INF2178 Midterm Project

Study on Factors Affecting Strip Search Rate

By Group 5:
Meiqi Li
Zhenyu Yuan

Faculty of Information
University of Toronto St.George

Table of Contents

1. Introduction

Stage 1

According to The Economist's Safe Cities Index 2021, Toronto ranks second, behind Copenhagen. However, crime is on the rise in Toronto, as shown by the key crime indicator statistics compiled by the Toronto Police Service, such as theft of motor vehicles, breaking and entering (Vilkhov Law, 2022).

Furthermore, as we know, Canada is a country of immigrants with multi-culture and in 2021, the number of hate crimes reported by police increased by 27% to 3,360. Compared to 2019, hate crimes have increased by 72% in the last two years. More hate crimes against religion (+67%) (including Jewish, Muslim and Catholic) and sexual orientation (+64%) are the main reasons for the national change, along with more incidents against race or ethnicity (+6%) (Moreau, G., 2022). Based on existing research, we are intrigued by examining the correlation between different races and crime rates.

Moreover, race is certainly not the only variable associated with crime rates. intimate partnership violence has also been a topic of concern in Toronto. According to Statistics Canada, Compared to Ontario (81%) and Canada (79%), Toronto had an overwhelming majority of female victims of intimate relationship abuse (83%) (Statistics Canada, 2020). Therefore, we also want to investigate whether male are more likely to commit crimes than females and what age groups are more likely to commit crimes?

Stage 2

There have been numerous reports and studies that suggest that there is a significant correlation between sex and perceived race discrimination and the strip search rate in Canada. In particular, people who identify as female or as belonging to certain perceived racial groups have reported experiencing higher rates of strip searches compared to other groups.

For example, a report by the Canadian Civil Liberties Association (CCLA) found that women, particularly those who identify as Indigenous, Black or racialized, were more likely to be strip-searched than men. The report also found that strip searches of women were often more invasive and degrading than those of men. This highlights how gender intersects with perceived race discrimination to exacerbate the strip search issue. Moreover, a study by the Ontario Human Rights Commission (OHRC) found that

people who identify as Black, Indigenous, or from other racialized groups were overrepresented in incidents where a strip search was conducted, compared to their representation in the population. The OHRC report also identified that strip searches were more likely to be conducted on individuals who had been arrested for less serious offenses, suggesting that strip searches were being used as a form of punishment or intimidation.

So, in this report, we are focusing on whether there is a significant issue with sex and perceived race discrimination in the strip search rate in Canada. Specifically, we aim to address three research questions related to the influence of demographic factors on strip search rates in Toronto by using power analysis, ANCOVA and logistic regression. First of all, we will utilize power analysis to determine the optimal sample size required to detect a statistically significant difference before conducting ANCOVA. Secondly, we will use ANCOVA to investigate how age factors impact strip search rates after controlling for the effect of the sex feature. Last but not least, we will use logistic regression to explore the relationship between perceived race, gender, age, and the likelihood of being strip searched, incorporating whether the case was booked by the police department as a covariate. The report will provide detailed descriptions of the results, including the Confusion Matrix and Prediction interval.

2. Literature Review

The literature "The usual suspects: police stop and search practices in Canada" examines the use of stop and search practices by police in Canada, with a particular focus on the impact of racial profiling. The authors (Wortley, S., & Owusu-Bempah, A. , 2011) use data in 2007 from a national survey of police officers and interviews with community members to explore the extent to which stop and search practices are influenced by race and ethnicity.

The literature highlights several key findings, including that people from racialized communities are disproportionately targeted for stop and search practices. The authors note that this is particularly true for Black and Indigenous people, who are more likely to be stopped and searched than other groups. They also note that there is a lack of transparency and accountability around stop and search practices, which can make it difficult to address concerns about racial profiling.

Furthermore, it’s worth mentioning that compared to white and Chinese counterparts, black Torontonians are often younger and less wealthy. Hence, it may not be racial prejudice but rather youth or poverty that accounts for why black people are more likely to be stopped and searched.

Therefore, we will use this literature to explore if there is any relationship between race and age and strip search by descriptive analysis and quantitative analysis.

In addition to exploring how race and age affect strip search rate, We are also curious if there is any gender bias when the police search people. Do they think women are safer? Or do they think men are more likely to commit crimes?

In the literature “Racial profiling in Canada”, the authors argue that racial profiling is a widespread problem in Canadian policing, with racialized individuals more likely to be stopped, questioned, and searched by police than non-racialized individuals. They also note that men are more likely to be subject to police stops and searches than women, and that racial profiling often leads to the violation of individual rights and freedoms (Wortley, S., & Tanner, J. , 2003).

Consequently, these studies have found that men tend to be more likely than women to be searched by police when committing crimes, and minorities or blacks are also more likely to be searched. These differences are related to a variety of factors including gender stereotypes, ethnic and immigrant background, police practices, and culture. In the following, we will conduct comprehensive research questions to see how these attributes (race, age, gender) affect the dependent variable (strip search rate).

For literature reviews on the previous project, we mainly discussed how race and age affected strip search rate. For this time, we are willing to focus on sex aspects at first.

The article "Strip-Searching of Women in Canada: Wrongs and Rights" provides an analysis of the legal and ethical issues surrounding strip searches of women in Canada. It highlights the harmful effects that strip searches can have on women's dignity, privacy, and mental health. It argues that strip searches are often unnecessary and disproportionate to the reasons for the search, and that they can be traumatizing for women who have experienced sexual violence or have a history of abuse.

The article also discusses the legal framework governing strip searches in Canada and the ways in which the law has failed to protect women from abusive and discriminatory strip searches. It argues that the current legal standards are too lenient and fail to provide meaningful safeguards against abusive strip searches.

Overall, the article sheds light on an important issue and raises important questions about the use of strip searches as a law enforcement tool. It underscores the need for greater attention to the rights of women in the criminal justice system and for reforms to ensure that strip searches are used only when necessary and with appropriate safeguards in place.

In my opinion, I strongly agree with the author. Is there a deficiency in Canadian law with regard to strip search? The Canadian government should adopt a higher legal standard for strip searches. For example, the police should be trained on gender sensitivity and trauma-informed practices, and the use of alternative search methods that respect women's privacy and dignity.

3. Research Objective and Questions

In the first stage, our study aims to investigate whether the decision to conduct strip searches differs based on certain factors and whether there are significant differences between groups. The core question is how demographic factors affect the rate of strip searches. Based on common sense and the content of the literature review, we selected gender, age, and perceived race as independent variables and the probability of strip searches as the dependent variable, dividing the cases into different groups.

In the second stage, we hope to test the validity of the preliminary analysis and research more into the correlations between different variables by identifying potential causality and assessing their predictive power. Our study will seek to examine how perceived race, gender, and age interact with strip search rates in line with existing literature. Our goal is to investigate how changes in one variable affect other variables and study the potential relationships between variables, such as how changes in age groups specifically affect the probability of being strip-searched for those who are arrested. We plan to conduct research to uncover any underlying patterns or trends that were not apparent in the initial analysis, with the ultimate goal of gaining a deeper understanding of the dataset and informing our decision-making process.

The following are our research questions:

Stage 1

- RQ1: Is the probability of an arrestee being subjected to a strip search related to their age, gender, or perceived race?

- RQ2: How do different perceived races along with genders affect the probability of an arrestee being subjected to a strip search? Do specific genders of minority groups of non-white races experience differential strip search possibilities when they are arrested?
- RQ3: How do different perceived races along with age affect the probability of an arrestee being subjected to a strip search? Does any age group of minority groups experience differential strip search possibilities when they are arrested?

Stage 2

- RQ4: How does the power of the study change with variations in sample size, effect size, or level of significance? At a specific level of power and significance, what is the optimal sample size required to detect statistically significant differences between groups? Does the sample we used meet this condition?
- RQ5: After controlling for whether the case was booked by the police department, is there a significant difference in the likelihood of experiencing strip searches based on race?
- RQ6: Based on the perceived race, gender, whether the case was booked by the police department and age of individuals, what factors are most strongly associated with strip searches?

2. Exploratory Data Analysis

In our proposed project, we will use the *Arrests and Strip Searches (RBDC-ARR-TBL-001)* dataset from Toronto Police Service Public Safety Data Portal shared with us by Professor Guha on GiHub for our research. This dataset is available on GitHub and can be found through the following link: https://raw.githubusercontent.com/Shinpai111/inf2178-expdesignfordatascience-w23-Zhenyu_Yuan/main/project%20dataset/Arrests_and_Strip_Searches_(RBDC-ARR-TBL-001).csv

A brief description of this dataset is available on Toronto Police Service Public Safety Data Portal and can be found through the following link: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

This dataset includes information related to all arrests and strip searches from Toronto city. According to the description from the Toronto Police Service Public Safety Data Portal, a strip search refers to a search conducted by a police officer on a person, which includes the removal of some or all clothing and a visual inspection of the body. This dataset contains 65276 entries and 25 attributes, and in this research, the main attributes we are focusing are:

Perceived_Race: The perceived race of an arrestee.

Age_group__at_arrest_: The age group to which the arrestee belongs.

Sex: The gender of the arrestee.

Booked: Whether the case was booked by the police department

StripSearch: Whether the arrestee was subjected to a strip search

Due to some data in the dataset requiring partial modification and the merging of some duplicate categories, we will provide a detailed description of these attributes below, and provide a table of these attributes in Table 1.

Variables

We employed four attributes as variables in this research:

Perceived_Race, *Age_group__at_arrest_*,*Sex*, and *StripSearch.*

The *Perceived_Race* variable contains 8 different categories: *White, Unknown or Legacy, Black, South Asian, Indigenous, Middle-Eastern, Latino, East/Southeast Asian*. The *Age_group__at_arrest_* variable contains 9 different categories: *Aged 17 years and younger, Aged 17 years and under, Aged 18 to 24 years, Aged 25 to 34 years, Aged 35 to 44 years, Aged 45 to 54 years, Aged 55 to 64 years, Aged 65 and older,* and *Aged 65 years and older*. The *Sex* variable contains 3 categories: *F, M, and U*. The *Booked* variable contains 2 different categories: *0 or 1*. The *StripSearch* variable contains 2 different categories: *0 or 1*, which indicates whether or not an arrestee has been subjected to a strip search.

