INF 2178 Final Project

A Study of Examining the Toronto Police Service's Strip Search Policy

By: Group 32

Si Cheng (1003834158),
Ruolan Zhang (1004711010)

Instructor: Prof. Shion Guha

Table of Contents

1. Introduction

1.1 Background Information

According to the Supreme Court of Canada, the strip search is described as "the removal or rearrangement of some or all of the clothing of a person so as to permit a visual inspection of a person's private areas, namely genitals, buttocks, breasts … or undergarments" (R. v. Golden). It has become a common method adopted by law enforcement during a police investigation. However, there has always been a lot of controversy around strip searches. Many people argue that it is against an individual's privacy right, and is a degrading procedure that may potentially cause psychic trauma. As a matter of fact, officers do not necessarily need to get consent from the searched person in order to conduct a strip search, nor do they need to get the assistance of a medical professional. These all reflect the unequal power between law enforcement and citizens, where police officers have the entire control over a person (Lemke). Apart from ethical concerns, it was reported that the police department in Ontario does not have a consistent regulation in terms of conducting a strip search. That is to say, there is no standard of what police officers can and cannot do during a strip search, and under what circumstance a strip search should be conducted (Admin). The report also showed the lack of training in the police department, as there was a serious deficiency regarding the strip search data collection and management, which makes this procedure even more untransparent (Admin). On top of that, from the existing data (Race and Identity-based Data Collection Strategy), implicit racial bias in the police service was also brought up to the table, as it was revealed that one out of every three individuals who underwent strip searches was Black people, and approximately one-third of Indigenous people arrested were subjected to strip searches.

Given the disputes surrounding strip searches, we were motivated to evaluate the current TPS strip search policy. Our assessment will focus on two areas of interest: the conduct of the strip search, and its resulting outcome.

1.2 Research Objective and Questions

The objective of our research is to examine the current policy of the Toronto Police Service (TPS) on strip searches conducted under various scenarios, by analyzing relationships between the decision of performing strip searches and various demographic and situational factors. This also includes the exploration of the impact of certain factors on the strip search frequency. In addition, through deeper investigation on relationships between the likelihood of finding items during a strip search and some demographic factors and type of offense committed, we aim to provide valuable insights into the use of strip search in the TPS and help identify areas for improvement in the corresponding policy and procedure.

Our analysis is based on a dataset of arrests and strip searches happening in the city of Toronto from 2020 to 2021, it will be structured around the following two research questions, which were developed based on the comprehension obtained from the background research, literature review, the statistical analysis with an informal, exploratory examination of the data, followed by the formal analysis including power test, ANCOVA test and logistic regression.

· **Research Question 1**: How are demographic factors (e.g., race, age, gender) and situational factors (e.g., month, whether or not booked, number of non-cooperative actions taken) related to the decision to conduct a strip search or not? In addition, how does the search frequency vary from different race groups controlling for potential covariate such as the number of non-cooperative actions taken during arrests.

· **Research Question 2**: Does the presence of items found (yes vs. no) during a strip search vary based on a person's race, age, gender, type of primary offenses they committed, and the month of the search?

2. Literature Review

Two research questions put different focuses on the dataset, "Arrests and Strip Searches", to dig into some meaningful criminal tendencies as well as potential biases associated with both demographic and situational factors in the criminology field. There are some published literature and reports analyzing similar perspectives as this study did during the past two decades. The article "Data, Denials, and Confusion: The Racial Profiling Debate in Toronto" written by Wortley and Tanner (2003) from the University of Toronto critically examines the ongoing debates about the phenomenon of "racial profiling", occurring in Toronto policies systems. This article specifically emphasized statistics comparison in the field of police stop and search practice between White and Black's people living in Toronto. It is reported that 50% of black high school respondents were stopped by Toronto police and conducted search practices more than once in the previous two years, which is a significantly higher statistic compared to 23% of Whites and 11% of Asians. Surprisingly, black people with higher social class and education are more likely to be stopped by the police compared to those with lower social class. According to the journal, although Toronto police denied the occurrence of racial disparity in policing, people of different races did not receive equal treatment from the perspective of police search practices, and some races are still overrepresented.

A British study conducted between May 1999 to September 2000 in north London collected data from people who were arrested and held in police custody in a single police station, including demographic characteristics, the reason for arrest, whether strip searches were conducted etc. It emphasized that the reason for the arrest is most likely the major factor for determining whether a strip search was conducted throughout every age group. Specifically, people associated with concealing evidence such as drugs and robbery own the highest two percentages of arrests that were strip-searched. This study also pointed out that the lack of trust and confidence towards minority ethnic communities, for example, African-Caribbean, triggered the local police to apply strip searches more frequently than other races. Different from our dataset, the outcome of arrest is another emphasized factor within their consideration. The same study shows arrests resulting in further charges have a 14% of strip search rate although the covariance is relatively lower.

In addition, a review of police strip searches in Ontario called “Breaking the Golden Rule” published in March 2019 mentions the percentage of items seized during the strip search practice. According to the table provided in the report, items were not found in 79% of strip search cases. Specifically, approximately 20% of cases of strip searches resulted in seizing drugs but nothing else. Despite the Supreme Court of Canada proposing strict guidelines on strip searches, the publisher also mentions the ethical issue induced by strip searches that some strip searches were conducted inappropriately or unnecessarily. There have been cases where strip searches were conducted as a form of punishment or humiliation, or where they were carried out without just cause or in a manner that was insensitive to the individual's privacy and dignity. Therefore, improving police accountability and system transparency are urgent affairs for local police in the province of Ontario.

3. Exploratory Data Analysis

3.1 Data Cleaning

To conduct analyses for our study, we created two sub-datasets that contain different dependent variables and explanatory variables regarding two different research questions. We start the data wrangling by merging “Age 17 years and under” and “Age 17 years and younger” together into one group, and merging “Aged 65 and older” and “Aged 65 years and older” together into one group under the Age column. Then we checked and dropped all rows with missing values since there were only 453 rows in total, consisting of 0.69% of the entire dataset. For our own area of interest, we also dropped the 9 rows where sex was displayed as ‘U’, and the 5056 rows where race was reported to be ‘Unknown/Legacy’. For the sub-dataset of our first research question, which was centered around relationships between strip searches and demographic and situational factors, we selected the five non-cooperative actions recorded during arrest and add them together into a numerical variable, then applied the ‘group by’ function to aggregate the dataset and calculate the mean number of non-cooperative actions taken regarding each group. On top of that, we merged all racial groups other than ‘White’ into a new group named ‘Color’. For the second sub-dataset, we reformatted values under the ‘Occurrence_Category’ column, checked the value counts of each level, and selected three main types of offense. In terms of the categorical variables, we used a label encoding technique to convert them into numbers and fed them into the logistic regression model for clearer interpretation.

3.2 Descriptive Statistics

Summary statistics are computed to show the sum-up features of each independent variable in the research question and provide a basic understanding of values recorded in the dataset, including mean, count, standard deviation and 95% confidence interval.

Diagram 3.1 presents a barplot illustrating the number of strip search cases for seven racial groups by gender. The x-axis represents race and the y-axis is the number of strip searches. White and Black groups have a significantly higher mean number of strip search cases compared to the rest racial groups. Specifically, female suspects are less frequently strip-searched than male suspects among all racial groups.

