Who Faces Housing Insecurity: An Analysis of Renters in U.S. Metropolitan Areas

IDC 6940 Capstone in Data Science
Student: Geun Sang Oh
Course Instructor: Ananda Mondal
Mentor: Florence George

***Abstract*—This study examines housing insecurity among renters in U.S. metropolitan areas using the 2021 American Housing Survey data. A composite index of housing insecurity was developed through factor analysis, and partial proportional odds (PPO) models identified its predictors across different income levels. Machine learning techniques, including Random Forest, XGBoost, Logistic Regression, and Support Vector Machine, were used to build a predictive model. Findings show that nearly half of renters face housing insecurity, with very low-income renters most vulnerable. Key predictors include race, gender, age, household size, the presence of seniors or disabled members, and community characteristics such as no good schools, high levels of crime, and high disaster risk. Policy interventions like subsidized housing and rental assistance significantly reduce housing insecurity. This study emphasizes the need for targeted policies for very low-income renters and further research on renter and community characteristics to better address severe housing insecurity.**

Keywords—housing insecurity, metropolitan areas, partial proportional odds model, random forest, xgboost, logistic regression, support vector machine

I. Introduction

A. Background

Housing insecurity is a significant issue in U.S. metropolitan areas, with 20% of renters in 25 major cities facing multiple forms of housing insecurity [1]. The number of very low-income renters who pay more than half of their income toward rent without housing assistance has averaged 8 million in the recent decades [2]. People experiencing homelessness rose rapidly from 582,462 to 771,480 between 2022 and 2024 after remaining steady from 2007 to 2021 [3]. Metropolitan areas has particularly high rates of chronic homelessness [3]. In short, renters in U.S. metropolitan areas, particularly those with low income, are increasingly vulnerable to housing insecurity.

B. Motivation and Goals

Housing insecurity is important among renters in U.S. metropolitan areas, especially among low-income households. Renters often face multiple forms of insecurity (e.g., forced moves, poor home conditions, high rent burdens), yet most research focuses on one aspect [1]. This research addresses the severity of housing insecurity by considering its multiple aspects.

The primary goals are to examine the extent and severity of housing insecurity among renters in U.S. metropolitan areas as well as to identify the key characteristics that are strong predictors of housing insecurity. By doing so, this study aims to inform policies that can improve renters' housing security.

C. Research Questions

I explore the following research questions: (1) How extensive and severe is housing insecurity in U.S. metropolitan areas, especially for low-income renters? (2) What characteristics best represent renters facing housing insecurity? (3) Do policy interventions (subsidized housing and rental assistance) contribute to addressing housing insecurity? (4) What characteristics best predict severe housing insecurity?

II. Dataset and Research Approach

A. Data

The data source is the 2021 American Housing Survey (AHS). The AHS is a longitudinal housing unit survey conducted biennially. The AHS, which is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau, collects information on various topics, including home conditions, occupant characteristics, and housing costs. The data can be downloaded from https://www.census.gov/programs-surveys/ahs/data.html. The 2021 AHS dataset includes a total of 64,141 housing units, from which I selected a sample of 22,027 housing units (renter households in metropolitan areas) for this research.

I grouped the sample by income levels, assuming that renter characteristics related to housing insecurity vary by income group. Using the 2021 federal poverty levels as a reference, I set 200% of the poverty level as the low-income baseline [4] and 400% as the modest-income baseline. I also divided the low-income category into two groups—low-income and very low-income—based on the 100% poverty level. Table 1 shows both the 2021 federal poverty levels for different family unit sizes and the corresponding annual income values for 100%, 200%, and 400% of the poverty levels, based on households where all members are aged 18 or older.

Table 2 presents the result of grouping the sample into the four income groups for renters, along with their corresponding observations.

TABLE 2. INCOME LEVEL CATEGORIES FOR RENTERS

Income Level	Income as Percent of Poverty Level	N
Very low-income	≤ 100%	6830
Low-income	> 100% and ≤ 200%	4887
Modest-income	> 200% and ≤ 400%	5318
Higher-income	> 400%	4992
Total		22027

B. Research Approach

I conducted this research in four phases, as shown in Fig. 1. In Phase 1, I explored housing insecurity predictors, which were selected from the AHS data. I presented the summary statistics for the variables (predictors) and conducted exploratory data analysis (EDA) on some key variables.

TABLE 1. 2021 FEDERAL POVERTY LEVELS

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 (unrelated individual):	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										18,145	36,290	72,580
Householder under 65 years	18,231	18,145	18,677								16,379	32,758	65,516
Householder 65 years and over	16,400	16,379	18,606								21,196	42,392	84,784
											27,949	55,898	111,796
Three people	21,559	21,196	21,811	21,831							33,705	67,410	134,820
Four people	27,740	27,949	28,406	27,479	27,575						38,767	77,534	155,068
Five people	32,865	33,705	34,195	33,148	32,338	31,843					44,606	89,212	178,424
Six people	37,161	38,767	38,921	38,119	37,350	36,207	35,529				49,888	99,776	199,552
Seven people	42,156	44,606	44,885	43,925	43,255	42,009	40,554	38,958			60,012	120,024	240,048
Eight people	47,093	49,888	50,329	49,423	48,629	47,503	46,073	44,585	44,207				
Nine people or more	56,325	60,012	60,303	59,501	58,828	57,722	56,201	54,826	54,485	52,386	14,097	28,194	56,388

Note. 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.
Source: U.S. Census Bureau [8]

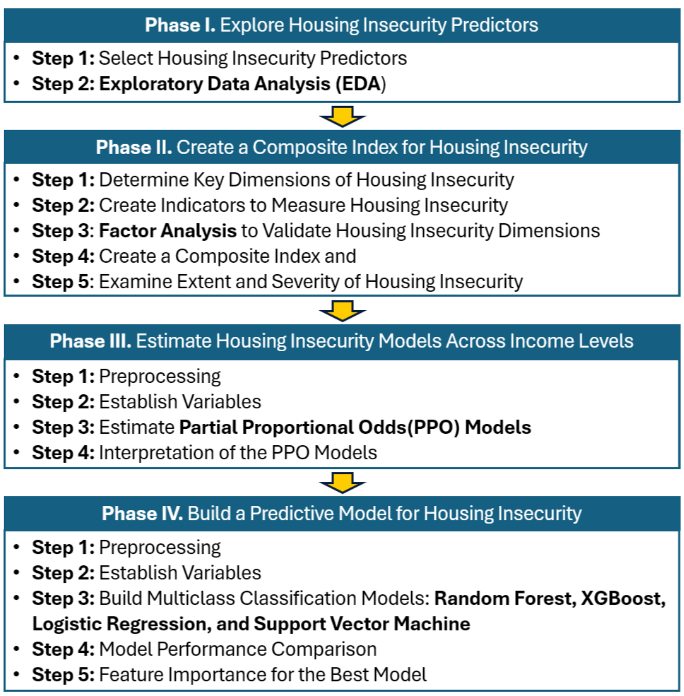

Fig. 1. Research Approach

In Phase 2, I created a composite index of housing insecurity. First, I determined key dimensions and created indicators to measure housing insecurity. I then conducted factor analysis to validate the dimensions and created the composite index by summing the indicators. This index was used to briefly examine the extent and severity of housing insecurity among renters in metropolitan areas. Also, this index underwent preprocessing before being used as a dependent variable in the subsequent phases.

