HEALTHCARE ANALYTICS
HEART DISEASE PREDICTION

Written By	Ankit Kashyap
Document Version	1.0
LastRevised Date	21-Dec-2021

PROJECT DETAIL

Project Title	HealthCare Analytics- Heart Disease Prediction
Technology	Business Intelligence
Domain	Healthcare
Project Difficulty level	Intermediate
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel, MS-Power BI

Objective:

"The goal of this project is to analyse the heart disease occurrence, based on a combination of features that describes the heart disease and develop an interactive Report/Dashboard based on the analysis.

Benefits:

➢ Detection of a heart disease at an early stage leads to prevention of more than 80% of potential related deaths

➢ Determine the relation and significance of the physical vitals in the development of a future heart disease.

➢ Create a patient profile for easy access of required information

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 analyse this health and medical data for better future preparation.

- A dataset is formed by taking into consideration some of the information of 303 individuals.

ARCHITECTURE
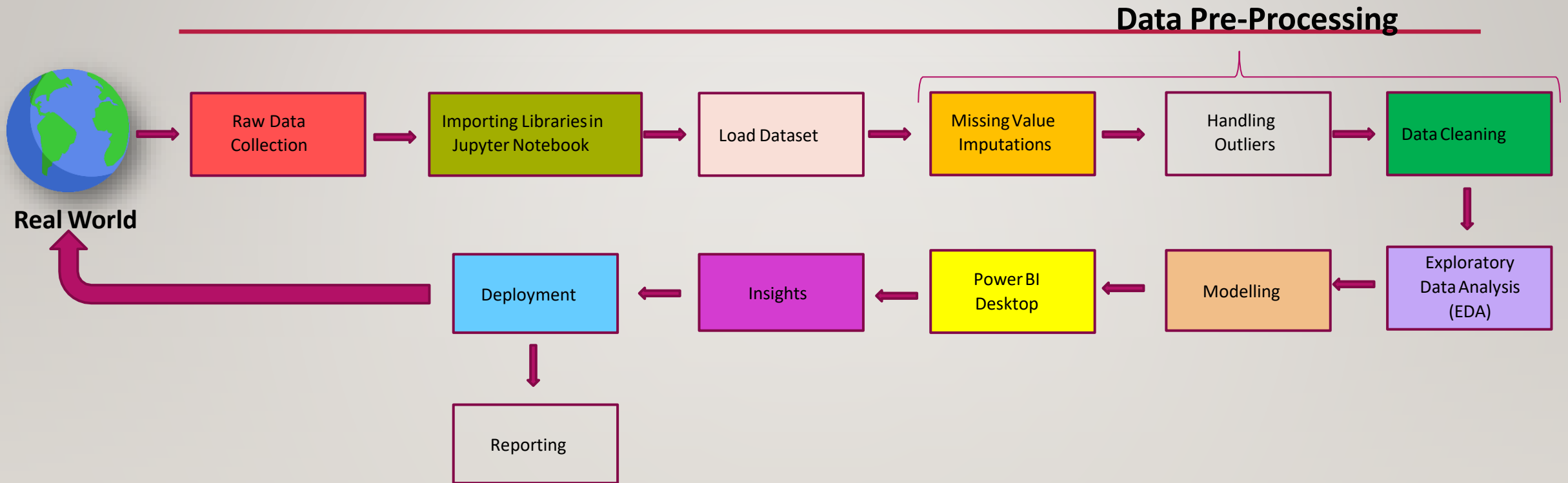
Data Description

age: The person's age in years

sex: The person's sex (1 = male, 0 = female)

cp: The chest pain experienced (Value 0: typical angina, Value 1: atypical angina, Value 2: non-anginal pain, Value 3: asymptomatic)

trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)

chol: The person's cholesterol measurement in mg/dl

fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)

restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)

thalach: The person's 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 (Value 0: upsloping, Value 1: flat, Value 3: down sloping)

- **ca:** The number of major vessels (0-3)

- **thal:** A blood disorder called thalassemia (0 = normal; 1 = fixed defect; 2 = reversable defect)

- **num:** Heart disease (0 = no, 1 = yes)

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 percent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.

