Introduction

Race- Based Data Collection, Analysis and Public Reporting Policy1 (Board Policy) dated September 2019 was approved by the Toronto Police Services Board on On September 19, 2019 to identify, monitor and address systemic disparities in policing. The Toronto Police Service Board had alleged to the public that they'll always do all the best to maintain fairness in all kinds of policing practices.

However, disparities are still the issues that people complain the most, especially related to police strip research incidents. Therefore, our team decides to investigate the potential relationships between various demographic factors and policing practices, with a particular focus on the degree of aggressiveness exhibited by criminals and the use of strip searches.

Based on the results obtained from T-tests, Tukey's Test, ANOVA tests, and Linear Regression, we conclude that age group and gender do account for disparities in terms of policing practices especially when police did strip research.

Literature Review

Phan, M. B., Panaitescu, M., & Rebelo, N. (2022, February). *Race & Identity Based Data Collection Strategy*. Online Reporting- Toronto Police Service. Retrieved February 26, 2023, from https://www.tps.ca/services/online-reporting/

On September 19 ,2019, the Toronto Police Service (Board) approved the Race-Based Data Collection, Analysis and Public Reporting Policy dated September 2019 to identify, monitor and address systemic racial disparities in policing.

Current situation at that time :

Police Search of Persons in Arrests Service procedure 01-02 Search of Persons governs and outlines possible risk factors for the search of persons, but the decision as to what type of search is appropriate should be accessed on a case-by-case basis. There are four levels of searches: protective, frisk, strip search, and body cavity and all searches must be done sequentially, and there must be reasonable justification for higher levels of searches.

The definition of strip search used by the Service is:

A thorough search of a person's clothing and non-physical search of the body that will often require removal or rearrangement of some, or all, of the person's clothing to permit a visual inspection of a person's private areas: namely the genitals, buttocks, breasts or chest, body cavity, and/or undergarments; the mouth was excluded from this definition despite being a bodily cavity.

(*R. vs. Golden 2001 SCC 83*)

In October 2020, the Service revised the Procedure 01-02, *Search of Persons*. The changes address office accountability, training, and data management. Some of these changes include:

1. All protective and frisk searches are captured on audio and video, wherever possible, to allow for transparency and accountability.
2. All strip searches must be authorized by a supervisor and documented.
3. Update of the information captured during a strip search.
4. All strip searches must be audited at a divisional and senior management level.
5. All officers must complete a robust training module of search of persons including a review of case law.

Then, author Dr. Mai B. Phan, Mihaela Dinca-Panaitescu and Nicole Rebelo conducted *Understanding Strip Searches in 2020 Methodological Report* in October 2022 to analyze the outcome of a change in policy of strip search. The following are the major findings.

To summarize their findings:

1. Strip search rates varied throughout 202 and were significantly reduced following changes in Search policy and procedures in October 2020.
2. Indigenous, Black and White people were over-represented in strip searches relative to their proportion in arrests. After policy changes, there was no longer any over-representation of Indigenous people in strip searches.
3. There were differences by race after accounting for repeat arrests, and arrests related to drug and weapons offenses.
4. There were differences in strip search rates across the city where arrests took place.

Research Objective and Questions

The research objective of this study is to investigate the potential relationships between various demographic factors and policing practices, with a particular focus on the degree of aggressiveness exhibited by criminals and the use of strip searches. By exploring these factors and relationships, we aim to better understand the potential discriminatory practices that may exist in law enforcement and to suggest areas for improvement.

RQ1: Do people's age and gender influence their degree of aggressiveness?
RQ2: Will the number of strip searches be influenced by people's gender and number of arrests?
RQ3: Will the probability of strip searches be influenced by people's gender?

The first research question seeks to understand whether age and gender play a role in the degree of aggressiveness. Understanding this can help law enforcement agencies identify any biases or stereotypes that may be present, as well as develop strategies to reduce violence, and better assess the potential threat posed by a criminal. For question 2 and 3, we want to explore whether the number of strip searches and the probability of being strip searched are influenced by an one’s gender and number of arrests. This is important because strip searches can be traumatizing experiences that should not be used or driven by personal prejudices. By studying whether discrimination exists, we can strive towards more equitable and just policing practices that promote fairness and better serve all members of society.

Method

To fulfill the research objective, we have used three methods in our analysis.

T-tests:to determine whether the means of two groups are significantly different from each other.

ANOVA test:to determine whether the means of three or more groups are significantly different from each other

Post-hoc test e.g, Tukey's HSD: to determine which groups are significantly different from each other following an ANOVA test.

Dataset Description

The dataset used for analyses comes from the Records Management System (R.M.S) that is used to record situations about arrests and strip searches. In normal procedure, people must first be arrested for a violation of law and then may be subject to a strip search. They also must be booked into custody at a police station if they are subject to a strip search. Occurrence records are sources of information about incidents including situations and individuals who were arrested.

The following table is the description of the dataset:

Dataset Name	Type	Description
Arrest_Year	Number	Year when a person is arrested
Arrest_Month	Number	Month period when a person is arrested
EventID	Number	Identification number of an event
ArrestID	Number	Identification number of an arrest
PersonID	Number	Identification number of a person
Perceived_Race	Text	Types of races of an arrested person
Sex	Text	Type of sex an arrested person
Age_group__at_arrest	Text	Age group of an arrested person
Youth_at_arrest__under_18_years	Text	Whether the age of a arrested person is a yonth(age<18)
ArrestLocDiv	Text	Aggregated to the Division level and refers to where the arrest took place within Division boundaries.

		Marked xx when arrest took place outside City of Toronto
StripSearch	Number	If strip search is conducted, 0 means no, 1 means yes.
Booked	Number	If custody is booked, 0 means no, 1 means yes.
Occurrence_Category	Text	The types of occurrence category
Actions_at_arrest__Concealed_i	Number	If concealed items happened at arrest, 0 means no, 1 means yes.
Actions_at_arrest__Combative__	Number	If combative, violent or spitter/biter happened at arrest, 0 means no, 1 means yes.
Actions_at_arrest__Resisted__d	Number	If resisted, defensive or escape risk happened at arrest , 0 means no, 1 means yes.
Actions_at_arrest__Mental_inst	Number	If Mental instability or possibly suicidal, 0 means no, 1 means yes.
Actions_at_arrest__Assaulted_o	Number	If assaulted officer happened at arrest,0 means no, 1 means yes.
Actions_at_arrest__Cooperative	Number	If cooptative happened at arrest, 0 means no, 1 means yes.
SearchReason_CauseInjury	Number	If cause injury is a search reason, 0 means no, 1 means yes.
SearchReason_AssistEscape	Number	If assist escape is a search reason,0 means no, 1 means yes.
SearchReason_PossessWeapons	Number	If possess weapons is a search reason, 0 means no, 1 means yes.

