

**Senac**

**Fecomércio  
Sesc**

## Data Science – Princípios e Técnicas

Agosto  
2024



# Quem sou eu?

# Data Science

---

Prof. Marco Mialaret

Doutorando em Ciência da Computação pelo CIN-UFPE.

Mestre em Matemática pela UFPB.

Licenciado em Matemática pela UFRPE e Tecnólogo em Banco de Dados pelo SENAC.

Atuo na área acadêmica superior desde 2011 e atuo como cientista de dados (~~quando sobra tempo~~).

# Data Science



Onde me encontrar:

<https://www.linkedin.com/in/marco-mialaret-junior/>

e

<https://github.com/MatmJr>

# A disciplina

# Data Science

---

## Ementa

Explorando a linguagem Python e as bibliotecas disponíveis.  
Ciência de Dados, conceitos, abrangência e mercado de trabalho.  
Estatística aplicada à Ciência de Dados.  
Principais conceitos relacionados: BI, BIG Data, Data Mining, IA, Machine Learning e Deep Learning.  
Definição de Problemas em DataScience.  
Modelagem de Dados para projetos de Ciência de Dados.

# Data Science

---

Preparação, limpeza e tratamento de dados.

Explorando a linguagem R em seus fundamentos aplicados à Ciência de Dados.

IA e ML em seus fundamentos aplicados à Ciência de Dados.

Bancos de dados NoSql e armazenamento distribuído.

Processamento Distribuído e ferramentas.

Princípios de Ciência de Dados associados à IA e Machine Learning.

# Data Science

---

## Objetivo Geral

Capacitar estudantes com conhecimentos introdutórios de Data Science, cobrindo desde os fundamentos teóricos até a aplicação prática de técnicas.



# Data Science



## Objetivos Específicos

- Introduzir os conceitos fundamentais de Data Science, sua importância no cenário atual de tecnologia e análise de dados, e explorar as principais áreas de aplicação.
- Explorar conceitos estatísticos essenciais aplicados à Ciência de Dados, incluindo probabilidade, inferência estatística, testes de hipóteses e análise exploratória de dados.
- Detalhar os conceitos de Business Intelligence, Big Data, Data Mining e Inteligência Artificial, e como eles se relacionam e contribuem para a Ciência de Dados.

# Data Science



- Ensinar como identificar, definir e estruturar problemas a serem resolvidos através de projetos de Ciência de Dados, focando na formulação de hipóteses e na definição de objetivos claros.
- Abordar técnicas de modelagem e estruturação de dados para análise, incluindo normalização, dimensionalidade e técnicas de redução de dados.
- Desenvolver habilidades práticas na preparação, limpeza e tratamento de dados para garantir a qualidade e precisão necessárias para análises confiáveis.

# Data Science

---

- Introduzir os fundamentos da linguagem R e sua aplicabilidade na análise estatística e modelagem de dados em projetos de Ciência de Dados.
- Ensinar os fundamentos da linguagem Python e o uso de bibliotecas específicas para Inteligência Artificial e Machine Learning, como Pandas, NumPy, Scikit-learn, TensorFlow e Keras.

# Data Science



---

## Conteúdo Programático

- 1 - Introdução à Ciência de Dados: Conceitos e Campo de Atuação
- 2 - Estatística Aplicada à Ciência de Dados
- 3 - Definição de Problemas e Modelagem de Dados em Ciência de Dados
- 4 - Preparação, Limpeza e Tratamento de Dados
- 5 - Python e Bibliotecas para Ciência de Dados

# Data Science



---

## Conteúdo Programático

6 - Bancos de Dados para Ciência de Dados: SQL, NoSQL, e Armazenamento Distribuído

7 - Aplicação de Machine Learning e Deep Learning em Ciência de Dados

8 - Introdução à Linguagem R e Aplicações em Ciência de Dados

# Avaliações e Materiais

# Engenharia de Software

---

## Formas de Acompanhamento

- Frequência nas aulas
- Participação em sala
- Evolução do aprendizado
- Avaliação do método

# Engenharia de Software



---

## Metodologia de Avaliação

**Avaliação diagnóstica:** frequência nas aulas; participação nos debates realizados durante as aulas; atendimento aos horários de aulas e aos prazos estabelecidos para entrega de atividades e projetos;



# Engenharia de Software

---

**Avaliação formativa:** realização das atividades de classe durante a aula; cumprimento e entrega das atividades propostas ao final das aulas; entrega das atividades de laboratório, boa expressão escrita textual; a percepção de problemas, raciocínio crítico e analítico e para elaboração do projeto com as diretrizes desenvolvidas na disciplina de Data Science;

# Engenharia de Software

3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	3 +	3 +		9	2 +	3 +		5	3 +	8,5 O	2 +	2 +	1 +	2 +	7	B
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E
3 +	1 +	3 +		7	3 +	3 +		6	3 +	8 B	3 +	3 +	2 +	2 +	10	E

# Data Science

---

Fim das férias!



