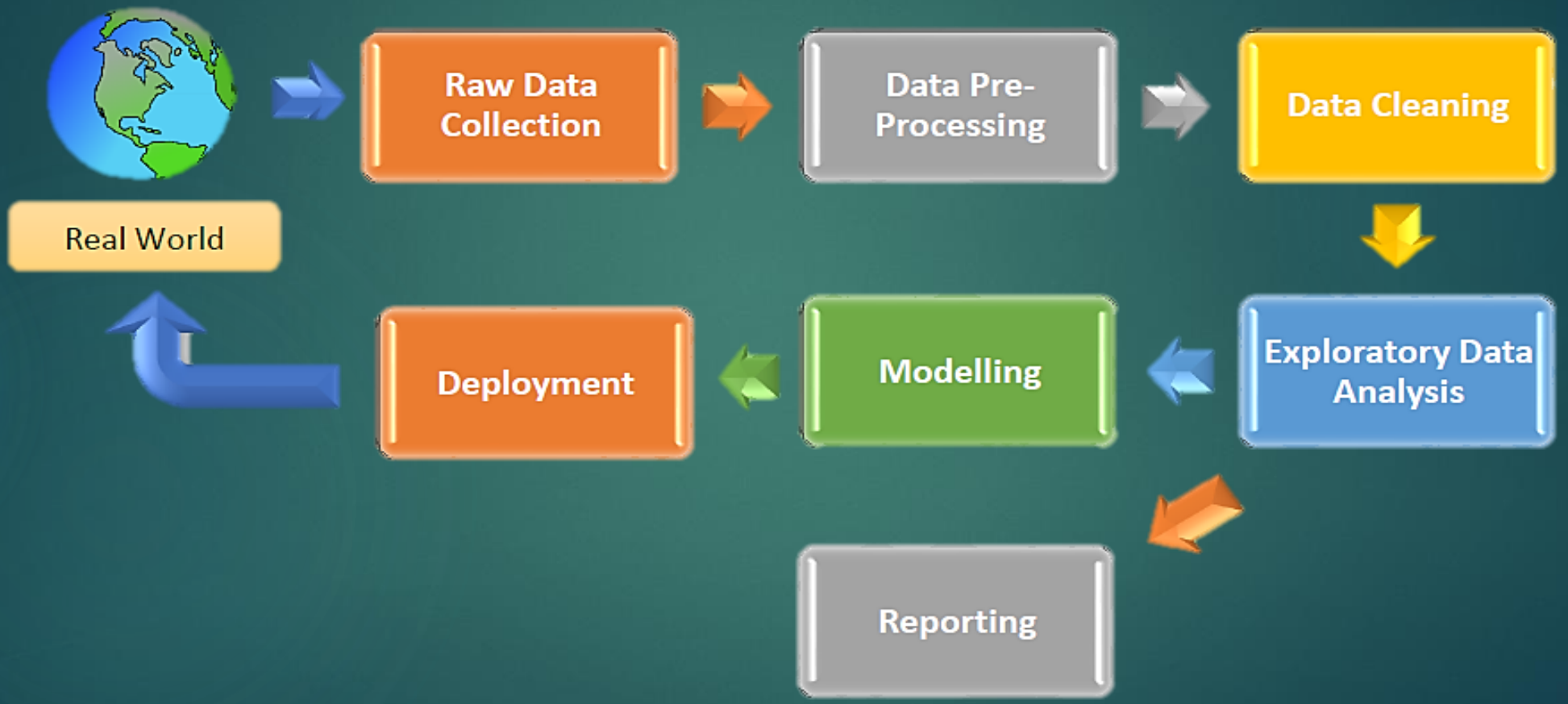
Heart Disease Diagnostics Analysis

Presented By:
Shivansh Kaushal
Chirag Sharma

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.
- The goal of this project is to analyze and to predict the probability of heart disease occurrence, based on a combination of features that describes the disease.
- To achieve the goal, we used a data set that is formed by taking into consideration some of the information of 303 individuals. The problem is based on the given information about each individual we have to calculate that whether that individual will suffer from heart disease or not.

