Analyzing Trends and Patterns of Racial Bias and Unfairness in the Toronto Police Service's Arrests and Strip Searches

Antoniette Fracassi and Amal Mohamed
Faculty of Information
INF2178: Experimental Design
Shion Guha
April 16, 2023

Introduction

Law enforcement institutions, particularly police services, have had a deep history of discrimination and racial bias. Research has proven that the complexity of contemporary policing culture, the enforcement of disciplinary practices, and the naturalization of violence against racialized populations, is shaped by imperialist projects that emerged from the era of slavery and colonization (Maynard, 2017). Although law enforcement institutions have implemented anti-racism training to combat this issue, the legacy of colonialism remains embedded within police culture. This has led to over-policing, hyper-surveillance, and the racial profiling of gendered and racialized groups across the globe. Consequently, the criminalization of race has resulted in a lack of trust towards law enforcement institutions.

Prior to the implementation of new policing strategies in 2020, law enforcement institutions in Canada were not mandated to collect race-based data during routine police patrols. In fact, Canadian policing services prided themselves on the fact that they do not see colour and that racism and discrimination were issues exclusive to the United States (Merritt, 2022). However, a recent demand for race-based data has helped shift this narrative and highlight the pervasiveness of systemic racism within policing institutions. As a result, the Toronto Police Service developed the "Race and Identity Based Data Collection Strategy" in 2020. The strategy mandates officers to select one of the following racial groups upon reports or arrests: Indigenous, Black, East/South Asian, South Asian, White, Latino, and Middle Eastern.

Results from the Toronto Police Service's new strategy are showcased in the "Arrests and Strip Searches" dataset, which was published in November 2022 and includes data collected from 2020 and 2021. From January 1st, 2020, all Police Services in Ontario commenced the collection of data on the race of individuals arrested or subjected to strip searches (Toronto Police Services, 2022). The goal of this strategy was to acknowledge and address occurrences of racism within routine police practice. Additionally, the Toronto Police Service considered this a new step towards implementing changes that will promote racial equality.

The primary objective of this research paper is to investigate whether or not there is evidence of racial and gender bias and/or unfairness in the Toronto Police Service's recent arrests and strip searches data, and if so, what are the specific racial groups that are most impacted. The specific research questions of this paper are detailed as follows:

Research Question 1: What is the relationship between race and strip searches? Is there a significant difference in the average number of strip searches performed between the Black and White individuals?

Research Question 2: How do age and gender affect the number of arrests and strip searches that are performed in Toronto?

Research Question 3: What is the relationship between the actions at arrests for individuals with mental instability across different racial and gender groups?

Literature Review

There is extensive literature on racial bias and unfairness in policing culture within Canada. As a result, this paper relies on three key sources. The first source is Robyn Maynard's book "Policing Black Lives: State Violence in Canada from Slavery to the Present". Maynard argues that incarceration and surveillance are rooted in colonial ideas of captivity and that early associations between Blackness and criminality worked to sustain white supremacy (Maynard, 2017). This paper offers historical context, specific to Black people and how they have been historically "made into criminals by the very policing strategies that target them" (Maynard, 2017, p. 87). This is an important source, as it provides relevant background information.

To further understand the meaning of racial profiling and the impact it has had on racialized communities, this paper is influenced by Scot Wortley and Julian Tanner's study, "Discrimination or 'good' policing?". The paper argues that racial profiling, which the authors define as the practice of targeting individuals as suspect based on their race or ethnicity, is a form of prejudice and violates basic human rights (Wortley & Tanner, 2004). The paper offers a detailed definition of racial profiling and discusses the negative impact racial profiling has had on various ethnic communities across Canada.

Lastly, Katherine Merritt's "Final Report and Recommendations on the Collection of Race Based Police Data in Nova Scotia: Submitted to the Wortley Report Research Committee", serves as a fundamental source for this research paper. Merritt analyzes the "stops and searches" data from both the UK and US and the results confirm Black and Hispanic individuals were more

prone to these searches than white people (Merritt, 2022). Merritt also touches on the limitations of crime statistics and they contribute to hyper-policing.