Sex: 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 ups 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 exercise 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.
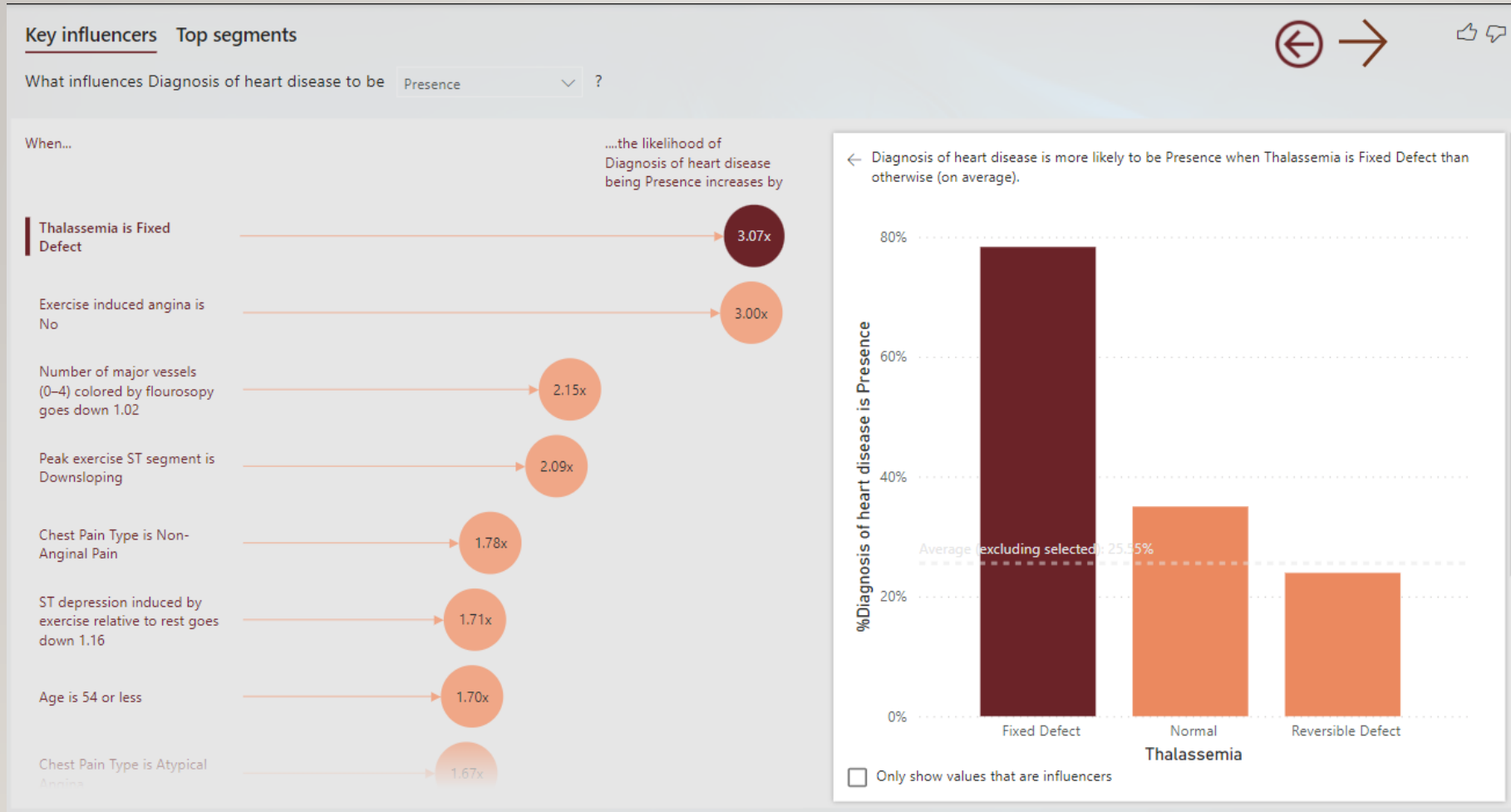
Observations/Insights
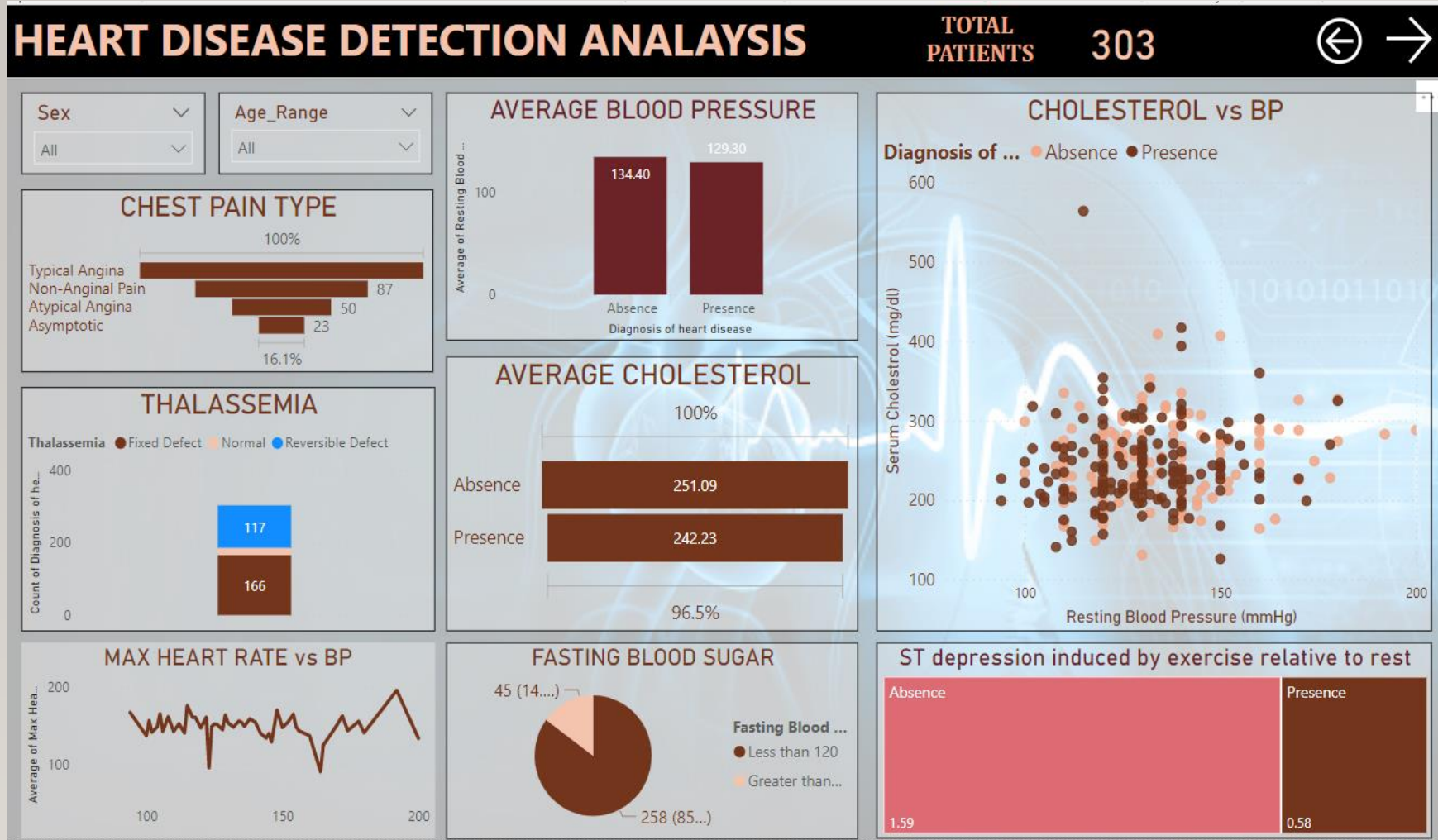
- Patients suffering from TYPICAL CHEST PAIN are more susceptible to have heart disease.
- Fixed Defect and Reversible defect form the bulk of patients.
- As BP increases the fluctuation in Max heart rate achieved becomes more erratic.
- Avg. BP and Avg. Cholesterol is not a significant factor for presence of heart disease.
- Cholesterol increases the likelihood of higher BP.
- ST Depression induced during exercise is 110% lower in presence of heart disease.