Diagram 3.1: *Barplot of Strip Search Incidence for Racial Groups by gender*

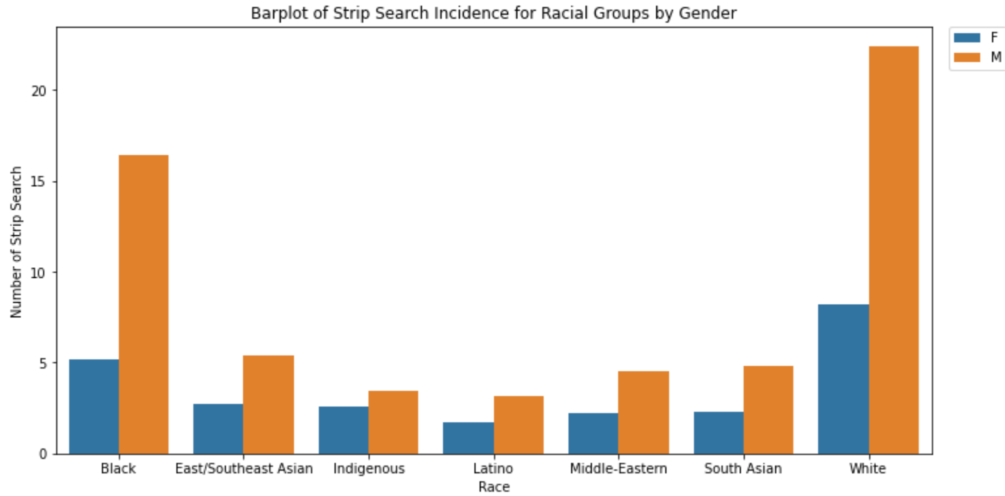
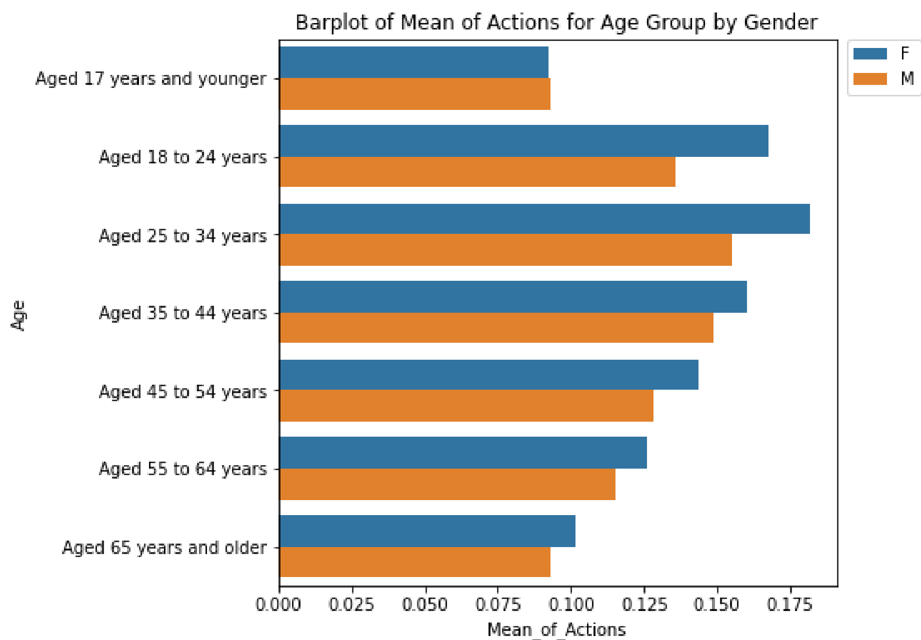


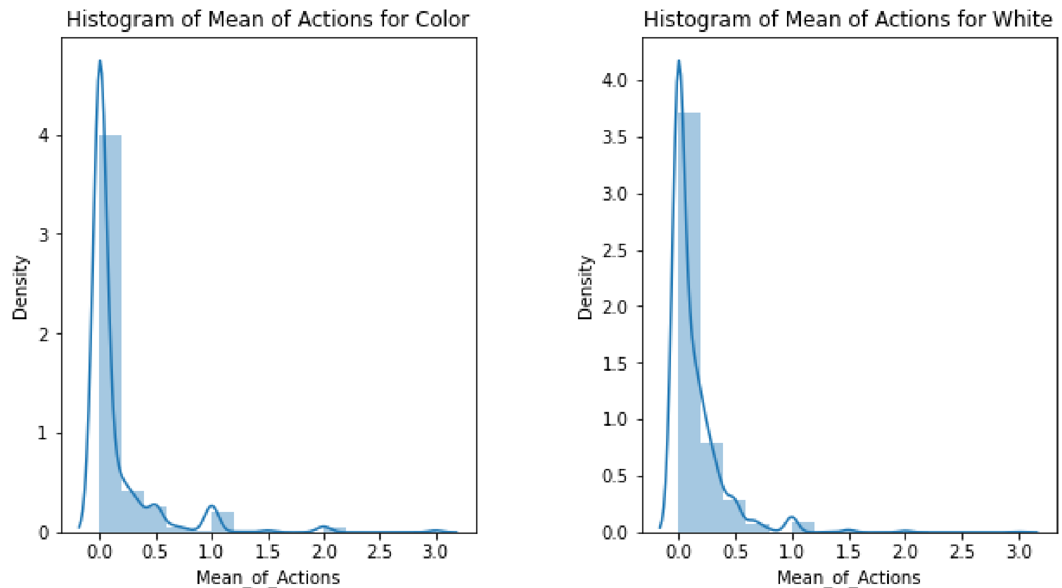
Diagram 3.2 is another barplot exploring the mean number of non-cooperative actions of suspects in nine different age groups during the arrest process. Another independent variable selected for this boxplot is Sex, and we are able to compare the difference in distributions of that of males and females. Suspects aged between 25 to 34 have the largest mean number of non-cooperative actions among all age levels. As shifting toward younger or older age levels, significant decreases in counts can be observed for both sexes. While females tend to actively resist arrest more frequently than males under every age group.

Diagram 3.2: *Barplot of Mean of Actions for Age Group by Gender.*



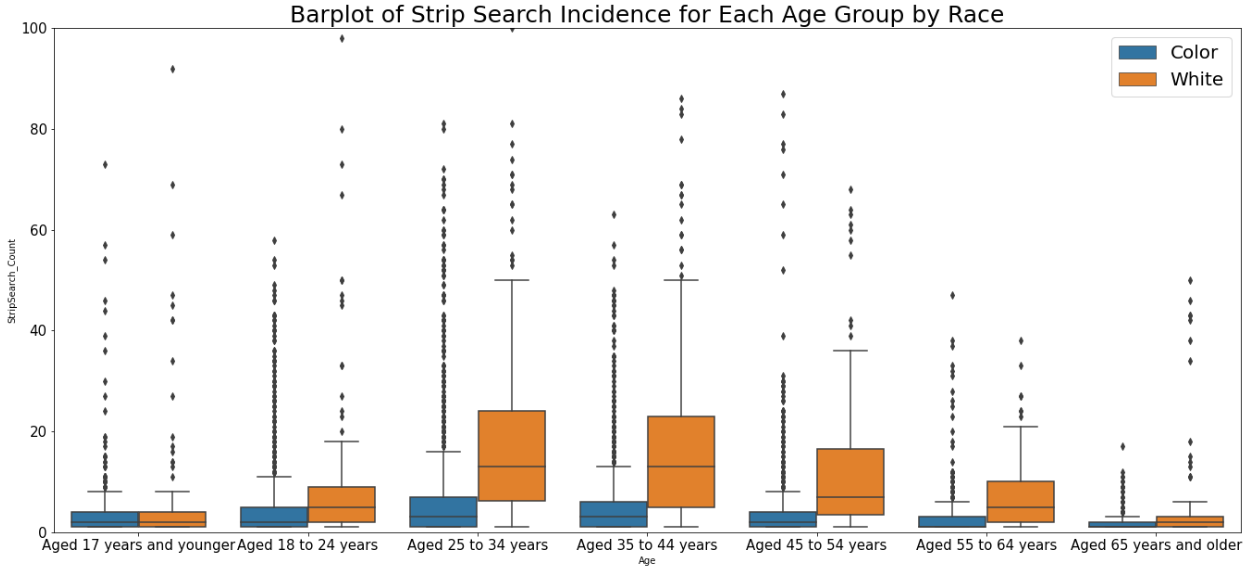
The following two histograms illustrate the distribution of the mean number of non-cooperative actions during the arrest process for two racial groups. There is no significant disparity between the two histograms in their distributions. One thing to be noticed is there are more people who belong to the Color group showing no non-cooperative actions compared to that White people. That is to say, the mean number of non-cooperative actions for the Color group is relatively smaller than that of the White group.

