



# Heart Disease Diagnostic Analysis

## Detailed Project Report

## Project Detail:

Project Title	Heart Disease Diagnostic – Analysis
Technology	Business Intelligence
Domain	Healthcare
Project Difficulty Level	Advanced
Programming Language Used	Python
Libraries Used	Numpy, Pandas, Matplotlib, Seaborn
Tools Used	Jupyter Notebook, Power BI, MS Excel

## Objective:

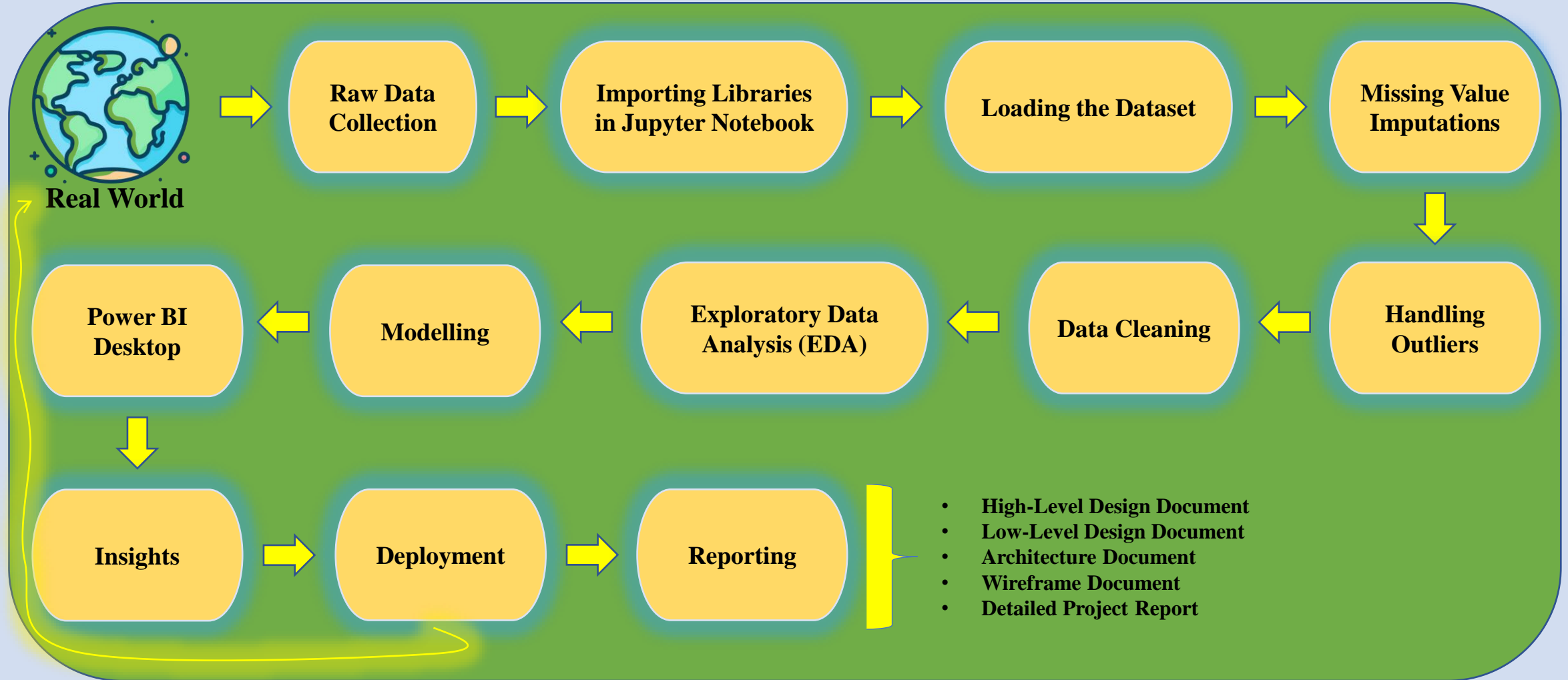
The goal of this project is to examine how often heart disease occurs by analyzing various factors linked to heart disease. This involves looking at different characteristics, such as age, blood pressure, and cholesterol levels, to better understand what might contribute to the risk of developing heart disease.

