



**Big Data**

Março  
2024



# Vamos nos conhecer



De onde eu vim e para aonde vamos?

# Quem sou eu?

# Big Data



---

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~~).

# Big Data

---

Onde me encontrar:

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

e

<https://github.com/MatmJr>

# A disciplina

# Big Data

---

## Ementa

Introdução ao Big Data, conceitos.

Infraestrutura, organização física e lógica, métodos de acesso.

Tratamento, transformação e limpeza de dados.

Armazenamento de Big Data, SQL, NoSQL e NewSQL.

Data Warehouse, data cleaning e interação.

Paralelismo, Clusters e Sistemas e plataformas distribuídas,

MapReduce, Hadoop e Spark.

Machine Learning e Deep Learning.

# Big Data

---

## Objetivo Geral

Capacitar profissionais com conhecimentos aprofundados em Big Data, cobrindo desde os fundamentos teóricos até a aplicação prática de técnicas e ferramentas de ponta no campo.



# Big Data



---

## Objetivos Específicos

- Introduzir os conceitos fundamentais de Big Data e sua importância no contexto atual da tecnologia da informação.
- Explorar a infraestrutura necessária para a implementação de soluções de Big Data, incluindo organização física e lógica e métodos de acesso.
- Desenvolver habilidades práticas no tratamento, transformação e limpeza de dados para garantir qualidade e precisão.

# Big Data

---

- Apresentar os diferentes modelos de armazenamento de dados, como SQL, NoSQL e NewSQL, e suas aplicações específicas.
- Discutir a arquitetura e funcionamento de Data Warehouses, além das técnicas de data cleaning e integração de dados.
- Explorar sistemas de processamento paralelo, clusters e plataformas distribuídas, incluindo MapReduce, Hadoop e Spark.
- Abordar as aplicações de Machine Learning e Deep Learning no contexto de Big Data, com foco na extração de insights e na tomada de decisões baseada em dados.

# Big Data



---

## Conteúdo Programático

- 1 - Introdução ao Big Data: Conceitos e Aplicações
- 2 - Infraestrutura para Big Data: Organização e Métodos de Acesso
- 3 - Processos de Tratamento e Limpeza de Dados
- 4 - Armazenamento e Gerenciamento de Big Data: SQL, NoSQL, NewSQL
- 5 - Data Warehousing e Estratégias de Interação com Grandes Volumes de Dados

# Big Data



---

## Conteúdo Programático

- 6 - Paralelismo e Computação Distribuída: Clusters, MapReduce, Hadoop, Spark
- 7 - Implementação de Machine Learning e Deep Learning em Ambientes de Big Data
- 8 - Visualização e Interpretação de Grandes Conjuntos de Dados
- 9 - Considerações Éticas e de Privacidade no Trabalho com Big Data

# Avaliações e Materiais

# Big Data



---

## Formas de Acompanhamento

- Frequência nas aulas
- Participação em sala
- Entrega das atividades

# Big Data

---

## Material de Estudo:

- Slides no Classroom do Aluno
- GitHub
- Bibliografia da Disciplina
- Vídeos
- Trabalhos
- Estudos de Casos
- Atividades



# Big Data



---

## Metodologia de Avaliação

- **Primeiro Processo Avaliativo** – Entrega das atividades e do projeto 1.
- **Segundo Processo Avaliativo** - Entrega das atividades e do projeto 2.



# Conhecendo vocês

# Big Data

---

Agora é com vocês!

- Nome
- Idade
- Você está procurando emprego? Em qual área?
- Um fato aleatório sobre você

# Big Data

---

Fim das férias!

# Git & GitHub na nossa disciplina

# Big Data

---

**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.

# Big Data

---

**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.

# Big Data

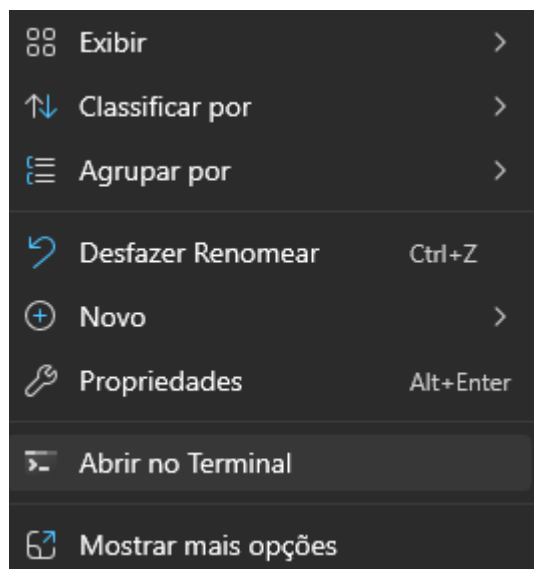
---

Exercício:

- 1 – Encontrar o usuário MatmJr no GitHub
- 2 – Encontrar o Repositório CursoBigData
- 3 – Acessar o Material da Primeira Semana

# Big Data

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

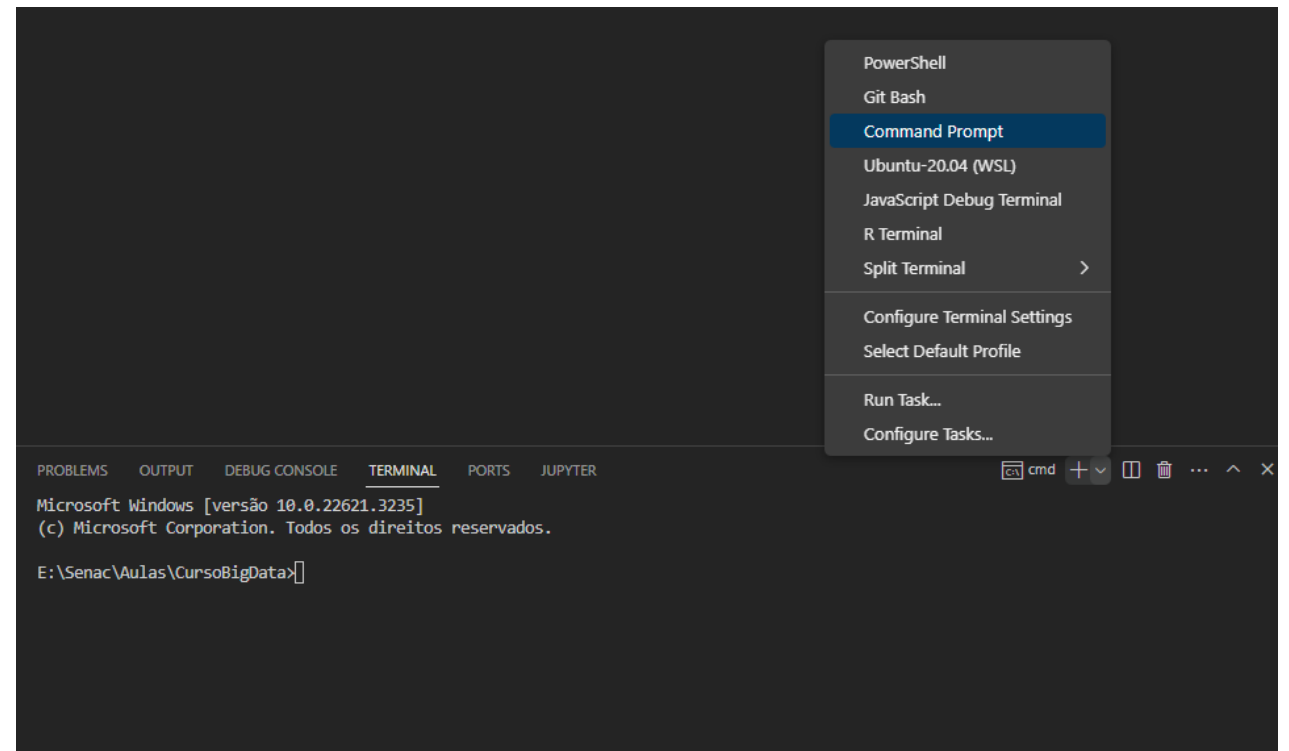
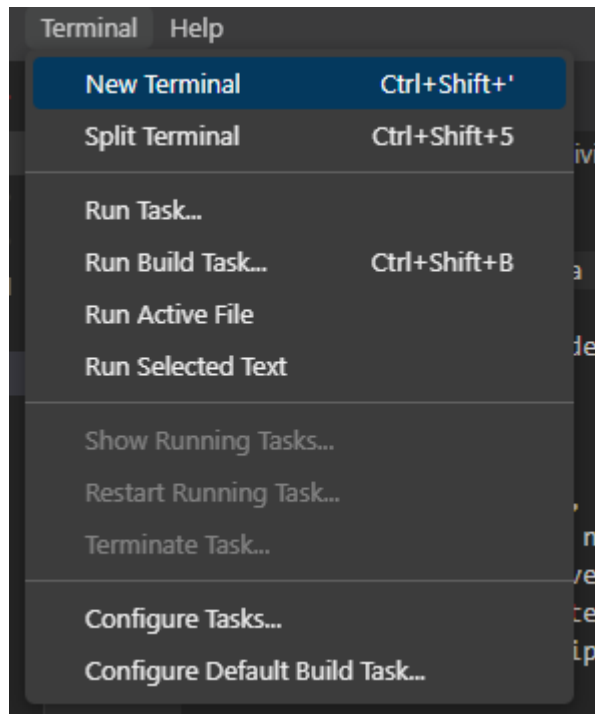


```
regamento de perfis pessoais e do sistema levou 729ms.  
PS E:\Senac\Aulas\Slides\BigData\BigData> |code .
```



