Project 4

Heart attack prediction

Group 3 - Bich, Jun, Oormi, Sujatha

Agenda

Introduction
Demo
Front End in-depth
Back End in-depth
Insights and future improvements

Project Proposal & Aim

Brief on project proposal, aim and importance

Proposal & Importance

The objective is to look at biological and circumstantial health indicators and analyse features such as age, BMI, family history, heart rate etc. to build a ML model that can predict the risk of heart attack, based on user input. We will also visualise the spread of heart attacks around the world. We will use a dataset from Kaggle.

Medical name: Myocardial infarction

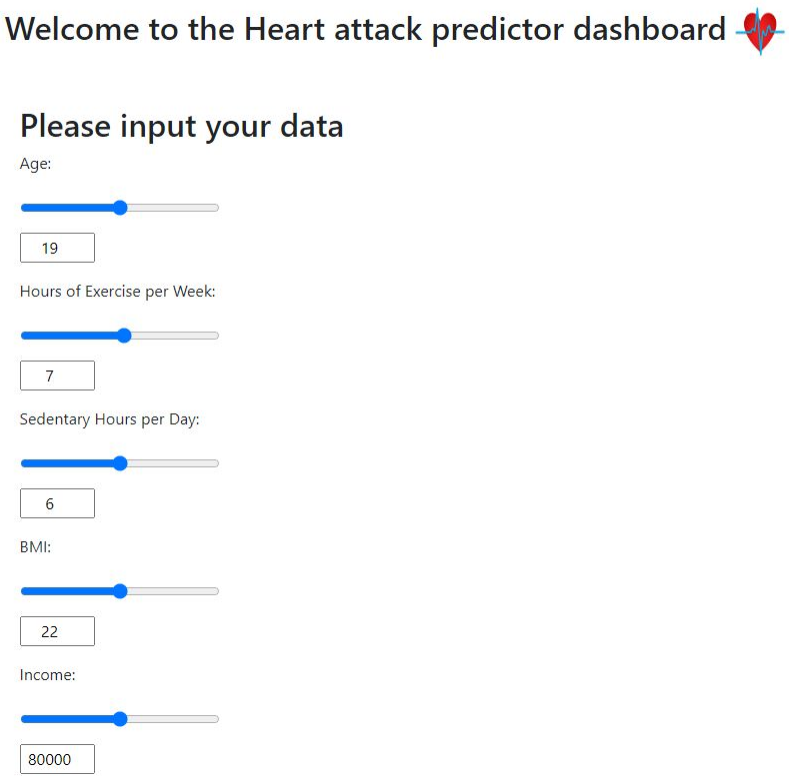
- One of the leading causes of death worldwide.
- Each year more than 19 millions people die from heart attack (about the population of all major Australian cities combined).
- Age doesn't discriminate, while it is more common in older people, young people are not immune.

Dataset

- 25 features, 8763 rows.
- Features include
  - Age,
  - Heart rate,
  - Blood pressure
  - Cholesterol,
  - Family history,
  - Diabetes,
  - Family History,
  - Hours of exercise per week,
  - Sedentary hours per day,
  - BMI,
  - Income etc.

