

Heart Diseases

Supervised by
Dr.Doaa Mohamed

Facilitator
Eng.Haneen Hossam

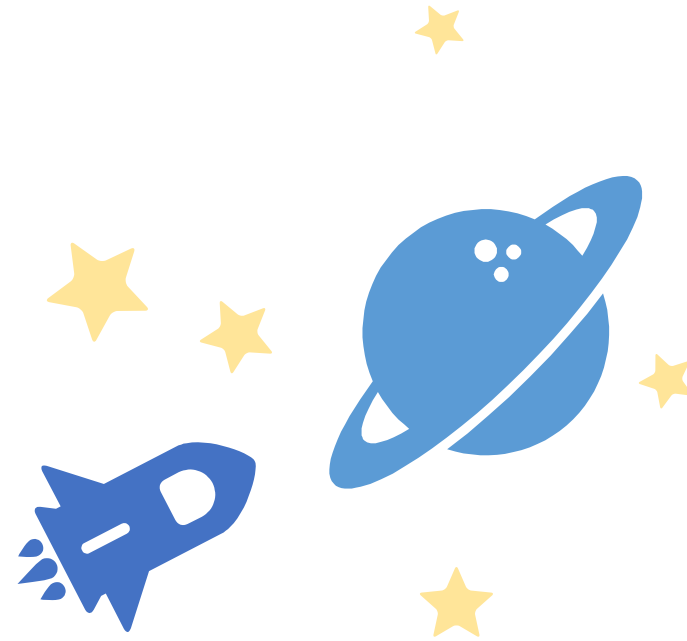
Team Members

- **Ahmed Mohsen**
- **Yomna Ramdan**
- **Hossam Galal**

Agenda:

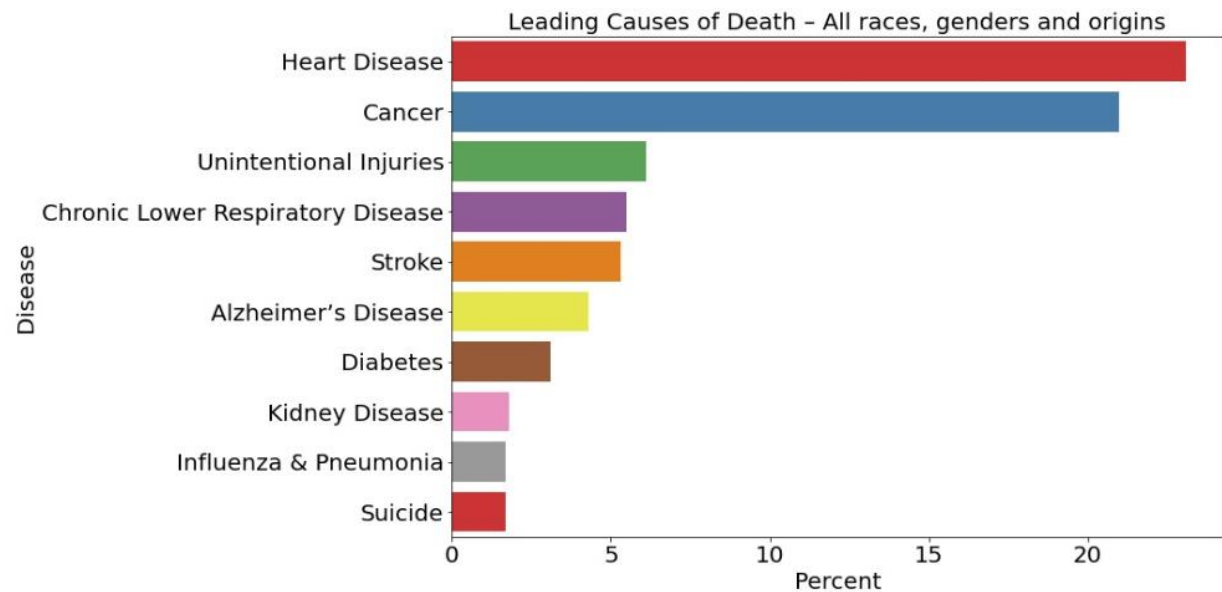
- Motivation
- About the problem
- Understanding data
- Business questions and Visualization
- Preprocessing & Modelling
- Business Solution
- Conclusions

1 Motivation



Motivation

- Heart disease can last a long time in a latent form, clinically not manifesting itself.

