Unveiling Chronic Disease in Singaporean Lifestyle

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

MINISTRY OF HEALTH SINGAPORE

Health Promotion Board

Healthier SG

Agenda

01 Chronic Disease in Singapore

02 Problem Statement

03 Data Collection, Cleaning and EDA

04 Model Evaluation and Tuning

05 Food Recommender on Healthy 365

06 Cost Benefit & Conclusion

Chronic Disease

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

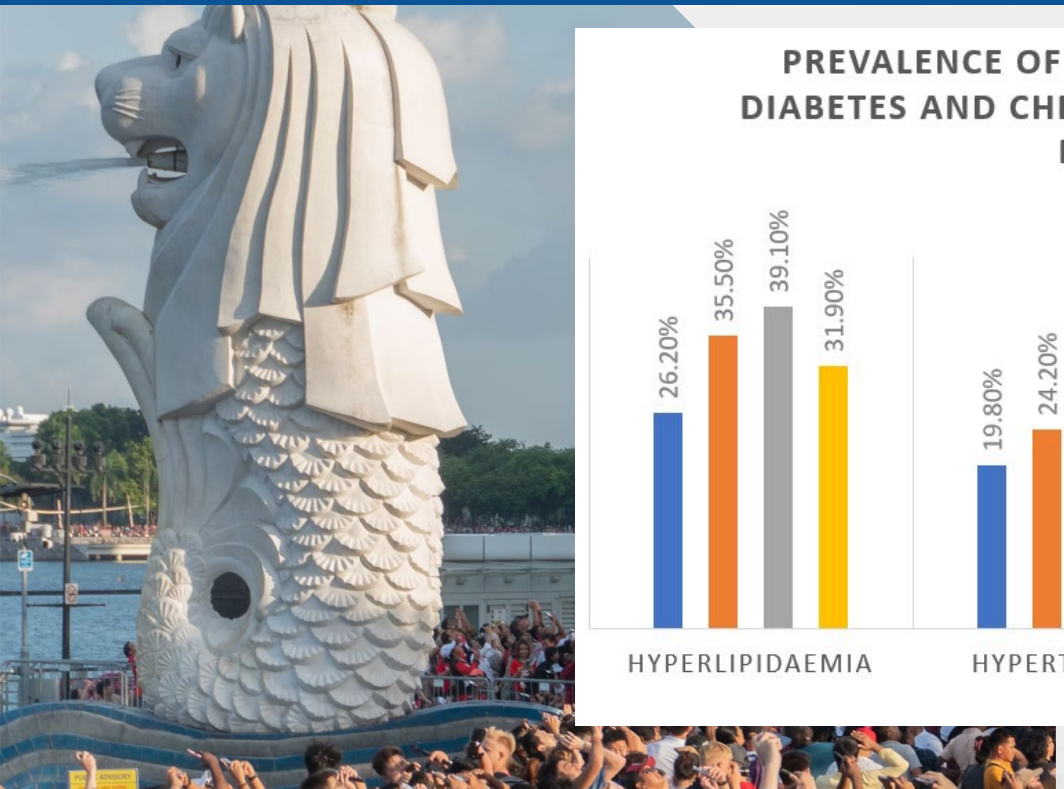
Example: Chronic Kidney Disease, Hyperlipidaemia, Diabetes, Hypertension, Depressive Disorder

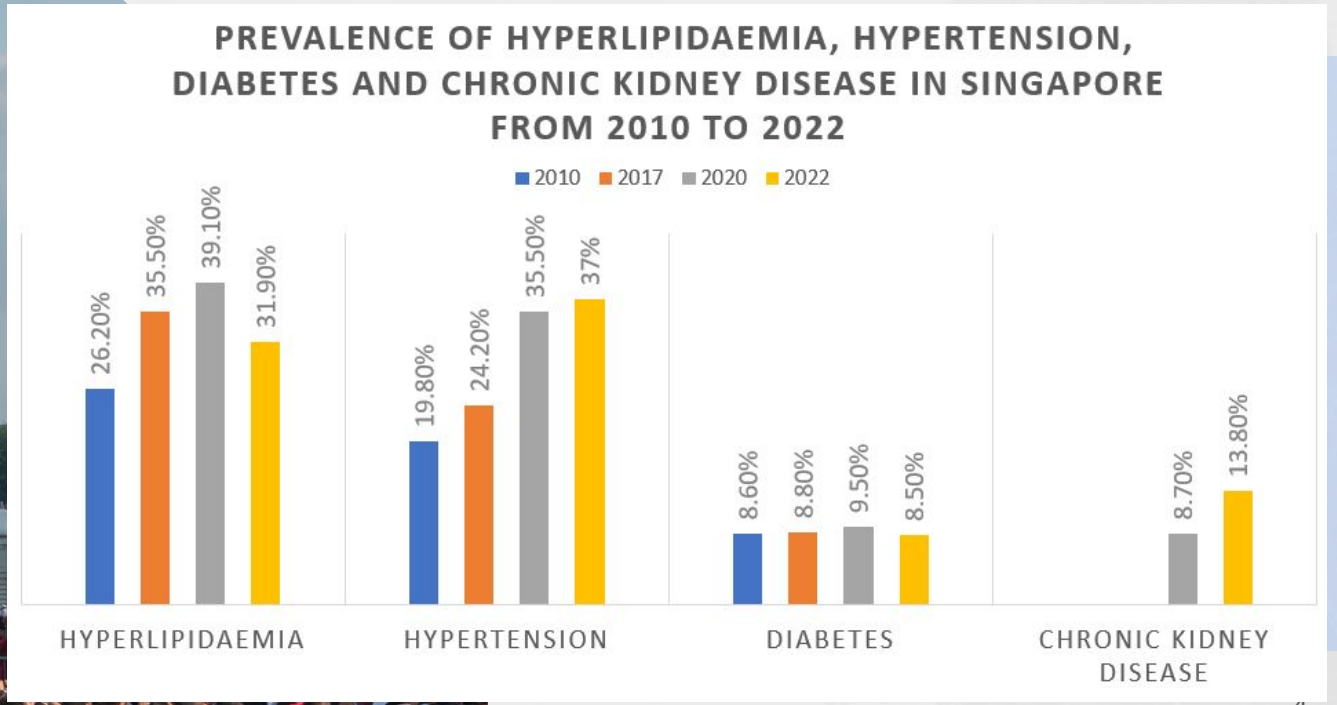
[Source: https://www.cdc.gov/chronicdisease/about/index.htm]

