

# Aula 1 - Apresentação da disciplina

Jayme Anchante

22 de fevereiro de 2021

## introdução ao curso

# apresentação pessoal

Economista: graduação e mestrado, especialização em ciência de dados

Cientista de dados: Contabilizei, 4all e Edenred

Escalador, ciclista e leitor

[github](#); [linkedin](#)

hora de vocês se apresentarem

Favor ligar a câmera e mantê-la ligada, o curso é para vocês, então seu feedback e sua participação é essencial!



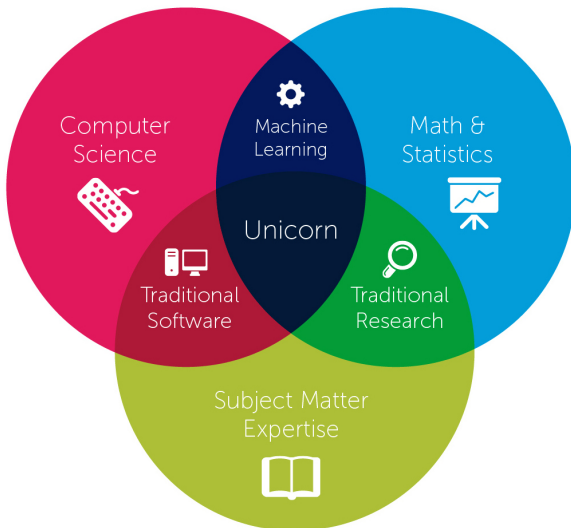
## o curso

- ▶ dia 1 (22/2): introdução, ciência de dados e programação
- ▶ dia 2 (23/2): processamento de dados com numpy e pandas
- ▶ dia 3 (24/2): aprendizado de máquina com scikit-learn
- ▶ dia 4 (25/2): visualização de dados com matplotlib e SQL/metabase
- ▶ dia 5 (01/3): projeto final

ciência de dados

# diagram de venn do cientista de dados

## Data Science



# cientista de dados moderno

## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



### PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

### COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



# profissão do cientista de dados

## Data Scientist: *The Sexiest Job of the 21st Century*

Meet the people who  
can coax treasure out of  
messy, unstructured data.  
by Thomas H. Davenport  
and D.J. Patil

**W**hen Jonathan Goldman arrived for work in June 2008 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But were users? At the time executives had expected, something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

38 Harvard Business Review October 2012



# casos de uso



programação com python

# por que python?

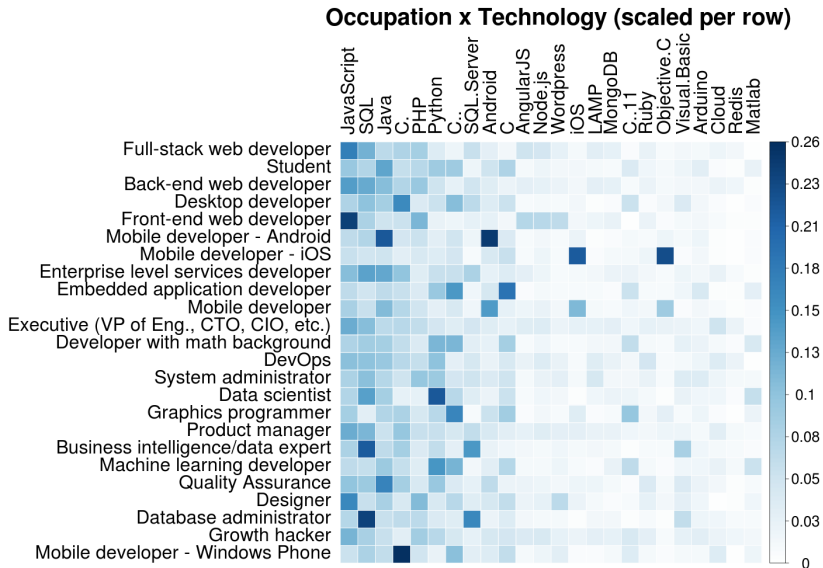


Figure 1: Pearson correlation Stack Overflow

o que é python



o que é anaconda



Acesse este [link](#) para baixar a distribuição para seu sistema

# teste de programação

[Acesse o teste aqui](#)

Favor não consultar internet ou outras pessoas

# revisão de Python

Livro base é o [Pense em Python - pense como um cientista da computação](#) por Allen B. Downey



exercícios de casa

# lista

Exercício 1.2.

Exercício 2.2.

Exercício 6.2.

Exercício 10.2.

Exercício 15.1.