CBSOData Documentation

Release 1.3

Jonathan de Bruin

Feb 13, 2021

Contents

1 Statistics Netherlands opendata API client for Python **3**
1.1 Installation 3
1.2 Usage 3
1.3 Command Line Interface 6

2 Reference **7**

Python Module Index **9**

Index **11**

Contents:

CHAPTER 1

Statistics Netherlands opendata API client for Python

Retrieve data from the open data interface of Statistics Netherlands (Centraal Bureau voor de Statistiek) with *Python*. The data is identical in content to the tables which can be retrieved and downloaded from StatLine. CBS datasets are accessed via the CBS open data portal.

The documentation of this package is found at this page and on readthedocs.io.

R user? Use cbsodataR.

1.1 Installation

From PyPi

```
pip install cbsodata
```

1.2 Usage

Load the package with

```
>>> import cbsodata
```

1.2.1 Tables

Statistics Netherlands (CBS) has a large amount of public available data tables (more than 4000 at the moment of writing). Each table is identified by a unique identifier (`Identifier`).

```
>>> tables = cbsodata.get_table_list()
>>> print(tables[0])
{'Catalog': 'CBS',
 'ColumnCount': 18,
 'DefaultPresentation': '_la=nl&_si=&_gu=&_ed=LandVanUiteindelijkeZeggenschapUCI&_
↪td=Perioden&graphType=line',
 'DefaultSelection': "$filter=((LandVanUiteindelijkeZeggenschapUCI eq '11111') or␣
↪(LandVanUiteindelijkeZeggenschapUCI eq '22222')) and (Bedrijfsgrootte eq '10000')␣
↪and (substringof('JJ',Perioden))&$select=LandVanUiteindelijkeZeggenschapUCI,␣
↪Bedrijfsgrootte, Perioden, FiscaalJaarloonPerBaan_15",
 'ExplanatoryText': '',
 'Frequency': 'Perjaar',
 'GraphTypes': 'Table,Bar,Line',
 'ID': 0,
 'Identifier': '82010NED',
 'Language': 'nl',
 'MetaDataModified': '2014-02-04T02:00:00',
 'Modified': '2014-02-04T02:00:00',
 'OutputStatus': 'Regulier',
 'Period': '2008 t/m 2011',
 'ReasonDelivery': 'Actualisering',
 'RecordCount': 32,
 'SearchPriority': '2',
 'ShortDescription': '\nDeze tabel bevat informatie over banen en lonen bij bedrijven␣
↪in Nederland, uitgesplitst naar het land van uiteindelijke zeggenschap van die␣
↪bedrijven. Hierbij wordt onderscheid gemaakt tussen bedrijven onder Nederlandse␣
↪zeggenschap en bedrijven onder buitenlandse zeggenschap. In de tabel zijn alleen de␣
↪bedrijven met werknemers in loondienst meegenomen. De cijfers hebben betrekking op␣
↪het totale aantal banen bij deze bedrijven en de samenstelling van die banen naar␣
↪kenmerken van de werknemers (baanstatus, geslacht, leeftijd, herkomst en hoogte van␣
↪het loon). Ook het gemiddelde fiscale jaarloon per baan is in de tabel te vinden.␣
↪\n\nGegevens beschikbaar vanaf: 2008 \n\nStatus van de cijfers: \nDe cijfers in␣
↪deze tabel zijn definitief.\n\nWijzigingen per 4 februari 2014\nDe cijfers van 2011␣
↪zijn toegevoegd.\n\nWanneer komen er nieuwe cijfers?\nDe cijfers over 2012␣
↪verschijnen in de eerste helft van 2015.\n',
 'ShortTitle': 'Zeggenschap bedrijven; banen, grootte',
 'Source': 'CBS.',
 'Summary': 'Banen en lonen van werknemers bij bedrijven in Nederland\nnaar land van␣
↪uiteindelijke zeggenschap en bedrijfsgrootte',
 'SummaryAndLinks': 'Banen en lonen van werknemers bij bedrijven in Nederland<br />
↪naar land van uiteindelijke zeggenschap en bedrijfsgrootte<br /><a href="http://
↪opendata.cbs.nl/ODataApi/OData/82010NED">http://opendata.cbs.nl/ODataApi/OData/
↪82010NED</a><br /><a href="http://opendata.cbs.nl/ODataFeed/OData/82010NED">http://
↪opendata.cbs.nl/ODataFeed/OData/82010NED</a>',
 'Title': 'Zeggenschap bedrijven in Nederland; banen en lonen, bedrijfsgrootte',
 'Updated': '2014-02-04T02:00:00'}
```

