



APRENDIZAGEM DE MAQUINAS - CIENCIA DA COMPUTACAO / UESB

# MODELO PREDITIVO PARA INSUFICIÊNCIA CARDÍACA

Jefferson Antônio T. Silva

João Paulo Sandes Brito





Desenvolver um modelo de Machine Learning capaz de prever a ocorrência de insuficiência cardíaca em pacientes, utilizando métricas presentes no dataset Heart Failure Prediction.

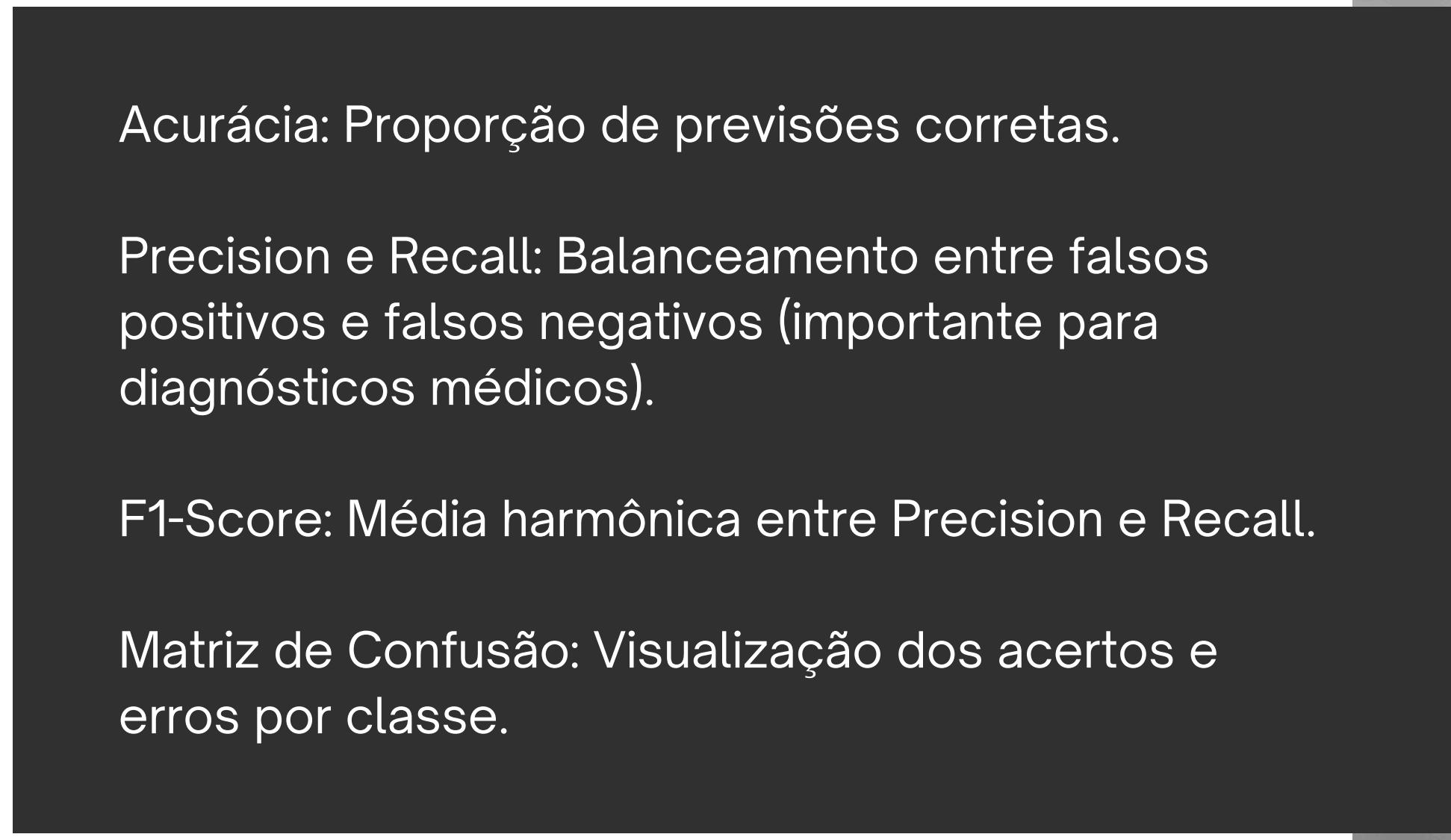
# Objetivo



# Justificativa



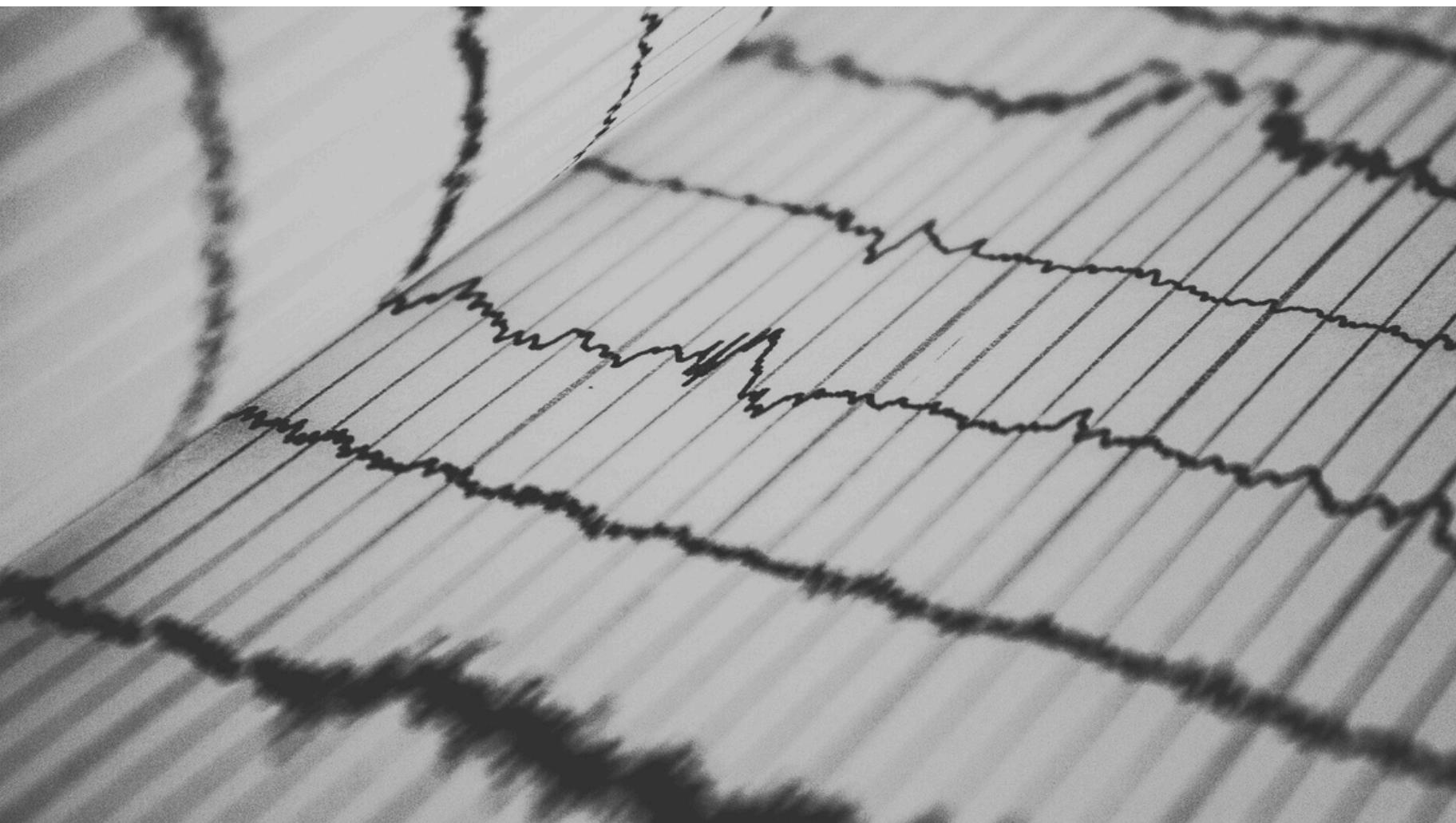
A previsão de insuficiência cardíaca é um problema relevante na área de saúde, com potencial para reduzir custos hospitalares e salvar vidas.



