



Python Data Analysis Mastery

Prof. Nicksson Freitas



Aula 02. Fundamentos de Data Science

Prof. Nicksson Freitas

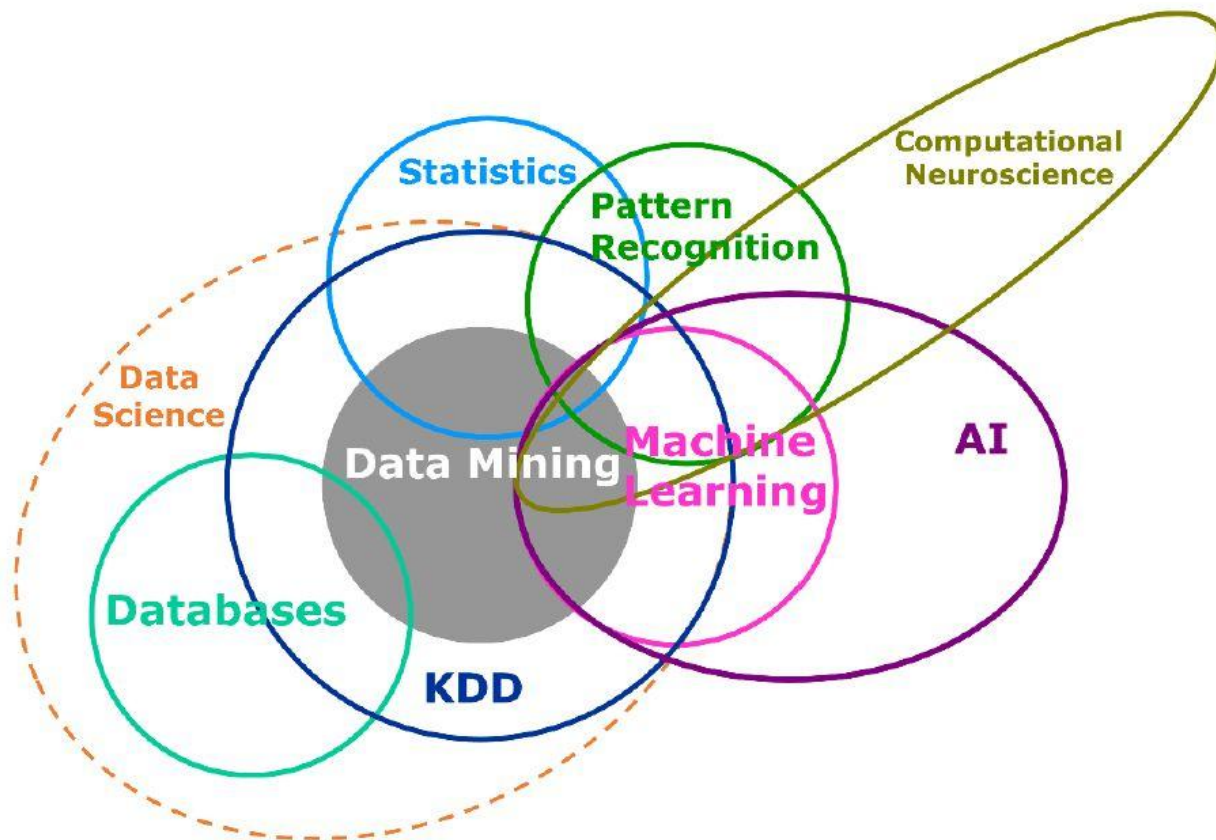
Objetivos

- Aprender os **princípios básicos da ciência de dados, mineração de dados e Big Data** e como aplicá-los a problemas reais.
- Entender o **papel do Cientista de Dados** no ambiente corporativo
- Conhecer o passo a passo para conseguir o **primeiro emprego**

