

TP2 : Manipulation et exploration d'une base de données

CIV8760 - Gestion de données en transport
Frédéric Chabot & François Bélisle

7 et 28 octobre 2022

Ce travail pratique vise à vous faire manipuler et explorer des données de transport à l'aide d'un logiciel de base de données. Ce travail vous permettra de mettre en pratique l'utilisation de logiciel pour la manipulation de base de données (SQLite ou similaire), d'extention spatiale (Spatialite), d'un système d'information géographique (SIG) tel QGIS ainsi que des logiciels permettant l'analyse de données (Python et/ou Excel).

1 Jeu de données

Les données utilisées sont les [collisions routières](#) survenues à Montréal entre 2012 et 2021. Il s'agit d'un sous-ensemble des données de la Société de l'Assurance Automobile du Québec (SAAQ), géo-référencées par la Ville de Montréal. Le jeu de données étant assez volumineux, il est fortement recommandé de commencer les analyses (tester vos requêtes SQL, comparer avec Excel, etc.) sur des sous-ensembles des données (par exemple, ne prendre que les valeurs de 2020 à l'aide de requêtes SQL). Notez cependant que pour répondre aux mandats, la base de données complètes doit être utilisée !

2 Mandats

Ce travail pratique est divisé en quatre étapes afin de mettre en pratique différentes compétences associées à la gestion des données en transport.

À noter que toutes requêtes SQL utilisées pour extraire les informations nécessaires doit être précisées dans le rapport. De plus, vous êtes libres d'utiliser l'outil que vous voulez pour faire les graphiques.

2.1 Modèle de données

Tout d'abord, il vous est demandé de discuter à savoir si le fichier de données disponible respecte les trois formes normales. Identifiez les entités auxquelles se rapportent les données du fichier, ainsi que leurs associations avec leurs cardinalités et fonctionnalités (les attributs ne sont pas demandés).

Plus précisément, résumez dans un (ou plusieurs) tableau à quelle entité correspond chaque variable. Au besoin, vous pouvez mettre ce tableau en annexe. À l'aide d'un schéma, présentez les caractéristiques du modèle E/A que vous proposez.

2.2 Analyse descriptives des données

Pour cette partie, il faut importer les données dans un logiciel de base de données (à l'aide du fichier csv disponible) pour en extraire l'information nécessaire. Veuillez présenter sous forme de tableau ou de graphique et commenter brièvement les informations suivantes :

- le nombre de collisions par an, par mois et par jour de semaine (incluant les jours de fin de semaine) ;
- la gravité des collisions (valeur numérique) par an, par jour de semaine (incluant les jours de fin de semaine) ;
- la gravité des collisions (valeur catégorielle) selon un facteur explicatif comme le genre de collision, le type de route, l'état de la chaussée, les conditions météorologiques, etc.

À noter que vous devez partir d'hypothèses (idéalement, provenant de la littérature) pour transformer les valeurs de gravité de catégorielles à numériques.

2.3 Analyse spatiale des données

Dans cette partie, une manipulation à l'aide d'un logiciel de base de données est aussi nécessaire. Cette analyse se fait au niveau des [limites administratives municipales](#) (entités municipales) dont le fichier est disponible sur le portail de données ouvertes de la Ville de Montréal. Il faudra aussi utiliser la [géobase du réseau routier](#) de la Ville de Montréal. Les questions suivantes peuvent ainsi être résolues avec un SIG (QGIS) ou un logiciel de base de données avec extension spatiale (Spatialite). En expliquant la méthode suivie, veuillez présenter sous forme de tableau ou de graphique (**pas de carte**) et commenter brièvement les informations suivantes :

- le nombre de collisions par secteur avec et sans prise en compte de la longueur du réseau routier (somme des polygones) par entité municipale ;
- la gravité des collisions (valeurs numériques) par entité municipale ;
- les implications (par exemple, nombre de blessés graves/légers) des usagers vulnérables par entité municipale.

2.4 Présentation cartographique de diverses analyses

Cette dernière partie vise à présenter certaines des informations précédentes sur une carte. Veuillez présenter et commenter brièvement :

- la distribution spatiale des accidents sous forme de carte de chaleur ([tutoriel](#) en anglais) en spécifiant les paramètres utilisés (noyau/“kernel” et rayon) ;
- la distribution spatiale des accidents reflétant leur gravité sous forme de carte de chaleur ;
- le nombre de collisions par entité municipale en prenant en compte la longueur du réseau.

Chaque carte doit inclure les éléments essentiels pour permettre une bonne compréhension.

3 Livrable

Ce travail se fait avec les mêmes équipes qu’au précédent travail. Veuillez contacter le chargé de laboratoire en cas de problème. Un rapport au format PDF ou Word, ne dépassant pas **15 pages**, doit faire état des mandats de ce travail pratique. La date de rendu est le **10 novembre 2022 à 23h59**. Le fichier doit être déposé en format électronique sur moodle.

Une attention particulière sera portée à la rédaction (les fautes de français seront sanctionnées tout comme une mauvaise organisation générale du travail), comptant au total pour deux points (10%) sur la note finale.