Data Science com Python para Todas

Data Science

python™

Data Science com Python para Todas

```
def maria_marinho():
    profession = ['System Analyst', 'Developer’]
    hobbies = ['Music', 'Shows', 'Piano', 'Movies & Series’]
    dogs = ['Iza', 'Ikky', 'Lino', 'Luke']
    email = 'mariamarinhos@gmail.com'
```

Data Science com Python para Todas

- ✓ Trajetória no mundo de TI, Data Science, Python e comunidades
- ✓ O que os dados me mostraram: projetos e desafios
- ✓ O que é Data Science?
- ✓ Perfil do Cientista de Dados
- ✓ Data Science com Python
- ✓ Mulheres cientistas de dados para se inspirar
- ✓ Guia de bolso para começar a aventura no mundo dos dados

Vem comigo, vai começar!

Minha História – anos 90

O começo: 1995

- Colegial Técnico em Processamento de Dados
- Qbasic, Pascal, Cobol, C
- Primeiro estágio: DOS, Windows 3.1, Clipper, dBase, Lotus 123

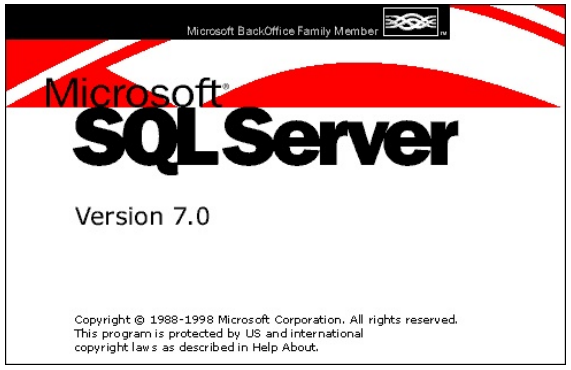
Minha História – anos 90

#Spoiler Stranger Things

- A foto ao lado mostra um código na linguagem **Basic** escrito pelo personagem Bob.

Minha História – anos 2000

- Bacharelado em Matemática com Informática
- Pós graduação em Educação Matemática
- Analista desenvolvedora
- Professora de Matemática



Minha História – anos 2010

- Primeiro contato com um software estatístico em uma empresa de Pesquisa de Mercado: SPSS
- Intercâmbio
- Analista Desenvolvedora em empresas do campo bancário e farmacêutico: automatização de processos em VBA

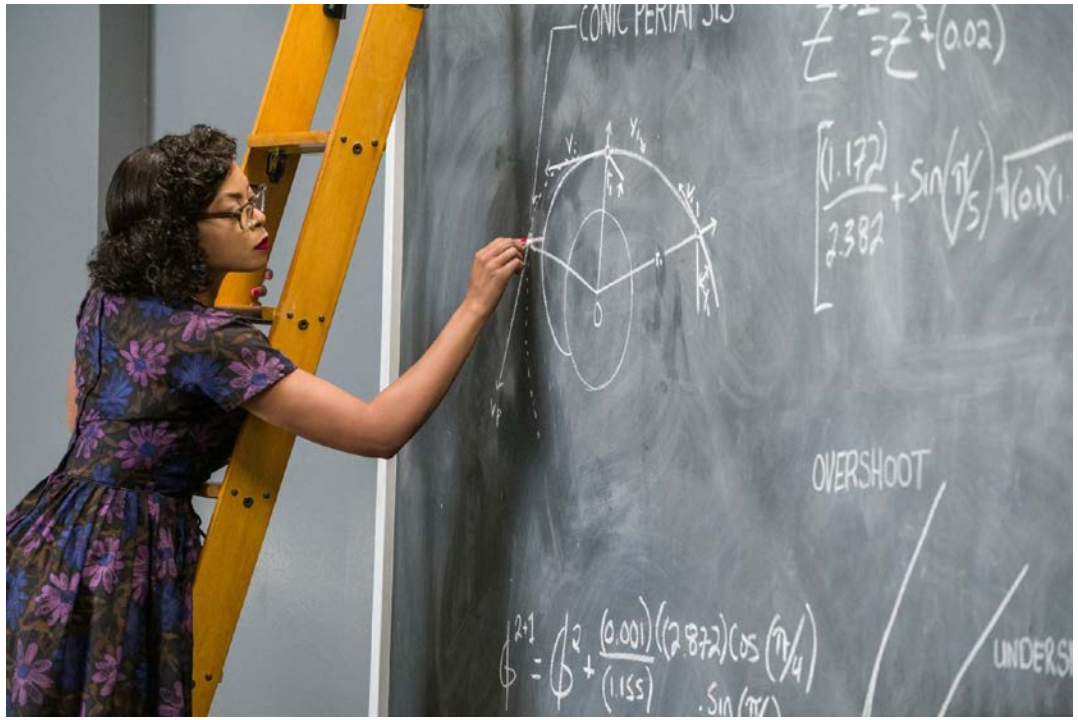
Minha História – era da Ciência de Dados

