



# **Social Determinants of Health Impact on Quality of Life in US Counties: A Predictive Analysis**

Team Beta  
Udumaga Onyeukwu, Miranda  
Gemme, and Steven Uzupis



# Table of Contents

- Background
- Methods
- Findings
- Conclusions
- References

## BACKGROUND: PURPOSE OF STUDY

“**Social determinants of health (SDOH)** are the conditions in which people are born, grow, live, work, and age. They include things like income, education, access to healthcare, and neighborhood environment.”<sup>1</sup>

SDOH have a significant impact on health outcomes.

This project will explore the relationship between SDOH and quality of life.

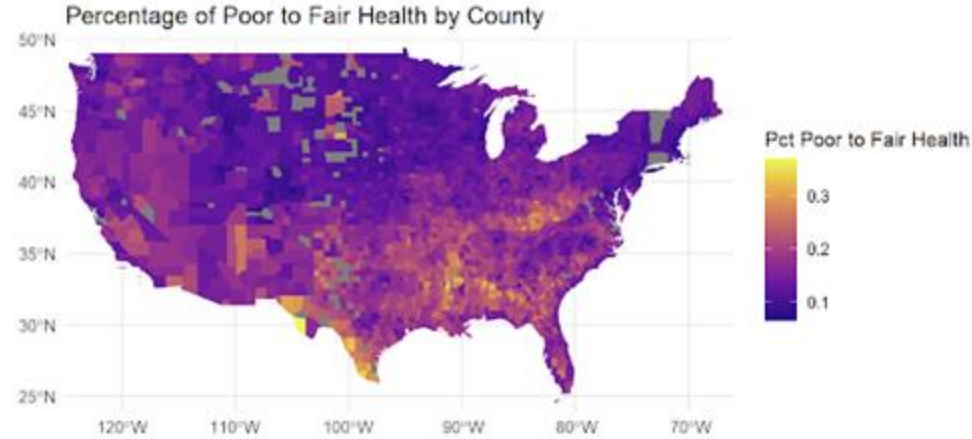
### Client

County health Officials such as members of the National Association of County and City Health Officials (NACCHO)

### Why the Client Should Care

Understanding the relationship between SDOH and quality of life enables NACCHO and local health departments to prioritize health improvements and equity, using data-driven insights to inform policy and resource allocation decisions.

Our analysis will rank the SDOH in order of importance and provide a tool to predict the likely effect of resource allocation decisions on the quality of life of citizens.



**The heat map shows the relative percentage of adults reporting “*POOR TO FAIR HEALTH*” across US counties.** (source data: *County Health Rankings dataset*)

## Background cont'd : Question, Hypothesis, and Predictions

### ❖ Question

**How do social determinants of health affect quality of life in different localities?**

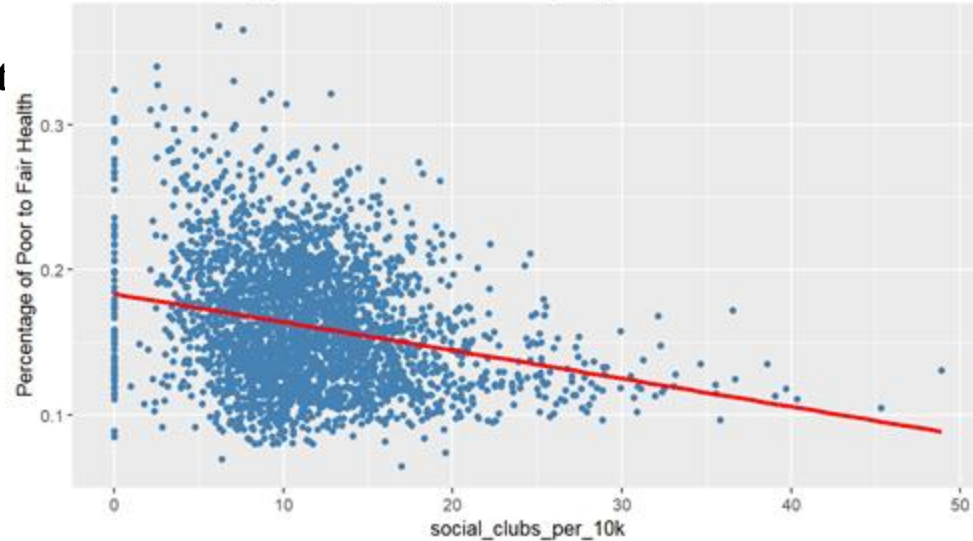
### ❖ Hypothesis

**Social determinants of health, such as economic stability, social connectedness, access to healthcare, and neighborhood environment, significantly predict the self-reported health status of citizens in US counties.**

