# **Disparities in factors affecting housing price:** a machine learning approach to the effects of built environment factors and house characteristics on single-family housing price

Yefu Chen[1]; Junfeng Jiao, Ph.D.[1]; Arya Farahi, Ph.D.[2]
[1]Community and Regional Planning, SOA; [2]Department of Statistics and Data Science

GOOD SYSTEMS
A UT Grand Challenge

URBAN INFO LAB
The University of Texas at Austin

TEXAS
The University of Texas at Austin
School of Architecture

## Introduction

The housing market is booming across the U.S., especially in Austin, Texas, and prices have exponentially increased in the past decade. These rapid increases have implications for housing affordability programs. The critical question is how to launch policies to achieve housing affordability and livability in Austin. As a first step, we model the residential real estate market and evaluate the built environment and residential characteristics. This information will guide urban planners to estimate the impact of changes in the built environment and urban form design on real estate values (Athey & Imbens, 2017; Binoy et al., 2021).

## Research Goal

There are two research questions:

- Modeling the impacts of built environment factors and house characteristics on single-family housing transaction prices
- Studying the dependency of these impacts to the neighborhood attributes

## Methods

We employ an interpretable machine learning framework to predict the sold price per square foot. There are three groups of independent variables: house characteristics (year built, types, stories, etc.), the surrounding built environment factors (public transit, new construction permission, and population density), and control variables (neighborhood attributes) in the global model.

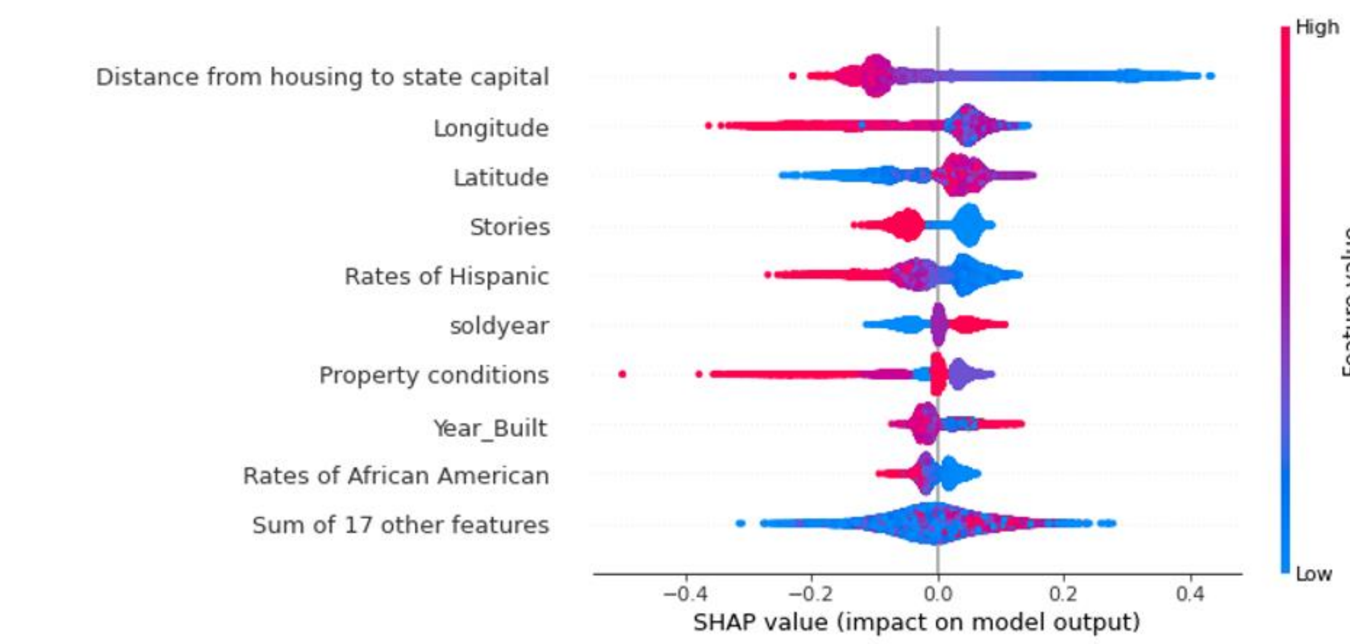
We study the dependency of these impacts to the neighborhood attributes by separating our data into six sub-groups:

- High-income neighborhoods: median household income > $75,887.
- Low-income neighborhoods: median household income < $75,887.
- High rates of white-only neighborhoods: the white-only population > 79%.
- High rates of Hispanic American neighborhoods: the Hispanic population > 33%.
- High rates of African American neighborhoods: African American population > 9%.
- Vulnerable neighborhoods: high nonwhite-only rates (> 29%) and lower median household income (<$75,887).

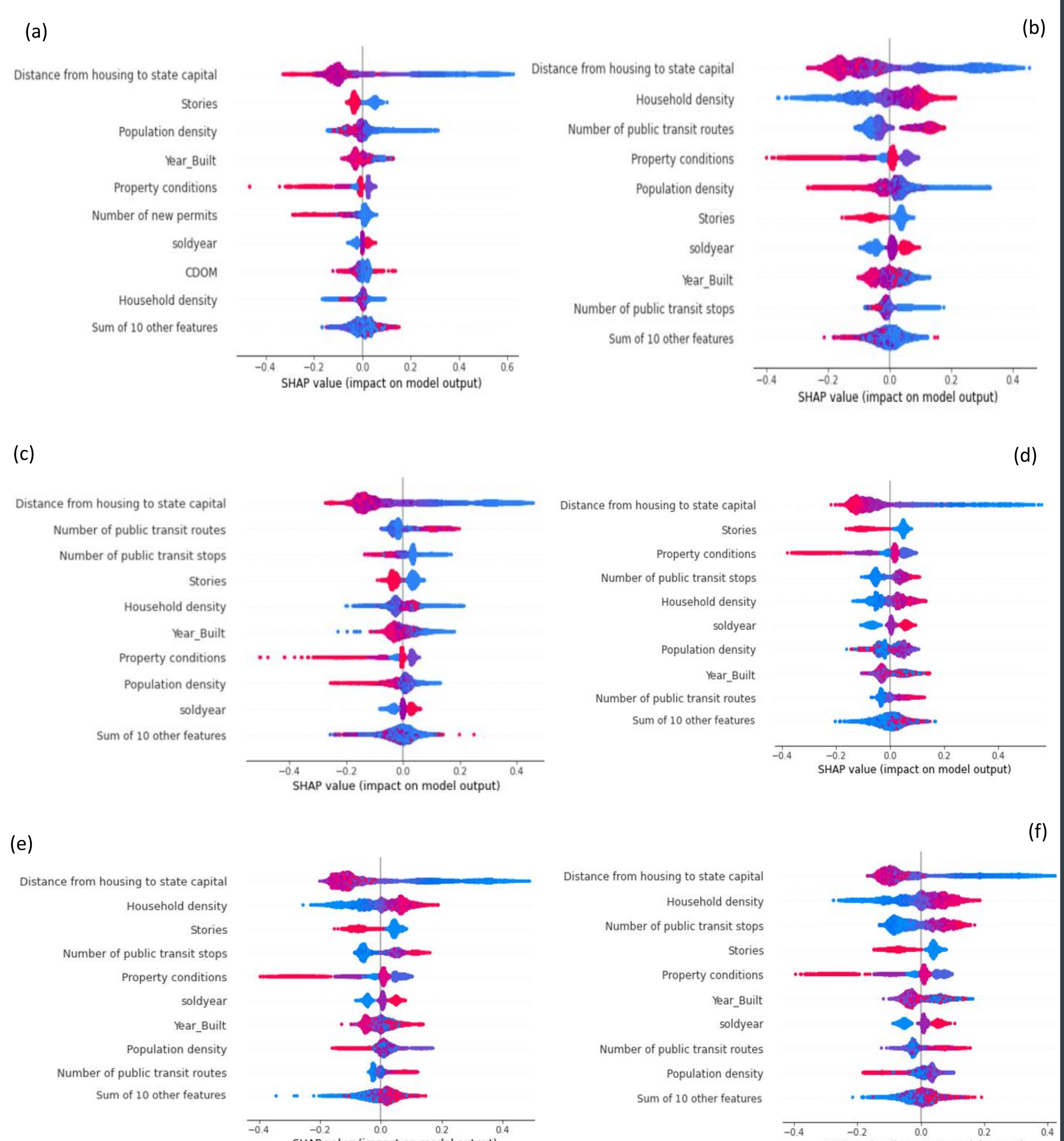
| Neighborhoods | High-income | Low-income | High rates of White-only | High rates of Hispanic American | High rates of African American | Vulnerable |
|---|---|---|---|---|---|---|
| **Explanatory variable: housing characteristics** | 0.11 | 0.14 | 0.09 | 0.15 | 0.14 | 0.14 |
| Year built | 0.03 | 0.04 | 0.03 | 0.04 | 0.03 | 0.04 |
| Type | | | | | | |
| Stories | 0.04 | 0.05 | 0.03 | 0.06 | 0.06 | 0.05 |
| CDOM | 0.02 | | | | | |
| Occupant type | | | | | | |
| Property conditions | 0.02 | 0.05 | 0.03 | 0.05 | 0.05 | 0.05 |
| Foundation | | | | | | |
| **Explanatory variable: surrounding environment** | 0.08 | 0.22 | 0.14 | 0.16 | 0.17 | 0.20 |
| Number of public transit routes | | 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 |
| 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 |

