Appendix

Introduction …….. *Introducing the project background*

Problem …….. *Identifying and Explaining the Problem / Task*

Methods …….. *Explaining the Methods taken to complete project*

Observations …….. *Highlighting Observations from the Project*

Results and Conclusions …….. *Highlighting the Findings and Conclusions*

References …….. *Showing the project references*

Project 1: ***Analyzing Factors that Could Help To Detect and Prevent Heart Disease***

Osmond C. Oke

Introduction

Heart disease, according to the CDC, is one of the leading causes of of death in the United States and has been the number one cause of death for the past 5 consecutive years. The term heart disease encapsulates many other elements, ranging from arrythmias to heart infections, and even congenital arterial defects, which is why it is able to claim so many lives.

Problem

Because a lot of heart diseases are irreversible, as although they cannot be remedied, they can still be managed, the best approach would be prevention in any possible shape or form. The simplest approach would be through simple lifestyle changes; by adopting healthy lifestyle habits. It would be advisable that everyone follow these practices in order to live with a better heart so to say. Now we know heart disease is indiscriminatory of age, sex, race etc., but who is more at risk of heart disease? From the given dataset, I analyzed the different attributes of individuals and tried to see what noticeable trends I could find within groups of people with heart disease and compared them against those without heart disease. The analysis of the dataset further prompted me to change my research question to ***Analyzing Factors that Could Help To Detect and Prevent Heart Disease,*** seeing as this was more a classification project than my initial idea of being able to predict heart disease in individuals.

Methods

I obtained a dataset from Kaggle which contained 303 entries of the health information of individuals, as well as whether there was a presence of heart disease or not. This information included:

The attributes included:

- **age** – The ages of the individuals
- **sex** – The sex of the individuals, with 1 representing males and 0 representing females
- **chest pain type (4 values)** – The different chest pains experienced with the values as follows (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
- **resting blood pressure** – The resting blood pressure of the individuals
- **serum cholestoral in mg/dl** – The measurement of the individuals' cholesterol
- **fasting blood sugar > 120 mg/dl** – The indicator of if individuals' fasting blood sugar was greater than 120mg/dl, with 1 representing true and 0 representing false

- **resting electrocardiographic results (values 0,1,2)** – The indicator of the individuals' resting ecg results, with 0 = normal, 1 = having ST-T wave abnormality, and 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria
- **maximum heart rate achieved** – The measurement of the maximum heart rate achieved
- **exercise induced angina** – The indicator of the exercised induced angina, with 1 = true and 0 = false
- **oldpeak** = ST depression induced by exercise relative to rest
- the slope of the peak exercise ST segment
- number of major vessels (0-3) colored by flourosopy
- **thal:** The indicator of a blood disorder called thalassemia, with different values of 3 = normal, 6 = fixed defect, and 7 = reversable defect
- **target** – The indicator of the heart disease, with 1 = the presence of heart disease and 0 = the absence of heart disease.

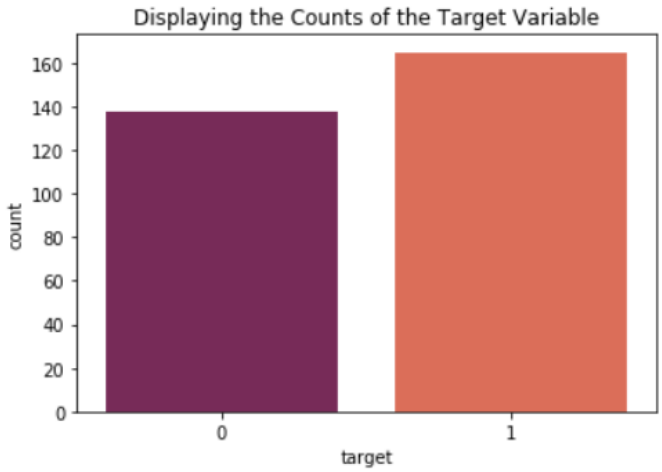
There was no need to create dummy variables, as all the categorical data was already converted to either binary, preassigned numerical values, which helped the analysis to run smoothly.

Observations

I started by examining the target variable to see how the numbers of those with heart disease compared to the number of those without heart disease. Of the total of 303 entries, it was observed that 138 of the individuals did not have heart disease, and the majority being 165 individuals showed a presence of a form of heart disease. This can be seen below in Figure 1:

Figure 1: Displaying the Counts of the Target Variable

```
The counts of the Target variable: 1    165
0    138
Name: target, dtype: int64
```



I then went on to inspect the heart disease distribution across other variables like sex, age (Figure 2) and the distribution of heart disease seen in the correlation between age and the maximum heart rates of individuals (Figure 3).

