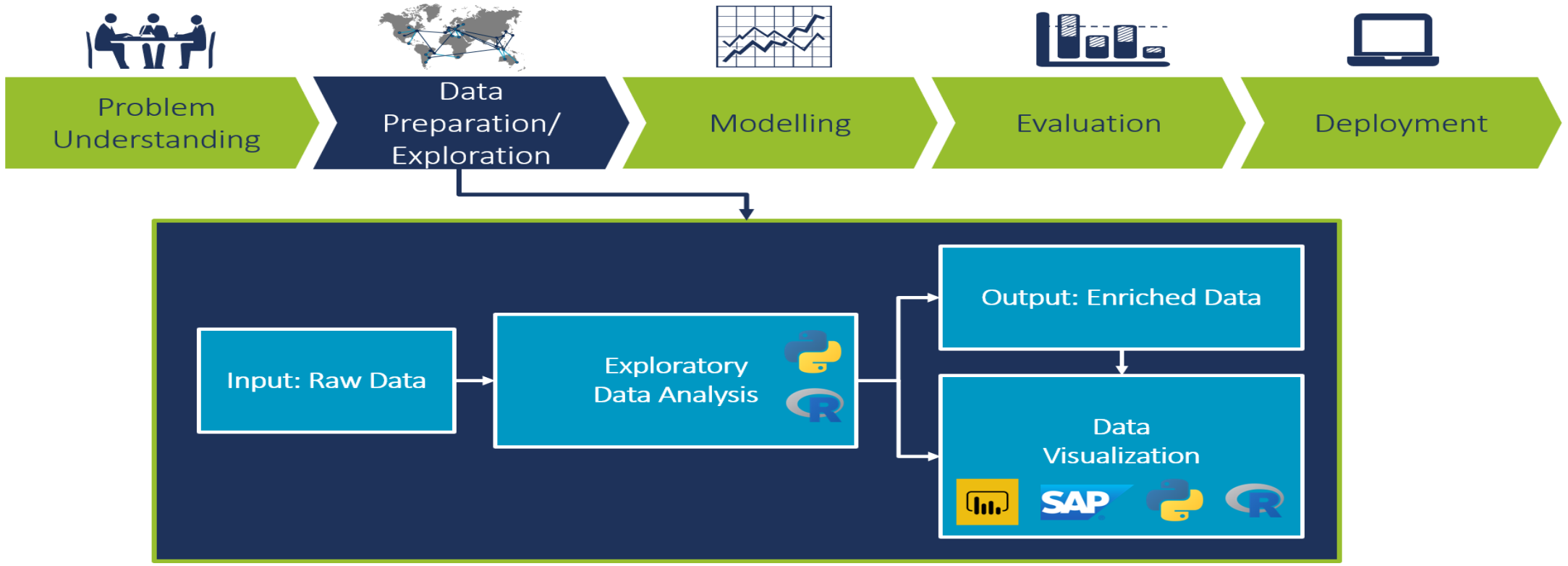
Heart Disease Data Analysis

Presented By:
Ashish Haldar

Objective

- Heart disease is perceived as the deadliest disease in the human life across the world. In particular, in this type of disease the heart is not capable in pushing the required quantity of blood to the remaining organs of the human body in order to accomplish the regular functionalities
- The main objective of this project is to analyze the data and to find out the key features which is contributing more to the occurrence of heart disease.
- The achieve this objective, we have used a data set consisting of information of 303 patients. The problem is based on the given information about each individual we have calculate that whether that individual will suffer from heart disease or not.

