Nutri-Grade labels: SINGAPORE
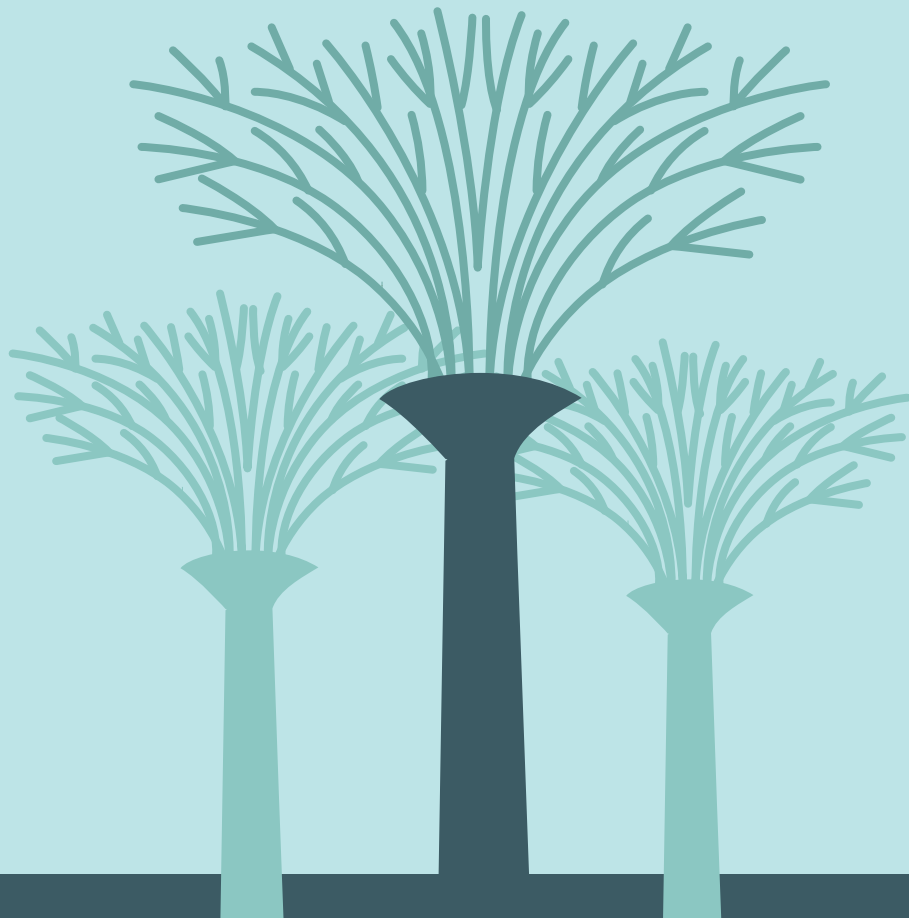
Viva La Wholegrain Muffin

TABLE OF CONTENTS

01

Understanding public health

Longer life expectancies

SINGAPORE: The life expectancy of Singaporeans is among the highest in the world at 81.4 and 85.7 for men and women respectively in 2019, according to the

According to other reports, with a combined life expectancy of 84.27 years, Singapore has the 7th highest life expectancy in the world.

If you think the life expectancy in Singapore is high, know that it's going to get even higher as it is expected to reach 85.4 years in 2040.

It seems like Singaporeans are enjoying higher life expectancies and are expected to live longer in the future

Poor life quality

old on average. While this is positive news, there is also a flip side to this statistic – and that's the fact that Singaporeans are spending more time in ill health, especially towards the end of life.

A person born in 2017 would live an estimated 12.5 per cent of his life in ill health, compared with 11.8 per cent for a person born in 1990

From the study, Singaporeans spent 10.6 years in ill health in 2017, which is 1.5 years longer than in 1990.

But in reality, they are not enjoying their life as they are plagued by medical conditions.

Top chronic illness

Hypertension

Hypertension or high blood pressure can damage your arteries by making them less elastic.
It's often associated with a diet high in **salt**.

Diabetes

People with diabetes are unable to fully utilize the glucose in their bloodstream because they either lack insulin in the body or their body cells have developed resistance to insulin.
It's often associated with a diet high in **sugar**.

Hyperlipidemia

High cholesterol is a condition in which fatty deposits develop in your blood vessels.
It's often associated with a diet high in **saturated fat.**

63.7% / 3.5 M

Singaporeans suffer from hypertension, diabetes and hyperlipidemia.
(Adults aged 18-69 years old)

3 ways to prevent chronic illness

Physical activity

Engage in at least 150-300 minutes of moderate-intensity aerobic activity in a week.