- Generated by chatGPT

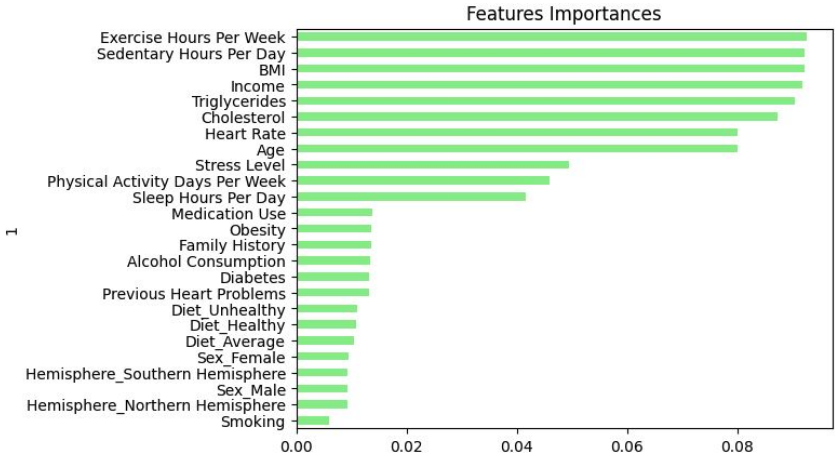
Dashboard Demo

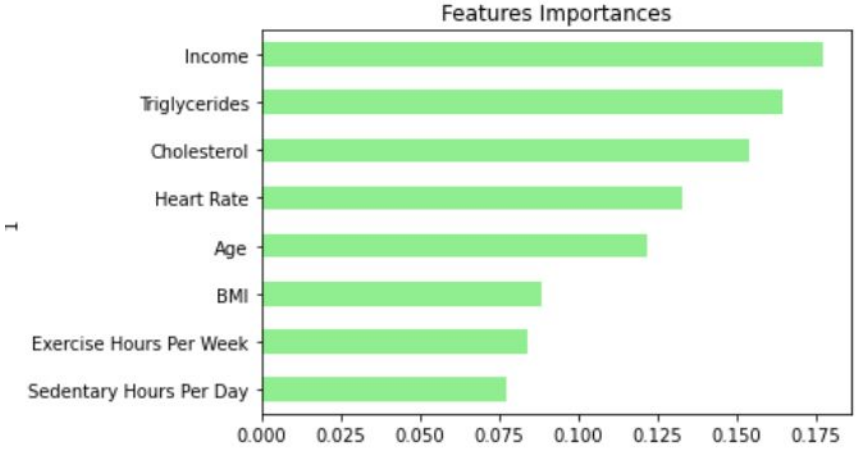


Project Development

Diving into the development of the front and back ends

Data Exploration & Data Delivery - Front End Mechanics





Designing Tool and Library:

HTML,CSS,JS and Bootstrap (4.5.2)

Framework: Flask (Python framework)