# Métricas de Avaliação do Modelo



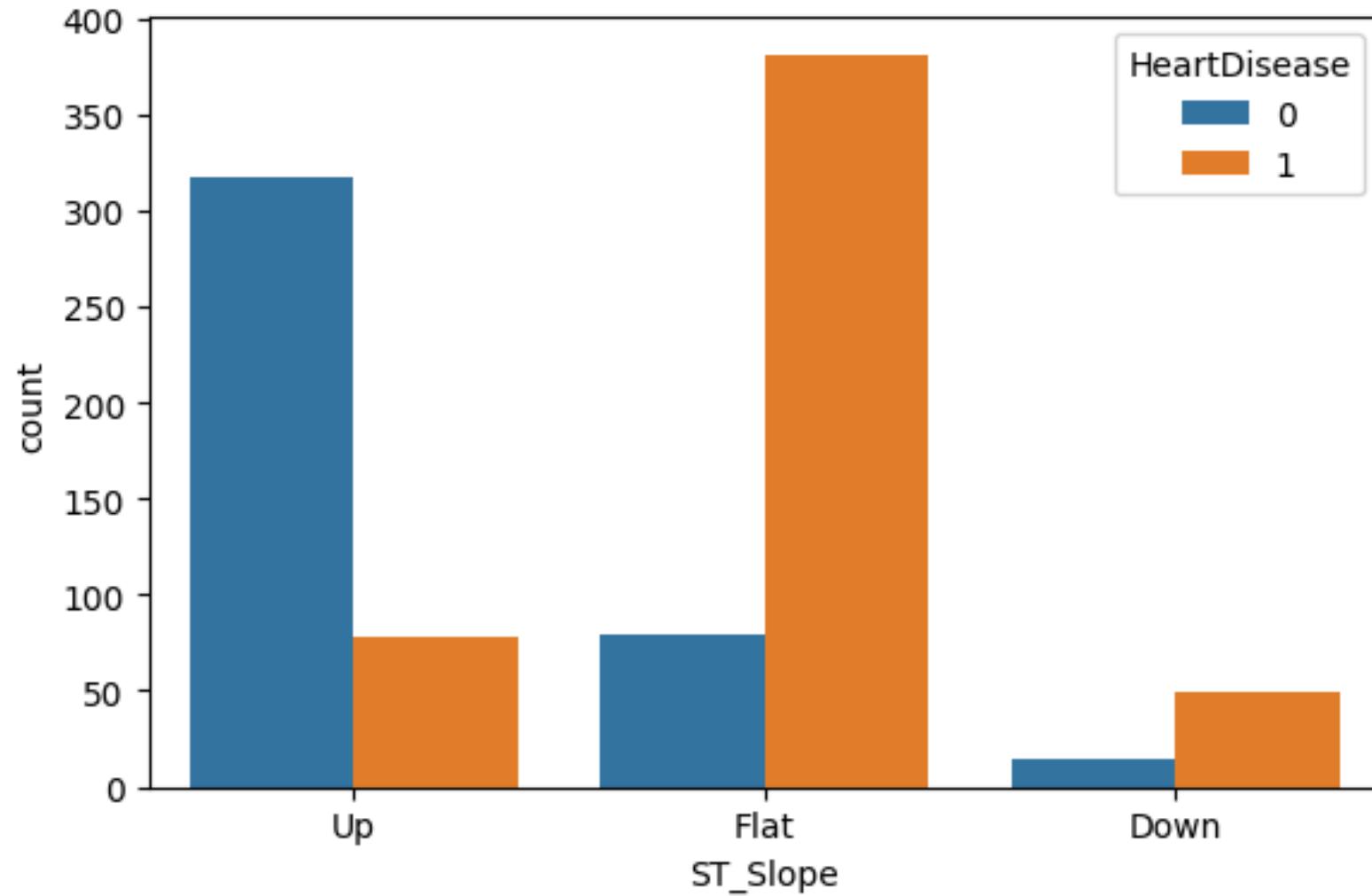
# Correlação de Variáveis



A matriz de correlação é uma tabela que mostra a relação linear entre variáveis numéricas.

1. Identifica padrões: Descobre quais variáveis estão associadas.
2. Detecta redundâncias: Variáveis muito correlacionadas podem ser removidas.
3. Guia análises: Direciona a escolha de features em modelos de machine learning.

