University of Toronto

Midterm Assignment

By:

Xinxuan Lin, Peggy Pu

Faculty of Information
INF2178
Group 15
February 28, 2023

Table of Contents

1. Introduction

Strip searches are a controversial practice in law enforcement that involve the complete removal of clothing from an individual in order to search for contraband or other items (BSB Solicitors, 2022). These searches can be invasive and humiliating, and have been the subject of much debate and criticism in recent years (Lemke, 2022). In particular, concerns have been raised about the disproportionate use of strip searches against certain demographics, including women, people of color, and those who are arrested at a young age (Lemke, 2022). In this report, we aim to explore the relationship between sex, race, age group, and the average number of being subjected to a strip search at the time of arrest. To achieve this goal, we will analyze a comprehensive dataset of arrest records to determine the incidence of strip searches across different sex, race, and age categories. Our analysis will uncover any discernible patterns or trends, allowing us to better understand the relationship between these demographic factors and the percentage of being strip searched. And we hope to shed light on the potential biases and disparities that exist within the criminal justice system with regard to strip searches, and to identify areas where reform may be needed.

1.1 Literature review

Several studies have examined the relationship between sex, race, age group at the time of arrest, and the probability of being strip searched. Research indicates that females are less likely to be strip searched than males. A study by Prison Policy Initiative (2019) found that although female making up a growing portion of arrests, male arrestees were still more likely to be strip searched than female arrestees. However, it should be noted that the reasons for the difference in strip search rates between males and females are not clear.

Another finding of the study is that women of color are more likely to be stopped and searched by police than white women. For example, Black women are 1.5 times more likely to be stopped and searched than white women, even though they are less likely to be carrying contraband (Prison Policy Initiative, 2019). This disparity in police stops and searches can have significant consequences for women of color, such as increased exposure to violence and trauma. Similarly, A study by Susan Nembhard and Lily Robin (2021) found that Black and Hispanic individuals were more likely to be strip searched than White individuals.

Finally, research indicates that the age group at the time of arrest may also play a role in the probability of being strip searched. Daniel M. Blonigen's article explores the relationship between personality traits and age-related crime trends. The research has consistently shown that crime rates decline as people age, and the decline is steeper for violent crimes than for nonviolent crimes (Blonigen, 2021).

1.2 Research Questions and objective

Our study aims to examine the relationship between sex, race, and age at the time of arrest and the average number of being subjected to strip searches. Our research questions are based on a thorough literature review and a preliminary analysis of the dataset, which is presented in the Descriptive Statistics and T-tests section below. Specifically, we will investigate the following research questions:

RQ1: How does the percentage of strip searches vary based on an individual's race and sex?

RQ2: How does the percentage of strip searches vary based on an individual's age?

By answering these research questions, we hope to provide a deeper understanding of the factors that influence strip searches in our dataset. We believe that exploring the factors, such as sex, race, and age, will shed greater light on the underlying patterns and disparities in the use of strip searches in law enforcement.

2. EDA

2.1 Descriptive Statistics

The dataset comprises 34,650 observations, with 27,434 males and 7,216 females. A breakdown of the number of arrests by gender and race during 2020 and 2021 is presented in Figure 1. Across all racial categories, males have a higher number of arrests than females, which could be attributed to their larger representation in the dataset. In terms of race, the majority of individuals arrested are white and black, followed by East/Southeast Asian, unknown or legacy race, South Asian, Middle-Eastern, Latina, and indigenous people. Notably, indigenous males and females have the lowest number of arrests among all the racial groups, which could be due to the distribution of the population across different racial groups in the Greater Toronto Area.

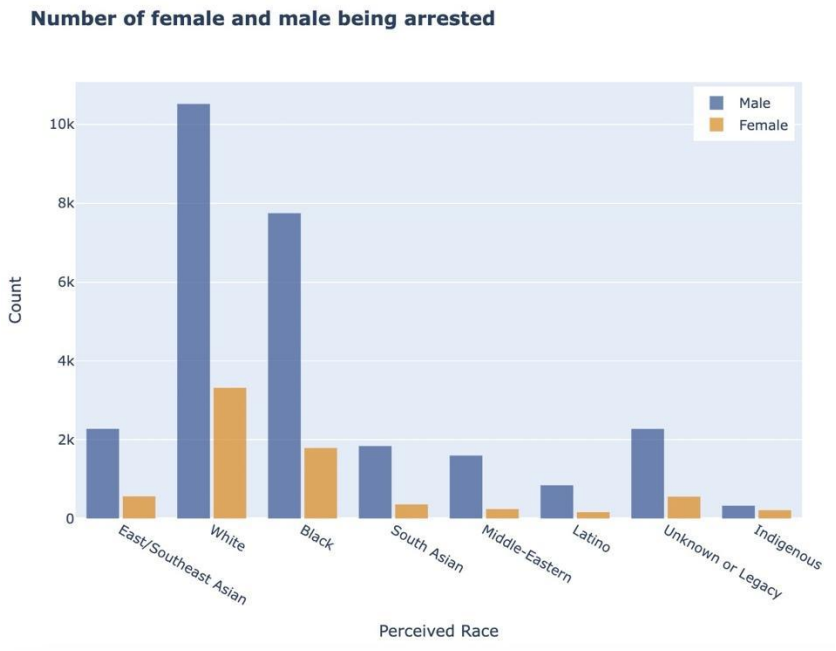

Figure 1: Distribution of males and females being arrested in each race group.

Figure 2 displays the total number of arrests and the percentage of strip searches for different racial and gender groups. Similar to the pattern seen in Figure 1, white and black individuals have the highest number of arrests, while indigenous people have the lowest. However, the right side of Figure 2 shows a different trend: black males have the highest probability of being strip searched, followed by indigenous males, white males, and other minority groups. It is surprising that both indigenous males and females have a relatively high percentage of strip searches compared to their total number of arrests, and this trend is also seen in white females. Even though the total number of arrests for white females is only about one third of the male arrests, their strip search percentages are almost the same. Furthermore, white females rank fourth in terms of the highest percentage of strip searches among all racial and gender groups. Females whose race is categorized as unknown or legacy also have a relatively high probability of being strip searched compared to other racial groups.

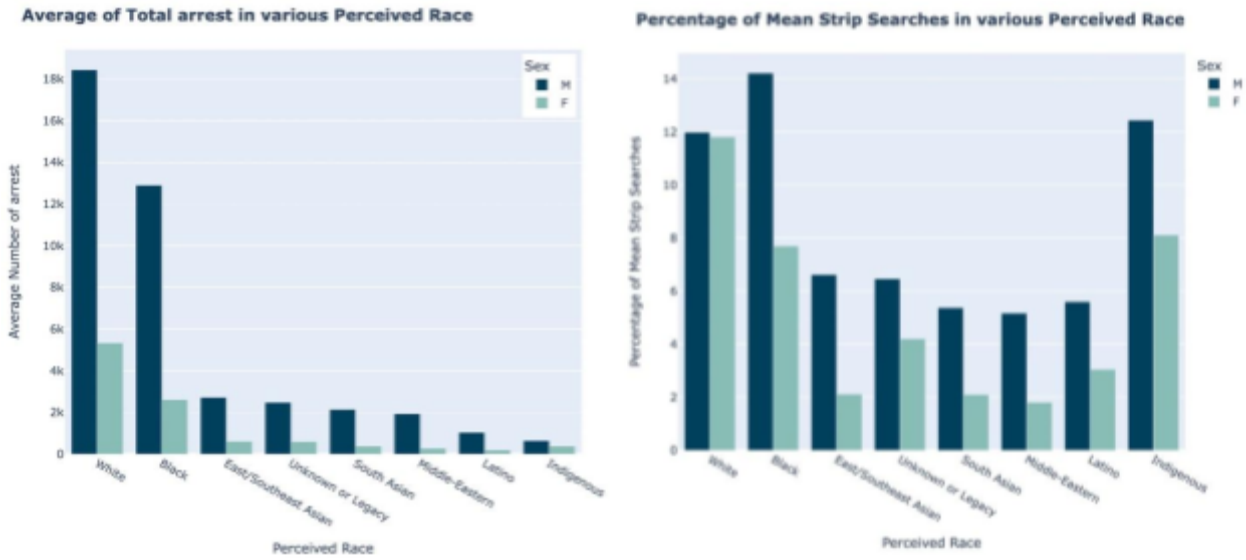

Figure 2: Number of total arrest and percentage of strip searches for different sex groups across the races

