Managing the Prevalence of **Diabetes** in Singapore

by MINISTRY OF HEALTH SINGAPORE in partnership with Health Promotion Board

Group 2: Clarence Mun, Conrad Aw, Syahiran Rafi
DSI-SG-42 / Project #4

Who are we?

MINISTRY OF HEALTH
SINGAPORE

Consultants
from MOH

Who are you?

Health
Promotion
Board

Public Engagement Team
from HPB

Agenda

01

Context & Problem Statement

02

Data Collection & Feature Engineering

03

Exploratory Data Analysis

04

Model Evaluation

05

Implementation

06

Cost-Benefit Analysis

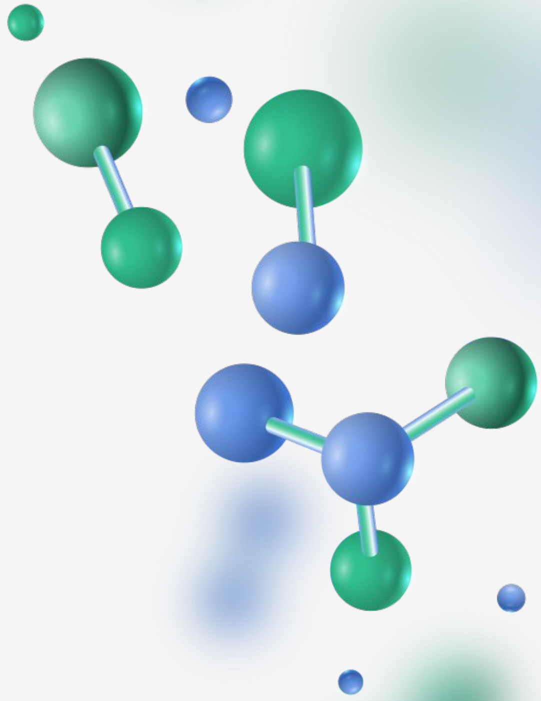
07

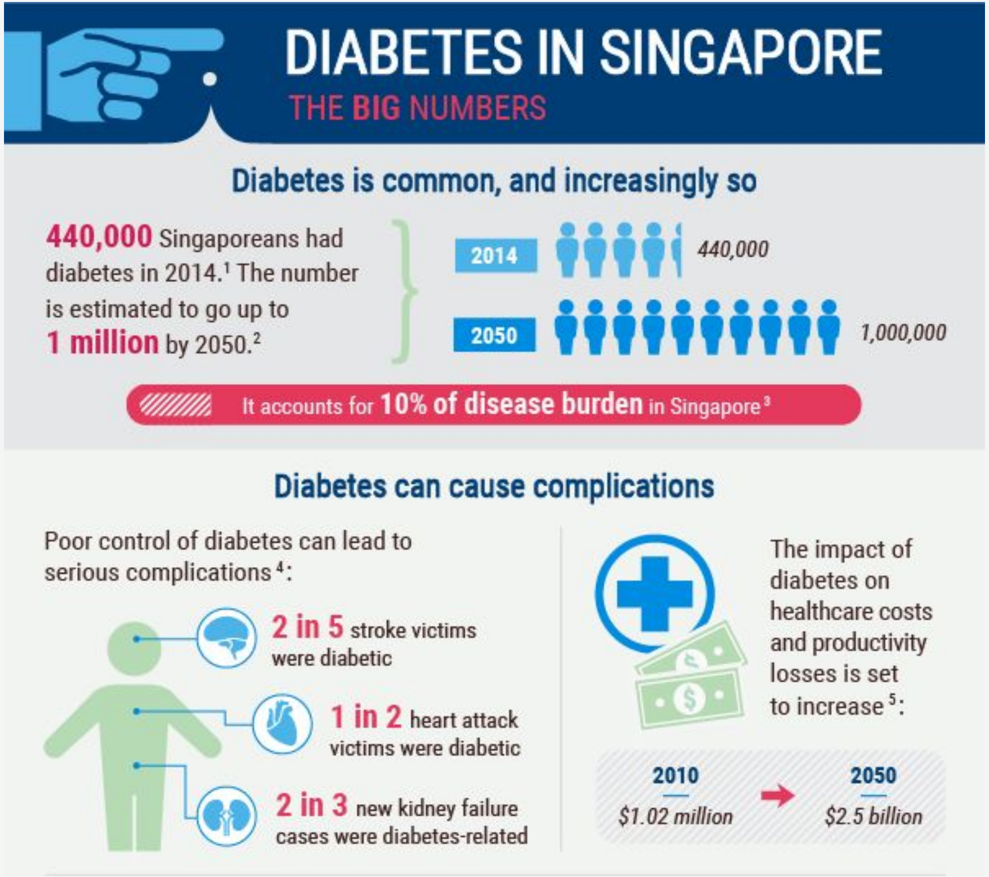
Conclusion & Recommendations


MYHEALTH
BY HPB

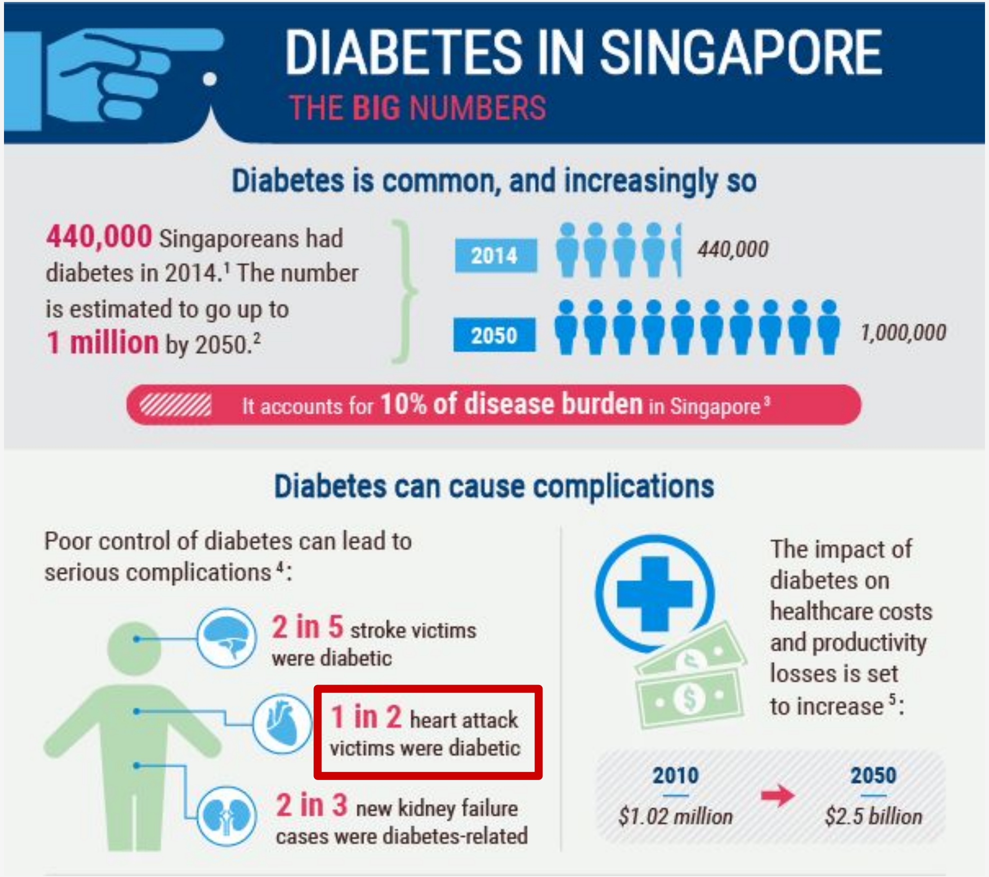
01

Context & Problem Statement





Source: diabetes.org.sg



Source: diabetes.org.sg

Problem Statement

According to the Ministry of Health, about one in three Singaporeans has a lifetime risk of developing diabetes. To address this challenge, we propose developing a data-driven solution that utilises healthcare data and predictive analytics to identify individuals at high risk of developing diabetes.

By leveraging classification algorithms and population health data, our solution aims to provide a risk assessment of diabetes for individuals to enable early detection and targeted intervention. Additionally, our solution also aims to equip individuals with the ability to make more informed nutritional choices by providing healthier drink suggestions based on their sugar content.

With this two-pronged approach, HPB is better positioned to manage diabetes among Singaporeans and reduce its associated healthcare burdens.

Who is Jasmine?

Jasmine is a 30-year-old **marketing executive** working in a fast-paced agency in Singapore. She feels that she is **generally healthy** as she has no major medical history, goes for a yearly health check-up and exercises at a spin studio 1-2 times a week.

What does Jasmine believe in?

Jasmine believes that **health is wealth**. She also believes that while access to good healthcare is a basic need, leading a healthy life **starts with the individual**.

Jasmine, 30

What are her goals?

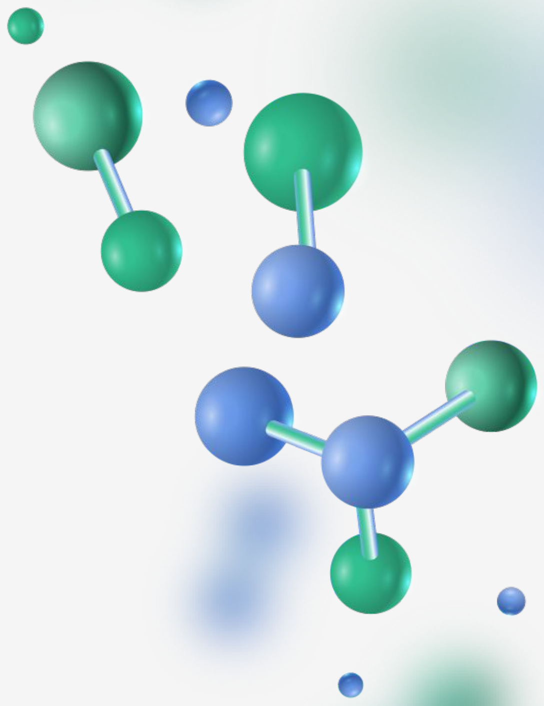
Jasmine hopes to **improve her overall well-being** by adopting healthier eating habits. She also wants to learn how better nutrition could help to **reduce her risk for certain chronic diseases**, particularly diabetes.

