Low Level Design

# Heart Disease Diagnostic Analysis

|  |  |
| --- | --- |
| **Written By** | Prakhyat Srivastava |
| **Document Version** | 1 |
| **Last Revised Date** | 13/04/2024 |

**DOCUMENT CONTROL**

## Change Record:

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **AUTHOR** | **COMMENTS** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **REVIEWER** | **COMMENTS** |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VERSION** | **REVIEW**  **DATE** | **REVIEWED BY** |  | **APPROVED BY** | **COMMENTS** |
|  |  |  |  |  |  |

**Contents**

1. **Introduction 04**
   1. **What is Low-Level Design Document? 04**

|  |  |  |
| --- | --- | --- |
| **1.2** | | **Scope 04** |
| **2.** | **Architecture 05** | |
| **3.** | **Architecture Description 08** | |
|  | **3.1** | **Data Description 08** |
|  | **3.2** | **Web Scrapping 08** |
|  | **3.3** | **Data Transformation 08** |
|  | **3.4** | **Data insertion into database 08** |
|  | **3.5** | **Connection with SQL server 08** |
|  | **3.5** | **Export Data from database 12** |
|  | **3.6** | **Deployment 12** |
| **4.** | **Unit** | **test cases 15** |

1. **Introduction**
   1. **What is Low-Level design document?**

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Heart Disease Diagnostic Analysis dashboard. LDD 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.

## 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

Cardiac ailments, encompassing a spectrum of disorders affecting the heart, have surged to the forefront of health concerns in India, mirroring a global uptick in prevalence and mortality. The statistics paint a stark picture—with heart disease mortality witnessing a staggering 34% increase from 1990 to 2016, escalating from 155.7 to 209.1 deaths per lakh of the population. In the face of such figures, heart disease stakes a claim as one of the leading causes of death across the nation. This escalating public health crisis calls for a proactive and data-centric approach to disease prediction and prevention. Leveraging a data-driven methodology, this project aims to refine the predictive modeling of heart disease, potentially revolutionizing research efforts and bolstering preventive measures to ensure enhanced longevity and well-being for India's populace.

