An analysis of gender bias in making traffic stops in Rhode Island.

By Ahmed Suhail
The University of Auckland

Abstract

Under the Sustainable Development Goals, Goal 5 proposed by the United Nations is to "Achieve gender equality and empower all women and girls." Inline with the goal of Gender Equality the United Nations has for sustainable development, this study assesses traffic stops in the United States for gender biases. The Stanford Open Policing Project made publicly available a dataset of about a hundred million stops in the United States, half a million of which from Rhode Island were analysed for gender discrimination in this study. Insights into the situation are provided, uncovering trends in male and female arrests. The study investigates potential bias in decisions to stop drivers, examining the rates at which drivers were arrested. A test for discrimination based on the principles by Nobel laureate Gary Becker was proposed, where the rates at which drivers are searched and the rates at which the search turns up contraband are examined. A Logistic Regression model with an accuracy of 96.33% is fit to the data to test for discrimination analysing the odds ratio of one being arrested if one is male. The study found that while males are more likely to be arrested for the same features, searches in males more often reveal contraband than in females (searches in males were 1.8 times more likely to turn up contraband), justifying male searches.

Introduction

Every person in the world deserves to reach their full potential, but disparities in the way they are treated in their daily lives and the disparities that their loved ones face, hinder this basic right. People have become used to seeing gender inequalities in their homes and communities every day. Unequal responsibilities for work in the home socialises children in the home to believe that these duties are women's only roles, thereby curtailing generational change and narrowing girls' ambitions. The world currently faces lots of problems because of the differences in the treatment of a specific gender. Women workers, in developing and third world countries, are often low-skilled and underpaid, with limited opportunities for growth in their careers. Household chores, the duty of caring for the family, and issues of safety oftentimes keep the girls out of school, while expectations of money drive boys out of school. Menstruation is still a taboo in most places in the world. In India, women are not encouraged to visit places of worship when they are menstruating, such as in the Sabrimala Temple ("What is the Sabarimala case?" 2018).

According to the United Nations, "Gender equality is not only a fundamental human right but a necessary foundation for a peaceful, prosperous and sustainable world. Unfortunately, at the current time, 1 in 5 women and girls between the ages of 15-49 have reported experiencing physical or sexual violence by an intimate partner within 12 months and 49 countries currently have no laws protecting women from domestic violence. Progress is occurring regarding harmful practices such as child marriage and FGM (Female Genital Mutilation), which has declined by 30% in the past decade, but there is still much work to be done to eliminate such practices completely.” There is a need to provide all genders equal access to education, decent work and equal representation in the social and political scenario. Implementation of policies to curb the disparities men and women face in different industries because of their gender needs to be addressed immediately.

Gender equality will be achieved when men and women will enjoy, across all sectors of the society, equal rights and opportunities. Women heavily underrepresent sectors such as governments, sports and economics while as in some countries, it is still socially unacceptable for a man to be a house maker, or participate equally in household chores. How do we measure Gender equality? According to the United Nations, which has a gender inequality index, based on the premise that "all too often, women and girls are discriminated against in health, education and the labour market with negative repercussions for their freedom". According to this index, Norway is currently one of the best places on the world for women, the United States stands at 5th and New Zealand at the 7th best.

Under the Sustainable Development Goals (Link: https://bit.ly/2PvS9eT), Goal 5 of proposed by the United Nations is to “Achieve gender equality and empower all women and girls”. Target 5.2 aims at the elimination of all forms of violence against all women and girls in public and private spheres, 5B at enhancing the use of enabling technology, “in particular information and communications technology, to promote the empowerment of women.”

On a typical day, police officers in the United States make more than 50,000 stops, more than 20 million motorists over a year (Davis, Whyde, & Langton, 2015). Around 10% of the people involved in traffic stops and 25% of the persons in street stops believe that police had appropriately behaved, with many of them resorting to filing complaints against the police (Langton, Durose, & Statisticians, 2013). The Stanford Open Policing Project “compiled and analysed a unique dataset detailing nearly 100 million traffic stops carried out by 21 state patrol agencies and 29 municipal police departments over almost a decade.” This dataset was built through a series of public records requests filed in all 50 states. The team reported evidence of widespread discrimination in decisions to stop and search drivers. It was reported that legalisation of recreational marijuana indicated “that such laws can have significant and unexpected downstream consequences on police behaviour.” The team concluded their findings with “insight into the differential impact of policing on minority communities on an unprecedented scale” (Pierson et al., 2017). To facilitate future analysis, the team redistributed these records in a standardised form.