# Problem Statement:

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. You are required to analyze this health and medical data for better future preparation.

Perform ETL (Extract, Transform, Load) on the heart disease diagnostic database, followed by exploratory data analysis (EDA) using Python. The focus is on extracting valuable insights such as heart disease rates, occurrences by gender and age, and comparing various attributes to uncover meaningful information. The analysis will aim to identify key metrics and factors that influence heart disease, revealing important relationships between the data attributes. A dashboard will be created using visualizations to present these findings clearly, offering a comprehensive view of the dataset. Additionally, research will be conducted to ensure well-informed conclusions and actionable insights are derived from the data.

# Architecture:



# Dataset Information:

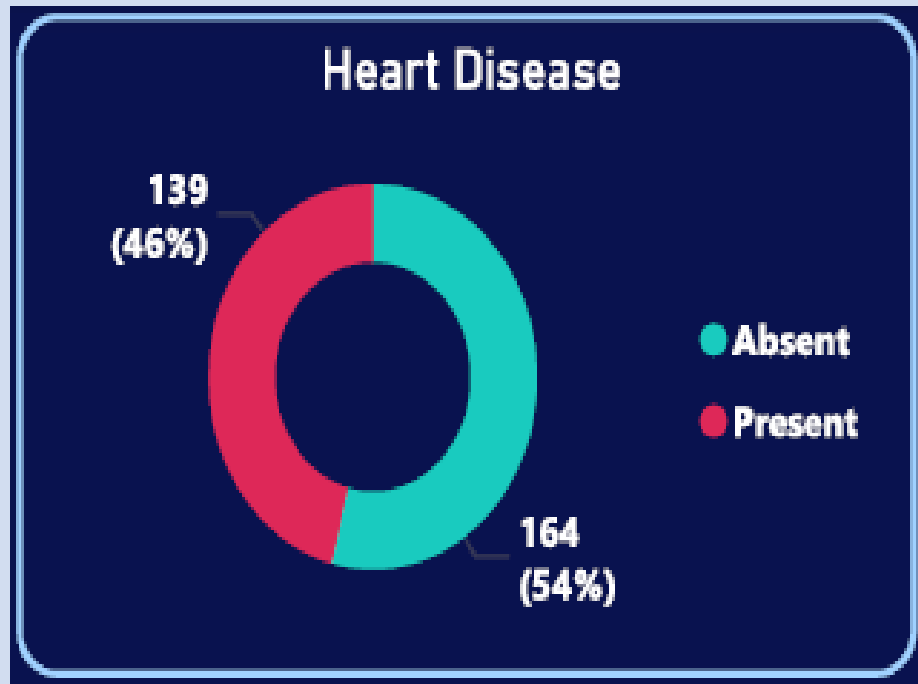
1. **age:** Age of the patient
2. **sex:** Sex of the patient (1 = male, 0 = female)
3. **cp:** Chest pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
4. **trestbps:** Resting blood pressure (in mm Hg)
5. **chol:** Serum cholesterol (in mg/dl)
6. **fbs:** Fasting blood sugar (> 120 mg/dl, 1 = true, 0 = false)
7. **restecg:** Resting electrocardiographic results (0: normal, 1: ST-T wave abnormality, 2: left ventricular hypertrophy)
8. **thalach:** Maximum heart rate achieved
9. **exang:** Exercise-induced angina (1 = yes, 0 = no)
10. **oldpeak:** ST depression induced by exercise relative to rest
11. **slope:** Slope of the peak exercise ST segment (1: upsloping, 2: flat, 3: downsloping)
12. **ca:** Number of major vessels (0–3) colored by fluoroscopy
13. **thal:** Thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect)
14. **num:** Diagnosis of heart disease (0 for no disease, 1 for yes)