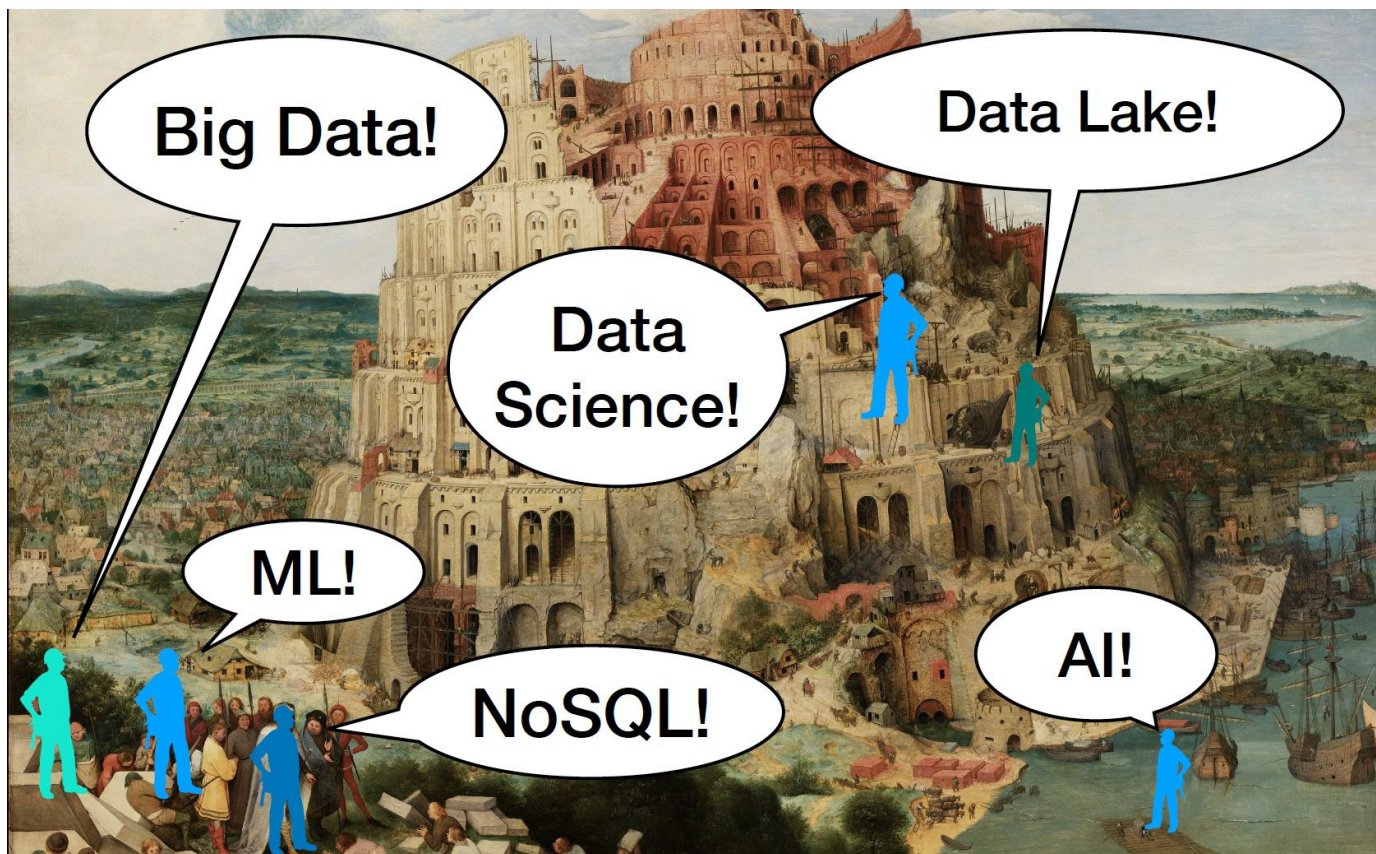
Resumo

1. O que é Data Mining?
2. O que é Machine Learning?
3. O que é Big Data?
4. O que é Ciência de Dados?
5. Carreira em Ciência de Dados
6. Por onde começar?

Contextualização



Contextualização



1. O que é Data Mining?

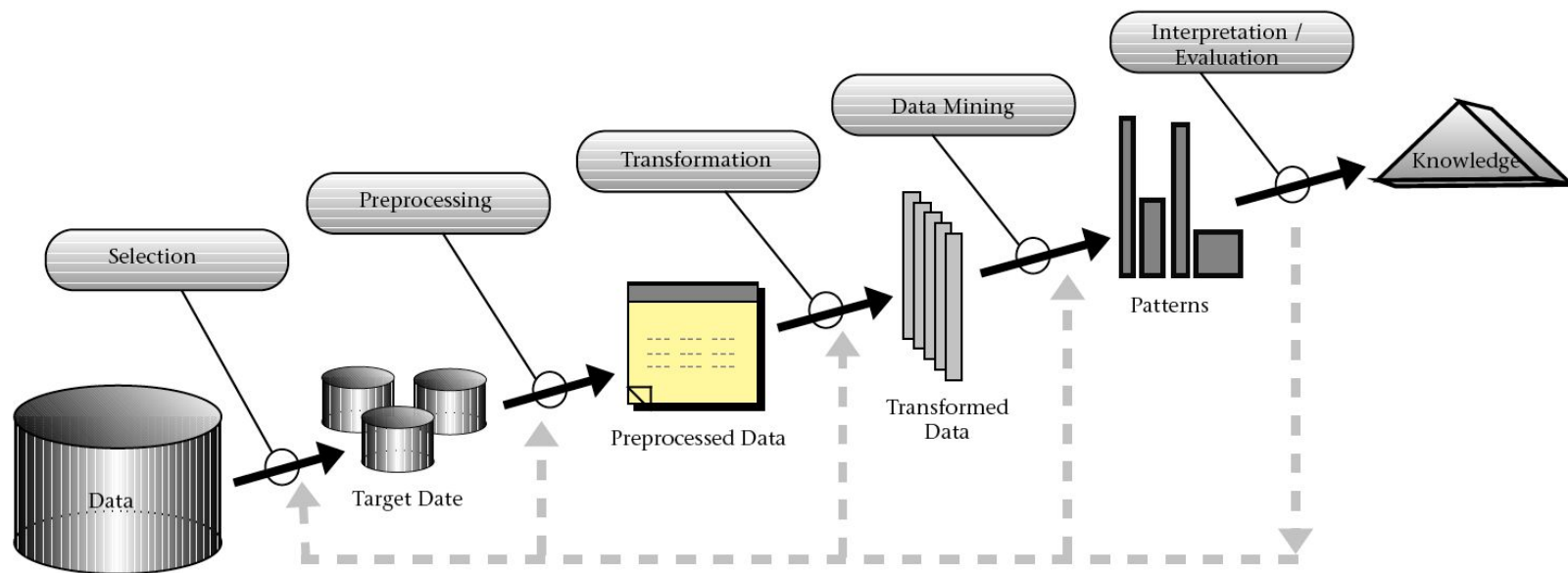
O que é Data Mining?

AI Magazine Volume 17 Number 3 (1996) (© AAAI)

From Data Mining to Knowledge Discovery in Databases

Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth

Processo de Mineração de Dados ou KDD



(FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996)

O que é KDD?

“

KDD é o processo não trivial, interativo e iterativo para identificar padrões em dados que sejam novos, válidos e potencialmente úteis e compreensíveis.