SearchReason_PossessEvidence	Number	If possess evidence is a search reason, 0 means no, 1 means yes.
ItemsFound	Number	If item is found,0 means no, 1 means yes.
ObjectId	Number	Identification number of a found item.

EDA

1 Actions at Arrest

We added up all six indicators of different actions at arrest and created a new variable called DegreeOfAction to represent the person's degree of aggressiveness when being arrested. Hence, the new variable ranged from 0 to 6.

1.1 Age & Sex vs. Actions at Arrest

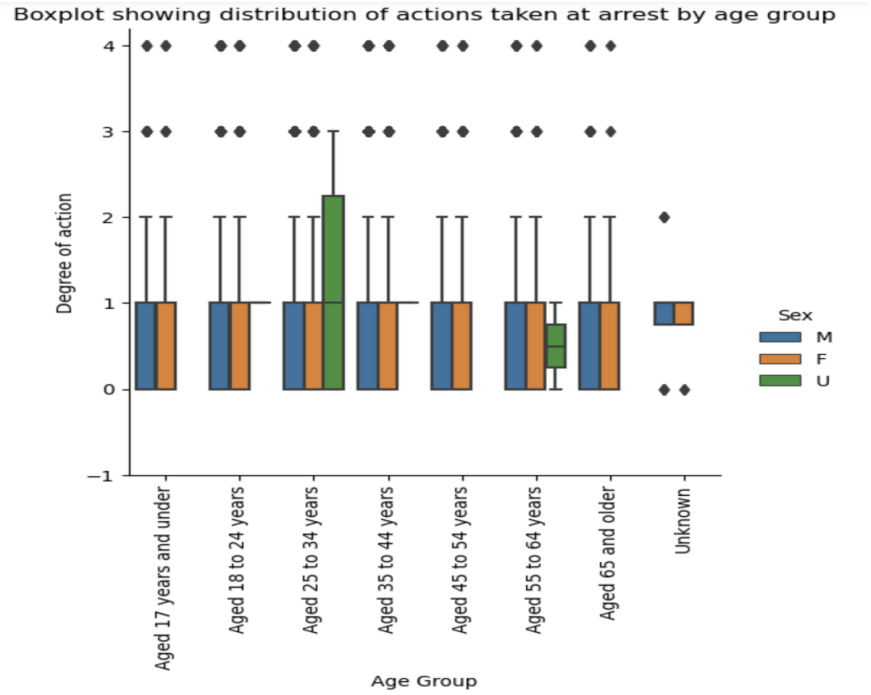

Figure 1: Boxplot for Distribution of Actions Taken at Arrest By Age Group

As we can see in Figure 1, all IQR are from 0 to 1 with no difference depending on Age Groups and Sex. Therefore, there is no obvious relationship between Age Group , Sex vs. Actions at Arrest.

1.2 Sex vs. Actions at arrest

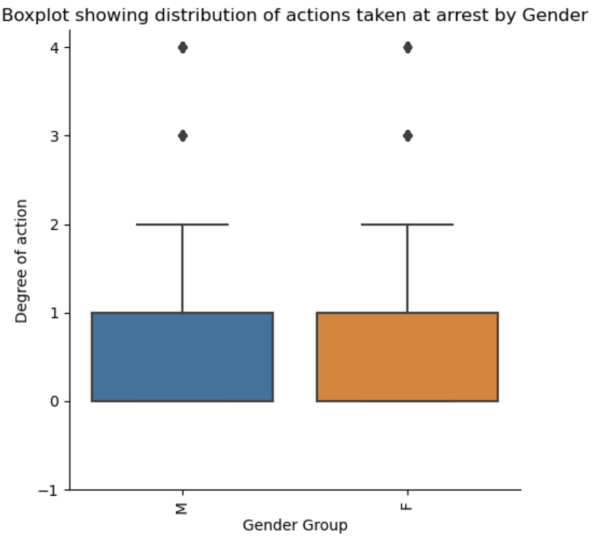

Figure 2: Boxplot for Distribution of Actions Taken at Arrest By Gender

As we can see in Figure 2, all IQR are from 0 to 1 with no difference depending on Gender Group. Therefore,there is no obvious relationship between Gender vs. Actions at Arrest.

T-test: Sex & Degree of Aggressiveness

H0: female mean degree of aggressiveness = male mean degree of aggressiveness

Ha: female mean degree of aggressiveness != male mean degree of aggressiveness

Before conducting the T-test, some assumption checks should be performed to validate the degree of appropriateness, which includes: independent individual measurements, normality, equal variance. The farther the real data is from these assumptions, the less power of test results, i.e. p-value. The details about test limitations due to the way the data is collected and measured will be furtherly discussed in the limitation section.

- Equal Variance: fulfilled, the standard deviation calculated for male and female groups are very close
- Independence: fulfilled, since we treat the each data element as an independent individual and gender is independence measurement for each data elements
- Normality: violated, as Figure 3 shown, both distributions are asymmetric (shewed to the right), that could not be a Gaussian distribution. However, the total number of

females (12617) and male (52650) in this dataset is large enough to apply CLT, the non-normality is not a big issue.

