



MACHINE LEARNING APLICADO À ANÁLISE DE DADOS

AULA 4 – 18/03/2021

ANÁLISE DE DADOS COM PANDAS



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...)

	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

The background is a blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural networks, with lines connecting to small circles.

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

← → ↻ <https://colab.research.google.com/drive/1KQVJBN0uLuR4viBEXc2k-gT8ZgRSYMHu#scrollTo=aIDfi1y7nTfY>

CO Untitled4.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

Arquivos

🔍 📁 📄 📁

📁 ...

📁 sample_data

📁

Arquivos

+ Código + Texto

1- Importe o pandas

```
[1] import pandas as pd
```

▶

2- Clique em Arquivos > ícone do drive

3- Conecte o Colab no seu drive

← → ↻ <https://colab.research.google.com/drive/1KQVJBN0uLuR4viBEXc2k-gT8ZgRSYMHu#scrollTo=aIDfi1y7nTfY>

Untitled4.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

Arquivos

drive

MyDrive

Aulas ML

Classroom

Colab Notebooks

Datasets

Relatórios Labee

Media de carros - Pági

campeonato-brasileiro...

municipios_brasileiro...

salarios.csv

solicitacoes-de-reconhe...

EXTENSÃO

Frequencia

KNN

Material Curso ML

cerberus

certificados

2020_10_14_16_08_14.m...

Avaliação de curso reform

Disco

Disponível: 69.52 GB

+ Código + Texto

```
[1] import pandas as pd
```

```
df = pd.read_csv('')
```

4- Localize o diretório do seu dataset

5- Quando localizar o arquivo do dataset, clique nos três pontos ao lado do seu nome

6- Clique em "copiar caminho"

7- Escreva esse comando e cole o caminho entre as aspas simples. Execute a célula.

8- Seu dataset foi carregado! 😊

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/) (<https://pandas.pydata.org/>)
- [5 modos de carregar arquivos no Google Colab – Mayra Azevedo](https://nymarya.github.io/5-modos-de-carregar-arquivos-no-google-colab/)
(<https://nymarya.github.io/5-modos-de-carregar-arquivos-no-google-colab/>)