Analyzing Traffic Collisions in American Cities

Group 16:
Mohammed Riaz, Jackson Weisner, Erica Hammond, Junha Liu, Prit Patel

mriaz5@uic.edu
Github:
riazmb01

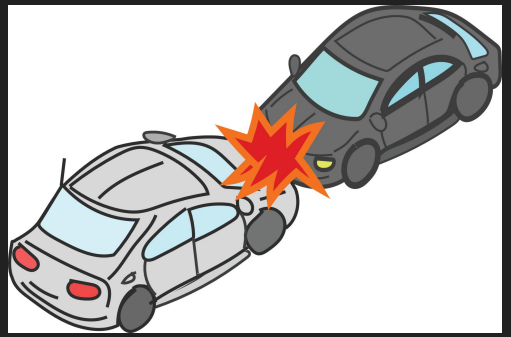
jweis4@uic.edu
Github:
jweis4

ehammo2@uic.edu
Github:
Hammond-EC

jliu238@uic.edu
Github:
junhaliu

ppate425@uic.edu
Github:
pppate425

The Problem



- What factors contribute to traffic collisions?
- We will explore three potential causes:
  - Weather
  - Time
  - Location
- We will explore the impact those potential causes in three cities (Chicago, New York, Los Angeles)
- It's important to understand how each city is impacted to better prepare drivers for risks
- Our results will also allow drivers to determine the safest conditions in their city to drive in

Relevant Data

- So far we have found traffic collision data in three major cities
  - Chicago
    - From the Chicago Data Portal, provided by the City of Chicago
    - Rows: 807K, Columns: 48
    - CSV file
  - New York
    - From NYC OpenData, provided by the NYPD (police department)
    - Rows: 4.15M, Columns: 25
    - CSV file
  - Los Angeles
    - Los Angeles Open Data, provided by LAPD
    - Rows: 604K, Columns: 18
    - CSV file

Expected Findings

- We will approach the problem as a general analysis of traffic incidents
- Proposed scope
  - Traffic accident data aggregated by cities
  - Data is assembled by State officials and parties of the incident
- Expected findings
  - High number of traffic collisions in locations that experience severe inclement weather and are highly populated. Particularly during winter and rush hour.
  - Exploring what times traffic collisions are most likely to occur
  - High-traffic areas will incur a higher number of incidents
  - Analysis of how/why certain infrastructure design can contribute to accidents
- Potential techniques
  - Bi-variate and multi-classification models, linear regression as well as visualizations of data insights (i.e., spatial visualization of incidents) - Significant cleaning and normalization would be a prerequisite for analysis
  - Sentiment analysis of testimonial data concerning the incidents

Envisioned system

- Our project will:
  - Determine what weather conditions cause more injuries in traffic crashes
  - Find the specific locations where collisions are more likely to occur
  - Find common times that collisions occur
  - Find high-risk areas where environmental hazards contribute to accidents
  - Spatial mapping and analysis of traffic accident location
- Most of the data will be presented statically, as well as interactive charts to drill down on specific data
- Models built off analysis techniques will display static charts and findings
- There will be interactive maps available which display the hotspots for traffic collisions, overlaid with weather data and other parameters that a User would be able to specify their visualized parameters
- Our first progress report will implement some static analysis and visualization of data using the techniques established (S.4)
- Included in that report will be our established dataset after all necessary cleaning and normalization of aggregate datasets

Credits

- New York dataset
  - https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Vehicles/bm4k-52h4/about_data
- Chicago dataset
  - https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if/about_data
- Los Angeles dataset
  - https://data.lacity.org/Public-Safety/Traffic-Collision-Data-from-2010-to-Present/d5tf-ez2w/about_data
- Image 1
- Image 2