FOOD AFFORDABILITY ANALYSIS

Madison Wozniak | December 11th, 2024

CONTENT

01 INTRODUCTION

CONTEXT

Food expenses are disproportionately high, placing an especially heavy burden on single mother households.

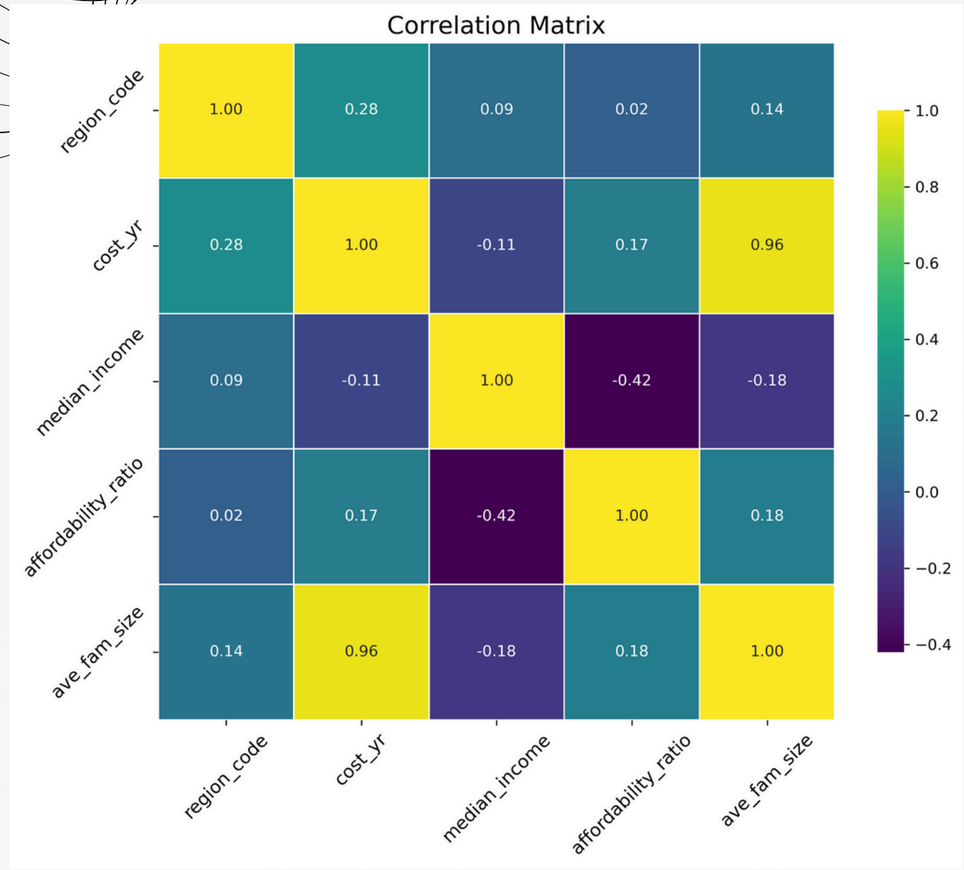
PROBLEM STATEMENT

Food affordability is a critical challenge for single mothers in California, with the gap between income and prices increasing every year.

