An Exploration of the Role of Race on Strip Searches Conducted by the Toronto Police Services

by

Shuang (Sharon) Liang, Tashfia Nidhi

A final paper submitted in conformity with the requirements

for the course of INF2178

Faculty of Information

University of Toronto

Table of Contents

1. Introduction

1.1 Overview

Since the dawn of mankind, power dynamics have played a vital role in shaping the societal practise of law enforcement. Law enforcement professionals such as police officers are vested with both immense responsibility and power to maintain societal safety and order. One of the many powers police officers hold is the power to conduct a strip search on an arrested or detained individual. In Canadian policing, a strip-search is referred to as "the removal or rearrangement of some or all of the clothing of a person as to permit a visual inspection of a person's private areas, namely genitals, buttocks, breasts… or undergarments" (Lemke, 2022). These highly intrusive police-led strip searches are usually conducted in the absence of consent of the searched person and not only does this practise strip away clothes from arrested individuals but it also strips away their dignity and autonomy. A 2019 report by the Ontario Independent Police Review Director indicated that Toronto Police Services have some of the highest strip search conducting rates in the nation as between 37% to 43% of all arrestees between 2014 and 2016 were strip searched (McNeilly, 2018). Several studies have highlighted the disproportionately high representation of Black and Indigenous ethnic demographics in datasets about police practises such as carding and strip searches. In light of these crucial race-related information about police practises, this research paper is aimed at utilizing a Toronto Police Department dataset to explore the role race plays both single handedly and concurrently with other factors in regard to the conduction of strip searches. Key findings from this research paper include the evidence of black and indigenous arrestees being more susceptible to being strip searched as well as the substantiation of factors like gender and an arrestee's behavior/actions in influencing their likelihood of being strip searched.

1.2 Literature Review

Research on racial and gender biases in policing practises has been conducted since decades, yet, the disparities and disproportionate representation and exploitation have persisted. Canada is no stranger to these sorts of statistics. Various studies have led to the findings that indigenous group (particularly females) and black racial demographic face the highest rates of intrusive police

practises such as strip searches. “Reforming the Strip Search in Canadian Law and Practice” by Benjamin L. Berger (2009) found that strip searches are frequently used by police in Canada even in cases where it’s not legally justified. In 2016, the Ontario Human Rights Commission (OHRC) released a report on the over-representation of Indigenous peoples in Ontario’s criminal justice system. The report also found that Indigenous peoples in Ontario are over-represented in strip search statistics, and that they are more likely to be subjected to a strip-search than non-Indigenous people (Robinson, 2016). The gruesome statistics associated with strip searches are not limited to Ontario. A 2018 report by BC Civil Liberties Association found that strip-searches were still being conducted in a degrading and inhumane manner for women (Mazur, 2021). A 2018 Toronto Star report highlighted that despite Black people making up 8.3% of Toronto’s demographic between 2010-2017, they accounted for 25.8% of all carding incidents, thus highlighting another police practise where there is evidence of discrimination (Rankin & Gillis, 2018).

Aside from demographic biases in police practises for parameters like racial group and gender, ethnographic studies have revealed that a suspect’s resistance and demeanor toward police behavior does influence police behavior towards them. However, hostility and demeanor towards police authority is another complex phenomenon that is heavily correlated with racial and socio-economic factors (Engel, 2009).

1.3 Dataset Description

This project utilizes a dataset openly released by the Toronto Police Service in November 2022 containing various information on the arrests and strip searches for 65 276 arrested individuals. This data was collected over the span of 2020 to 2021. There are a total of 24 attributes in the dataset, including demographic identifiers like perceived race, age category, sex as well as other information about the arrest such as the occurrence category, search reasons, actions at arrest and whether or not the booked individual was strip-searched. The format of data varies across the various variables; binary variables are used to indicate booking, strip search events, the different search reasons and actions at arrest. Categorical text variables are utilized for all demographic identifiers and occurrence categories. ID variables, arrest location, and the year of arrest were conveyed via numerical variables.

1.4 Research Questions

The data analysis conducted during the midterm revealed significant differences in police treatment between different racial groups. These findings alongside the background research that revealed indigenous and black demographics to experience higher rates of intrusive police practises led to this paper centering around the role of perceived race on strip search rates. However, midterm findings and EDA also did provide preliminary insights that there are other factors aside from race that has the potential to influence an arrestee's likelihood of being strip searched and this also served as an area of exploration. For the scope of this paper, the two specific research questions being answered are:

1) Is there a difference in the strip search rates (proportion) between the perceived race when controlling for the effect of actions at arrest?
2) How do gender, number of actions at arrest and racial group (black + indigenous vs. all other races) influence an arrested individual's likelihood of being strip searched?

2. Exploratory Data Analysis

2.1 Descriptive Statistics - General

To gain a better understanding of the processed and cleaned dataset, a summary chart (figure 1) showing the number of values for several non-numerical variables was generated. As evident in Figure 1, there is a total count of 65 074 data points, 8 different racial categories, 2 different gender groups, 7 age-group categories and 31 occurrence causes.

	Perceived Race	Sex	Age Group at Arrest	Occurrence Category
Count	65074	65074	65074	65074
Unique	8	2	7	31
Top	White	M	Aged 25 to 34 years	Assault
Frequency	27630	52499	20899	7724

Figure 1: Summarized Table of Non-Numerical Variables

The variable of key interest in this dataset is regarding the practice of strip-searches by the Toronto Police Department. Figure 2 (below) illustrates the number of data points (arrest events) that were associated with a strip searches event (represented by 1 on the x-axis) and those that weren't (represented by 0).

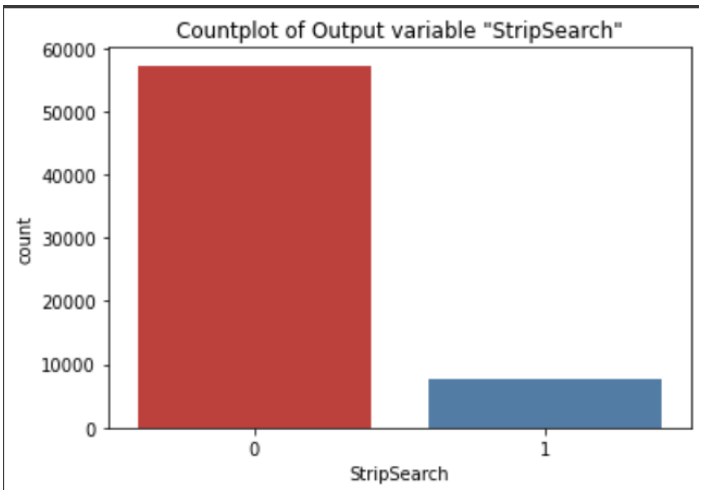

Figure 2: Count-plot of Strip Search Events

As illustrated in Figure 2, the vast majority of arrestees included in this dataset didn't undergo strip-searches. Less than 10 000 arrest events in this dataset led to the conduction of strip searches. While this plot may cause the number of people strip searched to appear relatively low, background research revealed that the Toronto Police Department has one of the highest rates of strip search practise in the country.

2.2 Race-related EDA

The most prominent theme that emerged based on background research revolved around the role of race in influencing police practice. This led to the selection of the variable 'perceived race' to be the pivotal variable of interest in the midterm data analysis. The EDA is aimed at better understanding the racial makeup and relationships with variables within the particular dataset as it relates to strip searches.

