University of Toronto

INF 2178H LEC0101: Experimental Design for Data Science

Final Project

Winter 2023

Instructor: Shion Guha

Group Number: Group 26

Group Members:
Chris Liu
Qingzhou Li

Academic Department:
Faculty of Information

Academic Program:
Master of Information

April 16, 2023

Table of content

Abstract

The purpose of this paper is to investigate the relationship between demographic factors, such as age and race, and the likelihood of being strip searched by the police. We used the *The Arrests and Strip Searches* dataset from Toronto Police Service and ran a descriptive analysis, t-tests, and one-way and two-way ANOVA to see if gender, age and race have an effect on the decision making of conducting a strip search. After that, to further research why the police would strip search those groups, we utilized power analysis, ANCOVA, logistic regression with a confusion model. According to the results, there is a significant difference between race, age and the possibility of being strip searched.

The study suggests that the Toronto Police Service needs to increase accountability, better supervise and coach officers, enhance training and development of performance metrics. The study highlights the benefits of using several statistical methods and logistic regression in analyzing strip search data, and these techniques can help to identify predictors and control for confounding variables. The research concludes by calling on the judiciary to treat people of every race and age equally and to further improve the regulations and rules for judicial execution. However, the study recognizes that further investigation is necessary to identify the underlying causes of the racial disparities and take effective measures for sustainable change. The limitations of the research are also acknowledged, and it suggests that repeated arrests could increase the likelihood of experiencing strip searches.

The code can be seen in the link of the colab.

Introduction

"In the case of Canadian policing, strip searching refers to when an arrested or detained person is compelled by police officers to expose their most intimate body areas" (Lemke, 2022a). In other words, a strip search is an act that is enforced when a person is arrested by the police without the consent of the person concerned. The purpose of the strip search is to maintain the safety and health of society and the judicial practice of the police on those who have potential threats.

However, whether it is necessary to conduct such frequent strip searches is a question we need to consider. "Supreme Court of Canada's recognition of the fact that strip searches are intrinsically humiliating, degrading, and dehumanizing, and bring up racial and sexual trauma in the process" (Lemke, 2022a). Toronto has a much higher rate of strip searches than other

provinces. “The Toronto police had strip-searched 37% to 43% of all arrestees between 2014 and 2016, whereas other large Ontario police forces reported strip-search rates of under 1% of people arrested” (Lemke, 2022a).

Our research found that demographic factors such as race, age group and gender influence police’s decisions on strip searches significantly. From a sociological perspective, we are interested in examining the impact of three demographic factors (gender, age and race) on the number of strip searches. The primary purpose of this report is to identify how those factors influence disparities in the number of strip searches and how to make improvements on judiciary policies acceptably. In addition, people start from being arrested to be booked in a police office, and finally be strip searched, so we would like to investigate the relationship within the sequential process to see if there is any pattern that people are strip searched. We also would like to identify if any factors would influence the possibility of items found when doing strip search on people who got arrested. There are many reasons to conduct strip search, so we want to gain insights about the thoroughness of the search is critical in determining the likelihood of an item being found.

Literature review

The Canadian police are authorized by law, both through statutes and common law, to perform personal searches, including strip searches. Although a frisk search is less invasive, a strip search can be highly intrusive and degrading. It is the responsibility of the police to provide justification for the strip search and ensure that it is conducted in a reasonable manner. The Supreme Court has provided guidelines for the police to follow during strip searches, as the manner in which the search is conducted plays a significant role in determining its reasonableness.

The article, *Policing Toronto: Strip Searching in a Divided City,* reveals that the high frequency of strip searches is strongly associated with racial disparities. “The recent race-based statistics on strip searches confirm the over-policing of Black people relative to the rest of the population. Even though Black people make up around 10 percent of Toronto’s population, they make up one in every three strip-searched people” (Lemke, 2022a). Black racial identity seems more likely to be strip-searched after getting arrested. Therefore, that evidence indicates racial disparities would influence police strip search decisions.

Moreover, according to a report published in 2019, Toronto Police continued to conduct illegal strip searches despite a court ruling prohibiting them from doing so without proper grounds (Smith, 2019). The report found that out of the 1,242 strip searches conducted by the police, only 23% were carried out in accordance with the established guidelines. The report also highlighted that black and Indigenous people were more likely to be subjected to strip searches. The Toronto Police Services Board acknowledged the issue and stated that steps were being taken to address the problem.

Also, in 2020, the Toronto Police reported that they were conducting far fewer strip searches following new rules and guidelines established in response to previous criticism (Gillis, 2020). According to the interim Chief of Police, the number of strip searches had decreased by 96%, from approximately 600 per month to just 25 per month. The new rules require officers to have reasonable grounds for conducting a strip search and to document the search in detail. The Toronto Police also implemented training programs for officers to ensure that they understand the proper procedures for conducting strip searches. These changes were implemented in response to concerns about the excessive and improper use of strip searches, particularly against Black and Indigenous individuals.

In addition, there are differences in the likelihood of strip searches by offence type. “Search rates and items found rates were highest for drug-related offences, followed by break & enter and weapons and homicide-related offences. While search rates were 6 lowest for arrests related to robberies & thefts when they did occur items found tended to be higher than average” (Toronto Police Service, n.d.). We found that most of the strip searches related to drugs and crimes are effective, and the police have a high probability of finding harmful items.

Original raw dataset

The dataset we used for this paper is *Arrests and Strip Searches,* which is from Toronto Police Service. This dataset records information about each arrest from January 2020 to December 2021. It contains 65276 arrest cases by 25 factors, including 7,801 strip searches and 34,483 booked cases. This dataset records basic information such as the gender, race and age group of individuals and details about situations and reasons for the arrest. During our research, after a person was arrested, they may be subject to a strip search. They must also be booked into custody at a police station as an intermediate outcome in the pathway to a strip search. We are researching demographic factors such as race and age that would influence the

likelihood of strip searches. Then, in the following sections, we will utilize those factors to discover their relationships with strip searches. This report provides EDA, methodologies, findings and discussions of the data analysis of this dataset.

Research Objectives and Questions

In Phase 1, we are interested in figuring out if the strip search decisions are differentiated based on certain factors, and if there is a significant variance between groups. Based on our research, we focus on the following research questions for our study. The core question is, how do demographic factors affect the number of strip searches? Based on some common sense, we choose some attributes to sort the cases into different groups based on demographic factors such as gender, age and race that influence disparities in the probability of strip searches.

In Phase 2, we would like to test the validity of the initial analysis and dig more deeply into the correlations between different variables by identifying potential causality and assessing their predictive power. We aim to investigate how changes in one variable may impact others and to examine the underlying relationships between variables. For example, it is considered whether there is a certain relationship between the strip search reason and the item found. We plan to conduct research to uncover any underlying patterns or trends that were not apparent in the initial analysis, with the ultimate goal of gaining a deeper understanding of the dataset and informing our decision-making process.

Phrase 1

1. Are there Gender disparities in reported strip search incidents? Gender disparities (comparisons between male and female): To what extent, if any, are certain racial groups more likely than others to experience a strip search?

2. Are there age group disparities in reported strip search incidents? Racial disparities (comparisons between different racial groups such as youth, adult, and middle-aged): To what extent, if any, are certain racial groups more likely than others to experience a strip search?

3. Are there racial disparities in reported strip search incidents? Racial disparities (comparisons between different racial groups): To what extent, if any, are certain racial groups more likely than others to experience a strip search?

4. Are there any interactions between the two factors listed above that lead to disparities in strip search incidents? The interaction disparities: To what extent, if any, are certain combinations more likely than others to experience a strip search?

Phrase 2

5. How does the power of a study vary as the sample size, effect size, or level of significance is changed? What is the optimal sample size required to detect a statistically significant difference between groups with a specified level of power and significance?

6. Is there a significant difference in the possibility of a strip search based on the perceived race of the person, after controlling for the possibility of being booked in the police office?

7. What factors are most strongly associated with the probability of finding items after being strip searched? Are there certain demographic or situational variables that increase the odds of an item being found?

EDA

Descriptive statistics

Raw dataset

From Figure 1 to 4, it is obvious that more males got arrested by the Toronto Police Service than females. White people and Black people account for most arrest cases, compared to other races. As for the age group, the adult group (25 - 44 years old) got arrested much more than other age groups, and the least age group is the elderly. Lastly, according to Figure 4, most arrests happen when people are committed in assault and other crimes against a person, which is followed by Robbery & theft and FTA/FTC, Compliance Check & Parollee crime groups. Crimes against children and homicide got the least arrest cases.