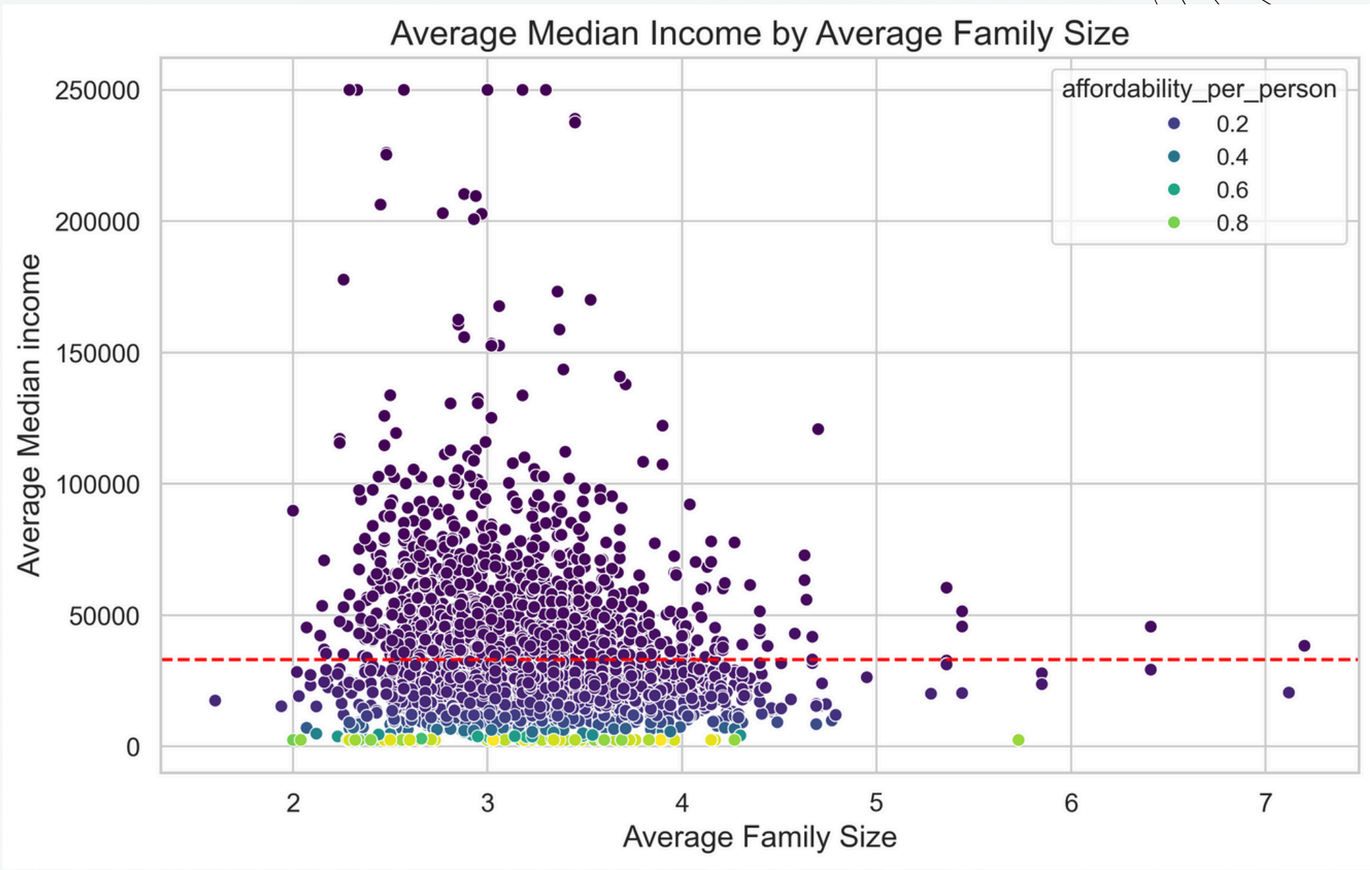
RESEARCH QUESTION

Can socioeconomic and demographic factors accurately explain and predict food affordability for single mothers in California, and how can these insights inform policy to improve economic outcomes?

