

Data Science – Princípios e Técnicas

Março
2024



Data Science



Onde me encontrar:

<https://www.linkedin.com/in/marco-mialaret-junior/>

e

<https://github.com/MatmJr>

Ciência de Dados, finalmente!

Data Science

“Um cientista de dados é alguém que sabe mais sobre estatística do que um cientista da computação e mais sobre ciência da computação do que um estatístico”

Data Science



A ciência de dados é uma área extremamente variada. Ela engloba desde profissionais que se assemelham à estatísticos até aqueles que são praticamente engenheiros de software. Alguns dominam o aprendizado de máquina, enquanto outros possuem conhecimento limitado sobre o tema.

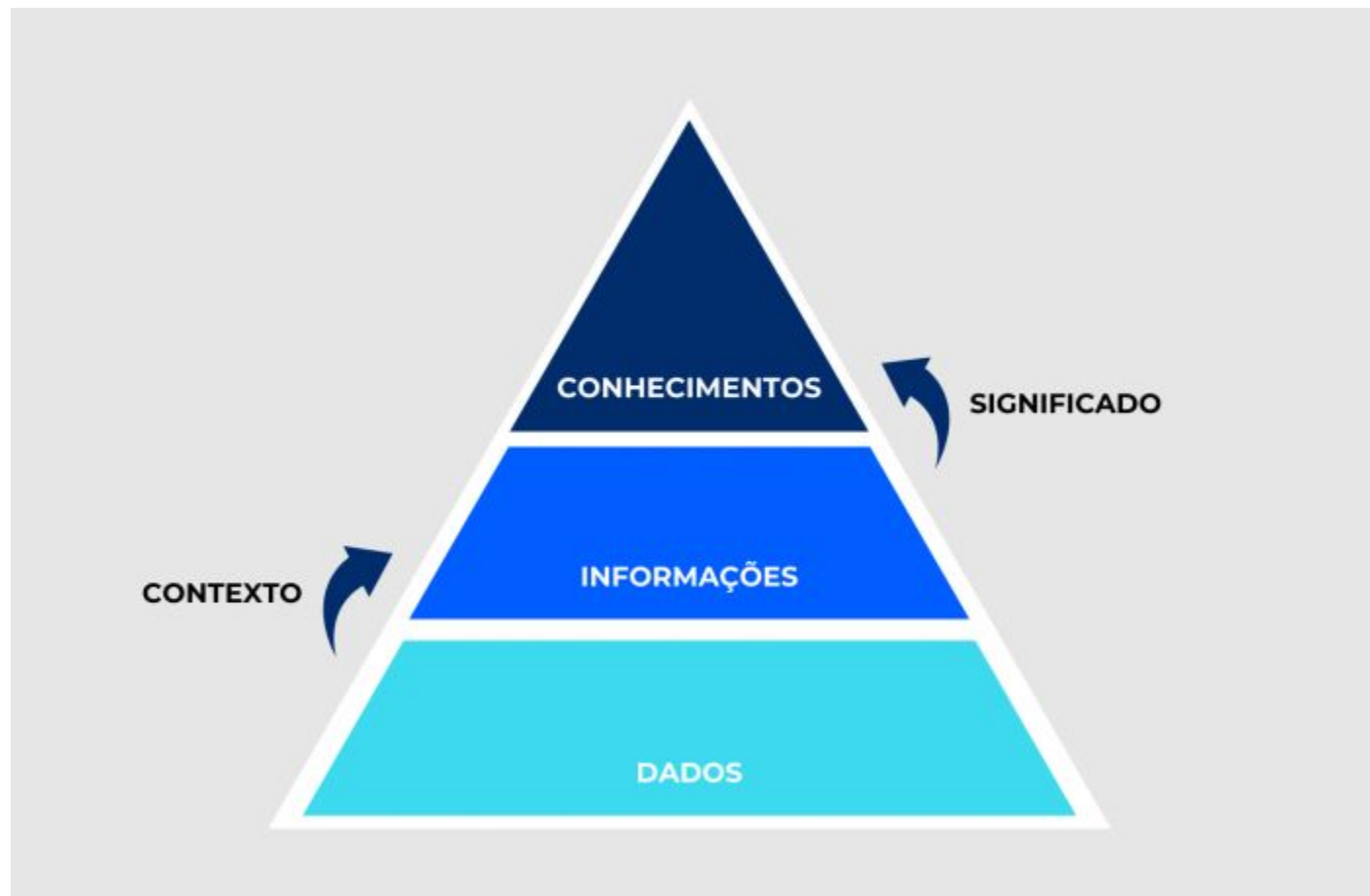
Data Science

O campo também inclui tanto indivíduos com doutorado e publicações acadêmicas quanto aqueles que nunca leram um artigo científico. Portanto, tentar definir ciência de dados de forma estrita é complicado, pois há sempre alguém no campo para quem a definição não se encaixa perfeitamente.

Data Science

Apesar da diversidade na ciência de dados, podemos tentar definir um cientista de dados como alguém que extrai conhecimento a partir de dados desorganizados. Atualmente, muitas pessoas se esforçam para transformar dados em conhecimento. E ajuda a gerar valor com esse conhecimento.

Data Science



Definições importantes

Data Science

Definição (Dados):

Coleções de medições, características ou fatos sobre um grupo.

Então, uma definição simples de ciência de dados é:

Definição(Ciência de dados):

O processo de extrair significado dos dados.

Data Science



Os dados consistem em pontos de dados:

Definição (pontos de dados):

Uma coleção de uma ou mais peças de informação coletadas sobre um único indivíduo ou entidade.

Cada ponto de dados pode conter variáveis e características.

Data Science

Definição(variáveis):

Características particulares, medições ou fatos que compõem um ponto de dados.

Data Science