Methods

Dataset Description

This paper used data from the Toronto Police Services. The dataset, "Arrests and Strip Searches", includes 65,276 detailed records of arrests and strip search incidents that occurred during 2020 and 2021 in the city of Toronto. Recorded individuals were each assigned an Event ID, Arrest ID, Person ID.

The dataset provides information on the individual's ethnic background, age group, and also details the use of force and the incidents associated with the "use of force" such as: carrying a weapon, date of the arrest or strip search, and age and gender information of the accused. The dataset utilizes both qualitative and quantitative data.

The quantitative data provides information related to a specific incident which can be expressed using numbers. The qualitative data provides details that assist the reader to understand and gain further insight into the specific occurrence. For instance, the type of crime committed such as "Fraud", "Theft", "Assault", etc. In addition, Arrest Month, Perceived Race, Sex, Age group at arrest are all classified as categorical variables. Individuals fell into the following age groups: 17 and under, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and above. Additionally, the racial categories included the following groups: Indigenous, Black, East/South Asian, South Asian, White, Latino, and Middle Eastern and unknown.

The dataset also provides insight into the Toronto Police Division where the incident or arrest took place. The Toronto Police Service has 17 geographical locations and police officers are dispatched accordingly. This information is useful in ascertaining if specific community groups are more targeted than others.

The programming environment used to conduct both exploratory data analysis and parametric statistical testing is Python. The integrated development environment used to write and run the python code is Jupyter Notebook. The packages used to format, visualize, and analyze the data include: scipy, matplotlib, statsmodel, seaborn, pandas, implotlib, plotly.express, and NumPy.

Statistical Techniques

In conjunction with various visualizations, this paper uses a wide range of statistical tests to not only conduct exploratory data analysis but to draw meaning from the dataset. We conducted 5 t-tests and 1 z-test to further examine the relationship between races, within age groups across different races, between gender, and within different types of arrests across genders. The significance level of the t-tests were set to 0.05.

We also used a two-way analysis of variance (ANOVA) to compare the means between the occurrences of arrests for individuals with mental instability across two groups: race and gender. To focus more on the relationship between the groups identifying as Black or White, we decided to conduct an analysis of covariance (ANCOVA) to dive deeper into research question 1: Is there a significant difference in the average number of strip searches performed between Black and White individuals? The purpose of this was to show the relationship would change when controlling for gender. And so, we decided on gender as the covariate, and kept race as the independent variable and the occurrence of a strip search as the dependent variable.

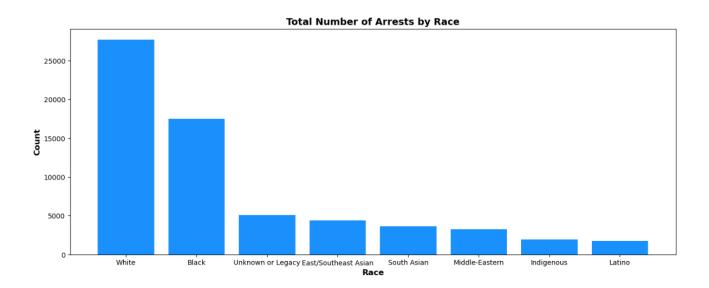
In addition, we conducted a post-hoc power analysis to assess the minimum sample size needed to yield a significant effect in our study. We set the alpha level to 0.05 and the level of power to 0.80. Post power analysis, we plotted a power curve to help visualize the relationship between power, effect and sample size.

Lastly, our study used logistic regression to further investigate research question 2. We used this statistical technique to examine the relationship between the occurrences of arrests and strip searches performed, age, and gender. We chose the binary outcome variable as the number of occurrences of arrests and strip searches, while treating age and gender as the predictor variables.

