

## Project proposal: **Predict cancer mortality rate by region**

Cancer is the one of the leading causes of death in the United States. Predicting cancer mortality rates by region and identifying important socioeconomic factors is important for policy makers and healthcare industries. The goal for this project is to analyze the effect of socioeconomic status on cancer mortality rate on the county level. This analysis includes data from [clinicaltrials.gov](https://clinicaltrials.gov), [cancer.gov](https://cancer.gov) & [census.gov](https://census.gov) to examine cancer trials, mortality, incidence and demographics. The general hypothesis is that poorer and underprivileged regions would have higher per capita cancer incidence and death rates.

The dataset was downloaded from [data.world](https://data.world). The data was aggregated from several sources. It consists of 34 variables and information for 3047 counties in the United states. I plan to train a few models including general linear models and non-linear models and identify the best model that can predict the cancer mortality rate. I The model will also be used to identify the major factors that influence cancer mortality rate.

This results of this experiment would be useful to healthcare providers, the general public as well as policy makers.