DARGAZANLI Nicolas

MAURICE Alexandre

TRITSCH Guillaume

TD I TP 1

**S2.04 – Exploitation d’une base de données**

**Rapport de l’Étape 1 : Prise en main de la base et réflexion statistique**

# Méthode :

Les trois différentes parties sont rédigées dans l’ordre de réalisation.

## Réalisation d’une vue dans mySQL :

Nous avons commencé par la création d’une requête regroupant toutes les données nous intéressant d’un point de vue statistique. On joint alors toutes les « tables » entre elles afin de ne former qu’un seul tableau, en prenant soin de ne pas garder les données les identifiants. Nous avons figé le résultat dans une vue contenant alors toutes les données. Cette vue peut ensuite être reprise éventuellement dans MS Excel ou avec la bibliothèque Pandas, mais nous n’avons pas pour objectif de l’utiliser pour la méthodologie.

## Utilisation de la fonction RECHERCHEV( ) de MS Excel :

La méthode RECHERCHEV( ) est utile quand nous avons besoin de réaliser une jointure entre deux feuilles (qui correspondent à des tables en SQL). Dans cette SAÉ nous l’utiliserons donc comme élément de jointure des deux identifiants de deux tables différentes. Dans sa forme la plus simple, la fonction RECHERCHEV a la signification suivante :

=RECHERCHEV(ce que vous voulez rechercher, où vous voulez le rechercher, le numéro de colonne dans la plage contenant la valeur à renvoyer, renvoyer une correspondance approximative ou exacte ( indiquée par VRAI ou FAUX).

En pratique :

* Nous importons deux tables (issues d’une exportation en .csv depuis PHPmyAdmin).
* Nous insérons dans la feuille qui va regrouper les données une nouvelle colonne contenant un libellé d’une table étrangère afin d’expliciter un identifiant.
* Saisie de l’adresse de la valeur cherchée, puis de la matrice située dans la table que l’on souhaite joindre en précisant la colonne que l’on souhaite garder, en attribuant à « valeur proche » la valeur FAUX.
* On peut ensuite étendre la valeur cherchée sur toute la plage restante

## Explications sur la bibliothèque Pandas de Python :

La bibliothèque Pandas permet de créer des DataFrames, une structure de données à deux dimensions, avec des colonnes de types différents, à la manière d’une feuille de calcul sous MS Excel, d’une table SQL ou d’un dictionnaire d’objets. Ces DataFrames sont très utilisés pour les études statistiques. Nous disposons dans cette bibliothèque d’une méthode merge( ) qui permet de faire la jointure deux DataFrames. De la même manière que la méthode JOIN en SQL nous donnons en paramètre les deux colonnes des DataFrames, qui représentent les identifiants. Cette méthode retourne un nouveau DataFrames avec la jointure des deux DataFrames.

# Dictionnaire de données :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nom** | **Signification** | **Type statistique** | **Type informatique** | **Table d’origine** | **Plage des valeurs** |
| **Nb\_blesses\_graves** | Nombre de blessés graves | Quantitative discrète | Entier (5) | MAccident | 0 à 6 |
| **Nb\_blesses\_legers** | Nombre de blessés légers | Quantitative discrète | Entier (5) | MAccident | 0 à 9, 24 |
| **Nb\_indemnes** | Nombre de personnes impliquées indemnes | Quantitative discrète | Entier (5) | MAccident | 0 à 9 |
| **Gravite** | Gravité de l’accident | Qualitative ordinale | Entier (5) | MAccident | 1 à 3 |
| **Nb\_morts** | Nombre de morts | Quantitative discrète | Entier (5) | MAccident | 0 à 2 |
| **LibelleCause** | Cause de l’accident | Qualitative nominale | Chaine de caractères | MCause | Annexe 1 |
| **Date\_format\_standard** | Date de l’accident | Qualitative ordinale | Date | MDate | 1984-01-01 5:30:00  Au 1998-04-30 14 :50 :00 |
| **Commentaire** | Informations supplémentaires concernant la date | Qualitative ordinale | Chaine de caractères | MDate | Jours fériés : Armistice, Ascension, Assomption, Fête du travail, Fête nationale, Jour de l'an, Lundi de Pâques, Lundi de Pentecôte, Noël, Toussaint, Victoire des alliés |
| **Libelle\_etat\_surface** | Description de l’état de la surface | Qualitative nominale | Chaine de caractères | MEtatSurface | Humide, Mouillee, Enneigee, Verglacee, Gras boueux, Gravillons, Sec Normal |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LibelleImplication** | Type de véhicule impliqué ou personne impliquée | Qualitative nominale | Chaine de caractères | MImplique | Pieton moins 10 ans, Pieton plus 10 ans,  Bicyclette, Cyclomoteur,  Moto, Voiture legere, Poids lourd, Bus,  Tram, Engin, Autre |
| **LibelleIntemperie** | Situation météorologique au moment de l’accident | Qualitative nominale | Chaine de caractères | MIntemperie | Beau temps, Pluie forte, Pluie legere, Neige, Grele, Brouillard, Vent fort, tempete, Inconnu |
| **Commune** | Code postal de la commune où a eu lieu l’accident | Qualitative nominale | Entier (10) | MLieu | Annexe 2 |
| **X** | Coordonnée x de l’accident | Quantitative continue | Entier (10) | MLieu | Trop de valeurs |
| **y** | Coordonnée y de l’accident | Quantitative continue | Entier (10) | MLieu | Trop de valeurs |
| **LibelleEclairage** | Description de l’éclairage | Qualitative nominale | Chaine de caractères | MLuminosite | Jour, Demi jour, Nuit eclairee, Nuit eclairee, insuffisant, Nuit sans eclairage |
| **Libelle\_luminosite** | Jour ou nuit | Qualitative nominale | Chaine de caractères | MLuminosite | Jour, Nuit |
| **Libelle\_type\_code\_etat\_surface** | Niveau de dangerosité de la surface | Qualitative ordinale | Chaine de caractères | MTypeEtatSurface | Normal, Dangereux,  T Dangereux |
| **LibelleType** | Type des deux parties impliquées | Qualitative nominale | Chaine de caractères | MTypeImplication | Humain, Deux roues, Véhicule, Autre raison |

