ANALYZING CARDIOVASCULAR HEALTH DISPARITIES

Understanding and mitigating cardiovascular health disparities and in-depth analysis of risk factors and health behaviors in the adult population.

Authors:

- Ali Bin Kashif
- Muhammad Aman

Presented To

Healthcare Management

Data Analysis Report **February 2024**

amanmuhammed0987@gmail.com

alibinkashif007@gmail.com

Acknowledgement

We would like to express our sincere gratitude to the following individuals for their valuable contributions to this data analysis project.

ALI BIN KASHIF

Aspiring Data Scientist, Data Analyst

I am a dedicated professional with a strong interest for data science and analytics. My journey has been marked by continuous learning and a commitment to solving complex problems. I believe in the transformative power of data to drive meaningful insights and impact businesses positively.

Currently pursuing a bachelors degree in Computer Science, I possess a solid foundation in programming, and have an ability to adapt and learn new technologies.

My enthusiasm for learning and my curiosity to explore the realm of data analytics, machine learning, and artificial intelligence have fueled my passion for this exciting career path. I am eager to contribute my skills to projects that challenge me and allow me to apply my knowledge.

MUHAMMAD AMAN

Undergraduate Data Scientist, Data Analyst

I am Muhammad Aman, currently an undergraduate student at NED University of Engineering and Technology in Pakistan. Pursuing a Bachelor's degree in Computer Science with a specialization in Data Science, I am deeply committed to explore the realm of data-driven insights and problem-solving.

At NED University, I'm committed to refining my skills and broadening my expertise in data science. Through challenging coursework and practical projects, I'm continuously endeavoring to enhance my comprehension of statistical analysis, machine learning algorithms, and data visualization techniques.

I'm excited about the possibilities of data science and how it can change industries for the better. My goal is to use my skills to make a positive difference in the world.

Table of Contents

Introduction

According to the **CDC**, heart disease is a leading cause of death for people of most races in the U.S. (African Americans, American Indians and Alaska Natives, and whites). About half of all people (47%) have at least 1 of 3 major risk factors for heart disease: high blood pressure, high cholesterol, and smoking. Other key indicators include diabetes status, obesity (high BMI), not getting enough physical activity, or drinking too much alcohol. Identifying and preventing the factors that have the greatest impact on heart disease is very important in healthcare. In turn, developments in computing allow the application of machine learning methods to detect "patterns" in the data that can predict a patient's condition.

Dataset Descripton

The dataset originally comes from the **CDC** and is a major part of the Behavioral Risk Factor Surveillance System (BRFSS), which conducts annual telephone surveys to collect data on the health status of U.S. residents.

As described by the CDC: "Established in 1984 with 15 states, BRFSS now collects data in all 50 states, the District of Columbia, and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world. The most recent dataset includes data from 2023.

Variables

The dataset comprises a comprehensive set of variables pertaining to individual health and healthcare practices. It covers demographic details like **race/ethnicity**, **age**, **height**, and **weight**, alongside lifestyle factors such as **physical activities**, **sleep duration**, **smoking**, and **alcohol consumption**. Health conditions including **heart attack**, **angina**, **stroke**, **asthma**, **diabetes**, and sensory impairments like **deafness** or **vision difficulties** are recorded. Additionally, it documents **difficulties in concentration**, **walking**, **dressing/bathing**, and **running errands**. Medical procedures such as **chest scans** and vaccination history for **flu**, **pneumonia**, and **tetanus** are also noted, along with **HIV testing** and **COVID-19** status. These variables offer insights into individual health status and healthcare behaviors for comprehensive analysis.

Problem Statement

The problem centers on **identifying** and **understanding** the **factors** that exert the most significant influence on **heart disease prevalence**. This exploration is crucial for healthcare initiatives and interventions aimed at prevention and management. The dataset encompasses a range of variables, including demographic information, health behaviors, chronic conditions, and COVID-19 status, providing a rich source for uncovering patterns, correlations, and disparities in cardiovascular health.

We're looking at big questions like what things make heart problems more likely, how mental and physical health are linked, what habits and steps help prevent heart issues, how other health problems affect the heart, and if issues like hearing or vision problems are connected to heart disease. We're also checking how smoking and e-cigarettes are used and how COVID-19 affects heart health.

Our main goal is to extract actionable insights from this wealth of data. We want to make plans that focus on specific things to help reduce differences in heart health and make everyone's heart healthier in the adult population.

Methodology

Data Source & Collection:

The data come from the BRFSS, a CDC initiative. It conducts telephone surveys annually to collect health information from residents. Random-digit-dialing techniques are used to reach a representative sample. Participants are asked standardized questions about their health and behaviors. The collected data are then analyzed to understand health trends in the population.

Analytical Techniques:

- **Descriptive statistics:** Summarizing and describing the main features of the dataset, such as mean, median, mode, standard deviation, and range.
- **Exploratory data analysis (EDA):** Using graphical and statistical techniques to explore the data, identify patterns, trends, and relationships among variables.
- **Correlation analysis:** Examining the strength and direction of relationships between variables to identify potential associations.

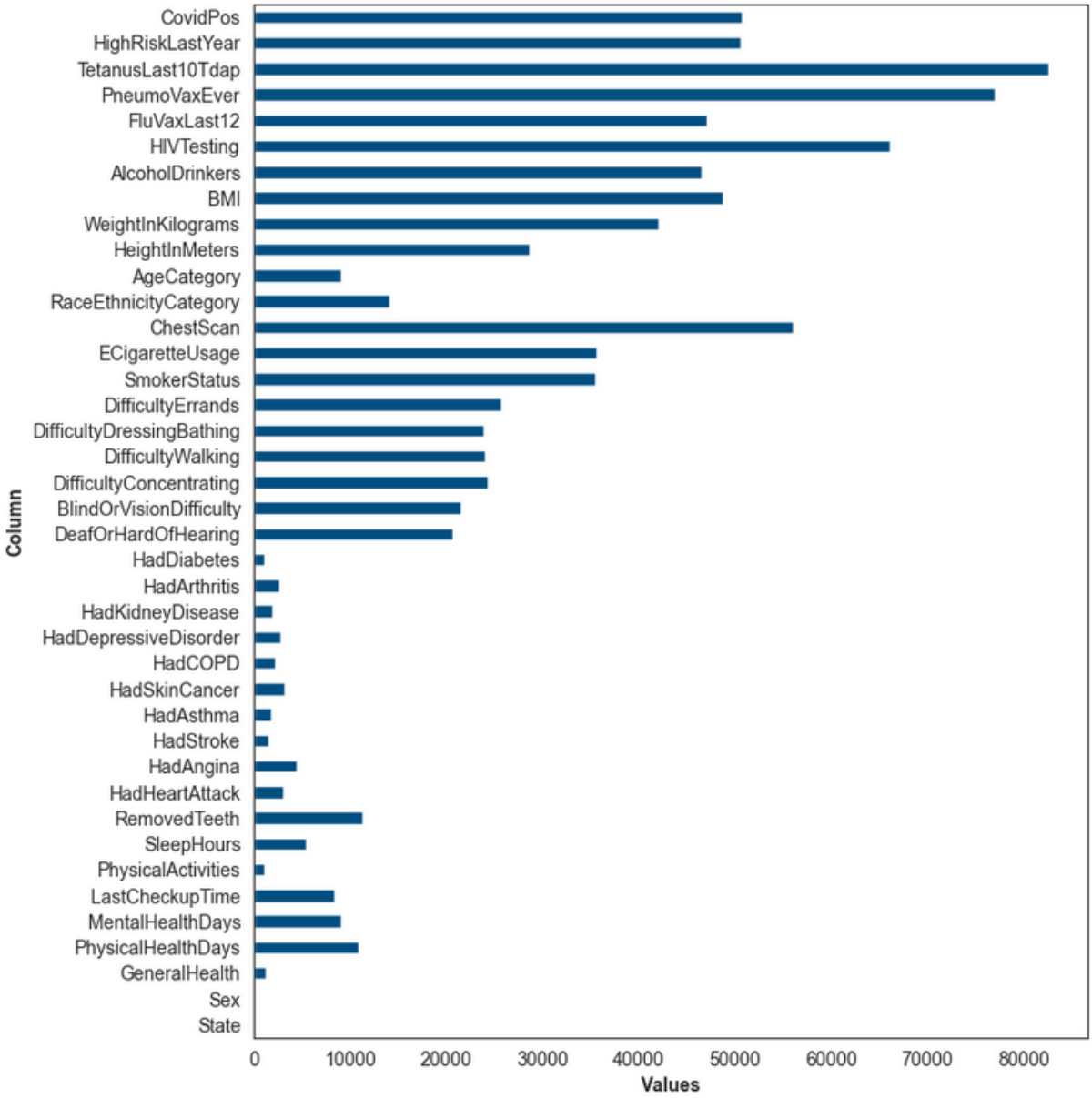
Tools Used:

- Analysis conducted using Python programming language and it's libraries: **pandas**, **matplotlib**, and **seaborn**.
- Pandas utilized for data manipulation and preprocessing.
- Matplotlib and seaborn employed for data visualization, aiding in trend identification and insight generation.

Methodology

Data Cleaning and Preprocessing:

- The dataset initially had 445132 records and 40 variables(columns).
- Handled null values and checked for redundant columns.

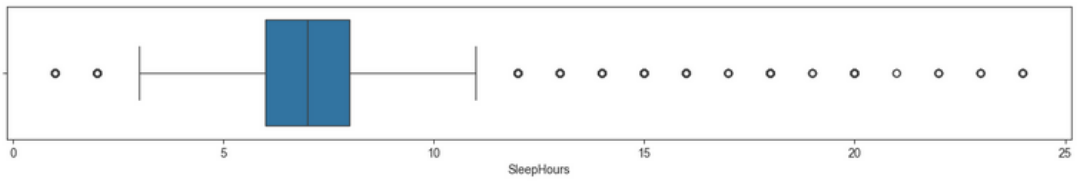


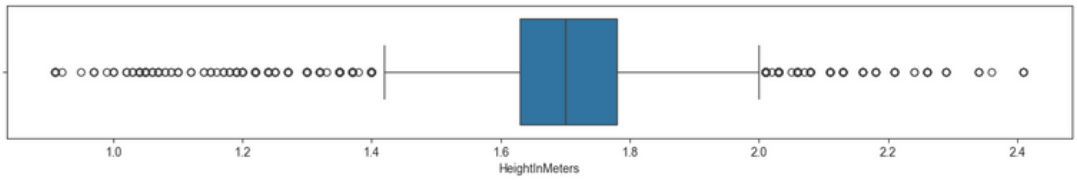
- There are columns which have large number of null values and also redundant, which means we don't need them in our analysis.
- Redundant columns were removed like **'HIVTesting'**, **'FluVaxLast12'**, **'PneumoVaxEver'**, **'TetanusLast10Tdap'**. Since they were also not the part of our scope.

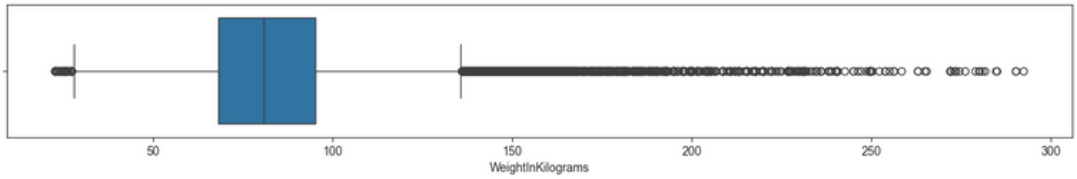
Methodology

Data Cleaning and Preprocessing (Continued):

- Dropped rows according to the null values in the **'HadHeartAttack'** column. As it is essential variable which indicates wether a person has heart disease or not.
- Not removing more null values, as different columns have null values on different places, so removing a row based on a null value of a specific column can lead to loss of useful values in other columns
- Identified unique values in all the columns.
- Some Values in the categorical columns were very long, so they were shortened which will help us in the analysis and don't take up much space in the charts. E.g: there was a value in HadDiabetes column "**No, pre-diabetes or borderline diabetes"** it was changed to **"Borderline"** only.
- Duplicate columns were dropped.
- Checked summary statistics like mean, median, percentiles, min, max, spread of the data and Identified outliers in numeric columns **PhysicalHealthDays**, **MentalHealthDays**, **SleepHours**, **HeightInMeters**, **WeightInKilograms**, **BMI**.
- There were anaomalous values in these columns, so those were removed, E.g: There were rows with more than 20 hours of sleep time, that is impossible. Same goes for other columns.

