

PYNSEE

Package python fournissant des outils pour faciliter la recherche et la collecte des données provenant de l'INSEE et de l'IGN

<https://pynsee.readthedocs.io/en/latest/index.html>

Hadrien Leclerc

CONSTAT DE DEPART

- Chercher et utiliser les données provenant du site insee.fr reste compliqué
- Dépendre des fichiers Excel téléchargés à la main signifie que les résultats sont plus difficilement reproductibles et partageables
- Créer des fonctions d'importation et de formatage des données

>> Besoin d'outils qui facilite l'obtention des données INSEE

The screenshot shows the INSEE website homepage. At the top left is the INSEE logo with the tagline "Mesurer pour comprendre". To the right is the text "Institut national de la statistique et des études économiques" and a search bar with the placeholder "Rechercher sur le site". A navigation menu below the header includes "STATISTIQUES ET ÉTUDES", "DÉFINITIONS, MÉTHODES ET QUALITÉ", "SERVICES", and "L'INSEE ET LA STATISTIQUE PUBLIQUE". Below the navigation menu is a descriptive sentence: "L'Institut national de la statistique et des études économiques collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises". The main content area features four key statistics: Population (68,04M), Inflation (+6,0%), Chômage (7,3%), and Croissance (+0,1%). At the bottom right, there is a "Accès rapide" button with an exclamation mark icon.

Institut national de la statistique et des études économiques

Mesurer pour comprendre

• Menu • Contenu • Blog • Espace presse • Aide • English

Rechercher sur le site

STATISTIQUES ET ÉTUDES DÉFINITIONS, MÉTHODES ET QUALITÉ SERVICES L'INSEE ET LA STATISTIQUE PUBLIQUE

L'Institut national de la statistique et des études économiques collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises

| | | | |
|---|---|---|--|
|  68,04^M Population |  + 6,0% Inflation |  7,3% Chômage |  + 0,1% Croissance |
|---|---|---|--|