- Descoberta da Ciência de Dados: um horizonte de possibilidades
- Cursos MOOC: Coursera
- Estudo sobre a Relação entre o Câncer de Mama e a Emissão de CO2
- Descoberta dos Meetups e Workshops das Comunidades de Tecnologia
- Bolsa de Estudos pela Udacity e Bertelsmann
- Monitora e Professora de cursos das PyLadies São Paulo

Data Science com Python para Todas

- O que os dados me mostraram?
- Projetos e Desafios

Cena do filme “Estrelas Além do Tempo”



Udacity Show Project: análise da base do ProUni

- **Projeto em grupo para o Bertelsmann Data Science Challenge Scholarship Course**

Universities' prices analysis based on Prouni database

- **Problem:** Is there any relevant difference between universities' prices grouped by UF (Federative Unit) for Computer Science graduation?
- **Hypothesis 1**: there isn't difference in monthly fees between UFs
- **Hypothesis 2:** there is difference in monthly fees between UFs

Datathon: Desafio de Dados sobre Saúde Pública no Brasil

- **Segurados de ambos sistemas: SUS e Planos de Saúde**

A análise tem como base os atendimentos de beneficiários de planos de saúde no Sistema Único de Saúde (SUS) e tem como objetivo responder a pergunta: Os planos de saúde ressarcem o SUS de acordo com a lei n° 9.656/1998?

HackMobilidade 2018

- **Análise de acidentes com pedestres (atropelamentos) biênio 2016-2017**

Tema do HackMobilidade 2018: usar ciência de dados para propor soluções de segurança para a mobilidade ativa feminina.

Equipe de 4 mulheres:

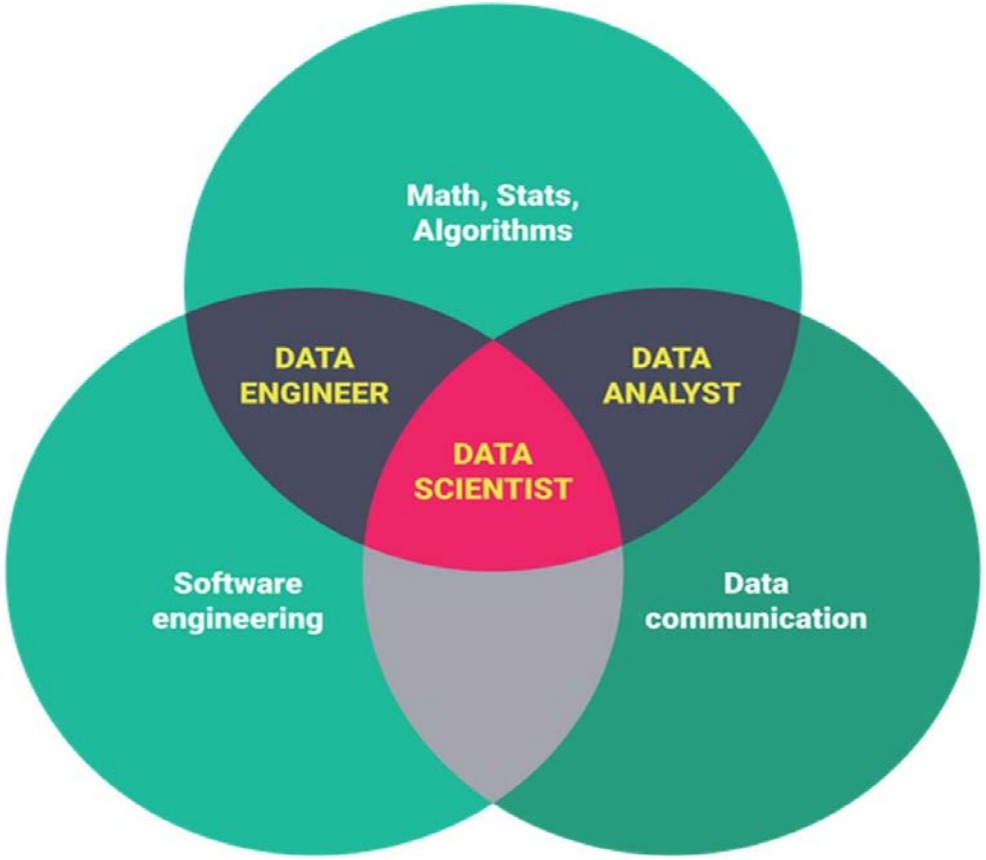
- Alissa Mune
- Fabíola Canedo
- Maria Marinho
- Monica Craveiro

