

Explorando o

mundo da ciência de dados!

Jéssika Ribeiro

mentorama.

Sobre mim

- graduação e mestrado em matemática
- saída do mestrado
- estudei mto por conta propria- cursos online e livros
- dificuldades computacao-business..
- busquei emprego e consegui a consultoria
- falar sobre consultoria
- falar sobre itau
- falar sobre picpay
- faltar tempo de exp
- o que aprendi em cada uma...o que mais me ajudou
- tipos de projetos que ja trabalhei, clientes grandes, diversos ramos..

Overview do curso

Conteúdo

1. O mundo de dados. (Jéssika)

A era da informação
Introdução à Ciência de Dados
Introdução a Python
Instalacao do ambiente

2. Fundamentos básicos de Python 1: (Jéssika)

Variáveis
Operadores
Estruturas básicas de dados
Estruturas de decisão e repetição
Métodos e Funções

3. Fundamentos básicos de Python 2: (Jéssika)

Módulos e pacotes
Introdução a orientação a objetos em Python
Classes e objetos
Tratamento de erros

4. Trabalhando com bancos de dados (Jéssika)

Conexão com banco de dados: SQLite3, MongoDB
Trabalho com banco de dados.
Leitura e gravação de dados.

5. Bibliotecas: Numpy e Pandas (Jéssika)

Overview sobre as bibliotecas.
Trabalhando com numpy.
Trabalhando com pandas.

6. Análise e Visualização de dados (Jéssika)

Estatística descritiva: Teoria
Estatística descritiva: Prática
Visualização de dados com matplotlib

7. Pré-processamento de dados (Fabio)

Preenchimento de dados faltantes
Normalização de dados
Detecção de outliers

8. Fundamentos de machine learning (Fabio)

O que é machine learning?
Principais desafios ao se trabalhar com machine Learning
Fundamentos básicos de machine learning

9. Machine learning na prática (Fabio)

Trabalhando em um projeto de data science com machine learning de ponta a ponta
Fundamentos de modelos de Regressão

10. Modelos de Classificação (Fabio)

Principais classificadores
Métricas para classificação
Análise de resultados

11. Implementando modelos de Machine Learning do zero (Fabio)

Regressão Linear
Regressão Polinomial
Regularização
Regressão Logística

12. Modelos de árvores e ensambles (Fabio)

Árvores de decisão
Random Forest
Modelos de Boosting

13. Clusterização (Fabio)

Aprendizado não supervisionado
KMeans
Métricas para KMeans
DBSCAN

14. Aplicações: Detecção de Anomalias (Fabio)

Fundamentação estatística para a aplicação
Modelos não supervisionados para detecção de anomalias

15. Aplicações: Sistemas de recomendação (Fabio)

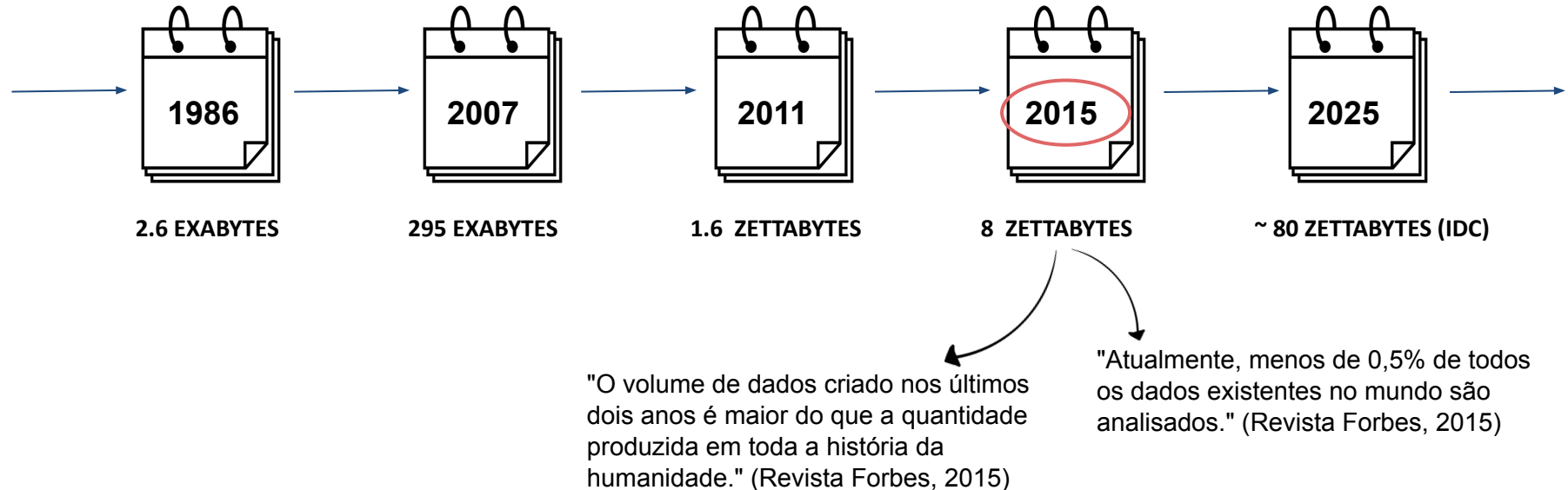
Recomendações baseadas nos usuários
Recomendações baseadas nos produtos

