Traffic Violations

—— in Maryland County

Group 11

Yichen Li
Yuyang Zhao
Shengqi Zhou
Yucong Hu

Outline

Identify dataset

Identity our dataset and describe the reasons we choose the dataset

.

Challenges with data

Concerns with the data and changes we expect to overcome

Process data

Acquire data, perform initial exploration and dimensional modeling

Business questions

Analyzing the data with three business questions.

Findings & Actions

Our group's business outcomes about data, and actions we implement

Part 1

Identify and describe your dataset

- Traffic Violations in Maryland County dataset
  1.04m record, 35 columns, 2012-2018, Kaggle
- Dimension: date ,hour, cause, personal information, consequence, law, vehicle, loaction
- Avoide future traffic risk, boost Social Welfare
- Diversity, comprehensive, related to daily life,nearby
- Tool: AWS jupyter, SQL, Python, R, Tableau, Googlemap

Issues:

1. Null values
2. Data contains 1.3 million records
3. Input errors -typo issues

Methods:

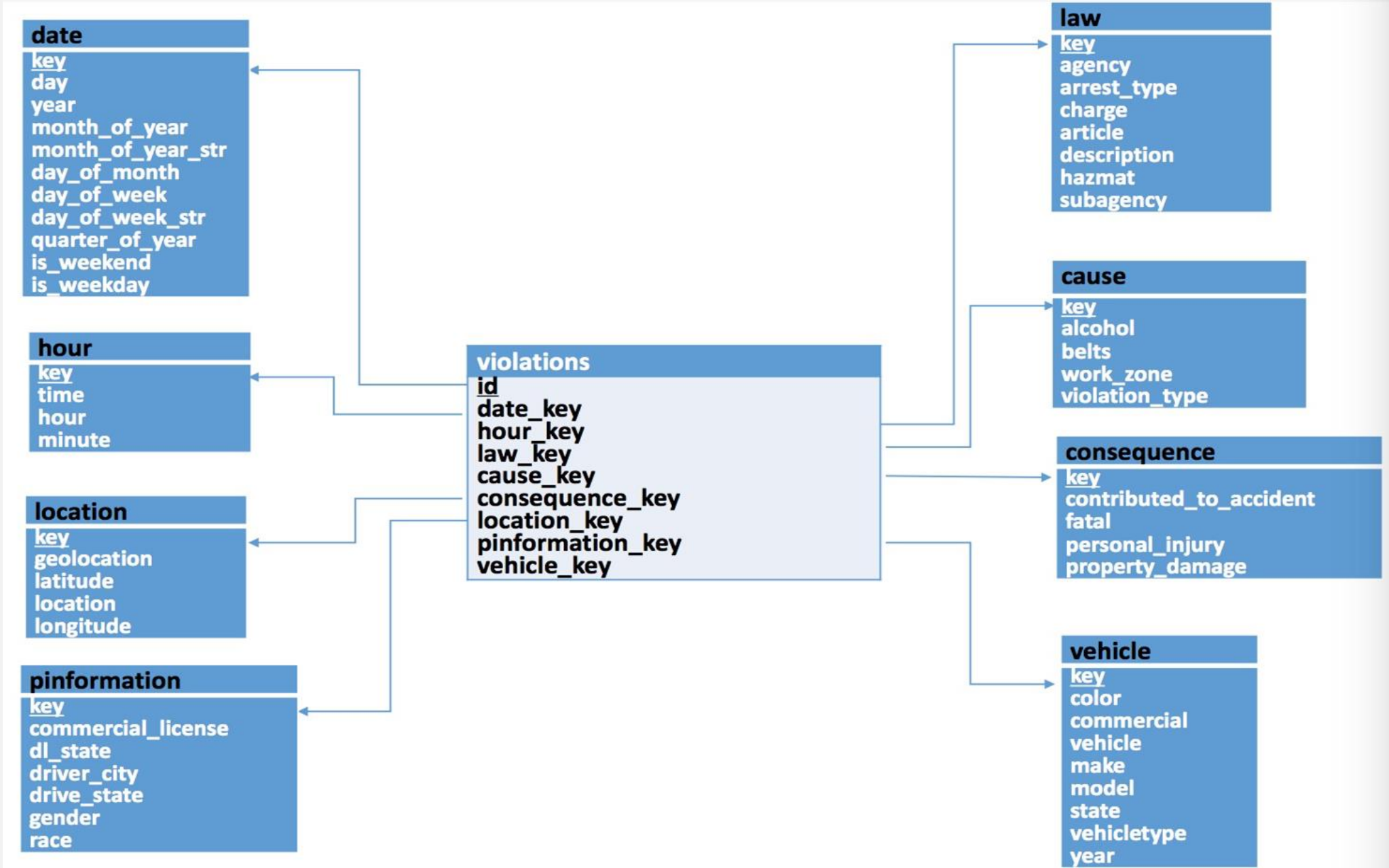
1. Remove & set default values
2. Filter
3. Case statement

