

# **SC1015**

# **Mini Project**

## **Identifying Hypertension**



Lab Group B133, Team 5

Yong Shao En Ernest (U2221153B), Wu Rixin (U2221172G) & Li Liyi (U2220985F)

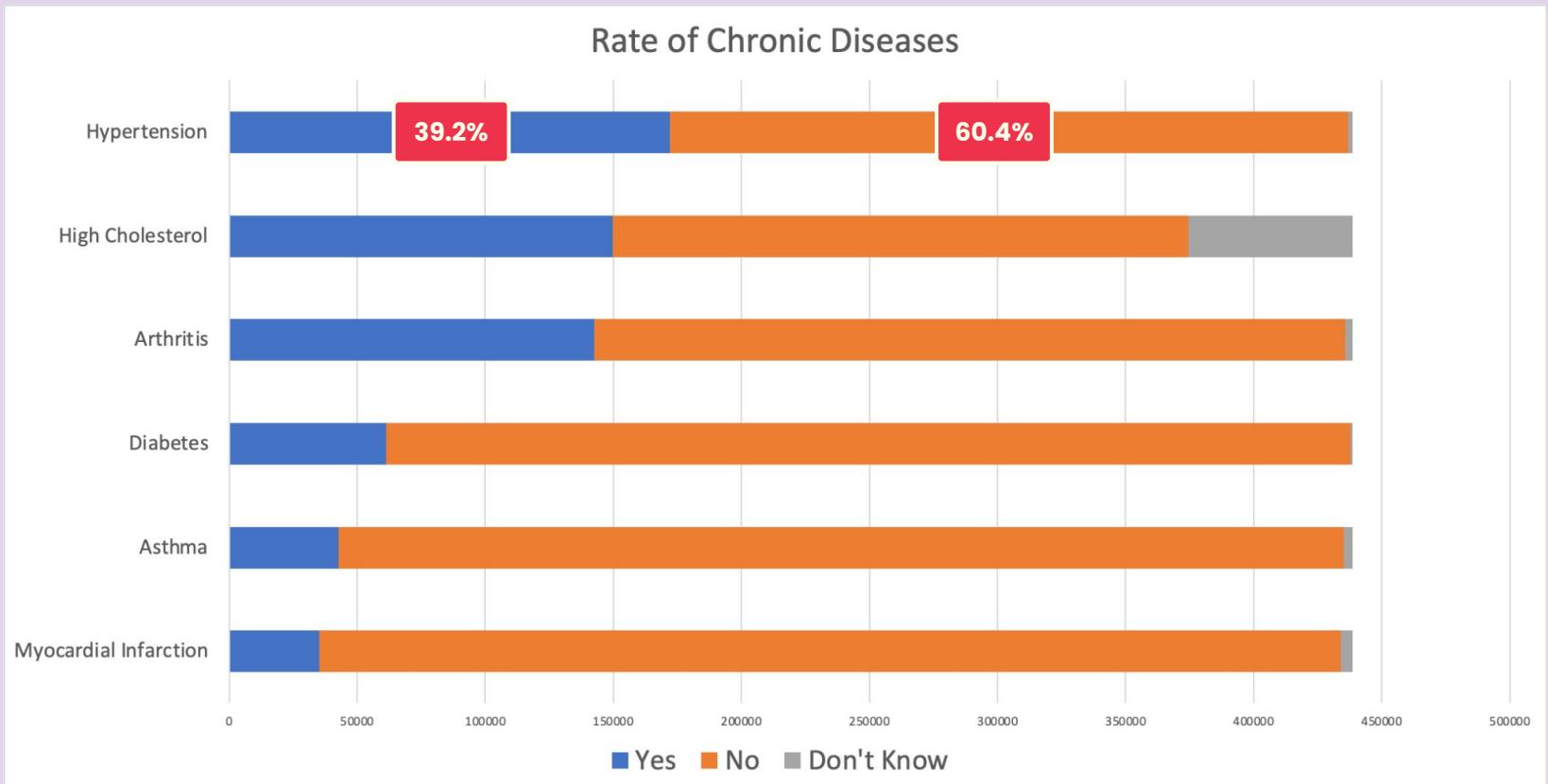
# Dataset Used

We selected data from the 2021 Behavioural Risk Factor Surveillance System Survey Data and Documentation conducted by US Centers for Disease Control and Prevention(CDC).



Centers for Disease Control and Prevention  
CDC 24/7: Saving Lives, Protecting People™

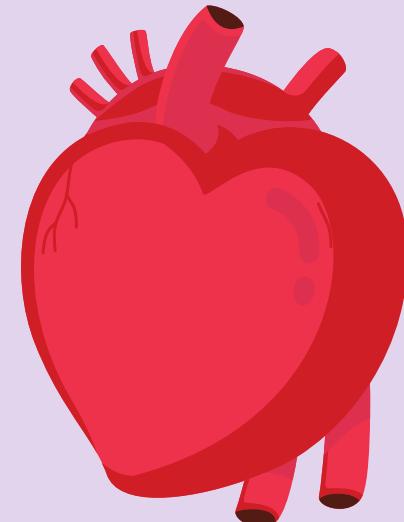
# Dataset Used



# Our Motivation

**35.5%** of  
Singaporeans have  
hypertension in 2020

**No. 1** risk factor of  
death globally



# Problem Definition

What are the variables  
**correlated** with hypertension,  
and how can we **identify**  
undiagnosed individuals  
suffering from hypertension?



# Data Extraction

Factors	
BMI	Alcohol
Physical Exercise	Smoker
Diabetes	Mental Health
High Cholesterol	Physical Health
Junk Food Intake	Race
Fruit Intake	Vegetables Intake
Education Level	



Relevant Variables	
_TOTINDA	FTJUDA2_
_BMI5	VEGEDA2_
DROCDY3_	MENTHLTH
AVEDRNK3	PHYSHLTH
_RFBING5	_RFHYPE6
CHOLMED3	_EDUCAG
FRENCHF1	DIABETE4
FRUTDA2_	_RFCHOL3



# Data Cleaning



01

Tackle missing and irrelevant values

	_TOTINDA	_BMI5	DROCDY3_	AVEDRNK3
0	2.0	1454.0	0.0	NaN
1	1.0	NaN	0.0	NaN
2	2.0	2829.0	0.0	NaN
3	1.0	3347.0	14.0	3.0
4	1.0	2873.0	0.0	NaN

# Data Cleaning



02

Create new variables by combining existing ones

**DROCDY3\_**

Drink occasions per day

**AVEDRNK3**

Number of drinks consumed

**Alcoholtake**

Weekly alcohol consumption

# Data Cleaning



03

Standardise units of measurement & adjust the decimal places for numeric variables

Value	Value Label
101 - 199	Days
201 - 299	Weeks
300	Less than once a month
301 - 399	Month / Year
555	Never
777	Don't know/Not sure
999	Refused
BLANK	Not asked or Missing

# Data Cleaning



04

Decode categorical variables based on data description

Question: Adults who reported doing physical activity or exercise during the past 30 days other than their regular job

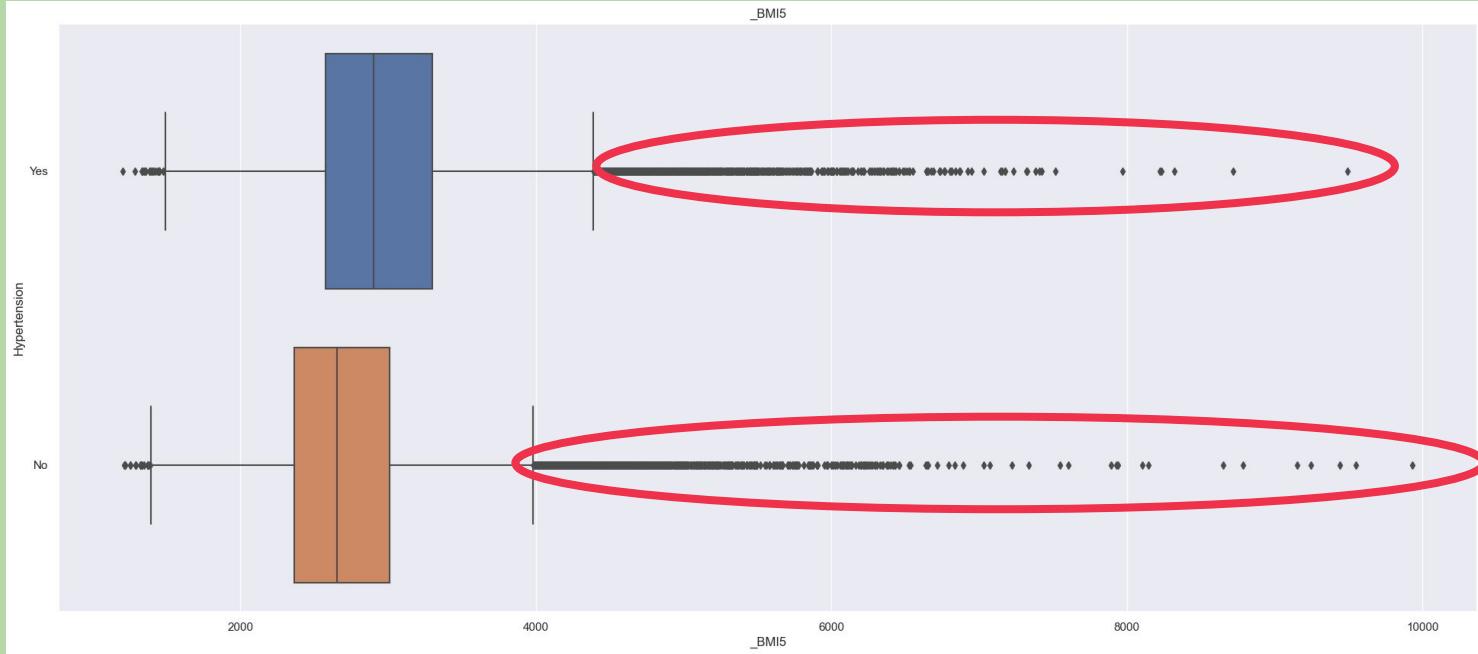
Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Yes	330,738	75.39	75.96
2	No	107,027	24.40	23.87
9	Don't know/Refused/Missing Notes: EXERANY2 = 7 or 9 or Missing	928	0.21	0.17

