Patterns and Motivations of Strip Searches:
An Exploration of Gender, Race, and Crime Typed biassed Policing in Toronto

Josua Lutian, Wenxuan Li

INF2178

16 April 2023

1. Introduction

Power and control have historically been points of tension in criminal justice systems throughout the world. Unfortunately, figures of authority have been known to misuse their station to enact power and command control over incarcerated individuals or potential prisoners. The infamous Stanford Prison Experiment by Philipp Zimbardo (1973), showed us that anyone who is given the tools and environment to fill the role of a figure of authority is capable of doing harm towards people. Although this experiment has been criticized for its disregard for research ethics, it remains as a powerful reflection of what is happening with our current justice system. Aside from assuming the role of an authority figure, a recent study showed that officers who experience higher levels of stereotype threat, moments where individuals are in situations that activates negative stereotypes about their group, also showed increased rates of support for use coercive policing (Trinker, Kerinson, & Goff, 2019). Seeing multiple excessive use of force cases portrayed in the media is now not so surprising knowing these mechanisms are playing in the background. In Ontario alone, an 18.2% increase in use of force cases was observed in a span of ten years, from 2010 to 2020 (Wortley et. al, 2021, p.19).

In 2001, Ian V. Golden, a black man, appealed his police assault charge to the Canadian Supreme Court, where he was strip searched three times throughout his arrest (R V. Golden, 2001). Although he was found guilty, his police assault charges were dropped in the light of the officers' excessive use of force. According to the Canadian Civil Liberties Association, Ontario's criminal system's usage of strip searches are dehumanizing and invasive in nature. In their recent constitutional challenge to the Supreme Court they describe strip searches as " unreasonable, overly broad, and grossly disproportionate to the purpose of the legislative scheme" (Canadian Civil Liberties Association v. Canada, 2022). It is not surprising to learn that officers have also found ways to circumvent the regulations that are meant to protect inmates from excessive use of strip searches by distorting the meaning of strip searches within their domain (Daems, 2014). What is more unjust is that, much like the R.V. Golden case, strip searches have been studied to be a site for racial profiling and in general, a tool for authority to enact discriminatory behaviour.

Past research showed strip searches are disproportionately used against racially marginalized groups, specifically Afro Caribbeans (Newburn, Shiner, & Hayman, 2004). In tandem with that, Newburn et. al (2004) explained that strip search percentage is much higher in crimes where there is perceived obfuscation of evidence (i.e. drug related crimes). This meant that Black individuals who committed crimes that were not typically violent in nature were more likely to be stripped and searched. On top of that, due to various factors like high unemployment rates and economic disadvantages, women are more likely to commit non-violent crimes like larceny, property, and fraud (Heimer, 2000). This could mean that women who identify as a racial minority might face the highest rates of strip searches compared to the rest of the population who have committed crimes. Moreover, in a Canadian qualitative paper, where the researchers explored the experience of female inmates, it was concluded that strip searches could qualify as sexual assault as the experience of it takes away a womans's agency to say no due to "due to power imbalances and fear of serious consequences." (Hutchinson, 2020).

Knowing that crime, gender, and race-based discrimination could potentially affect strip search rates, we want to explore and uncover the trends that happen here in Toronto. To explore this, we aim to answer these research questions using ANOVAs, ANCOVAs, and Logistic Regressions.

RQ1: Does perceived race influence strip search rates in Toronto?
RQ2: Does crime type influence strip search rates in Toronto?
RQ3: Does gender influence strip search rates in Toronto?
RQ4: Do perceived race, crime type, and gender of arrested individuals influence strip search rates in Toronto?
RQ5: Can the race, crime type, and gender of arrested individuals predict strip search occurrence?

To do so, we explored the Toronto Police Arrests and Strip Searches Database, which documents strip searches throughout the greater Toronto Area from 2020 - 2021. Our results showed that there is a significant difference in strip search rates between race groups, especially between Asian and Indigenous individuals. Although no significant difference in strip search proportions were found between crime type groups, it was found to be a significant covariate when paired with race. Finally, there were significant differences in strip searches between the two gender groups. Gender was also the only predictor that significantly impacted the prediction model for strip search occurrence, indicating that men are 3 times more likely to experience strip searches compared to women.

2. EDA

2.1 Data Cleaning

The first thing we did was to pick variables that would help us explore our research questions. We isolated the 'Arrest_Year' ,'Perceived_Race', 'Sex','Age_group__at_arrest_', 'Occurrence_Category', and 'StripSearch' variables, which we renamed to Year, Race, Sex, Age, Crime_Type, and Strip for easier reference. We used the column name "Sex" because that is what the original dataset has. However, in this paper we will be referring to it as Gender. Besides variables that are directly included in the research questions, we also include age to see some demographic differences that could provide nuance to our analysis if needed.

We aggregated levels within our predictor variables Race and Crime_Type. In the original data set Race has 8 levels and Crime_Type has 30 levels. As a public safety record, such a detailed record is a good thing; however, in this report, we simplified these two variables. For the race, we combine different levels into White, Black, Asian, Latino, Indigenous, and Unknown. This is quite conventional in crime and race research especially the ones who were largely investigating the dichotomy between White and Black groups. For the crime type, we were interested in exploring Violent and Non-Violent crimes. So we followed a paper that explored crime types (Pratt & Cullen, 2005) in three levels: Property crime, Violent crime, and Other crime. The specific classification of different levels into these categories follows (NIJ, 2023). We ended up aggregating Property and Other crime into a single category called “Non-Violent”.

The last step we did is to find continuous variables that can be used as the main outcome variable for the ANOVAs and ANCOVAs and as a covariate and for our ANCOVAs. In the beginning, we thought that the counts of strip searches and violent crimes could be good variables. However, we soon realised that if the total number of strip searches and violent crimes are not controlled. Using raw counts for between group comparisons would not be fair. Therefore, we choose to calculate the proportion of strip searches as the dependent variable, which is defined as :

$$\sum_{count} of\ strip\ searches\ (within\ each\ predictor\ levels)\ \div \sum_{total\ count} of\ arrests$$

We did the same to calculate the proportion of violent crimes:

$$\sum_{count} of\ violent\ crimes\ (within\ each\ predictor\ levels)\ \div \sum_{total\ count} of\ arrests$$

To create these new variables, we grouped the original dataset by the variables we picked in the first step, except Strip. Then we calculated the count of strip searches (entries that had a ‘1’ in the Strip column) and total arrests for each combination of levels. We did the same for Violent Crimes. Finally, we divided these values to compute the proportion of strip searches and proportion of violent crimes per group. We thought that this was a fairer comparison and reflection of how much strip searches occur within a specific group rather than relying on the count variable as it is directly affected by the magnitude of the populations or sample size of each level within the Race and Gender categories. Lastly, we took out any rows that had an NA or a U (Gender is Undefined) in the “Sex” Column leaving us with 324 rows of data representing 61751 viable arrest cases.

2.2 Descriptive Statistics

2.2.1 Gender

We first wanted to look at the different patterns of arrests when looking at the Gender category. According to Table 1, there are more male arrests in Toronto in 2020 and 2021 combined, comprising 81% of total arrests, compared to female arrests, which make up 19%.

Table 1. *Gender differences in Strip Searches*

Gender	Mean	StDev	Strip Searches	Total Arrest
Female	0.08	0.13	1283	12575
Male	0.11	0.12	6516	52499

This pattern is reflected in strip searches as well. In Figure 1 below, when controlling for race, crime type, and age, the average proportion of strip searches within female groups are 8%, while male groups average at 10%. This does not follow the findings and logic of previous research that expected women to be subjected to more strip searches.

