Sabrina Yang
PM-566: Introduction to Health Data Science

Final Project Write-up

Investigating Factors' Association with Heart Disease

Introduction

According to CDC, heart disease is the leading cause of death in the United States. 47% of the people in the United States have at least one of three key risk factors for heart disease: high blood pressure, high blood cholesterol, and smoking. Other key factors like BMI, not getting enough physical activity or drinking too much alcohol can also contribute to heart disease. The impact of these risk factors on heart disease is multifaceted.

High blood pressure, a condition where the force of blood against the artery walls is consistently too high, strains the cardiovascular system, increasing the likelihood of heart disease. Elevated blood cholesterol levels, particularly high levels of low-density lipoprotein (LDL) cholesterol, can lead to the buildup of plaque in arteries, narrowing blood vessels and hindering proper blood flow, thereby raising the risk of heart disease.

Smoking can damage blood vessels, decrease the oxygen supply to the heart, and accelerate the formation of arterial plaque. High Body Mass Index (BMI), indicative of obesity, puts additional strain on the heart and is associated with an elevated risk of heart disease. Insufficient physical activity contributes to heart disease by reducing cardiovascular fitness, impairing blood vessel function, and promoting weight gain. Excessive alcohol consumption is linked to an increased risk of heart disease, as it can raise blood pressure, contribute to obesity, and negatively impact heart function.

Thus, addressing these modifiable risk factors through lifestyle modifications and medical interventions is crucial for preventing and managing heart disease.

Question of Interest

How BMI, physical health, alcohol consumption (yes or no), smoking, physical activity associate with heart disease?

Methods

The dataset used in this study was acquired from Kaggle https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease/data. Analysis through assessing various plots, tables and graphs was performed to identify association between heart disease and risk factors, including the examination of the effect of high blood pressure, high blood cholesterol, smoking, BMI, physical activity, and alcohol consumption. The data was cleaned by replacing extreme values to 'NA's, shortening variable names, and creating new factor variable for better analysis. Skim function from skimr package was used to explore data. Dim, head and tail, summary, and table functions were used to check detailed observations. BMI was stratified into four groups (underweight, healthy, overweight, and obese) for better understanding of relationship with heart disease.

Data inspection and indicators

The data was from 2020 with 319795 data points and 18 variables. The heart disease health indicators based on the BRFSS 2015 Codebook https://www.cdc.gov/brfss/annual_data/2015/pdf/codebook15_llcp.pdf

- BMI: Body Mass Index.
- Alcohol drinking: Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week)
- Smoking: Have you smoked at least 100 cigarettes in your entire life?
- Physical activity: Adults who reported doing physical activity/exercise during the past 30 days other than their regular jobs
- Physical health: Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?

Data inspection includes: check dimensions using dim(heart), check the first few rows (headers) using head(heart), check the last few rows (footers) using tail(heart), and check variable names and types using str(heart). Rename variables to shorter length and all lower case.

- BMI is right skewed and has a range from 12.02 to 94.85 with mean of 28.33.
- Physical Health right skewed and has a range from 0 to 30 with mean of 3.37.
- Mental Health right skewed and has a range from 0 to 30 with mean of 3.90.

There is no missing value in the dataset.

Preliminary Results

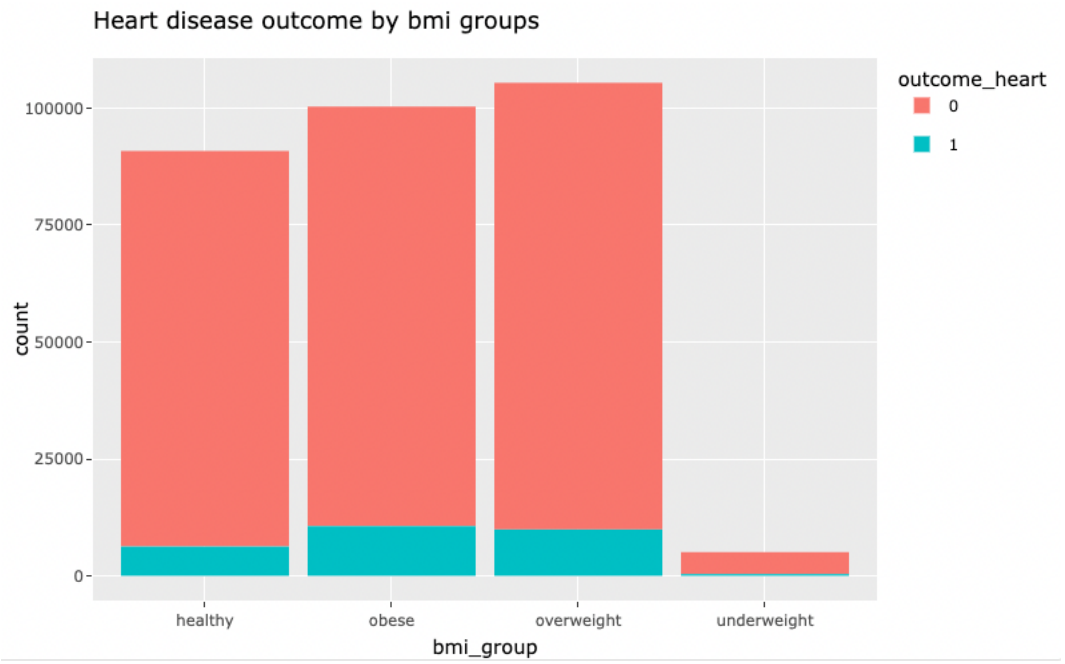

Figure 1: Heart disease outcome by BMI groups

Description: The figure of heart disease outcome by BMI groups depicts the distribution of individuals based on their categorized BMI group and occurence of heart disease. BMI was categorized into groups based on CDC standard. Individuals whose BMI is less than 18.5 would be in underweight group, individuals whose BMI is greater or equal to 18.5 but less than 25 would be in healthy group, individuals whose BMI greater or equal to 25 but less than 30 would be in overweight group, and individuals whose BMI greater or equal to 30 would be in obese group.

Findings: The data visualization reveals a predominant pattern wherein a significant proportion of individuals in the study exhibit are either obese or overweight. Within this demographic, it was observed that the majority of individuals diagnosed with heart disease are concentrated in the obese and overweight categories, signifying a potentially heightened risk of cardiovascular issues associated with higher body weight. Intriguingly, the underweight group comprises the fewest individuals with heart disease, suggesting a distinct association between body weight and the prevalence of heart-related conditions. This correlation underscores the relevance of weight management in cardiovascular health, emphasizing the need for targeted interventions and lifestyle modifications to address and mitigate the impact of excess weight on heart disease risk within the studied population.