Diet

Less sugar
Less saturated fat
Less salt

Healthy life choices

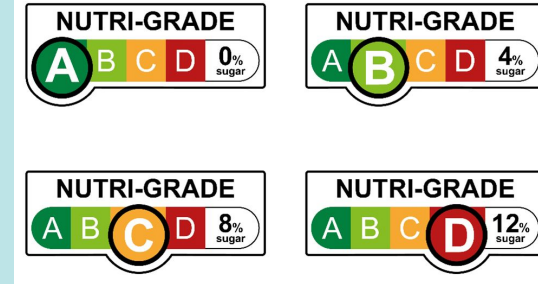Avoid tobacco and excessive drinking

Current measures

Healthier choice

There are many variations of this symbol but the words are too small to read.
Some products may be lower in sodium but higher in calories.
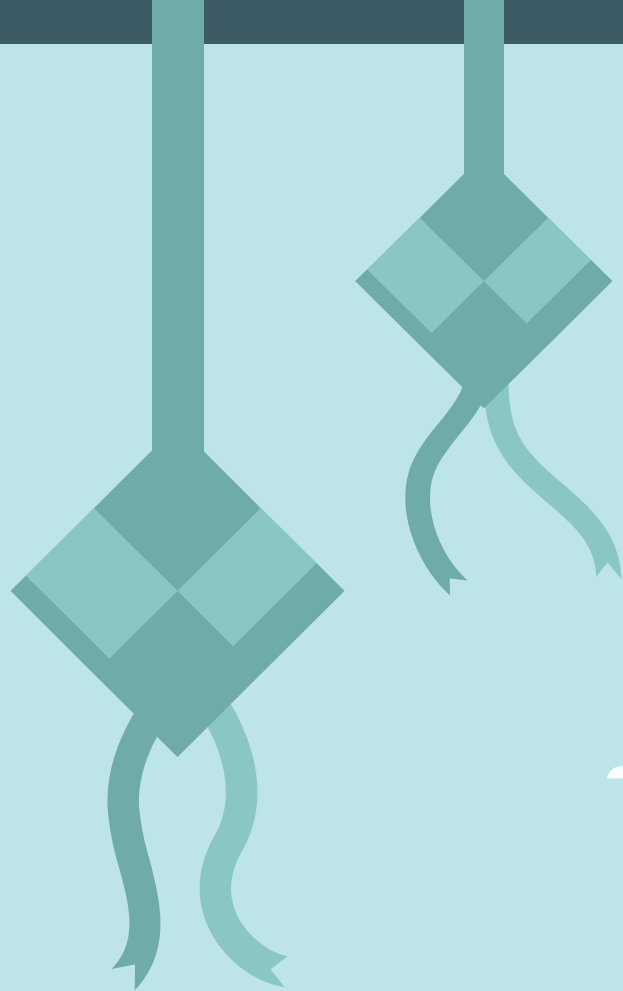

Nutri-grade

Has 4 different ranks but there is an over-reliance on sugar content.
Does not account for other nutrients.
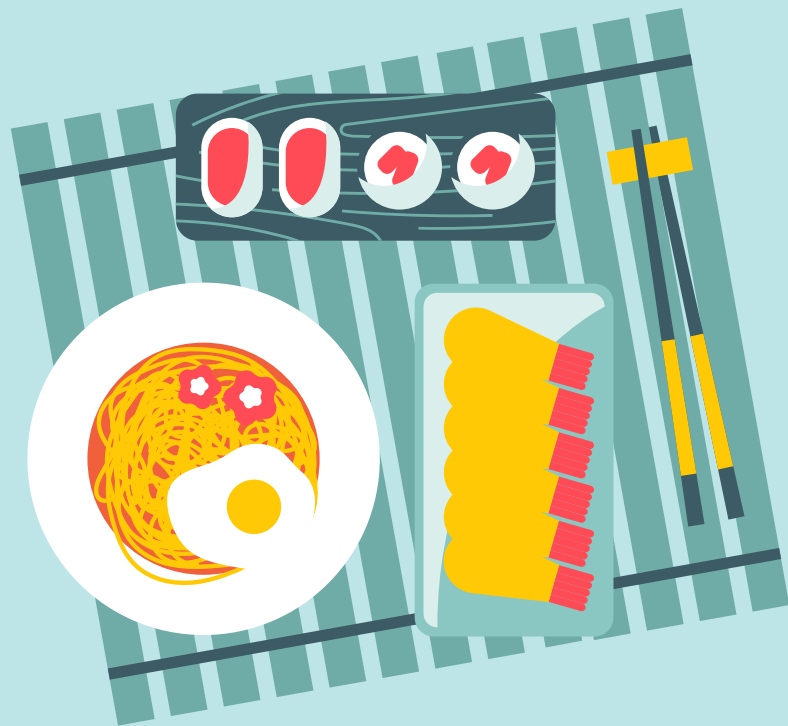
Problem Statement

The nutrigrade labels only take into account saturated fat and sugar and do not provide a holistic picture of the health of the drinks.

Is there a way to create a more comprehensive indicator of how healthy drinks are?

02

Analysis

Workflow

01

Data scraping

Getting the nutritional information from the NTUC website

02

Data cleaning

Ensuring that all the values are in the correct format

03

Analysis & Clustering

Exploring the data Using algorithms to find clusters

04

Modelling

Finding the best model to gain an accuracy of more than 95%

05

Prototype

How consumers can use our solution to make more accurately informed choices
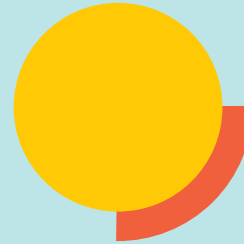
EDA: Composition

Goal: To find action points which focus on nutrients **not covered** by existing and proposed labels.

Finding Drink Categories

Through exploring dataset
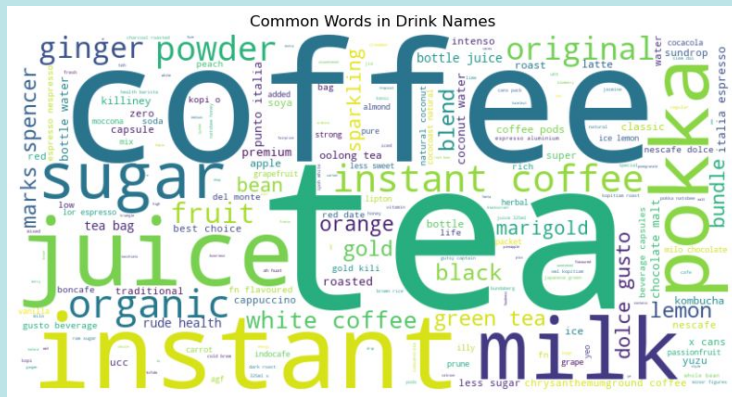
Finding Dietary Habit Gaps

Through secondary research

Tea, Coffee, Juices, and Milk are popular drinks

Drink Types

Coffee
Tea
Juices
Milk
Cola/Soda

Common Words in Drink Names

Brands

Pokka
Marigold
Yeo's
Coca Cola
F&N

Needs of Different Age Groups Overlap

Children (5 – 17)