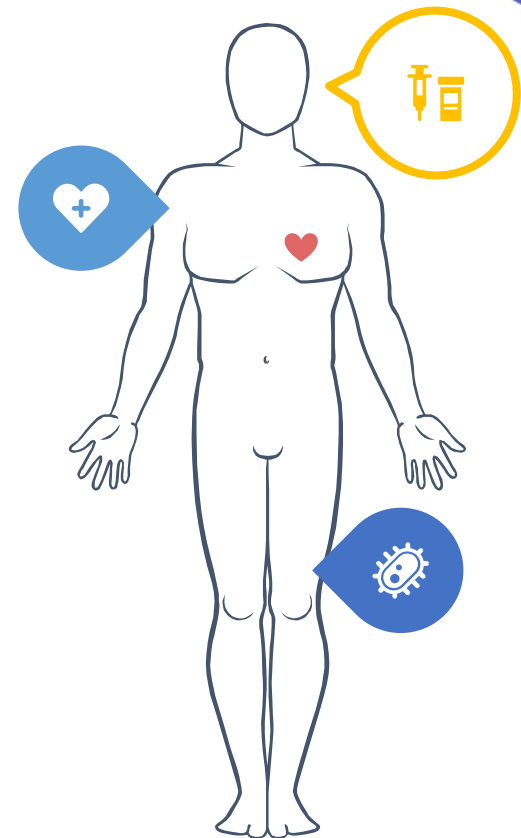


2

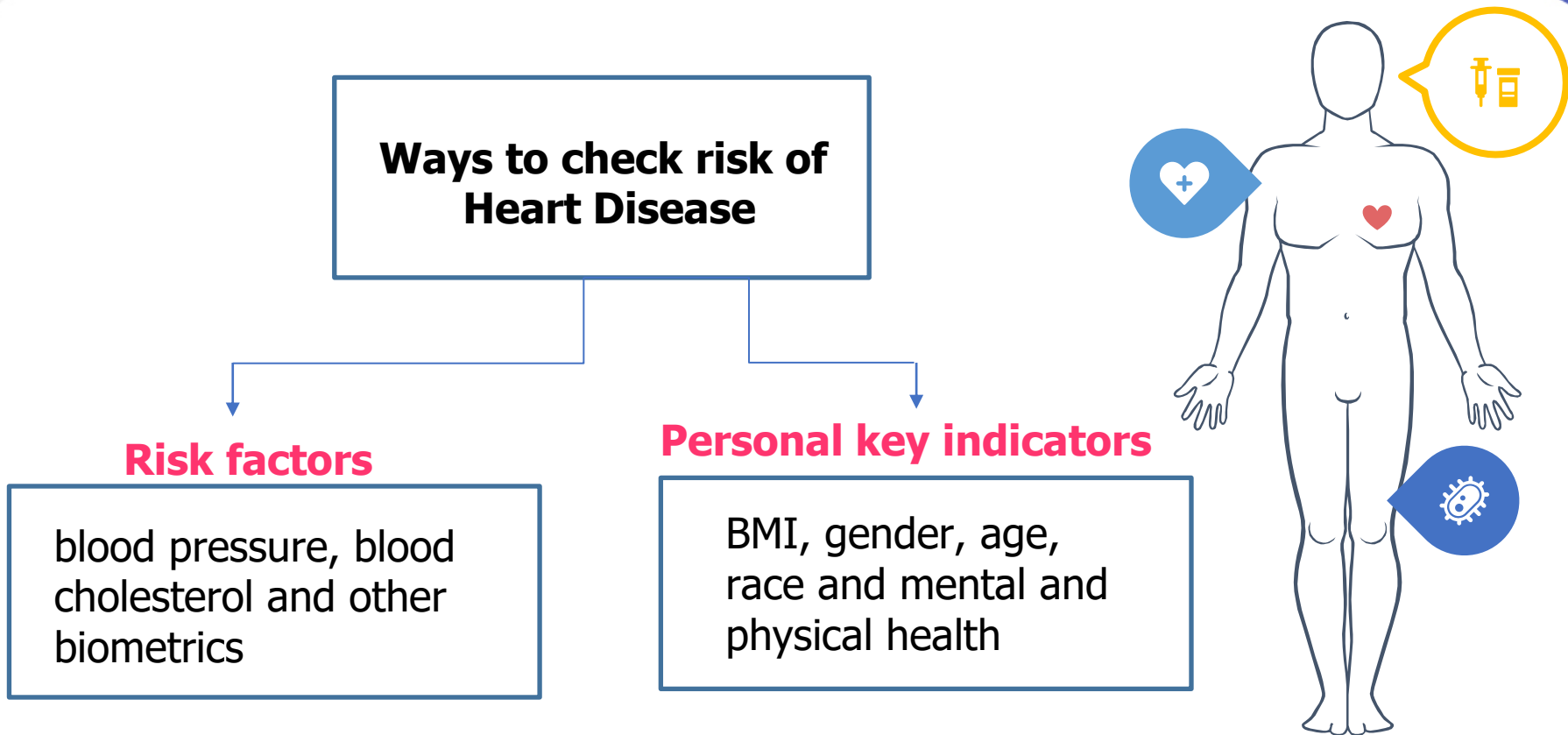
About the problem

About the problem

- The term “heart disease” refers to several types of heart conditions. our data represents the most common types, which are coronary artery disease (CAD) and Myocardial infarction (MI) which affects the blood flow to the heart.
- Decreased blood flow can cause a heart attack.

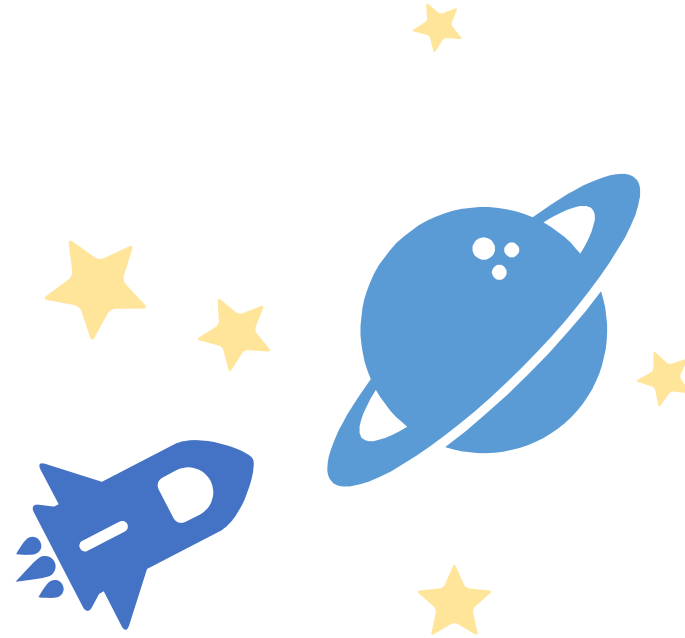


About the problem



3

Objectives



Objectives

- Find what personal factors may increase the risk of getting heart disease
- Increase awareness of people to Heart Disease and how to reduce the chance of getting it

4

Understanding Data

Understanding Data

BMI

Body mass index is a measure of body fat based on height and weight

Smoking

Have you smoked at least 100 cigarettes in your entire life?

Alcohol Drinking

Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week

Stroke

(Ever told) (you had) a stroke?

