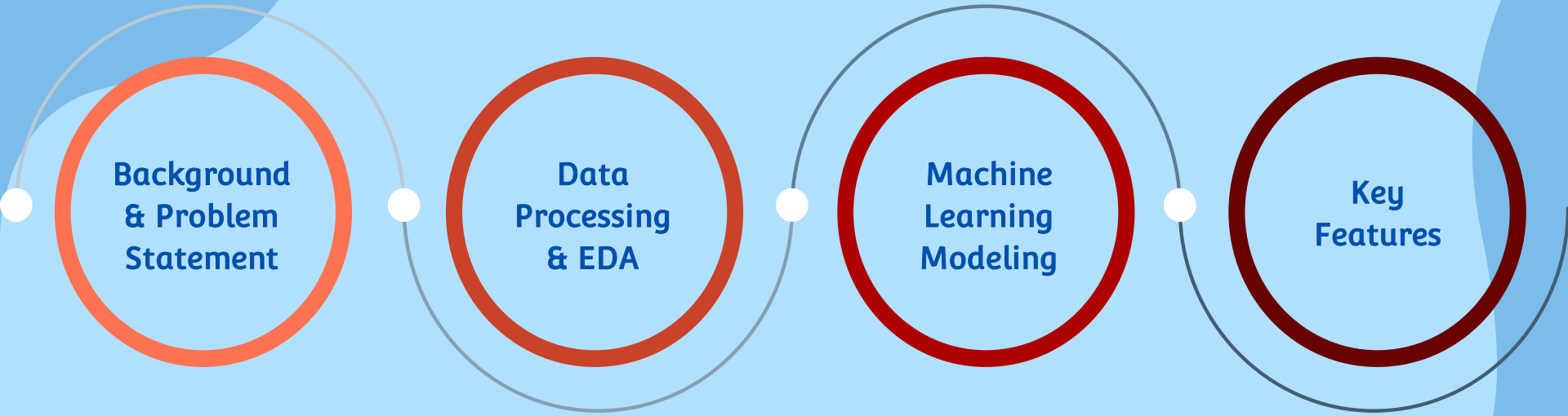
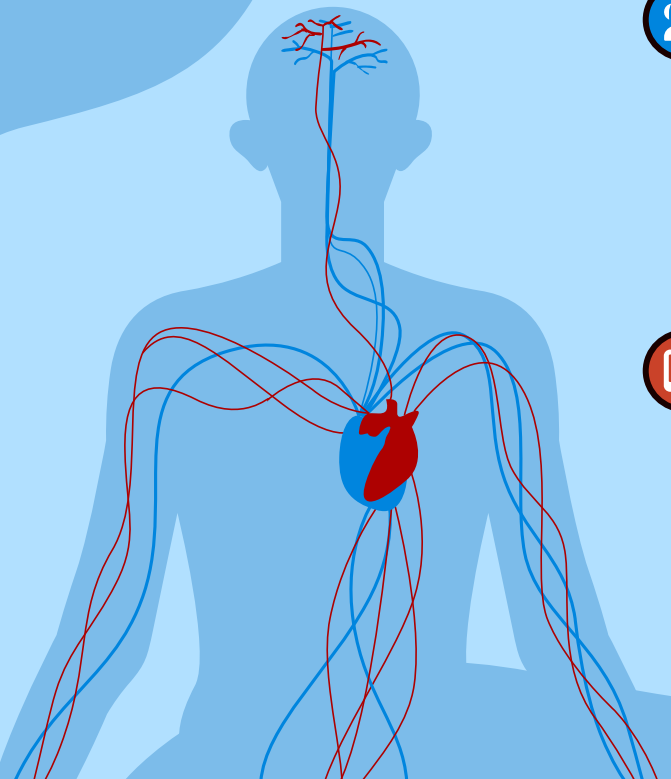
HEART DISEASE PREDICTION

Kate Yu | BrainStation | Capstone Project

Project Overview



Background



Fact of Heart Disease:

In every **36 sec**

one person dies in the United States from cardiovascular disease, that's 1 in every 4 deaths! [1]

Traditional Ways to Diagnose Heart Disease:

- Blood tests, electrocardiogram, cardiac computerized tomography scan etc. [2]

Time Consuming & Expensive & Invasive

[1]: "Heart Disease Fact" , *CDC government*
[2]: "Diagnosis Coronary Heart Disease", *NHS*

Problem Statement

Goal of this project:

Predict heart disease possibility based on current health status

Who can benefit from this?

