

Python e Dados: Como Transformar Informação em Poder

Prof. Me. Nisston Moraes Tavares de Melo





Roteiro

- Introdução ao poder dos dados
- Por que Python para dados?
- Manipulação de dados com Python
- Visualização de dados
- Análise de dados
- Automatização e pipelines de dados
- Desafios e ética no uso de dados



Introdução ao poder dos dados





Flexibilidade em Diferentes Áreas

- Pesquisa: Usado por cientistas e pesquisadores para análise e modelagem.
- Negócios: Aplicado em relatórios de desempenho e previsões de mercado.
- Tecnologia: Base para sistemas de recomendação, algoritmos de busca e otimização.
- Ciências: Ideal para análise de dados em biologia, física, economia, entre outros.



Introdução ao poder dos dados

- O Que São Dados e Por Que São Importantes?
 - Definição: Unidades de informação bruta que, quando processadas, se transformam em insights úteis.
 - Contextualização histórica: como a evolução tecnológica tornou a coleta e análise de dados mais acessíveis.
 - O papel dos dados no mundo atual: de pequenas empresas a grandes corporações, todos dependem de dados para tomar decisões.



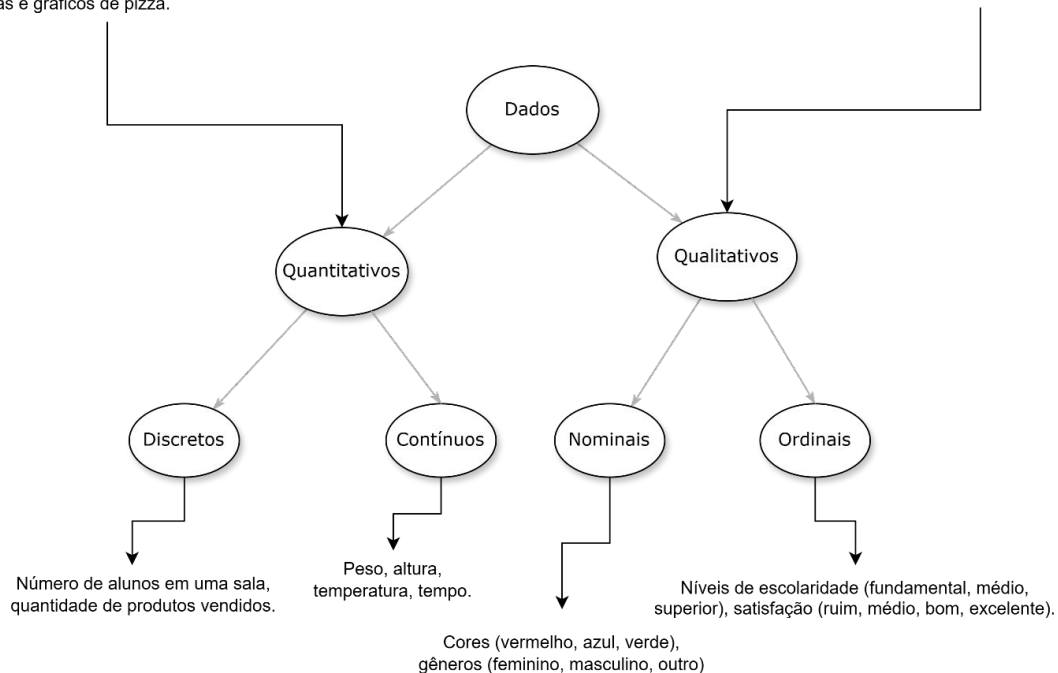
Tipos de Dados

Características

- Representados por textos ou rótulos.
- Não permitem cálculos matemáticos, mas podem ser analisados com frequência ou proporção.
- São ideais para gráficos de barras e gráficos de pizza.

Características

- Representados numericamente.
- Permitem cálculo de média, mediana, desvio padrão, etc.
- São ideais para gráficos como histogramas, gráficos de dispersão e boxplots.





Comparação Entre Dados Quantitativos e Qualitativos:

Característica	Quantitativos	Qualitativos
Formato	Números	Textos ou categorias
Subtipo	Discretos, Contínuos	Nominais, Ordinais
Análise	Média, desvio padrão	Frequência, proporção
Gráfico	Histogramas, dispersão	Barras, pizza

The background is a solid orange color. In the top-left corner, there are three vertical bars of varying heights, each composed of three overlapping circles. In the bottom-right corner, there are four vertical bars of increasing height from left to right, each also composed of three overlapping circles.

Por que Python para dados?