Diagram 3.3*: Histograms of Mean of Actions for Racial Group*



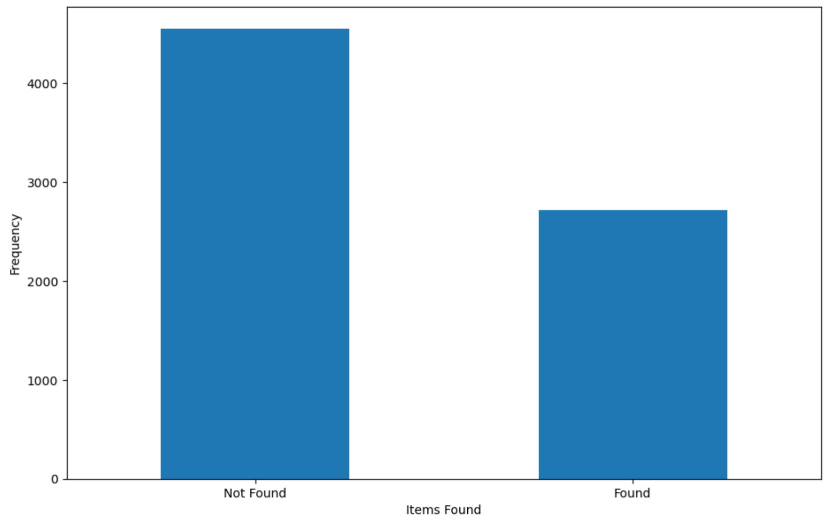
The following boxplot (Diagram 3.4) displays the number of strip search cases aggregated by age groups with nine levels. Another independent variable selected for this boxplot is a two-level racial group, so we are able to compare the difference in distributions of White and Color people through boxes and the position of whisker marks. Generally, we can see suspects aged between 25 to 34 and 35 to 44 are the age ranges that experienced the greatest number of strip searches. Specifically, White suspects located between 25 to 54 are the potential target cases of strip search compared to other age levels. As shifting toward younger or older age levels, significant decreases in counts can be observed for both racial groups.

Diagram 3.4*: Barplot of Strip Search Incidence for Each Age Group by Race*



The fifth diagram (Diagram 3.5) is a barplot showing the frequency of item presences during strip searches. Over the course of two years, 7265 strip searches were conducted, and items were discovered in 2716 cases, which accounts for 37.38% of the total. This is worth investigating as the percentage is considerable, and it is important to ensure that police officers are aware of the potential danger they may face while dealing with different individuals during arrests.

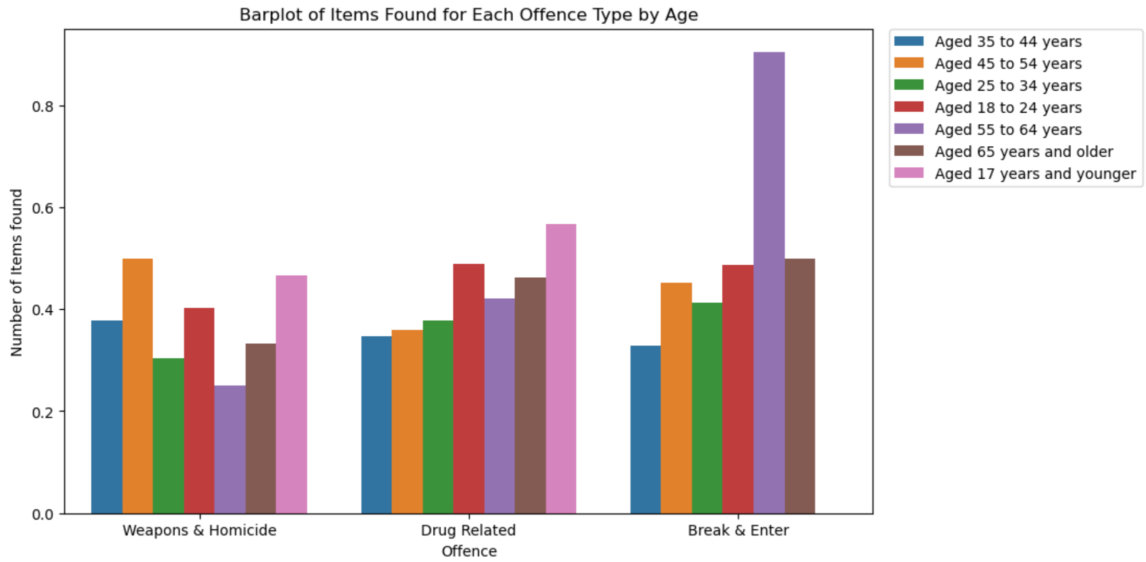
Diagram 3.5: *Barplot of the Frequency of Discovering item*



The sixth diagram (Diagram 3.6) is a barplot showing the number of items found for three main major offense types, grouped by seven age levels. Surprisingly, arrest cases associated with weapons and homicide found the least number of items on average during the strip search in general, but suspects who committed a break-and-enter crime found a relatively much higher number of items on average.

Specifically, break-and-enter suspects aged between 55 to 64 found significantly more items than any other levels. In general, older suspects are inclined to bring items when committing crimes.

Diagram 3.6*: Barplot of Number of Items Found for Each Major Offence Type by Age*



3.3 T-test: Strip Search Frequency vs. Race

3.3.1 Hypothesis

Hypothesis for two-sided two-sample t-test

- **H0 (Null Hypothesis)**:
  The mean number of strip search frequency does not vary from the White and Color groups.
- **Ha (Alternative hypothesis)**:
  The mean number of strip search frequency does vary from the White and Color groups.

3.3.2 Result

Race and the number of strip search cases
Calculated summary statistics display that the mean number of strip searches subjected to the White group (16.16) is greater than strip searches subjected to the Color group (6.06). To examine whether the difference is statistically significant, we apply Welch's t-test with non-equal variance. The resulting p-value for this test is 3.215e-24, which was much smaller than the chosen significance level of 0.05. Therefore, we have sufficient evidence to reject the null hypothesis that the true mean difference equals zero and conclude that there are significant differences between the White and Color groups regarding the strip search frequency.

4. Method

4.1 Dataset Description

The dataset used for this project is the Arrests and Strip Searches uploaded on Toronto Police Service (TPS) Public Safety Data Portal (Arrests and Strip Searches (RBDC-ARR-TBL-001)). The dataset contains details of all arrest and strip search cases that happened in 2020 and 2021, where the demographic data of people arrested, arrest divisions, actions at arrest, reasons for a strip search, and other relevant information were entered. The dataset has a size of 65277 and 25 columns, indicating that there were 65277 cases in total in these two years and information regarding 25 different attributes was recorded. Some of the important personal and procedural characteristics comprise the person's race, sex, age group, arrest location, whether being strip-searched, actions they took while being arrested, and the specific reason why they were subjected to strip searches. 12 out of 25 features are numerical, others are text. It is worth mentioning that most of the numerical features are in binary format, and all measurements are on a nominal scale.

4.2 Power Analysis

There are two main types of error that may happen while statistically testing a hypothesis. One is the Type I error, also known as the False Positive, and the other one is the Type II error, also known as the False Negative. The latter one in general is more important, as failing to reject the null hypothesis when there is a significant effect could result in serious consequences. Therefore, in order to make sure our statistical hypothesis testing performed later is valid, i.e., the Type II error is minimized, we need to perform the power analysis before the experiment to check how confident we can be regarding the ability of our model to correctly detect an effect if there exists any.

Theoretically, performing power analysis is crucial for ensuring that a study is adequately powered to determine meaningful differences between groups while minimizing the risk of Type I and Type II error.

Our power analysis will be conducted from two directions. One is to calculate the ideal sample sizes needed for both White and Color groups to detect a particular racial effect on the strip search frequency, given a significance level of 0.05, a power of 0.8 which is a commonly acceptable power in most experiments, and the effect size of the explanatory variable 'Race' using the Cohen's D metric. By determining the desired sample size for the following test, researchers are able to allocate resources efficiently and prevent wasting time and resources on unnecessary data.

The other direction is to calculate the statistical power based on the same significance level, effect size and our current sample sizes, to validate whether the probability of making Type II errors is acceptable in our study based on the current sample data

4.2.1 Power Test Result

The Cohen's D effect size of the explanatory variable 'Race' (two-level explanatory variable) was 0.424, calculated from the two independent White and Color racial samples. According to Cohen's criteria, this was a medium effect size, indicating that the size of the difference between White and Color groups is medium regarding the strip search frequency (outcome variable).

