Cities 140 (2023) 104432

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit, and density factors on single-family housing price

Yefu Chen [a], Junfeng Jiao [a,*], Arya Farahi [b]

[a] Community and Regional Planning, University of Taxes at Austin, United States of America
[b] Department of Statistics and Data Sciences, University of Taxes at Austin, United States of America

ARTICLE INFO

Keywords:
Housing price
Built environment
Neighborhood attributes
Machine learning

ABSTRACT

Profound insights have been gained into which characteristics determine housing prices. These characteristics reflect two different aspects: those which are correlated with the dwelling itself and those which are correlated with the location and the surrounding area. However, few studies precisely looked at the disparities and heterogeneity in these effects across neighborhoods with varied conditions. Also, there lacks studies focusing on the moderate-density cases where housing markets have drawn concerns recently. This study aims to fill this research gap by analyzing these disparities across neighborhoods with different economic and racial/ethnic conditions. Through machine learning approaches, we compare the disparities in the impacts of housing status, public transit services, and surrounding environment factors under seven conditions. Results indicate that the heterogeneity in economic conditions could be more significant than racial/ethnical conditions. Through comparison analysis, we call policymakers to need to adopt differentiated perspectives on housing price analysis, and future studies should consider the disparities in the impacts across neighborhoods.

1. Introduction

Housing is regarded as an essential factor in daily human life. On the one hand, owning a house can provide the proudness and happiness to individuals (Rohe et al., 2013). On the other hand, it may lead to the stress of repayment of loans, forcing buyers to spend less on other expenditures, such as education and medicine, when the houses are not affordable. Thus, keeping the good part of owning a house should be a critical theme for local authorities and researchers. They have explored the effects on housing affordability and developed myriad practical toolkits to affect housing values based on empirical studies. Based on their findings, housing status, public transit services, and the surrounding environment are regarded as powerful factors in housing values (Diao & Ferreira, 2010; Guo et al., 2016; Zhong & Li, 2016).

Although the important factors have been identified, there are demands for further studies focusing on moderate population density cases with appropriate investigations. There are research gaps worth noting. First, most previous studies focused on cases with high population density (Boslett et al., 2016; Diao & Ferreira, 2010; Wang et al., 2022), while lacking empirical studies on moderate-density cases. However, during the pandemic, there is a significant in-migration from the high population density areas, such as the East and West Coasts of the United States, to the moderate-density cases, raising concerns about the affordability in these areas (Florida, 2022; Sandoval, 2021). In addition, variations in the effects of built environment and public services on housing values have been observed between areas with different density statuses (Aurand, 2010; Dunse et al., 2013; Wu et al., 2018). In high-density areas, the accessibility of amenities like grocery stores, restaurants, and green spaces has a greater impact on housing values compared to moderate-density areas. This is because residents in high-density areas are more sensitive to the distance they have to travel to access these services. In contrast, residents in moderate-density areas are more accustomed to longer distances and tend to rely on driving, making larger catchments for these amenities (Gim, 2012; Holz-Rau et al., 2014; Richardson et al., 2012). Also, recent studies have noted the nonlinearity impacts on housing values (Gan et al., 2021; Rehman et al., 2020). They developed models to examine the global effects, but the disparities in effects across conditions of neighborhoods could be worth further consideration. Therefore, it is essential to examine the disparities in the impacts of factors affecting housing values in a moderate-density case to

* Corresponding author.
E-mail addresses: chenyf56@utexas.edu (Y. Chen), jjiao@austin.utexas.edu (J. Jiao), arya.farahi@austin.utexas.edu (A. Farahi).

https://doi.org/10.1016/j.cities.2023.104432
Received 13 November 2022; Received in revised form 18 May 2023; Accepted 4 June 2023
Available online 10 June 2023

fill these research gaps.

The City of Austin, Texas, can be a typical case of this study. Austin is regarded as a moderate-density city with around three thousand people per square mile, which is lower than famous cities on the East and West Coasts (City of Austin, 2021). Second, the housing affordability concerns in Austin are severe. According to the Census Bureau, the population reached 96,5872 in 2020, which increased by 1.5 % from 2019 and 22.2 % from 2010, higher than the average population increase rate in the United States. This significant population growth led to the increasing demand for housing and the cost of owning a house. However, the real estate market may fail to provide affordable housing. From 2019 to 2021, the average housing price increased by 12.8 %, while the number of housing sold declined by 15.5 % year-over-year (Redfin, 2022).

Austin authorities have worked on interventions considering the condition of the recent local real estate market. They provided bond and homestead exemption to financially support the house buyers (Eubank, 2021). Also, Austin launched a comprehensive transit-oriented development plan, Project Connect, which can enlarge the housing supplies (Smalley, 2021). However, these efforts are criticized. Critics mentioned that the factors in housing prices should be varied across neighborhoods with different sociodemographic characteristics (Heyman & Sommervoll, 2019; Toussaint-Comeau & Lee, 2018; Zhao & Ke, 2021), which are lacked in the plan. Given the history in Austin, the sociodemographic disparity in effects on housing values should be worth considering (Sullivan, 2018).

Regarding the research gaps and demands for study in Austin, we conducted a study in Austin to address existing research gaps and the need for housing studies. Our objective was to analyze the variations in factors that influence housing values across different neighborhood conditions using machine learning techniques. The study aims to contribute to the understanding of how housing status, public transit service, and built environment factors impact residential property values. Specifically, we focused on two research questions: 1) What is the overall impact of housing status, public transit, and density on single-family housing transaction prices? 2) How do housing status, public transit, and density affect housing prices in specific neighborhood conditions? Furthermore, our study has three notable contributions. First, it examines a case of moderate density, which has received limited attention in previous research. Second, we employ emerging machine learning algorithms to explore nonlinearity, providing a framework for future studies. Third, we investigate the disparities in house valuation factors, which can assist researchers and practitioners in developing more robust and accurate models for housing valuations. The remainder of the article is structured as follows. Section 2 presents a literature review of previous studies and research gaps. Section 3 describes the data, methodology, and analytic framework. Section 4 mentions the results of model selection and models in detail. In Section 5, we discussed our findings and policy implication suggestions.

2. Literature review

2.1. Models and factors in single-family housing values

The Hedonic Price Model (HPM) is a widely used method for estimating the value of residential properties (Cao & Wei, 2010; Funderburg & MacDonald, 2010; Rosen, 1974). According to the initial definition, HPM is based on the willingness of buyers to spend the extra money to satisfy themselves in the housing decision-making process, which causes the premium to change to properties in neighborhoods with hedonic characteristics. Based on findings, studies argued that HPM could be established by combining hedonic values, including housing status, neighborhood attributes, and locations (Gibbs et al., 2018; Huh & Kwak, 1997; Jones, 1988; Zhong & Li, 2016). The conceptual form of HPM is:

$$P = f(H, N, L, e)$$

where P refers to the value of a residential property. H refers to the housing status. N refers to the neighborhood attributes. L refers to the location information, and e refers to the error terms.

Recently, machine learning approaches have been considered in the HPM to address the limitations of traditional models, including low estimation accuracy and multicollinearity (Binoy et al., 2021). Previous studies have shown efficiency of machine learning approaches in housing studies, and they claimed the increases in accuracy and robustness of machine learning approaches to traditional models (Mullainathan & Spiess, 2017; Park & Bae, 2015; Truong et al., 2020). In addition, a study highlighted the importance of choosing the most suitable machine learning approach for modeling (Ho et al., 2021). They compared the performance of three approaches, namely, support vector machine, random forest, and gradient boosting machine, and argued that the comparison between performances of machine learning approaches should be critical.