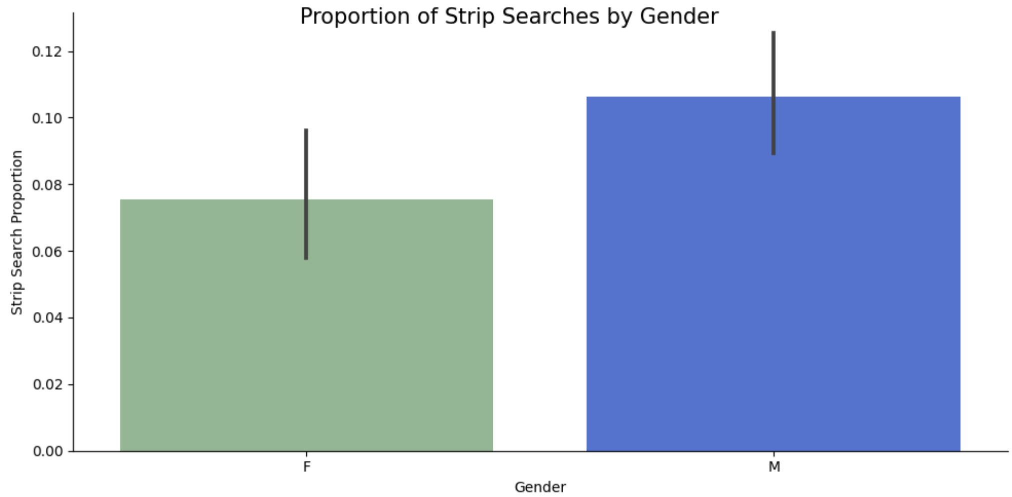


Figure 1. *Proportion of Strip Searches grouped by Gender*

Another surprising pattern that we found was when looking at the violent crimes grouped by gender, the average proportion of violent crimes in female groups is much higher, 34%, than that of male groups, 3%, (Table 2). This links to the claim that individuals who commit nonviolent crimes tend to be strip searched more than individuals who commit violent crimes.

Table 2. *Gender differences in Violent Crimes*

Gender	Mean	StDev	Violent Crimes
Female	0.34	1.1	217
Male	0.03	0.23	129

The only difference is that, unlike in past research, our data showed that the male groups, in proportion to all male crimes, commit less violent ones and thus were strip searched more. You can see this pattern in the bar chart below (Figure 2).

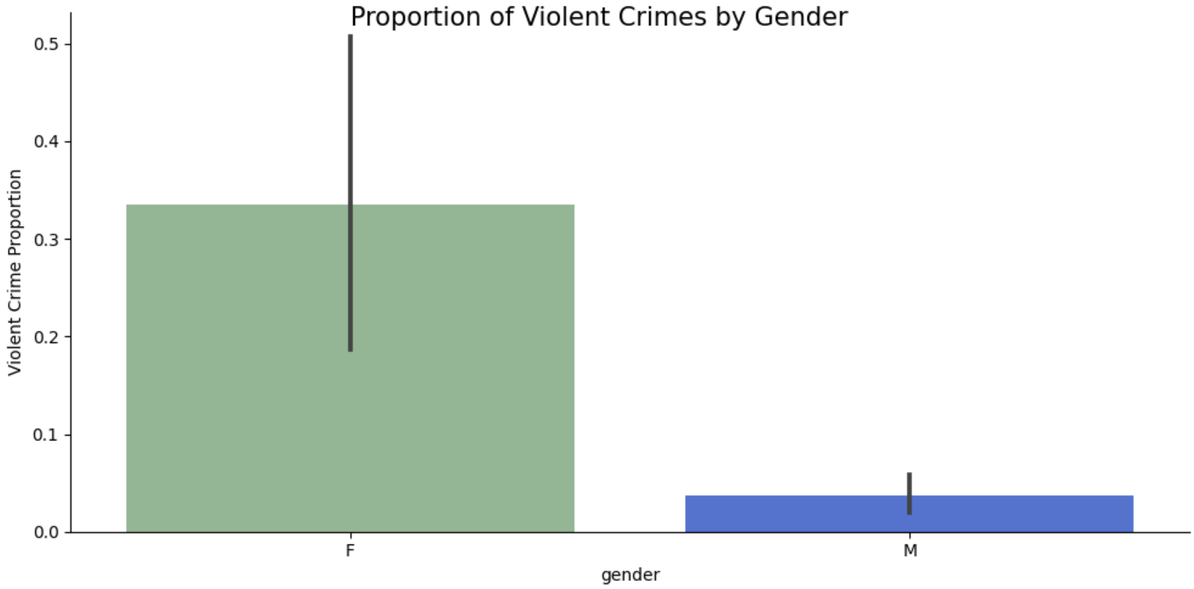


Figure 2. *Proportion of Violent Crimes grouped by Gender*

2.2.2 Race

We also wanted to see the general trends of Race within the context of strip searches. You can see in Table 3 that looking at the Strip Searches column, which uses the raw counts of strip searches, there is a disproportionate amount of White individuals represented. They also happen to be the largest population represented in the dataset.

Table 3. *Summary of Descriptive Statistics for Proportion of Strip Searches grouped by Race*

Race	Mean	StDev	Strip Searches	Total Arrest
Asian	0.06	0.07	826	11232
Black	0.10	0.11	2434	17487
Indigenous	0.13	0.18	306	1926
Latino	0.08	0.11	132	1759
Unknown	0.08	0.11	535	5044
White	0.10	0.11	3566	27635

So we turned to the proportion of strip searches by race for a more balanced way of representing the spread of strip searches within racial groups in Toronto.

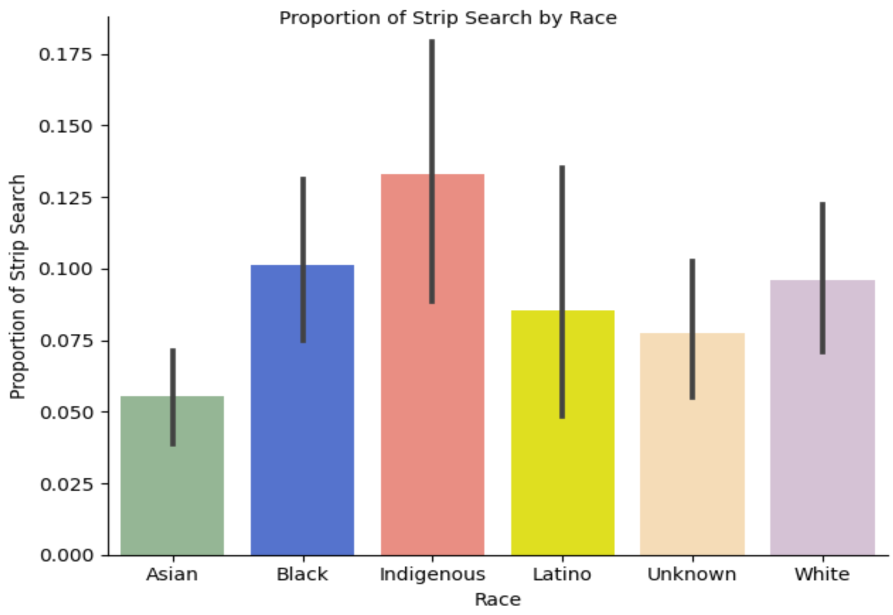


Figure 3. *Proportion of Strip Searches grouped by Race*

Figure 3 shows that Indigenous people, who account for the second smallest racial group for all arrests (3%), have the highest proportion of strip searches p = 306/1962 (0.16) among all races. While the Asian group, which is 17%, have the lowest proportion of strip searches p = 826/11232 (0.07).

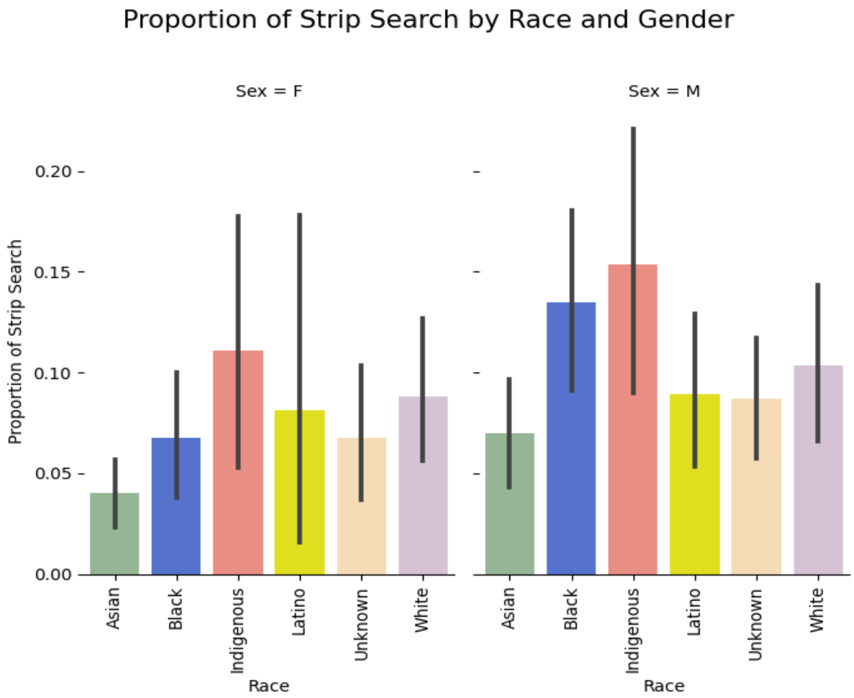


Figure 4. *Proportion of Strip Searches grouped by Race and Gender*

This overall pattern is also reflected when looking at strip searches grouped by both race and gender. Within both gender groups, male and female, the indigenous group represented the highest average proportion of strip searches while the Asian groups represented the lowest. As seen in previous visualisations, male proportions of strip searches are systematically higher than the female ones.

