**Milestone 1: Background**

Cardiovascular diseases (known as CVDs) are the leading cause of death globally, taking nearly 18 million lives per year, or 31% of deaths worldwide. 80% of CVD deaths are due to heart attacks, or strokes, and a third of these deaths occur prematurely in individuals under 70 years old. Heart failure is a common event, caused by CVDs. This data set contains 11 different variables or features that can be used to predict a possible heart disease. These include:

1. Age
2. Sex
3. Chest Pain Type (Typical angina, Atypical angina, non-anginal pain, asymptomatic)
4. Resting Blood Pressure, mm Hg
5. Cholesterol (serum cholestrol mm/dl)
6. Fasting blood sugar (1 >120mg/dl, 0 otherwise)
7. Resting Electrocardiogram results (Normal, ST-T wave abnormality, LVH showing probable or definite left ventricular hypertrophy by Estes’ criteria
8. Max Heart Rate achieved (60-202)
9. Exercise Angina (yes or no, angina caused by exercise)
10. Old Peak = ST (numeric value measured in depression)
11. ST\_slope or the slope of the peak exercise ST segment (upsloping, flat, or down)
12. Heart Disease (output of other 11, 1=has heart disease, 0 = normal

This Data set was created by combining different datasets that had already been independently available. It includes 5 data sets with data from Cleveland, Hungary, Switzerland, Long Beach Virginia, and from the Stalog (heart) data set. The total amount of records that this data set contains is 918 individuals.

**Analytics Objectives**

When attempting to see what influences CVDs, can we model the individual’s health indicators to analyze what combination of parameters makes an individual more likely to have a CVD?

What is the most influential factor of the listed health indicators that has the strongest correlation in determining whether an individual has a CVD?

What factors have the least impact or influence on determining whether an individual has a CVD?