Statistical and Machine Learning Analysis of Contributing Factors to Heart Disease

Jane Chea and Beth Gronski

CSE 163: Intermediate Python Programming

13 March 2023

Table of Contents

Table of Contents 2
Summary of research questions 3
Motivation 4
Dataset 4
Method 5
Results 7
Impact and Limitations 15
Challenge goals 16
Work Plan Evaluation 17
Testing 19
Collaboration 20

Summary of research questions

1. Which factor plays a more important role, or is most significant, in contributing to a patient's heart disease?
    a. After using our new models, we found that the number of major vessels (ca), the ST depression reading after exercise relative to rest (oldpeak), and chest pains (cp), specifically non-anginal and typical anginal, contributed the most to a patient developing heart disease. When looking at only the most significant values in the dataset, we found that in addition to everything previously listed, all types of chest pains including atypical anginal, nonanginal, and typical anginal chest pains were significant.
2. Can these factors create an accurate machine learning model that will predict with decent accuracy if a patient will get heart disease?
    a. We created two main machine learning algorithms focused on the general data and then the data with the most influential factors. When looking at the general data without sectioning, we were not able to find an accurate enough predictor of heart disease. With this algorithm as well, we changed the number of neighbors considered in the model to see if there would be changes to the accuracy, but it overall provided a lower accuracy than we were hoping. However, once we had sectioned the data off to the most influential factors, we were able to find a model that could predict heart disease more accurately.
3. Do the heart disease factors differ between locations within the data?
    a. We incorporated location into our analysis of most influential factors in determining whether a patient develops heart disease and we see that it is not significant in determining this. When filtering the data based on the US and Europe, we found that the US had the same most influential factors as our answer to the first question: the typical anginal and non-anginal chest pain (cp), the number of major vessels (ca), and the ST depression difference between exercise and rest (oldpeak), thereby increasing our confidence in those observed results. Europe could not be analyzed due to the limited amount of data present in the dataset.
4. How does gender and age affect which factors are most significant in determining if a patient will get heart disease?
    a. Using the statistical models again, we found that there was some variance when separating the data into specific criteria of gender and age, which we divided based on the available data into female and male and under and over 55, respectively. From this, we found that the factors that most

significantly influenced getting heart disease in females were non-anginal chest pain, the number of major vessels, and having an ST segment T abnormality in the ECG reading (restecg). For both males and the group under 55, we found the same factors that influenced heart disease as in the first question we answered: the number of major vessels (ca), ST depression difference between exercise and rest (oldpeak), and typical and non-anginal chest pains (cp). Finally, with the group over 55, we found that typical anginal chest pain (cp) and the number of major vessels (ca) were the only two factors that affected heart disease development.

Motivation

The CDC has cited heart disease as the leading cause of death in the United States. Heart disease consists of numerous types of heart conditions and has many different factors that are correlated with higher risks of developing these diseases. We care about this problem due to its prevalence and how it can commonly afflict people we care about. Knowing what lifestyle changes can be made to lower the risk of heart disease is useful and informative for choices that we make throughout our lives to remain physically healthy. Our research questions hope to provide some insight about these factors by comparing our machine learning predictions to predictions from medical professionals. This will hopefully allow us to understand what factors may be predictors for heart disease, so we and other people can make more informed choices based on those predictors.

Dataset

The dataset was created by Andras Janosi, M.D. from the Hungarian Institute of Cardiology in Budapest, William Steinbrunn, M.D. from the University Hospital, Zurich, Switzerland, Matthias Pfisterer, M.D. from the University Hospital, Basel, Switzerland, and Robert Detrano, M.D., Ph.D. from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. The dataset consists of 14 features: “id” (Unique id for each patient), “age” (Age of the patient in years), “origin” (place of study), “sex” (Male/Female), “cp chest pain type ([typical angina, atypical angina, non-anginal, asymptomatic]), “trestbps” resting blood pressure (resting blood pressure (in mm Hg on admission to the hospital)), “chol” (serum cholesterol in mg/dl), “fbs” (if fasting blood sugar > 120 mg/dl), “restecg” (resting electrocardiographic results), “thalach”: maximum heart rate achieved, “exang”: exercise-induced angina (True/ False), “oldpeak”: ST depression induced by exercise relative to rest, “slope”: the slope of the peak exercise ST segment, “ca”:

number of major vessels (0-3) colored by fluoroscopy, “thal”: [normal; fixed defect; reversible defect], and “num”: the predicted attribute. There are 920 rows, which represent 920 patients from the 4 different locations of the doctors who conducted this study and compiled this dataset. This dataset was updated 2 years ago, so it is fairly recent data. This data was collected from Kaggle, but it can also be found in the UCI data repository.

In consideration of limitations, this dataset only comes from patients in Cleveland, Hungary, the VA in Long Beach, and Switzerland. This limits the variety in patient data, so the data may not be representative of the world’s population and what attributes may contribute to them developing heart disease. Additionally, the dataset only provides predicted values for the patient’s heart disease level, so we are unsure if these are accurate. So our analysis will focus on analyzing which factors contribute to the predicted levels of heart disease. However, this does not discredit this analysis due to the credibility of the doctors, their extensive education and background in medicine.

https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data

Method

To answer the question of which factor plays a more important role in contributing to a patient’s heart disease, we will use the ‘statsmodels’ library from python. We will create a multiple regression model and then produce a summary of coefficients, standard deviations, p-values, t-values, and R-squared values to determine how significant each factor is. To accomplish this, we will be using the statsmodels functions to fit a regression model, specifically ols() and fit(). Within these functions we will compare the predicted level of heart disease to these factors: age, sex, cp, trestbps, chol, fbs, restecg, thalch, exang, oldpeak, and ca. We chose these variables because they have a very low number of NA values. From this model, we will use the library’s summary function to create a table of the results from our analysis. To discover which factor is most significant, we will first look at the R squared value, and this will tell us how well fit our model is in general. If it is a high R-squared value, such as 0.9 and above, we will know our model is good in predicting the heart disease level. Next we will look at the p-values and t-values of each variable. If the p-values are low, ideally 0.000, and the t-values are large, larger than the t-critical value (which will be discovered in the analysis), then we know this variable is statistically significant. Once we have evaluated which variables are statistically significant, we will then view their coefficients, and among those that are significant we will know that the one with the highest coefficient has the most influential effect on the level of heart disease.