### ❖ Predictions

**US counties with higher economic security, stronger social support infrastructure, better access to healthcare services, and safer, more accessible neighborhoods will report better overall health status than counties lacking these social determinants.**

Scatter Plot of social\_clubs\_per\_10k Vs 'Poor\_To\_Fair\_Health' Reported Per Count data from [www.ahrq.gov](http://www.ahrq.gov) and [www.countyhealthrankings.org](http://www.countyhealthrankings.org)



**The scatter plot above compares the percentage of adults reporting “*POOR TO FAIR HEALTH*” with number of social clubs per Ten Thousand citizens. (source data: *County Health Rankings dataset*)**

## Background cont'd :DATA SOURCES | DATA COLLECTION| DATA PROCESSING

- ❖ **Data Sources: Social Determinants of Health (SDOH) Database** provided by the **Agency for Healthcare Research and Quality (AHRQ)** and The **County Health Rankings dataset** is produced by the **University of Wisconsin Population Health Institute**
- ❖ **Data Collection:** SDOH data collected from AHRQ database and health outcome data collected from County Health Rankings dataset. Both datasets draw from existing sources such as American Community Survey(ACS) and National Census Data.
- ❖ **Data Processing:** Data is initially processed to ensure consistency and compatibility, missing values are imputed using appropriate methods or discarded depending on the level of missingness, and data is transformed as necessary to meet analysis requirements.



Links to data sources

1. <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html#download>
2. <https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation>

# Methods

Objective: Clean and simplify the data

How?

- Removed unnecessary and redundant features from the SDOH and CHR datasets.
- Combined the datasets using county FIPS codes, targeting approximately 3000 unique observations by initially removing features with significant NA values.
- Further refined by eliminating features with high correlations and collinear values.
- Applied zero variance feature removal, reducing an additional feature.

Why?

- Make it more manageable and eliminating redundancy while keeping the most relevant features.



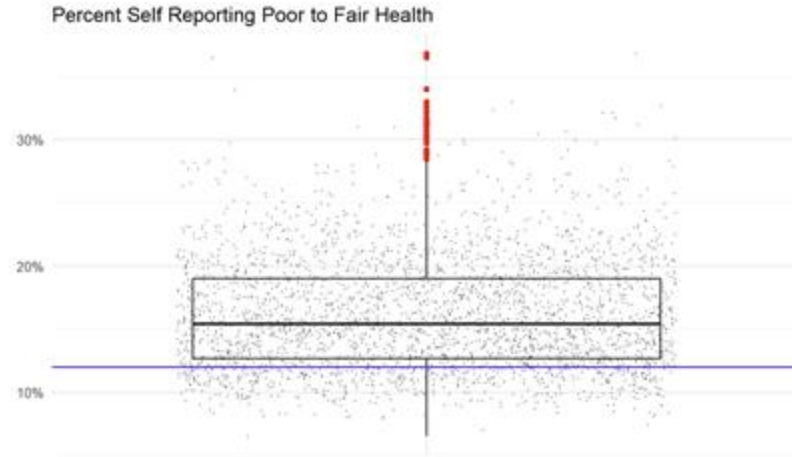
# Methods - Exploratory Data Analysis (EDA)

## Key Actions:

- Analyzed the distribution of the response variable (**pct\_poor\_to\_fair\_health**) as a **numerical continuous response**.
- Observed the key patterns and relationships between predictors and the response variable (**pct\_poor\_to\_fair\_health**).
- Conducted summary statistics and graphical analysis to understand the the patterns and data distribution.

## Insights:

- Created a more accurate measure of health status, focusing on regions where factors have the most significant impact.

