



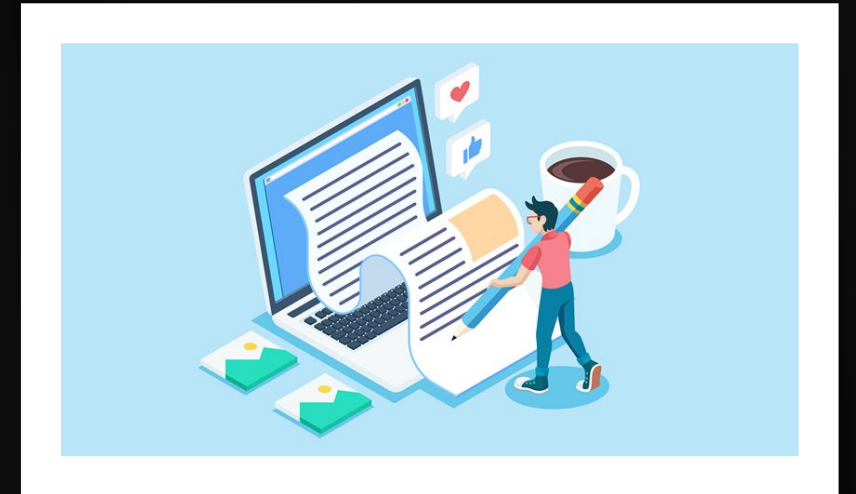
**UNINOVE**  
●●●●●

# Ciência de Dados

Prof. Dr. Edson Melo de Souza  
[souzaem@uni9.pro.br](mailto:souzaem@uni9.pro.br)

# Conteúdo Programático

- Introdução a Ciência de Dados e Google Colaboratory (Research Colab)
- Introdução ao Python
  - Operadores: Matemáticos, Lógicos e Relacionais
  - Desvios condicionais e repetição
  - Lista, dicionários e funções
- Computação Científica
  - Arrays e Matrizes
  - Álgebra Linear com NumPy
- Ciência de Dados
  - Análise de dados com pandas
  - Visualização de dados com pandas, matplotlib e seaborn
  - *Machine Learning* com Scikit-Learn (Modelos e Regressão Linear)
  - Web Scrap (raspagem de dados)





# Material de Apoio

## Github

O link está disponível em:  
Arquivos de Apoio na Central do Aluno

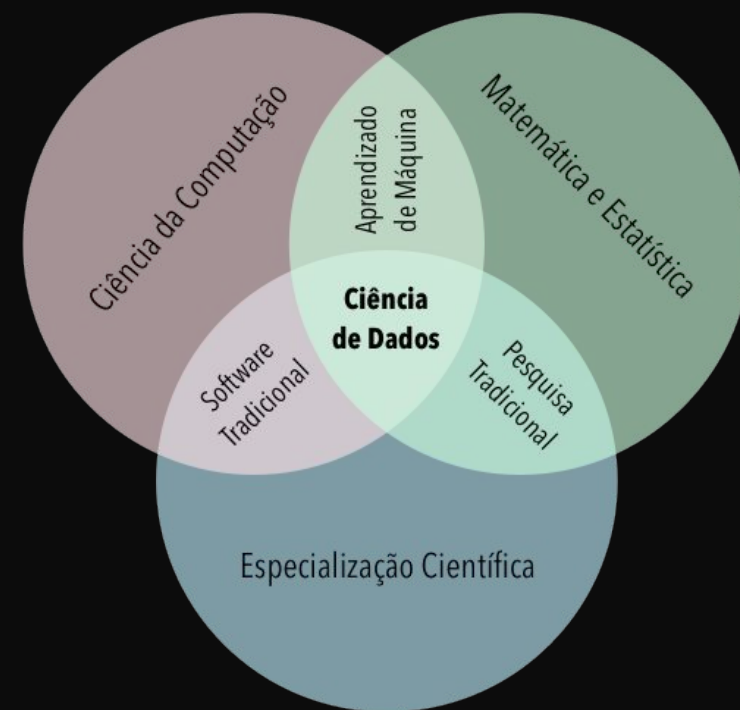
# Avaliações

- (A) 1 atividade digital individual com 10 questões de múltipla escolha (10 pontos)
- (B) 1 atividade em grupo com implementação de códigos (10 pontos)
- (C) 1 prova digital individual com 20 questões no final do semestre (10 pontos).

