Strip Searches: An Evidence-Based Practice or Another Profiling Exercise?

Karrie Chou, Zekun Zhao
INF2178: Experimental Design for Data Science
Winter 2023

1 Introduction

1.1 Background and Literature Review

In today's social climate, concerns about rising criminal activity within communities across North America have motivated politicians to implement policies that effectively support evidence-based policing practices (Rinehart Kochel, 2011). However, the public continues to see policing policies and their applications as efforts to further entrench discriminatory values about race, age, and sex, among others, within the police force (Hetey & Eberhardt, 2018). This is evidenced by the explosion of academic and activist work in the 1980s to 1990s which studied "arrest disparities under the war on drugs, showing disproportionate distributions of arrest and incarceration among minorities and subsequently raising concerns about racial profiling" (Rinehart Kochel, 2011). The growing interest in studying police forces and their practices has evolved today into calls to action to defund municipal police forces (Fine & Del Toro, 2022; Lum et al., 2022).

Related demands point to demographic differences between individuals who are encountered by police which potentially affect their treatment in the interaction. For instance, Hetey & Eberhardt review the impact of how activists present the realities of racial disparity's impact in policing (2018), Starr reviews the presence of gender disparity's effects on criminal court decisions (2012), and Camplain et al. "estimate the association between race/ethnicity and drug- and alcohol-related arrest outcomes" in their study (2020). But how do these demographic differences manifest in the actions that police take to apprehend and arrest individuals who are suspected of involvement in crime, and to gather evidence in order to build a potential case against them?

A method of obtaining evidence following an arrest is through a strip search. Strip searching is a widespread practice throughout North American police institutions, and is so commonly conducted due to instances of blanket strip search policies being applied to police divisions across the United States (Ha, 2010). Strip searching is a controversial practice whose ethics, constitutionality, and impact have been discussed at length in the North American social context; Kirkup describes strip searching as "[representing] a larger series of coercive mechanisms used to dissuade members of society from breaking from traditionally accepted 'norms'" (2009).

Generally, strip searching is seen by academics as a practice which reinforces subjective norms about gender and appearance through its administration (Kirkup, 2009), brings racialized individuals such as African-Canadians into contact with police which can "[lead to] emotional harm and death" (Mullings et al., 2016), and encourages police to continually violate individuals' rights to human dignity through the reinforcement of racism and misogyny, among other discriminatory lenses (Jones & Sheehy, 2021). Further, numerous studies and legal reviews have ruled that many cases of police carrying out the strip search process were unconstitutional despite the practice being widespread throughout North America.

Mullings et al. highlight that in Canada, strip searching is one part of an overall policy direction that discriminates against Black people (2016). In their study, they explore two narratives of Black individuals who came in contact with police, one of whom was strip searched without reasonable grounds being expressed by the officers. In this narrative, the detainee was unjustly apprehended under the suspicion that she possessed drugs, and was required to undergo a strip search on an open street; although she was released when no drugs were found on her person, she was sexually harassed by a male witness outside of the officers who conducted the strip search shortly following her release (Mullings et al., 2016). She requested that an inquiry into the conduct of the three officers be launched; however, the panel in charge of administering the inquiry "cleared all three police officers of discreditable conduct charges" (Mullings et al., 2016). Through the application of critical race theory, Mullings et al. situate this narrative in the context of systemic anti-Black racism, and highlight it as an instance in which existing Canadian prejudice against Black individuals is expressed through the criminal justice system's policies and ruling precedents (2016). This perspective raises the question of whether or not the decision to book an arrested individual for a strip search is an evidence-based practice. Are officers and legal systems "only [upholding] strip searches when there is an individualized, reasonable suspicion that a detainee is concealing contraband" (Ha, 2010), or does strip searching enable individual biases to affect officers' judgements and treatments of arrested individuals?

1.2 Research Objective and Questions

This study aims to dissect how the Toronto Police Service (TPS), a North American police institution, uses strip searches in its assessment of arrested individuals. The Toronto Police Service defines a strip search as "a search conducted by a police officer on a person, which includes the removal of some or all clothing and a visual inspection of the body" (2022).

Having established the academic contexts in which strip searching has been studied and discussed, our study aims to answer the following research questions:

> **RQ1:** Do officers use their perceptions of demographic characteristics to decide whether or not to conduct a strip search following the arrest of an individual?

RQ2: Can an officer's decision to strip search an arrested individual be classified as more of a profiling exercise than an evidence-based practice, causing harm to marginalized communities?

2 Exploratory Data Analysis (EDA)

2.1 About the Dataset

The data that we used for this study comes from the "Arrests and Strip Searches (RBDC-ARR-TBL-001)" dataset published by TPS's Public Safety Data Portal on November 10, 2021 (Toronto Police Service, 2022). The dataset includes data related to all arrests actioned by TPS between January 2020 and December 2021. There are n = 65,276 observations in the dataset. There are 24 variables in the raw data.

Table 2.1.1. All variables in the raw dataset RBDC-ARR-TBL-001.

Variable Name	Variable Description
Arrest_Year	int – The year the arrest took place in (either 2020 or 2021).
Arrest_Month	str – The quarter the arrest took place in (either Jan-Mar, Apr-June, July-Sept, or Oct-Dec).
EventID	int – An identifier to specify details of the arrest event.
ArrestID	int – An identifier to specify details of the arrest event.
PersonID	int – An identifier to specify details of the arrest event.
Perceived_Race	str – The profiled race of the arrested individual.
Sex	str – The profiled sex of the arrested individual.
Age_group__at_arrest_	str – The age of the arrested individual, listed as a category (either under 17, 18-24, 25-34, 35-44, 45-54, 55-64, over 65)
Youth_at_arrest__under_18_years_	str – A dummy which indicates whether the arrested individual is classified as a youth (under 18 years).
ArrestLocDiv	str – An identifier to specify the location of the arrest event.
StripSearch	int – A dummy which indicates whether the arrested

	individual was strip searched.
Booked	int – A dummy which indicates whether the arrested individual was booked at a police facility within 24 hours of their arrest.
Occurrence_Category	str – The reason for arresting the individual.
Actions_at_arrest_Concealed_items	int – A dummy which indicates whether the arrested individual was uncooperative with the arresting officers by performing a certain action.
Actions_at_arrest_Combative_violent_or_spitter/biter	int – A dummy which indicates whether the arrested individual was uncooperative with the arresting officers by performing a certain action.
Actions_at_arrest_Resisted_defensive_or_escape_risk	int – A dummy which indicates whether the arrested individual was uncooperative with the arresting officers by performing a certain action.
Actions_at_arrest_Mental_instability_or_possibly_suicidal	int – A dummy which indicates whether the arrested individual was uncooperative with the arresting officers by performing a certain action.
Actions_at_arrest_Assaulted_officer	int – A dummy which indicates whether the arrested individual was uncooperative with the arresting officers by performing a certain action.
Actions_at_arrest_Cooperative	int – A dummy which indicates whether the arrested individual was completely cooperative with the arresting officers.
SearchReason_CauseInjury	int – A dummy which indicates the reason an arrested individual was booked for a strip search.
SearchReason_ArrestEscape	int – A dummy which indicates the reason an arrested individual was booked for a strip search.
SearchReason_PossessWeapons	int – A dummy which indicates the reason an arrested individual was booked for a strip search.
SearchReason_PossessEvidence	int – A dummy which indicates the reason an arrested individual was booked for a strip search.

