-independent prices and different CVIs

Details	Avon Cluster	Farmington Cluster
CVI	889	793
Location-independent price	601202	773764
Centroid Location	(41.8088,-72.8882)	(41.7319,-72.8227)

The CVI of the Avon Cluster is higher because of the centroid's proximity to Avon High School. At the same time, the Location-independent price is higher for the Farmington Cluster due to an average 20% house size increase.

CONCLUSION

This paper adopts two different machine-learning models. The agglomerative hierarchical regression model creates and examines the Community Vitality Index and its close correlation with other community well-being indicators. CVI reflects a holistic view of the indicators showing a community's well-being.

NMR is an essential component of CVI, predicting population immigration. The logistic regression model provides a validation module to interpolate NMR, which is used to verify the CVI introduced by comparing it with the NMR value from the census. CVI can be updated based on the latest house sales data and does not have to wait for the 5-year census data to show the community's ever-changing dynamics.

The house price determination model derives CVI from house sales information. The level of details in the house sales information determines the CVI's granularity, such that the CVI's flexibility for both small and large areas allows users to create their search criteria on either a city, a county, or a region controlled by zip code combinations.

The CVI, along with other measures such as unemployment rate, poverty rate, household income, etc., can be used to measure the comparative economic well-being of communities and helps illuminate ground-level disparities across the country. These measures together can provide real-time insights into the spatial distribution of economic well-being during economic and social volatility.

DISCUSSION

Use Case of MODEL-PD

The figures below depict the average house price of each grid square by regions of Zipcode combinations. High prices are shown with lighter colors and low prices are displayed with darker colors.



Figure 3. Heatmap of average Connecticut suburban housing prices (Greater Hartford County)



Figure 4. Heatmap of average Texas housing prices (Plano, Houston)



Figure 5. Heatmap of average Minnesota suburban housing prices(Plymouth, Minnetonka)

Figure 3 shows a cluster of high-valued houses concentrated around (0, 5) and (6,0); the cluster for Figure 4 is at (1,8). The cluster for Figure 5 is at (2,1), which shows that the locational structure of data in different areas is similar. The similarity and high accuracy enable the utilization of MODEL-PD for any community. It will produce a community region and a CVI for it.

Predicting Migration Rate using CVI

MODEL-MC with the CVI can ascertain whether the population of a given area will increase and the extent of the increase. Extrapolating the people was done through logistic regression, as only four grades of NMR were recorded. These indexes are correlated reasonably well through their correlation on the quartiles. The logistic model can be visualized as such:

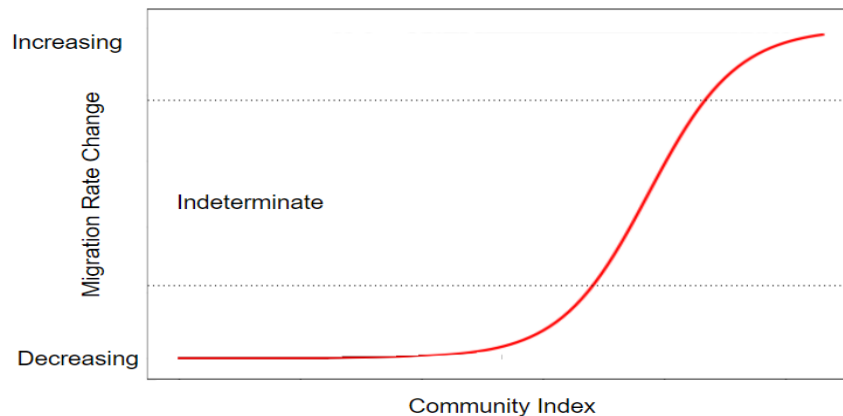


Figure 6. Comparison of CVI and NMR change

Figure 6 shows the relation between CVI and migration rate change. The index range is around 1200, with the chance of an increase in NMR having a threshold of 623. This graph used amortized data.

COMPARING CALCULATED CVI WITH OTHER COMMUNITY STATISTICS

Calculated CVI vs. Existing DCI

The CVIs for the cities of Greater Hartford County are displayed below, as well as the indexes produced through amortizing census information, the Distressed Community Index (DCI):

