TRAFFIC VIOLATIONS

INSY 6500 – Class project

DECEMBER 7, 2018

Mohammad Maydanchi
Mumen Rababah

Table of Figures

List of Tables

Owning a car in the United States is considered as a popular thing as around 95% percentage of an American family own a car [1], and if you asked any driver about his big fear when driving he may while tell you that being pulled over by policeman. However, trying to reduce the probability of being stopped be was the purpose behind this project by choosing a dataset related to traffic violations then going through some deep analysis to define relationships between being pulled over and other factors.

The dataset that has been selected for this project was published Montgomery County Government [2], which includes more than a million records with 35 attributes, these attributes varies between stop description date and time, race, gender, car model, locations, etc.

The analysis process was done using R and python by asking several research questions which will understanding the relationship between different fields. But before dealing with this step a checking process on the data was done to figure out if it needs to be cleaned first.

Starting the analysis process for this dataset by generating a question which are related to the distribution of data on states as shown in Figure 1.

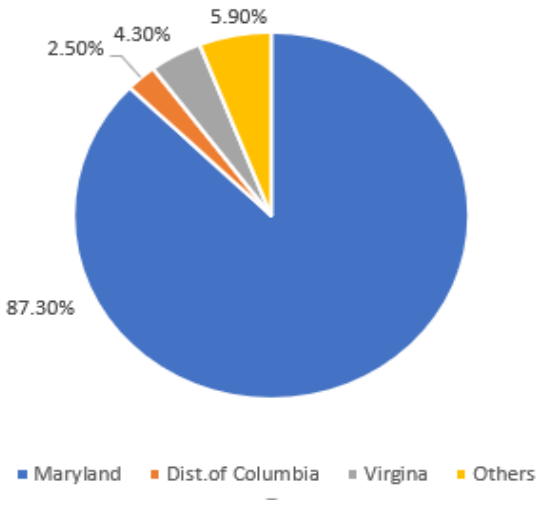

Figure 1: Distribution of stops on states.

Based on the previous figure its more reasonable to provide an analysis related to Maryland only as most of the records are there.

An expected question would be related to what are the most frequent causes for stopping? And most of us will expect that speed will be ranked first, but as you can see in Figure 2 that when a driver failed to obey properly to instructions that's on a placed traffic control device was ranked as the most frequent cause for being pulled over among more than 9000 different stops causes, and a simple example for this one is that when a driver ignore a no turn left sign.

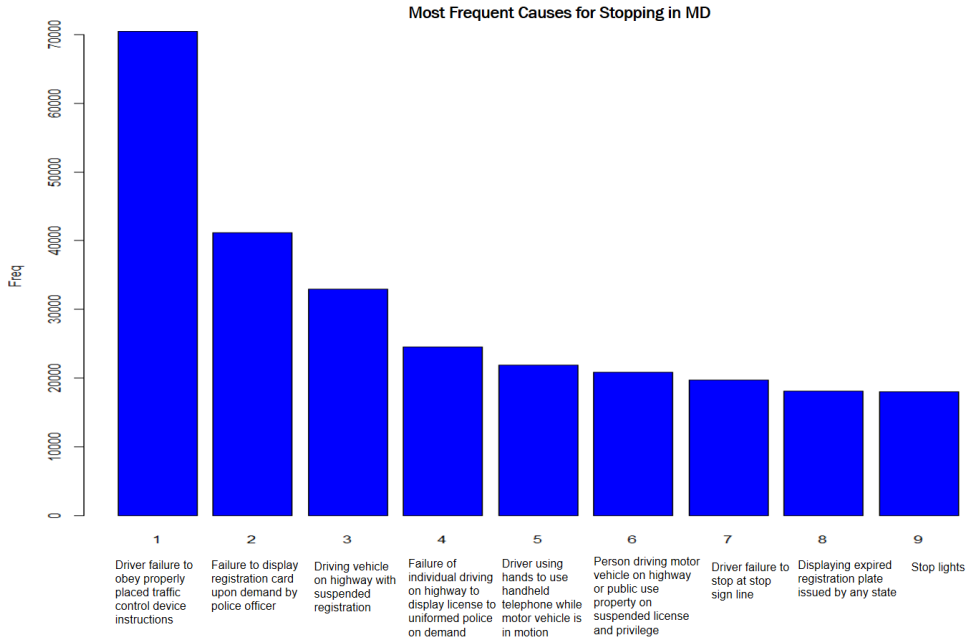

Figure 2: Most frequent causes for stopping in MD.