Results

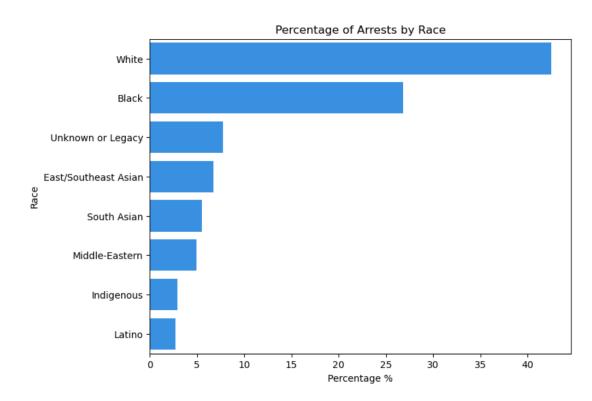
Exploratory Data Analysis: Visualizations

Figure 1: Bar graph comparing the number of total arrests across racial groups in Toronto during 2020 and 2021



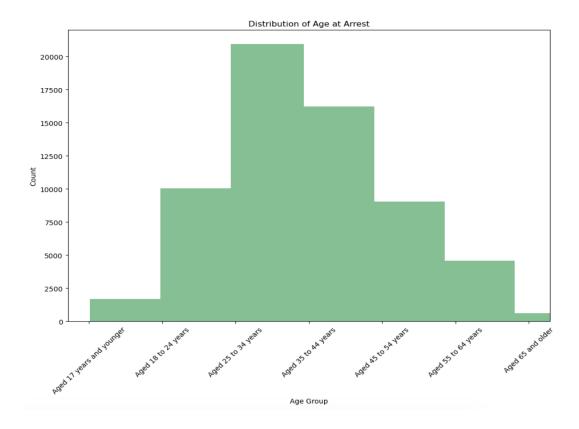
Prior to statistical analysis, this study explores the dataset and provides various visualizations to supplement some key insights. The total number of arrests (combining data from 2020 and 2021) for each racial group is presented in Figure 1. White individuals had the highest total number of arrests, which amounted to 27,723, compared to the other racial groups. Black individuals had the second highest number of recorded arrests, with a total of 17,526. Quite a large portion of the recorded arrests had an "unknown" racial category. The total number of arrested individuals that were not placed into a specific racial category is 5056, which is also larger than all of the other racial groups.

Figure 2: Bar graph comparing the percentage of arrests by each racial group in Toronto during 2020 and 2021



Additionally, Figure 2 also visualizes the relationship between arrests and race. However, this bar graph uses percentages to differentiate between the groups. As shown in Figure 2, approximately 42.4% of recorded arrests were from White individuals and 26.8% were from Black individuals. According to the 2021 Census Canada data, 44.3% of Torontonians do not identify as belonging to a racialized group. The remaining 55.7% identify as a racialized minority. In Toronto, only 8.9% of the population is Black and yet, 26.8% of arrests were of Black people (Statistics Canada, 2022). Additionally, only 0.8% of the population identifies as Indigenous and 2.9% of arrested individuals were of Indigenous.

Figure 3: Histogram showing the distribution of age groups at arrest



According to Figure 3, individuals between the ages of 25 and 34, make up the largest number of arrested individuals across the different age groups, with a total of 20,949. This is evident in the normal distribution of the graph. The mean age of arrested individuals falls in between the ages of 25 and 34. There is an upward trend, with a steady increase once it hits the 18 to 24 age group. An increase in the number of arrests progresses up until the ages of 25 and 34. This is followed by a decline with individuals aged 35 to 45 years. Arrested individuals aged 17 and under, along with those aged 65 and over, account for the smallest number of arrested occurrences.

