Exploring Transportation Insecurity in America with Household Census Data

90803 Final Presentation
Kaia Hu & Megan Ty

Table of Contents

MORE THAN HALF

of adults in the U.S. who are experiencing poverty also face transportation insecurity

14%

more black non-hispanic adults experienced transportation insecurity compared to their white counterparts

ONE IN FOUR

of adults in the US experience transit insecurity

Introduction to Transit Insecurity

What is Transit Insecurity?

The inability to travel in a **"safe and timely manner"** because of lacking **"materials, economic, or social resources"** [1].
Transit insecurity is measured through various indicators such as **cost burden of transportation, transit availability, and commuting times** to key places like work, school, and medical appointments [3].

Who does it affect?

This predominantly affects **marginalized communities** (e.g., low-income, elderly, POC), and usually affects their neighborhoods [2].

Dataset and Features

Commuting and Commuting Costs

Neighbourhood Ratings

Housing Unit Characteristics

Housing Mortgage and Rent Costs

Reasons for leaving Previous Residence

Demographics

American Housing Survey

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

Income

Exploratory Data Analysis

Correlation Plot of Commuting/Finance Related Features
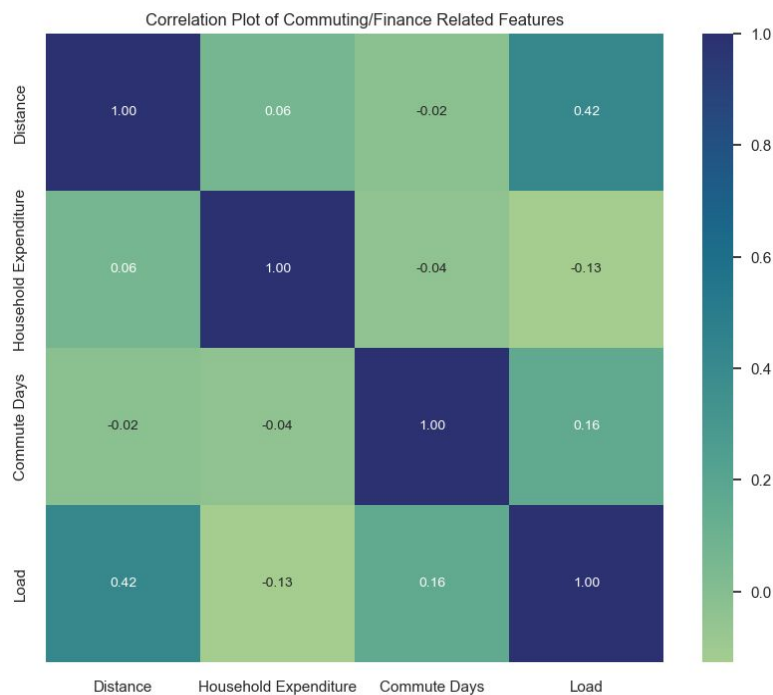
Finance/Commuting Features

Correlation Plot of Demographics
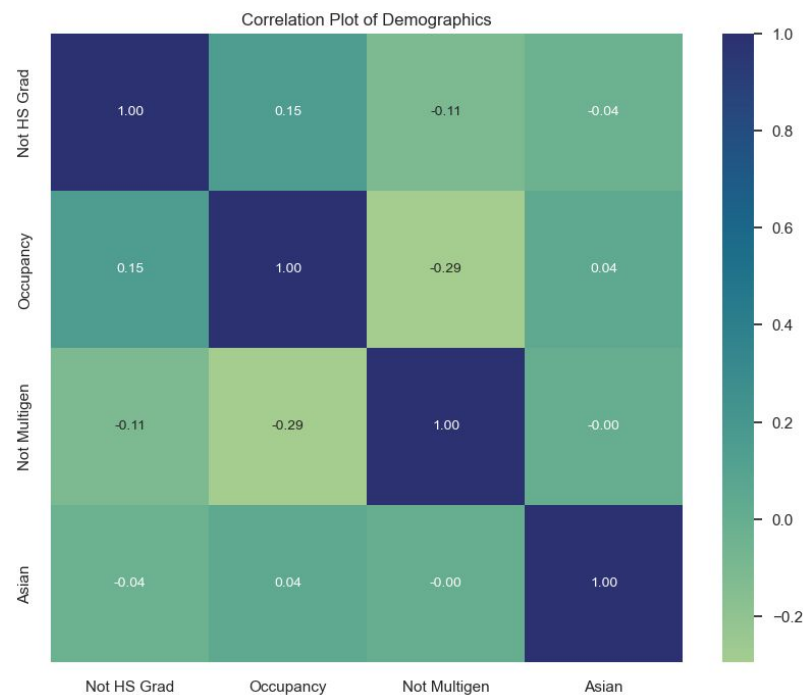
Demographics & Community Features