”

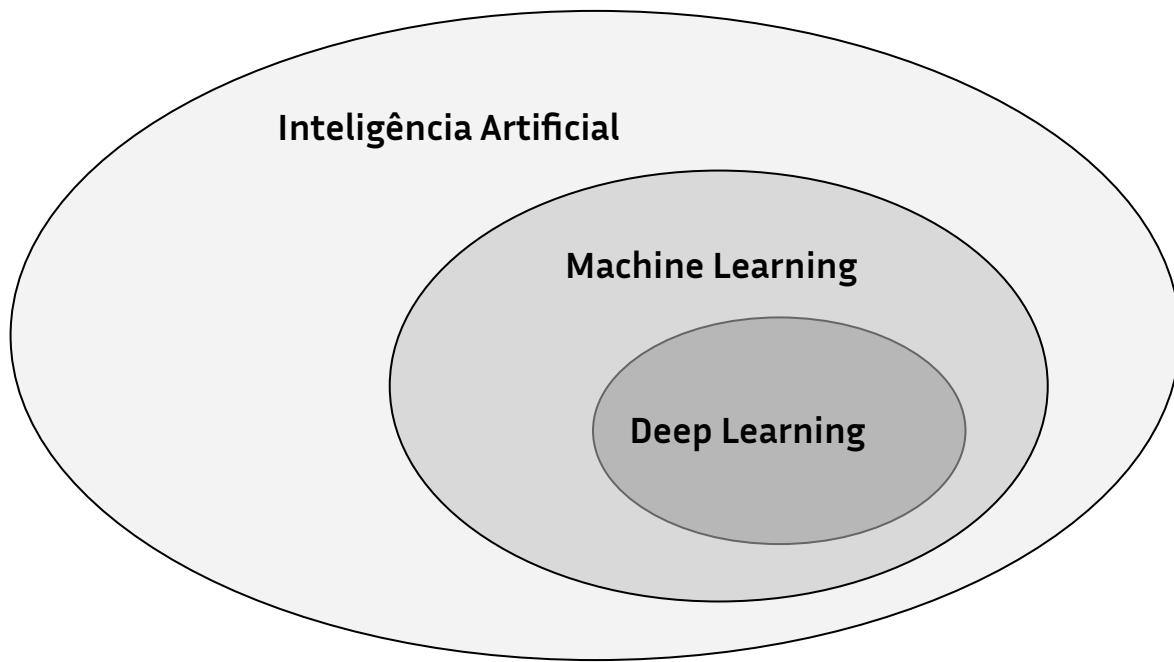
(FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996)

Resumindo Mineração de Dados

- Subcampo de Banco de Dados
- Data Mining é uma das etapas do processo de KDD (*Knowledge Discovery in Databases*)
- Uma ferramenta usada por humanos
- O que estudamos no campo de Data Mining?
 - Foca em como os dados são armazenados e acessados
 - como os algoritmos pode ser escalados
 - como os resultados podem ser interpretados e visualizados

2. O que é Machine Learning?

O que é Machine Learning?



O que é Machine Learning?

“

Machine Learning é um campo que estuda formas de fazer com que as máquinas aprendam sem serem programadas explicitamente

”

(Arthur Samuel, 1959)

O que é Machine Learning?

“

Machine learning é um programa de computador que aprende a partir de uma experiência E , em relação a alguma tarefa T e a uma medida de performance P e sua performance melhora a partir de E

”

(Tom Mitchell, 1997)

O que é Machine Learning?

	Data Mining	Machine Learning
Foco	Dados	Algoritmos
Grande Área	Bancos de Dados	Inteligência Artificial
Volume de Dados	Elevado	Pequeno
Intervenção Humana	Sim (toma muitas decisões)	Não (automático)
Aprendizagem/ Adaptabilidade	Não	Sim

3. O que é Big Data?

O que é Big Data?

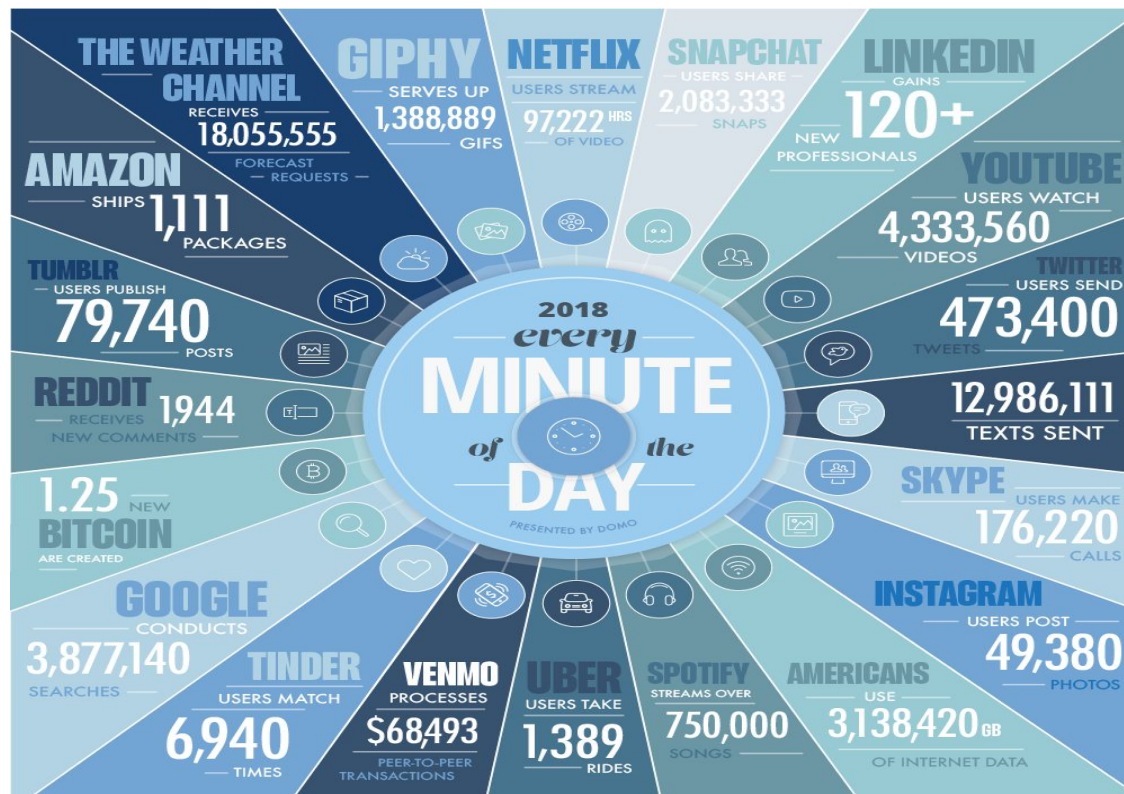
- O termo surgiu em 1997 para descrever um conjunto de dados cujo tamanho excedia os limites da capacidade de processamento de uma única máquina