2.2. Disparities in impacts on housing values

There have been numerous studies focusing on the housing valuation factors in the high-density areas; however, few specifically noted the situation in the moderate-density and low-density cases (Been et al., 2016; Binoy et al., 2021; Boslett et al., 2016; Saphores & Li, 2012; Shen et al., 2018). Meanwhile, the importance of factors (i.e., hedonic characteristics) on housing valuation should be different for the units in high-density and other cases. The importance of factors (i.e., hedonic characteristics) in housing valuation differs between high-density areas and other cases. In high-density areas, factors such as walkability, accessibility to amenities, and the presence of surrounding areas play a critical role since high-density residents place value on the ability to walk to shops, restaurants, and public transportation, considering them as important hedonic characteristics (Park et al., 2017; Wu et al., 2018). Additionally, the surrounding areas, including the neighborhood socioeconomic and physical characteristics, are considered hedonic characteristics in high-density areas (Grundström & Molina, 2016; Wassmer & Baass, 2006). On the other hand, in moderate-density and low-density areas, residents often prioritize personal space, natural surroundings, and suburban amenities. These residents prefer houses with larger lots, yards, and gardens, which provide more personal outdoor space, and proximity to parks, open green spaces, and natural areas is highly valued for recreational activities and a connection to nature (Cho et al., 2008; Dehring & Dunse, 2006; Marcus & Sarkissian, 1986). Moreover, suburban amenities like community facilities and car-dependent infrastructures play more significant roles in these areas compared to high-density areas since driving personal vehicle is more related to their daily travel behaviors (Cao, 2009; Dunse et al., 2013; Matthews & Turnbull, 2007).

Previous studies have noted the sociodemographic disparities in the impacts on housing values. The initial HPM indicates that housing values are determined by attributes, including location, housing status, year of built, amenities, and accessibility to public services, while it is worth noting that the disparities in the willingness of buyers to purchase attributes could be different (Bayer et al., 2007; Harrison & Rubinfeld, 1978). Harrison & Rubinfeld studied the impacts of air pollution and housing prices across neighborhoods with different economic conditions. The results indicated that high-income, higher than the local middle income, residents were willing to spend more money on reducing air pollution than low-income residents, pointing out a significant heterogeneity in neighborhood economic conditions affected how residents valued environmental attributes. Consistent with the results of Harrison & Rubinfeld, Bayer et al. found that the disparities in willingness for the performance of school districts can be different. Based on the census data, they highlighted that the high-income population was willing to pay for school performance, while the neighborhood with vulnerable populations refused to invest in education.

The spatial disparities in impacts on housing values are worth noting

(Been et al., 2016; Bitter et al., 2007; Song et al., 2019). Through geographic analysis, Bitter et al. pointed out that housing attributes' marginal impacts can vary across neighborhoods, indicating significant spatial heterogeneity from the city-level perspective. Focusing on the community-level perspective, Been et al. noted the significant spatial disparities in the impacts of historic districts in New York City and argued that the property values in Manhattan could increase when being defined as within historic districts; however, for those outside of Manhattan, only the properties locating on the boundaries can be associated with a higher price. In addition, Song et al. studied the spatial disparities of willingness to pay for dwelling in Beijing and found that housing status can have different impacts on housing values in different districts of the city. Furthermore, not only the disparities in attributes should be significant, but the heterogeneity in measurements of attributes can be worth noting (Osland et al., 2022). Osland et al. utilized GIS-derived data to examine the disparities and claimed that it should be significant across neighborhoods. They also argued that different measurements, such as the distance to and dominance of land use and public services, should have various importance across neighborhoods.

To sum up, it is important to investigate the housing valuation factors in the moderate-density case since it could be different than the findings which are mostly based on high-density case. Also, machine learning approaches are regarded as a potent method in the HPM of housing studies, and the disparities in impacts of attributes on housing values have been investigated as significant. However, few studies combined machine learning and heterogeneity, and most focused on high-density cases, such as New York City and Beijing, lacking attention to the moderate-density cases. Considering the sufficient of machine learning approaches and significant disparities, we applied five machine learning algorithms and examined the impacts of housing attributes on values focusing on six types of neighborhoods: high-income, low-income, high rates of white-only, high rates of Hispanic, high rates of African American, and general vulnerable neighborhoods.

3. Methods

3.1. Study area

We chose the City of Austin as the study area. Austin locates in the center of Texas. It is the county seat of Travis County and the capital city of Texas. Fig. 1 presents the locations of the Texas State Capitol, the city of Austin, and Travis County. According to the Census Bureau, 96,5872 residents are living in Austin, and the median household income (in 2019 dollars) was 75,887 dollars. White-only, African American, and Hispanic or Latino make up 72.63 %, 7.83 %, and 33.32 % of the population in our study area, respectively.

3.2. Data

Table 1 shows the data description. First, we obtained housing price

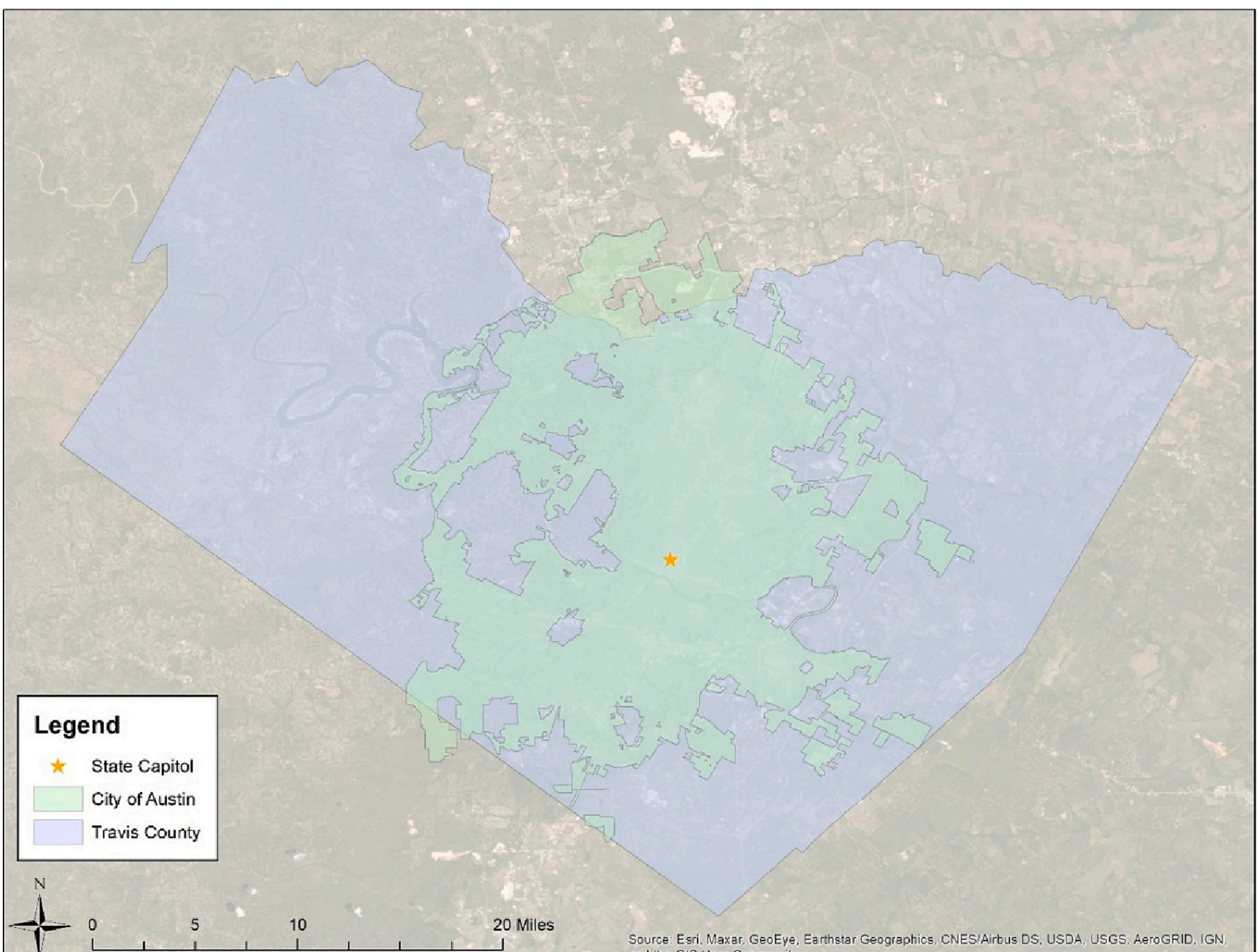

Fig. 1. Study area location.

Table 1
Data description (n = 25,135).

