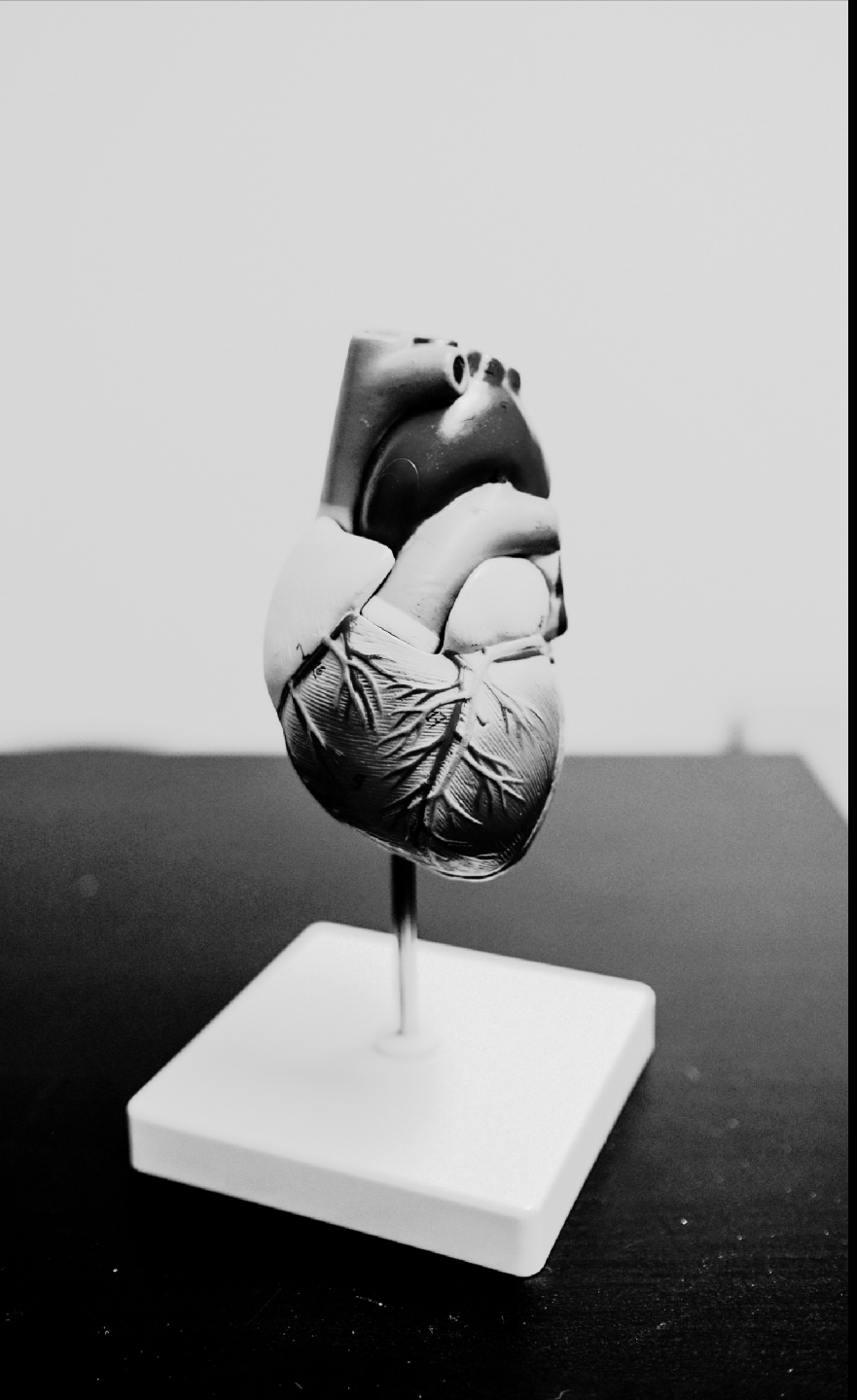
HEART DISEASE ANALYSIS

Fangle Xi

2022/4/26



MENU

1. Background Introduction
2. Data Pre-processing
3. Data Visualization & Statistical Analysis
4. Predictive Analysis & Machine Learning

