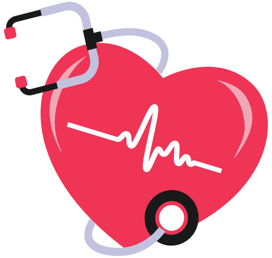
National Heart Centre Singapore
SingHealth

Project HeartBeat
Heart Disease Prevention

By: S01 Group 5

Lum Shi Zhen	U2210198H
Teo Man Ru Joleen	U2111561F
Poh Lee Tin	U2111981A
Sally Ngui Yu Ying	U2222782A
Tenia Xu Yuan	U2210808E



Shi Zhen

Table of **Contents**

Shi Zhen

01

Objectives

Shi Zhen

Significance of Cardiovascular Disease

18 million
global deaths

1 in 3 deaths

23 lives lost daily

Shi Zhen

Barriers to CVD treatment

Individual + Healthcare		Effects on CVD outcomes
Perceived high check up costs	→	Missed opportunities to prevent CVD
Fear of receiving a diagnosis		Increased treatment costs
Lengthy waiting times for diagnostic tests		Worse health outcomes

Shi Zhen

Business Problem & Opportunity

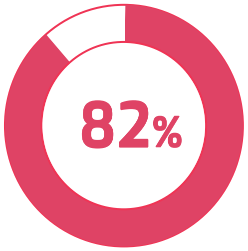
Lack of strategies for early detection of CVD

Opportunity

Create comprehensive predictive model to detect CVD

By utilizing analytical tools,

successful forecasts of chronic illnesses can reach up to



accuracy rate

National Heart Centre Singapore
SingHealth

Business Problem & Opportunity

Shi Zhen

Leverage on Data Analytics

Gain insights on patient demographics, risk factors that may indicate risk of CVD

Facilitate targeted preventive interventions

Enhance operational efficiency and cost-effectiveness for healthcare and individuals

Shi Zhen

Project HeartBeat

Objectives

- Predictive tool that relies on non-medical data
- Free & easily accessible for everyone
- Empowers individuals to be proactive towards their cardiovascular health

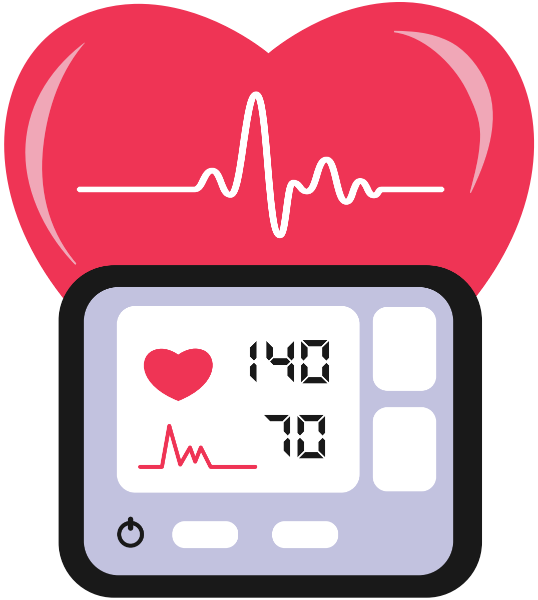
Benefits

- No need to wait and queue in hospitals
- Removes financial barriers
- Reduce overall CVD mortality rates through early prevention

Joleen

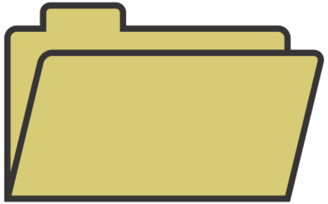
02

Data Sourcing & Exploration



Joleen

Data Sourcing



2022 Behavioural Risk Factor Surveillance System Survey Data (CDC)

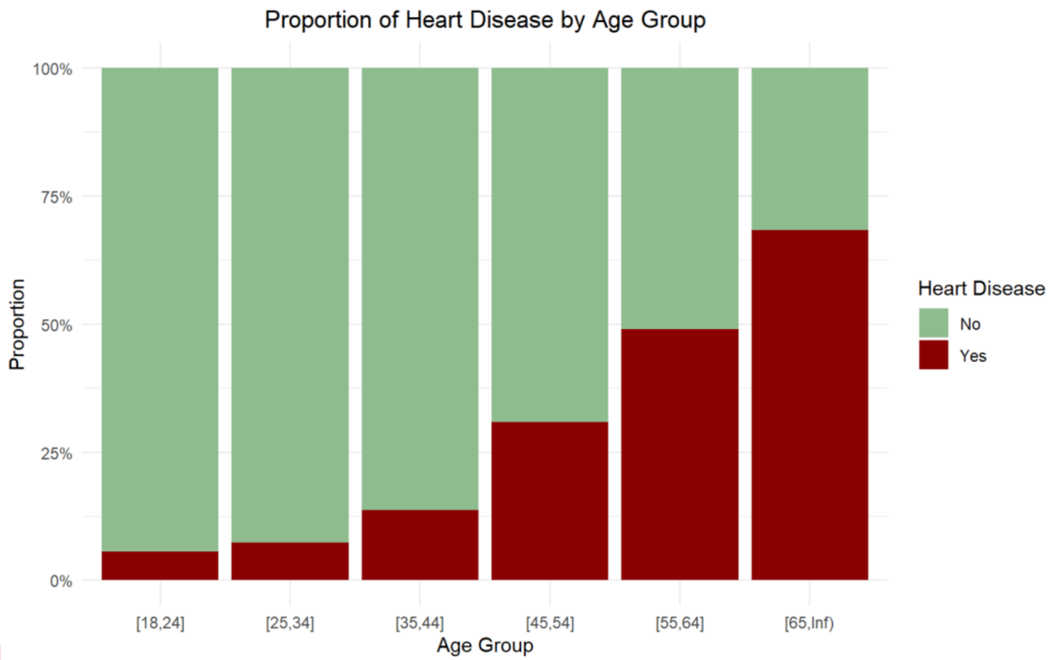
Categorized into

Factors related to CVD

1. Personal Particulars
2. Physical Characteristics
3. General Health Status
4. Health Habits & Behaviours
5. Health Issues & Illness History

Joleen

Heart Disease by Personal Factors

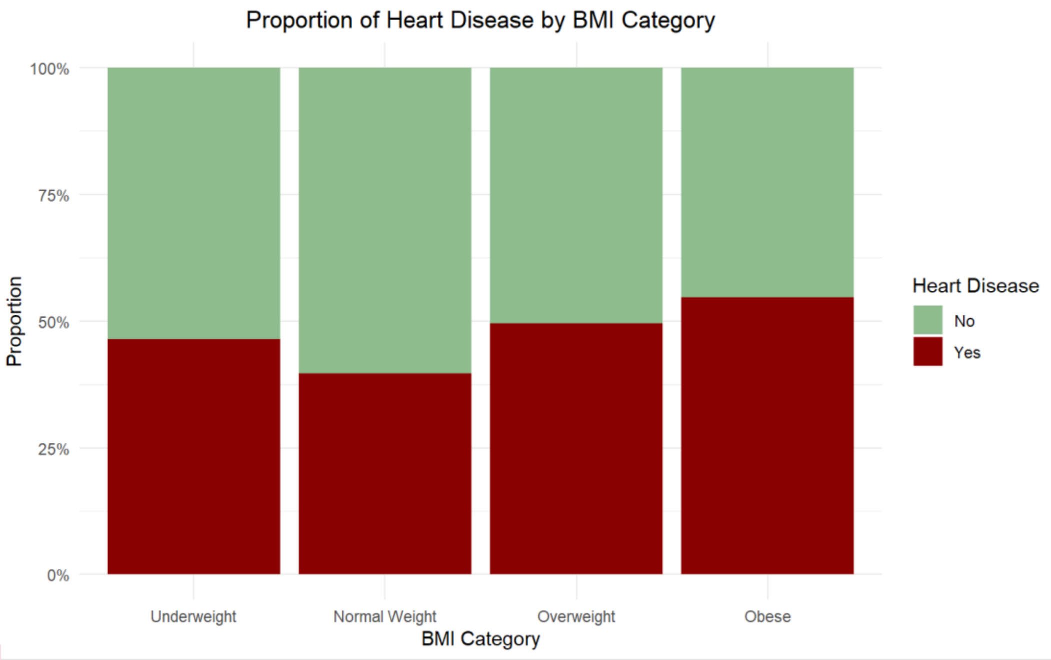


Data Exploration

- Age is a significant risk factor that compromises on cardiovascular system
- Myriad physiological changes such as increased oxidative stress

