

Tableur : analyse de données

1 Introduction

Dans ce TP, l'objectif est de présenter certaines méthodes d'analyse de données disponibles dans un tableur. De nouveau, toutes ces fonctionnalités se retrouvent à la fois sous Excel et sous OpenOffice avec de grandes similarités entre les interfaces.

L'idée de ce TP directif est de vous faire connaître les possibilités d'un tableur pour traiter automatiquement et facilement de grandes « bases de données ». Si dans le reste de votre formation ou dans votre vie professionnelle vous en avez l'utilité poussée, consulter les documentations !

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 de vous-d'étudiants des années passées). Pour interpréter ces séries de chiffre, vous pouvez consulter le questionnaire correspondant aux réponses dans le fichier `QuestionnaireUsage.pdf` (les numéros correspondent aux numéros à côté des cases).
- `ProductionCoop.xls` : données (imaginaires) de vente de légumes par une coopérative de trois agriculteurs

2 Filtrer et parcourir des données

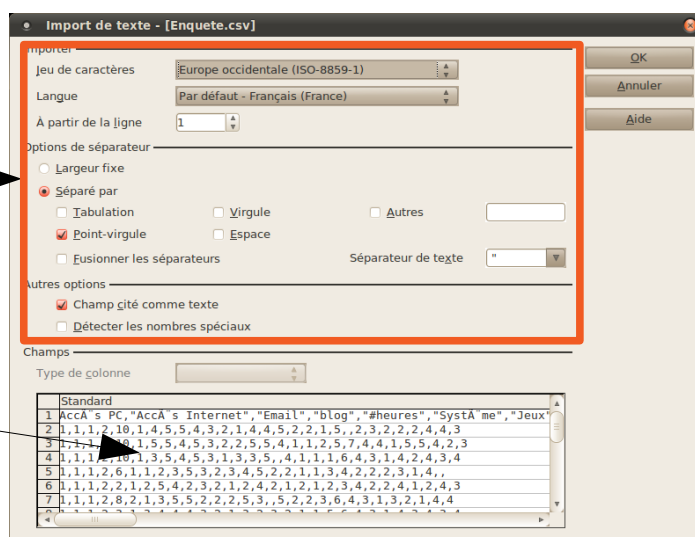
A Importation et exportation de données en format textuel

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 en mode Calc.
- Ouvrir le fichier `Enquete.csv` dans un éditeur de texte (*e.g.* Notepad++) et observer le contenu textuel du fichier,
- Ouvrir maintenant le fichier depuis votre tableur (**Fichier>Ouvrir**), 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



Lors que vous modifiez les configurations, la prévisualisation illustre la feuille de tableur qui sera construite après validation.

Q.1) Modifier la sélection des séparateurs pour indiquer comment décomposer les lignes en plusieurs colonnes contenant des chiffres. Utiliser la visualisation pour savoir comment sera interprété le fichier. 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.

B Utilisation de plage de données

Comme pour les noms de référence de cellule, on peut donner des noms à des **plages de données**. Contrairement à une référence, **une plage de données ne doit pas être organisée n'importe comment** : une plage de données désigne un ensemble de cellules définissant les caractéristiques d'un *individu sur une ligne*. Chaque **colonne** définit une caractéristique dont la cellule de la première ligne comprend le nom de la caractéristique. Cette façon d'organiser est **nécessaire** pour utiliser les outils de triage, de filtrage ou les tableaux croisés dynamiques, ...

Pour créer une plage de données (manipulation à faire pour les données d'enquête), il faut :

1. Sélectionner les données de la plage à définir (en incluant la première ligne contenant le nom de la colonne),
2. Aller dans le menu **Données>Définir la plage**,
3. Donner un nom explicite à votre plage de données, puis **OK** !

La fenêtre de navigation (accessible par la touche **F5**) affiche les plages de données déjà définies et permet de sélectionner les cellules de cette plage.

- ▶ **Définir une plage de données qui contient toutes les données de l'enquête (et choisir un nom approprié !)**
- ▶ Désélectionner les données et, ensuite, double-cliquer sur le nom de la plage de données dans le navigateur.

Que se passe-t-il ?

.....

C Tri et filtrage de données

Pour utiliser les fonctions de tri, d'auto-filtrage et de filtrage, il faut tout d'abord sélectionner toutes les données à manipuler (en particulier, il faut bien sélectionner toutes les colonnes et également les noms de colonne !).

Le plus simple pour réaliser correctement une sélection en vue d'utiliser les outils d'analyse de données est d'utiliser la plage de données qui a été définie plus haut. Ensuite, ces fonctionnalités sont accessibles dans le menu **Données**.

- ▶ 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 sélectionner 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** permet de faire 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 d'OpenOffice à quelle fonction 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*

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

« 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, **en dehors de la plage de données**.

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)

.....

D Sous-totaux automatiques

Fichier à utiliser : ProductionCoop.ods

Dans cette partie, l'objectif est de découvrir la fonctionnalité automatique de résumer d'un tableau de données. Cette fonctionnalité permet d'organiser les données par catégories et de calculer des sous-totaux par catégories. Cette fonctionnalité est intéressante pour visualiser rapidement un grand ensemble de données.