To answer the question of which factor is most significant at each predicted stage of heart disease, we will be conducting a similar analysis as to the one in the first question. Before we perform that analysis, we will filter our dataset into 5 smaller dataset. Each will be filtered by the predicted level of heart disease (0, 1, 2, 3, 4). 0 will be used as a control, and also to see the values that come out of that analysis for comparison. Then, following the steps in the first question, we will fit a multiple regression model and then analyze the results to find the most significant factor. After this analysis, we will compare which factors were most influential at each stage. If one factor dominates every category, then we will know this factor is the most significant. However, if we see variation we will then come to the conclusion that different factors may be more influential depending on the stage of heart disease.

To answer the question of whether these factors create an accurate machine learning model that will predict the level of heart disease in a patient, we will create this model and then compare the accuracy scores of the testing model. To begin, we will first need to change all of our categorical variables into dummy variables in order to create a machine learning regression model. Then, we define the features and labels from our dataset, and then separate these into training and testing data. We will then fit the model with the training data, and compute the accuracy of this model. From this analysis, we will adjust the max depth of our model to produce the most accurate outcome when considering our testing data. After this we will run our testing data, and determine the accuracy. If we achieve a high accuracy, we will have answered our question as yes we can use this model to accurately predict. Additionally, we will compare outputs from this machine learning model to our statsmodel as the concepts in creating both are very similar.

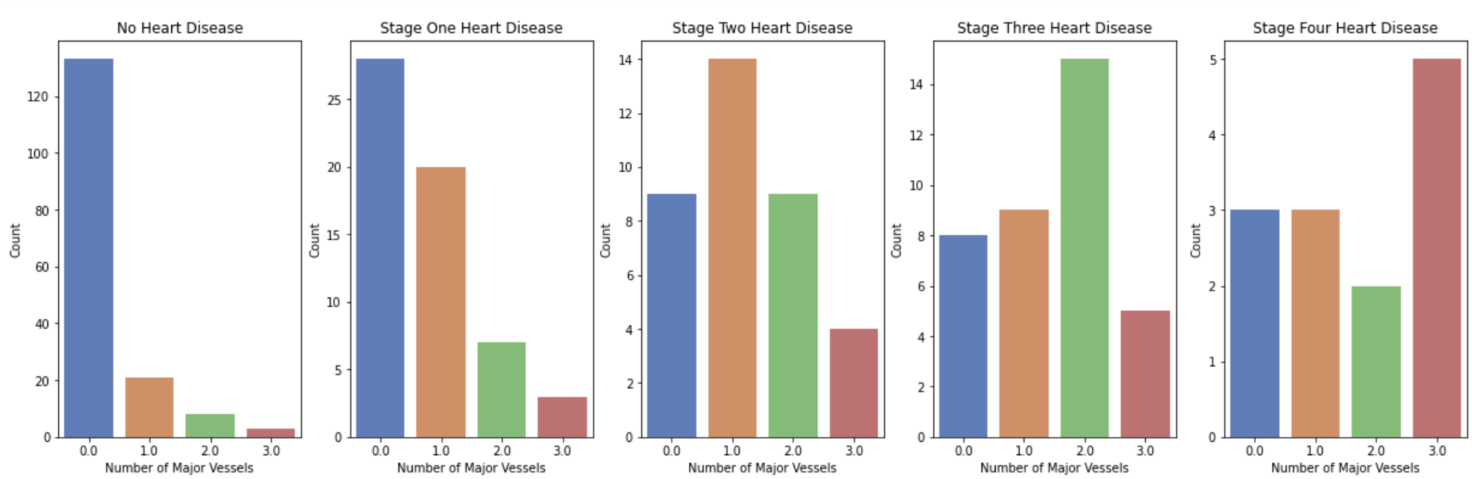
As we filtered our data in the second question on different stages in heart disease, we will answer the question of whether location plays a role by performing the same analysis but filtering by location. With this approach, we aim to discover if the results differ drastically depending on the location of the patients. Now with this analysis, we will be lenient on how much the models may differ. Once the model is split by location, the amount of patients per location will be small, especially considering that Cleveland and Hungary dominate the data set by 65%. If we see a significant difference, then we can determine that location plays a factor, but if there is minimal difference we cannot conclude that location plays a role.

If time allows, we will also try to answer the question of whether gender or age changes which factor is the most significant in contributing to heart disease. Like the location question before, we will filter the data by gender or age, as we did with location, and then perform the same statistical analysis. In consideration to gender, there are predominantly more males in this dataset than females, about 4 times more. So this analysis does not provide much representation to females. Also, the age range is significant, and any analysis done on age will have to be divided into age groups. This division into groups may create even more caveats to the accuracy of the analysis, and that is why this analysis will be less important and more for curiosity and exploration.

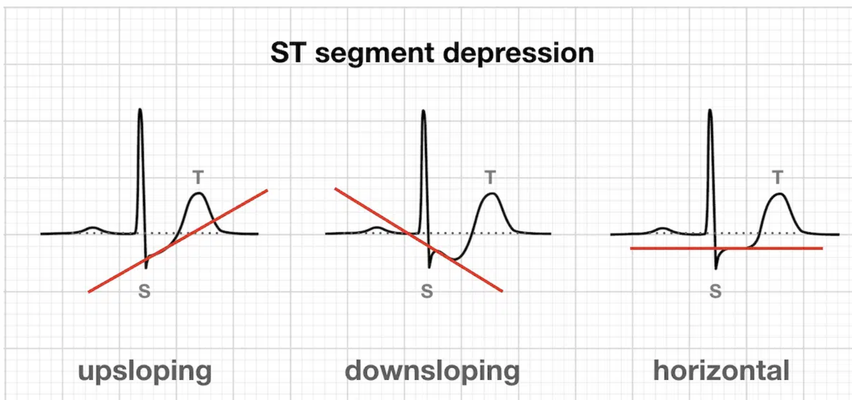
Results

The first question that we wanted to answer was what specific factors in our dataset contributed the most to heart disease. The model we used showed that the number of major vessels colored by fluoroscopy, the ST depression induced by exercise relative to rest, and the non-anginal and typical anginal chest pain types, were the most important factors contributing to heart disease. Each of these factors relate to the health of the heart in some shape or form, which is reassuring to see. However, to understand why these results make sense, we need to define terms and concepts within these factors.