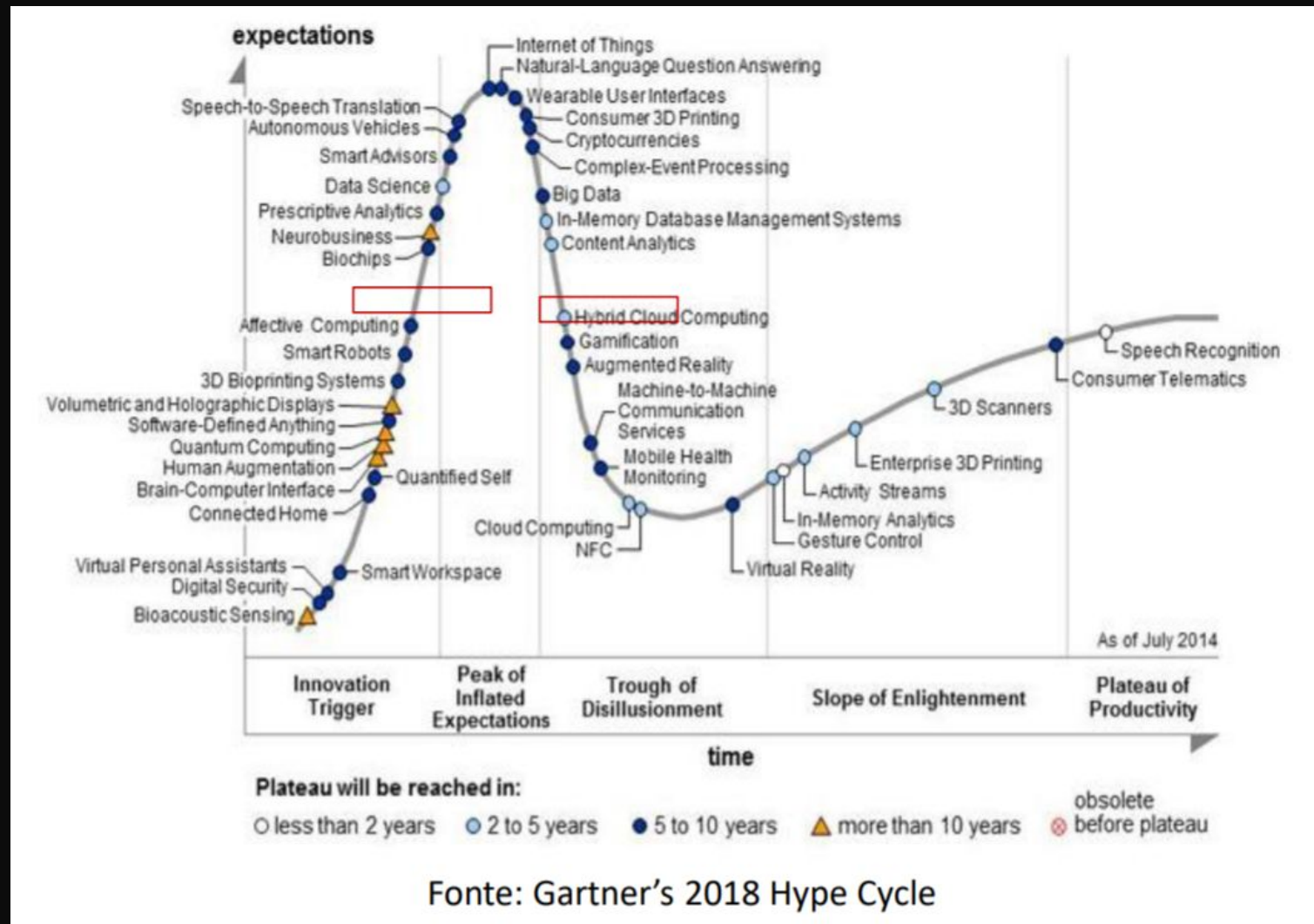
$$AV1 = (A+B+C)/3$$

# O que é Ciência de Dados

É uma área **interdisciplinar** voltada para o estudo e a **análise de dados** socioeconômicos, financeiros e sociais, estruturados e não-estruturados, que visa a extração de conhecimento, detecção de padrões e/ou obtenção de **insights** para possíveis tomadas de decisão (DHAR, 2013).



# Por que é tão interessante?





# MILAGRES DA ESTATÍSTICA

EU NÃO TENHO  
ONDE MORAR



EU TENHO  
DUAS CASAS



EXPURGO.COM.BR

# Análise de Dados Descritiva

- Este método é baseado em **uma** ou **várias questões** de pesquisa e não tem hipótese.
- Inclui a coleta de dados relacionados, organização, tabulação e descrição dos resultados.
- Uma análise descritiva básica envolve o **cálculo de medidas simples de composição e distribuição de variáveis**.



# Análise de Dados Preditiva

- É uma técnica analítica que usa dados, algoritmos e técnicas de *machine learning* para **antecipar** tendências e **realizar projeções**.
- A partir disso, decisões mais adequadas são tomadas de acordo com as expectativas.

# Análise de Dados Prescritiva

- É responsável por encontrar **uma solução** entre uma diversidade de variantes.
- Essa ferramenta usa diferentes **técnicas de simulação e otimização** para indicar o caminho que realmente deve ser seguido.

# Análise de Dados Diagnóstica

- Tem como finalidade compreender **o impacto de cada ação** dentro das estratégias concebidas.
- A proposta dessa análise é **destrinchar os dados** e concretizar todo o processo.
- Um exemplo é saber quem comprou, onde comprou, quando comprou, porque comprou ou, ainda, porque deixou de comprar.

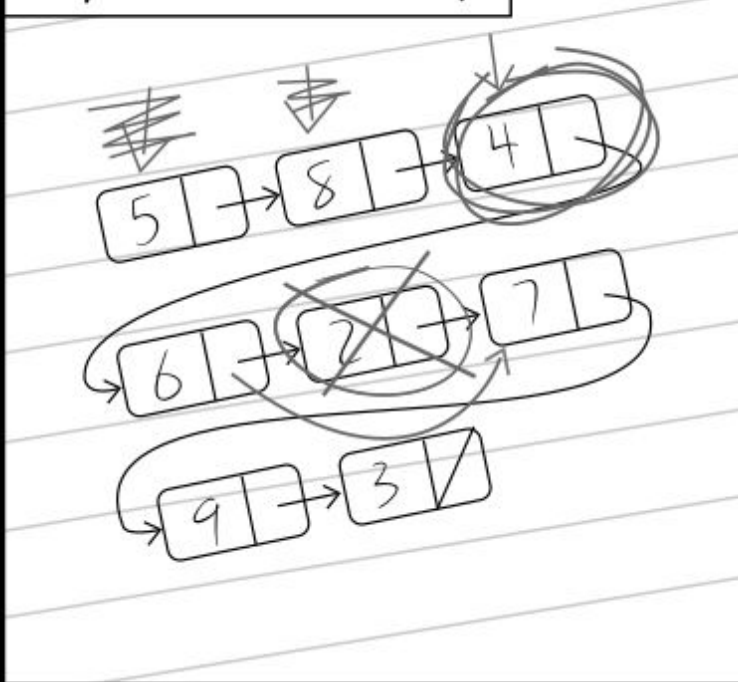
ALGUNS ANOS ATRÁS...

