

CARDIO VASCULAR RISK ANALYSIS

IE6600-COMPUTATION AND VISUALIZATION

FINAL PROJECT- GROUP 4



OUR TEAM

A Bunch of aspiring Data professionals



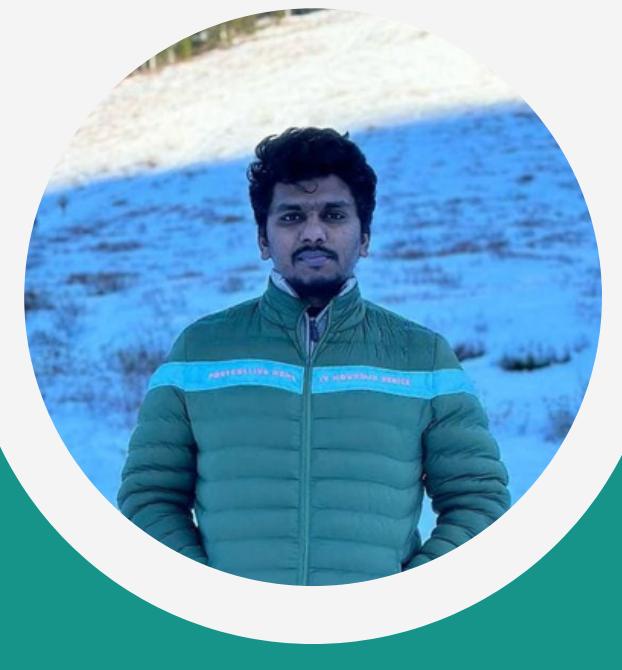
Shubham Gaur



Anjali Haryani



Jaamie Maarsh



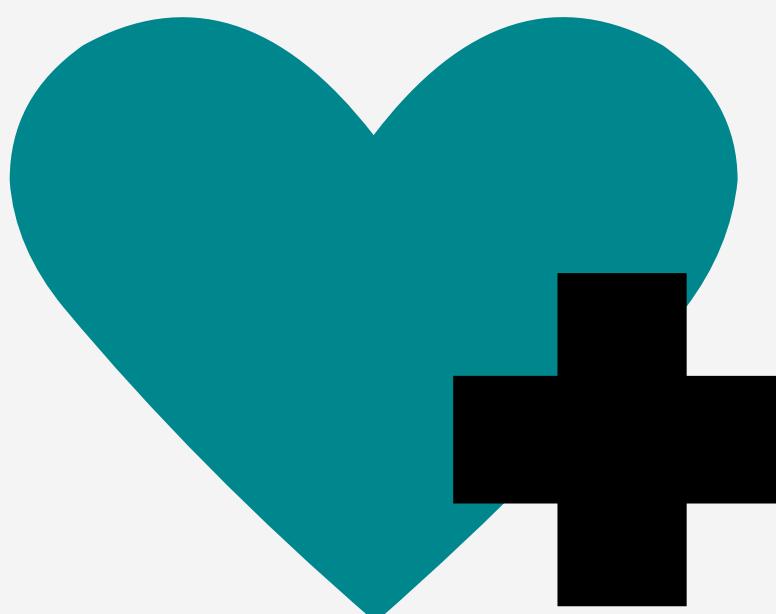
Kalyan Kumar



Payal Chavan

WHY CVD?

- One person dies every 33 seconds in the United States from cardiovascular disease.
- About 695,000 people in the United States died from heart disease in 2021—that's 1 in every 5 deaths



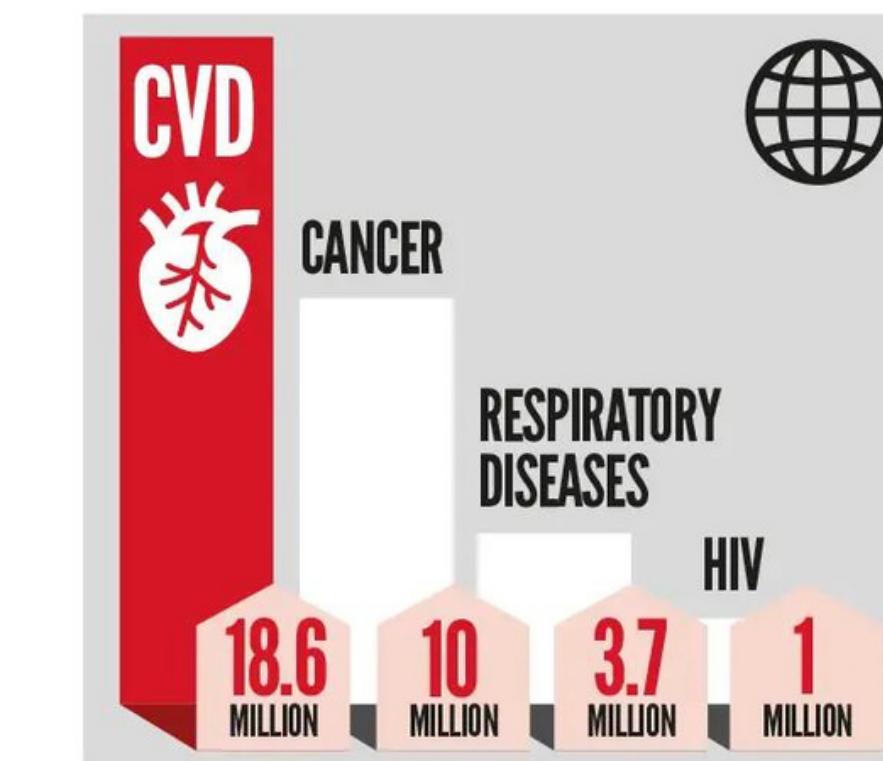
CARDIOVASCULAR DISEASE

THE WORLD'S NUMBER 1 KILLER

Cardiovascular diseases are a group of disorders of the heart and blood vessels, commonly referred to as **heart disease** and **stroke**.



GLOBAL CAUSES OF DEATH



RISK FACTORS FOR CVD





QUESTIONS TO BE ANSWERED...

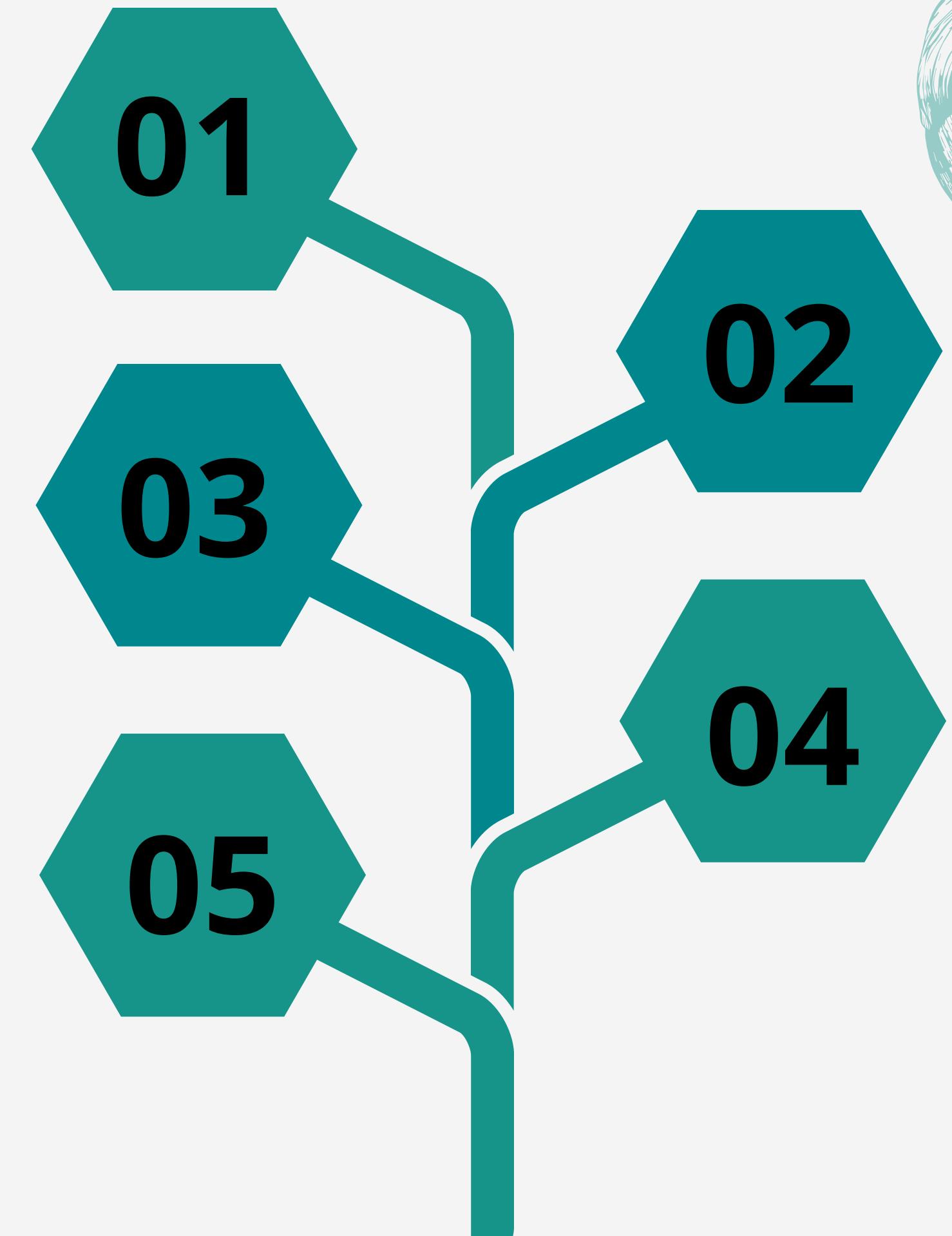
- Determining the prevalence of CVD across categories.
- Find out the major factors causing heart diseases?





