

# Estimativa e Análise da cor da Amstel

Jessica Cardoso



# Introdução

- ★ Há vários fatores que caracterizam uma cerveja, sendo um deles a sua coloração. Cervejaria de Ponta Grossa identificou um problema de assertividade de cor durante o processo de produção da cerveja.
- ★ Escalas:
  - EBC (European Brewing Convention – Convenção de Cervejeiros da Europa)
  - SRM (Standard Reference Method – Método de Referência Padrão)
- ★ O principal responsável pela coloração da cerveja é o malte



# Introdução

- ★ **Definição:** Cervejaria de Ponta Grossa identificou um problema de assertividade de cor durante o processo de produção da cerveja.
- ★ **Objetivo:** Melhorar o índice de coloração da cerveja
- ★ **Como:** Utilizando abordagens de análise de dados e machine learning para identificar e estimar a cor do mosto após o seu resfriamento

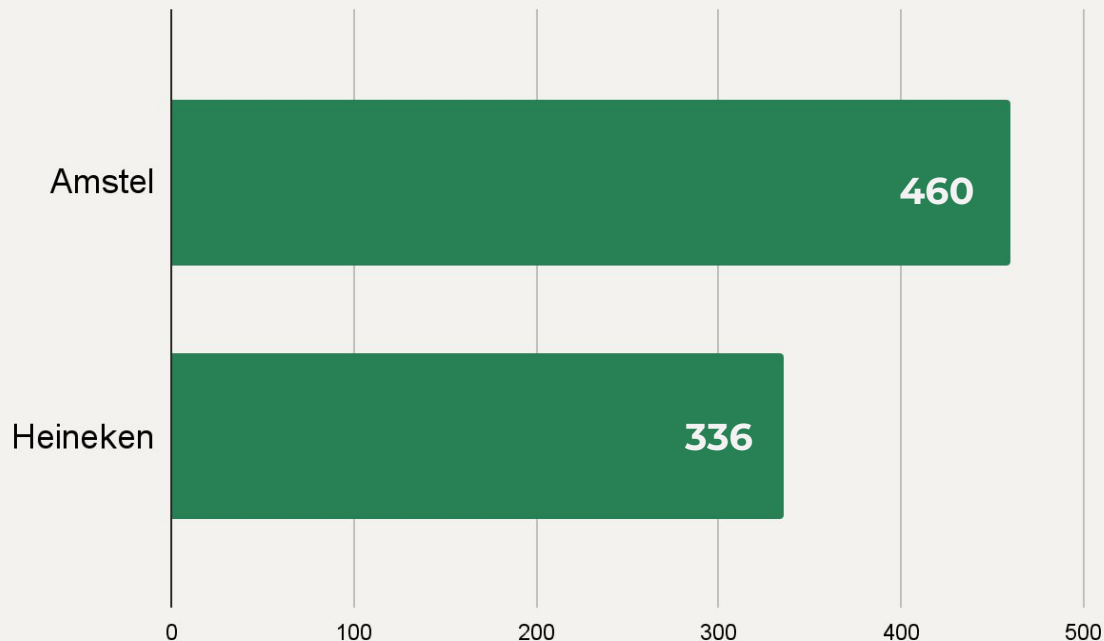


# Questões do estudo

- ★ Conseguimos estimar a coloração com as informações disponibilizadas?
- ★ O quão bem as técnicas respondem ao problema de coloração do mosto?
- ★ Como as variáveis contribuem para essa estimativa?

