Raspagem de dados | Twitter

Usando Python, mas não assusta.

Gustavo Sarti

sociólogo | educador

Redes sociais
@gmsarti
gmsarti@gmail.com

Coletivo
Máquina Tudo
maquinatudo.myportfolio.com

Por que o Twitter?

Relevância para o debate
+
Mais fácil de raspar dados

Raspagem de dados da internet (ou *web scraping*) é o processo de coletar dados públicos da internet.

O processo é diferente em cada site ou plataforma e exige um pouco de conhecimento específico sobre o site que será usado.

E esse Python?

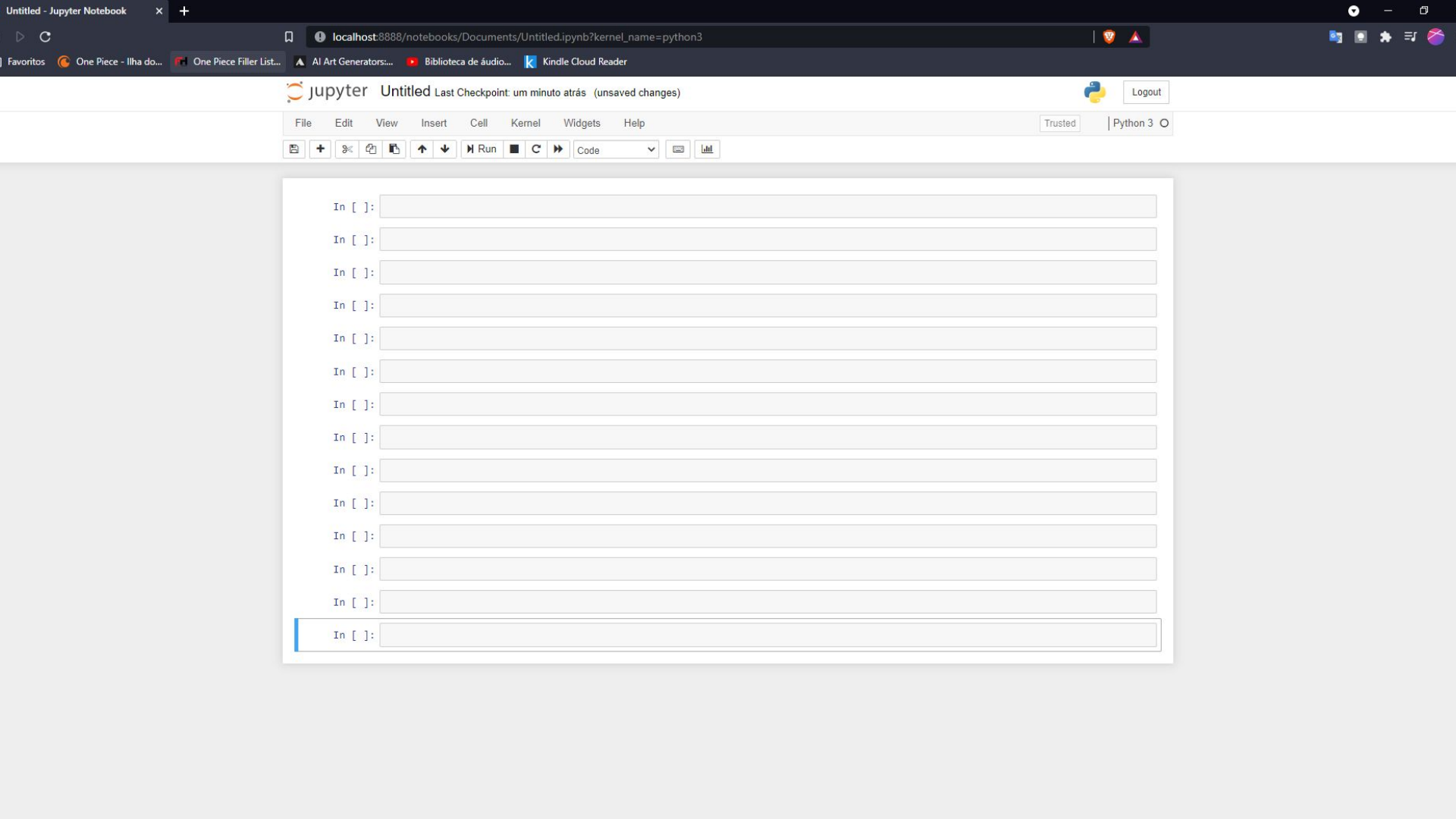
Linguagem de programação um pouco mais fácil, mas muito poderosa e ótima para trabalhar com dados

Programar é uma boa habilidade de se ter. No mínimo ela aumenta sua autonomia enquanto pesquisador.