Joleen

Heart Disease by Physical Characteristics

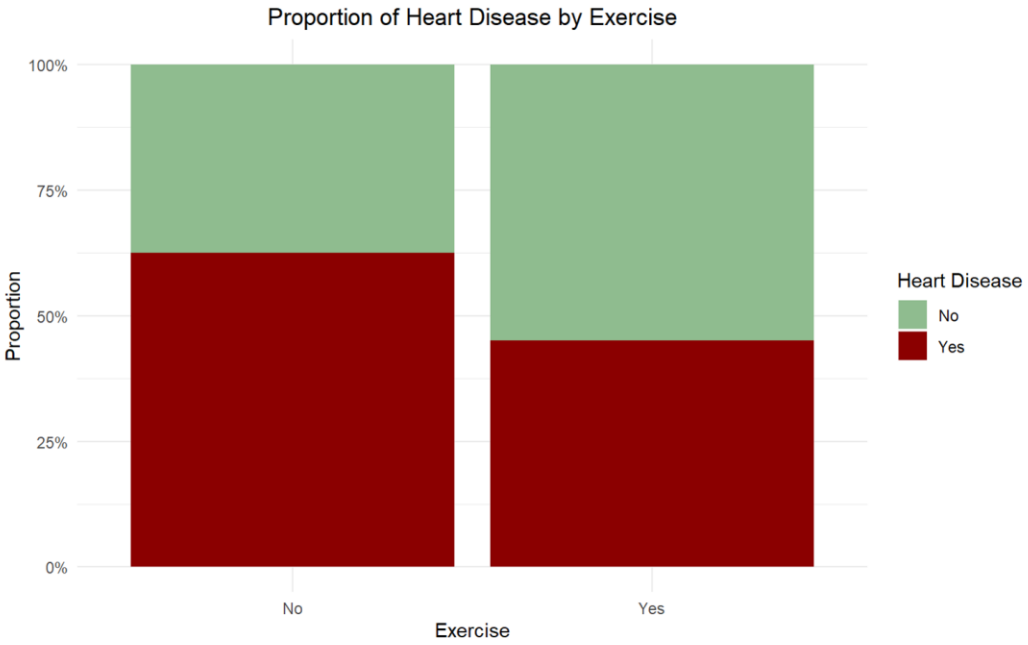


Data Exploration

- Obesity contributes to various physiological mechanisms
- Haemoglobin deficiency is more prevalent among underweight individuals

Joleen

Heart Disease by Health Habits & Behaviours

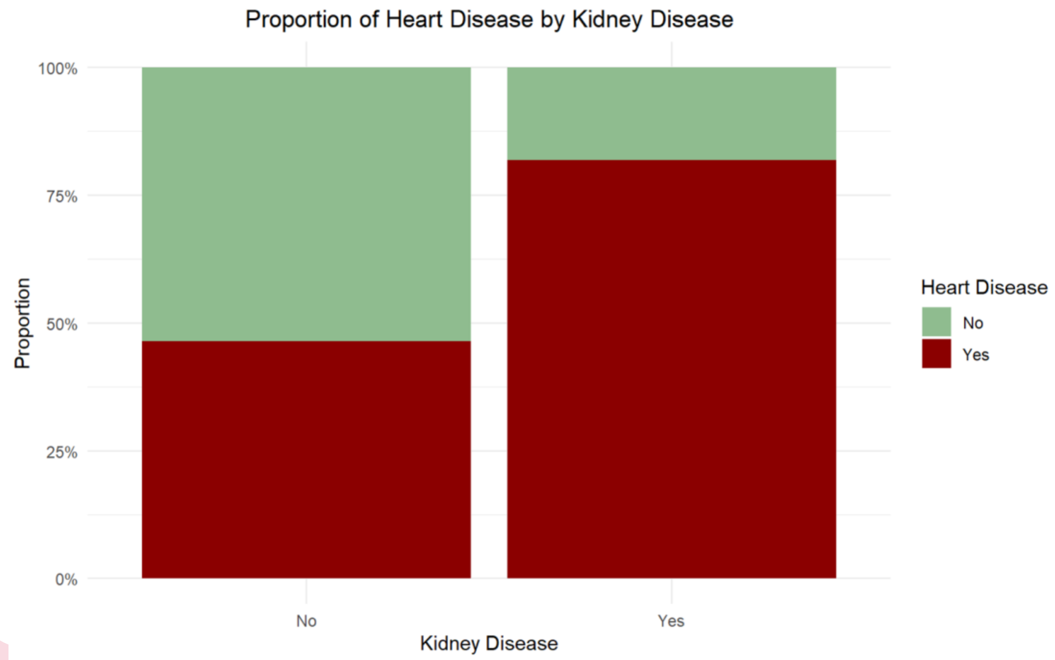


Data Exploration

- Physical inactivity can lead to other risk factors of heart disease such as diabetes, obesity etc.

Joleen

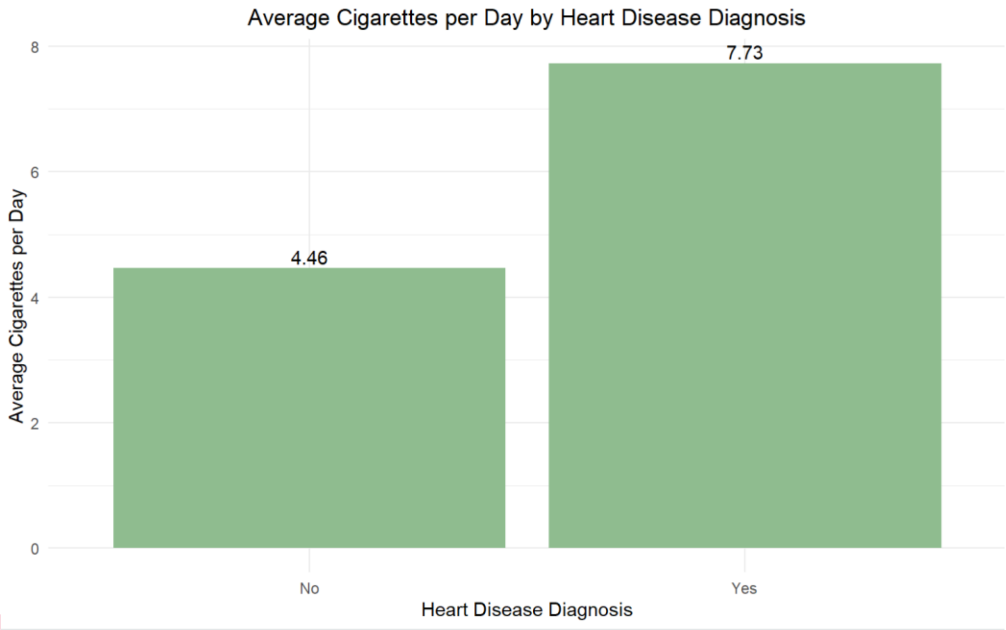
Heart Disease by Heart Issues & Illness History



