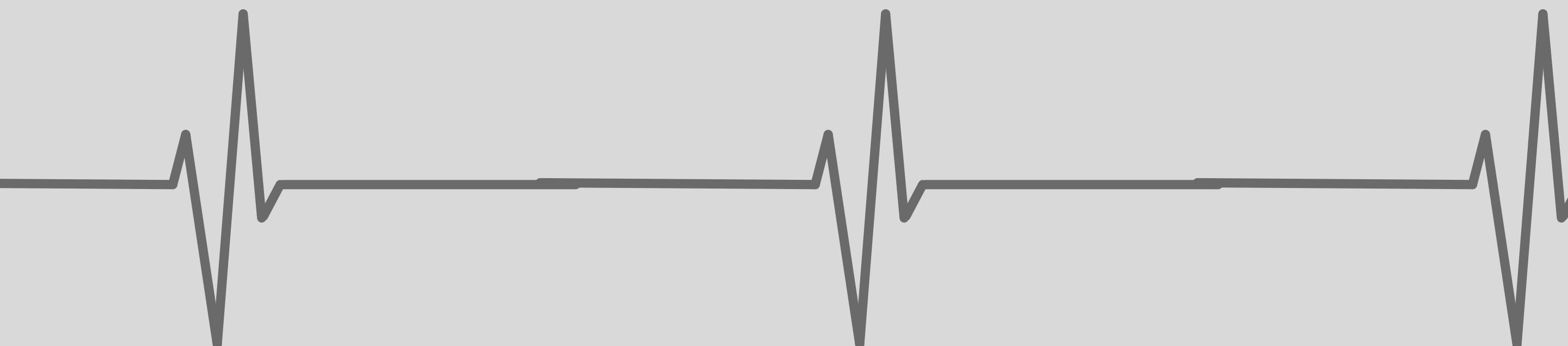
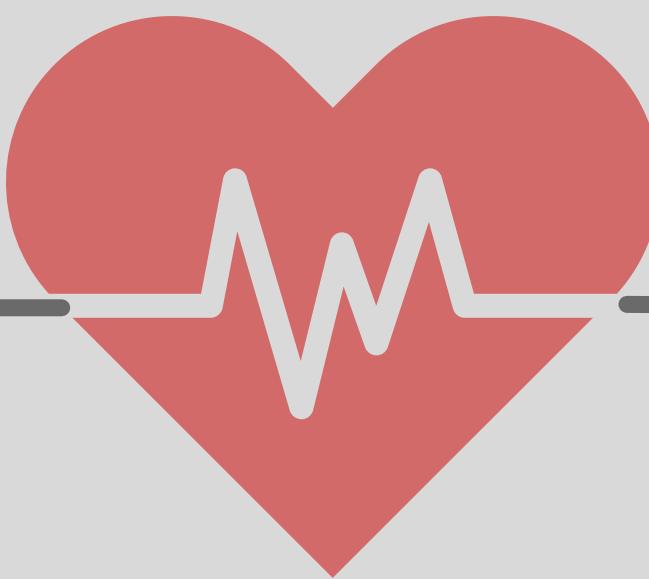


Health Disease Diagnosis

An analytic and predictive approach



Introduction:

- With the increasing environmental changes and rapid evolution, human bodies have become more prone and sensitive to deadly diseases than before.
- Cardiovascular disease refers to any disease that affects the cardiovascular system, including coronary heart disease, myocardial infarction or heart attack, stroke or cerebrovascular accident (CVA), vascular diseases, and/or peripheral arterial diseases.
 - One person dies every 33 seconds from a cardiac disease.
 - This increases the necessity to stop feigning ignorance and detect the disease to aid its diagnosis.