Variables	Mean (S.D.)/ count (%)	Data source
Outcome variable		
Sold price per square foot [numeric]	185.77 (70.15)	Austin realtor data
Explanatory variable: housing status		
Year built [numeric]	1987.25 (20.56)	Austin realtor data
Type [categorical]		
Resale	20,585 (81.90 %)	
Updated/remodeled	2584 (10.28 %)	
Under construction	656 (2.61 %)	
New	1264 (5.03 %)	
To be built	46 (0.18 %)	
Stories [numeric]	1.47 (0.53)	
CDOM [numeric]	47.78 (71.42)	
Occupant type [categorical]		
Vacant	11,302 (44.97 %)	
Owner	12,624 (50.22 %)	
Tenant	1209 (4.81 %)	
Property conditions [numeric]		
Poor (1)	710 (2.82 %)	
Fair (2)	1750 (6.96 %)	
Average (3)	2963 (11.79 %)	
Good (4)	9425 (37.50 %)	
Excellent (5)	10,287 (40.93 %)	
Foundation [categorical]		
Slab	22,778 (90.6 %)	
Pier & Beam	1682 (6.69 %)	
Others	675 (2.69 %)	
Explanatory variable: public transit		
Number of public transit routes [numeric]	2.90 (3.95)	Austin open portal
Number of public transit hubs [numeric]	0.02 (0.13)	
Number of public transit stops [numeric]	10.47 (14.28)	
Explanatory variable: surrounding environment		
Population density (per square mile) [numeric]	3515.05 (2700.17)	Austin open portal and American Community Survey
Household density (per square mile) [numeric]	1377.93 (1071.77)	
Residential road network (mile) [numeric]	16.66 (4.28)	
Distance to highway (mile) [numeric]	1.42 (1.56)	
Number of new permits [numeric]	10.07 (19.73)	
Controlling variable: time		
Timestamp [categorical]		Austin realtor data
2014	8579 (24.13 %)	
2015	8514 (23.96 %)	
2016	8042 (22.63 %)	
2017	8433 (23.73 %)	
Controlling variable: location		
Distance from housing to the state capital (mile) [numeric]	8.05 (3.49)	Austin open portal
Longitude [numeric]	[a]	
Latitude [numeric]	[a]	
Controlling variable: regional status		
Travis county population (in million) [numeric]	1.18 (0.02)	Travis County
Travis county GDP [numeric]	96.01 (6.71)	
Controlling variable: sociodemographic attributes		
Median household income [numeric]	85,108.3 (38,853.09)	American Community Survey
Rates of white only [numeric]	0.79 (0.13)	
Rates of Hispanic [numeric]	0.27 (0.09)	
Rates of African American [numeric]	0.06 (0.09)	

[a] Location information does not have statistical meaning; thus, we do not calculate the mean and standard deviation in the data description.

and housing status from the multiple listing services provided by the Austin realtors. In this study, we focused on single-family houses to avoid the impacts of different housing. However, the Austin realtors' raw dataset excludes the type of information. Thus, we referenced the Austin zoning code in 2015 and geographically intersected the sold housing and single-family residence to obtain the single-family housing transaction data. We acknowledged the mismatches of time points in 2014, 2016, and 2017 but assumed the bias should be minor since Austin did not grow rapidly at that time.

In the final dataset, we chose the sold price per square foot as the outcome variable. It was calculated as the total sold price divided by the livable area which is included in the housing status. Then, we used log transformation to standardize the outcome variable. Also, there are seven variables of housing status, including year built, types, stories, cumulative days on the market (CDOM), occupant type, property conditions, and foundation. Year built refers to when the residential property was built for the first time. Five types of sales, occupant types, and foundations are included in this study, and we treated them as the factor variable. However, since property conditions represent the level of properties, we treated it as numeric variables in the final modeling. In addition, the variable, stories, means the floor level in this single-family residential property, and CDOM refers to the listing activities. We also introduced the foundation type in the analysis. The raw data include slab, pier & beam, on stilts, and slab, among others, and we recategorized this factor into three types: slab, pier & beam, and others due to the amount of each category.

In addition, we considered the impacts of public transit services in this study since these factors were proven to be a strong predictor of housing valuation (Diao & Ferreira, 2010; Kim & Zhang, 2005; Orford, 2017; Shen et al., 2018). The public transit services from 2014 to 2017 is captured from the OpenMobilityData (OpenMobilityData, n.d.). We used 0.6-mile Euclidean buffers to capture the public transit factors, referencing previous studies (McGinn et al., 2007). Three public transit factors, namely, the number of public transit routes, the number of public transit hubs, and the number of public transit stops, are included.

The impacts of the surrounding environment, including new construction permits, road network, distance to highways, and density factors, are considered. First, we introduced the issued construction permits from 2014 to 2017 and captured the new and building construction permits in measuring the construction effects since they can present the local real estate market, which should play a significant role in housing values (Freemark, 2020; Glaeser & Gyourko, 2006). The amount of construction permits within 0.6 miles is calculated as the number of new permits. In addition, road network refers to the length of the road excluding highways within a 0.6-mile buffer surrounding properties, and distance to highways represents the Euclidean distance from properties to the closest highways. Both variables were collected from the 2015 database since data at other time points was not accessible. We acknowledge the limitation in the data mismatch but claim it as minor. In addition, density factors from the annual census report are calculated as the population or households divided by the area of census block groups, representing the neighborhood-level population and household density of observations.

Controlling variables are worth noting. First, we introduced the timestamp, when the property was sold, as the time-controlling variable. Also, we controlled the regional status, regional population, and gross domestic product to control the regional real estate market effects (Chen & Haynes, 2015; Davis, 2011; Tsatsaronis & Zhu, 2004). Previous studies pointed out that the centrality of houses should be a significant factor (Alonso, 1964). Accordingly, we introduced the distance to the Texas state capitol as the factor representing centrality in general. The detailed location factors, latitude, and longitude are considered for controlling the geographic heterogeneity (Song & Knaap, 2004). In addition, economic and sociodemographic factors from the 2014 to 2017 American Community Survey, including median household income, rates of white only, rates of Hispanic Americans, and rates of African Americans, are introduced in this study. The reason for choosing white only, Hispanic, and African Americans is that they are the three major racial/ethnical groups in Austin. Although this data is at the neighborhood level, which cannot directly represent the buyer's willingness, it has been used as the alternative when the buyers' information is inaccessible, which frequently happens (Saphores & Li, 2012; Spielman et al., 2014).

3.3. Methodology and analytic framework

This section presents the details concerning the algorithms and research process used in the study. First, we compared the performance of five machine learning approaches, namely, X-gradient boosting decision trees (XGBDT), random forests (RF), elastic net, lasso regression, and ridge regression, to find the best model fitting our data. Among these approaches, XGBDT and RF are ensemble decision tree models with performance advantages for large datasets (Afonso et al., 2019; Revend, 2020). Lasso and ridge regression techniques have been considered in previous studies (Lu et al., 2017; Madhuri et al., 2019). Lasso regression shows abilities in interpretations and robustness, given that it can automatically select variables. The advantage of ridge regression is robust performance in dealing with multicollinearity issues. The elastic net, regarded as the advanced model that combines the benefits of lasso and ridge regression techniques, has proven potential in housing valuations (Gabauer et al., 2020). All the models were developed in Python 3.7 through "statsmodels" and "sklearn" packages (Scikit-learn, n.d.; statsmodels, n.d.). We used the grid search for each model to determine the best combination of hyperparameters.

The datasets were divided into two groups, namely, the training dataset (from 2014 to 2016) and the test dataset (2017), based on when the properties were sold to examine the model performance. We also calculated the root mean square error (RMSE) to evaluate the fitting and predicting performance (Barnston, 1992). The equation of RMSE is given by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \left(z_{pi} - z_{ti}\right)^2 \Big/ n}$$

where z_{pi} refers to the predicted value i, and z_{ti} represents the actual value i. n equals the number of the observations.

Table 2 presents the RMSE on the training dataset and the RMSE using the training model to predict the test dataset. We found that

Table 2
Model performance.