Before conducting experiments with these variables, we dropped rows containing missing data from the selected variables. Also, since the categories 'Aged 17 years and younger' and 'Aged 17 years and under', as well as 'Aged 65 and older' and 'Aged 65 years and older', were duplicated within the *Age_group__at_arrest_* variable, we merged these duplicate categories and renamed the 'Aged 17 years and younger' category to 'Aged 17 years and under', and the 'Aged 65 and older' category to 'Aged 65 years and older'. After that, to facilitate our experiment, since the *Age_group__at_arrest_* variable from the original dataset was not sorted by the order of different age groups of the arrestees, we will factorize the different age stages in the 'Age' variable and in ascending order, starting from 'Aged 17 years and under' as 0, and sorting up to 'Aged 65 years and older' as 6." And finally, we excluded the *U* category from the *Sex* variable in our analysis since the observation size under the *U* category was insufficient compared to M and *F* and excluded the *Unknown or Legacy* category from the *Perceived_Race* variable since this category is not clear enough compared to other categories. After the previous preprocessing, our new data frame contains 60187 entries and 4 attributes. We can define the variables with more details as follows.

Table 1 Variables Table

Variables	**Categories**

Sex	M (Male)
	F (Female)
Age_group__at_arrest_	0 (indicating 'Aged 17 years and under')
	1 (indicating 'Aged 18 to 24 years')
	2 (indicating 'Aged 25 to 34 years')
	3 (indicating 'Aged 35 to 44 years')
	4 (indicating 'Aged 45 to 54 years')
	5 (indicating 'Aged 55 to 64 years')
	6 (indicating 'Aged 65 years and older')
Perceived_Race	White
	Black
	Latino
	East/Southeast Asian
	Indigenous
	Middle-Eastern
	South Asian
Booked	0, (case has not been booked in the police department)
	1, (case has been booked in the police department)
StripSearch	0, (arrestee has not been subjected to a strip search)
	1, (arrestee has been subjected to a strip search)

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Descriptive Statistics

The dataset provided on GitHub was available in one CSV file. To first understand the dataset file, we produced the following count charts based on the age group at arrest, gender, and race of the arrestees, showing whether they were strip-searched (Figure 1 and 2).

Figure 1

Number of cases with strip search has been applied and without strip search has been applied based on different perceived races of arrestees in the Arrests and Strip Searches dataset (broken down by perceived races).

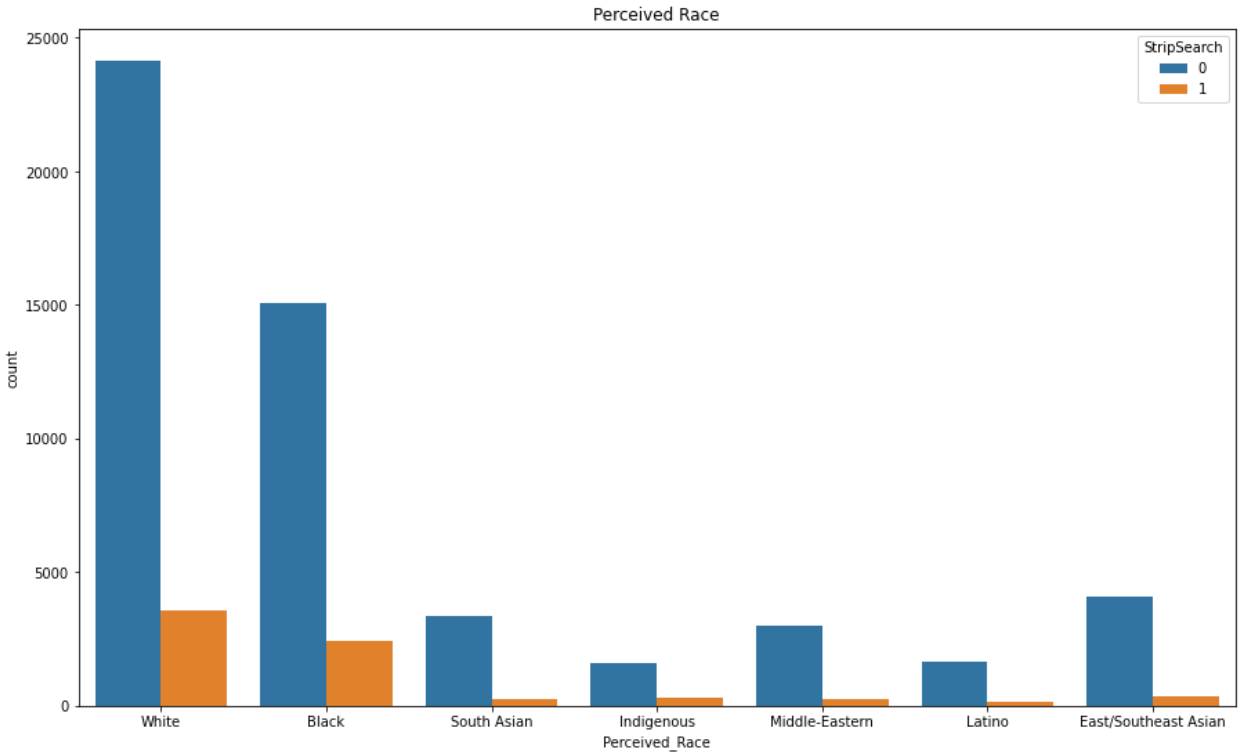
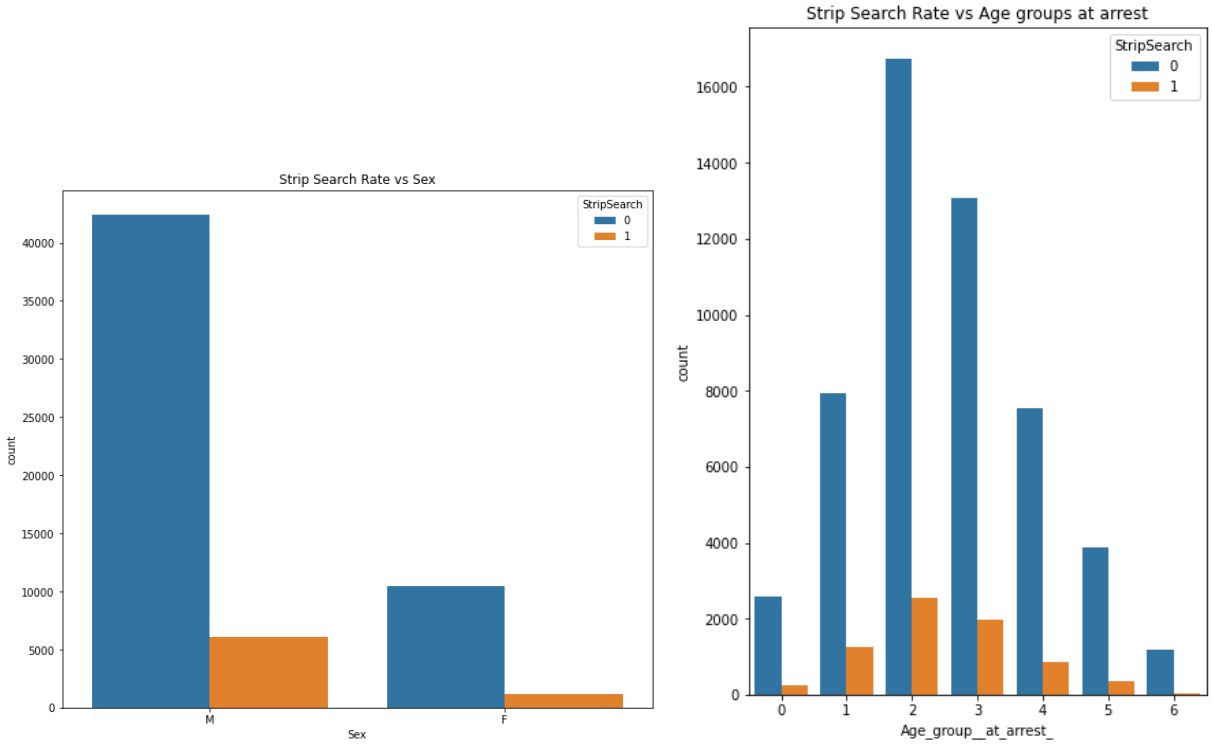


Figure 2

Number of cases with strip search has been applied and without strip search has been applied based on different gender and age groups at arrest in the Arrests and Strip Searches dataset (broken down by gender and different age groups).



We also looked at how arrestees' average strip search possibility differed for each perceived race, age group at arrest, and gender depending on the subject, by looked at how the mean value of *StripSearch* variable differed for each perceived race, age group at arrest, and gender (Table 2 and figure 3).

Table 2 Average Strip Search Possibility for Each Perceived Race, Age Group at Arrest, Gender

Perceived_Race	**Average Strip Search possibility**	**std**
Black	0.139	0.346
East/Southeast Asian	0.077	0.267
Indigenous	0.158	0.365
Latino	0.075	0.263
Middle-Eastern	0.070	0.256
South Asian	0.071	0.257
White	0.129	0.335
Age_group__at_arrest_		
0 (indicating 'Aged 17 years and under')	0.092	0.289
1 (indicating 'Aged 18 to 24 years')	0.136	0.342
2 (indicating 'Aged 25 to 34 years')	0.133	0.339

