TERM PROJECT– CASE STUDY
MATH 60611A

Specific instructions
- This assignment counts for 25% of the final grade
- Your report should be submitted on Zone cours no later than March 12th at 11:55 p.m.
- To be done in a team of 2 students.

Pedagogical goals
The MATH 60611A course is an advanced course. At this point in your studies, you should have the tools to solve real problems and come up with innovative solutions in a non-academic context. The purpose of this project is to put you in such a situation, and to expose you to a case whose problem is defined in a general way by decision-makers who do not have specific knowledge of data science.

Description of the case
You are a consultant hired by the City of Montreal as part of a road safety pilot project. Faced with the constant increase in the number of accidents involving pedestrians on its road network, the city decided to carry out a study to assess the level of risk at road intersections. To do this, a sample of 1864 signalized intersections was selected (see figure 1). This sample represents 80% of all signalized intersections on the island of Montreal. For each intersection, the city has the total number of accidents involving pedestrians and reported by the police (data from the SAAQ) in the last 10 years. In addition, Department of Transportation engineers collected the following data:
- Traffic data on the average daily flows of pedestrians and vehicles at each intersection
- Geometry data (design) for each intersection including the number of lanes, types of signals, layout for left turns, parking restrictions, etc…
- Road safety performance measures calculated by engineers for each intersection, based on the number of potential daily interactions between pedestrians and vehicles, depending on the directions and movements of the vehicles.

The objective of your mandate is to provide the city with a ranking of the 1864 intersections in terms of safety (from the most dangerous to the least dangerous), so that it can prioritize the riskiest intersections with the aim of improving infrastructure.
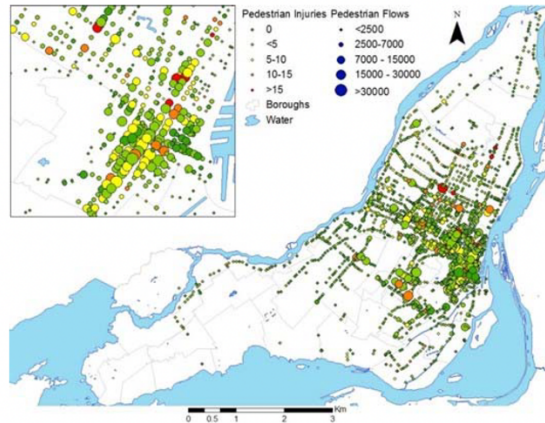
Figure 1 Signalized Intersections with Pedestrian Injuries and Flows

Specific objective
Your analysis should be based on a modeling of the accidents observed in the last 10 years (variable *acc*). You are completely free to choose the methods of your choice, there is no single answer to this problem and several solutions are adequate. Demonstrate creativity, depth in your analyses, and rigor in the implementation. The data you receive is reliable, but it is up to you what relevant information you want to use.

Data
The data is in the *data_final.csv* file. A dictionary of file variables is attached.

Deliverables
- A report of 8 pages maximum (no code) precisely describing your approach
- A file containing your cleaned R code to reproduce the results presented.
- A .csv file containing 2 columns: the ID of the intersection and its rank in terms of risk (a rank of 1 represents the most dangerous intersection).
- Two slides (ppt) summarizing to the city (non-technical audience) the results of your analysis.

Assessment
You will be evaluated on the following criteria:
- Depth of your analyzes
- Accuracy of your general approach
- Appropriate use of methods
- Clarity of the report
- Relevance of your analyzes for the city of Montreal
- Clarity and relevance of the results presented to the city (ppt slides)
- Reproducibility of your results: your code must be able to compile without errors and generate the ranks presented in the deliverables.