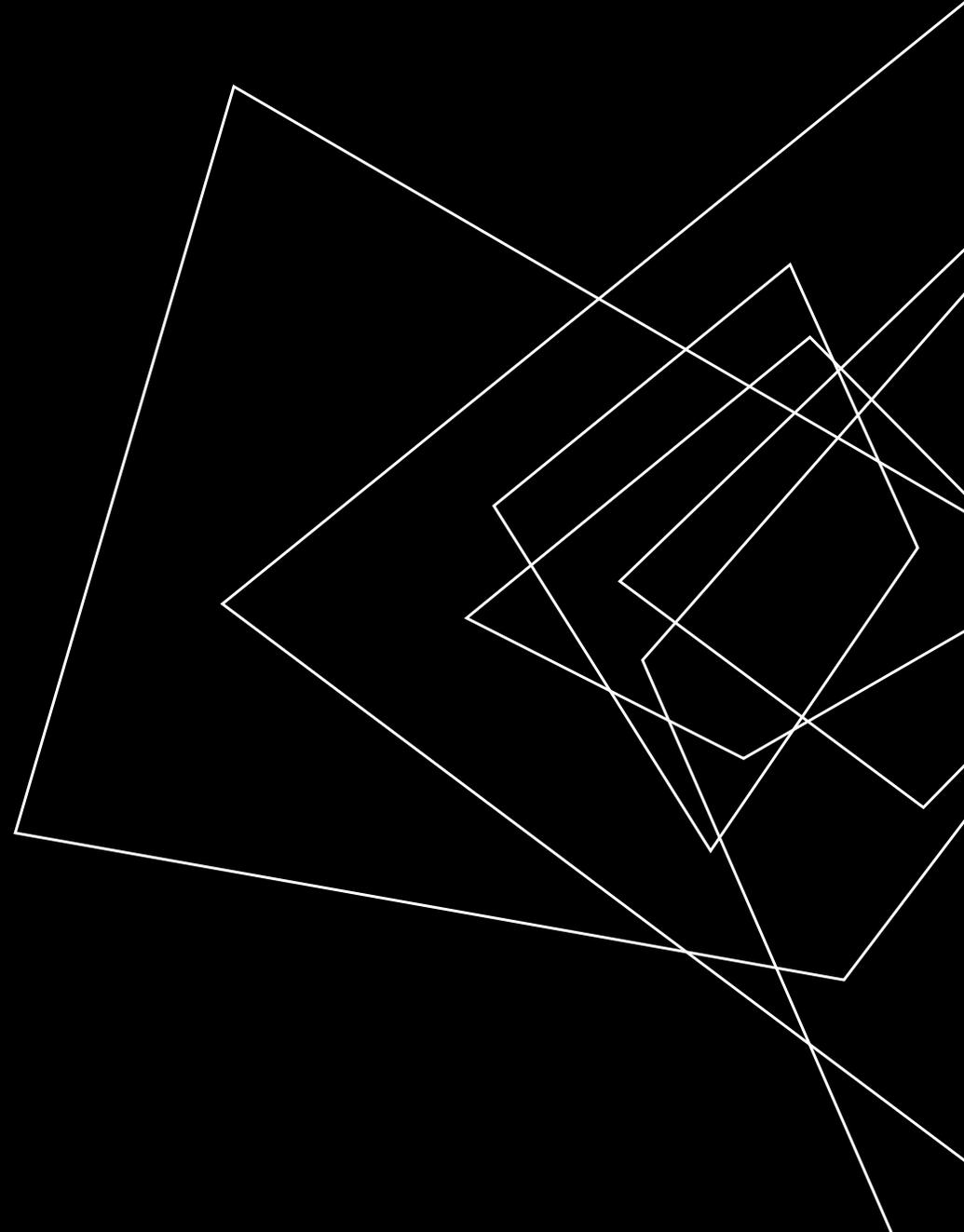
Accès rapide

PYNSEE PACKAGE

Installation

Données
disponibles

Exemples

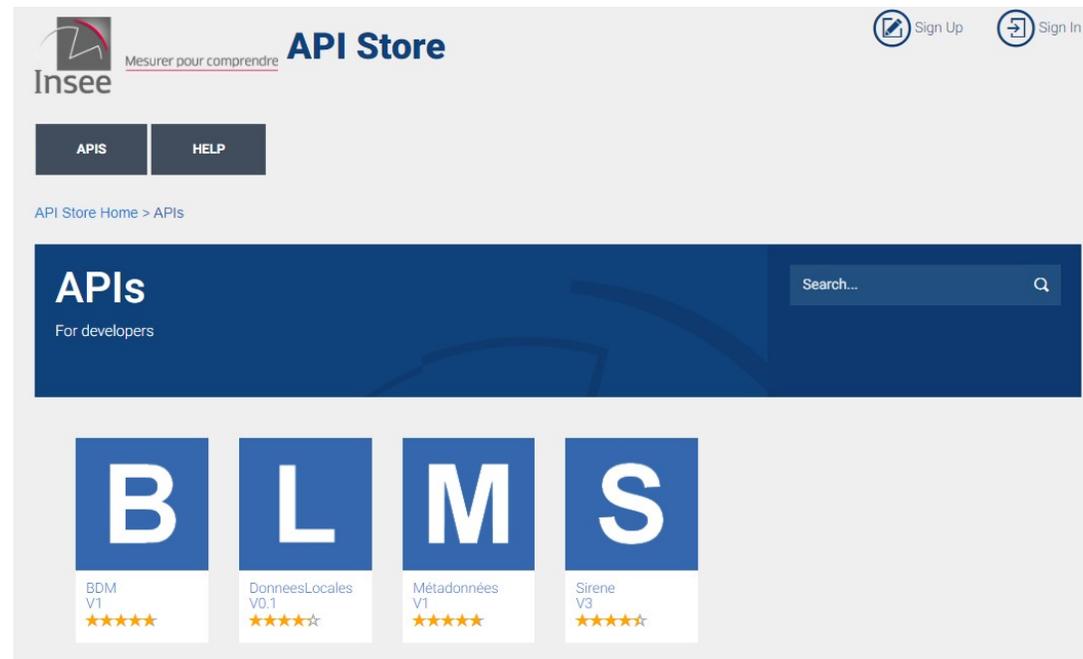


Installation du package pynsee et souscription aux APIs INSEE

```
# Download Pypi package  
pip install pynsee[full]
```

<https://api.insee.fr/catalogue/>

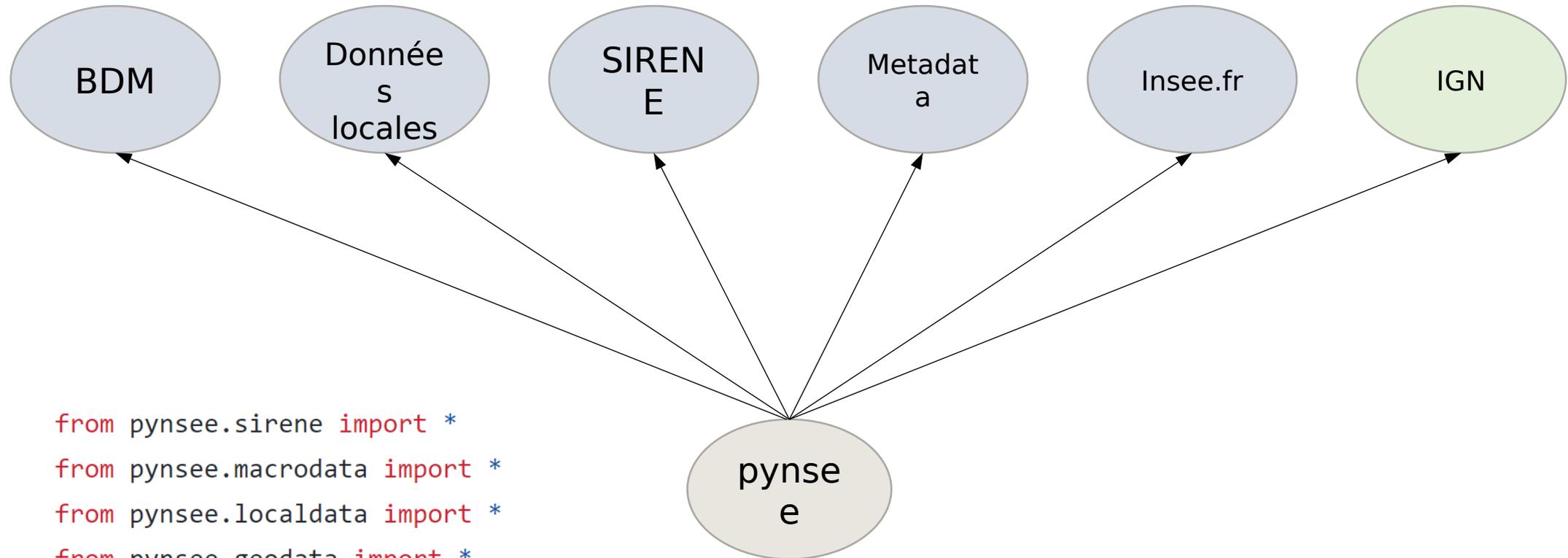
```
# Subscribe to api.insee.fr and get your credentials!  
# Save once and for all your credentials with init_conn function.  
# Then, functions requiring authentication will use the credentials saved locally on your machine by innit_conn  
from pynsee.utils.init_conn import init_conn  
init_conn(insee_key="my_insee_key", insee_secret="my_insee_secret")
```



The screenshot shows the INSEE API Store interface. At the top, there is the INSEE logo with the tagline "Mesurer pour comprendre" and the text "API Store". To the right, there are "Sign Up" and "Sign In" buttons. Below the header, there are two buttons: "APIS" and "HELP". A breadcrumb trail reads "API Store Home > APIs". A large blue banner contains the text "APIs" and "For developers", along with a search bar labeled "Search...". Below the banner, four API cards are displayed, each with a large letter in a blue square and a star rating below it:

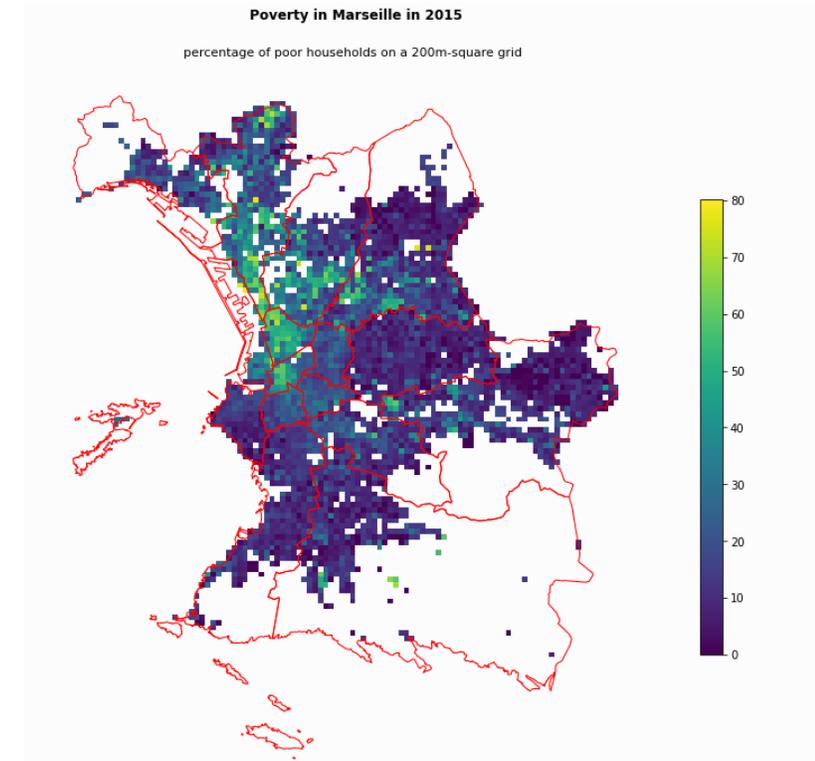
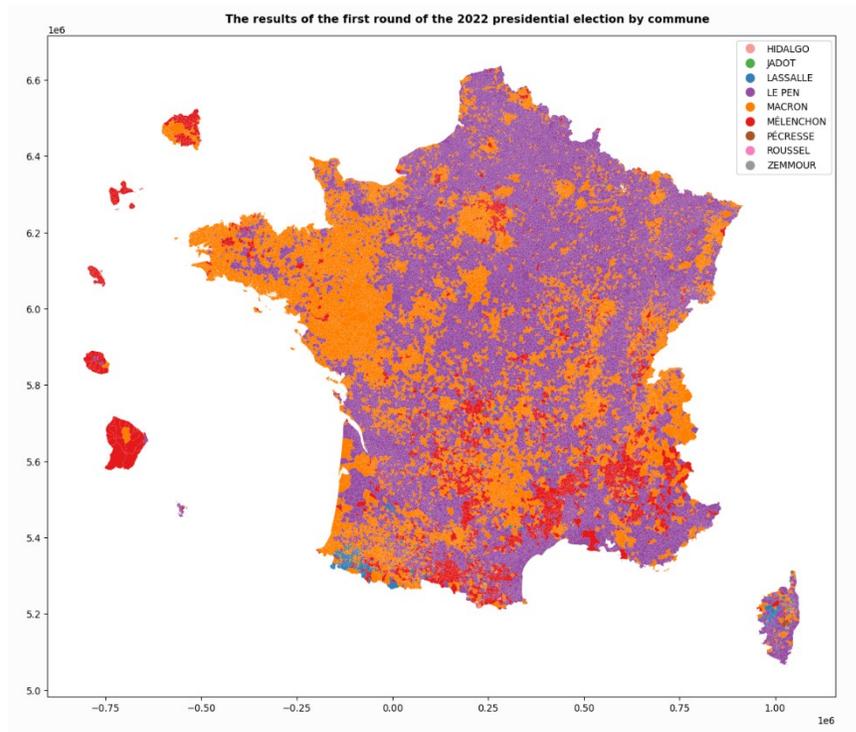
- B**: BDM V1, 5 stars
- L**: DonneesLocales V0.1, 4.5 stars
- M**: Métadonnées V1, 5 stars
- S**: Sirene V3, 5 stars

Les données accessibles : INSEE et IGN



```
from pynsee.sirene import *  
from pynsee.macrodata import *  
from pynsee.localdata import *  
from pynsee.geodata import *  
from pynsee.metadata import *  
from pynsee.utils import *  
from pynsee.download import *
```

IGN: contours géographiques des communes, départements et régions



```
from pyntee.geodata import get_geodata_list, get_geodata
```

```
# get geographical data list
```

```
geodata_list = get_geodata_list()
```

```
# get departments geographical limits
```

```
com = get_geodata('ADMINEXPRESS-COG-CARTO.LATEST:commune')
```

```
mapcom = mapcom.translate(departement = ['971', '972', '974', '973', '976'],  
                          factor = [1.5, 1.5, 1.5, 0.35, 1.5])
```

```
mapcom = mapcom.zoom(departement = ["75", "92", "93", "91", "77", "78", "95", "94"],  
                    factor=1.5, startAngle = math.pi * (1 - 3 * 1/9))
```

BDM : banque de données macroéconomiques – plus de 150 000 séries temporelles

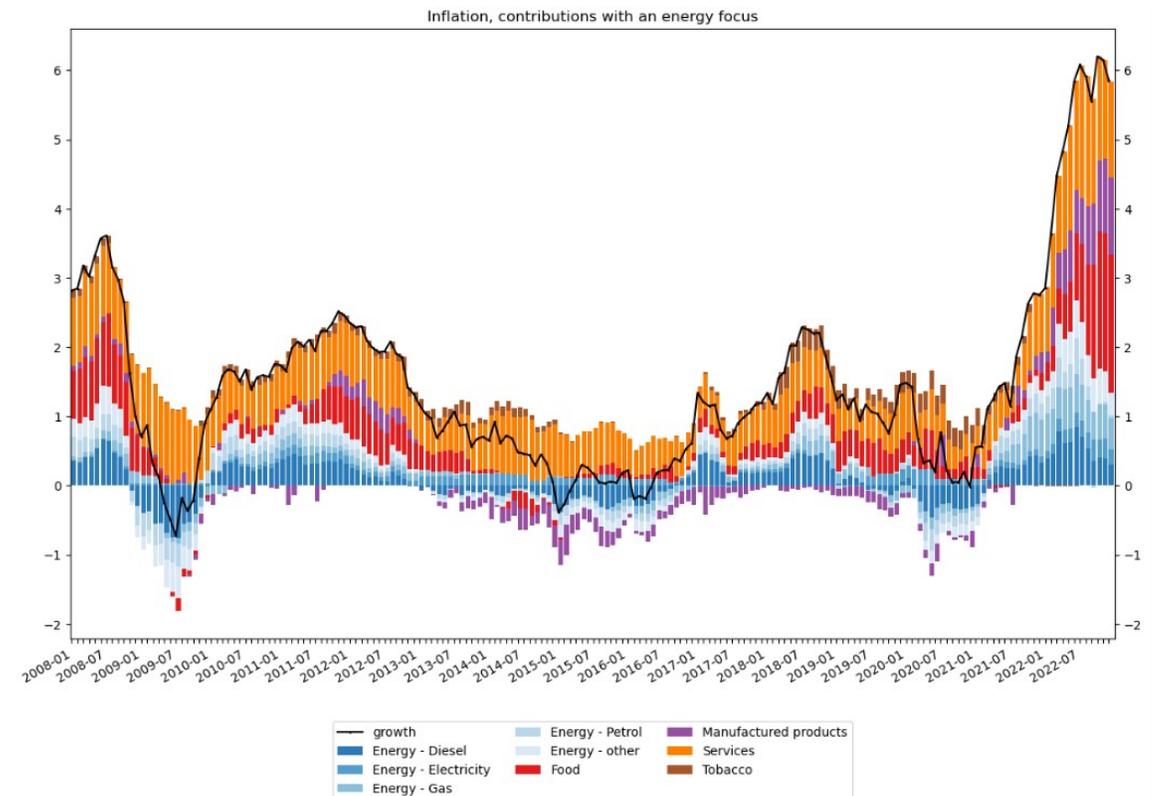
- Comptes nationaux (annuels et trimestriels) : PIB, consommation des ménages, investissement
- Balances commerciales et des paiements
- Données emploi et de chômage
- Enquêtes de conjoncture au près des ménages et des entreprises
- Comptes de l'Etat
- Démographie : naissances, décès et mariages
- Inflation et indices de prix
- Données régionales
- Indicateurs de développement durables

```
from pynsee.macrodatab.get_series_list import get_series_list
from pynsee.macrodatab.get_series import get_series
```

```
insee_data = get_dataset_list()
```

```
# download a list of inflation dataset series identifiers
series = get_series_list('IPC-2015')
```

