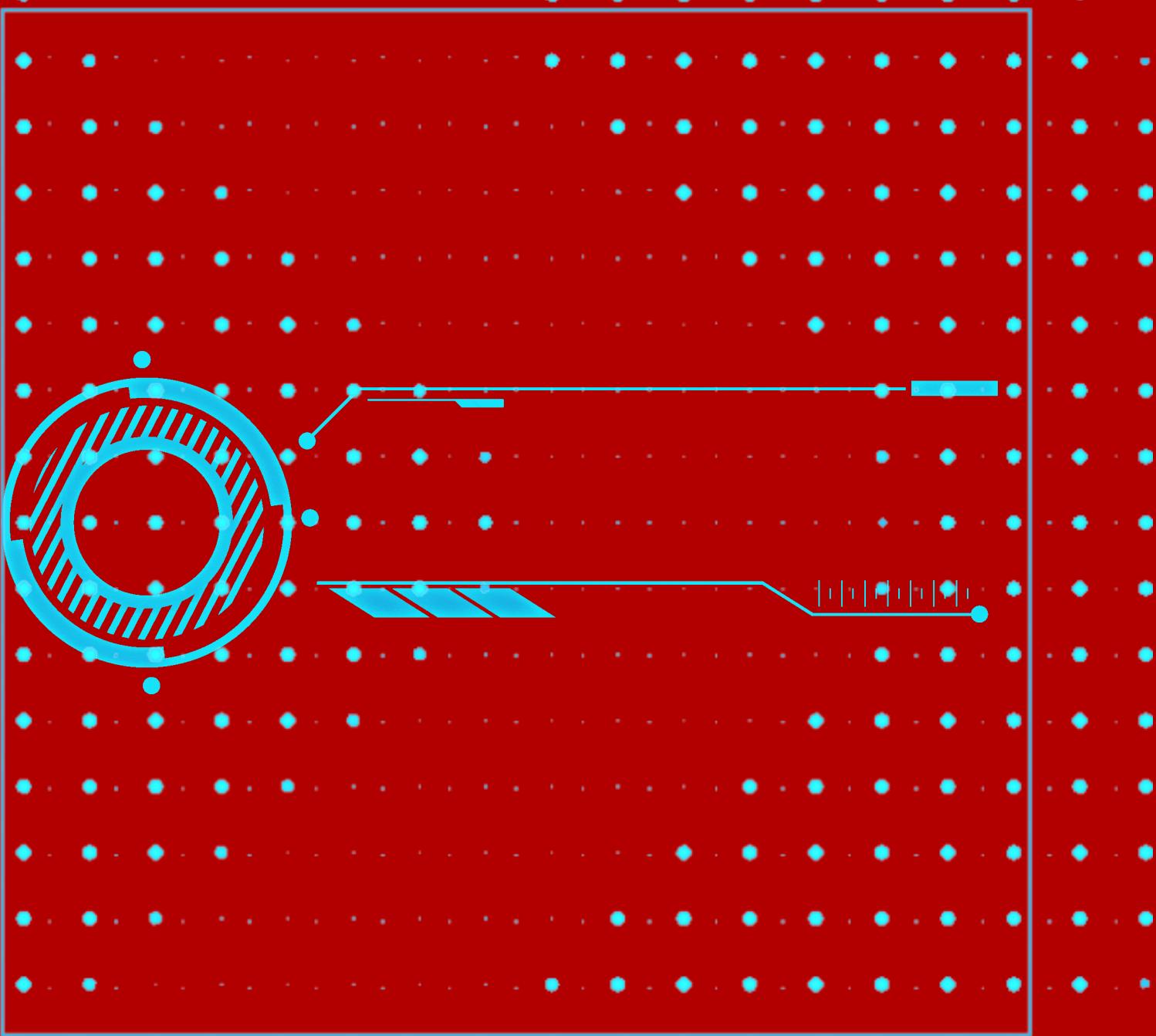




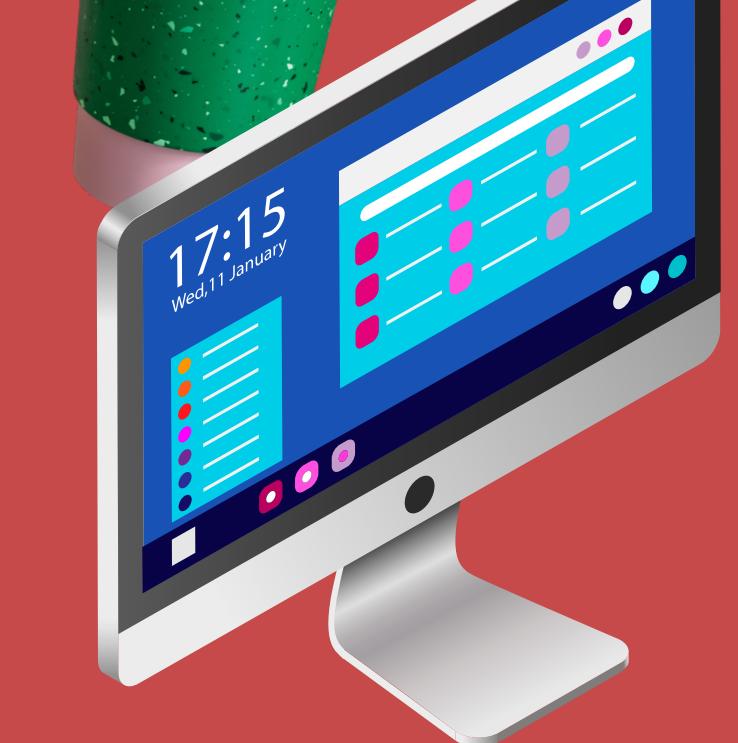
Dom Amaury Castanho

Introdução ao Data Science

Apresentado por: Prof. Ms. Ricardo Leme



Tópicos



Fatec
Itu
Dom Amaury Castanho

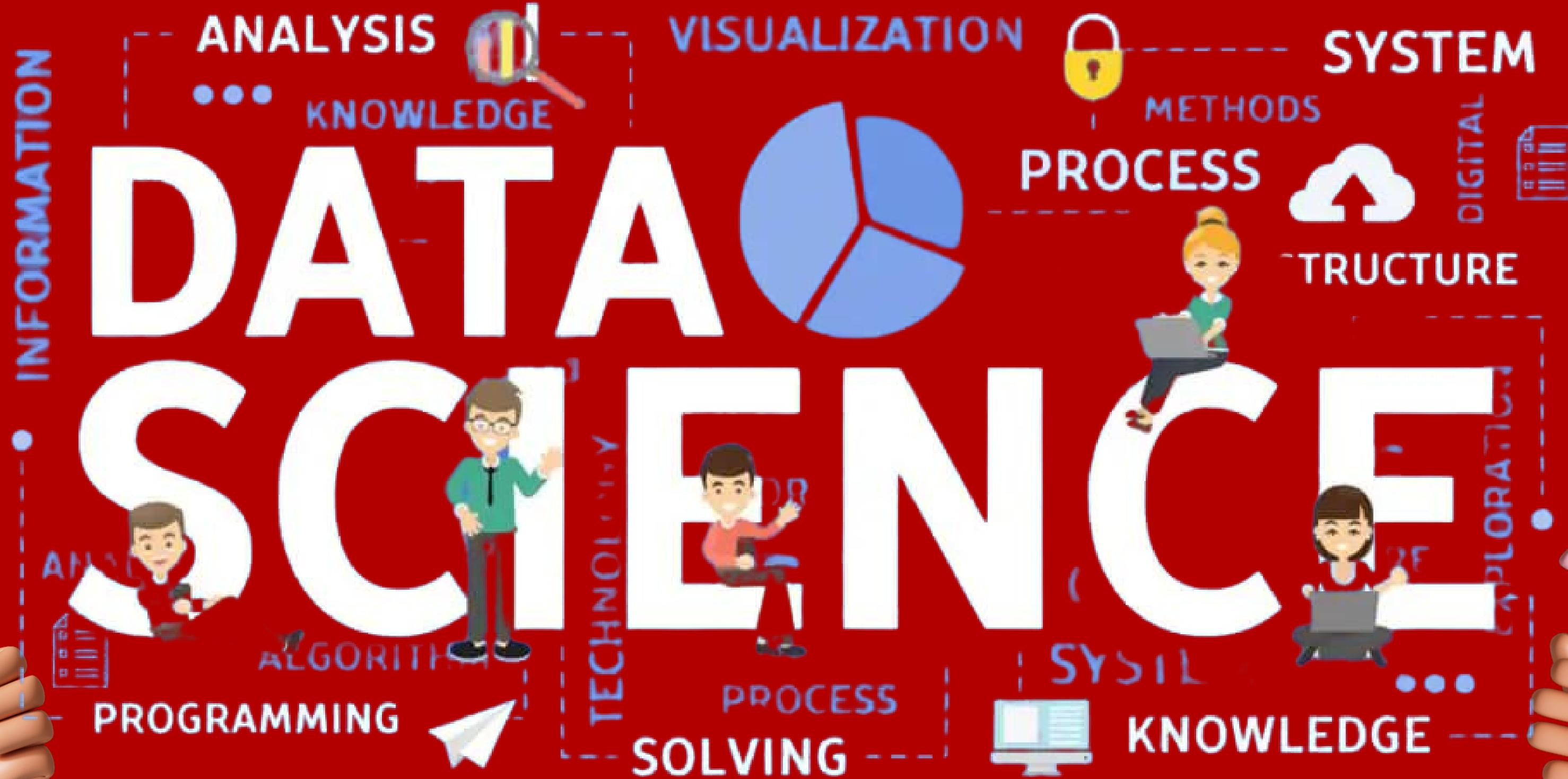
- Visão Geral do Data Science
- Google Collab
- Python para Data Science



Fatec

Itu

Dom Amaury Castanho



Valor do Dado



Data is the new oil. Clive Robert
Humby OBE - matemático e
empresário britânico no campo da
ciência de dados

Mudança do poder



Revolução Agrícola

Terra (até 1750)

Revolução Industrial

Meios de produção (até 1970)

Revolução Tecnológica

Informação (Século XXI)

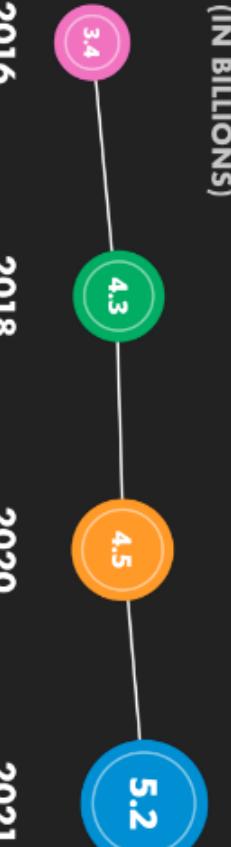
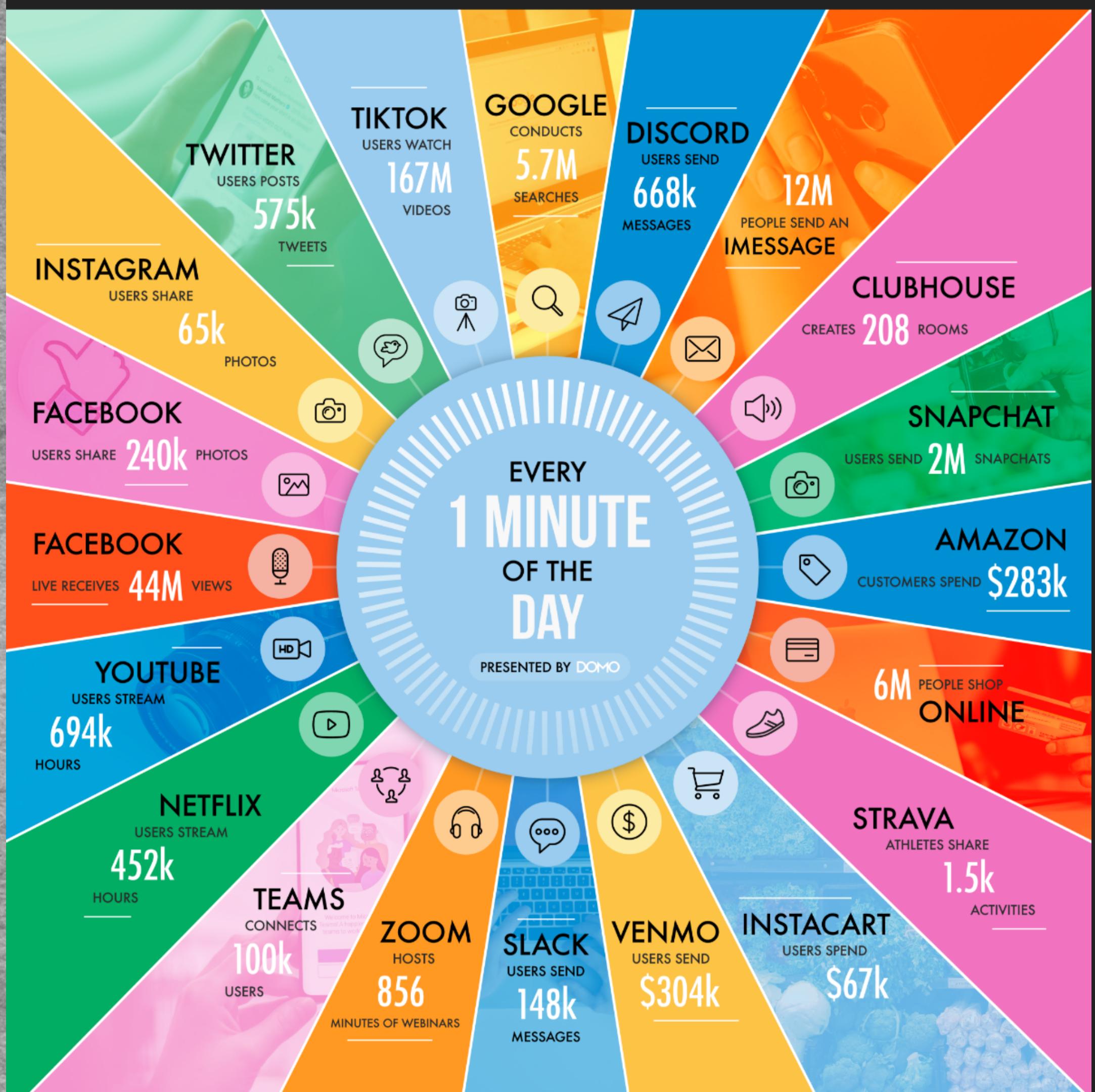


FatCat

Minuteman

Dom Amuny Castano

Ibu



As the world changes, businesses need to change too—and that requires data. Domo gives you the power to make data-driven decisions at any moment, on any device, so that you can make smart choices in a rapidly changing world. Every click, swipe, share, or like tells you something about your customers and what they want, and Domo is here to help you and your business make sense of all of it.

[Learn more at domo.com](https://www.domo.com)

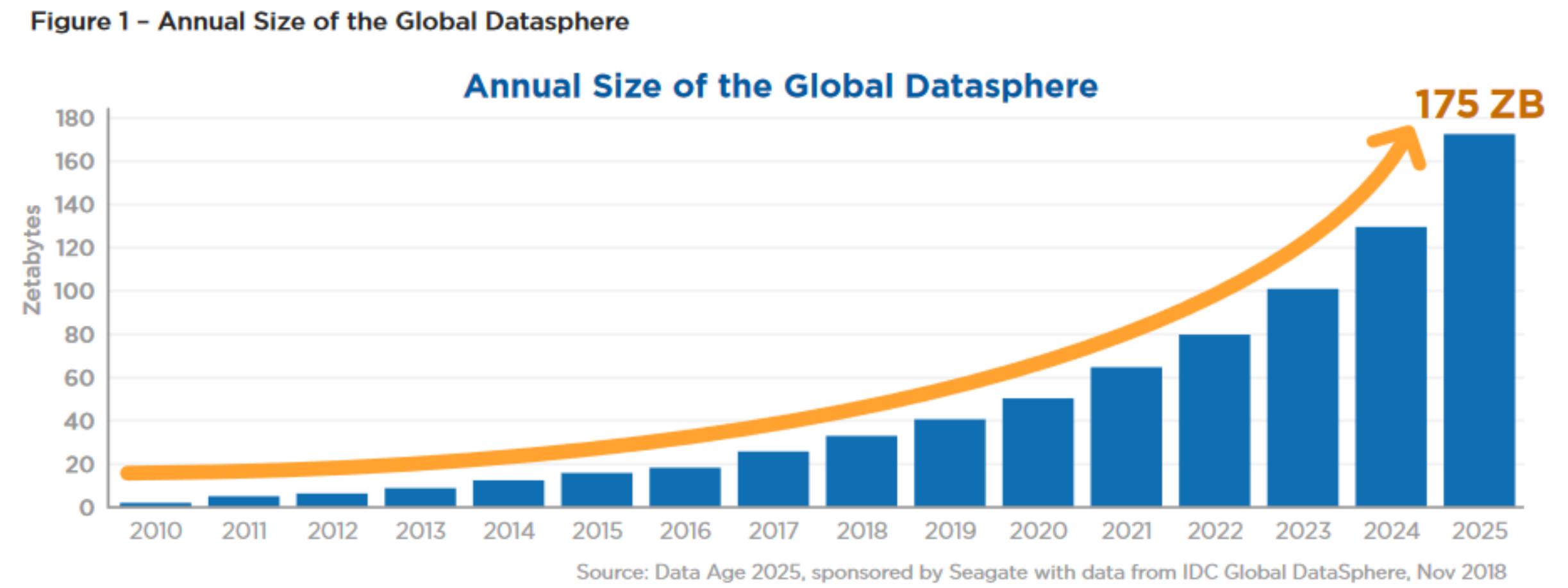
As of July 2021, the internet reaches 65% of the world's population and now represents 5.17 billion people—a 10% increase from January 2021. Of this total, 92.6 percent accessed the internet via mobile devices. According to Statista, the total amount of data consumed globally in 2021 was 79 zettabytes, an annual number projected to grow to over 180 zettabytes by 2025.

DOMO

SOURCES: LOCAL IQ, BUSINESS OF APPS, DUSTIN STOUT, HOOTSUITE, EXPANDED RAMBLINGS, INTERNET WORLD STATS, STATISTA, CNBC, BRANDWATCH, KILL THE CABLE BILL, YOUTUBE, KINSTA, THE VERGE, MANAGEMENT COMMUNICATION: A CASE ANALYSIS APPROACH, INTERNET LIVE STATS, SODA, STATISTA

Crescimento nos próximos anos

Estudo da Seagate mostra o crescimento exorbitante nos próximos anos

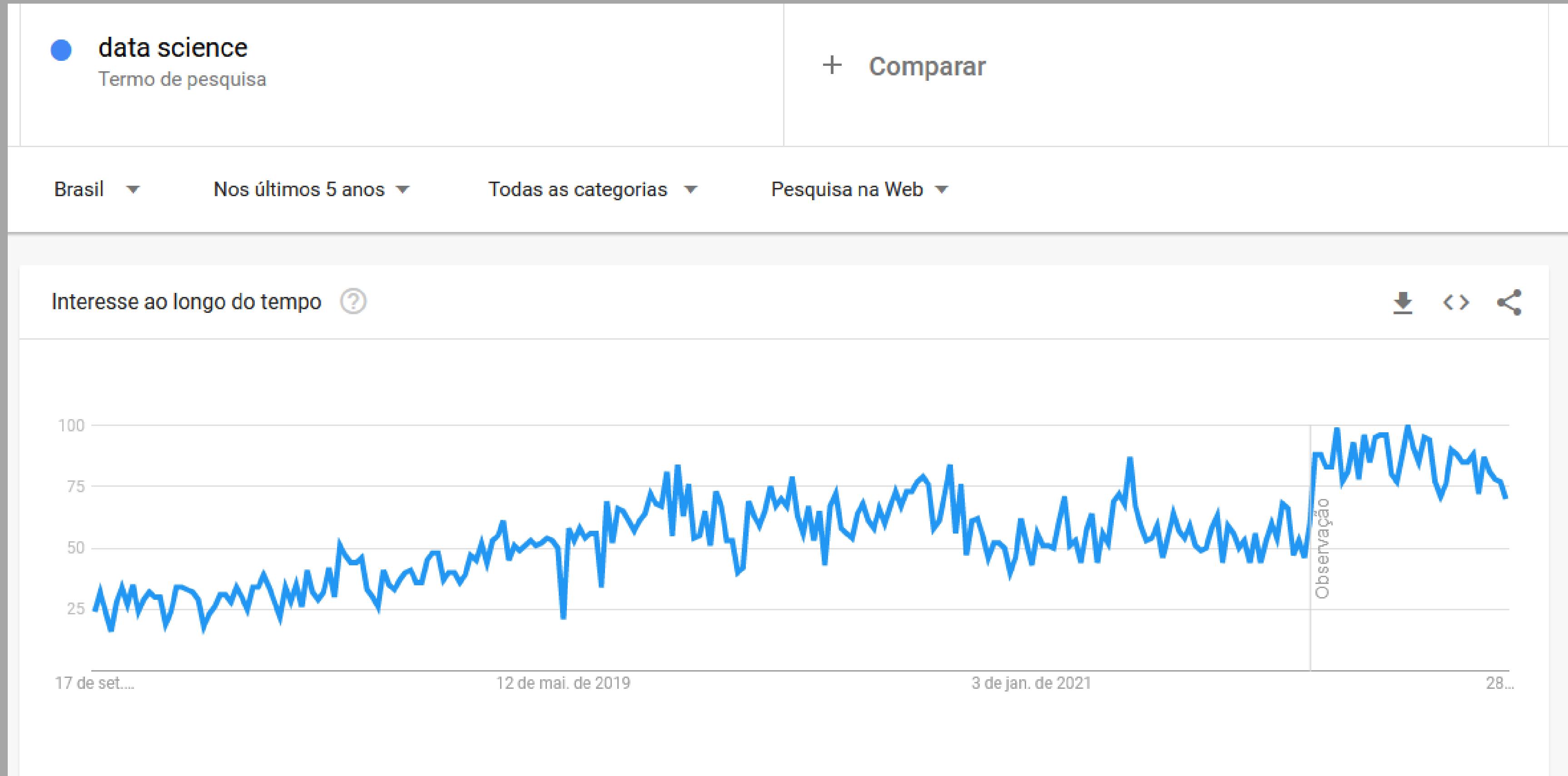


Fonte: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

O que é a Ciência de Dados?

É a prática de transformar dados brutos em insights de negócios utilizando métodos científicos.

Interesse pelo assunto



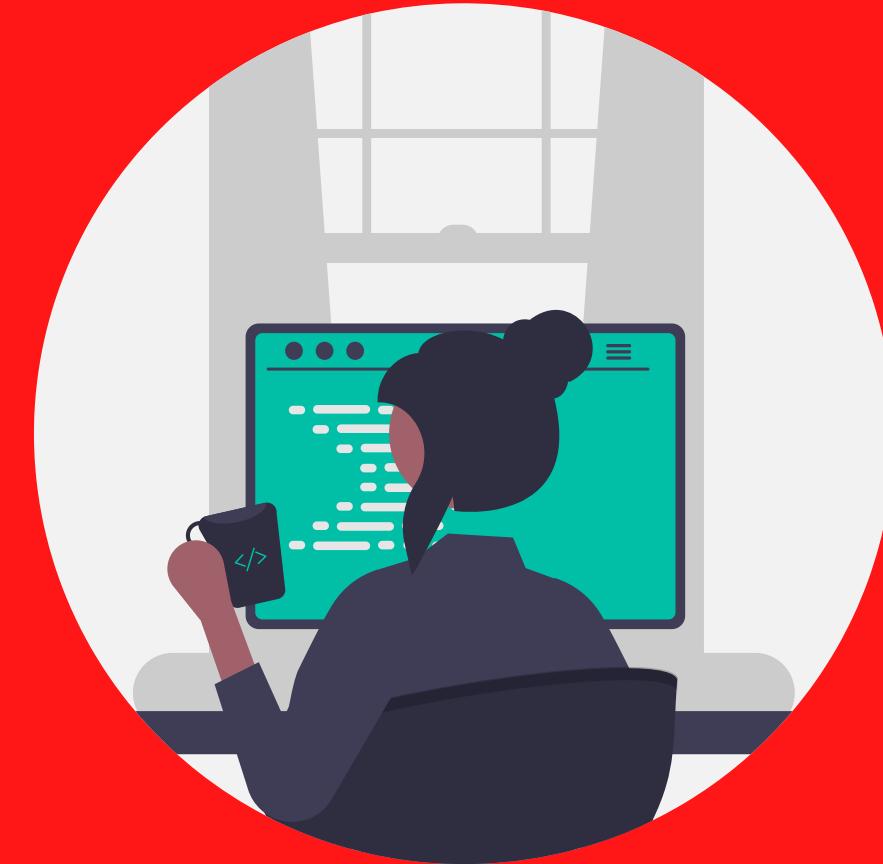


Sistemas de Recomendação

Análise de Crédito e Detecção de Fraudes

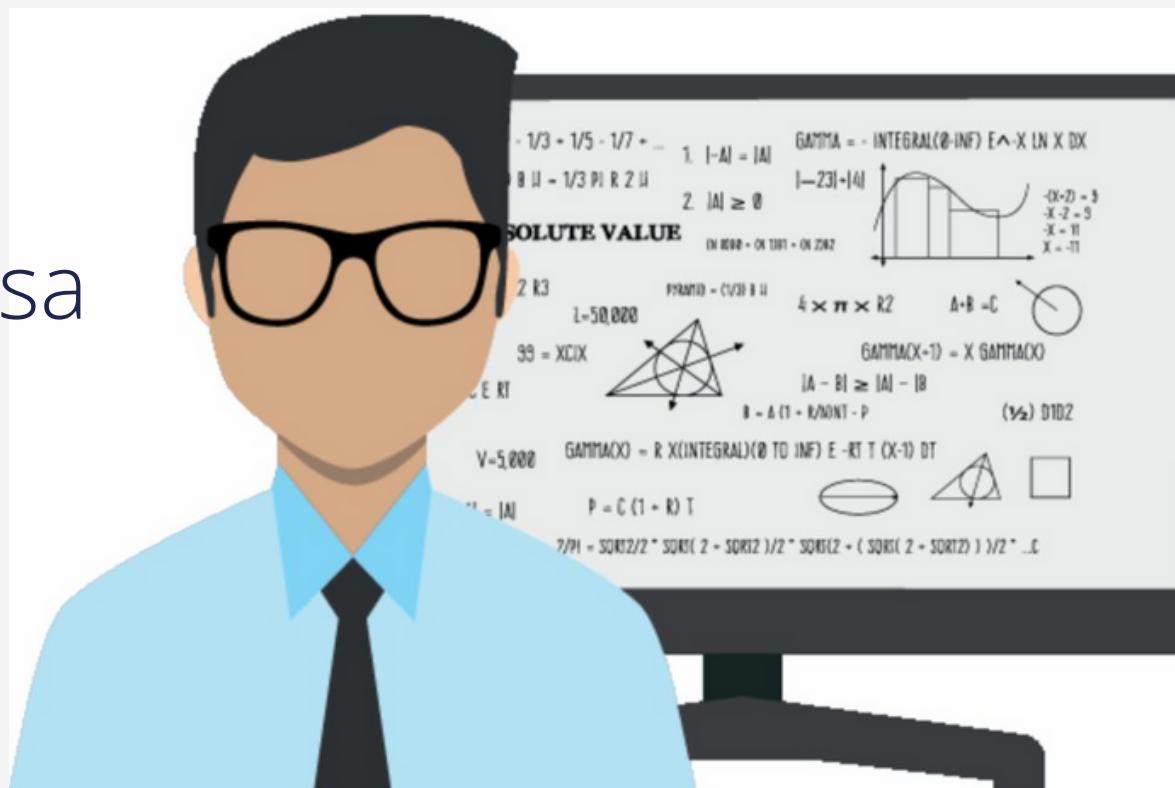
Marketing Direcionado

Logística



Função do Cientista de Dados

- Coletar grandes quantidades de dados
 - Transformar os dados brutos coletados em formatos adequados para análise e processamento
 - Resolver problemas relacionados aos negócios usando técnicas baseadas em dados
 - Implementação e teste de algoritmos de modelagem, classificação ou previsão ajustados aos requisitos do problema
 - Auxiliar na proposição de soluções para problemas de negócios graças ao conhecimento obtido a partir dos dados
 - Comunicar-se e colaborar com as áreas de TI e negócios
 - Procurar por ordem e padrões nos dados, além de detectar tendências que podem ajudar a base operacional de uma empresa
 - Apoio na elaboração de relatórios de estudos de mercado



Profissionais da Área

Engenheiro de Dados



- Coleta de dados
- Infraestrutura de armazenamento
- Qualidade dos dados

#sql, #nosql, #ETL #bigdata
#python #cloud

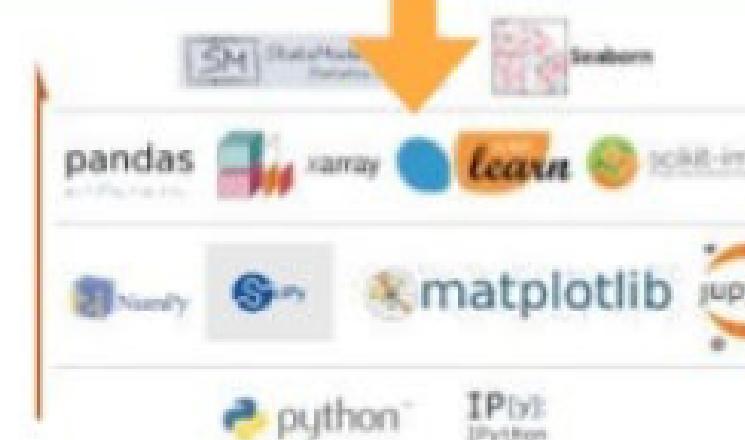


Cientista de Dados



- Análise estatística e ML
- Criação de Modelos
- Visualização dos resultados
- Geração de insights

#estatística #python #r
#machinelearning #dataviz



Engenheiro de Machine Learning



- Operacionalizar modelos
- Implementar modelos escaláveis
- Integração

#matemática #machinelearning
#cloud #MLops



Cientista de Dados – Análise Estatística

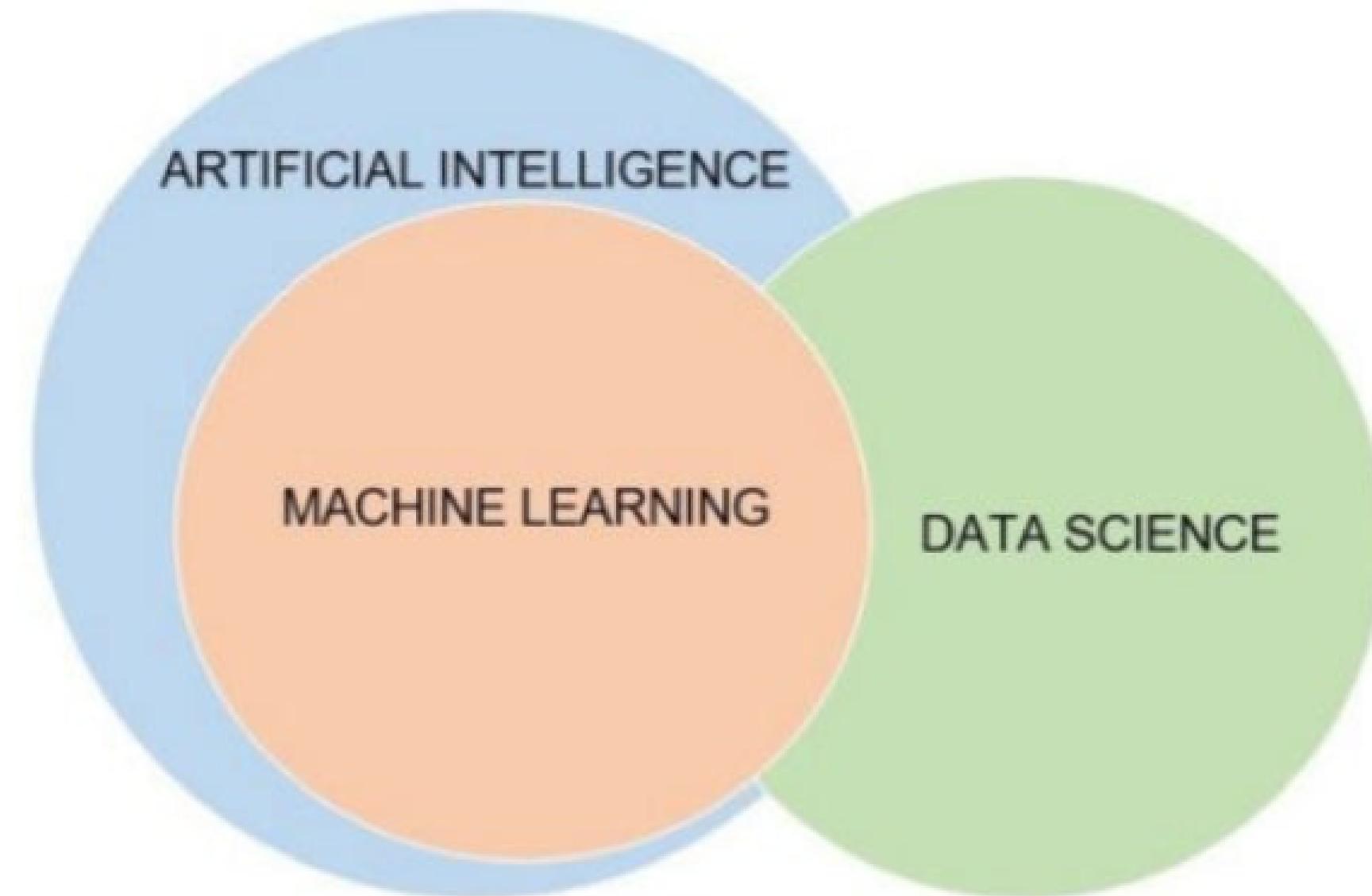
O cientista de dados não precisa saber tudo da área de Inteligência Artificial, apenas os métodos e algoritmos de extração de padrões nos dados (ML).

Cientista de Dados



- Análise estatística e ML
- Criação de Modelos
- Visualização dos resultados
- Geração de insights

#estatística #python #r
#machinelearning #dataviz



Criação de Modelos

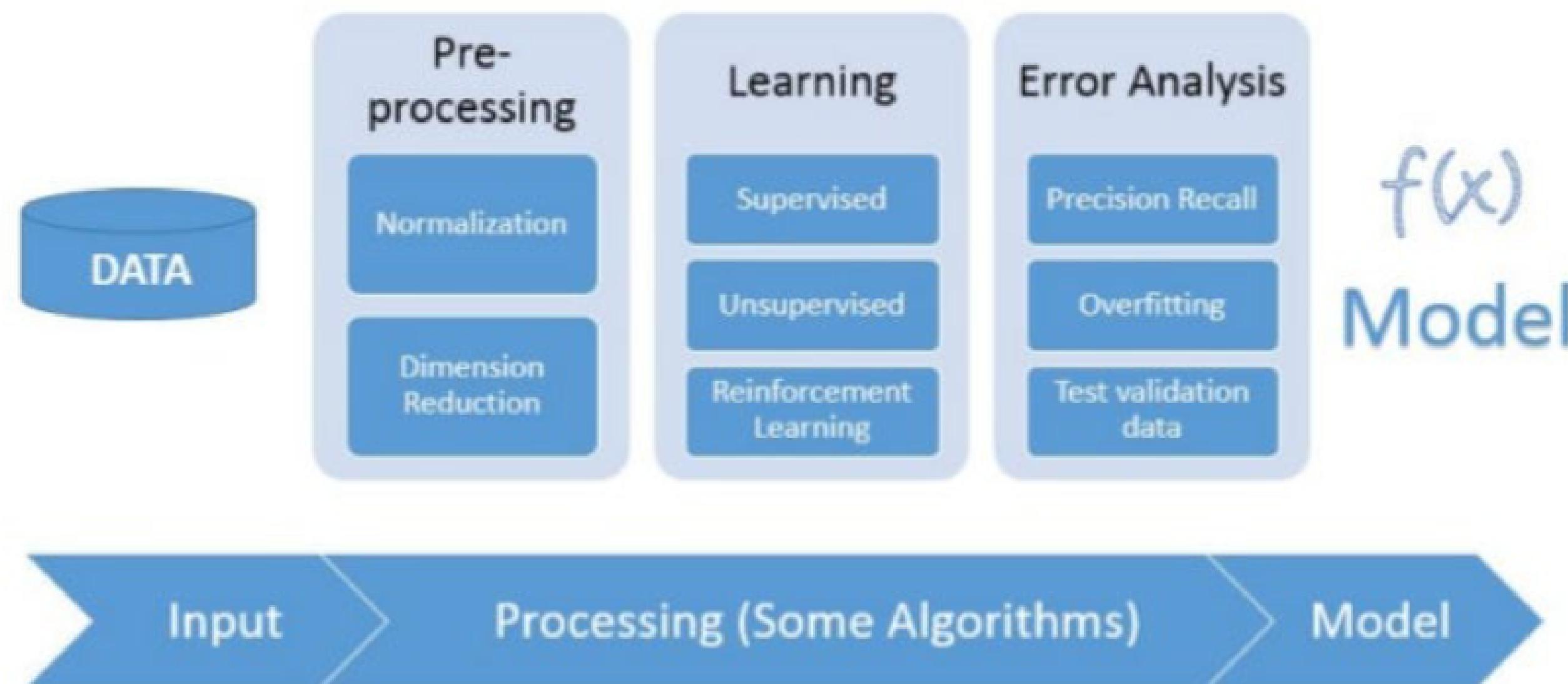
O modelo é a representação matemática dos dados, a partir do modelo é possível extrair os padrões nos dados para que esses padrões possam ser utilizados na resolução de problemas..

Cientista de Dados



- Análise estatística e ML
- **Criação de Modelos**
- Visualização dos resultados
- Geração de insights

#estatística #python #r
#machinelearning #dataviz



Visualização dos Resultados

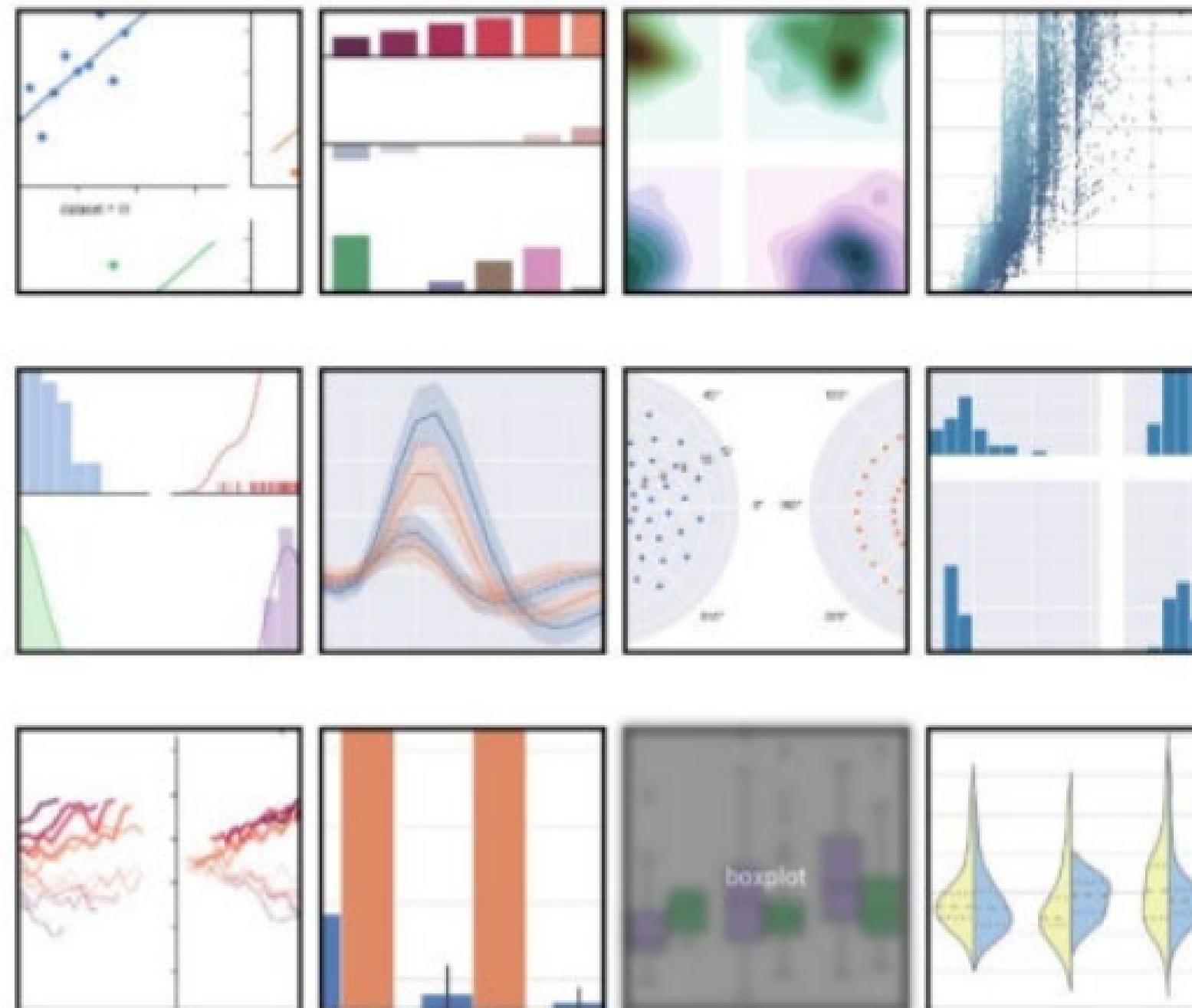
A visualização da informação **facilita o entendimento** e é pré-requisito para a **tomada de decisão** em qualquer projeto