Figure 4: Line graph showing the distribution of ages at arrest across racial groups

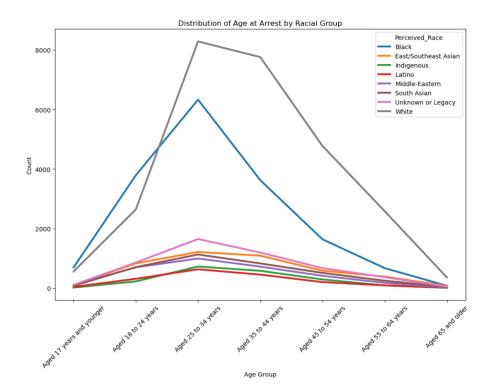


Figure 4 shows the rate of change using the information from Figure 3, by comparing the distribution of age against race, with regards to arrests. Upon first glance, there is a stark difference between the distribution of age with Black and White groups, compared to the remaining racial groups. Both lines are quite dynamic, with the majority of the arrested individuals falling between 25 and 34. While White individuals make up the majority of the offenders, there is a higher number of Black individuals that are arrested between 17 years old and 24 years old. Essentially, within this dataset, there are more Black youth that are arrested in comparison to the other racial groups.

Figure 5: Bar graph comparing the total number of strip searches that occurred across different racial groups

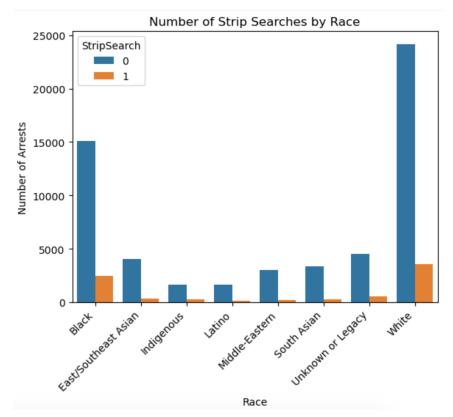


Figure 6: Bar graph comparing number of strip searches between Black and White groups

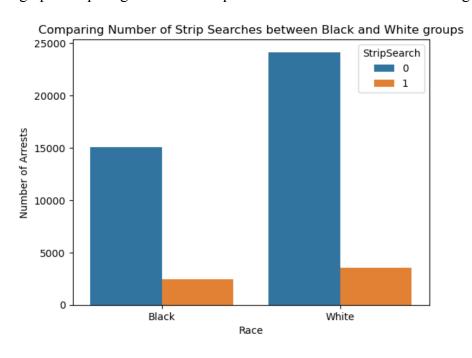
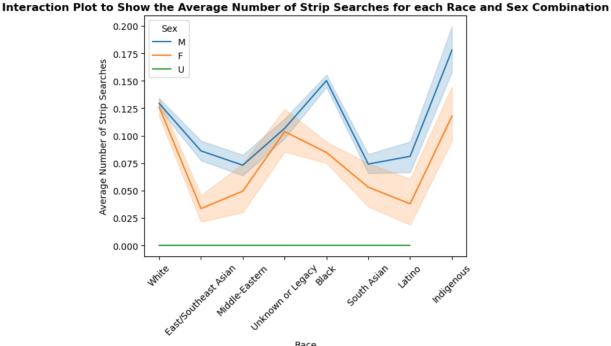


Figure 5 shows the number of strip searches that do not occur for every racial group and Figure 6 shows a close up of the relationship between race and strip searches with a focus on Black and White groups. The 0 denotes that a strip search did not take place, whereas 1 denotes that a strip search has occurred upon arrest. Out of the total 6000 number of strip searches that occurred, 3566 were of White individuals and 2434 were of Black individuals. There is approximately a 1100 difference between the two groups, even though Black individuals only make up 8.9% of the Toronto population.

Figure 7: Interaction plot showing the average number of strip searches for the two groups: race and sex



Although the interaction plot does not hold much weight in the conversation of statistical significance, it does help provide a comprehensive understanding of how various variables interact. To investigate the interplay between race, sex, and strip searches we used an interaction plot, where the y-axis represents whether or not a strip search took place and the x-axis represents perceived racial category. The third variable we used to represent hue is gender. As shown in Figure 7, it is evident that on average there are more strip searches that were recorded for men than for women. The plot also shows that the number of strip searches that happened to

White women, compared very closely to that of White men. Additionally, the plot closely intersects with both unknown genders for the arrested individuals that fall under the Unkown racial category.