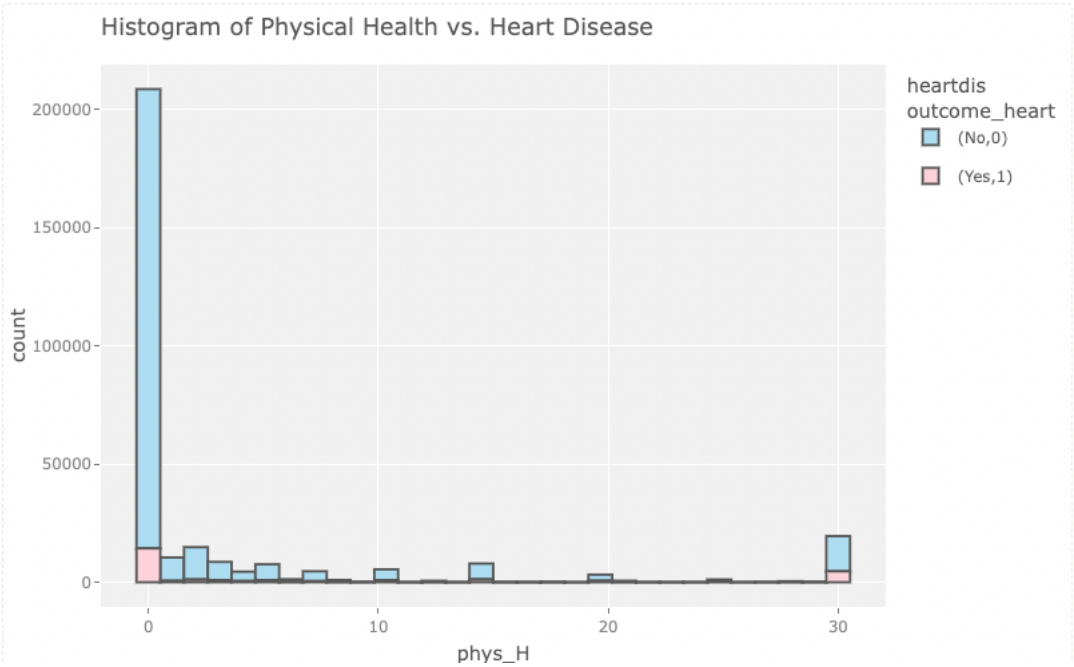

Figure 2: Physical Health vs Heart Disease

Description: The figure above illustrates the distribution of individuals based on their physical health status and the corresponding occurrences of heart disease and absence of heart disease within each category of physical health.

Findings: Based on the histogram visualization, the most abundant group is the people who consider themselves "good" for the past 30 days (0 for how many days during the past 30 days was your physical health not good?). Within that group (total of 208611 people), 194301 people have no heart disease (93.14% of the group) while 14310 have heart disease (6.86% of the group). Within a group of 7904 people who consider themselves as 15 days out of 30 days having "good" physical health, there were 6662 people without heart disease (84.3%) and 1242 people with heart disease (15.7%). Within a group of 19503 people who considered themselves 0 day out of 30 days having "good" physical health, there were 14884 people without heart disease (76.3%) and 4619 people with heart disease (23.7%). Remarkably, the data within each subgroup reveal a trend: as the number of days individuals perceive their physical health as "good" decreases, the proportion of individuals within that subgroup reporting the presence of heart disease increases.

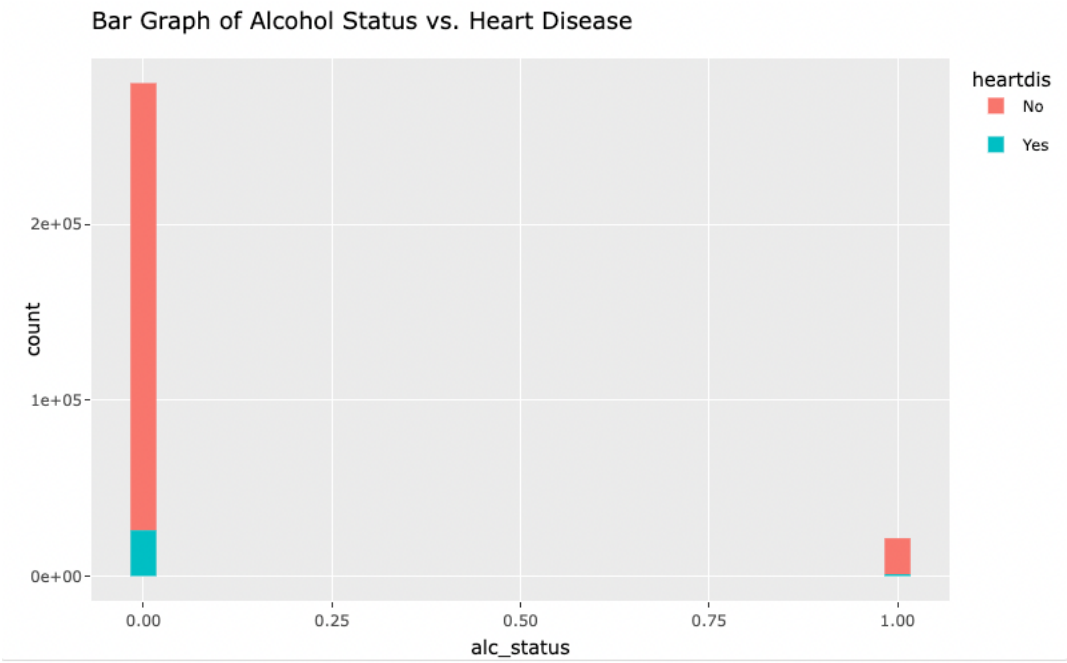

Figure 3: Alcohol Status vs Heart Disease

Description: The figure of alcohol status vs heart disease above illustrates potential relationship between alcohol status and heart disease. The y-axis simply a count of the individuals in the study in order to compare between alcohol status.

