



University of Toronto

**INF 2178 LEC0101: EXPERIMENTAL DESIGN FOR
DATA SCIENCE**

Midterm Project

Final Version

Winter 2023

Instructor: Prof. Shion Guha

Group Number: Group 61

Group Members:

Qiqi He

Yuqing Wu

February 27, 2023

Table of Contents

List of figures	3
List of tables	3
Abstract	4
Introduction	4
Literature Review	4
Dataset Description	6
Research Questions	6
Significance of Study	7
Outline of the Paper	7
Data Cleaning	7
Exploratory Data Analysis	10
Descriptive Statistics	10
T-tests	13
Methods	14
Results and Findings	16
Discussion	22
Limitations	23
Implications and Future Research	23
Conclusion	24
References	25

List of figures

Figure 1 - Histogram for checking null and not null values	8
Figure 2 - Pie Chart and Histogram of sex distribution	8
Figure 3 - Pie Chart and Histogram of sex distribution	8
Figure 4 - Histogram of Age_group_at_arrest_ distribution	9
Figure 5 - Histogram of Perceived_Race distribution	10
Figure 6 - Booked Count Correlation Heatmap	11
Figure 7 - Strip Count Correlation Heatmap	12
Figure 8 - Interaction Plot of Perceived_Race and Sex based on Booked Count	19
Figure 9 - Interaction Plot of Age_Group and Perceived Race based on Strip Count	21

List of tables

Table 1 - Race and Sex based on Booked count	11
Table 2 - Race and Age based on Strip count	13
Table 3 - Turkey HSD for Perceived Race based on Booked Count	17
Table 4 - Two-way ANOVA results for Perceived Race and Sex based on Booked Count	18
Table 5 - Turkey HSD for Perceived Race based on Strip Count	20
Table 6 - Two-way ANOVA results for Perceived Race and Age based on Strip Count	21
Table 7 - Turkey HSD for Age Group based on Strip Count	22

Abstract

This paper explores the relationship between “Booked”, “Race” and “Sex”, as well as the relationship between “Race”, “Age” and “StripSearch”, using the “Arrests and Strip Searches (RBDC-ARR-TBL-001)” (Arrests and Strip Searches (RBDC-ARR-TBL-001), 2022) dataset from the Toronto Police Service. Our paper provides a detailed description of the dataset, including information on the variables, data limitations or bias, or any other caveats that should be taken into account when analyzing and reporting the data. It provides exploratory data analysis on this dataset, including information on the number of arrests, number of booked and strip searches, demographic information on the individuals involved, and the relationship between variables. The paper aims to find out if there is a relationship between race, sex and the booked event, and if one’s race and age are related to a strip search by exploring any disparities that emerge and discussing their implications. It concludes by summarizing the main findings, stating the limitations, and suggesting avenues for future research.

Introduction

The arrest and strip search of individuals are common practices implemented by police departments worldwide to ensure public safety by identifying and apprehending individuals suspected of criminal activity. However, the implementation of these procedures has been widely criticized for their potential to create inequalities and discrimination in police practices. This study aims to investigate the relationship between an individual's demographic characteristics and the likelihood of being booked, as well as the relationship between an individual's demographic characteristics and the likelihood of being subjected to a strip search in the Toronto Police Service.

Literature Review

Police arrests and strip searches have been the subject of much research, with a focus on the impact of race, age, and gender on police practices. A growing body of literature has highlighted the overrepresentation of certain groups in police arrest and search statistics, suggesting that police practices may be influenced by implicit biases and stereotypes.

Research on the impact of race on police practices has been extensive. A study by Rosenfeld and Fornango (2013) found that African Americans were more likely to be arrested and subjected to searches than any other racial group. The study also found that the race of the officer did not significantly affect the likelihood of arrest, indicating that implicit biases may be at play in police practices. Similarly, a study by Alpert and Dunham (2004) found that black individuals were more likely to be arrested and searched than white individuals for the same offences.

While the research on police practices and race has primarily focused on the overrepresentation of African Americans in arrest and search statistics, there is also evidence of disparities affecting

other racial groups. One study by Eitle and Turner (2002) found that Native Americans were more likely to be arrested and searched than whites or African Americans, even after controlling for factors such as offence type and criminal history. The study suggests that Native Americans may be subject to bias and discrimination in the criminal justice system, which could be attributed to historical and ongoing patterns of discrimination.

Research has also shown disparities affecting Hispanic and Latino individuals. A study by Ghandnoosh (2015) found that Hispanics were more likely to be stopped and searched by the police than whites or African Americans, and were also more likely to experience the use of force during encounters with police. The study suggests that the high rates of immigration enforcement in Hispanic communities may contribute to the disparities observed.

Asian Americans have also been subject to disparities in police practices. A study by Lee, Yoo, and Park (2019) found that Asian Americans were less likely to be arrested and searched than whites or African Americans, but were more likely to experience non-lethal use of force during encounters with police. The study suggests that Asian Americans may be subject to stereotypes that portray them as non-threatening or passive, which may lead police to use force to assert their authority.

Age is another factor that has been shown to influence police practices. A study by Brunson (2007) found that young men were more likely to be arrested and searched than older men, indicating that age-based stereotypes may be at play in police practices. Similarly, a study by Piquero, Piquero, and MacDonald (2003) found that young adults were more likely to be arrested than juveniles or older adults, suggesting that police may be targeting this age group for surveillance and enforcement.

Gender is another factor that has been shown to influence police practices. A study by Miller and Schwartz (2018) found that women were less likely to be arrested and searched than men. The study also found that the gender of the officer did not significantly affect the likelihood of arrest, indicating that implicit biases may be at play in police practices.

In addition to individual-level factors, institutional and contextual factors have been shown to influence police practices. A study by Weitzer and Tuch (2006) found that the organizational culture of police departments influenced arrest and search practices, with more aggressive departments having higher rates of arrests and searches. Another study by Goff, Lloyd, Geller, Raphael, and Glaser (2016) found that racial bias was more likely present in departments with more negative public perceptions and lower levels of community trust.

Overall, the literature suggests that police practices are influenced by a complex interplay of individual-level, institutional, and contextual factors. The analysis of the Arrests and Strip

Searches dataset will provide important insights into the extent of disparities in police practices and help identify potential areas for intervention and reform.

Dataset Description

The data set used in this project is from the website of the Toronto Police Department, which provides information about all arrests and strip searches every three months from 2020 to 2021. You can find it through the following link:

<https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about>. The data set collected 65276 lines of crime data, with a total of 25 attributes, including object ID, the year of arrest, the month of arrest, the event ID, the arrest ID, the person ID, and the demographic information of the arrested (such as race, sex, and age group). The place of arrest, the type of crime, whether to register at the police station within 24 hours and the strip search is also included. At the same time, with regard to the action and search reasons at the time of arrest, the data set has some decision variables, such as whether to find Concealed items, mental ability or possibly feasible, AssistEscape, PossessWeapons, etc. However, there are many null values for the four attributes related to the search reason. ObjectID is unique for each line of the record. Among them, 12 are numeric variables and 12 are text variables. Numeric variables include binary values and numeric integer values.

