



# **Identificando sobrepço em compras públicas**

**Desenvolvimento de um banco de preços  
de licitações**



# Os Tribunais de Contas têm muitos dados

- Despesa
- Notas fiscais
- Licitações
- Relatórios de auditoria
- Denúncias de ouvidoria
- Votos, decisões
- Etc.

**Priorização:** banco de preços – identificar sobrepreço em processos licitatórios



# Mas como fazer?

## **Não há**

padronização  
de categorias  
de produtos

## **Não sabemos**

quantas  
categorias de  
produtos  
existem

## **O único campo**

minimamente  
confiável é a  
descrição do  
produto

## **Nem as propriedades**

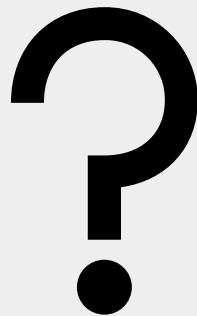
dentro de cada  
categoria, como  
apresentação,  
tipo, quantidade,  
etc

## **A descrição**

pode ser escrita  
de forma livre,  
inclusive com  
erros

## **Enfim**

é um problema  
complexo



# Machine learning:

## Aprendizado não supervisionado



# Alguns resultados:

## Pneu 275 80 225:

pneu 275 80 225 liso radial profundidade minim...

pneu 275 80 225 borrachudo radial profundidade...

pneu 275 80 r 225

pneu 275 80 225 borrachudo

pneu 275 80 225 liso

pneu 275 80 r 225 liso

pneu borrachudo 275 80 r 225

pneu 275 80 225 borrachudo

pneu 275 80 225 borrachudo

pneu 275 80 r 225 radial misto

pneu borrachudo 275 80 r 225

pneu 275 80 r 225 borrachudo

pneu 275 80 r 225 liso

pneu 275 80 r 225 borrachudo

pneu 275 80 r 225 liso

pneu 275 80 aro 225 radial liso diante

pneu 275 80 r 225 liso radial misto

pneu 275 80 225

pneu 275 80 r 225 borrachudo radial misto

## Óleo de soja 900 ml:

óleo de soja refinado 900 ml

óleo de soja refinado embalagem plastica trans...

óleo de soja garrafa pet 900 ml

óleo de soja 900 ml embalagem plastica caracte...

óleo de soja

óleo de soja lt 900 ml

óleo de soja refinado 900 ml

óleo de soja 900 ml

óleo de soja 900 ml

óleo de soja 900 ml

óleo de soja 900 ml

óleo de soja 900 ml

óleo de soja emb 900 ml

óleo de soja refinado - 900 ml

óleo de soja

óleo de soja

óleo de soja refinado - embalagem 900 ml

óleo de soja embalagem 900 ml

óleo de soja embalagem 900 ml

## Fita cirúrgica microporosa 25 cm 45 m :

fita cirurgica microporosa hipoalergica rolo medindo 25 45 m

fita cirurgica microporosa 25 cmx 45 m

fita micropore hipoalergenica 25 45 cm

fita cirurgica microporosa 25 cm 45 m

fita cirurgica microporosa 125 mm 45 m fita cirurgica micropo

fita hipoalergica microporosa 25 cm 45 m fita hipoalergica mic  
nao tecido a base de fibras de viscose resina acrilica e massa a  
tamanho 25 cm 45 m

fita cirurgica microporosa hipoalergica 25 cm 45 m

fita cirurgica microporosa 45 m 25 cm

fita micropore 25 cm

fita cirurgica microporosa hipo 12 cmx 45 m

fita micropore branco 25 cm 45 m

## Variações do mesmo produto:

### Azitromicina 500 mg:

azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg comprimido
azitromicina 500 mg
azitromicina 500 mg comprimido
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg sulcado
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg
azitromicina 500 mg blister fracionavel
azitromicina 500 mg