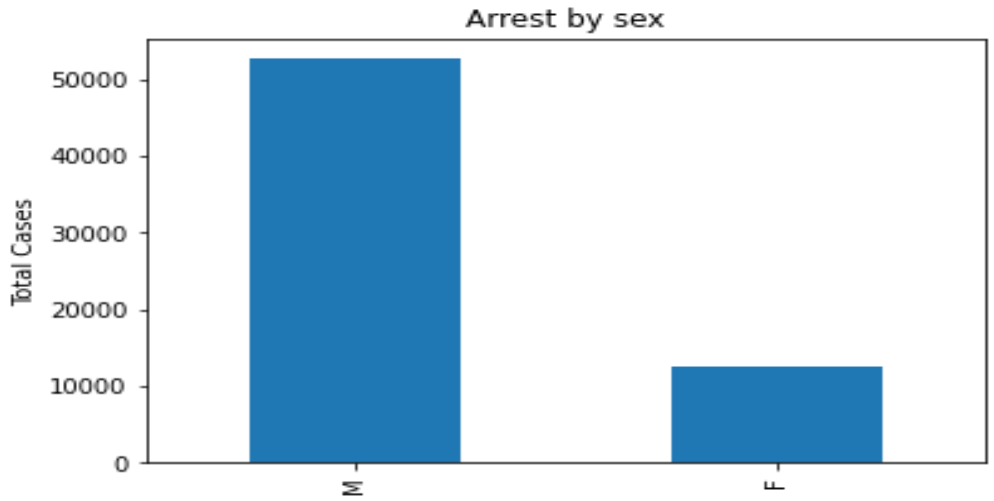

Figure 1.1 *Arrest cases by gender*

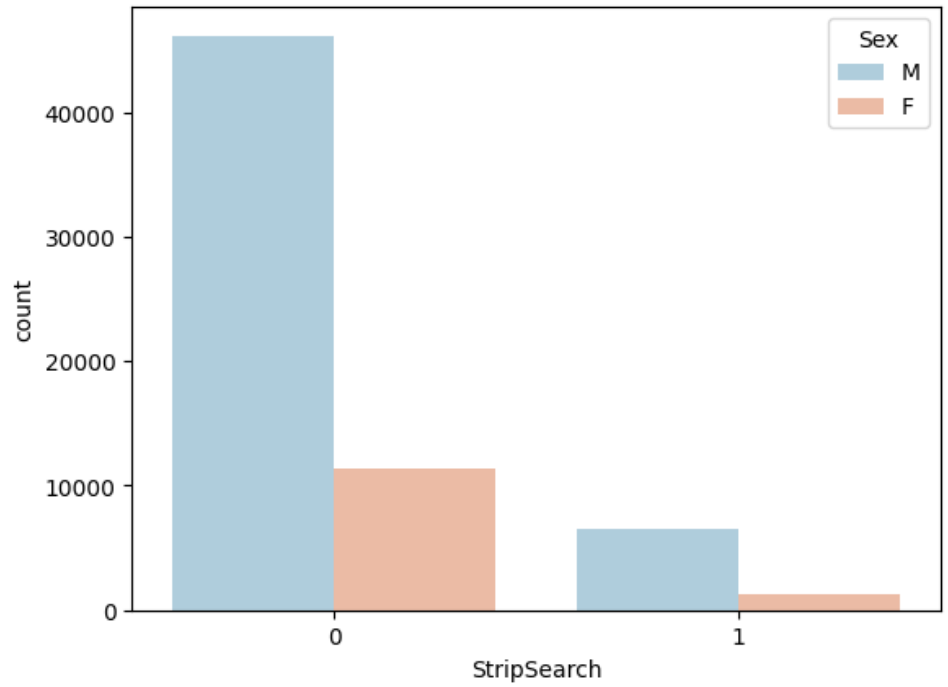

Figure 1.2 *stripsearch by gender*

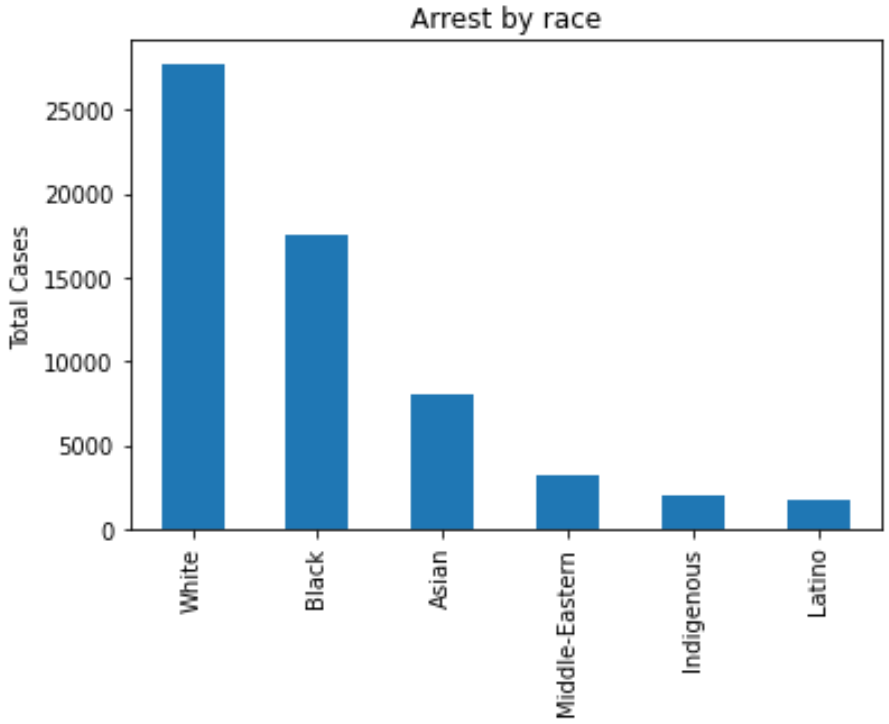

Figure 2.1 *Arrest cases by races*

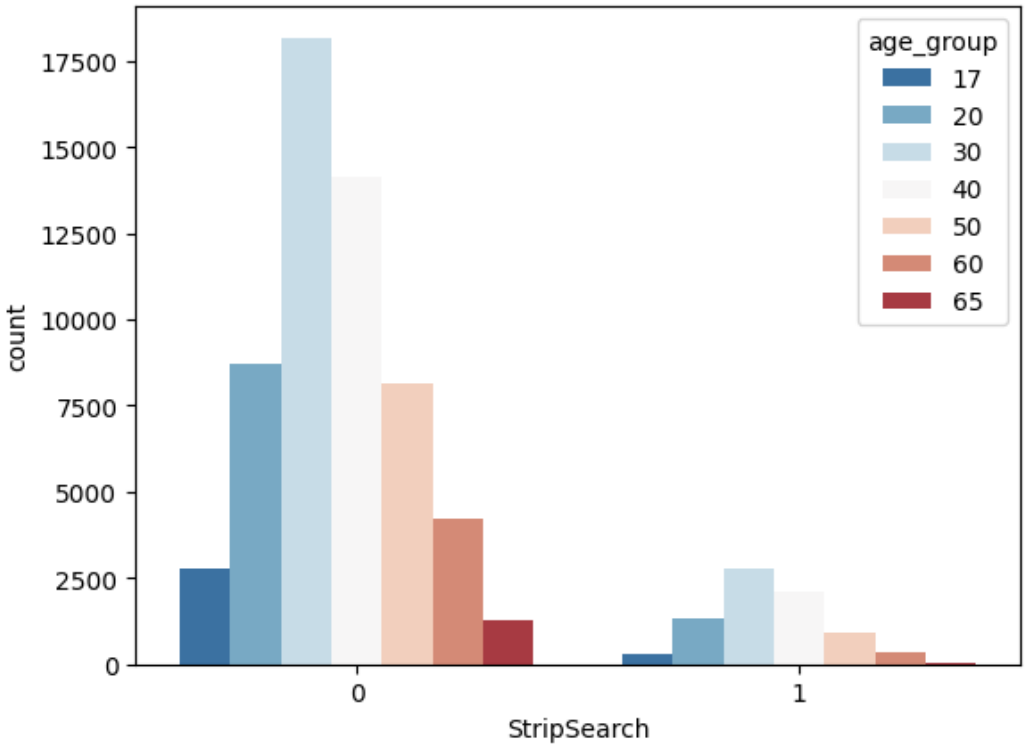

Figure 2.2 *stripsearch by races*

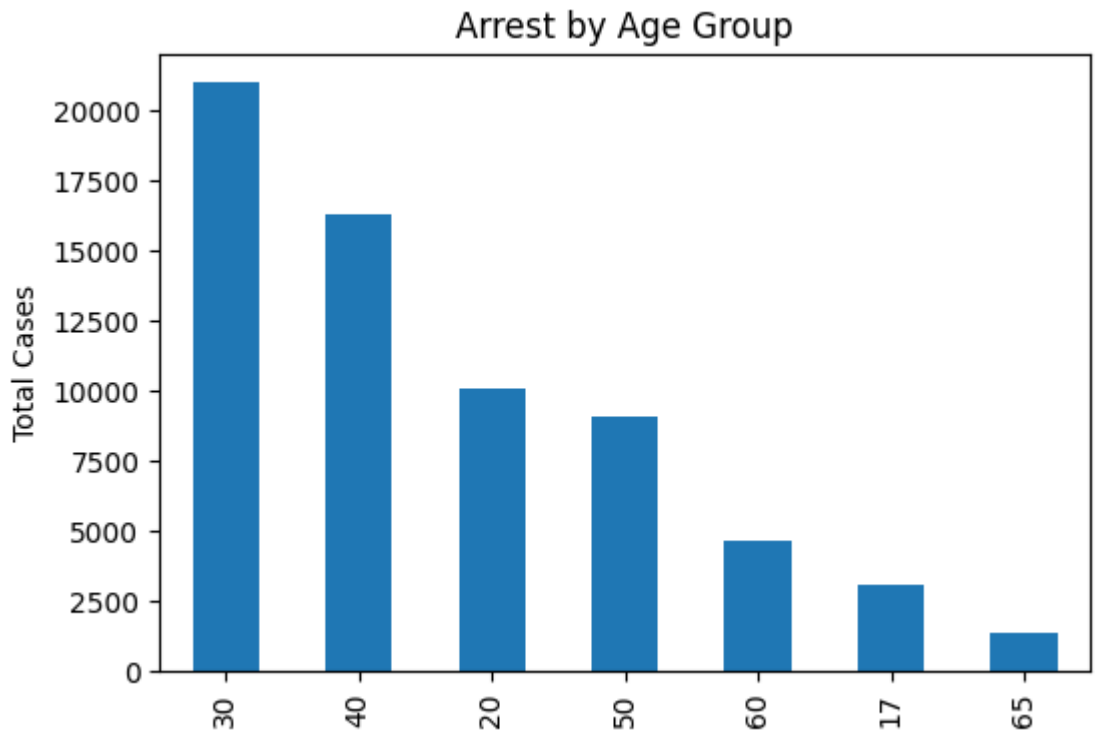

Figure 3.1 *Arrest cases by age group*

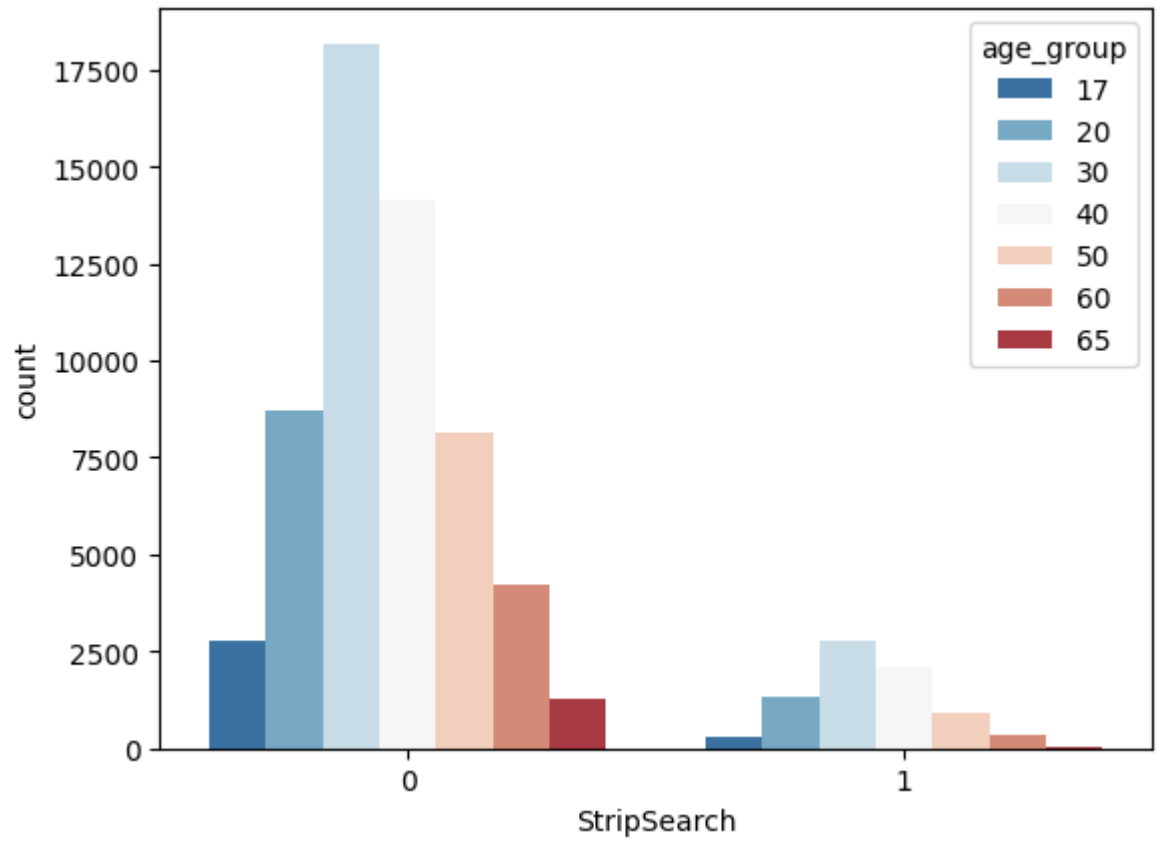

Figure 3.2 *stripsearch by age group*

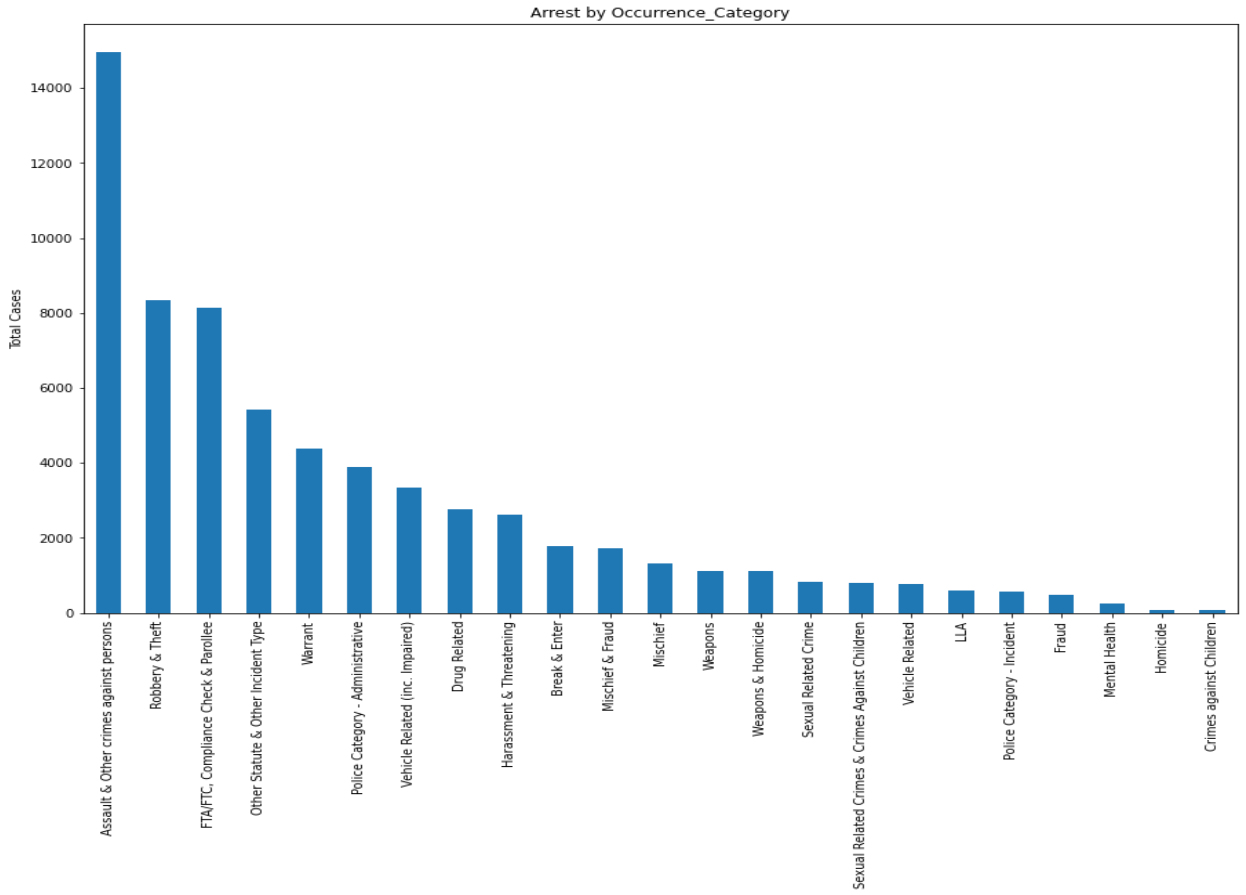

Figure 4 *Arrest cases by Occurrence category*

Stripsearch distribution

From Figure 5, we can find that the percentage of indigenous people who got strip research is the highest, compared with the strip search distribution of other race groups, especially the nearly 17.5% of indigenous males who got strip searched. The percentage of black males who got strip searched was followed by