Research Questions

This study seeks to explore two research questions, which are as follows:

- ***RQ1: How do booked counts differ by criminal's perceived race and sex in Toronto?***

The first research question aims to investigate how "Race" and "Sex" are related to "Booked". The study intends to determine whether a statistically significant relationship exists between an individual's gender and ethnicity and the likelihood of being arrested by the Toronto Police Service. Prior research has suggested that gender biases may play a role in police practices, with men being more likely to be arrested than women. However, studies have also found that this relationship may vary depending on the context, such as the nature of the crime and the race of the individual involved. Therefore, this study aims to examine this relationship in the context of the Toronto Police Service and contribute to the broader literature on the topic.

- ***RQ2: For Toronto criminals, will the number of strip checks vary depending on their age and perceived race?***

The second research question of this study is to explore the relationship between "Striped," "Race," and "Age." This study aims to determine whether a statistically significant relationship exists between an individual's ethnicity and age and the likelihood of being subjected to a strip search by the Toronto Police Service. Research in this area

has suggested that there may be implicit biases and stereotypes that influence police practices, particularly with respect to race and age. Studies have found that individuals from certain racial and ethnic groups are more likely to be arrested and subjected to strip searches and that young adults may be more likely to be targeted for surveillance and enforcement. Thus, this study aims to explore these relationships in the context of the Toronto Police Service and contribute to the broader literature on the topic.

Significance of Study

This study's findings have important implications for understanding potential biases and discrimination in police practices in Toronto, Canada. This study aims to provide essential insights into the extent of disparities in police practices and identify potential areas for intervention and reform. By examining the relationship between an individual's demographic characteristics and the likelihood of being booked or subjected to a strip search, this study aims to contribute to the broader literature on policing practices and help inform policy decisions.

Outline of the Paper

The paper is structured as follows: The introduction provides a comprehensive overview of the existing literature on police practices and the impact of race, age, and gender on policing and the basic information of the dataset. Then the data cleaning gives a general explanation of how to choose the appropriate dataset. The exploratory data analysis section presents an overview of the dataset used in this study and presents key descriptive statistics. The methods section provides a detailed description of the statistical methods used in this study, including ANOVA tests and post-hoc tests such as Tukey's HSD. The results and findings section presents the results of the analysis, including the relationships between the variables of interest. The discussion section provides a critical evaluation of the results and discusses the implications for policy and future research. The paper concludes with a summary of the key findings, limitations and recommendations for future research.

Data Cleaning

Considering that the dataset is huge and has 24 attributes, for the convenience of subsequent analysis, we need to preprocess the dataset. We first use the following histogram to show the counting of null and not null values for all 24 attributes in the dataset (Figure 1). By looking at Figure 1, blue represents a null value while orange represents not null value, we can see that there are 5 attributes 'SearchReason_CauseInjury', 'SearchReason_AssistEscape', 'SearchReason_PossessWeapons', 'SearchReason_PossessEvidence' and 'ItemsFound' that have more than 50000 null values, for the accuracy of the follow-up analysis results, these four columns of data will not be used in the later analysis. For a small amount of null value data of other attributes, we choose to drop the row corresponding to the null value.

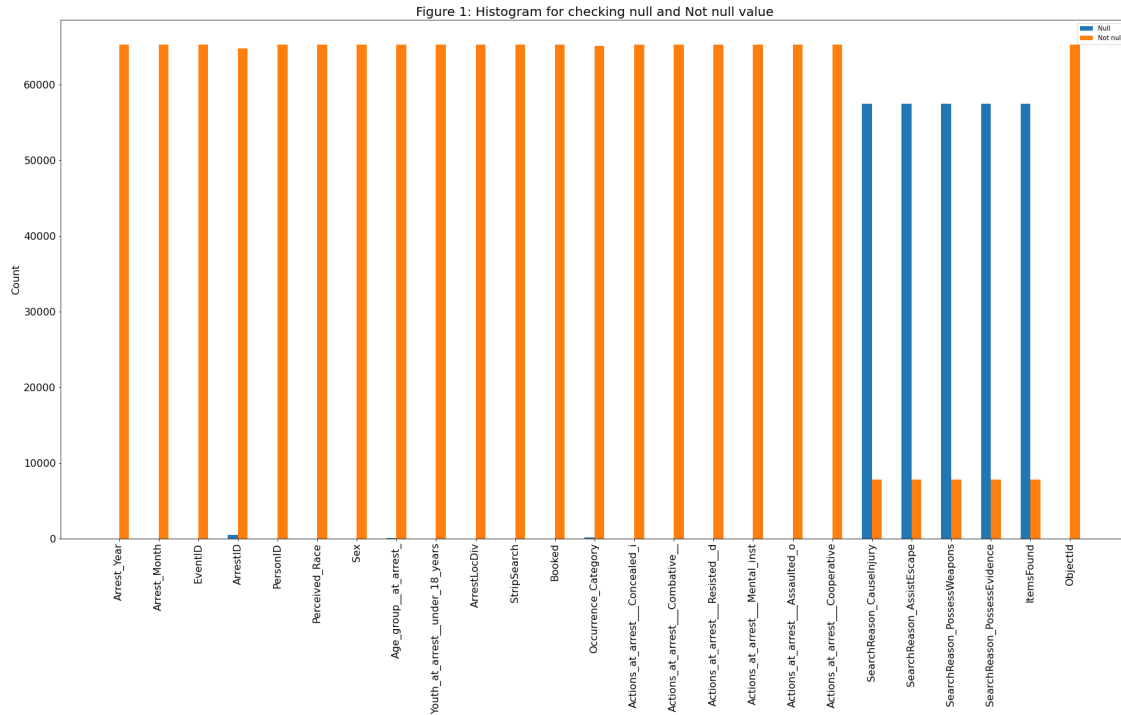


Figure 1: Histogram for checking null and not null values

After dropping columns with too many null values, considering that we want to discuss sex, age group and race for further analysis, the following plots show the distribution of these on the original dataset. Through Figure 2 and Figure 3, we can conclude that in this dataset, men account for the majority, while the union sex accounts for only 9, which is too small. Therefore, we removed union sex in the subsequent analysis. Not only that but also the bias of the data itself may cause certain limitations to our research.

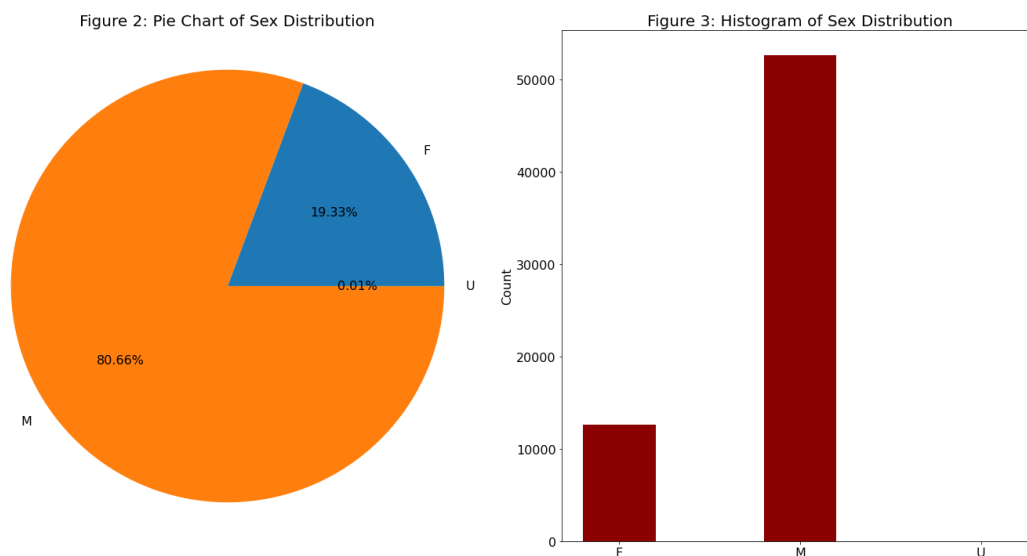


Figure 2 & Figure 3: Pie Chart and Histogram of sex distribution

Similarly, for the age group, we use the histogram to show its distribution. We can find that there are a total of 9 groups in Figure 4, but when we see the x-axis carefully, there exists overlapping in the categories. For example, “Aged 17 years and under” and “Aged 17 years and younger” are the same meaning, thus we combined these two groups into “Aged 17 years and under”. Similarly “Aged 65 and older” and “Aged 65 years and older” are merged into “Aged 65 years and older”. Meanwhile, we can find that among the criminals, those aged 25 to 34 are the most, followed by those aged 35 to 44.

