



UNINOVE
●●●●●

Ciência de Dados

Prof. Edson Melo de Souza

souzaem@uni9.pro.br



- Conteúdo da disciplina: <https://github.com/EdsonMSouza/covid19-2021-1>.
- Durante as aulas deixar o microfone desligado (*ligue somente para interagir*).
- Procurem não usar o *chat* para conversas paralelas.
- Links
 - <http://www.edsonmelo.com.br> (site)
 - <http://www.github.com/EdsonMSouza> (repositório de códigos)

Avaliações

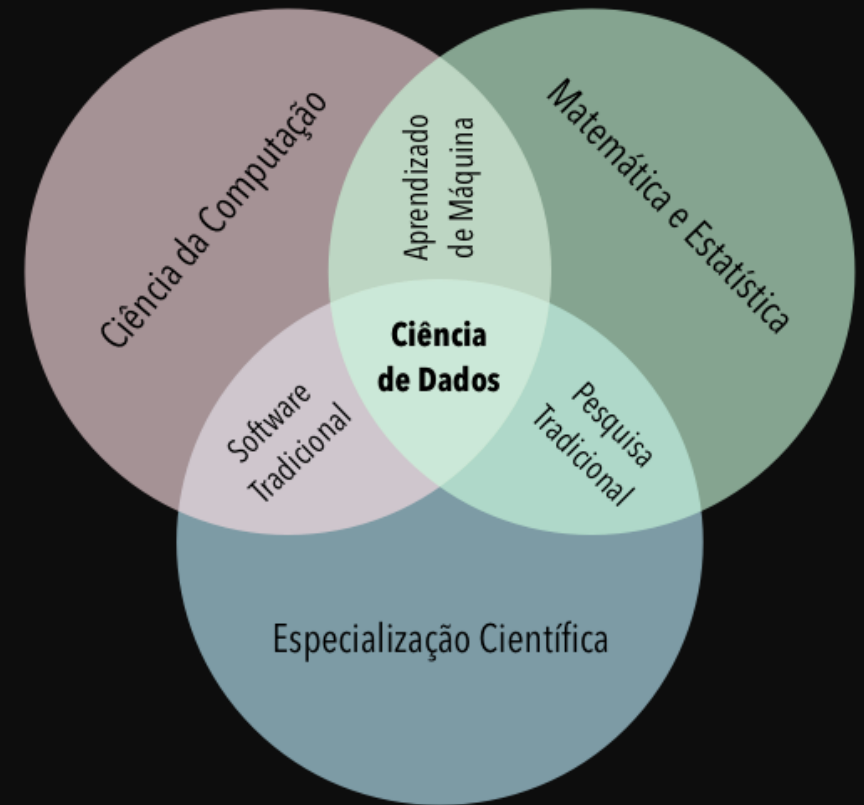
- Atividades digitais com questões de múltipla escolha e/ou implementação.
- Prova digital individual com 10 questões no final do semestre.
- Bônus aleatório durante as aulas (*máximo de um ponto no total*).

$$AV1 = \frac{Atividades + Prova}{2} + Bônus$$



O que é Ciência de Dados

- Do inglês *Data Science*, 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).



Tipos de Análise de Dados

- Descritiva;
- Preditiva;
- Prescritiva e;
- Diagnóstica.

Cada tipo permite trabalhar com os dados, buscando enxergar as informações neles contidas.

Análise Descritiva

- Este método é baseado em **uma** ou **várias questões** de pesquisa e não tem hipótese. Inclui a coleta de dados relacionados, depois organiza, tabula e descreve o resultado.
- Uma análise descritiva básica envolve o **cálculo de medidas simples de composição e distribuição de variáveis**.

Análise Preditiva

- É uma técnica analítica avançada que usa dados, algoritmos e *machine learning* para **antecipar** tendências e fazer projeções nos negócios.
- Busca analisar é **realizar projeções** mais sólidas, com um entendimento maior do que acontecerá no futuro.
- A partir disso, decisões mais adequadas são tomadas de acordo com as expectativas.