What's affecting her recently?

With an emphasis on career-building in recent years, **long working hours, high stress and irregular meals** are the norm for Jasmine. She fears that her current lifestyle could impact her health in the longer term.

02

Data Collection & Feature Engineering

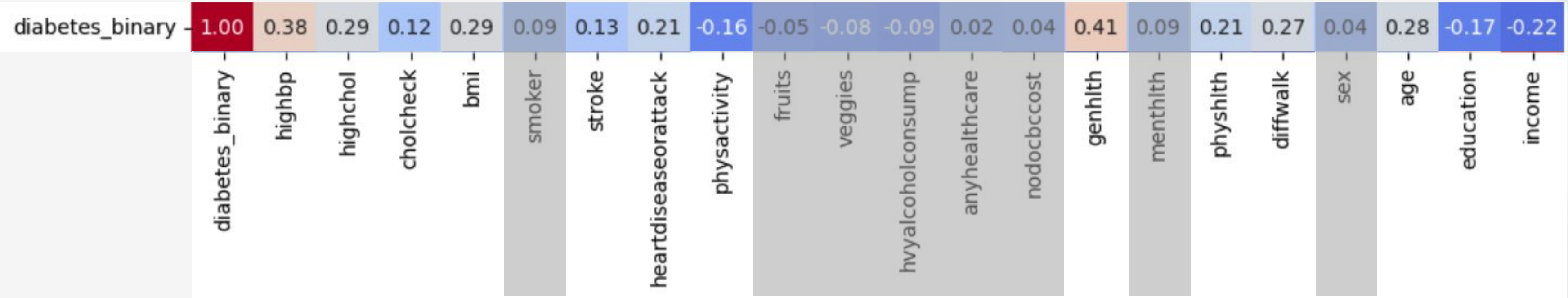


Our Datasets

The following datasets are obtained from the Behavioral Risk Factor Surveillance System (BRFSS) conducted by the Centers for Disease Control and Prevention (CDC) in 2015.

No. of features	No. of rows	Classes	Proportion of classes
21	70,692	0 - no diabetes or pre-diabetes 1 - diabetes	0 - 50% 1 - 50%

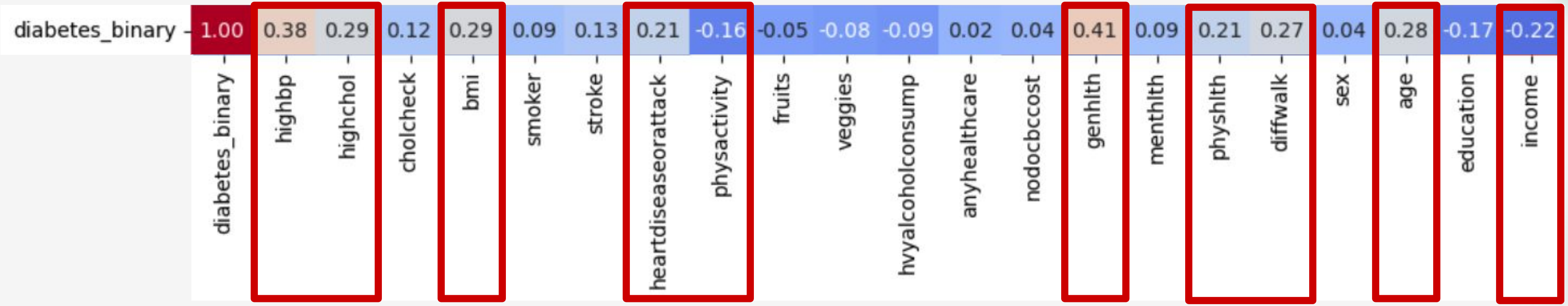
Correlation
between Diabetes and Other Features



dropped rows	
	smoker, fruits, veggies, heavyalcoholconsump, anyhealthcare, nodocbccost, menhtlh, sex

Correlation

between Diabetes and Other Features



diabetes_binary ⟷

highbp
highchol
bmi
heartdiseaseorattack
physactivity
genhlth
physhlth
diffwalk
age
income

Interaction Terms

genhlth_physhlth_interaction

bmi_highbp_diffwalk_interaction

age_highchol_heartdiseaseorattack_interaction

	genhlth	physhlth
genhlth	1.00	0.55
physhlth	0.55	1.00

Interaction Terms

```
genhlth_physhlth_interaction
```

```
bmi_highbp_diffwalk_interaction
```

```
age_highchol_heartdiseaseorattack_interaction
```

	bmi	highbp	diffwalk
bmi	1.00	0.24	0.25
highbp	0.24	1.00	0.23
diffwalk	0.25	0.23	1.00

Interaction Terms

genhlth_physhlth_interaction

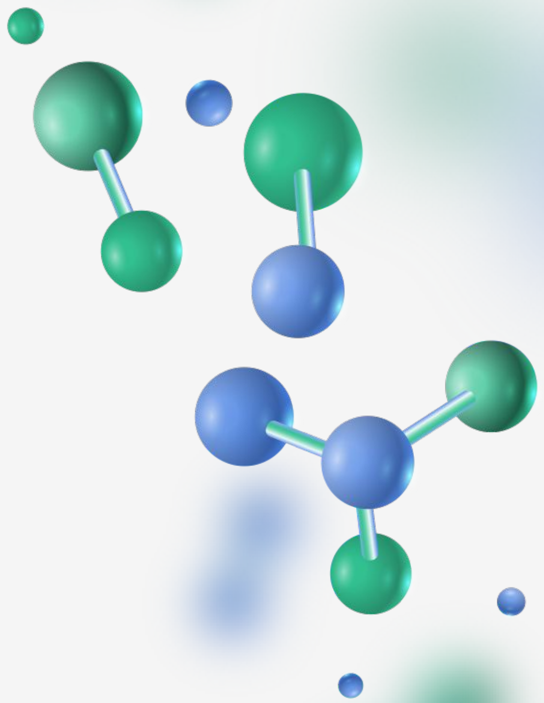
bmi_highbp_diffwalk_interaction

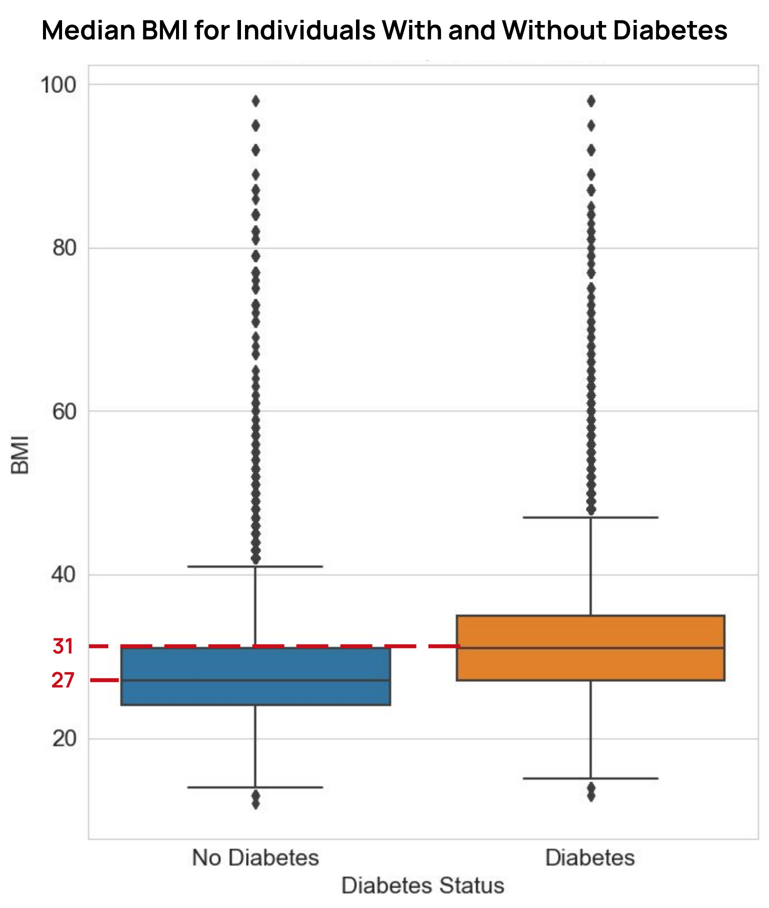
age_highchol_heartdiseaseorattack_interaction

	age	highchol	heartdiseaseorattack
age	1.00	0.24	0.22
highchol	0.24	1.00	0.18
heartdiseaseorattack	0.22	0.18	1.00

