SAMSUNG

Heart Diseases

Supervised by
Dr.Doaa Mohamed

Facilitator
Eng.Haneen Hossam

Together for Tomorrow!
Enabling People
Education for Future Generations

Team Members

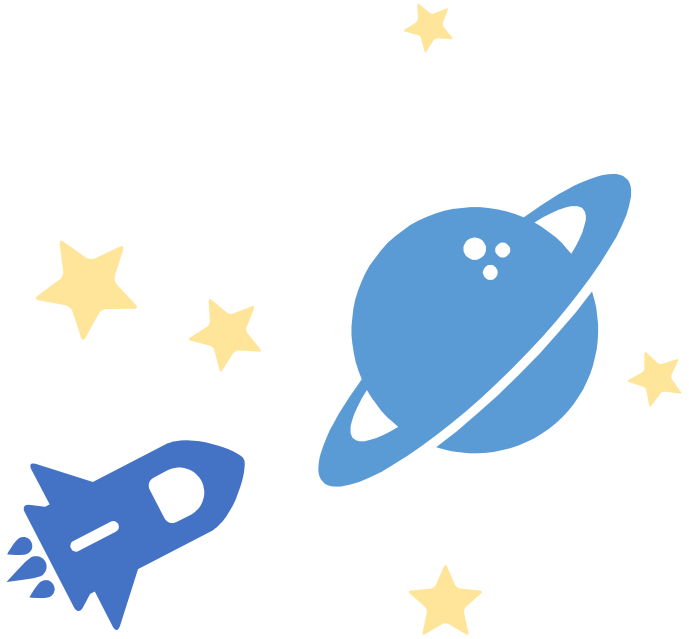
- **Ahmed Mohsen**
- **Yomna Ramdan**
- **Hossam Galal**

Agenda:

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

SAMSUNG

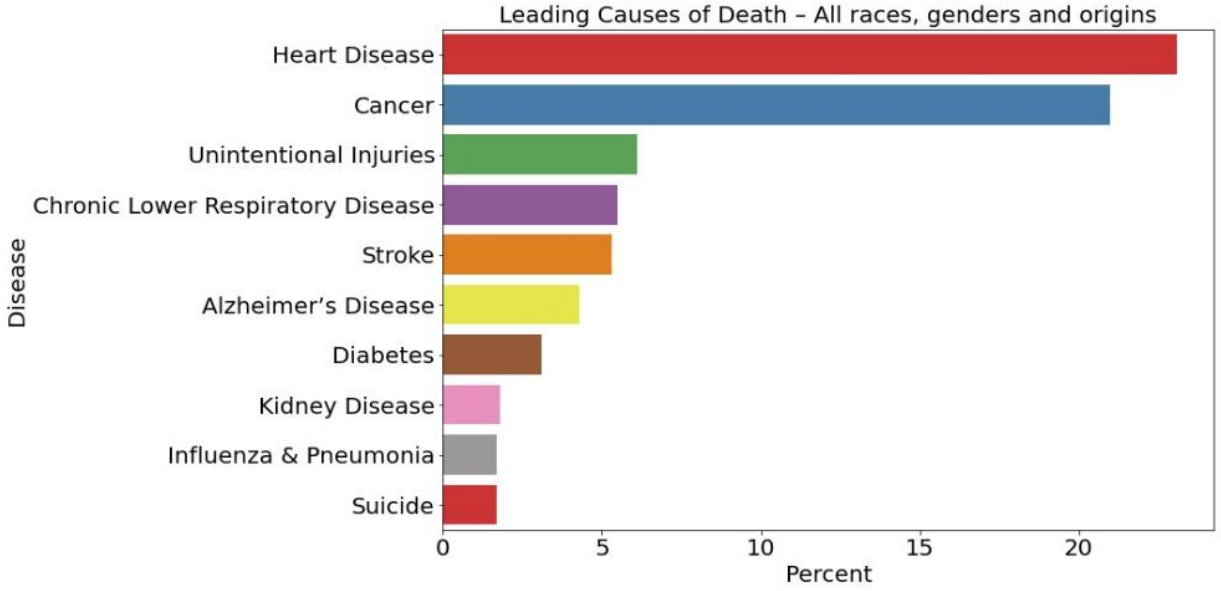
1 Motivation



Together for Tomorrow!
Enabling People
Education for Future Generations

Motivation

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



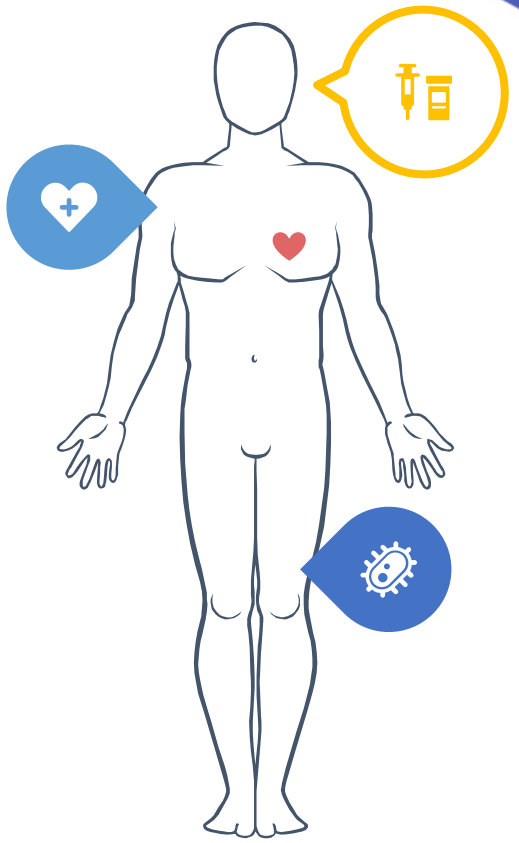
SAMSUNG

2 About the problem

Together for Tomorrow!
Enabling People
Education for Future Generations

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

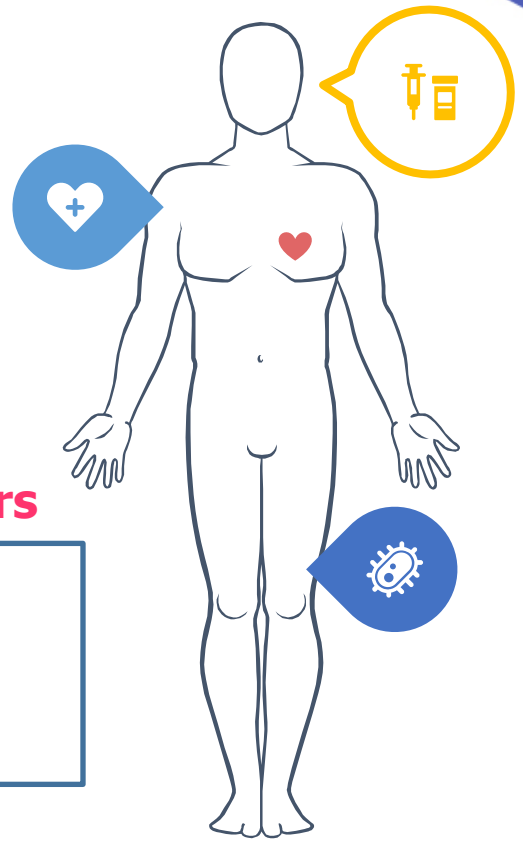
Ways to check risk of Heart Disease

Risk factors

blood pressure, blood cholesterol and other biometrics

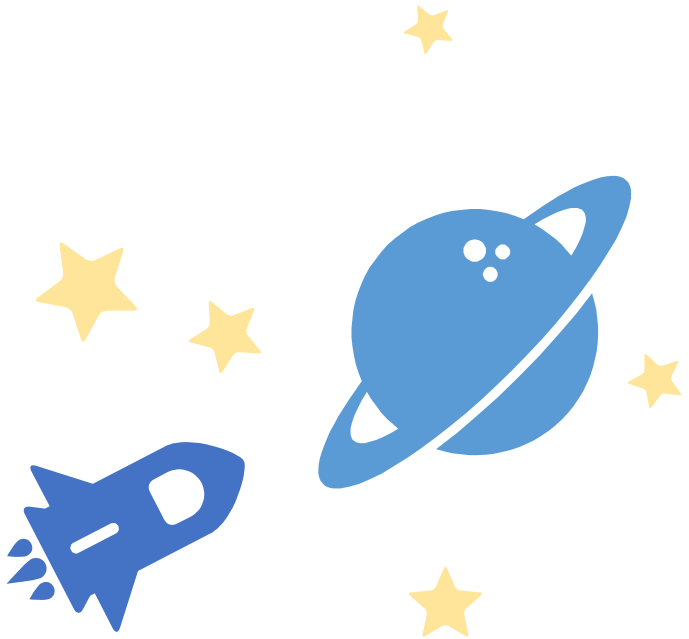
Personal key indicators

BMI, gender, age, race and mental and physical health



