|  |
| --- |
| Les Couturiers Inclusifs |
| Procédure pour Collecte, Exploration et Nettoyage, des Données. |
|  |

|  |
| --- |
| Quitterie ROUDAUT  23/10/2024 |

Table des matières

[I. Collecter les Données 2](#_Toc181472242)

[1. Identification des Données 2](#_Toc181472243)

[2. Clarification des Données 2](#_Toc181472244)

[3. Hiérarchisation des Données 2](#_Toc181472245)

[II. Nettoyer et Explorer les Données 4](#_Toc181472246)

[1. Formater ses Métadonnées 4](#_Toc181472247)

[1.1 Inspecter son Fichier 4](#_Toc181472248)

[1.2 Alléger le Poids du Fichier 5](#_Toc181472249)

[2. Structurer ses Données 6](#_Toc181472250)

[2. 1 Préparer les Données : 6](#_Toc181472251)

[2.2 Convertir les Données en Tableau 7](#_Toc181472252)

[2.3 Adapter les Données du Tableau 8](#_Toc181472253)

[2.4 Données Personnelles 9](#_Toc181472254)

[2.5 Résoudre les Valeurs Vides 10](#_Toc181472255)

[2.6 Adapter le tableau en fonction des besoins 10](#_Toc181472256)

[III. Uniformiser le Processus de Nettoyage en Respectant les Règles RGPD 12](#_Toc181472257)

[1. Standardisation du processus de Nettoyage 12](#_Toc181472258)

[2. Protéger ses Datasets 13](#_Toc181472259)

[2.1 Protéger ses Feuilles de Calculs 13](#_Toc181472260)

[2.2 Protéger la Structure du Classeur 14](#_Toc181472261)

[2.3 Protéger le Classeur 14](#_Toc181472262)

[3. Respecter les Règles RGPD 15](#_Toc181472263)

[Résumer 16](#_Toc181472264)

# Collecter les Données

## Identification des Données

La première étape lors de la collecte des données est l’**identification** de celles-ci ainsi que leurs **provenances** (portes d’entrée) afin de bien les répertorier.

Il faut pouvoir répondre aux questions suivantes :

* Est-ce une base de données d’un prestataire ?
* S’agit-il de données produites par mon entreprise ?
* Comment j’ai eu ces données ? Gratuitement, via Internet ? Sur un réseau social ? Sur un site administratif ? Physiquement, lors d’un événement, salon ?
* Comment ont-elles été collecté ? site web, par un formulaire d’inscription ou un formulaire de contact ?

## Clarification des Données

La seconde étape, la **clarification** des données permet de définir quels types de données nous avons :

* Données à **caractères personnels** (photo, date de naissance, mails, numéros de téléphone…)
* Données **publiques** (dossiers, rapports, statistiques…)
* Données **privées** (protégés par des droits…)

## Hiérarchisation des Données

La hiérarchisation des données va nous aider à savoir pour quelles raisons nous allons utiliser nos données. Quelles vont être leurs **intérêts** et l’**usage** que nous allons en faire.

Quels types de données nous avons ?

* **Données de base :** largement partagées (numéro de commandes, quantités de produits, valeurs…)
* **Données de références (ou données "maîtres")** : données stratégiques : Ce sont des données qui vont permettre de décrire et de rechercher d’autres informations, notamment des données de base.
* **Métadonnées** : « données » sur vos données : type de fichier, format, origine, date etc. 🡺 Maitriser ses métadonnées permet d’améliorer la gestion de celles-ci et permettre une interopérabilité.

Une fois les données hiérarchisées, nous pouvons créer un dictionnaire de données/nomenclature pour se repérer dans nos données.

Lors de cette collecte des données, il faut respecter les règles RGPD en vigueur. (Fournis par l’entreprise)

En résumé :

**Alignement Stratégique** : mise en place de processus uniformisés au sein de l’entreprise, qui vont permettre de faire concorder les objectifs et la volonté politique de la direction avec les actions des services opérationnels et du métier.

