

```
37     if path:
38         self._file = open(os.path.join(job_dir,
39                           path), "r")
40         self._file.seek(0)
41         self._fingerprints.update(self._file.read())
42
43     @classmethod
44     def from_settings(cls, settings):
45         debug = settings.getbool("DEBUG", False)
46         return cls(job_dir(settings), debug)
47
48     def request_seen(self, request):
49         fp = self.request_fingerprint(request)
50         if fp in self._fingerprints:
51             return True
52         self._fingerprints.add(fp)
53         if self._file:
54             self._file.write(fp + "\n")
```

# COMO SE TORNAR UM CIENTISTA DE DADOS?

CARLOS MELO



# CIENTISTA DE DADOS A PROFISSÃO DO FUTURO



PUBLICADO POR  S

PUBLICIDADE CORPORATIVA

## Mercad Science em TI

Por StartHub Press  
© 5 Jun 2019, 12h03 - Atualizado

Harvard Business Review

DATA SCIENCE

Data Scientist: The Sexiest Job of the 21st Century

by Michael H. Gottstein and C.J. Polk

Illustration by Daniel H. Pink

What Data Scientists Really Do, According to 25 Data Scientists

BY MICHAEL H. GOTTMAN AND C.J. POLK

Illustration by Daniel H. Pink

**W**hen Jonathan Goldmar arrived for work in June 2006 at LinkedIn, the business networking site, the place still had the start-up feel. The company had just under 500 members, and the number was growing quickly as existing members invited their friends and colleagues to

3/3

TOP STORIES FROM THE OCTOBER 2012 ISSUE

Harvard Business Review

BIG DATA

EXAME

Centista de dados: a profissão do futuro continua em alta

Por Michael H. Gottstein e C.J. Polk

Illustration by Daniel H. Pink

**W**hen Jonathan Goldmar arrived for work in June 2006 at LinkedIn, the business networking site, the place still had the start-up feel. The company had just under 500 members, and the number was growing quickly as existing members invited their friends and colleagues to

3/3

TOP STORIES FROM THE OCTOBER 2012 ISSUE

Harvard Business Review

BIG DATA



## SOBRE MIM

carlos melo

Hello, World!

Meu nome é **Carlos Melo**, e sou autor do blog sigmoidal.ai, um site dedicado ao universo **Python, Data Science**, e **Inteligência Artificial**.

Sou formado em Ciências Aeronáuticas e em Administração pela AFA, e Mestre em Ciências e Tecnologias Espaciais pelo ITA.

Como **piloto militar**, tive a oportunidade de servir durante 3 anos em um Esquadrão da Região Amazônica e por 4 anos como instrutor de voo da Academia da Força Aérea.

Atualmente, sou **cientista de dados** e **engenheiro de missão de satélite** no **Centro de Operações Espaciais** (COPE) em Brasília, onde analiso tanto os dados técnicos de satélites quanto as imagens obtidas por sensores ópticos.

Durante minha carreira, tive a oportunidade de trabalhar como **Data Scientist** no Comando de Operações Aeroespaciais (COMAE) e **Pesquisador** no Instituto de Estudos Avançados (IEAv).

Caso queria saber mais, me acompanhe nas redes sociais, pois estou sempre postando conteúdos nelas.



## PROFISSÃO CIENTISTA DE DADOS

Sabe quando as pessoas falavam 10 anos atrás que o inglês seria essencial para o mercado de trabalho? Pois é, o inglês se tornou requisito obrigatório para os bons empregos no Brasil e no Mundo.

Hoje eu digo: **quem não souber programar, provavelmente estará fora do mercado de trabalho em 10 anos.**

E sabe qual **a profissão** que se mostra a **mais promissora** para o futuro?  
A profissão de **CIENTISTA DE DADOS**.



## EM ATÉ 10 ANOS MAIS DA METADE DOS EMPREGOS SERÃO EXTINTOS

Especialistas do mundo inteiro estão certos de que **mais da metade dos empregos deixará de existir no horizonte de até 10 anos.**

Transporte autônomo, robôs que investem na Bolsa de Valores, e atendentes virtuais (*chatbots*) são alguns exemplos do que já é realidade hoje.