SAMSUNG

3 Objectives



Together for Tomorrow!
Enabling People
Education for Future Generations

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

SAMSUNG

4 Understanding Data

Together for Tomorrow!
Enabling People
Education for Future Generations

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

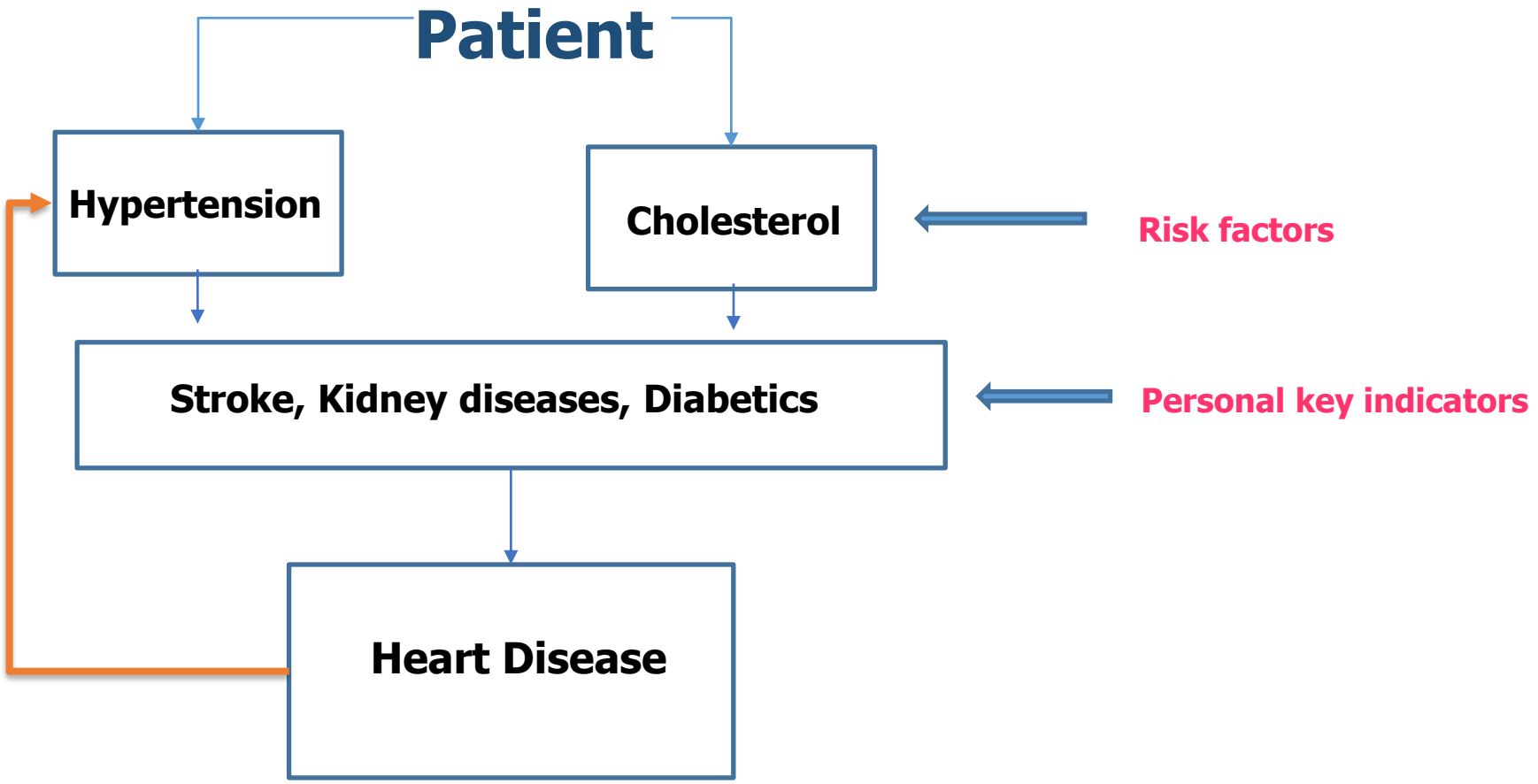
SAMSUNG

5

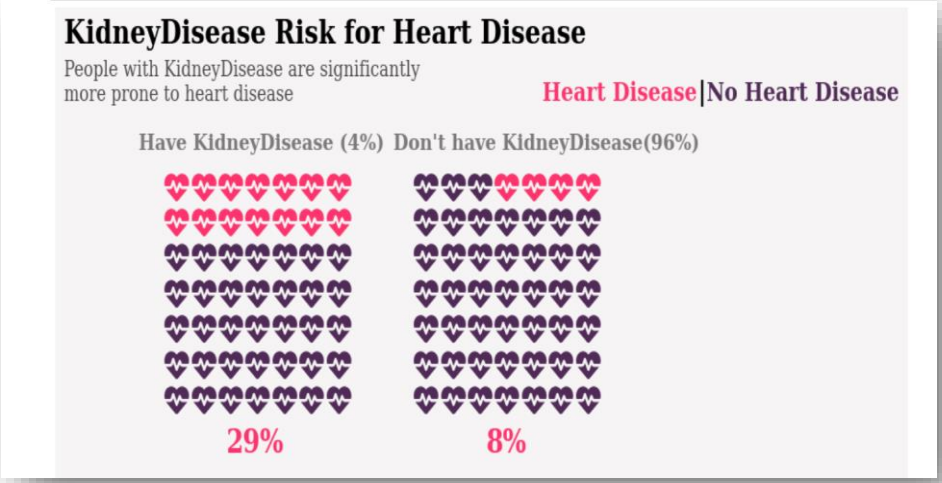
Business Question & Visualization

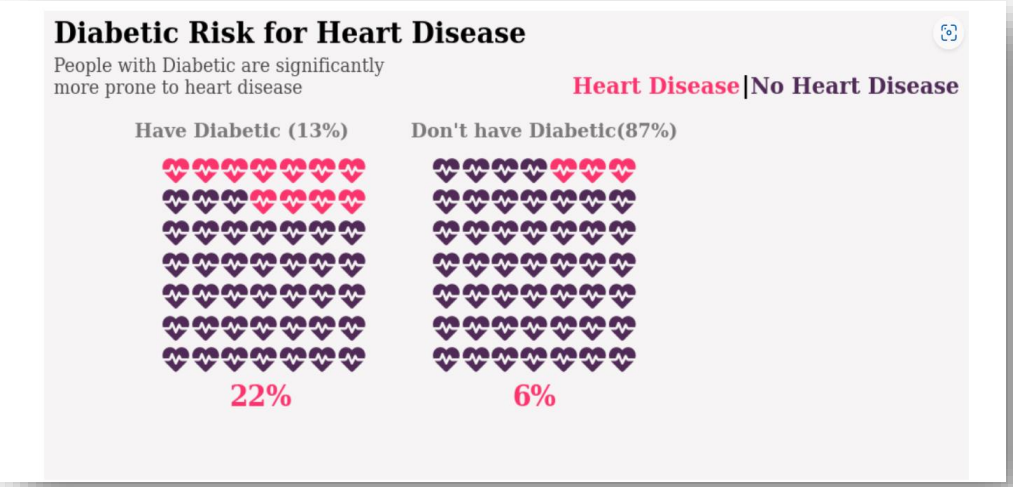
Together for Tomorrow!
Enabling People
Education for Future Generations

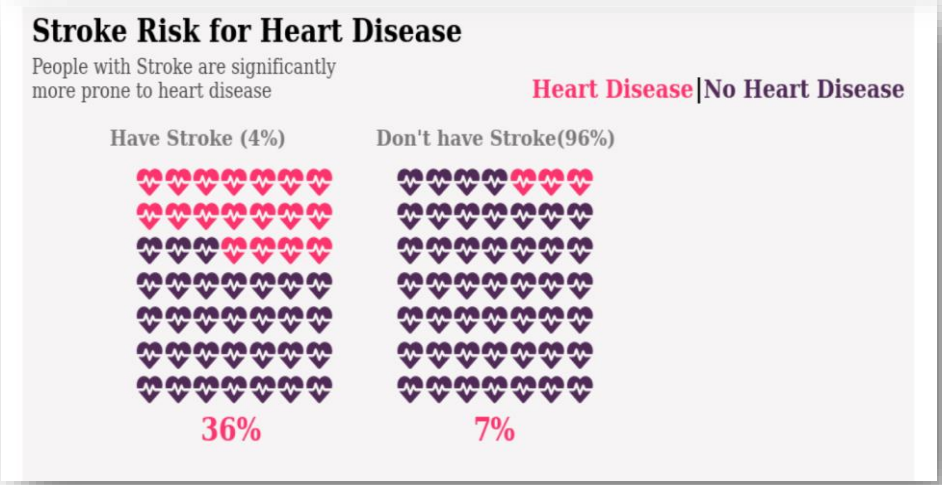
- **Are there any Personal key factors which indicate the risk as efficient as risk factors**



