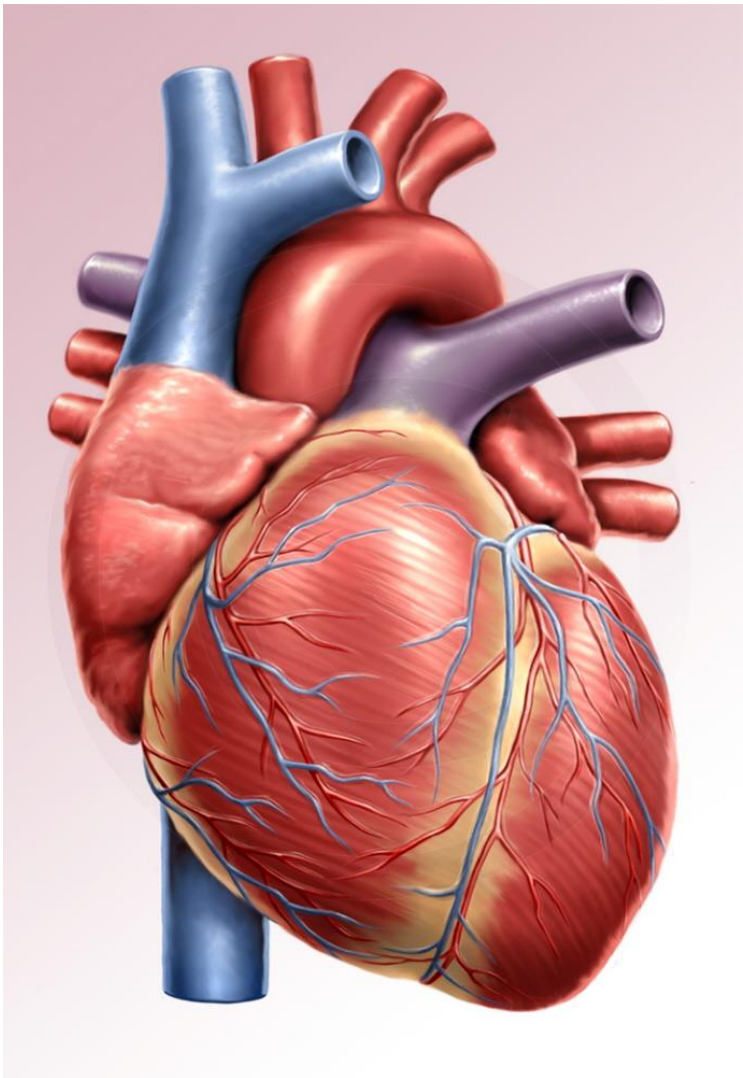
HEART DISEASE ANALYSIS



VERSION: 1.0

DATE 14-12-2022

RAVIKUMAR SURAM

DOCUMENT VERSION CONTROL

HEART DISEASE ANALYSIS
BUSINESS INTELLIGENCE PROJECT

DATE	VERSION	AUTHOR	CHANGE
14-12-2022	1.0	RAVIKUMAR SURAM	FIRST VERSION OF LLD

Contents:

1. Introduction

1.1 What is Low Level Design Document?

- The goal of the Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Heart Disease Analysis dashboard. LLD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 What is Scope?

- Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

1.3 Project Introduction

- Heart disease is the leading cause of death. The term “heart disease” refers to several types of heart conditions. The most common type of heart disease is coronary artery disease (CAD), which can lead to heart attack.
- Sometimes heart disease may be “silent” and not diagnosed until a person experiences signs or symptoms of a heart attack, heart failure, or an arrhythmia.
- Thus, Preventing Heart Disease is necessary and analysing the risk factors to predict the heart disease is need of an hour.

