Healthcare Analytics Project

Purpose: Find out how various predictors such as age, gender, obesity, smoking, drinking, immunization, diet, & exercise explain aggregate human health.

Questions:

* Should being overweight or obese count as categorical values, or should something like BMI be used?
  + If it’s the latter, should alternatives like Body Adiposity Index (BAI), Waist Circumference Measurement, Waist-to-Hip Ratio, Hydrostatic Weighing, or Body Fat Measuring be used.
  + For brief explanations of these, visit <https://www.thecalculatorsite.com/articles/health/alternatives-to-bmi.php>
* Should smoking be a simple yes/no value, or should rate of smoking be used?
* How should diet we quantified?
  + Perhaps diet could be divided into the different food groups, and average daily/weekly/monthly/yearly servings could be investigated to see how diet has an impact on human health.
  + It is possible that this will have a covariance effect with other potential predictors.
* How should exercise be quantified?
  + One idea is to quantify exercise based on average daily/weekly/monthly/yearly time spent.
  + From there, the exercise could be grouped into various types of muscle groups.
* What indicators should be used to measure human health?
  + Life expectancy?
  + Adult mortality rate?
  + Infant mortality rate?
  + Cost of a standard hospital stay?
  + Length of a standard hospital stay?
  + Cardiac revascularization rate?
  + Proportion of health care officials in area?

Possible Datasets:

<https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>

<https://ourworldindata.org/diet-compositions>

<https://ourworldindata.org/smoking>

<https://apps.who.int/gho/data/view.main.2463>

Compare Results Against

<https://ourworldindata.org/grapher/number-of-deaths-by-risk-factor>