SNT — Les données

Polycopié élève · parties D1 à D4

Julien Kergot

Mode d'emploi

- Le cours alterne des parties théoriques courtes et des exercices. On avance à son rythme.
- Les exercices sont numérotés de 1 à 36, comme sur le site. Le site donne aussi le code à copier et des **indices** dépliables.
- Difficulté : 🌶 prise en main · 🌶🌶 application · 🌶🌶🌶 plusieurs étapes · 🌶🌶🌶🌶 défi.
- Un fichier par exercice, nommé `NOM_Prenom_ex05.py`, rangé dans ton dossier SNT. Les réponses écrites vont en commentaire `# ...` ou dans le traitement de texte.

Partie D1 — Les données et leurs métadonnées

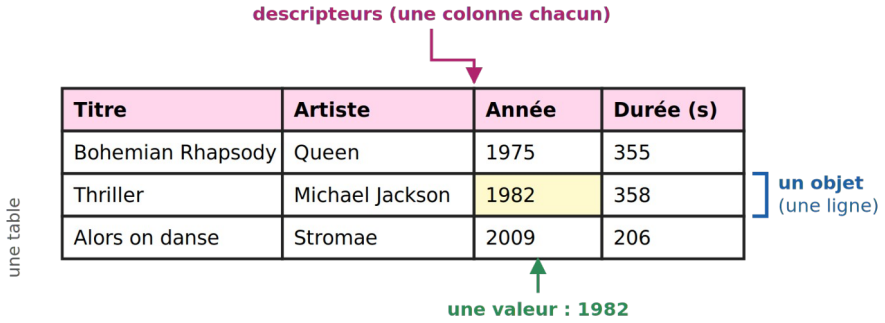
Objets, descripteurs, valeurs : à quoi ressemble une table, et que révèle-t-elle de toi ?

Ton téléphone, ton lycée, une application de météo, un réseau social : tous fabriquent et utilisent des **données**. Dans ce thème, tu vas apprendre à **lire une table de données**, à la **traiter** (avec un tableur puis avec Python), à repérer les **données personnelles** et à comprendre où sont rangés tes fichiers (le « cloud »).

Cours · Une table de données

Une **donnée** est une information représentée sous une forme qu'on peut conserver et traiter : un nombre, un texte, une date, une image… Quand on en a beaucoup, on les range dans une **table** :

- chaque **ligne** décrit un **objet** (ici, un titre musical) ;
- chaque **colonne** correspond à un **descripteur** (une caractéristique de l'objet : le titre, l'artiste, l'année…) ;
- chaque case contient une **valeur** (par exemple 1982 pour le descripteur « Année »).



Titre	Artiste	Année	Durée (s)
Bohemian Rhapsody	Queen	1975	355
Thriller	Michael Jackson	1982	358
Alors on danse	Stromae	2009	206

Une table : des objets (lignes), des descripteurs (colonnes), des valeurs (cases).

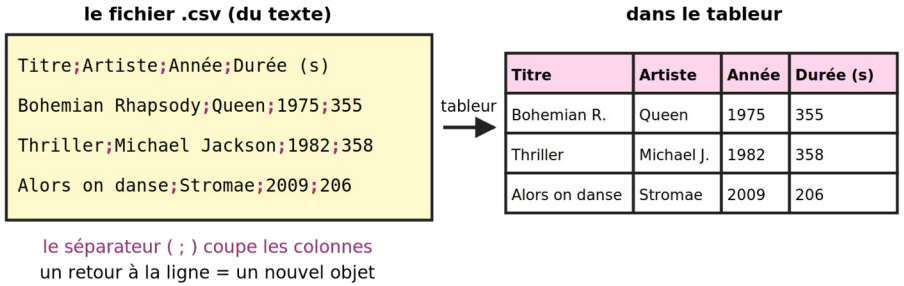
Attention

Une valeur seule ne dit rien : « 1982 » peut être une année, une durée, un prix… C'est le **descripteur** qui lui donne un sens. Et sans **unité**, un nombre est ambigu : la durée est-elle en secondes ou en minutes ?

Cours · Le format CSV

Un **fichier CSV** (*Comma-Separated Values*) est le moyen le plus simple d'échanger une table : c'est un fichier **texte**, que n'importe quel logiciel sait lire.

- une **ligne du fichier** = un objet (en général, la première ligne contient les descripteurs) ;
- un **séparateur** coupe les colonnes : le point-virgule `;` en France, la virgule `,` dans les pays anglophones ;
- le fichier ne garde **que les données** : ni formules, ni graphiques, ni couleurs. Pour les conserver, il faut enregistrer au format `.ods` ou `.xlsx`.



Le même contenu, vu dans un éditeur de texte puis dans un tableur.

Pourquoi le point-virgule en France ? Parce qu'on écrit les nombres décimaux avec une virgule (3,14) : la virgule ne peut donc pas servir de séparateur.

Attention aux accents

Si un fichier affiche Ann�e au lieu de Année, c'est un problème d'**encodage** : le logiciel ne lit pas le texte avec la même table de codes que celle utilisée pour l'écrire (voir le thème Cryptage et la table ASCII). À l'ouverture, il suffit de **choisir le bon encodage** (souvent « Europe occidentale » ou « Windows-1252 » pour des fichiers français anciens, « UTF-8 » aujourd'hui).

Cours · Les données personnelles

Une **donnée personnelle** est toute information qui se rapporte à une personne **identifiée ou identifiable**.

- **Directement** : ton nom, ta photo, ton numéro de téléphone, ton adresse électronique.
- **Indirectement** : une adresse IP, un identifiant de compte, ou même **plusieurs informations qui, croisées, ne désignent plus que toi** (classe, date de naissance, sport pratiqué…).

En Europe, le **RGPD** (règlement général sur la protection des données, 2018) encadre leur usage, et la **CNIL** veille à son application en France. Une organisation qui collecte tes données doit te dire pourquoi, n'en prendre que le nécessaire, et te laisser des droits :

- **accéder** à tes données et les **corriger** ;
- demander leur **effacement** ;
- t'**opposer** à certains usages (la publicité ciblée, par exemple).

Pour un service en ligne, l'accord d'un parent est nécessaire **avant 15 ans** en France.

À retenir

Supprimer le nom d'une table ne la rend pas **anonyme** : si les autres informations suffisent à reconnaître quelqu'un, ce sont encore des données personnelles.

Cours · Les métadonnées

Les **métadonnées** sont des **données sur les données** : elles décrivent un fichier ou une table sans en être le contenu.

- Pour un fichier : son nom, son type, sa **taille**, sa date de création ou de modification, parfois son auteur.
- Pour une photo : le modèle de l'appareil, la date et l'heure, les réglages, et parfois les **coordonnées GPS** du lieu de prise de vue (format EXIF, déjà croisé dans le thème Localisation).
- Pour une table : sa **source**, sa date de mise à jour, le territoire concerné. Ces informations sont souvent écrites sur les premières ou dernières lignes du fichier.

Attention

Les métadonnées peuvent en dire long : une photo envoyée telle quelle peut révéler **où tu habitais ou à quelle heure** tu étais quelque part. Certains services les effacent à la publication, d'autres non.

Cours · Les données ouvertes

Les **données ouvertes** (*open data*) sont des données publiques, **gratuites et réutilisables** par tous. Elles sont publiées par l'État, les collectivités, des organismes comme l'INSEE. Tu travailleras sur deux jeux de ce type : la **population française par âge** et les **victimes d'accidents corporels de la route**.

- data.gouv.fr : la plateforme française des données ouvertes : https://www.data.gouv.fr/
- INSEE : statistiques sur la population et la société : https://www.insee.fr/

- CNIL : protection des données personnelles : https://www.cnil.fr/

Exercices · Je lis une table

Exercice 1 — Objets, descripteurs, valeurs

Titre	Réalisateur	Année	Durée (min)
Le Voyage de Chihiro	Hayao Miyazaki	2001	125
Intouchables	Olivier Nakache et Éric Toledano	2011	112
Matrix	Lana et Lilly Wachowski	1999	136

1. Combien y a-t-il d'objets dans cette table ? Que représente chacun ?
2. Quels sont les descripteurs ?
3. Quelle est la valeur du descripteur « Durée » pour *Matrix* ?
4. Propose un descripteur à ajouter, et donne sa valeur pour *Intouchables*.

Réponses dans ton compte rendu (exercice 1)

Exercice 2 — Lire un CSV sans tableur

Voici le début et la fin du fichier `2023_Pyramide.csv`, ouvert dans un éditeur de texte :

2023_Pyramide.csv

```
Pyramide des ages au 1er janvier 2023, France;;;
Année de naissance;Âge révolu;Nombre d'hommes;Nombre de femmes
2022;0;345538;330929
2021;1;357351;343026
…
1923;99;3284;15028
1922 ou avant;100 ou plus;4277;25745
Total;;;
;;;
Mis à jour : janvier 2023;;;
Champ : France hors Mayotte jusqu en 2014 et France inclus
Mayotte à partir de 2014;;;
Source : Insee, estimations de population (résultats
provisoires à fin 2022);;;
```

1. Quel est le séparateur ?
2. Quelles lignes ne contiennent **pas** de données sur la population ? Que contiennent-elles ?
3. Quels sont les descripteurs ?
4. Combien d'hommes sont nés en 2021 ? Quel est leur âge révolu ?
5. Qui a produit ces données, et à quelle date ont-elles été mises à jour ?

Réponses dans ton compte rendu (exercice 2)

Exercice 3 — Des accents qui disparaissent

En ouvrant un fichier dans un éditeur de texte, une élève voit apparaître `Ann�e de naissance` à la place de `Année de naissance`.

1. Quelle est la cause du problème ?
2. Les données sont-elles perdues ? Comment les retrouver ?

Réponses dans ton compte rendu (exercice 3)

Exercice 4 — Créer mon propre fichier CSV

Avec le Bloc-notes (ou un éditeur de texte), crée un fichier `morceaux.csv` qui décrit **cinq morceaux de musique** que tu écoutes, avec quatre descripteurs : titre, artiste, année, durée en secondes. N'oublie pas la ligne des descripteurs, et sépare les colonnes par des points-virgules.

1. Ouvre ton fichier avec un tableur. Obtiens-tu bien quatre colonnes ? Sinon, que faut-il corriger ?
2. Dans le tableur, ajoute une colonne « Durée (min) » qui calcule la durée en minutes (formule `=D2/60`, à recopier vers le bas). Comment s'affichent les décimales ?
3. Enregistre sous `.ods` (ou `.xlsx`), puis rouvre le fichier `.csv` d'origine : la colonne ajoutée y est-elle ? Pourquoi ?

Fichier : NOM_Prenom_morceaux.csv

Exercices · Je protège mes données

Exercice 5 — Donnée personnelle ou non ? 🌶🌶

Pour chaque information, dis si c'est une **donnée personnelle** (directe ou indirecte) ou non, et justifie en une phrase :

1. ton prénom et ton nom ;
2. l'adresse IP de ton téléphone ;
3. la couleur de ton sac à dos ;
4. ta photo de profil ;
5. l'âge moyen des élèves d'une classe ;
6. ton identifiant de compte et son mot de passe ;
7. le nombre total de visites d'un site ce mois-ci.

Réponses dans ton compte rendu (exercice 5)

Exercice 6 — Une application un peu trop curieuse 🌶🌶

Une application de jeu demande l'accès à ta **localisation**, ton **micro**, ta **liste de contacts** et tes **photos**. Elle indique que c'est « pour améliorer ton expérience ».

1. Parmi ces accès, lesquels sont réellement nécessaires pour jouer ? Pourquoi ?
2. Que peux-tu faire avant d'accepter ?
3. Tu n'utilises plus l'application : quel droit te permet de faire supprimer ton compte et tes données ?

Réponses dans ton compte rendu (exercice 6)

Exercice 7 — Un sondage « anonyme » 🌶🌶🌶

Un sondage en ligne se dit « anonyme », mais il demande : le **code postal**, la **date de naissance** et le **sexe**. Aucun nom n'est demandé.

1. Dans ta classe, combien d'élèves ont la même date de naissance que toi ? Combien partagent en plus ton sexe ? Ces trois informations suffisent-elles à te reconnaître ?
2. Le sondage est-il vraiment anonyme ? Justifie.
3. Que peut-on changer pour que le sondage le soit davantage ?

Réponses dans ton compte rendu (exercice 7)

Exercices · Je lis les métadonnées

Exercice 8 — Les métadonnées d'un fichier 🌶

Fais un **clic droit** sur le fichier `morceaux.csv` de l'exercice 4, puis choisis **Propriétés** (ou **Lire les informations** sur Mac).

1. Note : le type de fichier, sa taille, sa date de création, sa date de dernière modification.
2. Quelles de ces informations sont des **métadonnées** ? Le contenu du fichier en est-il une ?
3. Ouvre le fichier, ajoute un morceau, enregistre : quelle information a changé ?

Réponses dans ton compte rendu (exercice 8)

Exercice 9 — Ce que dit une photo 🌶🌶

Ouvre l'outil lecteur de métadonnées de photo et clique sur **Charger la photo d'exemple** (le fichier `photo_exemple.jpg` est aussi dans l'archive Données). *Rien n'est envoyé sur Internet : le fichier est lu dans ton navigateur.*

1. Quel appareil a pris la photo ? À quelle date et à quelle heure ?
2. Que donnent les coordonnées GPS ? Retrouve le lieu avec une carte (voir le thème Localisation).
3. Si cette photo est envoyée **sans retirer ses métadonnées**, que peut apprendre une personne inconnue qui la reçoit ?
4. Défi : essaie avec une photo de ton téléphone. Contient-elle un lieu de prise de vue ?

Réponses dans ton compte rendu (exercice 9)

Exercice 10 — Les données ouvertes, qui les produit ? 🌶

Reprends l'exercice 2. Le fichier `2023_Pyramide.csv` est un jeu de **données ouvertes**.

1. Qu'est-ce que cela signifie pour toi, élève ?
2. Pourquoi est-il important que le fichier indique sa **source** et sa **date de mise à jour** ?
3. Cherche sur data.gouv.fr un jeu de données qui t'intéresse. Qui l'a publié, quand, et dans quel format ?

Réponses dans ton compte rendu (exercice 10)

Partie D2 — Traiter une table avec un tableur

Formules, références, filtres, graphiques : faire parler un fichier de données

Un fichier CSV s'ouvre dans un **tableur** (LibreOffice Calc, Excel…). Tu peux alors **chercher, filtrer, trier et calculer**, puis **représenter** les résultats par un graphique. Les fichiers de cette partie sont dans l'archive Donnees.zip ; le fichier `Eleves2022.csv` t'est donné en classe par ton professeur.

> **Un seul fichier pour toute la partie**
>
> Pour chaque fichier CSV, enregistre ton travail au format `.ods` (ou `.xlsx`) : `NOM_Prenom_donnees.ods`. Le format CSV ne garde ni les formules, ni les graphiques.

Cours · Les outils du tableur

- Une **cellule** se repère par sa colonne et sa ligne : `C3`. Une **plage** est un rectangle de cellules : `C3:C103`.
- Une **formule** commence par =. Quelques fonctions : `SOMME`, `MOYENNE`, `NB` (compte les nombres), `NB.SI` (compte les cellules qui valent une valeur donnée), `NB.SI.ENS` (même chose avec **plusieurs** critères).
- Le **filtre** n'affiche que les lignes qui répondent à une condition ; le **tri** les range dans un ordre.

Fonction	Exemple	Effet
SOMME	`=SOMME(C3:C103)`	additionne les valeurs de la plage
NB.SI	`=NB.SI(F2:F933;"ITALIEN")`	compte les cellules de la plage qui valent « ITALIEN »
NB.SI.ENS	`=NB.SI.ENS(A:A;"F";B:B;20)`	compte les lignes où A vaut « F » **et** B vaut 20

Cours · Recopier une formule : références relatives et absolues

Quand on **recopie** une formule vers le bas ou vers la droite, le tableur **décale** les références. C'est pratique… sauf quand on veut qu'une référence **ne bouge pas**. Le symbole $ sert alors à **figer** la colonne ou la ligne.

Écriture	Ce qui est figé	Recopiée d'une case vers la droite…
`C3`	rien (référence relative)	devient `D3`
`$E3`	la colonne E	reste `$E3` (la ligne, elle, change si on descend)
`E$3`	la ligne 3	devient `F$3`
`$E$3`	la colonne et la ligne	reste `$E$3`

> **Attention**
>
> Un tableur **ne prévient pas** quand une formule recopiée donne un résultat absurde. Vérifie toujours **l'ordre de grandeur** : ce résultat est-il raisonnable ? Et quand tu comptes avec `NB.SI`, les lignes **masquées par un filtre** sont comptées quand même.

Exercices · Population française : formules et références

Exercice 11 — Ouvrir la pyramide des âges 🌶️

Ouvre le fichier `2023_Pyramide.csv` dans un tableur. S'il te demande des options d'importation, choisis le **point-virgule** comme séparateur et l'encodage **Europe occidentale** (ou Windows-1252). Tu peux aussi regarder le fichier avec l'explorateur de tables.

1. Combien d'objets (lignes de population) contient la table ? Quels sont les descripteurs ?

2. Dans la cellule E3, fais calculer le nombre d'enfants nés en 2022 avec la formule `=SOMME(C3:D3)`. Quel est ce nombre ?
3. Recopie la formule de E3 dans E4. Quelle formule lis-tu dans E4 ? Que s'est-il passé ?
4. Recopie la formule jusqu'à la ligne 103. Combien de personnes ont 15 ans révolus (nées en 2007) ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 12 — Population totale 🌶🌶

1. Dans la cellule C104 (ligne « Total »), fais calculer le nombre total d'hommes en France avec une formule SOMME. Quelle formule as-tu écrite ? Quel est le résultat ?
2. Recopie cette formule dans D104. Quelle formule lis-tu ? Que s'est-il passé ?
3. Fais calculer la population totale. Y a-t-il plus d'hommes ou de femmes en France au 1er janvier 2023 ? Quel est l'écart ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 13 — Une proportion, et le piège du recopiage 🌶🌶

1. Dans F3, écris `=C3/E3` pour calculer la proportion de garçons parmi les enfants nés en 2022. Affiche le résultat en pourcentage avec une décimale. Que trouves-tu ?
2. Recopie la formule de F3 dans G3 pour avoir la proportion de filles. Quelle formule lis-tu dans G3 ? Le résultat est-il raisonnable ?
3. Corrige : écris dans F3 la formule `=C3/$E3`, puis recopie-la dans G3. Quelle formule lis-tu maintenant ? À quoi sert le $?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 14 — Autant d'hommes que de femmes ? 🌶🌶

1. Recopie les formules de F3:G3 jusqu'à la ligne 103, puis jusqu'à la ligne 104 (Total). Y a-t-il autant de garçons que de filles à la naissance ? Et autant d'hommes que de femmes dans toute la population ?
2. Existe-t-il un âge où il y a **exactement** autant d'hommes que de femmes ? Entre quels âges la situation s'inverse-t-elle ?
3. Sélectionne F3:G103 et insère un graphique qui montre la proportion d'hommes et de femmes **en fonction de l'âge**. Quel type de graphique choisis-tu, et pourquoi ?
4. Décris l'évolution observée. Propose une explication.

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 15 — Dessiner la pyramide des âges 🌶🌶🌶

Une **pyramide des âges** place les hommes à gauche, les femmes à droite, et l'âge en ordonnée. Le tableur ne sait tracer que des barres vers la droite : on fabrique donc les hommes **en valeurs négatives**.

1. Dans la colonne H, fais calculer l'**opposé** du nombre d'hommes (`=-C3`), puis recopie vers le bas jusqu'à la ligne 103.
2. Sélectionne l'âge (colonne B) et les colonnes H et D, puis insère un graphique en **barres horizontales empilées**. Mets des titres et une légende.
3. Sur l'axe horizontal, les nombres d'hommes s'affichent avec un signe moins. Comment les afficher sans ce signe ?
4. Décris la forme obtenue : à quel âge la population est-elle la plus nombreuse ? Que dire de la base de la pyramide ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercices · Un lycée : filtrer, compter, identifier

> **Un fichier fourni en classe**
>
> `Eleves2022.csv` décrit les élèves d'un vrai lycée général (sans les noms). Il n'est pas publié sur ce site, justement parce qu'il contient des données personnelles : tu verras plus loin que **retirer les noms ne suffit pas à protéger les élèves**. Chaque ligne est un élève ; les classes 201 à 212 sont des secondes, 1G1 à 1G8 des premières, TG1 à TG7 des terminales.

Exercice 16 — Ouvrir, filtrer, compter 🌶

1. Ouvre `Eleves2022.csv` dans le tableur. Quels sont les descripteurs ? Combien de lignes (d'élèves) contient-il ?
2. La colonne **Sortie** contient des dates pour certains élèves. Que signifie-t-elle ? Combien d'élèves y a-t-il **réellement** dans le lycée ?

3. Avec le filtre automatique, affiche uniquement les élèves de terminale qui ont l'**italien en LVB**. Combien sont-ils ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 17 — Les spécialités en première 🌶🌶

Les élèves de première sont aux lignes **434 à 697** du fichier. Leurs choix de spécialités sont dans les colonnes G à K (Option1 à Option5).

Avec la fonction `NB.SI` sur la plage `G434:K697`, compte les élèves de première pour chacune de ces dix spécialités : `MATHEMATIQUES`, `PHYSIQUE-CHIMIE`, `SC. ECONO.& SOCIALES`, `SCIENCES VIE & TERRE`, `HIST.GEO.GEOPOL.S.P.`, `LLC ANGL.MOND.CONT.`, `HUMAN.LITTER.PHILO.`, `SCIENCES INGENIEUR`, `LITT. ANGLAIS`, `NUMERIQUE SC.INFORM.`.

1. Écris la formule pour `MATHEMATIQUES`. Écris les dix résultats dans un tableau.
2. Représente ces résultats par un diagramme en barres, classé de la plus choisie à la moins choisie. Pourquoi un diagramme en barres convient-il ici ?
3. Un élève choisit trois spécialités. Pourquoi la somme des dix nombres n'est-elle pas égale au nombre d'élèves ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 18 — Les options en terminale 🌶🌶

Les élèves de terminale sont aux lignes **698 à 933**. Trois options sont proposées : `MATHS EXPERTES`, `MATHS COMPLEMENTAIRE` et `DRT&ENJX.MDE CONTEMP` (droit et grands enjeux du monde contemporain).

1. Compte les élèves de terminale pour chacune des trois options. Combien d'élèves n'ont aucune de ces trois options (en supposant qu'un élève n'en suit qu'une) ?
2. Représente la répartition par un diagramme en **secteurs**, en n'oubliant pas les élèves sans option.
3. Pourquoi le diagramme en secteurs convient-il ici, alors que ce n'était pas le cas pour les spécialités ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 19 — Anonyme ? Pas si vite 🌶🌶🌶

Dans le fichier, il n'y a **aucun nom**. Vérifions s'il est vraiment anonyme.

1. Avec la recherche du tableur (**Ctrl + F**, puis « Rechercher tout »), cherche `MUSIQUE`. Combien de cellules contiennent ce mot ? Regarde les lignes trouvées : combien d'élèves de **première** ?
2. Note les informations visibles sur la ligne de cet élève de première (classe, sexe, régime, LVB, options). Un élève de la classe concernée saurait-il de qui il s'agit ?
3. Le fichier est-il vraiment « anonyme » ? Que pourrait-on faire pour mieux protéger les élèves avant de le publier ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercices · Sécurité routière : compter, comparer, se méfier

Les données

`Securite_routiere1.csv` : âge des victimes d'accidents corporels de la route qui ne sont **pas sorties indemnes** (tuées ou blessées), pour les années **2015 à 2021** (une colonne par année). `Securite_routiere2.csv` : **sexe** (F ou G) et **âge** des victimes de l'année **2021**. Ces fichiers sont extraits du fichier national des accidents corporels (BAAC) : ils sont volumineux, tu peux donc t'en servir comme test de ton tableur.

Exercice 20 — Combien de victimes par an ? 🌶🌶

Ouvre `Securite_routiere1.csv`. Les années sont sur la ligne 3, les âges commencent à la ligne 4. Chaque colonne contient une année.

1. Avec la fonction NB (qui compte les nombres) sur la plage de la colonne A (`A4:A79732`), calcule le nombre de victimes en 2015. Recopie sur les autres colonnes. Écris les sept résultats.
2. Représente-les par un diagramme en colonnes. Que remarques-tu pour 2020 ?
3. Que donnerait `=NB(A:A)` ? Pourquoi le résultat serait-il faux de 1 ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 21 — Les accidents, âge par âge 🌶🌶🌶

On veut le nombre de victimes pour chaque âge et chaque année.

1. Dans la colonne `I`, écris « Âges » en `I3`, `0` en `I4`, puis une formule en `I5` qui, recopiée vers le bas, donne les âges suivants jusqu'à 110. Les années 2015 à 2021 sont écrites en `J3:P3`.
2. Dans `J4`, écris `=NB.SI(A:A;I4)`. Recopie-la en `J5` : que se passe-t-il ? Recopie-la en `K4` : est-ce correct ?
3. Corrige la formule de `J4` avec un `$` bien placé, puis recopie-la sur `J4:P114`. Quelle formule as-tu écrite ?
4. Combien de bébés (0 an), puis d'enfants de 1 an ont été victimes en 2015 ?
5. Quel est l'âge de la **plus âgée** victime du fichier, et en quelle année ?
6. Trace une **courbe** pour 2021 (colonne P). À quel âge y a-t-il le plus de victimes ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Exercice 22 — Hommes et femmes victimes 🌶🌶🌶

Ouvre `Securite_routiere2.csv` (colonne A : sexe, colonne B : âge ; les données commencent à la ligne 3).

1. Avec `NB.SI`, compte les victimes de sexe `G` puis de sexe `F`. Quel résultat obtiens-tu ?
2. Avec `NB.SI.ENS`, compte les garçons et les filles victimes à 20 ans, puis à 40 ans. Quelle formule as-tu écrite pour les garçons de 20 ans ?
3. Un camarade conclut : « Les hommes conduisent donc deux fois plus mal que les femmes. » Que réponds-tu ? De quelles informations aurait-on besoin pour comparer honnêtement ?

Fichier : NOM_Prenom_donnees.ods (un seul fichier pour toute la partie D2)

Partie D3 — Traiter une table avec Python

Lire un fichier CSV, compter, filtrer, croiser deux tables : le tableur ne fait pas tout

Un tableur est pratique pour explorer. Mais pour **répéter** un traitement, traiter un **très gros fichier** ou **croiser** des tables, un programme est plus sûr et plus rapide. Dans cette partie, tu vas refaire en Python une partie de D2, puis aller plus loin.

> **Avant de commencer**
>
> Mets dans **un même dossier** ton programme `.py` et les fichiers `2023_Pyramide.csv`, `Securite_routiere2.csv` (dans Donnees.zip) et `Eleves2022.csv` (donné en classe). Chaque exercice est un fichier `NOM_Prenom_ex05.py`. Tu auras besoin des boucles, des conditions et des fonctions du cours Python.

Cours · Lire un fichier dans un programme

Un fichier CSV est du **texte** : Python le lit ligne par ligne. Chaque ligne est un texte, qu'on **découpe** avec `.split(";")` pour obtenir ses colonnes. On range tout dans une **liste de listes** : la **table**.

lire_table.py

```
def lire_table(nom):
    """Lit un fichier CSV et renvoie la table : une liste de lignes,
    chaque ligne étant une liste de textes (un texte par colonne)."""
    table = []
    with open(nom, encoding="cp1252") as fichier:
        for ligne in fichier:
            table.append(ligne.strip().split(";"))
    return table
```

- `open(nom, encoding="cp1252")` ouvre le fichier ; `with … as fichier` le referme automatiquement. L'encodage `cp1252` lit les accents des fichiers français de ce thème.
- `.strip()` enlève le retour à la ligne de fin ; `.split(";")` coupe le texte à chaque point-virgule.
- Le résultat est une **liste** : `table[0]` est la première ligne, `table[2]` la troisième ; `table[2][3]` est la quatrième colonne de la troisième ligne.

essai.py

```
table = lire_table("2023_Pyramide.csv")
print(len(table))
print(table[1])
print(table[2])
```

console

```
108
['Année de naissance', 'Âge révolu', "Nombre d'hommes",
'Nombre de femmes']
['2022', '0', '345538', '330929']
```

> **Attention**
>
> Dans la liste, **tout est du texte** : `'345538'` n'est pas le nombre 345538. Pour calculer, il faut convertir avec `int(...)` (nombre entier). Et les indices **commencent à 0** : la colonne « Nombre d'hommes » est `ligne[2]`.

Cours · Boîte à outils

Besoin	En Python	Exemple
Convertir un texte en nombre	`int(texte)`	`int(ligne[2])`
Une partie de la liste (une **tranche**)	`liste[debut:fin]` (la fin est exclue)	`table[2:103]` : lignes d'indices 2 à 102
Tester l'appartenance	`valeur in liste`	`"MATHEMATIQUES" in ligne[6:11]`
Compter : un **compteur**	`n = 0` puis `n = n + 1` dans la boucle	`if ligne[1] == "F": n = n + 1`
Compter par catégorie : un **dictionnaire**	`dico[cle]` (voir ci-dessous)	`effectifs[age] = effectifs[age] + 1`

dictionnaire.py

```
effectifs = {}
for age in [20, 18, 20, 21, 20]:
    if age in effectifs:
        effectifs[age] = effectifs[age] + 1
    else:
        effectifs[age] = 1
print(effectifs)
```

console

```
{20: 3, 18: 1, 21: 1}
```

Un **dictionnaire** associe une **clé** (ici l'âge) à une **valeur** (ici l'effectif) : c'est un compteur par catégorie, sans avoir à connaître les catégories à l'avance.

Exercices · La population en Python

Exercice 23 — Lire la pyramide des âges

Dans un fichier, recopie la fonction `lire_table` du cours, puis ajoute :

ex01.py

```
table = lire_table("2023_Pyramide.csv")
print(len(table))
print(table[1])
print(table[2])
```

1. Exécute le programme. Que représente le premier nombre affiché ?
2. Quel est le **type** de `table[2][2]` ? Vérifie avec `print(type(table[2][2]))`.
3. Que faut-il faire pour utiliser `table[2][2]` dans un calcul ? Affiche `int(table[2][2]) * 2`.

Fichier : NOM_Prenom_ex23.py

Exercice 24 — Les totaux de la population

Avec une boucle sur `table[2:103]` (les 101 lignes de population), calcule :

1. le nombre total d'**hommes** et le nombre total de **femmes** ;
2. la proportion de femmes dans la population, en pourcentage arrondi à 0,1.

Fichier : NOM_Prenom_ex24.py

Exercice 25 — L'âge moyen des Français 🌶🌶🌶

L'**âge moyen** d'une population se calcule ainsi : pour chaque âge, on multiplie l'âge par le nombre de personnes de cet âge ; on additionne tous ces produits ; on divise par la population totale.

1. Écris un programme qui calcule l'âge moyen. (Attention à la ligne « 100 ou plus » : on prendra 100 ans.)
2. Écris la **fonction** `age_moyen(table)` qui renvoie ce nombre, puis affiche `round(age_moyen(table), 1)`.

Fichier : NOM_Prenom_ex25.py

Exercices · Le lycée en Python

Exercice 26 — Retrouver 916 et 39 🌶

Lis `Eleves2022.csv` avec `lire_table`. La première ligne contient les descripteurs : on la saute avec `table[1:]`.

1. Combien de lignes contient la table ? Combien d'élèves sont **encore présents** (colonne `Sortie` vide, soit `ligne[2] == ""`) ?
2. Combien d'élèves de **terminale** (classe qui commence par T : `ligne[0][0] == "T"`) ont l'italien en LVB (`ligne[5] == "ITALIEN"`) ? Compare avec ton résultat du tableur.

Fichier : NOM_Prenom_ex26.py

Exercice 27 — Une fonction qui compte 🌶🌶

1. Écris une fonction `compter(table, colonne, valeur)` qui renvoie le nombre d'élèves **présents** dont la colonne `colonne` vaut `valeur`.
2. Utilise-la pour compter les élèves présents par LVB (colonne 5) : `"ESPAGNOL"`, `"ALLEMAND"`, `"ITALIEN"`. Vérifie que la somme est 916.

Fichier : NOM_Prenom_ex27.py

Exercice 28 — Les filles et les maths : un tableur peine, Python non 🌶🌶🌶

Dans une table d'élèves, les options sont dans cinq colonnes (`ligne[6:11]`). Un tableur a du mal à compter « les élèves qui ont `MATHEMATIQUES` dans **l'une** des cinq colonnes ». Python, lui, sait : `"MATHEMATIQUES" in ligne[6:11]`.

1. Écris une fonction `part_filles(eleves, option)` qui, pour les élèves **présents** de **première et de terminale** (classe qui commence par `1` ou `T`) suivant cette option, renvoie le nombre d'élèves et le pourcentage de filles (`ligne[1] == "F"`).
2. Compare le pourcentage de filles parmi les élèves de première et de terminale, avec celui dans la spécialité `MATHEMATIQUES`, puis dans `MATHS EXPERTES`.
3. Teste aussi `NUMERIQUE SC.INFORM.`, `HUMAN.LITTER.PHILO.` et `PHYSIQUE-CHIMIE`. Qu'observes-tu ? Que peut-on en conclure ?

Fichier : NOM_Prenom_ex28.py

Exercices · Les accidents : un gros fichier

Exercice 29 — Garçons, filles, âges 🌶🌶

`Securite_routiere2.csv` contient 73 985 victimes de 2021 : une ligne = un sexe et un âge. Le fichier a deux lignes de titre : on les saute avec `table[2:]`.

1. Combien de victimes de sexe G et de sexe F ? Compare avec ton résultat du tableur.
2. Quel est l'**âge moyen** des victimes garçons, puis filles ?
3. Quel est l'âge de la **plus âgée** victime du fichier ?

Fichier : NOM_Prenom_ex29.py

Exercice 30 — Un histogramme en texte 🌶🌶🌶

Avec un **dictionnaire** `effectifs`, compte les victimes de chaque âge. Puis affiche, **par tranche de 10 ans** (0 à 9 ans, 10 à 19 ans…), le nombre de victimes et une barre de # : un # par tranche de 500 victimes.

1. Quel âge compte le plus de victimes ?

2. Quelle tranche d'âge compte le plus de victimes ? Que penses-tu de la forme obtenue ?

Fichier : NOM_Prenom_ex30.py

Cours · Croiser deux tables

Deux tables peuvent se **croiser** quand elles ont un **descripteur commun**. Ici : l'**âge**. La pyramide des âges dit **combien de personnes** ont chaque âge ; le fichier des accidents dit **combien de victimes** ont chaque âge. En divisant, on obtient un **taux** : le nombre de victimes **pour 100 000 personnes** de cet âge. Un taux permet de **comparer** des groupes de tailles différentes.

> **Attention**
>
> Les victimes sont celles de **2021** ; la population est celle du **1er janvier 2023**. Ce n'est pas la même année : les taux sont des **ordres de grandeur**, pas des mesures exactes.

Exercice 31 — Le taux de victimes par âge 🌶🌶🌶🌶

Construis, à partir de la pyramide, un dictionnaire `population` : la clé est l'âge, la valeur est le nombre de personnes (hommes + femmes). Avec `effectifs` (exercice 30), calcule pour chaque âge de 0 à 99 ans :

```
taux = 100000 * effectifs[age] / population[age]
```

1. À quel âge le taux est-il le plus élevé ? Quelle est sa valeur (arrondie à l'unité) ?
2. Affiche le taux à 10, 18, 30, 50, 70 et 90 ans. Que constates-tu ?
3. Compare le **nombre** de victimes à 70 ans et à 90 ans, puis le **taux** à ces âges. Pourquoi le taux est-il plus juste pour comparer ?
4. Calcule le taux global des **garçons** (victimes garçons ÷ population masculine × 100 000) et celui des **filles**. Quel est le rapport entre les deux ? Peut-on parler de « surmortalité masculine » ?

Fichier : NOM_Prenom_ex31.py

Exercices · Bilan

Exercice 32 — Tableur ou Python ? 🌶🌶

Pour chaque situation, indique l'outil que tu choisirais (tableur ou Python) et pourquoi :

1. Regarder rapidement un petit fichier et en faire un graphique.
2. Refaire chaque année le même calcul sur un nouveau fichier de 500 000 lignes.
3. Compter les élèves qui ont une option dans l'une de cinq colonnes.
4. Montrer à un camarade, sans programmation, comment tu es arrivé à un résultat.

Réponses dans ton compte rendu (exercice 32)

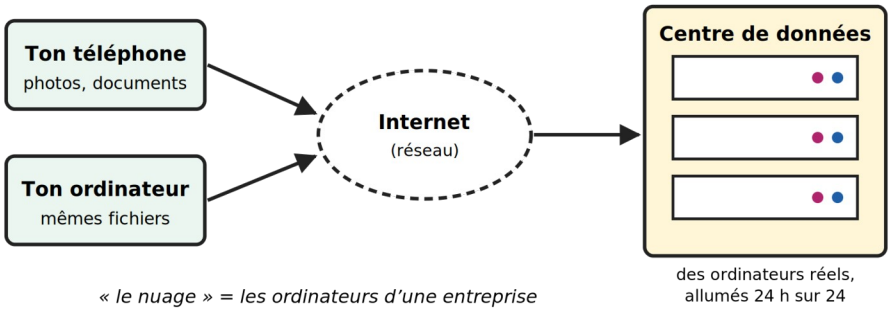
Partie D4 — Le cloud : stocker ailleurs

Où sont vraiment tes fichiers, que devient ton travail quand tu le « synchronises », et que coûte ce stockage ?