# Important Parameters:

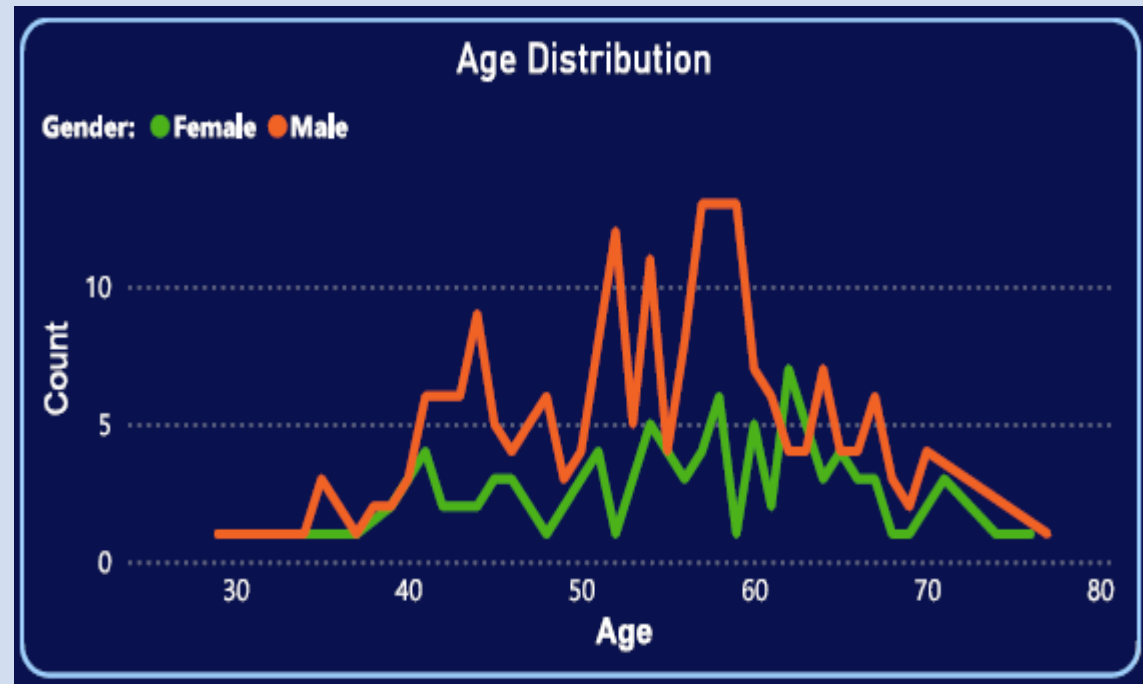
- ❖ **Age:** Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. It is estimated that 82 per cent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.
- ❖ **Gender:** Men are at greater risk of heart disease than pre-menopausal women. Once past menopause, it has been argued that a woman's risk is similar to a man's although more recent data from the WHO and UN disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.
- ❖ **Resting Blood Pressure:** Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.
- ❖ **Fasting Blood Sugar:** Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of heart attack.
- ❖ **Cholesterol:** A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also increases your risk of heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of heart attack.
- ❖ **Resting ECG:** For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercising ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.
- ❖ **Max heart rate achieved:** The increase in the cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.
- ❖ **ST Depression:** In unstable coronary artery disease, ST-segment depression is associated with a 100% increase in the occurrence of three-vessel/left main disease and to an increased risk of subsequent cardiac events. In these patients, an early invasive strategy substantially decreases death/myocardial infarction.

# Insights:

What is the nature of the population we are analyzing?



