# Low Level Design (LLD)

# Healthcare Analytics

# Heart Disease Prediction

Revision Number: 1.3
Last date of revision: 22/12/2021

Ankit Kashyap

# Document Version Control

| Date Issued | Version | Description | Author |
|---|---|---|---|
| **10th Dec 2021** | 1.0 | First Draft | Ankit Kashyap |
| **15th Dec 2021** | 1.1 | Added Proposed Solution and Dataset | Ankit Kashyap |
| **21st Dec 2021** | 1.2 | Added the results | Ankit Kashyap |
| **22nd Dec 2021** | 1.3 | Final Modifications | Ankit Kashyap |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Contents

# Abstract

Heart disease/ **cardiovascular disease**(**CVD**) is a class of diseases that involve the heart or blood vessels. It is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases have increased significantly over the past few decades in India it has become the leading cause of death in India. A study shows that from 1990 to 2021 the death rate due to heart diseases have increased around 39% from 155.7 to 217 deaths per 1 lakh population in India. The underlying mechanisms vary depending on the disease. This may be caused by high blood pressure, smoking, diabetes mellitus, lack of exercise, obesity, high blood cholesterol, poor diet, excessive alcohol consumption, and poor sleep, among others.

Thus, preventing heart diseases has become more than necessary. It is estimated that up to 90% of CVD may be preventable. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives. Detection of a CVD at an early stage leads to prevention of more than 80% of potential related deaths.

With the anecdotes of fairly and conventionally healthy people getting heart related illnesses, it is no langer an ‘old-age’ disease as previously thought and need dedicated research for prevention and potential cure.

# 1 Introduction

## 1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the heart disease prediction analysis technique. It will explain the necessary steps which have to be followed before any analysis can begin. The document would also describe the algorithms and techniques used to predict the presence and absence of the heart disease and present a comparative result for the same. 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. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

- Problem Understanding.
- Data Acquisition.
- Data Pre-Processing and Exploratory Analysis
- Development of models
- Auditing accuracy and retrain if require
- Finalizing the model
- Dashboard report for important activities

## 1.2 Scope

The LLD documentation presents the detailed structure of the heart disease prediction system for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.

## 1.3 Project Introduction

Heart disease (HD) is one of the most common diseases nowadays, and an early diagnosis of such a disease is a crucial task for many health care providers to prevent their patients for such a disease and to save lives.

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some data science techniques need to be used. The data analysis predicts the likelihood of patients getting heart disease. It enables significant knowledge. E.g. Relationships between medical factors related to heart disease and patterns, to be established. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases.

## 1.4 Constraints

Our analysis is done based on a limited dataset provided for a specific (14) features affecting heart disease. The analysis does not take into account any external interventions like underlying disease, type of medication used, lifestyle patterns, BMI value e.t.c.

## 1.5 Risks

Document specific risks that have been identified or that should be considered.

## 1.6 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

# 2 Technical specifications

## 2.1 Dataset

The dataset used in this article is the Cleveland Heart Disease dataset taken from the UCI repository.

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

Figure 1: Heart Disease Prediction dataset

The dataset consists of 303 individuals data. There are 14 columns in the dataset, which are described below.

1. ***Age***: displays the age of the individual.
2. ***Sex***: displays the gender of the individual using the following format :
   1 = male
   0 = female
3. ***Chest-pain type***: displays the type of chest-pain experienced by the individual using the following format :
   0 = typical angina
   1 = atypical angina
   2 = non — anginal pain
   3 = asymptotic
4. ***Resting Blood Pressure***: displays the resting blood pressure value of an individual in mmHg (unit)
5. ***Serum Cholestrol***: displays the serum cholesterol in mg/dl (unit)
6. ***Fasting Blood Sugar***: compares the fasting blood sugar value of an individual with 120mg/dl.