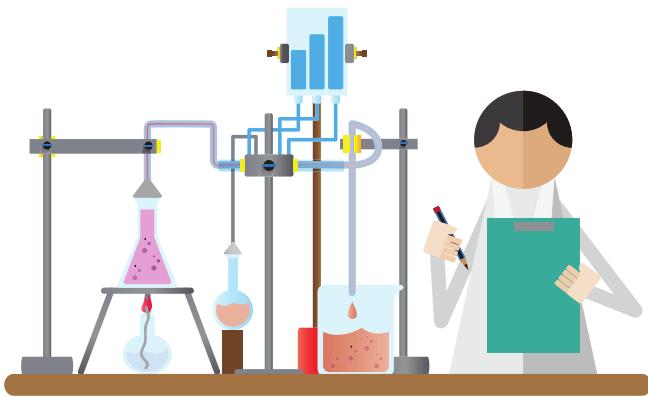
Gradativamente, **as máquinas vêm substituindo os humanos** em atividades cada vez mais complexas. Já existem até escritórios de advocacia e contabilidade substituindo a maior parte de seus recursos humanos por algoritmos de *Machine Learning*.

**Nenhuma indústria está livre da automação,** graças a técnicas avançadas como *Deep Learning*. Fato é que padrões podem ser identificados mesmo em tarefas não repetitivas.

Apesar de estar extinguindo várias profissões tradicionais, a Inteligência Artificial fez surgir uma demanda por um tipo de profissional ainda escasso no mercado, **o cientista de dados**



A **criatividade humana** nunca poderá ser substituída. Um cientista de dados que tenha as habilidades necessárias e senso crítico elevado, com certeza estará inserido nesse cenário futurista que se desenha.



# CIENTISTA DE DADOS:

## O PROFISSIONAL MAIS COBIÇADO DO SÉCULO 21

A profissão de **cientista de dados** (*data scientist*) foi chamada de **"a mais sexy deste século"** pela Harvard Business Review. Se você acompanha as notícias na área de tecnologia, com certeza já viu que as companhias precisam cada vez mais de pessoas que saibam transformar informações em produtos.

Se você já escutou falar dessa profissão, mas ainda não conseguiu entender o que ela faz, pense no cientista de dados como sendo **aquela pessoa que consegue pegar um grande número de dados, extrair informações e gerar conhecimento**.



Média salarial anual do cientista de dados no mundo.

Para você ter ideia de como esse mercado está aquecido, **empresas de recrutamento não estão conseguindo encontrar pessoas qualificadas para a profissão no Brasil** (que oferece salários de até 22 mil reais para especialistas na área).