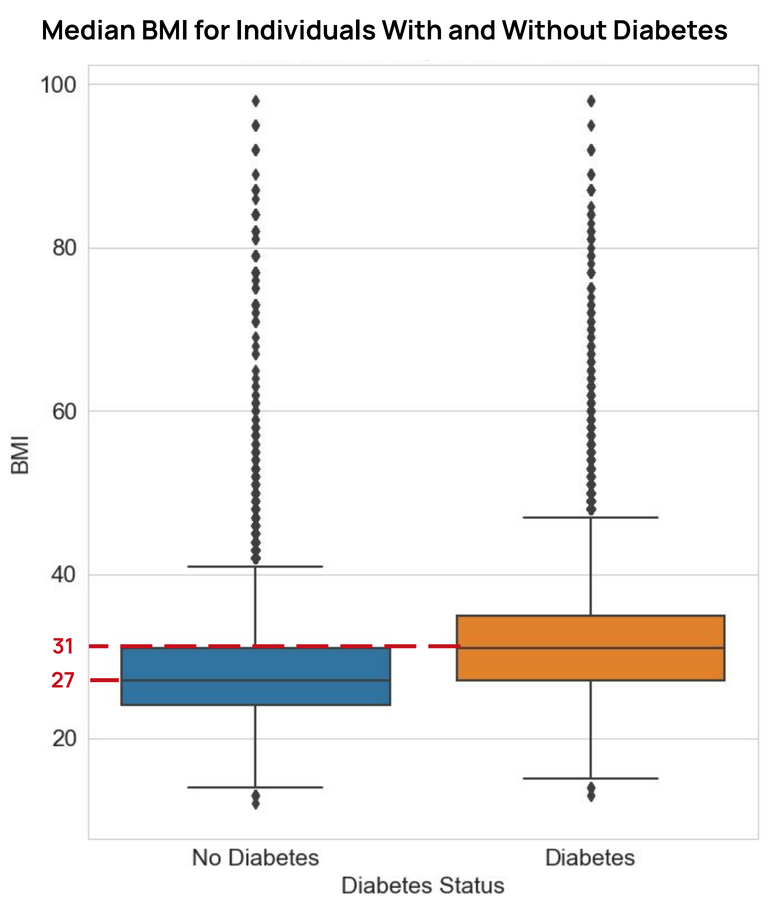
03

Exploratory Data Analysis



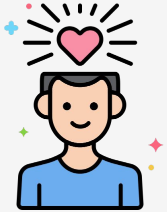


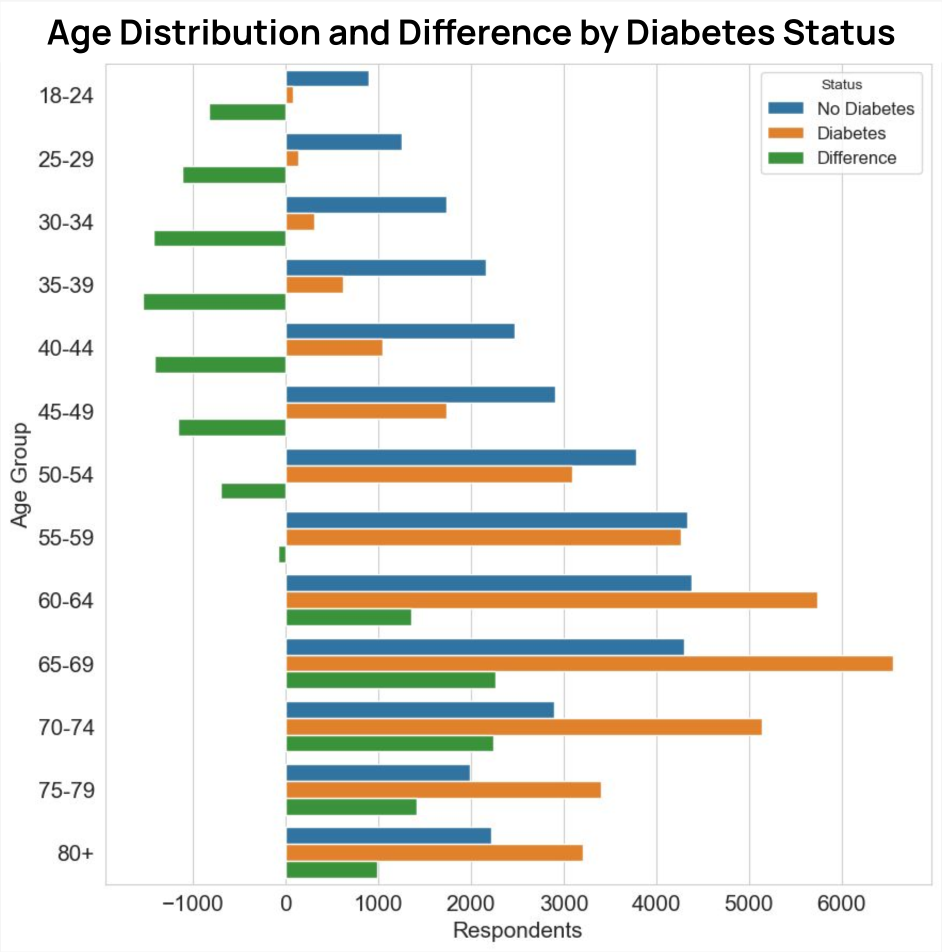
The median BMI is noticeably **higher in individuals with diabetes** compared to those without.



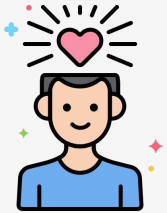
The median BMI is noticeably **higher in individuals with diabetes** compared to those without.

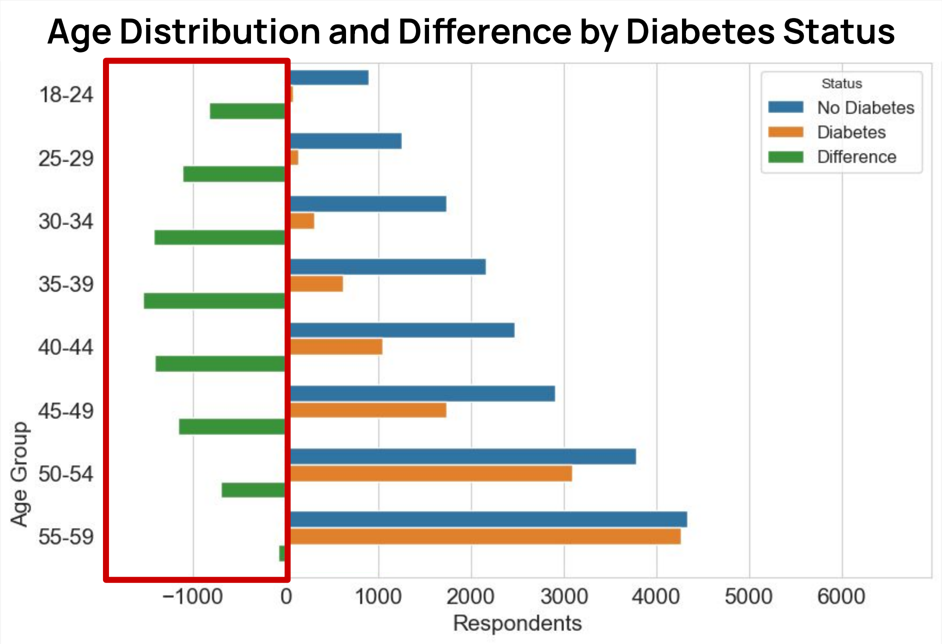
A **higher BMI** may be linked to an **increased risk of diabetes**, aligning with previous studies.



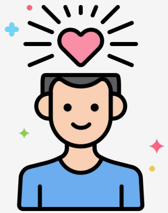


Orange - Blue = Green

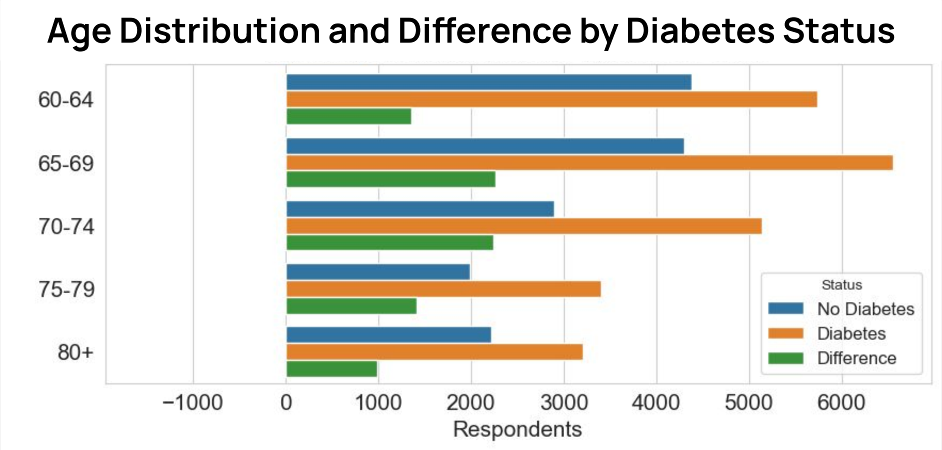




Generally, **below 60**, there is a **higher proportion of non-diabetics** compared to diabetics.

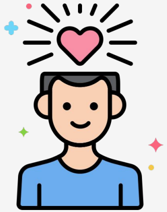


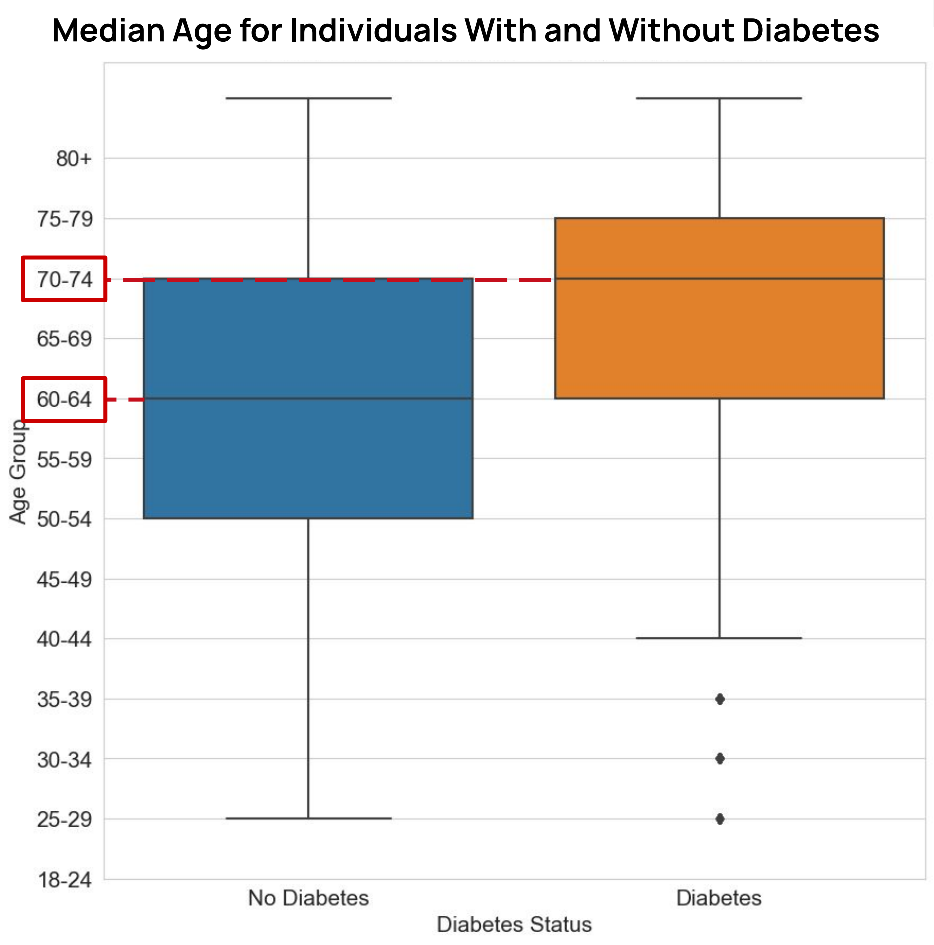
Prevalence of diabetes is notably **greater in older age groups**.