# ST\_Slope vs. HeartDisease

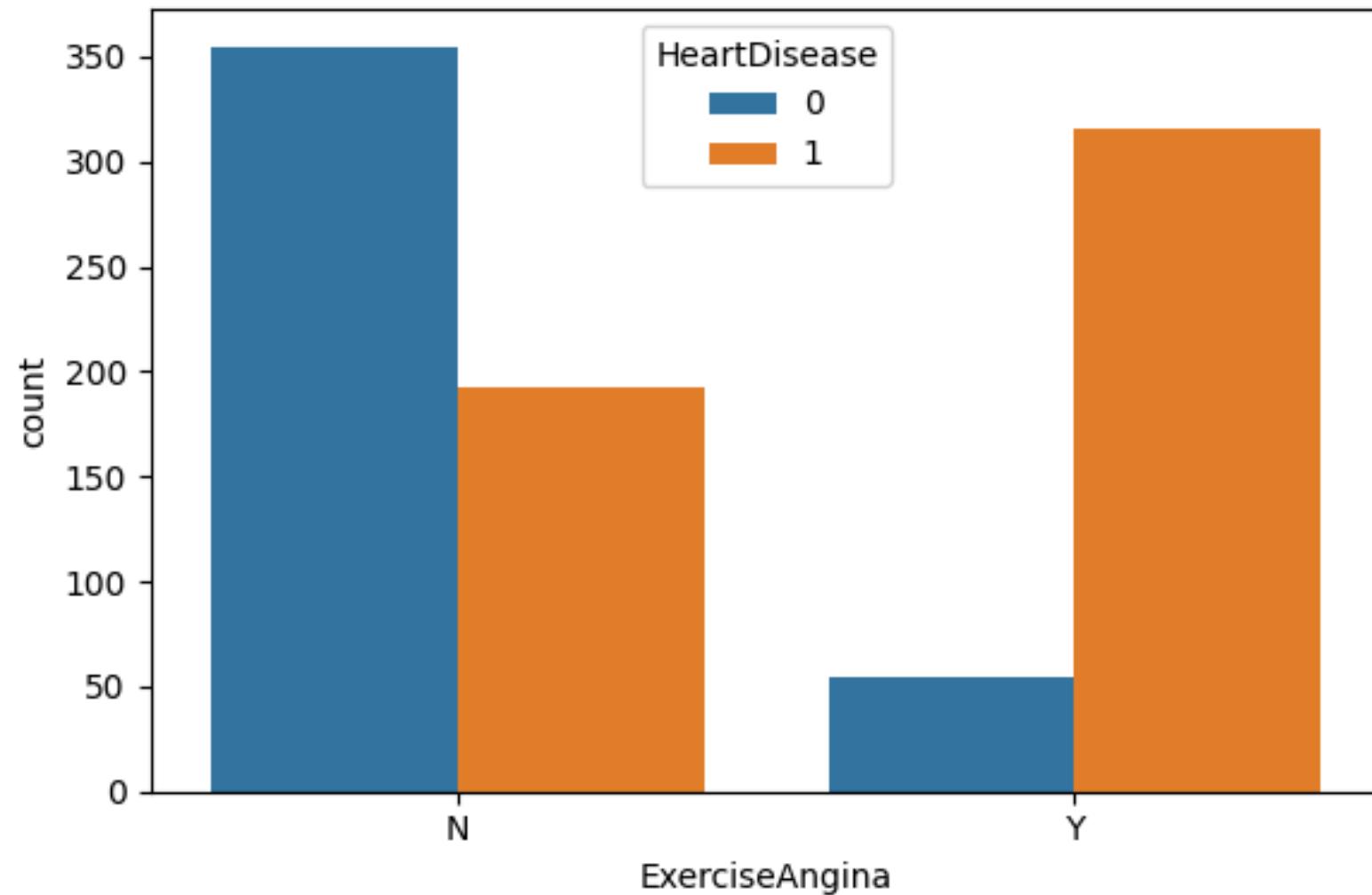


## ST\_Slope (Inclinação do Segmento ST) (-0.56)

Mostra que inclinações descendentes ("Down") ou planas ("Flat") estão fortemente associadas a maior risco cardíaco. Essas alterações indicam problemas no fluxo sanguíneo cardíaco.



# ExerciseAngina vs. HeartDisease

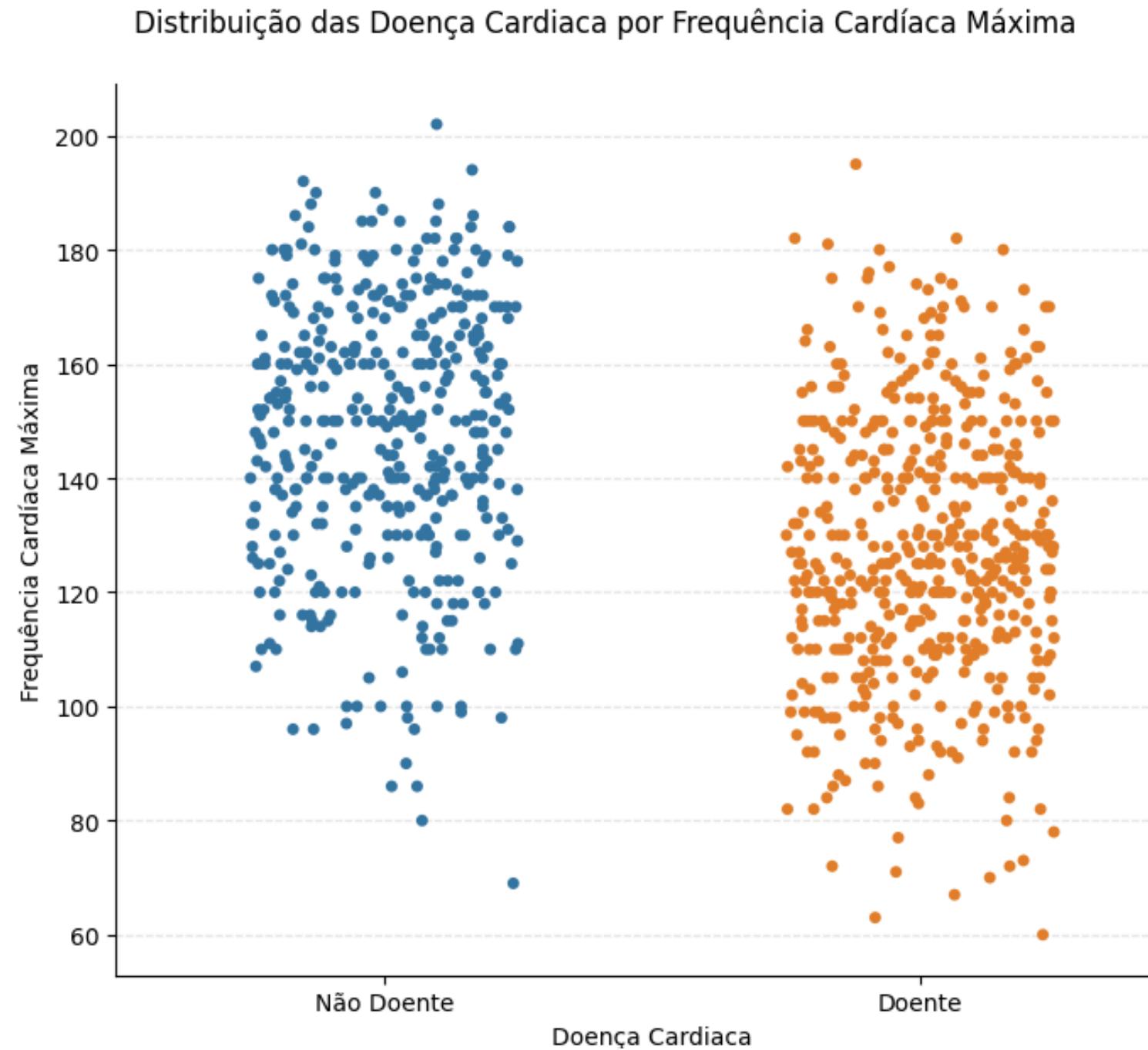


## ExerciseAngina (Dor durante o Exercício) (0.49)

Revela que quem experimenta angina durante exercício tem mais probabilidade de doença cardíaca, sendo um sinal clássico de obstrução arterial.