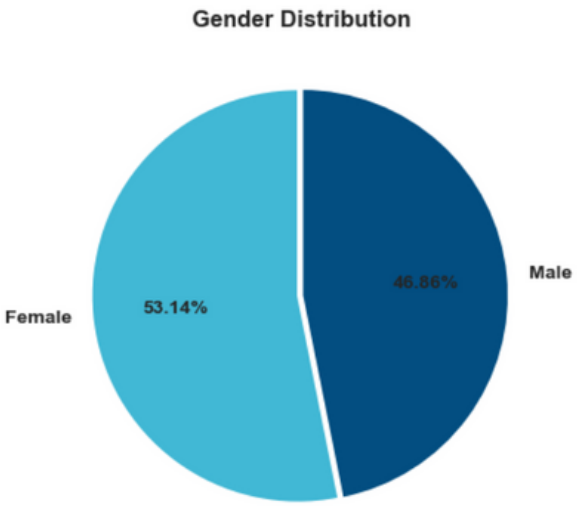




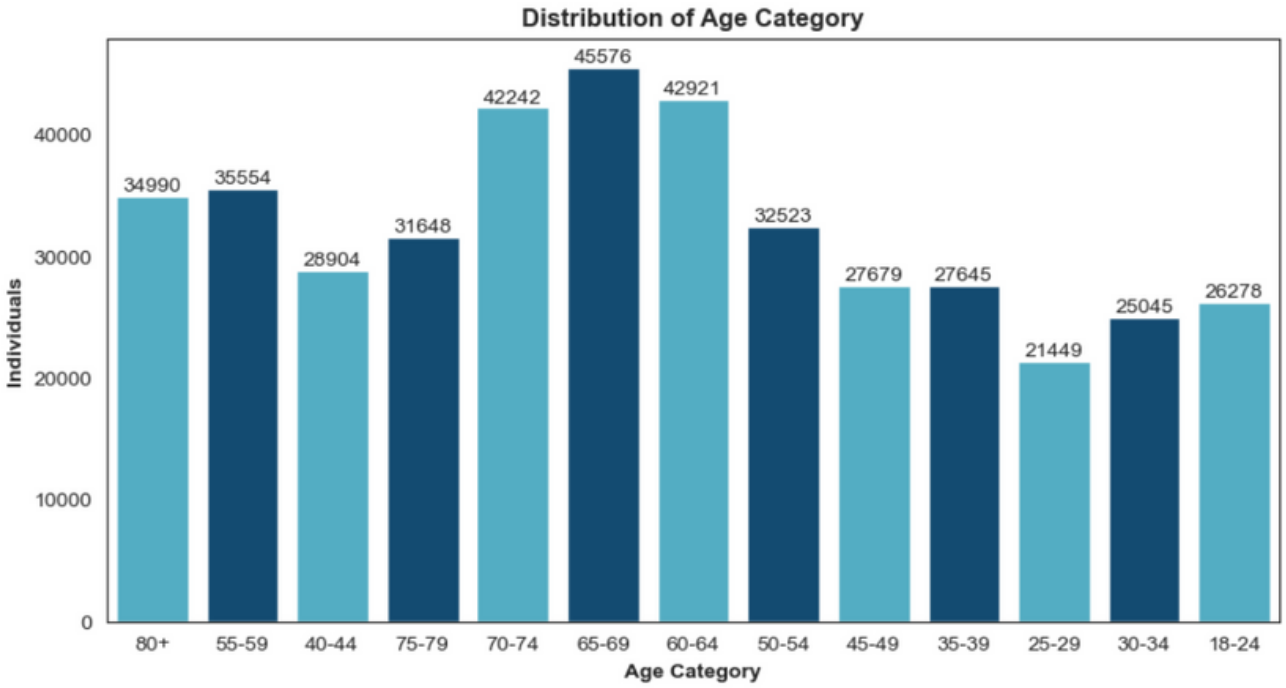


Demographic Variation & Risk Factor Patterns

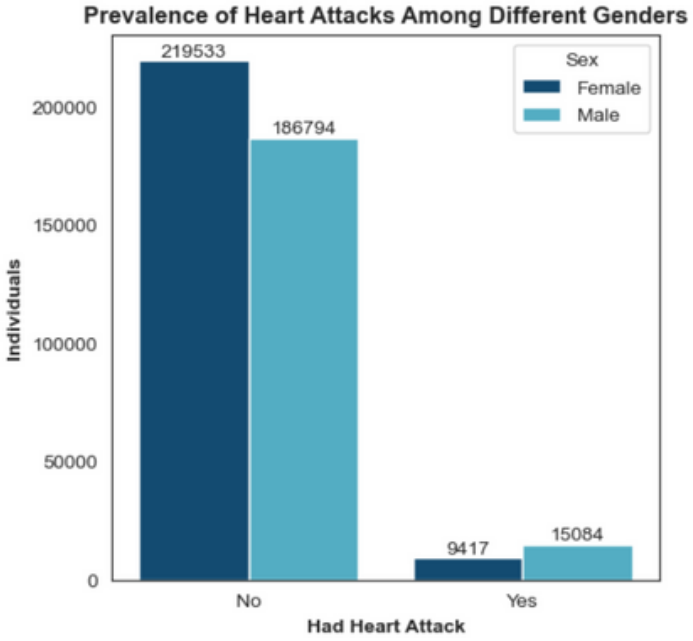
Understanding heart disease and its risk factors across demographics like sex, age, and race is crucial for addressing health disparities. Analyzing these variations helps identify high-risk groups and informs targeted interventions.



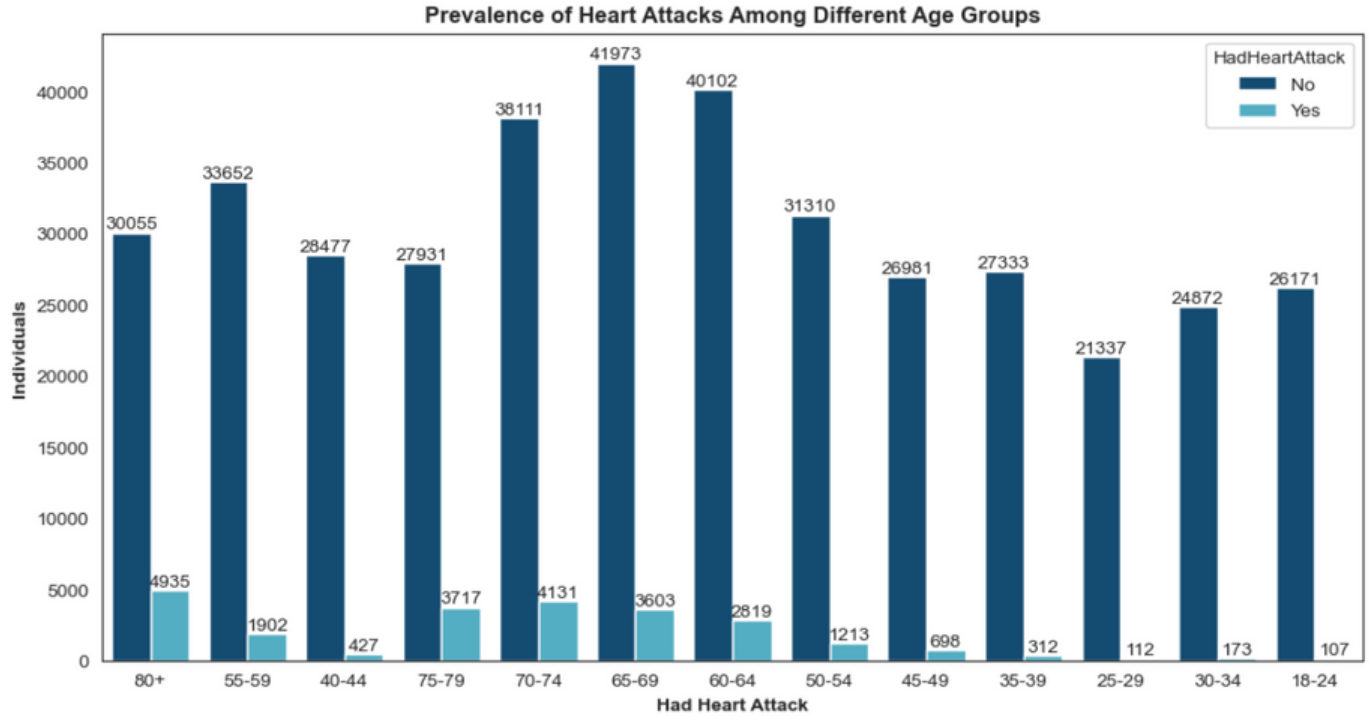
- **Females** outnumber males in the sampled population.
- The majority of individuals fall within the age range of **65-69** years.



Building on our understanding of demographics, let's explore the prevalence of heart disease among different genders and age groups.

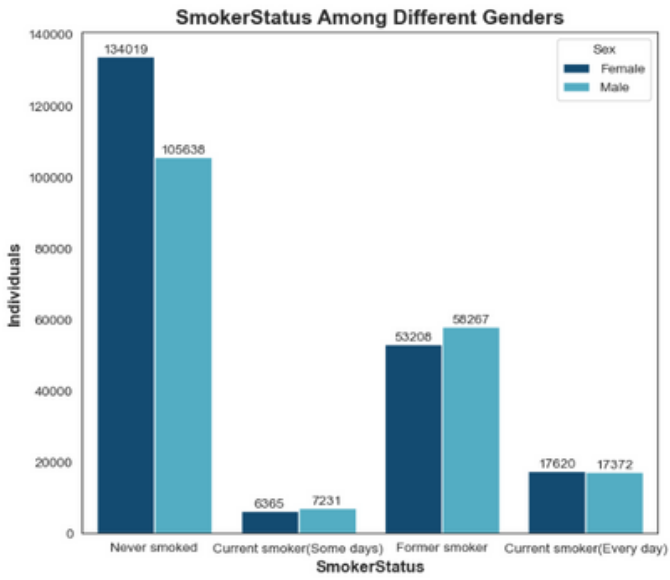


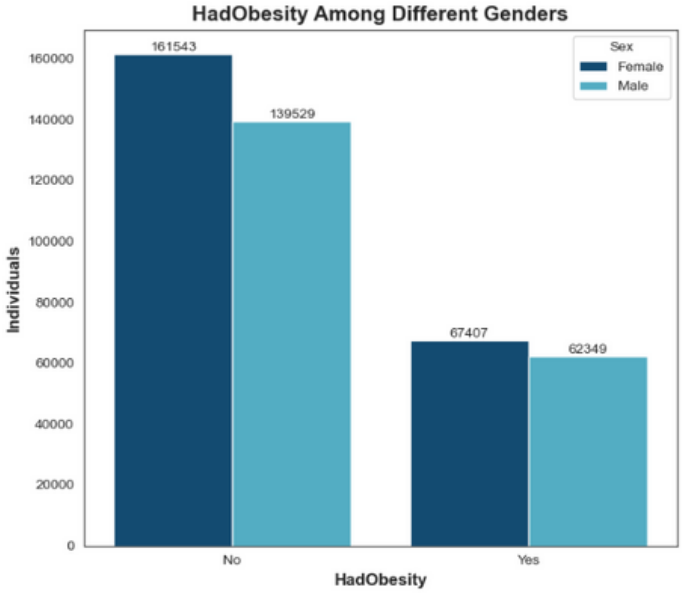
- Most individuals did not experience heart disease.
- Among those who did, **males** comprised the majority.

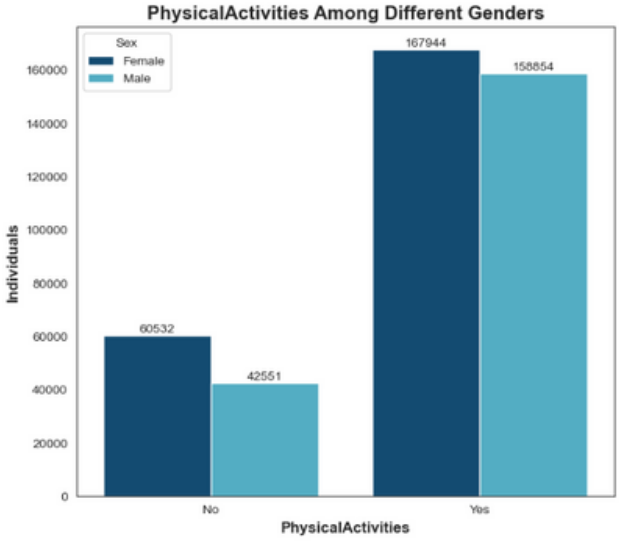


- Individuals above **80** years old exhibit a notable prevalence of heart diseases.
- Even though less common, some individuals between **18** and **24** also experience heart disease.

Expanding on our examination of demographics, we are now investigating the prevalence of various major risk factors among different genders and age groups.



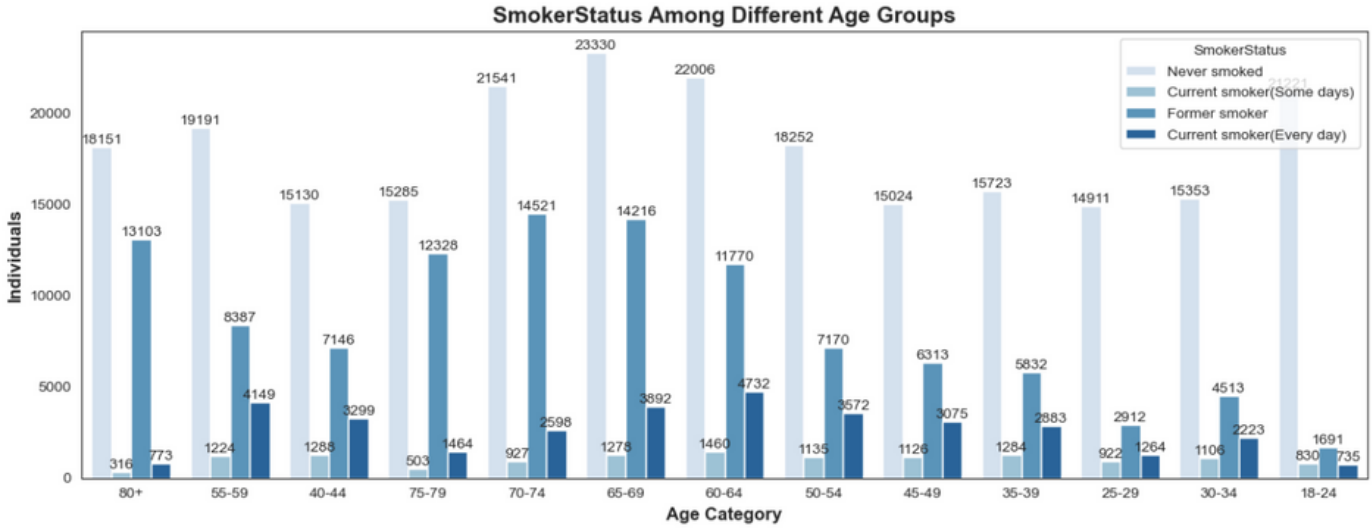




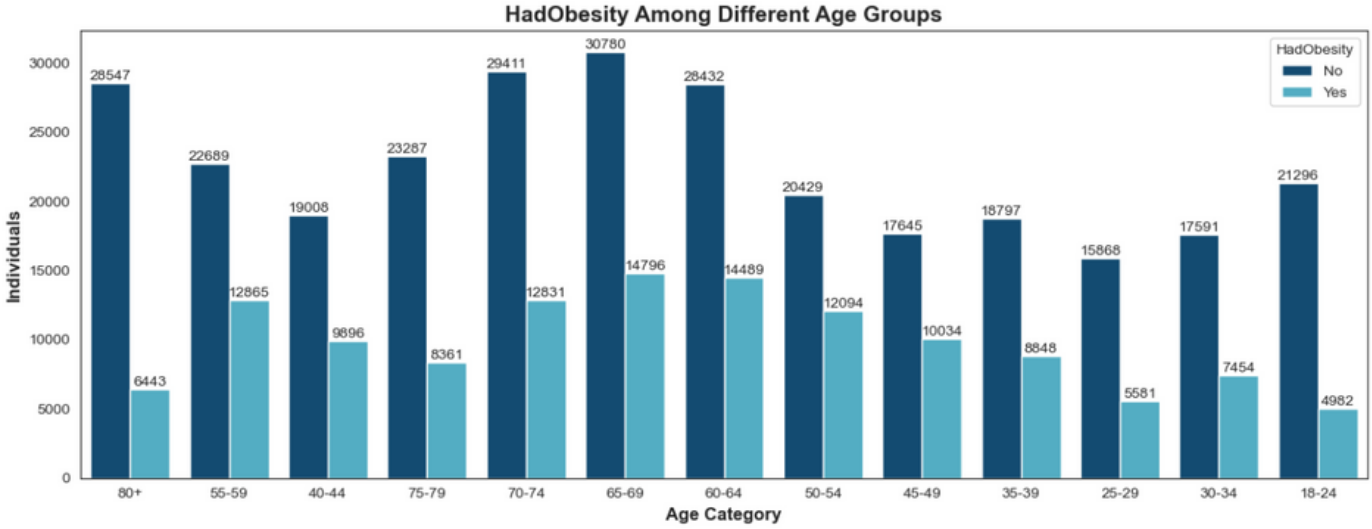
- The majority of individuals engage in physical activities, with **females** being the most prevalent among them.
- The majority of individuals have never smoked in their lives. However, **males** show a slight lead over females in the category of current smokers.
- Most people are not obese. However, among those who are, **females** are the most.

Some other Findings

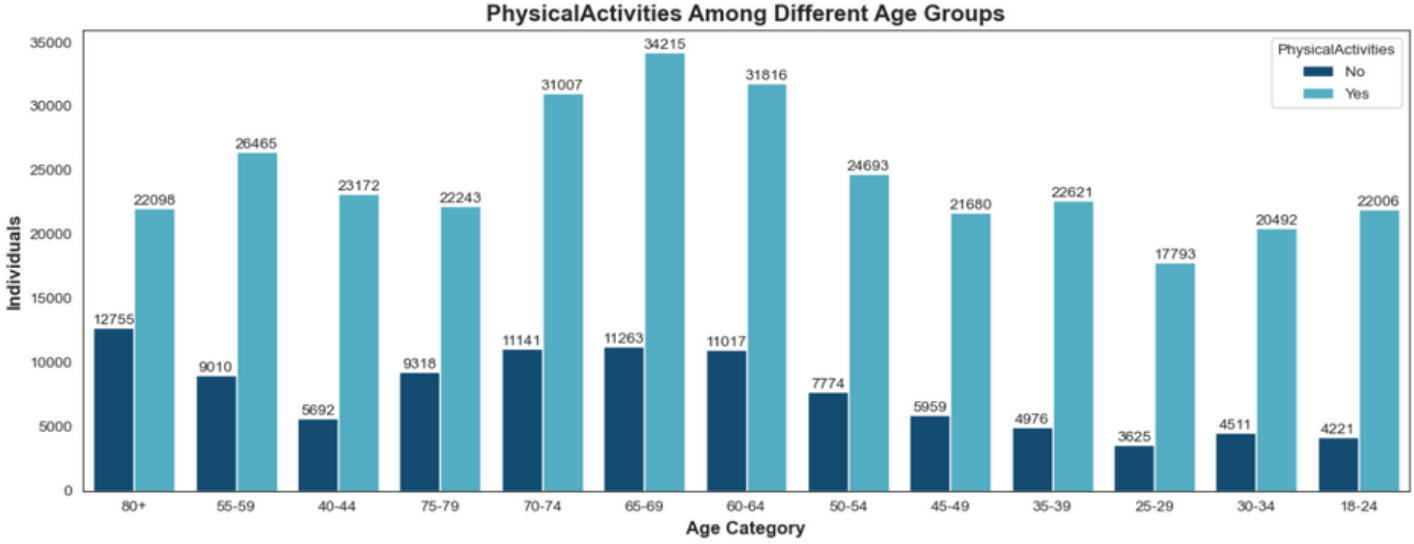
- The majority of individuals do not have diabetes. However, among those who do, females constitute the larger portion.
- Most people never use E-Cigarette. However, majority of the people who regularly use e-cigarettes are males.



- Most individuals aged between **60 and 64** engage in regular smoking, while those aged **70 to 74** comprise a larger proportion of former smokers.

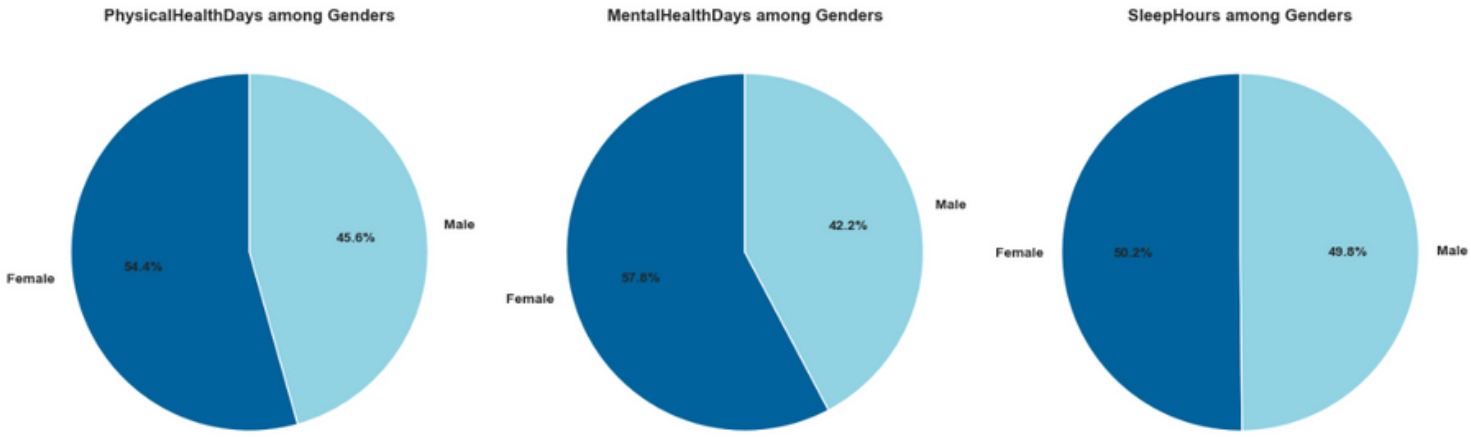


- People aged **65 to 69** have more obesity compared to other age groups. However, those between **18 and 24** don't have as much obesity.

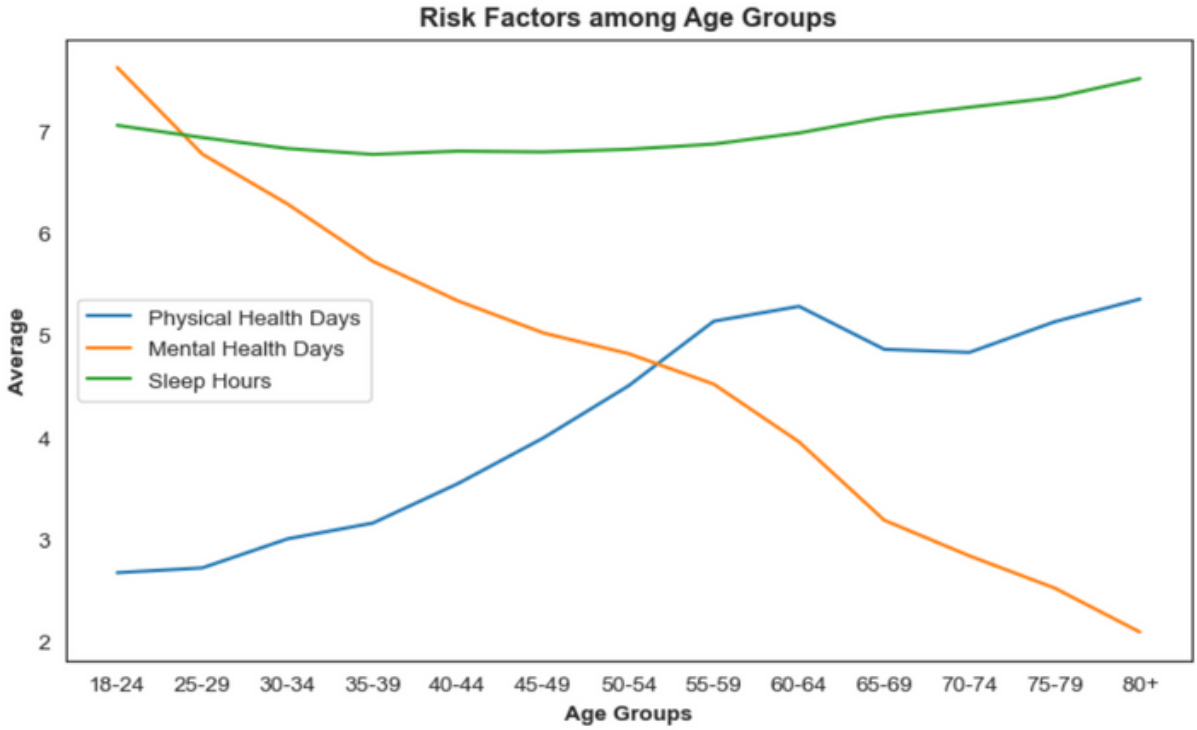


- Most individuals aged around **65-69** are physically active; however, those above **80** exhibit a larger portion of being physically inactive.

Now let's look at some other common risk factors among different genders and age groups, such as Physical Health Days, Mental Health Days, and Sleep Hours.



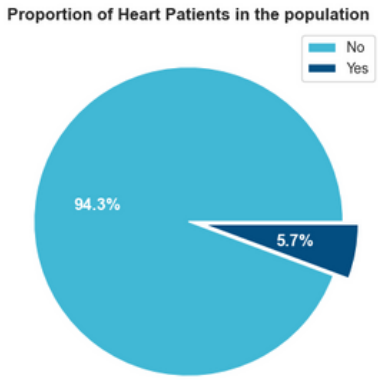
- **Females** experience a higher percentage of days with poor physical health and poor mental health compared to males. This means that, on average, females report more days of poor physical and mental health in the past **30** days than males.
- The average number of **sleep hours** is quite similar between males and females.

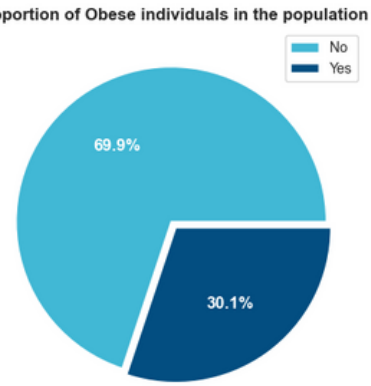


- As people get **older**, they experience more days with poor physical health in the past 30 days. However, as people age, they experience **fewer** days with poor mental health in the past 30 days.
- Additionally, older individuals, particularly those aged **65 to 80** and above, tend to sleep slightly more hours.

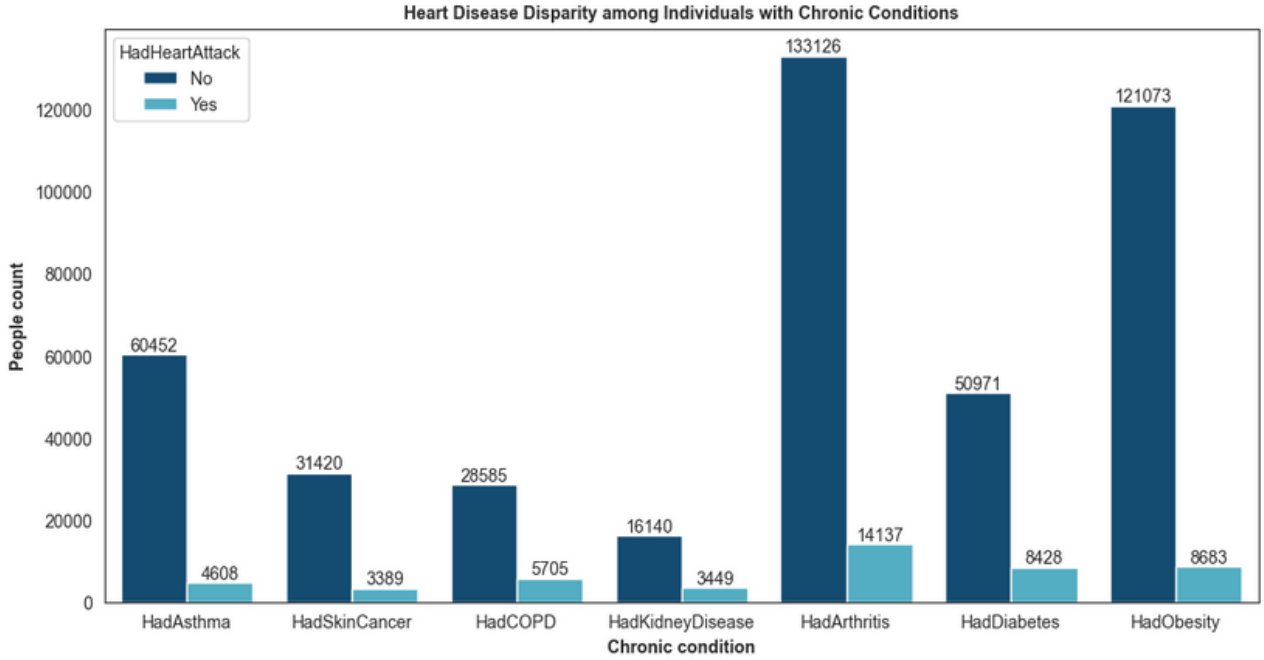
Impact of Chronic Conditions

Chronic conditions like diabetes, high blood pressure, obesity, and lung issues can greatly increase the risk of heart disease. Recognizing this connection is important for preventing and managing heart health risks. Some initial findings are:

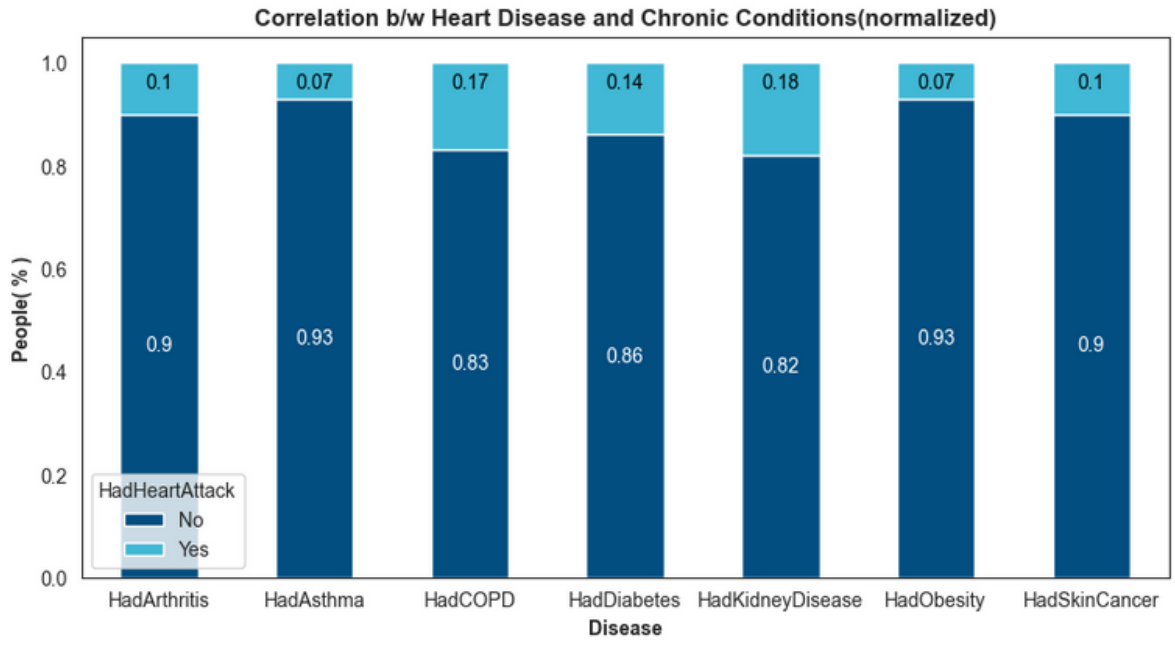




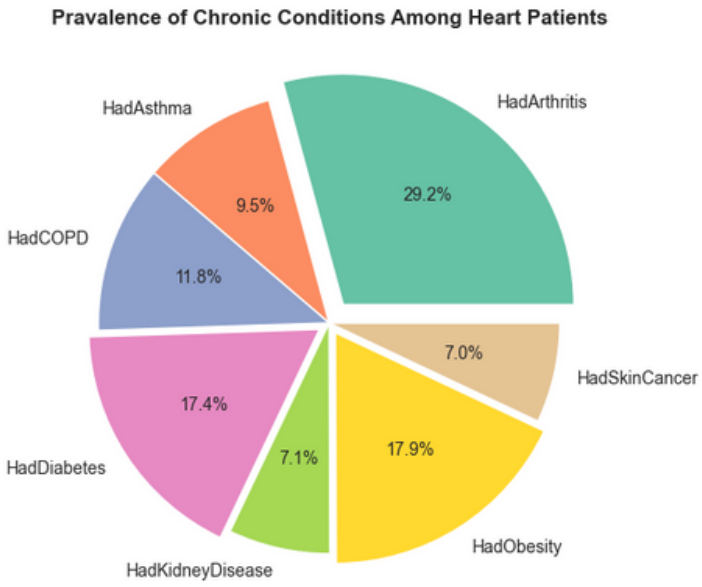
- Approx. 6% of the sample data population are heart patients.
- Almost 30% of the people are obese i.e their BMI is greater than 30.



- Here you can see that many people reported heart disease, previously had problems of mostly arthritis, diabetes and obesity.



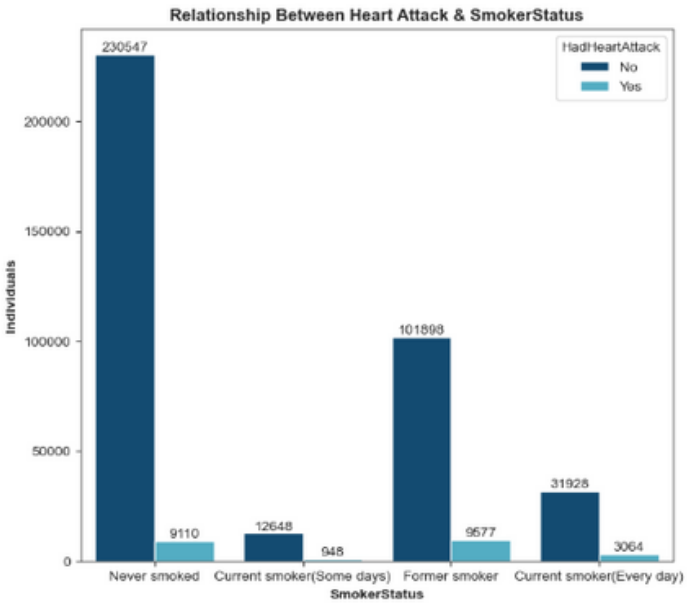
- Certain chronic conditions are not so strongly associated with cardiovascular health disparities.
- Individuals previously having chronic diseases such as COPD(Chronic obstructive pulmonary disease), diabetes, kidney diseases are more likely to have heart diseases as **17%**, **14%** and **18%** respectively of them are also heart patients.
- Other conditions like **arthritis(10%)**, **asthma(7%)**, skin **cancer(10%)**, **obesity(7%)**, also have a chance of developing a heart disease.

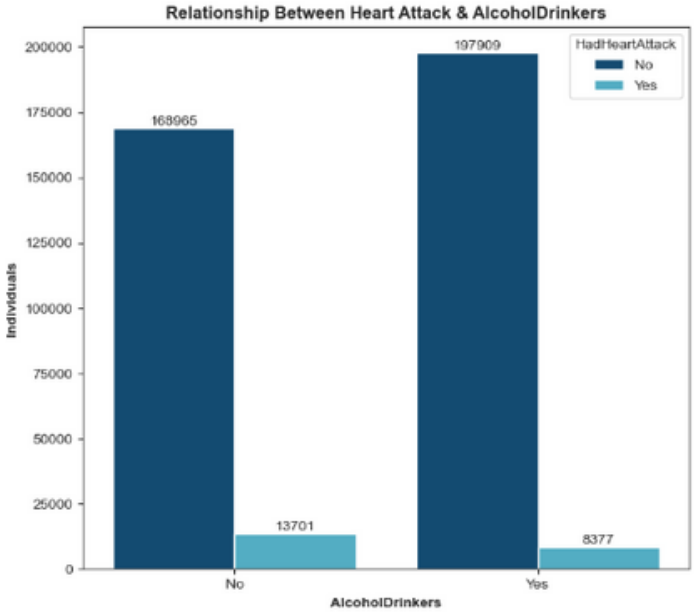


- Considering overall heart patients with chronic conditions previously, Nearly one third one them had arthiris before, around 17-18% are diabetic, other 18% are obese.

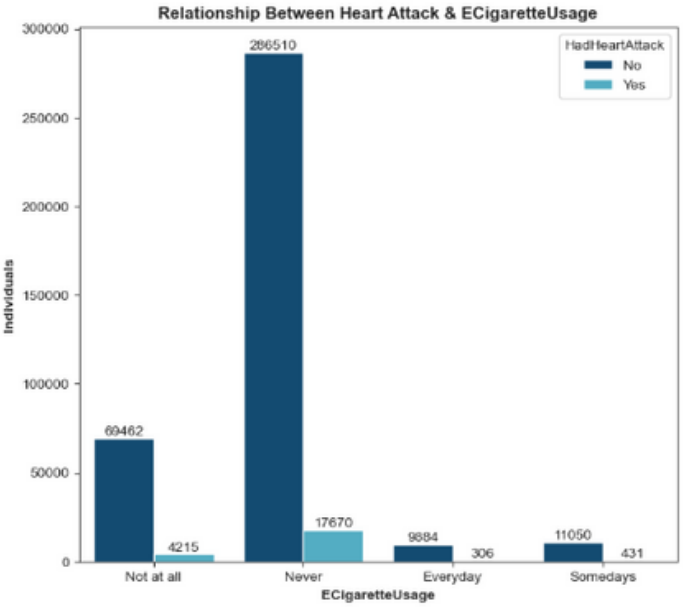
Behavioral Correlations

This section explores how health behaviors like physical activity, sleep, smoking and alcohol consumption relate to heart disease prevalence. We'll identify key behavioral factors contributing to cardiovascular health disparities.

