

**Educação  
Profissional  
Paulista**

Técnico em  
**Ciência de  
Dados**

# **Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn**

## **Pandas – Transformando Data Frame**

Aula 1

Código da aula: [DADOS]ANO1C2B4S25A1

**Bibliotecas: Pandas,  
NumPy, SciPy,  
Matplotlib e Seaborn**

## **Mapa da Unidade 5 Componente 3**

**Você está aqui!**  
Pandas –  
Transformando  
Data Frame

*semana*

**25**

Pandas – Combinando  
Data Frame

*semana*

**26**

*semana*

**23**

Pandas – Acesso  
e seleção

*semana*

**24**

Pandas – Ler e  
escrever

**Bibliotecas: Pandas,  
NumPy, SciPy,  
Matplotlib e Seaborn**

## **Mapa da Unidade 5 Componente 3**

**Você está aqui!**

Pandas – Transformando  
Data Frame

**Aula 1**

Código da aula: [DADOS]ANO1C2B4S25A1

**25**



## Objetivos da aula

- Conhecer a técnica de inspecionar o Data Frame com a biblioteca Pandas do NumPy.



## Recursos didáticos

- Recursos audiovisuais para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou internet;
- Software Anaconda/Jupyter Notebook instalado ou similar.



## Duração da aula

50 minutos.



## Competências técnicas

- Ser proficiente em linguagens de programação para manipular e analisar grandes conjuntos de dados;
- Usar técnicas para explorar e analisar dados, aplicar modelos estatísticos, identificar padrões, realizar inferências e tomar decisões baseadas em evidências.



## Competências socioemocionais

- Colaborar efetivamente com outros profissionais, como cientistas de dados e engenheiros de dados;
- Trabalhar em equipes multifuncionais, colaborando com colegas, gestores e clientes.







## Primeiras ideias

Há a sensação de que os elementos estão em movimento ou estáticos?

Você visualiza elementos que se mantêm ou se transformam?

Você conseguiu identificar números, dados, tabelas e *arrays* na imagem? Estão como em uma manipulação de dados?

## Ponto de partida

# Analizando – Escola Yannk

As notas da primeira prova dos alunos da turma A da Escola Yannk foram divulgadas. Os alunos realizaram os seguintes componentes:

C1 – Introdução à Ciência de Dados, ferramentas e jargões da área

C2 – Ética e responsabilidade em Inteligência Artificial

C3 – Programação aplicada à Ciência de Dados

C4 – Lógica, algoritmo e operação de planilhas eletrônicas

C5 – Banco de dados e computação em nuvem

Você, como técnico de dados, precisa realizar uma análise para saber quem precisa estudar mais e quem está com boa pontuação.



© Getty Images

Situação fictícia elaborada especialmente para o curso.



# Ponto de partida

Para isso, você abriu o arquivo abaixo e começou a analisar as informações.

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem
0	Micael Miguel Ehenvic Puzuho de Oliveira	6.0	3.0	7.0	4.0	6.0
1	Janaína Mupru Froiri	9.0	2.0	6.0	7.0	4.0
2	Álvaro Savo da Silva	3.0	7.0	7.0	2.0	5.0
3	Marlon Rosatto Neto	4.0	1.0	7.0	5.0	1.0
4	Sandra Inetirn	4.0	0.0	9.0	5.0	8.0
5	Charlene Sandra Klein de Simões	0.0	9.0	2.0	6.0	3.0
6	Gilberto Martinho Riga	8.0	2.0	4.0	2.0	6.0
7	Gilson de Albuquerque	4.0	8.0	6.0	1.0	3.0
8	Fred Jair dos Santos	8.0	1.0	9.0	8.0	9.0

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Apenas observando a planilha, você consegue responder às seguintes questões?

1. Qual a média do Gilberto?
2. Qual a média dos alunos de Programação aplicada à Ciência de Dados?
3. Qual a nota mais alta do Fred?
4. É possível mudar o nome dos componentes?

Você acha que é melhor utilizar Pandas para chegar a algumas respostas dessas questões?

