

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 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 `SOUS.TOTAL`)
- ▶ des tableaux croisés dynamiques
- ▶ des graphiques dynamiques

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

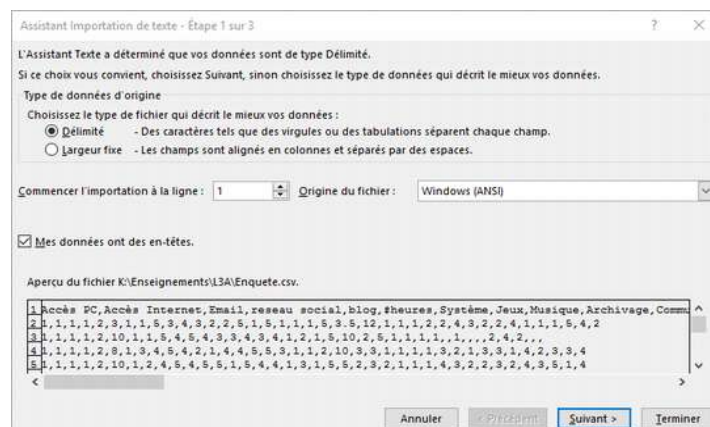
- ▶ `Enquete.csv` : données des réponses à un questionnaire sur vos usages des ordinateurs (menée auprès d'étudiants des années passées). Pour interpréter ces séries de chiffres, vous pouvez consulter le questionnaire correspondant aux réponses dans le fichier `QuestionnaireUsage.pdf` (**les numéros 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.

- ▶ Ouvrir un fichier Excel vide
- ▶ Dans le ruban Données, groupe Données Externes, utilisez le bouton « Fichier Texte » pour importer les données de votre fichier CSV. Cette fois-ci, une fenêtre apparaît pour définir la manière d'importer les données (préconfigurées automatiquement). Finalisez la configuration :
 - ▶ vérifier que vous utilisez un fichier « délimité »
 - ▶ le jeu de caractères doit être « Windows (ANSI) » (pour faire apparaître correctement les accents, **à vérifier** dans la visualisation au bas de la fenêtre)
 - ▶ cocher la case « Mes données ont des entêtes »



- ▶ valider avec **Suivant**
- ▶ Sélectionnez ensuite le séparateur virgule (observer les données pour justifier le choix), vous pouvez décocher les autres options, puis faire **Suivant**
- ▶ La fenêtre suivante permet de spécifier les natures des données de vos colonnes (Texte, nombres, etc.) : ne changez rien, faire **Terminer**
- ▶ Indiquez ensuite dans quelle cellule importer le tableau de données et voilà vos données importées correctement !

Mais ce n'est pas terminé ! Mon fichier contenait des nombres utilisant le point comme séparateur des décimales. Il faut changer cela pour que ces cellules soient reconnues comme des nombres et non comme des textes !

- ▶ Utiliser la combinaison de touches **Ctrl+H** (ou dans le Ruban Accueil, « Rechercher et remplacer ... ») et remplacer tous les points par des virgules.

Voilà, votre fichier est à présent bien formaté ! Vous pouvez maintenant enregistrer votre fichier **en format Excel**.

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 fonctionnalité souhaitée, Excel trouvera la sélection adéquate.

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

Q.1) 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.2) 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.3) Répondez aux questions suivantes en utilisant l'auto-filtre :

q.3.1) Combien¹ d'étudiants ne savent pas s'ils utilisent des logiciels libres ?

q.3.2) Les étudiants qui jouent souvent à des jeux vidéos (jeux) consomment-ils souvent de la musique (Musique) ?

q.3.3) Combien d'étudiants n'ont ni smartphone (blog), ni accès internet ?

*Q.4) 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.4.1) Combien d'étudiants sont de niveau élevé (autonome ou expert) et utilisent un ordinateur **moins de 11 heures** par semaine ?*

*q.4.2) 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 Excel** à quelle fonction correspond chacun des 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.5) (★)

q.5.1) À quelle fonction correspond chacun le numéro de fonction 9 ?

*q.5.2) Quelle formule permettra de répondre **automatiquement** aux questions « combien » précédentes ?*

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

- 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.**

Q.6) 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.7) 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)

C 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 ruban **Données**, icône **Sous-totaux**,
- Choisir les options suivantes :
 - à chaque changement de AGRICULTEUR,
 - Utilisation la fonction SOMME
 - Ajouter à un sous-total QUANTITE,
- Cliquer sur **OK**, les données ont été regroupées comme indiqué et à gauche des références de ligne se sont ajoutées des **+**, **-** permettant d'afficher ou de masquer les groupes.