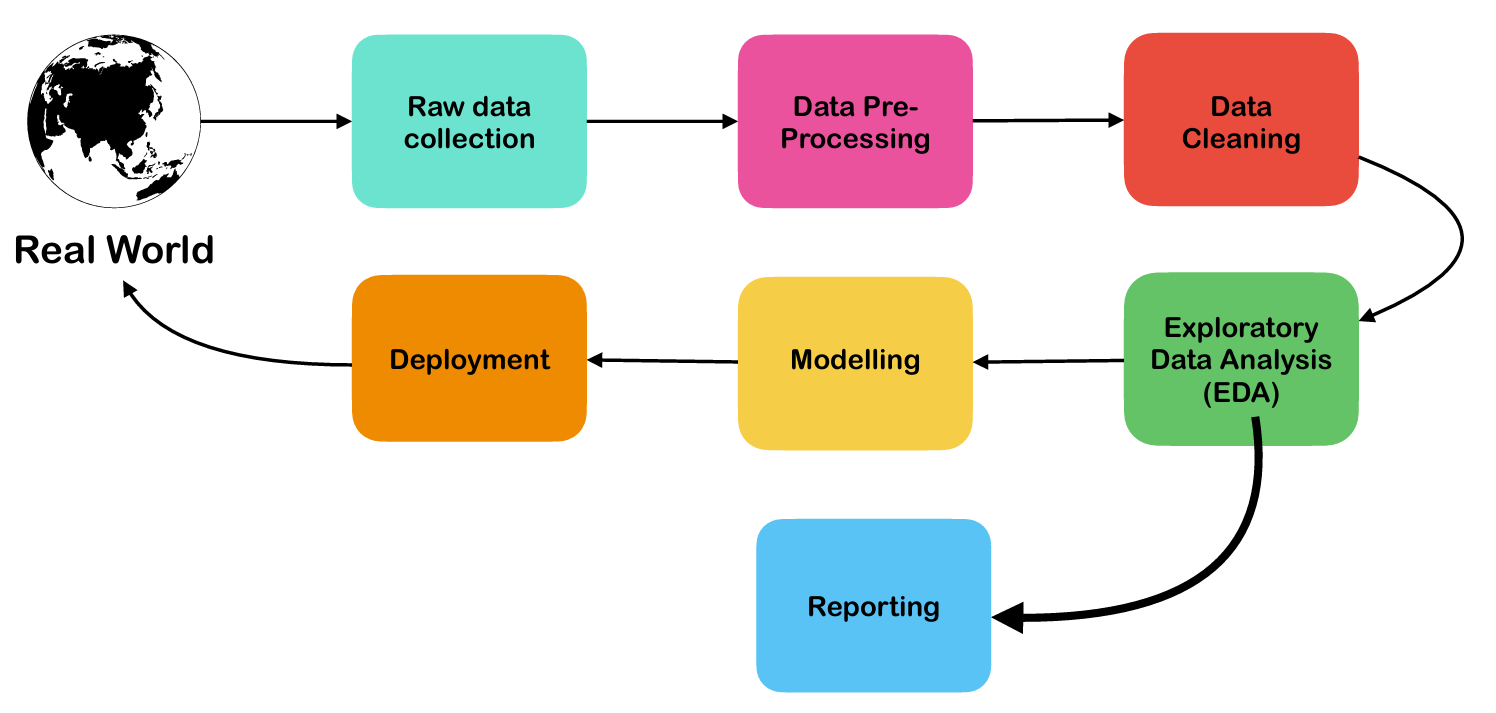
## 1.4 Problem Statement

## In an era where health stands as our most valuable asset, the ravages of the COVID-19 pandemic have starkly highlighted the vulnerability of human life regardless of societal stature. It is imperative, now more than ever, to delve deeply into medical datasets to arm ourselves for future medical challenges. This project calls for a thorough analysis of health data concerning heart diseases, encapsulating a diverse set of information from 303 individuals. The objective is to distill this data into meaningful patterns and predictive models that can serve as the cornerstone for robust, preemptive healthcare strategies and ensure a fortified stance against potential health crises.

## 1.3 Dataset Information

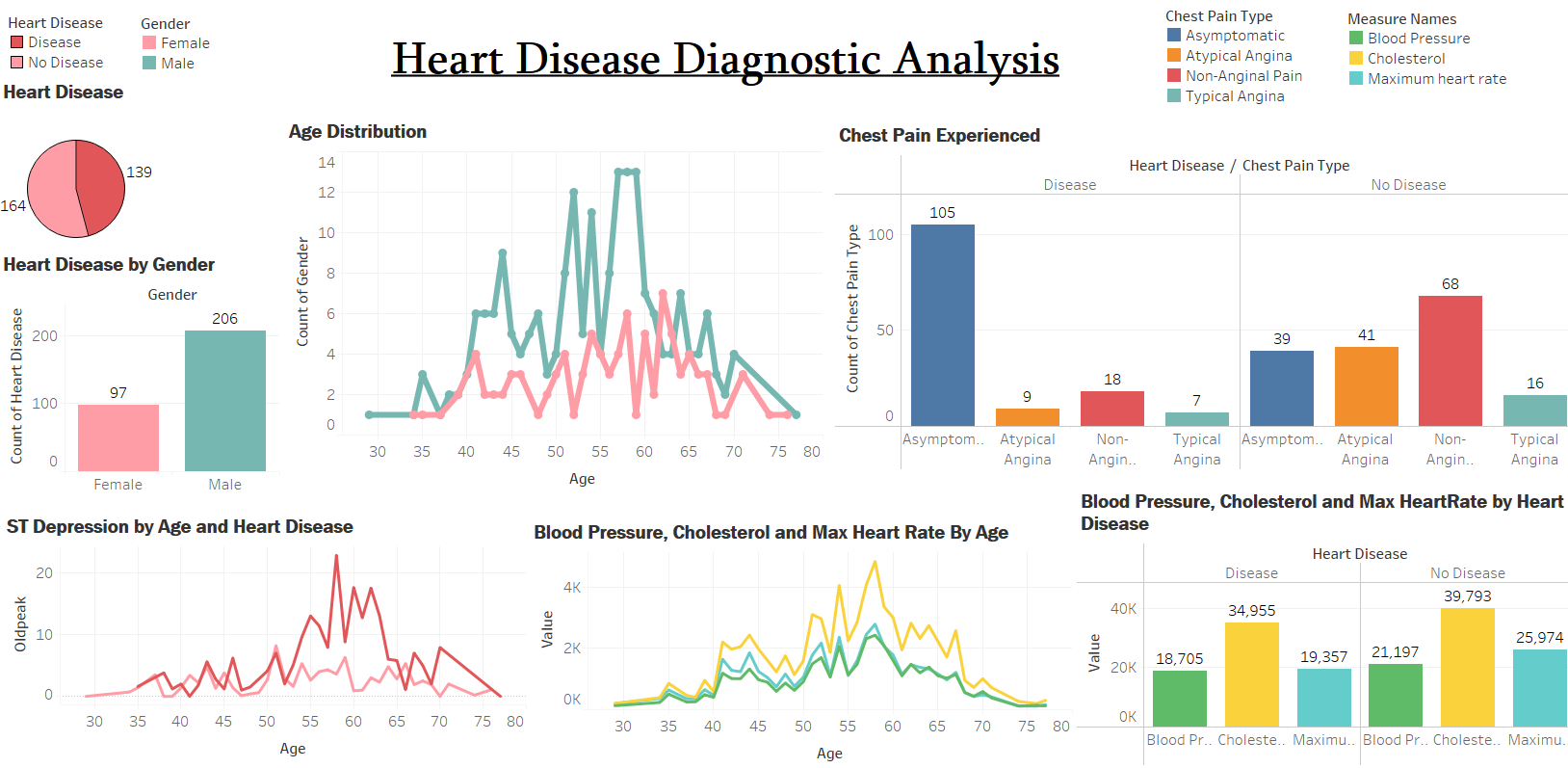
* **Age**: The patient's age in years.
* **Sex**: The patient's gender (1 = male, 0 = female).
* **Chest Pain Type (cp)**: Categorized into four types:
  + 1: Typical angina
  + 2: Atypical angina
  + 3: Non-anginal pain
  + 4: Asymptomatic
* **Resting Blood Pressure (trestbps)**: Measured in mm Hg at hospital admission.
* **Cholesterol (chol)**: Serum cholesterol level in mg/dl.
* **Fasting Blood Sugar (fbs)**: Indicates if fasting blood sugar is over 120 mg/dl (1 = yes, 0 = no).
* **Resting ECG (restecg)**: Resting electrocardiographic findings:
  + 0: Normal
  + 1: ST-T wave abnormality (inversions and/or ST elevation/depression > 0.05 mV)
  + 2: Likely left ventricular hypertrophy (Estes' criteria)
* **Max Heart Rate (thalach)**: Peak heart rate achieved during testing.
* **Exercise-Induced Angina (exang)**: Occurrence during exercise (1 = yes, 0 = no).
* **ST Depression (oldpeak)**: Measured in mm, induced by exercise relative to rest.
* **ST Segment Slope (slope)** during peak exercise:
  + 1: Upsloping
  + 2: Flat
  + 3: Downsloping
* **Major Vessels (ca)**: Number of major vessels seen in fluoroscopy (0-3).
* **Thallium Stress Test (thal)**:
  + 3: Normal
  + 6: Fixed defect
  + 7: Reversible defect
* **Disease Diagnosis (num)**: Angiographic disease status, ranging from no disease (0) to varying degrees (1-4).