# MaxHR vs. HeartDisease



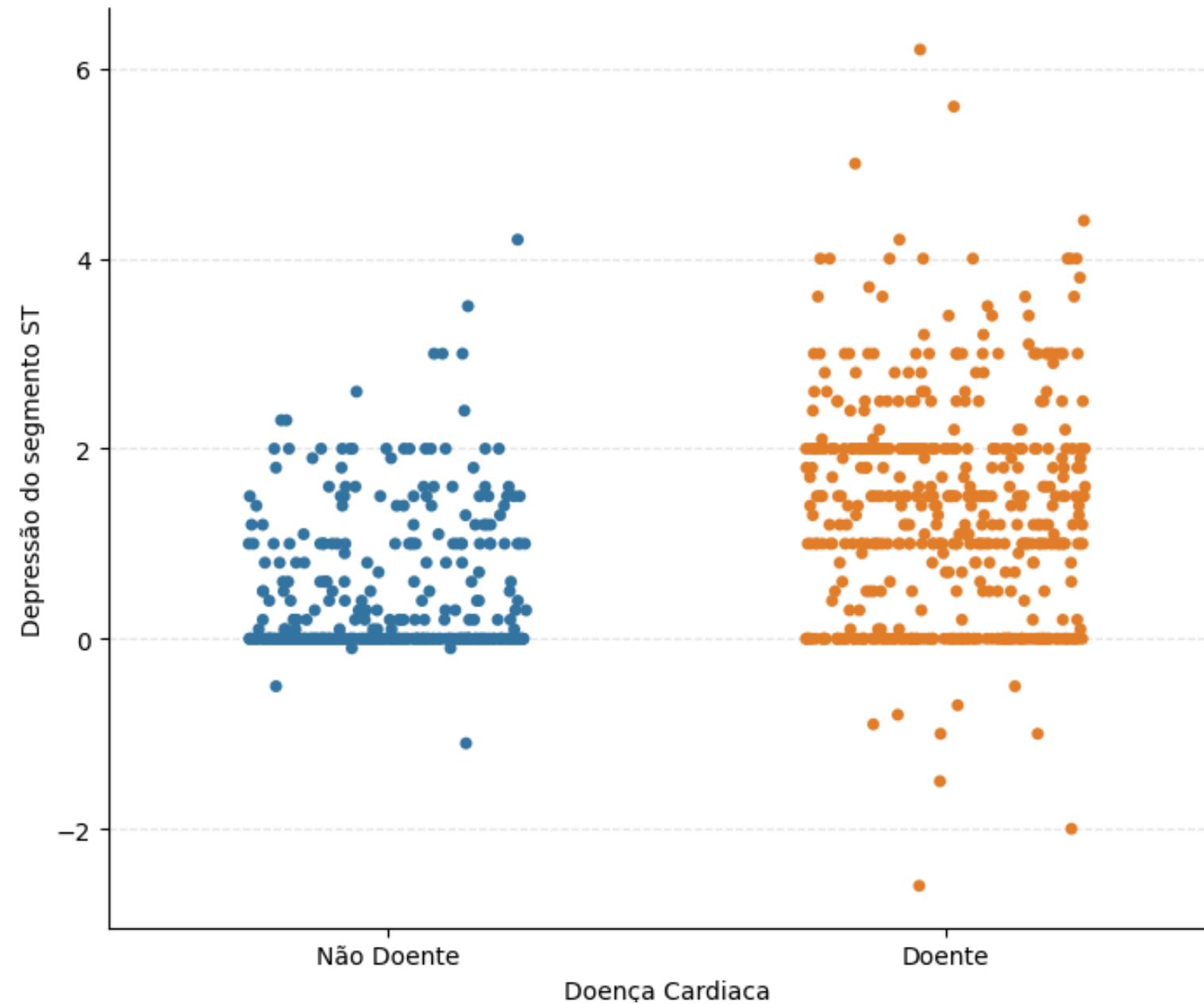
## MaxHR (Frequência Cardíaca Máxima) (-0.40)

Indica que valores mais baixos estão associados a maior risco. Isso sugere que capacidade cardiovascular reduzida pode ser um fator de risco importante.



# Oldpeak vs. HeartDisease

Distribuição das Doença Cardiaca por Depressão do segmento ST

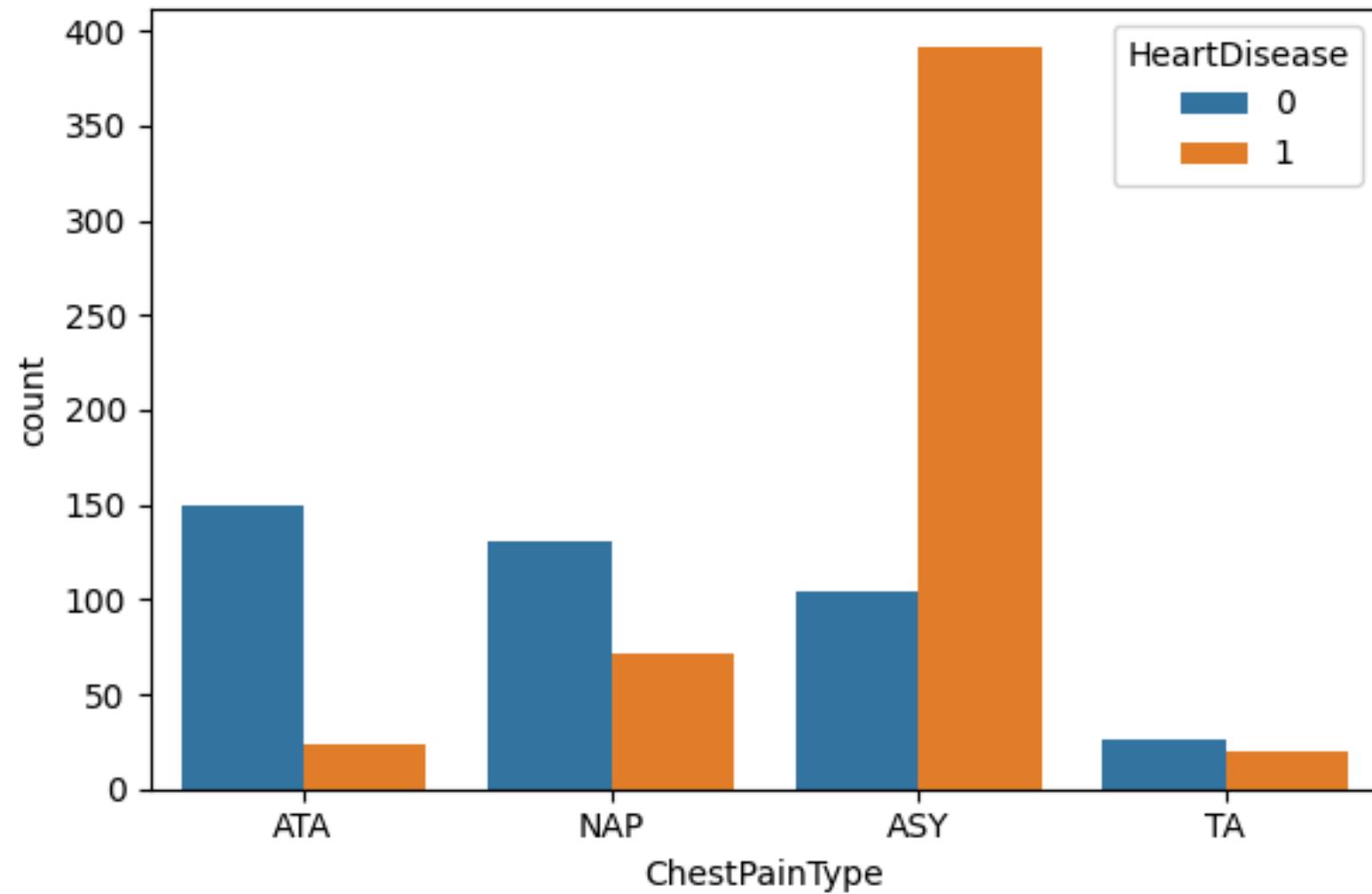


## Oldpeak (Depressão do ST) (0.40)

Valores mais altos indicam maior risco cardíaco, pois refletem anormalidades no suprimento de oxigênio ao músculo cardíaco.