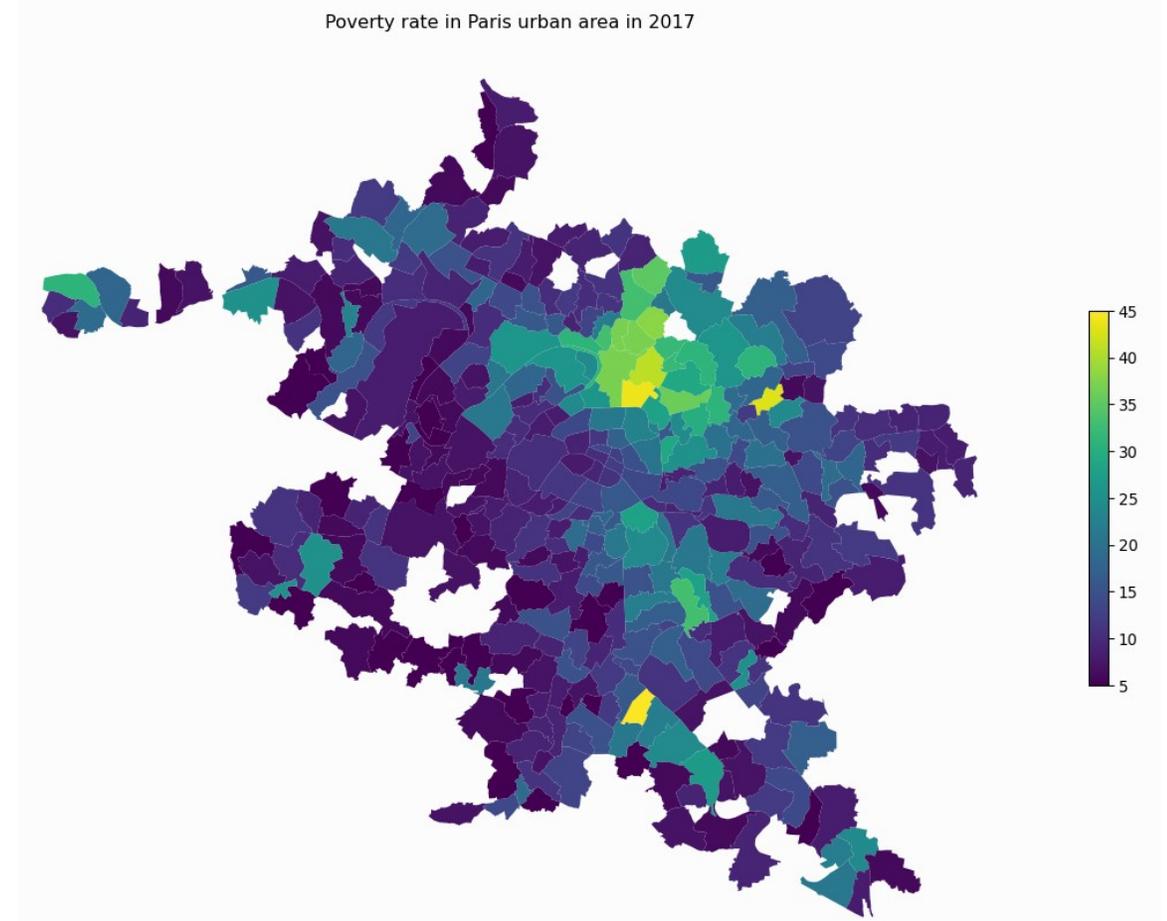
```
# get data
data = get_series(df_idbank.IDBANK)
```



Données locales : données aux échelles communale, départementale, régionale et nationale

- Recensement de la population
- Indicateurs de la base des données communales
- Etat civil
- Répertoire des entreprises
- Indicateurs de revenus
- Données sur le tourisme

```
from pynsee.localdata import get_local_metadata,  
  
# get numeric values from INSEE database  
dataParis = get_local_data(dataset_version='GEO2020FILO2017',  
                           variables = 'INDICS_FILO_DISP_DET',  
                           nivgeo = 'COM',  
                           geocodes = code_com_paris)
```



SIRENE : banque de données contenant des informations sur près de 32 millions d'établissements dont 13 millions sont actifs

-
- Numéro siret/siren (identifiants)
- Date de création et de fin d'activité
- Dénomination (noms)
- Catégorie d'activité selon la NAF
- Adresse
- Nombre d'employés (intervalle)
- Nom et prénom (entrepreneur individuel)

```
from pynsee.sirene import search_sirene
```

