

TRAVAIL DE SESSION – ÉTUDE DE CAS MATH 60611

1. Instructions générales

- Ce travail compte pour 35% de la note finale
- Il y a trois parties, avec les dates limites suivantes :
 - o 11 mars 2025 : partie 1 (rapport)
 - o 20 avril 2025 : partie 2 (diapositives)
 - o 20-25 avril 2025 : entrevue orale (dates à définir avec la professeure)
- Les remises de documents doivent se faire sur Zone cours
- À faire en équipe de 2 personnes.
- Les données à utiliser pour ce projet se trouvent dans le fichier *data_final.csv*. Un dictionnaire des variables du fichier est joint.

2. But pédagogique

Le cours MATH 60611 est un cours avancé. À ce stade de votre parcours, vous devriez avoir les outils requis pour résoudre des problèmes réels et proposer des solutions innovantes dans un contexte non académique. Le but de ce projet est de vous mettre dans une telle situation, et de vous exposer à un cas dont la problématique est définie de façon générale par des décisionnaires n'ayant pas de connaissances spécifiques en science des données. Ce projet va aussi vous donner l'opportunité de travailler avec des données spatiales, quelque chose que vous n'avez probablement pas étudié dans vos autres cours à HEC.

3. Description du cas

Vous êtes un consultant engagé par la Ville de Montréal dans le cadre d'un projet pilote en sécurité routière. Face à l'augmentation constante du nombre d'accidents impliquant des piétons sur son réseau routier, la ville a décidé de réaliser une étude permettant d'évaluer le niveau de risque des intersections routières. Pour ce faire, un échantillon de 1864 intersections signalisées a été sélectionné (voir figure 1). Cet échantillon représente 80% de toutes les intersections signalisées de l'île de Montréal. Pour chaque intersection, la ville dispose du nombre total d'accidents impliquant des piétons et rapportés par la police (données provenant de la SAAQ) dans les 10 dernières années. De plus, les ingénieurs du département des transports ont collecté les données suivantes :

- Données de trafic sur les flux moyens journaliers des piétons et véhicules à chaque intersection
- Données de géométrie (design) pour chaque intersection incluant le nombre de voies, types de signaux, aménagement pour les virages à gauche, restriction de stationnement, etc...
- Mesures de performances de sécurité routières calculées par les ingénieurs pour chaque intersection, basées sur le nombre d'interactions potentielles journalières entre des piétons et des véhicules, selon les directions et mouvements des véhicules.

L'objectif de votre mandat est de fournir à la ville un classement des 1864 intersections en termes de sécurité (des plus dangereuses aux moins dangereuses), afin qu'elle puisse prioriser les intersections les plus à risque dans le but d'améliorer les infrastructures.

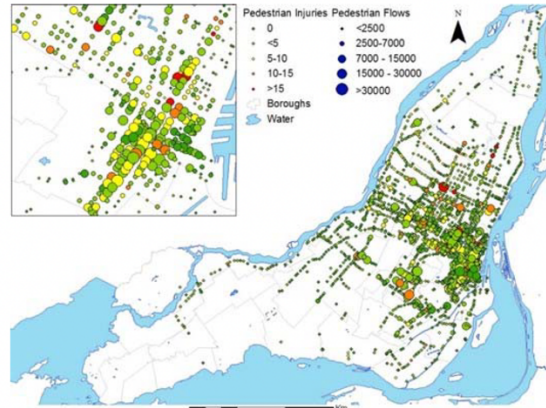


Figure 1 Signalized Intersections with Pedestrian Injuries and Flows

4. Objectif général

Votre analyse devra se baser sur une modélisation des accidents observés dans les 10 dernières années (variable *acc*). Vous êtes totalement libre de choisir les méthodes de votre choix, il n'y a pas une réponse unique à ce problème et plusieurs solutions sont adéquates. Faites preuve de créativité, de profondeur dans vos analyses, et de rigueur dans l'implémentation. Les données que vous recevez sont fiables, mais c'est à vous de choisir l'information pertinente que vous voulez utiliser.

5. Objectifs spécifiques et livrables

PARTIE 1 [50%]

Répondez à la question de recherche en utilisant les méthodes vues en classe (sélection de variable, régression non linéaire, méthodes basées sur les arbres, boosting), en ignorant la corrélation spatiale qui est présente dans les données.

Les livrables pour cette partie sont :

- Un rapport de 8 pages maximum (pas de code) décrivant précisément votre démarche
- Un fichier contenant votre code R nettoyé pour reproduire les résultats présentés.
- Un fichier .csv contenant 2 colonnes : le ID de l'intersection et son rang en termes de risque (un rang de 1 représente l'intersection la plus dangereuse).
- Trois diapositives (ppt) résumant à la ville (audience non technique) les principaux résultats de votre analyse.

PARTIE 2 [50%]

Répondez à la question de recherche en utilisant cette fois des méthodes spécifiquement développées pour les données spatiales, ou des méthodes qui permettent de prendre en compte la corrélation spatiale entre les intersections. Pour cette partie, vous aurez besoin

d'explorer la littérature sur l'analyse de données spatiales et fournir une solution innovante. Notez qu'il existe de multiples façons d'aborder ce problème et je m'attends à ce que chaque équipe fournisse une solution unique.

Les livrables pour cette partie sont :

- Un rapport détaillant votre stratégie d'analyse, les méthodes utilisées et vos résultats **sous un format de diapositives** (ppt ou Latex). Ces diapositives doivent être conçues pour être lues par la professeure, et donc doivent contenir un niveau de détail suffisant pour être compris sans une présentation formelle (vous pouvez prendre comme modèle les notes de cours fournies par la professeure). Le rapport devrait contenir au moins 30 diapositives. Il n'a pas d'instructions spécifiques sur la taille de police, mais les diapositives devraient être préparées comme si elles allaient être projetées sur un grand écran devant un large auditoire.
- Un fichier contenant votre code R nettoyé pour reproduire les résultats présentés.
- Un fichier .csv contenant 2 colonnes : le ID de l'intersection et son rang en termes de risque (un rang de 1 représente l'intersection la plus dangereuse).

6. Évaluation

La partie 1 sera évaluée sur les critères suivants :

- Profondeur de vos analyses
- Justesse de votre démarche générale
- Utilisation adéquate des méthodes
- Clarté du rapport
- Pertinence de vos analyses pour la ville de Montréal
- Clarté et pertinence des résultats présentés à la ville (diapositives ppt)
- Reproductibilité de vos résultats : votre code doit pouvoir compiler sans erreurs et générer les rangs présentés dans les livrables.

La partie 2 sera évaluée basée sur les diapositives fournies ainsi qu'une **entrevue orale** selon les critères suivants :

- Compréhension des méthodes utilisées
- Profondeur de vos analyses
- Utilisation adéquate des méthodes
- Clarté des diapositives
- Reproductibilité de vos résultats : votre code doit pouvoir compiler sans erreurs et générer les rangs présentés dans les livrables.

Remarque 1 : l'entrevue orale aura lieu en équipe, mais la note donnée pour la partie 2 du projet sera individuelle

Remarque 2 : vous devrez vous enregistrer pour l'entrevue et choisir une date (premier arrivé, premier servi). La professeure enverra des instructions à ce sujet.