Architecture



Dataset Description

- **age** : Age in years
- **sex** : Sex
  - 1 : Male
  - 0 : Female
- **cp** : Chest Pain Type
  - 1 : Typical Angina
  - 2 : Atypical Angina
  - 3 : Non-anginal Pain
  - 4 : Asymptomatic
- **trestbps** : Resting Blood Pressure (in mm Hg on admission to the hospital)
- **chol** : Serum Cholesterol in mg/dl
- **restecg** : Resting Electrocardiographic Results
  - 0 : Normal
  - 1 : Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria
- **fbs** : Fasting Blood Sugar > 120 mg/dl
  - 1 : True
  - 0 : False
- thalach : Maximum Heart Rate Achieved
- exang : exercise induced angina
  - 1 : Yes
  - 0 : No
- oldpeak : ST depression induced by exercise relative to rest
- slope : the slope of the peak exercise ST segment
  - 1 : upsloping
  - 2 : flat
  - 3 : downsloping
- ca : Number of major vessels (0-3) colored by fluoroscopy
- thal : Thalassemia
  - 3 : Normal
  - 6 : Fixed Defect
  - 7 : Reversable Defect
- num : Diagnosis of heart disease (angiographic disease status)
  - 0: < 50% diameter narrowing
  - 1: > 50% diameter narrowing

Features in Detail

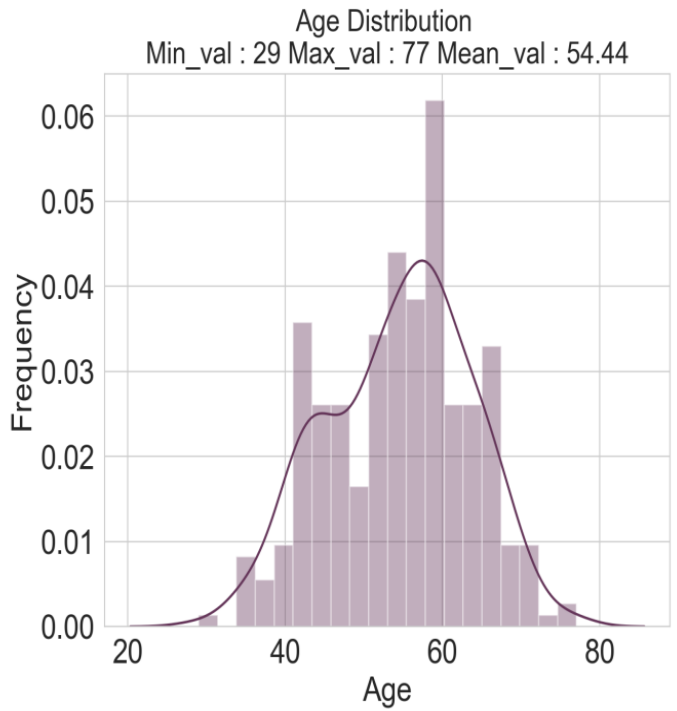
- **Age** : age plays a major role in the prediction of heart attack. As person becomes older the risk of damaged & narrowed arteries also increases. It weakens or thickens heart muscles that contributes to heart attack.
- **Gender** : men are generally at greater risk of heart disease. However, the risk for a women increases after menopause. The cause of disability & death in women is high after menopause. In the Cleveland dataset, 0 indicates male & 1 indicates female.
- **Chest pain (cp)** : The main symptom of heart attack is angina which is commonly called as chest pain. If the blood flow to the heart disease decreases, then the delivery of oxygen to heart muscles also decreases. Thus, there is a cause of discomfort, or painful feeling which is known as angina. The byproduct lactic acid builds up in the heart muscle because of less efficiency of heart when chest pain occurs.
- **Resting Blood Pressure** : Your blood pressure is recorded as two numbers: a) Systolic blood pressure (the first number) – indicates how much pressure your blood is exerting against your artery walls when the heart beats. b) Diastolic blood pressure (the second number) – indicates how much pressure your blood is exerting against your artery walls while the heart is resting between beats
- **Serum Cholesterol** : The cholesterol is classified as Low density lipoprotein (LDL) which is considered as Bad Cholesterol, High density lipoprotein (HDL) cholesterol which is generally considered as God Cholesterol, triglycerides and total cholesterol. The combined values of LDL, HDL and other lipids are called total cholesterol. The appropriate value of total cholesterol is less than 200 mg/dL. In Cleveland data set, serum cholesterol ranges from 160-410 mg/dL

