HEART DISEASE DIAGNOSTIC ANALYSIS - BI PROJECT

DETAILED PROJECT REPORT

BY:- SIDDHARTH JAIN

PROJECT DETAILS

Project Title	Heart Disease Diagnostic – Analysis
Technology	Business Intelligence
Domain	Healthcare
Project Difficulty level	Advanced
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel, MS-Power BI

OBJECTIVE OF THIS PROJECT

- The goal of the project is to analyze heart disease dataset which comprises various features involving that describes the disease using various analytical tools and derive insights from them and create a report.

PROBLEM STATEMENT OF THIS PROJECT

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

ARCHITECTURE

DATASET INFORMATION

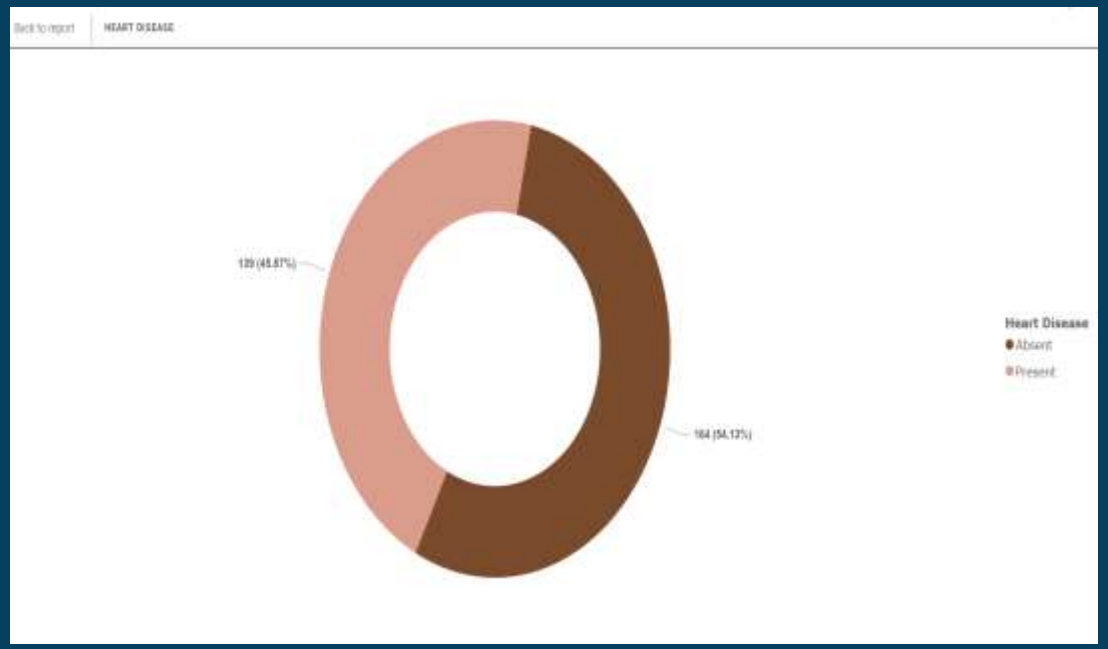
- **age**: The person's age in years
- sex: The person's sex (1 = male, 0 = female)
- cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
- trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)
- 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

