



# WINE CASE

## [ExE]mplarias Solutions

**Anderson Souza**  
*Cientista de Dados*

# WINE CASE

- **Produtos/serviços**
  - Vinhos nacionais e internacionais
- **Problema**
  - Prever o potencial de guarda

# 1. ANÁLISE EXPLORATÓRIA

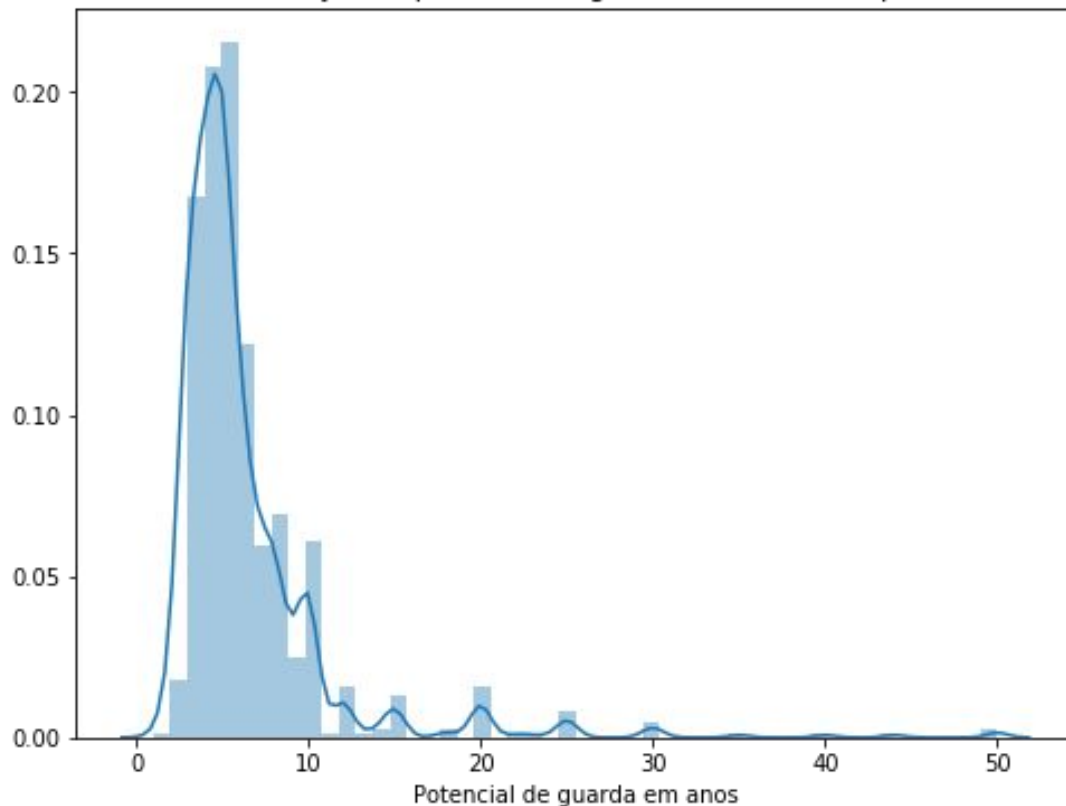
- **Ficha técnica dos vinhos**
- Harmonização
- Comentários do sommelier
- Notas de avaliação
- **Preços dos vinhos**

## 2. FICHA TÉCNICA

- Tipos de vinhos e uvas
- País, região e vinícola
- Teor alcoólico e temperatura
- Amadurecimento
- Classificação
- Visual, aroma e gosto

### 3. POTENCIAL DE GUARDA

Distribuição do potencial de guarda: amostra completa



- Valores variam de 1 - 50 anos
- 87% dos vinhos tem potencial de guarda menor que 10 anos

