

Exploring Racial Bias in Traffic Policing: An Analysis of Six Eastern States of USA

Mahfuzur Rahman Chowdhury

Matriculation number: 23305223

Methods of Advanced Data Engineering

Friedrich-Alexander-Universität Erlangen-Nürnberg

Introduction

- Research Question: Are there observable patterns in traffic stop occurrences and their outcomes based on the drivers' race?
- Motivation:
 - Importance of addressing potential racial biases in law enforcement.
 - Relevance to human rights and equality.
- Scope: Analysis of traffic stop data from **six eastern states** using the Stanford Open Policing Project dataset.

Data Pipeline

- Datasets:
 - Sourced from the Stanford Open Policing Project [1]
 - Licensed under the ODC-BY 1.0 [2]
 - Total rows: 9,494,312
 - Six states:
 - New Hampshire
 - Rhode Island
 - Connecticut
 - Vermont
 - Massachusetts
 - Virginia

[1] Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., Goel, S.: A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour* 4(7), 736–745 (2020). <https://doi.org/10.1038/s41562-020-0858-1>

[2] OpenDataCommons: OpenDataCommons Attribution License (ODC By) Summary—OpenDataCommons: legal tools for open data.

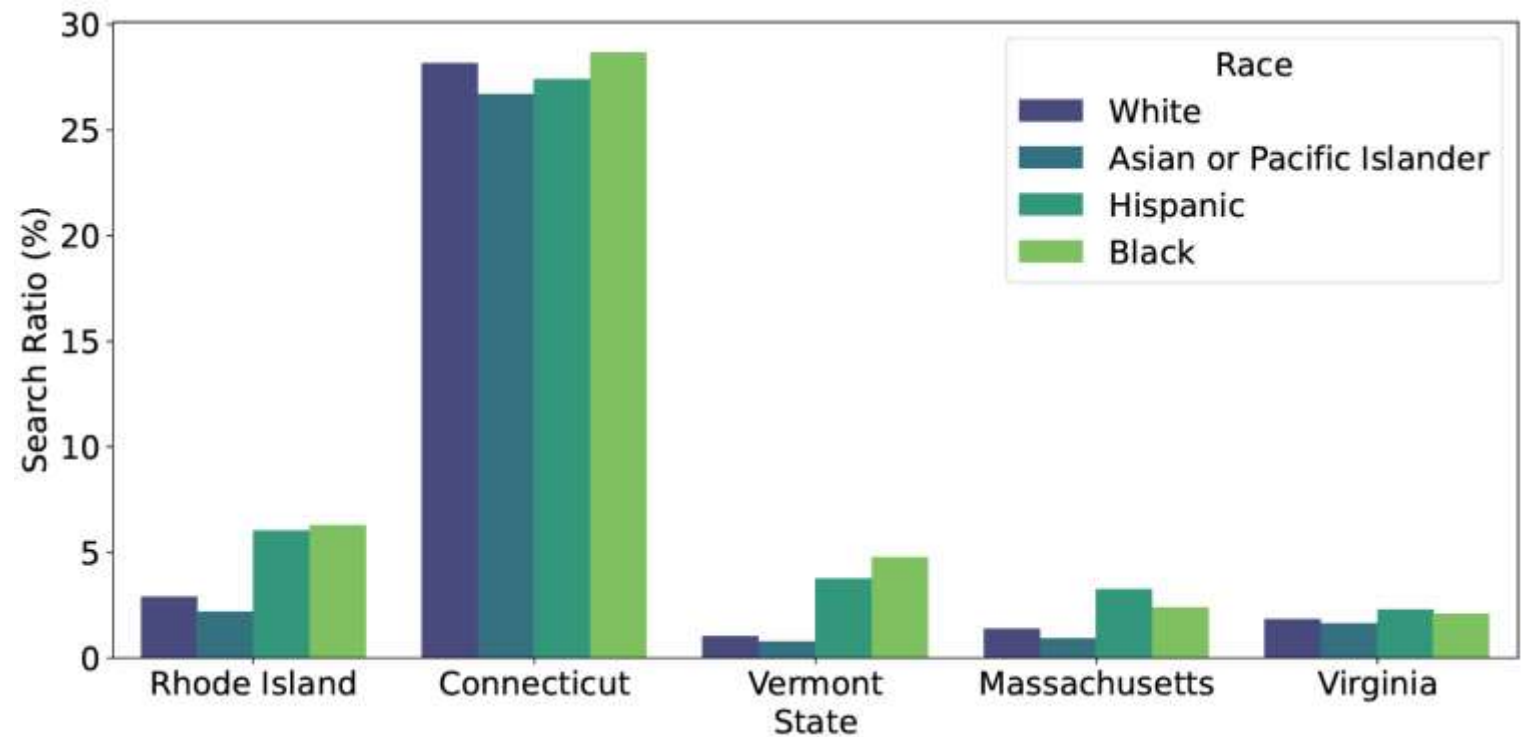
Data Pipeline

- Processing Steps:
 - Downloaded and unzipped .csv file
 - Standardized column names (Example: 'driver's race' to 'race')
 - Merged data into a unified SQLite database (1.08 GB)
 - Handled missing values (null value to '000' string value)