In Phase 3, I estimated partial proportional odds models across four income levels (very low, low, modest, and higher income) to examine the impact of renter characteristics and policy interventions on housing insecurity.

In Phase 4, I built a predictive model by comparing four multiclass classification models (Random Forest, XGBoost, Logistic Regression, and Support Vector Machine) and selecting the best-performing model. I then calculated feature importance for the selected model.

III. PHASE I: EXPLORE HOUSING INSECURITY PREDICTORS

A. *Select Housing Insecurity Predictors*

I selected renter characteristics (including householder and household characteristics), house characteristics, and community characteristics, policy intervention types, and regions as the predictors from the AHS data. Table 3 presents the predictors (variables) along with their summary statistics.

B. *Exploratory Data Analysis (EDA)*

I conducted exploratory data analysis (EDA) on the key predictors (variables) selected from the AHS data.

Fig. 2 shows the distribution of householder's race by income level. White householders make up 61.0% of all renters, while Black householders make up 28.2%. However, as income decreases, the proportion of Black householders increases, while that of White householders decreases.

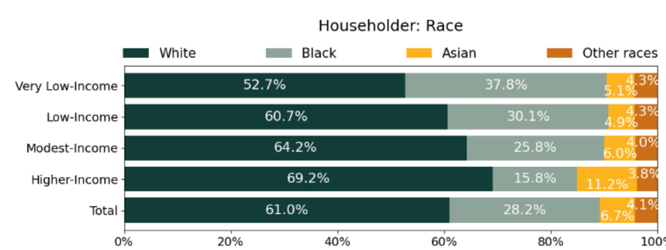

Fig. 2. Distribution of Householder's Race by Income Level

Fig. 3 indicates the distribution of householder's gender by income level. Female householders make up 58.5% of all renters. As income decreases, the proportion of female householders increases, while that of male householders decreases.

TABLE 3. VARIABLE SUMMARY STATISTICS

Variable		Category [Description for Continuous Variables]	Proportion [Mean for Continuous Variables]					Min-Max / S.D.
			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
	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 (incl. 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	
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
House	Age of house	Years since the house was built	53.5	52.3	50.1	47.4	51.0	0-102 / 27.1
	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, trailer, etc.	0.017	0.021	0.023	0.006	0.017	
Community (Self-Rated)	No good school	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	
Policy Intervention	Rent Controlled	Yes	0.031	0.024	0.022	0.044	0.030	(categorical)
		No	0.969	0.976	0.978	0.956	0.970	
	HUD Assisted	Subsidized unit (non-voucher)	0.370	0.212	0.059	0.017	0.180	(categorical)
		Rent-assisted unit (voucher)	0.259	0.162	0.068	0.025	0.138	
		Unassisted unit	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.

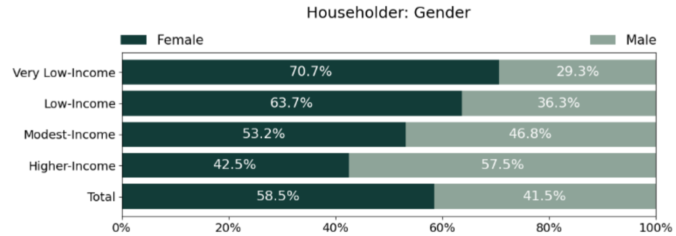

Fig. 3. Distribution of Householder's Gender by Income Level

Fig. 4 presents the distribution of householder's marital status by income level. 41.0% of all renters are never-married householders, 33.8% are separated (including widowed and divorced), and 25.3% are married householders. As income decreases, the proportion of separated householders increases, while that of married householders decreases.

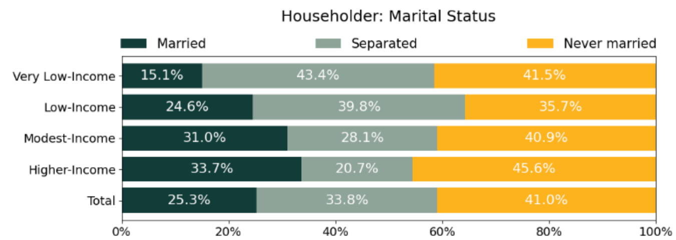

Fig. 4. Distribution of Householder's Marital Status by Income Level

Fig. 5 shows that the histograms of householder's age by income level. Unlike the very low- and low-income groups, where the age distribution is more evenly spread, the modest- and higher-income groups have a relatively higher proportion of younger householders, as clearly shown by the distinct differences in the histograms.

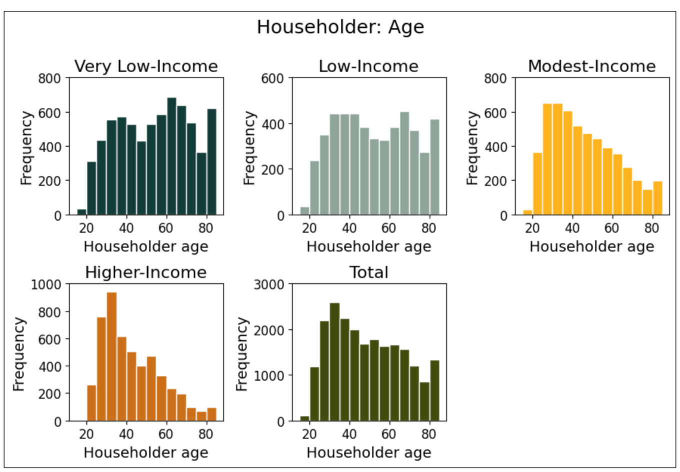

Fig. 5. Histograms of Householder's Age by Income Level

Fig. 6 indicates that the distribution of householder's education level by income level. Among all renters, 28.5% have college education, followed by 26.9% with high school education, 18.2% with education below high school, 17.3% with bachelor's degree, and 9.1% with master's degree or higher. As income decreases, the proportion of renters with high school education or less increases. Conversely, as income increases, the proportion with bachelor's degree or higher increases.

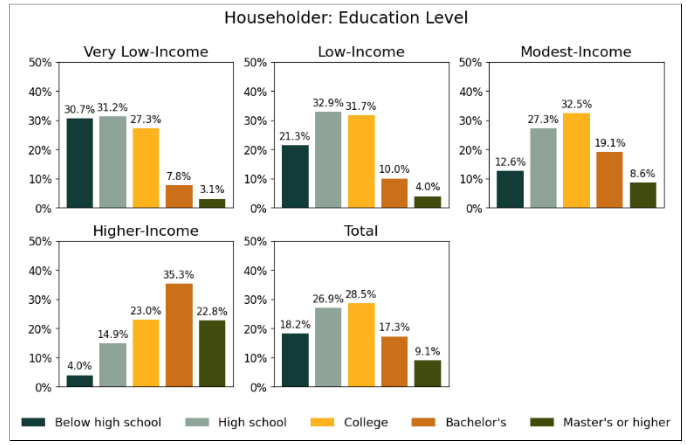

Fig. 6. Distribution of Householder's Education Level by Income Level

Fig. 7 shows the plots for household characteristics. Regarding length of stay in the current residence, 60.7% of renters have lived in their home for less than 5 years, 19.7% for 5 to 10 years, and 19.6% for more than 10 years. As income decreases, the proportion of renters with less than 5 years of stay declines, while the proportion with more than 10 years of stay increases. This suggests lower-income renters have lower residential mobility.