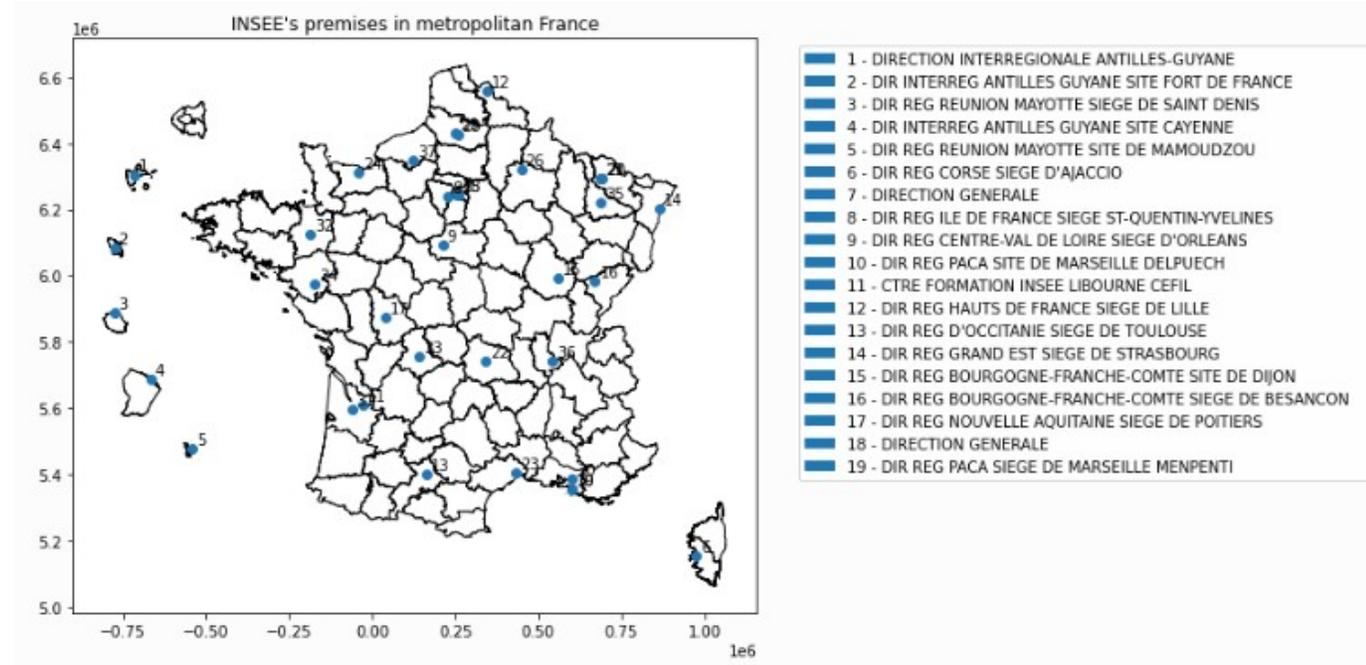
```
variable = ["denominationUniteLegale", 'sigleUniteLegale', 'categorieJuridiqueUniteLegale']
```

```
insee_pattern = "INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES"
```

```
pattern = [insee_pattern, 'INSEE', '7120']
```

```
# 7120 : Service central d'un ministère
```

```
data = search_sirene(variable = variable, pattern = pattern, kind="siret")
```



