

Heat Disease

Diagnostic Analysis

**Table Of Contents**

1. **Problem Statement**
2. **Objectives and Benefits**
3. **Steps Followed**
4. **Visualizations**

Objective and Benefits

Objectives:

* The dataset contains the records for the patients and their medical parameters details and the target variable whether they will suffer from heart disease or not.
* The aim of this project is to use the given data and perform ETL and data analysis to infer key metrics and patterns in the dataset. In addition to this, different visualizations are developed to depict meaningful relationships.

Benefits:

* The data analysis will reveal some common and unique patterns in the dataset related to the medical parameters.
* Data visualizations will enhance the understanding of the effect of the high or low of these features on the chances of heart rate and give a better chance of prediction

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
* **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)

**Steps Followed**

1. **Data Extraction:** This step involves extracting the data from different sources relevant to the problem statement or obtaining data from the client.
2. Data Pre-processing: Once the raw data is obtained, we need to ensure that the data is free from errors. We perform Exploratory Data Analysis followed by Data Cleaning which involves imputing missing values, removing duplicates, finding anomalies or outliers, and treating them.
3. Data Exporting: The preprocessed data is exported to a .csv file to be used for analysis.
4. Data Loading and Modification: The preprocessed data in the .csv file is loaded into the Tableau Desktop for analysis purposes and modified for simplicity purposes.
5. Data Analysis: Once the data is loaded, we perform the data analysis using Tableau features and store the visualizations in Tableau worksheets.
6. Deployment: The prepared visualizations are deployed on the Tableau Online Software where they will be available publicly

Data Extraction

The dataset used for analysis is the Heart Disease dataset provided by the UCI Repository. It actually contains 76 attributes out of which only 14 are used. We will be using the Cleveland dataset .

Dataset Source:

[https://archive.ics.uci.edu/ml/datasets/Heart+Disease](https://archive.ics.uci.edu/ml/datasets/Heart%2BDisease)

The dataset is available is a .csv file - ‘heart\_disease\_dataset.csv’

Data Preprocessing

After Exploratory Data Analysis carried out on the dataset we have certain observations with the dataset .

* + 1. There is no column in the dataset with missing values.
    2. There are a few columns which actually contain categorical values but have been incorrectly labeled as numeric. As a part of data preprocessing we will convert them to categorical values.
    3. There are a few columns which have unusual values / outliers. We will impute these values with the median / mode value obtained from the remaining values of the columns .

Data Exporting

Once the data has been cleaned in the data preprocessing stage, we will export the cleaned dataset into a new file with .csv format.

The new dataset file has name - ‘preprocessed\_heart\_disease\_dataset.csv

Dataset Loading and Modification

* The exported .csv dataset file – ‘preprocessed\_heart\_disease\_dataset.csv’ will be imported into Power Bi Public Desktop. Since this a .csv file, we will choose the ‘Microsoft Excel’ file option when prompted to import dataset into Powerbi.
* Since the dataset contains many categorical columns which store the categories in the form of integers we will convert these numbers into meaningful phrases which will be understandable to the viewer and also easy to understand the terms used in the visualizations
* These phrases are called as ‘Aliases’ and will be provided to the values available in the categorical columns as part of data modification

Data Analysis

* Once the data has been loaded into the Tableau Desktop software, we perform the analysis for the various medical parameters provided in the dataset and study the relationship between them .
* Based on these patterns, we try to draw approximate inferences about the data provided on the basis of visualizations created .
* We have made use of different aspects of Tableau like different charts, labeling, aliases, filtering, and actions based on user choice. We have created separate worksheets for each type of visualization which contains the chart and a caption as well which contains the summary of analyses drawn

Deployment

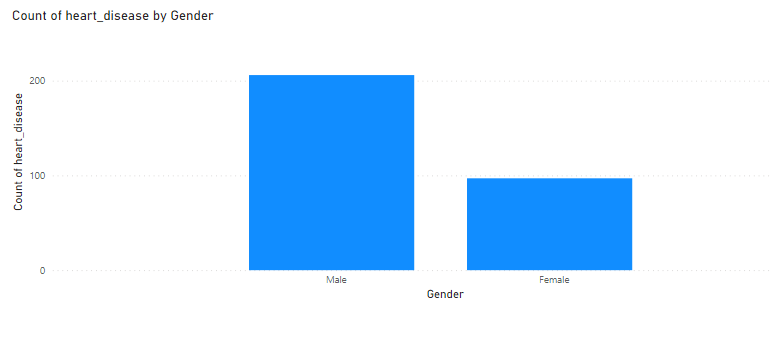
* All the different worksheets that have been created are compiled together into a Tableau workbook. Each worksheet is named based on the type of visualization performed in the chart .
* When we save all the worksheets on Tableau software, it connects to the Tableau Public Software via personal email id and credentials. The dashboard is uploaded onto the Tableau Public Software on personal profile and this is visible to public.
* The link for the worksheets is at:

<https://app.powerbi.com/view?r=eyJrIjoiNTE0NGE2NWMtNzQ3ZC00ZDBhLTg5NjYtODA4ZGQ1NmQwZjRiIiwidCI6IjI3ZjlmMWZlLWY3NzctNGMzNi1hYTQyLWM2YTY3MjBiODczMSJ9&pageName=ReportSection>

Visualizations

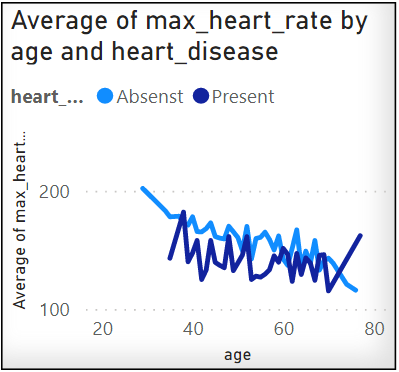
We have performed Exploratory Data Analysis using Jupyter notebook and Power Bi Desktop . The following important results came out after detailed analysis .