nearly 15%, but just under 10% of black females got strip searched. A few Asian females and Latino females got strip-searched, the number is around 3%. As for Figure 6, apparently around 13% of male adults and young adults experienced strip searches, which is higher than other groups. And the number of elderly people who are strip searched was the least, only accounting for 3% for males and 2% for females.

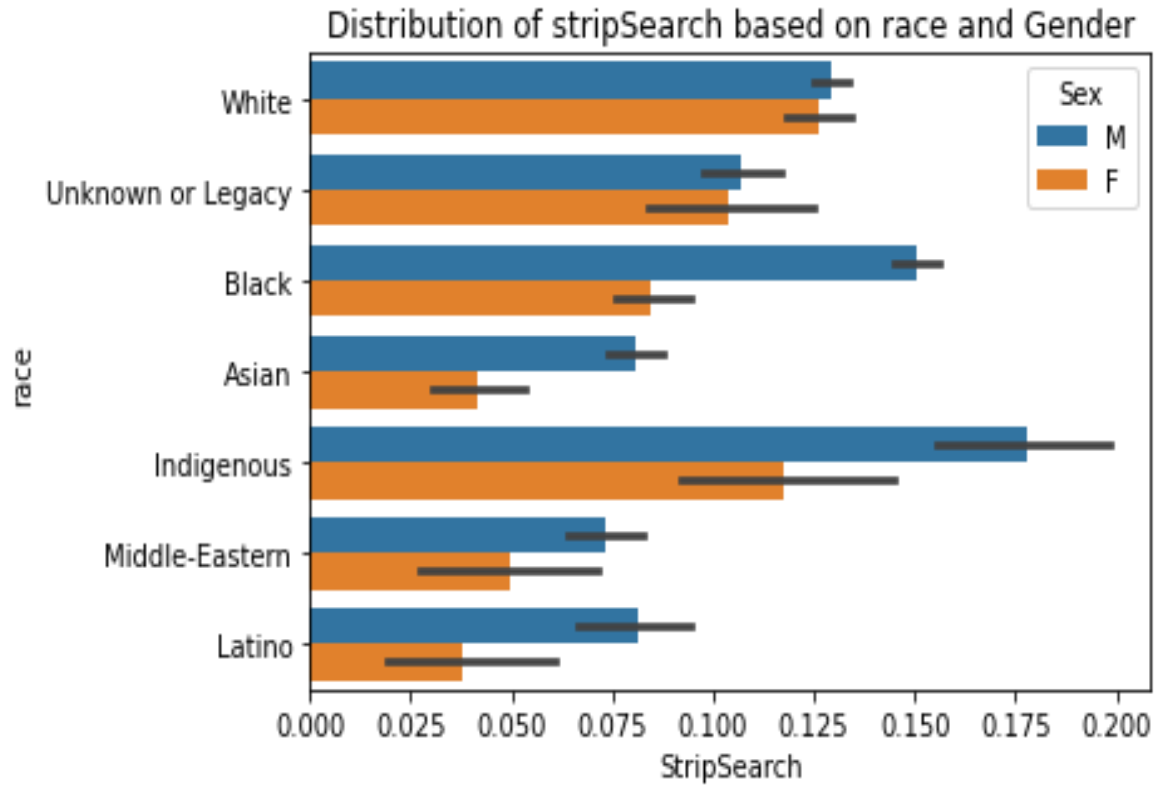

Figure 5 *strip search distribution based on gender & race*

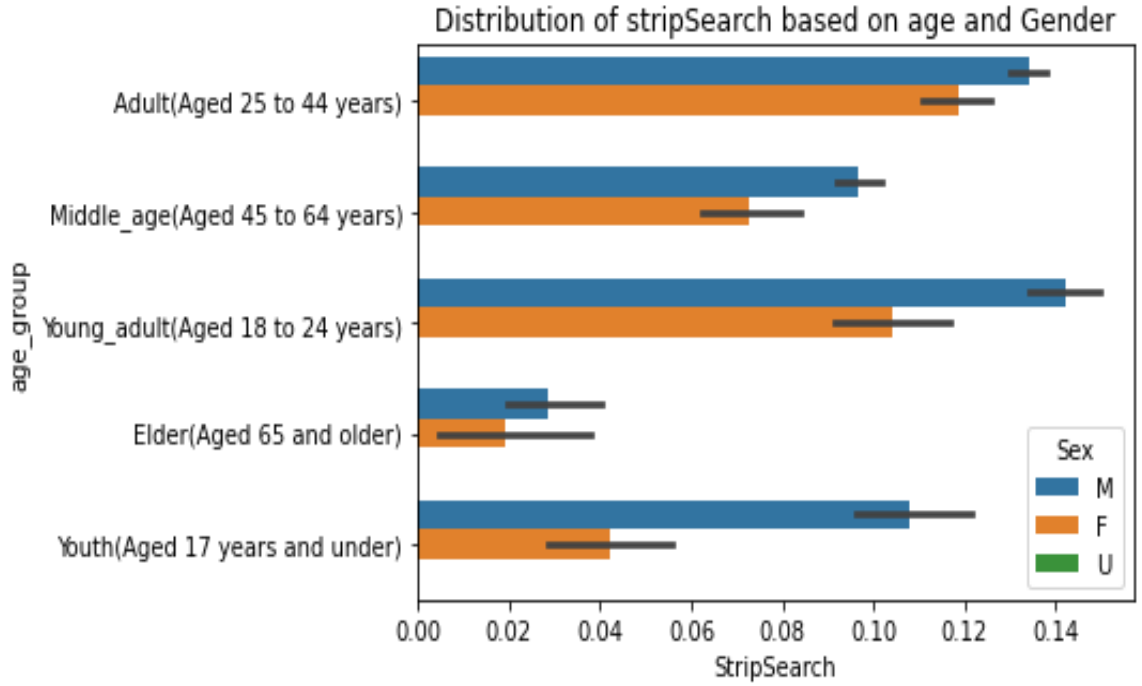

Figure 6 *strip search distribution based on gender & age_group*

Dataset using for further analysis

We assume the rate of stripsearch for a certain offence behaviour in a certain time would be the same for all arrests, so we calculate the strip search ratio by using the number of strip searches divided by the number of arrests that have happened.