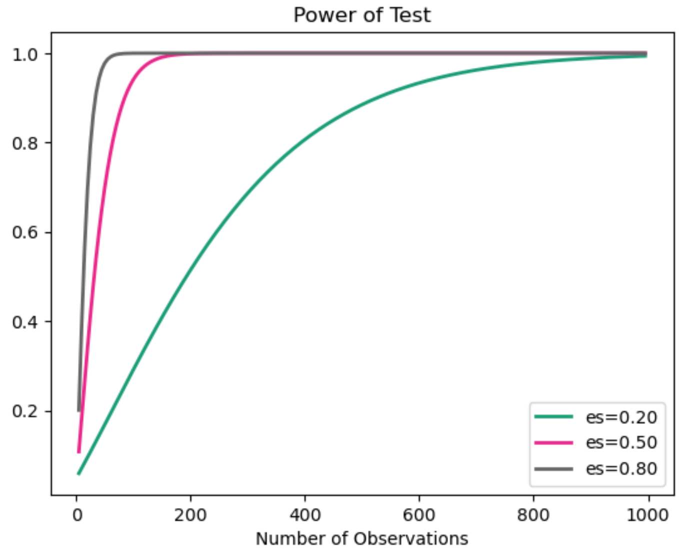
We then use this effect size, along with the 0.05 significance level (alpha), and a relatively large power of 0.8, to calculate the required minimum sample size, received 59 and 183 for White and Color groups, respectively. The power test result suggested that in order to observe an effect, a sample size of at least 59 is needed for the White group, while a sample size of at least 183 is needed for the Color group. The difference between the required sample sizes for the two racial groups is somewhat large. However, this is not a concern in our study, as our actual sample size of the White group is 1701, and our actual sample size of the Color group is 5323, both are far more than the minimum requirement. This also indicated that our sample is enough to assure a sufficient power (0.8), that is to say, our hypothesis testing result will be acceptable.

To further validate our finding, a second round of calculations was performed based on the actual sample sizes, the same significance level and effect size. We obtained a power of 1.00 for both White and Color groups, indicating that the probability of committing Type II (false negative) errors would be very small, which was desired and was consistent with our power test result.

Table 4.1: *Power Analysis Result*

	Effect Size	**Minimum Sample Size for Power of 0.8**	**Achieved Power**
Race - White	0.424	59	1.0
Race - Color		183	1.0

Diagram 4.1: *Power Curve*



The power plot (Diagram 4.1) shows the power increasing as the number of observations increases from small (0.20) to large (0.80) effect sizes, with steeper trends for higher effect sizes, holding a significance level of 0.05 (alpha).

As the effect size increases, the number of samples needed to receive a statistically significant result reduces. The power curves illustrated that in order to obtain very high power, the minimum required sample sizes for a medium effect size (0.50) sample and a large effect size (0.80) sample were similar, around 200. While the minimum required sample size for a small effect size sample (0.20) was greater than 1000. In our study, we have such large sample sizes for both White (1701) and Color (5323) racial groups that with a significance level of 0.05 and a power of 0.8, we have a great chance to detect a statistically significant result if the race effect exists at effect size (Cohen's D) of 0.4. Therefore, it can be expected that many of the further analyses in our study will be significant.

4.3 Analysis of Covariate (ANCOVA) Test

The ANCOVA test provides statistical support for part of the first research question examining the variation of strip search frequency across racial groups while controlling for the number of non-cooperative actions taken during arrests. Our group aims to shed light on the role of race in the frequency of strip searches and better understand factors that may contribute to the disparities in the application of strip searches during the arrest process. By investigating this question, the study can inform Toronto's local police and practices aiming at ensuring fair and equitable treatment during the arrest process.

Therefore, as an extension of one-way ANOVA, one-way ANCOVA is performed using the Pingouin package in Python to test whether there are significant differences between two independent groups when incorporating a covariate that is related to the dependent variable. We choose the number of strip searches as the dependent variable, while racial group and the mean number of non-corporate actions are categorical variables and continuous covariates respectively.

The ANCOVA hypothesis being tested are listed as following:

- **H0**: People who get arrested for all race groups have the same adjusted average of strip search frequency after counting their mean number of actions taken during arrests.
- **Ha**: People who get arrested for all race groups do not have the same adjusted average of strip search frequency after counting their mean number of actions taken during arrests.

4.4 Assumptions Check

Checking assumptions for Analysis of Covariate (ANCOVA) is essential to ensure the validity and reliability of statistical results. We are able to determine whether the data meet the criteria necessary for an accurate and meaningful interpretation of the results. In the ANCOVA model, this study is trying to explore how the strip search frequency varies from different race groups controlling for the covariate, and the number of non-cooperative actions taken during arrests. Generally, there are five main assumptions we have to check after fitting the model: linearity, normality, homogeneity of variances, independence of observations and homogeneity of regression slopes.

4.4.1 Independence of Observations

The dataset used for doing the ANCOVA test is a subset of “Arrests and Strip Searches” after transformation and aggregation. The independence assumption is automatically passed since all groups are mutually exclusive and there are no repeated measures, which means every observation is independent of each other. In other words, the strip search incidence of one suspect does not influence that of the other.

4.4.2 Normality

A statistical test called the Anderson-Darling Normality test is applied to check whether samples with different categories under the variable ‘race’ have normally distributed residuals. This test sets up the null hypothesis that data is not different from normal. The test statistics of the White group sample is approximately 340.45. We compare this value to each critical value that corresponds to each significance level to see if the test result is significant. For example, the critical value for alpha equals 0.1 is 1.089, and our test result is greater than this critical value. Therefore, this test statistic is significant at this level. After completing the comparison in every significance level, we can see the test statistics are significant for all levels, and we have sufficient evidence to reject the null hypothesis that sample data is normal. Thus, the normality assumption is violated in this case. The same procedure is performed on the sample of colored people, and the result is pretty much similar to that of the previous one which is not normally distributed. However, we will still carry out our test since a real-life dataset is impossible to be perfectly normal under any circumstance. The analysis result should not be rejected for the reason of non-normality, but have to mention this point as a limitation.

__Table 4.2__: Anderson-Darling Normality Test result for White group

Test Statistics: 340.4451602775962				
α = 0.01	**α = 0.025**	**α = 0.05**	**α = 0.1**	**α = 0.015**
1.089	0.916	0.785	0.654	0.575

__Table 4.3__: Anderson-Darling Normality Test result for Color group

Test Statistics: 1108.7218250593369				
α = 0.01	**α = 0.025**	**α = 0.05**	**α = 0.1**	**α = 0.015**
1.091	0.917	0.786	0.656	0.576

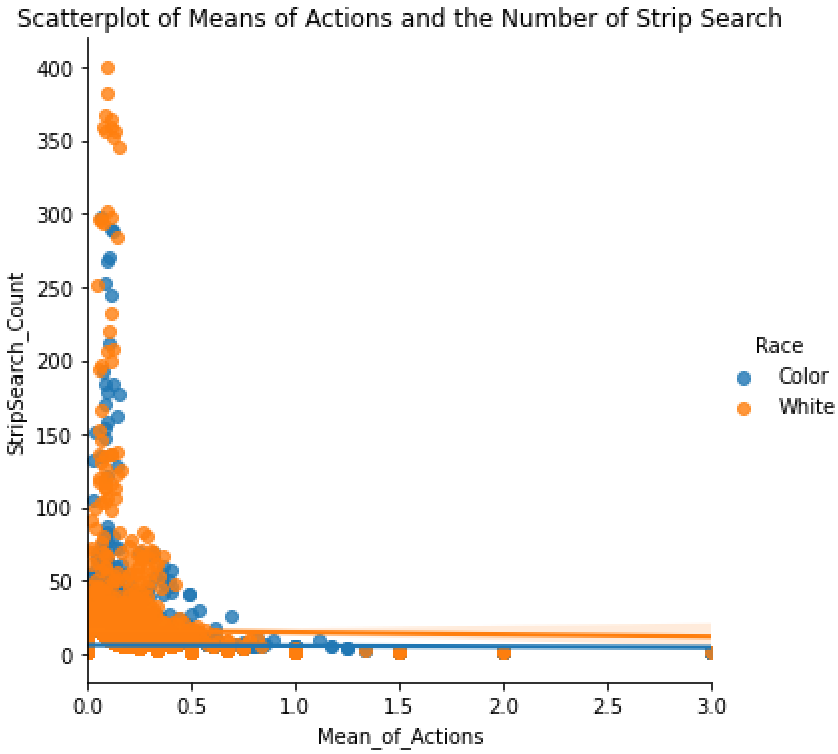
4.4.3 Homogeneity of Variances

As all data is drawn from non-normally distributed datasets as we check in the last section, Levene's Test is selected to check the homogeneity of variance for the variable 'race' with multiple levels. Applying Levene's Test, we propose the null hypothesis that each level of location division has an equal variance. The resulting p-value equals 0.7824443867697906 which is significantly larger than 0.05. As we don't have sufficient evidence to reject the null hypothesis, the homogeneity of variance assumption is passed for 'race'. Therefore, variances of residuals are the same across each group. This lowers the possibility of biased results and invalidity of the comparisons among groups.

