

Traitement d'une base de données grâce au package Pandas sur Python

- 1) **Présentation de Pandas et de notre base de données**
- 2) **Premières manipulations : aspects globaux, regroupement par régions**
- 3) **Evolution des températures moyennes de la France métropolitaine par mois et par année**
- 4) **Evolution des températures moyennes des régions de la France métropolitaine par mois et par année**

Nghiem Dinh Gia Nghi
Célia Zaidi
Sandra Peroumal

Université de Lille - M1 ENSP 2021/2022

1) Présentation de Pandas et de notre base de données

Pandas est une bibliothèque écrite Python qui permet de structurer des données de manière rapide et intuitive puis de les analyser. La structure des données se nomme DataFrame. Ces DataFrames permettent de stocker et de manipuler les données du tableaux.

Pandas permet entre autres d'importer et d'exporter des données, de créer des objets tests, d'inspecter et de sélectionner des données, de les filtrer, nettoyer, agréger mais également de faire de l'indexing, du boolean indexing, des jointures de tables, des statistiques et des graphiques. De ce fait, Pandas ambitionne de devenir l'outil de manipulation de données open source le plus puissant et le plus flexible celui ci étant déjà disponible dans de multiples langues.

Pour utiliser le package Pandas sur Python, il faut entrer la commande : `import pandas`, au début du notebook.

Dans le cadre de ce devoir, nos manipulations reposent sur une base de données nommée *Températures quotidiennes départementales*. Celle-ci a été publiée par le ministère de l'intérieure et s'étend sur une période allant du 1 janvier 2018 au 01 novembre 2021. Elle est disponible librement sur data.gouv.fr.

2) Premières manipulations : aspects globaux, regroupement par régions

Départements dont la moyenne est supérieure ou égale à 10°C.

```
data[ data["tmoy"] >= 10.0 ].head()
```

	date_obs	code_insee_departement	departement	tmin	tmax	tmoy
1	2018-04-15	24	Dordogne	3.45	20.20	11.82
2	2018-04-15	30	Gard	7.90	18.73	13.32
3	2018-04-15	35	Ille-et-Vilaine	8.40	16.55	12.48
4	2018-04-15	37	Indre-et-Loire	7.20	18.30	12.75
5	2018-04-15	49	Maine-et-Loire	8.20	17.10	12.65

5 premiers départements
et de leurs températures moyennes.

```
data.loc[ 0 : 4, ['departement', 'tmoy']]
```

	departement	tmoy
0	Creuse	9.65
1	Dordogne	11.82
2	Gard	13.32
3	Ille-et-Vilaine	12.48
4	Indre-et-Loire	12.75

```
data[ data["tmoy"] >= 10.0 ].head()
```

Régions avec les températures les plus élevées
et les plus faibles

```
data.loc[data['tmin'].idxmin()]
```

date_obs	2018-02-28
code_insee_departement	43
departement	Haute-Loire
tmin	-16.8
tmax	2.0
tmoy	-7.4
régions	Auvergne-Rhône-Alpes
Mois_Année	2018-02
Name: 15817, dtype: object	

```
data.loc[data['tmax'].idxmax()]
```

date_obs	2019-06-28
code_insee_departement	84
departement	Vaucluse
tmin	20.13
tmax	42.77
tmoy	31.45
régions	Provence-Alpes-Côte d'Azur
Mois_Année	2019-06
Name: 109634, dtype: object	

```
data.loc[data['code_insee_departement'].isin(['2A', '2B']), 'régions']="Corse"
```

```
data.iloc[ 30 : 35, 1 : 7 ]
```

	code_insee_departement	departement	tmin	tmax	tmoy	régions
30	93	Seine-Saint-Denis	8.20	17.60	12.90	Ile-de-France
31	44	Loire-Atlantique	8.52	14.94	11.73	Pays de la Loire
32	43	Haute-Loire	4.70	14.70	9.70	Auvergne-Rhône-Alpes
33	16	Charente	8.10	18.90	13.50	Nouvelle Aquitaine
34	57	Moselle	7.60	18.50	13.05	Grand Est