- Everyone!! Improve awareness of heart health
- Health organizations: make proactive treatments and distribute medical resource efficiently

Data Source

Original source: CDC Organization

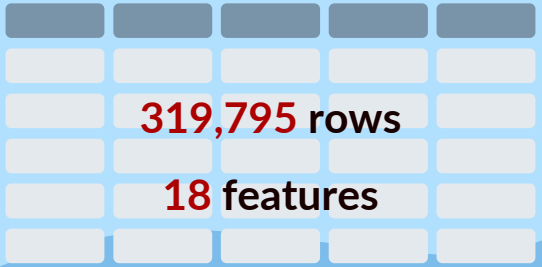
2020 annual CDC survey data of 400k adults related to their health status (300 columns)

CDC Centers for Disease Control and Prevention

Direct source: Kaggel

Condensed version with less features in a single csv file

kaggle



Data Preprocessing

Numeric Features

- Examples:
  BMI, Sleep Time
- Processing:
  Check distribution

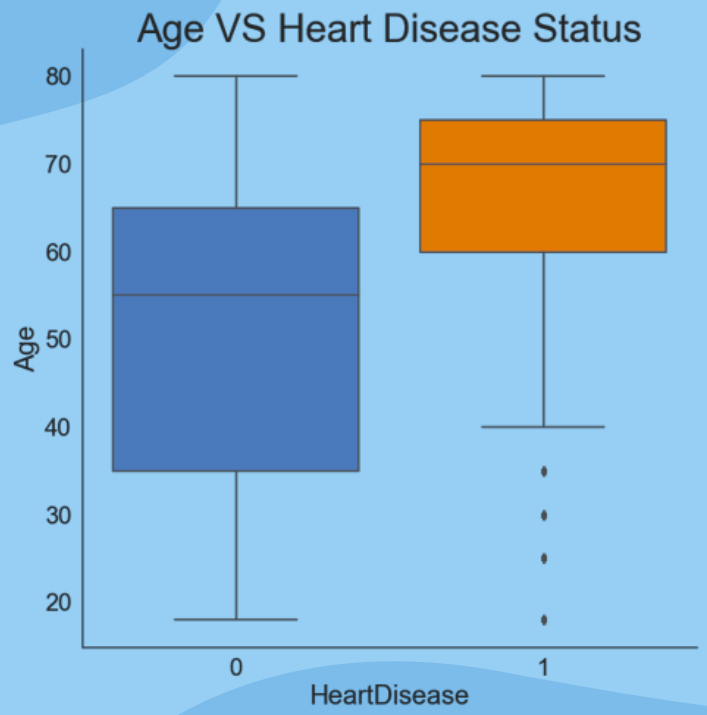
Binary Features

- Examples:
  Diabetic, Smoking
- Processing:
  Change to 1/0

Multi-classes Features

- Examples:
  Age Category, Race
- Processing:
  Change to a series of number, dummy variable

Exploratory Data Analysis



Age is the most important factor lead to heart disease.