4.4.4 Linearity

To check the third assumption of linearity, we have to ensure the relationship between the covariate "Mean_of_Actions" and the dependent variable "StripSearch_Count" is linear. This assumption can be checked by observing a scatterplot of the continuous variable(covariate) and dependent outcome, grouped by racial groups at the same time. The following scatterplot (Diagram 4.1) is the overview of all observations with regression lines plotted for White and Color groups. Observing the plot, although the Python package forces the regression line on the scatterplot to be linear, obviously the pattern of dots does not show any linearly-related trend. The same situation applies to both racial groups, in which 'Mean_of_Actions' and 'StripSearch_Count' are not linearly related. Failing in meeting the linearity assumption could result in a biased adjustment in the ANCOVA test.

Diagram 4.1: *Scatterplot of Means of Actions and the Number of Strip Search*



4.4.5 Homogeneity of Regression Slopes

Last but not the least, the final assumption of the ANCOVA test is that the relationship between the covariate and the dependent variable should be the same across all groups. After testing whether the relationship between the covariate and the dependent variable is linear, we have to check whether the regression line for each group is parallel to each other, which means they have the same slope. Since the dataset fails in meeting the linearity assumption, the assumption of homogeneity of regression slopes automatically fails.

4.5 Logistic regression

RQ1:

From the Diagram 3.1 bar plot in the EDA section, we observed that there was a significant difference in the mean number of strip search frequency regarding two gender groups, and also the White and Color racial groups. Also, Diagram 3.4 boxplot also showed that the strip search frequency varies for different age groups. In addition, the bar plot and the boxplot also showed that individuals of different races, ages, and gender had significantly different levels of uncooperating represented by their disobedient actions during the arrest, which suggested that demographic factors can affect the strip search.

We will build a logistic regression model to investigate the relationship between the strip search decision and a person’s demographics, arrest month, arrest division, whether being booked and the number of non-cooperative actions during the arrest.

RQ2:

From the plot (Diagram 3.6), we can see that items were found in 2716 out of 7265 total strip search cases, indicating that whether any items were found as a result of the strip search is also a question of concern. Therefore, for further exploration, we will build another logistic regression model to find out whether a person's demographics also play a role in the discovery of items, in addition to factors such as the type of offense committed and the month.
For both of the logistic regression models, we will employ the machine learning framework to split the sub-dataset into training and test datasets with suitable ratios. Then we will calculate the odds that represent the ratio of probabilities for positive and negative outcomes regarding each explanatory variable. Finally, we will use the trained model to make predictions based on the test data.

5. Result and Finding

RQ1:

How are demographic factors (e.g., race, age, gender) and situational factors (e.g., month, division, whether or not booked, number of non-cooperative actions taken) related to the decision to conduct a strip search or not? In addition, how does the search frequency vary from different race groups controlling for potential covariate such as the number of non-cooperative actions taken during arrests.

Logistic Regression Result

Coefficient, CI and OR

The result of the logistic regression centered around the first research question is shown in the Table 5.1 below. The values under the ‘P>|z|’ column suggest that all variables other than Month[T. July-Sept] (p = 0.082), Race[T.White] (p = 0.126), Age[T.Aged 18 to 24 years] (p = 0.124), Age[T.Aged 25 to 34 years] (p = 0.840), and Age[T.Aged 35 to 44 years] (p = 0.822) are statistically significant. The table also illustrates negative associations of the decision of conducting a strip search (yes or no) with variables Month[T.Oct-Dec] (coef = -0.166), Age[T.Aged 45 to 54 years] (coef = -0.239), Age[T.Aged 55 to 64 years] (coef = -0.456), and Age[T.Aged 65 years and older] (coef = -1.336).

The following result can be inferred from the odds ratio and its 95% confidence interval table generated using the logistic regression results:

1. The intercept indicates that the average odds of being searched when all of the explanatory variables are controlled is 0.017. I.e., the average odds of being searched for an under 18 female

racial minority who is booked and arrested in April to June, with zero non-cooperative actions during arrest, is 0.017.

2. Compared with April to June, the odds that people are subjected to strip searches are approximately 1.368 times higher in January to March, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of being strip-searched in January to March is 1.268 to 1.477 times more likely than in April to June. Note this effect is statistically significant.

3. Compared with April to June, the odds that people are subjected to strip searches are approximately 1.073 higher in July to September, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of the month, from July to September in particular, has an effect on the outcome (being strip searched) that is between 0.991 times less and 1.161 times more than in April to June. However, this effect is not statistically significant.

4. Compared with April to June, the odds that people are subjected to strip searches are approximately 0.191 times lower in October to December, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of being strip-searched in October to December is 0.169 to 0.216 times less likely than in April to June. Note this effect is statistically significant.

5. Compared with visible minorities, the odds that white people are subjected to strip searches are approximately 1.251 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of white people being strip-searched is 1.177 to 1.330 times more likely than the visible minorities. Note this effect is statistically significant.

6. Compared with females, the odds that males are subjected to strip searches are approximately 1.136 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of males being strip-searched is 1.049 to 1.212 times more likely than in females. Note this effect is statistically significant.

7. Compared with people under 18, the odds that people aged 18 to 24 are subjected to strip searches are approximately 1.140 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 18 to 24 in particular, has an effect on the outcome (being strip searched) that is between 0.965 times less and 1.348 times more than age under 18. However, this effect is not statistically significant.

8. Compared with people under 18, the odds that people of age 25 to 34 are subjected to strip searches are approximately 1.017 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 25 to 34 in particular, has an effect on the outcome (being strip searched) that is between 0.866 times less and 1.193 times more than age under 18. However, this effect is not statistically significant.

9. Compared with people under 18, the odds that people of age 35 to 44 are subjected to strip searches are approximately 1.019 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 35 to 44 in particular, has an effect on the outcome (being strip searched) that is between 0.866 times less and 1.199 times more than age under 18. However, this effect is not statistically significant.

10. Compared with people under 18, the odds that people of age 45 to 54 are subjected to strip searches are approximately 0.787 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of age 45 to 54 being strip-searched is 0.661 to 0.937 times less likely than age under 18. Note this effect is statistically significant.

11. Compared with people under 18, the odds that people aged 55 to 64 are subjected to strip searches are approximately 0.634 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of ages 55 to 64 being strip-searched is 0.518 to 0.776 times less likely than age under 18. Note this effect is statistically significant.

12. Compared with people under 18, the odds that people of age over 65 are subjected to strip searches are approximately 0.263 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of age over 65 being strip-searched is 0.173 to 0.399 times less likely than age under 18. Note this effect is statistically significant.

13. Compared with people who are not booked, the odds that booked people are subjected to strip searches are approximately 14.652 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of booked people being strip-searched is 13.212 to 16.249 times more likely than people not booked. Note this effect is statistically significant.

14. Increasing the number of non-cooperative actions is associated with an increased likelihood of a person being strip searched. The odds ratio is approximately 1.421. For each additional unit increase in the number of non-cooperative actions, the odds that the person is subjected to a strip search increases by about 1.421 times. Note this effect is statistically significant.

All metrics are highlighted more clearly in Table 5.2, including Coef, SE, Lower CI, Upper CI and OR.

