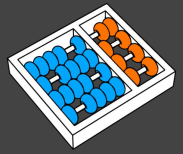
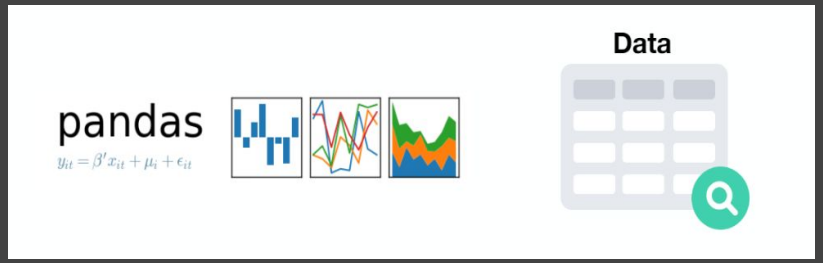


INTRODUÇÃO À PANDAS

Prof. Dr. Julio Cesar dos Reis
Instituto de Computação - Universidade Estadual Campinas

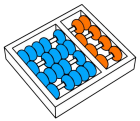
Slides originalmente criados pelo Prof. Allan M. de Souza e modificados pelo Prof. Julio Cesar dos Reis (IC/UNICAMP)

O que é pandas?



PANDAS

https://pandas.pydata.org/

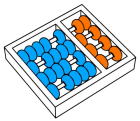


UNICAMP

PANDAS

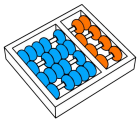
- Biblioteca open source (BSD-licensed), a qual fornece estruturas de **alta-performance** e fácil de usar
- Fornece ferramentas para análise e processamento de dados
- Fornece uma API eficiente para análise de dados em python
- Ferramenta poderosa e flexível para análise de dados

PANDAS FEATURES



UNICAMP

- Acessível para todos
- Free para uso
- Permite modificações (BSD-licensed)
- Ponderosa
- Fácil de usar
- Eficiente (rápida)



UNICAMP

ESTRUTURAS DE DADOS DO PANDAS

Existem duas estruturas de dados principais no pandas

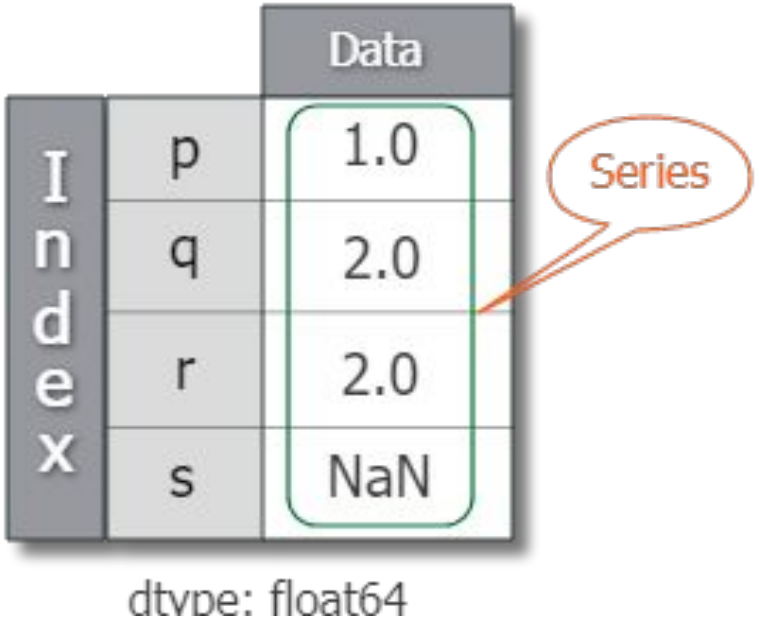
- Series
- DataFrames

PANDAS SERIES

- **Series** é um array de uma dimensão **rotulado** capaz de armazenar qualquer tipo de dados (integer, string, float, objetos, etc)

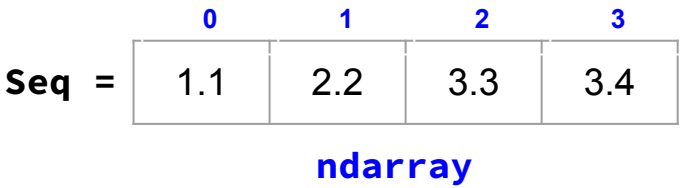
- Os labels podem ser indexados

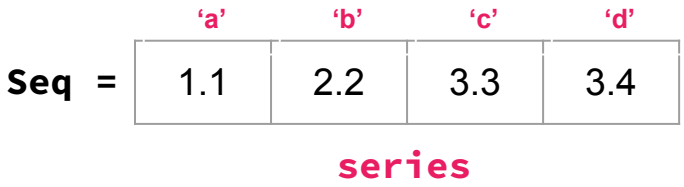
- Podemos criar **series** baseado em:
  - Listas
  - Dicionários
  - ndarrays