47%

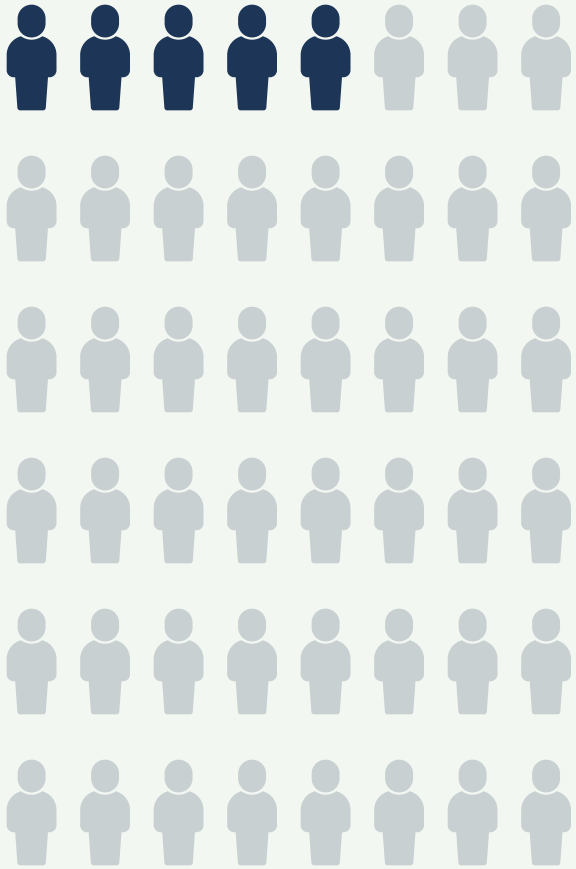
OF U.S. HAVE AT LEAST ONE OF THREE RISK HEART DISEASE FACTORS.


3.32 Pages Visit
Traffic Sources Overview
Direct Traffic
3,097.00 (40.4
Search Engine
2,910.00 (38.04%)
Referring Sites
1,642.00 (21.47%)
Visitors Overview
Apr 19
Apr 26
May 3
Visitors
2,958


ALCOHOLDRINKING
MENTALHEALTH
KIDNEYDISEASE
GENHEALTH
RACE
DIFFWALKING
SEX
GENHEALTH
BMI
DIABETIC
SLEEPTIME
SEX
ASTHMA
SMOKING
PHYSICALHEALTH
STROKE
PHYSICALHEALTH
SKINCANCER
HEARTDISEASE
ASTHMA
STROKE
RACE
MENTALHEALTH
BMI
PHYSICALACTIVITY
DIABETIC
ALCOHOLDRINKING
SMOKING
AGECATEGORY
SKINCANCER
PHYSICALACTIVITY
SLEEPTIME
DIFFWALKING
AGECATEGORY
KIDNEYDISEASE

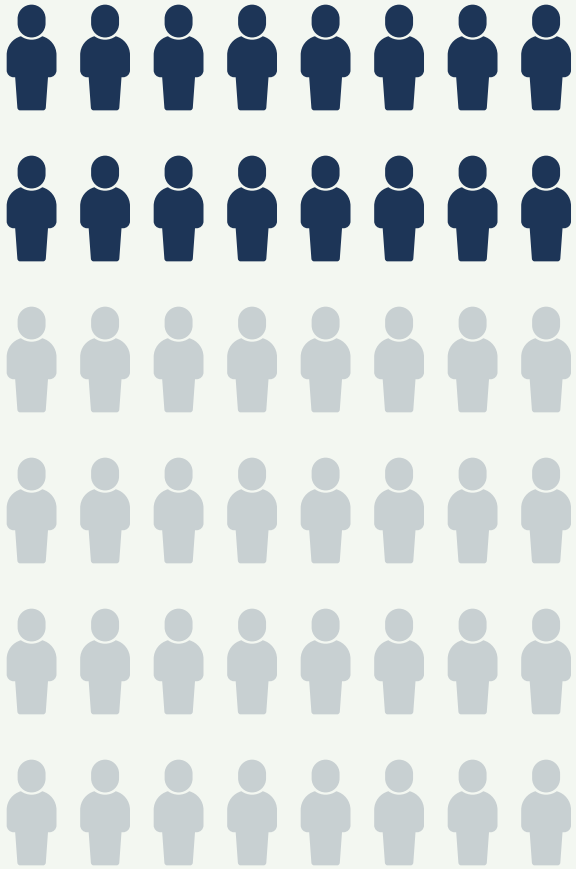
SO, USE UNDERSAMPLING METHODS TO RESAMPLE OUR DATASET!