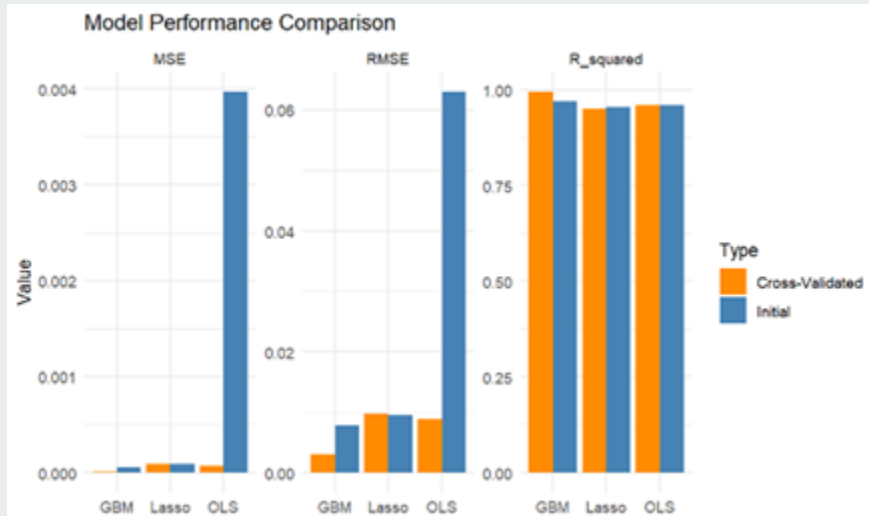


The box and whiskers plot above shows the spread of the response variable.

# Methods



Ideal Model: Gradient Boosting Machine (GBM)



## Models Tried:

- Ordinary Least Squares (OLS)
- Lasso Regression
- Gradient Boosting Machine (GBM)
- Random Forest

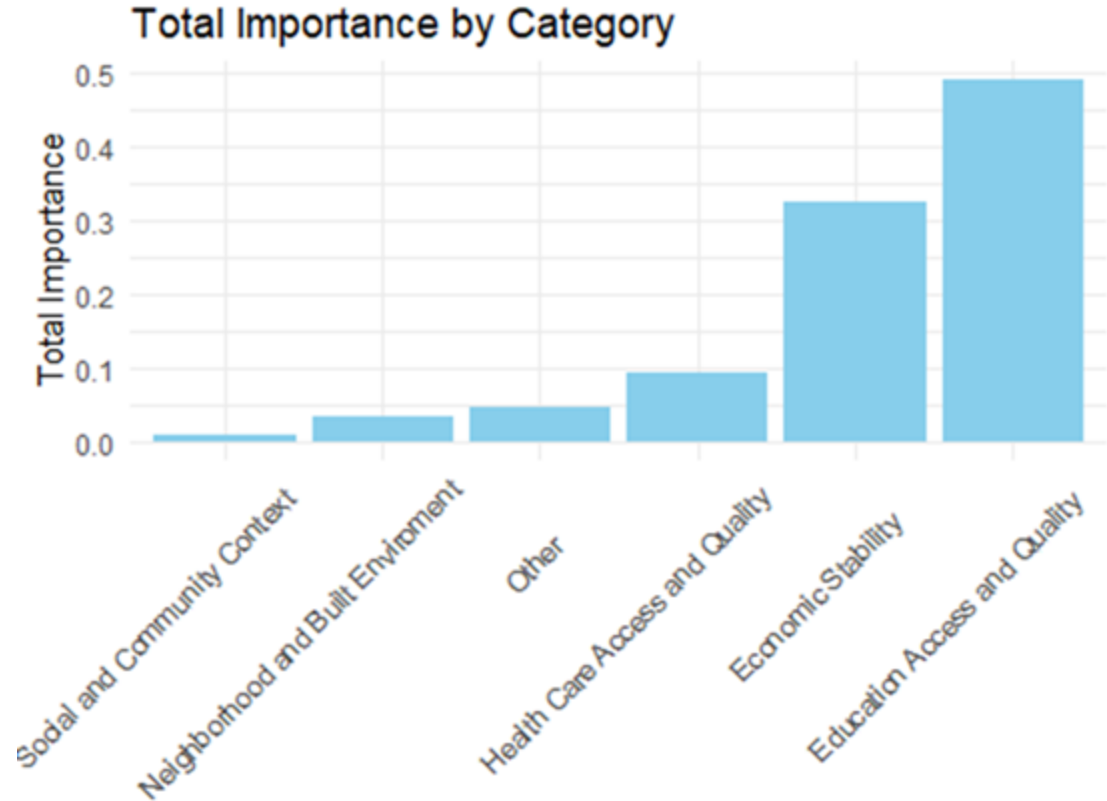
## Key Insights:

- The Gradient Boosting Machine (GBM) model worked best, balancing accuracy and reliability.
- The issue of overfitting was remediated by monitoring improvement by rounds of ten and tracking RMSE for change direction in measurement.