Adults (18–60)

Elderly (60+)

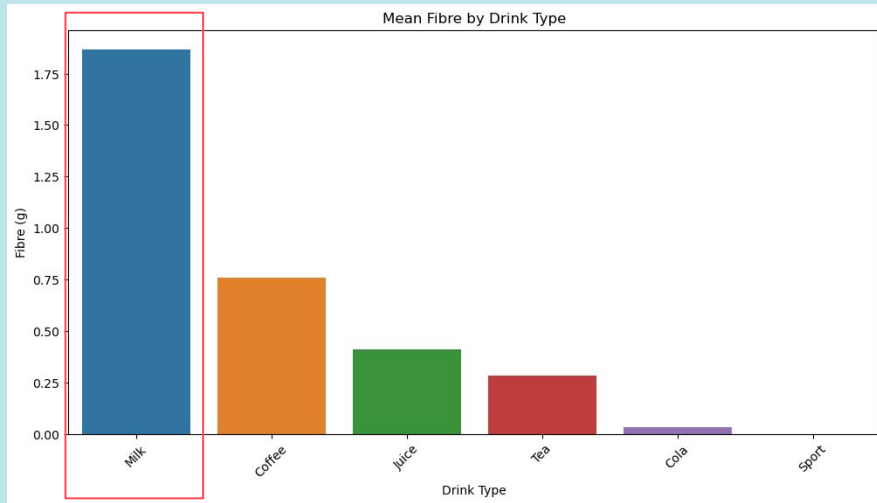⬆️ Dietary Fibre
    Protein

⬆️ Dietary Fibre
    Energy

⬆️ Dietary Fibre
    Protein

Overlapping requirements for the 3 groups, could help in making targeting more efficient.

References:
- https://www.straitstimes.com/singapore/4-in-5-elderly-singaporeans-who-are-at-risk-of-malnutrition-have-low-muscle-mass-study
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10421189/#:~:text=Inadequate%20dietary%20calcium%20intake%20was,in%20dietary%20fiber%20and%20calcium.
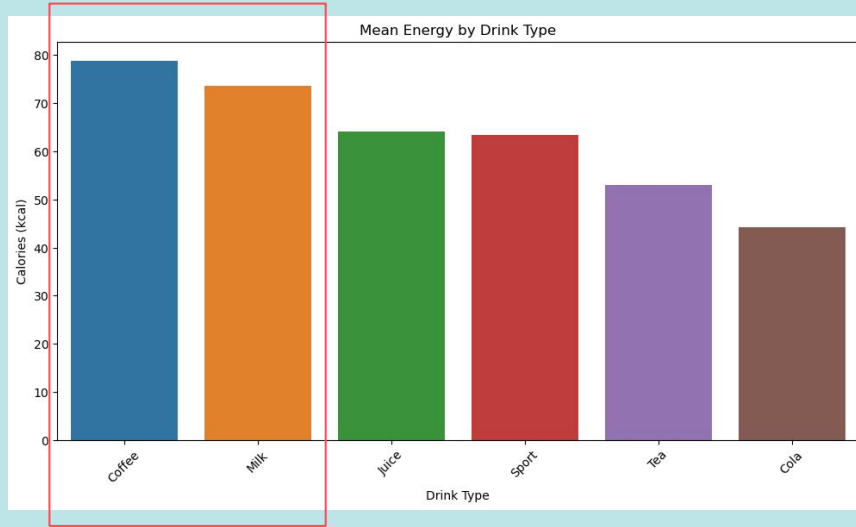- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7612427/
- https://www.healthhub.sg/live-healthy/recommended_dietary_allowances

Dietary Fibre

Mean Fibre by Drink Type

Milk is relatively high in Fibre content

Oat milk contains high Fibre

Note:
- Juices in general are not a significant sources of Fibre.
- However, pulpy juices can be relatively high in fibre content.

Orange juice with pulp is a good source of fiber. Studies have shown that fiber helps the body maintain digestive health and may lower your risk of

References:
- https://www.webmd.com/diet/health-benefits-orange-juice

Energy

Mean Energy by Drink Type

- Coffee provides relatively higher energy (due to Caffeine).

- Milk (especially whole milk) is also energy rich.

Protein

Mean Protein by Drink Type
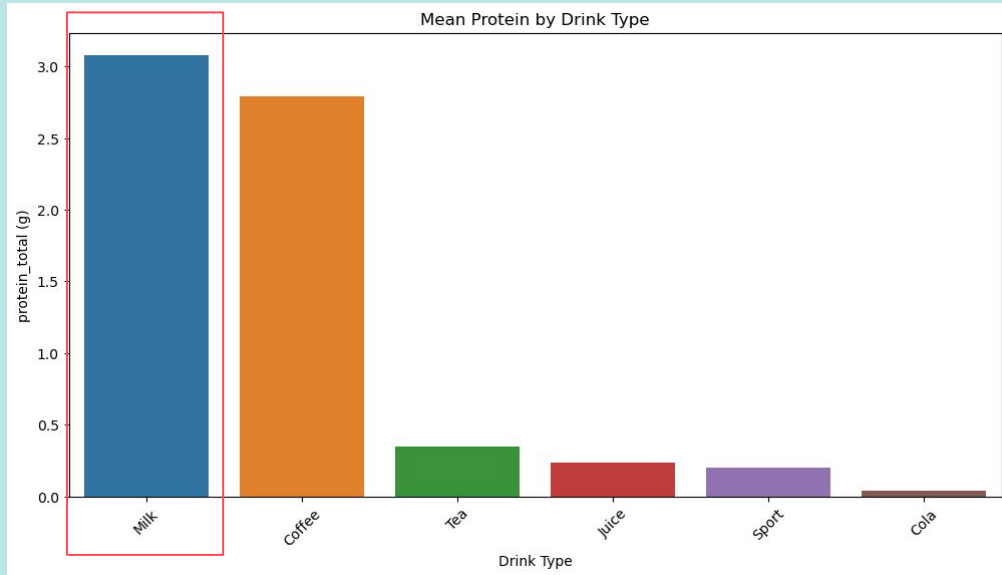
Note: Coffee by itself is not a significant source of protein, but Milk Coffees/Latte can be.