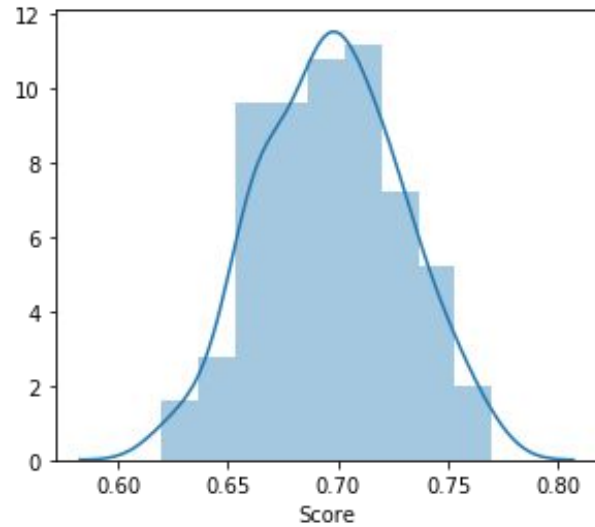
### 3. POTENCIAL DE GUARDA

- **Potencial de guarda em bins**

Bins	Nº de vinhos
(0, 5]	510
(5, 10]	282
(10, 20]	43
(20, 50]	18

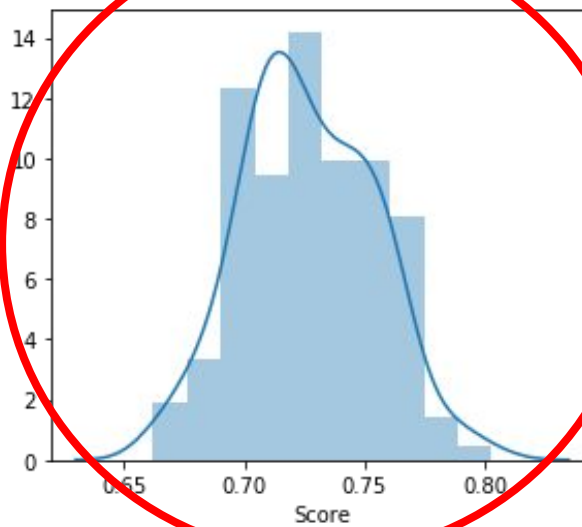
## 4. MODELAGEM

Decision Tree



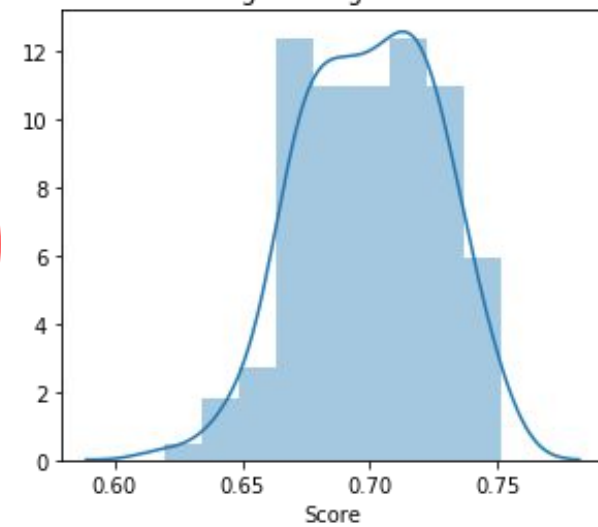
