AlzAware: Early Alzheimer’s Detection

by : GROUP 12 PHASE 5

INTRODUCTION

The AlzAware Project leverages predictive modeling and social determinants of health to detect early signs of Alzheimer's disease and related Dementias.

PROBLEM STATEMENT

Addressing Cognitive Decline: Predictive Modeling of Alzheimer’s Disease Through Social Determinants of Health.

MAIN OBJECTIVE

- **Develop a Predictive Model for Early Alzheimer's Detection Using Social Determinants.**

SPECIFIC OBJECTIVES

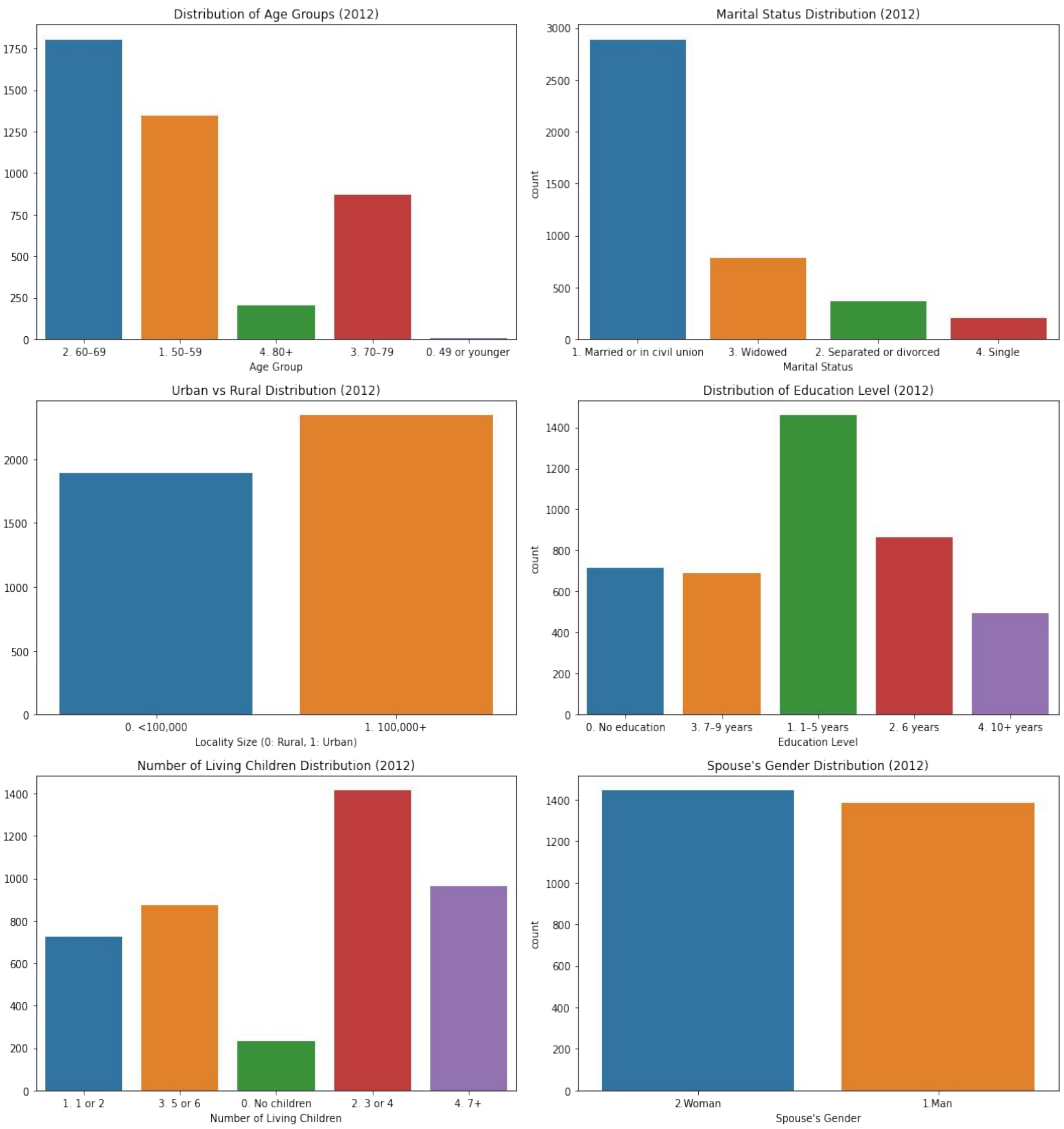
- Improved Detection: Predict AD/ADRD risk using non-clinical factors.
- Bias Mitigation: Ensure accuracy across diverse groups.
- Enhanced Accessibility: Utilize widely available social health data.
- Generalization Potential: Adaptable framework for global applications.