Table 5.1*: Result of Logistic Regression for RQ1*

	Coef	SE	P-value	Lower CI	Upper CI	OR
Intercept	-4.1021	0.099	0.000	0.013619	0.020083	0.016538
Month[T.Jan-Mar]	0.3134	0.039	0.000	1.267531	1.476540	1.368050
Month[T.July-Sept]	0.0703	0.040	0.082	0.991225	1.161203	1.072853

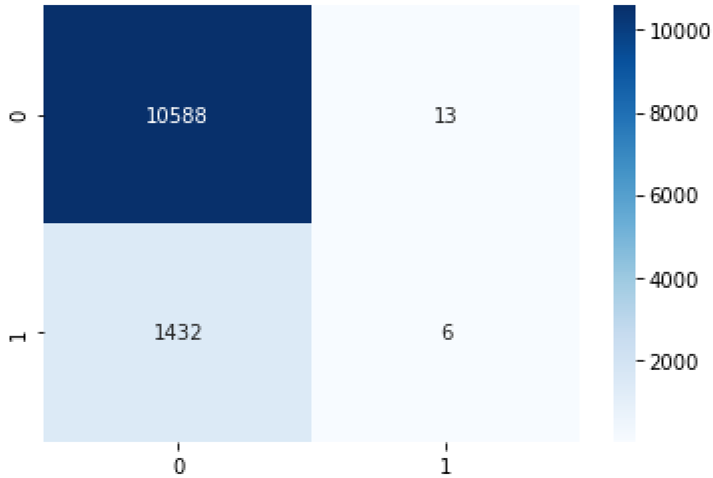
Month[T.Oct-Dec]	-1.6567	0.062	0.000	0.168842	0.215551	0.190772
Race[T.White]	0.2241	0.031	0.000	1.177292	1.329703	1.251179
Sex[T.M]	0.1278	0.041	0.002	1.048540	1.231502	1.136345
Age[T.Aged 18 to 24 years]	0.1314	0.085	0.124	0.964566	1.348445	1.140466
Age[T.Aged 25 to 34 years]	0.0164	0.082	0.840	0.866448	1.192687	1.016563
Age[T.Aged 35 to 44 years]	0.0186	0.083	0.822	0.865971	1.198566	1.018785
Age[T.Aged 45 to 54 years]	-0.2394	0.089	0.007	0.661434	0.936641	0.787100
Age[T.Aged 55 to 64 years]	-0.4555	0.103	0.000	0.518308	0.775760	0.634100
Age[T.Aged 65 years and older]	-1.3363	0.212	0.000	0.173358	0.398399	0.262804
Booked	2.6846	0.053	0.000	13.212092	16.248480	14.651840
Num_of_Actions	0.3513	0.027	0.000	1.348111	1.497515	1.420851

Model Prediction and Accuracy

The test accuracy of this logistic regression model is around 0.88 on average, indicating that the model correctly predicted whether or not a person will be strip searched 88% of the time.

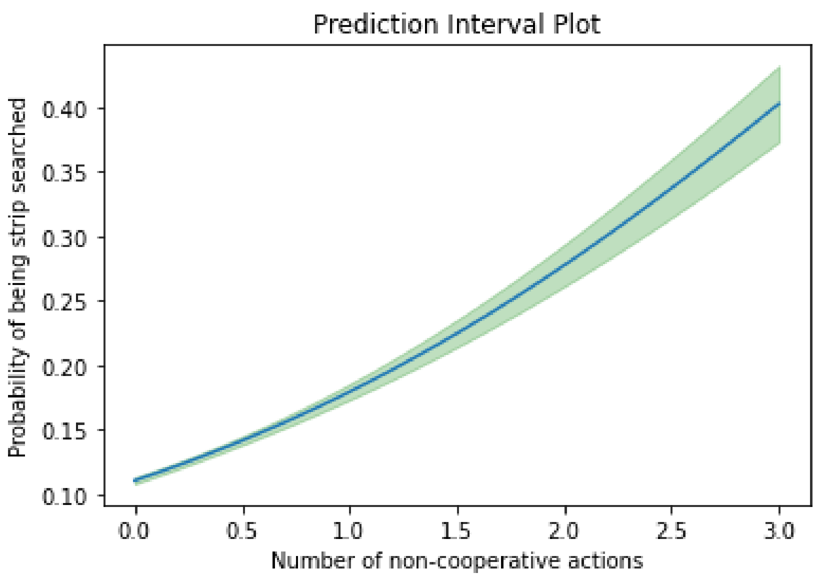
The confusion matrix (Figure x) below shows that among the 12039 test data points (20% of the sub-dataset), TN is 10588, TP is 6, FP is 13 and FN is 1432 (Type II error). This suggests that for 1438 people who actually need to be strip-searched, only 6 of them were asked to conduct one. Besides, for 10601 people that are not supposed to be searched, only 13 of them were subjected to a search incorrectly. That is to say, according to our logistic regression model, the probability of making a Type I (false positive) error is 0.12%, and the probability of making a Type II (false negative) error is 99.58%. This results in a significant issue, despite our model's relatively high accuracy (88%), there are still numerous people who have avoided strip searches when they were supposed to.

Diagram 5.1*: Confusion Matrix for RQ1*



Based on the prediction interval plot below, it appears that individuals who exhibit more violent behavior during arrests are more likely to be subjected to a strip search.

Diagram 5.2*: Prediction Interval Plot: Probability of strip search vs. Number of non-cooperative actions*



ANCOVA result

The result for one-way ANCOVA is shown in the Python summary table (Table 5.1) with the covariate included. Looking first at the covariate statistics, it is clear that the uncorrected p-value is larger than the selected significance level (p-unc = 0.474), using a significance level of α = 0.05. Therefore, the covariate does not adjust the association between the predictor and the dependent variable. In other words, the covariate does not have a significant effect on the dependent variable when accounting for the effect of the categorical predictor. We can also observe that the uncorrected p-value for 'Race' equals 2.170098e-51, which is significantly lower than the chosen significance level of 0.05. We can reject the null hypothesis that people who get arrested for all race groups have the same adjusted average of strip search frequency after counting their mean number of actions taken during arrests. This indicates that there is a significant effect of 'Race' on the 'StripSearch_Count' after accounting for the covariate. As a result, differences in the number of strip search cases between two racial groups are statistically significant, even when controlling for the effect of the covariate.

Table 5.2*: Result of One-Way ANCOVA Test for RQ1*

Source	SS	DF	F	p-unc	np2
Race	1.312154e+05	1	231.167258	2.170098e-51	0.031876
Mean_of_Actions	2.911625e+02	1	0.512952	4.738875e-01	0.000073
Residual	3.985269e+06	7021	NaN	NaN	NaN

RQ2:

Does the presence of items found (yes vs. no) during a strip search vary based on a person's race, age, gender, type of primary offenses they committed, and the month of the search?

Logistic Regression Result

Coefficient, CI and OR

The result of the logistic regression centered around the first research question is shown in the Table x.1 below. The values under the 'P>|z|' column suggest that only Month[T.Oct-Dec] (p = 0.001), Sex[T.M] (p = 0.021), Age[T.Aged 25 to 34 years] (p = 0.019), and Age[T.Aged 35 to 44 years] (p = 0.009) are statistically significant. The table also illustrates that only Race[T.White] (coef = 0.078) and Age[T.Aged 55 to 64 years] (coef = 0.241) have positive associations with the presence of items found during a strip search (yes or no).

The following result can be inferred from the odds ratio and its 95% confidence interval table generated using the logistic regression results:

1. The intercept indicates that the average odds of discovering an item during a strip search when all of the explanatory variables are controlled is 1.759. I.e., the average odds of discovering an item for an under-18 female racial minority who is committed to the Break & Enter offense and is arrested from April to June is 1.759.

2. Compared with April to June, the odds of discovering an item during a strip search from a person are approximately 0.903 times lower in January to March, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of the month, from January to March in particular, has an effect on the outcome (item found) that is between 0.687 times less and 1.187 times more than in April to June. However, this effect is not statistically significant.

3. Compared with April to June, the odds of discovering an item during a strip search from a person are approximately 0.960 times lower in July to September, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of the month, from July to September in particular, has an effect on the outcome (item found) that is between 0.730 times less and 1261 times more than in April to June. However, this effect is not statistically significant.

