

# Activité croiser des données

## Présentation du mini projet

Je vais utiliser des **données ouvertes** (Open Data) pour **localiser** des communes qui répondent aux critères suivants :

- **communes** entre **10 000** et **15 000** habitants ;
- **température** de la région de la commune qui **ne dépasse pas 30° en été** ;
- **température** de la région de la commune qui **ne descend pas au dessous de 18°**.

Pour cela ce projet, je vais **traiter plusieurs collections de données** et **croiser les données disponibles** pour obtenir les informations voulues.

## Je récupère la collection de données sur la population

Les **données** sur le dernier recensement de la population sont disponibles sur le site de **L'INSEE** :

- je vais sur la page d'accueil de L'INSEE : <https://www.insee.fr/fr/accueil> ;
- je survole avec la souris le lien **L'INSEE et la statistique publique** puis je clique sur **Réaliser le recensement de la population** ;
- dans **Pour en savoir plus**, je clique sur **Les résultats du recensement de la population** ;
- je vais dans **Populations légales 2016 de toutes les circonscriptions administratives** et je télécharge le **fichier d'ensemble (France hors Mayotte)** au **format xls**, pour l'ouvrir et l'enregistrer ensuite dans mon dossier SNT personnel.



- **Question 1** : je prends connaissance des **métadonnées** et je note la **taille du fichier sur disque** en **octets** et en **Mo**.
- **Question 2** : Quelle **opération de calcul** a permis d'afficher la **taille en Mo** à partir de la taille en octet.

## Je prépare le fichier sur la population

- Je **cherche** parmi les différentes collections de données situées dans des onglets différentes, la **collection** qui me paraît **la plus intéressante** pour connaître la population des communes ;
- je **supprime** toutes les autres collections de données ;
- dans la collection de données que j'ai gardée, je **supprime** les lignes d'entête pour ne garder que les lignes des **descripteurs** et des **données** ;



- **Question 3** : Quels sont **descripteurs** de cette collection qui seront **utiles** à mon projet ?
- **Question 4** : **Combien** de communes sont contenues cette collections de données



?

- **j'enregistre** mon travail au **format xlsx** sous un autre nom pour l'appeler **communes.xlsx**

## Je récupère la collection de données sur les températures quotidiennes régionales

- **j'accède** au site [www.data.gouv.fr](http://www.data.gouv.fr) pour récupérer les données de températures quotidiennes régionales au format cvs.
- en **ouvrant** le fichier csv avec Excel, je constate que les caractères accentués ne sont pas correctement interprétés. Il y a un **problème d'encodage des caractères**.
- Je **crée** un nouveau classeur Excel puis dans le menu **données**, je choisis **A partir d'un fichier texte/csv** ;
- je choisis dans mon dossier SNT le fichier **temperature-quotidienne-regionale.csv** ;
- lors de **l'importation** je sélectionne comme **origine** du fichier **65001 : Unicode (UTF-8)** puis je clique sur **Charger** ;



- **J'enregistre** le fichier ensuite dans mon dossier SNT personnel au **format xlsx** sous le nom **temperature.xlsx**



- **Question 5** : je prends connaissance des **métadonnées** et je note la **taille du fichier sur disque** du fichier temperature.xlsx.
- **Question 6** : Sur quelles périodes ont été relevé ces mesures de températures ?
- **Question 7** : Comment ces données ont-elles été obtenues ?

## Je prépare le fichier sur les températures

- J'ouvre le fichier **temperature.xlsx**. Les températures sont indiquées avec des valeurs décimales mais utilisent le **point comme séparateur décimal** et non pas la **virgule**.
- J'utilise la fonctionnalité d'Excel **Remplacer** pour remplacer tous les points par des virgules pour les colonnes des températures.
- \* J'utilise la fonctionnalité d'Excel **Filtres numériques** pour trouver les régions avec les températures minimales et maximales voulues ainsi que les température de cette année 2019.



- **Question 8** : quels sont les codes des régions dont les températures en 2019 ont pu être comprises sur une journée entre 18 et 30 degrés ?

## Je croise les données



- **Question 9** : quels sont les descripteurs communs aux deux tables **communes.xlsx** et **temperatures.xlsx** ?

- En utilisant les informations des 2 tables **communes.xlsx** et **temperature.xlsx** ainsi que les fonctions de filtre et de tri d'Excel répond aux question suivantes :



- \* **Question 10** : combien de communes de 10 000 à 15 000 habitants sont dans une région où les températures en 2019 sur une journée ont été comprises entre 18 et 30 degrés ?
- **Question 11** : quelles sont les communes de la Haute-vienne qui répondent à ces critères ?

## Je continue ...



Je reviens à l'accueil SNT du thème [Les données structurées et leur traitement](#)

From:

<https://siocours.lycees.nouvelle-aquitaine.pro/> - Les cours du BTS SIO

Permanent link:

<https://siocours.lycees.nouvelle-aquitaine.pro/doku.php/snt/donnee/acsvcroiser>

Last update: **2021/01/29 12:20**