## Construindo o **conceito**

# Lendo o arquivo

Abra o arquivo **notas\_CD.csv** com o Pandas:

```
1 # importar a biblioteca
2 import pandas as pd
3
4 # Ler o arquivo e salvar no dataframe df
5 df = pd.read_csv("notas_CD.csv")
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

**Observação:** o arquivo deve estar na mesma pasta em que você está abrindo o notebook.ipynb



Baixe o material de apoio.



Construindo  
o **conceito**

Vamos olhar o Data Frame?

O arquivo possui:

Informações pessoais

- Nome do aluno;
- Data de nascimento.

Informações educacionais:

- Notas em cada um dos componentes;
- Turma da qual o aluno faz parte;
- Escola em que estuda.

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem	Turma	Nasc	Escola
0	Micael Miguel Ehenvic Puzuho de Oliveira	9.0	7.5	9.5	19.0	2.500	Turma_5	30/01/2007	Escola 2
1	Janaína Mupru Froiri	9.0	4.0	9.0	20.0	1.250	Turma_3	18/09/2008	Escola 1
2	Álvaro Savo da Silva	9.0	8.0	7.0	15.0	1.875	Turma_4	06/07/2006	Escola 1
3	Marlon Rosatto Neto	6.0	5.5	10.0	16.0	0.125	Turma_1	16/10/2006	Escola 2
4	Sandra Inetirn	3.0	3.5	6.5	15.0	1.875	Turma_3	08/07/2006	Escola 1

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

## Construindo o **conceito**

# Inspeccionar o Data Frame

Vamos juntos explorar algumas informações sobre o Data Frame!

### **df.info()**

- O método info() fornece detalhes sobre as colunas, tipos de dados e quantidade de valores não nulos.
- Ele exibe informações como o nome das colunas, o número de entradas não nulas e os tipos de dados.

### **df.shape**

- A propriedade shape retorna uma tupla com o número de linhas e colunas do Data Frame.

### **df.dtypes**

- Atributo que mostra os tipos de dados de cada coluna.



## Construindo o conceito

# Data Frame: df.info()

Veja no nosso Data Frame:

```
1 # Ver informações do dataframe
2
3 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 100 entries, 0 to 99
```

```
Data columns (total 9 columns):
```

#	Column	Non-Null Count	Dtype
0	Nome	100 non-null	object
1	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	100 non-null	float64
2	C2 - Ética e Responsabilidade em Inteligência Artificial	100 non-null	float64
3	C3 - Programação Aplicada a Ciência de Dados	100 non-null	float64
4	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	100 non-null	float64
5	C5 - Banco de Dados e Computação em Nuvem	100 non-null	float64
6	Turma	100 non-null	object
7	Nasc	100 non-null	object
8	Escola	100 non-null	object

```
dtypes: float64(5), object(4)
```

```
memory usage: 7.2+ KB
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

## Construindo o conceito

# Data Frame: df.shape

Observe no nosso Data Frame cada coluna e seus tipos:

```
1 # Ver a quantidade de linhas e colunas de um df
2
3 df.shape
```

(100, 9)

```
1 # Ver
2 df.dtypes
```

Nome	object
C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	float64
C2 - Ética e Responsabilidade em Inteligência Artificial	float64
C3 - Programação Aplicada a Ciência de Dados	float64
C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	float64
C5 - Banco de Dados e Computação em Nuvem	float64
Turma	object
Nasc	object
Escola	object
dtype:	object

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Construindo  
o **conceito**

## Inspeccionar o Data Frame

<b>df.head()</b>	O método head() exibe as cinco primeiras linhas do Data Frame. Isso ajuda a ter uma ideia inicial dos dados.
<b>df.tail()</b>	Método tail() exibe as cinco últimas linhas do Data Frame.
<b>df.describe()</b>	O método describe() apresenta estatísticas resumidas para colunas numéricas. Ele inclui média, desvio padrão, mínimo, máximo e quartis.

Construindo  
o **conceito**

# Data Frame: df.head()

Veja como fica em nosso Data Frame.

```
1 df.head()
```

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem	Turma	Nasc	Escola
0	Micael Miguel Ehenvic Puzuho de Oliveira	9.0	7.5	9.5	19.0	2.500	Turma_5	30/01/2007	Escola 2
1	Janaína Mupru Froiri	9.0	4.0	9.0	20.0	1.250	Turma_3	18/09/2008	Escola 1
2	Álvaro Savo da Silva	9.0	8.0	7.0	15.0	1.875	Turma_4	06/07/2006	Escola 1
3	Marlon Rosatto Neto	6.0	5.5	10.0	16.0	0.125	Turma_1	16/10/2006	Escola 2
4	Sandra Inetirn	3.0	3.5	6.5	15.0	1.875	Turma_3	08/07/2006	Escola 1