O QUE É ISSO, ALONSO?  
VOCÊ NÃO APRENDEU COMO  
FAZ PARA BUSCAR UM ELEMENTO  
NUMA LISTA ENCADEADA?

HMMMM...  
NÃO...



"TÁ, DEIXA EU EXPLICAR!"



NÃO DÁ PRA DAR  
SÓ UM CTRL+L QUE  
NEM NO WORD?



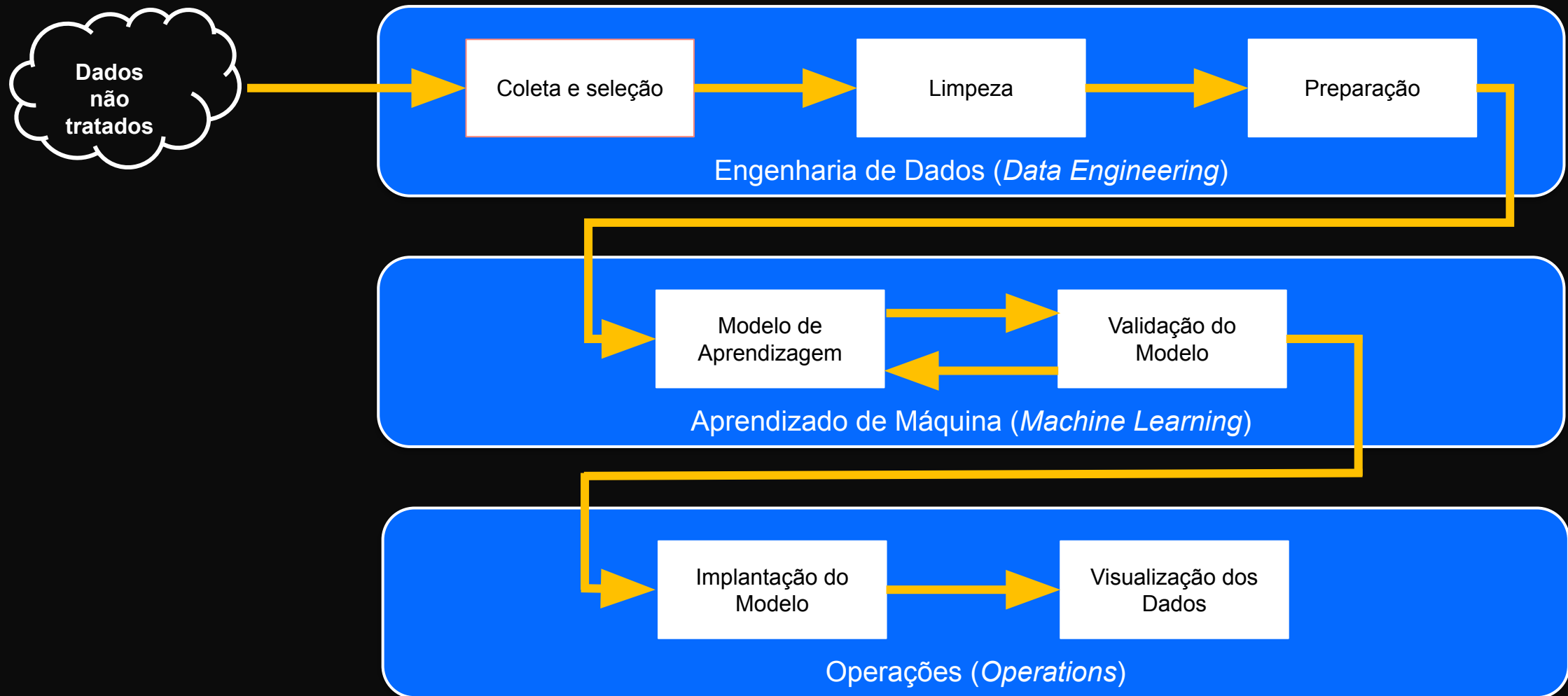
VIDA DE  
PROGRAMADOR

```
real historia;  
string sender = "alexandre";
```