By having more deep analysis regarding time and date, first an analysis about the frequency of stops during each season was done as shown in Figure 3.

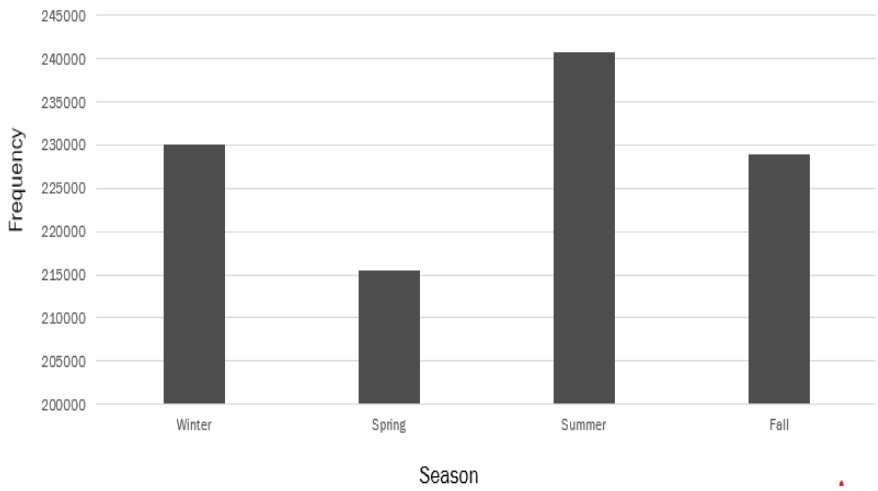

Figure 3: Distribution of stops based on season in MD.

Here, trying to define a relationship regarding number of stops in each season. As you can see in the results that Winter has 229,998 stops, Spring has 215,492 stops, Summer has 240,674 stops, and Fall has 228,868 stops. So, based on this a conclusion could be build that getting stops during Summer is more probable. Also, the frequency of stops for each month was calculated, and Figure 4 describes the distribution of stops based on month. The major finding was that during June the number of stops was the minimum.

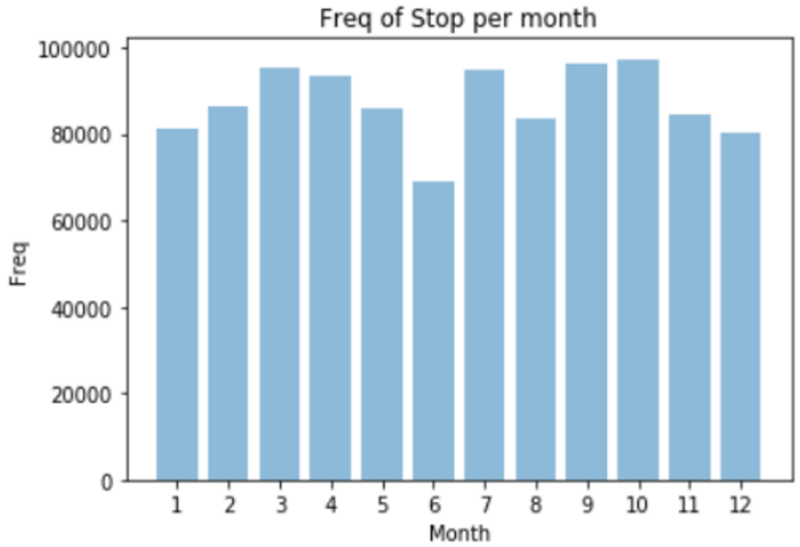

Figure 4: Frequency of stops per month in MD.