The dataset consists of 214,319 entries for people who were not subjected to strip searches (with a strip search percentage of 0) and 28,231 entries for those who were (with a strip search percentage greater than 0). As a result, the dataset is highly imbalanced, and the distribution is right-skewed with minimum, median, and interquartile range all equaling 0. Table 1 presents a comprehensive breakdown of the strip search percentages for various genders and races.

Table 1: statistics of percentage of strip searches across sex and race.

		count	mean	std	min	25%	50%	75%	max
Perceived_Race	Sex								
Black	F	1791.0	6.441413	22.737082	0.0	0.0	0.0	0.0	100.0
	M	7747.0	11.485870	28.665206	0.0	0.0	0.0	0.0	100.0
East/Southeast Asian	F	565.0	1.805310	13.219276	0.0	0.0	0.0	0.0	100.0
	M	2278.0	5.145648	20.842163	0.0	0.0	0.0	0.0	100.0
Indigenous	F	214.0	7.632399	24.874791	0.0	0.0	0.0	0.0	100.0
	M	328.0	10.002959	25.686455	0.0	0.0	0.0	0.0	100.0
Latino	F	164.0	2.235772	13.251846	0.0	0.0	0.0	0.0	100.0
	M	846.0	4.828605	20.340304	0.0	0.0	0.0	0.0	100.0
Middle-Eastern	F	244.0	1.844262	13.095107	0.0	0.0	0.0	0.0	100.0
	M	1600.0	4.463170	19.480398	0.0	0.0	0.0	0.0	100.0
South Asian	F	361.0	2.077562	14.037829	0.0	0.0	0.0	0.0	100.0
	M	1841.0	4.771246	20.326802	0.0	0.0	0.0	0.0	100.0
Unknown or Legacy	F	558.0	3.942652	19.246402	0.0	0.0	0.0	0.0	100.0
	M	2276.0	5.928320	22.904613	0.0	0.0	0.0	0.0	100.0
White	F	3319.0	8.453872	24.761885	0.0	0.0	0.0	0.0	100.0
	M	10518.0	8.413567	24.265143	0.0	0.0	0.0	0.0	100.0

According to Table 1, we can observe that there are differences in the mean and standard deviation among the various groups. However, the mean and standard deviation for white males and females do not exhibit a dramatic difference.

The large number of observations with 0 strip searches causes the interquartile range and whiskers to be compressed in Figure 3. As shown in the figure, data for individuals who have not undergone strip searches is also right-skewed, with strip search percentages greater than 0 being considered outliers. The median for both groups is 0, but the mean percentage of strip searches for individuals with negative actions at the time of arrest is higher than the other group. Furthermore, the data for individuals with negative actions is more widely dispersed and scattered than the other group.

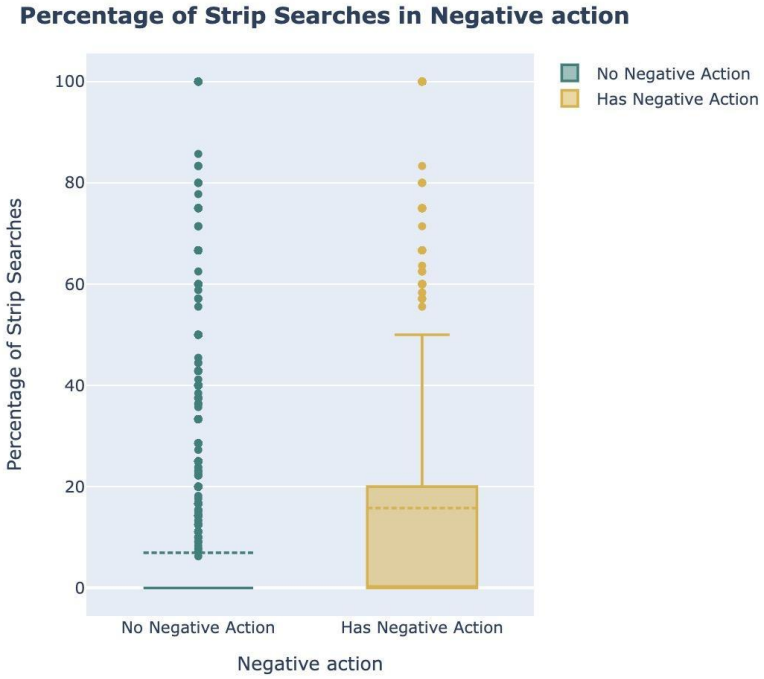

Figure 3: Boxplot to show the statistics of percentage of strip searches break down by negative action

Figure 4 shows that a large proportion of individuals in both youth categories had zero strip searches, causing the interquartile range and whiskers to be compressed. The median number of strip searches for both groups is zero. However, our analysis found that the mean percentage of individuals subjected to strip searches at the time of arrest was 6.06% for the youth category, which is lower than the other age group.

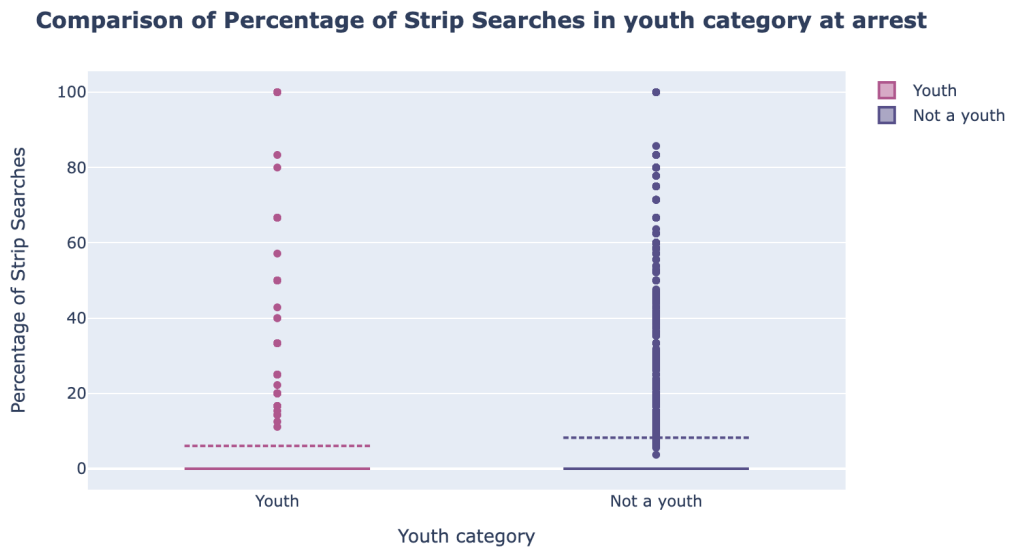

Figure 4: Boxplot to show the statistics of the percentage of strip searches in the youth category at arrest

Figure 5 illustrates the percentage of strip searches conducted for various age groups. Those aged 17 years and younger have the highest percentage of strip searches at 13.78%, while those aged 65 years and older have the lowest percentage of strip searches at 4.98%. This finding is consistent with previous research indicating that crime rates generally decrease as people age.

