



# Unveiling Chronic Disease in Singaporean Lifestyle

Presented by:  
GA DSI- 42 | Team 6 | Chung Yau, Gilbert, Han Kiong, Zheng Gang



MINISTRY OF HEALTH  
SINGAPORE



# Agenda

**01**

**Chronic Disease in  
Singapore**

**Model Evaluation  
and Tuning**

**04**

**02**

**Problem  
Statement**

**Food Recommender  
on Healthy 365**

**05**

**03**

**Data Collection,  
Cleaning and EDA**

**Cost Benefit &  
Conclusion**

**06**

# Chronic Disease

Diseases that last **1 year or more**  
and require **ongoing medical**  
**attention** or **limit activities** of daily  
living or both

Example: Chronic Kidney Disease, Hyperlipidaemia, Diabetes,  
Hypertension, Depressive Disorder

[Source: <https://www.cdc.gov/chronicdisease/about/index.htm>]

World Health Organization (WHO)  
reveals that non-communicable  
diseases are **claiming around three**  
**quarters of all lives lost each year.**

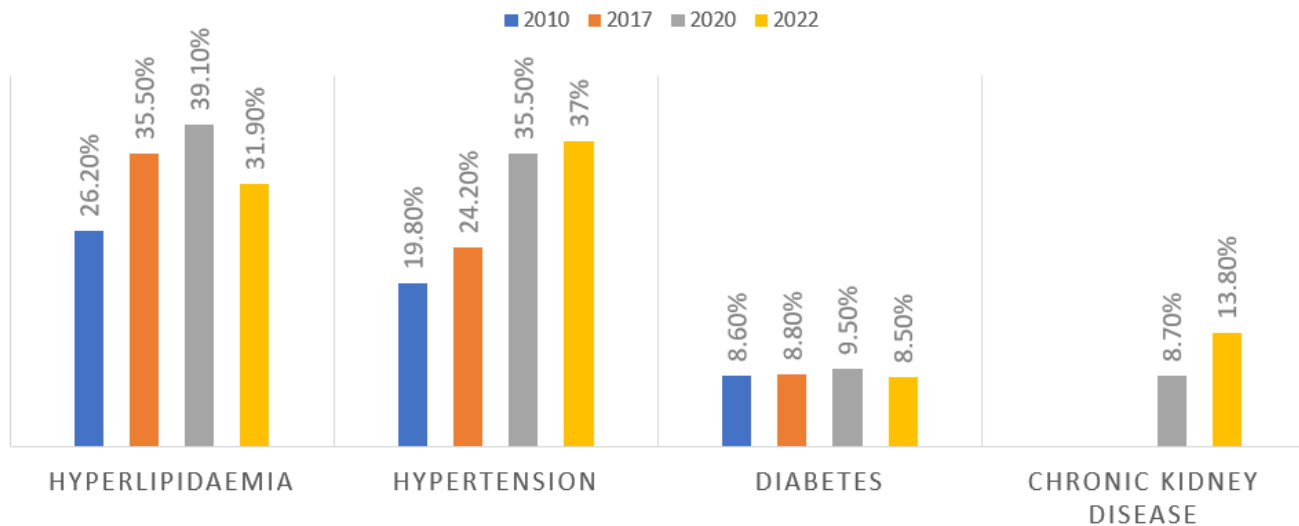
By around **2050**, chronic diseases  
will account for **86 per cent** of the **90**  
**million deaths yearly.**

(source: UN News on 19-May-2023  
<https://news.un.org/en/story/2023/05/1136832>)

# Prevalence of Chronic Disease in Singapore



**PREVALENCE OF HYPERLIPIDAEMIA, HYPERTENSION, DIABETES AND CHRONIC KIDNEY DISEASE IN SINGAPORE FROM 2010 TO 2022**



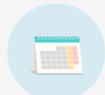
(Source: MOH, National Population Health Survey 2022)

# Risk Factors for Chronic Disease

## Chronic disease risk factors



### Ingrained risk factors



Age



Sex



Family genetics

### Behavioural risk factors



Poor nutrition



Lack of physical activity



Tobacco use



Excessive alcohol consumption

Four personal behaviors that can affect chronic diseases are:

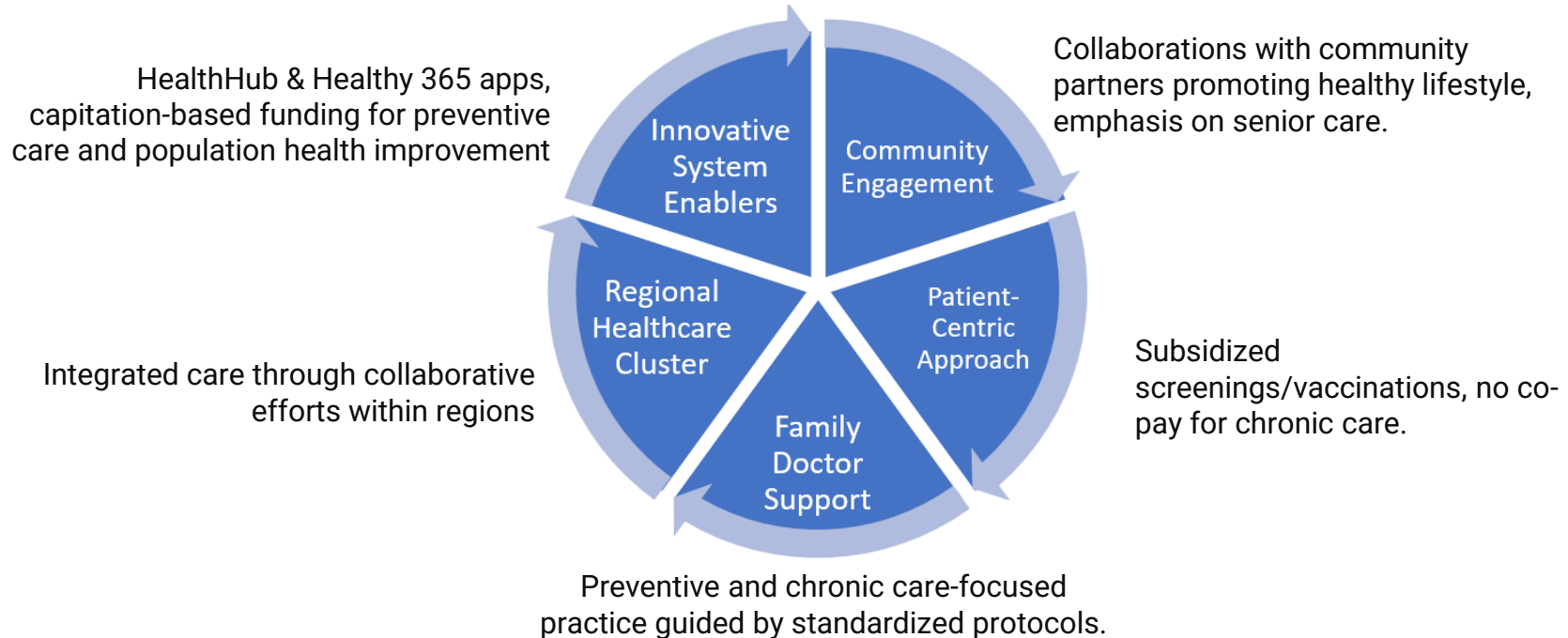
1. Lacking of **physical activities**,
2. Poor **nutrition**,
3. **Tobacco** use, and
4. Excessive **alcohol** use.