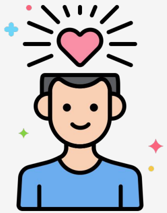
The majority of individuals **aged 60-79** with **diabetes outnumber those without diabetes**.

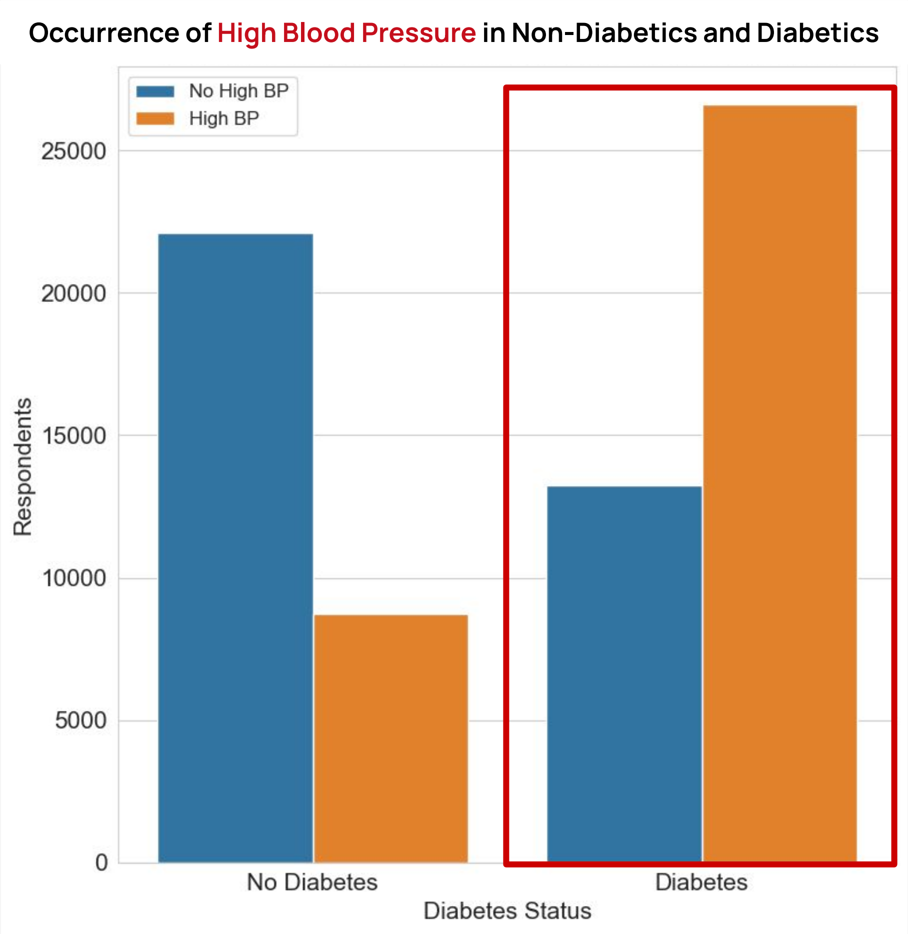
This trend shows that the **risk** for developing diabetes generally **increases with age**.



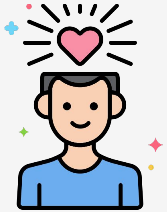


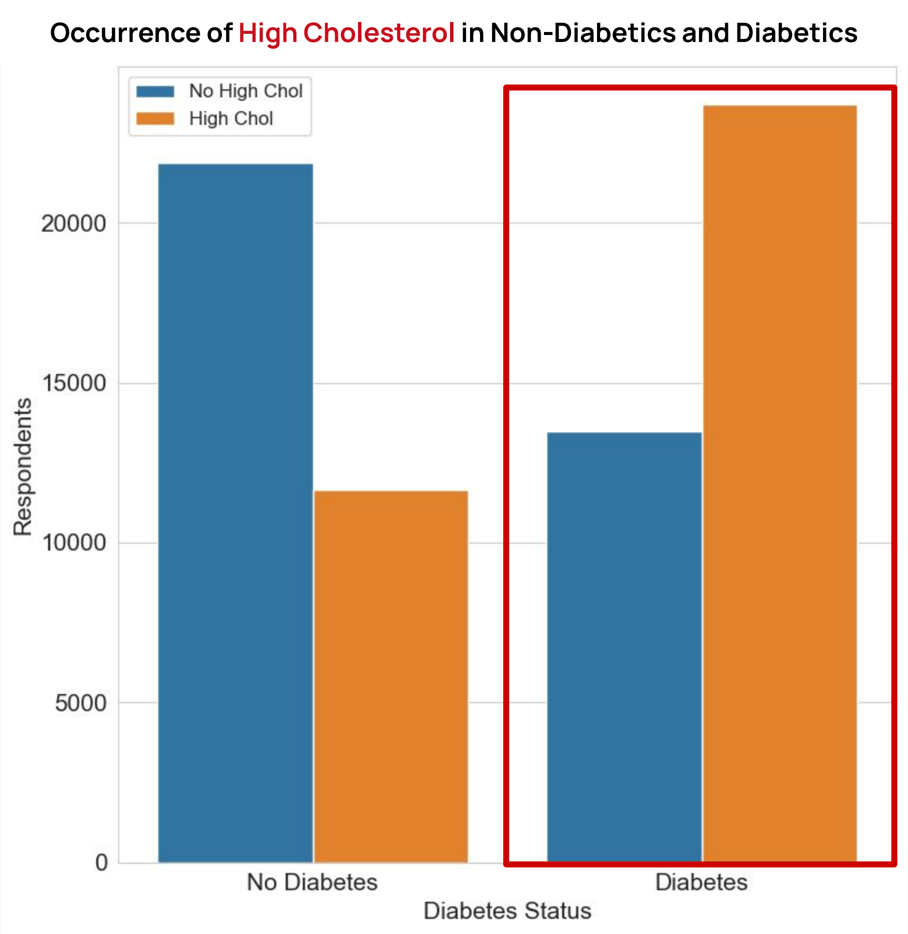
The **median age** of individuals with diabetes is **higher than those without diabetes**, reinforcing the idea that diabetes risk escalates as people age.





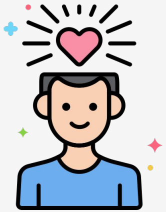
A substantial number of **individuals with diabetes also have high blood pressure**, highlighting the known link between these conditions.

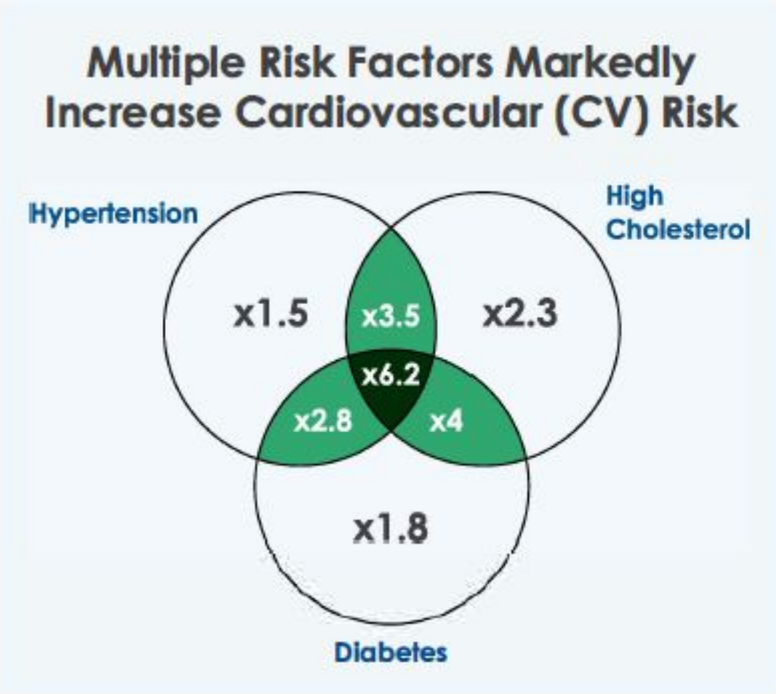




Similar to high blood pressure, **high cholesterol** appears to be **common among individuals with diabetes**.