Data Exploration

- Strain placed on the heart due to kidney dysfunction
- More effort required to circulate blood to kidney
- Presence of high blood pressure exacerbates the strain

Average Cigarettes Per Day

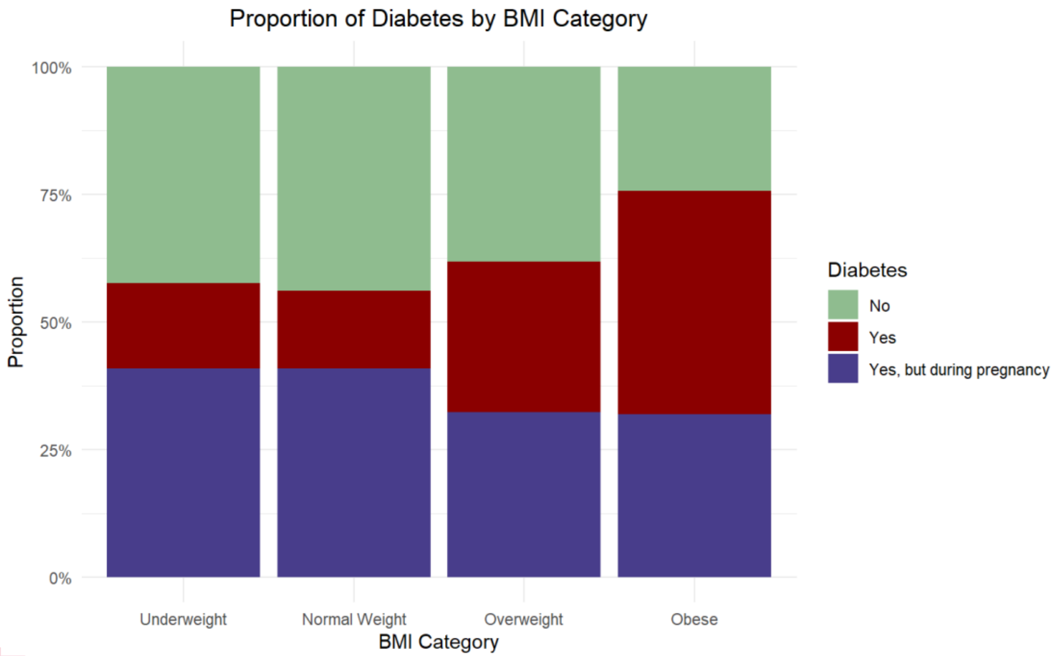


Data Exploration

- Compounds present in cigarette promote the accumulation of plaque in blood vessels
- Elevate heart rate and induce inflammation

Joleen

Diabetes by BMI Category

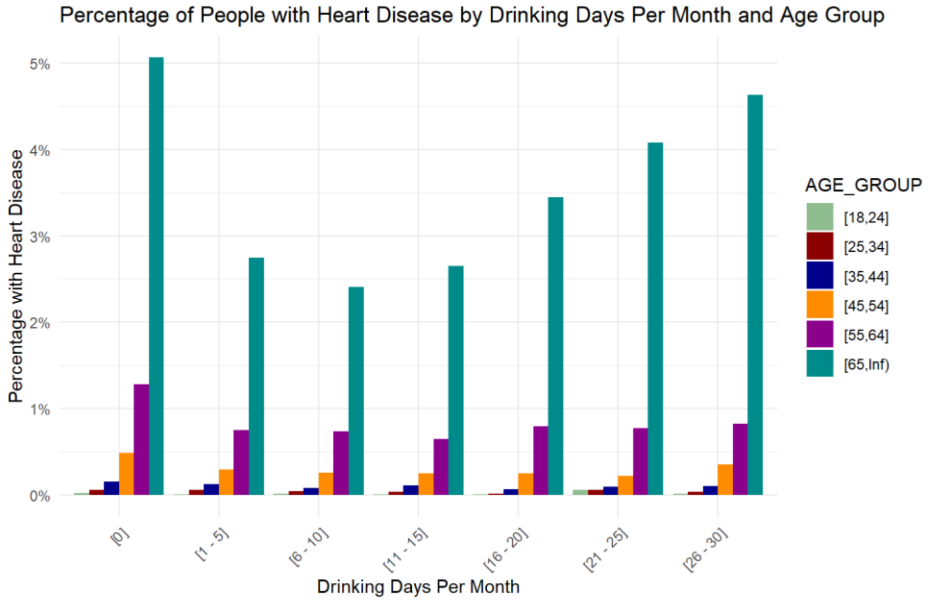


Data Exploration

- Excessive body fat accumulation is a key contributor to type 2 diabetes

Joleen

Heart Disease by Drinking Days Per Month and Age Group



Data Exploration

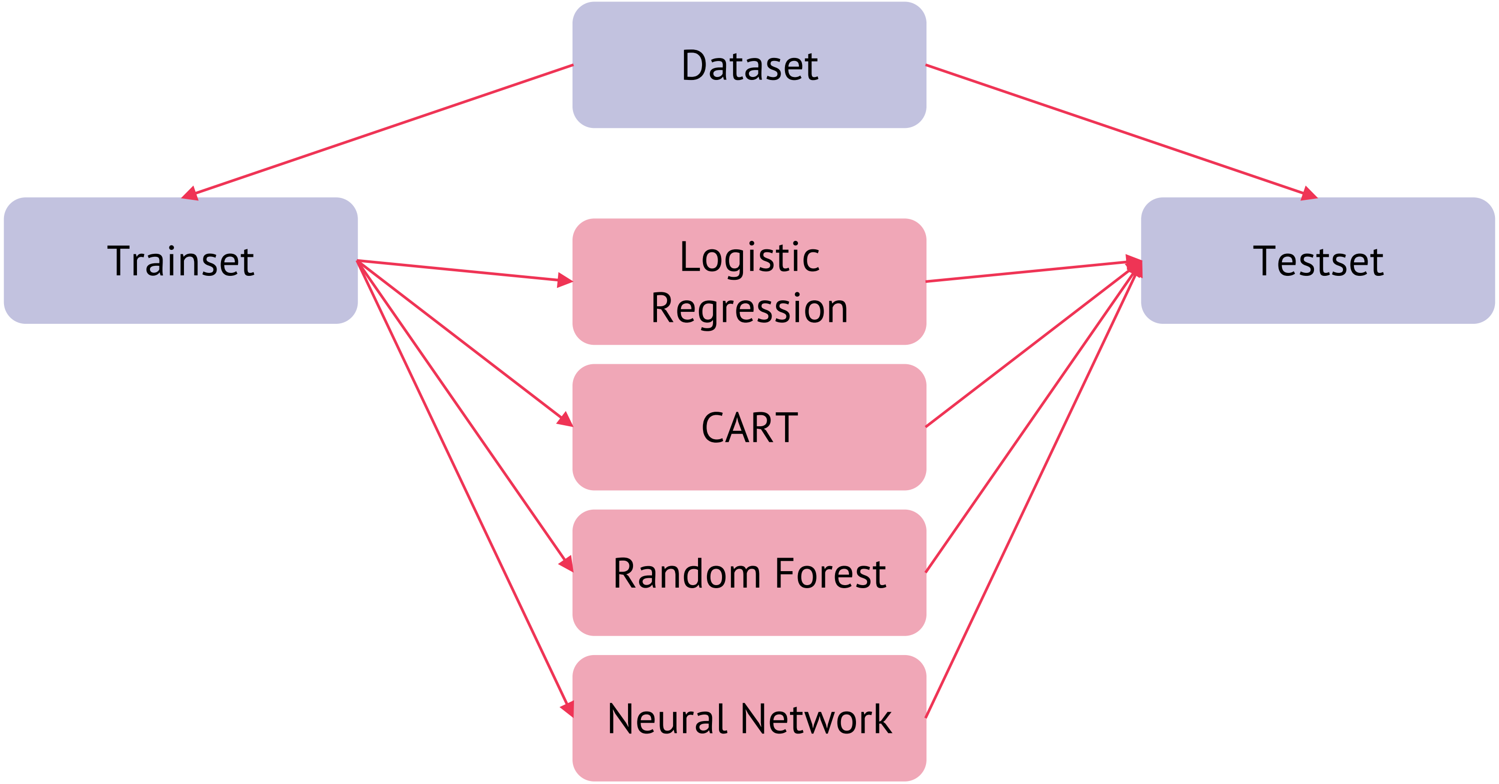
- Excessive alcohol consumption can increase risk of high blood pressure
- Dependent on the frequency and amount of alcohol consumed