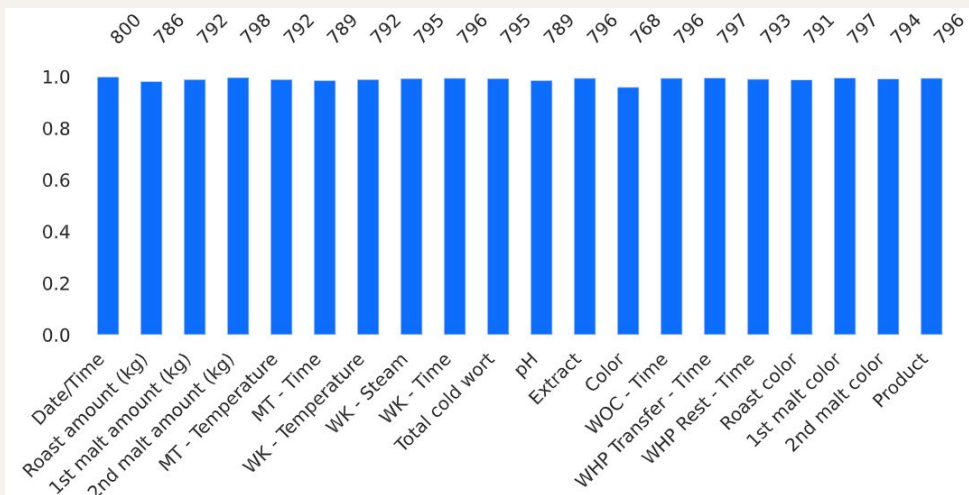


# Conjunto de Dados



# Conjunto de Dados

- ★ 801 observações
- ★ 20 variáveis
- ★ 168 células com dados faltantes



# Conjunto de Dados

Variável	Faltantes	Negativos	Extremos (+)
Roast Amount	15	18	0
1st malt amount	9 + 1 (zerado)	0	0
2nd malt amount (kg)	3 + 12 (zerado)	1	0
MT - Temperature	9	0	1
MT - Time	12	2	1?
WK - Temperature	9	1	2
WK - Steam	6	3	1
WK - Time	5	1	1?

# Conjunto de Dados

Variável	Faltantes	Negativos	Extremos (+)
Total cold wort	6	1	2
pH	12	2	0
Extract	5	1	1
<b>Color</b>	<b>33</b>	<b>12</b>	0
WOC - Time	5	0	1
WHP Transfer - Time	4	0	2?
WHP Rest - Time	8	0	1
1st malt color	4	1	0

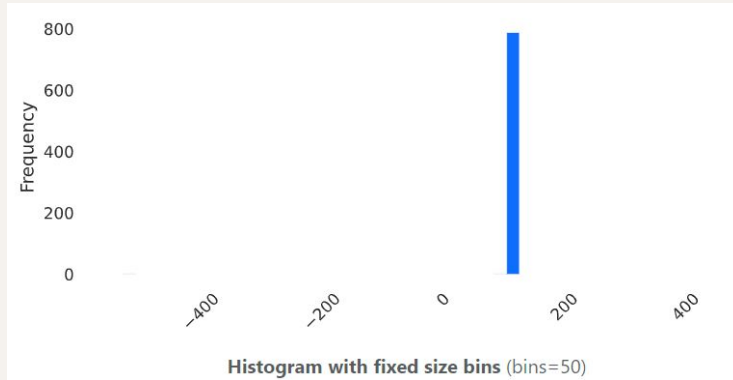


# Conjunto de Dados

Variável	Faltantes	Negativos	Extremos (+)
2nd malt color	7	3	0
Product	5	---	---
Roast color	10	1	1
Date/Time	1	---	---
Job ID	0	--	--

