



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.**

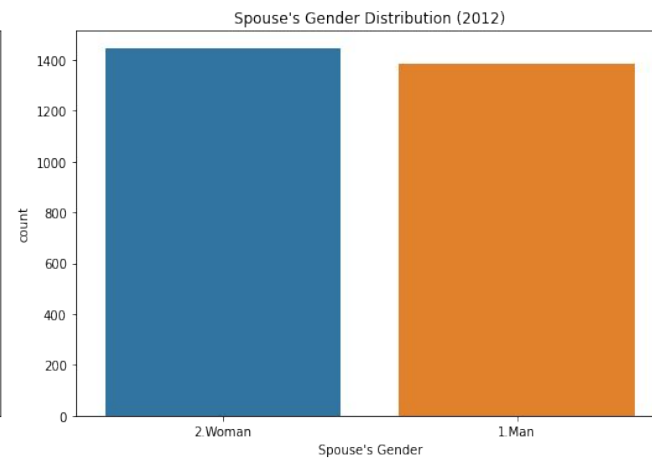
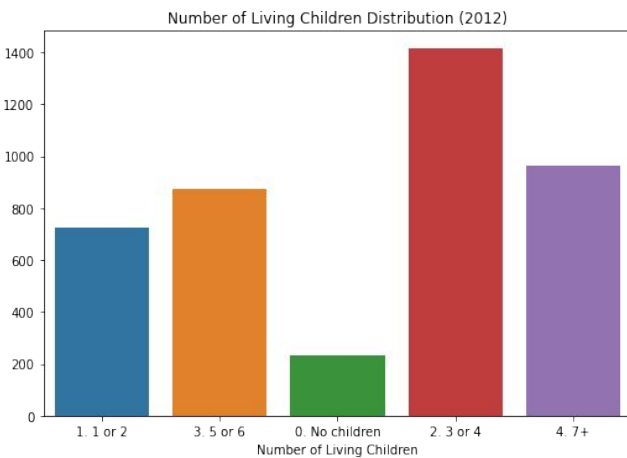
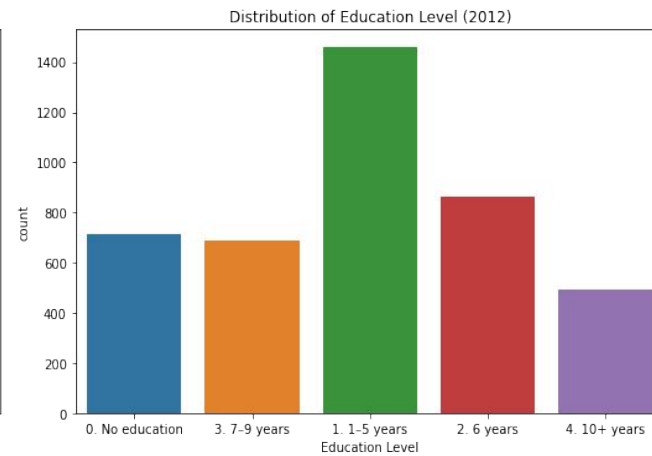
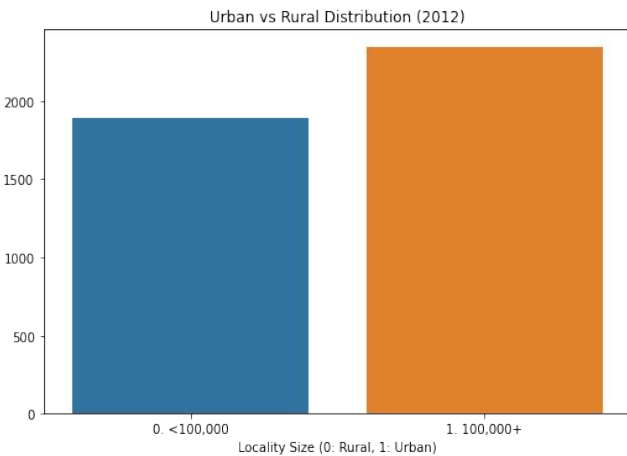
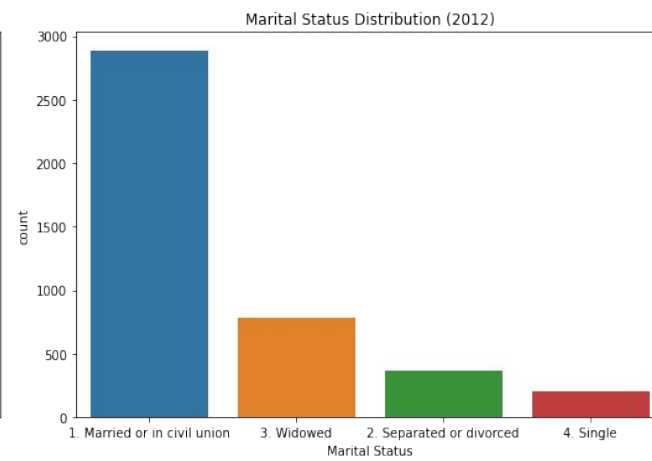
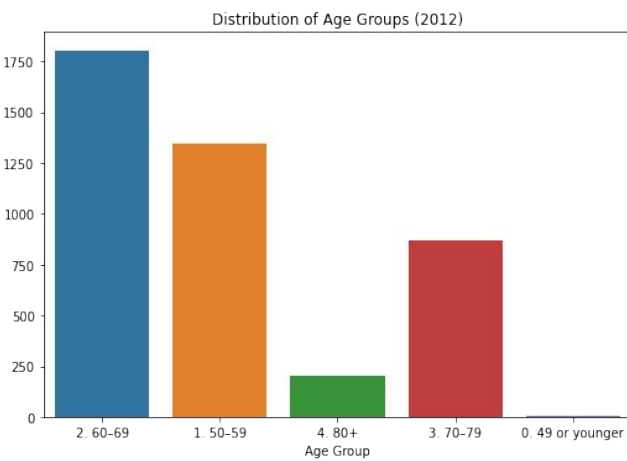