Comunidade Ativa e Suporte

- Recursos online: Fóruns como Stack Overflow e comunidades como Kaggle são ricas em conhecimento.
- Bibliotecas de código aberto: Constantemente atualizadas com novos recursos.
- Colaboração: Uso difundido em notebooks Jupyter, que permitem colaboração em tempo real e documentação integrada.



Integração com Outras Tecnologias

- Bancos de dados: Conexão com MySQL, PostgreSQL, MongoDB e outras tecnologias de armazenamento.
- APIs e Web: Integração com APIs REST e desenvolvimento de sistemas analíticos em frameworks como Django e Flask.
- Ambientes de Big Data: Compatível com Hadoop, Spark e AWS.



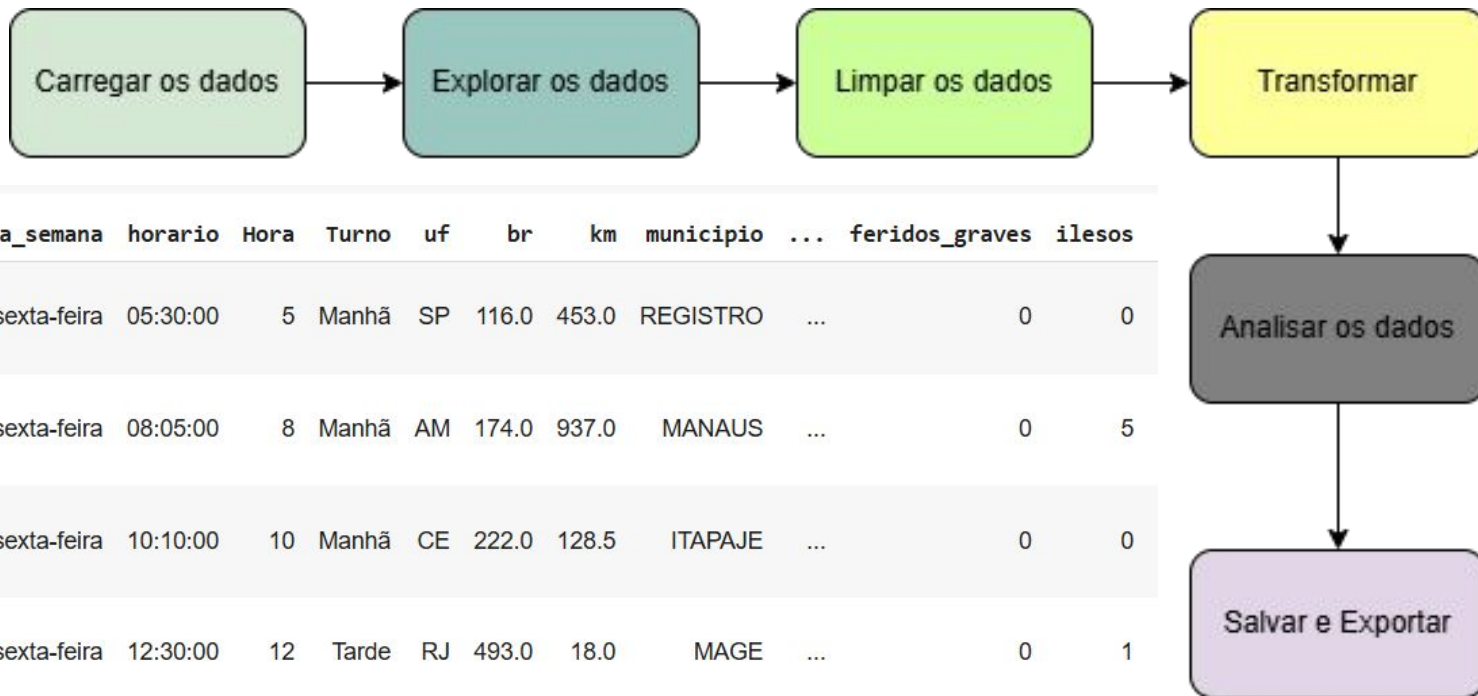
Manipulação de dados com Python



Fluxo de trabalho

- Carregar os dados: Importar de fontes como CSV, Excel, bancos de dados, APIs ou arquivos JSON.
- Explorar os dados: Visualizar a estrutura, tipos de dados e valores nulos.
- Limpar os dados:
 - Remover duplicatas.
 - Preencher ou excluir valores ausentes.
 - Corrigir erros de formatação.
- Transformar os dados:
 - Criar novas colunas ou atributos derivados.
 - Alterar o formato dos dados (ex.: pivot ou melt).
- Analisar os dados: Agrupar, filtrar e resumir para insights.
- Salvar ou exportar: Exportar os dados tratados para arquivos ou bases.

