



# Module 3 – Heart Disease

Andy Peng

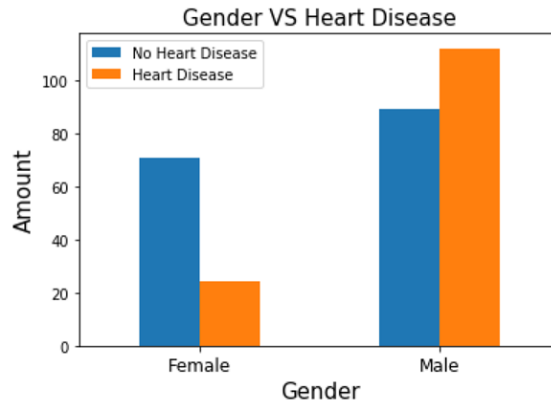
Hi my name is Andy Peng. Welcome to my presentation for Flatiron School Module 3 – Heart Disease. For the extent of this project, our stakeholders are a company that focuses on predicting whether a patient has heart disease given patient's medical records.

# Overview

- Relationship to heart disease
  - Gender
  - Thallium Stress Result
  - Age
- Modeling

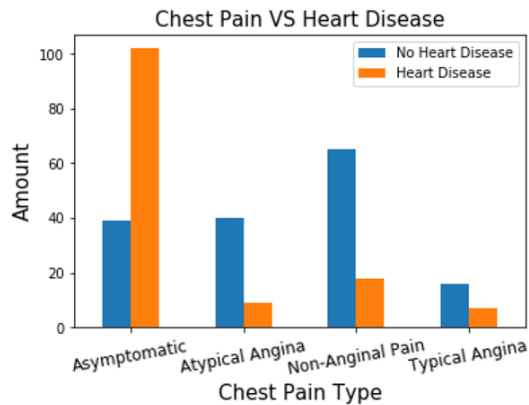
We managed to create a model that can predict heart diseases with great accuracy while also minimizing the chances of classifying a patient as false negative. Before we talk about the results of our model, we will be discussing about a few features and its relationship to heart disease.

# Does Gender affect Heart Diseases?



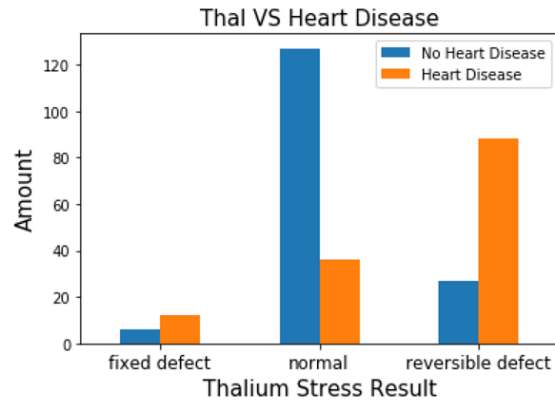
Does gender affect heart disease? Here we have a graph showing the amount of patients with and without heart diseases separated by the gender of the individual. As you can see male patients tend to have more heart diseases than female patients. This led me to question if there is any other feature that also appears in individuals with just heart disease.

# Chest Pain



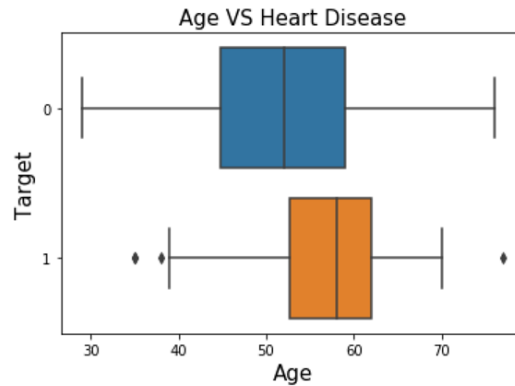
Further investigation shows that chest pain type is a feature that stands out the most. As you can see in this graph, majority of the patients with heart disease have asymptomatic chest pain type. 100 patients that have heart disease have asymptomatic chest pain compare to the second most 18 patients with non-anginal pain.

# Thallium Stress Results



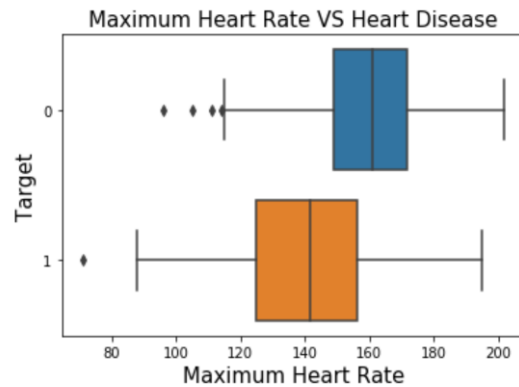
Thallium Stress Results. Do they affect heart disease? Patients with Thallium Stress Results as reversible defect tend to have a higher chance to have heart disease compared to patients with normal or fixed defect.