Figure 8: Histogram showing the number of arrests and searches by gender for both 2020 and 2021

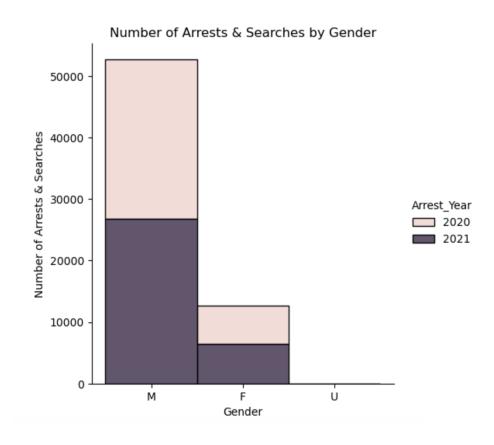


Figure 8 shows the Number of Arrests & Strip Searches by Gender. The chart shows a significant difference between Men and Women. In fact, 81% of Arrests and Strip Searches were male and 19% were women.

Figure 9: Line graph comparing the number of arrests between males and females in both 2020 and 2021

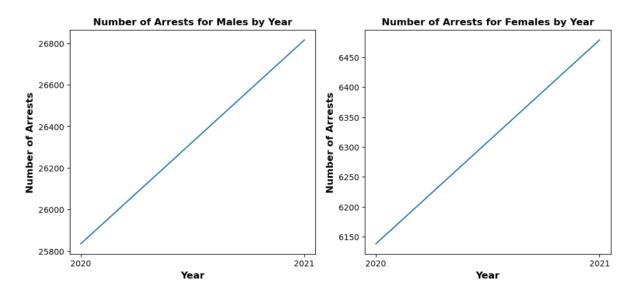


Figure 9 indicates an increase in Arrests & Strip Searches for both Males and Females from 2020 to 2021. The data was collected during the pandemic when society was undergoing factors such as disruptions, social & economic impact, health concerns, increase in death rates and increases in unemployment and inflation. These factors may have contributed to the increase in incidences.

T-tests

To test research question 1, this study compared the mean of strip searches between the two racial groups, Black and White, through an independent two-sample t-test. The independent variable was "perceived race" and the dependent variable was "total number of arrests". The null and alternative hypotheses, along with the results, are outlined as follows:

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

There are no differences in the average number of strip searches conducted by the Toronto Police between Black and White individuals.

Alternative hypothesis (H₁): $\mu_1 \neq \mu_2$

There is a difference in the average number of strip searches conducted by the Toronto Police between Black and White individuals.

Table 1: T-Test comparing number of strip searches between Black and White groups

Race	Sample Size	t-value	p-value
Black	17526	3.132	0.0017
White	27723		

Table 1 presents the result of the two-sample t-test. The p-value is 0.001, which is less than 0.05. From the p-value, we can conclude that the average number of strip searches between the arrested individuals who are Black is different from the average number of strip searches who are White. There is sufficient evidence to state that there is a difference between the two groups, and so, we reject the null hypothesis.

To test research question 2 and to further assess the results highlighted in Figure 4, this study compared the number of arrests between Black and White individuals that are 17 and under. The age group 17 and under was held constant, with the independent variable as "perceived race" and the dependent variable as "total number of arrests". The null and alternative hypotheses, along with the results, are outlined as follows:

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

There are no differences in the average number of arrests between Black and White individuals that are 17 and under.

Alternative hypothesis (H_1) : $\mu_1 \neq \mu_2$

There is a difference in the average number of arrests between Black and White individuals that are 17 and under.

Table 2: T-test comparing the number of arrests between Black and White individuals that are 17 and under

Race	Sample Size	t-value	p-value
Black	17526	12.278	1.343e-34
White	27723		

The t-value is 12.27, which is quite large. This is an indication that there is a significant difference in the number of strip searches that occurred between Black youth aged 17 and under, and White youth aged 17 and under. Additionally, the p-value is extremely small: 1.34e-34. This is extremely significant as it is much less than the threshold for statistical significance. As a result we reject the null hypothesis.