02 EXPLORATORY DATA ANALYSIS

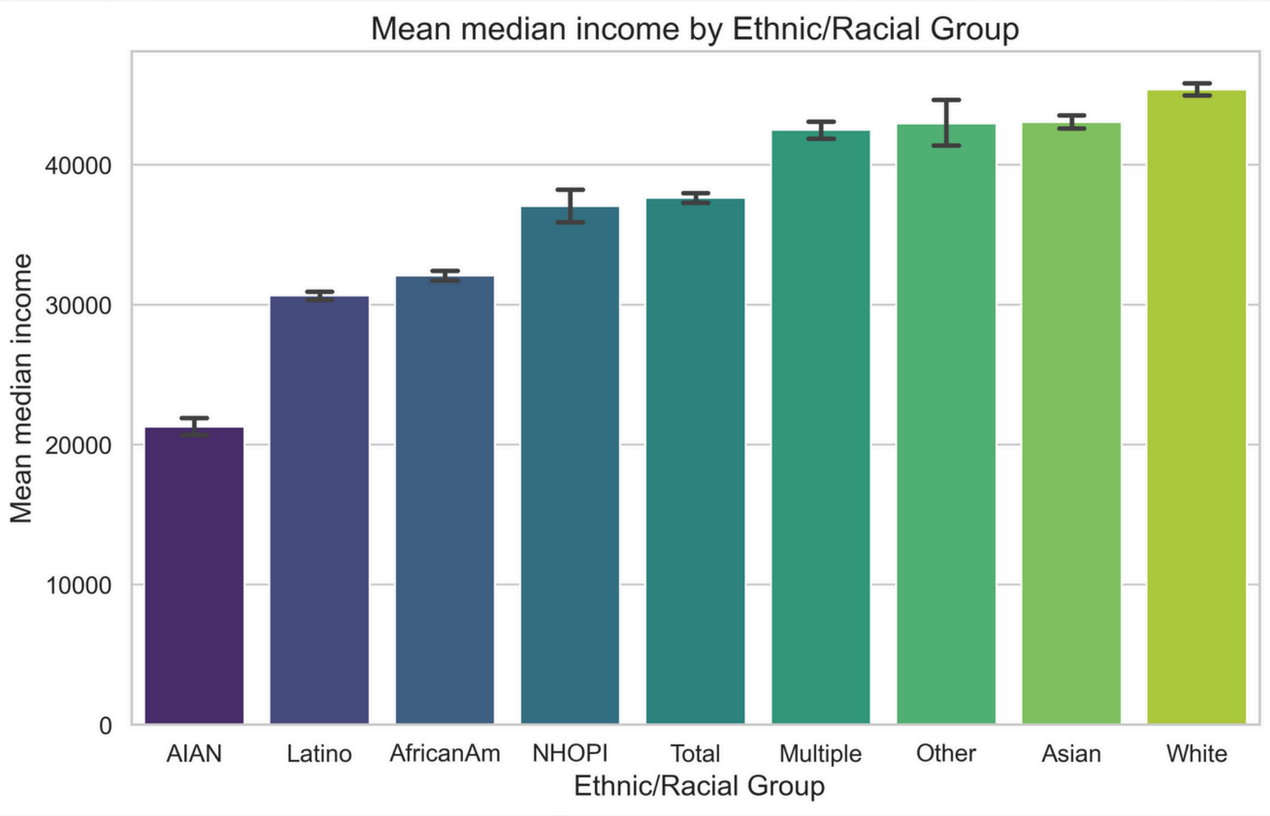


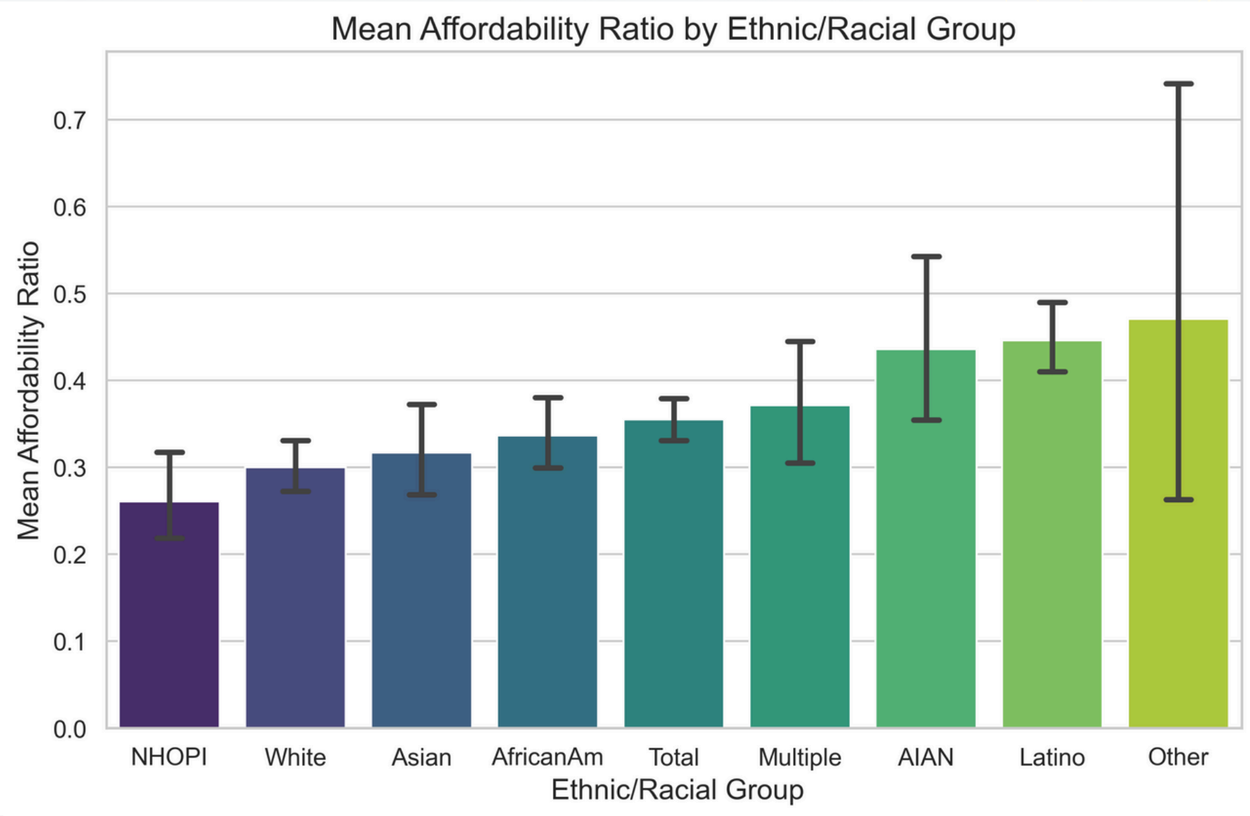
Strong positive correlation between family size and food costs



WIDE RANGE OF INCOMES MEDIAN INCOME

02 EXPLORATORY DATA ANALYSIS





There is an unexpected relationship between income and affordability when accounting for ethnic group

03 METHODS: FEATURE ENGINEERING

LATITUDE & LONGITUDE

Map region names in original data to latitude and longitude coordinates

FOOD COST INDEX (FDI)

Measurement from the Consumer Price Index to assess food costs against the national average

PREPROCESSING

Handle missing values, scale numeric features, and one-hot encode categorical features

03 METHODS: MODELS USED

KNN

Rationale:

Try to find meaningful groupings in the data

Challenges:

high computational cost

RIDGE REGRESSION

Rationale:

Assess linearity with a penalty term

Challenges:

Scored poorly when compared to tree-based models

DECISION TREES

Rationale:

Find important patterns between influential features

Challenges:

Prone to overfitting

XGBOOST

Rationale:

Implement an ensemble method to attempt to outperform the decision trees

Challenges:

most effective model

DNN

Rationale:

To explore intricate patterns that could not be captured by simpler models

Challenges:

Difficult to balance over/underfitting

04

RESULTS: XGBOOST

FITTING THE DATA

Handled overfitting better than Decision Trees and captured patterns better than linear regression

