Report Investigating the Possible Intersections of Race and Gender (when Controlling for Age) with Strip Searching Practices by the Toronto Police Service.

by Jiawei Liu and Vlada Gorchkova

1. Introduction

When one is arrested they may be subject to a strip search performed by the arresting officers, however, this is a controversial and potentially humiliating practice and it can be seen as a violation of autonomy, individual privacy, and human rights. This practice is usually performed on arrestees or inmates, and involves all clothing to be removed for a visual inspection by an officer, intended to search for contraband or weapons that may be hidden on or within the body. The Toronto Police Service (TPS), which is responsible for patrolling and enforcing law in the Toronto Area, includes this practice as one potential procedure that occurs when someone is arrested. However, in recent decades, it has been suggested by activists and research alike that the TPS may be unfairly biased against individuals of certain racial, gender, or class backgrounds [1]. The purpose of this paper is to investigate if this bias against certain communities extends to TPS' likelihood to strip search an individual, and the implications that this may have for Toronto, the TPS, and the justice system in Canada at large.

First, we will provide a literature review on strip searching, the Toronto Police Service's history with discrimination, and the questions this may pose for the dataset we will be analyzing. Following, we will introduce the dataset, and discuss how it was prepared for analysis. Finally, we will analyze the dataset according to research questions that we prepared as an outcome of the literature review, and discuss the findings, followed by their implications.

2. Literature Review

Generally speaking, many have investigated the impact of racism (specifically anti-Black racism) in policing, leading to the over-policing of Black communities, and ultimately larger proportions of inmates from these communities [2]. In Canada, the policing of Indigenous communities and furthermore the sexualization of Indigenous women when policed, has also been disproportionate compared to White populations in Canada [10], meaning that there are different (and sometimes intersecting) identities that can affect the way that a person will be treated by Toronto Police officers.

When an individual is arrested, Toronto Police officers may follow certain procedures after arrest based on the level of risk that they assume an arrestee poses. One of such procedures is that of strip searching, however, some activists argue that strip searches can be considered a form of state-sanctioned sexual assault, as inmates are unable to consent to the procedure [5]. In the case of arrestees, as they have not yet been processed or booked, this procedure that may be conducted on them can possibly be violating their dignity, if not their rights, regardless of any proved criminal wrongdoing.

In Toronto, this may be a particular issue as about 56% of the population of Toronto identified as visible minorities in 2021, with 9% of the population as Indigenous [3], meaning that there are a high number of people that are vulnerable in the face of over-policing and the use of force disproportionate to white communities. Recent reports suggest this conclusion, indicating that Black individuals are more likely to be strip searched compared to others [8], and as such, this is something worth investigating in our assessment as well.

Another angle to the disparate impact on communities that has been discussed in recent years is the use of police services in response to mental health crises or incidents. With the death of Regis Korchinski-Paquet [14], Ejaz Choudry [13], and Bobby Ramroop [9], many are now calling into question whether TPS behaves with disproportionate or inappropriate force when responding to calls regarding those suffering a mental health incident. Disability justice activists argue that racialized individuals are especially vulnerable when experiencing these incidents [4], and in 2022, special response teams were formed in Toronto to provide the opportunity for non-police related responses [12].

Another concern that has often been levied against policing institutions is that of over policing youth that leads to the concept known as the school-to-prison pipeline [8]. This is when youth and minors are disproportionately punished for petty or minor infractions, leading to them being funneled into the prison industrial complex from a young age. This is of further concern if the minor is strip searched, as at such a young age such treatment may be incredibly traumatizing, and while it is challenging to find Canadian data on the subject, in the United Kingdom, Black boys are disproportionately strip-searched compared to any other race or gender [6], adding another dimension to an already troubling practice. Additionally, there have been instances of violence from police officers at arrest against elderly people [15], especially against elderly people of Indigenous background [7] indicating that intersections of race and age and the frequency of strip search may have different incident rates based on an individual's identity.

Ultimately, the behavior, race, gender, or age of arrestees should not impact their treatment. Whether one is behaving erratically or otherwise, is old or is young, they should be treated with reasonable fairness and dignity, and instead only logical reasons should be used to evaluate further procedures and processing, including strip searches. As the dataset allows for the recording of such factors when someone is arrested, this could be worth investigating.

3. Research Questions

3.1 Research Questions

1. Whether the number of actions that an arrestee takes when being arrested differs between arrestees that have been strip searched and what ways this may interact with race, while holding age constant?
2. Whether the number of actions that an arrestee takes when being arrested differs between arrestees that have been strip searched and what ways this may interact with gender, while holding age constant?
3. Whether the number of search reasons provided for a strip search vary depending on the gender of an arrestee, while holding age constant?
4. Whether the number of search reasons provided for a strip search vary depending on the race of an arrestee, while holding age constant?
5. Do the number of search reasons, number of actions, gender, race, and age group (or some combination of these variables) have an impact on the likelihood of an individual to be strip searched?

3.2 Hypotheses

1. The race and number of actions that an arrestee takes will have an impact on their likelihood to be strip-searched, while holding age constant.
2. The gender and number of actions that an arrestee takes will have an impact on their likelihood to be strip-searched, while holding age constant.
3. The number of search reasons will vary depending on the gender of the arrestee that has been strip searched, while holding age constant.
4. The number of search reasons will vary depending on the race of the arrestee that has been strip searched, while holding age constant.
5. There is a significant relationship between the predictor variables and the likelihood to be stripsearched when constructing a logistic regression model.

4. Dataset

To further investigate this issue, we will be analyzing a dataset including all Toronto Police Service data related to strip searches following arrest from 2020 to 2021 [11], provided by a Private Member on the TPS Open Data Portal. Variables that are recorded when someone is arrested include whether they were strip searched, their race, sex, offense, age group, among other demographics variables.

Furthermore, there are two other kinds of variables that are of interest in this analysis. The first is Search Reasons that are provided for conducting the strip search, and these are listed as a boolean for one or more of four different reasons (Caused Injury, Assisted Escape, Possessed Weapons, and Possessed Evidence). The second is Actions at Arrest, also listed as a boolean for one or more of five different kinds of actions (Combative, violent or spitter/biter; Resisted, defensive or escape risk; Mental instability or possibly suicidal; Assaulted officer; and Cooperative). To enhance the comprehensiveness of our data analysis, we have introduced two additional columns in our dataset. These columns are named "total actions at arrest" and "total strip search reasons," and they serve to summarize the values of "Actions at Arrest" and "Search Reasons," respectively.

Regarding race, sex, and age characteristics, race is recorded as any of the following categories: Black, East/Southeast Asian, Indigenous, Latino, Middle-Eastern, South Asian, White, and Unknown or Legacy. Sex is recorded as either M(ale), F(emale), or U(known). Age is recorded by the arrestees age group: 'Aged 17 years and younger', '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 years and older'.

5 Exploratory Data Analysis

5.1 Descriptive Statistics

5.1.1 Barplot to show the frequency of main variables

To explore the data, we created bar plots to visualize the frequency distribution of the related variables, including Race, Sex, Age Group, and StripSearch. This allowed us to quickly identify any patterns or imbalances in the distribution of these variables in our dataset.