16. Gerenciamento de projetos de dados (Jéssika)

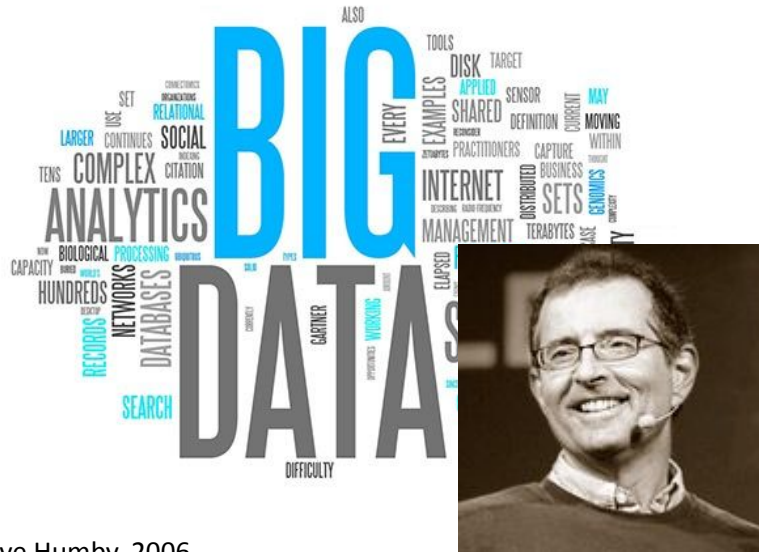
O que é um projeto de ciência de dados?
Ciclo de vida de um projeto de ciência de dados;
Definição das fases de desenvolvimento e deploy um modelo;
Boas práticas: Organização de arquivos, nomenclatura, documentação e versionamento.
Monitoramento

A era da informação

Sempre produzimos dados! O que mudou agora?



A era dos dados



Leaders

May 6th 2017 edition >

Regulating the internet giants

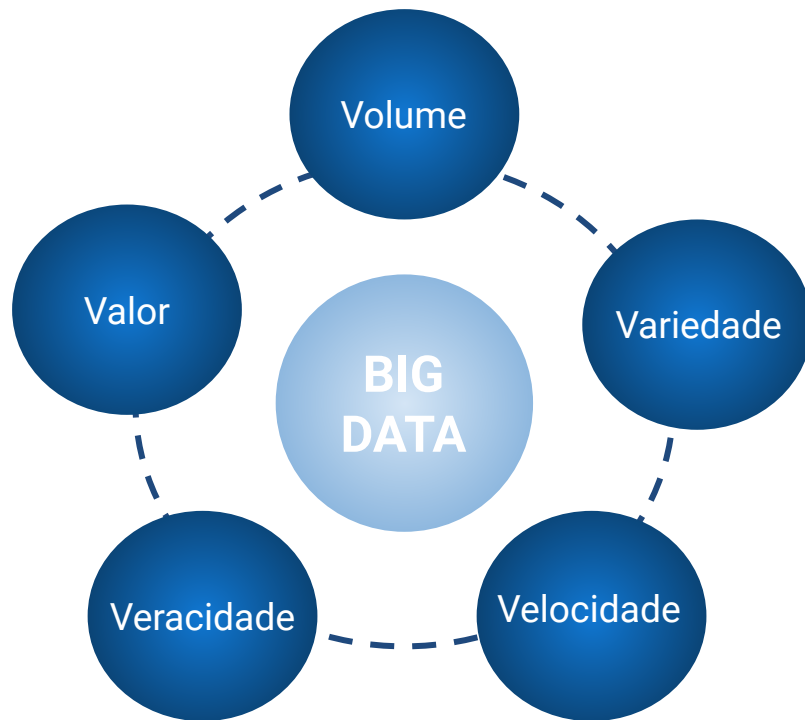
The world's most valuable resource is no longer oil, but data

Data is the new oil!

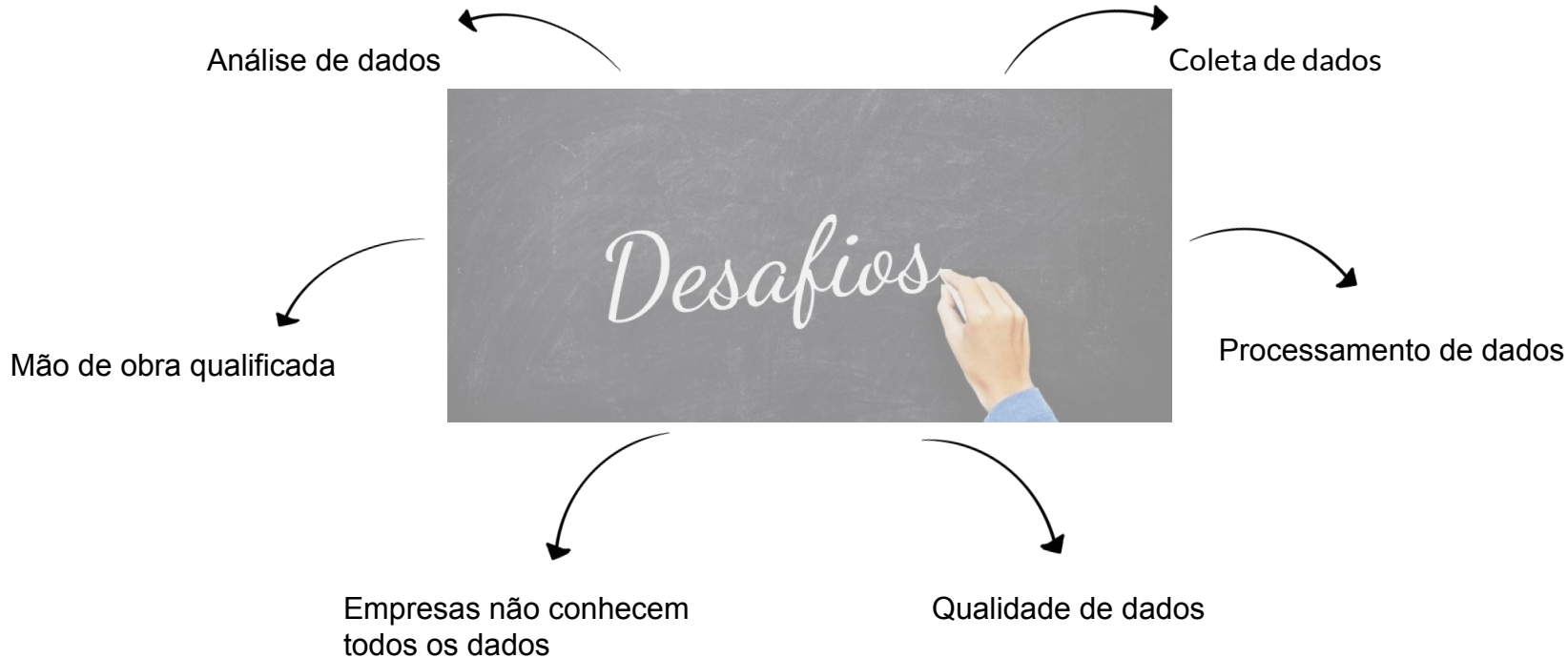
O que é BIG DATA?

