Heart Disease Analysis

By Unmilan Baruah

Introduction

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. This report is made to analyze the health and medical data, which is provided, for better future preparation. The following section is all about the purpose of this report. At the end, observations are recorded and conclusions are made.

Purpose

- To analyze the relationship between the **Target** variable, i.e. indication of heart problem, and the continuous variables: **Resting blood pressure**, **Serum cholestrol**, **Maximum heart rate achieved**, and **Oldpeak (ST depression caused by activity in comparison to rest)** while taking the **Age** of patients into account.
- To analyze the relationship between the **Target** variable, and the categorical variables: **Gender**, **Chest pain**, **Resting electrocardiographic results**, **Exercise induced angina**, **Fasting blood sugar**, **Slope of peak exercise ST segment**, **No. of major blood vessels colored by flourosopy**, and **Thalassemia**.

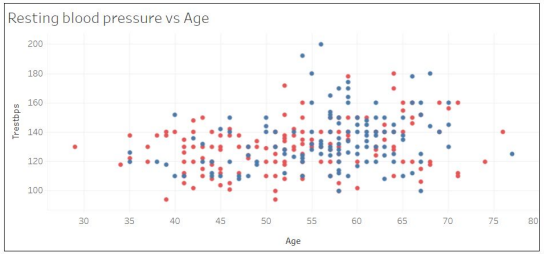
Methodology

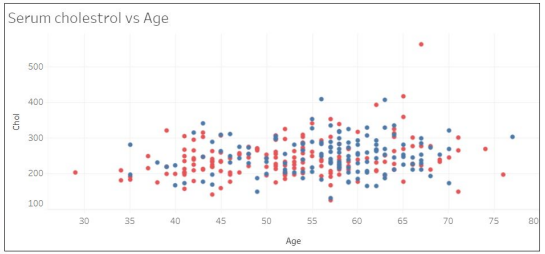
1. ETL (Extract, Transform, and Load): Collecting the data, transforming it into something which can be analyzed with accuracy, and loading the processed data for further analysis.
2. Data Analysis
3. Observations
4. Inference

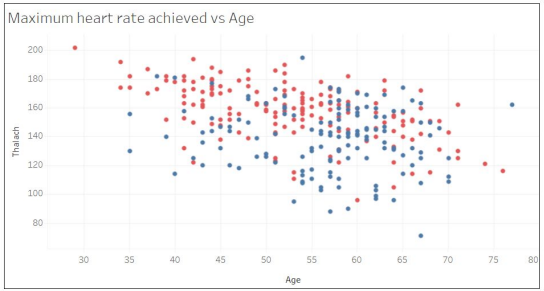
ETL (Extract, Transform, and Load)

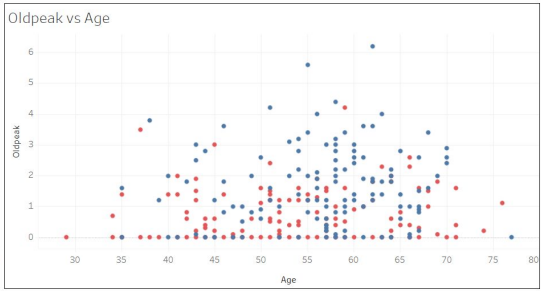
- The data source is named as “**Heart Disease data.csv**” and this CSV file has sales records from 2010 to 2017.
- There are total **1025** entries, but no null values. Therefore, data cleaning was not required.
- However, for analyzing the relationship between the **Target** and the **Categorical** variables, the file “**Dataset Details.txt**” is used and the data type of the **Categorical** variables are changed from integers to strings. This transformation was done in both Google Colab and Tableau.

