

POLYTECHNIQUE MONTRÉAL

TRAVAIL PRATIQUE 2

---

# Exploration et analyse avec une base de données

---

*CIV8760 : Gestion des données  
en transport*

*Chargé de TP :*  
Guillaume NEVEN

Automne 2024



# 1 Introduction

Ce travail pratique vise à vous faire manipuler, concevoir et explorer des données de transport à l'aide d'un logiciel de gestion de base de données. Ce travail vous permettra de mettre en pratique l'utilisation de logiciels pour la manipulation de base de données (SQLite ou similaire) avec extension spatiale (Spatialite).

## 2 Développement d'un modèle de données [15]

L'objectif de cette partie est de concevoir un modèle de données pour une agence de location de vélos en libre partage tel que Bixi dont la gestion est centralisée. Tout déplacement est constitué d'un temps de départ  $t_1$  à une station  $s_1$  jusqu'à une station  $s_2$  à temps  $t_2$  avec un vélo  $v$  par un utilisateur  $u$ . Les entités suivantes doivent être représentées :

- Déplacement
- Utilisateur
- Vélo (incluez au moins : agglomération d'origine)
- Station (incluez au moins : latitude, longitude, agglomération)
- Agglomération

Votre modèle doit constituer un ensemble cohérent (qui ne manque pas un élément important nécessaire au fonctionnement essentiel du système). Ajoutez des attributs communs aux différentes entités (ex: age, type de pneu, etc.).

- a) Proposez un modèle de données sous forme de diagramme Entité/Association impliquant toutes les entités listées ci-dessus. Ajoutez des attributs (en indiquant les identifiants) et les associations entre entités avec leurs cardinalités, minimale et maximale, et les fonctionnalités. [10]
- b) Traduisez le modèle Entité/Association en modèle relationnel. Indiquez clairement les clefs primaires et externes (et ce à quoi les clefs externes réfèrent), et proposez des types pour les attributs. [5]

## 3 Analyse descriptive d'une base de donnée [25]

Pour cete seconde partie, vous allez devoir, en utilisant SQL, améliorer et manipuler une base de données.

### 3.1 Base de données [15]

Pour cet exercice, nous allons utiliser les données des trajets bixi de 2023-2024 disponibles sur leur site. Téléchargez et groupez les années 2023 et 2024.

- a) Comment pourriez-vous réduire la redondance de cette base de donnée ? [3]
- b) Implémentez vos changements. [5]
- c) Les stations de Bixi sont identifiées par un *Nom* et par leur position. Par quel attribut devrions-nous les regrouper ? Testez les deux façons et comparez-les. Commentez les différences, comment les résoudre et quelle méthode vous allez utiliser. [3]
- d) Créez une colonne de type spatial représentant la position du point. Cet élément doit être lisible par QGIS. [2]

Veillez fournir toutes les commandes SQL utilisées pour répondre à ces questions (copier-coller, pas d'image (copie d'écran)).

### 3.2 Exploration [10]

Veillez répondre à toutes les questions suivantes en utilisant le langage SQL. Quand un graphique est demandé, vous pouvez soit utiliser l'outil de DB Browser, soit transférer les résultats de la requête dans un autre programme.

- a) Créez un graphique montrant les nombres de trajets journaliers en fonction du temps. [1]
- b) Créez un graphique montrant les nombres de trajets moyen par jour de la semaine. [1]
- c) Trouvez la station avec le plus grand nombre de trajets (somme des arrivées et départs), et suggérez une explication.[1]
- d) En utilisant QGIS, créez une carte de chaleur qui représente le nombre de trajets partant/arrivant à chaque station. [1.5]
- e) Sélectionnez une station de départ de votre choix, et créez une carte de chaleur des stations d'arrivée. Commentez. [1.5]
- f) Regardez les dix paires les plus communes: que remarquez-vous? Expliquez et donnez une explication. [2]
- g) Sélectionnez une paire ayant beaucoup de déplacements (top 10%), puis créez un graphe montrant le temps de trajet moyen pour la paire en fonction de l'heure. Faites le une fois pour les jours de semaine, et une fois pour les fins de semaine. [2]

## 4 Soumission du rapport

Ce TP se fait individuellement. Veuillez ne pas dépasser la longueur de 10 pages pour le rapport et garder à l'esprit qu'il s'agit d'un exercice.

Veuillez soumettre le rapport au format PDF d'ici au 24 octobre à 23h59 sur Moodle. Assurez-vous que le rapport est exempt d'erreurs grammaticales et comprend des graphiques précis et des explications claires. Des points seront retirés pour les erreurs de rédaction et les inexactitudes dans l'analyse des données. Si un modèle de langage est utilisé (par exemple, ChatGPT), vous devez divulguer son utilisation précise.