Research Question

This study accesses the traffic stop data released by the Stanford Open Policing Project for Gender Discrimination, motivated by the original search for racial discrimination by the team. The motivation behind making the data collected public was to "help researchers, journalists, and policymakers investigate and improve interactions between police and the public." The study investigates if police officers are discriminatory based on gender while making a traffic stop, specifically in Rhode Island in the United States of America. The hypotheses, motivated by personal experiences, was that girls are discriminated against when police officers are making stops, and girls are more likely to be stopped when driving. The study aims at concluding, with statistical evidence, if Police Officers in Rhode Island are discriminatory in making traffic stops against a specific gender. The goal is a successful examination and study of the rate at which male and female drivers are stopped and arrested. It also aims at evaluating the effects of date, race, and other features available on traffic stops by the police and analysing for discrimination in stops by gender using a test that effectively categorises discriminatory stops from non-discriminatory ones.

Related Work

The team at Stanford Open Policing Project worked on the same dataset and used a test by Grogger and Ridgeway (Grogger & Ridgeway, 2006), who proposed a test for racial profiling that does not require explicit external estimates of the risk set. The "veil of darkness" hypotheses, asserted that "police are less likely to know the race of a motorist before making a stop after dark than they are during daylight". Grogger and Ridgeway assumed that "racial differences in traffic patterns, driving behaviour, and exposure to law enforcement do not vary between daylight and darkness", and tested for racial profiling by "comparing the race distribution of stops made during daylight to the race distribution of stops made after dark". The test applied to the data found that black drivers comprise a smaller proportion of drivers stopped after sunset (Pierson et al., 2017), which is reminiscent of discrimination against people of colour in traffic stops. The team also inspected for the consequences for the legalisation of recreational marijuana on stop outcomes. Two patterns, in policing in Colorado and Washington, that had legalised recreational marijuana at the end of 2012 vs 12 states where marijuana was illegal, found out that "legalization reduced both search rates and misdemeanour rates for drug offences for white, black, and Hispanic drivers—though a gap in search thresholds persists" (Pierson et al., 2017).

An investigative news organisation, ProPublica, claimed that an algorithm used in courts across the United States to help make bail and sentencing decisions, COMPAS (which stands for Correctional Offender Management Profiling for Alternative Sanctions), is biased against defendants who are coloured. ProPublica gathered records for around 12000 defendants in Broward County who were assigned a score and faced repercussions from the usage of the algorithm COMPAS, from 2013 to 2014. According to the team at ProPublica, "the recidivism risk categories predicted by the COMPAS tool to the actual recidivism rates of defendants in the two years after they were scored, and found that the score correctly predicted an offender's recidivism 61 per cent of the time, but was only correct in its predictions of violent recidivism 20 per cent of the time" (Angwin, Laeson, Mattu, & Kirchner, 2016).

A study by (Skeem & Lowenkamp, 2016) tested the predictive validity of the Post-Conviction Risk Assessment tool that was developed by the federal courts to help probation and parole officers determine the level of supervision required for an inmate upon release. The authors found that the average risk score for black offenders was higher than for white offenders, but that concluded the differences were not attributable to bias.

Another paper by (Busse & Spielmann, 2004) empirically explored the international linkages between gender inequality and trade flows of a sample of 92 developed and developing countries. The results of the study indicated that gender wage inequality is "positively associated with a comparative advantage in labour-intensive goods, that is, countries with a larger gender wage gap have higher exports of these goods". The authors also found out that "gender inequality in labour force activity rates and educational attainment rates are negatively linked with a comparative advantage in labour-intensive commodities".

An analysis of the same data released by the Stanford Open Policing Project by Alex et. Al in (Chohlas-Wood, Goel, Shoemaker, & Shroff, 2018), concluded that people of colour also faced discrimination in non-moving violations. High stops in Nashville Police Department's found no immediate or long term response in serious crime committed in the area, with only less than 2% of the stops leading to arrests, often for licence violations and drug possession.