Methodology

- 01** Phase 01 Data reading and Understanding
- 02** Phase 02 Data Cleaning/Pre-processing
- 03** Phase 03 Approach and Methods
- 04** Phase 04 Data Visualisation & Analysis
- 05** Phase 05 Answering the question, Improvements & Learnings



DATA HANDLING



Data Source

- Database Source -Kaggle
- Other resources - Centres for Disease control and prevention



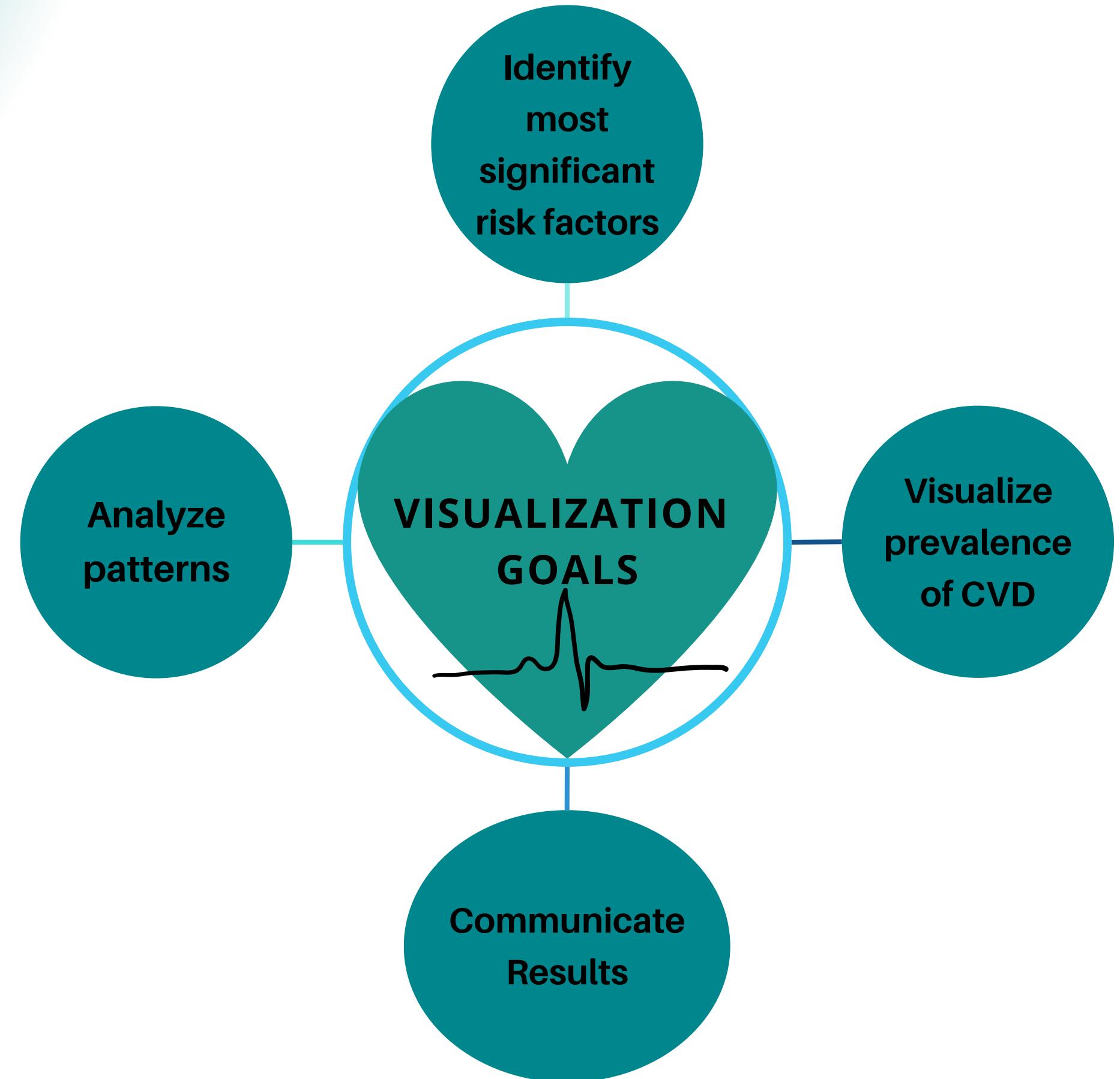
Data Preprocessing

- Null values
- Duplicate values
- Feature Engineering
- Outlier values



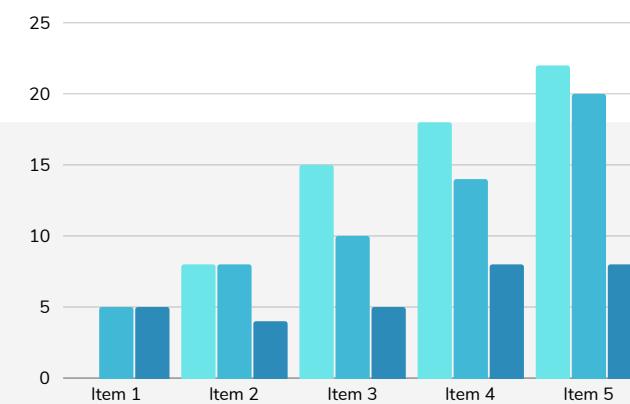
Nature of Data

- Binary values
- Uncertain data values
- Unclear bins



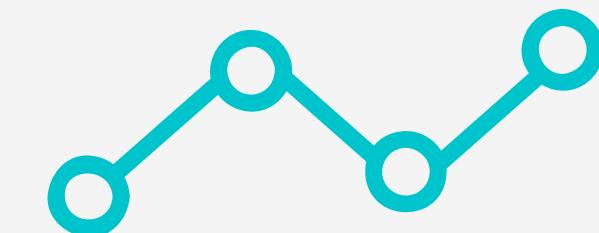
VISUALIZATION CHOICES

“Developing Charts, Analyzing Visualizations, and Driving Insights”



Bar Graphs

- Compare frequency of categorical data.
- Count Plots.



Line Graphs

- To analyze trends of different columns.

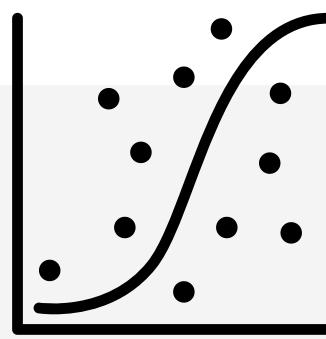


Pie Charts

- Percentage of the population.

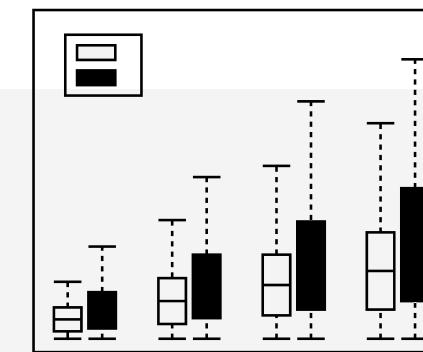
VISUALIZATION CHOICES

"More Charts, More Visualizations, and More Insights"



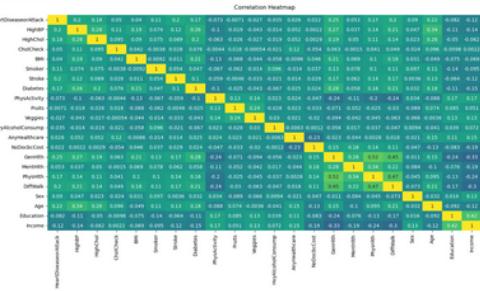
Scatter Plots

- To identify correlational relationships and patterns.



Box Plots

- To display the distribution of data, and identify outliers.



Heatmaps

- Correlation between different factors.

Design and Development

“Unveiling Patterns, Overcoming Challenges, and Driving Insights”



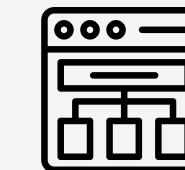
Define Objectives and Scope



Analyzing Data



Discuss Challenges Related to Data Quality and Consistency



Describe Data Preprocessing Techniques



Visuals and Insights



DESIGN CHALLENGES



Uncertain data values

- One Hot Encoding
Education, Income etc.



Binary values

- Converted to Categorical Data
Gender, Smoker etc.



Continuous Data

- Feature Engineering
BMI, Age

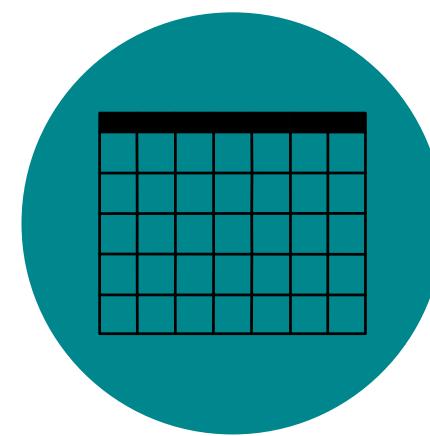
DATA VISUALISATION TOOLS

Applications that enable the creation of graphical representations of data analysis, and communication of complex information