# ChestPainType vs. HeartDisease

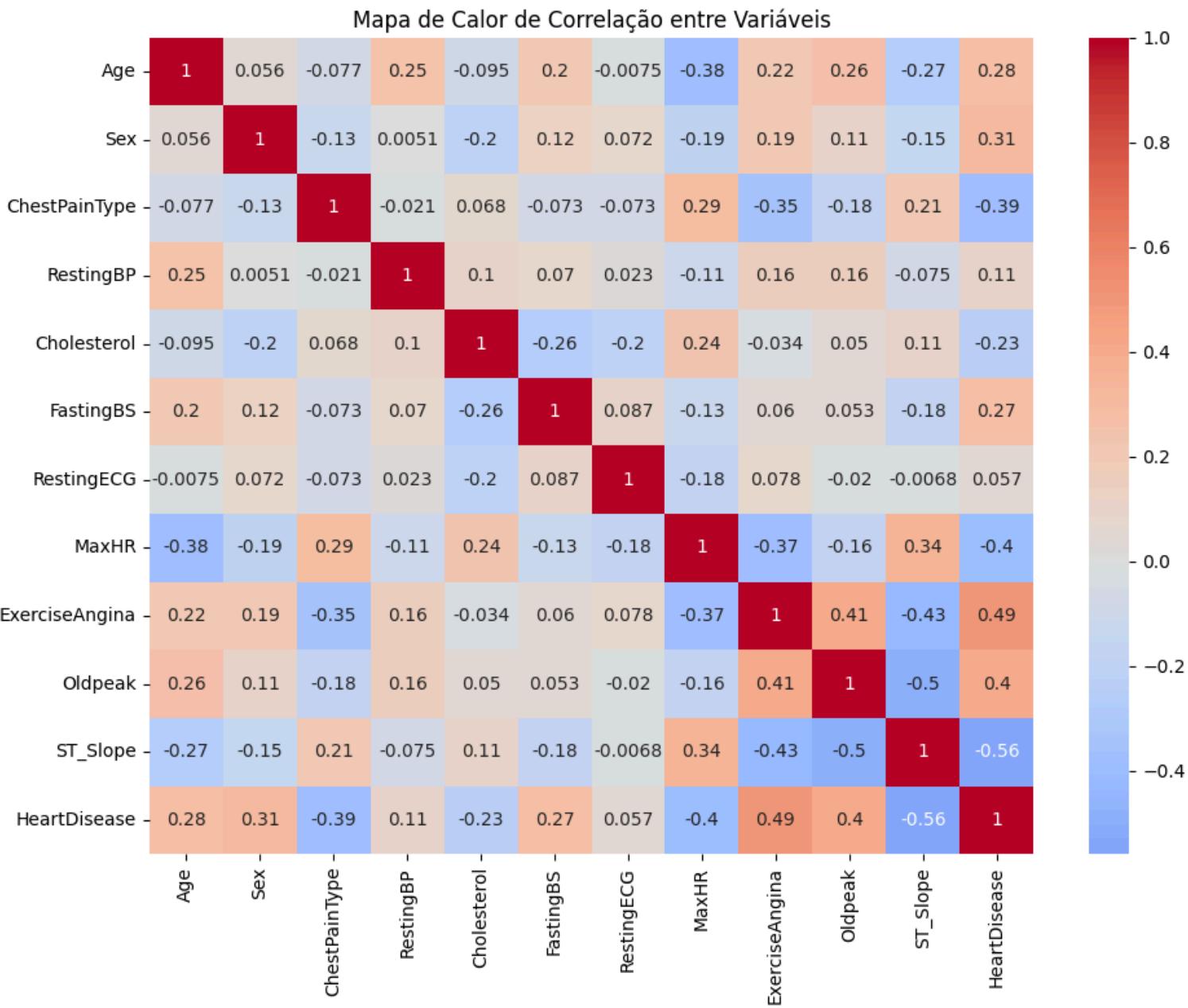


## ChestPainType (Tipo de Dor Torácica) (-0.39)

Mostra que dores atípicas e casos assintomáticos estão associados a maior risco, especialmente preocupante em diabéticos e idosos onde sintomas podem ser silenciosos.



# Correlações Notáveis



**ExerciseAngina ↔ ST\_Slope (-0.43):**

- Pacientes com angina durante exercício tendem a ter inclinação ST anormal

**ST\_Slope ↔ Oldpeak (-0.50):**

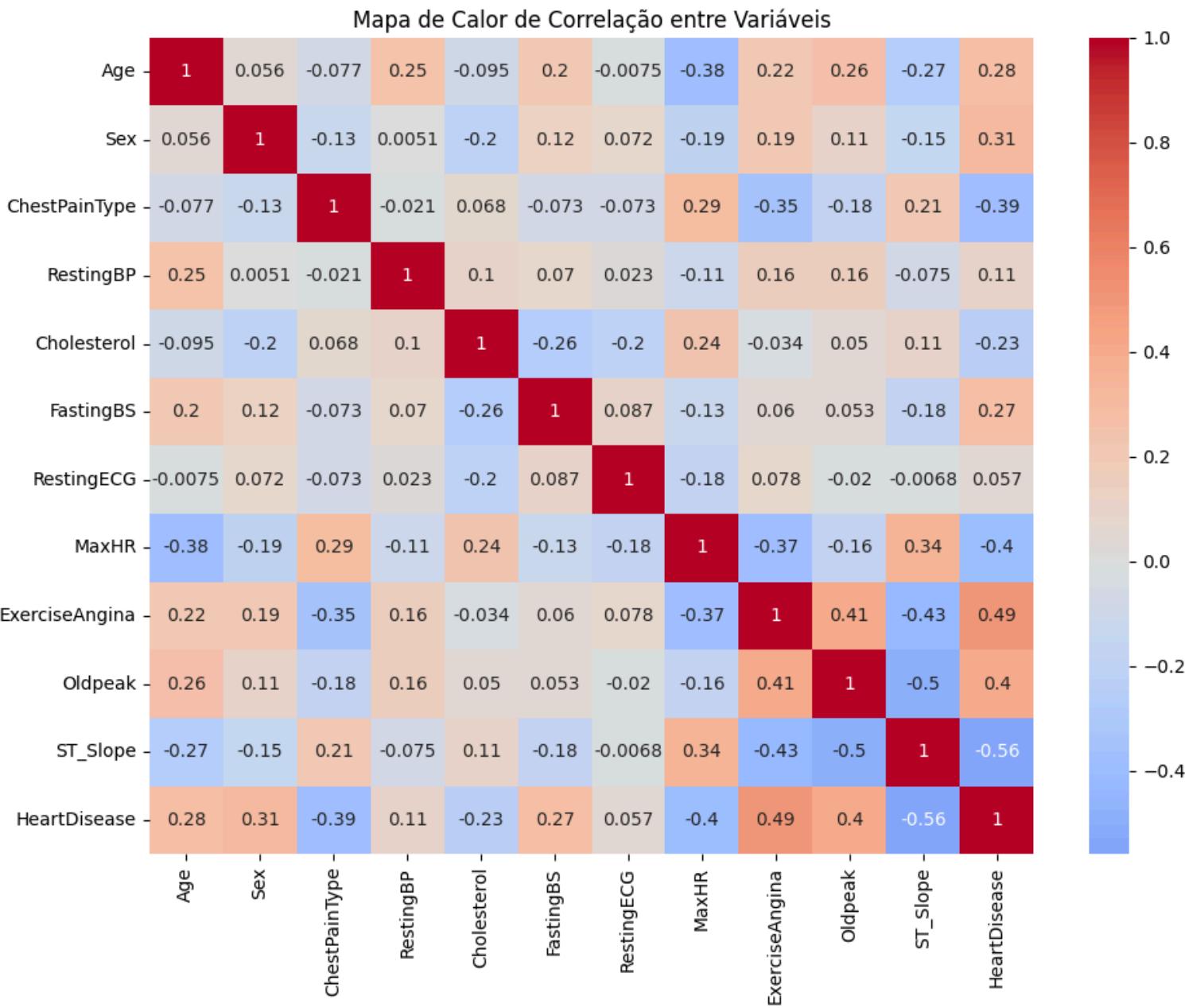
- Inclinação anormal do ST correlacionada com depressão do segmento ST (esperado clinicamente)