Raspar dados da internet nos possibilita acessar mais dados do que conseguiríamos processar em ferramentas de planilha.

Jupyter notebook

O artigo acadêmico está obsoleto e o futuro está aqui


Untitled - Jupyter Notebook
localhost:8888/notebooks/Documents/Untitled.ipynb?kernel_name=python3
Favoritos
One Piece - Ilha do...
One Piece Filler List...
AI Art Generators:...
Biblioteca de áudio...
Kindle Cloud Reader
jupyter
Untitled
Last Checkpoint: um minuto atrás
(unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code
In [ ]:

Jupyter Notebook

O Jupyter Notebook é um app de código aberto que roda no seu navegador no computador. Nele é possível criar e compartilhar documentos com código (em algumas linguagens de programação), equações, visualizações e texto narrativo.

Os principais usos são: limpeza e transformação de dados, simulação numérica, modelagem estatística, visualização de dados, aprendizado de máquina etc.

Se um dia você quiser instalá-lo em seu computador eu recomendo que faça isso usando a Anaconda. Mas vamos usar esses recursos todos sem precisar instalar nada e usando o poder computacional dos outros.

Google Colab

https://colab.research.google.com/notebooks/intro.ipynb

Kaggle

https://www.kaggle.com/

Programação mínima

O que são variáveis?

Operadores lógicos?

O que são comentários?

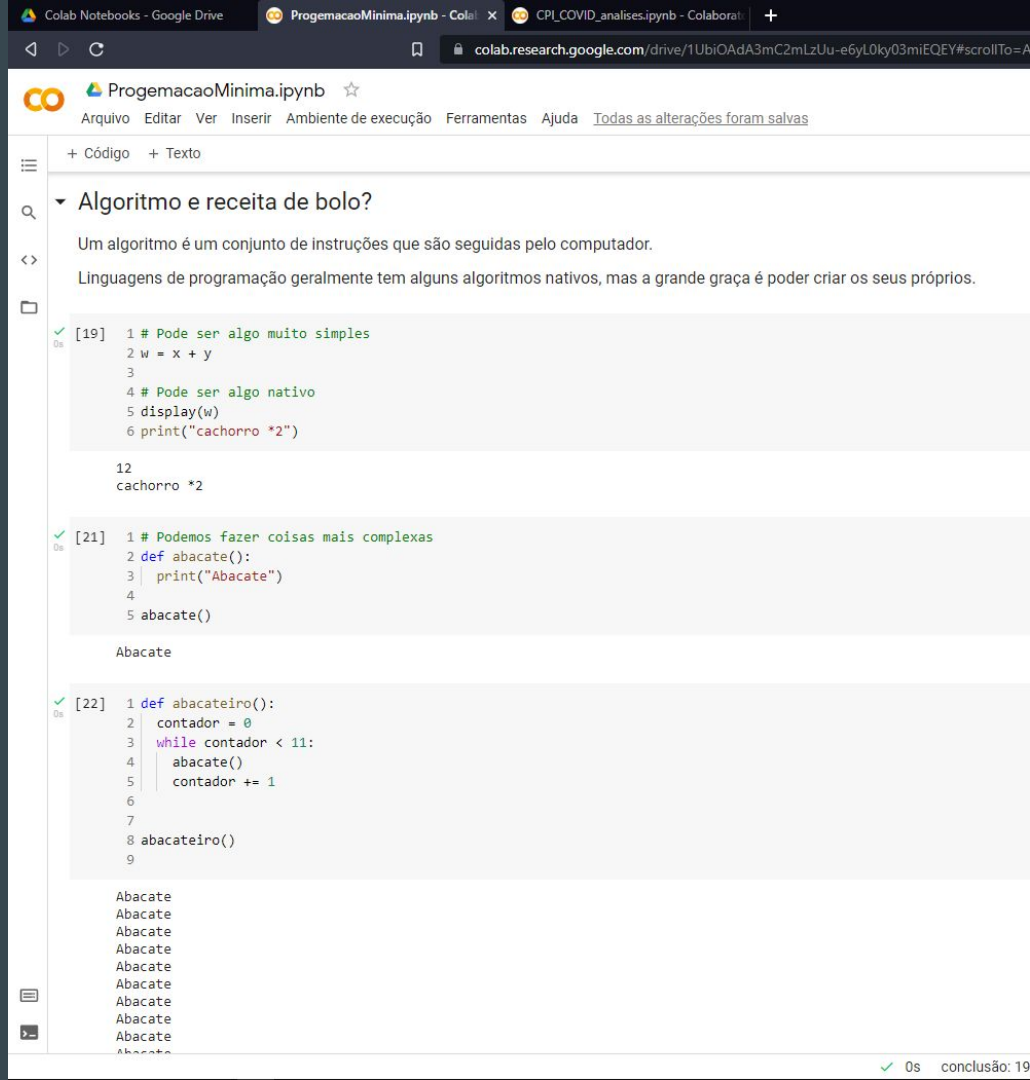
Algoritmo e receita de bolo?