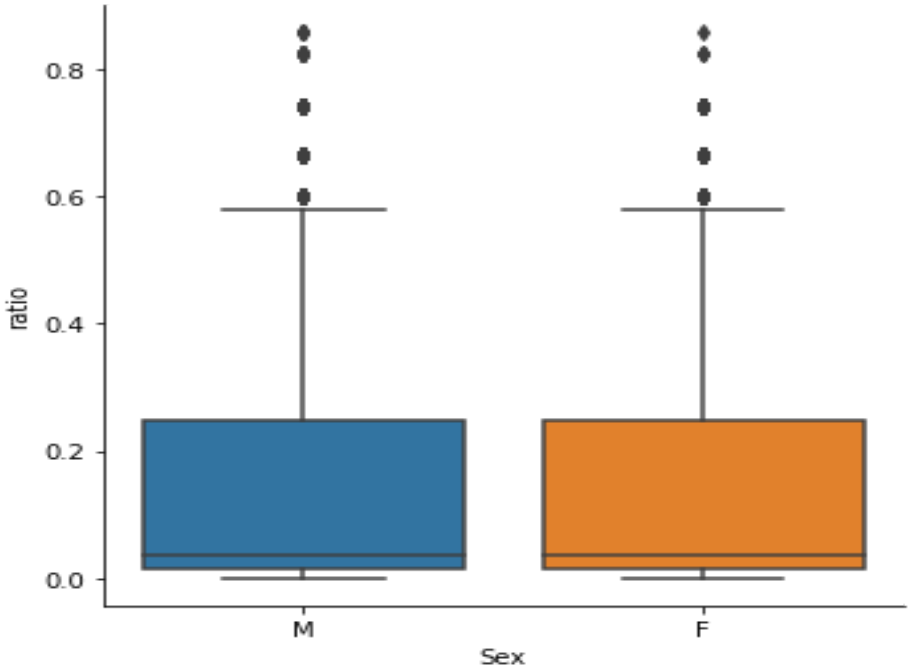

Figure 7 *strip search ratio by gender*

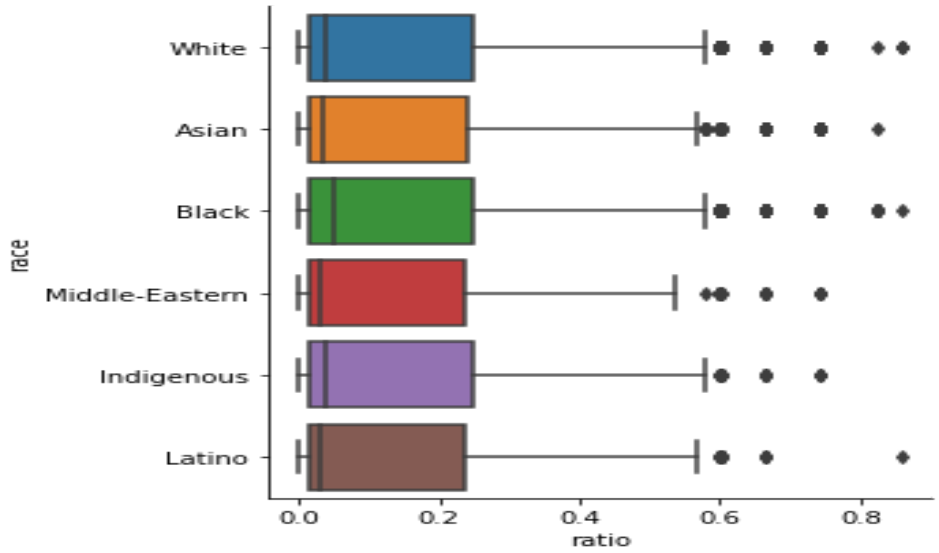

Figure 8 *strip search ratio by race*

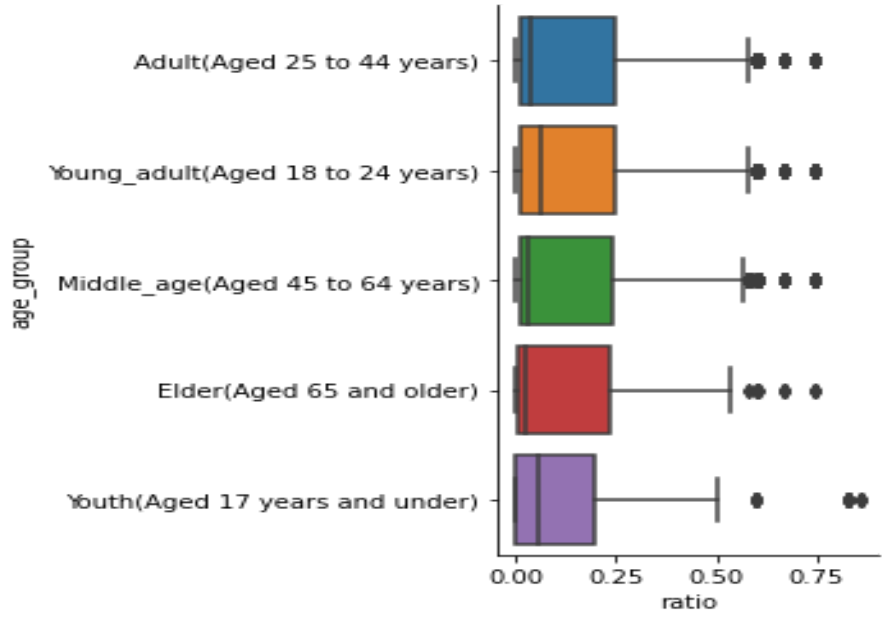

Figure 9 *strip search ratio by age_group*

According to figures 7 to 9, we can detect that the ratio is close to 0, which means most people are not easily strip searched. However, in some cases, people got a higher ratio to be strip searched. To be specific in certain figures, figure 7 shows there are no obvious differences between male and female, both groups experience some outliers. The similar pattern happens in people grouped by race, but white people, Black people and Asian people would come out with some higher ratios. As for age groups, the youth group will have a higher ratio possibility, but the ratio of other groups is close to the 0, especially the middle age and elderly people.

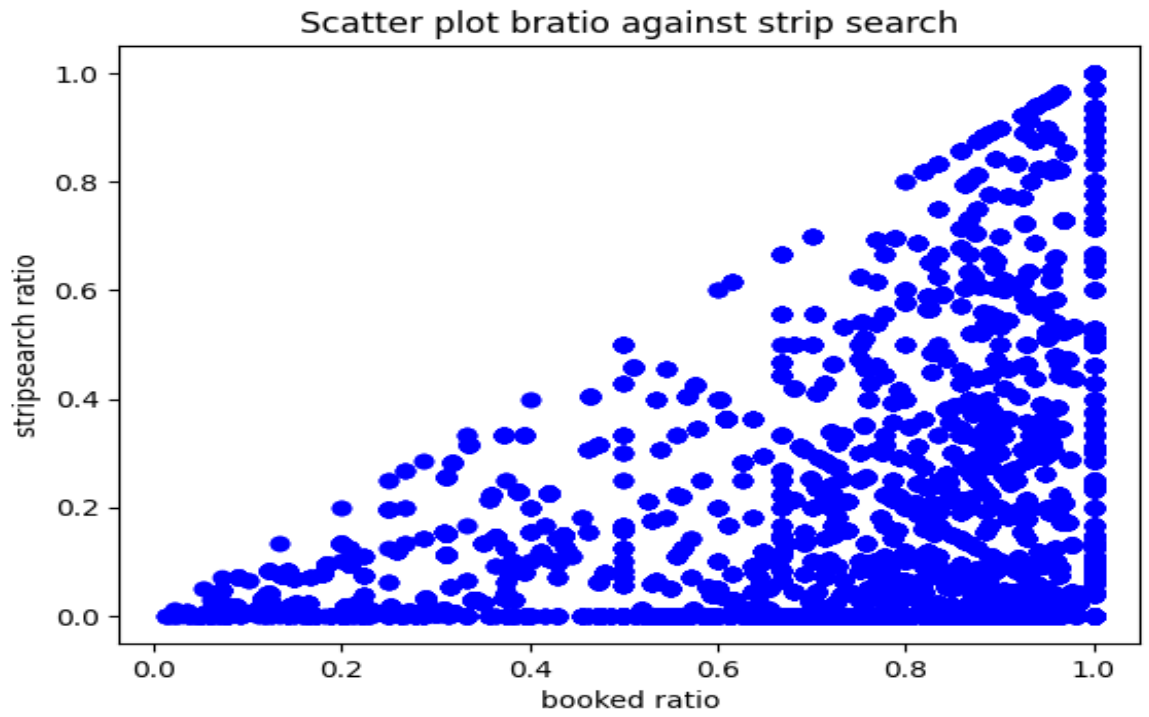

Figure 10 *Scatplot by bratio and ratio*

According to figure 10, the strip search turn over ratio increases with the booked ratio increase, which means there is a positive relationship between those two variables. On the other hand, the number of scatter plots is more concentrated toward 1. It make sense as the strip search would be implemented after the booked event, so there are more deeply connections between booked ratio and stripsearch ratio needed to be digged out.