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)

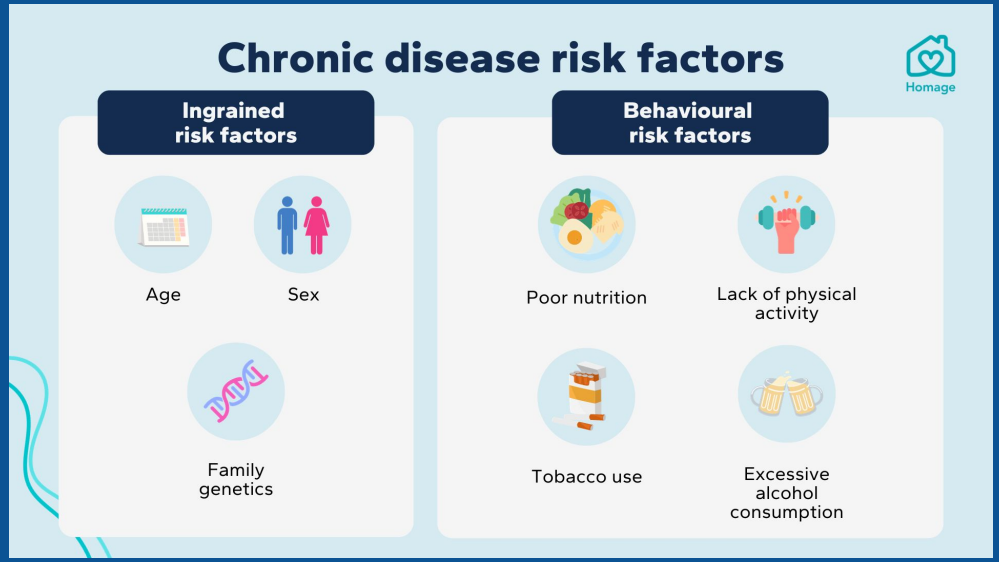
Prevalence of Chronic Disease in Singapore





(Source: MOH, National Population Health Survey 2022)

Risk Factors for Chronic Disease



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.

(Source: https://www.cdc.gov/nceh/tracking/topics/LifestyleRiskFactors.htm)

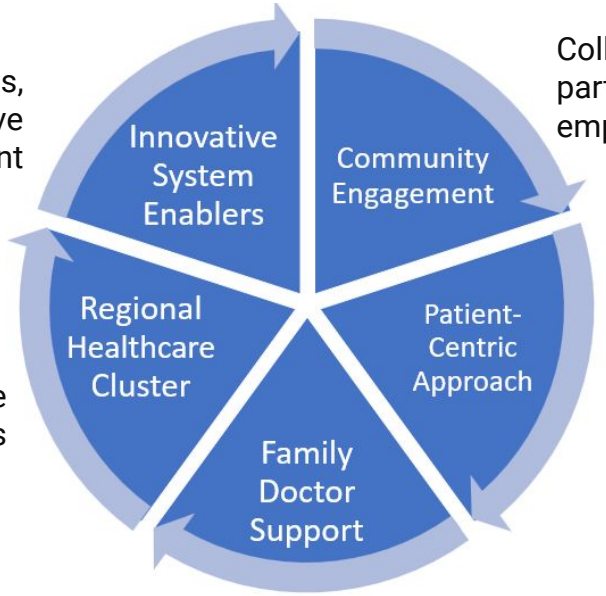
Prevention is better than cure - Healthier SG

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

(Source: The Straits Times, 25-Apr-2022)

Healthier SG

HealthHub & Healthy 365 apps, capitation-based funding for preventive care and population health improvement



Collaborations with community partners promoting healthy lifestyle, emphasis on senior care.

Integrated care through collaborative efforts within regions

Subsidized screenings/vaccinations, no co-pay for chronic care..

Preventive and chronic care-focused practice guided by standardized protocols.

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

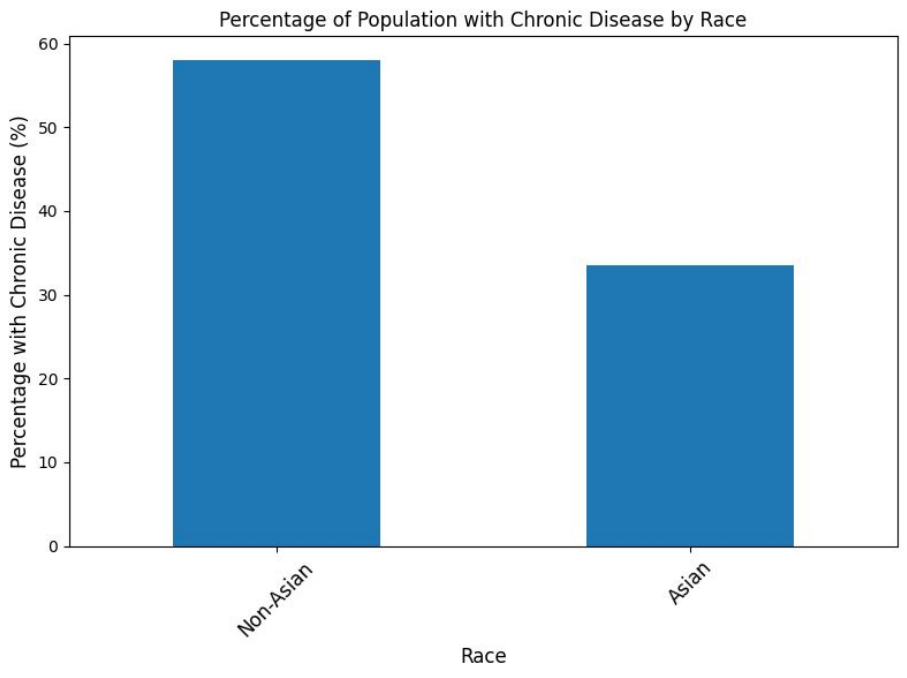
Questionnaire data with > 9k response with respondents of Asian descent.

Questionnaire includes questions about:

1. **4 personal behaviors congruent with previously mentioned:**
    a. Physical Activities
    b. Nutrition and Diet
    c. Alcohol Use
    d. Tobacco

2. **Whether an individual is diagnosed with a chronic disease by a medical professional.**

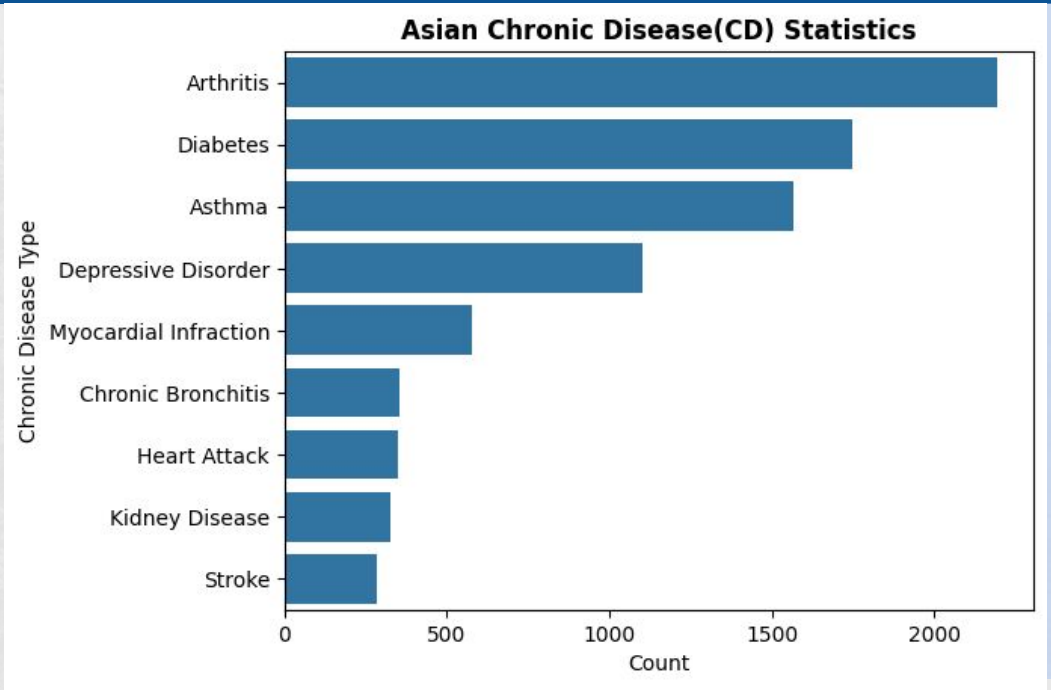
Why we do not include other races?



30% of Asian have chronic disease vs 60% for non-Asian

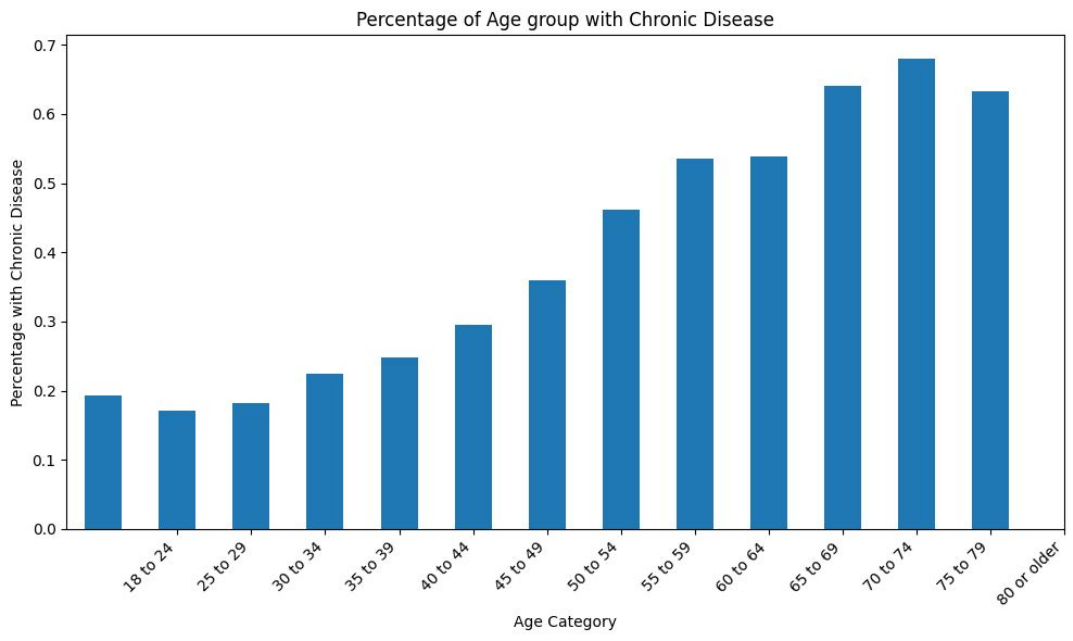
Action: Exclude non-Asian for our conversations and upcoming analysis.