Metadata : définition de concepts INSEE et nomenclature d'activité NAF

| | A10 | A129 | A17 | A21 | A38 | A5 | A64 | A88 | NAF1 | NAF2 | NAF3 | NAF4 | NAF5 | TITLE_NAF5_40CH_FR | TITLE_NAF5_65CH_FR | TITLE_NAF5_FR |
|-----|-----|------|-----|-----|-----|-----|-----|-----|------|------|------|-------|--------|--|---|---|
| 0 | AZ | A01Z | AZ | A | AZ | AZ | AZ1 | 01 | A | 01 | 01.1 | 01.11 | 01.11Z | Cult céréale, légumineuse, graine oléag. | Culture de céréales (sf riz) légumineuses, gra... | Culture de céréales (à l'exception du riz), de... |
| 1 | AZ | A01Z | AZ | A | AZ | AZ | AZ1 | 01 | A | 01 | 01.1 | 01.12 | 01.12Z | Culture du riz | Culture du riz | Culture du riz |
| 2 | AZ | A01Z | AZ | A | AZ | AZ | AZ1 | 01 | A | 01 | 01.1 | 01.13 | 01.13Z | Cult. légume, melon, racine & tubercule | Culture de légumes, de melons, de racines et d... | Culture de légumes, de melons, de racines et d... |
| 3 | AZ | A01Z | AZ | A | AZ | AZ | AZ1 | 01 | A | 01 | 01.1 | 01.14 | 01.14Z | Culture de la canne à sucre | Culture de la canne à sucre | Culture de la canne à sucre |
| 4 | AZ | A01Z | AZ | A | AZ | AZ | AZ1 | 01 | A | 01 | 01.1 | 01.15 | 01.15Z | Culture du tabac | Culture du tabac | Culture du tabac |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 727 | RU | S96Z | RU | S | SZ | GU | SZ3 | 96 | S | 96 | 96.0 | 96.09 | 96.09Z | Autres services personnels n.c.a. | Autres services personnels n.c.a. | Autres services personnels n.c.a. |
| 728 | RU | T97Z | RU | T | TZ | GU | TZ0 | 97 | T | 97 | 97.0 | 97.00 | 97.00Z | Act. ménage: empl. de person. domestique | Activités des ménages : employeurs de personn... | Activités des ménages en tant qu'employeurs de... |
| 729 | RU | T98Z | RU | T | TZ | GU | TZ0 | 98 | T | 98 | 98.1 | 98.10 | 98.10Z | Act. ménage : prod. biens (usage propre) | Activ. indifférenciées ménages : producteurs b... | Activités indifférenciées des ménages en tant ... |
| 730 | RU | T98Z | RU | T | TZ | GU | TZ0 | 98 | T | 98 | 98.2 | 98.20 | 98.20Z | Act. ménage : prod. serv. (usage propre) | Activ. indifférenciées ménages : produc. serv... | Activités indifférenciées des ménages en tant ... |
| 731 | RU | U99Z | RU | U | UZ | GU | UZ0 | 99 | U | 99 | 99.0 | 99.00 | 99.00Z | Act. organisations extraterritoriales | Activités des organisations et organismes extr... | Activités des organisations et organismes extr... |

```

from pynsee.metadata import get_activity_list
activity_A138 = get_activity_list('A138')
activity_NAF3 = get_activity_list('NAF3')
activity_NAF5 = get_activity_list('NAF5')

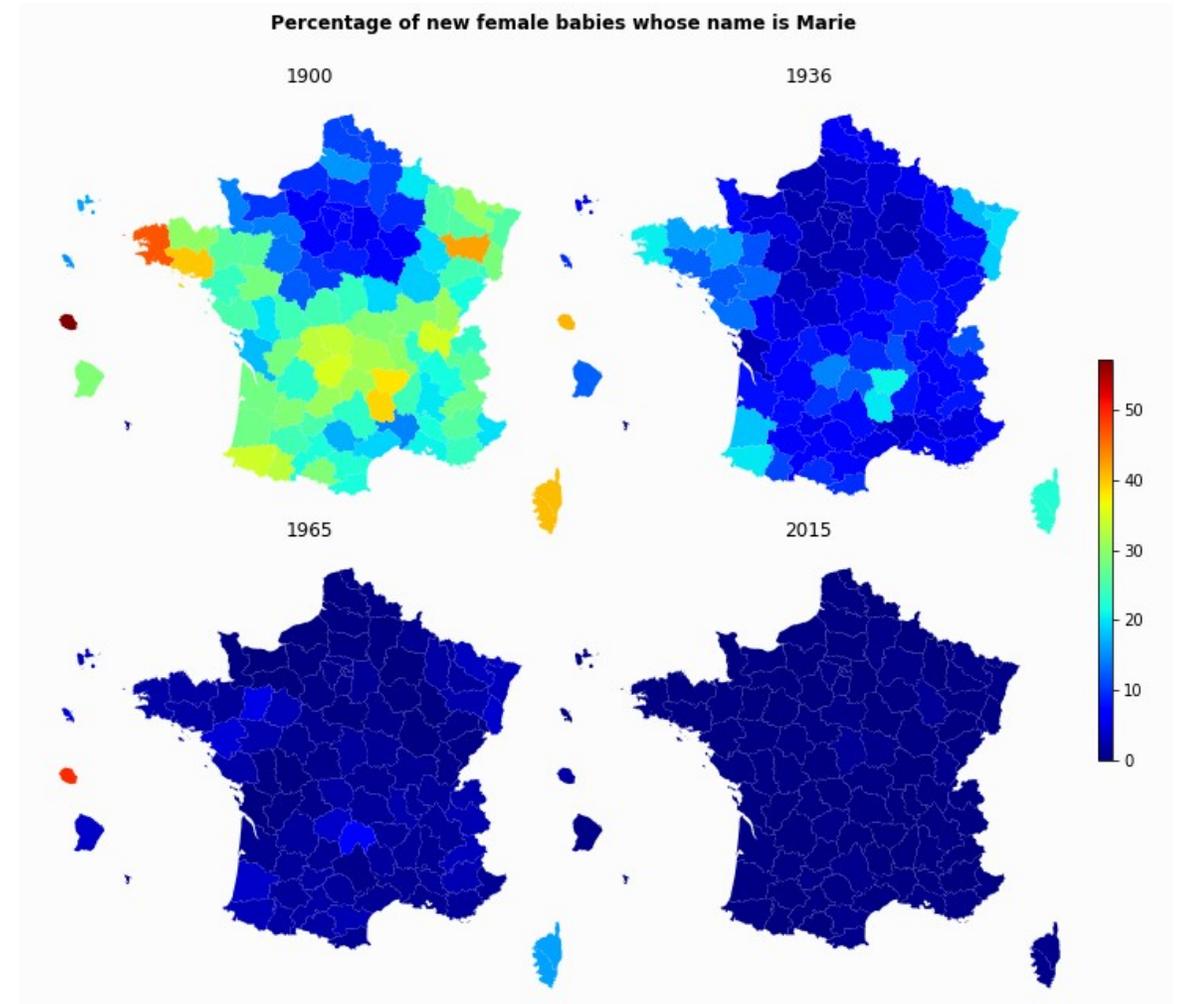
```

Insee.fr : liste de plus de 1300 fichiers disponibles

- Recensement de la population (RP)
- Estimations d'emploi (ESTEL)
- Revenus et pauvreté des ménages à échelle fine (Filosofi)
- Etat civil par commune
- Fichier des prénoms à la naissance
- Code officiel géographique des communes
-

```
from pyinsee.download import download_file, get_file_list
```

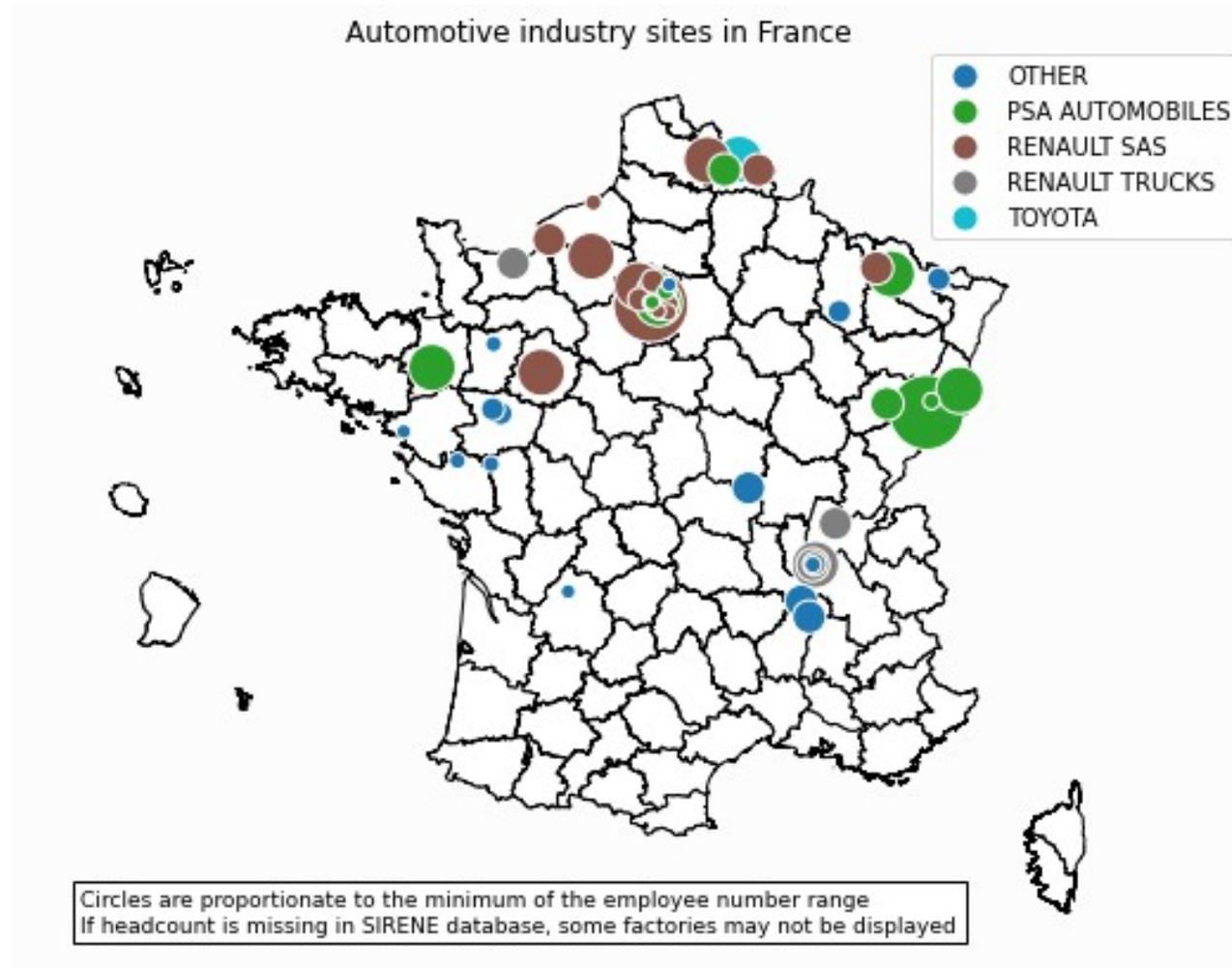
```
meta = get_file_list()  
dfraw = download_file("PRENOM_DEP")
```



