

IDC 6940 Capstone Course in Data Science

# Who Faces Housing Insecurity

An Analysis of Renters in U.S. Metropolitan Areas

Student: Geun Sang Oh

Course Instructor: Ananda M. Mondal, Ph.D.

Mentor: Florence George, Ph.D.

# Background

- 20% of renters in 25 U.S. metropolitan areas experience multiple forms of housing insecurity (e.g., forced moves, poor home conditions, high rent burdens) (Routhier, 2019)
- In recent decades, around 8 million very low-income renters are spending more than 50% of their income on rent, without any housing assistance (Alvarez & Steffen, 2023)
- From 2022 to 2024, people experiencing homelessness rose rapidly from 582,462 to 771,480, after remaining steady for years (de Sousa & Henry, 2024)

*In U.S. metropolitan areas, renters, particularly those with low income, are increasingly at risk of housing insecurity*

- Alvarez, T. A., & Steffen, B. L. (2023). *Worst Case Housing Needs: 2023 Report to Congress*. Office of Policy Development and Research. <https://www.huduser.gov/portal/portal/sites/default/files/pdf/Worst-Case-Housing-Needs-2023.pdf>
- de Sousa, T., & Henry, M. (2024). *The 2024 Annual Homelessness Assessment Report (AHAR) to Congress: PART 1: Point-in-Time Estimates of Homelessness*. Office of Community Planning and Development. <https://www.huduser.gov/portal/sites/default/files/pdf/2024-AHAR-Part-1.pdf>
- Routhier, G. (2019). Beyond Worst Case Needs: Measuring the Breadth and Severity of Housing Insecurity Among Urban Renters. *Housing Policy Debate*, 29(2), 235-249. <https://doi.org/10.1080/10511482.2018.1509228>

# Motivation and Goals

- Motivation
  - Renters often face multiple forms of housing insecurity, but most research focuses on one aspect (Routhier, 2019)
  - This research examines housing insecurity by considering its severity through its multiple aspects
- Research Goals
  - Examine the extent and severity of housing insecurity among renters in U.S. metropolitan areas
  - Apply statistical and machine learning methods to identify renter characteristics that predict housing insecurity
  - Inform policies that can improve renters' housing security

# Research Questions

- Research Questions
  - How extensive and severe is housing insecurity in U.S. metropolitan areas?
  - What characteristics best represent renters facing housing insecurity?
  - Do housing policies, such as subsidized housing and rental assistance, contribute to addressing housing insecurity?
  - What characteristics best predict *severe* housing insecurity?

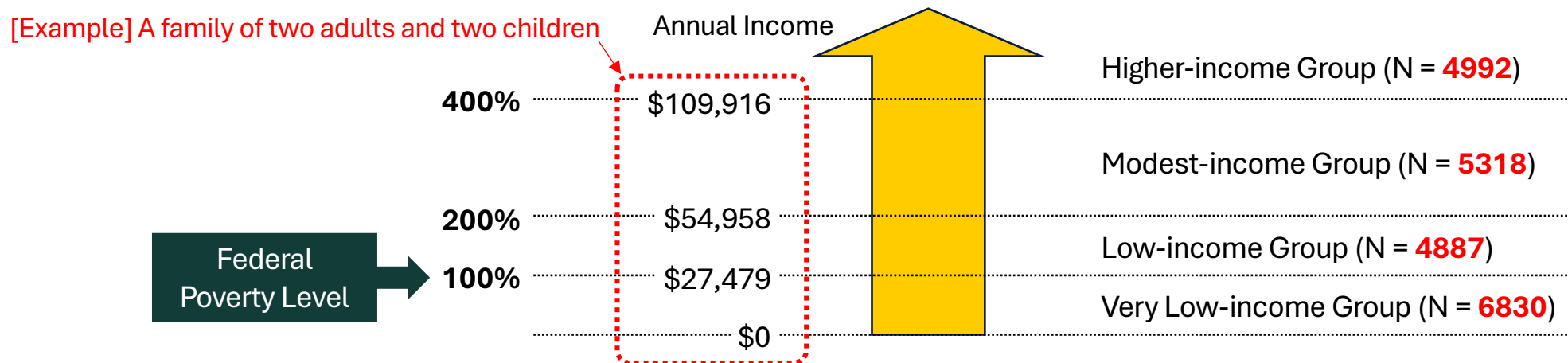
# Dataset

- Data source: 2021 American Housing Survey (AHS)
  - Supplier: Census Bureau & Department of Housing and Urban Development (HUD)
  - Survey topics: home conditions, household characteristics, housing costs, etc.
  - Number of observations: **64,141** housing units
- Selected sample: renter households in metropolitan areas
  - **22,027** housing units

Areas	Tenure (Houses)				
	Owned	Rented	Occupied without Rent	N/A (e.g., vacant units)	Total
Metropolitan Areas	28328	22027	561	6768	57684
Non-metropolitan Areas	3341	1655	146	1315	6457
Total	31669	23682	707	8083	64141

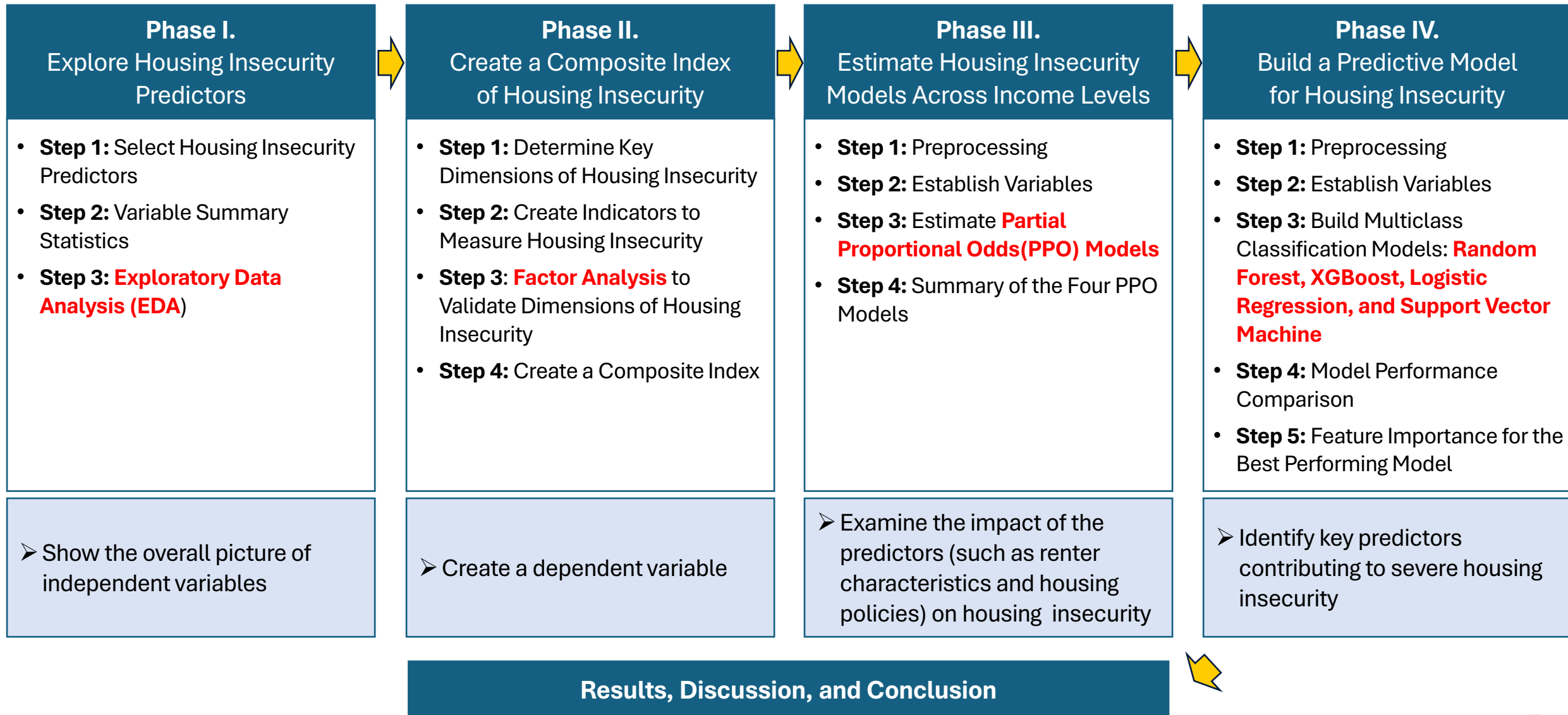
# Dataset (cont.)

- Divide the sample into four income groups
  - Assumption: renter characteristics related to housing insecurity may vary across income groups
  - Reference: 2021 federal poverty levels by annual income (Census Bureau, 2021; See [Appendix A](#) for details)
  - Criteria: Low-income baseline = 200% of federal poverty level (Kilduff, 2022)
    - I further divided the low-income level into two categories: very low-income and low-income
    - I defined 400% as the modest-income baseline



- U.S. Census Bureau. (2021). *Poverty Thresholds*. Retrieved October 14, 2024, from <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>
- Kilduff, L. (2022, January 31). *How Poverty in the United States Is Measured and Why It Matters*. Population Reference Bureau. <https://www.prb.org/resources/how-poverty-in-the-united-states-is-measured-and-why-it-matters/>

# Research Approach



# Phase I: Explore Housing Insecurity Predictors



# Step 1: Select Housing Insecurity Predictors

- Using AHS dataset, I selected several characteristics of renters, houses, and communities, along with housing policies and regions, which I used as housing insecurity predictors

## Predictors Selected from the AHS Data

Renter characteristics	Householder	Race, Spanish origin, Non-US citizen, Gender, Age, Education level, Marital status
	Household	Length of stay, Presence of young children (age<6), Presence of seniors (age≥65), Presence of disabled members, Number of household members
House characteristics	Age of house, Unit type	
Community characteristics (Self-rated by renters)	No good schools, No good public transportation, High levels of petty crime, High levels of serious crime, High disaster risk (such as flooding)	
Housing policies	Rent control, HUD* assistance	
Regions	US census division	

*Note.* The variables selected for this research were based on those used in Siskar and Evans (2021), where the authors variables related to utilized race/ethnicity, nativity, educational attainment, household demographics, and region.

\* HUD: U.S. Department of Housing and Urban Development

## Step 2: Variable Summary Statistics

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Row %				Col %				
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Householder	Race	White	0.268	0.221	0.254	0.257	0.610				(categorical)
		Black	0.416	0.237	0.221	0.127	0.282				
		Asian	0.238	0.163	0.217	0.381	0.067				
		Other races	0.326	0.230	0.235	0.209	0.041				
	Spanish origin	Yes	0.326	0.247	0.263	0.164	0.229				(categorical)
		No	0.305	0.215	0.235	0.245	0.771				
	Non-US citizen	Yes	0.266	0.241	0.268	0.225	0.127				(categorical)
		No	0.316	0.219	0.237	0.227	0.873				
	Gender	Male	0.219	0.194	0.273	0.314	0.415				(categorical)
		Female	0.375	0.241	0.219	0.165	0.585				
	Age	Householder's age	53.4	52.1	45.5	42.2	48.7	15	85	17.7	Years
	Education level	Below high school	0.523	0.260	0.167	0.050	0.182				(categorical)
		High school	0.359	0.271	0.244	0.125	0.269				
		College	0.296	0.246	0.275	0.183	0.285				
		Bachelor's	0.139	0.129	0.267	0.464	0.173				
	Marital status	Master's or higher	0.105	0.098	0.228	0.569	0.091				(categorical)
		Married	0.185	0.216	0.296	0.303	0.253				
		Separated (including widowed and divorced)	0.399	0.261	0.201	0.139	0.338				
		Never married	0.314	0.193	0.241	0.252	0.410				

Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.

See [Appendix B](#) for the table with column-wise percentages

## Step 2: Variable Summary Statistics (cont.)

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Row %				Col %				
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Household	Length of stay (Years in current residence)	Less than 5 years	0.261	0.203	0.262	0.274	0.607				(categorical)
		5 to less than 10 years	0.346	0.254	0.222	0.178	0.197				
		10 years or more	0.427	0.247	0.198	0.128	0.196				
	Presence of young children (age<6)	Yes	0.379	0.248	0.243	0.130	0.122				(categorical)
		No	0.301	0.218	0.241	0.240	0.878				
	Presence of seniors (age≥65)	Yes	0.414	0.298	0.182	0.106	0.248				(categorical)
		No	0.276	0.197	0.261	0.267	0.752				
	Presence of disabled members	Yes	0.474	0.276	0.164	0.086	0.279				(categorical)
		No	0.247	0.201	0.272	0.281	0.721				
	# of household members	Number of household members	2.16	2.33	2.36	2.06	2.22	1	19	1.48	Persons
House	Age of house	Years since the house was built	53.5	52.3	50.1	47.4	51.0	0	102	27.1	Years
	Unit type	One-family house	0.259	0.218	0.263	0.260	0.287				(categorical)
		Apartment	0.331	0.222	0.230	0.217	0.696				
		Mobile house, etc.	0.316	0.279	0.330	0.074	0.017				

Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.