# Prevention is better than cure - Healthier SG

Shifting healthcare focus to **preventive care** is difficult but right thing to do, says PM Lee



# Healthier SG



# Problem Statement

In Singapore, the increasing prevalence of **chronic diseases** presents a pressing public health concern, underscoring the need for proactive intervention strategies.

How can we use a classification model to **identify** individuals at high risk for **chronic diseases** based on their behavioral habits ? By doing so, we can enable **early detection** and provide **recommendations**, fostering a proactive approach to preventing various chronic diseases.





Conrius, a 30-year-old auditor at KPMG, is mindful of the health risks associated with his **foodie hobby** amidst his strenuous work schedule.

He seek **personalized** guidance, not generic online information.

Conrius aims to assess his lifestyle's **impact** on potential chronic diseases.

He desires **tailored** strategies to cultivate **healthier eating habits** while still enjoying occasional indulgences in delicious cuisine.



## Conrius

30 year old Auditor at KPMG

**Goal:** Gain **early insights** into his **health risks**, particularly the impact of **his** lifestyle choices, especially the love for good food.

**Motivation:** Aims for a **healthy life** to enhance career success and longevity, without sacrificing much on his enjoyment for food.

**External Factors:** No homecook food, **always eat out**

# Problem Statement

How can we **identify** individuals at high risk for chronic diseases based on their lifestyle data? By doing so, we can enable early detection and provide **recommendations**, fostering a proactive approach to preventing various chronic diseases.



## Three Steps Approach

Step 1: **Understanding** relationship between lifestyle choices and chronic diseases

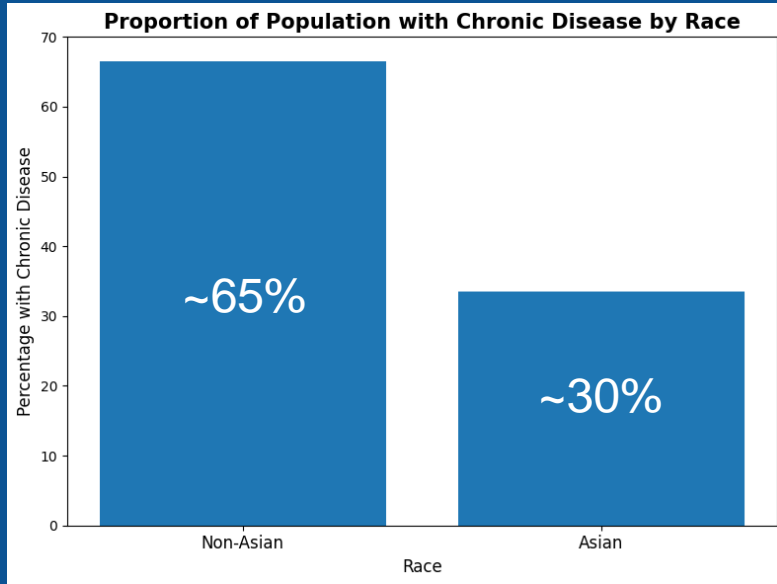
Step 2: **Identifying** individuals with high risk

Step 3: Providing **recommendations** (food use-case)



Step 1:  
**Understanding** relationship  
between lifestyle choices  
and chronic diseases

# Our Data

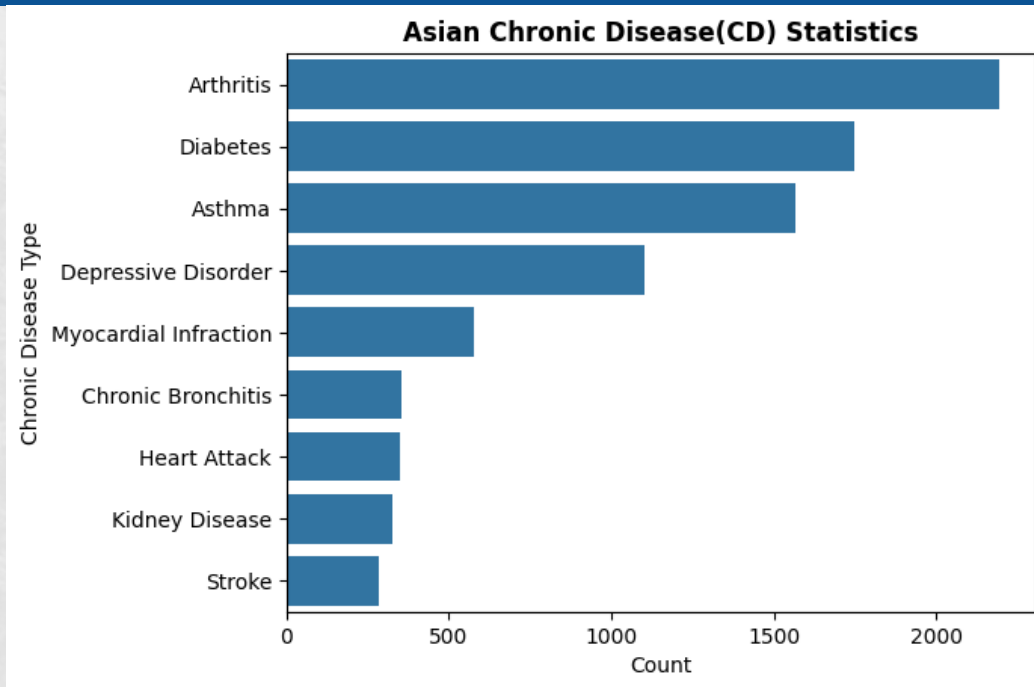


Questionnaire data with > 9k response with respondents of Asian descent.

Questionnaire includes questions about:

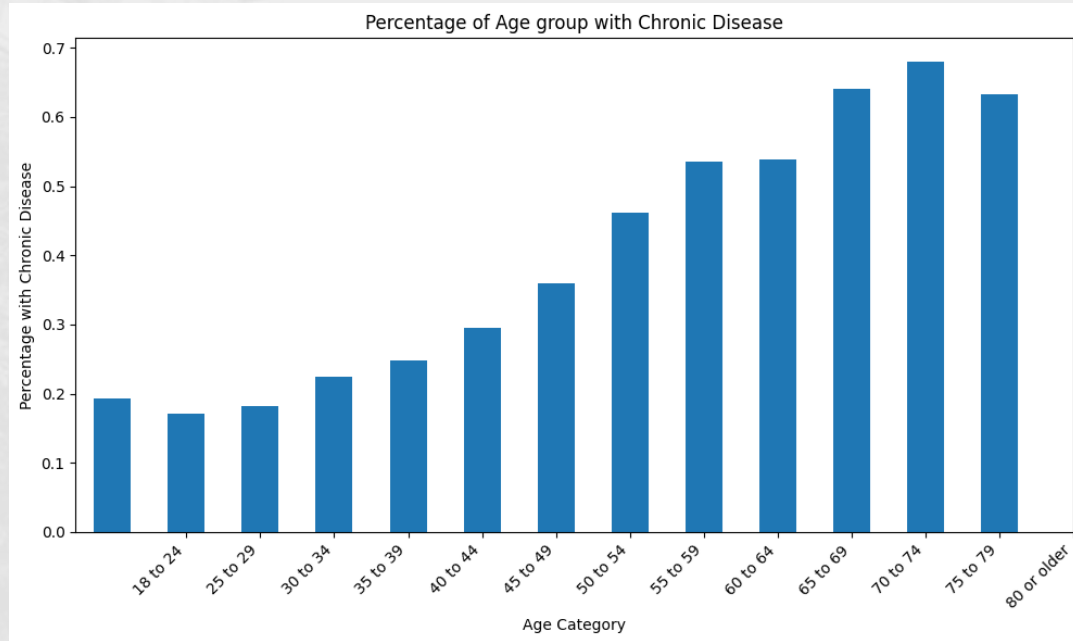
1. 4 personal behaviors congruent with previously mentioned:
  - a. Physical Activities
  - b. Nutrition and Diet
  - c. Alcohol Use
  - d. Tobacco
1. Whether an individual is diagnosed with a chronic disease by a medical professional.

