



escola  
britânica de  
artes criativas  
& tecnologia

## **Profissão: Cientista de Dados**

Análise de agrupamentos

# Análise de agrupamentos

Brincando com a ideia







# Redução de dimensionalidade

Em PCA fizemos redução de dimensionalidade nas colunas

Com Clusters fazemos redução de dimensionalidade nas linhas

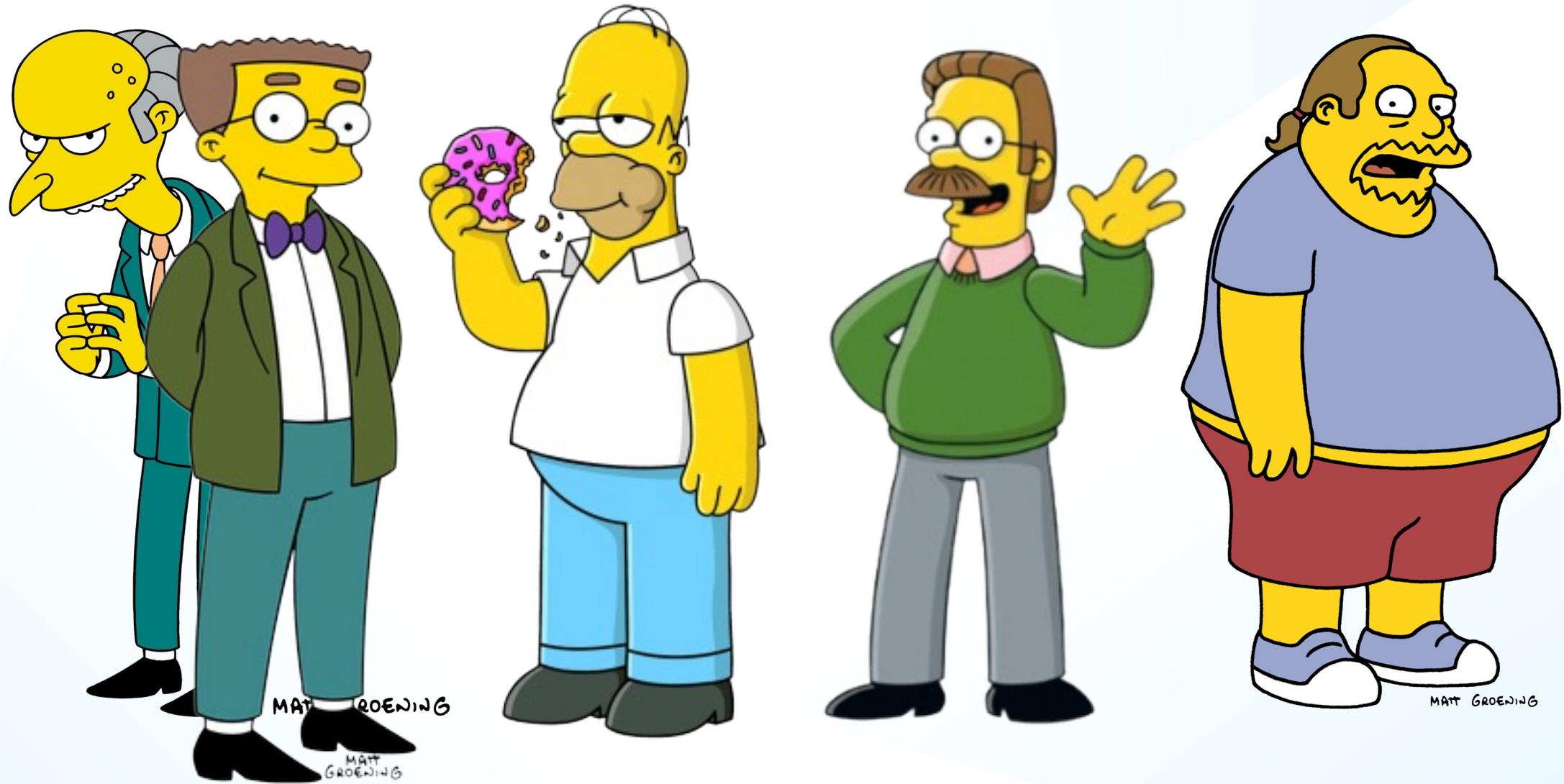
Ao invés de avaliar milhares de indivíduos, avaliamos grupos semelhantes

