Volume 3 Project Proposal

Question 1 Who will work on this project with you (remember groups must be at least 2 people and no more than three people)?
Zachary Evans, Daniel Treuhaft, and Teayoun Kim

Question 2 What the data source you will use?
https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who
This data set has a lot of information of life expectancy in different countries.

Question 3 Data sets you will use to try to answer the questions?

- Can we use data from other countries to predict if switching to universal health care would increase the life expectancy in a certain country?
- Does increased government expenditure on health care have a positive correlation with life expectancy?
- Does the country's GDP impact this correlation?
- What are the most important factors in determining life expectancy for a country? Can these factors be summarized well by a decision tree?
- How well can life expectancy be predicted by infant mortality rate?

Question 4 Data sets you will use to try to answer the questions.
We also would want to add one more column using our own research based on if the country has universal health care. https://www.internationalinsurance.com/health/countries-free-healthcare.php

Question 5 Techniques you think you will use?

- We can use AdaBoosted Trees to show which factors lead to higher or lower life expectancy.
- GeoPandas can help us find geographic correlations between the data.
- We will likely use PCA to find the factors most associated with life expectancy in general.

Question 6 Metrics you expect to use to decide how good your answers are
We hope to be able to use the current data set with which has a large change in later years to compare to our analysis of how things turned out for other countries

Question 7 How you will divide the work among the different team members
We can divide the work according to a correlation we are interested in. We also are planning to meet on Wednesdays at 5