Moreover, it was questionable to know which hours of day have gotten most stops as shown in Figure 5. Notice that we assumed 6 am -12 pm as morning, 12-6 pm as afternoon, 6-pm -12 am as night and 12-6 am as midnight.

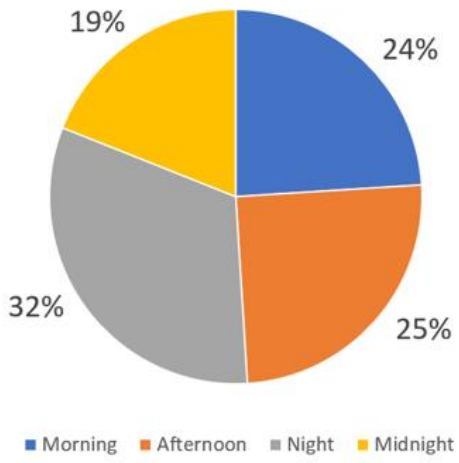

Figure 5: Frequency of stops based on time in MD.

Based on the above pie chart, the lowest percentage of stops was at midnight with 19%. Also, 25% of stops occurred during afternoon. But, most of the stops in Maryland occurred during night; It is reasonable as during night it is more possible for drivers to do not see signs. And, Even the number of drivers after midnight is supposed to be lower than anytime of a day but there is a good percentage of stops which was 19%.

A new research question was asked to know the number of different locations and which locations has the most frequent number of stops in Maryland, and to get that the below Figure 6 was developed which shows the most frequent location based on if that location has more than 750 stops, and the total number of different locations that was consider as stop location was 148,018 locations.

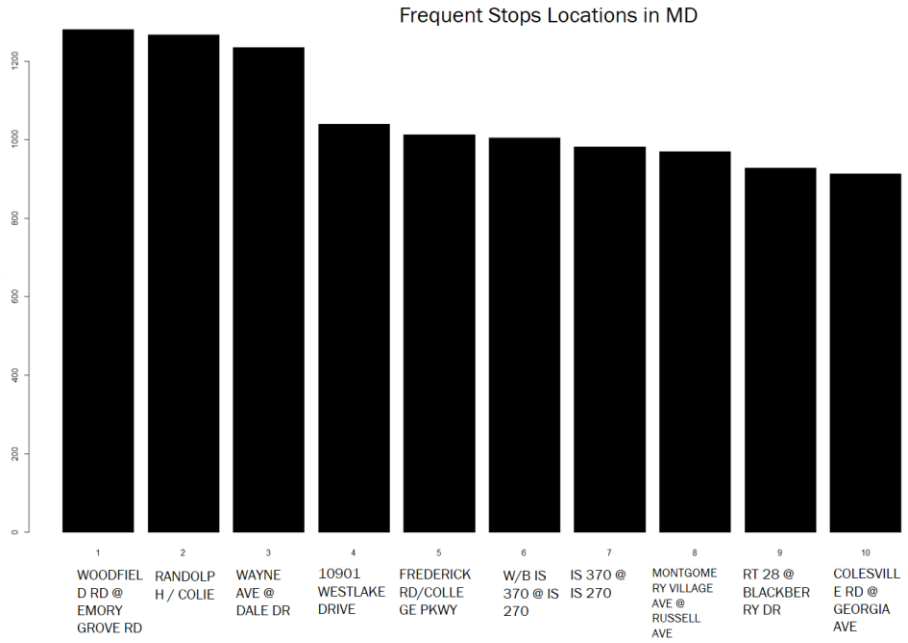

Figure 6: Frequent stops locations in MD.

Continuing with the location's analysis, an important information should be provided which about where are the locations with fatal cases? And the best answer of this one is to show all fata locations on the map as there are 101 locations with 212 fatal among the entire dataset. Figure 7 shows all these locations.

