# Machine Learning and Heart Disease

Using Biometric Data to Find Heart Disease

# Introduction to the Problem

Our target for this problem is identifying heart disease in patients based on available biomedical statistics such as their age and max heart rate. The stakeholders for this problem are in anyone who may be at risk for heart disease. Identifying heart disease early can lead to better outcomes for people suffering from the disease.

# Introduction to the Data

Each row in this data set represents one patient and their individual health data related to heart disease.

1. Age
2. Resting Blood Pressure
3. Cholesterol
4. Fasting Blood Sugar
5. Max Heart Rate
6. Oldpeak
7. Heart Disease

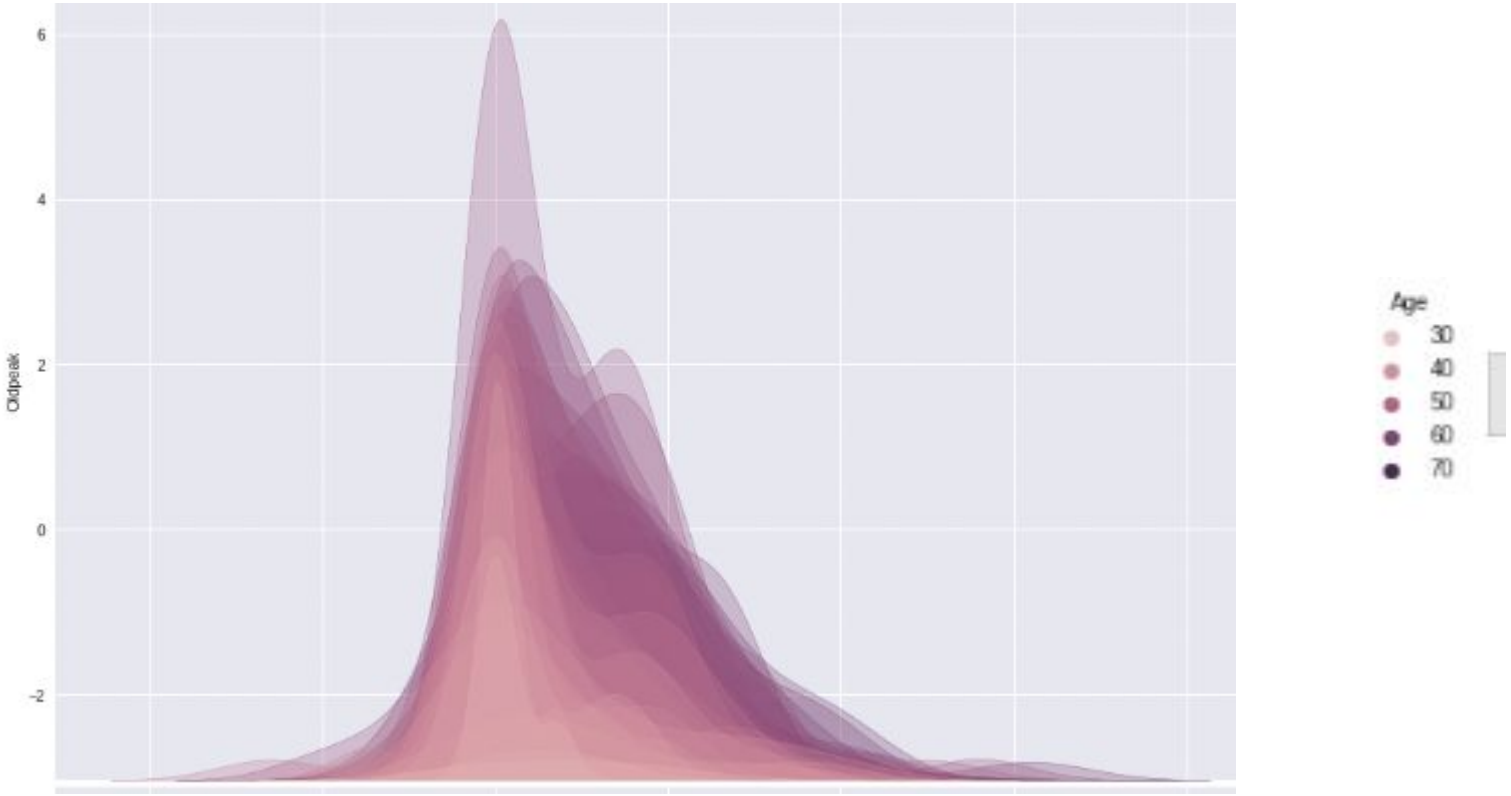
# Introduction to the Data

Each row in this data set represents one patient and their individual health data related to heart disease.

- Age
- Resting Blood Pressure
- Cholesterol
- Fasting Blood Sugar
- Max Heart Rate
- Oldpeak
- Heart Disease
- Sex
- Chest Pain Type
- Resting ECG
- Exercise Angina
- ST Slope

# Findings:

1. The highest correlation for heart disease in the features were: oldpeak and max heart rate
2. Oldpeak and max heart rate also correlated highly with age




Oldpeak paired with Age. As age increases so does oldpeak predicting heart disease.


Max heart rate pared with age: as age increases max heart rate decrease

# Logistic Regression Model Results

The most important metric for this model is recall. To reduce false negatives we must improve recall. The model have a recall score of 0.84  meaning it caught 84% of the false negatives. Reducing false negatives is very important, as a false negative will lead someone who is very ill to believe themselves healthy preventing them from seeking urgently needed medical care.

# Recommendations:

1. A recall score of 0.89 is high but more data in needed to better train the data and improve scores.
2. In addition, the data is too limited to be safely used as a training set for the model. With less than a thousand records and only 10 features it would not be safe to assume that this data can safely encompass all people. More data is needed for a more accurate model.