

TP2 : Manipulation et exploration d'une base de données

CIV8760 - Gestion de données en transport
Frédéric Chabot & Nicolas Saunier

6 Octobre 2023

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'extension 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 concernent le fichier csv des [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 l'information nécessaire doit être précisées dans le rapport. Veuillez vous référer à l'annexe pour un exemple de format attendu. Vous ne pouvez pas mettre de simples captures d'écran de vos requêtes. Il est ainsi obligatoire d'utiliser SQL pour traiter/filtrer les données avant de les extraire pour en faire des figures ou des tableaux. Les cartographies devront être faites à partir du logiciel QGIS. Enfin, vous êtes libres d'utiliser l'outil que vous voulez pour faire les graphiques/tableaux.

2.1 Analyse descriptives des données

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

- le nombre de collisions par an, par mois et par jour de semaine (incluant les jours de fin de semaine). Il est important de faire au moins un graphique pour chacun (an, mois et jour) ;

- indicateur de gravité des collisions (valeur numérique) par an et par jour de semaine (incluant les jours de fin de semaine). Il est important de faire au moins un graphique pour chacun (an et jour) ;
- 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. (gravité en fonction de ...). Vous devez proposer au moins trois graphiques (trois facteurs explicatifs).

À noter que vous devez partir d'hypothèses (idéalement, provenant de la littérature) pour transformer les valeurs de gravité de catégorielles à numériques. Aussi, assurez-vous d'utiliser la variable de gravité des accidents de façon numérique et non de façon catégorielle lorsque c'est demandé.

2.2 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 partiellement 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 les éléments suivants :

- le nombre de collisions par entité avec et sans prise en compte de la longueur du réseau routier (somme des longueurs des polygones, donc un nombre de collisions par kilomètre) par entité municipale ;
- le nombre de collisions par entité avec prise en compte d'une autre variable ;
- la gravité moyenne des collisions (valeurs numériques) par entité municipale ;
- les implications (par exemple, nombre de blessés graves/légers) des usagers vulnérables par entité municipale.

Vous devrez donc procéder à l'intersection (autrement dit, une forme de "jointure spatiale") de certaines couches sur QGIS pour en extraire une base de données modifiées. Ensuite, vous devrez faire vos traitements (comme par exemple, "GROUP BY") à l'aide de SQL. Notez qu'il est possible de directement faire vos requêtes SQL dans QGIS, donc nul besoin d'exporter une couche depuis QGIS ! Notez aussi que la prise en compte du réseau routier signifie de calculer une nouvelle valeur qui est le ratio du nombre de collisions et de la longueur du réseau.

2.3 Présentation cartographique de diverses analyses

Cette dernière partie vise à présenter certains éléments précédents sur une carte. Veuillez présenter et commenter au moins l'un des trois éléments suivant (vous pouvez aussi proposer une autre carte si vous en jugez la pertinence) :

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

La carte doit inclure les éléments essentiels pour permettre une bonne compréhension. De plus, vous devez présenter l'ensemble de l'île de Montréal dans une carte, mais n'hésitez pas à ajouter des cartes dans lesquelles il y a un "zoom" sur certaines parties de la ville.

3 Modalités

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 **20 pages**, doit faire état des mandats de ce travail pratique. La date de rendu est le **26 octobre 2023 à 23h59**. Le fichier doit être déposé en format électronique sur moodle.

Le nom du fichier doit porter la nomenclature suivante : EQ{numéro d'équipe}_TP{numéro du TP}_{semestre d'étude (A, H ou E)}{année d'étude}. Par exemple, "EQ01_TP1_A23".

Une attention particulière sera portée à la rédaction (les fautes de français seront sanctionnées tout comme une mauvaise organisation générale du travail), comptant pour un point (5%) sur la note finale.

Veuillez consulter le [Guide de rédaction pour ingénieur civil](#) disponible sur Moodle à la section Ressources.

4 Annexe

Exemple de format de requête accepté :

```
SELECT * FROM thisTable GROUP BY uneColonne
```

Au besoin, pour la section 2.2, vous pourriez utiliser d'autres fichiers de données concernant les arrondissements (entités) tel que le recensement. Sur le portail de données ouvertes de Montréal, vous pouvez retrouver celui de [2011](#).