Predicting Heart Disease

Alex Zieky & Amin Nazerzadeh

Data

- Kaggle (Originated from UCI, Machine Learning Repository)
- 1025 Observations
  - Hungarian Institute of Cardiology. Budapest
  - University Hospital, Zurich, Switzerland
  - University Hospital, Basel, Switzerland
  - V.A. Medical Center, Long Beach
  - Cleveland Clinic Foundation
- Target Variable: Heart Disease
  - **Arrhythmia**: Heart rhythm abnormality
  - **Atherosclerosis**: Harding of the arteries
  - **Cardiomyopathy**: Heart muscles harden or grow weak
  - **Congenital Heart Defects**: Heart irregularities that are present at birth
  - **Coronary Artery Disease**: Caused by the buildup of plaque in the heart's arteries
  - **Heart Infections**: Heart infections may be caused by bacteria,viruses, or parasites

Features

1. **Age**
2. **Sex**
3. **Chest Pain**
    a. Typical Angina
    b. Atypical Angina
    c. Non-Anginal Pain
    d. Asymptomatic
4. **Resting Blood Pressure**
5. **Cholesterol**
6. **Fasting Blood Sugar**
    a. Test is done in the morning to determine how much glucose (sugar) is in a blood sample after an overnight fast.
7. **Resting Electrocardiographic Results**
8. **Maximum Heart Rate Achieved**
9. **Exercise Induced Angina**
10. **Oldpeak**
    a. ST Depression induced by exercise relative to rest
11. **Slope**
    a. Slope of the peak exercise ST segment
12. **Number of Major Vessels**
    a. Fluoroscopy is used to help the healthcare provider see the flow of blood through the coronary arteries to check for arterial blockages.
13. **Thalassemia**
    a. Blood Disorder

Business Questions

1. Which features should doctors look at when trying to predict Heart Disease?
2. Can we help everyday people determine if they need to start monitoring their heart?
3. Can our model confidently predict if a patient has a Heart Disease?

Metrics

- Recall
  - Goal: Limit False Negatives

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

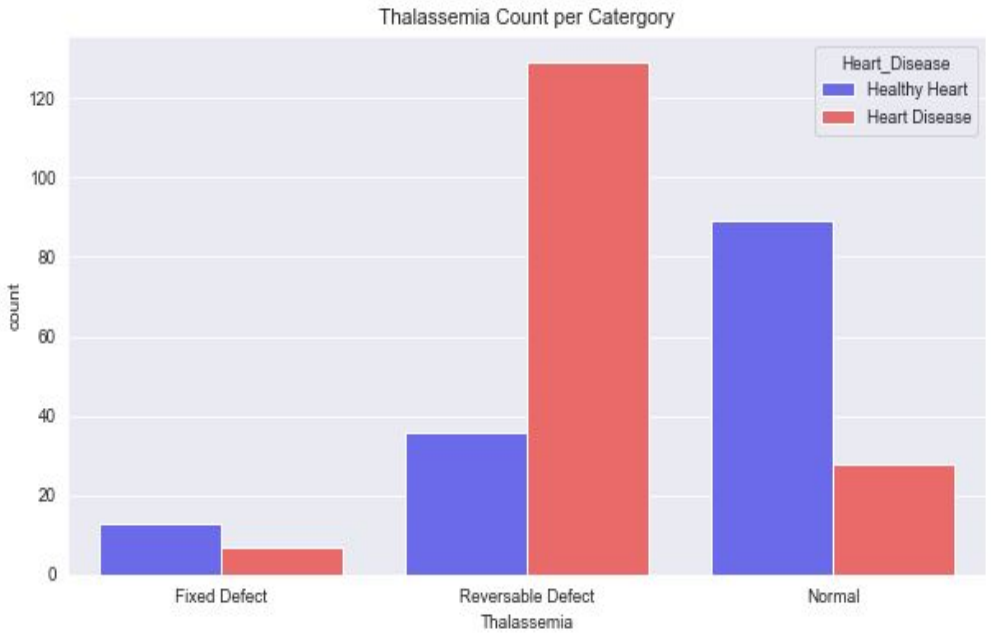
- F1 Score
  - Goal: Further evaluate our model

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Business Question 1: Most Important Features