	RMSE on training dataset	RMSE on predictions
XGBDT	0.115	0.154
RF	0.166	0.316
ElasticNet	0.189	0.268
Lasso regression	0.258	0.237
Ridge regression	0.190	0.251

XGBDT performed best in the training dataset and the prediction. Therefore, we used XGBDT as the primary algorithm in this study.

After the model performance was compared, we used XGBDT to build the full model. Also, in the results section, we only focused on the results of training data since their hyperparameters have been proven robust. In addition, we constructed six sub-models. The descriptions are given as follows:

1. High-income neighborhoods: single-family property transactions happened in census block groups where median household income was higher than the regional median ($75,887).
2. Low-income neighborhoods: single-family property transactions happened in census block groups where median household income was lower than the regional median ($75,887).
3. High rates of white-only neighborhoods: single-family property transactions happened in census block groups where rates of the white-only population were higher than the regional average (0.79).
4. High rates of Hispanic American neighborhoods: single-family property transactions happened in census block groups where rates of Hispanic population were higher than the regional average (0.33).
5. High rates of African American neighborhoods: single-family property transactions happened in census block groups where rates of African American population were higher than the regional average (0.09).
6. Vulnerable neighborhoods: single-family property transactions happened in census block groups with relatively high nonwhite-only rates (>0.29) and lower median household income (<$75,887).

Each sub-model takes all variables except the geography information (longitude and latitude) and sociodemographic attributes. The reason for this process is that we assign neighborhood backgrounds based on the census block group data in which the sociodemographic attributes variables are from. Data from the census block group can also have multicollinearity to the geography information data. We dropped them and executed the abovementioned process to avoid collinearity.

We adopted the "SHAP" method to interpret the results. SHAP (shapely additive explanations) is a game-theoretic, inspired method developed to make machine-learning models explainable (Chakraborty et al., 2021; Gardiner et al., 2021; Lundberg et al., 2020; Lundberg & Lee, 2017). It utilizes the classic Shapley values from game theory to locally explain the contribution of each factor to the output of a predictive model. SHAP values measure the importance of each factor in model output (Mazzanti, 2021). The equation is given as the follow (Chen et al., 2020).

$$SHAP_{feature}(x) = \sum_{\text{set:feature}\in\text{set}} \frac{|\mathrm{A}|!(\mathrm{p}-|\mathrm{A}|-1)!}{\mathrm{p}!}\left(\mathrm{val}(\mathrm{A}\cup\{\mathrm{x}\}) - \mathrm{val}(\mathrm{A})\right)\Big)$$

where p is the number of features. A represents the subset of the feature. x is the vector of feature values of an instance to be explained. $val(A)$ is the prediction for feature values in the set A.

4. Results

4.1. Global impacts of housing status, public transit, and density

Table 3 presents the average absolute SHAP values of the global model and sub-models, and Fig. 2 is the beeswarm summary plot of the global model. Focusing on the global model, Table 3 indicates that the location factors have the most collective predictive power, followed by housing status, indicating possible geographic heterogeneity affecting Austin housing values. Also, time stamps and housing status (stories, property conditions, and year built) are essential to predict housing values. In addition, the rate of Hispanics is the only important factor of sociodemographic attributes, and public transit routes are on the list. It is interesting noting that public transit is significant in predicting

Table 3
Average absolute SHAP values of top nine factors in the global model and sub-models.

	Global model	Sub-models					
	(n = 25,135)	High-income neighborhoods (n = 13,242)	Low-income neighborhoods (n = 13,517)	High rates of White-only neighborhoods (n = 16,076)	High rates of Hispanic American neighborhoods (n = 4330)	High rates of African American neighborhoods (n = 6777)	Vulnerable neighborhoods (n = 7558)
Explanatory variable: housing status	0.10	0.11	0.14	0.09	0.15	0.14	0.14
Year built	0.02	0.03	0.04	0.03	0.04	0.03	0.04
Type							
Stories	0.05	0.04	0.05	0.03	0.06	0.06	0.05
CDOM		0.02					
Occupant type							
Property conditions	0.03	0.02	0.05	0.03	0.05	0.05	0.05
Foundation							
Explanatory variable: public transit	0.02		0.09	0.08	0.08	0.08	0.10
Number of public transit routes	0.02		0.07	0.04	0.03	0.02	0.03
Number of public transit hubs							
Number of public transit stops			0.02	0.04	0.05	0.06	0.07
Explanatory variable: surrounding environment		0.08	0.13	0.06	0.08	0.09	0.10
Number of new permits		0.02					
Population density		0.04	0.05	0.03	0.04	0.02	0.03
Household density		0.02	0.08	0.03	0.04	0.07	0.07
Residential road network							
Distance to highway							
Controlling variable: Timestamp	0.03	0.02	0.04	0.02	0.04	0.03	0.04
Controlling variable: location	0.25	0.15	0.15	0.14	0.13	0.14	0.12
Distance from housing to the state capital	0.11	0.15	0.15	0.14	0.13	0.14	0.12
Longitude	0.08						
Latitude	0.07						
Controlling variable: regional status							
Travis county population							
Travis county GDP							
Controlling variable: sociodemographic attributes	0.04						
Median household income							
Rates of white only							
Rates of Hispanic	0.04						
Rates of African American							

housing values in our moderate-density case, which is similar to the results of previous studies (Shen et al., 2018).

Fig. 2 focuses on the top nine important factors and presents an information-dense summary of the impact on the output of the global model. It indicates that further from the Texas capital can decrease housing values, and the marginal effects decrease when the distance keeps increasing. This finding is consistent with previous studies (Kim & Zhang, 2005; Söderberg & Janssen, 2001).

Important factors of housing status are worth noting. Properties with more stories could be associated with lower transaction prices. Considering most of the houses in the final dataset have one and two stories, we argued that the prices of story single-family residential properties could be larger than those with two stories. In addition, the transaction prices of houses with excellent conditions could be larger than those with poor or fair conditions. However, this result needs further study since the distribution in property condition is not balanced. Finally, the impacts of the year of housing built are complicated, indicating the year built was not the priority concern of Austin housing buyers from 2014 to 2017.

The rates of Hispanic and the number of transit routes play a significant role in predicting housing values. In general, houses in neighborhoods with high Hispanic rates are associated with lower transaction prices, while transit routes could drive housing values. In addition, the nonlinearities in both impacts are significant. The middle value could be the threshold of Hispanic impacts, and the effect of public transit services turns insignificant when the number is low.

4.2. Impacts of housing status, public transit, and density across neighborhoods

Table 3 presents the sub-model results. First, focusing on the

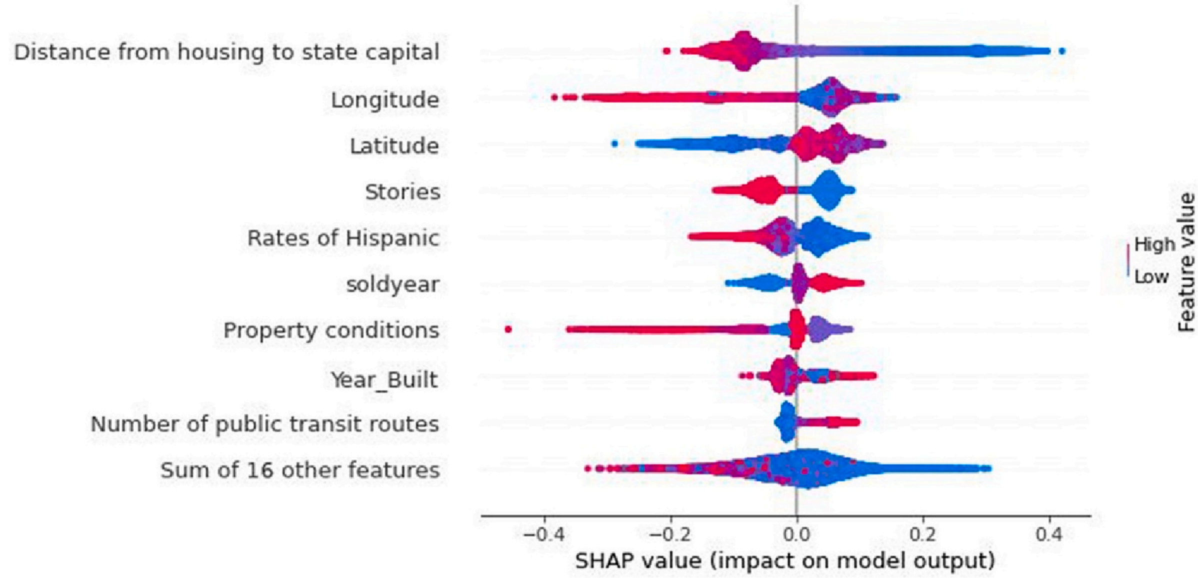

Fig. 2. SHAP values of the global model result.