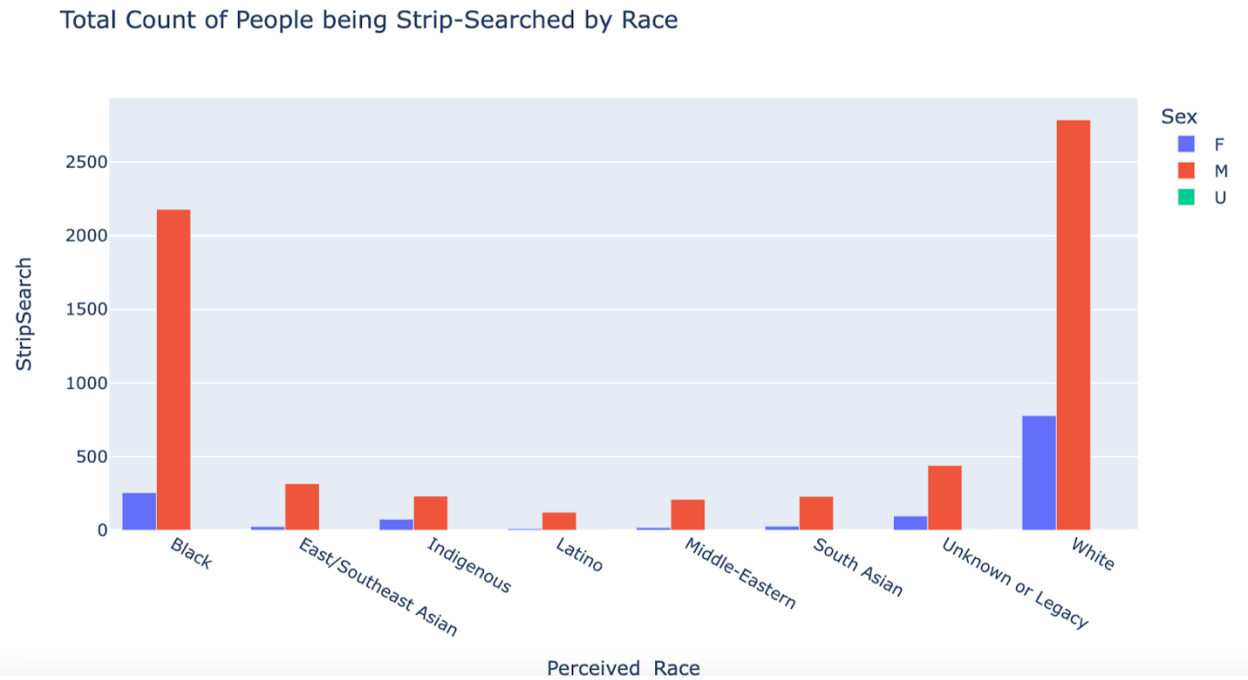

Figure 3: Bar Plot Illustrating the total count of strip searches conducted per perceived race by gender

Figure 3 illustrates the total counts of strip searches conducted under each different perceived racial group and gender after being grouped for the respective variables. It can be observed that the number of males being strip-searched is higher than the number of females across all racial groups. This is in line with Toronto crime statistics which show males committing a greater percentage of crimes than females. The perceived race that had the highest strip search counts was White, followed by Black while Latino had the lowest total counts for both genders. Keeping Toronto's population demographic in mind where white ethnicity is the largest ethnic group, it is no surprise that the perceived 'white' race has the highest strip search count. The 2nd largest perceived race represented in this dataset is 'Black' which doesn't align with the city of Toronto's ethnic makeup where ethnic demographics such as East Asians and South Asians have a greater population than Black.

These discrepancies and racial disproportions led to further exploration of the racial relationship associated with strip searches.

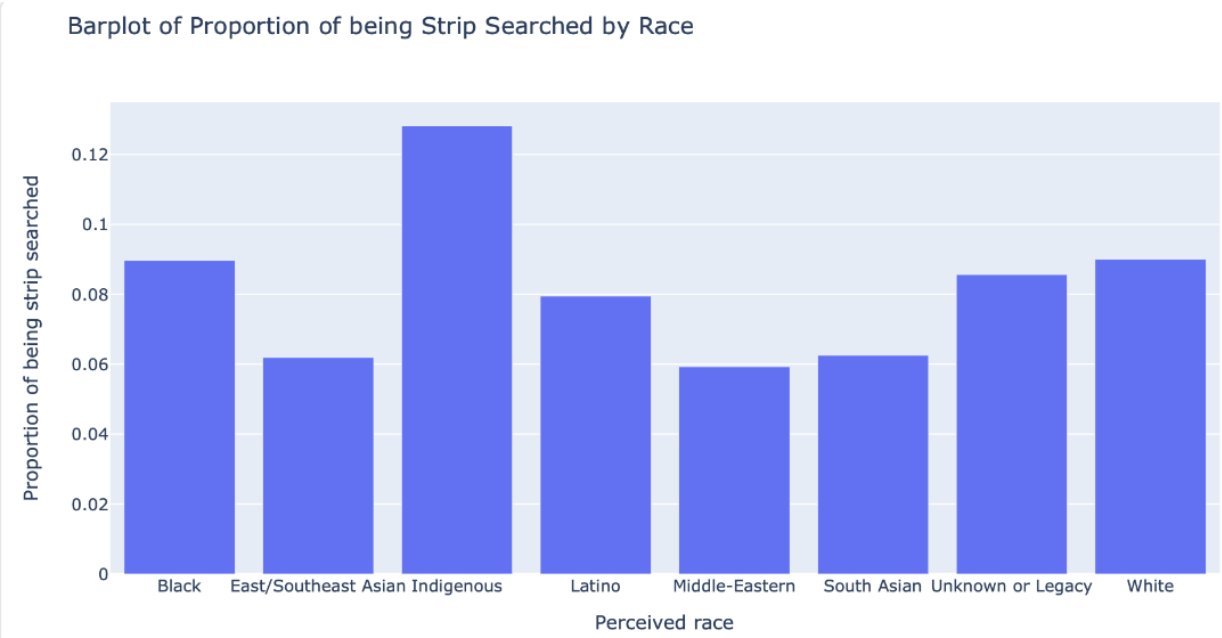

Figure 4: Bar Plot displaying the mean proportions of arrested individuals that were strip-searched by race

Looking at the mean proportions of arrested individuals that were strip-searched for each of the racial groups provides a comparative insight into the likelihood of an individual being strip-searched based on their race. Surprisingly, neither White nor Black racial groups had the highest proportion, but rather it is the Indigenous group that has the highest proportion of being strip searched for both genders. This is in line with the background research that revealed that not only in the present but even historically, Indigenous individuals have been one demographic that experienced overtly greater rates of intrusive and dehumanizing policing practises. The strip search proportions for White and Black demographics are quite similar and display the 2nd and 3rd highest strip search proportions according to this dataset.

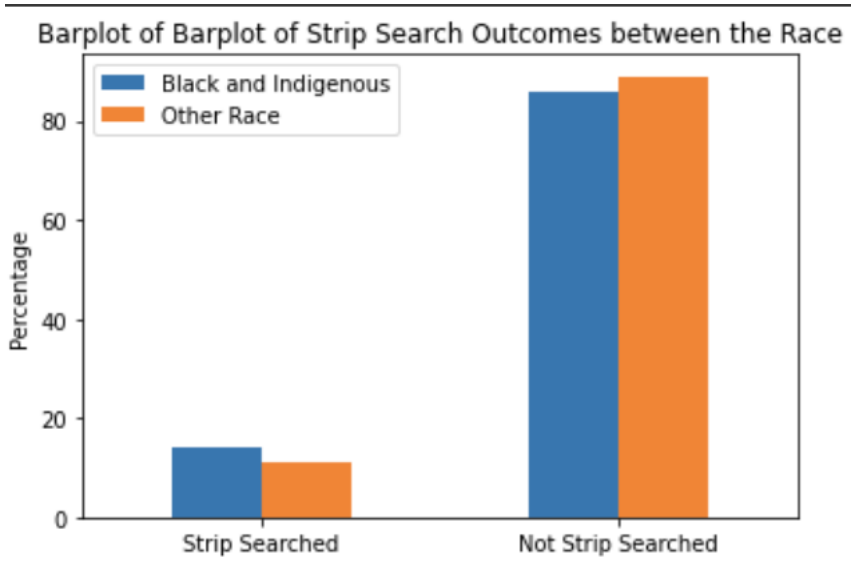

Figure 5: Barplot displaying the percentages of arrestees strip searched based on racial grouping

Background literature indicated Black and Indigenous groups to be the 2 primary racial groups that experience disproportionately higher rates of policing, and this bar plot supports this notion. The percentage of black and indigenous arrestees that were strip-searched is greater than the percentages of arrestees who were strip-searched that had any other perceived race.

