

**TASK**

**Exploratory Data Analysis on the Heart Disease**

**Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The dataset includes information on 303 patients, considering 14 different attributes related to heart disease. The dataset contains both numerical and categorical data. The dataset does not contain any outliers.

**DATA CLEANING**

Column names and the values of categorical data was changed to make the data easier to read and interpret. No columns or attributes were dropped from the dataset.

MISSING DATA

There was no missing or inconsistent entries in the dataset.

DATA STORIES AND VISUALIZATIONS

Important relationships that were discovered are that men and patients between 41 and 54 are the most likely groups to experience heart disease. Maximum heart rates above 150 bpm could indicate increased risk of heart disease. Heart disease patients most typically experience 'atypical angina' chest pain, and do not typically experience exercise induced angina chest pain. Heart disease can be characterized by a flat slope of the peak exercise short term segment. Thalassemia sufferers with a fixed defect should be closely monitored for heart disease. Patients have a higher chance of having heart disease if their cholesterol levels are above 3 mg/dl, and it is difficult to identify/ pick up heart disease from looking at blood pressure and cholesterol only.

Heart disease is caused by a combination of different factors, making it tricky to treat. Because no strong correlations could be identified in this study, further investigation should be done as to what attributes are correlated with and causes heart disease, such as smoking and weight. There is much room for this study to be expanded.

**THIS REPORT WAS WRITTEN BY : AMY MARAIS**