The proportion of renters with young children (under age 6) or seniors (age 65 or older) generally decreases as income rises. The proportion of renters with disabled members varies significantly across income groups: 42.7% in the very low-income group, 34.8% in the low-income group, 18.9% in the modest-income group, and 10.6% in the higher-income group. The average number of household members is highest in the modest-income group (2.36) and lowest in the higher-income group (2.06).

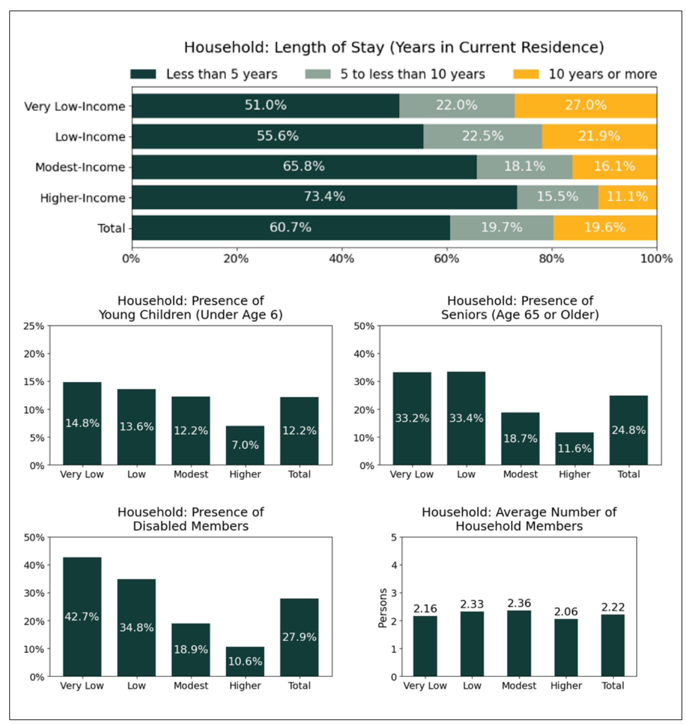

Fig. 7. Distributions of Household Characteristics by Income Level

Fig. 8 presents the plots for community characteristics. As income decreases, the proportion of renters living in communities with negative characteristics—such as no good schools, high levels of petty crime, high levels of serious crime, and high disaster risk—tends to increase. However, it is noteworthy that the proportion of renters living in communities with no good public transportation increases as income rises.

Fig. 8. Distributions of Community Characteristics by Income Level

Fig. 9 displays the distributions of policy interventions, including rent-controlled and HUD-assisted units. Only 3% of renters live in rent-controlled units, with the highest proportion in the higher-income group. This suggests a shortage of rent-controlled units, and those available may not be allocated to the most appropriate target population. In contrast, HUD-assisted units are primarily occupied by renters in lower-income renters. In the low-income group, 37.4% live in subsidized (21.2%) or rent-assisted units (16.2%), while in the very low-income group, 62.9% reside in subsidized (37.0%) or rent-assisted units (25.9%).

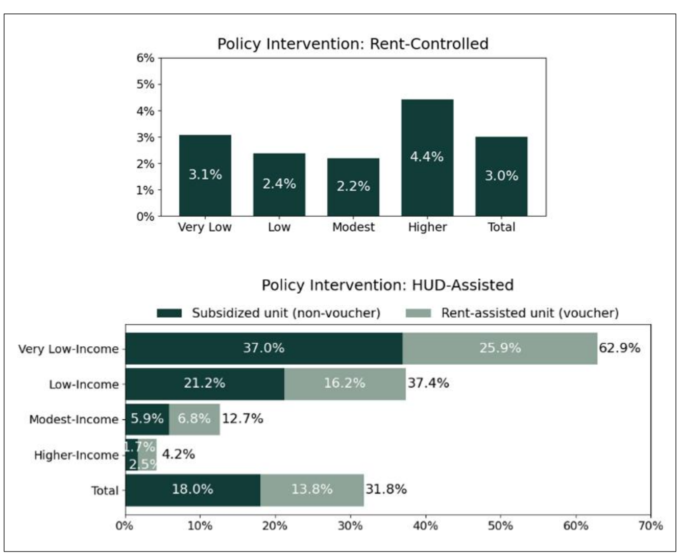

Fig. 9. Distributions of Policy Interventions by Income Level

IV. Phase II: Create a Composite Index of Housing Insecurity

A. Determine Key Dimensions of Housing Insecurity

Housing insecurity refers to a situation where a household lacks one or more elements necessary for "secure" housing, defined in terms of stability, fitness, adequacy, and affordability [5], [6]. Based on a review of the existing literature [1], [5], [6], [7], I determined key dimensions of housing insecurity as *Residential Instability*, *Unfitness and Inadequacy*, and *Unaffordability*, as shown in Fig. 10.

Residential Instability	Unfitness & Inadequacy	Unaffordability
• The household is at significant risk of involuntary displacement for *economic or non-economic reasons*.	• The housing unit *lacks sufficient physical attributes* to meet functional needs for health, security, and daily living activities.	• Housing costs are *not manageable over the long term*, compromising other essential needs for health and well-being.

Fig. 10. Three Key Dimensions of Housing Insecurity

B. Create Indicators to Measure Housing Insecurity

To measure the key dimensions of housing insecurity, I selected several variables from the 2021 AHS dataset based on [1] and transformed them into nine binary indicators: eviction risk, eviction threats, rent arrears, utility shut-off, inadequacy, upkeep, self-rated physical conditions, severe rent burden, and housing-induced poverty, as shown in Table 4.

TABLE 4. BINARY INDICATORS TO MEASURE HOUSING INSECURITY

Key Dimension	Binary Indicator	N	%[a]
Residential Instability	• ***Eviction Risk***: Highly likely to be evicted within 2 months	218	0.99
	• ***Eviction Threats***: Threatened with eviction in the last 3 months	452	2.05
	• ***Rent Arrears***: Unable to pay at least one of the last 3 months	1659	7.53
	• ***Utility Shut-Off***: Had utilities shut-off	196	0.89
Unfitness & Inadequacy	• ***Inadequacy***: Severely inadequate unit (see **Appendix** for details)	607	2.76
	• ***Upkeep***: 3 or more upkeep problems	815	3.70
	• ***Self-Rated Physical Conditions***: 4 or less on a scale of 10	1184	5.38
Unaffordability	• ***Severe Rent Burden***[b]: Gross rent[c] above 50% of income	7019	31.87
	• ***Housing-Induced Poverty***: Income below 133% of federal poverty level[d] (if only gross rent above 30% of income)	7158	32.50

[a] This represents the proportion of renter households in metropolitan areas with a value of 1 for the given indicator (Total N=22027).

[b] 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%.

[c] Gross Rent: Rent plus separate utility costs and related housing expenses (e.g., rental insurance)

[d] 133% of the federal poverty level is the eligibility threshold for Medicaid.

