Traffic violations

Data Science Study

*Data Science Geeks*

Lissette

Bhavesh

Chris

Viral

## Categories

## 

1. ​Introduction​
2. Data Set​
3. Tools
4. Proposed Work/Motivation​
5. ​ Summary of ​Results​
6. Applications
7. References

## Introduction

Driving self-owned cars is the most used form of transportation. For most of us, it is a daily task. However, according to CDC, it is also the third highest cause for death. Loads of accidents happen year-round. As our dependency on cars increases, its important to gather and study information on traffic violations to find any trends/causes that would diminish those causes. Also important for car insurance companies so they can use that information to establish changes in premiums.

## Dataset

While we found the dataset in Kaggle, where in turn it points to this link to show where they got it from:

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

The dataset contains 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 description of charge.

## Tools

We used Python as the programming language. We removed some columns like latitude and longitude, color and make of the car, and the license/drivers state. Following step in data cleanup was made usable and scrubbed with few Python commands like dropna. The two most important Python libraries used for our study were Matplotlib, which will be used to create various charts and graphs to represent the data that has been mined, and pandas, which was useful for reading in and working with large data sets.

## Proposed Work/Motivation

Some of the questions we were curious about in this study were:

* + Who gets more tickets (Males v. Female)?
  + Which race tends to get more violation tickets?
  + Which month of the year most violations happens?
  + What time of the day most violations happens?
  + What are the most common types of violations?

## 

## Summary of Results

Note: These are our main takeaways, to see code snippets and plots, please refer to the power point presentation.

* Our data showed that many more warnings and citations were given to men than they are to women.
* About two-thirds of female traffic stops are for speeding, whereas stops for male driver are more balanced among other violation categories. This doesn't mean that females speed more often than males.
* For all other types of violations, the stop rate is higher for males than for females, disproving our hypothesis. The stop rate varies by violation type, and the difference in stop rate between male and female is because they tend to commit different violations.
* White and Black have higher amount of violations than other races.
* According to our data, about 10% traffic stop by race are Native Americans followed by Asian 9.6% in our list.
* White population has biggest count of traffic ticket violations in Montgomery County.
* May was busiest month for years 2012-2014, while it moved to April in 2015 and March in 2016.
* Lowest amount of tickets seems to be given early in the year (January or February) from 2012-2015. The exception was 2016 when December became the year with the lowest ticket count (Happy Holidays!).
* We see a consistent trend where the busiest time for giving traffic tickets is 10PM – 12:00AM. Meanwhile, 5:00 AM was the least busy time for traffic tickets throughout all years.
* More than half of all violations are for speeding, followed by other moving violations and equipment violations.
* The rate of Alcohol-related stops is very minimal and did not meet our top ten, disproving our hypothesis that alcohol played a significant role in speeding.
* Most of the arrest made were initiated by mark patrol cars.

## Applications

The data that has been collected is a good first step and can be expanded and analyzed in many ways. For instance, what traffic policies may be more effective than others. Does the amount of traffic cops affect the number of tickets issue and if so by how much? How long do their shifts last and at what times they switch? Are they being evenly distributed throughout the county?

Insurance companies could use this type of information and determine whether they need to make changes in requirements. Driving instructors could use this kind of information to see if they can make improvements in their teachings.

Is there a car model/color most likely to get tickets? Is there an area where accidents and personal injuries happen most often and if so, why? Local government agencies could use this sort of information to look at future improvements in the roads, if our current laws are still applicable without needing adjustment.

Another possibility is finding data available on the total amount of drivers in the county and determine if there is a chance of bias. Do males get more tickets because there are more male drivers or are they getting tickets at a higher percentage despite there being more female drivers? Similar findings could be made in terms of race.

## Reference

[1] Anon. 2017. Montgomery County of Maryland Traffic Violations. (March 2017). Retrieved March 2, 2017 from https://catalog.data.gov/dataset/traffic-violations-56dda