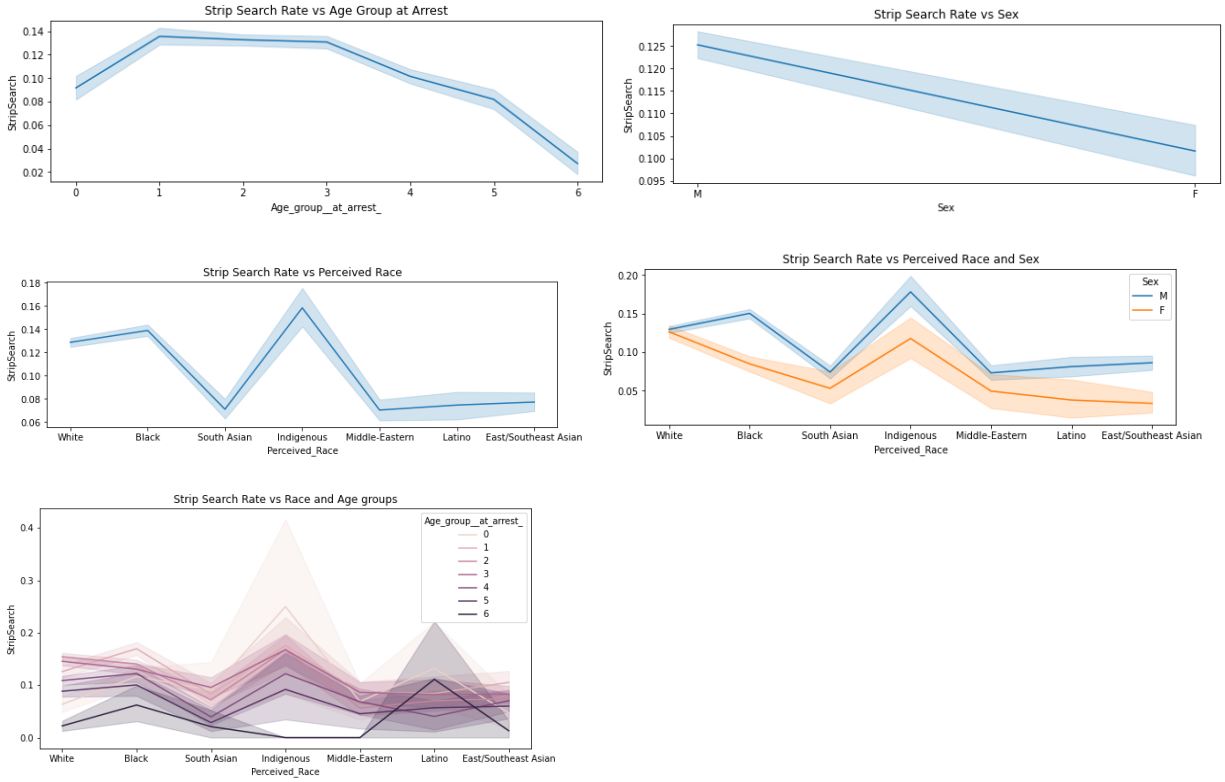
3 (indicating 'Aged 35 to 44 years')	0.131	0.337
4 (indicating 'Aged 45 to 54 years')	0.101	0.302
5 (indicating 'Aged 55 to 64 years')	0.082	0.274
6 (indicating 'Aged 65 years and older')	0.027	0.163
Sex		
F	0.102	0.302
M	0.125	0.331

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

From the line charts in Figure 3, we could see differences in the average possibility of an arrestee having been subjected to a strip search between different gender, age group at arrest and perceived race. We can see those arrestees in age groups 1 to 3, which is between 18 and 44 years old, are the most likely to be strip-searched, with strip-searched rates exceeding 0.12, higher than those in age groups 0, 4, and 5, which is below 17 years old or between 45 and 64 years old, with strip-searched rates around 0.1 ~ 0.08. Arrestees over 65 years old are the least likely to be strip-searched, with a strip-searched rate of less than 0.03. Moreover, we can see male arrestees are relatively more likely to be strip-searched than female arrestees, and arrestees who appear to be of South Asian, Middle Eastern, Latin American, and East/Southeast Asian descent are less likely to be strip-searched compared to those who are White, Black, or Indigenous.

At the same time, it is not surprising to see that the probability of male and female arrestees being strip-searched varies largely among almost all races except between White males and White females. The difference in the probability of being strip-searched between Black males and Black females is the largest among the same races, and the difference between Indigenous men and East/Southeast Asian women is the largest among all racial and gender disparities. Moreover, we also see that the probability of arrestees being strip-searched after arrest varies among almost all races in different age groups. Among them, the difference in strip-searched rates among different age groups of Indigenous individuals is the largest, with youth Indigenous arrestees having the highest risk rate among all racial and age groups, while Indigenous arrestees over 65 years old have the lowest frisk rate among all racial and age groups.

Figure 3



Looking at the line charts in Figure 3 and according to existing literature mentioned in the literature review section, perceived races, age groups, and gender groups among arrestees can affect the possibility of an arrestee being strip-searched, and it seemed possible that the combined effect of the attributes mentioned above could affect the possibility of an arrestee being strip-searched.

Furthermore, we generated a correlation heatmap of the numerical attributes in the dataset to look at the correlation between the attributes of the dataset (Figure4). The heat map showed that there were no noticeable differences in correlation between meaningful attributes.

Figure 4

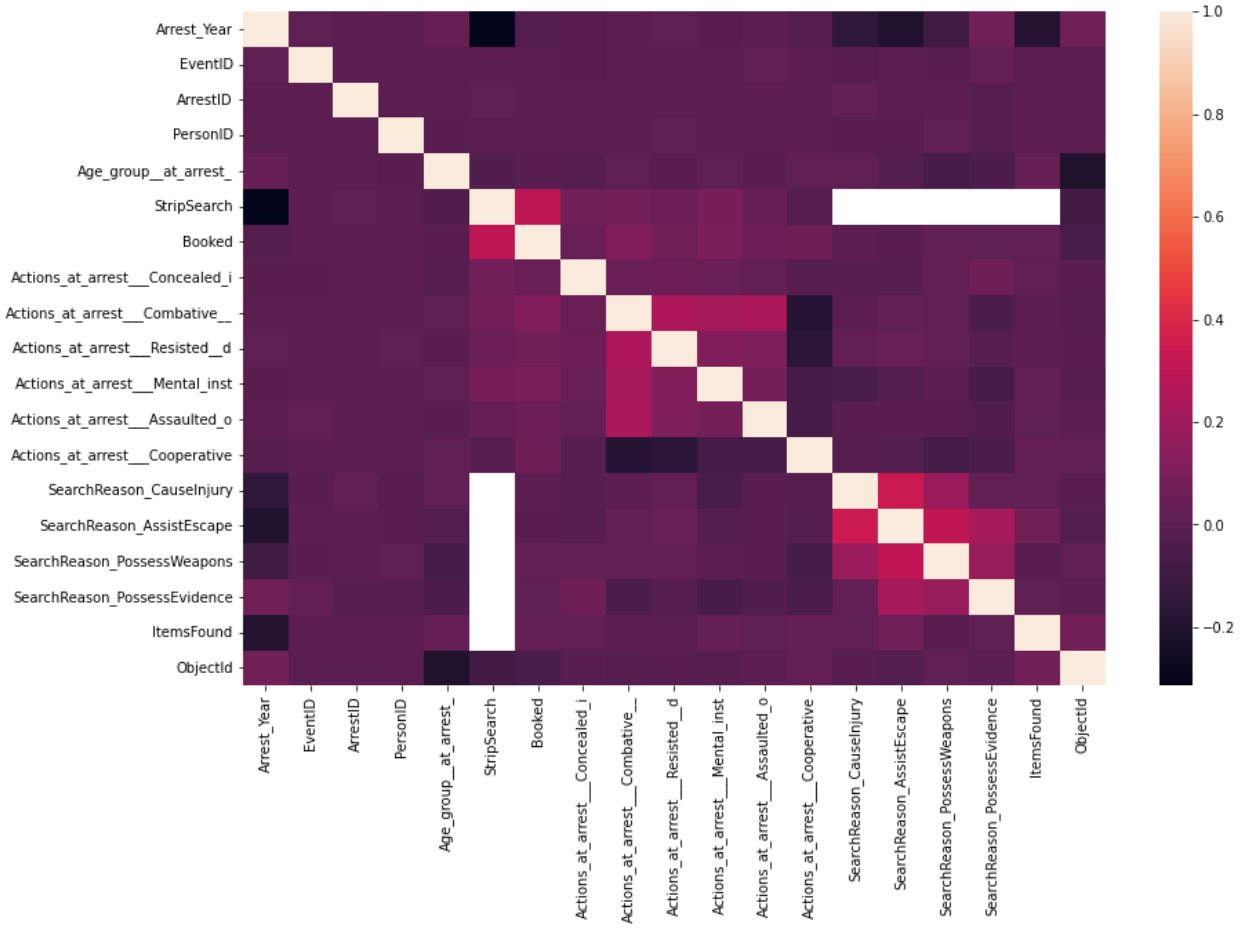
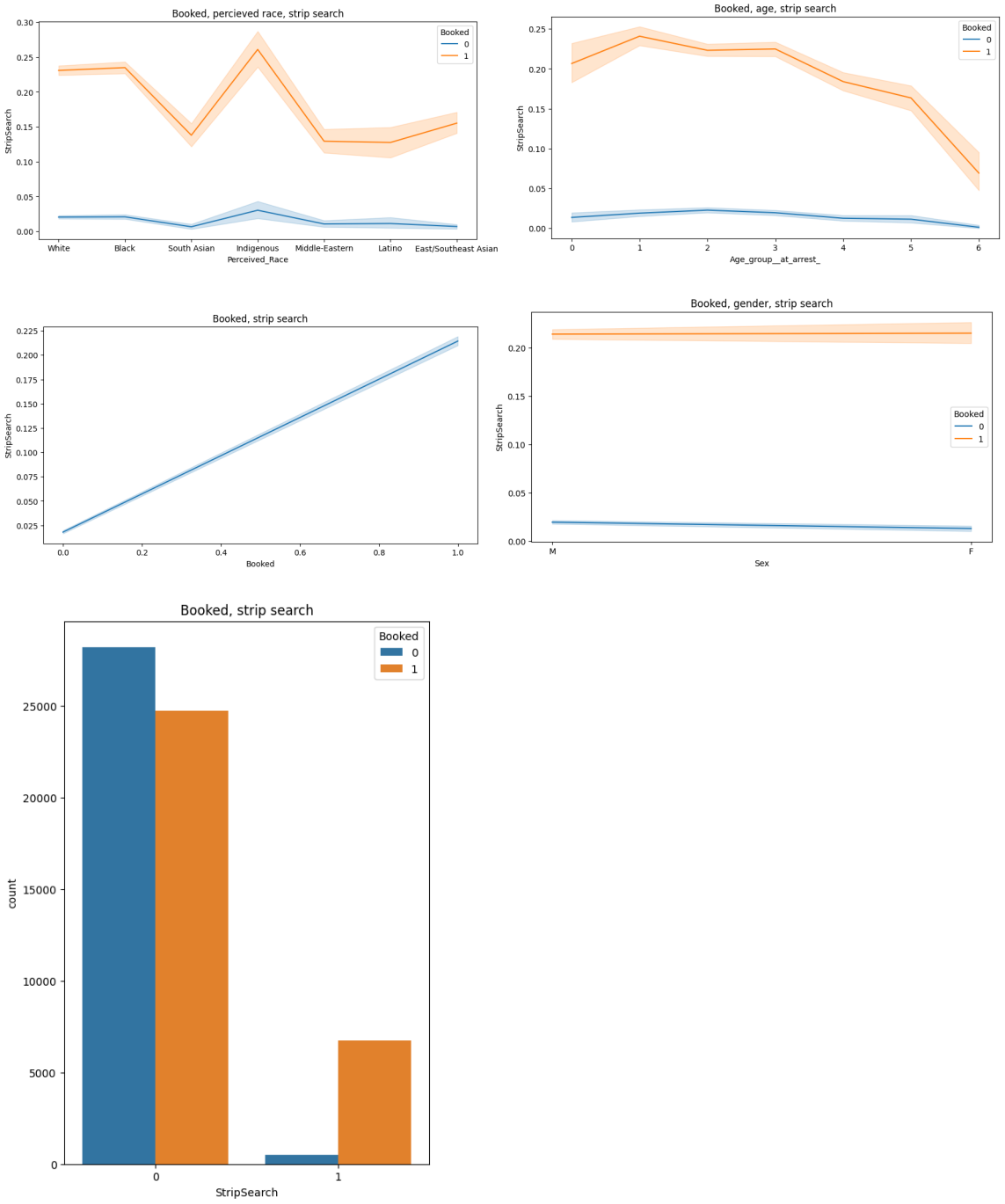