C. Factor Analysis to Validate Housing Insecurity Dimensions

To identify whether the underlying structure of the derived nine indicators aligns with the three dimensions of housing security, I conducted an exploratory factor analysis (EFA) using Stata 17. Since the indicators are binary, I used tetrachoric correlations to construct a correlation matrix. Tetrachoric correlations are specifically designed for binary variables and assume that the underlying relationship between two binary variables follows a bivariate normal distribution [8]. This approach ensures that the underlying structure is properly captured, accounting for the nature of the data, and provides a more reliable basis for validating the dimensions of housing security.

The resulting tetrachoric correlation matrix was then used in the EFA. I applied an oblimin oblique rotation, which allows the factors (dimensions) to be correlated. This rotation method was chosen because there seems to be little reason to require that the three key dimensions of housing insecurity be uncorrelated.

The EFA results confirm the three-dimensional structure is appropriate, with the nine indicators effectively representing the three dimensions. As shown in Fig. 11, the scree plot after EFA has an elbow point at Factor 4. Additionally, according to Kaiser's criterion, which recommends retaining only factors with eigenvalues greater than 1, this criterion is met up to Factor 3. These results support the adequacy of retaining three dimensions.

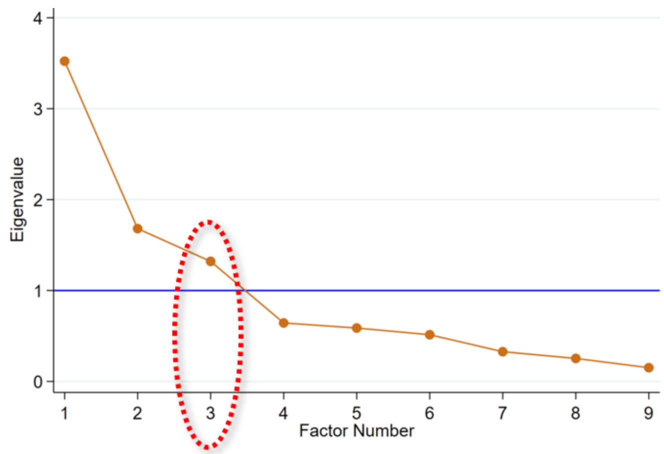

Fig. 11. Scree Plot after Exploratory Factor Analysis

Table 5 demonstrates the validity for the three dimensions, as evidenced by the high factor loadings of the nine indicators.

TABLE 5. FACTOR LOADINGS FOR THE THREE-FACTOR SOLUTION

Dimension	Variable	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-Off	**0.640**	-0.101	0.040	0.632
Unfitness & 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. 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.

Fig. 12 displays the visualized EFA results for the three key dimensions of housing insecurity. Among these dimensions, *Residential Instability* has a correlation coefficient of 0.507 with *Unfitness and Inadequacy*, suggesting a moderate likelihood that renters facing one of these issues may also face the other.

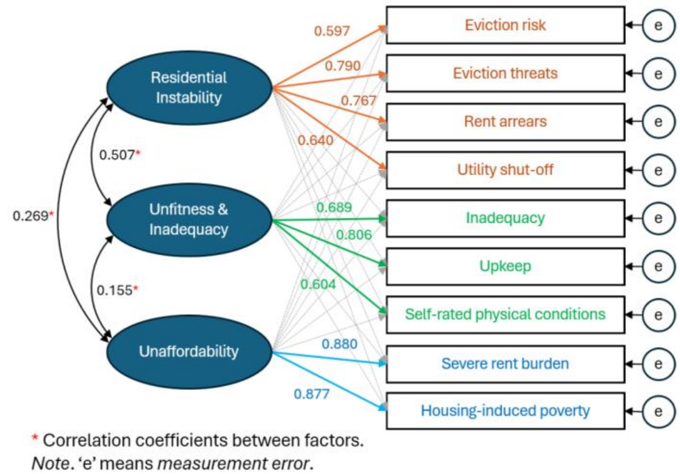

Fig. 12. Visualized Exploratory Factor Analysis Results

All of these results show that the nine indicators support the data structure underlying housing insecurity's three key dimensions and that they can be used to create a composite index of housing insecurity.

D. *Create a Composite Index*

I created the composite index of housing insecurity by summing all nine indicators. The values of this index range from 0 to 8. The higher the index value, the more severe the housing insecurity. A value of 0 is considered secure housing. A value of 1 or higher is considered insecure housing, and a value of 3 or higher can be considered severely insecure housing. Table 6 presents the frequency of renters by the composite index of housing insecurity and income level.

TABLE 6. FREQUENCY OF RENTERS BY THE COMPOSITE INDEX OF HOUSING INSECURITY AND INCOME LEVEL

Index Value		Very low-income	Low-income	Modest-income	Higher-income	Total
0		842	2053	3969	4466	11330
1		1164	1630	1099	440	4333
2		3646	948	180	62	4836
3+	3	778	182	56	21	1037
	4	277	49	9	2	337
	5	79	21	4	1	105
	6	27	1	1	0	29
	7	13	2	0	0	15
	8	4	1	0	0	5
Total		6830	4887	5318	4992	22027

E. *Examine Extent and Severity of Housing Insecurity*

Fig. 13 shows the distribution of renters by the composite index of housing insecurity by income level. Overall, 48.6% of renters face housing insecurity, including 22.0% who face two forms of housing insecurity (index value = 2) and 6.9% who face three or more forms (index value $\geq$ 3). In the higher-income group, 10.5% face housing insecurity, including 1.2% who face two forms and 0.5% face three or more forms. In contrast, in the very low-income group, 87.7% face housing insecurity, including 53.4% who face two forms and 17.2% face three or more forms. These findings highlight the strong relationship between income level and housing insecurity among metropolitan renters. As income decreases, the proportion of renters facing housing insecurity increases. While the extent and severity of housing insecurity increase moderately with a slight decrease in income, they rise sharply as income falls significantly.

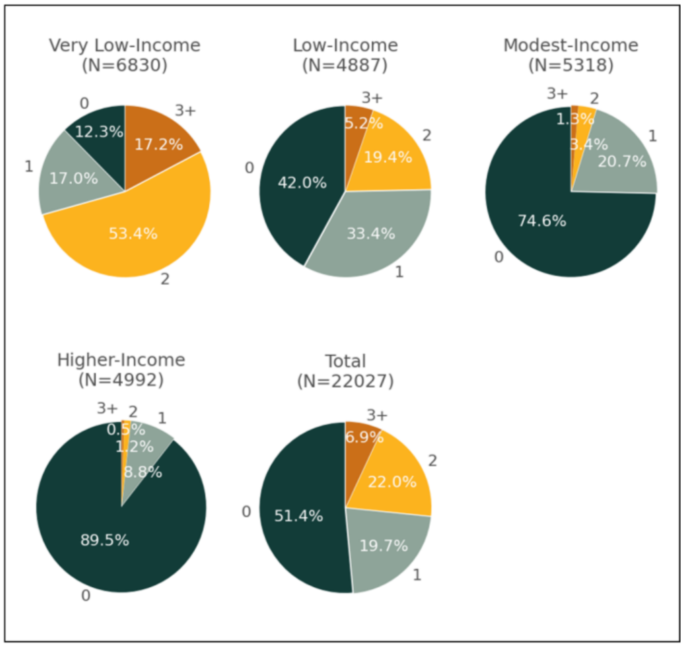

Fig. 13. Distribution of Renters by the Composite Index of Housing Insecurity and Income Level

V. Phase III: Estimate Housing Insecurity Models Across Income Levels

A. Preprocessing

Due to the small number of observations for certain index values in the composite index of housing insecurity, values of 3 and higher were consolidated into a single value of 3 for the very low- and low-income groups. Similarly, for the modest- and higher-income groups, values of 2 and higher were consolidated into a single value of 2. Fig. 14 shows the frequencies of the preprocessed composite index by income group. The index ranges from 0 to 3 for the very low-income and low-income groups, and from 0 to 2 for the modest-income and higher-income groups.

