

# TP3-2023

February 6, 2023

## 1 TP Evolution de la population des communes françaises

Bernard Uguen

Février 2023

### 1.1 Objectifs du TP

- Utiliser le module pandas et les structures de données qu'il introduit (au dessus de numpy)
  - Séries
  - DataFrame
- Pratiquer pandas sur un cas concret (études des communes françaises)
  - Lecture
  - Mise en forme, gestion des données invalides
  - Association entre différentes sources de données (merge)
  - Visualisation (scatter plot)

## 2 Etude des Communes Françaises

```
[10]: from IPython.display import Image
import numpy as np
```

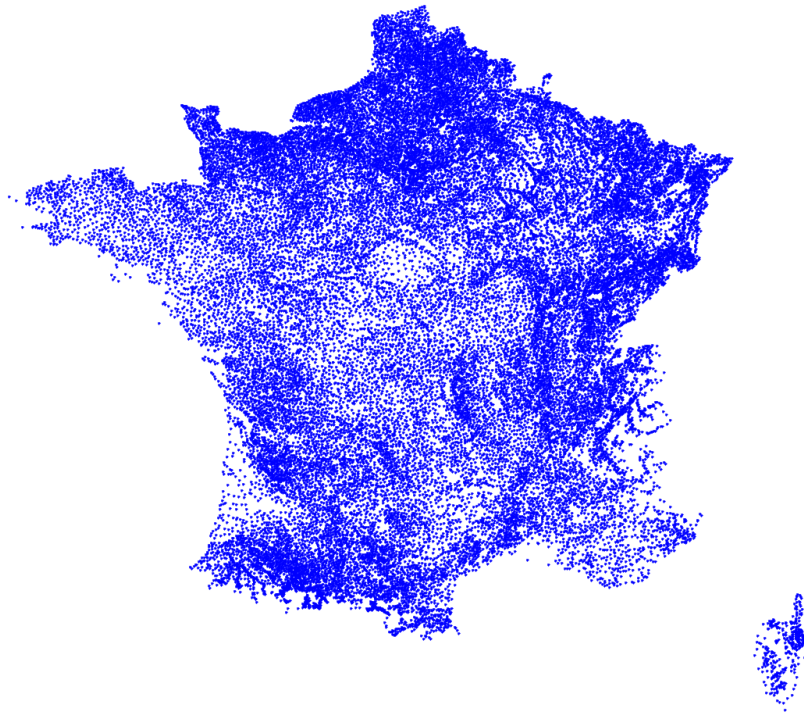
Vous disposez sur moodle d'une archive contenant différents fichiers sélectionnés pour les besoins du TP. En particulier:

- `commune1.csv` est un fichier qui contient l'évolution démographique de toutes les communes française depuis 1975 ainsi que les coordonnées géographiques (longitude, latitude) de la commune. Ce fichier est volumineux et découpé en 2 archives zip.
- `commune2.csv` est un fichier d'une autre source qui contient entre autres les codes postaux des communes
- `maires-25-04-2014.xlsx` contient la liste des maires de France.
- `insee.csv`

Sur la figure ci-dessous chaque point correspond à une commune.

```
[11]: Image('communes.png')
```

```
[11]:
```



## 2.1 Lecture de fichiers

- Utiliser les fonctions de lecture dédiées de pandas pour lire les différents jeux de données fournis
- `pd.read_csv`
- `pd.io.excel.read_excel`

## 2.2 Nettoyage et fusion (Merge) de tous les DataFrame

### 2.2.1 Affichage de la densité de population en France en 1975

### 2.2.2 Affichage des communes selon l'âge de leur maire

Utiliser la fonction `merge` pour fusionner les différents DataFrame fournis en un seul DataFrame contenant toutes l'information.

Cette étape nécessite au préalable d'identifier les colonnes permettant de merger et nécessite une harmonisation des colonnes pivots.

Différentes difficultés vont apparaître liés à des hétérogénéité dans les différentes sources de données.

Problèmes à attendre :

Plusieurs communes françaises portent le même nom, mais elles sont rarement dans le même département. Utiliser la commande `group_by` pour enchaîner des traitement par départements. La syntaxe des noms de communes diffère entre les différentes sources de données, il faudra donc veiller à harmoniser le nom de la commune dans les différents DataFrame.

