
County Database of Traffic Violations

— Chris Hill, Ankhbayar Jansan, —
Rachel Lewis, Chase Whyte

Project Description

Our dataset contains traffic violation information from all traffic violations issued in the Montgomery County of Maryland. The data has been collected from 2012 to 2016, and contains identifiers like the time of stop, location, accident, injury (if any), alcohol and/or drug (if any), make/model of car, violation type, race, gender, and charge.

Some interesting questions to keep in “mine”:

What are the top five streets for issuing traffic citations?

What cars are most likely to commit traffic violations in this county?

Are there certain times of the year/day that crashes are more likely to occur?

How does the percentage of traffic violations issued based on ethnicity compare with the demographic distribution of the county?

Prior Work in Area

Extensive research into applying data mining to traffic violations

Usually to predict future crashes or improve road safety

Utilize K-means clustering to predict where accidents are most likely to occur

Datasets

Dataset: <https://catalog.data.gov/dataset/traffic-violations-56dda>

The data set is found on the official U.S government's open data website (data.gov). We currently have the data downloaded, with cloud stored copies also available.

Proposed Work

Scrub the irrelevant attributes (latitude, longitude, and types of police patrols)

Perform data reduction based on year the violation occurred

Mine for different association rules and correlations

Create bar plots and graphs to show various trends among the traffic data

Tools

Primary tool for programming will be Python

Excel to open and scrub dataset

Matplotlib to create the graphs we need

Python-based tool named Orange for mining text

Evaluate Results

Characterize people who commit traffic violations and the cars they drive in the Montgomery county of Maryland

Mine data trends relating to the different times of the year/day that traffic violations and crashes occur

Observe how number of traffic violations per-year changes

Utilize this information to predict where crashes are most likely to occur and what environmental and situational factors influence this

Questions