03

MODELS

Models Developed

Lee Tin



Logistic Regression

Lee Tin

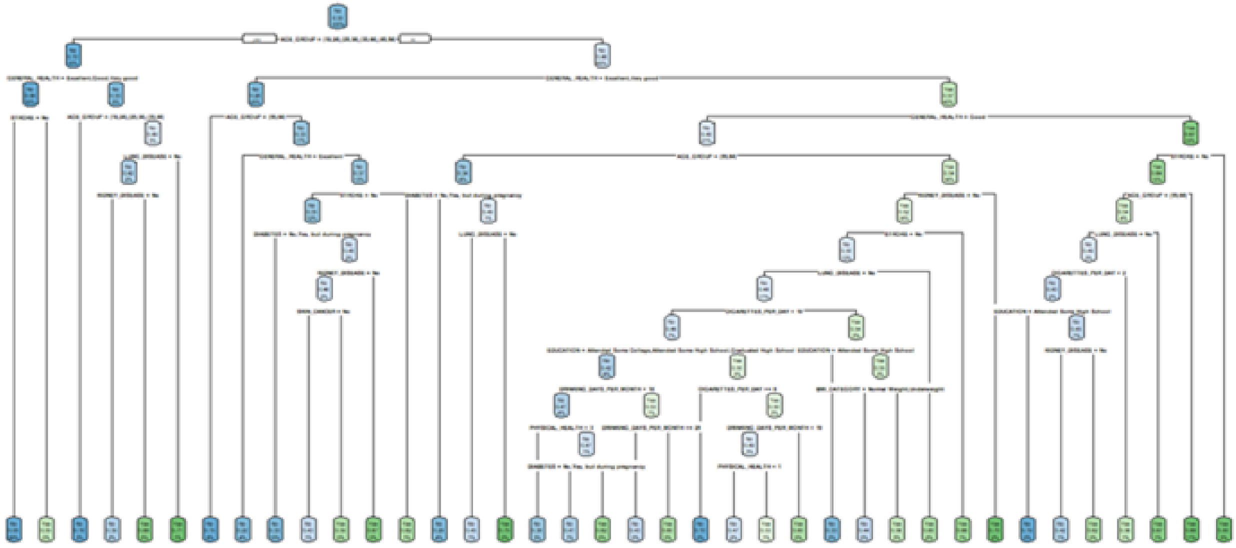
```
Call:
glm(formula = HEART_DISEASE ~ AGE_GROUP + EDUCATION + UNABLE_TO_AFFORD_MED +
    BMI_CATEGORY + GENERAL_HEALTH + PHYSICAL_HEALTH + CIGARETTES_PER_DAY +
    STROKE + SKIN_CANCER + OTHER_CANCER + LUNG_DISEASE + KIDNEY_DISEASE +
    ARTHRITIS + DIABETES, family = binomial, data = trainData,
    na.action = na.omit)
```

- Model built based on insights derived from EDA
- Backward stepwise selection was employed
- Checked for multicollinearity
- Variables with high p-values removed, due to lack of statistically significant impact on the target variable, resulting in the optimal logistic regression model

CART - Classification Tree

Lee Tin

Optimal Prune

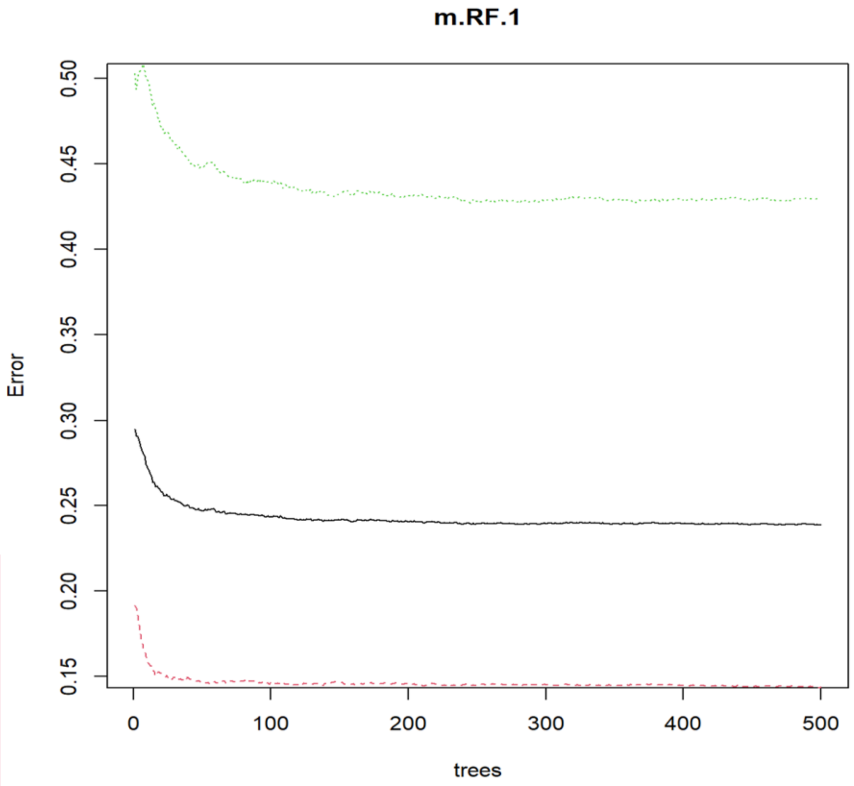


- Optimal complexity parameter determined based on <u>cross-validation</u> results
- Tree is <u>pruned</u> using this optimal CP of 0.00118497 to <u>38 terminal nodes</u>

