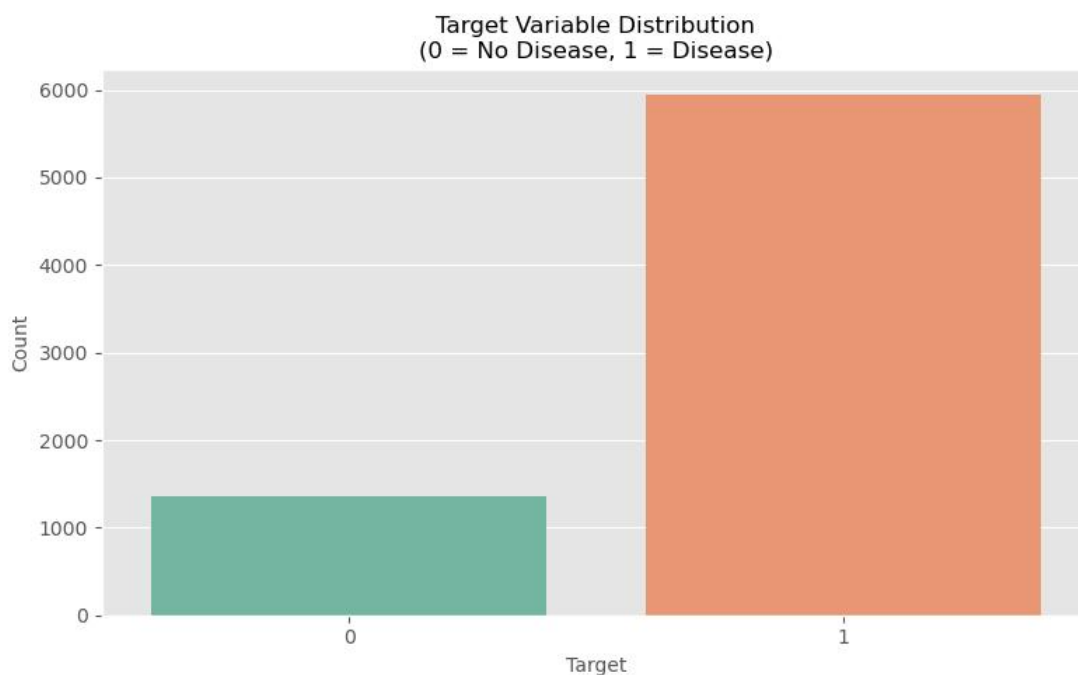


# Exploratory Data Analysis (EDA) of Heart Disease Dataset

## Introduction

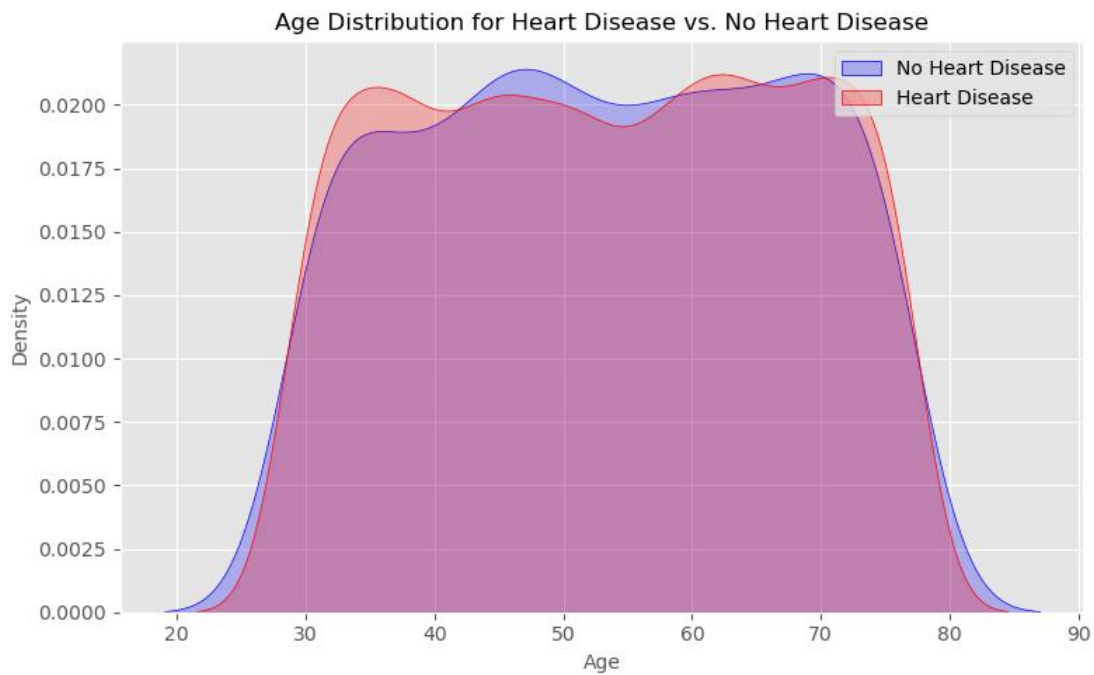
This Exploratory Data Analysis (EDA) looks into the heart disease dataset to uncover insights that could help in understanding the various factors of heart disease. The dataset consists of various features such as age, sex, chest pain type, blood pressure, cholesterol levels, etc. The relationship between these variables and the target variable—that is, the presence of heart disease—is visualised here to bring out the patterns and correlations that can inform predictive modelling and healthcare intervention.