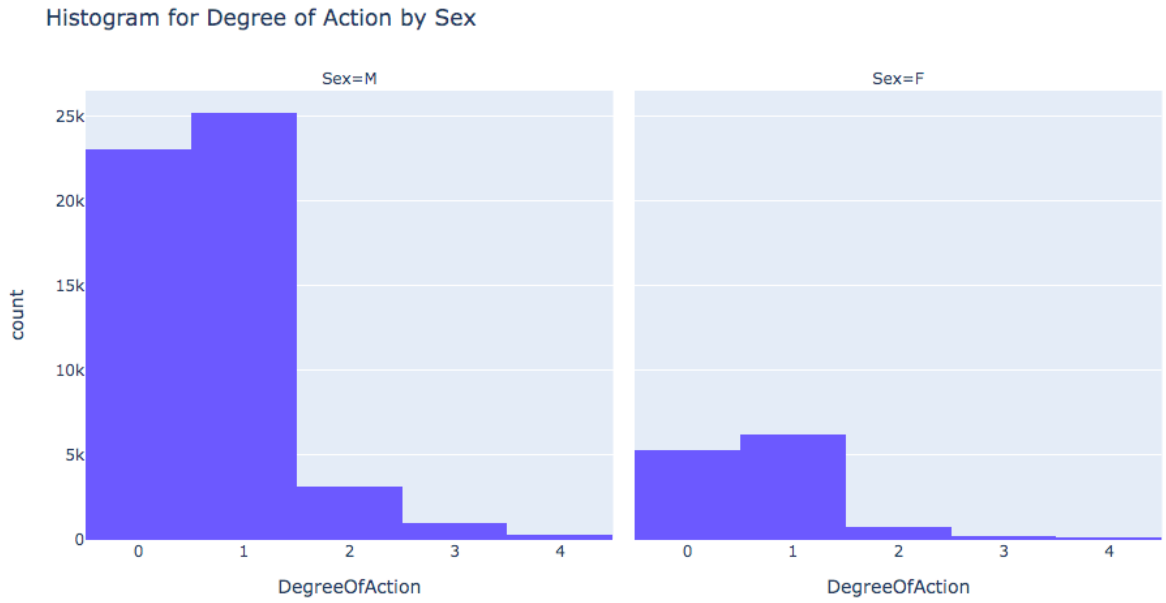

Figure 3: Histogram for Degree of Action by Sex

The results indicate that the mean degree of aggressiveness for females (M=0.70, SD=0.72) is higher than the mean degree of aggressiveness for male (M=0.68, SD=0.72). With alpha established at 0.05, this is a statistically significant difference as test statistic = 3.16 and the p-value (0.0015) is less than 0.05, 95% CI [-1.3, -0.29]. Therefore, we can reject the null hypothesis and conclude that there is a difference in degree of aggressiveness between sex.

1.3 Age vs. Actions at Arrest

As we can see in Figure 4, all IQR are from 0 to 1 with no difference depending on Age Group. Therefore,there is no obvious relationship between Age group vs Actions at Arrest.

Tukey's HSD test for Age Group

Similar to the previous t-test between sex and degree of aggressiveness, the independence and equal variance assumptions are met; normality is violated but can be ignored due to large sample size.

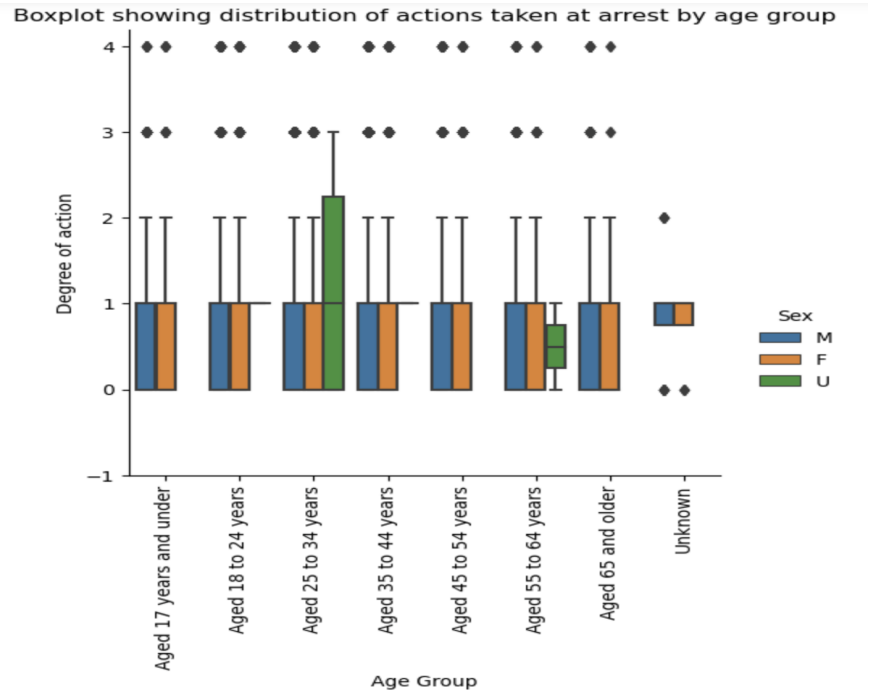

Figure 4: Boxplot for Distribution of Actions Taken at Arrest by Age and Sex

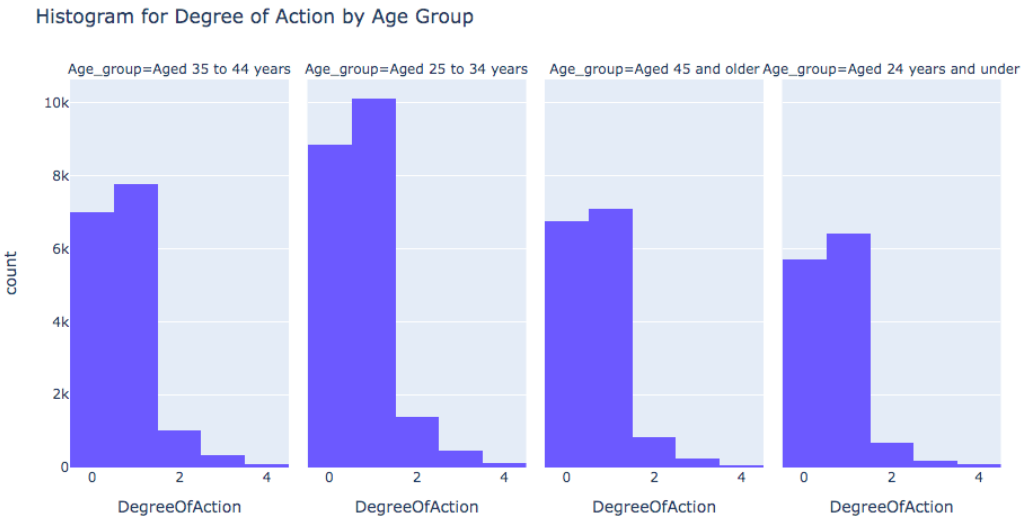

Figure 5: Histogram for Degree of Action by Age

The age category has 4 different levels: 'Aged 35 to 44 years', 'Aged 25 to 34 years', 'Aged 45 and older', 'Aged 24 years and under'. The Tukey's HSD test compared each pair of them to check the difference between means. Therefore, there are a total six hypothesis tests, where the null hypotheses are that the mean differences between each pair of groups equal to 0, $\mu_2 - \mu_1 = 0$ and alternative hypotheses are differences not equal to 0, $\mu_2 - \mu_1 \neq 0$.