Breakdown of Chronic Disease



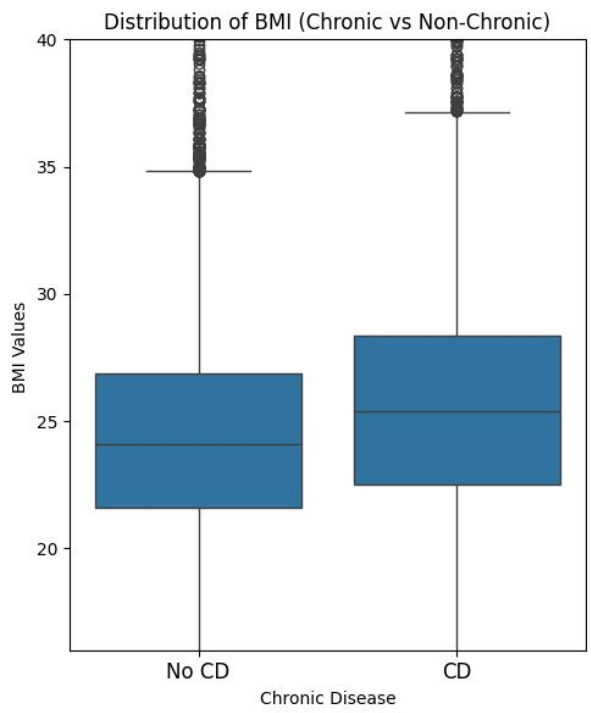
Arthritis, Diabetes and **Asthma** ranked top 3 in terms of frequency for the chronic disease troubling Asian.

Chronic Disease (Age Group)



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

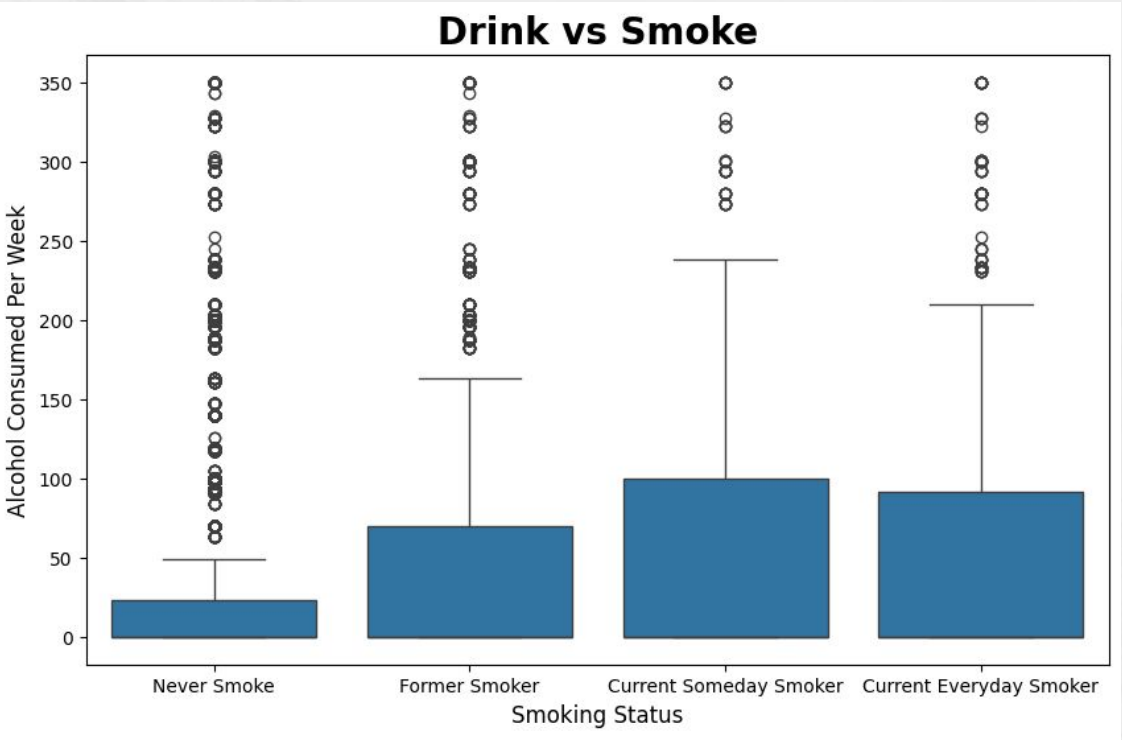
Chronic Disease (BMI)



Slight increase in BMI for individuals diagnosed with Chronic Diseases.