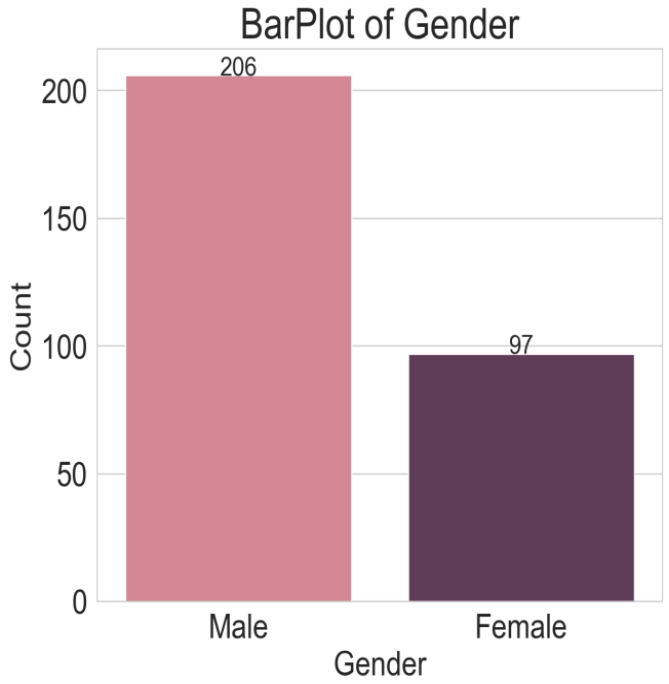
Features in Detail (Contd.)

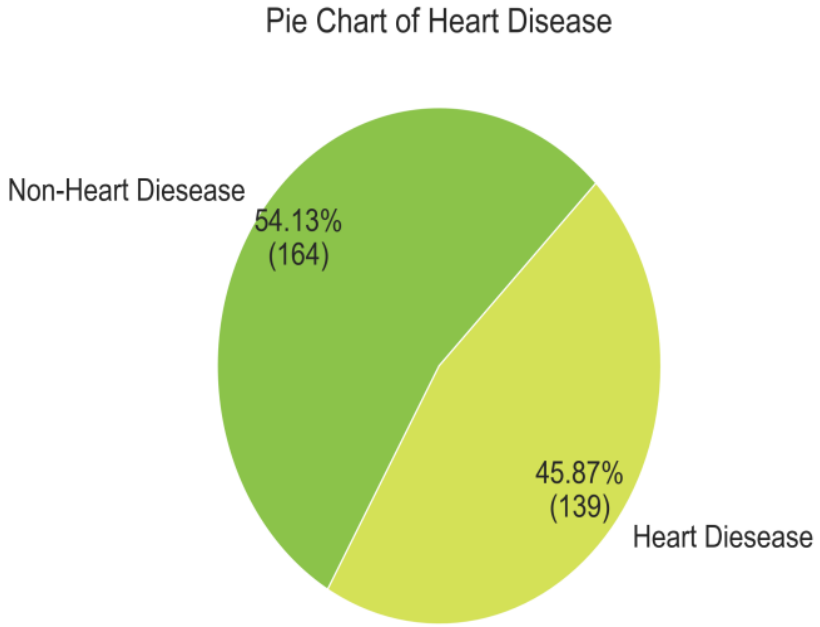
- **Fasting blood sugar** : Increased blood sugar is well-defined as fasting blood glucose of greater than 125 mg/dl. Diabetes increases the risk of heart attack. The role of increased heart attack is the same as that of other risk factors high cholesterol level & high blood pressure
- **Maximum heart rate (beats per minute)** : A normal heart rate is 50 to 100 beats per minute. Heart rate of patients is measured.
- **Exercise induced angina** : The exercise stress test called treadmill test is the most common method to diagnose the patients with suspected heart disease. In this test, the doctors monitor ECG along with the patients running on treadmill. This is done to identify the problems in heart because while exercising it is easy to identify in a patient. Reason is if anyone has 70-80% blockage in their arteries then it means the artery is supplying blood to the heart only 20-30% of the blood which is sufficient at rest. But when one is involved in any physical exercise the demand for blood increases and supply will not be sufficient to match the demand. This gives the symptoms of heart disease.
- **Thallium Test :** A thallium stress test is a nuclear imaging test that shows how well blood flows into the heart during exercise and at rest. In this test, thallium is carried to the bloodstream of the entire body and the perfusion defects are identified. People having perfusion defects during exercise have the high risk of MI.. In the HD data set, the normal value obtained in thallium stress test is coded as '3', fixed defect is encoded as '6' and reversible defect is encoded as '7'.
- Angiogram results : Angiography is a medical imaging method used to visualize the organs of the body like arteries, veins and the heart chambers. In this method, a radio opaque contrast agent is injected into the blood vessel and then imaging using fluoroscopy. From the obtained coronary angiogram, the number of major vessels (0-3) colored by fluoroscopy can be measured. The more number of vessels are are colored in number that indicates the severe MI.