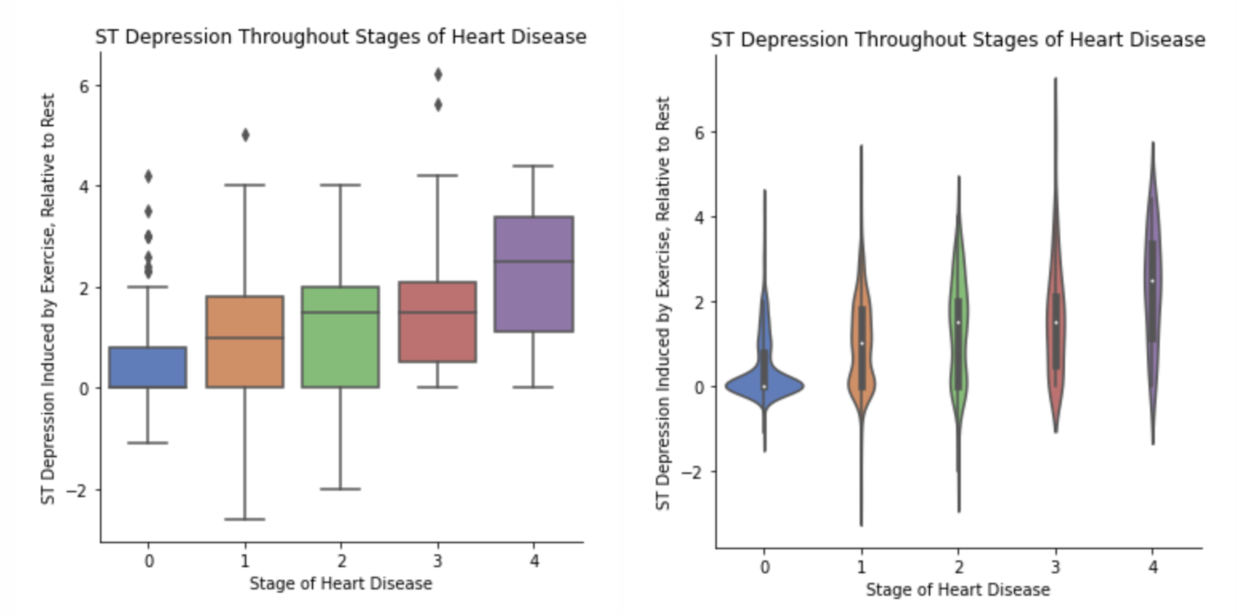
Fluoroscopy is a procedure that can look at the movement of the inside parts of the body using x-rays over a period of time (Johns Hopkins Medicine). They are used for a whole range of medical procedures, but most importantly, they can be used for cardiac catheterization, where doctors will view the blood flow through coronary arteries to evaluate the presence of arterial blockage (Johns Hopkins Medicine). As such, in our dataset, one of the factors most significantly influencing the development of heart disease is the number of major vessels observed using this technique. Although the dataset does not specify the major vessels observed, any damage where there is a blockage and allows us to see the vessel under fluoroscopy, could have detrimental effects to the normal functioning of the heart. From the data visualization below, more vessels observed are correlated with increased progression of heart disease.



Next, another significant factor was the ST depression induced by exercise relative to rest, which is usually measured in millimeters (mm) on the reading. This column of our dataset comes from electrocardiogram (ECG) data where the ST segment is a specific region seen when the heart contracts to expel blood from the ventricles (Burns). The ST segment relevant is the segment after the highest peak has descended and is starting to get back to the baseline. The attached visualization below was not coded in Python as it simply is a good means of understanding what our data is telling us.

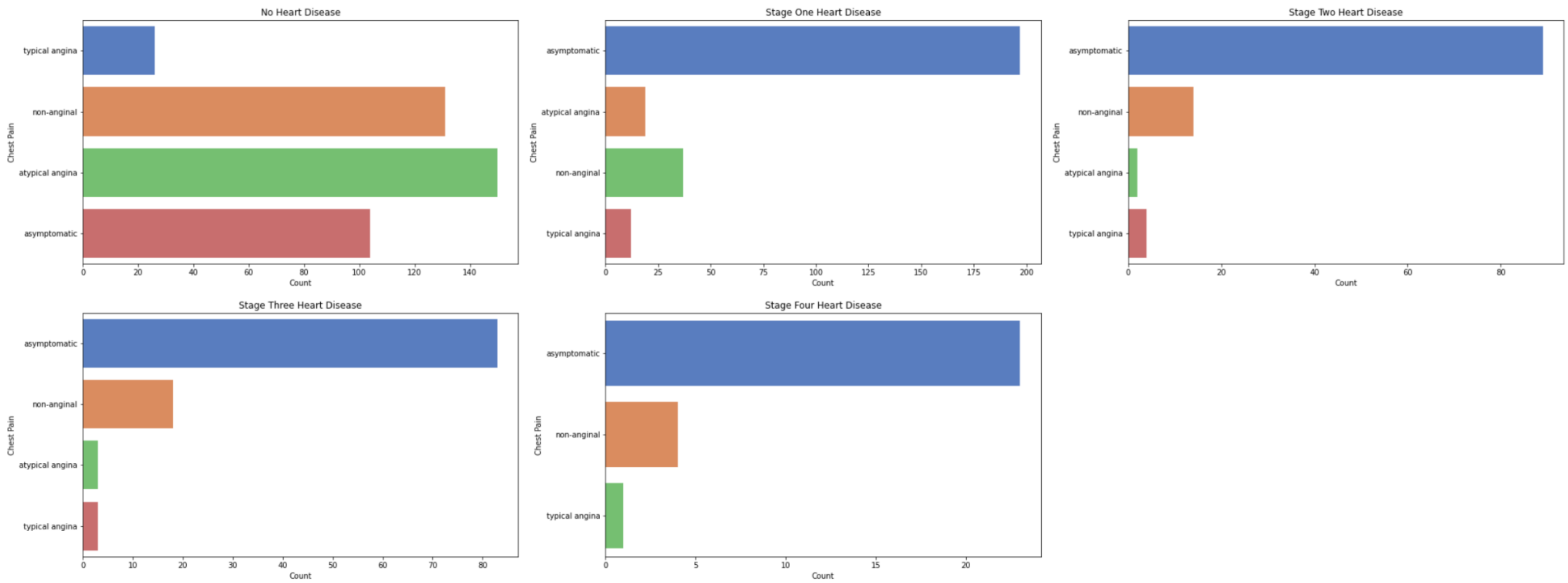


In our data, the values in this column show the difference between the ST segment between rest and when exercising, where positive values show the magnitude of the depression, while the negative values show the opposite. From the data visualization below, we can see that as the stage of heart disease progresses, the average ST depression difference gets larger. The graph on the left allows us to see the outliers in our data while the graph on the left allows us to see the distribution of points around a specific ST depression difference value. Exercise stress tests consider that a difference larger than 1 mm is highly indicative of coronary artery disease (Lim). This validates our data visualizations as there are a lot of data points for individuals with no heart disease at the value of zero for their ST segment depression difference, while the difference appears to increase as we progress through the stages of heart disease.

