



# Data Wrangling

## Ciência de Dados

**Centro Universitário Senac**

**Prof. Celso Crivelaro**

[celso.vcrivelaro@sp.senac.br](mailto:celso.vcrivelaro@sp.senac.br)

# Tipos de Tratamentos

**Limpeza de dados**



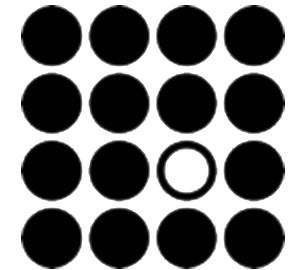
**Valores Faltantes**



**Remoção de Atributos**



**Tratamento de Outliers**



# Tipos de Tratamentos

**Limpeza de dados**



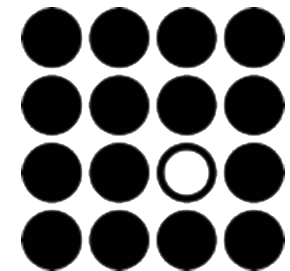
**Valores Faltantes**



**Remoção de Atributos**



**Tratamento de Outliers**



# Limpeza de dados

Remoção de dados inúteis

Valores fora do intervalo de análise

Dados incompletos ou dados com erro

Dados duplicados ou dados iguais representados de forma diferente

# Tipos de Tratamentos

**Limpeza de dados**



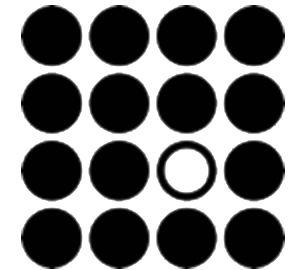
**Valores Faltantes**



**Remoção de Atributos**



**Tratamento de Outliers**



# Como proceder com valores faltantes?

Definir um valor padrão pré-estabelecido

Substituir dados ausente pela média ou moda do campo

Substituir por um valor aleatório baseado em uma distribuição

Substituir o dado baseado em outras característica do registro

# Tipos de Tratamentos

**Limpeza de dados**



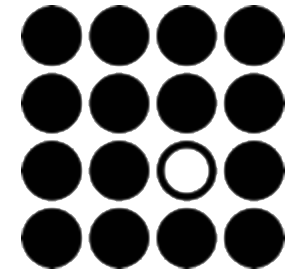
**Valores Faltantes**



**Remoção de Atributos**



**Tratamento de Outliers**



# Remover dados

Nem todos os dados precisam ser usados para modelagem ou gráficos

Dados que não influencia modelos ocupam espaço e podem gerar ruídos. Assim, podemos remover atributos

Usar técnica para descobrir dimensionalidade como *Principal Component Analysis - PCA*