Figure 1: Barplot to check the frequency of Race

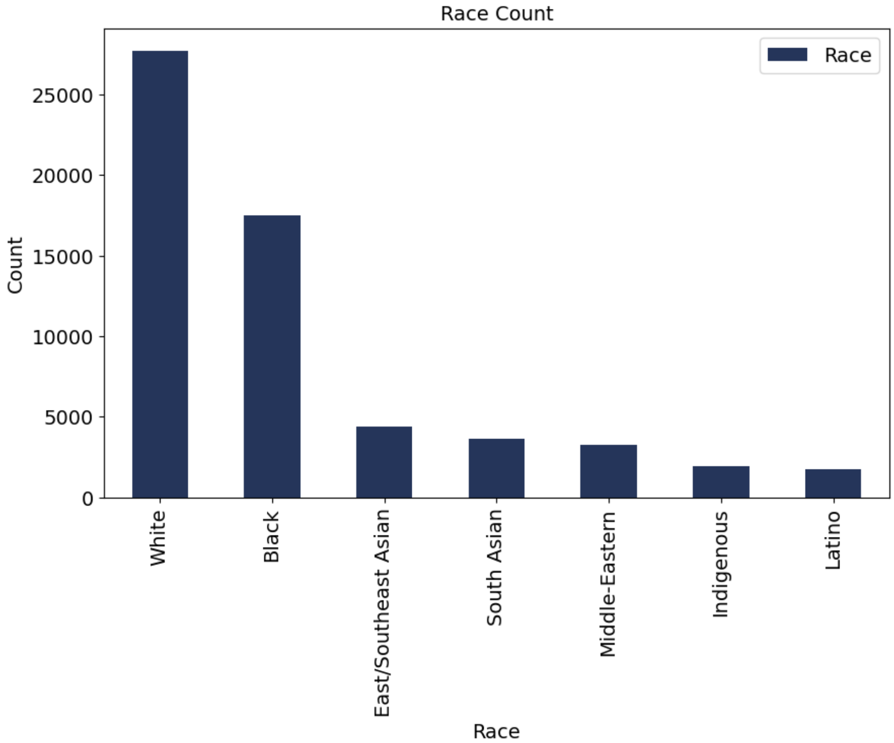


Figure 1 contains information on the race of the recorded arrestees. We created this table to gain an understanding of the racial composition of our dataset. The plot shows 7 distinct categories, and the majority of recorded arrestees are White or Black people, with an amount around 27500 and 17500, respectively.

Figure 2: Barplot to check the frequency of Gender

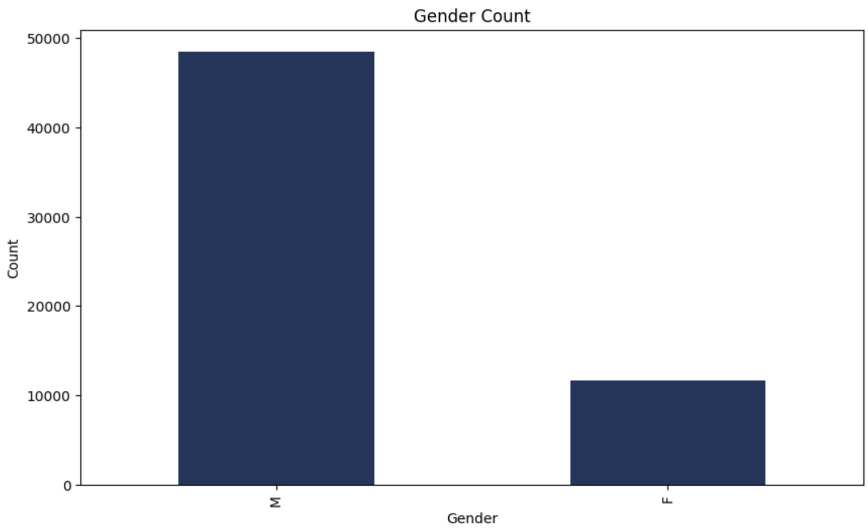


Figure 2 displays the gender distribution of arrestees. There are 4 times more male arrestees than females. The large difference in the number of male and female arrests highlights the potential impact of gender on the variable we are interested in, for example, the number of improper actions taken during an arrest. We will explore this relationship further later.

Figure 3: Barplot to check the frequency of Age Group

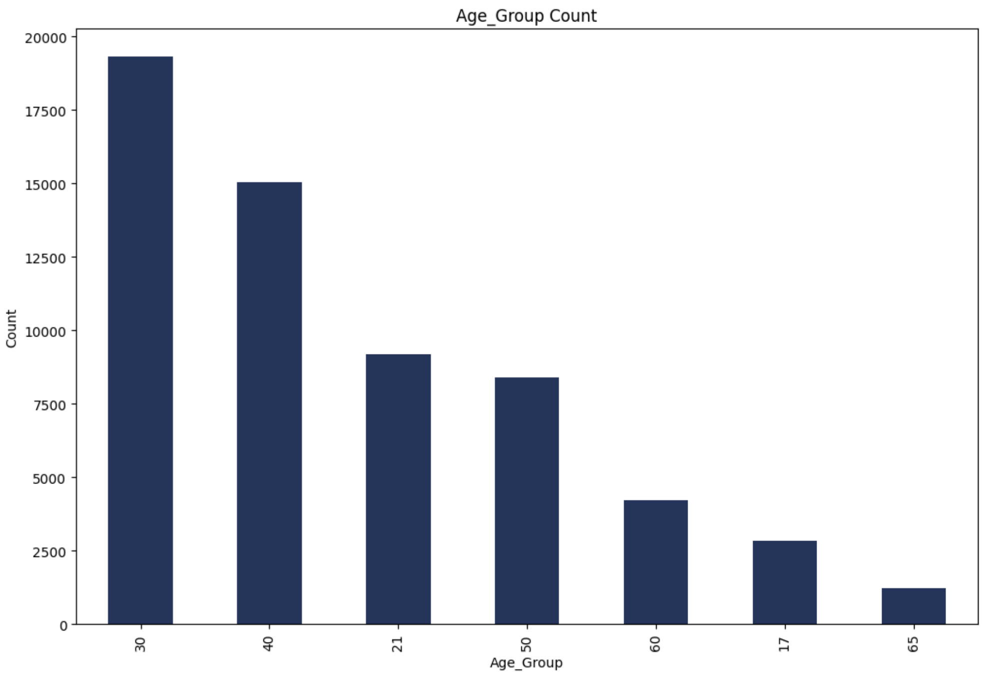


Figure 3 displays the age group distribution of arrestees. The majority of arrestees are aged between 30 to 40 years, in the original two age groups 25-34 and 35-44 years. While both young (under 17) and old (over 65) arrestees are fewer, especially the old individuals.