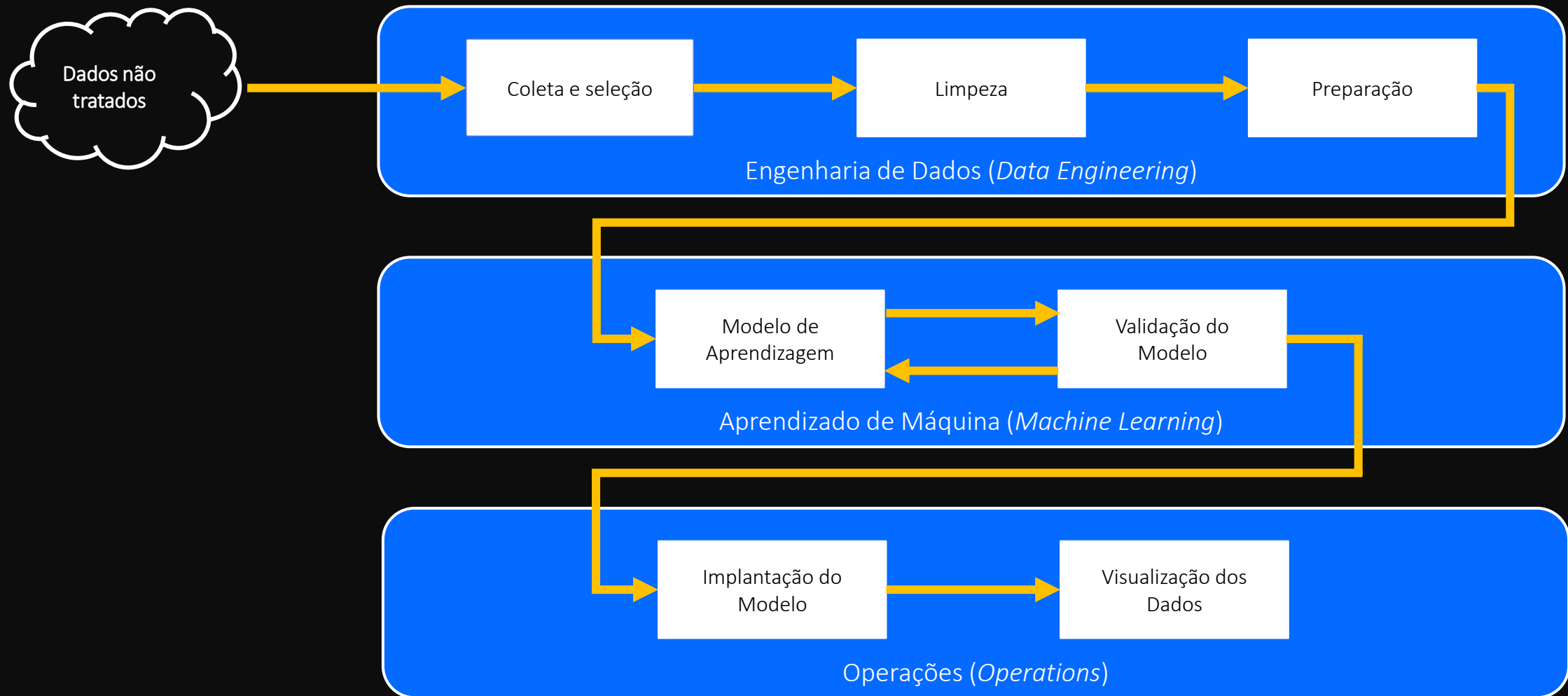
Análise Prescritiva

- É responsável por encontrar **uma solução** entre uma diversidade de variantes.
- Tem como objetivo otimizar recursos e aumentar a eficiência operacional.
- Essa ferramenta usa diferentes **técnicas de simulação e otimização** para indicar o caminho que realmente deve ser seguido.
- Por exemplo, uma ação de Marketing passa por uma análise prescritiva para que haja o entendimento de seu possível sucesso.