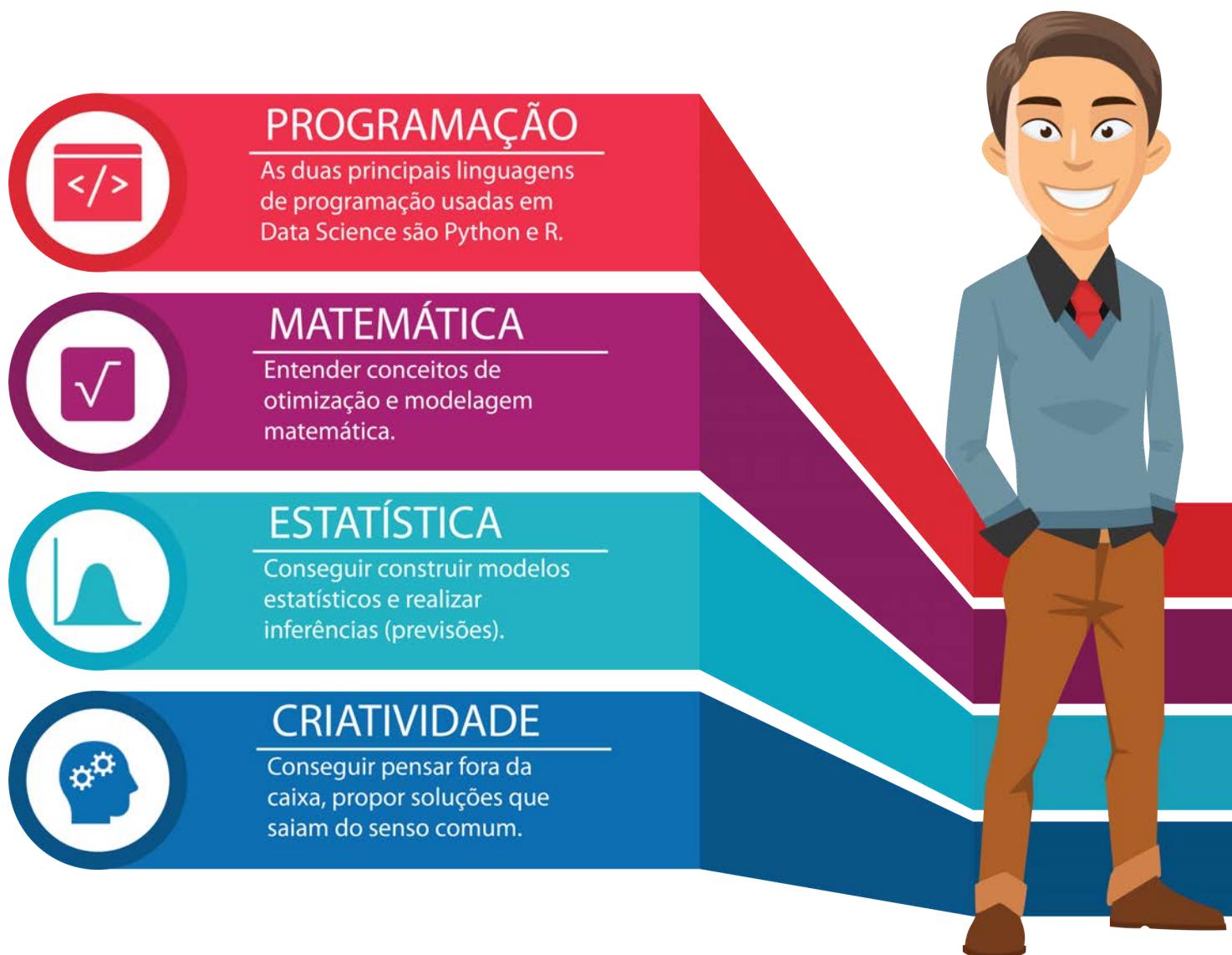
Já no exterior, a média salarial de um cientista de dados fica acima de **11 mil dólares por mês**, segundo o site *Glassdoor*, com ofertas de emprego em Startups e empresas como Facebook, Google e Microsoft.

# HABILIDADES DE UM CIENTISTA DE DADOS

Não existe uma regra sobre quais habilidades um aspirante a cientista de dados deve possuir. O que existe é um **consenso** sobre quais seriam as **principais habilidades**.

Saber **programar**, entender **conceitos matemáticos** (como otimização), compreender os principais **modelos estatísticos**, e pensar com **criatividade** ("pensar fora da caixa") são as características que **vão te destacar** daqueles que só sabem usar códigos prontos.

Hoje em dia, **Python é a linguagem mais usada em Data Science**, e também a mais procurada por empregadores das empresas. Acredite em mim, estude Python!





## O QUE É DATA SCIENCE?

**Nunca antes a humanidade produziu tantos dados como agora.** Cada vez mais nós estamos gerando e armazenando dados que têm um potencial gigantesco de serem transformados em valiosas informações.

Mas **como extrair informações** a partir de uma infinidade *bits* que inundam os bancos de dados em todo o mundo?

É aí que entra essa área conhecida como **DATA SCIENCE**, com suas técnicas e metodologias.

# APLICAÇÕES DATA SCIENCE

Em todas as áreas há a possibilidade de aplicar o Data Science. Seja para ver padrões ou para fazer análises preditivas

Seja para descobrir uma fraude em cartões de crédito, otimizar propagandas em mídias sociais, investir na Bolsa de Valores ou mesmo para criar um modelo preventivo contra falhas de comunicação com satélites em órbita, **Data Science está em todo lugar!**



Apesar de ser mais popular entre *Fintechs* e *Startups*, empresas tradicionais tem aumentado seus investimentos para construir *pipelines*, estruturar seus dados e aumentar o time de Ciência de Dados.

