

Le diagramme de la base de données et le script de création sont disponibles, analysez-les attentivement. Vous ferez tout d'abord la partie 2-A : Insertion des données. La partie 2-B : Requêtes SQL sera disponible lors de la prochaine séance d'autonomie.

**Astuce 1 :** Mettez dans un seul fichier .sql la création de la base, la manipulation des tables et les inserts. De cette manière, il vous sera plus facile de reprendre là où vous en étiez lors des séances de SAE.

**Astuce 2 :** Lorsque vous avez des équations dans votre tableur, n'oubliez pas de sauvegarder votre fichier au format .xlsx. Si, par mégarde, vous sauvegardez en .csv, vous perdrez toutes les équations.

**Astuce 3 :** Lorsque vous devrez supprimer des lignes, vous pouvez créer un attribut temporaire avec une valeur aléatoire entre 0 et 1, ordonner cette colonne, puis supprimer une portion de la table. De cette manière la suppression sera plus rapide et uniforme.

### Q1 : Manipulation des données

1. Ecrire le (ou les) ordre(s) SQL ajoutant automatiquement une valeur sur la clé primaire lors d'une insertion des tables relatives aux EtapeRH, ContratRH, Personne, et Projet.
2. Ajoutez deux attributs à la table d'association Exploiter, nommés TauxExploitation et Frequence. Ces éléments vont nous permettre de quantifier l'usage d'un équipement au sein d'un projet.
3. Ajoutez l'attribut ville à la table Personne.
4. De nouveaux type de salle et de type de publication peuvent être rajoutés. Supprimer les contraintes de vérification associée à ces attributs.

### Q2 : Insertion des données

1. Vous disposez d'un fichier INSERT.sql pour les tables de dimension TypeEquipement, TypeSalle, GuichetFinancement, TypePublication, Edition et Publication. Insérer les données du fichier .sql dans les tables de dimension avec pgAdmin. *Vous pouvez rajouter des éléments aux tables à votre convenance.*
2. Vous disposez d'un fichier INSERT.xlsx pour les tables ProcedureRH, StatutContrat, TypeContrat, Salle, Equipement, et les tables d'associations Suivre et Lie. Transformer les feuilles du classeur Excel en fichiers csv ou en fichiers sql. Insérer les données dans chaque table correspondante avec pgAdmin. *Vous pouvez rajouter des éléments aux tables à votre convenance.*

Nous allons pour la suite générer et insérer les données des tables Projet, Personne, ContratRH, et des tables d'associations EtapeRH, Affecter, Impliquer, Contribuer et Exploiter.

3. Générer et insérez les données de la table Projet :

Définissez environ 15 projets différents,

- Choisissez les noms à votre convenance,
- Une date de début de projet est définie aléatoirement entre 2024 et 2026,  
=ALEA.ENTRE.BORNES(2024,2026)
- Une durée prévisionnelle est définie aléatoirement entre 6 et 48 mois,
- Un budget est définie aléatoirement entre 10 000 et 1 million,
- Une date de dépôt du guichet est inférieure à la date de début de projet avec une différence aléatoire entre 10 et 90 jours,  
= date\_de\_début\_de\_projet - ALEA.ENTRE.BORNES(10,90)

- Et un numéro de guichet de financement pour chaque projet est à tirer aléatoirement entre les guichets de financement disponibles.

#### 4. Générez et insérez les données de la table Personne :

Définissez environ une centaine de personnes avec generatedata.com,

- Nom.
- Prénom.
- Mail personnelle composé du prénom et nom.
- Email professionnel se terminant par « univ-smb.fr ».
- Interne ou externe par un booléen aléatoire true/false.
- Ville française.
- Vous passerez ensuite sur excel pour générer le code INSEE et l'identifiant HALL. Le code INSEE sera construit par la concaténation de X éléments :

**Attention : soyez attentif à la taille des éléments qui doit être respectés, vous utiliserez la fonction =TEXTE(..., "00000") pour imposer la taille.**

