# Data Analytics Project

## Heart Disease Analysis Project

## Name:B.Mounika

## Abstract

Heart disease is the most major health issue that is suffering by many people all over the globe, some of the causes of heart diseases due to hypertension, diabetes, overweight, and an unhealthy lifestyle. This project of <u>Healthcare Analysis on Heart Disease Data</u> is aimed to explore the Heart Disease dataset .The objective is analyze the various features and their relationship with each other and find out their contribution towards getting a heart disease.

Various features such as **Age, Sex, Chest pain type, Blood pressure, Cholestrol,Fasting Blood sugar, Rest ECG, Thalach, Exercise enduced Angina,Major vessels,oldpeak,slope,thal** are present in the dataset.The goal of the project is to find all types of relationships between the features and come out with significant contributors to a heart disease.

# 1.introduction

- Problem statement under consideration.
- Solution strategy towards the problem.
- Architecture of the process.
- Implementation of the solution.
- Tools used.
- Analysis of the KPI's and coming out with suitable conclusions about the analysis.

The final analysis (dashboard) will be shared through Tableau public.

## 2. General Description

### 2.1 Product Perspective and Problem Statement

Healthcare domain is one of the leading domain. Using Data analysis tools and techniques to build a solution for a heathcare domain problem can be counted as a significant contribution to this domain.

The main objective of the project includes :

- Analyzing raw dataset.
- Perform data cleaning and pre processing operations on the raw data.
- Building visualizations on the cleaned features using a BI tool.
- Final conclusions.
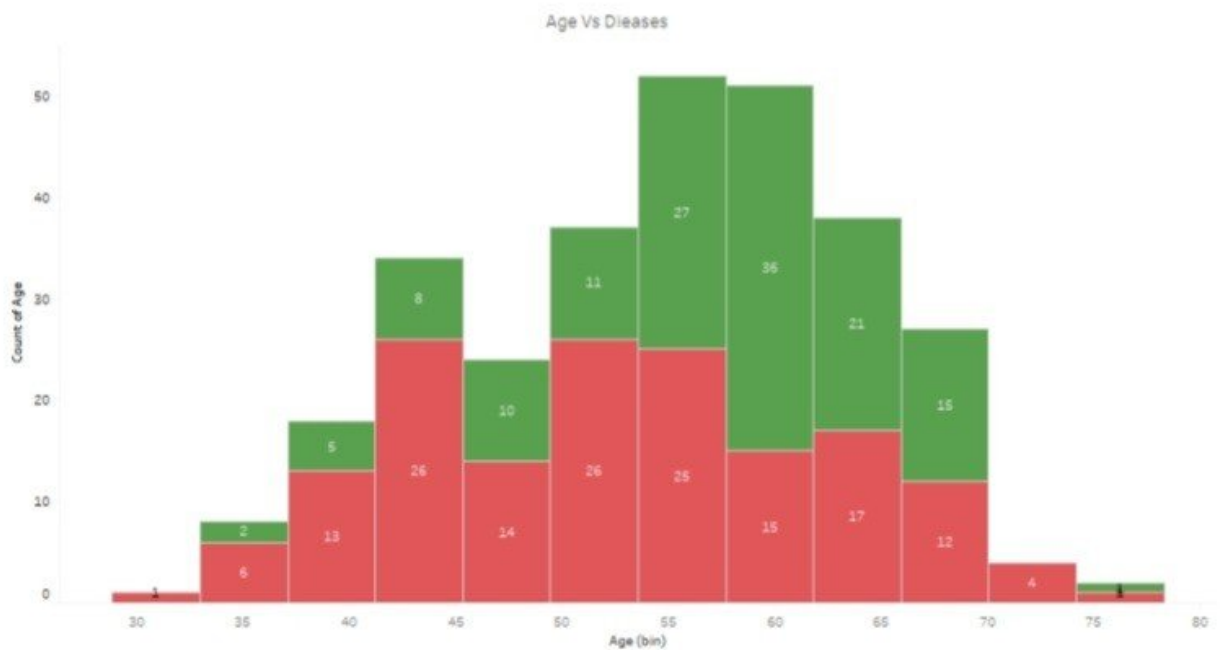- Deployment of the project on Tableau public.

## 4.1 Age Vs Diseases.



Fig. 1

- A histogram(Fig. 1) has been plotted to get an understanding of how age is putting an impact on heart dieases.
- The histogram is showing the distribution of age on the total dataset and how many people are having heart diseases.
- From the above graph, it can be concluded that the age range of **50 to 60 is having the highest chances of having heart diseases,** as the **red slope** is highest on the bin of 50 to 60. And it can also be considered because most older people can have heart diseases.
- Average age of the population is 55.

