Team 4: Heart Disease Indicators

DATS 6103: Summary Report

Professor Ning Rui

December 18, 2022

Predicting Heart Disease Using Health Indicators

**Introduction**

Heart Disease is one of the most common chronic diseases in the United States. Each year millions of Americans are impacted. For many, this means they incur the cost of expensive medical bills. For a smaller subsect though, it is even worse. In 2020, heart disease was the leading cause of death in the United States, claiming the lives of nearly 700,000 individuals. Because of how deadly the disease is, it is important to get out ahead of it and identify risk factors. Unfortunately, people often learn they have heart disease after a major, and potentially deadly symptoms such as chest pain, a heart attack, or cardiac arrest. As noted, there are risk factors people should consider before the point is reached where major symptoms become a problem. Some examples of risk factors include natural ageing, the building up of plaque in arteries, chronic inflammation, high blood pressure, and diabetes. Ideally individuals should assess themselves for risk factors before they suffer from one of the major symptoms.

Using the breadth of data from the Center for Disease Control’s Behavioral Risk Factor Surveillance System (BRFSS), we look at indicators of heart disease and attempt to create a model to assess an individuals risk for heart disease. The BRFSS is a telephone survey first conducted by the CDC in 1984. It is conducted each year and over 400,000 respondents are reached each year. Respondents are asked about a range of health-related risk behaviors such as if they smoke or if they get regular exercise. They are also asked about chronic conditions such as if they have heart disease. We use data from a cleaned up version of the 2015 survey that has 253,680 observations. Our data set has 22 variables, 21 features and the target. Most of our variables are binary, such as if an individual has high cholesterol or not. Our target variable, if an individual has heart disease or has had a heart attack, is also binary. Non-binary variables include BMI, education level, among others. The following chart presents every variable in the data set.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **HighBP​**​ | **HighChol​**​ | **CholCheck​**​ | **BMI​**​ | **Smoker​**​ | **Stroke​**​ | **Diabetes​**​ | **PhysActivity​**​ |
| **Fruits​**​ | **Veggies​**​ | **HvyAlcoholConsump​**​ | **Any​**​  **Healthcare​**​ | **NoDocbc​**​  **Cost​**​ | **GenHlth​**​ | **MentHlth​**​ | **PhysHlth​**​ |
| **DiffWalk​**​ | **Sex​**​ | **Age​**​ | **Education​**​ | **Income​**​ | **Target - HeartDiseaseorAttack​**​ | | |

Analysis

EDA and Graph: -

We can examine the distribution of the population depending on those with and without heart disease from the count plot shown above. Out of 253680 observations, 23893 people have heart-related illness, compared to 229787 who do not. In percentage words, we may say that 90.58 percent of people are in good health and 9.42 percent are suffering from a heart-related condition.

Chart, bar chart

Description automatically generated

Let's use the correlation matrix to visualize the relationship between the variables.

Chart

Description automatically generated

As the correlation matrix shows, the variables like GenHlth, Age, DiffWalk, HighBP, Stroke, PhysHlth, HighChol, Diabetes, Smoker, Sex, MentHlth, BMI show the positive correlation with the target variable or dependent variable HeartDiseaseorAttack where as the variables like CholCheck, NoDocbcCost, AnyHealthcare, Fruits, HvyAlcoholConsump, Veggies show no correlation with the target variable and the variables PhysActivity, Education and Income shows the negative correlation with the target variable.

Now, using various plots and statistical tests, let's examine how various variables impact the target variable while also providing answers to our SMART questions.

* SMART Question 1: - Is age associated with having heart disease or a heart attack?

First, the age variable value has been broken down into the following categories: -

Age categories:

# 1.0 -> 18 – 24 # 2.0 -> 25 – 31 # 3.0 -> 32 – 38 # 4.0 -> 39 - 44

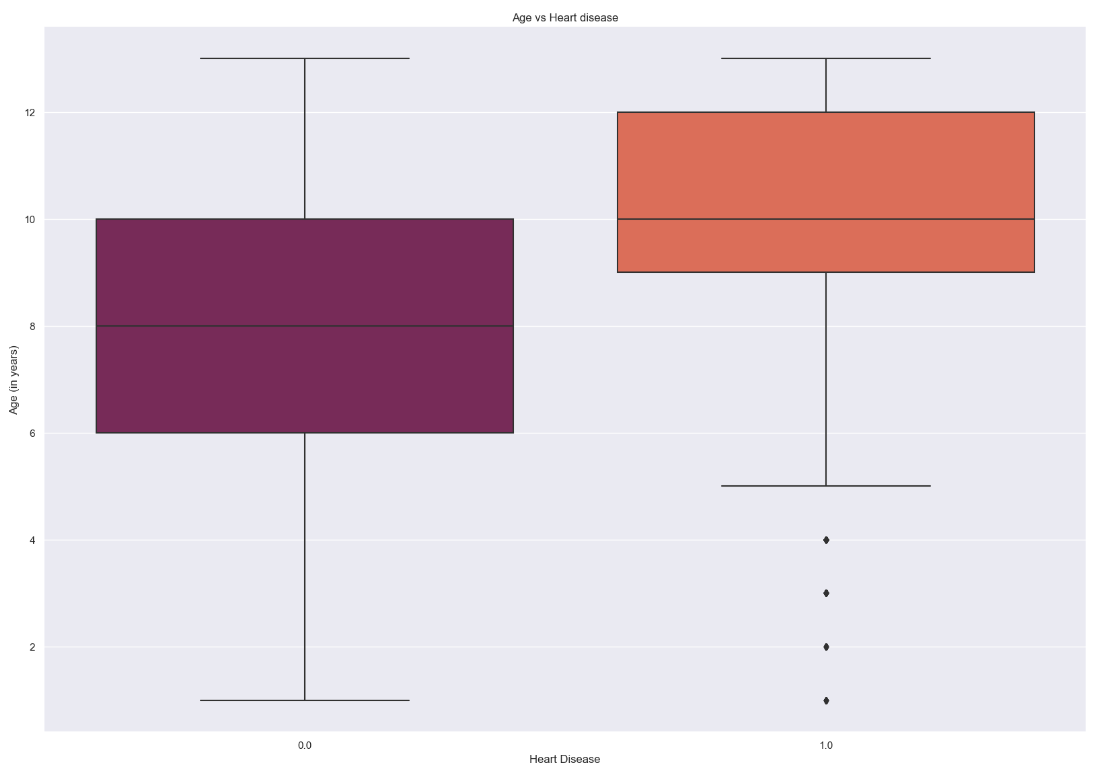
# 5.0 -> 45 – 51 # 6.0 -> 52 – 58 # 7.0 -> 59 – 65 # 8.0 -> 66 - 72

# 9.0 -> 73 – 79 # 10.0 -> 80 – 86 # 11.0 -> 87 – 93 # 12.0 -> greater than 93

More people fall into the age category 9 (those ages 73 to 79), which is followed by category 10 (those ages 80 to 86), which includes the next largest group of individuals. According to the box plot, the median age for those without heart disease is age category 8, which is between 66 and 72 years old, while the median age for those with heart disease is age category 10, which is between 80 and 86 years old. Therefore, if a person is older than 70, their risk of developing heart disease increases.

Chart, bar chart

Description automatically generated



* SMART Question 2: - Is having high BP associated with having heart disease or a heart attack?

People with high blood pressure make up 42.9% of the total observations. According to the count plot, those without high blood pressure typically have a lower risk of developing heart disease, whereas those with high blood pressure have an increased risk of developing or already having the condition.

Chart, bar chart

Description automatically generated

We can observe from the frequency table that 17928 out of 23893 persons have heart disease, or about 75% of the total. Therefore, we may conclude that High BP is one of the major factors influencing the target variable.

Graphical user interface, text, application, chat or text message

Description automatically generated

* SMART Question 3: - Is BMI associated with having heart disease or a heart attack?

BMI is often broken down into many ranges. Your BMI is considered underweight if it is less than 18.5. Your BMI falls into the Healthy Weight range if it is between 18.5 and 24.9. Your BMI is considered overweight if it is between 25.0 and 29.9. Your BMI is considered obese if it is 30.0 or greater. Most of the observations have BMIs greater than 25, as seen by the histogram. People with BMI over 25 make up over 79% of those with heart disease overall.

Chart, histogram

Description automatically generated

* SMART Question 4: - Is consuming veggies and fruits associated with having heart disease or a heart attack?

We can see from the fruit plots that those who don't eat fruit have a lower risk of developing heart disease, whereas those who do seem to have a higher risk of developing the condition. Similar evidence can be seen in the plot for vegetables, where it is revealed that people who eat vegetables have a higher risk of developing heart disease than those who do not.

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

* SMART Question 5: - Is having variable income associated with having heart disease or a heart attack?

The income is divided into 8 categories as follows:

# 1.0 -> less than $10,000 # 2.0 -> $20,000 # 3.0 -> $30,000

# 4.0 -> $40,000 # 5.0 -> $50,000 # 6.0 -> $60,000

# 7.0 -> $70,000 # 8.0 -> $75,000 more

According to the correlation matrix, "Income" and "Heart Disease or Attack" have a negative connection (score of -0.1). The bar graph shows that when income rises, some income groups' risk of having a heart attack stays the same or doesn't rise.

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

* SMART Question 6: - Is heavy alcohol consumption associated with having heart disease or a heart attack?

According to the correlation matrix, there is no connection between "Heavy Alcohol Consumption" and "Heart Attack or Disease." According to the bar plot, those who don't consume alcohol or who drink alcohol in moderation are highly unlikely to get heart disease or an attack.

Chart, bar chart

Description automatically generated

* SMART Question 7: - Is having high cholesterol associated with having heart disease or a heart attack? The value from the correlation matrix is .2, indicating there is some positive correlation between having high cholesterol and having heart disease or a heart attack. Looking at the graph below, we can see that the proportion with heart disease or attack is much higher for the group with high cholesterol versus the group without high cholesterol.

Chart, bar chart

Description automatically generated