Big Data and the Next Wave of InfraStress Problems, Solutions, Opportunities

Abstract:

Data storage is growing at a higher rate than ever before, and coupled with rapidly increasing demand for instant access, will cause great stress on both the physical and the human infrastructure of computing. System planners and administrators will soon face the interesting challenge of dealing with network and backup issues when office systems hold 100s of GB of disks, and larger servers reach 10s and 100s of TB and even PB. There will also be great opportunities in both research and commercial applications, but the problems must be understood, and solutions anticipated. This talk will give some examples, including some large customer problems that Silicon Graphics has been working on; and examine technology trends in storage capacities, access times, computer architectures, and bandwidths, to see what these portend over the next few years.

O que é Big Data?



O termo Big Data se popularizou...

- O termo foi impulsionado pelo novo conjunto de ferramentas e técnicas para desenvolver aplicativos modernos movidos a dados.



O que estudamos em Big Data?

- Captura de Dados
- Recuperação de Dados
- Consultar dados
- Compartilhamento de Dados
- Análise de Dados
- Visualização de Dados

4. O que é Ciência de Dados?

Definições de Data Science?

- Não existe uma única definição aceita por toda comunidade
- Não existe uma taxonomia padrão ou classificação
- Profissionais de diferentes áreas possuem visões diferentes
- Muitos termos inter-relacionados
 - Big Data, Data Science, Data Mining, Data Analysis, Machine Learning, Pattern Recognition, Analytics, Predictive Analytics, Big Data Analytics.

Definindo Data Science

“ Ciência de dados trata da coleta, integração, gerenciamento e exploração dos conjuntos de dados e da utilização do conhecimento adquirido com a finalidade de tomar decisões, **entender o passado/presente, prever o futuro e criar novos serviços/produtos.** ”

(Ozdemir, 2016)

Definindo Data Science

“

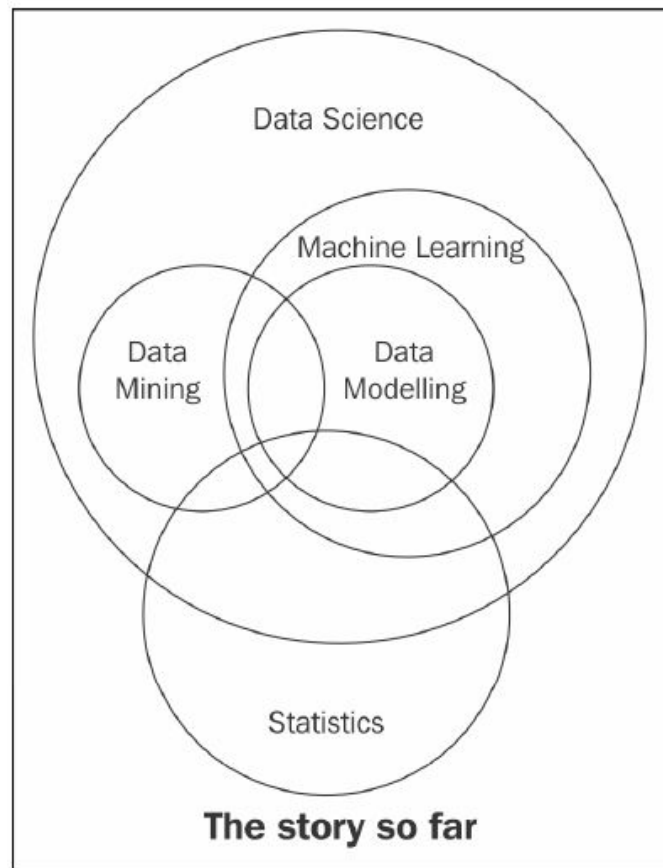
Data Science is the science which uses computer science, statistics and machine learning, visualization and human-computer interactions to collect, clean, integrate, analyze, visualize, interact with data to **create data products**.

”