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



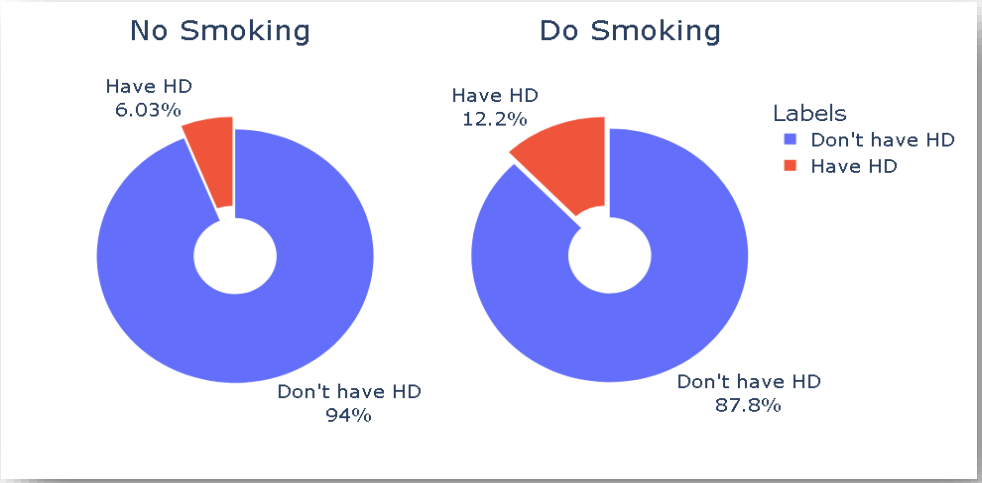


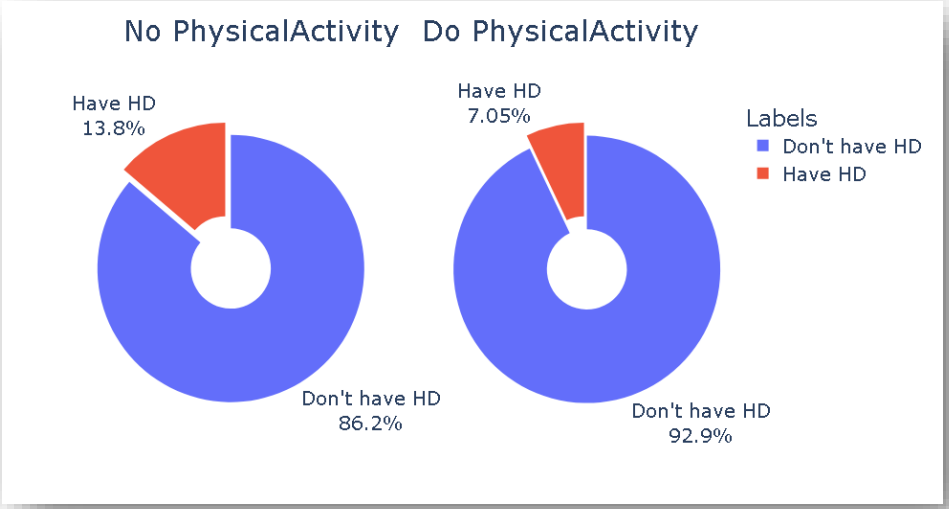


Insight:
People with common causes diseases are affected with Heart Disease up to 3 to 4 times