### 1. Target Variable distribution



The count plot shows that the dataset is slightly imbalanced, with a larger number of instances of patients with heart disease, represented by target = 1, than without, represented by target = 0.

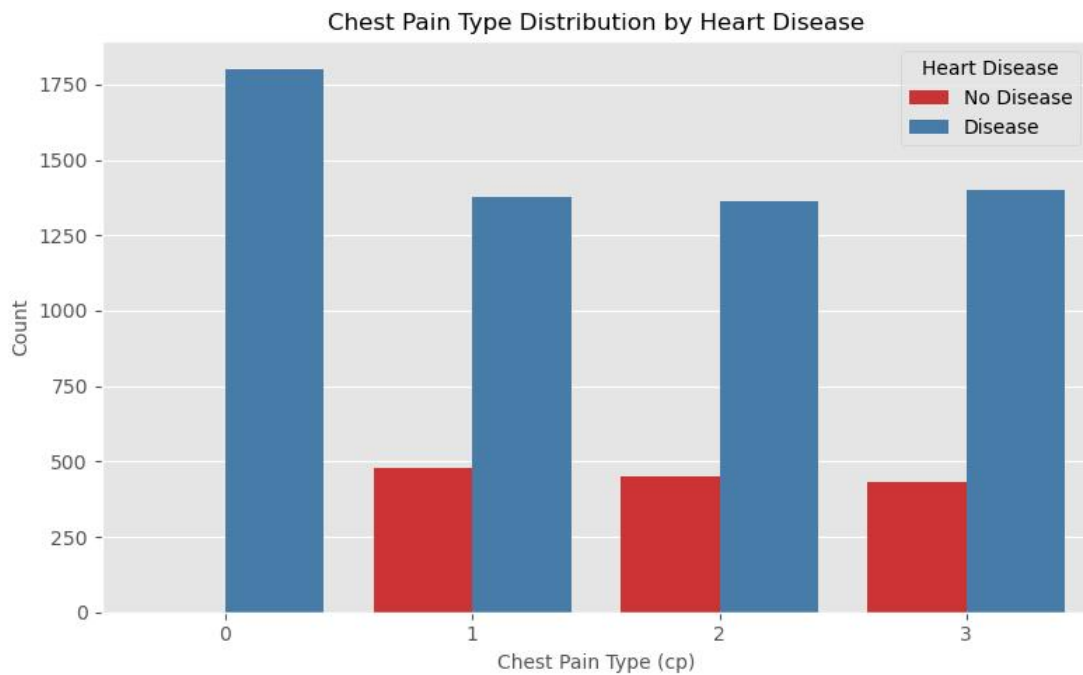
## 2. Age Distribution by Target



The KDE plot extends more to the right, indicating that there is indeed a subset of older patients who are at risk of having heart disease (target = 1).