Fluxo de manipulação dos dados



	id	data_inversa	dia_semana	horario	Hora	Turno	uf	br	km	municipio	...	feridos_graves	ilesos
0	331730	01/01/2021	sexta-feira	05:30:00	5	Manhã	SP	116.0	453.0	REGISTRO	...	0	0
1	331804	01/01/2021	sexta-feira	08:05:00	8	Manhã	AM	174.0	937.0	MANAUS	...	0	5
2	331815	01/01/2021	sexta-feira	10:10:00	10	Manhã	CE	222.0	128.5	ITAPAJE	...	0	0
3	331823	01/01/2021	sexta-feira	12:30:00	12	Tarde	RJ	493.0	18.0	MAGE	...	0	1
4	331843	01/01/2021	sexta-feira	14:40:00	14	Tarde	RJ	393.0	252.0	BARRA DO PIRAI	...	1	1



Visualização dos Dados



Visualização dos Dados

- Durante a manipulação, é útil gerar visualizações rápidas:

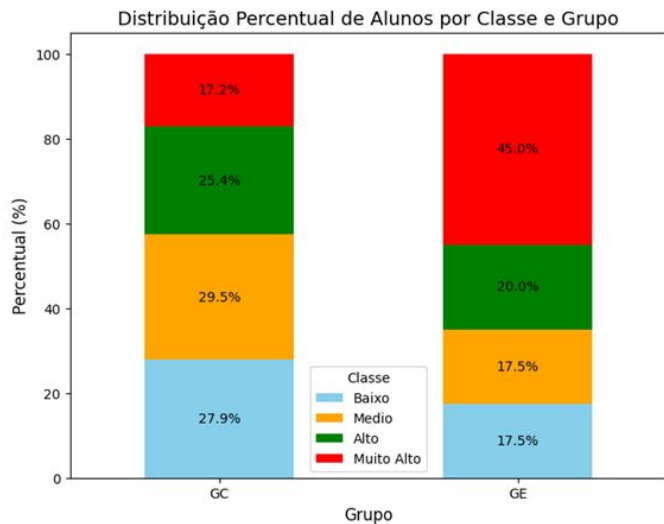
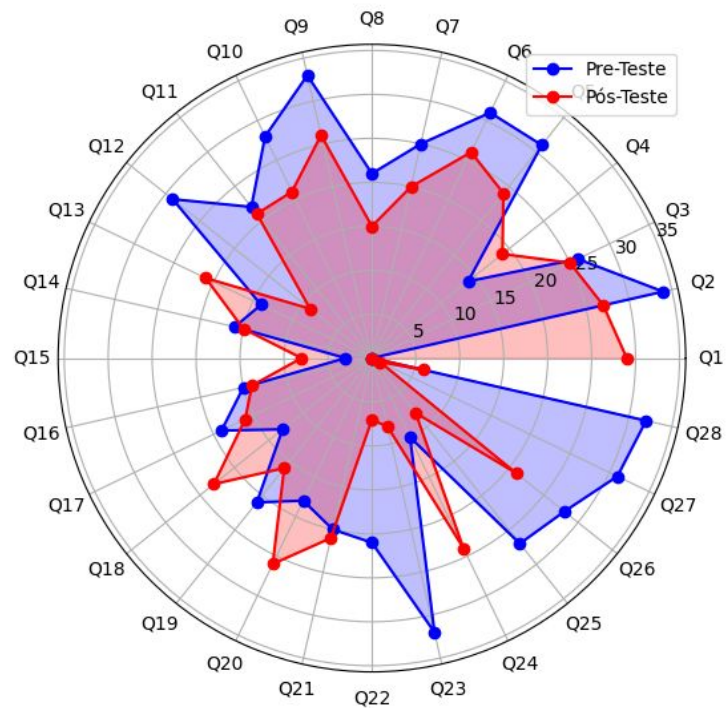


Gráfico de Radar das Questões do Grupo Experimental





Análise dos dados





Estatística dos dados

- A estatística é fundamental para transformar dados brutos em informações úteis, auxiliando na tomada de decisões embasadas em evidências. Seus tipos abrangem diferentes propósitos:
 - Descritiva
 - Exploratória
 - Inferencial
 - Preditiva
 - Explicativa
 - Prescritiva

