The data set that I am taking a look at is a repository of people who have gone in for a clinical heart checkup. These members believe that they are at risk of heart failure so they went to go in for a checkup. The data set that I am using can be found here at https://www.kaggle.com/andrewmvd/heart-failure-clinical-data. This data set is a .csv and it contains the age of the patient that went in for the clinical, their sex, whether they smoke, have anaemia, diabetes, high blood pressure or platelets in thier blood. This data set also has records of the patients creatine phosphokinase (CPK enzyme) levels, serum creatinine levels and serum sodium levels in the blood as well as the percentage of blood leaving the heart at each pump. Finally, the data set also records the patients follow up time since their first meeting where this data was recorded, and marks whether the patient died between the initial clinical and the follow-up. I would like to find if there are any significant factors that lead to heart failure, and if those methods are things that can help prevent heart failure.

Heart failure is the leading cause of death in the world, and they account for 31% of the deaths worldwide. One in 5 men and women develop heart failure during their lifetime. The human body is incredibly complex, and there are no straight forward warnings that can help us prevent heart failures. Being able to prevent something as prevalent as this would be a huge boon for the average life expectancy. As such, I am hoping that by analyzing this clinical data, I can make some analysis and see if there are certain conditions that raise the risk of heart failure significantly enough that they need to be watched. For example, does being a genetic male or female lead to a higher risk of heart failure? What about a high blood pressure? My father has high blood pressure and it would be terrible for him to be at risk of a heart failure as well. It is well known that if a person is unhealthy, they are at risk for more diseases that can take their life. Does being generally unhealthy make you significantly more at risk for heart disease, or can it be more genetics based? People know that smoking is incredibly unhealthy, but if people that found it hard to quit were shown that it is a leading factor to heart disease, would they be able to quit? Many of these conditions I will be analyzing can be monitored and even eliminated with proper notice and precaution, so it is reasonable to say that if these conditions can cause significant risks for heart disease and death, then heart disease can be effectively monitored itself.

Donald M. Lloyd-Jones, Martin G. Larson, Alexa Beiser, and Daniel Levy. 1999. Lifetime risk of developing coronary heart disease. *The Lancet* 353, 9147 (1999), 89–92. DOI:http://dx.doi.org/10.1016/s0140-6736(98)10279-9

Douglas S. Lee, Peter C. Austin, Jean L. Rouleau, Peter P. Liu, David Naimark, and Jack V. Tu. 2003. Predicting Mortality Among Patients Hospitalized for Heart Failure. *Jama* 290, 19 (2003), 2581. DOI:http://dx.doi.org/10.1001/jama.290.19.2581

Jennifer E. Ho et al. 2016. Predicting Heart Failure With Preserved and Reduced Ejection Fraction. *Circulation: Heart Failure* 9, 6 (2016). DOI:http://dx.doi.org/10.1161/circheartfailure.115.003116

Ken Gu. 1999. Diabetes and Decline in Heart Disease Mortality in US Adults. *Jama* 281, 14 (1999), 1291. DOI:http://dx.doi.org/10.1001/jama.281.14.1291

Michele Senni et al. 2013. Predicting heart failure outcome from cardiac and comorbid conditions: The 3C-HF score. *International Journal of Cardiology* 163, 2 (2013), 206–211. DOI:http://dx.doi.org/10.1016/j.ijcard.2011.10.071

Stanton A. Glantz. 1995. Passive Smoking and Heart Disease. *Jama* 273, 13 (1995), 1047. DOI:http://dx.doi.org/10.1001/jama.1995.03520370089043