Random Forest

Lee Tin

OOB Error against Number of Trees

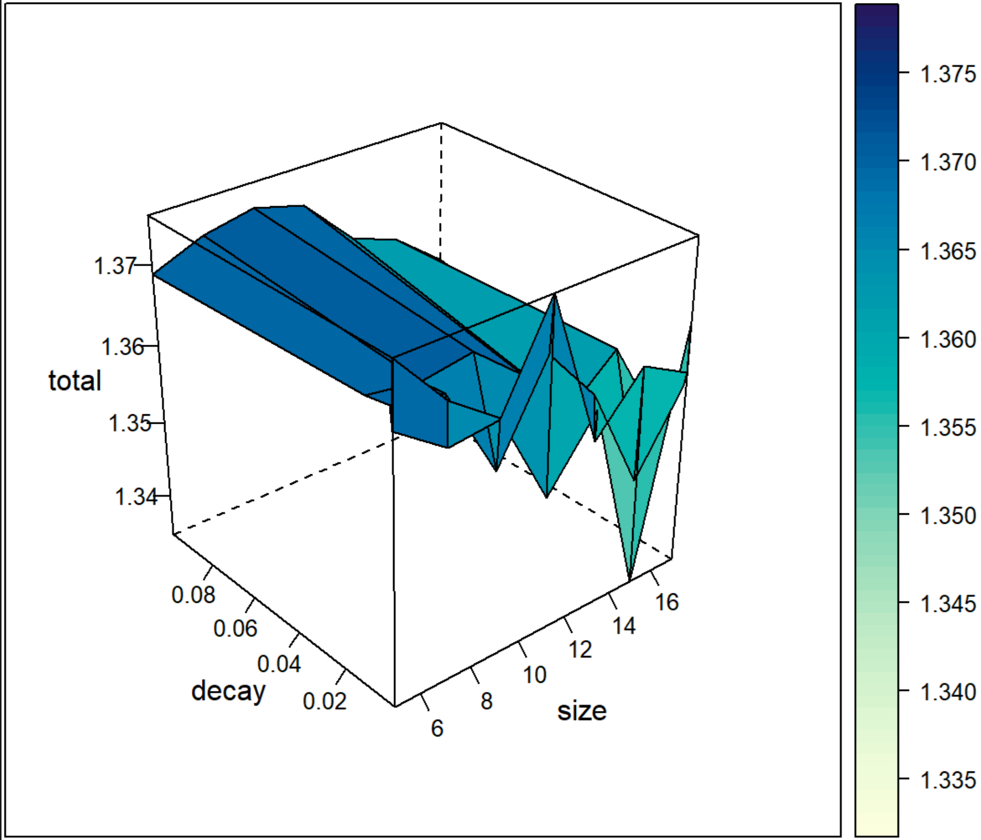


Model Evaluation

- Optimal model determined by minimising and stabilising the Out-of-bag (OOB) error
- Stabilized after 100 to 150 trees
- OOB error 23.92%

Neural Network

Lee Tin



Model Evaluation

- Through grid search, the <u>optimal size of 17 nodes</u> and <u>weight decay of 1e-05</u>, gives the <u>best model accuracy</u> and the <u>optimal Neural Network model</u>

Confusion Matrix Results

Lee Tin

Logistic Regression		Predicted	
		No	Yes
Actual	No	TN 8345	FP 2837
	Yes	FN 1319	TP 4282

CART		Predicted	
		No	Yes
Actual	No	TN 9524	FP 1678
	Yes	FN 2261	TP 3340

Random Forest		Predicted	
		No	Yes
Actual	No	TN 22254	FP 3882
	Yes	FN 5742	TP 7326

Neural Network		Predicted	
		No	Yes
Actual	No	TN 10823	FP 379
	Yes	FN 4168	TP 1433

Model Employed

Lee Tin

Models	TPR (%)	FNR (%)	FPR (%)	TNR (%)	Accuracy (%)
Logistic Regression	76.5	23.5	25.3	74.7	75.3
CART	59.6	40.4	15.0	85.0	76.6
Random Forest	56.1	43.9	14.9	85.1	75.5
Neural Network	25.6	74.4	3.4	96.6	72.9

Final Decision

- Logistic regression model has been selected as the most optimal model
- All 4 models have similar accuracy
- Logistic regression has the highest TPR of 76.5%

Business Insights from Models

Lee Tin

VARIABLE IMPORTANCE

Variable	Importance
AGE_GROUP [above 65]	18.815
GENERAL HEALTH [fair & poor]	26.030 25.663
STROKE	18.665

Insights

- These variables appear to have notable impacts on the likelihood of developing CVD
- Understanding the importance of these variables can enable more targeted and effective preventive strategies

Business Insights from Models

Lee Tin

ODDS RATIO FOR LOGISTIC REGRESSION

Variable	Odds Ratio
AGE_GROUP [above 65]	14.729
GENERAL HEALTH [poor]	7.600
STROKE	2.453

Insights

<u>Higher odds ratio</u>:

1) in AGE_GROUP for individual aged above 65
2) poor GENERAL_HEALTH
3) for individual with stroke

signifies a <u>heightened risk of CVD</u>

Prioritize interventions aimed at improving these variables can help with <u>resource allocation</u> and <u>reduce the risk of CVD</u>, ultimately <u>improve patient outcomes</u> and <u>cost efficiency</u>

Sally

04

NHCS’ Current Measures

Sally

Current Reactive Measures

Cardiac Tests — Supplemented with → **AI Tool**

- Echocardiogram (ECG),
- Cardiac Computed Tomography (CT) Scan,
- Exercise Stress Test (Treadmill Exercise) and more

- Gabor-Convolutional Neural Network (Gabor-CNN) algorithm
- Recognise ECG patterns to diagnose CVD
- 98.5% Accuracy

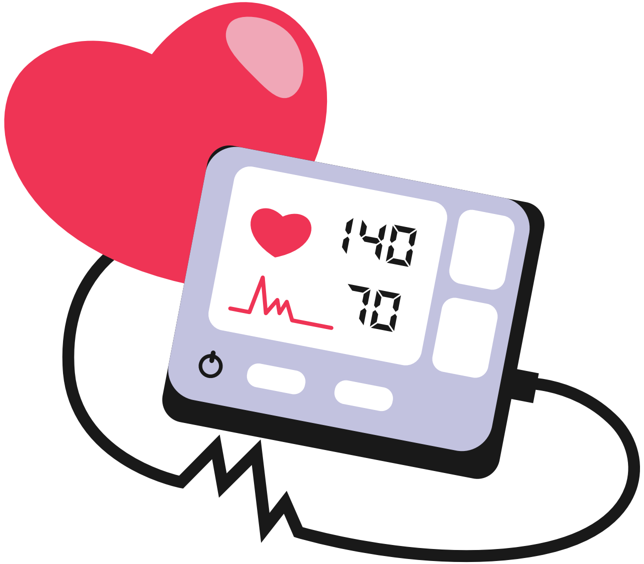
Sally

Limitations

Limited Improvement in Efficiency and Cost	• AI tool still requires ECG to be conducted • **Costly and time-consuming**
Lack of Early Prevention Measures	• Current focus on **reactive measures** • Potential delayed diagnosis for those at risk **only when they show CVD symptoms**
Limited Predictive Variables	• AI tool only analyses ECG signals • **Non-medical factors** also play a part in determining risk

05

Proposed Solution



Sally

Proposed Solution

Current Measures

Reactive

Supplemented with

Our Solution

Predictive

Preventive

Cardiac Health Monitoring and Prediction (CHAMP)

Sally

Overview of CHAMP App: 3Ps

Predictive Analytics Models

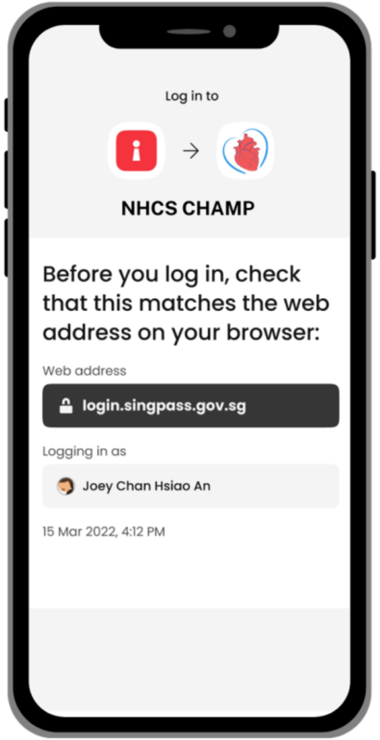
Personalised Preventive Measures

Public Education

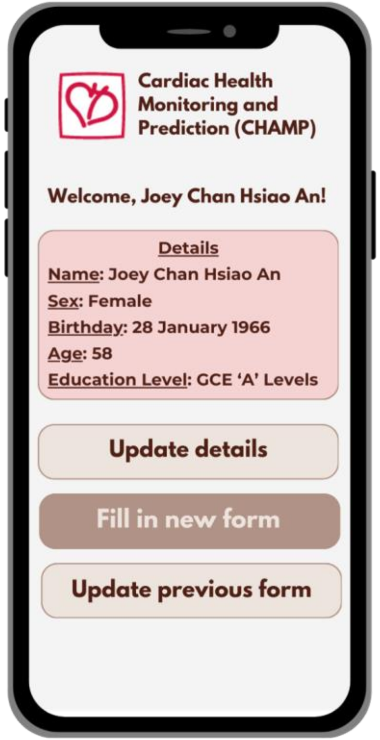
Predictive Feature 1: Retrieval of Input Factors