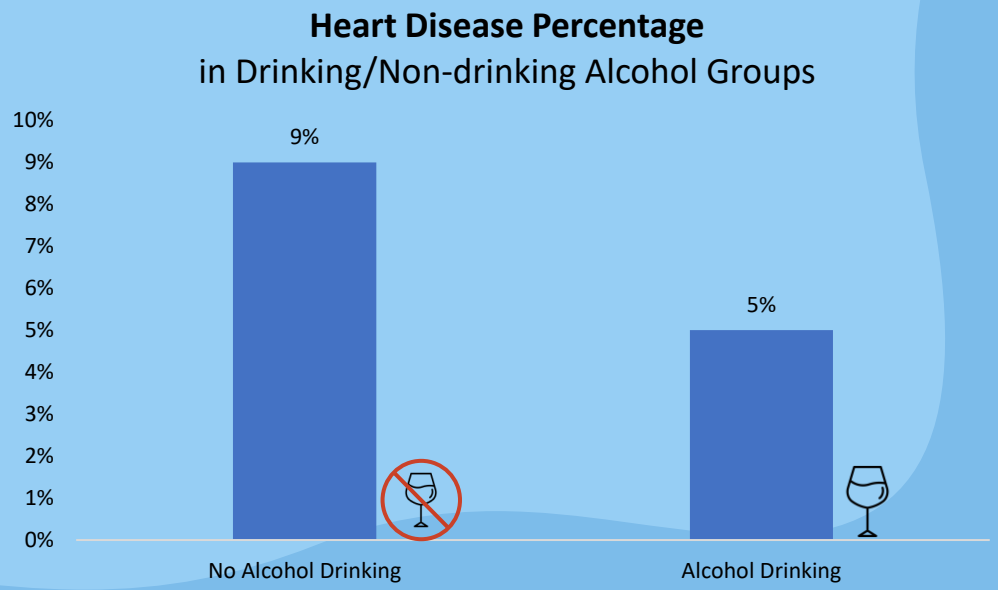
The average age of people who have heart disease is 66 years old, which is **15 years older** than the average of people who do not have the disease.

Exploratory Data Analysis

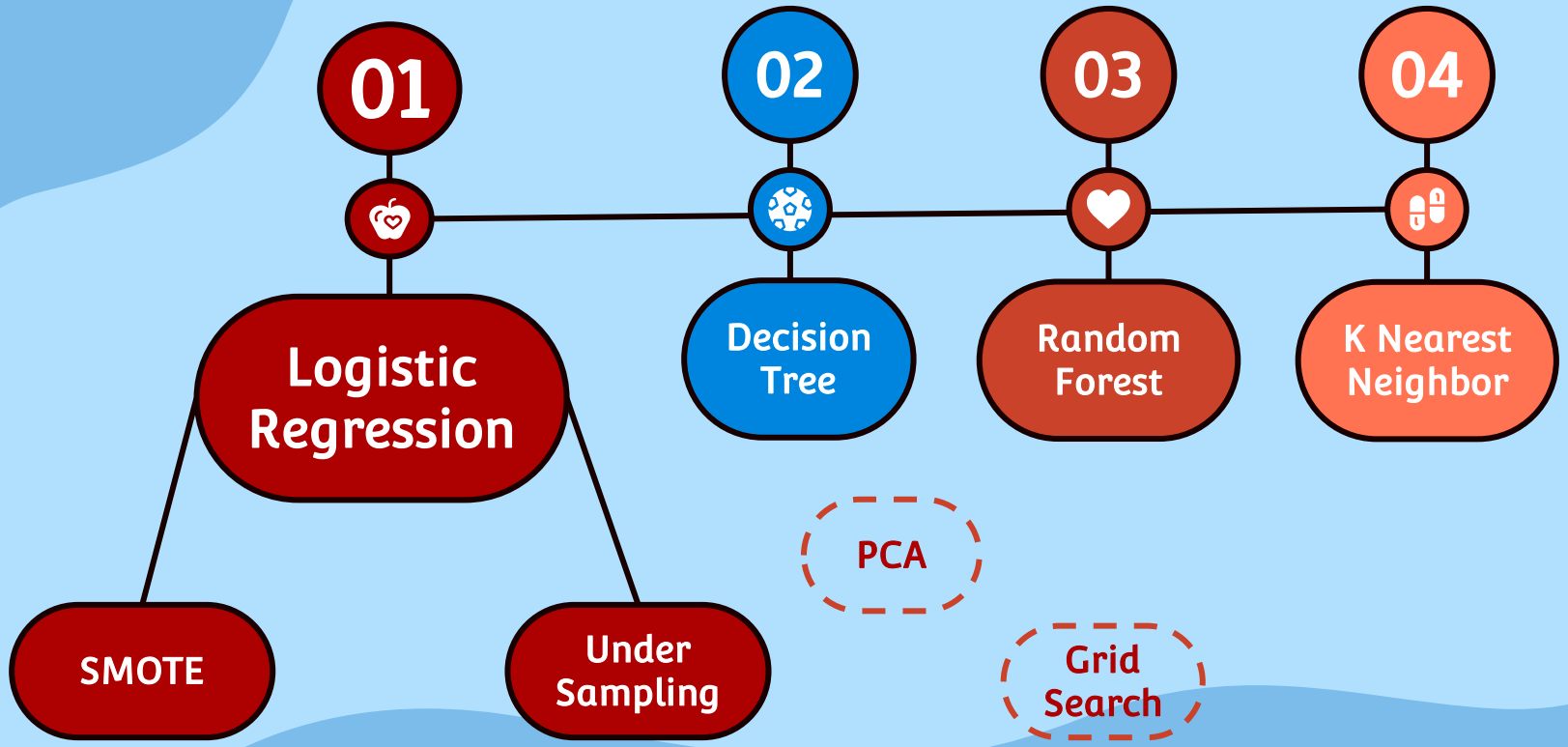
Drink Alcohol helps to **PREVENT** heart disease!!