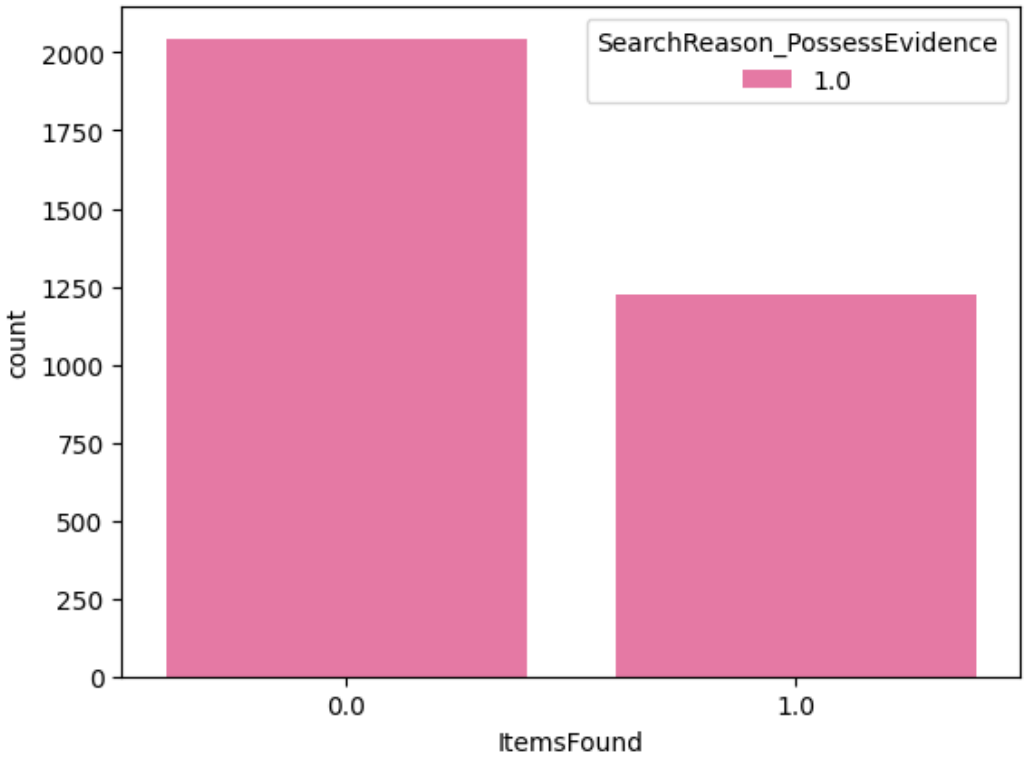

Figure 11.1 *itemfound by SearchReason_PossessEvidence*

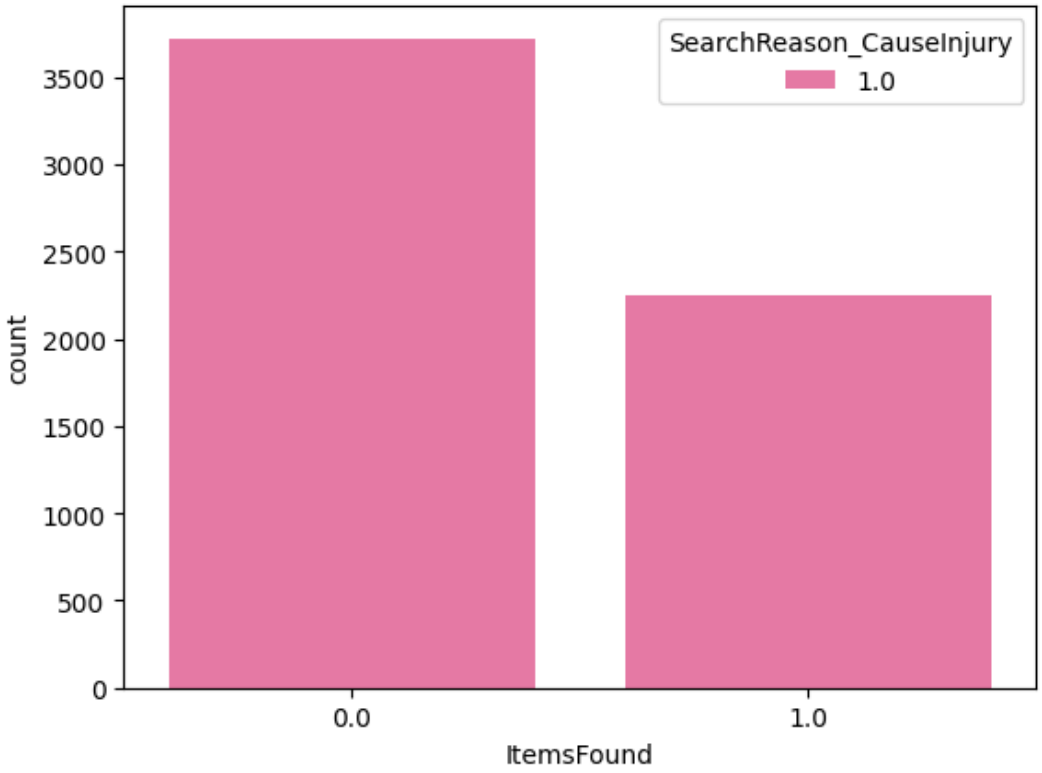

Figure 11.2 *itemfound by SearchReason_AssistEscape*

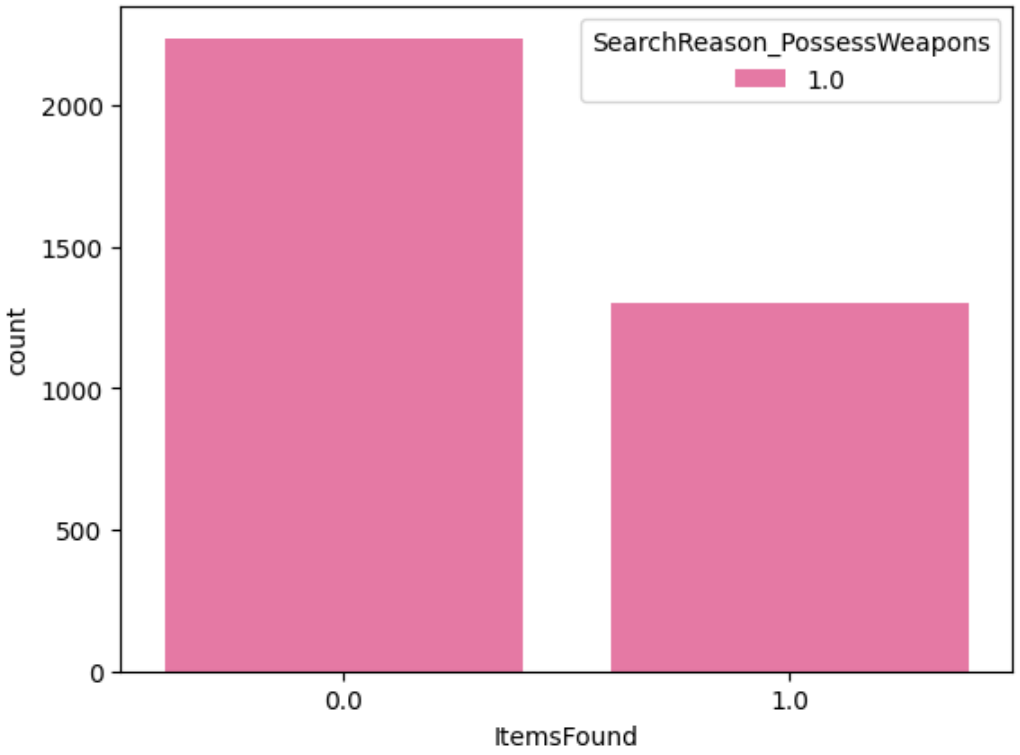

Figure 11.3 *itemfound by SearchReason_AssistEscape*

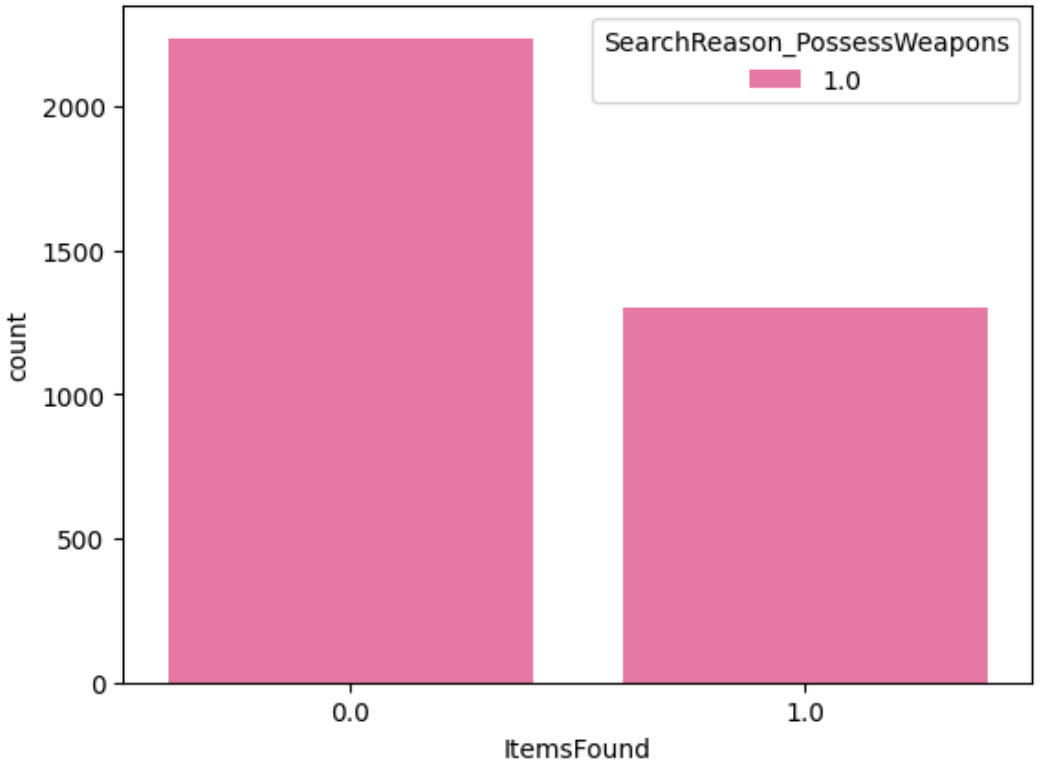

Figure 11.4 *itemfound by SearchReason_PossessWeapons*

From figure 11.1-figure 11.4, we can find that some items can be found in nearly 30% of strip

search cases. Depending on different strip search reasons, the actual case could be various. For example, high possibility of that item would be found when the stripsearch reason causes injuries. However, since police officers conduct the strip search with multiple reasons, it is required to make further research on the relationship between item found and strip search reasons.

T-tests

We conducted t-tests to compare the mean probability of strip searches for two sample groups and to confirm if there is a significant difference between the means of two groups and how they are correlated with each other. The p-value can be calculated to determine the likelihood of observing the observed difference or a more significant difference if the null hypothesis were true. If the p-value is below the significance level (usually 0.05), then the null hypothesis can be rejected, and it can be concluded that a significant difference exists in the number of strip searches between the two groups.

Gender

Null hypothesis: There is no difference between males and females in the likelihood of strip searches.
Alternative hypothesis: There is a difference in the likelihood of strip searches between males and females.

The results indicate that the mean of the number of strip searches for males (M=0.139, SD=0.169) is lower than for females (M=0.145, SD=0.171). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.02) is less than 0.05, 95% CI [-0.01, -0.0007]. Therefore, we can reject the null hypothesis that there is no difference in the probability of strip searches for males and females.

Table 1. Results of T-test by Gender

Variables	**Mean**	**SD**	**t**	**p-value**
Gender			-2.27	0.02 *
Male	0.139	0.169		
Female	0.145	0.171		

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Age group: *30s v.s. Other age groups*

Null hypothesis: There is no difference in the probability of strip searches between 30s adults and other age groups.

Alternative hypothesis: there is a difference in the probability of strip searches between 30s adults and other age groups.

The results indicate that the p-value (0.89) is greater than 0.05. Therefore, we cannot reject the null hypothesis that there is no difference in the ratio of strip searches for adults and other age groups.

Age group: *20s v.s. Other age groups*

Null hypothesis: There is no difference in the probability of strip searches between 20s adults and other age groups.

Alternative hypothesis: there is a difference in the probability of strip searches between 20s adults and other age groups.

The results indicate that the mean of the ratio of strip searches for young adults (M=0.19312, SD=0.01714) is lower than for other groups (M=0.1688, SD=0.03150). With alpha established at 0.05, this is a statistically significant difference as the p-value is much less than 0.05, 95% CI [0.01714, 0.0315]. Therefore, we can reject the null hypothesis that there is no difference in the ratio of strip searches for young adults and other age groups.

Table 2. Results of T-test by Age group

Variables	**Mean**	**SD**	**t**	**p-value**
Age group			-0.1321	0.89
30s	0.17232	0.24821		
Other Group	0.17267	0.23717		
Age group			6.6397	3.37e-11 ***

20s	0.19312	0.01714
Other Group	0.1688	0.03150

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Race: *White v.s. Other racial groups*

Null hypothesis: There is no difference in the probability of strip searches between white and other racial groups.

Alternative hypothesis: there is a difference in the probability of strip searches between white and other racial groups.

The results indicate that the mean ratio of strip searches for white (M=0.144, SD=0.172) is higher than for other groups (M=0.136, SD=0.167). With alpha established at 0.05, this is a statistically significant difference as the p-value is much less than 0.05, 95% CI [0.0039, 0.0114]. Therefore, we can reject the null hypothesis that there is no difference in the ratio of strip searches for white and other racial groups.

Race: *White v.s. Black*

Null hypothesis: There is no difference in the probability of strip searches between white and black.

Alternative hypothesis: There is a difference in the probability of strip searches between white and black.

The results indicate the p-value (0.00) is less than 0.05. Thus, we can reject the null hypothesis that there is no difference in the ratio of strip searches for white and black people arrested by the police service.

Table 3. Results of T-test by Race

Variables	**Mean**	**SD**	**t**	**p-value**
Race			4.01	0.00 ***
White	0.144	0.172		
Other race	0.136	0.167		

Race			-1.331	0.183
White	0.144	0.172		
Black	0.147	0.175		

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Booked ratio

Null hypothesis: There is no difference in the probability of strip searches between groups with different booked ratios.

Alternative hypothesis: There is a difference in the probability of strip searches between groups with different booked ratios.

The results indicate the p-value (0.000) is much lower than 0.05. Thus, we can reject the null hypothesis that there is no difference in the ratio of strip searches for white and black.

Table 4. Results of T-test by bratio

Variables	**Mean**	**SD**	**t**	**p-value**
Race			68.33	0.00 ***
High	0.195	0.249		
Low	0.06	0.089		

Methods

Since most of the raw data in the original dataset are categorical variables, it is difficult to conduct a precise t-test and ANOVA. Thus, the raw data are processed using multiple categorical groups to generate a new dataset with a continuous variable. We assume that the frequency of strip searches for a particular offence during a certain period is consistent for all arrests. Thus, we determine the strip search ratio by dividing the number of strip searches conducted by the total number of arrests made. Additionally, we eliminated cases of individuals not in custody from the original dataset since strip searches are only conducted on individuals who have been booked in the policy office.

In our case study phase 1, in addition to One-Way ANOVA to compare the means of the

individual groups, we also use Two-Way ANOVA to investigate combinations of factors that may be statistically significant. This will provide a more thorough data analysis and help us identify any significant differences or patterns between the groups.

The alpha value is set prior to performing the hypothesis test and determines the acceptable level of error. In this report, a common alpha value of 0.05 will be used. If the consequences of making an error are insignificant, a higher alpha value may be acceptable, whereas a lower alpha value should be used for cases where errors can have severe consequences, such as medical diagnoses or criminal sentencing.

Also, we process the p-value to measure the significance of evidence against the null hypothesis. The smaller the p-value, the stronger the evidence in favour of rejecting the null hypothesis. The p-value is then compared to the alpha value set for the statistical test. If the p-value is less than the alpha value we set, then the null hypothesis is rejected, whereas if the p-value is greater than it, the null hypothesis cannot be rejected.

In the phase 2 analysis stage, we tried to use power analysis, ANCOVA and Logistic regression with confusion matrix and prediction interval to conduct more comprehensive studies on the questions which can not be completed in initial studies.

The power of a study is the probability of correctly rejecting a null hypothesis when the alternative hypothesis is true. It depends on several factors, including the sample size, effect size, and level of significance. Power analysis is a useful tool for determining the optimal sample size required to detect a statistically significant difference in a study. By using power analysis, we can ensure that this report has sufficient power to detect meaningful effects, which can improve the reliability and validity of the research findings.

ANCOVA (Analysis of Covariance) is a statistical technique that is used to determine whether there is a significant difference between the means of two or more groups after controlling for the effects of one or more continuous variables, known as covariates. In the context of analyzing the race and strip search ratio, the booked ratio can be considered a potential covariate. The booked ratio is the proportion of people who are booked or arrested for a crime, and it may be related to both the race of individuals and the likelihood of being subjected to a strip search. For example, if individuals of a particular race are more likely to be arrested than others, this may confound the relationship between race and strip search. By including the booked ratio as a covariate in the ANCOVA, we can control for this potentially

confounding variable and obtain a clearer picture of the relationship between race and strip search. In essence, the ANCOVA with the booked ratio as a covariate allows us to determine if there is a significant difference in strip search rates between different races, after taking into account the potential influence of differences in the booked ratio. It can provide more reliable and accurate results by accounting for the effects of a potentially confounding variable, and thus improve the validity of the analysis.