# BIG DATA

## UM OCEANO DE DADOS

O aumento na quantidade de dados e no poder de processamento viabilizaram o *Data Science*.

Modelos matemáticos e estatísticos sempre estiveram presentes no mundo acadêmico e empresarial. Por que esse *boom* agora?!

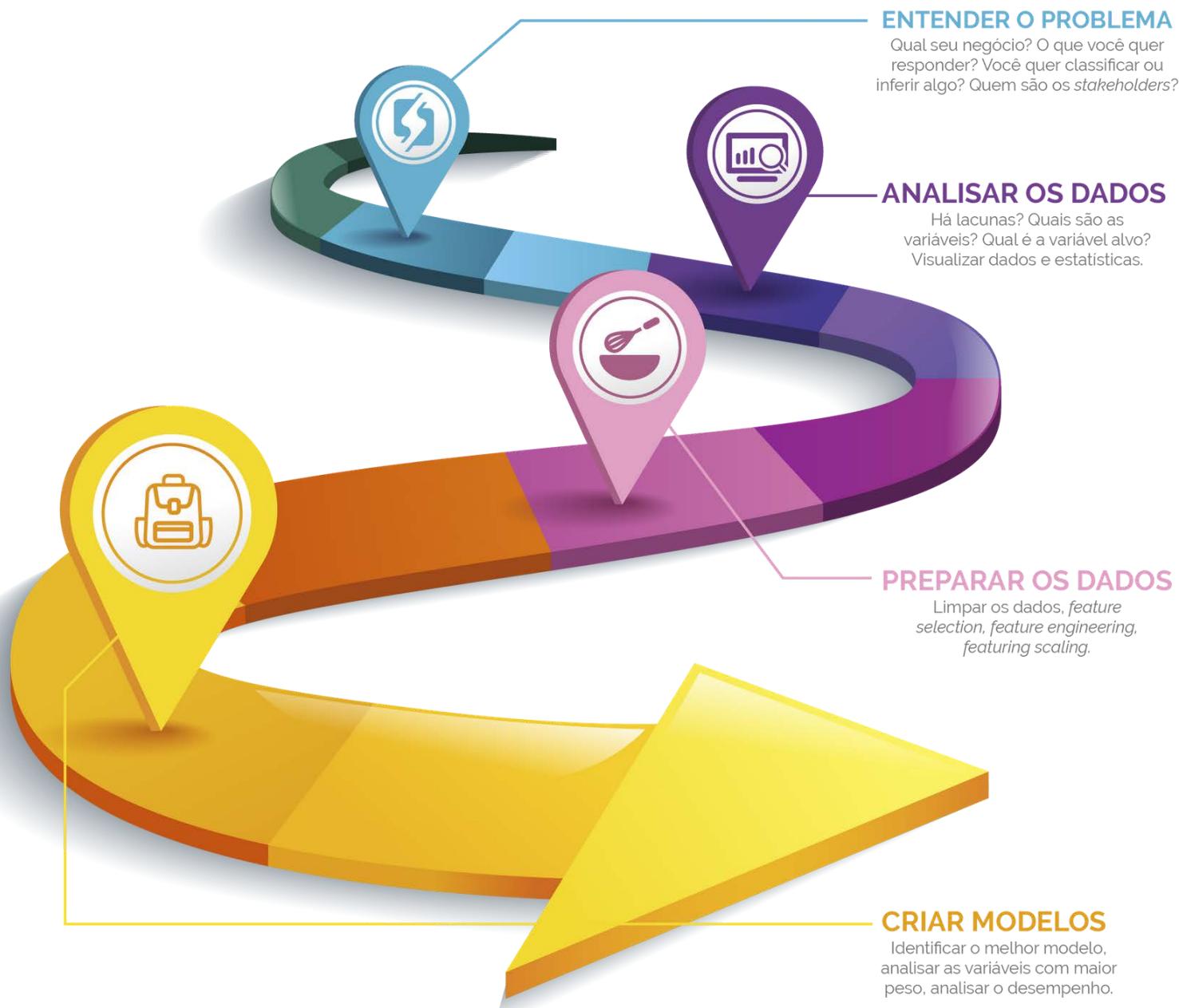
**A explosão** que temos acompanhado para o universo da Inteligência Artificial, *Machine Learning*, *Deep Learning* e *Data Science* **só aconteceu por causa de 2 fatores:**

