Data Analysis Project

Project Title: Heart Disease Data Analysis
Data Analyst: Shijin Ramesh
Client/Sponsor: Dataset obtained from Kaggle

Purpose:

This project involves exploring a heart disease dataset using data visualization and statistical techniques. The goal is to uncover key patterns and relationships among clinical features that contribute to the presence of heart disease. By performing Exploratory Data Analysis (EDA), the project aims to deliver meaningful insights that could support early detection and decision-making.

Scope / Major Project Activities:

Activity	Description
Data Collection	Load the dataset sourced from Kaggle
Data Cleaning	Handle missing values, remove outliers (e.g., cholesterol = 0)
Data Exploration	Understand distributions, identify trends using descriptive statistics
Feature Analysis	Explore key variables and their relationship with the target variable
Visualization	Use charts such as heatmaps, histograms, and violinplots to visualize trends
Correlation Analysis	Identify important features correlated with heart disease
Insight Generation	Write clear, actionable insights after each visual and analysis section
Reporting	Prepare project report and README with insights, code, and conclusions

---

This project does not include:

- Building predictive machine learning models (this is limited to EDA only)
- Clinical diagnosis or decision-making advice
- Longitudinal or time-series data modeling

Deliverables:

Deliverable	Description/Details
Cleaned Dataset	Refined dataset with outliers and nulls addressed
Jupyter Notebook (.ipynb)	Final Python notebook including code, visuals, and markdown cells
Project Report (.pdf)	Summary report for non-technical stakeholders
README.md File	Project overview for GitHub documentation
requirements.txt	Required Python libraries for running the project

Schedule Overview / Major Milestones:

Milestone	Expected Completion Date	Description/Details
Data Cleaning Completed	Day 1	Outliers removed and missing data handled
Exploratory Analysis Completed	Day 2	Visualizations and feature insights created
Final Report Compiled	Day 3	PDF and README finalized
GitHub Project Published	Day 4	Notebook, report, and requirements uploaded to GitHub

Estimated Completion Date:

Within 4 days of project start, assuming daily progress.