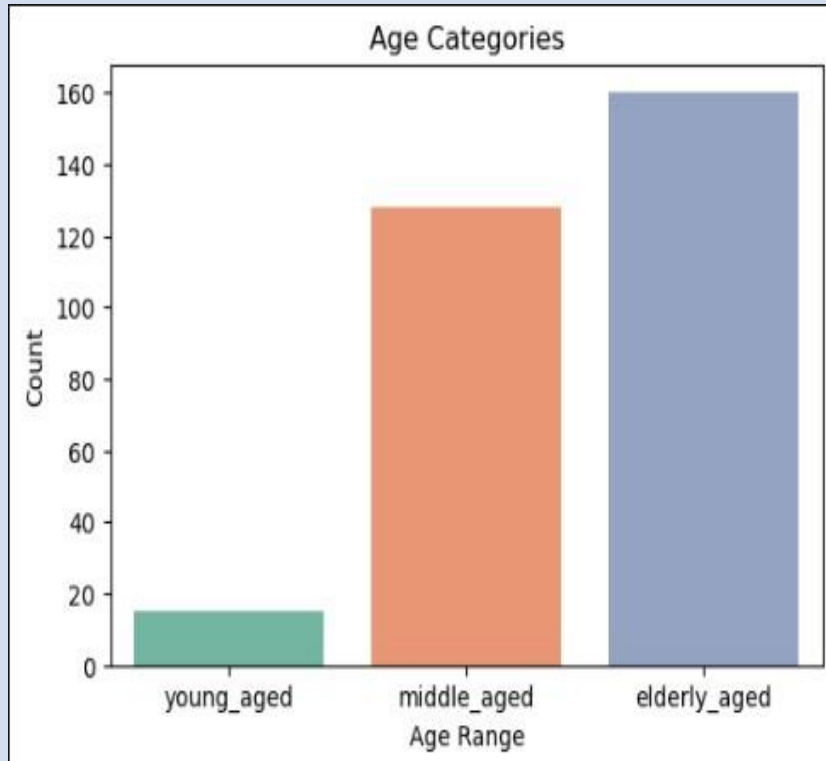
- Approximately 46% of individuals are affected by heart disease.



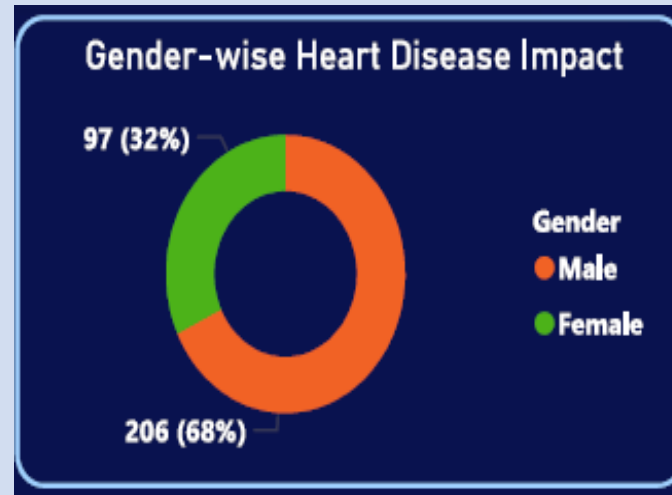
- Men aged 50 to 60 years exhibit a higher prevalence, while women predominantly fall within the 55 to 65 years age group.