Figure 2: Heart Disease Distribution by Age

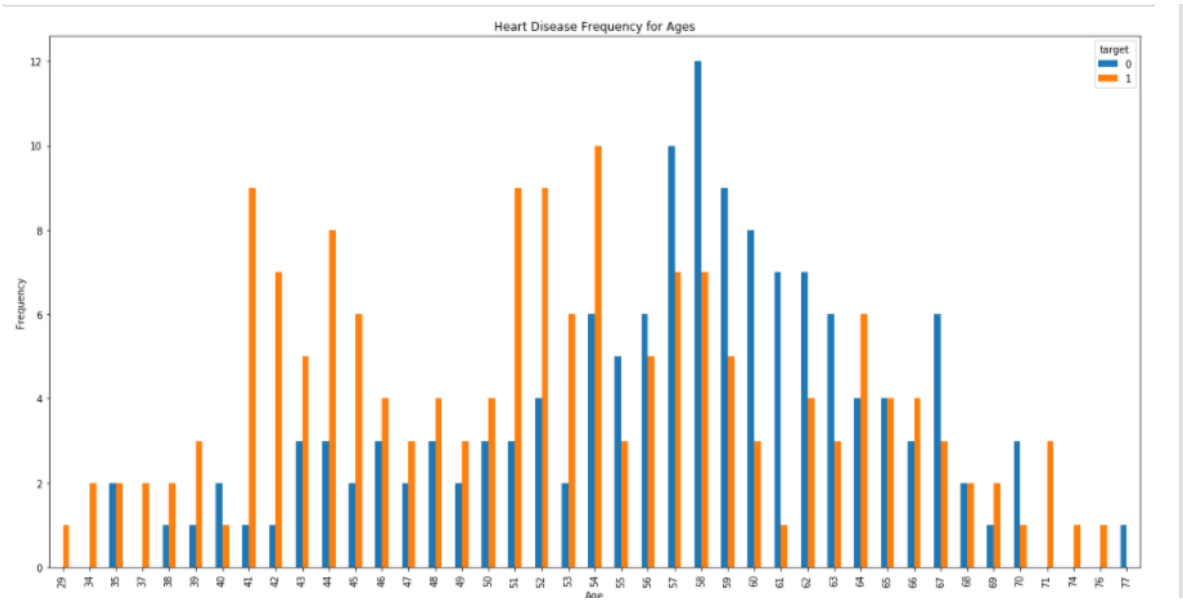
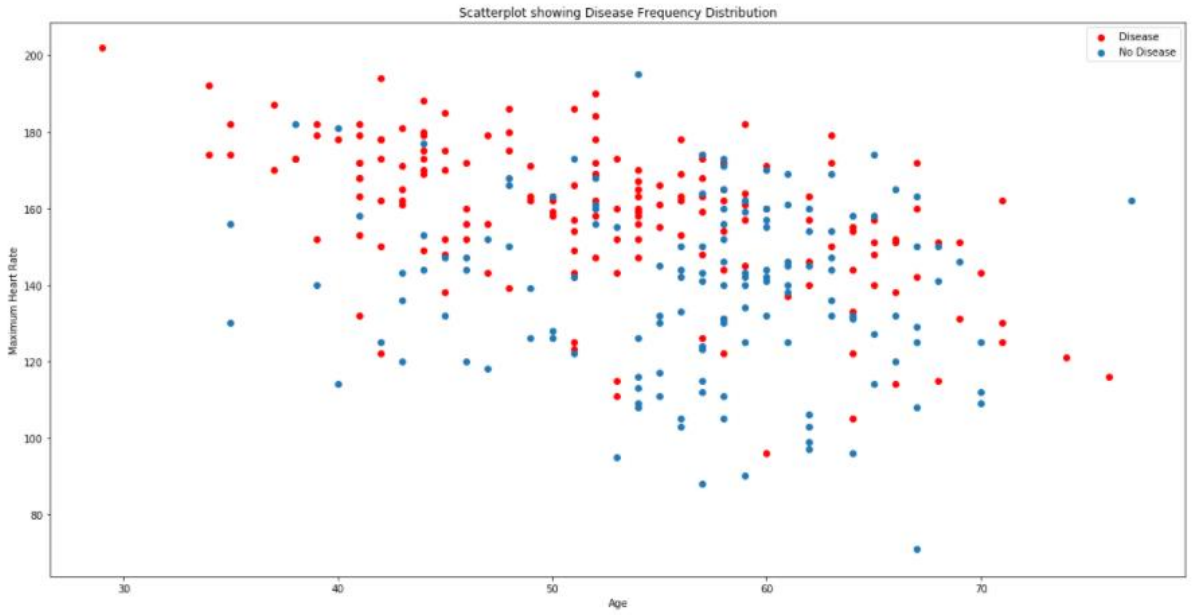