Análise Diagnóstica

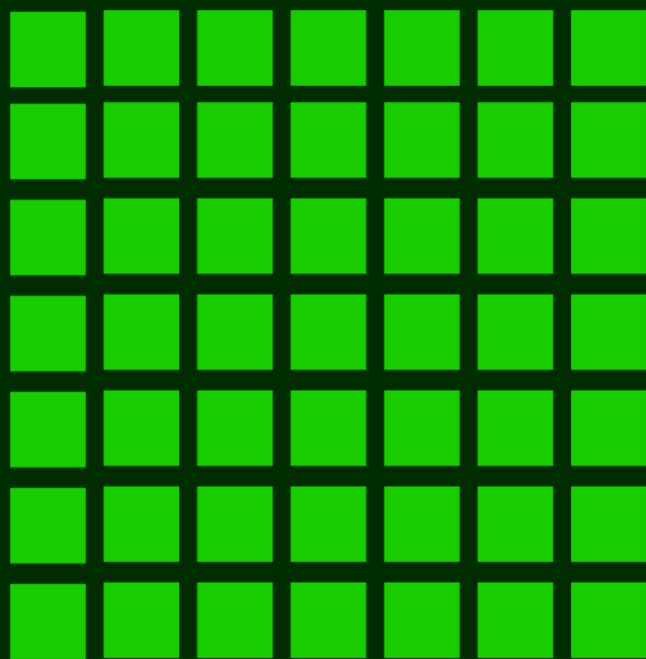
- Tem como finalidade compreender **o impacto de cada ação** dentro das estratégias concebidas.
- A proposta dessa análise é **destrinchar os dados** e concretizar todo o processo.
- Um exemplo é saber quem comprou, onde comprou, quando comprou, porque comprou ou, ainda, porque deixou de comprar.