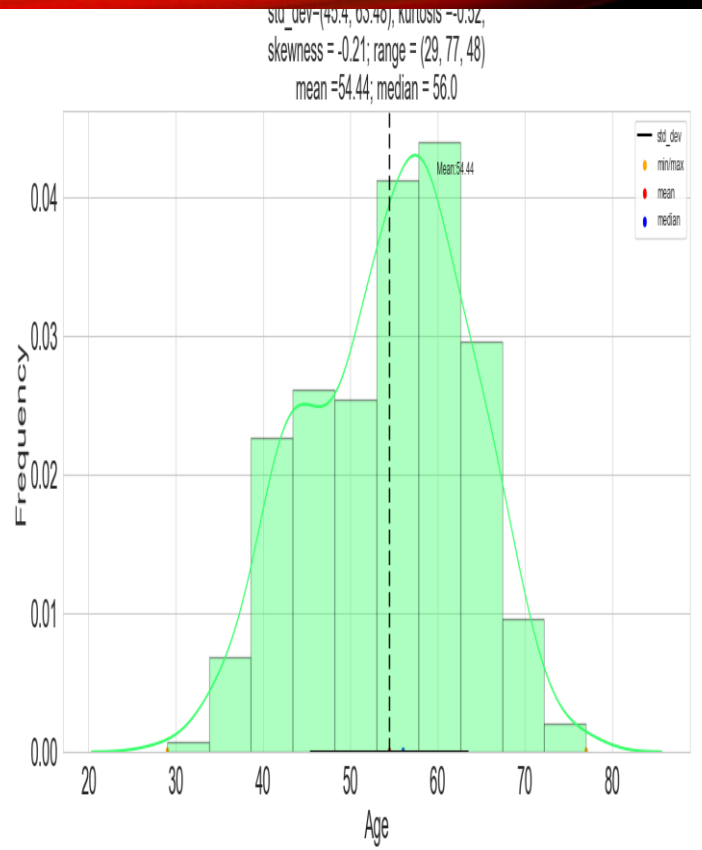
Architecture

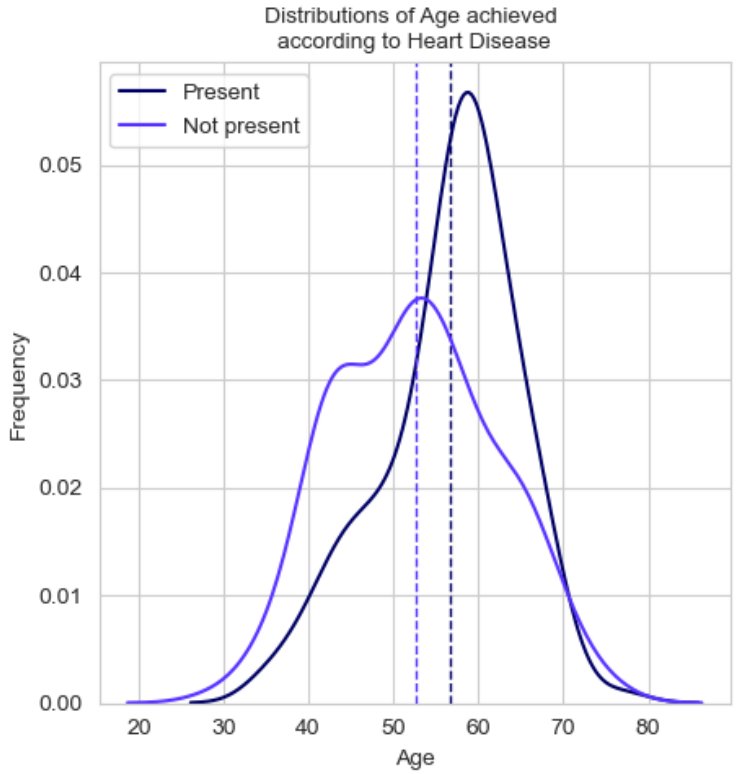


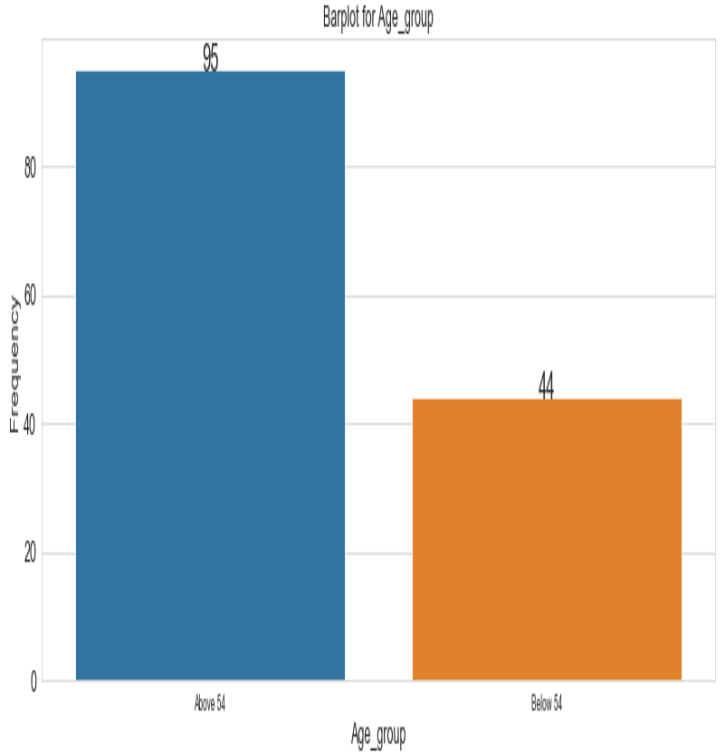
Dataset Description

Sl.No.	Feature Name	Feature Code	Description	Domain of value
1	Age	Age	Age of the person in years	28<Age<78
2	Gender	Sex	1 = Male 0=Female	1 0
3	Type of chest pain	Chest_pain	1. Atypical angina 2. Typical angina 3. Asymptomatic 4. non-angina pain	1 2 3 4
4	Resting blood pressure	Resting_BP	Mm Hg	94 to 200
5	serum cholesterol	cholesterol	Mg/dl	126 to 564
6	Fasting Blood Sugar	Blood_sugar	fasting blood sugar > 120 mg/dl	1=True 2=False
7	resting electrocardiographic result	Resting_ECG	0 = Normal 1=ST-T wave abnormality 2=Left Ventricular Hypertrophy	
8	maximum heart rate achieved	Heart_rate		71 to 202
9	exercise induced angina	Exer_induced_agina	1=Yes 0=No	
10	ST depression induced by exercise relative to rest	ST_depression		0 to 6.3
11	the slope of the peak exercise ST segment	Slope_STsegment	1= Up-sloping 2=Flat 3= Down-sloping	
12	number of major vessels (0-3) colored by flourosopy	Coronary Angiogram	0,1,2,3	
13	Thalium	Thal	3=normal 6=fixed defect 7=reversible defect	