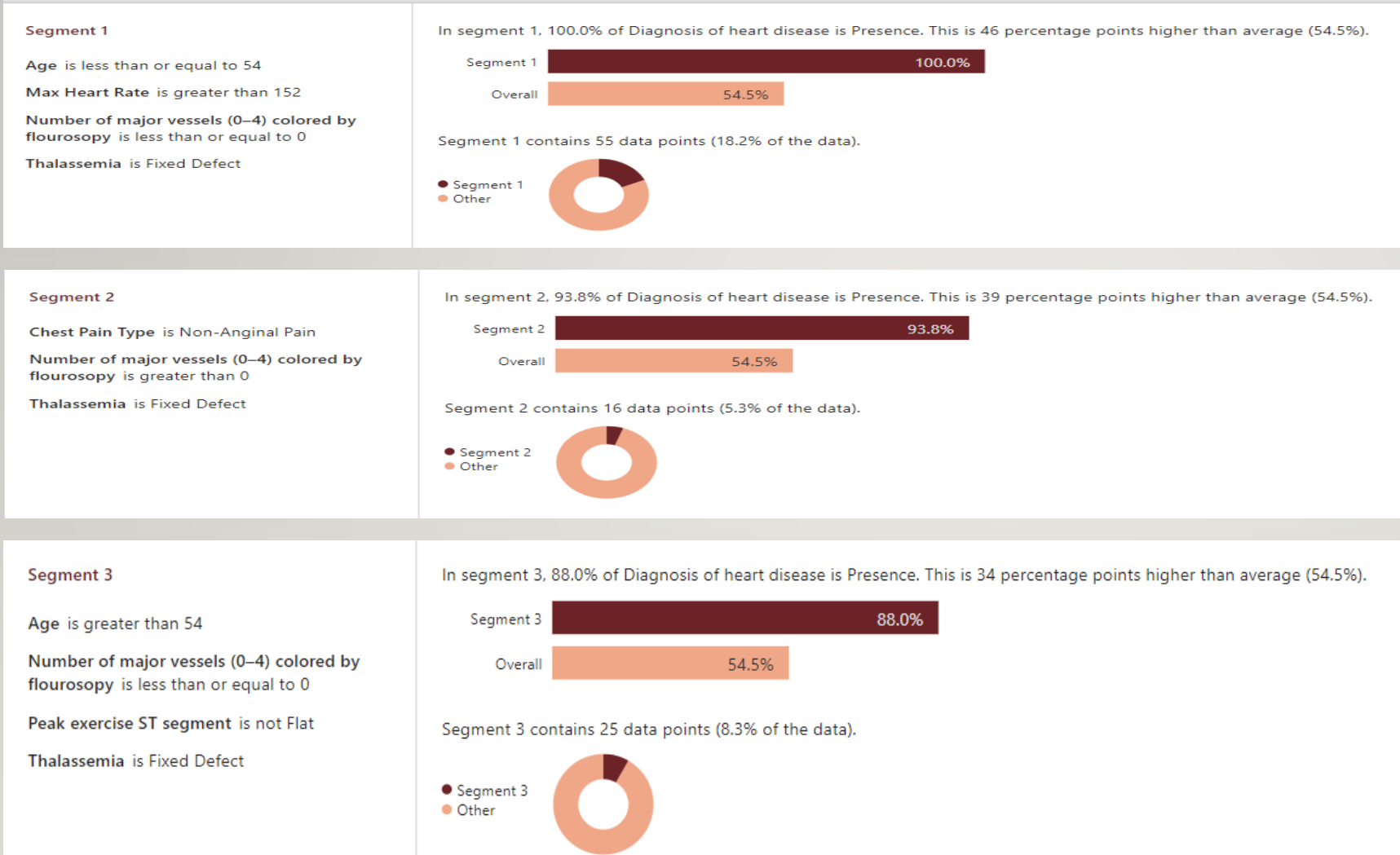# Observations/Insights

- Fixed Defect Thalassemia increases the risk of getting heart disease by 3.07 times.

