

Analyse de données avec un tableur

1 Introduction

Dans ce TP, l'objectif est de vous faire découvrir des possibilités de traitement automatique de grandes « bases de données » dans un tableur. De nouveau, toutes ces fonctionnalités se retrouvent à la fois sous Excel et sous LibreOffice avec de grandes similarités entre les interfaces. À l'issue de ce TP vous devrez être en mesure de savoir faire :

- des filtres de sélection, ainsi que des calculs sur des sélections (formule `SOMME.SI`)
- des tableaux croisés dynamiques
- des graphiques dynamiques

Les fichiers à utiliser (à récupérer sur le site du cours) :

- `Enquete.csv` : données des réponses à un questionnaire sur vos usages des ordinateurs (menée auprès d'étudiants des années passées). Pour interpréter ces séries de chiffres, vous pouvez consulter le questionnaire correspondant aux réponses dans le fichier `QuestionnaireUsage.pdf` (**les numéros saisis correspondent aux numéros à côté des cases dans le questionnaire**).
- `ProductionCoop.xls` : données (imaginaires) de vente de légumes par une coopérative de trois agriculteurs

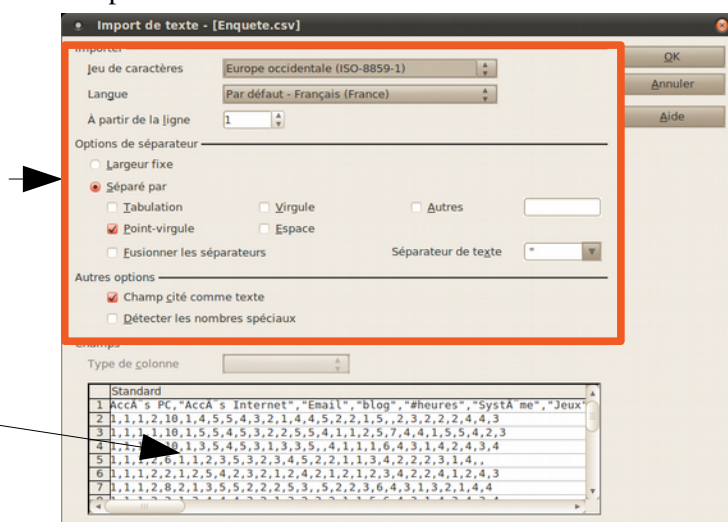
2 Filtrer et parcourir des données

A Importation et exportation de données en format textuel

Le format de fichier de données CSV (*Comma Separate Values*) est le plus simple des formats de données « tabulées » (*i.e.* sous forme de table) et donc généralement le plus fiable entre deux ordinateurs. Il retient les données dans un fichier texte sans aucun formatage ... uniquement les données brutes.

- Commencer par lancer LibreOffice
- Ouvrir maintenant le fichier depuis votre tableur (**Fichier>Ouvrir** depuis LibreOfficeCalc), une fenêtre apparaît pour définir la manière d'importer les données. La fenêtre fait apparaître en bas une prévisualisation du tableau de données

Configuration de la lecture du fichier CSV



Prévisualisation du tableau

Lorsque vous modifiez les configurations, la prévisualisation du bas de l'interface illustre la feuille de tableur qui sera construite après validation. Ceci permet de vérifier que les options correspondent à vos données. Vous utiliserez :

- ▶ Jeu de caractères : UTF8 (c'est celui qui va permettre d'avoir des accents corrects dans les entêtes)
- ▶ Langue : vous utiliserez l'anglais ou l'américain. La raison de ce choix est l'utilisation du « . » comme séparateur décimal dans le document (convention anglaise), pour que les cellules soient comprises comme un chiffre, il faut que la convention soit correcte, sinon votre cellule contiendra du texte (on le repère lorsque qu'une valeur telle que « 3.5 » est alignée à droite alors que les nombres sont alignés à gauche par défaut).
- ▶ Pour le séparateur de colonnes, modifier les choix des séparateurs pour indiquer comment décomposer les lignes en plusieurs colonnes contenant des chiffres. Utiliser la visualisation pour savoir comme sera interprété le fichier : **il faut que vous retrouviez bien toutes les colonnes de chiffres bien séparées.**

Q.1) Quel séparateur faut-il utiliser pour ouvrir correctement le fichier (c'est-à-dire que les données soient dans des colonnes bien séparées) ?