Etapas da Análise de Dados

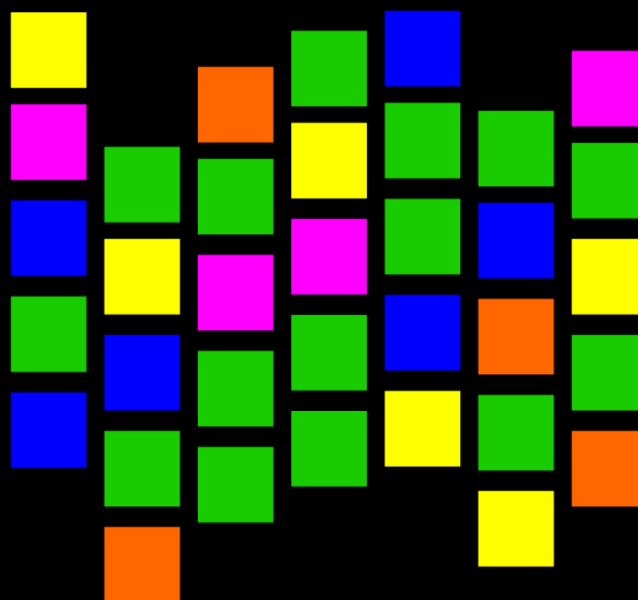


Tipos de Dados

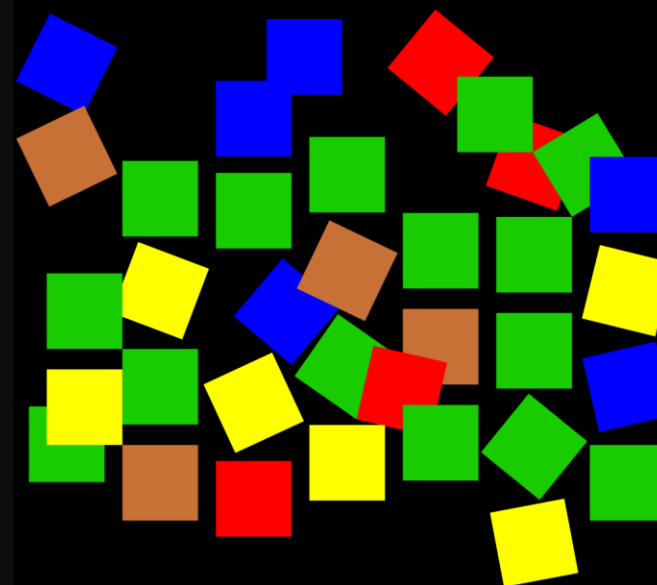
Dados Estruturados



Dados Semiestruturados

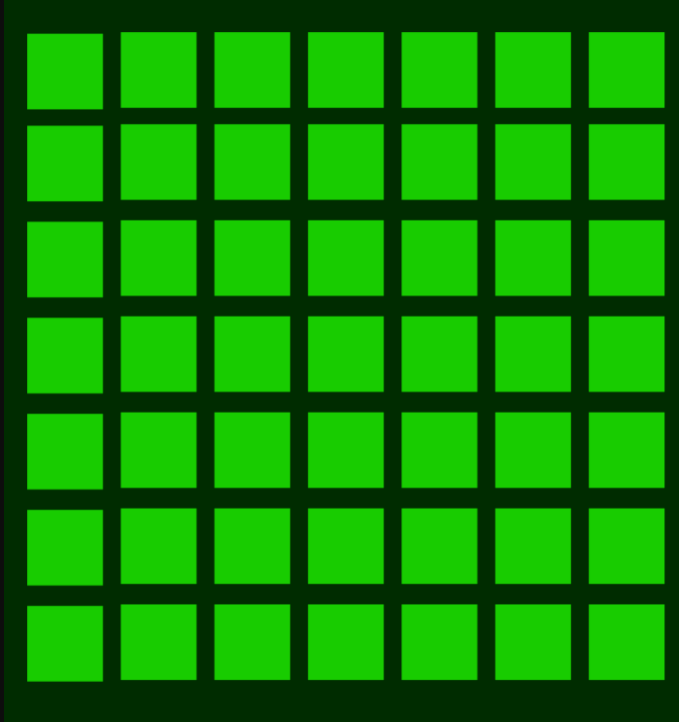


Dados Não estruturados



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.
- Exemplos:
 - XML;
 - JSON (*Javascript Object Notation*)
 - Bancos de dados relacionais.

Tipos de Dados



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

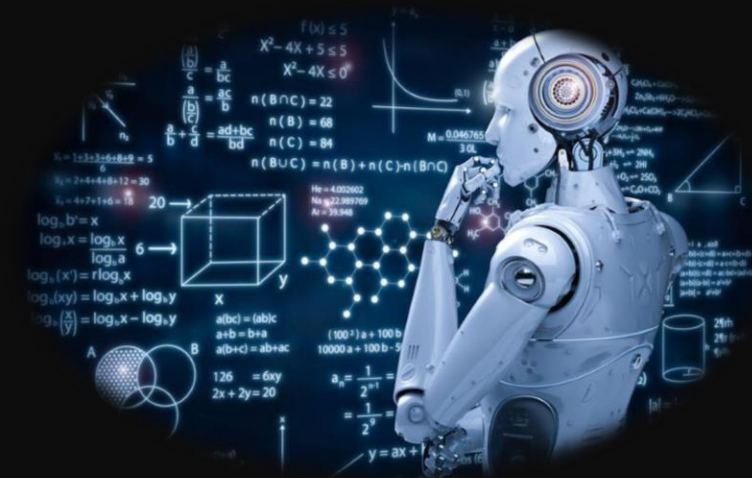
Limpeza de Dados ou *data cleaning*

- Nesta etapa são identificadas e corrigidas:
 - anomalias nos dados;
 - Dados faltantes;
 - Dados supérfluos ou desnecessários;
 - Inconsistências;
 - Entre outros.