HackMobilidade 2018

HackMobilidade 2018

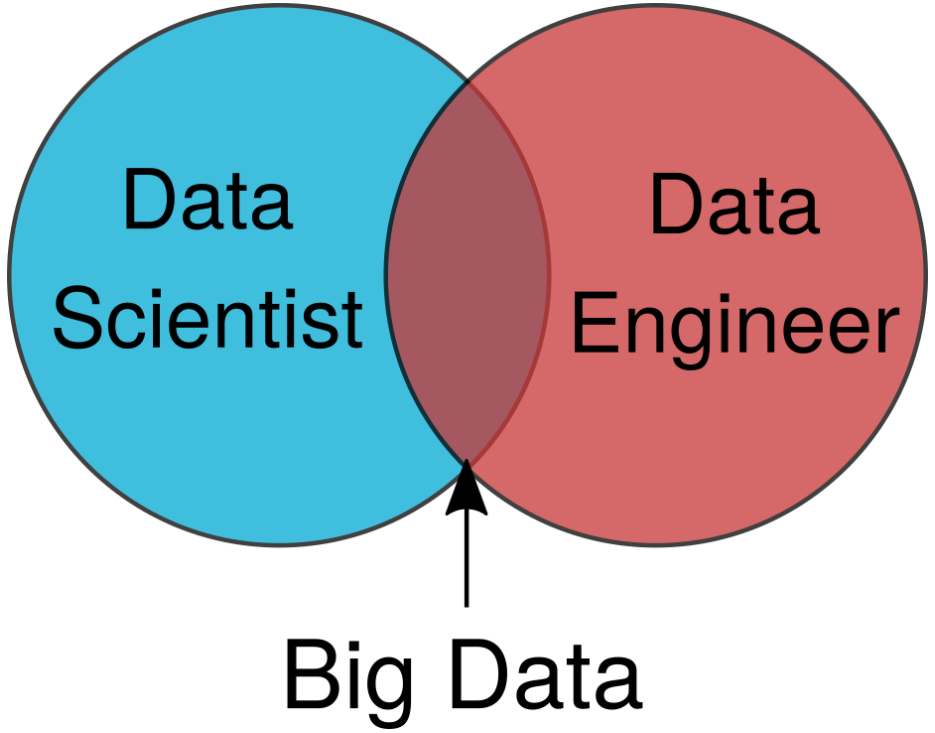
Mas afinal, o que é Data Science?

- **Data Science ou Ciência de Dados** é uma ciência interdisciplinar sobre o processamento de grandes conjuntos de dados usando métodos estatísticos para extrair insights sobre os dados brutos.



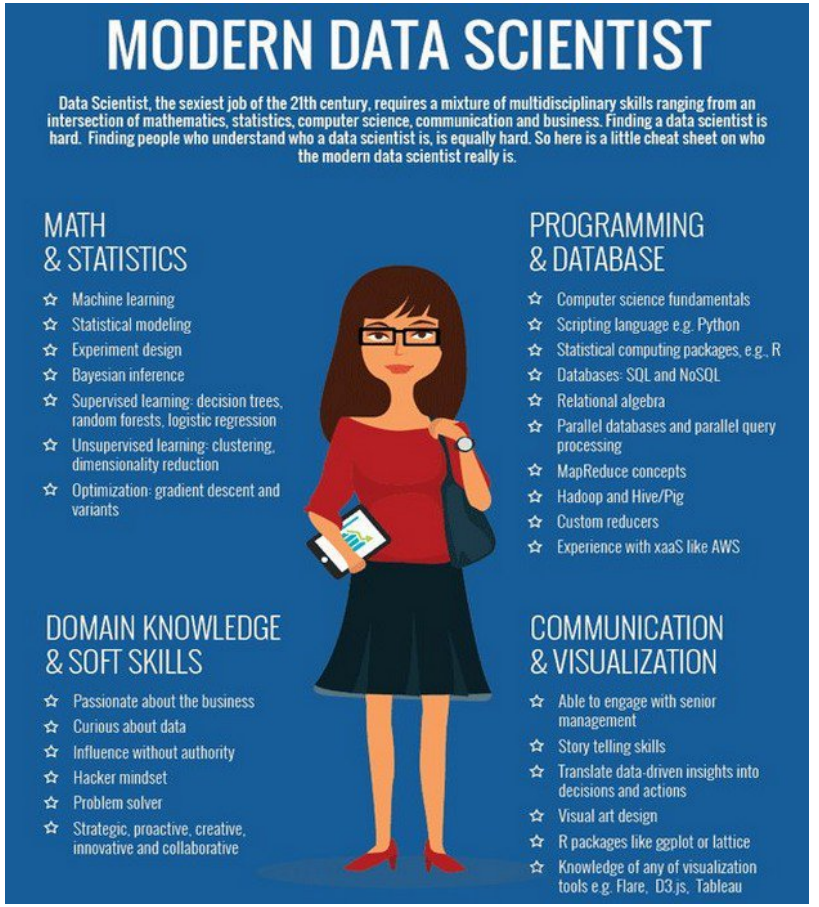
Data Science, Data Engineer: Big Data

- **Data Engineer ou Engenharia de Dados** é a área que se dedica a superar os “gargalos” de processamento de dados e problemas de manuseio de dados para aplicações que utilizam grandes volumes, variedades, e velocidades de dados.



Data Scientist

- **Data Scientist ou Cientista de Dados** é quem extrai insights de dados brutos (row data, messy data).