Definição(características):

Peças individuais de informação em um conjunto de dados. Enquanto variáveis tipicamente representam dados não processados ou brutos, características podem incluir tanto variáveis quanto versões processadas das variáveis.

Data Science

Definição(dados quantitativos):

Dados numéricos. Dados quantitativos podem ser discretos (como o número de pessoas em uma família) ou contínuos (como a média de notas).

Data Science

Exemplos de variáveis quantitativas:

- altura
- peso
- renda anual
- GPA da faculdade
- milhas dirigidas no trajeto para o trabalho
- temperatura
- velocidade do vento
- população

Data Science



Definição(dados qualitativos):

Dados não numéricos. Variáveis qualitativas são geralmente categorias não numéricas às quais os dados podem pertencer (como cor de cabelo). Algumas categorias podem ter uma ordem associada a elas, mas a ordem não implica uma natureza numérica para as categorias. Por exemplo, uma pergunta de pesquisa pode ter respostas que variam de Discordo Fortemente a Concordo Fortemente.

Data Science



Exemplos de variáveis qualitativas:

- cor do cabelo (loiro, castanho, preto, vermelho, cinza, ...)
- status atual da precipitação (sem precipitação, chovendo, nevando, granizando, ...)
- tipo de carro (sedan, coupé, SUV, minivan, ...)
- categorias de furacões ou terremotos
- modelo de smartphone usado para acessar um aplicativo móvel
- tipo sanguíneo

Data Science



Definição(pergunta de pesquisa):

Uma pergunta que pode ser respondida usando pesquisa, incluindo coleta de dados e análise.

Data Science



Definição(pergunta de pesquisa):

Uma pergunta que pode ser respondida usando pesquisa, incluindo coleta de dados e análise.

Data Science

Por exemplo:

- Mais educação se traduz em mais riqueza?
- O clima está mudando?
- Quão rápido o coronavírus COVID-19 estava se espalhando quando se tornou prevalente nos Brasil dezembro de 2020?

E o mercado de trabalho?