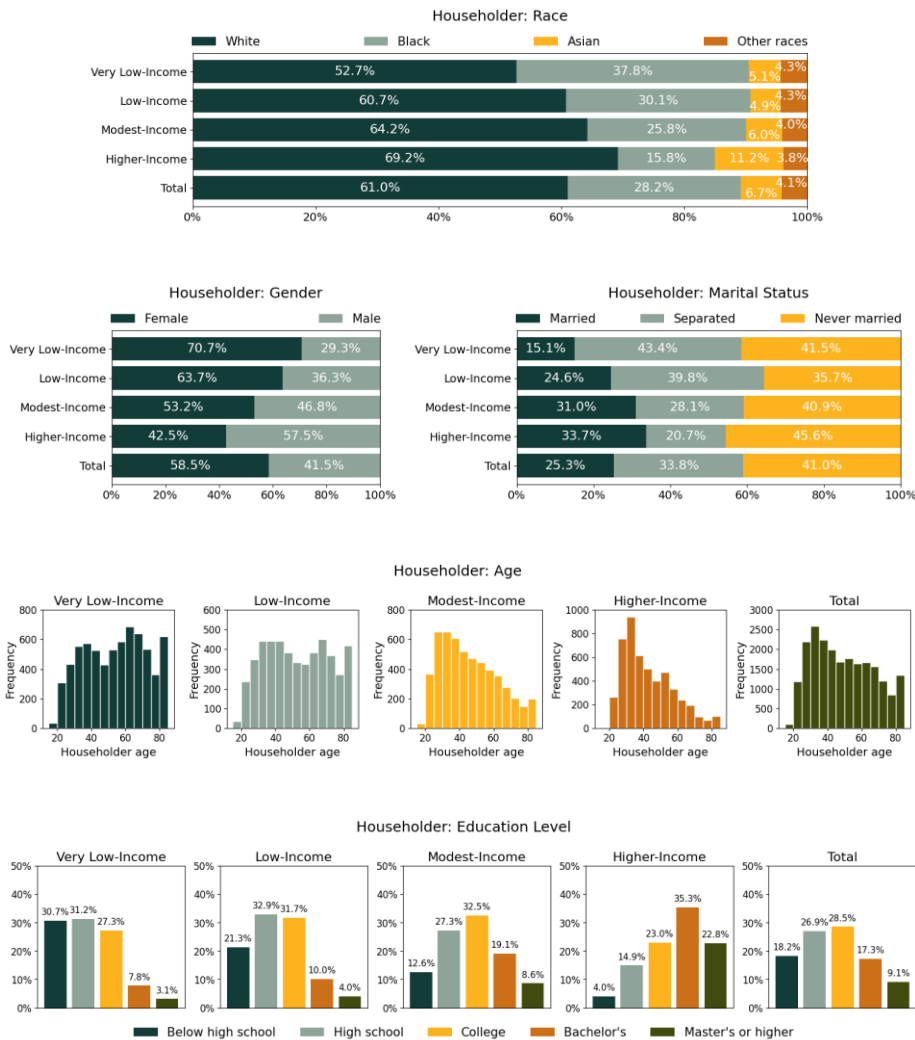
## Step 2: Variable Summary Statistics (cont.)

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Row %				Col %				
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Community (Self-Rated)	No good schools	Agree	0.374	0.226	0.234	0.166	0.096				(categorical)
		Disagree + no response	0.303	0.221	0.242	0.233	0.904				
	No good public transportation	Agree	0.269	0.217	0.260	0.254	0.273				(categorical)
		Disagree + no response	0.326	0.224	0.234	0.216	0.727				
	High levels of petty crime	Agree	0.377	0.237	0.222	0.164	0.288				(categorical)
		Disagree + no response	0.283	0.216	0.249	0.252	0.712				
	High levels of serious crime	Agree	0.464	0.231	0.185	0.120	0.145				(categorical)
		Disagree + no response	0.284	0.220	0.251	0.245	0.885				
	High disaster risk (such as flooding)	Agree	0.366	0.218	0.217	0.198	0.012				(categorical)
		Disagree + no response	0.304	0.222	0.244	0.230	0.898				
Housing Policy	Rent control	Yes	0.316	0.175	0.177	0.332	0.030				(categorical)
		No	0.310	0.223	0.243	0.223	0.970				
	HUD assistance	Subsidized housing	0.637	0.261	0.079	0.022	0.180				(categorical)
		Rental assistance (voucher)	0.582	0.260	0.118	0.040	0.138				
		No assistance	0.169	0.204	0.309	0.318	0.682				
Region	US census division	Northwest	0.348	0.223	0.220	0.209	0.206				(categorical)
		Midwest	0.354	0.250	0.237	0.159	0.144				
		South	0.328	0.222	0.249	0.201	0.354				
		West	0.242	0.207	0.249	0.302	0.296				

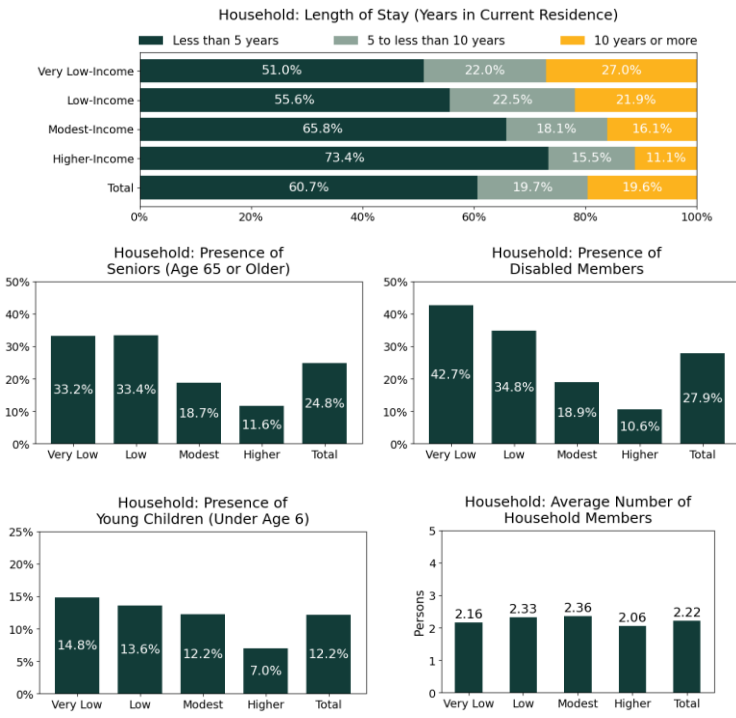
Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.

# Step 3: Exploratory Data Analysis (EDA)

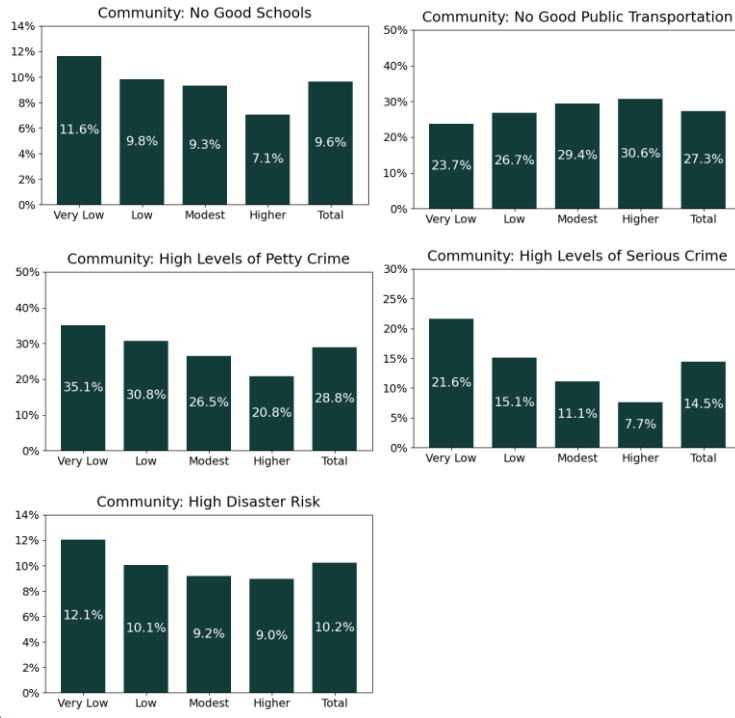
## Householder Characteristics



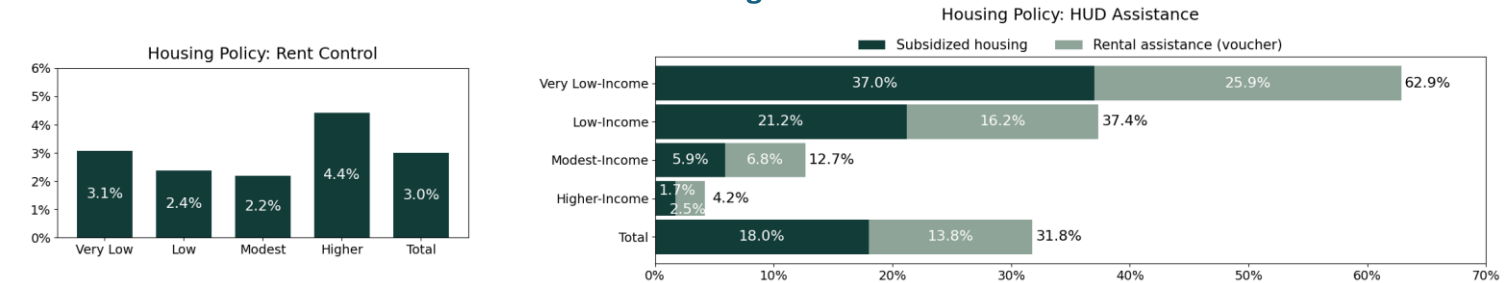
## Household Characteristics



## Community Characteristics



## Housing Policies



## Phase II:

# Create a Composite Index of Housing Insecurity

# Step 1: Determine Key Dimensions of Housing Insecurity

- A comprehensive review of existing literature
  - Cox et al., 2017; Leopold et al., 2016; Routhier, 2019; Watson & Cater, 2020

## Three Key Dimensions of Housing Insecurity

Residential Instability	Unfitness and Inadequacy	Unaffordability
<ul style="list-style-type: none"><li>• The household is <u>at significant risk of involuntary displacement</u> for economic or non-economic reasons</li></ul>	<ul style="list-style-type: none"><li>• The housing unit <u>lacks sufficient physical attributes to meet functional needs</u> for health, security, and daily living activities</li></ul>	<ul style="list-style-type: none"><li>• <u>Housing costs are not manageable over the long term</u>, compromising other essential needs for health and well-being</li></ul>

- Cox, R., Rodnyansky, S., Henwood, B., & Wenzel, S. (2017). *Measuring Population Estimates of Housing Insecurity in the United States: A Comprehensive Approach* (CESR-SCHAEFFER Working Paper No. 2017-012). Center for Economic and Social Research; Schaeffer Center for Health Policy & Economics.  
<https://dx.doi.org/10.2139/ssrn.3086243>
- Leopold, J., Cunningham, M., Posey, L., & Manuel, T. (2016, November 23). *Improving Measures of Housing Insecurity: A Path Forward*. Urban Institute.  
<https://www.urban.org/research/publication/improving-measures-housing-insecurity-path-forward>
- Routhier, G. (2019). Beyond Worst Case Needs: Measuring the Breadth and Severity of Housing Insecurity Among Urban Renters. *Housing Policy Debate*, 29(2), 235-249.  
<https://doi.org/10.1080/10511482.2018.1509228>
- Watson, N. E., & Carter, G. R. (2020). Toward Implementation of a National Housing Insecurity Research Module, *Cityscape*, 22(1), 227-248.  
<https://www.jstor.org/stable/26915495>

## Step 2: Create Indicators to Measure Housing Insecurity

- Nine binary indicators created using AHS data
  - Select AHS variables for each key dimension of housing insecurity, based on Routhier (2019)
  - Transform them into binary indicators

Key Dimension	Binary Indicator
Residential Instability	<ul style="list-style-type: none"><li>• <b>Eviction Risk:</b> Highly likely to be evicted within 2 months</li><li>• <b>Eviction Threats:</b> Threatened with eviction in the last 3 months</li><li>• <b>Rent Arrears:</b> Unable to pay at least one of the last 3 months</li><li>• <b>Utility Shut-Offs:</b> Had utilities shut-off</li></ul>
Unfitness and Inadequacy	<ul style="list-style-type: none"><li>• <b>Inadequacy:</b> Severely inadequate unit (see <a href="#">Appendix C</a> for details)</li><li>• <b>Upkeep:</b> 3 or more upkeep problems</li><li>• <b>Self-Rated Physical Conditions of Renters:</b> 4 or less on a scale of 10</li></ul>
Unaffordability	<ul style="list-style-type: none"><li>• <b>Severe Rent Burden*:</b> Gross rent** above 50% of income</li><li>• <b>Housing-Induced Poverty:</b> Income below 133% of federal poverty level*** (if only gross rent above 30% of income)</li></ul>

\* The Department of Housing and Urban Development (HUD) defines “rent burdened” as paying more than 30% of income for gross rent and “severely rent burdened” as paying more than 50%.

\*\* Gross Rent: Rent plus separate utility costs and related housing expenses (e.g., rental insurance)

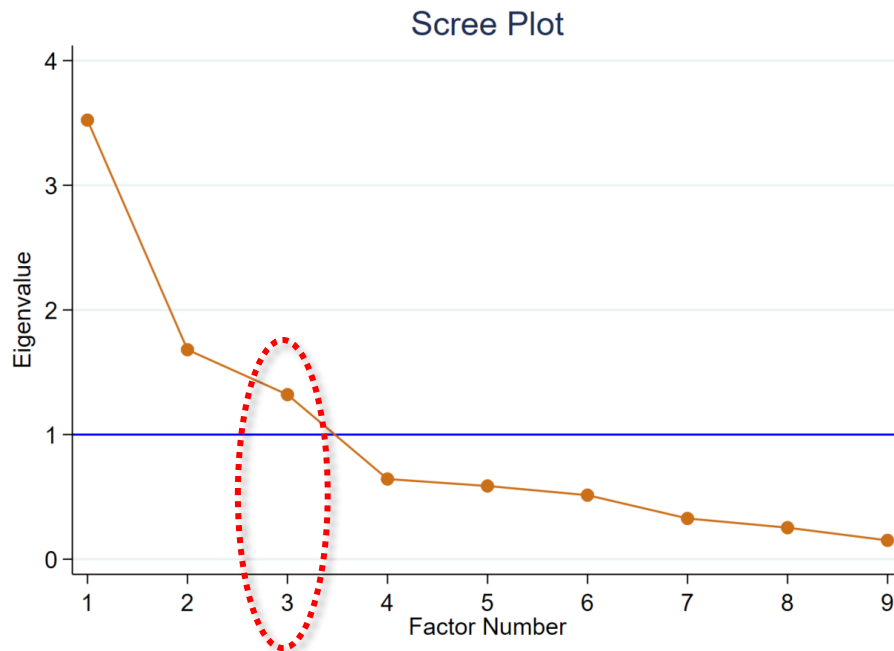
\*\*\* 133% of the federal poverty level is the eligibility threshold for Medicaid.