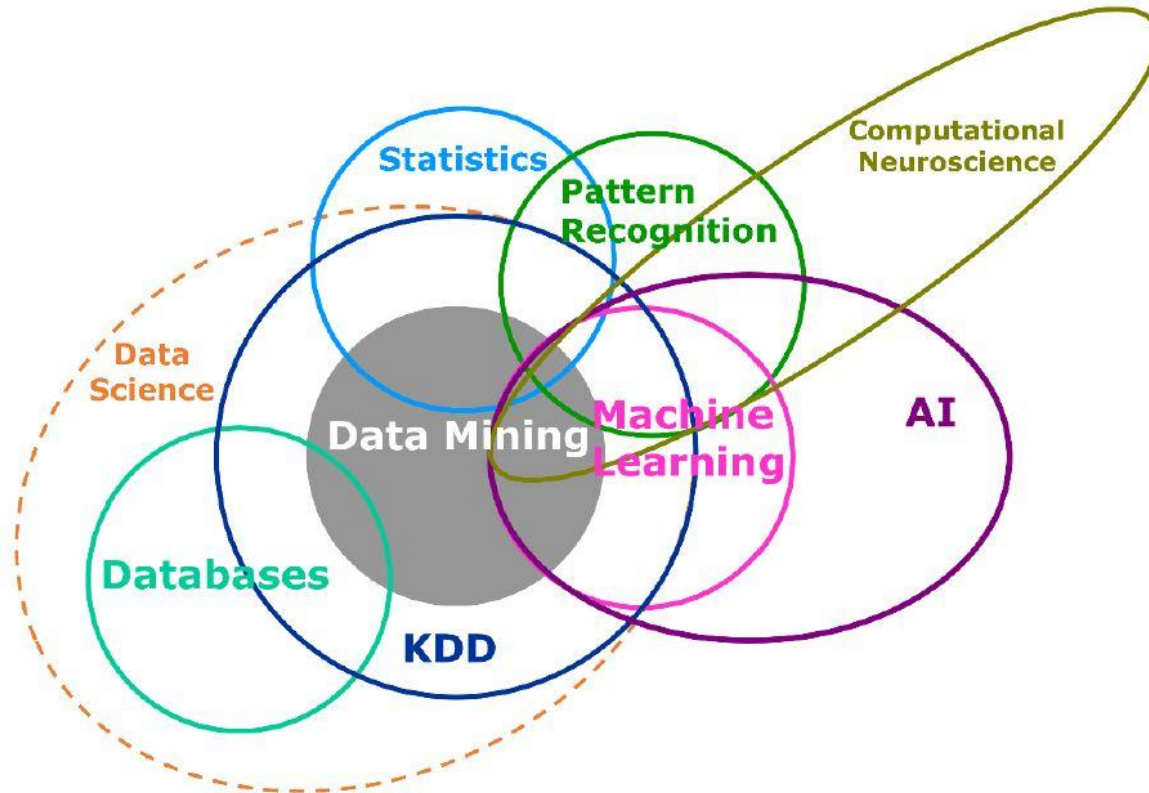
Autor Desconhecido

Principais áreas envolvidas

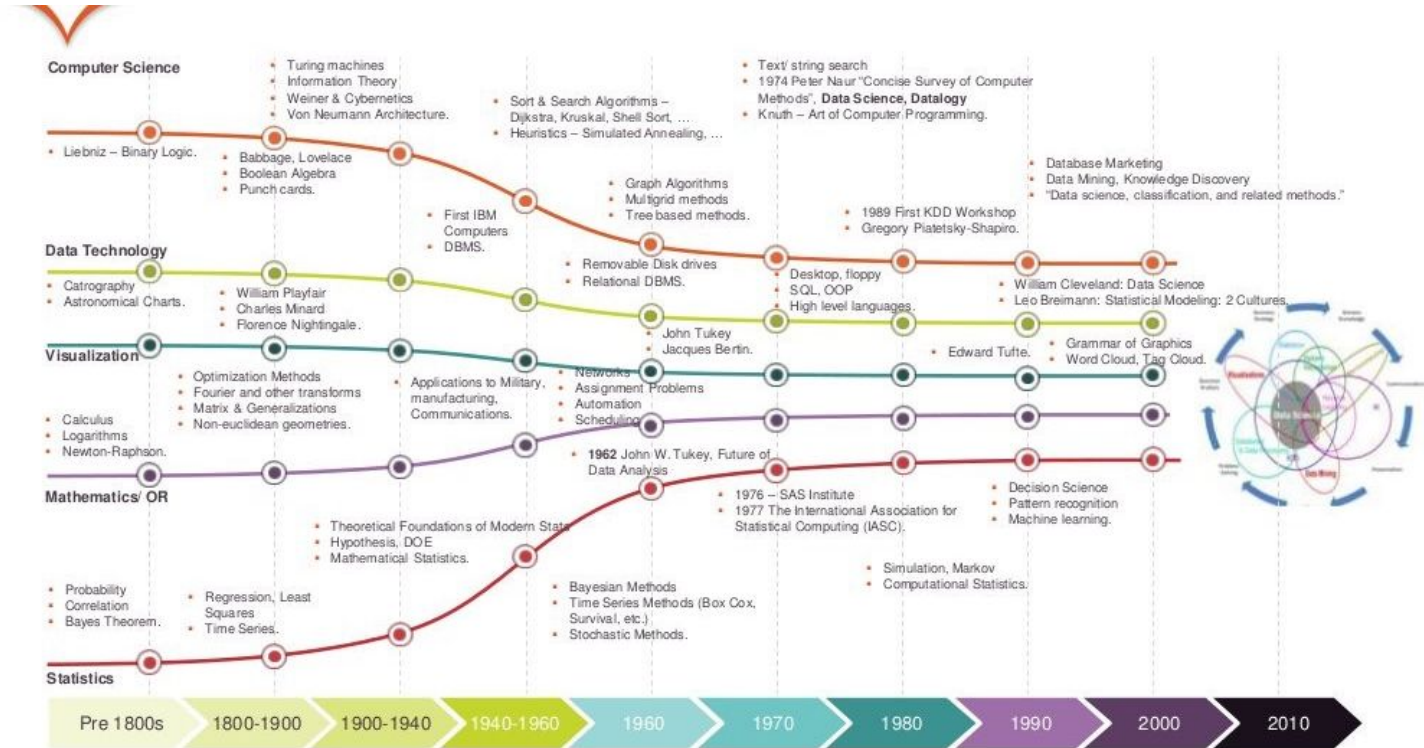
- Mineração de Dados
- Aprendizado de Máquina
- Matemática e Estatística
- Modelagem de Dados (Bancos de Dados)