# Data Cleaning

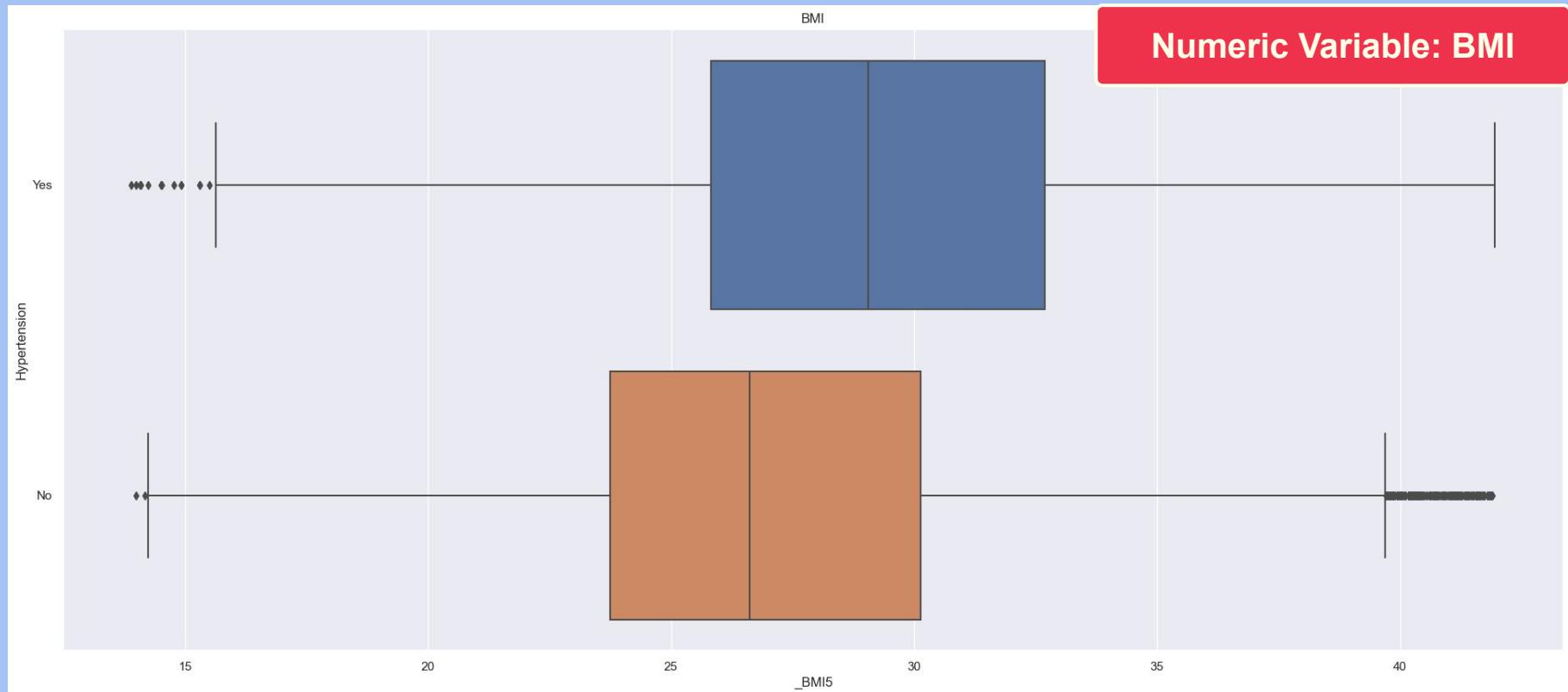


05

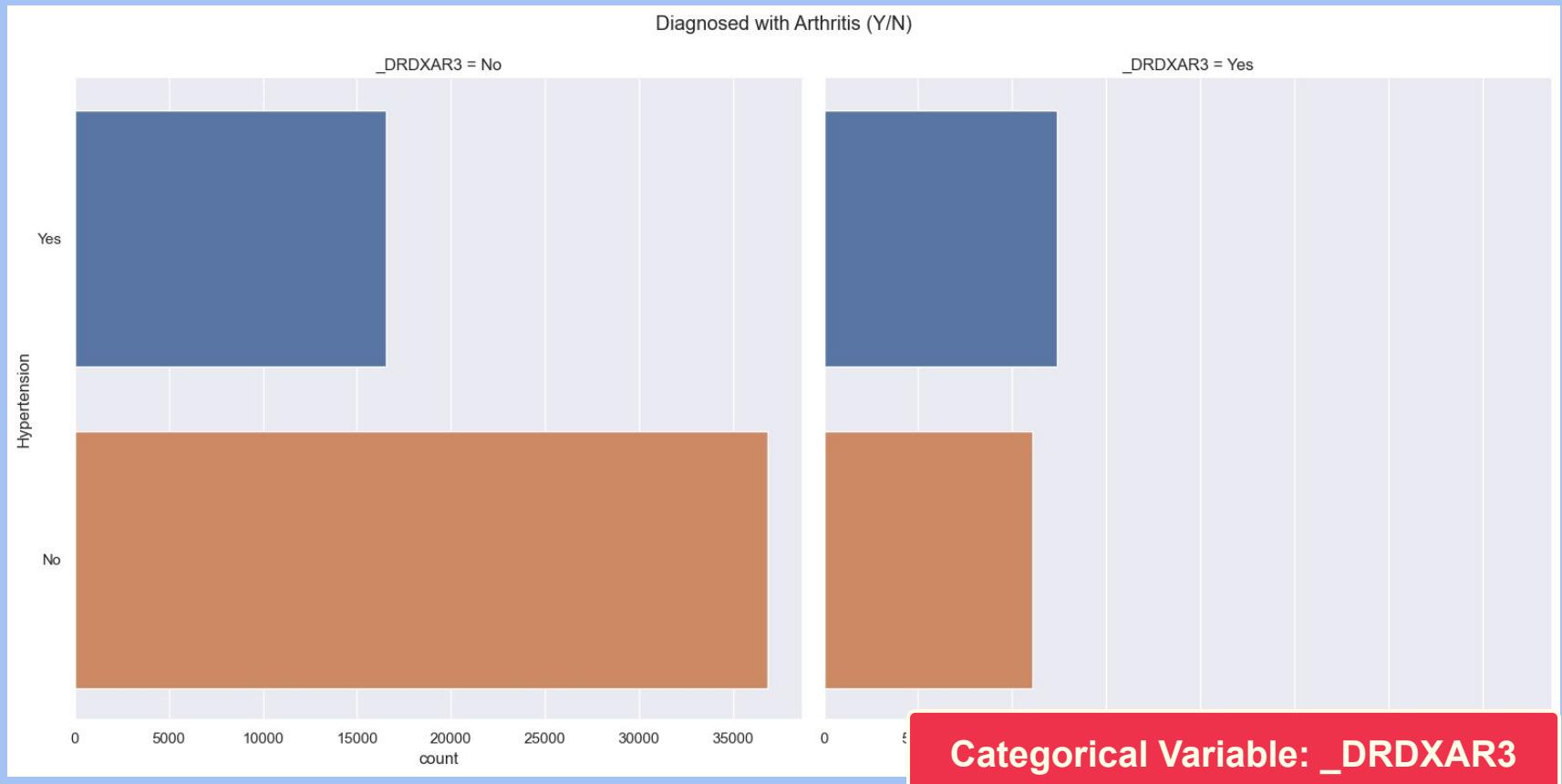
Identify and remove outliers for numeric variables



# Exploratory Data Analysis



# Exploratory Data Analysis



# Exploratory Data Analysis

## Response Variables

Hypertension

## Numeric Variables

\_BMI5

AlcholIntake

PHYSHLTH

\_AGE80

## Categorical Variables

\_TOTINDA

CHOLMED3

DIABETE4

\_RFCHOL3

\_MICHLD

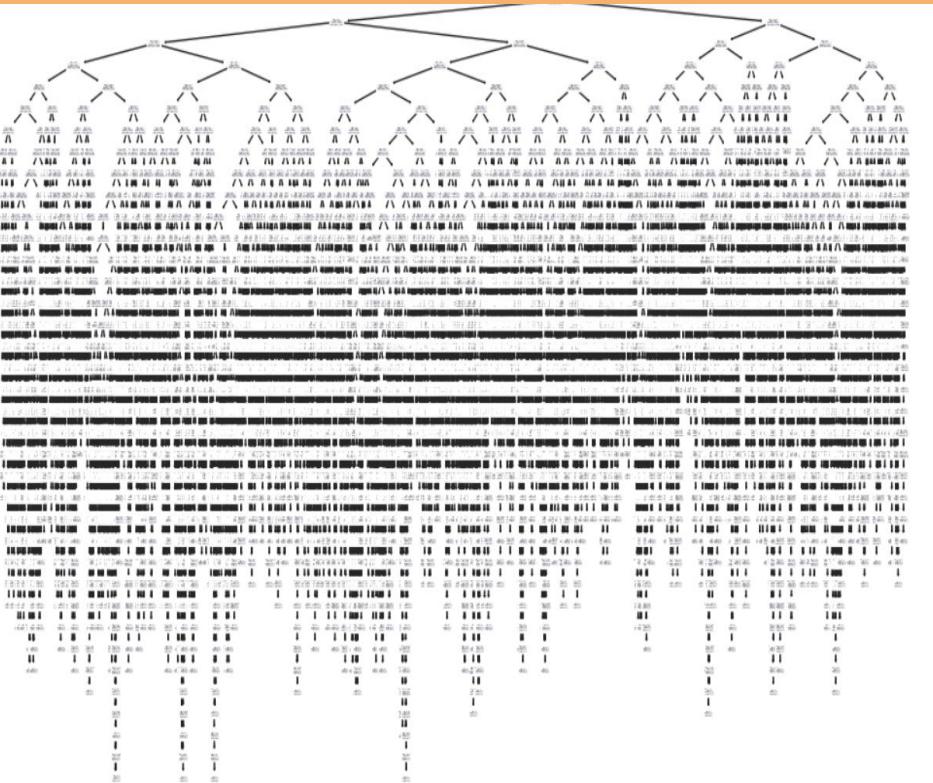
\_EDUCAG

\_DRDXAR3



# Model 1: Decision Tree

Initial Decision Tree

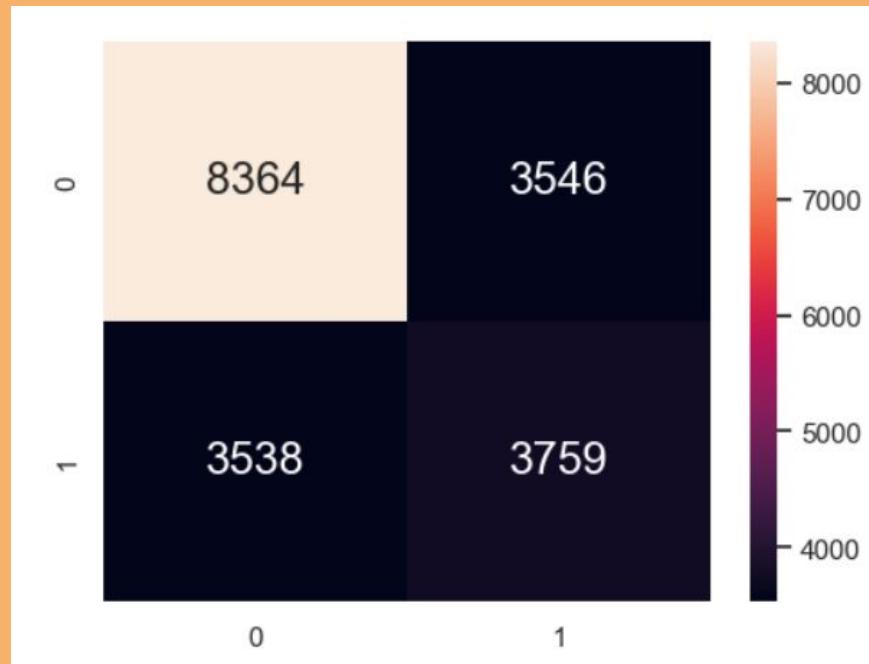


- Decision tree constructs a **model of decisions** and their **possible consequences**
- The tree can accurately **predict** the class or value of new, unseen instances
- It can handle **both numerical and categorical** variables
- Decision trees may suffer from overfitting when the depth level or the stopping criterion is not well-defined



