Capstone_Project_Two on Prediction of Heart Health

Heart attack statistics in US are very alarming as follows:
- • Heart disease is the leading cause of death in the United States.
- • One person dies every 36 seconds in the United States from cardiovascular disease.
- • About 655,000 Americans die from heart disease each year—that's 1 in every 4 deaths.
- • Heart disease costs the United States about $219 billion each year from 2014 to 2015. This includes the cost of health care services, medicines, and lost productivity due to death.

Data acquisition and EDA
A data file is acquired from Kaggle (Heart.csv) and cleaned from duplicates. The data frame has features like male/female, chest pain type, resting blood pressure, fasting blood sugar, resting ECG, exercise induced Angina. All these features are shown visually and numerically.

Predictions: Females have higher likelihood for heart attack than males, exercise induced angina has significant effect on heart health, recommendation is to go easy and slow. Blood sugar does not have major effect on the health of heart. ECG relationship on Heart attacks is also presented.

Machine learning algorithms like Logistic Regression, Naive Bayes, Support Vector Machines, K nearest Neighbors, Decision Trees and Random Forests are used to make predictions. Random Forests has the highest prediction accuracy of all being 86.89% with 200 trees.
It is important to note that Random Forests has 86.90% prediction accuracy with 2000 trees taking extremely long run time.

It is my hope this analysis helps people realize the importance of taking care of vascular health with good food, moderate exercise and keeping the mind calm to reduce stress