# Breakdown of Chronic Disease



**Arthritis, Diabetes and Asthma** ranked top 3 in terms of frequency for the chronic disease troubling Asian.

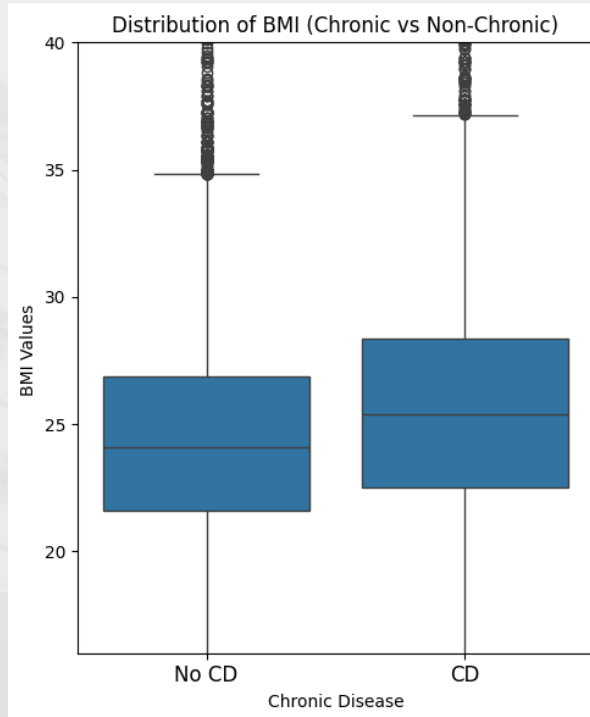
# Chronic Disease (Age Group)



There is an increasing trend of the proportion of people with chronic disease as they age stating the obvious



# Chronic Disease (BMI)

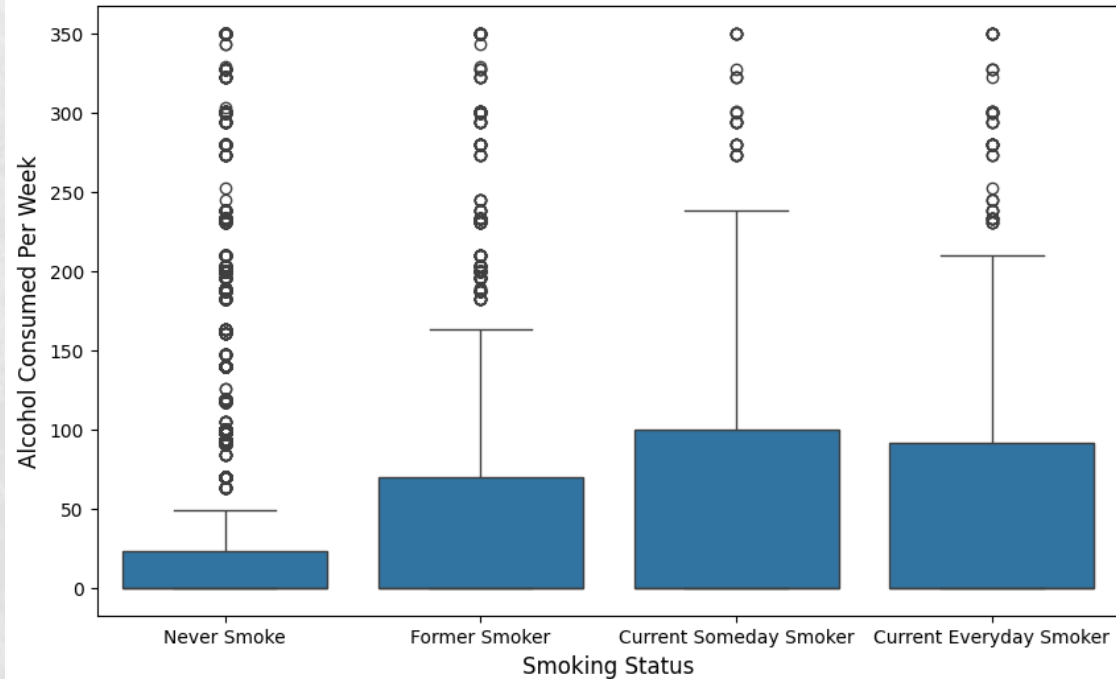


Slight increase in BMI for individuals diagnosed with Chronic Diseases.

