

SAE2.02 Exploitation d'une base de données

COMPTE RENDU

Lieu	Salle de réunion	Membres présents	
Date et heure	28/03/2022 à 11h40	Nicolas	DARGAZANLI
Absents	Aucun	Alexandre	MAURICE
Rédacteur du compte-rendu	Guillaume TRITSCH	Guillaume	TRITSCH
Date de rédaction	30/03/2022		Mme. Bruyère

OBJECTIF DE LA REUNION

Présenter le rapport de l'étape 1 comportant le tableau *individus-variables*, sur les données statistiques, et sur la problématique choisie.

SUJET DE DISCUSSION

Méthode utilisée pour mySQL, Python et Excel

En début de réunion, nous avons créé le tableau *individus-variables* à partir de SQL à l'aide de JOIN, et sur Excel à l'aide de RECHERCHEV().

On nous a donc apporté l'explication sur le fonctionnement de la bibliothèque Pandas, l'utilité des *dataframes* dans la manipulation de données, de l'utilisation de *read_table()*, et une piste pour permettre la création du tableau *individus-variable* (l'utilisation de la commande *merge*, similaire à une jointure sur SQL).

Détails des variables

Le type que nous avons donné à la variable *nb_blesses_legers* doit être justifié si nous continuons à la typé comme variable quantitative discrète. **Nous avons décidé de rester sur une variable de type quantitative discrète car en dehors de la valeur extrême 24**, dont la table ne comporte qu'une seule occurrence, la plage de valeur s'établit entre 0 et 9 (comme illustré sur la figure 1).

Figure 1 :

valeur de 'nb_blesses_legers'	nombre d'occurrence
0	4983
1	19614
2	3610
3	982
4	338
5	115
6	48
7	13
8	4
9	2
24	1

Nous devons aussi modifier la manière dont nous nommions les données, car le nom « libellé » apparaissait plusieurs fois. Nous avons donc mis à jour pour plus de précisions grâce à un modèle du type *libelle_<table>*.

Problématique

Nous devons préciser avec plus de finesse notre problématique, afin que les études réalisées soient le plus clair possible. Aussi, nous n'avons pas précisé qu'une telle problématique (l'étude de l'impact des jours particuliers de l'année sur le nombre d'accidents, leurs parties impliquées et leurs causes) nécessite une certaine flexibilité pour la durée de l'étude. Nous devons prendre en compte, en fonction du jour, une certaine marge exprimée en jours. Nous en venons alors à l'étude des ponts : il est possible de les étudier eux aussi pour voir et confirmer leur impact ou non sur le nombre d'accidents.

La même question se pose aussi pour les vacances scolaires, mais nous avons décidé après cette réunion de ne pas nous y attarder étant donné que les zones des vacances scolaires ont changé tous les ans pendant la durée totale de l'étude (de 1984 à 1998), rendant la tâche beaucoup plus compliquée.

Pour revenir à la manière dont nous allons étudier ces données, nous avons évoqué un éventuel Δ qui mettrait en valeur l'impact des jours particuliers sur le nombre d'accidents par rapports aux jours dits « normaux », en fonction de la moyenne totale des accidents par jour des 14 années, ramenée sur une année.

Pour cette étude, un lissage et l'utilisation de médianes peuvent être réellement intéressants. Pour ce qui est de la période étudiée, nous concentrer sur l'année plutôt que le mois se trouvait être plus intéressant.

Autre remarque

Chaque membre du groupe doit réaliser la prise de notes durant la réunion afin de réunir le plus d'informations, et d'éviter les risques de malentendus.

TRAVAIL A FAIRE

Nous devons, au plus tôt, envoyer le rapport retravaillé avec le tableau complet, l'explication du fonctionnement de la création du tableau *individus-variables* sous MS Excel, et une explication similaire pour *Pandas* sur Python, avec les détails concernant le fonctionnement d'un *dataframe*, des fonctions *merge()* et *read_table()*.