Ethical Considerations & Disclaimer

The authors recognize that the census data contains data on potentially sensitive information such as race, education, and disability, which was used to train the recommended models. We understand that including these demographic information into the training data may result in biased outcomes. As such, any findings and predictions from this project will be interpreted with the understanding of the potential effect the biased models may have on stakeholders.

Introduction - Key Questions

1

Financial Load

How can census data be used to build a regression model that predicts transportation cost load (% of household income that goes to transportation) and what factors within the census data significantly influence transportation costs?

2

Non-Financial Insecurity

How can census data be used to build a regression model that accurately predicts transportation insecurity and what factors within the census data significantly influence insecurity?

3

Common Groups

Can clustering analysis of census data effectively identify distinct demographic groups with unique transportation needs?
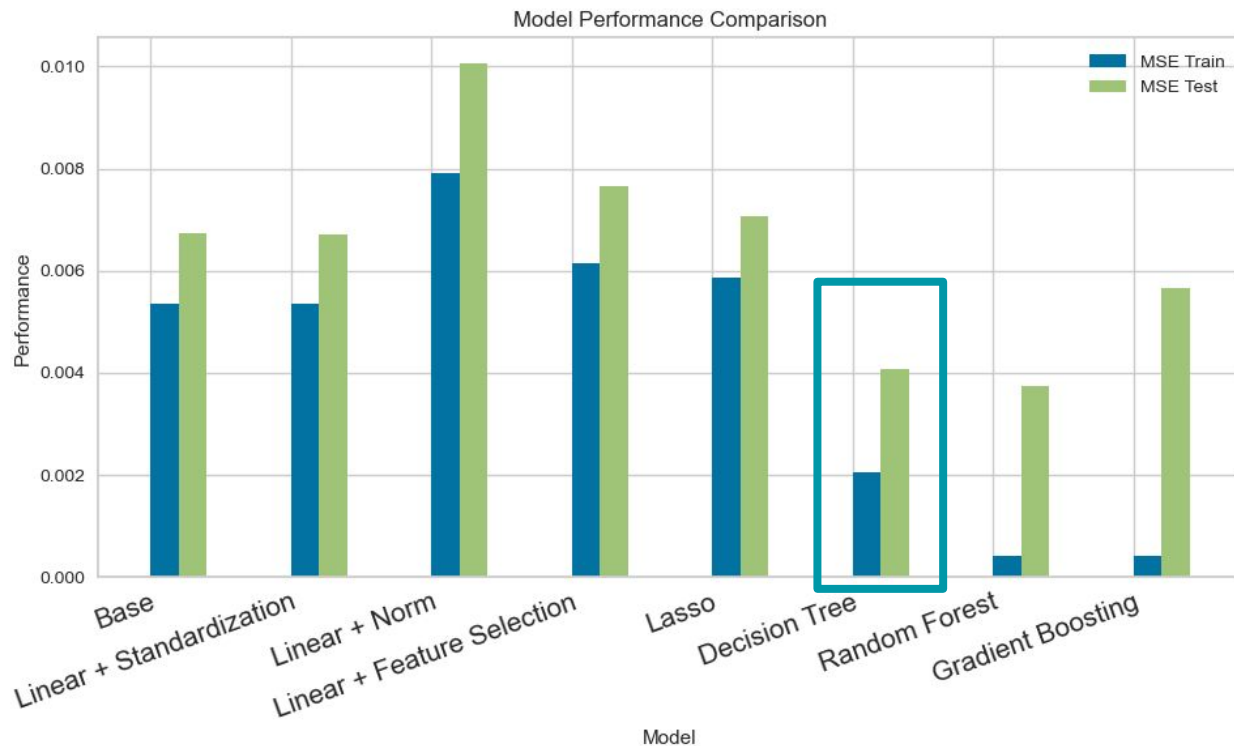
Question 1: Predict Transportation Cost Load

Can census data be leveraged to build a regression model that accurately <u>predicts transportation cost load (% of household income that goes to transportation)</u> and what factors within the census data significantly influence transportation costs, thereby aiding in the formulation of cost-effective transportation policies?