2.2.3 Crime Type

Figure 5 below represents the strip search counts across the pre-processed crime types. We can see that the top two crimes with the most strip searches are Assault-based (Violent crime). The crimes that yielded the least strip searches were Homicide (Violent) and Crimes against Children (Nonviolent).

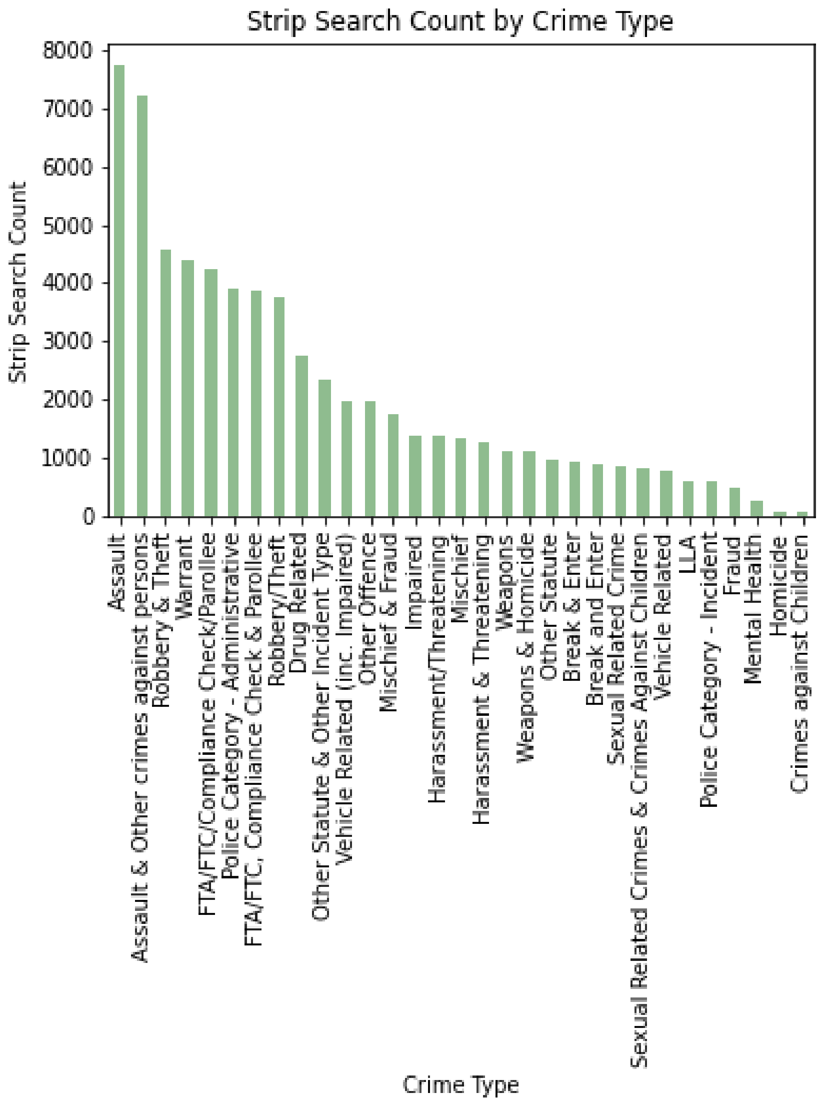


Figure 5. *Strip Search counts groups by original crime type*

We also plotted Crime_Type by proportion of strip searches (Figure 6) and found that there is a higher proportion of strip searches for violent crimes compared to nonviolent crimes. This is further supported with the calculated means (Table 4) showing the same pattern, where the average proportion of strip searches related to violent crimes is 10 %, while nonviolent crimes have an average proportion of strip searches of 9% .

Table 4. *Crime type differences in Strip Searches*

Crime Type	Mean	StDev	Strip Searches	Total Arrest
Violent	0.10	0.14	2684	23528
Nonviolent	0.09	0.11	5115	41546

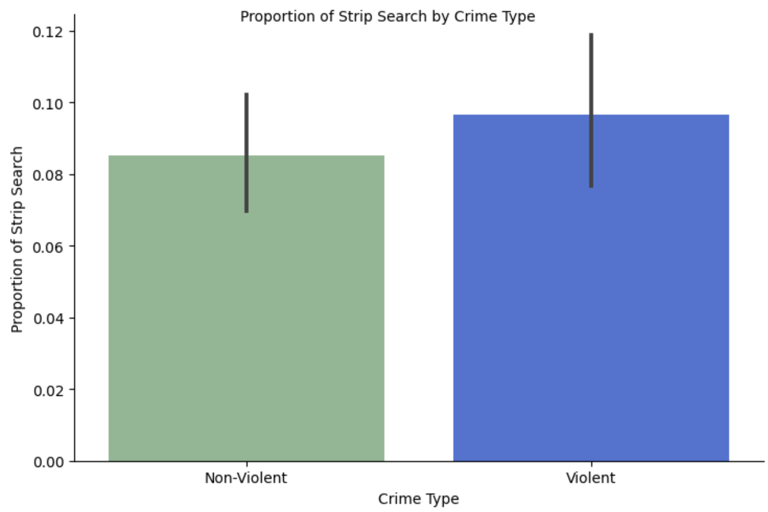


Figure 6. *Proportion of Strip Search grouped by crime type*

This is not what we expected due to some claims from previous research, where non-violent crimes were shown to have higher percentages of strip searches. We concur that there must be some other factor that contributes to higher strip search proportions in male groups.

Since we are investigating the relationship between the predictors, we also graphed the proportion of strip searches, race, and crime type together. Much like figures 4 and 3, Figure 7 shows that regardless of crime type, indigenous people tend to have higher proportions of strip searches compared to other race groups. The Asian group was also found to experience the lowest average proportion of strip searches, regardless of crime type.

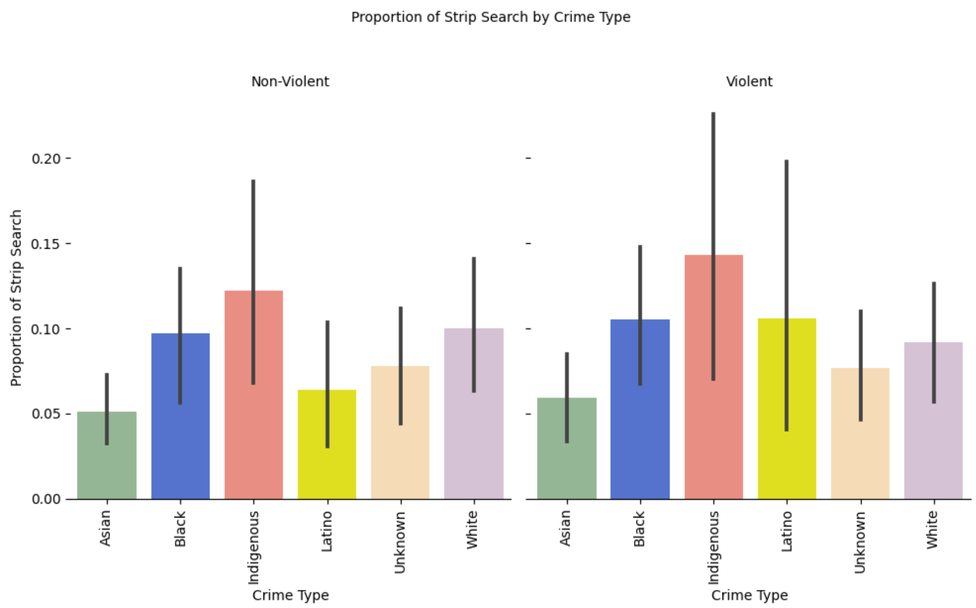


Figure 7. *Proportion of strip search grouped by crime type and race*

Again, a very similar pattern arises when crime type, in the form of a proportion of violent crimes, was plotted with race and gender (Figure 7). The biggest difference is the overwhelming difference between the proportion of violent crimes within the female group compared to the male group. The long error bars also suggest that there is so much more deviation from the mean in the female group. This drastic graph, can suggest that there might be a significant difference in proportion of violent crimes between gender groups.