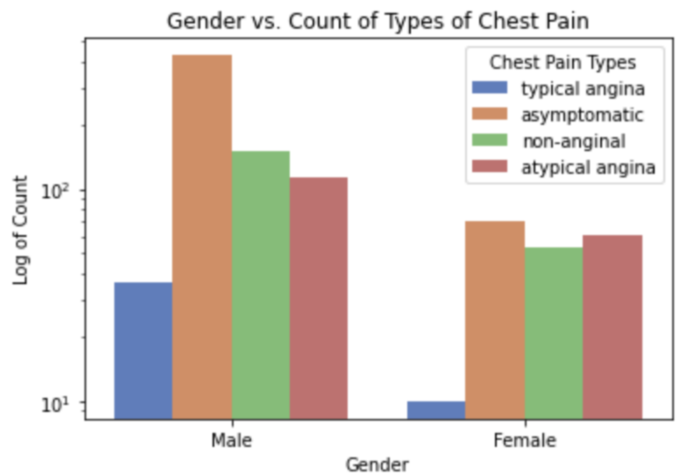


Finally, we also identified chest pain to be an influential factor for the development of heart disease. Chest pain is also referred to as angina and happens when there is discomfort due to the heart muscles failing to get enough oxygen-rich blood (American Heart Association). Medically, angina is known to be a symptom of other underlying heart problems, like coronary heart disease. To a patient experiencing angina, they can identify this when they feel a pressure or squeezing in their chest accompanied by discomfort around the (American Heart Association). There are numerous different types of angina with variation in the length, frequency, cause, and treatments (National Heart, Lung, and Blood Institute), and our dataset separates them by typical angina, atypical angina, non-anginal, and asymptomatic. Typical angina is characterized by "substernal chest discomfort of characteristic quality and duration", "provoked by exertion or emotional stress", and "relieved by rest and/or nitrates within minutes" (Gore). On the other hand, atypical angina is when chest pain meets only two of those three criteria, and non-anginal chest pain is when the chest pain only meets one or less of the criteria (Gore).

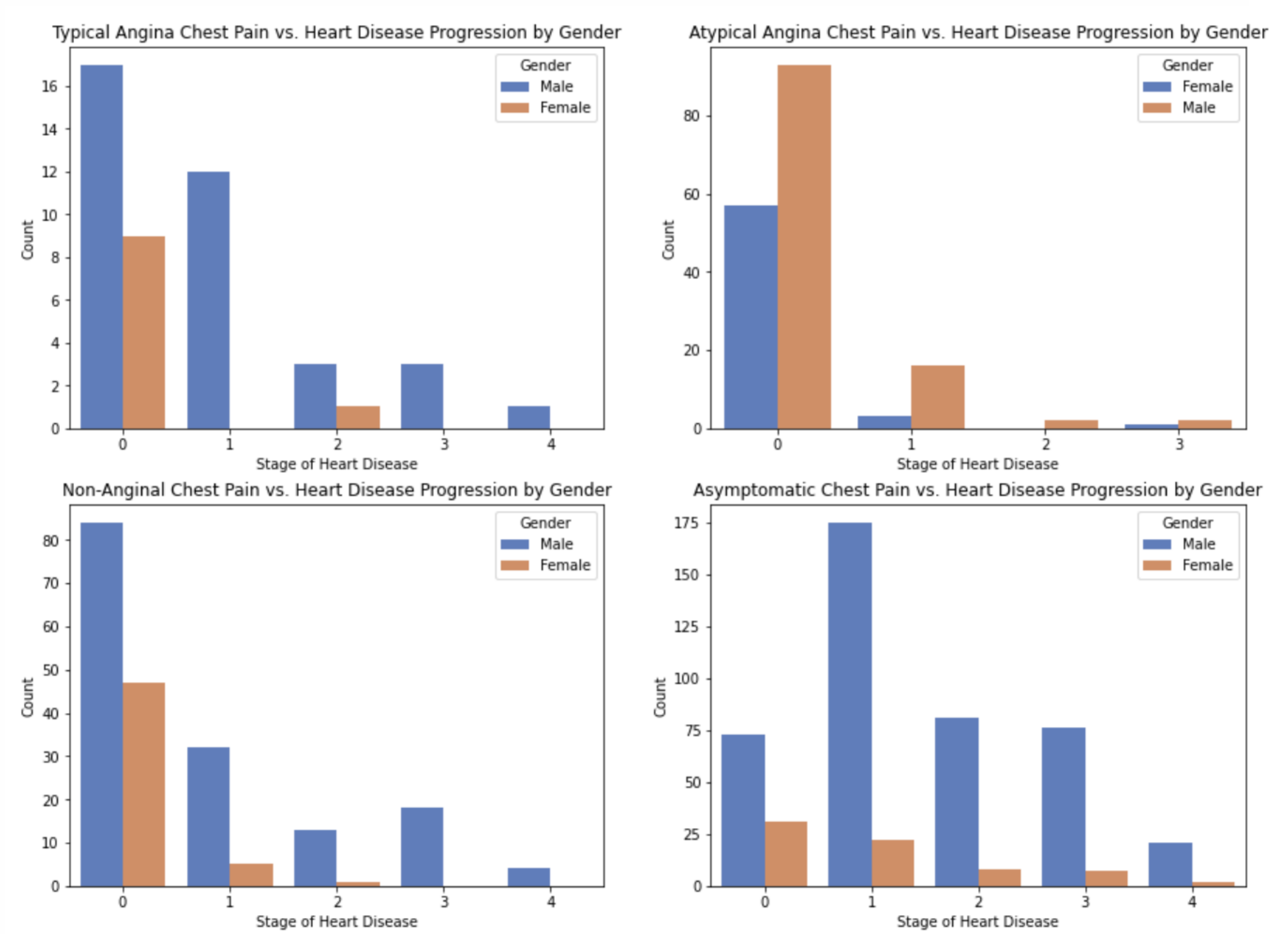
However, in regards to this, there has been some debate about the term 'typical' versus 'atypical' angina as clinicians seem to use this to classify chest pain differently between men and women. In the past, it would seem that typical angina was used to describe how men experienced angina whereas atypical angina was used to describe how women experienced it (Ham). Interestingly enough, MIT researchers used a machine learning model to classify symptoms experienced by both men and women and found that they both experience angina in the same way (Ham). Despite this, our dataset still focuses on this idea, so we will continue to use it here to analyze our results. With the visualization below, we can see that asymptomatic angina is the most common in all stages of heart disease, whereas atypical angina is the most common chest pain for people who did not develop heart disease.



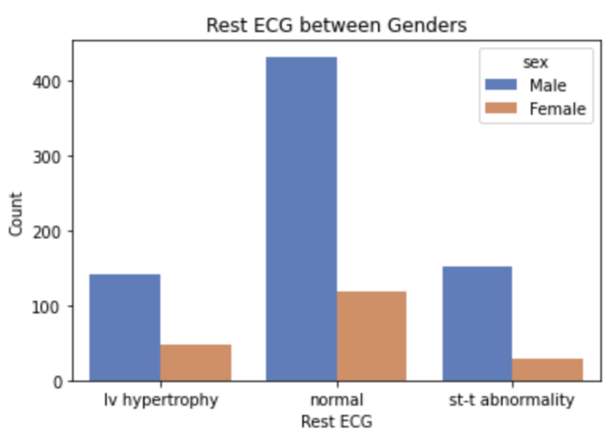
As we can see with the data visualization below, it shows count in a log scale as there were many more males than females in the dataset. This was one limitation that we discussed later on in our report as well. Despite this, we can see that the distribution of chest pain types is relatively similar for both genders, except females have more atypical anginal pain than non-anginal as compared it to males.

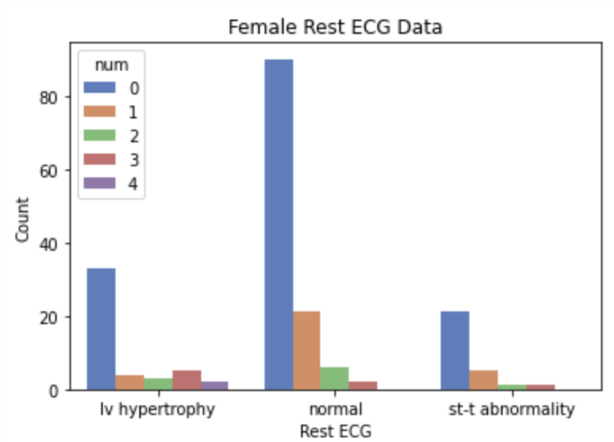


Considering the difference in chest pain between genders, one of our other research questions looked at the most influential factors between gender and age in general. The similarity between genders and our general dataset was the number of major vessels and non-anginal chest pain as an influential factor. The main difference between gender was that females also had their resting ECG reading with an ST-T abnormality as an influential factor whereas males also had typical anginal chest pains and an ST depression difference as influential factors. We can see with the data visualization below that the reason that typical anginal chest pains are not influential for females is because there is simply not enough data. All around for chest pain, there seems to be less data for females, so typical anginal chest pain may be influential if there had been more data to analyze.

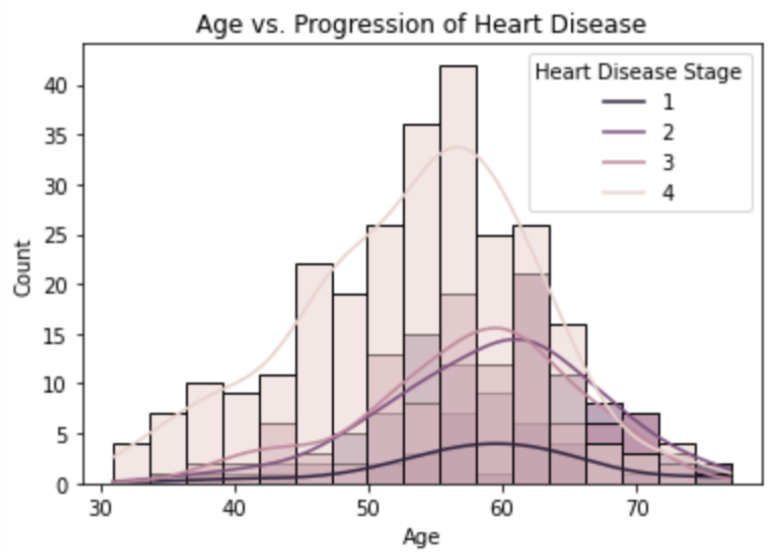


Another one of the differences between the genders is with females having their resting ECG, specifically with the ST segment T abnormality as an influential factor, whereas males did not. Looking at the data visualizations below, we can see that here is a large difference in the amount of resting ECG data between males and females. This could be a factor that plays into whether or not certain factors were more influential than others because rest ECG may have been influential to predicting heart disease in females since there were so few data points on them. These data points may have been skewed to having heart disease, when in actuality, this might not be a good predictor of heart disease in general. Furthermore, looking at the resting ECG data at females specifically on the right, we can see that there are very few data points of the ST-T abnormality, which was the influential factor. As such, we would conclude that this may be an influential factor of predicting heart disease in our dataset, but this might not be applicable to general heart disease factor data.

