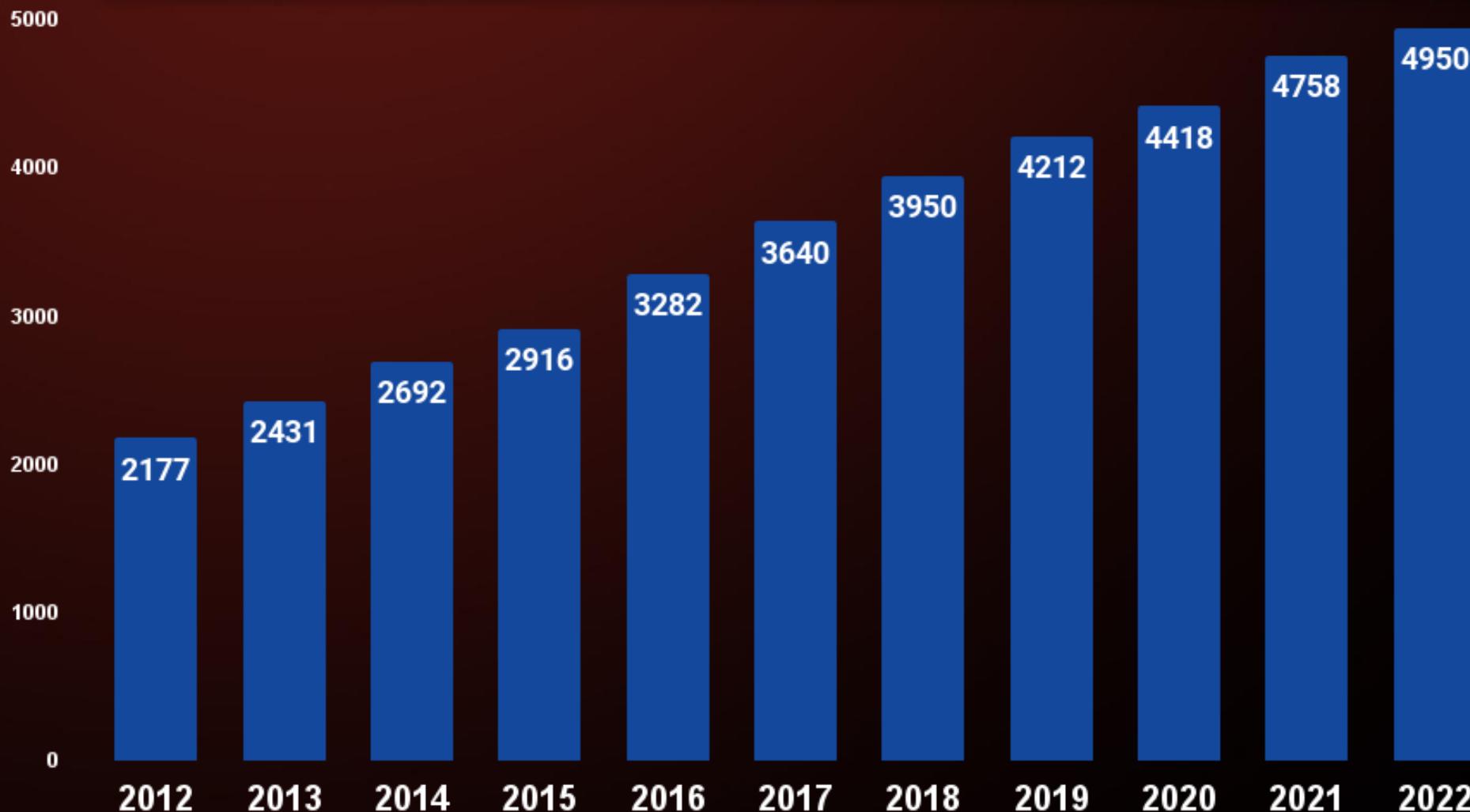


MINERAÇÃO DE DADOS

Introdução à Mineração de Dados

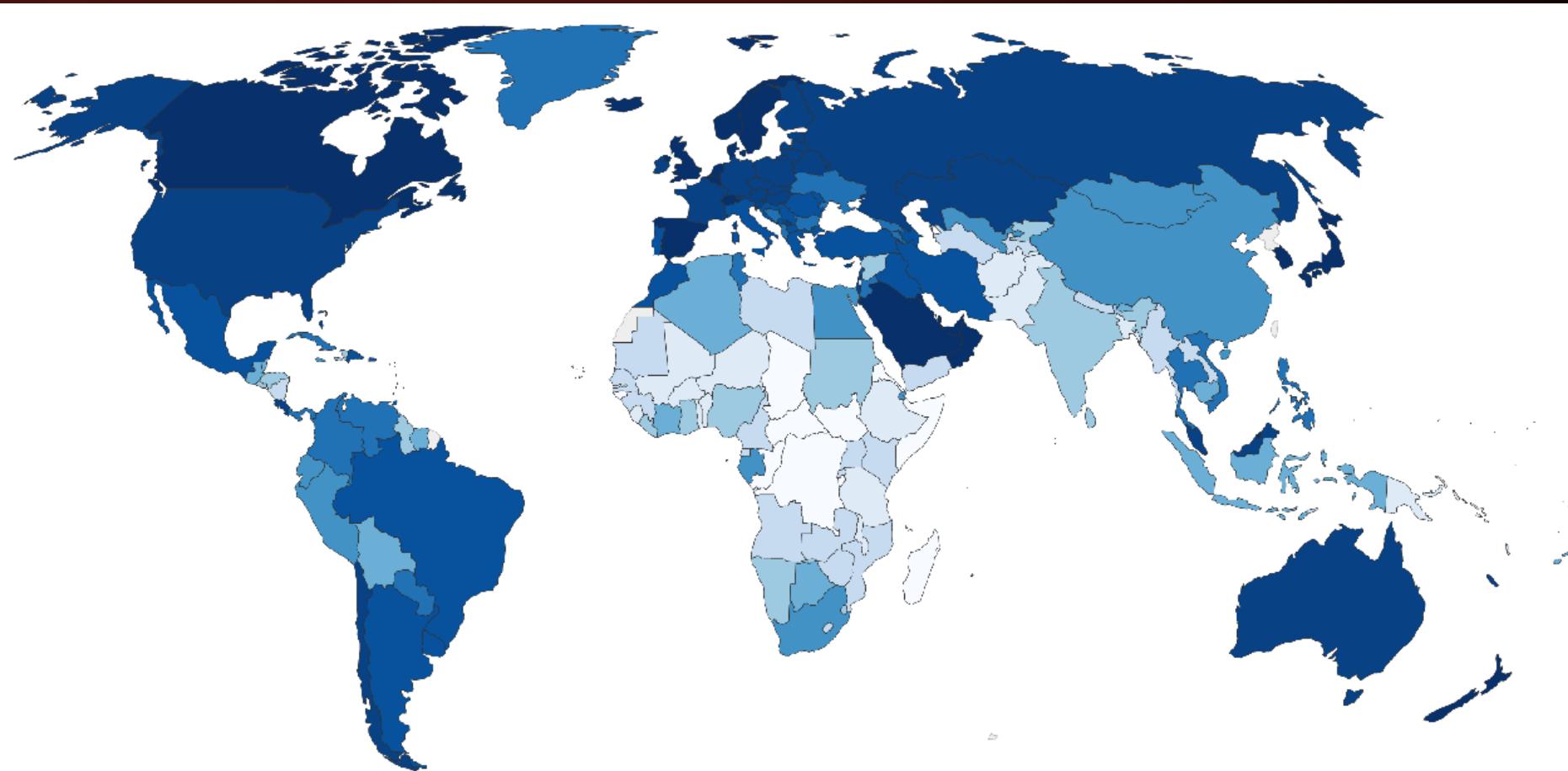
USUÁRIOS* DE INTERNET NO MUNDO



*em milhões

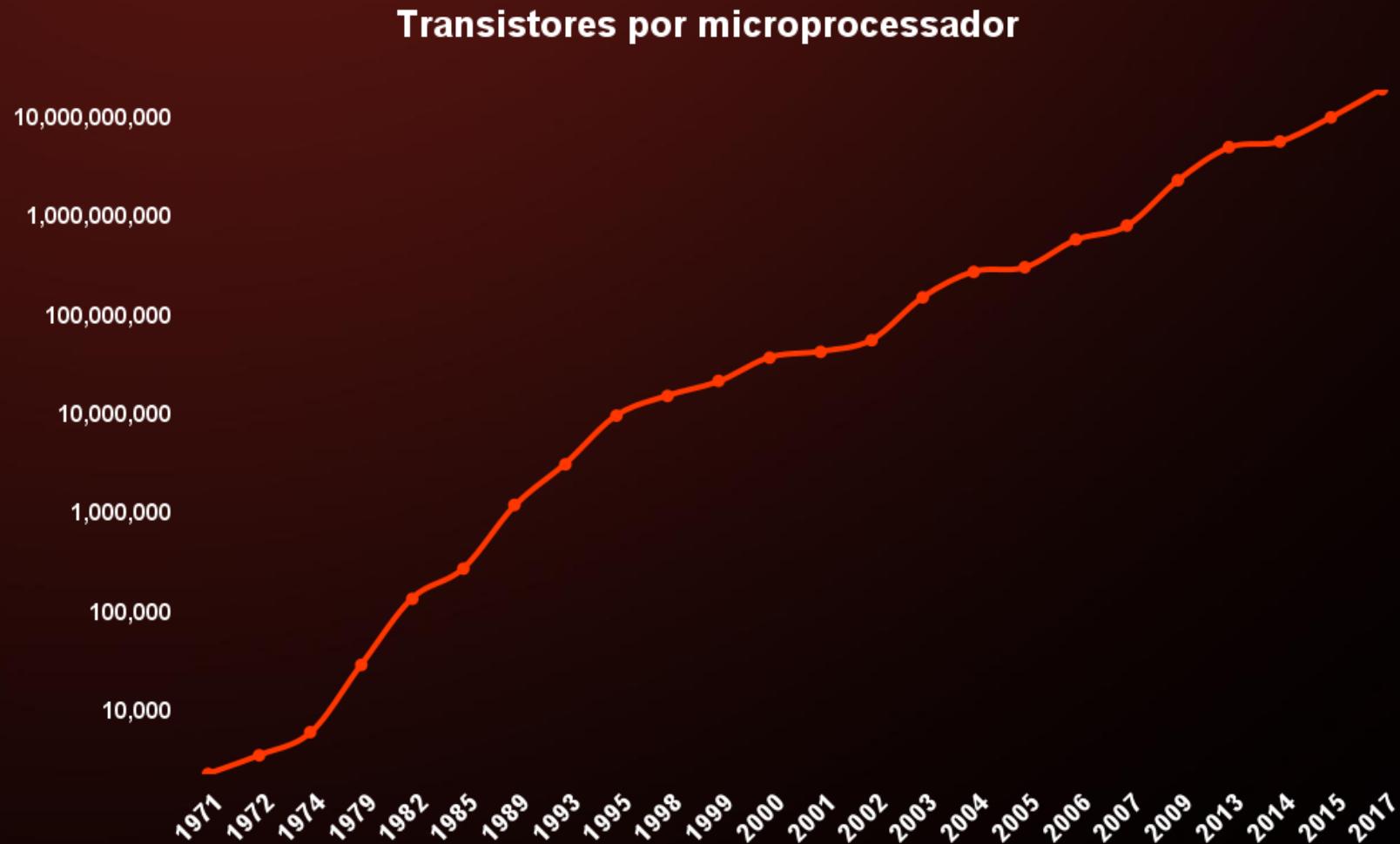
Fonte: DataReportal [1]

USUÁRIOS DE INTERNET NO MUNDO



Fonte: Our World in Data (2019) [2]

CAPACIDADE DE PROCESSAMENTO



Fonte: Our World in Data [3]

DADOS GERADOS POR MINUTO (2021)



Fonte: Statista [4]

EXPLOSÃO DE DADOS

- 79 zettabytes (Zb) gerados (2021):
 - 10^{21} bytes ou 1 bilhão de Tb
- Uma pessoa: 1.7 Mb por segundo (2020)
- Google: 3.5 bilhões de buscas por dia
- WhatsApp: 100 bilhões de mensagens por dia
- Gastos mundiais com Big Data (2022):
 - \$ 274,3 bilhões de dólares

