

# Pensamento Computacional e Lógica de Programação

Prof. Dr. Álvaro Campos  
Ferreira

# Web Scraping

Dados obtidos da página estão na linguagem HTML, que divide os dados no que chama de tags.

Para explorar as tags de um site, usa-se a ferramenta de desenvolvimento do navegador, F12 no Firefox.

# Web Scraping

Uma vez identificadas as tags de interesse, filtramos o resultado:

```
soup.findAll('div')  
for tag in soup.findAll('div'):  
    print(tag.text)
```

# Qualidade de dados

Dimensões da qualidade de dados:

- Completude ou integridade
- Conformidade
- Validade
- Acurácia e precisão

# Valores faltantes

Valores faltantes são representados por **nan** em DataFrames e não são a mesma coisa que valores nulos.

- Faltas (completamente) aleatórias
- Faltas não aleatórias

# Outliers

Outliers são valores encontrados fora de um intervalo razoável com relação às outras observações. É um ponto diferente dos demais, não necessariamente um erro. Sua identificação é importante e pode possuir informações interessantes sobre o sistema.

# Limpando dados numéricos e datas

# Limpando dados numéricos e datas

Vamos em geral salvar os resultados em novos DataFrames para garantir a integridade dos dados.

- Função `split()`
- Função `strip()`
- Compreensões de lista



# Bancos de dados e ORM

# Bancos de dados e ORM

Uma das tarefas do programador backend é fazer a conexão e o acesso ao banco de dados.

- Bancos de dados surgiram na década de 70 com SQL sendo adotada universalmente a linguagem para bancos de dados relacionais.

# Bancos de dados e ORM

Em geral, o banco de dados relacional será acessado utilizando SQL, e cada banco de dados diferente tinha uma sintaxe própria para isso.

- ORMs permitem o acesso abstrato a bancos de dados utilizando funções Python



INSTITUTO BRASILEIRO DE ENSINO,  
DESENVOLVIMENTO E PESQUISA