# Architecture



# Architecture Description

1. **Data Acquisition**: The dataset, a rich tapestry of clinical information related to heart diseases, is sourced from the publicly accessible UCI Machine Learning Repository. This repository is a well-known fountainhead of datasets pivotal for analytical studies across various domains. [UCI Heart Disease Dataset](https://archive.ics.uci.edu/ml/datasets/heart+disease)
2. **Data Pre-Processing**: The cornerstone of a predictive model is the quality of the data it learns from. In this phase, the data is refined to ensure the model has the most accurate information. This includes: a) Rectification of Null or Missing Values b) Stabilizing Skewed Data Distributions c) Identification and Amelioration of Outliers
3. **Data Sanitization**: The integrity of the dataset is reinforced through rigorous cleaning protocols aimed at excising duplicate entries, extraneous outliers, and renovating the dataset to an optimal structure. This entails: a) Purging of redundant or irrelevant records b) Filtration of extraneous outliers c) Renaming of attributes for clarity and relevance
4. **Exploratory Data Analysis (EDA)**: EDA is the analytical step where patterns are identified, anomalies are highlighted, hypotheses are tested, and assumptions are examined through statistical summaries and graphical representations, setting the stage for subsequent modeling.
5. **Reporting**: The essence of effective communication in data analytics is clear and intuitive reporting, especially for stakeholders from a non-technical background. Essential reports include: a) High-Level Design Document (HLD) b) Low-Level Design Document (LLD) c) Architecture Overview d) Wireframes for application design e) Detailed Project Report (DPR) f) Powerpoint Presentation encapsulating findings and insights
6. **Modeling**: Data Modeling involves scrutinizing data entities and their interrelations, focusing on organizing the data needed for business processes. This process does not center on the operations on the data but rather on the structure of the data itself, preparing it for effective storage and retrieval in a database.
7. **Deployment**: The final model, honed and vetted through the prior stages, is deployed on a visualization platform such as Tableau. This facilitates interactive analytics, allowing end-users to engage with the model and glean insights through dynamic and informative visualizations.



# Unit Test Cases

|  |  |
| --- | --- |
| **TEST CASE DESCRIPTION** | **EXPECTED RESULTS** |
| Gender Distribution Slicer | Upon interaction with the slicer, a dropdown should display the gender distribution options. |
| Age Category Analysis Slicer | Selecting an option from the slicer should present age categories relevant to heart disease data. |
| Relationship Between Age and BP | A graph should illustrate the trend of blood pressure across different age groups. |
| Cholesterol Levels Across Age Groups | A visualization should plot the average cholesterol levels by age, indicating the correlation between age and cholesterol. |
| Heart Disease Prevalence by Gender | The visualization should reveal the proportion of heart disease occurrences segregated by gender. |
| ST Depression Analysis | A detailed view should emerge showcasing ST depression values across various age categories. |
| Chest Pain Type Distribution | A chart should display the distribution of chest pain types within the patient data set. |
| Data Quality Checks | The data set should pass checks for null values, outliers, and correct data types with no errors reported. |
| Predictive Model Accuracy | The predictive model should demonstrate an accuracy level above the pre-defined threshold when tested with validation data. |