Figure 4: Barplot to check whether strip searched

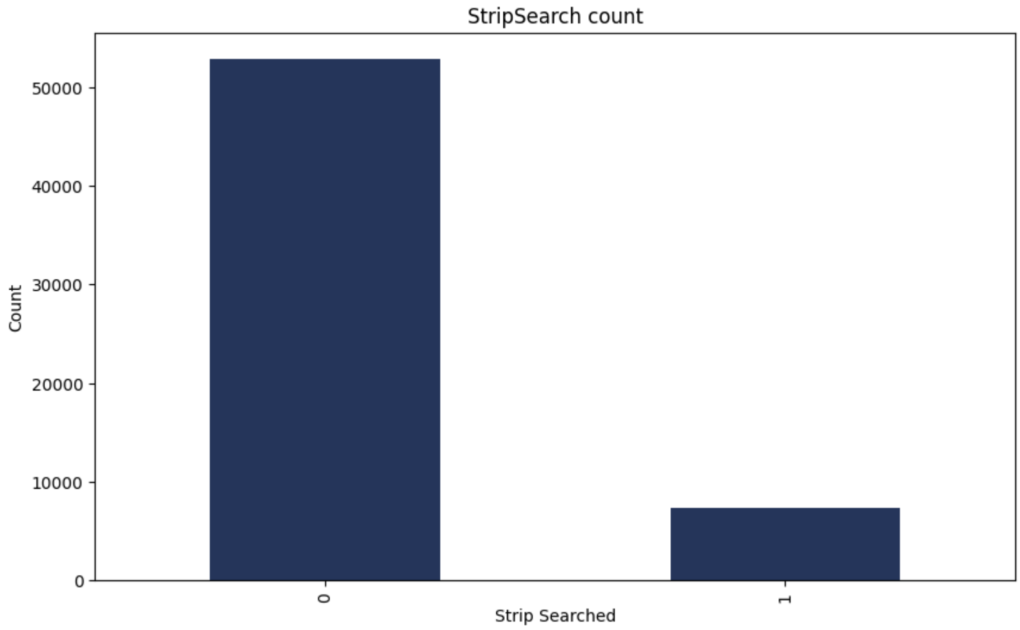


Figure 4 shows that the number of arrestees who will be strip searched is fewer than that without strip searches. Then, for the people who are strip-searched, we wondered whether race and gender will impact the probability to be strip-searched. Therefore, several pivot tables were created.

5.1.2 Pivot table for Strip search probability based on Race, Sex, and Age group

Table1

Race	Probability of strip-search		Sex	Probability of strip-search		Age Group	Probability of strip-search
Black	0.139		Female	0.102		17	0.092
East/South east Asian	0.077		Male	0.125		21	0.136
Indigenous	0.158					30	0.133
Latino	0.075					40	0.131
Middle-East ern	0.070					50	0.101
South Asian	0.071					60	0.082
White	0.129					65	0.027

Table 1 displays the probability of being strip-searched based on race, sex, and age group. We found for the Indigenous category, they are most likely (15.8%) to be strip-searched after being arrested. Followed by the White and the Black groups, the probabilities are 13.9% and 12.9%, respectively. However, from figure one, the amount of Indigenous arrestees is quite small, so the highest probability should be examined carefully. Then the probability of strip search for female arrestees (0.102) is a little bit lower than that of males (0.125), in other words, gender might be a factor to impact strip search. Furthermore, individuals aged between 30 to 40 years (age groups 25-34 and 35-44 years) are likely to be strip-searched, with the probability around 13%. Meanwhile, both young (under 17) and old (over 65) arrestees are less likely to be strip-searched, especially those over 65, with the smallest probability of 2.7%.

5.1.3 Boxplot

Figure 5: boxplot for Race, gender, and total search Reasons.

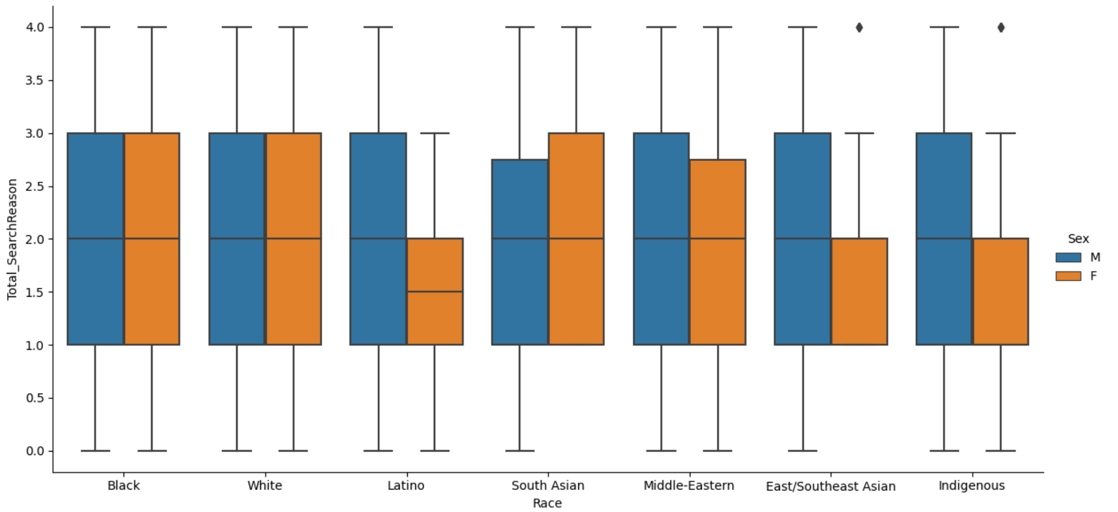
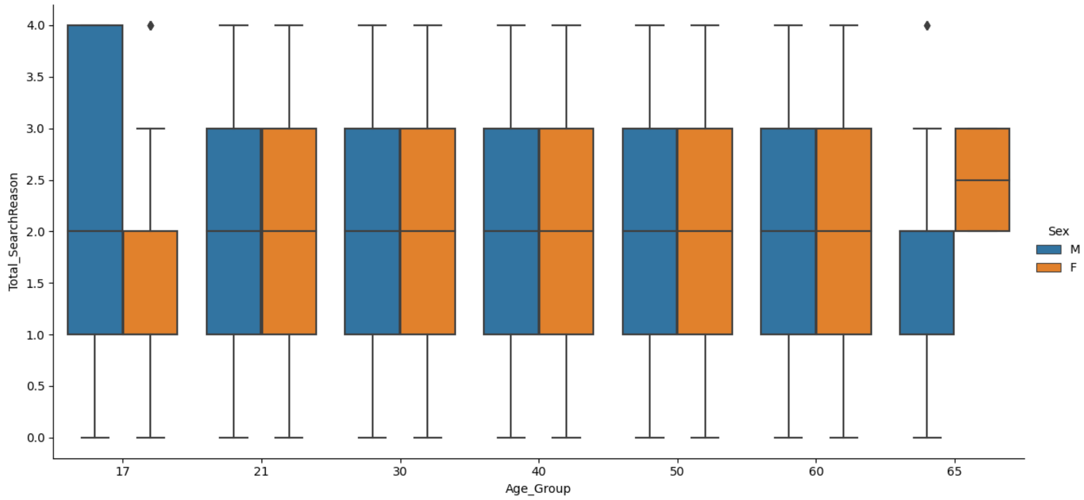


Figure 6: boxplot for age group, gender, and total search Reasons.