-
- ▶ Une fois que la configuration du format CSV est correcte, valider l'importation,
 - ▶ Finalement, enregistrer ce fichier au format .ods (utiliser un nom de fichier explicite). Ce fichier sera utilisé pour la suite du TP.

Ces opérations nécessitent parfois des ajustements délicats. Si besoin, un fichier *Enquete.ods* est fourni avec un contenu correct.

B Tri et filtrage de données

Les fonctionnalités de tri et de filtrage sont accessibles dans le menu **Données**. Comme dans toute la suite, elles s'appliquent sur des tableaux dont l'organisation des lignes et des colonnes est celle-ci :

- ▶ Les lignes sont des « individus » (à classer, à filtrer) (ici, les étudiants ayant répondu au questionnaire)
- ▶ Les colonnes représentent des attributs qui décrivent ces individus (ici, les réponses aux questions)
- ▶ La première ligne contient les noms des attributs (entête de colonne)

Pour utiliser les fonctions de tri, d'auto-filtrage et de filtrage, il faut bien s'assurer que toutes les données sont sélectionnées : Ne faites pas la sélection vous-même : placer votre curseur au milieu du tableau et sélectionner la fonctionnalités souhaitée, LibreOffice trouvera la sélection adéquate.

- ▶ Trier le tableau par le nombre croissant d'années d'utilisation d'un ordinateur.

Q.2) Quel est le(s) type(s) de B2I acquis par l'étudiant ayant pratiqué le moins d'année ?

-
- ▶ Trier le tableau suivant le temps passé par semaine (#heure) et le niveau auto-estimé (Niveau estimé).

Q.3) Quel(s) système(s) d'exploitation(s) (Windows, Mac ou Linux) utilise l'étudiant qui passe le plus de temps sur ordinateur par semaine et à l'auto-estimation de son niveau le plus élevé.

.....

.....

Le filtrage des données permet d'afficher les lignes (*i.e.* les individus) qui correspondent à une condition. L'auto-filtre est une fonctionnalité qui permet d'exprimer facilement des filtres simples et de les modifier. L'auto-filtre ajoute des menus déroulants dans les entêtes de colonnes. Ces menus permettent de sélectionner les valeurs à faire afficher. Si on affiche **Tout**, alors le filtre sur la colonne est désactivé. Vous pouvez combiner les filtres sur plusieurs colonnes.

Pour désactiver les auto-filtres, il faut vous placer dans la plage de données et re-cliquer dans le menu **Données>Filtre>AutoFiltre**.

Q.4) Répondez aux questions suivantes en utilisant l'auto-filtre :

q.4.1) Combien¹ d'étudiants ne savent pas s'ils utilisent des logiciels libres ?

q.4.2) Les étudiants qui jouent souvent à des jeux vidéos (jeux) consomment-ils souvent de la musique (Musique) ?

q.4.3) Combien d'étudiants n'ont ni smartphone (blog), ni accès internet ?

*Q.5) Répondez aux questions suivantes en utilisant des filtres standards (Menu **Données>Filtres>Filtres standard** ou **Filtre Standard** dans une des menus déroulants des AutoFiltres :*

*q.5.1) Combien d'étudiants sont de niveau élevé (autonome ou expert) et utilisent un ordinateur **moins de 11 heures** par semaine ?*

*q.5.2) Combien d'étudiants ont pratiqué **au moins** un peu de programmation ou **au moins** un peu d'applications scientifiques ?*

La fonction **SOUS.TOTAL** fait des calculs sur les données filtrées. Le premier paramètre de cette fonction est un numéro qui correspond à une fonction à calculer, et le second paramètre correspond à une plage de données (voir la documentation pour plus de détails).

► Rechercher dans l'aide de LibreOffice à quelles fonctions correspondent les différents numéros de fonction. Attention : il ne faut pas confondre la fonction **SOUS.TOTAL** (en majuscules) et le calcul de sous-totaux (*cf.* section suivante)

Q.6) (★)

q.6.1) À quelle fonction correspond chacun le numéro de fonction 9 ?