- **Big Data:** dados em grande volume, nos mais diversos formatos e gerados com velocidade cada vez maior. O acúmulo de dados estruturados e não-estruturados são ativos valiosos para qualquer Organização.
- **Poder de Processamento:** O custo menor e o crescimento exponencial no uso de GPU (Unidades de Processamento Gráfico) fez com que qualquer pessoa pudesse criar Redes Neurais Profundas do seu próprio notebook pessoal.



# QUAL O PROCESSO DO **DATA SCIENCE**

Entender o seu negócio, construir modelos e colocá-los em produção: tudo isso é Data Science



Muita gente que entra no mundo do *Data Science* acha que o principal é construir um modelo de *Machine Learning* para resolver seu problema.

Na verdade, **as etapas mais importantes estão bem antes** disso. *Data Science* significa identificar um problema, definir métricas, analisar dados, limpar e preparar esses dados, e só no final pensar em modelagem.

**±70%**

do tempo de um projeto será gasto nas fases de análise e preparação dos dados

Ser um cientista de dados significa dominar todas as áreas do projeto, além de saber comunicar os resultados para Decisores não-técnicos.

# QUEM PODE APRENDER DATA SCIENCE

Hoje nós vivemos em uma **sociedade cheia de esteriótipos** e premissas falsas. Muita gente me procura nas redes sociais para perguntar se alguém que faz direito, publicidade ou medicina pode aprender a programar.

**Eu sempre costumo afirmar: qualquer pessoa**, de qualquer área, com ou sem formação superior, **consegue aprender programação**. Digo mais, qualquer pessoa é capaz de **se tornar um cientista de dados**.

**"Só quem é de exatas pode aprender a programar?"**



## JEREMY HOWARD

- cientista de dados
- formado em Filosofia
- ex-Presidente do Kaggle
- fundador do fast.ai
- fundador da Enlitic

Um dos maiores cientistas de dados que eu conheço é o **Jeremy Howard**, ex-Presidente do Kaggle (maior site de competições de *Data Science* do mundo). Sabia que ele é **formado em Filosofia**?

Mais um exemplo? No **Projeto Serenata de Amor** (projeto aberto que usa ciência de dados para monitorar contas públicas) está um time que engloba **jornalistas e sociólogos**.

Falando também sobre **equilíbrio e inclusão** no Python, há iniciativas como **PyLadies**, que trouxe uma comunidade cujo propósito é instigar mais mulheres a entrarem na área tecnológica



## PYTHON

Quando se trata de **Inteligência Artificial, Machine Learning e Data Science**, 3 linguagens são comumente citadas:

- **Python**
- **R**
- **Matlab**

Indiscutivelmente, **Python se tornou a mais popular** entre todas, sendo inclusive um dos *hard skills* mais desejados em vagas de emprego para Inteligência Artificial.

# POR QUE APRENDER PYTHON

Python é uma linguagem poderosa, com uma grande comunidade e enorme variedade de bibliotecas



**Python é uma grande ferramenta**, e suas características têm feito a maioria dos cientistas de dados optarem por ela para fazerem suas análises.

Sua **curva de aprendizado** permite que mesmo iniciantes consigam fazer seus primeiros scripts em poucas horas.

Para **Data Science**, a linguagem tem **alta procura** no mercado, sendo **atlamente desejada** pelos empregadores e recrutadores.