significance, we found common important factors in predicting housing values, including year built, stories, property conditions, transit routes, population density, household density, timestamps, and distance to the state capital. However, CDOM and the number of permits are only regarded as important predictors, and transit stops are regarded as an insignificant factor in the high-income sub-model, while the significances of variables across the neighborhood with different sociodemographic attributes are similar. It could indicate that the heterogeneity in

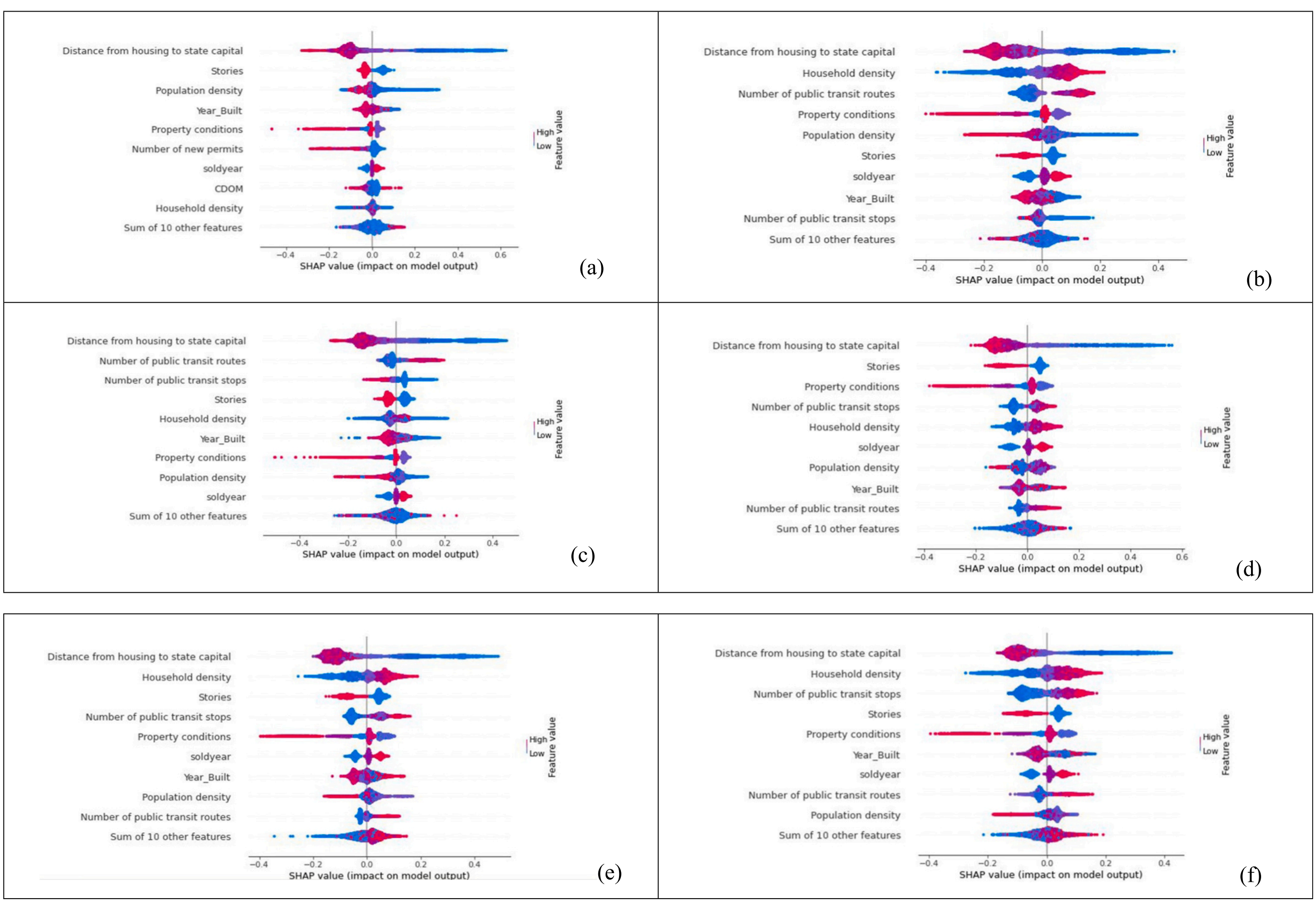

Fig. 3. SHAP values of the sub-models result.
*(a) high-income neighborhoods; (b) low-income neighborhoods; (c) high white-only rate neighborhoods; (d) high Hispanic American rate neighborhoods; (e) high African American rate neighborhood; (f) high vulnerable population rate neighborhoods.

the neighborhood with different economic conditions is more significant than between demographic conditions, which can be the support for researchers starting at the economic disparities (Harrison & Rubinfeld, 1978; Jin et al., 2022). Second, considering the value of significance, we concluded that low-income buyers might put more value on housing status and the surrounding environment when they decide to purchase houses, while white-only buyers may value less on them. In addition, the differences in significance values between minority groups are insignificant. Both value more on housing status and the surrounding environment more. Moreover, houses within vulnerable neighborhoods put more value on housing status, public transit, and the surrounding environment. This result is similar to the high-density case studies (Been et al., 2016; Binoy et al., 2021; Saphores & Li, 2012), but it is also worth noting that locations and housing status are two principal predictors, while the roles of other factors are relatively less critical in predicting housing values.

Comparing the beeswarm summary plots of high-income and low-income sub-models (Fig. 3 (a and b)), we found three significant differences. First, the range of SHAP values of sold year in the high-income model is smaller than in the low-income model, indicating that the willingness to purchase a house in the low-income neighborhoods is sensitive to changes in time. Second, the trend of household density is unclear in high-income neighborhoods, while it plays a positive role in low-income neighborhoods. It can be evidence that low-income buyers considered high-dense properties, like the apartment, as the driver of housing values. Third, public transit services are significant drivers in housing values in low-income neighborhoods, while it is insignificant in high-income areas, indicating that these facilities may raise the housing prices in low-income neighborhoods and lead to gentrification and displacement (Chapple & Zuk, 2020). Except for the differences in significance, it is worth noting that CDOM is significant only in high-income neighborhoods, implying that the impacts of time on the market are significant on housing prices in high-income neighborhoods. In most cases, extended time on the market can lead to relatively low, while some contrast cases are worth noting. For example, some single-family houses have red (high) CDOM but positive SHAP. We found these houses relatively costly (>$221.93 per square foot).

Comparing the results across sociodemographic attributes (Fig. 3 (c–e)), first, we found that the impacts of transit routes are constantly positive, while the disparities in transit stops are worth noting. In the minority gathering neighborhoods, its impact is positive but turns negative in the white-only neighborhoods. It indicates that minorities could rely more on public transit services than white-only (Tan et al., 2020). Second, similar to the low-income neighborhoods, the impact of household density is positive in minority-gathering neighborhoods, while its trend turns unclear in white-only neighborhoods. In addition, it is worth noting that the impacts of household density and population density are different in the minority-gathering neighborhoods, which may indicate that minorities prefer to live in neighborhoods with more households but are worried about the household size. Too large households could decline their willingness to purchase a house in minority-gathering neighborhoods.

Lastly, it is worth focusing on the vulnerable neighborhood model (Fig. 3 (f)). Both impacts of transit stops and routes are positive. It can support the claim that vulnerable populations rely on public transit (Kotval-K et al., 2023). In addition, the difference between impacts of household density and population density could be explained by the same reasons of minority-gathering neighborhoods. Given the definition of vulnerable neighborhoods, it is no surprise that the results are similar to the low-income and minority-gathering neighborhoods.