Findings: Among 280,136 individuals classified as non-heavy drinkers, the plot reveals that 26,120 individuals, constituting 9.32% of this group, report the presence of heart disease. On the other hand, among the 21,581 individuals classified as heavy drinkers, a lower proportion—1,141 individuals, or 5.29% of this particular group—acknowledge the occurrence of heart disease. These findings suggest a potential association between heavy alcohol consumption and a lower prevalence of heart disease within this specific population, underscoring the need for further investigation into the intricate relationships between drinking habits and cardiovascular health outcomes. For future analysis, it is crucial to consider additional factors such as type of alcohol consumed, individual variation in their susceptibility to the effects of alcohol to gain a comprehensive understanding of the impact of alcohol consumption on heart disease risk within distinct subgroups of the population.

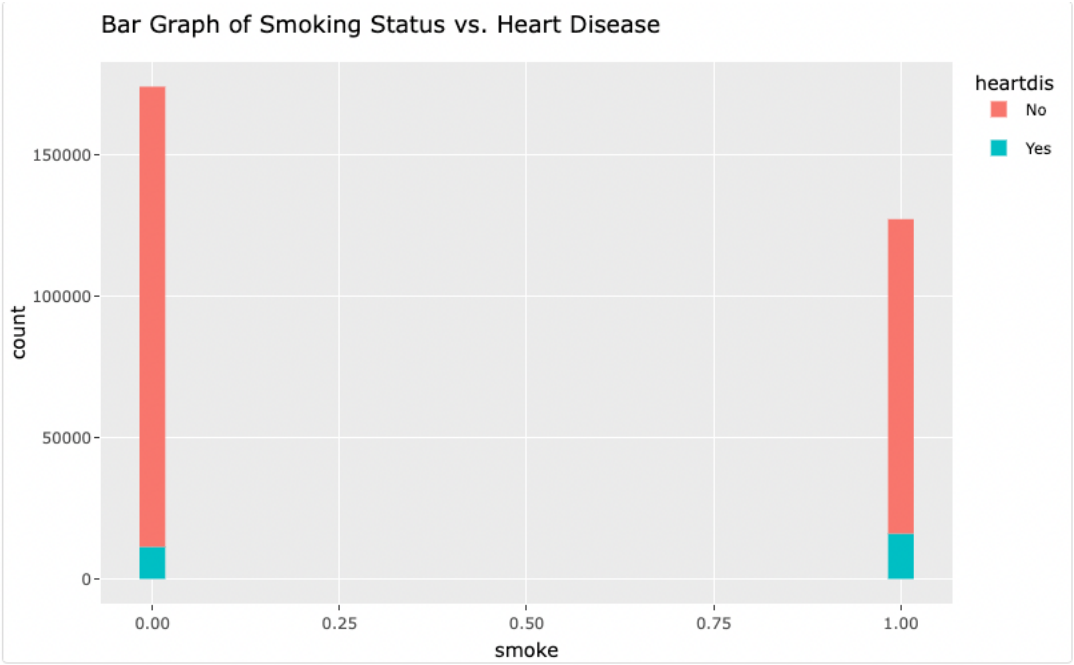

Figure 4: Smoking Status vs Heart Disease

Description: The figure of smoking status vs heart disease above depicts the distribution of individuals within the study, categorized by their smoking status (divided into smoking (1) and non-smoking (0) groups) and whether or not they have heart disease.

Findings: The stacked bar graph distinctly illustrates a clear comparison between the two groups based on their smoking status- among individuals with heart disease, the proportion of individuals who smoke is visibly larger when compared to the smoking population within the group without heart disease. This observation suggests a potential association between smoking and an increased likelihood of developing heart disease within the studied population.

