

Manipulação de dados com Pandas



Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC)
Introdução à Ciências de Dados - UFC *Campus* Sobral – 2023.1
Andressa Gomes Moreira





Manipulação de Dados

- Uma das habilidades mais importantes de um cientista de dados é manipular dados.
- Consiste em coletar, limpar, processar e consolidar os dados para uso na análise.



Pandas

- Ferramenta de código aberto usada para análise e manipulação de dados.
- Um dos pacotes mais importante para Data Science em Python.
- Construído sobre o pacote Numpy.





Estruturas de Dados

Series

- Matriz unidimensional capaz de conter qualquer tipo de dados (números de ponto flutuante, inteiros, objetos Python, strings etc).

DataFrame

- Estrutura de dados rotulada bidimensional com colunas de tipos diferentes. Cada coluna em um DataFrame é uma Serie.



Estruturas de Dados

	Coluna A
0	10
1	30
2	5

Series



	Coluna B
0	20
1	15
2	24

Series



	Coluna A	Coluna B
0	10	20
1	30	15
2	5	24

DataFrame



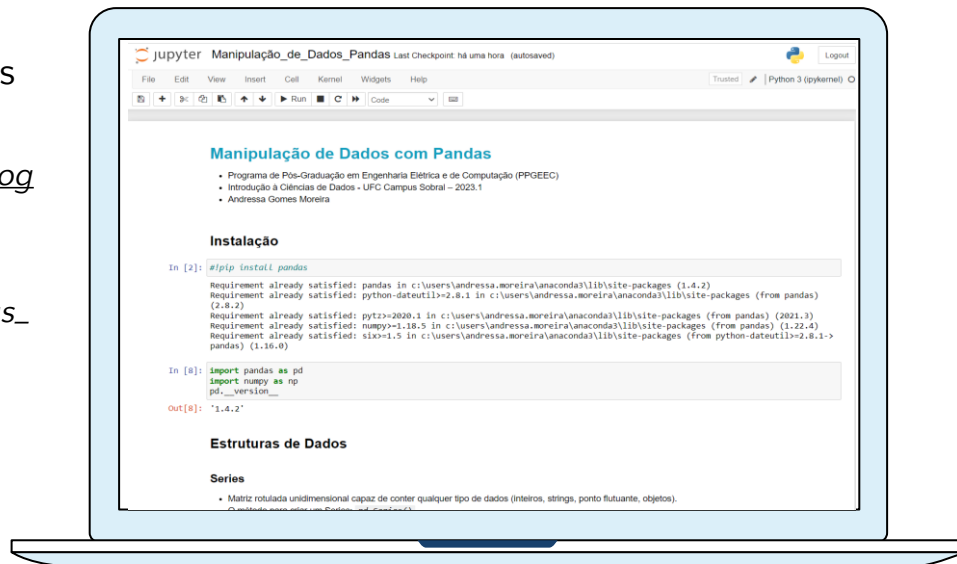
O que o Pandas pode fazer?

- Manipulação de dados ausentes
- Mutabilidade de tamanho: as colunas podem ser inseridas e excluídas
- Conversão de dados irregulares e indexados de maneira diferente em outras estruturas de dados (Python e NumPy) em objetos DataFrame
- Ferramentas robustas para carregar dados de arquivos simples (CSV), arquivos Excel, bancos de dados.



Aplicação Jupyter

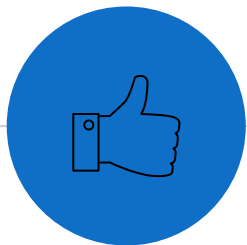
- Análise de dados Google Play Store Apps
- Dataset:
<https://www.kaggle.com/datasets/lava18/google-play-store-apps>
- Código - git clone
https://github.com/andressagomes26/ciencias_dados_estagio-docencia





Referências

- Documentação Pandas: [*https://pandas.pydata.org/*](https://pandas.pydata.org/)
- Python Data Science Handbook: [*http://www12.allitebooks.org/*](http://www12.allitebooks.org/)



Obrigada!

Alguma dúvida ?

● andressagomes@alu.ufc.br