These preliminary racial differences led to the conduction of some T-tests to explore whether these differences are statistically significant.

2.2.1 Welch's T-test: Strip search proportions between Indigenous background and all other racial groups

Based on Canada's longstanding history of authoritative injustices against Indigenous as well as preliminary EDA which illustrated a higher proportion of Indigenous arrestees being strip-searched, this Welch's T-test is aimed at checking whether there's a statistically significant difference between the mean proportion strip-searched between Indigenous arrestees and arrestees of the other 7 perceived races. The hypothesis that are being tested are:

H0 (Null Hypothesis): The mean strip-search proportions between indigenous individuals is the same as the combined mean strip-search proportions for all other racial groups.

HA (Alternative Hypothesis): There is a statistically significant difference in the mean strip-search proportions between Indigenous individuals and individuals of other races.

This Welch's T-test results indicate a p-value of 0.00104 which is lower than a critical value of 0.05 (95% confidence interval). This leads to the rejection of the null hypothesis. This indicates that there indeed is a statistically significant difference in the mean proportion of arrestees strip searched for Indigenous individuals (M=0.128, SD=0.234) and all other racial groups (M=0.0767, SD = 0.163).

2.2.2 Welch's T-test: Strip search proportions between Black + Indigenous groups and all other

Since Figure 6 displayed a higher percentage of Black + Indigenous arrestees being strip searched in comparison to all other racial groups, a Welch's T-test was conducted to check whether there's a statistically significant difference between the mean proportion of arrestees strip-searched between Black + Indigenous arrestees and arrestees of the other 6 perceived races. The hypothesis that are being tested are:

H0 (Null Hypothesis): The mean strip-search proportions between black + indigenous individuals is the same as the combined mean strip-search proportions for all other racial groups.

HA (Alternative Hypothesis): There is a statistically significant difference in the mean strip-search proportions between black + Indigenous individuals and individuals of other races.

This Welch's T-test results indicate a p-value of 0.000333 which is lower than a critical value of 0.05 (95% confidence interval). This leads to the rejection of the null hypothesis. This indicates that there indeed is a statistically significant difference in the mean proportion of arrestees strip searched for Black + Indigenous individuals (M=0.105, SD=0.188) and all other racial groups (M=0.0742, SD = 0.165).

2.2.3 One-Way ANOVA – Research Question from Midterm

Since preliminary T-tests and graphs have been elucidating towards the existence of racial biases existing in the Toronto Police Department's strip-search practices, a one-way ANOVA was conducted (midterm) to validate this. In this Anova, the perceived race behaved as the categorical predictor variable and the strip search counts behaved as the continuous response variable. The findings of the ANOVA revealed insights into whether any racial differences in strip search counts are statistically significant.

2.2.3.1 About the Categorical Predictor Variable

There are 8 different racial groups under the perceived race variable (the predictor variable). The 8 races are White, Black, East/Southeast Asian, Indigenous, Middle Eastern, Latino, South Asian and Unknown.

2.2.3.2 Hypothesis Testing

The following hypotheses are being tested:
H0 (Null hypothesis 1): There is no difference in the mean strip search counts between the different perceived racial groups.

H1 (Alternate Hypothesis 1): There is a difference in the mean strip search counts between at least 2 of the different perceived racial groups.

2.2.3.3 Checking Assumptions

Before conducting the Anova, the homogeneity of variance assumption was checked which revealed a ratio of 1.43. Since $1.43 < 2.00$, this variance homogeneity assumption is met.

2.2.3.4 ANOVA Results

	Degrees of Freedom	Sum of Square	Mean Square	F-value	P-value of the F statistic
C(Perceived_Race)	7.0	23815.384566	3402.197795	24.160518	4.749578e-32
Residual	2565.0	361194.128066	140.816424	NaN	NaN

Figure 6: Summarized Table of One-Way Anova Results

As illustrated in Figure 6, the degree of freedom in this particular Anova test was 7 and an F-value of 24.16 was attained. This relatively large F value means that the ratio of explained variance to unexplained variance is greater which elucidates towards the rejection of the null hypothesis. A p-value of 4.75e-32 was also attained which is the p-value associated with this particular F-distribution. As this value is much lower than a critical P value of 0.05 (indicating a 95% confidence interval), it can be said that there is a statistically significant difference in the mean strip search counts between at least 2 of the different perceived racial groups.

2.2.3.5 Note about Race-Related EDA

The data analysis (T-tests and One-Way ANOVA) conducted in the midterm revealed evidence for the existence of racial differences in strip searches conducted by the Toronto Police Department. The background research indicated these racial discrepancies in police practises primarily impact indigenous and black demographics at higher rates and in a negative manner. Thus far, this has been supported by further T-tests. However, it is essential to note recognize that police practices such as the conduction of a strip search are multi-factorial and usually the decision to conduct a strip search on someone is influenced by many elements and not just race itself.

2.2.4 Two-Way ANOVA - Research Question from Midterm

This Two-Way Anova (conducted for the midterm) was performed with the aim of exploring the combined effect of perceived race and occurrence category on strip search counts, a 2-way Anova was conducted. For this 2-way Anova, the racial group and occurrence category behaved

as the two categorical predictor variables while the strip search counts behaved as the continuous response variable.

2.2.4.1 About the Categorical Predictor Variables

-Perceived race → There are 8 different racial groups under the perceived race variable (the predictor variable).
-Occurrence category → There are 31 different types of occurrences under the occurrence category variable.

2.2.4.2 Hypotheses testing

There are 3 sets of hypotheses being tested with this two-way ANOVA. They are:

Racial groups

H0 (Null hypothesis 1): There is no difference in the mean strip search counts between the different perceived racial groups.

H1 (Alternate Hypothesis 1): There is a difference in the mean strip search counts between at least 1 of the different perceived racial groups.

Occurrence category:

H0 (Null hypothesis 2): There is no difference in the mean strip search counts between the different occurrence categories.

H1 (Alternate Hypothesis 2): There is a difference in the mean strip search counts between at least 1 of the occurrence categories.

Interaction:

H0 (Null hypothesis 3): There are no statistically significant differences in the mean strip count for the different perceived races that can be explained by the occurrence category and vice versa. There is no interaction.

H1 (Alternate hypothesis 3): There is at least one statistically significant difference in the mean strip count for the different perceived races that can be explained by the occurrence category and vice versa. There is interaction.