Preprocessed Composite Index for Housing Insecurity

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

Fig. 14. Frequencies of Preprocessed Composite Index by Income Group

B. Establish Variables

I established the dependent and independent variables to examine the impact of renter characteristics and policy interventions on housing insecurity across income levels, as shown in Fig. 15.

Dependent Var.	Preprocessed Composite Index		• **{0, 1, 2, 3}** for Very low-income group (N=6830) • **{0, 1, 2, 3}** for Low-income group (N=4887) • **{0, 1, 2}** for Modest-income group (N=5318) • **{0, 1, 2}** for Higher-income group (N=4992)
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
	Policy Intervention		Rent-controlled, HUD-assisted
	Region		US census division

Note. The numerical variables are underlined.

Fig. 15. Variables for Housing Insecurity Models Across Income Level

C. Estimate Partial Proportional Odds (PPO) Models

Partial proportional odds (PPO) model is an extended ordinal regression model that relaxes the proportional odds assumption for certain predictors, providing greater flexibility when the assumption is violated. The proportional odds assumption posits that the effect of each predictor is consistent across all levels of the ordinal dependent variable in ordered logit or proportional odds (PO) models. This assumption is a strict condition that is rarely met in practice. By allowing predictors to have varying effects at different levels of housing insecurity, the PPO model offers more flexibility.

I applied the PPO model to analyze the impact of renter characteristics and policy interventions on housing insecurity across four income groups, using *gologit2*, a user-written program for Stata [8]. Table 7, 8, 9, and 10 show the PPO models for the very low-income group, the low-income group, the modest-income group, and the higher-income group, respectively.

D. Interpretation of the PPO Models

Householder's race significantly impacts housing insecurity across all income groups. For Black renters compared to White renters, the relative risk of greater housing insecurity increases by 27.9% in the very low-income group, 28.3% in the low-income group, and 53.1% in the higher-income group. For racial minorities ('Other races') compared to the White, the relative risk of greater housing insecurity increases by 60.4% in the modest-income group.

Across all income groups, a master's degree or higher does not significantly impact housing insecurity. However, among the very-low-income group, it may be associated with a reduced risk of severe housing insecurity, though statistically insignificant.

Married householders face a lower housing insecurity only among low-income group. Compared to never-married householders, the relative risk of greater housing insecurity decreases by 23.3% in the low-income group.

Length of stay in the current residence affects housing insecurity only among the low-income group. In this group, renters have lived in their home for 5 to 10 years have an 18.1% lower relative risk of greater housing insecurity compared to those who have lived for less than 5 years.

TABLE 7. PPO MODEL FOR VERY LOW-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***
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
	Other races	**-0.081**	**0.186**	**0.922**	**-0.108**	**0.140**	**0.897**	**0.406***	**0.149**	**1.501**
	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
	Master’s or higher	**0.776**	**0.354**	**2.172**	**0.285**	**0.207**	**1.330**	**-0.454**	**0.230**	**0.635**
	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
Household	Presence of young children	0.010	0.085	1.010	0.010	0.085	1.010	0.010	0.085	1.010
	Presence of seniors	**0.004**	**0.102**	**1.004**	**-0.083**	**0.087**	**0.920**	**-0.483****	**0.106**	**0.617**
	Presence of disabled members	**-0.149**	**0.081**	**0.861**	**-0.344****	**0.062**	**0.709**	**0.348****	**0.074**	**1.416**
	Number of household members	**-0.077**	**0.033**	**0.926**	**-0.116****	**0.024**	**0.891**	**0.042**	**0.025**	**1.043**
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)									
	One-family house	**0.612****	**0.118**	**1.844**	**0.429****	**0.078**	**1.535**	**0.128**	**0.079**	**1.137**
	Mobile house	0.295	0.200	1.344	0.295	0.200	1.344	0.295	0.200	1.344
Community	**No good schools**	**0.004**	**0.125**	**1.004**	**0.084**	**0.093**	**1.088**	**0.484****	**0.092**	**1.623**
	No good public transportation	-0.044	0.059	0.957	-0.044	0.059	0.957	-0.044	0.059	0.957
	High levels of petty crime	**0.082**	**0.090**	**1.085**	**-0.014**	**0.071**	**0.986**	**0.564****	**0.080**	**1.758**
	High levels of serious crime	0.431**	0.072	1.538	0.431**	0.072	1.538	0.431**	0.072	1.538
	High disaster risk	**-0.071**	**0.121**	**0.931**	**0.188**	**0.091**	**1.207**	**0.427****	**0.091**	**1.533**
Policy Intervention	Rent controlled	-0.168	0.144	0.845	-0.168	0.144	0.845	-0.168	0.144	0.845
	HUD assisted (ref. Unassisted unit)									
	Subsidized unit (non-voucher)	**-2.116****	**0.143**	**0.120**	**-2.074****	**0.087**	**0.126**	**-0.702****	**0.086**	**0.496**
	Rent-assisted unit (voucher)	**-1.722****	**0.148**	**0.179**	**-1.570****	**0.090**	**0.208**	**-0.599****	**0.091**	**0.550**
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
	West	**0.012**	**0.103**	**1.012**	**0.040**	**0.081**	**1.041**	**-0.262***	**0.092**	**0.770**
(intercept)		3.805**	0.215	44.916	2.746**	0.168	15.576	-1.545**	0.159	0.213

Note. Number of observations = 6830. Log-likelihood at zero = -7319.59. Log-likelihood at convergence = -8181.13. Likelihood ratio χ^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.

TABLE 8. PPO MODEL FOR LOW-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***
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
	Age	**0.004**	**0.003**	**1.004**	**0.001**	**0.003**	**1.001**	**-0.011**	**0.005**	**0.989**
	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
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
	Presence of disabled members	**0.208***	**0.071**	**1.231**	**0.160**	**0.080**	**1.174**	**0.595****	**0.143**	**1.812**
	Number of household members	**-0.132****	**0.026**	**0.876**	**-0.063**	**0.027**	**0.939**	**-0.047**	**0.043**	**0.954**
House	**Age of house**	**0.000**	**0.001**	**1.000**	**0.002**	**0.001**	**1.002**	**0.009****	**0.002**	**1.009**
	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
	High levels of serious crime	**0.283***	**0.101**	**1.327**	**0.516****	**0.103**	**1.675**	**1.044****	**0.149**	**2.841**
	High disaster risk	0.317**	0.091	1.373	0.317**	0.091	1.373	0.317**	0.091	1.373
Policy Intervention	Rent controlled	0.417	0.173	1.517	0.417	0.173	1.517	0.417	0.173	1.517
	HUD assisted (ref. Unassisted unit)									
	Subsidized unit (non-voucher)	-1.076**	0.080	0.341	-1.076**	0.080	0.341	-1.076**	0.080	0.341
	Rent-assisted unit (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

Note. Number of observations = 4887. Log-likelihood at zero = -5585.91. Log-likelihood at convergence = -5879.94. Likelihood ratio χ^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.

