# Capstone 2 Project Proposal - US Life Expectancy at Birth in 2021

Problem Statement:

Predict Life Expectancy at Birth in the United States in 2021 using machine learning regression models.

Context:

Life Expectancy has often been used as a measure of a country's economic condition. Larger life expectancy or longer lifespans indicate that a population is able to progress from birth to old age with the ability to cope with or avoid detrimental or possibly fatal conditions. For example: A country with poor educational attainment and health care system for its population would have a low life expectancy.

Life Expectancy at Birth in the US reflects the varying socioeconomic conditions affecting the population. For this exercise, education, unemployment, and poverty datasets are used to build the model.

Criteria for success:

The criteria for success for this project will be to minimize the root mean squared error, RMSE. This will be accomplished by comparing the performance of several regression models. Each regression model will be optimized using GridSearchCV to find best performing model parameters.

Scope of solution space:

During the modeling process, the influence of education, poverty, and unemployment on life expectancy can be calcalcuted. Example: The median income has x percent of influence on life expectancy.

Constraints:

For this exercise, education, unemployment, and poverty datasets are used to build the model.

Data sources:

The data used in this project were obtained from the following websites:

1. CDC US Life Expectancy Data
2. USDA Economic Research Service
3. US Census State and County FIPS Codes