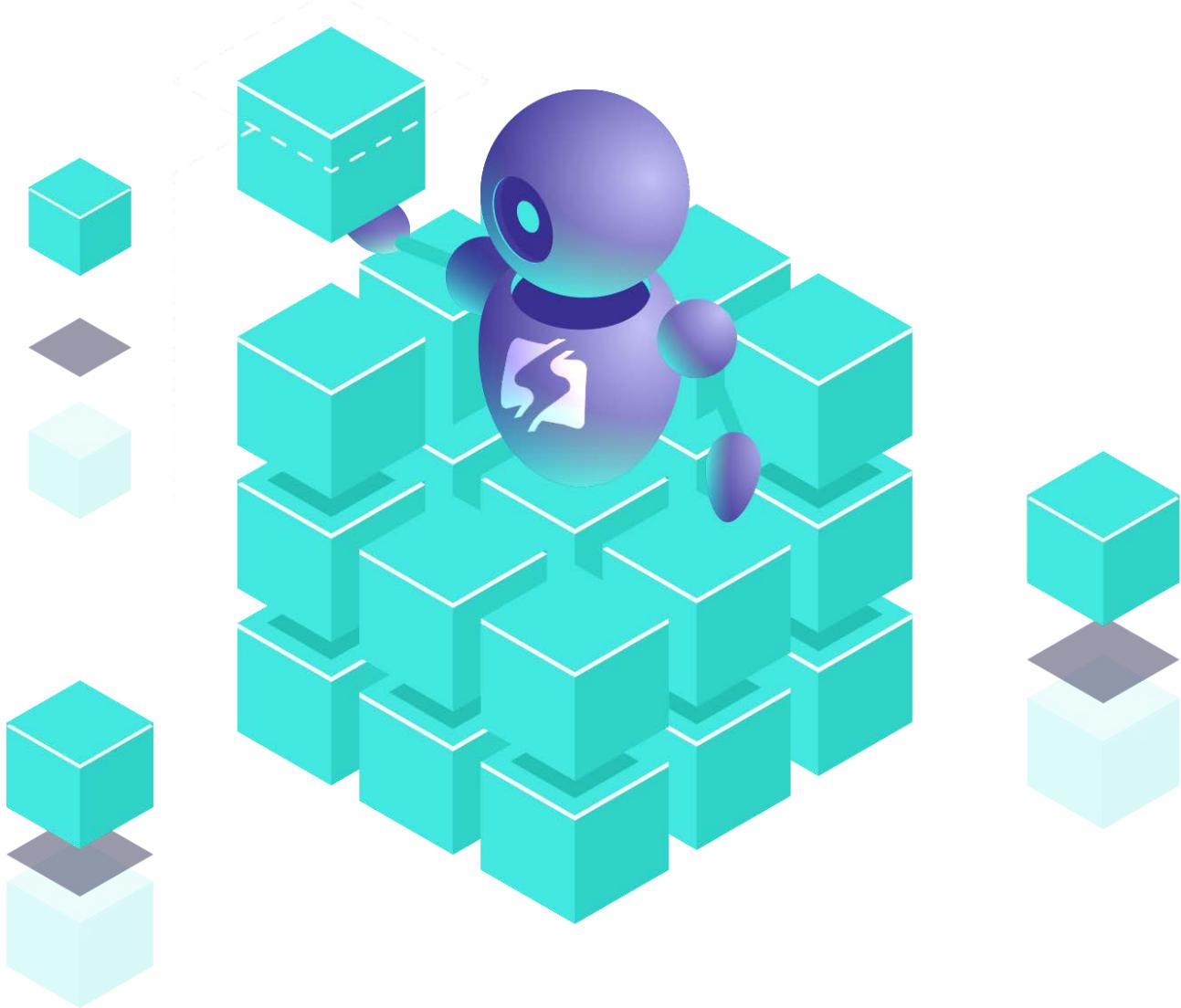


Outro ponto positivo para você, futuro cientista de dados, é que o **Python** é muito poderoso também quando se trata de **visualização de dados**.

A principal biblioteca para plotagem de gráficos é a **Matplotlib**, que consegue gerar facilmente gráficos como o de pizza, barras, linhas e histograma.

Se você quiser uma biblioteca de alto nível, com uma rica galeria de visualizações, séries temporais e outros tipos complexos de gráficos, pode contar ainda com **Seaborn**, **Plotly** e **Bokeh**.