5. Conclusion remarks

This study aims to explore the sociodemographic and economic disparities in impacts on housing values through machine learning approaches and provide an empirical study focusing on the moderate-density case to advance knowledge in housing valuations. To this end, a housing valuation model was developed by extracting a list of essential housing attributes from existing literature. Then, we identified the best model for this study and examined the impacts of housing status, public transit, and surrounding environment attributes on housing prices and precisely studied the differences across neighborhoods by answering two research questions. The findings of this study may shed greater light on understanding the behavioral aspects of the valuation process, as well as the importance of individual differences that could apply to broader settings such as different markets or diverse types of residential properties. They are also anticipated to contribute to the decision-making process on transit-oriented development and improvement of housing status.

Several takeaways from this study are worth noting. First, our study indicates that the disparities in the effects of public transit services are significant, consistent with those obtained in previous studies (Boyle et al., 2014; Zhao & Ke, 2021). In general, we found that public transit services should be significant in affecting housing values in low-income and minority-gathering neighborhoods, while it is insignificant in the high-income neighborhoods. Currently, the transit-oriented-development is widely accepted as a toolkit to expand housing market and provide affordable transit services (Shen et al., 2018), and there is an ambitious plan in Austin (Thompson & Sanders, 2022).

Although the city council have approved millions of dollars for anti-displacement, we encourage policymakers to precisely allocate the budget and facilities. In general, housing values of neighborhoods, except for the high-income, in Austin are sensitive to public transit improvement, and the amount of public transit routes is one driver. Since Project Connect plan to increase the bus routes, both regular bus and bus rapid transit, in Austin, the housing values within 0.6-mile corridor surrounding the increased bus lines should increase. To keep the Austin affordable, there should be financial supports implemented to anti-displacement. In addition, based on our findings, the demand of public transit services, both routes and stops, in low-income and minority-gathering neighborhoods is significant, while the house values in these neighborhoods are much more sensitive to public transit improvement than others. Thus, it would be better to locate transit services in these neighborhoods with extra financial supports (Anti-displacement network, n.d.).

Second, the disparities in valuing attributes among economic and sociodemographic neighborhoods are significant, and we further argue that the disparity in economic conditions is more significant than demographic conditions. We noted that previous studies focused on the economic disparity while recent studies concerning about the racial disparity (Harrison & Rubinfeld, 1978; Roberts et al., 2022). We encourage researchers to combine both and to examine the integrated disparities to provide a holistic result. In addition, we acknowledge the limitations in this study and call for research with better database and analytic approaches to test our claim.

Third, the density factors present complexity in affecting housing values. There are two density factors introduced in this study, and their role in predicting housing values is different. We found that household density is regarded as a driver in house values of most sub-models except for the high-income; however, the role of population density is negative in most cases. This is quite an interesting finding since researchers claimed the negative impacts of density factors on housing values (Glaeser et al., 2005; Lee, 2016). Our findings suggest that household density, especially in the low-income and minority-gathering neighborhoods, should be regarded as the new-urbanism neighborhoods with high public service accessibility, like gentrified areas. Compared to where they used to live, although the total area of the livable place is more considerable than high household dense areas, the livability and property conditions may not satisfy the demands of residents. In contrast, population density could be a direct measure of crowding conditions. Therefore, we encourage local authorities not only to consider the public service accessibility in minorities neighborhood with high density but also to promote the services in low-density areas. Although this sounds inefficient, it may provide a choice for vulnerable populations to decide whether they should relocate. For researchers, it would be better to distinguish the differences in the impacts of household density and population density, and we call further studies to test the generalizability of this finding.

Moreover, the analytic framework presented in this study can be applied to other projects to help understand the effect of housing status and other factors on the local housing market. In this study, we argue that housing policies should focus on ethnic minority communities to improve their housing resilience through land use planning and capital investment, as well as other social supports such as public housing, tax cuts, and financial support, given that they are more sensitive to the changes. For instance, Austin is working on investing in outdoor green spaces to promote life qualities to all citizens (Spearman, 2021). Meanwhile, the impacts of these investments may be varied across neighborhoods, and overwhelming green phenomenon, that is, the displacement led by green space investment, may even occur (Wolch et al., 2014). Finally, we encourage relevant authorities to apply the analytic framework for providing a precise investigation of the impacts of green spaces on housing values.

The limitations of this study should be considered. First, the exploratory nature of this research implies that the results should be interpreted with caution. For instance, we used the distance to the state capital to represent the hedonic choice. It may be good for single-centroid cities, but it should be modified under complex situations like New York City. One should be cautious in generalizing our findings to other metropolitan cities of similar size due to the inhomogeneous urban environmental factors. Besides, future research could evaluate other factors (Guo et al., 2016; Heyman et al., 2019; Zulkifli et al., 2017). Second, we studied three races, namely, white-only, Hispanic, and African Americans, while ignoring the situations of other races. We called for future studies should take a further step to provide a holistic version of the disparities in the effects of housing status, public transit, and density across all racial and ethnic groups. Moreover, we only considered the regional median situations as the threshold to divide groups, while we acknowledged that the real situations could be complicated. Thus, future studies should apply a more precise model, such as quartile or quantile, to identify the disparities. Finally, this model is designed for residential single-family housing and thus needs to be modified and completed for other types of residential properties, such as detached or semi-detached houses.

CRediT authorship contribution statement

YC led the overall research idea, oversaw data analysis, modeling, paper drafting, and revision.

JJ developed the data analysis, data modeling, paper reviewing, and revision.

AF guided data analysis and modeling and helped with paper revision.

Declaration of competing interest

The authors declare there is no conflict of interest in the whole paper development process.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to acknowledge the funding supports from NSF (1952193 and 2133302), UT Good Systems, and USDOT.

References

Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. In *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)* (pp. 389–400). https://doi.org/10.5753/eniac.2019.9300

Alonso, W. (1964). *Location and land use* (Reprint 2013 ed. ed.). Harvard University Press.

Anti-displacement network. (n.d.). Home. Small Business Anti-Displacement Network (SBAN). Retrieved March 11, 2023, from https://antidisplacement.org/.

Aurand, A. (2010). Density, housing types and mixed land use: Smart tools for affordable housing? *Urban Studies, 47*(5), 1015–1036. https://doi.org/10.1177/0042098009353076

Barnston, A. G. (1992). Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score. *Weather and Forecasting, 7*(4), 699–709. https://doi.org/10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy, 115*(4), 588–638. https://doi.org/10.1086/522381

Been, V., Ellen, I. G., Gedal, M., Glaeser, E., & McCabe, B. J. (2016). Preserving history or restricting development? The heterogeneous effects of historic districts on local housing markets in New York City. *Journal of Urban Economics, 92*, 16–30. https://doi.org/10.1016/j.jue.2015.12.002

Binoy, B. V., Naseer, M. A., Anil Kumar, P. P., & Lazar, N. (2021). A bibliometric analysis of property valuation research. *International Journal of Housing Markets and Analysis, 15*(1), 35–54. https://doi.org/10.1108/IJHMA-09-2020-0115

Bitter, C., Mulligan, G. F., & Dall'erba, S.. (2007). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems, 9*(1), 7–27. https://doi.org/10.1007/s10109-006-0028-7

Boslett, A., Guilfoos, T., & Lang, C. (2016). Valuation of expectations: A hedonic study of shale gas development and New York's moratorium. *Journal of Environmental Economics and Management, 77*, 14–30. https://doi.org/10.1016/j.jeem.2015.12.003

Boyle, A., Barrilleaux, C., & Scheller, D. (2014). Does walkability influence housing prices? *Social Science Quarterly, 95*(3), 852–867.