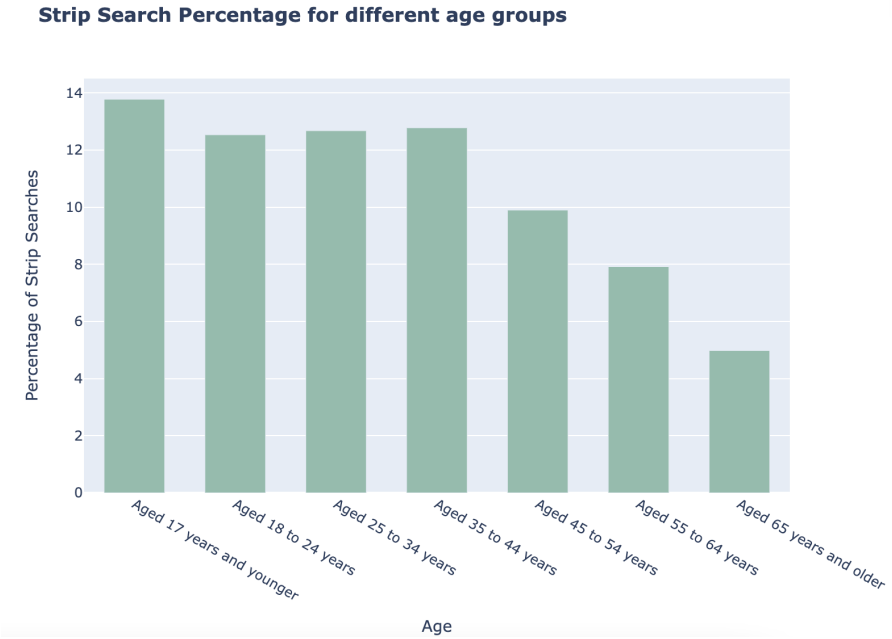

Figure 5: Barplot to show the statistics of the percentage of strip searches in different age groups at arrest

2.2 T-tests

We used Welch's t-test with categorical variables to assess whether there was a significant difference in mean strip search percentages between groups, given the imbalance in sample size. The categorical variables included perceived race, sex, youth status, negative action, and age group.

The assumptions for Welch's t-test are: (1) Independence errors; (2) continuous outcome variable; (3) categorical covariate with two levels; (4) outcome data for each group are approximately normality distributed. Since Welch's t-test is conducted, there is no assumption for residuals to have constant variance.

2.2.1 Assumption Validation

The assumption of independence of errors is satisfied since each row in the dataset corresponds to a single individual, ensuring that there are no duplicate rows representing the same individual with different characteristics such as perceived race, sex, youth status, and age group.

The percentage of strip searches is a continuous variable representing the average number of strip searches among their total arrest. Perceived race is categorized as white or not white, while sex is classified as male or female. Youth status is determined as either youth or not a youth, and age group is divided into below 34 or above 35 years old. Negative action is classified as either no negative action or has negative action. All five covariates are two-level categorical variables.
To assess the normality assumption, we conducted a Shapiro-Wilk test for each variable, and the results are presented in Table 2.

Table 2: Shapiro-Wilk test result for all covariates

Variable	Level	P-value
Perceived Race	White	< 0.01
	Not white	< 0.01
Sex	Male	< 0.01
	Female	< 0.01
Negative Action	Has negative action	< 0.01
	No negative action	< 0.01
Youth Status	Youth	< 0.01
	Not a youth	< 0.01
Age Group	Below 34 years old	< 0.01
	Above 35 years old	< 0.01

The p-values for all groups are lower than the significance level (α=0.05), indicating strong evidence to reject the null hypothesis that the outcome data for each group follows a normal distribution. Therefore, we can conclude that the normality assumption is not met.

However, one of the reasons that may have caused this issue is that the Shapiro-Wilk test is more suitable for sample sizes smaller than 5,000, whereas the sample size in each group is greater than that, which may have resulted in inaccurate p-values.

2.2.1.1 Perceived Race and Percentage of Strip Searches

We observed a difference in the pattern of the percentage of strip searches across perceived race groups compared to the trend in total arrests. Therefore, we used Welch's t-test to determine whether there is a significant difference in the percentage of strip searches between white and non-white people.

The hypothesis of the t-test are:

$Null\ Hypothesis\ (H_0)$: The population means of two independent levels: white people and non-white people, are equal.

$Alternative\ Hypothesis\ (H_a)$: The population means of two independent levels: white people and non-white people, are not equal.

The results of the Welch's t-test indicate that the mean percentage of strip searches for white people (M=8.42, SD=24.38) is higher than the mean percentage of strip searches for non-white people (M=7.47, SD=24.22). The p-value of 0.00035 is less than the significance level (α=0.05) with a 95% confidence interval of [0.43, 1.48]. Therefore, we have strong evidence to reject the null hypothesis and conclude that the mean percentage of strip searches differs between white and non-white people.

2.2.1.2 Sex and Percentage of Strip Searches

A Welch's t-test was performed to investigate whether there is a difference in the percentage of strip searches between male and female individuals. The hypotheses for the test are:

$Null\ Hypothesis\ (H_0)$: The population means of two independent levels: male and female, are equal.

$Alternative\ Hypothesis\ (H_a)$: The population means of two independent levels: male and female, are not equal.

The results of Welch's t-test indicate that the mean percentage of strip searches for male individuals (M=8.23, SD=24.78) is greater than that for female individuals (M=6.37, SD=22.28). The p-value (<0.00000) is less than the significant level ($\alpha = 0.05$), with 95% CI [1.26, 2.45]. Therefore, we have sufficient evidence to reject the null hypothesis that the mean percentage of strip searches is the same for male and female individuals.

2.2.1.3 Youth status and Percentage of Strip Searches

A Welch's t-test was performed to determine whether there is a difference in the percentage of strip searches between individuals who are classified as youth (below 17 years old) and those who are not. The hypothesis tested are:

$Null\ Hypothesis\ (H_0)$: The population means of two independent levels: youth and not a youth, are equal.

$Alternative\ Hypothesis\ (H_a)$: The population means of two independent levels: youth and not a youth, are not equal.

The results of Welch's t-test indicate that the mean percentage of strip searches for non-youth individuals (M=8.24 SD=24.36) is higher than the mean percentage of strip searches for youth (M=6.06, SD=21.54). The p-value (0.00001) is less than the significant level ($\alpha = 0.05$), with a 95% CI [-3.17, -1.12]. Therefore, there is strong evidence to reject the null hypothesis (H_0) that the mean percentage of strip searches is the same between youth and non-youth people.

2.2.1.4 Negative action and Percentage of Strip Searches

A Welch's t-test was carried out to determine if there is a difference in the percentage of strip searches between people who had a negative action at the time of arrest and people who did not have a negative action. The hypotheses are:

$Null\ Hypothesis\ (H_0)$: The population means of two independent levels: has negative action and no negative action, are equal.

$Alternative\ Hypothesis\ (H_a)$: The population means of two independent levels: has negative action and no negative action, are not equal.

The results of Welch's t-test show that the mean percentage of strip searches for people who had negative action (M=15.76 SD=29.40) is higher than the mean percentage of strip searches for people who did not have negative action (M=6.95, SD=23.47). The p-value (<0.00000) is less than the significant level ($\alpha = 0.05$), with a 95% CI [7.82, 9.82]. Thus, we have evidence to reject the null hypothesis that the mean percentage of strip searches is the same between

people who had negative action and people who did not have negative action at the time of arrest.

