The cause and factors of cardiovascular disease

Team Members - Divya, Anthony Elbers, Shebo, Steve, Brett

## Project Description / Outline

To discover the relationship and trends of cardiovascular disease with the different variables such as: age, smoking status, glucose level, physical activity, medical history, genetics, obesity.

focus on global audience.

## Research question to answer

* Risk factors for cardiovascular disease
* What are the correlations between the variables and cardiovascular disease?
* What is the percentage between gender and cardiovascular disease?

## Datasets to be used

<https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>

<https://www.kaggle.com/aiaiaidavid/cardio-data-dv13032020>

Things to do:

collect data relating to income datasets of cardiovascular disease

## Rough breakdown of task

1. Collect data
2. Email corresponding parties for the collection of data
3. Read the recent articles and complications relating the topic
4. Clean data
5. Discover trends and relationships
6. Visualisations, correlation regressions etc.

## Validations

Observations – Males are more likely to have cardiovascular disease than female

Data descriptions

* Input features

1. Objective; factual information
2. Examination; results of medical examination
3. Subjective; information given by the patient

Smokers more likely to get cardiovascular

Age

Cholesterol

Diabetes

Diet & Nutrition

Obesity

Physical activity

**Data indicators (heart disease)**

Gender – 1 = female, 2 = male

CHOLESTEROL – 1 = normal, 2 = above normal, 3 = high

Glucose **-** 1 = normal, 2 = above normal, 3 = high

* 1- Normal - below 140  
  2 -Above normal -140-199  
  3- High ->200

Smoke – 0 = non-smoker, 1 = smoker

Alcohol – 0 = non-drinker, 1 = drinker

Physical activity – 0 = no exercise, 1 = does exercise

Cardio disease – 0 = no cardio related disease, 1 = evidence of cardio disease

BMI to be calculated for this data

AP – is equal to blood pressure

**Data indicators (framingham )**

Gender – 0 = female, 1 = male

CHOLESTEROL – below 220 is normal

Glucose **-** 0 = no, 1 = yes (mus)

Smoker – 0 = non-smoker, 1 = smoker

Alcohol – 0 = non-drinker, 1 = drinker

Physical activity – 0 = no exercise, 1 = does exercise

Cardio disease – 0 = no cardio related disease, 1 = evidence of cardio disease

Bp – 0 = no blood pressure issues, 1 = blood pressure related issues

**What to do, Data Cleaning**

1. **Stages of analysis**

* Identifying data source,
* Get the data
* Familiarise data set

1. **Data cleaning**

* Calculate the BMI for heart disease
* Add a column for blood pressure, distinguish High Blood Pressure …
* Calculate and categorise glucose level (1,2,3)

1. **Correlations and Observation**

* Comparisons between gender and heart disease - Brett

Bar chart, histogram

* Comparisons between age and heart disease - Steve

Need to be binned: bar and histogram

* Comparisons between bmi and heart disease - Shebo

Scatter with correlation and regression, line graph

* Smoking vs heart disease - Divya

Bar Chart, line graph connected

* Alcohol vs heart disease - Anthony

Bar Chart, line graph connected

* Cholesterol vs heart disease - Shebo

Scatter with correlation and regression, line graph

* Physical activity vs heart disease - Anthony

Bar Chart, line graph connected

* Blood pressure and heart disease -Brett

Scatter with correlation and regression, line graph

Stages of analysis

Visualisations (6-8)

* Box plot for outliers
* Bar chart for gender
* Pie chart for gender

Presentation-

Intro – shebo and also providing plotting and creating some graphs