Be Heart Smart

The Healthy Healthcare Enthusiasts (Collaborators):
(Final-Project Group 7)

- Ayse Ozgun
- Pam Noble
- Subhangi Ghosh
- Krishnakali Sarkar

Cardiovascular Disease (CVDs)

Disorders of the heart and blood vessels including coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions.
Leading cause of death globally ~ 40% deaths in the US.

Leading Behavioral Risk Factors :

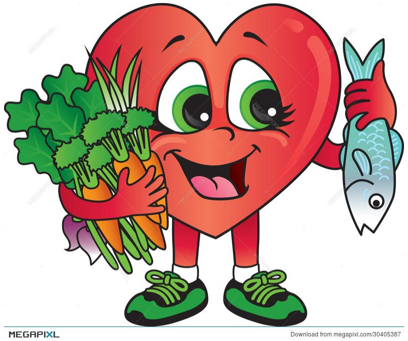
- Unhealthy diet,
- Physical inactivity
- Tobacco use
- Harmful use of alcohol

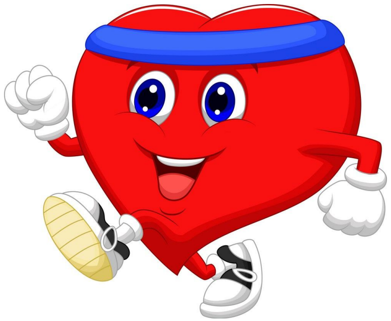
Effects of behavioral risk factors :

- Raised blood pressure,
- Raised blood glucose,
- Raised blood lipids,
- Overweight and
- Obesity.

A healthy heart is a happy heart

The purpose of this project is to spread awareness. Embracing a healthy lifestyle at any age can help prevent heart disease, and lower the risks for heart attack or stroke.





About the data

Website : Cardiovascular Disease dataset (Kaggle)

Description :

Three types of input features

- Objective
- Examination
- Subjective

Objective	**Examination**	**Subjective**
Age (days)	Systolic Blood Pressure	Smoking
Height (cm)	Diastolic Blood Pressure	Alcohol Intake
Weight (kg)	Cholesterol	Physical Activity
Gender	Glucose	

Target Variable : Presence or Absence of Cardiovascular Disease

Questions we hope to answer with the data:

- Is a person at risk of heart disease?
- What are the potential risk factors for heart disease--smoking, alcohol consumption, obesity, etc?
- Which factors are the best predictors of heart disease?

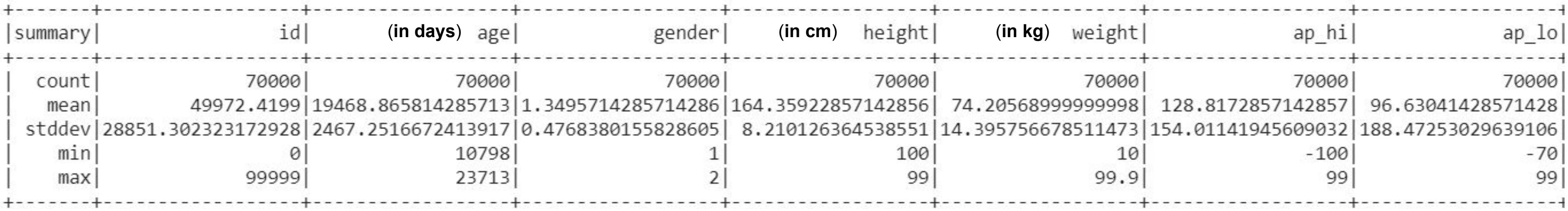
Classification model to predict risk (Yes/No) of heart disease based on different factors

- Supervised Machine Learning
  - Logistic Regression
  - Support Vector Machine
  - Random Forest
  - Gradient Boosting
- Basic Neural Network
- Deep Neural Network

Initial Assessment of Data

- ➢ Downloaded data has values separated by semicolon. Converted to csv using Microsoft Excel.
- ➢ 70000 observations
- ➢ 11 features