2.2.1.5 Age group and Percentage of Strip Searches

A Welch's t-test was performed to investigate whether the proportion of strip searches varied between individuals who were below the age of 35 at the time of arrest and those who were older than 35. The hypothesis is:

H_0: The population means of two independent levels: under 35 and above 35 years old, are equal.

H_a: The population means of two independent levels: under 35 and above 35 years old, are not equal.

The results of Welch's t-test show that the mean percentage of strip searches for people under 35 years old (M=9.02, SD=25.75) is greater than the mean percentage of strip searches for people over 35 years old (M=7.12, SD=22.44). The p-value (<0.00000) is less than the significant level ($\alpha = 0.05$), with a 95% CI [7.82, 9.82]. Therefore, we have evidence to reject the null hypothesis that the mean percentage of strip searches is the same between people who are under 35 years old and those who are older than 35.

2.2.1.6 T-test Conclusion

According to Welch's t-test findings presented above, there were statistically significant differences in the mean percentage of strip searches for different categories of perceived race, sex, negative action, youth status, and age groups.

3. Method

3.1 Dataset Description

The dataset used for this project was provided by the Toronto Police Service (Toronto Police Service, 2022).The dataset has 34,650 observations that are grouped by person ID, which is assumed to be the identification number assigned to those who have been arrested. It includes demographic information on each individual, such as their perceived race, sex, age group at the time of arrest, and youth status at the time of arrest. These attributes are categorical variables with at least two levels. Additionally, the dataset includes information about the arrest, including the number of strip searches, the strip search percentage (i.e., the average number of strip searches per person during 2020 and 2021), and any negative actions taken at the time of arrest, such as concealed items, violent behaviour, defensive or escape risk, and mental health

issues. Strip search counts the number of times each person was searched at the time of arrest without considering the total number of arrests, while strip search percentage represents the average number of strip searches per person during the two-year period. Negative action counts the number of inappropriate actions taken at the time of arrest for each person. To avoid any potential for misleading information, cases involving individuals with more than one perceived race and age group were removed from the dataset.

3.2 Research Methods and Assumptions

Section 2.2.1.6 concluded that there were significant differences in the mean percentage of strip searches across various levels categorized by youth status, negative action, age groups, race, and sex. However, since youth status and age groups both relate to age, only age groups will be considered in the study to avoid collinearity between variables.

Additionally, the sample size for the group with negative actions is significantly smaller, making it unrepresentative to accurately investigate the association between negative actions and the mean percentage of strip searches. Therefore, the focus of the research would be on examining the relationships between an individual's demographics such as race, sex and age groups, and the mean percentage of strip searches.

Figure 2 and t-test result showed the mean differences in percentage of strip searches between races and between sexes. To further investigate this, a two-way ANOVA was conducted to determine how the combination of sex and race affects an individual's percentage of strip searches.

Perceived race was treated as an 8-level categorical variable (East/Southeast Asian, White, Black, South Asian, Middle-Eastern, Unknown or Legacy, Latino, Indigenous) to examine the impact of each race on the outcome rather than only differentiating between white and non-white individuals. The sex variable was categorized as male or female.

In addition, a one-way ANOVA was performed to investigate how the percentage of strip searches is affected by an individual's age group. The age group is a categorical variable with 7 levels, including aged 17 years and under, aged 18 to 24 years, aged 25 to 34 years, aged 35 to 44 years, aged 45 to 54 years, aged 55 to 64 years, and aged 65 years and older.

To use ANOVA, the following assumptions were satisfied: (1) continuous outcome; (2) categorical explanatory variables; (3) independent errors; (4) normality assumption; and (5) constant variance. The same reasons explained in the t-test section supported the first three assumptions. The constant variance and normality of residuals assumptions were validated using Levene's test and Shapiro-Wilk test, respectively.

Due to the unequal sample size across groups, post-hoc tests were conducted using the Tukey-Kramer method to determine if the mean percentage of strip searches differs between specific levels of a variable. The assumptions for the post-hoc tests are the same as for ANOVA.

3.3 ANOVA tests

3.3.1 Research Question 1

The hypothesis are following:

> $Null\ Hypothesis\ 1$: The population means of two independent levels: male and female, are equal.
>
> $Alternative\ Hypothesis\ 1$: The population means of two independent levels: male and female, are not equal.
>
> $Null\ Hypothesis\ 2$:The population means of eight independent levels: East/Southeast Asian, White, Black, South Asian, Middle-Eastern, Unknown or Legacy, Latino, Indigenous, are equal.
>
> $Alternative\ Hypothesis\ 2$: The population means of eight independent levels: East/Southeast Asian, White, Black, South Asian, Middle-Eastern, Unknown or Legacy, Latino, Indigenous, are not equal.
>
> $Null\ Hypothesis\ 3$: The effect of sex on mean percentage of strip searches does not depend on the effect of perceived race, and vice versa.
>
> $Alternative\ Hypothesis\ 3$: There is an interaction effect between sex and perceived race on mean percentage of strip searches.

3.3.2 Research Question 2

The hypothesis are:

> $Null\ Hypothesis\ 1$: The population means of seven independent levels: aged 17 years and under, aged 18 to 24 years, aged 25 to 34 years, aged 35 to 44 years, aged 45 to 54 years, aged 55 to 64 years, and aged 65 years and older, are equal.

$Alternative\ Hypothesis\ 1$: The population means of seven independent levels: aged 17 years and under, aged 18 to 24 years, aged 25 to 34 years, aged 35 to 44 years, aged 45 to 54 years, aged 55 to 64 years, and aged 65 years and older, are not equal.

3.4 Post-hoc tests

3.4.1 Research Question 1

The hypothesis are:

$Null\ Hypothesis\ 1_{ij}$: There is no difference in mean percentage of strip search for $race_i$ and $race_j$, where i and j represent distinct races.

$Alternative\ Hypothesis\ 1_{ij}$: There is difference in mean percentage of strip search for $race_i$ and $race_j$, where i and j represent distinct races.

$Null\ Hypothesis\ 2_{ij}$: There is no difference in mean percentage of strip search for $race_i\ sex_a$ and $race_j\ sex_b$ where i and j represent race levels and a b represent sex levels.

$Alternative\ Hypothesis\ 2_{ij}$: There is difference in mean percentage of strip search for $race_i\ sex_a$ and $race_j\ sex_b$ where i and j represent race levels and a b represent sex levels.

Since the variable "sex" has only two levels, it can be determined whether there is a significant difference in the mean of the outcome between these levels based on the ANOVA results. Hence, we did not perform a Tukey-Kramer test for this variable.

3.4.2 Research Question 2

The hypothesis are:

$Null\ Hypothesis$: There is no difference in mean percentage of strip search for age_i and age_j where i and j represent distinct age groups.

$Alternative\ Hypothesis$: There is difference in mean percentage of strip search for age_i and age_j, where i and j represent distinct age groups.

4. Results/ Findings

4.1.1 Research Question #1 ANOVA

Since the dataset has unbalanced groups, we used type-III sums of squares to conduct an ANOVA.

Table 3: Two-way ANOVA for sex and race

	sum_sq	df	F	PR(>F)
Intercept	7.431182e+04	1.0	127.348682	1.754658e-29
C(Perceived_Race)	4.427517e+04	7.0	10.839237	9.877818e-14
C(Sex)	3.701695e+04	1.0	63.436212	1.706759e-15
C(Perceived_Race):C(Sex)	2.506953e+04	7.0	6.137403	3.428399e-07
Residual	2.020999e+07	34634.0	NaN	NaN

Table 1 displays that the p-values for perceived race and sex are both below the significance level of α=0.05, indicating that we have strong evidence to reject the null hypothesis that there is no difference in the mean percentage of strip searches for different groups of race and for different levels of sex. As a result, we can infer that perceived race and sex both have an impact on the percentage of strip searches conducted.