Insights

Distribution of our Population over Gender and Age



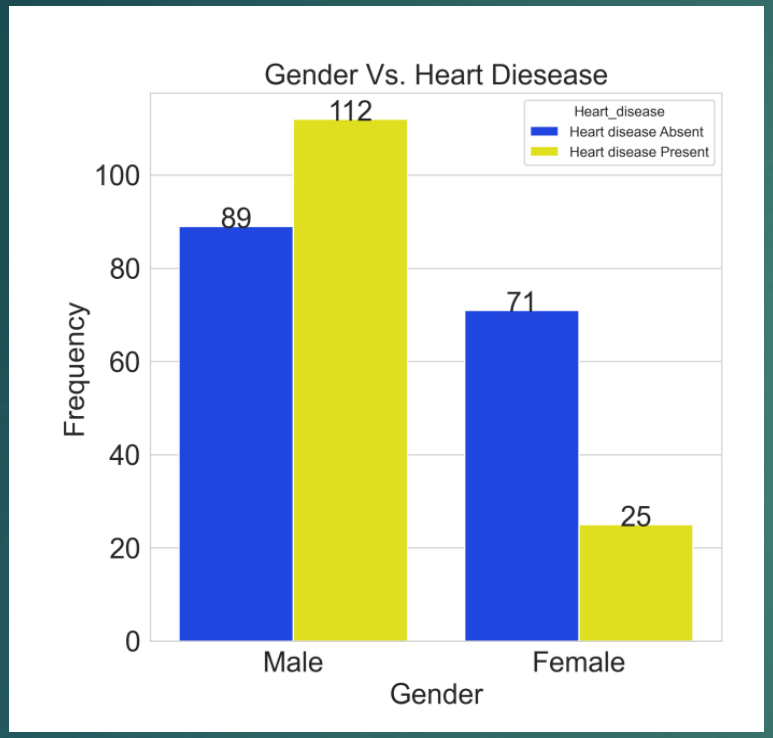




- Approximately 54% & 46% of the patients are Non-heart disease patients and Heart disease patients.
- Younger as well as older people are present in the dataset. Males are more in number than Females.