# Big Data

No VSCode:



# Big Data

---

Criando o ambiente:

```
E:\Senac\Aulas\CursoBigData>python -m venv venv
```

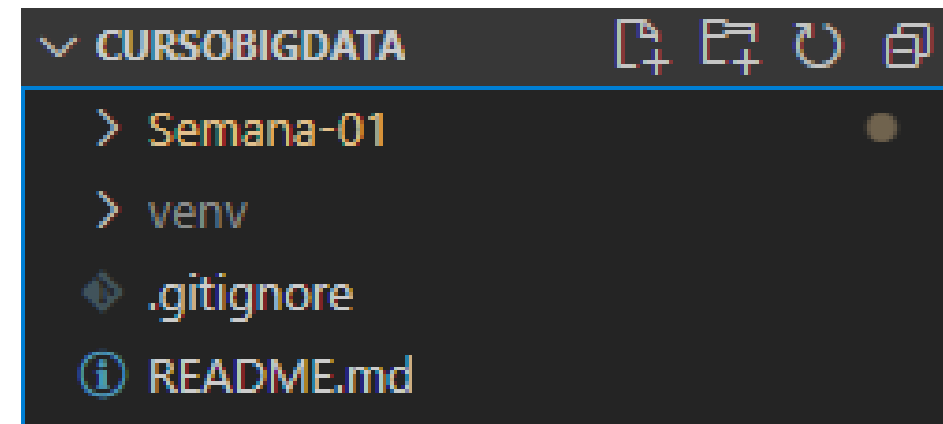
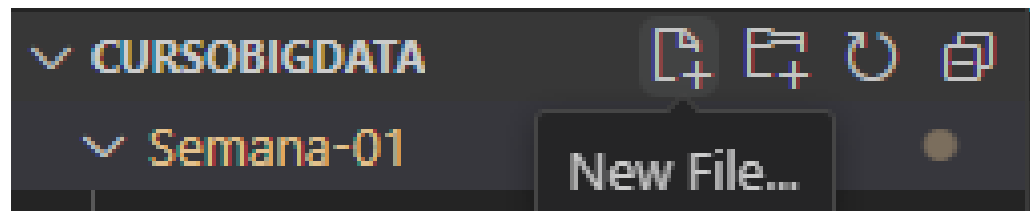
Ativando o ambiente:

```
E:\Senac\Aulas\CursoBigData>.\venv\Scripts\activate
```

```
(venv) E:\Senac\Aulas\CursoBigData>|
```

# Big Data

Criação de arquivos no VSCode:



# Big Data



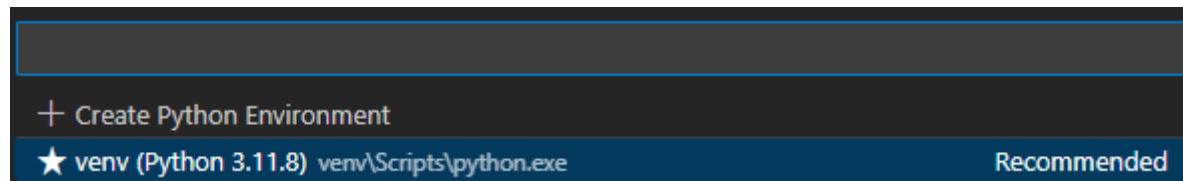
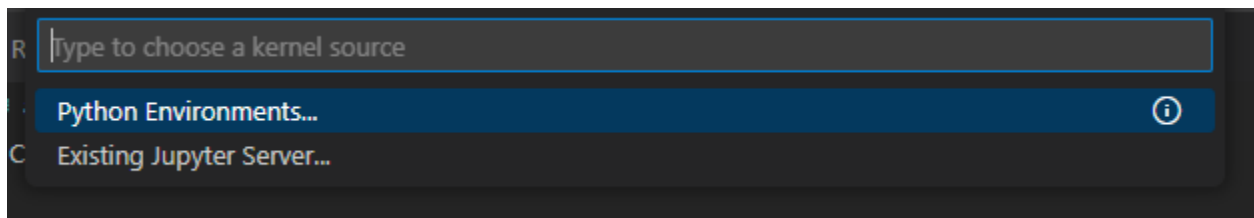
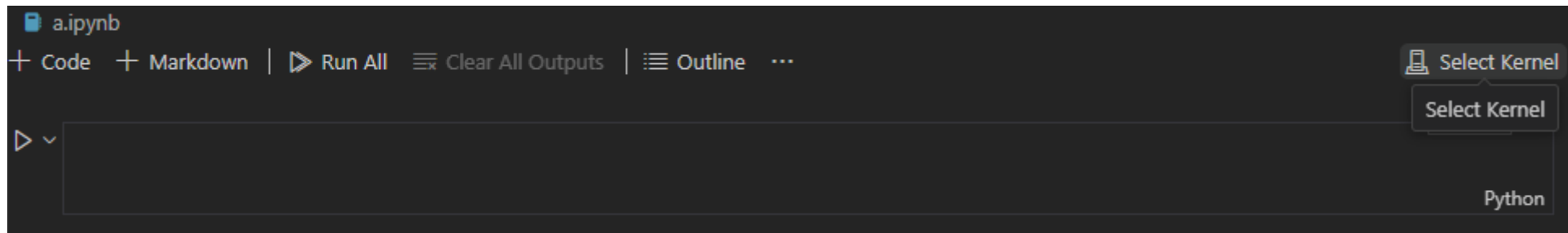
Crie um arquivo chamado aula01.ipynb.

No terminal, instale o pandas no ambiente virtual:

```
(venv) E:\Senac\Aulas\CursoBigData>pip install pandas
Collecting pandas
  Using cached pandas-2.2.1-cp311-cp311-win_amd64.whl.metadata (19 kB)
Collecting numpy<2,>=1.23.2 (from pandas)
  Using cached numpy-1.26.4-cp311-cp311-win_amd64.whl.metadata (61 kB)
```

# Big Data

## Ativando o ambiente virtual no notebook Jupyter



# Big Data

Na primeira célula do notebook Jupyter

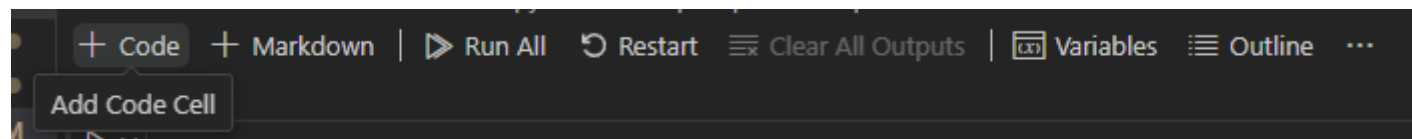
```
import pandas as pd

url = "https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/stuff/titanic.csv"

data = pd.read_csv(url)
```

[1] ✓ 2.4s

Crie uma nova célula:



# Big Data

Para visualizar os dados importados...

data								
✓ 0.0s								
Python								
Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare	
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000

# Dúvidas?

---





**Marco Mialaret, MSc**

**Telefone:**

**81 98160 7018**

**E-mail:**

**marcomialaret@gmail.com**