Some Insights

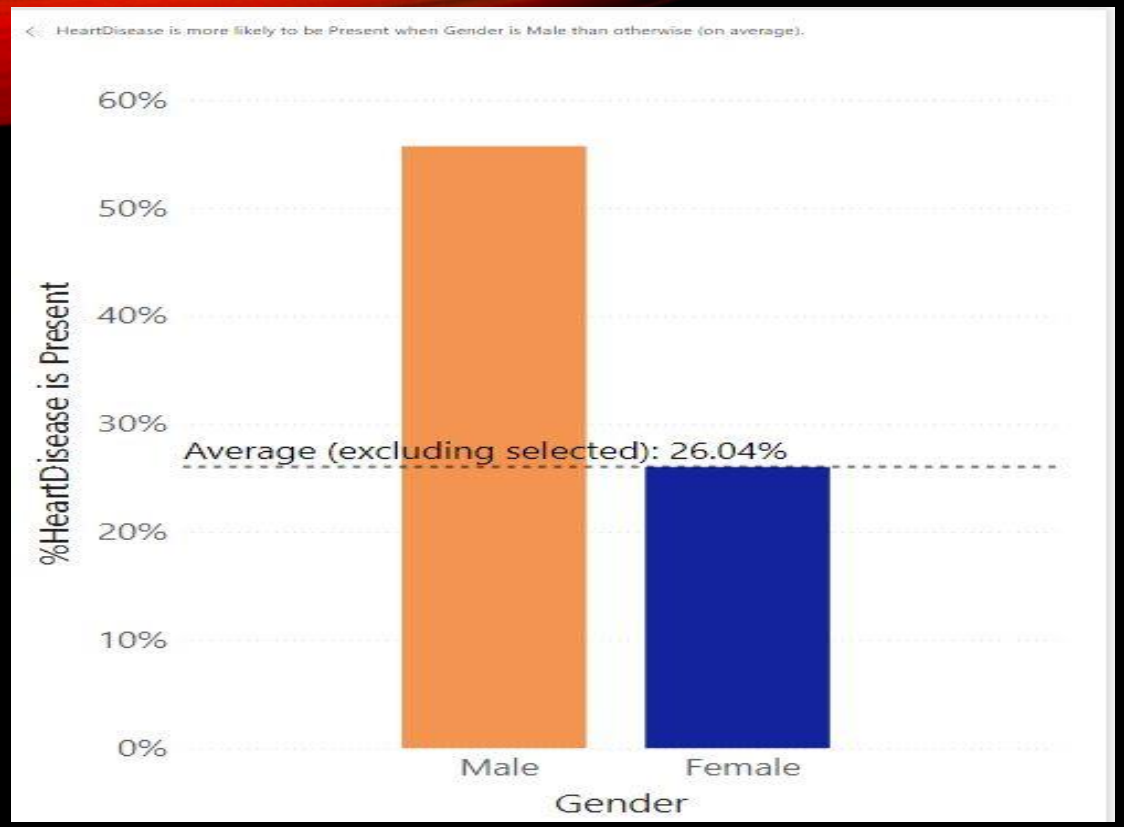


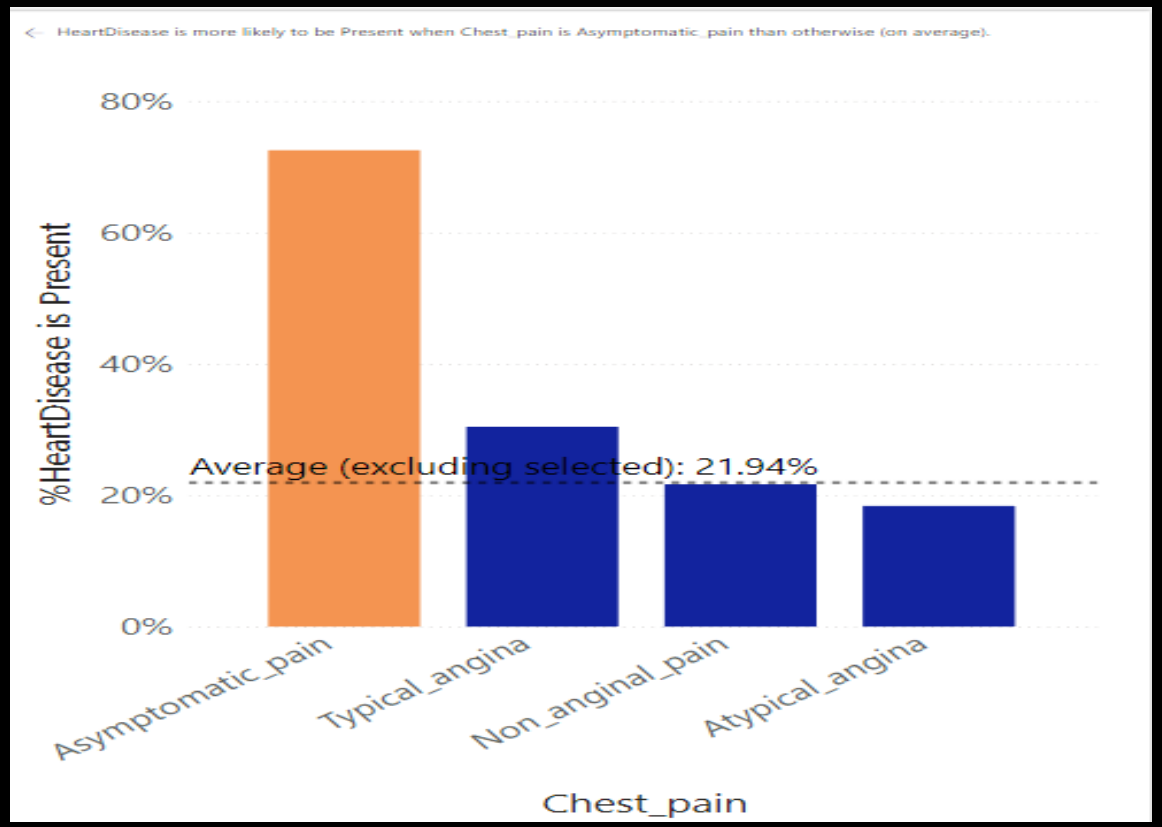




- People with age above 54(mean) is more prevalent to heart disease compared to below 54.