Part 2-3

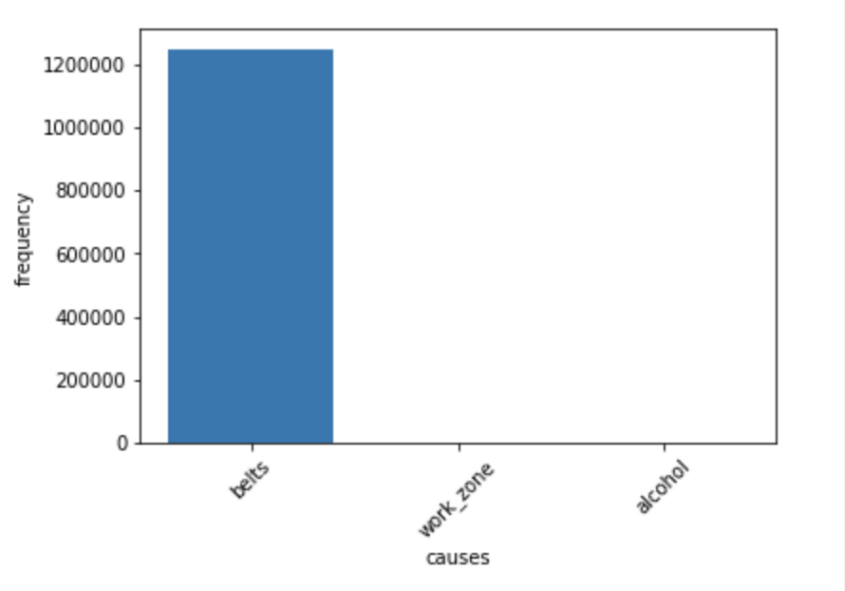
Challenges with data

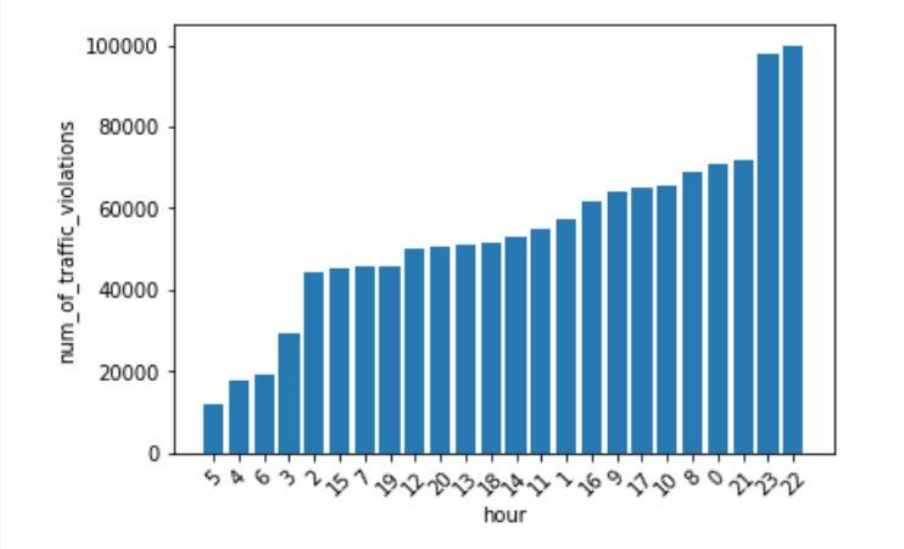
Data Wrangling and Data Cleaning

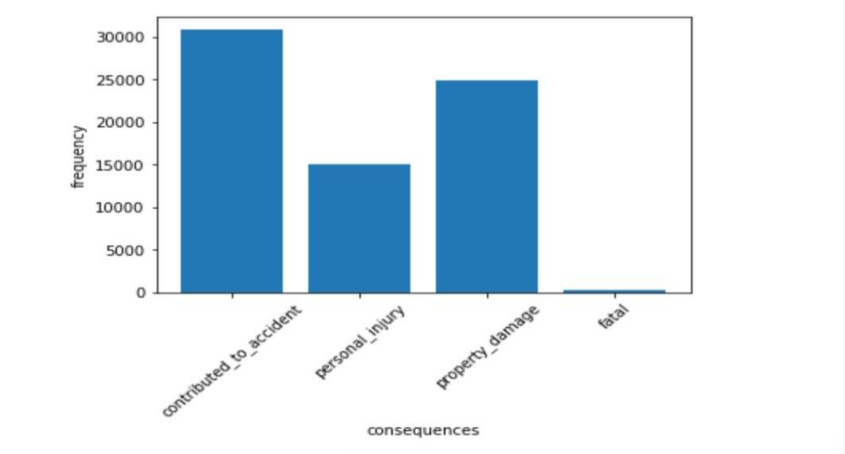
Creating the Relational Data Model

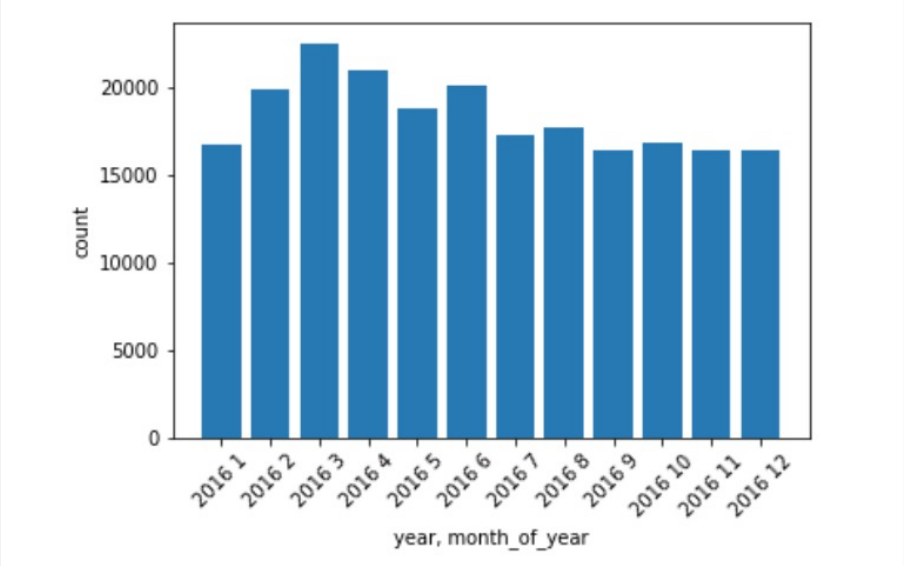


Pre-analysis









Part 4

Business contexts

Business Context

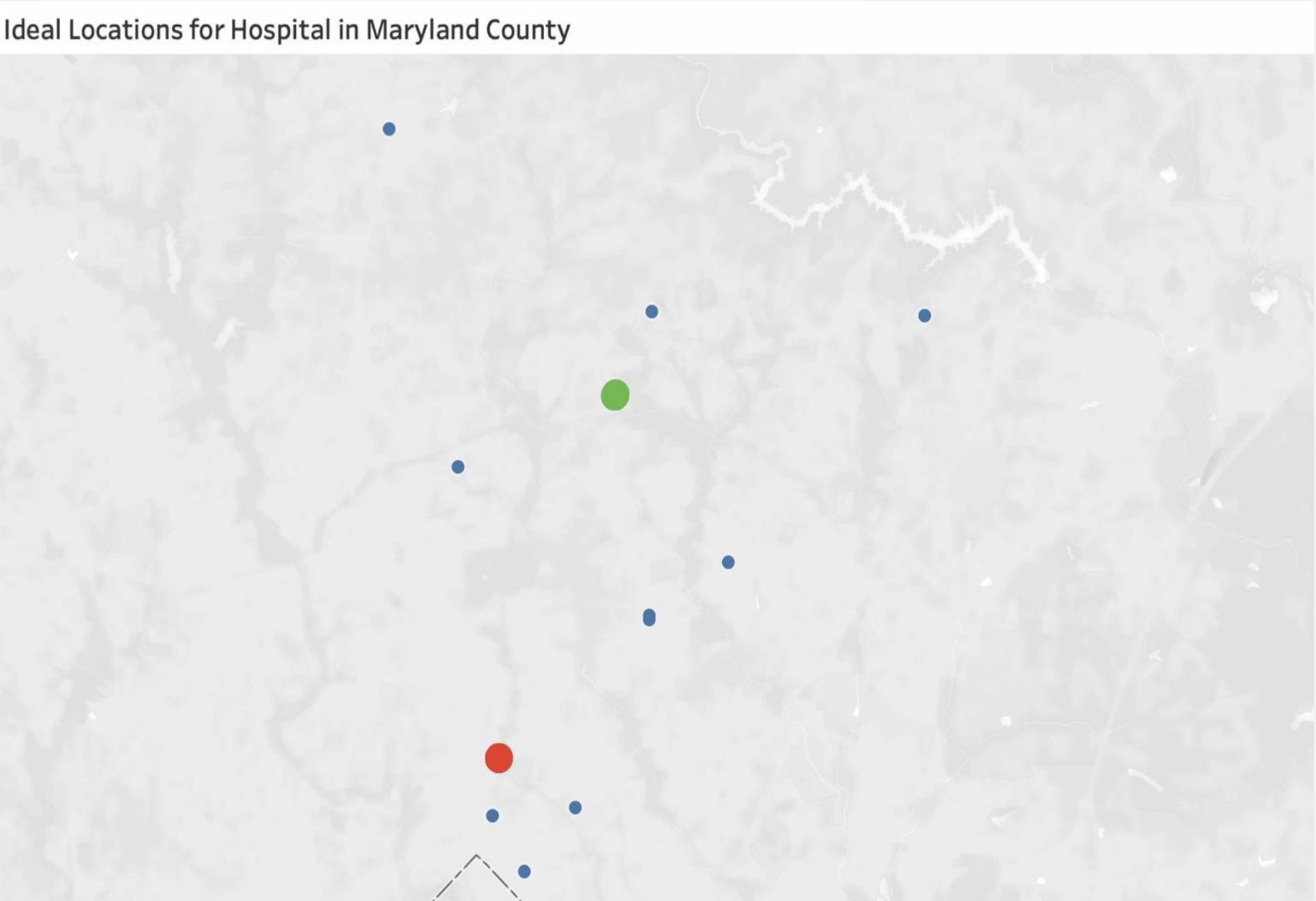
1. **If an investor wants to establish a hospital in the area we are researching, where would be an ideal location?**
2. **If department of transportation wants to decrease personal injury and number of violations in this area, which streets are the obvious ones the department should pay attention to and what actions could the department take?**