We also observe the same across all different chronic illness

# Bad Habits Come in Tandem

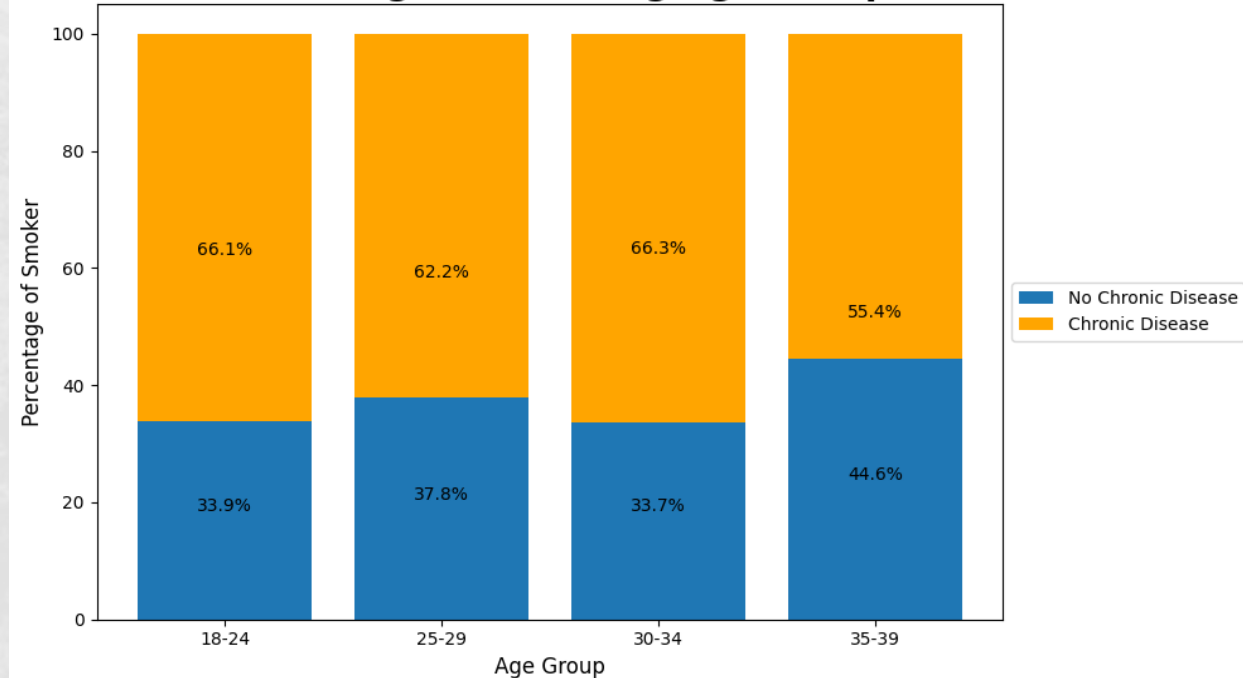
**Drink vs Smoke**



Smoking individuals tends to consume more alcohol

# Chronic Disease (Smoking)

**Percentage of Smoking Age Group**



Smoking young group (18-39)  
have higher percentage of  
having Chronic Disease



Step 2:  
**Identifying** individuals with  
high risk

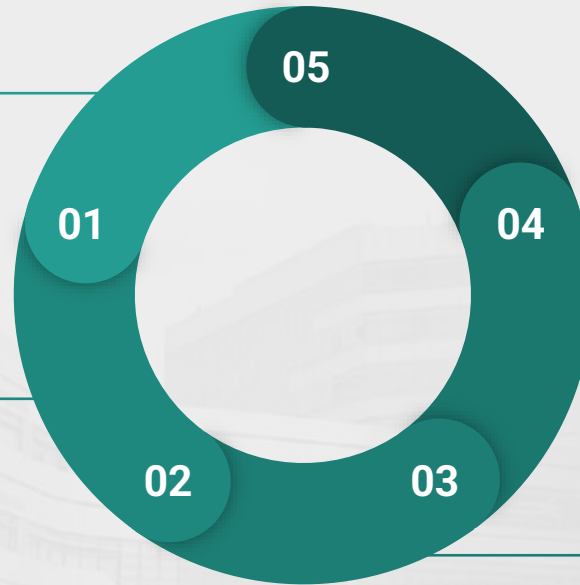
# Model Workflow

## Data Collection

US Behavioral Risk  
Factor Surveillance  
System (BRFSS) Survey  
- 2013 & 2015

## Data Preprocessing

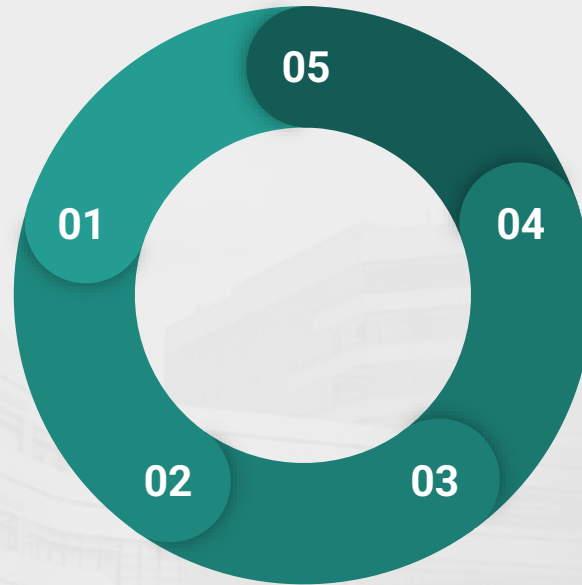
1. Removal of irrelevant features,
2. Dropping or imputing missing values



## Exploratory Data Analysis

1. Post cleaning to uncover patterns and insights,
2. Understand relationships between lifestyle choices and risk of chronic diseases,
3. Identify individuals at high risk based on behavioural patterns

# Model Workflow



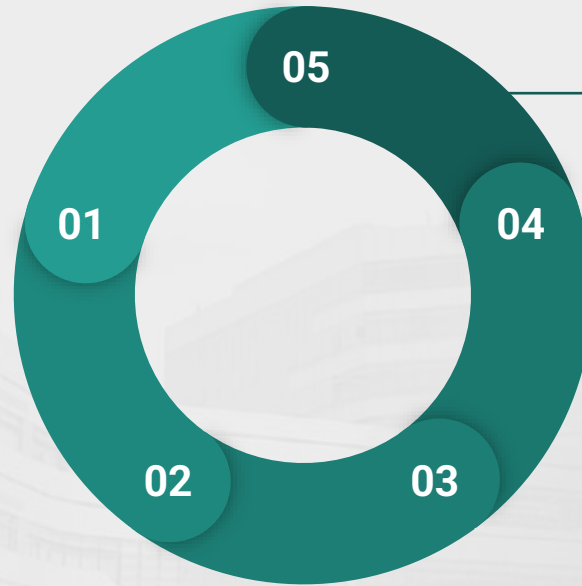
## Feature Engineering and Modelling

### Feature Engineering:

1. Generate polynomial and interaction features,
2. Standardized features using StandardScaler,
3. PCA for dimensionality reduction (n\_components = 30)

Balance classes for chronic diseases using RandomUnderSampler and ADASYN (oversampling)

# Model Workflow



## Model Selection and Hyperparameter Tuning

Model for tuning: Logistic Regression and Support Vector Classifier

1. Define pipeline for transformers and estimator
2. Define hyperparameters
3. Apply GridSearch to hyperparameter tune



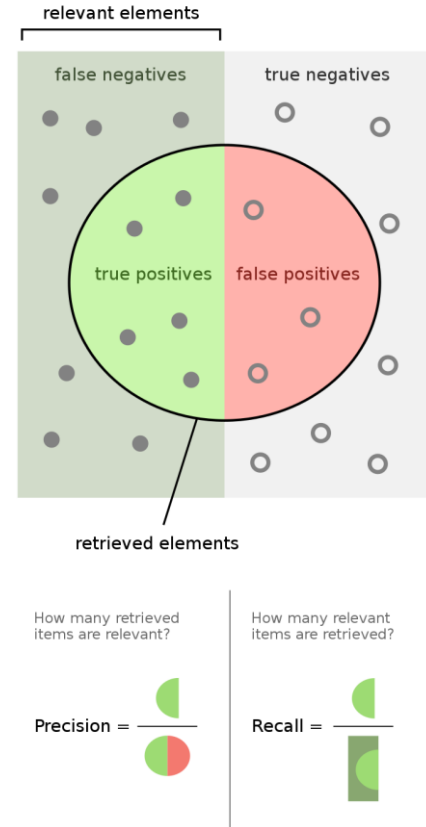
# Maximizing F1-Score

1. Balanced Trade Off - both false positive and false negative are equally undesirable.

- a. False positive (wrongly identify an individual as high risk) - false alarm causing unnecessary stress and testing or treatment
- b. False negative (wrongly identify an individual as low risk) - unable to provide timely intervention.

