
Artigo

[Henrique Dias](#) · jan 13, 2022 5min de leitura

[Open Exchange](#)

Como achar o dataset que você precisa?

Fala pessoal, tudo bem?

Espero que todos estejam bem, saudáveis e que tenham um excelente 2022!

Ao longo dos anos, eu trabalhei nos mais diferentes projetos e acabei me deparando com dataset super interessantes.

Mas, na maioria das vezes, os datasets utilizados para o trabalhar eram datasets dos clientes. Quando eu comecei a participar das competições nos últimos anos, eu comecei a vasculhar na web por datasets que eu possa chamar de meu

Eu acabei coletando alguns dados, mas eu estava pensando, "Esses datasets são o suficiente para ajudar as outras pessoas?"

Então, trocando ideias com o [@José Pereira](#) para essa competição, nós decidimos por uma abordagem utilizando uma perspectiva diferente.

Nós pensamos em oferecer uma variedade de datasets de qualquer espécie de 2 fonte de dados famosas. Dessa forma, nós podemos empoderar vocês para encontrarem e instalarem o dataset desejado de uma forma rápido, fácil e indolor.

Socrata

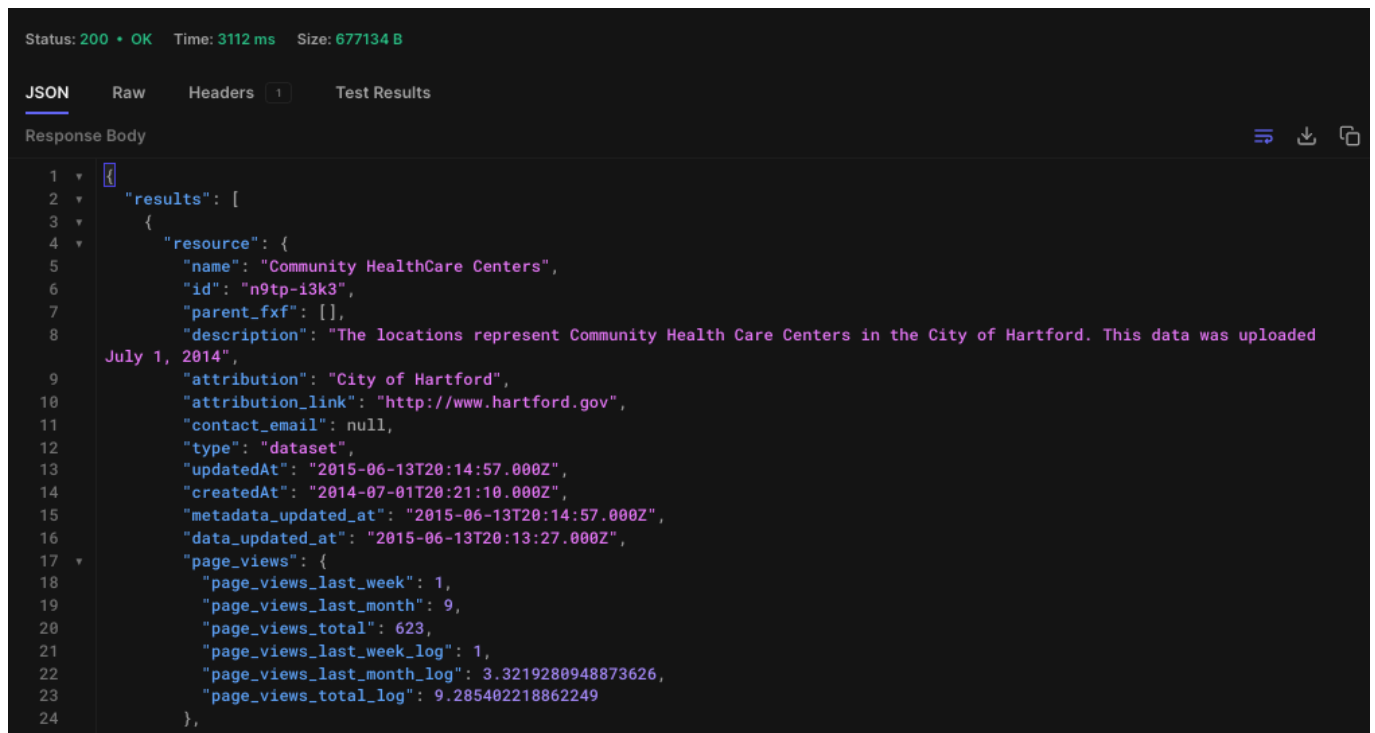
Socrata Open Data API permite que você de forma programática possa acessar uma variedade de dados abertos de governos, organizações sem fins lucrativos e ONGs de todo o mundo.

Para esse release inicial, nós estamos usando as APIs Socrata para pesquisar e fazer download de um dataset específico.