4. Compared with April to June, the odds of discovering an item during a strip search from a person are approximately 0.463 times lower in October to December, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item during a strip search from a person in October to December is 0.300 to 0.717 times less likely than in April to June. Note this effect is statistically significant.

5. Compared with the visible minorities, the odds of discovering an item during a strip search from a White person are approximately 1.081 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of race, for the White group in particular, has an effect on the outcome (item found) that is between 0.856 times less and 1.365 times more than the visible minorities. However, this effect is not statistically significant.

6. Compared with females, the odds of discovering an item during a strip search from a male are approximately 0.702 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item from a male is 0.520 to 0.947 times less likely than in females. Note this effect is statistically significant.

7. Compared with people under 18, the odds of discovering an item during a strip search from a person aged 18 to 24 are approximately 0.807 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 18 to 24 in particular, has an effect on the outcome (item found) that is between 0.458 times less and 1.423 times more than age under 18. However, this effect is not statistically significant.

8. Compared with people under 18, the odds of discovering an item during a strip search from a person aged 25 to 34 are approximately 0.510 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item from a person aged 25 to 34 is 0.292 to 0.893 times less likely than those aged under 18. Note this effect is statistically significant.

9. Compared with people under 18, the odds of discovering an item during a strip search from a person aged 35 to 44 are approximately 0.466 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item from a

person aged 35 to 44 is 0.263 to 0.828 times less likely than those aged under 18. Note this effect is statistically significant.

10. Compared with people under 18, the odds of discovering an item during a strip search from a person aged 45 to 54 are approximately 0.628 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 45 to 54 in particular, has an effect on the outcome (item found) that is between 0.336 times less and 1.174 times more than age under 18. However, this effect is not statistically significant.

11. Compared with people under 18, the odds of discovering an item during a strip search from a person aged 55 to 64 are approximately 1.273 times higher, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, from 55 to 64 in particular, has an effect on the outcome (item found) that is between 0.592 times less and 2.738 times more than age under 18. However, this effect is not statistically significant.

12. Compared with people under 18, the odds of discovering an item during a strip search from a person aged over 65 are approximately 0.915 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the effect of age, over 65 in particular, has an effect on the outcome (item found) that is between 0.300 times less and 2.797 times more than age under 18. However, this effect is not statistically significant.

13. Compared with people who commit Break & Enter offenses, the odds of discovering an item during a strip search from people who commit Drug Related offenses are approximately 0.950 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item from people who commit Drug Related offenses is 0.726 to 1.242 times more likely than people who commit Break & Enter offenses. However, this effect is not statistically significant.

14. Compared with people who commit Break & Enter offenses, the odds of discovering an item during a strip search from people who commit Weapons & Homicide offenses are approximately 0.792 times lower, controlling for other features. The 95% confidence interval indicates that we are 95% confident that the odds of discovering an item from people who commit Weapons & Homicide offenses is 0.574 to 1.093 times more likely than people who commit Break & Enter offenses. However, this effect is not statistically significant.

All metrics are highlighted more clearly in Table 5.3, including Coef, SE, Lower CI, Upper CI and OR.

Table 5.3*: Result of Logistic Regression for RQ2*

	Coef	**SE**	**P-value**	**Lower CI**	**Upper CI**	**OR**
Intercept	0.5645	0.346	0.103	0.891971	3.466952	1.758528
Month[T.Jan-Mar]	-0.1022	0.139	0.464	0.686922	1.186645	0.902847

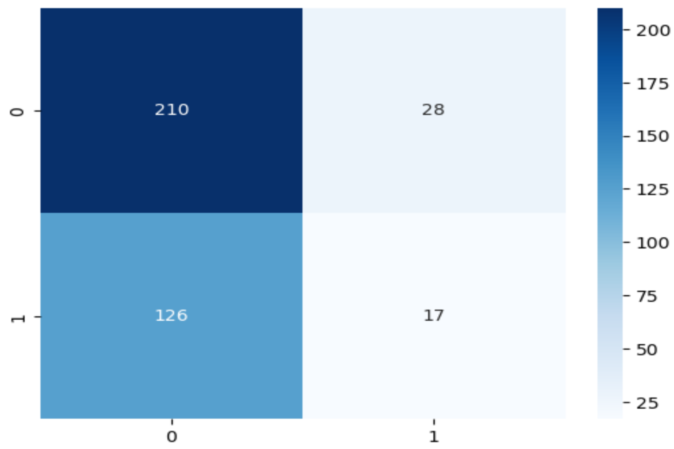
Month[T.July-Sept]	-0.0410	0.139	0.768	0.730342	1.261218	0.959788
Month[T.Oct-Dec]	-0.7690	0.222	0.001	0.299702	0.716735	0.463473
Race[T.White]	0.0780	0.119	0.512	0.856414	1.364759	1.081110
Sex[T.M]	-0.3540	0.153	0.021	0.520198	0.946926	0.701847
Age[T.Aged 18 to 24 years]	-0.2141	0.289	0.459	0.458053	1.422663	0.807252
Age[T.Aged 25 to 34 years]	-0.6727	0.286	0.019	0.291579	0.893265	0.510351
Age[T.Aged 35 to 44 years]	-0.7625	0.293	0.009	0.262799	0.828033	0.466483
Age[T.Aged 45 to 54 years]	-0.4660	0.319	0.145	0.335560	1.173543	0.627530
Age[T.Aged 55 to 64 years]	0.2411	0.391	0.537	0.591632	2.737816	1.272705
Age[T.Aged 65 years and older]	-0.0885	0.570	0.877	0.299522	2.797120	0.915314
Offence[T.Drug Related]	-0.0513	0.137	0.708	0.726214	1.242684	0.949976
Offence[T.Weapon & Homicide]	-0.2331	0.165	0.157	0.573769	1.093491	0.792093

Model Prediction and Accuracy

The test accuracy of this logistic regression model is around 0.5958 on average, indicating that the model correctly predicted the presence of items during strip searches 59.58% of the time.

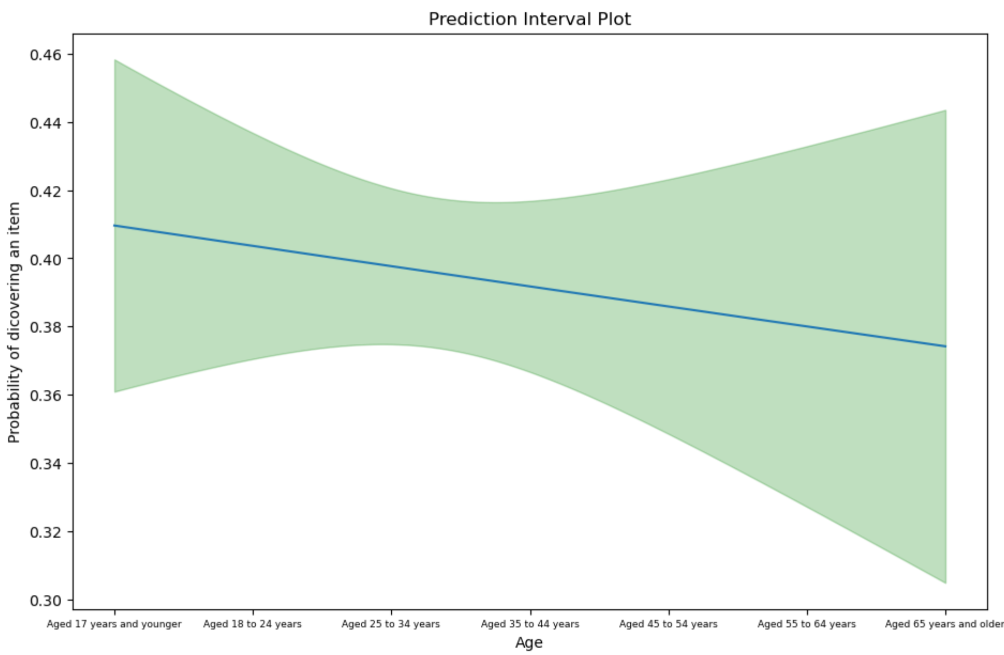
The confusion matrix (Diagram 5.2) below shows that among the 381 test data points (20% of the sub-dataset), TN is 210, TP is 17, FP is 28 and FN is 126 (Type II error). This suggests that for 143 people who actually carried an item, only 17 of them were detected. Besides, for 227 people who did not bring any item, 28 of them were detected to carry one incorrectly. That is to say, according to our logistic regression model, the probability of making a Type I (false positive) error is 11.76%, and the probability of making a Type II (false negative) error is 88.11%. These suggest that the model is inadequate as it substantially fails to identify the potential threat posed by individuals who may be carrying an item.