2.2.4.3 ANOVA Results

	Sum of Square	Degrees of Freedom	F-value	P-value of the F statistic
C(Perceived_Race)	25824.927551	7.0	33.041153	4.619568e-44
C(Occurrence_Category)	45449.806286	30.0	13.568283	1.042411e-61
C(Perceived_Race):C(Occurrence_Category)	56141.839924	21.0	2.394316	1.412906e-22
Residual	259602.481856	2325.0	NaN	NaN

Figure 7: Summarized Two-Way Anova Results

The perceived race predictor variable had a degree of freedom of 7, an F statistic of 33.04 and a p-value of 4.62e-44. As this p-value is much lower than the critical p-value of 0.05 (indicating a 95% confidence interval), the null hypothesis is rejected. Therefore, it can be said that there is a statistically significant difference in the mean strip search counts between at least 2 of the different perceived racial groups.

The occurrence category predictor variable had a degree of freedom of 30, and an F value of 13. 57 and a p-value of 1.04e-61. This relatively large F value means that the ratio of explained variance to unexplained variance is great which elucidates towards the rejection of the null hypothesis. Since the p-value is also lower than the critical p-value of 0.05 (indicating a 95% confidence interval), the null hypothesis is rejected. Therefore, it can be said that there is a statistically significant difference in the mean strip search counts between at least 2 of the different occurrence categories.

For the interaction term, the degree of freedom is 210, the F value is 2.39 and the p-value is 1.41e-22. As this p-value is lower than 0.05, the null hypothesis is rejected. Therefore, it can be said that there is an interaction effect occurring where the mean strip search count for one of the two predictor variables is influenced by the other predictor variable. This ANOVA finding is pivotal in understanding the interrelated role racial categories and occurrence category plays together in influencing an arrested individual's likelihood of being strip-searched.

2.2.4.4 Interaction Effect Plot

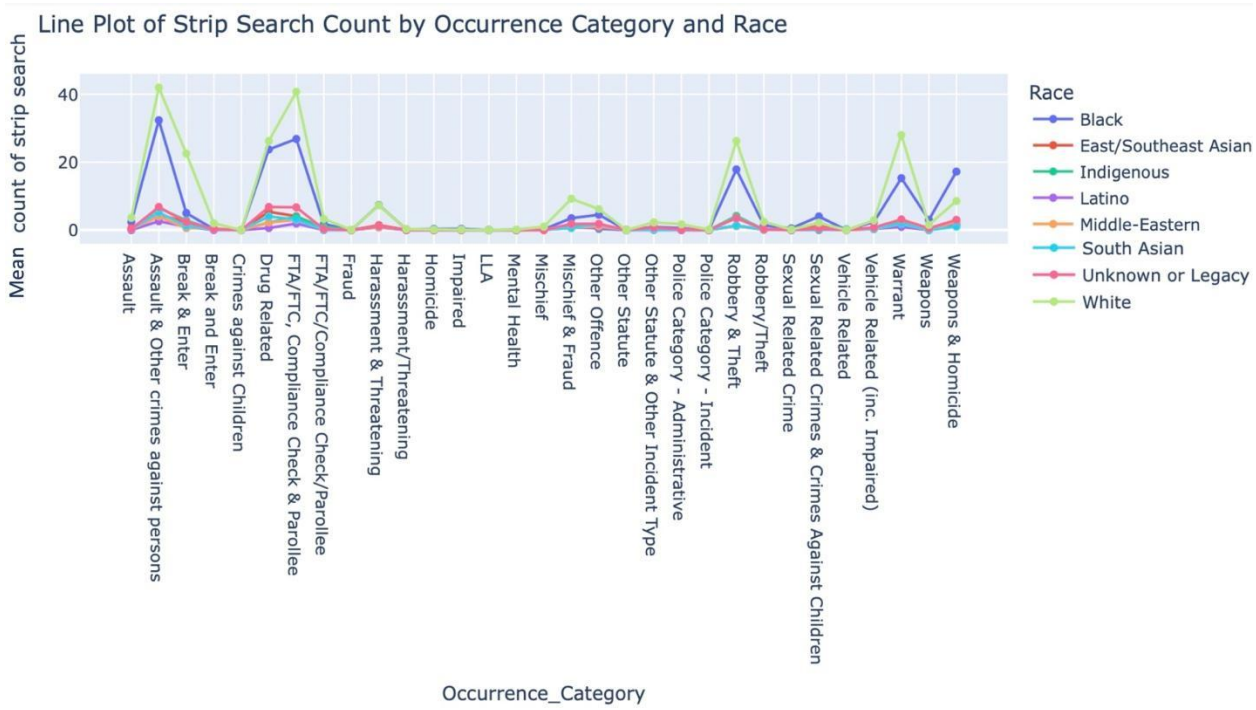

Figure 8: Interaction Effect Plot for Perceived Race and Occurrence Category with respect to Mean strip search count

A parallel line on the interaction plot indicates no interaction present while an intersecting point elucidates the presence of an interaction. In this plot, the highest interaction between perceived race and occurrence categories can be observed when the demographic is black and the occurrence category is Assault & other crimes against persons. This means that the highest number of strip searches were conducted when a Black individual was arrested for Assault or other crimes against persons or for FTA/FTC Compliance Check & Parollee. There is also

notable interaction when the demographic is white, and the occurrence category is assault and other crimes against persons.

This two-way ANOVA illustrated the interaction effect and relationship that can be present when talking about the conduction of something like strip-searches which are multifactorial in nature. Throughout this paper, a goal was to explore the relationships and role other predictor variables play alongside race in influencing an arrestee's likelihood of being strip-searched.

2.3 Actions-at-Arrest related EDA

As some exploration of the relationship between occurrence category and race on strip search counts have been explored in the midterm, we wanted to explore some other variable that provides valuable insight into the likelihood of someone being strip searched. Resistance and demonstration of offensive actions to police arrest is something that influences the harshness of the police action taken against the arrestee. Due to this, it is worth investigating the role resistance from arrestees plays in influencing their likelihood of being strip searched. Under the constraints of this data set, resistance will be represented by the proportion of total possible actions at arrest that were demonstrated by the arrestee.

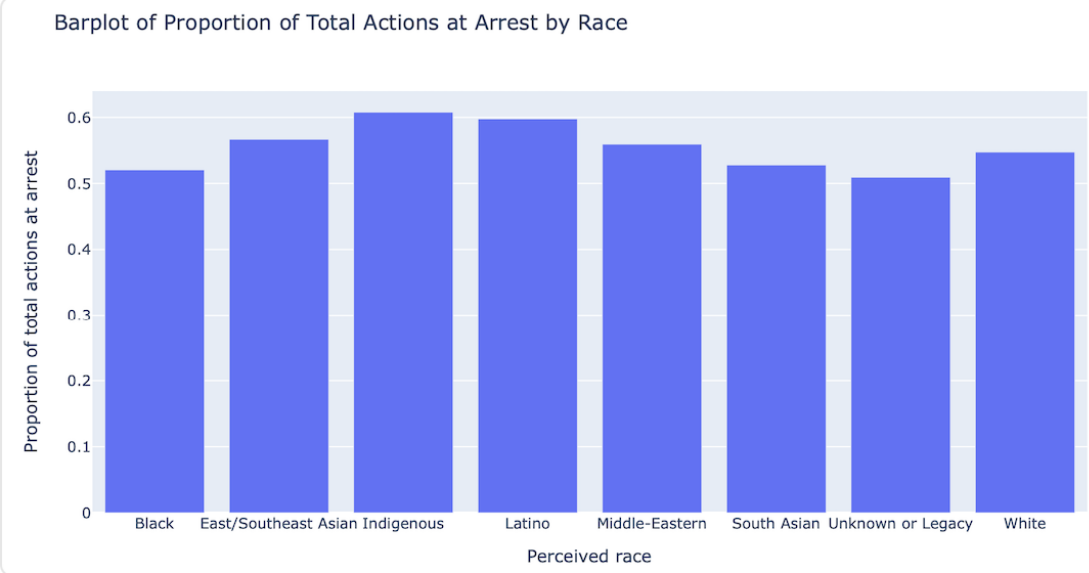

Figure 9: Mean Proportion of Total Actions-at-Arrest By Racial Groups

Figure 9 illustrates the average proportion of total possible actions-at-arrest that were conducted by each perceived race. This graph shows Indigenous arrestees to have the highest proportion (which reflects resistance) of total possible actions at arrest. This is interesting to note since Figure 5 also illustrated Indigenous arrestees experience the highest rate of the strip search. However, while Figure 5 and other EDA conducted illustrated Black demographics to also experience disproportionately greater strip search rates relative to other races, figure 12 also shows black arrestees to have one of the lower mean proportions of total possible actions-at-arrest. These findings raise some interesting notions about the multifactorial relationship between resistance, race and police practices.