Objective:

- The aim of this data science project focuses on determining the various health-related factors that are direct or indirect reasons for the increasing numbers of heart patients.
-
- By establishing relations between various attributes, the project also provides a model that serves to detect heart diseases, which in turn can reduce the number of cases by predicting them in their early stages

Main KPI's:

- KPI stands for key performance indicator, a quantifiable measure of performance over time for a specific objective.
- In this dataset, exploratory data analysis and accurate models provided top five determining KPI's. They are
 - chest pain type (cp),
 - number of major vessels (ca)
 - maximum heart rate achieved during stress test (thalach)
 - ST depression induced by exercise (oldpeak)
 - thalium stress result (thal)

Tools Used:

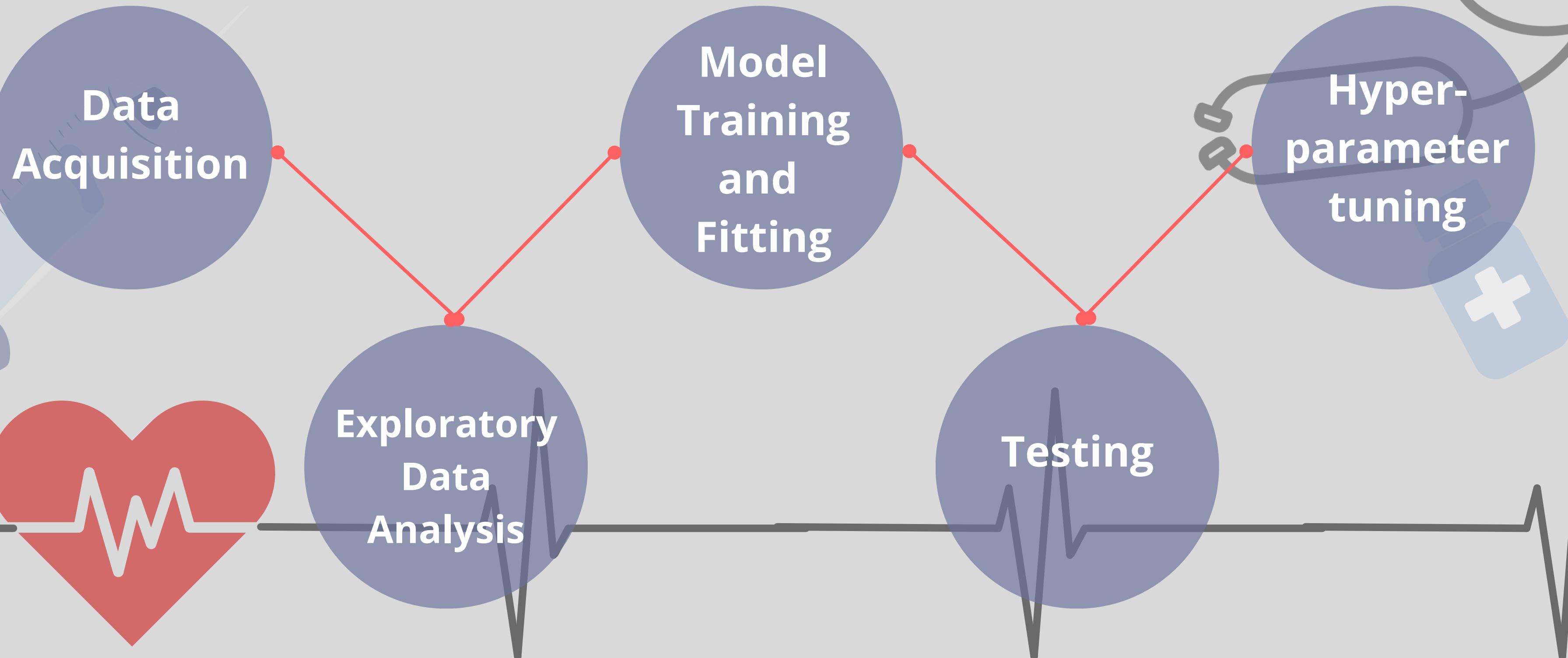
Data Processing

- Language : Python
- IDE : Jupyter Notebook
- Libraries :
 - numpy
 - pandas
 - scikitlearn
 - matplotlib
 - seaborn