# Git & GitHub na nossa disciplina

# Data Science

---

**Git** é um sistema de controle de versão distribuído que permite aos desenvolvedores rastrear e gerenciar mudanças no código fonte ao longo do tempo. É uma ferramenta essencial para a colaboração em projetos de desenvolvimento de software.

# Data Science

---

**GitHub** é uma plataforma de hospedagem para repositórios Git que facilita a colaboração entre desenvolvedores. Além de armazenar código, o GitHub oferece ferramentas para revisão de código, gerenciamento de projetos e integração com outras ferramentas de desenvolvimento.

# Hora do Exercício

# Data Science

---

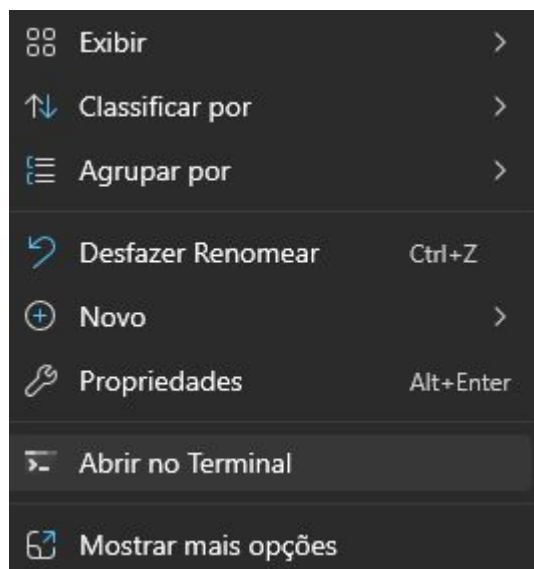
Exercício:

- 1 – Encontrar o usuário MatmJr no GitHub
- 2 – Encontrar o Repositório DataScienceSenac2024-2
- 3 – Clonar o repositório
- 4 – Acessar o Material da Primeira Semana