Outra biblioteca que facilita absurdamente durante todas as fases de um projeto de Data Science é o **Pandas**.



O **Pandas**, biblioteca construída com em cima de todas as vantagens do **NumPy**, um **canivete suíço** quando se trata de estrutura de dados.

Ela possibilita você usar uma **estrutura de dados** conhecida como **DataFrame**, que traz uma visualização agradável, em forma de matriz, com os nomes das colunas e dos índices.

Ao importar seus dados brutos em um *DataFrame*, você consegue realizar *slices*, identificar lacunas e valores faltantes, calcular informações estatísticas como média, mediana e desvio-padrão.

É possível até mesmo **plotar gráficos diretamente** da estrutura *DataFrame*, chamado o método adequado.

# ONDE ESTUDAR PYTHON

Há muitos cursos, livros e blogs de altíssimo nível para Python, com foco em *Data Science* e *Machine Learning*

Recebo muitas recomendações de amigos e seguidores do Instagram, porém resolvi deixar uma **lista de cursos e livros que eu fiz ou conheço** (e por isso posso indicar), sempre pensando em ***Data Science***.



## CURSOS

- **Data Scientist Nanodegree (Udacity)** - Os cursos da Udacity são, na minha opinião, os mais bem estruturados e completos do mercado hoje. Com foco na prática e em exercícios da vida real (que tal analisar dados do Airbnb ou da Starbucks), eles te deixam muito alinhado com o que o mercado espera de você.
- **Introduction to Computer Science and Programming Using Python (edX)** - Este foi meu primeiro curso de Python da vida, e recomendo para você. Gratuíto (caso você não queira o certificado), esse curso foi preparado pelo MIT, e tem um foco em Ciências da Computação, o que significa que os exercícios e teoria são mais bem preparados.

## LIVROS

- **Data Science do Zero (Joel Grus)**: Mais focado em ensinar conceitos de probabilidade, estatística e manipulação de dados, te dá uma excelente base antes de você começar a usar bibliotecas prontas (e não saber nem por que está usando).
- **Mãos à Obra. Aprendizado de Máquina com Scikit-Learn e Tensorflow (Aurélien Géron)** - Este livro é fantástico na minha opinião, passando não apenas os códigos, mas explicando com profundidade adequada o motivo de cada coisa. O autor se preocupa ainda em seguir uma metodologia para projetos de *Data Science*.
- **Data Science para Negócios (Foster Provost)** - Entrando nos conceitos matemáticos e técnicos o mínimo possível, o foco deste livro não é programação ou código, mas sim apresentar os princípios fundamentais do Data Science e estimular o pensamento crítico. Quantas vezes não queremos aprender a parte do código e esquecemos os fundamentos?!

## SITES

- **Sigmoidal** - Claro que não poderia deixar meu blog de fora. Estou tentando levar sempre conteúdos de alto nível, abrangendo as áreas de Data Science, Machine Learning, Deep Learning e Visão Computacional.
- **Python para Zumbis** - Começando do básico, a máxima é aprender a programar de um modo profissional e divertido, com a referência Fernando Masaroni.
- **Dan Bader** - Ótimos tutoriais de Python. Aprendi a escrever meus códigos de maneira mais Pythonic com os insights do autor.
- **Real Python** - Para quem está começando a programar, vale a pena acompanhar as dicas aqui, pois te ensinar a fugir do senso comum e pensar o código mais fora-da-caixa.





## CONCLUSÃO

Aprender a programar não acontece da noite para o dia. **Exige dedicação e foco.** Entretanto, nunca antes na história da humanidade, o conhecimento esteve tão disponível e democrático.

A internet possibilita que cursos inteiros de faculdades renomadas estejam disponíveis, assim como vídeos e artigos em blogs (ou Medium).

**Mais de metade das profissões tendem a desaparecer nos próximos 10.** Você tem a sorte de saber disso com antecipação.

Que tal reservar algumas horinhas na semana para se dedicar a algo como **Python e Data Science?**

**Nunca se esqueça que daqui a 10 anos, você vai desejar ter começado hoje ;)**

# SIGMOIDAL

POWERED BY DEEP NEURAL NETWORKS

