**Social Effects on Health and Quality of Life for Older Adults**

Preliminary Project Proposal

Team Lead: Udumaga Onyekwy

Recorder: Steve Uzupis

Spokesperson: Miranda Gemme

Data Science Master’s Program, Merrimack University

DSE6311 Data Science Capstone

Dr. Katherine Geist

July 8, 2024

**Executive Summary:**

As the population in the US ages, there is the potential for illnesses in older Americans to become a drag on all aspects of life in the US especially, as we lag other advanced nations in key health indicators. According to the CDC, “Aging increases the risk of chronic diseases like dementia, heart disease, type 2 diabetes, arthritis, and cancer. Older adults are also more vulnerable to severe illness from infections, flu, and pneumonia.” A potential negative to the health of our elderly can be attributed to societal effects due to the conditions lived here that are unique to this society. This project will analyze demographics and data available at a county level in the United States to determine how social factors impact health and the quality of life for older adults.

**Needs Fulfilled:**

This study would fulfill a need to understand how social factors influence the elderly. This information would allow for individuals to identify confounding factors that may lead to a future lower quality of life and correct for them. Where individuals cannot modify these factors, a community organization could work with healthcare providers to find proper interventions. At a larger scale, government policy could be implemented that would influence negative and positive, social factors.

**Project Outline:**

A dataset obtained from the Agency for Healthcare Research and Quality, a Department of Health and Human Services website, contains numerous observations for social statistics about the quality of life for citizens of that county. A few of these observations are state, county name, population in below the poverty line, number of houses, sex, citizenship, spoken language, race, children guardianship, single-family households, access to computer devices and internet access, job sectors, Gini coefficient, household income, access to utilities, mortgage, Medicaid access, commute, access to medical services, land area, population density, and days with excess heat.

These factors can be analyzed to determine expected health outcomes for older adults in those communities. The response is proposed to be a negative health outcome by calculating the percentage of the population aged 65 and older likely to have two or more illnesses as determined using the CDC data on the prevalence of disabilities by demographic groups.

The response shall be calculated using CDC data available at the state level for the presence of disabilities in the population aged 65 and older. The dataset contains the response for the percent population of a state that is positive for a disability. For this study, it will be assumed that the responses in this dataset are uniform across each state at the county level. Using Bayes theorem, the percent of the population that has a negative quality of life in the age group 65 and older will be calculated and a variable will be calculated.

**Hypothesis and Prediction:**

The hypothesis is that the social determinants of health, including economic stability, social connectedness, access to healthcare, and neighborhood environment, significantly impact the quality of life experienced by older adults.

The prediction is that older adults with greater economic security, stronger social support networks, better access to healthcare services, and who live in safer, more accessible neighborhoods will report higher overall quality of life compared to older adults lacking in these social determinants of health.

**Data:**

Predictors:

* [Social Determinants of Health Database | Agency for Healthcare Research and Quality (ahrq.gov)](https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html)
* Currently, the predictors consist of 137 unique categories and 3229 unique observations
* A dataset is loaded into a github repo with specific observations highlighted for use

Response:

* [DHDS - Prevalence of Disability Status and Types by Demographic Groups, 2021](https://data.cdc.gov/Disability-Health/DHDS-Prevalence-of-Disability-Status-and-Types-by-/qjg3-6acf/about_data)
* A calculated regression response that is calculated from a percentage from seven unique disabilities

**Analysis Plan:**

The data sets will be loaded into R to be cleaned, joined, and a response variable will be created. An initial data exploration will be performed to see if any observations are unique. This initial exploration will include finding relations within the data to the response. A best subset of predictors will be identified and used to find an ideal model for prediction using but not limited to Linear Models, Quadratic and Complex Models, and Neural Networks.

**Potential Issues and Ideal Outcome:**

Issues that may arise that could adversely affect this hypothesis are bad assumptions about the response variable. Data available on disabilities was found at either the state level broken down by age demographics or at the county level with no age demographics. It is assumed that this protects individuals from their data being identified and used in adverse ways.

**Sources:**

<https://www.nia.nih.gov/about/aging-strategic-directions-research/goal-society-policy>

<https://www.cdc.gov/cdi/indicator-definitions/older-adults.html#:~:text=Aging%20increases%20the%20risk%20of%20chronic%20diseases%20like,severe%20illness%20from%20infections%2C%20including%20flu%20and%20pneumonia>.