- What is the problem you are attempting to solve?

I am trying to solve the problem of how to predict heart disease. I have family members who have been diagnosed with heart disease and I wanted to help find the predictors to see if they may get it. About 610,000 people die of heart disease in the United States every year–that's 1 in every 4 deaths. Also, heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2009 were in men.This isn't a fool-proof way to determine if someone will get heart disease but checking via data can certainly give us a better idea of what to look for to help prevent heart disease. By being able to find preventative measures against heart disease, I want to cut these diagnoses by 20% if not more.

- How is your solution valuable?

Currently, heart disease is the number 1 killer of adults.If we can reduce that in any way, it's a win for humanity. Those that use my analysis will be able to see what actually causes heart disease rather than speculation. Currently, we assume that age and cholesterol are the primary factors to heart disease and I believe that assumption is incorrect. Users will take the results of this analysis and find better ways for heart disease prevention.

- What is your data source and how will you access it?

I am retrieving my data from Kaggle who got it from the University of California, Irvine. Here is a link to the dataset: https://www.kaggle.com/nikitagrec/heart-disease-uci-analysis/data

- What techniques from the course do you anticipate using?

Supervised and unsupervised learning The data has variables such as cholesterol, age, and sex that could be possible factors in diagnosing heart disease. These are the previously assumed factors in heart disease, I want to see what the data says.My methods will include random forest classifying, KNN classifying, Gradient Boosting, and utilizing Keras. The specialization I choose from Unit 6 is Tensorflow and Keras and anticipate what that will do to my data. Keras is used for deep learning and a highly sought after skill so I wanted to incorporate that in my analysis.

- What do you anticipate to be the biggest challenge you'll face?

A potential hurdle is that my data may not accurately represent the population seeing as my sample size is rather small (1000 rows).