In non-alcohol drinking group, there are **1.8 times** more possibility of getting heart disease compared to alcohol drinking group.

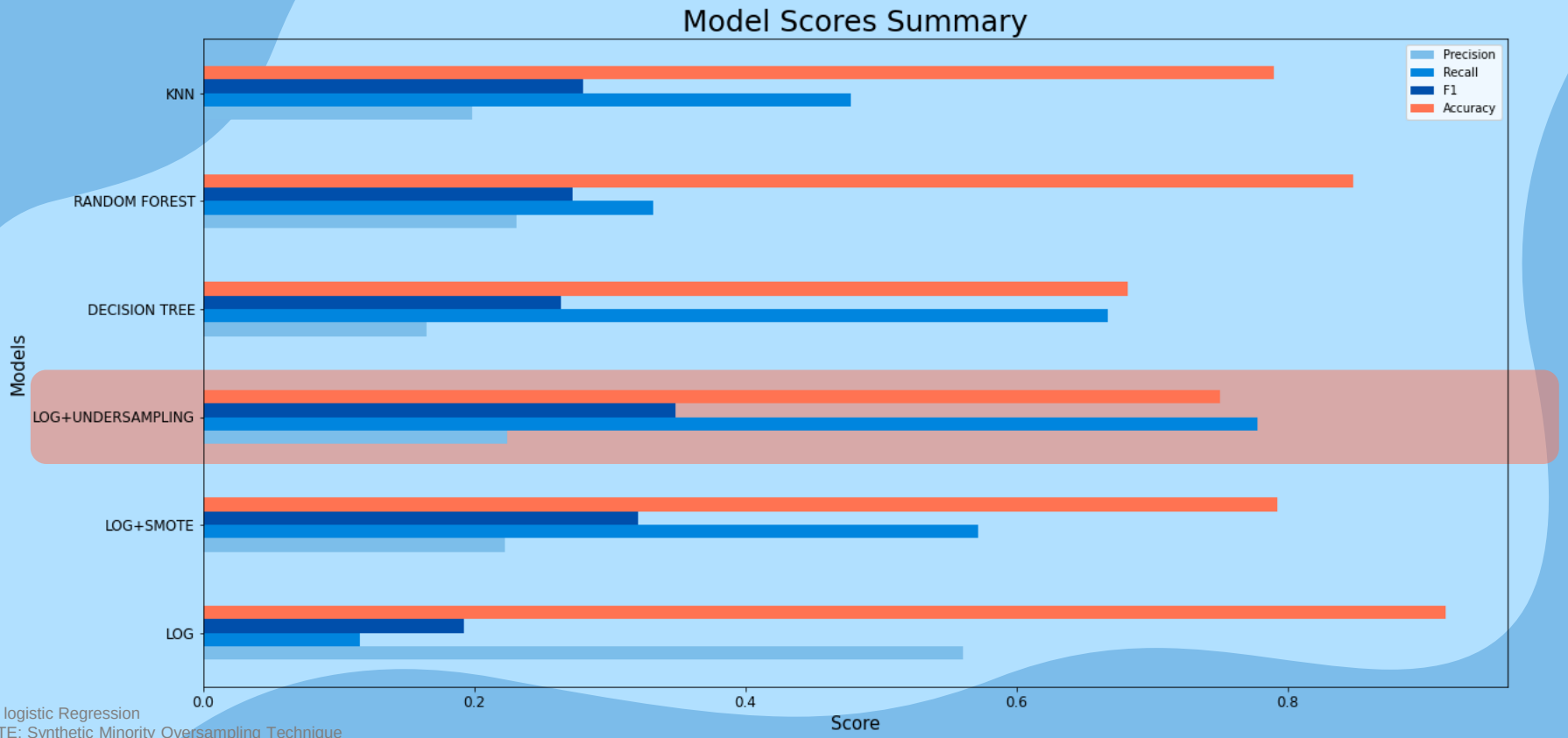
	No Heart Disease	Heart Disease
No Alcohol Drinking	91%	**9%**
Alcohol Drinking	95%	**5%**