*q.6.2) Quelle formule permettra de répondre **automatiquement** aux questions « combien » précédentes ?*

► Utiliser la fonction **SOUS.TOTAL** pour calculer les moyennes pour les colonnes #heures, #Années et Niveau estimé. Vous ajouterez pour cela des formules utilisant la fonction **SOUS-TOTAL** dans des cellules, en bas du tableau, **séparé d'une ligne du reste du tableau pour éviter que cela soit considéré comme une nouvelle ligne de données.**

1 Il faut parfois répondre aux « combien » en comptant manuellement le nombre de ligne affichée !

Q.7) Donner les niveaux moyens des étudiants ayant respectivement le B2I Collège, le B2I Lycée, le B2I École et sans B2I ? (donner une formule et 4 résultats)

.....

Q.8) Le niveau moyen des étudiants qui passent plus de 6h/sem est-il supérieur à celui des étudiants qui passent moins de 6h/sem ? (donner la formule nécessaire pour répondre à la question, avec 2 manipulations)

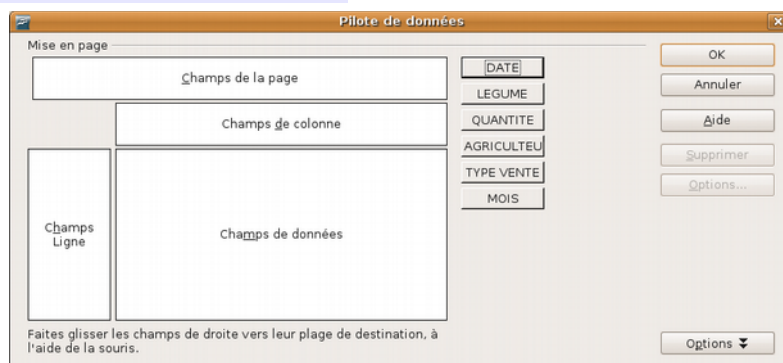
.....

3 Tableaux croisés dynamiques

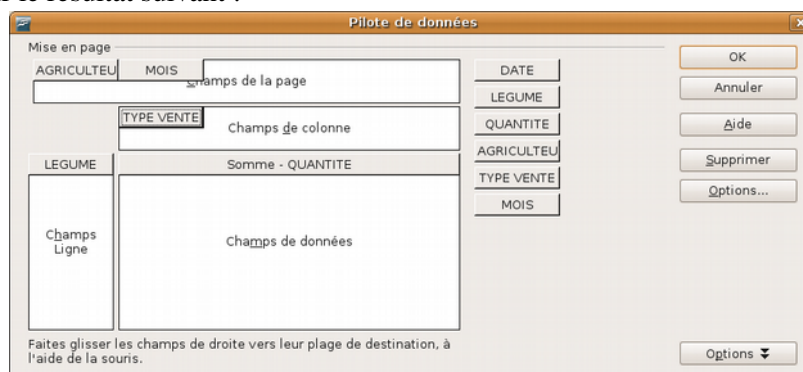
Fichier à utiliser : ProductionCoop.ods

Les tableaux croisés dynamiques permettent de faire une synthèse d'une grande masse de données. On parle de « tableaux croisés » car le tableau contient des données qui sont calculées en croisant les valeurs de deux variables. Ces tableaux sont dits « dynamiques » car les données qui sont prises en compte dans le tableau croisé peuvent être filtrées sans avoir à reconstruire le tableau. Les tableaux croisés dynamiques ne sont pas créés à l'aide de formules, mais à l'aide d'un assistant graphique !