Multiple Comparison of Means - Tukey HSD, FWER=0.05					
group1	group2	meandiff	p-val	lower	upper
Aged 24 years and under	Aged 25 to 34 years	0.0394	<0.001	0.0188	0.06
Aged 24 years and under	Aged 35 to 44 years	0.0244	0.0203	0.0027	0.0461
Aged 24 years and under	Aged 45 and older	-0.0186	0.134	-0.0407	0.0035
Aged 25 to 34 years	Aged 35 to 44 years	-0.015	0.1892	-0.0343	0.0043
Aged 25 to 34 years	Aged 45 and older	-0.058	<0.001	-0.0778	-0.0382
Aged 35 to 44 years	Aged 45 and older	-0.043	<0.001	-0.0639	-0.0221

Table 1: Tukey's HSD Test Result: Age Group and Degree of Aggressiveness

The results in Table 1 indicate that with alpha established at 0.05: (1) there are statistically significant differences between 'Age<=24' and '25<= Age <= 34', 'Age<=24' and '35<= Age <= 44', '25<= Age <= 34' and 'Age>=45', '35<= Age <= 44' and 'Age>=45, since the p-values (< 0.01) are all less than 0.05, thereby we reject null hypotheses; (2) there is no statistically significant difference between 'Age<=24' and 'Age>=45' or between '25<= Age <= 34' and '35<= Age <= 44', since the p-values are greater than 0.05, thereby we fail to reject these two null hypotheses. In conclusion, there is a statistically significant difference between gender and between age groups for degree of aggressiveness, thus we will use these two variables for Two-way ANOVA.

2 Strip Search Counts

Strip Search Count variable is obtained by counting the total number of times a person (data elements with the same unique ID) was searched.

2.1 Gender vs. Strip Search Counts

As we can see from Figure 6, male people have more strip search counts than male.Therefore, there seems to be a relationship between sex and strip search counts.

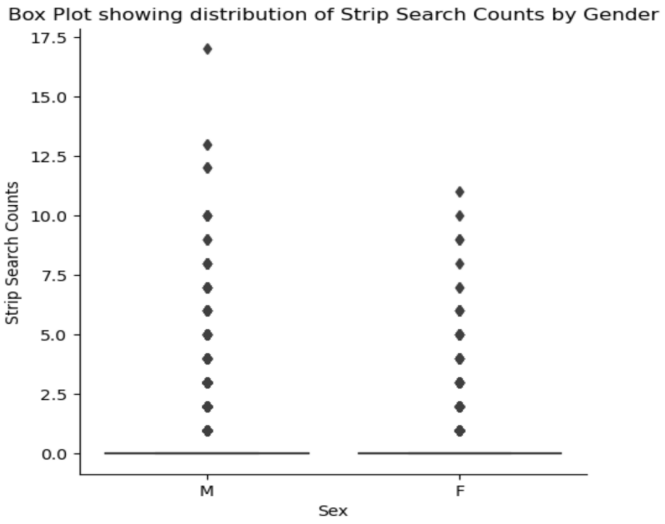

Figure 6: Boxplot for Distribution of Strip Search Counts by Gender

T-test: Gender & Strip Search Counts

H0: mean Strip Search Counts for female = mean Strip Search Counts for male

H1: mean Strip Search Counts for female != mean Strip Search Counts for male

Check Assumptions: (1) Independence: We analyzed the total number of time one person was strip-searched, that people are independent with each other and take independent measurements. (2) Normal distributed: As Figure 7 showed, both distributions are asymmetric (shewed to the right). However, the total number of female and male in this dataset is 12617 and 52650, respectively, then by CLT, the non-normality is not a big problem. (3) Equal variance: Two groups' variance is almost the same, while we use the Welch's t-test which doesn't require equal variance assumption to be met.

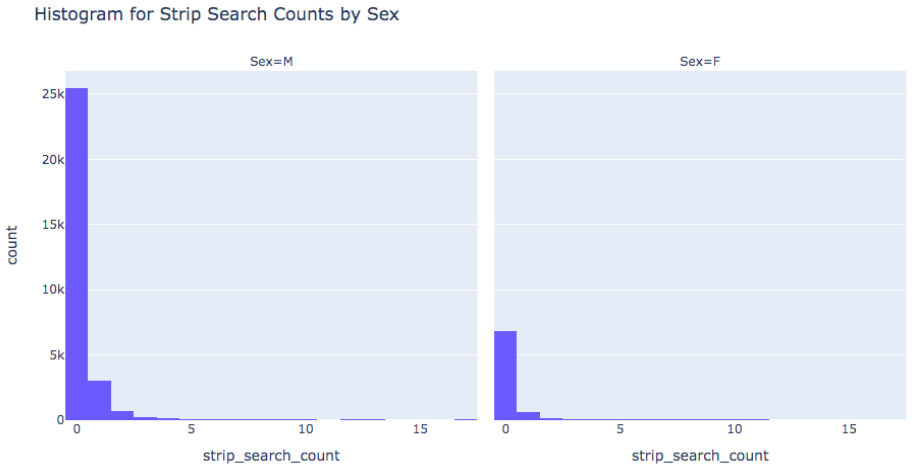

Figure 7: Histogram for Strip Search Counts by Sex

The test results indicate that the mean strip searches for females (M=0.168, SD=0.619) is lower than the mean strip searches for male (M=0.220, SD=0.709). With alpha established at 0.05, this is a statistically significant difference as test statistic = 6.36 and the p-value is less than 0.05, 95% CI [0.04, 0.07] for mean difference between male and female. Therefore, we can reject the null hypothesis and conclude there is a difference in strip search counts between sex.

2.2 Arrest times vs. Strip Search Counts

As we can see from Figure 8, the more arrest times a person has, the more strip searches will be conducted.Therefore, there seems to be a relationship between arrest times and strip search counts.