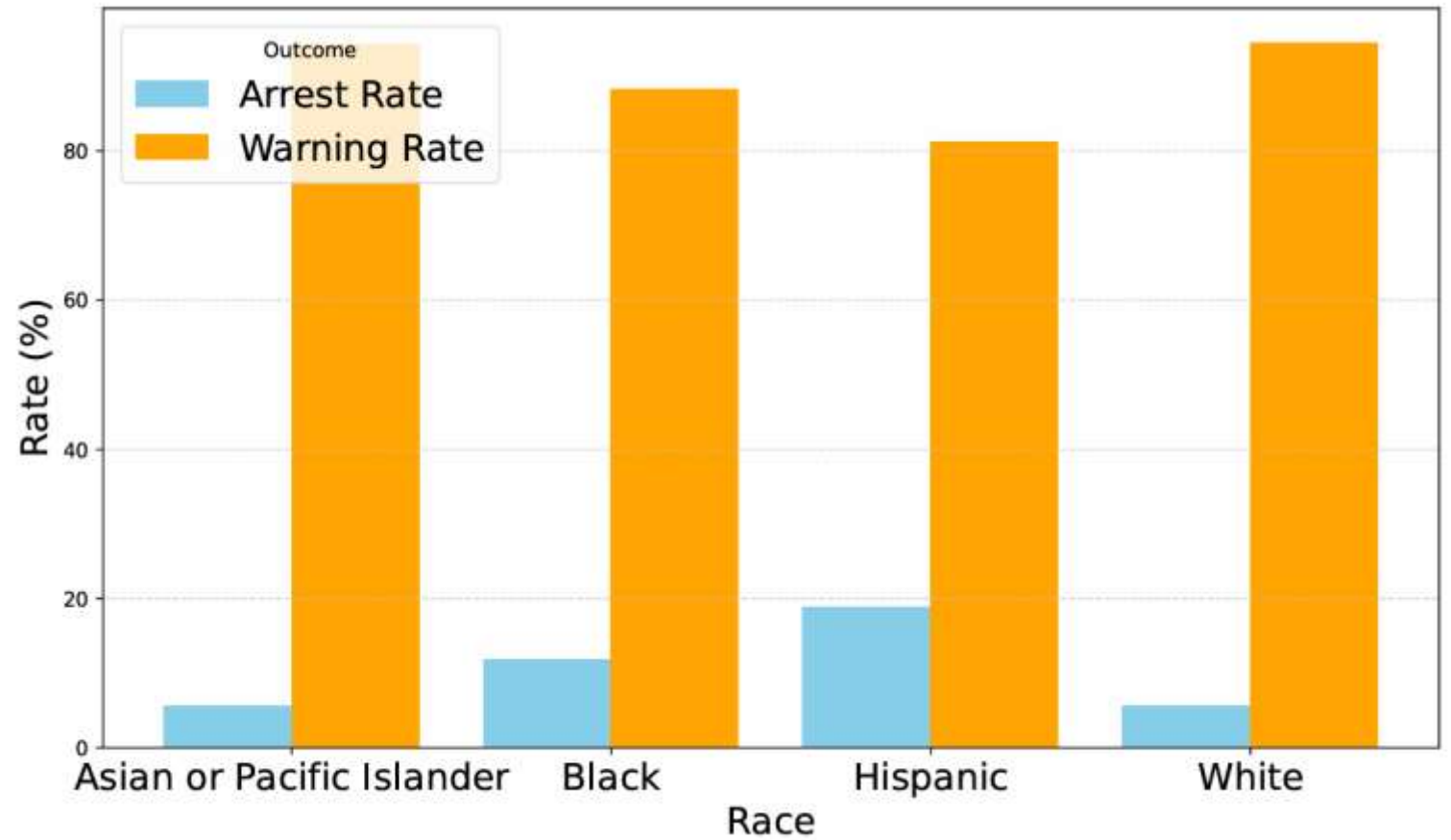
Challenges in Data Handling

- Large Dataset Issues:
 - High computational costs for manipulation and visualization.
- Missing Data:
 - Some columns have large number of null values.
 - Variations in data columns across state datasets.

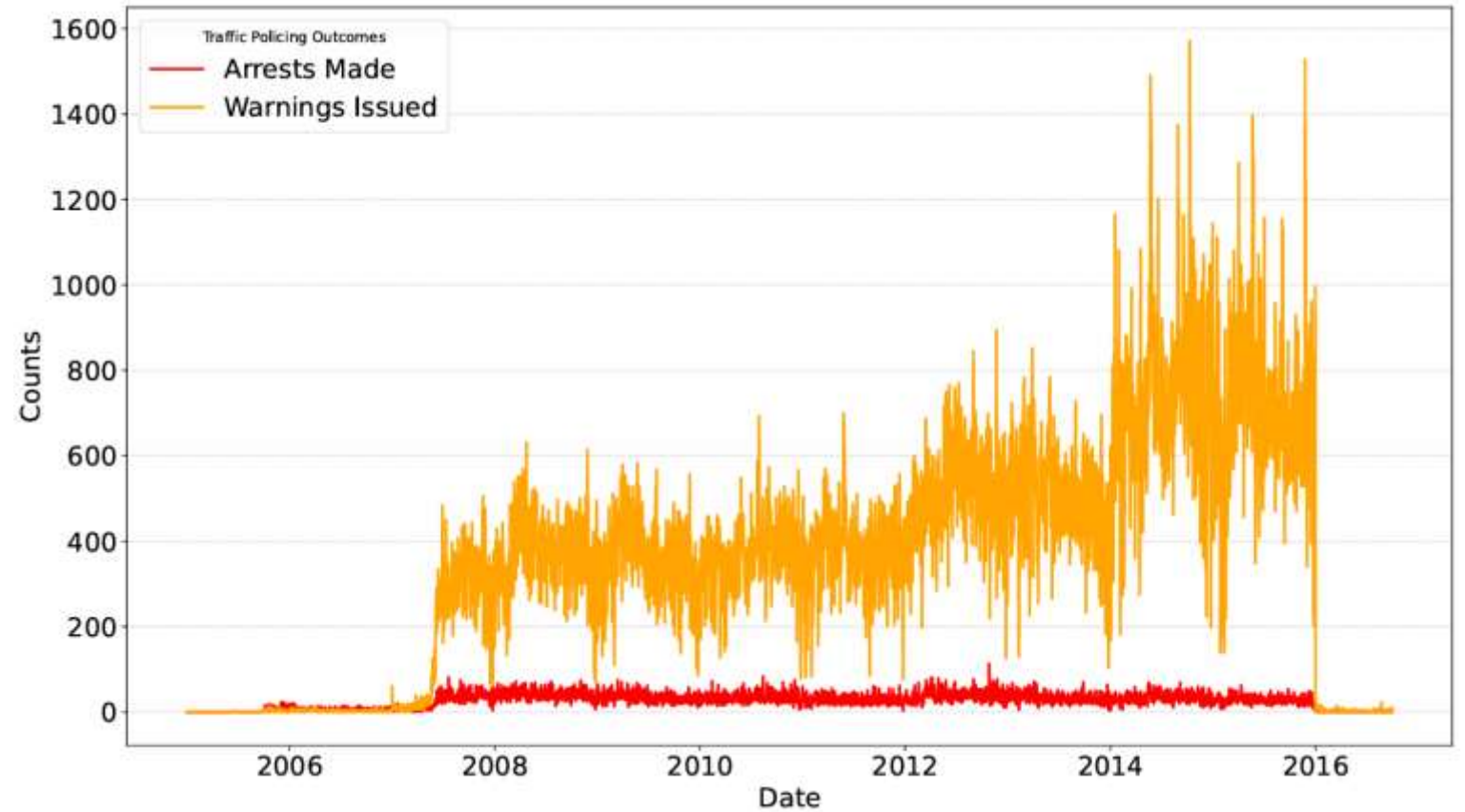
Search Conducted Ratio by State and Race (Conducted vs Not Conducted)