TIPOS DE ESTRUTURAS DOS DADOS

- **Estruturados:**
 - Banco de dados, planilha
- **Semiestruturados:**
 - Marcadores (tags), email
- **Não estruturados:**
 - PDF, texto, imagens

TIPOS DE ATRIBUTOS



TIPOS DE ATRIBUTOS



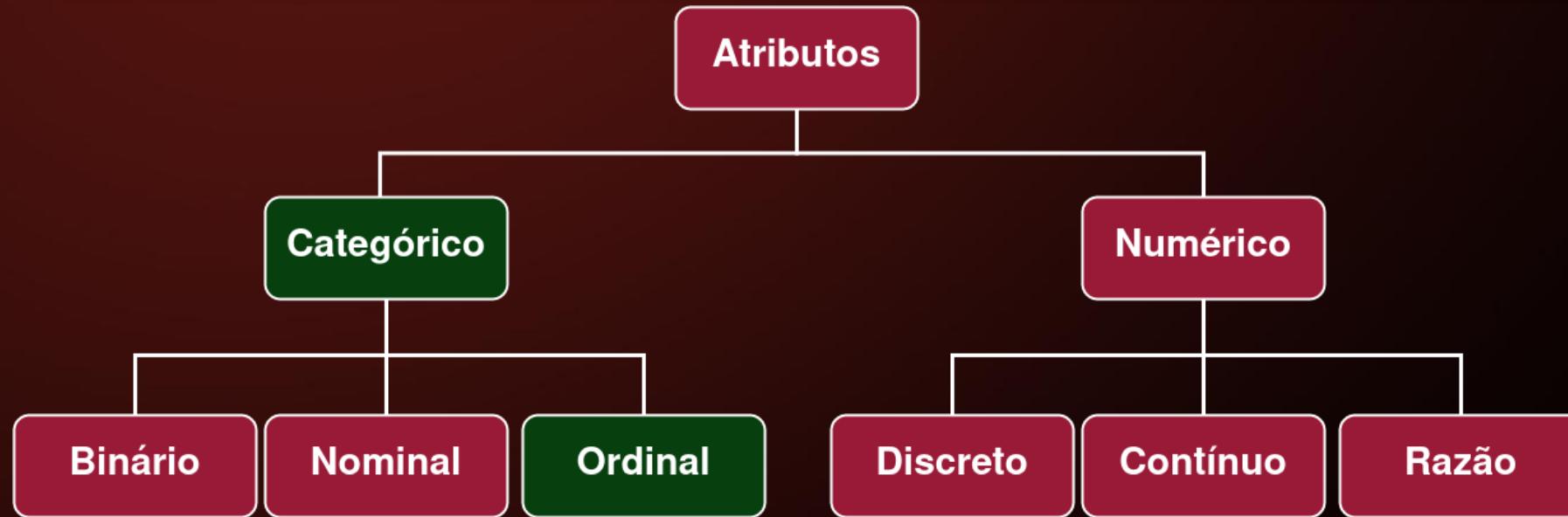
estado, aposentado

TIPOS DE ATRIBUTOS



estado civil, tipo de contato

TIPOS DE ATRIBUTOS



nível educacional, perfil de investimento

TIPOS DE ATRIBUTOS



idade, filhos

TIPOS DE ATRIBUTOS



peso, distância, salário

TIPOS DE ATRIBUTOS



distância entre objetos, proporção

valores contínuos ou discretos

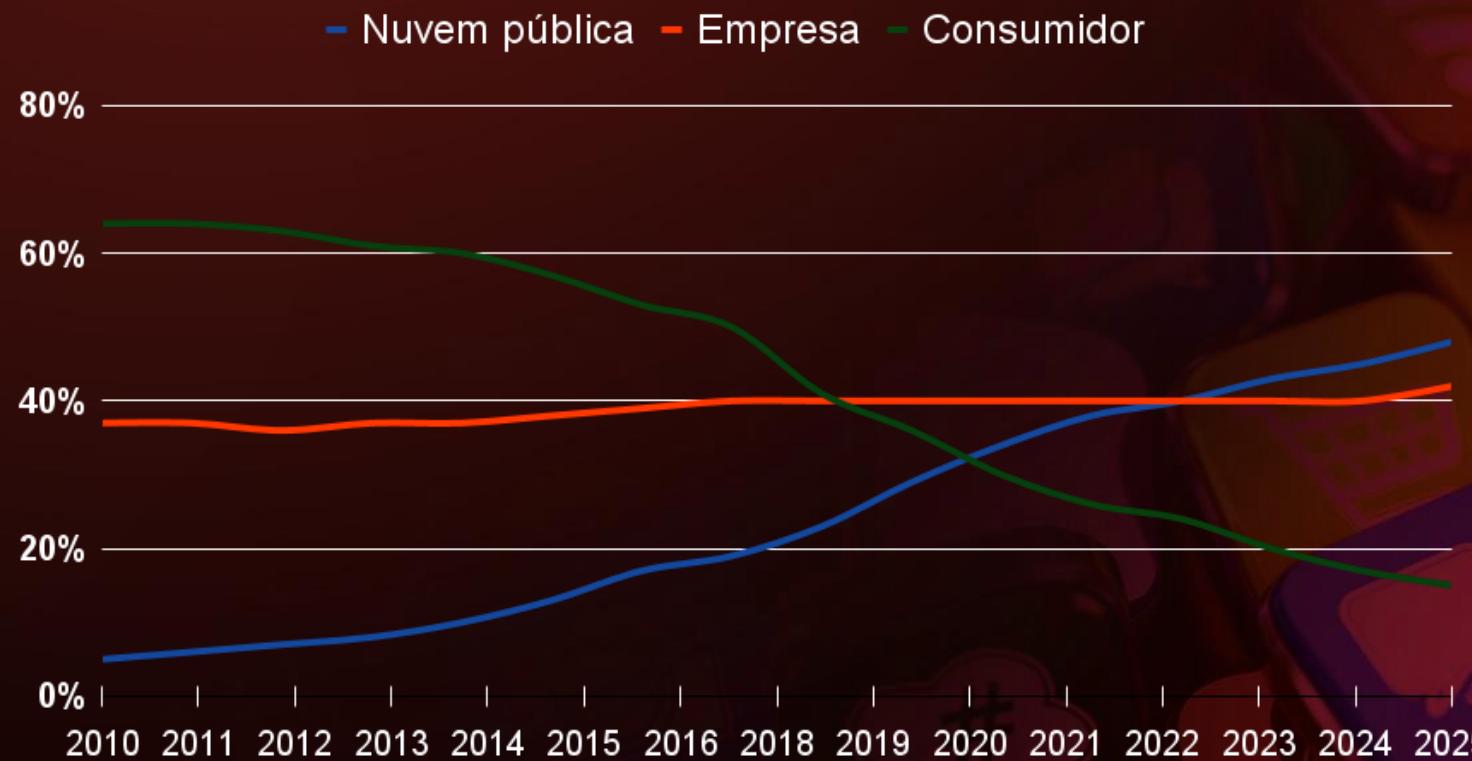
EXPLOSÃO DE DADOS

- 91% das 1000 maiores empresas mundiais investem em IA e Big Data
- 95% das empresas precisam lidar com dados desestruturados
- 80% dos dados gerados são desestruturados
- Até 2026: aumento de 28% de empregos na área de ciência de dados [6]

ONDE ESTÃO OS DADOS?

2022: maioria dos dados armazenados em nuvem

Distribuição de armazenamento de dados



Fonte: Visual Capitalist [7]

TIPOS DE DADOS

- Multimídia:
 - **Texto: formulários, postagens, notícias, buscas**
 - **Imagens, áudio, vídeo**
- Dados estruturados:
 - **Formulários, cookies, dados da aplicação**
- Dispositivos IoT:
 - **GPS, medições de sensores**

TIPOS DE DADOS

- Ciência:
 - DNA, experimentos diversos
- Dados pessoais:
 - Saúde, localização, governo
- Dados comerciais:
 - Transações em geral:
 - Compras, ações

