

# Rising Cost Of Home Ownership

## Big Data Dunces

Adnan - [asidd5@uic.edu](mailto:asidd5@uic.edu) - asidd5 (github)

David - [dboda2@uic.edu](mailto:dboda2@uic.edu) - DavidBodansky (github)

Philip - [pkeit2@uic.edu](mailto:pkeit2@uic.edu) - pkeith4 (github)

Sufyan - [ssidd5@uic.edu](mailto:ssidd5@uic.edu) - sufyan sid5 (github)

Yanjaa - [yganz2@uic.edu](mailto:yganz2@uic.edu) yanjaa-g (github)

Project link: <https://github.com/uic-ds-fall2025/class-project-big-data-dunces>



# Rising Cost Of Homes And Unaffordability

- **Big Idea:**
  - Why does homeownership feel more unattainable compared to a couple decades ago?
- **Why should you care?**
  - Gen Z and Millennials are currently affected the most by changes in housing prices as a growing number of people cannot afford to buy homes
- **Why we chose this problem:**
  - We chose this problem because of its growing impact on our generation, and so that we could thoroughly analyze it and understand how to deal with it both collectively and individually
- **Hypothesis:**
  - It is a combination of artificial market inflation driven by the private sector, lack of financial literacy, and the result of 2 economic crashes in the past 2 decades





**Data -** All of our data is already out there waiting to be processed in the form of Mass Surveys collected and organized by NGOs and the U.S Census Bureau. Most datasets we have found so far are very large in record count and size

### **Main Housing Centric Data Sets**

- American Community Survey (ACS PUMS) - Hundreds of Person and Housing Level Variables such as age, race, education, marital status, housing tenure, rent, mortgage status, taxes, utilities, household income, earnings, occupations
- American Housing Survey - More detail about housing quality
- Housing Affordability Data System - Derived Dataset with indicators related to affordability and cost-burden
- Housing Vacancy Survey - Quarterly stats on homeownership rates by age group

Datasets for financial literacy:

- <https://catalog.data.gov/dataset/money-management-and-financial-literacy>
- <https://finrafoundation.org/nfcs-data-and-downloads>
- <https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market>
- <https://www.kaggle.com/datasets/camnugent/california-housing-prices?resource=download>
- <https://www.kaggle.com/datasets/stealthtechnologies/regression-dataset-for-household-income-analysis>

# Solution

Our goal is to identify the key factors that are contributing to this feeling of unaffordability for young Americans within the U.S. Housing market. By examining housing and financial data we hope to determine a general trend in the market and find out if homes are as affordable for current generations as they have been for generations in the past.

We will compare historical trends in homeownership rates, income-to-home-price ratios, and housing cost burdens across generations, in order to assess whether current young Americans face greater challenges entering the housing market compared to prior cohorts.

Scope:

- United States and focus on young adults as defined by Census age groups.
- Homeownership, affordability, income, debt burden, and regional variation.
- Policy interventions (e.g., subsidies, tax credits) may be noted but will not be the primary focus.
- Financial Literacy Rates

Next Steps:

- Acquire raw datasets from Census ACS PUMS, HUD CHAS/HADS, and FHFA House Price Index.
- Clean and merge datasets by geography and time period.
- Define affordability metrics (e.g., cost burden >30% of income, time-to-down-payment savings).
- Conduct exploratory data analysis (EDA) and visualization to establish baseline patterns.
- Build models to compare generational affordability.



# Roles/Considerations

- Finding Data Sources: David, Philip
- Cleaning: David, Adnan, Yanjaa
- Statistical Analysis: Sufyan
- Visualization: Adnan
- Machine Learning Applications: Philip
- Deadlines:
  - Finding Data Sources/Cleaning and Proposal Video - October 12th
  - Statistical Analysis/Visualization - October 25th
  - Statistical Analysis and Progress Report - November 9th
  - Machine Learning Applications - November 23rd
  - Project Presentation - November 30th
  - Final Project Report - December 9th