SPECIFIC OBJECTIVES

- 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.

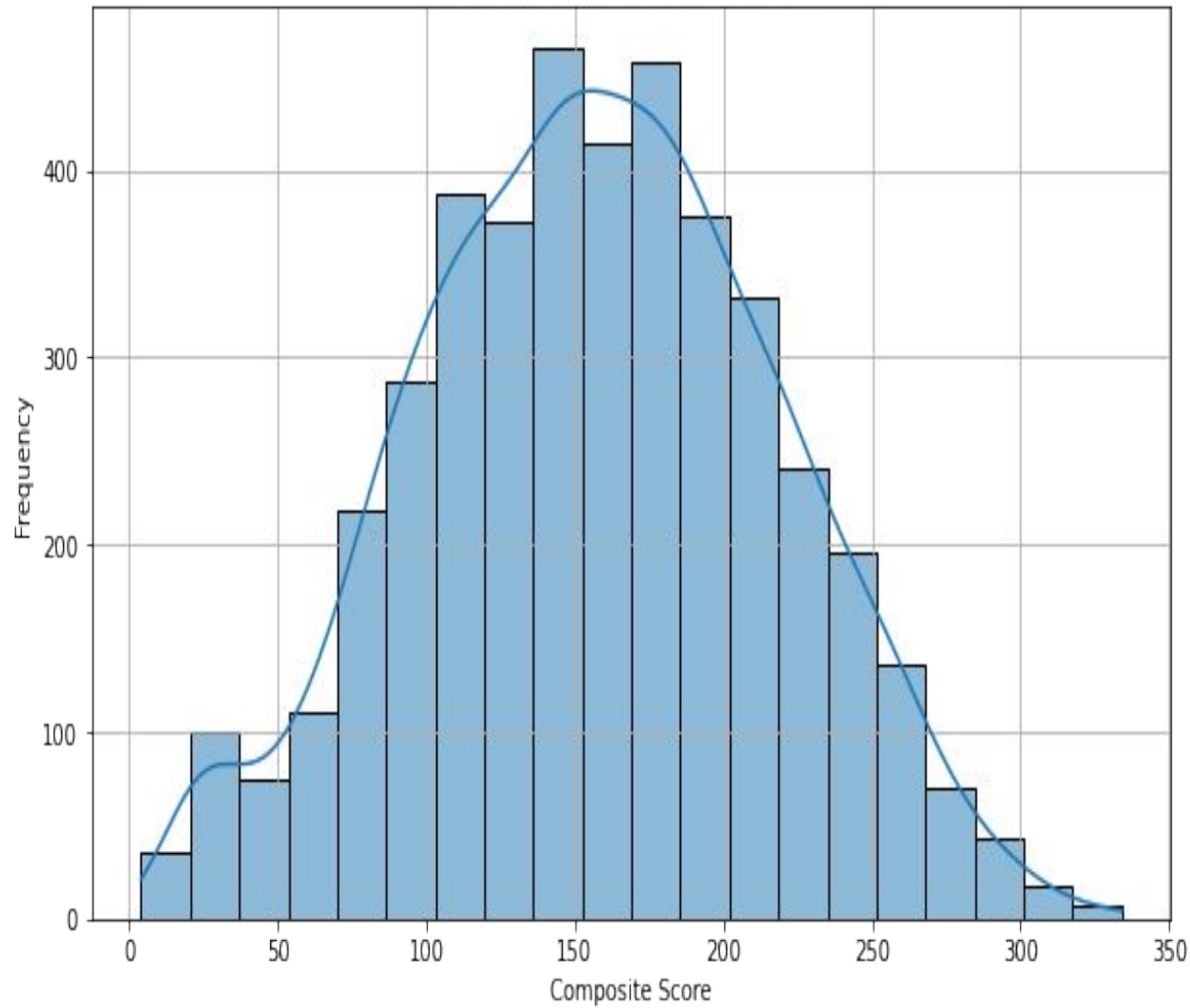
DATA UNDERSTANDING

- 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.



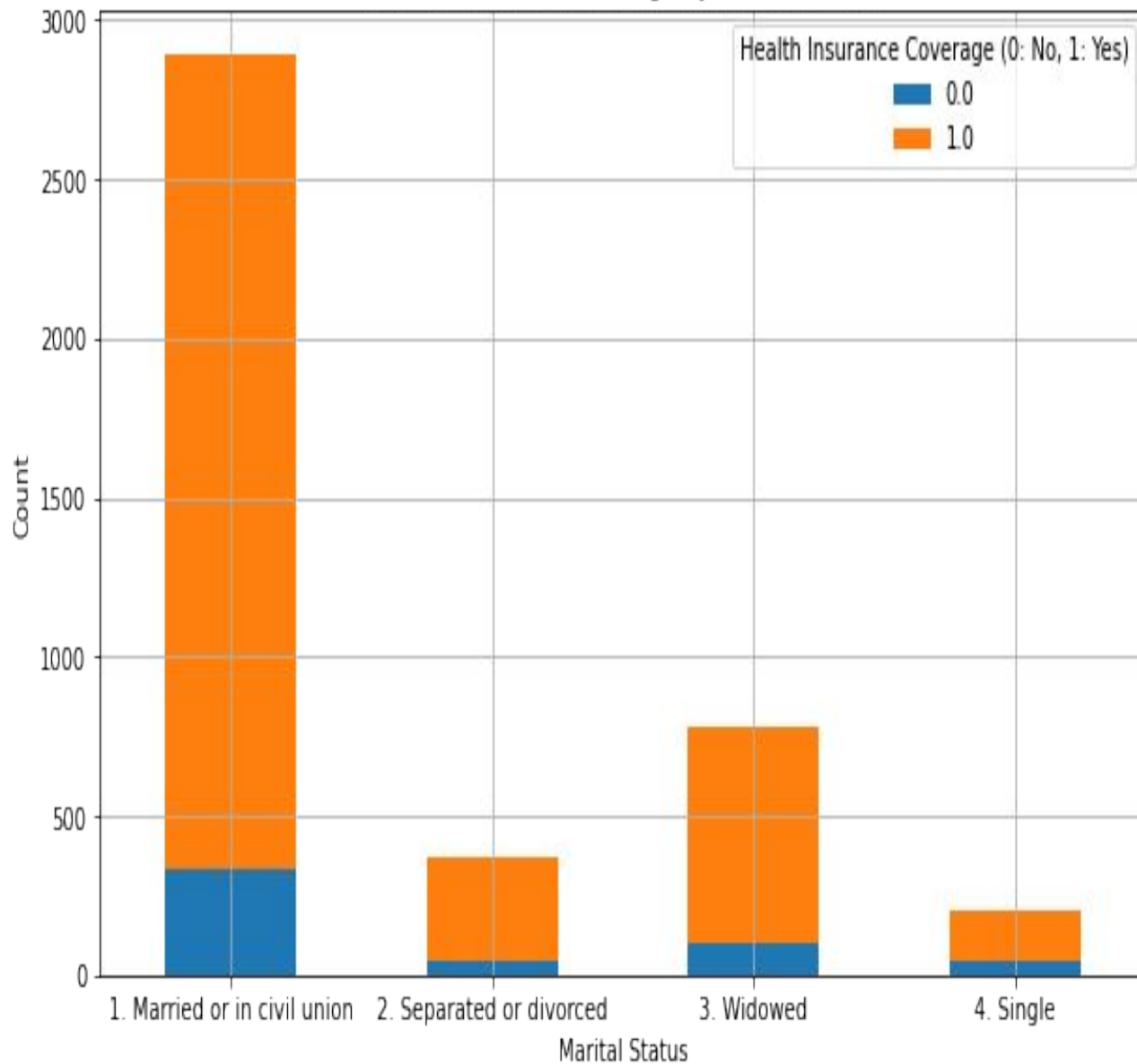
- Majority aged 60–69, mostly married/civil union.
- Urban residents slightly outnumber rural.
- Common education: 1–5 years; most have 3–4 children.
- Spouse gender nearly equal, slightly more women.