Physical Health

Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30

Mental Health

I Thinking about your mental health, for how many days during the past 30 days was your mental health not good?

DiffWalking

Do you have serious difficulty walking or climbing stairs?

Sex

Are you male or female?

Age Category

Fourteen-level age category

BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory
16.60	Yes	No	No	3.0	30.0	No	Female	55-59
20.34	No	No	Yes	0.0	0.0	No	Female	80 or older
26.58	Yes	No	No	20.0	30.0	No	Male	65-69

Understanding Data

Race

Imputed race/ethnicity value

Diabetic

(Ever told) (you had) diabetes?

Physical Activity

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

GenHealth

General health condition

Sleep Time

On average, how many hours of sleep do you get in a 24-hour period?

Asthma

(Ever told) (you had) asthma?

Kidney Diseases

Not including kidney stones, bladder infection or incontinence, were you ever told you had kidney disease?

Skin Cancer

(Ever told) (you had) skin cancer?

Race	Diabetic	PhysicalActivity	GenHealth	SleepTime	Asthma	KidneyDisease	SkinCancer
White	Yes	Yes	Very good	5.0	Yes	No	Yes
White	No	Yes	Very good	7.0	No	No	No
White	Yes	Yes	Fair	8.0	Yes	No	No

5

Business Question & Visualization

- **Are there any Personal key factors which indicate the risk as efficient as risk factors**

KidneyDisease Risk for Heart Disease

People with KidneyDisease are significantly more prone to heart disease

Heart Disease|No Heart Disease

Have KidneyDisease (4%) Don't have KidneyDisease(96%)



29%

8%

Diabetic Risk for Heart Disease

People with Diabetic are significantly more prone to heart disease

Heart Disease|No Heart Disease

Have Diabetic (13%) Don't have Diabetic(87%)



22%

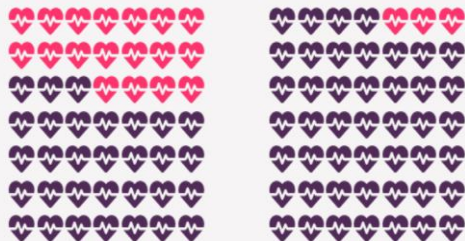
6%

Stroke Risk for Heart Disease

People with Stroke are significantly more prone to heart disease

Heart Disease|No Heart Disease

Have Stroke (4%) Don't have Stroke(96%)



36%

7%

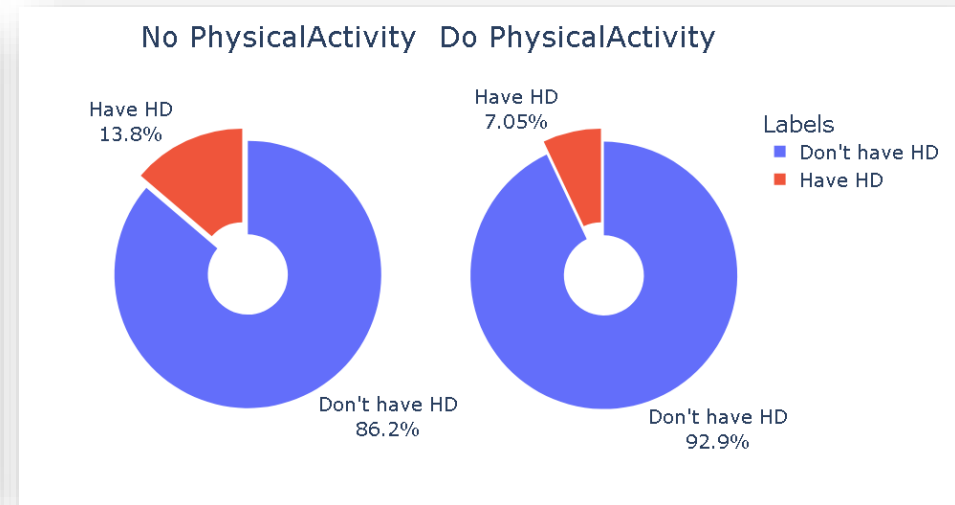
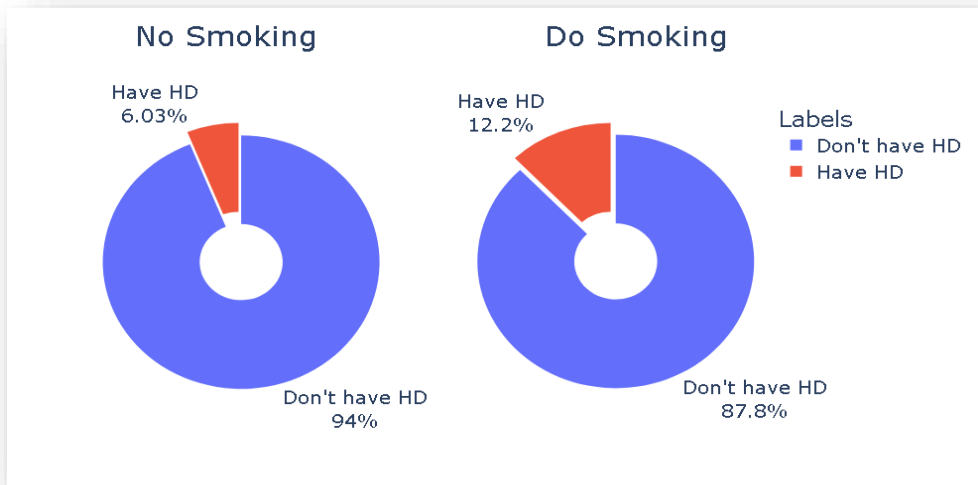
Insight:

People with common causes diseases are affected with Heart Disease up to 3 to 4 times

Solution:

The group affected by these diseases should be more careful than others

- **What are usual activities that people do and can increase the chance of getting Heart Diseases ?**



