	<b>M1 MACIA</b> <b>Programmation de la Données :</b> <b>Projet Population</b>	
		Année 2025-26

## Sujet

On souhaite produire des études et statistiques sur la population française d'après les divers recensements faits entre 1968 et 2020 avec leur répartition géographique par villes.

## Génération de la Base données

On a les fichiers csv suivants (pour rappel, les fichiers csv sont éditables sous Libre Office Calc pour mieux voir leur contenu, mais vous ne devez pas les modifier à partir de Calc) :

- *populationSerieHistorique2020.csv* : données de population
- *populationMetaDataSerieHistorique2020.csv* : explication des champs du fichier précédent
- *populationDepartementsFrance.csv* : liste des départements et leur région par numéro

Installer les librairies du connecteur MySQL pour Python, pandas et SQLAlchemy grâce à la commande conda suivante dans la console Spyder :

***conda install mysql-connector-python sqlalchemy pandas matplotlib***

Tester le script fourni *BDDConnexion.py* avec le fichier *pokemonbis.csv*. Vérifier la table fournie et sa structure.

En vous inspirant du script exemple, créer la base de données « Population » sous MySQL issue des fichiers csv ci-dessus, éventuellement retravaillés, afin d'avoir une table de données pour population, une des villes pouvant être regroupées par département et région.


Normaliser le MCD puis modifier le MPD pour qu'il corresponde.

Pour cela, vous pourrez si nécessaire :

- Créer/modifier/supprimer les tables, voir utiliser des tables temporaires.
- Créer/modifier/supprimer les champs des tables et leurs types de données pour les simplifier et/ou les optimiser.
- Créer les relations nécessaires entre les tables pour optimiser les requêtes futures.

## Requêtes et graphiques

1. Réaliser les requêtes suivantes :
  - a. Liste des populations en 2020 avec le nom de ville, département, région.
  - b. Évolution de la population française de 1968 à 2020.
  - c. Liste des populations en 2020 par département / région avec leurs noms (2 requêtes).
  - d. Population de Paris au total et par arrondissement. Quel est le problème ? Corriger et vérifier que ce cas n'est pas produit ailleurs. Revoir la question 2.
  - e. Liste des 10 villes / départements / régions ayant cru le plus de 1968 à 2020.
  - f. Liste des 10 villes / départements où on naît / meurt le plus.
  - g. Liste des 10 villes / départements avec la plus grande/petite densité.
  - h. Comparaison pour 2020 des naissances / décès / mouvements de population par département / région (2 requêtes). (Mouvements =  $\text{deltapop}(1968/2020) - (\text{nais-deces})$ ).
  - i. Comparaison par recensement des naissances / décès / mouvements de population de la France.

	<b>M1 MACIA</b> <b>Programmation de la Données :</b> <b>Projet Population</b>	
		Année 2025-26

2. Imaginez 5 autres requêtes comprenant, sur l'ensemble, au moins une opération de regroupement, une jointure et une requête imbriquée.
3. Créer des graphiques représentant ces résultats avec la librairie Python Matplotlib.

### **À rendre par groupe de 3 personnes :**

Vous rendrez l'ensemble du projet sous forme d'un seul fichier compressé (format 7z ou zip uniquement) portant votre nom et comprenant :

- Les scripts SQL et les programmes Python permettant de générer la BDD Population, d'afficher les réponses aux questions et générant les graphiques.
- Une sauvegarde des graphiques générés au format png ou jpeg.
- Un petit rapport au format PDF de 10 pages maximum avec l'explication du script, comment l'exécuter, des imprime-écran des résultats attendus et un bilan de ce projet.