Análise Estatística

```
graph TD; A([Análise Estatística]) --> B[Descritiva]; A --> C[Exploratória]; A --> D[Inferencial]; A --> E[Preditiva]; A --> F[Explicativa]; A --> G[Prescritiva]; B --> B1[Utiliza medidas de tendência central (média, mediana, moda), medidas de dispersão (desvio padrão, variância, amplitude) e gráficos (histogramas, boxplots, etc.)]; C --> C1[Explorar os dados para identificar padrões, relações e outliers. Geralmente utiliza visualizações gráficas (scatterplots, heatmaps) e técnicas para descobrir insights.]; D --> D1[Fazer generalizações ou conclusões sobre uma população com base em uma amostra. Envolve hipóteses, testes estatísticos, e intervalos de confiança.]; E --> E1[Prever valores futuros ou tendências. Baseia-se em modelos estatísticos ou machine learning. Envolve técnicas como regressão linear, árvores de decisão, séries temporais.]; F --> F1[Identificar relações de causa e efeito. Busca entender "por que" algo acontece, indo além de correlações para explorar causalidade.]; G --> G1[Sugerir ações ou decisões com base em dados. Utiliza modelos preditivos aliados a otimização para propor estratégias.];
```

Descritiva

Utiliza medidas de tendência central (média, mediana, moda), medidas de dispersão (desvio padrão, variância, amplitude) e gráficos (histogramas, boxplots, etc.)

Exploratória

Explorar os dados para identificar padrões, relações e outliers. Geralmente utiliza visualizações gráficas (scatterplots, heatmaps) e técnicas para descobrir insights.

Inferencial

Fazer generalizações ou conclusões sobre uma população com base em uma amostra. Envolve hipóteses, testes estatísticos, e intervalos de confiança.

Preditiva

Prever valores futuros ou tendências. Baseia-se em modelos estatísticos ou machine learning. Envolve técnicas como regressão linear, árvores de decisão, séries temporais.

Explicativa

Identificar relações de causa e efeito. Busca entender "por que" algo acontece, indo além de correlações para explorar causalidade.

Prescritiva

Sugerir ações ou decisões com base em dados. Utiliza modelos preditivos aliados a otimização para propor estratégias.



Automatização e pipelines de dados





Etapas de um pipeline de dados

- Coleta de Dados:
 - Conexão com APIs, bancos de dados, arquivos CSV/Excel, etc.
 - Automação de scraping de dados.
- Transformação e Limpeza:
 - Padronização e limpeza de dados (tratamento de valores nulos, duplicados).
 - Enriquecimento de dados com informações externas.
 - Aplicação de transformações customizadas (normalização, agregação, etc.).
- Carregamento de Dados:
 - Persistência em bancos de dados ou data warehouses.
 - Geração de arquivos (CSV, JSON, etc.).
 - Integração com ferramentas de visualização (ex.: Power BI, Tableau).



Ferramentas e tecnologias

- Ferramentas para ETL (Extract, Transform, Load):
 - Apache Airflow.
 - Talend.
 - Apache NiFi.
 - Prefect.
- Soluções em nuvem:
 - AWS Glue.
 - Google Cloud Dataflow.
 - Azure Data Factory.
- Bibliotecas Python:
 - Pandas para manipulação de dados.
 - Luigi para criação de pipelines.
 - PySpark para processamento de grandes volumes.
 - Dask para computação paralela.



Exemplo Prático

- Web Scraping com Python
 - <https://github.com/nisstn/coletadadoswebscraping>



Desafios e ética no uso de dados





Introdução aos Desafios e à Ética no Uso de Dados

- O que são dados e por que eles são tão valiosos na era digital?
- A relação entre tecnologia, dados e sociedade.
- Por que a ética no uso de dados é importante?
- Exemplos históricos de mau uso de dados e suas consequências.

Privacidade e Proteção de Dados

- O que é privacidade de dados e por que ela está em risco?
- Leis e regulamentações sobre proteção de dados:
 - GDPR (Europa).
 - LGPD (Brasil).
 - CCPA (Califórnia).
- Desafios no cumprimento de normas de privacidade:
 - Coleta de consentimento.
 - Retenção e exclusão de dados.
 - Transferência internacional de dados.
- Exemplos práticos de violação de privacidade (ex.: vazamentos de dados).





Segurança de Dados

- Principais ameaças à segurança dos dados:
 - Ataques cibernéticos (phishing, ransomware, etc.).
 - Vazamentos de dados em grandes corporações.
- Técnicas para proteger dados:
 - Criptografia.
 - Controle de acesso.
 - Testes regulares de segurança.
- Riscos associados ao uso de dispositivos IoT e dados em nuvem



Exemplos Práticos do Uso de Dados





Negócios: Otimização de Vendas e Marketing

- Cenário: Uma loja online utiliza dados de navegação e compras para personalizar recomendações.
- Exemplo real: A Amazon, que recomenda produtos com base no histórico de compras e preferências dos usuários.
- Impacto: Aumento das vendas por meio de estratégias de upselling e cross-selling.