# Model 1: Decision Tree

Confusion Matrix of  
Initial Model



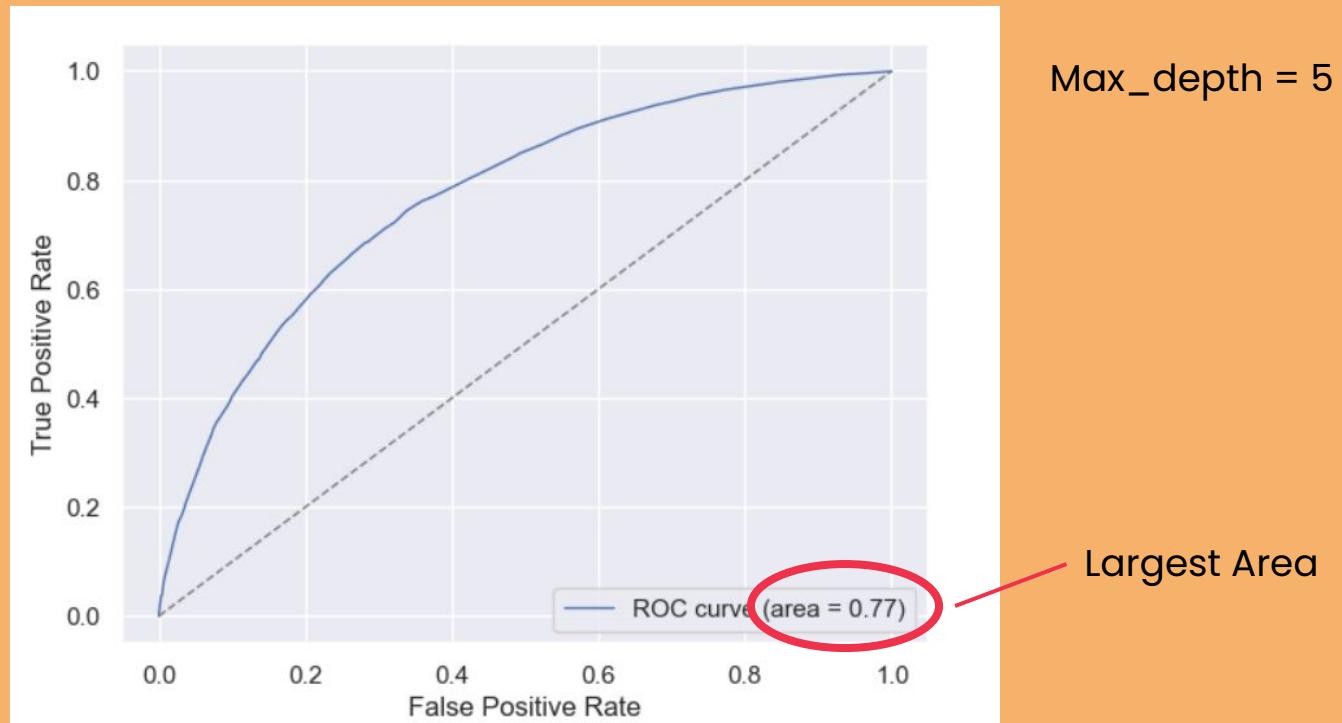
TPR = 0.51514  
FPR = 0.29773

Prediction  
Accuracy  
= 0.60871



# Model 1: Decision Tree

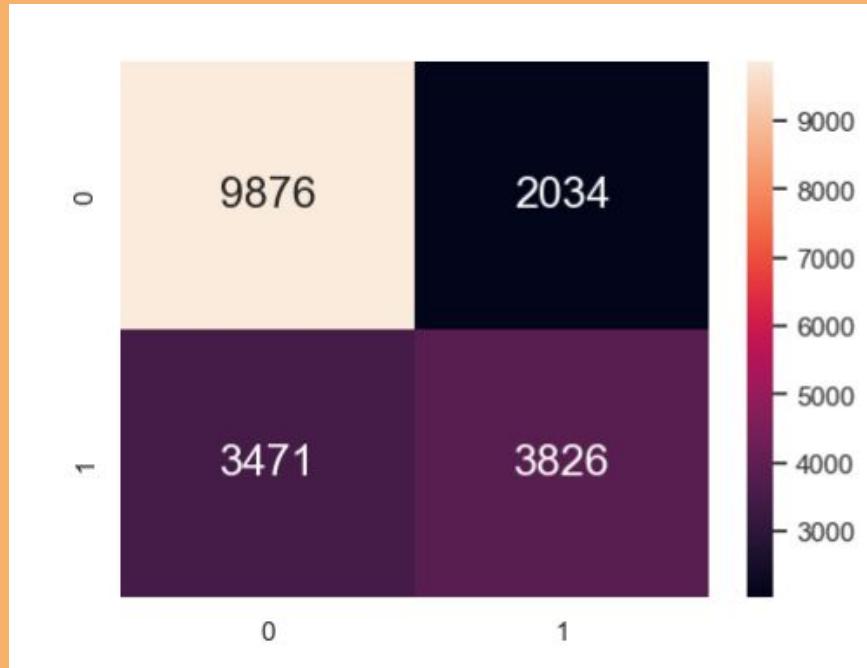
ROC Curve





# Model 1: Decision Tree

Confusion Matrix after  
Optimisation

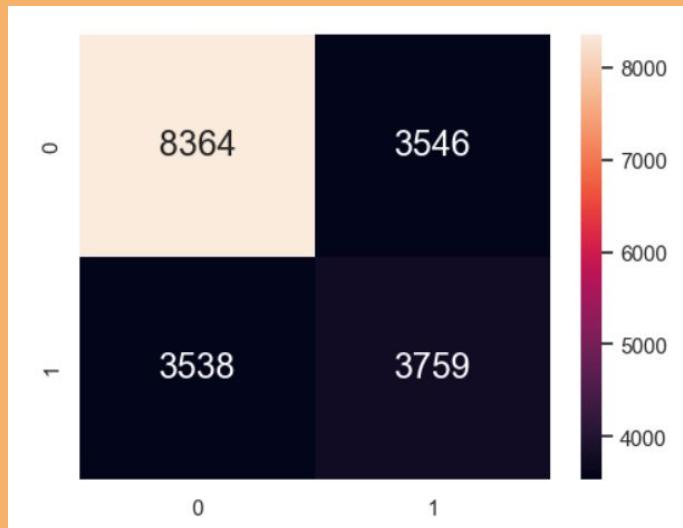


Prediction  
Accuracy  
= 0.67677



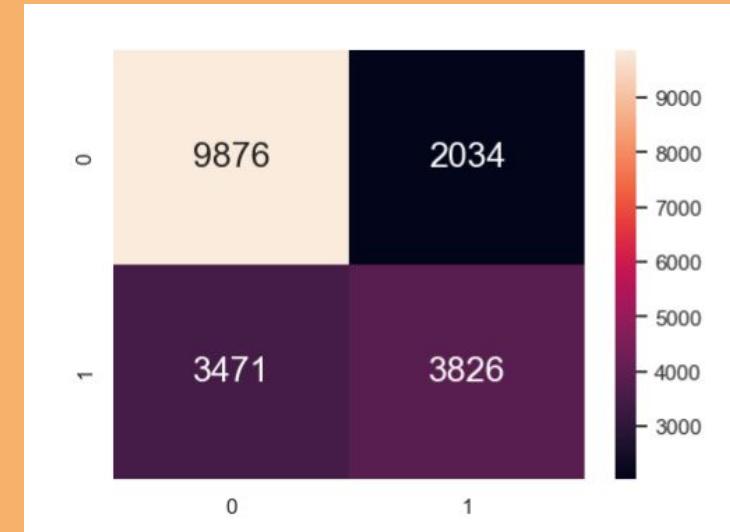
# Model 1: Decision Tree

Confusion Matrix of  
Initial Model



Prediction Accuracy = 0.60871

Confusion Matrix after  
Optimisation



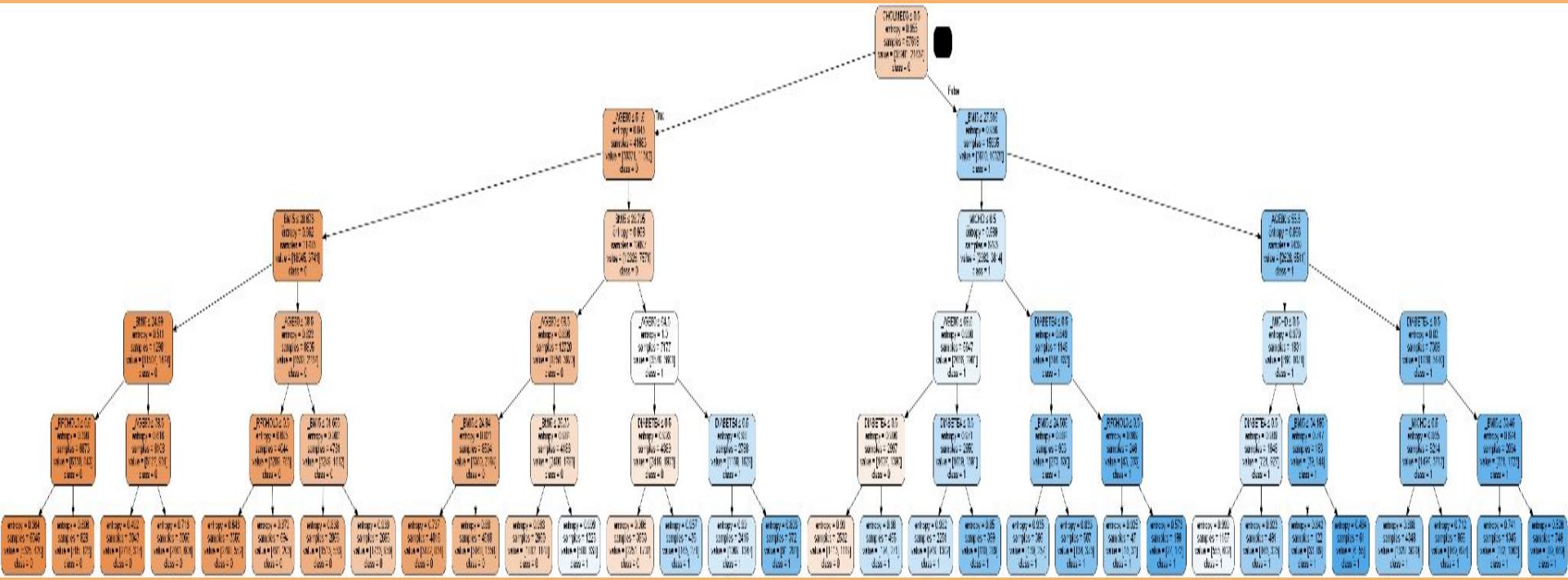
Prediction Accuracy = 0.67677

Improvement: 6.88%



# Model 1: Decision Tree

Optimised Decision Tree





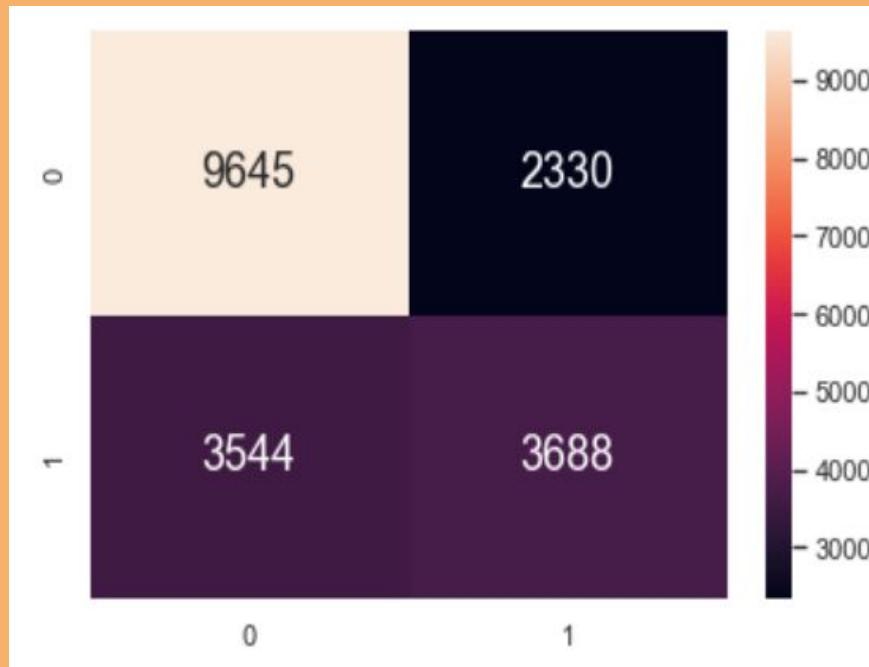
## Model 2: Random Forest

- Random Forest takes random data points from random variables to come up with **multiple decision trees**.
- Multiple decision trees allows the **strengths and weaknesses** of each tree to be balanced out by the other trees.
- The output of each tree is then combined to make a final prediction with a **greater accuracy** than a single decision tree
- Suitable for **large** datasets with a **mix** of categorical and numerical data



## Model 2: Random Forest

Confusion Matrix of  
Initial Model



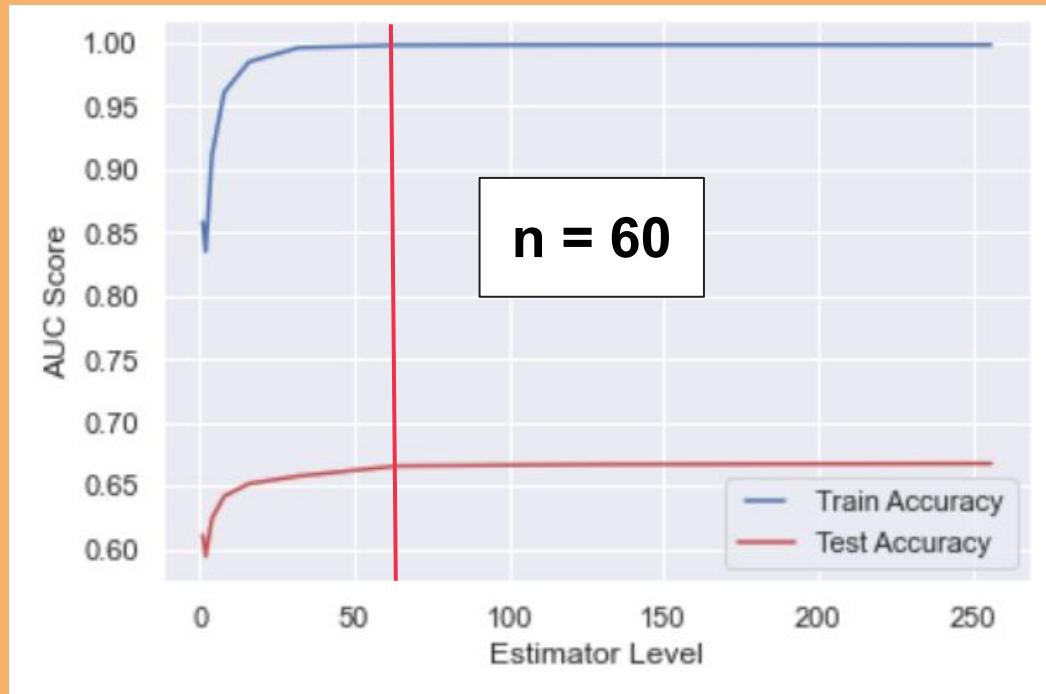
TPR = 0.50996  
FPR = 0.19457

Prediction  