DATA UNDERSTANDING

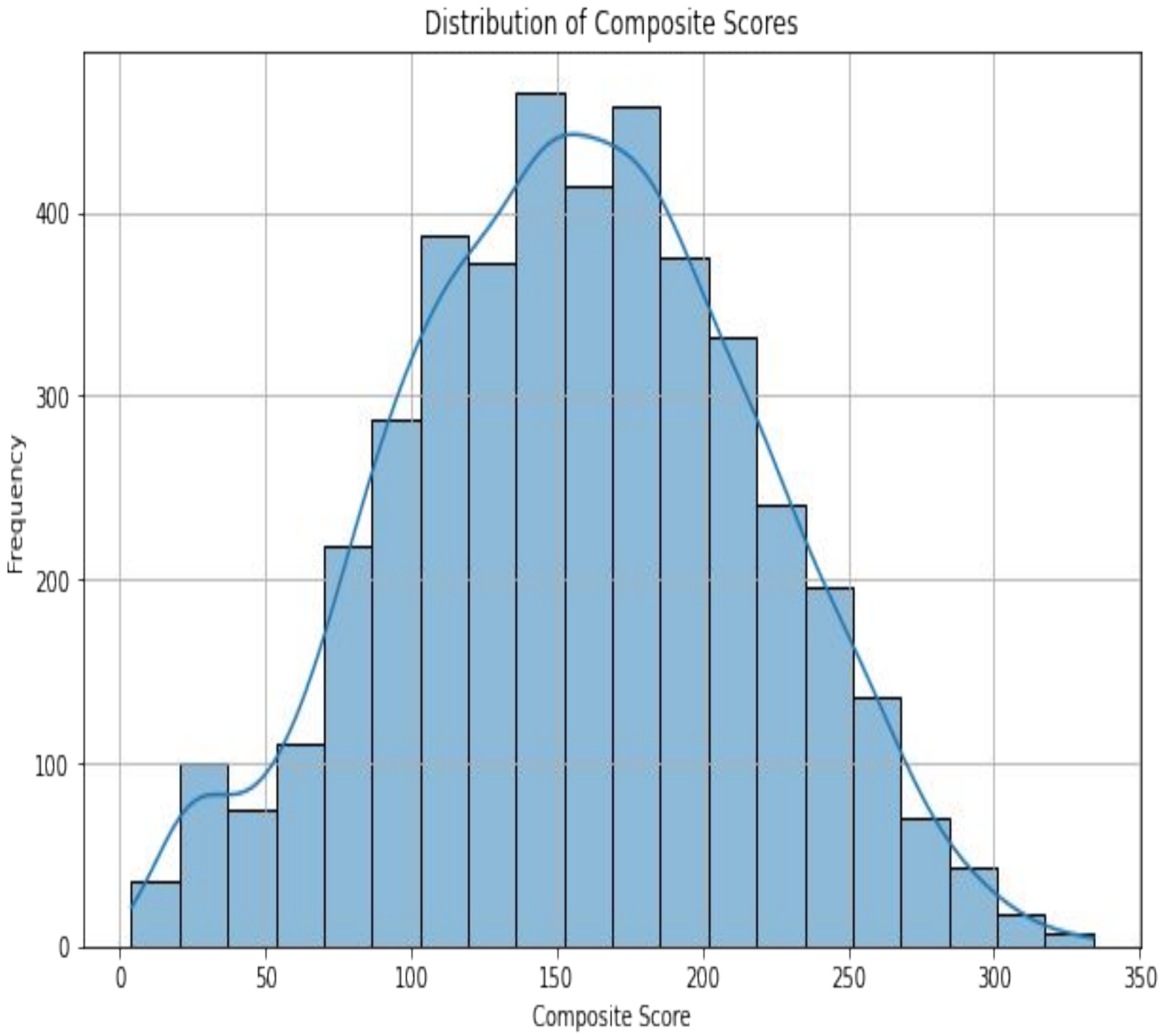
- Source: Mexican Health and Aging Study (MHAS).