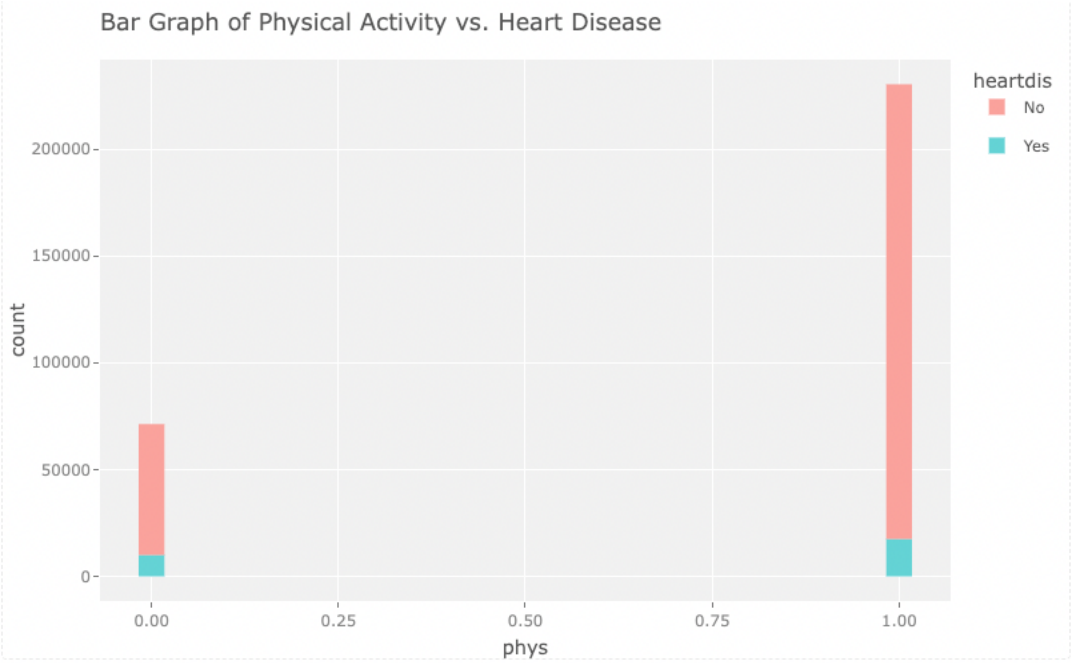

Figure 5: Physical Activity vs Heart Disease

Description: The figure of physical activity vs heart disease above illustrates the distribution of individuals within the study, categorized by their physical status (divided into physical activity (1) and no physical activity (0) groups) and whether or not they have heart disease.

Findings: Among the individuals who reported no physical activity or exercise beyond their regular job duties in the past 30 days, a total of 9,882 individuals are identified as having heart disease. This translates to a proportion of heart disease within this subgroup as 13.9% (9882 / (9882 + 61,423)). In contrast, within the cohort that engages in physical activity, a total of 17,379 individuals have been identified as having heart disease. The proportion of heart disease within this group is 7.54% (17379 / (17379 + 213033)). The proportions indicate that the prevalence of heart disease is notably higher among individuals who report no physical activity compared to those who engage in physical activity.

Conclusion

According to CDC and aside from the key risk factors like high blood pressure, high blood cholesterol, and smoking, several other medical conditions and lifestyle choices can put people at a higher risk for heart disease which include but not limited to physical inactivity, and overweight and obesity.

Building on this foundation, the exploratory data analysis conducted as part of the study further examined the relationship between specific variables and heart disease

risk. The findings reveal that certain factors exhibit a positive association with heart disease. Specifically, physical inactivity, a higher Body Mass Index (BMI), smoking, and the overall status of physical health demonstrate a positive relationship with the presence of heart disease in the studied population.

This comprehensive assessment underscores the multifactorial nature of heart disease, highlighting that a spectrum of health conditions and lifestyle choices beyond the traditional risk factors plays a role in influencing cardiovascular outcomes. By identifying these associations, this analysis provides insights that can inform targeted interventions, preventive strategies, and public health initiatives. It also emphasizes the importance of considering a broad range of factors when assessing an individual's risk for heart disease and underscores the need for comprehensive and personalized approaches to cardiovascular health promotion and disease prevention.