Leveraging Risk Scores to Reduce Heart Attacks in the Workplace

Group 2
29 November 2023
Team Leader: Skanda Sharma
Team Members: Matthew Dominguez, Vansh Deep, Jaden Le, Benjamin Guberman

Agenda

- Introduction
- Problem Statement
- Objectives
- Methods
- Descriptive statistics
- Data Visualization
- Results
- Conclusion and Recommendation
- Lessons learned
- Challenges

Introduction

- What is a Heart Attack?
- Heart Attacks : an increasing trend
  - Workplace Stress

Problem Statement

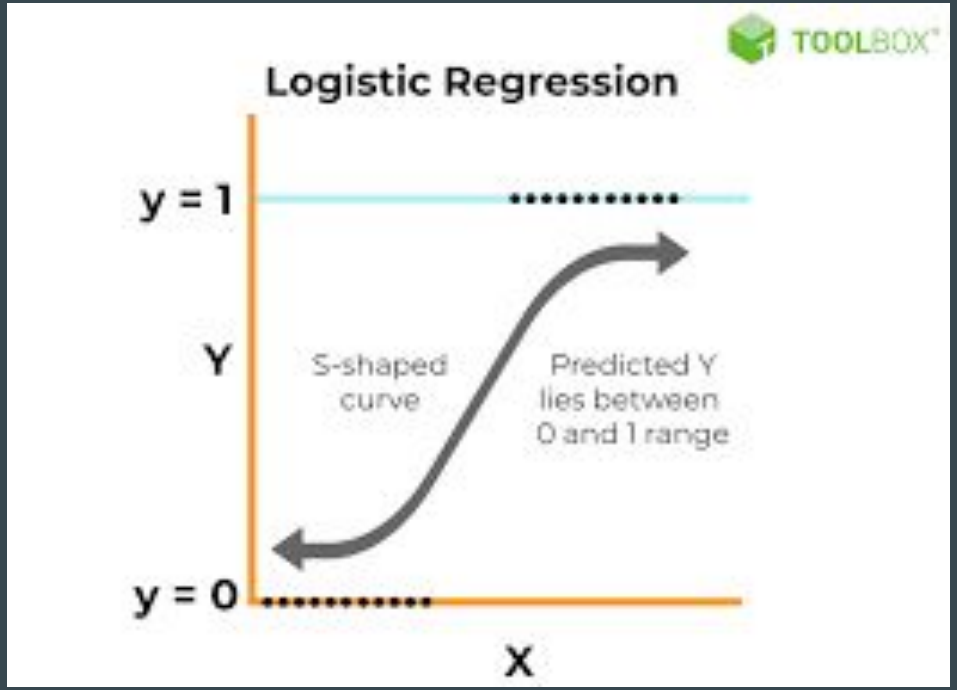
- Companies are losing high value employees due to sudden heart attacks
    - Decreased workplace productivity
    - Increased hiring and training costs

Objective Function

- Recognize key factors contributing to heart attacks
- Develop a model which predicts the risk score % using these factors
- Implement the results for companies to make changes in workplace and employee health