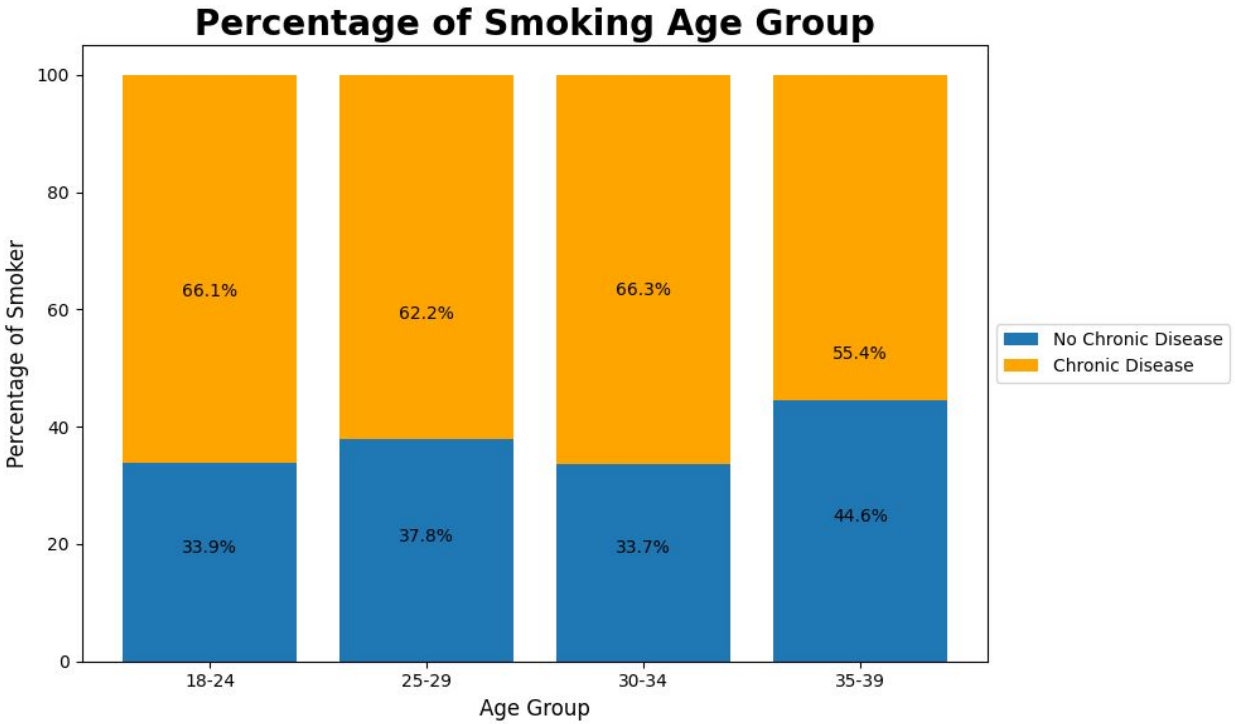
We also observe the same across all different chronic illness

Bad Habits Come in Tandem

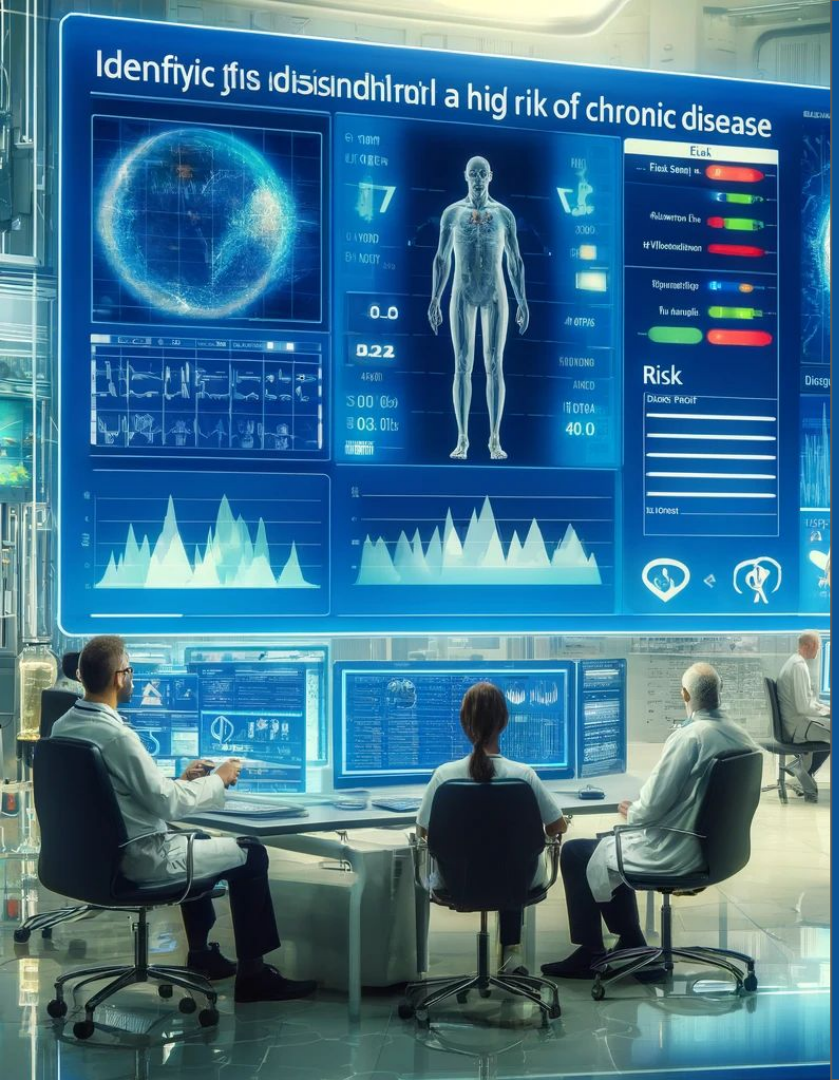


Smoking individuals tends to consume more alcohol

Chronic Disease (Smoking)