Since we are interested in whether the number of search reasons provided for a strip search varies depending on the race, gender, and age group of an arrestee, Figures 5 and 6 above visualize the relationship between gender, race/age group, and the total number of strip search reasons for arrestees, respectively. The range of search reasons is from 0 to 4.

First, we examined the relationship between gender and total search reasons based on race/ age group. For the majority of races, the distributions of the total search reasons for males and females were different, while for the two largest racial groups, Black and White, the data distributions were almost the same. However, for the majority age groups, the distributions of the total search reasons for males and females are the same, except for arrests under 17 years or over 65 years.

Next, we looked at the relationship between race/age group and total search reasons based on gender. Among almost all races, 50% of the data (Interquartile range) is distributed from 1-3 search reasons, while only IQR of females for Latino, East/Southeast Asian, and Indigenous are from 1-2 search reasons, and 4 search reasons tend to be considered as outliers for them. Similar to races, for the age groups, 50% of the data (Interquartile range) is distributed from 1-3 search reasons while IQR for males under 17 years is 1-4 reasons, and that of females is 1-2 reasons.

5.2 Chi-square test

5.2.1 Chi-square test: Race and StripSearch

In this case, Race and Strip Search are both categorical variables, since a t-test requires continuous data, it is not appropriate for testing the relationship between a categorical and binary variable. Therefore, to determine if there is a significant association between Race and Strip Search, we conducted a chi-square test, which is the appropriate statistical test for this scenario. The hypothesis was:

H_0: There is no significant association between Race and Strip Search.

H_a: There is a significant association between Race and Strip Search.

For the result, we got a chi-square statistic of 371.781, and a p-value of $3.242*10^{-77}$ which was smaller than 0.05. Therefore, the null hypothesis can be rejected, and we can conclude there was a significant association between race and being strip searched.

5.2.2 Chi-square test: Sex and StripSearch

Similar to the case before, a chi-square test was conducted for Sex and StripSearch. The hypothesis were:

H_0: There is no significant association between Sex and Strip Search.

H_a: There is a significant association between Sex and Strip Search.

We received a chi-square statistic of 49.508 and a p-value of $1.975*10^{-12}$ which was smaller than 0.05. Therefore, the null hypothesis can be rejected, and we can conclude there was a significant association between Sex and being strip searched.

5.2.3 Chi-square test: Age group and StripSearch

A chi-square test was conducted for the Age group and StripSearch. The hypothesis were:

H_0: There is no significant association between Age group and Strip Search.

H_a: There is a significant association between Age group and Strip Search.

We received a chi-square statistic of 271.101 and a p-value of $1.261*10^{-55}$ which was smaller than 0.05. Therefore, the null hypothesis can be rejected, and we can conclude there was a significant association between the Age group and being strip-searched.

5.3 T-Tests

5.3.1 T-Test: Race and Total Actions at arrest

Since it was a two-sample t-test, we chose White and Black these two representative races (the majority of arrestees). The purpose was to explore whether there is a significant difference in the mean total actions at arrest for people who are strip searched between White and Black arrestees.

Prior to conducting the t-test, we performed a Levene test to evaluate the homogeneity of variances between the two groups. The hypothesis were:

H_0: The variance between groups of White and Black is equal.

H_a: The variance between groups of White and Black is unequal.

The resulting p-value was $1.006*10^{-5}$, smaller than the significance level of 0.05, indicating that there was enough evidence to reject the null hypothesis, and there was a significant difference to the variances of the two groups. Therefore, we should conduct a Welch's t-test. The hypothesis for Welch's t-test were:

H_0: The sample means of the total number of actions at arrest for White and Black arrestees are equal.

H_a: The sample means of the total number of actions at arrest for White and Black arrestees are not equal

Table 2: Welch's t-test result for Race and Total actions at arrest

Race	sample means	sd.	t-statistic	p-value	CI	dof
White	0.230	0.576	-4.312	$1.649*10^{-5}$	[-0.103,-0.039]	4763
Black	0.301	0.657				

The results from Table 2 indicate that the mean of total actions at arrest for the White group (M=0.2302, SD=0.576) is slightly lower than that of the Black group (M=0.301, SD=0.657). With alpha established at 0.05, this is a statistically significant difference as the p-value ($1.649*10^{-5}$) is less than 0.05, and 95% CI [-0.103,-0.039]. Therefore, we can reject the null hypothesis that there is a significant difference in the number of total actions at arrest between the White and the Black who are strip-searched.

5.3.2 T-Test: Sex and Total Actions at Arrest

To explore the relationship between gender and the total number of actions at arrest, we conducted a similar Levene test to check the homogeneity of variances, the resulting p-value was 0.426, which was greater than the significance level of 0.05, indicating that we did not have enough evidence to reject the null hypothesis and the variances of the two groups can be assumed to be equal. Then proceeded to conduct a normal two-sample t-test. Our hypotheses for the t-test were:

H_0: The sample means of the total number of actions at arrest for Female and Male arrestees are equal.

H_a: The sample means of the total number of actions at arrest for Female and Male arrestees are not equal

Table 3: T-test results for Gender and Total actions at arrest

Sex	sample means	sd.	t-statistic	p-value	CI	dof
Female	0.244	0.599	-0.796	0.426	[-0.053,0.022]	7262
Male	0.260	0.610				

The results from Table 3 indicate that the mean of total actions at arrest for females (M=0.244, SD=0.599) is slightly lower than that of the Black(M=0.260, SD=0.610). With alpha established at 0.05, there's not enough evidence to say the sample means are statistically significant difference as the p-value (0.426) is larger than 0.05, and 95% CI [-0.053,0.022]. Therefore, we can not reject the null hypothesis and there does not exist a significant difference in the number of total actions between females and males who are strip-searched.

5.3.3 T-Test: Race and Total Search Reasons

To investigate the relationship between race and the total number of strip search reasons, we still chose to focus on the White and Black racial groups in our analysis. The resulting p-value from Levene's test was 0.165, which was greater than the significance level of 0.05, indicating that the variances of the two groups can be assumed to be equal. We then proceeded to conduct a normal two-sample t-test. Our hypotheses for the t-test were:

H_0: The sample means of the total number of reasons to perform strip searches for White and Black arrestees are equal.

H_a: The sample means of the total number of reasons to perform a strip search for White and Black arrestees are not equal

Table 4: t-test result for Race and Total Search reasons

Race	sample means	sd.	t-statistic	p-value	CI	dof
White	1.965	1.213	-2.018	0.044	[-0.127,-0.002]	5998
Black	2.029	1.206				

The results from Table4 indicate that the mean total strip search reasons for the White (M=1.965, SD=1.213) are slightly lower than that of the Black(M=2.029, SD=1.206). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.044) is less than 0.05, and 95% CI [-0.127,-0.002]. Therefore, we can reject the null hypothesis that there is a significant difference in the number of total reasons for strip searches between the White and the Black.

5.3.4 T-Test: Gender and Total Search Reasons

To explore the relationship between gender and the total number of strip search reasons, similarly, we conducted a Levene test at first, the resulting p-value was 0.314, which was greater than the significance level of 0.05, indicating that the variances of the two groups can be assumed to be equal. Then proceeded to conduct a normal two-sample t-test. Our hypotheses for the t-test were:

H_0: The sample means of the total number of reasons to perform strip searches for Female and Male arrestees are equal.