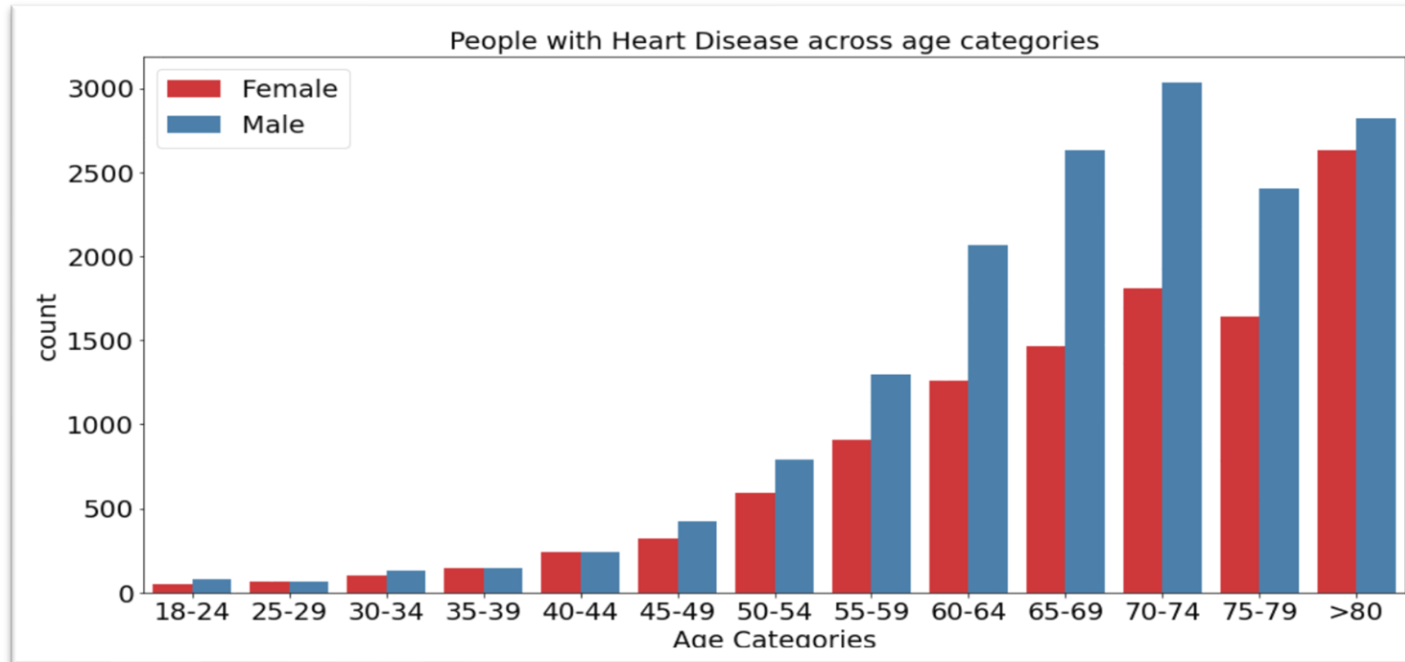
Insight:

Smokers and inactive people are twice as likely to get Heart Disease, Alcohol Drinking turned out to be not a factor

Solution:

Educating people about the importance of physical activates and quitting smoking

- Who is more vulnerable to Heart Diseases between genders across all ages?



Insight:

- As people get older, their risk of heart disease is increased significantly
- Males are clearly more prone to heart disease

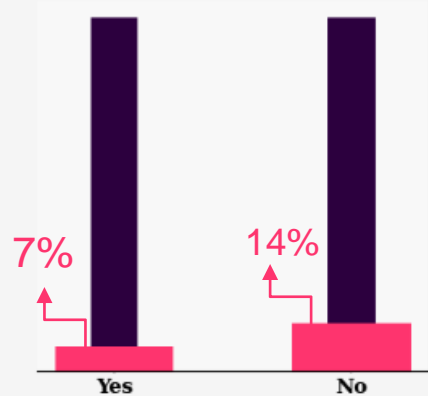
Solution:

Aged people should have a regular check-up to check their health especially males

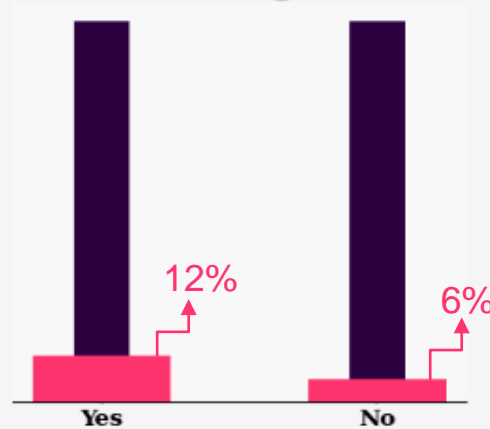
- What are the most habits/factors that can be modified and have high impact in getting Heart Disease?