**ExerciseAngina ↔ Oldpeak (0.41):**

- Angina durante exercício associada a maior depressão do ST



# Correlações Moderadas



**Age ↔ HeartDisease (0.28)**

- Idade avançada aumenta risco

**Sex ↔ HeartDisease (0.31)**

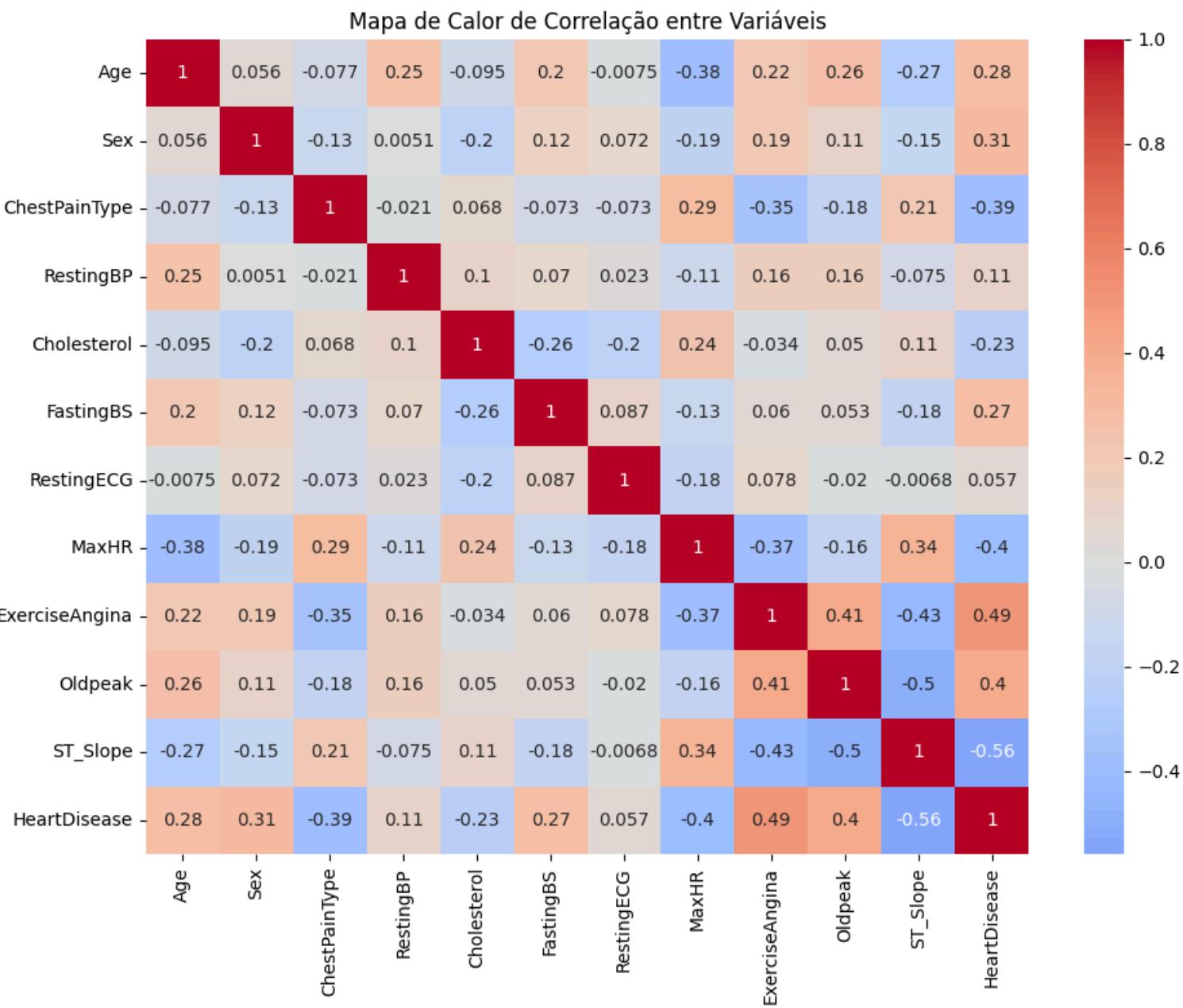
- Sexo masculino (M=0) associado a maior risco

**MaxHR ↔ Age (-0.38)**

- Idade avançada → Frequência cardíaca máxima mais baixa



# Surpresas e Alertas



## Cholesterol ↔ HeartDisease (-0.23):

- Correlação negativa inesperada!
- Possíveis causas: viés amostral

## RestingBP ↔ HeartDisease (0.11):

- Sugere que outros fatores são mais determinantes

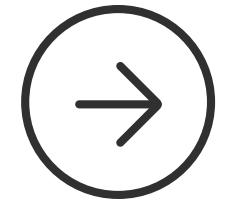
## FastingBS ↔ HeartDisease (0.27):

- Glicemia elevada em jejum aumenta risco cardíaco (esperado)