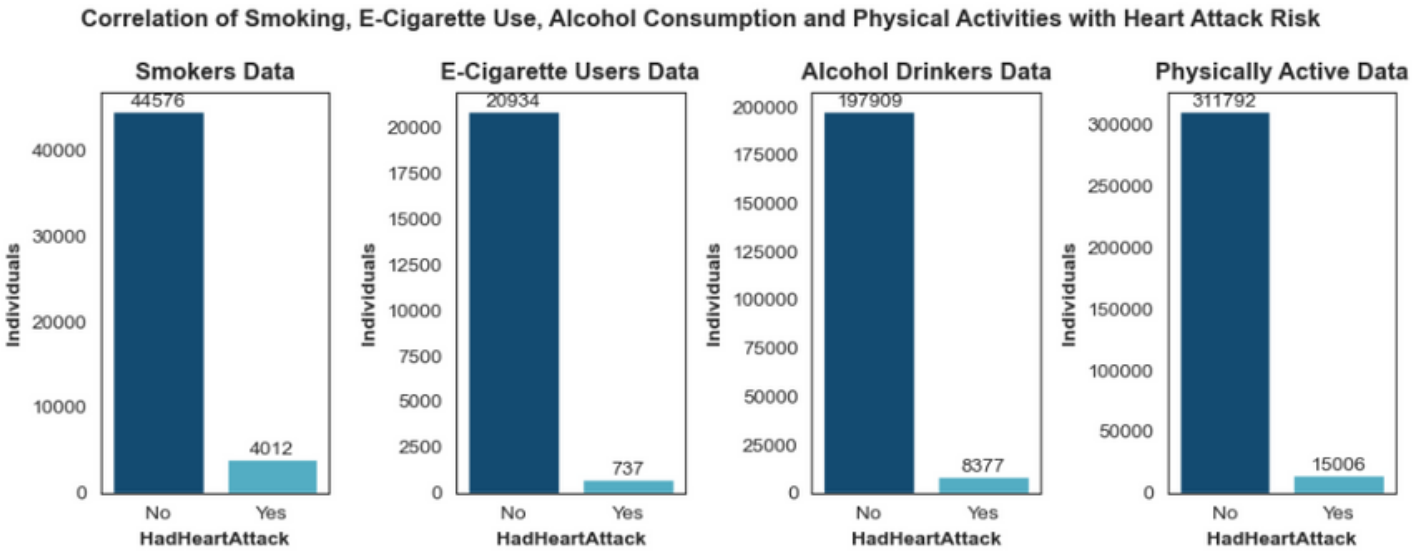


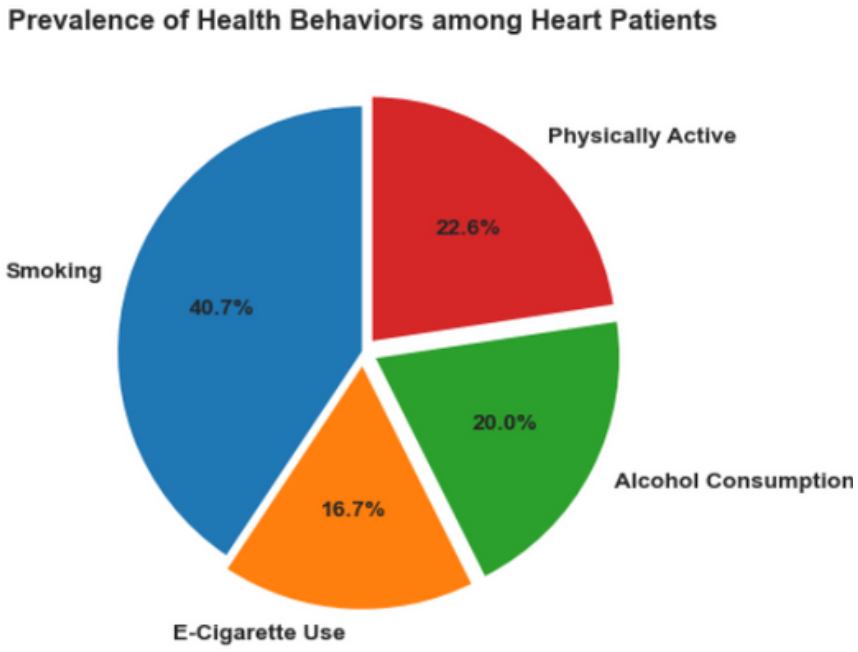


- Individuals who used to **smoke** in the past are often affected by heart attacks.
- Additionally, it's noteworthy that individuals who have never used **e-cigarettes** and do not consume **alcohol** also experience heart attacks, which is quite surprising.

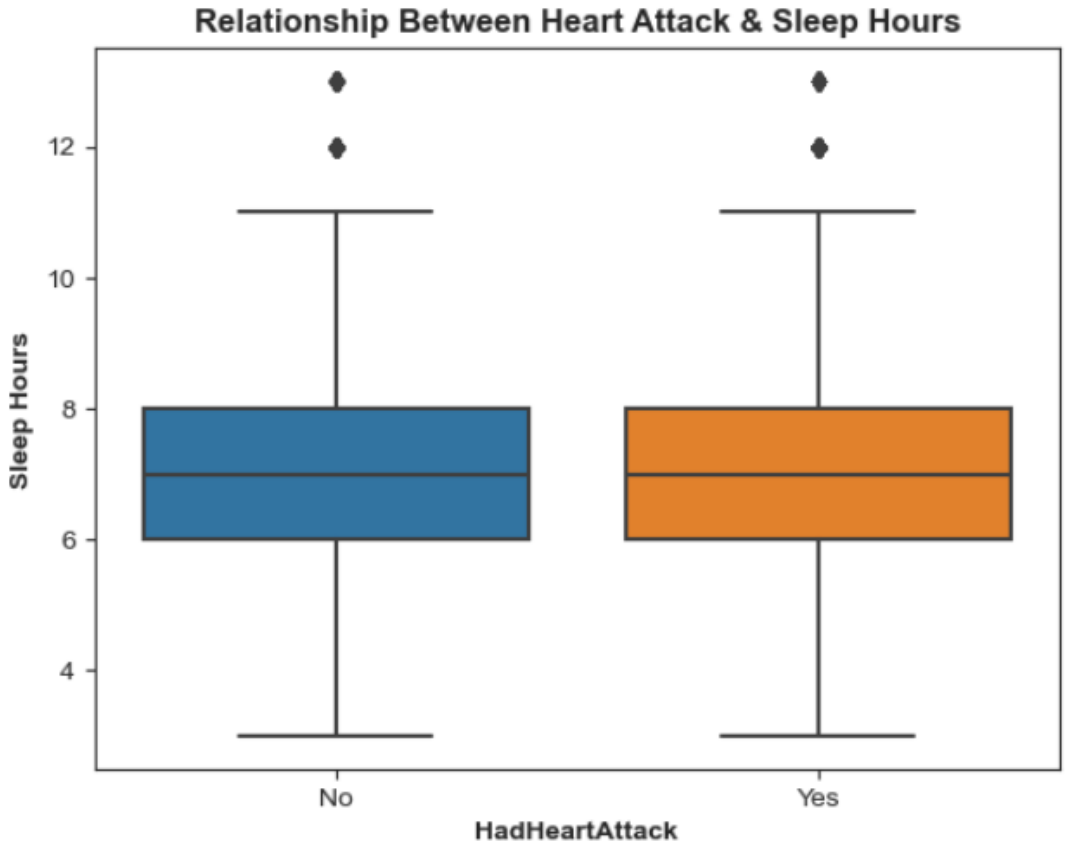


Now, we will investigate how the presence of these factors is connected to the prevalence of heart disease.

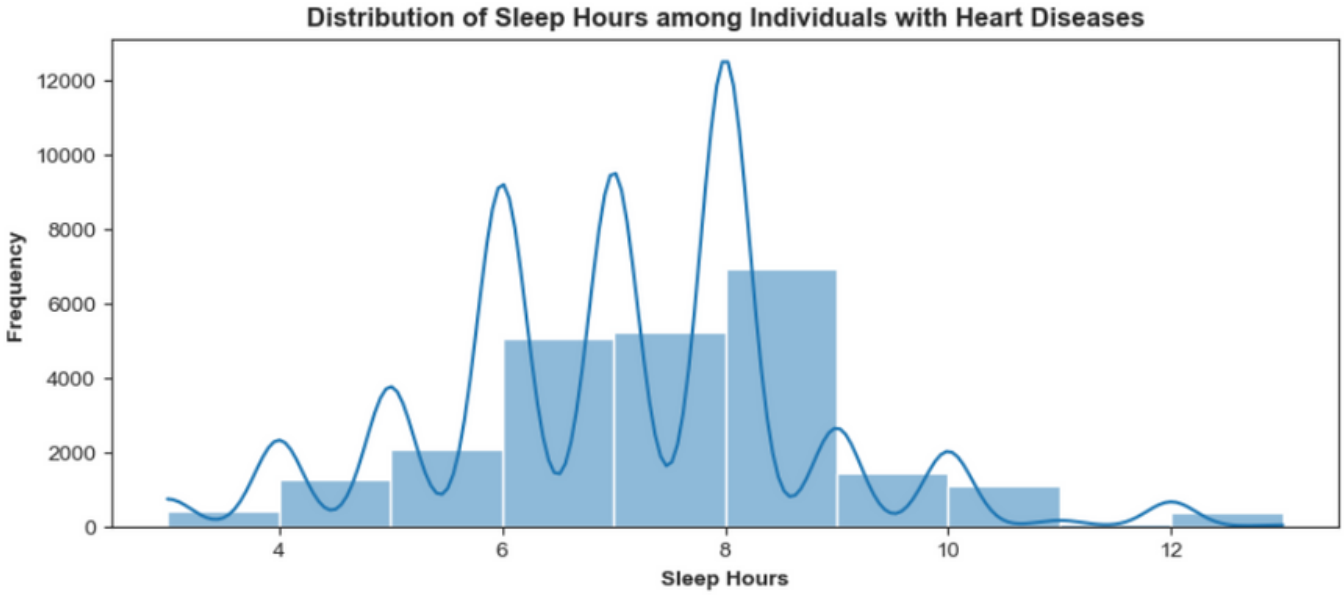




- Considering the overall population of heart patients, approximately **two-fifths** are smokers, while around **one-fourth** engage in physical activity.
- Additionally, **20%** consume alcohol, and **16.7%** use e-cigarettes.



The box plot illustrates that the distribution of sleep hours is **similar** between individuals who have experienced a heart attack and those who have not. So, now we will explore the sleep hours among those who have heart diseases

