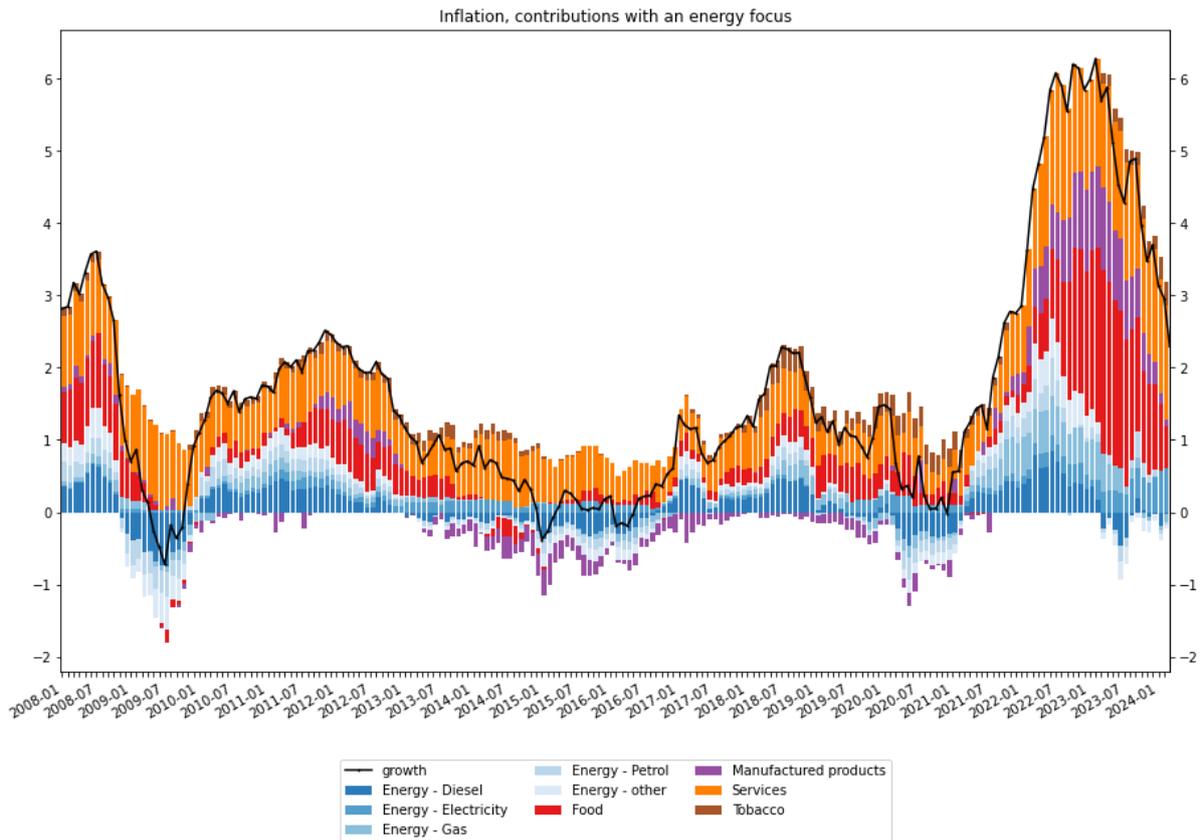


EXPLOITER LES DONNEES PUBLIQUES DE L'INSEE

AVEC PYTHON, MINICONDA ET L'API PYNSEE.

GERARD (gerard093)



L'objet du présent document est de partager des notes prises au cours de l'installation des API de l'INSEE dédiées à l'exploitation de ses bases de données (OPEN DATA) Les environnements de travail utilisés sont open source, et les codes source présentés ici sont donc libres de droits, mais restent néanmoins à la base la propriété des sites web de publication. Le but de ce document est de faciliter l'accès aux données publiques pour des débutants.

Notes d'installation de Miniconda

Le logiciel est un environnement de travail très utilisé en data analytics. Cette application consomme toutefois moins de ressource qu'Anaconda, et permet de gérer des configurations spécifiques à un projet. L'intérêt de ce produit est de gérer automatiquement les dépendances entre packages installés.