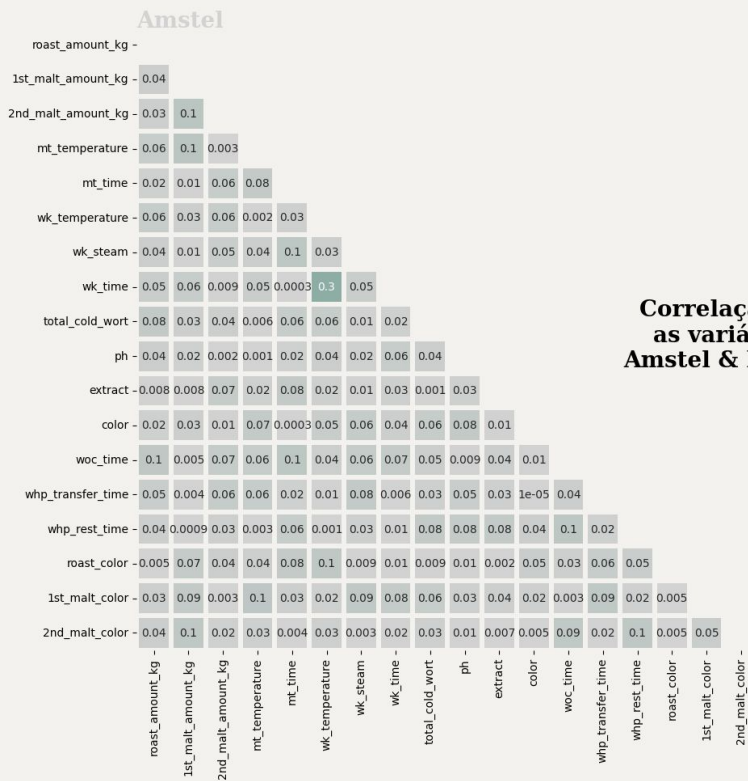
# Conjunto de Dados

WK Temperature



Top-10 valores

Value	Count	Frequency (%)
525.8830336	1	0.1%
524.7513928	1	0.1%
106.8762728	1	0.1%
106.6875467	1	0.1%
106.6657065	1	0.1%
106.5323315	1	0.1%
106.527875	1	0.1%
106.5231026	1	0.1%
106.4960784	1	0.1%
106.4536312	1	0.1%



## Correlação entre as variáveis da Amstel & Heineken

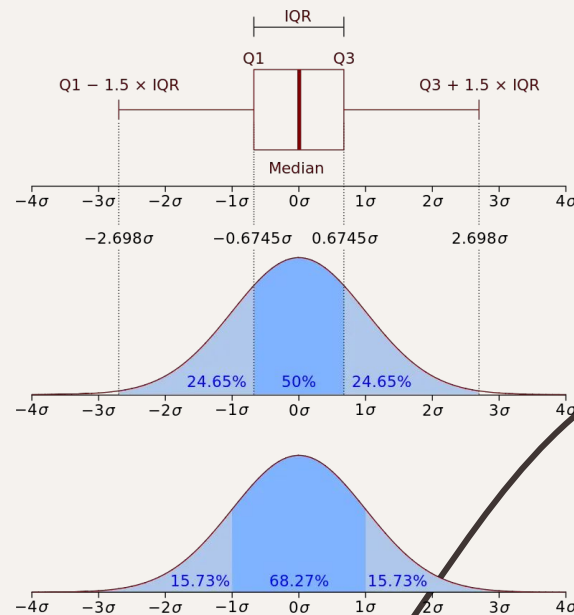
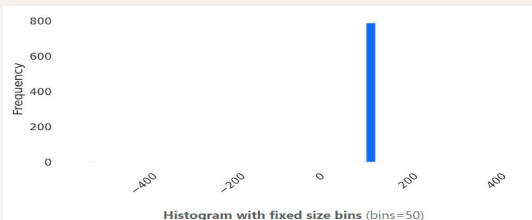


A temperatura do cozedor de mosto e o tempo que o lote fica no cozedor de mosto tem uma correlação de 0.3 para a Amstel

# Estratégias para tratamento dos dados

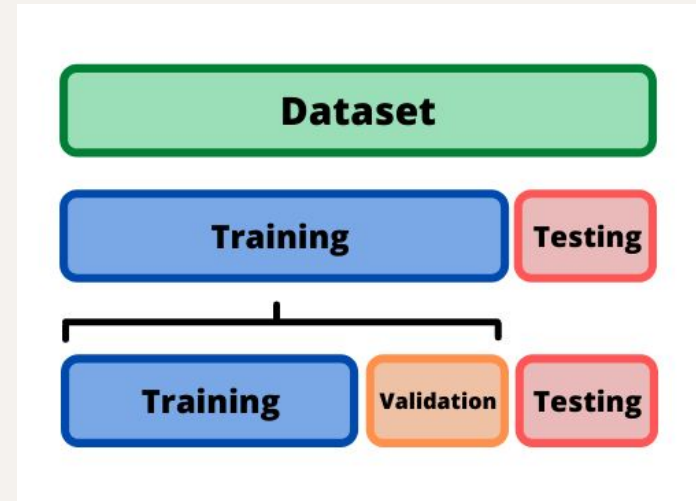
- ❖ Usar o método de distância interquartil, com valores 0,05 e 0,95 para identificar outliers.
- ❖ Substituir os negativos por zero
- ❖ Limitar os valores usando IQR
- ❖ Substituir esses valores por NaNs e preencher usando abordagens para lidar com valores ausentes.