Based on the observed gender differences shown in our EDA Analysis, a second t-test was completed to measure statistically significant differences between gender and the various "Arrest" categories. The null and alternative hypothesis are outlined as follows:

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

There are no significant differences between the average number of resists upon arrest between men and women

Alternative hypothesis (H1):
$$\mu_1 \neq \mu_2$$

There are significant differences between the average number of resists upon arrest between men and women.

Table 3: T-test comparing the average number of resists upon arrest between Men and Women

Gender	Sample Size	t-value	p-value
Male	52650	2.615	0.008
Female	12617		

The P-value result is less than 0.05 therefore, we can reject the null hypothesis and conclude that we have sufficient evidence to say that the alternative hypothesis is true. The result confirms that the difference in men and women who resist arrest is statistically significant.

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

There is no significant difference between men and women who were arrested as a result of assault.

Alternative hypothesis (H₁): $\mu_1 \neq \mu_2$

There is a significant difference between men and women who were arrested as a result of assault.

Table 4: T-test comparing the average number of assaults as cause of arrest between men and women

Gender	Sample Size	t-value	p-value
Male	52650	-3.089	0.002
Female	12617		

P-value result is less than 0.05 therefore, we can reject the null hypothesis and conclude that we have sufficient evidence to say that the alternative hypothesis is true. The result confirms that the difference in men and women who are arrested due to Assault is statistically significant.

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

A significant difference between men and women arrested due to mental instability does not exist.

Alternative hypothesis (H₁):
$$\mu_1 \neq \mu_2$$

A significant difference between men and women arrested due to mental instability exists.

Table 5: T-test comparing the average number of arrests due to mental instability between men and women

Gender	Sample Size	t-value	p-value
Male	52650	-4.646	0.000003
Female	12617		

The P-value result is greater than 0.05 therefore, we fail to reject the null hypothesis and conclude that we do not have sufficient evidence to say that the alternative hypothesis is true. A p-value result greater than 0.05 indicates statistically the result is insignificant.

Z-Test

From the results of our exploratory data analysis in the first phase of this project, there were several results that indicated disproportionate rates of arrests and strip searches across racial groups. As a result, our group decided to conduct a two-proportion z-test to determine whether there is a significant difference between the proportion of white people that had to undergo a strip search vs.the proportion of black people that had to undergo a strip search.. The z-score is -3.11. The negative sign of the z-score means that the difference between the two groups is 3.1 standard deviation points away from the hypothesized difference which was no difference (stated in the null hypothesis). This is an indication that the null hypothesis can be rejected and that there is strong statistical difference between the proportion of strip searches that occurred for Black and White individuals arrested by the Toronto Police Service.

Analysis of Variance (ANOVA)

A two-way ANOVA test was conducted to compare any differences in means across the varying racial and gender groups. The independent variables are race and gender and the dependent variable is the average number of arrests for individuals experiencing mental instability. The null and alternative hypotheses are outlined as follows:

Null hypothesis (
$$H_0$$
): $\mu_1 = \mu_2$

There is no difference between the mean values of people experiencing mental instability at arrest for race and gender.

Alternative hypothesis (H₁): $\mu_1 \neq \mu_2$

There is a difference between the mean values of people experiencing mental instability at arrest for race and gender.

Prior to interpreting the ANOVA results, we first checked the assumptions. The first assumption, normality, was assessed by plotting normal probability. Figure 10 showcases the results.

Figure 10: Normality distribution plot for ANOVA residuals

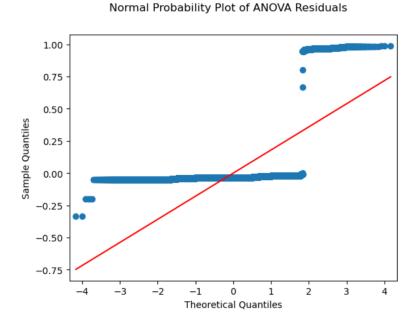


Figure 10 shows that there is no normal distribution. This means that the ANOVA test violated one of the assumption principles. To make sure that this was not a false negative regarding the normality assumptions of the residuals, we also conducted a Shapiro-wilk test. However, the results also showed that there was no normal distribution. We interpreted this as

the ANOVA test not being the best statistical model for our research question. As a result, a post-hoc test was not conducted.