Thalassemia

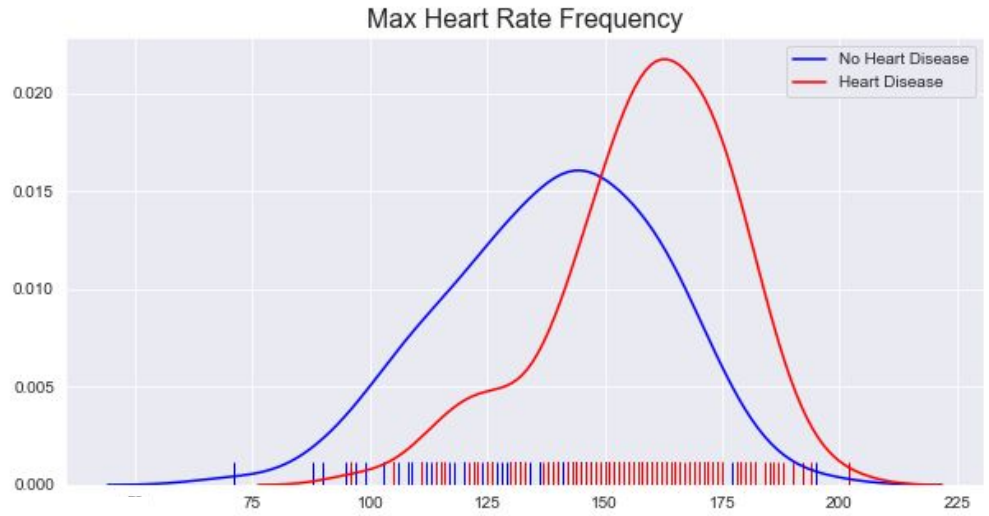
- Hereditary blood disorder in which the body makes an abnormal form or inadequate amount of Hemoglobin (the protein in red blood cells that carries oxygen)
- The disorder results in large number of red blood cells being destroyed, which leads to anemia

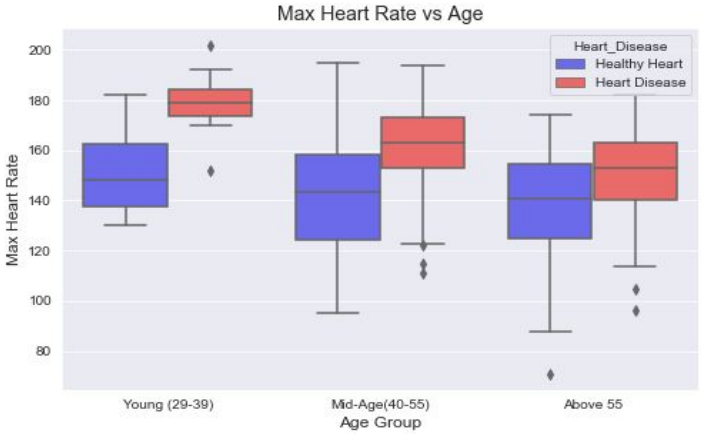


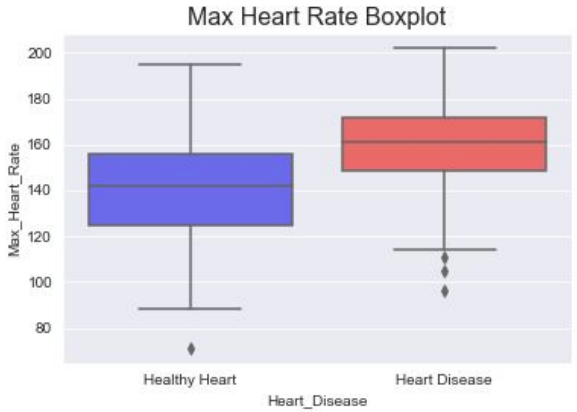
Business Question 1: Most Important Features

Max Heart Rate

- Heart rate is the speed of the heartbeat measured by the number of contractions of the heart per minute
- Healthy Patients vs Patients with Heart Disease
- Max Heart Rate vs Age