PYTHON

- Data Preprocessing
- Custom Analytics
- Flexible Data Manipulation

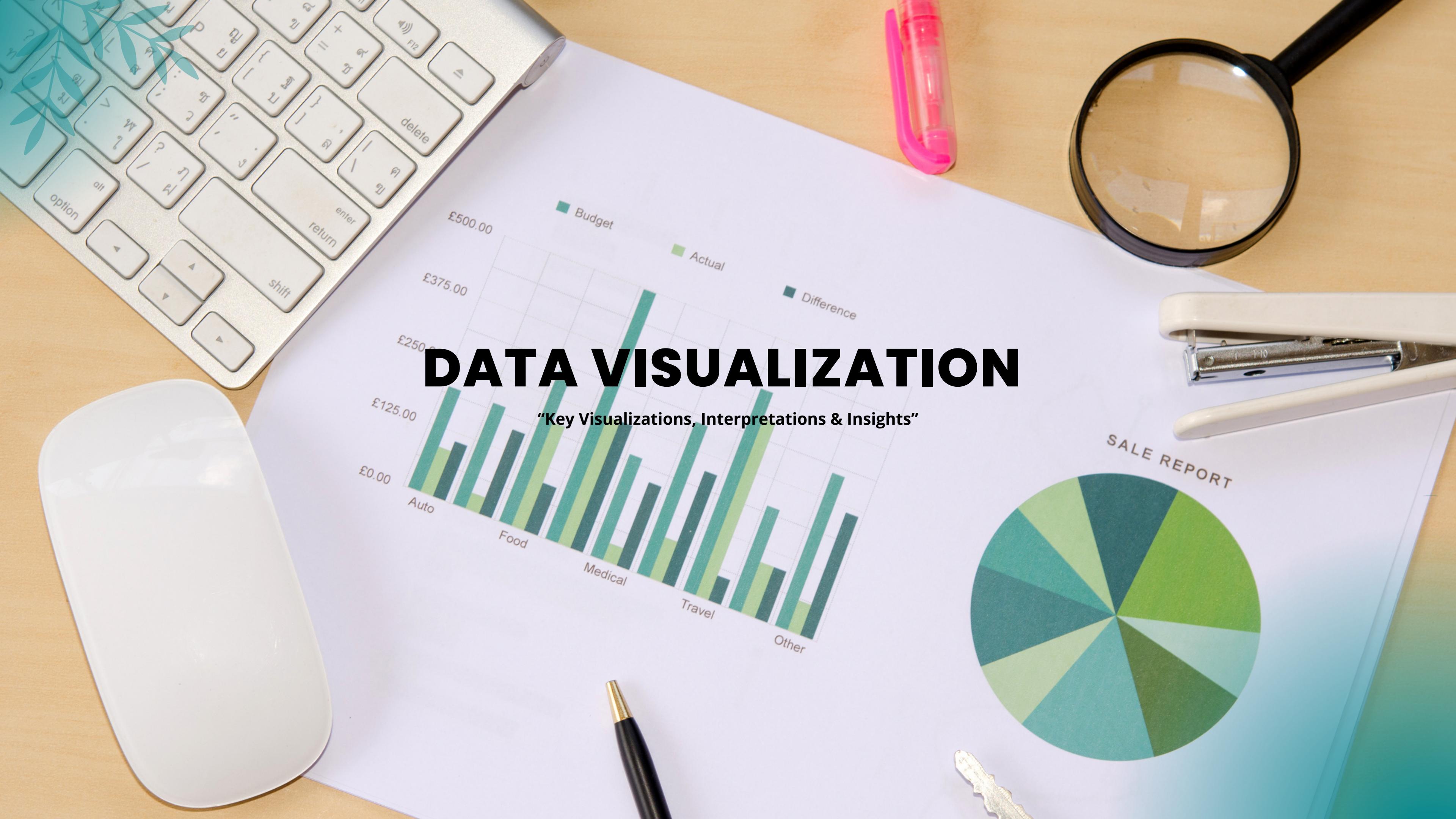
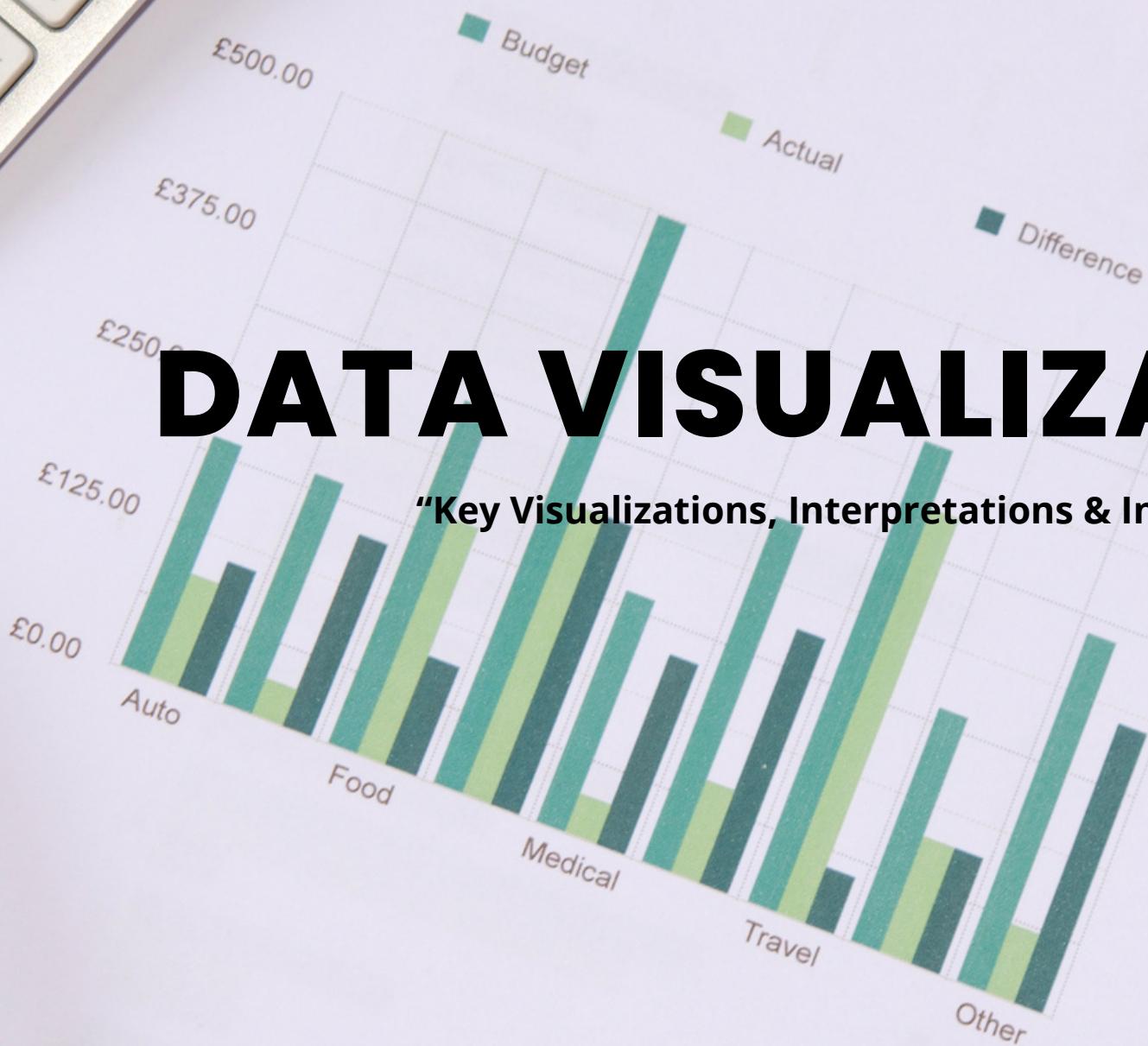


TABLEAU

- Easy Drag-and-Drop Interface
- Beautiful Visuals
- Explore Easily

DATA VISUALIZATION

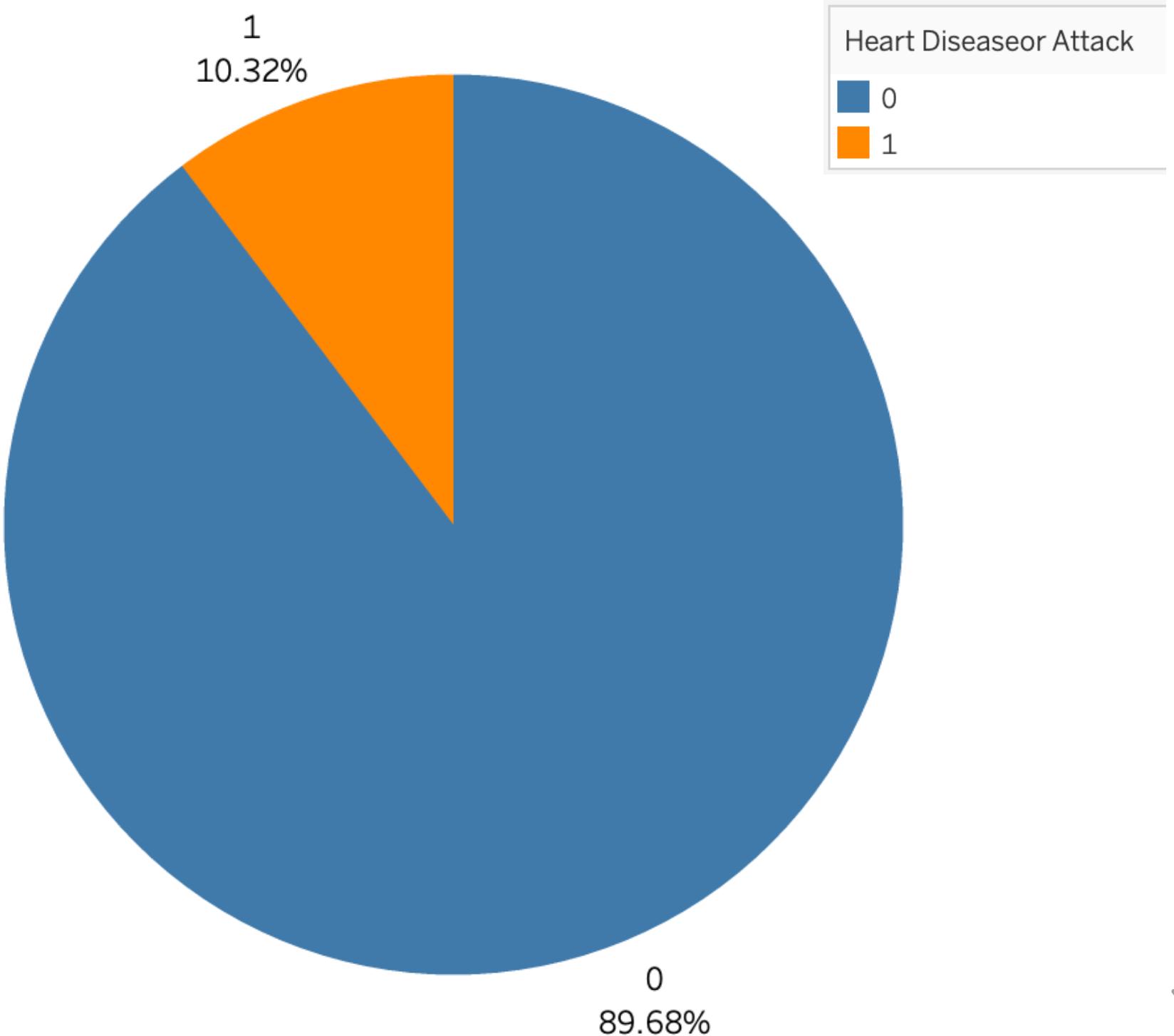
“Key Visualizations, Interpretations & Insights”



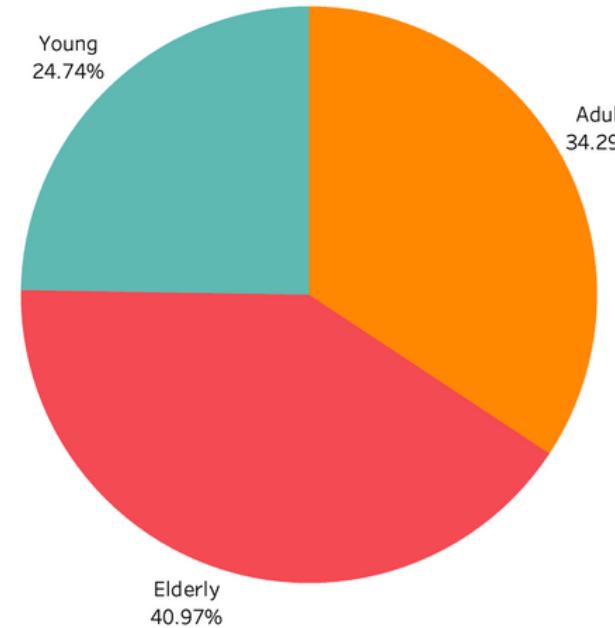
DISTRIBUTION OF HEART ATTACKS

Prevalence of Heart Disease/Heart Attack:

- 10.3% of the population has a history of heart disease or heart attack.

