

- Objectifs :**
- Utiliser un site de données ouvertes pour sélectionner et récupérer des données.
 - Réaliser des opérations de recherche, filtre, tri ou calcul sur une ou plusieurs tables ; observer les différences de traitement selon le logiciel choisi pour lire le fichier : tableur, programme Python.

1. Activité avec le tableur

1.1 Obtenir les données

Les données ouvertes (Open Data en anglais) sont des informations accessibles librement et gratuitement, sous la forme de fichiers respectant des formats inter-opérables.

La finalité est de donner la possibilité à tout citoyen, toute entreprise ou association d'utiliser ces données numériques à ses propres fins d'analyse pour en extraire l'information désirée.

Une partie de ces données sont publiques, par exemple le site <https://www.data.gouv.fr/fr/> contient un grand nombre de données publiques ou le site de l'Insee <https://www.insee.fr/fr/accueil>. Ces données sont librement réutilisables.

En général, le site qui fournit le fichier des données donne aussi une description des données de la table (description du fichier avec définition des descripteurs)

1. Sur le site <https://www.data.gouv.fr/fr/>, faire une recherche pour trouver les données relatives aux *temps de parole des hommes et des femmes à la télévision et à la radio*.

2. Télécharger le fichier CSV relatif aux *moyennes par années et par chaînes* **20190308-years.csv** correspondant.

3. Ouvrir le fichier **20190308-years.csv** avec Libre Office en choisissant **Jeu de caractères : Unicode : UTF-8** et **Langue : Français(Suisse)**.

Remarque : Lorsque vous ouvrez le fichier avec Libre Office, la boîte de dialogue vous montre un aperçu du fichier : vous remarquerez que ici les nombres sont écrits avec le point comme séparateur décimal. Si vous choisissez **Langue : Français(France)**, les nombres ne seront pas reconnus comme des nombres mais comme des chaînes de caractère : impossible alors de faire des calculs.

Donner une description du contenu du fichier.

1.2 Trier les données

Trier les données (voir [fichier d'aide](#)) pour répondre aux questions suivantes :

1. En quelle année le taux d'expression des femmes a-t-il été le plus important ? Préciser le nom du média et le taux.
2. En 2019, quel est le nom du média pour lequel le taux d'expression des femmes est le plus important ? Préciser le taux.
3. En quelle année, le taux d'expression des femmes a-t-il été le plus important sur un média public ? Préciser le nom du média et le taux.

1.3 Filtrer les données

Une autre façon d'obtenir des informations est de filtrer les données pour n'afficher que celles qui nous intéressent. Des outils de filtres permettent de ne faire afficher que certaines lignes d'une feuille de calcul suivant certains critères.

1. Reprendre le fichier **20190308-years.csv** et retrouver les réponses aux trois questions posées précédemment en utilisant les filtres.

2. Vous pouvez aussi choisir la fonction **Filtre Standard**. Choisir **Données** puis **Plus de filtres** puis **Filtre standard...**

3. La boîte de dialogue **Filtre standard** s'affiche et vous permet de définir un filtre standard selon des critères que vous pouvez choisir dans les menus déroulants.

Répondre à la question suivante en utilisant un filtre standard :

En 2018, quel est le nom du média pour lequel le taux d'expression des femmes est le plus important ?

1.4 Croisement des données

Lorsqu'on dispose de deux tables qui ont au moins un descripteur en commun, on peut obtenir une nouvelle information, on peut créer une nouvelle table en rajoutant un ou plusieurs descripteurs absents dans une des deux tables.

1. Prenons les deux tables suivantes :

Nom	Ville	Entrées
Musée des Beaux-Arts	Lyon	334 459
Musée d'Art Roger Quillot	Clermont-Ferrand	78 386
Musée Dauphinois	Grenoble	76 413
Musée Alpin	Chamonix	35 747
Musée d'art contemporain	Lyon	135 000

Ville	Région	Habitants
Grenoble	Isère	160 649
Lyon	Rhône	513 275
Chamonix	Haute Savoie	8 906
Clermont-Ferrand	Puy-de-Dôme	141 398

d'après Françoise Tort.

- a. Quel est le nombre d'habitants de la ville du musée Alpin ?
- b. Dans quelle région se trouve le musée qui a le plus petit nombre d'entrées ?
- c. Combien y a-t-il de musées dans la ville qui a le plus grand nombre d'habitants ?

2. Dès que les tables contiennent un nombre important de données, il est impossible de faire des recherches manuellement. Nous allons travailler sur un fichier de trois tables.

Télécharger le fichier [activité_regions_2025.ods](#) sur le site de SNT et l'ouvrir avec LibreOffice Calc.

Ce fichier a été obtenu à partir de plusieurs fichiers de données ouvertes disponibles sur le site de l'INSEE.

3. Dans la feuille **departement**, on souhaite rajouter le nom des régions, ce qui est possible puisque les deux collections départements et régions ont en commun le descripteur donnant le numéro des régions.

- a. Dans la cellule H1 de la feuille **departement**, tapez le nom du descripteur à rajouter par exemple **NomRegion**.
- b. Dans la cellule H2, taper la formule **=RECHERCHEV(B2;region.A\$2:F\$19;4;0)** (voir [fichier d'aide](#) pour comprendre la formule)
- c. Recopier la formule de la cellule **H2** vers le bas jusqu'à la dernière ligne remplie. Pour cela, copier la cellule **H2** (ctrl+c) , sélectionner la cellule **H3** et appuyer en même temps sur **Ctrl + Maj + fin** (ce qui sélectionne la plage **H3:H976**) puis coller (ctrl+v)
- d. Rajouter le nom du département dans la feuille **villes_france** en utilisant la feuille **departement** puis y rajouter le nom de la région.
- e. Utiliser un filtre pour ne voir que les villes de votre département dans la feuille "**villes_france**".
- f. Insérer une nouvelle feuille que vous appellerez "**Calvados**" ou "**Manche**" ou "**Orne**" et y copier-coller les lignes sélectionnées par le filtre.

g. Dans cette dernière feuille, calculer le nombre total d'habitants en 2010 et 2012.

h. Donner l'altitude minimale et l'altitude maximale.