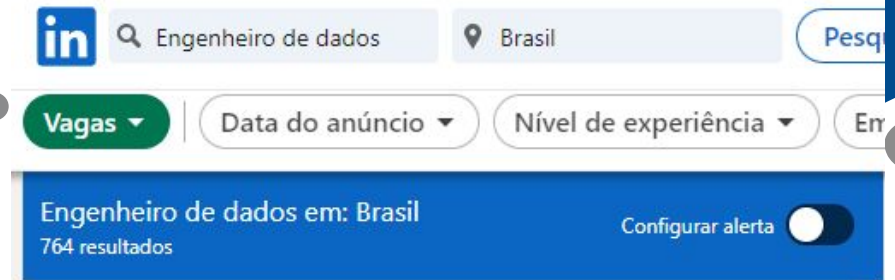
Data Science

Segundo a IDC, cerca de 2,5 quintilhões de dados são gerados diariamente. O grande desafio para as empresas é converter essa imensa quantidade de dados brutos em informações valiosas e insights úteis para alcançar objetivos específicos, papel fundamental desempenhado pela Ciência de Dados.

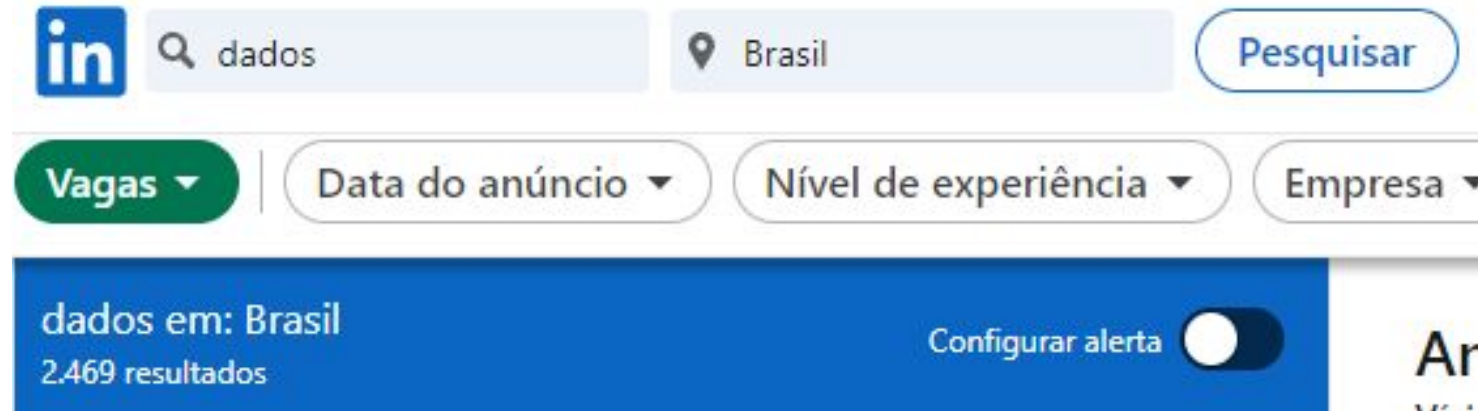
Data Science

Este cenário suscita um interesse crescente em entender o mercado de Ciência de Dados e avaliar o potencial de investimento nessa carreira.

Data Science



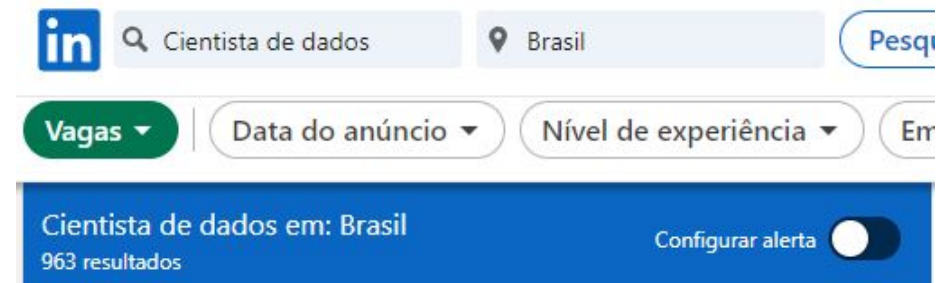
LinkedIn search results for "Engenheiro de dados" in Brazil. The search bar shows "Engenheiro de dados" and "Brasil". The results bar displays "Engenheiro de dados em: Brasil" with "764 resultados" and a "Configurar alerta" toggle switch.