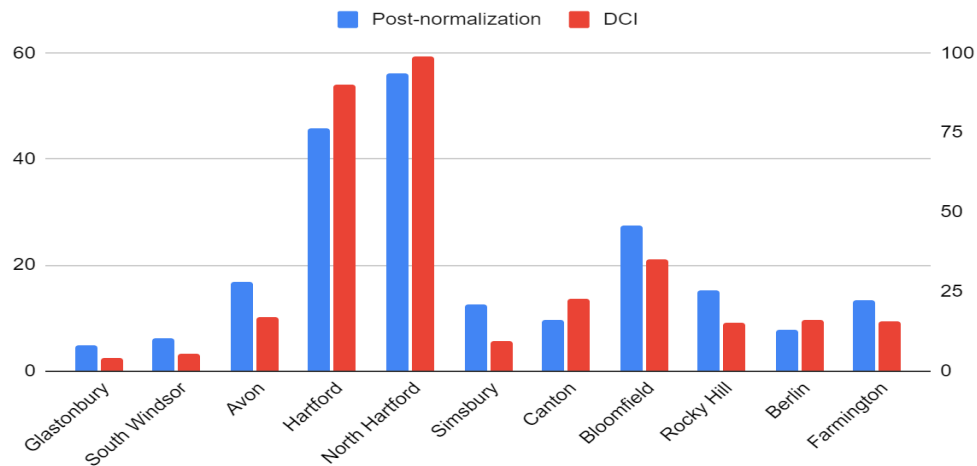


Figure 7. Comparison of DCI and Scaled CVI⁶

These indexes are correlated with each other in terms of their ranking, showing how the index produced from housing prices is accurate. However, because the vitality index used an exclusive method to produce the index, location-independent variables still needed to be removed, leading to a base value of around 1,000 in the index. This is subtracted after population normalization from the normalized value. Here is a graphical depiction:

Here, the rankings are pretty well correlated, showing that this index complements the indices produced from census information. The scale on the left describes the post-normalization scaled CVI and the scale on the right represents the DCI.

School quality was a way to compare these indexes in terms of their order, followed by crime and air quality.

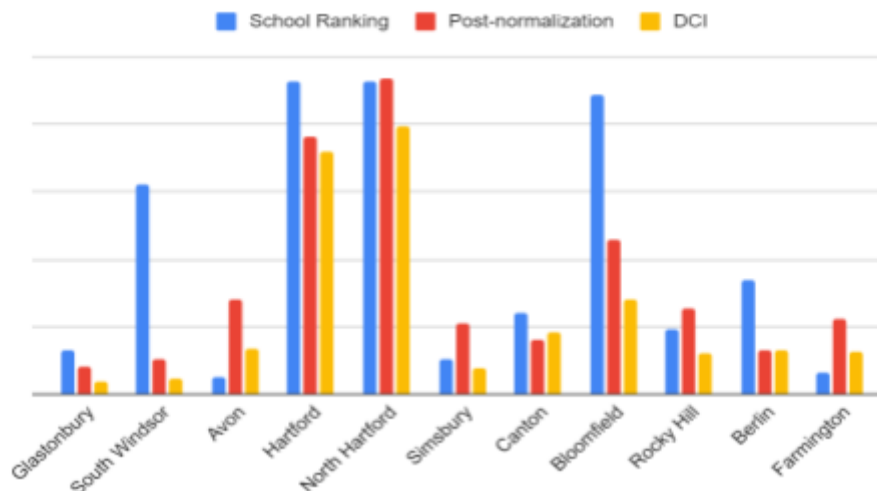


Figure 8. Comparison of School Ranking, Scaled CVI, DCI

⁶ Normalization was conducted by dividing the unprocessed index by the city's population.

Figure 8 shows that school quality is pretty well correlated with the CVI, with both indexes coinciding with the rankings of the schools.⁷ and ⁸ The school ranking, DCI, and normalized CVI are scaled to fit in one graph.

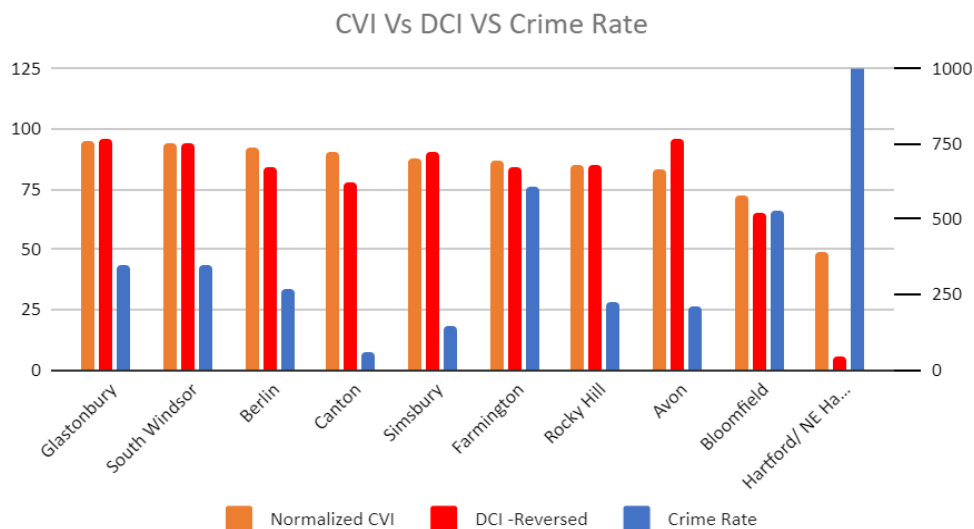


Figure 9. Comparison of Crime Rate, Scaled CVI, DCI⁹

There is a slight trend in terms of the crime rate.¹⁰ The crime rates are similar in communities with a higher vitality index or lower DCI. However, crime rates in areas with high DCI or low vitality index are much larger. Hartford was not included due to its exceptionally high crime rate and DCI preventing scaling of the other crime rates.