If fasting blood sugar > 120mg/dl then : 1 (true)
else : 0 (false)

7. ***Resting ECG*** : displays resting electrocardiographic results
   0 = normal
   1 = having ST-T wave abnormality
   2 = left ventricular hyperthrophy
8. ***Max heart rate achieved*** : displays the max heart rate achieved by an individual.
9. ***Exercise induced angina*** :
   1 = yes
   0 = no
10. ***ST depression induced by exercise relative to rest***: displays the value which is an integer or float.
11. ***Peak exercise ST segment*** :
    0 = upsloping
    1 = flat
    2 = downsloping
12. ***Number of major vessels (0–4) coloured by flourosopy*** : displays the value as integer or float.
13. ***Thal*** : displays the thalassemia :
    0,1 = normal
    2 = fixed defect
    3 = reversible defect
14. ***Diagnosis of heart disease*** : Displays whether the individual is suffering from heart disease or not :
    0 = absence
    1 = present

# 3 Technology stack

| | |
|---|---|
| **Backend Programming** | Python |
| **Dataset** | Comma Separated Value file |
| **BI Reporting** | Microsoft Power BI |

# 4 Proposed Solution

## 4.1 Architecture

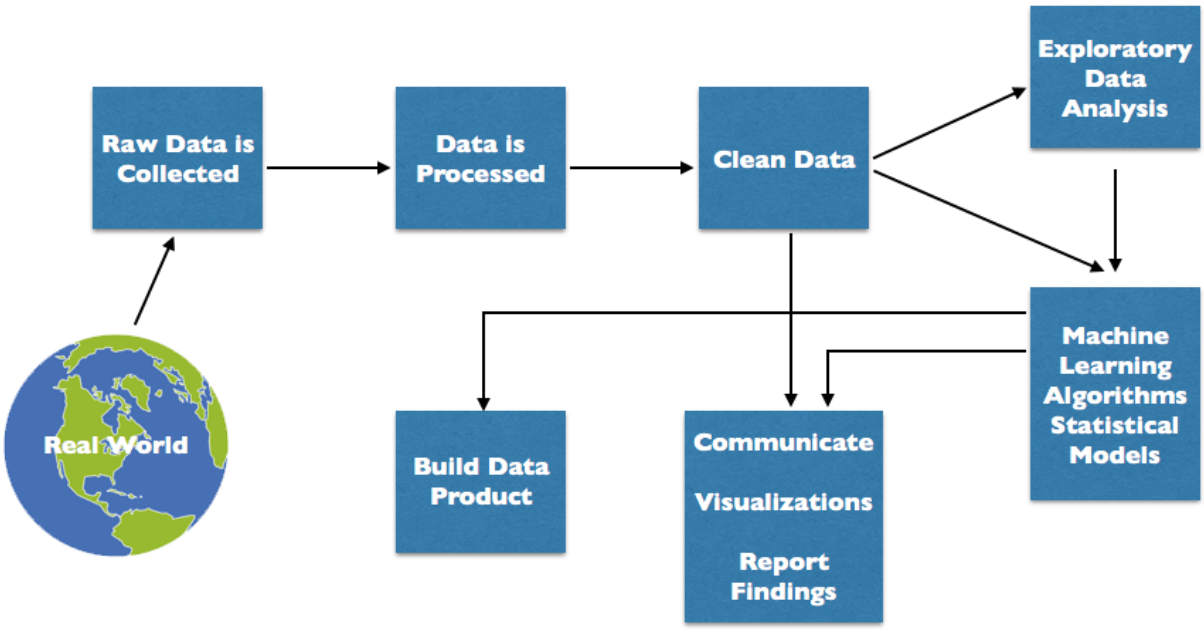


Figure 2: Project Architecture

### 4.1.1 Architecture Description

1. **Raw Data Collection:** The Dataset was taken from iNeuron's Provided Project Description Document.
   https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA22LW0Lt3Y-Q8

2. **Data Pre-Processing**
   Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data feeded to the model to train.
   This Process includes-
   a) Handling Null/Missing Values
   b) Handling Skewed Data
   c) Outliers Detection and Removal

   **Data Cleaning**

   Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

   d) Remove duplicate or irrelevant observations

   e) Filter unwanted outliers

   f) Renaming required attributes

3. **Exploratory Data Analysis**

   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.