#2142

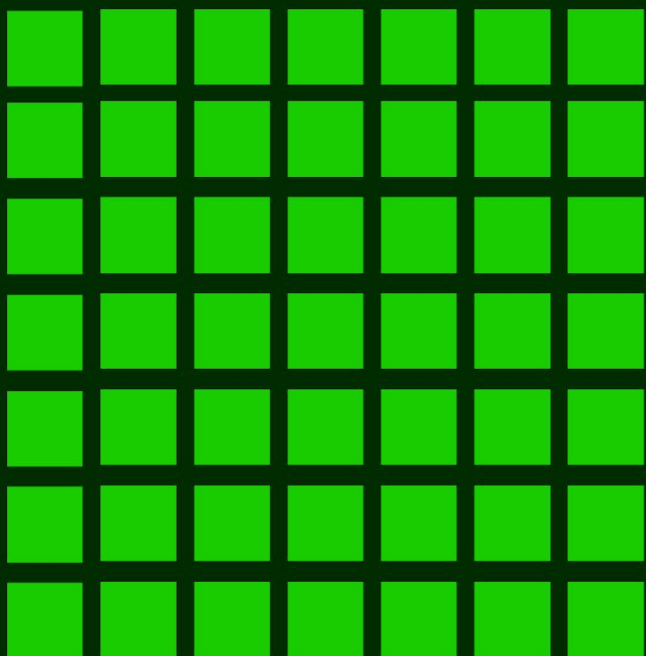


# Etapas da Análise de Dados

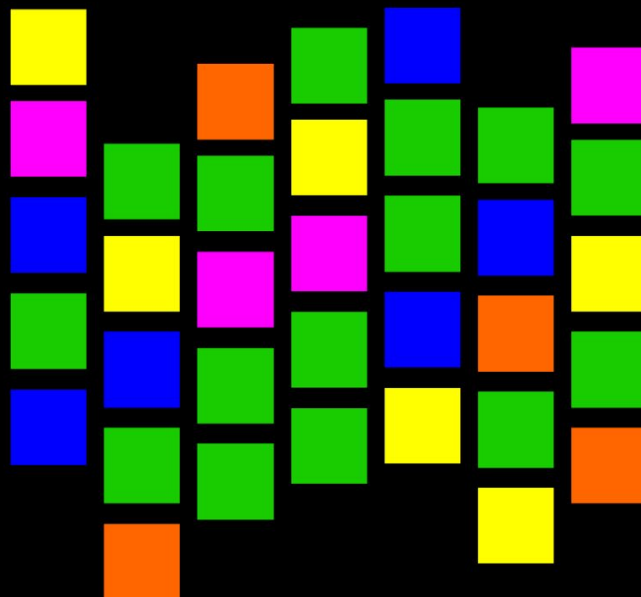


# Tipos de Dados

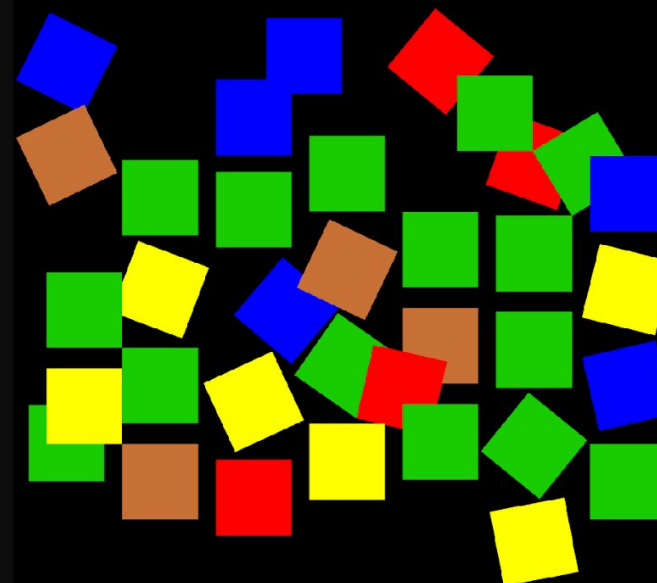
Dados Estruturados



Dados Semiestruturados

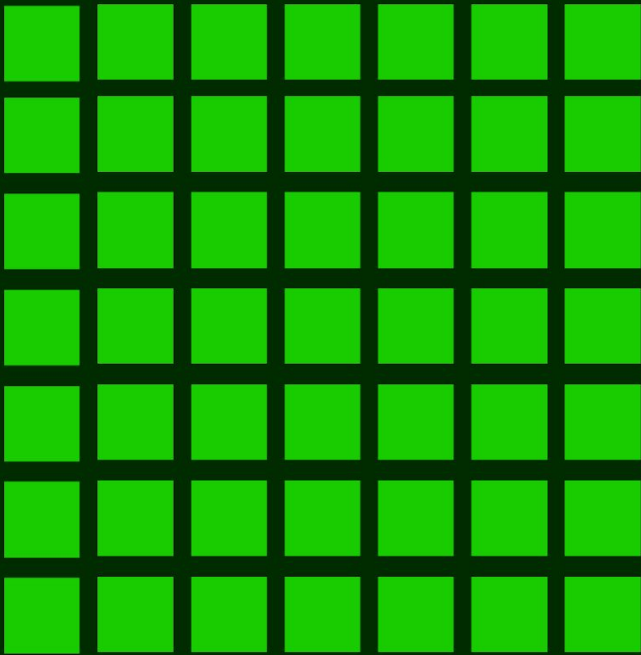


Dados Não estruturados



# Tipos de Dados

## Dados Estruturados

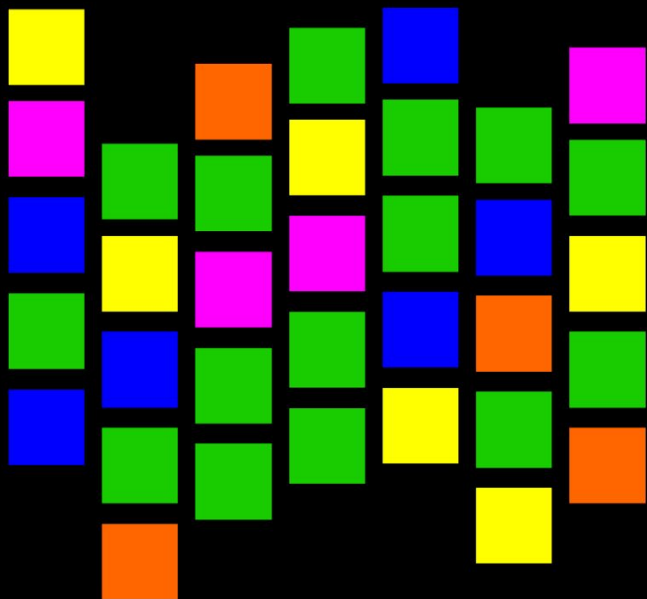


- São dados organizados de forma homogênea, no formato tabular.
- Exemplos:
  - Planilhas;
  - Arquivos CSV (texto separado por vírgula);
  - Bancos de dados relacionais.



# Tipos de Dados

## Dados Semiestruturados



- São dados heterogêneos, possuindo estrutura irregular (não tabular).
- Exemplos:
  - XML;
  - JSON (*Javascript Object Notation*).

# Tipos de Dados



- Não possuem padrão definido, podendo ser obtidos de diferentes fontes.
- Exemplos:
  - Textos;
  - Documentos;
  - Áudio;
  - Imagens.