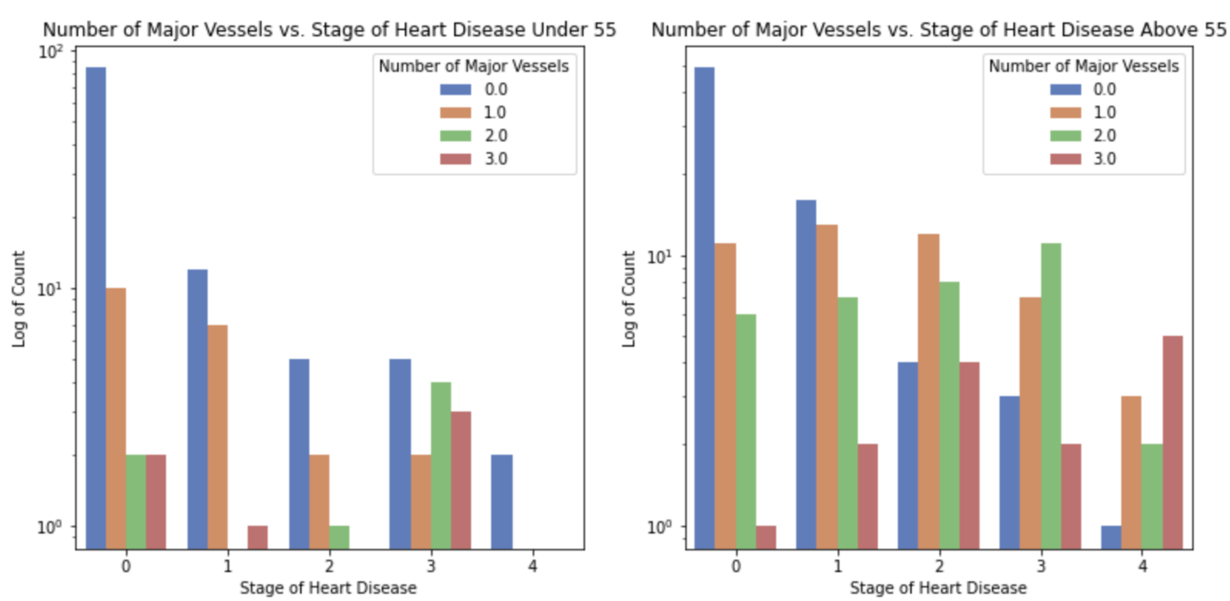




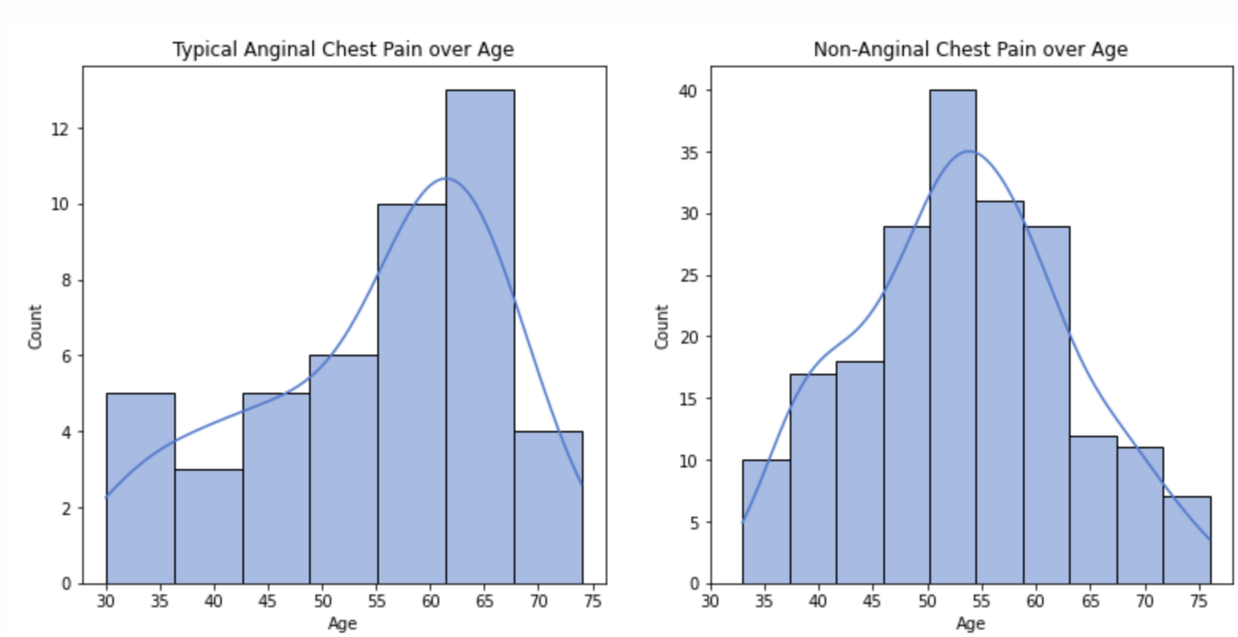
Moving onto age, we separated the groups for under 55 and over 55 due to the median age of the dataset being 54 and then rounded to the nearest nice number. Again, for the group under 55, we observed that the most influential factors were the number of major vessels, the ST depression difference between exercise and rest, and the typical anginal and non-anginal chest pain. On the other hand, we found that those over 55 only had typical anginal chest pain and the number of major vessels as the only two most influential factors for developing heart disease. As we can see with the data visualization below, there seems to be a later onset for later stages of heart disease. This is what we had expected since we know that older people are more at risk for heart disease and may also have other comorbidities that make their development of a higher stage of heart disease more likely (Rodgers). Calculating the means of the ages in the progression of heart disease we find that the average age for stage one is 53.5, for stage two is 57.6, for stage three is 59.2, and for stage four is 59.2.



When looking at the number of major vessels since that was one of the most influential factors on heart disease development, we can see that for those above 55, it is more likely for them to have a higher number of observed major vessels. For example, for heart disease stage four, under 55 individuals do not have any major vessels observed while almost all of the individuals above 55 have more than one major vessel observed. As such, it is reaffirmed that the number of major vessels observed could be a good indicator for heart disease.