These preliminary observations will be further explored by some T-tests.

2.3.1 Welch's T-test: Proportions of Total possible actions-at-arrest between the Indigenous group and all other perceived races

As a result of the observations from Figure 12, a Welch's T-test was conducted to check whether there's a statistically significant difference between the mean proportion of total possible actions at arrest between Indigenous arrestees and arrestees of the other perceived races. The hypothesis that are being tested are:

H0 (Null Hypothesis): The mean proportion of actions-at-arrest between indigenous individuals is the same as the mean proportion of actions-at-arrest for all other racial groups.

HA (Alternative Hypothesis): There is a statistically significant difference in the mean proportion of actions-at-arrest between indigenous individuals and individuals of other races.

This Welch's T-test results indicate a p-value of 0.0219 which is lower than a critical value of 0.05 (95% confidence interval). This leads to the rejection of the null hypothesis. This indicates that there indeed is a statistically significant difference in the mean proportion of actions-at-arrest between indigenous individuals (M=0.608, SD=0.412) and all other racial groups (M=0.545, SD = 0.300).

2.3.2 Welch's T-test: Proportions of Total possible actions-at-arrest between the Black + Indigenous group and all other perceived races

Due to literature findings, a considerable amount of the EDA conducted thus far grouped Indigenous and Black demographics together and compared a variable of interest to other racial groups. The same approach was taken to determine where there's a statistically significant difference between the mean proportion of total possible actions at arrest between Indigenous + Black arrestees and arrestees of the other perceived races. The hypothesis that are being tested are:

H0 (Null Hypothesis): The mean proportion of actions-at-arrest between black + indigenous individuals is the same as the mean proportion of actions-at-arrest for all other racial groups.

HA (Alternative Hypothesis): There is a statistically significant difference in the mean proportion of actions-at-arrest between black + indigenous individuals and individuals of other races.

This Welch's T-test results indicate a p-value of 0.745 which is higher than a critical value of 0.05 (95% confidence interval). This leads to a failure to reject the null hypothesis. This indicates that there is no statistically significant difference in the mean proportion of actions-at-arrest between black + indigenous individuals (M=0.554, SD=0.331) and all other racial groups (M=0.549, SD = 0.307).

However, there are some other factors that could've potentially contributed to this failure to reject the null hypothesis. For example, while indigenous groups have the highest proportion of total possible actions at arrest conducted, they have the lowest total number of actions-at-arrest conducted (sum of all indigenous arrestees' actions at arrest counts). For black arrestees, the opposite is observed as they have lower proportions of total possible actions at arrest but one of the highest total summed number of actions-at-arrest conducted. The grouping of these two demographics as a result could contribute to a flawed t-test.

2.4 Gender-based EDA

Gender is another variable that could potentially provide interesting information about another demographic predictor influencing strip search likelihoods. Background research also shone light on how females from minority groups such as Indigenous and black demographics are more susceptible to experiencing intrusive police practices.

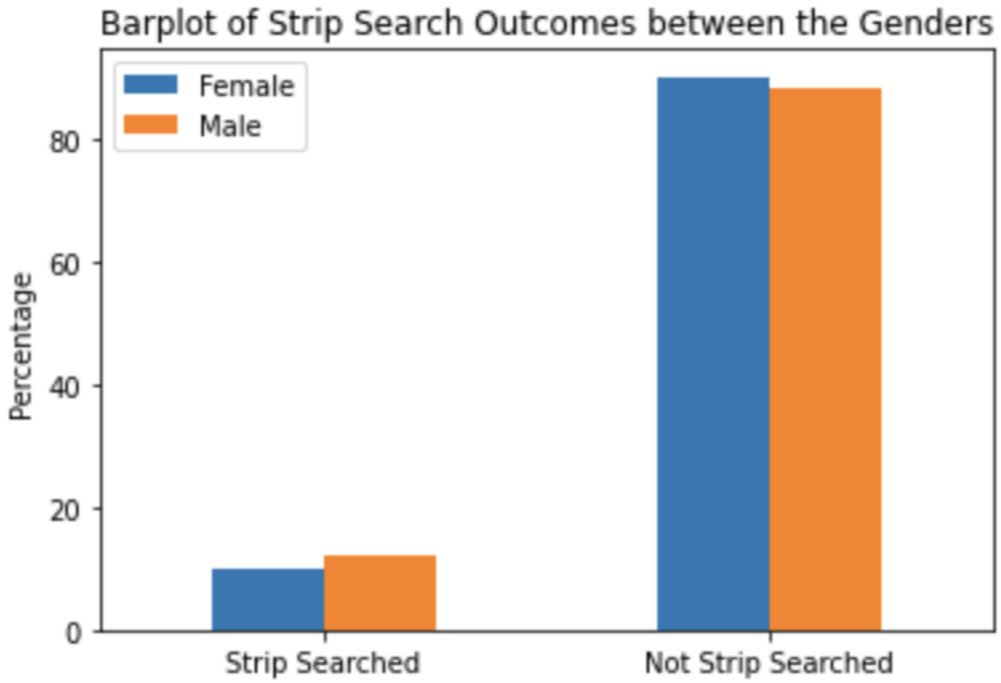

Figure 10: Bar plot of Strip Search Outcomes between the Genders

Based on Figure 10, it can be observed that a slightly greater proportion of males were strip-searched in comparison to females. According to background literature, there generally tends to be a greater percentage of male representation in crime-related statistics which is supported by Figure 3 which illustrates how the number of strip searches conducted in this dataset was greater for males across all racial groups.

2.4.1 Welch's T-test: Strip search proportions between male and female

The gender-related observations discussed above led to the conduction of this Welch's T-test which is aimed at checking whether there's a statistically significant difference between the mean

proportion strip-searched between male and female groups. The hypothesis that are being tested are:

H0 (Null Hypothesis): The mean strip-search proportions between male arrestees is the same as the mean strip-search proportions for females.

HA (Alternative Hypothesis): There is a statistically significant difference in the mean strip-search proportions between male and female arrestees.

This Welch's T-test results indicate a p-value of 0.0464 which is lower than a critical value of 0.05 (95% confidence interval). This leads to the rejection of the null hypothesis. This indicates that there indeed is a statistically significant difference in the mean proportion of arrestees strip searched for males (M=0.0874, SD=0.165) and females (M=0.0733, SD = 0.180).

2.5 Remarks about EDA

The series of T-tests and ANOVAs conducted for the midterm revealed the presence of a racial influence on the likelihood of being strip searched. Further EDA aimed towards zooming into racial groups (black + indigenous, all others) provided more evidence for not only strip search rates but also for other noteworthy parameters such as actions at arrest. These weren't surprising as background research did reveal racial groups (especially black and indigenous groups) to be impacted by racially biased treatment from police officers. To further confirm and consolidate the findings of these preliminary tests, race will be explored in tandem with other factors, such as gender, and actions at arrest, that can influence police conductance of strip searches.

