

Project 1

William Davis

6/11/2022

Background

- In America, 1 in every 4 males suffer from heart disease related deaths each year.
- This attributed to 357,761 deaths in 2019. I chose the Heart Attack Analysis and Prediction Dataset to help scientist and medical professional prevent future events of heart attack.
- This dataset will predict the likelihood of a patient going into cardiac arrest.

```
heart_data = read.csv("heart.csv", sep = ",")  
heart_data$cp <- as.factor(heart_data$cp)  
heart_data$sex <- as.factor(heart_data$sex)  
heart_data$age <- as.factor(heart_data$age)  
heart_data$restecg <- as.factor(heart_data$restecg)  
heart_data$exng <- as.factor(heart_data$exng)  
heart_data$fbs <- as.factor(heart_data$fbs)  
heart_data$output <- as.factor(heart_data$output)  
heart_data$thall <- as.factor(heart_data$thall)
```

The heart attack analysis and prediction dataset, used many fields that needed to cleaned and converted to factors. In this dataset the following values were converted into factors.

- age
- sex
- cp (chest pain)
- restecg (resting electrocardiographic results)
- exng (exercise induced angina)
- fbs (fasting blood sugar > 120 mg/dl)
- thall
- output (likely-hood of having a heart attack)

Feature Engineering

To validate the data values for thalach (maximum heart rate achieved) and trtbps (resting blood pressure), The values have been normalized and used in the predictive model.

```
meanMaxHeartRateAtt = mean(heart_data$thalachh)
s = sd(heart_data$thalachh)
heart_data$MaxHeartNormal = (heart_data$thalachh - meanMaxHeartRateAtt)/s

meanRestBlPress = mean(heart_data$trtbps)
s = sd(heart_data$trtbps)
heart_data$RestBloodPressNormal = (heart_data$trtbps - meanRestBlPress)/s

diabetes = heart_data$fbs

diabetesChance <- character()

for(i in 1:length(diabetes)){
  if(diabetes[i] == 1){
    diabetesChance[i] <- 1L
  }else if(diabetes[i] == 0){
    diabetesChance[i] <- 0L
  }
}

heart_data$twiceAsLikely = diabetesChance
```

According to the CDC, men with diabetes are twice as likely to enter cardiac arrest. With this in mind, a patient with a fbs above 120 is likely to have diabetes. The feature “twiceAsLikely” has been created and added to the dataset to contribute to the predictive model with 1 = twice as likely, and 2 = not twice as likely.

Data Sampling and Predictive Modeling

```
heart_dataSelect <- select(heart_data, age, sex, caa, cp, chol,
                           trtbps, fbs, thall, thalachh, exng,
                           output, MaxHeartNormal,
                           RestBloodPressNormal, twiceAsLikely)

samples <- createDataPartition(heart_data$output,
                             times = 1,
                             p = 0.80,
                             list = F)

heart_dataSelectTrain <- heart_dataSelect[samples,]
heart_dataSelectTest <- heart_dataSelect[-samples,]

control <- trainControl(method = "repeatedcv",
                        number = 10,
                        repeats = 2,
                        search = "grid")

cl <- makeCluster(3, type = "SOCK")
registerDoSNOW(cl)
set.seed(1234)
caret.cv <- train(output ~.,
                   data = heart_dataSelectTrain,
                   method = "multinom",
                   preProcess = ("center"),
                   trControl = control)

## # weights: 58 (57 variable)
## initial value 168.434765
## iter 10 value 79.089752
## iter 20 value 72.172628
## final value 72.170874
## converged
stopCluster(cl)

pred <- predict(caret.cv, heart_dataSelectTest)
```

Multinomial Logistic Regression was used to determine the likelihood of a patient entering cardiac arrests. The predictive model was created using the caret and doSnow libraries. Multiple clusters were created to reduced the time to compute the results.