# Limpeza de Dados ou *Data Cleaning*

Nesta etapa são identificadas e corrigidas:

- Anomalias nos dados;
- Dados faltantes;
- Dados supérfluos ou desnecessários;
- Inconsistências;
- Entre outros.

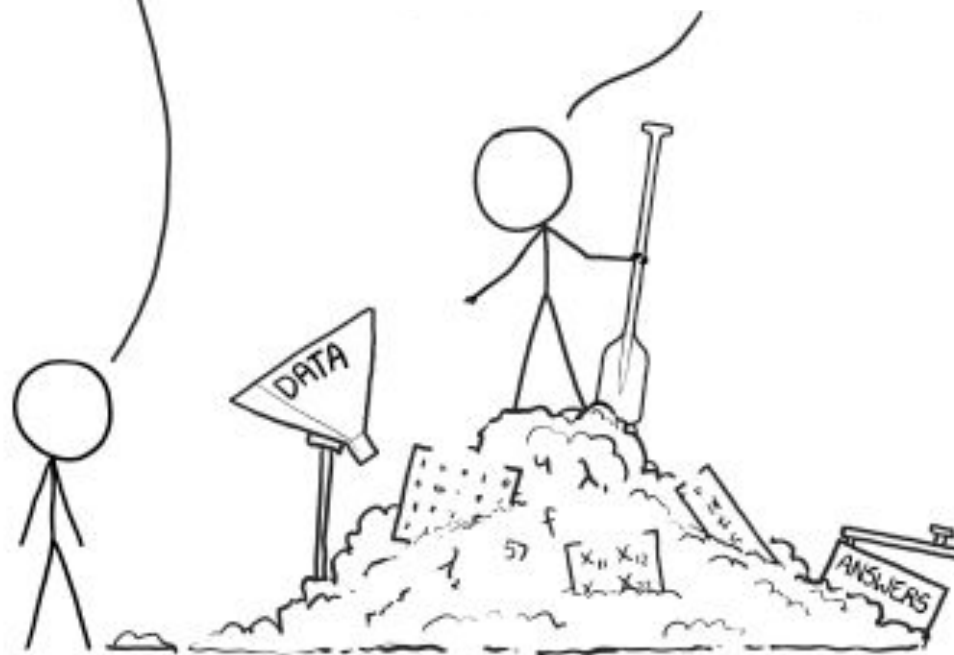


PRECISAMOS ENTREGAR LOGO NOSSAS ANÁLISES.

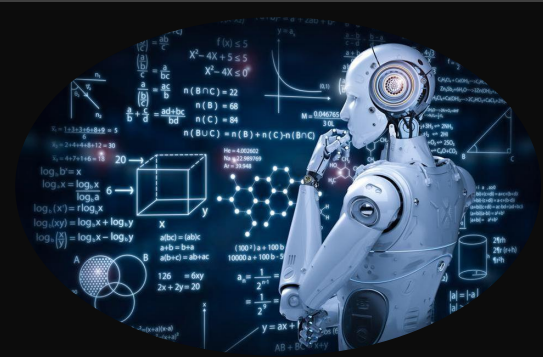
FIQUE TRANQUILO, JÁ ESTOU COLETANDO OS  
DADOS, DAÍ É SÓ LIMPAR E PROCESSAR.

MAS JÁ TEMOS DADOS SUFICIENTES?

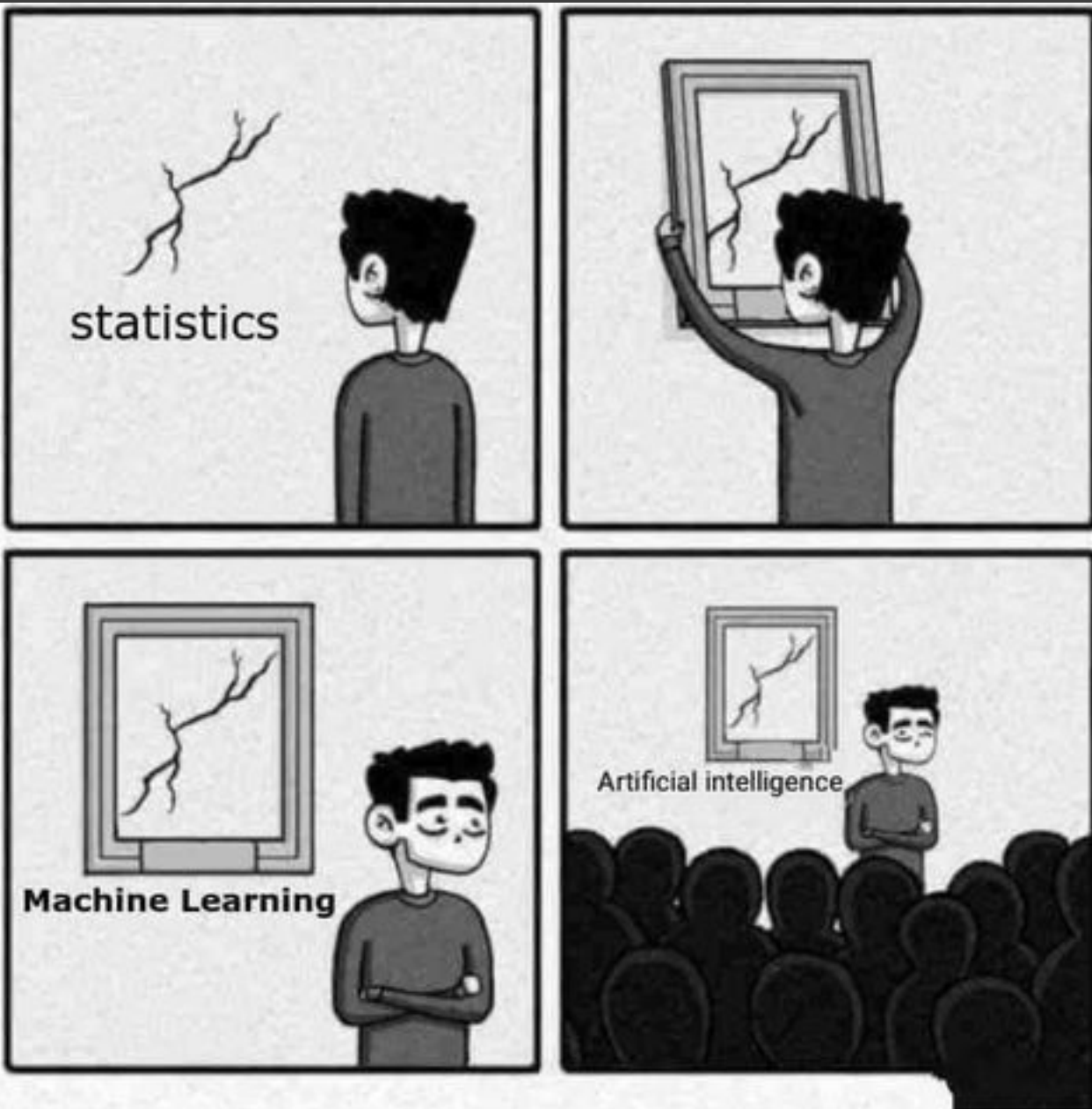
SE NÃO FALTAR NADA, SIM!!!