LinkedIn search results for "dados" in Brazil. The search bar shows "dados" and "Brasil". The results bar displays "dados em: Brasil" with "2.469 resultados" and a "Configurar alerta" toggle switch.



LinkedIn search results for "Big Data" in Brazil. The results bar displays "Big Data em: Brasil" with "1.492 resultados" and a "Configurar alerta" toggle switch.



LinkedIn search results for "Cientista de dados" in Brazil. The search bar shows "Cientista de dados" and "Brasil". The results bar displays "Cientista de dados em: Brasil" with "963 resultados" and a "Configurar alerta" toggle switch.

Cientista x Analista x Eng. de Dados

Data Science



Embora possam parecer similares, essas Três áreas têm sim uma diferença entre elas.

A **Ciência de Dados** abrange muitos modelos e métodos científicos, matemáticos e estatísticos, além de ferramentas para analisar e manipular dados.

Data Science

Já a **Análise de Dados** é mais específica. Isso porque, ao invés de procurar conexões entre dados, ela conta com um objetivo específico no momento em que estão ordenando.

Essa área envolve o penteamento de dados para encontrar as informações que podem ser aplicados para o sucesso da empresa de alguma maneira.

Data Science

A **Engenharia de Dados** é responsável por preparar, coletar e organizar os dados. Depois disso é que entra a área de Ciência de Dados para analisá-los.



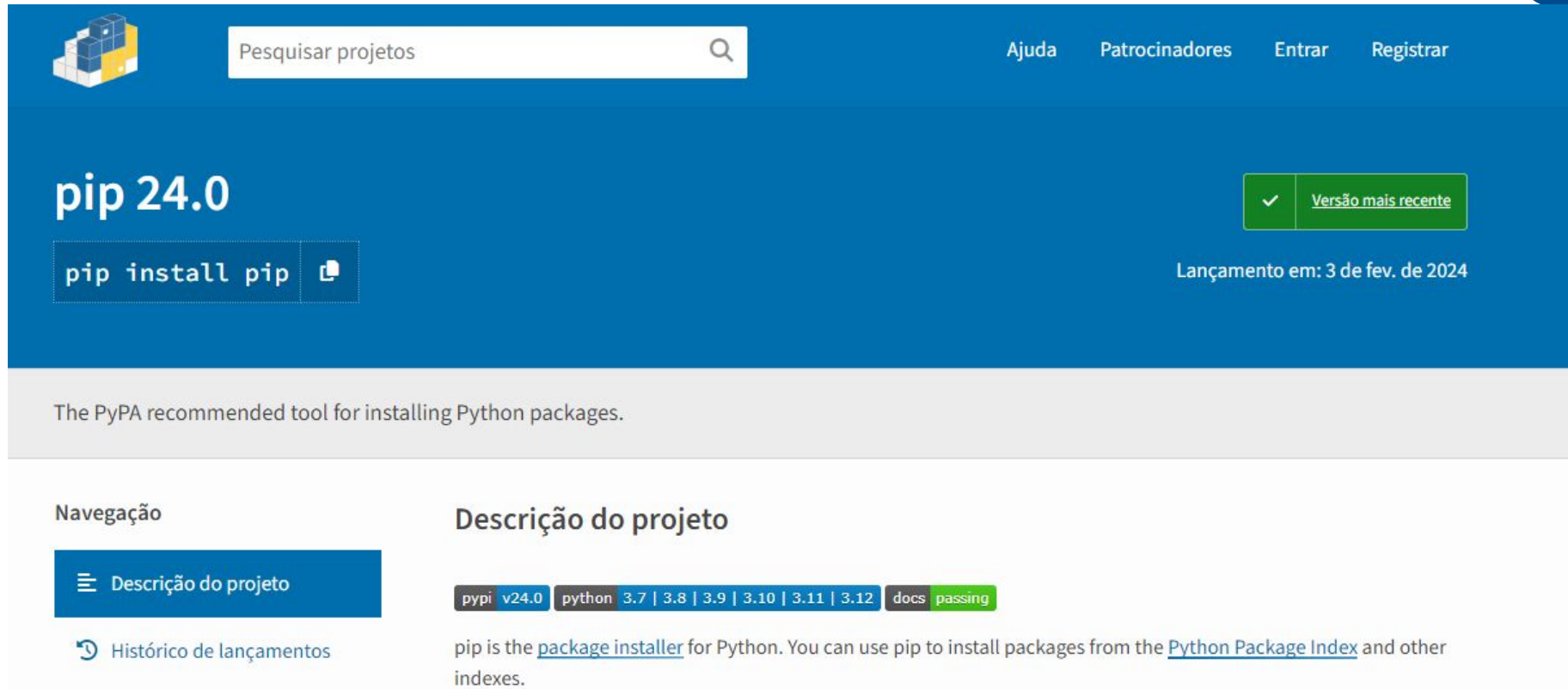
E por onde começamos?