- It seems like coffee and milk are good sources of protein.

- Milk should be considered as a better protein source.

The short answer: no, coffee is not a good source of protein.

However, protein content depends on the type of coffee. For instance, one cup (about 6 fluid ounces) of black coffee contains approximately 0.21 grams of protein, which is not meaningful, to say the least.

When you add sugar, cream, milk or any other sweeteners to your coffee to improve taste or richness, you may increase its protein content.

For example, one cup of caffe latte contains an average of 4.41 grams of protein, which is a significant contribution to your daily recommended intake.

Action Points: to the groups

Children (5 – 17)

- Push beverages like milk (especially oat milk) and pulpy juices to increase fibre intake.

- Milk will also help increase the protein intake, reducing risk of malnutrition.

Adults (18–60)

- Push beverages like coffee and whole milk to increase energy and fibre intake.

- As with children, pulpy juices can also be targeted towards this age group.

Elderly (60+)

- Push and target milk/milk based beverages to help increase protein intake and fight muscle decay.

- Milk will also help increase fibre intake. Juices may not be ideal (due to the sugar content), but specific teas (like matcha) could also help to increase fibre intake.

Action Points: To NTUC

- NTUC house brands do not occur frequently in the data.

- Cause for concern given that the data has been scraped from the NTUC website.

- Promote drink products more aggressively, like ensuring their beverages are among the first few search results on their website.

- Try a differentiation strategy on the basis of the nutrient content of their drinks. For instance, they could produce drinks rich in fibre or protein and market them to the above demographics.
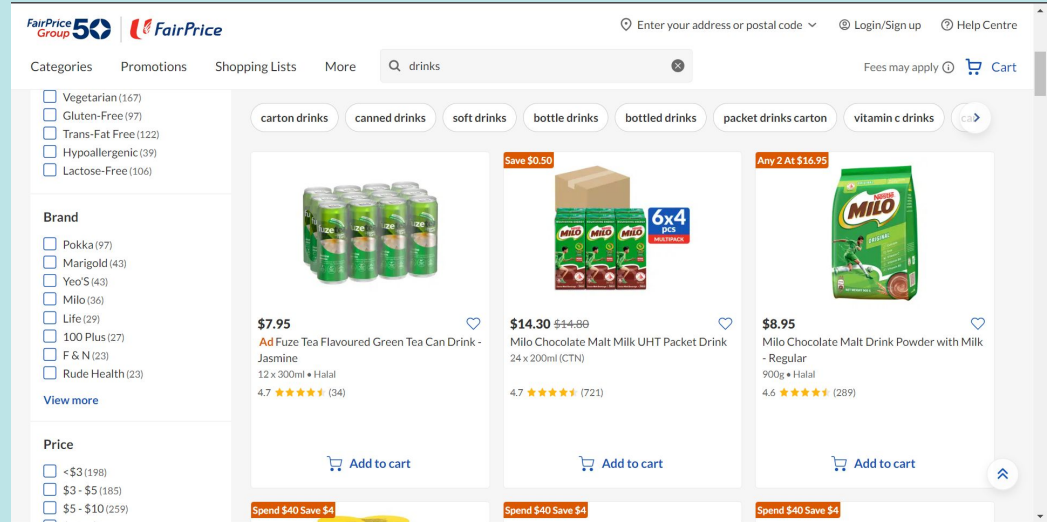
03

Machine Learning
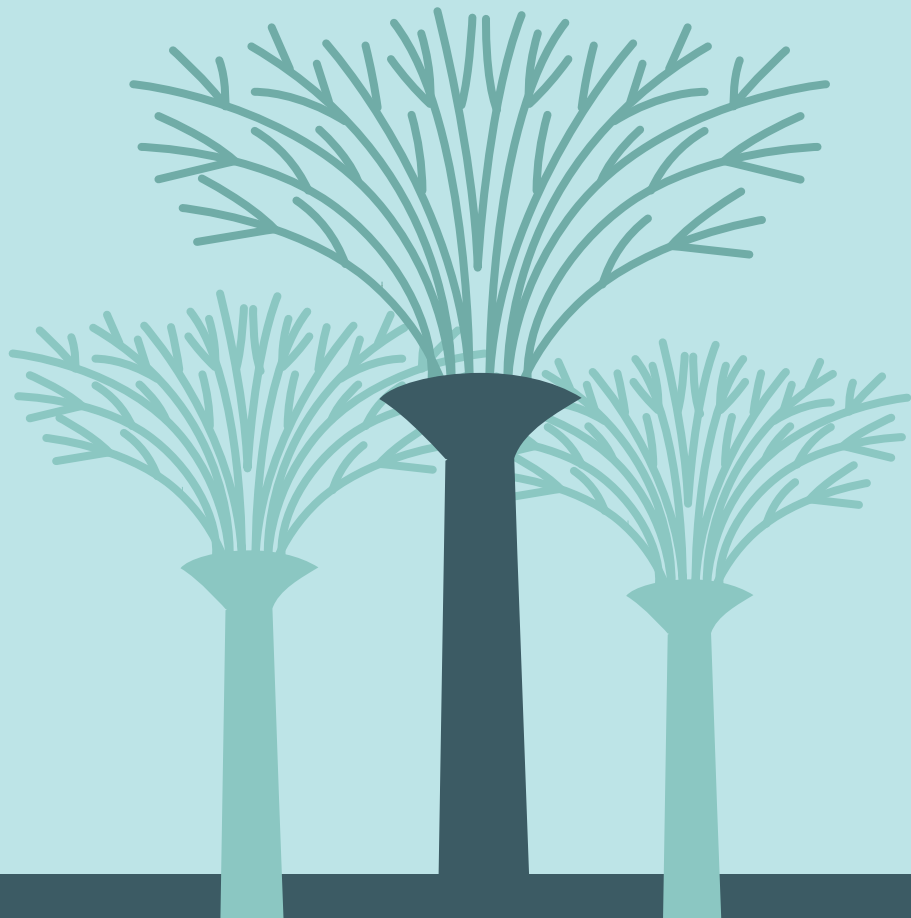
Standard Scaler

- **Equal Importance *:*** Ensures all features have equal importance by scaling them to the same range