ItemsFound	int – A dummy which indicates whether the strip search of an arrested individual resulted in items related to the crime event being found on their person.

Before we started our EDA, we transformed several features into predictors that could easily be used with a generalized linear model (GLM). We created three new variables:

- **Age_group_clean**, which takes all entries from the Age_group variable but combines the values "Aged 17 and under" with "Aged 17 years and under", and "Aged 65 and older" with "Aged 65 years and older" under one label each to reduce redundancies in mapping.
- **Time_series**, which aggregates the Arrest_Year and Arrest_Month variables in a discrete array, with numbers 1 through 8 representing the year and quarter of each arrest event in chronological order (e.g. a value of 1 in 'Time_series' means that the arrest took place in Q1 (January to March) of 2020, a value of 2 means that the arrest took place in Q2 (April to June) of 2020, and so on).
- **'Actions_at_arrest'**, which takes each of the six variables representing cooperative or uncooperative actions which affected the interaction between the arresting officer and arrested individual, and counts the number of uncooperative actions that the arrested individual took during the arrest.

Further, the 'Occurrence_Category' variable included multiple entries which represented more than one distinct reason for arresting an individual, as well as entries which were encoded differently but represented the same reason. We streamlined several of these entries; for instance, the values "Assault & Other crimes against persons" and "Assault" were counted as part of the same occurrence category, and the same thing happened with "Robbery & Theft" and "Robbery/Theft". In order to simplify subsequent analysis, we counted the number of data entries corresponding to each category and excluded observations whose 'Occurrence_Category' value did not fall into one of the 6 most popular categories:

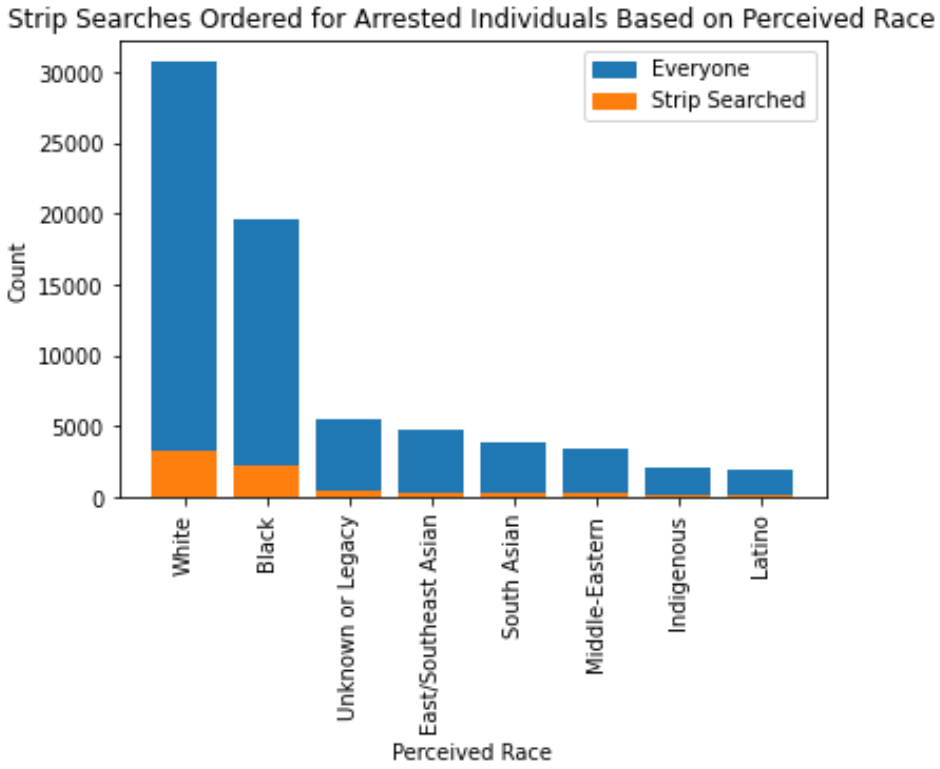
- Assault.
- Robbery and theft.
- Warrant.
- FTA/FTC/Compliance Check/Parollee.
- Police administration related.
- Drug related.

From the dataset, we also excluded any observations which had missing values in Perceived_Race and Age_group__at_arrest, as well as observations where Sex was neither Male nor Female. Following these exclusions, our dataset contained n = 64,615 entries that we used in our analysis.

2.2 Initial Visualizations

First, we looked at the perceived race of all individuals in our cleaned dataset, as well as the breakdown of perceived races for all individuals who were strip-searched.

Figure 2.2.1. Stacked bar chart depicting the number of arrest incidents involving an individual of a given perceived race (blue), as well as the number of arrest incidents which had a strip search associated with them that involved an individual of a given perceived race (orange).



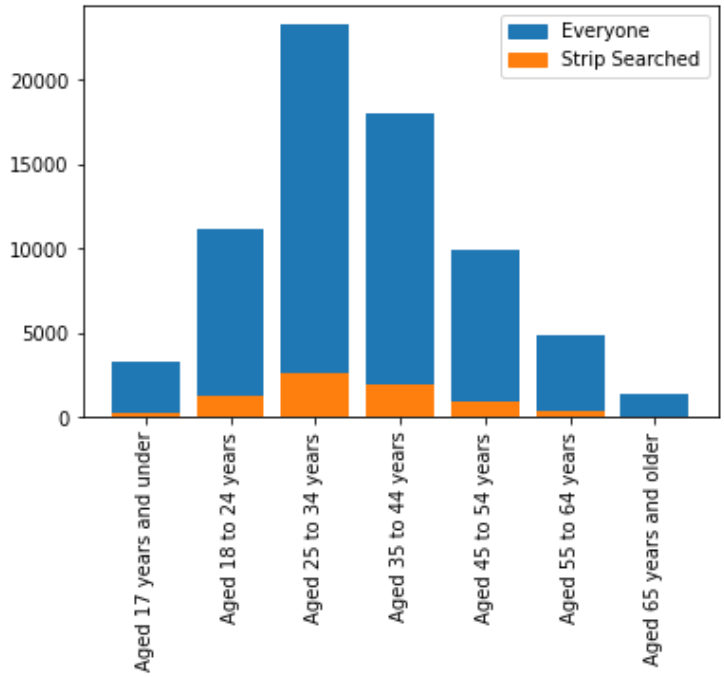
In the context of this dataset, we interpret legacy arrests as those involving individuals who had previously been arrested, and therefore already had entries in the dataset which made reference to perceived race. From the stacked bar chart, we see that more arrested individuals who were profiled as Black, White, or Indigenous by the arresting officers were subjected to a strip search, compared to individuals of other or unknown races, as well as legacy arrests. The dataset is also subject to over-representation of White and Black arrested individuals, as evidenced by the number of arrest events involving individuals who were perceived as such by officers. These findings are corroborated by Table 2.2.2.

Table 2.2.2. Proportions of arrested individuals who were strip searched, grouped by perceived race.

Perceived Race	Proportion (rounded to 3 decimals)
Black	0.132
White	0.122
Indigenous	0.112
Unknown or Legacy	0.099
South Asian	0.080
Middle-Eastern	0.077
East/Southeast Asian	0.074
Latino	0.071

These findings allow us to begin building a case for strip search decision making being partially influenced by the racial profiling habits of officers, since a POC subset of individuals documented in the dataset holds the highest proportion of strip searched individuals. However, we wanted to further assess similar breakdowns of strip searching in other demographic groups, such as age, in order to develop the foundational assumptions on which to build our final model.