*Big Data é um conceito que descreve o fenômeno de um **grande conjunto** de dados, **estruturados ou não**, que são gerados a cada **segundo**, por **diversas fontes** e que precisam de **métodos não tradicionais** para armazenamento e processamento. Esses dados permitem que informações sejam cruzadas e analisadas para gerar **insights**.*

Os famosos V's do Big Data



Os desafios do Big Data

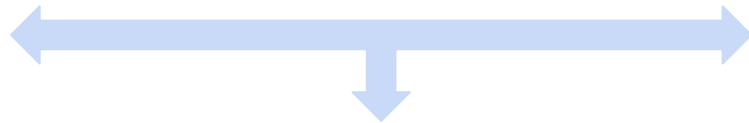


BIG DATA vs DATA SCIENCE

"Coletar é diferente de extrair Informação útil"

BIG DATA: Coletar, armazenar e gerenciar grandes volumes de dados

DATA SCIENCE: Extrair, analisar e gerar informação para auxiliar na tomada de decisão.



Quanto mais informação útil e de qualidade, melhor serão os modelos e análises

Data Science

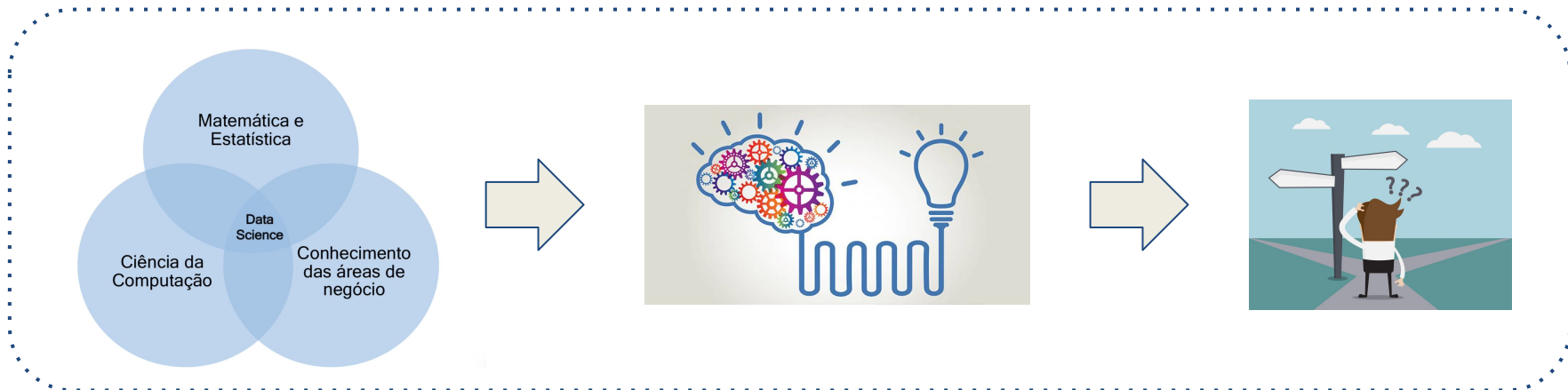
Data Science

Ciência de dados (em [inglês](#): *data science*) é uma área *interdisciplinar* voltada para o estudo e a análise de dados econômicos, financeiros e sociais, *estruturados e não-estruturados*, que visa a *extração de conhecimento*, *detecção de padrões* e/ou obtenção de variáveis para possíveis *tomadas de decisão*. (Wikipedia)

Data Science é explorar o big data para prever o futuro!



Data Science



Área multidisciplinar cujo objetivo é extrair conhecimento dos dados para auxiliar a tomada de decisão

O famoso cientista de dados



Artwork: Tamar Cohen, Andrew J. Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2012 issue

"What kind of person does all this? What abilities make a data scientist successful? Think of him or her as a hybrid of data hacker, analyst, communicator, and trusted adviser. The combination is extremely powerful—and rare."

Quem é esse profissional?

**Visão de
Negócio**

Cientista da computação

Matemático

Estatístico

Técnico

Crítico

Analítico

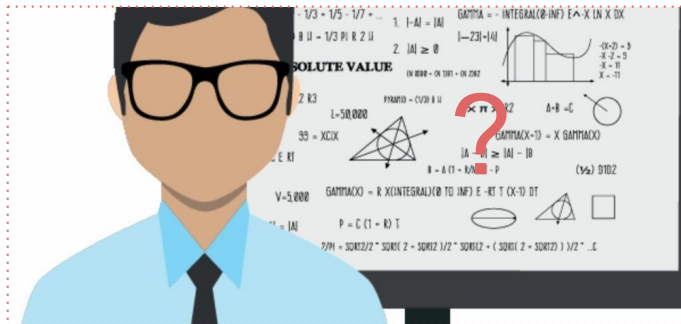
**Lifetime
learner**

Criativo

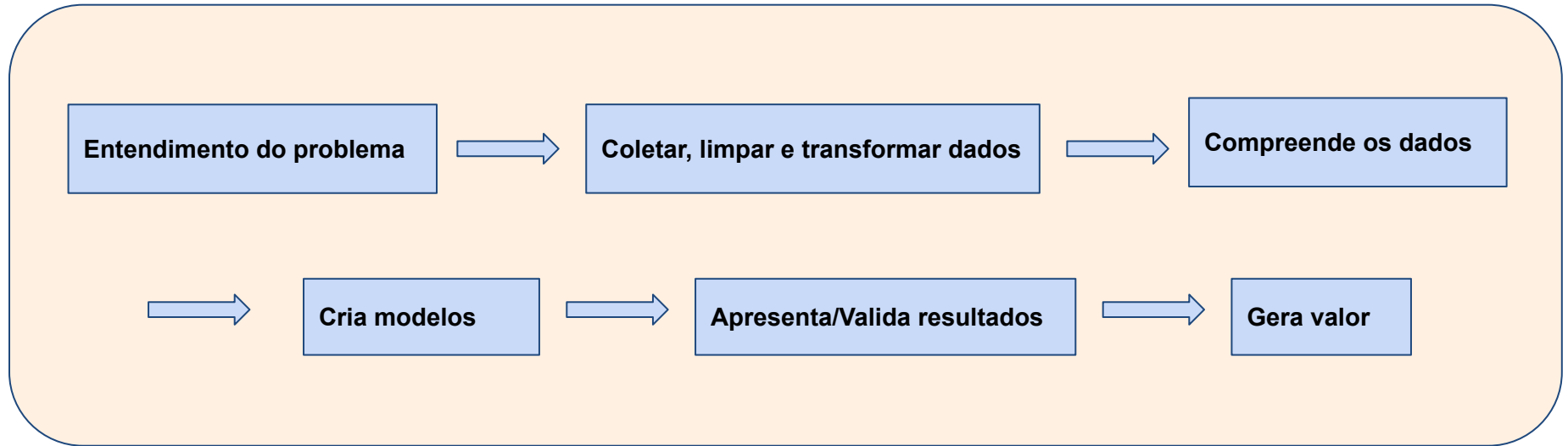
Comunicativo

Curioso

Consultor



Dia-a-dia de um cientista



Ciência de dados na vida real

Grandes players



Estratégias de preço e campanhas de marketing



Recomendação de filmes



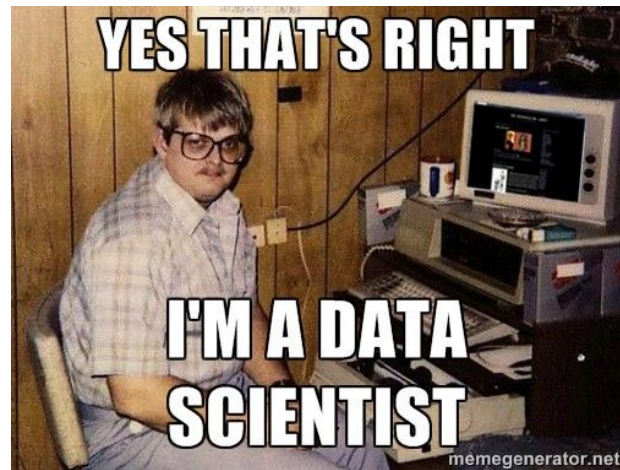
Personalização de experiência de compra.

Aplicações comuns

- Predição de demanda
- Logística
- Sistemas de recomendação de produtos
- Detecção de fraude
- Sistemas de busca mais inteligentes
- Otimização de campanhas

Porque se tornar um cientista?

- Profissões mais promissoras do século XXI;
- Desde startups à empresas mais conservadoras, todas estão criando suas áreas de dados com grandes equipes de ciência de dados.
- Desde 2018 a busca cientistas de dados já era muito alta nos EUA, com salários que giram em torno dos 100k anuais.
- O mercado de ciência de dados está extremamente aquecido.



