**Project Title:**

**Heart Disease Risk Prediction and Analysis**

**Problem Statement:**

Cardiovascular diseases (CVDs) remain the leading cause of death globally, accounting for nearly one-third of total deaths each year. The increasing prevalence is influenced by factors such as sedentary lifestyles, unhealthy diets, stress, smoking, and genetic predispositions. Despite significant medical advancements, many individuals remain unaware of their risk levels until complications occur.

This project aims to analyze heart health data to identify key risk factors, understand geographical and age-wise trends, and predict the likelihood of heart-related health issues in the future. The insights will be presented through an interactive dashboard that supports public health awareness and early intervention strategies.

**Objective:**

The objective of this project is to analyze and predict heart-related health issues by identifying potential risk factors and forecasting the likelihood of heart attacks or related fatalities in the future.

Our goal is to assist in early detection, prevention, and better healthcare decisions through data-driven insights.

**Description:**

This project explores the relationships between various health indicators — including age, blood pressure, cholesterol, glucose, BMI, and lifestyle habits — to uncover meaningful patterns.  
By applying statistical and machine learning techniques, we aim to:

* Identify key factors influencing heart disease risks
* Predict the probability of heart attacks or cardiac events
* Visualize mortality trends and risk distributions
* Provide actionable insights to support preventive healthcare

**Data Requirements:**

We need datasets that include:

* Patient demographics: age, gender, region
* Medical indicators: blood pressure, cholesterol, heart rate, glucose, BMI
* Lifestyle factors: smoking, alcohol consumption, diet, physical activity
* Health outcomes: diagnosed heart disease (yes/no), risk level

To Find(Data Source):

* Kaggle: Heart Disease UCI Dataset, Framingham Heart Study, Cardiovascular Disease Dataset
* WHO / CDC / [HealthData.gov](http://healthdata.gov/) / India’s data portals: For regional statistics
* Optional: Synthetic lifestyle survey data if needed

**Software & Tools:**

|  |  |
| --- | --- |
| **Task** | **Tools** |
| Data Cleaning & Analysis | Python (pandas, numpy) |
| Data Visualization | matplotlib, seaborn, plotly |
| Machine Learning | scikit-learn, XGBoost |
| Dashboard | Tableau, Power BI |
| Version Control | GitHub |
| Project Management | Jira |
| Documentation | Confluence |

**Development Steps:**

1. Define Problem Statement
2. Data Gathering
3. Data Preprocessing
4. Exploratory Data Analysis (EDA)
5. Model Building & Evaluation
6. Dashboard & Visualization
7. Documentation & Final Presentation

**Deliverables:**

* Cleaned dataset
* ML model (with evaluation metrics)
* Interactive dashboard
* Project documentation & presentation
* GitHub repository with code

***“Prevention is better than cure — let's predict before it’s too late.”***