- Also Fixed Defect Thalassemia has 40% higher chance of getting a heart disease as compared to other forms of Thalassemia.

Observations/Insights
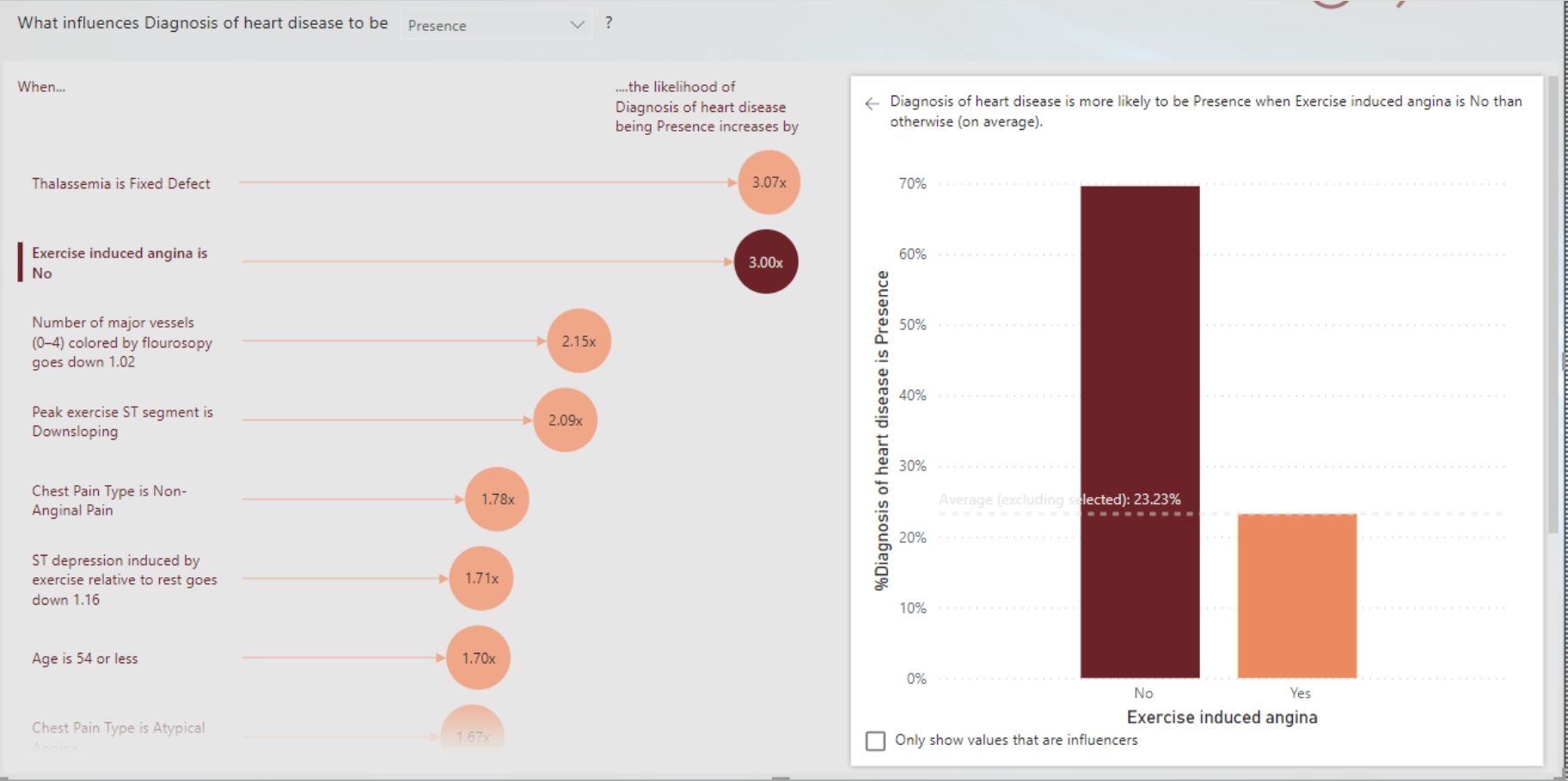
If the **patient has angina** and it is **not induced due to exercise,** then the risk of getting a heart disease is 3 times higher.

