# Correlation Between Heart Activity, Certain Biomedical Traits and Heart Disease

## Data Acquisition

Data acquired is directly from Kaggle repository.

## Data Cleaning Steps

1. Renaming column headers to make them more clear.
   The column names used in the data were very unclear because most of them were abbreviated. I renamed each column to the full clinical name given to that attribute for clarity.

2. Checking consistency and correct data types.
   After implementing the property for data frames, dtype, the data type of each columns was determined and were satisfactory, all columns were numerical. The two data types found were integers and floats.

3. Identifying missing values
   The isnull and sum function was used to determine a count of the total number of missing values in each column. There are no missing values in this data set, however in the case of missing values two methods that are frequently used to correct null values including replacing them with a mean value or deleting the row with null.

4. Identifying and handling outliers.
   Out of the 14 attributes there were 4 attributes which features values were unconfined and capable of having outliers. The zscore was determined using the stats module. This gives the number of standard deviations from the mean with 0 being identical to the mean and 1 being 1 standard deviation above the mean. Threshold value to locating outliers in this module is 3 times the standard deviation. There were 8 values located and for the whole data set and they were dropped.