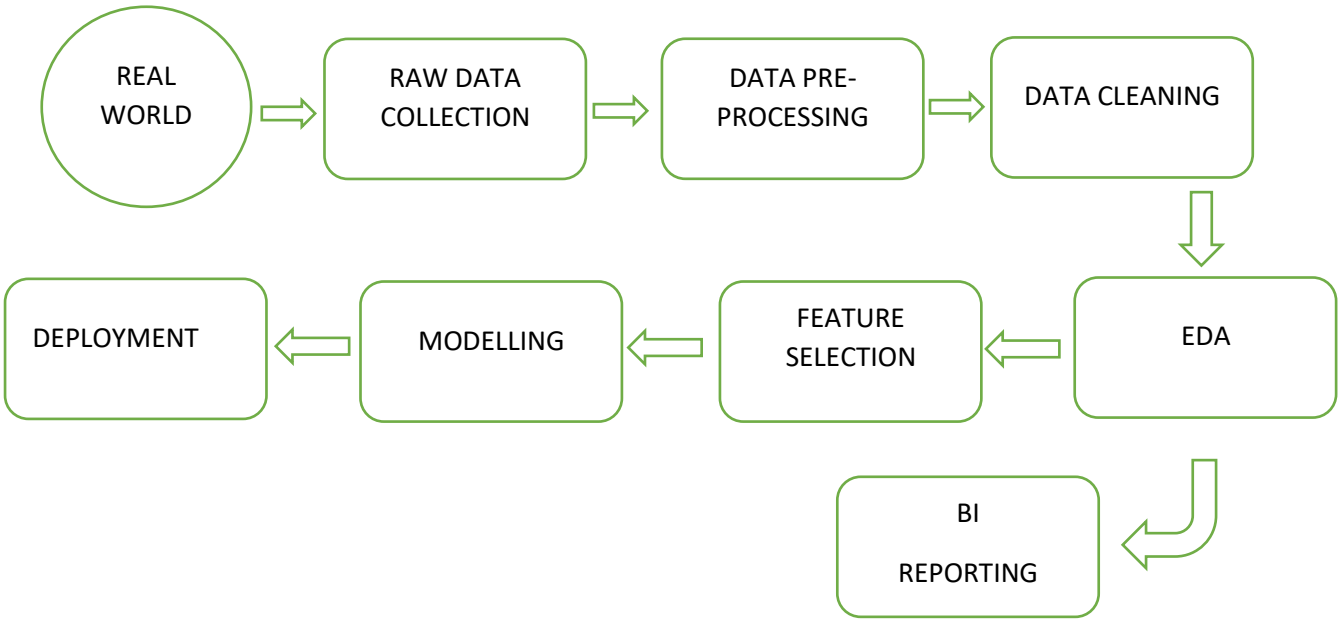
2. Problem Statement

- The goal of this project is to analyse what are the factors that are responsible for heart disease and probability of occurrence of the heart disease based on a combination of features that describes the condition of the patients.
- To do EDA and to extract the meaningful insights from the data based on the given information.

3. 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: downsloping)
- 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)

4. Architecture



4.1 Architecture Description

1. **Raw Data Collection**
   The Dataset is available at
   https://drive.google.com/file/d/1CI7Jb-d3rhccfQhvDUIGxHCaAGaP9-Uk/view?usp=sharing

2. **Data Pre-Processing**
   Before building any model, it is necessary to clean the data to feed the correct data to the model to bring meaningful insights and build a model and predict. Model performance depends on the quality of data is feed to the model to train. This Process includes

   a) Handling Null/Missing Values

   b) Handling Skewed Data

   c) Outliers Detection and fixing

3. **Data Cleaning**
   Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
   a) Remove duplicate or irrelevant observations
   b) Fixing the outliers
   c) Renaming required attributes

4. **Exploratory Data Analysis (EDA)**
   Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

5. **Reporting**
   Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy and self-explanatory report because your model will be used by many stakeholders who are not from technical background.
   a) Low Level Design Document (HLD)
   b) Low Level Design Document (LLD)
   c) Architecture
   d) Wireframe
   e) Detailed Project Report
   f) Power Point Presentation

6. **Modelling**
   Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

7. Deployment

Power BI Dashboard.