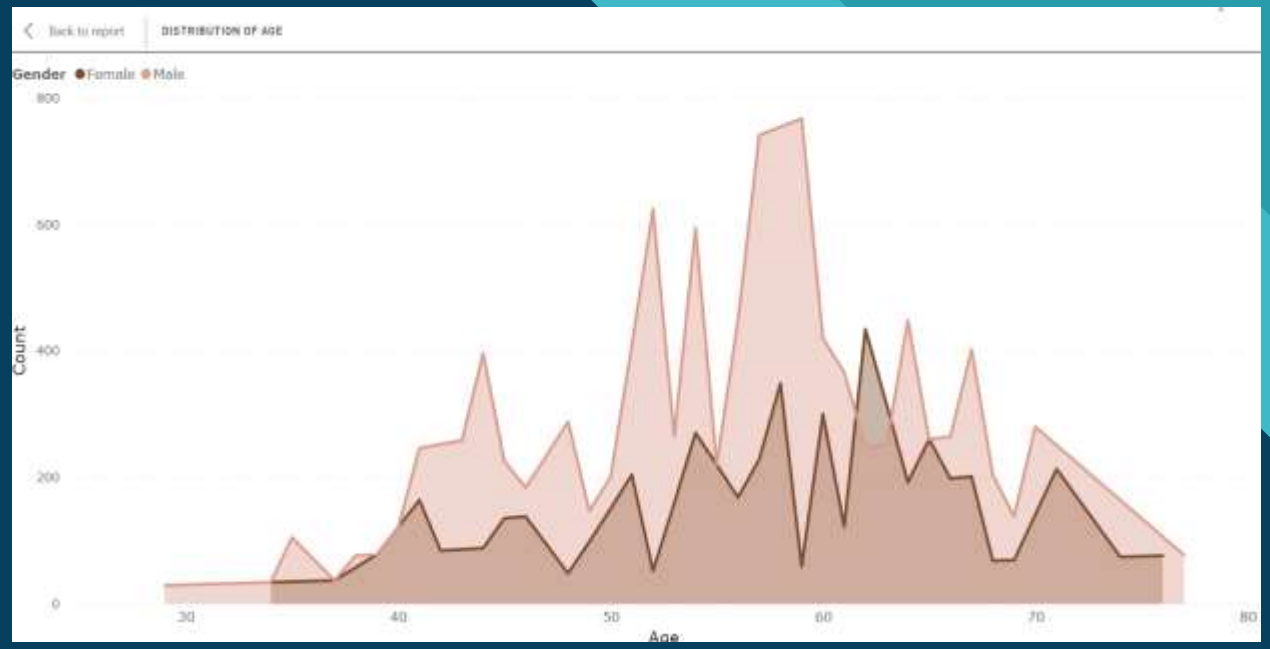
CONTINUATION

- **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 1: upsloping, Value 2: flat Value 3: down sloping)
- **ca:** The number of major vessels (0-3)
- **thal:** A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
- **num:** Heart disease (0 = no, 1 = yes)