Most importantly, the interaction between perceived race and sex is significant (F=6.14, p=0.0000003). Therefore, we have evidence to reject the null hypothesis and conclude that the effect of a change in sex on the mean percentage of strip searches depends on the level of perceived race, and vice versa.

4.1.2 Interaction plots

Figure 6 illustrated the means of percentage of strip searches of perceived race and sex. Two lines are not parallel; therefore, the interaction effect is significant between perceived race and sex.

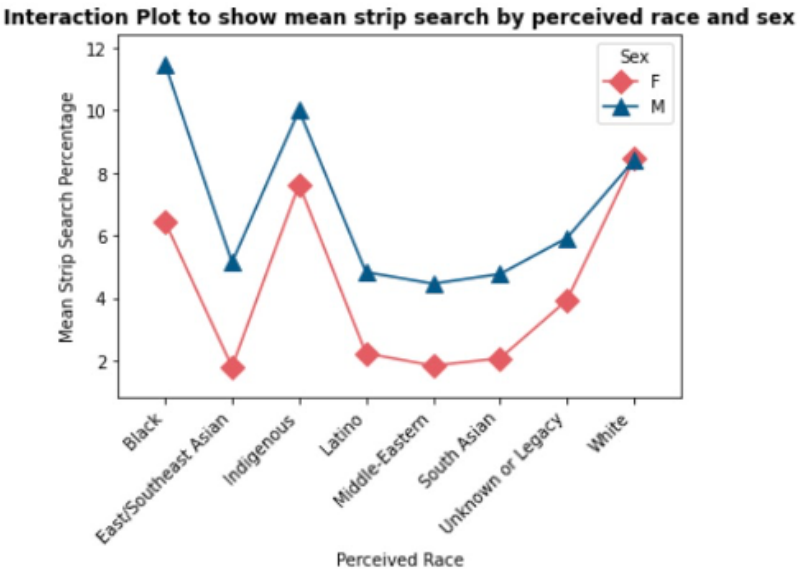

Figure 6: Interaction plot for sex and race

People whose race were perceived as black have higher mean percentage of strip searches, followed by indigenous, white, and other coloured groups. For all non-white racial groups, male individuals have higher mean percentage of strip searches than female individuals. But sex difference is smaller for people whose race were perceived as indigenous, and unknown or legacy. For white people, the mean percentage of strip searches are almost the same between male and female individuals.

4.1.3 Post-hoc test

Tukey-Krama test was conducted for the main effect, perceived race, and the results for pairs that are significant are shown in Table 4. Full results can be seen in Appendix i.

Table 4: Tukey comparison for perceived race

	group1	group2	Diff	Lower	Upper	p-value
0	East/Southeast Asian	White	3.9414	2.4337	5.4491	0.0010
1	East/Southeast Asian	Black	6.0568	4.4923	7.6214	0.0010
6	East/Southeast Asian	Indigenous	4.5852	1.1534	8.0169	0.0013
7	White	Black	2.1154	1.1410	3.0898	0.0010
8	White	South Asian	4.0936	2.4137	5.7735	0.0010
9	White	Middle–Eastern	4.3066	2.4915	6.1217	0.0010
10	White	Unknown or Legacy	2.8859	1.3762	4.3956	0.0010

11	White	Latino	4.0156	1.6291	6.4021	0.0010
13	Black	South Asian	6.2090	4.4779	7.9401	0.0010
14	Black	Middle–Eastern	6.4220	4.5594	8.2846	0.0010
15	Black	Unknown or Legacy	5.0013	3.4348	6.5677	0.0010
16	Black	Latino	6.1311	3.7082	8.5539	0.0010
21	South Asian	Indigenous	4.7373	1.2265	8.2482	0.0011
24	Middle–Eastern	Indigenous	4.9503	1.3728	8.5278	0.0010
26	Unknown or Legacy	Indigenous	3.5296	0.0970	6.9623	0.0388
27	Latino	Indigenous	4.6594	0.7608	8.5580	0.0071

From Table 4, the mean percentage of strip searches of black individuals is significantly different from every other group excluding the indigenous people. The mean outcome of indigenous people is significantly different from Latino, Middle-Eastern, South Asian, Ease/Southeast Asian, and unknown or Legacy. The mean outcome of white people is significantly different from every other group excluding indigenous people.

Table 5 shows the results of the Tukey-Kramer test conducted for the interaction effect of perceived race and sex, and only few significant pairs are presented. The complete findings are available in Appendix ii.

Each row of Table 5 demonstrates that the mean percentage of strip searches is significantly different between the two combinations listed. For instance, the first three rows indicate that the mean percentage of strip searches for East/Southeast Asian males is significantly different from that of White males and females, and Black males. The mean outcome of East/Southeast Asian females is significantly different from that of White males and females, Black males and females, males whose race is unknown or legacy, and indigenous males. The mean outcome of White males is significantly different from that of Black males, Latino males, and both males and females for South Asian, Middle-Eastern, and individuals with unknown or legacy races. The mean outcome of White females is significantly different from that of Black and Latino males, and both sexes for South Asian, Middle-Eastern, and individuals with unknown or legacy races. The mean outcome of Black males is significantly different from that of Black females, and both sexes for South Asian, Middle-Eastern, Latino, and individuals with unknown or legacy races. The mean outcome of both sexes of South Asian and Middle-Eastern, and females with unknown race are significantly different from indigenous males.

Table 5: Tukey test result for interaction effect (only significant combinations)

	group1	group2	Diff	Lower	Upper	p-value
1	(East/Southeast Asian, M)	(White, M)	3.267919	1.355113	5.180725	0.001000
2	(East/Southeast Asian, M)	(White, F)	3.308225	1.056193	5.560256	0.001000
3	(East/Southeast Asian, M)	(Black, M)	6.340222	4.367455	8.312990	0.001000
15	(East/Southeast Asian, F)	(White, M)	6.608257	3.033758	10.182756	0.001000
16	(East/Southeast Asian, F)	(White, F)	6.648563	2.881620	10.415506	0.001000
17	(East/Southeast Asian, F)	(Black, M)	9.680561	6.073619	13.287502	0.001000
18	(East/Southeast Asian, F)	(Black, F)	4.636103	0.642243	8.629964	0.006979
23	(East/Southeast Asian, F)	(Unknown or Legacy, M)	4.123011	0.232536	8.013485	0.025256
27	(East/Southeast Asian, F)	(Indigenous, M)	8.197649	2.451963	13.943335	0.001000
30	(White, M)	(Black, M)	3.072304	1.833068	4.311539	0.001000
32	(White, M)	(South Asian, M)	3.642321	1.551217	5.733424	0.001000
33	(White, M)	(South Asian, F)	6.336004	1.905516	10.766493	0.001000
34	(White, M)	(Middle-Eastern, M)	3.950397	1.729307	6.171487	0.001000
35	(White, M)	(Middle-Eastern, F)	6.569304	1.209335	11.929274	0.002816
36	(White, M)	(Unknown or Legacy, M)	2.485246	0.571749	4.398743	0.001000
37	(White, M)	(Unknown or Legacy, F)	4.470914	0.875201	8.066628	0.002175
38	(White, M)	(Latino, M)	3.584961	0.627009	6.542914	0.003429
42	(White, F)	(Black, M)	3.031998	1.314872	4.749124	0.001000
44	(White, F)	(South Asian, M)	3.682627	1.277313	6.087940	0.001000
45	(White, F)	(South Asian, F)	6.376310	1.789149	10.963471	0.001000
46	(White, F)	(Middle-Eastern, M)	3.990703	1.471564	6.509842	0.001000
47	(White, F)	(Middle-Eastern, F)	6.609610	1.119429	12.099791	0.003846
48	(White, F)	(Unknown or Legacy, M)	2.525552	0.272934	4.778170	0.011788
49	(White, F)	(Unknown or Legacy, F)	4.511220	0.724140	8.298300	0.004597
50	(White, F)	(Latino, M)	3.625267	0.437431	6.813104	0.009559
54	(Black, M)	(Black, F)	5.044457	2.874299	7.214615	0.001000
55	(Black, M)	(South Asian, M)	6.714624	4.568536	8.860713	0.001000
56	(Black, M)	(South Asian, F)	9.408308	4.951604	13.865012	0.001000
57	(Black, M)	(Middle-Eastern, M)	7.022701	4.749768	9.295634	0.001000
58	(Black, M)	(Middle-Eastern, F)	9.641608	4.259949	15.023267	0.001000
59	(Black, M)	(Unknown or Legacy, M)	5.557550	3.584113	7.530987	0.001000
60	(Black, M)	(Unknown or Legacy, F)	7.543218	3.915252	11.171184	0.001000