**$0.70 \pm 0.03$**

Random Forest



**$0.73 \pm 0.03$**

Logistic Regression



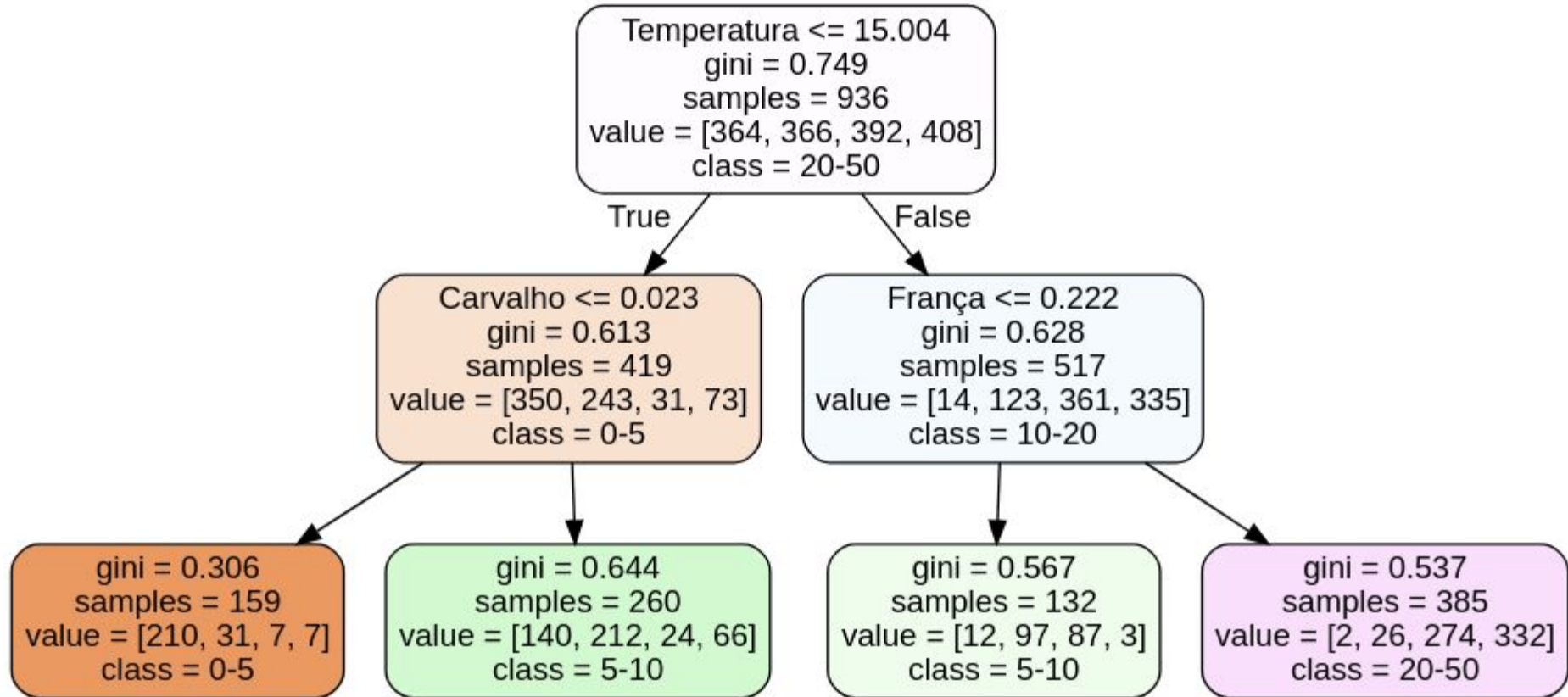
**$0.70 \pm 0.03$**

## 4. RANDOM FOREST

- Machine Learning
- Uma floresta de árvores de decisões
- Cada árvore de decisão
  - Perguntas e respostas
  - Separação das classes