2.2.4 Power Analysis

Before we tested if the proportion of strip searches differed significantly between the different levels of our three predictors: Sex, Race, and Crime_Type, we conducted a power analysis to see if our sample size is suitable given our parameters. Using the Sex predictor, which has two levels, our calculated effect size (Cohen's D) was 0.25 or a medium effect size.

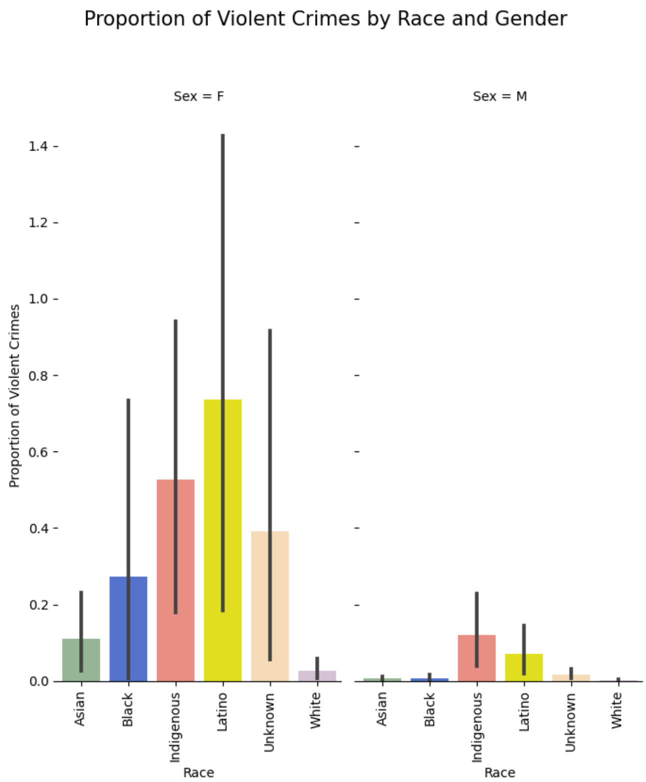


Figure 8. *Proportion of violent crimes grouped by race, and gender*

Table 5. *Power Analysis Sample Sizes*

Gender	Sample Size Needed	Actual Sample Size
Female	258	163
Male	265	167

After calculating the effect size, we also determined the suggested sample size by setting our alpha level = 0.05 with a statistical power of 80%. The calculations indicated that for the Female Group a sample size of 258 was needed, while a sample size of 265 was needed for the Male Group. Our actual sample sizes, which can be seen in Table 5, were 163 and 167 respectively. This means that each group does not have the minimum sample size to produce reliable results according to the parameters set by the power analysis. However, we will continue with the analyses and keep this in mind as a limitation. The blue marker in the figure below represents our data's statistical power with a sample size of 167, Cohen's D of 0.25, and an alpha level of 0.05. This does not fit into the power analysis. So we calculated the power that we need keeping our alpha level and effect size constant, while using 258 as our sample size. The red marker represents this calculation, which indicated that with our effect size and alpha level constant, we would need a power of 0.81.

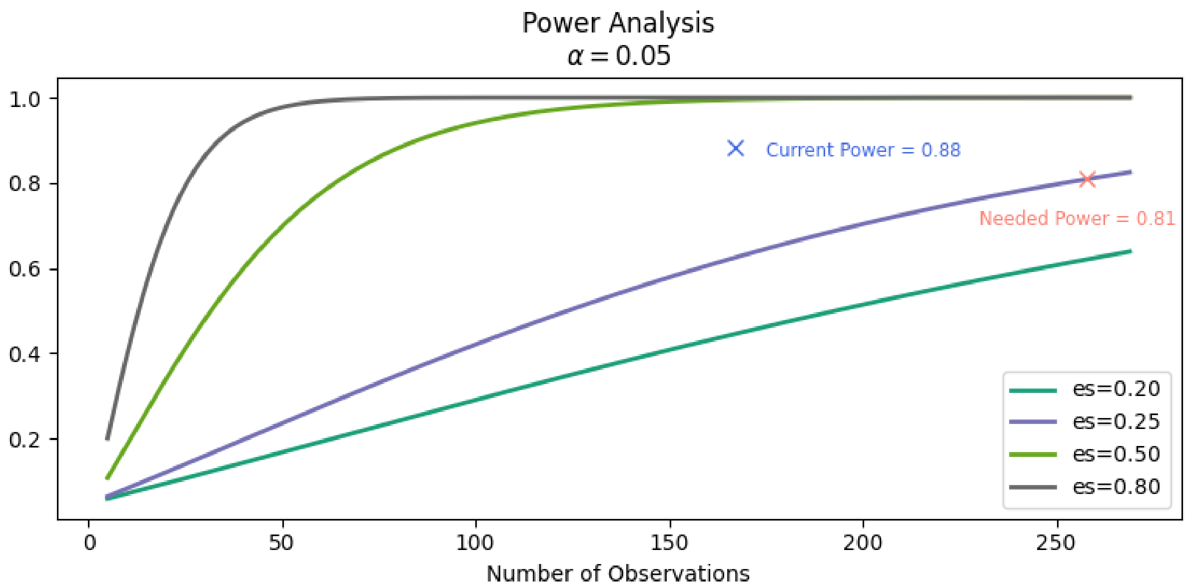


Figure 9. *Power Analysis Plot*

2.3 T-Tests

From the EDA, we saw that there are differences in mean proportions of strip searches within the different predictors of Gender, Race, and Crime_Type. The visualisations show that the Indigenous group had the highest proportion of strip searches amongst all the racial groups, while the Asian group had the lower average proportion of strip searches. Additionally, female groups had a higher proportion of strip searches compared to the male groups. And lastly, nonviolent crimes had a lower proportion of strip searches than violent crimes. We wanted to investigate whether these differences are significant. To examine that, we conducted 3 one-tailed two sample t-tests with Bonferonni corrections between the different levels of gender, race, and crime type.

2.3.1 Gender

We conducted the t-test to check if gender influences the proportion of strip searches. Our hypotheses were as follows:

H0 (null hypothesis): The mean proportion of strip searches of Female and Male groups is equal.
HA (alternative hypothesis): The mean proportion of strip searches of Female and Male groups is not equal.

The t-test showed that there was a significant difference in proportion of strip searches between Male ($M = 0.08$, $SD = 0.13$) and Female groups ($M =0.11$, $SD = 0.12$), at p-value of $\alpha = 0.03$. This means that the mean proportion of strip searches between the male and female groups were significantly different. We were able to reject the null hypothesis.

2.3.2 Race t-tests

For the Race predictor we wanted to understand race influences the proportion of strip searches. The hypotheses we tested were:

H0 (null hypothesis): The mean proportion of strip searches of the six racial groups, Asian, Black, Indigenous, Latino, Unknown, and White, is equal.
HA (alternative hypothesis): The mean proportion of strip searches of the six racial groups, Asian, Black, Indigenous, Latino, Unknown, and White, are not equal.

In the initial t-test, the results indicated that there were 4 comparisons that produced significant differences in mean proportions of strip searches. The proportion of strip searches in the Asian group ($M = 0.06$, $SD = 0.06$) were significantly lower compared to the proportion of strip searches of the Indigenous ($M= 0.13$, $SD = 0.18$), $p = 0.003$, Black ($M= 0.10$, $SD= 0.11$), $p = 0.01$, and White ($M = 0.10$, $S = 0.10$), $p = 0.01$ groups. Additionally, the Indigenous group's ($M = 0.13$, $SD = 0.18$) mean proportion of strip searches were significantly higher than the Unknown ($M = 0.08$, $SD = 0.09$), $p = 0.04$ groups. However, since we did 15 t-tests we also corrected the p-value using the Bonferroni correction. After this correction, with an adjusted alpha level of .0003 per test (0.04/15) none of the comparisons yielded a significant difference. Thus we cannot reject the null hypothesis.

2.3.2 Crime Type t-tests

Lastly, we conducted t-tests to find if there are significant differences between mean proportion of strip searches between the two crime types: Violent and Nonviolent.