Saúde: Dados salvando vidas

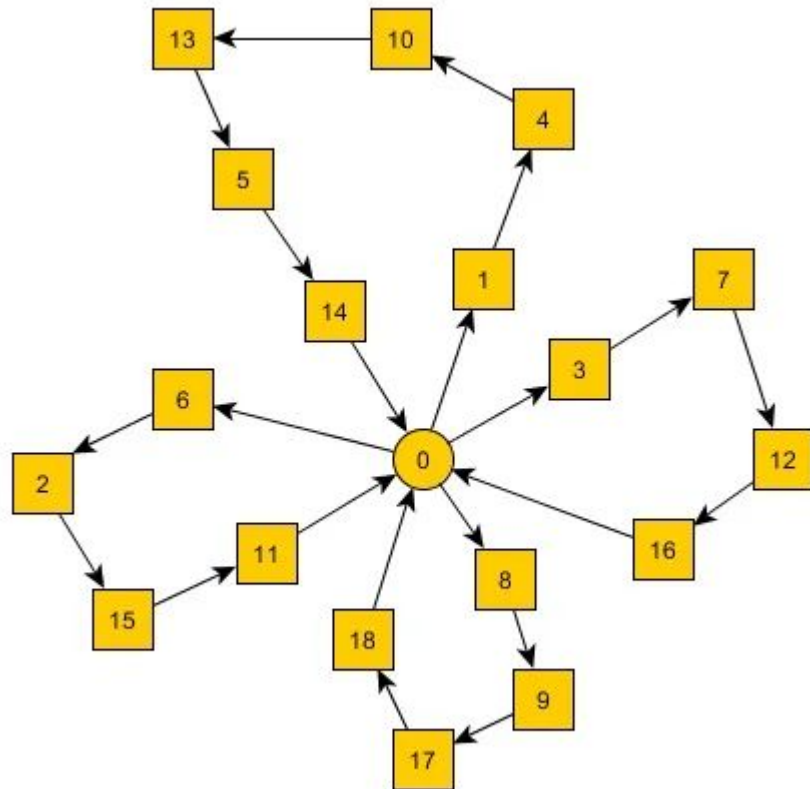
- Cenário: Um hospital coleta dados de pacientes para prever surtos de doenças ou monitorar condições críticas.
- Exemplo real: Durante a pandemia de COVID-19, modelos preditivos baseados em dados ajudaram a prever a disseminação do vírus.
- Impacto: Melhora na alocação de recursos e no planejamento de respostas de emergência.





Transportes: Redução de Custos e Aumento da Eficiência

- Cenário: Empresas de logística usam dados para otimizar rotas.
- Exemplo real: A UPS utiliza algoritmos para determinar as melhores rotas, economizando milhões de dólares em combustível.
- Impacto: Redução de custos operacionais e menor impacto ambiental.





Entretenimento: Personalização de Conteúdo

Cenário: Plataformas de streaming recomendam filmes e séries com base no comportamento do usuário.

Exemplo real: A Netflix usa análises de dados para criar e sugerir conteúdo popular, como a série Stranger Things.

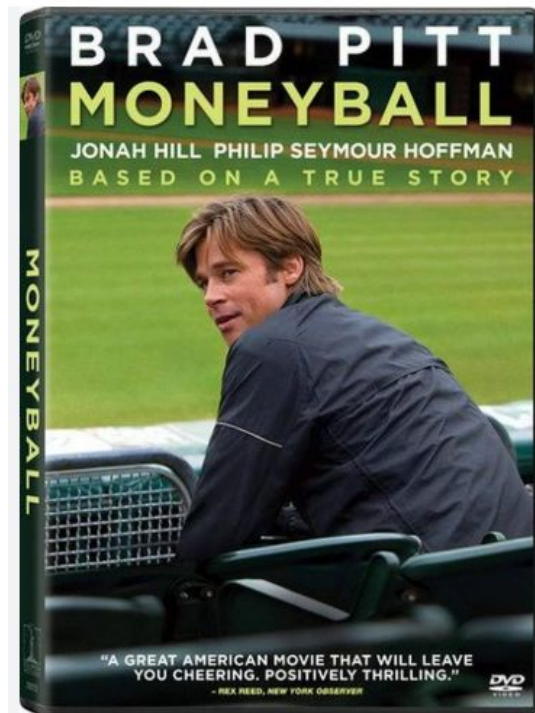
Impacto: Maior retenção de clientes e fidelidade à marca.

Esportes: Desempenho e Estratégia

Cenário: Times esportivos utilizam dados para planejar jogos e treinar jogadores.

Exemplo real: O filme Moneyball (baseado em uma história real) demonstra como estatísticas ajudaram um time de beisebol com orçamento limitado a alcançar o sucesso.

Impacto: Decisões mais estratégicas baseadas em desempenho, em vez de intuição.