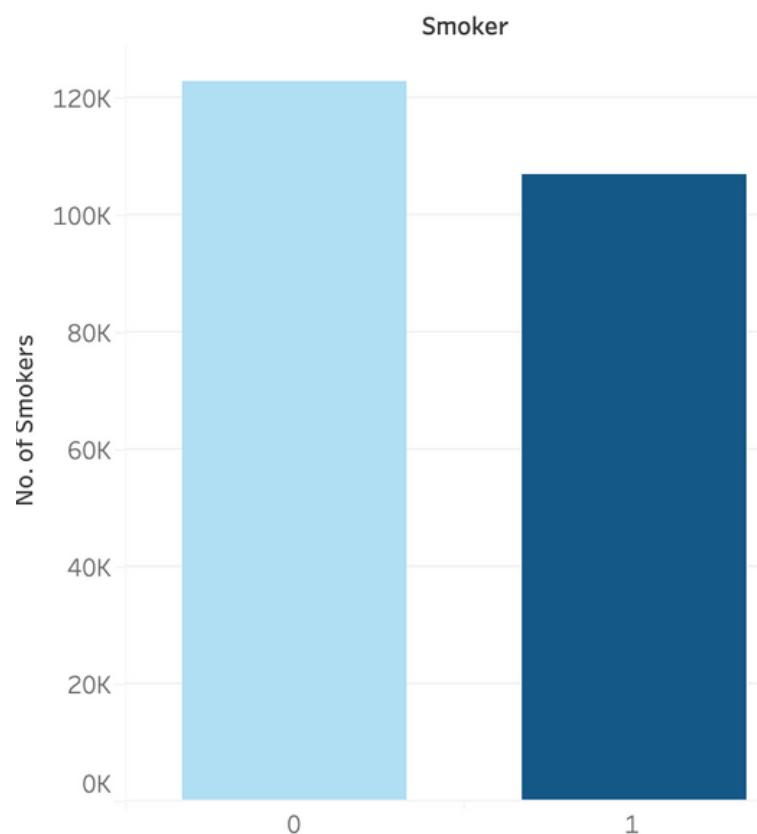


DISTRIBUTION OF DIFFERENT FACTORS



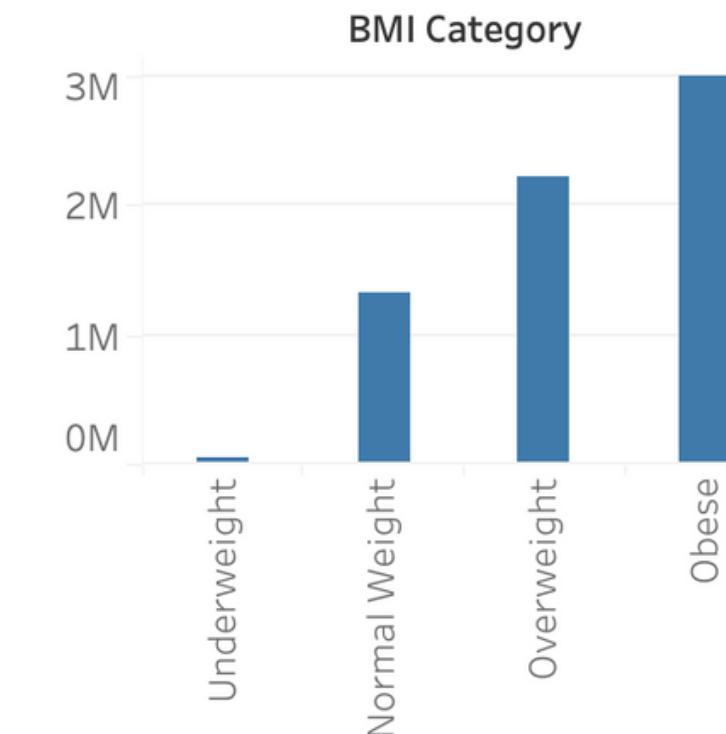
AGE DISTRIBUTION

- Elderly age group: 41%
- Young age group: 24.7%



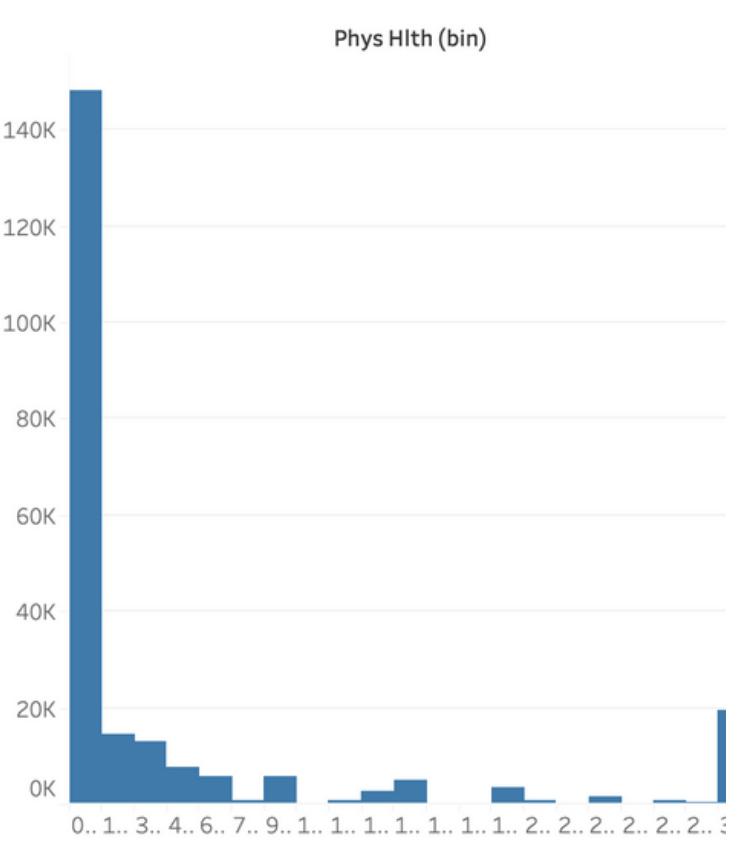
SMOKER/NON-SMOKER

- Higher count of non-smokers.
- Emphasizes the prevalence of non-smoking individuals