# Nettoyer et Explorer les Données

Le nettoyage des données est une étape essentielle avant toute analyse. Certaines données peuvent provenir de plusieurs sources/portes d’entrées et peuvent donc être inexactes, incohérentes… Le nettoyage des données correspond au **processus de correction des erreurs identifiables dans le fichier de données**.

Ce nettoyage va nous permettre d’avoir des données de qualité ; améliorer la cohérence et la fiabilité des données ; ce qui est essentielle pour l’analyse.

## Formater ses Métadonnées

Après avoir collecté les données, et repérer les champs de données pertinents, nous pouvons préparer notre base de fichier.

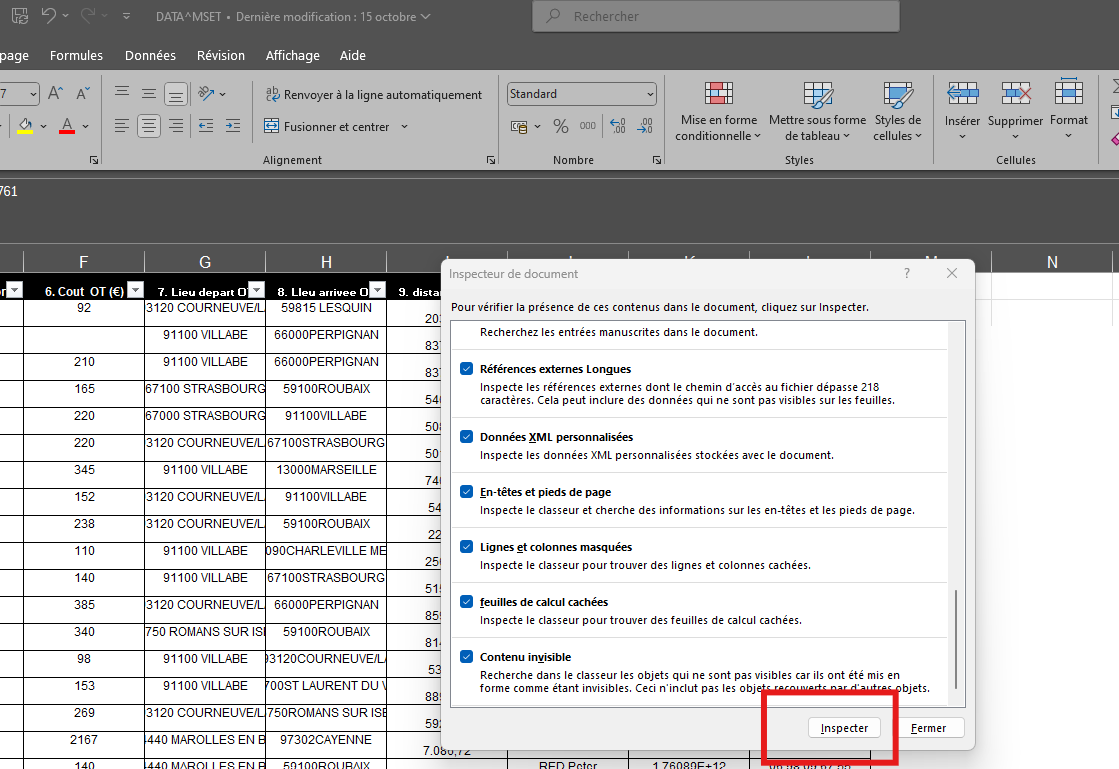
L’objectif est d’identifier les données essentielles.