Bibliotecas?

Link pro notebook na imagem ali ->

Acesse e copie para o seu Drive.

Clique para aprender mais.



Começando os trabalhos

Nossa ideia é fazer um sistema que colete informações no Twitter e nos entregue alguma informação útil.

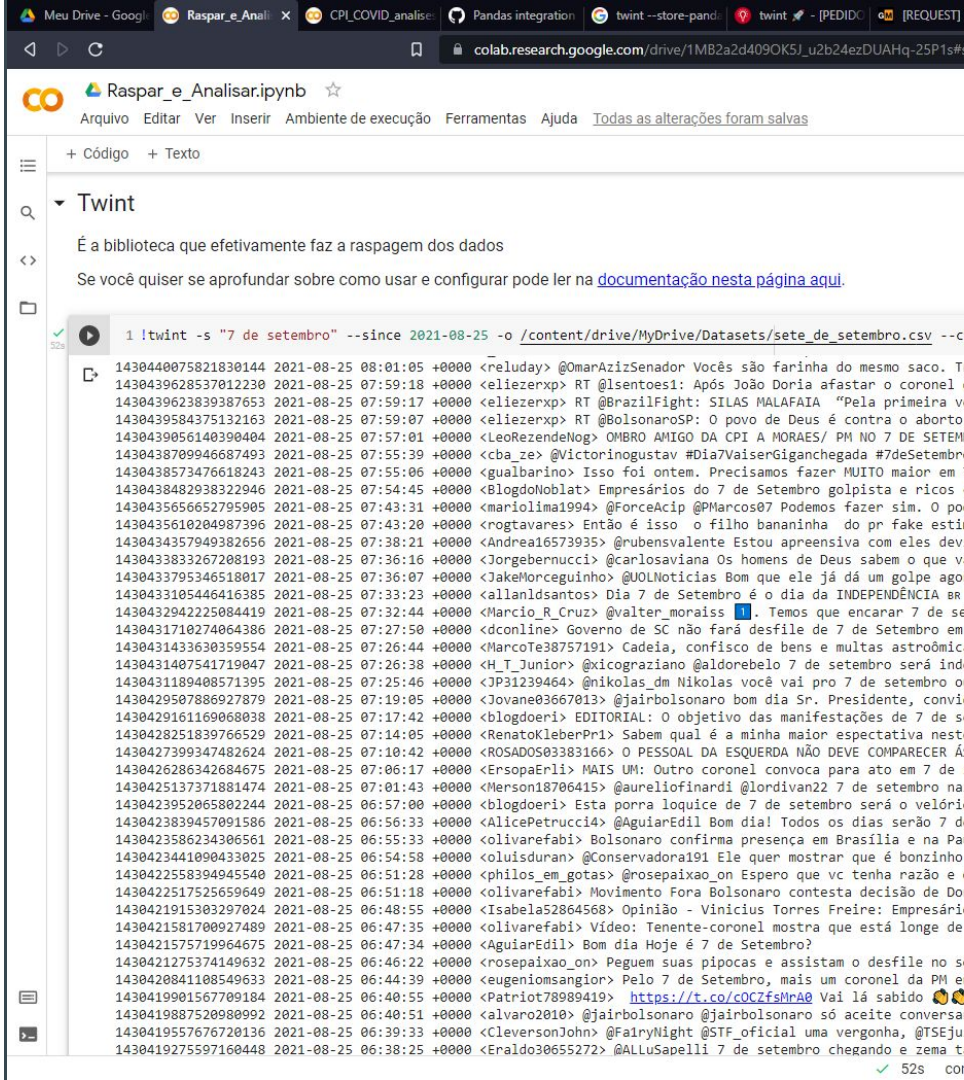
Para isso existe um ciclo típico de trabalho com dados, geralmente iniciando de uma *Questão de pesquisa*.

Reduzir tempo despendido em cada etapa mais mecânica libera mais tempo para análise e reflexão.

Máquina de salsicha

Um sistema que recolhe dados não estruturados e nos entrega alguma informação

Link para o notebook ->



Aprofundando

Existe um risco em seguir mantendo afastadas as redes sociais de ferramentas digitais de pesquisa e análise. Programar em Python ou R nos dá muitas opções fantásticas, mas não são as únicas. Vale a pena conhecer ferramentas como Tableau, Power BI ou mesmo os bons e velhos editores de planilhas.

Para relembrar como os notebooks de hj funcionam ficam alguns vídeos.

Contatos

Twitter | LinkedIn
Github | Youtube

boa sorte ;)