Figure 5 presents a series of effects of whether a case is booked on other independent and dependent variables.

Figure 5



In the above figure, we can observe that, despite no significant difference in the number of cases booked and not booked, the probability of detainees being searched in booked cases has significantly increased.

T-Test

After the descriptive statistics, we ran multiple Welch's t-tests with the categorical attributes in the dataset. Before running the tests, we checked that the following assumptions were fulfilled: (1) a nominal

two-level explanatory variable; (2) a quantitative outcome variable; (3) a normality assumption; and (4) independence of the errors. Seeing as we ran Welch's t-test, equal variance among the residuals was not assumed. The following paragraphs outline the results of noteworthy results from our t-tests.

Perceived_Race and StripSearch

We computed the mean strip search rates for arrestees from different perceived races, and found that compared with White arrestees, the mean strip search rate was higher for Black arrestees and Indigenous arrestees, but was lower for South Asian, Middle-Eastern, Latino, and East/Southeast Asian arrestees. Subsequently, we conducted 6 Welch's T-tests to analyze whether the strip search rate (outcome variable) differs between different perceived race groups compared with White arrestees.

The hypotheses being tested for each t-test are the following:

> For each t-test, for the White arrestee group and each of the minority(Black, South Asian, Indigenous, Middle-Eastern, Latino, East/Southeast Asian) arrestee groups:
>
> H0 (Null Hypothesis): The population means of the two independent groups, White arrestee group and the specific minority arrestee group are equal.
>
> HA (Alternative Hypothesis): The population means of the two independent groups, White arrestee group and the specific minority arrestee group are not equal.

The results indicate that:

For t-test between white and black arrestee groups:

The mean strip search rate for white arrestees (M=0.129, SD=0.335) is lower than for black arrestees (M=0.139, SD=0.346). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0019) is less than 0.05, 95% CI[-0.017, -0.004]. Therefore, we reject the null hypothesis of this t-test.

For t-test between white and South Asian arrestee groups:

The mean strip search rate for white arrestees (M=0.129, SD=0.335) is higher than for South Asian arrestees (M=0.071, SD=0.257). With alpha established at 0.05, this is a statistically significant difference as the p-value (1.108e-33) is less than 0.05, 95% CI[0.048, 0.067]. Therefore, we reject the null hypothesis of this t-test.

For t-test between white and Indigenous arrestee groups:
The mean strip search rate for white arrestees (M=0.129, SD=0.335) is lower than for Indigenous arrestees (M=0.158, SD=0.365). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0005) is less than 0.05, 95% CI[-0.046, -0.013]. Therefore, we reject the null hypothesis of this t-test.

For t-test between white and Middle-Eastern arrestee groups:
The mean strip search rate for white arrestees (M=0.129, SD=0.335) is higher than for Middle-Eastern arrestees (M=0.070, SD=0.256). With alpha established at 0.05, this is a statistically significant difference as the p-value (8.3467e-32) is less than 0.05, 95% CI[0.049, 0.068]. Therefore, we reject the null hypothesis of this t-test.

For t-test between white and Latino arrestee groups:
The mean strip search rate for white arrestees (M=0.129, SD=0.335) is higher than for Latino arrestees (M=0.075, SD=0.263). With alpha established at 0.05, this is a statistically significant difference as the p-value (3.5784e-16) is less than 0.05, 95% CI[0.041, 0.067]. Therefore, we reject the null hypothesis of this t-test.

For t-test between white and East/Southeast Asian arrestee groups:
The mean strip search rate for white arrestees (M=0.129, SD=0.335) is higher than for East/Southeast Asian arrestees (M=0.077, SD=0.267). With alpha established at 0.05, this is a statistically significant difference as the p-value (5.2727e-30) is less than 0.05, 95% CI[0.0426, 0.0602]. Therefore, we reject the null hypothesis of this t-test.

Gender and StripSearch

We computed the mean strip search rates for arrestees from different genders and found that compared with males, the mean strip search rate was lower for female arrestees. To further examine this, we conducted a Welch's T-test to analyze whether the strip search rate (outcome variable) differs based on their gender (two-level explanatory variable). The hypotheses being tested are the following:

> H0 (Null Hypothesis): The population means of the two independent groups, the male arrestee group and the female arrestee group are equal.

HA (Alternative Hypothesis):The population means of the two independent groups, male arrestee group and female arrestee group, are not equal.

Our results indicate that the mean strip search rate for male arrestees (M=0.125, SD=0.331) is higher than for female arrestees (M=0.102, SD=0.302). With alpha established at 0.05, this is a statistically significant difference as the p-value (9.0501e-14) is less than 0.05, 95% CI[0.0175, 0.0299]. Therefore, we can reject the null hypothesis.

Age group at arrest and StripSearch

Before conducting t-tests, there are 7 categories of age classification, and if t-tests are performed one by one, the total number of t-tests will be too large. In order to simplify the number of t-tests, based on the number of arrests for different age groups, we temporarily divide age groups into two groups: those over 34 years old and those under 34 years old, with a relatively similar number of cases on both sides. Then, t-tests will be conducted on these two groups.

We computed the mean strip search rates for arrestees from different age groups and found that compared with arrestees older than 34 years old, the mean strip search rate was higher for arrestees younger than 34 years old. To further examine this, we conducted a Welch's T-test to analyze whether the strip search rate (outcome variable) differs based on arrestees' age groups (two-level explanatory variable). The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, arrestees older than 34 years old and arrestees younger than 34 years old are equal.

H0 (Null Hypothesis): The population means of the two independent groups, arrestees older than 34 years old and arrestees younger than 34 years old are not equal.

Our results indicate that the mean strip search rate for arrestees older than 34 years old (M=0.111, SD=0.314) is lower than for arrestees younger than 34 years old (M=0.130, SD=0.336). With alpha established at 0.05, this is a statistically significant difference as the p-value (6.1221e-13) is less than 0.05, 95% CI[-0.0243, -0.0139]. Therefore, we can reject the null hypothesis.

Booked and StripSearch

We computed the mean strip search rates for arrestees from whether a case has been booked and found that compared with booked, the mean strip search rate was lower for not booked cases. To further examine this, we conducted a Welch's T-test to analyze whether the strip search rate (outcome variable) differs based on whether a case has been booked (two-level explanatory variable). The hypotheses being tested are the following:

> H0 (Null Hypothesis): The population means of the two independent groups, booked case group and not booked case group are equal.
>
> HA (Alternative Hypothesis):The population means of the two independent groups, booked case group and not booked case group, are not equal.

Our results indicate that the mean strip search rate for male arrestees (M=0.214, SD=0.410) is higher than for female arrestees (M=0.018, SD=0.133). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0) is less than 0.05, 95% CI[0.1916, 0.2011]. Therefore, we can reject the null hypothesis.

T-test concluding remarks

The above t-test results presented some important findings from the dataset. In line with existing literature, we found that arrestees' mean strip search possibilities were statistically different depending on arrestees' perceived race groups, gender and age groups. It was found that whether or not a case is booked also has a significant impact on whether a person will be subjected to a strip search or not.

3. Methods

In stage 1, table 1 from the exploratory Data Analysis section outlines the specific details of the variables that will be used. We provide a brief summary of variables here:

The Sex variable retains two categories, F and M, indicating female and male arrestees.

The Age_group__at_arrest_ variable retains seven categories, 0 ~ 6, representing seven different age groups from low to high.

The Perceived_Race variable still contains 7 different categories, which are *White, Black, South Asian, Indigenous, Middle-Eastern, Latino, East/Southeast Asian*, indicating the perceived race of an arrestee.

The StripSearch variable contains 2 different categories, 0 or 1, which indicate whether or not an arrestee has been subjected to a strip search.

Due to space limitations and the nature of the t-test itself, we only conducted one t-test in the age group at arrest attribute and only compared white arrestees with other minority arrestee groups in the perceived race attribute during the t-test stage. In the following research, based on our research questions, we want to perform two 2-way ANOVAs to further analyze the effect of age group at arrest, gender and perceived race on mean strip search, and provide Tukey's HSD as post-hoc tests for each ANOVA tests. After that, we will provide 2 interaction plots in the Result section to further explain our findings.
In the following two 2-way ANOVA tests, we will respectively analyze the effects of perceived race and gender on the mean strip search, and the effects of perceived race and age group at arrest on the mean strip search.

The alpha value is set prior to perform the hypothesis test and determines the acceptable level of error. In this report, a common alpha value of 0.05 will be used. We process the p-value to measure the significance of evidence against the null hypothesis. The smaller the p-value, the stronger the evidence in favor of rejecting the null hypothesis. The p-value is then compared to the alpha value set for the statistical test. If the p-value is less than the alpha value we set, then the null hypothesis is rejected, whereas if the p-value is greater than it, the null hypothesis cannot be rejected.

In stage 2, we tried to use power analysis, ANCOVA, and Logistic regression with confusion matrix and prediction interval to conduct more comprehensive studies on our research questions based on the literature review.