## 4.2 Blood Sugar Vs Blood Pressure.



Blood Sugar Vs Blood Preassure

Diseases

Negative           Positive

Fbs(Fasting blood sugar)
- False
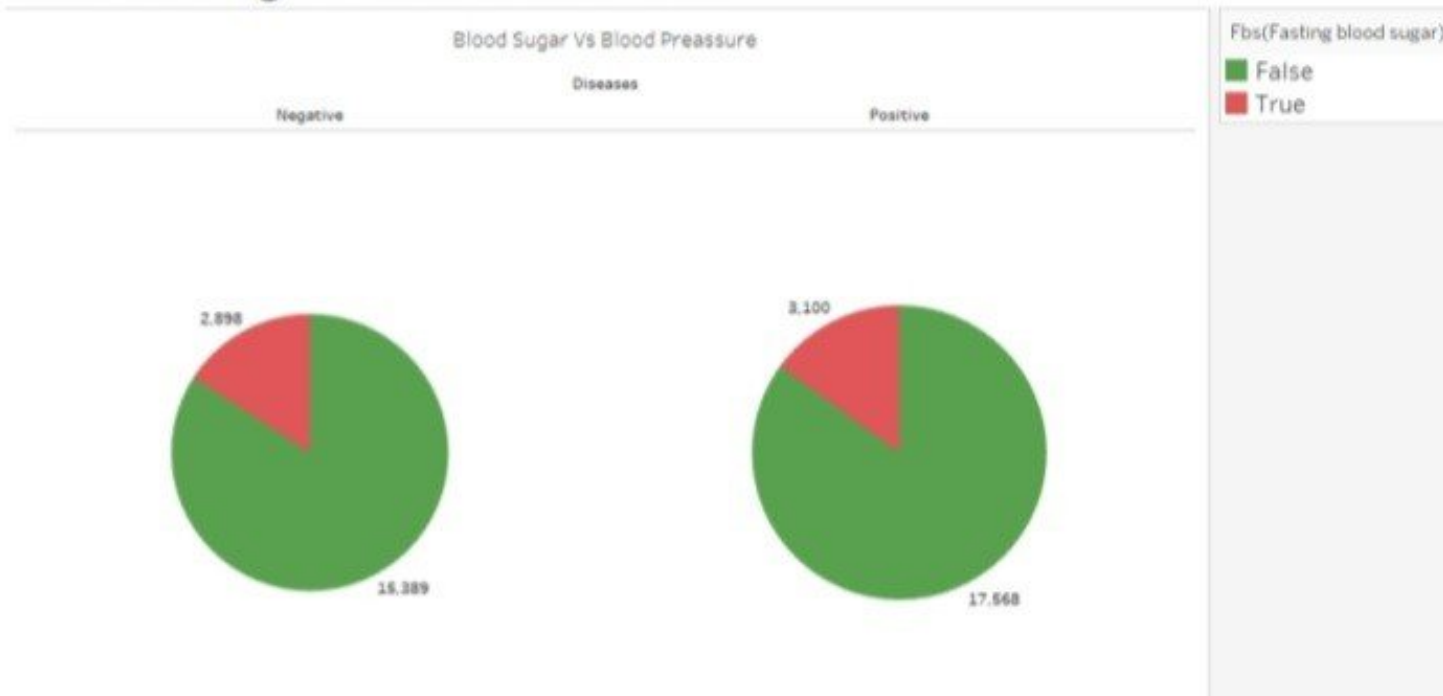- True

2.898

3.100

16.389

17.568

Fig. 2

- Impact of blood pressure and sugar in heart diseases, here we have plot two pie plots to get the understanding of the relationship between the same.
- The above figure(Fig. 2) shows the impact of blood pressure and blood sugar of individual patients in heart diseases. It is founded that people with **high blood pressure and blood sugar have more chances of having heart diseases**.
- And the population is also bigger as we can notice the right pie chart is bigger than the left one, that is an indication the population of having high blood pressure and blood sugar. And it falls under Positive Heart Diseases.