- Routhier, G. (2019). Beyond Worst Case Needs: Measuring the Breadth and Severity of Housing Insecurity Among Urban Renters. *Housing Policy Debate*, 29(2), 235-249.  
<https://doi.org/10.1080/10511482.2018.1509228>



# Step 3: Factor Analysis to Validate Dimensions of Housing Insecurity

- **Exploratory Factor Analysis (EFA)** to examine underlying factor structure of data
- EFA confirms the appropriateness of three dimensions (factors) and nine indicators
  - The scree plot and Kaiser's criterion (eigenvalue > 1) demonstrate the adequacy of the **three** factors
  - Factor loadings exhibit the validity of **the three key dimensions supported by the nine indicators**



Factor Loadings for the Three-Factor Solution

Dimension	Indicator	Factor 1	Factor 2	Factor 3	Uniqueness
Residential Instability	Eviction Risk	0.597	0.095	-0.025	0.586
	Eviction Threats	0.790	0.123	-0.052	0.285
	Rent Arears	0.767	0.006	0.019	0.399
	Utility Shut-Offs	0.640	-0.101	0.040	0.632
Unfitness and Inadequacy	Inadequacy	0.017	0.689	0.030	0.505
	Upkeeps	-0.046	0.806	-0.012	0.388
	Self-Rated Physical Conditions	0.067	0.604	0.012	0.587
Unaffordability	Severe Rent Burden	0.014	-0.032	0.880	0.227
	Housing-Induced Poverty	-0.013	0.058	0.877	0.218
Variance		2.599	2.190	1.742	
Proportion of Variance Explained		0.561	0.473	0.376	

Note. Tetrachoric correlation matrix was used for binary variables. Tetrachoric correlations are specifically designed for binary variables and assume that the underlying relationship between two binary variables follows a bivariate normal distribution (StataCorp, 2021). Oblique rotation was used, assuming factors can be correlated. 'Uniqueness' is the variance unique to the variable, not shared with other variables. The smaller the 'Uniqueness,' the stronger the variable's relevance in explaining the underlying factors.

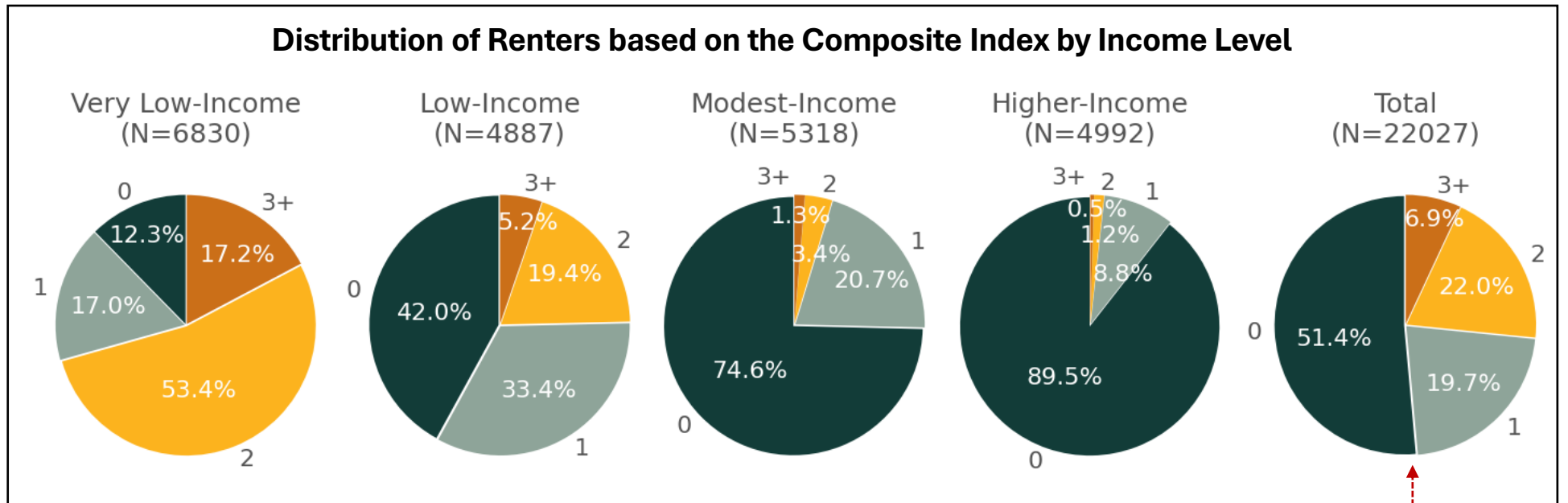
# Step 4: Create a Composite Index

- Create a Composite Index of Housing Insecurity
  - An index that sums all nine indicators (range: 0 to 8)
  - A value of 1 or higher: housing insecurity
  - A value of 3 or higher: severe housing insecurity
  - Will serve as the dependent variable in the next phases

**Distribution of All Renters  
based on the Composite Index**

Index Value	N	%
0	11330	51.44
1	4333	19.67
2	4836	21.95
3+	3	4.71
	4	1.53
	5	0.53
	6	0.13
	7	0.07
	8	0.02
Total	22027	100.00

**Distribution of Renters based on the Composite Index by Income Level**



## Phase III:

# Estimate Housing Insecurity Models Across Income Levels

# Step 1: Preprocessing

- Adjust the composite index to account for small observation counts in certain index values:
  - For very low- and low-income groups, values of 3 and higher were combined into value 3
  - For modest- and higher-income groups, values of 2 and higher were combined into value 2

Very Low-Income  
(N=6830)

→

Value	0	1	2	3
Obs.	842	1164	3646	1178

Low-Income  
(N=4887)

→

Value	0	1	2	3
Obs.	2053	1630	948	256

Modest-Income  
(N=5318)

→

Value	0	1	2
Obs.	3969	1099	250

Higher-Income  
(N=4992)

→

Value	0	1	2
Obs.	4466	440	86

Preprocessed Composite Index of Housing Insecurity

# Step 2: Establish Variables

## Dependent and Independent Variables

Dependent Var.			Preprocessed Composite Index { <ul style="list-style-type: none"><li>• {0, 1, 2, 3} for Very low-income group (N=6830)</li><li>• {0, 1, 2, 3} for Low-income group (N=4887)</li><li>• {0, 1, 2} for Modest-income group (N=5318)</li><li>• {0, 1, 2} for Higher-income group (N=4992)</li></ul>
Independent Vars.	Renter	Householder	Race, Spanish origin, Non-US citizen, Gender, <u>Age</u> , Education level, Marital status
		Household	Length of stay, Presence of young children (age<6), Presence of seniors (age≥65), Presence of disabled members, <u>Number of household members</u>
	House		<u>Age of house</u> , Unit type
	Community		No good schools, No good public transportation, High levels of petty crime, High levels of serious crime, High disaster risk
	Housing Policy		Rent control, HUD assistance
	Region		US census division

*Note.* The numerical variables are underlined.

# Step 3: Estimate Partial Proportional Odds (PPO) Models

- **Partial Proportional Odds (PPO) Model**

- To estimate the impact of the predictors on housing insecurity across four income groups
  - An extended ordinal regression model that offers flexibility by allowing predictors to have varying effects across levels of housing insecurity
  - Four models in total (one for each income group)
- 
- Tool: *gologit2*, a user-written program for Stata (Williams, 2006)

# Step 4: Summary of the Four PPO Models

- Relative Risk Ratios for Statistically Significant Variables

Variable / Category			Very low-income	Low-income	Modest-income	Higher-income
Householder	Race (ref. White)	Black	1.279	1.283		1.531
		Other races	1.501 (S)		1.604	
	Age		0.991			
	Education level (ref. College)	Bachelor's				0.389
	Marital status (ref. Never Married)	Married		0.767		
Household	Length of stay (ref. Less than 5 years)	5 to less than 10 years		0.819		
	Presence of seniors		0.617	0.719		
	Presence of disabled members		1.416 (S)	1.231, 1.812	1.387	1.702
	Number of household members		0.891	0.876	0.856	
House	Age of house		1.003	1.009		1.006
	Unit type (Ref. Apartment)	One-family house	1.535, 1.844		1.263	
		Mobile house		0.414		
Community	No good schools		1.623 (S)	1.345	1.360	1.661
	High levels of petty crime		1.757 (S)			2.766
	High levels of serious crime		1.538	1.327, 1.675, 2.841	1.844, 4.023	1.982
	High disaster risk		1.533 (S)	1.373	2.293	
Housing Policy	HUD assistance (ref. No assistance)	Subsidized housing	0.120, 0.126, 0.496	0.341		
		Rental Assistance	0.179, 0.208, 0.550	0.492		

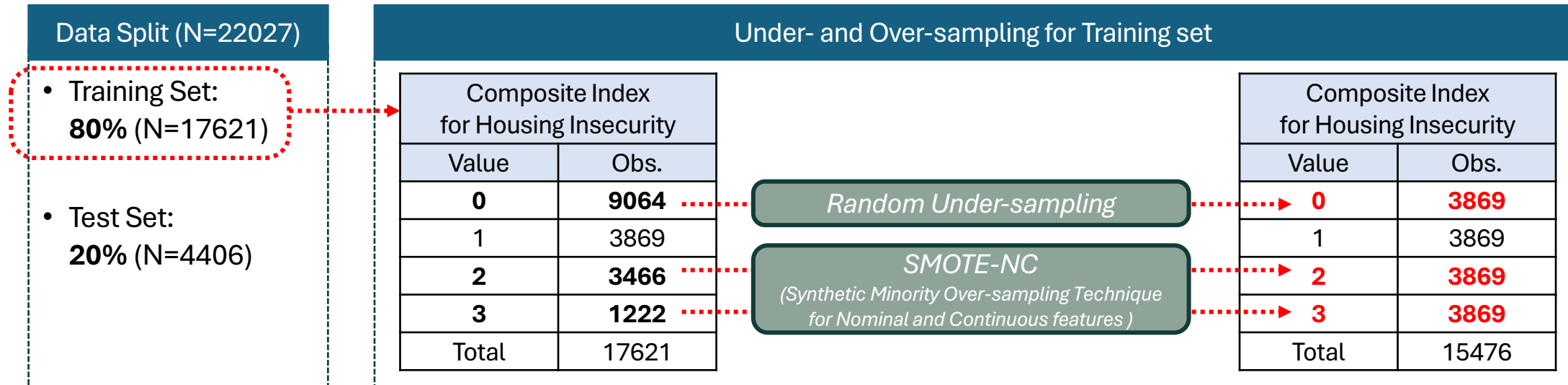
Note. The variables presented here are statistically significant at the 0.01 level. Red colors indicate variables that increase the risk of housing insecurity, while green colors indicate variables that decrease the risk of housing insecurity. 'S' indicates variables that only affect the risk of severe housing insecurity.

## Phase IV: Build a Predictive Model for Housing Insecurity



# Step 1: Preprocessing

- Combine all four income groups into one dataset
  - For the composite index of housing insecurity, values of 3 and higher were combined into value 3
- Adjust the variable settings
  - Add the 'ln\_income' variable (the natural logarithm of the 'income' variable)
  - Standardize numerical variables
- Split the data and balance the training set



# Step 2: Establish Variables

## Dependent and Independent Variables

Dependent Var.			Preprocessed Composite Index (Values: {0, 1, 2, 3} for Total Renters)
Independent Vars.	Renter	Householder	Race, Spanish origin, Non-US citizen, Gender, <u>Age</u> , Education level, Marital status
		Household	Length of stay, Presence of young children (age<6), Presence of seniors (age≥65), Presence of disabled members, <u>Number of household members</u> , <u>ln Income</u>
	House		<u>Age of house</u> , Unit type
	Community		No good schools, No good public transportation, High levels of petty crime, High levels of serious crime, High disaster risk
	Housing Policy		Rent control, HUD assistance
	Region		US census division

Note. The numerical variables are underlined and have been standardized.