Logistic regression is a type of statistical analysis commonly used in machine learning for predicting the probability of a categorical outcome based on one or more predictor variables. In the context of a strip search and item found a dataset, logistic regression can be used to predict the probability of finding an item on a person during a strip search based on various factors such as the person's gender, age, ethnicity, etc. Logistic regression is particularly well-suited for binary classification tasks, which is often the case in a strip search and item-found dataset. In other words, logistic regression can help predict whether or not an item will be found on a person during a strip search. It can also provide insights into which factors are most predictive of finding an item during a strip search.

RQ1-3: Based on our research and statistically significant t-test results, our preliminary analysis of the dataset is that demographic factors, especially age and race are correlated with the likelihood of strip search. In order to explore our research questions 1-3, we will conduct a One-Way ANOVA to test whether there is a significant difference in the mean of the probability of strip search between different groups. If the results are statistically significant, Post hoc comparisons would be used to examine the means by which specific groups are different from each other.

RQ4: For research question 4, we intend to investigate the interactive effect between two independent variables: race and age. Then, we will run Two-Way ANOVA. If the results are statistically significant, post hoc comparisons would be used again to see which groups are different from one another.

RQ5: To determine the optimal sample size required to detect a statistically significant difference before conducting ANCOVA, a power analysis will be used. To perform a power analysis, the following steps are taken in this report. The effect size is typically expressed as Cohen's d, which is the difference between the means of the groups divided by the pooled standard deviation. Then, The level of significance (alpha) is the probability of rejecting the null hypothesis when it is true. It is typically set at 0.05, which corresponds to a 5% chance of

making a Type I error. The power (1 - beta) is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It is typically set at 0.80 or 0.90, which corresponds to an 80% or 90% chance of detecting a true effect. The sample size can be calculated using power analysis formulas or online calculators. For example, the G*Power software can be used to calculate the sample size based on the effect size, level of significance, power, and other parameters. Blow is the power curve.

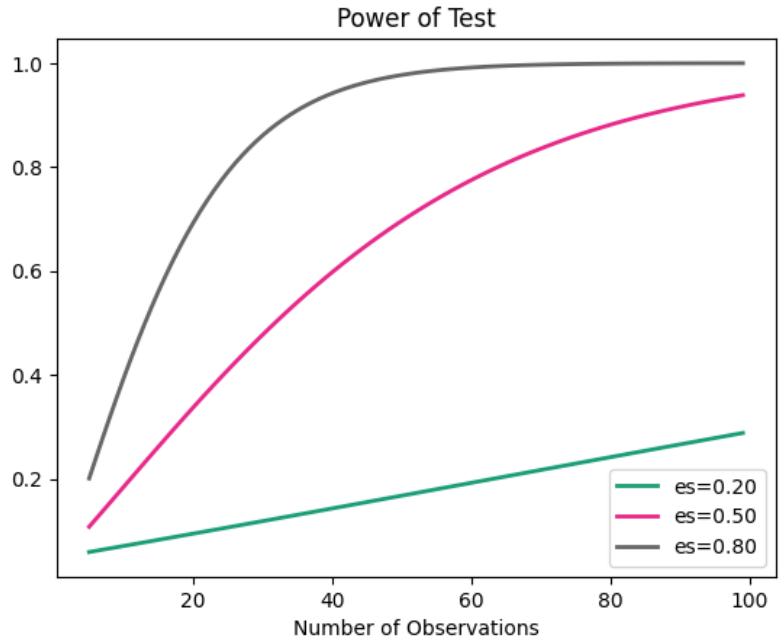

Figure 12 *Power curve*

The power curve allows us to visualize how changes in sample size can impact their ability to detect a meaningful effect. By examining the power curve, researchers can identify the minimum sample size needed to achieve a desired level of statistical power, Additionally, the power curve can help researchers understand the trade-off between sample size and effect size.

RQ6: Through the previous analysis using ANOVA, we found that demographic factors such as race and age do affect the ratio of strip search. However, the significance of the results may be affected by the activity of being booked. In order to further study how these factors affect the ratio of strip search , we set the ratio of being booked as a control variable. We will use ANCOVA to test whether the independent variable still affects the dependent variable after removing the effect of the covariate.

RQ7: To determine the relationship between the likelihood of items found and stipsearch, we will adopt logistic regression and set the probability of finding items as the dependent variable, and the ratio of strip search and search reasons as independent variables. The logistic regression model is estimated using the Logit function from the statsmodels module, which returns a LogitResult object. This object is used to obtain the predicted probabilities by calling the prediction method on the model object, passing the confusion matrix as an argument.

The confidence intervals for the predicted probabilities are estimated using the delta method, which is based on the Taylor series approximation of the standard errors of the predicted probabilities. The standard errors are obtained by multiplying the gradient of the predicted probabilities with respect to the regression coefficients by the covariance matrix of the regression coefficients, which is available in the cov_params attribute of the LogitResult object. The gradient is computed as a matrix of partial derivatives of the predicted probabilities with respect to the regression coefficients, evaluated at each observation. The 95% confidence intervals are then computed by adding and subtracting the product of the standard errors and a critical value of 1.96, which corresponds to a 95% confidence level assuming a normal distribution.

Results and Findings

One-Way ANOVA

Race

Null hypothesis: There are no significant ratio differences between different racial groups who are in custody of the policy office.
Alternative hypothesis: There are significant ratio differences between racial groups who are in custody of the policy office.

At 95% confidence interval, f statistics is 26.76 and the p value is much less than 0.05. Thus, the results are statistically significant. We can reject the null hypothesis and accept the alternative hypothesis that there is significant difference between racial groups in the probability of strip searches.

In addition, we ran post hoc tests to confirm where the differences occurred between groups. Table 3 presents 9 pairs of groups with significant differences in the ratio of strip search,

especially between Asian and Black/White, Black and Latino/Middle Eastern, White and Latino/Middle Eastern, and Indigenous and Middle Eastern.

Table 5. Results by Race of One-Way ANOVA

Variables		**sum_sq**	**df**	**F**	**p-value**
Race		3.824	5.0	26.759	4.3E-27 ***
Asian	Black				0.001 ***
Asian	Middle-Eastern				0.0128 *
Asian	White				0.001 ***
Black	Latino				0.001 ***
Black	Middle-Eastern				0.001 ***
Indigenous	Latino				0.0018 *
Indigenous	Middle-Eastern				0.001 ***
Latino	White				0.001 ***
Middle-Eastern	White				0.001 ***

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Age Group

Null hypothesis: There is no significant ratio difference between age groups booked in the policy office.
Alternative hypothesis: There are significant ratio differences between age groups who are booked in the policy office.

At 95% confidence interval, f statistics is 31.61 and the p value is much less than 0.05. Therefore, we reject the null hypothesis and significant differences in the likelihood of being strip searched do exist between age groups. By running the Post Hoc test, we can see that pairwise differences exist in 8 combinations. There are 5 groups with the most significant difference, namely, adult and young adult/youth, and young adults and elder/middle

aged/youth (Table 5).

Table 6. Results by Age group of One-Way ANOVA

Variables		sum_sq	df	F	p-value
Age group		3.628	4.0	31.613	2.5e-26 ***
17	20				0.0091
20	30				0.00
20	40				0.00
20	50				0.00
20	65				0.00
20	65				0.00 ***
30	65				0.004 *
40	65				0.0064 ***

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Two-Way ANOVA

Table 7. Results of Two-Way ANOVA

Variables	sum_sq	df	F	p-value
Age group	4.08	6.0	12.40	5.15e-14***
Race	11.19	5.0	40.77	5.72e-42***
Race: Age group	3.08	30.0	1.87	2.586e-03***

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Strip Search ratio & age_group

Null hypothesis: There is no difference in the ratio of Strip Search for arrested persons who

are in different age groups.

Alternative hypothesis: There is a difference in the ratio of Strip Search for arrested persons who are in different age groups.

At a 95% confidence interval, the f statistic is 29.63 and p value (1.23e-24) is less than 0.05. Therefore, we can reject the null hypothesis and support the alternative hypothesis. There is a difference in the probability of strip search for people arrested and in custody grouped by age.

Strip Search Ratio & Race

Null Hypothesis: There are no differences in strip search ratio for arrest and custody cases grouped by races.

Alternative hypothesis: There is a difference in strip search ratio for arrest and custody cases by races.

At 95% confidence interval, f statistics is 28.59 and p-value is much less than 0.05. Then, we reject the null hypothesis and conclude that the differences in the likelihood of strip search exist for cases grouped by race.

Interaction between Age Group & Race

Null hypothesis: There is no interaction effect between age group and race in determining the strip search decisions. That is the effect of either age group or race is independent of the levels of each other.

Alternative hypothesis: There is an interaction effect between age group and race for police in making the strip search decision.

At a 95% confidence interval, f statistics is 3.573 and the p-value is much less than the alpha level of 0.05, thus, we reject the null hypothesis and accept the alternative hypothesis. There is an interaction effect between age group and race in making the strip search decision.

According to Table 7, there are 80 groups with significant differences.

Table 8. Post HOC Comparisons of Two-Way ANOVA

Group 1	Group 2	meandiff	p-adj
17Black	30Asian	-0.0517	0.0045

20Asian	30Asian	-0.0688	0
20Asian	30Latino	-0.0636	0.0185
20Asian	30Middle-Eastern	-0.071	0.0001
20Asian	40Asian	-0.0466	0.0318
20Asian	40Latino	-0.071	0.0202
20Asian	40Middle-Eastern	-0.06	0.0273
20Asian	50Asian	-0.0544	0.0313
20Asian	50Middle-Eastern	-0.07	0.0376
20Asian	65White	-0.0667	0.0397
20Black	20Latino	-0.0812	0.0122
20Black	20Middle-Eastern	-0.06	0.0025
20Black	30Asian	-0.0859	0
20Black	30Black	-0.0331	0.0001
20Black	30Latino	-0.0806	0
20Black	30Middle-Eastern	-0.09	0
20Black	40Asian	-0.0637	0
20Black	40Black	-0.037	0.0002
20Black	40Latino	-0.0881	0
20Black	40Middle-Eastern	-0.08	0
20Black	40White	-0.0264	0.0119
20Black	50Asian	-0.0715	0
20Black	50Black	-0.0418	0.0053
20Black	50Latino	-0.1074	0.0039
20Black	50Middle-Eastern	-0.09	0.0001
20Black	50White	-0.0391	0
20Black	60Asian	-0.0781	0.0001

20Black	60White	-0.0375	0.0058
20Black	65Asian	-0.1115	0.0124
20Black	65White	-0.0838	0
20White	30Asian	-0.067	0
20White	30Latino	-0.0617	0.0067
20White	30Middle-Eastern	-0.07	0
20White	40Asian	-0.0448	0.0037
20White	40Latino	-0.0692	0.01
20White	40Middle-Eastern	-0.06	0.0102
20White	50Asian	-0.0526	0.0083
20White	50Middle-Eastern	-0.07	0.0224
20White	65White	-0.0649	0.0208
30Asian	30Black	0.0529	0
30Asian	30Indigenous	0.08	0
30Asian	30White	0.0654	0
30Asian	40Black	0.0489	0
30Asian	40Indigenous	0.06	0.0188
30Asian	40White	0.0595	0
30Asian	50Black	0.0441	0.0136
30Asian	50Indigenous	0.08	0.023
30Asian	50White	0.0469	0
30Asian	60Black	0.0667	0.0055
30Asian	60White	0.0484	0.0003
30Black	30Middle-Eastern	-0.055	0.0004
30Indigenous	30Latino	-0.0726	0.0118
30Indigenous	30Middle-Eastern	-0.08	0.0002