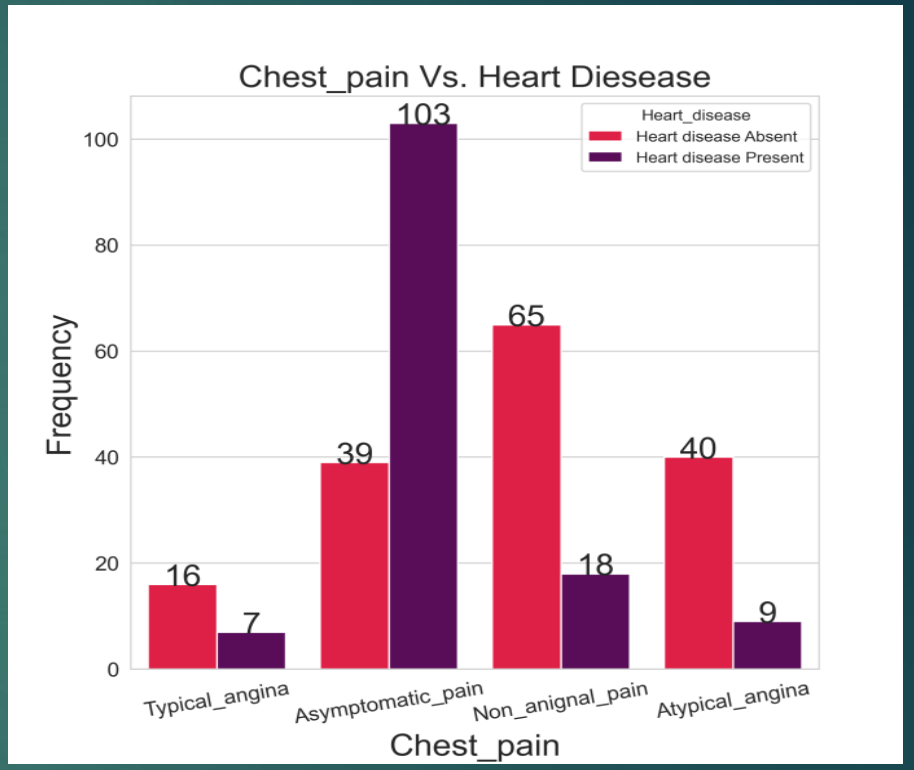
Insights (contd.)

Heart Disease vs Gender



- Both Men & Women are at greater risk of heart disease but the number for males is higher overall (indicated in below image). Gender is very important feature since the Critical value was coming very high using chi-square test.