REFERENCE

1. "The Accuracy of Past Projections." *National Academies Press*, 2000, nap.nationalacademies.org/read/9828/chapter/4. Accessed 28 Aug. 2023.
2. Azimlu, Fateme, et al. "House Price Prediction Using Clustering and Genetic Programming along with Conducting a Comparative Study." *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, July 2021. Crossref, <https://doi.org/10.1145/3449726.3463141>.
3. Indiana University Public Policy Institute. "Community Vitality Index." 15 Nov. 2015. PDF. https://archives.iupui.edu/bitstream/handle/2450/10109/ThrivingComm_Vitality_Web.pdf

⁷ School rankings were obtained from Niche.com

⁸ Northeast Hartford does not have a school district.

⁹ Note that the data for Hartford/NE Hartford overshoots the graph

¹⁰ Crime rate was obtained from Census Information

4. Johnson, Reid. "Building the Neural Zestimate." *AI, Machine Learning and Research*, edited by Reid Johnson, Zillow, 23 Feb. 2023, www.zillow.com/tech/building-the-neural-zestimate/. Accessed 23 Aug. 2023.
5. "Labor Market Trends and Local Job Strategies." *Labor Market Trends and Local Job Strategies*, Community Economics, June 1997, api.aae.wisc.edu/pubs/pdf/cenews/ce248.pdf. Accessed 23 Aug. 2023.
6. Schwartz, Amy Ellen, & Susan Wachter. 'COVID-19's Impacts on Housing Markets: Introduction'. *Journal of Housing Economics*, vol. 59, 2023, p. 101911, <https://doi.org/10.1016/j.jhe.2022.101911>.
7. Vigor, Jacob L. "Immigration, Housing Markets, and Community Vitality." *Cato Journal*, edited by Jacob L. Vidgor, Cato Institute, Sept. 2017, www.cato.org/cato-journal/fall-2017/immigration-housing-markets-community-vitality#the-challenge-to-discerning-impact-in-data. Accessed 28 Aug. 2023.
8. "2023 Home Buyers and Sellers Generational Trends Report" National Association of REALTORS® Research Group
9. "Vania Ceccato & Mats Wilhelmsson (2020) Do crime hot spots affect housing prices? Nordic Journal of Criminology, 21:1, 84-102, DOI: 10.1080/2578983X.2019.1662595"
10. "Crime and Housing Prices" Keith Ihlanfeldt, Tom Mayock, Florida State University, February, 2009"
11. "Unemployment and the US Housing Market during the Great Recession" Pavel Krivenko, Stanford Economics, February 2018
12. "The Real Estate Mantra –Locate Near Public Transportation," October 2019
13. Niche. "2023 Best School Districts in Connecticut." 2023 Best School Districts in Connecticut, Niche, Apr. 2023, www.niche.com/k12/search/best-school-districts/s/connecticut/?page=5. Accessed 10 Sept. 2023.
14. Parker, Will, and Nicole Friedman. "Zillow Quits Home-Flipping Business, Cites Inability to Forecast Prices." *The Wall Street Journal*, edited by Will
15. Parker and Nicole Friedman, www.wsj.com/articles/zillow-quits-home-flipping-business-cites-inability-to-forecast-prices-11635883500. Accessed 10 Sept. 2023.

16. "Studying Community Vitality & Revitalization: Synthesizing, Case Study Research for Practical Application" (2012). Curtis, Jennifer and Cunningham, Mariah, Muskie School Capstones and Dissertations. 14. https://digitalcommons.usm.maine.edu/muskie_capstones/14
17. National Academies of Sciences, Engineering, and Medicine. 2020. Review of the New York City Watershed Protection Program. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25851>.
18. Distressed Community Index. Economic Innovation Group, 2017, eig.org/wp-content/uploads/2017/09/2017-Distressed-Communities-Index.pdf. Accessed 10 Sept. 2023.
19. "Key: highway." Key: highway, Open Street Map, [wiki.openstreetmap.org/wiki/Key: highway](http://wiki.openstreetmap.org/wiki/Key:highway). Accessed 10 Sept. 2023.
20. "Community Index." Fourth Economy Community Index, Fourth Economy, fourtheconomy.shinyapps.io/Community_Index. Accessed 10 Sept. 2023.
21. Flippen, Alan. "Where Are the Hardest Places to Live in the U.S.?" The Upshot: By Numbers, edited by Alan Flippen, NYTimes, 26 Apr. 2014, www.nytimes.com/2014/06/26/upshot/where-are-the-hardest-places-to-live-in-the-us.html#. Accessed 13 Sept. 2023. The source that the Indiana CVI used for its data and general approach to CVI.
22. Office of Commission Auditor. 3 Mar. 2023, www.miamidade.gov/govaction/legistarfiles/Matters/Y2023/230464.pdf. Accessed 15 Sept. 2023.