30Indigenous	40Asian	-0.0556	0.03
30Indigenous	40Latino	-0.08	0.0115
30Indigenous	40Middle-Eastern	-0.07	0.0173
30Indigenous	50Asian	-0.0634	0.02
30Indigenous	50Middle-Eastern	-0.08	0.0202
30Indigenous	65White	-0.0757	0.02
30Latino	30White	0.0601	0.0022
30Latino	40White	0.0543	0.0165
30Middle-Eastern	30White	0.0676	0.00
30Middle-Eastern	40Black	0.0511	0.0063
30Middle-Eastern	40White	0.0617	0.00
30Middle-Eastern	50Indigenous	0.0834	0.0457
30Middle-Eastern	50White	0.049	0.0096
30Middle-Eastern	60Black	0.0689	0.0257
30Middle-Eastern	60White	0.0506	0.0238
30White	40Asian	-0.0432	0.0001
30White	40Middle-Eastern	-0.06	0.0035
30White	40Latino	-0.0676	0.005
30White	50Asian	-0.051	0.0016
30White	50Middle-Eastern	-0.07	0.0138
30White	60Asian	-0.0576	0.0313
30White	65White	-0.0633	0.011
40Asian	40White	0.0373	0.0046
40Latino	40White	0.0617	0.0255
40Middle-Eastern	40White	0.052	0.0252
40White	50Asian	-0.0451	0.0181

Interaction plot

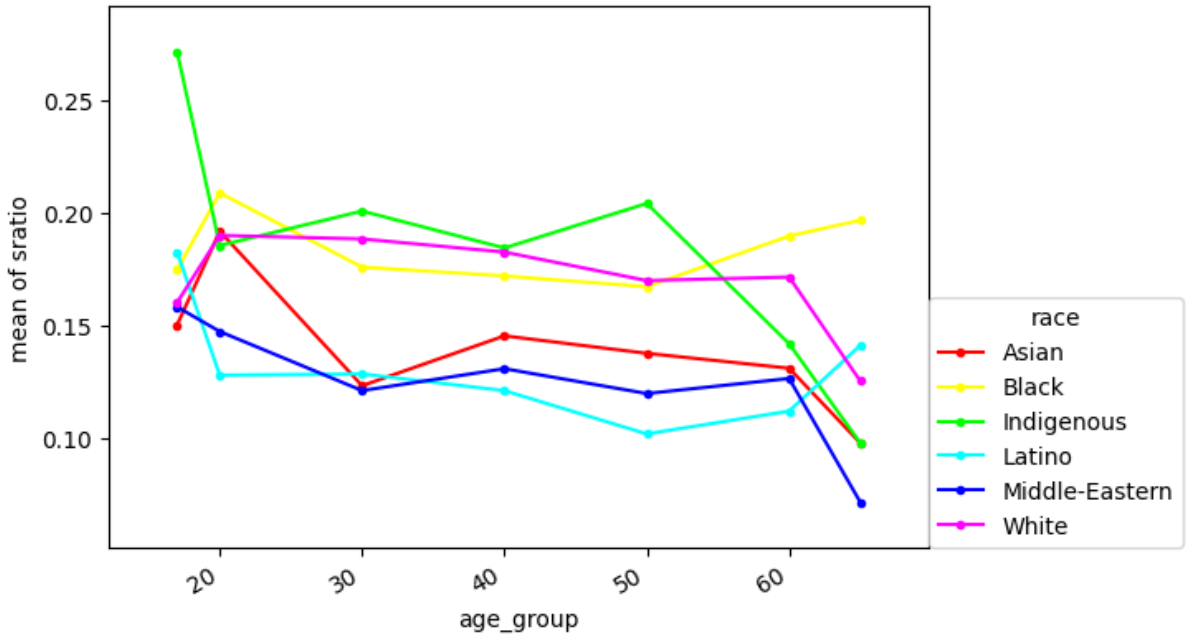

Figure 13 *Interaction plot*

While the interaction plot does not provide any information on the statistically significant difference, the plot showed: (a)generally black people face a higher ratio of strip search especially for older people and young adults, but youth can expect a lower ratio; (b) there is a big fluctuation in the indigenous group; for instance, the elder indigenous would have a relatively low ratio of strip search, but the indigenous adults would have a relatively high strip search ratio; and (c) most of the racial groups follow the similar pattern that they got relatively high ratio if they are adults; on the other hand, they would have a relatively low ratio when they get older or they are youth.

Power Analysis

Statistical analysis only relying on p-value is not comprehensive, because researchers may make an error, such as type l error and type ll error. Power analysis can help to judge the sample size required for the experiment, given an expected significant level, effect size and statistical power. To test the validity of the study we did based on our available sample size, we conducted power analysis to confirm the required sample size, and calculate whether the probability of correctly detecting an effect has reached the standard.

Table 9. Results of Power Analysis for ANCOVA

Setup: sig.level = 0.05; effect size = 0.0374; power = 0.8

	Sample Size	Actual Size
White	9730	27723
Other Race	13180.000	37553
Power	0.800	1.0

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

According to the result of power analysis, to reach a power of 0.8, the minimum sample size needed for white people is 9730, and the required number of samples in other races is 13179 to have a significant p-value in the t-test. Obviously, the actual size (27723 and 37553 respectively) far exceeds the required sample size.

Table 10. Results of Power Analysis for Logistic Regression

Setup: sig.level = 0.05; effect size = 0.3139; power = 0.8

	Sample Size	Actual Size
High sratio	121	11486
Low sratio	241	22997

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

To reach a power of 0.8, the minimum sample size needed for high sratio is 121, and the required sample size for low sratio is 241 to have a significant level. Obviously, the actual size (11486 and 22997 respectively) far exceeds the required sample size.

ANCOVA

Table 11. Result of ANCOVA

Source	SS	DF	F	p-unc	np2
race	7.497478	6	39.718278	1.59E-48	0.003638
bratio	432.051047	1	13732.8761	0.00E+00	0.173832
Residual	2053.401459	65268			

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

Null hypothesis: People from all race groups do not have different strip search turnover ratio by controlling the booked turnover ratio

Alternative hypothesis: People from all race groups have different strip search turnover ratio by controlling the booked turnover ratio

Statistical interpretation Interpretation p-unc = "uncorrected p-value" for race is less than 0.05. We can reject the null hypothesis that each of all race groups have the same strip search turnover ratio by controlling the booked turnover ratio, even after controlling for the booked turnover ratio.

Practical interpretation We hypothesized that the ethnicity background would be able to predict a strip search turnover ratio. From our results, we see that there is a statistically significant relationship between race and strip search turnover ratio when controlling for their booked turnover ratio. This raises interesting insights. If we consider race as a proxy for socioeconomic status, the race present in the dataset may be able to account for the relationship between the possibility of being strip searched. The ethnicity background may be influencing the possibility of strip search right now but it is still possible it influences their further arrest (further down the road).

Logistic Regression

Model results

Since our dependent variable in this study is the probability of finding items. Logistic regression is a machine learning method used to solve binary classification problems to estimate the likelihood of one event. The output below is the summary of the logistic regression model, which includes 5 predictor variables, namely, sratio, searchreason_causeInjury, searchreason_assistescape, searchreason_possessweapons, and searchreason_possessevidence.

Table 12. Results of Logit Regression

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-1.0389	0.075	-13.788	0 ***	-1.187	-0.891
sratio	1.119	0.1	11.155	0 ***	0.922	1.316

SearchReason_CauseInjury	-0.035	0.068	-0.517	0.605	-0.168	0.098
SearchReason_AssistEscape	0.3471	0.062	5.589	0 ***	0.225	0.469
SearchReason_PossessWeapons	-0.0948	0.057	-1.663	0.096	-0.206	0.017
SearchReason_PossessEvidence	-0.0892	0.056	-1.587	0.113	-0.199	0.021

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 13. Odd ratio

	Lower CI	Upper CI	OR
Intercept	-1.0389	0.075	-13.788 ***
sratio	1.119	0.1	11.155 ***
SearchReason_CauseInjury	-0.035	0.068	-0.517
SearchReason_AssistEscape	0.3471	0.062	5.589 ***
SearchReason_PossessWeapons	-0.0948	0.057	-1.663
SearchReason_PossessEvidence	-0.0892	0.056	-1.587

Source: The *Arrests and Strip Searches* dataset from Toronto Police Service

*** p < 0.001; ** p < 0.01; * p < 0.05.

We perform a logistic regression to examine the effects of sratio, SearchReason_CauseInjury, SearchReason_AssistEscape, SearchReason_PossessWeapons, and SearchReason_PossessEvidence, on the likelihood that item would be found when conducting a strip-search. Based on the results interpreted from table x., the sratio and SearchReason_AssistEscape are statistically significant. To better interpret the results of logistic regression, we convert the log odds to an odd ratio.

The odd ratio of sratio is 11.155 and the p-value is less than 0.05, so sratio is significantly associated with the probability of finding items. As sratio increases by one unit, the likelihood of finding an item during a search will increase by 11.155 times, holding other variables constant. In addition, for the variable searchreason, the odd ratio of searchreason_assistescape is 5.589 and the p value is much less than 0.05, which is

statistically significant. When the search reason is to assist escape, the probability of finding items will increase by 5.589 times, holding other variables constant.

The p-values for SearchReason_CauseInjury, SearchReason_PossessWeapons, and SearchReason_PossessEvidence are greater than 0.05, indicating that they are not significant predictors. Then, sratio and search reason of assist escape are significant predictors.

Assess the model

According to our test, the model has a test accuracy of approximately 0.63, which means it correctly predicted the class for about 63% of the instances in the test set. The confusion matrix shows the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) predictions made by the model.

Table 14. Confusion Matrix

Confusion Matrix	Negative	Positive
Negative	931	48
Positive	522	60

The given confusion matrix shows the results of a binary classification model. The model predicted for 1039 samples.

- 931 samples were classified as negative (0) and were actually negative (true negative).
- 48 samples were classified as positive (1) but were actually negative (false positive).
- 522 samples were classified as negative (0) but were actually positive (false negative).
- 60 samples were classified as positive (1) and were actually positive (true positive).

These metrics can help evaluate the performance of the model and identify areas for improvement. For instance, in this case, the low recall suggests that the model is not very effective at identifying positive samples. This could be due to the model's threshold for classifying samples as positive being too high, leading to a large number of false negatives. To improve the recall, the threshold could be lowered or other model parameters could be adjusted.