- **Improved performance** : Algorithms like K-means clustering, performs better when features are on a common scale.

- **Consistency** : Offers a standardized way to compare data across different units or scales.

Elbow Method

- **A technique used to determine the optimal number of clusters(k) for a dataset**

- **Produces a plot of the sum of squared distances versus the number of clusters. The elbow point suggests the optimal k value**

Elbow Method for Optimal Number of Clusters
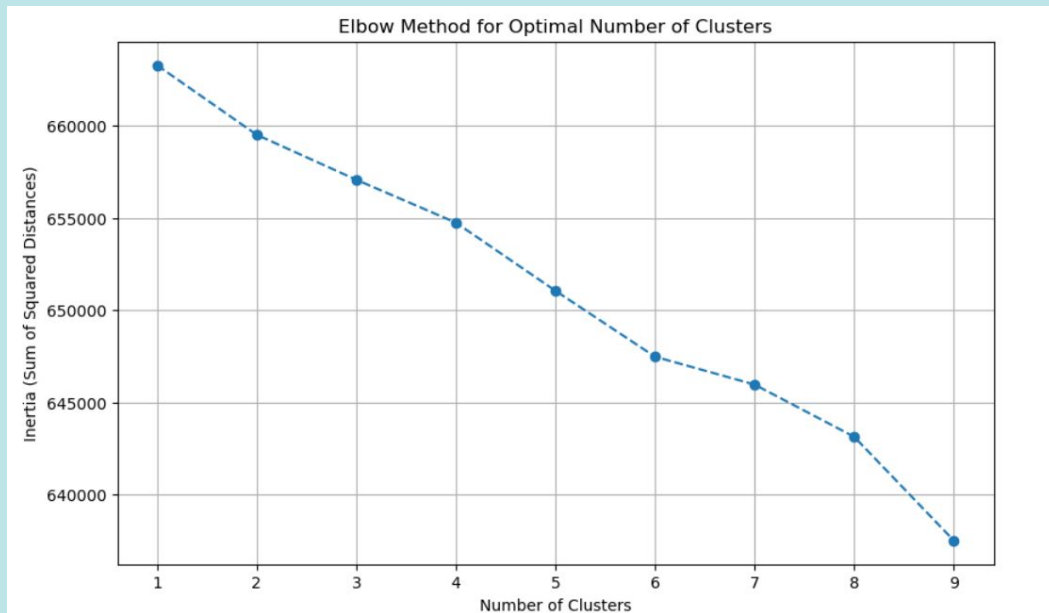
K-Means Clustering

- **In this analysis, K-means groups drinks by similar attributes**

- **The algorithm picks a set number of center points and assigns each drink to the nearest one, creating clusters**

- **Similar drink attributes cause indistinct clusters in visuals**
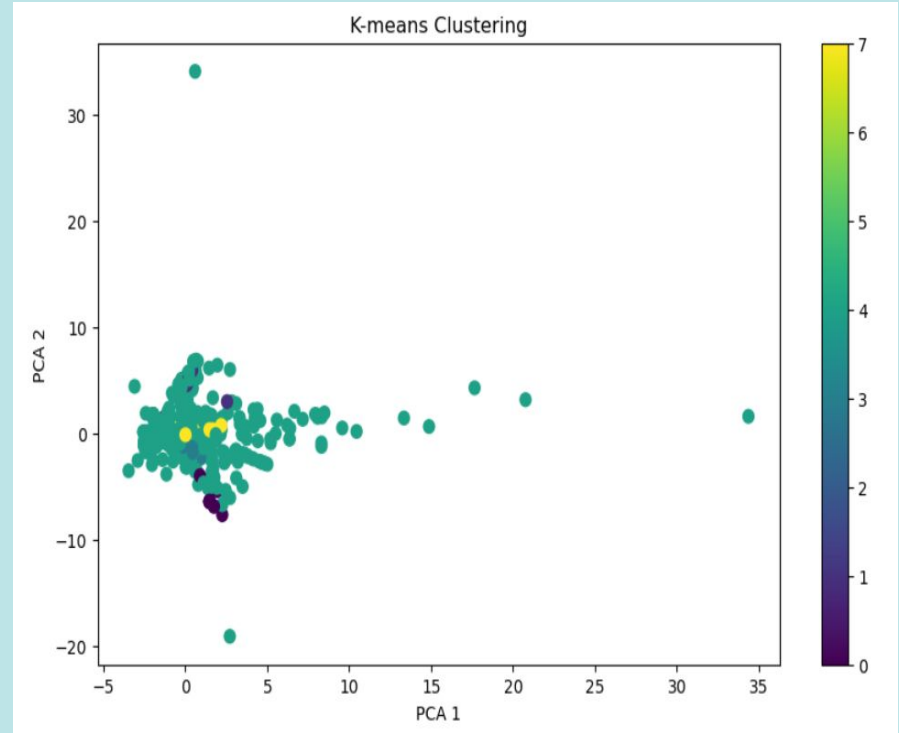
Nutrigrade (Revised)

- **Holistic Evaluation : By considering sugar, saturated fat and sodium, we capture comprehensive view of the drink's health impact.**

- **Clear consumer awareness: Using grades "A" to "D" makes the health metric easily understandable for consumers**

- **Worst-Case Scenario: By classifying based on the least healthy nutrient, we ensure consumers are aware of any potential health risk even if other nutrients are at healthier levels**

Why Use Models?

- **Adaptability : Machine learning models can adapt to new data. If drink formulations change or new drinks enters the market, the model can adjust without needing to manually redefine thresholds.**

- **Granular Insights: While thresholds give broad categories, a model could provide more granular or nuanced insights about the healthiness of a drink.**

- **Future Proofing: Health guidelines and research can evolve. A model can be retrained on new guidelines and research findings, ensuring our evaluations remain current.**

424 Total Drinks

Nutrigrade	A	B	C	D
	115	50	85	174

SMOTE (Simple Minority Over-sampling Technique)

Purpose — It is used to address class imbalance by generating synthetic samples in the feature space

Mechanism — SMOTE selects 2 or more similar instances from the minority class and creates a new instance that is a blend of the selected instances

Advantages — It helps improve model performance by making the class distribution more balanced.

Adaptability — SMOTE is like a chameleon; It can change fit into many different situations, just like how a chameleon changes its colours to fit into different surroundings. Its flexible and can work with various types of data

After SMOTE

	A	B	C	D
Nutrigrade	174	174	174	174

696 total samples

METRICS

Accuracy

- Shows overall correctness of the model

- Reflects true successes over total attempts

- Main metric for general model performance

F1 - Score

- Can provide deeper insights than accuracy alone

- Highlights false positives and negatives

- Essential for imbalanced datasets

Random Forest

What does it do? :

- _Teamwork_ : Uses multiple mini - decisions to make a final decision

- _Knows Important Info_ : Can tell which facts(ingredients) matter the most

- _Doesn't Assume Too Much_ : Uses multiple checks to avoid wrong guesses

XGBoost

What does it do? :

- **_Gets Smarter Step-By-Step_ : Learns from mistakes and tries to do better each time**

- **_Fair to All Data_ : Even if some data are rare, XGBoost ensures it's considered**

- **_Avoids Being Too Sure_ : Uses techniques to prevent jumping into conclusions**

Model comparison

	Train	Test	F1-Score
Random forest	**100%**	**96%**	**95%**
XGBoost	**100%**	**99%**	**99%**

Model comparison (Tuned)

	Train	Test	F1-Score
Random forest	100%	94	95%
XGBoost	100%	99%	99%

XGBoost

★★★★★

High Test Accuracy

XGBoost achieved a remarkable 99% accuracy on test data, indicating its excellent prediction capability

★★★★★

Efficient Run Time

The model is optimized for performance, ensuring quick predictions without compromising quality
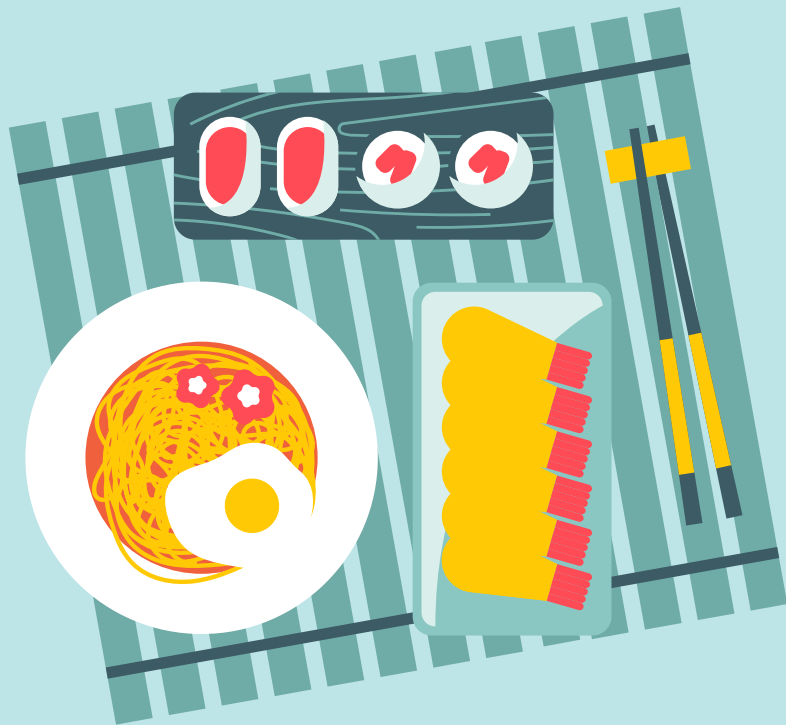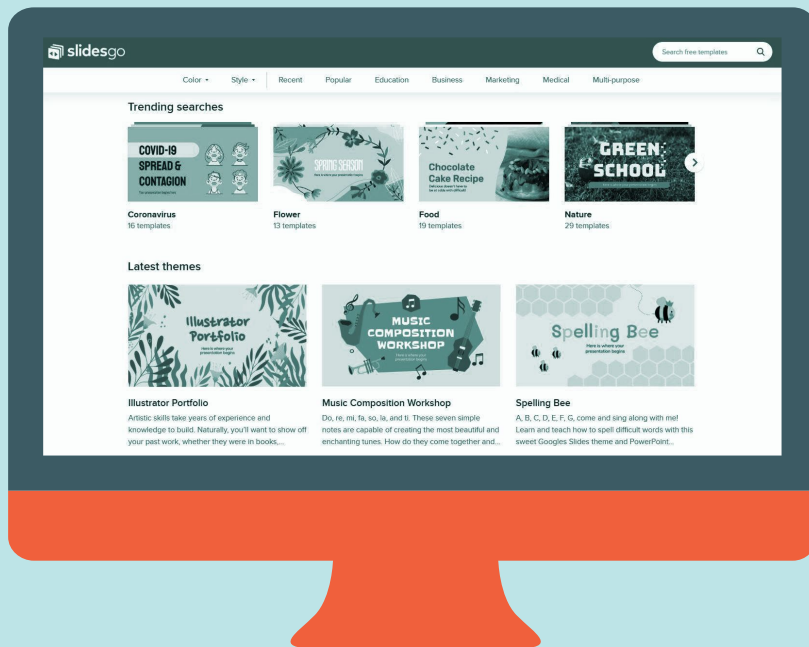
★★★★★