2.6 Statistical Power Analysis

Prior to conducting the ANCOVA and logistic regression model, it is imperative to ensure that the appropriate sample size is chosen for desired effect size. A Cohen's D for the two major racial groups (Black + Indigenous, All others) that were utilized in the logistic regression model had a value of 0.0935 when alpha was set at 0.05 and the statistical power was set at 0.8. To

ensure a high statistical power (0.8) was achieved, there is a necessity for a sample size of 3009.34 for the “All other races” perceived race group and 1279.43 for the “Black + Indigenous” group. It was ensured that the sample size for these 2 groups in the Logistic regression models exceed these values.

	Suggested sample size	Actual Sample size
Black and Indigenous Group	1279.434	19413
All other racial group	3009.336	45661

Figure 11: Summary of Statistical Power Analysis for Desired Power of 0.8

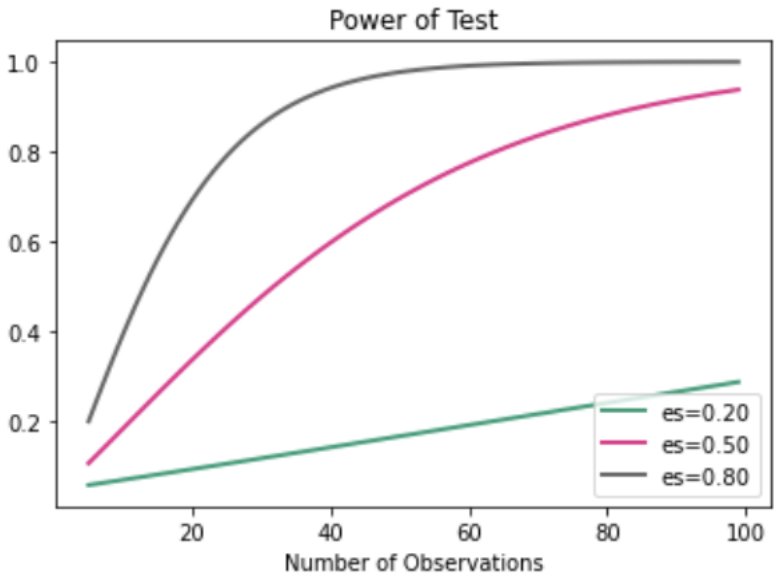

Figure 12: Sample Power Curve

As illustrated in figure 12, there is a relationship between sample size, effect size and statistical power. With an increase of observations, an increased statistical power and effect size can be observed. For the purpose of this paper, to ensure a large effect size, the sample chosen exceeds the suggested sample sizes thus allowing for high statistical power. A higher statistical power is crucial in ensuring that accurate conclusions about a population are drawn from the sample data analyzed.

3. Method

Based on the exploratory data analysis, background research, midterm findings and personal interest, the primary focus of this paper will continue to be on the relationship between race and strip search conduction while also taking into account various other influential parameters. Prior to delving into the specific research question, a power analysis was conducted to determine the appropriate sample size necessary for desired effect size. For each statistical test conducted, the underlying assumptions were checked. The table below summarizes the research questions and the specific methods utilized to answer each research question.

Research Question	**Method**
Is there a difference in the strip search rates (proportion) between the different perceived races of arrestees when controlling for the number of actions at arrest?	- One Way ANCOVA - DV: strip search proportions - IV: perceived race - Covariate: proportion of total number of actions at arrest demonstrated by arrestee
How do gender, number of actions at arrest and racial group (black + indigenous vs. all other races) influence an arrested individual's likelihood of being strip searched?	- Binary Logistic Regression - DV: Being strip-searched (yes or no) - IV_1: racial group (black + indigenous, all others) - IV_2: gender (male, female) - IV_3: Total number of actions at arrest - A prediction model was created and accuracy parameters were determined - Random under-sampling was used for splitting the dataset due to class imbalance problem

Figure 13: Table of Research Questions and selected method of exploration

4.Results

3.1 Research Question 1

3.1.1 One-Way ANCOVA Overview

Since the aim of the first research question is to explore the difference in strip search rates between the difference in perceived race of arrestees when the effect of total actions at arrest is being controlled, a one-way ANCOVA was conducted to answer the question. In this ANCOVA, the continuous dependent variable was the proportion of arrestees strip-searched, the predictor categorical variable was the different perceived racial groups of arrestees and the continuous covariate/control was the proportion of the total possible actions at arrest that the arrestee displayed.

3.1.2 One-Way ANCOVA Assumption Checks

There were several assumptions that were checked to ensure the validity of the ANCOVA results. The Shapiro test was conducted to ensure that the residuals followed a normal distribution. With a Shapiro score of 0.39 and a p-value of 0, this assumption was not met. Another assumption for ANCOVA is the homogeneity of regression slopes (referring to the interaction between the covariate and independent variable), and this assumption was met. The 3rd assumption is to ensure the absence of multicollinearity. The VIF values were determined and these values didn't exceed 5, thus satisfying the assumption. Assumption 4 is regarding sample size sufficiency which was met as the sample size for this study exceeded 500.

3.1.3 About the Categorical variable

-Perceived race → There are 8 different racial groups under the perceived race variable (the predictor variable).

3.1.4 Hypothesis Testing

Null Hypothesis (H_0): There is no statistically significant difference in the proportion of arrestees being strip searched between the different races, after controlling for the effect of the arrestee's actions at arrest.

Alternative Hypothesis (H_A): There is a statistically significant difference in the proportion of arrestees being strip searched between the different races, after controlling for the effect of the arrestee's actions at arrest.

3.1.5 ANCOVA Results

Factor Names	Sum of Squares	Degrees of Freedom	F-values	Uncorrected P-values	Partial Eta-squared
Perceived Race	0.932664	7	4.584296	0.000042	0.012549
Proportion of total actions at arrest	0.223572	1	7.692432	0.005586	0.003037
Residual	73.386466	2525	NaN	NaN	NaN

Fig 14: Summarized One-Way ANCOVA Results Table

For the perceived race variable, a degree of freedom of 7 was obtained indicating the presence of 8 groups under the variable. An F-value of 4.58 and an uncorrected p-value of 0.000042 is observed for perceived race which is lower than the alpha threshold of 0.05, thus leading to the rejection of the null hypothesis. This indicates that there is a significant effect of perceived race on the proportion of arrestees strip-searched when controlling for the total number of actions at arrest. An np2 (partial eta-squared) value of 0.0125 is observed which indicates that 1.25% of the variance in the mean proportion of people strip-searched can be explained by the perceived race after controlling for actions at arrest. The sum of squares associated with perceived race is 0.933.

For the proportion of total possible actions at arrest, the degree of freedom is 1, the F value is 7.69 and the uncorrected p-value is 0.0056. This leads to the rejection of the null hypothesis as the p-value is lower than 0.05, meaning that there is a statistically significant effect of the covariate (proportion of an arrestee's total actions at arrest) on the strip search rates of arrestees. The np2 (partial eta-squared) value is 0.00304 indicating that 0.304% of the variance in mean strip search proportions can be explained by the actions at arrest when controlling for race.

3.2 Research Question 2

3.2.1 Logit Regression Overview

After conducting EDA on several different potential predictor variables, the aim of this logistic regression model is to put together these different elements together and evaluate how they interact together to influence an arrestee's likelihood of being strip-searched. The specific research question this regression model is aimed at addressing is: How do gender, actions at arrest and perceived race influence an arrested individual's likelihood of being strip searched? The dataset was split into an 80-20 split for training and testing purposes and a logit regression model was generated for both the training and testing dataset. Since the EDA revealed that there was a class imbalance of data for the outcome of interest, a random-under-sampling method was used to split data points into train and test sets.