Data Science



Bibliotecas no Python são coleções de módulos e funções que facilitam a programação ao oferecer funções e métodos prontos para serem usados. Isso significa que, em vez de escrever cada detalhe do código do zero, você pode usar bibliotecas para realizar tarefas complexas ou rotineiras de forma mais eficiente e com menos código.

Data Science



The screenshot shows the PyPI project page for pip 24.0. The header is blue with a search bar labeled "Pesquisar projetos" and navigation links: "Ajuda", "Patrocinadores", "Entrar", and "Registrar". The main content area is blue and features the text "pip 24.0" in large white font. Below this is a button with the command "pip install pip" and a copy icon. To the right, there is a green checkmark icon and a link "Versão mais recente". Below this, it says "Lançamento em: 3 de fev. de 2024". A light gray box contains the text "The PyPA recommended tool for installing Python packages." The bottom section is white and divided into two columns. The left column, titled "Navegação", has a blue button "Descrição do projeto" and a link "Histórico de lançamentos". The right column, titled "Descrição do projeto", shows a list of versions: "pypi v24.0 python 3.7 | 3.8 | 3.9 | 3.10 | 3.11 | 3.12 docs passing". Below this, it says "pip is the [package installer](#) for Python. You can use pip to install packages from the [Python Package Index](#) and other indexes."

Pesquisar projetos

Ajuda Patrocinadores Entrar Registrar

pip 24.0

pip install pip

Versão mais recente

Lançamento em: 3 de fev. de 2024

The PyPA recommended tool for installing Python packages.

Navegação

Descrição do projeto

Histórico de lançamentos

Descrição do projeto

pypi v24.0 python 3.7 | 3.8 | 3.9 | 3.10 | 3.11 | 3.12 docs passing

pip is the [package installer](#) for Python. You can use pip to install packages from the [Python Package Index](#) and other indexes.

Data Science



About us ▾ Getting started

pandas

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

[Install pandas now!](#)

Getting started

- [Install pandas](#)
- [Getting started](#)

Documentation

- [User guide](#)
- [API reference](#)
- [Contributing to pandas](#)
- [Release notes](#)

Community

- [About pandas](#)
- [Ask a question](#)
- [Ecosystem](#)

Data Science

Vamos acessar o google colab:

<https://drive.google.com/file/d/1VFy0kbwg-H5THkWol416niDf-if4rTHD/view?usp=sharing>

Dúvidas?



Marco Mialaret, MSc

Telefone:

81 98160 7018

E-mail:

marco.junior@pe.senac.br