Accuracy  
= 0.69417



## Model 2: Random Forest

ROC Curve & AUC Score





## Model 2: Random Forest

Confusion Matrix after  
Optimisation

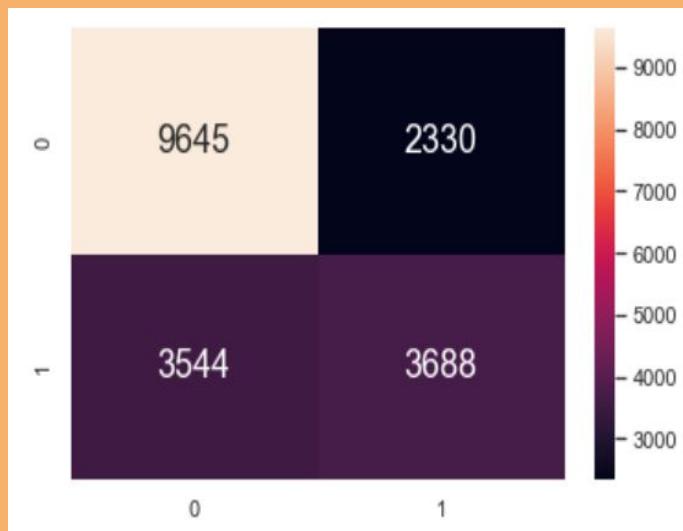


Prediction  
Accuracy  
= 0.70141



## Model 2: Random Forest

Confusion Matrix of  
Initial Model



Prediction Accuracy = 0.69417

Confusion Matrix after  
Optimisation



Prediction Accuracy = 0.70141

Improvement: 0.72%



## Model 3: Logistic Regression

- Logistic Regression **predicts** the output of a categorical variable based on one or more independent variables.
- It **reveals** the interrelationships between different variables and their impact on outcomes
- This helps us make **accurate predictions**



# Model 3: Logistic Regression

Confusion Matrix of  
Initial Model

TPR = 0.52389  
FPR = 0.14882





# Model 3: Logistic Regression

Confusion Matrix after  
Hyperparameter Optimisation

TPR = 0.52401  
FPR = 0.14875



Prediction  
Accuracy  
= 0.73042



# Model 3: Logistic Regression

Confusion Matrix of  
Initial Model



Prediction Accuracy = 0.73033

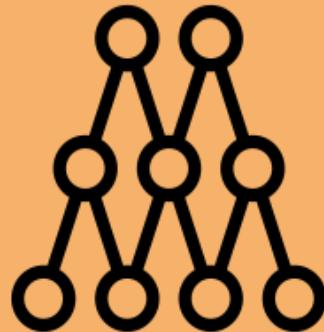
Confusion Matrix after  
Optimisation



Prediction Accuracy = 0.73042

# Best Model: Logistic Regression

Decision Tree



Random Forest



Logistic Regression



<b>Accuracy</b>	0.67677	0.70141	0.73042
<b>TPR</b>	0.52433	0.52959	0.52401
<b>FPR</b>	0.17078	0.19482	0.14875

