

# Predictive Model Based on Homelessness



## Team members

Syeda Aqeel

(sharoo@bu.edu)

Shiheng Xu

(timxu5@bu.edu)

Samritha Ravikumar

(samritha@bu.edu)

Kunshu Yang

(luckykun@bu.edu)

Renjie Fan

(renjief@bu.edu)

**BOSTON**  
**UNIVERSITY**



# INTRODUCTION

BOSTON  
UNIVERSITY

## Project Overview:

- Develop a community-level predictive model for homelessness using data from 2007–2023.
- Focus on identifying structural community-level factors (e.g., rent levels, economic conditions) rather than individual-level predictors.
- Analyze approximately 400 Continuums of Care (CoC) funded by the U.S. Department of Housing and Urban Development (HUD).



**Dr. Tom Byrne**

Associate Professor at BU  
Department of Social Welfare  
Policy & Charles River Campus  
Advisor



**Dr. Molly Richard**

Postdoctoral Associate at BU  
Center for Innovation in Social  
Science

## Project Goal:

- Forecast trends in homelessness.
- Identify key community-level predictors influencing homelessness.

## The Big impact:

### Understanding Homelessness:

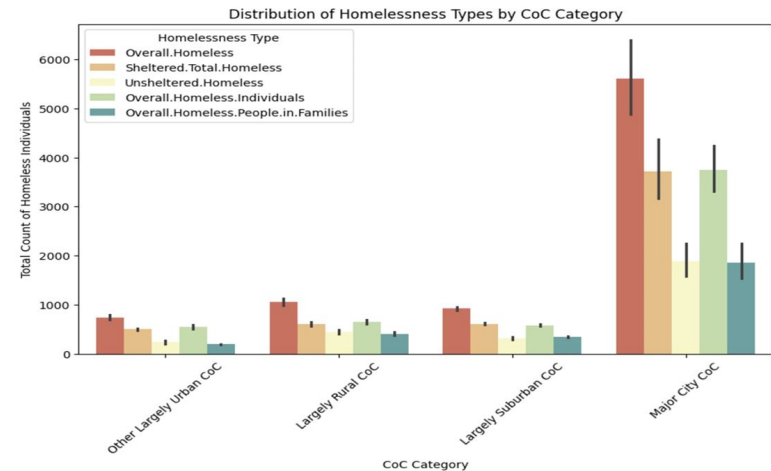
- Explore community characteristics that influence homelessness.
- Provide insights into trends and patterns over time.

### Policy Support:

- Help policymakers and organizations optimize resource distribution.
- Improve the effectiveness of homelessness interventions.

### Data-Driven Solutions:

- Enable early identification of high-risk communities.
- Support the design of long-term strategies to reduce homelessness.



# DATA DESCRIPTION



BOSTON  
UNIVERSITY

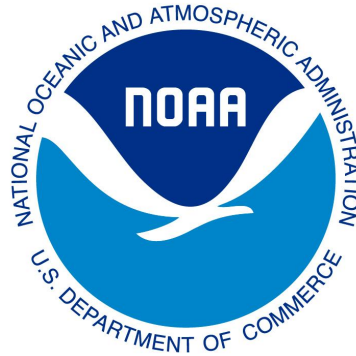
# DATA SETS

YEAR

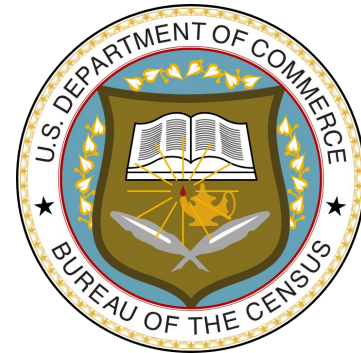
2010–2023



Housing Inventory Count  
Point-in-Time Count  
U.S. Department of Housing  
and Urban Development