TABLE 9. PPO MODEL FOR MODEST-INCOME GROUP

Variable / Category		0 vs 1,2,3 Coef.	0 vs 1,2,3 S.E.	0 vs 1,2,3 Relative Risk Ratio	0,1 vs 2,3 Coef.	0,1 vs 2,3 S.E.	0,1 vs 2,3 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
Household	Presence of young children	0.106	0.118	1.112	0.106	0.118	1.112
	Presence of seniors	**-0.005**	**0.115**	**0.995**	**-0.492**	**0.209**	**0.611**
	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
	High levels of serious crime	**0.612****	**0.111**	**1.844**	**1.392****	**0.154**	**4.023**
	High disaster risk	**0.214**	**0.108**	**1.238**	**0.830****	**0.166**	**2.293**
Policy Intervention	Rent controlled	0.266	0.204	1.305	0.266	0.204	1.305
	HUD assisted (ref. Unassisted unit)						
	Subsidized unit (non-voucher)	-0.100	0.141	0.905	-0.100	0.141	0.905
	Rent-assisted unit (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

Note. Number of observations = 5318. Log-likelihood at zero = -3658.40. Log-likelihood at convergence = -3486.14. Likelihood ratio χ^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.

TABLE 10. PPO MODEL FOR HIGHER-INCOME GROUP

	`	0 vs 1,2,3			0,1 vs 2,3		
		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
	Bachelor's	**-0.180**	**0.132**	**0.836**	**-0.943***	**0.313**	**0.389**
	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
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
	High levels of petty crime	**0.297**	**0.127**	**1.346**	**1.017****	**0.234**	**2.766**
	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
Policy Intervention	Rent controlled	0.165	0.207	1.180	0.165	0.207	1.180
	HUD assisted (ref. Unassisted unit)						
	Subsidized unit (non-voucher)	0.399	0.274	1.491	0.399	0.274	1.491
	Rent-assisted unit (voucher)	-0.407	0.304	0.666	-0.407	0.304	0.666
Region	US census division (ref. South)						
	Northeast	**0.180**	**0.148**	**1.197**	**0.738***	**0.250**	**2.092**
	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

Note. Number of observations = 4992. Log-likelihood at zero = -1915.21. Log-likelihood at convergence = -1791.92. Likelihood ratio χ^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.

Among the very low-, low-, and modest-income groups, the relative risk of greater housing insecurity generally decreases with each additional household member. However, the effect of each additional household member is not significant in the higher-income group.

Households with seniors face a lower relative risk of housing insecurity among the very low- and low-income groups. Compared to households without seniors, the relative risk of severe housing insecurity (index value ≥ 3) decreases by 38.3% in the very low-income group, while the relative risk of greater housing insecurity decreases by 28.1% in the low-income group.

Households with disabled members face a higher relative risk of severe housing insecurity (index value ≥ 3) across all income groups. Compared to households without disabled members, the relative risk of severe housing insecurity increases by 41.6% in the very low-income group, 81.2% in the low-income group, 38.7% in the modest-income group and 70.2% in the higher-income group.

Across all income groups, renters in communities with high levels of serious crime or no good schools face a higher relative risk of greater housing insecurity. Among the 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, this effect is not significant in the higher-income group.

Subsidized housing and rental assistance greatly reduce housing insecurity for the very low- and low-income groups. For subsidized (non-voucher) units compared to unassisted units, the relative risk of greater housing insecurity decreases by 50.4% to 88.0% in the very low-income group and 65% in the low-income group. For rent-assisted (voucher) units compared to unassisted units, the relative risk of greater housing insecurity decreases by 45.0% to 82.1% in the very low-income group and 50.8% in the low-income group.

VI. PHASE IV: BUILD A PREDICTIVE MODEL FOR HOUSING INSECURITY

A. Preprocessing

I consolidated all four renter groups (very low-, low-, modest-, and higher-income) to build a predictive model, before combining values of 3 and higher in the composite index of housing insecurity were combined into a single value of 3. As a result, the frequencies corresponding to each index value (class frequencies) of the dependent variable are as follows: 11330 for index value = 0, 4333 for index value = 1, 4836 for index value = 2, and 1528 for index value 3.

The independent variables are largely the same as those in Phase III, with the exception that the ‘ln_Income’ variable (the natural logarithm of the ‘Income’) was added and the numerical variables (‘Householder’s age,’ ‘Number of household members,’ ‘ln_Income,’ and ‘Age of house’) were standardized.

Next, I split the entire dataset (N=22027) into a training set (80%, N=17621) and a test set (20%, N=4406). To balance the composite index’s observations across classes. I applied *Random Under-sampling* and *Synthetic Minority Over-sampling Technique Nominal and Continuous features* (*SMOTE-NC*) to the training set. Before the under- and over-samplings, the class frequencies for the composite index of housing insecurity were 9064 for class 0, 3869 for class 1, 3466 for class 2, and 1222 for class 3. After the samplings, the frequencies were balanced to 3869 for each class, increasing the total number of observations in the training set to 15476.

B. Establish Variables

I established the dependent and independent variables to build a predictive model, as shown in Fig. 16.

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
	Policy Intervention		Rent-controlled, HUD-assisted
	Region		US census division

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

Fig. 16. Variables for Building a Predictive Model

C. Buil Multiclass Classification Models

I explored four multiclass classification models—Random Forest, XGBoost, Logistic Regression, and Support Vector Machine (SVM)—to build a single predictive model for housing insecurity. Random Forest is an ensemble learning algorithm that uses multiple decision trees to reduce overfitting and is capable of modeling complex, non-linear relationships. XGBoost, known for its high performance, employs gradient boosting techniques to efficiently handle large datasets while delivering fast and accurate predictions. Logistic Regression is a simple algorithm that is computationally efficient, trains quickly, and performs well as a baseline, even when dealing with a high number of features. Finally, SVM is effective in handling high-dimensional data and works well for both linear and non-linear classification tasks. To implement these models, I used the *scikit-learn* and *XGBoost* libraries in Python.

TABLE 11. 5-FOLD CROSS-VALIDATION RESULTS FOR THE MODELS

Model	Key Hyperparameters	Training & Validation Accuracies
Random Forest	n_estimator = 800 max_depth = 7 max_leaf_nodes = 30 max_features = 0.85 max_samples = 0.20	Train Accuracy=0.615 Valid Accuracy=0.596
XGBoost	n_estimators = 550 learning_rate = 0.01 max_depth = 20 gamma = 15 subsample=0.99 colsample_bytree = 0.79 reg_lambda = 0.05	Train Accuracy=0.642 Valid Accuracy=0.629
Logistic Regression	solver = ‘lbfgs’ C = 10 penalty = ‘l2’	Train Accuracy=0.603 Valid Accuracy=0.592
Support Vector Machine (SVM)	kernel = ‘rbf’ C = 100 gamma = 0.001	Train Accuracy=0.616 Valid Accuracy=0.599

For the under- and over-sampled training set, I used 5-fold cross-validation to optimize hyperparameters and improve the performance of the four models. Table 11 presents the results of hyperparameter tuning for each model, as determined through 5-fold cross-validation. After tuning the hyperparameter, these models were evaluated on the test set to determine the best-performing one for predicting housing insecurity.