9%

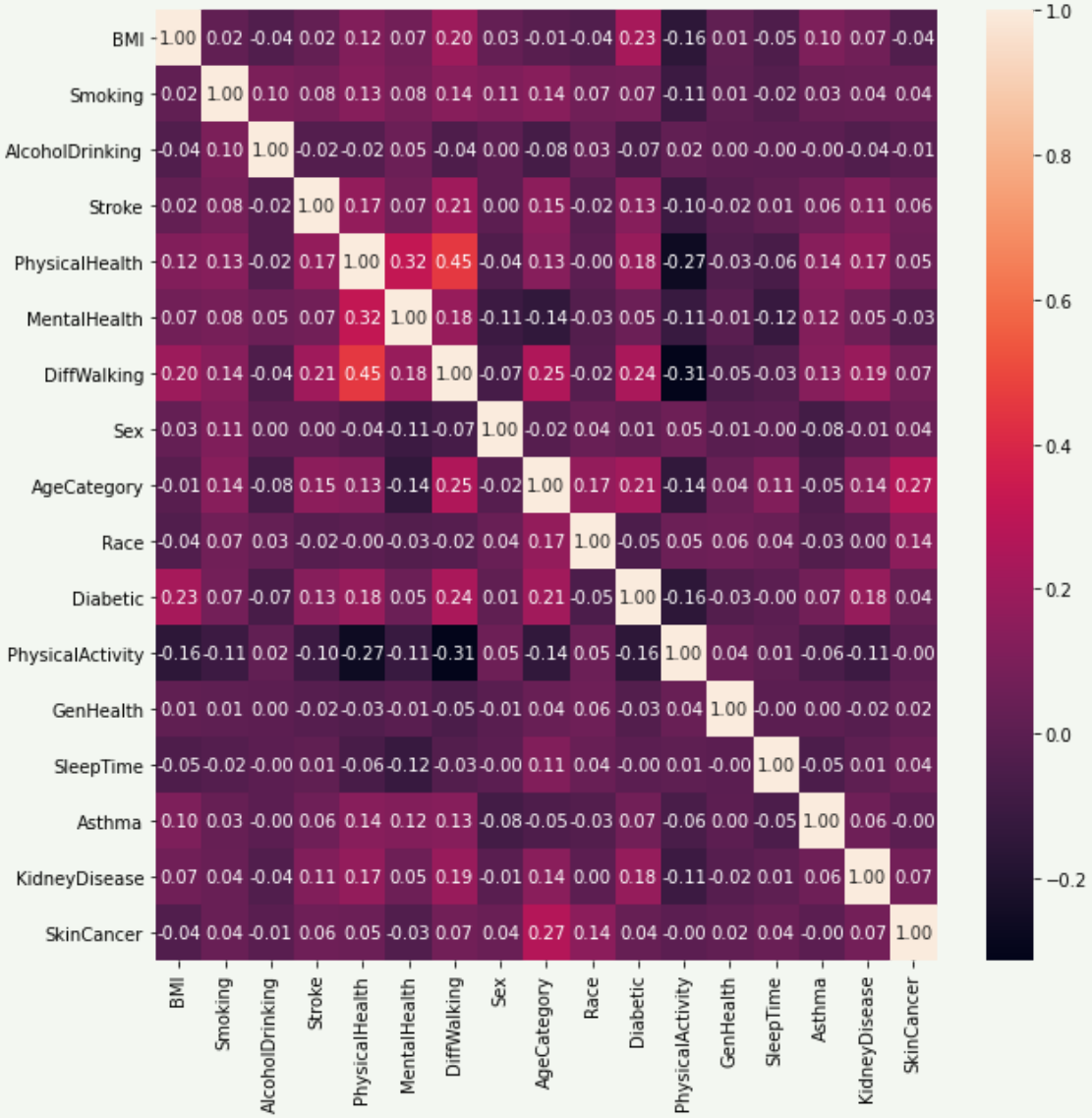


In the original data set, there is only about 9% objects have heart disease, so it may exits unbalance problem

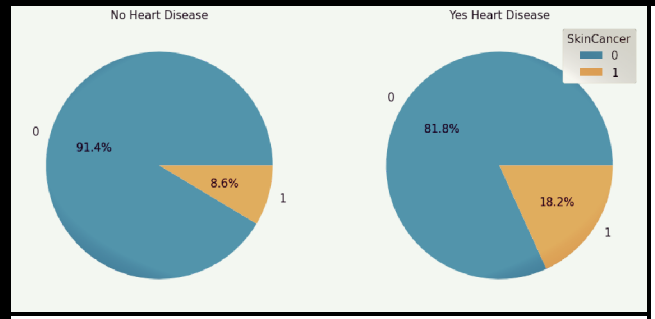
33%

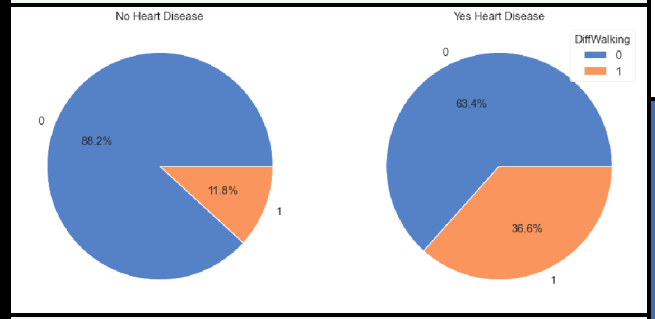


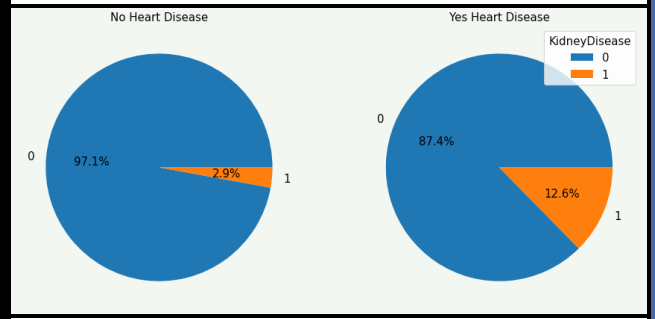
After resampling the data-set, there are about 33% objects having heart disease.