The surrounding environment surprisingly becomes crucial, although it is insignificant in the global result. It indicates that surrounding environmental factors are significant in housing prices predictions.

## Figures and Results




Distance from home to the state capital is the most important feature on average, and its impact on transaction prices can be negative, which means properties far from the state capital are more likely to have lower transaction prices. Also, flat houses may have relatively high transaction prices than stacked houses. The impacts of property conditions and the year built are complicated since there is no clear trend. Finally, the properties within neighborhoods with relatively high rates of minorities may have relatively low transaction prices.



* (a) high-income neighborhoods; (b) low-income neighborhoods; (c) high white-only rate neighborhoods; (d) high Hispanic American rates neighborhoods; (e) high African American rate neighborhood; (f) high vulnerable population rate neighborhoods

## Conclusion

This study investigates impacts on single-family residential property values in a fast-growing market – Austin housing. To this end, a housing valuation model was developed by extracting a list of essential housing characteristics and surrounding built environment. The results show that house status factors, including the year of housing, stories, and property conditions, are common vital predictors. The important surrounding environment attributes are density factors. Besides, public transit plays a significant role in most neighborhoods except the high-income.

Based on these results, several takeaways are worth noting.

1. Our interpretable ML framework may be used in other empirical studies. One can apply a similar process to explore the impacts on property values, while due to the inhomogeneous urban environment factors, one should be cautious in generalizing our findings to other cases.
2. Decision-makers pay attention to the positive role of public transit service on single-family residential property values. When public transit authorities are working on the public transit investment planning, they need to be careful since the single-family residential house prices can be increased due to new public transit facilities, which can also lead to increases in the property tax. Due to increases in property tax, vulnerable populations have to find a new place to live, causing displacements, since they cannot afford these increases. Hence, we encourage local authorities to cooperate with communities in case of potential transit-led displacements.
3. Based on the results, 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

## Future Directions

In follow-up works, we plan to:

1. Study the impact of other points of interest to housing value.
2. Study structural inequalities in homeownership
3. Explore how AI can assist redesigning the housing affordability programs

## Acknowledgments

This work was supported by the Good Systems Grand Challenge at The University of Texas at Austin.

## Key References

- Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. Journal of Economic Perspectives, 31(2), 3–32. https://doi.org/10.1257/jep.31.2.3
- 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

The University of Texas at Austin
Bridging Barriers
Office of the Vice President for Research