## 4. ÁRVORE DE DECISÃO

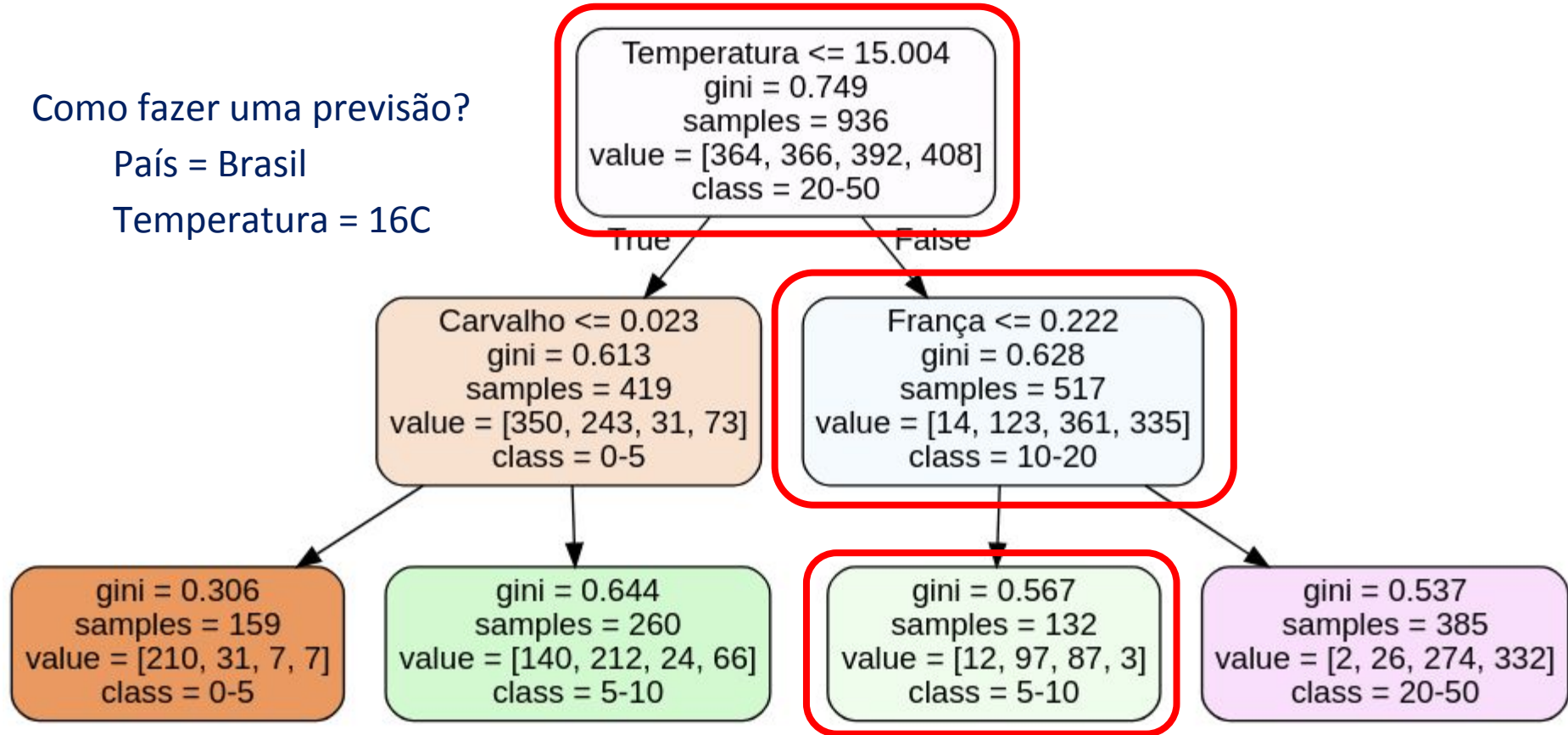


# 4. ÁRVORE DE DECISÃO

Como fazer uma previsão?