Salários de Cientista De Dados

125 Salários Atualizado em 9 de jul 2020

Salário médio

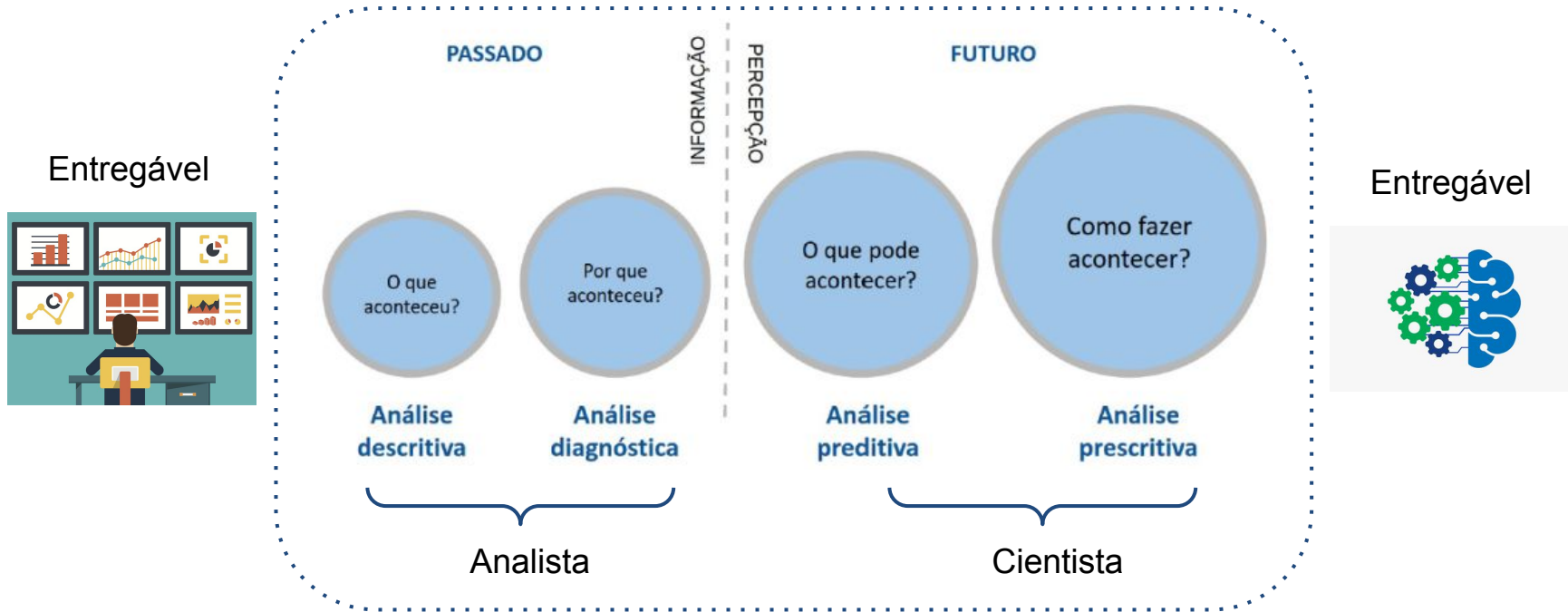
R\$ 7.285 / mês



Quanto ganha um Cientista De Dados?

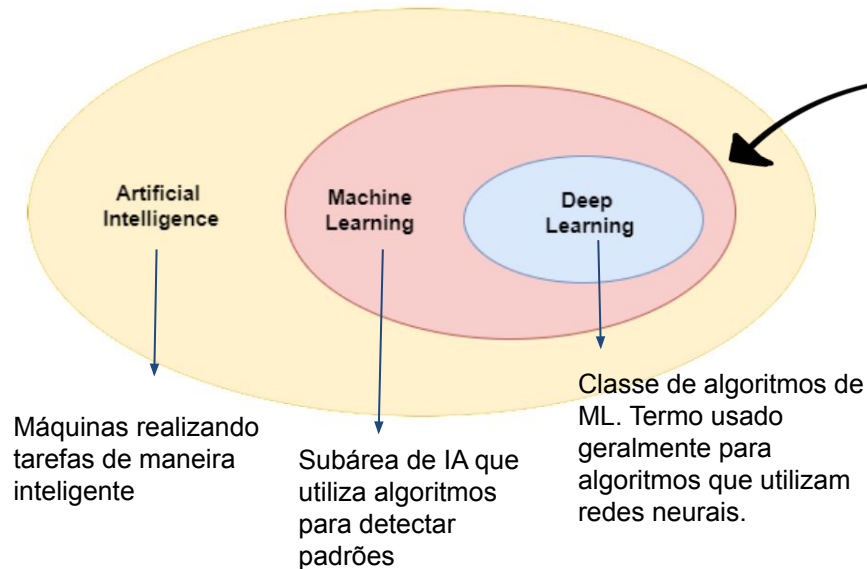
O salário médio nacional de Cientista De Dados é de R\$7.285 em Brasil. Filtre por localização para ver os salários de Cientista De Dados na sua região. As estimativas de salários têm como base os 125 salários enviados de forma sigilosa ao Glassdoor por funcionários com o cargo de Cientista De Dados. [Menos](#)

O analista e o cientista:



Data Buzzwords

Qual a diferença entre IA, Machine Learning, Deep Learning e Data science?