- Pour le code INSEE :
  - Le sexe (taille 1) est défini aléatoirement pour les personnes entre 1 et 2,
  - L'année de naissance (taille 2) est définie aléatoirement entre 1955 et 2009,
  - Le mois de naissance (taille 2) est défini aléatoirement
  - Le lieu de naissance (taille 5), pour cela récupérer la base officielle des codes postaux INSEE du gouvernement. Et compléter le lieu de naissance en récupérant aléatoirement un code postal parmi les 39 193 codes disponibles. Pour cela, ajoutez la liste officielle dans une nouvelle feuille Excel, puis à partir de la formule ci-dessous récupérer une valeur aléatoire de la liste en générant un index *n* aléatoire,
   
`INDEX(Tableau1[Entete1], n)` permet d'extraire la nième valeur de la colonne Entete1
  - Et enfin, l'ordre de naissance sur le même lieu (taille 3) est défini aléatoirement.
- Pour le code HALL vous concatènerez une chaîne de caractères "hal-" et un code aléatoire de taille 7 ou 8.

#### 5. Générez et insérez les données de la table ContratRH :

Définissez environ 150 contrats :

- Attribuez ces contrats à des numéros de personnes aléatoire entre 1 et 90.
- Une date de début est définie aléatoirement entre 1985 et 2024.
- Une date de fin se compose de la date de début plus un nombre aléatoire de jours entre 30 et 5 000, vous en supprimerez une trentaine.
- Un statut de contrat est défini aléatoirement entre 1 et 5.
- Et enfin, un type de contrat aléatoire entre 1 et 3.

Nous allons maintenant pouvoir générer les tables d'associations EtapeRH, Affecter, Impliquer, Contribuer et Exploiter. Pour cela, vous réaliserez des produits cartésiens en sql (`CROSS JOIN`).

```
SELECT ...
FROM Table1
      CROSS JOIN Table2
ORDER BY 1, 2;
```

Un produit cartésien de deux tables, Table1 et Table2, appelé également ensemble-produit, est l'ensemble de tous les couples dont la première composante appartient à la Table1 et la seconde à la Table2. Nous y appliquerons ensuite une loi normale dans les attributs d'associations dépendante de certaines valeurs. Par conséquent, il est quelque fois nécessaire d'ajouter des jointures, voire des filtres, au produit cartésien afin de restreindre les résultats à certaines contraintes.

#### 6. Générez et téléchargez les cinq tables d'associations (les nombres de lignes donnés ici sont à titres indicatifs) :

- ❖ Les étapes RH dépendent des procédures RH et des contrats RH, mais également des dates de début et fin de contrat. *Vous obtiendrez près de 3200 lignes.*
- ❖ Les affectations dépendent des personnes et des salles. *Vous obtiendrez 3700 lignes.*
- ❖ Les implications relient les personnes à des projets. *Vous obtiendrez 1500 lignes.*
- ❖ Des contributions sont établies entre les personnes et les publications. Vous récupérerez également le statut de la personne au travers de son contrat. Vous obtiendrez 6000 lignes.
- ❖ Et enfin, l'exploitation des équipements dépendent des projets, mais également des types d'équipements. *Vous obtiendrez 225 lignes.*

Il est maintenant temps de compléter les produits cartésiens et de préciser certains attributs sur excel.

7. Générez et insérez les données de la table d'association EtapeRH :

- Supprimer aléatoirement 1/3 des lignes de votre table d'association. Veuillez supprimer complètement au moins une procédure RH et un contrat RH.
- A partir de la date de début et date de fin de contrat, définissez aléatoirement une date de mise à jour. Dans le cas où la date de fin est nulle, définissez une date de mise à jour aléatoire entre la date de début et la date d'aujourd'hui.
- Et enfin, un statut d'étape aléatoire entre les 3 choix possibles.