Perfil do Cientista de Dados

- **Rachel Schutt**, uma das autoras do livro "**Doing Data Science**", criou um gráfico para visualizar a si mesma como cientista de dados.

- **Reflexão:**

- Onde está o seu perfil de cientista de dados no momento?
- Onde você gostaria de estar daqui meses ou anos?

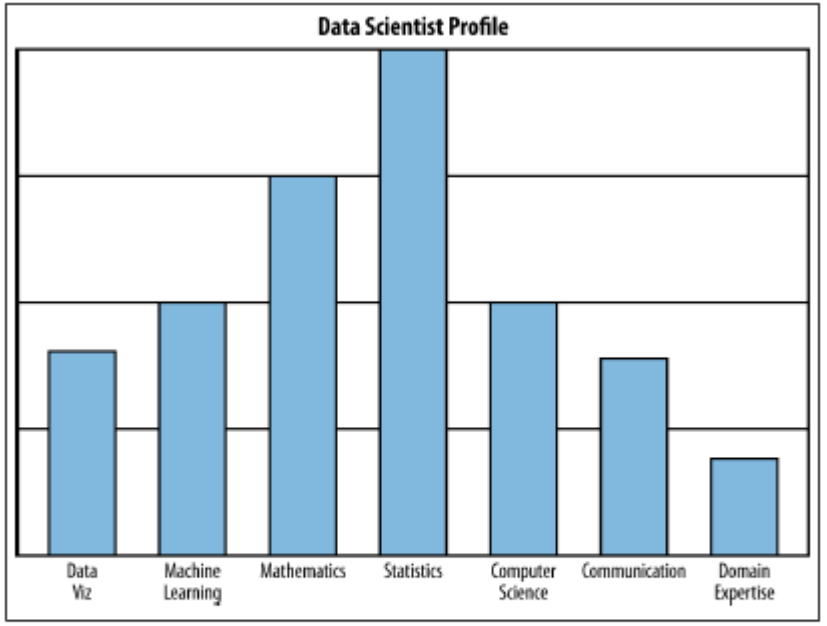
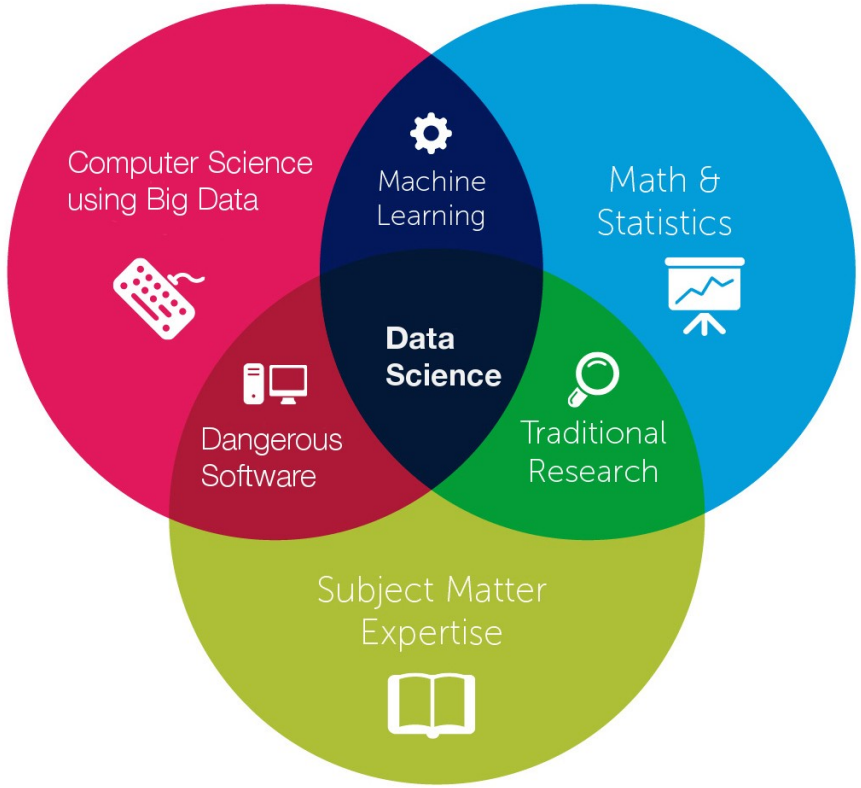


Figure 1-2. Rachel's data science profile, which she created to illustrate trying to visualize oneself as a data scientist; she wanted students and guest lecturers to "riff" on this—to add buckets or remove skills, use a different scale or visualization method, and think about the drawbacks of self-reporting