3.2.2 About the variables

In this regression model, there are 3 predictor variables. The predictor variable perceived race is a categorical variable consisting of 2 groups; black + indigenous, and all other races. Gender is another categorical predictor variable which also consists of 2 groups; male and female. The final predictor variable is the proportion of total possible actions at arrest demonstrated by an arrestee and this is a continuous variable. These first 2 variables were chosen as they would allow for crucial revelations regarding the role of demographic identifiers in influencing police practices such as strip searches whereas the 3rd variable (proportion of total possible actions at arrest demonstrated) represents resistance to police authority which undoubtedly has practical implications guiding a police officer's treatment of an arrestee. The dependent variable in this logit model is a binary outcome to represent a strip search event (1 = strip search conducted, 0 = strip search not conducted).

3.2.3 Logit Regression Results for Training Dataset

Dependent Variable:	StripSearch	**Number of Observations:**	12478
Model:	Logit	**Df Residuals:**	12474
Method	MLE	**Df Model:**	3
Date:	Wed, 12 Apr 2023	**Pseudo R-squared:**	0.01147
Time:	00:08:42	**Log-Likelihood:**	-8549.7
Converged:	True	**LL-Null:**	-8549.0
Covariance Type:	nonrobust	**LLR p-value:**	8.981e-43

	Coefficient	Standard Error	Z-scores	P-values	[0.025	0.975]
Intercept	-0.4097	0.047	-8.636	0.000	-0.503	-0.317
total_actions_at_arrest	0.3208	0.029	11.022	0.000	0.264	0.378
Sex_Male	0.1326	0.047	2.804	0.005	0.040	0.225
Black_and_Indigenous	0.3040	0.039	7.804	0.000	0.228	0.380

Figure 15: Summarized Logit Regression Results for Training Dataset

For the training data for this logit regression model, the degree of freedom of the residuals is 12474. There is also a Pseudo R-squared value of 0.01147 which is relatively low thus indicating a poorly fitted model for the data set.

The coefficients, standard error, confidence intervals and statistical significance for the intercept and each predictor variable are also noted in Figure 15. All features used in this model have a p-value that's lower than 0.05, thus indicating that each feature is statistically significant in influencing the likelihood of a strip search event. The intercept coefficient is – 0.4097, and the standard error is 0.047. For the total actions at arrest, a coefficient of 0.3208 and a standard error

of 0.029 is observed. For gender, there is a coefficient of 0.1326 and a standard error of 0.047. For the perceived race, the black and indigenous category was set as the default and it had a coefficient of 0.3040 and a standard error of 0.039.

For a more efficient interpretation of the coefficients, the odds ratio for each of the predictors was determined.

Category	Odds Ratio
Intercept	0.652482
Total Actions at Arrest	1.350582
Male Gender	1.192495
Black and Indigenous	1.324080

Figure 16: The Odds Ratio for Training Dataset

As evident in Figure 16, total actions at arrest has an odds ratio of 1.35 which indicates that when other predictor variables are held constant, a one-unit increase in the proportion of total actions at arrest will increase the odds of being strip-searched by a factor of 1.35. For gender (male), the odds ratio is 1.19, meaning that when all other predictors are held constant, males are about 1.19 times more likely than females to be strip searched. For the perceived race, when other predictors are held constant, black or indigenous arrestees are 1.32 times more likely to be strip searched.

3.2.4 Logit Regression Results for Test Dataset

Dependent Variable:	StripSearch	**Number of Observations:**	3120
Model:	Logit	**Df Residuals:**	3116
Method	MLE	**Df Model:**	3
Date:	Thu, 13 Apr 2023	**Pseudo R-squared:**	0.009808
Time:	15:42:38	**Log-Likelihood:**	-2140.9
Converged:	True	**LL-Null:**	-2162.2
Covariance Type:	nonrobust	**LLR p-value:**	3.280e-09

	Coefficient	Standard Error	Z-scores	P-values	[0.025	0.975]
Intercept	-0.4141	0.096	-4.295	0.000	-0.603	-0.225
total_actions_at_arrest	0.2462	0.058	4.236	0.000	0.132	0.360
Sex_Male	0.2347	0.095	2.461	0.014	0.048	0.422
Black_and_Indigenous	0.3178	0.077	4.113	0.000	0.166	0.469

Figure 17: Summarized Logit Regression Results for Test Dataset

For the training data for this logit regression model, the degree of freedom of the residuals is 3116 as a total of 3120 data points were used (20% of overall data). There is also a Pseudo R-squared value of 0.0098 which is relatively low thus indicating a poorly fitted model for the data set.

The coefficients, standard error, confidence intervals and statistical significance for the intercept and each predictor variable are also noted in Figure 17. All features used in this model have a p-value that's lower than 0.05, thus indicating that each feature is statistically significant in influencing the likelihood of a strip search event. The intercept coefficient is – 0.414, and the standard error is 0.096. For the total actions at arrest, a coefficient of 0.2462 and a standard error of 0.058 is observed. For gender, there is a coefficient of 0.2347 and a standard error of 0.095.

For the perceived race, the black and indigenous category was set as the default and it had a coefficient of 0.3178 and a standard error of 0.077.

For a more efficient interpretation of the coefficients, the odds ratio for each of the predictors was determined for the test data as well.

Category	Odds Ratio
Intercept	0.621630
Total Actions at Arrest	1.327376
Male Gender	1.304855
Black and Indigenous	1.402865

Figure 18: The Odds Ratio for Test Dataset

As evident in Figure 18, total actions at arrest has an odds ratio of 1.33 which indicates that when other predictor variables are held constant, a one-unit increase in the proportion of total actions at arrest will increase the odds of being strip-searched by a factor of 1.33. For gender (male), the odds ratio is 1.30, meaning that when all other predictors are held constant, males are about 1.30 times more likely than females to be strip searched. For the perceived race, when other predictors are held constant, black or indigenous arrestees are 1.40 times more likely to be strip searched.

3.2.5 Model Assessment

An accuracy score was determined for both the test and training models.

Train Accuracy	0.541
Test Accuracy	0.531

Figure 19: Accuracy scores for test and training models

The accuracy scores for both the training and test models were similar. It can be said that the training model can accurately classify whether someone will be strip searched accurately 54.1% of the time on the data it has seen. On examples that the model hasn't seen, an accuracy rate of 53.1% is achieved.

3.2.6 Confusion Matrix

Confusion Matrix for the test and training models were generated and they are summarized below.

		Actual Values	
		Positive	Negative
Predicted values	Positive	2729	3537
	Negative	2191	4021

Figure 20: Confusion Matrix for Train Model

		Actual Values	
		Positive	Negative
Predicted values	Positive	637	896
	Negative	567	1020

Figure 21: Confusion Matrix for Test Model

In both confusion matrices, it can be observed that the models had the greatest number of true negatives, followed by false positives. The accuracy rates of the models align with these confusion matrix values as there are a considerable amount of data points that fit under false positives and false negatives, thus indicating that both models have limited ability to accurately

classify whether or not someone will be strip searched based on gender, perceived race and actions at arrest. However, this will be subject to change every time the model is running.

3.2.7 Confidence Intervals

The tables below states the confidence intervals for the odds ratio of the test and training models. These values indicate the odds ratio range where there's a 95% chance of a population's odd ratio being located.

