Case Assignment: Toronto Road Safety

Rotman MMA On-Campus Datathon 2018

Case

The City of Toronto is Canada's most populous city and a focal point of development and growth in Canada. As one of the most populous municipalities in North America, significant efforts are being made across all levels of government to manage the challenges posed by Toronto's increased urbanization. This case will aim to data and analytics to address some of the challenges posed to the City of Toronto.

Geotab

To address these challenges, more and more cities have been turning towards better use of data, analytics and most recently nascent Internet of Things (IoT) technologies. Geotab Inc. is a leader in GPS fleet management solutions collecting over 2 billion data points per day from sensors installed in road vehicles across North America, including Toronto.

Geotab's collects a wide variety of IoT data including GPS telematics on hyper-local weather data, detailed accelerometer data, and even cell phone coverage. Geotab's advanced analytics (https://data.geotab.com/) provides higher-level datasets from their detailed sensor data. For this case, they have provided (see detailed description in the next section):

- **Hazardous Driving Area Dataset**: Identifies areas for which harsh braking and accidents have occurred based on aggregated onboard sensor data for vehicles that have their solution installed.
- **Road Impediments Dataset:** Identifies areas with potential road impediments by aggregating their z-axis (up/down) accelerometer data.

Note that both data sets are aggregations over the vehicles (and fleets) using Geotab GPS devices in a given location.

One of Geotab's initiatives involves a Smart City program where their datasets can be utilized to increase safety, mobility, efficiency and citizen engagement.

Toronto Police Service

The Toronto Police Service is the first municipal police service created in North America and among one of the oldest in the English-speaking world. It is also one of the leaders in data analytics with their Business Intelligence and Analytics Unit supporting operational and strategic initiatives at TPS. They also maintain an open data portal (http://data.torontopolice.on.ca/) which is intended to both improve transparency and confidence as well as improve public safety in Toronto.

- One key strategic area for the TPS is traffic safety as it relates to public awareness and enforcement. Recent data on traffic collisions resulting in serious injuries or death have been released publically by the Toronto Police Service in order to collaborate broadly on approaches to traffic safety and to use data and analytics to drive decisions:
  - **Toronto Police Services Killed or Seriously Injured Dataset**: includes all traffic collision events in the City of Toronto where a person was either Killed or Seriously Injured (KSI) from 2007 – 2017.

This dataset includes specific information about each KSI incident over the specified period. While this makes the KSI dataset more specific with respect to the most serious types of incidents compared to Geotab's Hazardous Driving Areas Dataset (which relates to all incidents of harsh breaking – many of which may not have resulted in accidents at all), it is also, by necessity much sparser geographically, since the number of KSI incidents is quite limited. It is hoped that by combining the two datasets new insights on the areas with high potential for traffic accidents may emerge.

City of Toronto data

Traffic safety around schools, hospitals, places of worship, parks and other buildings with heavy public usage are of particular concern. To help you identify these areas two data sets maintained by the City of Toronto are available:
- **City of Toronto One Address Repository**: the authoritative address database, which includes exact geographical location of each city address, as well as the name, where it exists, of the location (e.g., church, school, park, etc.)
- **City of Toronto Centreline Intersection Dataset**: identifies all intersections in the city of Toronto

Problem Statement

The main objective of this case is to understand the factors that affect traffic safety in the City of Toronto utilizing the two datasets provided by Geotab and the Toronto Police Service. The aim of the project is to extract insights from the data and provide specific recommendations to the Toronto Police Service to increase traffic safety.

Some of the managerial questions you may wish to examine include:
- Are there relationships between locations with high number of harsh breaking incidents, road impediments, and KSI incidents? To what extent are these relationships predictive? How can they be used in pro-active accident prevention initiatives?
- Is there evidence that the relationship between KSI incidents and the factors measured in Geotab data sets has been changing over time?
- Are there "sensitive" areas (e.g. in close proximity to schools, hospitals or other public buildings) that are particularly risky with respect to more serious accidents?

Outcomes from this datathon can potentially help TPS Business Intelligence and Analytics Unit with several initiatives related to safety and traffic enforcement:

- Examine public safety data through new lenses
- Speeding Reduction Campaigns
- Red Light Camera Locations
- Vision Zero Safety Measures Enforcement & Collaboration
- Service Gaps
- Leveraging IoT datasets to gain further insights

Submission

Your submission should be in the form of a presentation to the Toronto Police management addressing some or all of the questions above. The presentation should be limited to 10 pages and non-technical in nature; it can be followed by an Appendix containing more technical descriptions of your work. While you can also submit your code in a separate file, the presentation should be self-contained (i.e., it should not be necessary for the reader to open your code).

The presentation will be judged on:
- Managerial relevance of your work
- Soundness of your analysis
- Effectiveness of you communication, particularly data visualization
- Actionability of your conclusions and recommendations.

Datasets

There are five datasets for this case. Please see the accompanying "Data Dictionary" document and "Data Exploration" notebooks.

- Geotab Hazardous Driving Area Dataset (https://data.geotab.com/urban-infrastructure/hazardous-driving)
- Geotab Road Impediments Dataset (https://data.geotab.com/urban-infrastructure/road-impediments)
- Toronto Police Services Killed or Seriously Injured Dataset (http://data.torontopolice.on.ca/datasets/ksi)
- City of Toronto One Address Repository (https://portal0.cf.opendata.inter.sandbox-toronto.ca/dataset/one-address-repository/)
- City of Toronto Centreline Intersection Dataset (https://portal0.cf.opendata.inter.sandbox-toronto.ca/dataset/centreline-intersection/)

Hints/Suggestions

For examples on how to manipulate the data along with some examples, please see the "Sample Analysis" notebooks.

Here are some sample questions that you may want to address:

- Which are the most accident/hazardous-prone locations?
- What factors affect accidents/hazards-prone locations (e.g. proximity to certain buildings, roadway features, vehicle types, road impediments, elevation)?
- Which accident types are most important to focus on? (e.g. vehicle, speeding, visibility, time of day, injury age etc.)
- How do certain segments compare to the average for the ward, city or surrounding areas?

General Tips:

- Do NOT try to bite off too much: it is MUCH better to focus on one specific question and investigate it in some depth than to obtain limited insights and results on a wide range of topics.
- Perform exploratory analysis to find a particular problem to focus on. Most likely, no actionable insight can be extracted at the entire city-level.
- Use a combination of exploration, segmentation and modelling to come to a relevant conclusion.
- Make use of the geographic data to derive insights by visualizing it on a map (see sample analyses).
- Ensure your presentation is non-technical but backed with solid data analyses.
- Keep in mind that your primary goal is not to build many analytical models and. Rather it is to provide actionable model-backed insights and recommendations.
- There is no requirement to use all data available to you. Only the data relevant to your chosen analytical focus should be used.
- Keep data limitations in mind. Each dataset you are provided with is limited in some way. While these limitations do not make the data useless (all real-life data has some limitations), they are important to keep in mind when arriving at conclusions and recommendations.