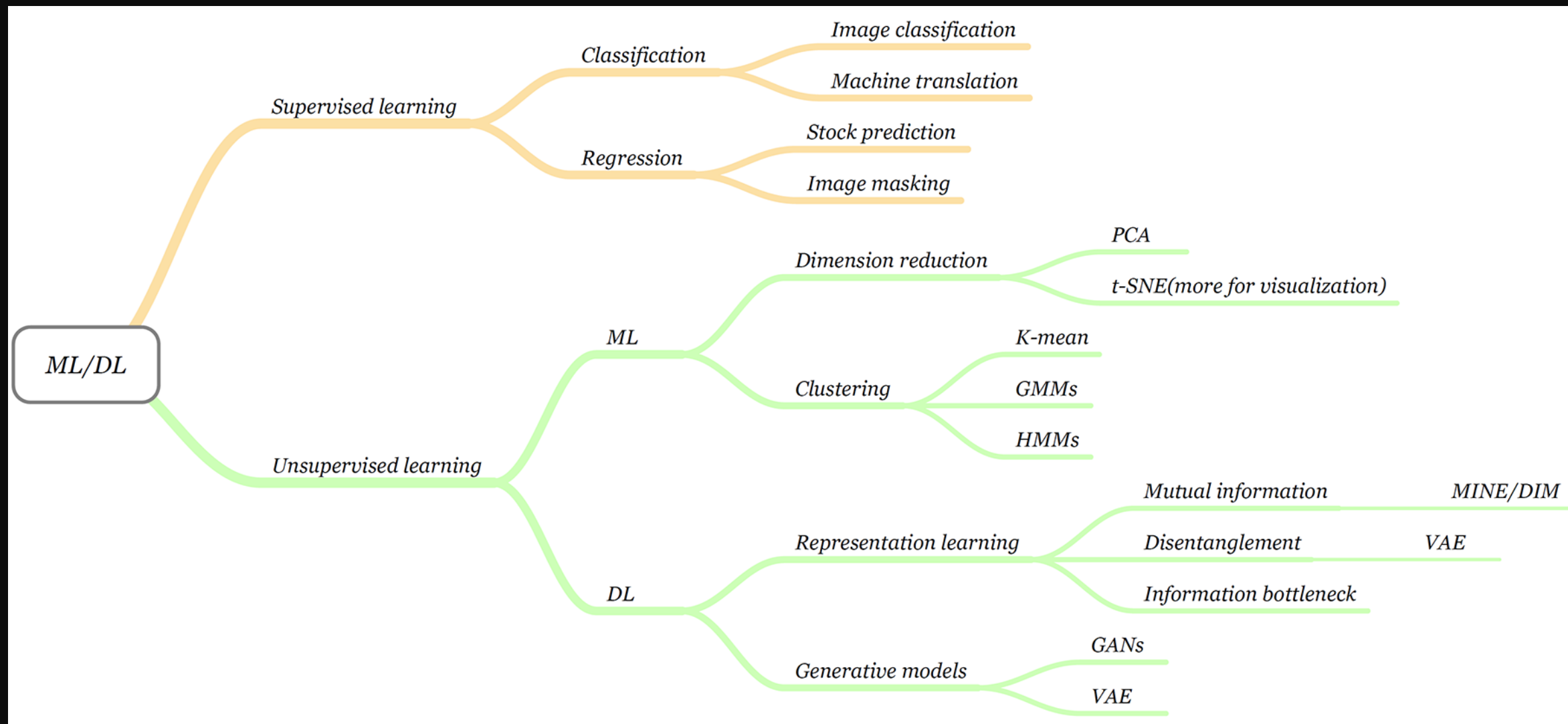


Modelo de Aprendizagem

- É uma das técnicas utilizadas na Inteligência Artificial que usa algoritmos baseados em **matemática** e **estatística** para realizar a tarefa de aprendizado.
- *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 que temos disponíveis atualmente, é necessário automatizar tarefas, simulando o comportamento humano.



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.