61	(Black, M)	(Latino, M)	6.657265	3.660189	9.654341	0.001000
62	(Black, M)	(Latino, F)	9.250098	2.718728	15.781468	0.001000
82	(South Asian, M)	(Indigenous, M)	5.231713	0.271009	10.192416	0.026861
90	(South Asian, F)	(Indigenous, M)	7.925396	1.611515	14.239277	0.001826
97	(Middle–Eastern, M)	(Indigenous, M)	5.539789	0.522907	10.556671	0.014640
103	(Middle–Eastern, F)	(Indigenous, M)	8.158696	1.161187	15.156206	0.006511

4.1.4 Two-way ANOVA assumptions

According to the results of the Shapiro-Wilk test, we have enough evidence to reject the null hypothesis and conclude that the residuals are not normally distributed (W=0.44, p=0.0).

To test the equal variance assumption, we conducted Levene's test. The results indicate that we have evidence to reject the null hypothesis that the variances of different groups are equal. Therefore, the assumption of constant variance is violated (statistic=97.03, p<0.01).

4.2.1 Research Question #2 ANOVA

Table 6: One-way ANOVA for age

	sum_sq	df	F	PR(>F)
C(Age)	1.001966e+05	6.0	28.505908	3.286485e-34
Residual	2.122498e+07	36231.0	NaN	NaN

Table 6 displays that the p-values for age is below the significance level of α=0.05, indicating that we have strong evidence to reject the null hypothesis that there is no difference in the mean percentage of strip searches for different groups of age. As a result, we can conclude that age have an impact on the percentage of strip searches performed.

4.2.2 Post-hoc test

Table 7: Tukey test results

group1	group2	meandiff	p-adj	lower	upper	reject
Aged 17 years and younger	Aged 18 to 24 years	4.3457	0.001	2.4306	6.2608	True
Aged 17 years and younger	Aged 25 to 34 years	3.1411	0.001	1.3226	4.9596	True
Aged 17 years and younger	Aged 35 to 44 years	2.591	0.001	0.7278	4.4542	True
Aged 17 years and younger	Aged 45 to 54 years	1.1576	0.5759	-0.803	3.1182	False
Aged 17 years and younger	Aged 55 to 64 years	-0.2721	0.9	-2.4128	1.8686	False
Aged 17 years and younger	Aged 65 years and older	-3.2345	0.0133	-6.0641	-0.4049	True
Aged 18 to 24 years	Aged 25 to 34 years	-1.2046	0.0283	-2.3365	-0.0727	True
Aged 18 to 24 years	Aged 35 to 44 years	-1.7547	0.001	-2.957	-0.5523	True
Aged 18 to 24 years	Aged 45 to 54 years	-3.1881	0.001	-4.5364	-1.8397	True
Aged 18 to 24 years	Aged 55 to 64 years	-4.6178	0.001	-6.2167	-3.0188	True
Aged 18 to 24 years	Aged 65 years and older	-7.5802	0.001	-10.0257	-5.1346	True
Aged 25 to 34 years	Aged 35 to 44 years	-0.5501	0.6839	-1.5917	0.4916	False
Aged 25 to 34 years	Aged 45 to 54 years	-1.9835	0.001	-3.1907	-0.7762	True
Aged 25 to 34 years	Aged 55 to 64 years	-3.4132	0.001	-4.8951	-1.9313	True
Aged 25 to 34 years	Aged 65 years and older	-6.3756	0.001	-8.7463	-4.0049	True
Aged 35 to 44 years	Aged 45 to 54 years	-1.4334	0.0158	-2.707	-0.1598	True
Aged 35 to 44 years	Aged 55 to 64 years	-2.8631	0.001	-4.3995	-1.3267	True
Aged 35 to 44 years	Aged 65 years and older	-5.8255	0.001	-8.2306	-3.4204	True
Aged 45 to 54 years	Aged 55 to 64 years	-1.4297	0.1416	-3.0829	0.2235	False
Aged 45 to 54 years	Aged 65 years and older	-4.3921	0.001	-6.8735	-1.9108	True
Aged 55 to 64 years	Aged 65 years and older	-2.9624	0.0154	-5.5884	-0.3365	True

As shown in table 7, the mean percentage of strip searches of people aged 18 to 24 years is significantly different from other age groups. The mean of people aged 17 years and younger is significantly different from people aged 18 to 24, people aged 25, people aged 35 to 44 and people aged 65 years and older. The mean outcome of people aged 25 to 34 years is significantly different from every other group excluding the group aged 35 to 44 years.

4.2.3 One-way ANOVA assumptions

FIgure 6 shows the residuals do not lie on the theoretical normal line, we see that the sample values are generally curve off along the line which indicates the violation of normality assumption.

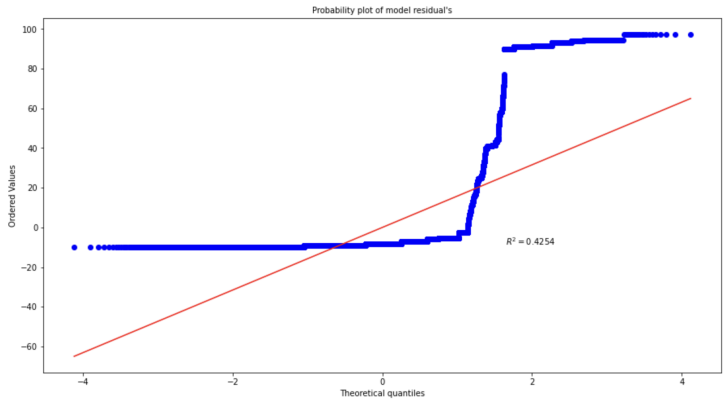

Figure 6: Normal QQ plot of model standardized residuals

To test the equal variance assumption, we conducted Levene's test. The results indicate that we have evidence to reject the null hypothesis that the variances of different age groups are equal. Therefore, the assumption of constant variance is violated (statistic=110.94, $p<0.01$).

5. Discussion

Our results presented both perceived race and sex have a significant impact on the percentage of strip searches conducted. This finding is supported by the p-values, which are below the significance level of $\alpha=0.05$, indicating strong evidence to reject the null hypothesis. Moreover, the interaction between perceived race and sex is also significant, suggesting that the effect of a change in sex on the mean percentage of strip searches is dependent on the level of perceived race, and vice versa.

