Analysis to Determine

Impact of Venues on Collision Occurrence in the City of Ottawa

Sonia Singh
March 22, 2020

Image: Canadian Flag projected on the city of Ottawa's Parliament Building

Abstract

This report summarizes a data science-based analysis of Ottawa collisions to determine whether or not there is a co-relation between collision occurrence and nearby venues. A co-relation would allow city planners to take these into consideration when approving and designing roads and intersections near specific types of venues.

Table of Contents

1 Introduction

1.1 Background

Collisions represent a significant impact in major cities and a large number of studies have been undertaken to analyze causes of collisions with a view towards actions that can be taken to reduce the number of collisions.

Many of these studies observe that collisions are impacted by a variety of factors including but not limited to:

- Road conditions
- Weather conditions
- Traffic volume
- Driver behavior
- Driver distractions

Some of the factors are relatively well studied, and it is understood that driver behavior and distracted driving are strong contributors towards collision occurrences.

1.2 Gap

Amongst published papers, there is currently minimal analysis of the potential impact of nearby venue types such as restaurants, coffee shops, bars, or stores on collision occurrences.

Given that distracted driving is considered to be a strong factor contributing to collisions, the types of venues in specific locations could be impacting driver distraction.

My hypothesis is that the number of collisions at each location have a reasonable probability of being co-related to the number of types of venues at each location. For example, the number of venues related to food, liquor, or shopping could impact the driving behaviors such as individuals being distracted or being in a rush due to their visits to these venues.

1.3 Focus of this Analysis

Collisions in the Ottawa area will be analyzed to determine whether or not such a co-relation exists. If this analysis indicates that collision occurrences are impacted by nearby venues, this

will be of use for city planners when approving and designing roads and intersections near specific types of venues.

2 Data

2.1 Data Sources

The analysis required both collision data and venue data which was obtained from the following sources:

1. The City of Ottawa's open data site includes 2018 tabular transportation collision data. The dataset includes a tabular list of all the collisions that occurred within the Ottawa area in 2018.
    a. Accuracy of the dataset:
        - Each reportable collision occurring on public roadways is sent to the City of Ottawa and is validated at least once.
        - Approximately 50% of the records are validated once again by a senior staff. Additionally, many queries are run on the database looking for errors.

    b. Features included in the dataset:
        - X and Y coordinate format is projected in MTM Zone 9, NAD83 (CSRS)
        - KML and CSV/XLS formats are projected in latitude, longitude (WGS84)
        - Date
        - Time
        - Location description (RD1 @ RD2 or RD from RD 1 to RD 2)
        - Classification of collision (non-fatal, fatal, property damage only)
        - Collision location (Intersection, non-intersection, at/near private driveway)
        - Pedestrians involved
        - Road surface condition (Ice, wet, dry snow...)
        - Environment (Clear, rain, snow…)
        - Light (daylight, dawn, dusk…)
        - Initial impact type (Angle, turning movement, rear-end…)
        - Traffic control (stop, traffic signal, no control…)
        - Latitude and longitude

    c. Location of dataset:

- https://open.ottawa.ca/datasets/ec6f9c7d3e214fcebd8d97111f1804df_0?geometr y=-77.066%2C45.217%2C-74.542%2C45.890

2. Foursquare API offers real-time access to their global database of rich venue data and can be used to extract venue data for specific geographical locations. The "explore" endpoint was used to extract a list of venues near each collision location.
    a. Features included in the dataset that were used for this analysis:
        - Venue name
        - Venue category
        - Venue location in latitude and longitude

3. GoogleMaps was used to obtain latitudes and longitudes for specific venues that were investigated.

2.2 Data Cleansing

- Collision data had a feature called "TRAFFIC_CONTROL_CONDITION" with missing data. Since this feature included a value of "00 – Unknown", all missing data was set to "00 – Unknown".
- Venue data had a feature called "Venue category" which had a high level of granularity, and included values such as "Restaurant", "Japanese Restaurant", "Noodle House", etc. This level of granularity was too high, so I created a new feature called "Venue type" to allow grouping at a lower level of granularity, such as grouping all of the above examples into "Eatery". A total of 238 venue categories were grouped into 16 venue types for ease of analysis.

2.3 Feature Selection

Collision data features included in the analysis:

1. The City of Ottawa's open data site includes 2018 tabular transportation collision data. The dataset includes a tabular list of all the collisions that occurred within the Ottawa area in 2018.
    - Location description (RD1 @ RD2 or RD from RD 1 to RD 2)
    - Classification of collision (non-fatal, fatal, property damage only)
    - Collision location (Intersection, non-intersection, at/near private driveway)
    - Pedestrians involved

- Road surface condition (Ice, wet, dry snow...)
- Environment (Clear, rain, snow…)
- Light (daylight, dawn, dusk…)
- Initial impact type (Angle, turning movement, rear-end…)
- Traffic control (stop, traffic signal, no control…)
- Latitude and longitude

2. Venue data features included in the analysis:
   - Venue type, which is derived from rolled up values of venue category
   - Latitude and longitude

2.4 How the Data is Used

The collision data includes 14485 incidents at 5585 locations. Since FourSquare places a limitation on the number of calls that can be made in a day and in an hour, a subset of the collision data is used for analysis instead of using the full dataset.

A subset of the collision locations close to the Ottawa city center are selected for analysis. For each of these locations, corresponding venue data are obtained to create a dataset that can then be analyzed.

3 Methodology

4 Results

5 Discussion

6 Conclusion

7 References