Data Analysis

- Tableau Public

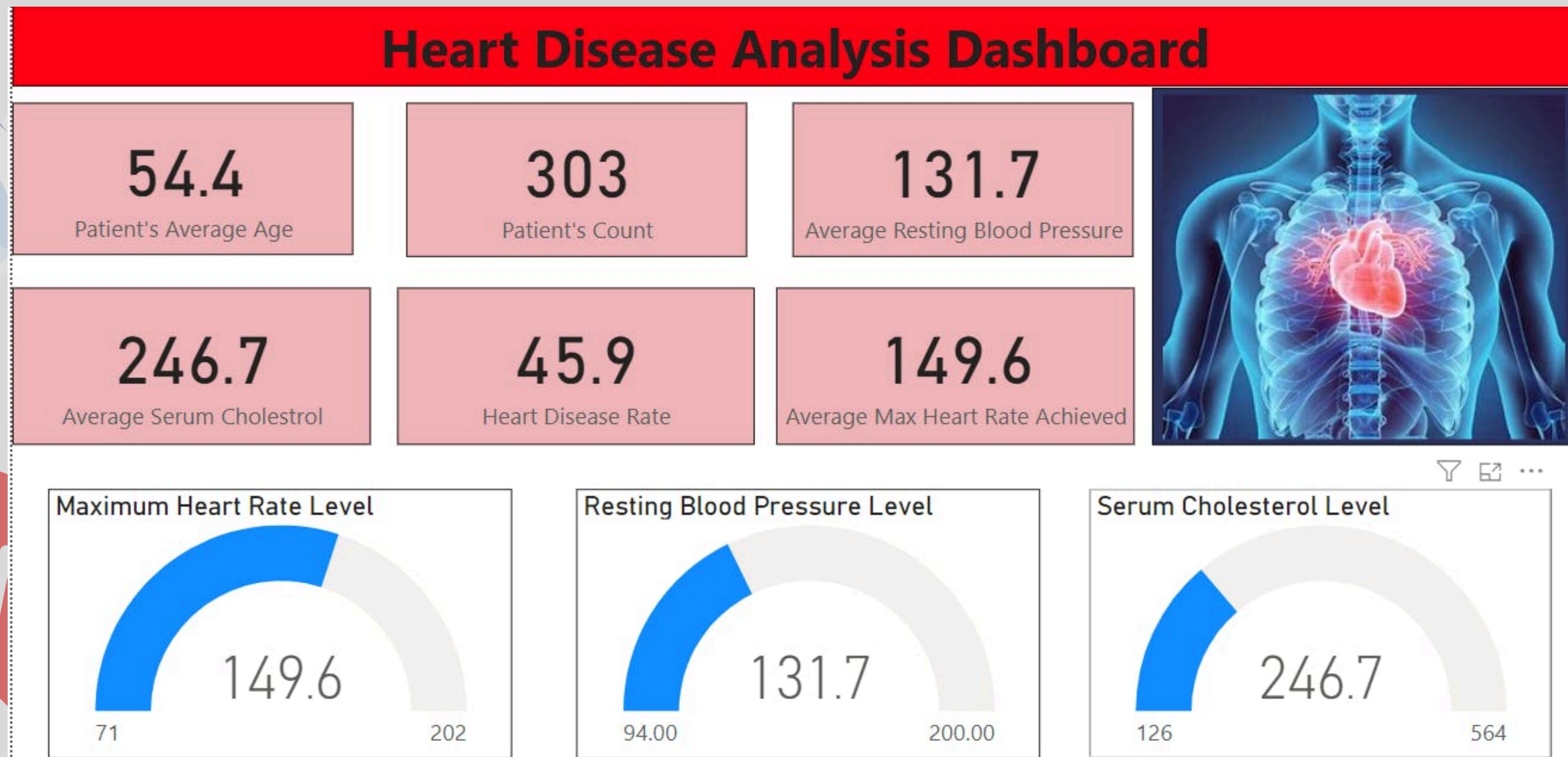
Process Involved:



Model Choosen:

- The data containing various patient records was split into two parts for trainings, and later fitted into 4 different models.
- Keeping the accuracy score of the model to be the basis for selection, the Random Forest Classifier Model showed the highest accuracy score of 97%.
- The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees. Random Forests are particularly well-suited for handling large and complex datasets, dealing with high-dimensional feature spaces, and providing insights into feature importance.