Assessing bias in traffic stop decisions

Methodology:

Logistic Regression is conventionally used for Predictive Analytics, to predict whether and with what likelihood an event will occur in the future. Unlike linear regression, it is a classification algorithm. Many people, such as the authors of the book (Kleinbaum & Klein, 2002) introduce it as a predictive analytic technique for when the dependent variable is binary. Another typical use of Logistic Regression, however, is the use of Odds ratio to predict discrimination, as talked about by the authors in the same book cited above (Kleinbaum & Klein,2002) and for better understanding in the article by (Ed, n.d.). Logistic regression will be used to test for discrimination, as was done by the team at Stanford Open Policing Project, where the odds ratio is used to describe the type of relationship between the dependent variable, which in our case will be male or female. This study, hence going to use Logistic Regression in a not so conventional way, to test for discrimination. The idea will be further explained in a further section, when the test is described in more detail and results documented. An in-depth analysis of traffic stops data in Rhode Island on the Stanford Open Policing Project's website was done over three iterations using different technological stacks. Open Source libraries for Python such as

- Pandas: Software library written for the Python programming language for data manipulation and analysis
- NumPy: A library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with an extensive collection of high-level mathematical functions to operate on these arrays.
- Scikit-learn: A machine learning library for Python, featuring algorithms such as Logistic Regression, K-Means Clustering, k-Nearest Neighbours etc.
- Matplotlib: A Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments

A similar test, as described in the paper was also conducted in IBM SPSS Modeler and using PySpark in an Amazon EC2 instance.

Data description

The state patrol from Rhode Island had provided data of all vehicular stops from January 2005 to December 2015. The available data consisted of the following features:

▪ raw_row_number: A record ID for reference to the raw data

▪ Date: Stop date in the format of YYYY-MM-DD

▪ Stop Time: Time of the Stop, in a 24 hour, HHMM format.

▪ Stop Location: Textual, non-standardised location.

▪ Driver Race: Race of the stopped subject.

▪ Driver Age: Age of the stopped subject.

▪ Driver Sex: Gender of the stopped subject.

▪ Search Conducted: Boolean value that indicates whether a search was conducted.

▪ Contraband Found: Boolean value that indicates if contraband was found

▪ Citation Issued: Boolean value indicating outcome, if a citation was issued.

▪ Warning Issued: If a stop resulted in the issue of warning, this Boolean was selected.

▪ Frisk Performed: If a quick Frisk was performed, this Boolean value was set true

▪ Arrest Made: If the person were arrested, this Boolean value was set true.

	raw_row_number	date	time	zone	subject_race	subject_sex	department_id	type	arrest_made	citation_issued	...	contraband_weapons	contrab
0	1	2005-11-22	11:15:00	X3	white	male	200	vehicular	False	True	...	NaN	
1	2	2005-10-01	12:20:00	X3	white	male	200	vehicular	False	True	...	NaN	
2	3	2005-10-01	12:30:00	X3	white	female	200	vehicular	False	True	...	NaN	
3	4	2005-10-01	12:50:00	X3	white	male	200	vehicular	False	True	...	NaN	
4	5	2005-10-01	13:10:00	X3	white	female	200	vehicular	False	True	...	NaN	

5 rows × 24 columns

Data Preparation

A variety of manual and automated tasks were done to prepare the data for a test that would help assess bias in traffic stops. Following is a table that represents missing values in the entries in the dataset:

raw_row_number	0
date	10
time	10
zone	10
subject_race	29073
subject_sex	29097
department_id	10
type	0
arrest_made	29073
citation_issued	29073
warning_issued	29073
outcome	35841
contraband_found	491919
contraband_drugs	493693
contraband_weapons	497886
contraband_alcohol	508464
contraband_other	0
frisk_performed	10
search_conducted	10
search_basis	491919
reason_for_search	491919
reason_for_stop	29073
vehicle_make	191564
vehicle_model	279593

Since the data was rich vertically, the entries that had gender missing, whose discrimination we were tested for were dropped for analysis. Only data that the officer would have known before making the stop was fed to the model, and hence features such as the type of contraband found whether a citation was issued or an arrest made, were dropped. Linearly related features were also dropped for better performance of the Logistic Regression Model. Other types of steps for preparation included correcting the format of the data types, encoding the data using label encoder class from scikit-learn for modeling etc., snippets of which can be seen from the outputs below:

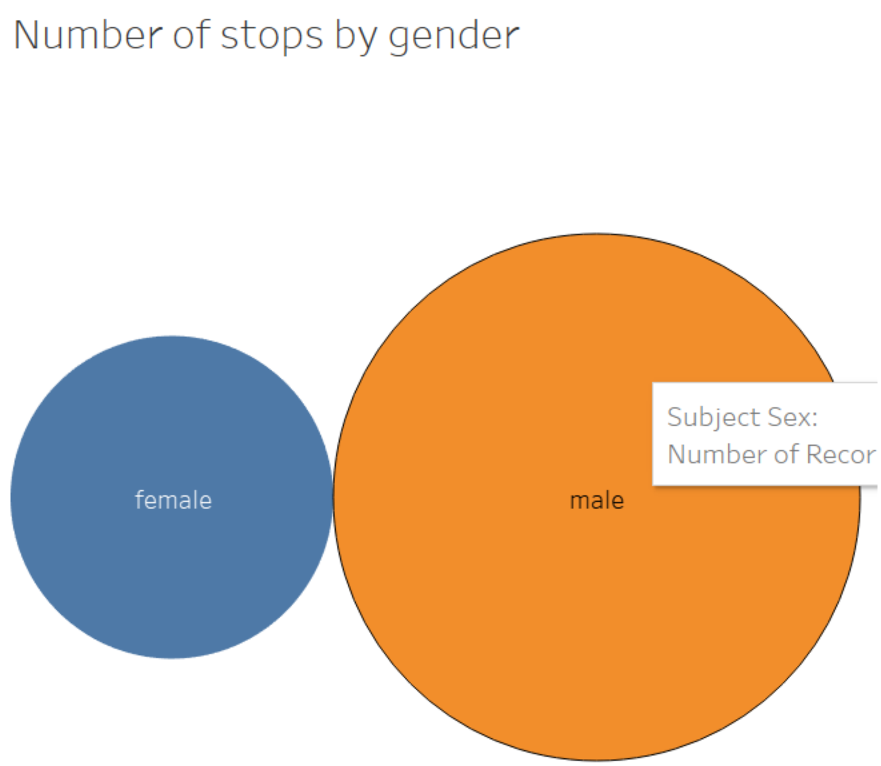
```
ri['subject_sex'] = ri.subject_sex.astype('category')
ri['subject_race'] = ri.subject_race.astype('category')
ri['arrest_made'] = ri.arrest_made.astype('bool')
ri['citation_issued'] = ri.citation_issued.astype('bool')
ri['warning_issued'] = ri.warning_issued.astype('bool')
ri['frisk_performed'] = ri.frisk_performed.astype('bool')
ri['reason_for_stop'] = ri.reason_for_stop.astype('category')
ri['search_conducted'] = ri.search_conducted.astype('bool')
ri['contraband_found'] = ri.search_conducted.astype('bool')
```

The following features were fed to the logistic regression model for the test.

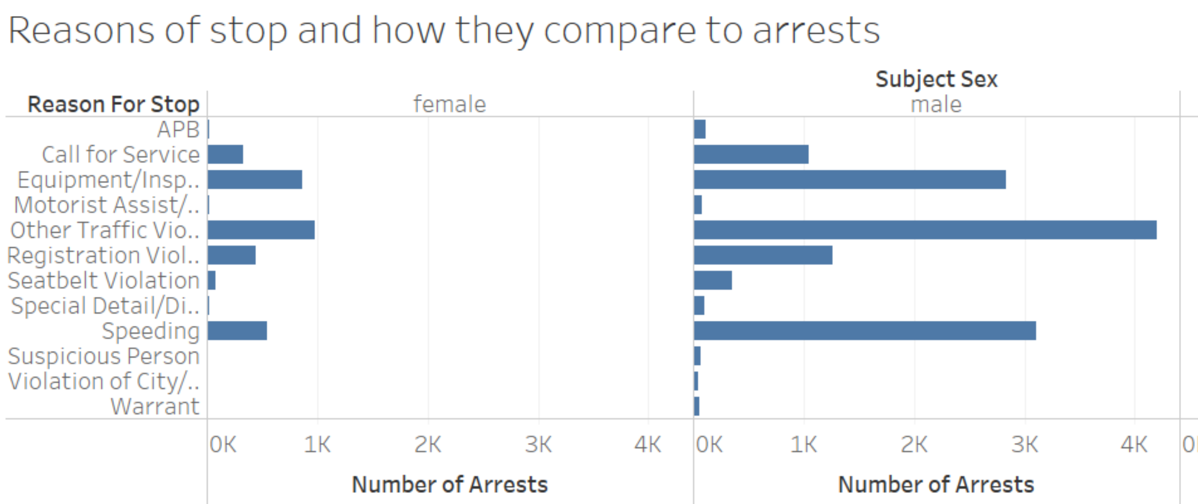
	date	zone	subject_race	subject_sex	arrest_made	contraband_found	frisk_performed	search_conducted	reason_for_stop
0	112	4	4	1	False	False	False	False	8
1	60	4	4	1	False	False	False	False	8
2	60	4	4	0	False	False	False	False	8
3	60	4	4	1	False	False	False	False	8
4	60	4	4	0	False	False	False	False	8