Average Temperature  
National Oceanic &  
Atmospheric Administration

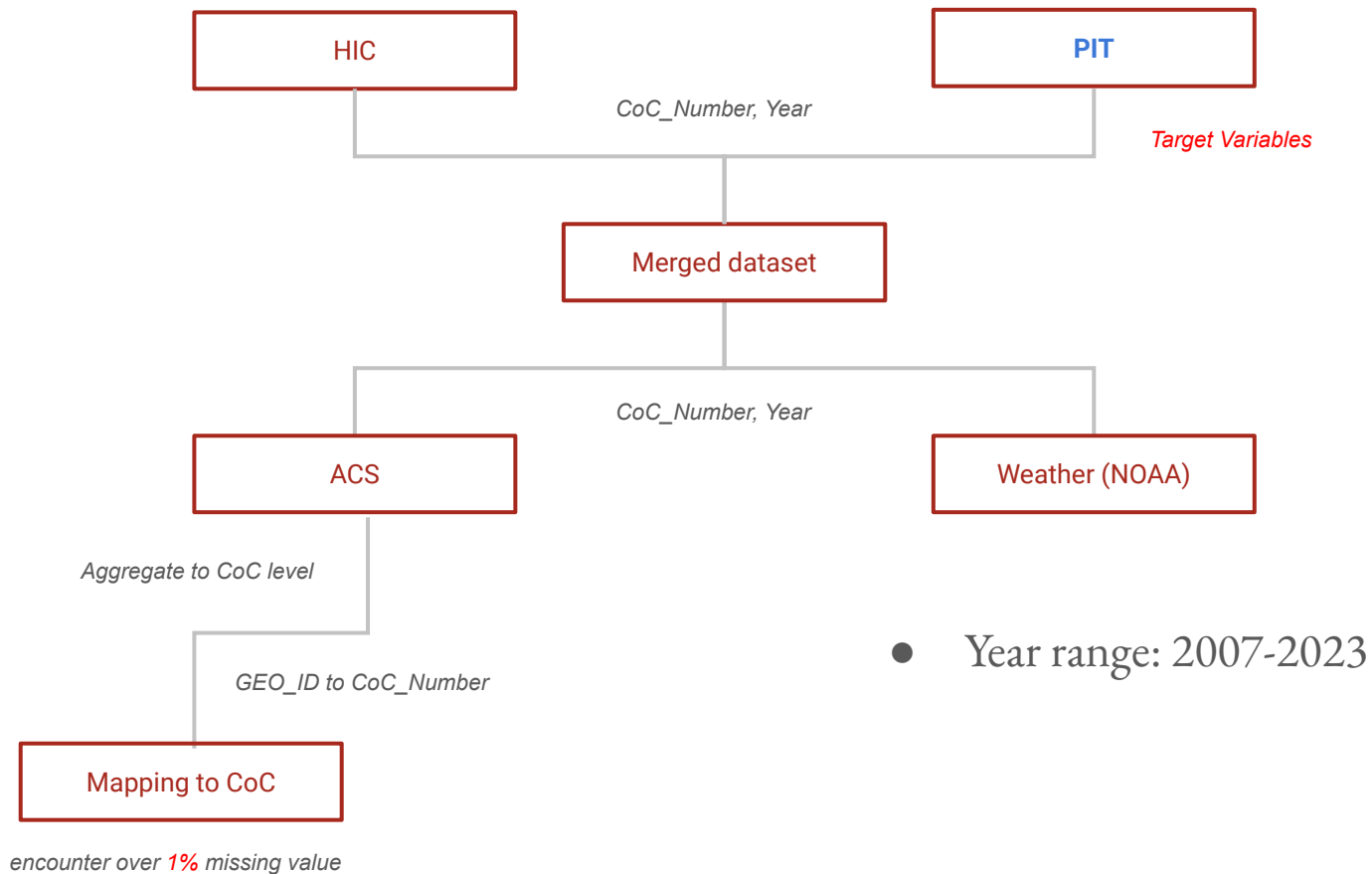


American Community Survey  
U.S. Census Bureau