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

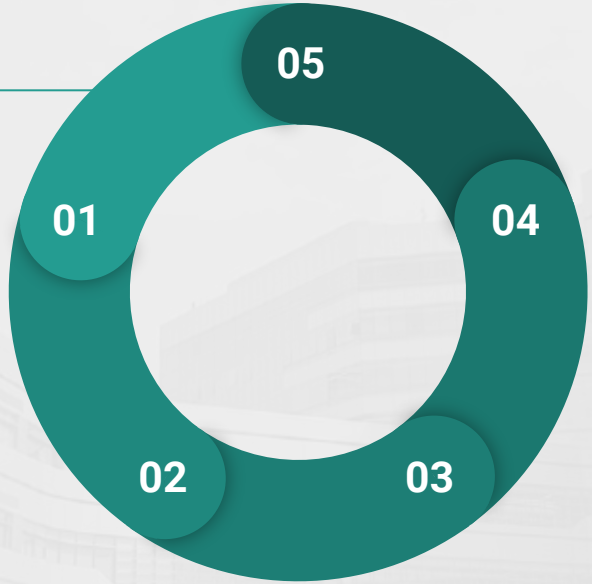


Step 2:
Identifying individuals with high risk

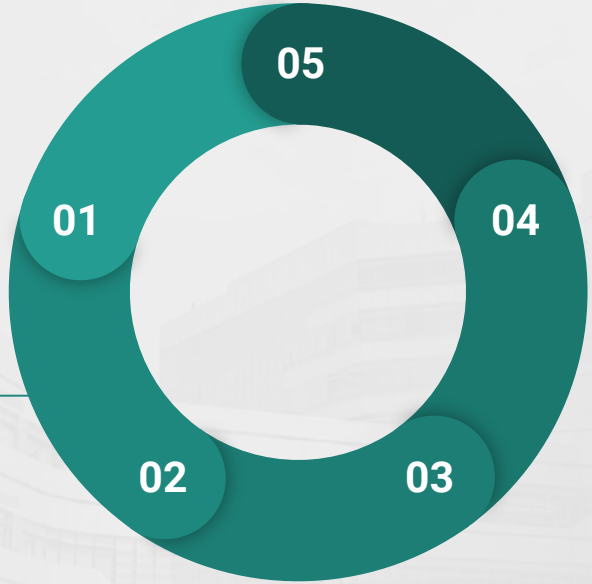
Model Workflow

Data Collection

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



Model Workflow



Data Preprocessing

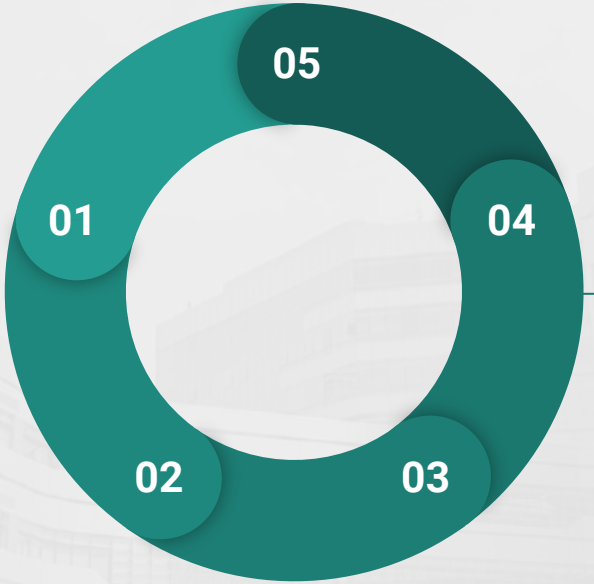
1. Removal of irrelevant features,
2. Dropping or imputing missing values

Model Workflow

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

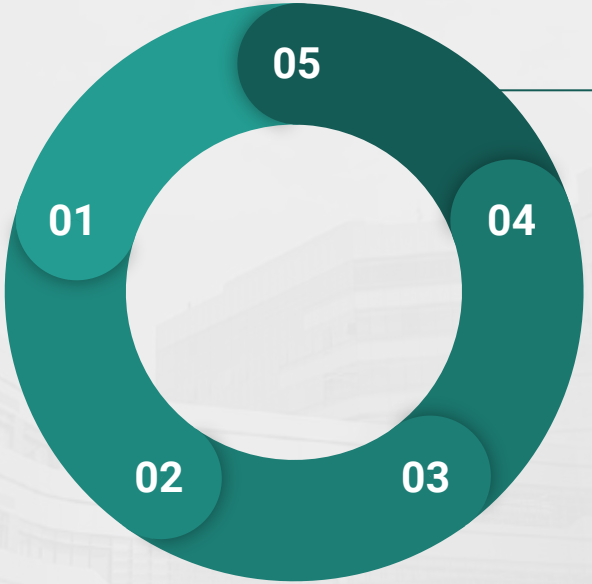


Feature Engineering and Modelling

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

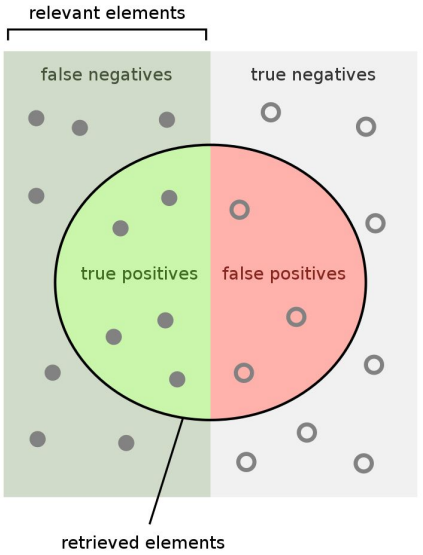
Model for tuning: Logistic Regression and Support Vector Classifier

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.

2. 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

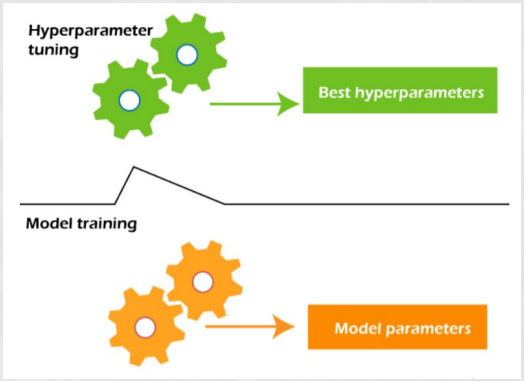
Metric	Support Vector Classifier		Logistic Regression		
	Train	Test	Train	Test	
Accuracy	68.34%	68.12%	71.80%	72.00%	Better Accuracy
Precision	52.62%	52.34%	57.87%	57.97%	Better Precision
Recall	68.48%	69.50%	62.56%	63.74%	
F1 Score	59.51%	59.71%	60.13%	60.72%	Better F1 score

We chose Logistic Regression due to:

Logistic Regression - Before vs After Tuning

Metric	Before Tuning			After Tuning		
	Train		Test	Train		Test
Accuracy	69.61%	2.67%	72.28%	71.80%	0.20%	72.00%
Precision	72.68%	14.48%	58.20%	57.87%	0.10%	57.97%
Recall	62.84%	2.64%	65.48%	62.56%	1.18%	63.74%
F1 Score	67.40%	5.78%	61.62%	60.13%	0.59%	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)

How it works?