INSIGHTS

- **What Kind of Population do we have?**

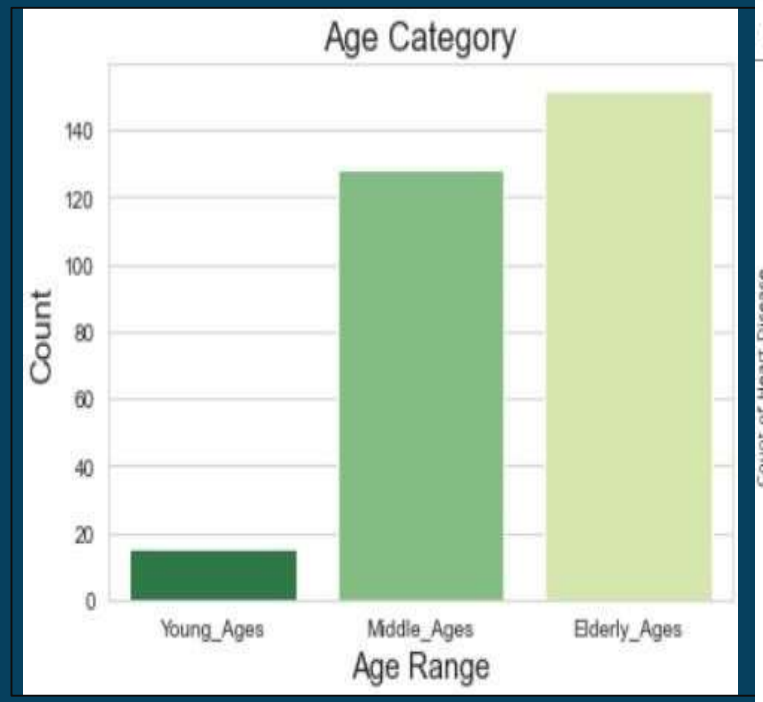


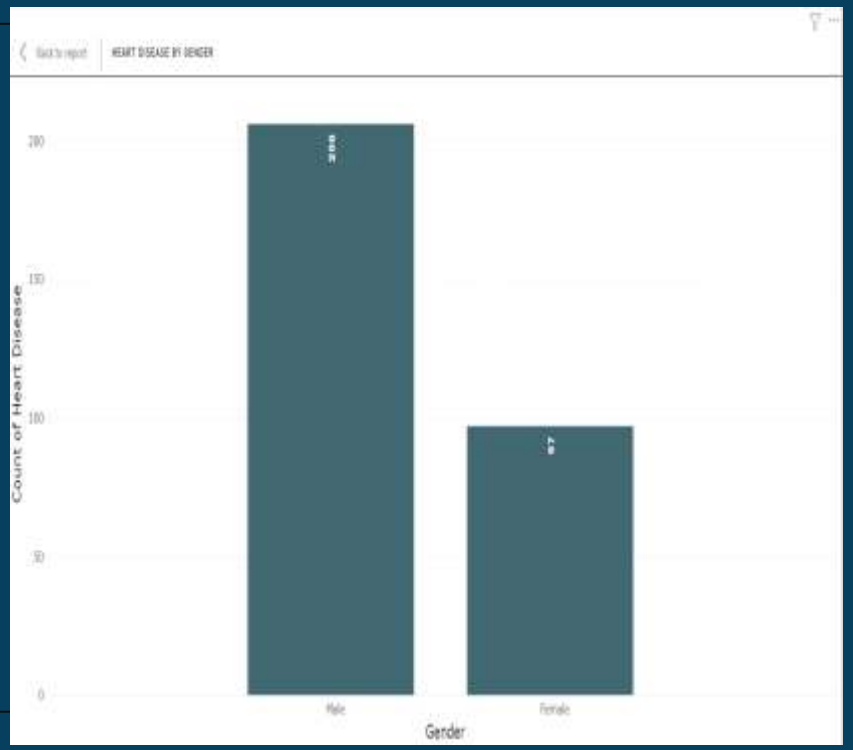


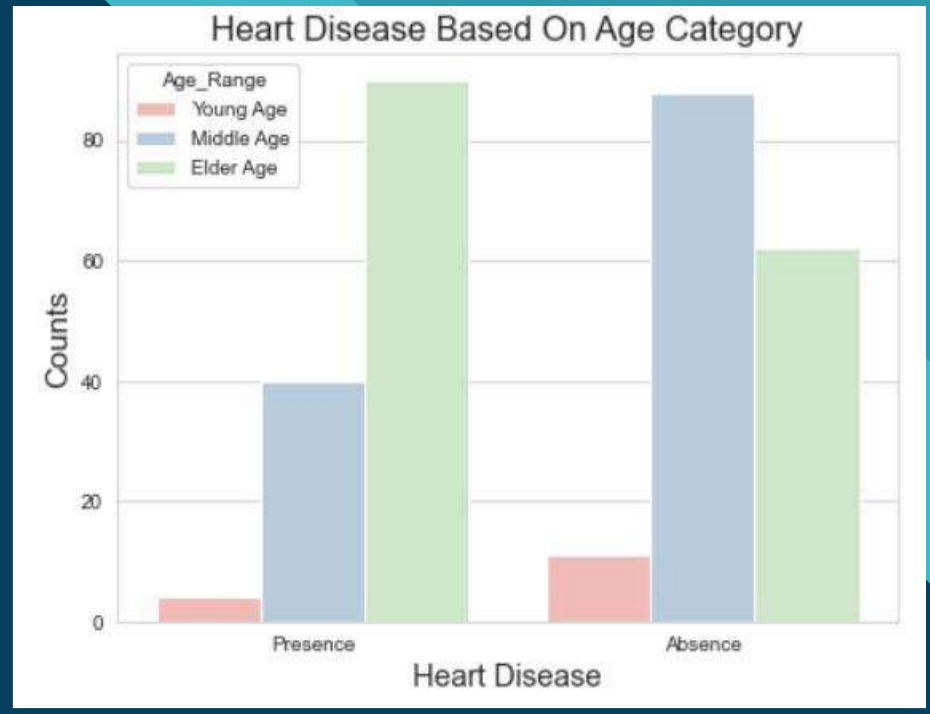
- 45.87% People suffering from heart disease.
- More men are from age category >50 and females are from category >55

