The dataset I have identified for the second capstone project is a Heart Disease Dataset sourced from Kaggle. It has 14 variables: age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting ECG results, maximum heart rate achieved, exercise induced angina, oldpeak (ST depression induced by exercise relative to rest), slope of the peak, number of major vessels colored by fluoroscopy, presence/type of heart defect, and the presence of heart disease in a patient.

1. S: Preventative medical is an approach that can be taken for individuals who are known to be predisposed to a certain condition. By creating a predictive model based on vital statistics and individual health factors, doctors may be able to determine if somebody has a chance of developing heart disease later in life.
2. M: The Heart Disease Dataset contains vital statistics and health information on both individuals who were diagnosed with heart disease and who were not diagnosed with heart disease. Using the information included in the dataset, a model can be built to see how the different variables relate to each other and to an overall diagnostic outcome.
3. A: After verifying that there are no issues with the dataset, using exploratory data techniques the Heart Disease Dataset can be analyzed to achieve the goal of this project.
4. R: This is a worthwhile analysis to conduct. It would provide insight into the future of a patient’s health and allow them to potentially make conscious health-based decisions that could prevent the onset or severity of disease.
5. T: A timeline of no more than 1 month would be appropriate to perform the analysis, given that the analysis techniques are already established and there is only a single dataset. If this project were to be expanded, the validity and strength of additional datasets would need to be considered, but for this single one it would not take more than 1 month.

The criteria for success for this project would be determining if the dataset is suitable for analysis and then the construction of a predictive model based on the information in the dataset. There are still some constraints; the dataset only has 14 variables of an innumerous potential vital statistics and comorbidities so it may not be inclusive of other important factors that could contribute. It also only has about 1000 observations, which may not be strong enough to build a predictive health model compared to a larger dataset. The only required datasource for this project would be the initial Heart Disease Dataset, but if the analysis were to be expanded further than other datasets could be included, such as BRFSS, MEDSIS, or other health based datasets and surveys.