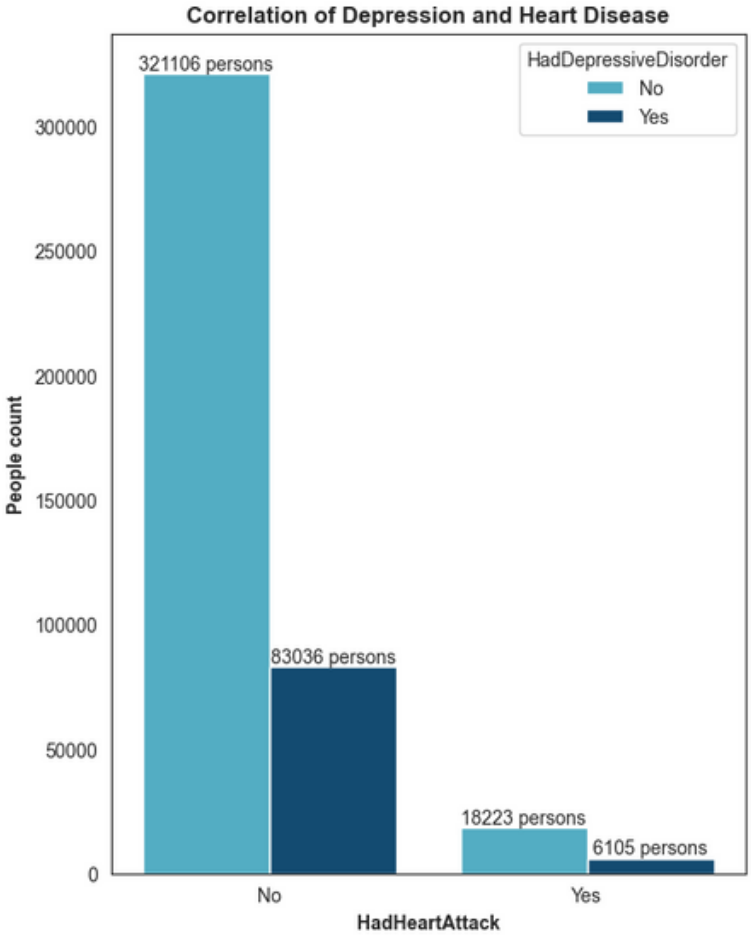


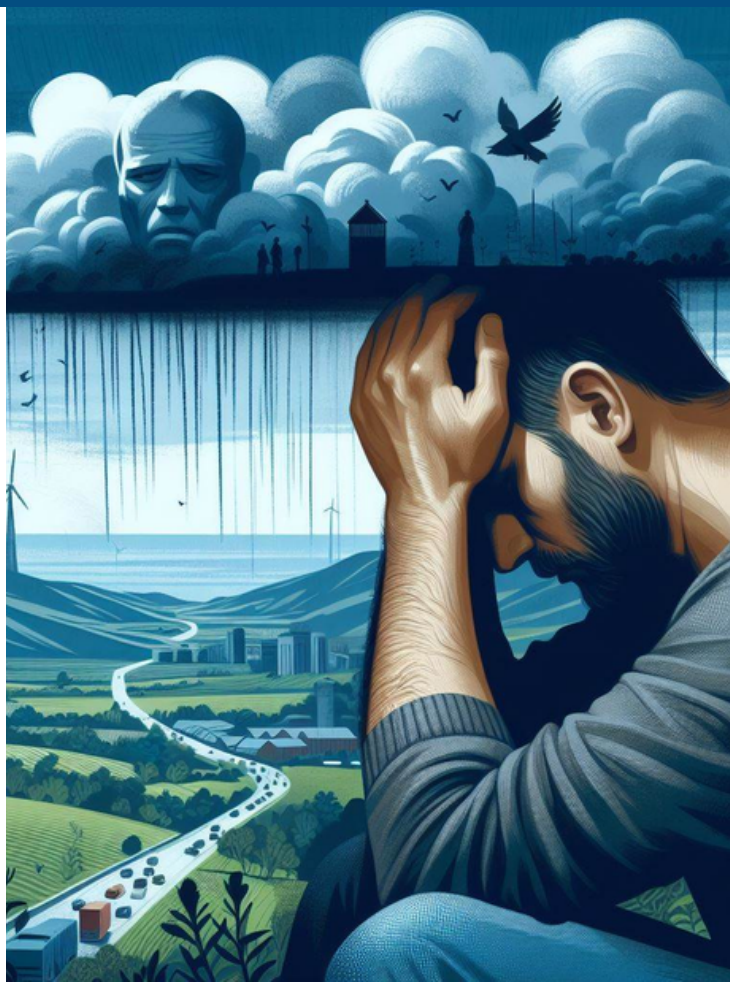
- The data seems to be slightly **skewed to the right**.
- It suggests that most people with heart diseases tend to sleep for about **6 to 8 hours.**

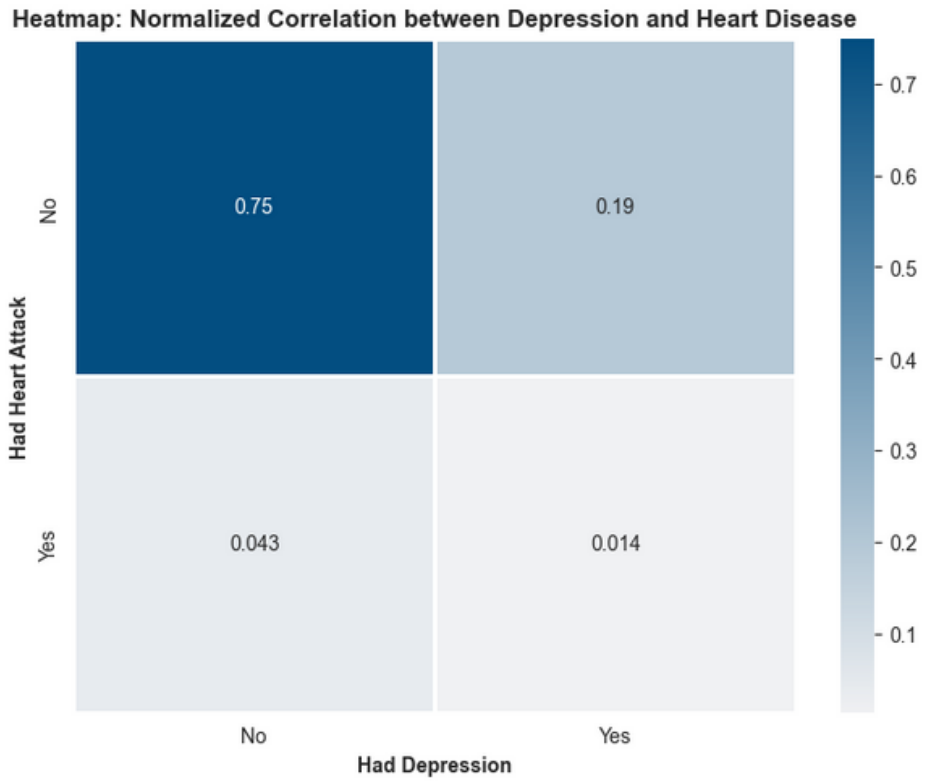
Effect of Mental Health on Heart Disease

In this section, we will examine the relationship between mental health and heart disease, investigating how psychological well-being impacts cardiovascular health. Mental health conditions like stress, anxiety, and depression can significantly influence the risk of heart disease by contributing to physiological changes such as increased blood pressure and inflammation. Understanding this connection is crucial for developing effective prevention and management strategies to promote heart health

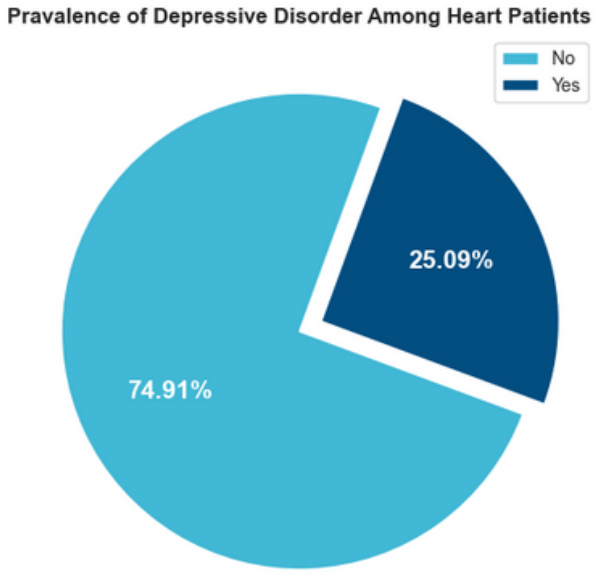
After analysing the poplulation data, we came to know that **20%** or **1 out of 5** people are suffering from depressive disorder. Among people having heart problems, a significant big portion of 6000+ have depressive disorders also.



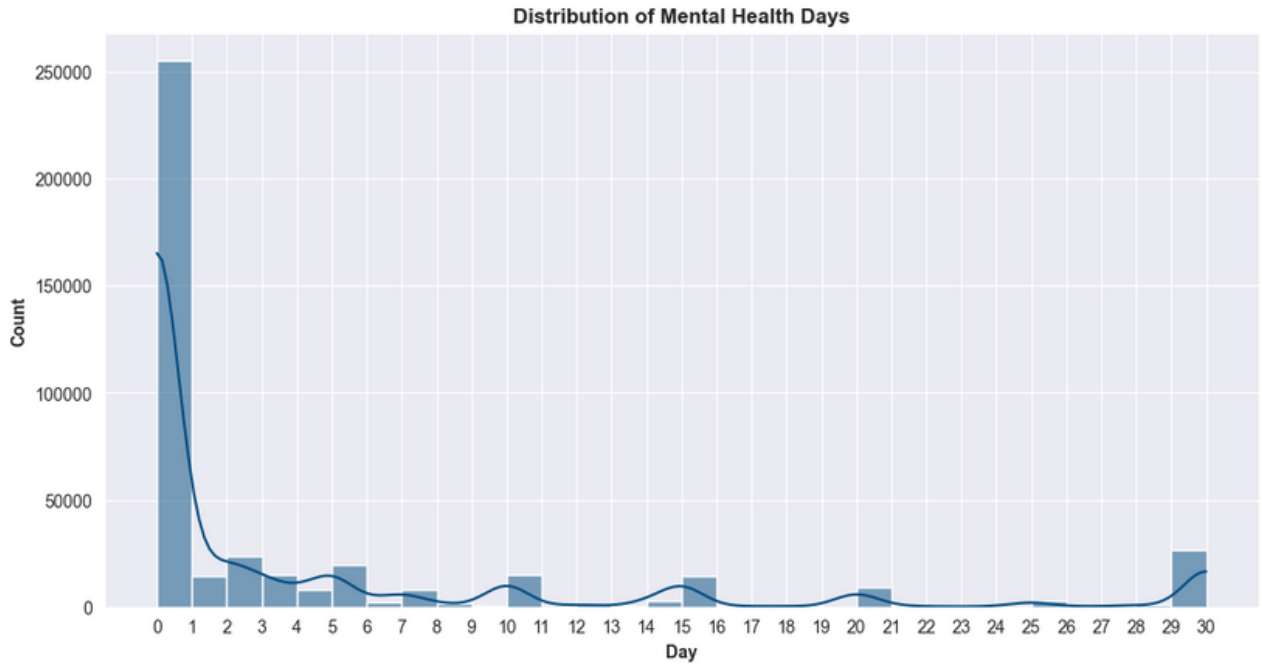




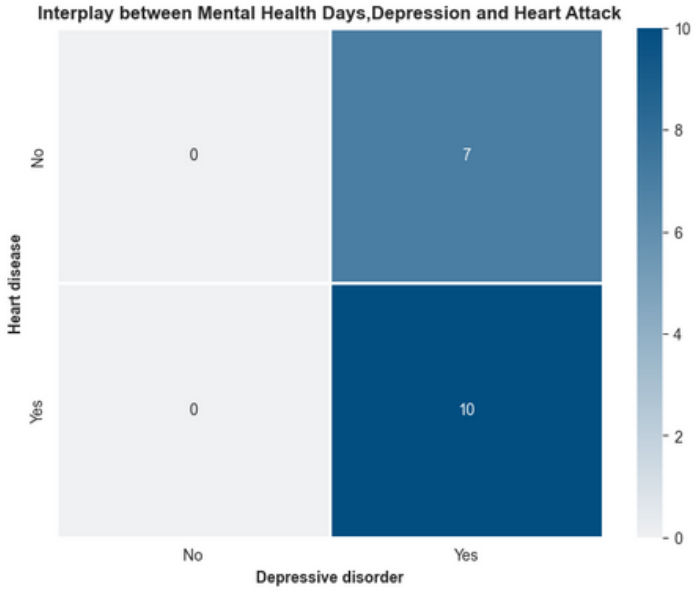
- Mostly **75%** people donot have depressive disorder and also have a good heart health.
- **19%** people are suffering from depression but also don't have a heart disease.
- Around **4%** people don't have depressive disorder but facing heart disease.
- Very few people i.e around **1.5%** have faced depressive disorder before developing heart problems.



- If we consider all the heart patients, then **approx. 25%** or **one out of four people** are facing depressive disorders.