A close-up photograph of a white mesh produce bag filled with red and yellow apples. The bag is tied at the top with green elastic bands. It sits on a silver digital scale. In the background, shelves filled with various fruits like oranges and lemons are visible in a supermarket setting.

# Supermercado

Ex1: Como são esses clients?

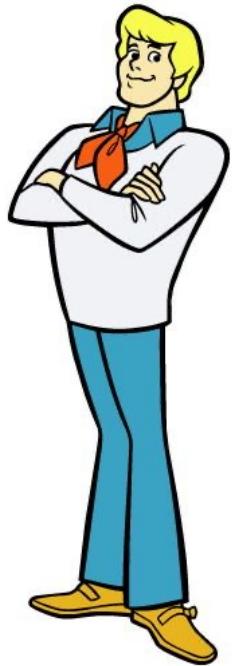
# Redução de dimensionalidade





# Streaming

Ex 2: Como são esses clientes?



# Análise de agrupamentos

*Classificação*

# Classificação dos algoritmos



## Supervisionados

- Regressão
- GLM
- GLMM
- Support vector machines
- Naive Bayes
- K-nearest neighbors
- Redes Neurais
- Decision Trees



## Não supervisionados

- K-Means
- Métodos hierárquicos
- Mistura Gaussiana
- DBScan
- Mini-Batch-K-Means

# Classificação dos algoritmos



## Métodos Machinelârnicos

- Árvores de decisão
- Bagging
- Boosting
- K-NN
- Redes Neurais
- Support Vector Machines