Dados no Cotidiano

Trânsito: Aplicativos como Waze e Google Maps utilizam dados de tráfego em tempo real para sugerir rotas mais rápidas.

Financeiro: Instituições bancárias usam dados para avaliar crédito e prevenir fraudes.

Educação: Plataformas de ensino online ajustam o conteúdo com base no progresso e nas necessidades dos alunos.



Como Python Está Inserido Nesse Cenário?

Python facilita a análise e manipulação de grandes volumes de dados.

A combinação de simplicidade e poder das bibliotecas de Python o torna a escolha ideal para iniciantes e profissionais avançados.

Exemplo prático: Análise de tendências em dados de vendas de uma padaria ou visualização de crescimento de clientes ao longo do tempo.





História e Criação

- Criador: Guido van Rossum.
- Ano de Criação: 1991.
- Motivação: Desenvolver uma linguagem fácil de aprender, com sintaxe clara e foco em produtividade e legibilidade.
- Origem do Nome: Inspirado no grupo de comédia britânico Monty Python's Flying Circus.



Características principais

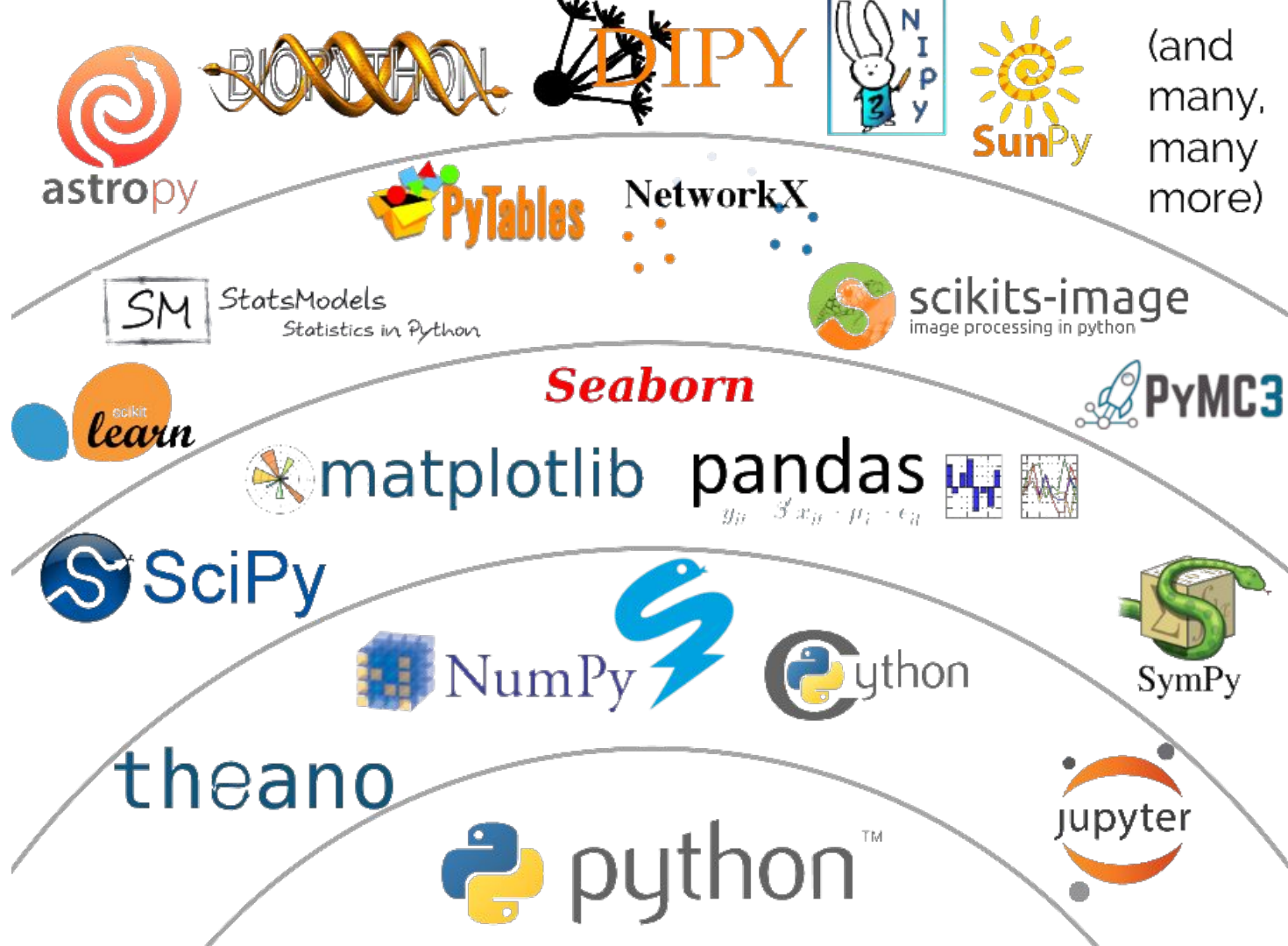
- Sintaxe Simples e Intuitiva: Ideal para iniciantes e projetos complexos.
- Linguagem Multiparadigma: Suporta programação procedural, orientada a objetos e funcional.
- Ampla Biblioteca Padrão: Fornece ferramentas prontas para manipulação de dados, redes, processamento de texto, entre outros.
- Portabilidade: Funciona em múltiplas plataformas (Windows, macOS, Linux).
- Comunidade Ativa: Milhões de desenvolvedores no mundo todo.