- Sélectionner les données du tableau à l'aide de la plage de données prédéfinie,
- Lancer l'assistant de construction de sous-totaux par le menu **Données>Sous-totaux**,
- Choisir les options suivantes :
 - 1er groupe : grouper par AGRICULTEUR, et calculer la somme des QUANTITE,
 - 2e groupe : grouper par TYPEVENTE, et calculer la somme des QUANTITE,
 - 3e groupe : grouper par MOIS, et calculer la somme des QUANTITE,
- Cliquer sur **OK**, les données ont été regroupées comme indiquer et à gauche des références de ligne se sont ajoutés des **+**, **-** permettant d'afficher ou de masquer les groupes.
- Cliquer sur les petits **1**, **2**,... **5** en haut à gauche

Q.9) Indiquer la quantité de légume vendue par chaque agriculteur de la coopérative ?

.....

Une fois cette manipulation effectuée, vous devez supprimer les sous-totaux pour la suite du TP. Pour cela, sélectionner de nouveau votre plage de données et retournez dans le menu **Données>Sous-totaux**, pour désactiver les sous-totaux, vous indiquez que vous souhaitez grouper par « Aucun » à tous les niveaux.

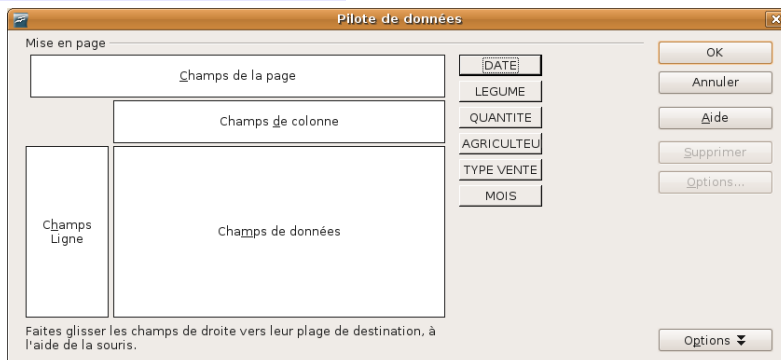
3 Tableaux croisés dynamiques

Fichier à utiliser : ProductionCoop.ods

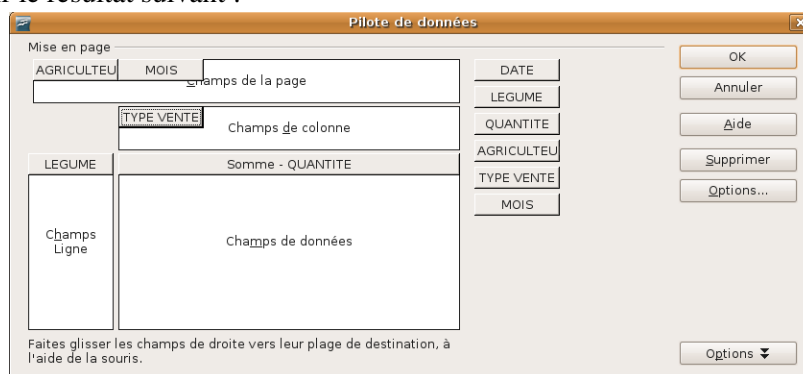
Les tableaux croisés dynamiques permettent de faire une synthèse d'une grande masse de données. On parle de « tableaux croisés » car le tableau contient des données qui sont calculées en croisant les valeurs de deux variables. Ces tableaux sont dits « dynamiques » car les données qui sont prises en compte 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 (à l'aide de la plage de données prédéfinie),
- Lancer l'assistance tableau (nommé **Pilote de données** sous OpenOffice) 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.10) Répondre aux questions de compréhension suivantes à partir du tableau croisé dynamique. L'utilisation des menus déroulants sera parfois requis.

q.10.1) Combien de navets ont été vendu en magasin par l'ensemble des agriculteurs ?

q.10.2) Quelle est la somme totale de carotte vendu par les agriculteurs ?

q.10.3) Combien de concombres ont été vendus en été (juillet et août) à des magasins par l'ensemble des agriculteurs ?

q.10.4) 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.11) Comment construire un tableau ?


.....

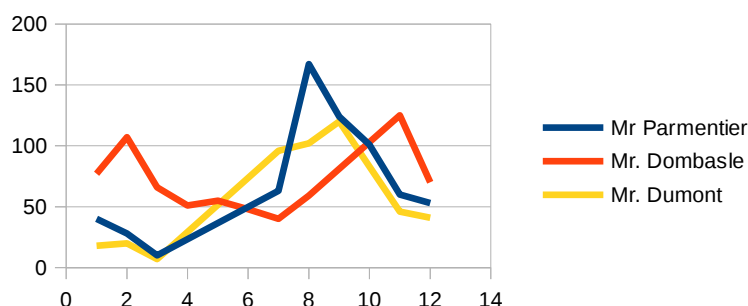
- Modifier le tableau de sorte à savoir si les quantités moyennes vendues aux différentes structures sont équivalentes entre les agriculteurs ?

Q.12) 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 légume produites 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 !