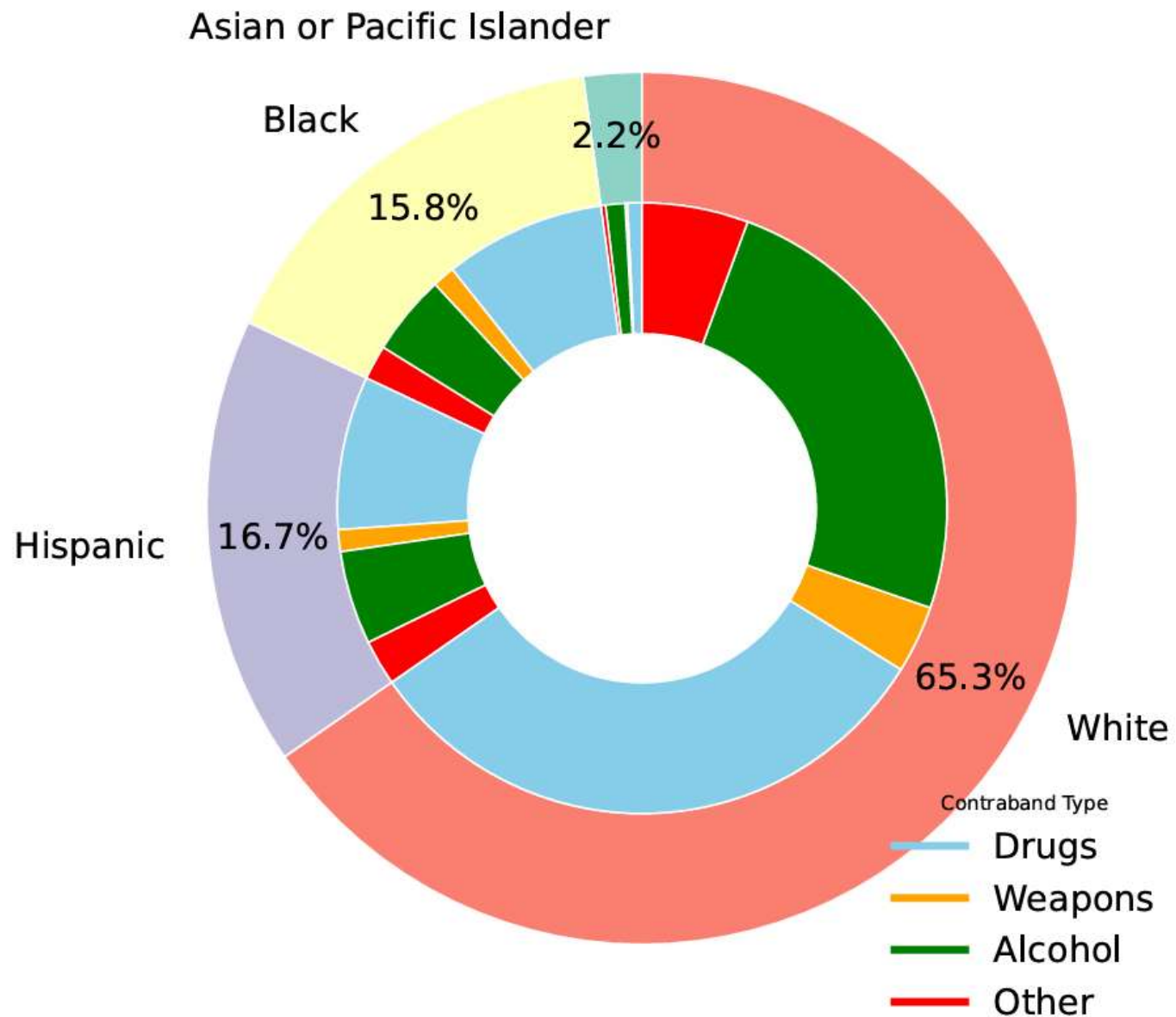
Arrest vs. Warning Rates by Race



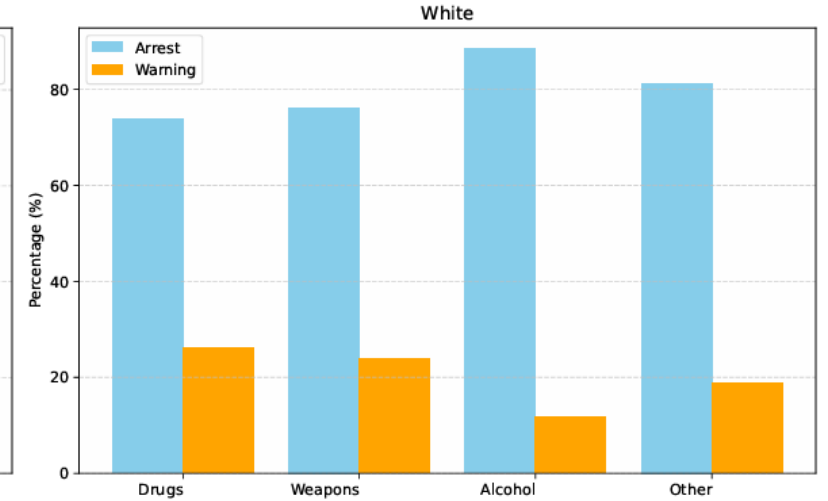
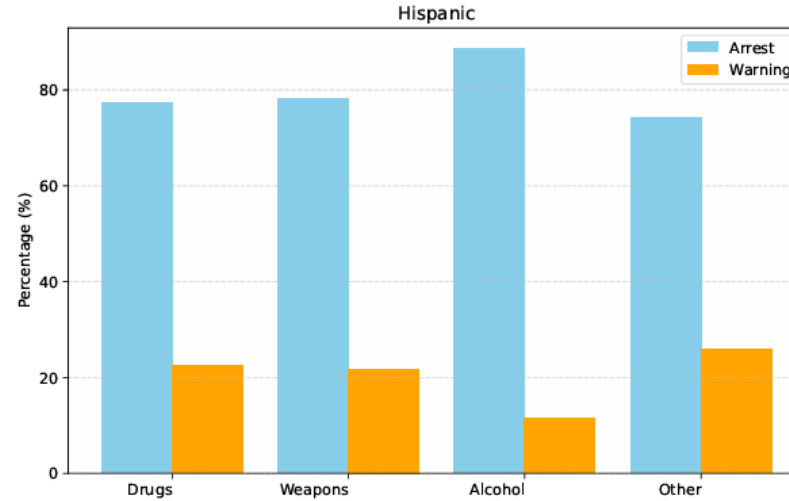
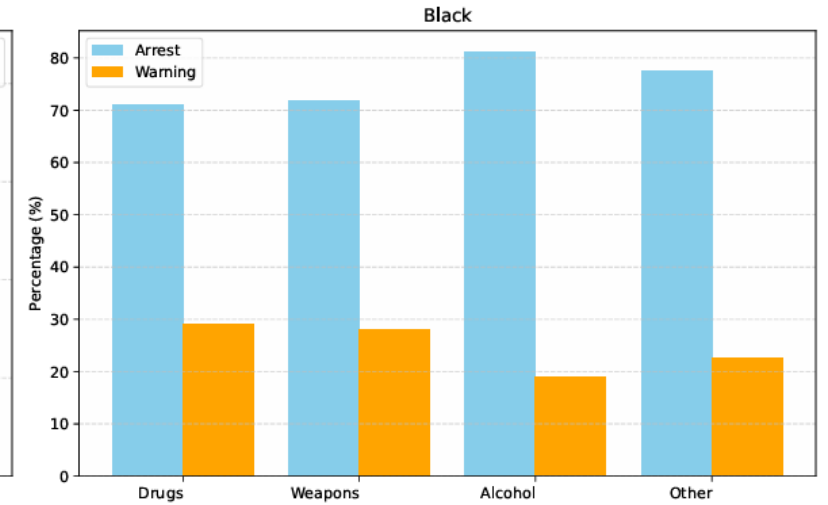
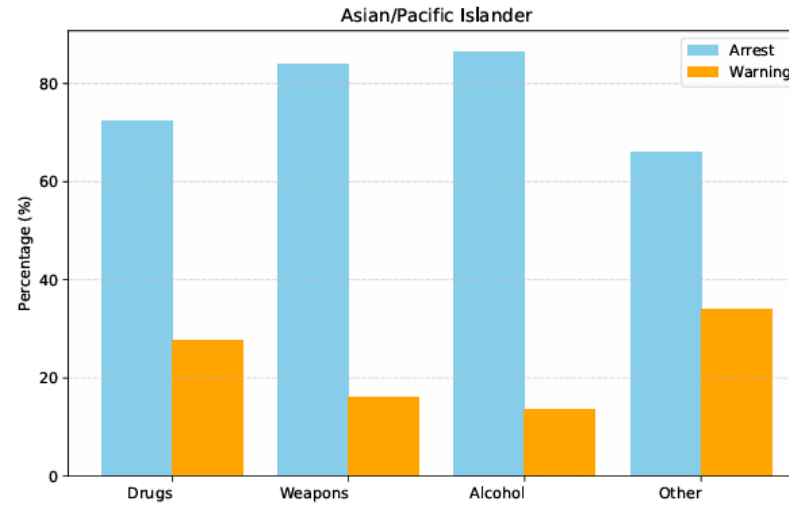
Timeline of Arrests and Warnings Issued



Contraband Distribution



Contraband Outcome Analysis by Race



Limitations

- Data Constraints:
 - Only six states analyzed, limiting generalizability to the entire US
 - Missing contextual variables
- Computational Limitations:
 - Difficulty in scaling analysis to include all 42 states in the dataset

Conclusions

- Observations:
 - Evidence of racial disparities in traffic stops and outcomes.
 - Higher arrest rates for Black and Hispanic drivers, even with lower overall contraband rates.
- Future Work:
 - Incorporate more states and contextual variables for a comprehensive study.
 - Analysis reason for stop.

Thankyou
Any Question?