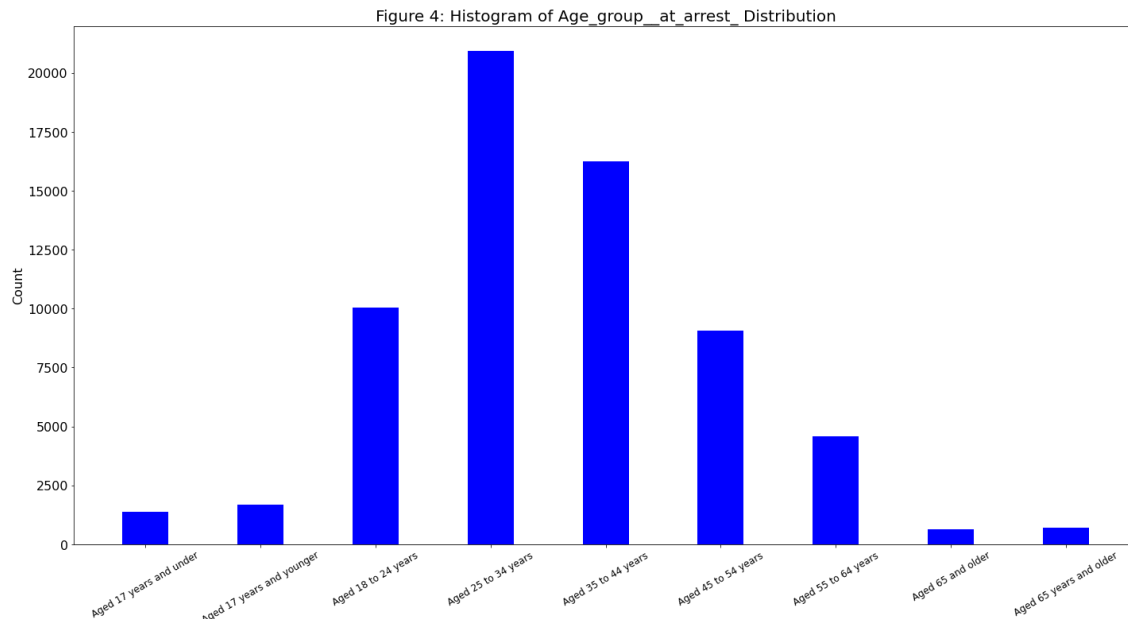


Figure 4: Histogram of Age_group_at_arrest_distribution

From Figure 5, it can be seen that there are eight racial types, of which white people account for the largest number, and white and black people account for the majority of the race. This does not mean that the data has bias because the most commonly reported races in Toronto were White (50.2%), East Asian (12.7%), South Asian (12.3%), and Black (8.5%) according to the racial demographic data of Toronto in 2021. (Toronto Population 2021/2022, n.d.) Therefore, we think that the data on perceived race is logical and not very biased in analysis.

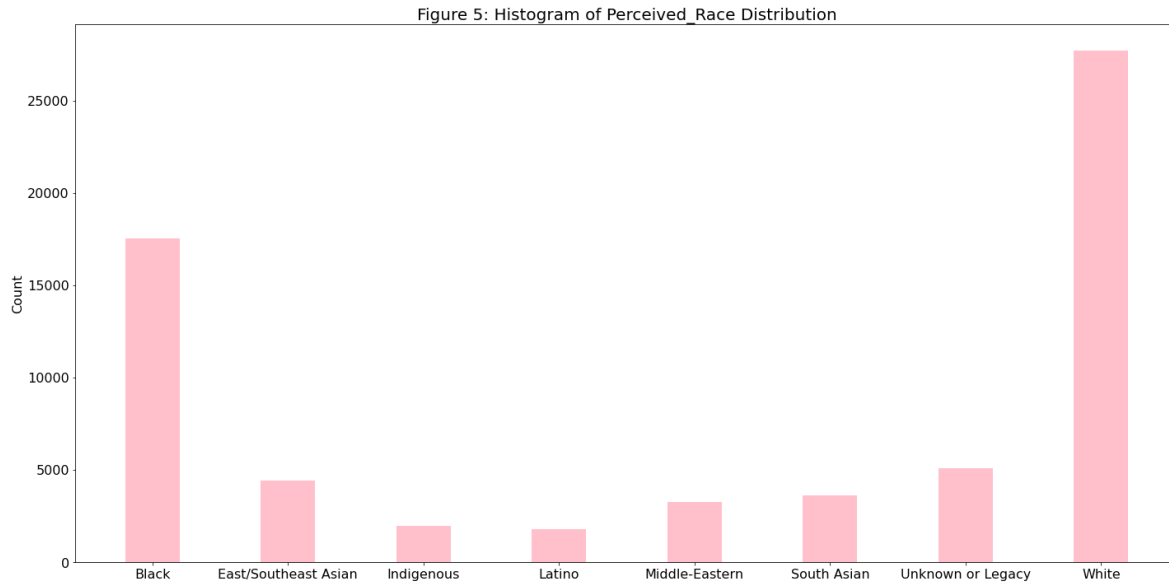


Figure 5: Histogram of Perceived_Race distribution

For Actions_at_arrest attributes, we found that there are too many 0 among them, which we think there are too few available data for reasonable statistical analysis, so we also removed them from the final data set. And for 'Occurrence_Category', by grouping them into a table, we find that there are 31 kinds of crime which is too much for us to split and merge. And the objectID is the label of each record and has no other special meaning, so we also removed this column. Finally, we just keep 8 attributes, which are 'Arrest_Year', 'Arrest_Month', 'Perceived_Race', 'Sex', 'Youth_at_arrest__under_18_years', 'Age_group__at_arrest_', 'Booked' and 'StripSearch' for our further analysis in this project.

Exploratory Data Analysis

Descriptive Statistics

Based on our general thought and research question, we need to have a general understanding of sex, age group and race based on the booked counting. Firstly, we want to learn about whether booked count will associate with sex and race, thus we provide a correlation heatmap with the booked count (Figure 6). Through the heatmap, it is clear that race and sex are the two variables with the highest correlation with the booked count. As a result, we believe that it is reasonable for us to study whether the theme of booked count is different due to race and gender.