H_a: The sample means of the total number of reasons to perform a strip search Female and Male arrestees are not equal

Table 5: t-test result for Gender and Total Search reasons

Sex	sample means	sd.	t-statistic	p-value	CI	dof
Female	1.871	1.180	-3.611	0.0003	[-0.213,-0.063]	7269
Male	2.009	1.212				

The results from Table6 indicate that the mean total strip search reasons for females (M=1.871, SD=1.180) is slightly lower than that of the Black(M=2.009, SD=1.212). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0003) is less than 0.05, and 95% CI [-0.213,-0.063]. Therefore, we can reject the null hypothesis that there is a significant difference in the number of total reasons for strip searches between females and males.

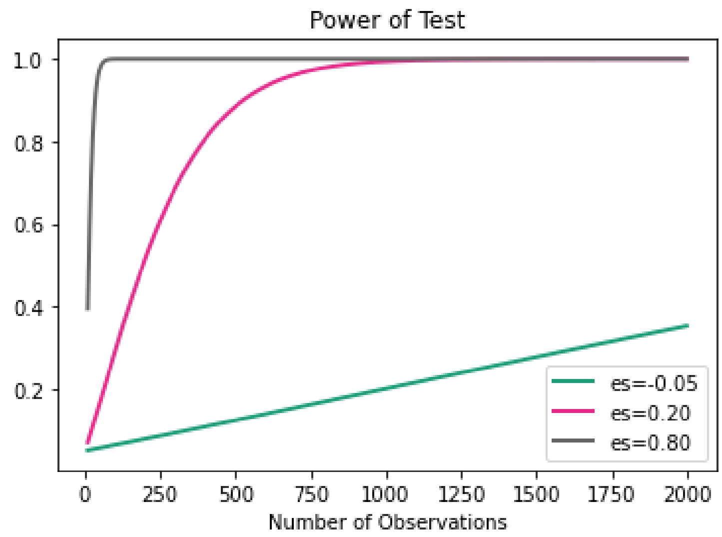
5.4 Power Analysis for T-Tests of Race VS. Total Search Reason

To perform the power analysis, we first defined functions to calculate Cohen's D as the effect size for the T-test. The effect size (Cohen's D) for the search reason based on race was -0.0531. Generally, a Cohen's D effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 or greater is considered large; thus, -0.0531 is a very small value. This indicates that the difference between the groups is unlikely to be significant. Additionally, the p-value for the T-test is 0.044, which is very close to 0.05. We can still use -0.0531 as the effect size to calculate the proposed sample size for the selected races (White and Black). We determined the required sample size using the obtained effect size and set the statistical power at 80%. The results showed that a sample size of 4692 was necessary for Black individuals who were strip-searched, while 6874 was needed for White individuals. This is significant because the dataset provided sample sizes of 2434 and 3566, respectively, which may affect the reliability of the results.

We then calculated the statistical power based on the actual effect size (-0.0531) and sample size, with a significance level of 0.05. The power of the T-test was 0.744, which indicates a 25.6% probability of committing a Type II error. To increase the statistical power, we can expand the sample size and control for more confounding variables to reduce variability.

Lastly, we generated a power curve plot to visualize how the power changed with different effect sizes as the sample size increased.

Figure 7: Power graph for t-test between Gender and Total Search Reasons.



From the graph, we observed that if the effect size is 0.8, 100 observations are sufficient to achieve a power of 1.0, while for an effect size of 0.2, we need approximately 1000 observations to reach a power of 1.0. Regrettably, for the actual effect size in this case, with 2000 observations, the power is only 0.4. This is why, despite having a total sample size of 6000, the power remains at 0.744.

6 Research Design and Methods

6.1 Research Design

As the dataset was externally generated, we were unable to enact data collection processes of our own and had only the resulting dataset to work with. To begin our analysis, we had to clean the dataset to make it intelligible, for instance by removing data points with the "U" marker for sex (of which there were only 6) and changing the value of *Age_Group* from categories to numerical integers. Following this, in the course of exploratory data analysis, we determined that feature engineering needed to be performed and created two new parameters, as demonstrated in the Exploratory Data Analysis. The first was *'Total Search Reasons'* which recorded the number of search reasons provided to conduct the search, and the second was *'Total Actions at Arrest'* which recorded the number of negative actions that the arrestee performed in the course of the arrest. This would create variables that could undergo t-testing and further statistical analysis.

6.2 Methods

In our analysis, statistical significance was determined at an alpha level of 0.05. There was no weighting of factors in determining statistical significance - the limitations of such an approach are discussed in section 8.2. T-tests were conducted to identify patterns in the dataset worthy of investigation. Moreover, ANCOVA and Logistic Regression were conducted to identify subsequent differences and predict the probability of a strip search occurring. The power analysis would be conducted before the test to find the proposed sample size, and also after the test to validate the findings by checking the statistical power.

6.3 Power analysis before ANCOVA

We conducted a power analysis to determine the recommended sample size for each group in our ANCOVA study. In this study, we have two groups: one for the treatment and one as a control group, which serves as a covariate. We set the desired effect size (Cohen's f) at 0.25, the target power at 0.8, and the significance level at 0.05. Based on these parameters, our power analysis yielded a proposed sample size of 128 participants per group.

7 Results and Discussion

7.1 ANCOVA

Following the results of the t-test and chi-square test in exploratory data analysis, we were interested in determining if there exist differences in the mean between all race categories that are recorded in this dataset and Actions at Arrest/Total Seatch Reasons for people who strip-searched. We conducted an ANCOVA test and set Age_group as the covariate, since after EDA, we suspected Age might influence Dependent and Independent Variables.

7.1.1 Race and Actions at Arrest for people who strip-searched, holding Age constant

H_0: There are no significant differences in mean actions at arrest between any races while holding age constant.

H_A: There exist differences in mean actions at arrest between any of the races while holding

Table 6: ANCOVA Race and Total Actions

≈	SUM OF SQUARES	DEGREES OF FREEDOM	F	Uncorrected p-values.	Partial eta-squared
RACE	11.486780	6	5.195426	0.000017	4.277730e-03
Age_Group	0.000023	1	0.000063	0.993681	8.645020e-09
Residual	2673.764511	7256	NaN	NaN	NaN

The uncorrected p-value for Race is smaller than 0.05. We can reject the null hypothesis that there exist differences in mean actions at arrest between any of the races after controlling for the age of individuals.

Moreover, the p-value for Age_Group is 0.994, bigger than 0.05, which indicates controlling age groups might not contribute to making more accurate results in this case.

7.1.2 Sex and Actions at Arrest when Controlling Age

As the t-tests did not indicate that there were significant differences, we wanted to explore whether ANCOVA would have different results when controlling Age groups.

H_0: There is no difference in total actions at arrest means between the sexes while holding age constant.

H_A: There exists a difference in total actions at arrest means between the sexes while holding age constant.