1.2.2 Info

Get information about a table with the `get_info` function.

```
>>> info = cbsodata.get_info('82070ENG') # Returns a dict with info
>>> info['Title']
'Caribbean Netherlands; employed labour force characteristics 2012'
>>> info['Modified']
'2013-11-28T15:00:00'
```

1.2.3 Data

The function you are looking for!! The function `get_data` returns a list of dicts with the table data.

```
>>> data = cbsodata.get_data('82070ENG')
[{'CaribbeanNetherlands': 'Bonaire',
  'EmployedLabourForceInternatDef_1': 8837,
  'EmployedLabourForceNationalDef_2': 8559,
  'Gender': 'Total male and female',
  'ID': 0,
  'Periods': '2012',
  'PersonalCharacteristics': 'Total personal characteristics'},
 {'CaribbeanNetherlands': 'St. Eustatius',
  'EmployedLabourForceInternatDef_1': 2099,
  'EmployedLabourForceNationalDef_2': 1940,
  'Gender': 'Total male and female',
  'ID': 1,
  'Periods': '2012',
  'PersonalCharacteristics': 'Total personal characteristics'},
 {'CaribbeanNetherlands': 'Saba',
  'EmployedLabourForceInternatDef_1': 1045,
  'EmployedLabourForceNationalDef_2': 971,
  'Gender': 'Total male and female',
  'ID': 2,
  'Periods': '2012',
  'PersonalCharacteristics': 'Total personal characteristics'},
 # ...
]
```

The keyword argument `dir` can be used to download the data directly to your file system.

```
>>> data = cbsodata.get_data('82070ENG', dir="dir_to_save_data")
```

1.2.4 Catalogs (dataderden)

There are multiple ways to retrieve data from catalogs other than 'opendata.cbs.nl'. The code below shows 3 different ways to retrieve data from the catalog 'dataderden.cbs.nl' (known from Iv3).

On module level.

```
cbsodata.options.catalog_url = 'dataderden.cbs.nl'
# list tables
cbsodata.get_table_list()
# get dataset 47003NED
cbsodata.get_data('47003NED')
```

With context managers.

```
with cbsodata.catalog('dataderden.cbs.nl'):
    # list tables
    cbsodata.get_table_list()
    # get dataset 47003NED
    cbsodata.get_data('47003NED')
```

As a function argument.

```
# list tables
cbsodata.get_table_list(catalog_url='dataderden.cbs.nl')
# get dataset 47003NED
cbsodata.get_data('47003NED', catalog_url='dataderden.cbs.nl')
```

1.2.5 Pandas users

The package works well with Pandas. Convert the result easily into a pandas DataFrame with the code below.

```
>>> data = pandas.DataFrame(cbsodata.get_data('82070ENG'))
>>> data.head()
```

The list of tables can be turned into a pandas DataFrame as well.

```
>>> tables = pandas.DataFrame(cbsodata.get_table_list())
>>> tables.head()
```

1.3 Command Line Interface

This library ships with a Command Line Interface (CLI).

```
> cbsodata -h
usage: cbsodata [-h] [--version] [subcommand]

CBS Open Data: Command Line Interface

positional arguments:
  subcommand  the subcommand (one of 'data', 'info', 'list')

optional arguments:
  -h, --help  show this help message and exit
  --version   show the package version
```

Download data:

```
> cbsodata data 82010NED
```

Retrieve table information:

```
> cbsodata info 82010NED
```

Retrieve a list with all tables:

```
> cbsodata list
```

1.3.1 Export data

Use the flag `-o` to load data to a file (JSON lines).

```
> cbsodata data 82010NED -o table_82010NED.jl
```

CHAPTER 2

Reference

Statistics Netherlands opendata API client for Python

Python Module Index

C

cbsodata, 7

Index

C

`cbsodata` (*module*), 7