

# **Análise, construção e visualização de dados**

Prof. Dr. Álvaro Campos  
Ferreira

[alvaro.ferreira@idp.edu.br](mailto:alvaro.ferreira@idp.edu.br)

# Vocabulário

**Programar é uma  
atividade criativa**

# Pandas e DataFrames

# Pandas e Dataframes

Pandas vem de 'Panel Data' (dados em painel um termo da econometria para dados multidimensionais estruturados).

- É adequado para trabalhar com dados não numéricos e numéricos
- Adequado para lidar com dados proveniente de webscrapping
- Lida bem com dados incompletos

# Pandas e Dataframes

Vamos usar o Pandas para acessar os nossos dados.

```
import pandas as pd  
file = "meu_arquivo.csv"  
df = pd.read_csv(file)  
print(df)
```

# Análise exploratória

# Análise exploratória

Obtendo um sumário descritivo estatístico de um DataFrame inteiro e de colunas selecionadas.

Usando a função `describe()`

É possível utilizar as funções de estatística descritiva diretamente no DataFrame.



# Análise exploratória

O Pandas possui funções para facilitar a descrição de dados.

```
df.describe()
```

```
df.plot()
```

# Onde encontrar dados?

# Onde encontrar dados?

- 1) Google Data Set Search
- 2) OpenDataSUS
- 3) Kaggle
- 4) Webscrapping



INSTITUTO BRASILEIRO DE ENSINO,  
DESENVOLVIMENTO E PESQUISA