Detailed Project Report (DPR)

Health Analytics on Heart Disease Data

Last Revised Date: 31/08/2021

Document Version Control

Version	Date	Author	Comments
0.1	31-Aug-2021	Subhabrata Nath Samruddhi Deshmukh Manish Tiwari	
0.2			

Reviews:

Version	Date	Reviewer	Comments

Approval Status:

Version	Review Date	Reviewed By	Approved By	Comments

Index

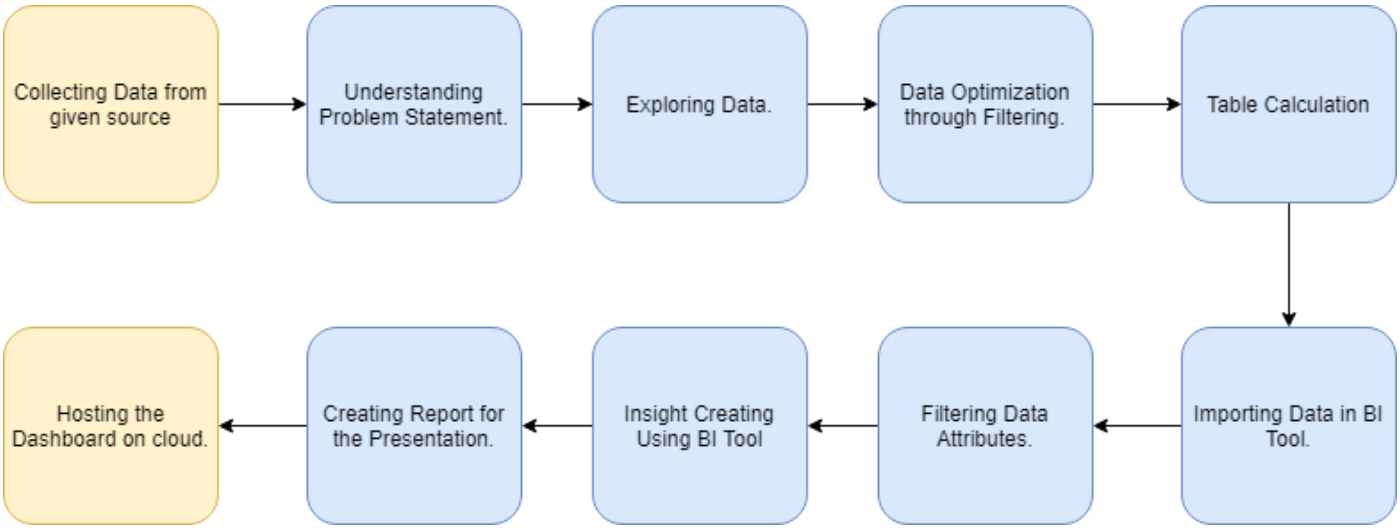
1. Objective

Perform data analysis and find the relationship between how a patient's basic health condition is dependent on having Heart Disease or not for individual patients. Heart Diseases analysis is about to find to get understanding the relationship between a patient's age, blood sugar, blood pressure, heart rate are putting impact on having heart diseases. Data analysis in healthcare assists in predicting diseases, improving diagnosis, analyzing symptoms, provide appropriate medicines, improving the quality of care. Minimizing cost, extending the life span, and reduce the death rate of heart patients.

1.1 Benefits:

- To get an understanding of the various health conditions of individual patients.
- How age is affecting other diseases.
- By analyzing different health conditions it will be possible to get a prediction of having heart disease or not.
- Help Doctors to give understanding about all health conditions in one place by providing a granular report.
- Predicting heart diseases that can save time and cost before doing a detailed medical examination of the patients.
- Generating a physical laboratory report can be time-consuming, costly, and computationally intensive. And BI reports can reduce time and expenses.

Architecture Design



1.2 Architecture Description/ Process Flow

1.2.1 Collecting Data:

Dataset is already available on the project dashboard. We have used the given dataset for building the project.

1.2.2 Understanding Problem Statement:

We have performed exploration on Heart Diseases data and have found insightful information from it. It involved analysis of various features and understanding contribution of every feature towards a possible heart disease in the future. After analysis/viz building on tableau , we have created a Dashboard for user interaction and a Detailed project report for publishing our findings.

1.2.3 Exploring Data:

Data cleaning operations performed on raw data using plotting library and python data manipulation library like pandas for these tasks.

1.2.4 Data Optimization through Filtering:

We have optimized the given dataset before starting the analysis. E.g- dropping duplicate columns and reducing the dataset dimensions as needed, encoding our data into numerical to categorical or vice versa as needed.