1. Sensitivity to Class Imbalance.

- a. Approx 30% of Asian is medically diagnosed with a chronic disease



# Baseline Model Evaluation - Train and Test Scores

Model	Accuracy		Precision		Recall		F1 Score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	69.61%	72.28%	72.68%	58.20%	62.84%	65.48%	67.40%	61.62%
Support Vector Classifier	71.86%	71.19%	74.77%	56.60%	65.99%	65.39%	70.10%	60.68%
Random Forest	99.84%	65.01%	99.89%	48.92%	99.79%	66.02%	99.84%	56.20%
Gradient Boosting Classifier	72.98%	70.32%	75.24%	55.26%	68.50%	66.67%	71.71%	60.43%
XGBoost	96.77%	65.66%	98.69%	49.62%	94.79%	65.30%	96.70%	56.39%

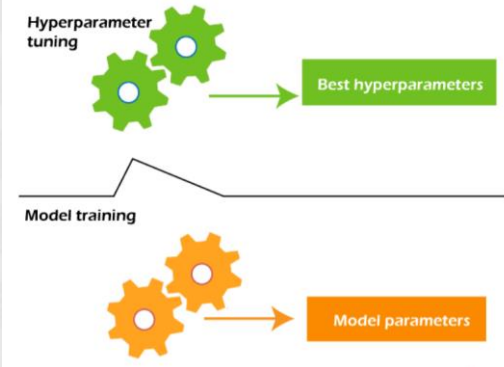
# Model Evaluation - Hyperparameter Tuning

Metric	Support Vector Classifier		Logistic Regression		We chose Logistic Regression due to:
	Train	Test	Train	Test	
Accuracy	68.34%	68.12%	71.80%	72.00%	Better Accuracy
Precision	52.62%	52.34%	57.87%	57.97%	Better Precision
Recall	68.48%	69.50%	62.56%	63.74%	
F1 Score	59.51%	59.71%	60.13%	60.72%	Better F1 score

# Logistic Regression - Before vs After Tuning

Metric	Before Tuning		After Tuning	
	Train	Test	Train	Test
Accuracy	69.61%	72.28%	71.80%	72.00%
Precision	72.68%	58.20%	57.87%	57.97%
Recall	62.84%	65.48%	62.56%	63.74%
F1 Score	67.40%	61.62%	60.13%	60.72%

Reduced overfitting!!





## Conrius

30 year old Data Analyst

**“Now that I know I am at risk, can you also recommend me some well balanced dishes that is suited for my lifestyle?”**



## Step 3: Providing **recommendations** (food)



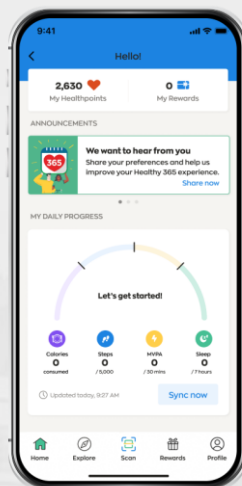


# Healthy 365 - Encourage Healthier Lifestyles among Singaporean

**App to monitor physical activity,  
food intake, and overall health**



**Gamified health challenges to  
motivate active and healthier  
lifestyles**



**Rewarding health goal through  
Health Points and promotional  
vouchers**



**Personalized health info**







# Healthy 365 - Encourage Healthier Lifestyles among Singaporean

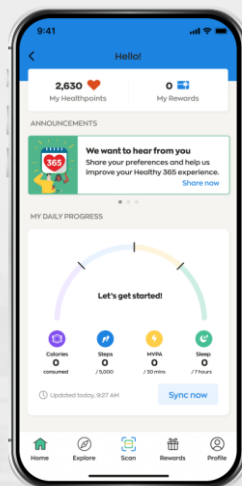
App to monitor physical activity,  
food intake, and overall health