Resumindo Data Science

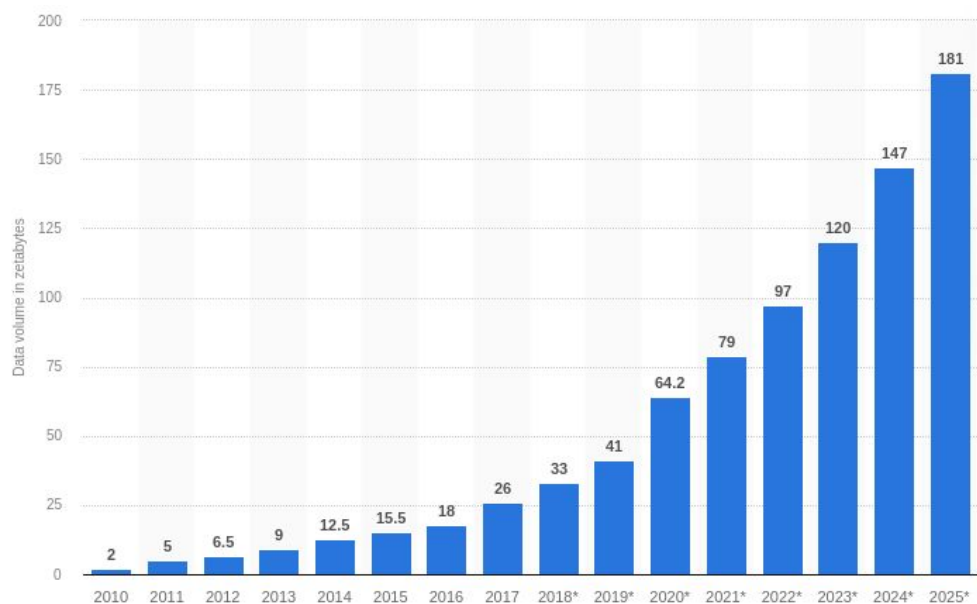


Resumindo Data Science



Por que Data Science se popularizou?

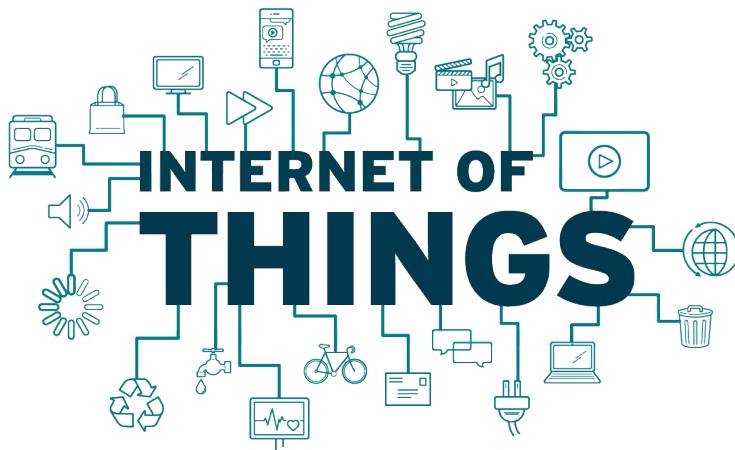
- 1. Aumento na coleta e volume de dados



Fonte: <https://www.statista.com/statistics/871513/worldwide-data-created/>

Por que o termo Data Science se popularizou?

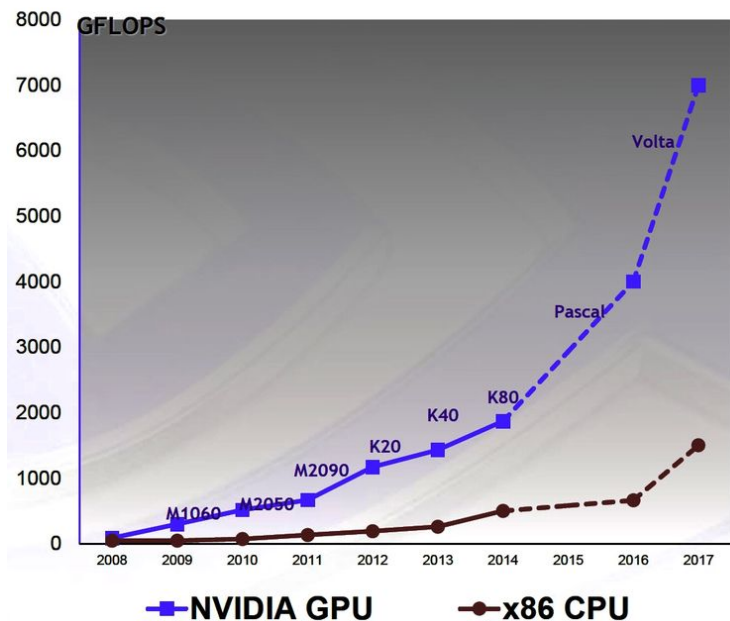
- 2. Crescimento no número de dispositivos móveis



Fonte: <https://www.statista.com/statistics/871513/worldwide-data-created/>

Por que o termo Data Science se popularizou?

- 3. Aumento no poder computacional



Modelos recentes - Nvidia A100



9.700 GFLOPS em FP64

19.500 GFLOPS em FP64 Tensor Core

É uma tendência na Ciência da Computação



5. Carreira em Ciência de Dados?

Quais são as principais habilidades?



Resumindo o dia a dia de um Cientista de Dados

- Coletar, gerenciar e armazenar dados
- Processar e tratar dados
- Analisar e estudar os dados
- Criar e gerenciar modelos de machine learning
- Criar sistemas e produtos a partir de dados
- Apresentar, documentar e discutir resultados

Onde eu consigo emprego?