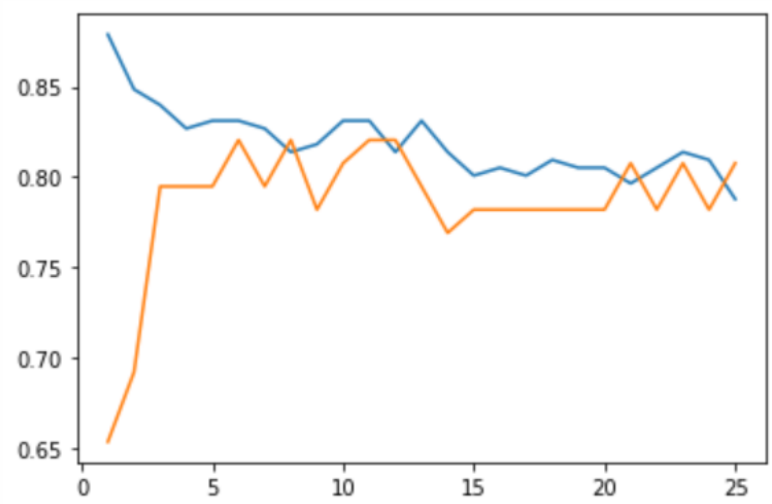


Another factor to look at between ages is the chest pain types since they did not have the same chest pain types that were influential to the progression of heart disease. With the two data visualizations below, we can see that typical angina seems to have a trend of later onset of roughly around 60 years of age whereas with non-anginal chest pain, we can see that the trend for onset is closer to 55 years old. This would help explain why perhaps non-anginal appeared as a more influential data point for under 55 while it was not for above 55.



Before moving onto the machine learning algorithms that we utilized, our last question of exploration regarding influential factors was surrounding how location may change the most influential factors in the data. When looking at the date overall, we find that location is not one an influential factor in determining heart disease progression. The dataset only includes location data from Cleveland, Long Beach, Hungary, and Switzerland. As such, we cannot fully rule out location as a determinant of heart disease, but we can only work within the confines of our dataset. Breaking up the data into the US and Europe, we were able to find that the number of major vessels, ST depression difference between exercise and rest, and typical anginal and non-anginal chest pains were also the most influential factors. This works to help to reaffirm our most influential factor results from the general analysis in question one.

In our machine learning analysis, we discovered that creating a model only using our most influential variables (`ca`, `cp`, and `oldpeak`) we were able to create a much more accurate and reliable model than when considering all variables. We conducted our analysis by working with the KNeighborsClassifier model from the sklearn library. We determined that the ideal number of neighbors for analysis was 5. We found this number by first filtering down our data to these 3 variables, removing NA values, and then running our model through a for loop to find the number of neighbors with the highest accuracy. We then computed a confusion matrix with that ideal number and gathered information on the model through analysis scores on the matrix. This model with 5 nearest neighbors has an accuracy score of 84%, and a recall score of 81%. The recall score is very important in our analysis because it considers false-negatives. False-negatives are most dangerous when working with heart disease data, because that would mean the model reported the patient had no heart disease, when they in-fact did.



Here is the plot I created to find the ideal number of neighbors in our machine learning model. The x-axis represents the number of neighbors. The y-axis is the accuracy score of each model. Lastly, the blue line is our training data and the orange line is our testing data. When finding the best model we wanted the number of neighbors where the orange line is highest.

Impact and Limitations

There are limitations with the dataset that we have chosen. Firstly, there were a limited number of locations that were included in the dataset. For example, although we sectioned off our data into information in the US or Europe, we were unable to arrive at analyzable data for Europe due to the limited number of data present. Furthermore, these areas do not reflect a majority of the world since they were taken from just four main areas. If we wanted this machine learning model to accurately predict heart disease development from all over the world, location is likely going to be a big factor. If a location has higher rates of smoking for example, even if the patients themselves were not smokers, secondhand smoke could also increase the risk of heart disease. As such, in terms of location, only those that live in a similar environment, with similar socioeconomic levels, access to healthcare, pollution levels, etc. would be the ones who would be able to use our algorithm.

Another limitation of our dataset was the number of values present for females versus males. Compared to the males in our dataset, the females were not as equally represented and this could have affected the ways in which our machine learning algorithm can accurately predict the progression and development of heart disease of females. Though males would be able to benefit from this algorithm, females would be excluded and may make certain healthcare decisions that could be detrimental to their health because they are following suggestions based off of a largely male dataset. This bias against women is already present in healthcare, with many of the procedures and advice provided by doctors for women to be inaccurate to their biology because they were studied from men. As such, if doctors do begin using algorithms like this where gender is not taken into account, this could be quite harmful to women.

Additionally, one of our most influential variables 'ca' which exhibits the number of major vessels each patient has, had the most NA values in the dataset. Because of this, for much of our machine learning analysis we were working with a limited number of patients (about 200-300) compared to the approximately 900 patients in the data. Without knowing why these values were excluded, we cannot be sure of the accuracy in our influential variables.

Lastly, the `num` value that we use in our analysis to work with the models, find the most significant value, and create our new model, is already a predicted value from the research conducted in this study (detailed in our dataset limitations). Because of this, our results are purely based on this research and do not factor in the actual outcome of these patients aside from what is provided in the research. Because of this, the implications of our study are building off of this research and providing more analysis on their predictions. This does limit the scope of our impact, but it does not minimize it. With our models, this research can be built off of and more research into our most influential variables can be conducted.

Challenge goals

Machine Learning

We will be sectioning our current dataset into a testing and training dataset that is characteristic of machine learning. Both the testing and training dataset will consist of a column that outlines whether or not the patient was predicted by medical professionals to develop heart disease and what specific stage the disease may have progressed to. We believe that our project and research questions will address this challenge goal since we will be using machine learning to predict if a patient would be diagnosed with heart disease given the other attributes, in conjunction with the heart disease progression.

Our analysis focused on the KNeighborsClassifier, which will use input from neighbors around our point to determine whether the particular patient the point represents will have heart disease based on the variables we input. We expanded on this goal by not only creating one KNeighbors model, but creating several to find the most ideal model for our data. We used our knowledge of for loops to create hundreds of models testing the most effective number of neighbors. We also created several models by changing variables and removing variables that were not determined to be statistically significant.

We also demonstrated an excellent understanding of the material by going further and analyzing our models using a confusion matrix and accompanying scores of accuracy, recall, precision, and f1. This allowed us to better comprehend our results and produce the most effective model for predicting heart disease in patients based on this data.

New Library

A new library we will utilize is statsmodels and will help us ensure that our results are statistically significant. This will be done through utilizing the library's functions to find correlation coefficients, which will determine the relationship of attributes and whether or not it serves as a good predictor of developing heart disease. This library will be important for our machine learning challenge goal too as it will help us determine the attribute that we can focus on to predict heart disease progression.

We stayed consistent with this goal by learning about the statsmodel library and utilizing its OLS (ordinary least squares) model and analyzing this information through the summary table. However, we expanded on this goal by learning more about the statistical values presented in the table and how to analyze them to learn more about the results of our data. This statistical knowledge we developed helped us determine which of our variables was most statistically significant which led to our more accurate model in the machine learning aspect of our project.