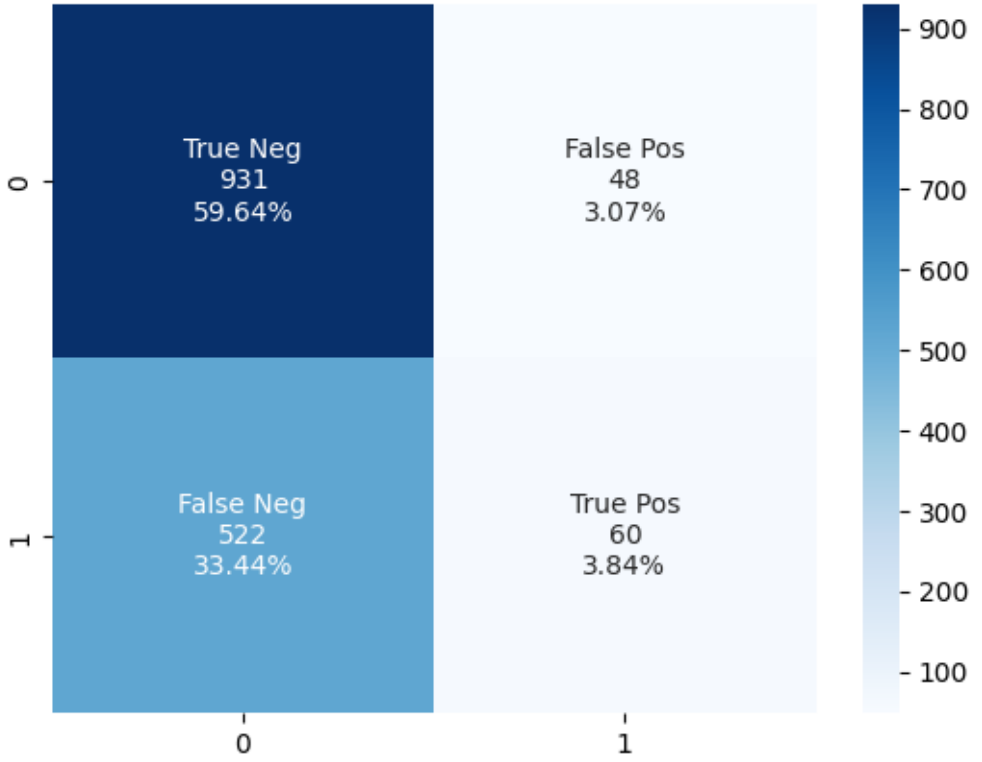

Figure 14 *Confusion Matrix Plot*

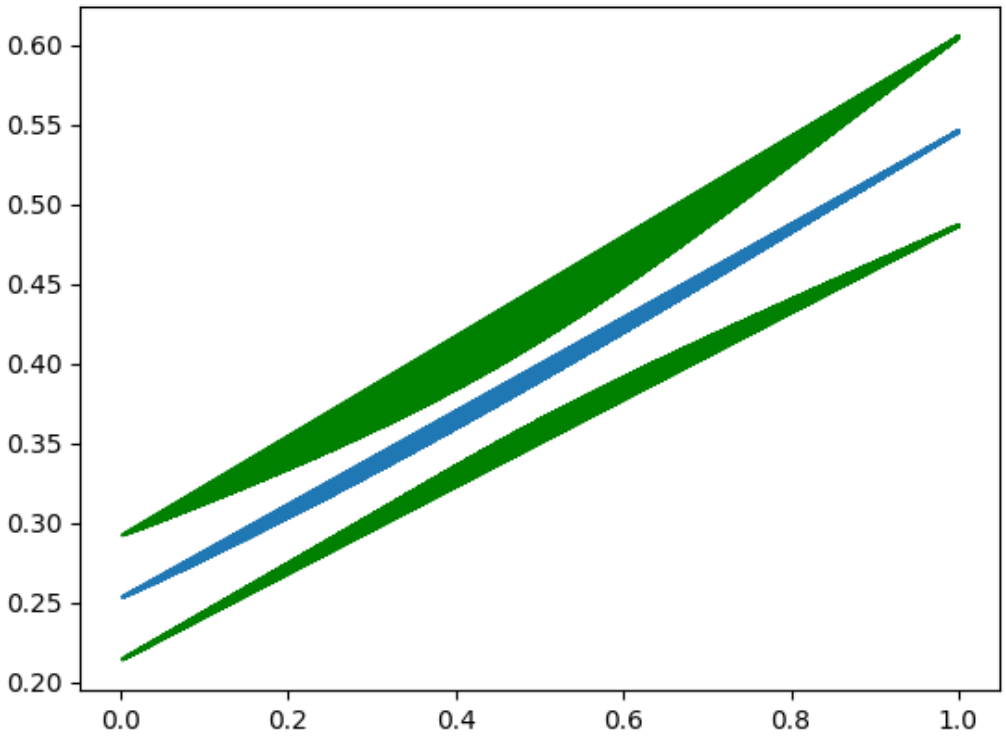

Figure 15 *Prediction interval and confidential interval*

Binary response variable that indicates whether an item was found (1) or not (0), the logistic regression model estimates the probability of finding an item as a function of the sratio variable. The predicted interval of finding an item for each value of sratio are shown in blue on the plot. As the graph shows the model can predict the possibility of item_found between around 0.25 and 0.5. The confidence intervals for the predicted probabilities are also shown in green. The confidence intervals indicate the range of values within which the true probability of finding an item is likely to fall with a certain level of confidence (in this case, 95%).

Interpretation of the plot depends on the specific problem being addressed and the context of the sratio variable. However, in general, the plot can be used to identify any relationship between the sratio variable and the probability of finding an item. If the plot shows a clear pattern, such as an increasing or decreasing trend, it suggests that the sratio variable is a significant predictor of the probability of finding an item. On the other hand, if the plot shows no clear pattern or random fluctuations around a constant value, it suggests that the sratio variable is not a strong predictor of the probability of finding an item.

Discussion

Preliminary suppositions

Through analysis, we found that differences in race and age do affect the probability that the police will conduct strip searches on arrested persons. For example, whites and blacks are more likely to be searched than other racial populations. In addition, adults are more likely to be searched than youth and elderly people. These biases are partly due to historical factors and the distribution of Toronto's population, such as a higher proportion of whites and blacks than other races.

Continuous studies

By using the ratio of being booked as a covariate in ANCOVA, we obtained more reliable and accurate results by taking into account the influence of the booked ratio, which is the proportion of people who are booked or arrested for a crime. By using ANCOVA and setting the booked ratio as a covariate, we got a clearer picture of the relationship between race and strip search, controlling potential confounding variables, accounting for the effects of a continuous variable, and improving the validity of the analysis.

In addition, applying logistic regression to our study allows us to identify the predictors which are most strongly associated with the outcome, such as the reasons for conducting a strip search or the likelihood of finding contraband. Also, it helps us identify which predictors are most influential in determining the outcome of interest and can be used to make predictions about future outcomes. Furthermore, logistic regression can be used to control for the influence of confounding variables, such as demographic or contextual factors, that may affect the relationship between the predictors and the outcome. By including these variables in the analysis, we can obtain a more accurate estimate of the effect of the predictors on the outcome. More specifically, given the results of logistic regression, we found that the odds of finding illegal items is associated with the search reason which is to assist escape and the practice of strip searching. That helps to inform policies and procedures related to strip searches. For example, the police can search suspects based on specific search reasons such as assisting escape and reduce the number of searches in other categories.

We must admit that the judiciary act of strip search has caused dissatisfaction among the public. Strip searches are traumatic. Strip searches represent a significant invasion of privacy and are often a humiliating, degrading and traumatic experience. Racialized people and women can experience being strip searched as akin to a sexual assault" (Lemke, 2022). In addition to females, strip searches have a severe adverse impact on youth and children. "Further research has shown that strip searches, performed even as intended, can cause children to experience anxiety, depression, loss of concentration, sleep disturbances, difficulty performing in school, phobic reactions, shame, guilt, and other lasting emotional scars" (Shah & Fireman, 2021). At the same time, strip searches are inefficient to some extent. "In May 2014, the Toronto police chief at the time reported to the Police Services Board that only in two percent of strip and cavity searches did police find any items and only a fraction of those found objects posed a risk" (Lemke, 2022). The purpose of conducting strip searches is to maintain social order, but the harm caused by this practice far outweighed the benefits it brings to society.

These findings provide us with new insights into strip searches and how age group, ethnicity and the reasons for search related to the outcomes. As a result, the Toronto Police Service has a better understanding of search practices and is making changes to reduce disparities during controlled situations such as arrests and searches. At the same time, it is suggested to conduct efficient strip searching on specific searching categories such as the category of assisting

escape in order to reduce meaningless judicial actions. Based on these findings, it has identified areas for improvement, including increased accountability, better supervision and coaching of officers, enhanced training, and the development of performance metrics to monitor progress transparently. The authority is committed to ongoing engagement with communities and members to ensure the safety of citizens while not harming human rights.

Limitations and Conclusion

There are still limitations to our research. Since the population ratio of male is far more than the ratio of females, and the number of whites is more than the number of other races, this dataset has a certain bias and will have a certain impact on the statistical results. The analysis considers incidents of arrests, bookings, and strip searches, which could happen to the same individual multiple times. The goal is to determine whether repeated arrests increase the likelihood of experiencing strip searches, especially for offences such as drug or weapon-related crimes. It is important to note that racial disparities' findings are insufficient evidence of systemic racial bias. Further investigation, possibly through additional data sources and methods, is necessary to identify the underlying causes of these differences and take effective measures for sustainable change. Furthermore, a majority of White and adult respondents compared to other races and age groups may indicate a lack of representation in our study. Also, there are some motivations beyond our knowledge and data availability to understand why police will conduct strip searches.

Our research found that differences in demographic factors such as age and race will make the probability of being strip searched different. At the same time, the interaction between them also affects the judiciary's decisions. However, it is unclear whether the judgement based on such factors effectively improves the success rate of the strip search. Therefore, we call on the judiciary to treat people of every race and age equally. At the same time, the judiciary should further improve the regulations and rules for judicial execution. We know that strip-searching is not only frustrating for those being frisked, it is also a waste of our judicial department's resources.

While ANCOVA with the booked ratio as a covariate can help to control for the influence of a potentially confounding variable, it is not without limitations. One limitation is that it assumes linearity between the covariate and the dependent variable, which may not always be

the case in real-world data. Another limitation is that ANCOVA may not completely eliminate the influence of confounding variables, as there may be other unmeasured variables that are related to both the covariate and the dependent variable. Additionally, ANCOVA assumes homogeneity of regression slopes, which means that the relationship between the covariate and the dependent variable is the same across all levels of the independent variable. If this assumption is violated, ANCOVA may produce biased results. Finally, ANCOVA requires a large sample size to achieve sufficient power, and violations of the assumptions can lead to reduced power and less reliable results.

Furthermore, there are several limitations to logistic regression when applied to the analysis of strip search reasons and items found. One limitation is that logistic regression assumes linearity between the predictors and the outcome, which may not always hold in real-world data. For example, there may be non-linear relationships between the strip search reason and the likelihood of finding contraband. Another limitation is that logistic regression assumes independence of observations, which may not be the case in certain situations. For example, strip search data from the same officer or facility may be correlated, violating the assumption of independence and potentially leading to biased results. A further limitation is that logistic regression is limited to analyzing binary outcome variables, which may not capture the full complexity of strip search data. For example, there may be multiple types of contraband found or multiple reasons for conducting a strip search, which cannot be fully captured by a binary outcome variable. Finally, logistic regression assumes that the predictors are measured without error, which may not always be the case in practice. For example, strip search reasons or contraband found may be subject to measurement error or misclassification, leading to biased estimates.

References

Gillis, W. (2020, November 24). *'clearly, we were doing it wrong': Toronto police are doing far fewer strip searches under strict new rules, interim chief says*. thestar.com. Retrieved April 16, 2023, from https://www.thestar.com/news/gta/2020/11/23/clearly-we-were-doing-it-wrong-toronto-police-doing-far-fewer-strip-searches-under-new-rules-interim-chief-says.html

Lemke, M. (2022a, July 6). *Policing toronto: Strip searching in a divided city - the bullet*. Socialist Project. Retrieved February 28, 2023, from https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/

Lemke, M. (2022b, June 16). *Strip searches are ineffective, unnecessary and target racialized Canadians*. The Conversation. Retrieved February 28, 2023, from https://theconversation.com/strip-searches-are-ineffective-unnecessary-and-target-racialized-canadians-185187

Shah, R. S., & Feierman, J. (2021, October 12). *Strip-Searching Children Is State-Imposed Trauma*. Americanbar.org. Retrieved February 28, 2023, from https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/empowering-youth-at-risk/strip-searching-children-is-state-imposed-trauma/

Smith, P. (2019, April 13). *News*. Toronto police continue practice of illegal strip searches: report. Retrieved April 16, 2023, from https://dailyhive.com/toronto/toronto-police-continue-illegal-strip-searches-2019

Toronto Police Service. (n.d.). *Strip Searches: Measurement & Outcomes*. Online Reporting-Toronto Police Service. Retrieved February 28, 2023, from https://www.tps.ca/services/online-reporting/

Toronto Police Service. (2022, November 10). *Arrests and strip searches (RBDC-arr-TBL-001)*. Toronto Police Service Public Safety Data Portal. Retrieved February 28, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about