Gamified health challenges to  
motivate active and healthier  
lifestyles



Chronic disease risk calculator  
through lifestyle questionnaire



Rewarding health goal through  
Health Points and promotional  
vouchers



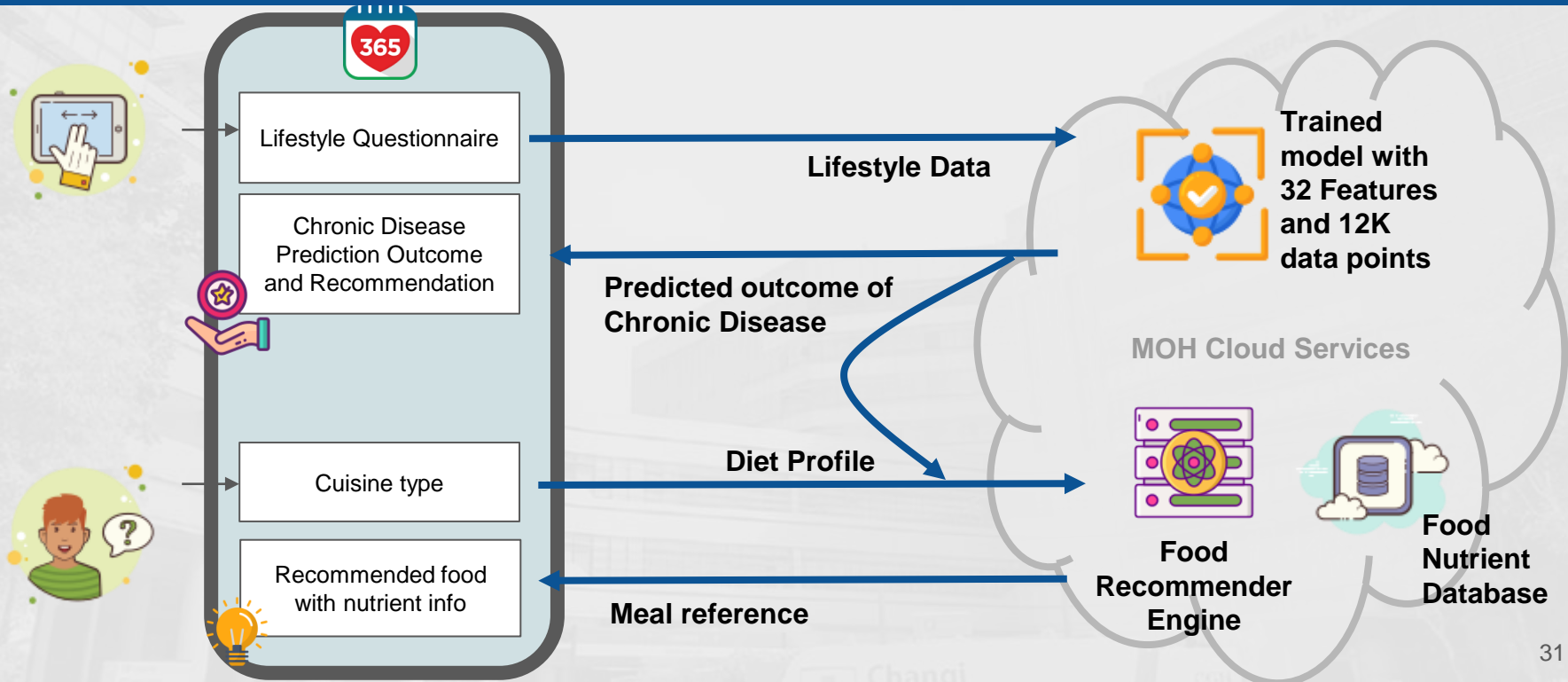
Personalized health info



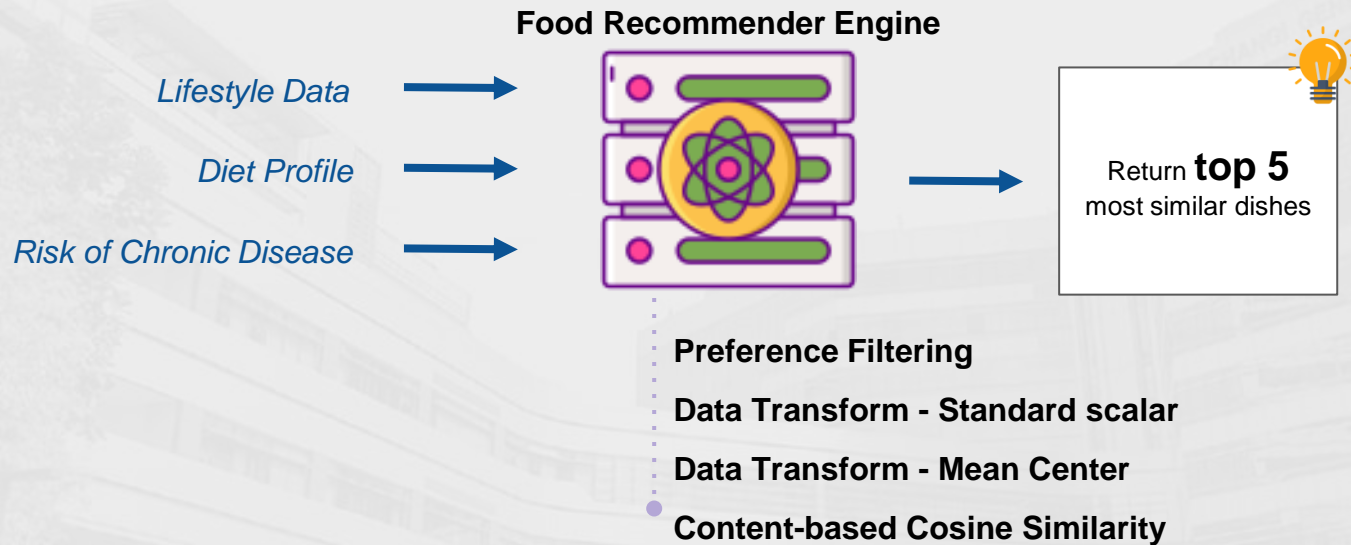
Health enhanced meal  
recommendation



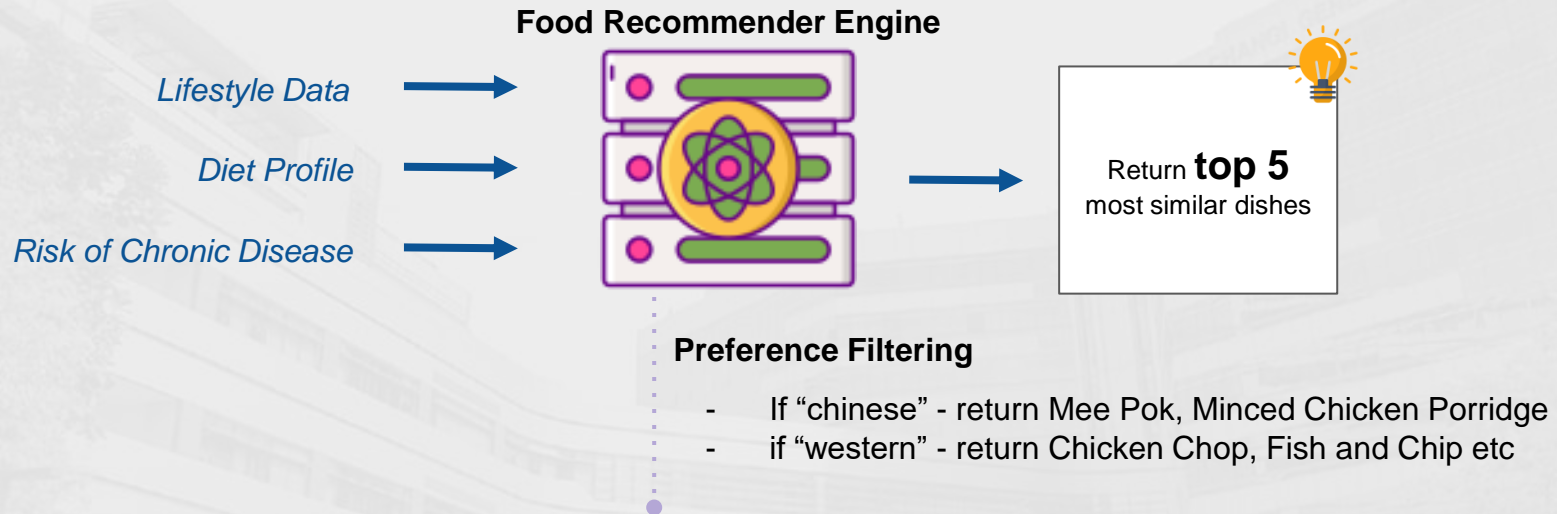
# Recommender Engine in Action



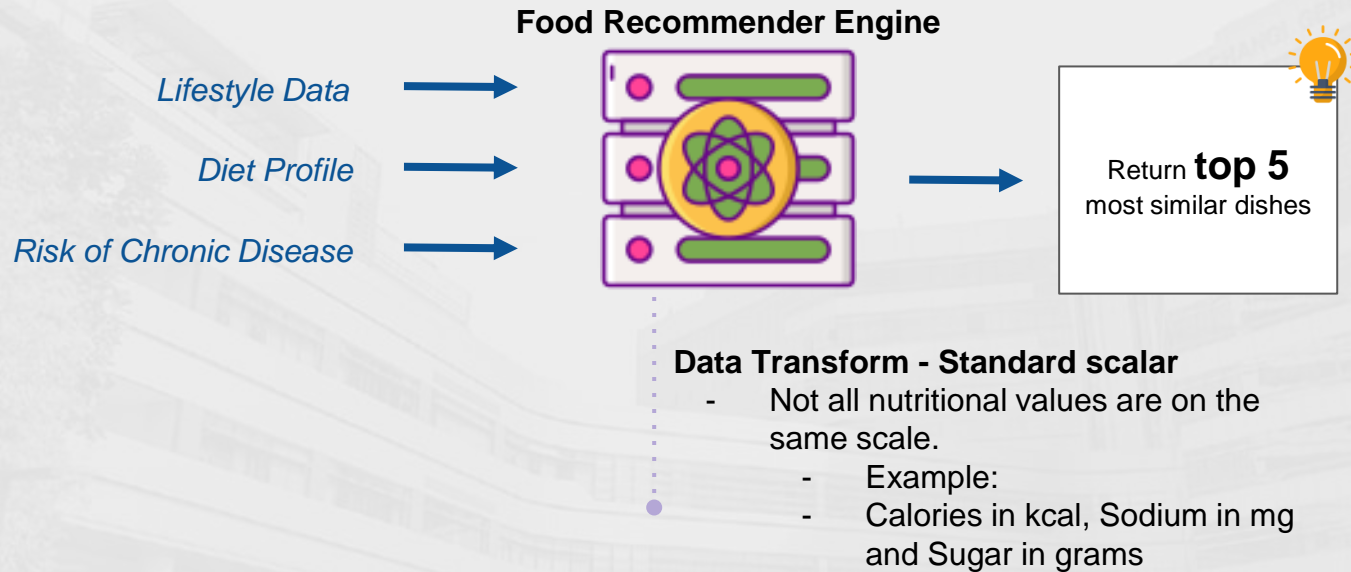
# Recommender Engine in Action



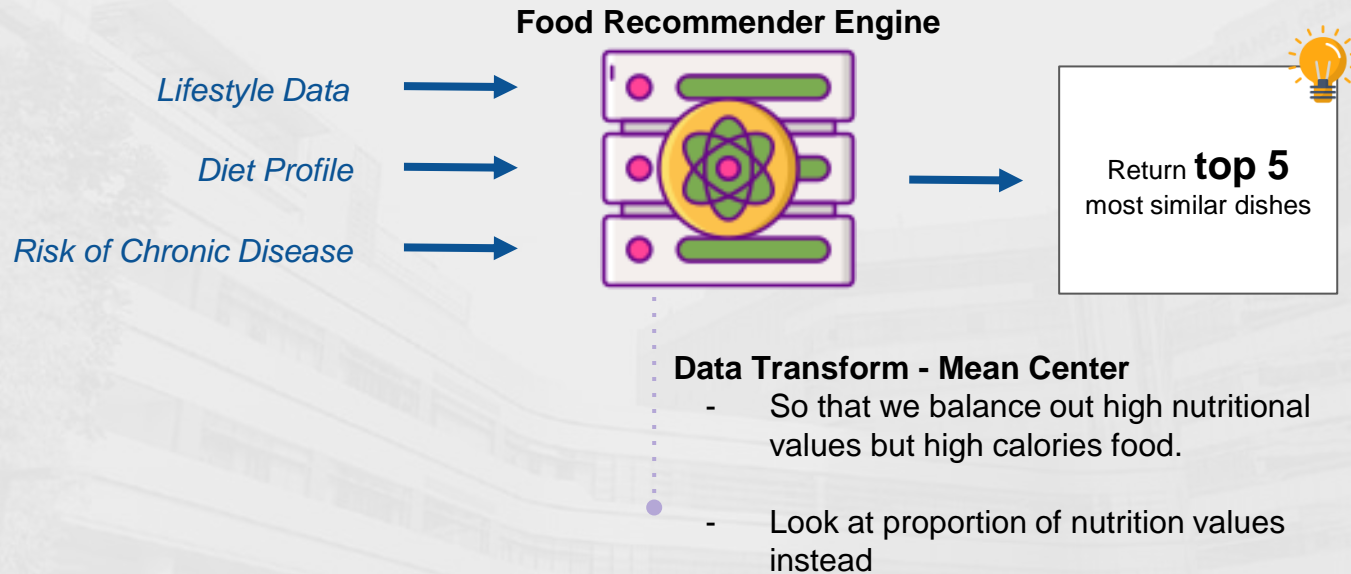
# Recommender Engine in Action



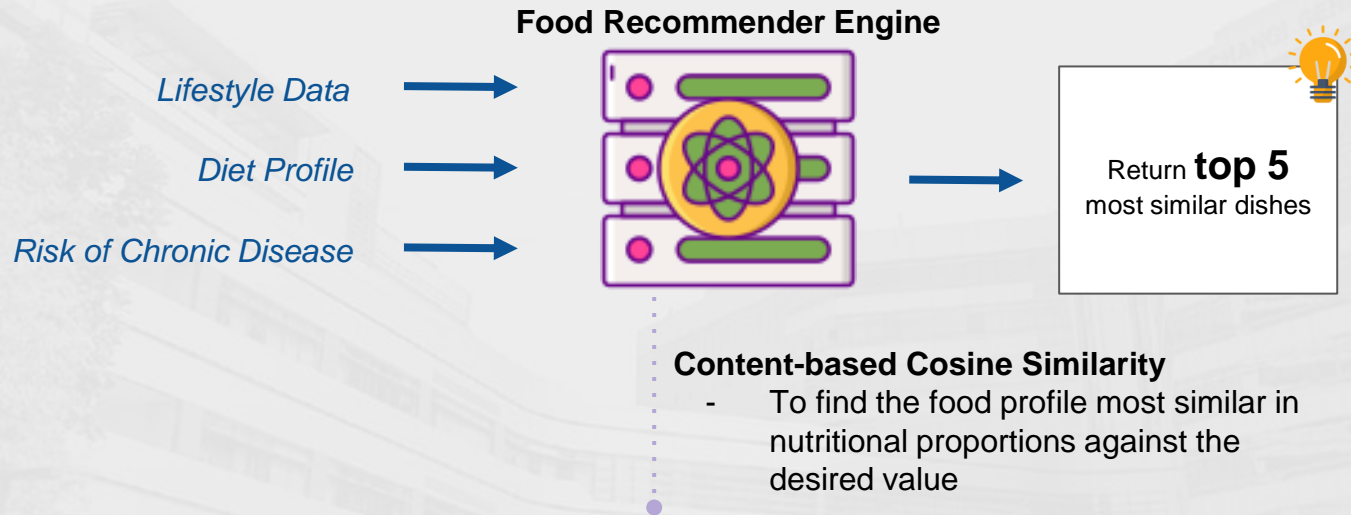
# Recommender Engine in Action



# Recommender Engine in Action

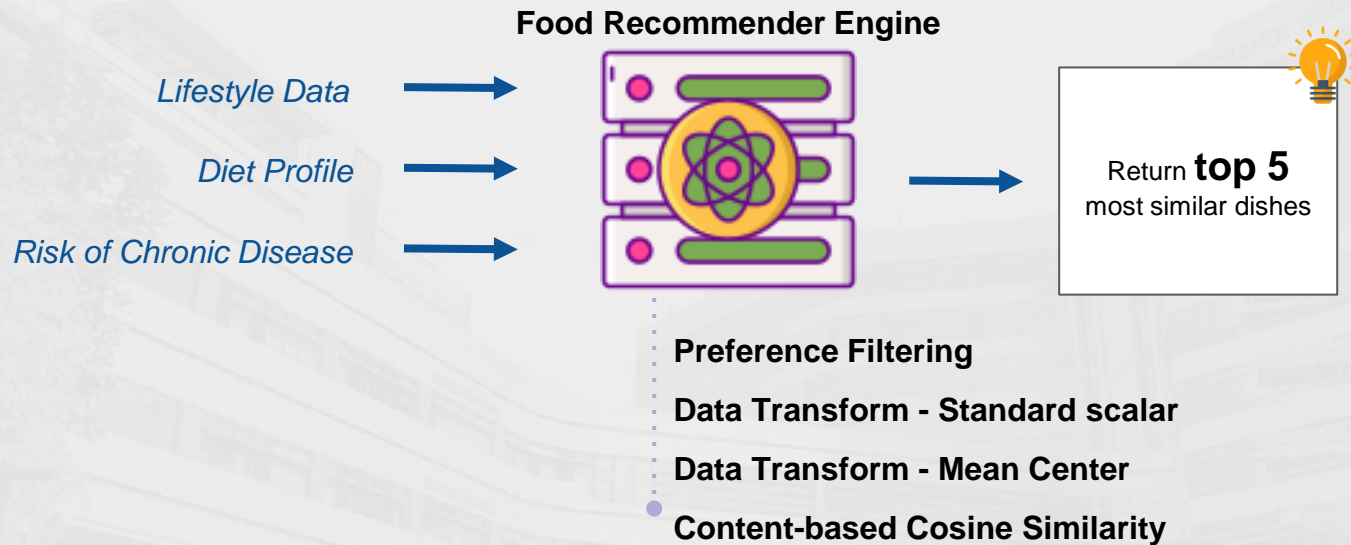


# Recommender Engine in Action

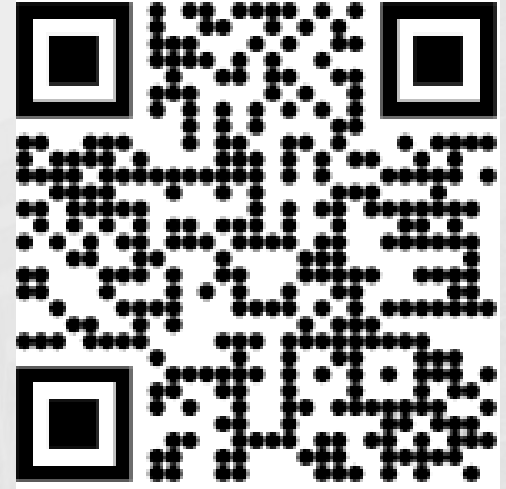




# Recommender Engine in Action



# App Demo



Healthy365 - Chroniscope

*Scan me to try out !!*



# Cost-Benefit Analysis

# Recommender Feature - Cost

Label	2024	2025	2026	2027
% of chronic disease singapore (multimorbidity) <sup>[1]</sup> (A)	16.30%	16.30%	16.30%	16.30%
Users on healthy 365 (based on 2017) (B)	84000	88200	92610	97241
<b>% Users on health hub with multimorbidity (C = A*B)</b>	<b>13692</b>	<b>14377</b>	<b>15095</b>	<b>15850</b>
Marketing cost (per user level) (D)	\$70.00	\$72.45	\$74.99	\$77.61
<b>Total Marketing spend based on 1% conversion (E = D*1%*C)</b>	<b>\$9,584.40</b>	<b>\$104,158.47</b>	<b>\$113,194.21</b>	<b>\$123,013.81</b>
Data gathering effort (model training) (F)	\$50,000.00	\$20,000.00	\$20,000.00	\$20,000.00
Data gathering effort (recommender) (G)	\$50,000.00	\$10,000.00	\$10,000.00	\$10,000.00
App development (inclusive of backend works) (H)	\$500,000.00	\$51,750.00	\$53,561.25	\$55,435.89
App maintenance (I)	\$129,600.00	\$134,136.00	\$138,830.76	\$143,689.84
<b>Assuming 1% conversion</b>	<b>Projected Total Cost</b>			