1.2.5 Table Calculation:

We have calculated some new fields which represents more granular results than raw data. We have found the relationship between attributes and performed the calculation as required.

1.2.6 Importing Data In BI Tool:

We have used Tableau for creating the Visualization and Dashboard. Tableau is a drag and drops application for creating Dashboard and it is very much efficient in handling a large amount of data. We can plot and show different types of relationships between attributes.

1.2.7 Filtering Data Attributes:

We have used different types of filters on our data set for plotting purposes and getting the insight information. Using filters can speed up information finding and also it can produce more variety and granular reports.

1.2.8 Insight Creation Using BI Tools:

This is the most crucial part of the complete process. Here we have found meaningful information by going deep into our dataset. We have plot different types of plots for every attribute on the Heart Diseases dataset to show the relationship between a patient's health condition depends on having heart disease or not.

1.2.9 Creating Report for the Presentation:

After we find all the possible information that can be found from the dataset we have created a detailed report for presenting and publishing our works. We have included all the visual plots and key findings from the plots.

1.2.10 Hosting the Dashboard on Cloud:

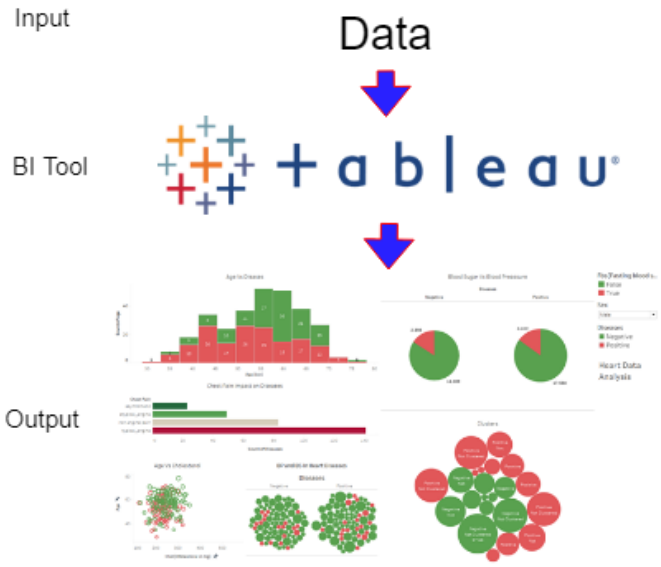
After creating a Detailed Project report and creating the dashboard we have hosted our dashboard on the cloud platform for global use. For this, we have used the Tableau Public Platform. Here we can host our dashboard and it will create a sharable link for access globally.

2. Data Manipulation

The data is required for the analysis is provided and we have done data preprocessing to remove duplicate rows, impute missing values, and encoding numerical values to categorical values for producing a more readable report.

3. BI Tools

We have used Tableau for doing the analysis. We load our data into tableau and perform various data analyses and try to find the relationship between different attributes.



4. Key Findings

We have come up with the following information after the analysis, and we have created a few plots to explain the information and we have given a constant color on all the plots for maintaining stability.

Negative.

Positive.

We have maintained the color code in all of the plots.