Method

- Logistic Regression

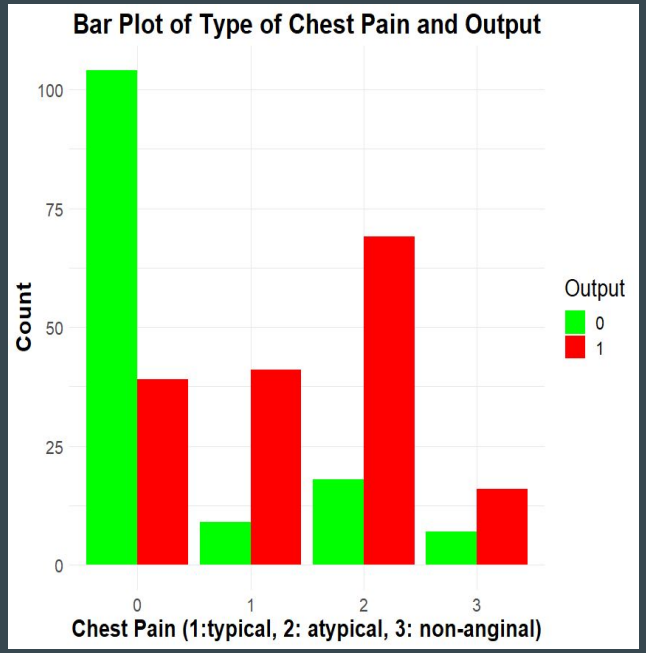


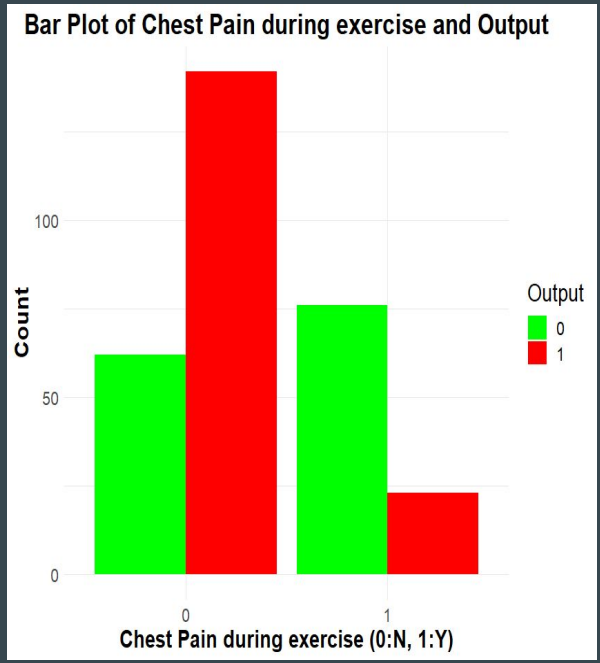
Descriptive Statistics

```
> summary(data)
 sex      cp        thalachh       exng      oldpeak          caa          output
 0: 96   0:143   Min.   : 71.0   0:204   Min.   :0.00   Min.   :0.0000   0:138
 1:207   1: 50   1st Qu.:133.5   1: 99   1st Qu.:0.00   1st Qu.:0.0000   1:165
         2: 87   Median :153.0           Median :0.80   Median :0.0000
         3: 23   Mean   :149.6           Mean   :1.04   Mean   :0.7294
                 3rd Qu.:166.0           3rd Qu.:1.60   3rd Qu.:1.0000
                 Max.   :202.0           Max.   :6.20   Max.   :4.0000
>
```

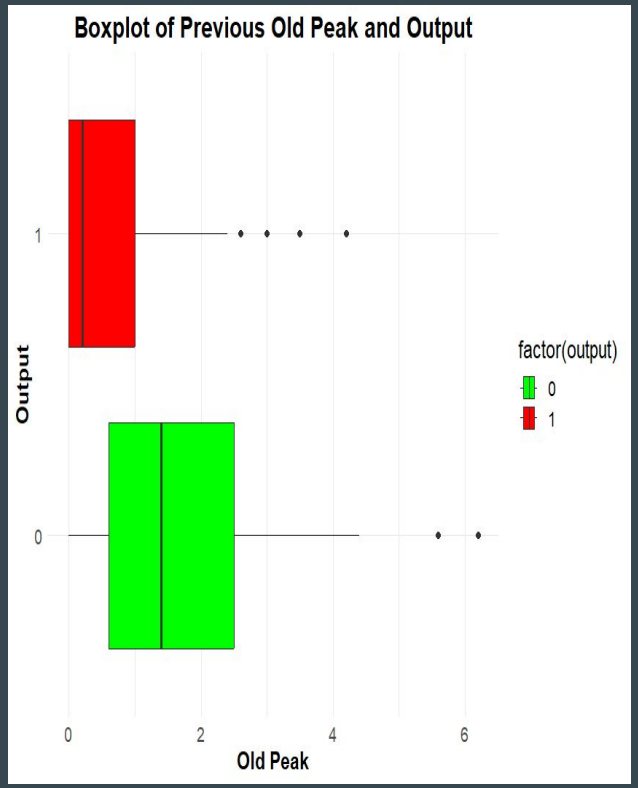
- CP ~ Type of chest pain (1: typical angina, 2: atypical angina, 3: non-anginal pain)
- Thalachh ~ Max Heart Rate Achieved
- Exng ~ Exercise induced angina (Exercise that makes the heart work hard)
- Old Peak ~ Amount of oxygen heart muscles receive (ECG machine) **
- Caa ~ # of major blood vessels in the heart
- Output(Y) ~ 0: Healthy, 1: At Risk for Heart Attack

