Summary of Data Cleanup and Exploration:

Going into this project we spent a lot of time trying to figure out how and what to analyze. The best way that we felt this was possible was to begin by pulling the data from the csv and analyzingt what columns and data were being presented in this dataset. We found that a lot of values in the data were binary meaning (1 for yes and 0). We also noticed that different metrics were being used, age was listed by days rather than years. So, our first course of action was to convert these values so that they could be presented more clearly. Once we had our dataframe cleaned up we begin our individual work. We each took one or more lifestyle factors and our goal was to extract information to show whether particpants participated in those activities or if they fell under a certain category (ie weight ranges, BMI Indexes, age ranges). We compared those findings to whether or not those participants had cardiovascular disease or not (findings are summarized below).

Finally Summary and Conclusion:

With our project we sought to find out how lifestyle factors may affect the probability that someone will have Cardiovascular Disease. Going into this we assumed that people who lived less “healthy” lifestyles were more likely to have Cardiovascular Disease. Our findings differed slightly from our initial assumptions. What we found was that individuals who were overweight and/or individuals who were older were more likely to have Cardiovascular Disease. However, individuals who smoke and/or consume alcohol did not differ much in numbers compared to those who don’t participate in those activities.

Through our analysis we also learned that our dataset was skewed in certain aspects. We wanted to take gender into consideration for this but found that there were at least double the amount of male particpants than there were female. As for particpants that were drinkers and or smokers we found that a majority of participants did not participate in either of those activities.