<https://docs.anaconda.com/free/miniconda/miniconda-install/>

Le logiciel est disponible sur les environnements Windows, Mac et Linux

Une fois l'installation de miniconda terminée, il est possible de contrôler le bon fonctionnement

tapez la ligne de commande (dans votre console) :

[conda activate](#)

Si ceci ne provoque pas une réponse correcte dans votre console, vérifiez le contenu de votre variable PATH qui doit contenir le chemin d'accès aux ressources contenues dans miniconda3/condabin

tapez maintenant

[conda list](#)

la liste des packages installés apparaît. Vérifiez que vous avez bien le package pip qui vous sera utile.

Création d'un environnement de travail

L'étape suivante est de créer un environnement spécifique au projet, lequel contiendra les dépendances nécessaires. Créons par exemple l'environnement pytest88 :

[conda create -n pytest88](#)

Comme le message en réponse l'indique, pour pouvoir accéder à ce nouvel environnement, il suffit de taper :

[conda activate pytest88](#)

Tout ce qui sera effectué en terme de gestion de package sera désormais effectué dans l'environnement auquel on est désormais connecté, et ceci jusqu'à la demande de sortie qui s'effectue par la commande [conda deactivate](#)

Documentation

Au delà des ressources documentaires disponibles sur le site web de miniconda, il est utile de télécharger le document pdf : <https://anaconda.cloud/conda-cheatsheet>

Ce document contient les commandes de base utilisables sous miniconda.

Installation de Spyder sous conda

Spyder est un éditeur puissant utilisé par les scientifiques pour gérer les codes sources. Il est composé à gauche d'un éditeur de fichier source, et à droite d'une fenêtre d'exploration des variables et données en mémoire, des graphiques, et en dessous de cette fenêtre, d'un environnement d'exécution d'IPython. Il suffit donc de copier les lignes de code du source et de les coller dans la fenêtre IPython, puis de faire enter pour provoquer l'exécution du bloc de lignes copiées

L'installation de Spyder se fait par la commande [conda install spyder](#) et pour lancer spyder il suffit de taper [spyder](#) ou [spyder &](#) (linux)

Il est possible d'enlever certains packages comme mysql qui consomment de la place et sont difficiles à gérer : [conda remove -n pytest88 mysql](#)

La plupart des packages sont gérés par conda install, mais ce n'est pas le cas de l'API de l'INSEE. Aussi faudra t-il taper cette fois la commande : [pip install pynsee](#)

Ceci est conforme à la méthode recommandée par l'INSEE sur son site : <https://pypi.org/project/pynsee/>

Il est recommandé d'aller sur le site, ci dessus.

A partir de ce site racine, il est utile de cliquer sur lien vers les api de l'insee, pour ouvrir dans un nouvel onglet du navigateur : <https://api.insee.fr/catalogue/>

Sur cet onglet, cliquer sur le bouton «Créer son compte»

A partir du moment où le compte est créé, il faut se connecter à partir des données de connexion transmises par l'INSEE, pour créer un projet personnel. A ce projet, on rattachera les ressources data proposées par l'INSEE (BDM, Données locales, Métadonnées, Sirene) et on demandera en naviguant dans les onglets du projet la génération de mots de passe («credentials») Attention à bien renseigner le nombre de jours où les mots de passe seront valides - par exemple en remplaçant 1 par 365 !

Ce sont les identifiants et mots de passe ainsi générés par l'INSEE qui vont être utilisés dans le code python ci-après :

```
# Subscribe to api.insee.fr and get your credentials!  
# Save your credentials with init_conn function :  
  
from pynsee.utils.init_conn import init_conn  
init_conn(insee_key="my_insee_key",insee_secret="my_insee_secret")  
  
# Beware : any change to the keys should be tested after having cleared the cache  
# Please do : from pynsee.utils import clear_all_cache; clear_all_cache()
```

UN EXEMPLE FOURNI PAR L'INSEE

L'insee fournit une série d'exemples qui exploitent ses Api, afin d'en souligner l'intérêt. En suivant le lien suivant, on peut obtenir un graphique qui permet d'analyser les contributions à l'inflation :

https://pynsee.readthedocs.io/en/latest/examples/example_inflation_contribution.html