<b>Total Spend (E + F + G + H + I)</b>	<b>\$739,184.40</b>	<b>\$320,044.47</b>	<b>\$335,586.22</b>	<b>\$352,139.54</b>

We are aggressive with our cost.

## Assumptions:

Data gathering effort - Chargeback by HPB for data gathering efforts

App Development - Aggressive cost for a simple feature given the complexity of government projects

App Development - 10% allocated to change request subsequent years

App maintenance - 1 x full stack engineer catered with monthly salary 8k

Marketing Cost - Taking mean cost of an intervention program per pax

[1] MOH, National Population Health Survey 2022

Assuming 3.5% increase in cost due to inflation. [Link](#) to detailed analysis

# Recommender Feature - Benefit

Label	2024	2025	2026	2027
Estimated annual societal cost per pax <sup>[2]</sup> (A)	\$15,148.00	\$15,678.18	\$16,226.92	\$16,794.86
Per Capita Govt Health Expenditure <sup>[3]</sup> (B)	\$2,674.00	\$2,767.59	\$2,864.46	\$2,964.71
<b>Net Chronic Disease Spending per pax (C = A - B)</b>	<b>\$12,474.00</b>	<b>\$12,910.59</b>	<b>\$13,362.46</b>	<b>\$13,830.15</b>
% of chronic disease singapore (multimorbidity) <sup>[1]</sup> (D)	16.30%	16.30%	16.30%	16.30%
Users on healthy 365 (based on 2017) (E)	84000	88200	92610	97241
<b>% Users on healthhub with multimorbidity (F = D*E)</b>	<b>13692</b>	<b>14377</b>	<b>15095</b>	<b>15850</b>
<b>Assuming 1% conversion</b>	<b>Projected Benefits</b>			
<b>Benefit = 1% * F * C</b>	<b>\$1,707,940</b>	<b>\$1,856,104</b>	<b>\$2,017,121</b>	<b>\$2,192,106</b>

We are conservative in the benefits.

## Assumptions:

% of people with chronic disease remains constant due to governments proactive intervention

Users on Healthhub takes a conservative growth of 5%

Per successful prevention of one user with multi-morbidity reduces the cost for a person to the baseline per capita govt health expenditure