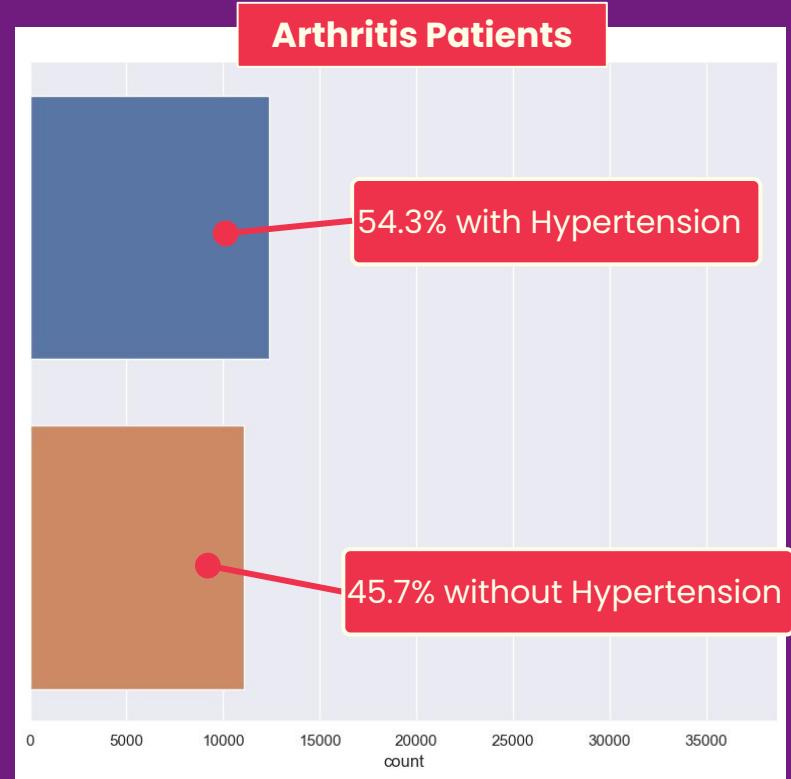
# Findings & Data-Driven Insights



## Insight 1

Affirm

Arthritis is highly correlated with hypertension, but no explanation can be found as of now



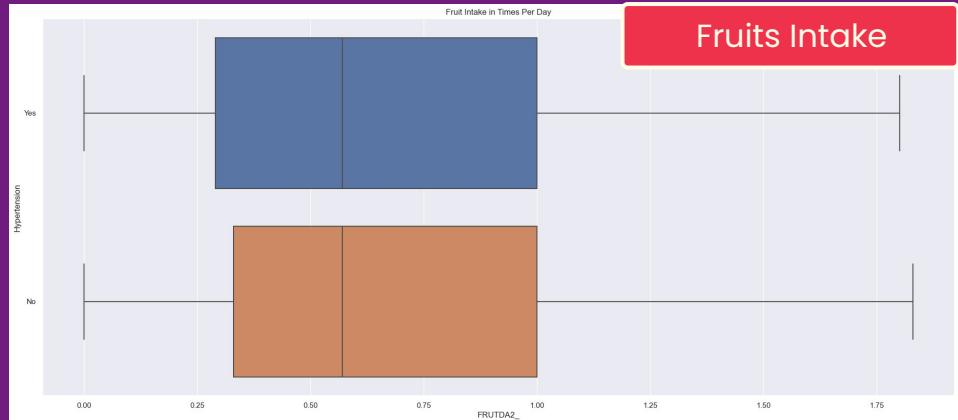
# Findings & Data-Driven Insights



## Insight 2

Debunk

People with and without hypertension have the same average intake of fruits and vegetables



# Findings & Data-Driven Insights

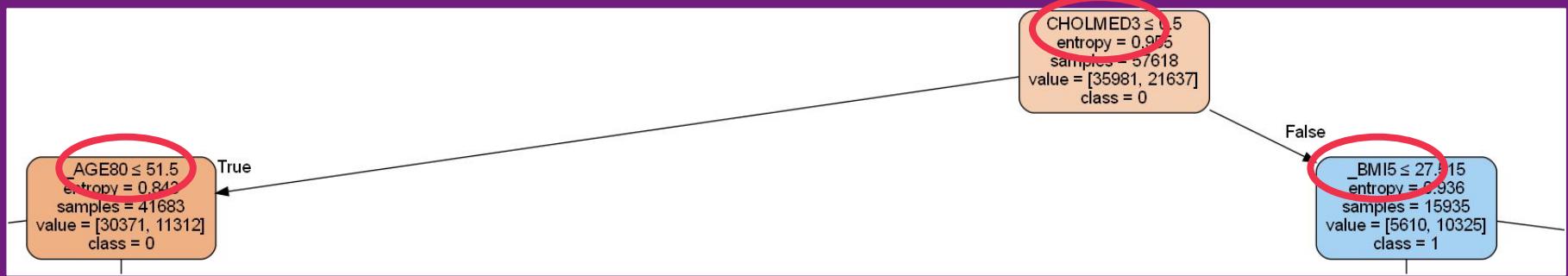


## Insight 3

Discover

For all the races included in the survey except Black, the proportion of people with hypertension is smaller.

# Findings & Data-Driven Insights



## Insight 4

Observe

Most frequently used factors for classification models:

1. Cholesterol
2. BMI
3. Age



# Future Recommendations

## 1. Cholesterol Levels

Focus on reducing consumption of alcohol and food high in saturated fats, thus reducing cholesterol levels

## 2. Body-Mass-Index

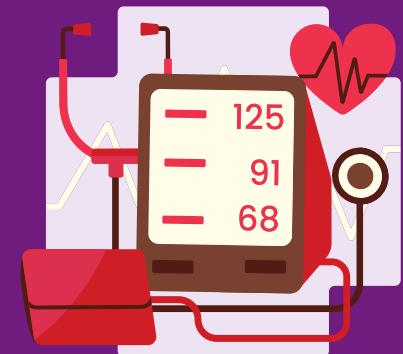
Encourage healthy eating habits and physical exercise

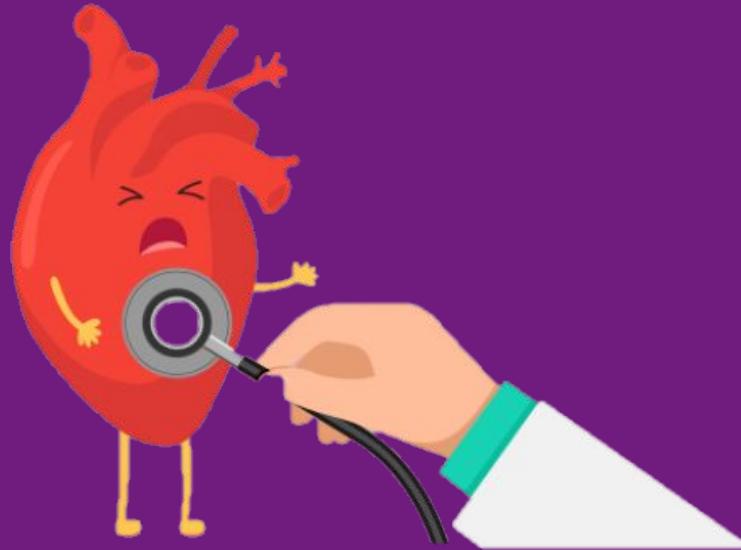
## 3. Age

Health campaigns can be targeted more towards the elderly

# Limitations & Recommendations

- **Data is collected from the US:**  
Demographic and lifestyle factors may differ from Singapore
- **Genetic factors:**  
A more in-depth survey can be conducted to identify factors contributing to hypertension in Singapore





# Thank You!

Done by:

Yong Shao En Ernest (U2221153B)

Wu Rixin (U2221172G)

Li Liyi (U2220985F)

# Findings & Data-Driven Insights



## Debunk

People with and without hypertension have the similar average daily consumption of junk food



Junk Food Consumption

01

# What is hypertension?

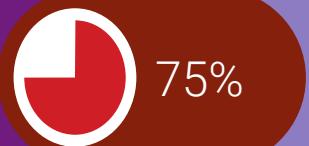
+ You can enter a subtitle here if you need it +



# Content Page

<b>Your Motivation</b>	Motivation Dataset Used Problem Definition
<b>Set The Stage</b>	Data Cleaning Exploratory Data Analysis Initial Insights  (Present your Exploratory Data Analysis and some initial data-driven Insights from the dataset. You MAY also mention how you are planning to set up the Analysis / ML problem for this case. You MUST mention how you collected / curated / cleaned / prepared the data for this problem. Did you only use tools and techniques learned in this course? What ELSE did you learn / try?)
<b>Core Analysis</b>	Machine Learning:  If you used ML (regression, classification, or something else); mention mainly WHICH one(s). You may now briefly CLARIFY why and how the ML problem(s) aim(s) to solve your objective. How did you apply ML technique(s) to SOLVE your problem? Which model(s), how and why? Did you only use tools and techniques learned in this course? What ELSE did you learn / try?
<b>Finish Strong</b>	Findings, Insights & Conclusion  What is the OUTCOME of your project? Did it solve your original problem? Anything interesting? What are your data-driven INSIGHTS and

# Initial Insights



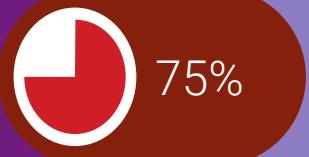
- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat

- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat



- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat

- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat



- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat

- Donec risus dolor porta venenatis
- Pharetra luctus felis
- Proin vel tellus in felis volutpat