Interpretation

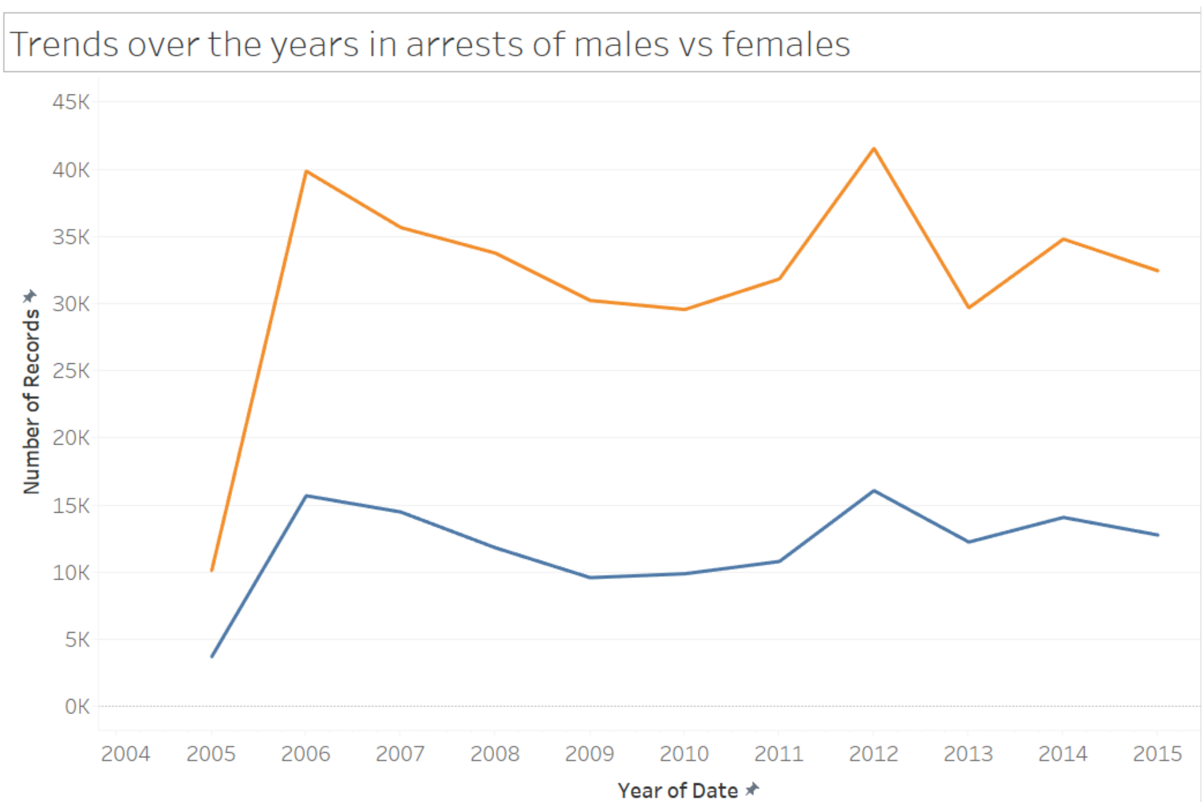
Note that the scope of this study did not allow for comparisons with the base population of Rhode Island, and the following insights, though representative of the disparity in stops is by no means a complete test of discrimination. We found out from the data that male drivers were stopped more often than female drivers.



These numbers are a starting point for understanding racial disparities in traffic stops, but they do not, in and of themselves, provide evidence of racially disparate treatment. For example, if the population of drivers in Rhode Island consists of more males than females, that could explain the higher stop rates seen for males. Males were often arrested more than females for almost every type of violation, which can be seen from the following visualisation.