Figure 3: Distribution of heart disease seen in the correlation between age and the maximum heart rates

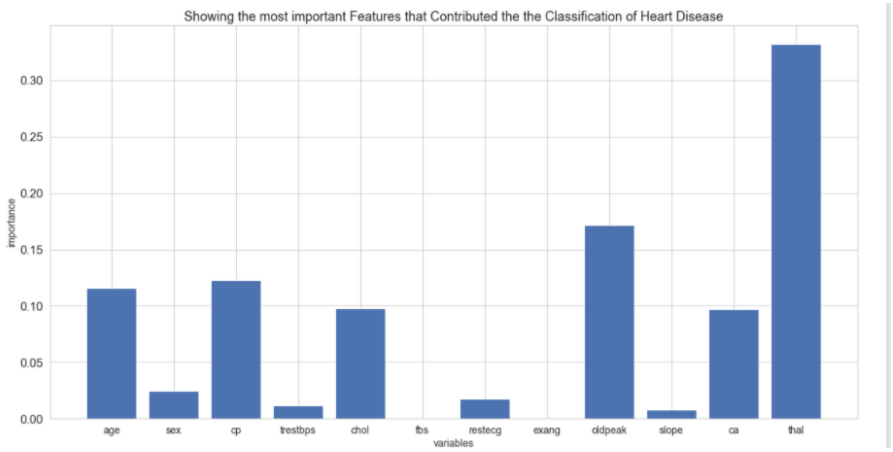


I observed that heart disease was prevalent in men as opposed to women, so I looked further to see the distribution across the ages. Expecting to find the highest heart disease numbers at the extreme end of the age spectrum, I was surprised to see that the highest heart disease numbers were seen in the age range of men in their early 40s to mid 50s (41-54 years old), with the highest numbers coming from 54-year-old

individuals. This observation in Figure 2 feeds into validating Figure 3, as it can be observed that there is a higher presence of heart disease in that same range, and that correlates with the higher heart rates observed in the data set as well.

As there were numerous variables involved in this dataset, I decided to run an analysis so too which attributes had a greater effect on the outcome of the target variable, the presence of heart disease. This was done in at attempt to draw public attention to these factors in particular, and in turn lead to better heart health. The results can be seen in Figure 4.

Figure 4: Displaying The Significance of Variables to Heart Disease Prevention



It was observed that the most important attributes to pay attention to were the age, chest pain, cholesterol, the depression, the number of major blood vessels, and the maximum heart rate.

Results & Conclusions

I ran different regression models on the data and used the accuracy score as the determining metric. The accuracy of a model justifies the usage of that model on your data, as the score determines to what degree the data is a fit for the model. The scores of these regression models were as follows:

Regression Model	Accuracy Score
Logistic Regression	83.61%
Naïve Bayes	85.25%
KNN	90.16

It was observed that the KNN Classifier was the best fit for our data.

In conclusion, as much as these factors all contribute to heart disease, that are not the ultimate determining factors, as some heart diseases can even be inherited; this is just the best course from our given attributes.

References

1. "Heart Disease." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 19 Jan. 2021, www.cdc.gov/heartdisease/index.htm#:~:text=Heart%20disease%20is%20the%20leading,can%20lead%20to%20heart%20attack.
2. Murphy SL, Xu J, Kochanek KD, Arias E. Mortality in the United States, 2017. NCHS data brief, no 328. Hyattsville, MD: National Center for Health Statistics; 2018.
3. "Heart Disease." Mayo Clinic, Mayo Foundation for Medical Education and Research, 9 Feb. 2021, www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118.
4. Donovan, Robin. "Heart Disease: Risk Factors, Prevention, and More." Healthline, Healthline Media, 27 Feb. 2020, www.healthline.com/health/heart-disease.
5. Wedro, Benjamin. "5 Types of Heart Disease Symptoms, Early Signs, Treatment & Causes." MedicineNet, MedicineNet, 31 Aug. 2020, www.medicinenet.com/heart_disease_coronary_artery_disease/article.htm.
6. "What Is Cardiovascular Disease?" Www.heart.org, www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease.
7. "Heart Disease." MedlinePlus, U.S. National Library of Medicine, 12 Feb. 2021, medlineplus.gov/heartdiseases.html.
8. "Coronary Heart Disease." National Heart Lung and Blood Institute, U.S. Department of Health and Human Services, www.nhlbi.nih.gov/health-topics/coronary-heart-disease.
9. Welch, Ashley, and Kaitlin Sullivan. "What Is Heart Disease? Symptoms, Causes, Diagnosis, Treatment, and Prevention: Everyday Health." EverydayHealth.com, www.everydayhealth.com/heart-disease/.
10. Richard N. Fogoros, MD. "Heart Disease." Verywell Health, www.verywellhealth.com/heart-disease-4014709.