**BOSTON**  
UNIVERSITY

# DATA MERGING





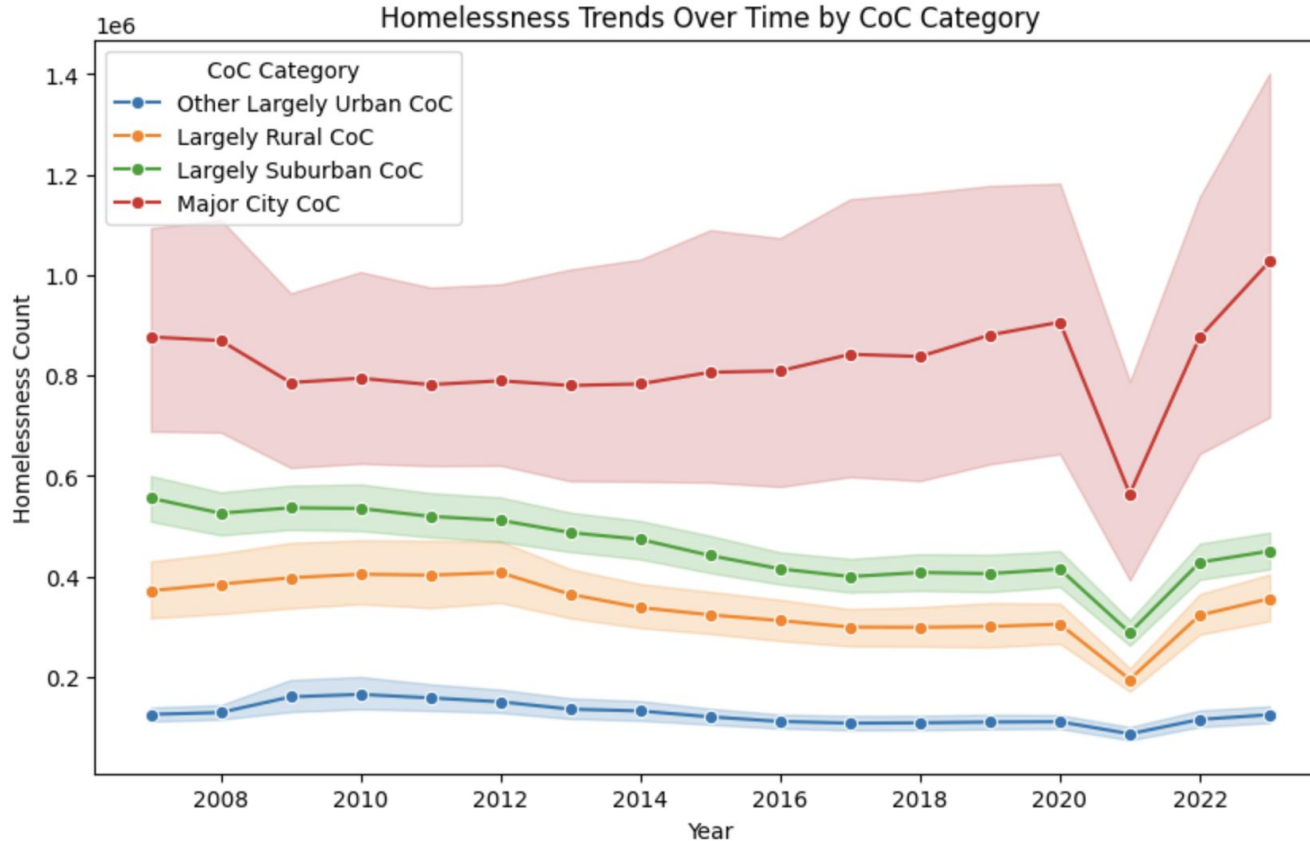




# EXPLORATORY DATA ANALYSIS

BOSTON  
UNIVERSITY

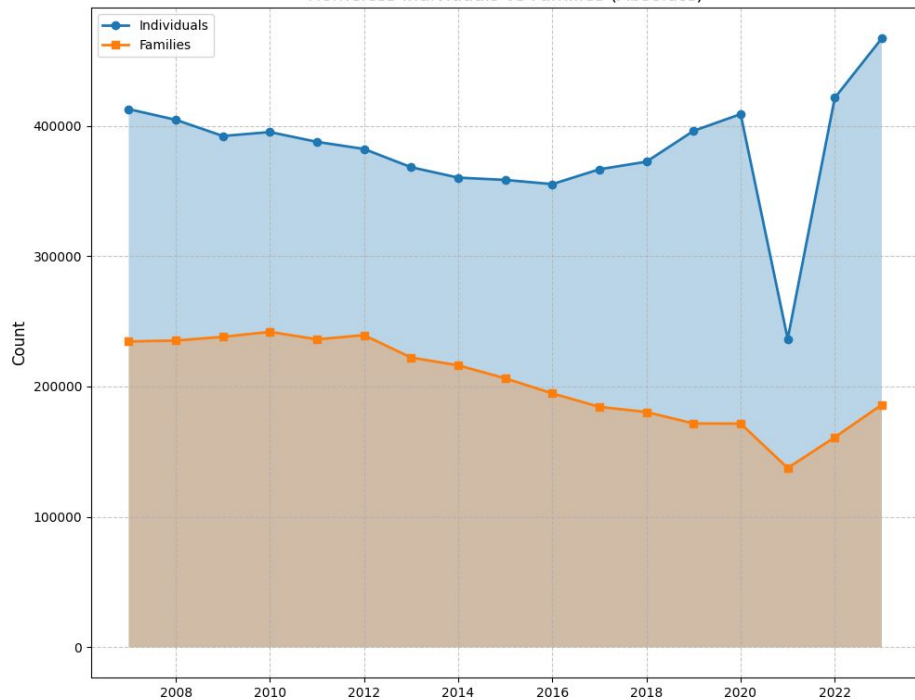