Analysis of Covariance (ANCOVA)

We conducted an ANCOVA test to assess research question 1 from a different perspective by using a covariate. We wanted to answer the question: Is there a significant difference in the average number of strip searches performed between Black and White individuals, when controlling for gender? The purpose of this was to show if the relationship would change when controlling for gender. And so, we decided on gender as the covariate, and kept race as the independent variable and the occurrence of a strip search as the dependent variable.

Table 6: Analysis of covariance for the effect of gender on race and the average number of strip searches

Sources of Variation	Sum of Squares (SS)	Degrees of Freedom (DF)	F-statistic	P-unc	Partial eta squared (np2)
Race_Ind	0.8636	1	7.51	0.0061	0.0002
Sex_Ind	4.3813	1	38.12	0.0000	0.0008
Residuals	5198.8	45241	NaN	NaN	NaN

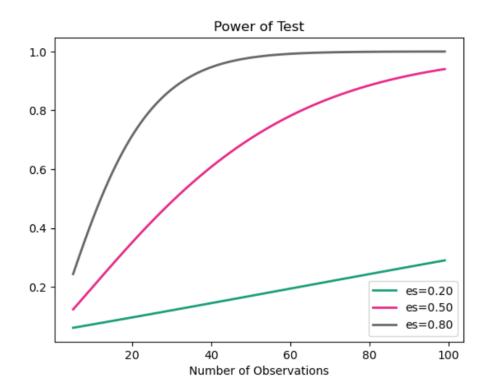
As shown in Table 6, the uncorrected p-value for Race_Ind is less than 0.05. This means that we can reject the null hypothesis that even when controlling for gender, there is a difference between the average number of strip search occurrences between the two racial groups.

Power Analysis

To assess the appropriate sample size needed to conduct statistical testing, our group performed a power analysis. We decided to focus on research question 1 and calculate the minimum sample size that would be needed to avoid a Type II error and to reject the null hypothesis when it is wrong. To begin our power analysis, we calculated the proportion of arrests that occurred for the racial groups: Black and White. As shown in line 49 of our code, the total number of white individuals that were arrested was 27723 and the proportion of these arrests that

involved a strip search was approximately 21%. For Black people, the total number of arrests were 17526 and the proportion of those arrests that involved a strip search was 12%. We calculated the effect size and it was -0.03. The negative sign is an indication that there is not much of a difference in the proportion of strip searches that occurred between the groups. Following the effect size calculation, we computed the sample size and conducted a power analysis, with the statistical power at 80%. The sample size needed for the Black racial group is 14,125 and the sample size needed for the White racial group is 22,343. This means that with a statistical significance threshold of 0.05 and a statistical power of 0.8, there would need to be a minimum sample size of 14,125 for Black individuals and a minimum sample size of 22,343 for White individuals. This indicates that the actual sample sizes are sufficient enough to use for testing and would not affect the reliability of the results. Below is the power curve:

Figure 11: Power curve



Logistic Regression

This study used logistic regression to investigate the relationship between the binary variable, whether or not a strip search was conducted, and two predictor variables, age and gender. The logistic regression model was used to predict the probability of a strip search occurring upon arrest based on race, while holding gender as a constant. The coefficient for the variable race is 0.0859. As a result, for every one-unit increase in predicted log of odds, the occurrence of a strip search being conducted increases by 0.0859. This means that the likelihood of strip searching occurring when a police officer encounters a Black person is higher than for a White person. The p-value for race is 0.007, which is less than 0.05 and so the coefficient value is statistically significant and not a result of chance. To further interpret the results, we calculated the odds ratio. The odds of a strip search for a White individual are: 0.14, whereas it is 0.15 for the odds of a strip search occurring for a Black person. The odds ratio of 1.081, which is <1 and suggests that the odds of a strip search happening to a Black individual upon arrest is higher than that of a White individual.