- Most of the people are in the age group of 45 to 63
- Distribution of Age based on whether disease is present or not is same i.e. their mean is following same distribution

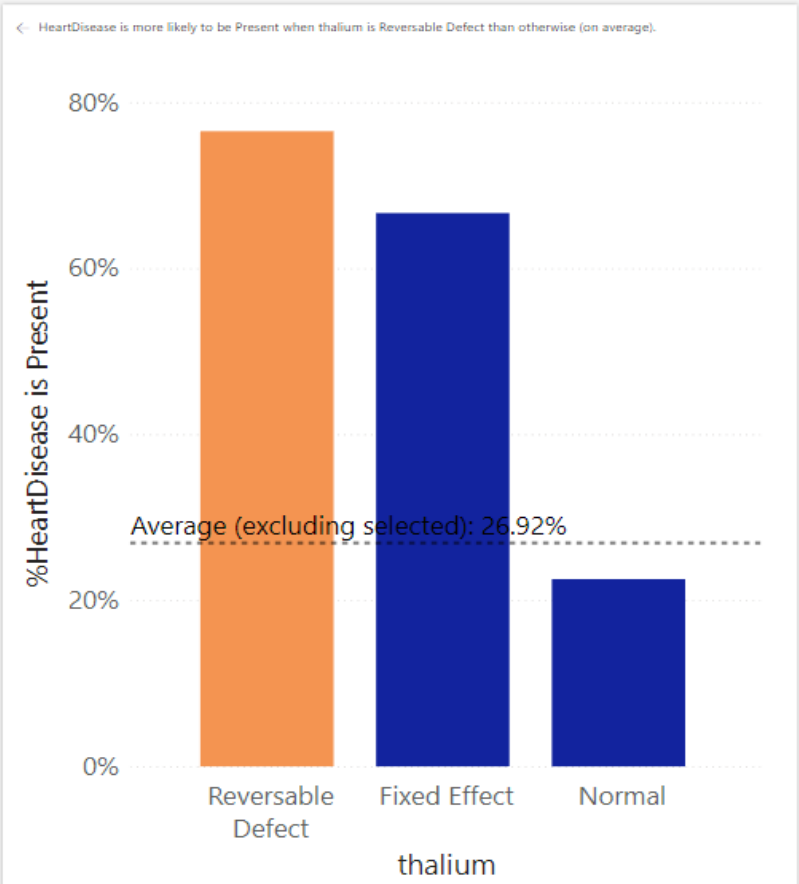
Some Insights(Cont.)