Top 10 Geolocations and Locations with high violation frequency

geolocation	count
(39.045425, -76.9907366666667)	551
(39.0462766666667, -76.990695)	431
(39.109775, -76.91044)	327
(39.0056183333333, -77.0123283333333)	268
(39.11061, -76.9897983333333)	253
(39.057555, -76.9678266666667)	210
(39.0039, -77.036485)	155
(39.077695, -77.046385)	128
(39.1492783333333, -77.06662)	128
(38.9920483333333, -77.0271333333333)	124

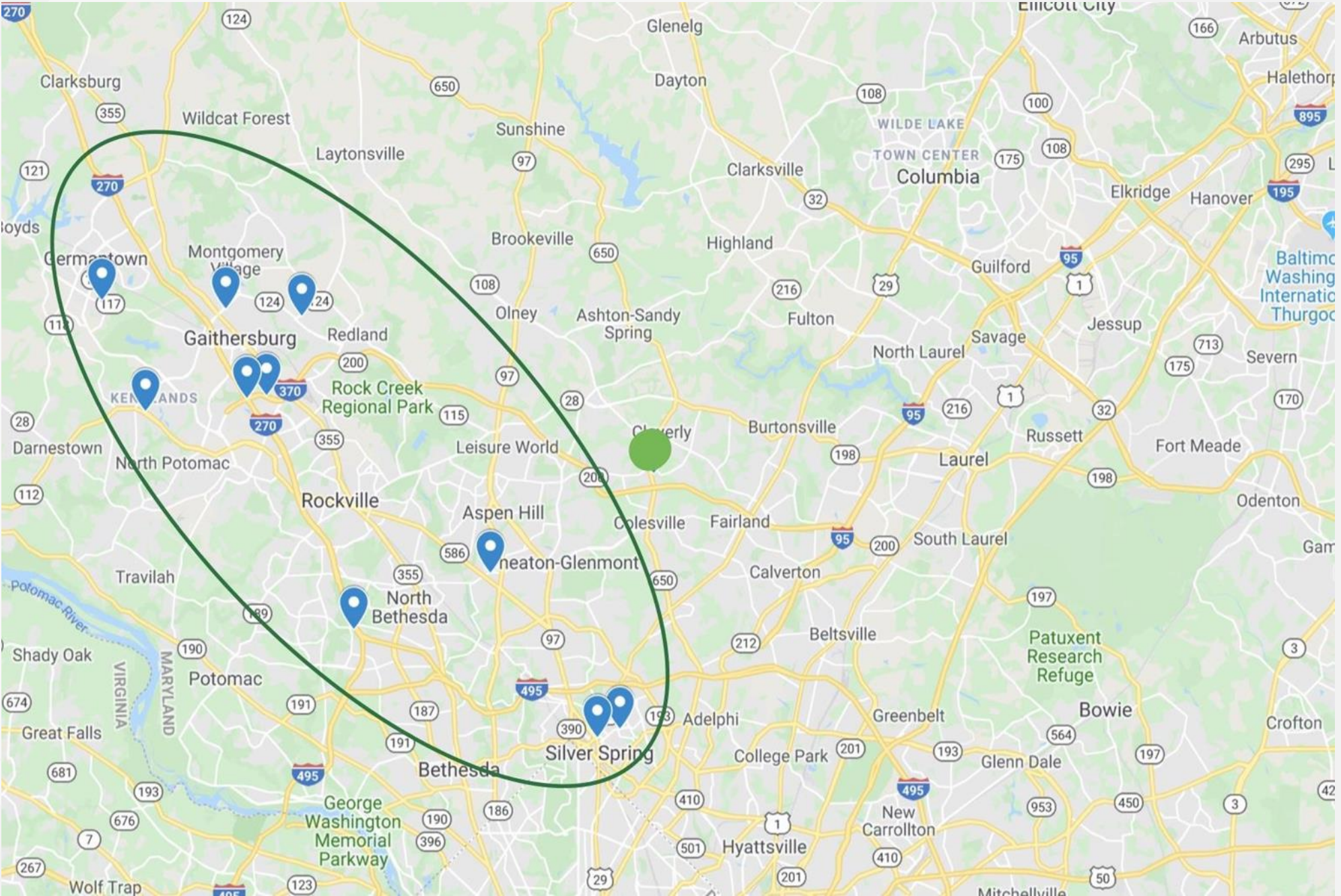
location	count
IS 370 @ IS 270	1981