Avantages du package python pynsee:

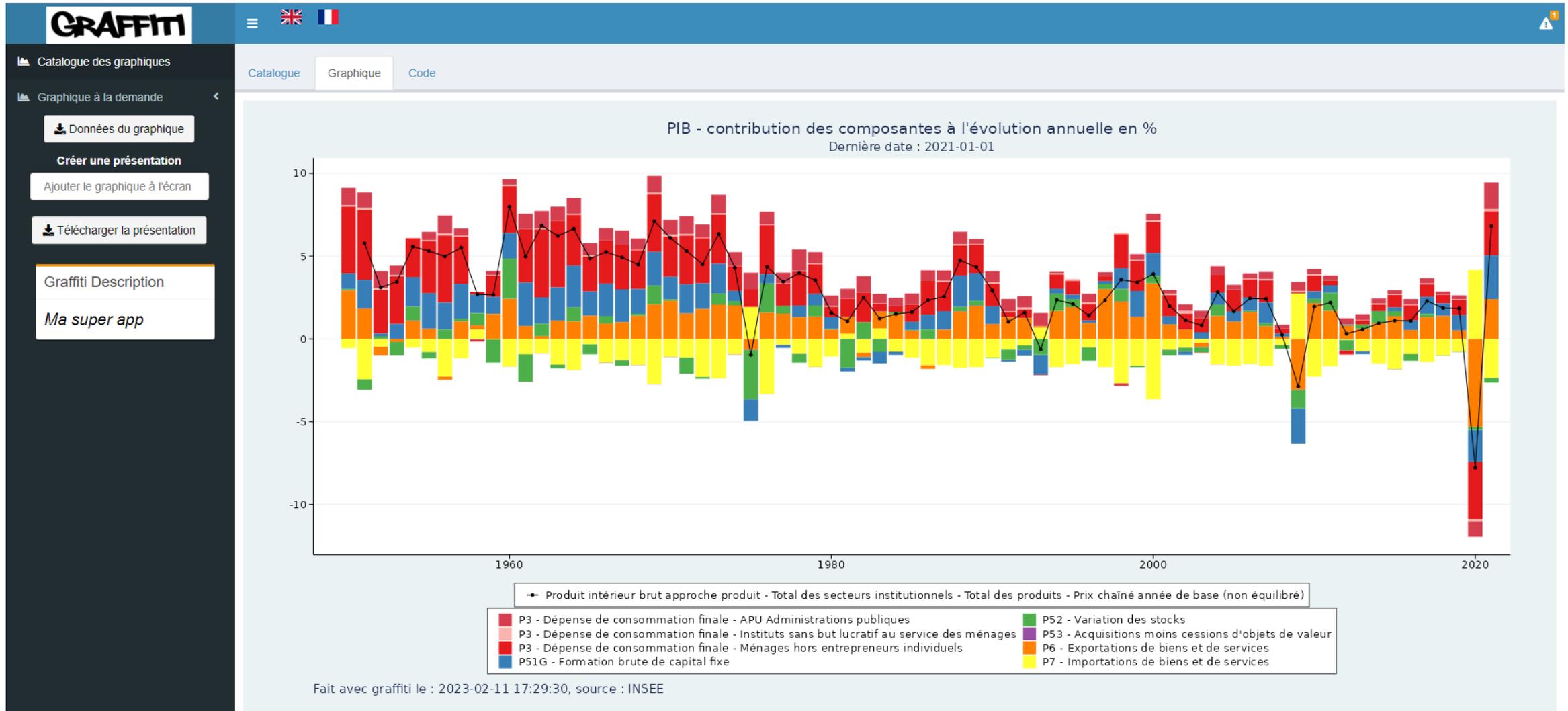
- Simple d'utilisation, logique commune : fonctions de recherche et de collecte des données
- Dernière données disponibles
-
- Travail dont les résultats sont reproductibles et facilement partageables
-
- Données sont sauvées en local sur la machine : confort pour l'utilisateur
-
- Réunion dans un même logiciel de plusieurs sources données, ce qui facilite leur utilisation conjointe
-
- Tests exhaustifs sur les fonctions, couverture > 90%
-
-
-

Exemple : Quelle est la géographie de l'industrie automobile en France?



Graffiti : application expérimentale de visualisation des données de la BDM

<https://graffiti.lab.sspcloud.fr>



Futurs développements:

- Refonte de la documentation en ligne pour enlever les notebooks jupyter
- Passage à une version de shapely > 2.0
- Adaptation à Sirene V4
- Exemples d'utilisation à des fins de modélisations
- Rendre accessible le package sur conda-forge
-

Mettons en commun nos efforts: toutes les contributions sont les bienvenues!