A picture is worth  
a thousand words

# Countries with the highest incidences



## Mercury

It's the closest planet to the Sun

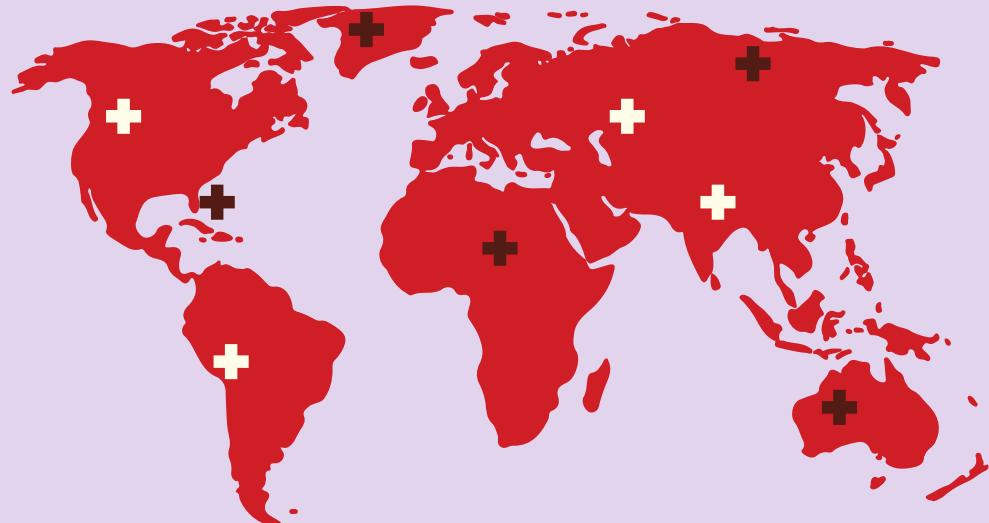


## Mars

Mars is actually a very cold place

**50,000,363**

total cases



# Treatment

+ 01

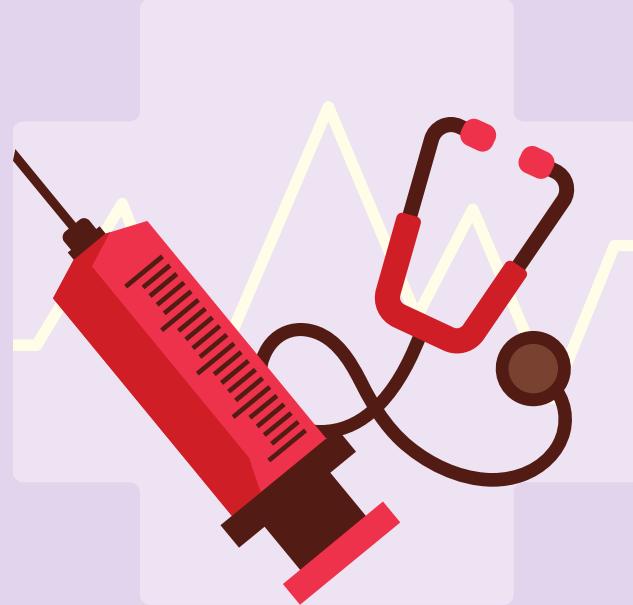
## Mercury

It's the closest planet to the Sun and the smallest in the Solar System

+ 02

## Venus

Venus has a beautiful name and is the second planet from the Sun



# **Awesome words**

# 150,000

- + Big numbers catch your audience's attention +

# Symptoms

**Mercury**

It's the closest planet to the Sun

**Venus**

Venus is the second planet from the Sun

**Mars**

Mars is actually a very cold place

**Jupiter**

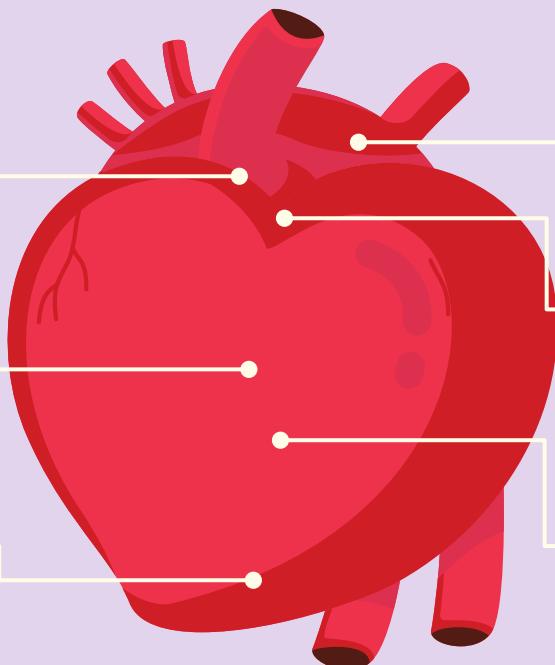
Jupiter is the biggest planet of them all

**Saturn**

Saturn is composed of hydrogen

**Neptune**

It's the farthest planet from the Sun





## Other symptoms you should know

Mercury is the closest planet to the Sun and the smallest one in the Solar System. This planet's name has nothing to do with the liquid metal, since Mercury was named after the Roman messenger god

# Findings

## Evolution



Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

## Jupiter

Jupiter is the biggest planet of them all

## Neptune

It's the farthest planet from the Sun

## Saturn

It's a gas giant and has several rings

125,000

Big number

# Expert recommendations

Jupiter



20%

Jupiter is the biggest planet of them all

Neptune



80%

It's the farthest planet from the Sun



75%

A circular progress bar with a red outer ring and a white inner circle, indicating 75% completion.

Venus



Venus is the second planet from the Sun



Mercury

It's the smallest planet of them all



50%

A circular progress bar with a red outer ring and a white inner circle, indicating 50% completion.

# Physical examination

+ 01

## Venus

Venus is the second planet from the Sun

## Jupiter

It's the biggest planet in the Solar System

+ 03

## Mars

Despite being red, Mars is a cold place

## Saturn

Saturn is a gas giant and has several rings

02 +

# Comparaison



Symptoms	Venus	Mars	Jupiter
Venus is the second planet from the Sun	Yes	Yes	No
Saturn is composed of hydrogen and helium	Yes	No	Yes
Jupiter is the biggest planet of them all	No	No	Yes

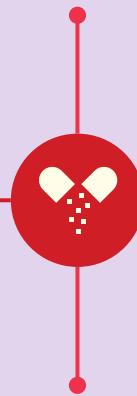
# Disease evolution

**Year 1**



Mercury is the  
smallest planet

**Year 2**



Jupiter is the  
biggest planet

**Year 3**



Mars is actually a  
very cold place

**Year 4**



Neptune is very  
far from the Sun

# Alternative treatments



## Mercury

Mercury is the closest planet to the Sun



## Jupiter

Jupiter is the biggest planet of them all



## Venus

Venus is the second planet from the Sun



## Saturn

It's composed of hydrogen and helium



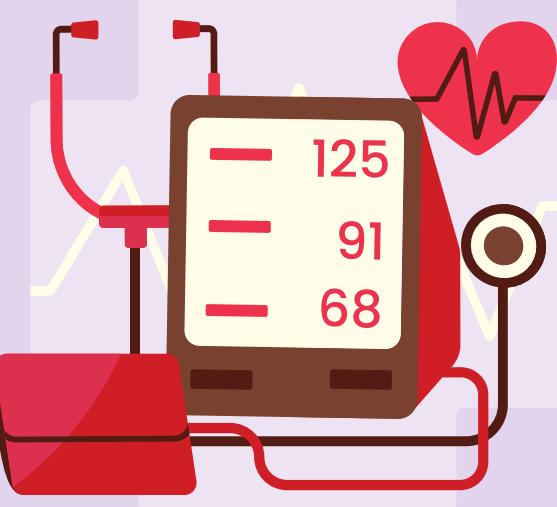
## Mars

Mars is actually a very cold place



## Neptune

It's the farthest planet from the Sun



04

# How to prevent hypertension?

+ You can enter a subtitle here if you need it +

# Preventions



## Dr. Lisa Jones

"Mercury is the closest planet to the Sun and the smallest of them all"



## Dr. Louis Smith

"Neptune is the farthest planet from the Sun. It's also the fourth-largest"



## Dr. John Lewis

"Venus has a beautiful name and is the second planet from the Sun"



## Dr. Jena Doe

"Mars is full of iron oxide dust, which gives the planet its reddish cast"

# Post-preventions



50%

## Jupiter

It's the biggest planet of them all



10%

## Mars

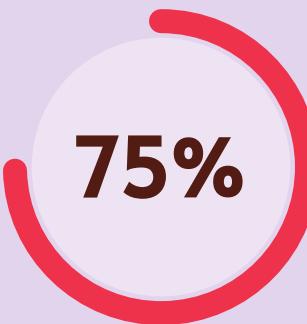
Mars is actually a very cold place



25%

## Mercury

It's the closest planet to the Sun

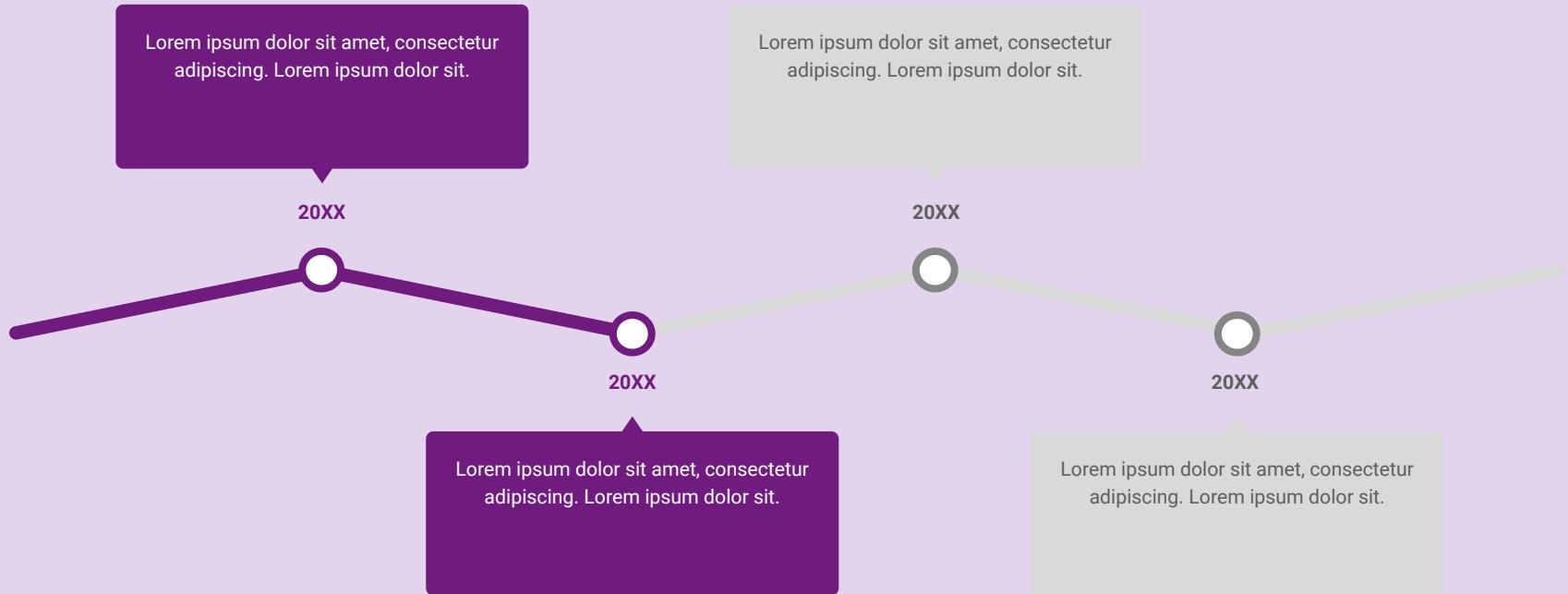


75%

## Neptune

It's the farthest planet from the Sun

# Exploratory Data Analysis



# Contraindications and indications

# You shouldn't



# You should

# Research in progress

## Mercury



2005

It's the closest planet to the Sun

## Mars



2010

Despite being red, Mars is a cold place

## Venus



2015

Venus is the second planet from the Sun

## Jupiter



2017

It's the biggest planet of them all

## Pluto



2019

It's now considered a dwarf planet

## Neptune



2021

It's the farthest planet from the Sun

# 50%

+ Despite being red,  
Mars is a cold place +

# 150,000

+ Mars is actually a  
very cold place +

# 20,000

+ Venus is the second  
planet from the Sun +

# 75%

+ Neptune is very far  
from the Sun +

# Conclusions

Venus has a beautiful name and is the second planet from the Sun. It's terribly hot, even hotter than Mercury, and its atmosphere is extremely poisonous. It's the second-brightest natural object in the night sky after the Moon