# Step 3: Build Multiclass Classification Models

- Four Models
  - **Random Forest:** An ensemble learning algorithm that uses multiple decision trees to reduce overfitting and can model complex, non-linear relationships
  - **XGBoost:** Provides high performance through gradient boosting and is fast and efficient in handling large datasets
  - **Logistic Regression:** A simple algorithm that is computationally efficient, trains quickly, and performs well as baseline, even when dealing with a high number of features
  - **Support Vector Machine (SVM):** Effectively handles high-dimensional data and works well for both linear and non-linear classification tasks
- Tools: *scikit-learn* and *XGBoost* libraries in Python

## Step 3: Build Multiclass Classification Models (cont.)

- I used 5-fold cross-validation to optimize hyperparameters and improve model performance

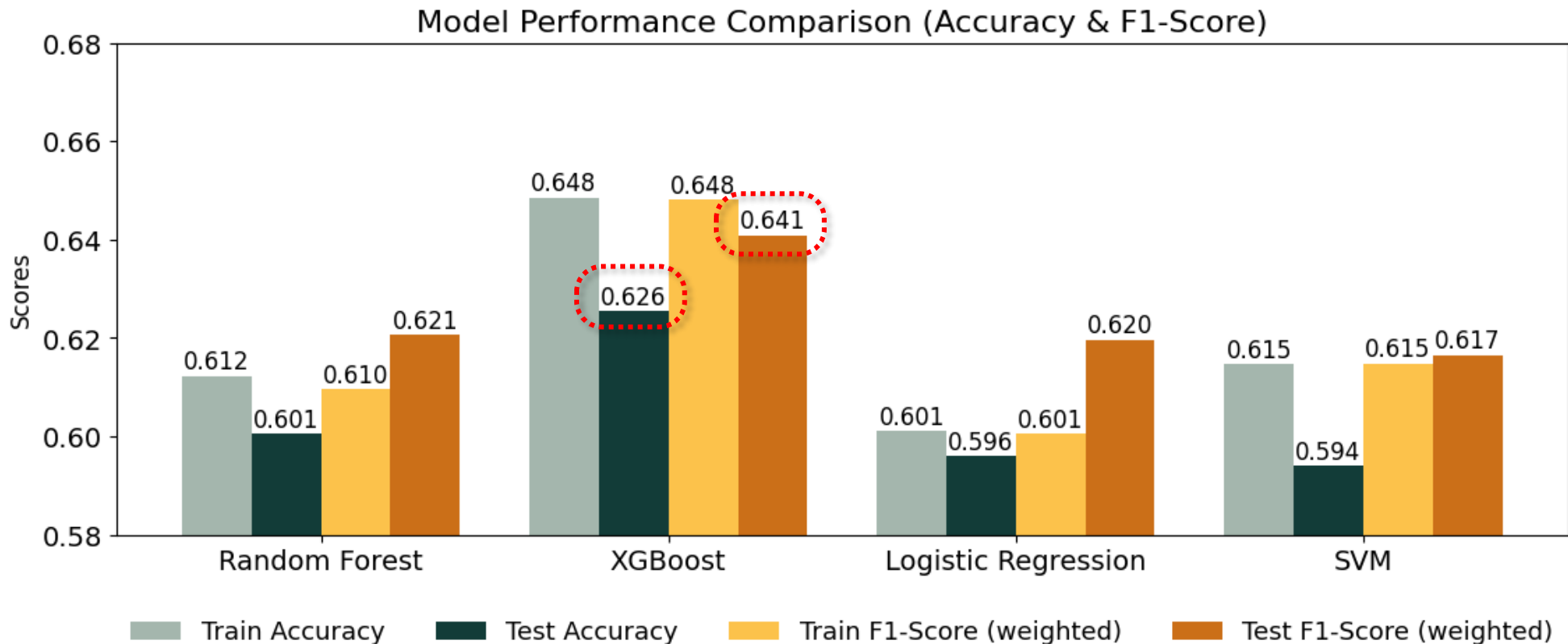
	Random Forest	XGBoost	Logistic Regression	Support Vector Machine (SVM)
Key Hyperparameters	<ul style="list-style-type: none"><li>n_estimator = 800</li><li>max_depth = 7</li><li>max_leaf_nodes = 30</li><li>max_features = 0.85</li><li>max_samples = 0.20</li></ul>	<ul style="list-style-type: none"><li>n_estimators = 550</li><li>learning_rate = 0.01</li><li>max_depth = 20</li><li>min_child_weight = 6</li><li>gamma = 15</li><li>subsample = 0.99</li><li>colsample_bytree = 0.79</li><li>reg_lambda = 0.05</li></ul>	<ul style="list-style-type: none"><li>solver = 'lbfgs' *</li><li>C = 10</li><li>penalty = 'l2'</li></ul>	<ul style="list-style-type: none"><li>kernel = 'rbf' **</li><li>C = 100</li><li>gamma = 0.001</li></ul>
Training & Validation Accuracies (CV=5)	<ul style="list-style-type: none"><li>Train Accuracy = 0.615</li><li>Valid Accuracy = 0.596</li></ul>	<ul style="list-style-type: none"><li>Train Accuracy = 0.642</li><li>Valid Accuracy = 0.629</li></ul>	<ul style="list-style-type: none"><li>Train Accuracy = 0.603</li><li>Valid Accuracy = 0.592</li></ul>	<ul style="list-style-type: none"><li>Train Accuracy = 0.616</li><li>Valid Accuracy = 0.599</li></ul>

\* lbfgs : Limited-memory Broyden-Fletcher-Goldfarb-Shanno Algorithm

\*\* rbf: Radial Basis Function

## Step 4: Model Performance Comparison

- XGBoost achieved the highest Test Accuracy (0.626) and Test F1-Score (0.641)

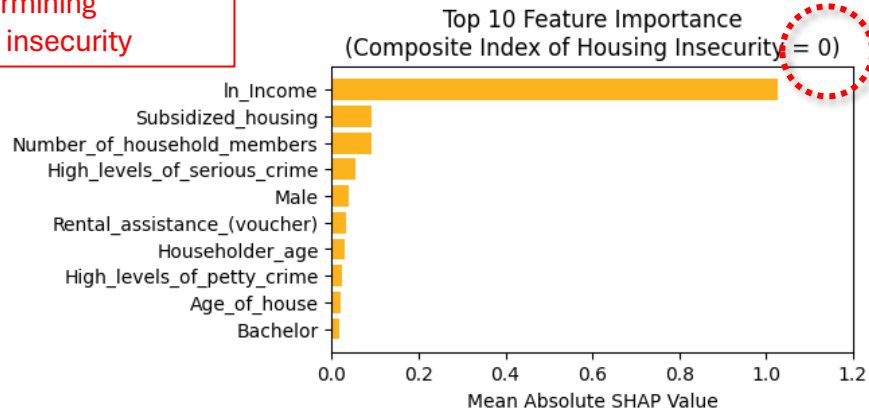


Note. 'Train Accuracy' and 'Train F1-score (weighted)' were calculated using the whole training set, including the validation set.

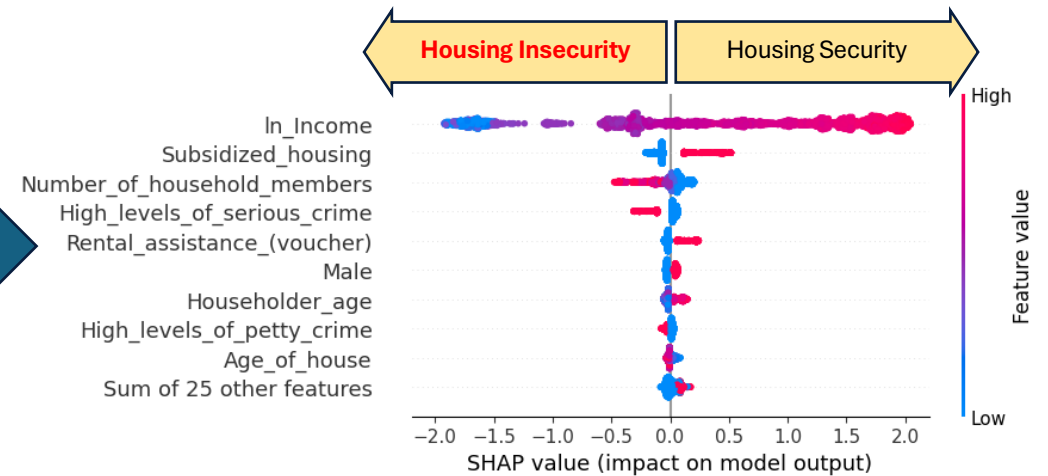
# Step 5: Feature Importance for the Best Performing Model (XGBoost)

- Feature Importance for XGBoost by *SHAP*\* values (the composite index = 0 and 3+)

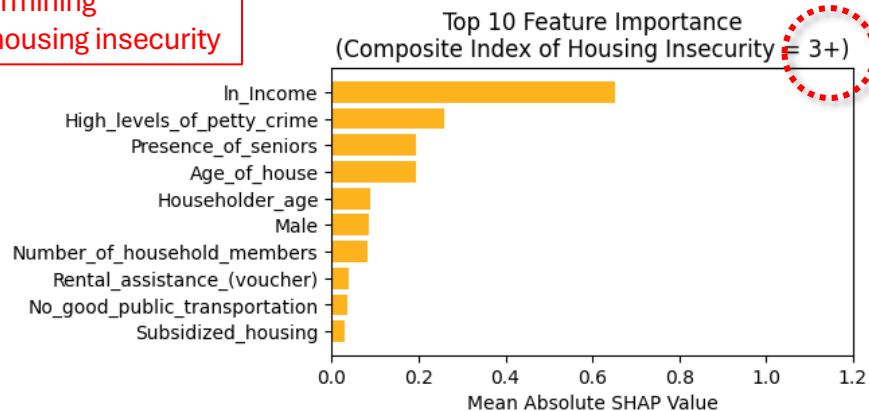
For determining  
Housing insecurity



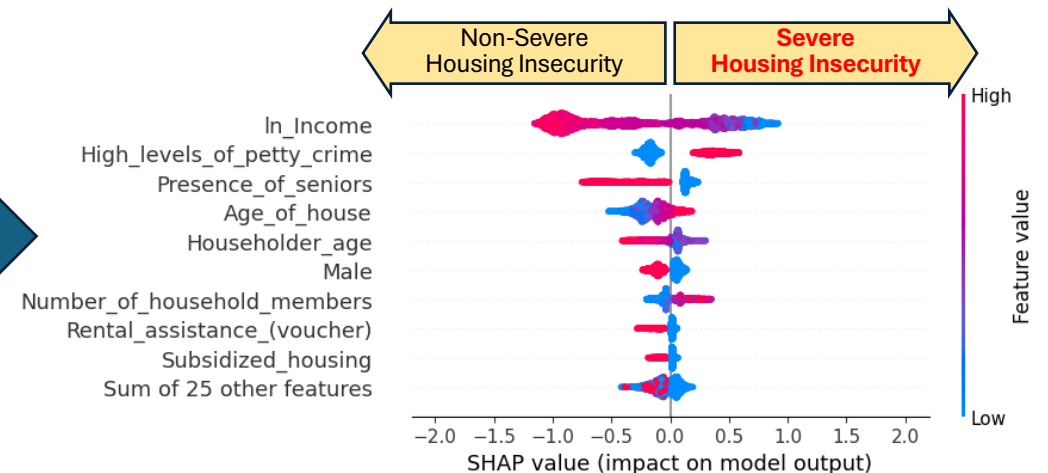
Beeswarm Plot



For determining  
Severe housing insecurity



Beeswarm Plot

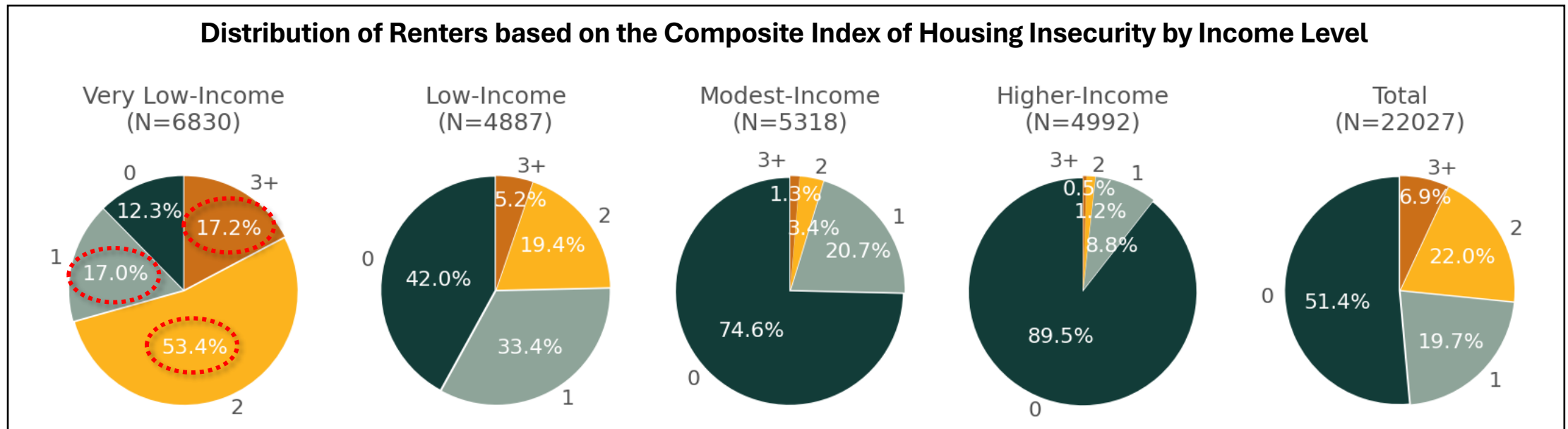


\* **SHAP (SHapley Additive exPLANations)**: A method used to estimate how each feature contributes to the prediction, based on Shapley values from game theory. The Shapley values represent the weighted average of a feature's marginal contributions across all possible subsets of features. SHAP values prevent distortions and misinterpretations caused by scaling features (Marsh, 2023). The Python library *shap* was used to calculate these values.