Further analysis of the data in Table 1 reveals that individuals whose race was perceived as black have a significantly higher mean percentage of strip searches than other racial groups.
This finding is consistent with previous research that has shown that disproportionate use of strip searches against people of color (Lemke, 2022). Interestingly, the sex difference in the mean percentage of strip searches is smaller for people whose race was perceived as indigenous, unknown, or legacy. For white people, the mean percentage of strip searches is almost the same between male and female individuals.

The Tukey-Kramer test for the interaction effect of perceived race and sex shows that there are significant differences in the mean percentage of strip searches between different combinations of race and sex. For example, the mean percentage of strip searches for East/Southeast Asian males is significantly different from that of White males and females, and Black males. Similarly, the mean outcome of East/Southeast Asian females is significantly different from that of White males and females, Black males and females, males whose race is unknown or legacy, and indigenous males.

Finally, Table 6 shows that age also has a significant impact on the percentage of strip searches performed. This finding is supported by the p-value, which is below the significance level of $\alpha=0.05$. However, the report does not provide further analysis of the differences in mean percentage of strip searches across age groups. Overall, the findings suggest that perceived race, sex, and age all have a significant impact on the percentage of strip searches conducted. These results highlight the need for further research and policy changes to address racial and gender disparities in police practices, such as strip searches.

Our study has several limitations, the t-test's normality assumption is not met because our sample size exceeds 5000, which is the threshold at which the Shapiro-Wilk test is reliable. Thus, the p-value may be unreliable. The normality and equal variance assumptions for ANOVA are both violated, casting doubt on the accuracy of the ANOVA results. The smaller sample size of individuals who have been strip searched means that it may not be representative enough to

properly study the relationship between the mean percentage of strip searches and other variables. The use of a single dataset, which may not be representative of all law enforcement agencies in Canada. Lastly, our analysis did not consider other variables that may impact the likelihood of strip searches, such as the severity of the crime or the location of the arrest. Despite these limitations, our study sheds light on the disproportionate use of strip searches in law enforcement, which may have implications for civil liberties and human rights. Moving forward, it is important for law enforcement agencies to reevaluate their policies and procedures regarding strip searches and ensure that they are not influenced by implicit biases or stereotypes.

Future implications include using logistic regression to predict the probability of being subjected to strip searches based on age, race, and sex, considering the method of data collection. Instead of treating strip searches as a categorical variable, we can treat it as a binary variable. Additionally, future research should investigate other variables that may impact the likelihood of strip searches such as offensive actions towards police officers during arrest have an impact on the probability of being searched, and explore potential solutions to reduce the use of strip searches in law enforcement while still maintaining public safety.

6. Conclusion

Our analysis reveals significant disparities in the use of strip searches in law enforcement based on sex, race, and age group. Our findings indicate that individuals whose race is perceived as black are subject to a higher percentage of strip searches, while individuals whose race is perceived as white have a lower percentage of strip searches. Moreover, male individuals of all non-white racial groups have a higher percentage of strip searches than female individuals, but the sex difference is smaller for individuals whose race is perceived as indigenous or unknown/legacy. The results also show that age is a significant factor in the percentage of strip searches conducted. Furthermore, the interaction effect between perceived race and sex is significant, which suggests that the effect of a change in sex on the percentage of strip searches depends on the level of perceived race and vice versa. The Tukey-Kramer test results further illustrate the marked differences in the mean percentage of strip searches between various racial and sex combinations.

The findings of this study have important implications for law enforcement agencies and policymakers. Our results suggest that racial profiling and gender bias may play a role in strip search practices, which may lead to unjust treatment of individuals from certain racial and gender groups. Therefore, it is crucial for law enforcement agencies to examine their policies and procedures for strip searches and ensure that they are not discriminatory or biased.
In conclusion, this study sheds light on the complex relationship between perceived race, sex, age, and strip search practices, and underscores the need for further research and interventions to address potential biases in law enforcement practices.

References

Lemke, M. (2022, June 17). *Strip searches are ineffective, unnecessary and target racialized Canadians*. Phys.org. Retrieved February 26, 2023, from https://phys.org/news/2022-06-ineffective-unnecessary-racialized-canadians.html

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

Police powers: Strip searching. BSB Solicitors. (2022, August 19). Retrieved February 26, 2023, from https://www.bsbsolicitors.co.uk/blog/police-powers-strip-searching/

Prison Policy Initiative. (2019, May 14). *Policing women: Race and gender disparities in police stops, searches, and use of Force*. Prison Policy Initiative. Retrieved February 26, 2023, from https://www.prisonpolicy.org/blog/2019/05/14/policingwomen/

Susan Nembhard and Lily Robin. (2021, August). Racial and Ethnic Disparities throughout the Criminal Legal System A Result of Racist Policies and Discretionary Practices. Urban Institute. Retrieved February 26, 2023, from https://www.urban.org/sites/default/files/publication/104687/racial-and-ethnic-disparities-throughout-the-criminal-legal-system.pdf

Daniel M. Blonigen. (2021). Explaining the relationship between age and crime: Contributions from the developmental literature on personality. Clinical Psychology Review, Volume 30, Issue 1, 89-100.

Arrests and strip searches (RBDC-arr-TBL-001). Toronto Police Service Public Safety Data Portal. (2022, November 10). Retrieved February 23, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Appendix

i. Tukey test result for main effect: Perceived Race

	group1	group2	Diff	Lower	Upper	p-value
0	East/Southeast Asian	White	3.9414	2.4337	5.4491	0.0010
1	East/Southeast Asian	Black	6.0568	4.4923	7.6214	0.0010
2	East/Southeast Asian	South Asian	0.1522	−1.9264	2.2307	0.9000
3	East/Southeast Asian	Middle−Eastern	0.3652	−1.8241	2.5545	0.9000
4	East/Southeast Asian	Unknown or Legacy	1.0555	−0.8880	2.9991	0.6953
5	East/Southeast Asian	Latino	0.0742	−2.6079	2.7563	0.9000
6	East/Southeast Asian	Indigenous	4.5852	1.1534	8.0169	0.0013
7	White	Black	2.1154	1.1410	3.0898	0.0010
8	White	South Asian	4.0936	2.4137	5.7735	0.0010
9	White	Middle−Eastern	4.3066	2.4915	6.1217	0.0010
10	White	Unknown or Legacy	2.8859	1.3762	4.3956	0.0010
11	White	Latino	4.0156	1.6291	6.4021	0.0010
12	White	Indigenous	0.6437	−2.5623	3.8498	0.9000
13	Black	South Asian	6.2090	4.4779	7.9401	0.0010
14	Black	Middle−Eastern	6.4220	4.5594	8.2846	0.0010
15	Black	Unknown or Legacy	5.0013	3.4348	6.5677	0.0010
16	Black	Latino	6.1311	3.7082	8.5539	0.0010
17	Black	Indigenous	1.4717	−1.7615	4.7048	0.8546
18	South Asian	Middle−Eastern	0.2130	−2.0983	2.5243	0.9000
19	South Asian	Unknown or Legacy	1.2077	−0.8723	3.2877	0.6273
20	South Asian	Latino	0.0780	−2.7046	2.8605	0.9000
21	South Asian	Indigenous	4.7373	1.2265	8.2482	0.0011
22	Middle−Eastern	Unknown or Legacy	1.4207	−0.7699	3.6114	0.5042
23	Middle−Eastern	Latino	0.2910	−2.5753	3.1572	0.9000
24	Middle−Eastern	Indigenous	4.9503	1.3728	8.5278	0.0010
25	Unknown or Legacy	Latino	1.1298	−1.5535	3.8130	0.9000
26	Unknown or Legacy	Indigenous	3.5296	0.0970	6.9623	0.0388
27	Latino	Indigenous	4.6594	0.7608	8.5580	0.0071