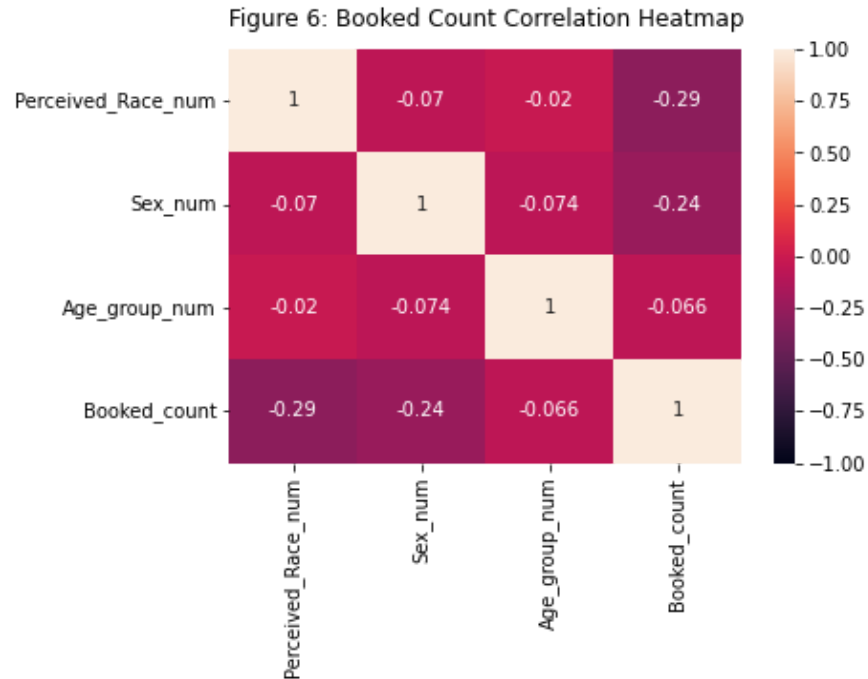


Figure 6: Booked Count Correlation Heatmap

The following table shows the mean booked count based on race and sex (Table 1). We can find that the mean booked count is different in different race groups and sex groups, the average number of booked times of black men and white men is significantly higher than that of other races. And the frequency of booking in men is higher than that in women. However, this may be due to the fact that the number of males is more than that of females, and the number of blacks and whites is higher than that of other races in this data set. So, we need to do more statistical analysis to prove this phenomenon.

Table 1: Race and Sex based on Booked count

Perceived_Race		Sex	Booked_count
Black		F	16.617284
		M	78.528302
East/Southeast Asian		F	3.650794
		M	20.260870
Indigenous		F	5.578947
		M	9.934211
Latino		F	3.285714
		M	10.493827
Middle-Eastern		F	2.562500
		M	16.955056
South Asian		F	3.634615
		M	18.904762
Unknown or Legacy		F	5.873016
		M	21.978723
White		F	28.000000
		M	107.065421

Similarly, based on our research question 2, we want to have a general look at the correlation between strip count (Figure 7). As we can see, race and strip count are the highest correlated attributes, and sex is the second. However, through our first research question, we have already analyzed race and sex and their interaction. So for this research question, we will choose to analyze the relationship between strip count and race and age. This correlation heatmap just shows the age under 18 years and above, which does not reveal age groups. In our analysis, we will further subdivide the age into groups to find the relationship between groups and strip count.

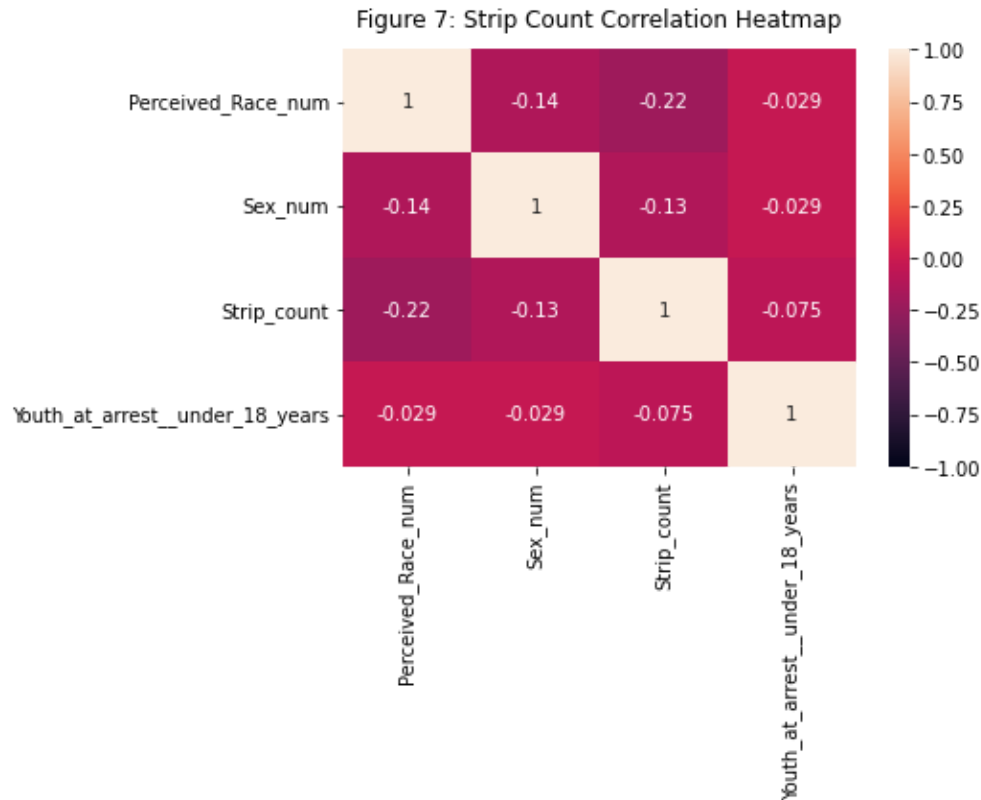


Figure 7: Strip Count Correlation Heatmap

In addition to the correlation heatmap (Figure 7), we can also see from the data whether there are differences between strip count, race and age. Table 2 shows us the average number of strip counts for different ages and races. It is obvious that there is an average difference between different groups. It can be seen that the average stripping inspection frequency of not a youth is significantly higher than that of minors, and the average stripping inspection rate of blacks and whites is significantly higher than that of other races. However, considering the uneven distribution of people of different races and ages in the data itself, we can not draw a conclusion from this table, and we need follow-up ANOVA analysis.

Table 2: Race and Age based on Strip count

Perceived_Race	Youth_at_arrest__under_18_years	Strip_count
Black	Not a youth	24.021053
	Youth (aged 17 years and under)	11.692308
East/Southeast Asian	Not a youth	6.054545
	Youth (aged 17 years and under)	2.666667
Indigenous	Not a youth	4.838710
	Youth (aged 17 years and under)	1.500000
Latino	Not a youth	3.236842
	Youth (aged 17 years and under)	1.800000
Middle-Eastern	Not a youth	4.280000
	Youth (aged 17 years and under)	2.800000
South Asian	Not a youth	5.255319
	Youth (aged 17 years and under)	2.000000
Unknown or Legacy	Not a youth	7.253521
	Youth (aged 17 years and under)	3.500000
White	Not a youth	28.975207
	Youth (aged 17 years and under)	4.615385

T-tests

Considering that the above heatmap could only reveal the correlations between different attributes, we decide to conduct independent sample t-tests on some of the attributes. The t-test is a statistical test used to determine whether there is a significant difference between the means of two groups or populations. Because t-tests have some assumptions to run before we do the test, we have already checked the following assumptions: (1) Normality: The distribution of the population should be normal or approximately normal. (2) Homogeneity of variance: The variances of the two groups being compared should be equal. (3) Independence: The observations in each group should be independent of each other. This means that there should be no relationship between the observations in each group. Before we conduct ANOVA tests and post-hoc tests, we first do the t-tests for the dataset on booked count and strip count. There is a total of two t-tests we made in this project.