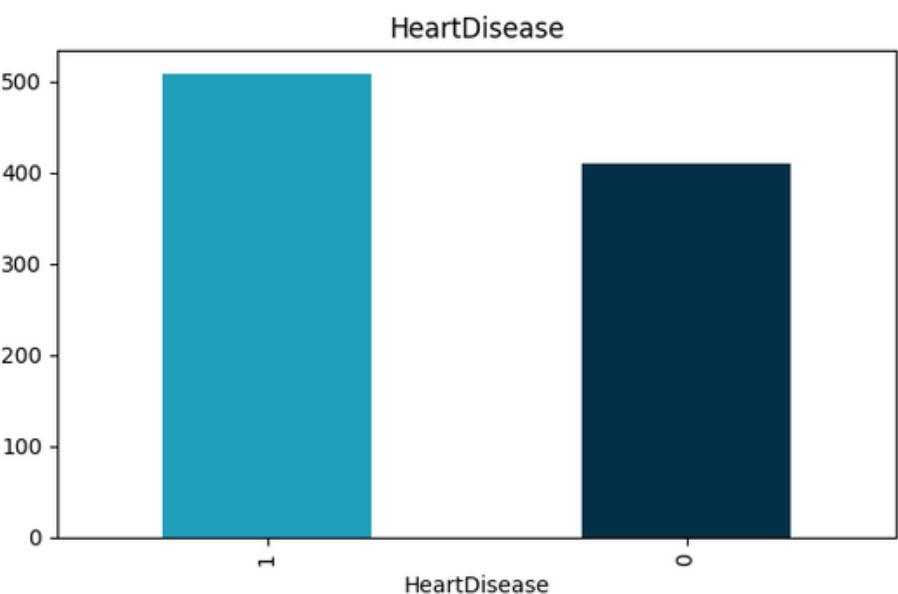
# pre- processam ento



## Tratamento de Outliers (variável RestingBP)

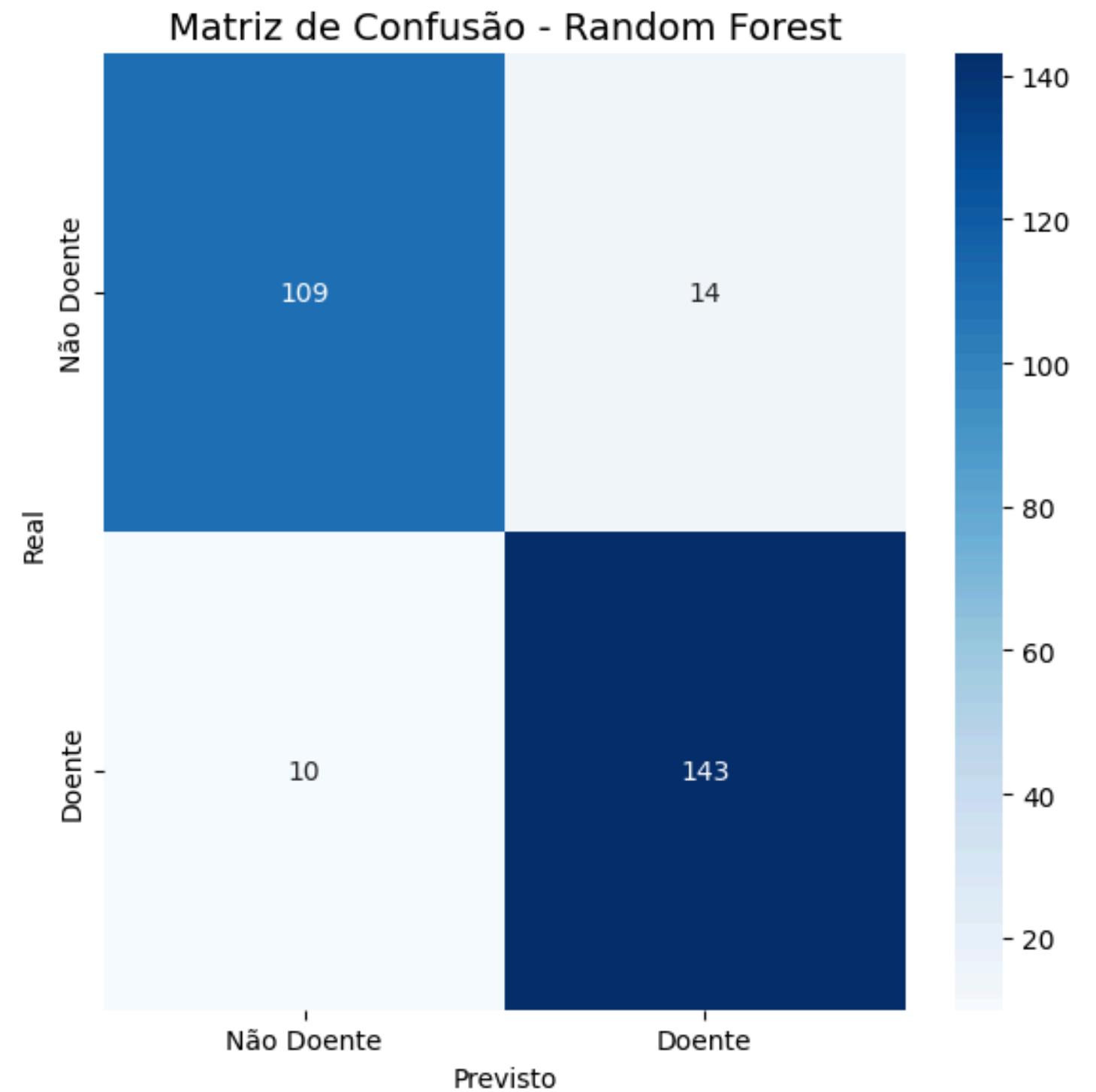
Substituir outliers por  
valores como a mediana

## Balanceamento da variável ALVO (Target)

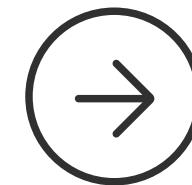


## Verificar os valores únicos em cada variável

Age : 50	RestingECG : 3
Sex : 2	MaxHR : 119
ChestPainType : 4	ExerciseAngina : 2
RestingBP : 67	Oldpeak : 53
Cholesterol : 222	ST_Slope : 3
FastingBS : 2	HeartDisease : 2



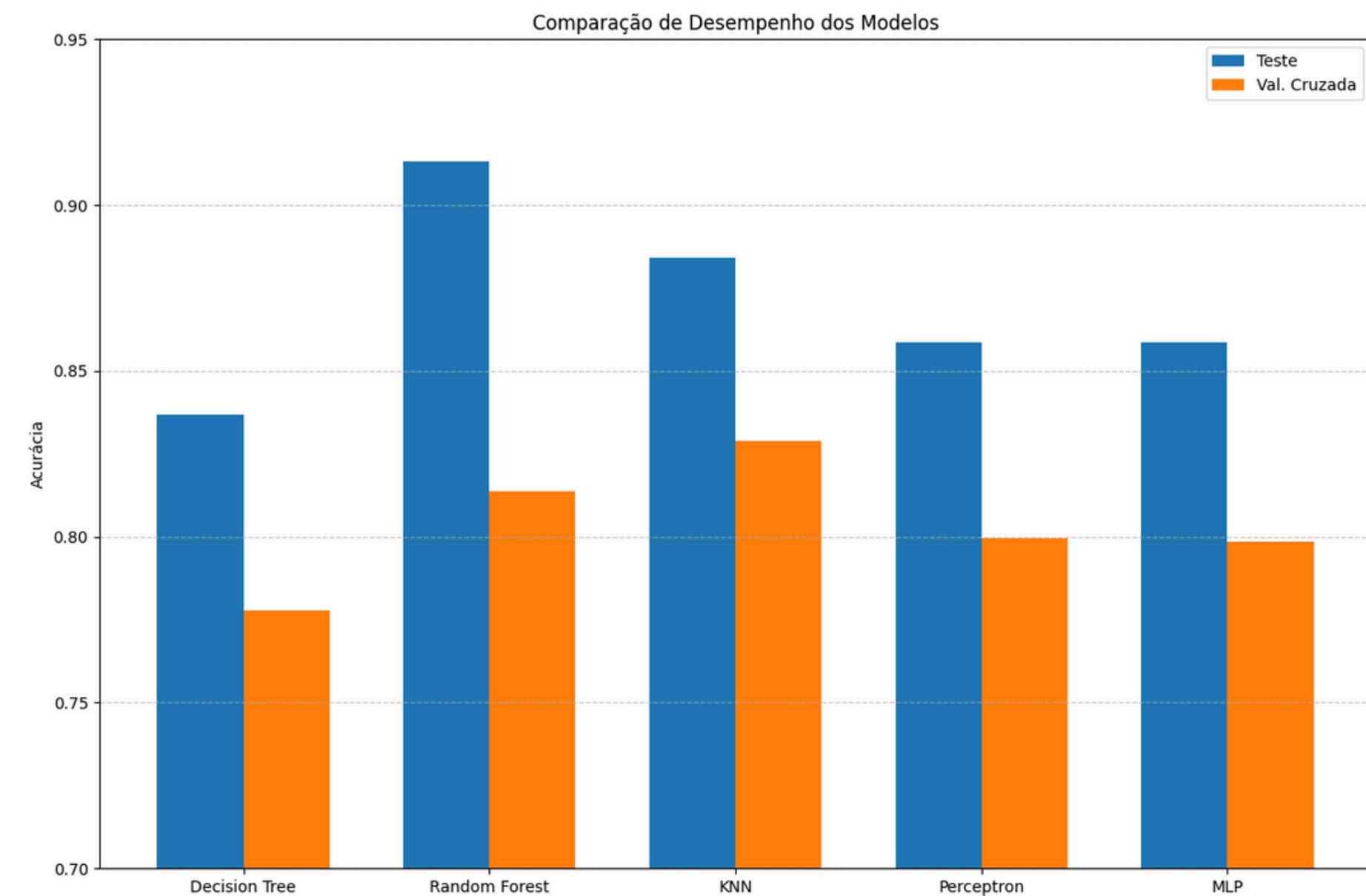
Acurácia: 0.9130  
Precisão: 0.9131  
Recall: 0.9130  
F1-Score: 0.9129  
Validação Cruzada (5-fold):  $0.8136 \pm 0.0457$



