

# Final Project Title

By Yuta Chung-Umemoto, Jason Chalmers, Hammad  
Mohammad, and Dylan Steinhauer

# Topic

-The topic we chose is CVDs (Cardiovascular diseases). We are using a dataset that contains predictive factors for heart failure.

# Reason for Choosing Topic

-We chose this topic because it could potentially help people. We wanted to do something that felt important and could be a benefit to society. Heart disease is a serious issue that potentially affects a large number of people. In fact, Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. People who are at risk need early detection and management. This where a machine learning model can be greatly beneficial. We hope that our analysis and observations can at least help with this issue.

# Description of Source of Data

-Our dataset was found on kaggle. This dataset was created by combining five other datasets based on eleven common features. The five datasets used were Cleveland, Hungarian, Switzerland, Long Beach (VA), and the Stalog heart dataset. This is currently the largest heart disease dataset currently suitable for analysis.

# Questions We Hope to Answer with the Data

-Heart failure is a common event caused by CVDs (Cardiovascular diseases) and this dataset contains 11 features that can be used to predict a possible heart disease. The main question we want answered is whether a person has a heart disease that can lead to heart failure. In order to find this out, we want to get more specific insights into which features play the biggest roles in heart diseases. For example, what are the top three features from the dataset in showing a correlation to CVDs? In contrast, which three features contribute the least to heart diseases? Is heart failure more of a concern in males or females? How much of a risk factor is age? Which age ranges are most susceptible to CVDs and heart failure? What is the disparity in people who experience normal angina with those who only experience exercise-induced angina? What role does this disparity play in possible heart failure?

# Description of the Data Exploration Phase

# Description of the Analysis Phase

# Technologies, Languages, Tools, and Algorithms Used

-We used a variety of tools, languages, technologies and algorithms to accomplish our task. We used Jupyter Notebook and Pandas to load our dataset into a dataframe in order to look at and clean the data. For the machine learning aspect, we used a handful of sklearn modules such as `train_test_split`, `StandardScaler`, `OneHotEncoder`, `accuracy_score`, `LogisticRegression`, and `SVC`. We also used Tensorflow.