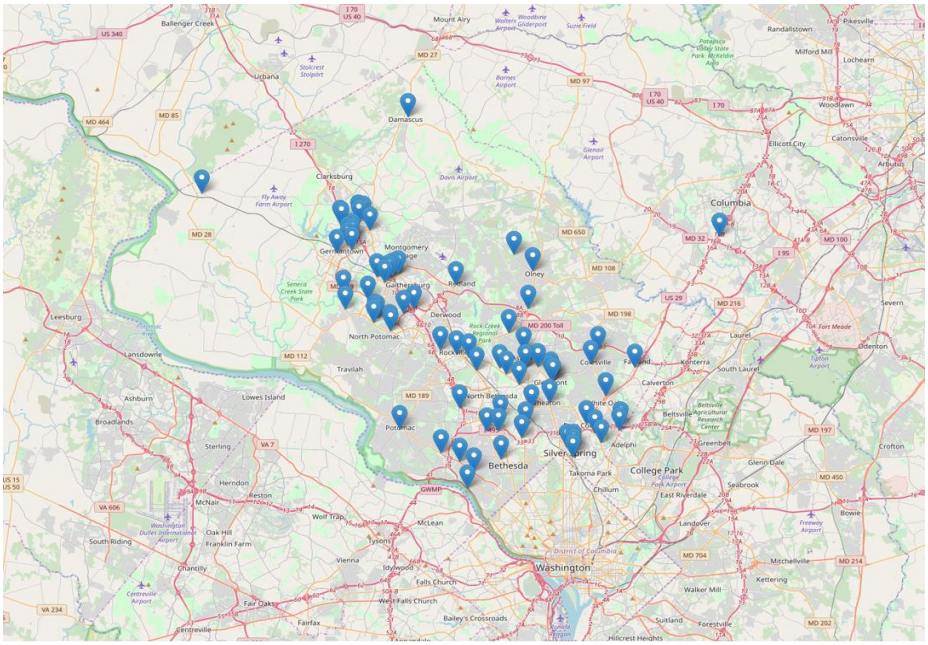

Figure 7: All fata locations.

One of the important questions that a lot of people is asking is if there is a relationship between alcohol and different attributes such as fatal, personal injuries, and properties damage, and the results were based on Table 1, 2 and 3 that alcohol doesn't has a significant effect on these attributes.

Table 1: Effect of alcohol on fatal.

Alcohol vs fatal	No	Yes
No	913271	212
Yes	1549	0

Table 2: Effect of alcohol on personal injuries.

Alcohol vs personal injury	No	Yes
No	902246	11237
Yes	1508	41

Table 3: Effect of alcohol on properties damage.

Alcohol vs property damage	No	Yes
No	894856	18627
Yes	1391	158

Another interesting relationship could be defined between arrest type and race. As you can see in Figure 8 the title of more than 82% of arrested types was 'A - Marked Patrol' and 'Q - Marked Laser' with 9.2% was in second place.

Between 2012 to 2018, 285,174 black people and 284,942 white people are arrested for A - Marked Patrol. And there is no significant difference between these 2 majority races for this type. By comparing other types between these two, the same results will be gotten as shown in Figure 9.

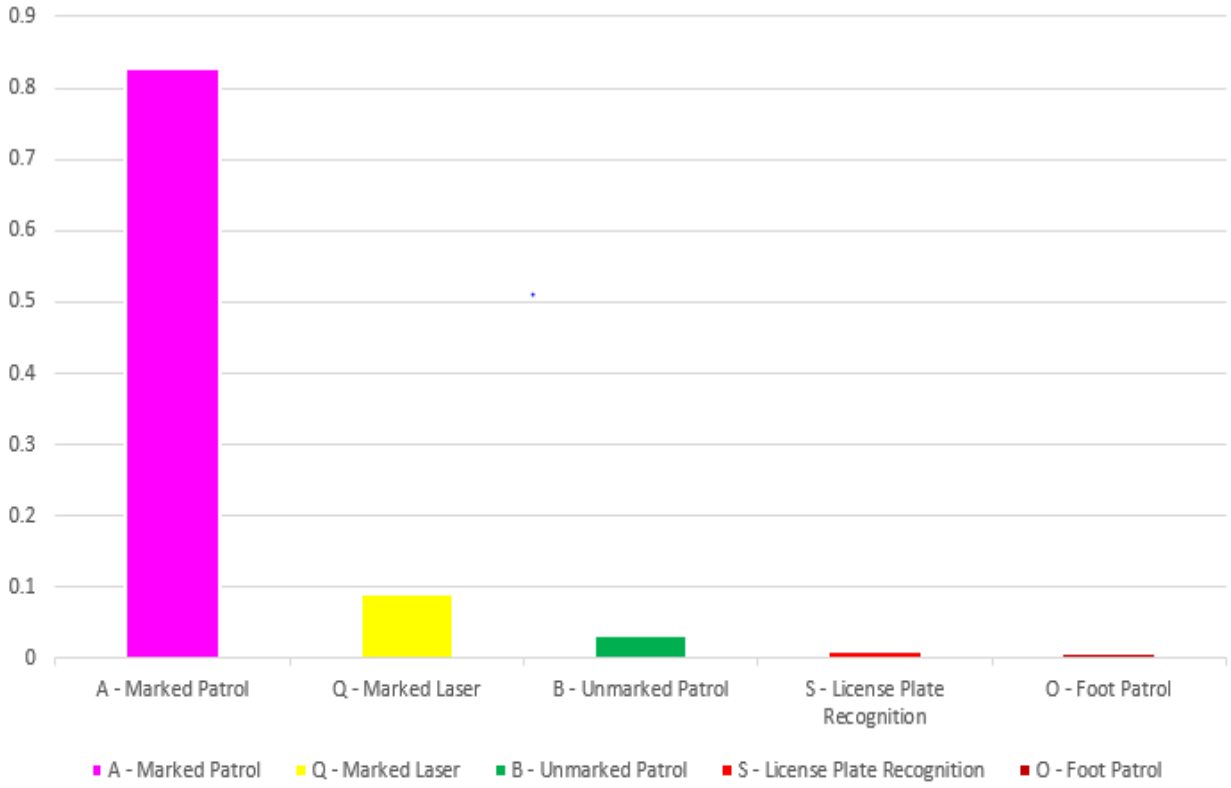

Figure 8: Most frequent arrest types.

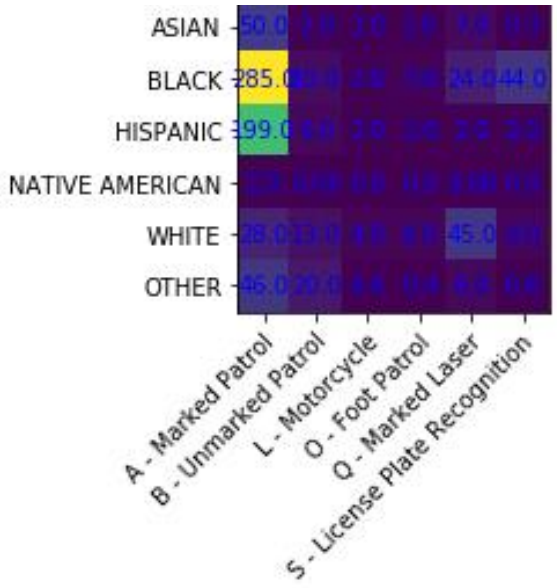

Figure 9: Relationship between Arrest Type and Race.

Caring about rules is different between men and women, so for the next question it was trying to know which gender got most stops? And the answer was that number of stops which men got was more than twice the number of stops which women got as demonstrated in Figure 10. Also, in 2018 the ratio of female in US is 49% and for men this number is 51%. [3] Moreover, approximately we have same ratios for 2012-2017. So, based on our results and these ratios we can make conclusion that women care about rules more than men.

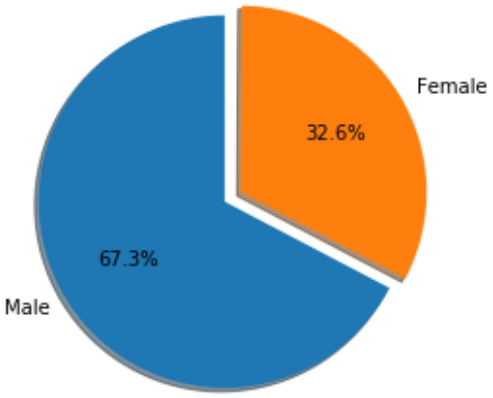

Figure 10: Relationship between Gender and Violation.

Another interesting question was that do owners of new cars care more about rules than owners of old cars? Notice that the assumption here that an old car means a car which was produced 10 years before its stop.

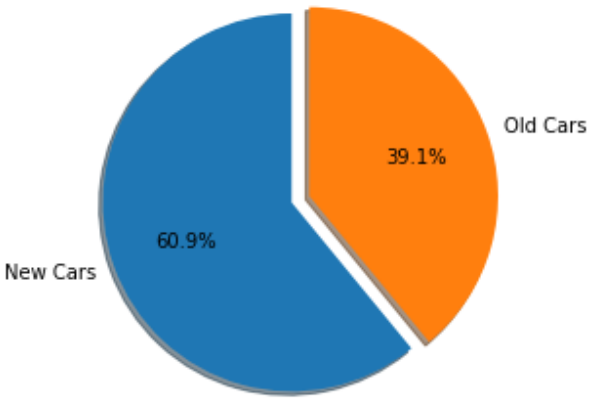

Figure 11: Old cars owners stops vs new cars owners stops.

Based on Figure 11 we can draw the conclusion that 60.9% of violated cars are new and 39.1% of them are old.

Driving rules are different between state, as a last question it's interesting to know that do drivers who are living in a specific state and have gotten driving license from that state care more about rules than people who have gotten driving license from other states and they are just traveling to that specific state?

Based on Figure 12, drivers who come from other states care more about rules. But this information is not enough as most of transportations in a specific state are done with people who are living in that state.

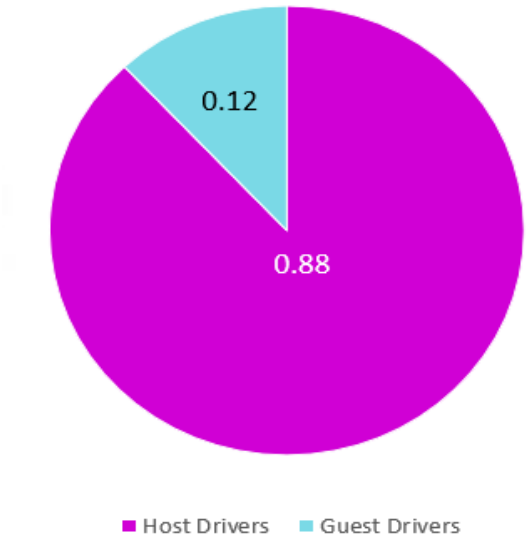

Figure 12: Host drivers stops vs guest drivers stops.

To sum up, dealing with traffic violations is interesting as you can define some patterns and relationships between different attributes, and all of these will be very useful in the future as can be used for further actions that can reduce the number of fatal, personal injuries, and properties damage. In addition, drivers can use these results to avoid future fault and reducing the probability

of being pulled over, also it helps then being more focus when they drive close to a certain location that has a bad history for other drivers.

References:

1. Robin Chase. “Car-sharing Offers Convenience, Saves Money and Helps the Environment”. U.S. Department of State, Bureau Of International Information Programs. link: https://photos.state.gov/libraries/cambodia/30486/Publications/everyone_in_america_own_a_car.pdf

2. “Traffic violations”. https://www.montgomerycountymd.gov/

3. “Population Distribution by Gender”. https://www.kff.org/other/state-indicator/distribution-by-gender/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D