Robustness

XGBoost handles various data types, missing values and is less prone to overfitting, making it a reliable choice for diverse datasets
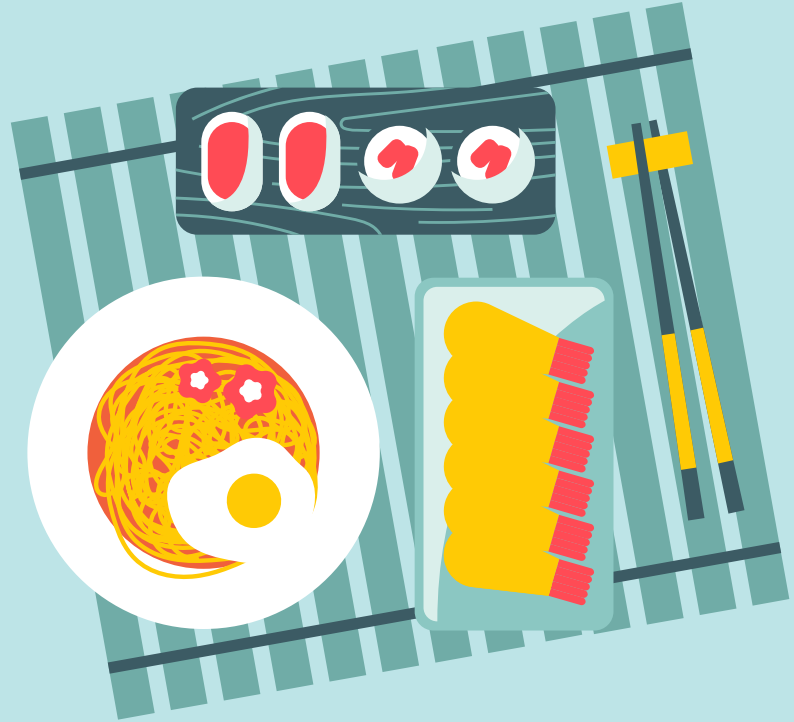
04

Our solution

Online portal

Upload a picture of your drink and get a more accurate health label!

04

Cost/Benefit

Underlying Reasoning + Assumptions (½)

- **<u>Costs:</u>**

- The HPB spent **~SGD 152 Million** on Health Programmes in 2022. There are **35 active programmes** on the HPB website currently.
- Assuming an even split, the HPB spent **~2.85%** of the total cost on Nutrigrade.
- It is assumed that companies **will have to repackage** their products to include Sodium information.
- **Yeo's** was taken to represent the local RTD industry as it represents **~1/3rd of market share** and is a local brand.
- Studies have shown that the average packaging cost is **15% of Cost of Goods Sold (COGS)**. Yeo's had a COGS of **~250.6 Million in 2022.**

References:
- https://www.hpb.gov.sg/docs/default-source/pdf/hpb-2022_2023-annual-report.pdf?sfvrsn=bec1a971_2
- http://yeos.com.sg/wp-content/uploads/2023/03/Yeo-Hiap-Seng-Limited_AR-2022.pdf
- https://www.healthhub.sg/programmes
- https://sec.org.sg/pdf/reports/unlocking-sustainable-packaging-opportunities-in-Singapore.pdf
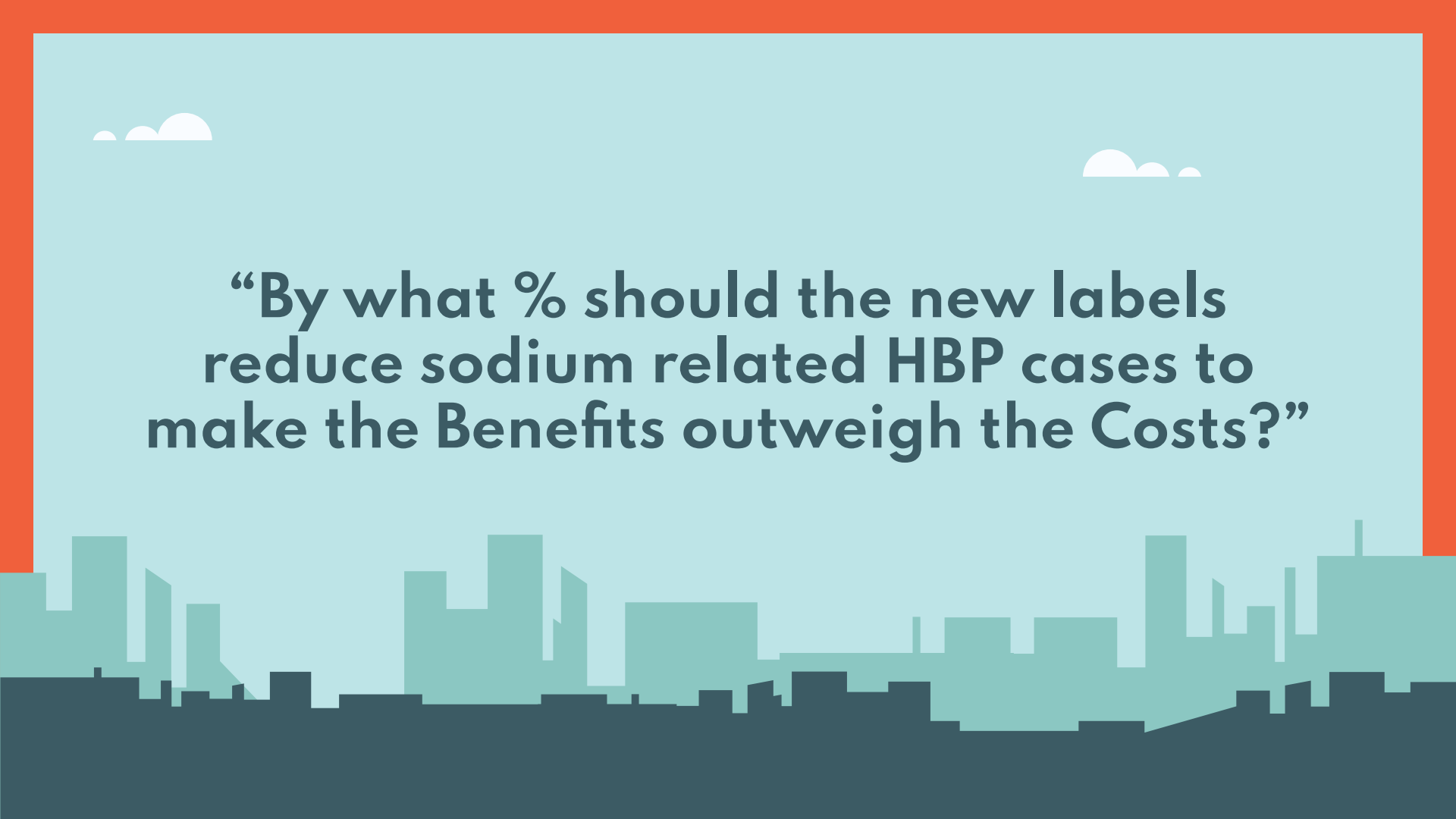
Underlying Reasoning + Assumptions (²⁄₂)