4.1 Age Vs Diseases.

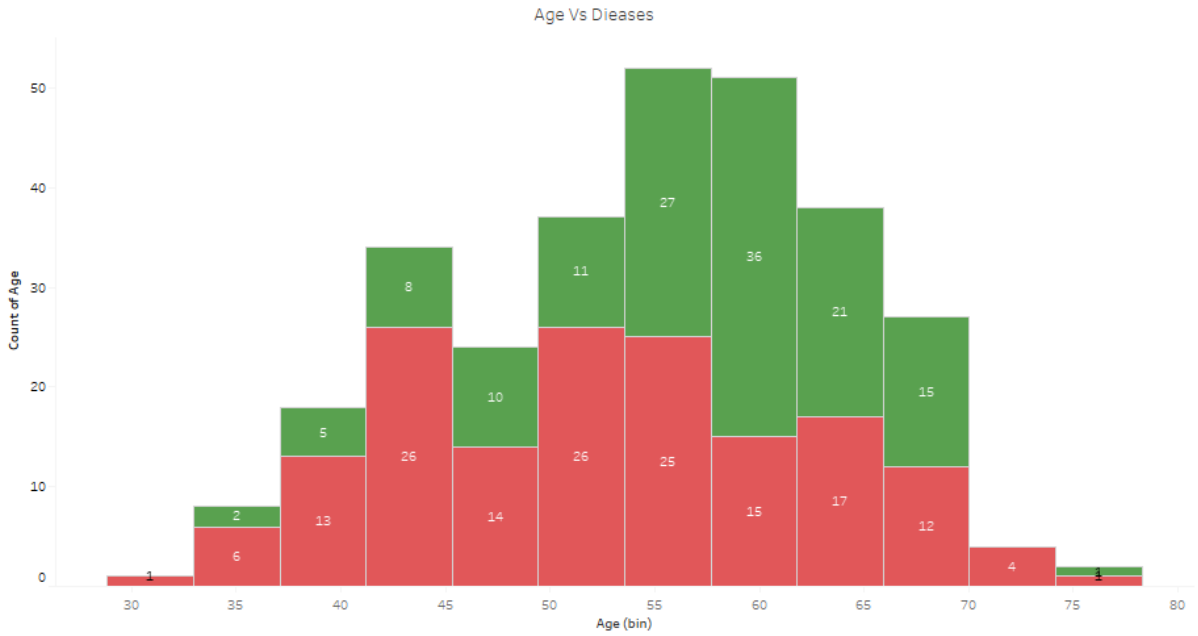


Fig. 1

- 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.

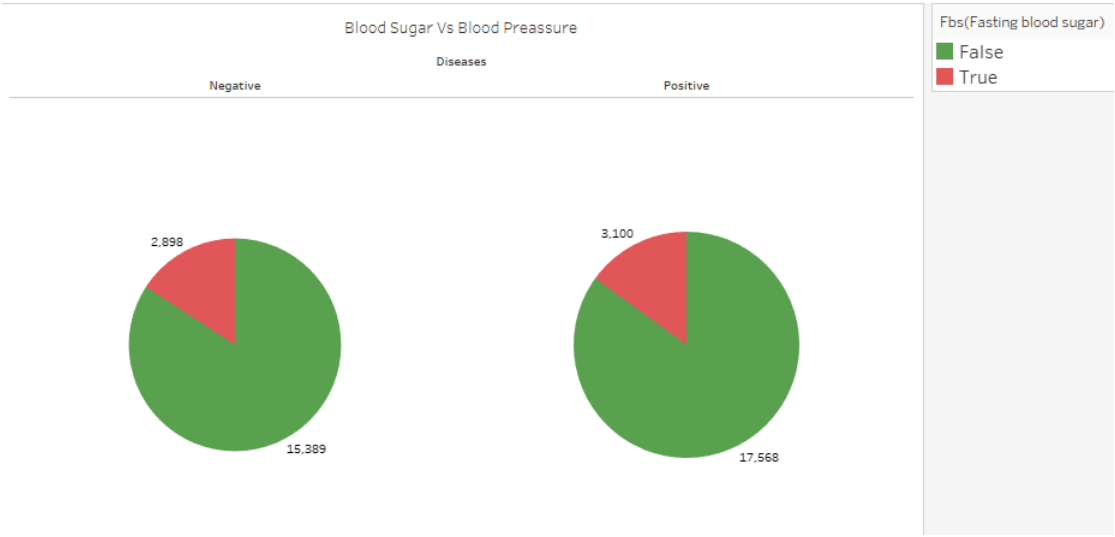


Fig. 2

- 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

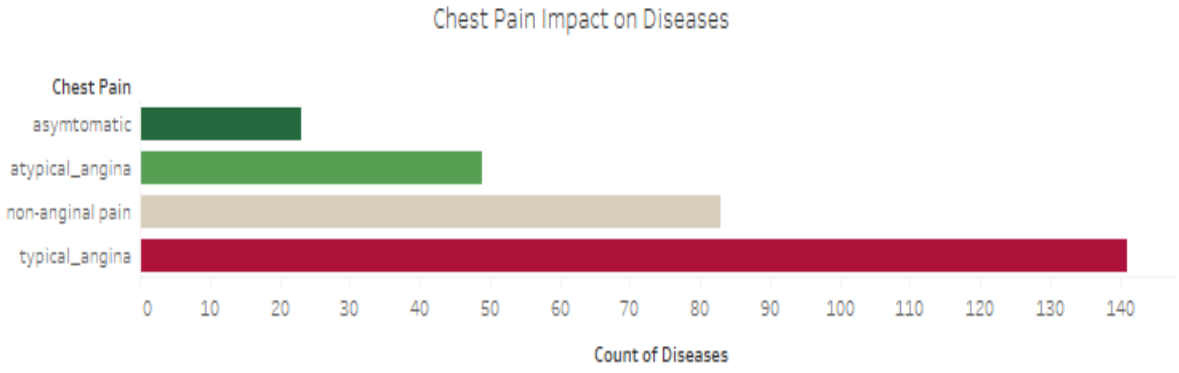


Fig. 3

- 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.