Figure 2.2.3. Stacked bar chart depicting the number of arrest incidents involving an individual of a given age group (blue), as well as the number of arrest incidents which had a strip search associated with them that involved an individual of a given age group (orange).



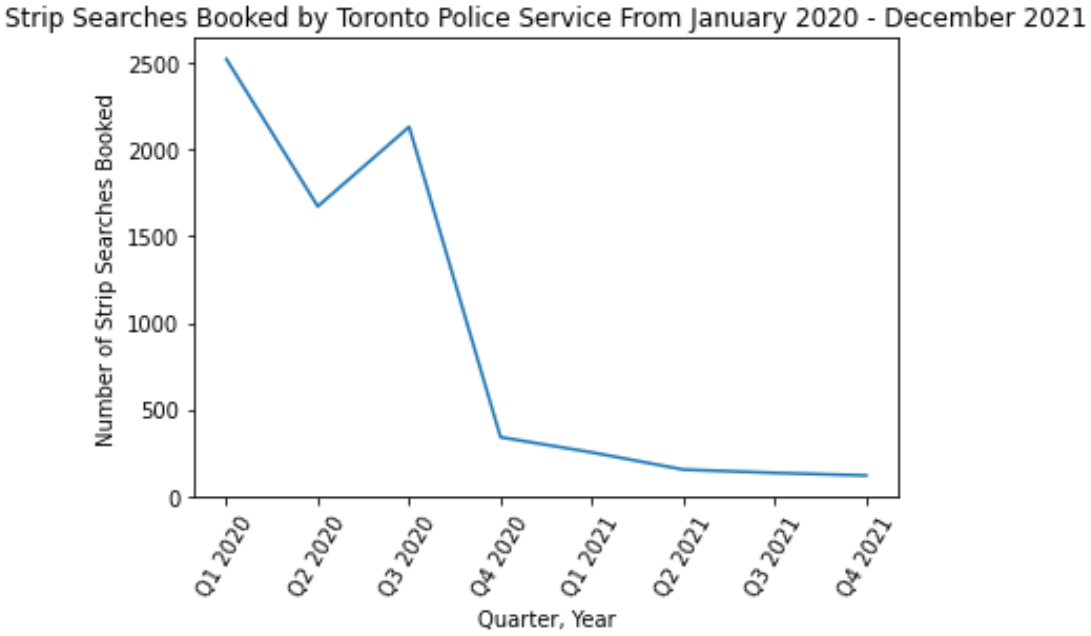
The largest group of arrested individuals in the dataset fall into the "Aged 25 to 34 years" category, matching the largest group of strip searched individuals. Noting that there are more arrest events in the dataset involving individuals who fall into this age range, we can partially attribute the nature of this particular distribution of strip search decisions to covariance with the size of each age group that is represented in the dataset.

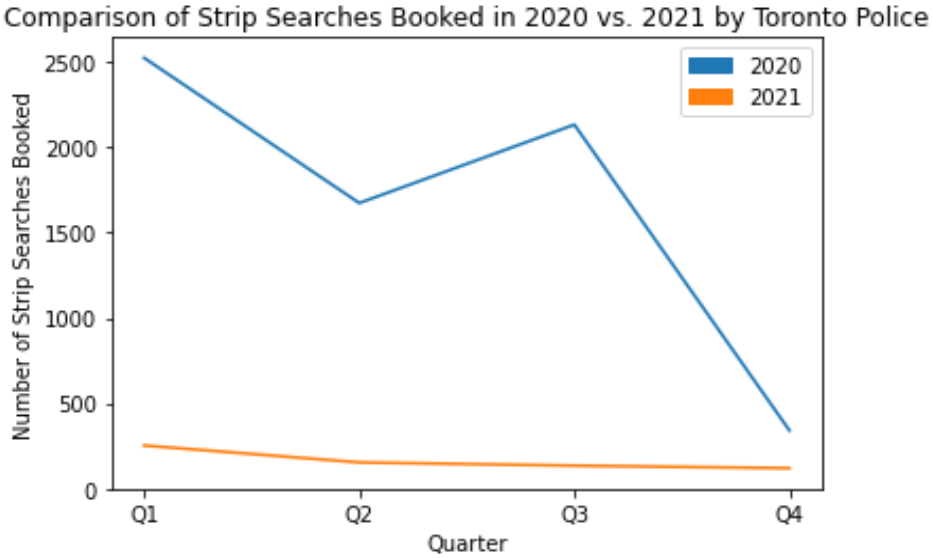
Table 2.2.4. Proportions of arrested individuals who were strip searched, grouped by age.

Age Group	Proportion (rounded to 3 decimals)
Aged 17 years and under	0.125
Aged 18 to 24 years	0.123
Aged 25 to 34 years	0.128
Aged 35 to 44 years	0.096
Aged 45 to 54 years	0.074
Aged 55 to 64 years	0.087
Aged 65 years and older	0.027

Beyond looking at the proportions of strip search decisions among individuals in different demographic groups, we also wanted to assess how the number of strip search decisions made by officers changed over the time during which the data was collected.

Figure 2.2.5. The number of strip searches booked by TPS officers from January 2020 - December 2021. Top: a single line graph depicting the number of strip searches booked over time. Bottom: a stacked line graph depicting the number of strip searches booked over 4 quarters in 2020 (blue) and 2021 (orange).

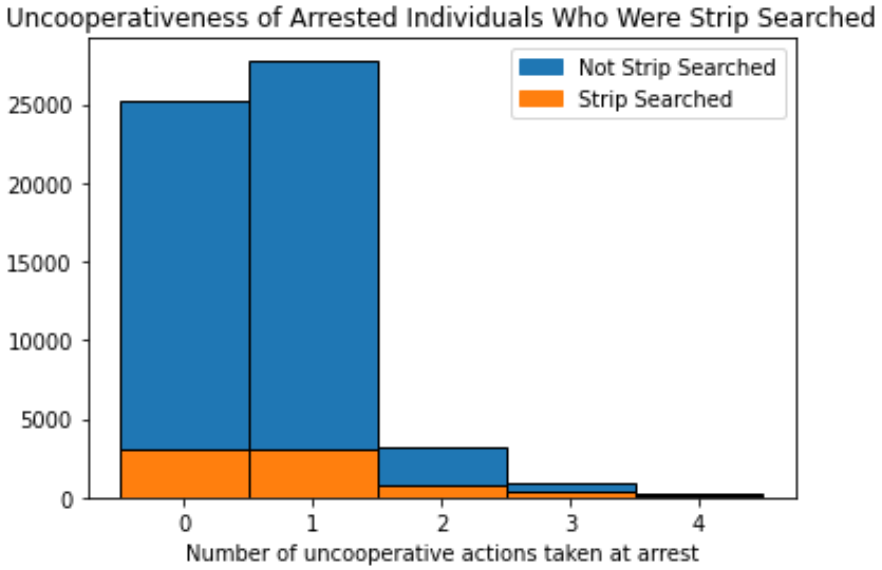




We see that in 2020, more strip searches were booked for arrested individuals in all quarters than in 2021.

Finally, we wanted to assess the uncooperativeness of arrested individuals who were involved in each event recorded in the dataset using our transformed 'Actions_at_arrest' variable.

