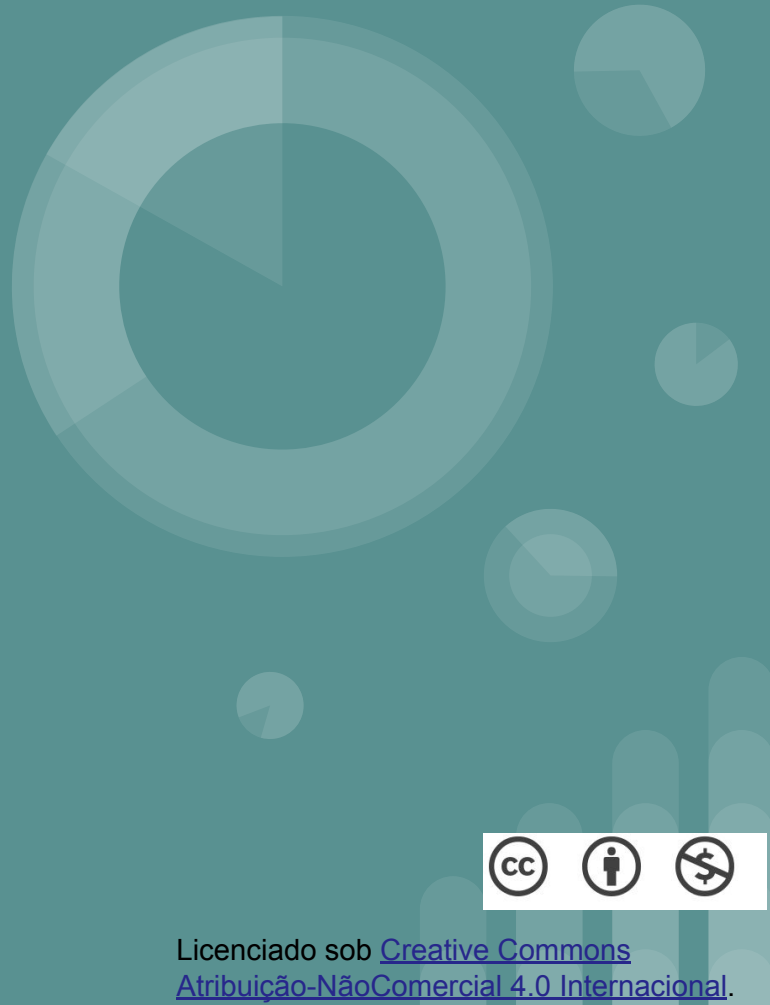


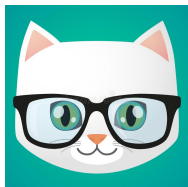
Ciência de Dados

uma visão geral





Sobre eu:



Mestre em **Ciências Sociais**, gerando o Léo e participante da comunidade **PyLadies São Paulo**. Me apaixonei pela análise de dados e decidi migrar de área profissional. Hoje trabalho como **Analista de Dados** em uma empresa ligada ao agronegócio.



Agenda do dia:

- ☐ Ciência de Dados e o Cientista de Dados
- ☐ Como fazemos essa ciência?
- ☐ Olhando a modelagem
- ☐ E o Aprendizado de Máquina (ML para os íntimos)?
- ☐ Onde vemos ML?
- ☐ Temos tempo para debate?
- ☐ Sugestões de Leitura

O que é ciência de dados e quem a faz?





Uma definição em construção:

Data Science ou **Ciência de Dados** é uma ciência relativamente nova que utiliza técnicas de diversas áreas, já consolidadas, para **extrair de dados brutos, insights** e/ou soluções para diferentes tipos de problemas.



A Ciência de Dados produz insights que você pode usar para entender e melhorar seu negócio, seus investimentos, sua saúde e até mesmo seu estilo de vida Usar a ciência de dados é como ver no escuro

Pierson, L. - ***Data Science for Dummies***






Quem faz ciência de dados?