- Years: Data from 2003, 2012 (training), and 2016, 2021 (evaluation).

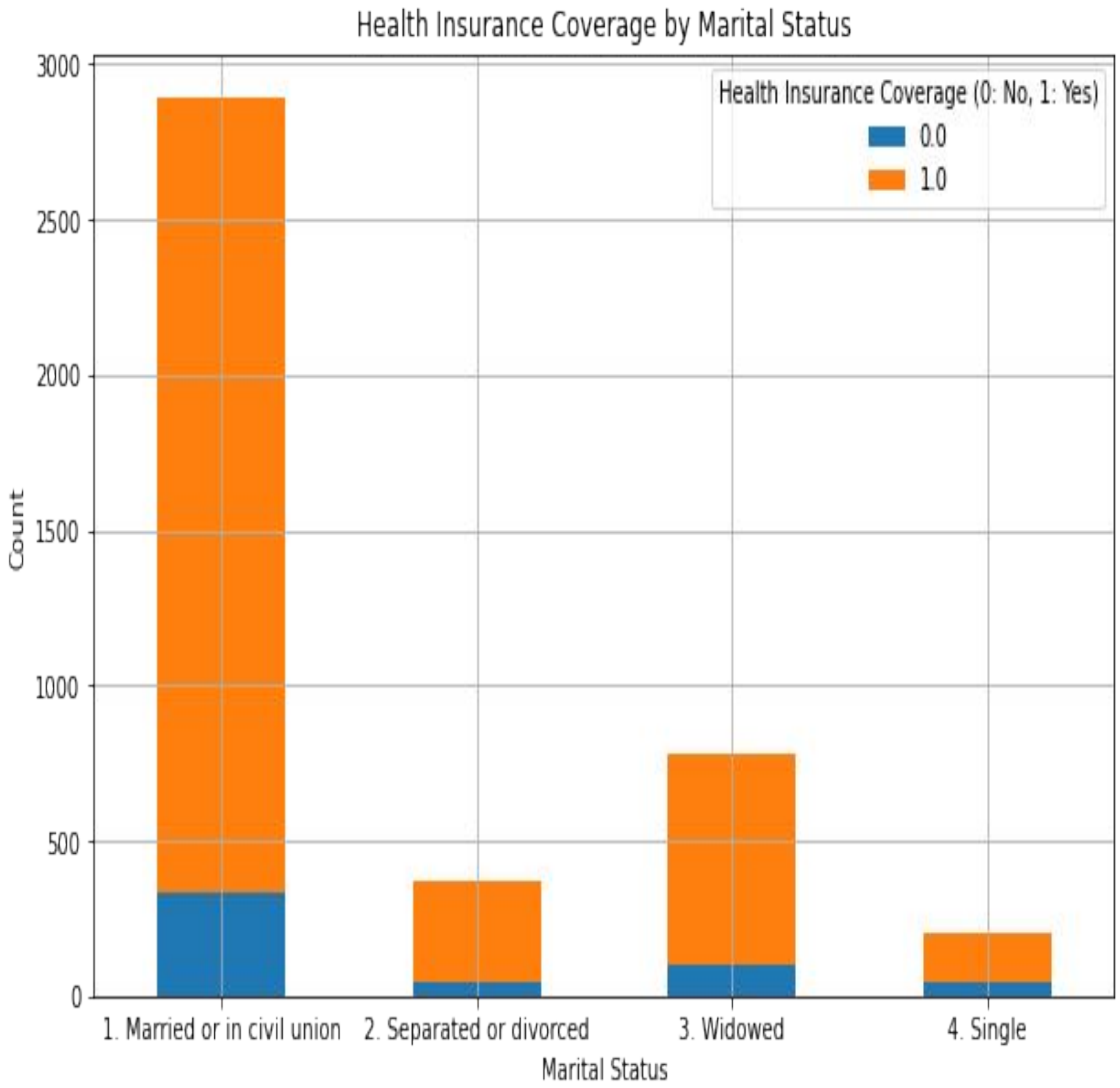
- Key Data: Demographics, socioeconomic factors, health metrics, lifestyle behaviors, and cognitive scores.