	Lower CI	Upper CI
Intercept	0.577457	0.694052
Total Actions at Arrest	1.289691	1.445662
Male Gender	1.150817	1.381419
Black and Indigenous	1.144516	1.331247

Figure 22: Train Model Odds Ratio Confidence Interval

	Lower CI	Upper CI
Intercept	0.514812	0.750612
Total Actions at Arrest	1.174921	1.472622
Male Gender	1.103094	1.600085
Black and Indigenous	1.205066	1.633130

Figure 23: Test Model Odds Ratio Confidence Interval

3.2.8 Prediction Interval Plot

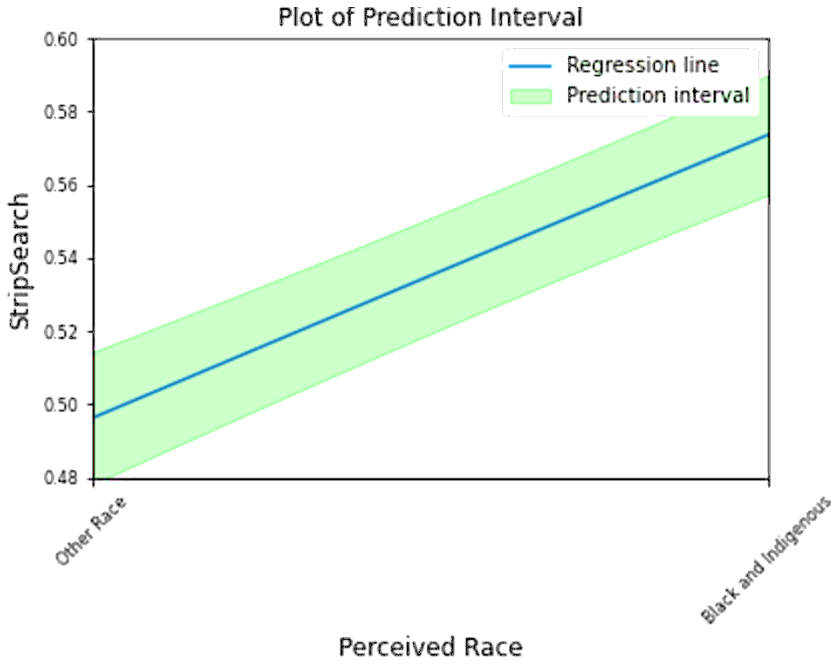

Figure 24: Prediction Interval Plot for Perceived Race

The prediction interval plot above illustrates the range of values (in green) one can say with 95% confidence that a new observation will fall within this range. For this example plot, the categorical perceived race variable acts as the x-axis and the y-axis illustrates the strip search proportion when the other predictor variables (gender and total actions at arrest) are being controlled. The increasing trend observed in this prediction interval plot supports all previous findings regarding higher strip search rates for black and indigenous arrestees.

5. Discussion

This paper was aimed at exploring the various factors that influence an arrestee's likelihood of being strip-searched, including demographic factors such as gender, race and other factors such as an arrestee's behavior during the arrest. There were several key findings that came out of the various statistical tests conducted. Not only are these findings essential in understanding this dataset, but they provide valuable insights into Toronto's policing practises. Preliminary EDA suggested that these factors do indeed play a role in influencing strip search rates conducted by the Toronto Police Department. Based on the EDA, previous findings and

background research, an ANCOVA was conducted to answer the first research question of this paper which is regarding the differences in strip search rates between the different perceived races when the number of actions is being controlled. According to this ANCOVA which is rendered statistically significant, there does appear to be a significant role that Race plays in influencing a police officer's decision to conduct a strip search even after taking into account the role of resistance behavior displayed by an arrestee. This is a crucial finding and can't be looked at in isolation from socio-political context. One of the EDA findings includes Indigenous arrestees having the highest rate of total possible actions displayed at arrest (Figure 12), and often at times, statistics such as these are utilized to justify intrusive and harsh police treatments without being held accountable for racially-divisive practices. The findings of the ANCOVA help confirm that despite controlling for demonstration of resistance from arrestees, there is still evidence of racial biases in the rates of strip searches that negatively affect black and indigenous racial groups.

The second research question was aimed at determining the likelihood of someone being strip searched based on demographic factors such as their perceived race, their gender and other factors such as the demonstration of hostility/resistance (based on total actions at arrest). Two logistic regression models were generated for training and testing purposes. Both models revealed that being a male, being black or indigenous and demonstrating any hostile actions at arrest increases an arrestee's likelihood of being strip searched. These sorts of insights are essential in raising awareness about the problematic trends that the Toronto Police Department is contributing towards so that they can work proactively to address these issues in their police practices. The train and test models have an accuracy rate of 54.1% and 53.1% respectively for predicting whether or not someone will be strip-searched based on the three parameters discussed. These rather low accuracy rates tell us about the dangers of utilizing real-life datasets with limited information on different predictors to make accurate and unbiased predictions for situations like this.

These crucial findings about the Toronto Police Department provide valuable insights into police practises, however, there are several limitations that hinder the validity and accuracy of the statistical findings of this analysis. Several of the assumptions for the ANCOVA and logistic regressions weren't met which impacts not only the accuracy of the findings but also the generalizability of the points made regarding the relationships between race, gender, actions at

arrest and strip search rates. Furthermore, despite considering more predictors than the midterm investigation, the comprehensiveness of the approach is still very limited due to a multitude of reasons. A lack of data for various parameters, and poor data distribution are just some of the factors that pose limitations to this data analysis.

Despite the limitations, the key findings do align with existing literature and notions regarding police practises especially with respect to the treatment of minority groups. There are various societal and political implications for these findings. Studies such as this one indicate the need for acknowledgement of these biases as well as the necessity for proactive policies to combat the racial over-representation and biases that exist in policing practises.

5. Conclusion

With the recent resurgence of movements like Black Lives Matter, there has been an increased light shone on various controversial policing practises, specifically policing practises on minority groups. The aim of this research was to investigate the role of race and other essential factors in strip searches conducted in Toronto between 2020 and 2021. The first research question aimed to investigate whether the perceived race of an arrested individual still influences their likelihood of being strip searched even when the effect of hostile actions at arrest is controlled. It was discovered that race still continued to play a significant role despite taking into account an arrestee's demeanour and behavior towards police officers. The 2nd research question was aimed at building a logistic regression prediction model to determine the likelihood of someone being strip searched based on their race, gender and actions at arrest. It was revealed that while these parameters do influence strip search likelihoods, they are not sufficient to predict the likelihoods with a high degree of accuracy. This highlights the limitations and potential for bias perpetuation when models such as these are implemented in actual law enforcement settings. However, these studies do provide insight into law enforcement practises within society and the findings can be used to guide policy changes in these areas.

6.Citations

Engel, R. S. (2003). Explaining suspects' resistance and disrespect toward police. *Journal of Criminal Justice*, *31*(5), 475–492. https://doi.org/10.1016/s0047-2352(03)00052-7

Lemke, M. (2022, July 18). *Policing toronto: Strip searching in a divided city - the bullet*. Socialist Project. Retrieved February 26, 2023, from https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/

Mazur, D. (2021, December 10). *Policing indigenous communities*. BC Civil Liberties Association. Retrieved February 28, 2023, from https://bccla.org/2017/12/policing-indigenous-communities/

Rankin, J., & Gillis, W. (2018, December 10). *Black people 'grossly overrepresented,' more likely to be hurt or killed by Toronto police, racial profiling report finds*. thestar.com. Retrieved February 28, 2023, from https://www.thestar.com/news/gta/2018/12/10/blacks-grossly-overrepresented-more-likely-to-be-hurt-or-killed-in-interactions-with-toronto-police-racial-profiling-interim-report-finds.html

Robinson, R. (2016, June 22). *2016 Report*. Office of the Independent Police Review Director. Retrieved February 28, 2023, from https://www.oiprd.on.ca/news/annual-repo