-



Contributors 6



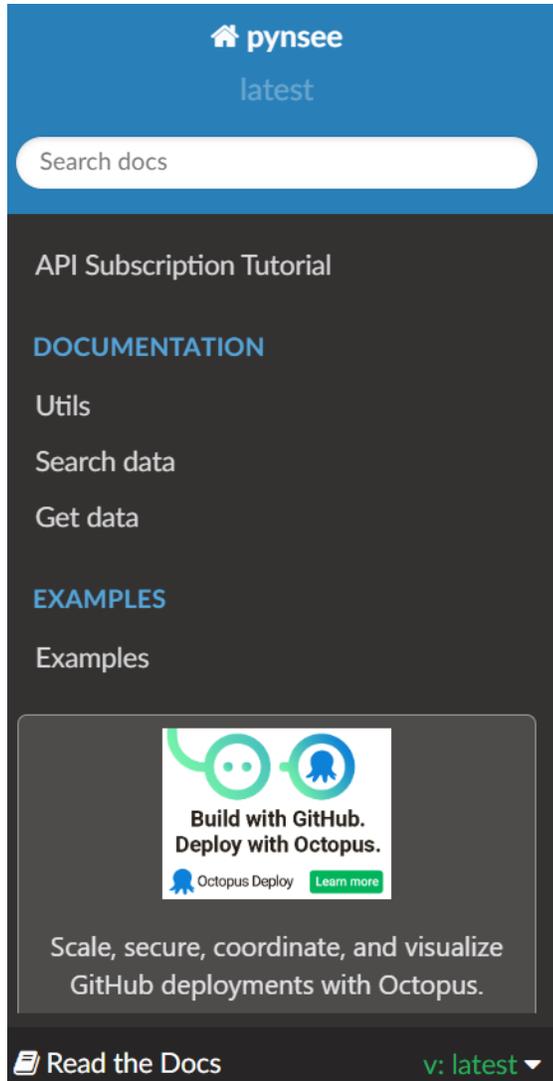
Remerciements:

Lino Galiana
Raphael Adjrad
Thomas Grandjean
Tanguy Fardet

équipes qui travaillent
sur les APIs à l'INSEE et à l'IGN

Documentation en ligne

<https://pynsee.readthedocs.io/en/latest/index.html>



The sidebar of the pynsee documentation website. At the top, it features the 'pynsee' logo and the word 'latest'. Below this is a search bar labeled 'Search docs'. The main navigation menu includes 'API Subscription Tutorial', 'DOCUMENTATION', 'Utils', 'Search data', and 'Get data'. Under 'EXAMPLES', there is a link to 'Examples'. At the bottom, there is a promotional banner for Octopus Deploy with the text 'Build with GitHub. Deploy with Octopus.' and 'Scale, secure, coordinate, and visualize GitHub deployments with Octopus.' The footer of the sidebar contains a 'Read the Docs' button and a version selector set to 'v: latest'.

» Introduction to pynsee package

[Edit on GitHub](#)

Introduction to pynsee package

pypi package 0.1.1 pynsee package tests passing examples tests passing codecov 92% docs failing
python 3.7 | 3.8 | 3.9 | 3.10 code style black downloads 99/month

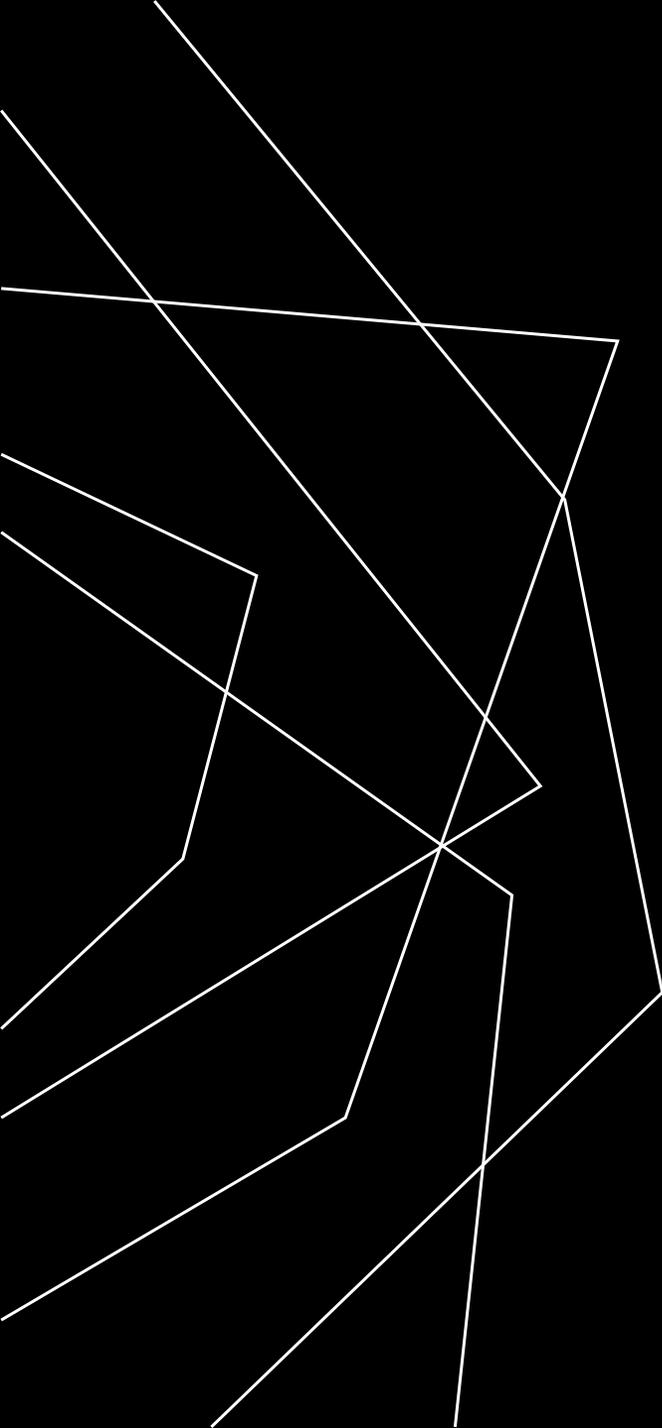
pynsee package contains tools to easily search and download French data from INSEE and IGN APIs.

pynsee gives a quick access to more than 150 000 macroeconomic series, a dozen datasets of local data, numerous sources available on insee.fr as well as key metadata and SIRENE database containing data on all French companies. Have a look at the detailed API page api.insee.fr.

This package is a contribution to reproducible research and public data transparency. It benefits from the developments made by teams working on APIs at INSEE and IGN.

Installation & API subscription

The files available on insee.fr and IGN data, i.e. the use of *download* and *geodata* modules, do not require authentication. Credentials are necessary to access some of the INSEE APIs available through *pynsee* by the modules *macrodata*, *localdata*,



MERCI

Hadrien Leclerc

leclerc.hadrien@gmail.com

Github: [hadrilec](#)