8. Générez et insérez les données de la table d'association Affecter :

- Supprimez aléatoirement environ 1/3 de la table d'association. Veuillez supprimer complètement au moins une personne et une salle.
- La date de début est définie aléatoirement entre 1990 et aujourd'hui.
- La date de fin est conditionnée :
  - o si l'affectation à une salle a débuté il y a moins d'un an, alors la date de fin sera laissée à NULL,
  - o dans le cas contraire, la date de fin est définie aléatoirement entre la date de début et la date d'aujourd'hui moins un an.
- Triez par ordre croissant le numéro de personne et supprimez au moins 1 personne de la table d'association.

9. Générez et insérez les données de la table d'association Contribuer :

- Supprimez aléatoirement environ 1/3 de la table d'association. Veuillez supprimer complètement au moins une publication et une personne.
- Sur une nouvelle feuille Excel, construisez un tableau avec les types de contribution :
  - o Relecteur,
  - o Données,
  - o Méthode,
  - o Ecriture,
  - o Idée,
  - o Expérimentation,
  - o Analyse
- Définissez aléatoirement un type de contribution pour chaque ligne.
- Triez par ordre croissant le numéro de personne et supprimez au moins 1 personne de la table d'association.

10. Générez et insérez les données de la table d'association Impliquer :

- Supprimez aléatoirement environ 1/3 de la table d'association. Veuillez supprimer complètement au moins une personne et un projet.
- Sur une nouvelle feuille Excel, construisez un tableau avec les rôles :
  - o Chef de projet
  - o Rédacteur
  - o Coordinateur
  - o Développeur
  - o Testeur
  - o Consultant
  - o Responsable de la communication

- Administrateur
- Triez par ordre croissant le numéro de projet et supprimez au moins 1 projet de la table d'association.

11. Générez et insérez les données de la table d'association Exploiter :

- Supprimez aléatoirement environ 1/3 de la table d'association. Veuillez supprimer complètement au moins un équipement et un projet.
- Sur une nouvelle feuille Excel, construisez une loi normale pour le taux d'exploitation suivant ces paramètres :

Type équipement	Moyenne	Ecart type
1 : Ordinateur de bureau	Entre 65 et 85	Entre 5 et 12
2 : Station de travail	Entre 15 et 65	Entre 1 et 15
3 : Serveur de calcul	Entre 10 et 35	Entre 3 et 10
4 : Drone	10	Entre 2 et 11

- Ensuite, vous procéderez comme dans l'exemple des fichiers EXPLICATION - Exemple application distribution normale xlsx et pdf à savoir :
  - En utilisant la fonction `ALEA.ENTRE.BORNES`, pour définir à chaque ligne de votre table d'association une espérance et un écart-type associés à l'attribut. Pour cela, vous aller devoir rechercher la valeur minimale et la valeur maximale de l'espérance pour en obtenir une valeur aléatoire. Puis, refaire la même opération pour l'écart-type.  
`=ALEA.ENTRE.BORNES(RECHERCHEV(..., FAUX), RECHERCHEV(..., FAUX))`
  - Générez ensuite les valeurs aléatoires contraintes par ces 2 paramètres en utilisant la fonction `LOI.NORMALE.INVERSE.N`
  - Convertissez les valeurs produites en nombres entiers.
  - Corriger les valeurs pour respecter une plage de valeurs comprises entre 0 et 100.
- Supprimez encore quelques lignes.
- Effacez par-ci par-là quelques valeurs de nombres et indiquez à la place de la valeur numérique, un commentaire sur l'attribut textuel tiré de cette liste :
  - Ponctuelle,
  - Régulière,
  - Permanente.

	A	B	C	D
1	numequipem	numprojet	TauxExploitati	Frequence
2	3	3	81	
3	3	4	60	
4	3	5		Régulière
5	3	6	74	
6	3	7	83	
7	3	8	78	
8	3	9	57	
9	3	10		Permanente
10	4	8	66	
11	4	9	51	
12	4	10	56	
13	4	11	20	
14	4	12	28	
15	4	13		Ponctuelle
16	4	14	53	
17	4	15	31	
18	5	1	78	