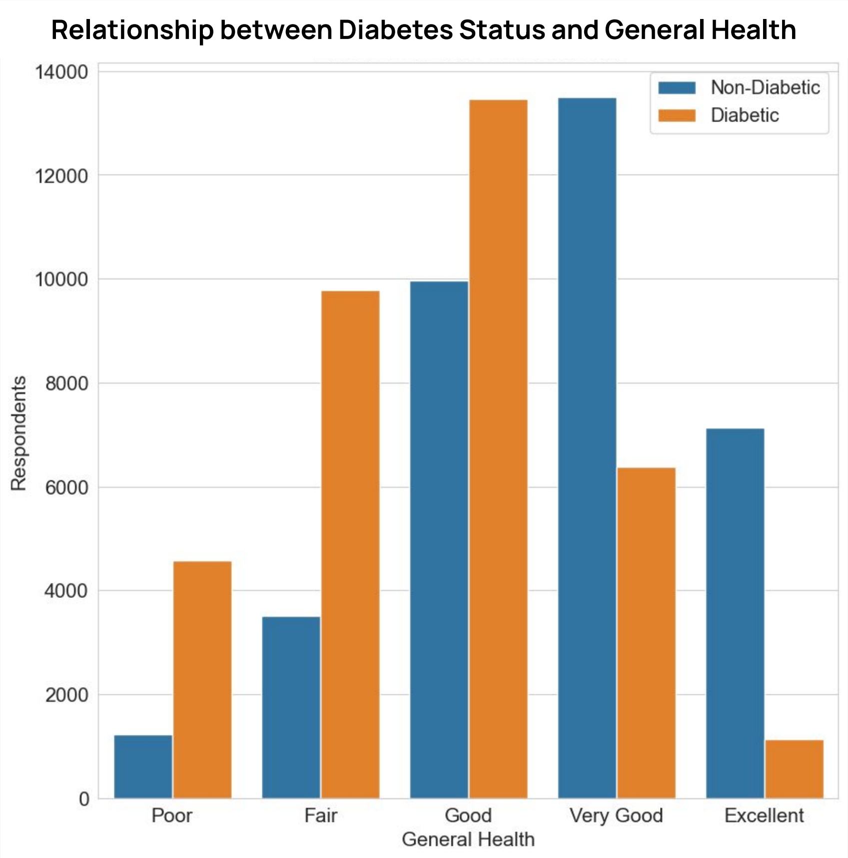
This reinforces the established connection between lipid levels and diabetes risk.



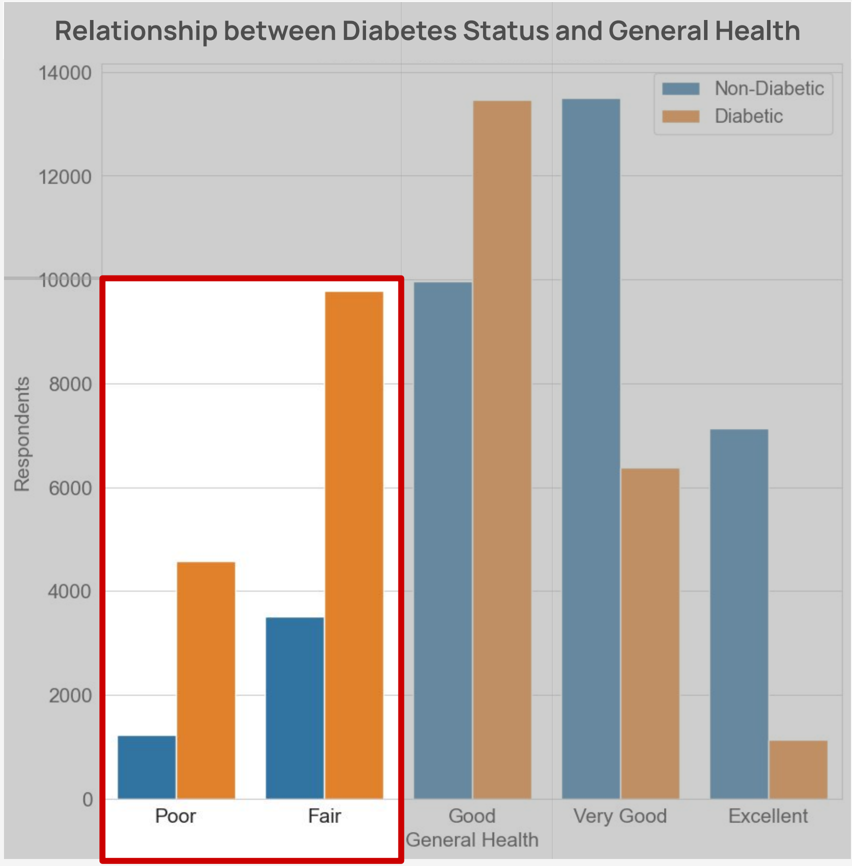


If left unmanaged, these chronic conditions could lead to **heart disease** and **stroke** with staggering long-term implications, impacting quality of life.

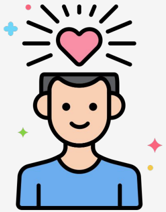
Source: Singapore Heart Foundation, DrAnywhere

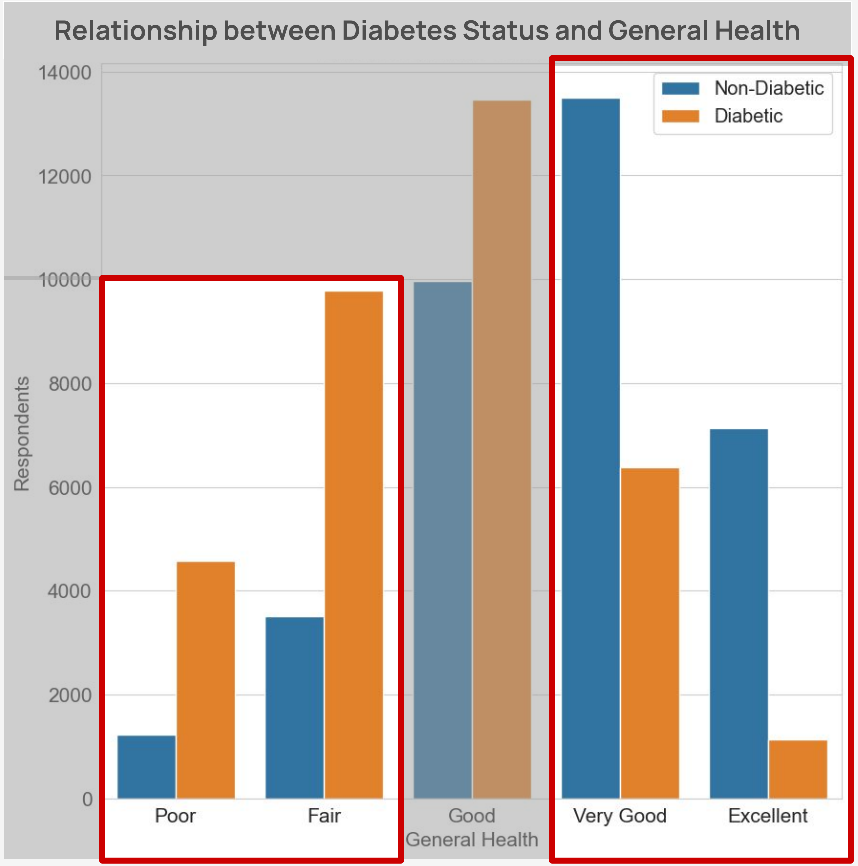


Relationship between Diabetes Status and General Health
Non-Diabetic
Diabetic
Respondents
14000
12000
10000
8000
6000
4000
2000
0
Poor
Fair
Good
Very Good
Excellent
General Health

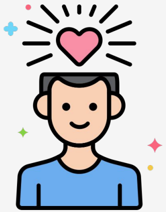


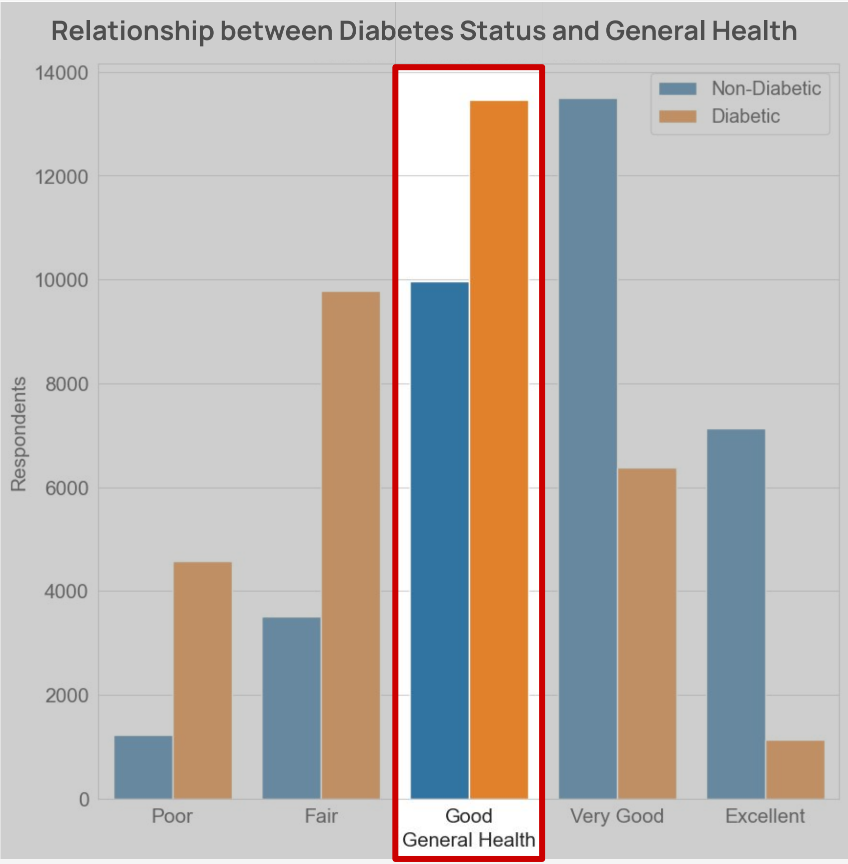
Prevalence of diabetes is **higher** amongst individuals with **poorer general health**





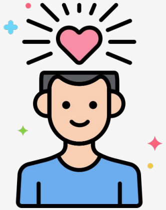
Prevalence of diabetes is **higher** amongst individuals with **poorer general health**, and **lower** amongst individuals with **better general health**.