H0 (null hypothesis): The mean proportion of strip searches of the three crime types, Other, Property, and Violent, are equal.
HA (alternative hypothesis): The mean proportion of strip searches of the three crime types, Other, Property, and Violent, are equal.

After calculating each group's mean proportions of strip searches, we found that there were no significant results. The proportion of strip searches in Violent (M = 0.10, SD = 0.14) was not significantly different from the Nonviolent groups (M = 0.09, SD = 0.11), only at a p-value of $\alpha = 0.41$. Thus we were not able to reject the null hypothesis.

3. Methods

3.1 Data Description

The dataset we analysed for this report was the *Arrests and Strip Searches (RBDC-ARR-TBL-001),* taken from the Toronto Police Service, Public Safety Data Portal (*Arrests and Strip Searches,* 2021). It includes all recorded information related to all arrests and strip searches that officers from Toronto Police Service conducted from 2020 to 2021 (arrests that are not in Toronto are identified with an 'xx' entry in the 'ArrestLocDiv' variable). It contains 65276 rows and 25 columns. The variables represented are year, month, IDs (i.e., event ID, arrest ID, person ID, and information ID), race, sex, age group, youth, location, strip search, booked, crime type, etc. The dataset also included observational data of strip search reasons and actions in arrest coded as binary categories. Aside from ID and Time variables, all of the data were categorical in nature.

3.2 Analytic Approach

In order to answer our four research questions and investigate the patterns we found through our exploratory data analysis, we first performed a power analysis to see if our data had enough sample sizes per group for us to test and conclude from our analyses with an established statistical power of 80%, an alpha level of $\alpha = 0.05$, and a calculated effect size of 0.25.

We then performed multiple one-way ANOVAs to check for any significant differences between the average proportion of strip searches (outcome variable) and the levels within Gender, Race, and Crime Type. After that, we ran Tukey's HSD post-hoc tests, when our main ANOVA test produced significant results, to check which specific levels were significantly different from other levels in a given categorical variable.

We then conducted multiple two-way ANOVAs looking to see if different pair-wise comparison of the predictors produced significant interactions that influenced the proportion of strip searches together. Regardless of significance, we plotted these in some interaction plots to view the relationships between the three predictors.

To gain a more nuanced understanding of our questions, we performed two ANCOVAs, to see the impact of Race and Gender on the proportion of strip search, when controlling for the proportion of Violent Crimes. Lastly, we created a Logistic Regression model using the three predictors: Gender, Race, Crime Type, to see how much they all influence the outcome of being strip searched or not. We divided the grouped dataset into test and training sets, which created and fitted a model to predict strip search occurrence.

4. Results

4.1 ANOVA

4.1.1 One-way ANOVAs

Since we had significant t-test results for the Gender and Race predictors, we also wanted to perform one-way ANOVAs and the Tukey's post-hoc test to see if we can really reject the null hypothesis for these two.

The two one-way analysis of variance for Gender (F = 5.0, p-value = 0.03) and Race (F = 2.4, p-value = 0.04) both resulted in significant overall difference. Thus, we performed Tukey's post-hoc test to compare the differences between the levels of the predictors. The post-hoc test for Race showed that, among all fifteen comparisons, only the one between the Asian and Indigenous group indicated significant results (Mean Difference = 0.08, p-value = 0.02). Similarly, the Tukey's test result for the Gender predictor also indicated a significant difference between the average proportion of strip searches between the Male and Female groups (Mean Difference = 0.03, p-value = 0.03). This means that for both, one-way ANOVAs we are able to reject the null hypothesis.

4.1.2 Two-way ANOVA

The past sections looked at Gender, Race, and Crime Type as individual predictors to the proportion of strip searches. To explore if the three have a significant interaction or are influencing the proportion of strip searches together, we performed three two-way ANOVAs checking the influence of each pair of predictors: Race & Crime Type, Race & Gender, and Gender & Crime Type.

In congruence with the one-way ANOVAs, in all three two-way ANOVAs Race (p-value = 0.04) and Gender (p-value = 0.03) produced significant main effects indicating that they both influence the proportion of strip searches. However, none of the three interactions produced significant results (Appendix I).

Table 6. *Results of Two-way ANOVAs Interactions*

Interaction	p-value
Race & Crime Type	0.92
Race & Gender	0.83
Gender & Crime Type	0.69

4.2 ANCOVA

Following the ANOVAs we also conducted two ANCOVAs to answer the question of if these predictors are influencing the proportion of strip search while controlling for Crime Type. We transformed Crime Type into a continuous variable called proportion of violent crimes and used that as the covariate for both ANCOVAs.

The first ANCOVA was conducted to determine whether there is a significant difference in mean strip search proportion between the different race groups while keeping violent crime proportion constant. It was found that there was significant mean difference in strip search proportion across different levels of race ($F = 2.67$, $p = 0.02$). The proportion of violent crimes was also found to be significantly related to the proportion of strip searches ($F = 5.90$, $p = 0.02$).

Table 7. *ANCOVA Results for Race and Proportions of Violent Crimes*

Source	DF	F	p-value
Race	5	2.67	0.02
Violent_prop	1	5.90	0.02
Residual	323	NaN	NaN

The second ANCOVA was conducted to determine whether there is a significant difference in mean strip search proportion between the male and the female groups, while keeping violent crime proportion constant. According to Table 8, there were no significant differences found for gender, while controlling for proportion of violence.

Table 8. *ANCOVA Results for Gender and Proportions of Violent Crimes*

Source	DF	F	p-value
Gender	1	3.44	0.06
Violent_prop	1	3.00	0.08

Residual	327	NaN	NaN

4.3 Logistic Regression

Lastly, we conducted a logistic regression to explore the relationship between all three predictors: gender, race, and crime type on the odds of strip searches occuring.

According to Table 9, only the gender predictor showed significant results, with a pvalue of less than = 0.05. Our results showed that while controlling for other predictors: race and crime type, every time an the arrest individual is a male or perceived as male, they have a 3.12 times increase in odds of being strip searched than women (SexCode is a binary categorical variable with 0 = Female, and 1= Male). The rest of the predictors: race and crime type, did not result in a significant p-value.

Table 9. Odds Ratios

Source	**Coef.**	**SE**	**p-value**	**Odds Ratio**
Intercept	0.07	0.30	0.80	1.08
RaceCode	0.08	0.08	0.32	1.08
SexCode	1.14	0.28	< 0.05	3.12
CrimeCode	-0.21	0.27	0.45	0.81

The created logistic regression model has 72% accuracy, 73% precision, 98% recall. Overall, our model was able to correctly predict the occurrence of strip searches about 3 out of 4 times for the whole test set. You can see in the confusion matrix in Figure 10, that the model's prediction was mostly true positives (47) but also had quite a number of false negatives (17).

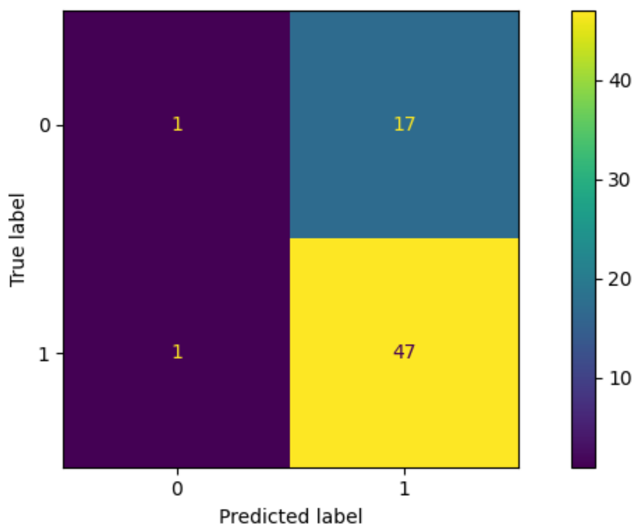


Figure 10. Confusion matrix for predicted and true values

Figures 11 through 13 represent the plotted confidence intervals of the observed strip search odds and prediction intervals for our models predicted odds. You can see that because of our predictors categorical nature, clusters of points can be seen. When we assessed our model, it showed that 98% of our predicted values fall inside our prediction intervals across all predictions.

For all the predictors, the confidence intervals show that in general 95% of the average observed values fall between 0.55 and 0.78. This is different from the prediction intervals, which indicate that 95% of the time we could expect our model's estimates to fall between -0.24 and 1.57.