Cao, M., & Wei, J. (2010). Valuation of housing index derivatives. *Journal of Futures Markets, 30*(7), 660–688. https://doi.org/10.1002/fut.20438

Cao, X. (2009). Disentangling the influence of neighborhood type and self-selection on driving behavior: An application of sample selection model. *Transportation, 36*(2), 207–222. https://doi.org/10.1007/s11116-009-9189-9

Chakraborty, D., Başağaoğlu, H., & Winterle, J. (2021). Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Systems with Applications, 170*, Article 114498. https://doi.org/10.1016/j.eswa.2020.114498

Chapple, K., & Zuk, M. (2020). Chapter 4 - Transit-oriented displacement: The role of transit access in the housing market. In E. Deakin (Ed.), *Transportation, land use, and environmental planning* (pp. 55–79). Elsevier. https://doi.org/10.1016/B978-0-12-815167-9.00004-9.

Chen, L., Yao, X., Liu, Y., Zhu, Y., Chen, W., Zhao, X., & Chi, T. (2020). Measuring impacts of urban environmental elements on housing prices based on multisource data—A case study of Shanghai, China. *ISPRS International Journal of Geo-Information, 9*(2), Article 2. https://doi.org/10.3390/ijgi9020106

Chen, Z., & Haynes, K. E. (2015). Impact of high speed rail on housing values: An observation from the Beijing–Shanghai line. *Journal of Transport Geography, 43*, 91–100. https://doi.org/10.1016/j.jtrangeo.2015.01.012

Cho, S.-H., Poudyal, N. C., & Roberts, R. K. (2008). Spatial analysis of the amenity value of green open space. *Ecological Economics, 66*(2), 403–416. https://doi.org/10.1016/j.ecolecon.2007.10.012

City of Austin. (2021). Austin by the numbers: City releases tract-level analysis of 2020 census data for Austin metro service area. https://www.austintexas.gov/news/austin-numbers-city-releases-tract-level-analysis-2020-census-data-austin-metro-service-area.

Davis, L. W. (2011). The effect of power plants on local housing values and rents. *The Review of Economics and Statistics, 93*(4), 1391–1402. https://doi.org/10.1162/REST_a_00119

Dehring, C., & Dunse, N. (2006). Housing density and the effect of proximity to public open space in Aberdeen, Scotland. *Real Estate Economics, 34*(4), 553–566. https://doi.org/10.1111/j.1540-6229.2006.00178.x

Diao, M., & Ferreira, J. (2010). Residential property values and the built environment: Empirical study in the Boston, Massachusetts, Metropolitan Area. *Transportation Research Record, 2174*(1), 138–147. https://doi.org/10.3141/2174-18

Dunse, N., Thanos, S., & Bramley, G. (2013). Planning policy, housing density and consumer preferences. *Journal of Property Research, 30*(3), 221–238. https://doi.org/10.1080/09599916.2013.795992

Eubank, B. (2021, November 30). Austin leaders discuss housing affordability, supply during special-called meeting. https://www.kvue.com/article/money/economy/boomtown-2040/austin-housing-affordability-supply-meeting/269-1470a21f-2461-4075-986f-dbec4bb751f2.

Florida, R. (2022, September 8). How the 'Rise of the Rest' Became the 'Rise of the Rents'. Bloomberg.Com. https://www.bloomberg.com/news/features/2022-09-08/why-did-housing-costs-explode-during-the-pandemic.

Freemark, Y. (2020). Upzoning Chicago: Impacts of a zoning reform on property values and housing construction. *Urban Affairs Review, 56*(3), 758–789. https://doi.org/10.1177/1078087418824672

Funderburg, R., & MacDonald, H. (2010). Neighbourhood valuation effects from new construction of low-income housing tax credit projects in Iowa: A natural experiment. *Urban Studies, 47*(8), 1745–1771. https://doi.org/10.1177/0042098009356122

Gabauer, D., Gupta, R., Marfatia, H., & Miller, S. M. (2020). *Estimating U.S. housing price network connectedness: Evidence from dynamic elastic net, lasso, and ridge vector autoregressive models (SSRN scholarly paper no. 3660950)*. Social Science Research Network. https://doi.org/10.2139/ssrn.3660950

Gan, L., Ren, H., Xiang, W., Wu, K., & Cai, W. (2021). Nonlinear influence of public services on urban housing prices: A case study of China. *Land, 10*(10), Article 10. https://doi.org/10.3390/land10101007

Gardiner, L.-J., Rusholme-Pilcher, R., Colmer, J., Rees, H., Crescente, J. M., Carrieri, A. P., … Hall, A. (2021). Interpreting machine learning models to investigate circadian regulation and facilitate exploration of clock function. *Proceedings of the National Academy of Sciences, 118*(32), Article e2103070118. https://doi.org/10.1073/pnas.2103070118

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing, 35*(1), 46–56. https://doi.org/10.1080/10548408.2017.1308292

Gim, T.-H. T. (2012). A meta-analysis of the relationship between density and travel behavior. *Transportation, 39*(3), 491–519. https://doi.org/10.1007/s11116-011-9373-6

Glaeser, E. L., & Gyourko, J. (2006). *Housing dynamics (working paper no. 12787)*. National Bureau of Economic Research. https://doi.org/10.3386/w12787

Glaeser, E. L., Gyourko, J., & Saks, R. E. (2005). Why have housing prices gone up? *American Economic Review, 95*(2), 329–333. https://doi.org/10.1257/000282805774669961

Grundström, K., & Molina, I. (2016). From Folkhem to lifestyle housing in Sweden: Segregation and urban form, 1930s–2010s. *International Journal of Housing Policy, 16*(3), 316–336. https://doi.org/10.1080/14616718.2015.1122695

Guo, Y., Agrawal, S., Peeta, S., & Somenahalli, S. (2016). Impacts of property accessibility and neighborhood built environment on single-unit and multiunit residential property values. *Transportation Research Record, 2568*(1), 103–112. https://doi.org/10.3141/2568-15

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management, 5*(1), 81–102. https://doi.org/10.1016/0095-0696(78)90006-2

Heyman, A. V., Law, S., & Berghauser Pont, M. (2019). How is location measured in housing valuation? A systematic review of accessibility specifications in hedonic price models. *Urban Science, 3*(1), Article 1. https://doi.org/10.3390/urbansci3010003

Heyman, A. V., & Sommervoll, D. E. (2019). House prices and relative location. *Cities, 95*, Article 102373. https://doi.org/10.1016/j.cities.2019.06.004

Ho, W. K. O., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research, 38*(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558

Holz-Rau, C., Scheiner, J., & Sicks, K. (2014). Travel distances in daily travel and long-distance travel: What role is played by urban form? *Environment and Planning A: Economy and Space, 46*(2), 488–507. https://doi.org/10.1068/a4640

Huh, S., & Kwak, S.-J. (1997). The choice of functional form and variables in the hedonic price model in Seoul. *Urban Studies, 34*(7), 989–998. https://doi.org/10.1080/0042098975691

Jin, T., Cheng, L., Liu, Z., Cao, J., Huang, H., & Witlox, F. (2022). Nonlinear public transit accessibility effects on housing prices: Heterogeneity across price segments. *Transport Policy, 117*, 48–59. https://doi.org/10.1016/j.tranpol.2022.01.004

Jones, L. E. (1988). The characteristics model, hedonic prices, and the clientele effect. *Journal of Political Economy, 96*(3), 551–567.

Kim, J., & Zhang, M. (2005). Determining transit's impact on Seoul commercial land values: An application of spatial econometrics. *International Real Estate Review, 8*(1), 1–26.

Kotval-K, Z., Wilkinson, A., Brush, A., & Kassens-Noor, E. (2023). Impacts of local transit systems on vulnerable populations in Michigan. *Urban Science, 7*(1), Article 1. https://doi.org/10.3390/urbansci7010016

Lee, J.-S. (2016). Measuring the value of apartment density? The effect of residential density on housing prices in Seoul. *International Journal of Housing Markets and Analysis, 9*(4), 483–501. https://doi.org/10.1108/IJHMA-08-2015-0047

Lu, S., Li, Z., Qin, Z., Yang, X., & Goh, R. S. M. (2017). A hybrid regression technique for house prices prediction. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 319–323). https://doi.org/10.1109/IEEM.2017.8289904

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), Article 1. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Madhuri, C. H. R., Anuradha, G., & Pujitha, M. V. (2019). House price prediction using regression techniques: A comparative study. In *2019 International Conference on Smart Structures and Systems (ICSSS)* (pp. 1–5). https://doi.org/10.1109/ICSSS.2019.8882834

Marcus, C. C., & Sarkissian, W. (1986). *Housing as if people mattered: Site design guidelines for the planning of medium-density family housing.* University of California Press.