Question 1 Performance and Insights

Model Performance Comparison

Best Performing Model
Decision Tree

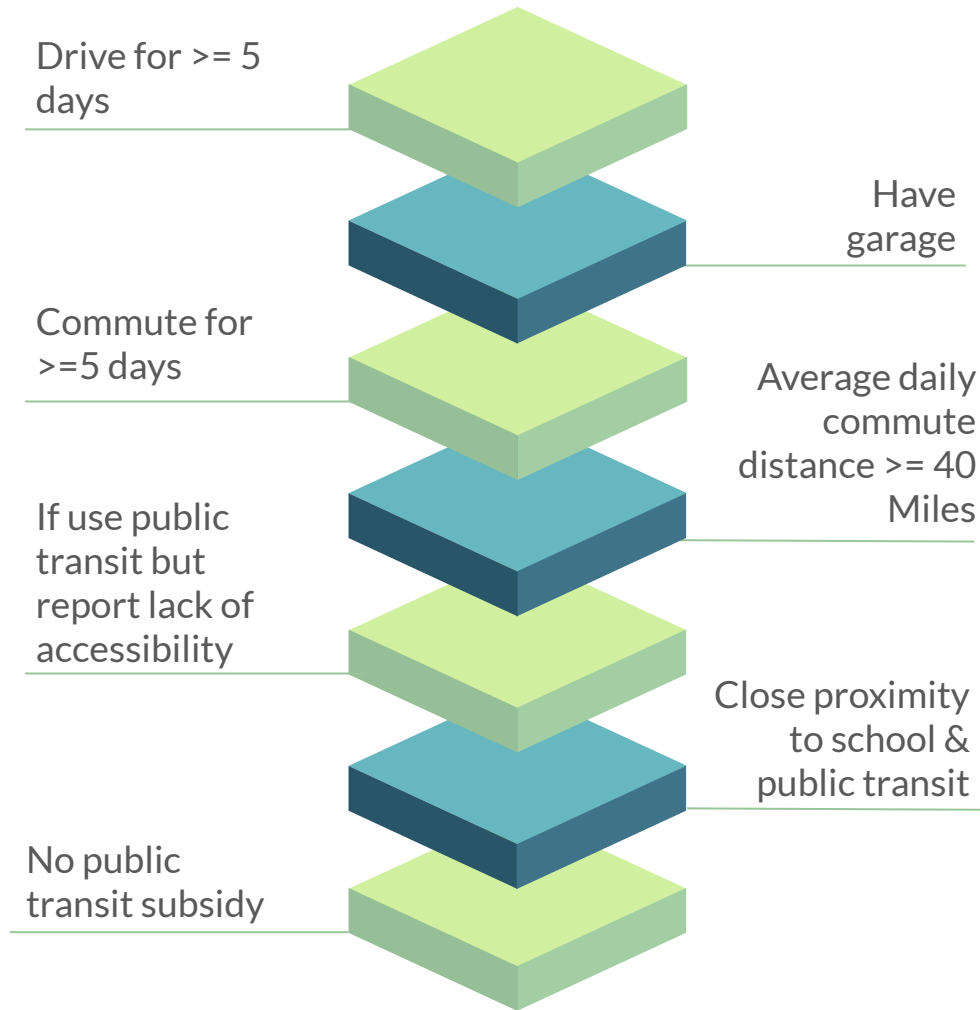**MSE Test Score**
0.00479

R2 0.788 ->0.614

Key Features
1. Commuting distance
2. % of total income that goes to household costs
3. Total monthly housing costs
4. No. of Driving Days