Results

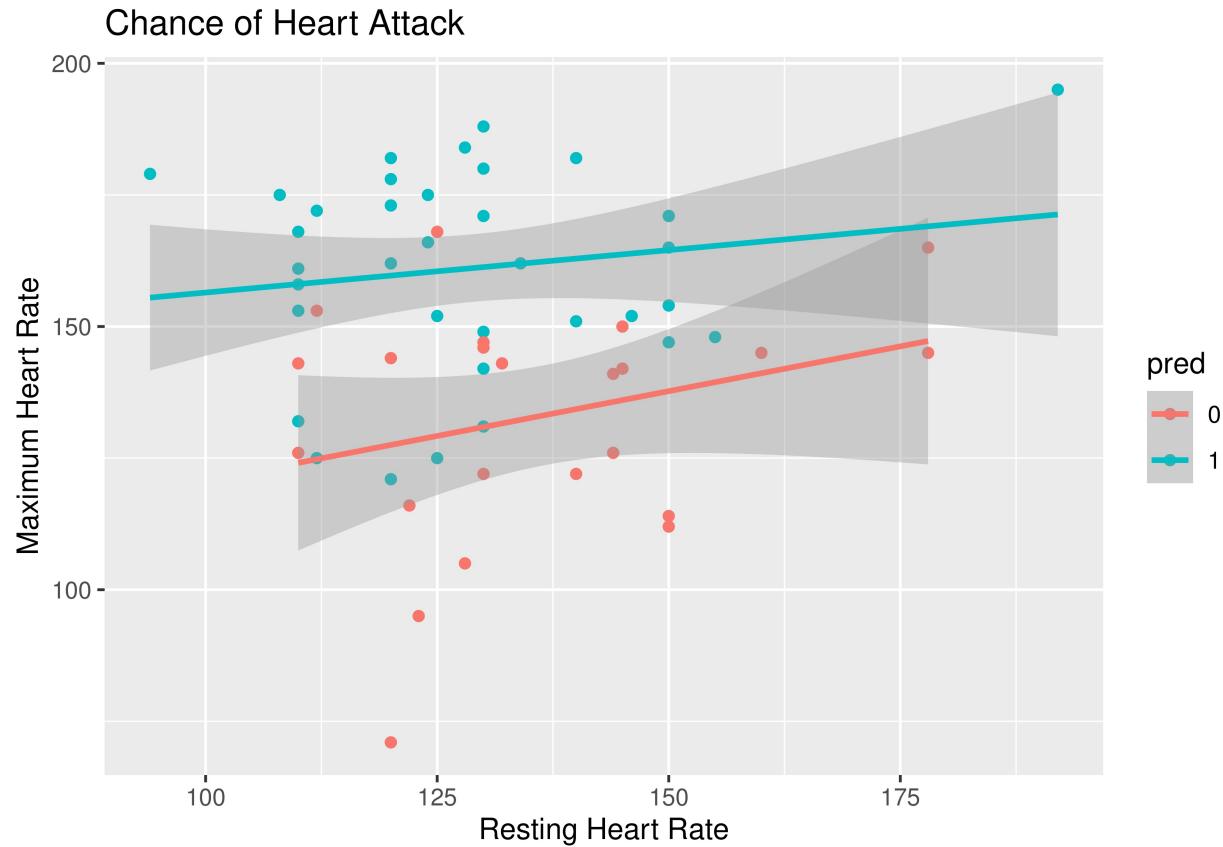
A confusion matrix was created to verify the accuracy of the data.

```
confusionMatrix(pred, heart_dataSelectTest$output)

## Confusion Matrix and Statistics
##
##             Reference
## Prediction  0   1
##          0 18   6
##          1   9 27
##
##                  Accuracy : 0.75
##                         95% CI : (0.6214, 0.8528)
##      No Information Rate : 0.55
##      P-Value [Acc > NIR] : 0.001116
##
##                  Kappa : 0.4898
##
## McNemar's Test P-Value : 0.605577
##
##                  Sensitivity : 0.6667
##                  