The power of a study is the probability that it will correctly reject the null hypothesis when the alternative hypothesis is true. In other words, it is the likelihood that a study will detect a statistically significant difference or effect if it really exists in the population being studied. The power of a study depends on several factors, including the sample size, the significance level, and the effect size. A study with high power is more likely to detect a true effect, whereas a study with low power may fail to detect a true effect, leading to a false negative result. Power analysis is a useful tool for determining the optimal sample size needed to detect a statistically significant difference in a study, and for ensuring that a study has sufficient power to detect meaningful effects. By using power analysis, researchers can improve the reliability and validity of their research results.

ANCOVA is a statistical technique used to determine if there is a significant difference in means between two or more groups after controlling for the effects of one or more continuous variables (covariates). When analyzing the relationship between age groups, gender, and strip search ratios, age groups can be considered a potential covariate. By including age groups as a covariate in ANCOVA, we can control this potentially confounding variable and obtain a clearer picture of the relationship between gender and strip search. Essentially, ANCOVA with *Booked* as a covariate allows us to determine if there is a significant difference in strip search rates between different case groups whether a case has been booked or not. By considering the impact of potentially confounding variables, it can provide more reliable and accurate results, thus improving the effectiveness of the analysis.

Logistic regression is a statistical technique used to model the relationship between a binary outcome variable (i.e., either 0 or 1) and one or more predictor variables. When analyzing strip searches and factors such as age groups, gender, whether the case was booked by the police department, and perceived race, logistic regression can be used to predict which factor is most likely to lead to a person being subjected to a strip search.

In stage 1:

ANOVA tests and Tukey's HSD tests design

Perceived race and Age group at arrest on strip search

The main effect of perceived race (set 1):

H0 (Null Hypothesis): The population means of different perceived races are equal.

H0 (Null Hypothesis): The population means of different perceived races are not equal.

The main effect of age group at arrest (set 2):

H0 (Null Hypothesis): The population means of different age groups at arrest are equal.

H0 (Null Hypothesis): The population means of different age groups at arrest are not equal.

The main effect of perceived race and age group at arrest (set 3):

H0 (Null Hypothesis): The effect of perceived race on strip search between different age groups at arrest are equal.

H0 (Null Hypothesis): The effect of perceived race on strip search between different age

groups at arrest are not equal.

A two-way ANOVA will be performed to analyze the effect of perceived races and different age groups at arrest on strip search.
After we have conducted the above ANOVA test, if the interaction effect is statistically significant, we will use the Tukey HSD test as a post-hoc test to further analyze our research question.

Perceived race and Gender at arrest on strip search

The main effect of perceived race:

H0 (Null Hypothesis): The population means of different perceived races are equal.
H0 (Null Hypothesis): The population means of different perceived races are not equal.

The main effect of age group at arrest:

H0 (Null Hypothesis): The population means of different gender groups at arrest are equal.
H0 (Null Hypothesis): The population means of different gender groups at arrest are not equal.

The main effect of perceived race and age group at arrest:

H0 (Null Hypothesis): The effect of perceived race on strip search between different gender groups at arrest are equal.
H0 (Null Hypothesis): The effect of perceived race on strip search between different gender groups at arrest are not equal.

A two-way ANOVA will be performed to analyze the effect of perceived races and different gender groups at arrest on strip search.
After we have conducted the above ANOVA test, if the interaction effect is statistically significant, we will use the Tukey HSD test as a post-hoc test to further analyze our research question.

In stage 2:

RQ4: Before conducting ANCOVA, we will use power analysis to determine the optimal sample size required to detect a statistically significant difference. To perform power analysis, the following steps will

be taken in this report. Effect size is typically expressed as Cohen's d, which is the difference between the means of the groups divided by the pooled standard deviation. Then, the level of significance (alpha) is the probability of rejecting the null hypothesis when it is true. It is typically set at 0.05, which corresponds to a 5% chance of making a Type I error. The power (1 - beta) is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It is typically set at 0.80 or 0.90, which corresponds to an 80% or 90% chance of detecting a true effect. Sample size can be calculated using power analysis formulas or online calculators. For example, the G*Power software can be used to calculate the sample size based on the effect size, level of significance, power, and other parameters. The power curve is shown in Figure 6 in the result session.

RQ5: Through the previous use of ANOVA, we found that demographic factors such as perceived race, age group, and sex do indeed influence the rate of strip searches. However, the meaning of the results may be affected by the activity of being booked. To further investigate how these factors impact the strip search rate, we will use the proportion of being booked as a covariate and perform ANCOVA to test whether the independent variables still have an effect on the dependent variable after controlling for the covariate.

RQ6: To investigate the relationship between perceived race, gender, whether the case was booked by the police department, and age of individuals with the likelihood of being strip searched, we will use logistic regression with the probability of being strip searched as the dependent variable and age, gender, and race as independent variables, and incorporate whether the case was booked by the police department as a covariate in the analysis. We will describe the results using the Confusion Matrix and Prediction interval.

4.Results

In stage 1:

Perceived race and Age group at arrest on strip search

ANOVA test result is shown in Table 3, alpha established at 0.05.
Simple main effects analysis showed that perceived races did have a statistically significant effect on strip search. ($p = 1.233e-76$). Null hypothesis (set1) is rejected.
Simple main effects analysis showed that age groups at arrest did have a statistically significant effect on strip search. ($p = 6.529e-55$). Null hypothesis (set2) is rejected.

A two-way ANOVA revealed that there was a statistically significant interaction between the effects of perceived races and different age groups at arrest (p = 1.072e-08). Null hypothesis (set 3) is rejected.

Table 3 2-Way ANOVA Test Result Perceived Race and Age group at arrest on StripSearch

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	38.835453	6.0	61.700552	1.233072e-76
C(Age_group__at_arrest_)	28.150854	6.0	44.725196	6.529425e-55
C(Perceived_Race):C(Age_group__at_arrest_)	11.039273	36.0	2.923142	1.072262e-08
Residual	6308.658931	60138.0	NaN	NaN

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

While all null hypotheses are rejected, to further investigate the effects, a Tukey HSD test was conducted to compare all possible pairs of means.

Table 4 *Post HOC Comparisons of Two-Way ANOVA*

Group 1	Group 2	Meandiff	P-adj
FBlack	FEast/Southeast Asian	-0.0512	0.0087
FBlack	FWhite	0.0415	0.001
FBlack	FWhite	0.0415	0.001
FBlack	MBlack	0.0654	0.001
FBlack	MIndigenous	0.0933	0.001
FBlack	MWhite	0.0447	0.001
FEast/Southeast Asian	FIndigenous	0.0841	0.001
FEast/Southeast Asian	FWhite	0.0927	0.001
FEast/Southeast Asian	MBlack	0.1166	0.001
FEast/Southeast Asian	MEast/Southeast Asian	0.0526	0.0043
FEast/Southeast Asian	MIndigenous	0.1445	0.001

FEast/Southeast Asian	MWhite	0.0958	0.001
FIndigenous	MIndigenous	0.0604	0.0094
FLatino	FWhite	0.0884	0.0012
FLatino	MBlack	0.1123	0.001
FLatino	MIndigenous	0.1402	0.001
FLatino	MWhite	0.0915	0.001
FMiddle-Eastern	FWhite	0.0768	0.001
FMiddle-Eastern	MBlack	0.1007	0.001
FMiddle-Eastern	MIndigenous	0.1286	0.001
FMiddle-Eastern	MWhite	0.0799	0.001
FSouth Asian	FWhite	0.0732	0.001
FSouth Asian	MBlack	0.0971	0.001
FSouth Asian	MIndigenous	0.125	0.001
FSouth Asian	MWhite	0.0764	0.001
FWhite	MBlack	0.0239	0.001
FWhite	MEast/Southeast Asian	-0.0401	0.001
FWhite	MIndigenous	0.0518	0.001
FWhite	MLatino	-0.0451	0.001
FWhite	MMiddle-Eastern	-0.0532	0.001
FWhite	MSouth Asian	-0.0522	0.001
MBlack	MEast/Southeast Asian	-0.064	0.001
MBlack	MLatino	-0.069	0.001
MBlack	MMiddle-Eastern	-0.0771	0.001
MBlack	MSouth Asian	-0.0761	0.001
MBlack	MWhite	-0.0208	0.001
MEast/Southeast Asian	MIndigenous	0.0919	0.001
MEast/Southeast Asian	MWhite	0.0432	0.001
MIndigenous	MLatino	-0.0969	0.001
MIndigenous	MMiddle-Eastern	-0.105	0.001
MIndigenous	MSouth Asian	-0.104	0.001

MIndigenous	MWhite	-0.0487	0.001
MLatino	MWhite	0.0482	0.001
MMiddle-Eastern	MWhite	0.0563	0.001
MSouth Asian	MWhite	0.0553	0.001

The result suggests that there were significant differences in means of strip search rates for the relatively low number of all pairs of races and age groups.

Perceived race and Gender at arrest on strip search

ANOVA test result is shown in Table 5, alpha established at 0.05.
Simple main effects analysis showed that perceived races did have a statistically significant effect on strip search. (p = 5.526e-82). Null hypothesis (set1) is rejected.
Simple main effects analysis showed that different gender groups at arrest did have a statistically significant effect on strip search. (p = 4.658e-17). Null hypothesis (set2) is rejected.
A two-way ANOVA revealed that there was a statistically significant interaction between the effects of perceived races and different age groups at arrest (p = 4.419e-13). Null hypothesis (set 3) is rejected.