Sex and Booked Count

The first one is for sex based on the booked count. We compare the mean of booked count between males and females and want to know whether they are different. The hypothesis of this test is as follows:

H0: The population mean of Males and Female's booked count are equal

Ha: The population mean of Males and Female's booked count are not equal

After calculation, the p-value is $5.67923465780535e-25$, which is statistically significant at a $p=0.05$ level. We can conclude that there is evidence that male and female booked count means are not equal, which is the same as the previous heatmap and table conclusions.

Age and Strip Count

The second one is for age based on strip count, we compare the mean of strip count between adults and youth, and want to know whether there exists a difference. And the hypothesis of this test is as follows:

H0: The population mean of Not a youth and youth under 18 years' strip count are equal

Ha: The population mean of Not a youth and youth under 18 years' strip count are not equal

After calculation, the p-value is $1.1316728582652387e-08$, which is statistically significant at a $p=0.05$ level. We can conclude that there is evidence that not a youth and youth under 18 years' strip count means are not equal, which is the same as the previous heatmap and table conclusions.

From these two t-tests result, we hope to further explore how these variables affect booked and strip count. As for age, the result of the t-test only shows the difference in the frequency of undressing examinations between adults and adults. We hope that we can group the age more carefully and make a more in-depth analysis.

Methods

In this project, we mainly use One-way ANOVA tests, Two-way ANOVA tests and Tukey's HSD tests. ANOVA test is mainly used to compare the means of three or more groups and determine if there is a significant difference in the means of the groups being compared. Before we conduct the ANOVA tests, three assumptions have been checked: (1) Normality: The distribution of the population should be normal or approximately normal. (2) Homogeneity of variance: The variances of the two groups being compared should be equal. (3) Independence: The observations in each group should be independent of each other. This means that there should be no relationship between the observations in each group. Tukey's HSD test is a multiple comparison test that compares all possible pairs of means to determine which pairs are significantly different from each other. In addition to the above three assumptions, Tukey's HSD tests also need to meet (4) Homogeneity of sample sizes: The sample sizes in each group should be approximately equal.

Tests for Research Question 1

For research question 1, One-way ANOVA tests are performed on the difference of booked count means between the groups of perceived race. The hypothesis is as follows.

RQ1 One-way ANOVA: Perceived Race and Booked Count

H0: There is no difference in the booked count means of the different levels of perceived race

Ha: The booked count means the different levels of perceived race are not the same

Then, Tukey's HSD test is used to identify which perceived race groups' difference is significant. It can control the family-wise error rate, which is the probability of making at least one false positive result in multiple comparisons. Finally, we conduct a Two-way ANOVA test for both perceived race and sex on the booked count, including the interaction between perceived race and sex to find whether these variables will affect each other. The two-way ANOVA hypothesis is as follows.

RQ1 Two-way ANOVA: Perceived Race, Sex and Booked Count

Hypothesis 1:

H0: There is no difference in the booked count means of the different levels of perceived race

Ha: The booked count means the different levels of perceived race are not the same

Hypothesis 2:

H0: There is no difference in the booked count means of the different levels of sex

Ha: The booked count means the different levels of sex are not the same

Hypothesis 3:

H0: There exists no interaction effect between perceived race and sex

Ha: There exists an interaction effect between perceived race and sex

Tests for Research Question 2

Similarly for research question 2, at first we use two One-way ANOVA tests on perceived race groups and strip count, and age groups and strip count. The hypothesis is as follows.

RQ2 One-way ANOVA 1: Perceived Race and Strip Count

H0: There is no difference in the strip count means of the different levels of perceived race

Ha: The strip count means the different levels of perceived race are not the same

RQ2 One-way ANOVA 2: Age Group and Strip Count

H0: There is no difference in the strip count means of the different levels of age groups

Ha: The strip count means the different levels of age groups are not the same

Then for both of them, Tukey's HSD tests are applied for race groups and age groups to find which group has a significant difference. Lastly, a Two-way ANOVA test is performed on

perceived race and age group to check whether an interaction exists between them. The two-way ANOVA hypothesis is as follows.

RQ2 Two-way ANOVA: Perceived Race, Age Group and Strip Count

Hypothesis 1:

H0: There is no difference in the strip count means of the different levels of perceived race

Ha: The strip count means the different levels of perceived race are not the same

Hypothesis 2:

H0: There is no difference in the strip count means of the different levels of age groups

Ha: The strip count means the different levels of age groups are not the same

Hypothesis 3:

H0: There exists no interaction effect between perceived race and age group

Ha: There exists an interaction effect between perceived race and age group

This chapter only introduces the assumptions and principles related to tests. All the results and analyses of ANOVA tests and post-hoc tests are shown in the next section.

Results and Findings

The analysis conducted in this study reveals important findings regarding the relationship between demographic factors and the booking and strip search rates in the Toronto Police Service. The results of the ANOVA tests and post-hoc tests indicate that there are significant differences in the booking and strip search rates across different demographic groups, highlighting potential issues of bias and discrimination in policing practices.

R1 Race and Booking: One-way ANOVA and Turkey

The results of the one-way ANOVA show a statistically significant effect of race on book count ($p < 0.001$). The effect of race was found to be significant, having p equal to $1.96e-38$, indicating that individuals from different racial groups have different book counts on average.

The Tukey HSD test (Table 3) revealed that there were significant differences in book count between the different racial groups. When reject is True, the average number of booked counts between the two groups is different. Specifically, Black, South Asian, and Indigenous had significantly higher book counts than White. East Asian also had a significantly higher book count than White individuals, but the effect size was relatively small.