Sally

Biometric Login



Automatic Retrieval of Details & Form History

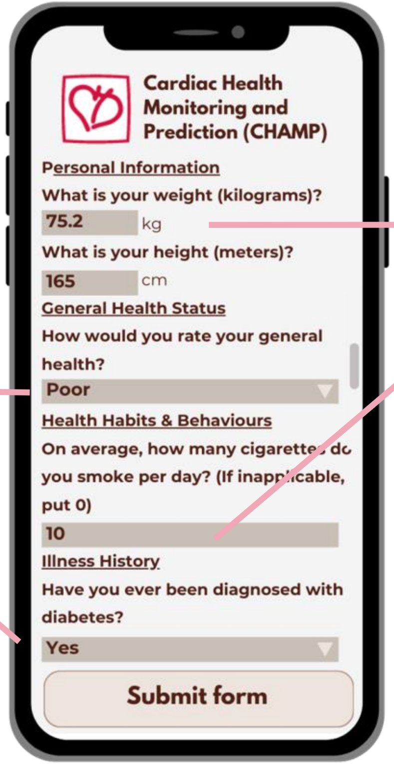


Predictive Feature 2: Risk Assessment Form

Sally

Categorical Variables

Select category from dropdown

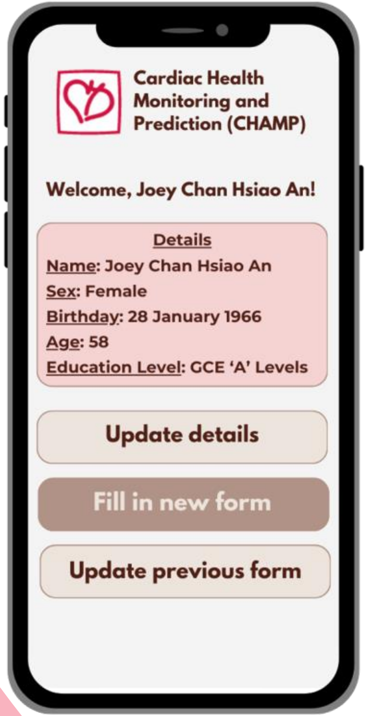


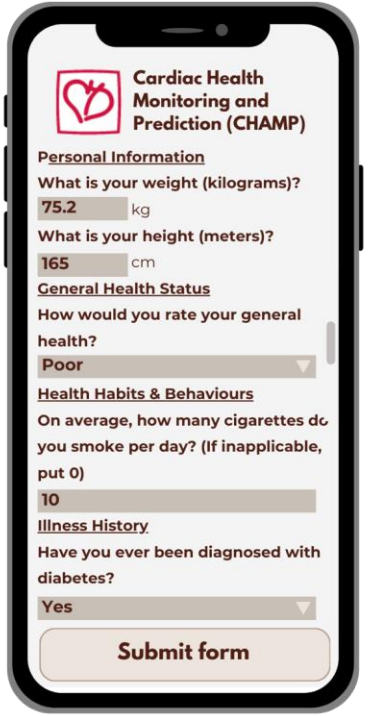
Continuous Variables

Fill in numerical value

Predictive Feature 3: Utilisation of Logistic Regression Model

Sally





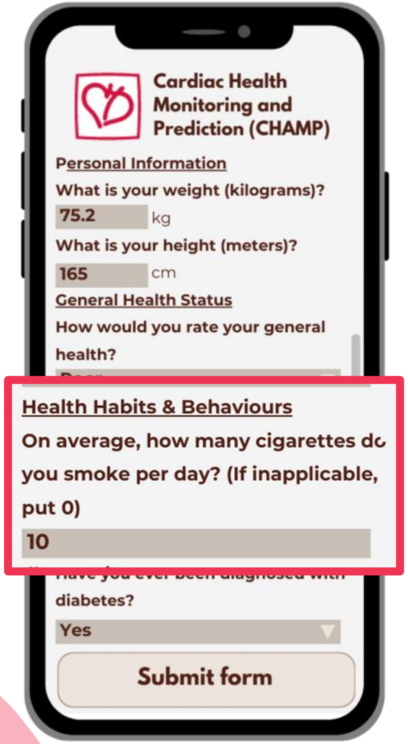
Running our Logistic Regression Model...

Results generated

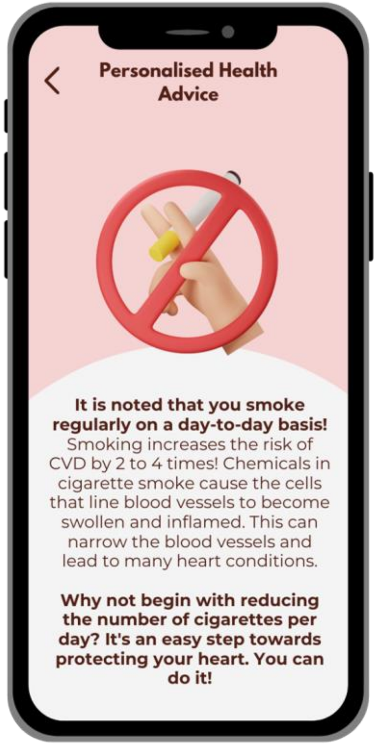
At risk for CVD: Yes

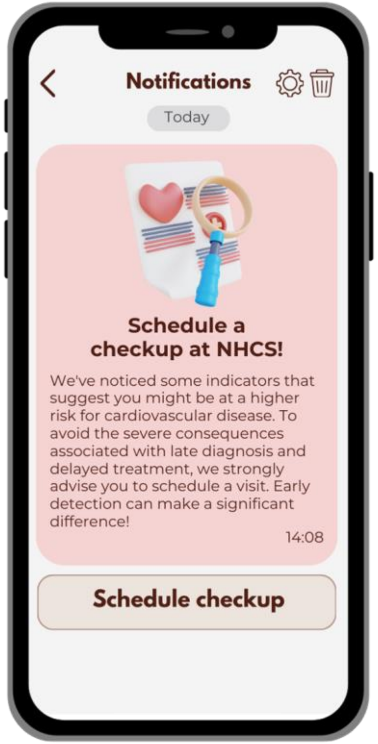
Probability: 0.805

Preventive Features: At risk for CVD



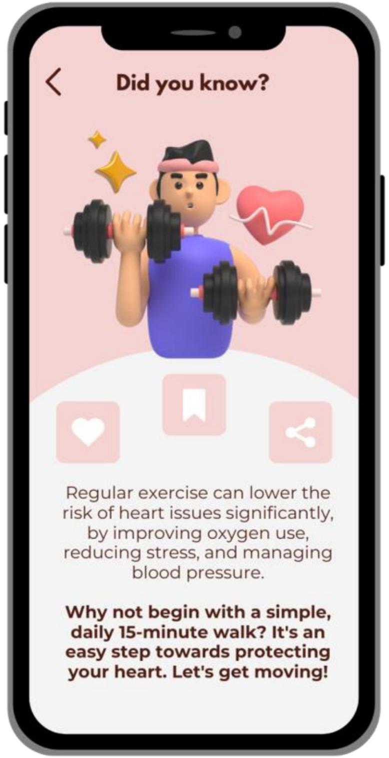
1. Personalised Health Advice





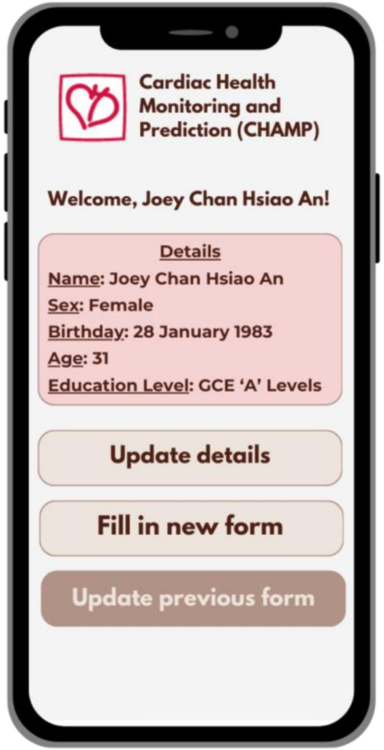
2. Scheduled Check-up Notifications

Preventive Features: Not at risk



- **Educational information** about CVD
- **Raise awareness** even if they are not at risk