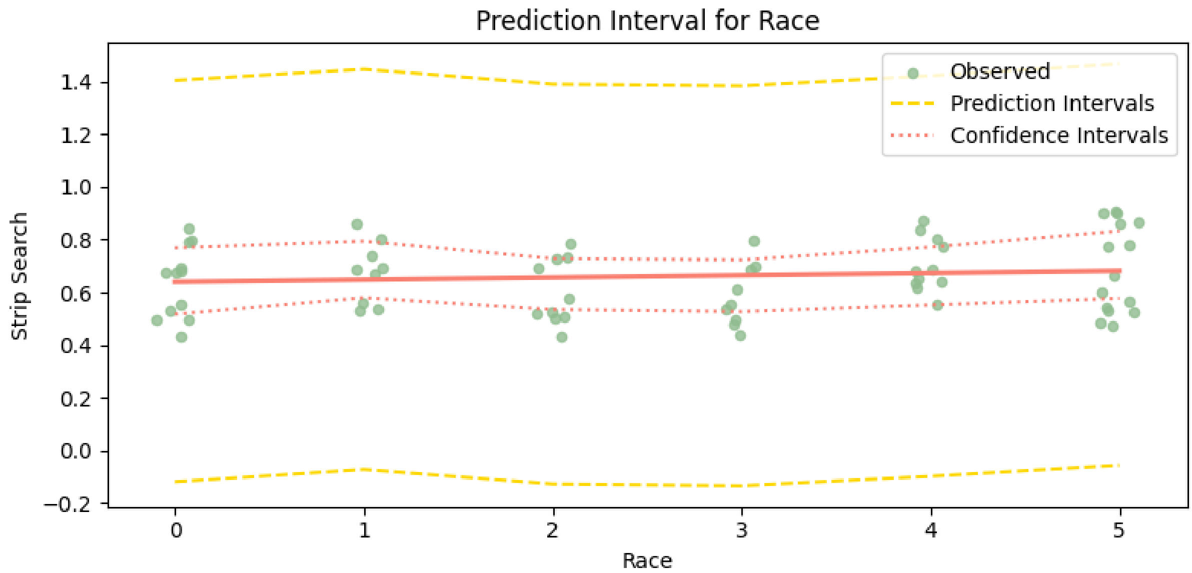


Figure 11. Prediction Interval for Race

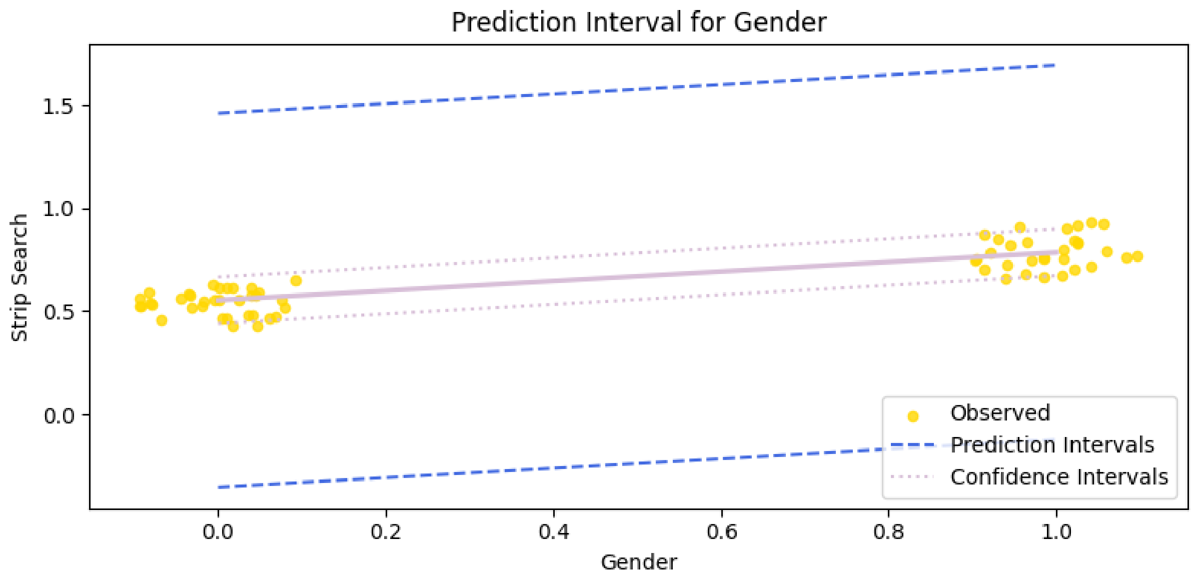


Figure 12. Prediction Interval for Gender

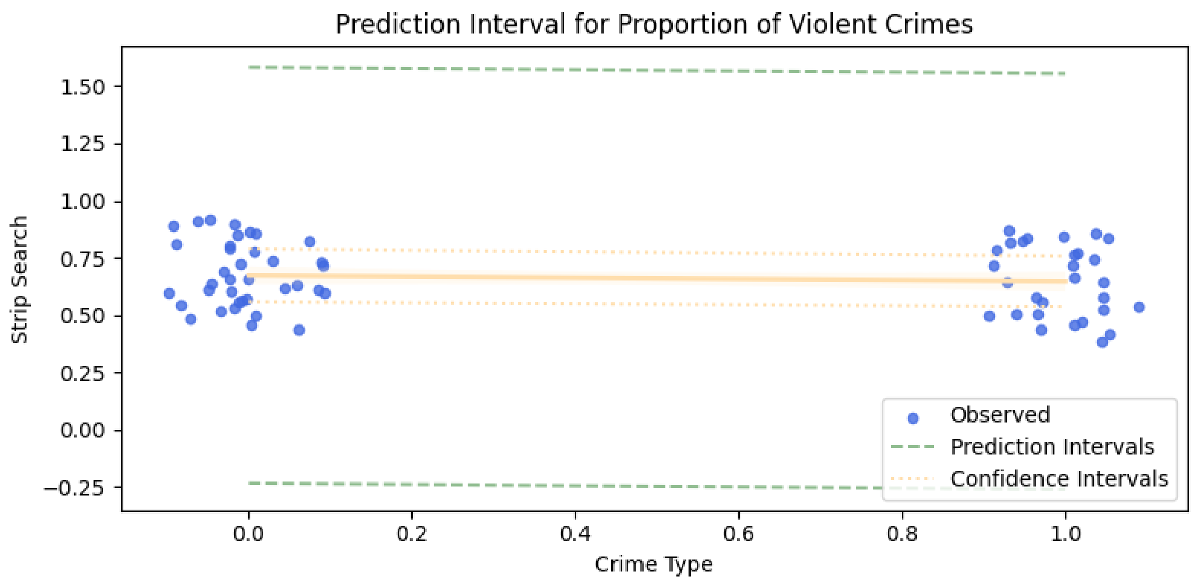


Figure 13. Prediction Interval for Violent Crimes

5. Discussion

Our analysis of variances show that race does impact the proportion of strip searches in Toronto, which indicates potential discrimination by the police showing different treatment of strip searches on different racial groups. This supports previous research that found that police tend to target specific race groups during arrests (Newburn, Shiner, & Hayman, 2004). Unlike the findings of previous research where the Black people tend to be targeted by the authorities, our post hoc test indicates that the major difference in proportion of strip searches is between Asian and Indigenous groups, where Indigenous people tend to be strip searched more. Other groups faced similar rates of strip searches regardless of their race.

However, our ANOVA results showed that crime type does not significantly influence the proportion of strip searches. This is in contrast with previous research explaining that people who commit Property crimes (non-violent) usually experience more strip searches compared to Violent crimes because the police suspect that the individual is concealing evidence (Newburn, Shiner, & Hayman, 2004). This is perhaps due to our interpretation of what nonviolent crimes could be. In categorising crimes in the dataset, we consulted the National Institute of Justice (NIJ, 2023), where drug-related crimes were not categorised under the

Property crime, contrasting from the study mentioned above. Drug crimes represented the 9th highest strip searched crime type in the dataset (Figure 4).

Interestingly, our last ANOVA test indicated that gender has a significant influence on the proportion of strip searches, where men experience strip searches more than women. This is against our inference based on past research, which states that women would be subjected to more strip searches because they tend to commit more non-violent crimes (Heimer, 2000) and individuals who committed nonviolent crimes are more likely to be strip searched (Newburn, Shiner, & Hayman, 2004). Our data showed the opposite: 1) men were subjected to more strip searches, 2) individuals who committed violent crimes were stripped searched more, and 3) violent crimes did not have a significant effect on strip searches. So there must be something else that creates this pattern of gendered strip searches in Toronto.