**Azitromicina 600 mg, frasco 15 ml,  
suspensão oral:**

azitromicina 600 mg sus
azitromicina 600 mg po para suspensao oral emb...
azitromicina po susp oral 40 mg ml
azitromicina po para suspensao oral 200 5 ml
azitromicina 600 mg frasco com 60 ml copo
azitromicina po p suspensao oral 600 mg diluen...
azitromicina susp 600 mg ml
azitromicina susp 600 mg 15 ml
azitromicina po para suspensao oral 600 mg
azitromicina 600 mg
azitromicina 600 mg
azitromicina po para susp oral 600 mg diluente...
azitromicina po p susp oral 40 mg ml
azitromicina susp 600 mg 15 ml
azitromicina 600 mg po oral
azitromicina 600 mg susp
azitromicina po susp oral
azitromicina po para suspensao oral 40 mg ml f...
lincomicina 600 mg
azitromicina po para suspensao oral 600 mg 15 ml

**Azitromicina 40 mg, frasco 15 ml,  
suspensão oral:**

azitromicina 40 mg ml po para suspensao oral
/azitromicina 40 mg ml po para suspensao oral f...
azitromicina 40 mg ml
azitromicina 40 mg ml
azitromicina 40 mg ml
azitromicina suspensao oral 40 mg ml
azitromicina 40 mg ml
azitromicina 40 mg ml po susp
azitromicina 40 mg ml suspensao
azitromicina 40 mg ml po para suspensao oral
azitromicina 40 mg ml po para suspensao oral
azitromicina 40 mg ml po para suspensao oral 1...
azitromicina 40 mg ml po para suspensao oral
azitromicina 40 mg ml po p sus 15 ml fr plas d...
azitromicina 40 mg ml
azitromicina 40 mg ml
azitromicina 40 mg ml
azitromicina 40 mg ml po para suspensao oral 1...
azitromicina 40 mg ml po para suspensao oral

# NLP – processamento de linguagem natural

1

## Limpeza

- Transforma para minúsculas;
- Retira pontuação;
- Mantém stopwords.

2

## Treino FastText

- Treino rede neural;
- Gera vetores de 300 dimensões;
- Palavras similares tem vetores próximos.

3

## Vetores das sentenças

- Sentence embeddings com fórmula própria;
- Média das palavras da sentença;
- Primeiras palavras são mais importantes;
- Números com peso um pouco maior.

# Redução de dimensionalidade

## Clustering

com alta dimensionalidade não funciona direito

## Para captar

a relação entre as palavras precisamos número considerável de dimensões

## UMAP

é o algoritmo que conseguiu realizar um bom trabalho pra reduzir a dimensionalidade:

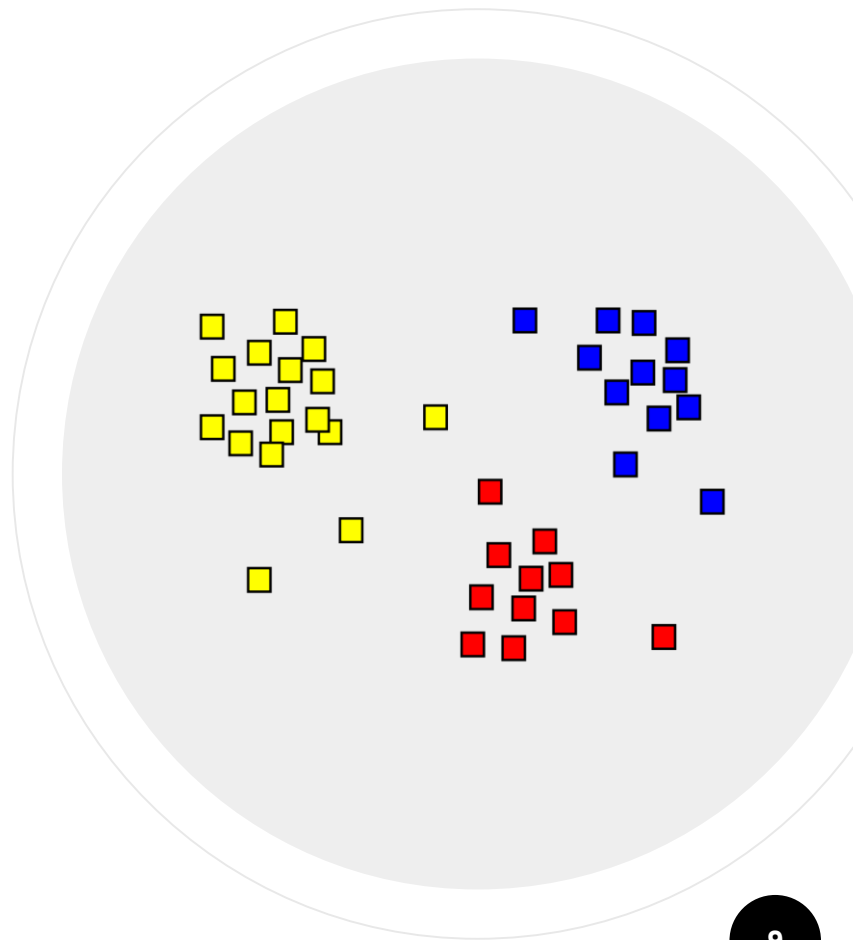
- Métrica de distância de cosseno;
- Não linear.





# Agrupamento – clustering com HDBSCAN

- Não linear
- Por densidade
- Não importa “formato” do grupo
- Parâmetro – tamanho mínimo do grupo: 30 itens



# Limpeza pós clusterização

1

## Exclui outliers

- Outliers do HDBSCAN
- Grupos com itens de apenas um órgão (escrita específica).

2

## Une grupos parecidos

- Ajuste fino;
- Sentence embeddings com peso maior para os números;

3

## Encaixa quem ficou de fora

- Por cosine similarity.

# Separação de grupos

## **Gasolina:**

- Gasolina
- Gasolina comum
- Gasolina aditivada

## **Amendoim:**

- Amendoim
- Amendoim 500g
- Amendoim descascado
- Amendoim sem casca
- Amendoim tipo 1
- Amendoim cru 1 kg



The background is black. A large white circle is centered on the slide. To the left of the white circle, there are two overlapping circles: a larger grey one and a smaller white one. To the right of the white circle, there are several concentric white circles of varying sizes.

# **Até agora somente texto**

Onde entram:

- preço
- unidades de medida?

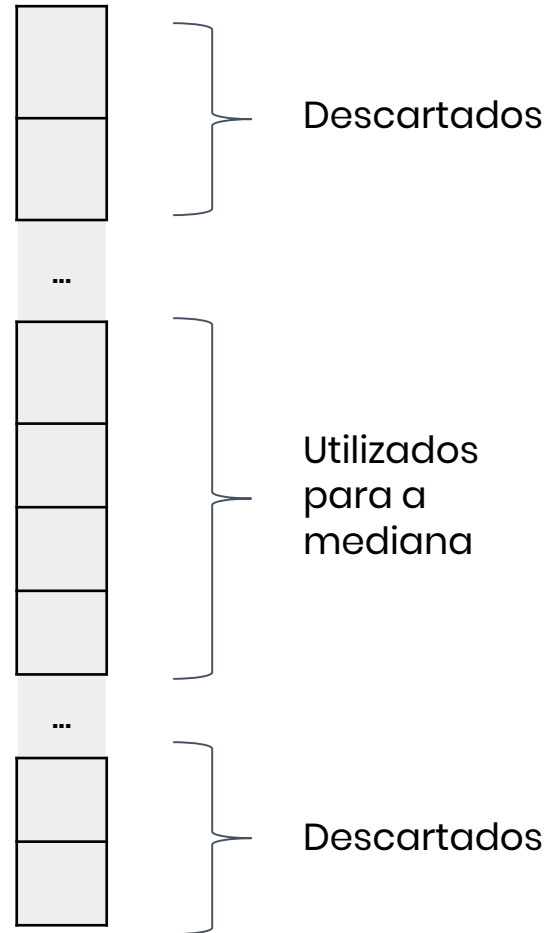
# TCE-RS possui sistema de licitações

- Órgãos públicos informam licitações
- Sete anos de dados
- Informam, dentre outros, descrição, preço homologado e **unidades de medida!**
- Possibilita comparar preços de coisas iguais.