LinkedIn

WORKANA

Upwork

glassdoor

99jobs.com



codementor

stackoverflow
JOBS



AngelList

Foque no LinkedIn (Brasil)



Desafio no time Amazon Brasil

Olá, Nicksson Ckayo. Tudo bem?

Quero te convidar a participar do nosso processo para o time Amazon São Paulo. Estamos crescendo cada vez mais rápido em escala mundial e esse time em específico será a chave principal no crescimento do Brasil.

O pacote de remuneração e benefícios são diferenciados e bastante atrativos.



B3 - Oportunidade Cientista de Dados

Olá, Nicksson Ckayo, tudo bem?

Gostáramos de conversar com você sobre uma oportunidade de trabalho que temos em aberto de Cientista de dados.

Qual a sua pretensão salarial?
Caso tenha interesse em obter mais informações, por favor encaminhe seus contatos para que possamos conversar.

Obrigada

Aryane Gabrielle da Conceição
Atração & Seleção / Tech Recruiter



Stellantis | Data Analysis Specialist

Olá, Nicksson Ckayo!
Tudo bem?

Tenho uma oportunidade de Data Analysis Specialist na Stellantis para atuar na área de Supply Chain e o seu perfil me chamou a atenção. O projeto tem previsão de duração de 9 meses e a vaga tem fortes chances de se tornar efetiva após este período.

Buscamos por um profissional com os seguintes requisitos:

- Experiência em projetos de Integração de Dados;
- Sólida expertise em análise de dados;
- Vivência em linguagem de programação (diferencial: Python, SQL);
- Inglês fluente (interação global).

Estamos trabalhando com uma faixa salarial de até R\$ 14.000. Esta vaga faz sentido para o seu momento atual? Se sim, podemos agendar um bate papo online?

Abraço!



Segue mais informações

Trata-se de uma vaga remota de Ciência de Dados para atuar junto à equipe que atende o contrato da Anatel.
Contratação CLT + benefícios.

Local Brasil: Home Office

Sênior: 05 anos de experiência
Salário CLT: R\$18.193,03

Contratação somente CLT

Requisitos:

Superior completo;
Experiência comprovada de 05 anos na área de atuação ou equivalente.

(Essa comprovação pode ser pela CTPS ou contrato de trabalho ou declaração de atividades assinada pelo contratante)

Foque no LinkedIn (Estrangeiro)

Oportunidade Internacional - Machine Learning

Nicksson Ckayo,
em parceria com a Revelo Network, estou buscando um Engenheiro de Software (Machine Learning) e que tenha inglês avançado, trabalhando para uma empresa de pesquisa de conversação personalizada e de inteligência artificial com sede em Los Angeles, Califórnia.

Remuneração em dólar.

Ótima oportunidade de carreira.

Vaga 100% remoto.

Tem ou conhece alguém que possa ter interesse?

Hi Nicksson,

Thrilled to help you find your next challenge.

You can visit our page, here you'll find all our available positions and their job description with the technologies needed: <https://careers.virtualmind.com/hiring?origin=recruiting&suborigin=CG>

For the Mid Level positions the salary range is from 4.500K to 6K and for the Senior positions from 6K to 7.500K, depending on your experience and seniority.

Hi from Meta! - PhD Opportunities

Hi Nicksson Ckayo,

Hope you are doing well.

My name is Ana Gabriela and I currently support Meta's Machine Learning team to grow.

I recently came across your LinkedIn profile and wanted to check if you would be interested in starting a conversation with us about PhD career opportunities at Meta!

At Meta you would apply your accumulated talent towards working with the world's richest collection of social data to solve challenging problems in areas like Search Ranking, Natural Language Processing, Graph

Hi Nicksson,

We are after a Data Scientist to work on a REMOTE Full-time or Part-time contract.

Client is US based.

Our budget is 40 usd/ hour. (Negotiable)

How is your spoken English (rate of 0- 10) and Availability?

Hello Nicksson Ckayo - Data Scientist

Hi Nicksson Ckayo, how are you doing?

I'm Andrea, Talent Advocate at Celerative. Right now we are looking for a Data Scientist to join a project with an analytics-based approach to cyber insurance, specifically quantifying specific exposure to ransomware attacks. Bridging the gap between technology and insurance by using AI-powered analytics that tracks individual exposure in real-time to create smarter insurance products.

THURSDAY

Harshita Wadhvani (She/Her) • 7:21 PM

Hi,

I hope you are doing well, I found your profile as a great match for one of our clients.

I have an amazing Remote Financial Data and Analytics Analyst opportunity for you. Kindly revert back if interested.

O que eu preciso fazer?

- Aprender as habilidades necessárias para um Cientista de Dados Jr
- Comprovar experiência nos processos seletivos
 - Projetos em portfólio
 - Artigos científicos
 - Artigos na internet
 - Vídeos
- Mostrar a capacidade analítica para resolver problemas usando dados
- Boa comunicação

6. Por onde começar?

Por onde começar?

1. Programação para análise de dados (Python, Pandas, Scikit-learn, Seaborn)
2. Mineração de dados
3. Matemática e Estatística
4. Visualização de Dados
5. Bancos de Dados
6. Machine Learning
7. Big Data
8. Cloud (AWS, GCP, Azure)

O que cobrimos neste curso?