Cientista de Dados



- Análise estatística e ML
- Transformação dos dados
- **Visualização dos resultados**
- Geração de insights

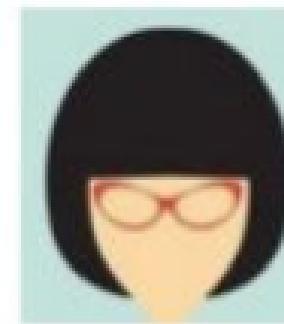
#estatística #python #r
#machinelearning #dataviz



Coleta, Limpeza e Organização

Um cientista de dados gasta aproximadamente ~80% do tempo de uma análise coletando, limpando e organizando dados.

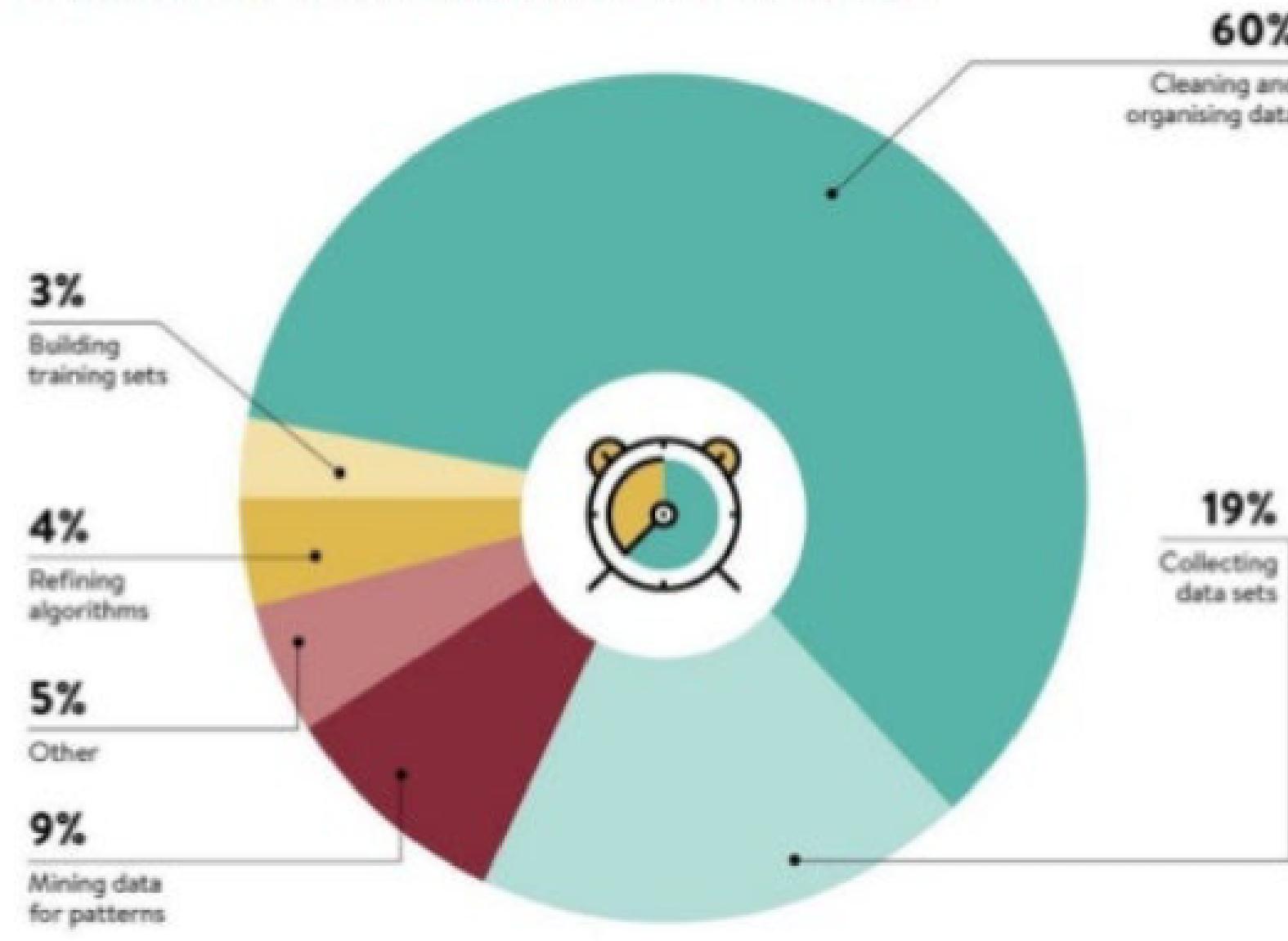
Cientista de Dados



- Análise estatística e ML
- Criação de Modelos
- Visualização dos resultados
- Geração de insights

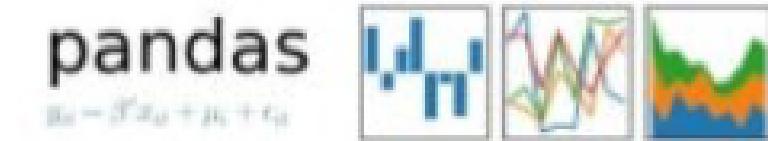
#estatística #python #r
#machinelearning #dataviz

WHAT DATA SCIENTISTS SPEND THE MOST TIME DOING



Source: CrowdFlower 2016

Dicas para iniciantes na área

- **Estude estatística e cálculo.** Não existe análise de dados sem conhecimento em estatística e não existe Machine Learning sem cálculo.
- Aprenda uma **linguagem de programação** voltada à análise de dados e machine learning **como R ou Python**.
- **Aprenda a usar** bem pelo menos uma biblioteca de machine learning **como scikit-learn, spark, H2O...**
- Utilize plataformas como **Kaggle**, fóruns e **comunidades** para trocar experiência e aprender análises públicas.