# Results, Discussion, and Conclusion

## How extensive and severe is housing insecurity in U.S. metropolitan areas?

- Almost half of all renters (**48.6%**) face housing insecurity; **6.9%** face severe housing insecurity (three or more indicators)
  - Very low-income group: **87.7%** face housing insecurity; **17.2%** face severe housing insecurity
- Policies need to pay particular attention to the very low-income group (renters below the poverty level) given their disproportionately high risk of housing insecurity





## What characteristics best represent renters facing housing insecurity?



## Do housing policies contribute to addressing housing insecurity?

- Subsidized housing and rental assistance greatly reduce the risk of housing insecurity for the very low- and low-income groups
  - For the very low-income renters in subsidized housing, the relative risk of housing insecurity is only **12%** of the relative risk for those without assistance
- The two housing policies are especially effective for households **on the brink of** housing insecurity, rather than those already facing it

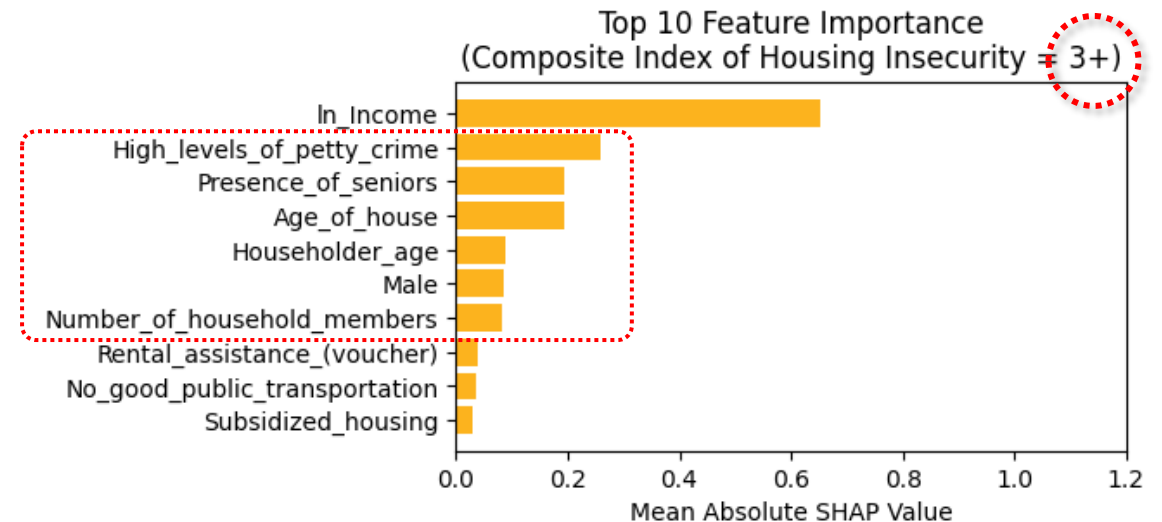
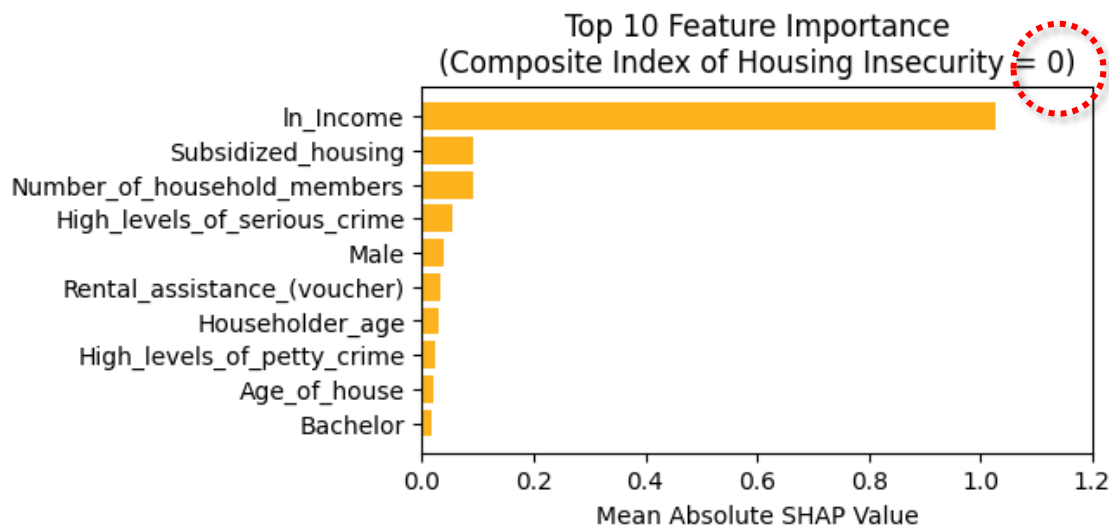
**Coefficients of HUD assistance variables for very low- and low-income groups in PPO models**

Income Group	Variable / Category	0 vs 1,2,3			0,1 vs 2,3			0,1,2 vs 3		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Very low	HUD assistance (ref. No assistance)									
	<b>Subsidized housing</b>	<b>-2.116**</b>	<b>0.143</b>	<b>0.120</b>	<b>-2.074**</b>	<b>0.087</b>	<b>0.126</b>	<b>-0.702**</b>	<b>0.086</b>	<b>0.496</b>
	<b>Rental assistance (voucher)</b>	<b>-1.722**</b>	<b>0.148</b>	<b>0.179</b>	<b>-1.570**</b>	<b>0.090</b>	<b>0.208</b>	<b>-0.599**</b>	<b>0.091</b>	<b>0.550</b>
Low	HUD assistance (ref. No assistance)									
	Subsidized housing	-1.076**	0.080	0.341	-1.076**	0.080	0.341	-1.076**	0.080	0.341
	Rental assistance (voucher)	-0.709**	0.083	0.492	-0.709**	0.083	0.492	-0.709**	0.083	0.492

Note. \*\* denotes significance at the 0.001 levels.

# What characteristics best predict *severe* housing insecurity?

- Income is most important in determining housing insecurity
- However, for *severe* housing insecurity (value  $\geq 3$ ), the importance of income decreases
  - relative to other characteristics such as 'High levels of petty crime,' 'Presence of seniors,' and 'Householder's age and gender'
- Future research requires a close examination of the householder characteristics, household composition, and community conditions among renters experiencing severe housing insecurity



# Conclusion

- This research has examined housing insecurity among renters in U.S. metropolitan areas by employing statistical and machine learning approaches
- This research highlights the value of analyzing housing insecurity by income level, as different income groups could face distinct challenges and vulnerabilities
- This research contributes to the housing studies literature by:
  - Identifying key characteristics of renters facing housing insecurity
  - Informing the need for policies that prioritize very low-income renters
  - Demonstrating the complementary use of statistical and machine learning approaches

Thank You

# References

- Alvarez, T. A., & Steffen, B. L. (2023). *Worst Case Housing Needs: 2023 Report to Congress*. Office of Policy Development and Research. <https://www.huduser.gov/portal//portal/sites/default/files/pdf/Worst-Case-Housing-Needs-2023.pdf>
- Cox, R., Rodnyansky, S., Henwood, B., & Wenzel, S. (2017). *Measuring Population Estimates of Housing Insecurity in the United States: A Comprehensive Approach* (CESR-SCHAEFFER Working Paper No. 2017-012). Center for Economic and Social Research; Schaeffer Center for Health Policy & Economics. <https://dx.doi.org/10.2139/ssrn.3086243>
- de Sousa, T., & Henry, M. (2024). *The 2024 Annual Homelessness Assessment Report (AHAR) to Congress: PART 1: Point-in-Time Estimates of Homelessness*. Office of Community Planning and Development. <https://www.huduser.gov/portal/sites/default/files/pdf/2024-AHAR-Part-1.pdf>
- Eggers, F. J., & Moumen, F. (2013). *Housing adequacy and quality as measured by the AHS* (Prepared for U.S. Department of Housing and Urban Development and Office of Policy Development and Research). Econometrica, Inc. <https://www.census.gov/content/dam/Census/programs-surveys/ahs/publications/HousingAdequacy.pdf>
- Kilduff, L. (2022, January 31). *How Poverty in the United States Is Measured and Why It Matters*. Population Reference Bureau. <https://www.prb.org/resources/how-poverty-in-the-united-states-is-measured-and-why-it-matters/>
- Leopold, J., Cunningham, M., Posey, L., & Manuel, T. (2016, November 23). *Improving Measures of Housing Insecurity: A Path Forward*. Urban Institute. <https://www.urban.org/research/publication/improving-measures-housing-insecurity-path-forward>
- Marsh, E. (2023, January 31). *Calculating XGBoost Feature Importance*. Medium. <https://medium.com/@emilykmarsh/xgboost-feature-importance-233ee27c33a4>
- Routhier, G. (2019). Beyond Worst Case Needs: Measuring the Breadth and Severity of Housing Insecurity Among Urban Renters. *Housing Policy Debate*, 29(2), 235-249. <https://doi.org/10.1080/10511482.2018.1509228>
- Siskar, T., & Evans, M. (2021). Predicting Mobility: Who Is Forced to Move? *City & Community*, 20(2), 141-159. <https://doi.org/10.1177/1535684120981010>
- StataCorp. (2021). *Stata 17 Base Reference Manual*, Stata Press.
- U.S. Census Bureau. (2021). *Poverty Thresholds*. Retrieved October 14, 2024, from <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>
- Watson, N. E., & Carter, G. R. (2020). Toward Implementation of a National Housing Insecurity Research Module, *Cityscape*, 22(1), 227-248. <https://www.jstor.org/stable/26915495>
- Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1), 58–82.

# Appendix A

- 2021 Federal Poverty Levels and the Values of 100%, 200%, and 400% of the Levels

## Federal Poverty Levels for 2021

(Unit: Dollars)

Size of family unit	Weighted average thresholds	Related children under 18 years									Percents of poverty levels*		
		None	One	Two	Three	Four	Five	Six	Seven	Eight or more	100%	200%	400%
One person:	13,788												
Under 65 years	14,097	14,097									14,097	28,194	56,388
65 years and over	12,996	12,996									12,996	25,992	51,984
Two people:	17,529												
Householder under 65 years	18,231	18,145	18,677								18,145	36,290	72,580
Householder 65 years and over	16,400	16,379	18,606								16,379	32,758	65,516
Three people	21,559	21,196	21,811	21,831							21,196	42,392	84,784
Four people	27,740	27,949	28,406	27,479	27,575						27,949	55,898	111,796
Five people	32,865	33,705	34,195	33,148	32,338	31,843					33,705	67,410	134,820
Six people	37,161	38,767	38,921	38,119	37,350	36,207	35,529				38,767	77,534	155,068
Seven people	42,156	44,606	44,885	43,925	43,255	42,009	40,554	38,958			44,606	89,212	178,424
Eight people	47,093	49,888	50,329	49,423	48,629	47,503	46,073	44,585	44,207		49,888	99,776	199,552
Nine people or more	56,325	60,012	60,303	59,501	58,828	57,722	56,201	54,826	54,485	52,386	60,012	120,024	240,048

\* I calculated the annual income values for 100%, 200%, and 400% of the poverty levels, based on households where all members are aged 18 or older..

- U.S. Census Bureau. (2021). *Poverty Thresholds*. Retrieved October 14, 2024, from <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>

# Appendix B

- Variable Summary Statistics (Column-wide Percentage) (1/3)

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Householder	Race	White	0.527	0.607	0.642	0.692	0.610				(categorical)
		Black	0.378	0.301	0.258	0.158	0.282				
		Asian	0.051	0.049	0.060	0.112	0.067				
		Other races	0.043	0.043	0.040	0.038	0.041				
	Spanish origin	Yes	0.240	0.254	0.250	0.166	0.229				(categorical)
		No	0.760	0.746	0.750	0.834	0.771				
	Non-US citizen	Yes	0.109	0.138	0.141	0.126	0.127				(categorical)
		No	0.891	0.862	0.859	0.874	0.873				
	Gender	Male	0.293	0.363	0.468	0.575	0.415				(categorical)
		Female	0.707	0.637	0.532	0.425	0.585				
	Age	Householder's age	53.4	52.1	45.5	42.2	48.7	15	85	17.7	Years
	Education level	Below high school	0.307	0.213	0.126	0.040	0.182				(categorical)
		High school	0.312	0.329	0.273	0.149	0.269				
		College	0.273	0.317	0.325	0.230	0.285				
		Bachelor's	0.078	0.100	0.191	0.353	0.173				
		Master's or higher	0.031	0.040	0.086	0.228	0.091				
	Marital status	Married	0.151	0.246	0.310	0.337	0.253				(categorical)
		Separated (including widowed and divorced)	0.434	0.398	0.281	0.207	0.338				
		Never married	0.415	0.357	0.409	0.456	0.410				

Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.



# Appendix B

- Variable Summary Statistics (Column-wide Percentage) (2/3)

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Household	Length of stay (Years in current residence)	Less than 5 years	0.510	0.556	0.658	0.607	0.607				(categorical)
		5 to less than 10 years	0.220	0.225	0.181	0.197	0.197				
		10 years or more	0.270	0.219	0.161	0.196	0.196				
	Presence of young children (age<6)	Yes	0.148	0.136	0.122	0.070	0.122				(categorical)
		No	0.852	0.864	0.878	0.930	0.878				
	Presence of seniors (age≥65)	Yes	0.332	0.334	0.187	0.116	0.248				(categorical)
		No	0.668	0.666	0.813	0.884	0.752				
	Presence of disabled members	Yes	0.427	0.348	0.189	0.106	0.279				(categorical)
		No	0.573	0.652	0.811	0.894	0.721				
	# of household members	Number of household members	2.16	2.33	2.36	2.06	2.22	1	19	1.48	Persons
House	Age of house	Years since the house was built	53.5	52.3	50.1	47.4	51.0	0	102	27.1	Years
	Unit type	One-family house	0.240	0.282	0.313	0.329	0.287				(categorical)
		Apartment	0.743	0.697	0.664	0.666	0.696				
		Mobile house, etc.	0.017	0.021	0.023	0.006	0.017				

Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.

# Appendix B

- Variable Summary Statistics (Column-wide Percentage) (3/3)

Variable		Category [ Description for Continuous Variables ]	Categorical (%) / Continuous (Mean)					Min	Max	S.D.	Unit
			Very Low Income (N=6830)	Low Income (N=4887)	Modest Income (N=5318)	Higher Income (N=4992)	Total (N=22027)				
Community (Self-Rated)	No good schools	Agree	0.116	0.098	0.093	0.071	0.096				(categorical)
		Disagree + no response	0.884	0.902	0.907	0.929	0.904				
	No good public transportation	Agree	0.237	0.267	0.294	0.306	0.273				(categorical)
		Disagree + no response	0.763	0.733	0.706	0.694	0.727				
	High levels of petty crime	Agree	0.351	0.308	0.265	0.208	0.288				(categorical)
		Disagree + no response	0.649	0.692	0.735	0.792	0.712				
	High levels of serious crime	Agree	0.216	0.151	0.111	0.077	0.145				(categorical)
		Disagree + no response	0.784	0.849	0.889	0.923	0.885				
	High disaster risk (such as flooding)	Agree	0.121	0.101	0.092	0.090	0.012				(categorical)
		Disagree + no response	0.879	0.899	0.908	0.910	0.898				
Housing Policy	Rent control	Yes	0.031	0.024	0.022	0.044	0.030				(categorical)
		No	0.969	0.976	0.978	0.956	0.970				
	HUD assistance	Subsidized housing	0.370	0.212	0.059	0.017	0.180				(categorical)
		Rental assistance (voucher)	0.259	0.162	0.068	0.025	0.138				
		No assistance	0.371	0.626	0.873	0.958	0.682				
Region	US census division	Northwest	0.231	0.207	0.188	0.190	0.206				(categorical)
		Midwest	0.164	0.162	0.141	0.101	0.144				
		South	0.374	0.355	0.365	0.314	0.354				
		West	0.231	0.276	0.305	0.395	0.296				

Note. The highlighted categories indicate the highest observation for each variable and serve as the reference group.