D. *Model Performance Comparison*

Fig. 17 displays a performance comparison across the four models, showing the accuracies and F1-scores for both the training and test sets for each model. 'Train Accuracy' and 'Train F1-score (weighted)' were calculated using the entire training set, including the validation set. XGBoost achieved the highest Test Accuracy (0.626) and Test F1-Score (0.641).

Fig. 17. Model Performance Comparison (Accuracy & F1-Score)

Fig. 18, 19, 20, and 21 present the confusion matrix and evaluation metrics for the four models on the test set. XGBoost demonstrates superior performance not only in accuracy and F1-score but also in ROC (Receiver Operating Characteristic) AUC (Area Under the Curve).

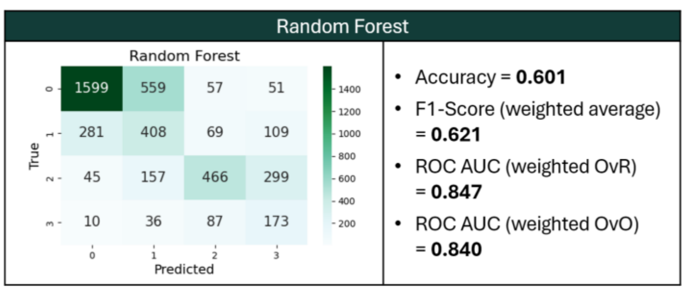

Fig. 18. Random Forest Test Set Evaluation

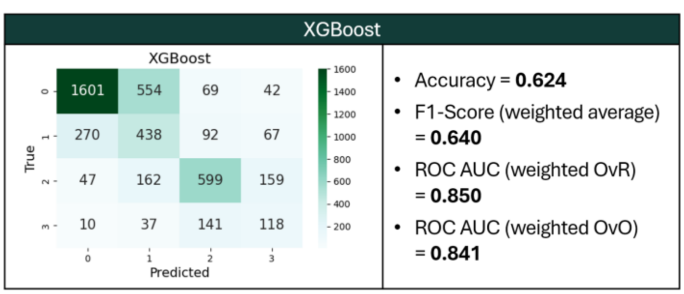

Fig. 19. XGBoost Test Set Evaluation

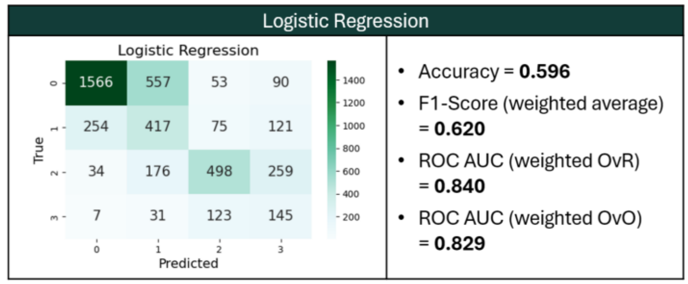

Fig. 20. Logistic Regression Test Set Evaluation

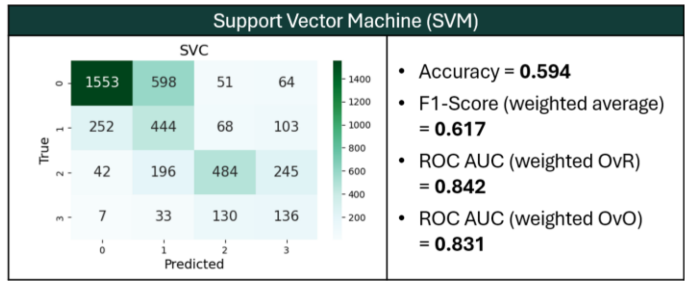

Fig. 21. Support Vector Machine Test Set Evaluation

E. *Feature Importance for the Best Model*

I examined the feature importance for XGBoost, the best-performing model, using SHAP (Shapley Additive exPlanations) values. SHAP is a method for explaining the output of machine learning models by quantifying the contribution of each feature to the model's prediction. SHAP values are calculated based on Shapley values, which represent the weighted average of a feature's marginal contributions across all possible subsets of features. This approach ensures scale-invariance, meaning that SHAP values prevent distortions and misinterpretations caused by scaling features [10]. I used the Python library *shap* to calculate these SHAP values.

Fig. 19 displays the feature importances for housing insecurity across different class values. Specifically, for the composite index of housing insecurity = 0 (class value = 0), the feature importance plot identifies the top 10 features most influential in determining whether renters face housing insecurity. For the composite index of housing insecurity = 3+ (class value = 3), the plot highlights the top 10 features influencing the likelihood of facing severe housing insecurity.

In the feature importance plot for the composite index of housing insecurity = 0, income ('ln_Income') is the most influential feature in determining housing insecurity, followed by 'Subsidized_unit,' 'Number of household members,' 'High levels of serious crime,' 'Householder age,' 'Householder gender,' and 'Rent-assisted unit.' For the composite index of housing insecurity = 3+, income remains the most influential feature in determining severe housing insecurity. However, while the importance of income has relatively decreased, other features such as 'High levels of petty crime,' 'Presence (or absence) of seniors,' 'Age of house,' 'Householder gender and age,' and 'Number of household members' show a greater importance compared to other plots. These findings suggest that, in addition to income, both renter and community characteristics play a key role in severe housing insecurity among renters.