PANDAS SERIES vs NDARRAYs

- A principal diferença entre ndarrays e series é que nas séries os índices podem ser **rotulados**
- *Ndarrays* só permite a indexação por números inteiros
- Caso desejamos acessar o elemento 2.2 da sequência abaixo:
  - Usando um **ndarray** podemos acessar usando com Seq[1]
  - Usando uma **series** podemos acessar usando com Seq[‘b’]

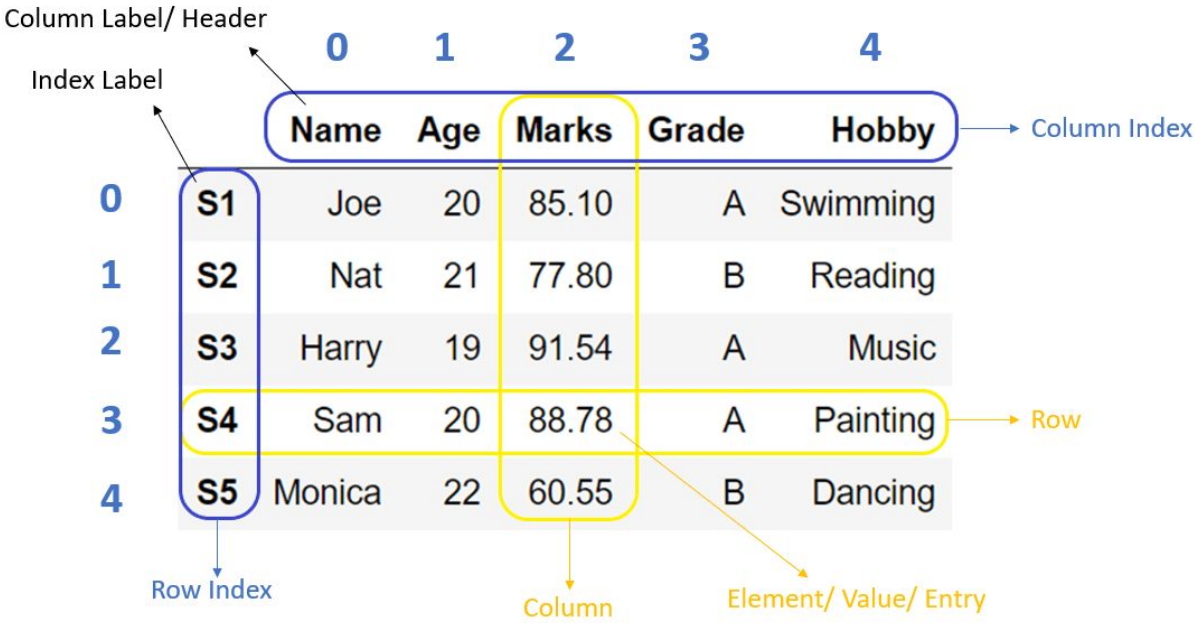




PANDAS DATAFRAMES

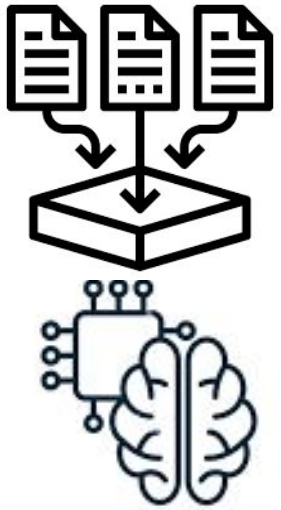
- Estrutura de dados no formato de tabela com rótulos nas linhas e nas colunas

- Podem ser indexados tanto por linha quanto por coluna



POR QUE USAR PANDAS?

- Simples de usar
- Integrado com diversas outras ferramentas de ciência de dados
- Ajuda a preparar os dados para usar no aprendizado de máquina



Prática

- Funções úteis
- Tipos de dados
- Importar e exportar dados
- Análise inicial de dados
- Visualizar e selecionar dados
- Manipular dados