**COMO ANALISAR ESSA QUANTIDADE
MASSIVA DE DADOS?**

MINERAÇÃO

Atividade de extração de materiais valiosos

Fonte: herbert2512, pixabay



MINA

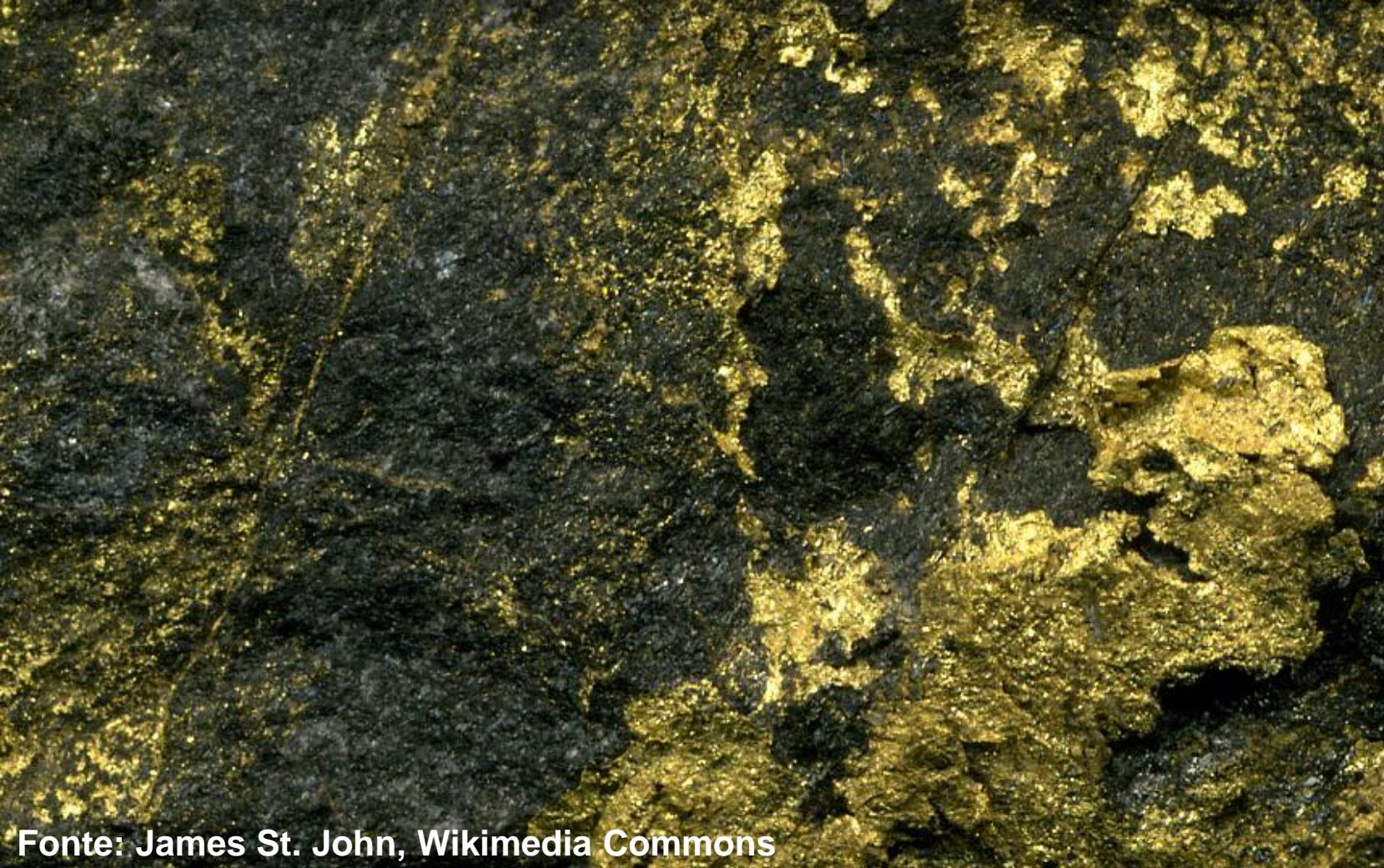
Local com concentração de substâncias



Fonte: Gus Jewel, Wikimedia Commons

SUBSTÂNCIA

Ouro, diamante, cobalto, urano, cobre etc.