WAYNE AVE @ DALE DR	1894
W/B IS 370 @ IS 270	1884
WOODFIELD RD @ EMORY GROVE RD	1584
RANDOLPH / COLIE	1494
10901 WESTLAKE DRIVE	1452
RT 28 @ BLACKBERRY DR	1443
MONTGOMERY VILLAGE AVE @ RUSSELL AVE	1360
CLOPPER RD E/B @ ORCHARD HILLS DR	1323
COLESVILLE RD @ GEORGIA AVE	1242

Map based on Geolocation (Points)



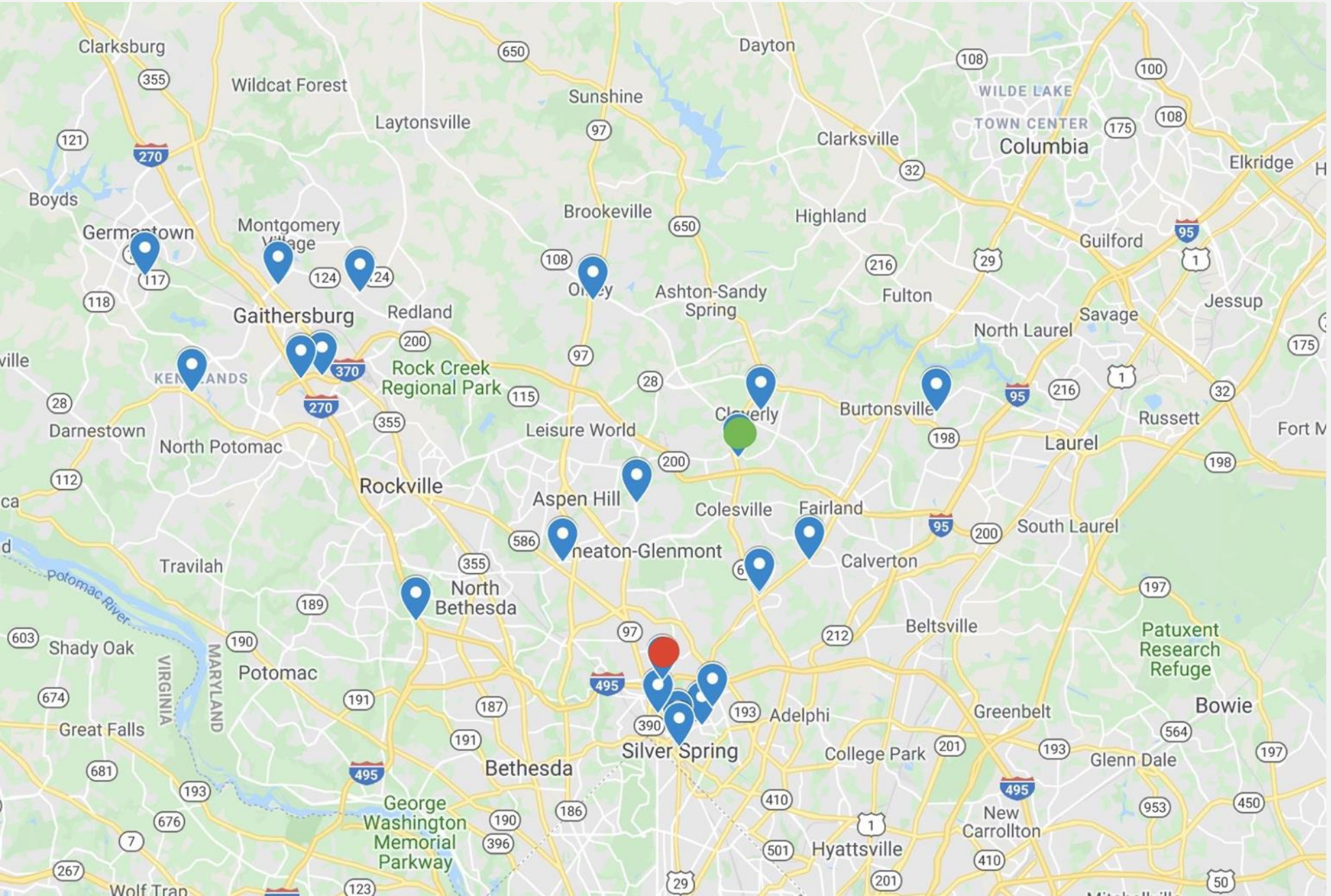
- Blue Points
  Highest violations frequency
- Red Point
  Ideal Hospital location 1 (Holy Cross Hospital)
- Green Point
  Ideal Hospital location 2