# Healthy 365 App - Cost Benefit Analysis

Year	1% Conversion	2% Conversion	3% Conversion
Year 1	\$968,756	\$2,667,111	\$4,365,467
Year 2	\$1,629,802	\$3,477,989	\$5,323,727
Year 3	\$1,783,409	\$3,794,329	\$5,800,239

[Link](#) to detailed analysis





## Conclusion and Recommendations



# Conclusion

## Effective Prediction with Logistic Regression

Achieves a balanced F1 score, highlighting precision and recall in predicting chronic disease risks based on lifestyle data, avoiding overfitting.

## Customized Lifestyle Recommendations

Utilizes predictive insights to tailor lifestyle adjustments, exemplified by the integration with a food recommender app for diet optimization.

## Importance of Early Detection and Prevention

Emphasizes the critical role of early detection and promoting healthy lifestyles in reducing the incidence and impact of chronic diseases.

# Limitation and Recommendation

## Inadequate Lifestyle Factor Consideration:

- **Limitation:** The existing dataset lacks other chronic disease risk factors, such as sleep patterns, mental health, and genetic predispositions.
- **Recommendation:** Data collection can encompass a wider range of lifestyle factors. Integrating more diverse data points like sleep quality, mental health statuses, and genetic markers can sharpen the model's predictive accuracy and enable the provision of more holistic lifestyle recommendations.

# Limitation and Recommendation

## Nutritional Information Gaps:

- **Limitation:** The model currently does not account for critical nutritional details like Glycemic Index and Dietary Fiber, which are significant for blood sugar management.
- **Recommendation:** Proactively expand data collection through online resources or partnerships with research labs to include these nutritional factors, enhancing the model's utility and accuracy in providing dietary advice



Thank you