- Heart Disease percentage

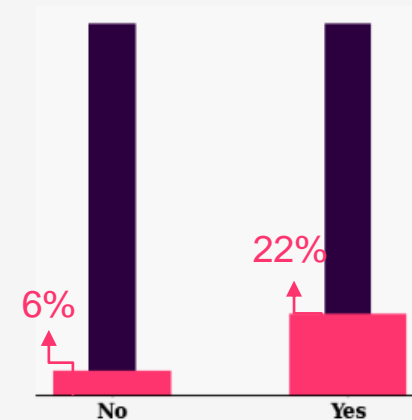
Physical Inactivity Risk



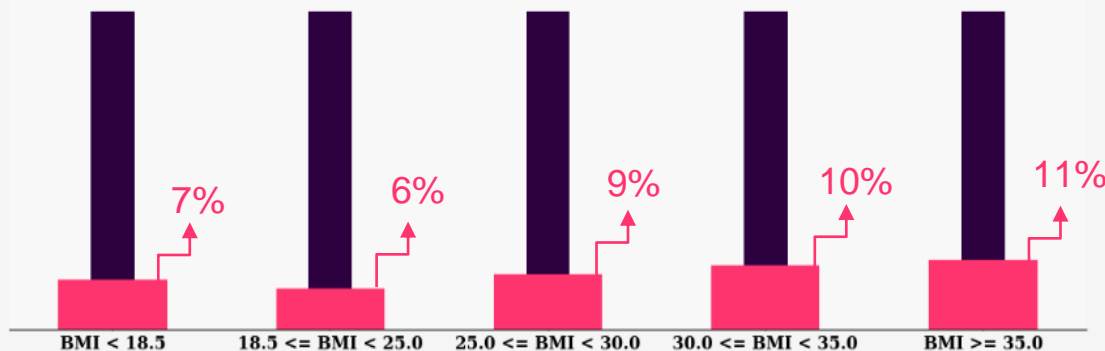
Smoking Risk



Diabetic Risk



BMI Risk



Insight:

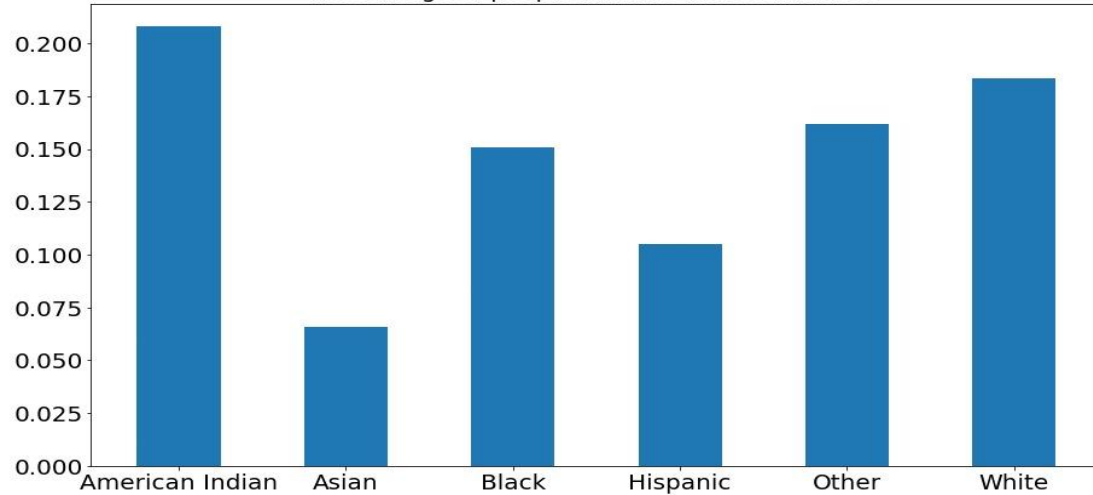
NOT managing Diabetic, Physical Activity, BMI and smoking and can increase the risk of getting Heart Disease up to 2-3 times

Solution:

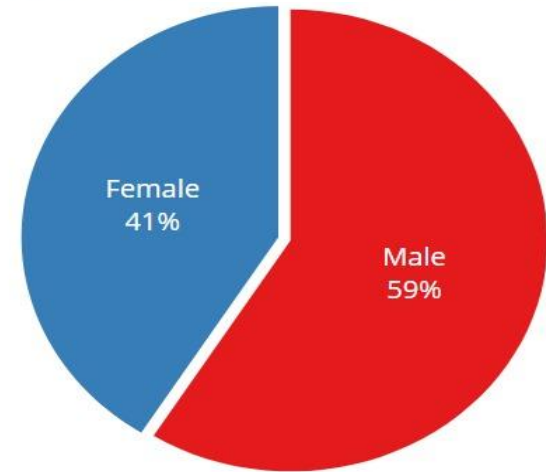
A balanced, calorie-controlled diet is the ticket to a healthy BMI and can decrease the Diabetic risk as well.

If I don't have those bad habits, am I safe?

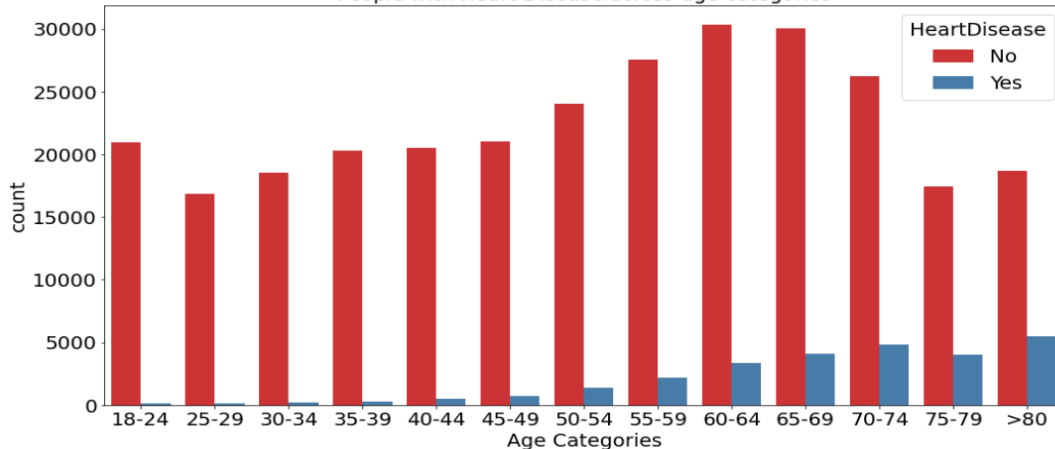
Percentage of people with HD across all Races



Percentage of people with heart disease for each Gender



People with Heart Disease across age categories



Insight:

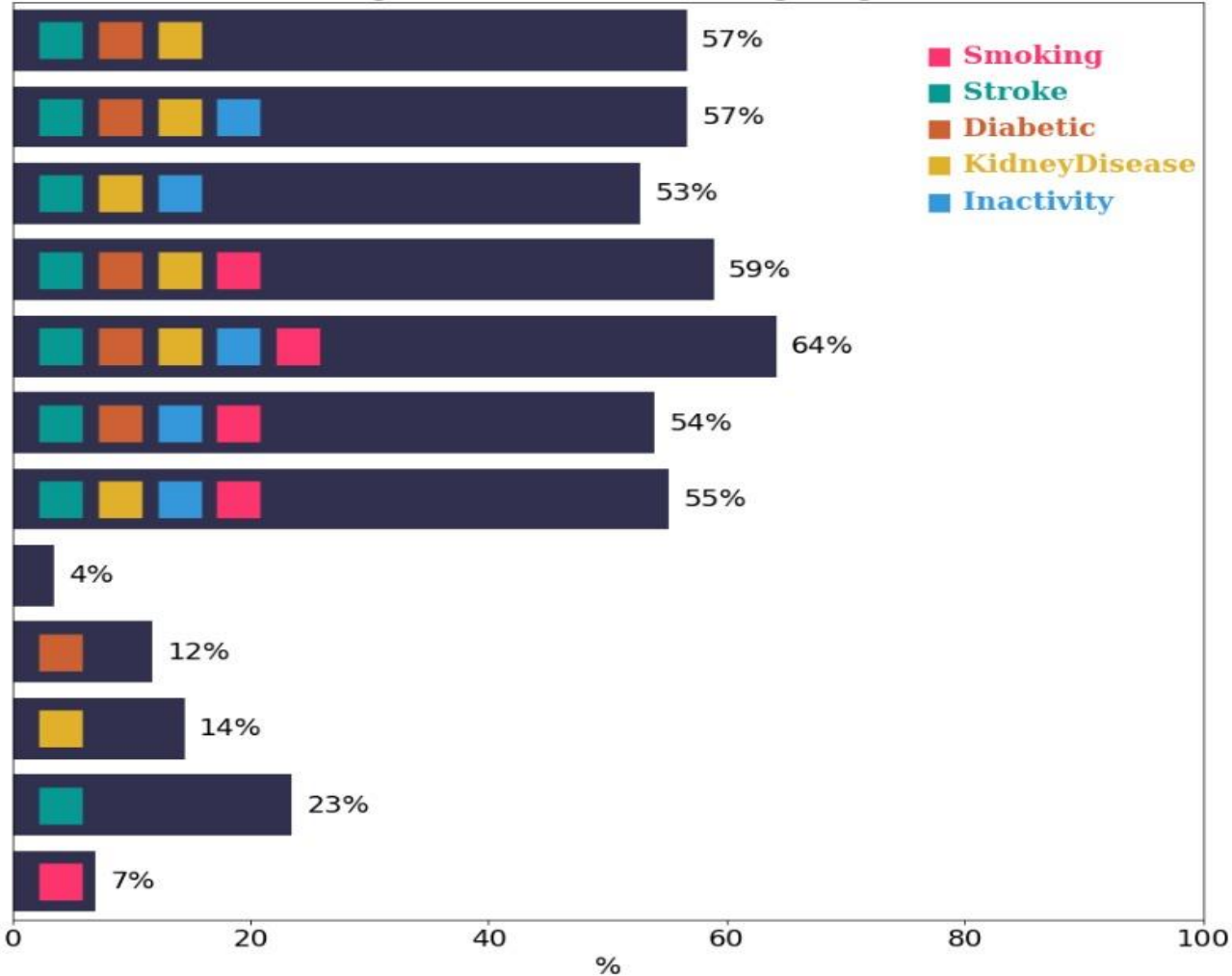
it's shown that there is some non-modifiable factors which can increase the risk of getting Heart Disease like: Gender, Age, Race

Solution:

Increasing people's awareness more in areas that contain the most vulnerable races

What is the risk of getting Heart Disease if you are exposed to more than 1 factor?

Having more than 1 factor risk of getting HD



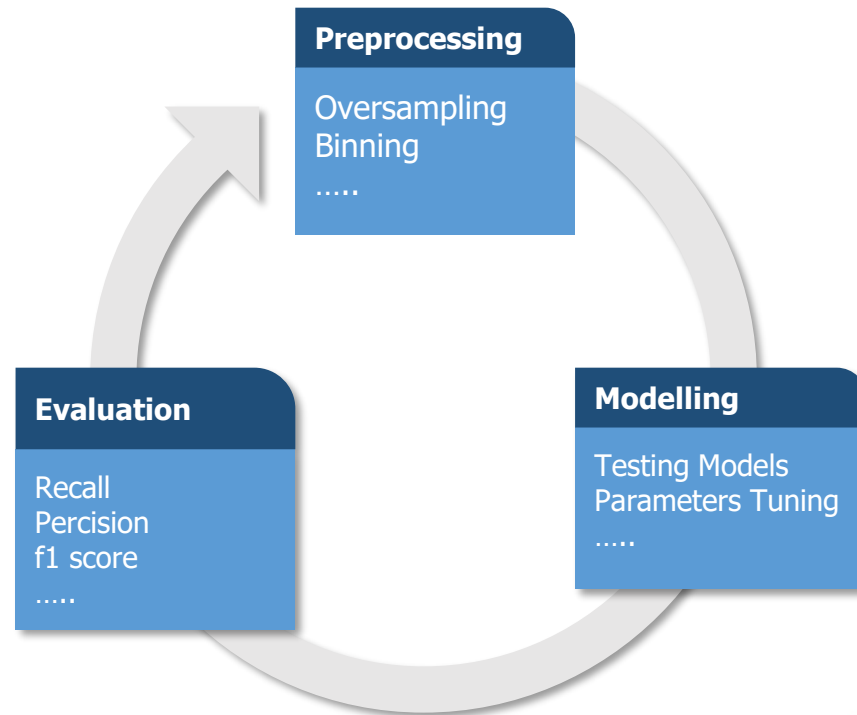
Insight:

Having more than 1 factor can increase the chance of getting Heart Disease significantly

Solution:

People affected with more than one factor must necessarily do a check-up to ensure their safety

Pre-processing & Modelling



Pre-processing

● Missing values

HeartDisease	0
BMI	0
Smoking	0
AlcoholDrinking	0
Stroke	0
PhysicalHealth	0
MentalHealth	0
DiffWalking	0
Sex	0
AgeCategory	0
Race	0
Diabetic	0
PhysicalActivity	0
GenHealth	0
SleepTime	0
Asthma	0
KidneyDisease	0
SkinCancer	0

Data is Clean , don't have missing values

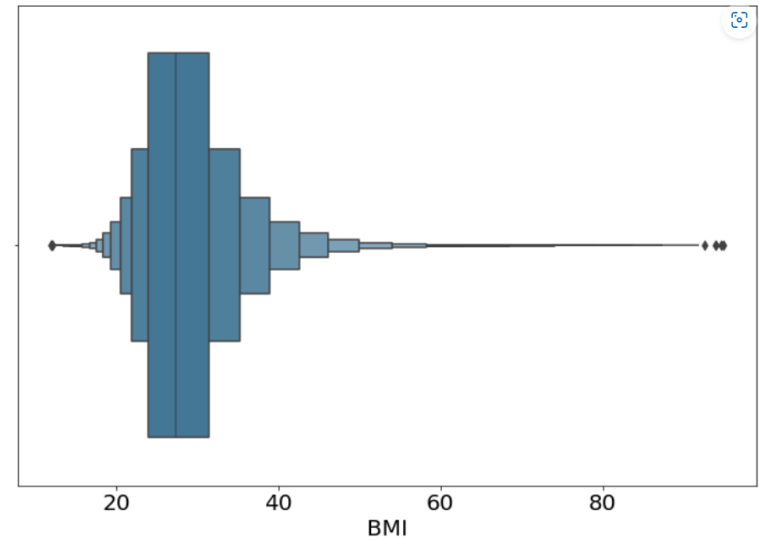
● Duplicated data

Number of duplicated records = 18078

It is obvious to have duplicate records because almost all features are categorical and

● Duplicated data

The only column could have outliers is BMI, and its values ranging from 12-94.8 which is a real range and we can consider not outliers on it



Pre-processing

● Binning

Overweight ($25.0 \leq \text{BMI} < 30.0$)	114355
Normal weight ($18.5 \leq \text{BMI} < 25.0$)	97778
Obese ($30.0 \leq \text{BMI} < 35.0$)	61169
Extremely Obese ($\text{BMI} \geq 35.0$)	41379
Underweight ($\text{BMI} < 18.5$)	5114

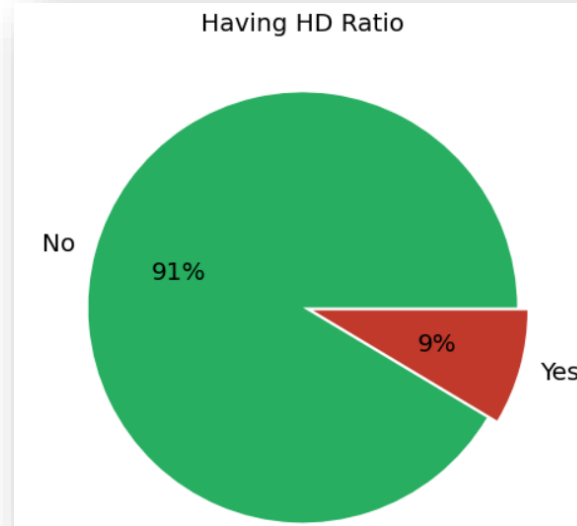
In real life, BMI is divided into finite categories, so for better insights and to increase efficiency, BMI is categorized using binning

● Encoding

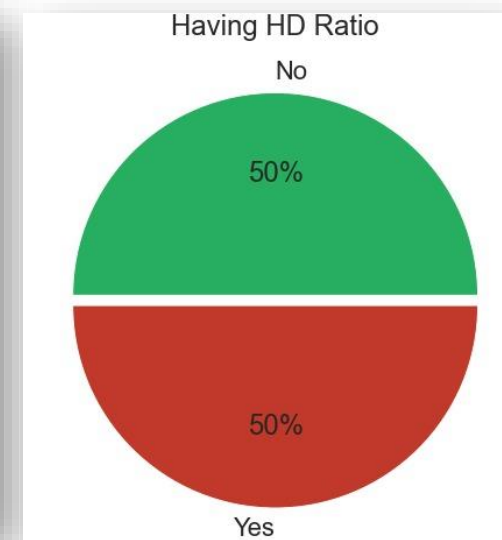
- Encoding features with no ordering by one-hot encoding
- Encoding features with ordering by label encoding