Perfil do Cientista de Dados

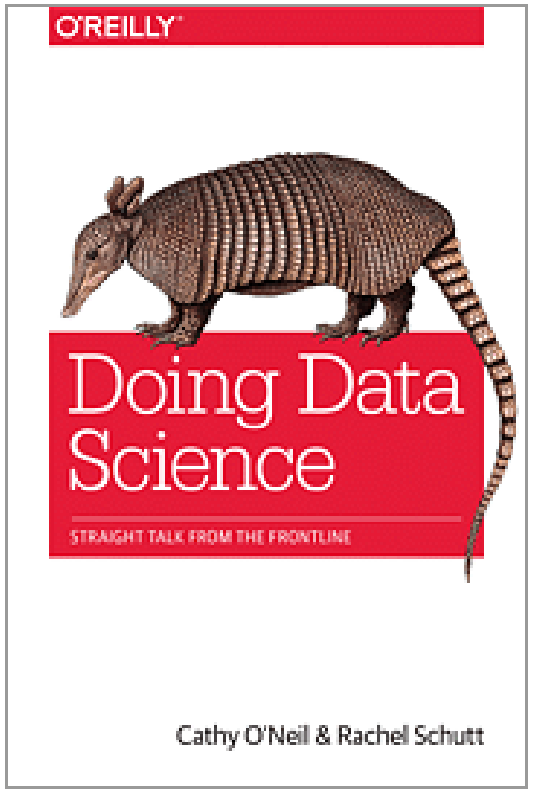
- Ciência da Computação
- Matemática
- Estatística
- Machine Learning (Aprendizagem de Máquina)
- Conhecimento da área que será “investigada”
- Habilidades de comunicação e apresentação
- Visualização de dados



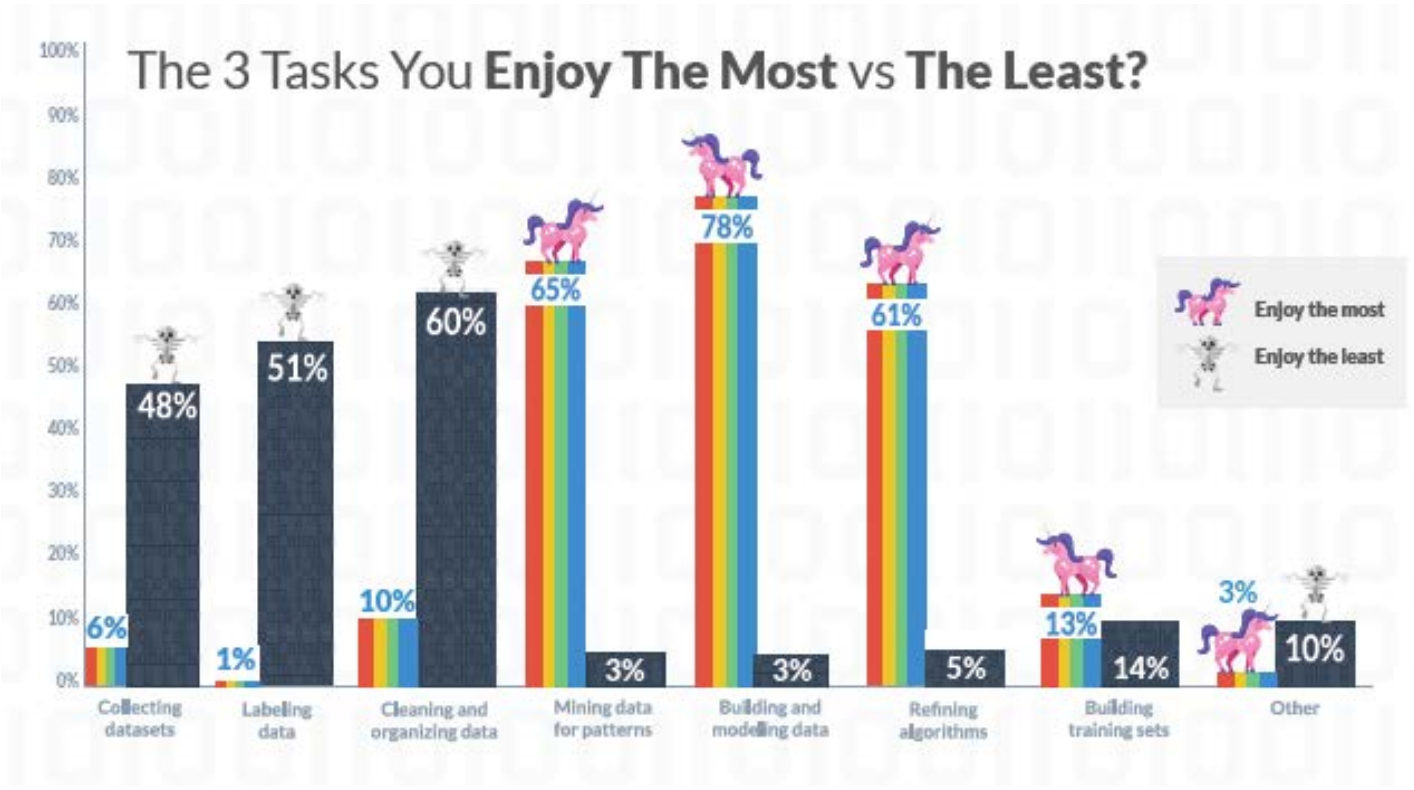
Perfil do Cientista de Dados

- **Opinião das autoras do livro “Doing Data Science”:**

“Uma equipe de ciência de dados funciona melhor quando diferente habilidades (perfis) são representadas em diferentes pessoas, porque ninguém é bom em tudo. Isso nos faz pensar se poderia ser mais vale a pena definir uma “equipe de ciência de dados”, do que definir um cientista de dados”.