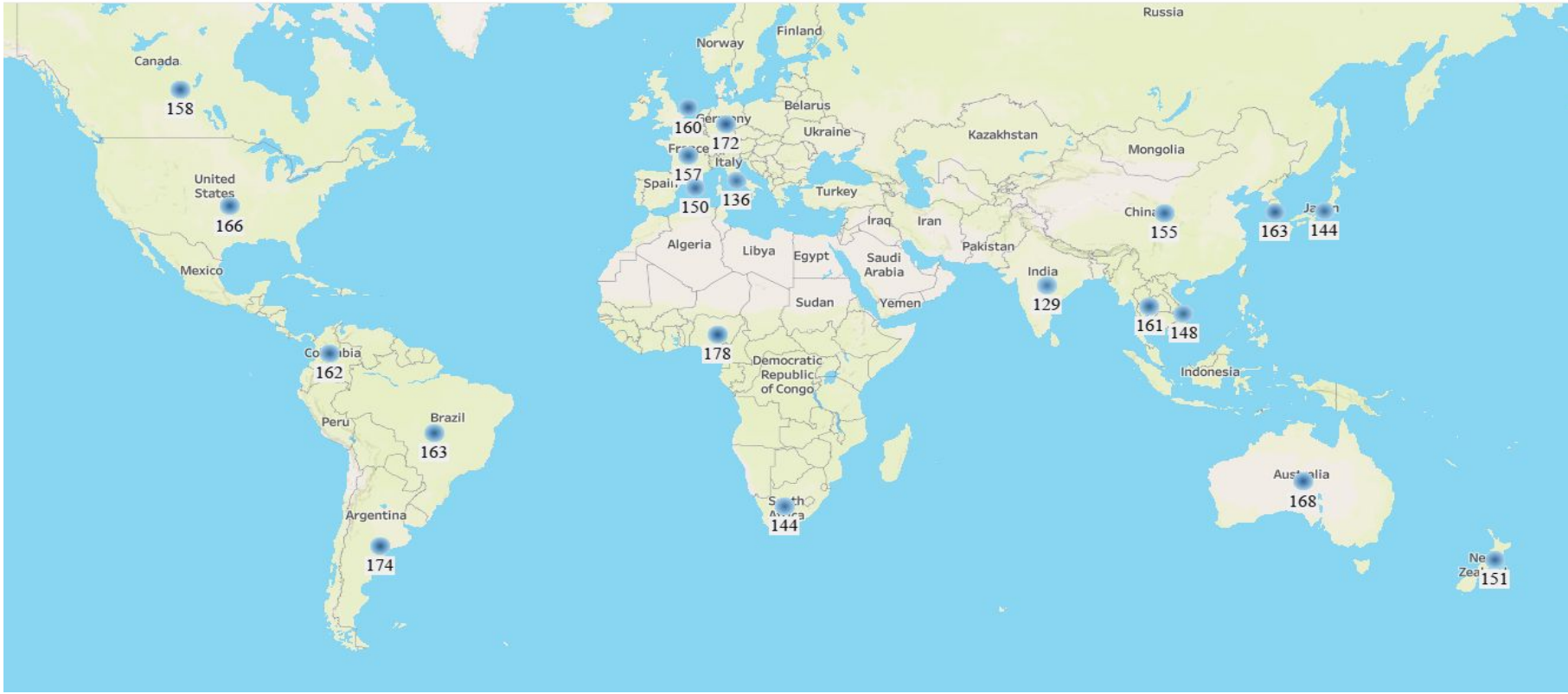
Web hosting: Python Anywhere

Dashboard Visualisation:

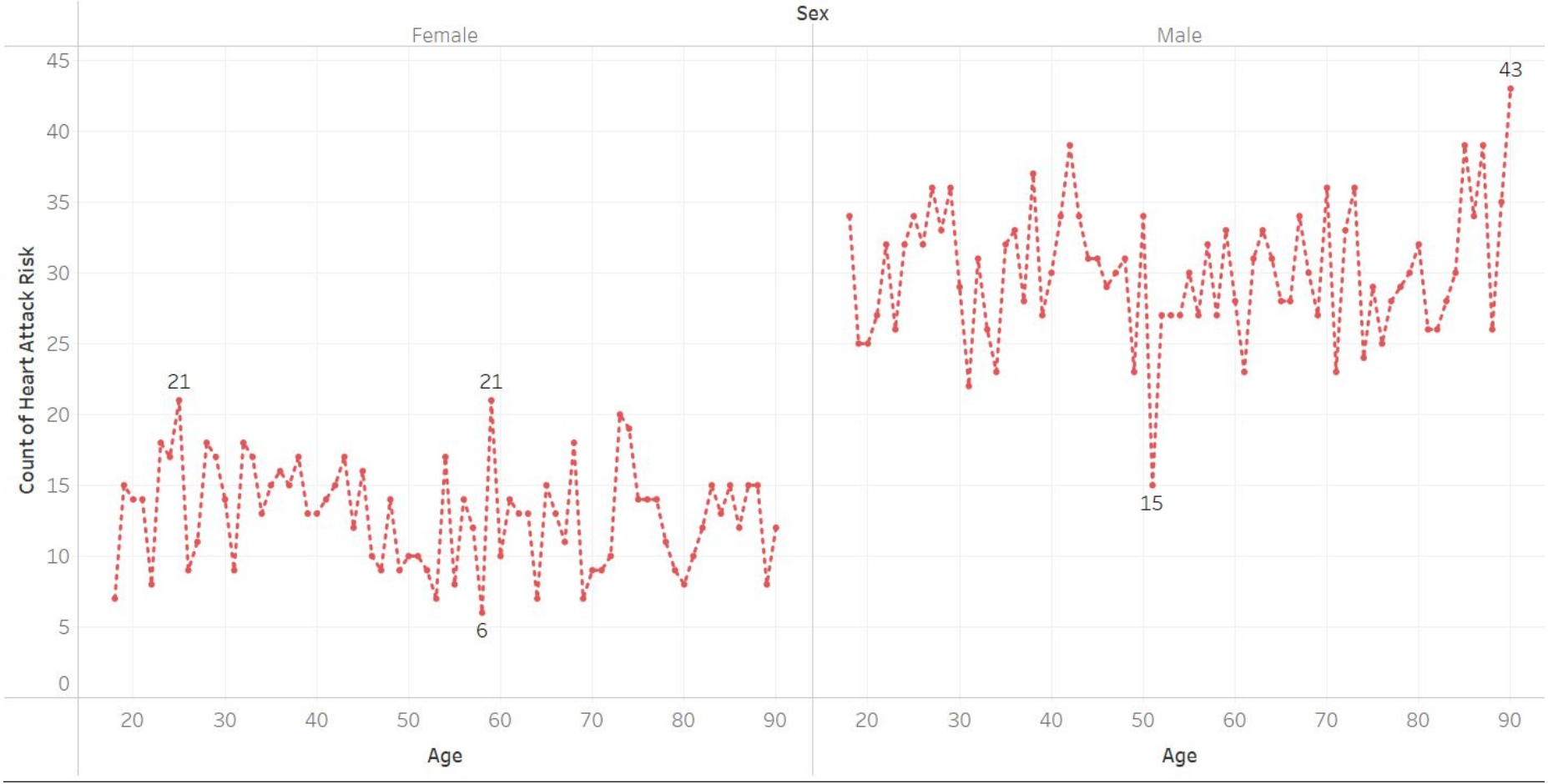
Tableau, Matplotlib & Seaborn

Visualtisation from the Dashboard Demonstration

Global Map View

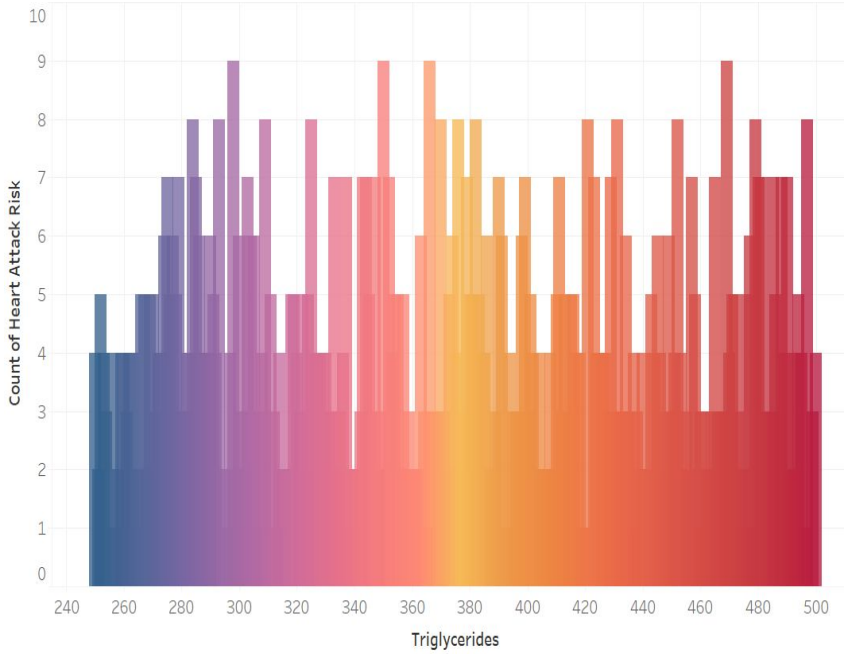


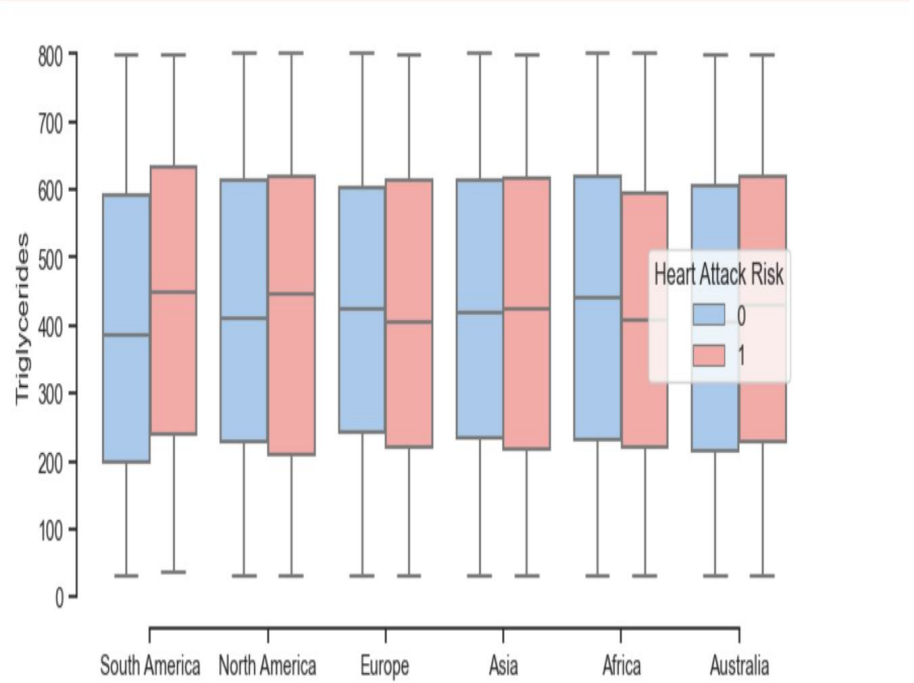
Gender View


Sex