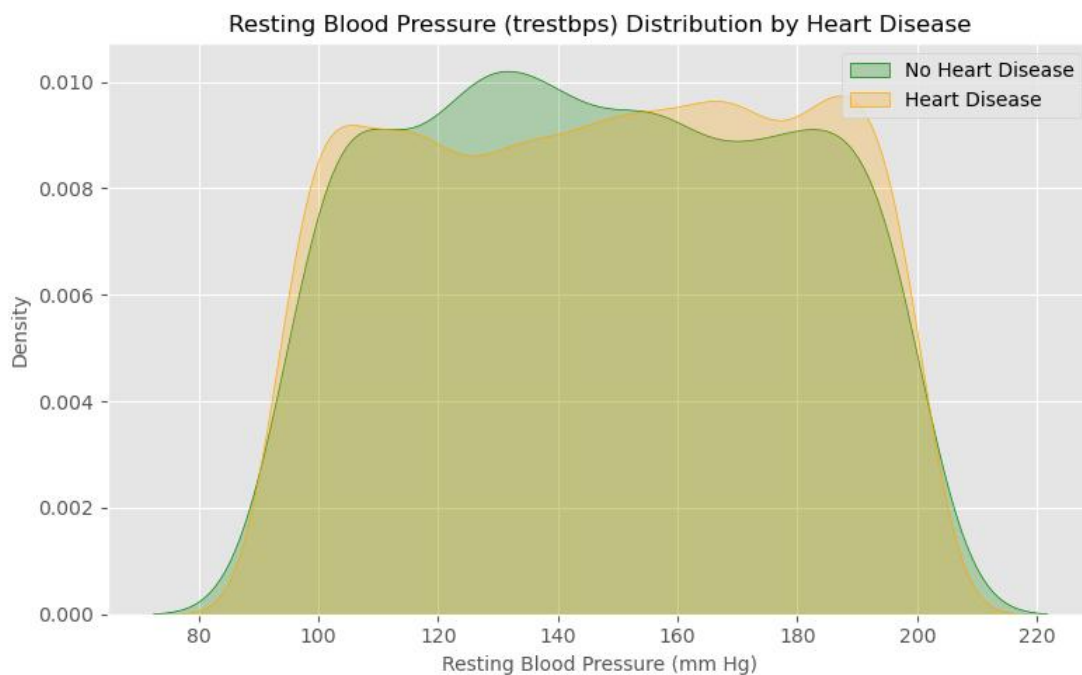
We can also observe that the distributions overlap to some degree, indicating that perhaps other factors come into play, such as genetics, lifestyle, and medical history.

### 3. Chest Pain Type (cp) vs Target



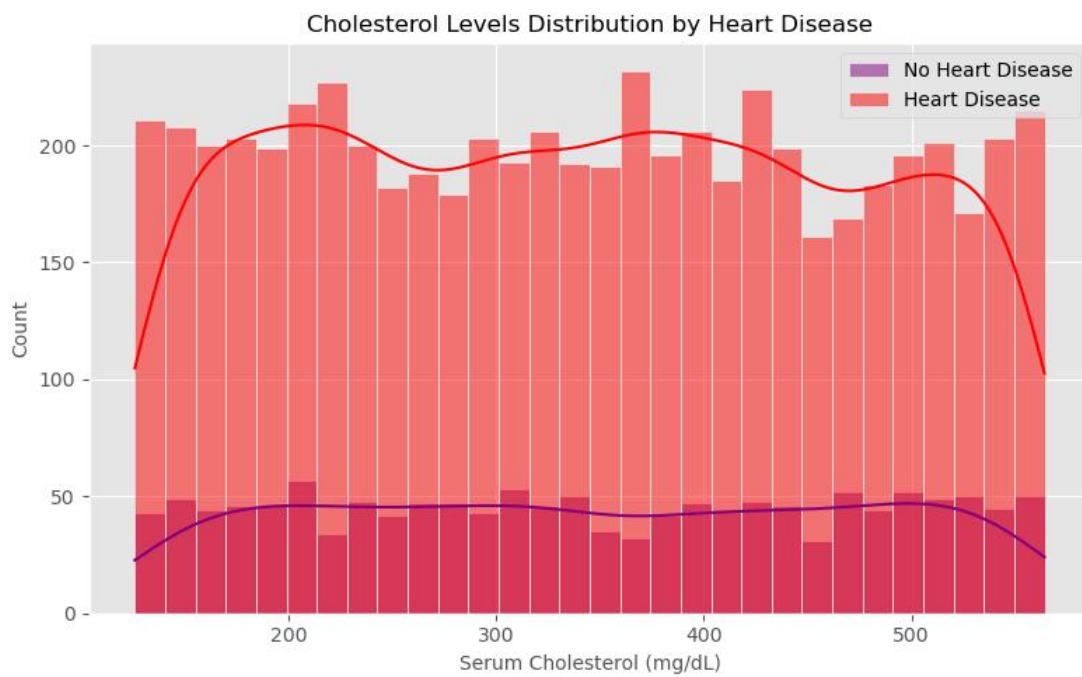
Types of chest pain vary between people without heart disease (target = 0) and people with heart disease (target = 1). In people without heart disease, the target is 0, and the type of chest pain is mainly no chest pain—Type 0. Chest pain Types 1, 2, and 3 are more associated with heart disease; target = 1. This indicates that there is some relationship between chest pain and heart disease. However, since some people without heart disease also have chest pains, this symptom cannot be used alone during the prediction of the disease.