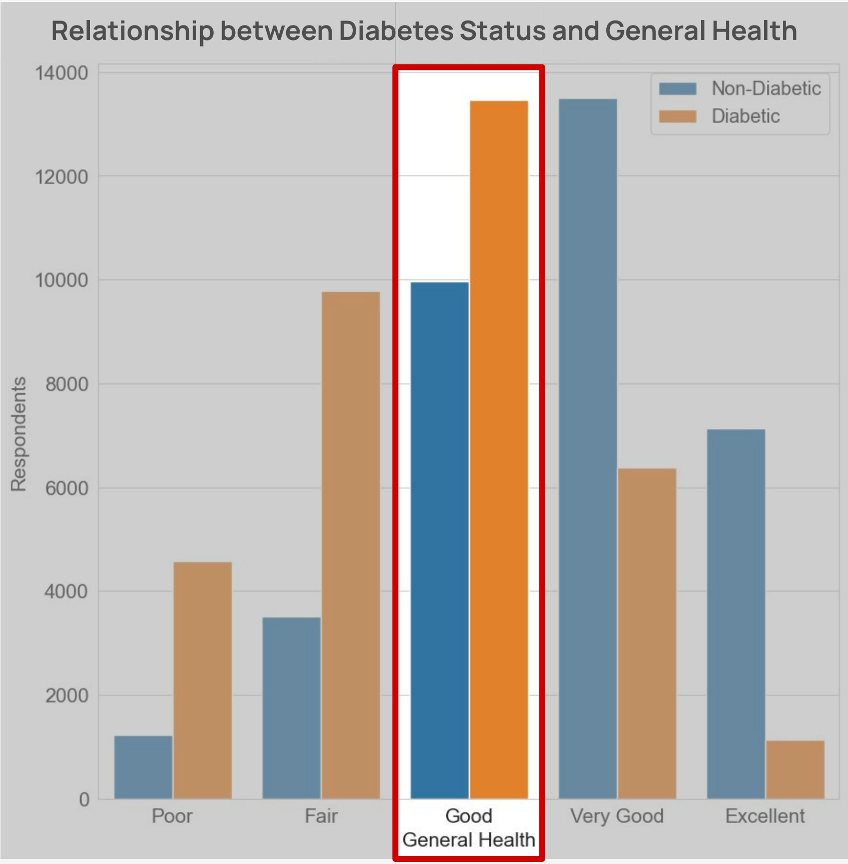




There is a **higher proportion of diabetics** amongst individuals with **“good” general health**.

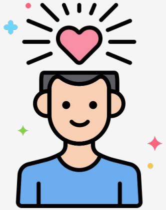
What does this mean?





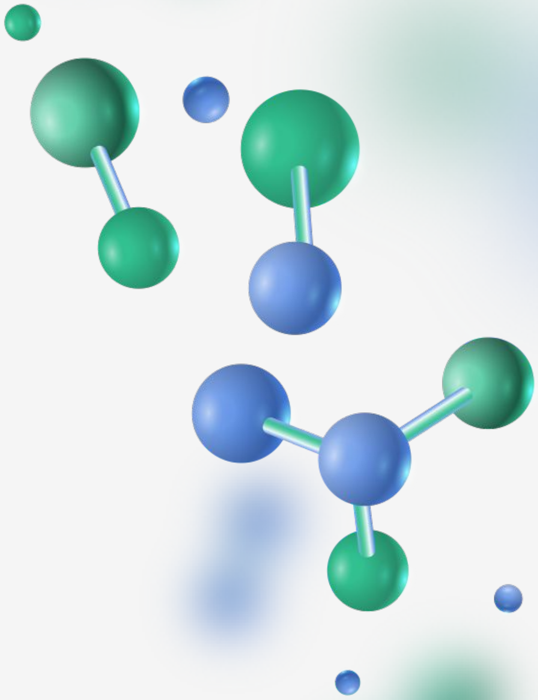
Quality of life for diabetics may not be impacted so significantly with **effective management** of the disease.

With early detection and/or intervention, symptoms can be better managed.



04

Model Evaluation



Initial Models

	Train Score	Cross-Validation Score
Logistic Regression	0.747	0.747
Random Forest	0.965	0.722
XGBoost	0.774	0.745
Decision Tree	0.965	0.662
Gradient Boost	0.753	0.751
Support Vector Machine	0.746	0.745
Neural Network	0.751	0.745

Initial Models

	Train Score	Cross-Validation Score
Logistic Regression	0.747	0.747
~~Random Forest~~	0.965	0.722
XGBoost	0.774	0.745
~~Decision Tree~~	0.965	0.662
Gradient Boost	0.753	0.751
Support Vector Machine	0.746	0.745
Neural Network	0.751	0.745

Choosing the Baseline Model

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Logistic Regression	0.745	0.733	0.770	0.719	0.751
XGBoost	0.744	0.726	0.784	0.704	0.754
Gradient Boost	0.747	0.729	0.788	0.707	0.757
Support Vector Machine	0.741	0.716	0.798	0.684	0.755
Neural Network	0.746	0.709	0.835	0.658	0.767

Choosing the Baseline Model

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Logistic Regression	0.745	0.733	0.770	0.719	0.751
XGBoost	0.744	0.726	0.784	0.704	0.754
Gradient Boost	0.747	0.729	0.788	0.707	0.757
Support Vector Machine	0.741	0.716	0.798	0.684	0.755
Neural Network	0.746	0.709	0.835	0.658	0.767

Choosing the Baseline Model

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Logistic Regression	0.745	0.733	0.770	0.719	0.751
XGBoost	0.744	0.726	0.784	0.704	0.754
Gradient Boost	0.747	0.729	0.788	0.707	0.757
Support Vector Machine	0.741	0.716	0.798	0.684	0.755
Neural Network	0.746	0.709	0.835	0.658	0.767

Choosing the Baseline Model

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Support Vector Machine	0.741	0.716	0.798	0.684	0.755
Neural Network	0.746	0.709	0.835	0.658	0.767

SVM has better **precision** and **specificity**.

Choosing the Baseline Model

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Support Vector Machine	0.741	0.716	0.798	0.684	0.755
Neural Network	0.746	0.709	0.835	0.658	0.767

NN has better **accuracy**, **sensitivity** and **F1-score**.

Shortlisted Model

	GridSearch Runtime
Support Vector Machine	› 90 mins
Neural Network	approx. 15 min

Neural Network: Before and After Tuning

	Pre-Tuning Score	Post-Tuning Score	Percentage Change
Accuracy	0.746	0.748	+0.15%
Precision	0.709	0.722	+1.77%
Sensitivity	0.835	0.805	-3.55%
Specificity	0.658	0.690	+4.85%
F1-Score	0.767	0.761	-0.74%

While there is a drop in sensitivity, the post-tuning **sensitivity** score (0.805) **still exceeds those of the other models** we considered.

Neural Network: Before and After Tuning

	Pre-Tuning Score	Post-Tuning Score	Percentage Change
Accuracy	0.746	0.748	+0.15%
Precision	0.709	0.722	+1.77%
Sensitivity	0.835	0.805	-3.55%
Specificity	0.658	0.690	+4.85%
F1-Score	0.767	0.761	-0.74%

While there is a drop in sensitivity, the post-tuning **sensitivity** score (0.805) **still exceeds those of the other models** we considered. The 0.74% drop in **F1-score** should not affect the model's performance significantly.

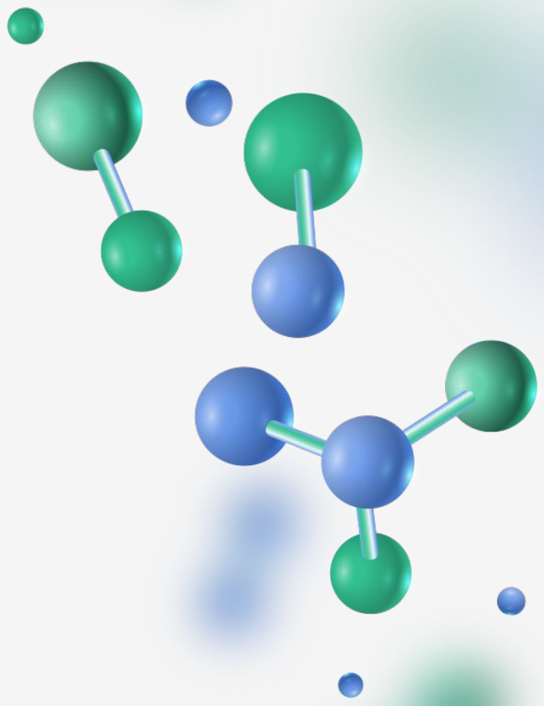
Neural Network: Before and After Tuning

	Pre-Tuning Score	Post-Tuning Score	Percentage Change
Accuracy	0.746	0.748	+0.15%
Precision	0.709	0.722	+1.77%
Sensitivity	0.835	0.805	-3.55%
Specificity	0.658	0.690	+4.85%
F1-Score	0.767	0.761	-0.74%

While there is a drop in sensitivity, the post-tuning **sensitivity** score (0.805) **still exceeds those of the other models** we considered. The 0.74% drop in **F1-score** should not affect the model's performance significantly. Three other performance metrics – **accuracy**, **precision** and **specificity** – increased, making the model **more well-balanced overall**.

05

Implementation



Who is Jasmine?

Jasmine is a 30-year-old **marketing executive** working in a fast-paced agency in Singapore. She feels that she is **generally healthy** as she has no major medical history, goes for a yearly health check-up and exercises at a spin studio 1-2 times a week.

What does Jasmine believe in?

Jasmine believes that **health is wealth**. She also believes that while access to good healthcare is a basic need, leading a healthy life **starts from the individual**.

Jasmine, 30

What are her goals?

Jasmine hopes to **improve her overall well-being** by adopting healthier eating habits. She also wants to learn how better nutrition could help to **reduce her risk for certain chronic diseases**, particularly diabetes.

What's affecting her recently?

With an emphasis on career-building in recent years, **long working hours, high stress and irregular meals** are the norm for Jasmine. She fears that her current lifestyle could impact her health in the longer term.

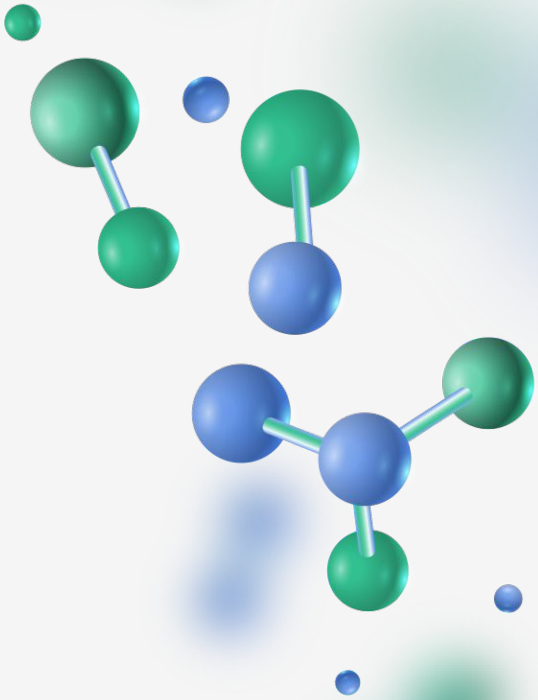
MYHEALTH
BY HPB

- Diabetes Risk Assessment
- Sugar Content Detector

App Demo

06

Cost-Benefit Analysis



Development Costs

Personnel costs	• Data scientists: Assume an average salary of SGD 80,000 per year. • Software developers: Assume an average salary of SGD 70,000 per year. • Project managers: Assume an average salary of SGD 90,000 per year.
Technology and infrastructure	Servers, software licenses, cloud services, etc.
Research and data acquisition	Costs associated with gathering and purchasing data, particularly for the diabetes predictive model.
App development	• Design, user interface, and user experience costs. • Development of OCR technology or licensing existing technology.
Testing and quality assurance	Costs associated with beta testing, pilot studies, etc.
Marketing and promotion	Costs to promote the app to ensure adequate user base.

Operational Costs

Maintenance	Ongoing server costs, app updates, and troubleshooting.
Support staff	Customer service and technical support.

Direct vs. Indirect Benefits

Direct Benefits	Indirect Benefits
Improved Health Outcomes: Early detection and management of diabetes can significantly reduce the cost of healthcare associated with the disease.	**Increased Productivity:** Healthier individuals contribute more effectively to the economy.
Cost Savings for Healthcare System: Reducing the incidence and severity of diabetes can lead to substantial savings in medical costs.	**Public Health Data:** Data collected can be used for further research and improvement in health policies.
	Educational Value: The app can raise awareness and educate the public on healthy eating habits.

Cost vs. Benefit Summary

Assumptions:

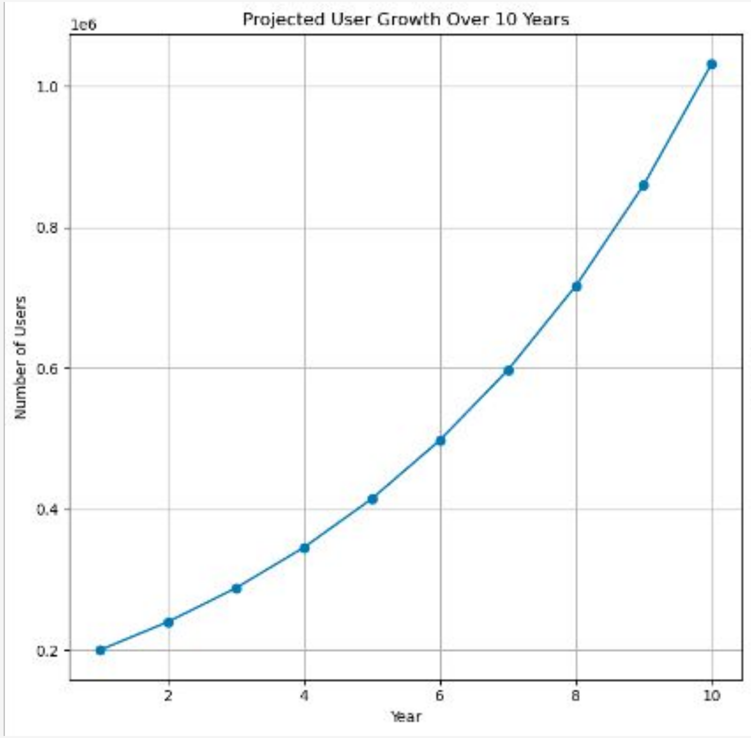
- Assume 3 years of development before launch.
- Assume maintenance costs are 20% of the initial development cost annually.
- Assume 200,000 active users by the third year post-launch.

Costs	Benefits
Development Team: • 2 Data Scientists • 3 Software Developers • 1 Project Manager Operational Yearly Costs: • Maintenance, support staff. Miscellaneous Costs: • Marketing • Licensing fees for OCR tech	Healthcare Cost Reduction: • Based on studies, early diabetes intervention can save approximately SGD 5,000 per patient per year. Productivity Gains: • Reduced sick days and higher employment rates among healthier populations.

Cost vs. Benefit Summary

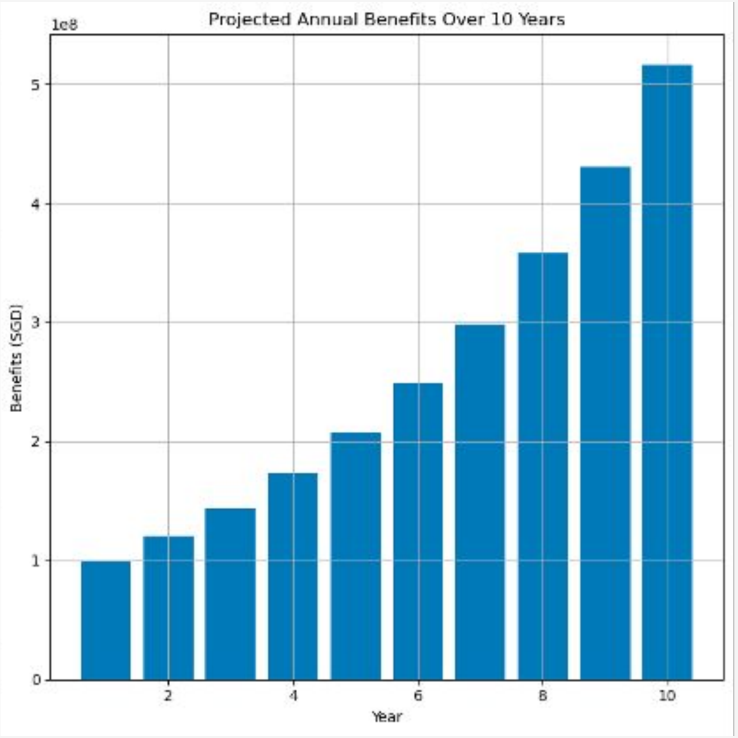
Cost Breakdown	Benefit Breakdown
Total Development Costs (over 3 years): **SGD 1,380,000** Other Development Costs (licenses, technology, data acquisition, etc.): **SGD 690,000** Total Operational Costs (for the first 3 years post-launch): **SGD 1,242,000** Marketing and Other Costs (one-time): **SGD 276,000** Total Costs Over 6 Years: **SGD 3,588,000**	Early diabetes management saves about **SGD 5,000 per patient per year**. Assume early detection and improved management impact **10% of users per year**. **200,000 active users** by the third year post-launch, **growing at 20% annually** thereafter.

Summary of Benefits



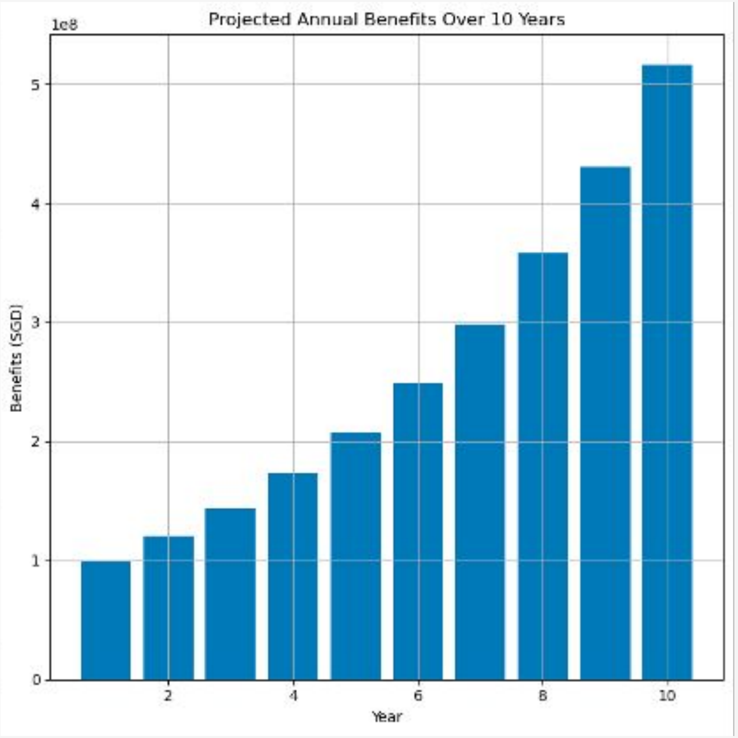
Projected User Growth:
Starting from 200,000 users in the third year, we expect a 20% annual growth rate. This growth reflects the increasing adoption and reach of the app.

Summary of Benefits



Projected Annual Benefits:
These are calculated based on the assumption that 10% of users benefit from the early management of diabetes, resulting in healthcare savings of SGD 5,000 per patient per year.