INSIGHTS

- **Who Suffers from Heart Disease?**



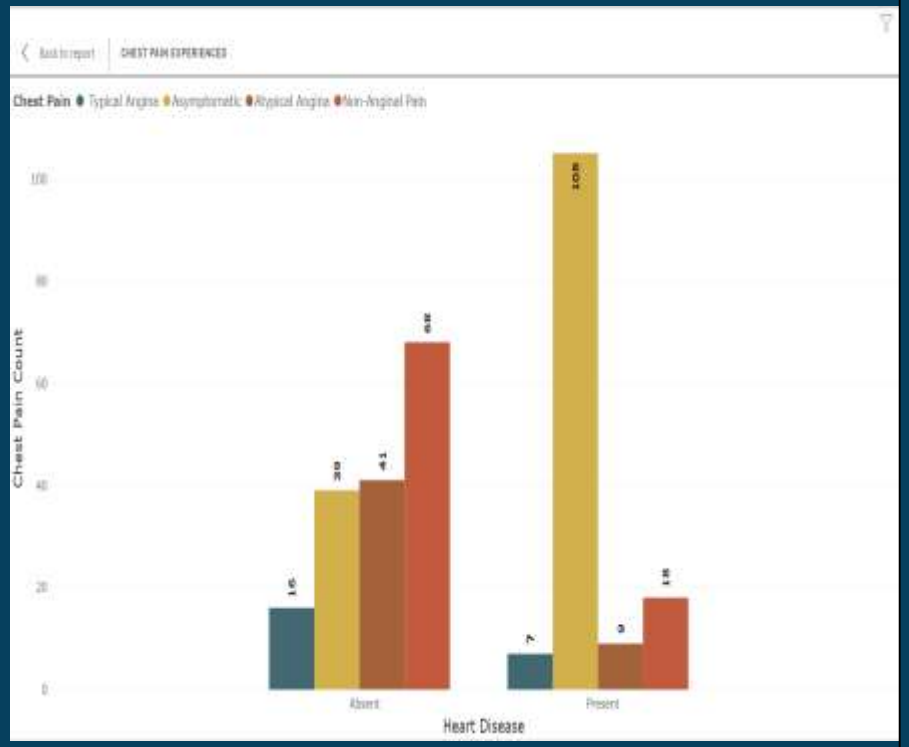


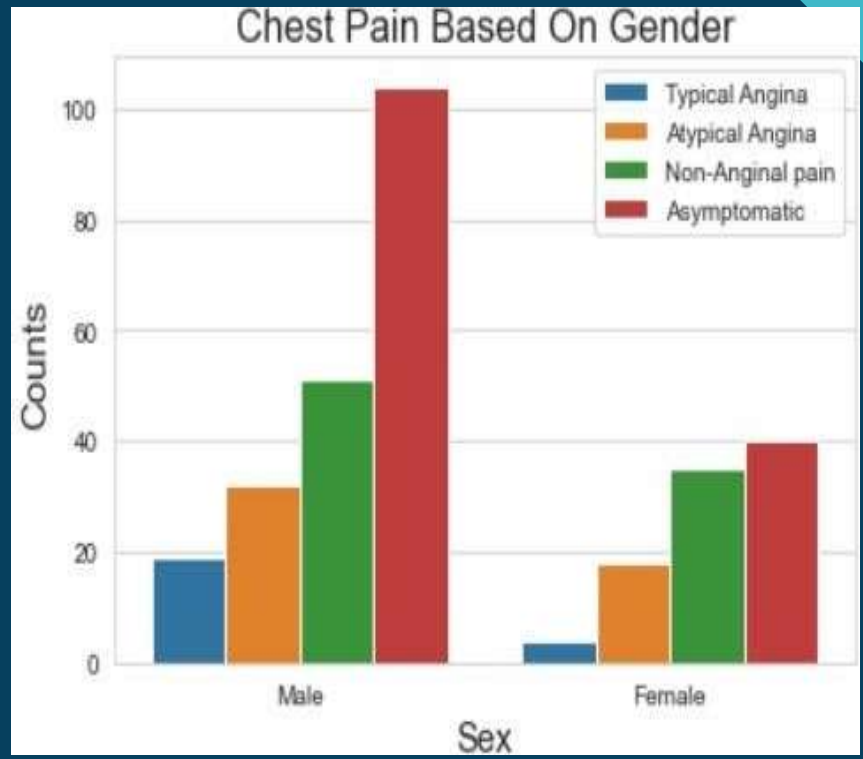


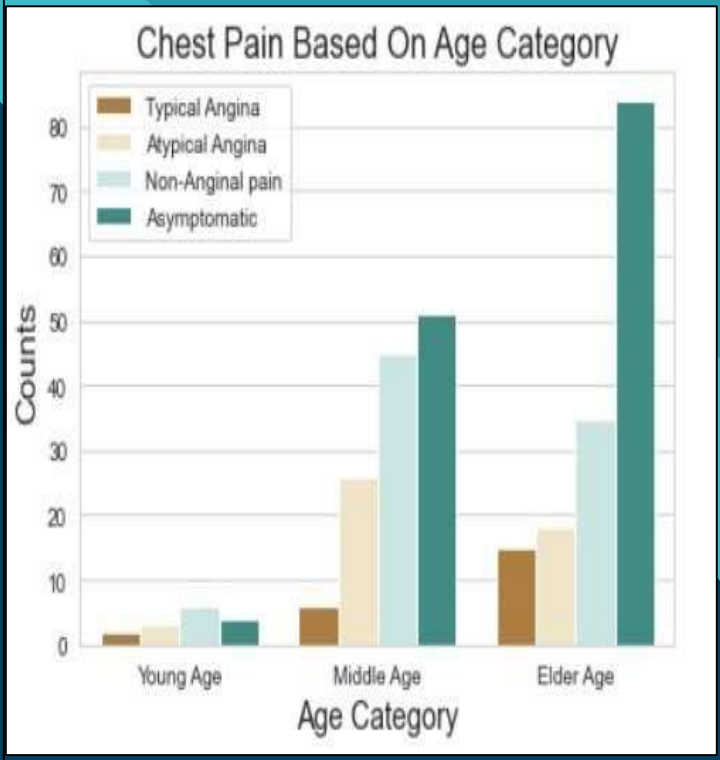
- 45.87% People suffering from heart disease.
- More men are from age category >50 and females are from category >55