Ecossistema e aplicações

- Principais Usos:
 - Análise de Dados e Ciência de Dados (pandas, NumPy, Matplotlib).
 - Desenvolvimento Web (Django, Flask).
 - Inteligência Artificial e Machine Learning (TensorFlow, PyTorch, scikit-learn).
 - Automação e Scripting.
 - Desenvolvimento de Jogos (Pygame).
 - Computação Científica.
- Empresas que Usam Python:
 - Google, Facebook, Netflix, Instagram, Spotify, NASA, entre outras.

Ecosistema e aplicações





Recursos e Comunidades

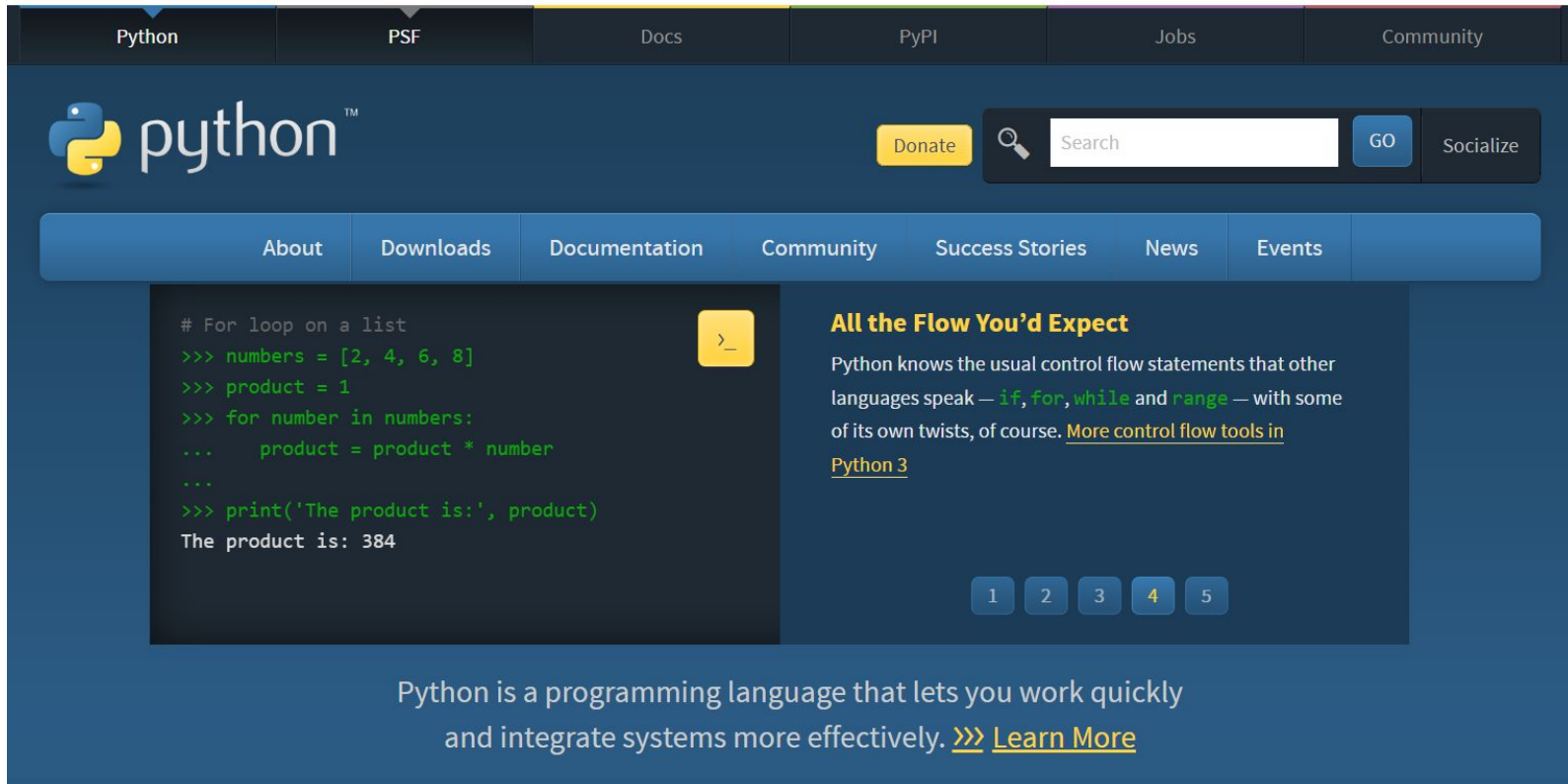
- Grupos de Discussão e Fóruns:
 - Python.org (site oficial).
 - Stack Overflow (milhares de perguntas e respostas).
 - Reddit: Subreddits como r/Python.
- Conferências Globais:
 - PyCon (eventos anuais em várias regiões).
 - DjangoCon (focada em Django).
- Cursos e Tutoriais:
 - Plataformas como Codecademy, Coursera, edX e YouTube oferecem conteúdos para todos os níveis.