# Plummeted amid COVID - limited data availability



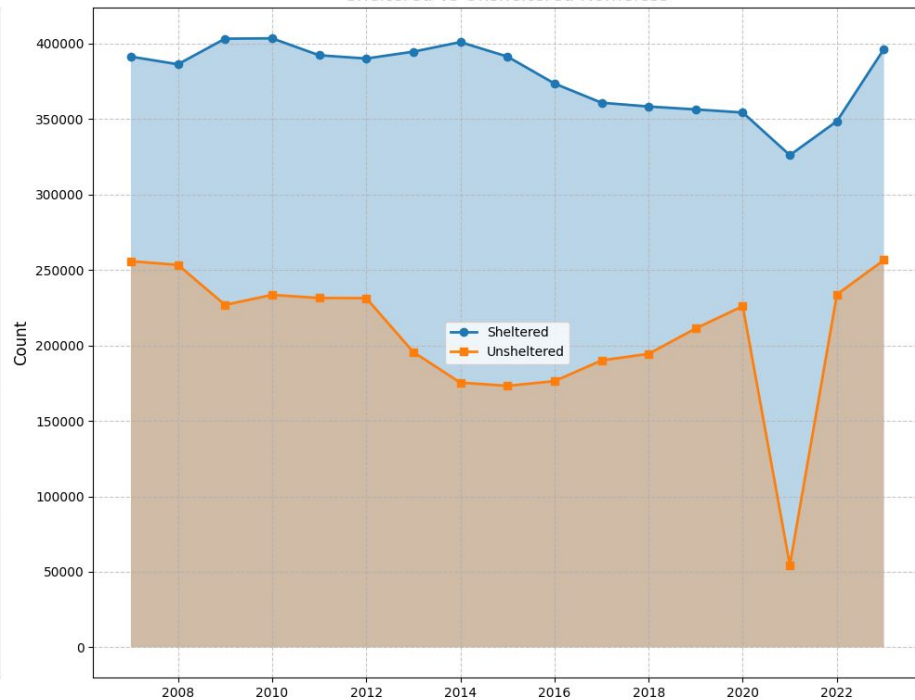
# Rate in Individuals and Sheltered dropped amid COVID