Work Plan Evaluation

1. First, we will determine the significant attribute that may contribute the most to the development of heart disease using statistical analysis. We will also use this type of analysis to identify the attributes associated with each stage of heart disease progression.
    a. Estimated time: 3 hours
2. Second, we will use the machine learning regression model to predict whether or not a patient will develop heart disease using the most significant attributes from the previous task. We will also be checking the max depth of the tree to ensure the highest accuracy of the machine learning regression model.
    a. Estimated time: 5 hours
3. Lastly, we will perform more statistical tests to look at how location, gender, and age could affect the prediction of someone developing heart disease.
    a. Estimated time: 2 hours

Each person will be working on their designated task in terms of developing and testing the code. When a person is assigned to the task, they will work to learn any additional information needed to ensure that their code can answer the research questions they have been assigned to. We have divided the tasks up after deciding the estimated time for each task that also includes the amount of time a person would need to take to learn the appropriate code and how to best use the library or the regression model. One person will do tasks 1 and 3 that are answering questions 1, 2, 4, and 5, and the other person will do task 2, which answers question 3. For coordinating work, we will be frequently communicating and checking in to see where each person is in their task. In case one team member needs support on one task that is unexpectedly challenging, we will meet up in person to talk through the code and also review what part of the task is particularly more challenging. We will work together to identify other potential ways to finish the task first. If either one of us are unable to come up with a solution, we can both ask the professor and/or the TAs for any advice they might have with solving the task.

When starting our analysis on the most significant factors in the data, we were moving effectively on track to finish the question within our given 3 hour timeframe. We dedicated 1.5 hours to work together to understand the output results of the statsmodels library, and how to compute the multiple regression OLS model and the summary table. Then one member started to incorporate that model in with our data, and ultimately interpret the results and discover the most significant values within our data.

However, when it came time to filter our data and work with our statistical analysis based on location, gender, and age, we came across several barriers. We discovered that the statsmodel OLS needs a significant number of observations, and when our data was filtered in certain ways the number was not met. Through our analysis we came across many errors when computing this, and through several hours of research we were unable to determine how to fix this problem for our location question. This added approximately 2 hours to our analysis of the data based on filtration.

Lastly, we redesigned our plan for the machine learning aspect of our project and focused on learning about the KNeighbors model within sklearn. Due to this change, we added an hour to our work time to learn more about the model, and how to work with it properly considering our data. Yet, when we put it into practice we were able to complete the models in 4 hours rather than our predicted 5, so we ultimately stuck to our 5 hour timeframe for our machine learning goal.

Overall, we did not consider time for if a very serious problem arises during our analysis. As a group, we were able to better budget our time and still complete the project within the timeframe of the class due to our starting of the project a couple weeks beforehand.

Testing

For our models, testing was conducted through numerous runs of models and varying divisions among testing and training data. When working with our new library and discovering our most statistically significant values, we tested our results by working with different versions of our data. Not only did we test all of the data, but we also analyzed how our results changed based on location, gender, and age. With each of these analyzes we produced very similar results. In consideration to gender, our results differed slightly but we believe that to be in consideration to the disproportionate ratio of men to women (which is detailed further in our limitations). Through our education on the statsmodels library we also gained knowledge on analyzing the results of the summary table, which we explain within the coding analysis in the Jupyter notebooks. This provides further insight to our audience, which can give them more confidence in our results. Due to the reproduction of our results across a variety of filtered data and our thorough analysis, we feel confident in our variables' significance and accuracy.

When working with our KNeighbors machine learning model, we ensured to split our data into testing and training datasets. We also ran our model several times within our Jupyter notebook to ensure we were getting similar results, and the model was drastically varying with each run. Additionally, we computed several confusion matrices on these models. This analysis further looked into the results of the model and provided more insight to how they work, which gives our audience more confidence in our results. We are confident in our results based on these facts.

Collaboration

American Heart Association. (2022, December 5). Angina (chest pain). www.heart.org. Retrieved March 13, 2023, from https://www.heart.org/en/health-topics/heart-attack/angina-chest-pain

Burns, E., & Buttner, R. (2022, March 16). The ST Segment. Life in the Fast Lane • LITFL. Retrieved March 13, 2023, from https://litfl.com/st-segment-ecg-library/

Gore, J. M. (2010, July 7). Typical Angina vs. Atypical Chest Pain. NEJM Journal Watch: Summaries of and commentary on original medical and scientific articles from key medical journals. Retrieved March 13, 2023, from https://www.jwatch.org/jc201007070000002/2010/07/07/typical-angina-vs-atypical-chest-pain

Ham, B. (2019, November 6). Machine learning shows no difference in angina symptoms between men and women. MIT News | Massachusetts Institute of Technology. Retrieved March 13, 2023, from https://news.mit.edu/2019/machine-learning-shows-no-difference-angina-symptoms-between-men-and-women-1106

Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (2021). UCI Heart Disease Data. Retrieved 2023, from https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data.

Johns Hopkins Medicine. (2019, August 14). Fluoroscopy Procedure. Johns Hopkins Medicine. Retrieved March 13, 2023, from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/fluoroscopy-procedure

Lim, Y. C., Teo, S.-G., & Poh, K.-K. (2016, July). ST-segment changes with Exercise Stress. Singapore medical journal. Retrieved March 13, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958709/

National Heart, Lung, and Blood Institute. (2022, May 24). Angina (Chest Pain) Types. National Heart Lung and Blood Institute. Retrieved March 13, 2023, from https://www.nhlbi.nih.gov/health/angina/types#:~:text=The%20types%20of%20angina%20are,unstable%2C%20microvascular%2C%20and%20variant.

Rodgers, J. L., Jones, J., Bolleddu, S. I., Vanthenapalli, S., Rodgers, L. E., Shah, K., Karia, K., & Panguluri, S. K. (2019, April 27). Cardiovascular risks associated with gender and aging. Journal of cardiovascular development and disease. Retrieved March 13, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616540/

Xu, J., Murphy, S. L., Kochanek, K. D., & Arias, E. (2022, December 8). Mortality in the United States, 2021. Centers for Disease Control and Prevention. Retrieved February 14, 2023, from https://www.cdc.gov/nchs/products/databriefs/db456.htm