# Modelo de Aprendizagem



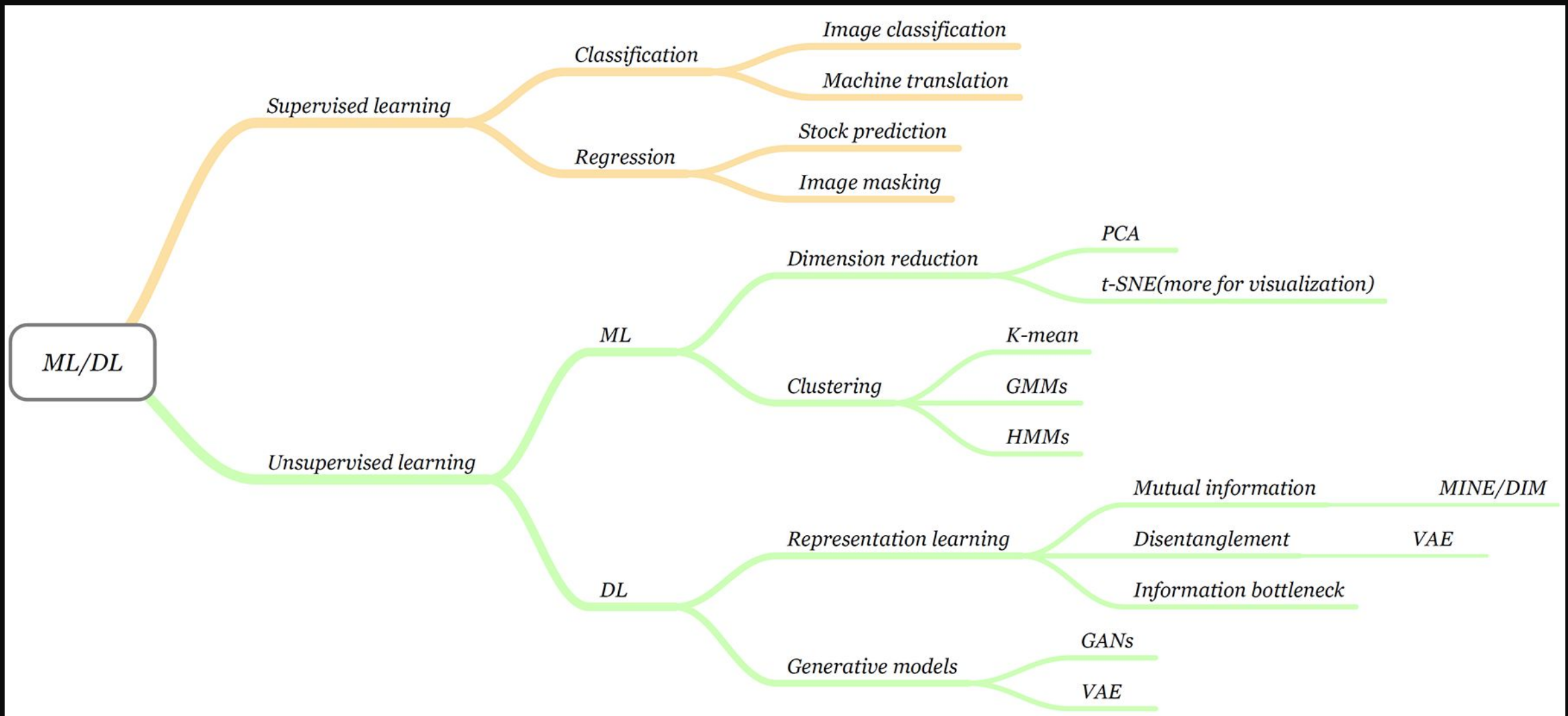
- É uma das técnicas utilizadas na Inteligência Artificial que usa algoritmos baseados em **matemática** e **estatística** para realizar a tarefa de aprendizado de máquina (*Machine Learning*).
- O *Machine Learning* surge da necessidade de processar e obter informação útil a partir dos dados e, uma vez que é inviável realizar manualmente o processamento e análise da grande quantidade de dados existentes atualmente.



*Hype* é o **exagero de algo**, ou em marketing uma estratégia para enfatizar alguma coisa, ideia ou um produto.

É um assunto que está dando o que falar, é algo que está na moda e que é comentado por todo mundo.