## Emergency housing initiatives , temporary shelters

Homeless Individuals vs Families (Absolute)



Sheltered vs Unsheltered Homeless





# DATA MODELING

BOSTON  
UNIVERSITY

## FEATURES

1. Total Population
2. Median Gross Rent
3. Median Household Income
4. Poverty Rate
5. Vacancy Rate
6. Unemployment Rate
7. Cost Burdened Rate
8. Renter Household\_Rate
9. Total Year-Round Beds (ES, TH, SH)
10. Average Temperature

## TARGET VARIABLES

1. Overall Homelessness Rate Per 1000 Individuals
2. Overall Homelessness Individuals Rate Per 1000 Individuals
3. Overall Homelessness People in Families Rate Per 1000 Individuals
4. Unsheltered Homelessness Rate Per 1000 Individuals
5. Sheltered Homelessness Rate Per 1000 Individuals

# RESULTS

MODELS	HOMELESSNESS TYPE	RMSE	R <sup>2</sup>
Linear Regression	Overall	0.0017	0.3453
	Individuals	0.0014	0.2753
	Families	0.0075	0.2843
	Unsheltered	0.0012	0.1981
	Sheltered	0.0010	0.4701
Random Forest	Overall	0.0009	0.7866
	Individuals	0.0007	0.7802
	Families	0.0003	0.8177
	Unsheltered	0.0007	0.7001
	Sheltered	0.0004	0.8798

Models	Homelessness Type	RMSE	R <sup>2</sup>
Histogram-Gradient Boosting	Overall	0.38	0.58
	Individuals	0.35	0.59
	Families	0.31	0.41
	Unsheltered	0.36	0.58
	Sheltered	0.33	0.52
Extreme Gradient Boosting (XGBOOST)	Overall	0.43	0.42
	Individuals	0.43	0.18
	Families	0.38	0.18
	Unsheltered	0.42	0.45
	Sheltered	0.41	0.28