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 pourrez utiliser le bouton « Supprimer tout ».

La fonction réalise un sous-total à chaque changement, mais de regroupe pas nécessairement toutes les lignes avec une valeur d'attribut identique (comme le fait LibreOffice). Pour que les sous-totaux s'affichent correctement, il est parfois nécessaire de commencer par trier vos données selon la colonne de regroupement. Vous pourrez faire le test avec un regroupement sur la colonne des types de vente.

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 !

- Placer le curseur dans le tableau de données,
- Lancer l'assistance tableau (depuis le ruban **Insertion**, groupe **Tableaux**, icône **Tableau Croisé Dynamique** (au singulier). Choisissez la création du tableau sur une nouvelle feuille de calculs.

► Une nouvelle feuille est créée et le menu de droite apparaît, comme représenté à droite

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

FILTRES AGRICULTEUR MOIS		COLONNES TYPE VENTE	
LIGNES LEGUME		VALEURS Somme de QUANTITE	

Champs de tableau croisé d... x

Choisissez les champs à inclure dans le rapport :

Rechercher

☐ DATE
☐ LEGUME
☐ QUANTITE
☐ AGRICULTEUR
☐ TYPE VENTE
☐ MOIS
 PLUS DE TABLES...

Faites glisser les champs dans les zones voulues ci-dessous :

FILTRES	COLONNES
LIGNES	VALEURS

☐ Différer la mise à jour de la disp...

- Le tableau construit dans l'illustration va croiser les données LEGUME avec TYPE VENTE
- Ce tableau va contenir la somme des QUANTITE. Vous pouvez modifier la fonction d'agrégation en cliquant sur le menu déroulant (menu Paramètres des champs de valeurs)
- AGRICULTEUR et MOIS vont être utilisées pour filtrer le contenu du tableau à l'aide des menus déroulants au dessus du TCD (aspect « dynamique »)

Q.8) Répondre aux questions de compréhension suivantes à partir du tableau croisé dynamique. L'utilisation des menus déroulants sera parfois requis.

q.8.1) Combien de navets ont été vendu en magasin par l'ensemble des agriculteurs ?

q.8.2) Quelle est la somme totale de carotte vendu par les agriculteurs ?

q.8.3) Combien de concombres ont été vendus en été (juillet et août) à des magasins par l'ensemble des agriculteurs ?

q.8.4) 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 : est-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 ?

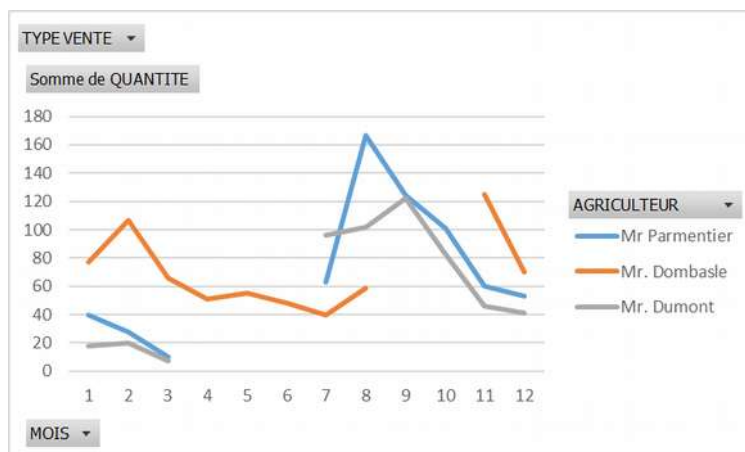
Q.9) Comment construire un tableau ?

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

Q.10) 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.
- Positionner ensuite votre curseur sur le tableau croisé dynamique et utiliser l'icône **Graphique croisé Dynamique** dans le groupe **Outils** du ruban **Analyse**.
- Sélectionner le type de graphique souhaité (courbe), le graphique devrait alors être créé.
 - Si besoin, modifier votre tableau croisé dynamique pour que les lignes et colonnes correspondent bien



- Modifier le filtre Type de vente pour comprendre l'intérêt de l'outil !
- Finalement, on souhaite que le tableau contienne une ligne pour chaque mois, même si il n'y a pas de vente pour ce mois. Hors, le comportement par défaut du tableau croisé dynamique est de supprimer les lignes sans données. Pour modifier ce comportement, faites un **clic-droit sur le champ mois** donné dans la fenêtre de configuration du tableau croisé dynamique, puis choisir le menu **Paramètre de champs**, dans l'onglet **Disposition et ...** décocher la case **ne pas afficher les lignes sans données**.