Descriptive stats on the continuous variables
(Notice the range of values)



```
+-------+------------------+------------------+------------------+------------------+------------------+------------------+------------------+
|summary|                id|    (in days)  age|            gender|  (in cm)   height|  (in kg)   weight|             ap_hi|             ap_lo|
+-------+------------------+------------------+------------------+------------------+------------------+------------------+------------------+
|  count|             70000|             70000|             70000|             70000|             70000|             70000|             70000|
|   mean|        49972.4199|19468.865814285713|1.3495714285714286|164.35922857142856| 74.20568999999998| 128.8172857142857| 96.63041428571428|
| stddev|28851.302323172928|2467.2516672413917|0.4768380155828605| 8.210126364538551|14.395756678511473|154.01141945609032|188.47253029639106|
|    min|                 0|             10798|                 1|               100|                10|              -100|               -70|
|    max|             99999|             23713|                 2|                99|              99.9|                99|                99|
+-------+------------------+------------------+------------------+------------------+------------------+------------------+------------------+
```

Data Pre-processing, Exploratory Data Analysis and Data Processing

Data Pre-processing:

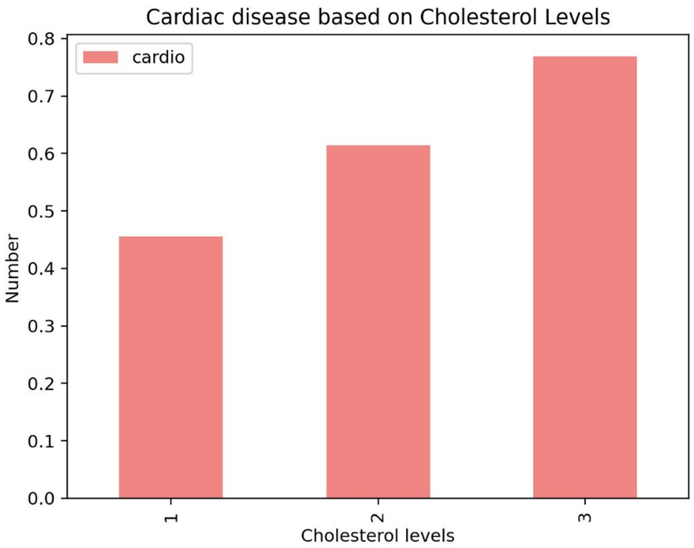
- ➔ 70,000 observations
  - ◆ Few observations have values not observed in human adults (eg. diastolic bp: 11000)
  - ◆ Negative values (eg. systolic bp: -150)
  - ◆ Categorical variables given values (eg. Glucose: 1-normal, 2-above normal, 3-well above normal)
- ➔ Various reasons for above numbers
- ➔ Observations with probable values for human adults will be retained
  - ◆ Height: 135 - 215 cm
  - ◆ Weight: 40 - 200 kg
  - ◆ Systolic bp: 90 - 230
  - ◆ Diastolic bp: 40 - 180
- ➔ Decision will taken with respect to negative numbers during Data Processing. May keep the absolute value but change sign, or may remove the datapoint entirely

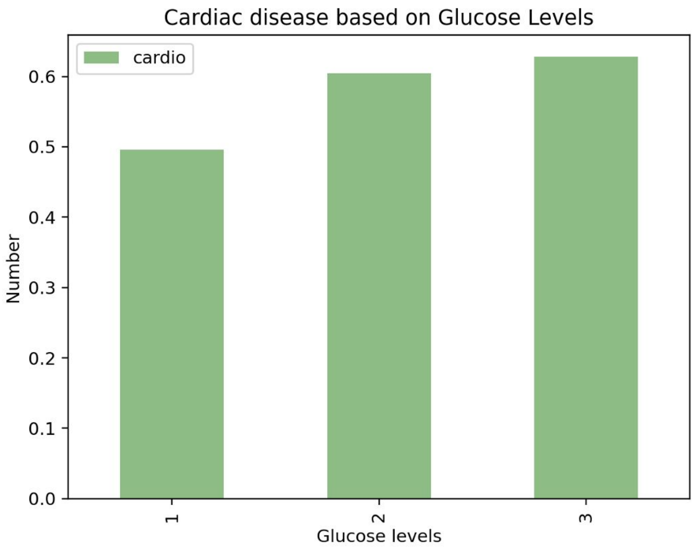
Initial trial of data pre-processing in Excel had brought down the total number of observations to 60,510.

Data Pre-processing, Exploratory Data Analysis and Data Processing

Exploratory Data-Analysis

Performed on the initial trial pre-processed data on Excel


Cardiac disease based on Cholesterol Levels
🟥 cardio
Number
Cholesterol levels


Cardiac disease based on Glucose Levels
🟩 cardio
Number
Glucose levels