Table 3: Turkey HSD for Perceived Race based on Booked Count

Multiple Comparison of Means – Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-38.2016	0.001	-55.504	-20.8991	True
Black	Indigenous	-43.6436	0.001	-61.7116	-25.5756	True
Black	Latino	-43.3923	0.001	-62.2181	-24.5665	True
Black	Middle-Eastern	-39.7988	0.001	-57.712	-21.8856	True
Black	South Asian	-38.6451	0.001	-56.5962	-20.6939	True
Black	Unknown or Legacy	-36.1953	0.001	-53.4374	-18.9532	True
Black	White	17.1583	0.0267	1.088	33.2286	True
East/Southeast Asian	Indigenous	-5.442	0.9	-24.2692	13.3852	False
East/Southeast Asian	Latino	-5.1907	0.9	-24.7463	14.3649	False
East/Southeast Asian	Middle-Eastern	-1.5973	0.9	-20.276	17.0814	False
East/Southeast Asian	South Asian	-0.4435	0.9	-19.1586	18.2716	False
East/Southeast Asian	Unknown or Legacy	2.0062	0.9	-16.0299	20.0424	False
East/Southeast Asian	White	55.3599	0.001	38.4405	72.2793	True
Indigenous	Latino	0.2513	0.9	-19.9848	20.4874	False
Indigenous	Middle-Eastern	3.8447	0.9	-15.5452	23.2347	False
Indigenous	South Asian	4.9985	0.9	-14.4266	24.4236	False
Indigenous	Unknown or Legacy	7.4483	0.9	-11.3235	26.22	False
Indigenous	White	60.8019	0.001	43.1004	78.5034	True
Latino	Middle-Eastern	3.5934	0.9	-16.5046	23.6914	False
Latino	South Asian	4.7472	0.9	-15.3846	24.8791	False
Latino	Unknown or Legacy	7.197	0.9	-12.3053	26.6992	False
Latino	White	60.5506	0.001	42.0763	79.0249	True
Middle-Eastern	South Asian	1.1538	0.9	-18.1274	20.4349	False
Middle-Eastern	Unknown or Legacy	3.6035	0.9	-15.0193	22.2263	False
Middle-Eastern	White	56.9572	0.001	39.4137	74.5006	True
South Asian	Unknown or Legacy	2.4497	0.9	-16.2096	21.1091	False
South Asian	White	55.8034	0.001	38.2211	73.3856	True
Unknown or Legacy	White	53.3536	0.001	36.496	70.2113	True

R1 Gender and Booking: One-Way Anova

The one-way ANOVA test conducted on the booking rates by gender showed a statistically significant difference ($p = 1.59e-17$, $p < 0.01$). These findings suggest that gender plays a significant role in the booking process and that males are more likely to be booked and brought into custody than females.

R1 Gender and Booking: Two-Way Anova

The booked count means for different levels of perceived race are not the same; the booked count means for different levels of Sex are not the same; there exists an interaction effect between perceived race and sex. The results of the two-way ANOVA (Table 4) show a statistically significant effect of both race and sex on book count ($p < 0.001$). The effect of race was found to be significant ($p=2.07e-57$, $p < 0.001$), indicating that individuals from different racial groups have different book counts on average. The effect of sex was also found to be

significant ($p = 1.18e-28$, $p < 0.001$), indicating that males have a higher book count on average than females.

Table 4: Two-way ANOVA results for Perceived Race and Sex based on Booked Count

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	7.420622e+05	7.0	45.514238	2.074719e-57
C(Sex)	3.025305e+05	1.0	129.889540	1.175224e-28
C(Perceived_Race):C(Sex)	2.325323e+05	7.0	14.262320	6.770281e-18
Residual	2.822914e+06	1212.0	NaN	NaN

The interaction plot between Perceived_Race and Sex based on Booked Count shows how the effect of Perceived_Race and Sex on the Booked Count variable changes depending on the values of the other variable.

Looking at the plot (Figure 8), it can be seen that for both male and female individuals, Black individuals have the highest booked count, followed by Indigenous, White, and then all other races. However, the effect of race on booked count is stronger for male individuals compared to female individuals. Specifically, the booked count for Black males is substantially higher than the booked count for males of all other races. In contrast, the difference in booked count between Black and non-Black females is not as pronounced.

This suggests that there is an interaction effect between Perceived_Race and Sex on the Booked Count variable. That is, the effect of race on booked count varies depending on the individual's sex. This highlights the importance of considering the joint effects of multiple variables when analyzing the relationship between demographic characteristics and police practices (see Figure 8).

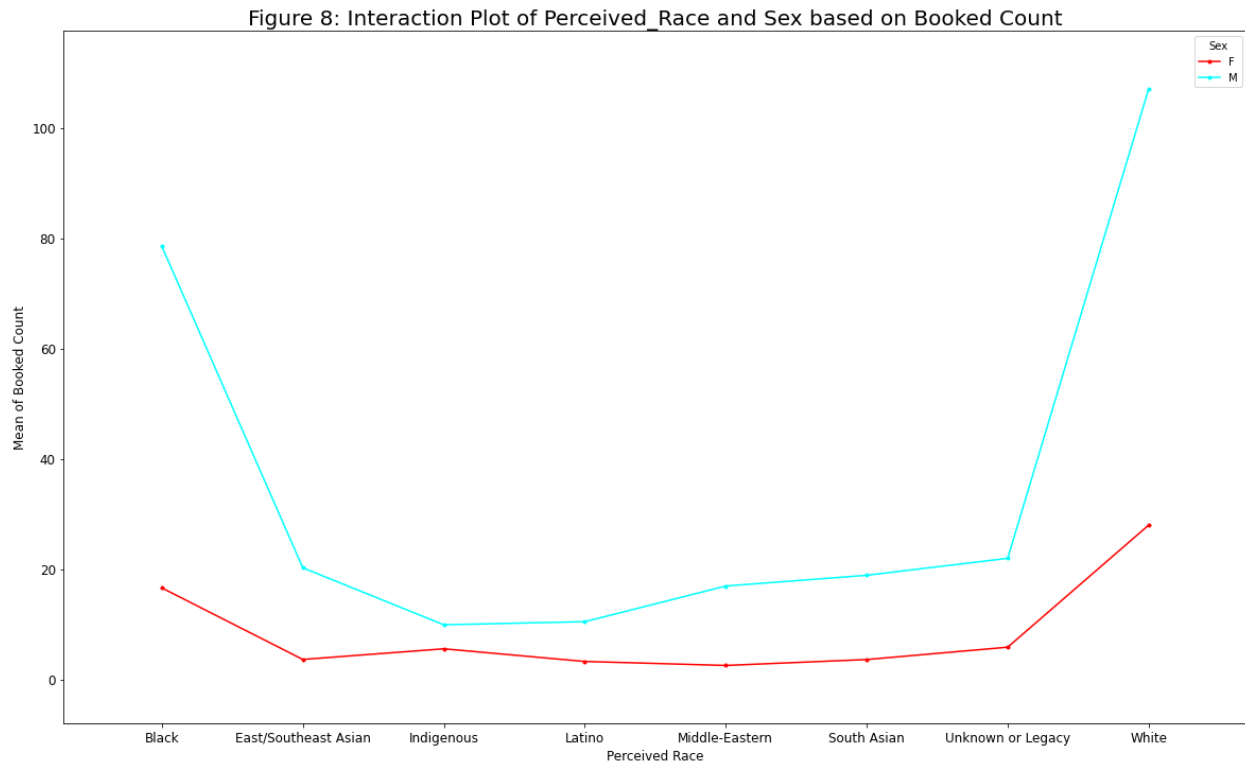


Figure 8: Interaction Plot of Perceived_Race and Sex based on Booked Count

R2: Race and Strip Search

The one-way ANOVA test conducted on the strip search rates by race showed a statistically significant difference ($p = 1.056e-07$, $p < 0.01$). The results of the post-hoc test using Tukey's HSD (Table 5) revealed that the strip search rate for black individuals was significantly higher than that for white and other racial groups. Additionally, the strip search rate for white individuals was significantly higher than that for other racial groups. These findings suggest that race plays a significant role in the strip search process and that black individuals are more likely to be subjected to strip searches than individuals of other races.