Data Scientist ou **Cientista de Dados** geralmente possuem algum conhecimento ou domínio sobre alguma das áreas ao lado e conseguem garimpar diamantes no meio do caos dos dados.

Inspirado na apresentação de Maria Marinho sobre Ciência de Dados - Aniversário do PyLadies 2018

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 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
- PROGRAMMING & DATABASE**
 - ☆ Computer science fundamentals
 - ☆ Scripting language e.g. Python
 - ☆ Statistical computing packages, 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
- DOMAIN KNOWLEDGE & SOFT SKILLS**
 - ☆ Passionate about the business
 - ☆ Curious about data
 - ☆ Influence without authority
 - ☆ Hacker mindset
 - ☆ Problem solver
 - ☆ Strategic, proactive, creative, innovative and collaborative
- 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

Perfil de uma cientista de dados:

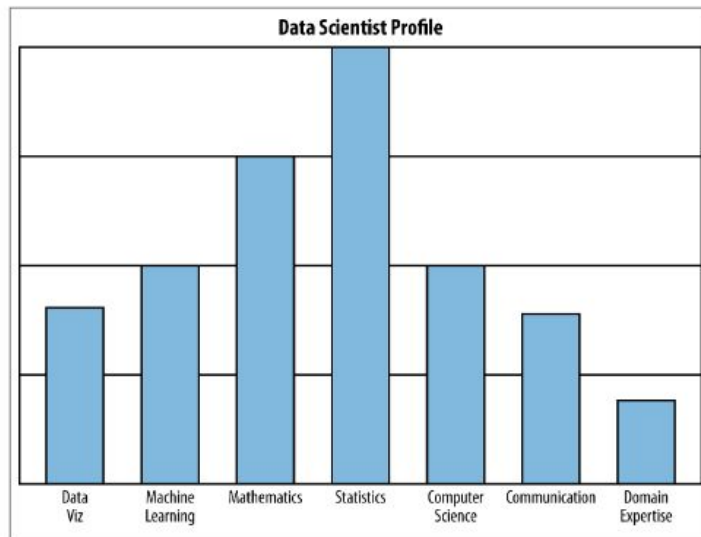
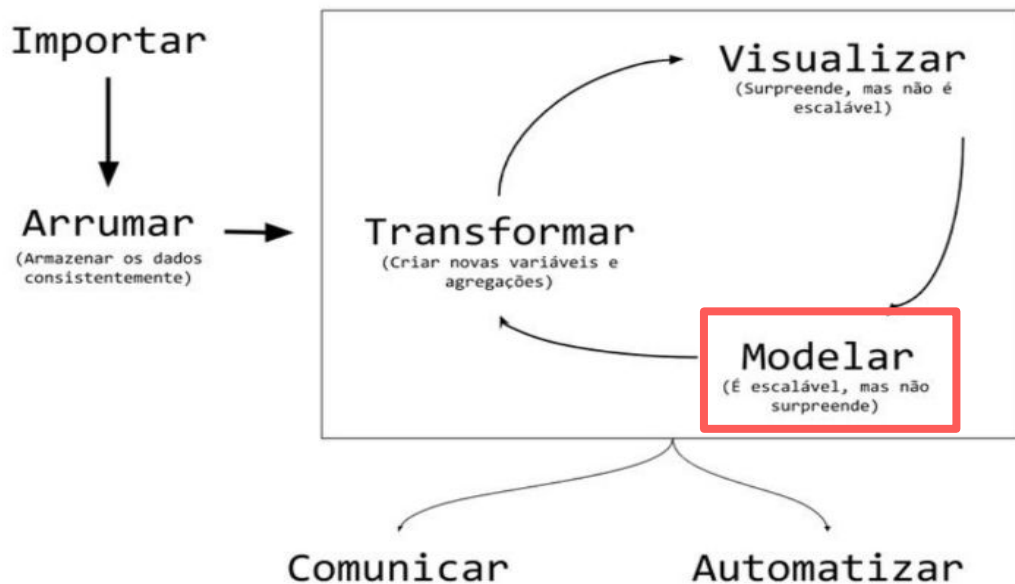


Figure 1-2. Rachel's data science profile, which she created to illustrate trying to visualize oneself as a data scientist; she wanted students and guest lecturers to "riff" on this—to add buckets or remove skills, use a different scale or visualization method, and think about the drawbacks of self-reporting

Como se faz?