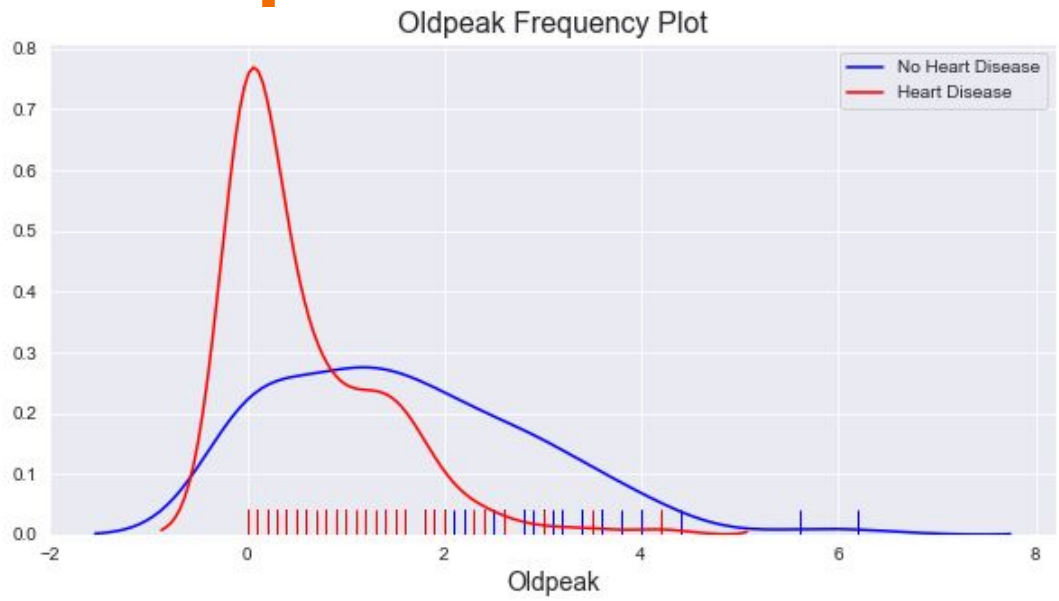


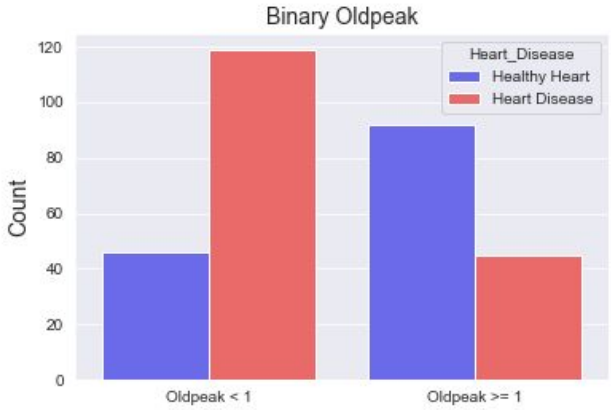


Business Question 1: Most Important Features

Oldpeak

- ST depression test refers to a finding on an electrocardiogram, wherein the trace in the ST segment is abnormally low below the baseline.
- Scoring Scale