Table 5: Turkey HSD for Perceived Race based on Strip Count

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-16.6577	0.0356	-32.7084	-0.6071	True
Black	Indigenous	-17.9007	0.0104	-33.3054	-2.4959	True
Black	Latino	-19.4673	0.0206	-37.2462	-1.6883	True
Black	Middle-Eastern	-18.3916	0.0151	-34.7245	-2.0586	True
Black	South Asian	-17.5947	0.0297	-34.2369	-0.9525	True
Black	Unknown or Legacy	-15.576	0.0291	-30.2819	-0.8701	True
Black	White	4.0749	0.9	-8.675	16.8248	False
East/Southeast Asian	Indigenous	-1.2429	0.9	-18.9885	16.5026	False
East/Southeast Asian	Latino	-2.8095	0.9	-22.6511	17.032	False
East/Southeast Asian	Middle-Eastern	-1.7339	0.9	-20.2909	16.8231	False
East/Southeast Asian	South Asian	-0.937	0.9	-19.7668	17.8927	False
East/Southeast Asian	Unknown or Legacy	1.0817	0.9	-16.0607	18.2241	False
East/Southeast Asian	White	20.7326	0.0014	5.2356	36.2296	True
Indigenous	Latino	-1.5666	0.9	-20.8894	17.7562	False
Indigenous	Middle-Eastern	-0.4909	0.9	-18.4922	17.5104	False
Indigenous	South Asian	0.3059	0.9	-17.9764	18.5883	False
Indigenous	Unknown or Legacy	2.3247	0.9	-14.2145	18.8639	False
Indigenous	White	21.9756	0.001	7.1485	36.8026	True
Latino	Middle-Eastern	1.0757	0.9	-18.995	21.1463	False
Latino	South Asian	1.8725	0.9	-18.4506	22.1956	False
Latino	Unknown or Legacy	3.8913	0.9	-14.8792	22.6617	False
Latino	White	23.5422	0.001	6.2614	40.823	True
Middle-Eastern	South Asian	0.7969	0.9	-18.2741	19.8678	False
Middle-Eastern	Unknown or Legacy	2.8156	0.9	-14.5914	20.2226	False
Middle-Eastern	White	22.4665	0.001	6.6773	38.2557	True
South Asian	Unknown or Legacy	2.0187	0.9	-15.6788	19.7162	False
South Asian	White	21.6696	0.0013	5.5607	37.7785	True
Unknown or Legacy	White	19.6509	0.001	5.5513	33.7505	True

R2: Age and Strip Search

The one-way ANOVA test conducted on the strip search rates by age showed a statistically significant difference ($p = 0.0045$, $p < 0.01$), indicating that individuals of different ages have different strip search counts on average.

Moving forward to interaction plots. From the plot (Figure 9), we can see that the lines for each age group are not parallel, indicating that there is an interaction effect between the age group and perceived race on the strip count. This means that the effect of perceived race on the strip count varies depending on the age group of the individual.

Figure 9: Interaction Plot of Age_Group and Perceived Race based on Strip Count

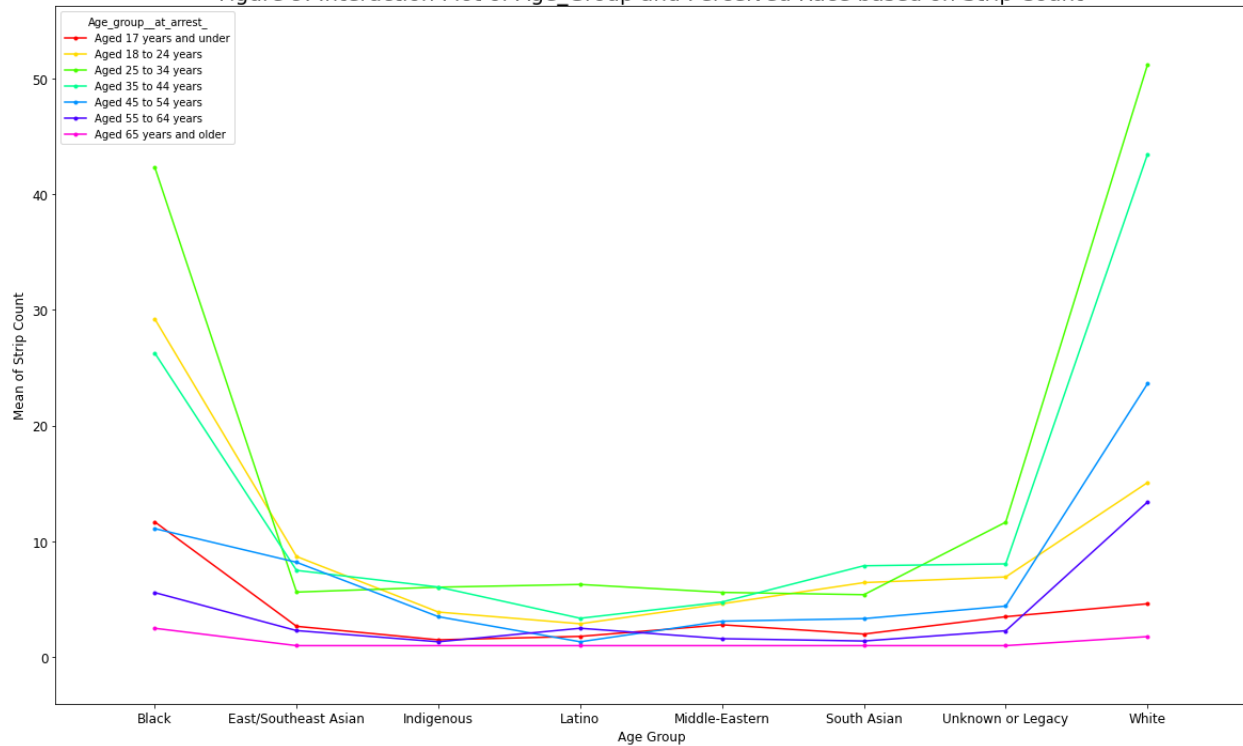


Figure 9: Interaction Plot of Age_Group and Perceived Race based on Strip Count

To formally test for the interaction effect, we can conduct a two-way ANOVA with age group and perceived race as the independent variables and strip count as the dependent variable. The ANOVA result shows that both age group and perceived race have a significant effect on the strip count ($p < 0.01$). Furthermore, the interaction effect between age group and perceived race is also significant ($p < 0.01$). This confirms that the effect of perceived race on the strip count varies depending on the age group of the individual. The results of the ANOVA are shown in Table 6 below:

Table 6: Two-way ANOVA results for Perceived Race and Age based on Strip Count

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	91549.224879	7.0	12.725180	1.049841e-11
C(Age_group_at_arrest_)	37840.034586	6.0	6.136314	7.815658e-05
C(Perceived_Race):C(Age_group_at_arrest_)	32257.022272	42.0	0.747278	8.722386e-01
Residual	553963.904724	539.0	NaN	NaN

The strip count means for different levels of perceived race are not the same; the strip count means for different levels of Age are not the same; there exists an interaction effect between perceived race and age.

Next, we perform a post-hoc Tukey test (Table 7) to investigate the specific differences in strip search means between different age groups. The results of the Tukey test indicate that there is a statistically significant difference between the mean strip search count for the 25-34 age group and the mean strip search count for the 55-64 age group (p-value = 0.0417). However, there are no other statistically significant differences in strip search means between any of the other age groups.