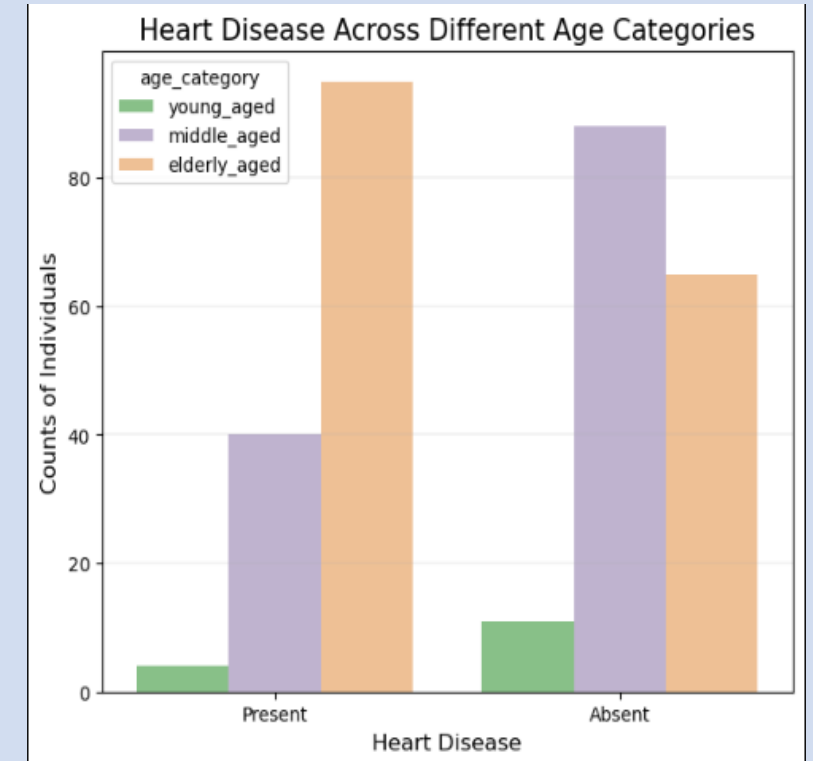
## Who is most affected by heart disease?



- The population has a higher proportion of Elderly People. (Individuals aged 55 and above)

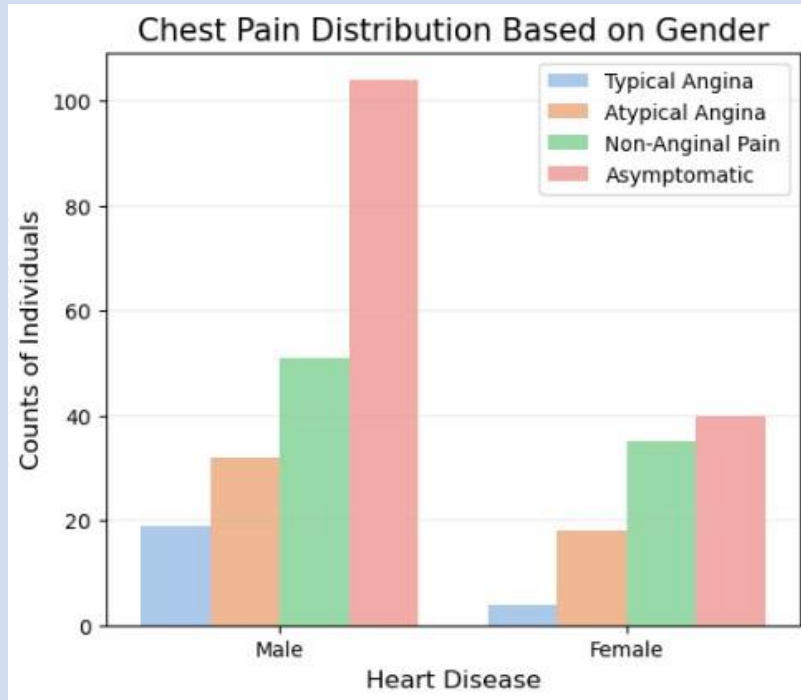


- Men have a higher susceptibility to heart disease compared to women.