Random Forest

Observa-se que todos os modelos obtiveram uma acurácia de teste superior à média da validação cruzada, o que pode indicar overfitting leve — especialmente no caso do Random Forest, cuja diferença entre teste e validação cruzada é a maior.

Já o KNN apresentou desempenho mais equilibrado entre os dois cenários, sugerindo maior generalização. O gráfico permite comparar rapidamente a robustez e a estabilidade de cada modelo ao longo do treinamento.

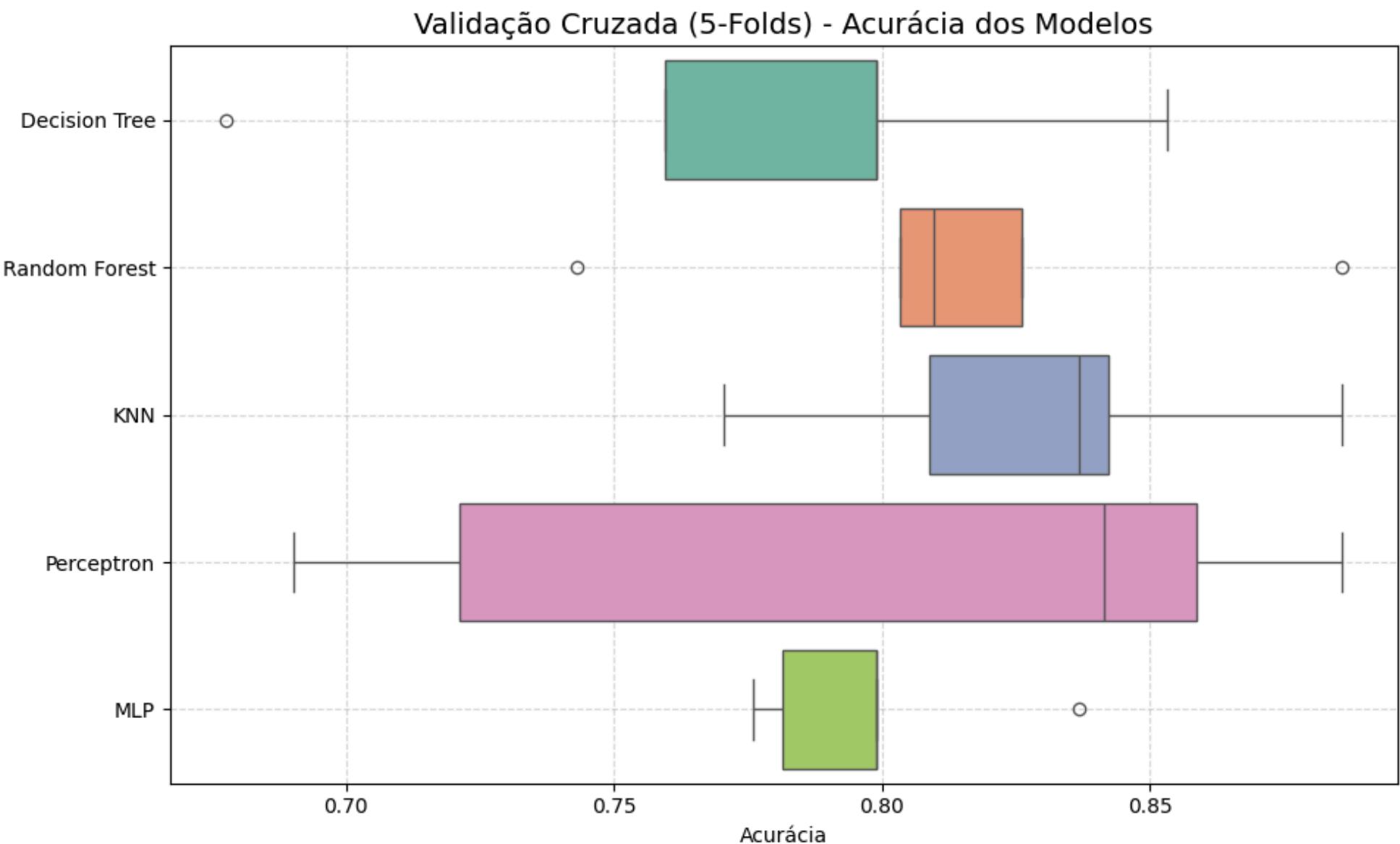


# Acurácia em Teste vs. Validação Cruzada

O gráfico de vela (boxplot) exibe a variação da acurácia dos cinco modelos de classificação após validação cruzada com 5 folds.

Cada barra representa a distribuição dos resultados por modelo, indicando a mediana (linha central), os quartis (caixa), a amplitude (linhas horizontais) e possíveis outliers (pontos fora das extremidades).

Observa-se que o Perceptron apresentou a maior variação de desempenho, indicando instabilidade nos diferentes folds, enquanto o Random Forest teve desempenho mais consistente e relativamente alto. Modelos como KNN e Decision Tree mostraram variação moderada, e o MLP teve acurácia mais baixa e estável.



# Variação da acurácia



# OBRIGADO

Jefferson Antônio T. Silva

João Paulo Sandes Brito



Este trabalho demonstra que, com os dados corretos e modelagem adequada, o aprendizado de maquina pode transformar insights em ações precisas.