# Machine Learning

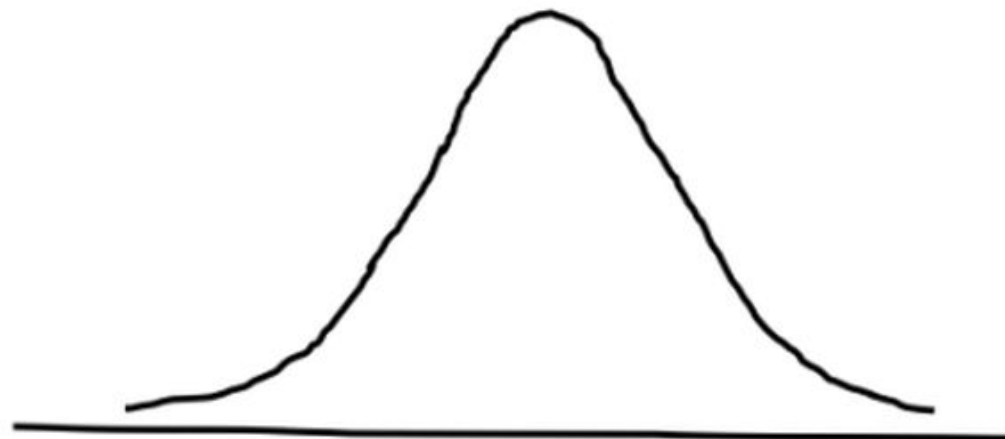




# Visualização dos Dados

- Consiste na representação gráfica de informações e dados.
- Usa elementos visuais, como diagramas, gráficos e mapas.
- A visualização de dados é uma forma acessível de ver e entender exceções, tendências e padrões nos dados.





Normal Distribution



Paranormal Distribution

# Tecnologias para Análise de dados

- **Python** - <https://www.python.org>  python™

- **Julia** - <https://julialang.org/>



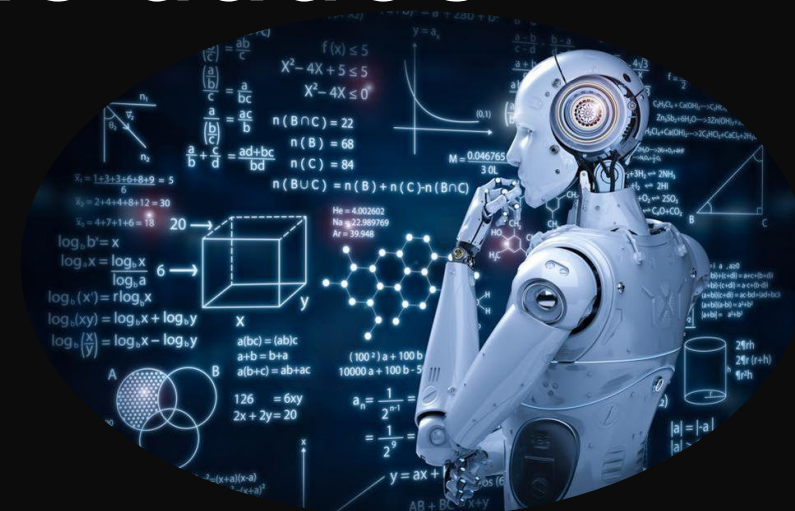
- **R** - <https://www.r-project.org>



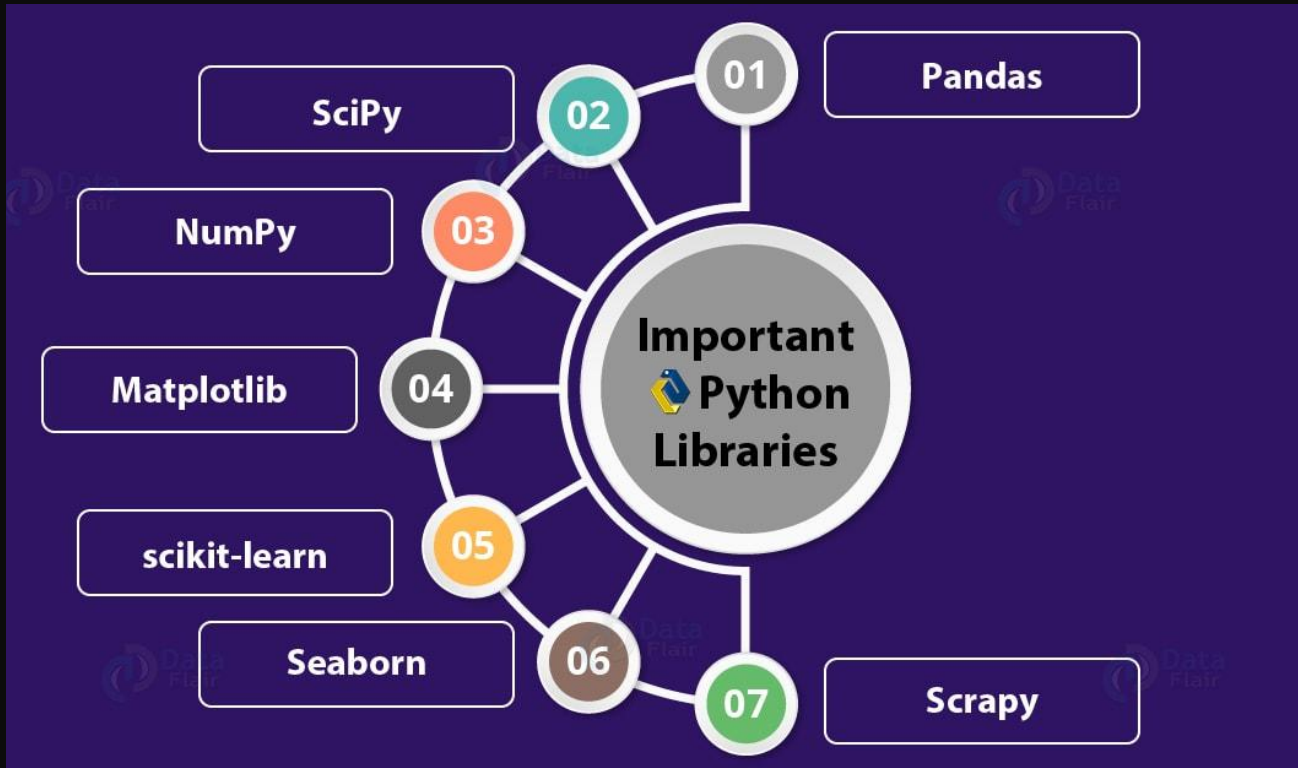
- **Tableau** (lê-se tablô) - <https://www.tableau.com>



- Entre outras.