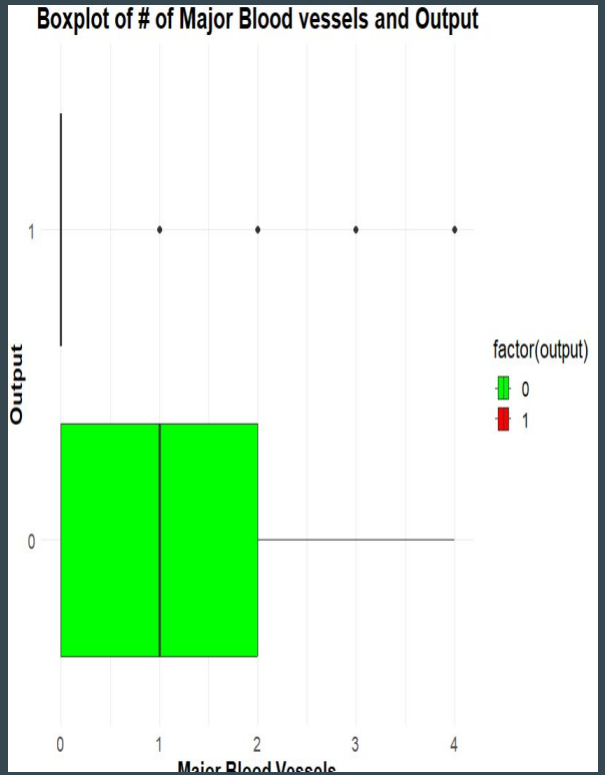
Data Visualization - Categorical Variables

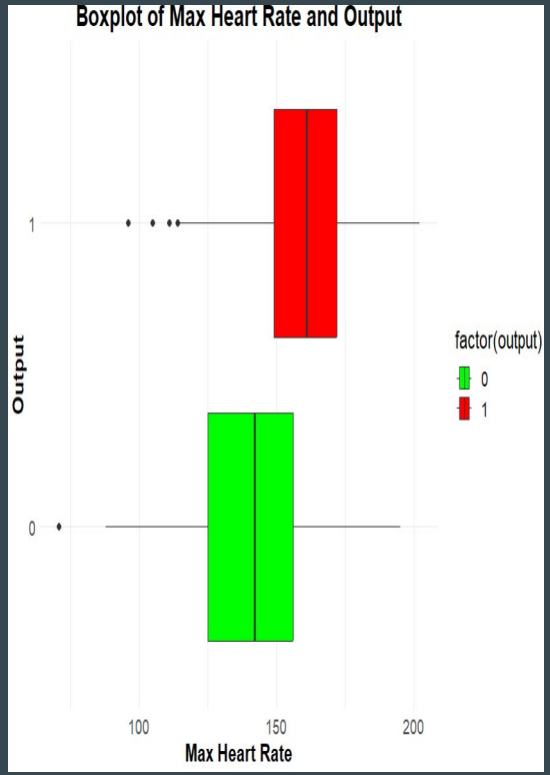




Data Visualization - Continuous Variables







Data Analysis

- Binomial Logistic Regression
  - Leveraging the probability an employee might be at a risk for heart attack
  - Alpha = .05

- Probability (Y= 1) = $\frac{e^{-1.41 - 1.52x1 + 1.22x2 \ldots\ldots -.72x7}}{1 + e^{-1.41 - 1.52x1 + 1.22x2 \ldots\ldots -.72x7}}$

```
(Intercept)        sex1         cp1         cp2         cp3    thalachh        exng1
 -1.4121152  -1.5244929   1.2214219   1.9518797   1.8507347   0.0220054  -1.0204647
    oldpeak         caa
 -0.7014139  -0.7225997
>
```

Results

- McFadden's Pseudo R squared
  - r^2 = .44
- Example Risk Score
  - Anyone above a 75% is at a risk for heart attack

```
> # Example usage
> sex1 <- 1
> cp1 <- 0
> cp2 <- 0
> cp3 <- 0
> thalachh <- 185
> exng <- 0
> oldpeak <- 1
> caa <- 3
> predicted_probability <- calculate_probability(sex1, cp1, cp2, cp3, thalachh, exng, oldpeak, caa)
> cat(predicted_probability, "%")
15 %
```

Conclusion and Recommendations

Heart rate risk predictors:

- Chest pain
- Sex
- Maximum heart rate achieved
- Exercise
- Amount of oxygen heart muscles receive
- Number of major blood vessels

Business Recommendations:

- Management can assess individual heart attack risk using our model
- Implement monthly screenings for high blood pressure and other vital factors
- Provide access to therapeutic counseling or programs

Lessons Learned

1. Age, resting blood pressure, cholesterol, fasting blood sugar, resting heart rate, and slope show no significant impact on heart attack risk.
2. There is a significant reward for exercise, stretching, or moving from time to time.
3. Gender/sex contributes to the significance of an individual's heart attack risk
4. Learned how to accurately conduct a regression analysis problem.

Challenges

- Scheduling Conflicts
- Time Management
- Understanding the Data
- Changing Our Statistical Approach/Method to the Problem

Thank you!