**HOW DIFFERENT FACTORS LEAD TO HEART DISEASE?**

Date: August 1, 2022

By: Emy Varghese

1. **Intro**

Data Analyst are increasing in the healthcare sector everyday as data drives the future of healthcare. Almost, 20 million people died from cardiovascular diseases in 2019, representing 32% of all global deaths. So, we chose the heart disease dataset to understand what factors lead to heart disease. In this project, we will use the Excel software to answer the question on what factors lead to heart disease.

**2.0 Understanding the dataset**

Load the dataset: [Heart Dataset](https://www.kaggle.com/datasets/zeeshanmulla/heart-disease-dataset)

*Information on the columns:*

**age**: age

**sex**: sex

**cp**: chest pain type (4 values)

**trestbps:** resting blood pressure

**chol**: serum cholesterol in mg/dl

**fbs**: fasting blood sugar > 120 mg/dl

**restecg**: resting electrocardiographic results (values 0,1,2)

**thalach**: maximum heart rate achieved

**exang**: exercise induced angina. Stable angina is usually triggered by physical activity. When you climb stairs, exercise or walk, your heart demands more blood, but narrowed arteries slow down blood flow

**oldpeak**: oldpeak = ST depression induced by exercise relative to rest

**slope**: the slope of the peak exercise ST segment. The ST/heart rate slope (ST/HR slope), has been proposed as a more accurate ECG criterion for diagnosing significant coronary artery disease (CAD)

**ca**: number of major vessels (0-3) colored by fluoroscopy.

**thal**: Thalassemia is an inherited blood disorder characterised by less oxygen-carrying protein (haemoglobin) and fewer red blood cells in the body than normal.3 = normal; 6 = fixed defect; 7 = reversable defect

**target**: Heart disease (0 = no, 1 = yes)

*The first four rows of the dataset:*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **age** | **sex** | **cp** | **trestbps** | **chol** | **fbs** | **restecg** | **thalach** | **exang** | **oldpeak** | **slope** | **ca** | **thal** | **target** |
| 2 | 52 | 1 | 0 | 125 | 212 | 0 | 1 | 168 | 0 | 1 | 2 | 2 | 3 | 0 |
| 3 | 53 | 1 | 0 | 140 | 203 | 1 | 0 | 155 | 1 | 3.1 | 0 | 0 | 3 | 0 |
| 4 | 70 | 1 | 0 | 145 | 174 | 0 | 1 | 125 | 1 | 2.6 | 0 | 0 | 3 | 0 |
| 5 | 61 | 1 | 0 | 148 | 203 | 0 | 1 | 161 | 0 | 0 | 2 | 1 | 3 | 0 |

**3.0 Data Cleaning**

1. Filter the header:
   1. Click row 1
   2. Go to the data tab and click the filter button
2. Remove duplicates
   1. In the same data tab, click the remove the duplicate button
   2. Click select all, and it will tell how many duplicates there were
   3. Now the total number of observations are 303.
3. Add Age Bracket
   1. Left Click column B and insert a new row called age bracket
   2. Make if statement stating IF(A2>54,"Old", IF(A2>=31,"Middle Age", IF(A2<31,"Adolescent","invalid")))
   3. So, B2 will give Middle age since A2 is 52 years old
   4. Then copy the statement all the way to the bottom.
4. Replace sex
   1. Click column C
   2. In the home tab, click the find and select button
   3. Click the replace button, and change the search from by rows to by columns
   4. In the find what section type 1 and replace it with Male and click the replace all button
   5. Do the same for 0 and replace it with Female
5. Replace target
   1. Click column O
   2. In the home tab, click the find and select button
   3. Click the replace button, and change the search from by rows to by columns
   4. In the find what section type 1 and replace it with Has Heart Disease and click the replace all button
   5. Do the same for 0 and replace it with No Heart Disease
6. Replace exang
   1. Click column J
   2. In the home tab, click the find and select button
   3. Click the replace button, and change the search from by rows to by columns
   4. In the find what section type 1 and replace it with Yes and click the replace all button
   5. Do the same for 0 and replace it with No

**4.0 Data insights**

There are 164 people that have heart disease from the data collected. Middle and Old age people were more affected with heart disease.

Males are more likely to have heart disease. Chest Pain has direct correlation to people that have heart disease.

For trestbps: resting blood pressure anything above 130-140 is a cause of concern. Gender does not impact the target so bp is neutral for the genders.

Cholesterol levels that are above 200 is a cause of concern. The person having ST-T wave abnormality are more likely to have to heart disease.

Thalach is the maximum heart rate a person achieved, and if that heart rate reached above 140 then they are more likely to have a heart disease.

Heart disease patient are more likely to have a old peak value of 0.6. Fixed defect type Thalessemia are more effected by heart disease

**5.0 Data Visualization**

****

**My work :** [**file:///Users/emyvarghese/Documents/heart dataset.xlsx**](file:///Users/emyvarghese/Documents/heart%20dataset.xlsx)