Diagram 5.3*: Confusion Matrix for RQ2*



Based on the prediction interval plot below, it appears that the probability of the presence of an item decreases as the age increases.

Diagram 5.4*: Prediction Interval Plot: Probability of item found vs. Age*



6. Discussion

As public members living in Toronto, people inevitably get in touch with and rely on police sectors for their daily demands, including but not limited to crime prevention, law maintenance and community policing. Ensuring the fairness of police services and preventing police brutality are able to efficiently earn public trust towards law enforcement as well as expose positive influences on the relationship between citizens and police officers. This study provides an analysis of arrest cases in the Toronto area and mainly focuses on the topic of accessing the legitimacy of local police conducting strip searches on suspects with different demographic backgrounds and situational factors. According to research, annually over 22000 strip search cases were conducted by law enforcement in Ontario, while not all of the cases met guidelines proposed by Canadian Charters, which means some of the police inappropriately conducted police searches. One of the ultimate goals of this report is to identify whether there exist biases in deciding to strip-search suspects or not.

To ensure citizens in Toronto are treated with respect and dignity in public security systems, analysis is performed on race-based strip search data to test whether Toronto police follow their proposed bias-free principles. Synthesizing results we get from t-test and one-way ANCOVA, which investigate whether there is a significant difference in the number of strip search cases between the White group and visible minorities, the statistically significant result confirms the existence of potential racial profiling phenomenon. However, this study reverses the conclusion we extracted from the literature when combined with statistics from the first logistics regression. The odds for the White racial group to be strip-searched were found to be 1.251 times higher than that of the visible minorities (People of color) all over the time, which contradicts the literature review talking about racial profiling of Black people in the Toronto area. The occurrence of this disparity might be the difference in dependent variables, the number of strip search cases and the rate of the strip search. This could be one of the limitations of this study in that only the frequency of the strip search is computed but not combined with the arrest frequency to calculate the strip search rate for two racial groups. Since population bases for each racial group living in Toronto are significantly dissimilar, and the number of arrests frequency is unknown in our analysis, more strip search cases do not stand for a higher search rate for a specific race. 43.5% of the population living in Canada are White whereas only 9.6% of it are Black in 2021. We would consider adding statistics about arrest frequency for each racial group to compute strip search rate in the next step, so we are able to better compare the test statistics.

Our analysis reveals that in addition to race, gender, age and some other factors also significantly affect the decision to conduct a strip search. Our first logistic regression model shows that females are less likely to be strip-searched, and older individuals (aged over 45) are also less likely to be searched. The odds decrease as the age increases. This aligns with the common perception that males are more likely to engage in suspicious behavior, and that strip-searching females can be uncomfortable and humiliated. According to disciplines proposed by the government, the search officers must be of the same gender as the suspect to participate in the strip search in a private area. However, there were many reports from females who were searched that Toronto police sometimes did not strictly follow the guidance, which became one of the concerns of the public about how to ensure the legitimacy of every strip search case. Generally, strip search in the Toronto area is lenient with female people compared to males.

Additionally, older individuals are generally less likely to commit offenses. We also found that the odds of being searched are significantly lower during October to December, and much higher during January to March, despite both being cold seasons. The reasons behind this phenomenon are unclear and require further investigation. One speculative reason is that the period from October to December includes several major holidays such as Halloween, Thanksgiving Days and Christmas, which may result in a decrease in search rates due to the increase in family gatherings. Police enforcement and resource allocation may be inadequate at this moment. Our second logistic regression model suggests that disadvantaged groups, such as teenagers, females, and elders, are more likely to have items discovered during a search. Individuals who commit Break & Enter offenses are also more likely to carry items, potentially due to the need for tools during such crimes. Notably, the odds of discovering an item from October to December are significantly lower, which may be due to a lower frequency of strip searches during that time period.

However, our models are not very effective as the high Type II error rates indicate that we failed to detect most individuals who should have been subjected to a strip search. This creates a dilemma, as we aim to include individuals who potentially need to be searched while maintaining the expectation of citizens that strip searches are conducted precisely and only when necessary. Thus, further analysis is necessary to improve our current models. Another concern for the above analysis is that three of the five assumptions for ANCOVA are actually violated. But we have to acknowledge that for a real-world dataset, data scientists cannot expect the distribution to be perfectly normal, and they have to accept this imperfection. A remedy for non-normal distribution is to transform the dataset using algorithms into a normal shape, but this method does not apply to every dataset. A remedy for non-normal distribution is to transform the dataset using algorithms into a normal shape, but this method does not apply to every dataset. The violation of linearity and homogeneity of the regression slope should be treated with careful consideration, for example, adding interaction terms or using nonparametric methods to improve the study.

7. Conclusion

Our study demonstrates that demographic factors, such as race, gender, and age, significantly impact the decision to conduct strip searches. The results suggest that police officers must ensure fairness in strip searches across different racial groups, as some potential suspects may avoid the procedure based on their gender and age. Additionally, situational factors such as the month, booking status, and the number of non-cooperative actions also affect the likelihood of being searched to some extent. Individuals with a higher frequency of violent actions and those who are booked are more likely to be searched, and more searches occur between April and June. Our analysis highlights the need for police officers to exercise caution when arresting teenagers and females, as there is a higher chance of carrying items, even if those items are not intended for harm. Furthermore, police officers must be vigilant when dealing with individuals who commit Break and Enter offenses, as they are more likely to carry items.

In summary, our analysis offers some useful insights into the current strip search policy in the Toronto Police Service, which can help the public better understand its workings. Additionally, the findings of our study may serve as a reference for improving policing practices in Toronto.

8. Reference

Admin, OIPRD. OIPRD Releases Report on Police Strip Searches in Ontario. 21 03 2019. <https://www.oiprd.on.ca/oiprd-releases-report-on-police-strip-searches-in-ontario/>.

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

Breaking the golden rule - office of the independent police review director. (n.d.). Retrieved April 13, 2023, from https://www.oiprd.on.ca/wp-content/uploads/OIPRD_Breaking-the-Golden-Rule_Report_Accessible.pdf

Cheng, S., Zhang, R. (2023). INF 2178 Midterm Project: Exploration of the Arrest Pattern and Performance of the Toronto Police Service [Unpublished paper]. Faculty of Information, University of Toronto

Data, denials, and confusion: The racial profiling debate in Toronto. (2003, July). Retrieved April 13, 2023, from https://utpjournals.press/doi/10.3138/cjccj.45.3.367

GeeksforGeeks. (2022, February 22). *How to perform an Ancova in python*. GeeksforGeeks. Retrieved April 13, 2023, from https://www.geeksforgeeks.org/how-to-perform-an-ancova-in-python/

Lemke, Monika. Policing Toronto: Strip Searching in a Divided City. 06 07 2022. <https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/>.

Newburn, T., Shiner, M., & Hayman, S. (2004). RACE, CRIME AND INJUSTICE? Strip Search and the Treatment of Suspects in Custody. The British Journal of Criminology, 44(5), 677–694. http://www.jstor.org/stable/23639161

One-way Ancova in SPSS statistics. How to perform a one-way ANCOVA in SPSS Statistics | Laerd Statistics. (n.d.). Retrieved April 13, 2023, from https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics.php

Race and Identity-based Data Collection Strategy. 15 June 2022. <https://data.torontopolice.on.ca/pages/race-based-data>.

R. v. Golden. No. 27547. Supreme Court. 06 12 2001.

Understanding analysis of covariance (ANCOVA) - Northern Arizona University. (n.d.). Retrieved April 13, 2023, from https://oak.ucc.nau.edu/rh232/courses/EPS625/Handouts/ANCOVA/Understanding%20ANCOVA.pdf

Use and interpret ANCOVA in SPSS. Accredited Professional Statistician For Hire. (n.d.). Retrieved April 13, 2023, from https://www.scalestatistics.com/ancova.html