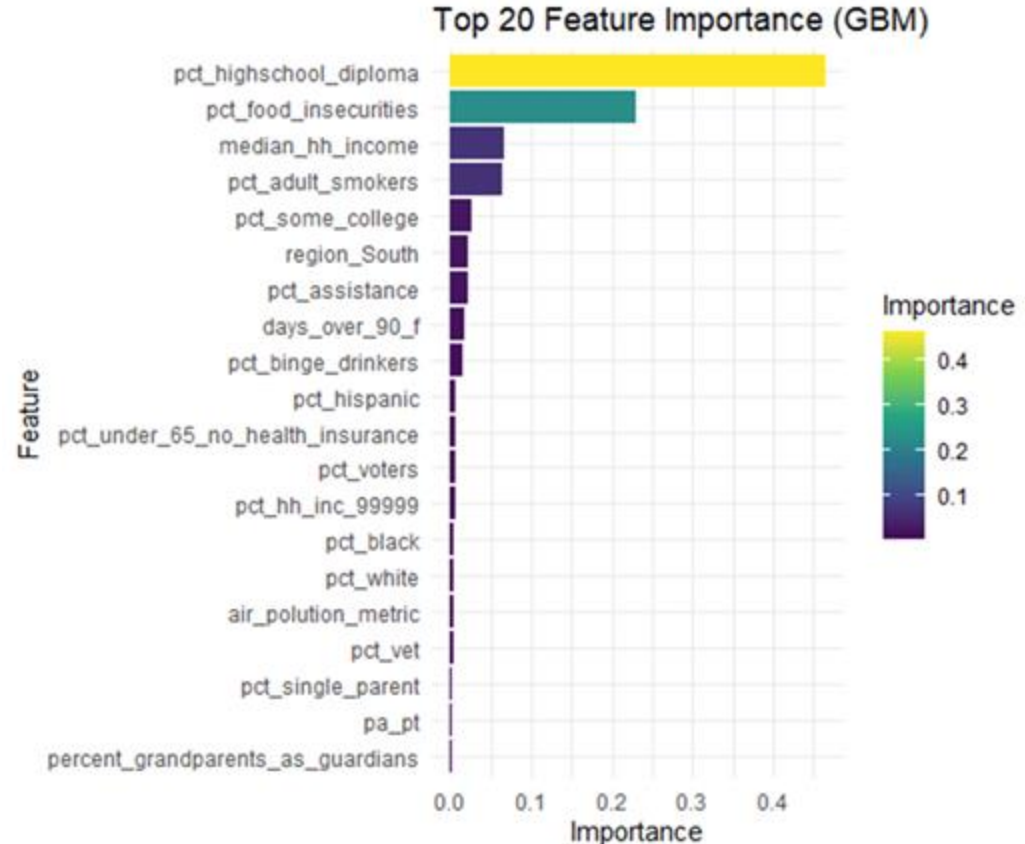
## FINDINGS: WHAT THE MODEL TELLS US ABOUT THE BROAD CATEGORIES

- Understanding categories of social determinants of health to for importance.
- Predictors importance does not directly correlate to a response outcome, but does represent it's relative importance to other predictors .
- The Other category predictors such as the region, sex, and race, while having some importance, are not a social determinant.