BMI DISTRIBUTION

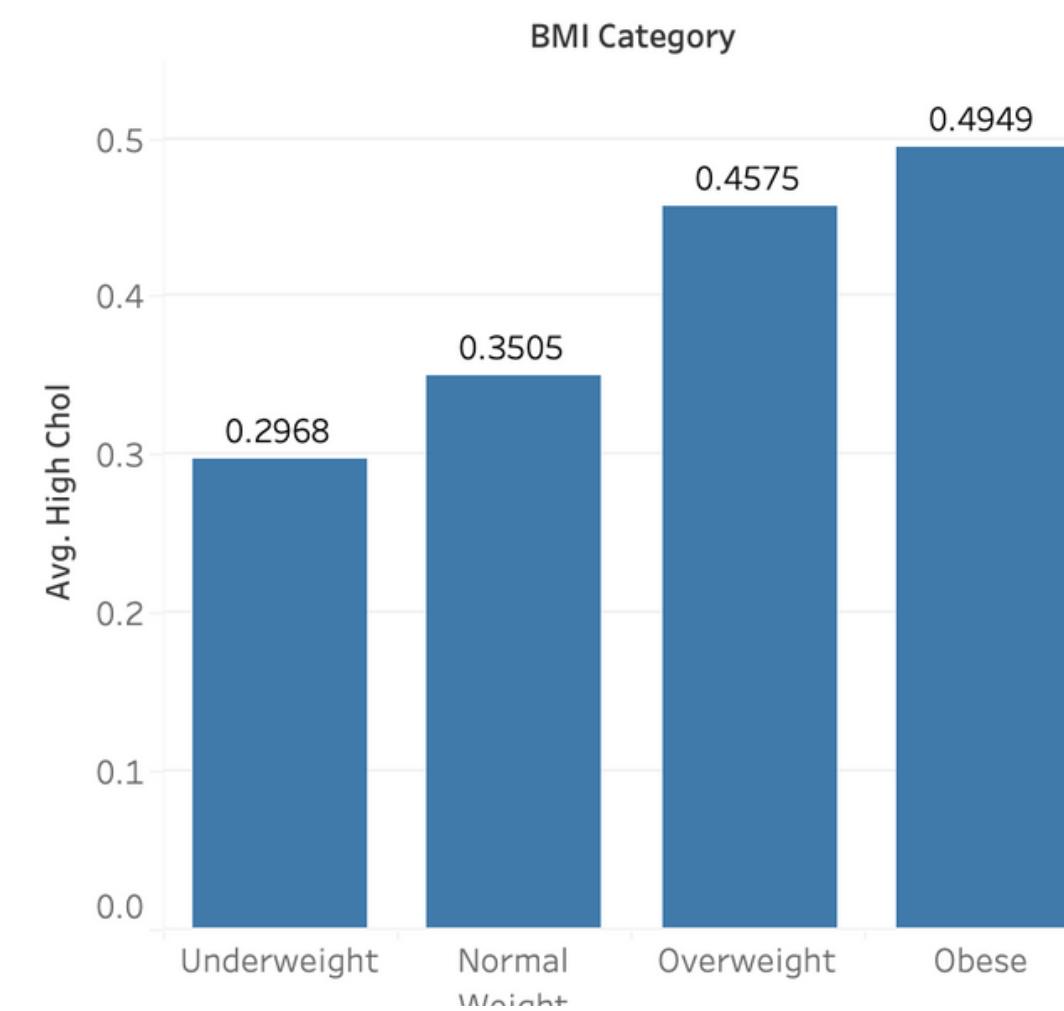
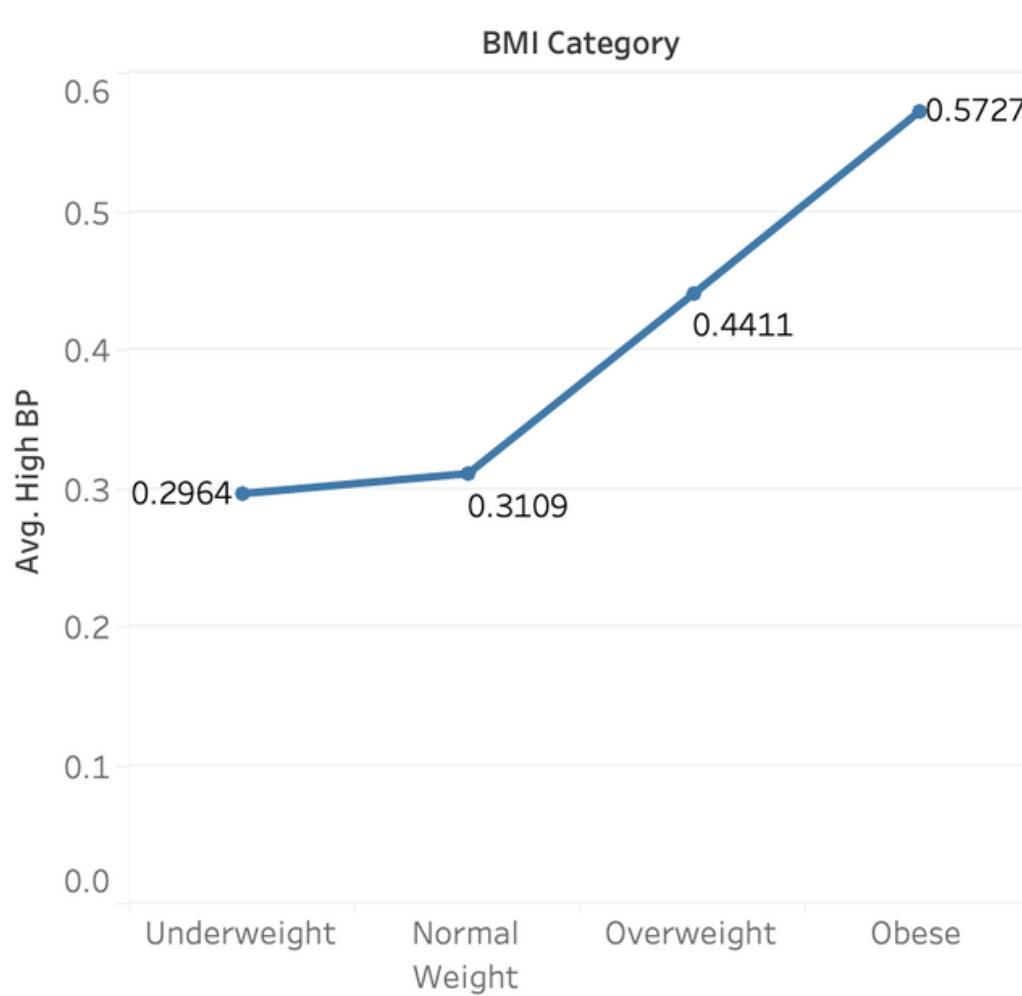
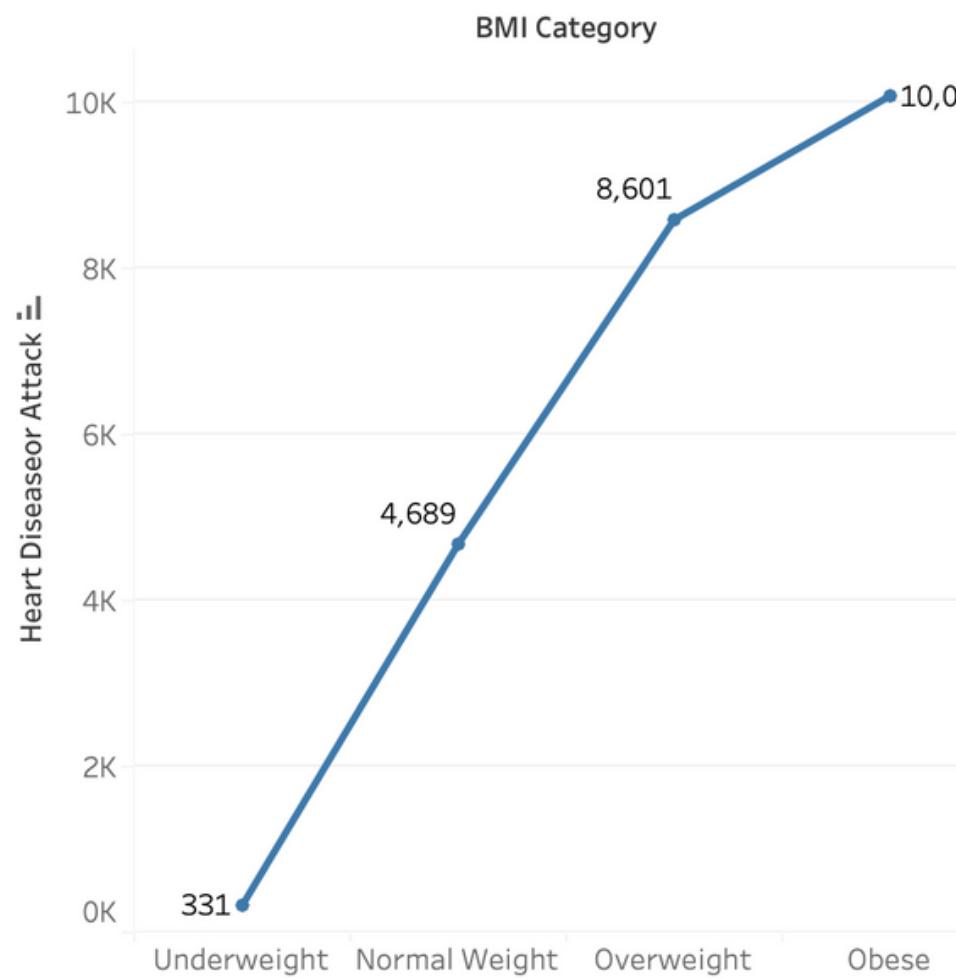
- 'Obese' and 'Overweight': Approx. 9000 individuals
- 'Underweight': < 500 patients
- 'Normal Weight': Around 6000 patients



PHYSICAL HEALTH DISTRIBUTION

- 0 indicates people are not physically active.
- Highlights a prevalent lack of physical activity in the population.

DISTRIBUTION OF HEART ATTACKS ACROSS BMI CATEGORIES



BMI CATEGORY

- Correlation with increased heart disease.
- Higher BMI values indicate potential risk for heart-related problems.