# Does age affect Heart Diseases?



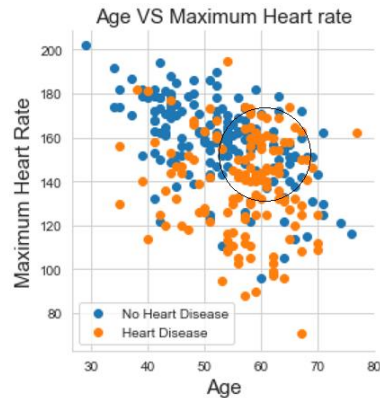
Does age affect Heart Diseases? Based off the boxplot here we can see that individuals with heart disease overall tend to be older than individuals without heart disease. But what other feature are affected when one gets older? Cholesterol, Maximum Heart Rate and Blood Sugar Levels.

# Maximum Heart Rate



There are many other features that are affected by old age, but we will be focusing on maximum heart rate. As an individual gets older, their heart won't be able to pump at the same maximum heart rate as a 20-year-old. By graphing a box plot of maximum heart rate of patients with and without heart disease we can see that patients with heart disease tend to have a lower maximum heart rate.

# Maximum Heart Rate



Now let's look at the scatter plot of age versus maximum heart rate. We can see there is a negative correlation between these two features. As you get older you have a lower maximum heart rate. But also notice that majority of patients with heart disease are older and have a lower maximum heart rate.

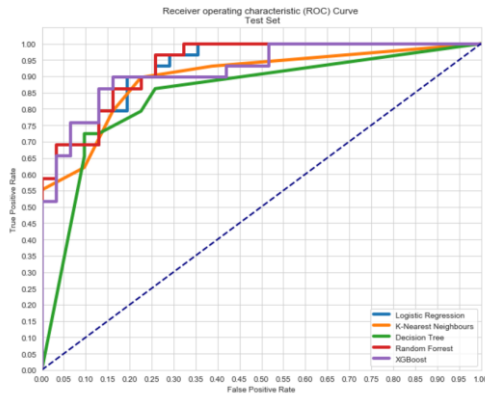


# Modeling

- Logistic Regression
- K Nearest Neighbors
- Decision Tree
- Random Forest
- XGBoost
- GridSearchCV

For modeling we used a combination of Logistic Regression, K Nearest Neighbors, Decision Tree, Random Forest, XGBoost and GridSearchCV to get the best model for our problem.

# Results



- AUC - Random Forest with GridSearchCV
- Accuracy/ F1 Score/ Recall - XGBoost
- Precision - Decision Tree with GridSearchCV

And here are the results. Here is a ROC curve displaying the AUC results for the different models we did. For best AUC values we used Random Forest with GridSearchCV. For Accuracy/F1 Score/Recall we used XGBoost. Finally for Precision we used Decision Tree with GridSearchCV.

# Recommendations

- XGBoost for Recall Score

## Base Model

	Model	Test Accuracy	Test F1	Test Precision	Test Recall	AUC Score
0	Logistic Regression	0.83	0.82	0.85	0.79	0.931034
1	K-Nearest Neighbours	0.80	0.79	0.81	0.76	0.894883
2	Decision Tree	0.75	0.74	0.75	0.72	0.749166
3	Random Forest	0.80	0.78	0.84	0.72	0.927697
4	XGBoost	0.87	0.86	0.86	0.86	0.924360

## Model Tuning

	Model	Test Accuracy	Test F1	Test Precision	Test Recall	AUC Score
0	Logistic Regression	0.83	0.82	0.85	0.79	0.931034
1	K-Nearest Neighbours	0.82	0.81	0.82	0.79	0.895996
2	Decision Tree	0.80	0.77	0.87	0.69	0.845940
3	Random Forest	0.78	0.75	0.83	0.69	0.935484
4	XGBoost	0.85	0.84	0.86	0.83	0.919911

Since we are working to classify patients on whether they have heart disease or not. Our goal would be to minimize the chances of us telling a patient they don't have heart disease when they indeed do have heart disease. In other words we want to maximize our recall score, which is exactly what we obtained by using XGBoost.

## Recommendations

- Gender
- Asymptomatic Chest Pain
- Reversible Defect
- Age & Maximum Heart Rate

Here are some features we should focus on when identifying whether a patient has heart disease or not. Gender – Males tend to have a higher chance at having heart disease than Females. Asymptomatic Chest Pain – When patients have this chest pain type, they tend to have a higher chance at having heart disease than other individuals. Reversible Defect – From the Thallium Stress Results, more patients with this result have heart disease than other results. Age & Maximum Heart Rate – As individuals get older and their maximum heart rate starts to decrease, they will have a higher chance of having heart disease.

## Next Steps

- Features to Explore

- Genetic Disorders
- Body Weight Percentages
- Patient's Diet

- Gather More Data

For our next step we can explore other key features such as whether the patient have genetic disorder, their body weight percentages and their diet. If we want to improve our model, gathering more data would help us do a better job at predicting whether a patient has heart disease or not.



# Thank You

Thank You