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 ?**

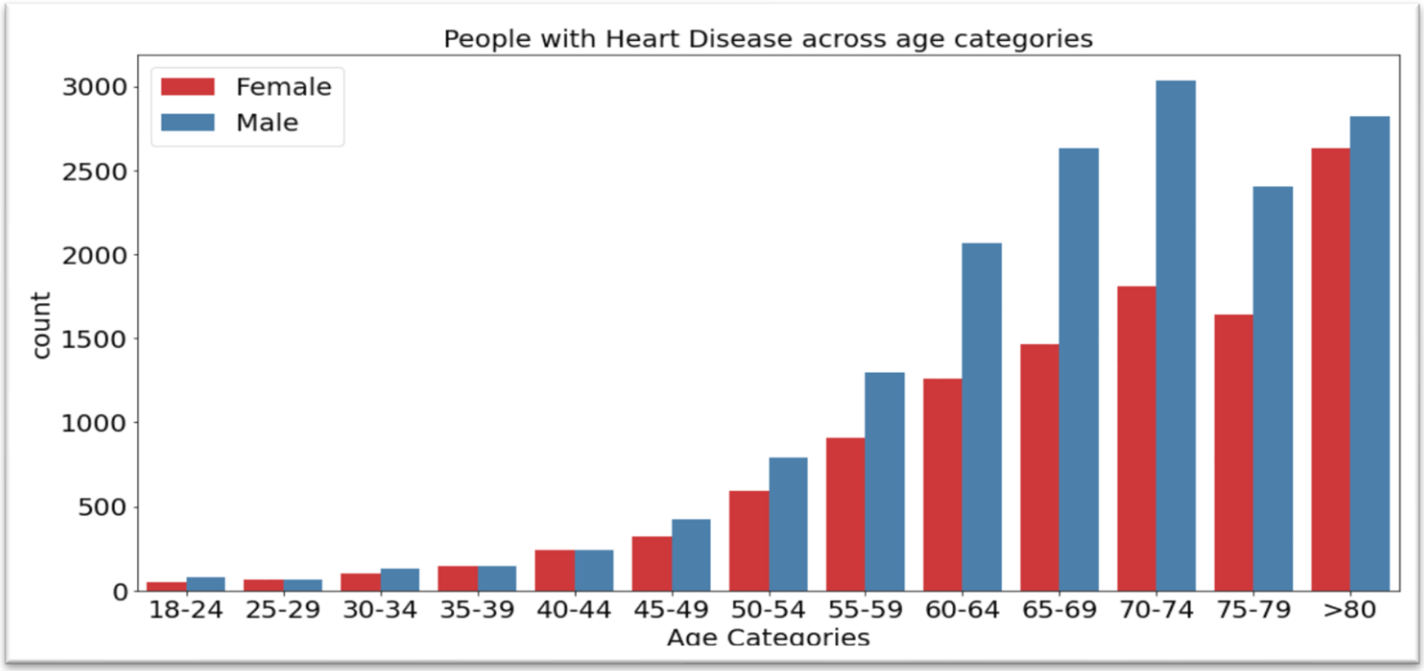




Insight:
Smokers and inactive people are twice as likely to get Heart Disease, Alcohol Drinking turned out to be not a factor

Solution:
Educating people about the importance of physical activates and quitting smoking

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



Insight:

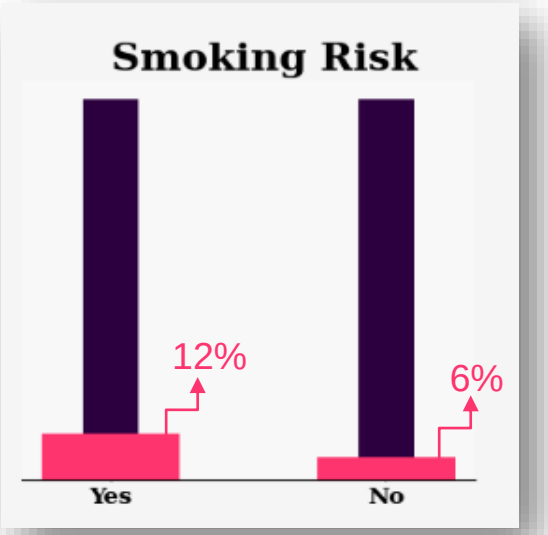
- As people get older, their risk of heart disease is increased significantly
- Males are clearly more prone to heart disease

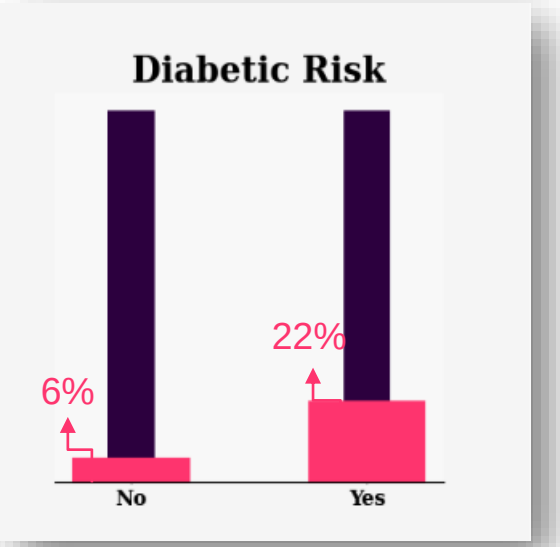
Solution:

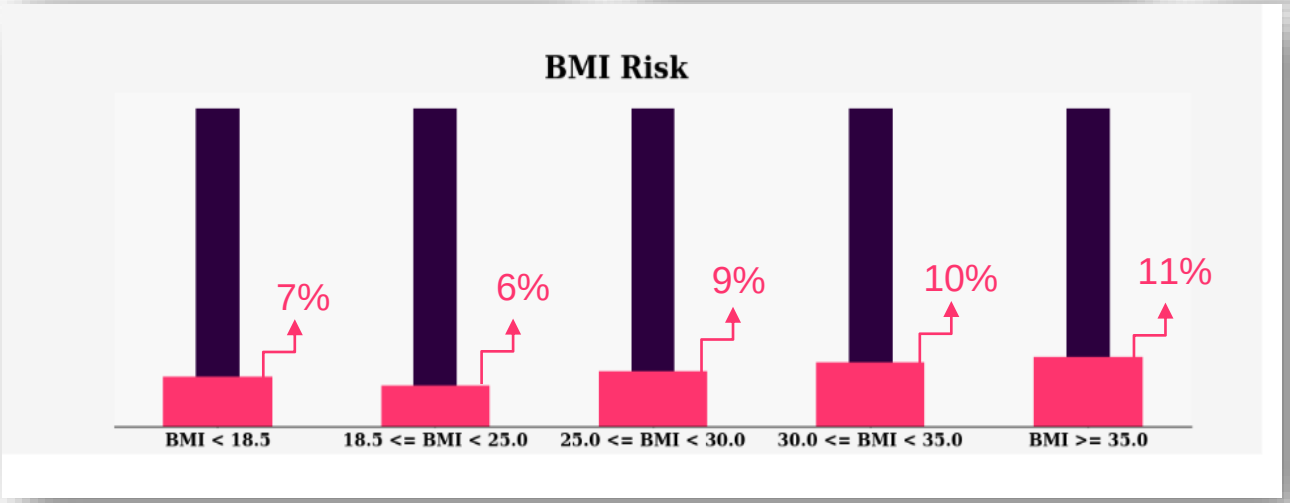
Aged people should have a regular check-up to check their health especially males