- **Benefits:**

- **37% (around 2.2 Million people!)** of SG is population affected by high blood pressure (HBP).
- Studies show that **~26.4% (~587K)** of HBP cases are attributable to **excess sodium intake.**
- The **average cost** of admitting a patient for HBP is **$2440.** Total cost of sodium related HBP cases will be ~**SGD 1.43 Billion**.
- It is assumed that the **benefit to the people** will be the reduction in the sodium related HBP cases.
- The benefit of reducing the number of sodium related HBP cases by **1%,** for e.g., is **SGD 14.33 Million.**

References:
- https://www.myheart.org.sg/health/risk-factors/high-blood-pressure/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6770596/
- https://www.valuechampion.sg/average-cost-cardiovascular-disease-treatment-singapore

"By what % should the new labels reduce sodium related HBP cases to make the Benefits outweigh the Costs?"

Cost Breakdown

Item	SGD
Implementation Costs (to the HPB) (2.85% of toal Programme Costs)	4.3 Million
Compliance Cost (to companies) (packaging cost is 15% of COGS)	37.6 Million
Total Cost	41.9 Million

Computing Benefits

Item	SGD
Total cost of sodium related HBP cases	1.43 Billion
Benefit of reducing sodium related HBP cases (by 1%, for e.g.)	14.33 Million
Percentage reduction required to break even	???

Finding the Target

BREAK EVEN POINT: Total Benefits - Total Cost = 0

Required % reduction **= 2.92% = ~3%**

The new label will be worthwhile if a **3% reduction** in sodium related HBP cases is achieved.

Nutri-Grade labels reduced median sugar % by ~3% in **4 years.** Target time span should be **similar or better.**

For maximum effect: New labels **must not be limited** to drinks, must also include **processed food.**

Problem

The nutrigrade labels only take into account saturated fat and sugar and do not provide a holistic picture of the health of the drinks.

Is there a way to create a more comprehensive indicator of how healthy drinks are?

Solution

An online portal that users and businesses can use to upload nutritional information and get live labels.

- Provide users with more accurate labels

- Reduce the amount of sodium intake

- Reduce the chance of Singaporeans getting chronic illnesses

Steps forward

01

Improve the model

Expand the drink catalogue

Create more accurate labels and better recommendations

02

Release the new label

Allow companies to upload their drinks and we will return their new label

03

Update

Update the labels by conducting blind experiments to measure it's effectiveness

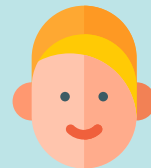Introduce health tax of 10% on unhealthy drinks

Context

Who are we?

We are part of the data science team at HPB

Who are you?

You are the management of HPB. You are planning to update the nutri-grade labels.

Future work

01
Include more drink categories
We can expand our drink catalogue to include fresh drinks and alcohol.

02
Include more nutrients
We can include more ingredients such as calcium and vitamins to create better labels.

03
Health tax
We could put a 10% tax on unhealthy drinks and use that money to subsidise healthy drinks

04
Recommendation Improvements
Extracting prices, ratings and comments

Future work:
Health tax

03

Health tax

A 1 peso per litre (about 10%) tax on SSBs introduced in 2014 has successfully sustained sales reduction over time, resulting in savings of $4 in healthcare costs per dollar spent on tax implementation.

Future work:
Recommender Improvements

Recommendation Improvements

Extracting prices, ratings and comments.

Extracting Comments will be using a TF-IDF matrix instead together with nutritional similarity matrix

```python
# create the tf-idf matrix
@st.cache
def create_tfidf_matrix(listings):
    tfidf = TfidfVectorizer(stop_words='english')
    tfidf_matrix = tfidf.fit_transform(listings['description'])
    return tfidf_matrix

# calculate the cosine similarity
@st.cache
def calculate_cosine_similarity(tfidf_matrix):
    cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
    return cosine_sim
```

THANKS!

Do you have any questions?