Ciclo da Ciência de Dados:



Modelagem



Mas antes, o algoritmo:

Sequência finita de **instruções**
PyLadies São Paulo - **Workshop Básico 1**

Bolo Super Fofo



Ingredientes do Bolo Fofo:

3 colheres de sopa de manteiga;
4 ovos;
1 xícara de leite;
3 xícaras de farinha de trigo;
2 xícaras de açúcar;
1 colher de sopa de fermento em pó;
1 pitada de sal.



Modelagem:

1. Processo pelo qual o escultor obtém **um modelo**, geralmente em argila ou cera, que será posteriormente fundido; moldagem.
2. Conjunto de exercícios físicos que se destinam a fortalecer os músculos e a **modelar o corpo**.
3. Criação de **modelos em computador**.





Juntando tudo:

Ingredientes



Preparo



Modelo



Resultado



Machine Learning:





O que é?

Machine Learning, **Aprendizado de Máquina** em português, é um método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da **inteligência artificial** baseado na ideia de que sistemas podem aprender com dados, **identificar padrões e tomar decisões com o mínimo de intervenção humana**

SAS - Machine Learning

Transforma Dados em Conhecimento:



Dados de Treinamento



Processamento



Modelo / Decisões automatizadas

Aplicações:





Medicina:

- reconhecimento de imagens e exames;
- cruzamento de informações de prontuários;
- monitoramento de pacientes e disparo de alarmes em caso de necessidades.





Agronegócio:

- tratamento de imagens de satélite apontando maior ou menor biomassa;
- identificação de pragas através de imagens de satélite;
- recomendação de híbridos conforme região que se deseja plantar;
- recomendação de taxa de sementes por região que se deseja plantar.





Setor Financeiro:

- perfil de cliente;
- cessão de crédito;
- fraudes.





Administração Pública:

- distribuição de processos no STF;
- controle epidemiológico, aplicativo ARBO no caso da Dengue em GO;
- reconhecimento facial nas unidades prisionais em MG





Participação Social:

- Parlamentaria - acompanhamento das tramitações de lei e perfil dos parlamentares
- Operação Serenata de Amor - fiscaliza os reembolsos efetuados pela Cota para Exercício da Atividade Parlamentar (CEAP).
- Elas no Congresso - medição de como cada deputada (o) e senadora (o) tem atuado em leis importantes para os direitos das mulheres no Brasil.



**Caminhamos para um
mundo utópico ou
distópico?**





Sugestões de Leitura:

- <https://saudedasaude.anahp.com.br/hospital-do-futuro-como-o-avanco-da-tecnologia-esta-revolucionando-a-saude/>
- <https://www.goias.gov.br/servico/35-saude/120189-tecnologia-auxilia-na-batalha-contr-o-aedes.html>
- <http://www.agenciaminas.mg.gov.br/noticia/sistema-prisional-de-minas-implanta-tecnologia-de-reconhecimento-facial>
- <https://serenata.ai/>
- <https://parlametria.org.br/home>
- <https://www.elasnocongresso.com.br/>



Sugestões de Leitura:

- <https://www.politize.com.br/distribuicao-processos-stf/>
- https://www.sas.com/pt_br/insights/analytics/machine-learning.html
- <https://tecnoblog.net/247820/machine-learning-ia-o-que-e/>
- <https://canaltech.com.br/inteligencia-artificial/como-a-inteligencia-artificial-vai-transformar-o-agronegocio-109188/>

Livros:

Data Science Para Leigos

Doing Data Science