Table 7: Logistic Regression Results

Dependent Variable		Df Residuals	Pseudo R-squ.	Log- Likeli hood:	Converged	LL-Null	LLR p-value
Strip Search	36199	36197	0.00025	-14116	True	-14120	0.006481

	Coef.	Std err.	z	P> z	[0.025	0.975]
Constant -1.9177	0.020	-95.562	0.000	-1.957	-1.878	
Race_Ind	0.0859	0.032	2.710	0.007	0.024	0.148

Discussion

The results of both the exploratory data analysis and statistical testing show that Black people are significantly over-represented in arrests and strip searches (p < 0.05), while accounting for only 8.9% of the Toronto population. Additionally, both the z-test and power analysis gave us insight that this is an accurate test. There is a large amount of White people who were also arrested and strip searched, however, it is imperative that these results are interpreted in relation to population and demography statistics. Racial bias and unfairness is prevalent within the Toronto Police Service. The disproportionate representation of Black people, particularly, black youth, can be heavily associated with hyper-policing and a policing practice referred to as carding (Maynard, 2017). The practice of carding originated in Toronto during the 1950s, which allowed police officers to stop people on the street and request identification and personal information for no reason. Although discounted, this practice is deep-rooted within the Toronto Police Service system (Maynard). The over-surveillance of racialized individuals is not given as much attention as police brutality, but is equally as harmful. The results of this study also show that there is a statistical significance difference between Black youth aged 17 and under, and White youth aged 17 and under (p < 0.001). There is a large number of Black youth that interact with law enforcement institutions. However, this cannot be understood by discussing the impact of poverty, discrimination, the school to prison pipeline, and various other factors. Racial profiling increases the chances that a Black person interacts with law enforcement and adds to crime statistics that assume Black people are more "deviant" than other racial groups. Crime statistics cannot be analyzed in a vacuum, but rather placed within the context of race, gender, socioeconomic status, and history. Transparent race-based data is transformative and can help in challenging racist narratives.

In addition, the results show that females account for a substantially lower Strip Search and Arrests compared to men. The focus was to confirm if there was a significant statistical difference between gender based on the different arrest categories. The results confirm that a significant difference does exist.

Some of the limitations of this study included the large portion of arrested individuals that were not categorized within a particular racial group. The Toronto Police Service did not

provide an explanation for this in the dataset summary. This is significant because unknown racial groups account for the third largest number of arrests. It is possible that such a large number of non-responses for race can contribute to bias within the data. For this reason, this study did not remove this from the dataset and decided to explore the relationship between the unknown racial category and the other racial groups. Additionally, this study is also very limited and only provides a small snapshot of a larger issue. With such a robust dataset, there were a lot more opportunities to dive deeper into the relationship between race and the variables actions at arrest and type of offences. It is a known fact that racialized groups are arrested at a higher rate for petty crimes, compared to their white counterparts. It would have been interesting to explore this in the context of Toronto.

Conclusion

The "Arrests and Strip Searches" dataset from the Toronto Police Service provides an opportunity to explore how the police interact with the Toronto population. Through descriptive statistics and hypothesis testing, this study showed that racialized groups, particularly Black people, are disproportionately arrested and strip searched. Additionally, this study also assessed the relationship between arrests and strip searches with relation to gender. We concluded that there is statistical significance between men and women when it comes to resistance during arrests.

References

- Maynard, R. (2017). Arrested (in)justice. *Policing Black lives: State violence in Canada from slavery to the present*. Fernwood Publishing.
- Merritt, K. (2022). Final Report And Recommendations On The Collection Of Race-Based Police Data In Nova Scotia: Submitted To The Wortley Report Research Committee. 3-14
- Statistics Canada. (2022). Census of Population: Citizenship, Immigration, Ethnic Origin, Visible Minority Groups (Race), Mobility, Migration, Religion
- Statistics Canada. (2022). Census of Population: Housing and Indigenous Peoples
- Wortley, S., & Tanner, J. (2004). Discrimination or 'good'policing? The racial profiling debate in Canada. Our Diverse Cities, 1(Spring), 197-201.