Table 5 2-Way ANOVA Test Result Perceived Race and Gender group at arrest on StripSearch

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	41.599256	6.0	65.875129	5.526450e-82
C(Sex)	7.421914	1.0	70.518499	4.657528e-17
C(Perceived_Race):C(Sex)	7.354067	6.0	11.645644	4.418843e-13
Residual	6333.073078	60173.0	NaN	NaN

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

While all null hypotheses are rejected, to further investigate the effects, a Tukey HSD test was conducted to compare all possible pairs of means. Since the output table of the Tukey HSD test is too large, due to page limitation, we will put the test results into an external document and provide a link to the document.
Link of the Tukey HSD test results: https://docs.google.com/document/d/1cCHhcaYUOtfXJiDQgC9tRo4Ds_p7KIp4sb_3TsNkR6s/edit

Table 6 Sample Post HOC Comparisons of Two-Way ANOVA

Group 1	Group 2	Mean Diff.	p-adj
0Black	1Black	0.053	0.001
0Black	2Middle-Eastern	-0.0601	0.0096
0Black	4South Asian	-0.0771	0.0051
0Black	6White	-0.0943	0.001
0East/Southeast Asian	1Black	0.1226	0.0014
0East/Southeast Asian	2Indigenous	0.1205	0.0114
0East/Southeast Asian	2White	0.1074	0.0163
0East/Southeast Asian	3Indigenous	0.1208	0.0166
0Middle-Eastern	1Black	0.101	0.0127
0White	1Black	0.1061	0.001
0White	1Indigenous	0.1121	0.0034
0White	1White	0.0623	0.001
0White	2Black	0.0768	0.001
0White	2Indigenous	0.104	0.001
0White	2White	0.0909	0.001
0White	3Black	0.0669	0.001

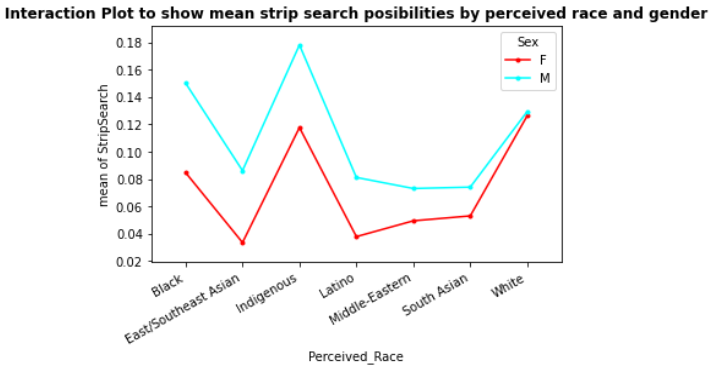
0White	3Indigenous	0.1043	0.001
0White	3White	0.082	0.01
0 White	4 Black	0.0586	0.0087
1 Black	1East/Southeast Asian	-0.0643	0.001
1 Black	1 Latino	-0.0855	0.0076
1 Black	1Middle-Eastern	-0.0833	0.001
1 Black	1 South Asian	-0.087	0.001
1 Black	1 White	-0.0438	0.001
1 Black	2 Black	-0.0292	0.0101
1 Black	2 East/Southeast Asian	-0.0953	0.001
1 Black	2 Latino	-0.0994	0.001
1 Black	2 Middle-Eastern	-0.1132	0.001
1 Black	2 South Asian	-0.0977	0.001
1 Black	3 Black	-0.0392	0.001
1 Black	3 East/Southeast Asian	-0.0867	0.001
1 Black	3 Latino	-0.0874	0.001
1 Black	3 Middle-Eastern	-0.0834	0.001
1 Black	3 South Asian	-0.074	0.001
1 Black	4 Black	-0.0474	0.001

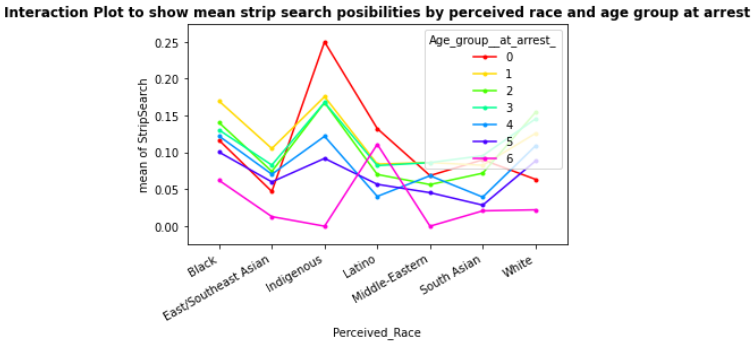
1 Black	4 East/Southeast Asian	-0.0992	0.001
1Black	4Latino	-0.1292	0.001
1Black	4Middle-Eastern	-0.1012	0.001
1Black	4South Asian	-0.1302	0.001
1Black	4White	-0.0607	0.001
1Black	5Black	-0.0692	0.001
1Black	5East/Southeast Asian	-0.1096	0.001
1Black	5Middle-Eastern	-0.1242	0.001
1Black	5South Asian	-0.141	0.001
1Black	5White	-0.0812	0.001
1Black	6Black	-0.1075	0.0297
1Black	6East/Southeast Asian	-0.1566	0.001
1East/Southeast Asian	2White	0.0491	0.0265
1East/Southeast Asian	6White	-0.0831	0.001
1Indigenous	2East/Southeast Asian	-0.1014	0.0158
1Indigenous	2Latino	-0.1055	0.0249
1Indigenous	2Middle-Eastern	-0.1192	0.001

1Indigenous	2South Asian	-0.1037	0.0115
1Indigenous	4East/Southeast Asian	-0.1052	0.0304
1Indigenous	4Latino	-0.1353	0.0167
1Indigenous	4South Asian	-0.1362	0.001
1Indigenous	5East/Southeast Asian	-0.1156	0.0196
1Indigenous	5Middle-Eastern	-0.1302	0.0499
1Indigenous	5South Asian	-0.1471	0.001
1Indigenous	6East/Southeast Asian	-0.1627	0.0017
1Indigenous	6White	-0.1535	0.001
1Middle-Eastern	2 Black	0.0541	0.0242
1Middle-Eastern	2 Indigenous	0.0813	0.0023
1Middle-Eastern	2 White	0.0681	0.001
1Middle-Eastern	3 Indigenous	0.0815	0.0073
1Middle-Eastern	3 White	0.0592	0.0038
1 South Asian	2 Black	0.0578	0.0069
1 South Asian	2 Indigenous	0.085	0.001

The result suggests that there were significant differences in means of strip search rates for around half of all pairs of races and gender groups.

Figure5





Although above interaction plots in Figure 5 do not provide any information on statistically significant differences, the plots showed that: (a) the average number of strip searches conducted on male arrestees is greater than that of female arrestees in all populations; (b) there is a noticeable big difference in the likelihood of strip searches conducted on arrestees of all races and genders; (c) the difference in the likelihood of strip searches conducted on white male and female arrestees is the least noticeable; (d) male arrestees of East/Southeast Asian, Latino, Middle-Eastern, and South Asian descent even have lower strip search rates than female arrestees of other races. (e) there are noticeable differences in the likelihood of strip searches conducted on arrestees of all races and age groups; (f) Indigenous juvenile arrestees have the highest strip search rate; (g) overall, arrestees aged 65 and above are the least likely to be subjected to strip searches.

Back to our RQs:

RQ1: Is the probability of an arrestee being subjected to a strip search related to their age, gender, or perceived race?

According to our research results, the probability of an arrestee being subjected to a strip search is related to their age, gender and perceived race. Both the ANOVA tests we performed have indicated such results,

and such finding is consistent with our literature review, which has shown the strip search likelihood is different from different age, gender and perceived races.

RQ2: How do different perceived races along with genders affect the probability of an arrestee being subjected to a strip search? Do specific genders of minority groups of non-white races experience differential strip search possibilities when they are arrested?

According to our research results, different perceived races along with genders do affect the probability of an arrestee being subjected to a strip search. Our two-way ANOVA test on perceived race and gender groups on strip search indicated such results, and the Tukey HSD test after the two-way ANOVA test further helped us to summarize in-depth which races and genders are more likely to be searched relative to white people, and which races and genders are less likely to be searched relative to white people. Overall, Black and Indigenous males are more likely to be subjected to body searches compared to White males, while minority females are not more likely to be searched compared to White females. Such findings are consistent with our literature review.

RQ3: How do different perceived races along with age affect the probability of an arrestee being subjected to a strip search? Does any age group of minority groups experience differential strip search possibilities when they are arrested?

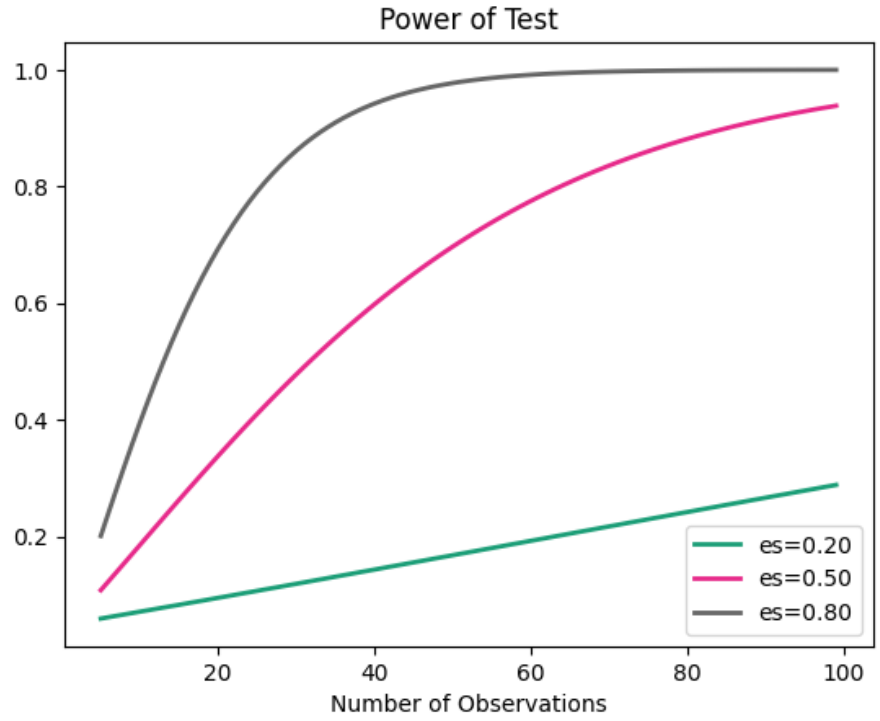
According to our research results, different perceived races along with age groups do affect the probability of an arrestee being subjected to a strip search. Our two-way ANOVA test on perceived race and age groups at arrest on strip search indicated such results, and the Tukey HSD test after the two-way ANOVA test further helped us to summarize in-depth which races and age groups are more likely to be searched relative to white people, and which races and genders are less likely to be searched relative to white people. Overall, compared with perceived races along with gender, perceived races along with age groups are less likely to affect the strip search likelihood of an arrestee, despite the ANOVA test result being still significant. In this case, indigenous minors who are arrested have the highest rate of being subjected to strip searches.