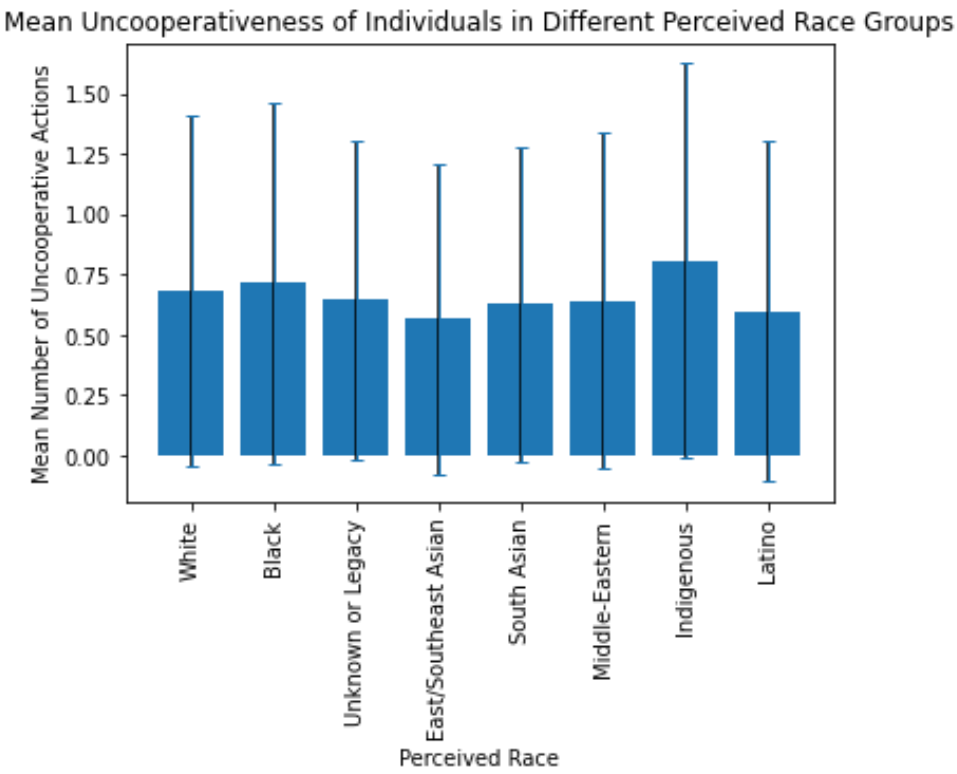
Figure 2.2.6. A stacked histogram of the distribution of 'Actions_at_arrest' for every arrested individual represented in the dataset (blue), and for strip searched individuals (orange).



Both histograms are right-skewed, indicating that more arrested individuals are cooperative with officers at the time of their arrest. However, the most concentrated part of both distributions is at the bin representing one distinct uncooperative action. Further, for both distributions, the median number of labeled uncooperative actions that took place for each arrest event in the dataset is 1. This means that on average, arrested individuals are still displaying some degree of resistance towards officer interactions at the time of their arrest.

We also assessed how uncooperativeness differed across the racial groups that were profiled by arresting officers.

Figure 2.2.7. A bar plot of the mean number of uncooperative actions taken by arrested individuals in each perceived race group represented in the dataset.



We see from this bar graph that at the time of their arrest, Indigenous and Black individuals display the highest resistance to officers. At this stage, we don't make assumptions about whether or not this resistance stems from inherent aversion towards police, the arrest event, or the perceived sequence of events following arrest.

2.3 Hypothesis Development and T-Tests

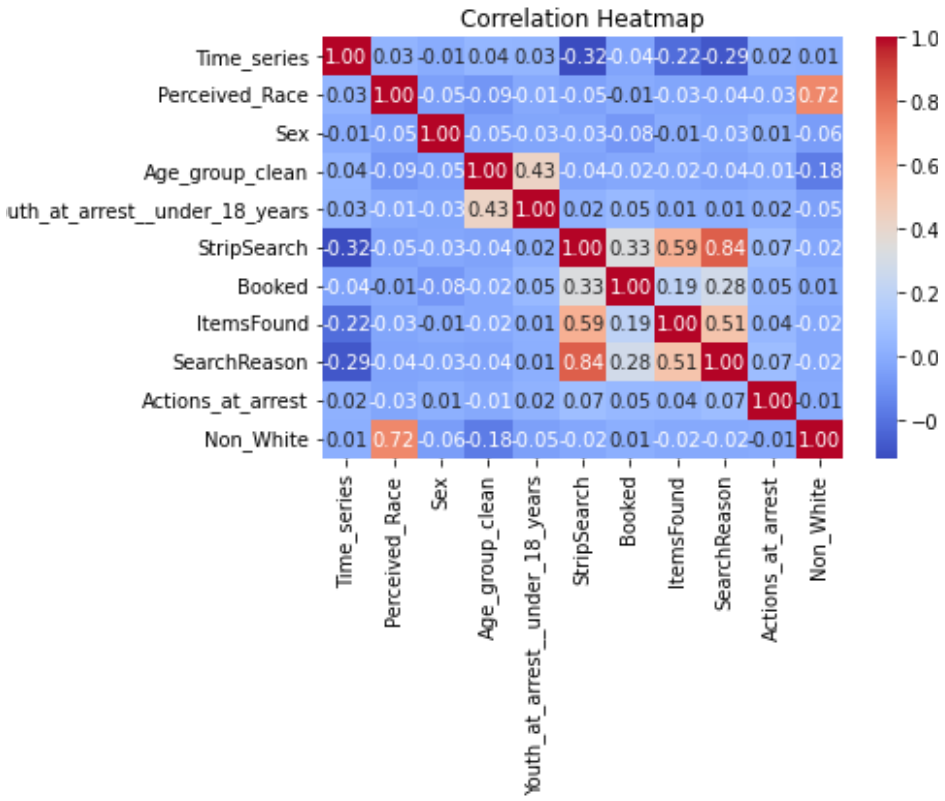
The visualization process of our EDA highlighted several variables of interest to use in a potential model that addresses our three research questions. The variables we considered include:

- Actions_at_arrest
- Perceived_Race
- Age_group_clean
- Youth_at_arrest__under_18_years
- Sex
- StripSearch
- SearchReason
- Time_series

To address our research questions, we were most interested in using 'StripSearch' and 'Actions_at_arrest' as dependent variables, and then having our other variables of interest

potentially serve as predictors following successful null hypothesis significance testing. To support our assumption that the variables we select for our final models are independent of each other to reduce the impact of collinearity, we created a correlation heatmap to assess the strength of the relationships between each pair of variables.

Figure 2.3.1. Correlation heatmap of model variables of interest.



Across many of the variable pairs, correlation was found to be close to 0, meaning that we can safely make the assumption that any pair of variables we pick to be predictors in our final models are independent of each other.

To test whether there are significant differences between the behaviours of arrested individuals in different perceived race groups, age groups, and sex groups, we conducted multiple Welch's two-tailed t-tests with significance level $\alpha = 0.05$ for our two dependent variables of interest. The Welch's t-test does not make the assumption that the two groups are drawn from the same population distribution, which works for our purposes as we are comparing groups with inherent differences in their demographic characteristics.

Table 2.3.2. Welch's t-test results for significance testing of the difference in uncooperativeness between arrested individuals who were strip searched and those who were not strip searched.