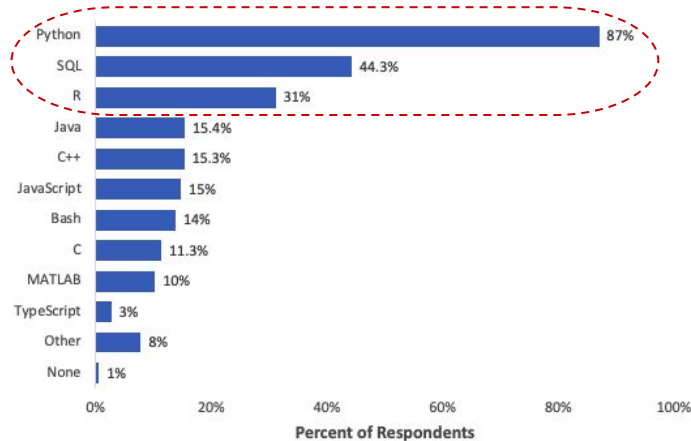


ML e DL são áreas de conhecimento que cientistas de dados utilizam para entendimento dos dados e construção de modelos.

Introdução ao Python

Linguagens

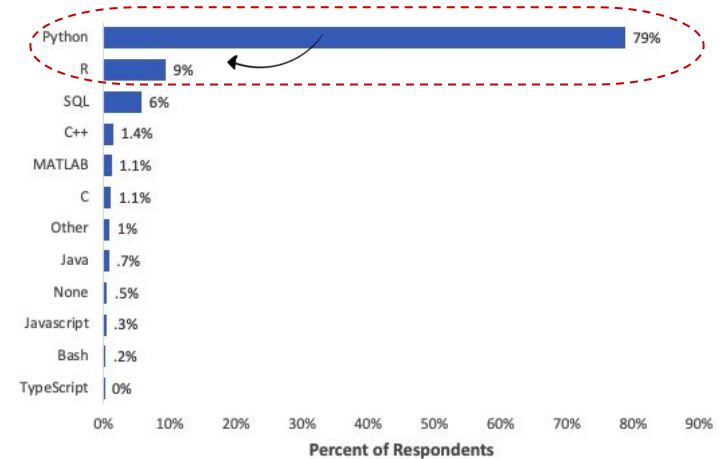
What programming languages do you use on a regular basis?



Note: Data are from the 2019 Kaggle ML and Data Science Survey. You can learn more about the study here: <https://www.kaggle.com/c/kaggle-survey-2019>.

A total of 19717 respondents completed the survey; the percentages in the graph are based on a total of 14762 respondents who provided an answer to this question.

What programming language would you recommend an aspiring data scientist to learn first?



Note: Data are from the 2018 Kaggle ML and Data Science Survey. You can learn more about the study here: <https://www.kaggle.com/c/kaggle-survey-2019>.

A total of 19717 respondents completed the survey; the percentages in the graph are based on a total of 14377 respondents who provided an answer to this question.

Python

- Python foi criada por Guido van Rossum em 1991, no Centro de Matemática e Computação, em Amsterdã;
- Possui uma sintaxe concisa e clara, cujo objetivo é ser simples e de fácil aprendizado;
- Suporta a maioria das técnicas da programação orientada a objeto.
- Case sensitive: sensível a letras maiúsculas e minúsculas.
- Interpretada
- Linguagem de alto nível;
- Tipagem dinâmica e forte;
- Grande comunidade ativa;
- Pacotes exclusivos para Ciência de Dados (Numpy/Pandas/ scikit-learn)
- Foi considerada pelo público a 3ª linguagem "mais amada", de acordo com uma pesquisa conduzida pelo *site* Stack Overflow em 2018;
- Bibliotecas escritas em subversões, (3.4 e 3.5 por exemplo) geralmente são compatíveis;
- Entre versões diferentes (2 e 3 por exemplo) geralmente são incompatíveis