PERFORMANCE METRIC

rMSE Score: 0.0153

TIME

27 seconds

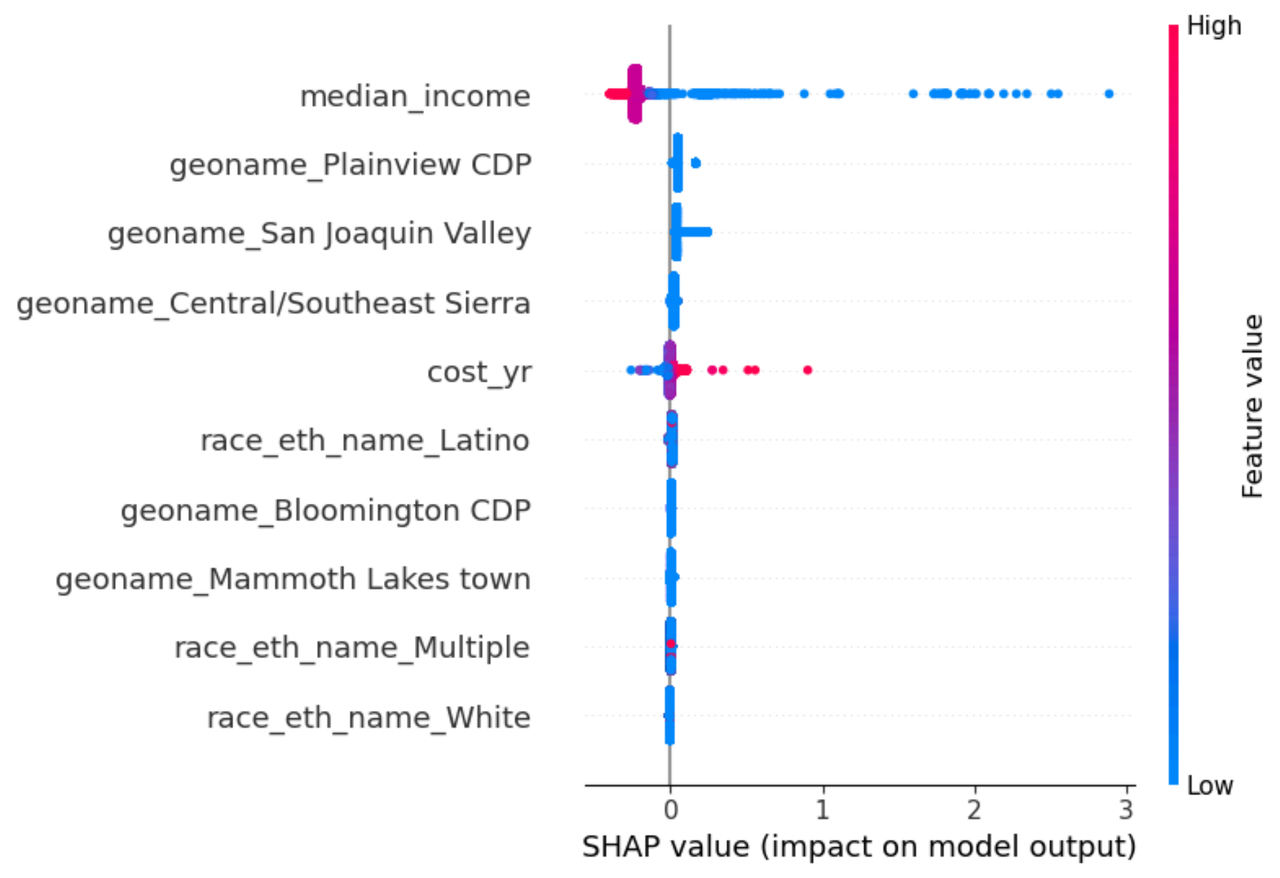
04 RESULTS: FEATURE IMPORTANCE

SHAP

Global Insights:

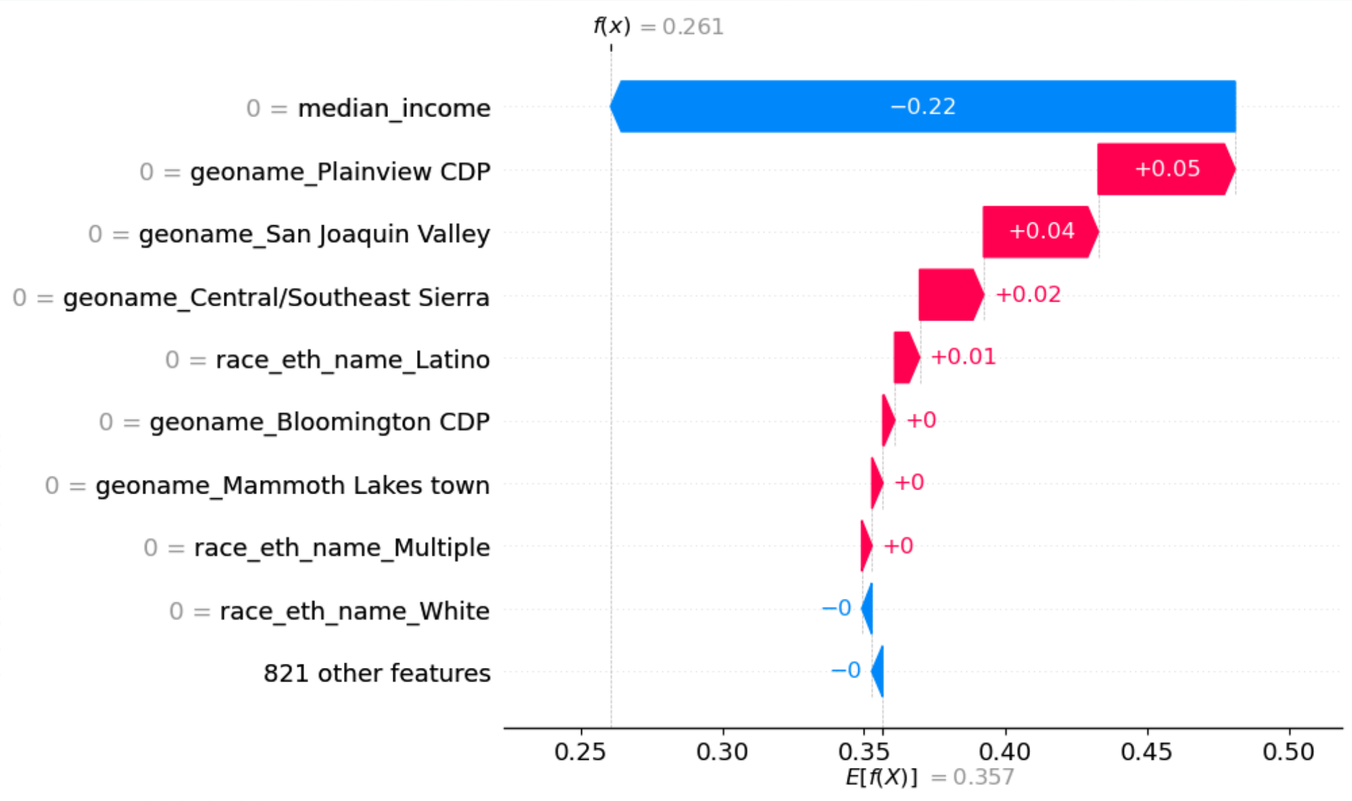
Median income reduces affordability

Geographic features proved to be highly influential



04 RESULTS: FEATURE IMPORTANCE

SHAP



Individual Interpretation:

XGBoost's predicted affordability ratio for this individual is 0.261, which is lower than the model's baseline prediction of 0.357 due to the negative and positive impacts of features such as median income and geographic location.