H_0	$\mu_{Actions\_at\_arrest,\ Strip\ Searched} = \mu_{Actions\_at\_arrest,\ Not\ Strip\ Searched}$
H_a	$\mu_{Actions\_at\_arrest,\ Strip\ Searched} \neq \mu_{Actions\_at\_arrest,\ Not\ Strip\ Searched}$
test statistic	14.998
p-value	3.312×10^{-50}

Table 2.3.3. Welch's t-test results for significance testing of mean uncooperativeness in the subset of individuals who were not strip searched. T-testing was done across 3 different demographic divisions based on 'Perceived_Race', 'Youth_at_arrest__under_18_years_', and 'Sex'.

	White vs. Non-White	**Youth vs. Adult**	**Female vs. Male**
H_0	$\mu_{Actions\_at\_arrest,\ White} = \mu_{Actions\_at\_arrest,\ non\text{-}White}$	$\mu_{Actions\_at\_arrest,\ Youth} = \mu_{Actions\_at\_arrest,\ Adult}$	$\mu_{Actions\_at\_arrest,\ Female} = \mu_{Actions\_at\_arrest,\ Male}$
H_a	$\mu_{Actions\_at\_arrest,\ White} \neq \mu_{Actions\_at\_arrest,\ non\text{-}White}$	$\mu_{Actions\_at\_arrest,\ Youth} \neq \mu_{Actions\_at\_arrest,\ Adult}$	$\mu_{Actions\_at\_arrest,\ Female} \neq \mu_{Actions\_at\_arrest,\ Male}$
test statistic	2.947	5.42	4.344
p-value	0.003	6.345×10^{-8}	1.409×10^{-5}

Table 2.3.4. Welch's t-test results for significance testing of mean uncooperativeness in the subset of individuals who were strip searched. T-testing was done across 3 different demographic divisions based on 'Perceived_Race', 'Youth_at_arrest__under_18_years_', and 'Sex'.

	White vs. Non-White	**Youth vs. Adult**	**Female vs. Male**
H_0	$\mu_{Actions\_at\_arrest,\ White} = \mu_{Actions\_at\_arrest,\ non\text{-}White}$	$\mu_{Actions\_at\_arrest,\ Youth} = \mu_{Actions\_at\_arrest,\ Adult}$	$\mu_{Actions\_at\_arrest,\ Female} = \mu_{Actions\_at\_arrest,\ Male}$
H_a	$\mu_{Actions\_at\_arrest,\ White} \neq \mu_{Actions\_at\_arrest,\ non\text{-}White}$	$\mu_{Actions\_at\_arrest,\ Youth} \neq \mu_{Actions\_at\_arrest,\ Adult}$	$\mu_{Actions\_at\_arrest,\ Female} \neq \mu_{Actions\_at\_arrest,\ Male}$
test statistic	-3.089	3.498	-0.943
p-value	0.002	0.0005	0.346

Table 2.3.2 shows there is a significant difference in the degree of uncooperativeness between arrested individuals who were strip searched and those who were not. This pushed us to conduct Welch's t-tests for different subsets of our data, splitting it between those who were not strip searched and those who were.

From Table 2.3.3, we infer that in situations where the strip search decision has not been made, 1) arrested individuals who are perceived as White are more likely to be uncooperative than those who are perceived as non-White, 2) youth are more uncooperative than adults, and 3) arrested individuals who identify as female are more uncooperative than those who identify as male.

From Table 2.3.4, we observe that among arrested individuals who were strip searched, the difference in mean uncooperativeness between groups that were defined by their perceived race and age were statistically significant. This is an indicator that both of these predictors would explain some of the differences in uncooperativeness that we see among strip searched individuals. Both perceived race and age group divisions were also statistically significant in explaining differences in uncooperativeness between individuals who were not strip searched. However, group division by sex was only statistically significant in explaining differences in uncooperativeness among individuals who were not strip searched.

Based on the results of our null hypothesis significance testing, we decided to use 'Perceived_Race', 'Age_group_clean', and 'Sex' as our predictors in the subsequent analysis.

3 Research Design and Methodology

3.1 One-Way ANOVA Analysis

To address RQ1, we first conducted four one-way ANOVA tests. The first test compared the means of strip searches conducted on individuals of different racial groups, specifically white, black, and Asian individuals, to determine if there were any statistically significant differences between them. The second test compared the means of strip searches conducted on male and female individuals to determine if there were any significant gender-based differences.

The third test examined whether there was a significant effect of time series on the likelihood of being selected for a strip search, while the fourth test examined whether there was a significant effect of search reason on the likelihood of being selected for a strip search.

3.2 Two-Way ANOVA Analysis

To further investigate the potential impact of race and gender on strip search decisions and address RQ2, we conducted four two-way ANOVA tests.

The first test used Coop_at_arrest_count, a variable which classified arrest events by whether or not Actions_at_arrest was equal to 0, as the dependent variable, and Perceived_Race and Sex as the independent variables. This test aimed to determine if either perceived race or sex differences contributed to any significant differences in whether or not an individual was cooperative with officers.

In the second two-way ANOVA test, we examined the potential impact of perceived race and sex on the number of strip searches conducted following an arrest.

For the third two-way ANOVA test, we investigated the potential impact of perceived race and youth or adults on the number of uncooperative behaviors displayed at the time of arrest. We grouped the data by Perceived_Race and Youth_at_arrest__under_18_years_, and counted the number of occurrences where Actions_at_arrest was equal to 1, which we defined as the measure of uncooperative behavior.

The fourth two-way ANOVA was performed to investigate the effects of both time series and search reason on the strip search count.

3.3 Post-Hoc Testing

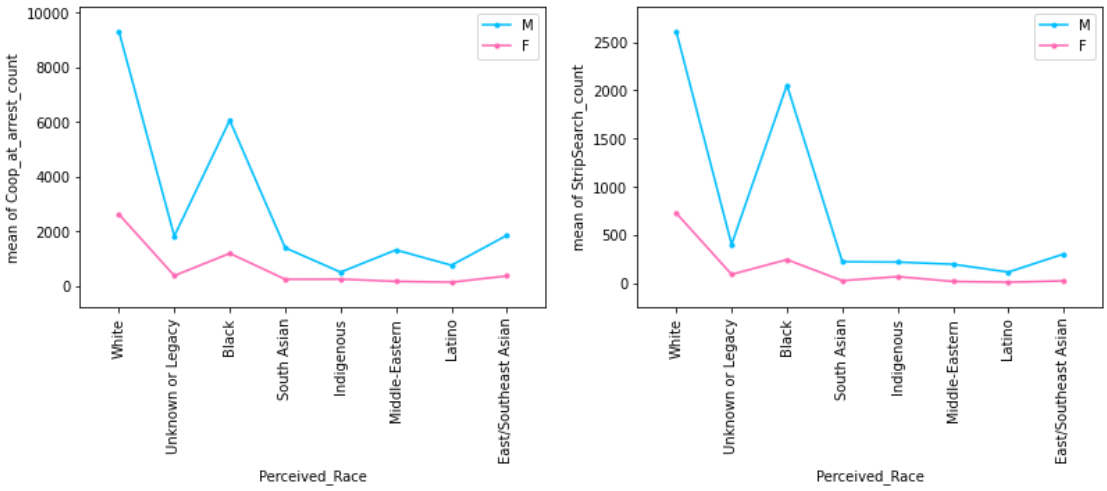
To verify the ANOVA test, we conducted the first post hoc test using Tukey's HSD to compare the mean difference in cooperative behavior count between different perceived race groups. The results suggest that there were no significant differences between any of the groups ($p > .05$ for all pairwise comparisons). Therefore, we cannot reject the null hypothesis that there is no significant difference in cooperative behavior count between different perceived race groups. Similarly, we also cannot reject the null hypothesis that there is no significant difference in cooperative behavior count between males and females. (See Appendix)