# Appendix C

- [AHS Definition of Physical Inadequacy](#)

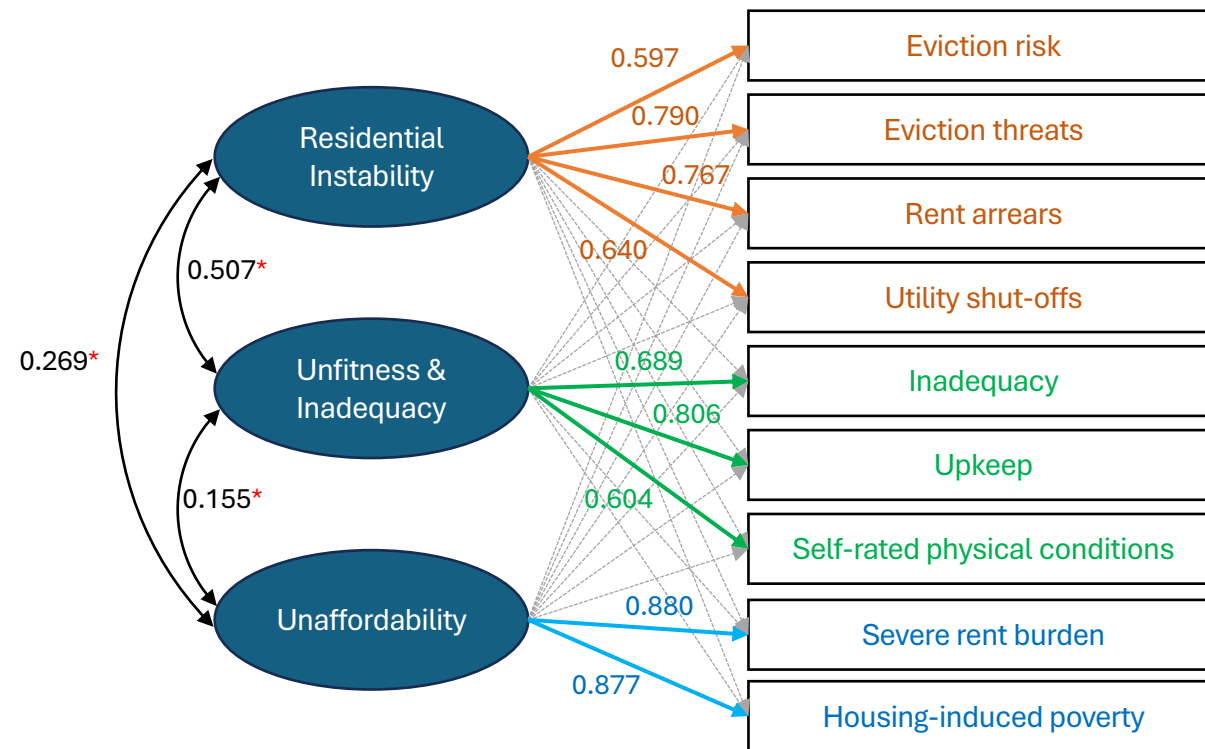
**A housing unit is considered severely inadequate if any of the following criteria apply:**

1. Unit does not have hot and cold running water.
2. Unit does not have a bathtub or shower.
3. Unit does not have a flush toilet.
4. Unit shares plumbing facilities.
5. Unit was cold for 24 hours or more and more than two breakdowns of the heating equipment have occurred that lasted longer than 6 hours.
6. Electricity is not used.
7. Unit has exposed wiring, not every room has working electrical plugs, and the fuses have blown more than twice.
8. Unit has five or six of the following structural conditions:
  - a. Unit has had outside water leaks in the past 12 months.
  - b. Unit has had inside water leaks in the past 12 months.
  - c. Unit has holes in the floor.
  - d. Unit has open cracks wider than a dime.
  - e. Unit has an area of peeling paint larger than 8 by 11 inches.
  - f. Rats have been seen recently in the unit

Source: Eggers and Moumen (2013, p. 2)

# Appendix D

- Visualized EFA result: Relationships between the key dimensions and the indicators



\* Correlation coefficients between factors.

# Appendix E

- PPO models for the very low-income group (1/2)

Variable / Category		0 vs 1,2,3			0,1 vs 2,3			0,1,2 vs 3		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Householder	Race (ref. White)									
	Black	0.246**	0.059	1.279	0.246**	0.059	1.279	0.246**	0.059	1.279
	Asian	-0.053	0.119	0.948	-0.053	0.119	0.948	-0.053	0.119	0.948
	<b>Other races</b>	<b>-0.081</b>	<b>0.186</b>	<b>0.922</b>	<b>-0.108</b>	<b>0.140</b>	<b>0.897</b>	<b>0.406*</b>	<b>0.149</b>	<b>1.501</b>
	Spanish origin	0.099	0.068	1.104	0.099	0.068	1.104	0.099	0.068	1.104
	Non-US citizen	0.146	0.090	1.157	0.146	0.090	1.157	0.146	0.090	1.157
	Age	-0.009**	0.003	0.991	-0.009**	0.003	0.991	-0.009**	0.003	0.991
	Male	-0.103	0.056	0.902	-0.103	0.056	0.902	-0.103	0.056	0.902
	Education level (ref. College)									
	Below high school	-0.075	0.065	0.928	-0.075	0.065	0.928	-0.075	0.065	0.928
	High school	-0.130	0.063	0.878	-0.130	0.063	0.878	-0.130	0.063	0.878
	Bachelor's	-0.094	0.102	0.911	-0.094	0.102	0.911	-0.094	0.102	0.911
	<b>Master's or higher</b>	<b>0.776</b>	<b>0.354</b>	<b>2.172</b>	<b>0.285</b>	<b>0.207</b>	<b>1.330</b>	<b>-0.454</b>	<b>0.230</b>	<b>0.635</b>
	Marital status (ref. Never married)									
	Married	0.100	0.084	1.105	0.100	0.084	1.105	0.100	0.084	1.105
	Separated	0.007	0.060	1.007	0.007	0.060	1.007	0.007	0.060	1.007
Household	Length of stay (ref. Less than 5 years)									
	5 to less than 10 years	-0.112	0.063	0.894	-0.112	0.063	0.894	-0.112	0.063	0.894
	10 years or more	0.041	0.064	1.042	0.041	0.064	1.042	0.041	0.064	1.042

Note. Number of observations = 6830. Log-likelihood at zero = -7319.59. Log-likelihood at convergence = -8181.13. Likelihood ratio  $\chi^2$  (57) = 1723.08 (0.000). **Predictors that violate the proportional odds assumption are highlighted in bold.** \*\* and \* denote significance at the 0.001 and 0.01 levels, respectively.

# Appendix E

- PPO models for the very low-income group (2/2)

(continued)

Variable / Category		0 vs 1,2,3			0,1 vs 2,3			0,1,2 vs 3		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Household	Presence of young children	0.010	0.085	1.010	0.010	0.085	1.010	0.010	0.085	1.010
	<b>Presence of seniors</b>	<b>0.004</b>	<b>0.102</b>	<b>1.004</b>	<b>-0.083</b>	<b>0.087</b>	<b>0.920</b>	<b>-0.483**</b>	<b>0.106</b>	<b>0.617</b>
	<b>Presence of disabled members</b>	<b>-0.149</b>	<b>0.081</b>	<b>0.861</b>	<b>-0.344**</b>	<b>0.062</b>	<b>0.709</b>	<b>0.348**</b>	<b>0.074</b>	<b>1.416</b>
	<b>Number of household members</b>	<b>-0.077</b>	<b>0.033</b>	<b>0.926</b>	<b>-0.116**</b>	<b>0.024</b>	<b>0.891</b>	<b>0.042</b>	<b>0.025</b>	<b>1.043</b>
House	Age of house	0.003**	0.001	1.003	0.003**	0.001	1.003	0.003**	0.001	1.003
	Unit type (ref. Apartment)									
	<b>One-family house</b>	<b>0.612**</b>	<b>0.118</b>	<b>1.844</b>	<b>0.429**</b>	<b>0.078</b>	<b>1.535</b>	<b>0.128</b>	<b>0.079</b>	<b>1.137</b>
	Mobile house	0.295	0.200	1.344	0.295	0.200	1.344	0.295	0.200	1.344
Community	<b>No good schools</b>	<b>0.004</b>	<b>0.125</b>	<b>1.004</b>	<b>0.084</b>	<b>0.093</b>	<b>1.088</b>	<b>0.484**</b>	<b>0.092</b>	<b>1.623</b>
	No good public transportation	-0.044	0.059	0.957	-0.044	0.059	0.957	-0.044	0.059	0.957
	<b>High levels of petty crime</b>	<b>0.082</b>	<b>0.090</b>	<b>1.085</b>	<b>-0.014</b>	<b>0.071</b>	<b>0.986</b>	<b>0.564**</b>	<b>0.080</b>	<b>1.758</b>
	High levels of serious crime	0.431**	0.072	1.538	0.431**	0.072	1.538	0.431**	0.072	1.538
	<b>High disaster risk</b>	<b>-0.071</b>	<b>0.121</b>	<b>0.931</b>	<b>0.188</b>	<b>0.091</b>	<b>1.207</b>	<b>0.427**</b>	<b>0.091</b>	<b>1.533</b>
Housing policy	Rent control	-0.168	0.144	0.845	-0.168	0.144	0.845	-0.168	0.144	0.845
	HUD assistance (ref. No assistance)									
	<b>Subsidized housing</b>	<b>-2.116**</b>	<b>0.143</b>	<b>0.120</b>	<b>-2.074**</b>	<b>0.087</b>	<b>0.126</b>	<b>-0.702**</b>	<b>0.086</b>	<b>0.496</b>
	<b>Rental assistance (voucher)</b>	<b>-1.722**</b>	<b>0.148</b>	<b>0.179</b>	<b>-1.570**</b>	<b>0.090</b>	<b>0.208</b>	<b>-0.599**</b>	<b>0.091</b>	<b>0.550</b>
Region	US census division (ref. South)									
	Northeast	0.131	0.069	1.140	0.131	0.069	1.140	0.131	0.069	1.140
	Midwest	-0.161	0.073	0.852	-0.161	0.073	0.852	-0.161	0.073	0.852
	<b>West</b>	<b>0.012</b>	<b>0.103</b>	<b>1.012</b>	<b>0.040</b>	<b>0.081</b>	<b>1.041</b>	<b>-0.262*</b>	<b>0.092</b>	<b>0.770</b>
(intercept)		3.805**	0.215	44.916	2.746**	0.168	15.576	-1.545**	0.159	0.213

# Appendix E

- PPO models for the low-income group (1/2)

Variable / Category		0 vs 1,2,3			0,1 vs 2,3			0,1,2 vs 3		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Householder	Race (ref. White)									
	Black	0.249**	0.068	1.283	0.249**	0.068	1.283	0.249**	0.068	1.283
	Asian	0.029	0.134	1.030	0.029	0.134	1.030	0.029	0.134	1.030
	Other races	0.275	0.133	1.317	0.275	0.133	1.317	0.275	0.133	1.317
	Spanish origin	-0.036	0.077	0.965	-0.036	0.077	0.965	-0.036	0.077	0.965
	Non-US citizen	0.133	0.092	1.142	0.133	0.092	1.142	0.133	0.092	1.142
	<b>Age</b>	<b>0.004</b>	<b>0.003</b>	<b>1.004</b>	<b>0.001</b>	<b>0.003</b>	<b>1.001</b>	<b>-0.011</b>	<b>0.005</b>	<b>0.989</b>
	Male	-0.052	0.059	0.950	-0.052	0.059	0.950	-0.052	0.059	0.950
	Education level (ref. College)									
	Below high school	-0.061	0.080	0.941	-0.061	0.080	0.941	-0.061	0.080	0.941
	High school	-0.053	0.068	0.949	-0.053	0.068	0.949	-0.053	0.068	0.949
	Bachelor's	0.069	0.098	1.072	0.069	0.098	1.072	0.069	0.098	1.072
	Master's or higher	0.268	0.140	1.307	0.268	0.140	1.307	0.268	0.140	1.307
	Marital status (ref. Never married)									
	Married	-0.265*	0.086	0.767	-0.265*	0.086	0.767	-0.265*	0.086	0.767
	Separated	0.030	0.074	1.030	0.030	0.074	1.030	0.030	0.074	1.030
Household	Length of stay (ref. Less than 5 years)									
	5 to less than 10 years	-0.200*	0.071	0.819	-0.200*	0.071	0.819	-0.200*	0.071	0.819
	10 years or more	-0.170	0.076	0.843	-0.170	0.076	0.843	-0.170	0.076	0.843

Note. Number of observations = 4887. Log-likelihood at zero = -5585.91. Log-likelihood at convergence = -5879.94. Likelihood ratio  $\chi^2$  (43) = 588.07 (0.000). **Predictors that violate the proportional odds assumption are highlighted in bold.** \*\* and \* denote significance at the 0.001 and 0.01 levels, respectively.