4. **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) High Level Design Document (HLD)
   b) Low Level Design Document (LLD)
   c) Architecture
   d) Wireframe
   e) Detailed Project Report
   f) Power Point Presentation

5. **Modelling**

   Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the data 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.

6. **Deployment**
   We created a Power BI Dashboard and published it on Power BI Service

**Size of the Dataset:** 303 patients data with 14 features describing the presence/absence of disease

**Result of the Model**

| Model name | Precision | Test Accuracy |
|---|---|---|
| **Logistic Regression** | 91.5% | 83% |
| **Decision Tree** | 73.5% | 75.4% |
| **Random Forest** | 85% | 81.3% |
| **Support Vector Classification** | 85.2% | 70% |
| **Naïve Bayes** | 92% | 85% |

# 5 Dashboards

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators.


Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the presence of heart disease.

The final dashboard based on the model result and prediction is shown below:

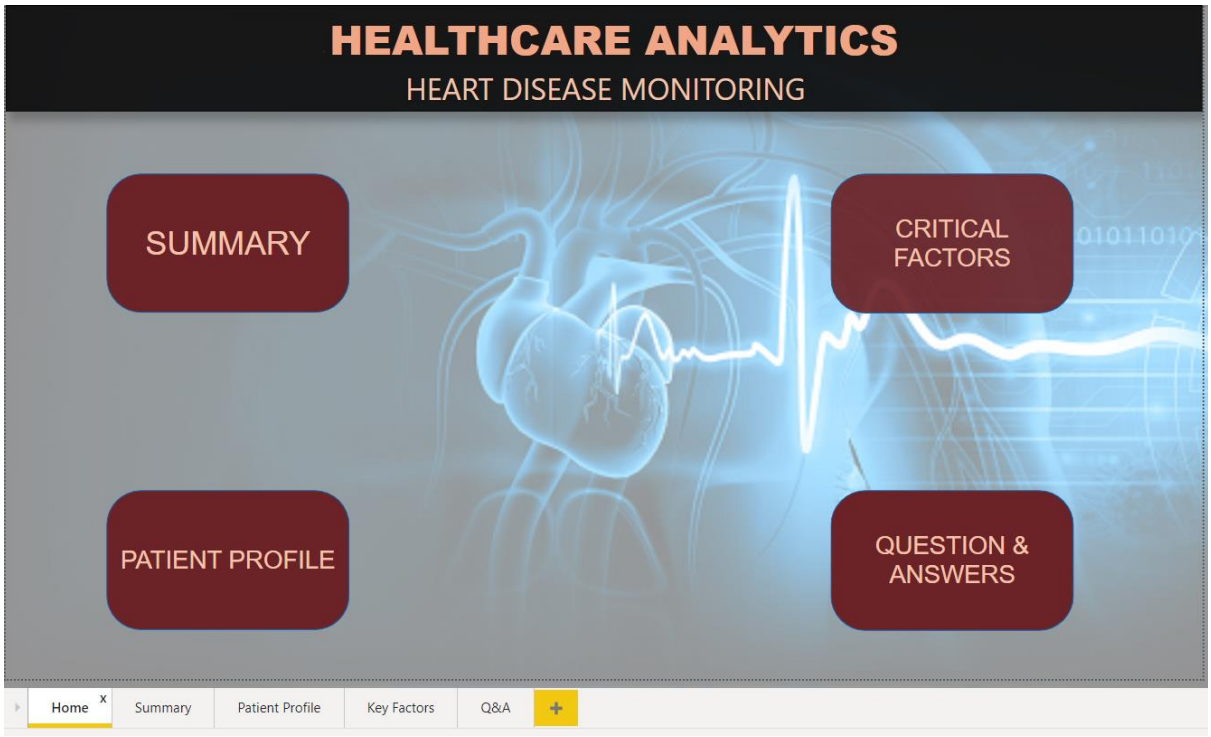


Figure 3. Final Dashboard

The Dashboard is published on Power BI service and an auto-refresh mode has been set so that the dashboard keeps on updating as the real time data loads into the log file.