#### 4. Resting blood pressure (trestbps) Distribution by Heart Disease



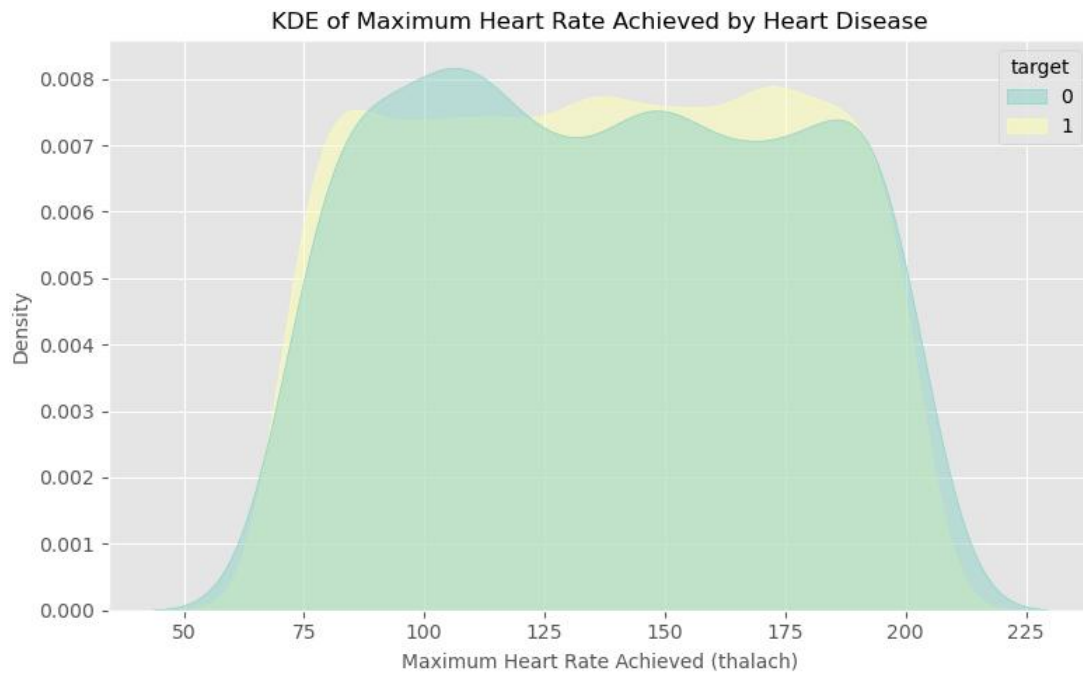
The distributions of resting blood pressure for people without and with heart disease overlap considerably. On the average, though, people with heart disease have somewhat higher blood pressure and the people with heart disease extend further to the right, which suggests that there may be a sub-population of people with truly very high resting blood pressure who are particularly at risk of heart disease.

## 5. Cholesterol Levels (chol) Distribution by Target



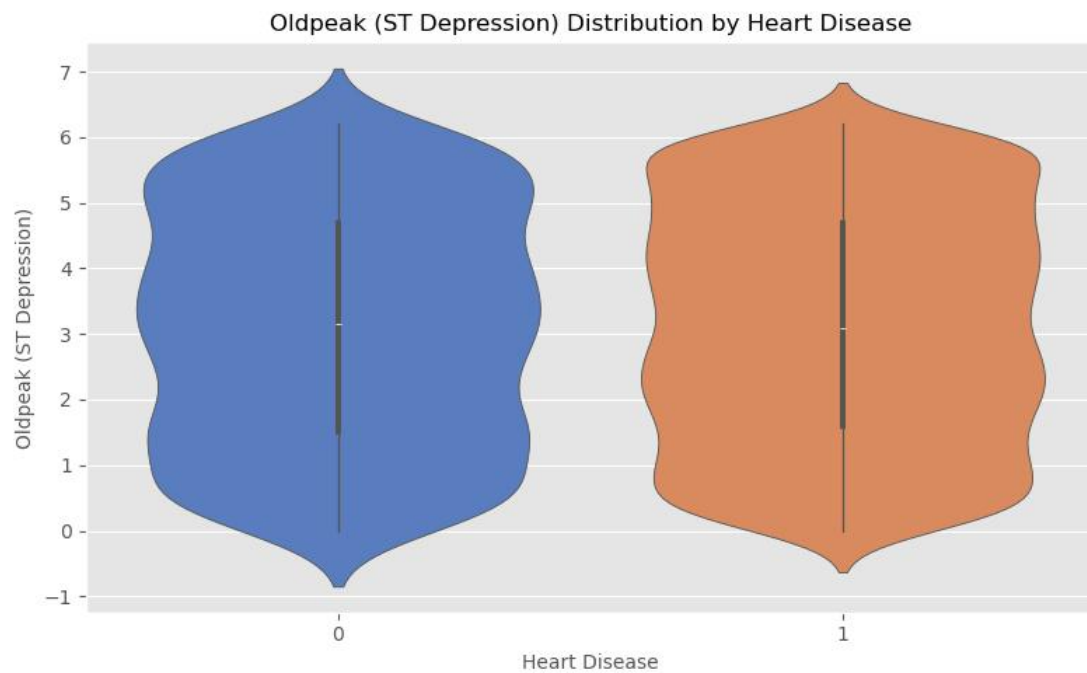
In the plot, we can see that people with heart disease generally have high cholesterol levels compared to people with low cholesterol levels.