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Construindo  
o conceito

# Data Frame: df.tail()

Em nosso Data Frame:

1 df.tail()

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem	Turma	Nasc	Escola
95	Tiago Jotreson de Oliveira	2.5	7.0	9.5	9.0	0.375	Turma_4	02/03/2009	Escola 1
96	Táles Isapra de Queiroz	7.5	1.5	9.5	15.0	2.500	Turma_1	18/10/2007	Escola 1
97	Paulínia Denise de Souza Loesuni	7.5	10.0	7.5	20.0	2.500	Turma_4	16/10/2008	Escola 3
98	Marcelo de Barbosa Gomes	3.5	10.0	8.5	9.0	2.500	Turma_3	19/03/2007	Escola 1
99	Suzanne Eoson do Rio	5.5	1.0	5.0	15.0	0.125	Turma_4	20/10/2008	Escola 1

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

## Construindo o conceito

# Data Frame: df.describe()

Em nosso Data Frame:

```
1 df.describe()
```

	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem
count	100.000000	100.000000	100.000000	100.000000	100.000000
mean	6.705000	7.260000	8.320000	15.840000	1.448750
std	2.490482	2.996193	1.575571	3.454524	0.952153
min	0.000000	0.500000	4.000000	8.000000	0.000000
25%	5.000000	5.000000	7.500000	13.000000	0.500000
50%	7.250000	8.500000	8.750000	16.000000	1.625000
75%	9.000000	10.000000	9.500000	19.250000	2.500000
max	9.000000	10.000000	10.000000	20.000000	2.500000

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

**Observe que apenas vieram as colunas numéricas.**



Construindo  
o **conceito**

## Inspeccionar o Data Frame

<b>df.sort_values(by="Nome")</b>	Classificar o Data Frame com base na coluna "Nome".
<b>df['Turma'].value_counts()</b>	Conta a quantidade de ocorrências de cada valor categórico na coluna "Turma".
<b>df['Escola'].unique()</b>	Retorna os valores únicos da coluna "Escola".
<b>df['Escola'].nunique()</b>	Retorna o número de escolas únicas na coluna "Escola".

## Construindo o conceito

# Data Frame: `df.sort_values(by="Name")`

Observe como os **índices** estão em nosso Data Frame:

```
1 df.sort_values(by="Name")
```

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem	Turma	Nasc	Escola
90	Aloísio dos Santos	9.0	8.0	9.0	11.0	0.250	Turma_5	14/05/2007	Escola 3
48	Ana Mikaella Ximenes	5.0	10.0	7.5	16.0	0.375	Turma_2	26/03/2009	Escola 1
62	Anakin Barnabé Weber da Costa	9.0	10.0	6.0	10.0	0.000	Turma_5	22/01/2008	Escola 1
45	Anne Laila de Leão Prates	6.5	2.5	10.0	11.0	2.500	Turma_4	03/03/2009	Escola 2
21	Arthur Otávio de Moraes Maos Sifro Júnior	9.0	9.0	9.0	17.0	2.250	Turma_3	02/06/2007	Escola 3

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

## Construindo o **conceito**

# Inspecionar o Data Frame

Demais métodos no nosso Data Frame:

```
1 df['Turma'].value_counts()
```

```
Turma
Turma_5    20
Turma_3    20
Turma_4    20
Turma_1    20
Turma_2    20
Name: count, dtype: int64
```

```
1 df['Escola'].unique()
```

```
array(['Escola 2', 'Escola 1', 'Escola 3'], dtype=object)
```