Matthews, J. W., & Turnbull, G. K. (2007). Neighborhood street layout and property value: The interaction of accessibility and land use mix. *The Journal of Real Estate Finance and Economics, 35*(2), 111–141. https://doi.org/10.1007/s11146-007-9035-9

Mazzanti, S. (2021, April 21). SHAP explained the way I wish someone explained it to me. Medium. https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30.

McGinn, A. P., Evenson, K. R., Herring, A. H., & Huston, S. L. (2007). The relationship between leisure, walking, and transportation activity with the natural environment. *Health & Place, 13*(3), 588–602. https://doi.org/10.1016/j.healthplace.2006.07.002

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

OpenMobilityData. (n.d.). Capital Metro GTFS. Retrieved March 11, 2023, from https://transitfeeds.com/p/capital-metro/24?p=36.

Orford, S. (2017). *Valuing the built environment: GIS and house price analysis*. Routledge. https://doi.org/10.4324/9781315235134

Osland, L., Östh, J., & Nordvik, V. (2022). House price valuation of environmental amenities: An application of GIS-derived data. *Regional Science Policy & Practice, 14*(4), 939–959. https://doi.org/10.1111/rsp3.12382

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications, 42*(6), 2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040

Park, J. H., Lee, D. K., Park, C., Kim, H. G., Jung, T. Y., & Kim, S. (2017). Park accessibility impacts housing prices in Seoul. *Sustainability, 9*(2), Article 2. https://doi.org/10.3390/su9020185

Redfin. (2022). Austin housing market: House prices & trends. https://www.redfin.com/city/30818/TX/Austin/housing-market?utm_source=google&utm_medium=ppc&utm_term=aud-849690325664:kwd-938524868521&utm_content=456699713039&utm_campaign=1024480.

Rehman, M. U., Ali, S., & Shahzad, S. J. H. (2020). Asymmetric nonlinear impact of oil prices and inflation on residential property prices: A case of US, UK and Canada. *The Journal of Real Estate Finance and Economics, 61*(1), 39–54. https://doi.org/10.1007/s11146-019-09706-y

Revend, W. (2020). Predicting house prices on the countryside using boosted decision trees. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-279849.

Richardson, A. S., Boone-Heinonen, J., Popkin, B. M., & Gordon-Larsen, P. (2012). Are neighbourhood food resources distributed inequitably by income and race in the USA? Epidemiological findings across the urban spectrum. *BMJ Open, 2*(2), Article e000698. https://doi.org/10.1136/bmjopen-2011-000698

Roberts, M., Glenk, K., & McVittie, A. (2022). Urban residents value multi-functional urban greenspaces. *Urban Forestry & Urban Greening, 74*, Article 127681. https://doi.org/10.1016/j.ufug.2022.127681

Rohe, W. M., Zandt, S., & McCarthy, G. (2013). The social benefits and costs of homeownership: A critical assessment of the research. In *The affordable housing reader* (pp. 196–212).

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55.

Sandoval, E. (2021, November 27). How Austin became one of the least affordable cities in America. The New York Times. https://www.nytimes.com/2021/11/27/us/austin-texas-unaffordable-city.html.

Saphores, J.-D., & Li, W. (2012). Estimating the value of urban green areas: A hedonic pricing analysis of the single family housing market in Los Angeles, CA. *Landscape and Urban Planning, 104*(3), 373–387. https://doi.org/10.1016/j.landurbplan.2011.11.012

Scikit-learn. (n.d.). scikit-learn: Machine learning in Python. Retrieved May 15, 2022, from https://scikit-learn.org/stable/.

Shen, Q., Xu, S., & Lin, J. (2018). Effects of bus transit-oriented development (BTOD) on single-family property value in Seattle metropolitan area. *Urban Studies, 55*(13), 2960–2979. https://doi.org/10.1177/0042098017729078

Smalley, S. (2021, November 30). Project Connect agreement sees third and final approval. Austin Monitor. https://www.austinmonitor.com/stories/2021/11/project-connect-agreement-sees-third-and-final-approval/.

Söderberg, B., & Janssen, C. (2001). Estimating distance gradients for apartment properties. *Urban Studies, 38*(1), 61–79. https://doi.org/10.1080/00420980123880

Song, H., Wilhelmsson, M., & Zheng, M. (2019). Buyer's willingness to pay for dwellings with different orientations. *International Journal of Strategic Property Management, 23*(6), 450–467.

Song, Y., & Knaap, G.-J. (2004). Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics, 34*(6), 663–680. https://doi.org/10.1016/j.regsciurbeco.2004.02.003

Spearman, K. (2021, July 9). How Austin is investing in green spaces for future generations. Tribeza. https://tribeza.com/austin-green-spaces/.

Spielman, S. E., Folch, D., & Nagle, N. (2014). Patterns and causes of uncertainty in the American Community Survey. *Applied Geography, 46*, 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002

statsmodels. (n.d.). Introduction—Statsmodels. Retrieved May 15, 2022, from https://www.statsmodels.org/stable/index.html.

Sullivan, Z. (2018, March 29). Confronting Austin's history of racial segregation. https://nextcity.org/urbanist-news/confronting-austins-history-of-racial-segregation.

Tan, S., Fowers, A., Keating, D., & Tierney, L. (2020, May 15). Amid the pandemic, public transit is highlighting inequalities in cities. Washington Post. https://www.washingtonpost.com/nation/2020/05/15/amid-pandemic-public-transit-is-highlighting-inequalities-cities/.

Thompson, K., & Sanders, J. (2022, March 3). Austin OKs $41 million in Project Connect anti-displacement efforts. KXAN Austin. https://www.kxan.com/news/local/austin/austin-oks-41-million-in-project-connect-anti-displacement-efforts/.

Toussaint-Comeau, M., & Lee, J. (2018). Determinants of housing values and variations in home prices across neighborhoods in cook count. ProfitWise News and Views. https://www.chicagofed.org/publications/profitwise-news-and-views/2018/determinants-of-housing-values-and-variations-in-home-prices-across-neighborhoods-in-cook-county.

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science, 174*, 433–442. https://doi.org/10.1016/j.procs.2020.06.111

Tsatsaronis, K., & Zhu, H. (2004). What drives housing price dynamics: Cross-country evidence (SSRN scholarly paper no. 1968425). Social Science Research Network. https://papers.ssrn.com/abstract=1968425.

Wang, Y., Wu, K., Zhao, Y., Wang, C., & Zhang, H. (2022). Examining the effects of the built environment on housing rents in the Pearl River Delta of China. *Applied Spatial Analysis and Policy, 15*(1), 289–313. https://doi.org/10.1007/s12061-021-09412-4

Wassmer, R. W., & Baass, M. C. (2006). Does a more centralized urban form raise housing prices? *Journal of Policy Analysis and Management, 25*(2), 439–462. https://doi.org/10.1002/pam.20180

Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning, 125*, 234–244. https://doi.org/10.1016/j.landurbplan.2014.01.017

Wu, J., Song, Y., Liang, J., Wang, Q., & Lin, J. (2018). Impact of mixed land use on housing values in high-density areas: Evidence from Beijing. *Journal of Urban Planning and Development, 144*(1), 05017019. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000422

Zhao, Y., & Ke, J. (2021). The impact of shared mobility services on housing values near subway stations. *Transportation Research Part D: Transport and Environment, 101*, Article 103097. https://doi.org/10.1016/j.trd.2021.103097

Zhong, H., & Li, W. (2016). Rail transit investment and property values: An old tale retold. *Transport Policy, 51*, 33–48. https://doi.org/10.1016/j.tranpol.2016.05.007

Zulkifli, S. N. A., Hamsa, A. A., Noor, N. M., & Ibrahim, M. (2017). Evaluation of land use density, diversity and ridership of Rail Based Public Transportation System. *Transportation Research Procedia, 25*, 5266–5281. https://doi.org/10.1016/j.trpro.2018.02.053