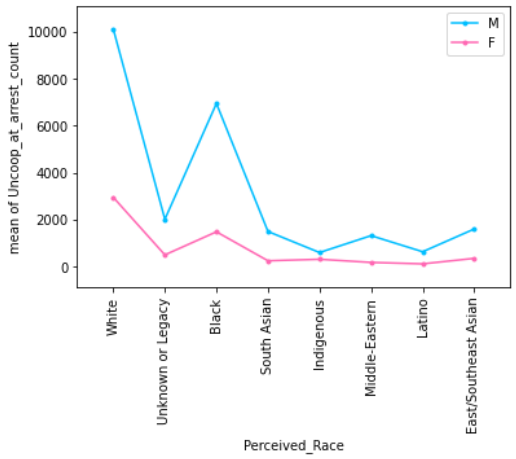
Based on the second Tukey's HSD test results and table, we can conclude that there is no significant difference in the mean StripSearch_count between any of the groups at a significance level of 0.05. All pairwise comparisons between groups have p-values greater than 0.05, indicating that we cannot reject the null hypothesis of equal means for any comparison, which indicates that we cannot reject the null hypothesis that there is no significant difference in strip search count between different races. Similarly, we also cannot reject the null hypothesis that there is no significant difference in strip search count between males and females. (See Appendix)

The third post hoc test compares the mean difference in the "Uncoop_at_arrest_count" variable between different levels of the "Perceived_Race" variable and the "Youth_at_arrest__under_18_years" variable. The results show that there are no significant differences in "Uncoop_at_arrest_count" between any of the racial groups. However, the test does reveal a significant difference in "Uncoop_at_arrest_count" based on whether the arrested individual is under 18 years old or not. This means that being a teenager at the time of arrest has a significant influence on the level of uncooperative behavior displayed during the arrest.

The fourth post hoc test compares the mean difference in strip search counts over time and search reasons. Tukey HSD test results show that there is a significant difference in mean StripSearch counts between several pairs of Time_series groups. Specifically, groups 1 and 4, 1 and 5, 1 and 6, 1 and 7, and 1 and 8 are significantly different from each other, which indicates that police's attitudes towards strip-search did change. However, there are no significant differences in mean StripSearch counts between any pairs of SearchReason groups.

Figure 3.3.1. Interaction plots for each of the four two-way ANOVAs.





There is a lack of intersections between both lines in the first three interaction plots, and in each plot, the pairs of lines had overall similar shapes. To reconcile the conclusions between the first and third interaction plots (for Coop_at_arrest_count and Uncoop_at_arrest_count), we must recognize that the dataset contained overwhelmingly more arrest events involving male-profiled individuals than female-profiled individuals. Concretely, we can infer that in all of the perceived race groups that were documented by TPS, arrested individuals who identified as female were less likely to be booked for a strip search.

The fourth interaction plot involves multiple intersections between each level of the number of search reasons that were documented. However, the shared decrease over time for all levels of search reason numbers serves as further confirmation that officers are booking less strip searches overall, and that the reason for booking an individual for a strip search is not entirely dependent on facts about the arrest event.

4 Results

4.1 One-Way ANOVA Analysis Interpretation

Both first and second ANOVA tests yielded p-values greater than 0.05; therefore, we cannot reject the null hypothesis. In other words, there is no evidence to suggest that individuals of different races or genders are treated unequally when it comes to strip searches. These findings suggest that police officers are not explicitly biased when making strip search decisions based on race or gender. However, it is important to note that this does not necessarily rule out the potential for implicit bias or other forms of discrimination.

The p-values for the third test were all less than 0.05, indicating that we can reject the null hypothesis and conclude that the police attitude towards strip searches has indeed changed over time. However, the p-value for the fourth test was greater than 0.05, indicating that we cannot reject the null hypothesis and conclude that search reason does not have a significant impact on the likelihood of being selected for a strip search.

To assess the accuracy of our one-way ANOVA analysis, we used Tukey's HSD post-hoc test. The results of the Tukey test also indicated that we cannot reject the null hypothesis for any of the racial or gender groups, which is consistent with the results of our one-way ANOVA analysis. This provides further evidence to support the conclusion that individuals of different races or genders are not treated unequally when it comes to strip searches.

4.2 Two-Way ANOVA Analysis Interpretation

Table 4.2.1. Output of the first two-way ANOVA test, rounded to 3 decimals.

	sum_sq	df	F	PR(>F)
Perceived_Race	2.568×10^7	1.0	7.724	0.016
Sex	1.957×10^7	1.0	5.883	0.032
Perceived_Race:Sex	8.269×10^6	1.0	2.487	0.141
Residual	3.991×10^7	12.0	NaN	NaN

The ANOVA table for this test showed that Perceived_Race had a significant main effect on Coop_at_arrest_count (F = 7.72, p = .017). Additionally, Sex had a significant main effect on Coop_at_arrest_count(F = 5.88, p = .032). However, the interaction effect between Perceived_Race and Sex was not significant (F = 2.49, p = .141). These results suggest that both perceived race and sex can have a significant impact on the number of actions taken at the time of arrest. The lack of a significant interaction effect suggests that the effect of sex is consistent across different racial groups. However, it is important to note that the effect size of these differences may be small and that there may be other non-demographic factors that contribute to how cooperative an individual would be while being arrested.

Table 4.2.2. Output of the second two-way ANOVA test, rounded to 3 decimals.

	sum_sq	df	F	PR(>F)

Perceived_Race	2.526×10^6	1.0	8.042	0.015
Sex	1.510×10^6	1.0	4.806	0.049
Perceived_Race:Sex	8.814×10^5	1.0	2.806	0.120
Residual	3.769×10^6	12.0	NaN	NaN

The table reveals that perceived race had a significant main effect on the number of strip searches conducted ($F = 8.04$, $p = .015$), as did sex ($F = 4.81$, $p = .049$). However, the interaction effect between perceived race and sex was not significant ($F = 2.81$, $p = .120$). These results suggest that both perceived race and sex may play a role in strip search decisions, but that the effect of race is slightly stronger than that of sex. Again, it is important to note that the effect size of these differences may be small and that there may be other non-demographic factors that contribute to the decision to conduct a strip search.

Table 4.2.3. Output of the third two-way ANOVA test, rounded to 3 decimals.

	sum_sq	**df**	**F**	**PR(>F)**
Perceived_Race	3.487×10^7	1.0	6.570	0.025
Sex	4.873×10^7	1.0	9.183	0.010
Perceived_Race:Sex	2.919×10^7	1.0	5.501	0.037
Residual	6.368×10^7	12.0	NaN	NaN

The ANOVA output shows both perceived race and sex had a significant main effect on the number of uncooperative behaviors ($p < .05$). These findings suggest that perceived race and sex may both influence the level of uncooperative behaviors displayed at the time of arrest.