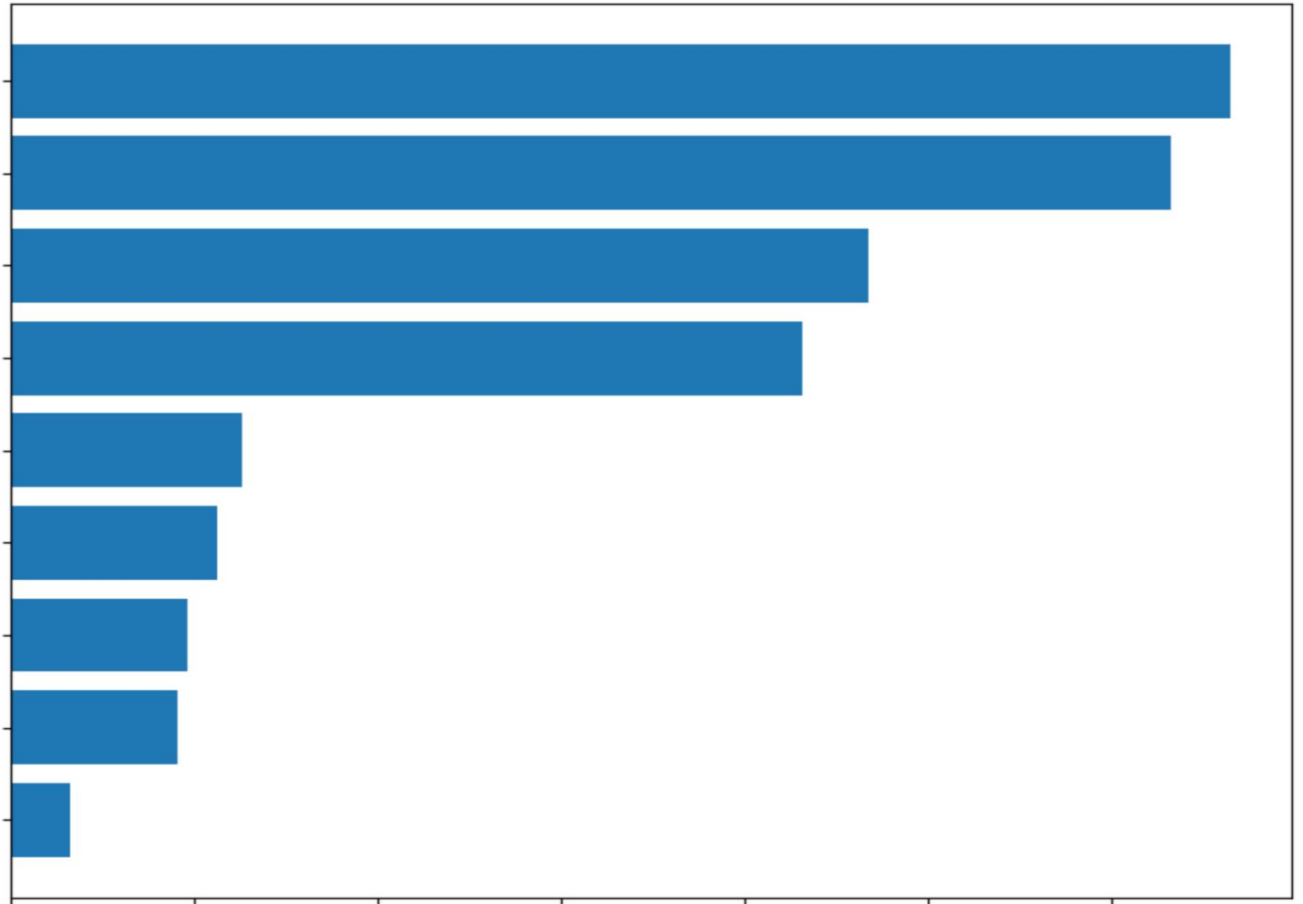
Feature Importance: Overall Homeless

Features

Renter\_Household\_Rate  
Total Year-Round Beds (ES, TH, SH)  
Total Population  
Average Temperature  
Cost\_Burdened\_Rate  
Median Household Income  
Vacancy\_Rate  
Unemployment\_Rate  
Poverty\_Rate

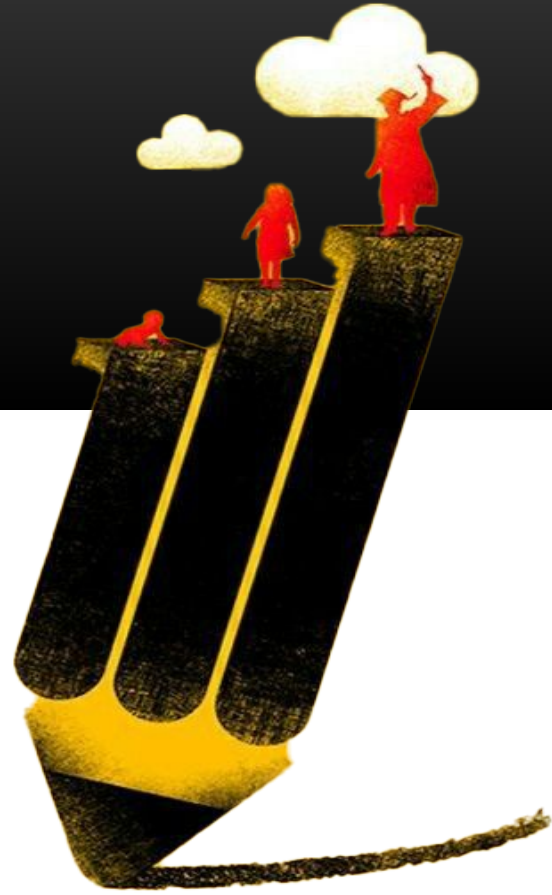
0.00 0.05 0.10 0.15 0.20 0.25 0.30

Mean Importance





# CONCLUSION



## Research Achievement:

- Created comprehensive dataset through extensive data integration
- Successfully developed several community-level homelessness prediction models
- Established strong predictive relationship between community factors and homelessness