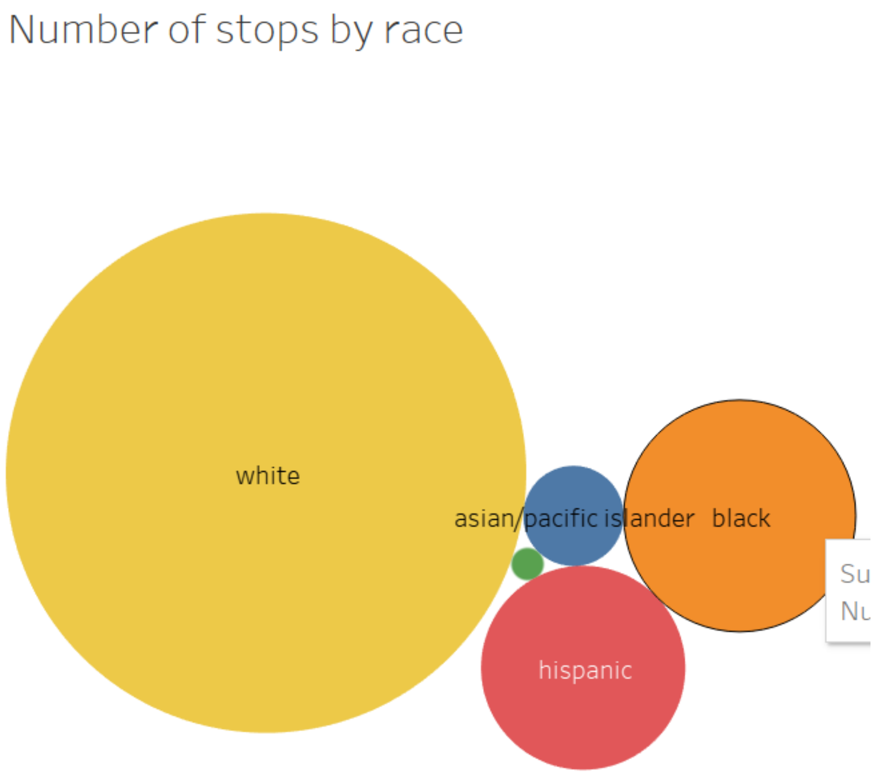
The trends in arrests of males and females seemed to follow the same pattern, as can be seen in the visualisation below, with male arrests being plotted in orange and female in blue.



Controlled for the zone in which males were stopped, the results were the same, with males being more likely to be stopped in every zone.

zone	subject_sex	
K1	female	0.016528
	male	0.038785
K2	female	0.013018
	male	0.031662
K3	female	0.023989
	male	0.052726
X1	female	0.005181
	male	0.013020
X3	female	0.013057
	male	0.030472
X4	female	0.026174
	male	0.059040

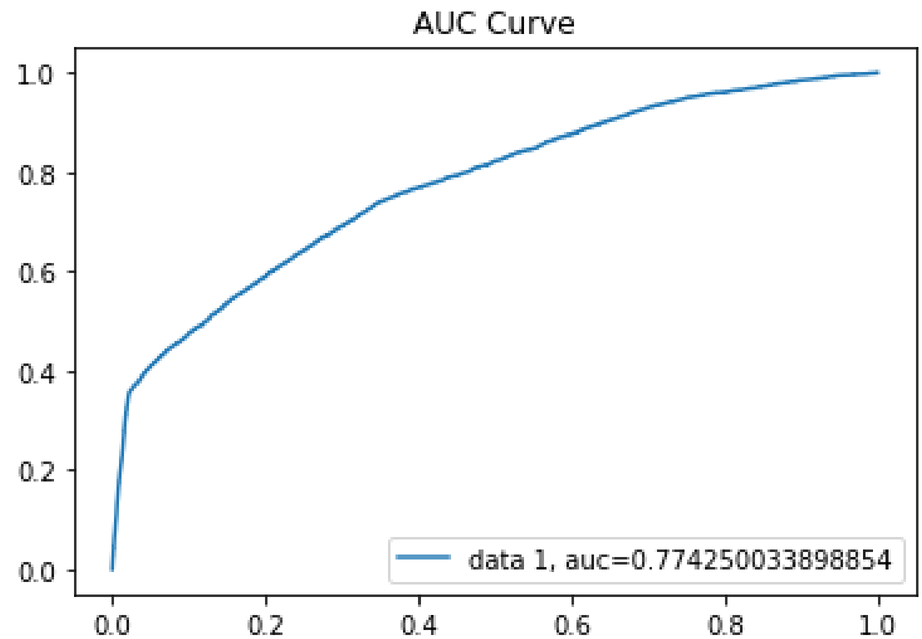
The data was also analysed for stop rates by race, which were as follows.