- Majority aged 60–69, mostly married/civil union.
- Urban residents slightly outnumber rural.
- Common education: 1–5 years; most have 3–4 children.
- Spouse gender nearly equal, slightly more women.

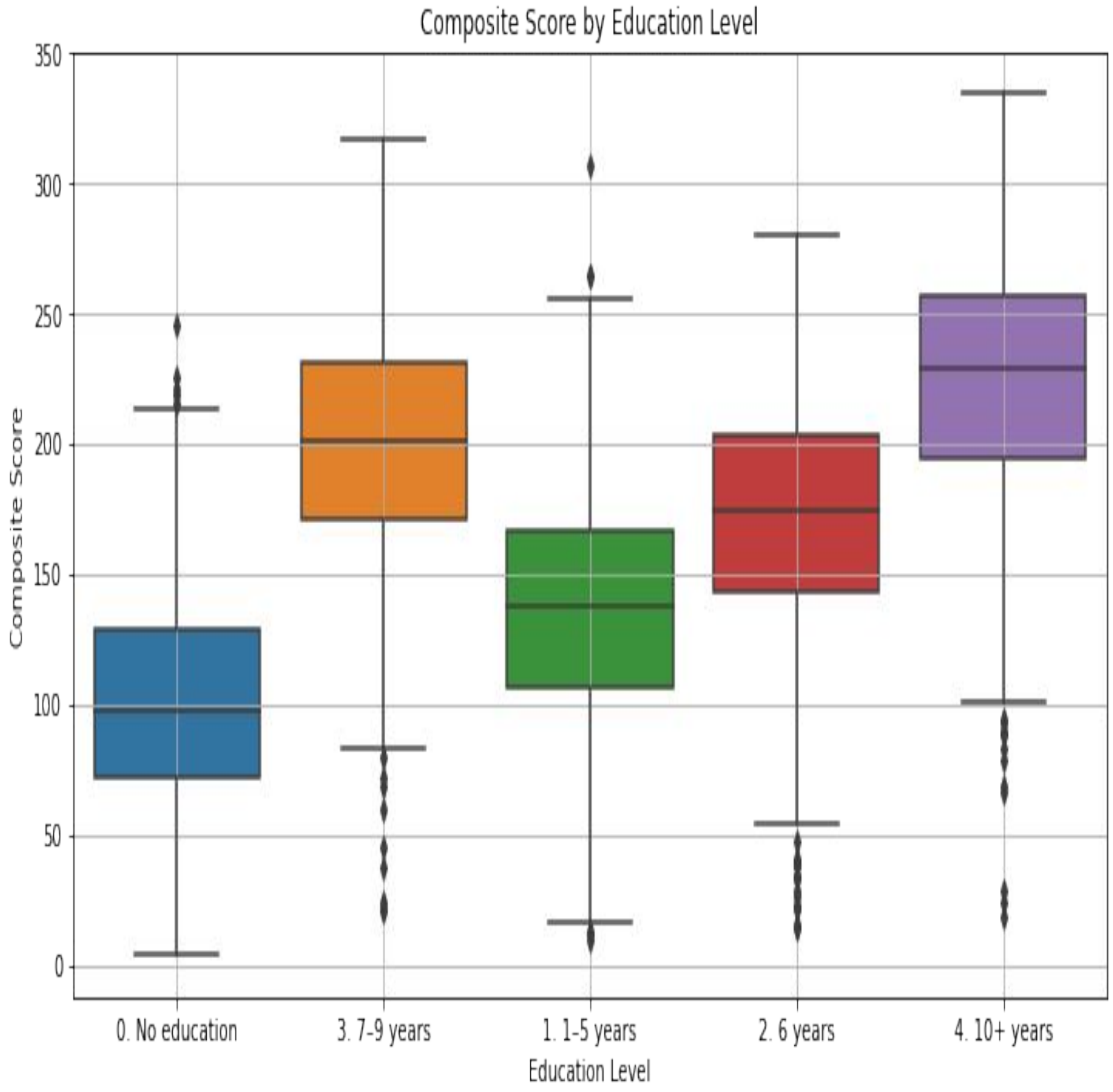


- The graph shows the distribution of the composite score, which aggregates various health and lifestyle domains.
- Analyzing this score can reveal patterns or trends in overall health across the population in your dataset.