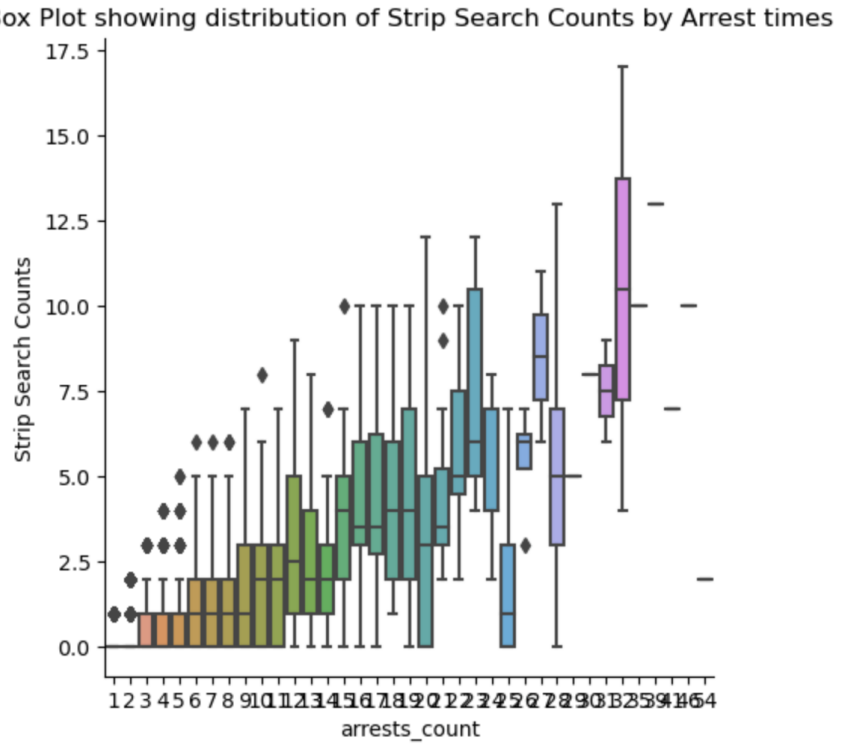

Figure 8: Boxplot for Distribution of Strip Search Counts by Arrest times

T-test: Number of Arrests & Strip Search Counts

H0: mean Strip Search Counts for people arrests once = mean Strip Search Counts for people arrest more than once

H1: mean Strip Search Counts for people arrests once != mean Strip Search Counts for people arrest more than once

Same as the assumption check for t-test between sex and strip search counts, the assumption for independence is met; normality is violated but can be ignored due to large sample size; equal variance assumption is violated by the way we divide the two groups, but not required due to the adoption of Welch's T-test.

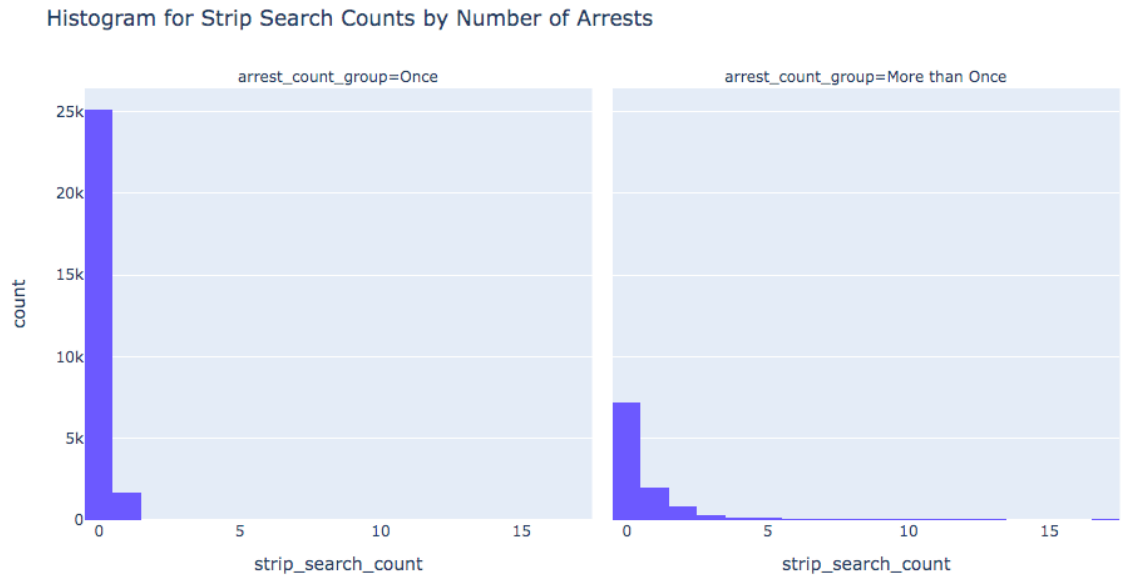

Figure 9: Histogram for Strip Search Counts by Number of Arrests

The test results indicate that the mean strip searches for people arrested once (M=0.062, SD=0.24) is lower than the mean strip searches for people arrested more than once (M=0.582, SD=1.16). With alpha established at 0.05, this is a statistically significant difference as test statistic = -45.67 and the p-value is less than 0.05, 95% CI [-0.54, -0.498]. Therefore, we can reject the null hypothesis and conclude there is a difference in strip search counts between arrest times.

2.3 Race vs. Strip Search Counts

As we can see from Figure 10, most of strip search counts are fall in 0 to10 and there is no major relationship between strip search counts and difference in races.

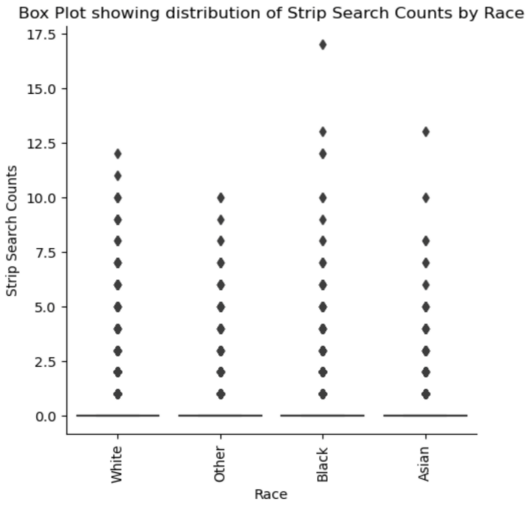

Figure 10: Boxplot for Strip Search Counts by Race

Tukey's HSD test for Race

Same as the assumption check for t-test between sex and strip search counts, the independence assumption is met; normality is violated but can be ignored due to large sample size; equal variance assumption is fulfilled as shown in Table 2.