# Our team



**Timmy Jimmy**

You can speak a bit  
about this person here



**Jenna Doe**

You can speak a bit  
about this person here

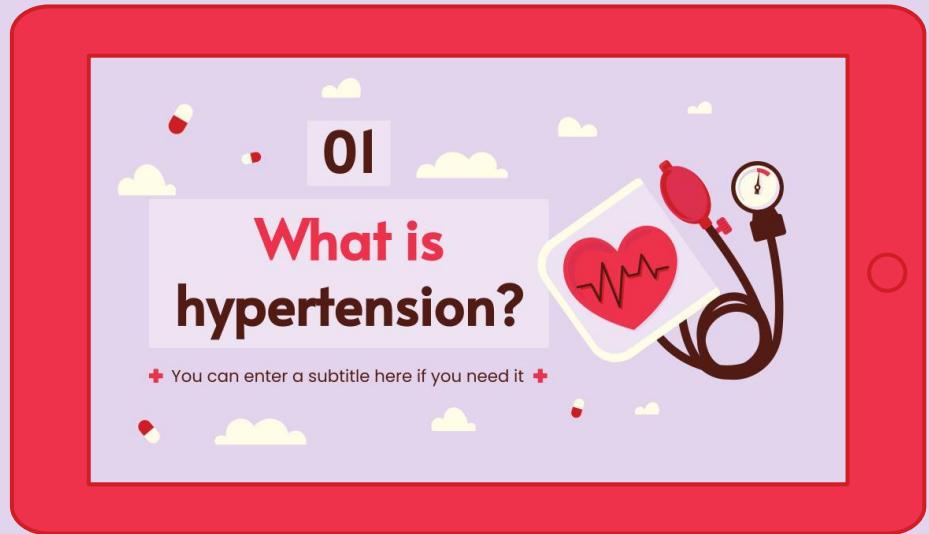
## More information in our web

You can replace the image on the screen with your own work. Just right-click on it and select "Replace image"



### World Hypertension Day

+ Here is where your presentation begins +



## Get now our tablet app

You can replace the image on the screen with your own work. Just right-click on it and select "Replace image"

# Get now our mobile app

You can replace the image on the screen with your own work. Just right-click on it and select "Replace image"



# References

- AUTHOR (YEAR). *Title of the publication*. Publisher



# Thanks!

**+ Do you have any questions? +**

[youremail@freepik.com](mailto:youremail@freepik.com)

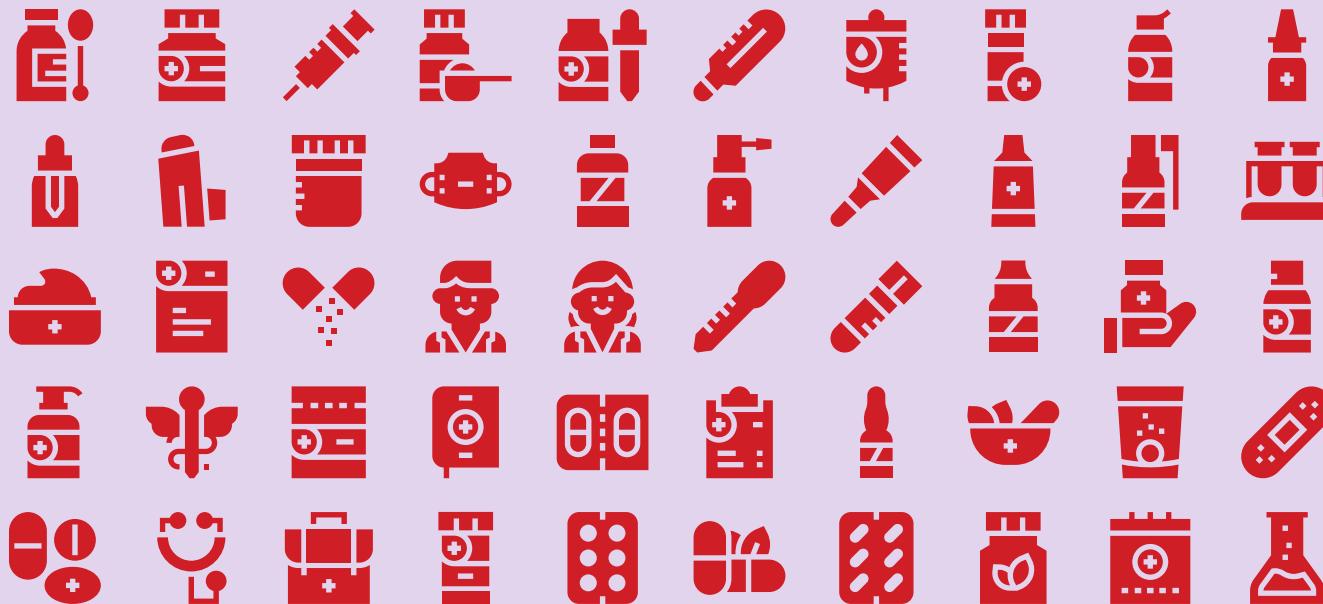
+91 620 421 838

[yourcompany.com](http://yourcompany.com)



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

# Medicaments icon pack



# Resources

Did you like the resources on this template? Get them for free at our other websites:

## Vectors

- World health day illustration with heart
- Flat world hypertension day illustration
- Flat world hypertension day illustration
- Flat world hypertension day illustration
- Hand drawn world hypertension day illustration
- World health day illustration with clipboard and first aid kit

# Resources

Did you like the resources on this template? Get them for free at our other websites:

## Photos

- Close up on health worker
- Medium shot doctor with stethoscope
- Medium shot smiley doctor with coat
- Young male pediatrician working at his office
- Blonde female doctor in her office
- Portrait of smiling doctor sitting on desk
- Doctors talking and checking notes
- Close-up of a man's hand sitting on bed measures his blood pressure gauge

## Icon

- Medicaments Icon Pack

# Instructions for use

If you have a free account, in order to use this template, you must credit Slidesgo by keeping the Thanks slide. Please refer to the next slide to read the instructions for premium users.

## **As a Free user, you are allowed to:**

- Modify this template.
- Use it for both personal and commercial projects.

## **You are not allowed to:**

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Instructions for use (premium users)

As a Premium user, you can use this template without attributing Slidesgo or keeping the "Thanks" slide.

## You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the "Thanks" slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

## You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Fonts & colors used

This presentation has been made using the following fonts:

## **Alata**

(<https://fonts.google.com/specimen/Alata>)

## **Roboto Slab**

(<https://fonts.google.com/specimen/Poppins>)

#521c15

#fffde8

#ef324b

#ef324b

#cf1e25

#794230

#eee3f3

##eadcf1

#e3d4ed

# Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out how it works.