**A RENDRE DANS LE CASIER /INFO1\_SAE204/[Votre Dossier]/Partie 2 :**

- **Tous les fichiers Excel et CSV créés ou générés.**
- **Vous réaliserez une documentation expliquant les scripts SQL que vous avez utilisé pour la manipulation des données, mais aussi pour l'insertion des données, tel que l'ordre d'insertion ou les produits cartésiens.**
- **Rendre le script de sauvegarde de la base de données en sélectionnant les options suivantes :**

○ **Fichier plat (« Plain ») :**

Sauvegarde (Base de données: SCODOC)

General Options de sauvegarde

Nom de fichier	BD sujet1 TP11.sql
Format	Fichier plat
Niveau de compression [0...9]	
Encodage	Sélectionner un objet...
Nombres de tâches	
Rôle	Sélectionner un objet...

○ **Options :**

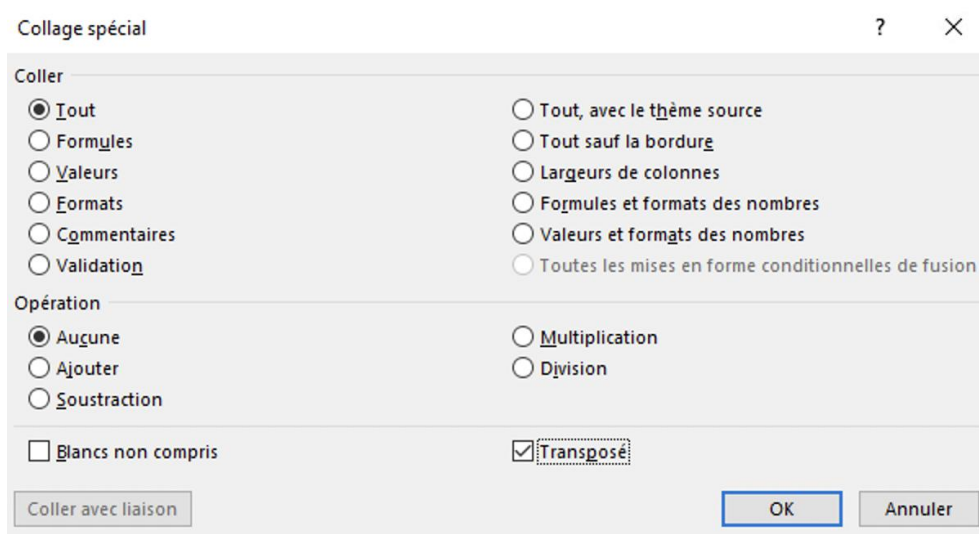
Sauvegarde (Base de données: SCODOC)

General Options de sauvegarde

<b>Sections</b>	
Pre-data	No
Post-data	No
<b>Type d'objets</b>	
Données uniquement	No
Blobs	No
<b>Ne pas enregistrer</b>	
Propriétaire	Oui
Tablespace	Oui
Commentaires	Oui
<b>Requête</b>	
Utiliser les commandes INSERT avec noms de colonnes (~column-inserts)	Oui
Utiliser l'instruction CREATE DATABASE	No
Charger Via la partition racine	No
<b>Disable</b>	
Déclencheur	No
<b>Divers</b>	
Avec OID(s)	No
Forcer les guillemets doubles des identifiants (~quote-all-identifiers)	No