Data Analysis

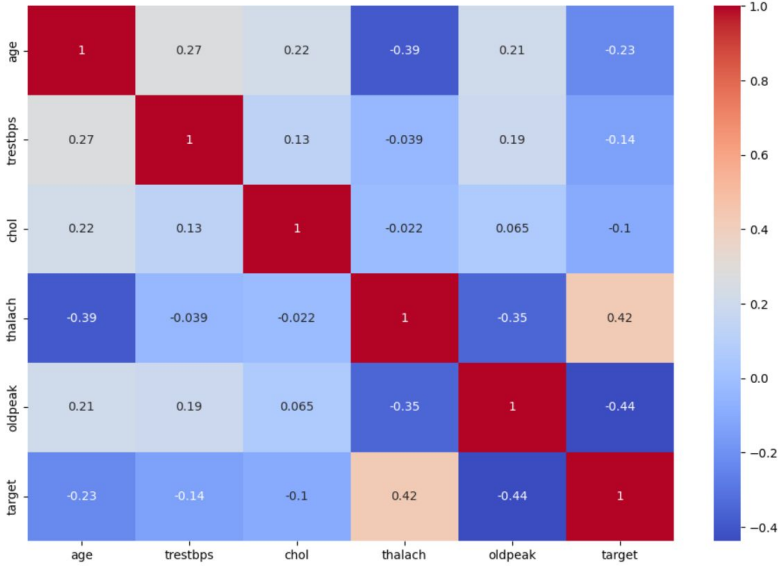




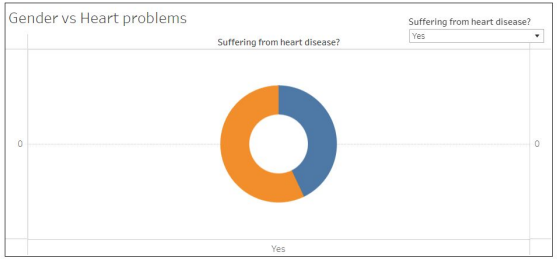


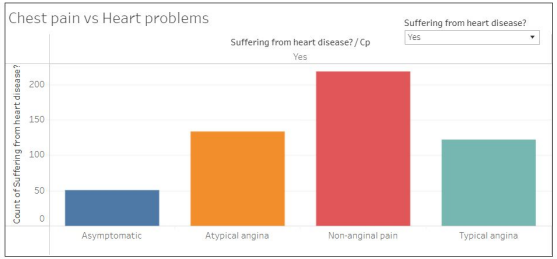


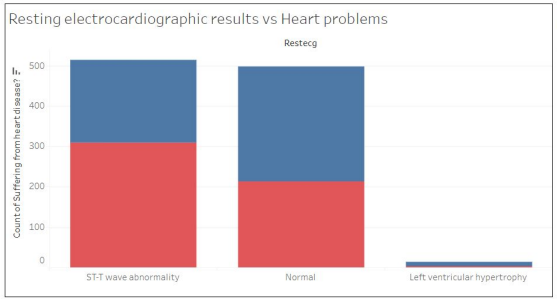
Data Analysis

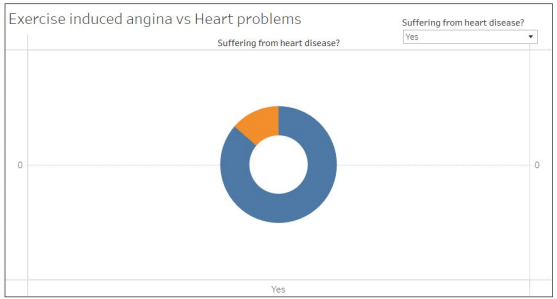


Data Analysis

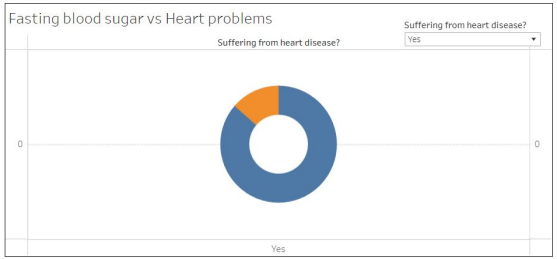


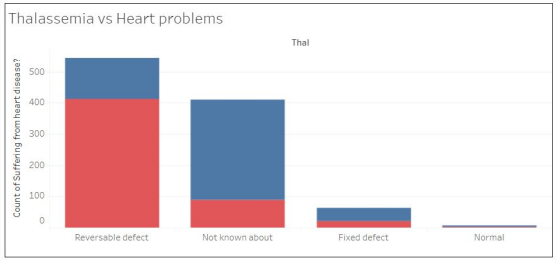


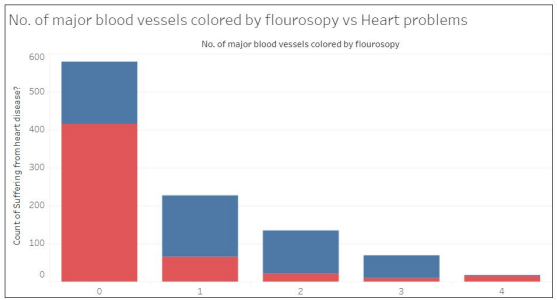


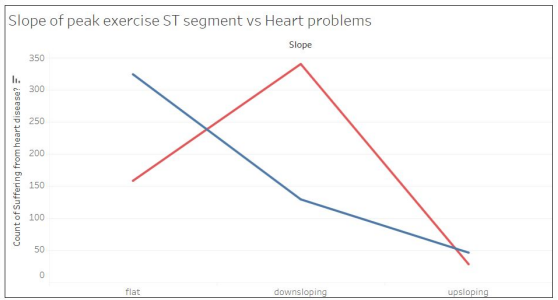


Data Analysis

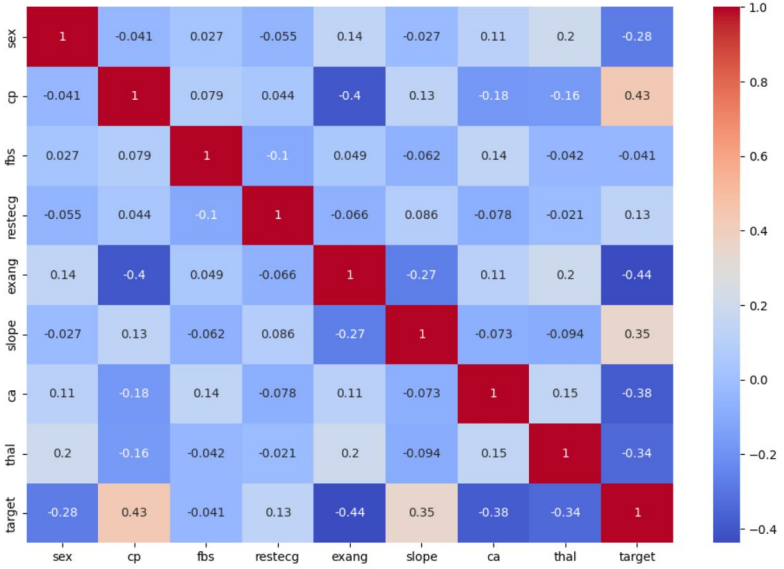








Data Analysis



Observations

- The correlation is less than 0.25/-0.25 or 25 percent between the **Target** and the variables like: **Age**, **Resting blood pressure**, **Serum cholestrol, Fasting blood sugar**, and **Resting electrocardiographic results**.
- The correlation is negative and above 25 percent between the **Target** and the variables like: **Oldpeak**, **Gender**, **Exercise induced angina**, **No. of major blood vessels colored by flourosopy**, and **Thalassemia**.
- The correlation is positive and above 25 percent between the **Target** and the variables like: **Maximum heart rate achieved**, **Chest pain**, and **Slope of peak exercise ST segment**.

Inference

- For those variables having correlation less than 25 percent with the **Target** variable implies that **Age**, **Resting blood pressure**, **Serum cholestrol, Fasting blood sugar**, and **Resting electrocardiographic results** has either no impact on chances of suffering from heart diseases or the impact is negligible compared to other variables when taken into account.
- For those variables having positive correlation and more than 25 percent with the **Target** variable clearly shows that increase in **Maximum heart rate achieved**, **Chest pain**, and **Slope of peak exercise ST segment** means increased chances of heart diseases.

Inference

- From the analysis it can also be observed that health conditions like: **Oldpeak**, **Exercise induced angina**, **No. of major blood vessels colored by flourosopy**, and **Thalassemia** can be symptoms of some other kind diseases but not necessarily of heart diseases. This is because of the negative correlation of above 25 percent of these health conditions with the **Target** variable.
- Based on the analysis of all the records, men suffer more than women. However, the percentage of women suffering from heart diseases is more than percentage of men. This indicates that women are more prone to heart problems than men.

Thank You