# Mediana saneada

- Retira-se os maiores e os menores valores da série de preço;
- Até que o coeficiente de variação da série fique abaixo do esperado;
- Assim, excluem-se os outliers de preço, mediana fica mais robusta.
- Possibilita comparar preços de coisas iguais.



# Nova consulta:

- Novas licitações que entram pelo sistema – alerta automático para auditores em caso de sobrepreço
- Consulta textual aberta em webapp, para auditores e aberto para a sociedade

## **Limpa e transforma**

o texto em um vetor

## **Vê qual é o grupo**

mais próximo por similaridade de cosseno

## **Calcula**

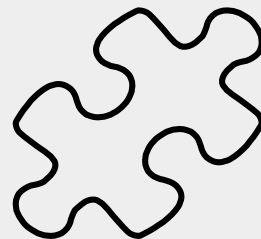
preço mediano com base nos filtros aplicados:

- Quantidade
- Região
- Intervalo de datas



# Resultados práticos preliminares

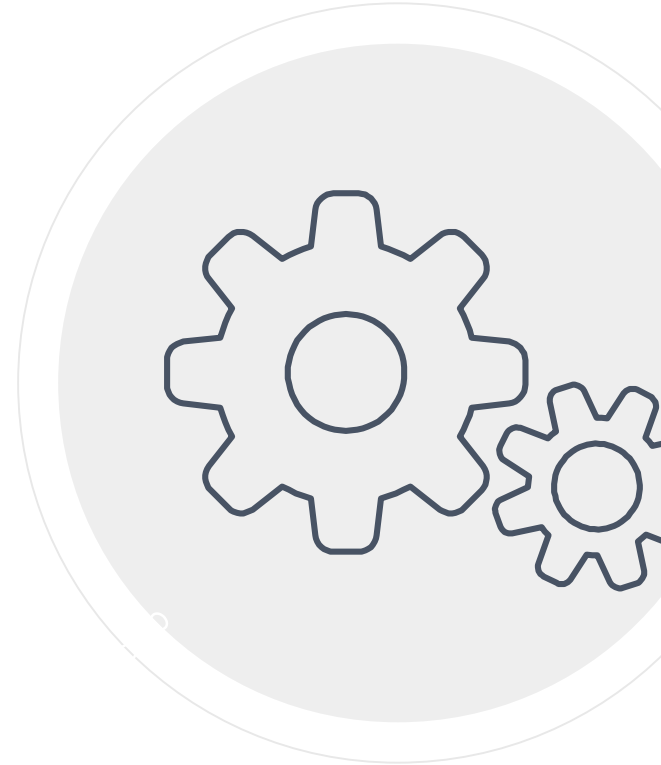
- Medicamentos
- Compras comuns recorrentes
- Quanto mais dados, melhor!





# Estado do projeto

- Implementado: interface/webapp para uso de auditores, em teste para correção de bugs – bom feedback.
- Em desenvolvimento: sistema de alertas automáticos de sobrepreço para novas licitações
- Futuro: portal de acesso externo / sociedade.





# Obrigado!

**Alexandre Gandini, TCE-RS.**

**Questões? Contato:**

**Email:** [agandini@tce.rs.gov.br](mailto:agandini@tce.rs.gov.br) / [agandini@gmail.com](mailto:agandini@gmail.com)

**Linkedin:** [linkedin.com/in/alexgand](https://www.linkedin.com/in/alexgand)