Map based on Location (Streets)



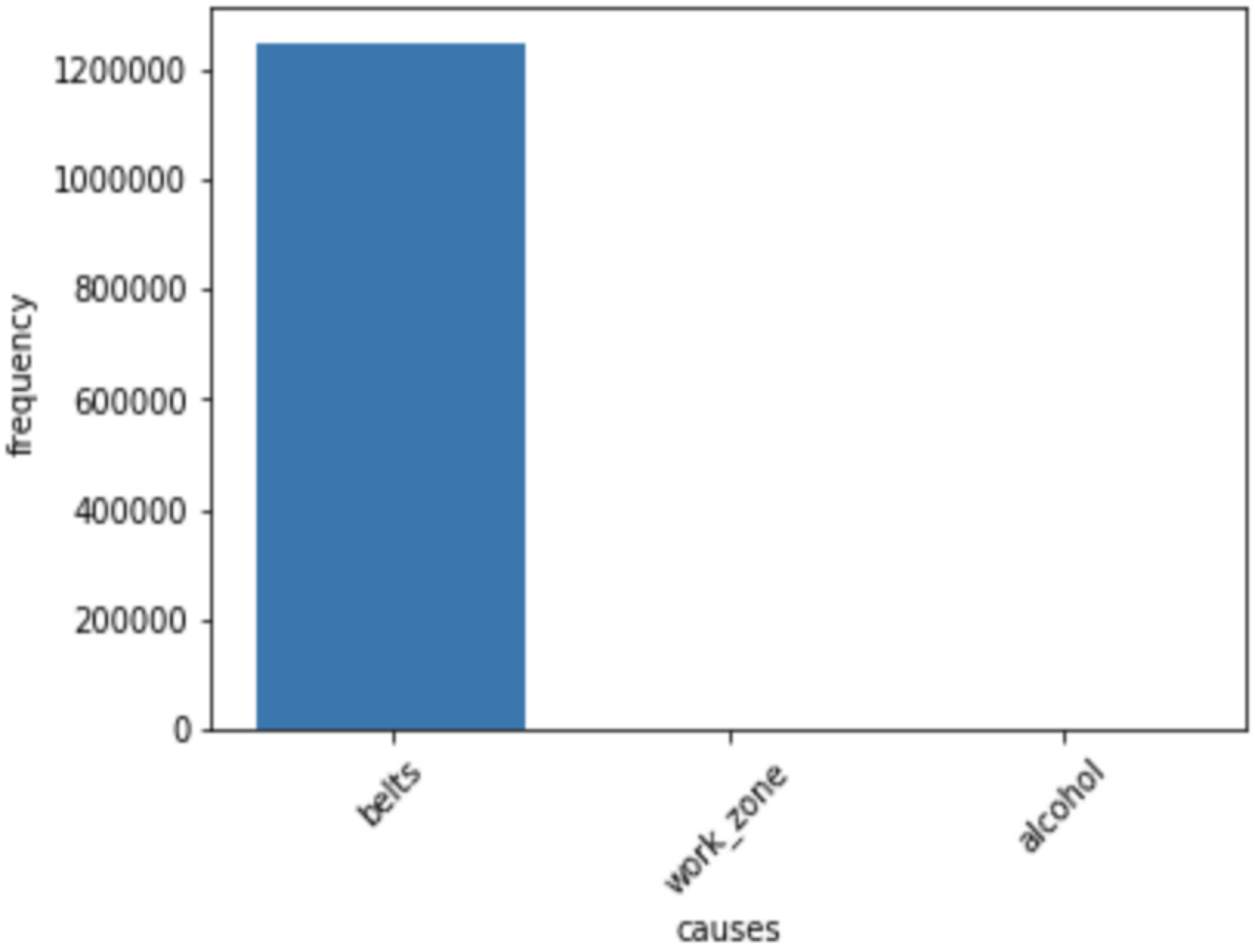
- Gaithersburg and Silver Spring
- Build stop signs
- Broadcast to remind drivers of traffic conditions