	Mean	Standard deviation
White	0.206	0.683
Black	0.209	0.702
Asian	0.210	0.667
Others	0.215	0.708

Table 2: Mean and standard deviation of strip search counts for four race groups

The race category has 4 different levels:'White', 'Other', 'Black', 'Asian'. The tukey's HSD test compared each pair of them to check the difference between means. Therefore, there are a total six hypothesis tests, where the null hypotheses are that the mean differences between each pair of groups equal to 0, $\mu_2 - \mu_1 = 0$, and alternative hypotheses are the mean differences not equal to 0, $\mu_2 - \mu_1 \neq 0$.

As shown in Table 3, we fail to reject all null hypotheses, which indicates that for the number of strip searches, there is no difference between race. Thus we will not use this variable for ANOVA tests.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Asian	Black	-0.0004	1.0	-0.0343	0.0335	False
Asian	Other	0.0048	0.9869	-0.0319	0.0416	False
Asian	White	-0.0035	0.9928	-0.0358	0.0289	False
Black	Other	0.0052	0.9638	-0.0229	0.0334	False
Black	White	-0.003	0.985	-0.0251	0.019	False
Other	White	-0.0083	0.8498	-0.0345	0.018	False

Table 3: Tukey's HSD Test for Race and Strip Search Counts

3 Strip Search Probability

The strip search probability variable is obtained by taking the fraction between 'Total number of times a person was searched' and 'Total number of times a person was arrested'.

3.1 Gender vs Strip Search Prob

As we can see from Figure 11, strip search probability is different between male and female. It seems to be a relationship between sex and strip search probability.

Figure 11: Boxplot for Strip Search Probability by Sex

T-test: Gender & Strip Search Prob

H0: mean Strip Search Prob for female = mean Strip Search Prob for male

H1: mean Strip Search Prob for female != mean Strip Search Prob for male

For assumption check, the independence assumption is met that we analyzed based on each individual criminal with independent measurement; normality assumption is violated, while the sample size for both genders is large enough to apply CLT; the equal variance assumption is met, while we use the Welch's t-test which doesn't require equal variance for testing.

The test results indicate that the mean strip searches for females (M=0.068, SD=0.222) is lower than the mean strip searches for male (M=0.087, SD=0.247). With alpha established at 0.05, this is a statistically significant difference as test statistic = 6.65 and the p-value is less than 0.05, 95% CI [0.014, 0.025] for difference between male and female. We can reject the null hypothesis and conclude there is a difference in strip search probability between sex.

3.2 Race vs Strip Search Prob

As we can see from Figure 12, the shapes of the plot are roughly the same. Therefore, there is no major relationship between strip search prob and difference in races.

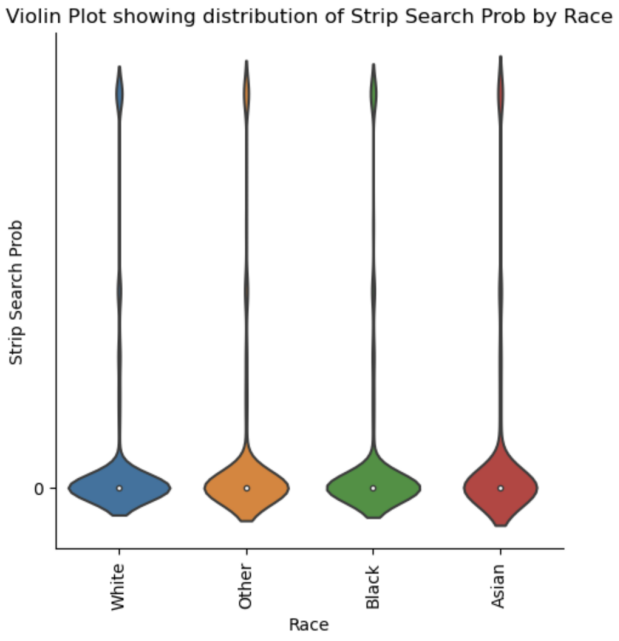

Figure 12: Violin Plot for Strip Search by Race

Tukey's HSD test for Race

The independence assumption is met; normality assumption is violated, while the sample sizes are large enough to apply CLT; the equal variance assumption is met by Table 4.

	Mean	Standard deviation
White	0.083	0.242
Black	0.082	0.240
Asian	0.087	0.246
Others	0.083	0.242

Table 4: Mean and standard deviation of strip search probability for four race groups

Same as section 2.3, the Tukey's HSD test conducts a total of six hypothesis tests, where the null hypotheses are that the differences between each pair of race is equal to 0, and alternative hypotheses are that the differences are not equal to 0.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Asian	Black	-0.0049	0.7133	-0.0168	0.007	False
Asian	Other	-0.0042	0.8309	-0.0171	0.0086	False
Asian	White	-0.004	0.7986	-0.0153	0.0073	False
Black	Other	0.0006	0.9983	-0.0092	0.0105	False
Black	White	0.0009	0.9911	-0.0068	0.0086	False
Other	White	0.0002	0.9999	-0.0089	0.0094	False

Table 5: Tukey's HSD Test for Race and Strip Search Probability

We fail to reject all null hypotheses, which indicates that for the probability of strip searches, there is no difference between race. Thus we will not use this variable for ANOVA tests.

Results

Research Question 1

Do people's age and gender influence their degree of aggressiveness?

Interaction plots

While the interaction plot does not provide any information on statistically significant difference, the plot showed: (1) Female criminals have a higher average degree of action(aggressiveness) compared to male criminals; (2) Older criminals (aged 45 and older) have a lowest average degree of action(aggressiveness) compared to other criminals; (3) Criminals aged between 25-34 have a highest average degree of action(aggressiveness) compared to other criminals; (4) There was a noticeable difference in average degree of aggressiveness between female and male criminals who were under 24 years old or above 45 years old; (5) Overall, the trend for two lines are similar, kind of parallel, but we will first run the interaction model to check if the interaction is indeed not significant.