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

- **Heart Disease percentage**



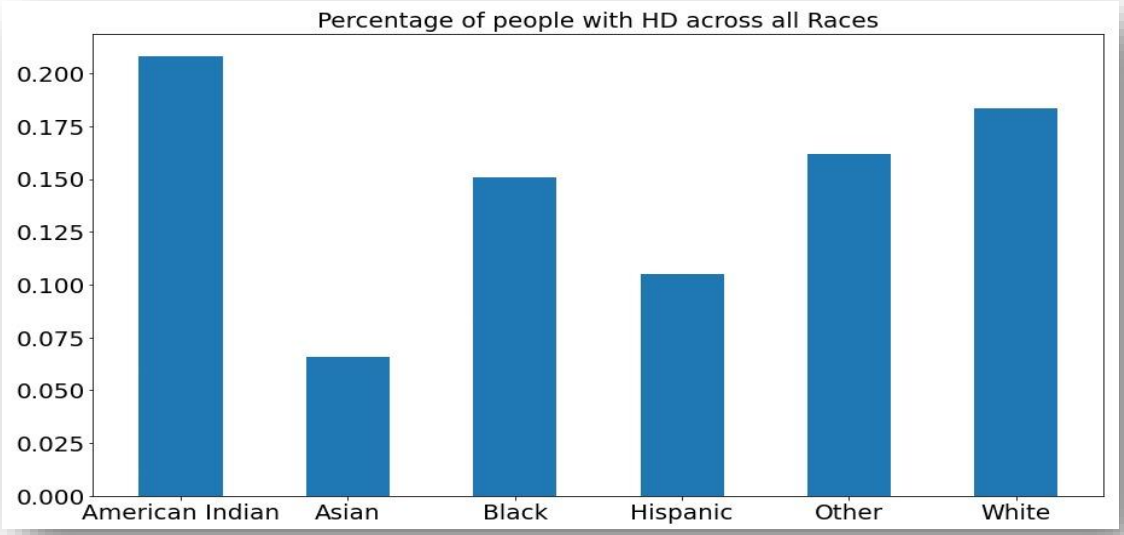


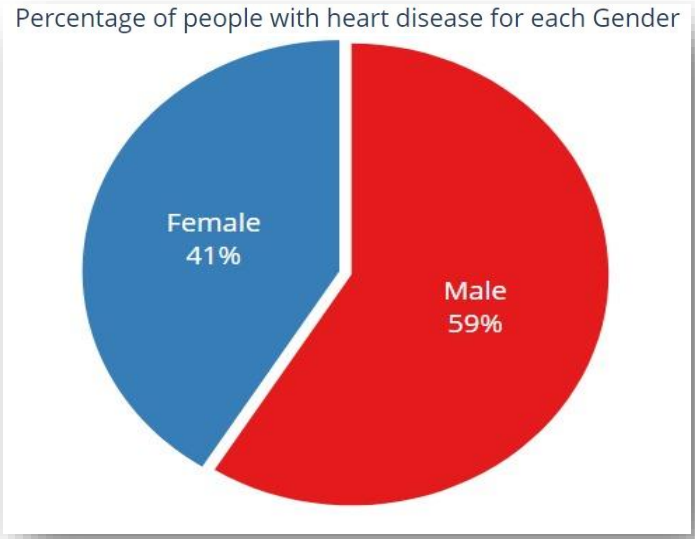


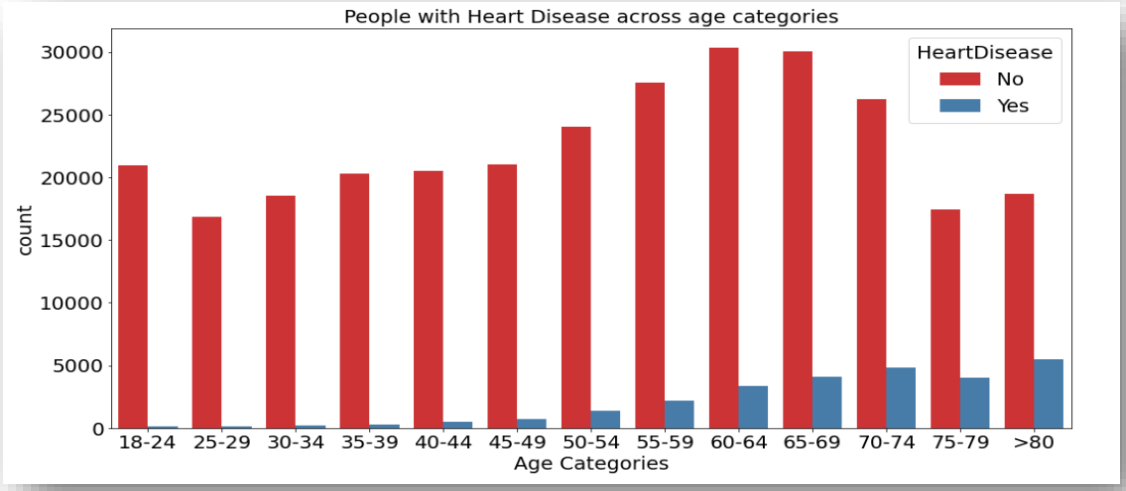
Insight:
NOT managing Diabetic, Physical Activity, BMI and smoking and can increase the risk of getting Heart Disease up to 2-3 times

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?