Perfil do Cientista de Dados



Data Scientist Report 2017 - CrowdFlower

Data Science com Python, por quê?

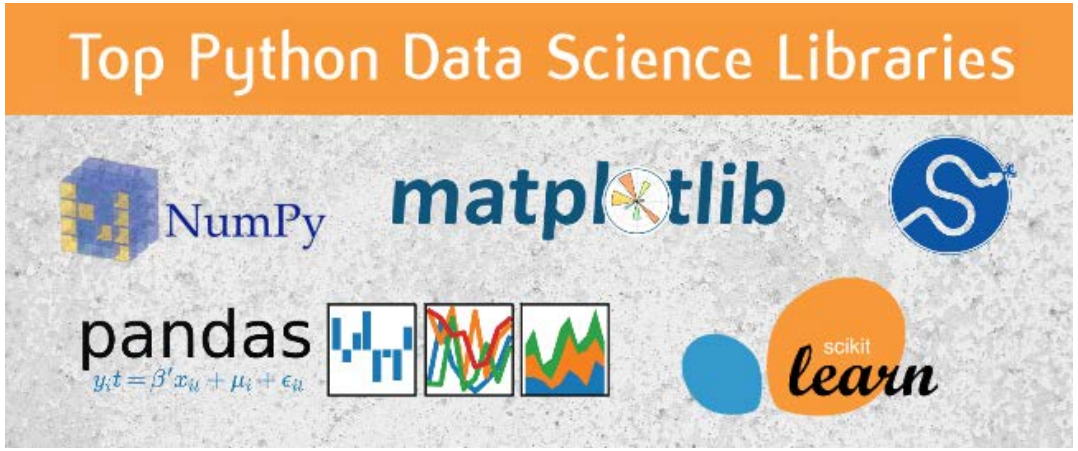
O **Python** tem vários **recursos** que o tornam adequado para aprender (e fazer) ciência de dados:

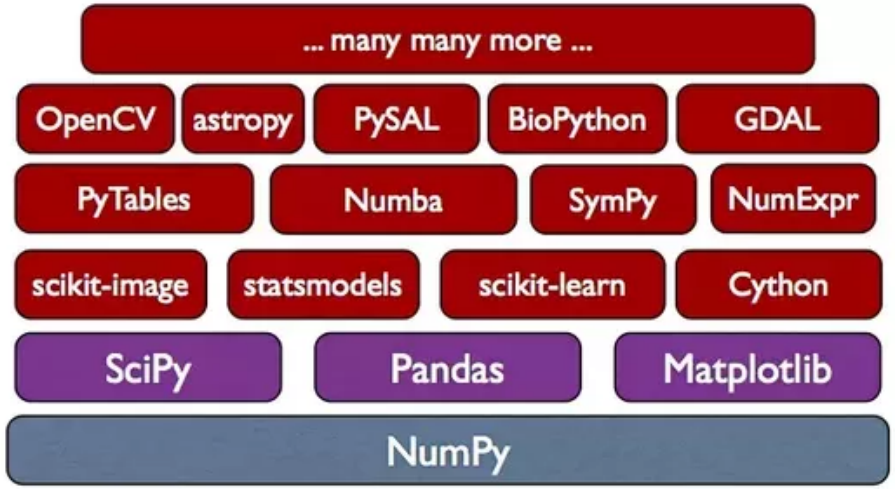
- É grátis ☺
- É relativamente simples codificar
- Tem grande intensidade computacional e **poderosas bibliotecas** de análise de dados
- Se integra bem com outras bases de dados e ferramentas usadas pelo Engenheiro de Dados (como o Hadoop e Spark, por exemplo).

Data Science com Python

Há muitas bibliotecas de dados, frameworks, módulos e toolkits

que implementam eficientemente algoritmos e técnicas de Data Science.





Bibliotecas do Python para Data Science

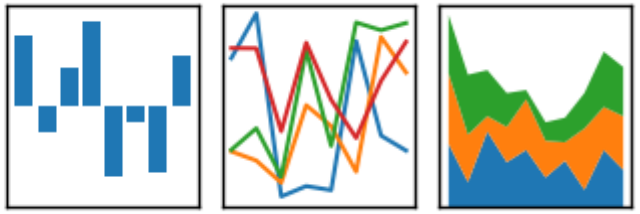
- **NumPy:** suporte para **Python numérico**. A característica mais poderosa de NumPy é o **array n-dimensional**. Esta biblioteca também contém funções básicas de **álgebra linear**, **transformações de Fourier**, capacidades avançadas de números aleatórios e ferramentas para integração com outras linguagens de baixo nível, como Fortran, C e C ++.