Specificity  : 0.8182
##      Pos Pred Value : 0.7500
##      Neg Pred Value : 0.7500
##                  Prevalence : 0.4500
##      Detection Rate : 0.3000
##      Detection Prevalence : 0.4000
##      Balanced Accuracy : 0.7424
##
##      'Positive' Class : 0
##

ggplot(heart_dataSelectTest,
       aes(x = trtbps,
            y = thalachh,
            col = pred)) +
  geom_point() +
  geom_smooth(method = lm) +
  ggtitle("Chance of Heart Attack ") +
  xlab("Resting Heart Rate") +
  ylab("Maximum Heart Rate")

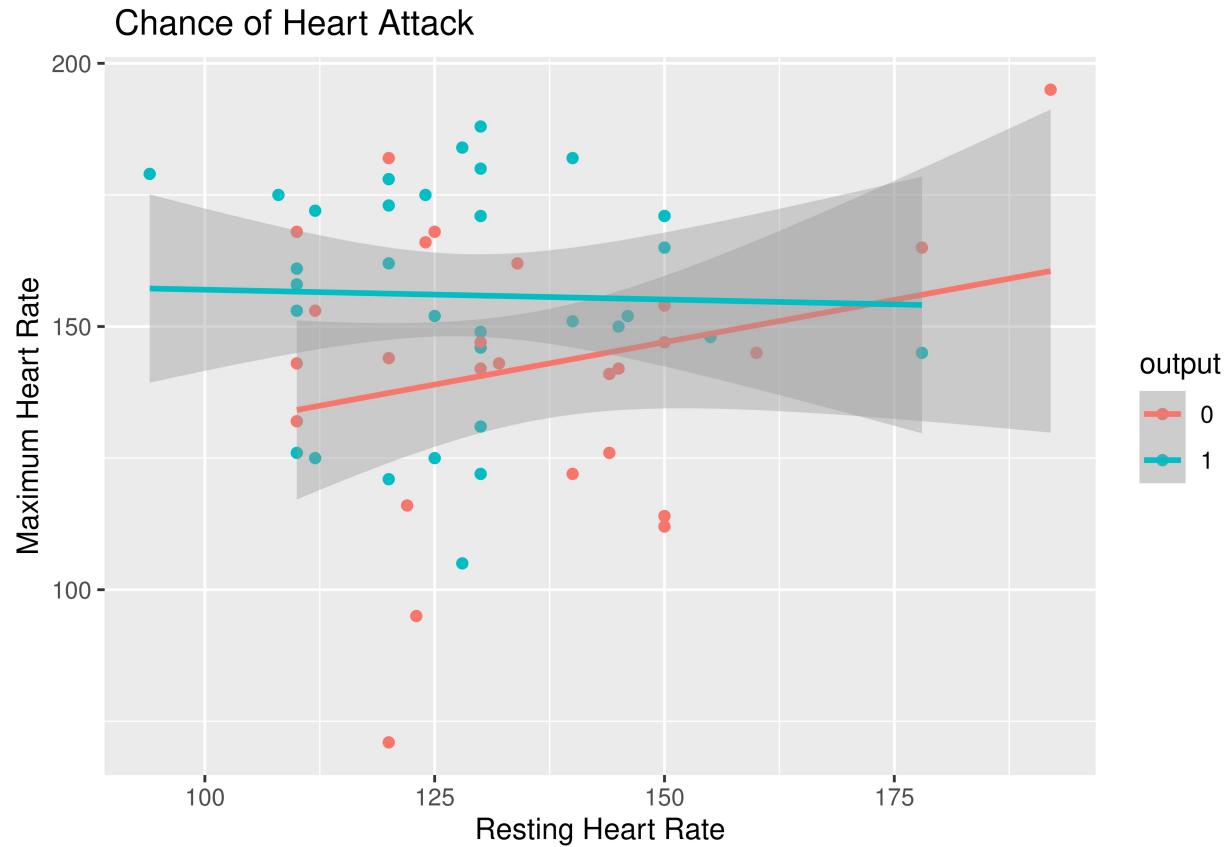
## `geom_smooth()` using formula 'y ~ x'
```