Provides lifestyle information and preference

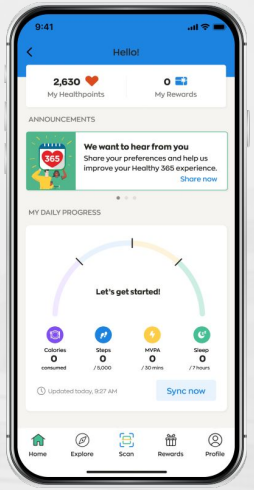
Identify dishes most similar to ideal nutrition needs

Provides top 5 food recommendation based on user's profile

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

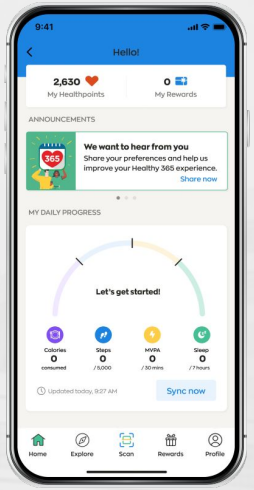


Rewarding health goal through Health Points and promotional vouchers

Personalized health info

Healthy 365 -
Encourage Healthier Lifestyles among Singaporean

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



Rewarding health goal through Health Points and promotional vouchers

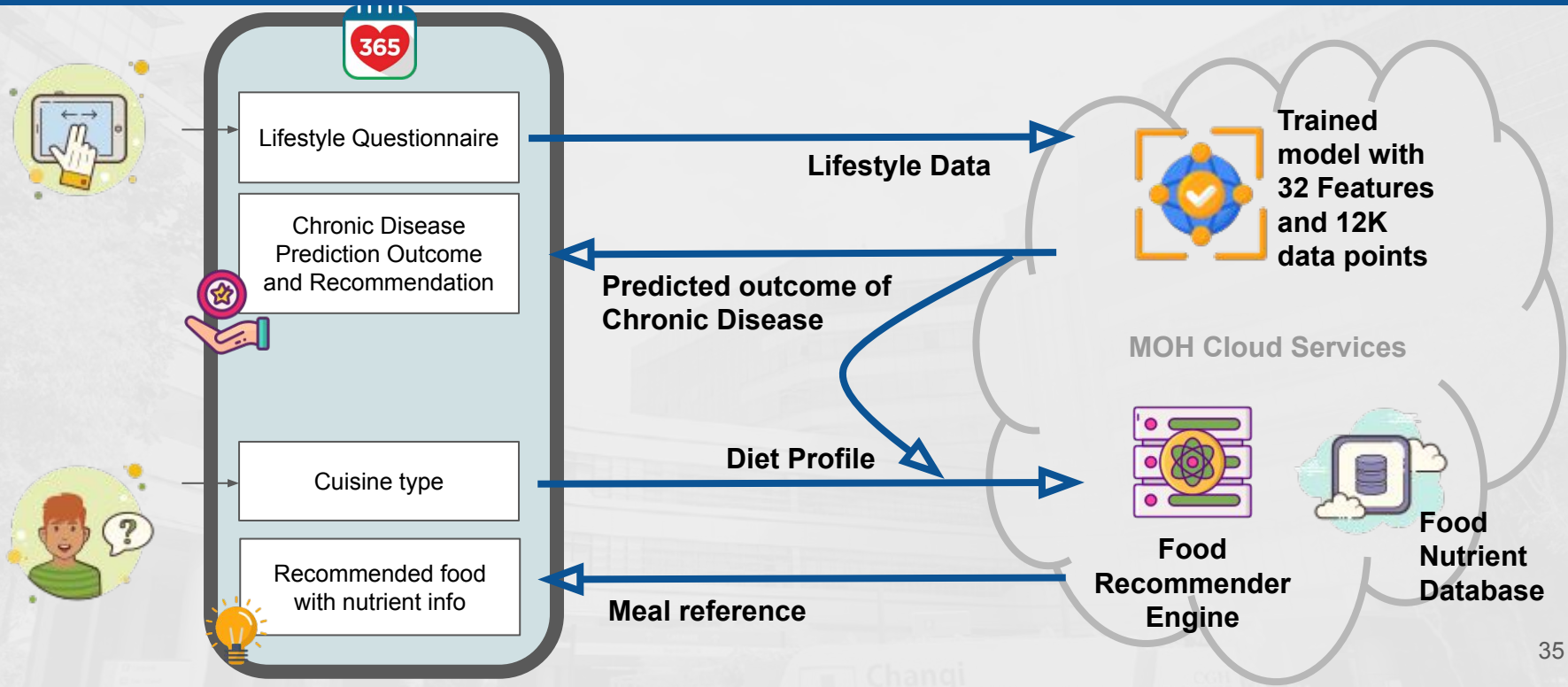
Gamified health challenges to motivate active and healthier lifestyles