Although we suspected that there could be patterns after seeing some in our EDA (Figure 8), our two-way ANOVAs that compared the three predictors pair-wise between, indicated that there is no significant interaction between race and crime type, race and gender, and gender and crime type.

We then performed two ANCOVAs for gender and race to see if they impact the proportion of strip searches, using the proportion of violent crimes as a covariate. As expected from our two-way ANOVAs, there was no significant effect when testing for gender whilst controlling the proportion of violent crimes. This indicates that there is no difference in mean proportion of strip searches across the two gender groups, whilst keeping crime type controlled. This further supports the ANOVA findings where crime type was found to not be the mechanism that creates a gender bias in strip searches.

When we created an ANCOVA model for race as the factor and proportion of violent crimes as the covariate, it showed that there was a significant impact of race and crime type on the proportion of strip searches. This supports findings from previous research that found that race and crime type both influence strip searches (Newburn, Shiner, & Hayman, 2004). This also reflects our previous ANOVA results where race was found to significantly impact strip searches both on its own and with the two other predictors. Surprisingly, the proportion of violent crimes, which is a transformed version of crime type (from categorical to continuous), also showed a significant relationship with the proportion of strip searches when it was a covariate for race. This would mean that even when you control for crime type, there is still a significant difference in proportion of strip searches among the race groups and that crime type is an important predictor for strip searches as well.

Lastly, we conducted a logistic regression to create a model using gender, race, and crime type as categorical predictors, to predict the occurrence of strip searches. Gender was the only predictor that significantly impacted the model. It could be inferred that male groups have a 3.11 times higher odds than the female group to be stripped searched. Our model had an accuracy of 72% accuracy, 73% precision, and a 98% recall. These performance metrics showed that the model does a good job in predicting strip search occurance in Toronto, with its accuracy and precision to be on par with industry standards. However, the model also produced 25% of false negative predictions. In context, if this model were to be used to decide whether an individual were to be strip searched, 25% of the sample population would be wrongly strip searched. Furthermore, with gender being the only significant predictor to this model, the 98% of arrested individuals that would be put into this model, who are found to be “true positives” might be stripped searched based on gender alone. This is quite a drastic difference from 14%, which is the total percentage of strip searched individuals in Toronto based on the dataset.

All in all, this implies that being a man or perceived as a man, over all other predictors, significantly raises the odds of an individual to be strip searched in an arrest situation in Toronto. Tying that into our other results, we can see that the difference in average proportion of strip searches within groups is also influenced by that group’s race and crime type. Even though both of these predictors, according to our model, does not significantly predict being stripped searched or not, the patterns of strip searches indicate that these factors are still used by figures of authority when deciding to strip search someone. This puts Indigenous men, regardless of their crime type, more susceptible to strip searches, which indicates that there is some racial profiling, gender-based, and crime-based discrimination in the current strip search operations of the Toronto Police Service.

5.1 Limitations:

All the data in the dataset were originally categorical in nature and had to be transformed for our analysis. This is a large limitation as it left us to make a lot of assumptions in how we wanted to represent the data in our research. The people who collected and used this data, the

Toronto Police, might also have a different data schema than us or the other researchers in literature. Additionally, fitting categorical data into analyses like ANOVAs and t-tests, which are for continuous data, is not the conventional way to do these investigations. A dataset like this could possibly yield more accurate results when run through a Chi-Square test, where two categorical predictors are compared. This was however, less important when doing ANCOVAs and Logistic Regressions as categorical variables were allowed.

Another limitation with our data is that our sample data did not fit a normal distribution. After running a Shapiro-Wilks test, for all levels of the three predictors, the p-values were less than 0.05. This meant that we were able to reject the null hypothesis for the test indicating that our data is not normalised. This could have affected our ANOVA results and could be the reason why our residuals errors were quite high.

The binary categories we used: violent and nonviolent, strip searched or not, and gender: male & female; leads to a lot of 0-valued numbers. So in trying to create a continuous variable like Strip_prop and Violent_prop from these created a few different limitations. For example, taking into account our outcome variable, strip search. The value of the proportion directly connected to the count of strip searches within the specific group. If there were no strip searches, which was often, then the calculated proportion would be 0. This is the same when we created the proportion of violent crimes as a continuous variable for our ANCOVAs. What this results in is clustering in your data and a general lower reliability of your models. You can actually see this quite vividly in our prediction interval graphs, where the clusters were visualised. Thus, we cannot trust the prediction intervals we calculated since our data is clustered in a way that most of our data points are not even within the limits of our intervals.

Additionally, the fields in the dataset can also be misleading. The Race predictor was originally called “Percieved_Race". This means that the officer who collected the information about the individual assumed the person’s race upon arrest. This was not a self report, which could have been more accurate. Additionally, individual “Person_IDs” fields are captured multiple times within the dataset, possibly as a repeat arrest for the same person. Unfortunately, unique “Person_IDs” do not all have the same value in their ‘Percieved_Race’ field. We interpreted this in two ways: 1) the “Person_ID” was used twice for two different people or 2) the officers who noted the arrest were not able to assume the same race.

Lastly, one large limitation is that our grouped data did not pass the power analysis we performed. This means that we cannot confidently make claims from our results. We would need to increase our power by increasing our sample sizes or even finding a more suitable dataset to answer our research questions. In the same logic, the power test we used was meant for t-tests and indirectly, ANOVAs. We used Cohen D as our effect size and that is only useful for t-tests and calculating Cohen's F for ANOVAs (Cohen's D/2). Next time, we need to perform more robust power testing for all the types of analyses we envision using to make sure that we can be confident that our research has reached our established statistical power, alpha level, effect size, and sample size.

6. Conclusion

In this report, we focused on a dataset retrieved from the Toronto police service on the arrest information between 2020 to 2021. Based on the dataset and a literature review, we tried to answer four research questions:

RQ1: Does perceived race influence strip search rates in Toronto?
RQ2: Does crime type influence strip search rates in Toronto?
RQ3: Does gender influence strip search rates in Toronto?
RQ4: Do perceived race, crime type, and gender of arrested individuals influence strip search rates in Toronto?

After our EDA and statistical analyses, we uncovered that in Toronto gender, race, and crime type all influence the proportion of strip searches in quite nuanced ways. Although there are no significant interactions between the three predictors, further analyses showed that race as an individual factor, is a strong predictor for strip searches, regardless of crime type. Crime type, although it does not have a significant direct influence on strip searches, plays a role as a covariate that strongly impacts strip searches. Lastly, gender on its own is a strong predictor for proportion of strip searches and the only feature that significantly impacts the model that predicts strip search occurrence.

Moving forward with these findings, we can explore the dataset's other variables that indicate reasons for strip searches and actions during arrest. This can give a more nuanced picture of the motivations of officers and the reactions of people who were arrested in Toronto, adding to a greater understanding of the arrest protocols in Toronto. Since there was no interaction found between the three predictors, which can be inferred from literature, we need to continue to investigate other factors to produce a better model. We can look at other well researched predictors like Year, and Age, as they influence Booking and Arrest rates.

Our research only sheds some light on a limited cross-section of the Toronto Police Service operations, however we were able to unveil that gender, race and crime type affected strip searches in Toronto. This indicates that racial profiling, gender based, and crime type based discrimination could potentially be happening in Toronto, especially amongst Indigenous men, regardless of crime. It is important for us to further examine our authorities' biases through studies like this in order to create policies driven by equity and justice, rather than prejudice.

7. References

Arrests and Strip Searches (RBDC-ARR-TBL-001). (n.d.). Retrieved February 27, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Daems, T. (2014). ‘ *Ceci N’est Pas Une Fouille À Corps* ’: The Denial of Strip Searches in Belgian Prisons. In M. Deflem (Ed.), *Sociology of Crime, Law and Deviance* (Vol. 19, pp. 75–94). Emerald Group Publishing Limited. https://doi.org/10.1108/S1521-613620140000019003

Heimer, K. (2000). Changes in the Gender Gap in Crime and Women’s Economic Marginalization. *CONTINUITY AND CHANGE*, *1*.