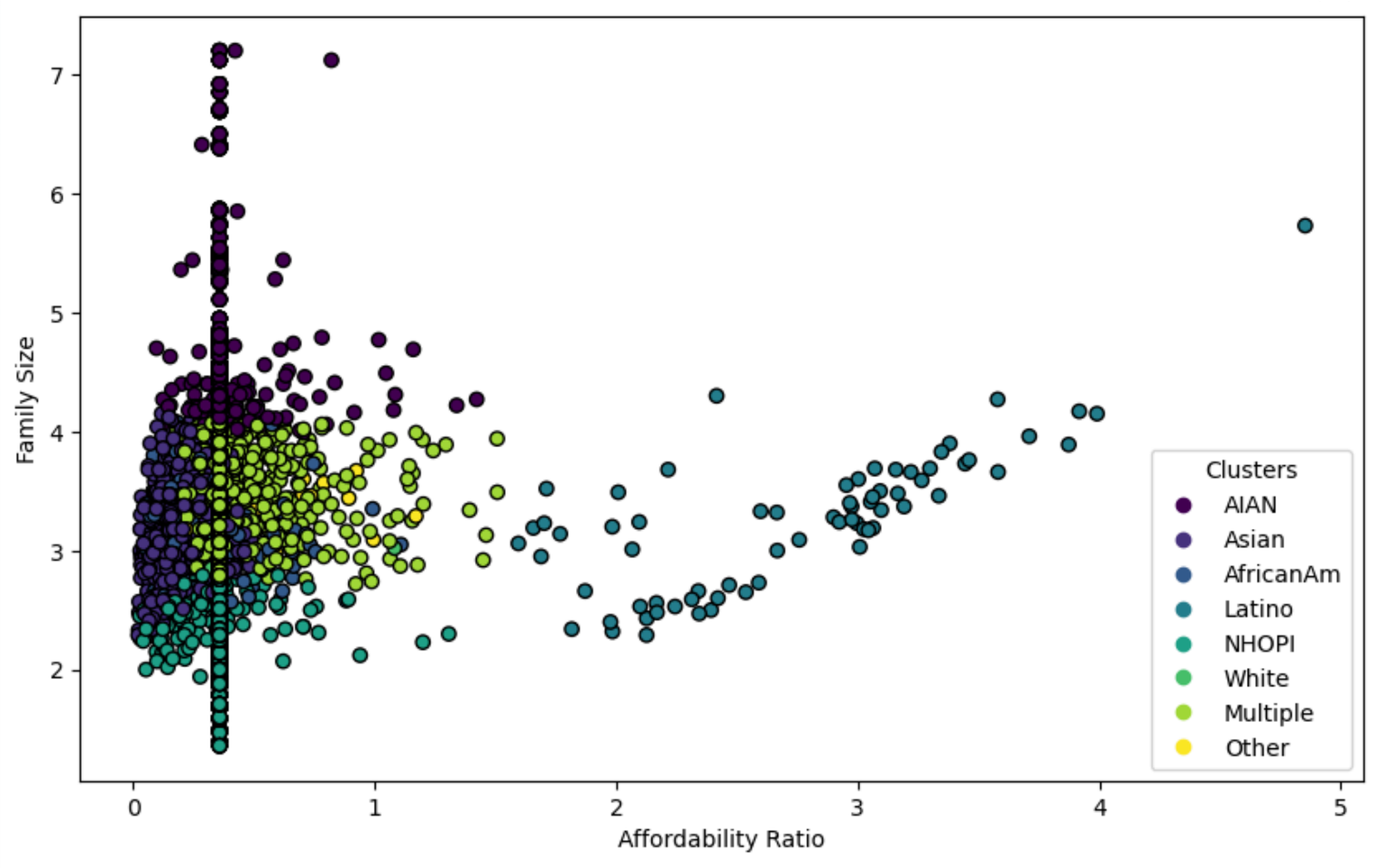
04

RESULTS: ANOMALY DETECTION AND CLUSTERING

There are distinguishable separations between ethnic groups.

Affordability tends to vary the most for Latino single mothers.

Affordability remains constant for AIAN, NHOPI and Multiple ethnic groups with various family sizes.



AIAN: American Indian/Alaskan Native
NHOPI: Native Hawaiian or Pacific Islander

04 CONCLUSION

KEY TAKEAWAYS

- XGBoost effectively modeled affordability with income and location being dominant factors
- SHAP analysis allows for interpretability on the individual and global scale
- Clustering allows for targeted intervention

NEXT STEPS

- Knowing the importance of location, stronger exploration of coordinate data could be insightful
- Test additional ensemble methods and refine anomaly detection

05 RECOMMENDATIONS

- Expand subsidies and food assistance in high-need areas.
- Combine these outcomes with additional data to uncover additional systemic location-based affordability barriers.
- Compare results to affordability of dual-income families or male-headed households to understand potential gender-related income disparities.