WK Temperature



# Metodologia

- ★ 20% da amostra da Amstel como teste final
- ★ Divisão posterior de treino, teste e validação para experimentar estratégias



# Modelos

- ★ Extra Trees
- ★ Random Forest
- ★ Decision Tree
- ★ Regressão Linear
- ★ KNN
- ★ Outros

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$$

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}}$$

# Processar saída do modelo/dados

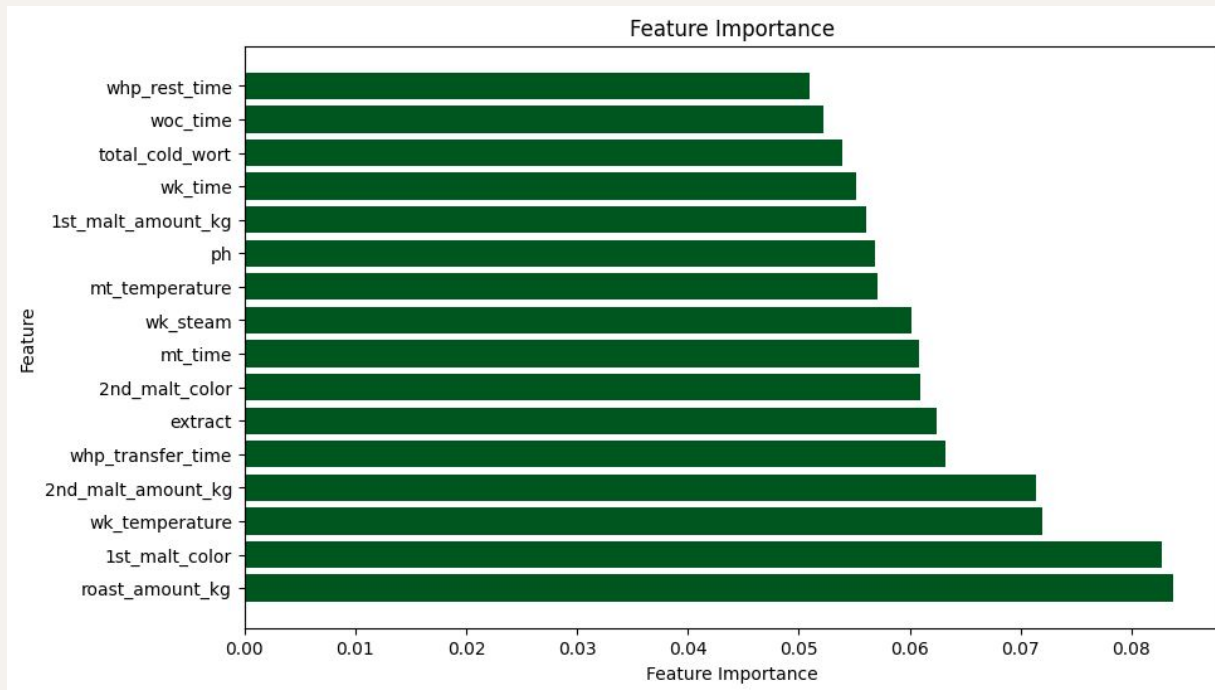
- ★ Uso o Optuna para buscar o normalizador utilizado
  - Standard Scaler
  - MinMax Scaler
  - Robust Scaler
- ★ Para a saída:
  - Além dos padrões do Sklearn
  - Logaritmo
- ★ Preenchimento de dados:
  - KNN
  - Iterativo
  - Simples (média, mediana, moda)

# Resultados para Amstel

	MSE	RMSE	R2	MAE	MAPE
Extra Trees (Ambas)	<b>68.04</b>	8.25	0.09	5.93	0.52
Extra Trees (Amstel)	68.30	8.30	0.08	6.16	0.49
Média	77.93	8.83	-0.04	6.45	0.48



# Importância das Features



MSE: 74.07  
RMSE: 8.61  
 $R^2$ : 0.01  
MAE: 6.42  
MAPE: 0.54

# Outras Tentativas

- ★ Foram experimentados diversas abordagens, dentre as quais:
- ★ Construção de *features*
  - Novas *features* são construídas a partir das existentes por meio de operações de soma subtração, etc, entre pares de features.
  - Seleção de features filtra as mais relevante dentre as criadas
- ★ Discretização de *features*
  - As *features* de valores numéricas são discretizadas em bins, se tornando variáveis categóricas

Nenhuma das abordagens obteve melhores resultados



# Thanks

Do you have any questions?

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

Please keep this slide for attribution