[8]:

```
!pwd
```

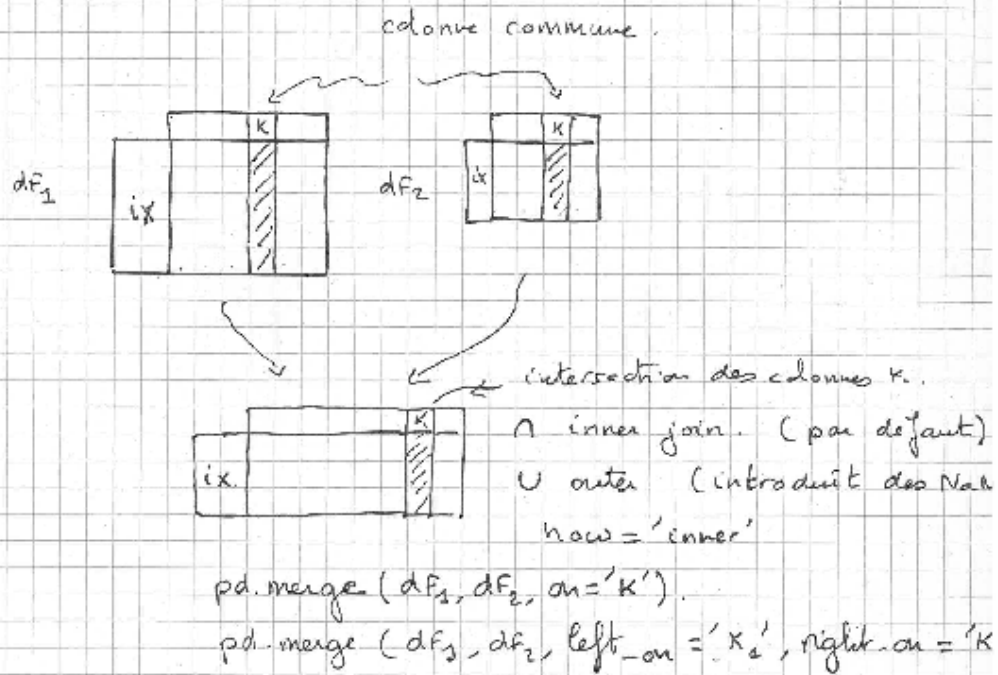
```
/mnt/e0p2/home/uguen/Bureau/TimeLine/2019/01/DE/BOURDIN-DUCHALET
```

[12]:

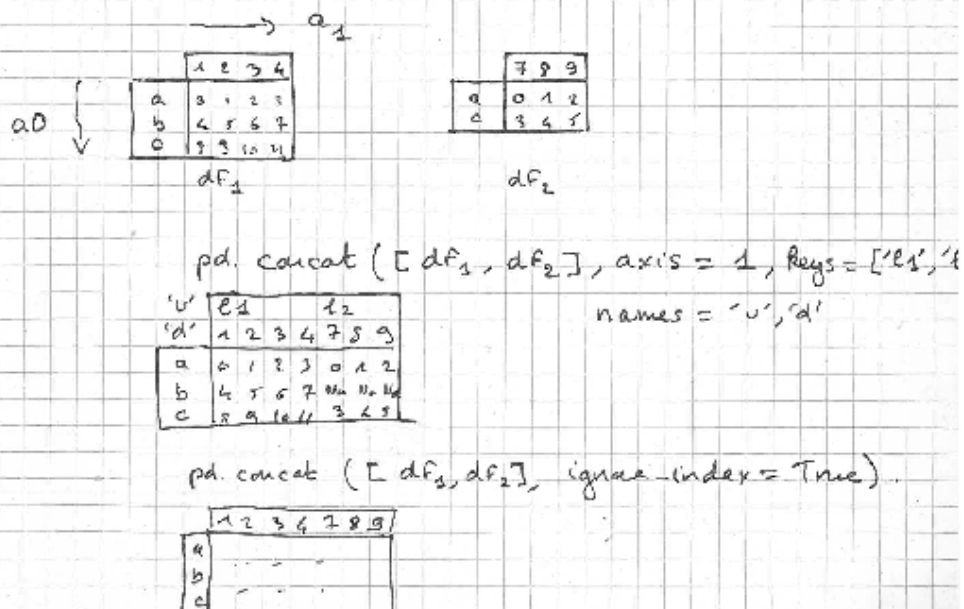
```
Image('merge.png')
```

[12]:

## Merge sur colonne



## concaténation pd.concat



## 2.3 Analyse demandées

- Construire une fonction qui évalue le taux de croissance de la population entre deux années consécutives.
- Visualiser avec la fonction `plt.scatter` la croissance ou la décroissance de la population
- Proposer des visualisations de la répartition des communes en fonction de différents critères (superficie, prénom du maire, ...)

## 2.4 Ressources

Dans la suite pour rechercher la documentation qui vous sera nécessaire pour réaliser le TP, utiliser les ressources suivantes :

[Basiques de Pandas](#)

[Tutoriel pandas](#)