The Dashboard showcases the multiple insights that has been drawn from the log files as f

1. **Summary:** The dashboard contains the statistical summary of the dataset and the visual displaying the effect of different feature to heart diseas

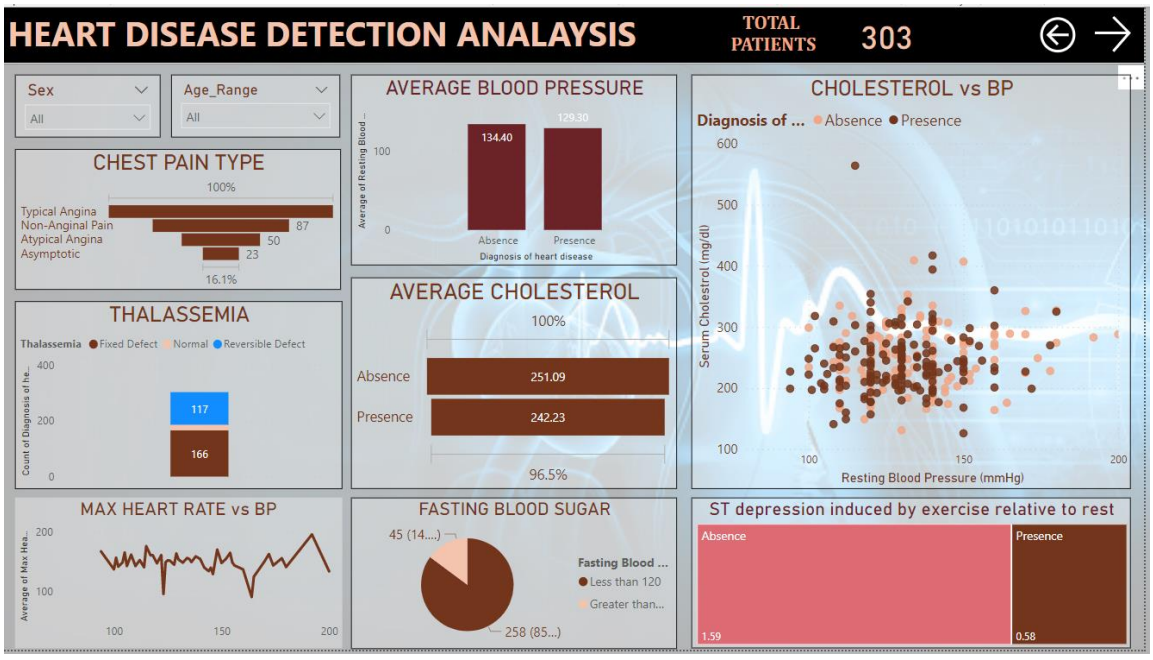


Figure 4: Heart Disease prediction Summary Dashboard

2. **Critical Factors:** This dashboard picks out the most important feature resulting in the presence of a heart disease and the magnitude at which it affects the presence.