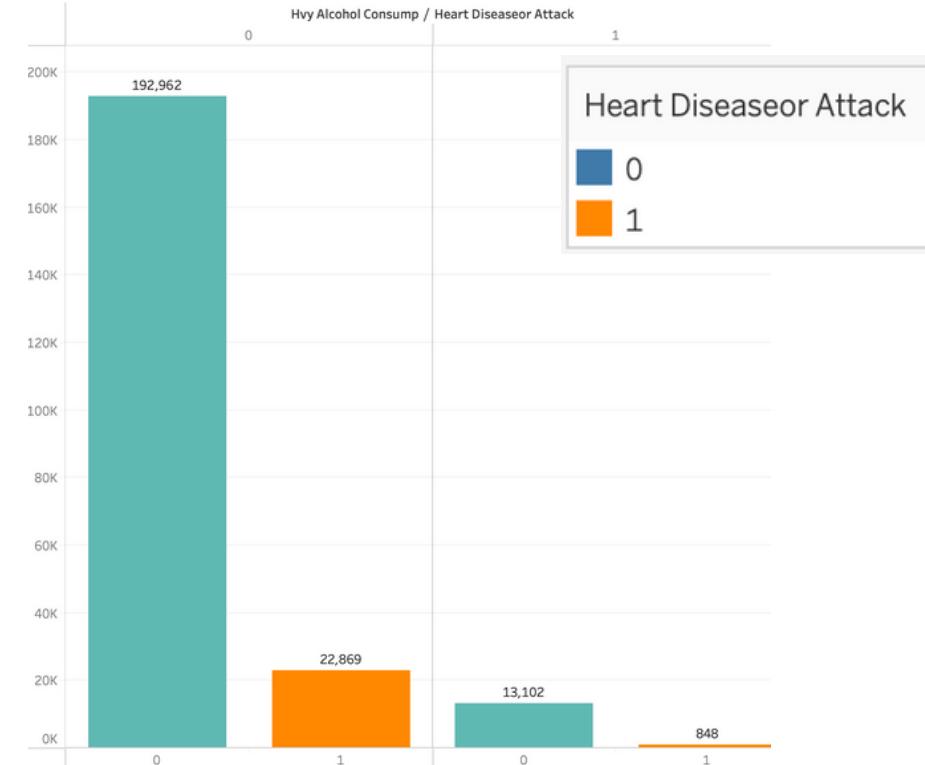
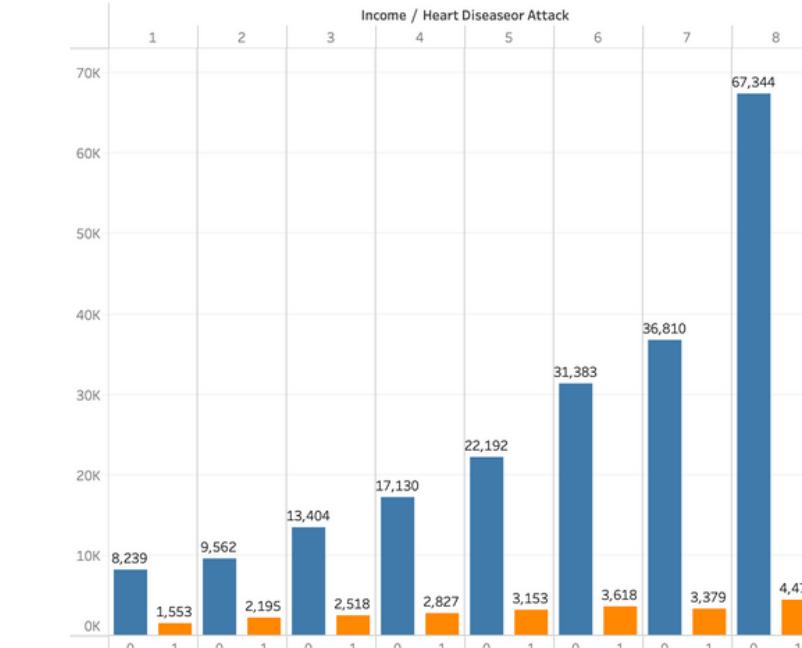
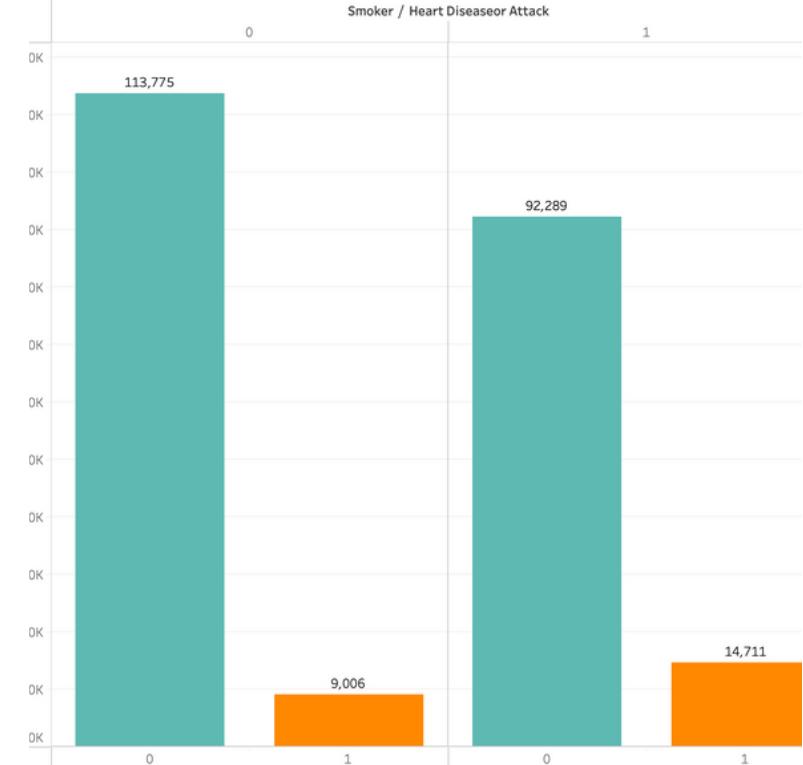
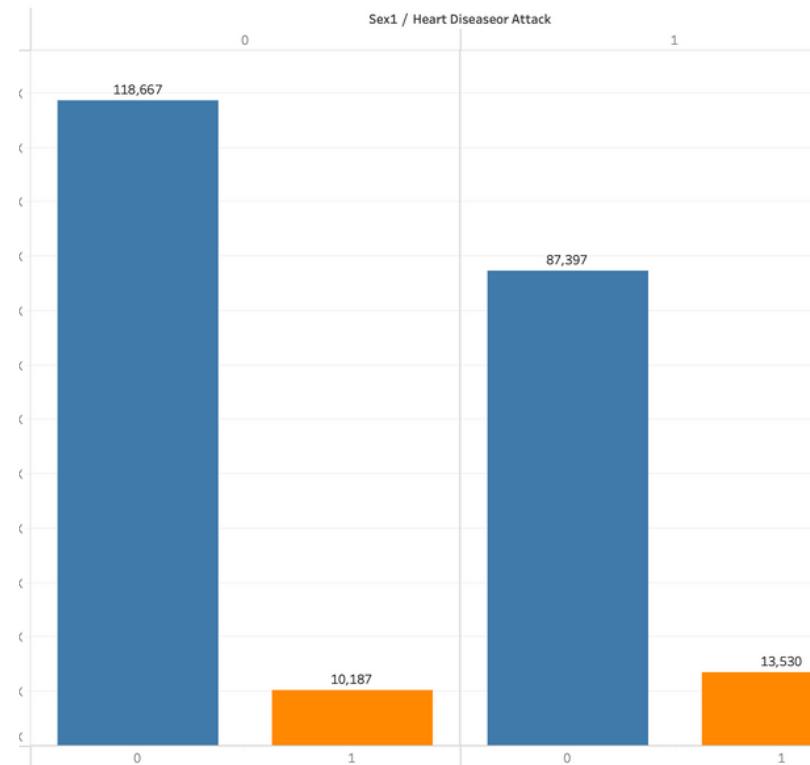
BMI CATEGORY WITH HIGH BP

- Clear upward trend in high blood pressure levels
- Emphasizes the substantial link between weight and the risk of heart attacks.

BMI CATEGORY WITH HIGH CHOL

- Highlights link between weight categories.
- Underscores the importance of the relationship between weight, cholesterol, and heart attack risk.

LIFESTYLE FACTORS INFLUENCING HEART ATTACKS



GENDER /HEART ATTACK

- Females predominate, signaling higher male susceptibility to heart attacks.
- Larger female dataset emphasizes the need for balanced gender inclusion in research.

SMOKER /HEART ATTACK

- Non-smokers outnumber smokers, indicating a slightly higher risk of heart disease among smokers.
- Addressing smoking-related risks is crucial, particularly in populations with higher smoking prevalence.

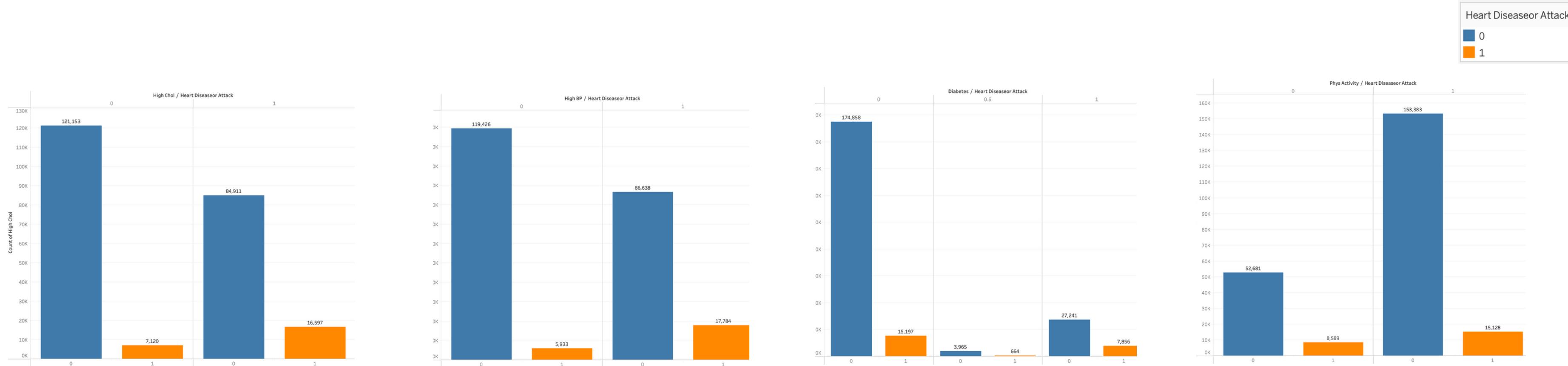
INCOME LEVELS /HEART ATTACK

- The dataset shows a higher likelihood of heart disease with increasing income levels.
- Correlation underscores the importance of considering socioeconomic factors in health outcomes.

HEAVY ALCOHOL/HEART ATTACK

- Heavy alcohol consumption shows an unexpected association with lower heart disease rates, warranting further investigation for nuanced discussions.

HEALTH FACTORS INFLUENCING HEART ATTACKS



HIGH CHOLESTEROL /HEART ATTACK

- Few individuals exhibit high cholesterol levels.
- Despite low prevalence, a strong association exists with increased heart disease.

HIGH BP /HEART ATTACK

- High blood pressure is linked to a higher proportion with a history of heart disease.
- Emphasizes high blood pressure's potential role in cardiovascular health outcomes.