## Key Findings:

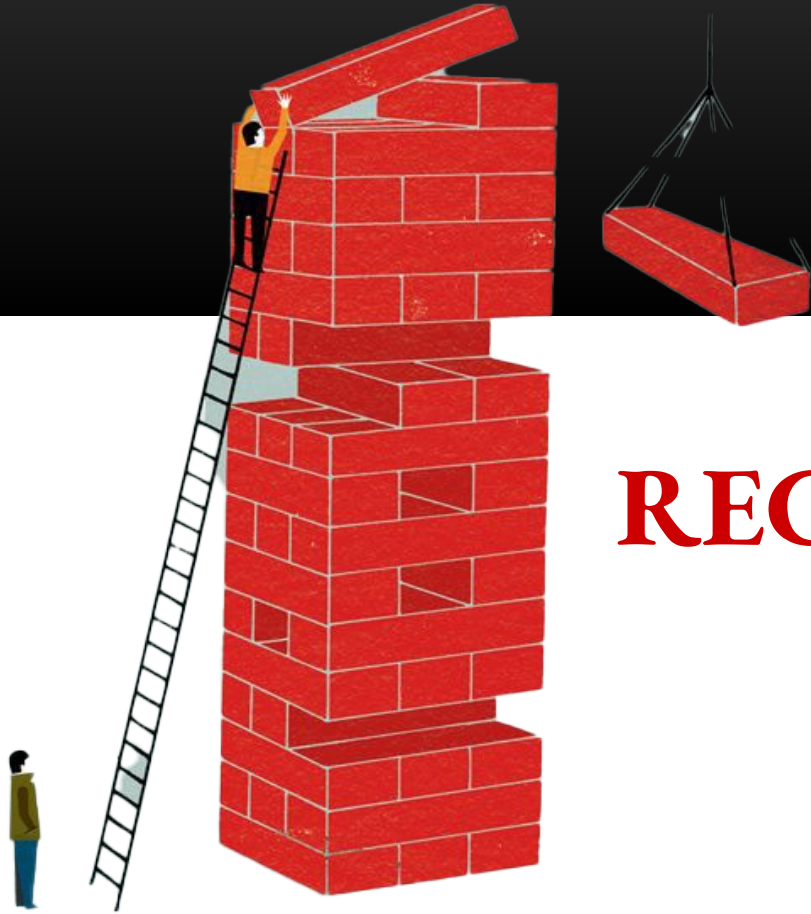
### Strong Predictive Model:

- The Histogram-Based Gradient Boosting model effectively predicted homelessness rates, demonstrated superior performance across homelessness types, with RMSE as low as 0.31 for family homelessness.

### Key Drivers Identified:

- Renter Household Rates and Total Year-Round Beds emerged as the most impactful features, followed by population and average temperature.
- Median household income, unemployment rates, and poverty rates contributed to nuanced community-level predictions.





# RECOMMENDATIONS

BOSTON  
UNIVERSITY

## Recommendations:

### Policy Recommendations:

- Target High-Risk Communities, focus interventions on communities with elevated median rent burdens and unemployment rates.
- Rent Control Policies ,implement rent stabilization measures to reduce housing cost pressures in vulnerable areas.

### Model Enhancements :

- Incorporate real-time variables: adopt dynamic factors like weather data in model to refine predictions.
- Explore model ensemble approaches: combine the strengths of The Histogram-Based Gradient Boosting with other models to achieve more robust predictions