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

Newburn, T., Shiner, M., & Hayman, S. (2004). RACE, CRIME AND INJUSTICE? Strip Search and the Treatment of Suspects in Custody. *The British Journal of Criminology*, *44*(5), 677–694.

Pratt, T. C., & Cullen, F. T. (2005). Assessing Macro-Level Predictors and Theories of Crime: A Meta-Analysis. *Crime and Justice*, *32*, 373–450.

Property Crimes. (n.d.). National Institute of Justice. Retrieved February 27, 2023, from https://nij.ojp.gov/topics/crime/property-crimes

R. v. Golden—SCC Cases. (n.d.). Retrieved February 27, 2023, from https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/1924/index.do

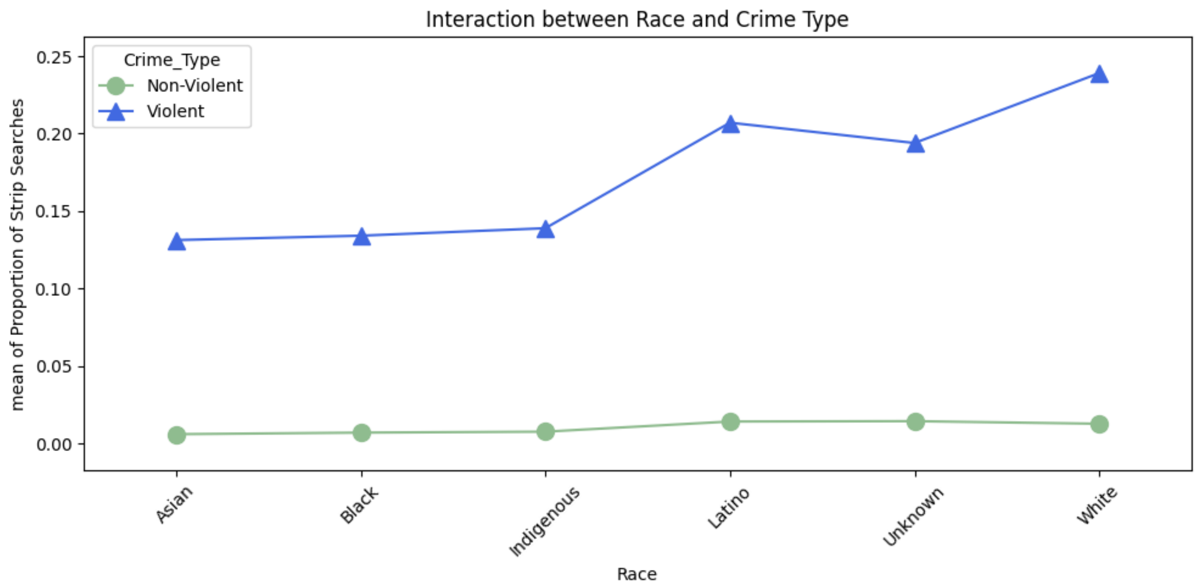
Trinkner, R., Kerrison, E. M., & Goff, P. A. (20190715). The force of fear: Police stereotype threat, self-legitimacy, and support for excessive force. *Law and Human Behavior*, *43*(5), 421. https://doi.org/10.1037/lhb0000339

Violent Crime. (n.d.). National Institute of Justice. Retrieved February 27, 2023, from https://nij.ojp.gov/topics/crimes/violent-crime

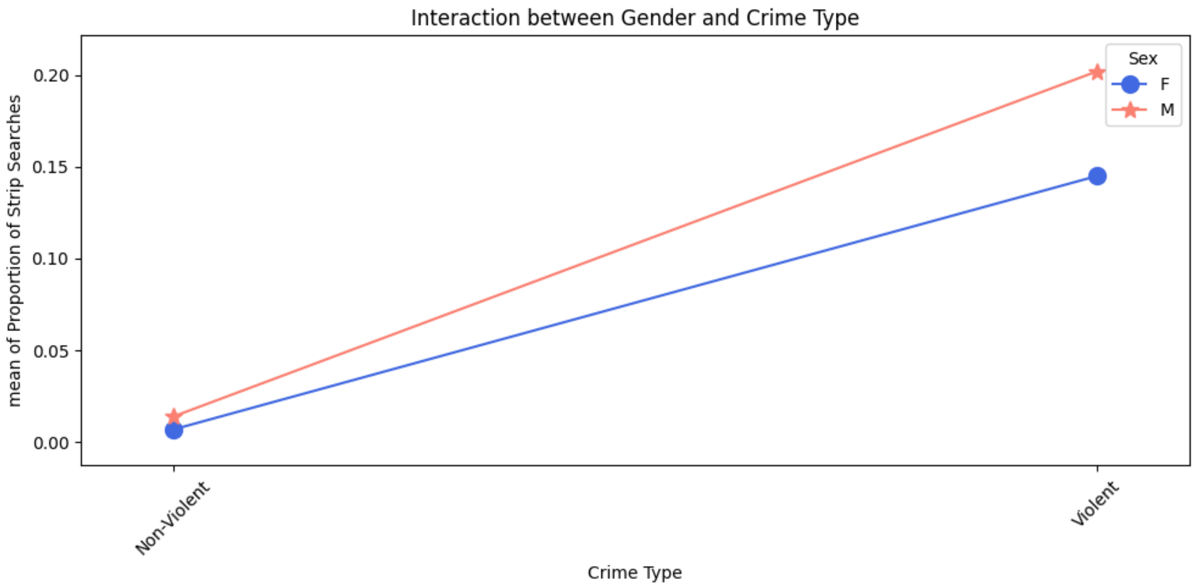
Wortley, S. (n.d.). *POLICE USE OF FORCE IN ONTARIO: An Examination of Data from the Special Investigations Unit*.

8. Appendix

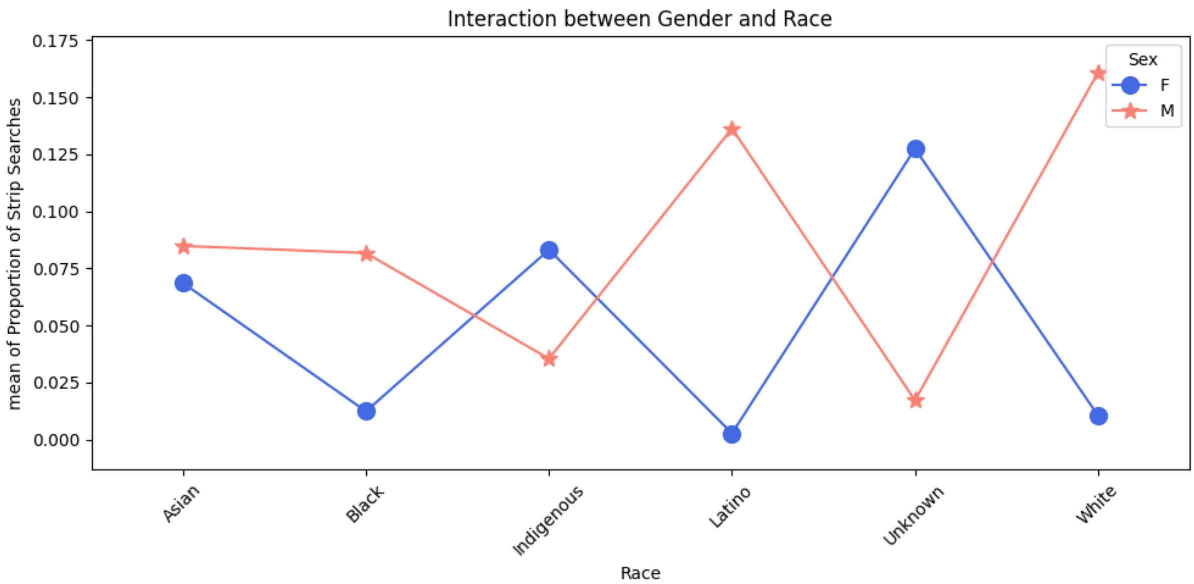
I. Interaction plots for ANOVAs



The interaction plot shows that there Violent crimes have systematically higher strip search proportions than Non-Violent crimes across all race groups.



Regardless of the gender, there is an increase in the proportion of strip searches moving from non-violent to violent crimes. Males also have a higher proportion of strip searches than females.



The interaction plot above shows that there is a lot of overlap between male and female proportions of strip searches across all races. Male groups who were perceived as Indigenous or Unknown have a higher strip search proportion, while female groups who were Asian, Black, Latino, and White represented higher proportions of strip searches.