In stage 2:

RQ4: How does the power of the study change as sample size, effect size, or significance level changes? What is the optimal sample size needed to detect statistically significant differences between groups at a given power and significance level? Do the samples we use to satisfy that condition?

Relying solely on P-value in statistical analysis is not comprehensive, as researchers may make errors such as Type I and Type II errors. Power analysis can help determine the sample size needed for the experiment at the expected level of significance, effect size, and statistical power. In order to test the effectiveness of our study based on the existing sample size, we conducted a power analysis to confirm the required sample size and to calculate whether the probability of detecting the effect correctly met the standard.

Figure 6 *Power Curve*



The power curve can help us visualize how changes in sample size affect their ability to detect a significant effect. By examining the power curve, we can identify the minimum sample size needed to achieve the desired level of statistical power. Additionally, the power curve can help us understand the trade-off between sample size and effect size.

From this figure, we can observe that with the same sample size, as the effect size increases, the power of the test also increases. Meanwhile, with the same power, a larger effect size requires less sample size. The effect of a larger effect size on power is particularly sensitive when the sample size is very small. To test a hypothesis more accurately, it is ideal to have a higher sample size and an effect size.

Table 7. Results of Power Analysis for ANCOVA

Setup: sig.level = 0.05; effect size = 0.632; power = 0.8

	Sample Size	Actual Size
Not Booked	38.520	28720
Booked	42.204	31467
Power	0.800	1.0

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

According to the result of power analysis, to reach a power of 0.8, the minimum sample size needed for Not Booked cases is 38.520, and the required number of samples in Booked cases is 42.204 to have a significant p-value in the t-test. Obviously, the actual size (28720 and 31467 respectively) far exceeds the required sample size.

RQ5: After controlling for whether the case was booked by the police department, is there a significant difference in the likelihood of experiencing strip searches based on race?

ANCOVA

Table 8. Result of ANCOVA

Source	SS	DF	F	p-unc	np2
Perceived_Race	30.693849	6	53.282864	7.41e-66	0.005284
Booked	570.115557	1	5938.138906	0.00E+00	0.089812
Residual	5777.733502	60179			

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Null hypothesis: There is no significant difference in the mean scores of strip search rate among arrestees of different race groups after controlling whether the case is booked or not in the police station as a covariate.

Alternative hypothesis: There is a significant difference in the mean scores of strip search rate among arrestees of different race groups after controlling whether the case is booked or not in the police station as a covariate.

Statistical interpretation: Interpretation p-unc = “uncorrected p-value” for Perceived_Race is less than 0.05. We can reject the null hypothesis that there is no significant difference in the mean scores of strip search rate among arrestees of different race groups after controlling whether the case is booked or not in the police station as a covariate.

Practical interpretation: We hypothesized that the ethnic background would be able to predict the strip search rate of an arrestee. From our results, we see that there is a statistically significant relationship between race and strip search rate when controlling for whether the case is been booked or not. The results show that Perceived_Race has a significant effect on the outcome variable (as indicated by the large F-value and very low p-value), but the effect size is small (np2 = 0.005284). Booked also has a significant effect on the outcome variable, with a much larger effect size (np2 = 0.089812). The residual SS and DF represent the unexplained variance in the outcome variable after accounting for the effects of the two predictor variables.

RQ6: Based on the perceived race, gender, whether the case was booked by the police department and age of individuals, what factors are most strongly associated with strip searches?

Logistic Regression

Since the dependent variable in this study is the probability of being strip searched, we need to use logistic regression to deeply analyze RQ6.

Model results

Table 9. Results of Logit Regression

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-3.7235	0.067	-55.811	0.000	-3.854	-3.593
Sex[T.M]	0.1406	0.040	3.495	0.000	0.062	0.219
Perceived_Race[T.East/Southeast Asian]	-0.5340	0.071	-7.555	0.000	-0.673	-0.395
Perceived_Race[T.Indigenous]	0.2375	0.079	3.015	0.003	0.083	0.392
Perceived_Race[T.Latino]	-0.6927	0.106	-6.535	0.000	-0.901	-0.485
Perceived_Race[T.Middle-Eastern]	-0.7623	0.085	-8.926	0.000	-0.930	-0.595
Perceived_Race[T.South Asian]	-0.6528	0.079	-8.252	0.000	-0.808	-0.498

Perceived_Race[T.White]	0.0549	0.034	1.601	0.109	-0.012	0.122
Age_group__at_arrest_	-0.1129	0.012	-9.284	0.000	-0.137	-0.089
Booked	2.6570	0.052	51.478	0.000	2.556	2.758

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Table 10. Odd ratio

	Lower CI	Upper CI	OR
Intercept	0.021189	0.027523	0.024150
Sex[T.M]	1.063712	1.245406	1.150980
Perceived_Race[T.East/Southeast Asian]	0.510421	0.673374	0.586263
Perceived_Race[T.Indigenous]	1.086646	1.479760	1.268059
Perceived_Race[T.Latino]	0.406366	0.615727	0.500210
Perceived_Race[T.Middle-Eastern]	0.394674	0.551613	0.466591
Perceived_Race[T.South Asian]	0.445805	0.607882	0.520574
Perceived_Race[T.White]	0.987753	1.129936	1.056455
Age_group__at_arrest_	0.872217	0.914796	0.893253
Booked	12.882381	15.771248	14.253815

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

The logistic regression results presented in Tables 8 and 9 are based on an analysis of data collected from the Arrests and Strip Searches dataset (RBDC-ARR-TBL-001). The dependent variable in the regression analysis is the probability of being strip-searched, and the independent variables are sex, perceived race, booked status, and search reason.

Table 9 displays the results of the logistic regression analysis in terms of the estimated coefficients, standard errors, z-scores, and p-values for each independent variable. The results indicate that sex, perceived race, booked status, and search reason are significant predictors of the probability of being strip

searched. Specifically, the coefficient for sex is 0.1406 with a p-value of 0.040, indicating that males are more likely to be strip-searched compared to females. The coefficients for the perceived race categories indicate that individuals who are East/Southeast Asian, Latino, Middle-Eastern, or South Asian are less likely to be strip-searched compared to individuals who are perceived as Black, and Indigenous are more likely to be strip-searched compared to individuals who are perceived as Black. The coefficient for the booked status variable is positive (2.6570), indicating that individuals who were booked by the police are more likely to be strip searched. The coefficient forAge_group__at_arrest_ variable is negative (-0.1129), indicating that individuals are less likely to be strip searched when they are elder,

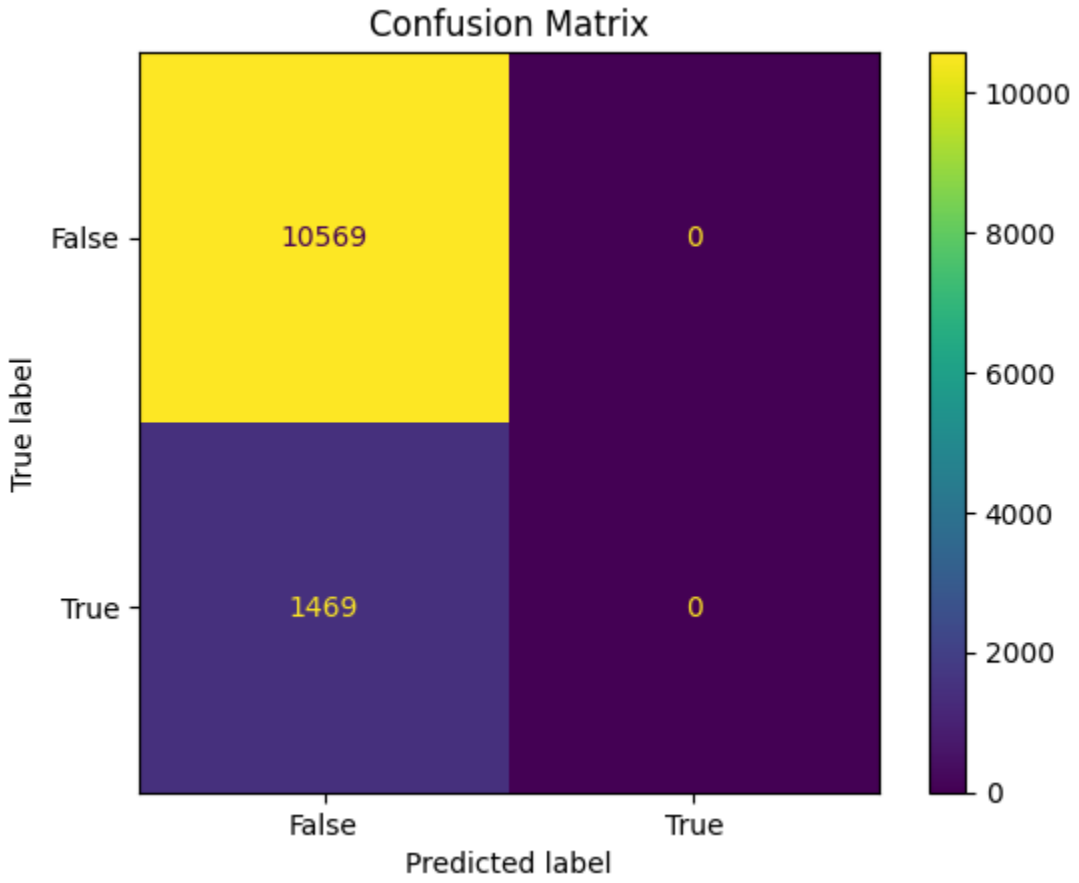
Table 10 presents the results of the logistic regression analysis in terms of the odds ratios and their 95% confidence intervals (CIs). The odds ratios provide a measure of the strength and direction of the relationship between each independent variable and the probability of being strip-searched. The odds ratio for sex is 1.151 with a 95% CI ranging from 1.064 to 1.245, indicating that males are 1.151 times more likely to be strip searched compared to females. The odds ratios for the perceived race categories indicate that individuals who are East/Southeast Asian, Latino, Middle-Eastern, or South Asian are less likely to be strip searched compared to individuals who are perceived as Black. The odds ratio for booked status is 14.253815 with a 95% CI ranging from 12.88 to 15.77, indicating that individuals who were booked by the police are 14.25 times more likely to be strip searched. Finally, the odds ratio for different age groups is 0.89 with a 95% CI ranging from 0.87 to 0.91, indicating that individuals who were from elder age groups are 0.89 times less likely to be strip searched whenever they belong to an older age group.