Table 7: ANCOVA Sex and total Actions

≈	SUM OF SQUARES	DEGREES OF FREEDOM	F	Uncorrected p-values.	Partial eta-squared
Sex	0.252174	1	0.681950	0.408943	0.000094
Age_Group	0.394527	1	1.066913	0.301678	0.000147
Residual	2684.999117	7261	NaN	NaN	NaN

The uncorrected p-value for Sex is 0.410, bigger than 0.05. We can not reject the null hypothesis that males and females result in the same mean of the total number of actions even controlling for the age of individuals.

7.1.3 Race and Total Search Reasons when controlling Age

As the significant level of the t-test for Race vs.Total Search Reason is not high (p-value: 0.04), we wondered about the performance of Ancova while controlling the Age.

H_0: There is no difference in total search reason means between any races while holding age constant.

H_A: There exists a difference in means between any races while holding age constant.

Table 8: ANCOVA Race and Total Search Reasons

≈	SUM OF SQUARES	DEGREES OF FREEDOM	F	Uncorrected p-values.	Partial eta-squared
Race	9.881744	6	1.132734	0.340213	0.000936
Age_Group	26.361787	1	18.130890	0.000021	0.002493
Residual	10550.013329	7256	NaN	NaN	NaN

The uncorrected p-value for Race is 0.340, bigger than 0.05. We can not reject the null hypothesis that races result in the same mean of the total number of search reasons even controlling for the age of individuals.

The p-value for Age_Group is smaller than 0.05, which indicates controlling age groups will contribute to making more accurate results in this case. Age_Group has relations to both DV and IV, that's why we received different results from the t-test.

7.1.4 Sex and Total Search Reasons when Controlling Age

The hypotheses were as follows:

H_0: There is no difference in total search reason means between sexes while holding age constant.

H_A: There exists a difference in total search reason means between sexes while holding age constant.

Table 9: ANCOVA Sex and Total Search Reasons

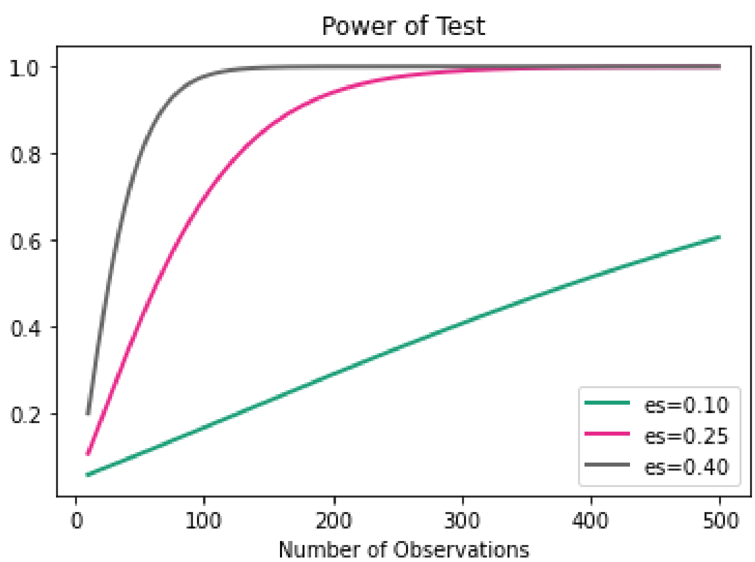
≈	SUM OF SQUARES	DEGREES OF FREEDOM	F	Uncorrected p-values.	Partial eta-squared
Sex	20.459487	1	14.095284	0.000175	0.001937
Age_Group	32.308600	1	22.258568	0.000002	0.003056
Residual	10539.435615	7261	NaN	NaN	NaN

The uncorrected p-value for Sex is 0.000175, smaller than 0.05. We can reject the null hypothesis that there exist differences in the mean amount of search reasons between males and females after controlling for the age of individuals. The p-value for Age_Group is smaller than 0.05, which indicates controlling age groups will contribute to making more accurate results in this case.

7.1.5 Power analysis to test the statistical power of ANCOVA

We maintained the desired effect size at 0.25 and the significance level at 0.05. However, we used the actual sample size of 7,264, representing the number of individuals subjected to strip searches by the police. With these parameters, we obtained the maximum power of 1.0, indicating a 100% probability of correctly rejecting the null hypothesis when it is false. It is important to note that the effect size might vary across different cases, which could impact the statistical power of the test. The figure below illustrates how the power of ANCOVA changes in response to variations in sample size and effect size.

Figure 8:power graph for t-test between gender and total search Reasons.



From the graph, we can conclude that as the effect size increases, smaller sample sizes are needed to achieve the desired statistical power (usually 0.8). For example, 60 and 130 observations are sufficient to achieve a power of 0.8 for effect sizes of 0.4 and 0.25, respectively. However, for an effect size of 0.4, even with 500 observations in each group, the power remains only 0.6. This illustrates the importance of considering both effect size and sample size when designing a study to ensure that the desired level of statistical power is achieved.

7.2 Logistic Regression

Based on our findings, we decided to implement logistic regression analysis to determine if there is a relationship with the identity characteristics we have outlined and investigated previously in this report and an individual being strip searched by the Toronto Police upon being arrested. Of all the previous features investigated, Total Search Reasons was not used as a feature as that would indicate someone has already been searched. Total Actions Taken at Arrest, Gender, Race and Age Group were included in the analysis.

Our hypotheses for this model were:

> H_0: There is no relationship between the variables outlined and whether an individual is strip searched by the Toronto Police department upon arrest (H0: $\beta1 = \beta2 = \beta3, = \beta4 = 0$).
>
> H_a: There is a relationship between the variables outlined and whether an individual is strip searched by the Toronto Police department upon arrest (H0:At least one βi is not equal to 0).

The results of the logistic regression analysis are presented in the following tables.

7.2.1 Logistic Regression Results

	Co-Eficcient	*Std Error*	*Z Score*	*P>|z|*	*[0.025*	*0.975]*
Age Group	-0.0491	0.001	-58.616	0.000	-0.051	-0.047
Total Actions	0.4782	0.025	18.883	0.00	0.429	0.528
Race (White)	0.0612	0.028	2.204	0.028	0.007	0.116
Sex (Male)	-0.3880	0.028	-13.893	0.000	-0.443	-0.33

Using this results table, we can reject the null hypothesis ($p<0.05$). All of the coefficients are not equal to zero, and so we can confirm that there is a relationship between the total actions at arrest, gender, race, and age group as they relate to whether someone will be strip searched on arrest.

With this in mind, we generated the odds ratios for all of the variables to interpret the relationship that they have with an individual's likelihood to be strip searched.