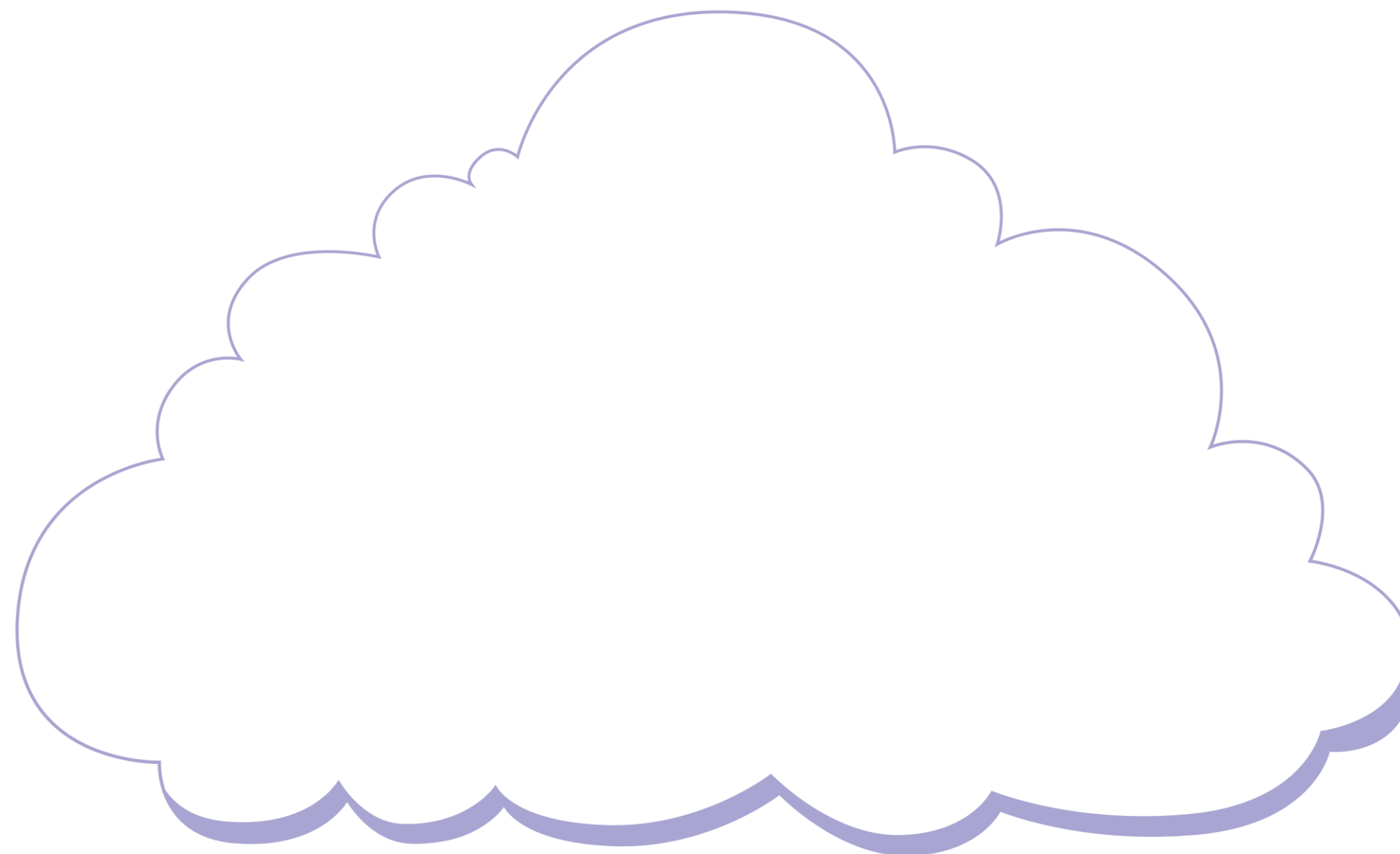
```
1 df['Escola'].nunique()
```

3

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



# Nuvem de palavras



© Getty Images

O que nós  
**aprendemos  
hoje?**



© Getty Images

O que nós  
**aprendemos  
hoje?**

## Então, ficamos assim:

- 1** Vimos alguns métodos de inspecionar o Data Frame, como head, tail, describe e info;
- 2** Vimos os métodos dtypes, unique, nunique, value\_counts, shape e suas atuações;
- 3** Esses métodos nos dão informações sobre o Data Frame, a série e a estrutura dos dados.



# Saiba mais

**Já pensou que o Pandas está mais perto de você do que você imagina?**

Netflix, Airbnb e Google são empresas que usam essa ferramenta!

Ficou curioso? Conheça sobre:

ALMEIDA, M. *Pandas Python: o que é, para que serve e como instalar*, 16 ago. 2023. Disponível em: <https://www.alura.com.br/artigos/pandas-o-que-e-para-que-serve-como-instalar>. Acesso em: 11 jul. 2024.



# Referências da aula

MCKINNEY, W. *Python para análise de dados: tratamento de dados com Pandas, NumPy & Jupyter*. São Paulo: Novatec, 2023.

PANDAS. *Pandas documentation*, 10 abr. 2024. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 11 jul. 2024.

Identidade visual: imagens © Getty Images

**Educação  
Profissional  
Paulista**

Técnico em  
**Ciência de  
Dados**