**Laboratório 01 – Dataframe**

# **Objetivos**

* Apresentar e utilizar o pacote **pandas** do Python;
* Aprender a carregar uma base de dados;
* Visualizar os dados;
* Começar uma análise exploratória dos dados;

# **Recursos**

* Google Colab: <https://colab.research.google.com/>
* Pandas para Python: <https://pandas.pydata.org/>
* UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.php>
* Seaborn para Python: <https://seaborn.pydata.org/>

# **Introdução**

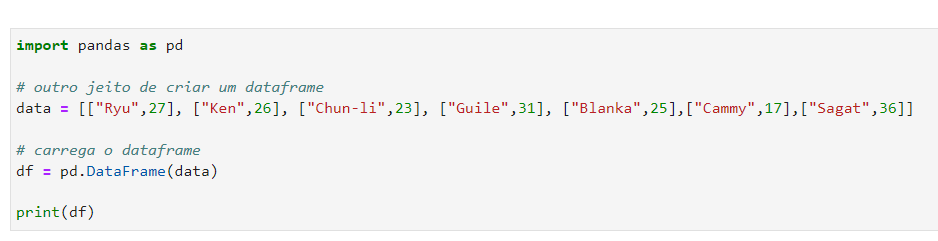
Um dataset nada mais é que um conjunto de dados. Vamos utilizar um pacote do Python chamando **pandas** para trabalhar com tabelas de conjuntos de dados. O nome dessas tabelas no pandas é **dataframe;**

Para mais informações, leia a documentação em <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>;

Primeiro vamos criar um dataframe:



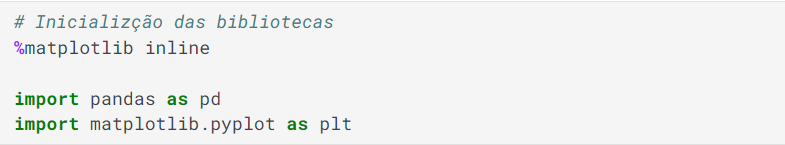
Pergunta: Qual o resultado do print?

Vamos criar esse mesmo dataset, mas de uma forma diferente: 

Pergunta: Qual o resultado do print agora?

LEIA A DOCUMENTAÇÃO EM <https://pandas.pydata.org/docs/index.html#>

Importante: Devido a questões de dependências e métodos depreciados, se faz necessário o uso das diretivas abaixo:

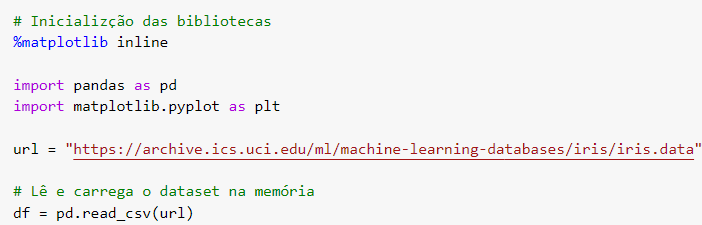


# **Datasets**

Podemos escolher qualquer base de dados disponíveis na internet, ou até mesmo criar nosso próprio dataset.

Vamos simplificar essa etapa e começar analisando uma base pequena e muito famosa chamada ”**íris**” que esta disponíveis em: <https://archive.ics.uci.edu/ml/datasets/Iris>

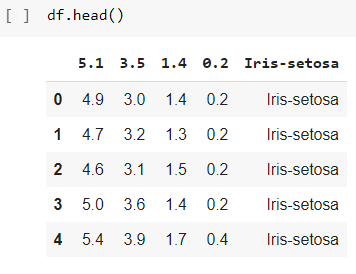
Conheça outros datasets: <https://archive.ics.uci.edu/ml/datasets.php>



# **Conhecendo os Dados**

Essa etapa é muito importante, CONHECER OS DADOS!

Quanto mais você conhece a base de DADOS maior a possibilidade de extrair INFORMAÇÕES úteis para tomada de decisão.



Note que a primeira linha não é com os nomes das colunas ou atributos(variaveis) e sim de dados (valores). Por padrão os dados da primeira linha são importados como atributos.

Vamos adicionar um cabeçario ao nosso dataframe. No repositório oficial é dito que as variaveis são: