Traffic Stops Disparities

A Comparison of Vermont with Florida

SPRINGBOARD - CAPSTONE PROJECT 1
JITENDRA AGARWAL
JAN 25, 2018

**** Data provider: - Open policing project by Stanford**

Problem Overview

Every day, Police pull over more than 50,000 drivers making the traffic stop as the most common police interaction.

There is a social perception of significant disparities in policing. These disparities can occur for many reasons: differences in driving behavior, to name one. But, we will check if we find evidence that bias also plays a role.

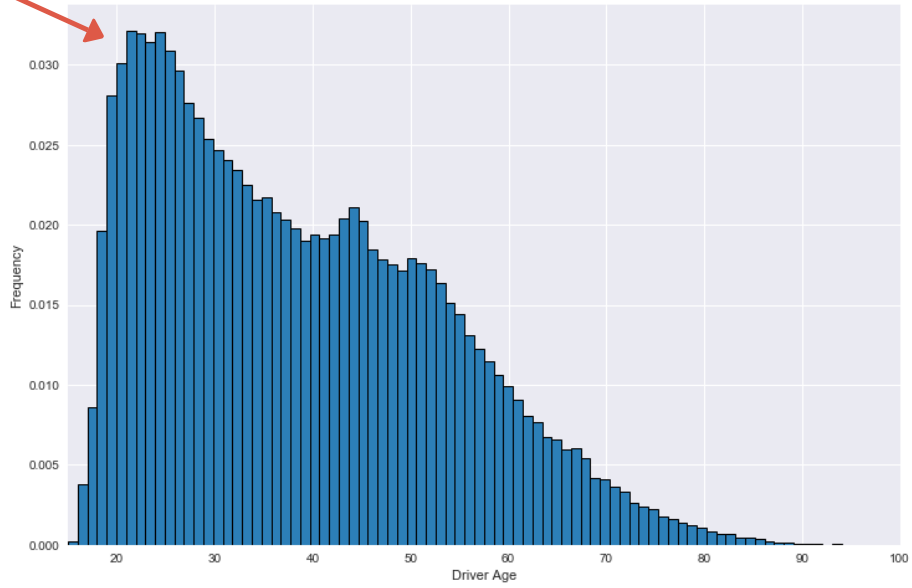
The Goal

The Goal is to analyze the patterns in traffic stops and highlight if there are disparities. It will conclude by offering recommendations for improving law and order condition.

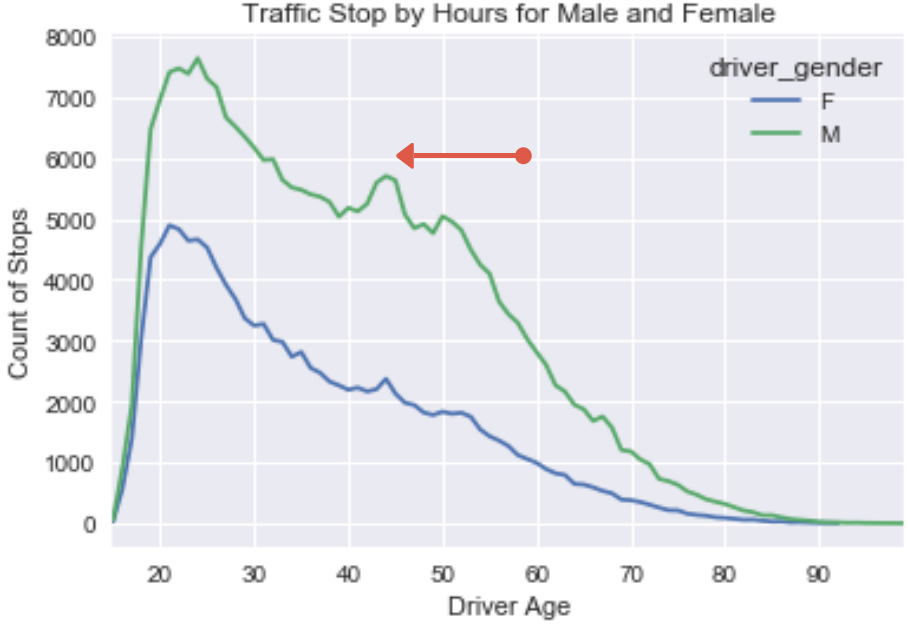
It will also provide a recommendation for private parties like insurance companies.

Exploratory Data Analysis and Trends

Florida 2015 - Distribution of Traffic Stops by Age and Gender



- The most Frequent Traffic stops involve Drivers of Age between 21 and 24

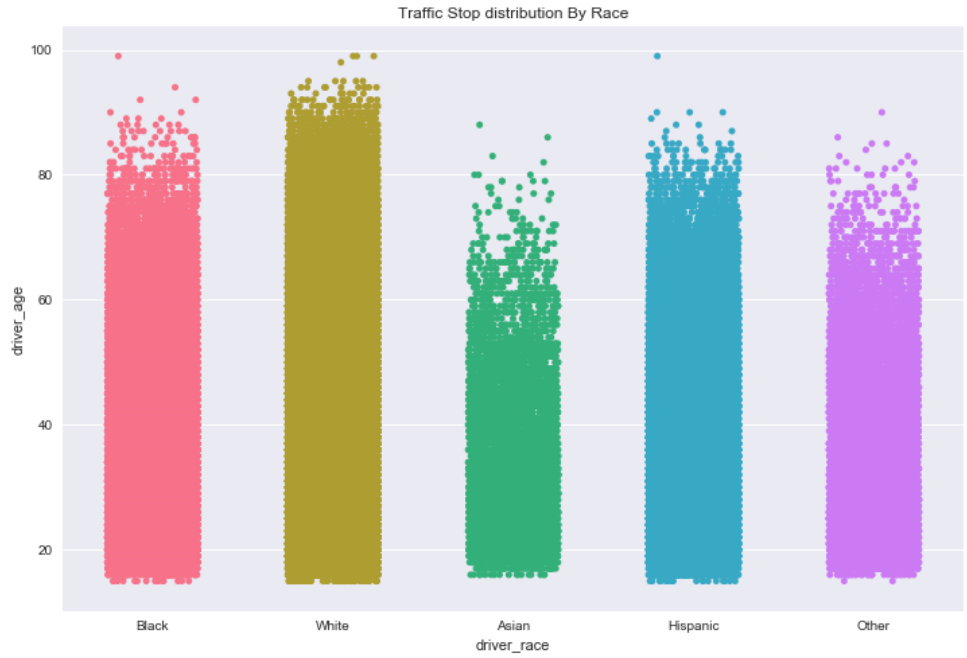


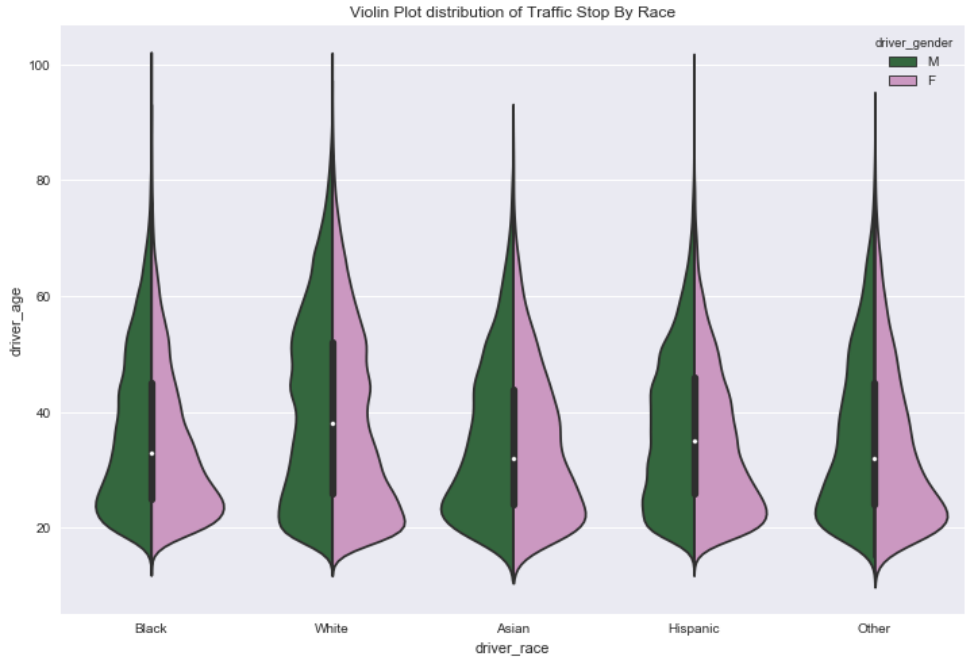
- The trend of traffic stop by age is same for male and female but male between 40-45 have a little higher rate of traffic stops

Florida 2015 - Distribution of Traffic Stops by Age, Race and Gender

- Traffic stop pattern shown by strip scatter plot of Driver age is equally distributed for all the races but it is interesting to see that Asian Seniors drivers are very less likely involved in the traffic stops.

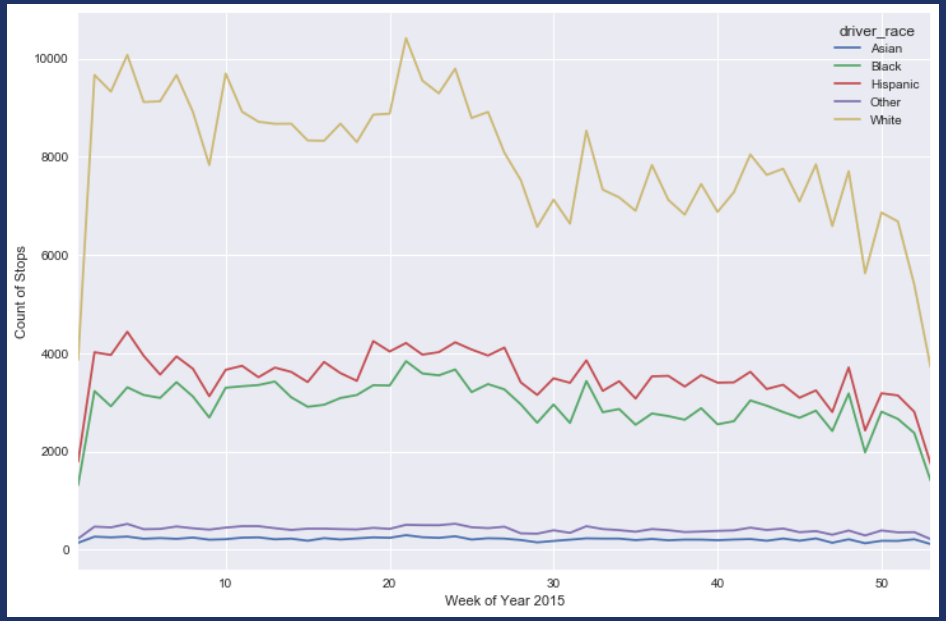
- Taking the Gender also in the consideration, it is visible that younger female drivers are more likely to get pulled over compared to men, especially for minority races. Violin plot width for 60+ is very thin for Minority races.



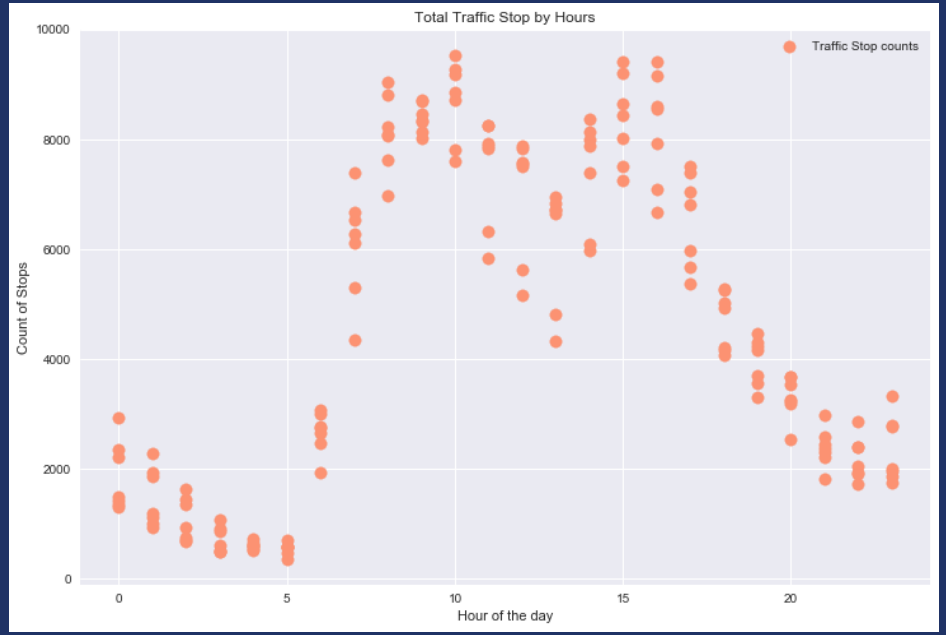


Florida 2015 - Trends of traffic stop over the year and during the time of the day

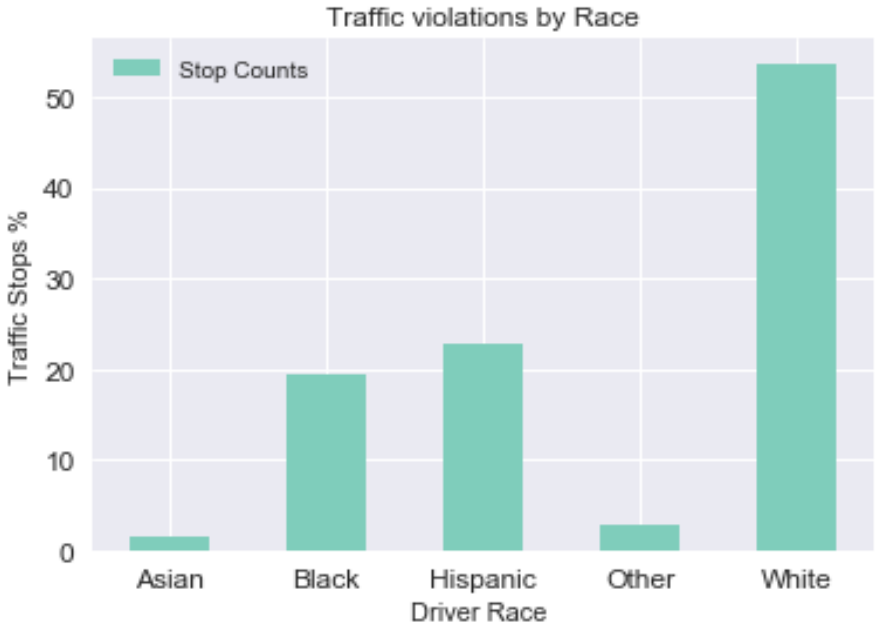
- Holiday weeks (Week No. 51 and 52) see a marginal drop in the traffic stop counts and it is notably high for white drivers during summer breaks. This could be due to holiday time and more tourist drivers coming to Florida.



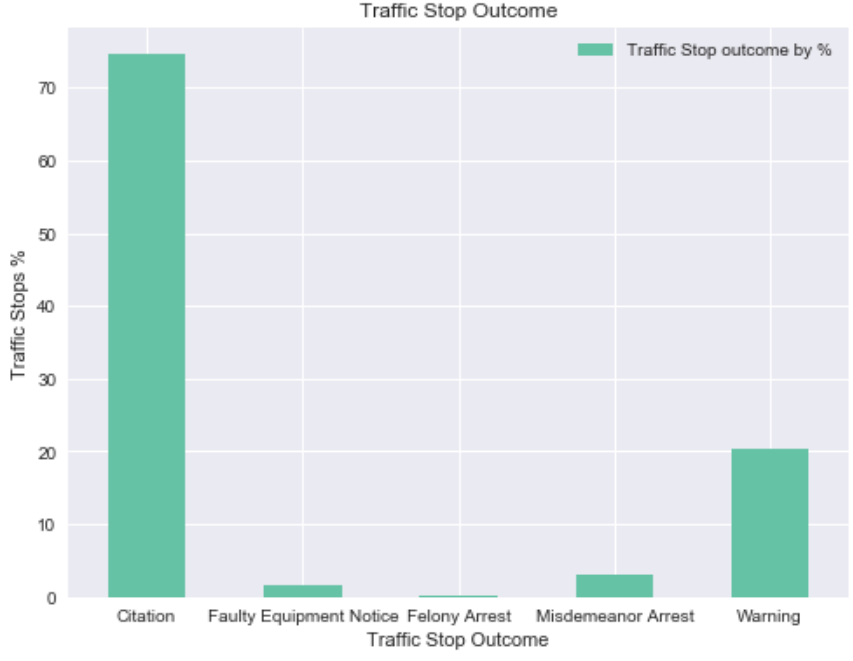
- The most common time of Traffic stop is between 10-11 AM and 3-5 PM. It is very less likely to get pullover during morning rush hour of 7-8 AM.



Florida 2015 - Distribution of Traffic Stops by Stop outcome and race

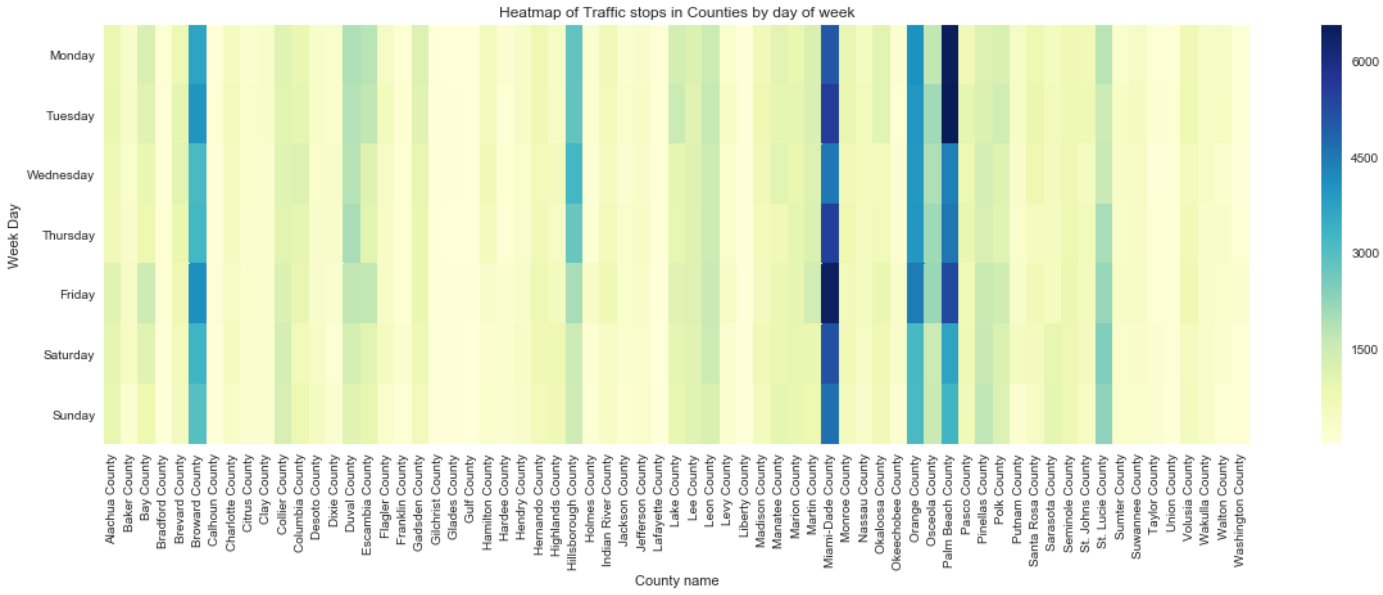


- More than 52% drivers who got pulled over are of majority Race.
- After Normalizing and Taking the State Race demographic into the account, It is seen that there is 8% more traffic stops for a driver from Minority races.



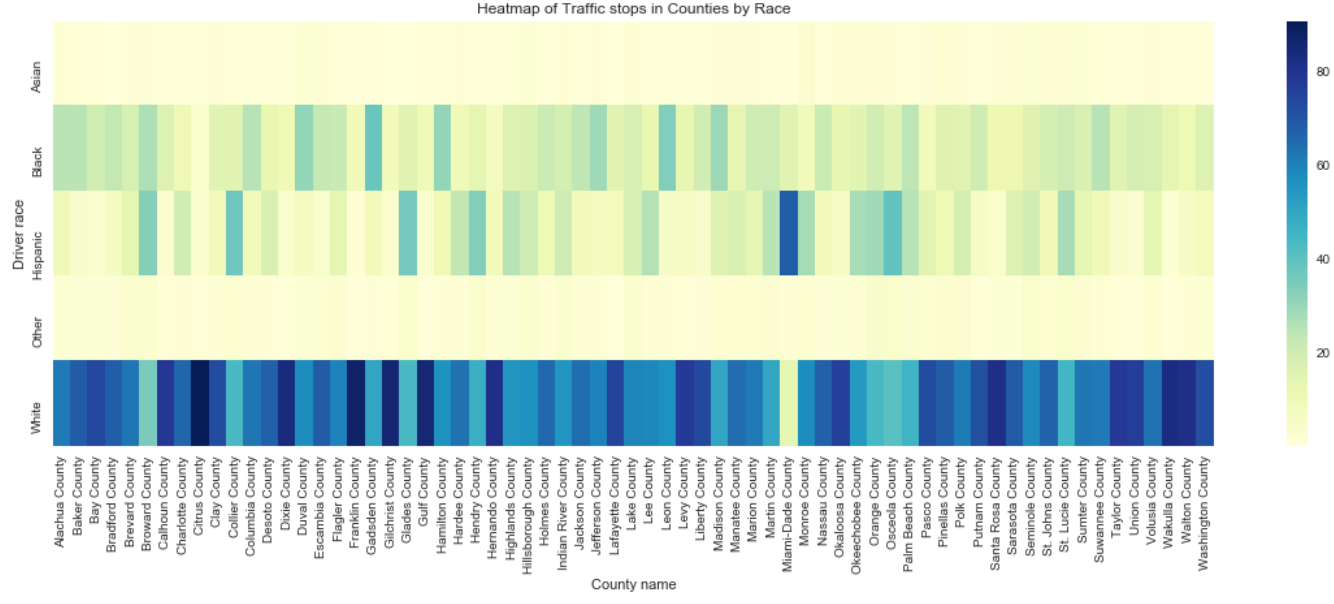
- The most common outcome of traffic stop is a Ticket, however, close to 20% were let go with a warning only
- After Normalization, it was observed that there is 12% more chance to get a ticket after a traffic stop if Driver belongs to Minority races.

Florida 2015 - Distribution of Traffic Stops by County, Weekday and Race

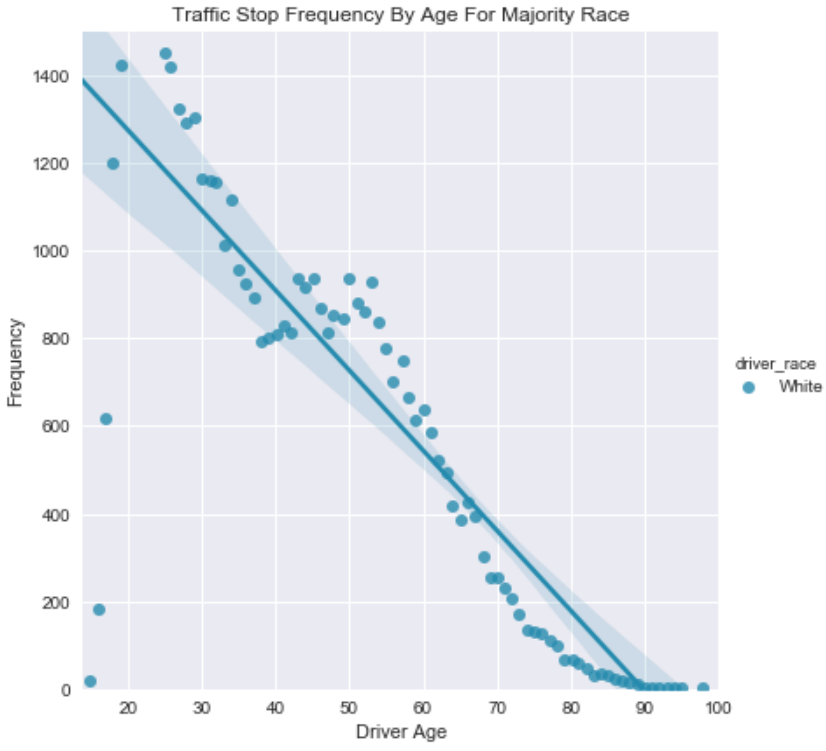


- No of Traffic Stops are very high for counties with the high density of population like Miami-Dade County.
- Interestingly Palm beach has a very high Stop rate for Monday and Tuesday

- Traffic pullover % by Race is consistent and Majority race Driver are most common to get pulled over
- Miami has a very high pullover % for Hispanic drivers.



Data Story



Linear Correlation for Non-White Drivers traffic stops with age:-

- The Linear Correlation between traffic stop and age is even sharper for non-white drivers.

- Black and Hispanic drivers of age 21-30 are pullover more than the average.

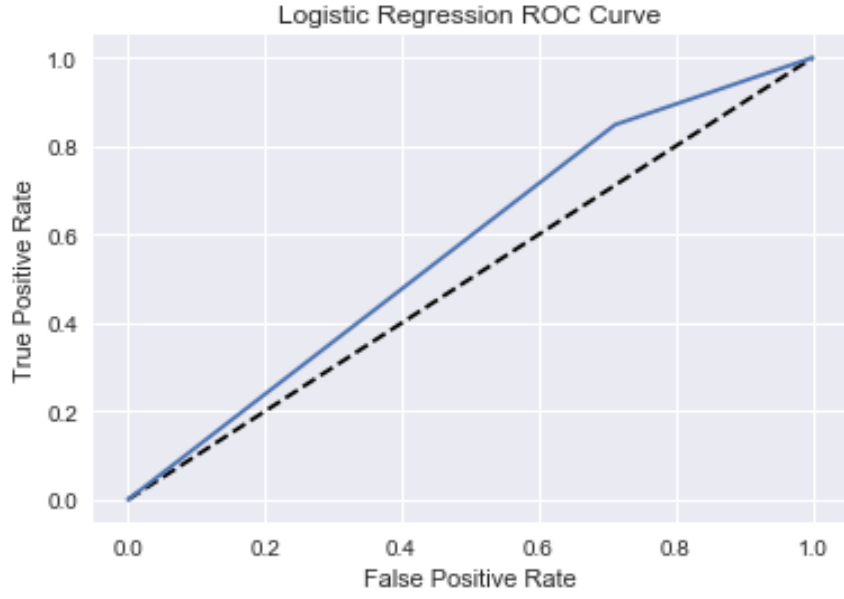
Linear Correlation of White Drivers Traffic stops with age:-

- The number of traffic stops drops when the age of driver increases.

- White Drivers with age between 21-24 has a high likelihood of being pulled over than anyone else.

Logistic regression model

This is a classification problem because stop outcomes is a predefined set. based on the lasso model there was no significant feature for predicting the target and for this type of categorical data we selected a logistic regression model to see if we can predict the outcome of traffic stops and the ROC AUC score is 0.65



	Precision	Recall	f1-score	Support
0	0.55	0.29	0.38	5068
1	0.65	0.85	0.74	7970
Avg/total	0.61	0.63	0.6	13038

- - The Logistic regression model has the optimistic AUC of ROC as 65.33%
- - The Mean of 5 fold cross validation score is also in same range as .6514
- - Recall (TP / (TP+FN)) - Probability of making right prediction from random example is .63
- - Precision (TP / (TP+FP)) - probability of making right prediction from positive examples is .61

There are 65% possibility that we can accurately predict the outcome of a random traffic stop based on Driver's gender, Race, Age, Day and Location provided as feature.

FL vs VT

Comparing Florida with Vermont

Key Statistic	Florida	Vermont
Traffic stop % by race	White 52%, Others 48%	White 94%, Others 6%
Citation %	34%	75%
Most common Day	Thursday and Monday	Thursday
Most common age	22-24	21-24
Most Common Driver profile	White male under 24	White male under 24
Most common time of Day	10AM - 11 AM, 3PM - 5 PM	5 PM- 6 PM

- Florida and Vermont have very different driver demographic and very different population density areas. The analysis shows few key differences and many common patterns.
- There is a key difference in the no of tickets issues but it turned out that Vermont state data point collection was flawed and the most of the traffic tickets issued were captured as a written warning.

Conclusion

- Every traffic stop is different in many ways and outcome of one stop is not related to another. It is not statistically possible to accurately predict the outcome of a traffic stop, however, based on the patterns a highly likely scenario can be established for each case.
- Statistical analysis and pattern may not be useful for a driver as each case is different and there are a lot of Data points not captured in this study which can influence the outcome of traffic stop but this statistical analysis can be used by government and private parties for policy or business decisions.

Recommendations

- More and More Young male drivers under age 23 are pulled over across the country for speeding or other movement-related traffic violations. it's the time that community and social volunteer consider to spend more time with young generation educating on this.
- Insurance and tech companies can explore more patterns from this data and use them to intuitively engage the young drivers into safe driving and use technology to deploy resource more efficiently.

Key Assumptions

- We selected Florida and Vermont because have very different driver demographic and very different population density areas. An assumption was made during data wrangling that the records we dropped due to missing value are randomly spread across all segments and do not impact the outcome by any mean.
- There is a key difference in the data collection for Vermont state. The most part of traffic ticket issues was captured as a written warning. we used this assumption that a written warning is actually a warning only.
- The factors impacting the traffic outcome is not captured completely. officer's training level, Driver's experience of driving, repeated offenders etc were not measured in the study.
- We assume that there is no bias exists based on the officer profile and all officer responds same to a given traffic stop scenario,

Standford Open Policing Project

- The Stanford Open Policing Project — a unique partnership between the Stanford Computational Journalism Lab and the Stanford School of Engineering — is changing that. Starting in 2015, the Open Policing Project began requesting such data from the state after state. The project has collected and standardized more than 100 million records of the traffic stop and search data from 31 states.

- Data from 20 states, comprising more than 60 million state patrol stops, are sufficiently detailed to facilitate rigorous statistical analysis. The result? The project has found significant racial disparities in policing. These disparities can occur for many reasons: differences in driving behavior, to name one. But, in some cases, it finds evidence that bias also plays a role.