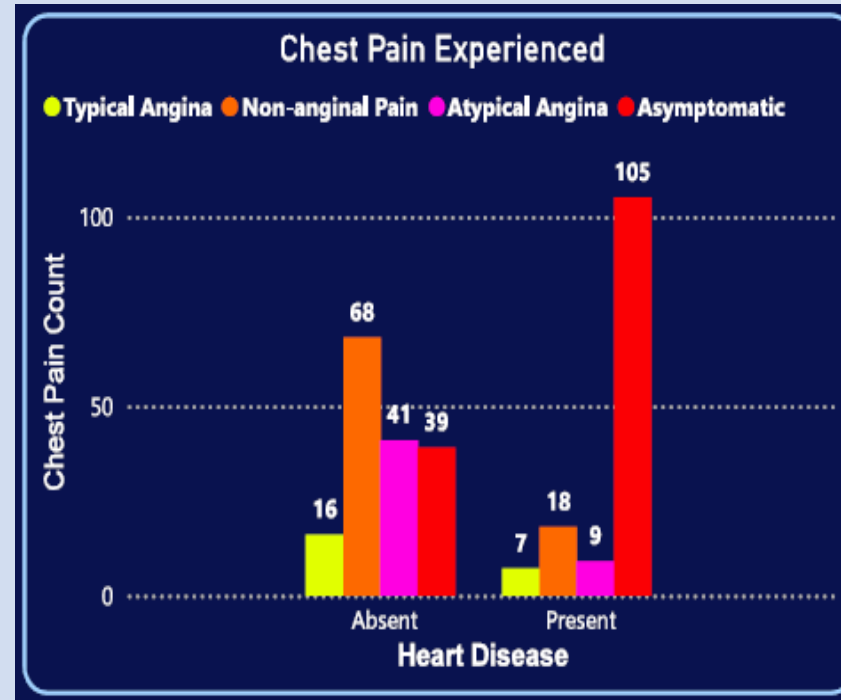


- People over the age of 55 (Elderly People) are at a greater risk of developing heart disease.

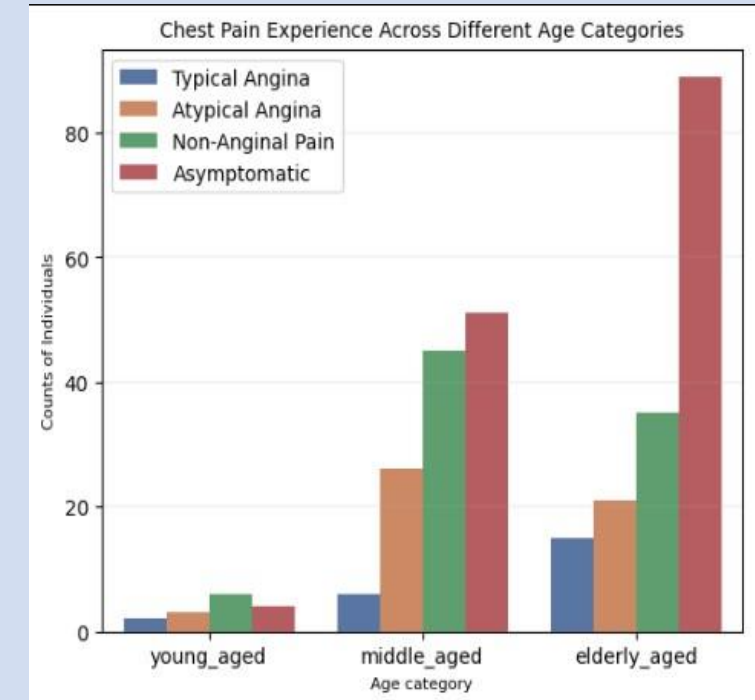
## Chest Pain Experienced by Patients



- It is evident that a larger proportion of men experience asymptomatic chest pain.