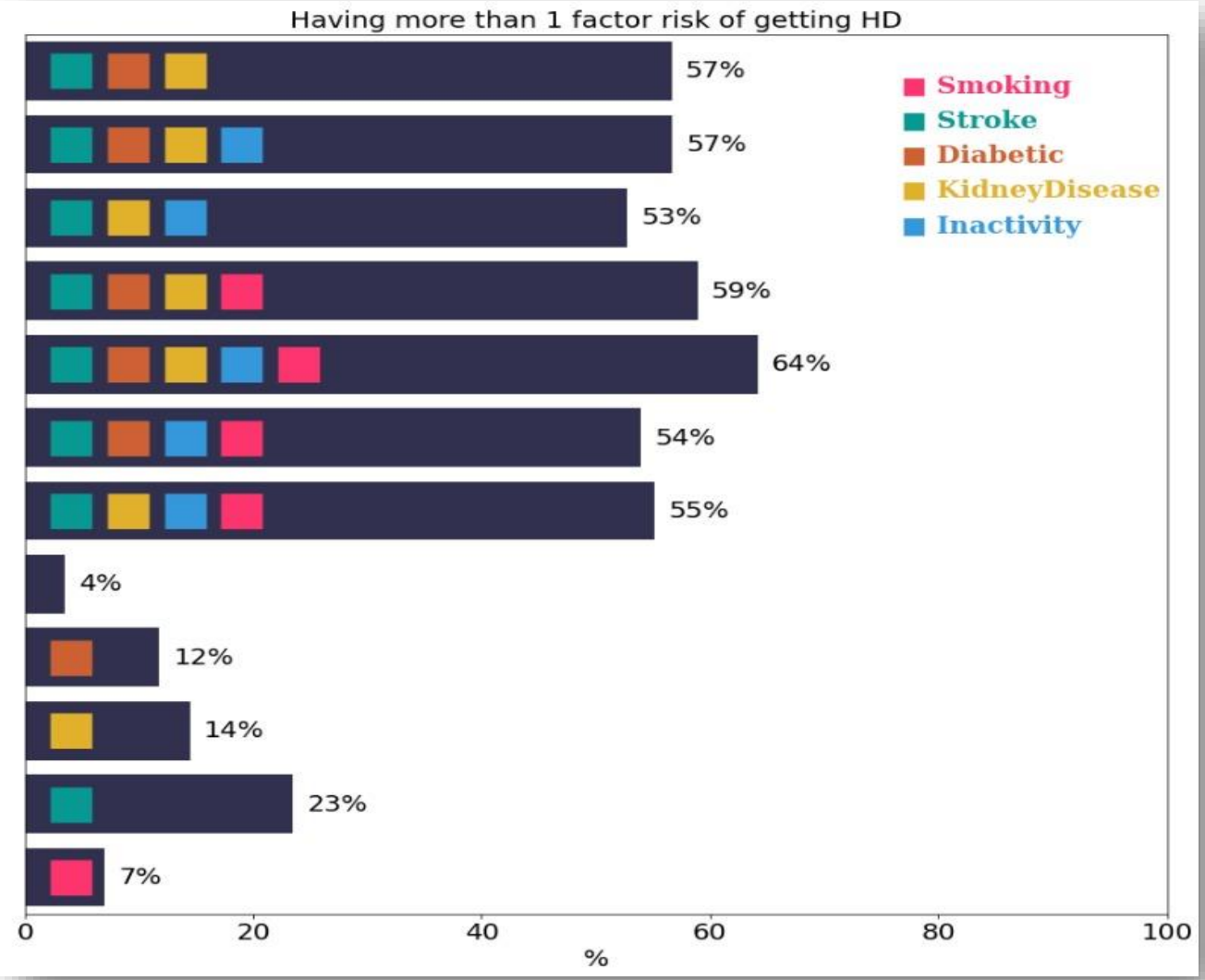




Insight:
it's shown that there is some non-modifiable factors which can increase the risk of getting Heart Disease like: Gender, Age, Race

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?



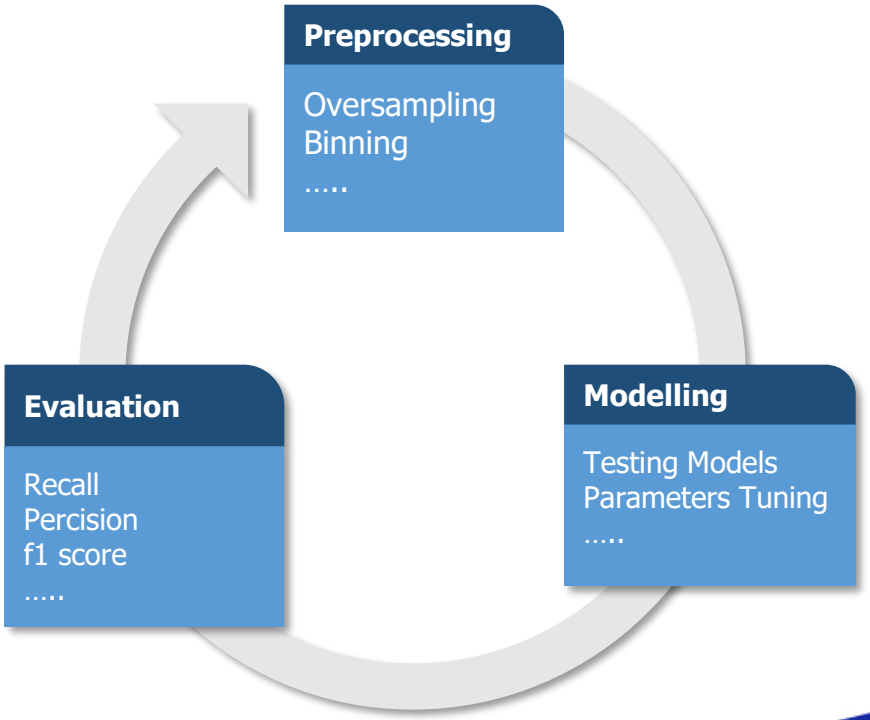
Insight:
Having more than 1 factor can increase the chance of getting Heart Disease significantly