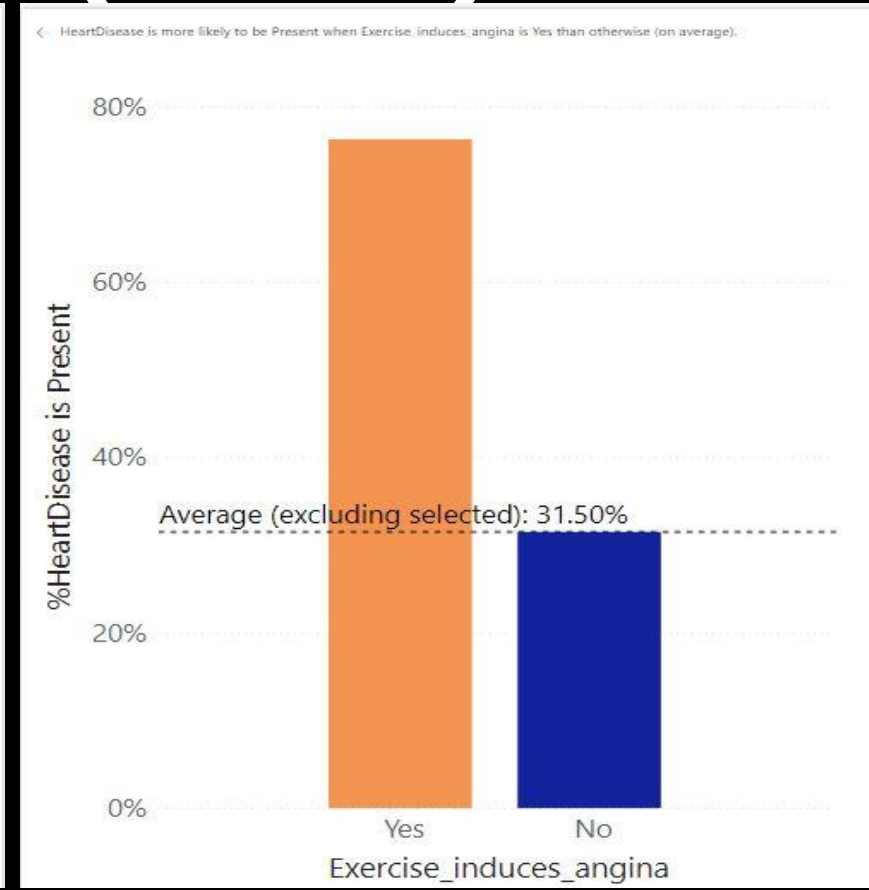


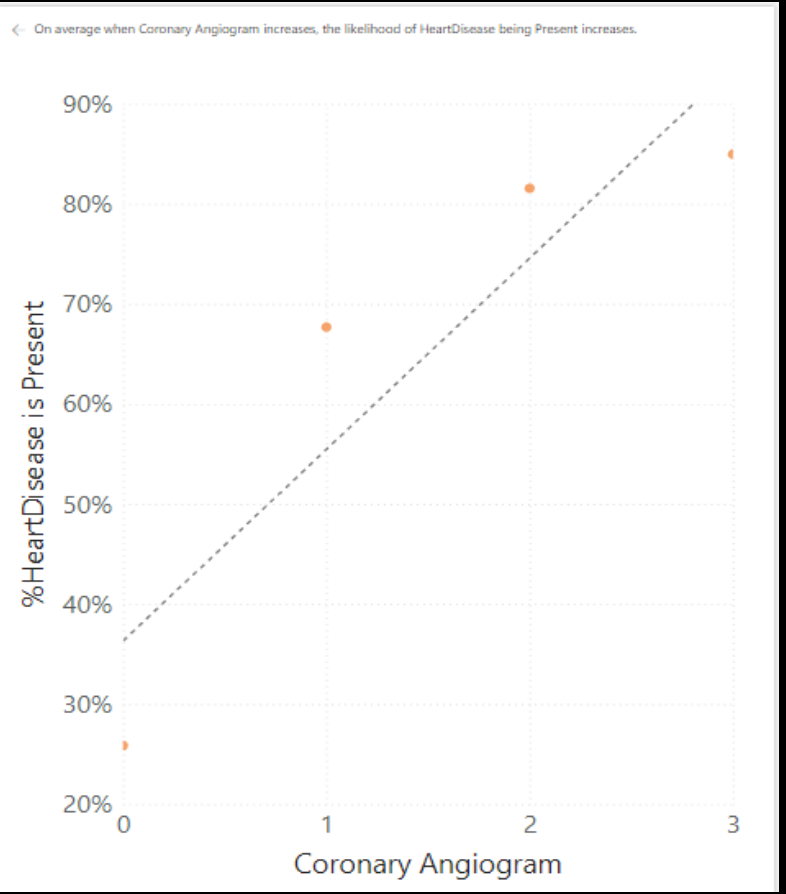


- Heart Disease is more likely to be present when Gender is male than otherwise(In average)
- When chest pain type is asymptomatic, heart disease is 3.31 time more likely to present compared to all other values of chest pain and this is even more when the person has blood sugar i.e. 5.06 times more likely.

Some Insights(Cont.)