Curiosidades

- Primeira Versão: Python 0.9.0 (incluía suporte a classes e funções).
- Mascote: Embora não seja oficial, a comunidade associa Python a uma cobra píton.
- Open Source: Python é de código aberto, mantido pela Python Software Foundation (PSF).
- Número de Desenvolvedores: Estimado em mais de 10 milhões de programadores.
- Linguagem Popular: Consistentemente classificada como uma das linguagens mais populares em rankings como TIOBE e Stack Overflow.
- Frameworks e Bibliotecas: Mais de 300 mil pacotes disponíveis no PyPI (Python Package Index).

Site oficial: <https://www.python.org/>



The image is a screenshot of the Python.org homepage. At the top, there is a navigation bar with links for Python, PSF, Docs, PyPI, Jobs, and Community. Below this is a dark blue header featuring the Python logo, a 'Donate' button, a search bar with a 'GO' button, and a 'Socialize' button. A secondary navigation bar contains links for About, Downloads, Documentation, Community, Success Stories, News, and Events. The main content area is split into two columns. The left column displays a code snippet in a dark-themed editor with a yellow prompt icon, showing a for loop that calculates the product of a list of numbers. The right column has a section titled 'All the Flow You'd Expect' with text about Python's control flow statements and a link to 'More control flow tools in Python 3'. At the bottom of the right column is a pagination control with buttons for 1, 2, 3, 4, and 5, where 4 is highlighted. The footer contains the text: 'Python is a programming language that lets you work quickly and integrate systems more effectively. >>> [Learn More](#)'.

Python

PSF

Docs

PyPI

Jobs

Community

python™

Donate

Search

GO

Socialize

About

Downloads

Documentation

Community

Success Stories

News

Events

```
# For loop on a list
>>> numbers = [2, 4, 6, 8]
>>> product = 1
>>> for number in numbers:
...     product = product * number
...
>>> print('The product is:', product)
The product is: 384
```

>_

All the Flow You'd Expect

Python knows the usual control flow statements that other languages speak — `if`, `for`, `while` and `range` — with some of its own twists, of course. [More control flow tools in Python 3](#)

1 2 3 4 5

Python is a programming language that lets you work quickly and integrate systems more effectively. >>> [Learn More](#)

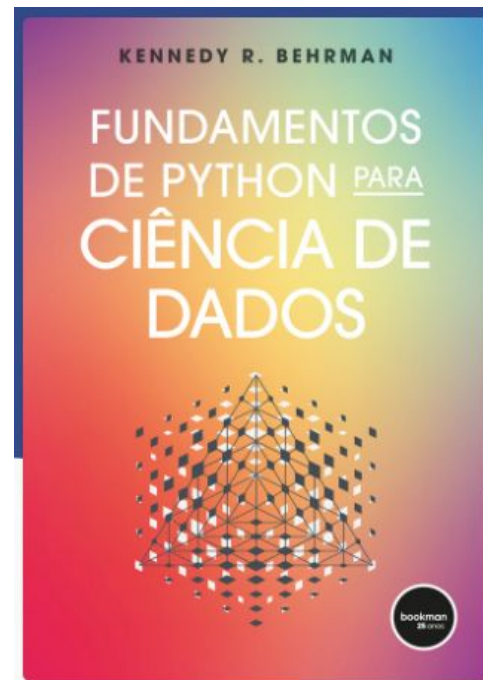
Indicação de Filmes



Indicação de Filmes



Indicações de livros



Exemplo Prático



Dados de acidentes de trânsito em BR

- [Link](#) para o GitHub
- [Link](#) para o Google Colab

Obrigado!