Combination of Geolocation and Location



- The margin of Maryland County with narrow road
- The crossroad with narrow width
- Downtowns with bumpy road

Causes of Violations



frequency	causes
1247956	belts
248	work_zone
2208	alcohol

For most of the violations, we found the drivers did not fasten their seatbelts

Part 4

Business contexts

Business Context

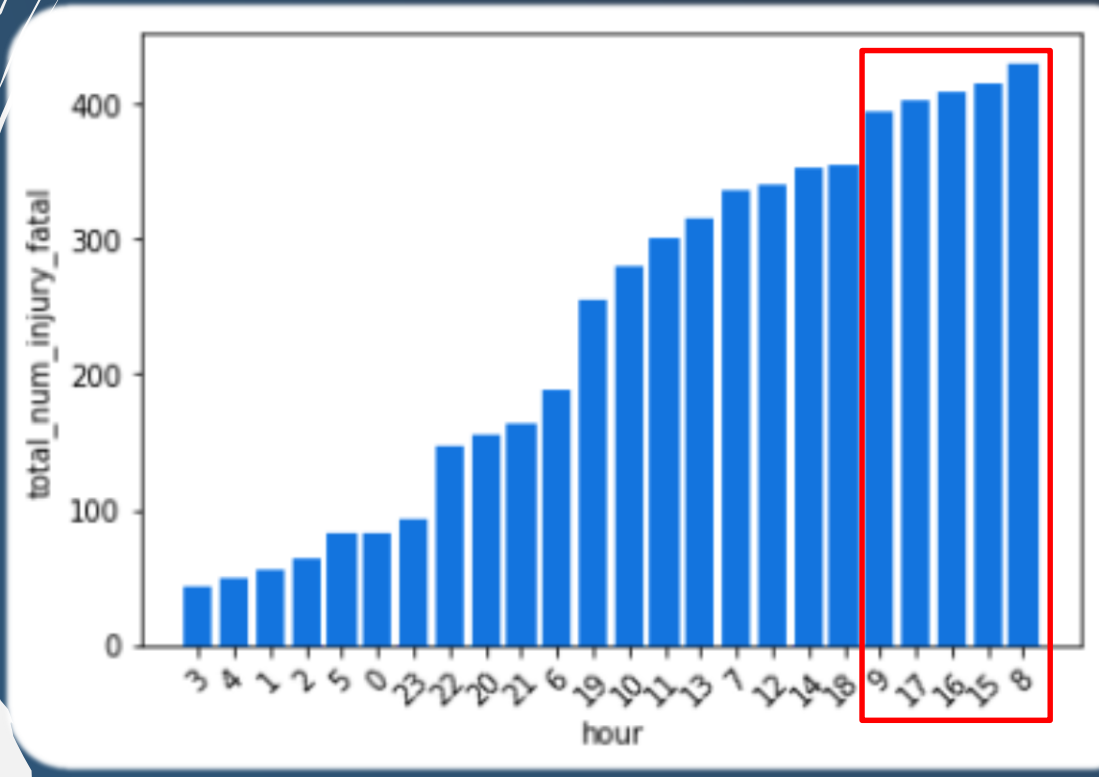
3. **If American Red-Cross want to arrange care workers to help people in traffic violations, how should they arrange resources?**
4. **If a company wants to buy vehicles for commercial use, what considerations should they take into?**