Tecnologias para Análise de dados

- SAS - https://www.sas.com/pt_br/trials.html



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



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



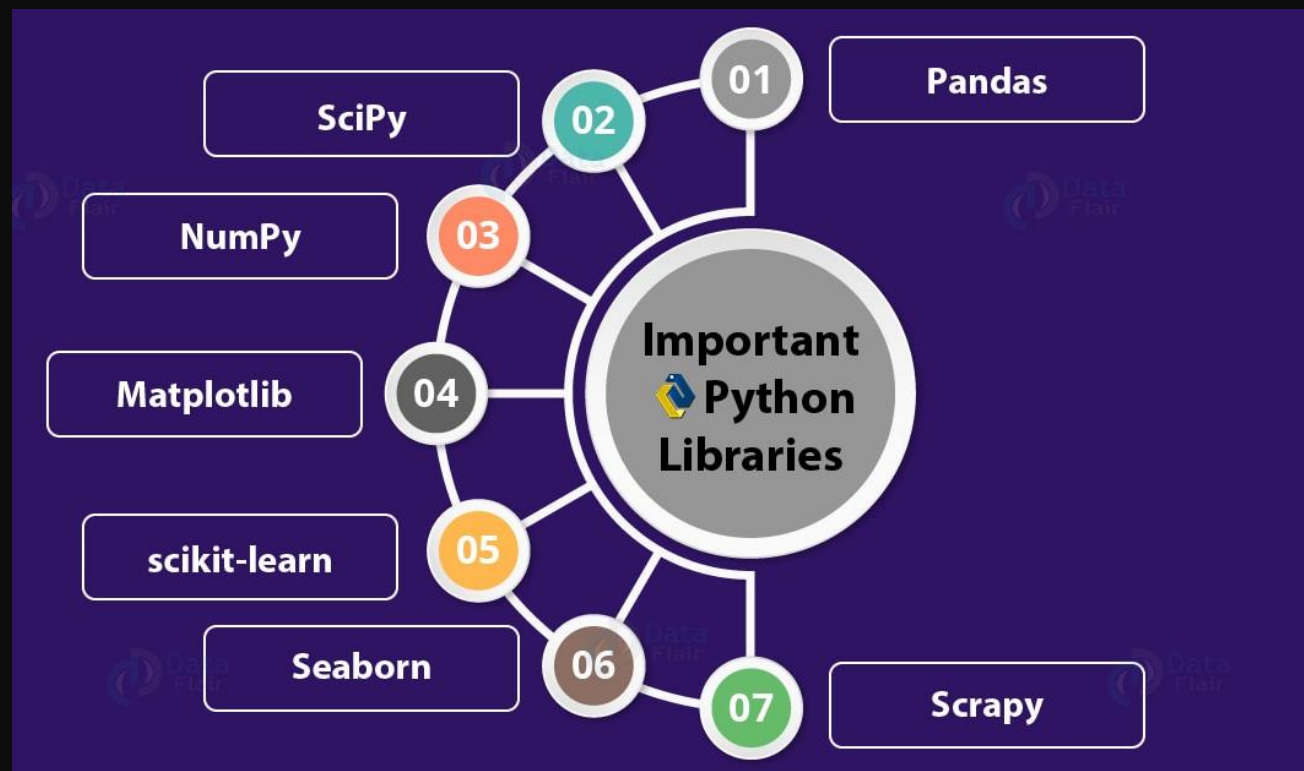
- Tableau (lê-se tablô) - <https://www.tableau.com>



- Entre outras.