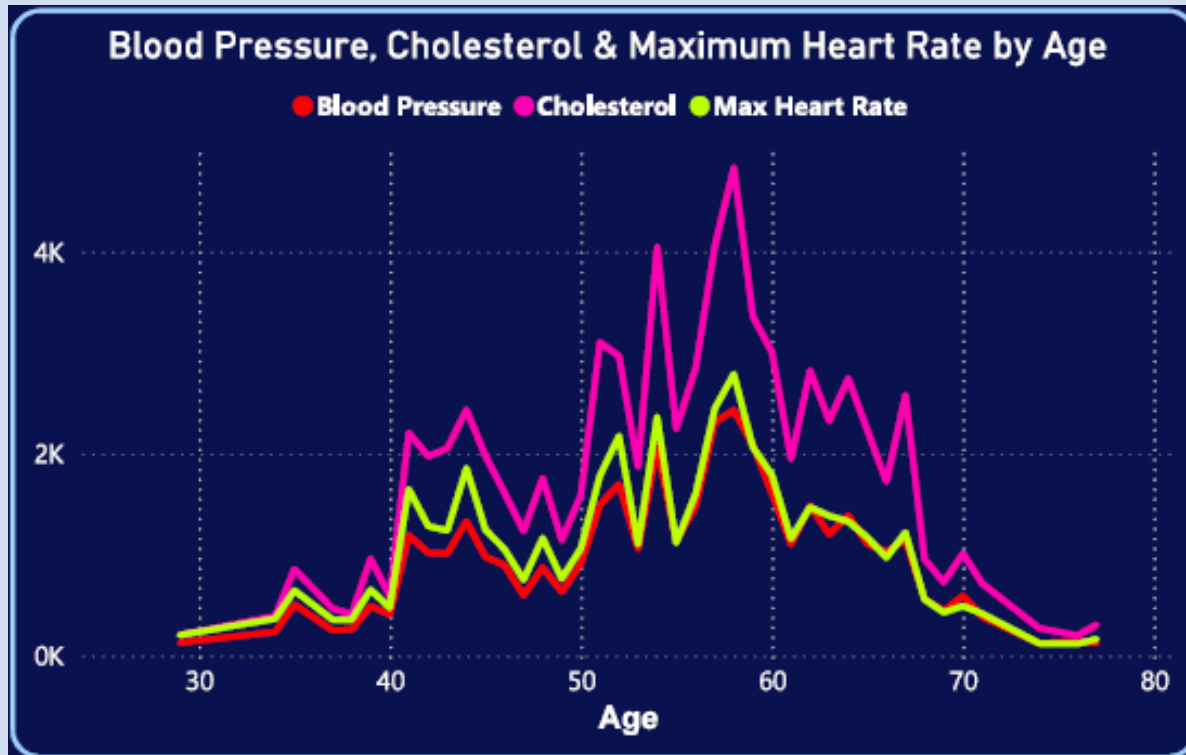
- It seems people having asymptomatic chest pain have a higher chance of heart disease.



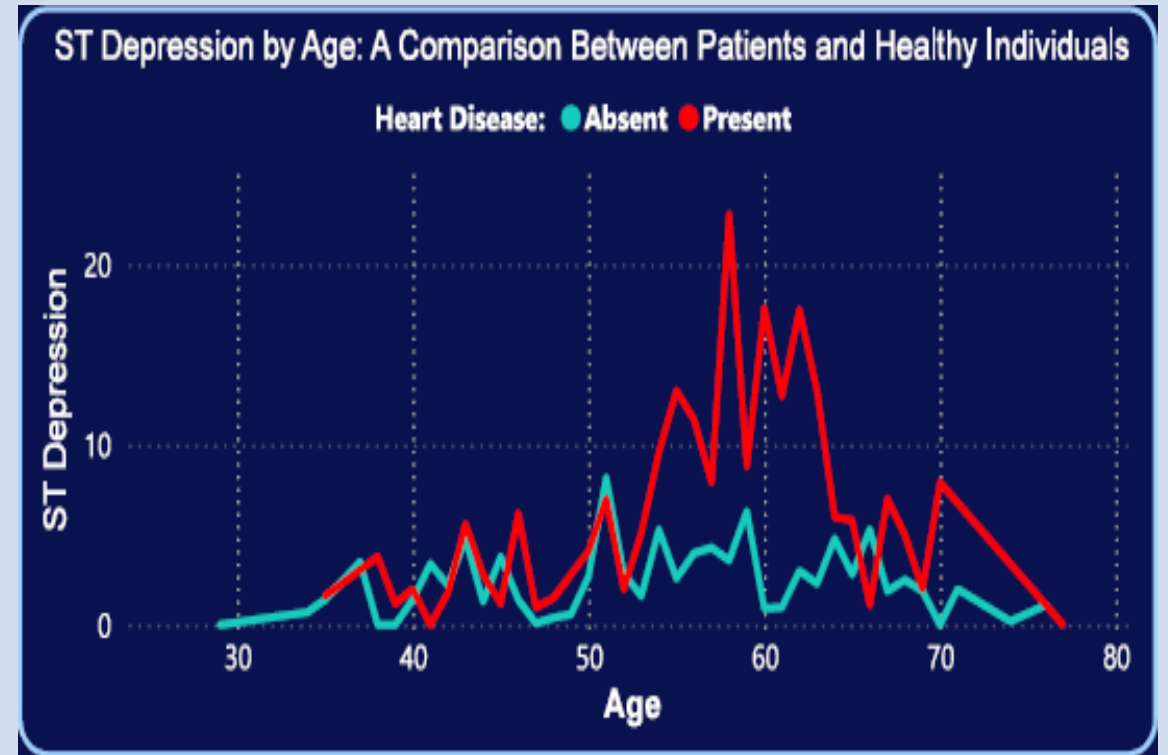
- The elderly age group shows a significantly high incidence of asymptomatic chest pain.