1. **Programação para análise de dados (Python, Pandas, Scikit-learn, Seaborn)**
2. **Mineração de dados**
3. **Matemática e Estatística**
4. **Visualização de Dados**
5. Bancos de Dados
6. Machine Learning
7. Big Data
8. Cloud (AWS, GCP, Azure)

Considerações Finais

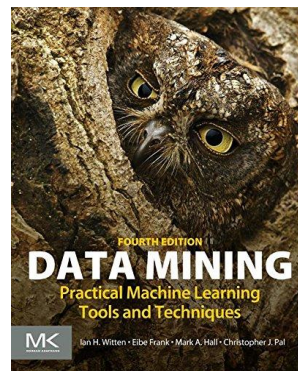
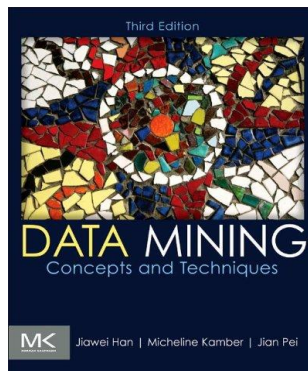
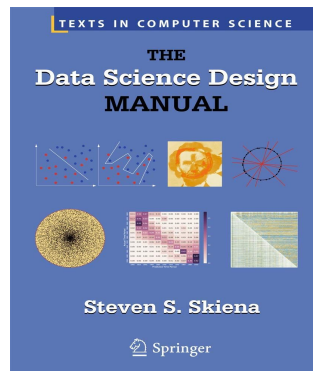
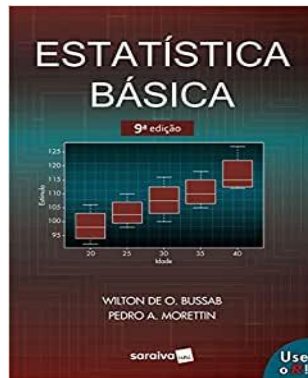
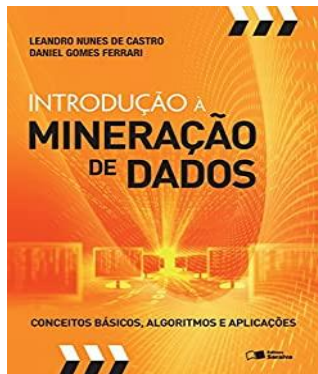
Considerações finais

- Você aprendeu o que é mineração de dados, big data, aprendizado de máquina e suas relações com Ciência de Dados.
- Você conheceu os diferentes papéis do Cientista de Dados no ambiente corporativo
- Conheceu as diferentes plataformas para conseguir o primeiro emprego
- Conheceu o passo a passo e as habilidades necessárias para se tornar um Cientista de Dados

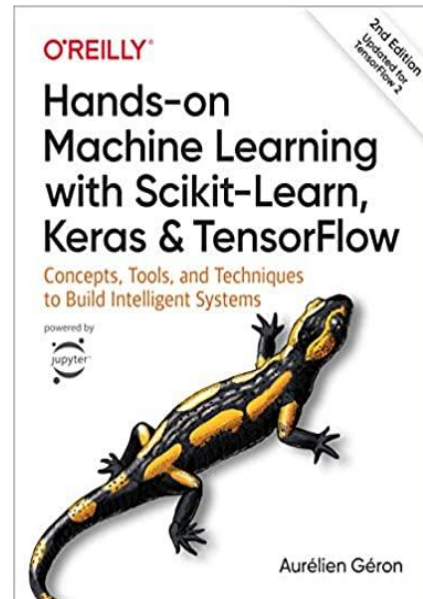
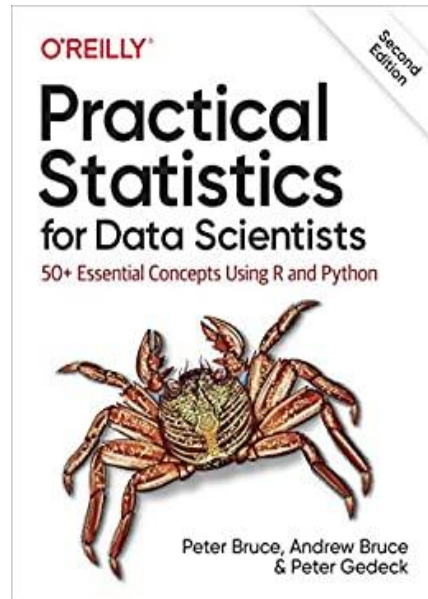
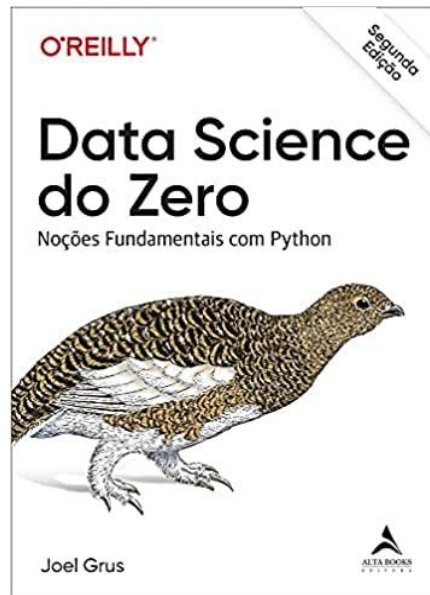
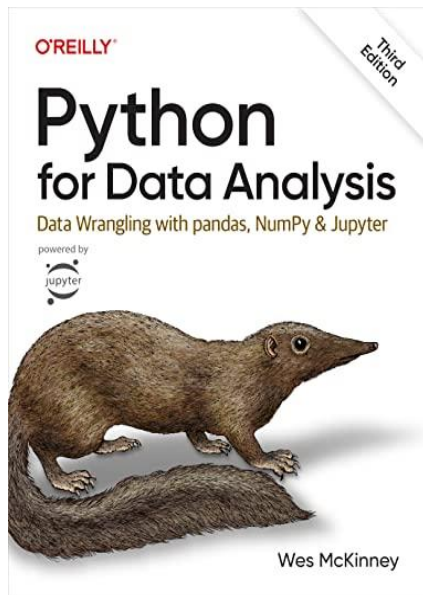
Referências



Bibliografia fundamental



Bibliografia técnica



OBRIGADO