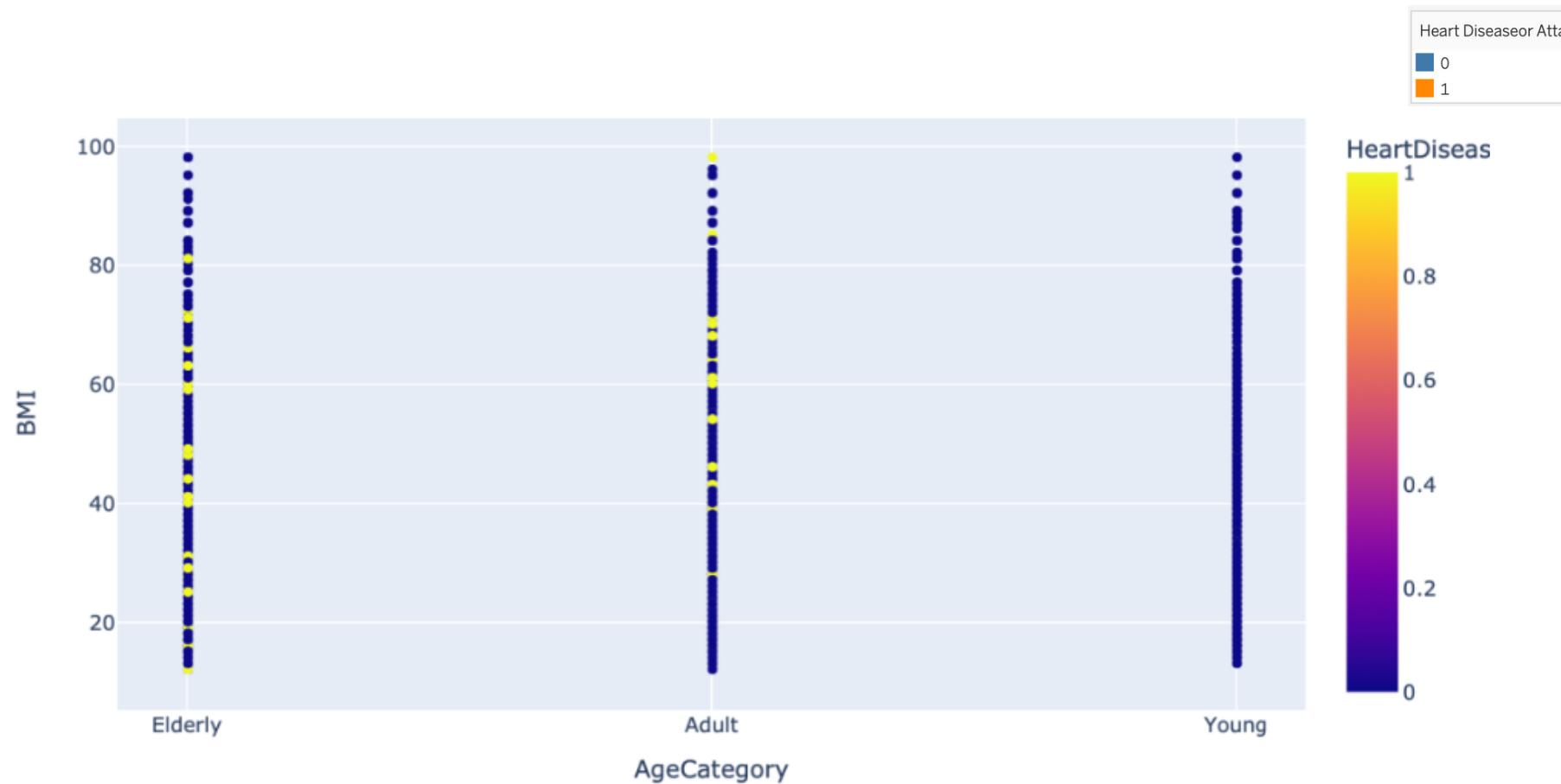
DIABETES/HEART ATTACK

- The dataset reveals a scarcity of individuals with diabetes, indicating lower prevalence.
- A larger proportion of those without diabetes have a history of heart disease.

PHYSICAL ACTIVITY /HEART ATTACK

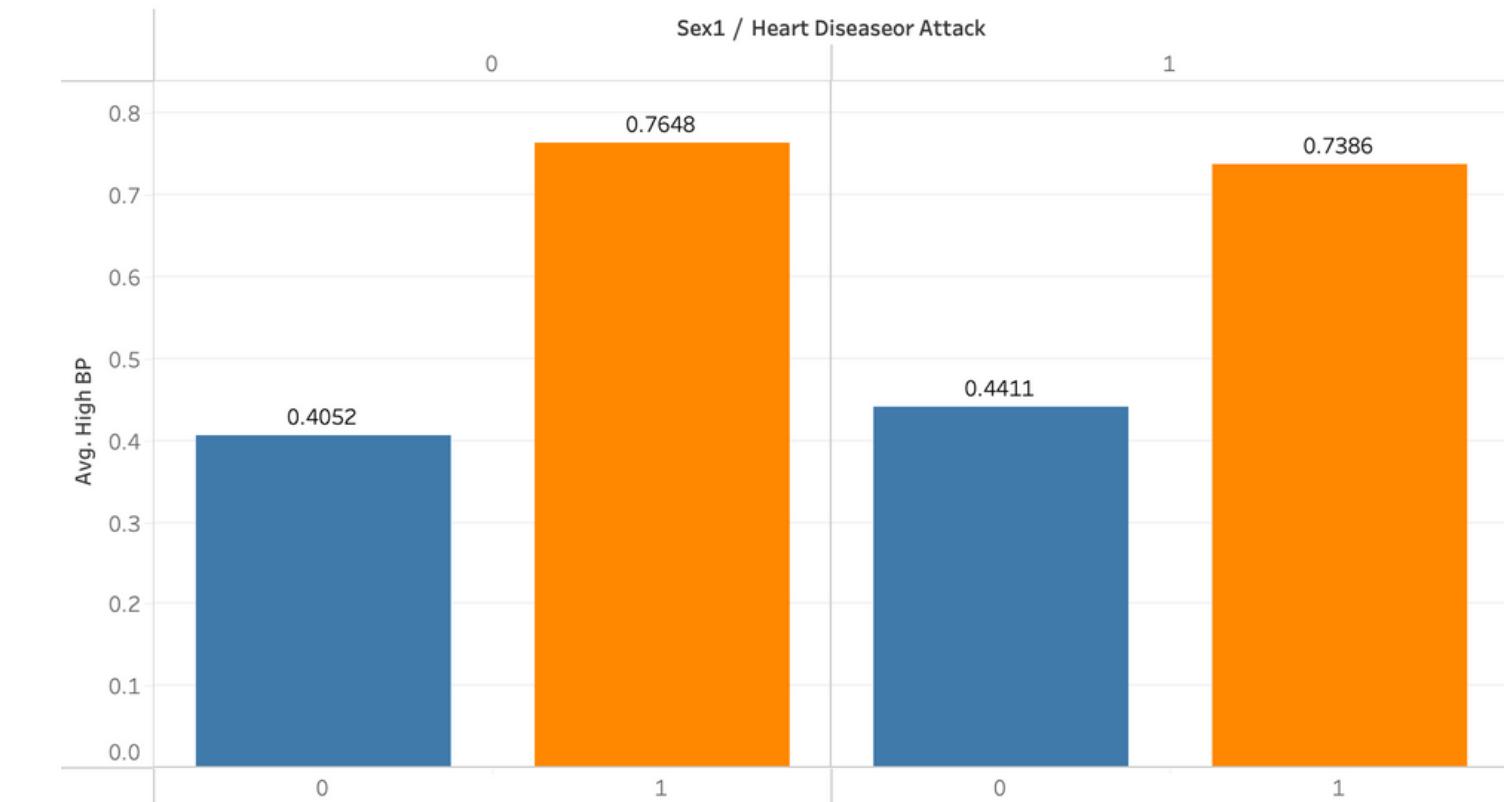
- Regular physical activity slightly increases heart disease susceptibility, highlighting a nuanced relationship.
- Understanding specific activities and durations is crucial for personalized heart health recommendations.

CORRELATION BETWEEN FACTORS INFLUENCING HEART ATTACKS



AGE & BMI /HEART ATTACK

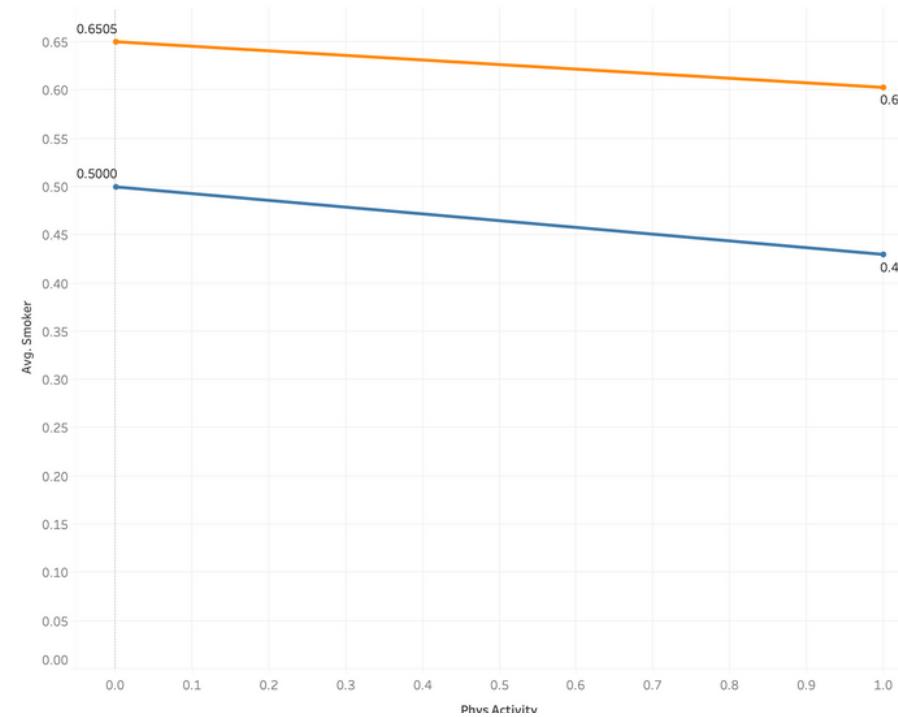
- Scatterplot indicates higher heart disease risk in elderly.
- 'Adults' also show risks, influenced by activity and lifestyle.
- High BMI and age strongly link to heart disease.