# Appendix E

- PPO models for the low-income group (2/2)

(continued)

Variable / Category		0 vs 1,2,3			0,1 vs 2,3			0,1,2 vs 3		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Household	Presence of young children	-0.028	0.094	0.972	-0.028	0.094	0.972	-0.028	0.094	0.972
	Presence of seniors	-0.330**	0.095	0.719	-0.330**	0.095	0.719	-0.330**	0.095	0.719
	<b>Presence of disabled members</b>	<b>0.208*</b>	<b>0.071</b>	<b>1.231</b>	<b>0.160</b>	<b>0.080</b>	<b>1.174</b>	<b>0.595**</b>	<b>0.143</b>	<b>1.812</b>
	<b>Number of household members</b>	<b>-0.132**</b>	<b>0.026</b>	<b>0.876</b>	<b>-0.063</b>	<b>0.027</b>	<b>0.939</b>	<b>-0.047</b>	<b>0.043</b>	<b>0.954</b>
House	<b>Age of house</b>	<b>0.000</b>	<b>0.001</b>	<b>1.000</b>	<b>0.002</b>	<b>0.001</b>	<b>1.002</b>	<b>0.009**</b>	<b>0.002</b>	<b>1.009</b>
	Unit type (ref. Apartment)									
	One-family house	0.066	0.066	1.068	0.066	0.066	1.068	0.066	0.066	1.068
	Mobile house	-0.882**	0.211	0.414	-0.882**	0.211	0.414	-0.882**	0.211	0.414
Community	No good schools	0.296*	0.094	1.345	0.296*	0.094	1.345	0.296*	0.094	1.345
	No good public transportation	-0.077	0.064	0.926	-0.077	0.064	0.926	-0.077	0.064	0.926
	High levels of petty crime	0.143	0.071	1.154	0.143	0.071	1.154	0.143	0.071	1.154
	<b>High levels of serious crime</b>	<b>0.283*</b>	<b>0.101</b>	<b>1.327</b>	<b>0.516**</b>	<b>0.103</b>	<b>1.675</b>	<b>1.044**</b>	<b>0.149</b>	<b>2.841</b>
	High disaster risk	0.317**	0.091	1.373	0.317**	0.091	1.373	0.317**	0.091	1.373
Housing policy	Rent control	0.417	0.173	1.517	0.417	0.173	1.517	0.417	0.173	1.517
	HUD assistance (ref. No assistance)									
	Subsidized housing	-1.076**	0.080	0.341	-1.076**	0.080	0.341	-1.076**	0.080	0.341
	Rental assistance (voucher)	-0.709**	0.083	0.492	-0.709**	0.083	0.492	-0.709**	0.083	0.492
Region	US census division (ref. South)									
	Northeast	0.174	0.082	1.189	0.174	0.082	1.189	0.174	0.082	1.189
	Midwest	-0.290**	0.086	0.748	-0.290**	0.086	0.748	-0.290**	0.086	0.748
	West	0.203*	0.073	1.225	0.203*	0.073	1.225	0.203*	0.073	1.225
(intercept)		0.735**	0.169	2.085	-0.927**	0.176	0.396	-2.969**	0.290	0.051



# Appendix E

- PPO models for the modest-income group (1/2)

Variable / Category		0 vs 1,2			0,1 vs 2		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Householder	Race (ref. White)						
	Black	0.212	0.083	1.236	0.212	0.083	1.236
	Asian	0.205	0.141	1.228	0.205	0.141	1.228
	Other races	0.473*	0.153	1.604	0.473*	0.153	1.604
	Spanish origin	-0.023	0.089	0.977	-0.023	0.089	0.977
	Non-US citizen	0.038	0.106	1.039	0.038	0.106	1.039
	Age	0.003	0.003	1.003	0.003	0.003	1.003
	Male	-0.023	0.067	0.977	-0.023	0.067	0.977
	Education level (ref. College)						
	Below high school	0.007	0.115	1.007	0.007	0.115	1.007
	High school	0.004	0.084	1.004	0.004	0.084	1.004
	Bachelor's	0.053	0.094	1.055	0.053	0.094	1.055
	Master's or higher	0.288	0.121	1.334	0.288	0.121	1.334
	Marital status (ref. Never married)						
	Married	-0.112	0.095	0.894	-0.112	0.095	0.894
	Separated	-0.053	0.092	0.949	-0.053	0.092	0.949
Household	Length of stay (ref. Less than 5 years)						
	5 to less than 10 years	-0.047	0.089	0.954	-0.047	0.089	0.954
	10 years or more	-0.093	0.099	0.911	-0.093	0.099	0.911

Note. Number of observations = 5318. Log-likelihood at zero = -3658.40. Log-likelihood at convergence = -3486.14. Likelihood ratio  $\chi^2$  (36) = 344.53 (0.000). **Predictors that violate the proportional odds assumption are highlighted in bold.** \*\* and \* denote significance at the 0.001 and 0.01 levels, respectively.

# Appendix E

- PPO models for the modest-income group (2/2)

(continued)

Variable / Category		0 vs 1,2			0,1 vs 2		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Household	Presence of young children	0.106	0.118	1.112	0.106	0.118	1.112
	<b>Presence of seniors</b>	<b>-0.005</b>	<b>0.115</b>	<b>0.995</b>	<b>-0.492</b>	<b>0.209</b>	<b>0.611</b>
	Presence of disabled members	0.327**	0.084	1.387	0.327**	0.084	1.387
	Number of household members	-0.156**	0.033	0.856	-0.156**	0.033	0.856
House	Age of house	-0.001	0.001	0.999	-0.001	0.001	0.999
	Unit type (ref. Apartment)						
	One-family house	0.233*	0.077	1.263	0.233*	0.077	1.263
	Mobile house	0.191	0.225	1.211	0.191	0.225	1.211
Community	No good schools	0.308*	0.108	1.360	0.308*	0.108	1.360
	No good public transportation	-0.168	0.076	0.845	-0.168	0.076	0.845
	High levels of petty crime	0.145	0.084	1.156	0.145	0.084	1.156
	<b>High levels of serious crime</b>	<b>0.612**</b>	<b>0.111</b>	<b>1.844</b>	<b>1.392**</b>	<b>0.154</b>	<b>4.023</b>
	<b>High disaster risk</b>	<b>0.214</b>	<b>0.108</b>	<b>1.238</b>	<b>0.830**</b>	<b>0.166</b>	<b>2.293</b>
Housing policy	Rent control	0.266	0.204	1.305	0.266	0.204	1.305
	HUD assistance (ref. No assistance)						
	Subsidized housing	-0.100	0.141	0.905	-0.100	0.141	0.905
	Rental assistance (voucher)	-0.077	0.130	0.926	-0.077	0.130	0.926
Region	US census division (ref. South)						
	Northeast	0.512**	0.100	1.668	0.512**	0.100	1.668
	Midwest	-0.338*	0.115	0.713	-0.338*	0.115	0.713
	West	0.288**	0.084	1.333	0.288**	0.084	1.333
(intercept)		-1.281**	0.173	0.278	-3.495**	0.192	0.030

# Appendix E

- PPO models for the higher-income group (1/2)

Variable / Category		0 vs 1,2			0,1 vs 2		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Householder	Race (ref. White)						
	Black	0.426**	0.129	1.531	0.426**	0.129	1.531
	Asian	-0.091	0.183	0.913	-0.091	0.183	0.913
	Other races	0.222	0.233	1.248	0.222	0.233	1.248
	Spanish origin	0.172	0.134	1.188	0.172	0.134	1.188
	Non-US citizen	0.024	0.162	1.024	0.024	0.162	1.024
	Age	0.001	0.005	1.001	0.001	0.005	1.001
	Male	-0.040	0.097	0.961	-0.040	0.097	0.961
	Education level (ref. College)						
	Below high school	0.258	0.229	1.295	0.258	0.229	1.295
	High school	-0.039	0.151	0.962	-0.039	0.151	0.962
	<b>Bachelor's</b>	<b>-0.180</b>	<b>0.132</b>	<b>0.836</b>	<b>-0.943*</b>	<b>0.313</b>	<b>0.389</b>
	Master's or higher	0.007	0.144	1.007	0.007	0.144	1.007
	Marital status (ref. Never married)						
	Married	-0.095	0.129	0.909	-0.095	0.129	0.909
	Separated	-0.011	0.140	0.989	-0.011	0.140	0.989
Household	Length of stay (ref. Less than 5 years)						
	5 to less than 10 years	0.093	0.132	1.097	0.093	0.132	1.097
	10 years or more	0.302	0.149	1.352	0.302	0.149	1.352

Note. Number of observations = 4992. Log-likelihood at zero = -1915.21 . Log-likelihood at convergence = -1791.92. Likelihood ratio  $\chi^2$  (36) = 246.58 (0.000). **Predictors that violate the proportional odds assumption are highlighted in bold.** \*\* and \* denote significance at the 0.001 and 0.01 levels, respectively.

# Appendix E

- PPO models for the higher-income group (2/2)

(continued)

Variable / Category		0 vs 1,2			0,1 vs 2		
		Coef.	S.E.	Relative Risk Ratio	Coef.	S.E.	Relative Risk Ratio
Household	Presence of young children	-0.030	0.213	0.971	-0.030	0.213	0.971
	Presence of seniors	0.275	0.178	1.317	0.275	0.178	1.317
	Presence of disabled members	0.532**	0.138	1.702	0.532**	0.138	1.702
	Number of household members	-0.049	0.053	0.952	-0.049	0.053	0.952
House	Age of house	0.006**	0.002	1.006	0.006**	0.002	1.006
	Unit type (ref. Apartment)						
	One-family house	0.156	0.113	1.169	0.156	0.113	1.169
	Mobile house	0.750	0.531	2.117	0.750	0.531	2.117
Community	No good schools	0.507**	0.155	1.661	0.507**	0.155	1.661
	No good public transportation	-0.116	0.113	0.890	-0.116	0.113	0.890
	<b>High levels of petty crime</b>	<b>0.297</b>	<b>0.127</b>	<b>1.346</b>	<b>1.017**</b>	<b>0.234</b>	<b>2.766</b>
	High levels of serious crime	0.684**	0.161	1.982	0.684**	0.161	1.982
	High disaster risk	0.338	0.148	1.402	0.338	0.148	1.402
Housing policy	Rent control	0.165	0.207	1.180	0.165	0.207	1.180
	HUD assistance (ref. No assistance)						
	Subsidized housing	0.399	0.274	1.491	0.399	0.274	1.491
	Rental assistance (voucher)	-0.407	0.304	0.666	-0.407	0.304	0.666
Region	US census division (ref. South)						
	<b>Northeast</b>	<b>0.180</b>	<b>0.148</b>	<b>1.197</b>	<b>0.738*</b>	<b>0.250</b>	<b>2.092</b>
	Midwest	-0.440	0.204	0.644	-0.440	0.204	0.644
	West	0.103	0.123	1.108	0.103	0.123	1.108
(intercept)		-2.854**	0.272	0.058	-5.131**	0.325	0.006

# Appendix F

- Interpretation of the PPO models

- Black renters face greater housing insecurity than White renters across all income levels
  - Black renters compared to White renters, the relative risk of greater housing insecurity increases by 27.9% for very low-income renters, 28.3% for low-income renters, 23.6% for modest-income renters, and 53.1% for higher-income renters
- ‘Other races’ face greater housing insecurity than White renters among modest-income renters.
  - For renters of ‘Other races’ compared to White renters, the relative risk of greater housing insecurity increases by 60.4% among modest-income renters
- Across all income levels, a Master’s degree or higher does not significantly impact housing insecurity
  - However, among very-low-income renters, a Master’s degree or higher may be associated with a reduced risk of severe housing insecurity, though statistically insignificant

# Appendix F

- Interpretation of the PPO models (cont.)
  - Married householders face a lower relative risk of greater housing insecurity only among low-income group
    - For married householders compared to never-married householders, the relative risk of greater housing insecurity decreases by 23.3% among low-income renters
  - Length of stay in the current residence affects housing insecurity only among low-income renters
    - Among low-income renters, households that have stayed in their current residence for five to less than ten years have an 18.1% lower relative risk of greater housing insecurity compared to those who have stayed for less than five years
  - Among very low-, low-, and modest-income renters, the relative risk of greater housing insecurity generally decreases with one additional household member
    - However, the effect of one additional household member is not significant among higher-income renters

# Appendix F

- Interpretation of the PPO models (cont.)
  - Households with seniors face a lower relative risk of housing insecurity among very low- and low-income renters
    - Among very low-income renters, for households with seniors compared to households without seniors, the relative risk of severe housing insecurity (index value  $\geq 3$ ) decreases by 38.3%.
    - Among low-income renters, for households with seniors compared to households without seniors, the relative risk of greater housing insecurity decreases by 28.1%
  - Households with disabled members face a higher relative risk of severe housing insecurity (index value  $\geq 3$ ) across all income levels
    - For household with disabled members compared to households without disabled members, the relative risk of severe housing insecurity increases by 41.6% for very low-income renters, 81.2% for low-income renters, 38.7% for modest-income renters and 70.2% for higher-income renters