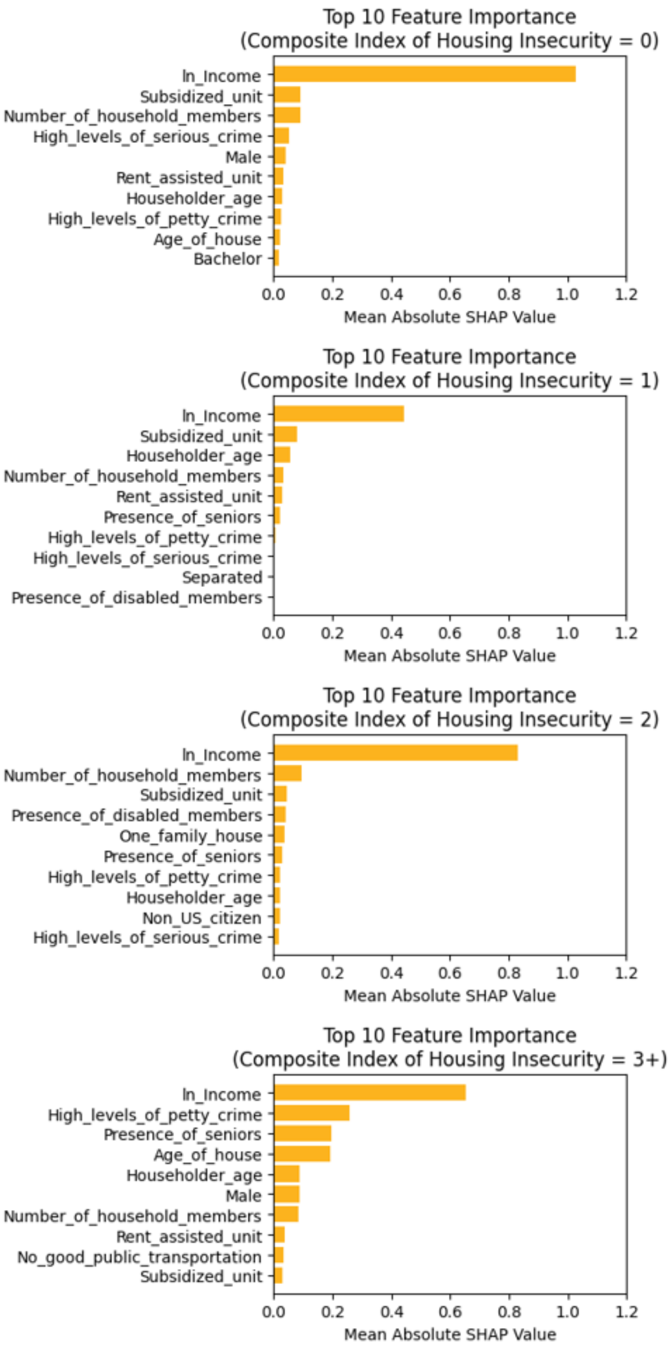

Fig. 22. Feature Importance by Class

VII. Discussion and Conclusion

A. *Discussion #1: How extensive and severe is housing insecurity in U.S. metropolitan areas?*

In U.S. metropolitan areas, nearly half of all renters (48.6%) face some form of housing insecurity: 6.9% facing severe housing insecurity—defined as encountering three or more forms of housing insecurity—while the remaining 41.7% face one or two forms. Among the low-income group, 58.0% face housing insecurity: 5.2% facing severe housing insecurity while remaining 52.8% face one of two forms. Among the very low-income group, 87.7% face housing insecurity: 17.2% facing severe housing insecurity while 70.4% face one or two forms. Given the disproportionately high risk faced by the very low-income renters, policies need to prioritize this group (renters below the poverty level) to address their severe housing insecurity.

B. *Discussion #2: What characteristics best represent renters facing housing insecurity?*

Householder's race (Black or other racial minorities), disabled members, communities with no good schools or high levels of serious crime face greater housing insecurity at all income levels. Among lower-income renters (the very low- or low-income groups), unmarried householders, households without seniors or fewer members, and those living in communities with high disaster risk are more vulnerable.

Length of stay in the current residence appears to be related to housing insecurity only among the low-income group. For renters who have lived in their home for 5 to 10 years face lower housing insecurity, compared to those who have lived for less than 5 years.

Interestingly, having a master's degree or higher does not significantly reduce housing insecurity. However, among the very low-income group, holding a master's degree or higher may be associated with only a lower risk of severe housing insecurity compared to obtaining only a college education, although this relationship is not statistically significant.

C. *Discussion #3: Do policy interventions contribute to addressing housing insecurity?*

Subsidized housing and rental assistance significantly reduce the risk of housing insecurity among the very low- and low-income groups. These two policy interventions are most effective for households on the brink of housing insecurity, rather than for those already face it, particularly within the very low-income group. In this group, renters in subsidized units face an 88.0% reduction in the relative risk of housing insecurity compared to housing security, relative to those in unassisted units. Additionally, renters in rent-assisted units face an 82.1% reduction in the relative risk, relative to those in unassisted units.

D. *Discussion #4: What characteristics best predict severe housing insecurity?*

Income is the most important characteristic in determining both housing security and severe housing insecurity. However, its importance decreases when determining severe housing insecurity compared to housing insecurity. Other renter and community characteristics—such as 'High levels of petty crime,' 'Presence (or absence) of seniors,' 'Householder gender and age,' and 'Number of household members'—contribute more to determining severe housing insecurity than to determining housing insecurity, even though their contribution is still lower than that of income.

This suggests the possibility that renter and community characteristics play a crucial role in renters facing severe housing insecurity. To better understand the characteristics contributing to severe housing insecurity, future research should closely examine the household and community conditions of renters who are facing this issue.

E. *Conclusion*

This research examined housing insecurity among renters in U.S. metropolitan areas and identified key characteristics that influence housing insecurity. The findings indicate that nearly half of renters face housing insecurity, with very low-income renters being particularly vulnerable. Regardless of income level, the characteristics contributing to housing insecurity include the householder's race (Black or racial minority), the presence of disabled members, and communities with no good schools or high levels of serious crime. Among lower-income renters, unmarried status, the absence of seniors, fewer household members, and high disaster-risk communities increase the likelihood of housing insecurity. For severe housing insecurity, compared to (general) housing insecurity, characteristics such as high levels of petty crime, the presence (or absence) of seniors, householder gender and age, and household size may play a more significant role. Subsidized housing and rental assistance are effective in reducing housing insecurity, particularly for lower-income renters.

This research contributes to a deeper understanding of housing insecurity among metropolitan renters, emphasizing the need for policy interventions that prioritize very low-income renters. By identifying renter and community characteristics linked to housing insecurity, this study supports efforts to identify those most at risk. Future research should explore the impact of these characteristics on housing insecurity to provide valuable insights for policy development.

References

[1] G. Routhier, "Beyond Worst Case Needs: Measuring the Breadth and Severity of Housing Insecurity Among Urban Renters," *Housing Policy Debate,* vol. 29, no. 2, pp. 235-249, 2019.

[2] T. A. Alvarez and B. L. Steffen, "Worst Case Housing Needs: 2023 Report to Congress," Office of Policy Development and Research, 2023.

[3] T. de Sousa and M. Henry, "The 2024 Annual Homelessness Assessment Report (AHAR) to Congress: PART 1: Point-in-Time Estimates of Homelessness," Office of Community Planning and Development, 2024.

[4] L. Kilduff, "How Poverty in the United States Is Measured and Why It Matters," 31 1 2022. [Online]. Available: https://www.prb.org/resources/how-poverty-in-the-united-states-is-measured-and-why-it-matters/. [Accessed 20 10 2024].

[5] J. Leopold, M. K. Cunningham, L. Posey and T. Manuel, "Improving Measures of Housing Insecurity: A Path Forward," Urban Institute, 2016.

[6] N. E. Watson and G. R. Carter, "Toward Implementation of a National Housing Insecurity Research Module," *Cityspace,* vol. 22, no. 1, pp. 227-248, 2020.

[7] C. Robynn, S. Rodnyansky, B. Henwood and S. Wenzel, *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, 2017.

[8] StataCorp, Stata 17 Base Reference Manual, College Station: Stata Press, 2021.

[9] R. Williams, "Generalized ordered logit/partial proportional odds models for ordinal dependent variables," *The Stata Journal,* vol. 6, no. 1, pp. 58-82, 2006.

[10] E. Marsh, "Calculating XGBoost Feature Importance," Medium, 31 1 2023. [Online]. Available: https://medium.com/@emilykmarsh/xgboost-feature-importance-233ee27c33a4. [Accessed 3 3 2025].

[11] F. J. Eggers and F. Moumen, "Housing adequacy and quality as measured by the AHS," Econometrica, Inc., 2013.

[12] U.S. Census Bureau, "Poverty Thresholds," 2021. [Online]. Available: https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html. [Accessed 14 10 2024].

Appendix

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 [11]