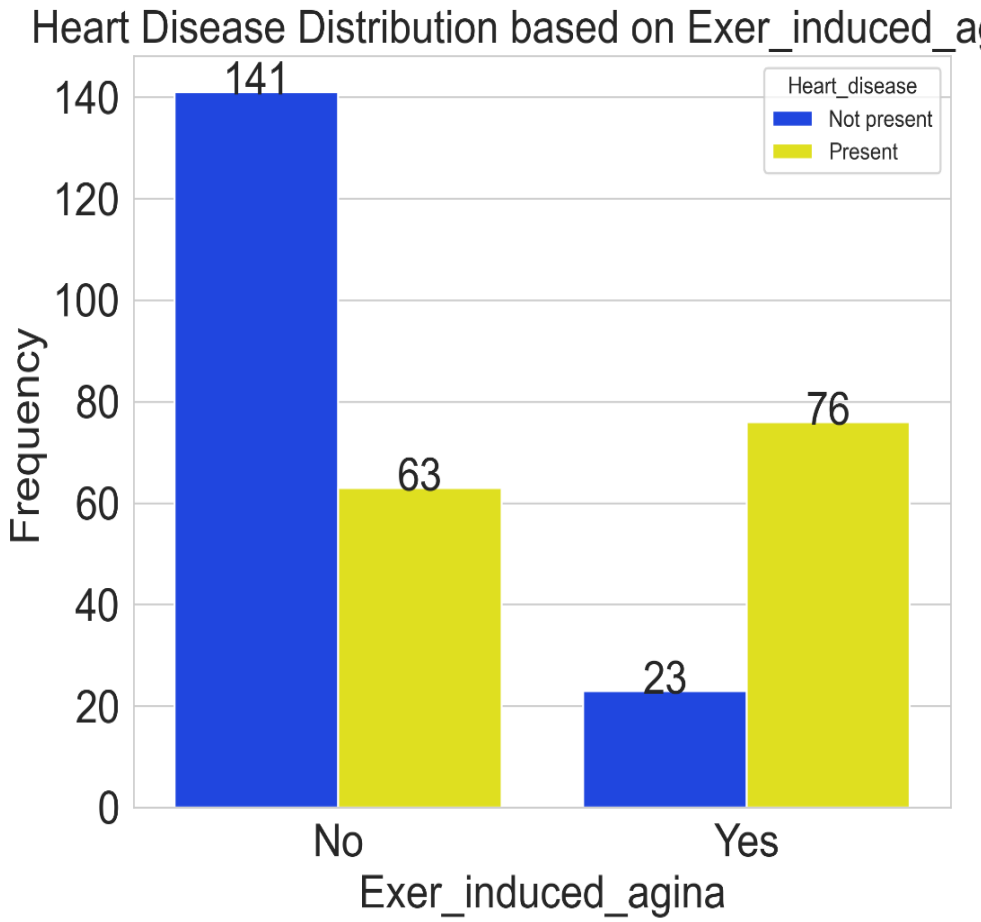


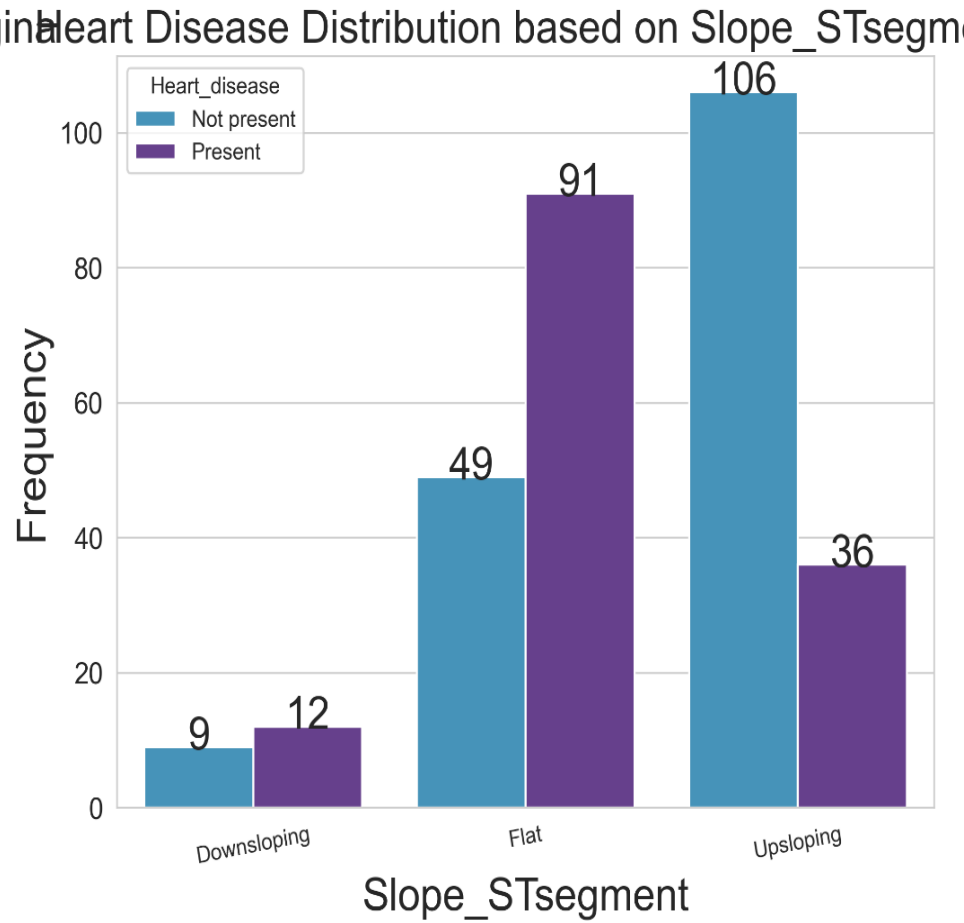




- When thallium is reversible defect the likelihood of heart disease being present increase by 2.84 times.
- When exercise induced angina is yes then the likelihood of heart disease being present increased by 2.42 times.
- On average when coronary angiogram increases the likelihood of heart disease being present increases.

Some Insights (Cont.)





Some Insights (Cont.)

- More number of heart disease patients experience angina while doing treadmill test, where as 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 import factor.
- Slowly up sloping 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

- Age distribution including gender
- Gender distribution based on Heart disease
- Chest pain type based on heart disease
- Exercise induced angina according to heart disease
- Blood Pressure, Cholesterol Level and Maximum Heart Rate of People According to their Age and Heart Disease Patients.
- ST Depression Experienced by People According to their age and heart disease

Conclusion

- From the above analysis it is referred that Age, maximum heart rate, Blood Pressure, cholesterol plays a vital role in influencing the heart disease more along with some important symptoms like **if a person has Asymptomatic pain, gender is male, thallium has reversible defect, Slope_Stsegment is Flat and Exercise_induced agina is yes then chances of a person being affected by disease is more compared to other symptoms/test.**
- Heart Stroke and vascular disease are the major cause of disability and premature death. Chest Pain is the key to recognize the heart disease. In this work, the heart disease are predicted by considering major factors with four type of chest pain.
- 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, Matplotlib, Seaborn and SciPy libraries in Pandas.