Lorsqu'un package est manquant, lors de l'exécution du programme dans la console de spyder, derrière

```
import matplotlib
```

on a le message d'erreur

Traceback (most recent call last):

```
Cell In[1], line 1  
import matplotlib
```

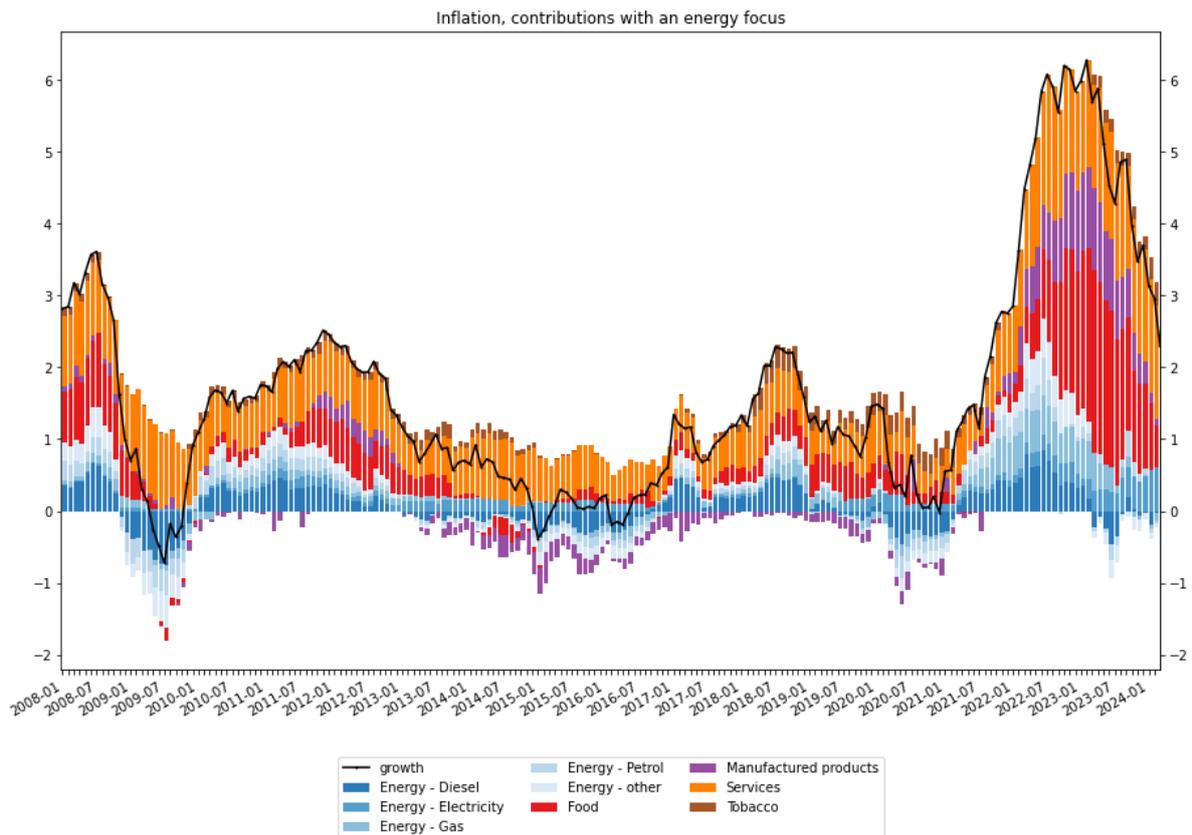
ModuleNotFoundError: No module named 'matplotlib'

La solution consiste, à partir de la ligne de commande de conda, à installer le package manquant par

```
conda install matplotlib
```

Il faut répéter cette opération pour tous les packages manquant.

Les graphes produits se trouvent dans la fenêtre sous l'onglet graphe, à droite, sur spyder. Voici un exemple :



CONCLUSION :

Les API de l'Insee fonctionnent donc sous l'environnement de Miniconda et/ou Anaconda, en Avril 2024. Vous pouvez utiliser les recommandations ci dessus pour gérer et installer les applications, les packages et les Api, afin d'enrichir vos analyses économiques.

A vous de jouer !