Machine Learning Models



Model Evaluation



LOG: logistic Regression
SMOTE: Synthetic Minority Oversampling Technique
KNN: K Nearest Neighbour

Key Features

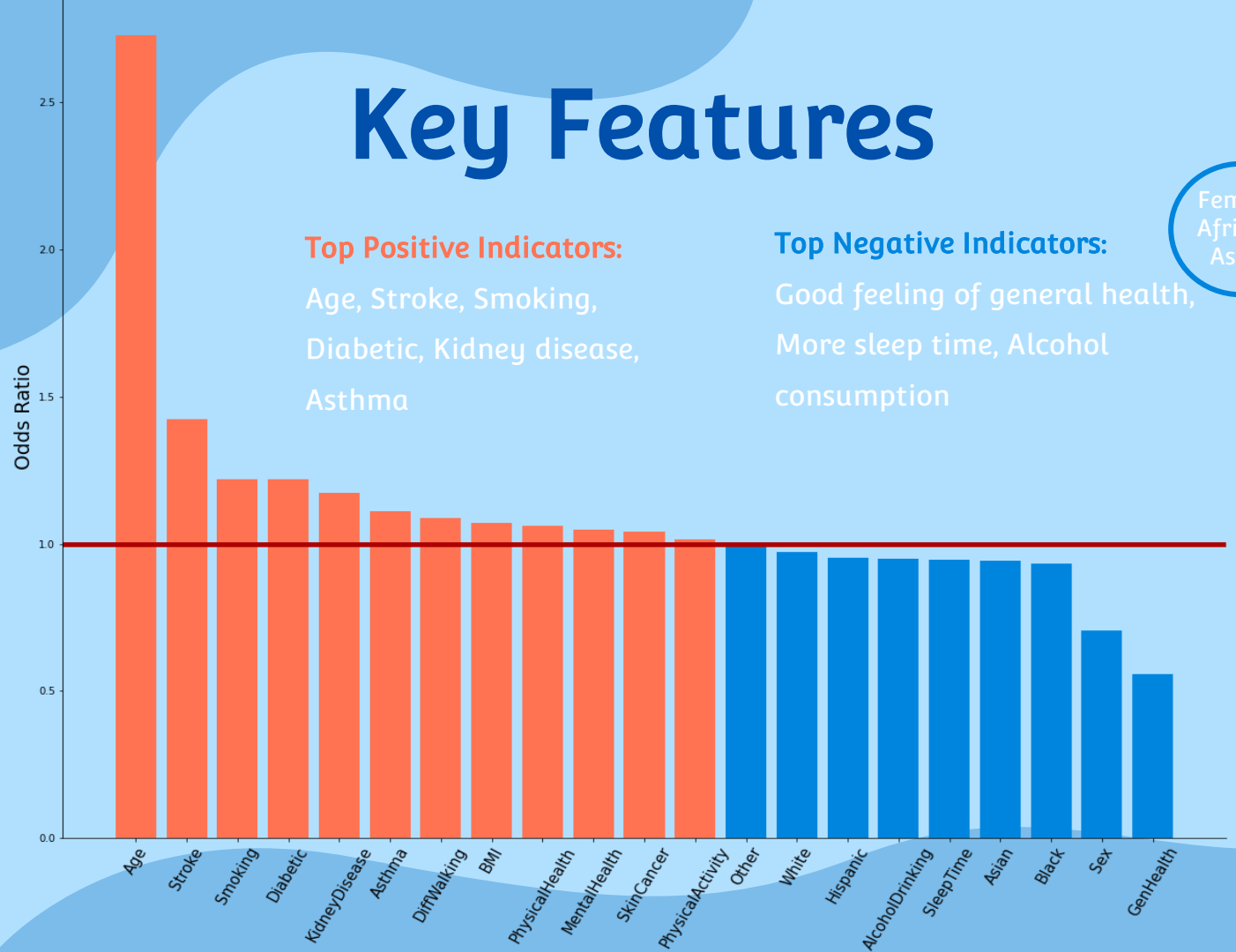
Top Positive Indicators:

Age, Stroke, Smoking, Diabetic, Kidney disease, Asthma

Top Negative Indicators:

Good feeling of general health, More sleep time, Alcohol consumption

Female, African, Asian



Application

app_building_code - Streamlit

localhost:8501

💔 Current Health Status 💔

How old are you?
33
18 80

How tall are you? (cm)
175
150 200

How much do you weight? (kg)
80
40 120

Average sleep time
8
4 12

Days in a month feel physical health is NOT good
0
0 31

Days in a month feel mental health is NOT good
0
0 31

What sex are you?
Female

Heart Disease Prediction APP

In every 36 seconds,

one person dies in the United States from **cardiovascular disease**.

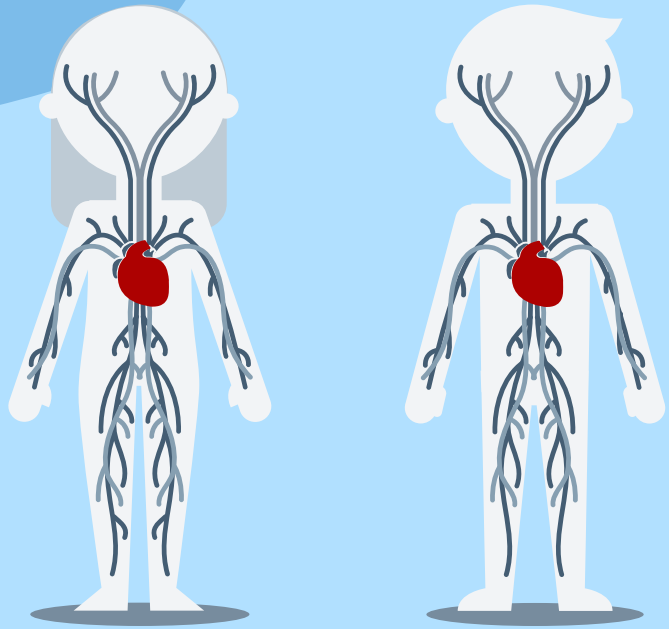
About 659,000 people in the United States die from heart disease each year—that's **1 in every 4 deaths**!

This application is build based on 32k responses from CDC questionaire, and applied optimal machine learning model, in order to help you predict the possibility of getting heart disease more accurately.

(CDC: Centers for Disease Control and Prevention)

💝 Click Here to Test 💝

Made with Streamlit



THANKS!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**