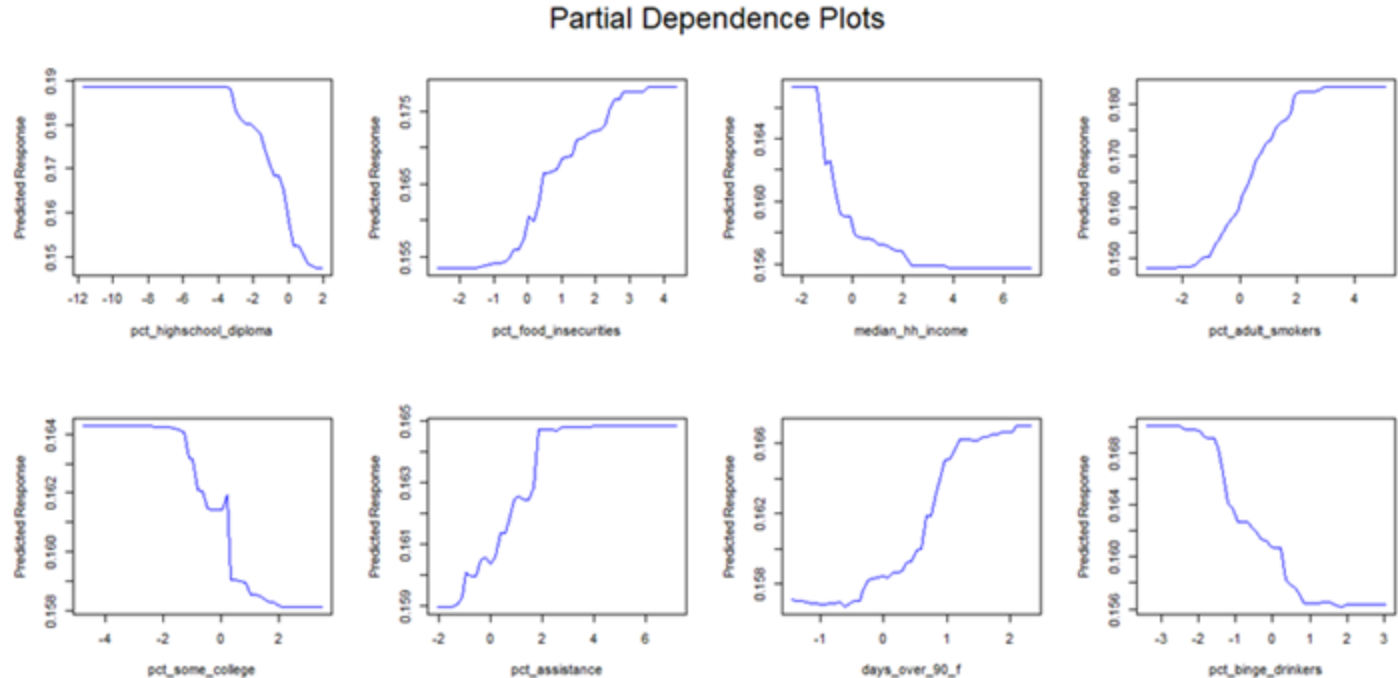
## FINDINGS: WHAT THE MODEL TELLS US ABOUT INDIVIDUAL PREDICTOR IMPORTANCE

- Key features to predict response using GBM Model
- Obtaining a highschool diploma is the largest predictor of counties with poor to fair health.



## FINDINGS: WHAT THE MODEL TELLS US ABOUT THE PREDICTORS INFLUENCE ON THE OUTCOME

- Direction of key predictors from GBM model.
- Modeled data is scaled and centered, so not directly correlate to the predicted response.
- Percent binge drinkers is not predicting what was expected, that counties with less binge drinkers would be healthier.



## CONCLUSION: ANSWERING THE QUESTION HOW TO IMPROVE QUALITY OF LIFE

- Focus on education access and quality, support communities that have lower on time high school graduation rates.
- Economic stability is next category of high importance to focus on. A primary goal of ensuring access to healthy foods for individuals in communities of need.
- Secondary economic stability goal of increasing access to higher paying jobs by attracting businesses to counties with workforce presence,



## FINDINGS: UNDERSTANDING THE MODEL AND NEXT STEPS

- Survey data inherently has features with high multicollinearity, reduce dimensions to reduce bias.
- Reduce features with high multicollinearity using VIF (variance inflation factor) to target features with Adjusted VIF values greater than 5.
- Review predictors used to build model, by analyzing with VIF as first step in feature reduction.

Feature	GVIF	Df	Adjusted_VIF
hh_tot_workers	383.79158	1	19.590599
weighted_population	380.54148	1	19.507472
pct_white	93.81389	1	9.685757
pct_black	72.37672	1	8.507451



# References

- Dan, L., & Valliant, R. (n.d.). [PDF] Condition Indexes and Variance Decompositions for Diagnosing Collinearity in Linear Model Analysis of Survey Data | Semantic Scholar. <https://www.semanticscholar.org/paper/Condition-indexes-and-variance-decompositions-for-Liao-Valliant/42c3f41b2fde70c023df4b9f0ebe0e62566dbeed>
- National data & documentation: 2010-2022 | County Health Rankings & Roadmaps. (n.d.). Retrieved from [www.countyhealthrankings.org](https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation/national-data-documentation-2010-2022) website: <https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation/national-data-documentation-2010-2022>
- Social Determinants of Health Database. Content last reviewed June 2023. Agency for Healthcare Research and Quality, Rockville, MD. <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html>