Health Insurance by Marital Status:

- Married or Civil Union: Majority have health insurance; few are uninsured.
- Widowed: Most have coverage, but uninsured rates are higher than married individuals.
- Separated/Divorced & Single: Lower overall counts, with insurance coverage less prevalent compared to married individuals.



Education and Performance:

- Higher education leads to better scores.
- Low scores and high variability in no education group.
- Significant improvement with 6+ years of education.
- Outliers in higher education reflect other influencing factors like socio-economics.

Pipeline Integration

Key steps we took:

- Defining preprocessing pipeline.
- Selecting models (e.g., Gradient Boosting, Random Forest).
- Training models using integrated pipelines.
- Evaluating performance using metrics like RMSE and R^2.

Model Evaluation

- Gradient Boosting Regressor: Lowest RMSE: 40.6148 and Highest R^2: 0.5652
- Random Forest Regressor: Slightly higher RMSE: 40.9076 and R^2: 0.5589
- XGBoost Regressor: Competitive RMSE: 40.1440 and R^2: 0.5752

Strategies of Improving RMSE

- Feature Engineering
- Hyperparameter Tuning:
- Gradient Boosting and Random Forest optimizations.
- Ensemble Methods:
- Combine predictions using stacking or blending.
- Dimensionality Reduction:
- PCA or RFE for noise reduction.

SHAP Analysis

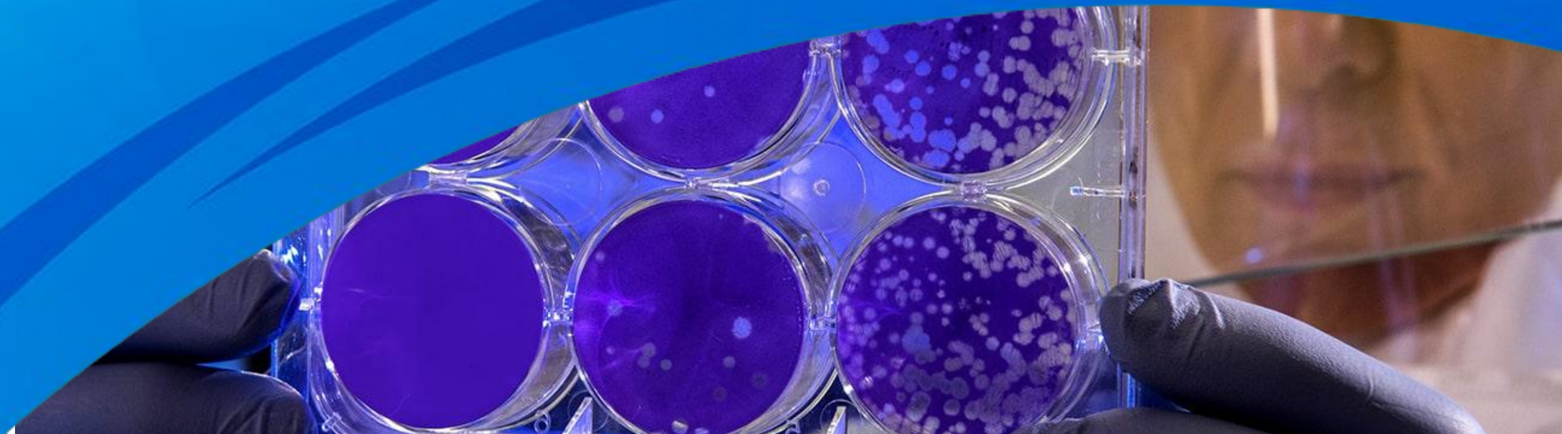
- Preventive Care Index.
- ADL/IADL Progression.
- Parental Education.
- Household Income.
- Global Health Ratings.
- Social Engagement.
- Education Levels.

Result: Final Model

- Best Model: Stacked Model

CONCLUSION:

- Our objective of building a predictive model can be achieved using the stacked model



Thank You!

Any Questions?