● Imbalancing

Before Over-Sampling



After Over-Sampling Using SMOTE



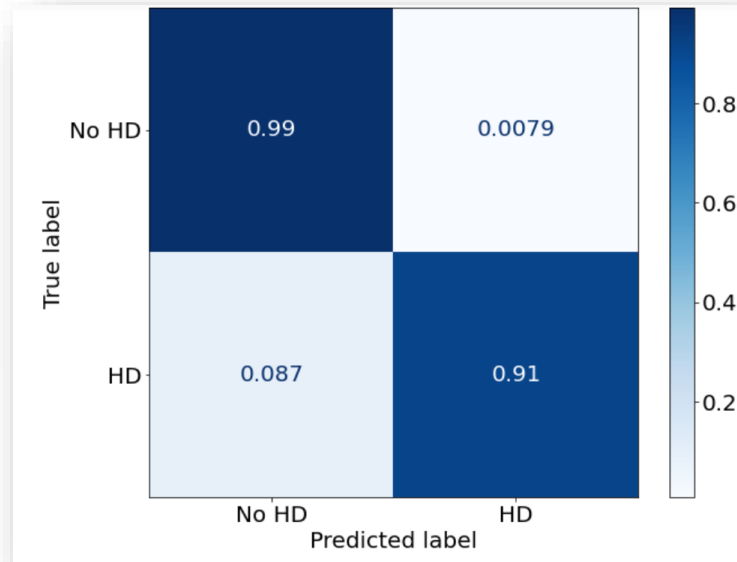
Modelling

All models comparison

	Model	Train Score	Test Score	Recall	Precision	f1-score
0	DT	0.942	0.933	88.55	97.75	92.92
1	Logistic	0.938	0.938	90.27	97.16	93.59
2	SGD	0.946	0.946	89.29	100.00	94.34
3	RF	0.948	0.942	91.07	97.16	94.02
4	Ada	0.931	0.933	90.16	96.13	93.05
5	XGB	0.954	0.953	91.28	99.17	95.06

XGBoost was the best model with ability to predict people with HD with accuracy of 91 %and people without HD with accuracy of 99 %

XGBoost Confusion matrix



Classification Report

	precision	recall	f1-score	support
0	0.91959	0.99212	0.95448	73105
1	0.99145	0.91325	0.95074	73106
accuracy			0.95268	146211
macro avg	0.95552	0.95269	0.95261	146211
weighted avg	0.95552	0.95268	0.95261	146211

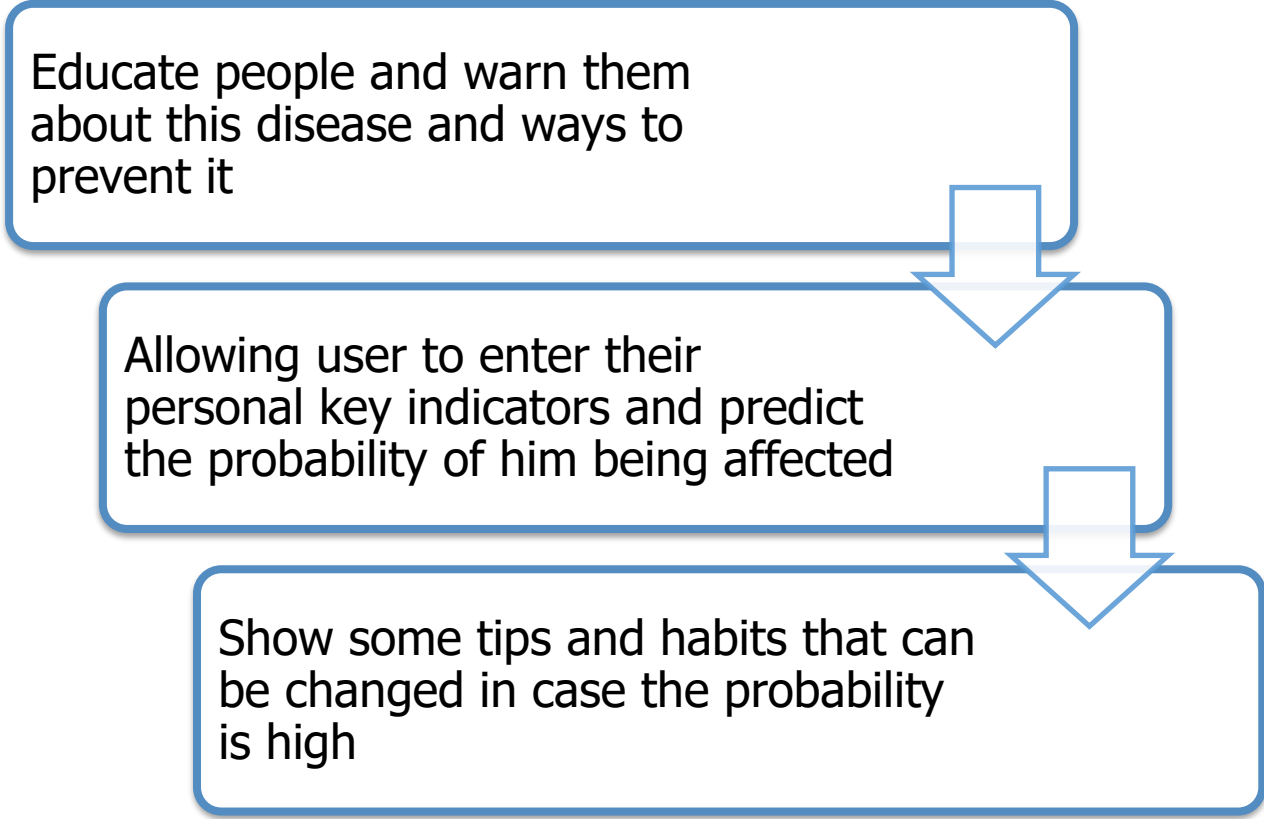
7

Business Solution

Business Solution

Based on our previous Business solution and model, we concluded that this project can be implemented as an application providing the following features

Educate people and warn them about this disease and ways to prevent it



```
graph TD; A[Educate people and warn them about this disease and ways to prevent it] --> B[Allowing user to enter their personal key indicators and predict the probability of him being affected]; B --> C[Show some tips and habits that can be changed in case the probability is high];
```

Allowing user to enter their personal key indicators and predict the probability of him being affected

Show some tips and habits that can be changed in case the probability is high

7

Conclusion

Conclusion

People with Heart disease has been increase a lot recently so awareness have to increase more

Giving up bad habits can help a lot in preventing Heart Disease

Having one or more unmodifiable factor increases the rate, so managing modifiable factors have to be done

SAMSUNG

THANKS!

Any question?

Together for Tomorrow!
Enabling People

Education for Future Generations



SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations

©2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.