Fonte: James St. John, Wikimedia Commons

FERRAMENTAS DE MINERAÇÃO

Usadas para extração dos minerais



FERRAMENTAS DE MINERAÇÃO

Usadas para extração dos minerais



Fonte: danielalejandro1712, pixabay

MINERAL

Substância extraída para uso



RESULTADO

Finalidade de uso do mineral com valor agregado



Fonte: PJ, Wikimedia Commons

MINERAÇÃO DE DADOS

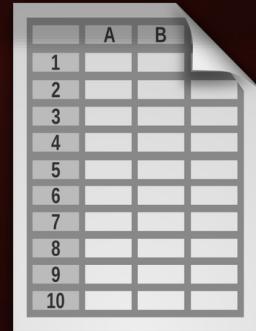
Mina:
Base de dados



Mineral:
Informação



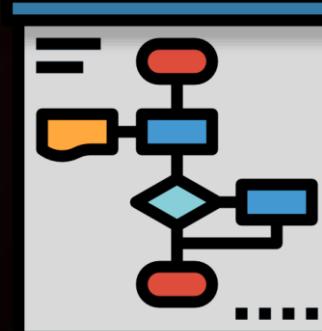
Substância:
Dados



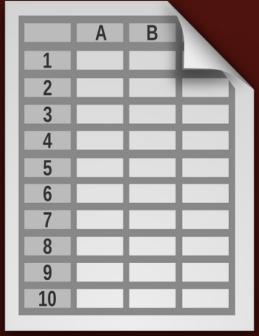
Resultado:
Conhecimento



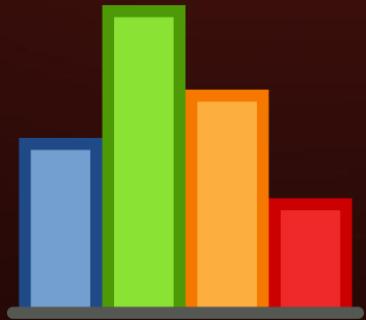
Ferramentas:
Algoritmos



DADOS x INFORMAÇÃO x CONHECIMENTO



1000 milibares
5,1 m/s; 95°
30 °C
poucas
1000 mts



Pressão atmosférica: 1000 milibares
Velocidade e direção do vento: 5,1 m/s; 95°
Temperatura do ar: 30 °C
Nuvens: poucas
Visibilidade: 1000 mts



A probabilidade de chuva é baixa, então posso ir à praia!

MINERAÇÃO DE DADOS

"Dados são o novo ouro"



Fonte: Stevebidmead, pixabay

REFERÊNCIAS

Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações: Cap. 1: Introdução à Mineração de Dados.

Leandro Nunes de Castro e Daniel Gomes Ferrari. Editora Saraiva, 2016.

[1] Digital 2022: Global Overview Report, Simon Kemp.

DataReportal: <https://datareportal.com/reports/digital-2022-global-overview-report>

[2] Share of the population using the internet. Our World in Data:

<https://ourworldindata.org/grapher/share-of-individuals-using-the-internet?time=latest>

[3] Moore's Law: The number of transistors on microchips has double every two years.

Our World in Data:

<https://ourworldindata.org/technological-change#:~:text=Moore's%20Law%20is%20the%20observation,to%20the%20number%20of%20transistors>

[4] A Minute on the Internet in 2021. Statista:

<https://www.statista.com/chart/25443/estimated-amount-of-data-created-on-the-internet-in-one-minute>

REFERÊNCIAS

[5] Big Data Statistics 2022: Facts, Market Size & Industry Growth. Jason Wise, Earthweb:
<https://earthweb.com/big-data-statistics>

[6] 25+ Impressive Big Data Statistics for 2022. Christo Petrov, TechJury:
<https://techjury.net/blog/big-data-statistics>

[7] Charting the Massive Scale of the Digital Cloud. Carmen Ang, Visual Capitalist:
<https://www.visualcapitalist.com/wp-content/uploads/2020/07/size-of-cloud-large.html>