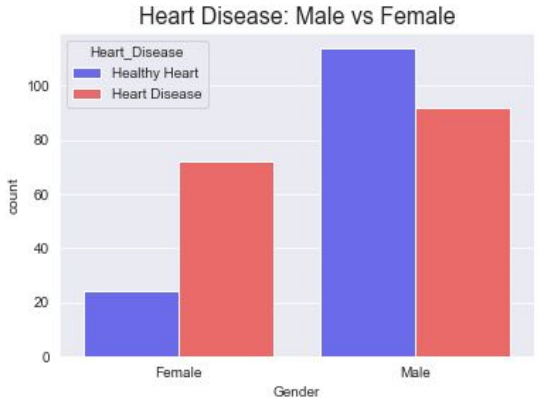


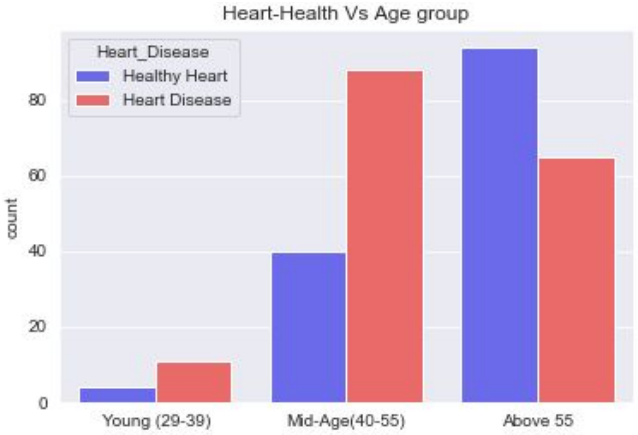


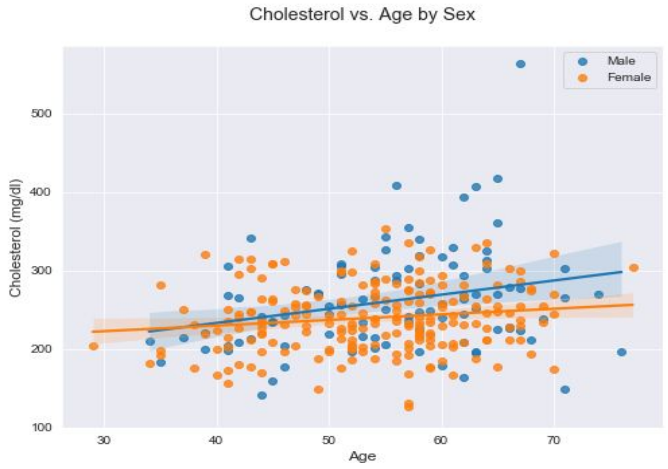
Business Question 2: Common Features

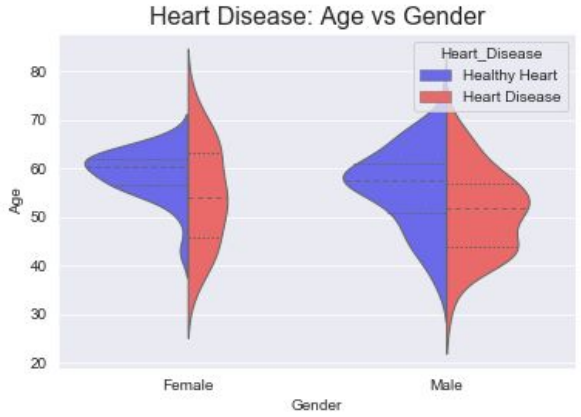
Age & Gender

- When to start monitoring your heart
- Male vs Female
- Age Groups





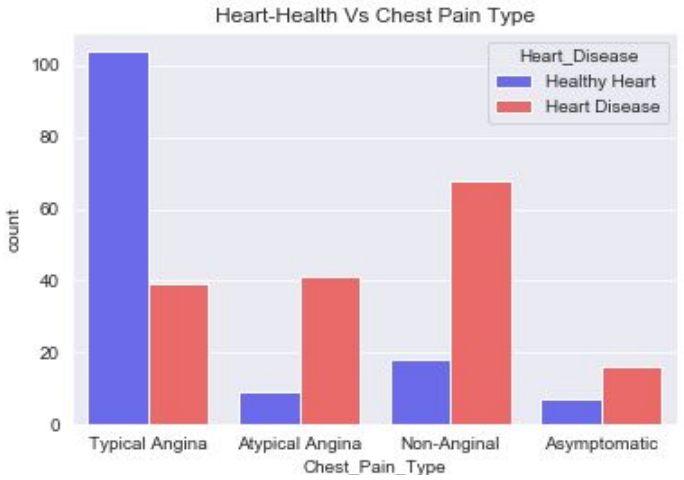


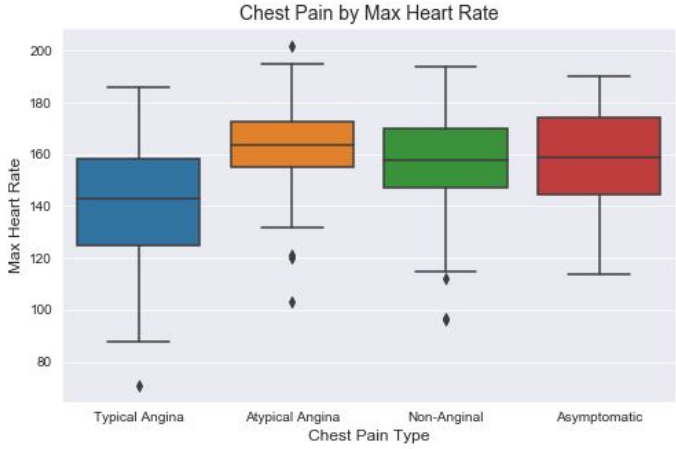


Business Question 2: Common Features

Chest Pain

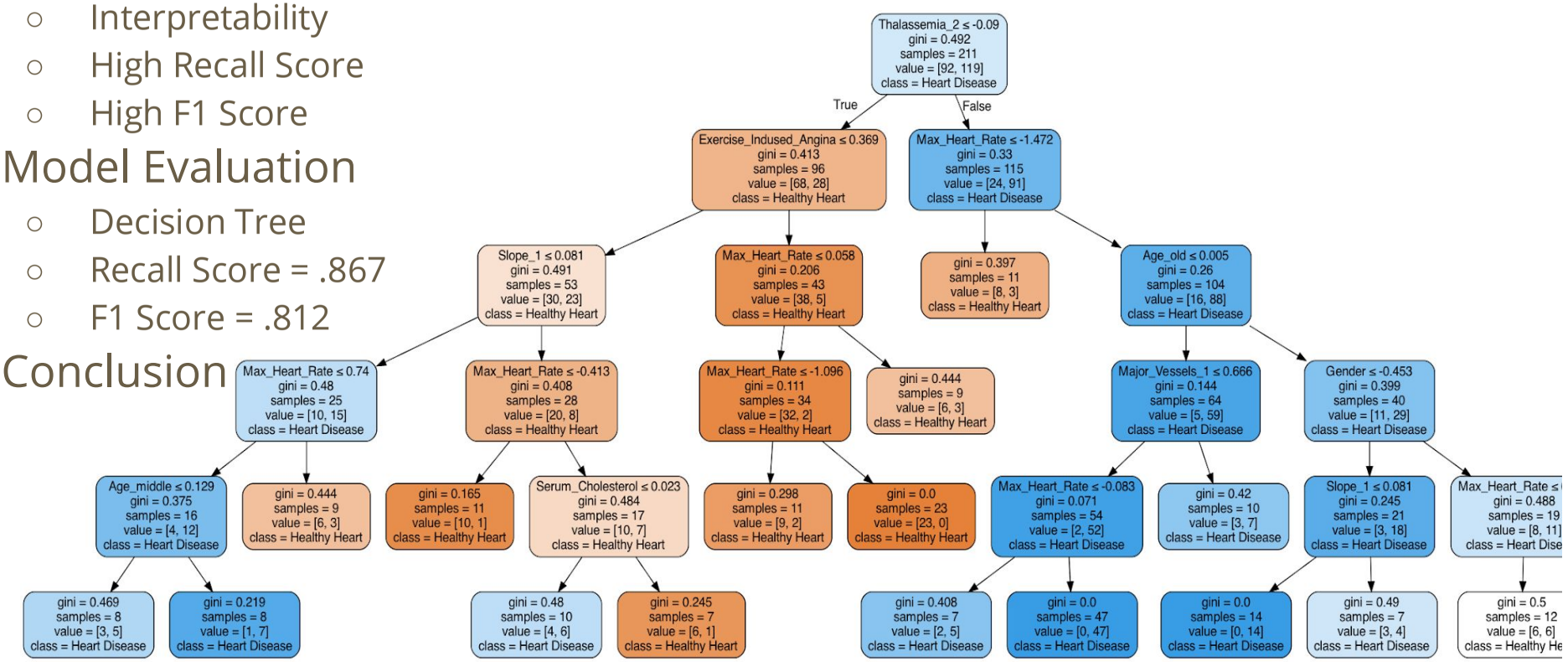
- Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood.
- Two Types
  - Typical
    - Heaviness, pressure, weight, vise-like aching, burning, tightness.
    - Relatively Predictable
    - Lasts 3-15 min
  - Atypical
    - Sharp Pain
    - Random Onset
    - Lasts seconds





Business Question 3: Predicting Heart Disease

- Model Goal
  - Interpretability
  - High Recall Score
  - High F1 Score
- Model Evaluation
  - Decision Tree
  - Recall Score = .867
  - F1 Score = .812
- Conclusion



Applications

Suggestion 1: Both male and female above 40 do regular checking for heart diseases so it can preventable at early stages.

Suggestion 2: When a patient is checking in, measure the importance features to identify any heart diseases.

Suggestion 3: Patients with Max Heart Rate above 160 should schedule regular check-ins with their doctor.

Further Analysis

To continue this project, we would suggest:

- Find similar data sets that have lot more features to work with
- Re-collecting data as more becomes available
- Automating collection and analysis methods to always have up to date information
- Consult with a healthcare expert to gain more knowledge about the dataset and features

Thank You

Any Questions?