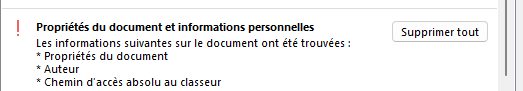
### 1.1 Inspecter son Fichier

Inspecter son fichier pour voir s’il n’y a pas de problèmes sur le fichier de base :

* Fichier > Information> Vérifier l’absence de problèmes > Inspecter le document >
* Cliquer sur « oui »
* Une image contenant texte, logiciel, Logiciel multimédia, Logiciel de graphisme

  Description générée automatiquement
* Cliquer sur inspecter :

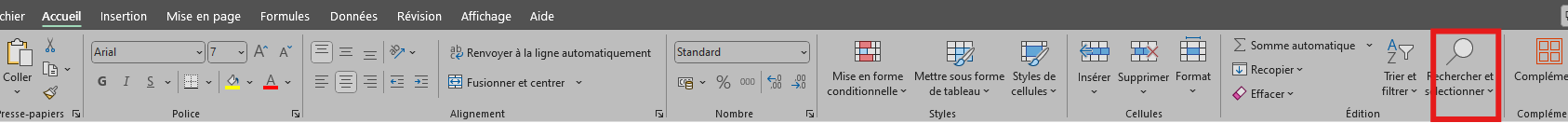




Avant de supprimer des données, il faut vérifier celles-ci. Excel nous donne ici des indications, à nous de juger, si nous devons prendre des actions ou non.

### 1.2 Alléger le Poids du Fichier

* Cliquer sur « Rechercher et sélectionner » > Sélectionner les cellules > Dernière cellule



Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

Description générée automatiquement

* Si elle ne contient aucune donnée > nous pouvons supprimer la cellule
* Clic droit sur la cellule, bouton “Supprimer”.
* Sélectionnez "Décaler les cellules vers la gauche" ou “Décaler les cellules vers le haut" selon votre besoin.
* Cliquez sur “OK”.
* Enregistrez le fichier.
* Une image contenant texte, capture d’écran, Police, nombre

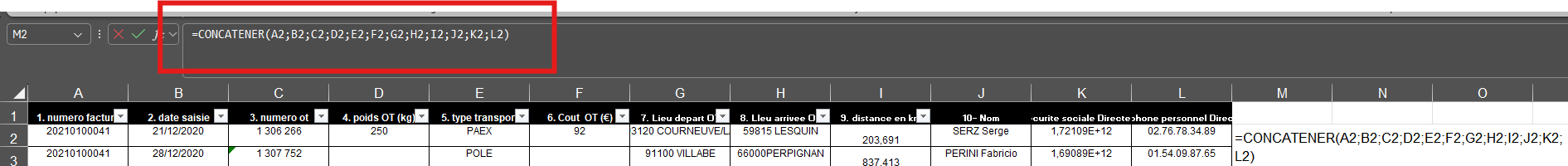
  Description générée automatiquement

## Structurer ses Données

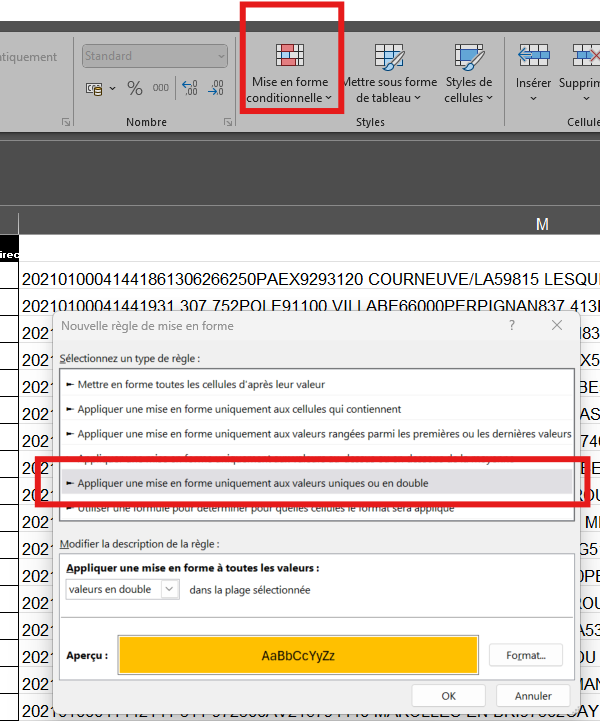
### 2. 1 Préparer les Données :