In conclusion, the logistic regression results suggest that sex, perceived race, booked status, and age are significant predictors of the probability of being strip searched. Specifically, males are more likely to be strip searched compared to females, while individuals who are East/Southeast Asian, Latino, Middle-Eastern, or South Asian are less likely to be strip searched compared to individuals who are perceived as Black. Additionally, individuals who were booked by the police are more likely to be strip-searched, while individuals who were from elder age groups are less likely to be strip searched. These findings highlight the need for further research and policy interventions to address the racial and gender disparities in strip searches. At the same time, we can see that whether the case was booked by the police station was the largest factor affecting the chances of a strip search, while whether the arrested person was Middle Eastern was the second largest factor.

Assess the model

According to our test, the model has a test accuracy of approximately 0.878, which means that the model correctly classified 87.8% of the test data. Figure 7 shows the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) predictions made by the model.

Figure 7

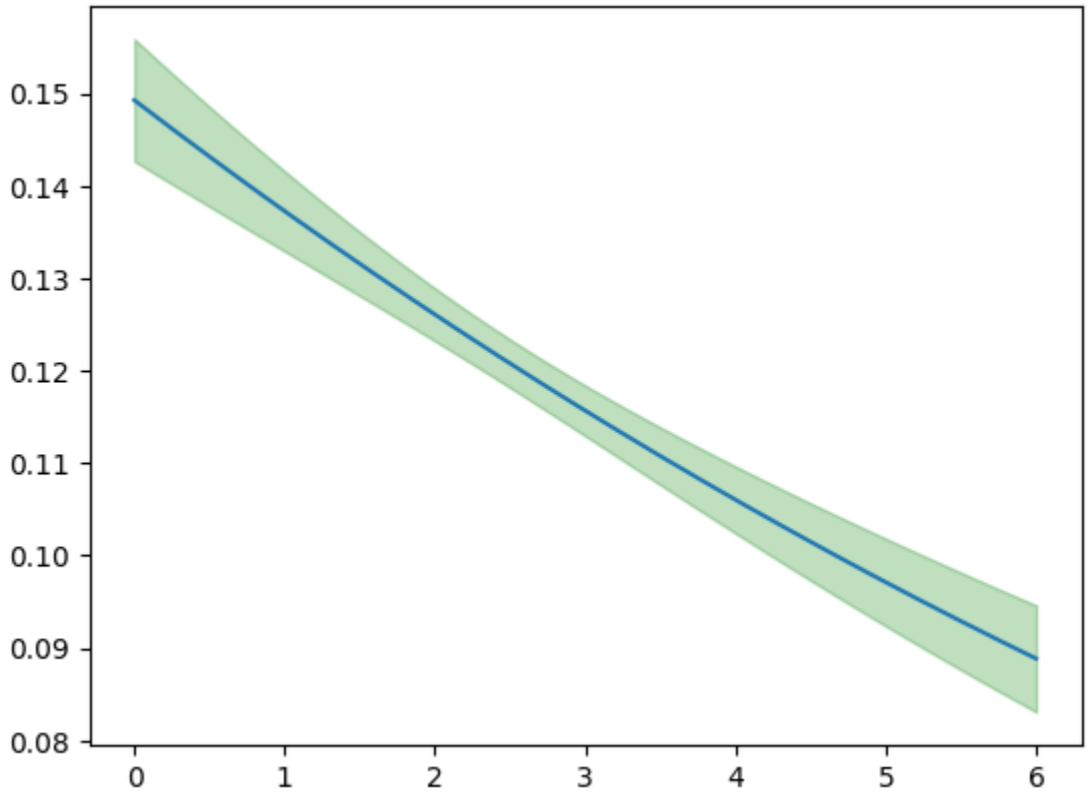


The confusion matrix shows that the model made correct predictions for all of the instances in the first class (True Negatives = 10,569), but it failed to predict any of the instances in the second class (False Negatives = 1,469). This indicates that the model is biased towards the first class and has poor predictive performance for the second class. Therefore, it is important to evaluate the model's performance using additional metrics such as precision, recall, and F1-score, which can provide more insights into the model's predictive ability for each class. Furthermore, it is necessary to investigate the cause of this bias

and explore possible solutions such as using a different model or adjusting the class weights to balance the performance of the model for both classes.

In conclusion, while the overall test accuracy of the model is high, the model's poor predictive performance for the second class suggests that further analysis and improvement of the model's performance are required before it can be used for reliable predictions in a real-world scenario.

Figure 8 *Prediction interval*



Binary response variable that indicates whether an arrestee is being strip searched or not. The logistic regression model estimates the likelihood of whether an arrestee would be strip searched or not. The predicted interval of strip search versus different age groups is shown above. As the figure shows the model can predict the possibility of strip search between around 0.08 and 0.16.

5. Discussion and Limitation

Our experiment has the following limitations:

- The sample size of perceived races is not very equal, and there are relatively few crime cases of minority groups compared with black or white. So it's limited to compare the strip search rate under different sample sizes.
- The given dataset has some empty data in our mail attributes, which lacks a certain level of accuracy. And we have dropped the U category from the *Sex* attribute since its sample size is relatively insufficient.
- Due to the lack of a reliable codebook, we need to assume the specific meanings of some attributes and their subcategories within a dataset, which may cause unclear aspects in our experimental results and display.
- After our logistic regression, our confusion matrix shows the result of a binary classification model. In our case study, the low recall indicates that the model is not very effective at identifying positive samples. This may be due to the model's threshold for classifying samples as positive being too high, leading to a large number of false negatives.

Preliminary Suppositions and Implications

Through our project, we hope to find out what combination of features affects strip search likelihood between perceived race, age, and gender groups. The main focus in our research is still perceived race, which is the reason why this attribute appeared in both ANOVA tests. This project will provide insight into how different perceived race, age and gender groups interact with strip search likelihood. Knowledge gained from this project will contribute to existing literature and help policymakers and police departments enhance fairness in law enforcement.

6. Conclusion

Stage 1:

The test results of our analyses revealed significant differences in the mean likelihood of being subjected to a strip search across different racial groups, age groups, gender groups, combined age and racial groups, and combined gender and racial groups. Specifically, we found that certain groups were more likely to be subjected to strip searches than other groups. The finding highlights the need for further investigation into potential inequalities in the criminal justice system.

Stage 2:

In subsequent experiments, we divided the cases into two groups based on whether they were recorded at the police station and conducted power analysis to confirm that our sample size was sufficient to achieve the desired statistical power. Experiments with sufficient sample size can increase our ability to detect effects and thus increase the reliability and repeatability of experiments. After using whether the case was booked as a covariate, we used ANCOVA to test whether the relationship between perceived race and whether the arrestee was subjected to a strip search was still significant. The results showed that the relationship between the two was still significant after registering the case as a covariate. Meanwhile, the covariate also showed a significant relationship with the dependent variable.

In the final logistic regression, we explored in depth the relationship between gender, race, age, and whether the arrestee was subjected to a strip search, and found that whether the case was booked with the police had the strongest association with whether the arrestee was subjected to a strip search. Arrestees in booked cases were more likely to be strip-searched, but the number of booked cases was relatively small compared to booked cases. Afterwards, we can use Confusion Matrix and Prediction interval to further validate the logistic regression model. The Confusion Matrix can help us calculate metrics such as accuracy, precision, recall, and F1 score of the model, thus providing a better understanding of the model's predictive ability. The Prediction interval can help us calculate the range and confidence level of the model's predictions for the target variable, thus providing a better understanding of the model's prediction effectiveness and reliability.

We believe that for the sake of social order and justice, it is important for more cases to be booked with the police and for arrestees to receive fair treatment, and to avoid explicit or implicit discrimination against arrestees based on race, age, or gender. This is an area where the Toronto Police Department needs to focus its efforts in the future.

7. References

Toronto crime rates in 2022. Vilkhov Law. (2022, September 22). Retrieved February 27, 2023, from https://vilkhovlaw.ca/toronto-crime-rates/

Moreau, G. (2022, August 2). This annual JURISTAT article presents findings from the 2021 uniform crime reporting survey. It examines trends in the volume and seriousness of police-reported crime for both violent and non-violent offenses at the national, provincial/territorial and census metropolitan area levels. specific violations, such as homicide, sexual assault, fraud and breaking and entering are examined, as well as trends in youth accused of crime. Government of Canada, Statistics Canada. Retrieved February 27, 2023, from https://www150.statcan.gc.ca/n1/pub/85-002-x/2022001/article/00013-eng.htm

Safe cities profile series: Key indicators by census metropolitan area Toronto, Ontario. Safe Cities profile series: Key indicators by census metropolitan area - Toronto, Ontario. (2020, May 15). Retrieved February 27, 2023, from https://www150.statcan.gc.ca/n1/pub/85-002-x/2020001/article/00001/toronto-eng.htm

Monika Lemke PhD Candidate. (2023, January 3). Strip searches are ineffective, unnecessary and target racialized Canadians. The Conversation. Retrieved February 27, 2023, from https://theconversation.com/strip-searches-are-ineffective-unnecessary-and-target-racialized-canadians-185187

Wortley, S., & Owusu-Bempah, A. (2011). The usual suspects: police stop and search practices in Canada. *Policing & Society*, *21*(4), 395–407. https://doi.org/10.1080/10439463.2011.610198

Tator, C., Henry, F., Smith, C., & Brown, M. (2006). *Racial Profiling in Canada: Challenging the Myth of" a Few Bad Apples"*. University of Toronto Press. https://books.google.ca/books?hl=en&lr=&id=n9CiinqsWIIC&oi=fnd&pg=PP1&dq=Racial+profiling+in+Canada&ots=qhuCCPuROU&sig=152PCb-KGN6x7LOduT0IgkQuENE#v=onepage&q=Racial%20profiling%20in%20Canada&f=false

Psutka, M., & Sheehy, E. (2016). Strip-searching of women in canada: wrongs and rights. *Canadian Bar Review, 94(2),* 241-280.

https://login.library.utoronto.ca/index.php?url=https://heinonline.org/HOL/P?h=hein.journals/canbarev94&i=241

8.Appendices

Tukey HSD test result

https://docs.google.com/document/d/1cCHhcaYUOtfXJiDQgC9tRo4Ds_p7KIp4sb_3TsNkR6s/edit