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

SAMSUNG

6

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

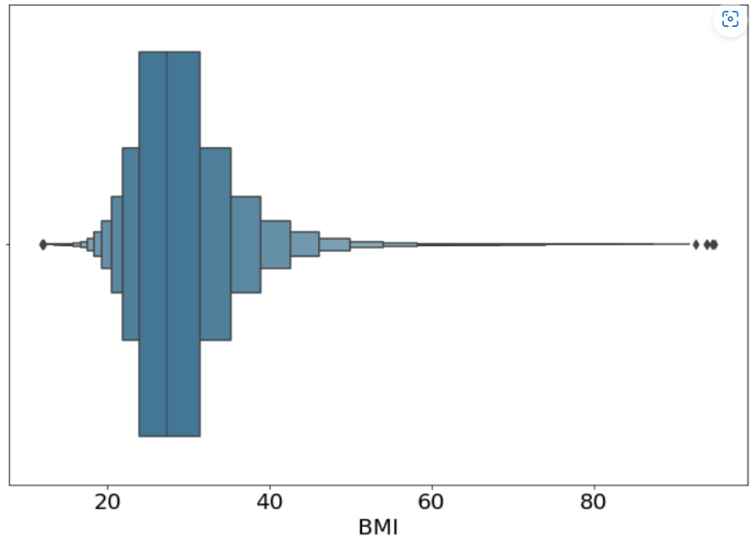
Duplicated data

```
Number of duplicated recrods = 18078
```

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

Outliers 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 <= BMI < 30.0)	114355
Normal weight (18.5 <= BMI < 25.0)	97778
Obese (30.0 <= BMI < 35.0)	61169
Extremly Obese (BMI >= 35.0)	41379
Underweight (BMI < 18.5)	5114

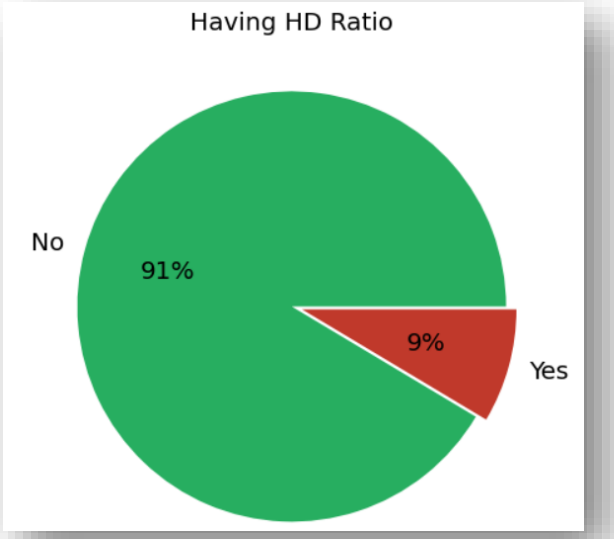
In real life, BMI is divided into finite categories, so for better insights and to increase efficiency, BMI is categorized using binning

Encoding

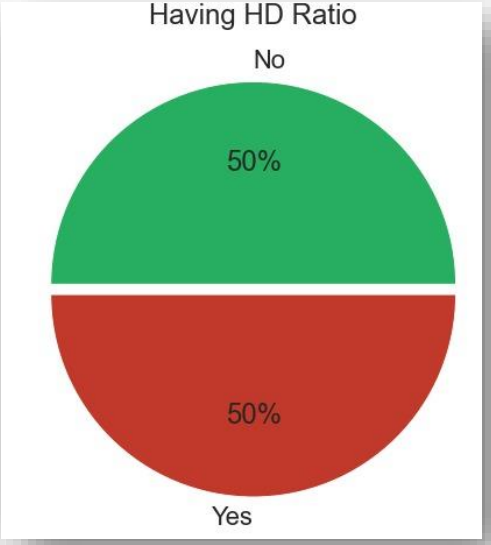
- Encoding features with no ordering by one-hot encoding
- Encoding features with ordering by label encoding

Imbalancing

Before Over-Sampling



After Over-Sampling Using SMOTE

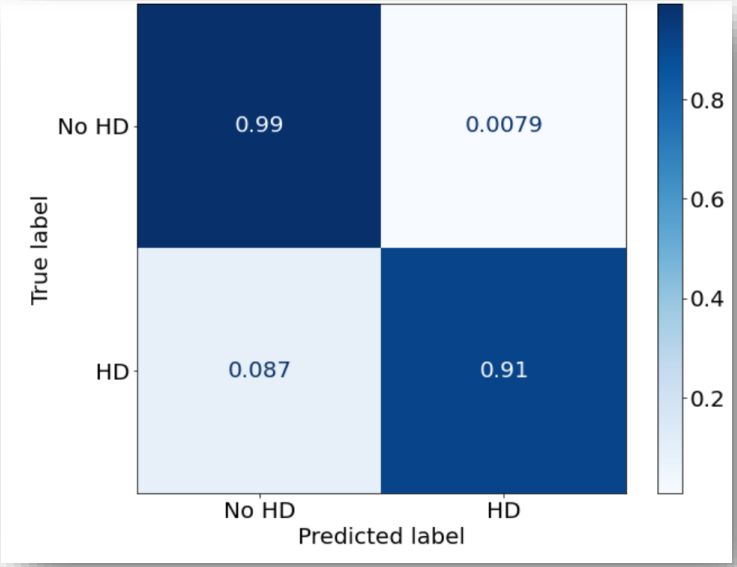


Modelling

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 Confusion matrix



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

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

SAMSUNG

7 Business Solution

Together for Tomorrow!
Enabling People
Education for Future Generations

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

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

SAMSUNG

7 Conclusion

Together for Tomorrow!
Enabling People
Education for Future Generations

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