* Supprimer les espaces inutiles à l’aide de la fonction SUPPRESPACE ()
* Détecter les doublons : créer une colonne concaténation, CONCATENER ()

Ex :

****

* Tirer la formule sur toute la plage.
* Sélectionner la colonne, faire une mise en forme et sélectionner la mise en forme conditionnelle :



* S’il y a des doublons, les supprimer :
* Données > Supprimer les doublons

### 2.2 Convertir les Données en Tableau

Avoir ses données sous forme de tableau uniforme facilite la recherche car elles peuvent être structurées, standardisées. Cela va également, nous permettre un traitement plus rapide, facilité la visualisation, le reporting, détecter les anomalies et l’interopérabilité.

* Sélectionner la plage de données, cliquer sur insertion tableau :

Une image contenant texte, capture d’écran, logiciel, nombre

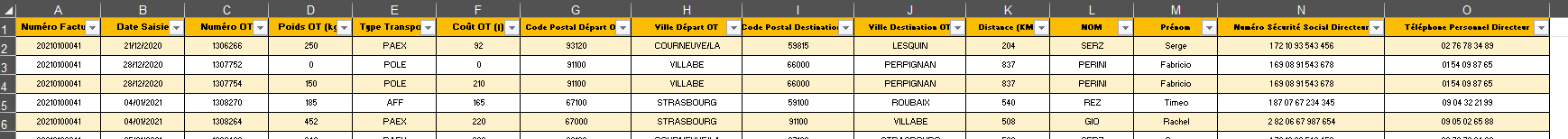
Description générée automatiquement

* Nommer le tableau :

Une image contenant texte, capture d’écran, affichage, logiciel

Description générée automatiquement

* Possibilité de mettre les tableaux en couleurs pour une meilleure lecture.



### 2.3 Adapter les Données du Tableau

* Vérifier que les données sont au bon format. Les triangles verts indiquent que le format n’est pas le bon.

Ex : colonne C « Numéro OT » sélectionner la colonne > clic droit et faire convertir en nombre.

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

* S’assurer que les datas en valeurs ont le même format et décimal.
* Ex : colonne D, nous avons des formats différents : Une image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquement
* Sélectionner la colonne,
* Contrôle H remplacer «, » par «  »

Une image contenant texte, nombre, ligne, capture d’écran

Description générée automatiquement

* Faire de même avec les autres colonnes jusqu’à uniformisation.
* Si le séparateur décimal est un point au lieu d’une virgule, la fonction **CNUM ()** ne fonctionne pas. Nous pouvons utiliser à la place la fonction **VALEURNOMBRE ()**.
* EX : créer une nouvelle colonne et utiliser la formule VALEUR NOMBRE :

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

* Copier/coller en valeur les datas de la nouvelle colonne et supprimer la colonne 5 de travail.
* On peut remarquer que nous n’avons plus de cases vides et qu’elles ont été remplacées par un zéro.
* Faire le même process pour toutes les colonnes contenant des valeurs.
* Checker les fautes d’orthographes des titres de colonnes et les modifier si besoin.

### 2.4 Données Personnelles

* Numéro de téléphone : garder le format texte ou indiquer un format spécifique permet de conserver toute la précision du numéro de téléphone et il est affiché exactement comme entré.
* Cela est particulièrement utile pour les numéros contenant des espaces ou des traits d’union, ou pour les formats internationaux.
* Pour utiliser un format spécifique :

Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

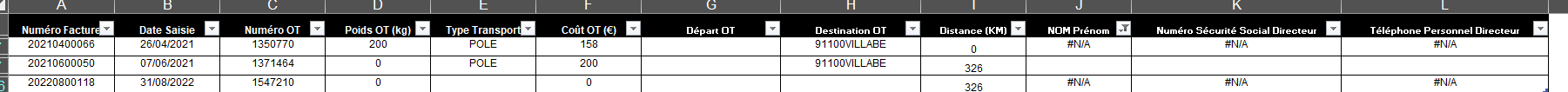
Description générée automatiquement

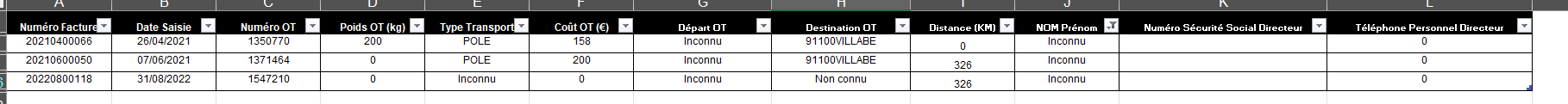
* Idem par exemple pour la sécurité sociale

### 2.5 Résoudre les Valeurs Vides

Pour avoir un dataset complet, il ne faut pas de valeurs vides. Nos données étant déjà sous forme de tableau il est facile de les repérer et de les modifier. Soit en ajoutant « 0 », ou « inconnu » selon la valeur que l’on souhaite donner à la cellule et le contexte, en filtrant sur les cellules vides.

EX :





### 2.6 Adapter le tableau en fonction des besoins

Dans certains cas, nous pouvons avoir plusieurs informations dans une même cellule. Il peut être intéressant en fonction de notre objectif de séparer ses données. (Une donnée par cellule). Par exemple : les colonnes G et I nous donnent plusieurs indications, nous pouvons les séparer en 4 colonnes avec le code postal et la ville ; que ça soit pour le départ comme pour l’arrivée.

* Utiliser les fonctions GAUCHE, DROITE, NB.CAR, TROUVE...

EX :

* **Pour extraire le code postal** :

=GAUCHE (A1 ; TROUVE (" "; A1) -1)

Cette formule extrait les premiers caractères jusqu'au premier espace (c'est-à-dire le code postal).

* **Pour extraire la ville**

=DROITE (A1 ; NBCAR(A1) -TROUVE (" "; A1))

* Cela extrait tout après le premier espace.

Une image contenant texte, capture d’écran, ligne, logiciel

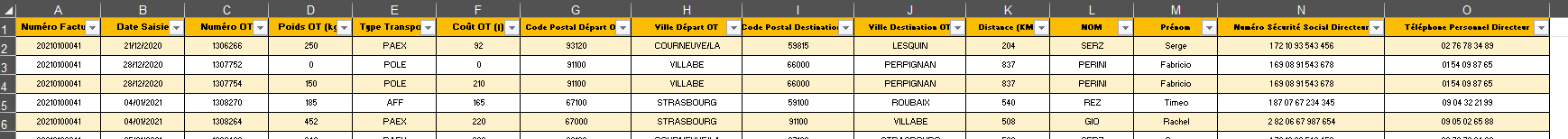
Description générée automatiquement

* Copier-Coller en valeur et supprimer la colonne G et convertir en nombre la colonne code Postal départ.
* Faire de même pour les autres colonnes.
* Faire un check avec les filtres si les codes postaux sont bien à 4 chiffres si ce n’est pas le cas, faire les modifications avec soit des formules, soit à la main si elles sont mineures.

Autres exemples :

=GAUCHE (L2 ; TROUVE ((" "; L2) -1) 🡺 nous donnes le nom de famille

=DROITE (L2 ; NBCAR(L2) -TROUVE ((" "; L2)) 🡺 nous donnes le prénom



Les datas sont à présent **nettoyées** et **prêtes** à être utiliser.

En résumé :

* Le nettoyage des données est un passage obligatoire avant l’utilisation de celles-ci. Plusieurs étapes doivent être respecter :

## Uniformiser le Processus de Nettoyage en Respectant les Règles RGPD

### Standardisation du processus de Nettoyage

Une fois le nettoyage effectué, il est important de garder des données correctes et de qualités.