INSIGHTS

- **Chest Pain Experienced By Patients**



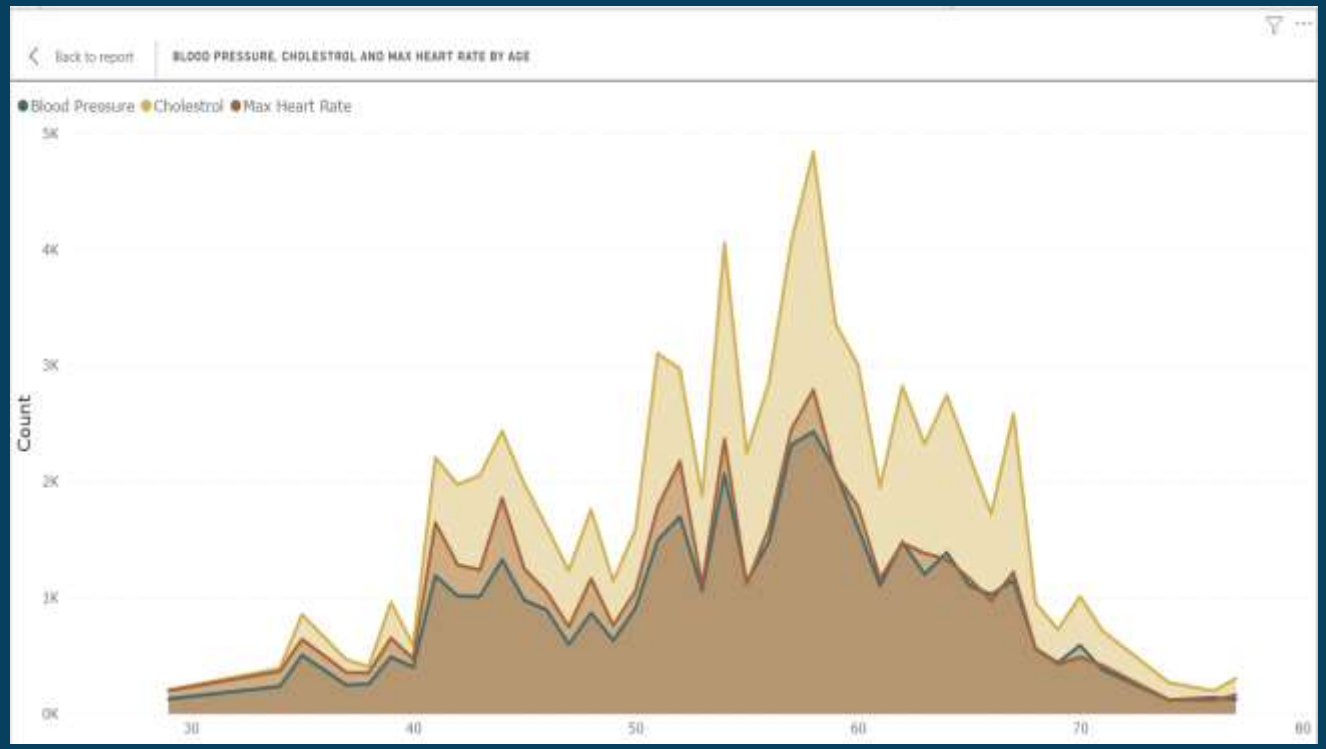


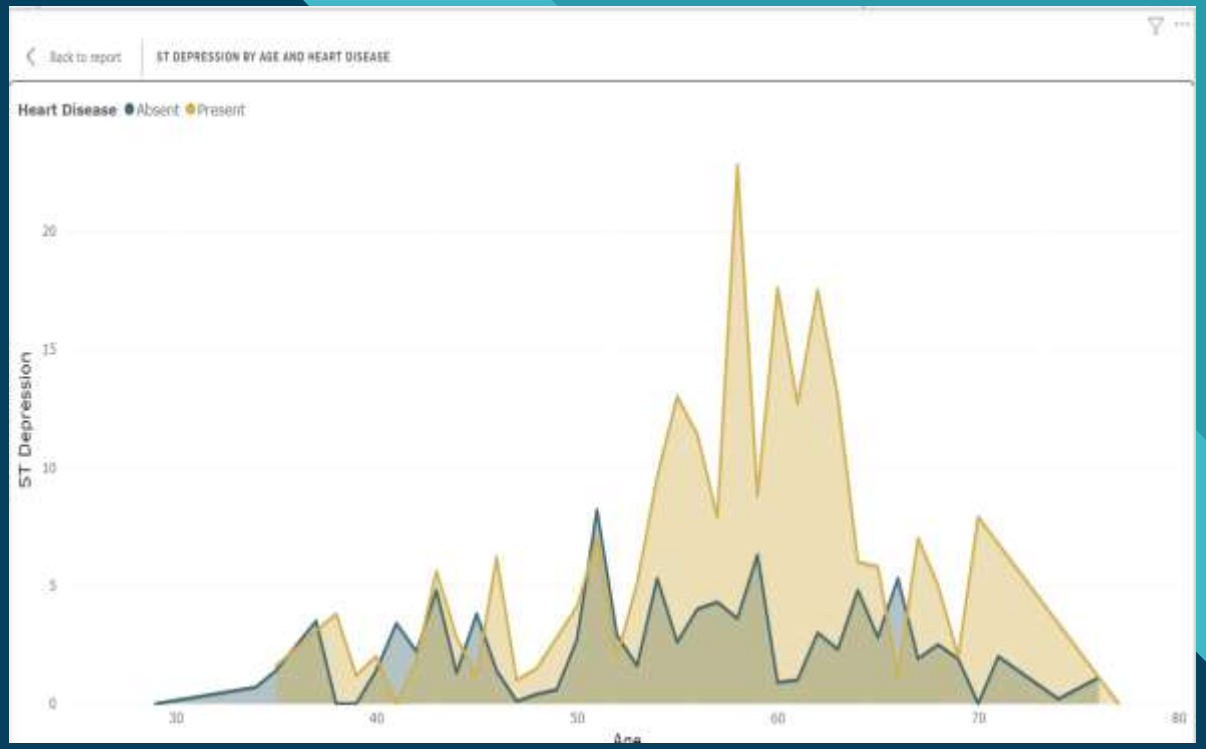


- It seems people having asymptomatic chest pain have a higher chance of heart disease.
- We can see that a higher number of men are suffering from Asymptomatic type of Chest Pain
- There is very high number of Asymptomatic Pain in Elderly age Category

INSIGHTS

Other symptoms people experience in heart disease





- Here we can observe that Blood Pressure increases between age of 50 to 60 and somehow continue the pattern till 70.
- Similarly, Cholesterol and maximum heart rate Increasing in the age group of 50-60.
- we can observe from here that ST depression mostly increases between the age group of 30-40.

KEY PERFORMANCE INDICATOR (KPI)

- Percentage of People Having Heart Disease
- Age Distribution including Gender
- Gender Distribution Based on Heart Disease
- Chest Pain Experienced by People Suffering from 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

- 45.87% People suffering from heart disease.
- Elderly Aged Men are more (50 to 60 Years) and Females are more in 55 to 65 Years Category
- Males are more prone to heart disease.
- Elderly Aged People are more prone to heart disease.
- People having asymptomatic chest pain have a higher chance of heart disease.
- High number of cholesterol level in people having heart disease.
- Blood Pressure increases between age of 50 to 60 and somehow continue till 70.
- Cholesterol and maximum heart rate Increasing in the age group of 50-60.
- ST depression mostly increases between the age group of 30-40.

Q & A

Q1) What's the source of data?

Ans) The Dataset was taken from I Neuron'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 Architecture slide 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.

Q & A Continue

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

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

THANK YOU