Summary of Benefits



Projected Cumulative Benefits Over 10 Years:

SGD 2,595,868,211

With the total costs over the first 6 years amounting to approximately SGD 3.59 million, and cumulative benefits over 10 years reaching about SGD 2.60 billion, the project presents a significant return on investment primarily due to the potential healthcare savings and improved public health outcomes.

Summary of Benefits

This table encapsulates the key financial aspects of the project over its developmental and operational phases, along with the projected cumulative benefits over a 10-year period following its launch.

Description	Timeframe	Amount (SGD)
Total development costs	First 3 years	1,380,000
Other development costs	First 3 years	690,000
Total operational costs	First 3 years	1,242,000
Marketing and other costs	One-time	276,000
Total costs over first 6 years	Up to year 6	3,588,000
Annual benefits (year 3 to year 12)	Year 3 to 12	2,595,868,211
ROI	After 10 years	72,248.61%

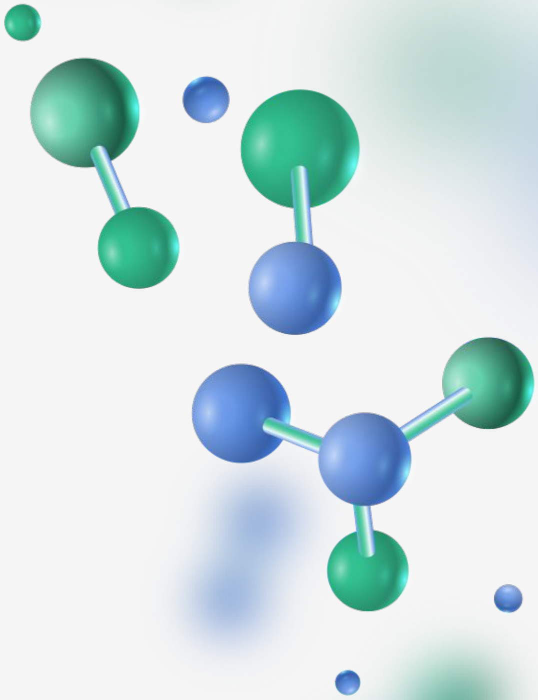
Other Considerations

The estimated return on investment (ROI) for the diabetes predictive app, though exceptionally high at over 72,000%, may not be entirely realistic due to optimistic assumptions in several areas:

Impact Scale: The assumption that 10% of users will annually achieve significant health outcomes and cost savings might be overly optimistic. Real-world factors such as patient adherence and diverse health conditions could affect these outcomes.	**User Adoption:** The projected growth rates and user base are ambitious. Achieving widespread adoption requires significant efforts and is influenced by factors like user trust, app effectiveness, and integration with healthcare systems.	**Cost Estimates:** Development and operational costs might be underestimated, especially if unforeseen technical or regulatory challenges arise. Additionally, costs related to compliance with health data regulations may not have been fully considered.
Healthcare Savings: The assumed savings of SGD 5,000 per patient per year may not apply universally across different stages of diabetes or vary with healthcare system differences. Savings are also dependent on patient compliance and other health issues.	**Economic Conditions:** The analysis doesn't account for variable economic factors such as inflation, changes in healthcare policy, or economic downturns, which could impact both costs and benefits.	

07

Conclusion & Recommendations



In 2016...

THE STRAITS TIMES

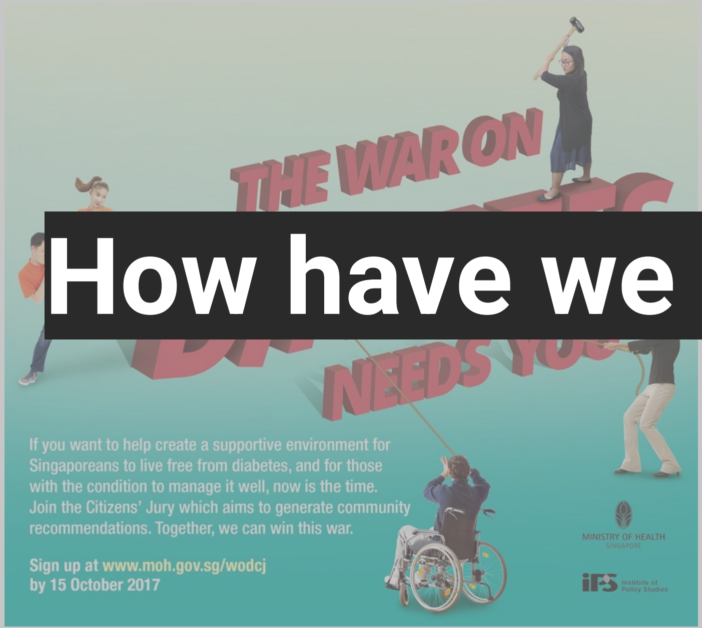
Parliament: Health Minister Gan Kim Yong declares 'war on diabetes', new task force set up

A diabetic patient undergoing dialysis treatment at Kim Keat Dialysis Centre. ST PHOTO: DESMOND WEE

UPDATED APR 14, 2016, 07:11 AM

Source: GovTechSG, The Straits Times, Apr 2016

In 2016...




THE STRAITS TIMES

Parliament: Health Minister Gan Kim Yong declares 'war on diabetes', new task force set up

A diabetic patient undergoing dialysis treatment at Kim Keat Dialysis Centre. ST PHOTO: DESMOND WEE

UPDATED APR 14, 2016, 07:11 AM

How have we fared since then?

Source: GovTechSG, The Straits Times, Apr 2016

In 2021...

THE STRAITS TIMES

Slight increase in diabetes prevalence despite 5-year war against disease

For the period of 2019 to 2020, the crude prevalence of diabetes was 9.5 per cent, an increase from 8.8 per cent in 2017. ST PHOTO: DESMOND WEE

UPDATED NOV 19, 2021, 07:22 AM

Source: The Straits Times, Nov 2021

In 2023...

Best News Website or Mobile Service • Digital Media Awards Worldwide 2022

Top Stories Latest News Discover Singapore Asia Commentary Sustainability

War against diabetes: Doctors seeing rise in patients below 40 due to lifestyle habits, early screening

More than 400,000 people in Singapore live with diabetes, with the number projected to rise to 1 million by 2050.

Sherlyn Seah & Calvin Yang

15 Nov 2023 05:52PM

Source: CNA, Nov 2023

Problem Statement

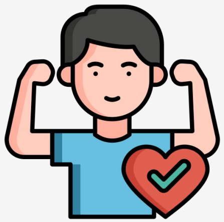
According to the Ministry of Health, about **one in three Singaporeans** has a lifetime risk of developing diabetes. To address this challenge, we propose developing a data-driven solution that utilises healthcare data and predictive analytics to **identify individuals at high risk of developing diabetes**.

By leveraging classification algorithms and population health data, our solution aims to provide a risk assessment of diabetes for individuals to enable early detection and targeted intervention. Additionally, our solution also aims to equip individuals with the ability to make more informed nutritional choices by providing healthier suggestions for everyday food products.

With this two-pronged approach, HPB is better positioned to **manage diabetes among Singaporeans** and **reduce its associated healthcare burdens**.

Conclusion

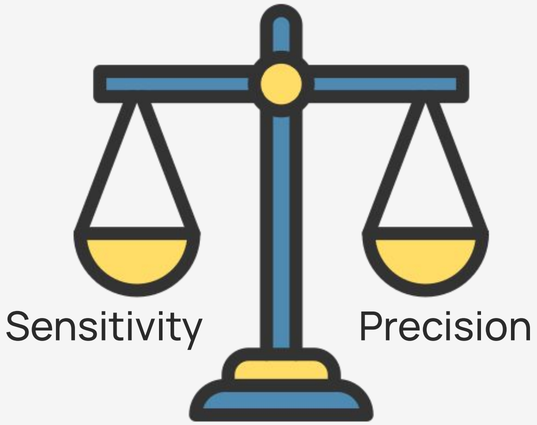
As Singapore continues to wage “war against diabetes”, there is still more that can be done to educate and empower individuals to take ownership of their own health and well-being.



By allowing individuals to conduct their own risk assessment for diabetes, the strain on our healthcare system may potentially be reduced while enabling early detection and/or intervention for high-risk individuals.

Simultaneously, we hope to encourage the general public to be more conscious of their own dietary habits by making it easier to identify healthier products with our app.

A Consideration for the Future



As we trained our model to **maximise sensitivity**, precision suffered slightly as a result. This means that individuals who may have **lower to no risk of diabetes may still be flagged** as being of higher risk.

For a disease detection model, having **low precision** may not be ideal due to ethical concerns related to **false diagnoses**.

While our model **does not claim to formally diagnose diabetes**, a **balance of sensitivity and precision** should ultimately be sought for the model to be serviceable to the general public.

Recommendations (Modelling)

Collect new data	Gather relevant data from **local participants/patients**. Some features might be worth exploring in the local context, e.g., ethnicity, family history. From the US data, we know which features to collect or focus on, thereby increasing the efficiency of the data collection process.
Improve feature selection and engineering	Based on the new data collected, and with additional features, we could potentially build a **more robust model** with increased possibilities in the feature engineering stage.
Address class imbalance	It is likely that the new data collected will still be imbalanced. Hence, we could explore **other sophisticated techniques to address class imbalance** beyond those we have already tried.

Recommendations (Optical Character Recognition)

Employ more advanced OCR models	Experiment with other robust OCR engines or models such as **OpenAI GPT-4**, **Google Cloud Vision API**, **Amazon Textract**, etc. Consider fine-tuning or training OCR models on **specific nutrition label datasets** to improve recognition accuracy for domain-specific content.
Integration of NLP tools	Use **Natural Language Processing (NLP) techniques** to analyse and validate extracted text based on semantic rules (e.g., expected nutrient formats, valid ingredient names).

LET'S BEAT DIABETES

Thanks

Any questions?