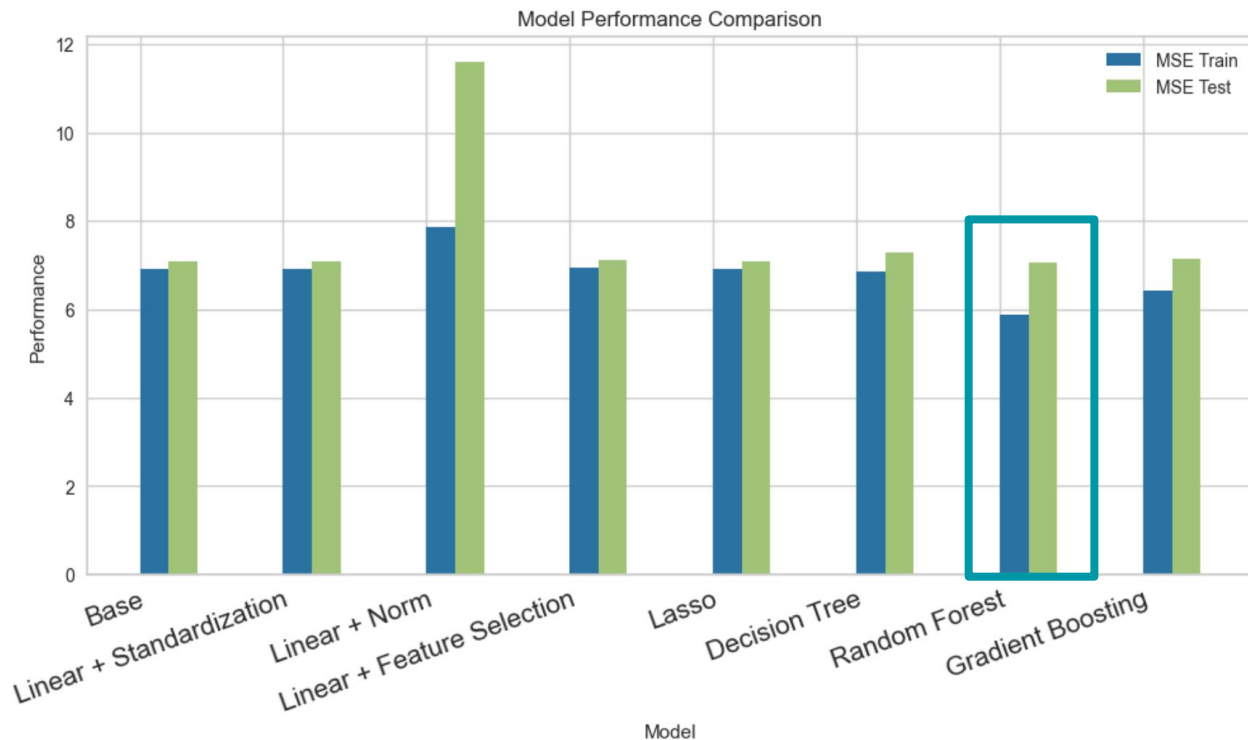# Question 2: Predict Non-Financial Transportation Stress

Can census data be leveraged to build a regression model that accurately predicts <u>non-financial transportation insecurity</u> and what factors within the census data significantly influence transportation costs, thereby aiding in the formulation of cost-effective transportation policies?

Building Our Insecurity Metric

Drive for >= 5 days

Have garage

Commute for >=5 days

Average daily commute distance >= 40 Miles

If use public transit but report lack of accessibility

Close proximity to school & public transit

No public transit subsidy

Question 2 Performance and Insights

Model Performance Comparison

Best Performing Model
Random Forest

MSE Test Score
7.063

R2 0.273 →0.105

Key Features
1. Commute Type (categorized as drive for all commuting)
2. Market evaluation of house
3. Total monthly housing cost
4. Year built