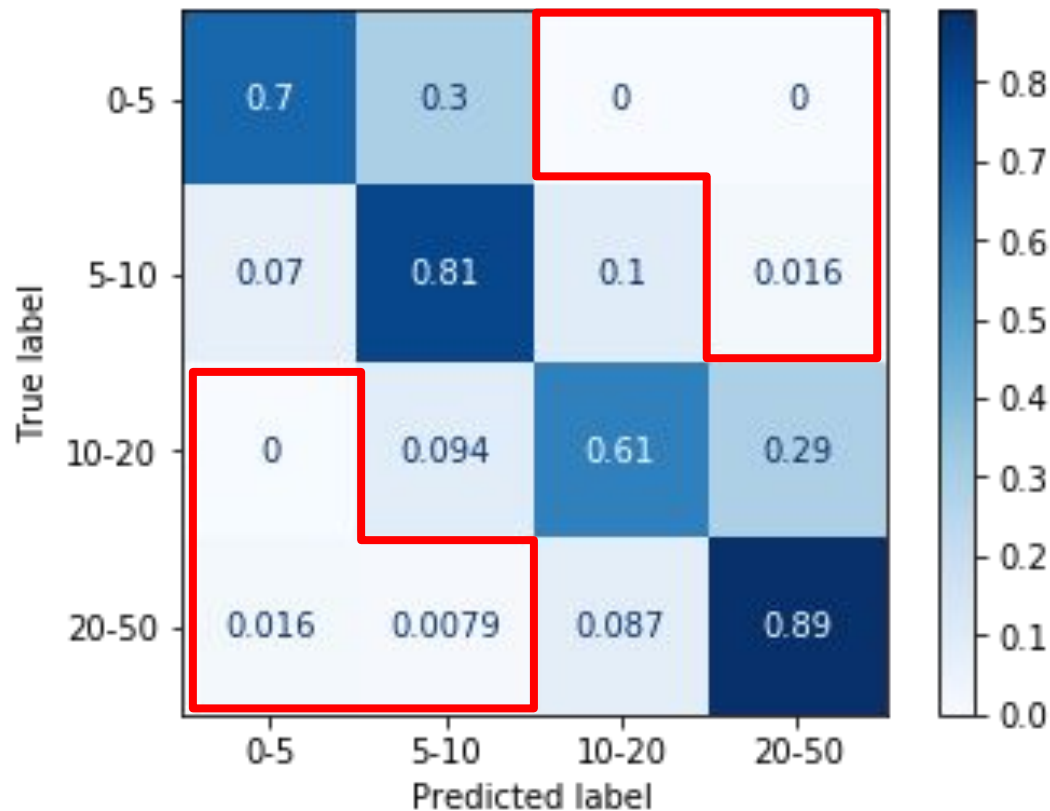
País = Brasil

Temperatura = 16C

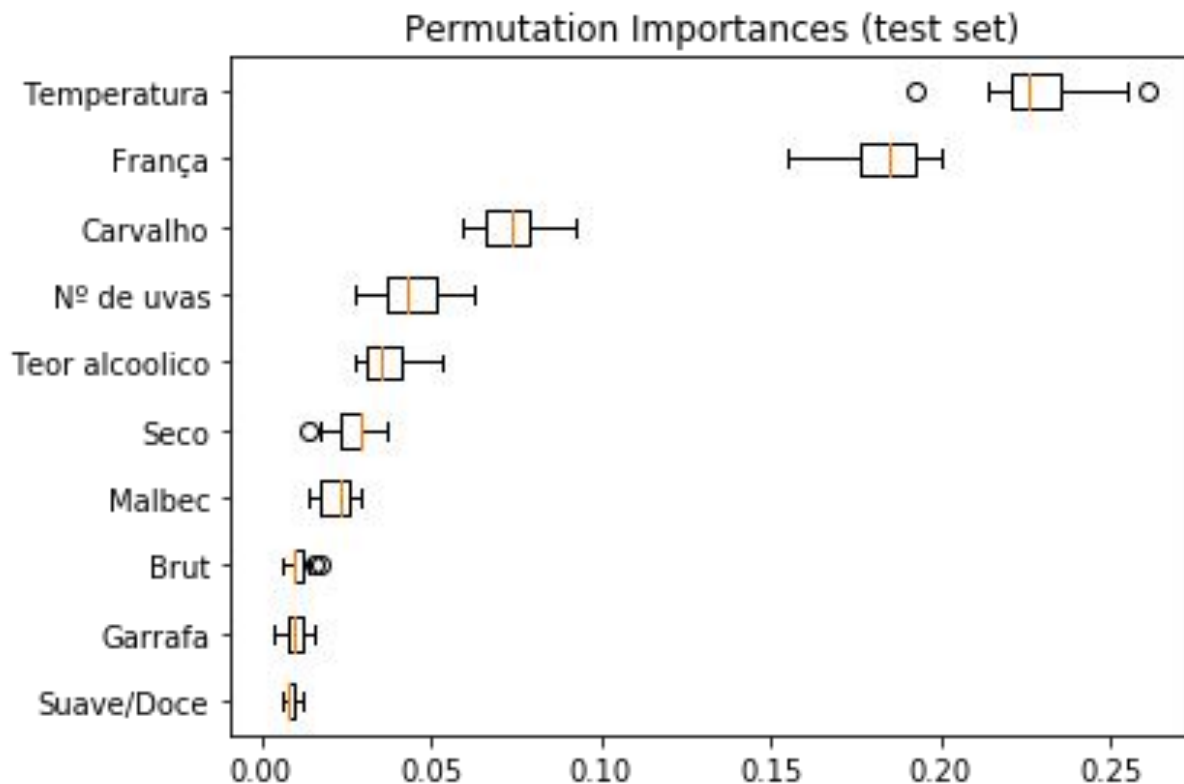


## 5. DESEMPENHO DO MODELO

- Matriz de confusão
  - acertos e erros



# 5. AS CARACTERÍSTICAS MAIS IMPORTANTES

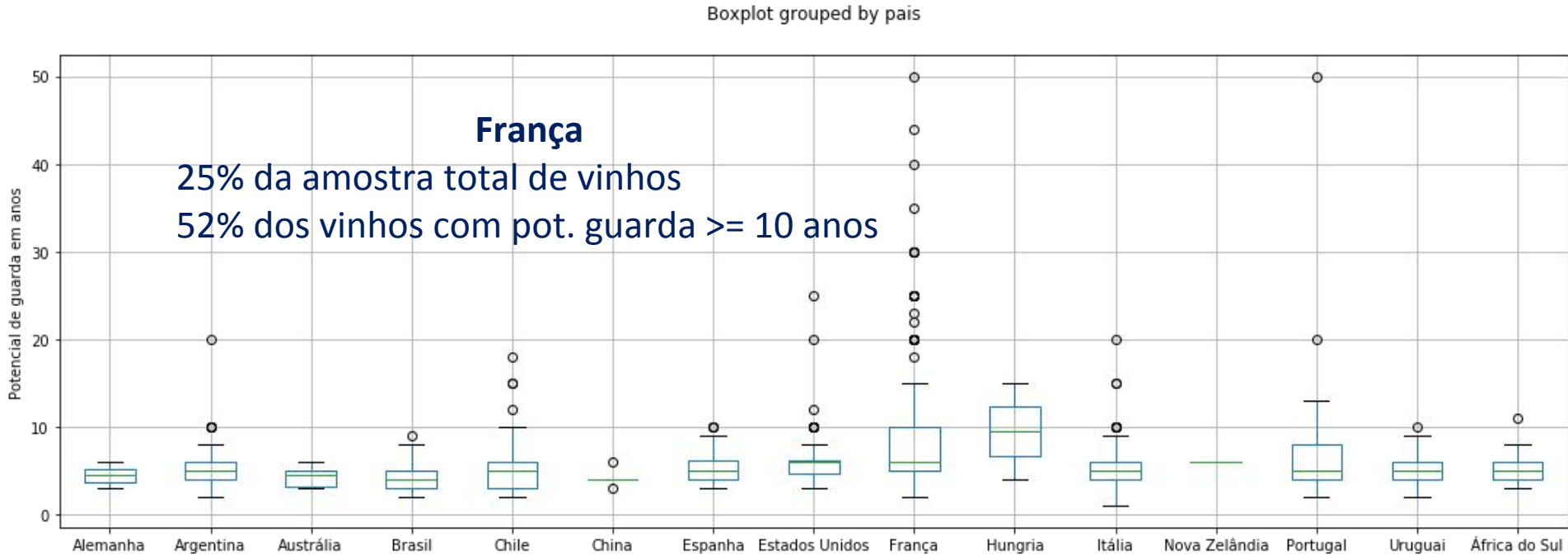


## 5. TEMPERATURA

- **Potencial de guarda em bins**