Regroupement des départements en
fonction de leurs régions

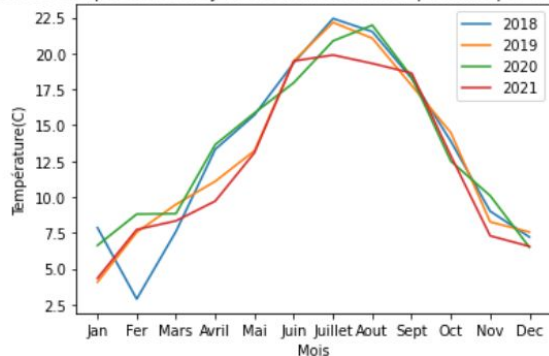
Par soucis d'espace, sur cette diapositive vous
trouverez seulement la ligne de code de la
Corse. Les 12 autres sont dans notre notebook.

3) Evolution des températures moyennes de la France métropolitaine par mois et par année

```
import matplotlib.pyplot as plt
```

```
fig,ax=plt.subplots()
ax=plt.plot(par_mois_2018['Mois'],par_mois_2018['tmoy'],label='2018')
ax=plt.plot(par_mois_2019['Mois'],par_mois_2019['tmoy'],label='2019')
ax=plt.plot(par_mois_2020['Mois'],par_mois_2020['tmoy'],label='2020')
ax=plt.plot(par_mois_2021['Mois'],par_mois_2021['tmoy'],label='2021')
ax=plt.xlabel('Mois')
ax=plt.ylabel('Température(C)')
ax=plt.title('Evolution des températures moyennes en France métropolitaine par mois et par année')
ax=plt.legend()
plt.show()
```

Evolution des températures moyennes en France métropolitaine par mois et par année



Pour produire ce graphique, nous avons fait plusieurs autres étapes que vous retrouverez dans le notebook, à savoir :

- Créer une nouvelle colonne "Mois_Année" qui reprend la colonne "Date_obs" en sélectionnant que le mois et l'année
- Regrouper les températures de toutes les régions par mois et le calcul de la température moyenne
- Regrouper les températures par année

Nous nous rendons compte que de manière générale, l'évolution des températures de la France Métropolitaine est la même sur les 4 années étudiées. A noter quelques particularités, comme de très faibles températures moyennes en février 2018, ou encore des températures moyennes moins élevées pour l'été 2021.

4) Evolution des températures moyennes des régions de la France métropolitaine par mois et par année

```
from matplotlib.pyplot import figure
```

```
ax[6,0].plot(par_mois_2018_PACA['Mois'],par_mois_2018_PACA['tmoy'],label='2018')
ax[6,0].plot(par_mois_2019_PACA['Mois'],par_mois_2019_PACA['tmoy'],label='2019')
ax[6,0].plot(par_mois_2020_PACA['Mois'],par_mois_2020_PACA['tmoy'],label='2020')
ax[6,0].plot(par_mois_2021_PACA['Mois'],par_mois_2021_PACA['tmoy'],label='2021')
ax[6,0].set_title("Provence-Alpes-Côte d'Azur")
plt.xlabel('Mois')
plt.ylabel('Température(°C)')
plt.legend()
plt.show()

fig.set_size_inches([15,30])
fig.savefig("parregion.png",dpi=300)
```

Pour produire ce graphique, nous avons fait plusieurs autres étapes que vous retrouverez dans le notebook, à savoir :

- Regrouper chaque température (tmoy, tmin et tmax) pour chacune des périodes (Mois_Année) par régions
- Répertoire les températures pour chaque région (qui sont renommées), sélectionner chaque année et renommer chaque mois
- Attribuer des coordonnées à chacune des températures moyennes des 13 régions pour les 4 années étudiées

Nous nous rendons compte que de manière générale, l'évolution des températures des 13 régions de la France métropolitaine est la même (augmentation de janvier à juillet, diminution de août à décembre). Cependant, quelques régions ont des particularités, comme par exemple la Corse en 2021 qui a eu, de façon globale, des températures plus faibles tout au long de l'année par rapport aux années précédentes.