# Appendix F

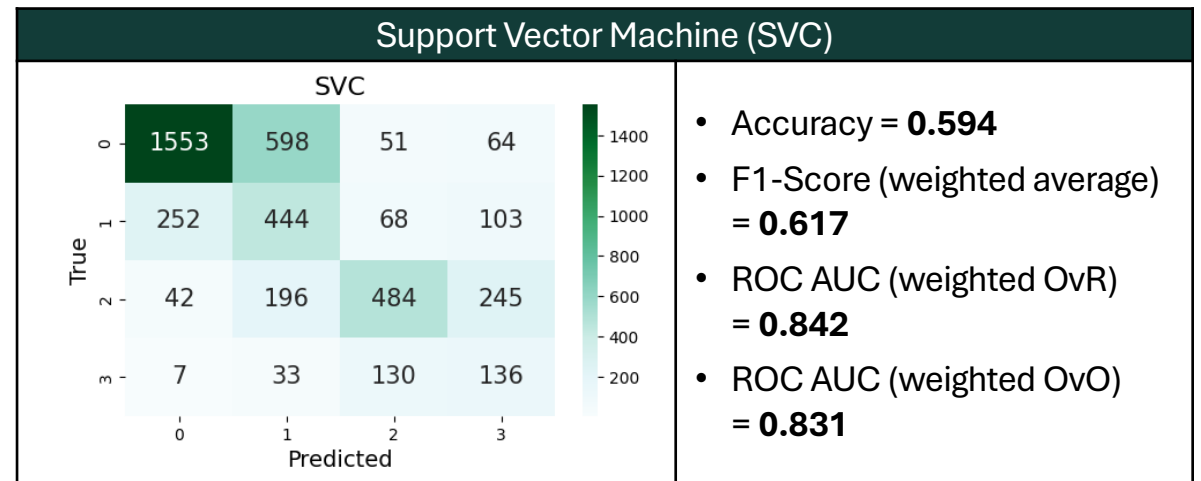
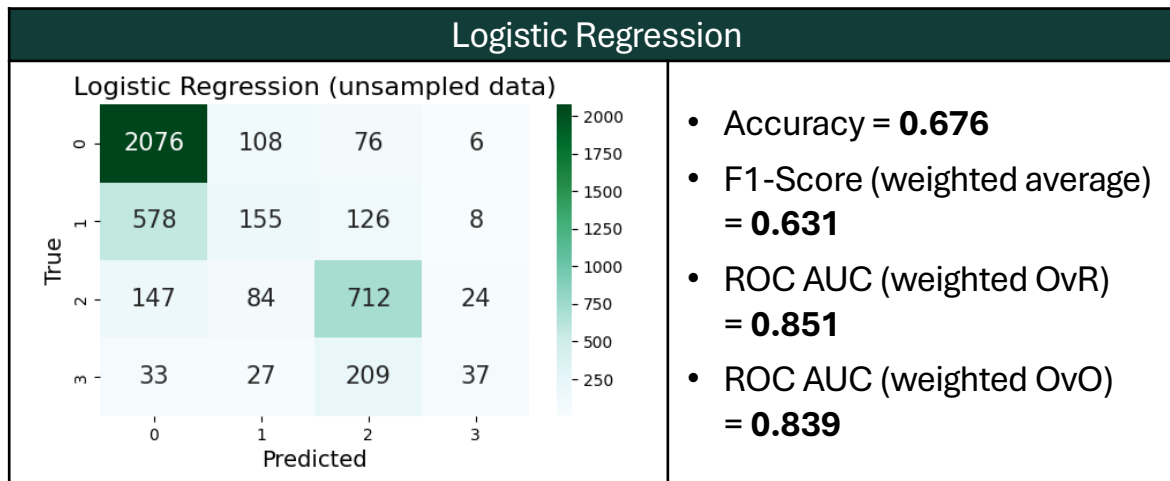
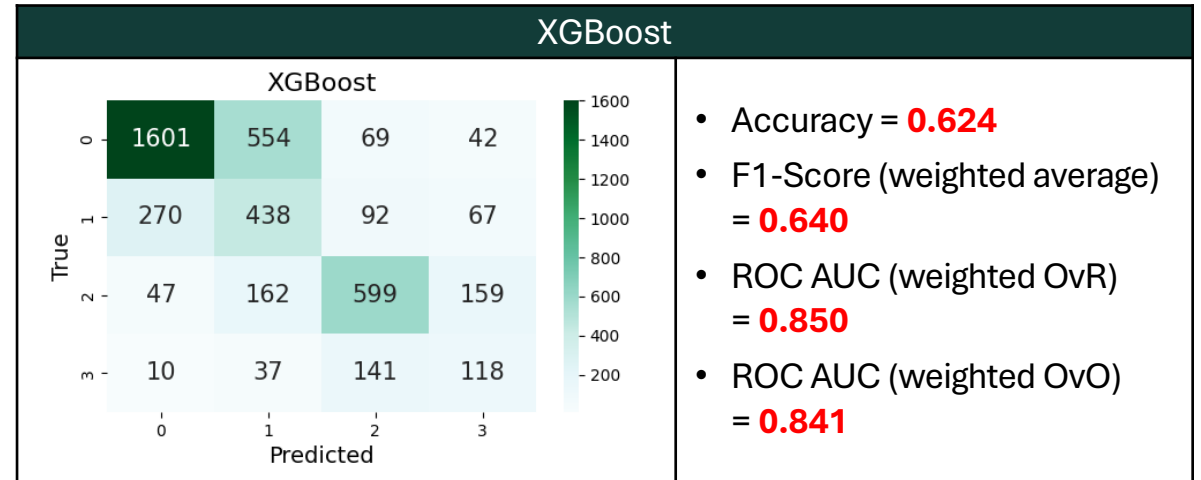
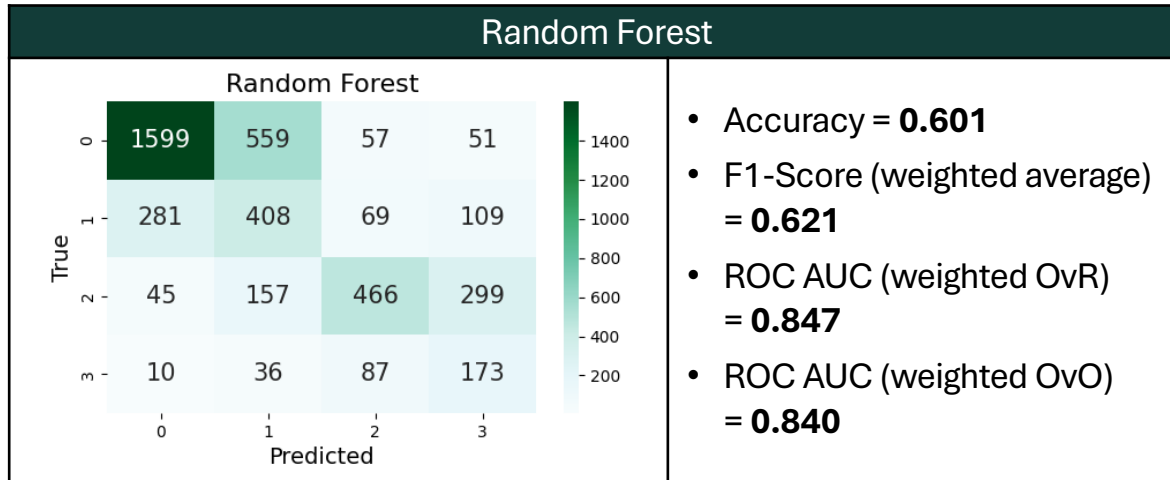
- [Interpretation of the PPO models \(cont.\)](#)

- Across all income levels, renters in communities with high levels of serious crime or no good schools face a higher relative risk of greater housing insecurity
- Among very low- and low-income groups, renters in communities with high disaster risk face a higher relative risk of greater or severe housing insecurity
  - However, the effect of high disaster risk is not significant among higher-income renters
- Subsidized housing and rental assistance (vouchers) greatly reduce housing insecurity for very low- and low-income renters, but have little effect on modest- and higher-income groups
  - For renters in subsidized housing compared to those with no assistance, the relative risk of greater housing insecurity decreases by 50.4% to 88.0% for very low-income renters and by 65% for low-income renters
  - For renters with rental assistance compared to those with no assistance, the relative risk of greater housing insecurity decreases by 45.0% to 82.1% for very low-income renters and by 50.8% for low-income renters



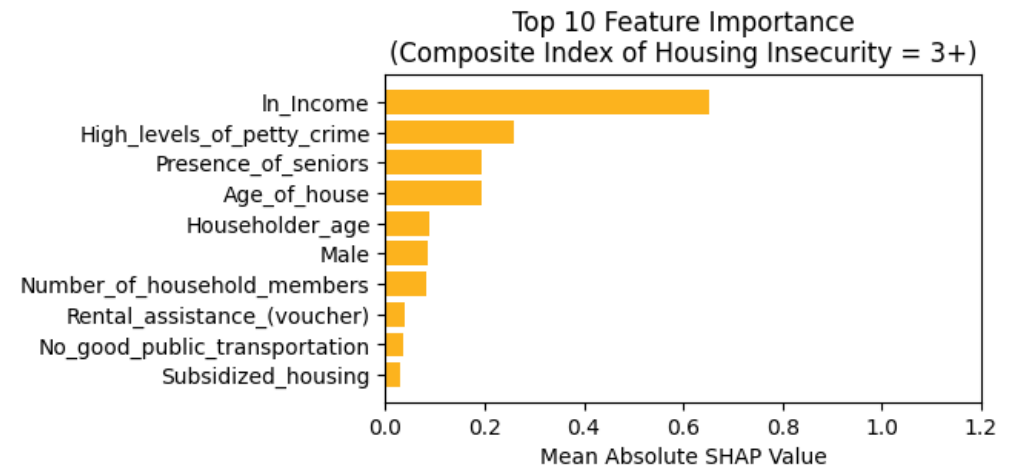
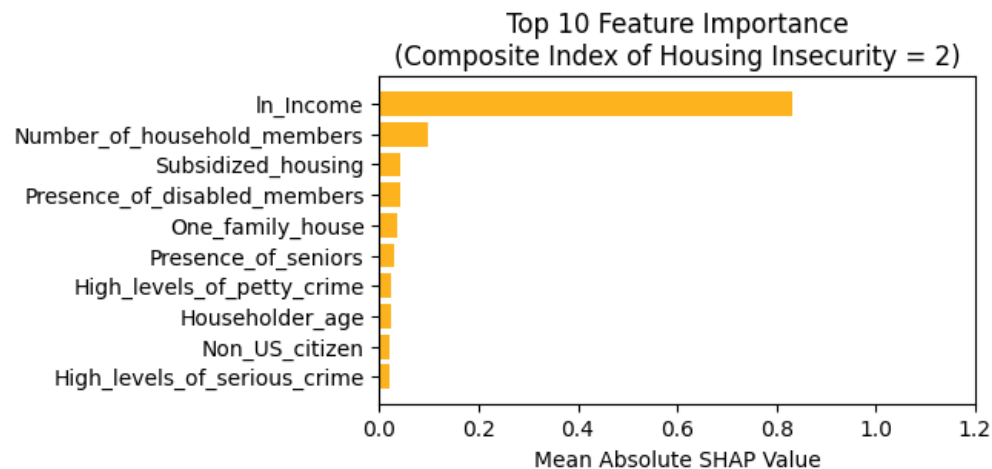
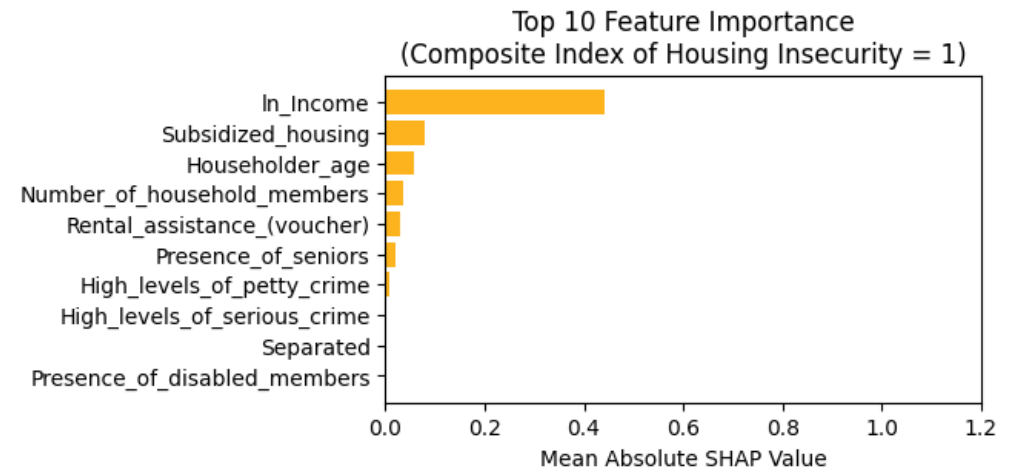
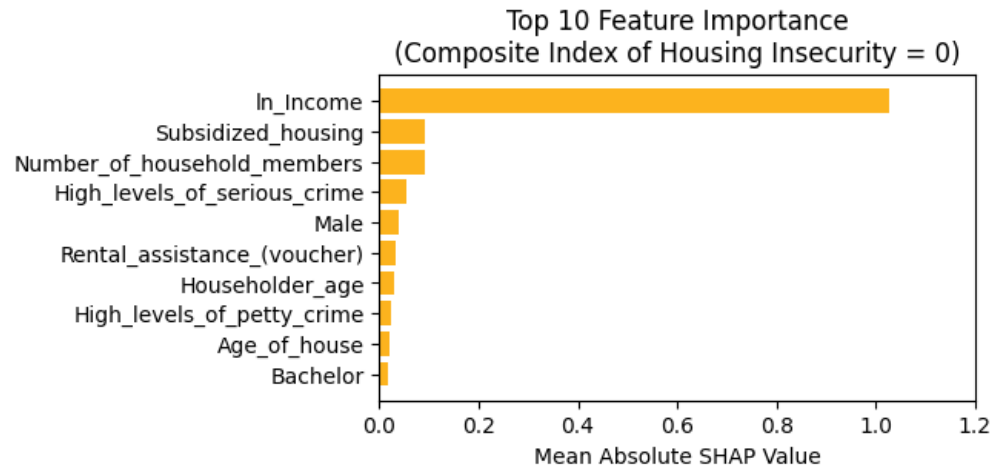
# Appendix G

- Confusion matrices and evaluation metrics on the test set



# Appendix H

- Feature Importance Plots for XGBoost by *SHAP* values



# Appendix H

- Beeswarm Plots for XGBoost by *SHAP* values