Table 7: Turkey HSD for Age Group based on Strip Count

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Aged 17 years and under	Aged 18 to 24 years	7.0785	0.8471	-9.3007	23.4576	False
Aged 17 years and under	Aged 25 to 34 years	15.8072	0.0532	-0.1161	31.7306	False
Aged 17 years and under	Aged 35 to 44 years	12.0033	0.2949	-4.1084	28.115	False
Aged 17 years and under	Aged 45 to 54 years	5.0989	0.9	-11.941	22.1388	False
Aged 17 years and under	Aged 55 to 64 years	0.4711	0.9	-17.7434	18.6855	False
Aged 17 years and under	Aged 65 years and older	-3.62	0.9	-28.164	20.9241	False
Aged 18 to 24 years	Aged 25 to 34 years	8.7288	0.3991	-3.997	21.4546	False
Aged 18 to 24 years	Aged 35 to 44 years	4.9249	0.9	-8.0358	17.8856	False
Aged 18 to 24 years	Aged 45 to 54 years	-1.9795	0.9	-16.0775	12.1184	False
Aged 18 to 24 years	Aged 55 to 64 years	-6.6074	0.8572	-22.1045	8.8897	False
Aged 18 to 24 years	Aged 65 years and older	-10.6984	0.7757	-33.2994	11.9026	False
Aged 25 to 34 years	Aged 35 to 44 years	-3.8039	0.9	-16.1836	8.5758	False
Aged 25 to 34 years	Aged 45 to 54 years	-10.7083	0.2289	-24.2741	2.8574	False
Aged 25 to 34 years	Aged 55 to 64 years	-15.3362	0.0417	-30.3507	-0.3216	True
Aged 25 to 34 years	Aged 65 years and older	-19.4272	0.1334	-41.7001	2.8457	False
Aged 35 to 44 years	Aged 45 to 54 years	-6.9044	0.7281	-20.6908	6.8819	False
Aged 35 to 44 years	Aged 55 to 64 years	-11.5323	0.2745	-26.7464	3.6819	False
Aged 35 to 44 years	Aged 65 years and older	-15.6233	0.3778	-38.0313	6.7846	False
Aged 45 to 54 years	Aged 55 to 64 years	-4.6278	0.9	-20.8218	11.5661	False
Aged 45 to 54 years	Aged 65 years and older	-8.7189	0.9	-31.8032	14.3655	False
Aged 55 to 64 years	Aged 65 years and older	-4.091	0.9	-28.0555	19.8734	False

Discussion

The results of the statistical tests indicate that there are significant differences in the booking and strip search rates across different demographic groups. These findings raise important questions about the potential for bias and discrimination in policing practices and highlight the need for further research and policy reform. The following sections discuss the implications of these findings and suggest avenues for future research.

Limitations

There are several limitations to this study. First, the dataset only includes information on individuals who were arrested and brought into custody by the Toronto Police Service and therefore does not provide a complete picture of police interactions with the public. Second, the dataset does not include information on the circumstances surrounding each arrest, which may influence the likelihood of a booking or strip search. Finally, the dataset is limited to a specific

time period and geographic area (Toronto), which may limit its generalizability to other locations and time periods.

Implications and Future Research

The findings of this study have important implications for policy and practice in the Toronto Police Service and other law enforcement agencies. The results suggest that certain demographic groups, including males, Black individuals, and younger individuals, may be disproportionately impacted by police booking and strip search practices. These practices can have serious consequences for individuals and communities, including loss of freedom, harm to personal dignity, and negative impacts on mental health and well-being.

To address these issues, there is a need for further research and policy reform. One potential avenue for future research is to explore the underlying factors that contribute to these disparities in policing practices. For example, the researchers could examine the role of implicit biases and stereotypes in shaping police behaviour, as well as the impact of systemic factors such as racial profiling and the over-policing of certain communities. By better understanding the factors that contribute to disparities in policing practices, researchers and policymakers can develop evidence-based interventions to address these issues and promote more equitable and just policing practices.

In addition to research, there is a need for policy reform to address the issues highlighted in this study. For example, policies could be developed to reduce the use of strip searches and to limit the circumstances under which they can be conducted. Additionally, policies could be developed to increase transparency and accountability in policing practices, such as requirements to report on the demographics of individuals subjected to booking and strip searches. These policies could help to promote greater equity and fairness in policing practices and reduce the potential for bias and discrimination.

Another potential avenue for policy reform is to increase diversity and cultural competency within law enforcement agencies. Research has shown that police officers from diverse backgrounds are better able to build trust and rapport with the communities they serve, and that cultural competency training can help officers to better understand and navigate cultural differences. By increasing the diversity and cultural competency of law enforcement agencies, it may be possible to reduce the potential for bias and discrimination in policing practices and promote more equitable and just policing practices.

Finally, it is important to recognize that the issues highlighted in this study are part of a broader pattern of systemic racism and inequality in our society. To address these issues, there is a need for broader social and political change that goes beyond the realm of law enforcement. This could include efforts to reduce economic and social inequality, increase access to education and

healthcare, and promote diversity and inclusion in all areas of society. By addressing the root causes of systemic inequality, it may be possible to reduce the potential for bias and discrimination in all areas of our society, including law enforcement.

Conclusion

The results of this study provide important insights into the relationship between demographic factors and booking and strip search rates in the Toronto Police Service. The findings suggest that gender, race, and age play significant roles in these practices, and highlight potential issues of bias and discrimination in policing. These findings underscore the need for further research and policy reform to address these issues and promote more equitable and just policing practices. By working to address the root causes of systemic inequality and promoting greater equity and fairness in all areas of our society, we can help to create a safer and more just world for all people.

References

- Alpert, G. P., & Dunham, R. G. (2004). The effect of race and ethnicity on traffic stops and searches in Phoenix, Arizona. *Policing: An International Journal of Police Strategies & Management*, 27(3), 389-416.
- Arrests and strip searches (RBDC-arr-TBL-001). Toronto Police Service Public Safety Data Portal. (n.d.). Retrieved February 27, 2023, from <https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about>
- Brunson, R. K. (2007). "Police don't like Black people": African American young men's accumulated police experiences. *Criminology & Public Policy*, 6(1), 71-102.
- Eitle, D., & Turner, R. J. (2002). Exposure to official bias and socialization into deviance: The case of marijuana use. *Justice Quarterly*, 19(1), 113-146.
- Ghandnoosh, N. (2015). The war on marijuana in black and white. The Sentencing Project. Retrieved from <https://www.sentencingproject.org/publications/the-war-on-marijuana-in-black-and-white/>
- Goff, P. A., Lloyd, M. H., Geller, A., Raphael, S., & Glaser, J. (2016). The science of justice: Race, arrests, and police use of force. *Journal of Social Issues*, 72(4), 802-823.
- Lee, H. Y., Yoo, J. W., & Park, J. (2019). The effects of race/ethnicity on police use of force: An analysis of officer-involved shootings in the United States. *Journal of Ethnicity in Criminal Justice*, 17(2), 101-120.
- Miller, J., & Schwartz, J. (2018). Gender differences in police stops and searches: Evidence from the British Crime Survey. *European Journal of Criminology*, 15(5), 537-554.
- Rosenfeld, R., & Fornango, R. (2013). The impact of police stops on precinct robbery and burglary rates. *Criminology*, 51(4), 919-948.
- Toronto Population 2021/2022. (n.d.). Canada Population. Retrieved from <https://canadapopulation.org/toronto-population/>
- Weitzer, R., & Tuch, S. A. (2006). *Race and policing in America: Conflict and reform*. Cambridge University Press.