GENDER & HIGHBP /HEART ATTACK

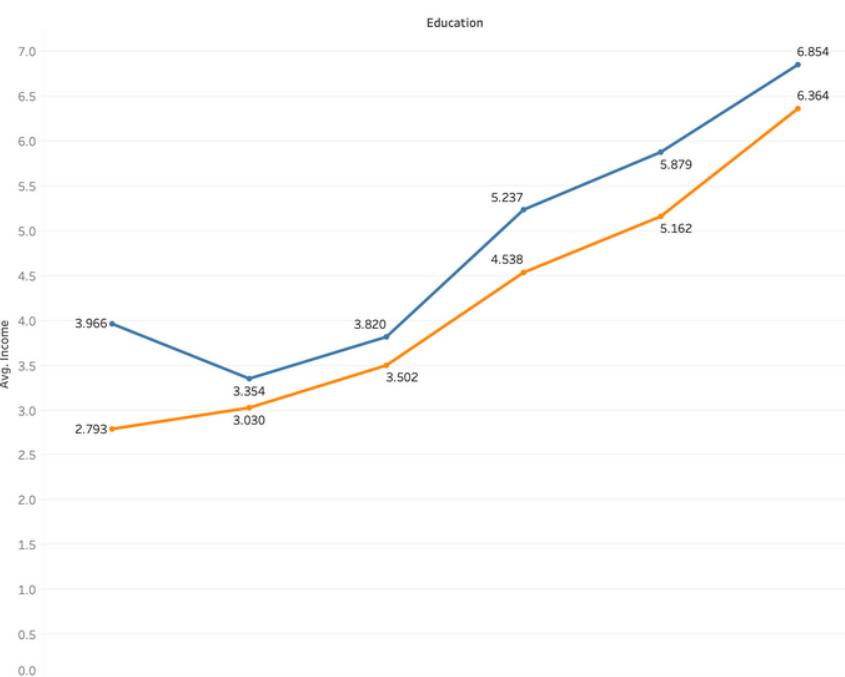
- High blood pressure elevates cardiovascular risk.
- Women have extra risk, showing gender differences.
- Recognizing this is crucial for tailored prevention and care

CORRELATION BETWEEN FACTORS INFLUENCING HEART ATTACKS



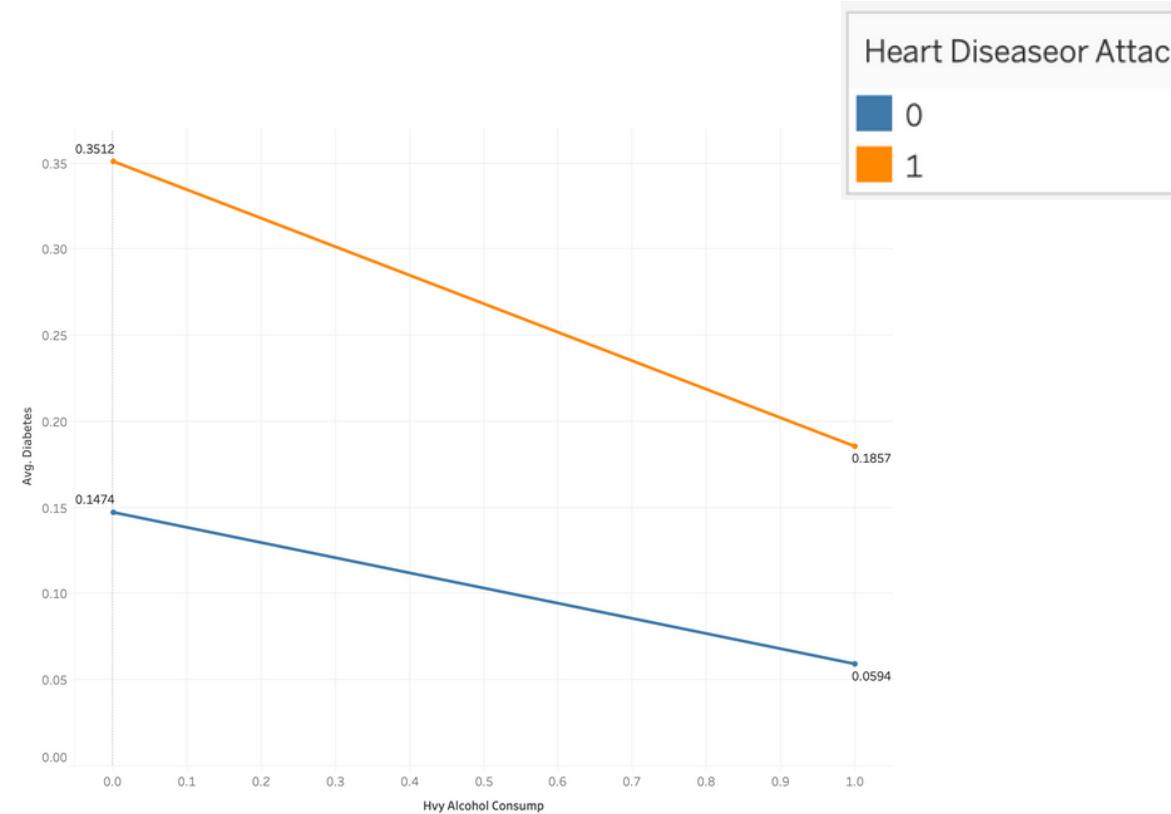
PHYSICAL ACTIVITY /AVG SMOKER

- Smoking and exercise are inversely related. This means that people who smoke are less likely to exercise.
- This relationship highlights the complex interplay of lifestyle factors that can impact cardiovascular health.



EDUCATION /AVG INCOME

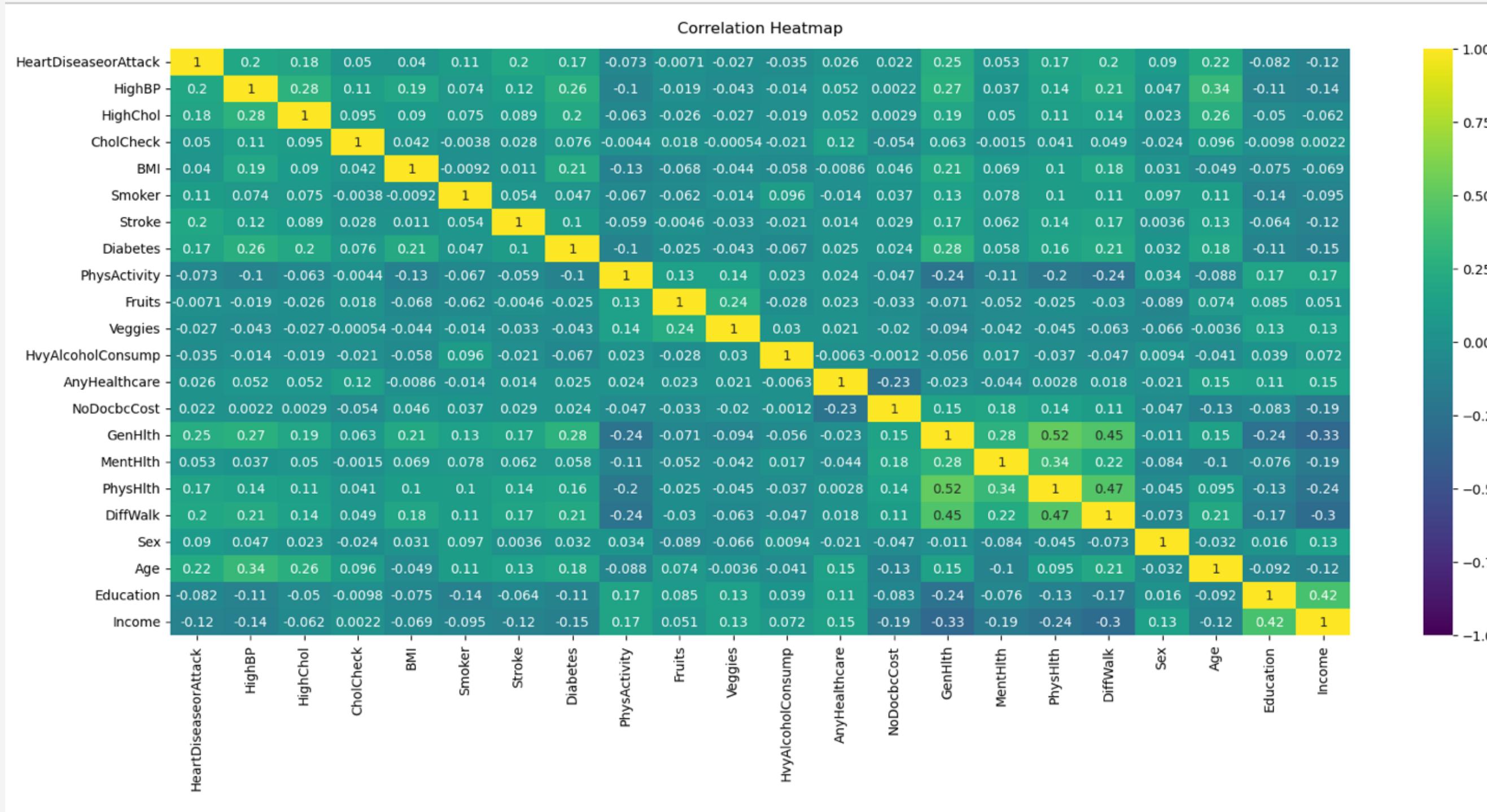
- Higher education and income are positively correlated, suggesting a complex interplay between socioeconomic factors and health outcomes.
- The link between education, income, and stress emphasizes the importance of considering these factors in promoting cardiovascular health.



ALCOHOL CONSUMPTION /AVG DIABETES

- The line plot indicates less alcohol consumption linked to more diabetes.
- Highlights a complex interaction affecting cardiovascular outcomes.

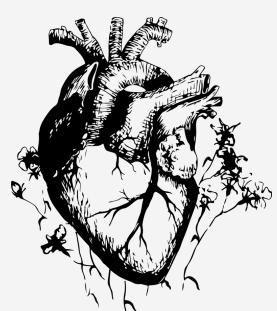
CORRELATION MATRIX



- Age, BMI, High BP, High Chol, Physical health emerges as pivotal determinant in direct proportion to heart disease.
- Physical Activity, Income, Education, Heavy Alcohol Consumption are negatively correlated with heart disease.
- High BP & Age, High Chol & Age having positive correlation between them.
- Income & general health, Income & Physical health having negative correlation between them



KEY FINDINGS



10.3% population
having heart disease



Major Factors

- Higher BMI
- Smoking
- High Cholesterol
- High BP
- Age



Minor Factors

- Gender
- Education
- Income
- Physical Activity



Factors not Impacting Heart Health

- Diabetes
- Heavy Alcohol
Consumption

IMPROVEMENTS



Dataset Selection

- Binary Values
- Uncertain data values
- Low percentage of the population with heart conditions



Physical & Mental Health

- Data range 0 - 30. Hard to interpret



Detailed Study

- Physical Activity
- Heavy Alcohol Consumption
- Diabetes

Learnings

Team Collaboration & Communication

Technical Skills

Data Analysis & Interpretation

Heart Health factors

Problem Solving



Thank You





QUESTION TIME!!