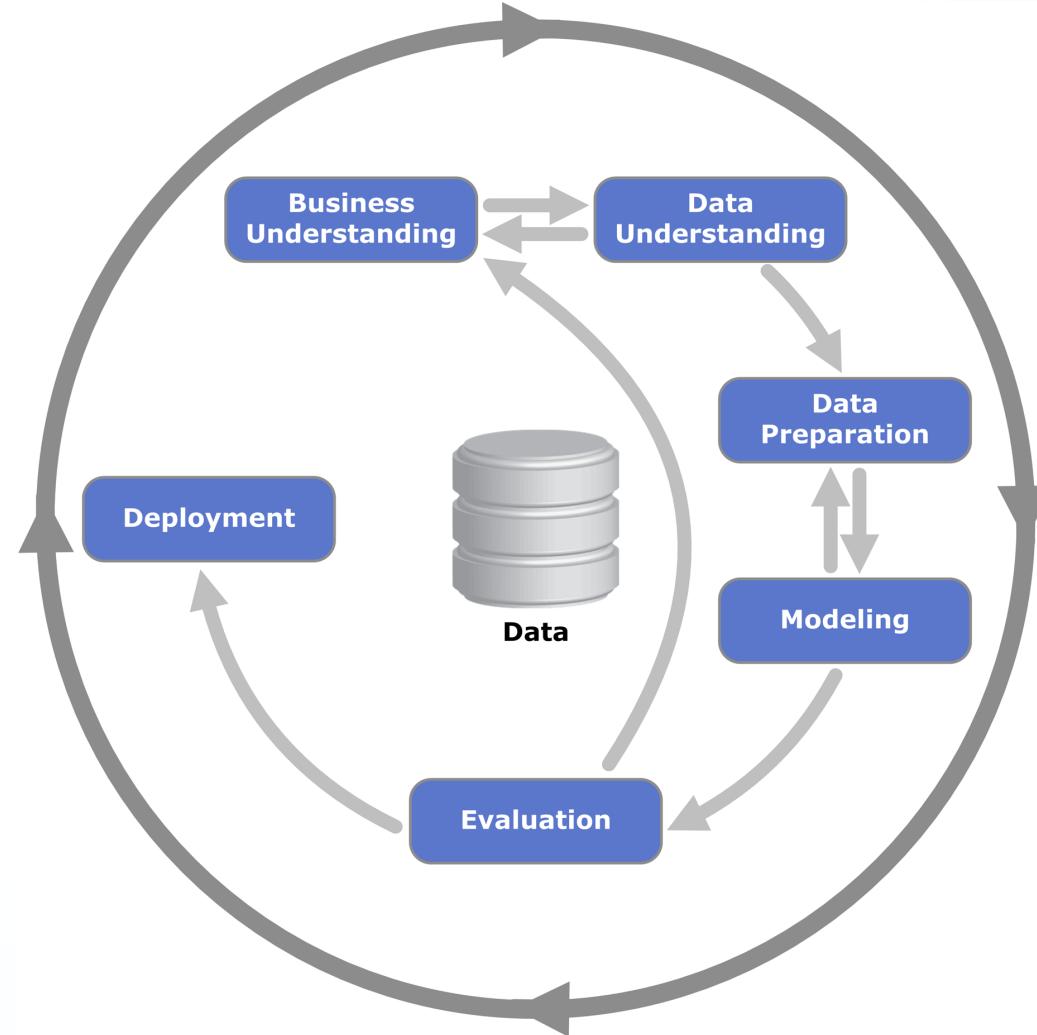
## Métodos Estatísticos

- Regressão
- GLM
- GLMM
- ANOVA

Estamos aqui!

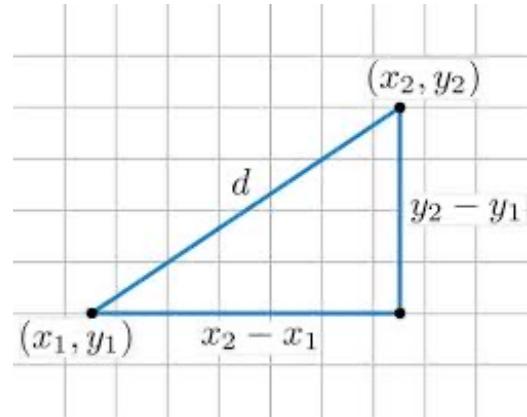
# CRISP-DM

---

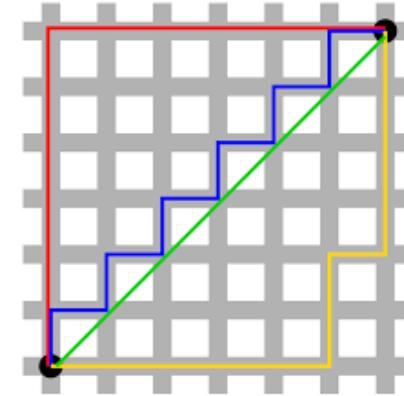


Fonte: <https://www.the-modeling-agency.com/crisp-dm.pdf>

# Definições de distância

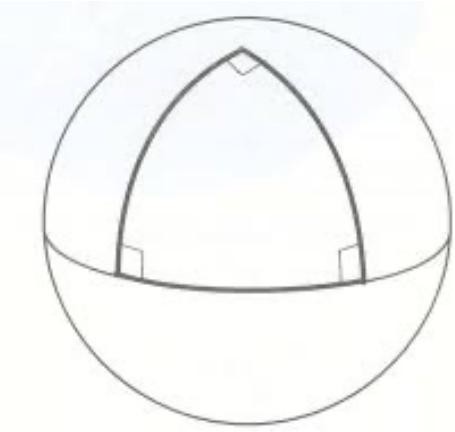


Distância Euclidiana



Distância de Manhattan

Distâncias Não Euclidianas



Distância de Mahalanobis