7.2.2 Odds Ratio Results

	Lower CI	*Upper CI*	*OR*
Age Group	0.9506	0.9537	0.9521
Total Actions	1.5350	1.6952	1.6131
Race	1.0068	1.1226	1.0631
Sex (Male)	0.6422	0.7166	0.6784

Overall, the results of the logistic regression re-affirmed that there is a significant relationship ($p<0.05$) between the identity and actions of an arrestee and whether they are strip searched. With regards to identity, age (OR=0.9251, 95% CI = 0.9506-0.9537) and sex (OR=0.6784, 95% CI = 0.6422-0.7166) both had a negative association with being strip searched. That is to say that men are 33% less likely than women to be strip searched (according to the logistic regression model), and as for an arrestee's age, for every unit increase in age, the odds of being strip searched decrease by about 5%. Race (OR=1.0631, 95% CI = 1.0068-1.1226) and Total Actions at Arrest (OR=1.6131, 95% CI = 1.5350-1.6952) however had a positive association with being strip searched. Interpreted plainly, a white arrestee is about 1% more likely to be strip searched than an individual of any other race. As well, for every unit increase in total actions at arrest (so for every additional action taken) the likelihood of being arrested increases by about 6%.

To determine if this model generalizes well, we tested its accuracy and generated a confusion matrix on the test dataset. The dataset had an 80-20 split for training and test data.

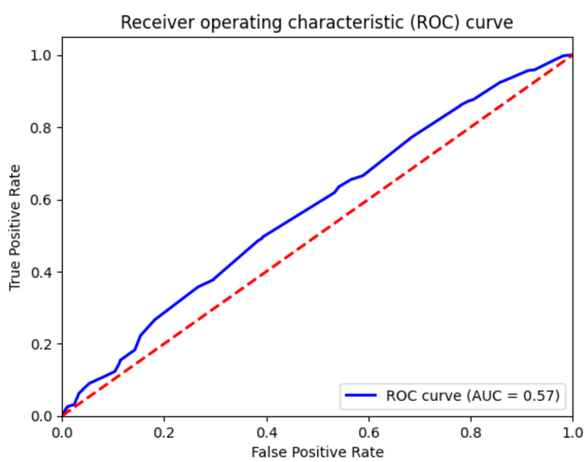
The test accuracy of the model was determined to be 0.8772, however, the F1 Score is very low at 0.0067. This result can also be seen in the confusion matrix listed below:

7.2.3 Confusion Matrix

	Pred. no SS	*Pred. SS*
True no SS	10555	14
True SS	1464	5

As can be seen, the model performed poorly in identifying positive cases - that is to say identifying cases where the arrestee was strip searched. This can further be evidenced by the receiver operating characteristic curve pictured below:

Figure 9: ROC Curve for Logistic Regression Model



With an AUC value of 0.57, the model is only slightly better at distinguishing between positive and negative cases than random chance.

Based on the model's performance, a prediction interval was prepared and plotted. However as the model's performance was questionable, this information is not very informative. The plot of the prediction interval can be found below, as well as a plot of the first 100 values to give a sense of the shape of predictions, where the data points are more intelligible.

Figure 10: Logistic Regression Model with Upper and Lower Bound

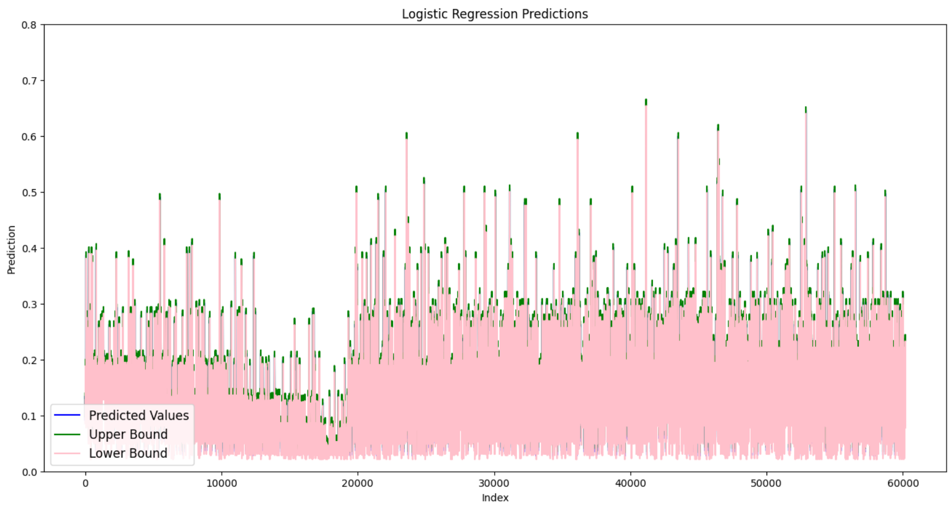
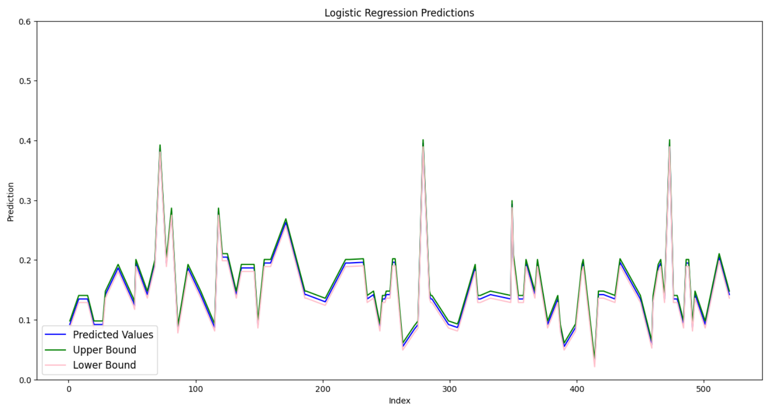


Figure 10: Logistic Regression Model with Upper/Lower Bound for the First 100 Predictions



Ultimately, this model errs towards classifying entries as negative cases rather than positive ones and as there are many false negatives, the Type II error in this model is incredibly high.

8. Discussion

The ANCOVA tests conducted offer valuable insights into the research questions initially presented, elucidating the relationships between race, gender, and actions during arrests or strip searches while taking age into account as a covariate.

The analysis revealed that racial differences might affect the number of actions taken by arrestees during the arrest process, regardless of their age. However, it is essential to emphasize that the age group itself did not significantly contribute to the model in this specific context, as indicated by the p-value for Age_Group exceeding 0.05.

Furthermore, the ANCOVA results showed no significant disparities in the mean actions at arrest between genders when age was held constant. This finding suggests that male and female arrestees exhibit similar behavior during arrests, irrespective of their age.

Regarding the research question about Race vs. Total Search Reasons, the ANCOVA test indicated that race did not significantly impact the number of reasons provided for strip searches. Interestingly, the t-test for the same research question yielded a p-value below 0.05, suggesting that race could influence the number of reasons for strip searches. However, in the ANCOVA, controlling for age led to a more accurate result, as the p-value for Age_Group was below 0.05.

Lastly, the ANCOVA test revealed significant differences in the mean number of search reasons between males and females when age was held constant. This outcome suggests that gender may indeed affect the number of reasons provided during strip searches, with age playing a crucial role in the model.

This finding was reflected in our logistic regression model, as the gender of a participant had a statistically significant effect on the likelihood of them being strip searched. It also found that total actions at arrest, race, and age group also had a statistically significant effect on the likelihood of a participant being strip searched, findings that were not necessarily evident in ANCOVA testing. Ultimately, however, since the logistic regression model was incredibly likely to commit Type II error - these findings should be taken with a grain of salt.

