MACHINE LEARNING APLICADO À ANÁLISE DE DADOS

AULA 4 - 18/03/2021

ANÁLISE DE DADOS COM PANDAS

He pandas

- Pandas é uma biblioteca open source que permite fácil análise e tratamento de dados estruturados de variadas formas.

POR QUE USAR PANDAS?

• É uma excelente opção para Data Munging/Wrangling (Manipulação de dados)

• É largamente utilizado na fase de preparação de dados

 Possui funcionalidades de manipulação de dados de séries temporais de alta performance

PANDAS + NUMPY

• Estruturas de dados de alto nível

Funcionalidades de séries temporais

Tratamento flexível de Dados Missing

• "Merge" e outras operações relacionais com SQL

O QUE VEREMOS EM PANDAS:

- SERIES (Criando e manipulando séries)
- DATAFRAMES (Criando e manipulando dataframes)

SERIES

• É um array unidimensional que contém um array de dados e um array de labels, chamado índice

```
Ser2 = Series([10,20,30,40], index = ['a', 'b', 'c', 'd'])

Ser2

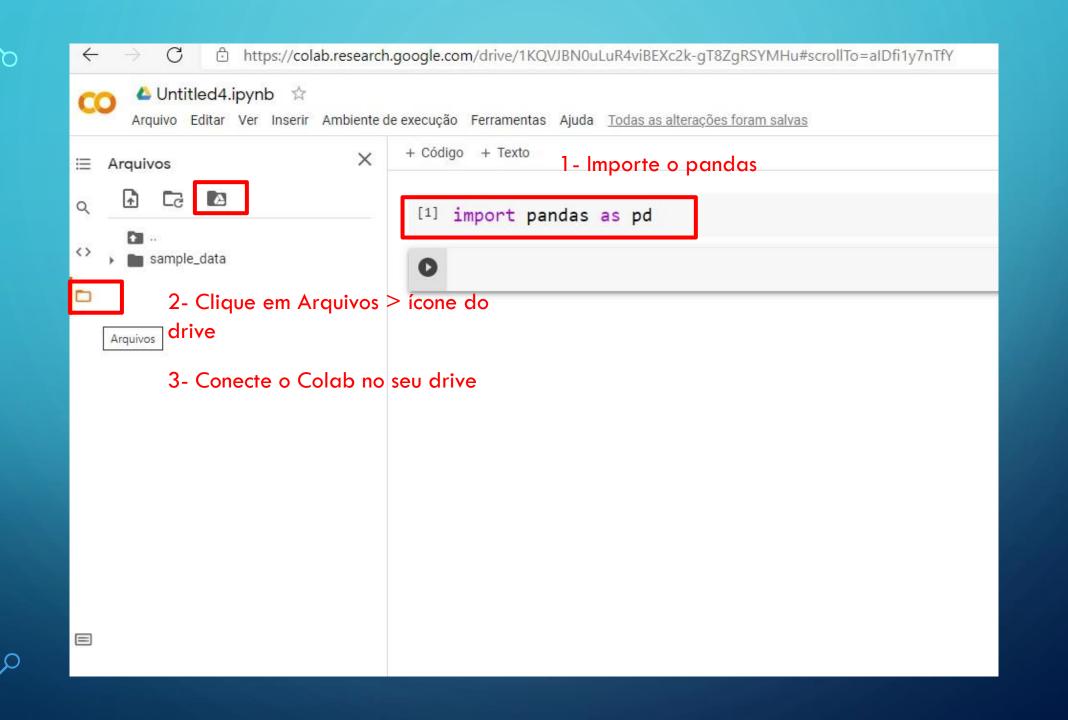
a 10
b 20
c 30
d 40
dtype: int64
```

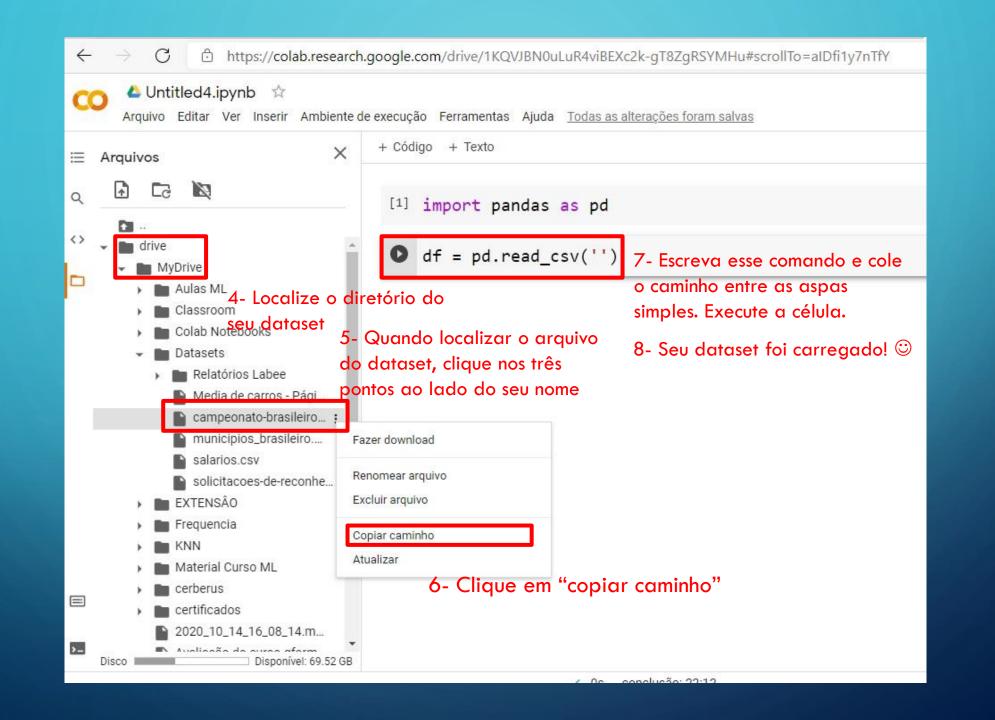
DATAFRAMES

• É uma estrutura de dados semelhante a uma planilha do Excel, dispondo de linhas e colunas que podem ter diferentes tipos de valores (numéricos, strings, booleanos, etc...)

0	turma			
		Aluno	Idade	Nota
	0	Alice	10	8.5
	1	Bruna	12	9.9
	2	Carlos	11	9.8
	3	Daiane	14	7.0
	4	Emily	13	5.0

COMO CARREGAR PARA O COLAB UM ARQUIVO DO TIPO CSV QUE ESTÁ ARMAZENADO EM SEU DRIVE:





OUTRAS FORMAS DE CARREGAR UM ARQUIVO PARA O SEU COLAB:

5 modos de carregar arquivos no Google Colab – Mayra Azevedo

(https://nymarya.github.io/5-modos-de-carregar-arquivos-no-google-colab/)

LINKS ÚTEIS

- Documentação oficial Pandas (https://pandas.pydata.org/)
- <u>5 modos de carregar arquivos no Google Colab Mayra Azevedo</u> (https://nymarya.github.io/5-modos-de-carregar-arquivos-no-google-colab/)