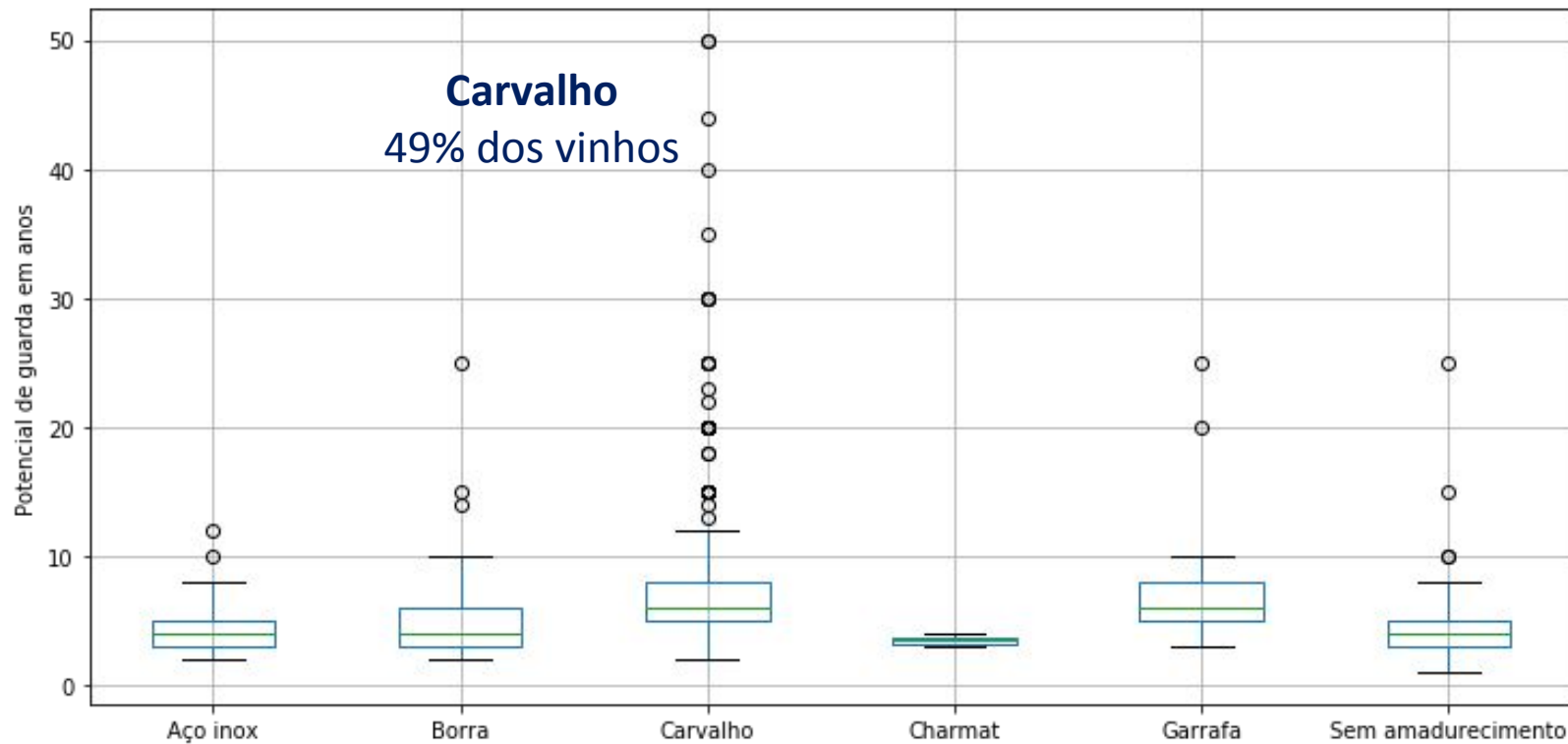
Bins	Temperatura (°C)
(0, 5]	5 - 16
(5, 10]	7 - 18
(10, 20]	8 - 16
(20, 50]	9 - 17

# 5. PAÍS, REGIÃO E VINÍCOLA



# 5. LOCAL DE AMADURECIMENTO

Boxplot grouped by Local



## 6. AVALIAÇÃO DO MODELO

- Modelos predictivos:
  - Machine Learning vs Estatístico

(Random Forest)

(Média por país e tipo)

Acurácia: % de acertos	
Machine Learning	Estatístico
72.6%	54.0%



## 6. AVALIAÇÃO DO MODELO

- Erro na previsão
  - Prejuízo financeiro

## 6. AVALIAÇÃO DO MODELO

- Erro na previsão
  - Prejuízo financeiro

Prejuízo (preço p/ não-parceiros)	
Machine Learning	Estatístico
- US\$ 109 902,59	- US\$ 168 129,28