9. Conclusion, Limitations, and Next Steps

9.1 Conclusion

Ultimately, our findings in this research paper reaffirm the findings in the literature review that the race and gender of an individual significantly affect their likelihood of being strip searched upon arrest, when controlling for age group. For race, this was demonstrated in the total actions taken at arrest, and for gender this was demonstrated in the total search reasons provided for the strip search. Our logistic regression model supported such a conclusion and the coefficients for age group, race, gender, and total actions at arrest all indicated that they had an effect on the likelihood of an individual to be strip searched.

9.2 Limitations

The first limitation speaks to how robust the collected data is. For some parameters, it could be subject to the arresting officers' discretion and so could result in biased data. For instance where one may be upset at being arrested, an officer might record that the arrestee is combative or violent when they merely voice their displeasure, as mentioned in the discussion. There is no way of knowing how an arrestee was actually behaving at the time of their arrest. Similarly, the reasons for initiating a search could be inaccurate or the number of search reasons could be inflated, as an officer might have no reason to believe that a particular arrestee has a weapon, but is likely not required to defend their reason before conducting the search. As such, they are free to provide as many or as few reasons as they like for the search, as there is no 'rulebook' for such reasons specifically, but instead relies on the arresting officer's better judgment, which as determined in the literature review, can be biased.

Our manipulation of the dataset also poses a limitation. As no exact age is recorded for arrestees we converted the categorical variable of 'age group' to a continuous one by taking the mean age of the age group for each arrestee. Both our ANCOVA model and the Logistic Regression model treat this variable as continuous (as age is indeed a continuous variable), however, as we performed such feature engineering, there are only 7 unique variations for this variable. This likely affects the performance of both of these models, and some unique insights are likely missed on account of it.

Another limitation is that we chose an alpha value of 0.05. In some instances, if we had chosen another value, and therefore another power, for instance with an alpha value of 0.1, then for some sections (for instance race and search reasons), we would have found statistically significant results. As this is sensitive data, we chose a more conservative statistical power so that we do not erroneously commit a Type 1 error.

Furthermore, the types of explanatory variables we could use were limited. The dataset had limited features, and so any model that is generated from it may lack explanatory power. As we did not create this dataset ourselves, there is nothing we can do to mitigate this, except for accept that the models may more poorly reflect a reality in the underlying population.

The final limitation is regarding the model creation. Logistic Regression models are not always well suited at predicting events that occur relatively infrequently. In the dataset that we have used, there were 52923 (~88% of the dataset) arrestees that were not strip searched, compared to 7264 (~12%) that were strip searched, meaning that we have an imbalanced dataset. For such types of data it is typically recommended to use models that are better suited for handling imbalance data, such as Support Vector Machines.

9.3 Next Steps

To investigate this question further there are several courses of action that we would recommend. The first, as with our previous investigation, we would recommend connecting external datasets to this analysis. Investigating census data and seeing the ways in which Toronto's population may or may not be represented, and may or may not co-vary in the strip-search dataset would be of interest to us.

Another recommended course of action is using different modeling techniques to generate a prediction model from this data. As mentioned previously in 9.2 Limitations, a Logistic Regression model is not always well suited for imbalance datasets. Instead Random Forest or Support Vector Machines could be applied to the dataset for possibly increased accuracy.

5. References

[1] Deloitte LLP. 2022. *Workplace Well-Being, Harassment and Discrimination Review: Report prepared for The Toronto Police Services Board and the Toronto Police Service*. Retrieved from https://www.tps.ca/media/filer_public/d4/7f/d47fd156-41f5-4ec0-a2b1-1002d9682baf/06b1320e-15a0-48a2-a676-af88640fc5f6.pdf

[2] Department of Justice Government of Canada. 2022. Overrepresentation of Black People in the Canadian Criminal Justice System. Retrieved February 27, 2023 from https://www.justice.gc.ca/eng/rp-pr/jr/obpccjs-spnsjpc/index.html

[3] Statistics Canada Government of Canada. 2022. Profile table, Census Profile, 2021 Census of Population - Toronto, City (C) [Census subdivision], Ontario. Retrieved February 23, 2023 from https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

[4] Kelly Hayes. 2023. Disability Justice Organizers Dream Big and Resist a Culture of Disposability. *Truthout*. Retrieved February 23, 2023 from https://truthout.org/audio/disability-justice-organizers-dream-big-and-resist-a-culture-of-disposability/

[5] Jessica Hutchison. 2020. “It’s Sexual Assault. It’s Barbaric”: Strip Searching in Women’s Prisons as State-Inflicted Sexual Assault. *Affilia* 35, 2 (May 2020), 160–176. DOI:https://doi.org/10.1177/0886109919878274

[6] Isabella Kwai. 2022. Strip-Searches of Children by London Police Are Called ‘Deeply Concerning.’ *The New York Times*. Retrieved April 12, 2023 from https://www.nytimes.com/2022/08/08/world/europe/london-police-strip-search-minors.html

[7] Brandi Morin. 2020. Chief Allan Adam on being beaten by police and Indigenous rights. Retrieved April 12, 2023 from https://www.aljazeera.com/features/2020/7/15/chief-allan-adam-on-being-beaten-by-police-and-indigenous-rights

[8] Jason P. Nance. 2015. Students, Police, and the School-to-Prison Pipeline. *Wash. Univ. Law Rev.* 93, 4 (2016 2015), 919–988.

[9] Shanifa Nasser. Family of Toronto man who fell to his death says police locked them in room, didn’t let them speak to him | CBC News. *CBC News*. Retrieved February 23, 2023 from https://www.cbc.ca/news/canada/toronto/taresh-bobby-ramroop-fall-death-1.6630688

[10] Pamela Palmater. 2016. Shining Light on the Dark Places: Addressing Police Racism and Sexualized Violence against Indigenous Women and Girls in the National Inquiry. *Can. J. Women Law* 28, 2 (July 2016), 253–284. DOI:https://doi.org/10.3138/cjwl.28.2.253

[11] Private Member. Arrests and Strip Searches (RBDC-ARR-TBL-001). *Toronto Police Service Public Safety Portal*. Retrieved February 23, 2023 from https://data.torontopolice.on.ca/pages/open-data

[12] Nadine Yousif. 2022. Toronto’s first-ever mental health crisis response teams — without police — to launch in March. *thestar.com*. Retrieved February 23, 2023 from https://www.thestar.com/news/gta/2022/01/19/torontos-first-ever-mental-health-crisis-response-teams-without-police-to-launch-in-march.html

[13] 2020. Police kill Canadian man during mental health check. *BBC News*. Retrieved February 23, 2023 from https://www.bbc.com/news/world-us-canada-53141449

[14] Who polices the police? SIU under microscope after Regis Korchinski-Paquet report. *The Toronto Star*. Retrieved February 23, 2023 from https://www.thestar.com/podcasts/thismatters/2020/08/28/who-polices-the-police-siu-under-microscope-after-regis-korchinski-paquet-report.html

[15] Victim of assault by Calgary police officer dies before constable sentenced - Calgary | Globalnews.ca. *Global News*. Retrieved April 12, 2023 from https://globalnews.ca/news/7991917/calgary-dalia-kafi-dead/