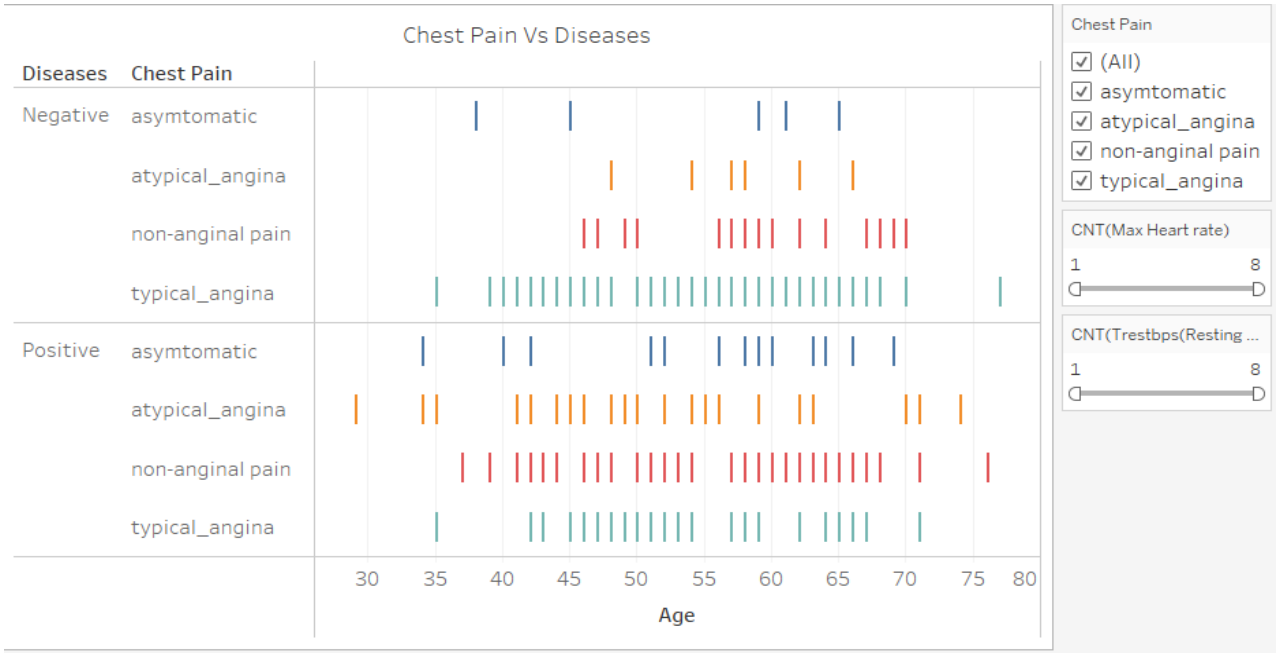


Fig. 4

- 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

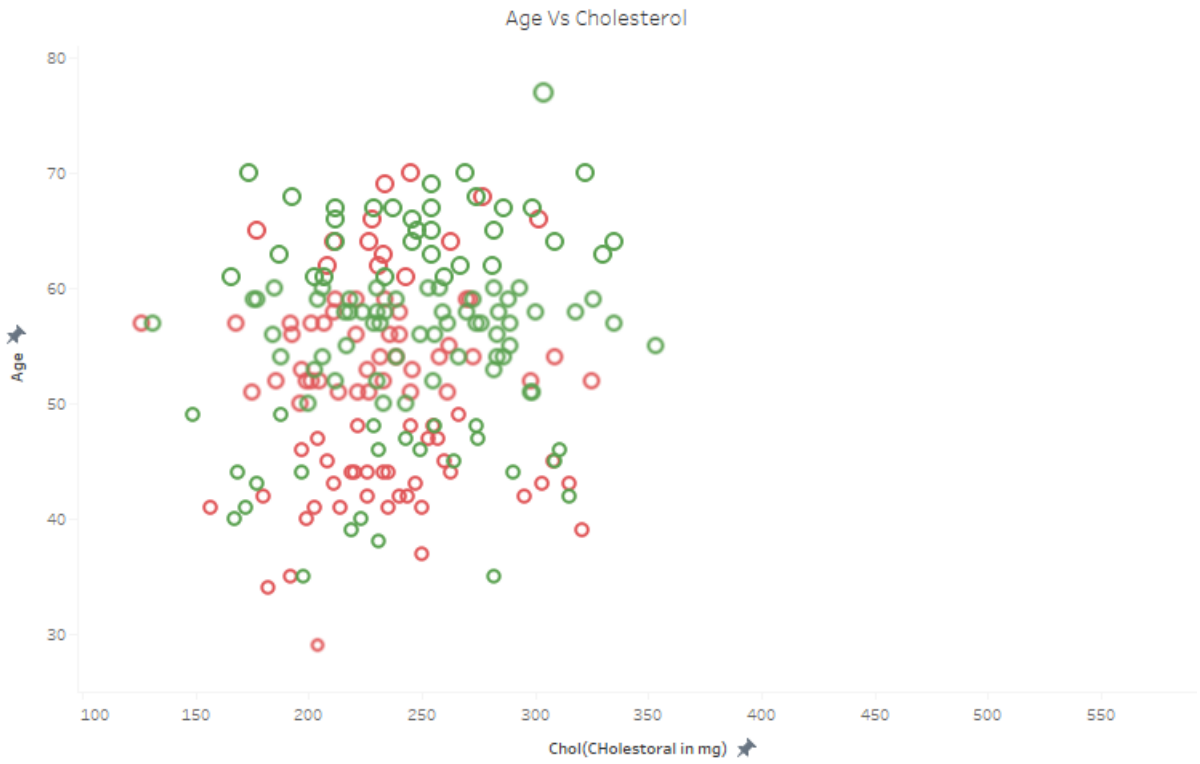


Fig. 5

- 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.

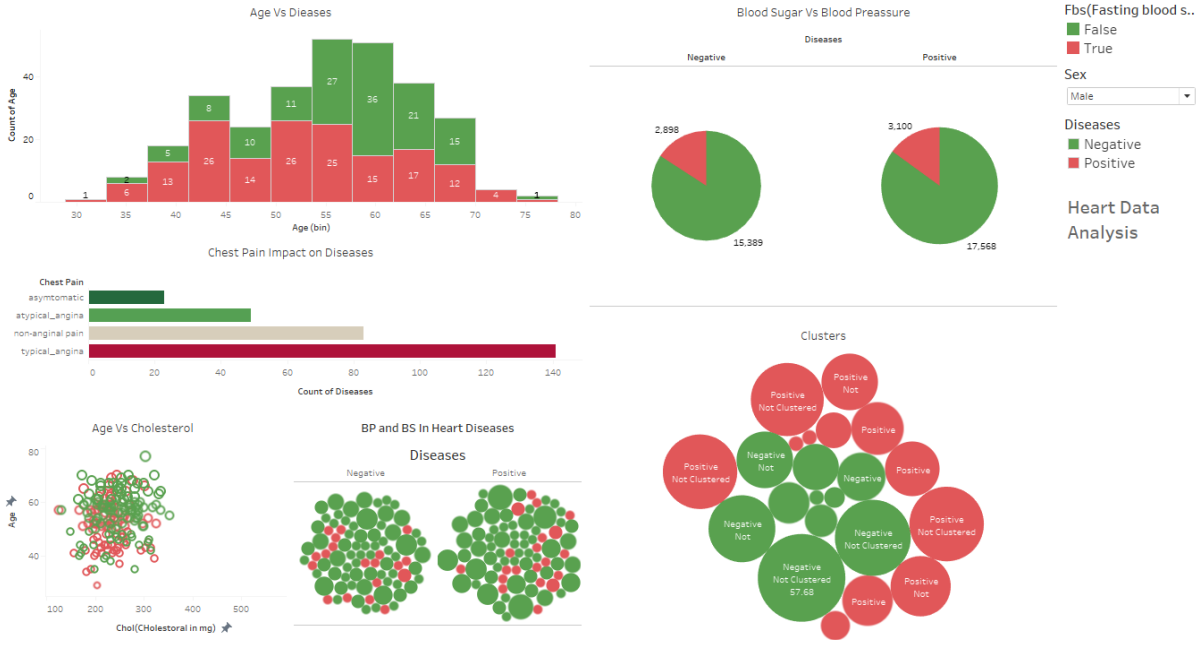


Fig. 6

5. Question and Answers (Q&A)

5.1 From where did you collect the dataset?

The data set is collected from Kaggle. The original dataset contained 76 attributes, but all published experiments refer to using a subset of 14 attributes from them. The link to the data set is - https://www.kaggle.com/ronitf/heart-disease-uci

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.

5.6 What is the final result that you are showing?

As an outcome of the project, we build a dashboard and putting individual analysis, and adding a filter with them to get more details about the analysis. And we have deployed the dashboard on Tableau public for global accessibility.

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.**