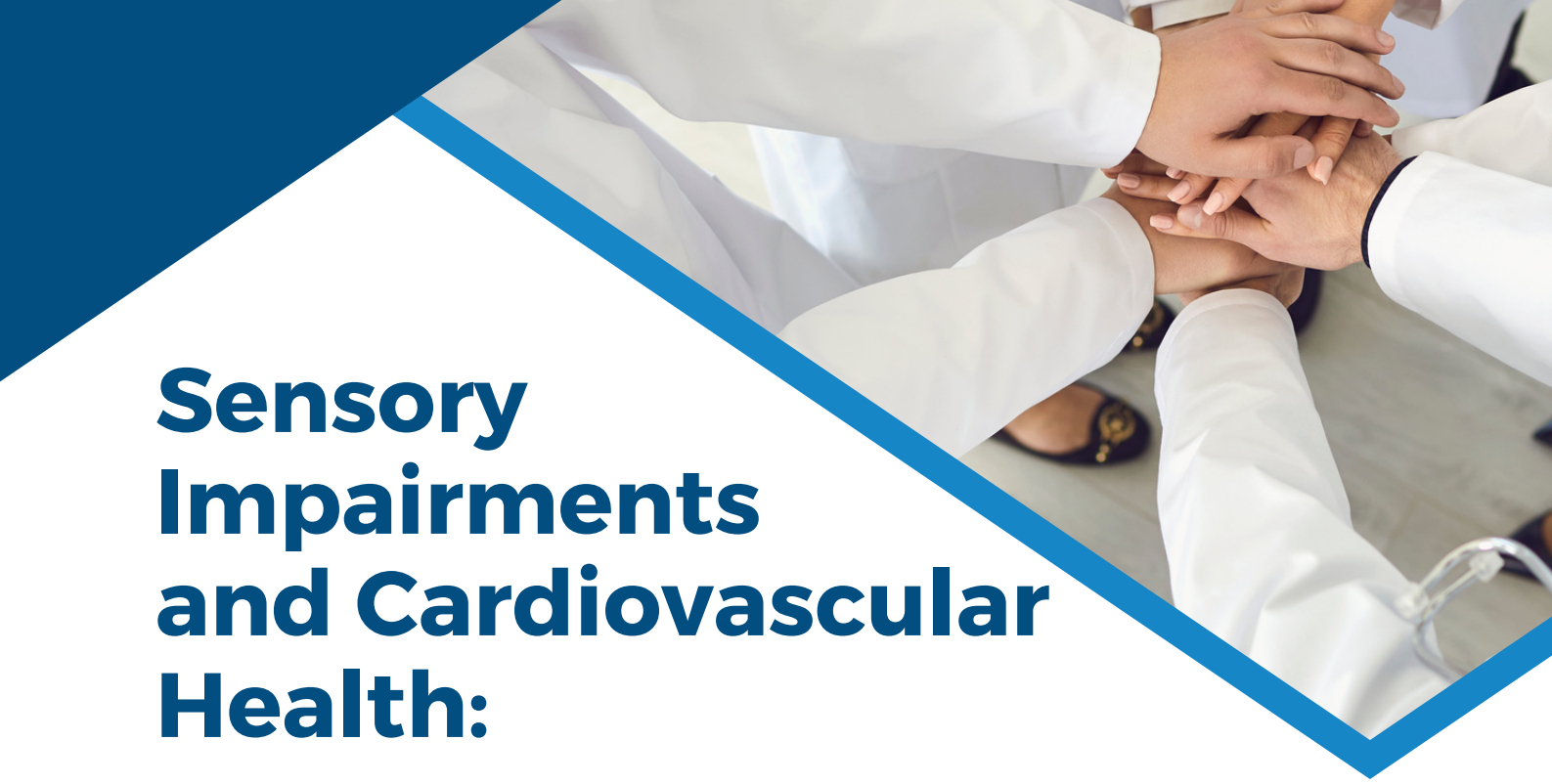


- The data seems to be **highly skewed to the right**.
- Most of the people i.e more than **250,000**, didn't face any issue related to mental health in past 30 days.
- There are also people who are still facing mental health problems from the **past 30 days**.



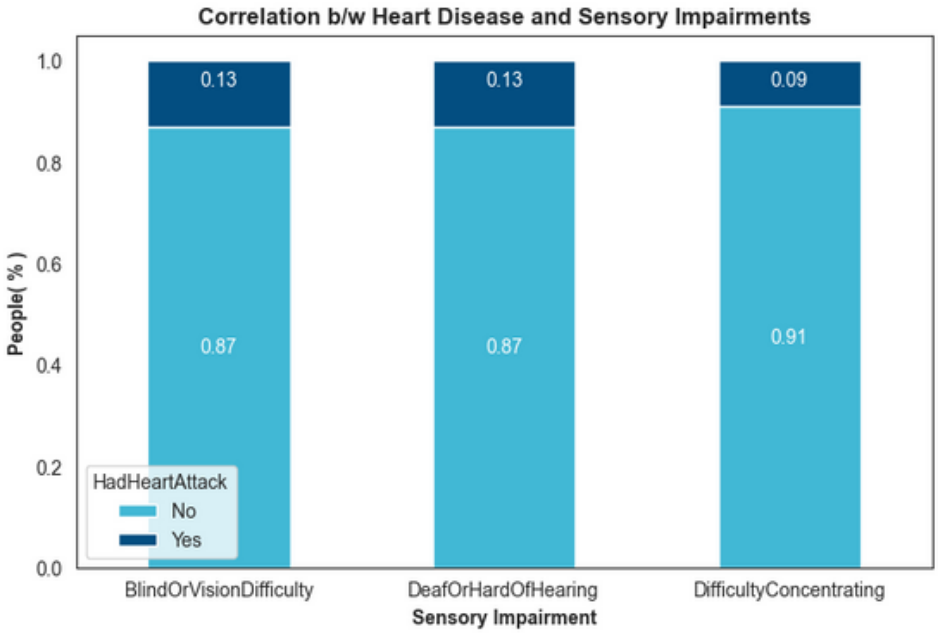
*"In datasets with **highly skewed distributions**, particularly when most values cluster towards the left with a long tail extending to the right, the **median** serves as the most suitable **measure of central tendency**. Unlike the mean, which can be heavily influenced by outliers, the median remains robust, providing a more accurate depiction of the typical value within the dataset."*

- Individuals with both depressive disorders and heart problems tend to experience more bad mental health days, with a **median** of **10 days**.
- Whereas, those with depressive disorders alone typically report **a week** of bad mental health.

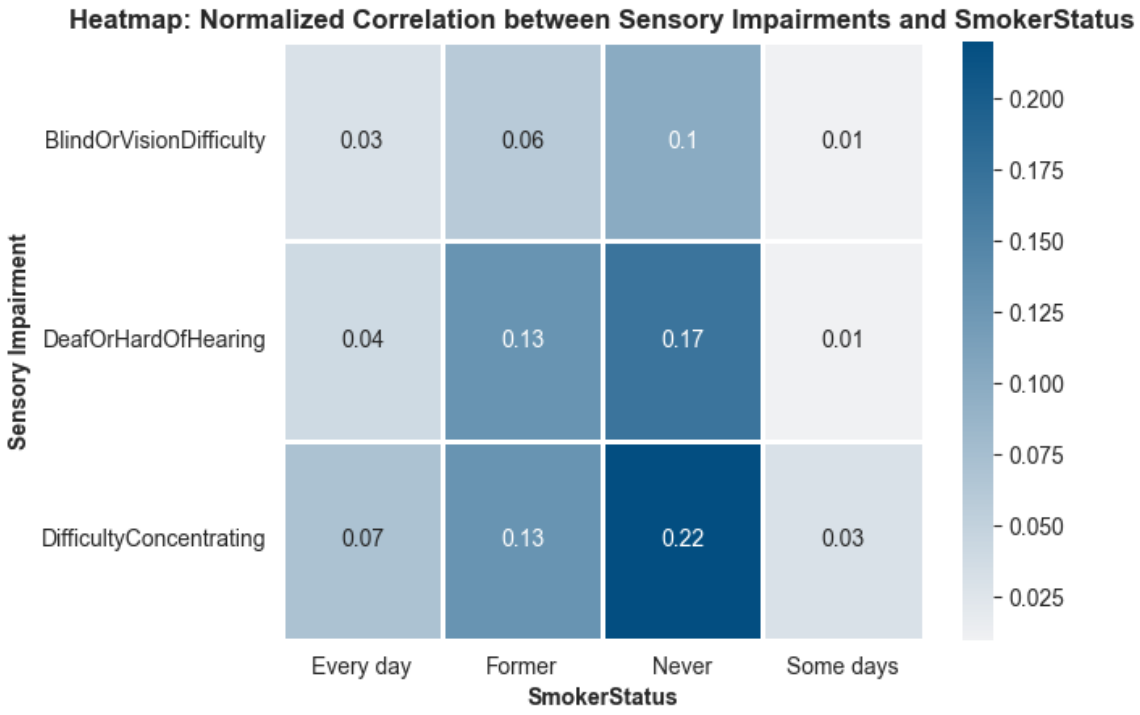


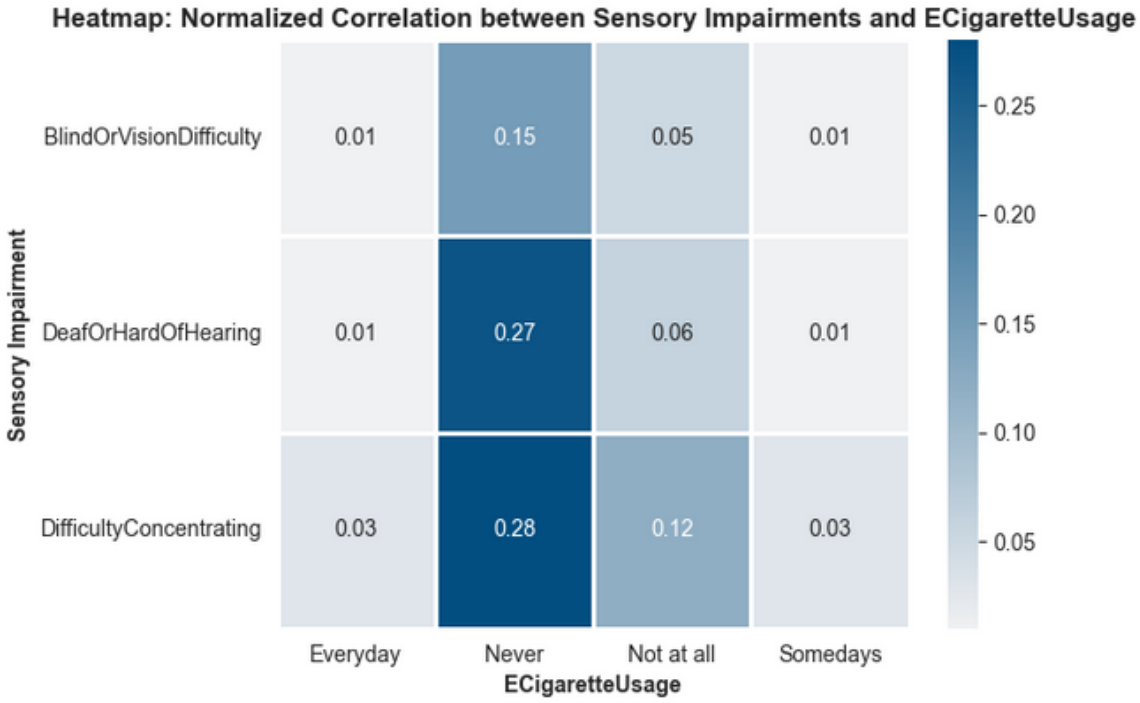
Sensory Impairments and Cardiovascular Health:

The section is data analysis report delves into the relationship between sensory impairments and cardiovascular health, focusing on two key aspects. Firstly, it examines the impact of sensory impairments, such as deafness and vision difficulties, on heart disease prevalence and associated risk factors. By analyzing trends and correlations within the dataset, the report aims to uncover whether individuals with sensory impairments are at a heightened risk of heart disease and its contributing factors.

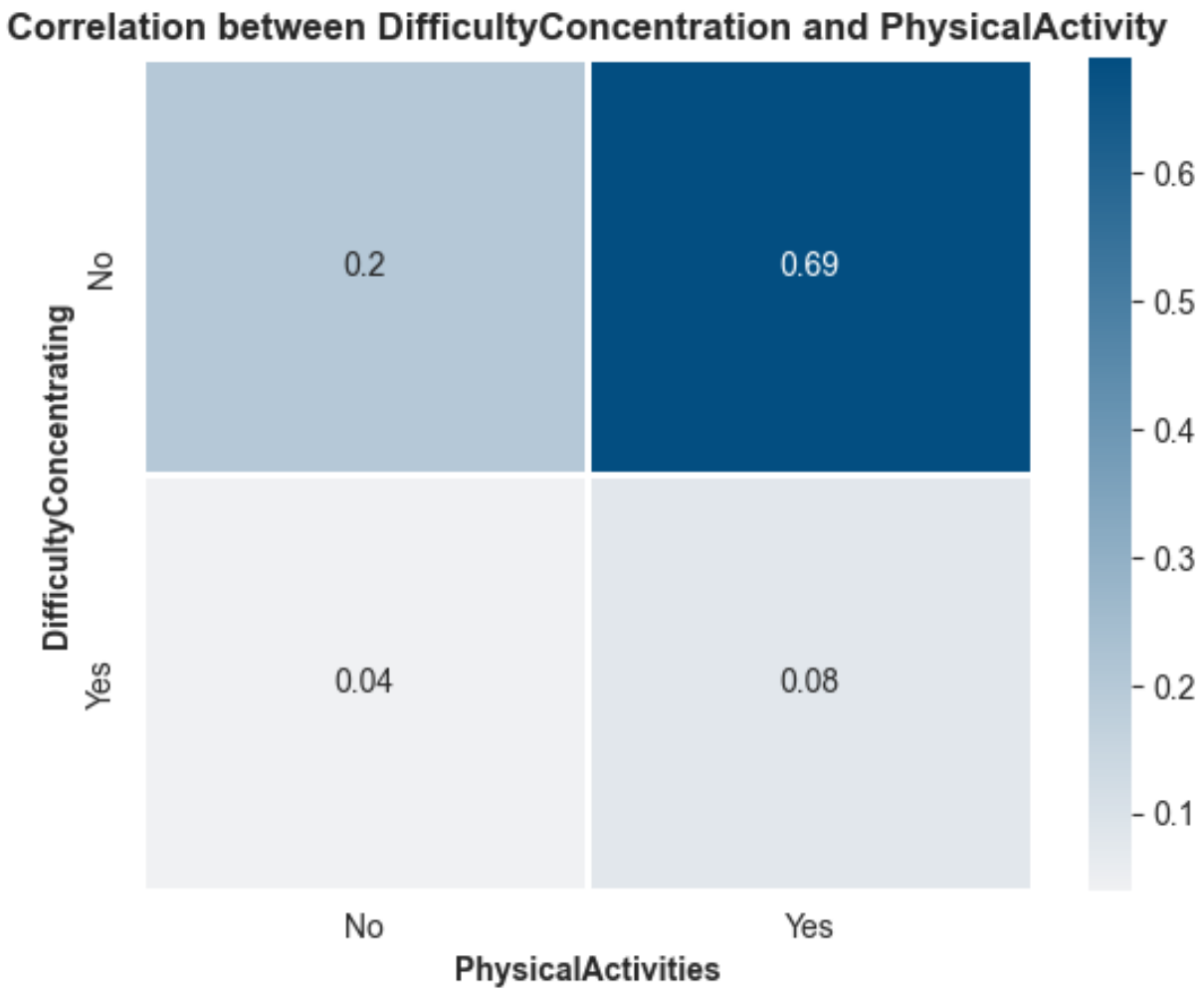


- **13%** of blind and deaf people are heart patients.
- **9%** people having are difficulty in concentration facing heart disease.
- Sensory impairment doesn't have any siginifact impact on heart disease pravalence.