Pana



Amico



Bro



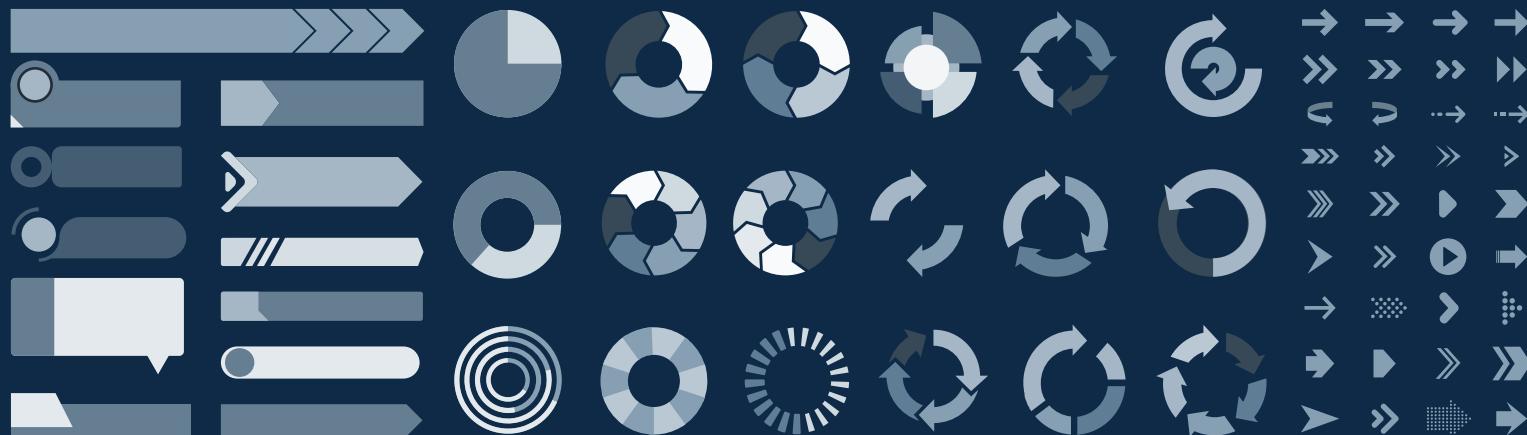
Rafiki

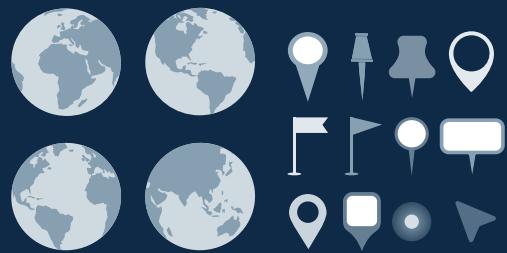


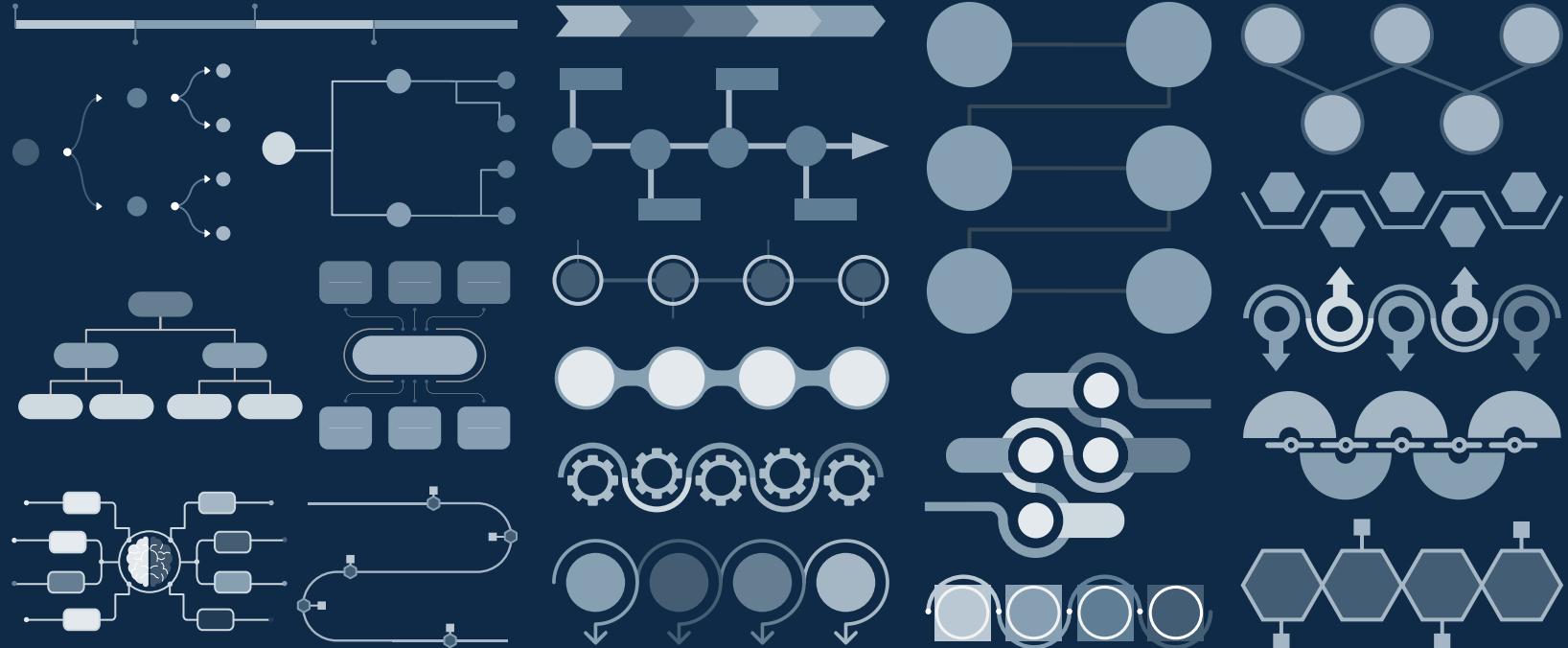
Cuate

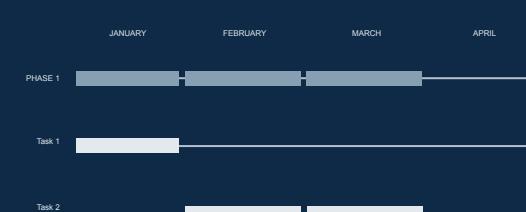
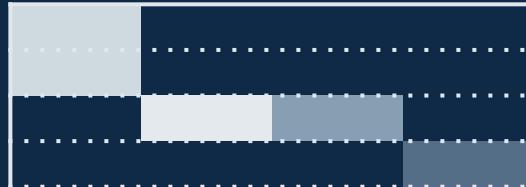
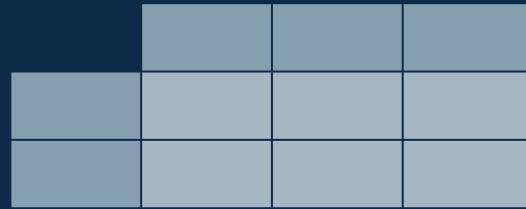
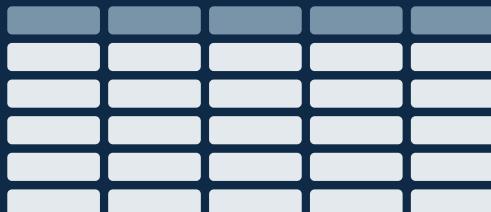
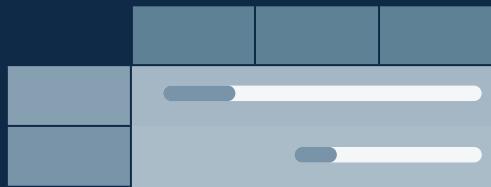
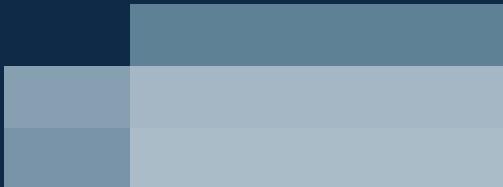
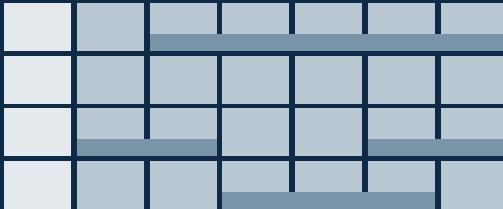
# Use our editable graphic resources...

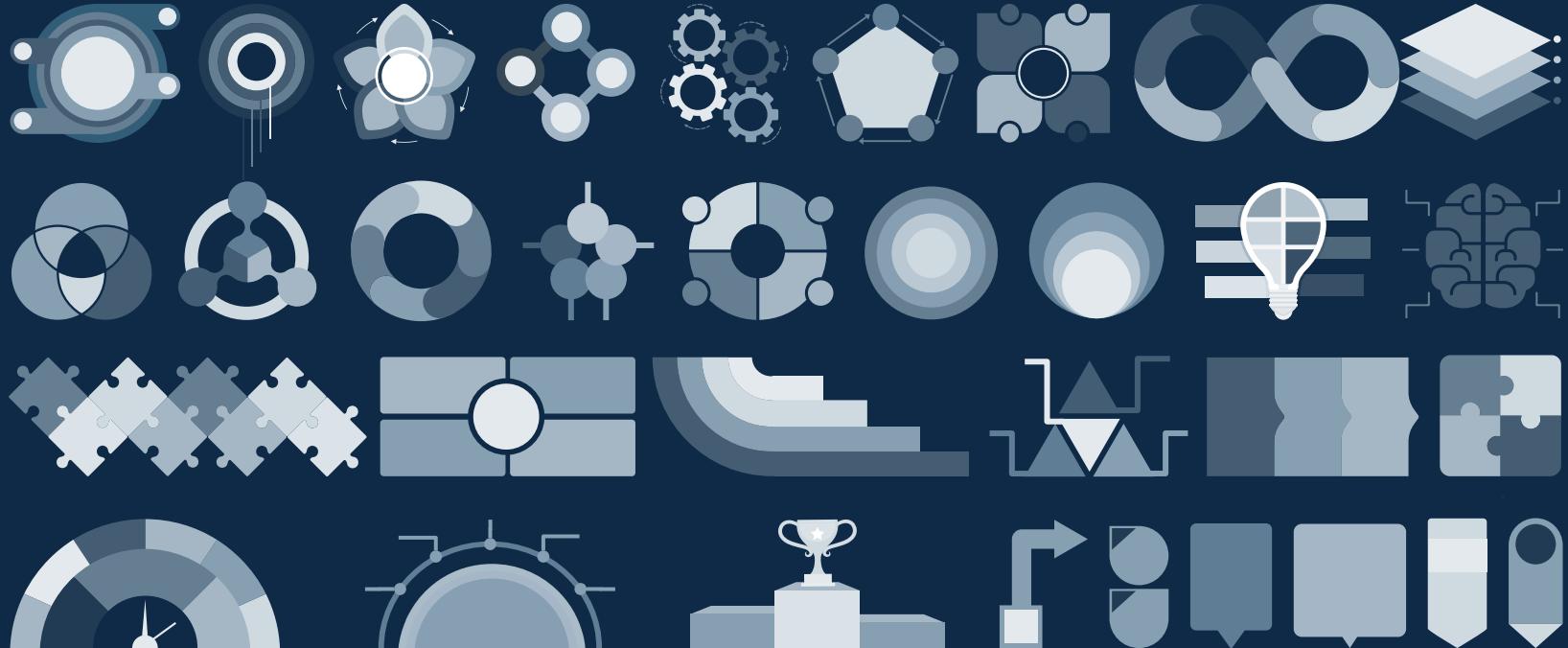
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on Slidesgo.

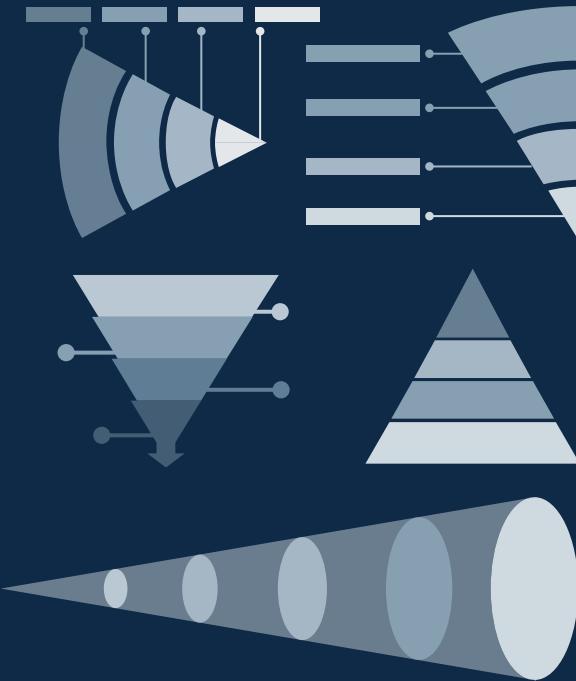
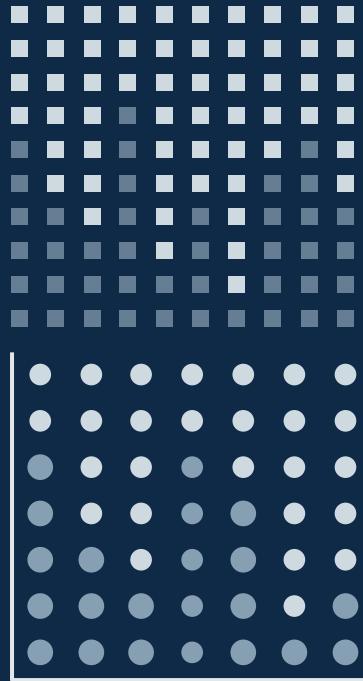












# ...and our sets of editable icons

You can resize these icons without losing quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



## Educational Icons



## Medical Icons



## Business Icons



## Teamwork Icons



## Help & Support Icons



## Avatar Icons



## Creative Process Icons



## Performing Arts Icons



# Nature Icons



# SEO & Marketing Icons