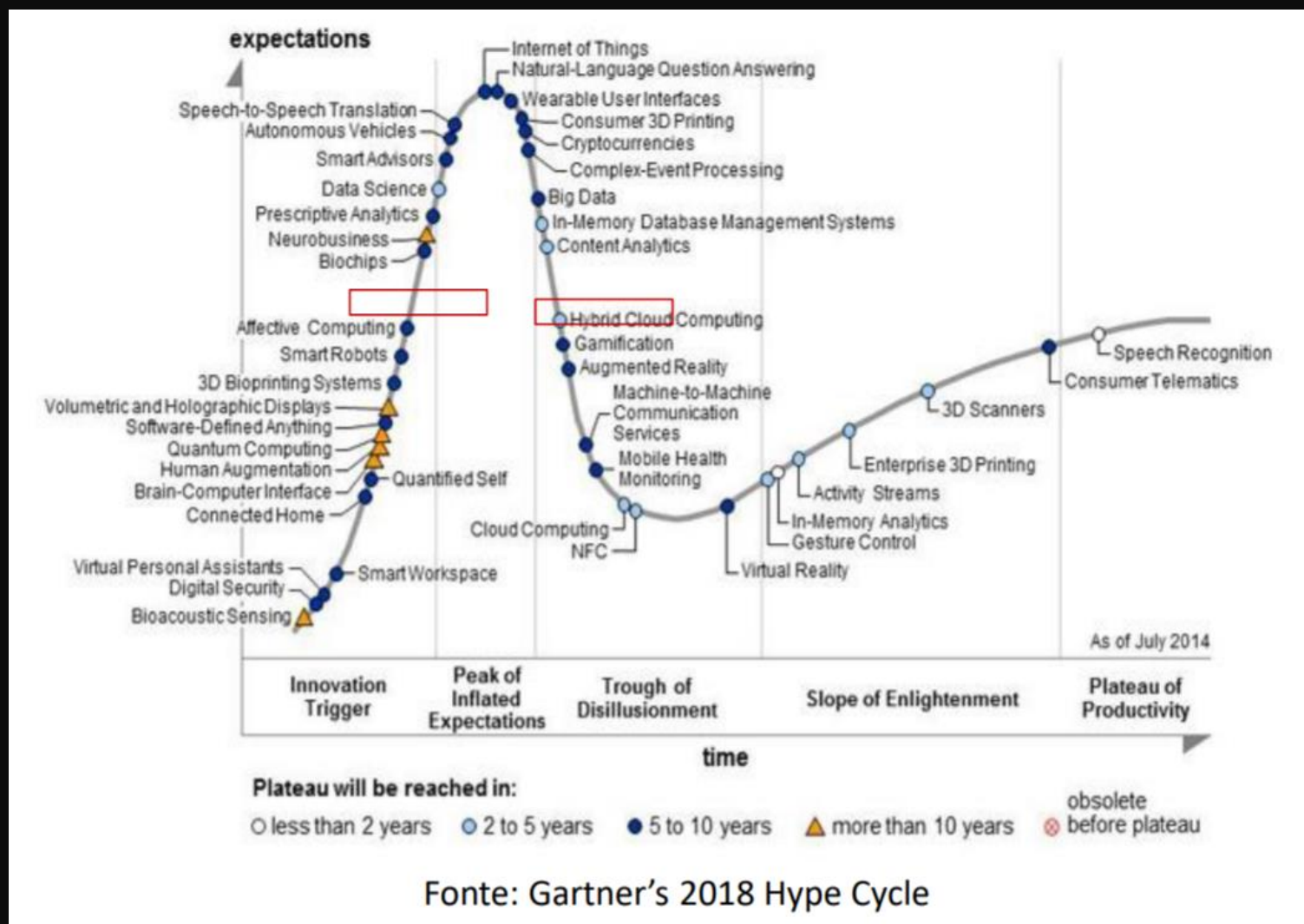


Bibliotecas Python



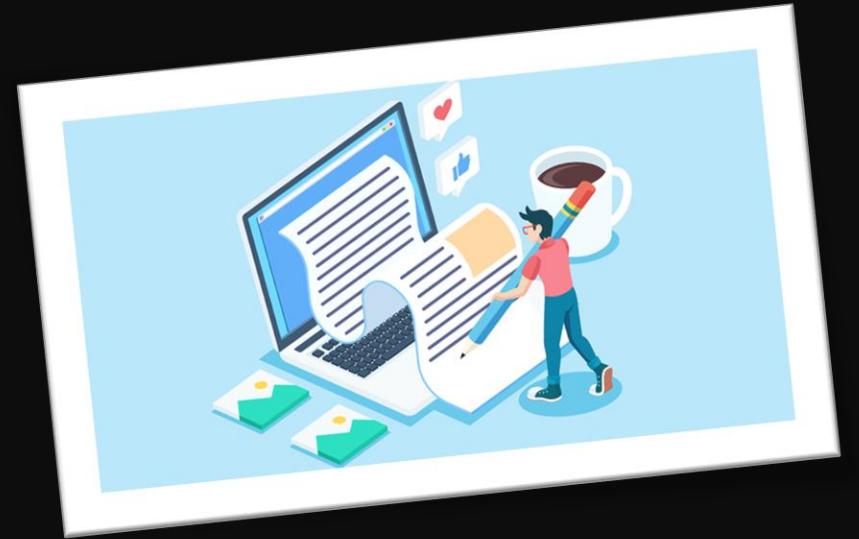
- As bibliotecas nos ajudam a implementar soluções sem muito esforço, e também evitando “reinventar a roda”.