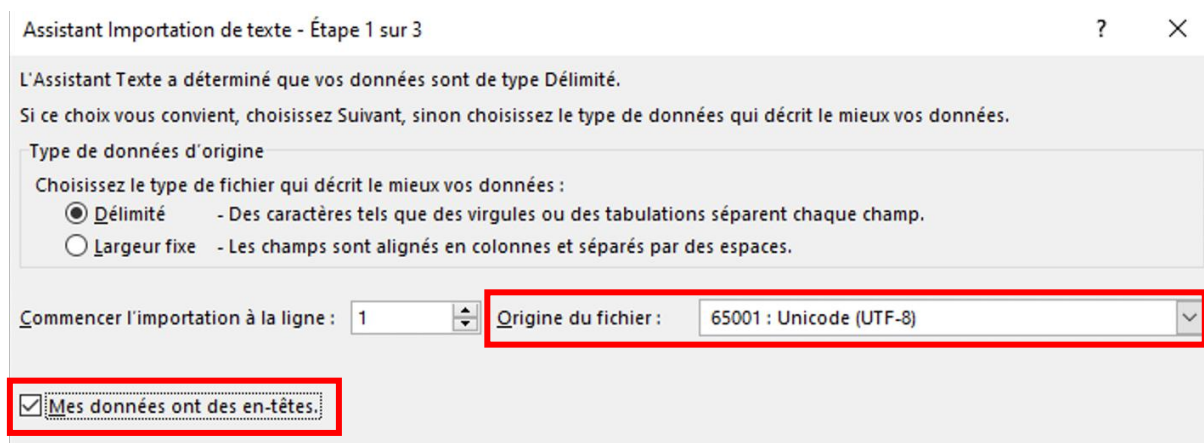
## Trucs et astuces :

- Dans Excel, pour copier des données d'une ligne et les transformer en colonne, copier les valeurs en lignes, puis *Bouton droit de la souris* > *Collage spécial*, sélectionner « Transposé »

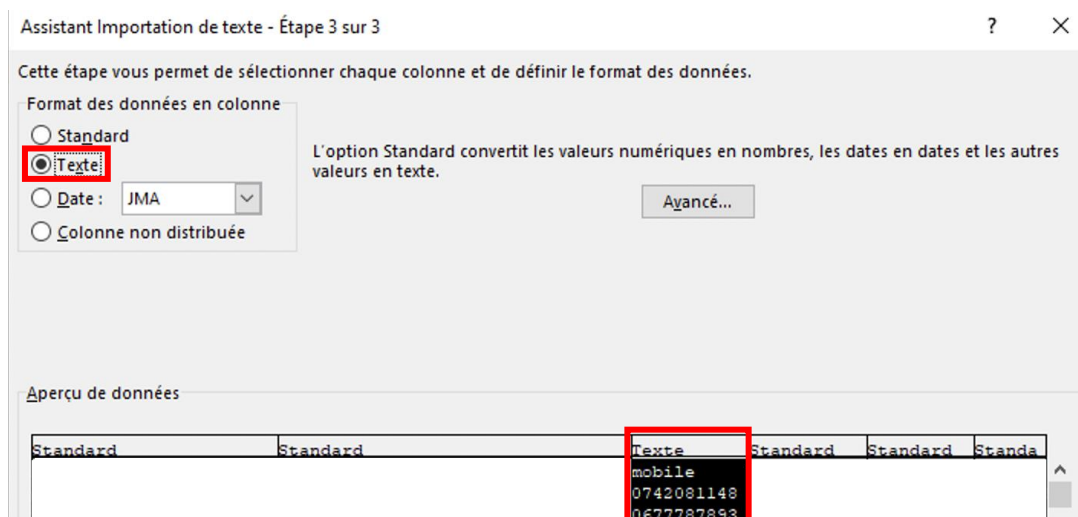


Idem pour colonne -> ligne.