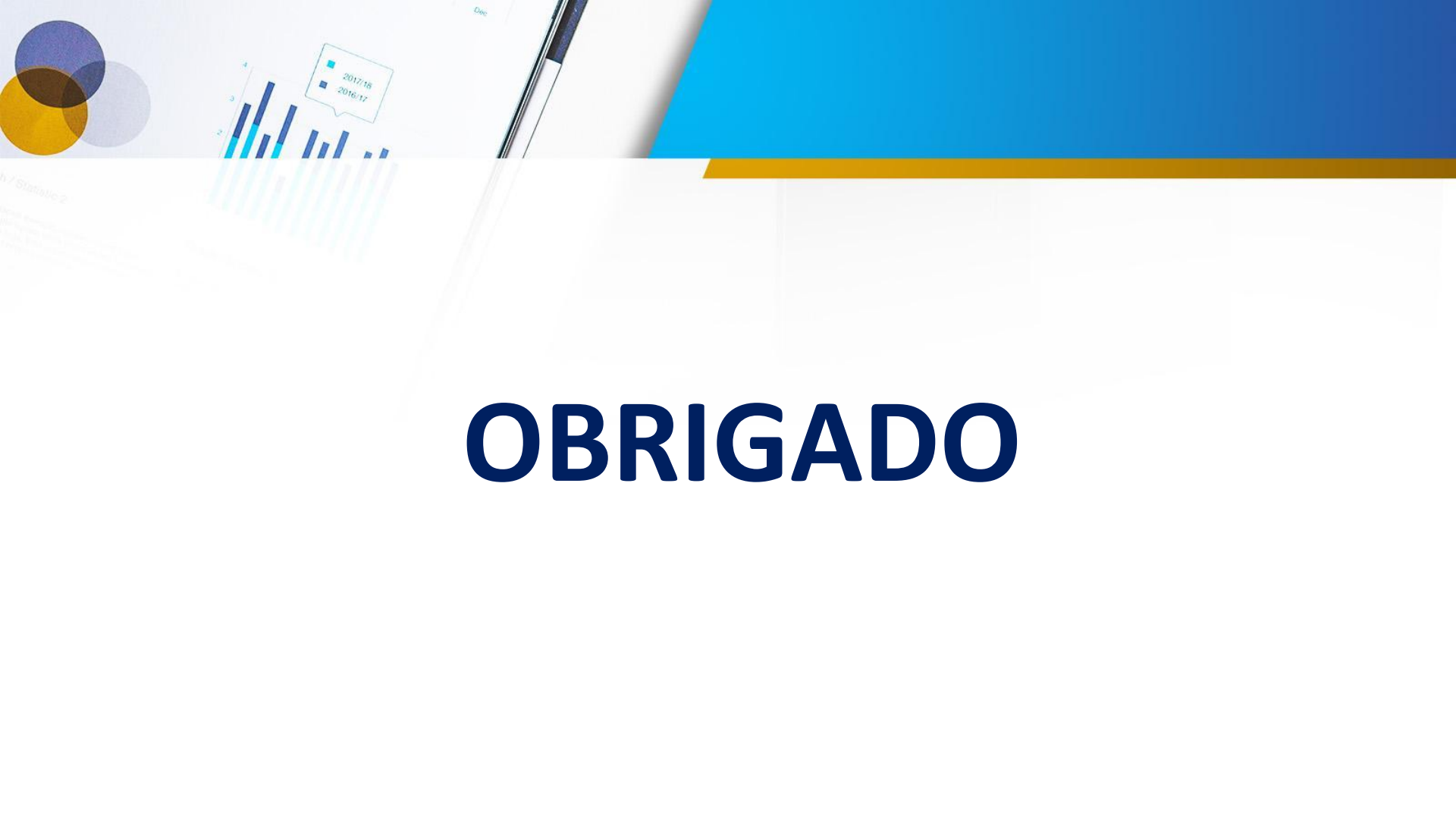
**ECONOMIA DE  
35%**

## 7. CONCLUSÕES

- **O que este modelo predictivo entrega?**
  - **Múltiplas faixas de potencial de guarda**
  - **Personalizado de acordo com o cliente**
  - **Desempenho melhor que os clássicos**
  - **Reduz o prejuízo financeiro**

## 8. PRÓXIMOS PASSOS

- **Como melhorar as previsões?**
  - **Composição química**
    - concentração de açúcares, ácidos (pH) e taninos
  - **Fatores externos**
    - umidade do ar



**OBRIGADO**