Bibliotecas do Python para Data Science

- **Pandas:** para operação e **manipulação de dados estruturados**. É amplamente utilizado para **preparação de dados**. A biblioteca Pandas foi adicionada há relativamente pouco tempo no Python e têm sido fundamentais para impulsionar o uso do Python na comunidade de cientistas de dados.

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Bibliotecas do Python para Data Science

- **SciPy:** suporte para **Scientific Python**. SciPy é construída sobre NumPy e é uma das bibliotecas mais úteis para uma variedade de ciências de alto nível e engenharia como transformação de Fourier discreta, Álgebra Linear e otimização e matrizes esparsas.

Bibliotecas do Python para Data Science

- **Matplotlib:** para traçar **grande variedade de gráficos**, desde histogramas até gráficos de calor.

- **Scikit Learn:** para a **Machine Learning ou Aprendizagem de Máquina**. Construído sobre NumPy, SciPy e matplotlib, esta biblioteca contém uma grande quantidade de ferramentas eficientes para aprendizado de máquina e **modelagem estatística**, incluindo classificação, regressão, clustering e redução de dimensionalidade.

Ferramentas do Python para Data Science

- **O que é Anaconda?**

A Anaconda é uma **distribuição Python** (ou R) que possui uma série de **ferramentas para Ciência de Dados**, Análise Preditiva, Computação Científica e Machine Learning. Inclui o núcleo da linguagem Python (no caso da versão Python da distribuição) e, ainda, mais de **100 bibliotecas Python**, um editor de código chamado **Spyder**, o **Jupyter notebook** e o **Conda** que é um gerenciado de pacotes do Anaconda.

Mulheres Brasileiras Cientistas de Dados

- **Jessica Temporal**

Python Developer, Data Scientist and Bachelor in Biomedical Informatics

@jesstemporal

- **Patricia Novais**

Physicist, Astrophysicist, Data Scientist, Ballerina, PyLady, Overleaf Advisor!

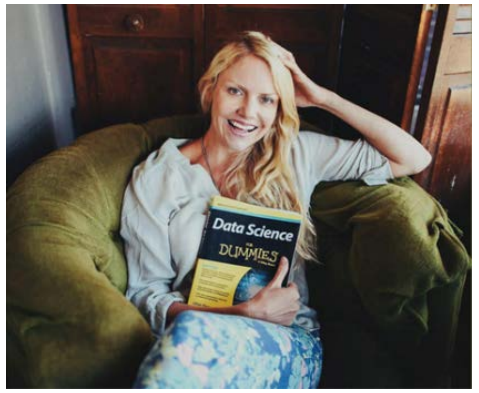
@alphapaty

Mulheres Cientistas de Dados

- **Lillian Pierson**

Data scientist and professional environmental engineer

@BigDataGal



- **Rachel Schutt**

Managing Director at BlackRock where she leads Data Science

www.linkedin.com/in/rachelschutt

Mulheres Cientistas de Dados

- **Hilary Manson**

GM for Machine Learning at @Cloudera. Founder at @FastForwardLabs. Data Scientist in Residence at @accel. I ♥ data and cheeseburgers.

@hmason

- **Lorena Mesa**

A diretora da Fundação Python Software é também coorganizadora da PyLadies Chicago e da Tech Ladies.

@looooorenanicole

Guia de Bolso:
“Data Science com Python Para Você”

- **Estatística**
- Curso Udacity: Introdução à Estatística Descritiva
- Curso Udacity: Introdução à Estatística Inferencial
- Livro: Guia Mangá de Estatística ❤
- Blog: O Estatístico

- **Python Básico**
- Curso Coursera/USP (Português): Introdução à Ciência da Computação com Python Parte 1
- Curso Coursera/USP (Português): Introdução à Ciência da Computação com Python Parte 2
- Curso SoloLearn: Python
- Curso Kaggle: Python
- Curso em vídeo (Português): Python para Zumbis

Guia de Bolso:
“Data Science com Python Para Você”

- **Data Science**
- Curso Data Science Academy: Introdução à Ciência de Dados 2.0
- Podcast: Pizza de Dados
- Blog: Cientista de Dados com GIFs

Guia de Bolso:
“Data Science com Python Para Você”

Data Science com Python

- Curso Data Science Academy (Português): Python Fundamentos para Análise de Dados
- Cursos Kaggle: Pandas, Machine Learning e Data Visualization
- Trilha Cognitive Class (3 cursos): Applied Data Science with Python
- Disciplina de Python do curso de Engenharia Civil da UFPR(4 cursos em Português): Introdução à Computação Científica com Python
- Curso em vídeo LabHacker (para leigos/Português):

  Análise de Dados em Python: Aula 01

  Análise de Dados em Python: Aula 02
- Blog (Português): Tutorial completo para aprender Data Science com Python do zero
- Blog: Get start with Pandas