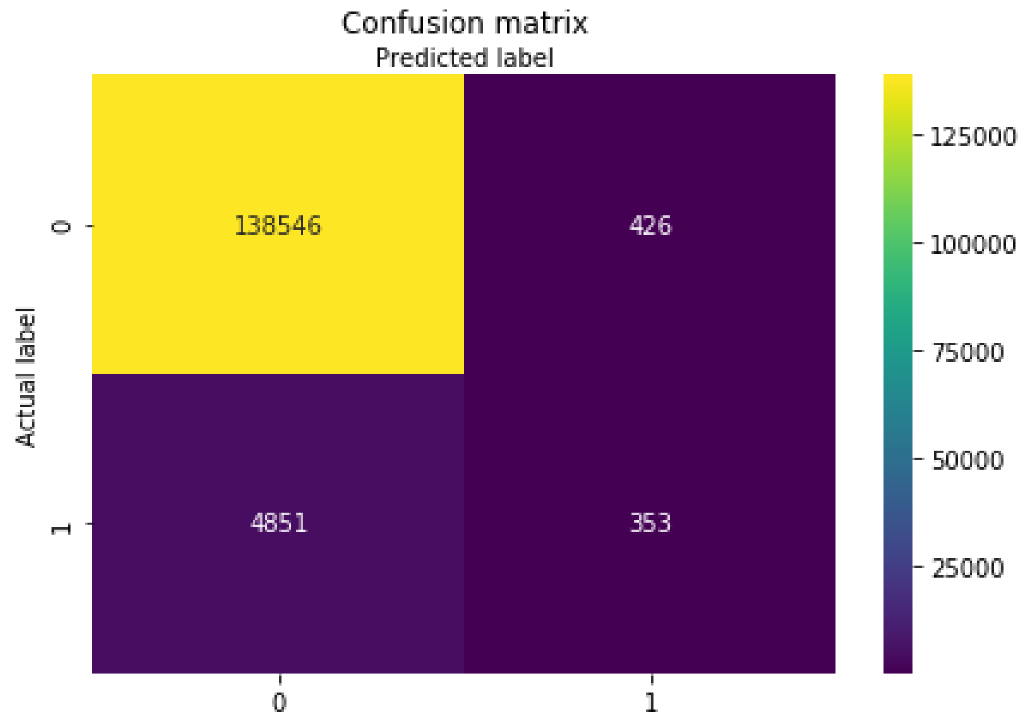


The outcome test

To circumvent omitted variable bias in traditional tests of discrimination, Nobel prize winner Gary Becker, in (Becker, 2012) proposed a test. The outcome test is based not on the search rate, but on the hit rate: the proportion of searches that successfully turn up contraband. Becker argued that even if minority drivers are more likely to carry contraband, absent discrimination, searched minorities should still be found to have contraband at the same rate as searched whites. If searches of minorities are less often successful than searches of majority, it suggests that officers are applying a double standard, searching minorities based on less evidence. For the test conducted in this study, this means that if searches conducted in males turn up contraband at the same rates as females, then officers are not biased in the way they are making traffic stops.

On the basis of the theory developed in the outcome test, a Logistic Regression model was fit on the prepared data to test for discrimination. 6 healthy, linearly independent features, which the officer knew prior to making the stop were fed to test for discrimination, and gather the importance of sex whilst making the stop. Universally accepted ratios of a train-test split of 70-30 were used, in order to make sure enough data was available for training and results were well validated. The metrics function from scikit-learn was used to evaluate the model, which indicated the model was well fit with high accuracy of 96.33%. Area Under the curve was 0.774, and a confusion matrix plotted using seaborn indicated excellent performance, as can be seen from the image. The high accuracy of the data was attributed to the data being rich horizontally and tuned hyperparameters for the Logistic Regression.





The logistic regression model was fit and coefficients evaluated using NumPy. A calculation of the odds ratio exhibit that the odds ratio of the feature subject sex was 1.897. The label encoder class had encoded males as 1 and females as 0. A positive relationship (greater than 1) of the odds ratio indicated that as “increases” from 0 to 1 (based on our encoding, is male) the odds of contraband being found are 1.8 times than for females. While males are more likely to be arrested for the same features, searches in males more often reveal contraband than in females, justifying male searches.

```
#Check confidence values in equation of LR
logreg.coef_
```

```
array([[-2.23173814e-04, -1.10745421e-02, -1.74829490e-01,
         6.40523796e-01,  2.78508645e+00, -1.56715692e-01]])
```

```
#Calculate the Odds Ratio
import numpy as np
np.exp(logreg.coef_)
```

```
array([[ 0.99977685,  0.98898655,  0.83960017,  1.89747451, 16.20121837,
         0.85494709]])
```