Female
Male
Count of Heart Attack Risk
45
40
35
30
25
20
15
10
5
0
21
21
6
43
15
20
30
40
50
60
70
80
90
Age
Age

Triglyceride Trend

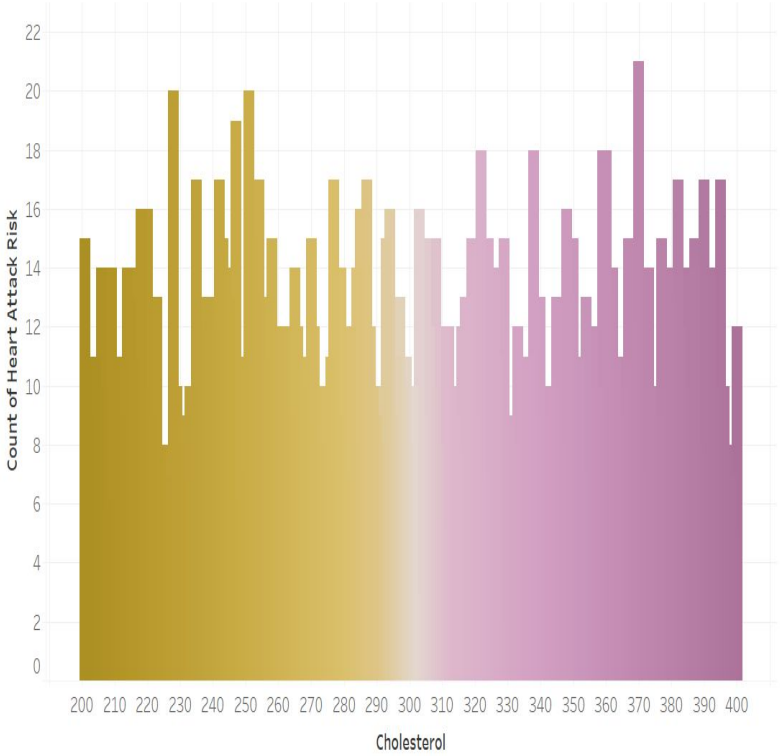
Healthy level: Below 150 Borderline: 150 to 199 High: 200 to 499