Por que é tão interessante?

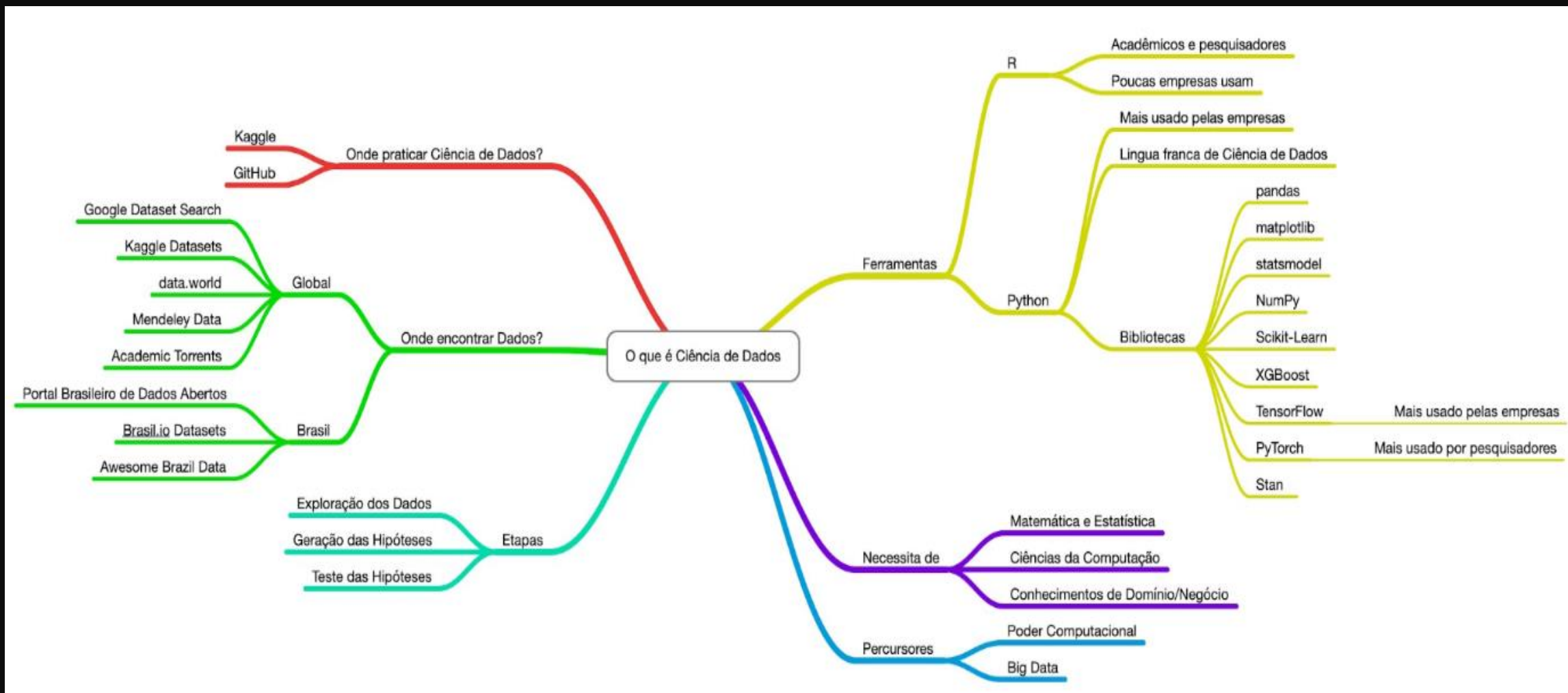


Conteúdo Programático

- Introdução ao Git (Github) e Google Colaboratory (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
- Análise de Dados com pandas
- Visualização de Dados com pandas, matplotlib e seaborn
- Machine Learning com Scikit-Learn (Regressão Linear e Logística)
- Web Scrap com requests e BeautifulSoup4



Mapa Mental



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.