Table 4.2.4. Output of the fourth two-way ANOVA test, rounded to 3 decimals.

	sum_sq	**df**	**F**	**PR(>F)**
Time_series	1.110×10^6	1.0	54.857	9.686×10^{-9}
SearchReason	2.450×10^2	1.0	0.012	0.913
Time_series:SearchReason	7.991×10^3	1.0	0.395	0.534

Residual	7.283×10^7	36.0	NaN	NaN

The results showed that the interaction between time series and search reason was not significant ($p > 0.05$), indicating that there was no significant combined effect of these two factors on the strip search count. However, the main effect of Time_series was significant ($p < 0.05$), indicating that the police attitude towards strip searches changed over time. The main effect of search reason was not significant ($p > 0.05$), indicating that search reason did not have a significant impact on the strip search count.

5 Discussion

Looking at our EDA results, ANOVA tests, post-hoc tests, and interaction plots, it appears that the attitude of Toronto police officers towards strip searches has changed between January 2020 and December 2021. Police officers have decided to perform fewer strip searches overall, but it is unclear whether this is due to public pressure or internal reform. Further, the reason for this shift in attitude remains unknown.

Although police officers have adopted a more cautious approach towards strip searches from early 2020 to the end of 2021, it does not necessarily mean that they have been conducting strip searches based solely on reasoning about the facts of the crime that took place which resulted in an arrest. Results from the two-way ANOVA test and post-hoc test indicate that the variable SearchReason does not have a significant effect on the number of strip searches that were conducted in a given time period. This suggests that police officers may not be entirely basing their decision to conduct a strip search on evidence, and that some degree of personal judgment is used when making the decision to conduct a strip search.

Notably, from our initial t-tests, we found that youth were more uncooperative than adults during arrest regardless of whether they were exposed to a strip search decision. Moreover, youth also underwent more strip searches than adults. Through further t-tests, we also found that police officers did find more items on strip searched youth related to the crimes they were arrested for. This finding raises concerns about heightened rates of youth crime, which could factor into how police profile individuals when determining who to arrest and who to book for a strip search or other evidence collection practices.

Moreover, the results of our t-tests indicate that police officers are more likely to conduct strip searches on individuals who exhibit uncooperative behavior during an arrest. This finding raises important questions about the factors that influence police decisions to conduct strip searches, and whether these decisions are being made in a fair and unbiased manner. Perhaps officers' judgements are impaired by recency bias following an arrest, and they subconsciously link acts

of uncooperativeness during the arrest event with the implication that the arrested individual is hiding something, which they use to justify their decision to book a strip search.

5.1 Limitations

The principal limitations with our analysis methodology stems from the fact that the raw data we used was entirely categorical. The continuous variables we created were proxies of phenomena we wanted to assess, but they don't necessarily capture the entire scope of how such concepts manifest in the real world. One such phenomenon was uncooperativeness, which we measured using 'Actions_at_arrest'. The raw data defined uncooperativeness by using documentation of very specific actions the arrested individual took or perceptions the officer had of them. Surely, uncooperativeness with an officer can be expressed in many different ways outside of what was documented in the dataset.

Another limitation with the dataset was in its inherent construction. Each observation represented an arrest event that took place between January 2020 and December 2021, regardless of if the arrested individual was a repeat offender. This complicates our analysis, especially because the dataset was generated using information that officers on duty report about the arrested individual. The self-reported information for one arrest event involving a given individual often conflicted with that in other arrest events involving that same individual, making it difficult to consolidate data and determine the "true" value of any one predictor, especially for Perceived_Race, Age_group_clean, and Sex, our three final predictor variables. We decided not to filter the data to only include arrests with unique first-time offenders because we were interested in using StripSearch as a dependent variable, but this was done with the understanding that the distributions of our predictors may be skewed in favour of repeat offenders as a result.

6 Conclusion

RQ1: Do officers use their perceptions of demographic characteristics to decide whether or not to conduct a strip search following the arrest of an individual?

Based on the information provided, it is difficult to conclusively determine if police officers use their perceptions of demographic characteristics to decide whether or not to conduct a strip search. However, we found that SearchReason did not have a significant influence on the number of strip searches. Additionally, we found that police conducted more strip searches on youth and on individuals who were uncooperative during the arrest process. Further investigation may be needed to fully answer the research question.

RQ2: Can an officer's decision to strip search an arrested individual be classified as more of a profiling exercise than an evidence-based practice, causing harm to marginalized communities?

Since SearchReason did not have a statistically significant influence on the number of strip searches that were conducted by TPS in a given time period, an officer's decision on whether or not to strip search an individual might not be solely based on hard evidence about facts from the crime that led to the arrest event. Therefore, there may be a profiling element involved in the process of determining whether or not a strip search should be booked. Furthermore, the fact that police officers are more likely to conduct strip searches on uncooperative individuals and youth raises concerns about potential harm to marginalized communities, and opens up room for us to conduct research on how these harms influence the interactions between police officers and individuals from these marginalized communities.

7 References

Camplain, R., Camplain, C., Trotter, R. T., Pro, G., Sabo, S., Eaves, E., Peoples, M., & Baldwin, J. A. (2020). Racial/Ethnic Differences in Drug- and Alcohol-Related Arrest Outcomes in a Southwest County From 2009 to 2018. *American Journal of Public Health*, *110*(S1), S85–S92. https://doi.org/10.2105/AJPH.2019.305409

Ha, D. (2010). Blanket Policies for Strip Searching Pretrial Detainees: An Interdisciplinary Argument for Reasonableness Note. *Fordham Law Review*, *79*(6), 2721–2760.

Hetey, R. C., & Eberhardt, J. L. (2018). The Numbers Don't Speak for Themselves: Racial Disparities and the Persistence of Inequality in the Criminal Justice System. *Current Directions in Psychological Science*, *27*(3), 183–187. https://doi.org/10.1177/0963721418763931

Jones, D., & Sheehy, E. (2021). R v Desjourdy: A Narrative Of White Innocence And Racialized Danger. *The Canadian Bar Review*, *99*(3), Article 3. https://cbr.cba.org/index.php/cbr/article/view/4714

Kirkup, K. (2009). Indocile Bodies: Gender Identity and Strip Searches in Canadian Criminal Law*. *Canadian Journal of Law and Society / La Revue Canadienne Droit et Société*, *24*(1), 107–125. https://doi.org/10.1017/S0829320100009790

Lum, C., Koper, C. S., & Wu, X. (2022). Can We Really Defund the Police? A Nine-Agency Study of Police Response to Calls for Service. *Police Quarterly*, *25*(3), 255–280. https://doi.org/10.1177/10986111211035002

Mullings, D. V., Morgan, A., & Quelleng, H. K. (2016). Canada the Great White North where Anti-Black Racism Thrives: Kicking Down the Doors and Exposing the Realities. *Phylon (1960-)*, *53*(1), 20–41.