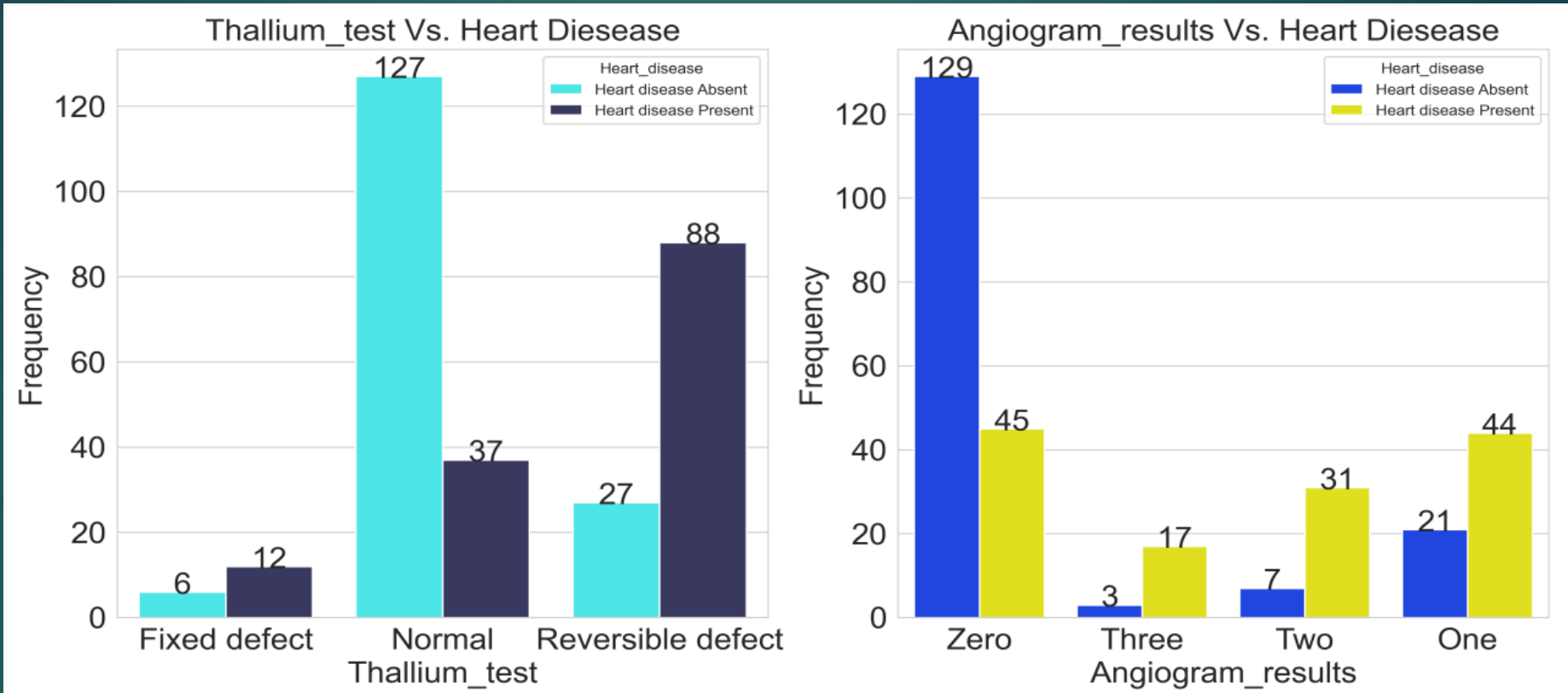
Heart Disease vs Chest Pain



- Asymptomatic chest pain is the main cause of heart disease among all the types of chest pains. Chest pain is a very important variable which was indicated by high chi-square critical value.

Insights (contd.)

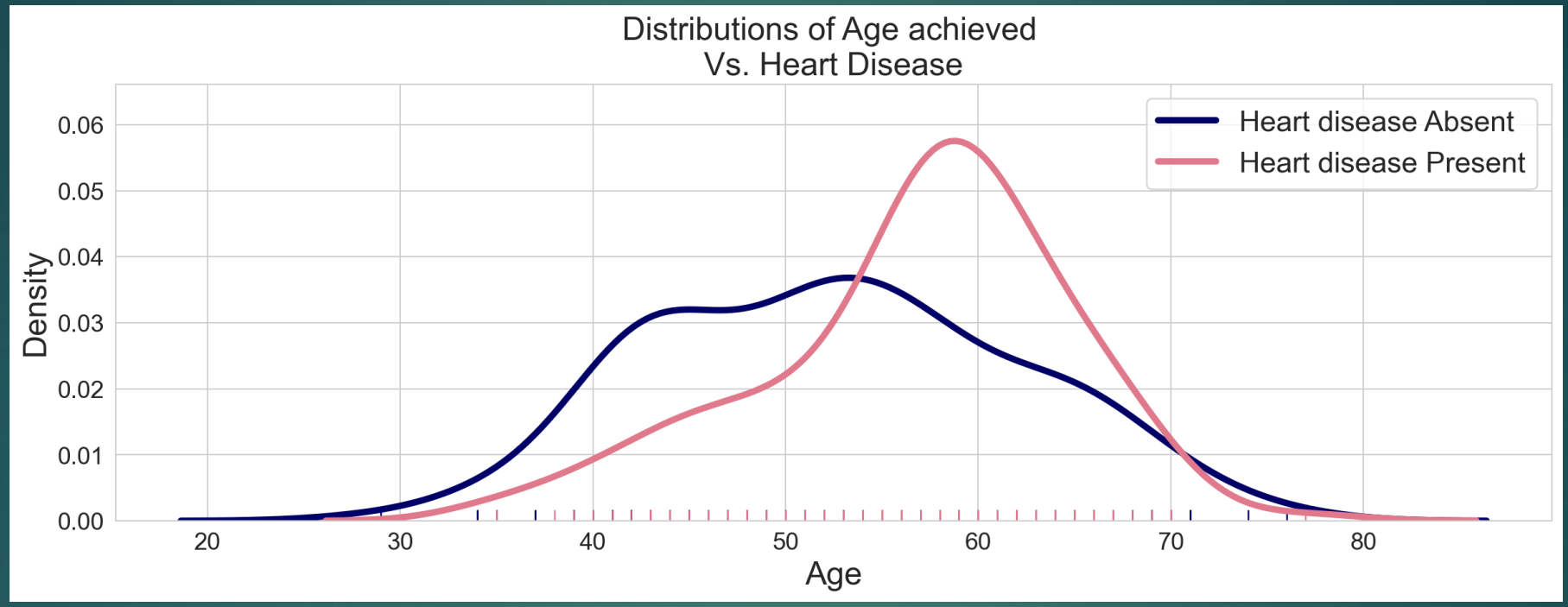
What do Angiogram results & Thallium test say about heart disease?