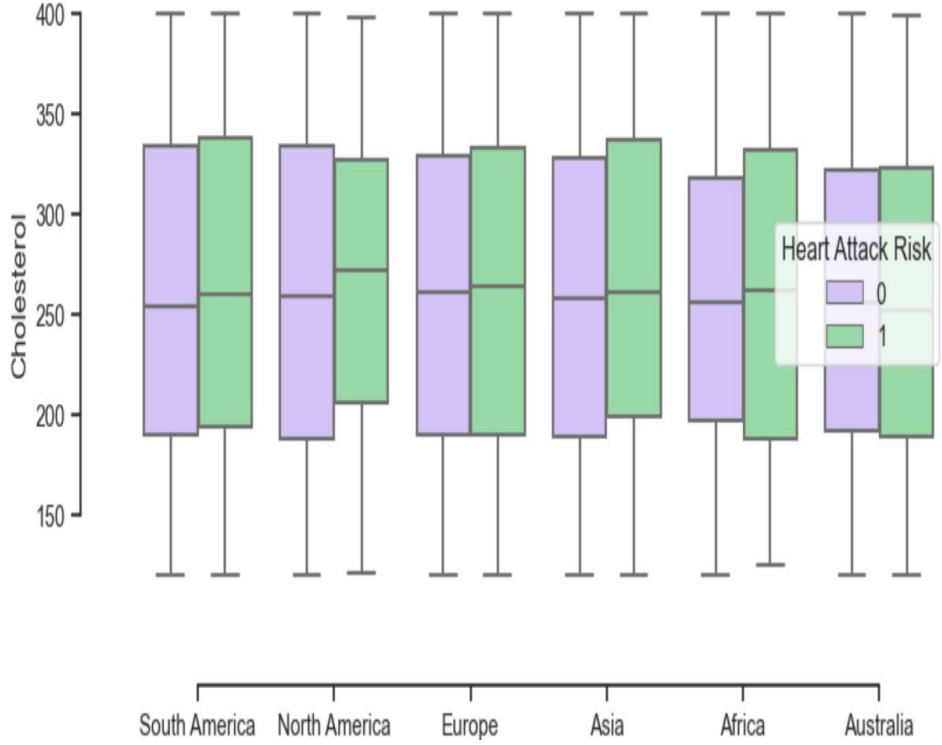




Cholesterol Trend

Healthy level: Total Below 200





Extract, Transform, Load - Back End Mechanics

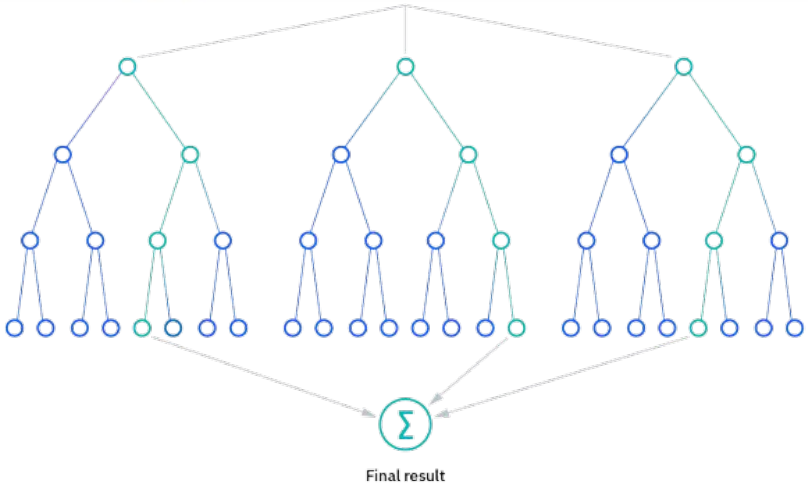
	Age	Cholesterol	Heart Rate	Exercise Hours Per Week	Sedentary Hours Per Day	Income	BMI	Triglycerides
0	67	208	72	4.168189	6.615001	261404	31.251233	286
1	21	389	98	1.813242	4.963459	285768	27.194973	235
2	21	324	72	2.078353	9.463426	235282	28.176571	587
3	84	383	73	9.828130	7.648981	125640	36.464704	378
4	66	318	93	5.804299	1.514821	160555	21.809144	231

Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	1013	112
Actual 1	543	85

Confusion matrix Accuracy Score : 0.6263548203080433
Classification Report

	precision	recall	f1-score	support
0	0.65	0.90	0.76	1125
1	0.43	0.14	0.21	628
accuracy			0.63	1753
macro avg	0.54	0.52	0.48	1753
weighted avg	0.57	0.63	0.56	1753

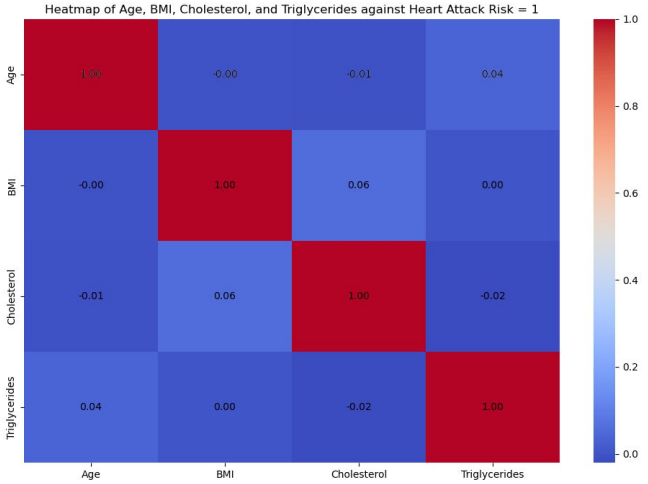


Insights & Summary

Insights into challenges, limitations and conclusion

Challenges

- Data delivery:
  - Static visuals
  - Unable to bring the strong correlation in Visualisation
- ETL
  - Noisy data and weak correlations - accuracy impacted
  - High false negatives
- What did we learn
  - Data delivery; Bootstrap, CSS framework, JS, user experience, API recall (Flask), security
  - Traditional ML; Linear regression, random forest, xgb boost
  - Hyperparameter tuning, features engineering
  - Artificial neural networks was not the best application for our dataset

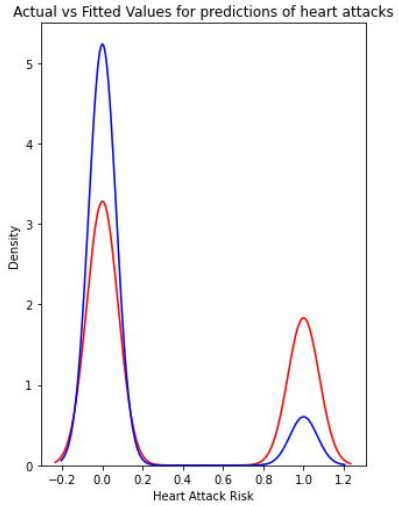


Hypothesis testing:

Actual	Decision: H_0 true (Fail to reject)	Decision: H_0 false (Rejecting H_0)
H_0 true	TRUE NEGATIVE Correct decision: Confidence level (prob $1-\alpha$)	FALSE POSITIVE Type I Error: Significance level/Size (α) (prob α)
H_0 false	FALSE NEGATIVE Type II Error: fail to reject (prob β)	TRUE POSITIVE Correct decision: Power (prob $1-\beta$)

Limitations & Further Exploration

- Limitations
  - The research uses a chatPGPT dataset from Kaggle
  - Did we build a model that can accurately predict heart attack risk?
    - Accuracy
    - Precision
    - Recall
    - F1 scores



Further exploration -

- Expand and explore real /other heart health features and explore dataset correlation
- Training on real datasets and improve correlations, accuracy of prediction
- Incorporation of real time predictions and visualisations to increase education
- Observations of differences / trends in populations and Anthropology

Summary

This project has important applications to real world heart health research, innovation and education with consideration to patient specific information, lifestyle choices and socioeconomic factors.

Society Benefits	Personal Benefits
Prevention of Cardiovascular Diseases (CVDs)	Prevention of Cardiovascular Diseases (CVDs)
Reduced Healthcare Costs (society)	Reduced Healthcare Costs (individual)
Improved Quality of Life	Improved Quality of Life
Public Health Initiatives	Improved Quality of Life
Increased Awareness and Education	Personal Empowerment
Research and Innovation	Better Emotional Well-being
Addressing Health Disparities	Personalized Health Management
Supporting Personalised Medicine (benefits both society and individuals)	Enhanced Productivity and Performance
	Improved Sleep
	Positive Impact on Family and Community