- Sélectionner les données du tableau
- Lancer l'assistance tableau (nommé **Pilote de données** sous LibreOffice) dans le menu **Données>Table de Pilote>Créé**, la fenêtre ci dessous apparaît :



- Faire glisser les noms des listes de données se trouvant à droite vers les différents champs de sorte à avoir le résultat suivant :



- Le tableau construit va croiser les données LEGUME avec TYPE VENTE
- Le tableau va contenir la somme des QUANTITE. Vous pouvez modifier la fonction de calcul en cliquant sur **Somme - QUANTITE**.
- AGRICULTEUR et MOIS vont être utilisées pour filtrer le contenu du tableau (aspect « dynamique »)

- ▶ Dans les options, indiquer de mettre le « Résultat à » une nouvelle feuille de données.
- ▶ Cliquer sur **OK** ! De nouvelles informations ont été automatiquement ajoutées dans la nouvelle feuille de calcul. Parmi les cellules ajoutées, certaines disposent d'un menu déroulant avec lesquels vous pouvez modifier le contenu du tableau.

Q.9) Répondre aux questions de compréhension suivantes à partir du tableau croisé dynamique. L'utilisation des menus déroulants sera parfois requis.

q.9.1) Combien de navets ont été vendu en magasin par l'ensemble des agriculteurs ?

q.9.2) Quelle est la somme totale de carotte vendu par les agriculteurs ?

q.9.3) Combien de concombres ont été vendus en été (juillet et août) à des magasins par l'ensemble des agriculteurs ?

q.9.4) Combien de légumes a vendu Mr Dombasle aux grandes surfaces sur toute la période ?

- ▶ Construire un nouveau tableau dynamique qui permette d'afficher les quantités vendues en fonction des agriculteurs et des différentes structures. Dynamiquement, on pourra modifier les légumes comptabilisés dans la quantité vendue. Ce tableau permet de répondre une question telle que : es-ce que tous les agriculteurs vendent (tous les légumes ou les carottes ou les carottes et les poireaux, ...) aux différentes structures avec des répartitions similaires ?

Q.10) Comment construire un tableau ?


.....

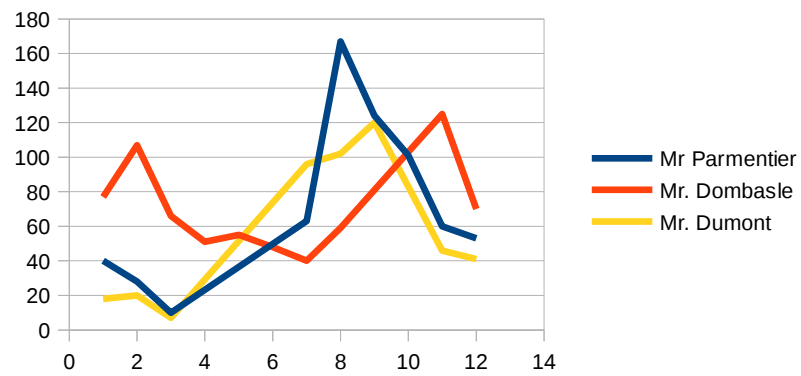
- ▶ Modifier le tableau de sorte à savoir si les quantités moyennes vendues aux différentes structures sont équivalentes entre les agriculteurs ?

Q.11) Quelle est la modification nécessaire ?

.....

Nous allons maintenant créer un graphique dynamique, c'est-à-dire un graphique dont le contenu est construit dans un tableau dynamique.

- ▶ Créer un tableau dynamique sur une nouvelle feuille de données. Ce tableau doit contenir les quantités de légumes produits par agriculteur (en colonne) et par mois. Le filtrage dynamique se fera sur les légumes.
- ▶ Ensuite faites un **clic-droit>Activer** sur l'un des champs pour revenir à la configuration du tableau. Dans la boîte de configuration du tableau, double-cliquer sur MOIS et cocher la case « Afficher les éléments sans données » (faire de même avec AGRICULTEUR). Cette modification des options permet d'assurer que le tableau dynamique aura toujours la même taille !
- ▶ Finalement, créer un graphique à partir des données du tableau :
 - ▶ Sélectionner les colonnes MOIS et les colonnes pour chacun des trois agriculteurs,
 - ▶ Lancer la création d'un graphique avec l'icône ,
 - ▶ Créer un graphique à l'allure ci-dessous ! Attention : il faut bien penser à donner des noms à vos séries en allant les chercher dans les bonnes cellules.



- Modifier le filtre pour comprendre l'intérêt de l'outil !