ii.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black & F	Black & M	5.0445	0.001	2.8743	7.2146	True
Black & F	East/Southeast Asian & F	-4.6361	0.007	-8.63	-0.6422	True
Black & F	East/Southeast Asian & M	-1.2958	0.9	-3.9097	1.3182	False
Black & F	Indigenous & F	1.191	0.9	-4.7956	7.1776	False
Black & F	Indigenous & M	3.5615	0.5051	-1.4096	8.5327	False
Black & F	Latino & F	-4.2056	0.7128	-10.9584	2.5471	False
Black & F	Latino & M	-1.6128	0.9	-5.0658	1.8402	False
Black & F	Middle-Eastern & F	-4.5972	0.2758	-10.2454	1.0511	False
Black & F	Middle-Eastern & M	-1.9782	0.5531	-4.8255	0.8691	False
Black & F	South Asian & F	-4.3639	0.1204	-9.1391	0.4114	False
Black & F	South Asian & M	-1.6702	0.7457	-4.4173	1.0769	False
Black & F	Unknown or Legacy & F	-2.4988	0.7131	-6.5116	1.5141	False
Black & F	Unknown or Legacy & M	-0.5131	0.9	-3.1275	2.1014	False
Black & F	White & F	2.0125	0.2455	-0.4144	4.4393	False
Black & F	White & M	1.9722	0.1006	-0.1436	4.088	False
Black & M	East/Southeast Asian & F	-9.6806	0.001	-13.2875	-6.0736	True
Black & M	East/Southeast Asian & M	-6.3402	0.001	-8.313	-4.3675	True
Black & M	Indigenous & F	-3.8535	0.604	-9.5892	1.8822	False
Black & M	Indigenous & M	-1.4829	0.9	-6.1489	3.1831	False
Black & M	Latino & F	-9.2501	0.001	-15.7815	-2.7187	True
Black & M	Latino & M	-6.6573	0.001	-9.6543	-3.6602	True
Black & M	Middle-Eastern & F	-9.6416	0.001	-15.0233	-4.2599	True
Black & M	Middle-Eastern & M	-7.0227	0.001	-9.2956	-4.7498	True
Black & M	South Asian & F	-9.4083	0.001	-13.865	-4.9516	True
Black & M	South Asian & M	-6.7146	0.001	-8.8607	-4.5685	True
Black & M	Indigenous & F	-3.8535	0.604	-9.5892	1.8822	False
Black & M	Indigenous & M	-1.4829	0.9	-6.1489	3.1831	False
Black & M	Latino & F	-9.2501	0.001	-15.7815	-2.7187	True
Black & M	Latino & M	-6.6573	0.001	-9.6543	-3.6602	True
Black & M	Middle-Eastern & F	-9.6416	0.001	-15.0233	-4.2599	True
Black & M	Middle-Eastern & M	-7.0227	0.001	-9.2956	-4.7498	True
Black & M	South Asian & F	-9.4083	0.001	-13.865	-4.9516	True
Black & M	South Asian & M	-6.7146	0.001	-8.8607	-4.5685	True
Black & M	Unknown or Legacy & F	-7.5432	0.001	-11.1712	-3.9153	True
Black & M	Unknown or Legacy & M	-5.5575	0.001	-7.531	-3.5841	True
Black & M	White & F	-3.032	0.001	-4.7491	-1.3149	True
Black & M	White & M	-3.0723	0.001	-4.3115	-1.8331	True
East/Southeast Asian & F	East/Southeast Asian & M	3.3403	0.1944	-0.5498	7.2305	False
East/Southeast Asian & F	Indigenous & F	5.8271	0.1669	-0.8167	12.4709	False
East/Southeast Asian & F	Indigenous & M	8.1976	0.001	2.452	13.9433	True
East/Southeast Asian & F	Latino & F	0.4305	0.9	-6.9112	7.7721	False
East/Southeast Asian & F	Latino & M	3.0233	0.6031	-1.4738	7.5204	False
East/Southeast Asian & F	Middle-Eastern & F	0.039	0.9	-6.3017	6.3796	False

East/Southeast Asian & M	South Asian & F	-3.0681	0.6429	-7.7569	1.6208	False
East/Southeast Asian & M	South Asian & M	-0.3744	0.9	-2.9684	2.2196	False
East/Southeast Asian & M	Unknown or Legacy & F	-1.203	0.9	-5.1126	2.7066	False
East/Southeast Asian & M	Unknown or Legacy & M	0.7827	0.9	-1.6704	3.2357	False
East/Southeast Asian & M	White & F	3.3082	0.001	1.0562	5.5603	True
East/Southeast Asian & M	White & M	3.2679	0.001	1.3551	5.1807	True
Latino & M	White & M	3.585	0.0034	0.627	6.5429	True
Middle-Eastern & F	Middle-Eastern & M	2.6189	0.9	-3.0697	8.3075	False
Middle-Eastern & F	South Asian & F	0.2333	0.9	-6.6264	7.093	False
Middle-Eastern & F	South Asian & M	2.927	0.9	-2.7121	8.5661	False
Middle-Eastern & F	Unknown or Legacy & F	2.0984	0.9	-4.2542	8.451	False
Middle-Eastern & F	Unknown or Legacy & M	4.0841	0.4681	-1.4916	9.6597	False
Middle-Eastern & F	White & F	6.6096	0.0038	1.1194	12.0998	True
Middle-Eastern & F	White & M	6.5693	0.0028	1.2093	11.9293	True
Middle-Eastern & M	South Asian & F	-2.3856	0.9	-7.2084	2.4372	False
Middle-Eastern & M	South Asian & M	0.3081	0.9	-2.5209	3.1371	False
Middle-Eastern & M	Unknown or Legacy & F	-0.5205	0.9	-4.5899	3.5488	False
Middle-Eastern & M	Unknown or Legacy & M	1.4652	0.8908	-1.2352	4.1655	False
Middle-Eastern & M	White & F	3.9907	0.001	1.4716	6.5098	True
Middle-Eastern & M	White & M	3.9504	0.001	1.7293	6.1715	True
South Asian & F	South Asian & M	2.6937	0.8402	-2.0707	7.4581	False
South Asian & F	Unknown or Legacy & F	1.8651	0.9	-3.7256	7.4558	False
South Asian & F	Unknown or Legacy & M	3.8508	0.2611	-0.8384	8.5399	False
South Asian & F	White & F	6.3763	0.001	1.7891	10.9635	True
South Asian & F	White & M	6.336	0.001	1.9055	10.7665	True
South Asian & M	Unknown or Legacy & F	-0.8286	0.9	-4.8285	3.1713	False
South Asian & M	Unknown or Legacy & M	1.1571	0.9	-1.4374	3.7516	False
South Asian & M	White & F	3.6826	0.001	1.2773	6.0879	True
South Asian & M	White & M	3.6423	0.001	1.5512	5.7334	True
Unknown or Legacy & F	Unknown or Legacy & M	1.9857	0.9	-1.9243	5.8956	False
Unknown or Legacy & F	White & F	4.5112	0.0046	0.7241	8.2983	True
Unknown or Legacy & F	White & M	4.4709	0.0022	0.8752	8.0666	True
Unknown or Legacy & M	White & F	2.5256	0.0118	0.2729	4.7782	True
Unknown or Legacy & M	White & M	2.4852	0.001	0.5717	4.3987	True
White & F	White & M	-0.0403	0.9	-1.6882	1.6076	False