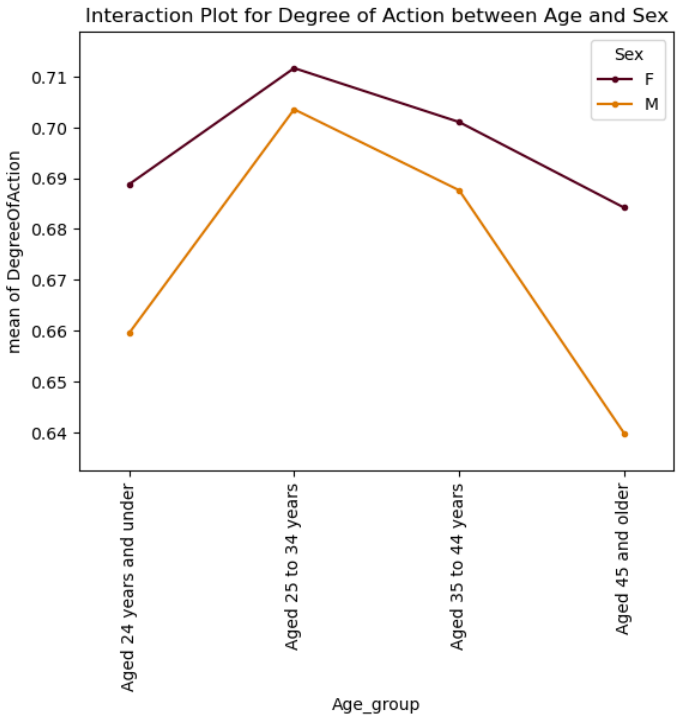

Figure 13: Interaction Plot for Degree of Action between Age and Sex

For assumptions checking, same as we discussed for T-test, the equal variance and independence errors assumptions are met, while for normality, as the histograms below, all distributions are asymmetric (shewed to the right). This is one limitation of our test. But we can conduct monotonic power transformation to make it more symmetric.

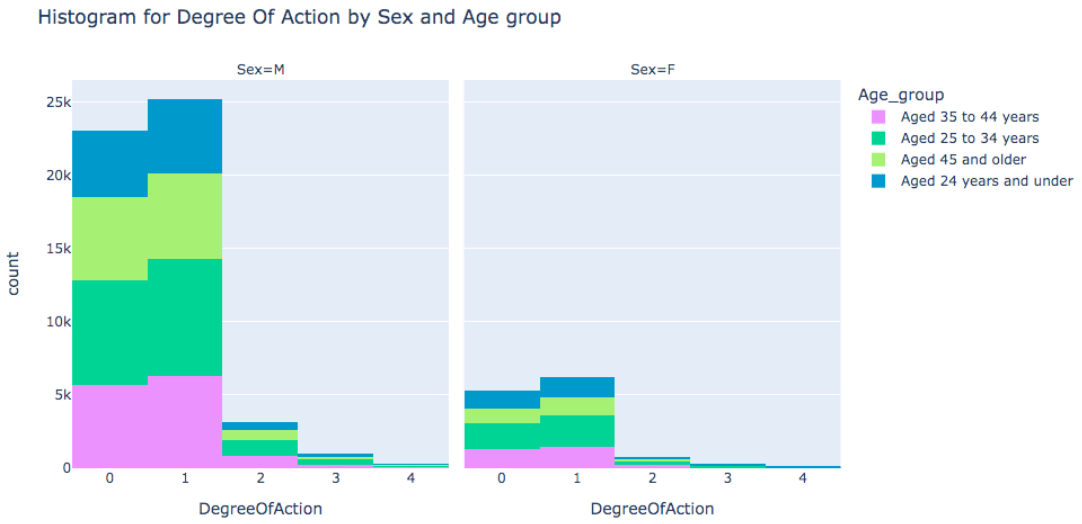

Figure 14: Histogram for Degree of Action by Sex and Age

Age_group	Sex	Variance
Aged 24 years and under	F	0.532
	M	0.482
Aged 25 to 34 years	F	0.522
	M	0.55
Aged 35 to 44 years	F	0.561
	M	0.531
Aged 45 and olde	F	0.476
	M	0.479

Table 6: Variance calculated for all combinations of Age and Sex

Two - Way ANOVA: Interaction Model

1. *H0: The population means of the four independent age groups are all equal*
   Ha: At least one age group has different population mean
2. *H0: The population means of the two independent sex groups are equal*
   Ha: Mean degree of aggressiveness of male is different from which of the female

3. *H0: There is an additive relationship between Age and Sex in their effects on the outcome*
   Ha: There is an interactive relationship between Age and Sex in their effect on the outcome

	sum_sq	DF	F	PR
C(Age_group)	33.16	3	21.38	<0.001
C(Sex)	4.67	1	9.022	0.003
C(Age_group):C(Sex)	1.99	3	1.28	0.278
Residual	33737.16	65235		

Table 7: Interaction ANOVA result for Age and Sex Effect on Degree of Aggressiveness

The results of a two-way ANOVA showed that the interaction between the explanatory variables Sex and Age is not significant (F=1.28, p=0.277), so we have no evidence to reject the additive model, and we conclude that Age effects on the outcome are the same for both genders, and gender effects on the outcome are the same for all four levels of Age. Therefore it is appropriate to re-fit the ANOVA with an additive model rather than an interactive model.

Two - Way ANOVA: Additive Model

1. *H0: The population means of the four independent age groups are all equal*
   Ha: At least one age group has different population mean
2. *H0: The population means of the two independent sex groups are equal*
   Ha: Mean degree of aggressiveness of male is different from which of the female

	sum_sq	DF	F	PR
C(Age_group)	33.16	3	21.38	<0.001
C(Sex)	4.67	1	9.022	0.003
Residual	33739.15	65238		

Table 8: Additive ANOVA result for Age and Sex Effect on Degree of Aggressiveness

Using a significance level of alpha = 0.05, we can conclude that:

1. Both Age and Sex affect degree of aggressiveness which have p-value < 0.05.

2. From ANOVA at least one level of age differs from the other three, and this is true separately for males and females because the additive model is an adequate model. But we cannot make further important statements about which levels of age are significantly different
3. By linear regression, we have estimated mean for female criminals 0.02 higher than male criminals, for equivalent age ($p = 0.003 < 0.05$). Similarly, estimated mean for people aged 25 to 34 is 0.04 higher than people aged below 24, controlling gender; estimated mean for people aged 45 or above is 0.02 lower than people aged below 24, controlling gender
4. ANOVA table also shows that the standard deviation in any group is approximately 0.72 (square root of 33739.15/65238)

Research Question 2

Will the number of strip searches be influenced by people's gender and number of arrests?