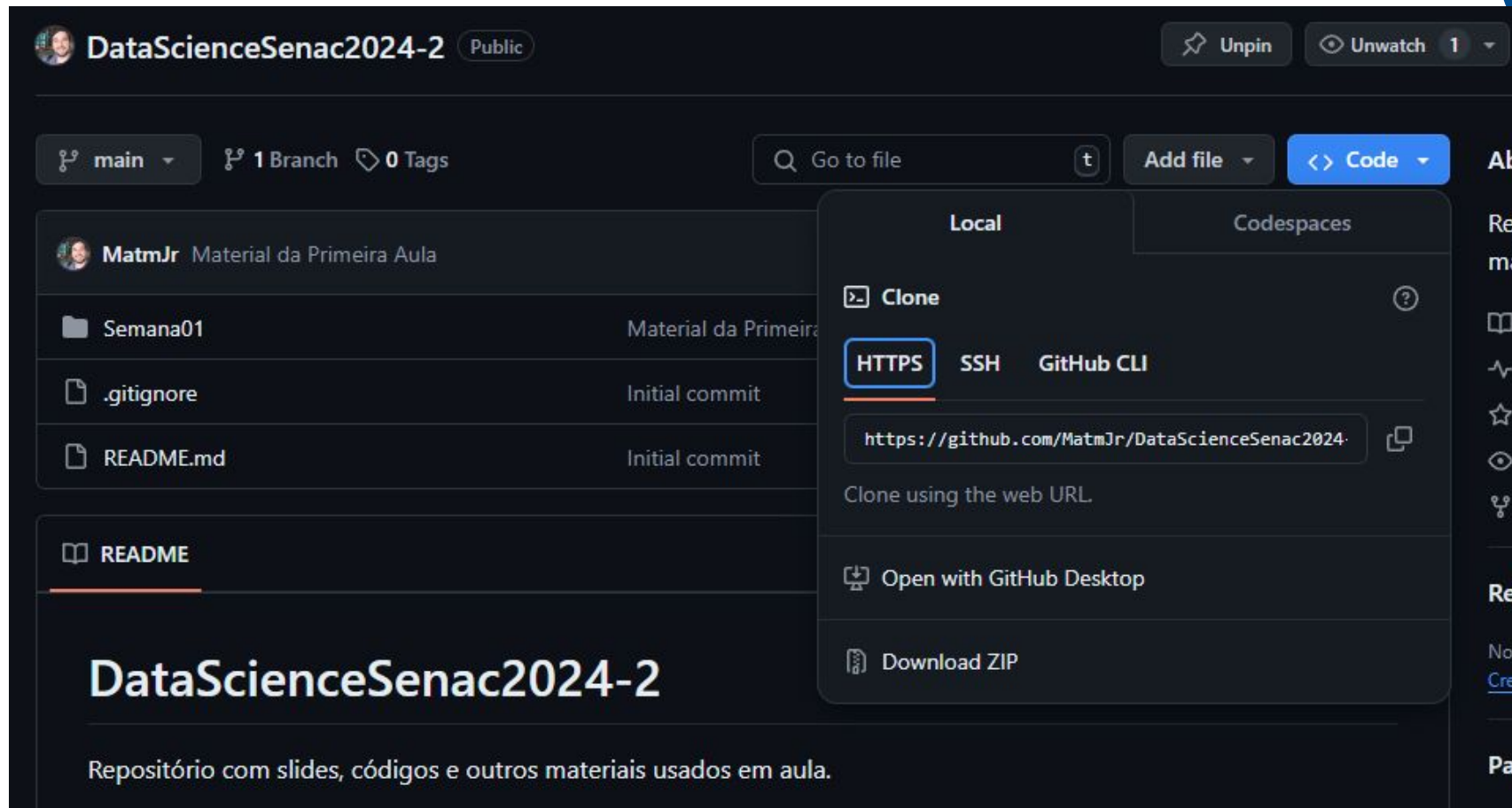
# Data Science

Após criar a pasta chamada DataScience, clicar com o botão do lado direito e selecionar a opção abrir no terminal



```
PS E:\Senac\Aulas\CursoDataScience> code .
```

# Data Science



# Data Science

---

```
x = 1
y = 0.1324

print("x vale ",x)
print("y vale ",y)
```

# Data Science

---

```
x = "Data "  
y = "Science"  
print(x + y)
```

# Data Science

---

```
x = "Data "  
y = "Science"  
print(x + y)
```

# Data Science

---

## INTEIROS

Em Python 3, números inteiros possuem precisão arbitrária, sem limite fixo de faixa e a ocupação de memória cresce conforme necessário. É possível declarar números inteiros em formatos decimais, octais, hexadecimais e binários.

# Data Science

---

*Para octais, adiciona-se um prefixo 0 ao número (ex: 0123), e para hexadecimais, utiliza-se o prefixo 0x ou 0X (ex: 0xFFFFFFFF ou 0X006699).*

# Data Science

---

## FLOAT

Representa números reais e/ou que possuem sinal de expoente (e ou E). Esses números são comumente chamados de números de ponto flutuante.



# Data Science

---

Por exemplo: 0.00213, .0015, 1.1265 e 3.01e23 (o mesmo que  $3.01 \times 10^{23}$ ).

# Data Science

---

## BOOL

Os valores do tipo bool representam dois valores: True (igual ao int 1) e False (igual ao int 0).

# Data Science

---

## NONE TYPE

NoneType é o tipo de None, uma constante embutida do Python que, assim como True e False, e é frequentemente utilizada para representar a ausência de um valor, similar ao null na linguagem C e derivadas. Exemplo: `a = None` (o mesmo que null em Java).

# Data Science

---

Uma variável do tipo string, basta que coloquemos entre aspas simples, duplas ou triplas

**Em Python, tudo é objeto, assim as Strings são objetos que tem embutidos vários métodos.**

# Data Science

## OPERADORES

Aritméticos	Comparação	Lógicos
+	==	and
-	!=	or
*	>	not
/ ou // (parte inteira)	<	
%	>=	
+= -= *= /=	<=	
**	in in not	
is		

# Data Science

---

Podemos fazer conversões para resolver os eventuais problemas com tipagem:

```
a = float(22/5)
b = int(4.5)
c = int(3.9)
d = int(0xff563)
e = float(int(3.9))
f = int(float(3.9))
g = int(float(3))
h = round(3.9)
i = round(3)
j = int(round(3.9))
```

# Dúvidas?

---