## 6. Maximum Heart Rate Achieved (thalach) vs Target



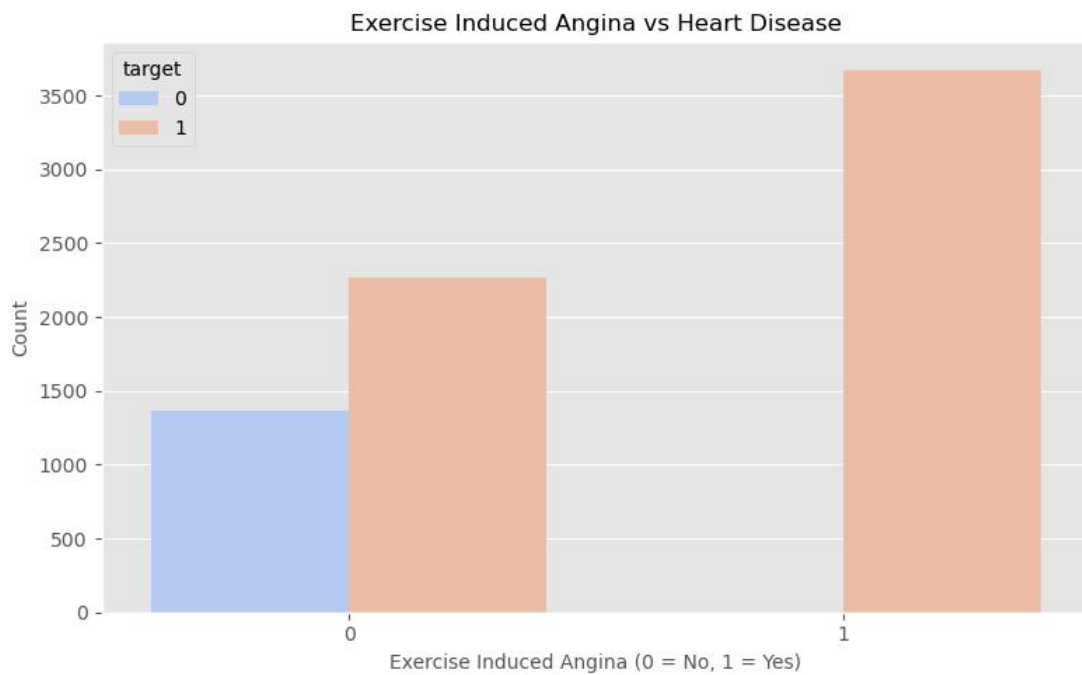
The KDE plots show that there is a lot of similarity in maximum heart rates for people with and without heart disease. We can also see from the plot that individuals with heart disease tend to have slightly lower maximum heart rates.

## 7. Oldpeak (ST Depression) Distribution by Target



The violin plot shows similar oldpeak values in both the heart disease and the non-heart disease patients, which suggests there might be some sort of relation between oldpeak and heart disease, although perhaps not too strong an indicator on its own.

## 8. Exercise Induced Angina vs Target



The chart illustrates that individuals with heart disease are more likely to report exercise-induced angina compared to those without heart disease. A larger number of heart disease patients (target = 1) experience angina during exercise, while individuals without heart disease (target = 0) generally do not.



## 9. Feature Summary