Interaction plots

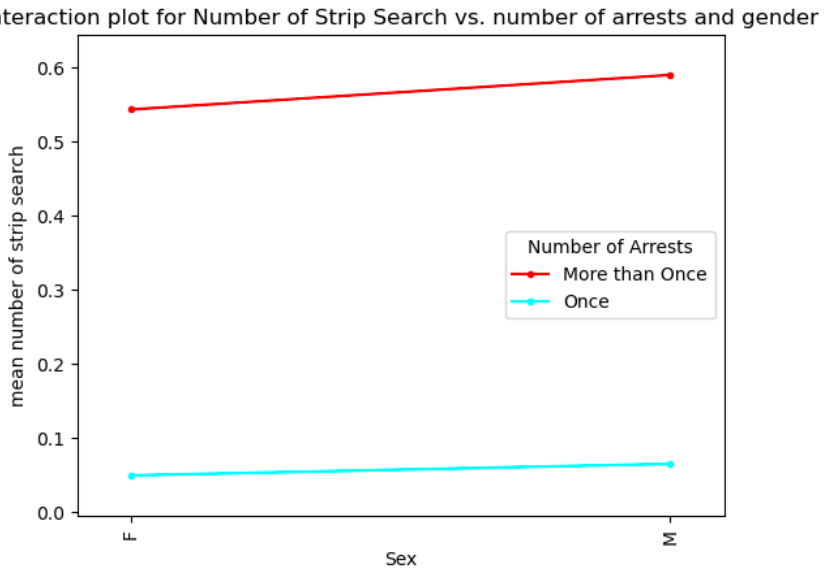

Figure 15: Interaction Plot for Number of Strip Search and Number of Arrests and Sex

The plot showed: (1) People arrested more than once will have higher mean of strip searches compared to people only arrested once; (2) Male criminals tend to have a higher mean of strip searches compared to female criminals; (3) Overall, the trend for two lines are similar and parallel, thus we can just conduct the additive ANOVA model (the interactive model result can be found in Appendix).

Similar to the previous ANOVA analysis, we checked that both equal variance and independence errors assumptions are fulfilled, while for normality, all distributions are asymmetric (shewed to the right), which is one of the limitations of our test.

Two - Way ANOVA: Additive Model

1. *H0: The population means of the two independent Arrests Count groups are all equal*
   Ha: Mean strip searches for people arrested only once is different from people being arrested more than once
2. *H0: The population means of the two independent sex groups are equal*
   Ha: Mean strip searches for male is different from which of female

	sum_sq	DF	F	PR
C(Arrest_count_group)	2036.2	1	4811.11	<0.001
C(Sex)	3.3	1	7.8	0.005
Residual	15795.9	37322		

Table 9: Additive ANOVA result for Arrest Counts and Sex Effect on number of strip search

Using a significance level of alpha = 0.05, we can conclude that: (1) Both Arrest counts and Sex affect strip search counts which have p-value < 0.05; (2) Since both variables only have two levels, which means there is statistical difference between male & female, between people arrested only once and more than once, same result from the T-test. And from the t-test, we know that mean strip search counts for male, and for people arrested more than once is higher than the opposite groups.

Research Question 3

Will the probability of strip searches be influenced by people's gender?

One-Way ANOVA

Since we only have one independent variable here, which has only two levels. The result of one-way ANOVA is exactly the same as the T-test result.
H0: mean Strip Search Prob for female = mean Strip Search Prob for male

H1: mean Strip Search Prob for female != mean Strip Search Prob for male
We got test statistic s = 39.135 with associated p-value = 3.998e-10, which means there is a statistically significant difference between gender on the probability of strip search.

Conclusion and Discussion

Based on the results obtained from T-tests, Tukey's Test, ANOVA tests, and Linear Regression, we can now answer our three research questions. Both Age and Sex did affect degree of aggressiveness (DOA), where the estimated DOA for female is 0.02 higher than male. The two most different groups on DOA are people aged 25 to 34 and people aged below 24, where the former is 0.04 higher than the latter. The differences are statistically significant, while not very significant in practice. Furthermore, the number of strip searches is not influenced by race but by gender and number of arrests, while the effect of gender and number of arrests do not interact with each other towards the number of strip searches.

We are 95% confident that: mean strip search probability for male is between 0.014 and 0.025 higher than females; mean strip search count for people arrested once is between 0.498 and 0.54 lower than people arrested more than once.
In conclusion, our research indicates that while gender may not have a significant impact on an individual's degree of aggressiveness, it does appear to play a role in the use of strip searches, as well as age group. This highlights the need for more equitable and carefully considered policing practices that take into account the potential biases and discrimination within law enforcement. By working towards more fair policing strategies, we can help promote public safety and create a more inclusive society.

Limitation

In our dataset, some individuals were arrested multiple times. To analyze the associated measurements, we treated each arrest as a separate individual in some tests, leading to non-independent and related data and results. Additionally, for the quantitative outcome variable required by ANOVA test, we created four variables to approximate continuous variables: Degree of Aggressiveness, # of arrests, # of strip searches, and strip search probability. However, these variables are actually discrete rather than continuous, potentially

leading to inappropriate and misleading ANOVA results. Furthermore, most of the variables in the dataset are skewed, violating the normality assumption required for both T-test and ANOVA, which may result in non-interpretable p-values and incorrect results, while the large sample size and central limit theorem can mitigate this issue. Alternatively, in the future, we could address this problem by applying a monotonic power transformation to the data.

Appendix

RQ2: Two-way Interaction Anova

	sum_sq	DF	F	PR
C(Arrest_count_group)	2036.2	1	4811.11	<0.001
C(Sex)	3.3	1	7.8	0.005
C(Arrest_count_group):C(Sex)	1.09	1	2.58	0.108
Residual	15795.9	37322		

Reference

Phan, M. B., Panaitescu, M., & Rebelo, N. (2022, February). *Race & Identity Based Data Collection Strategy*. Online Reporting- Toronto Police Service. Retrieved February 26, 2023, from https://www.tps.ca/services/online-reporting/

Office of Independent Police Review Director.(2019). Breaking the Golden Rule: A Review of the Police Strip Searches in Ontario. https://www.oiprd.on.ca/news/systemic-reviews/