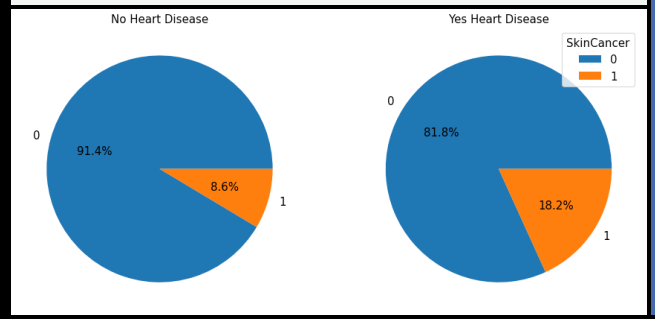


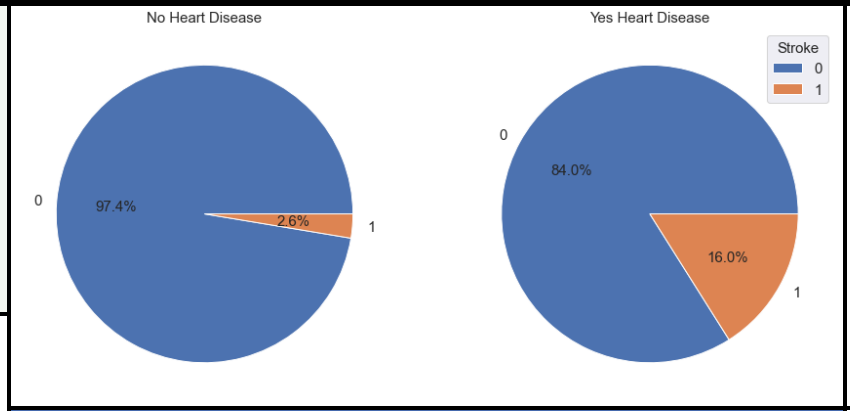
THE CORRELATIONS BETWEEN VARIABLES ARE SMALL. -> INDEPENDENT

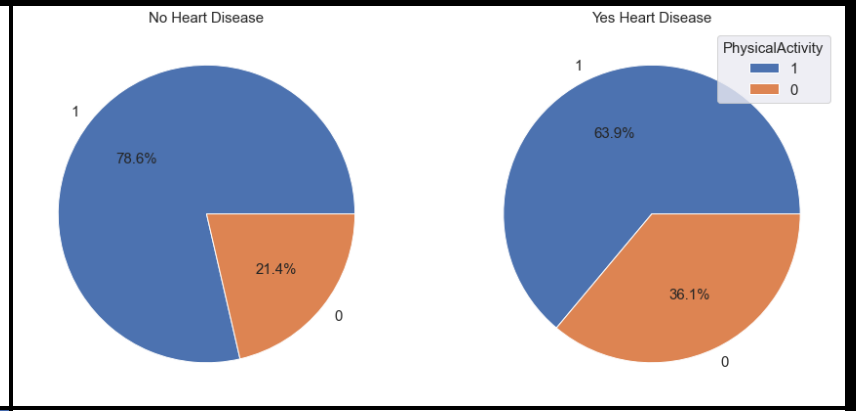












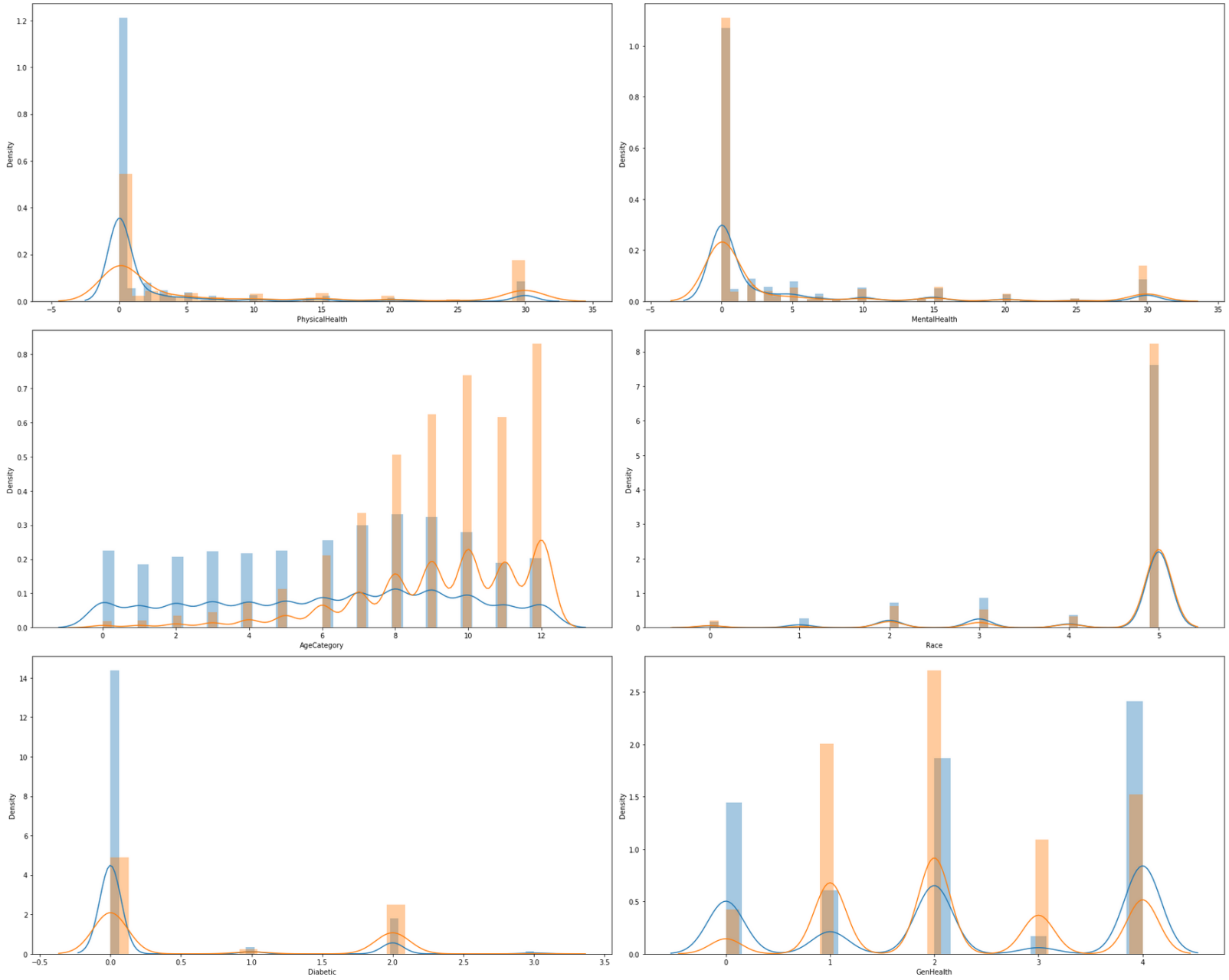
BINARY FEATURES & HEART DISEASE

From the above pie-charts, we can learn that the heart disease patients are more likely to have skin cancer, kidney disease, asthma, walking difficulty and stroke than people with health heart. Therefore, there may be some relationships between the other diseases and heart diseases.