- We notice that more number of non-heart disease patients with zero/no defect in their major vessels that supply blood, oxygen & nutrients to the heart than heart disease patients. But as the number of major vessels blocked increases we observe more number of heart diseases patients than with non-heart disease patients. These patients need an immediate attention (indicated by angiogram results). We observe a greater number of heart disease patients with reversible defect (indicated by Thallium test). And both features are important indicated by a very high chi-square critical value.

Insights (contd.)

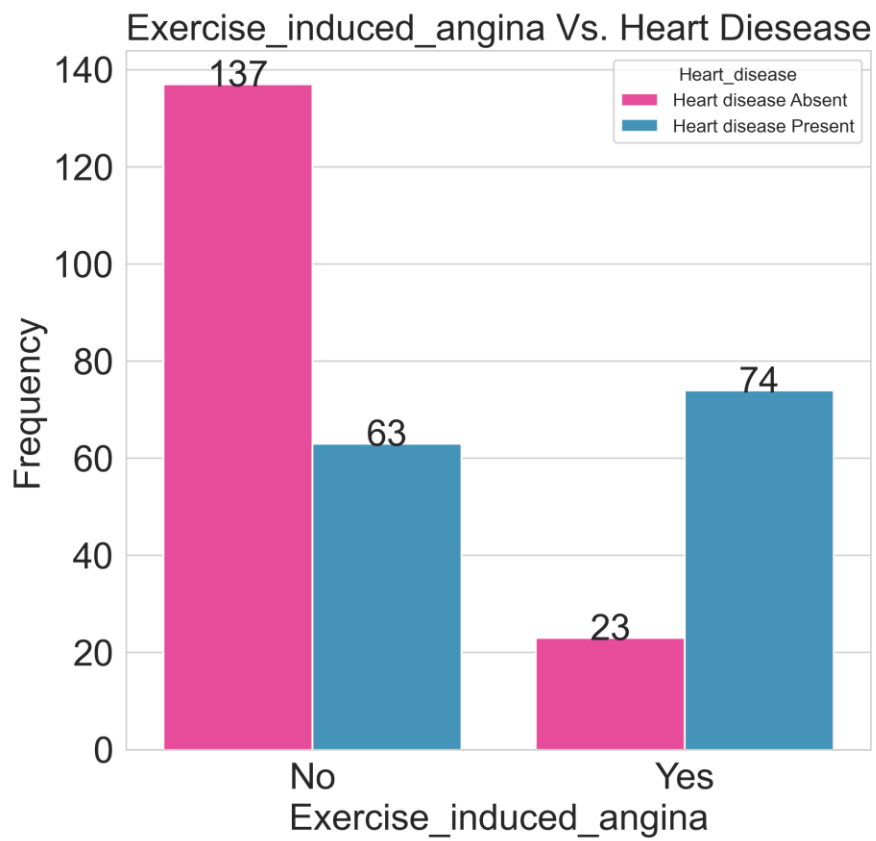
Heart Disease vs Age of Patient

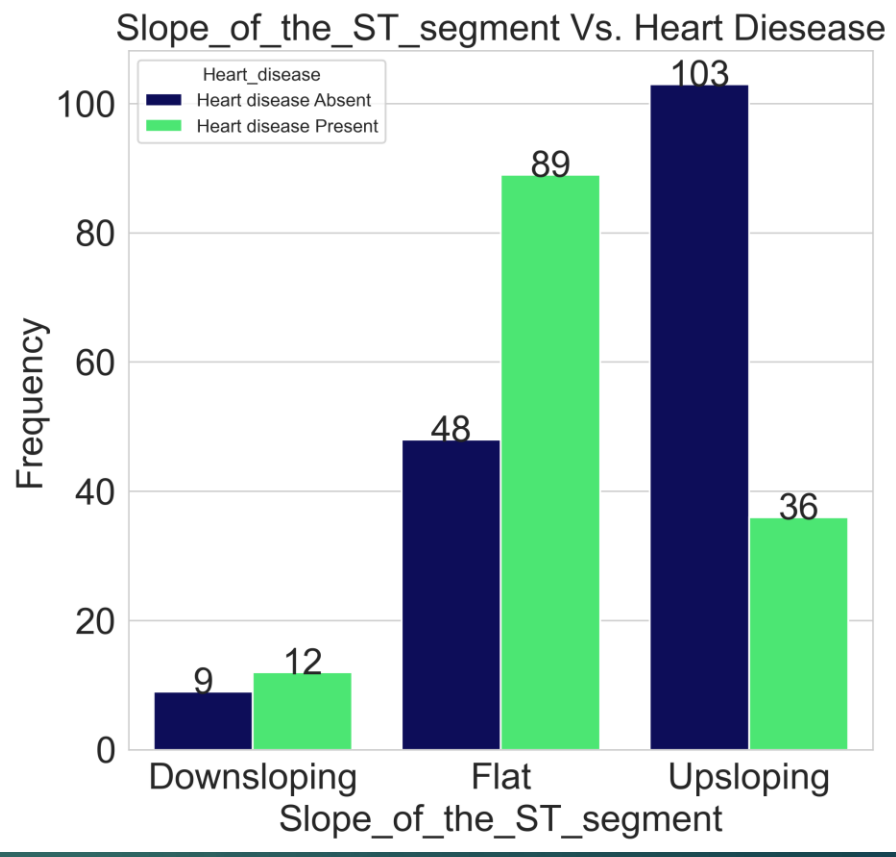


- We can clearly see that within 50 to 70 years old age more number of heart disease patients than with age less than 50 years old (indicated in above plot). We have broken down the Age variable into 2 categories/groups/bins, i.e., age less than 50 years & age greater than 50 years.

Insights (contd.)

Heart Disease vs Results of Treadmill Test and Slope of ST Segment





Insights (contd.)

- From previous two slides we can say:
- More number of heart disease patients experience angina while doing treadmill test, whereas less number of heart disease patients are experiencing angina at rest. Since the chi-square critical value is coming very high, therefore, we conclude Exercise induced angina to be an important feature.
- Slowly upsloping ST segment usually indicates heart attack. Horizontal ST Segment depression is considerable abnormal response. Down sloping ST Segment depression represents severe heart attack. We observe more number of patients having abnormal Slope of the ST segment (indicated by Flat slope of the ST segment). Number of patients with down-sloping Slope of ST segment is considerably very less than Flat slope segment patients.
- We found Age group variable to be important which was indicated by chi-square test of independence.
- We also calculated the percentage of patients having heart disease given age is above 50 years old & percentage of patients not having heart disease which were coming to be roughly 46% & 53& respectively. The percentage of patients having heart disease & patients not having heart disease were found to be roughly 31% & 69% given the age is less than 50 years old.