Mock Dashboard



Metrics Overview

Analysis-I

Analysis-II

Analysis-III

Analysis-IV

Analysis-V

Analysis-VI

Pie Chart Analysis

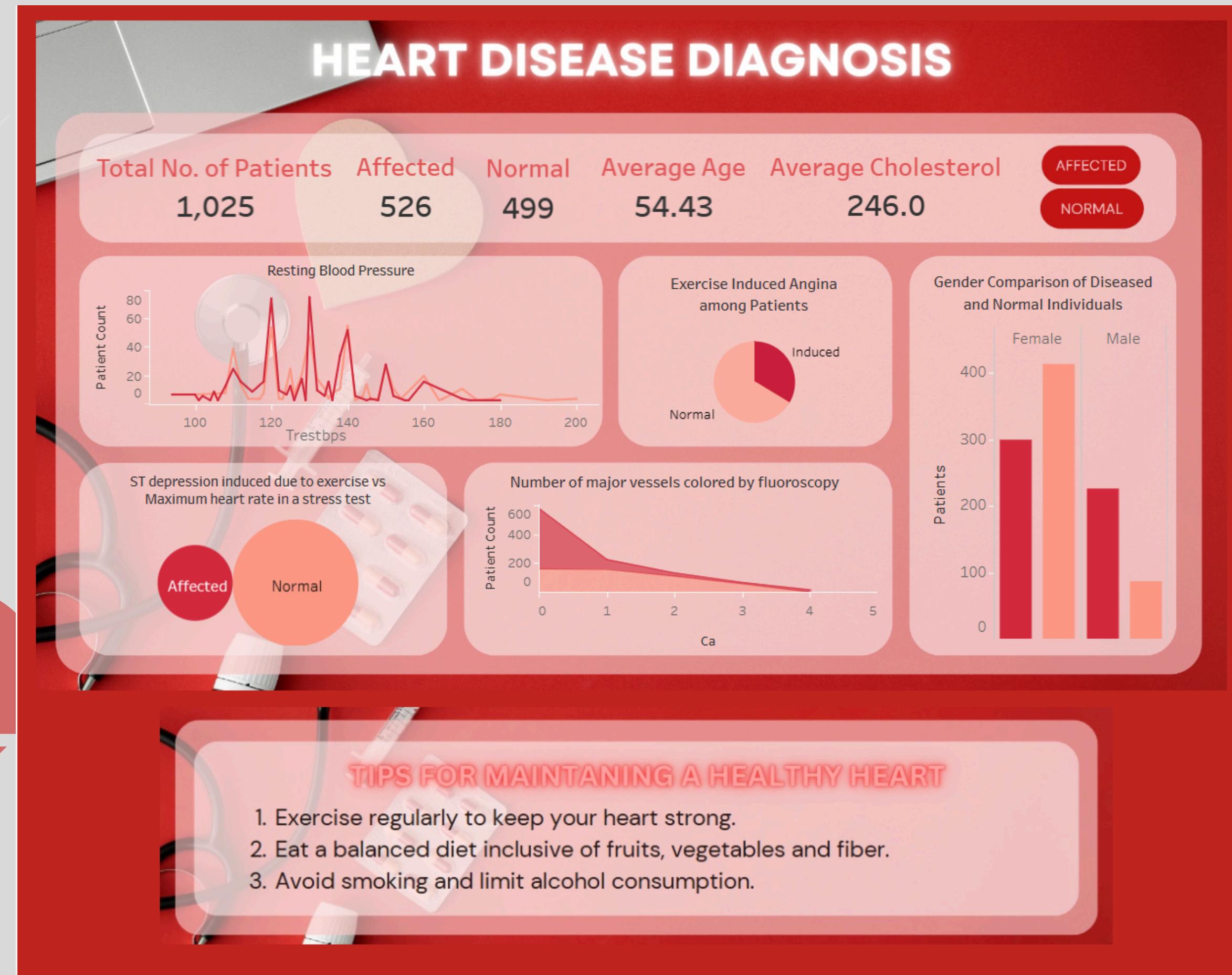
Line Chart Analysis

Numerical Analysis

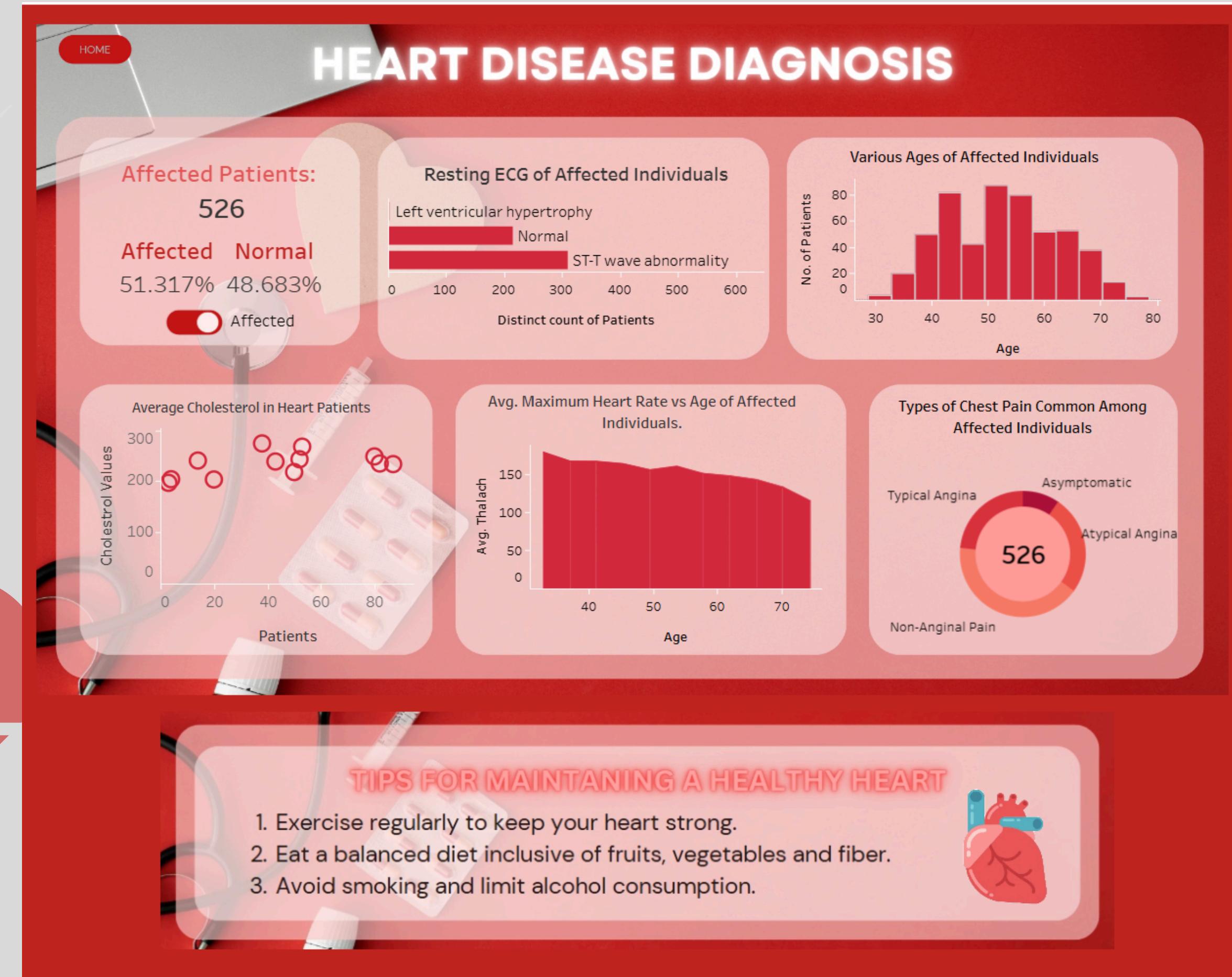
Statistical Analysis

Interactive Filters And Navigation

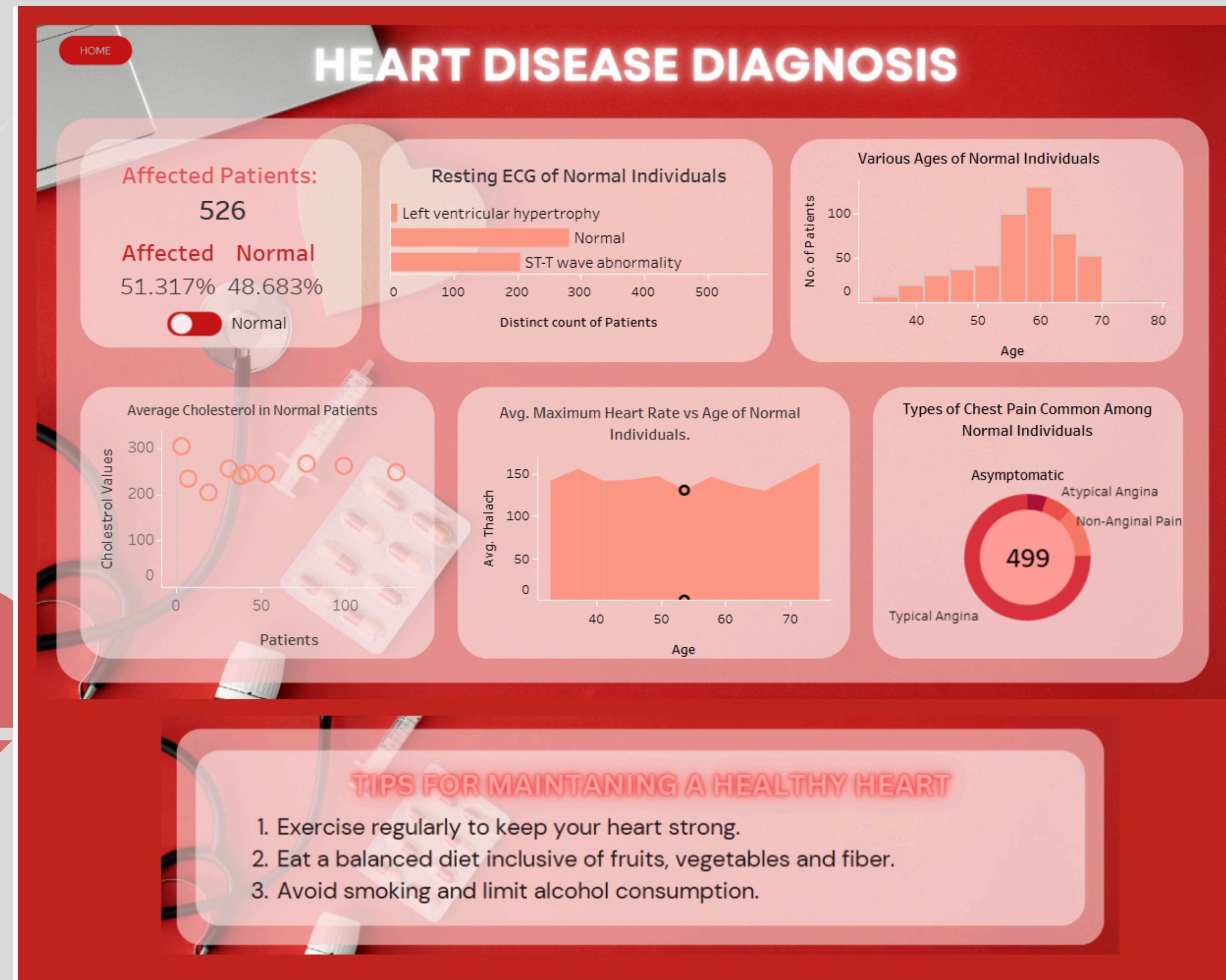
My Dashboard:



My Dashboard:



My Dashboard:



Links:

These are the links for the python and tableau project:

[Python Project Github Link](#) (Click)

[Tableau Link](#) (Click)

Conclusion:

- Out of the 13 attributes that determine the target attribute, chest pain type (cp), number of major vessels (ca), maximum heart rate achieved during stress test (thalach), Thalium stress result (thal) and ST depression induced by exercise (oldpeak) prove to be the most influential factors.
- The Random Forest Classifier Model has shown the highest accuracy with a score of 0.977, hence this model was chosen to experiment.
- This heart disease diagnosis model can be used for various predictive analysis to obtain maximum accurate results.