Business Context

- NGO and volunteers

Analysis

- SQL Query



Findings

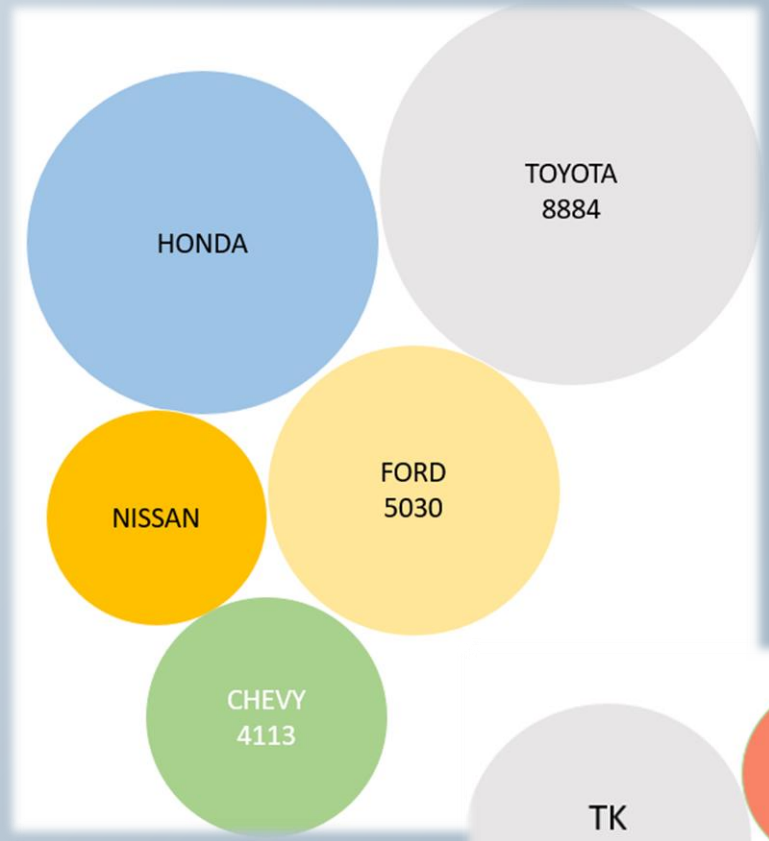
- Most likely occur in peak hours of people commute

Actions

- Set up emergency resettlement sites during peak hour of traffic
- Arrange care workers during 8-9 and 15-17

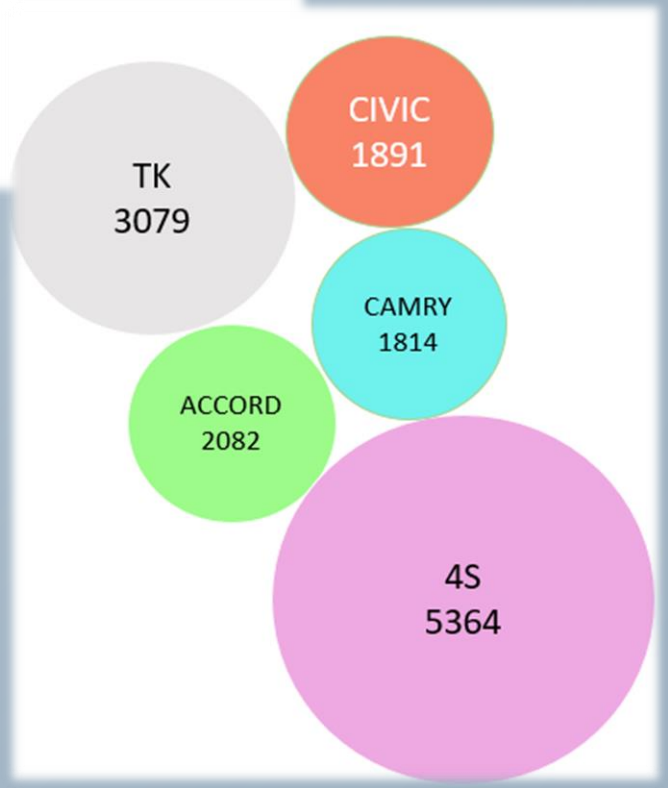
Business Context

- Vehicles characteristics
- Relationship with violations



Analysis

- Violations happened most in what kind of models and manufacturers



Findings

- Top 5 Model
- Top 5 Manufacturers

Actions

- Insurance plans

THANK YOU

Q&A

Attestation
All project team member contributed equally and meaningfully to the project and have completed knowledge on any part of it