- These observations conclude that as a person becomes older the risk of damaged & narrowing arteries also increases. It also weakens or thickens heart muscles that contributes to Ischemic heart disease & thus lead to heart attack. We also see that patients with Non-heart disease are higher than heart disease patients but with not much greater difference

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
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

- Gender, Age, Cholesterol, Resting blood pressure, Maximum heart rate, Chest pain & some test results like Thallium test, Angiogram results, Exercise induced angina, slope of ST segment are the important factors in determining the Heart disease.
- As people becomes older they need to maintain their blood pressure, cholesterol level, Heart rates and they should visit to a doctor as well to get check their health check. To avoid any heart disease they can do some of the following things like avoid smoking, do exercise, avoid high fat consumption diet and adopt low fat diet, eat raw green vegetables, maintain their stress level. In short, people should change their lifestyle & adopt healthy habits.

Q n A

Q1) What's the source of data?

Ans) The Dataset was taken from iNeuron's Provided Project Description Document.
https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA22LW0Lt3Y-Q8

Q2) What was the type of data?

Ans) The data was the combination of numerical and Categorical values.

Q 3) What's the complete flow you followed in this Project?

Ans) Refer slide 3rd for better Understanding

Q4) What techniques were you using for data?

Ans) - Removing unwanted attributes
- Visualizing relation of independent variables with each other and output variables.
- Removing outliers
- Cleaning data and imputing if null values are present.
- Converting Numerical data into Categorical values.

Q5) What were the libraries that you used in Python?

Ans) I used Pandas, NumPy and Matplotlib and Seaborn libraries in Pandas.

Thank You