- Fewer individuals report **difficulty concentrating** among **smokers** compared to **non-smokers**.
- **Smoking** may improve **concentration** and attention due to nicotine's effect on the body. **Nicotine** withdrawal symptoms can lead to increased concentration after smoking.
- People with blindness and deafness are less likely to smoke, with only around **1%-4%** of individuals in these groups reporting smoking habits.
- Same readings for e-ciggarete and vape users.

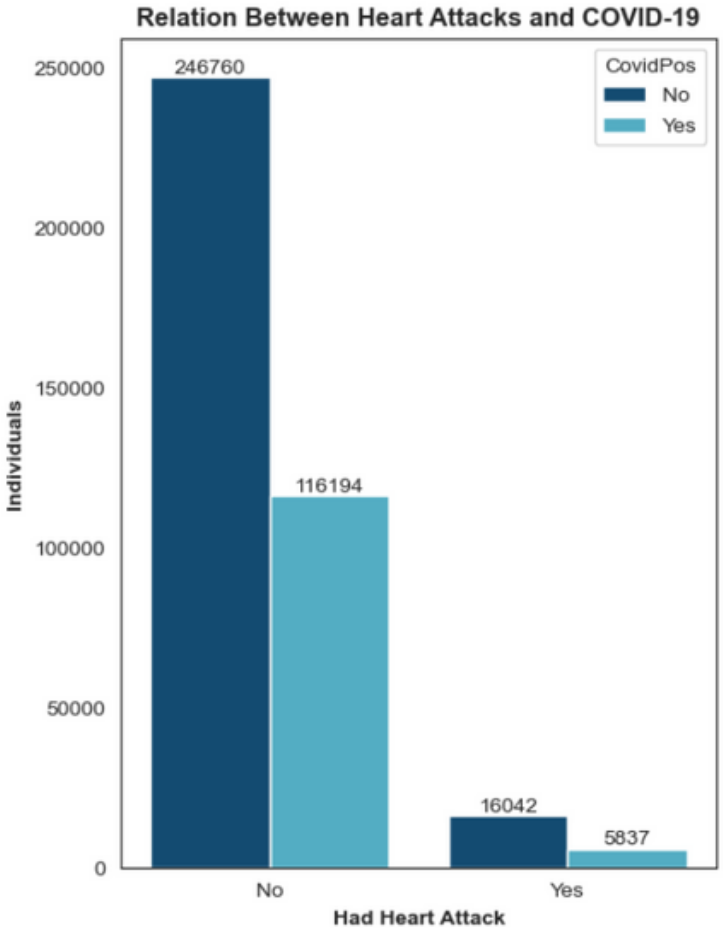


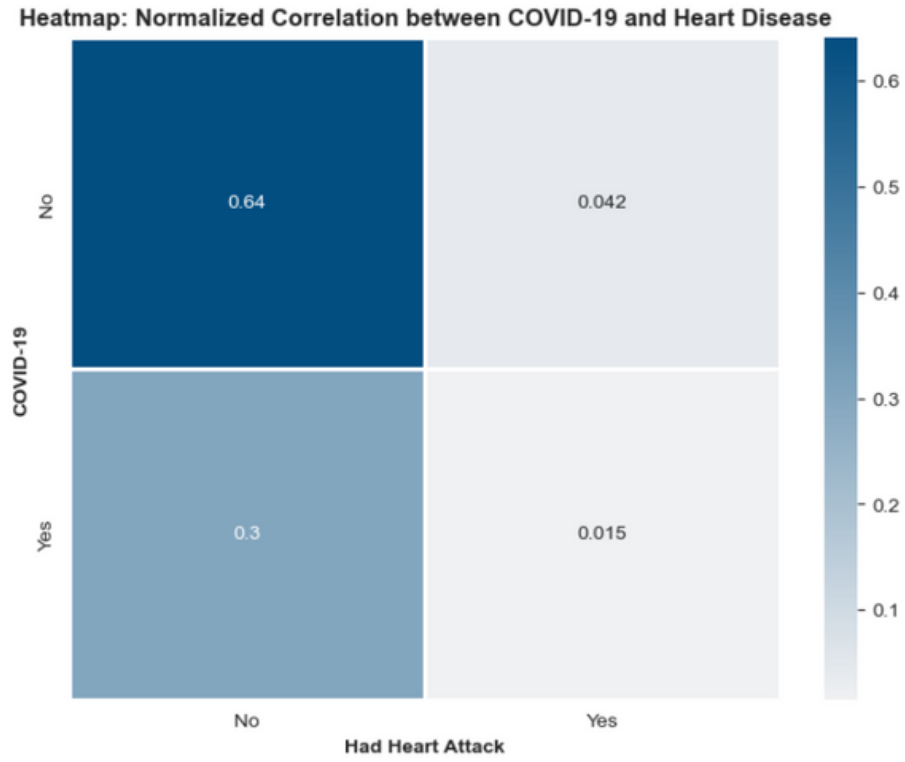
- **70%** people are physically active and donot have any difficulty in concentration.
- When you do any type of intense physical activity, it causes blood to flow to the brain. This in turn fires up your neurones and promotes cell growth, particularly in the hippocampus. This means that physical exercise can improve your concentration and help you focus

Impact of COVID-19

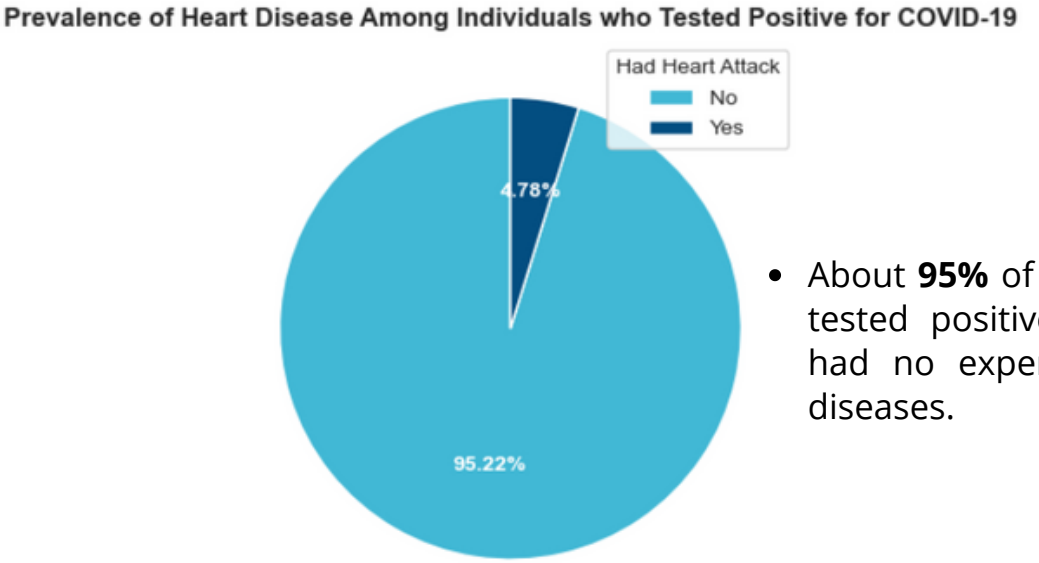
This section investigates the impact of COVID-19 on cardiovascular health, examining its potential association with heart disease prevalence. By exploring if individuals who tested positive for COVID-19 demonstrate distinctive patterns in cardiovascular health, we aim to shed light on the broader implications of the pandemic on heart health.

- Upon reviewing the population data, we found that around **one-third**, or **32%**, had contracted COVID-19.
- Within the demographic of individuals experiencing heart issues, a considerable number, exceeding **5500**, were also affected with COVID-19.





- The majority, around **64%**, of people do not have COVID-19 and maintain good heart health.
- About **30%** of those who tested positive for COVID-19 do not have heart disease.
- Approximately **4%** of individuals do not have COVID-19 but still experience heart disease.
- A small percentage, around **1.5%,** had COVID-19 before developing heart problems.



- About **95%** of individuals who tested positive for COVID-19 had no experience of heart diseases.

Hypothesis Questions

Can you indicate which variables have a significant effect on the likelihood of heart disease?

After analyzing the dataset, we came to know that the variables which have significant effect on the likelihood of heart disease are:

- Having Obesity.
- Having Angina pain.
- Having Depressive Disorder
- Smoking
- Having Arthritis.
- Having Diabetes.
- Lack of physical activity.

How many people have at least 1 of 3 major risk factors for heart disease: no physical activity, high cholesterol, and smoking?

Almost **50%** of the population have atleast 1 major risk factor for heart disease.

Key Findings

Here are some important key findings of the whole analysis.

- **Demographic Analysis**: Females outnumber males in the sampled population, and the majority of individuals fall within the age range of 65-69 years.

- **Heart Disease Prevalence**: Most individuals did not experience heart disease, but among those who did, males comprised the majority, especially among individuals above 80 years old.

- **Behavioral Patterns**: The majority of individuals engage in physical activities, and most have never smoked in their lives. However, males show a slight lead over females in the category of current smokers.

- **Chronic Conditions**: Certain chronic conditions like arthritis, diabetes, and obesity are strongly associated with cardiovascular health disparities.

- **Mental Health Impact**: Approximately 20% of individuals are suffering from depressive disorders, and a significant portion of heart patients also have depressive disorders.

- **COVID-19**: Around one-third of the population had contracted COVID-19, with a majority not having heart disease. However, a small percentage experienced heart disease post-COVID-19 infection.

Recommendations

Here are some reccomendations for the health care management.

- Implement targeted screening programs aimed at identifying individuals at high risk for heart disease based on demographic factors such as age, gender, and socioeconomic status. This can help healthcare providers identify at-risk populations early and initiate preventive interventions.

- Launch public health campaigns to promote healthy lifestyle behaviors such as regular physical activity, balanced diet, smoking cessation, and stress management. Engage community organizations, schools, and workplaces to create supportive environments for adopting and maintaining healthy behaviors.

- Develop comprehensive chronic disease management programs that address the complex needs of individuals with comorbid conditions such as diabetes, obesity, and hypertension. These programs should focus on patient education, medication management, lifestyle modification, and regular monitoring to optimize disease control and prevent complications.

- Enhance mental health support services, particularly for individuals with depressive disorders, to improve overall cardiovascular health outcomes.

- Partner with community organizations and local stakeholders to develop community-based interventions targeting vulnerable populations. This may include offering health education workshops, cooking classes, fitness programs, and access to healthy food options in underserved areas.

- Invest in ongoing research and surveillance efforts to monitor trends in heart disease prevalence, risk factors, and outcomes. Use data-driven insights to inform evidence-based interventions and policy decisions aimed at improving cardiovascular health at the population level.

Conclusion

In conclusion, our analysis of cardiovascular health indicators among the adult population revealed important insights into the prevalence of heart disease and associated risk factors. We found that certain demographic factors, lifestyle behaviors, and chronic conditions play significant roles in shaping cardiovascular health outcomes. Females outnumbered males in our sample, and the majority of individuals were aged 65-69 years. While most individuals did not experience heart disease, males comprised the majority among those who did. Chronic conditions such as arthritis, diabetes, and obesity were associated with an increased likelihood of heart disease. Mental health disorders, particularly depressive disorders, were also prevalent among individuals with heart disease. Overall, our findings highlight the importance of targeted screening, promotion of healthy lifestyle behaviors, and integrated care approaches in addressing cardiovascular health disparities. By prioritizing these strategies, healthcare management can work towards improving heart health outcomes and reducing disparities among diverse populations.