HEART DISEASE &

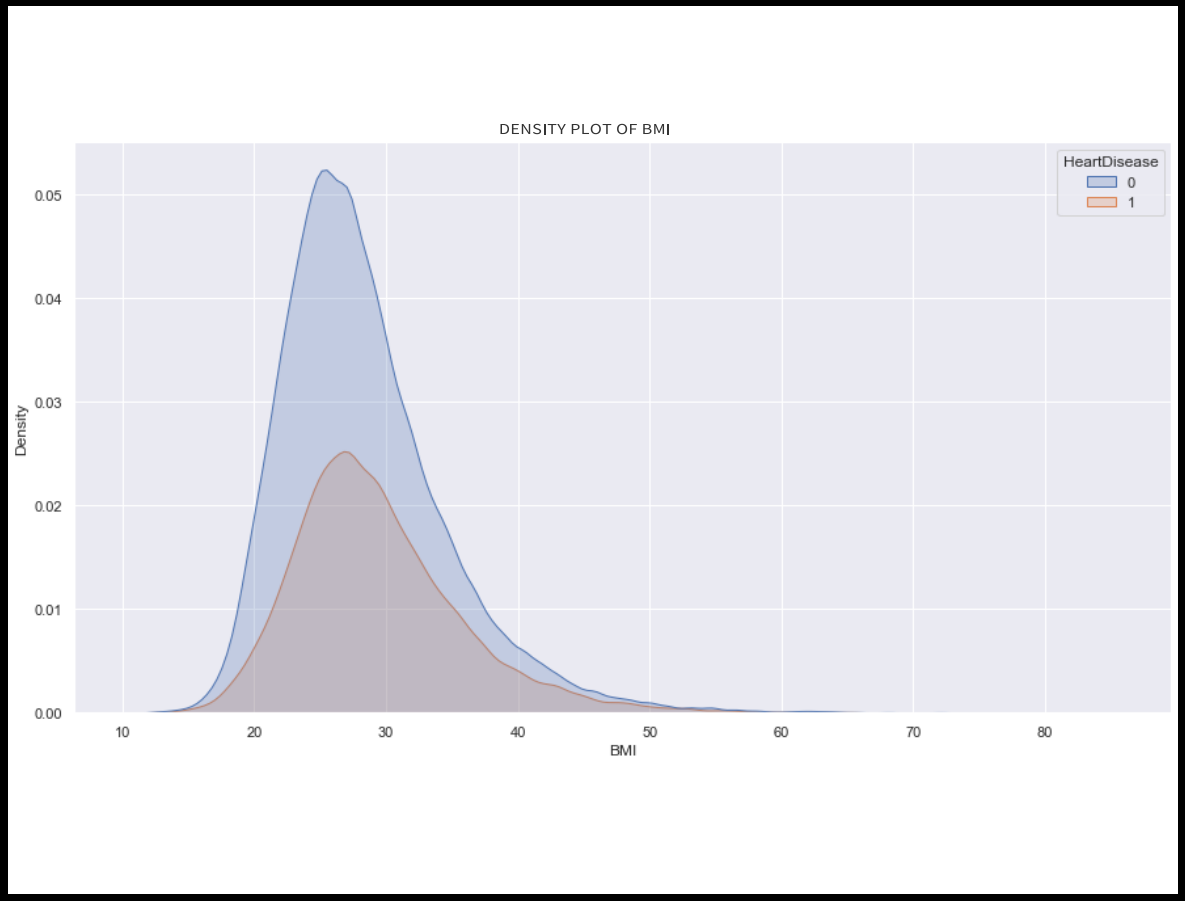
DISCRETE FEATURES

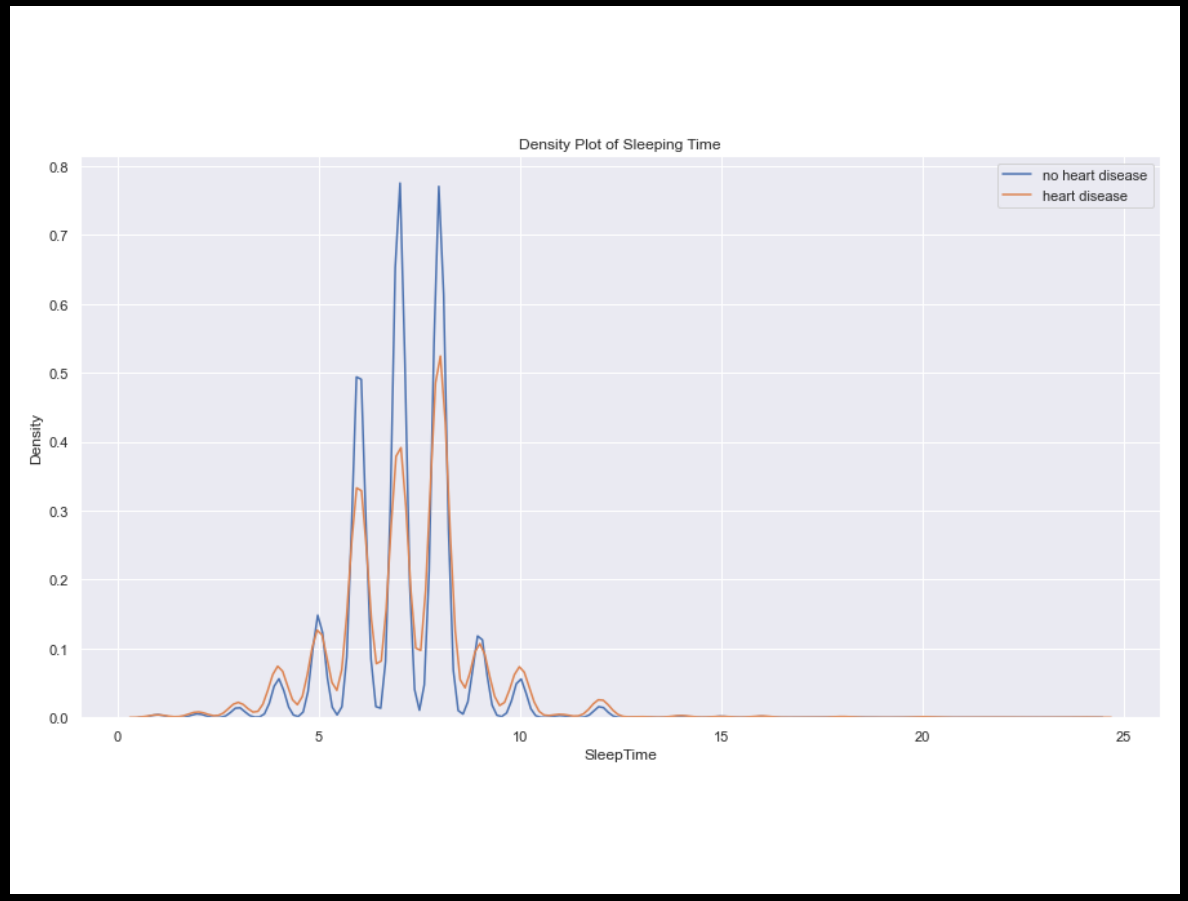
By comparing the distribution of each discrete features, we learned that the heart disease is not related to race, mental health and physical health. With the age growth, people are more likely to have heart problems. From the last plot, we can see that people with not very good genhealth(1-Fair, 2-Good, 3-poor) have high risk to have heart disease. What's more,it is more possible for people with heavier diabetic to have heart disease.