## Annexes :

**Annexe 1 :** Causes humaines, Circule sur le trottoir, Défaut de maîtrise, Demi-tour, Dépassement à droite, Dépassement dangereux, Dépassement en 3eme position, Dépassement en carrefour, Dépassement en virage, Dépassement interdit ou dangereux, Eblouissement par les phares, Ecart sur le côté, Eclairage insuffisant du véhicule, En intersection, En section, Entre ou sort de véhicule en stationnement, Entre sur la chaussée, Heurte un obstacle inerte, Heurte un obstacle mobile, Heurte un véhicule en stationnement, Heurte véhicule en stationnement interdit, Incident mécanique, Indéterminée, Infirme, Ivresse, Joue ou travaille sur la chaussée, Maitrise du véhicule, Malaise, Manœuvre sur parking, Marche arrière pour stationner, Marche sur la , haussée, Mauvais positionnement (changement de file), Non-respect des panneaux d’interdiction, Non-respect des signaux, Non-respect du piéton en carrefour, Non-respect du piéton en section, non-respect du stop, Non-respect d'une balise, non-respect feux tricolores, Non-respect priorité à droite, Non-respect , Priorité à droite (avec feux clignotant), Non-respect priorité de face, Ouverture d'une portière, Perte de contrôle, Piéton, Piéton descendant d'un t.c., Piéton montant dans un t.c., Queue de poisson, Quitte le stationnement, Roule à gauche, Roule en marche arrière, Sens interdit, Stationnement, Tourne, Droite interdit, Tourne à gauche interdit, Traverse hors passage, Traverse sans précaution, Va stationner à gauche.

**Annexe 2 :**

59009, 59013, 59017, 59044, 59051, 59056, 59090, 59098, 59106, 59128, 59143, 59146, 59152, 59163, 59173, 59193, 59195, 59196, 59201, 59202, 59208, 59220, 59247, 59250, 59252, 59256, 59275, 59278, 59279, 59281, 59286, 59298, 59299, 59303, 59316, 59317, 59320, 59328, 59332, 59339, 59343, 59346, 59350, 59352, 59355, 59356, 59360, 59367, 59368, 59378, 59386, 59388, 59410, 59421, 59426, 59437, 59457, 59458, 59470, 59482, 59507, 59508, 59512, 59522, 59523, 59524, 59527, 59550, 59553, 59560, 59566, 59585, 59598, 59599, 59602, 59609, 59611, 59636, 59643, 59646, 59648, 59650, 59653, 59656, 59658, 59660, 59670.

# Thème choisi :

## Problématique :

Nous souhaitons nous intéresser aux accidents qui se sont déroulés aux alentours d’une date particulière. Cela implique éventuellement une étude avec une durée variable en fonction de la date, puisque nous pensons par exemple que les fêtes de fin d’année peuvent avoir plus d’impact sur la conduite qu’un jour férié « classique » comme l’Assomption.

## Questions :

Afficher sur une année la moyenne du nombre d’accidents par jours. Le comparer avec la moyenne du nombre d’accidents sur toute la période entre 1984 et 1998. Un Δ peut être utilisé pour mieux mettre en valeur ces écarts.

Étudier les types de parties impliquées lors d’un accident des dates particulières, et le comparer avec des jours dits « normaux » ou peu influencés. On pourrait éventuellement étudier aussi la cause de ces accidents. S’intéresser à une période succédant un jour particulier peut être intéressant.

Comparer les heures « les plus à risque » de produire un accident un jour férié et les comparer avec les heures les plus risquées.