Pode utilizar a ferramenta para API da sua preferência [Postman](#), [Hoppscotch](#)

```
GET> https://api.us.socrata.com/api/catalog/v1?only=dataset&q=healthcare
```

This endpoint will return all healthcare related datasets, like the image below:



Agora, basta pegar o ID. Neste caso o ID é: "n9tp-i3k3"

Vamos para o terminal

```
IRISAPP>set api = ##class(dc.dataset.importer.service.socrata.SocrataApi).%New()
IRISAPP>do api.InstallDataset({"datasetId": "n9tp-i3k3", "verbose":true})
```

```
Compilation started on 01/07/2022 01:01:28 with qualifiers 'cuk'
Compiling class dc.dataset.imported.DsCommunityHealthcareCenters
Compiling table dc_dataset_imported.DsCommunityHealthcareCenters
Compiling routine dc.dataset.imported.DsCommunityHealthcareCenters.1
Compilation finished successfully in 0.108s.
```

```
Class name: dc.dataset.imported.DsCommunityHealthcareCenters
Header: Name VARCHAR(250),Description VARCHAR(250),Location VARCHAR(250),Phone_Number
VARCHAR(250),geom VARCHAR(250)
Records imported: 26
```

Depois do comando acima, você pode conferir o seu dataset prontinho para uso!

Kaggle

Kaggle, uma subsidiária do Google LLC, é uma comunidade online de cientistas de dados e profissionais de aprendizado de máquina. O Kaggle permite que os usuários encontrem e publiquem conjuntos de dados, explorem e construam modelos em um ambiente de ciência de dados baseado na Web, trabalhem com outros cientistas de dados e engenheiros de aprendizado de máquina e participem de competições para resolver desafios de ciência de dados.

Em junho de 2017, Kaggle anunciou que ultrapassou 1 milhão de usuários registrados, ou Kagglers, e em 2021 tem mais de 8 milhões de usuários registrados. A comunidade abrange 194 países. É uma comunidade diversificada, desde aqueles que estão começando até muitos dos pesquisadores mais conhecidos do mundo.

Comunidade pequenininha hein!?

Para usar os conjuntos de dados do Kaggle, você precisa se registrar no [site](#). Depois disso, você precisa criar um token para usar a API do Kaggle.

Agora, da mesma forma que fizemos com o Socrata, você pode utilizar a API para fazer a pesquisa e download do dataset.

```
GET> https://www.kaggle.com/api/v1/datasets/list?search=appointments
```

No Kaggle, ao invés de id utilizaremos o ref. Nesse exemplo o valor de ref é: "joniarroba/noshowappointments"

Os parâmetros abaixo "your-username", e "your-password" são parâmetros que o Kaggle fornece quando você cria o token para API.

```
IRISAPP>Set credentials = ##class(dc.dataset.importer.service.CredentialsService).%New()  
( )
```

```
IRISAPP>Do credentials.SaveCredentials("kaggle", "<your-username>", "<your-password>")
```

```
IRISAPP>Set api = ##class(dc.dataset.importer.service.kaggle.KaggleApi).%New()
```

```
IRISAPP>Do api.InstallDataset({"datasetId":"joniarroba/noshowappointments", "credentials":"kaggle", "verbose":true})
```

```
Class name: dc.dataset.imported.DsNoshowappointments
```

```
Header: PatientId INTEGER,AppointmentID INTEGER,Gender VARCHAR(250),ScheduledDay DATE  
,AppointmentDay DATE,Age INTEGER,Neighbourhood VARCHAR(250),Scholarship INTEGER,Hiper  
tension INTEGER,Diabetes INTEGER,Alcoholism INTEGER,Handcap INTEGER,SMS_received INTE  
GER,No-show VARCHAR(250)
```

```
Records imported: 259
```

Pronto! Mais um dataset prontinho para uso

Graphic User Interface

Para facilitar as coisas, estamos oferecendo uma interface gráfica para instalar o dataset. Mas isso é algo que quero falar em nosso próximo artigo. Enquanto isso, você pode conferir uma prévia abaixo enquanto estamos dando aquele talento, antes do lançamento oficial:

Vídeo Demo

Se você está se perguntando como é fazer o download de um dataset grande, da uma olhada nesse vídeo mostrando como é.... Agora... se +400.000 registros não forem o suficiente, que tal 1 MILHÃO DE REGISTROS?! Assista!

<https://www.youtube.com/embed/0T8wXRsaJso>

[Isso é um link incorporado, mas você não pode ver conteúdo incorporado diretamente no site, porque recusou os cookies necessários para acessá-lo. Para ver o conteúdo incorporado, você precisa aceitar todos os cookies nas suas Definições de cookies]

Votação

Como achar o dataset que você precisa?

Published on InterSystems Developer Community (<https://community.intersystems.com>)

Se você curtiu o app, por favor vote em iris-kaggle-socrata-generator!

<https://openexchange.intersystems.com/contest/current>

[#Concurso](#) [#Dados não estruturados](#) [#Importação e Exportação de Dados](#) [#Modelo de Dados](#) [#Modelo de Dados Documento](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)
[Confira o aplicativo relacionado no InterSystems Open Exchange](#)

URL de origem: <https://pt.community.intersystems.com/post/como-achar-o-dataset-que-voc%C3%AA-precisa>