O ambiente de desenvolvimento

Anaconda

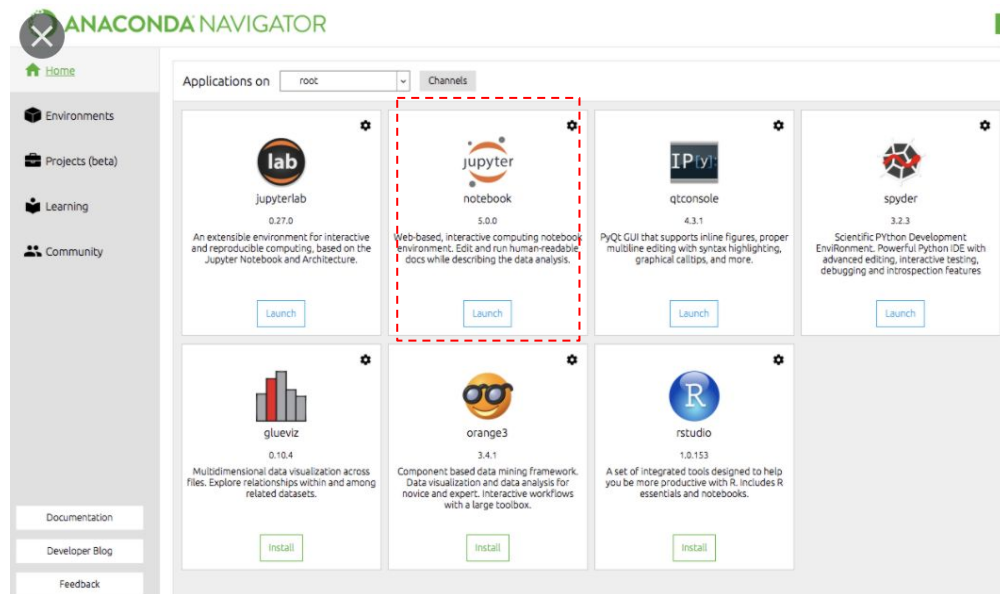


Ferramentas para análise
de dados em um único
arquivo

Praticidade

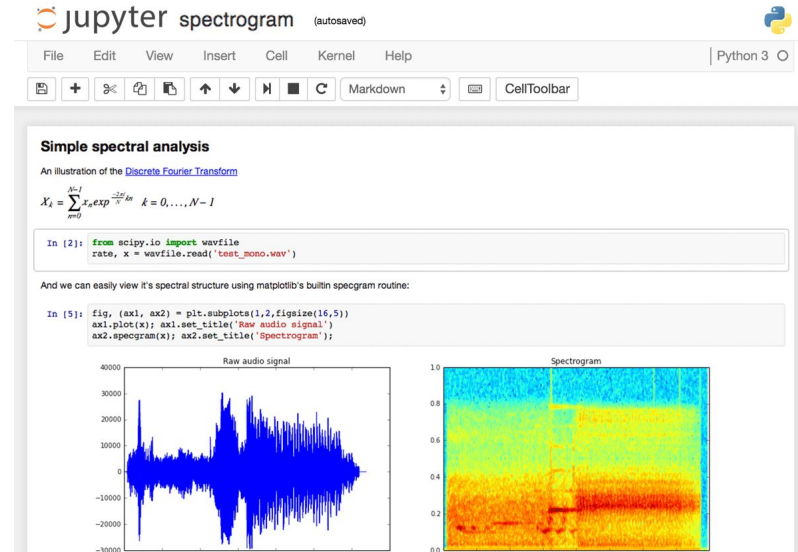
Windows, Linux, e macOS

Free e
Open-Source



Jupyter notebook

Ambiente computacional web;
Códigos, gráficos, textos;
Sequências de células de códigos;
Organização;
Didático;



Curiosidades...

 jupyter = Julia + Python e R



Galileu que registram a descoberta das luas de Júpiter.

Instalando o ambiente

- Acessar o site: <https://www.anaconda.com/products/individual>;
- No fim da página (Anaconda Installers), escolher qual instalador de anaconda de acordo com seu sistema operacional;
- NNF (Instalador next-next-finish)
- Abrimos o anaconda navigator;
- Launch jupyter;