Grupo de Estudo de Ciência de Dados das PyLadies São Paulo

- O grupo iniciou as atividades em 07/2018.
- Atualmente estamos testando caminhos para uma aprendizagem focada nos conceitos de **Estatística com o Python na prática** (bibliotecas para Ciência de Dados: Numpy e Pandas).
- **Teleconferências** via Zoom toda terça-feira às 21 h.
- E também temos um canal no **Slack**, porque somos chiques. ☺
- "Agora é melhor que nunca" – Zen do Python

Data Science

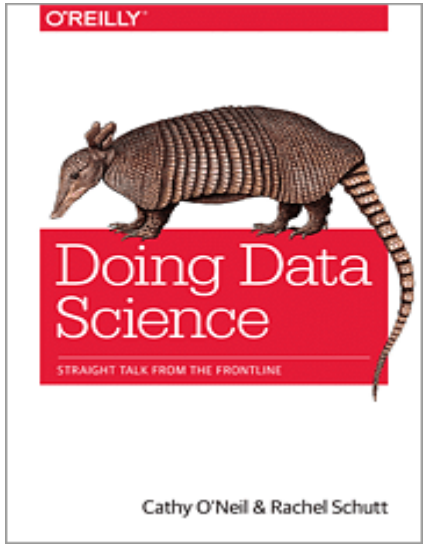
pyladies São Paulo

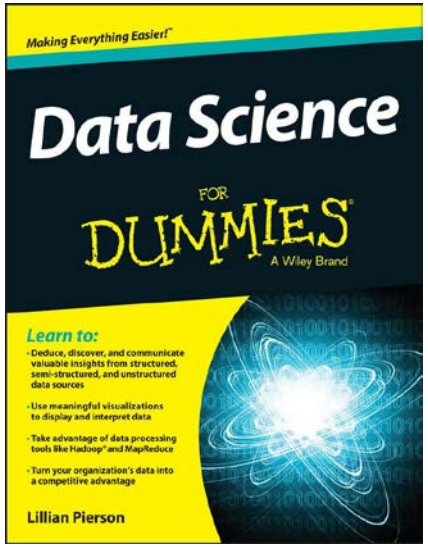
Referências

- **Livros**

Doing Data Science, autoras: Cathy O'Neil & Rachel Schut

Data Science For Dummies, autora: Lillian Pierson





Referências

- **Sites**

https://www.vooo.pro/insights/um-tutorial-completo-para-aprender-data-science-com-python-do-zero/

http://felipegalvao.com.br/blog/2016/02/29/manipulacao-de-dados-com-python-pandas/c

https://www.quora.com/What-is-the-relationship-among-NumPy-SciPy-Pandas-and-Scikit-learn-and-when-should-I-use-each-one-of-them

https://dadosedecisoes.com.br/anaconda/

Referências

- **Figuras**

Data_Scientist_Report_2017_CrowdFlower

https://www.kisspng.com/png-data-science-data-analysis-analytics-big-data-data-5415739/

https://www.oreilly.com/ideas/data-engineers-vs-data-scientists

http://datadriven.tv/blog/modern-data-scientist-infographic/

http://www.discoversdk.com/blog/top-python-data-science-libraries

https://medium.com/@sunnerli/get-start-with-pandas-822db89705c9

https://www.fullstackpython.com/scipy-numpy.html

https://softwareengineeringdaily.com/2016/02/01/matplotlib-with-ben-root/

http://wasduk.com/old-floppy-disk/

https://medium.com/cutshort/how-to-become-a-data-scientist-a-detailed-step-by-step-guide-635b079937e2

Referências

- **Figuras**

Data_Scientist_Report_2017_CrowdFlower

https://www.kisspng.com/png-data-science-data-analysis-analytics-big-data-data-5415739/

https://www.oreilly.com/ideas/data-engineers-vs-data-scientists

http://datadriven.tv/blog/modern-data-scientist-infographic/

http://www.discoversdk.com/blog/top-python-data-science-libraries

https://medium.com/@sunnerli/get-start-with-pandas-822db89705c9

https://www.fullstackpython.com/scipy-numpy.html

https://softwareengineeringdaily.com/2016/02/01/matplotlib-with-ben-root/

http://wasduk.com/old-floppy-disk/

https://medium.com/cutshort/how-to-become-a-data-scientist-a-detailed-step-by-step-guide-635b079937e2

Just do it!

If you want to go fast,

go alone.

If you want to go far,

go together.

- African Proverb -

Just do it!

And if the music stops
There's only the sound of the rain
All the hope and glory
All the sacrifice in vain
And if love remains
Though everything is lost
We will pay the price,
But we will not count the cost

E se a música parar
Há apenas o som da chuva
Toda esperança e glória
Todo o sacrífico em vão
Se o amor permanecer
Embora tudo esteja perdido
Nós pagaremos o preço,
Mas não nos importaremos com o custo

(Bravado, Rush, 1991)

Muito obrigada!

- **Maria Marinho**

mariamarinhos@gmail.com

https://datascienceforeverybody.tumblr.com

https://github.com/MaryMS