Observations/Insights

Segment 1

Age is less than or equal to 54

Max Heart Rate is greater than 152

Number of major vessels (0–4) colored by flourosopy is less than or equal to 0

Thalassemia is Fixed Defect

In segment 1, 100.0% of Diagnosis of heart disease is Presence. This is 46 percentage points higher than average (54.5%).

Segment 1	100.0%
Overall	54.5%

Segment 1 contains 55 data points (18.2% of the data).

● Segment 1
● Other

Segment 2

Chest Pain Type is Non-Anginal Pain

Number of major vessels (0–4) colored by flourosopy is greater than 0

Thalassemia is Fixed Defect

In segment 2, 93.8% of Diagnosis of heart disease is Presence. This is 39 percentage points higher than average (54.5%).

Segment 2	93.8%
Overall	54.5%

Segment 2 contains 16 data points (5.3% of the data).

● Segment 2
● Other

Segment 3

Age is greater than 54

Number of major vessels (0–4) colored by flourosopy is less than or equal to 0

Peak exercise ST segment is not Flat

Thalassemia is Fixed Defect

In segment 3, 88.0% of Diagnosis of heart disease is Presence. This is 34 percentage points higher than average (54.5%).

Segment 3	88.0%
Overall	54.5%

Segment 3 contains 25 data points (8.3% of the data).

● Segment 3
● Other

- The patient with the below features have all turned out to have developed heart disease:
  - age<**54**
  - Max heart rate **>152 bpm**
  - No. of vessels colored by flourosopy **=0**
  - Thalassemia is **fixed**

- The patient with the below features have 93.8% turned out to have developed heart disease:
  - Chest Pain type is **Non-Anginal**
  - No. of vessels colored by flourosopy **>0**
  - Thalassemia is **fixed**

- The patient with the below features have 88% turned out to have developed heart disease:
  - Age>**54**
  - Peak exercise ST segment is **NOT FLAT**
  - No. of vessels colored by flourosopy **=0**
  - Thalassemia is **fixed**