Faça parte da maior comunidade de Data Science do Brasil!

Cresça junto com profissionais de data science, machine learning, big data e inteligência artificial

Inscreve-se na newsletter quinzenal

Digite seu melhor email

Inscreve-se



26 de agosto de 2022

Democratização de Dados e AI no Grupo Boticário - Data Hackers Podcast 57

Um ano depois desde nosso primeiro episódio juntos, o pessoal do Grupo Boticário volta ao podcast do Data Hackers para falar o que aconteceu nesses últimos meses. Philipp Nguyen (Chief Data Officer), Jordy Rocha (Lead ML Engineer), e Thamires Teixeira (Analista de

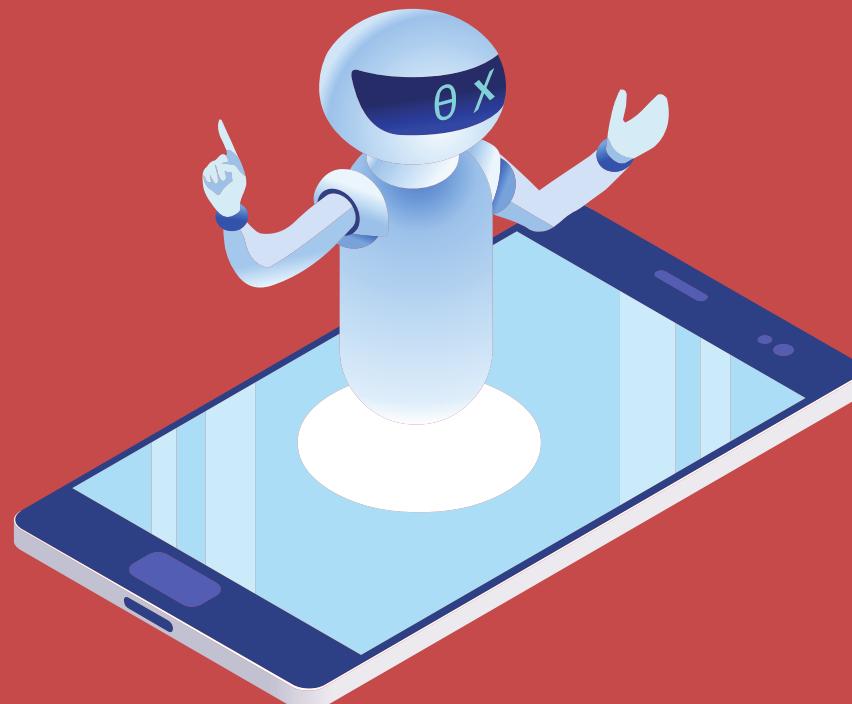
Ver mais

Ouvir agora



<https://www.datahackers.com.br/>

Posicionamento Agressivo da Google no suporte a IA



Machine Learning

- Biblioteca TensorFlow é *open-source*.
- Ela permite a criação e treinamento de redes neurais para detectar e decifrar padrões e correlações, análogo à forma como humanos aprendem e raciocinam
- Introdução, acesse: <https://youtu.be/t5z5lyrb-7s>



Data Science

- Criação do Google Colab que é totalmente baseado no Jupyter notebook.
- Permite escrever código Python no seu próprio navegador
- Fácil de compartilhar
- Com acesso gratuito ao processamento de GPU
- Suporta várias bibliotecas adicionais

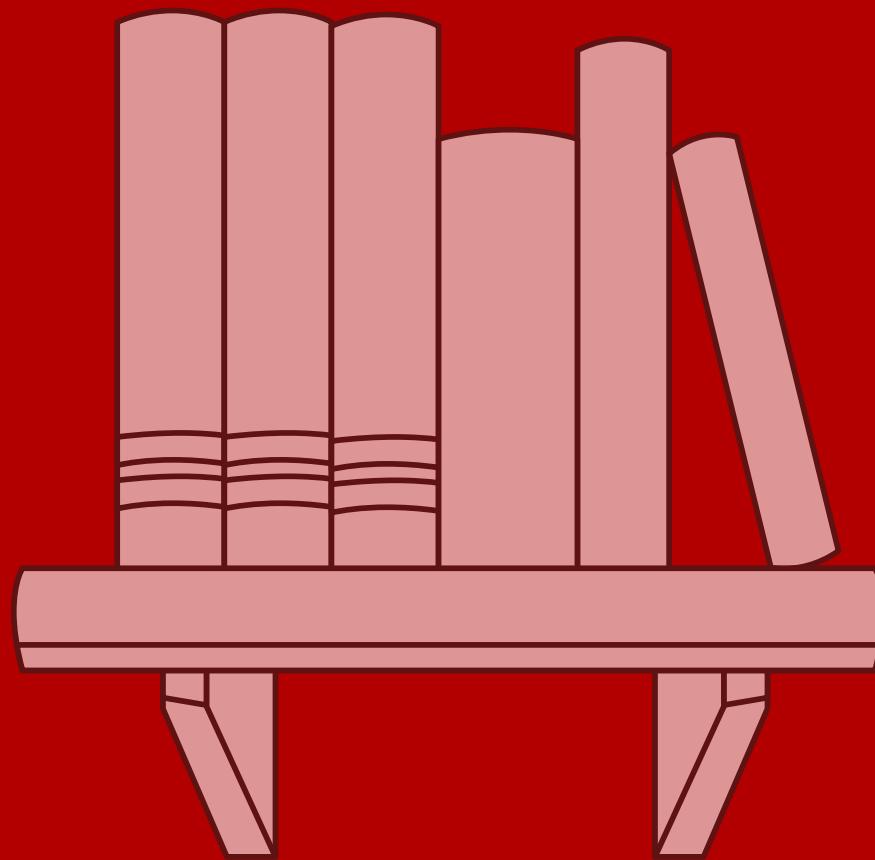


Referências

DOMO. Data Never Sleeps 9.0. Disponível em:

<https://www.domo.com/learn/infographic/data-never-sleeps-9>. Acesso em:
13 set. 2022.

SANTANA, Marlesson. Ciência de Dados: A revolução na tomada de decisões. Disponível em: <https://www.datahackers.com.br/blog>. Acesso em:
13 set. 2022.



A close-up photograph showing a person's hands kneading several pieces of dough on a light-colored, floured work surface. The hands are covered in flour, and the dough is being shaped into long, cylindrical rolls. In the background, there are more pieces of dough and some blue and green containers. The lighting is bright, highlighting the texture of the dough and the flour.

Mão na
Massa!