# Tipos de Tratamentos

**Limpeza de dados**



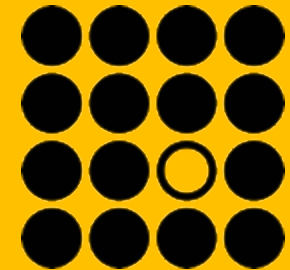
**Valores Faltantes**



**Remoção de Atributos**

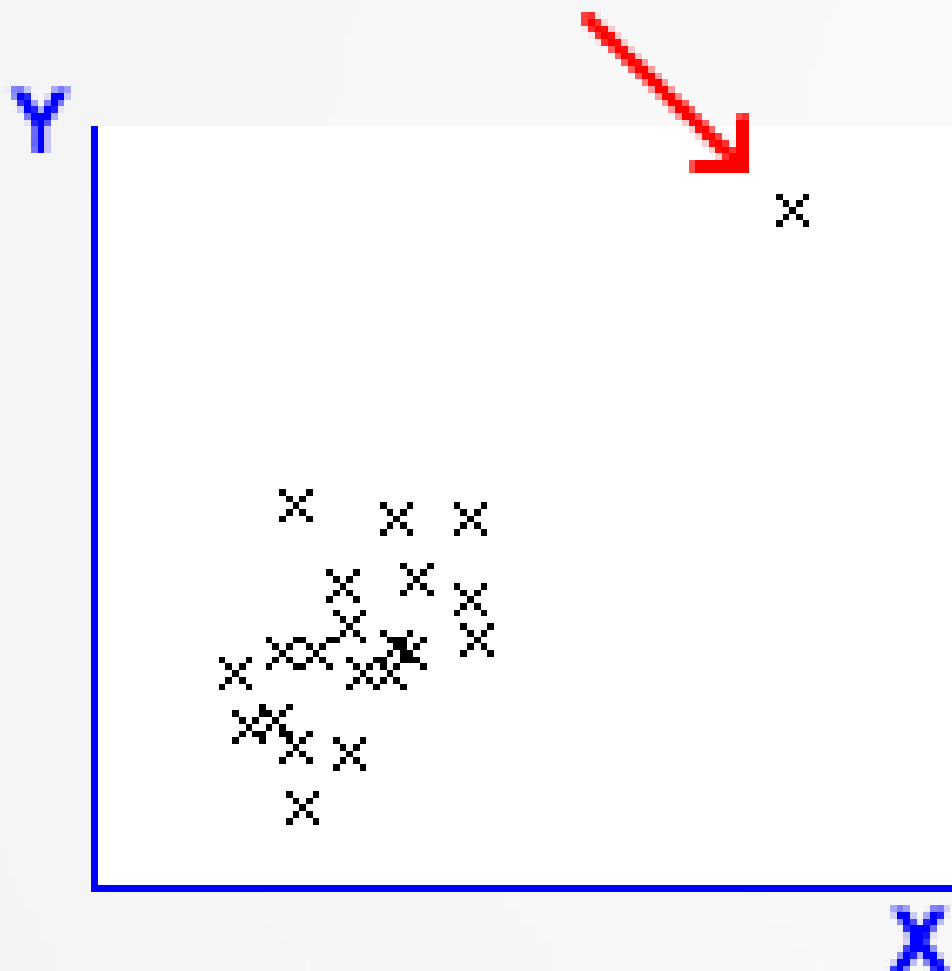


**Tratamento de Outliers**

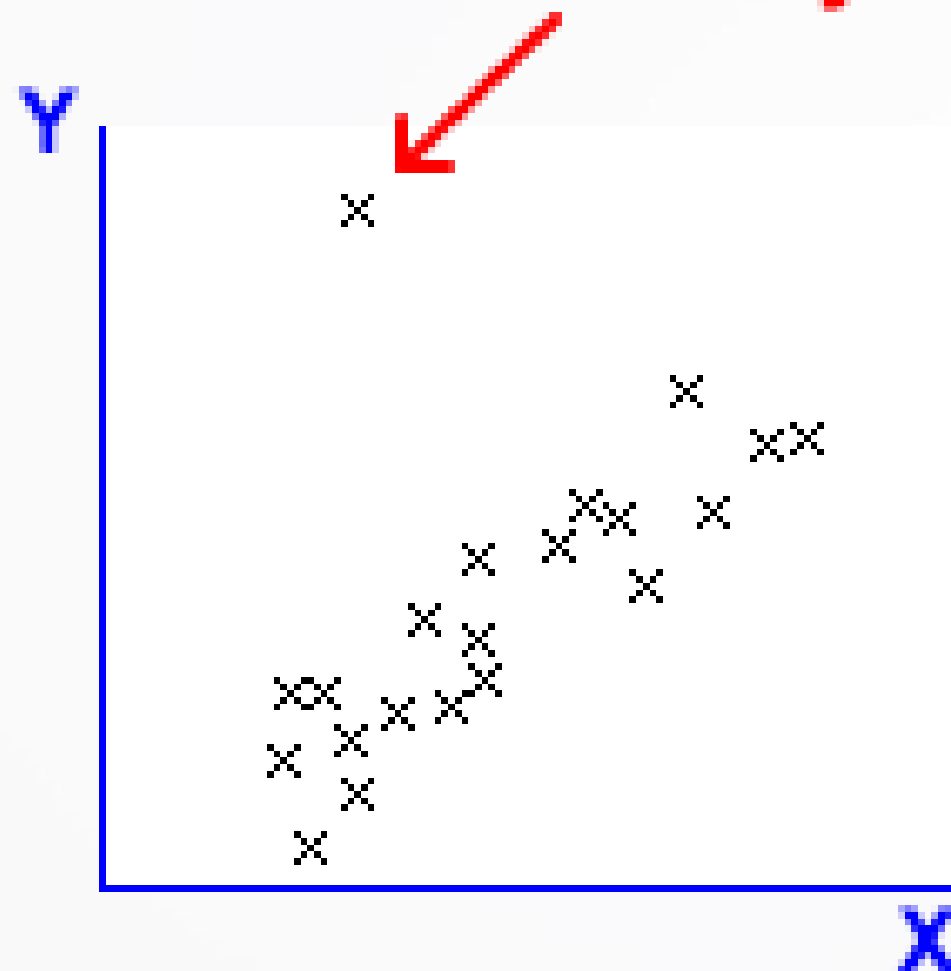


# O que é um outlier?

Outlier in both X and Y



Outlier in Y only



# Quando acontece um outlier?

Dados que contêm muita importância: Ex. Livro do Harry Potter

Supersensibilidade de sensores. Ex.: Temperatura

Erros

Acontecimentos novos ou foram com comum



Outliers são bons ou ruins?

# Quando acontece um outlier?

Caso bons

Detecção de fraudes

Casos novos

Casos ruins

Procura de padrões

Regressão linear e extrapolação



# Muito Obrigado!

**Centro Universitário Senac**

**Prof. Celso Crivelaro**  
[celso.vcrivelaro@sp.senac.br](mailto:celso.vcrivelaro@sp.senac.br)