* Définir quelles données sont utilisées ?
* Quand elles sont nécessaires ?
* Qui est responsable de la maintenance ?
* Fréquence de la maintenance ? (Journalière, hebdomadaire, mensuelle ?

Il est nécessaire de définir également la durée de vie des données, pour ne pas être enseveli avec des données inutiles.

Durée de vie :

Une image contenant texte, capture d’écran, Police, logo

Description générée automatiquement

* Pour ne pas supprimer une donnée, nous pouvons utiliser le processus d’**anonymisation.** Soit rendre impossible l’identification d’une personne.
* Exemple : utiliser le 1 chiffre de numéro de sécurité sociale pour définir un genre, sans le prénom de la personne.

## Protéger ses Datasets

Les mesures de sécurité à mettre en place dépendent de deux choses : de la vision du risque, et des risques réels (pour les personnes concernées, pour l’entreprise…) qui pèsent sur les données.

Les Datasets peuvent être utilisés par plusieurs personnes au sein d’une même entreprise. Il est donc important de protéger ses données, et fichiers :

### 2.1 Protéger ses Feuilles de Calculs

* Par défaut chaque cellule est protégée dans Excel.
* Pour avoir une protection opérationnelle à la feuille de calcul il faut enlever cette protection à la cellule.

Ex :

* Clique droite sur une cellule
* Format de cellule
* Protection
* Décocher la case « verrouillée »
* OK

Une image contenant texte, capture d’écran, logiciel, Police

Description générée automatiquement

* Aller dans l’onglet Révision
* « Protéger »
* Cliquer sur l’icône Protéger la feuille. Une image contenant texte, conception

  Description générée automatiquement

Une image contenant texte, Police, logiciel, Logiciel multimédia

Description générée automatiquement

* Saisir un mot de passe :
* Une image contenant texte, Appareils électroniques, capture d’écran, nombre

  Description générée automatiquement
* Laisser la sélection proposée par défaut
* La feuille de calcul est maintenant protégée

Pour enlever la protection, il suffit de retourner dans l’onglet révision > protéger et cliquer sur l’icône qui a changé « ôter la protection de la feuille »

### 2.2 Protéger la Structure du Classeur

Pour protéger le nom, ordre des onglets, la visibilité des onglets il faut protéger la structure :

* Onglet Révision
* Protéger
* Cliquer sur l’icone Une image contenant texte, conception

  Description générée automatiquement
* Excel propose de saisir un mot de passe + une confirmation.

### 2.3 Protéger le Classeur

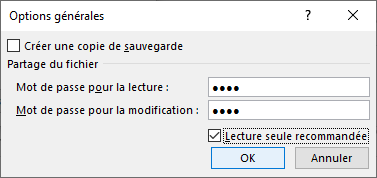
Mettre un mot de passe lors de l’ouverture d’un fichier, permet seulement aux utilisateurs autorisés de l’ouvrir.

* Aller dans l’onglet Accueil
* Enregistrer sous
* Cliquer sur « Outils »
* Sélectionner « Options générales »

Une image contenant texte, capture d’écran, logiciel, Page web

Description générée automatiquement

* Une boite de dialogue apparait avec deux possibilités de mot de passe
  + 1 : l’utilisateur doit avoir le mot de passe pour ouvrir le fichier
  + 2 : l’utilisateur doit avoir le mot de pour pouvoir modifier le fichier
* Il est important de choisir un mot de passe fort, avec des majuscules, minuscules, caractères spéciaux…
* Cliquer sur ok



* Le fichier est protégé.

## Respecter les Règles RGPD

En fonction de la sensibilité des données certaines règles doivent être respectés.

Les règles RGPD rendre en vigueur, et sont communiqué par l’entreprise.

Exemple de règles à respecter :

Une analyse de risques et d’impact relative à la protection des données doit être prise en compte dans le processus de standardisation.

# Résumer