Preventive Features: For everyone



Quarterly reminders to update Risk Assessment Form

- **Continuous monitoring** to **track improvements**
- **Dynamic adjustments** in health advice

Sally

Review of CHAMP App

Predictive Analytics Models

Personalised Preventive Measures

Public Education

Value Proposition

User-Friendly Interface

Simplified Registration

Freely Available

Preventive Approach

Risk Prediction

Lifestyle Integration

Tenia

Feasibility

Initial Investment

Software Development and Singpass integration

Free Accessibility

Eliminate financial barriers for users

Financial Empowerment

Enabling proactive preventive measures without cost burdens.

Early Detection

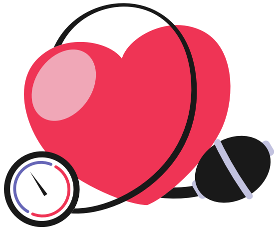
Facilitates early detection of CVD

Long-Term Cost Saving

Long-term benefits outweigh short-term costs

Tenia

Feasibility

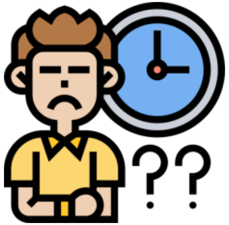


CVD Mortality Rate

Indicates **effectiveness in reducing Cardiovascular Disease (CVD) mortality**

Cost of Treatment

Assess the app's **effectiveness in reducing healthcare expenses** for users

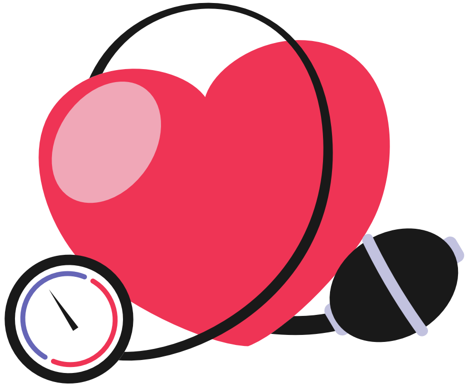


Waiting Time

Evaluate **improvements in healthcare accessibility** facilitated by CHAMP

06

Improvements & Conclusion



Tenia

Limitations & Improvements

	Limitation	Improvement
Inaccurate Information	Accuracy reliant on **completeness** of user-provided data.	Integrate wearable devices for real-time health data and schedule periodic checkups for validation.
Subjectivity of Self-Assessment	User input introduces **variability** in predictions.	Provide structured prompts for classification, reducing subjectivity and improving accuracy.
Limited Comprehensiveness of Risk Factors	Current dataset **lacks crucial factors** like dietary habits.	Access local datasets with broader factors and utilize user data for enhanced predictive ability.

Conclusion

Enhance Clinical Capabilities

Optimal Patient Care

Cost-Efficiency and Sustainability

Reputable Leader in Cardiovascular Care

Tenia

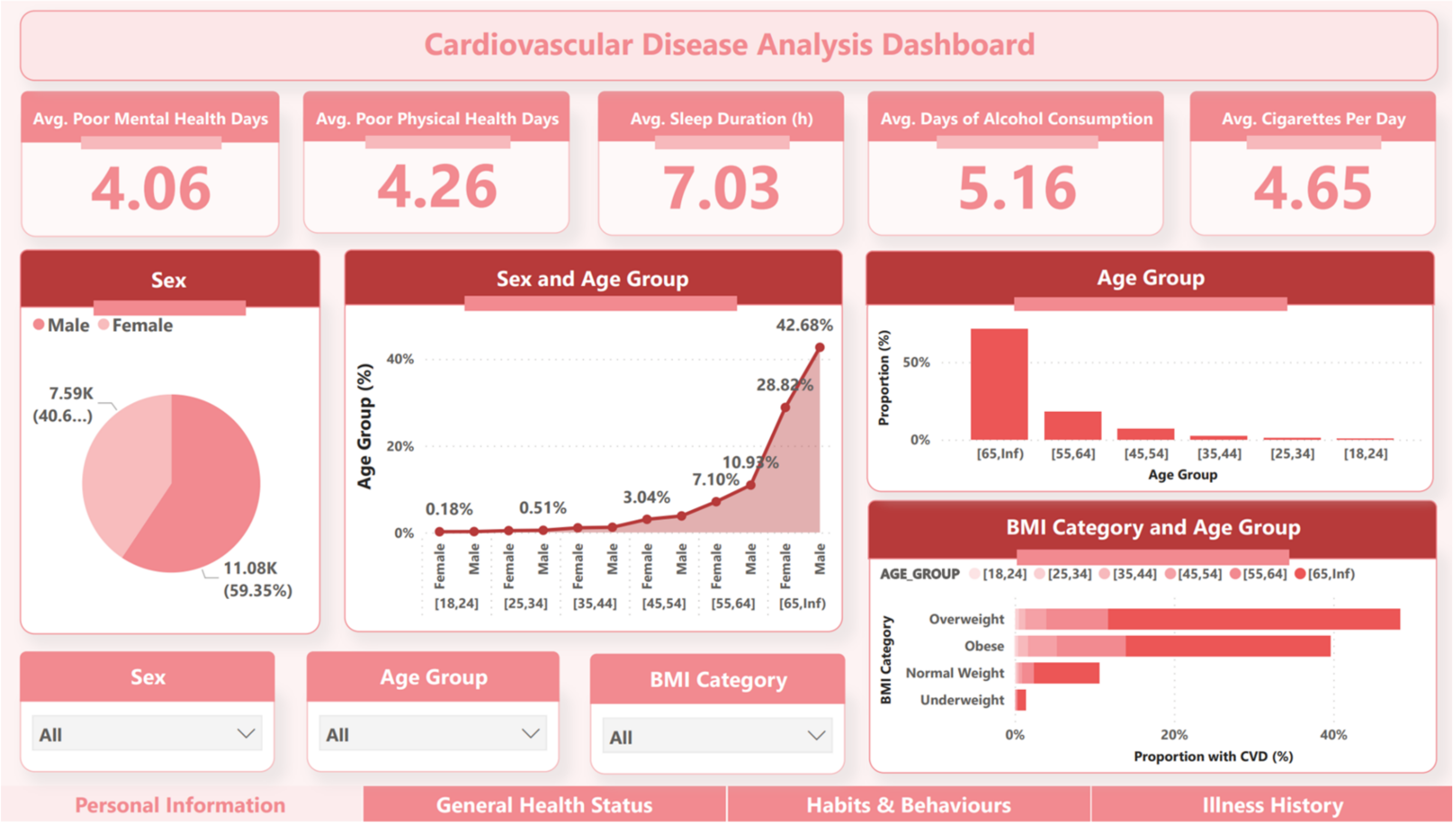
PowerBI Demonstration

Thank You!

Do you have any questions?