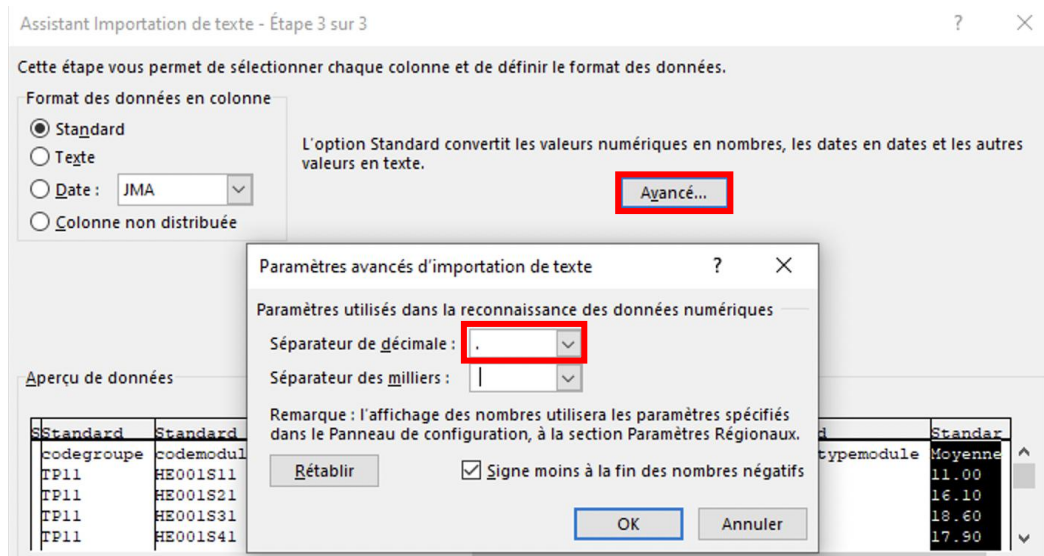
- Lorsque vous importez un fichier CSV avec Excel, penser à choisir l'origine adéquate du fichier (jeu de caractères) :



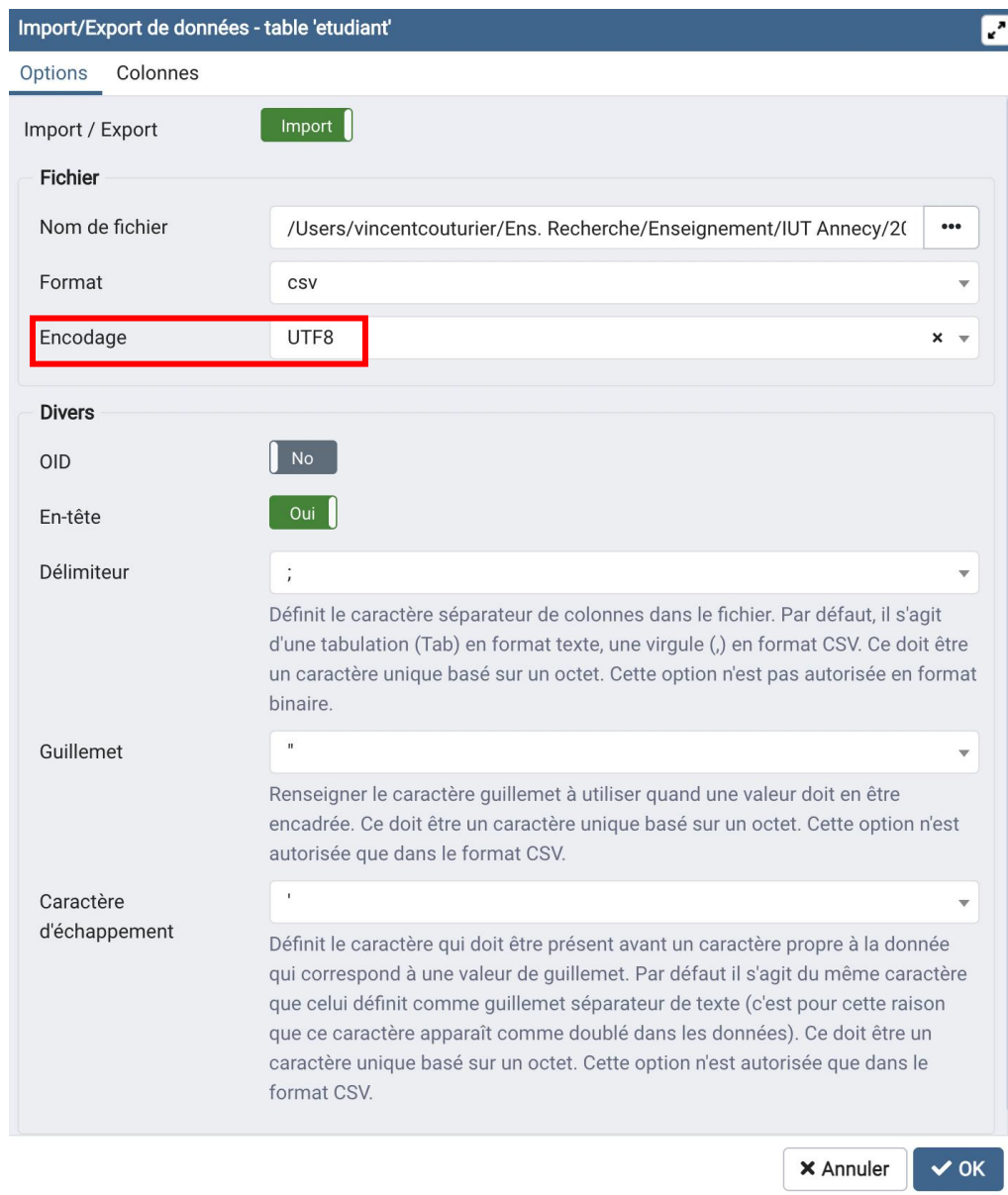
- Si le champ commence par 0 comme un n° de téléphone, pour le conserver il faudra choisir le format « Texte » lors de l'importation du CSV :