Distribution of Composite Scores



- 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.

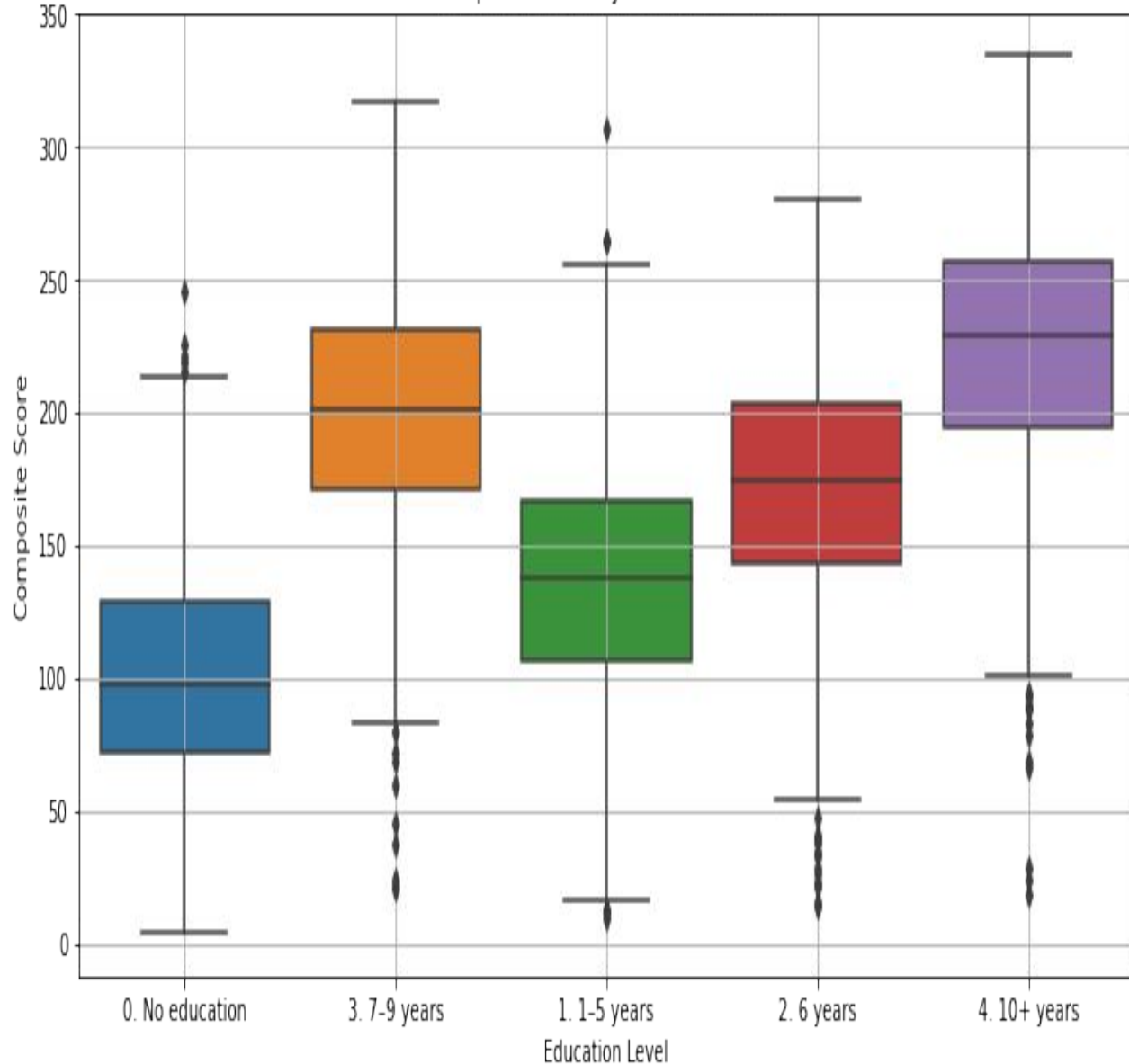
Health Insurance Coverage by Marital Status