Personalized health info

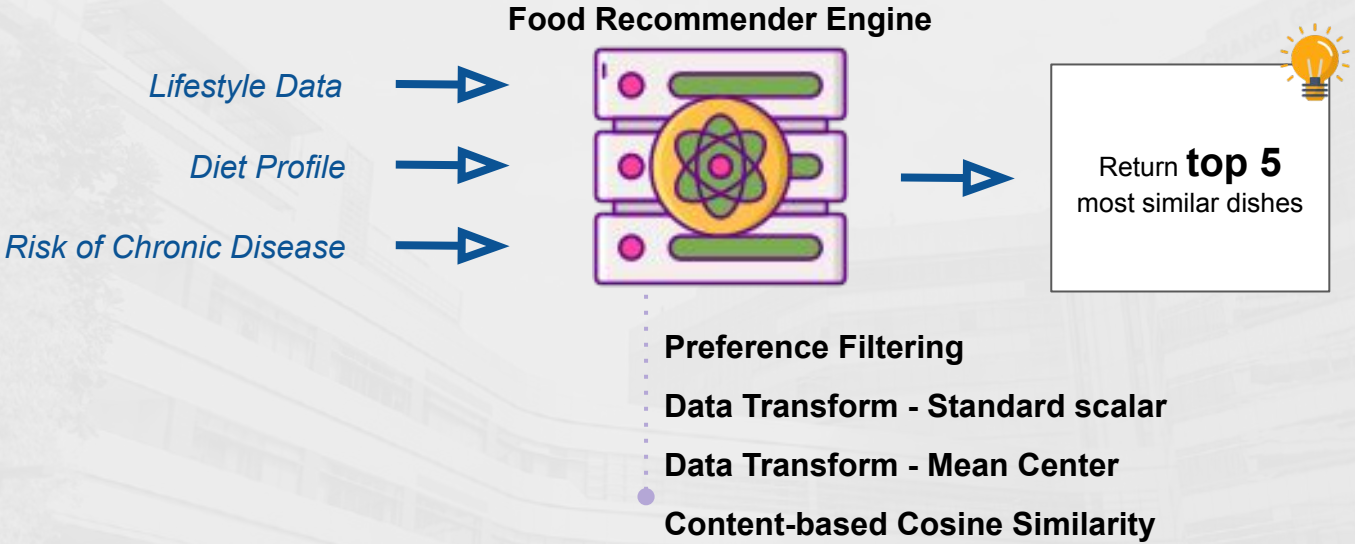
Chronic disease risk calculator through lifestyle questionnaire

Health enhanced meal recommendation

Recommender Engine in Action



Recommender Engine in Action



App Demo

Healthy365 - Chroniscope

Scan me to try out !!

Cost-Benefit Analysis

Benefit via prevention + early intervention

Aspects of Society

Quality of Life

Economy

Healthcare System

Emotional Well Being -
Helps improve mental health by easing the need of constant management of symptoms and potential progression

Encourages Healthier Behaviors -
Starting early will prevent significant lifestyle changes that can be difficult to incorporate into one's daily routine

Longevity -
Extends life expectancy, promoting active lifestyle well into old age

Qualitative

Benefit via prevention + early intervention

Aspects of Society

Quality of Life

Economy

Healthcare System

Better country's productivity

Study shows that cost of lifestyle risk accounts for SGD$1.98 billion loss mainly driven by productivity loss per year.

Reduced Healthcare Costs

Lower the costs associated with treating advanced stages of chronic conditions

(Source: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-023-16198-2)

Benefit via prevention + early intervention

Aspects of Society

Quality of Life

Economy

Healthcare System

Alleviate Burden on Healthcare Resources -
Reducing the strain on healthcare facilities, freeing up resources to handle acute cases more effectively

Data Driven Insights -
Preventive healthcare generates valuable data, as guide to decisions on health policies

Health Equity -
Bridge health disparity gap whereby all population segments have access to essential health services

Recommender Feature - Cost

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

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

Assuming 3.5% increase in cost due to inflation. Link to detailed analysis

Quantitative

Recommender Feature - Benefit

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

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

Assuming 3.5% increase in cost due to inflation. Link to detailed analysis

Healthy 365 App - Cost Benefit Analysis

Assuming 1% conversion…..

Year	Benefit (SGD)	Cost (SGD)	Net Benefit (SGD)
Year 1	$1,707,940	$739,184.40	$968,756
Year 2	$1,856,104	$320,044.47	$1,629,802
Year 3	$2,017,121	$335,586.22	$1,783,409

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

Appendix

Quantitative

Singapore Healthcare Spending - Escalating

Healthcare Spending[1][2]	Health Promotion and Preventive Healthcare Spending [2][3]	Annual cost per capita on Chronic Disease [4]
Government social spending on healthcare ● 2010 - $3.7b ● 2019 - $11.3b ● 2023 - $16.68b ● 2030 - $27b (estimate)	Government spend on Health Promotion and Preventive Healthcare ● 2022 - $810.33m ● 2023 - $933.59m 2022 - Healthier SG programme Setup cost $1b	Median annual healthcare cost per capita for patients with chronic diseases (year 2021) ● 1 disease (14.8%) - $332 ● 2 diseases (22.8%) - $457 ● 3 diseases (26%) - $652 ● 4 diseases (20%) - $827 ● >4 diseases (16.4%) - $1,021

[1] The Straits Times 17-Oct-2022 https://www.straitstimes.com/singapore/budget-2022-healthcare-spending-to-form-bulk-of-govt-expenditure-by-2030
[2] Head O Ministry of Health 2024 https://www.mof.gov.sg/docs/librariesprovider3/budget2024/download/pdf/35-moh-2024.pdf
[3] Today Online 4-Oct-2022 https://www.todayonline.com/singapore/s1b-setup-cost-healthier-sg-ong-ye-kung-2010556
[4] https://annals.edu.sg/healthcare-cost-of-patients-with-multiple-chronic-diseases-in-singapore-public-primary-care-setting/