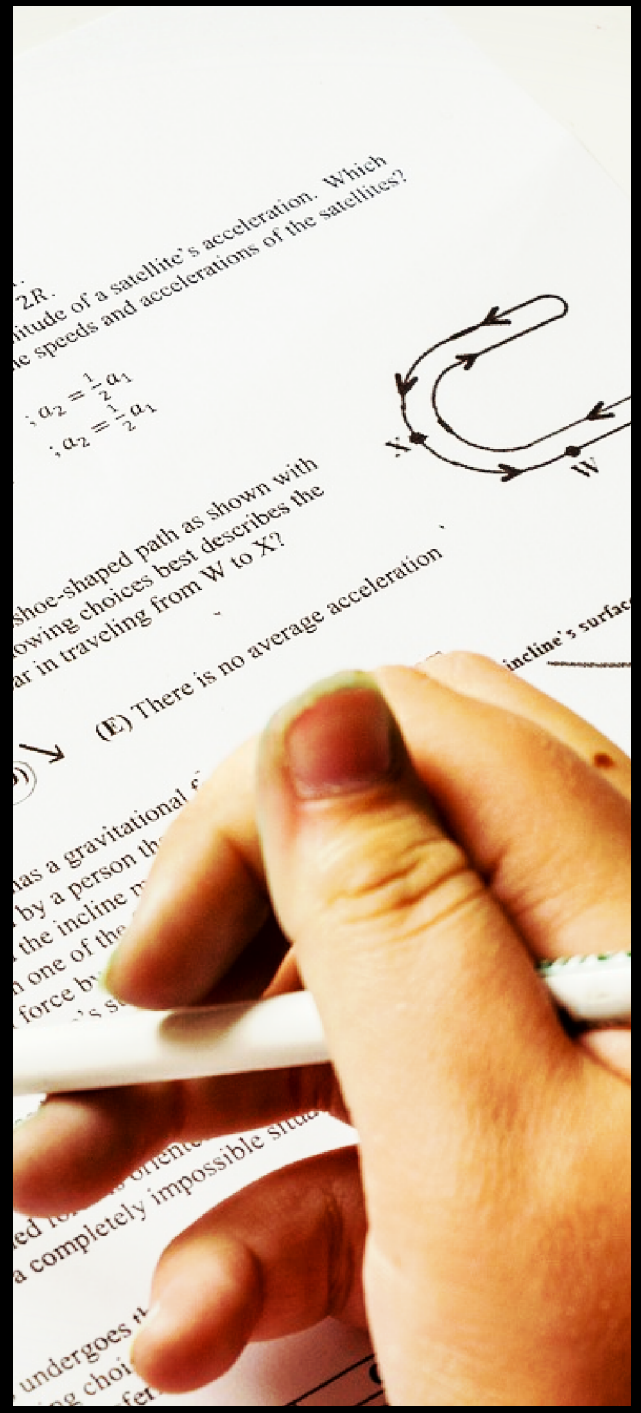
HEART DISEASE &

CONTINUOUS FEATURES

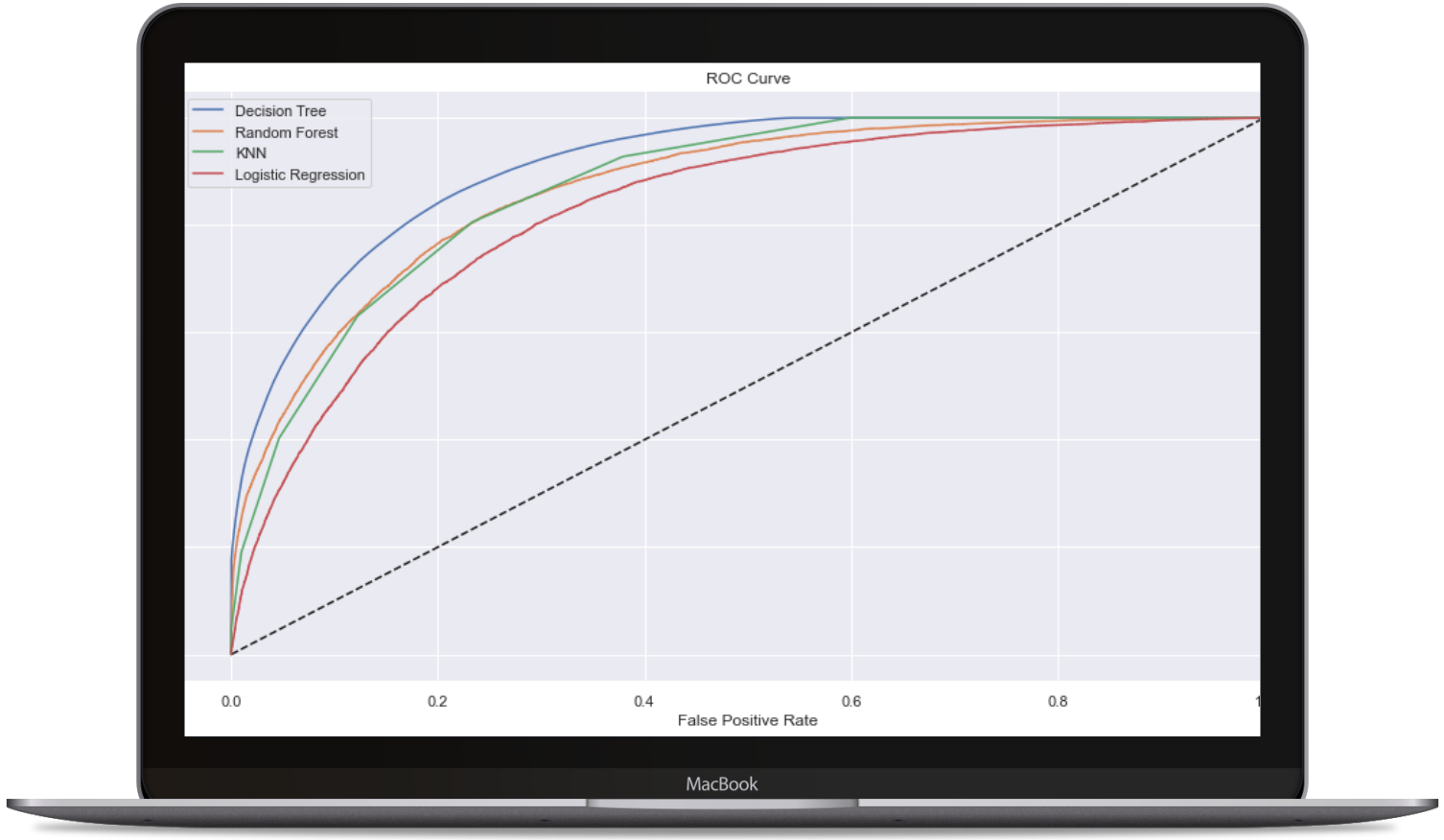




MANOVA TEST


2R.
itude of a satellite's acceleration. Which
e speeds and accelerations of the satellites?
$a_2 = \frac{1}{2}a_1$
$a_2 = \frac{1}{2}a_1$
shoe-shaped path as shown with
owing choices best describes the
ar in traveling from W to X?
(E) There is no average acceleration
incline's surface
has a gravitational
by a person th
the incline
one of the
force by
completely impossible situa
undergoes

PREDICTIVE MODELS AND MACHINE LEARNING



- In our experiment, we compare decision tree, random forest, knn and logistic regression algorithm.
- We use the Randomized Search Cross-validation to find the most suitable parameters for each machine learning algorithm.
- By comparing their ROC curves, the decision tree model('criterion': 'entropy', 'max_depth': None, 'max_features': 8, 'min_samples_leaf': 7) is better than the other three models.

CONCLUSION

For the indicators of heart disease

The heart disease is highly related to other different kinds of disease, like skin cancer, stroke, difficult-walking, asthma and kidney disease. Therefore, we advice people with related medical conditions to do heart screening.

For model selection and usage

In practice, we can use the logistic regression model to explain the relationship between the target variable and the features, and we can use the decision tree model to do the prediction.

QUESTION AND ANSWER PART | APRIL 2022

QUESTION + ADVICE

REFERENCE

- KAMIL PYTLAK (2022, Feb 16). Personal Key Indicators of Heart Disease.Retrieved April 27, 2022, from https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease

- Centers for Disease Control and Prevention.(2021, September 27). About heart disease.Centers for Disease Control and Prevention.Retrieved April 27, 2022, from https://www.cdc.gov/heartdisease/about.htm

- WE ARE FISH BUT WE CAN FLY, (2022, April 11). Heart Disease Data Understanding. Retrieved April 27, 2022, from https://www.kaggle.com/code/hiiammrcat/heart-disease-data-understanding

- SAKSHI,(2022, April 10). Heart Disease Indicators:ML Algorithms Comparison.Retrieved April 27, 2022, from https://www.kaggle.com/code/gladerunner/heart-disease-indicators-ml-algorithms-comparison/notebook

  Datacamp.Supervised Learning with scikit-learn. Retrieved April 27, 2022, from https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn

THANKS FOR YOUR PATIENCE