Tes photos, tes documents, tes messages ne sont plus seulement sur ton téléphone ou ton ordinateur : beaucoup sont rangés « dans le **cloud** » (le **nuage**). Cette partie est **courte** : elle peut se faire en fin de séance 3 ou en travail personnel.

Cours · Le nuage, ce sont des ordinateurs

Le **cloud** (ou informatique en nuage) désigne le fait de **stocker des données ou de faire des calculs sur des ordinateurs distants**, accessibles par Internet. Ces ordinateurs sont des **serveurs**, rassemblés par milliers dans des bâtiments appelés **centres de données** (*data centers*).



Le cloud : des fichiers rangés sur des serveurs, accessibles depuis plusieurs appareils.

- **Avantages** : on retrouve ses fichiers depuis n'importe quel appareil, on partage facilement, on dispose d'une copie si l'appareil est perdu.
- **Inconvénients** : il faut une connexion Internet, on dépend d'une entreprise, et les données sont **hébergées dans un pays** dont les lois s'appliquent (le RGPD en Europe).

Cours · Synchroniser n'est pas sauvegarder

La **synchronisation** maintient **les mêmes fichiers** sur plusieurs appareils : quand tu modifies un fichier sur l'un, la modification apparaît sur les autres.

Attention

Une suppression se synchronise aussi. Si tu effaces un fichier dans un dossier synchronisé, il disparaît partout. La synchronisation n'est donc **pas** une sauvegarde. Une **sauvegarde** est une copie **indépendante**, qui survit à une erreur, à une panne ou à une attaque.

La règle du 3-2-1

Pour des données importantes : **3** copies, sur **2** types de supports différents, dont **1** copie hors de chez soi (dans le cloud, par exemple).

Cours · Le stockage a un coût énergétique

Un centre de données fonctionne **jour et nuit** : il consomme de l'électricité pour faire tourner les serveurs et pour les **refroidir**. Plus on stocke, envoie et regarde de données, plus il en faut. Mais le numérique, ce sont aussi les **appareils** : selon les études de l'ADEME et de l'Arcep, leur **fabrication** pèse très lourd dans l'empreinte du numérique. (Ton professeur peut te renvoyer vers la séquence « Numérique, IA & écologie » pour aller plus loin.)

- On exprime l'énergie en **kilowattheures** (kWh) : un appareil de **1 kW** qui fonctionne **1 heure** consomme **1 kWh**.
- Pour un appareil de puissance P en kW utilisé pendant `t` heures : énergie = `P × t`.
- **Unités de stockage** : 1 ko = 1 000 octets ; 1 Mo = 1 000 ko ; 1 Go = 1 000 Mo ; 1 To = 1 000 Go.

Exercices · Je réfléchis au cloud

Exercice 33 — Combien de place ? 🌶

Une photo prise avec un téléphone pèse environ **4 Mo**. Un compte gratuit offre **15 Go** d'espace.

1. Combien de photos peut-on ranger dans 15 Go ?
2. Une minute de vidéo pèse environ **100 Mo** (hypothèse de l'exercice). Combien de minutes de vidéo peut-on ranger dans 15 Go ?
3. Combien de Mo de photos prends-tu par mois, à peu près ? En combien de mois ton espace gratuit serait-il plein ?

Réponses dans ton compte rendu (exercice 33)

Exercice 34 — Synchroniser, oui ; sauvegarder, non 🌶🌶

Léa synchronise son dossier « Devoirs » entre son ordinateur et son téléphone grâce à un service de cloud. Elle supprime par erreur un devoir sur son ordinateur.

1. Le devoir existe-t-il encore sur son téléphone ? Pourquoi ?
2. Léa pense : « Mes fichiers sont dans le cloud, donc ils sont sauvegardés. » Explique pourquoi elle se trompe.
3. Applique la règle du 3-2-1 : propose une organisation pour protéger les photos de ta famille.

Réponses dans ton compte rendu (exercice 34)

Exercice 35 — Que consomme un serveur ? 🌶️🌶️🌶️

Hypothèses de l'exercice (ce ne sont pas des mesures) : un serveur a une puissance de **300 W** (0,3 kW) et fonctionne **24 h sur 24**, toute l'année. Un centre de données en contient **10 000**.

1. Calcule l'énergie consommée par **un serveur en une journée**, en kWh.
2. Calcule l'énergie consommée par **un serveur en un an** (365 jours).
3. Calcule l'énergie consommée par les **10 000 serveurs en un an**, en kWh puis en GWh (1 GWh = 1 000 000 kWh).
4. Les centres de données doivent aussi être **refroidis**. Cette consommation s'ajoute-t-elle ou se déduit-elle des résultats précédents ? Que cela change-t-il au résultat ?

Réponses dans ton compte rendu (exercice 35)

Exercice 36 — Où ranger quoi ? 🌶️🌶️

Pour chaque type de fichier, dis **où** tu le rangerais (uniquement sur ton appareil, dans le cloud, ou les deux) et **pourquoi** :

1. tes devoirs en cours ;
2. un scan de ta carte d'identité ;
3. les photos d'un voyage en famille ;
4. un journal intime.

Réponses dans ton compte rendu (exercice 36)