# 4.3 The count of Heart Diseases case depends on the Different Chest pain



Chest Pain Impact on Diseases
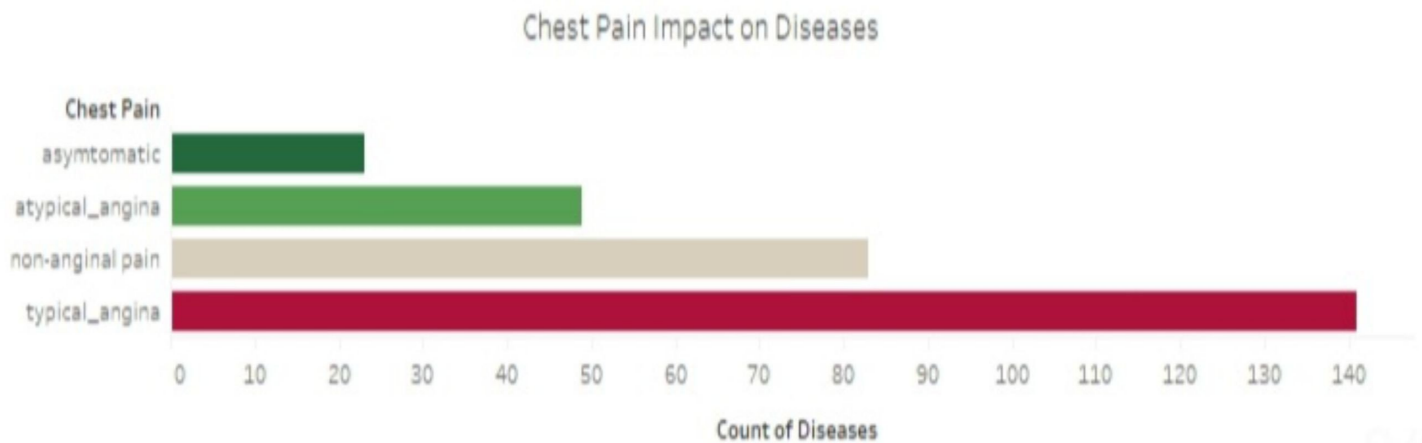
- From the above plot, we can get the count of total heart rate cases to depend on the different types of chest pain.
- And it can be concluded that chest pain type **"typical angina" is major contributor towards having heart diseases**.
- From this, we can say if a patient has "typical angina" chest pain, his/she may have a high chance of having heart disease.
- And the "asymptomatic" type of chest pain is causing the lowest chance of having heart diseases.

## 4.4 Analysis of Different types of chest pain with heart disease.



Fig. 4

- The above plot (Fig. 4) is showing the count of different types of chests that are putting an impact on heart diseases with age.
- Here the filter is used to get an understanding of different types of chest pain impact on heart diseases with heart rate, The filters are applied on dimension category is called filter, and the filter applied to measures is called a quantitative filter.
- Here the chest pain type is a categorical filter and the blood pressure, the heart comes under a quantitave filter.
- With the help of a slider, the user can change the measurement and type to predict heart disease.

## 4.5 Age Vs Cholesterol



Fig. 5

- We plot individual cholesterol vs Age to the impact of cholesterol is putting impact on age and heart disease.
- The above plot(Fig.5) is not showing any meaningful full relation between age vs cholesterol and heart diseases, the scatter does not have any correlation with any attribute.
- From this, we can conclude the **Cholesterol may not be a good attribute to determine the heart diseases.**

## 4.5 Dashboard

We have created a dashboard by merging all the plots altogether and the dashboard will provide a more granular report and we have applied a filter in every plot to produced filter results on every plot.



Fig. 6

# 5. Question and Answers (Q&A)

## 5.1 From where did you collect the dataset?

collecting the dataset from the internet which was available in google chrome and the dataset is available in a link

## 5.2 What are the most useful attribute in Heart Data Analysis?

There were a total of 14 attributes present in the data set and all are presenting the past health record of individual patients with targeting having heart diseases or not. After doing analysis we came to know that **Age, Blood sugar, Blood Pressure, and Chest Pain** are putting an impact on Heart Diseases, as we find some strong relation between them.

## 5.3 What type of analysis is you have done with the dataset?

The objective was to find the relationship between different attributes and find out the impact of past health records on heart disease. We have plotted attribute by using different approaches to get insight information. We have a plotted a histogram of Age, a scatter plot with age and cholesterol, and a pie chart with diseases and blood sugar. From these plots, we find the relation between health conditions.

## 5.4 How do you manage data encoding?

As the input data comes with all numerical values and it did not consist of any categorical value. For encoding numerical values to categorical values, we take help various reports and

existing analyses and follow dataset description to get the detailed information. And based on that we encode the numerical values.

**5.5 What are techniques were you used for data pre-processing?**

For raw data pre-processing we have used python programming language and python based library. We have to remove missing values and remove duplicate rows that were present in the dataset and also remove and restore some attributes values.

# 6. Conclusion

Heart diseases are the major causes of disability and causing uncertain death. We performed an analysis on the "Heart Disease Analysis" dataset and concluded that :

- **Chest pain is putting more impact on heart diseases of individuals along with age, blood sugar, and blood pressure.**

The moto was to perform heart diseases analysis to give an overview understanding of the relationship and how health conditions are matching with others. It can be less precise than physical laboratory analysis but it can give early-stage warnings that are observed from the record and it is very less expensive and less time-consuming than laboratory analysis.

- **From getting overview doctors can start further treatment for the patient based on his record. And provide some healthy diet plans to the patients to avoid certain diseases.**
- **Exercise routines can be incorporated on an early basis.**
- **Prior treatment can be started based on such analysis.**