Asymptomatic chest pain refers to chest pain that does not cause or display any noticeable symptoms of heart disease.

## Additional Symptoms Associated with Heart Disease



- We can observe a rise in blood pressure between age 50 and 60, which persists through the 70s.
- Similarly, cholesterol levels and maximum heart rate also increase in the 50 to 60 age group.



- ST depression is an electrocardiogram (ECG) finding where the ST segment trace appears abnormally lower than the baseline.
- The plot indicates that ST depression is more pronounced in the age groups of 30-40 and near about 55.

# Key Performance Indicators (KPIs):

1. Percentage of People Having Heart Disease
2. Age Distribution including Gender
3. Gender Distribution Based on Heart Disease
4. Chest Pain Experienced by People Suffering from Heart Disease Especially Asymptomatic Chest Pain.
5. Blood Pressure, Cholesterol Level and Maximum Heart Rate of People According to their Age and Heart Disease Patients.
6. ST Depression Experienced by People According to their age and heart disease.

## Conclusion:

- ☐ Around 46% of people are suffering from heart disease.
- ☐ A higher proportion of elderly men fall within the 50 to 60-year age group, while a greater number of elderly women are in the 55 to 65-year age range.
- ☐ Males are more prone to heart disease.
- ☐ Elderly people are more prone to heart disease.
- ☐ People having asymptomatic chest pain have a higher chance of heart disease.
- ☐ The high number of cholesterol levels in people having heart disease.
- ☐ Blood pressure increases between the ages of 50 to 60 and somehow continues till 70.
- ☐ Cholesterol and maximum heart rate Increase in the age group of 50-60.
- ☐ ST depression is more pronounced in the age groups of 30-40 and near about 55.

# Q & A

**Q1) What is the source of the data?**

**Ans:** The dataset was sourced from the project description document provided by iNeuron.

**Q2) What type of data was used?**

**Ans:** The data comprised a combination of numerical and categorical values.

**Q3) What was the complete flow you followed in this project?**

**Ans:** Please refer to slide 5 for a comprehensive understanding of the workflow.

**Q4) What techniques were employed for data processing?**

**Ans:**

- Removal of unwanted attributes.
- Visualization of relationships between independent variables and output variables.
- Identification and treatment of outliers.
- Data cleaning and imputation of null values.
- Conversion of numerical data into categorical values.

**Q5) What libraries did you use in Python?**

**Ans:** I utilized the Pandas, NumPy, Matplotlib, and Seaborn libraries within the Pandas framework.

Thank you!