Mark Perez

Problem statement: It is possible to predict heart failure from correlation between mortality and other features?

Context: 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. In the United States alone, this accounts for about 6.2 million adults. Heart failure is defined as the inability for the heart to pump enough blood and oxygen to support other organs in your body. There are actions that are commonly known to increase the risk, such as, smoking tobacco, excessive alcohol, little physical activity, and eating unhealthy. The dataset used contains other features and the count of mortality of the recorded events. It would be beneficial to hospitals and other health companies if they had additional information for Cardiovascular disease prevention.

Criteria for success: Develop a model, most likely a random forest, that can accurately predict the mortality of somebody with cardiovascular disease based on the features in the dataset. A successful result would be to determine which features correlate with heart failure and to determine the leading feature. Then, relay this information to the healthcare workers and business owners to further spread the information.

Scope of solution space: This assignment will encompass other EDA analyses that have been used in previous Capstone assignments. It will be useful to visualize each feature with box plots or histograms to have an idea to see what correlations can be seen early on.

Constraints: There are no apparent constraints at the moment.

Stakeholders: Business owners and health workers

Data sources: Data uploaded to Kaggle. License: Attribution 4.0 International (CC BY 4.0)

Citation: https://www.kaggle.com/andrewmvd/heart-failure-clinical-data
https://www.cdc.gov/heartdisease/heart_failure.htm