1. **Number of female heart patient vs male heart patient .**



From this findings we came to know that the probability of heart diseases are seen to be high in men than woman .

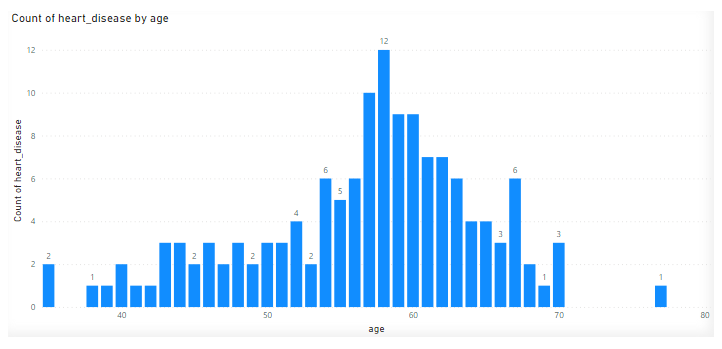
1. **Average of max\_heart\_rate by age and heart disease.**



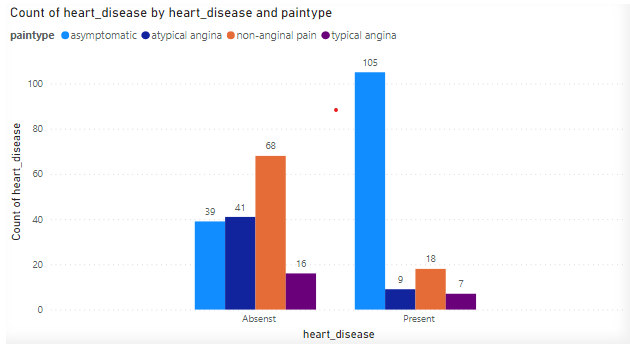
This line graph shows people having lower max\_heart\_rate have a greater risk of heart diseases.

1. **Count of heart disease by age**

This graph shows heart disease happens to most of the people having flat at the age of 50-60 .

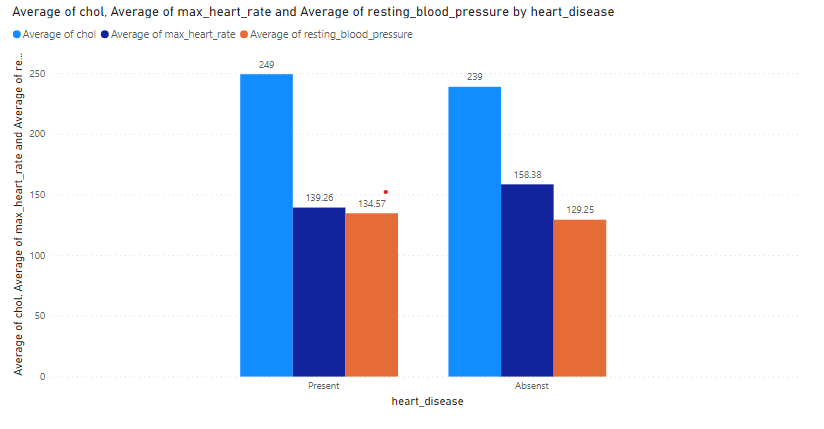


1. **Count of heart disease by pain type**



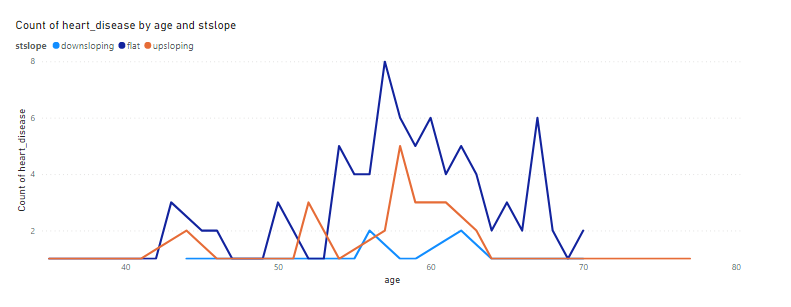
This bar plot shows that 80% of the heart disease are asymptomatic in nature . So we need to be a precaution and without any symptoms also need to do heart check up at least once in a month .

1. **Average cholesterol, max\_heart\_rate and blood pressure by heart disease.**



Again this bar plot shows people having lower max\_heart\_rate have greater risk of heart disease and having more than 130 mmHg blood pressure at age 42 have also high risk of heart disease .

1. **Count of heart disease by age and ST slope**



The line graph slows people having ST slope high are having a greater risk of heart disease .

Key Performance Indicator (KPI)

KPIs (Key Performance Indicators)

Key indicators displaying a summary of the heart disease and its relationship with different metrics

1. Variation of Count of people having hear disease .
2. Variation of ‘chol’ over age .
3. Variation of resting\_blood\_pressure by age .
4. Variation of max\_heart\_rate by age and heart\_disease.
5. Variation of Count of heart\_disease by age.
6. Variation of Average st\_depression by age.
7. Variation of count of heart disease by pain type .

Conclusion

* 45.87% of People suffer 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 cholesterol levels in people having heart disease.
* Blood Pressure increases between the age of 50 to 60 and somehow continues till 70.
* Cholesterol and maximum heart rate Increased in the age group of 50-60.

ST depression mostly increases between the age group of 30-40

Thank You

Thank You