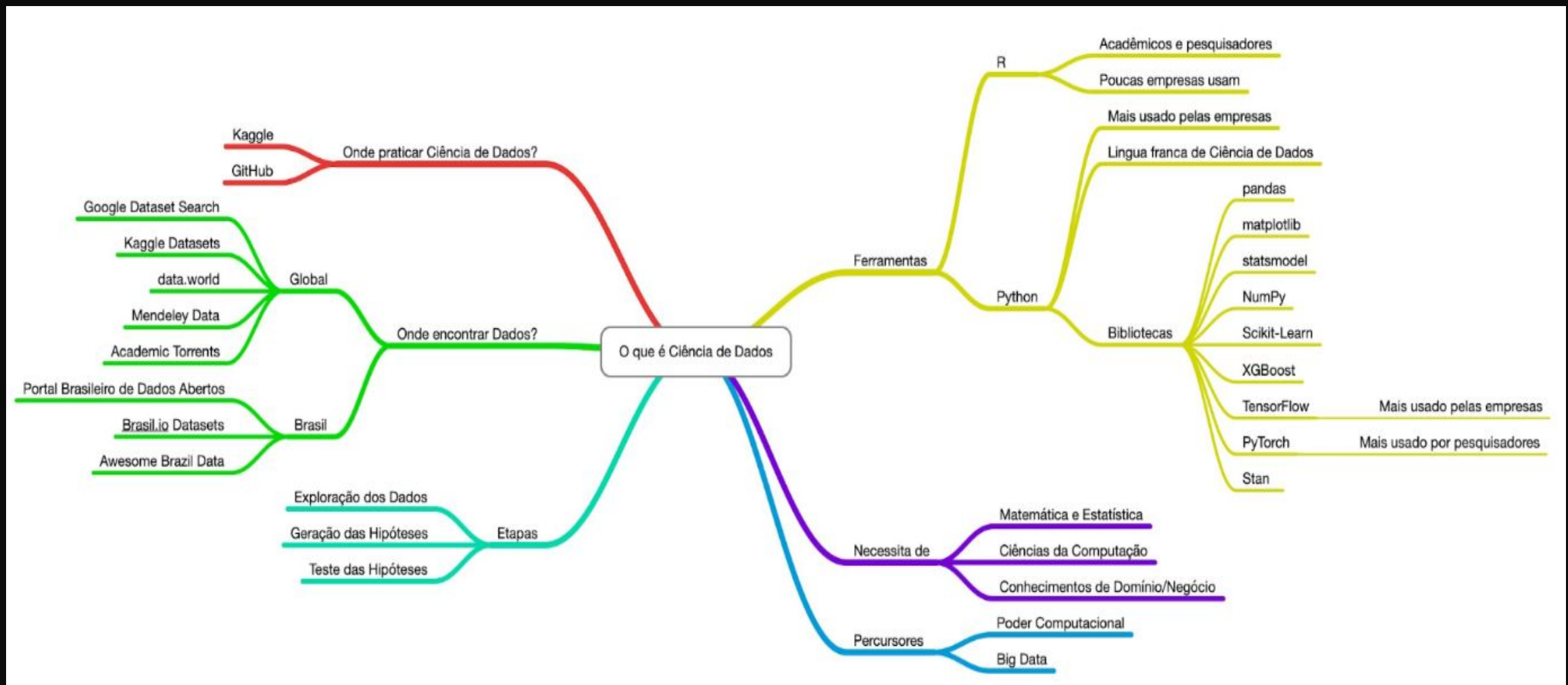


# Bibliotecas Python



As bibliotecas nos ajudam a implementar soluções sem muito esforço, e também evitando “reinventar a roda”.

# Mapa Mental sobre Ciência de Dados



# Bibliografia Recomendada

Acesso via Biblioteca Digital da Uninove

**Análise de Dados (Use a Cabeça) -**

<https://learning.oreilly.com/library/view/use-a-cabeca/9788576084686>

**Data Science Para Negócios -**

<https://learning.oreilly.com/library/view/data-science-para/9788576089728>

**Data Science do Zero -** <https://learning.oreilly.com/library/view/data-science-do/9788576089988>

**Learn Data Analysis with Python: Lessons in Coding -**

<https://learning.oreilly.com/library/view/learn-data-analysis/9781484234860/>

**Python for Data Analysis -** <https://learning.oreilly.com/library/view/python-for-data/9781491957653/>

**Python (Use a Cabeça) -** <https://learning.oreilly.com/library/view/use-a-cabeca/9788576087434>

# Referências

Campos, Claudinei José Gomes. "Método de análise de conteúdo: ferramenta para a análise de dados qualitativos no campo da saúde." *Revista brasileira de enfermagem* 57.5 (2004): 611-614.

Carpineto, Claudio, and Giovanni Romano. *Concept data analysis: Theory and applications*. John Wiley & Sons, 2004.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. *Applied survey data analysis*. CRC press, 2017.

Lattin, James, J. Douglas Carroll, and Paul E. Green. "Análise de dados multivariados." *São Paulo: Cengage Learning* 475 (2011).

Rice, John A. *Mathematical statistics and data analysis*. Nelson Education, 2006.





*That's all Folks!*