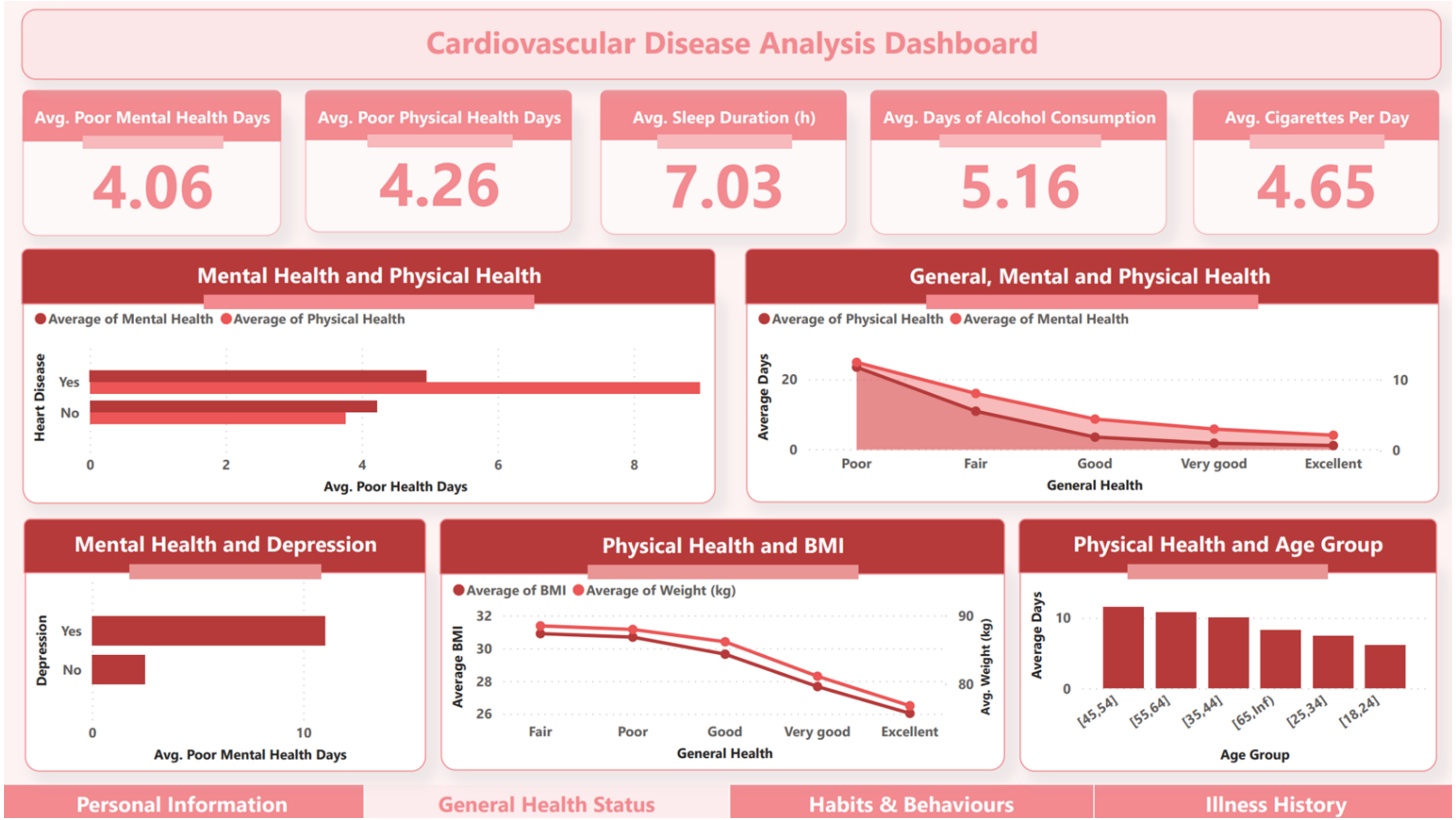
Appendix: Dashboard

Tenia



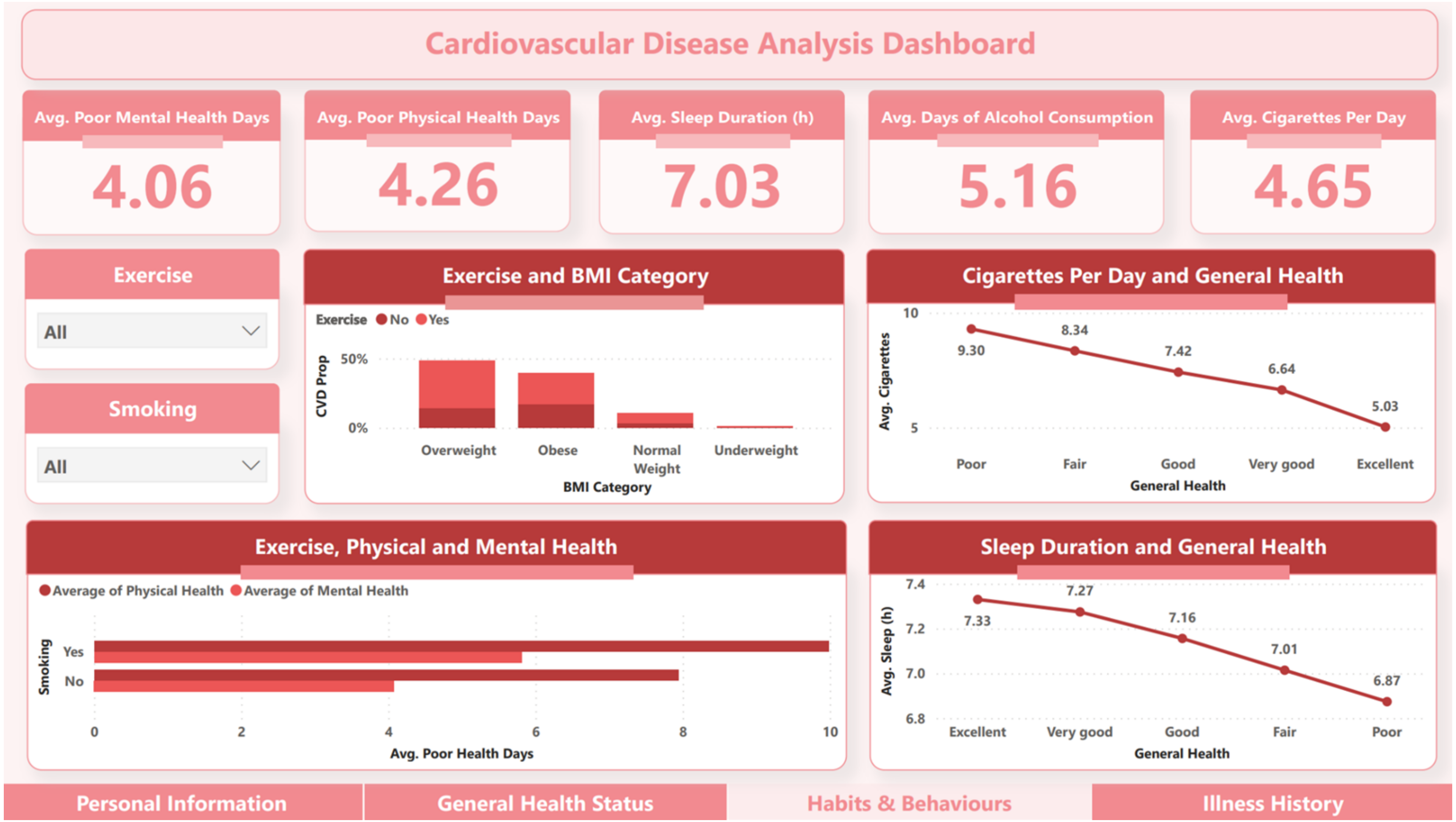
Appendix: Dashboard

Tenia



Appendix: Dashboard

Tenia



Appendix: Dashboard

Tenia