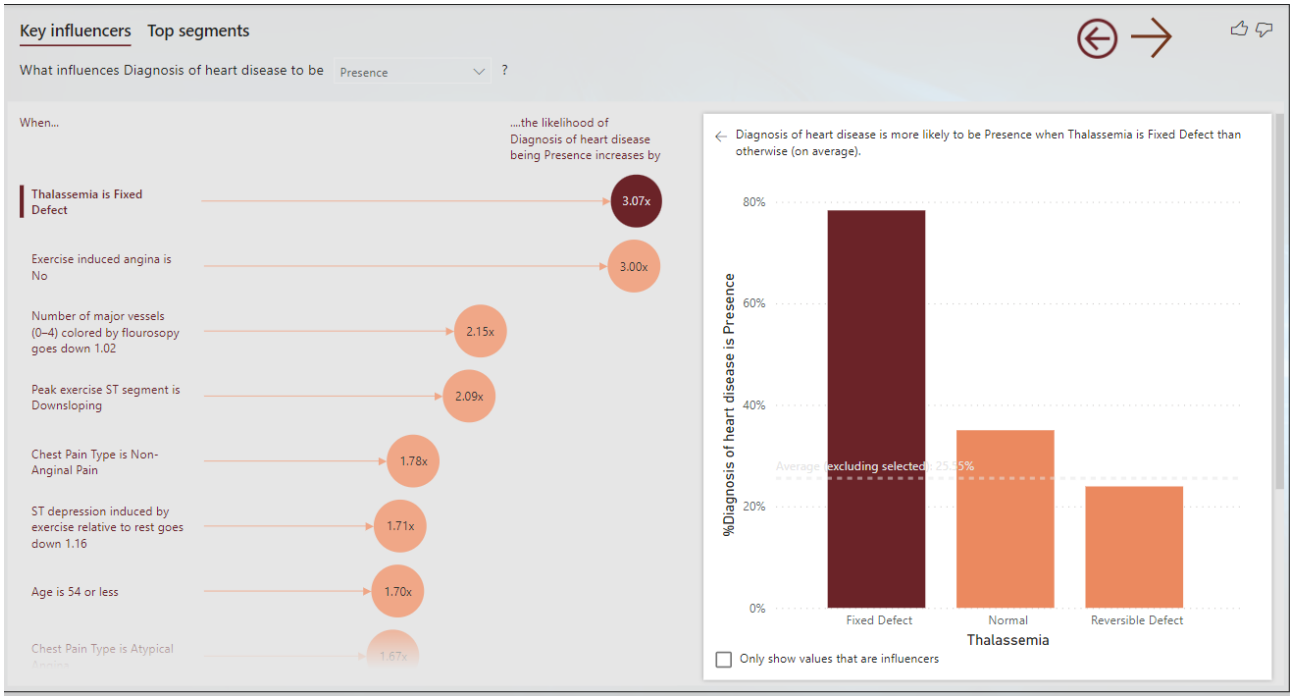


Figure 6: Critical Factors Dashboard

3. **Patient Profile:** This dashboard shows the profile of a patient that has a disease displaying the average age, BP level e.t.c

