Third Party Signs and Traffic Accidents

Shawn Mills - 304420

Research Question

Does the presence of third-party signs affect the volume of traffic accidents in Toronto?

Literature Review/Background

The concept of distracted driving is certainly far from a new phenomenon. For as long as vehicles have been on roads, there have been visual distractions for those behind the wheel. These distractions can come in many forms, including other vehicles, pedestrians, animals, advertisements, music, and most recently cell phones. While there have been an abundance of news stories and awareness campaigns towards the dangers of cell phones and texting, discussions surrounding the effects of excess signage are currently minimal by comparison.

The city of Toronto enacted its Sign By-law in 2010. Under these by-laws, signage is divided into two categories:

- First-party signs – those signs that are used to identify the business or service at the location
- Third-party signs – those signs that are used to advertise goods or services that are not related to the businesses on premises

Some examples of articles addressing signs as a driving distraction include:

The role of roadside advertising signs in distracting drivers - https://www.researchgate.net/publication/222691471_The_role_of_roadside_advertising_signs_in_distracting_drivers

Digital signage is a distracted driving hazard - https://www.myparkingsign.com/blog/digital-signage-distracted-driving/

Data and Description

Datasets being considered for this analysis include the following:

Third Party Signs Inventory: https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/business/#9b6f952e-52d7-1fc4-51f6-4ad6bc913218

- This dataset will provide the locations of all licensed third-party signs in the city

Toronto Police Service Public Safety Data Portal:
http://data.torontopolice.on.ca/datasets/automobile/data

- This dataset will provide details on the traffic collisions throughout the city

SimplyAnalytics:

https://simplyanalytics.com/

- Data from this site may prove beneficial in determining the populations within each region of Toronto.

Twitter (time permitting):

- Extracting and analyzing text relating to signs or billboards in Toronto may be used to determine if the public views these signs as entertaining, informative, or distracting

Other data required but not yet found:

- Identification of each zone in Toronto (residential vs. commercial)
- Kilometers of roadways within each FSA region in Toronto

Proposed Methodology

Geospatial analytics:

Referencing the following tutorial -
https://www.datacamp.com/community/tutorials/geospatial-data-python

Investigate use of the following python libraries:

- Geopandas
- RTree
- GDAL
- Fiona
- Shapely
- PySAL

Using the above, sign locations and traffic accidents will be overlaid on the map of Toronto.

Text analytics (topic modeling/sentiment analysis – time permitting):

For Twitter data:

- Tweepy
- Search API

Using the above, tweets will be reviewed to determine sentiment and word frequency.

Timeline as assigned:

Week 1:

- Signup for GitHub, research python libraries and data APIs. Start developing a topic area/dataset for inquiry.
- I highly recommend not doing the boilerplate Kaggle projects. Look for something more exciting and local.
  - For example, rental prices scraped from Kijiji from the GTA
  - Twitter volumes and actor interaction during Raptors games in Toronto
  - News sentiment and topic analysis for the GTA
- **Come in with several ideas by first class**.

Week 2:

- Begin data collection, assembling methodology for Exploratory Data Analysis
- ETL work into data warehousing
- Begin exploring data and documenting issues/limitations/needs understanding needs/limitations regarding research question (might need to adjust question, scope, data, etc.).

Week 3:

- Submission of EDA sprint (via GitHub/Moodle)
- €Collect/Augment/Refine project according to Sprint 1 findings
- Develop several analytical methodologies (these should be methods you are interested in learning – like Topic modelling or sentiment analysis, or social network analysis, etc.)

Week 4:

- The halfway point.
  - All data collection should be completed
  - The methodology should be finalized

Week 5

- Submission of Sprint 2 to GitHub/.
    - Includes codebase, report (brief), and plan for analysis

Week 6

- Start final sprint prepare for final submission.

Week 7

- Work on final project.

Week 8

- Final Submission.
    - Final report – pdf/word doc
        - 3-5 pages of background, research question, methodology, results, conclusion
    - GitHub Repo
        - Well documented – meaning someone who comes across it randomly should know exactly what it is, how to re-create it, and what your repo contains
    - Final notebooks/analysis/results
        - Each sprint notebook documented, detailed, and refined.