Question 3: Predict Non-Financial Transportation Stress

Can clustering analysis of census data effectively identify distinct demographic groups with different transportation accessibility and security levels?
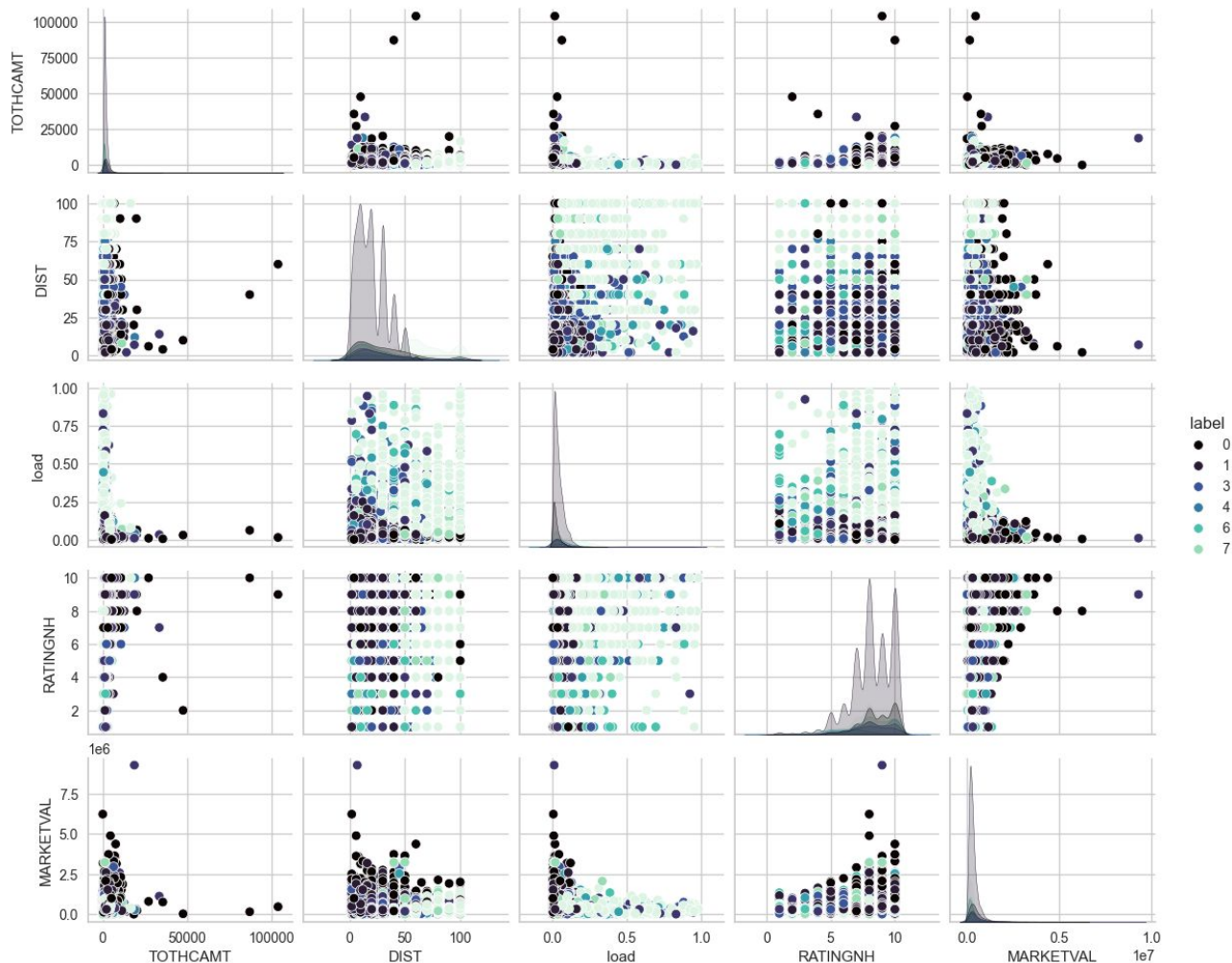
Monthly Total Housing Cost

Distance to Work

Transportation Load

Neighborhood Rating

House's Market Evaluation

Cluster 0

- Highest property values & housing cost
- Highest neighborhood rating
- Longest average commute distance yet lowest transportation burden (potentially due to low commute days ~3 days/week)

Cluster 6

- Highest average economic commute burden
- Lowest neighborhood rating

Cluster 5

- Lower education level – 851 households with people without HS degree
- Lowest property value
- The most commute days (>5) and high driving days
- 0 household close to any public transit

Cluster 7

- Highest number of disabilities in household
- Near-average transit metrics
- High car usage
- 0 household near public transit

Conclusion

1
Financial Load
Demographics are a poor predictor of transit cost load. Commuting and financial information are critical features.
Policy Implication: Prioritize collecting and analyzing data related to commuting behaviors and financial circumstances

2
Non-Financial Insecurity
Very poor model performance overall potentially due to the fact that all transit-related features were taken out to build the insecurity metric.

3
Common Groups
Very clear separation in demographics regarding transportation resource and accessibility. Overall groups with lower income and education level suffer more transportation insecurity. However, the story isn't that simple. (Ex. high disability ≠ the worst insecurity; different insecurity for urban vs rural vs suburban)
Policy Implication: Policy targeting specific demographics; transit subsidy is proven effective; public transit needs to be more accessible for disabled people.

Future Steps

Temporal Analysis
- Obtain data over longer period to identify long-term trends/changes
- Identify "warning" features, which signal impending changes when they fluctuate.
- Shed insights on effects of electric vehicle adoption and shifts towards automobile-centric or pedestrian-centric cultures.

Geospatial Analysis
- Spatial patterns and hotspots are key in transit insecurity, evident from the fact that commuting distance is important feature
- Adding location will aide in predictive modelling, and targeting spatial interventions

Controlling for Bias & Equity
- Examine model outcomes across different demographic groups
- Reselect features in a manner that minimizes biases or potential feature reconstruction

Thank you!