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.

Composite Score by Education Level



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.

Performance Overview

- Best Baseline Model: Random Forest (RMSE: 37.8982)
- Linear Regression: RMSE of 41.5095, showing potential for improvement.
- Feature Selection: Improved RMSE to 37.4956, refining model focus.



Key Insights

- Hyperparameter Tuning: Minimal improvement; baseline Random Forest performed best.
- Feature Selection: Improved model performance ($R^2 = 0.5966$).
- Dimensionality Reduction (PCA): Reduced accuracy (RMSE = 48.99, $R^2 = 0.3113$).



Linear Regression Insights

- Polynomial Features: RMSE 37.5991, $R^2 = 0.6137$.
- Ridge Regression: RMSE 22.4754, $R^2 = 0.861$
(explained ~86% variance).



Feature Importance Analysis

- Key Predictors: Education-related variables and age.
- Education: Positive impact
- Age: Negative impact
- Simplification: Removing negligible features improves accuracy.



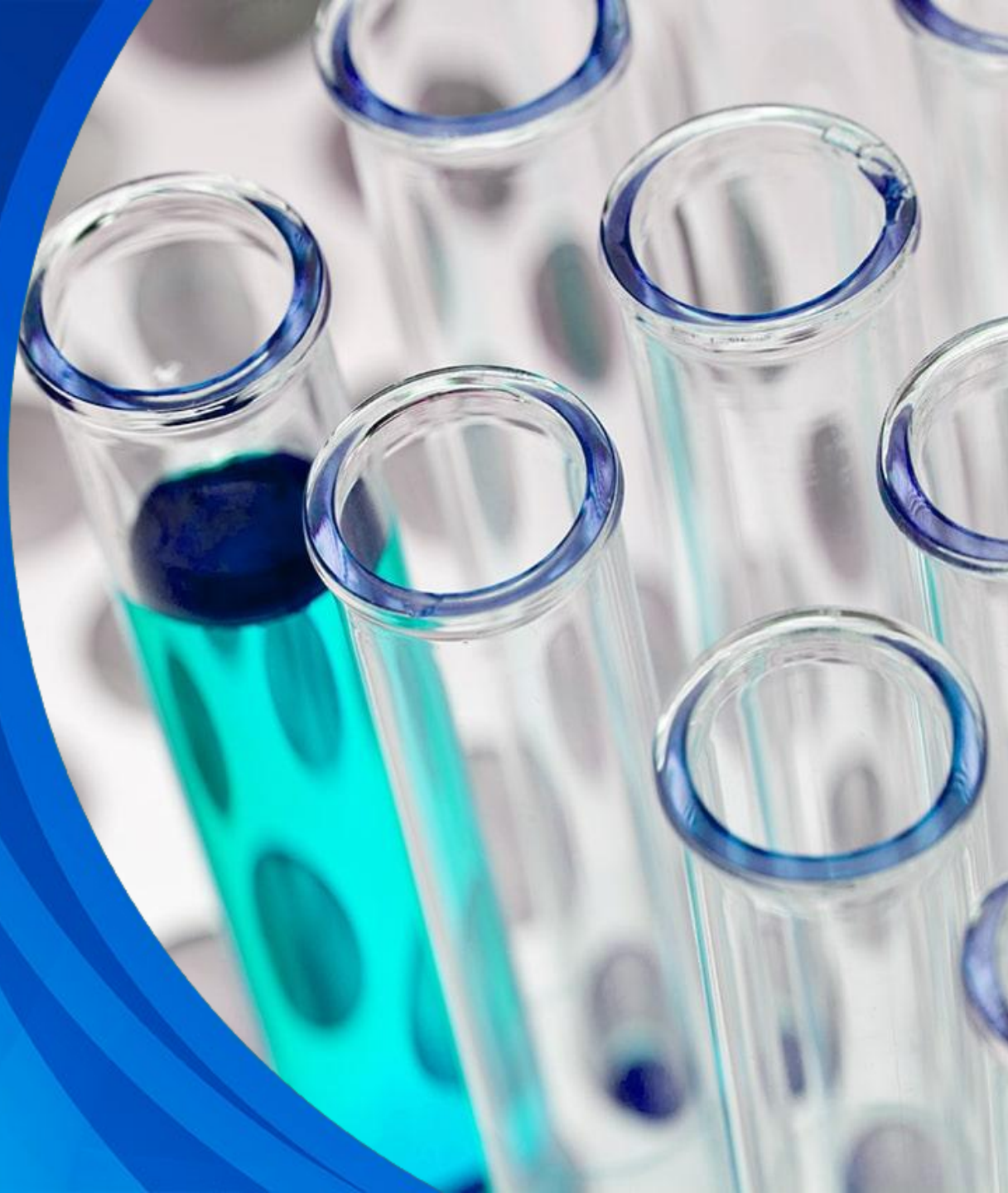
Modelling Conclusion

- Random Forest and Ridge Regression demonstrate strong predictive performance.
- Feature selection and regularization improved accuracy and reduced complexity.



Result: Final Model

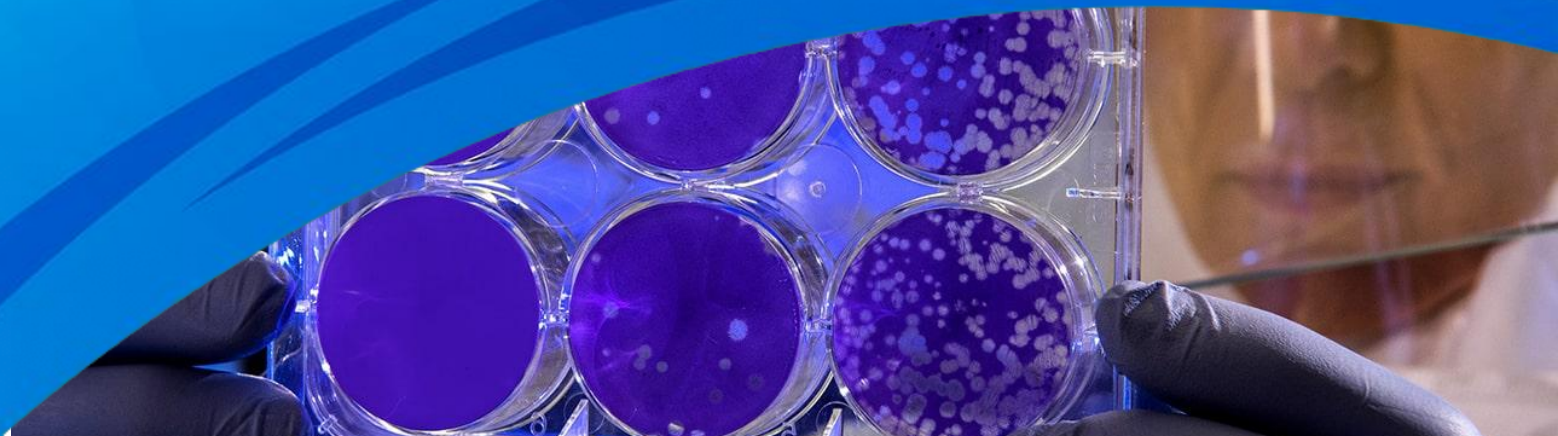
- Best Model: Ridge
Model(RMSE: 22.48)





CONCLUSION:

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



Thank You!

Any Questions?

