

TD1 | TP1

S2.04 - Exploitation d'une base de données

Réunion avec M. Bruyère Lundi 21/03/2022 à 12h

Compte rendu étape 1Correction des éléments du rapport 1 :

Après avoir exporté les fichiers CSV de la base de données :

- La procédure pour reconstituer un tableau individu-variable grâce à Excel est :

- Il faut d'abord mettre chaque feuille CSV sur une seul et unique feuille Excel
- Ensuite, il faut utiliser la fonction « recherchev() » afin de pouvoir lier chaque identifiant à la feuille correspondante pour avoir accès à une variable dont on a besoin mais qui se trouve sur une autre feuille.
- Après avoir lié tous les codes et donc obtenu toutes les variables donc on avait besoin, il suffit de supprimer les colonnes qui contiennent les identifiants afin d'obtenir le tableau individu-variable.

- La procédure pour reconstituer un tableau individu-variable sur python est :

- Nous allons utiliser la bibliothèque « Pandas » qui va nous permettre de manipuler nos données grâce à des data frames.
- Il faut maintenant lire les fichiers CSV et les importer dans des data frames grâce à la fonction « read_table() » de Pandas.
- Ensuite, il faut concaténer/faire une jointure entre les data frames à l'aide de la fonction « merge() » de Pandas afin de constituer le tableau individu-variable.
- Et enfin, il suffit de supprimer les identifiants qui ne nous servent plus afin d'obtenir le tableau individu-variable grâce à la fonction « drop() ».

Liste des variables et analyse (nom, type, intervalle ou liste de modalités) :

| table_origine | nom_variable | type_variable | intervalle ou modalité |
|------------------|--------------------------------|-----------------------|--|
| MAccident | nb_blesses_graves | quantitative discrète | 0 à 6 |
| | nb_blesses_legers | quantitative discrète | 0 à 24 |
| | nb_indemne | quantitative discrète | 0 à 9 |
| | gravite | qualitative ordinale | 0 à 3 |
| | nb_mort | quantitative discrète | 0 à 2 |
| MEtatSurface | libelle_etat_surface | qualitative ordinale | nombre de modalités : 8 (Humide, Mouillée ...) |
| MTypeEtatSurface | libelle_type_code_etat_surface | qualitative ordinale | T_Dangereux, Dangereux, Normal |
| MLieu | commune | qualitative nominale | nombre de modalités : 87 (communes du Nord) |
| | x | qualitative nominale | 632527 à 666057 |
| | y | qualitative nominale | moins 314394 à 343412 |
| Mluminosite | libelle | qualitative nominale | Jours, Demi-jour, Nuit éclairée, Nuit éclairée insuffisant, Nuit sans éclairage. |
| | type_luminosite | qualitative ordinale | 1 ou 2 (On modifiera en 0 ou 1 pour faire des pourcentages) |
| | libelle_luminosite | qualitative ordinale | Jour, Nuit |
| Mcause | libelle | qualitative nominale | nombre de modalités : 28 (Stationnement, Maitrise du vehicule ...) |
| Mdate | DateFormatStandard | qualitative ordinale | 1984-01-01 05:30:00 à 30/04/1998 14:50:00 |
| | Commentaire | qualitative nominale | nombre de modalités : 12 (en comptant NULL). (Précise si il y a des jours fériés par exemple le Lundi de Pâques) |
| Mimplique | libelle | qualitative nominale | nombre de modalités : 10 (Pieton -10 ans, Tram Engin...) |
| Mintemperie | libelle | qualitative nominale | nombre de modalités : 8 (Beau temps, Neige ...) |
| MTypeImplication | libelleType | qualitative nominale | Autre raison, humain, deux roues, véhicule |

Test sur la variable gravité :

0 → jamais de blessés graves ni de morts , très rarement de blessés légers, un nombre d'indemne normal.

1 → jamais de blessés graves ni de morts, souvent des blessés légers et des indemnes

2 → tout le temps des blessés graves, quelques fois des blessés légers et des indemnes, pas de mort

3 → tout le temps des morts, quelques fois des blessés légers, des indemnes et des blessés graves (problème sur la variable nb_mort)

Amélioration de la problématique :

Suite à notre réunion, nous avons conclu qu'il était plus intéressant d'étudier une part d'accident au lieu d'un nombre d'accident. De plus, pour éviter de trop limiter nos recherches, on a décidé d'étudier chaque type de véhicule et plus uniquement les vélos. Notre nouvelle problématique est donc :

Le moment de l'année (jours fériés / fêtes / saisons / week-end / vacances ...) est-il un facteur d'accident ? A quel moment de l'année y-a-t-il la plus grande part d'accidents par type de véhicule ? Pendant les jours de congé la gravité des accidents est-elle plus élevée ?

Compétences améliorées lors de cette SAE :

- Compétence 4 Gérer des données de l'information : car nous exploitons un type complexe de données : une base de données.
- Compétence 6 Travailler en équipe informatique : car nous sommes en équipe ce qui nous permet de nous partager le travail et ainsi gagner du temps.