Results

An investigation of nearly half a million traffic stops across Rhode Island in the United States does not reveal conclusive evidence of discrimination. A unique perspective on working with police data, following from the structure of work done at The Stanford Open Policing Project was provided, testing for discrimination. The study concludes by contradicting initial insights that more males are arrested in Rhode Island, following from the analysis of a Logistic Regression Model built testing for justification of searches. It was found that while males are more likely to be arrested for the same features, searches in males more often reveal contraband than in females, justifying male searches. Insights into the trends in arrests in Rhode Island, controlling for factors such as zone, nature of the crime were provided.

Conclusions

This study concludes by offering recommendations for actions that need to be done for better analysis of crime data, as analysis of crime data is difficult because of the sensitive nature of the information. If more data were available, the analysis could have delved deeper into it and possibly uncovered more insights, which would have helped both the citizens and the police to work for a better environment. All countries of the world should start making data from police stops public, with controlling the publications of private features, such as name and number plates of vehicles etc. Data from individual stops should be collected whilst making the stop that should include the date and time of the stop, and attention should be paid to police officers well recording the data. An officer should record his rationale for making the stop and any proceedings after that. Heads of the departments should ensure the integrity of the data collected. Automated measures should be developed to make this process easy. Police departments should analyse this data internally for improving their functions, and also make this data publicly available for better, open policing. Such practices would help us minimise the gender gap and better the relations of police and civilians.

Future Work

A comparison with the base population of Rhode Island, though possible, could not be done due to the limitations of the scope of this project. An analysis with the base population of Rhode Island could be done to better the insights and be able to see differences better. However, census data for this purpose is not very relevant, as traditional census datasets only have data about the population of an area, not the population and statistics on drivers in that area. A comparison with the drivers that drive through Rhode Island would help us analyse for gender bias in a better way.

Acknowledgement

I would like to thank the team at Stanford Open Policing Project for making the data collected publicly available for reproducibility and letting other people conduct analysis. I thank my teacher, David Sundaram and the tutors Daniel Wilson, Shahzad Khan and Aldrich Rasso for their guidance and help with this project.

References

Angwin, J., Laeson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Becker, G. S. (2012). The economic way of looking at Life. *Revista Universidad EAFIT*, *29*(89), 7–21.

Busse, M., & Spielmann, C. (2004). Gender inequality and Trade. *GMHC Treatment Issues : The Gay Men's Health Crisis Newsletter of Experimental AIDS Therapies*, *18*(5–6), 1–2.

Chohlas-Wood, A., Goel, S., Shoemaker, A., & Shroff, R. (2018). *An Analysis of the Metropolitan Nashville Police Department's Traffic Stop Practices*. 1–10. Retrieved from https://www.nyclu.org/en/stop-and-frisk-data

Davis, E., Whyde, A., & Langton, L. (2015). *Contacts Between Police and the Public*. (October), 1–33. Retrieved from https://www.bjs.gov/content/pub/pdf/cpp15.pdf

Ed. (n.d.). There's Nothing Odd about the Odds Ratio: Interpreting Binary Logistic Regression. Retrieved from https://www.statisticssolutions.com/theres-nothing-odd-about-the-odds-ratiointerpreting-binary-logistic-regression/

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, *101*(475), 878–887. https://doi.org/10.1198/016214506000000168

Kleinbaum, D. G., & Klein, M. (2002). Logistic Regression A Self-Learning Text. In Survival

Langton, L., Durose, M., & Statisticians, B. (2013). *Police behavior during traffic and street stops*. (September 2013). Retrieved from https://www.bjs.gov/content/pub/pdf/pbtss11.pdf

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Ramachandran, V., Phillips, C., & Goel, S. (2017). *A large-scale analysis of racial disparities in police stops across the United States*. 1–10. Retrieved from http://arxiv.org/abs/1706.05678

Skeem, J., & Lowenkamp, C. (2016). *Risk, Race, & Recidivism: Predictive Bias and Disparate Impact*. Retrieved from https://dx.doi.org/10.2139/ssrn.2687339

What is the Sabarimala case? (2018). *The Indian Express*. Retrieved from https://indianexpress.com/article/what-is/what-is-the-sabarimala-case-5376596/