Observations/Insights
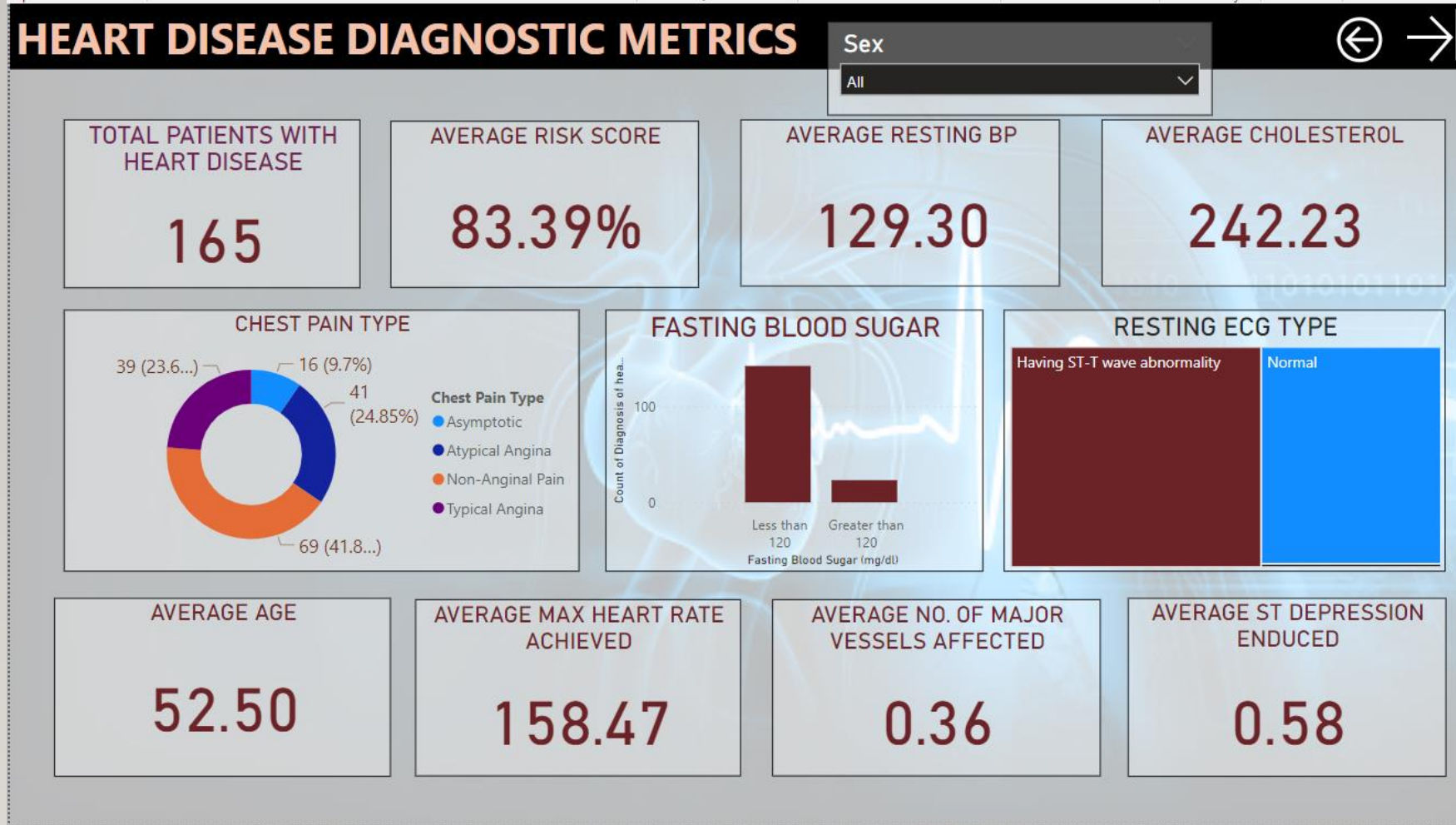
- Total Patients with heart disease is 165 (54.6%) of total 303 subjects.
- Average Risk of developing heart disease is 83.39%
- Average Resting BP for people with heart disease is 129.3 mmHg.
- Average Serum Cholesterol for people with heart disease is 242.33 mm/dl
- Average age for people with heart disease is 52.5 years
- Average max heart rate achieved for people with heart disease is 158.47 bpm
- Average no. of major vessels affected for people with heart disease is 0.36.
- Average ST depression induced due to exercise for people with heart disease is 0.58

KEY PERFORMANCE INDICATOR

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

1. From target value we can say that our dataset is amost balanced with 54% of patients having heart disease and 46 % of patients not having heart disease.

2) Females have higher chances of having heart disease than females.

3) Patients with age >55 years and having resting blood suger(i.e in diastolic state) in range 121-140 mm Hg have higher chances of heart disease. patients with age group 40 to 45 have little chances and age below 40 has negligiable cgances of having a heart disease.

4) Patients suffering from heart disease are mostly in age group of 55-65 years.

CONCLUSION

5) Higher cholesterol means higher chances of heart disease. it plays an important role in determining heart problems. With age cholesterol increases and level 200- 350 mg/dl are of concern.

6) Patients showing symmetric T wave in Resting electrocardiographic measurement are more likely to suffer from a heart disease.

7) Patients who are likely to suffer from heart disease have higher maximum heart rates(rate between 140-160) whereas patients who are not likely to suffer from heart disease are having lower maximum heart rates.

8) Exercise induced anginal pain slightly lower chances of getting heart disease than without exercise induced pain.

Q & 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/1M5z7z1NmWar7y1eFs67orfjqHL0iSViL?usp=sharing

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 5th 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 Categorical values into dummy variables to build the model.

Q 6) What were the libraries that you used in Python?

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