Rinehart Kochel, T. (2011). Constructing Hot Spots Policing: Unexamined Consequences for Disadvantaged Populations and for Police Legitimacy. *Criminal Justice Policy Review*, *22*(3), 350–374. https://doi.org/10.1177/0887403410376233

Starr, S. B. (2012). *Estimating Gender Disparities in Federal Criminal Cases* (SSRN Scholarly Paper No. 2144002). https://doi.org/10.2139/ssrn.2144002

Toronto Police Service (2022). *Arrests and Strip Searches (RBDC-ARR-TBL-001)* [Data set]. Toronto Police Service Public Safety Data Portal. https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/explore

8 Appendix

Figure 8.1. Post-hoc Tukey's HSD test results for the first two-way ANOVA. First table: comparison of the means of cooperation among different profiled racial groups. Second table: comparison of the means of cooperation among different sex groups.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-4876.0	0.4208	-13505.584	3753.584	False
0	2	-2345.5	0.9	-10975.084	6284.084	False
0	3	-5156.0	0.3647	-13785.584	3473.584	False
0	4	-5600.5	0.2878	-14230.084	3029.084	False
0	5	-5233.0	0.3502	-13862.584	3396.584	False
0	6	-5530.5	0.2989	-14160.084	3099.084	False
0	7	-4867.0	0.4227	-13496.584	3762.584	False
1	2	2530.5	0.9	-6099.084	11160.084	False
1	3	-280.0	0.9	-8909.584	8349.584	False
1	4	-724.5	0.9	-9354.084	7905.084	False
1	5	-357.0	0.9	-8986.584	8272.584	False
1	6	-654.5	0.9	-9284.084	7975.084	False
1	7	9.0	0.9	-8620.584	8638.584	False
2	3	-2810.5	0.876	-11440.084	5819.084	False
2	4	-3255.0	0.7771	-11884.584	5374.584	False
2	5	-2887.5	0.8589	-11517.084	5742.084	False
2	6	-3185.0	0.7927	-11814.584	5444.584	False
2	7	-2521.5	0.9	-11151.084	6108.084	False
3	4	-444.5	0.9	-9074.084	8185.084	False
3	5	-77.0	0.9	-8706.584	8552.584	False
3	6	-374.5	0.9	-9004.084	8255.084	False
3	7	289.0	0.9	-8340.584	8918.584	False
4	5	367.5	0.9	-8262.084	8997.084	False
4	6	70.0	0.9	-8559.584	8699.584	False
4	7	733.5	0.9	-7896.084	9363.084	False
5	6	-297.5	0.9	-8927.084	8332.084	False
5	7	366.0	0.9	-8263.584	8995.584	False
6	7	663.5	0.9	-7966.084	9293.084	False

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-2211.625	0.0747	-4674.7732	251.5232	False

Figure 8.2. Post-hoc Tukey's HSD test results for the second two-way ANOVA. First table: comparison of the means of cooperation among different profiled racial groups. Second table: comparison of the means of cooperation among different sex groups.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-1422.5	0.4675	-4060.2822	1215.2822	False
0	2	-519.5	0.9	-3157.2822	2118.2822	False
0	3	-1545.0	0.3845	-4182.7822	1092.7822	False
0	4	-1525.5	0.3973	-4163.2822	1112.2822	False
0	5	-1562.0	0.3736	-4199.7822	1075.7822	False
0	6	-1606.5	0.346	-4244.2822	1031.2822	False
0	7	-1505.5	0.4106	-4143.2822	1132.2822	False
1	2	903.0	0.844	-1734.7822	3540.7822	False
1	3	-122.5	0.9	-2760.2822	2515.2822	False
1	4	-103.0	0.9	-2740.7822	2534.7822	False
1	5	-139.5	0.9	-2777.2822	2498.2822	False
1	6	-184.0	0.9	-2821.7822	2453.7822	False
1	7	-83.0	0.9	-2720.7822	2554.7822	False
2	3	-1025.5	0.7549	-3663.2822	1612.2822	False
2	4	-1006.0	0.769	-3643.7822	1631.7822	False
2	5	-1042.5	0.7425	-3680.2822	1595.2822	False
2	6	-1087.0	0.7101	-3724.7822	1550.7822	False
2	7	-986.0	0.7836	-3623.7822	1651.7822	False
3	4	19.5	0.9	-2618.2822	2657.2822	False
3	5	-17.0	0.9	-2654.7822	2620.7822	False
3	6	-61.5	0.9	-2699.2822	2576.2822	False
3	7	39.5	0.9	-2598.2822	2677.2822	False
4	5	-36.5	0.9	-2674.2822	2601.2822	False
4	6	-81.0	0.9	-2718.7822	2556.7822	False
4	7	20.0	0.9	-2617.7822	2657.7822	False
5	6	-44.5	0.9	-2682.2822	2593.2822	False
5	7	56.5	0.9	-2581.2822	2694.2822	False
6	7	101.0	0.9	-2536.7822	2738.7822	False

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-614.375	0.1082	-1382.1716	153.4216	False

Figure 8.3. Post-hoc Tukey's HSD test results for the third two-way ANOVA. First table: comparison of the means of cooperation among different profiled racial groups. Second table: comparison of the means of cooperation among different age groups.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-5273.5	0.8009	-19730.4098	9183.4098	False
0	2	-2317.0	0.9	-16773.9098	12139.9098	False
0	3	-5659.5	0.7497	-20116.4098	8797.4098	False
0	4	-6076.0	0.6943	-20532.9098	8380.9098	False
0	5	-5783.0	0.7333	-20239.9098	8673.9098	False
0	6	-6156.5	0.6837	-20613.4098	8300.4098	False
0	7	-5555.5	0.7635	-20012.4098	8901.4098	False
1	2	2956.5	0.9	-11500.4098	17413.4098	False
1	3	-386.0	0.9	-14842.9098	14070.9098	False
1	4	-802.5	0.9	-15259.4098	13654.4098	False
1	5	-509.5	0.9	-14966.4098	13947.4098	False
1	6	-883.0	0.9	-15339.9098	13573.9098	False
1	7	-282.0	0.9	-14738.9098	14174.9098	False
2	3	-3342.5	0.9	-17799.4098	11114.4098	False
2	4	-3759.0	0.9	-18215.9098	10697.9098	False
2	5	-3466.0	0.9	-17922.9098	10990.9098	False
2	6	-3839.5	0.9	-18296.4098	10617.4098	False
2	7	-3238.5	0.9	-17695.4098	11218.4098	False
3	4	-416.5	0.9	-14873.4098	14040.4098	False
3	5	-123.5	0.9	-14580.4098	14333.4098	False
3	6	-497.0	0.9	-14953.9098	13959.9098	False
3	7	104.0	0.9	-14352.9098	14560.9098	False
4	5	293.0	0.9	-14163.9098	14749.9098	False
4	6	-80.5	0.9	-14537.4098	14376.4098	False
4	7	520.5	0.9	-13936.4098	14977.4098	False
5	6	-373.5	0.9	-14830.4098	14083.4098	False
5	7	227.5	0.9	-14229.4098	14684.4098	False
6	7	601.0	0.9	-13855.9098	15057.9098	False

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	3490.5	0.0366	251.1824	6729.8176	True

Figure 8.4. Post-hoc Tukey's HSD test results for the fourth two-way ANOVA. First table: comparison of the means of cooperation among different time periods. Second table: comparison of the means of cooperation among different numbers of search reasons.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0.0	1.0	175.0	0.4981	-136.6129	486.6129	False
0.0	2.0	176.0	0.4928	-135.6129	487.6129	False
0.0	3.0	65.75	0.9	-245.8629	377.3629	False
0.0	4.0	45.875	0.9	-265.7379	357.4879	False
1.0	2.0	1.0	0.9	-310.6129	312.6129	False
1.0	3.0	-109.25	0.8324	-420.8629	202.3629	False
1.0	4.0	-129.125	0.7314	-440.7379	182.4879	False
2.0	3.0	-110.25	0.8273	-421.8629	201.3629	False
2.0	4.0	-130.125	0.7263	-441.7379	181.4879	False
3.0	4.0	-19.875	0.9	-331.4879	291.7379	False