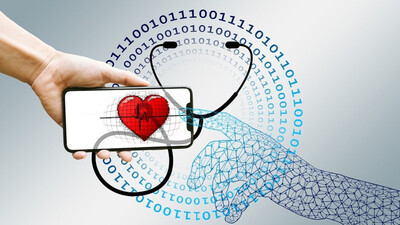


Insurance Premium Prediction

Low-Level Design (LLD)

Internship Project



Rahul Sharma

**DOCUMENT VERSION CONTROL**

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| **03-06-2022** | **1.0** | Introduction, Problem statement | **Rahul Sharma** |
| **08-06-2022** | **1.1** | Dataset Information Architecture | **Rahul Sharma** |
| **13-06-2022** | **1.2** | Final Revision | **Rahul Sharma** |

**Contents**

Document Version Control……………………………………………… 2

1. Introduction…………………………………………………………….4

1.1 What is a Low-Level Design Document?......................................... 4

1.2 Scope………………………………………………………………..4

1.3 Project Introduction………………………………………………... 4

1. Problem Statement.…………………………………………………… 5

2. Dataset Information…………………………………………………... 5

3. Architecture………………………………………………………… ...6

4. 4.1 Architecture Description ………………………………… ……….8

**1. INTRODUCTION**

**1.1 What is a 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 Diagnostic Analysis dashboard. LLDD describes the class diagrams, methods, and relations between classes and program 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.

**Project Introduction**

This project is to give people an estimate of how much they need to be based on their health situations. After that, customers can work with any health insurance carrier and its plans and perks while keeping the projected cost from our study in mind. This can assist a person in concentrating on the health side of an insurance policy rather than the ineffective part.

**2. Problem Statement**

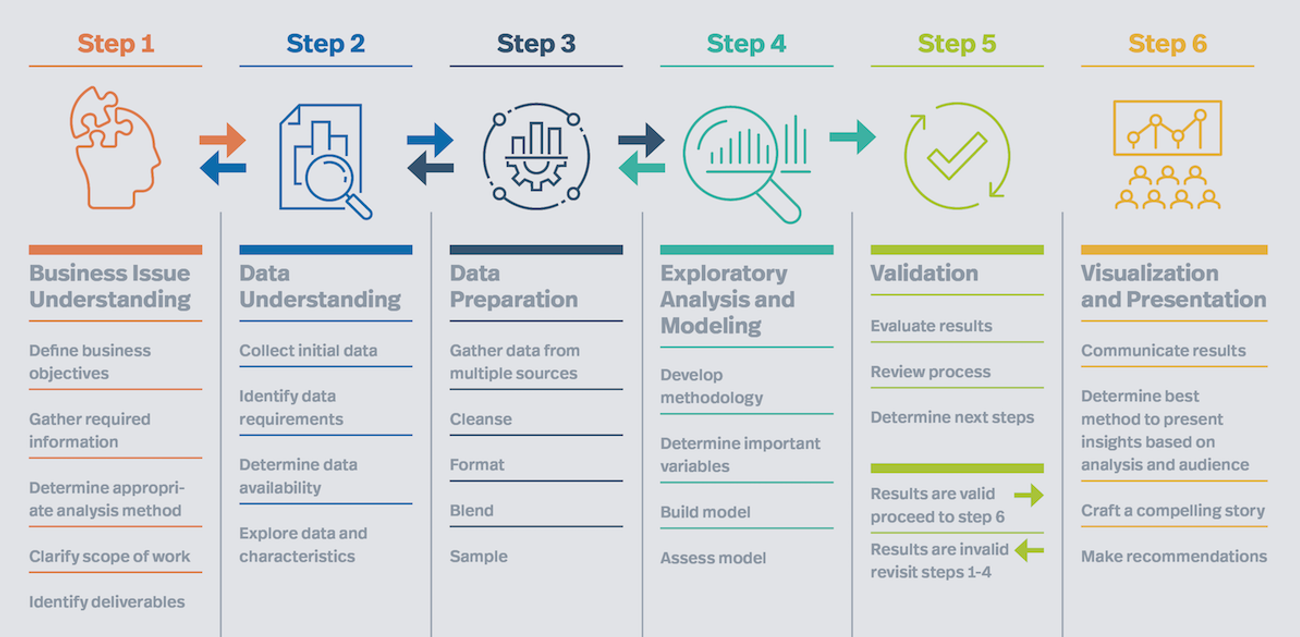
The goal of this project is to give people an estimate of how much they need to be based on their health situations. After that, customers can work with any health insurance carrier and its plans and perks while keeping the projected cost from our study in mind. This can assist a person in concentrating on the health side of an insurance policy rather than the ineffective part.

**3. Dataset Information**

The insurance\_data.csv dataset contains 1338 observations (rows) and 7 features (columns). The dataset contains 4 numerical features (age, BMI, children, and expenses) and 3 nominal features (sex, smoker, and region) that were converted into factors with numerical values designated for each level.

The purpose of this exercise is to look into different features to observe their relationship and plot a multiple linear regression based on several features of individuals such as age, physical/family condition, and location against their existing medical expenses to be used for predicting future medical expenses of individuals that help medical insurance to decide on charging the premium.

**4. Architecture**

****

**4.1 Architecture Description**

1. Raw Data Collection

The Dataset was taken from ineuron’s Provided Project.

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 fed to the model to train. This Process includes

a) Handling Null/Missing Values

b) Handling Skewed Data

c) Outliers Detection and Removal

3. 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) Filter unwanted 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 hypotheses, and check assumptions with the help of summary statistics and graphical Representations.

5. Reporting

Reporting is a most important and underrated skill in the data analytics field. Because being a Data Analyst, you should be good at easy and self-explanatory reports because your model will be used by many stakeholders who are not from a technical background.

a) High-Level Design Document (HLD)

b) Low-Level Design Document (LLD)

c) Architecture

d) Wireframe

e) Detailed Project Report

f) PowerPoint Presentation

6. Modelling

Data modeling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze 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 on what operations we have to perform.