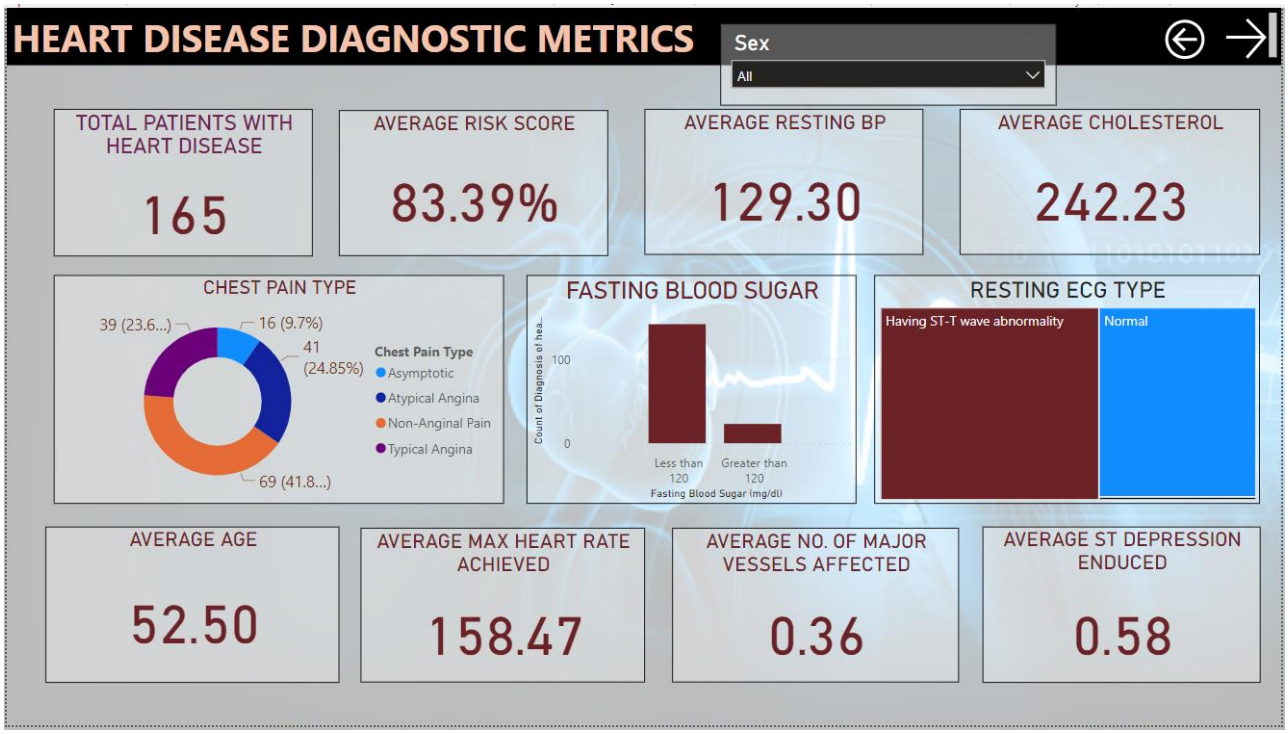


Figure 6: Patient Profile Dashboard

4. **Question & Answers:** This is a NLP enabled visual that allows end user to quickly create visuals based on the required specification.

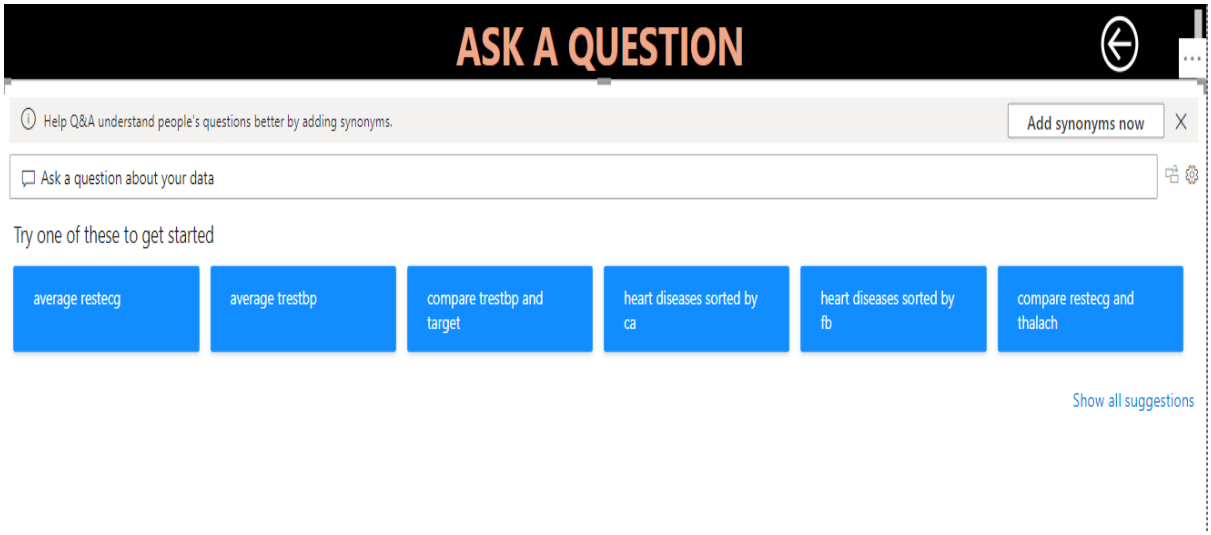


Figure 7: Q&A Dashboard

# 6 Exceptional scenarios

| Step | Exception | Mitigation | Module |
|---|---|---|---|
| **10th Dec 2021** | 1.0 | First Draft | Ankit Kashyap |
| | | | |

# 7 Key performance indicators (KPI)

Key indicators displaying a summary of the Housing Price and its relationship with different metrics

1. Percentage of People Having Heart Disease
2. Age Distribution including Gender
3. Gender Distribution Based on Heart Disease
4. Chest Pain Experienced by People Suffering from Heart Disease
5. Blood Pressure, Cholesterol Level and Maximum Heart Rate of People According to their Age and Heart Disease Patients.
6. ST Depression Experienced by People According to their age and heart disease.