- Si le fichier à importer contient des données décimales, cliquer sur le bouton « Avancé » et modifier le séparateur décimal :



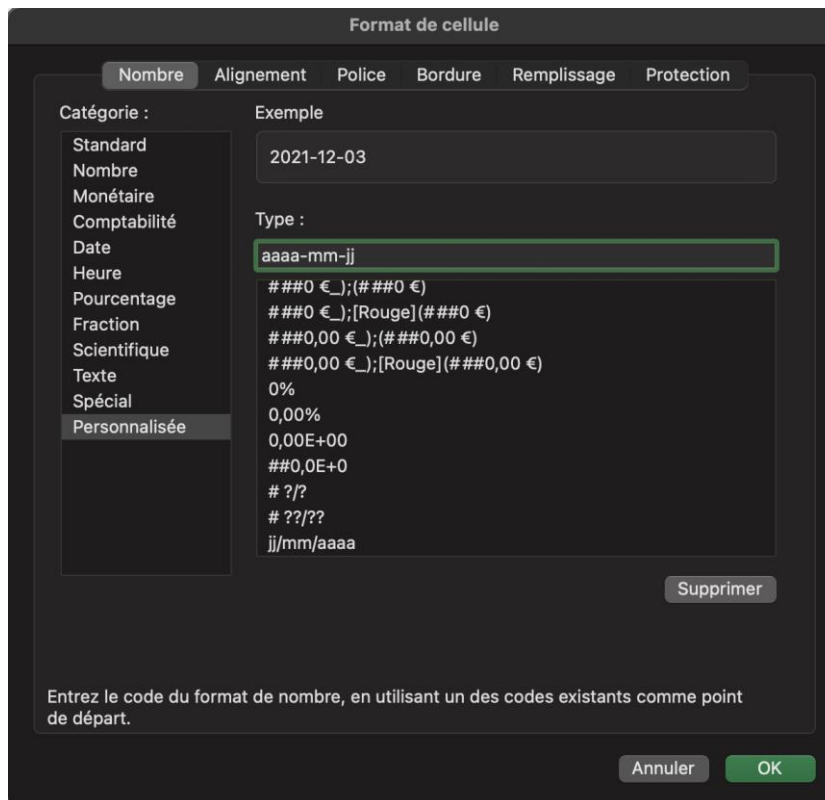
- Lorsque vous importez un fichier CSV dans PostgreSQL, indiquer l'encodage :





Ne pas oublier de remplacer la virgule « , » des décimaux par des « . » avant d'importer les données dans PostgreSQL.

- Appliquer le format de date `aaaa-mm-jj` dans Excel et `y-LL-dd` dans Generatedata.  
Exemple avec Excel :



- Pour enregistrer le résultat d'une requête dans un fichier CSV, utiliser le bouton  de PgAdmin4 (ou touche F8).