Plotting the resting heart rate vs the max rate, we can see that we predicted that if the as the maximum heart rate increases the chance of a patient entering cardiac arrest increase, even if the resting heart rate is lower.

```
options(knitr.duplicate.label = "allow")
ggplot(heart_dataSelectTest,
       aes(x = trtbps,
           y = thalachh,
           col = output)) +
  geom_point() +
  geom_smooth(method = lm) +
  ggtitle(" Chance of Heart Attack ") +
  xlab("Resting Heart Rate") +
  ylab("Maximum Heart Rate")

## `geom_smooth()` using formula 'y ~ x'
```

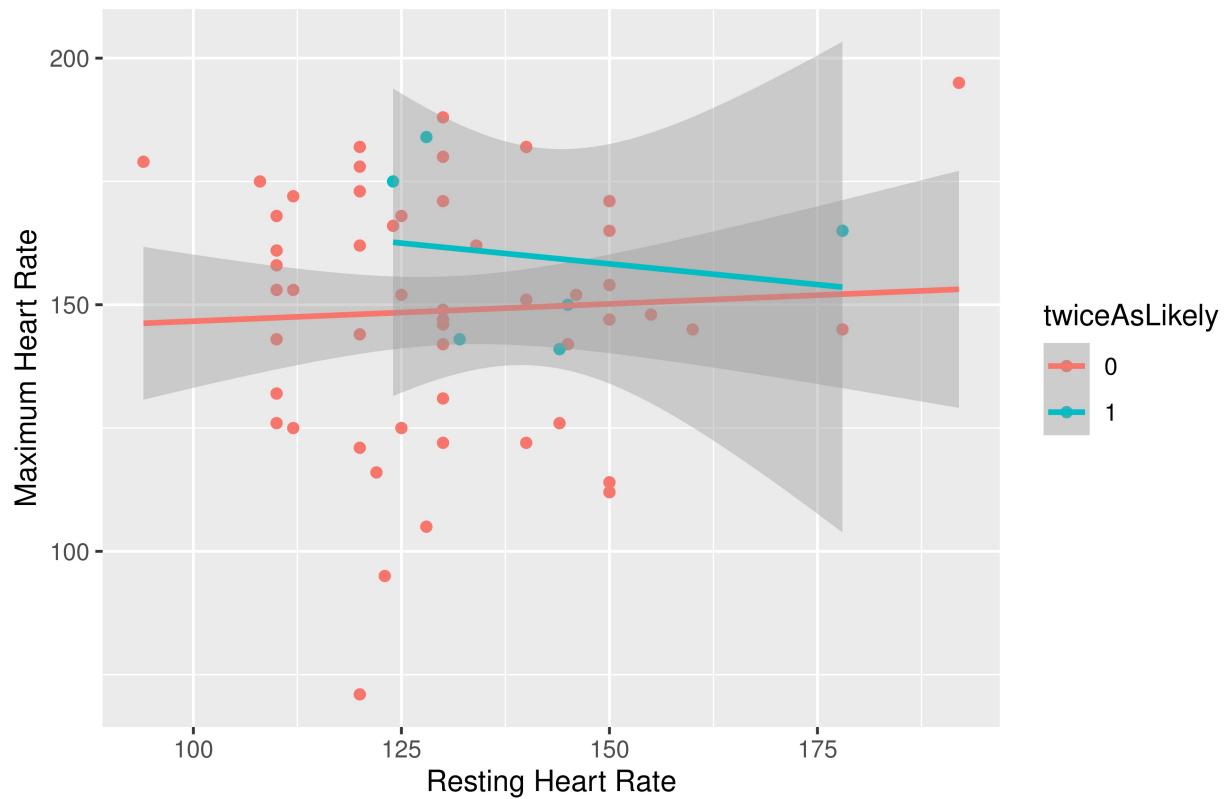


```
## The same is true for the initial un-predicted outcome.
```

```
ggplot(heart_dataSelectTest,
       aes(x = trtbps,
           y = thalachh,
           col = twiceAsLikely)) +
  geom_point() +
  geom_smooth(method = lm) +
  ggtitle("Chance of Heart Attack ") +
  xlab("Resting Heart Rate") +
  ylab("Maximum Heart Rate")

## `geom_smooth()` using formula 'y ~ x'
```

Chance of Heart Attack



#Dataset

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/discussion/234843>

#References

D. (2020, May 28). Multinomial logistic regression With R. R-Bloggers. <https://www.r-bloggers.com/2020/05/multinomial-logistic-regression-with-r/#:%7E:text=Overview%20%E2%80%93%20Multinomial%20logistic%20Regression,to%20use%20ordinal%20logistic%20regression>. Dalpiaz, D. (2020, October 28). Chapter 21 The caret Package | R for Statistical Learning. Github.Io.

<https://daviddalpiaz.github.io/r4sl/the-caret-package.html> Diabetes and Your Heart. (2021, May 7). Centers for Disease Control and Prevention.

<https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html#:~:text=If%20you%20have%20diabetes%C2%20you,are%20to%20have%20heart%20disease>. ECG test - Better Health Channel. (2022). ECG Test - Better Health Channel. Retrieved June 13, 2022, from [https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/ecg-testHeartAttack Analysis & Prediction Dataset. \(2021, March 22\). Kaggle](https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/ecg-testHeartAttack Analysis & Prediction Dataset. (2021, March 22). Kaggle).

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/discussion/234843> Men and Heart Disease. (2022, March 8). Centers for Disease Control and Prevention.

[https://www.cdc.gov/heartdisease/men.htm#:~:text=About%201%20in%2013%20\(7.7,men%20have%20coronary%20heart%20disease.&text=Half%20of%20the%20men%20who,disease%20had%20no%20previous%20symptoms.&text=Even%20if%20you%20have%20no,at%20risk%20for%20heart%20disease](https://www.cdc.gov/heartdisease/men.htm#:~:text=About%201%20in%2013%20(7.7,men%20have%20coronary%20heart%20disease.&text=Half%20of%20the%20men%20who,disease%20had%20no%20previous%20symptoms.&text=Even%20if%20you%20have%20no,at%20risk%20for%20heart%20disease). Prediabetes - Diagnosis and treatment - Mayo Clinic. (2022, March 17).

Mayoclinic.Org. [https://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/drc-20355284#:~:text=Fasting%20blood%20sugar%20test&text=L\)%20of%20blood.-,In%20general%3A,tests%20is%20diagnosed%20as%20diabetes](https://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/drc-20355284#:~:text=Fasting%20blood%20sugar%20test&text=L)%20of%20blood.-,In%20general%3A,tests%20is%20diagnosed%20as%20diabetes)